



# Cartography M.Sc.

## Master thesis

### Visualizing Mobile Network Data

A User-Centred Design Approach  
Connecting Visual Analytics to  
Urban Public Transportation Planning

Laura Verena Klasen



2020



# Visualizing Mobile Network Data

A User-Centred Design Approach Connecting  
Visual Analytics to Urban Public Transportation Planning

submitted for the academic degree of Master of Science (M.Sc.)  
conducted at the Department of Aerospace and Geodesy  
Technical University of Munich

Author: Laura Verena Klasen  
Study course: Cartography M.Sc.  
Supervisor: Dr.-Ing. Mathias Jahnke (TU Munich)  
Reviewer: MSc. Wangshu Wang (TU Vienna)  
Cooperation: Dr. Georg Fuchs – external Supervisor  
Fabian Patterson – Advisor  
Fraunhofer IAIS

Chair of the Thesis  
Assessment Board: Prof. Dr. Liqiu Meng

Date of submission: 10.10.2020

# Statement of Authorship

I hereby declare that I am the sole author of the submitted master's thesis entitled:

## **Visualizing Mobile Network Data**

**A User-Centred Design Approach Connecting  
Visual Analytics to Urban Public Transportation Planning**

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 10.10.2020

Laura Verena Klasen

## Acknowledgements

I would like to thank my supervisor DR. MATHIAS JAHNKE at TUM, my external supervisor DR. GEORG FUCHS and advisor FABIAN PATTERSON at Fraunhofer IAIS for their support and helpful comments throughout this semester.

Thank you as well to all professionals who took the time to participate in my interviews and user tests. Your feedback was valuable for this thesis and made it possible to connect two disciplines.

Thank you JULIANE CRON and everyone else who goes out of their way to make this unique Master's Programme possible.

Last but most importantly: thank you from the bottom of my heart to my wonderful friends and family who fill my life with love and laughter.

## Abstract

Urban regions are becoming increasingly crowded, which places heavy demands on public transportation networks. Movement data stemming from mobile network data is becoming cheap and easy to obtain and much research has been done within the field of Visual Analytics on how to best visualize these data for analytic tasks. Yet, collaboration with the public transportation sector has been lacking.

This work addresses that research gap by developing an interactive visualization concept for attributed flow data stemming from mobile network data to be used by professionals in the field of urban public transportation planning. A user-centred design approach is taken, starting with expert user interviews and resulting in an evaluation of the developed prototype. The design of the visualizations is based on an exploration of the historical development of visualizations of temporal data as well as recent developments within Visual Analytics.

The results show that the incorporation of attributed flow data stemming from mobile network data can help improve several issues of the current framework used for public transportation planning and would therefore be of great value. However, as only the first iteration of a user-centred design cycle could be realized within the time allocated for this research, several improvements need to be incorporated and evaluated in further cycles in order for a well working product to be developed.

**Keywords:** *Visual Analytics, Urban Public Transportation Planning, User-Centred Design, Extended Mobile Network Data, Spatiotemporal Data*

# Contents

- List of Figures iii
- List of Tables v
- 1 Introduction 1**
  - 1.1 Motivation . . . . . 1
  - 1.2 Research Identification . . . . . 2
  - 1.3 Thesis Structure . . . . . 3
- 2 Scientific Background 4**
  - 2.1 Visual Analytics of Spatio-Temporal Data . . . . . 4
  - 2.2 Visualization of Temporal Data Through Time . . . . . 5
  - 2.3 User-Centred Design & Usefulness Evaluation . . . . . 13
  - 2.4 Transport Planning: Related Work and Research on the Study Area . . . . . 14
- 3 Methodology 15**
- 4 Case Study & Results 17**
  - 4.1 Context . . . . . 17
  - 4.2 Current Analysis Practices in Urban Public Transportation Planning . . . . . 19
    - 4.2.1 Expert User Interview Preparation . . . . . 19
    - 4.2.2 Results of the Expert User Interviews . . . . . 19
  - 4.3 Data Exploration and Visualization . . . . . 24
    - 4.3.1 Visualization Concept . . . . . 24
    - 4.3.2 Data Handling . . . . . 33
    - 4.3.3 Data Visualization . . . . . 36
  - 4.4 User Testing . . . . . 43
    - 4.4.1 The Test Set-Up . . . . . 43
    - 4.4.2 Results of the Evaluation . . . . . 46
  - 4.5 Implementation of Test Results . . . . . 48
    - 4.5.1 Changes in the Notebooks . . . . . 49
    - 4.5.2 Further Possible Changes . . . . . 50
- 5 Discussion 52**
  - 5.1 Results with Regard to the Scientific Background . . . . . 52
  - 5.2 Limitations of the Proposed Visualization Concept . . . . . 53
- 6 Conclusion & Outlook 54**
- Bibliography 56**
- Appendix 57**
  - Interview Transcripts . . . . . 58
    - Interview I . . . . . 58
    - Interview II . . . . . 69
    - Interview III . . . . . 78
  - Structured Collection of Statements from the Interviews . . . . . 94

Screenshots from Interviews . . . . .	98
Data Tables . . . . .	104
User Test Transcripts . . . . .	110
User Test I . . . . .	110
User Test II . . . . .	138
Overview of the Visualizations used for the User Tests . . . . .	165
Questionnaire . . . . .	166
Questionnaire Results . . . . .	168

# List of Figures

- 2.1 Biographical chart by Joseph Priestley and a modern adaptation . . . . . 6
- 2.2 Details from Charles-Joseph Minard's flow maps depicting rail traffic. . . . . 7
- 2.3 Flow map series by Minard depicting cotton imports . . . . . 8
- 2.4 Detail from flow maps by Minard depicting transportation on water and railways . 8
- 2.5 Rose Diagram by William Farr on Temperature and Mortality of London . . . . . 9
- 2.6 Rose diagram on mortality causes by Florence Nightingale . . . . . 9
- 2.7 Coxcomb diagrams on Covid-19 cases by Kenneth Field . . . . . 10
- 2.8 Economic chart by William Playfair . . . . . 11
- 2.9 Polish-American System of Chronology by Elizabeth Palmer Peabody . . . . . 12
- 2.10 Plot of a sound recording by Yvonne F. Phillips . . . . . 12
- 2.11 Prototyping schema . . . . . 13
  
- 3.1 Workflow . . . . . 15
  
- 4.1 Dataflow . . . . . 18
- 4.2 Load carpet visualization as used by planners . . . . . 20
- 4.3 Daily graph visualization as used by planners . . . . . 21
- 4.4 Dashboard interface as used by planners . . . . . 22
- 4.5 Manually generated maps as used by planners . . . . . 23
- 4.6 Basic user interface concept . . . . . 25
- 4.7 Overview of visualizations: first stage . . . . . 27
- 4.8 Overview of visualizations: second stage . . . . . 28
- 4.9 Overview of visualizations: third stage . . . . . 29
- 4.10 Colour scheme . . . . . 29
- 4.11 Overview of visualizations: fourth stage . . . . . 30
- 4.12 Movement patterns on two different days in radial and horizontal bar charts . . . . 31
- 4.13 Triangular representation of a sample of the OD dataset . . . . . 32
- 4.14 Flow maps; first stage . . . . . 35
- 4.15 Heatmaps; first stage . . . . . 36
- 4.16 Coxcomb diagrams of the whole dataset; first stage . . . . . 36
- 4.17 Radial and horizontal bar charts showing flow data; first stage . . . . . 37
- 4.18 Flow map of anomalies; second stage . . . . . 38
- 4.19 Bar charts showing anomalies; second stage . . . . . 39
- 4.20 Flow maps showing OD data; third stage . . . . . 40
- 4.21 Radial and horizontal bar charts for OD data; third stage . . . . . 41
- 4.22 Flow map showing anomalies in OD data; fourth stage . . . . . 42
- 4.23 Bar charts showing anomalies in the OD dataset; fourth stage . . . . . 43
- 4.24 Interface used for user testing . . . . . 45
- 4.25 Overview of visualizations used for user testing . . . . . 45
- 4.26 Heatmaps with changes . . . . . 48
- 4.27 Radial bar charts with changes . . . . . 49
  
- A.1 Load carpet . . . . . 98
- A.2 Daily graph . . . . . 99
- A.3 Commuter flows Landsberg . . . . . 99
- A.4 Commuter flows . . . . . 100



A.5	Dashboard interface by Teralytics . . . . .	101
A.6	Dashboard interface by Teralytics . . . . .	102
A.7	Dashboard interface by Teralytics . . . . .	103
A.8	Answers to the first part of the questionnaire . . . . .	171

# List of Tables

- 4.1 Modes of Transportation . . . . . 18
- 4.2 Stages of the User Test . . . . . 44
  
- A.1 Preprocessed flow data . . . . . 104
- A.2 All flow data in one data frame . . . . . 105
- A.3 All unique cells for plotting on maps . . . . . 105
- A.4 Aggregated flows for plotting on maps . . . . . 105
- A.5 Aggregated flow anomalies for plotting on maps . . . . . 106
- A.6 Aggregated flows for plotting on bar charts . . . . . 106
- A.7 Aggregated flow anomalies for plotting on bar charts . . . . . 107
- A.8 Preprocessed OD data . . . . . 107
- A.9 All OD data merged with cell polygons in one data frame . . . . . 108
- A.10 All OD flows for plotting on maps and bar charts . . . . . 108
- A.11 All OD anomalies for plotting on maps and bar charts . . . . . 109

# 1 Introduction

## 1.1 Motivation

The temporal relation of events to each other has always been a major concern of humanity. For culture to form, it is vital to have an understanding about what happened where and when. By passing on knowledge to future generations, we allow for a continuous progress of understanding and invention to happen. Without knowing about what has been, we cannot make sense of the now and have a hard time shaping the future.

Our understanding of time is largely influenced by our understanding of events as effects of a cause. This concept of causality is closely linked to the linearity we experience to be inherent in time. It is therefore not surprising that the most common and intuitive graphical element we use to visualize temporality is a line. We draw a line from the past to the future with a dot for the present moment, we curve a line on a map to show the way we went while time was passing, we close the line to a circle to illustrate the cyclic character of an hour, a day, a week or a year.

While the timeline in all its different shapes is ubiquitous today and engraved in the language we use to talk about time, the history of the development of the timeline has also depended on our ability to measure time. The invention and development of increasingly precise time measurement has led to the possibility to zoom into minute intervals of time and explore what happens at tiny scales. It also enables us to synchronize our clocks across the globe and communicate with each other, locate ourselves on the surface of the earth and organize and coordinate our lives around time in a way that was far from reality in the past.

Before train travel made the need for synchronized clocks apparent, time measurement was a local phenomenon, building on the sun's path across the sky. This path can be directly projected onto the surface of sun dials that reveal the cyclic course of a day and a year. Time was, and essentially still is, dictated by Earth's rotation and its orbit around the sun.

Studying the course of time used to lie within the domain of *chronology*, but temporal changes in space have since become a subject of further fields of study, inspired by rapid environmental and cultural changes including a revolutionary view from physics on the relation between time and space. In fact, already half a century before the theories of ALBERT EINSTEIN reshaped the way we think about time on a fundamental level, EMMA WILLARD understood that how we relate to events in time on a more superficial level is, or should be, similar to how we relate to space. Beneath her *Temple of Time* (1846), she wrote:

“The attempt to understand chronology by merely committing dates to memory, is not only painful, but it is as useless as to learn latitudes and longitudes, without the study of maps. As in geography, the relation of any place to all other places is what is important to know; so in chronology, the relation which any given event bears to others constitutes the only useful knowledge. [...] By putting the course of time into perspective, the disconnected parts of a vast subject are united into one, and comprehended at a glance; – the poetic idea of “the vista of departed years” is made an object of sight; and when the eye is the medium, the picture will, by frequent inspection, be formed within, and forever remain, wrought into the living texture of the mind.”  
– Emma Willard (1846)

Fast forward to our digital age of big data, the study how we interact with and analyse data visually, has developed into the science of *Visual Analytics*, with strong ties to other data- and

geosciences such as *Cartography*. As everything tends to happen *somewhere* and *sometime*, most of the data we collect has a *spatiotemporal* component. Studying ways to effectively visualize spatiotemporal data for interactive exploration and analysis is therefore of major importance. One example of spatiotemporal data is mobile network data. How information on human movement stemming from such data can be visualized for the use within urban public transportation planning will be the topic of this study.

## 1.2 Research Identification

### Research Problem and Context

As urban regions are becoming increasingly crowded and public transportation is indispensable, the analysis of temporal variations of the use of public transportation becomes vital to effective planning. With the omnipresence of smartphones, movement data is now rather easy to obtain and there has been significant research within the field of Visual Analytics concerning its analytic visualization. However, collaboration with the transportation sector has been lacking (Andrienko, Andrienko, Chen, Maciejewski & Zhao, 2017) which is why this work adopts an interdisciplinary approach led by the principle of user-centred design. While the main scientific context is the field of Visual Analytics, the context of urban public transportation planning is taken into consideration as well, as it constitutes the professional field of work of the intended users. Since the focus is on the temporal component of the visualized data, a wider historical perspective on the visualization of time and temporal data is taken to guide the development of these visualizations.

### Research Objectives and Research Questions

The aim of this research is to develop and evaluate an interactive visualization concept for attributed time-variable flow data in the context of urban public transportation planning. The data used is provided by Fraunhofer IAIS as this work is part of an ongoing project on *Extended Mobile Network Data (xMND)* concerned with making mobile network data accessible and usable for public transportation planning. Within the scope of this project, this work is located at the end of the data-pipeline as it deals with the usability of the visualizations of the data. Planners and decision makers working at the public transportation associations in Munich were contacted for interviews and user tests to ensure a user-centred design approach.

The research is divided into the two interrelated research objectives (1. and 2.) which are further structured by four sub-objectives (1.1, 1.2, 2.1 and 2.2). Each of the sub-objectives is tackled by answering the corresponding research questions (a) to (k). The research objectives and questions are as follows.

1. Exploration of different ways of visualizing spatio-temporal data for planning purposes
  - 1.1 Context assessment
    - (a) How have time and temporal data been visualized historically?
    - (b) What is the state-of-the-art of visualizing movement-data in Visual Analytics?
  - 1.2 Exploration and Visualization of time-variable attributed flow graph data
    - (c) How is the data at hand structured?
    - (d) What kind of visualizations are applicable in the specific case?
    - (e) How can those visualizations be generated, taking the intended level of interactivity into account?
2. Evaluation of the usefulness of a visualization concept for flow data and OD data for urban public transportation planning

- 2.1 Assessment of current analysis practices in urban public transportation planning
  - (f) How are the intended users presently working?
  - (g) Which visualizations, if any, are they used to?
  - (h) How well does the current framework align with the users' needs and analytical questions?
- 2.2 User-test of a visualization concept for flow data and OD data for the evaluation of its usefulness within urban public transportation planning
  - (i) Which methods for testing usefulness are appropriate in this specific case?
  - (j) How good is the usefulness of the visualizations?
  - (k) How can the visualizations be improved to possibly increase usefulness?

The scope of this research is limited by the available data and expert users as well as the allocated time. Hence, it does not attempt to compare the usefulness of different data sets or ways of generating these. Neither is the aim to develop a ready-to-use tool for decision makers within urban public transportation planning but rather to contribute to making such a development possible.

## Hypothesis

The Hypothesis guiding this research is that extended mobile network data in the form of attributed flow data and origin destination data can be useful for transportation planning if a user centred visualization concept is applied.

## 1.3 Thesis Structure

The present work is structured within 6 chapters. The *Introduction* outlines the motivation, identification and structure of the thesis. For the *Scientific Background*, a literature review on the topics of *Visual Analytics*, the *Visualization of Temporal Data through Time*, *User-Centred Design & Usefulness Evaluation* and *Transport Planning* are explored, with a focus on the study area of the examined case study and the types of visualizations applied within it. This chapter answers research objective 1.1. The third chapter describes the *Methodology* used to meet the research objectives laid out in the introduction. It is organised in two nested cycles, corresponding to the two main research objectives. Then, the *Case Study* is presented, alongside the *Results*. For the case study, the *Context* of the thesis is provided and, subsequently, the remaining research objectives are answered. Results presented within the fourth chapter are discussed in the fifth chapter. Finally, a conclusion is drawn and an outlook is given, relating the results to a wider viewpoint.

## 2 Scientific Background

### 2.1 Visual Analytics of Spatio-Temporal Data

Visual Analytics can be defined as

“an iterative process that involves information gathering, data preprocessing, knowledge representation, interaction and decision making. The ultimate goal is to gain insight in the problem at hand which is described by vast amounts of scientific, forensic or business data from heterogeneous sources. To reach this goal, visual analytics combines the strengths of machines with those of humans.”

(Keim, Mansmann, Schneidewind, Thomas & Ziegler, 2008)

Visual Analytics has its roots in many different fields, including Data Analysis, Scientific Visualization, Information Visualization, Visual Data Mining, Geovisualization and GIScience (Andrienko et al., 2010; Cui, 2019). Out of these many roots, Geosciences have an especially long history in displaying data visually, that is, in the form of cartographic maps. When concerned with movement data, the spatial dimensions displayed on a map have to be complemented with a temporal one. As there is no intrinsic visual representation of time, different approaches have been made to facilitate interactive visualization with spatio-temporal data which, of course, build on the historic approaches discussed above.

The availability of interfaces in the form of digital displays has made dynamic three-dimensional visualizations possible. These can of course be used to display objects with a spatial extent but they can also facilitate the display of movement data. As the relevant spatial components of movement on Earth can often be confined to two dimensions, the third spatial dimension being fixed by Earth’s surface, time and space can be effectively visualized using a total of three dimensions. The result is a spacetime-cube as described in *Visual Analytics of Movement* (Andrienko, Andrienko, Bak, Keim & Wrobel, 2013, p. 4).

However, visualizations that make use of three dimensions but are physically confined to a two-dimensional display always face the problem of occlusion. The visibility of different parts of larger datasets depends on the rendered size of other parts of the dataset and, thus, intense interaction or supplementary visualizations are necessary to acquire a correct understanding of the data (Andrienko et al., 2013, p. 127).

To enable the visual exploration and analysis of attributed flow data, a cartographic map along with visually linked time graphs or temporal bar charts is useful. Using maps to display spatial data for analysis purposes is essential since they offer substantial contextual information and put different parts of the dataset into spatial relation to each other. A map showing aggregated movements, i.e. flows, is called a *flow map*. While maps facilitate the acquisition of a spatial overview, time graphs or temporal bar charts offer the complementary temporal overview needed to obtain a more complete understanding of a spatio-temporal dataset. Multiple visualizations showing different aspects of the same dataset are linked by ensuring visual consistency, e.g. using the same colours in all visualizations to represent the same categories. Within a full functioning user-interface, this linkage is enforced by consistent and simultaneous reactions of all visualizations to user interactions such as the changing of filter options. Interactive filter options furthermore help reduce visual clutter and occlusion on the map and enable the user to relate different components of the data to each other. (Andrienko et al., 2013, chapter 4)

## 2.2 Visualization of Temporal Data Through Time

There are many ways of visualizing spatio-temporal data today. To understand where and when they originated, this section will give an overview over their historical development, focusing on visualizations that were used for the case study.

In *Cartographies of Time*, Rosenberg and Grafton (2012) illustrate the history of the timeline which, in Europe, was for a long time driven by the attempt to make sense of the story of creation, as described in the Bible. Humans wanted to understand the process of creation and relate themselves to it. With the resurfacing of sciences other than theology in the 16<sup>th</sup> century, chronologers and astronomers found that astronomical measurements, which dated back centuries, could help align the historical accounts from different cultures to a continuous string of time, dictated by universal astronomical events such as eclipses. By doing this, it became apparent that the creation, as described in the Bible, was after all not the first thing that had happened on Earth. Chinese records of eclipses dating back to ancient times confirmed the existence of civilizations long before the events recorded in the Bible (Rosenberg & Grafton, 2012, p. 69).

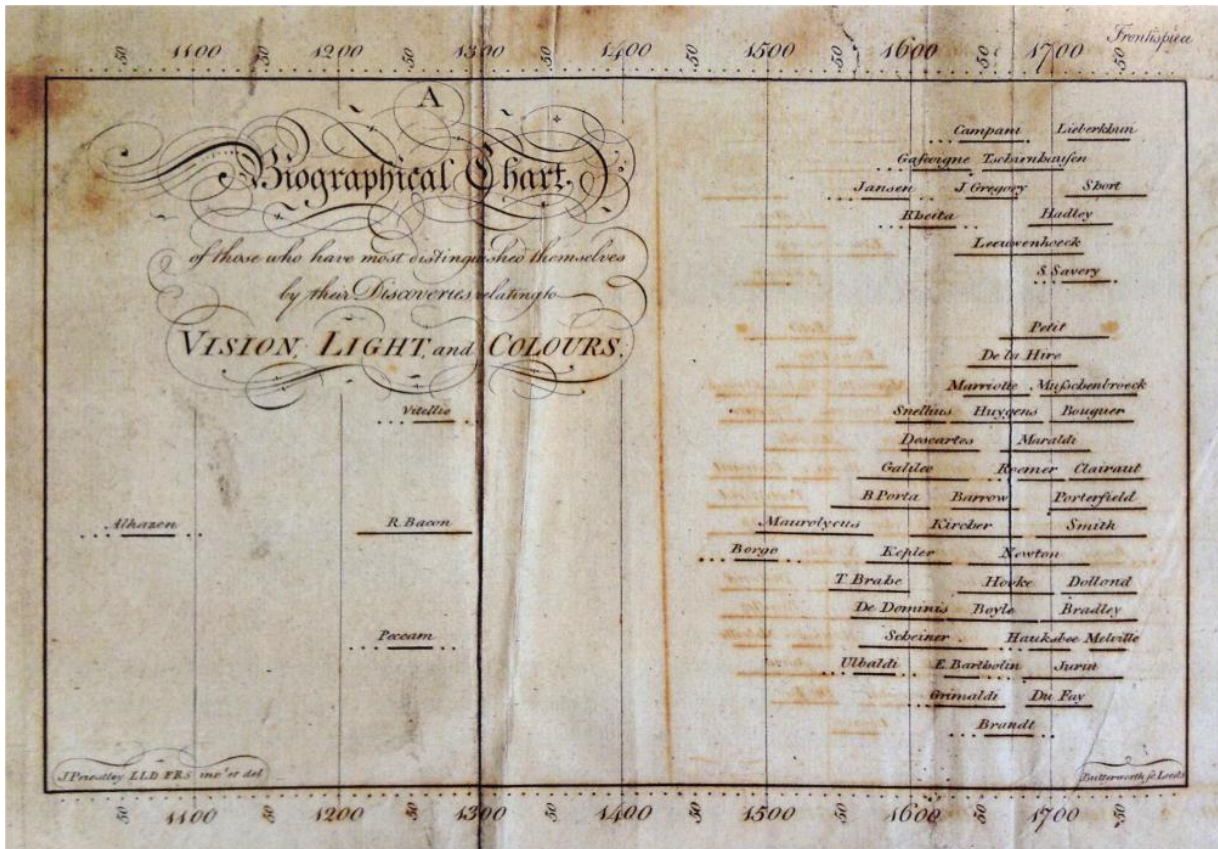
One of the first people to graphically connect astronomical data to events in time was GERARDUS MERCATOR, the Cartographer who coined a projection invented for navigation which is now used with maps all across the internet. Criticized for the severe distortion it imposes on the size of different landshapes, its adequacy is subject to many discussions (for further details see Battersby, Finn, Usery & Yamamoto, 2014). In his 1569 *Chronology*, Mercator introduced a dynamic timescale as he changed the temporal resolution depending on the amount of detail in available information while staying consistent by always relating time to the sun's position. (Rosenberg & Grafton, 2012, p. 63 and following)

Before graphic inventions such as Mercator's could emerge and take root with the help of printing techniques, the representation of temporal data was constrained to a tabular layout containing written information. To this day, tables seem to be the most efficient way to conserve data. The person who established the tabular layout in the field of Chronology was EUSEBIUS OF CAESAREA, who lived during the 4<sup>th</sup> century. In his *Chronicle* he laid out temporal information from different sources in separate columns which facilitated tasks that involved looking up and comparing different pieces of information (Rosenberg & Grafton, 2012, p. 15-16).

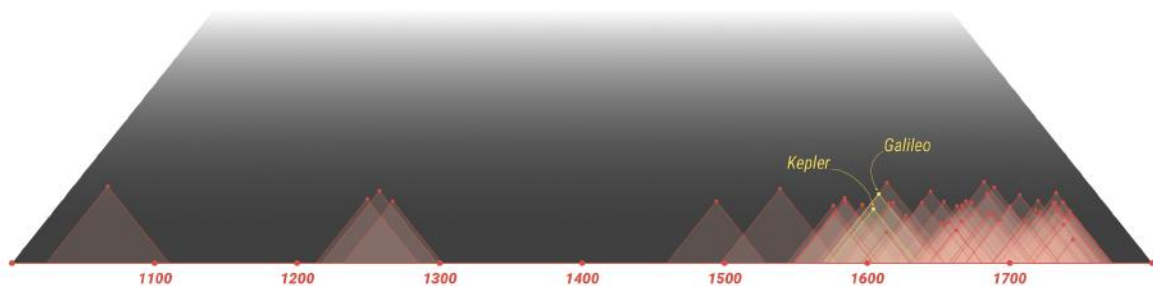
However, it took another two centuries after Mercator before JOSEPH PRIESTLEY would use graphic variables not only as a scale but to actually symbolize datapoints according to their temporal dimension. In 1765, he created *A Chart of Biography* that depicted lines representing the lives of influential people (for a reduced version see figure 2.1a). The lines, labelled with the respective names, start at the person's birth and end at their death, aligned to a horizontal scale. When specific dates were uncertain, Priestley used dots to indicate this uncertainty. (Rosenberg & Grafton, 2012, p. 117 and following)

Yet another century after Priestley, with the world becoming increasingly interconnected, CHARLES JOSEPH MINARD shaped the technique of representing movement of quantities in *flow maps*. His work, explored in detail by Sandra Rendgen in *The Minard System* (2018), was fuelled by the rise of statistics in science and a new abundance of statistical data (Rendgen, 2018, p. 8). Apart from his famous 1869 *Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812–1813* (Rendgen, 2018, p. 154-157), he repeatedly visualized more ordinary types of movement such as *Travelers Passing Through Grenoble Junction* in 1846 and *European Railroad Travel in 1862* (see figure 2.2 for details).

Today, flow maps are used widely but algorithms that produce well-readable maps are still being developed. The difficulty lies in drawing lines automatically while preventing occlusion (Phan, Xiao, Yeh & Hanrahan, 2005). Jenny et al. (2018) identified and tested design principles for manually created maps in a user study. Carefully applying emerging technologies like virtual and aug-



(a)



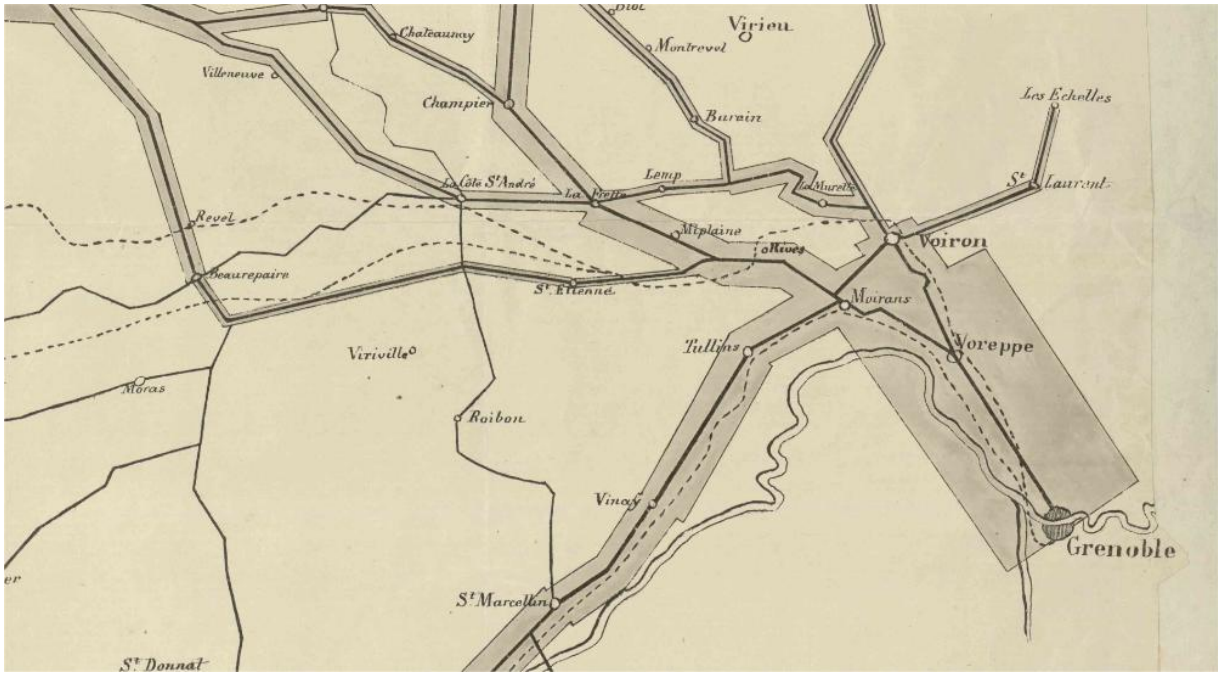
(b)

Figure 2.1: (a): *A Biographical Chart* by Priestley (1772), figure taken from Rosenberg and Grafton (2012, p. 18), similar to Priestley's bigger version titled *A Chart of Biography*. (b): The same data translated to a diagrammatic representation of interval space as described by Kulpa (1997). It shows the lifetimes of scientists from the area of optics between the 11<sup>th</sup> and 18<sup>th</sup> century.

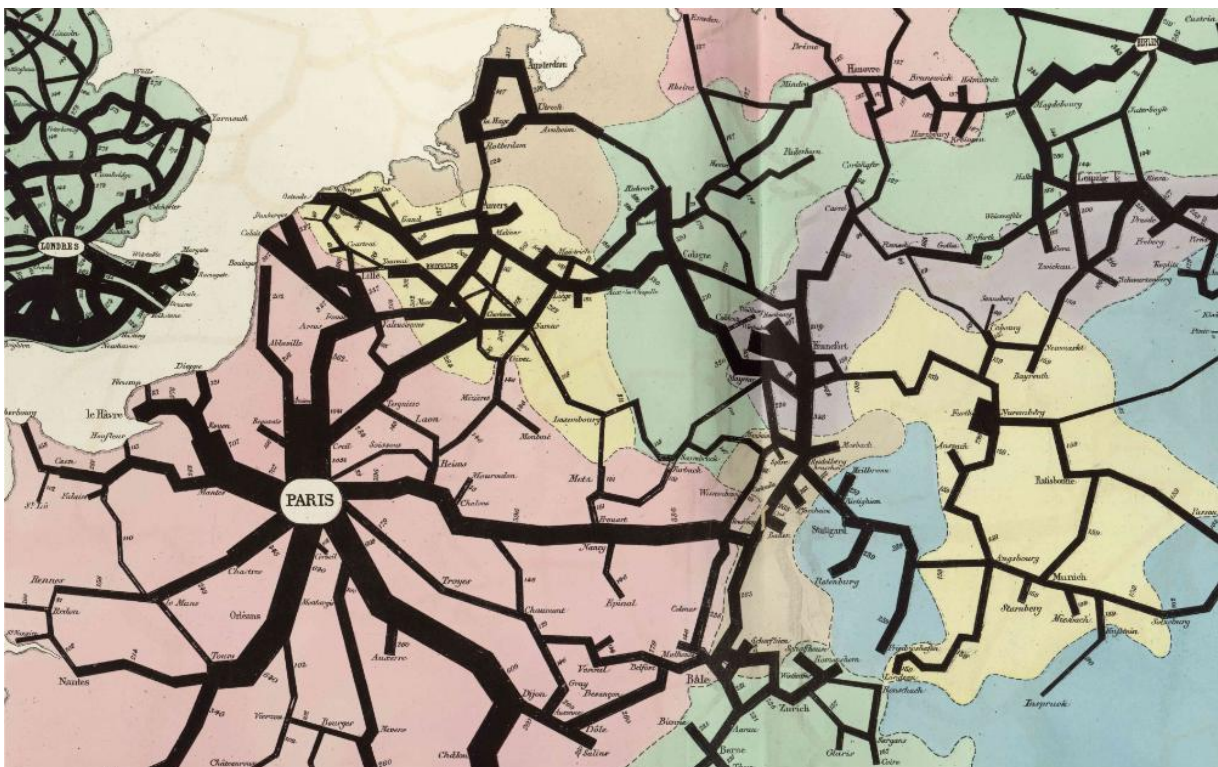
mented reality to flow maps can help reduce visual clutter and thus facilitate more effective map reading (Yang et al., 2019).

The essential visual innovation present in Minard's maps is the direct depiction of quantified movement on a map. It proves so effective because every visual variable used directly represents data: trajectories are represented by lines, while the widths of the different lines depict the quantity associated with the respective trajectory. To incorporate further information, the lines can be coloured according to a specific quality of the moving quantity. In a series of maps showing *European Cotton Imports* between 1858 and 1865 (see figure 2.3), Minard used colour to indicate the origin of each trajectory. Colour can also indicate the type of transport, as in Minard's series of maps depicting the *Circulation of Goods on French Railroads and Waterways* between 1850 and 1862 (see figure 2.4 for a detail). Both of these series of maps furthermore facilitate an analysis of the temporal development of the movement: The individual maps are easily compared and rough





(a)



(b)

Figure 2.2: Details from Minard's flow maps depicting (a) *Travelers Passing Through Grenoble Junction* (1846) and (b) *European Railroad Travel in 1862* (1865). The line width shows the number of passengers travelling along the corresponding connection. (Minard, 1870)<sup>1</sup>

changes in quantity and spatial distribution can be detected intuitively, without comparing numbers in a table (see figure 2.3).

Another effective use of basic visual variables can be observed in the depiction of intervals

<sup>1</sup>[https://patrimoine.enpc.fr/document/ENPC01\\_Fol\\_10975](https://patrimoine.enpc.fr/document/ENPC01_Fol_10975), accessed 2020/09/18

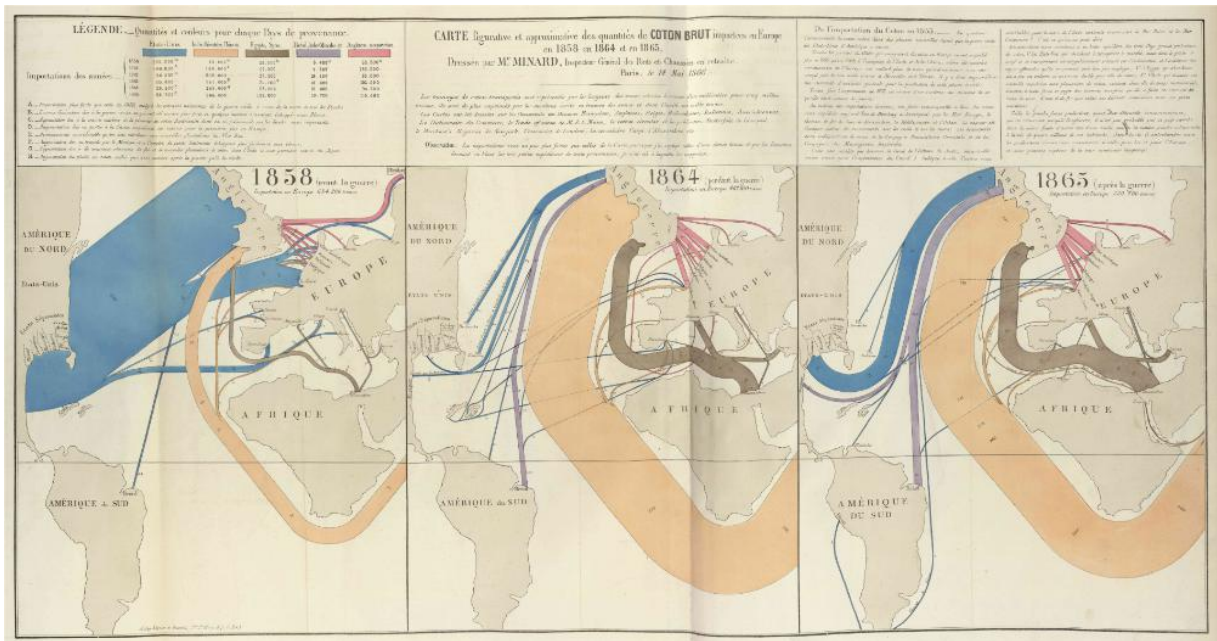


Figure 2.3: Parts of a series of flow maps by Minard depicting *European Cotton Imports between 1858 and 1865* (1866). Colour is used to indicate the region of origin, line width corresponds to the imported quantity. (Minard, 1870)<sup>2</sup>

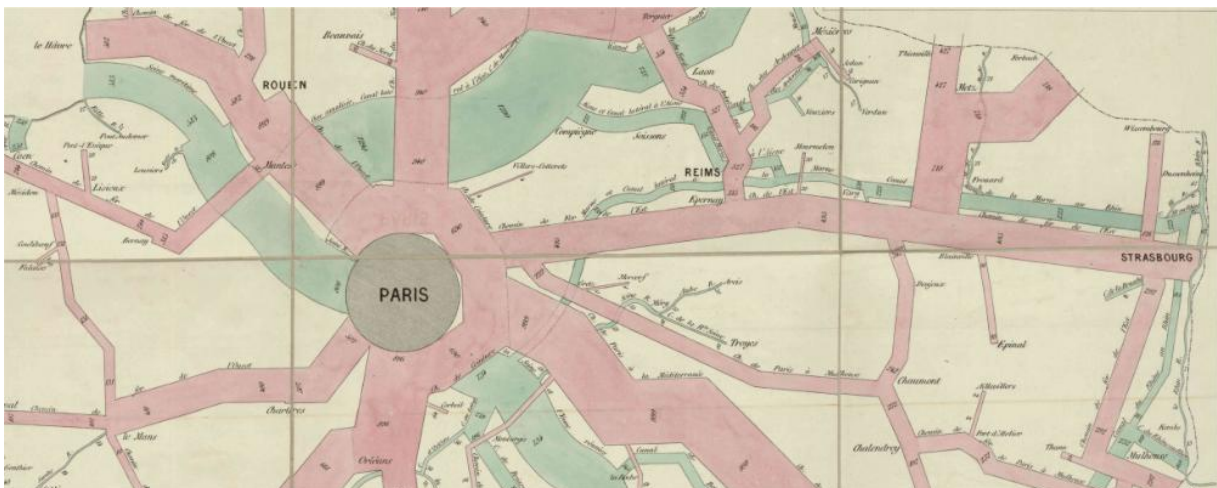


Figure 2.4: Detail from a series of flow maps by Minard depicting the *Circulation of Goods on French Railroads and Waterways between 1850 and 1862*. Here, colour is used to indicate the mode of transportation: greenish blue for waterways and red for railroads. (Minard, 1863)<sup>3</sup>

as points in a two-dimensional space. Any interval in time (or, more generally, any interval in a scalar variable) is characterized by two numbers, the start and the end point. While intervals are commonly represented by a line connecting those two points, as done by Priestley, this only makes use of one dimension within the chart, leaving the other one arbitrary. Priestley uses the vertical axis in *A Chart of Biography* to sort his entries categorically, which works fine but still leaves the continuity of a scalar axis unused and implies a ranking between and within the categories which might be undesirable. Making use of the diagrammatic representation of interval space, as described by Kulpa (1997), the same information can be presented using points instead of lines (see figure 2.1b). The length of intervals can be compared via the vertical position of the points and overlapping of intervals can be detected at a glance. Of course, compared to the

<sup>2</sup>[https://patrimoine.enpc.fr/document/ENPC01\\_Fol\\_10975](https://patrimoine.enpc.fr/document/ENPC01_Fol_10975), accessed 2020/09/18

<sup>3</sup>[https://patrimoine.enpc.fr/document/ENPC01\\_4\\_10975](https://patrimoine.enpc.fr/document/ENPC01_4_10975), accessed 2020/09/18

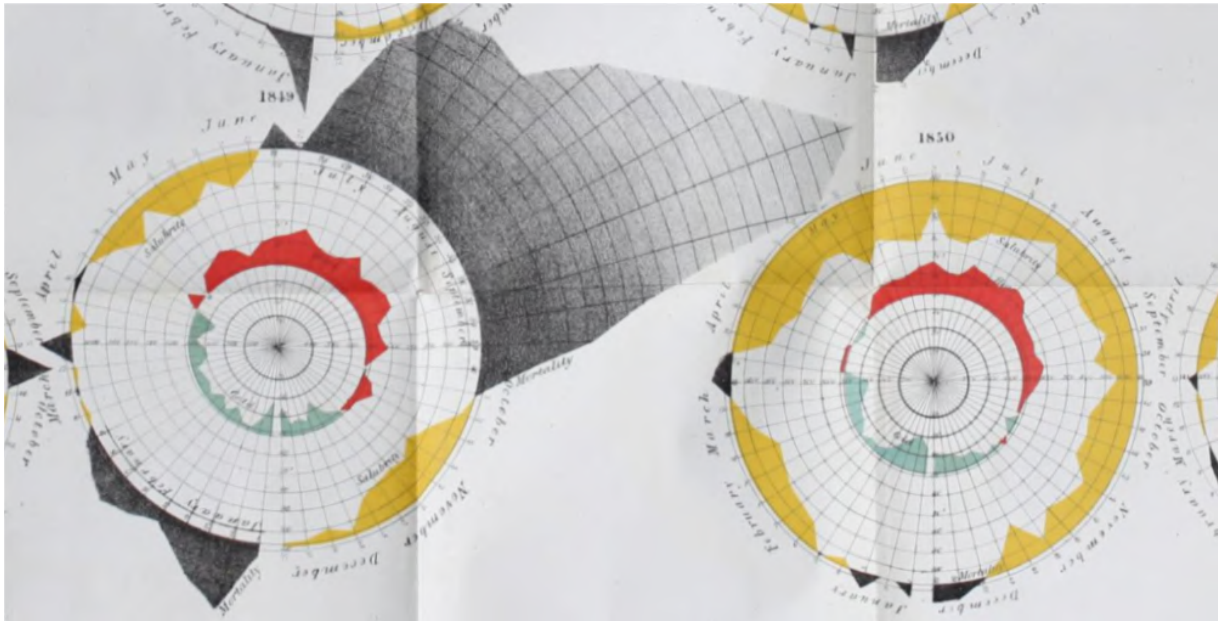


Figure 2.5: Rose Diagram by William Farr on *Temperature and Mortality of London* between 1840 and 1850. (Great Britain. General Register Office et al., 1852)<sup>4</sup>

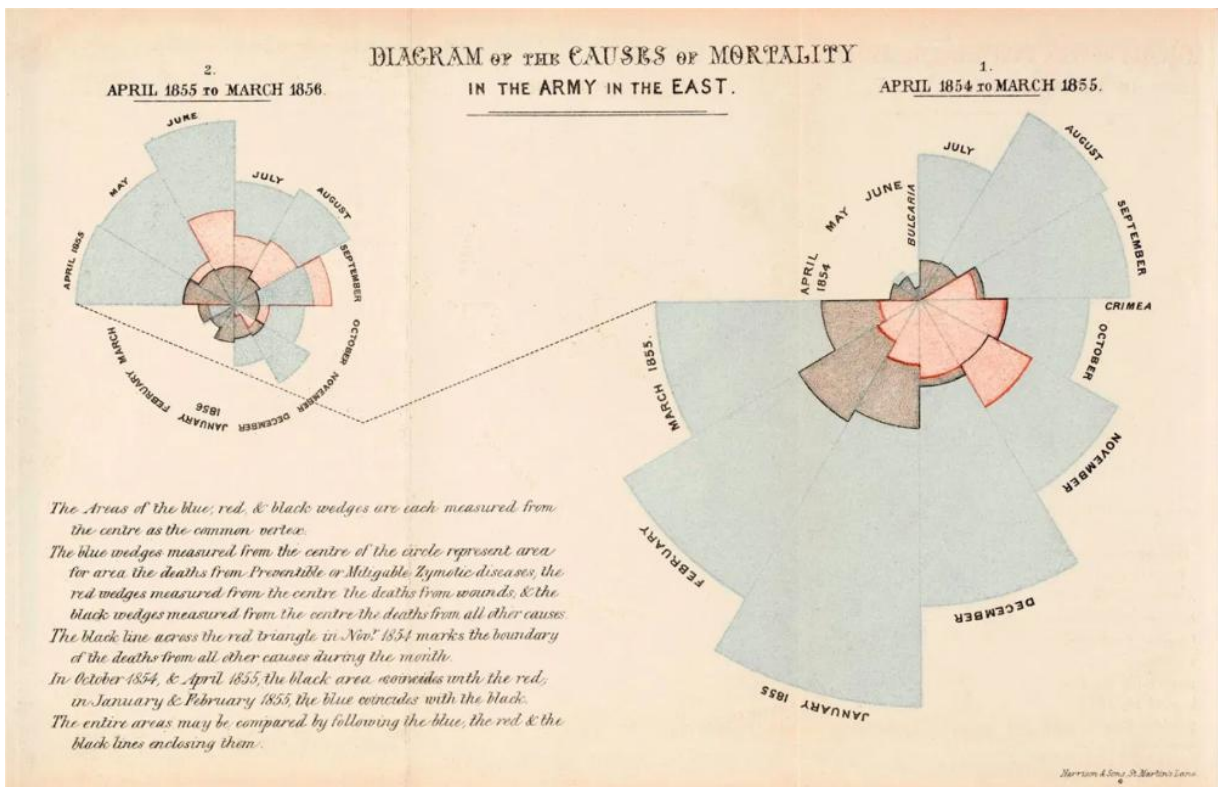


Figure 2.6: *Diagram of the Causes of Mortality in the Army in the East* between 1854 and 1856 by Florence Nightingale (1857). Nightingale used this visualization to promote her sanitation strategy during the Crimean War. (Popova, 2017)<sup>5</sup>

original chart, interactivity is needed to handle any visual clutter introduced by the direct projection of the data onto a two-dimensional space. The uncertainty that was visualized in the original chart can be encoded as well, as described in *Modelling imperfect time intervals in a two-dimensional*

<sup>4</sup><http://archive.org/details/b21308251>, accessed 2020/09/18

<sup>5</sup><https://www.brainpickings.org/2017/10/09/w-e-b-du-bois-diagrams>

space (Qiang, Delafontaine, Asmussen, Stichelbaut & de Weghe, 2010). While this representation appears to be rather new and not very well established in our everyday visual language for time, the connection between intervals and triangles had already been made in the 16<sup>th</sup> century by PAULUS CRUSIUS, who, a decade after Mercator had published *Chronology*, fit temporal information in terms of intervals into a triangular table (Rosenberg & Grafton, 2012, p. 65).

What about recurring intervals of time? As the cyclic motion of Earth, both its own rotation as well as its orbit around the sun, seems to be responsible for our perception of time, one obvious consequence is that we tend to organize time in a cyclic manner. Hours, days, weeks and years are structures we have built upon Earth’s orbital motions in space and our need for sleep, to be able to organize our lives with an overview of time in mind. It makes sense, then, to also display temporal data with an underlying or expected periodicity in a circular manner. Two people who famously did so in the 19<sup>th</sup> century were WILLIAM FARR and FLORENCE NIGHTINGALE who designed *Rose Diagrams* depicting deaths over time. One of these was investigating the seasonality of Cholera by displaying eleven years of data on the *Temperature and Mortality of London* (see figure 2.5 for a detail) and others to promote sanitation strategies, including a *Diagram of the Causes of Mortality in the Army in the East* during the Crimean War (see figure 2.6).

This type of chart, also known as *Coxcomb Diagram*, is essentially a radial bar or area chart showing quantities along a recurring interval, e.g. a year. Several instances of these roses can be compared effectively and seasonal patterns can be detected. Compared to a simple timeline, there is no definitive beginning or end in the graph which diminishes the effect of the visual cut present in a collection of yearly timelines showing the same data. On the other hand, a decision has to be made on where to place the underlying cut in the data within the diagram. A trend in the data which spans over several of the periodic intervals will manifest in the graph as abrupt changes in quantity. A rose diagram is therefore only apposite as long as the seasonality within the data outweighs possible linear trends. One way to mitigate this problem is to display the data along a spiral instead. However, this quickly results in problematic differences in scale since each cycle is given more space than the preceding one. To analyse only the seasonality of the data without accounting for differences between several cycles, the data can be aggregated over several cycles instead which results in a diagram of a “typical” cycle.

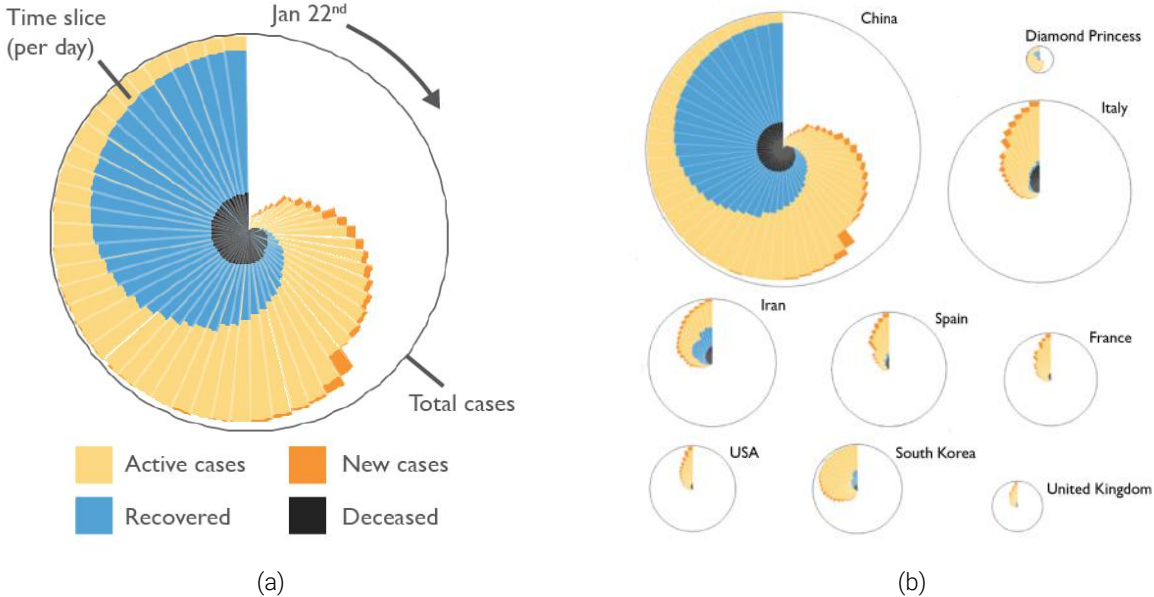


Figure 2.7: Coxcomb diagrams by Kenneth Field showing Covid-19 cases. An example with a legend (a) and as used on an interactive map (b). (Field, 2020)<sup>6</sup>

A recent application of coxcomb diagrams was done by Field (see figure 2.7). In this case, the underlying data is not of a cyclic manner but the circular shape of the chart allows for it to act as a proportional symbol on a map. This way, different areas can be compared to each other, both quantitatively, regarding the total number of cases, as well as qualitatively. The variation in shape and colour gives clues to the development through time as well as the current state. A spikey yellow shape indicates a recent outbreak with mostly active cases while a rounder shape that is dominated by a blue colour encodes a decelerated outbreak where most of the infected people have recovered.

For data, however, where an ongoing, non-cyclic trend or change is of interest, a horizontal line graph is more fitting. Building upon Priestley's historial charts, WILLIAM PLAYFAIR was among the first to create line graphs showing time-series. Displaying quantitative data in charts with a uniform scale in the 18<sup>th</sup> century, he laid the foundation for many types of charts that followed. (Rosenberg und Grafton 2012, p. 136 - 138; Tufte 2001, p. 32)

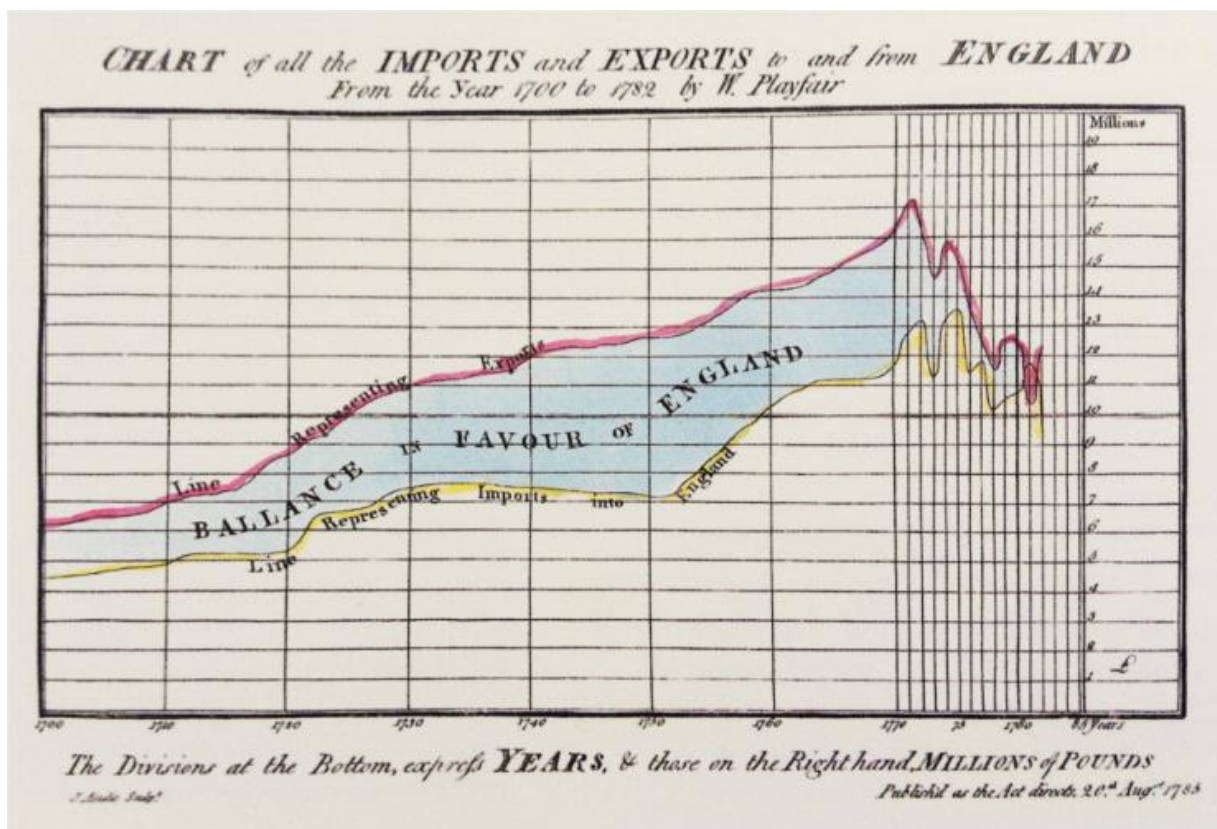


Figure 2.8: Economic chart by William Playfair (1785) with a constant temporal scale. Figure taken from Tufte (2001, p. 32).

To gain an overview of a whole century and memorize its events, the *Polish System* was developed by JÓZEF BEM and adapted and promoted by ELIZABETH PALMER PEABODY in the 19<sup>th</sup> century (Rosenberg & Grafton, 2012, p. 203-205). The chart consists of one square for each year, made up of nine individual blank squares and laid out in a ten-by-ten grid. It is to be filled with different colours and shapes that encode the type of events that took place in each year (see figure 2.9). A modern but simplified version of the Polish System can be found in *temporal heatmaps* as used by Yvonne F. Philips (2018), see figure 2.10. Temporal entities are represented by single pixels or little patches that are coloured according to a time-dependent variable. Put together, the graphic offers a visually simple yet data-dense representation of a large amount of time.

<sup>6</sup><https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/mapping-coronavirus-coxcombs>, accessed 2020/10/02

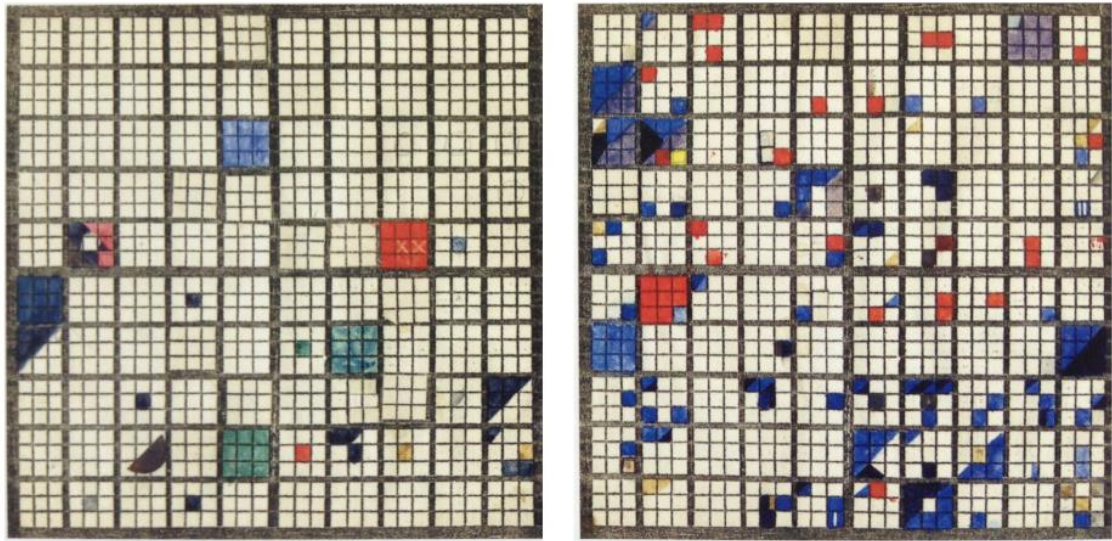


Figure 2.9: The *Polish-American System of Chronology, Reproduced with Some Modifications*, from *General Bem's Franco-Polish Method* published by Elizabeth Palmer Peabody in 1850. Used as a mnemonic device, every set of nine squares is filled with coloured shapes according to the types of events that happened in a given year. Figure taken from Rosenberg and Grafton (2012, p. 205)

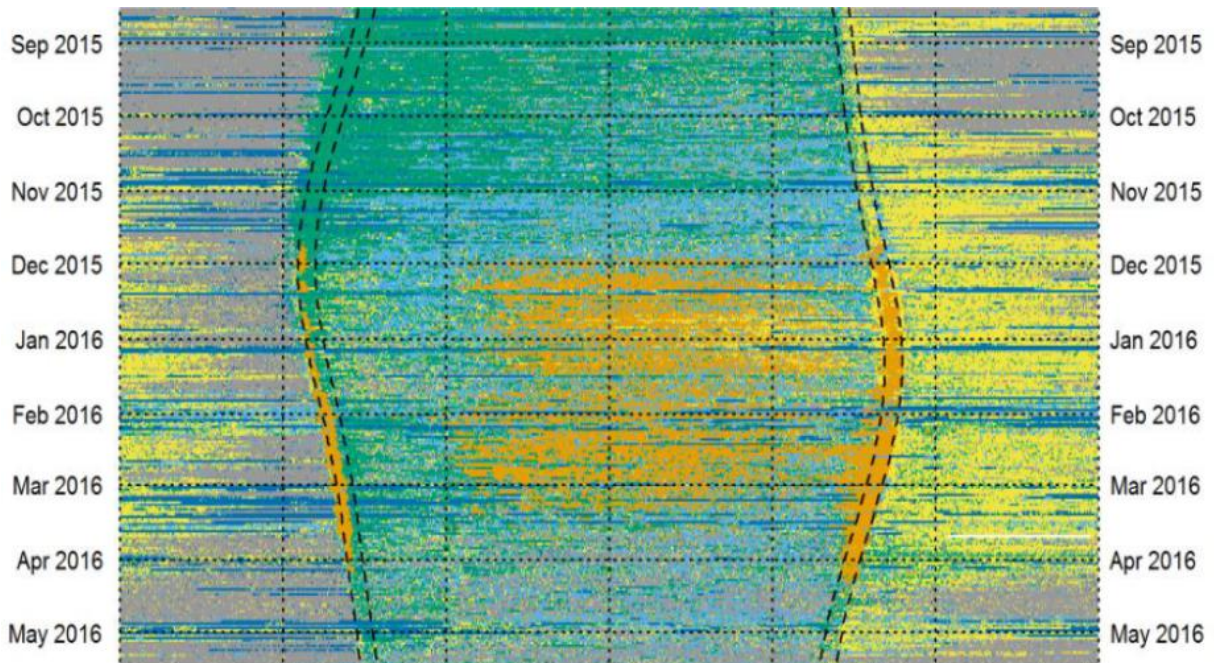


Figure 2.10: Detail from a plot of a sound recording at the Woondum National Park in Queensland, Australia, by Yvonne F. Phillips. Each minute in the recording is assigned a class corresponding to its dominant acoustic state. Different colours represent different classes: green for birds, yellow for orthoptera, orange for cicadas, blue for rain, light blue for wind, gray for quiet and pink for aircrafts. The curved lines represent civil-sunrise, sunrise, sunset and civil sunset. (Phillips, 2018, p. 123)

In the 20<sup>th</sup> century, the way we view time itself changed dramatically. As ALBERT EINSTEIN discovered the theory of Special and General Relativity, the fundamental similarities in the natures of time and space were finally acknowledged. This became visible in the spacetime-diagrams, also known as Minkowski-diagrams named after the Mathematician and Physicist HERMANN MINKOWSKI. Space is reduced to one dimension, the horizontal axis, while the temporal dimen-

sion is drawn vertically. The slope of a line in this diagram directly corresponds to the velocity of the moving object it represents. An introduction to spacetime-diagrams can be found in the first chapter of *A First Course in General Relativity* (Schutz, 2009).

Another radical development of the 20<sup>th</sup> century was, of course, the invention of computers. Today, in the digital era, generating shapes that make up diagrams based on numbers is easier and faster than ever. While the use of tables for the storage and transfer of data is still ubiquitous, the way we interact with data has changed dramatically. Fast visualization and the modification of visualizations through interaction enables humans and machines to work together in new and efficient ways. For this collaboration to be successful, however, the human who is supposed to interact with the machine should be at the center of the development of the machine. This way of designing is known as *User-Centred Design* and will be explored in the following section.

## 2.3 User-Centred Design & Usefulness Evaluation

*User-Centred Design* which, in the case of human users, is also known as *Human-Centred Design*, can be defined as the “process that ensures that the designs match the needs and capabilities of the people for whom they are intended” (Norman, 2013, p. 9). To apply user-centred design, it is essential that the user is included in the design process from the very beginning. Norman (2013, p. 222) describes this process as an iterative cycle consisting of *Observation*, *Idea generation*, *Prototyping* and *Testing*.

The *Usefulness* of a product can be understood as the sum of the product’s *Utility*, i.e. whether the product *can* achieve what it is designed to achieve and the product’s *Usability*, i.e. how well one can make use of the product’s functionality. *Usability* can be further broken down into the components *Learnability*, *Efficiency*, *Memorability*, *Errors* and *Satisfaction*. To be of high usability, a product’s functionality should therefore be easy to learn, efficient to use and easy to remember. It should lead to few errors and should be subjectively pleasing. (Nielsen, 1993, p. 25)

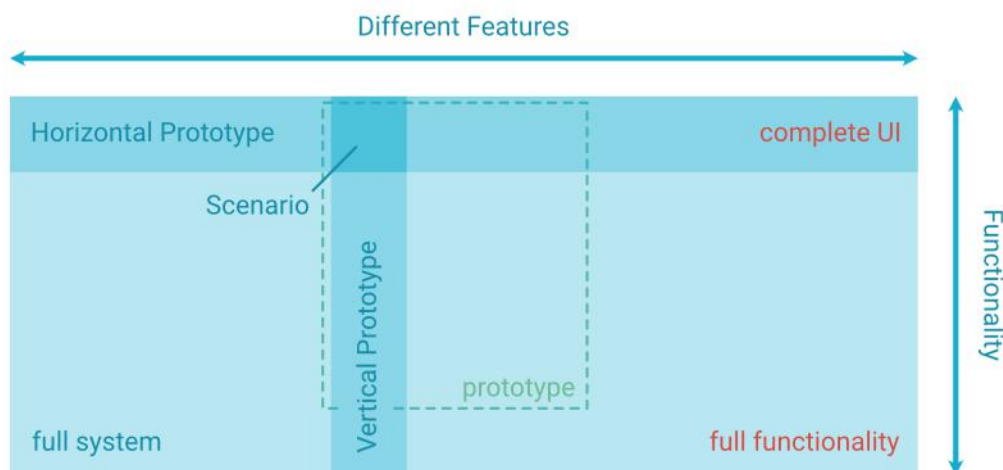


Figure 2.11: Prototyping schema based on Nielsen (1993, p. 94)

To reach that goal, Nielsen (1993, ch. 4) proposes the *Usability Engineering Lifecycle*. This approach is more detailed than that proposed by Norman but it follows the same basic principles. Starting with the user, a prototype is developed and evaluated iteratively. After assessing the user’s needs and established the goal which the final product is to achieve, the design of a prototype starts. Before iterating and developing one single design approach, several alternative basic ideas are developed by different designers and evaluated (*parallel design*). Based on these different approaches and their evaluation, a prototype is then developed further through an iterative pro-

cess. Compared to a fully usable product, a prototype lacks either part of the features of the final product but is fully functional within the features it has (*vertical prototyping*) or demonstrates all features in a full user interface while lacking actual functionality (*horizontal prototyping*), as shown in figure 2.11. The intersection of a vertical and a horizontal prototype is called a *scenario*. A scenario is neither fully functional nor does it offer all final features but it simulates the interaction process along a pre-defined course. The development of both, the horizontal and the vertical prototype, as well as scenarios is efficient and the output can be tested before investing much time and resources in the development of a final product.

To evaluate the usability of a prototype or system, different testing methods can be applied, a central one being *Thinking Aloud*. Nielsen states that it “may be the single most valuable usability engineering method” (1993, p. 195). The test user is asked to verbalize their thoughts while interacting with the system. While it is not ideal for measuring the performance of a system, it works well in gathering rich qualitative data from a small numbers of participants. Nielsen (1994) recommends including  $4 \pm 1$  test subjects in a thinking aloud study, as this already unveils about three fourths of the usability problems present.

As for ways to get information on user requirements, a straightforward option is interviewing the intended users. Especially in cases where the scope of possible answers is not yet well-defined, interviews offer the possibility of reacting to unanticipated answers or clearing up any misunderstandings. Interviews are time-consuming in both execution and analysis but deliver in-depth qualitative data. When the frame of the anticipated answers is well-defined or many participants need to be reached, e.g. to obtain significant quantitative data on certain aspects, questionnaires can be used. They require a more thorough preparation but are more easily analysed and repeated. (Nielsen, 1993, 7.2 & 7.6)

## 2.4 Transport Planning: Related Work and Research on the Study Area

A recent international survey of planning practices (te Brömmelstroet, Skou Nicolaisen, Büttner & Ferreira, 2017) investigated experiences with transportation models among professionals in the field of transportation planning. Finding a lack of trust towards those models, the authors propose “bringing models closer to those who use their outputs”. Assuming that this can be applied to any data used for planning, exploring how to best present these data for the purpose of planning becomes essential.

The *Metropolitan Region of Munich* was formed in 2007 to help facing the challenge of the city’s growing population. Two years later, the *TUM Accessibility Atlas* was initiated. Based on a multi-modal transport network and structural land-use data, the tool is being developed continuously, focusing on participatory development and incorporating up-to-date data and transparent indicators to support collaborative decision making. So far, human movement data stemming from mobile network data has not been incorporated in the tool. (Büttner, Kinigadner, Ji, Wright & Wulfhorst, 2018)

While exploratory research in both, the field of transportation and that of visual analytics has been fruitful in past years, collaboration between the two is still lacking. Andrienko, Andrienko, Chen et al. (2017) call for intensified interdisciplinary collaboration in order to successfully address existing and emerging complex problems in transportation.



### 3 Methodology

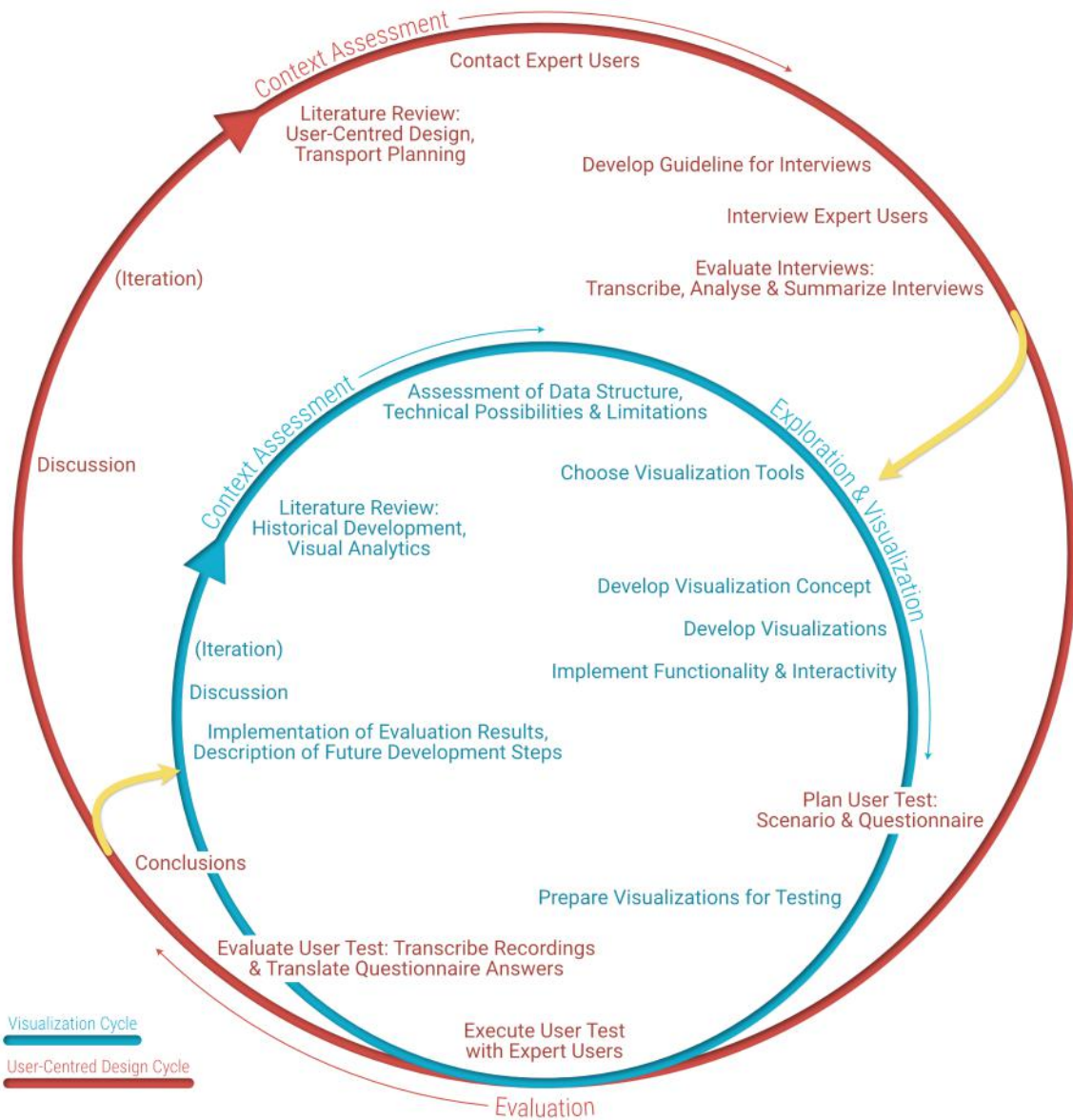


Figure 3.1: Workflow along two interconnected methodological cycles

The research can be broken down into two interconnected methodological cycles as shown in figure 3.1: an outer cycle concerned with user-centred design which corresponds to research objective 2 and an inner cycle corresponding to research objective 1 which covers the conceptualization and technical development of interactive visualizations. Both cycles start with an assessment of the context (research objectives 1.1 and 2.1) through a literature review and an analysis of the specific context of the case study. Within the inner cycle, the context assessment is mainly based on a literature review and a familiarization with the data structure and the scope of technical possibilities. For the outer cycle, the context assessment is lead by interviews with expert users who

work in the field of public transportation planning. The interviews are recorded, transcribed and summarized to establish the users' needs and support the development of a scenario.

Based on the results of research objectives 1.1 and 2.1, research objectives 1.2 and 2.2 are tackled. First, a rough user interface is developed based on the insights gained from the expert interviews. Then, a prototype is developed mostly vertically (see green rectangle in figure 2.11 on page 13) with the user interface in mind. This means that most of the intended functionality is realized for a limited amount of data while the superficial interactions between different parts of the intended user-interface are simulated. The data used for testing the prototype is prepared with a specific analytic scenario in mind and is therefore limited to supporting specific analytic tasks. Each interactive visualization is in itself fully functioning and uses real data. The interlinking of the individual visualizations is simulated by choosing corresponding filters within each visualization.

The visualizations are then tested for their usefulness in groups of two or three colleagues to answer the research questions associated with research objective 2.2. The usefulness of the concept, consisting of its utility, i.e. *to what extent can the visualized data help meet the users' needs*, and its usability, i.e. *to what extent is the way the data is presented helpful with tackling different tasks*. Each user test starts with a guided walk through the prototype with sequences of free interaction during which the participants are asked to verbalize their thoughts. During this part, each user switches between an observatory and an interactive role which is to encourage the participant to reflect on aspects of the prototype from different perspectives. The possibility to converse with colleagues and refer to examples from everyday working life should stimulate the discussion of arising issues. While testing with several colleagues at once might lead to a lively discussion, it can also lead to single opinions dominating it. To make up for that, the practical part is followed by a questionnaire which the expert users fill in individually. Here, they are asked to reflect on the practical part, summarize their impressions and state any points of critique or feedback. Finally, a discussion is held to recapitulate and summarize the individual experiences. During the first and third part, the audio is recorded to be transcribed and analysed.

The results from the user test are evaluated qualitatively and partly implemented into the code. Further implementations for a possible next iteration of the methodological cycles are described.

## 4 Case Study & Results

In this chapter, the case study is presented. Based on the scientific background described in chapter 2, the research questions listed in chapter 1 will be investigated with the methods explained in chapter 3. First, the context of the case study is given in section 4.1, followed by the investigation of research objective 2.1 in section 4.2. Then, the exploration and visualization of the data (research objective 1.2) is presented in section 4.3 and the evaluation of the resulting visualizations is described in section 4.4. Finally, a partial implementation of the results of the evaluation is described in section 4.5, along with possible future improvements.

### 4.1 Context

The context of this case study is the xMND Project which is funded by the *Federal Ministry of Transport and Digital Infrastructure (BMVI)*<sup>1</sup> of Germany and jointly carried out by the *Telefónica Germany NEXT GmbH*<sup>2</sup>, *civity Management Consultants GmbH & Co. KG*<sup>3</sup>, *Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS*<sup>4</sup> and *MotionTag GmbH*<sup>5</sup> (see *Projekt xMND | Telefónica Deutschland* (2018), *Projekt xMND: Wie Mobilfunkdaten die Verkehrsmodellierung revolutionieren könnten | civity Management Consultants* (2018) and Schubert (2019)). By combining *mobile network data (MND)* with models based on volunteered tracking data of a higher granularity, this project aims to make the resulting *extended mobile network data (xMND)* usable for public transportation planning. One of the regions studied within the xMND project is the Munich metropolitan area, which has also been chosen as the geographic region of study for this thesis. In Munich, public transportation is run by two associations, the *Munich Transportation Corporation (MVG)*<sup>6</sup> and the *Münchner Verkehrs- und Tarifverbund GmbH (MVG)*<sup>7</sup>.

#### Data Preprocessing

The data that is being visualized within this work was previously processed by the different collaborators within the xMND project (see figure 4.1). The raw mobile network data was first separated into individual *trips* representing movement of individual phones through the mobile network. Intersecting these trips with volunteered tracking data, a set of mobile network trips with additional information on the actual mode of transportation was obtained. Furthermore, geometries representing the mobile network were created based on the locations of the mobile network towers. To this end, Voronoi-polygons were generated and enriched with context data stemming from *OpenStreetMap*<sup>8</sup>, such as the number of bus stops, the length of train paths within the cell or the existence of a motorway within the cell. Based on this, a classification model was developed at Fraunhofer IAIS with the help of which, the data for this thesis was prepared. Each trip was assigned one of the modes shown in table 4.1, according to which one was most probably the actual mode of transportation, the person carrying the registered phone took.

---

<sup>1</sup><https://www.bmvi.de>

<sup>2</sup><https://www.telefonica.de>

<sup>3</sup><https://civity.de>

<sup>4</sup><https://www.iais.fraunhofer.de>

<sup>5</sup><https://motion-tag.com>

<sup>6</sup><https://www.mvg.de>

<sup>7</sup><https://www.mvv-muenchen.de>

<sup>8</sup><https://www.openstreetmap.org>

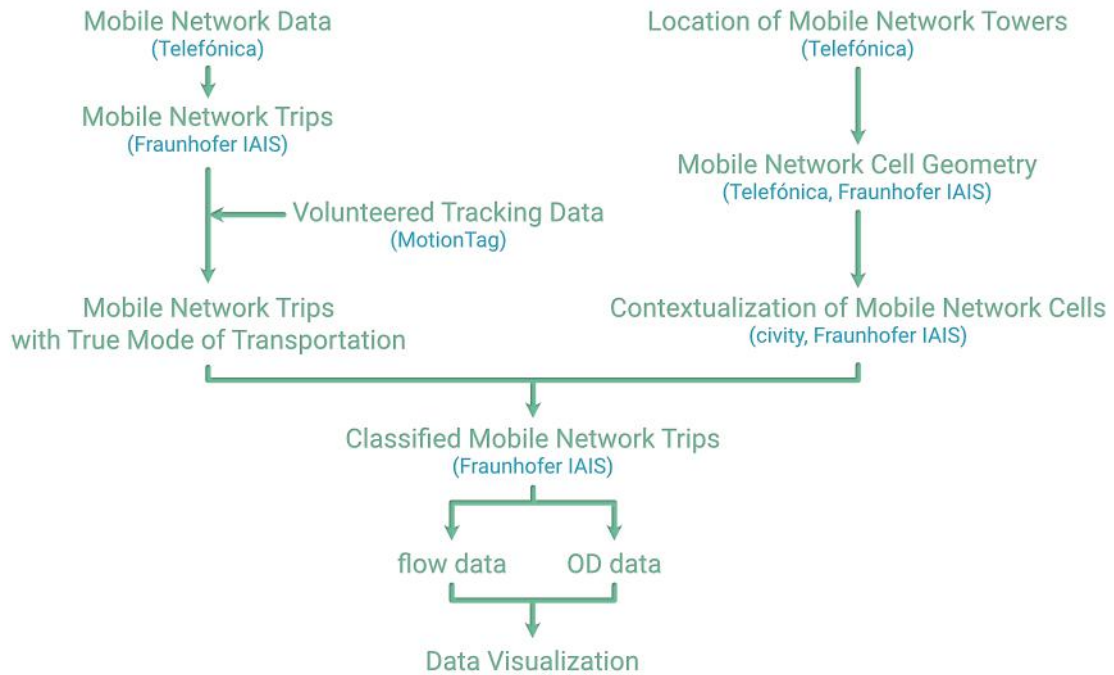


Figure 4.1: The flow of data from raw mobile network data to the visualizations presented in this thesis.

mode IAIS	mode Thesis	explanation
Mode::Bus	Bus	The majority of trip segments was probably taken by bus
Rail	Rail	The majority of trip segments was probably taken by train (excl. subway and tram)
Mode::Subway	UBahn	The majority of trip segments was probably taken on the subway
Mode::Tram	Tram	The majority of trip segments was probably taken by tram
-	public	The sum of all movements on a specific relation that were assigned one of the modes above
privat	privat	Trips that could not be assigned a public mode
moves	moves	The sum of all movements on a specific relation

Table 4.1: The different modes assigned to each trip during preprocessing (first column) and data handling (second column) and their explanations (third column)

Two different types of aggregation were done on the modelled data, resulting in two datasets: flow data and origin-destination data (OD data), both consisting of hourly data from the week of the 11<sup>th</sup> to the 19<sup>th</sup> of March 2019. For the flow-dataset, the trips were split into individual movements between cells, while for the OD dataset, only the first and last position of each trip were taken into account. Working on this aggregated level, both in the temporal and the spatial dimensions, was necessary to adhere to data privacy regulations.

## 4.2 Current Analysis Practices in Urban Public Transportation Planning

The goal of this section is to find answers to the research questions associated with the research objective 2.1: (f) *How are the intended users presently working?*, (g) *Which visualizations, if any, are they used to?* and (h) *How well does the current framework align with the users' needs and analytical questions?* To reach this goal, three interviews with a total of five experts were conducted. The aim was to reach planners from different associations and with differing planning goals and purposes. The interviews were semi-structured by a guideline of questions and the audio was recorded for later transcription and analysis.

### 4.2.1 Expert User Interview Preparation

Before the interviews were carried out, a guideline of questions and topics was compiled to keep track of what was to be assessed during the interview. These questions were divided into the two main topics *how* does planning work and *what* is being analysed and planned:

**How:** The planning method

- How is the planning process structured?
- How often do you plan/adjust plans?
- What kind of information is used for planning?
- How is that information or data prepared, i.e. in what form do you see the data?
- Are you using any visualizations? If yes, how do these look?
- Which units (number of people?) do you use when planning?

**What:** The planning content

- What is the goal of successful planning?
- Which questions arise in the planning process?
- Which temporal resolutions are relevant for answering the questions?
- Which spatial resolutions are relevant for answering the questions?
- Are there any questions or issues that cannot be answered sufficiently within the current framework due to a lack of information or data quality?

The experts' answers to these questions are the key to answering the research questions (f), (g) and (h).

The first expert (Interview I) works in the data department of one of the main public transportation associations in Munich. They are responsible for processing any data and for testing scenarios within a traffic model to serve the planners' needs of the association. The second expert (Interview II) works as a planner at the same association, mostly concerned with planning the city's bus network. Experts A, B and C (Interview III) work at the other big public transportation association within the department of conceptual planning. They are currently concerned with assessing the sensibility of expanding the association's network to further districts while they are also responsible for providing passenger information via different channels.

### 4.2.2 Results of the Expert User Interviews

In this section, the main results of the interviews are presented. The audio recordings were transcribed and annotated (see Appendix: Interview I, II and III). Main statements were extracted and structured according to topics suggested in the interview question guideline (see Appendix: Structured Collection of Main Statements from Interviews). The results are summarized accord-

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	<b>Linie 55, Putzbrunn, M.-Haslbeck-Straße - Ostbahnhof, Richtung: Rück, Last</b>																	
2	Stand 2019-09-25																	
3																		
4	Datenquelle: ZDMS, AggregationsID A 07 190415 190728 BT alle, 19.04.2019 - 28.07.2019																	
5																		
6	Bearbeiter: obermaier.sebastia																	
		<b>Last Mittelwert</b>																
7		<b>Erste Abfahrtszeit</b>	05:03:30	05:23:30	05:43:00	06:03:00	06:23:00	06:43:00	07:03:00	07:23:00	07:43:00	08:03:00	08:23:00	08:34:30	09:03:00	09:23:00	09:34:30	09:55:30
8		<b>Linie</b>	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
9		<b>Kurs</b>	4	6	8	4	6	8	4	6	8	4	5	6	8	4	5	6
10	<b>Putzbrunn, M.-Haslbeck-Straße</b>													2			2	
11	<b>Putzbrunn, W.-v.-Braun-Straße</b>													2			3	
12	<b>Putzbrunn, Kirchenzentrum</b>													3			4	
13	<b>Putzbrunn, Brennerei</b>													5			4	
14	<b>Putzbrunn, Waldkolonie</b>													5			4	
15	<b>Putzbrunn, Seemannsiedlung</b>													6			5	
16	<b>Am Hain</b>													6			6	
17	<b>Waldheimplatz</b>													6			6	
18	<b>Waldperlach</b>																	
19	<b>Waldperlach</b>		1	1	1	2	3	1	2	5	2	2	2	7	7	5	7	4
20	<b>Waldheimplatz</b>		2	1	2	3	6	3	4	9	5	4	5	10	10	9	10	4
21	<b>Eulenspiegelstraße</b>		2	1	3	3	7	3	4	9	6	5	6	11	12	11	12	4
22	<b>Im Gefilde</b>		3	2	5	4	8	3	5	15	8	7	8	14	14	12	13	4
23	<b>Tribulaunstraße</b>		4	6	6	5	8	4	6	15	8	8	9	14	14	13	14	4
24	<b>Klinikum Neuperlach</b>		11	17	10	10	10	8	8	18	11	12	14	18	20	18	20	4
25	<b>Sudermannallee</b>		14	18	12	11	11	9	9	19	13	13	18	21	21	21	23	4
26	<b>Kafkastraße</b>		14	20	13	12	11	11	11	19	14	14	19	23	22	23	25	4
27	<b>Neuperlach Zentrum</b>		4	6	5	4	3	5	3	7	4	4	7	7	7	11	10	4
28	<b>Thomas-Dehler-Straße</b>		5	6	5	4	4	6	4	7	4	5	7	7	8	12	11	4
29	<b>Friedhof Perlach</b>		5	7	6	4	4	6	4	7	4	5	7	7	7	11	11	4
30	<b>Wilhelm-Hoegner-Straße</b>		5	7	6	5	3	5	4	7	4	5	7	8	7	12	11	4
31	<b>Pfanzeltplatz</b>		8	9	7	7	5	6	7	8	7	8	8	10	11	17	14	7
32	<b>Weddigenstraße</b>		11	13	9	9	11	9	10	14	13	13	13	17	18	24	21	12
33	<b>Adilostraße</b>		12	13	10	10	11	10	10	14	14	13	14	20	19	27	22	15
34	<b>Woferlstraße</b>		13	13	11	11	14	11	12	15	18	17	20	25	26	32	28	18
35	<b>Diakon-Kerolt-Weg</b>		15	13	11	13	14	13	16	17	20	21	25	32	34	38	40	27
36	<b>Ramersdorf</b>		16	12	11	13	15	12	16	17	20	21	25	32	38	41	43	26
37	<b>Karl-Preis-Platz</b>		12	9	8	11	13	10	13	17	16	18	21	28	31	36	36	25
38	<b>Anzinger Straße</b>		10	9	9	12	15	10	13	17	16	19	20	27	30	34	37	29
39	<b>Orleansstraße</b>		9	9	8	11	15	10	12	17	16	18	20	27	29	33	34	28
40	<b>Ostbahnhof</b>		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4.2: A *load carpet* visualization as used by planners. Vertically, stations along a specific bus line are listed. One column is given to each journey along the line and averaged numbers of passengers between two consecutive stations are given as numbers. The cells are coloured according to the load and peak hours are indicated by a red border.

ing to the research questions (f), (g) and (h). All mentioned figures are collected in the Appendix (Screenshots from Interviews).

### (f) How are the intended users presently working?

As a base for their work, the planners use a variety of data and seek new opportunities, such as using mobile network data, albeit with varying experiences. In interviews I and II, the mentioned standard approach consists in analysing counting data from technical counting devices along with a small amount of manual survey data. This data is checked and preprocessed by the data-department and is made available through a database and spreadsheets that include two different kinds of visualizations (figures 4.2 and 4.3). Proposed spatial or temporal changes in the transportation network are tested on a traffic model using a software called *VISUM*<sup>9</sup> which simulates passenger flows across the network and is based on a multitude of data. In interview I, a disappointing experience with using mobile network data was described.

The planners from interview III have recently started using movement data extracted from mobile network data in a dashboard solution provided by the company *Teralytics AG*<sup>10</sup> (figure 4.4) along with official workplace statistics including forecast data which are being visualized manu-

<sup>9</sup><https://www.ptvgroup.com/en/solutions/products/ptv-visum>

<sup>10</sup><https://www.teralytics.net>

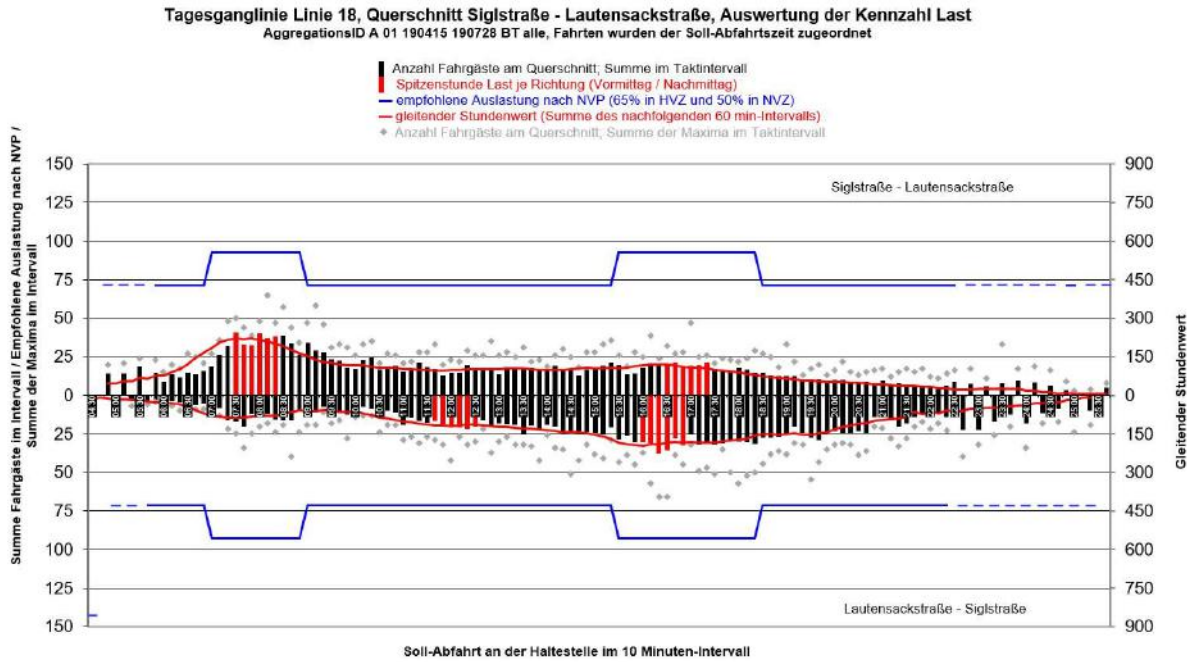


Figure 4.3: A *daily graph* visualization as used by planners. Passenger counts throughout a day between two consecutive stations along a tram line in both directions are shown as bars above and below the horizontal axis. Each bar represents a ten-minute interval containing one or two journeys. Bars within peak hours are coloured red. A red line shows the moving hourly value and a blue line indicates the recommended load. Grey dots represent maximum values within the aggregated data.

ally (figure 4.5). They use these data to assess current movement patterns and try to analyse the travel purpose of movers in order to determine whether expanding the current network to further municipalities makes sense.

In all cases, the data is mainly used to find and inspect spatial and temporal maxima, i.e. times and spatial relations where comparatively many people travel so that the current network is not meeting the people’s demand for appropriate transportation. It is also used to investigate cases that are of interest based on other input such as passenger complaints (interview II) or a requested expansion of the network (interview III).

**(g) Which visualizations, if any, are they used to?**

The planners use different kinds of visualizations as an interface to their data. In interviews I and II, the two main interfaces that are described are a so-called “load-carpet” (GER: “Belastungstepich”) as seen in figure 4.2 and a “daily graph” (GER: “Tagesganglinie”) as seen in figure 4.3. Both of these visualizations are generated within a spreadsheet containing the relevant data extracted from a database. The data is aggregated and averaged over several weeks for a specific type of weekday.

The load-carpet (figure 4.2) shows passenger counts for one direction of a specific bus-line. It is based on a table structure where each row represents a segment between two consecutive stations on the line. Each column represents one journey, so the temporal resolution varies throughout the day as it is defined by the temporal density of journeys. Each cell contains a number stating the average number of passengers on the segment given by the row at the time given by the column. A colour scheme is used to colour the cells according to the passenger load and daily peak hours are marked by red cell borders.

The daily graph (figure 4.3) shows passenger counts for one segment of the network, i.e. the connection between two consecutive stations on one tram-line, in both directions. Bars represent the number of passengers and are coloured red during peak load hours. Bars pointing upwards correspond to one direction of travel; bars pointing downwards correspond to the reverse direction. There is one bar for each ten-minute-interval, which means that intervals with more than one journey show summed values and intervals without any journeys are left without a bar. Further elements of the graph are a blue line displaying the recommended load, a red line showing the moving hourly value of passengers and grey dots above the bars indicating maximum values within the aggregated data.

Neither of these visualizations include any map. In interview II, a browser-based tool showing current load-data was mentioned to have been introduced recently in conjunction with the ongoing pandemic. This tool shows load data as coloured lines representing the network's segments on a zoomable map. A colour scale from yellow to red is used to indicate the load on each segment. This visualization is, however, not yet used on a daily basis.

Another interface that the planner from interview II has started to use is the map-based interface provided by a VISUM traffic model. This model is used to test proposed changes in the network by simulating passenger flow throughout the city. So far, the interaction with the model itself has been limited to the data-department but planners are starting to use it more directly as well. Passenger-flows are represented as lines on a map. The width of these lines represents the amount of people travelling. The spatial extent of the network is divided into traffic cells, each containing at most one underground stop and up to three bus stops.

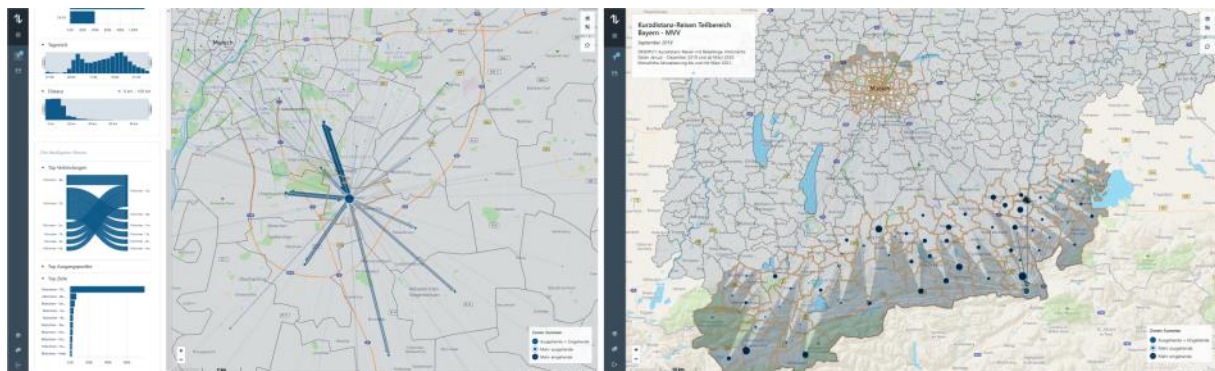


Figure 4.4: A dashboard interface showing movement data as used by planners. It consists of a map showing movement with arrows and bubbles of varying width and size and simple supplementary charts on the left. For more and larger screenshots see figures A.5, A.6 and A.7 in the appendix (pages 101-103)

In contrast to the first two interviews, the planners from interview III have as of recently started to use a dashboard solution provided by Teralytics (figure 4.4). The dashboard visualizes movement data extracted from mobile network data on a map as bubbles and arrows as well as through simple supporting graphs on the left of the interface. The different visualizations are interlinked, facilitating interactive filtering and highlighting of the data. The spatial aggregation level changes dynamically when zooming but can manually be set to a constant value. Data from different months can be compared, differences are then visualized with diverging colours on the map. While the planners mention that they are still getting used to the new interface, they do seem to appreciate having a map-based visualization as a first step of interacting with the data and the new level of interactivity that is provided.

As for other data, such as official workplace statistics, visualizations are generated manually, atop a base map created with a GIS-programme (see figure 4.5). These visualizations also use bubbles on and lines or arrows between different municipalities to indicate flows of commuting people. The width of the lines and the size of the bubbles indicate the number of people moving.



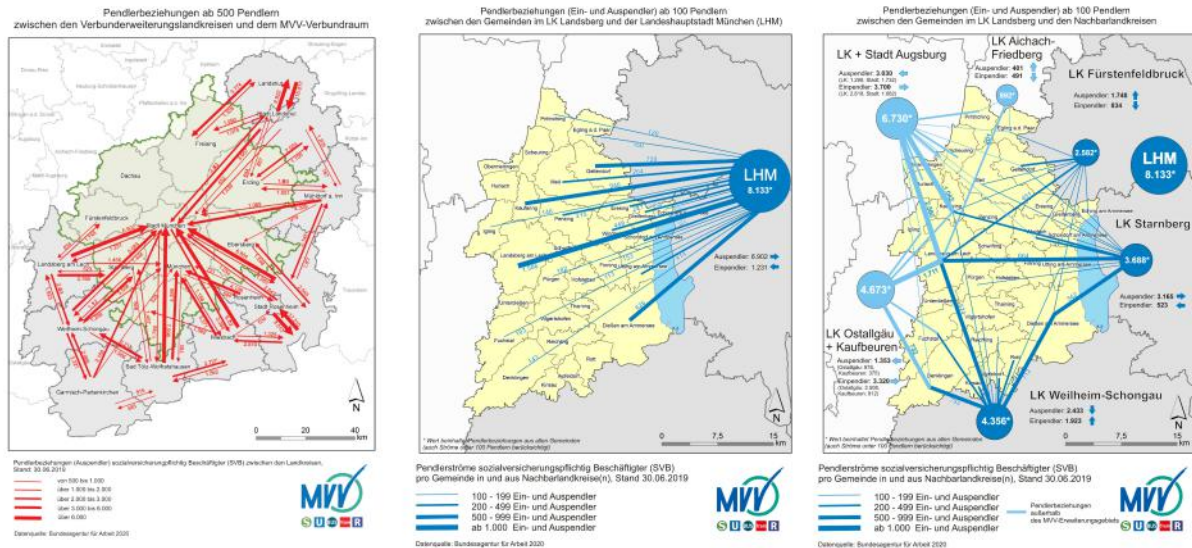


Figure 4.5: Manually generated maps as used by planners. Arrows and bubbles of varying width and size show numbers of commuters between different municipalities. For larger versions see figures A.3 and A.4 in the appendix (pages 99 and 100).

Colours are used to indicate different categories of municipalities. These static visualizations are aimed at presenting results of data analysis rather than as a means of interacting with the data.

#### (h) How well does the current framework align with the users' needs and analytical questions?

In general, the users seem to be content with their current set-up when concerned with taking on their day-to-day tasks. However, several issues have been mentioned in all interviews that are expected to possibly be resolved by incorporating mobile network data in general (interviews I and II) or by improving the data processing and visualization of the data (interview II and III).

In interviews I and II, one of three main analytical questions that cannot be answered within the current framework is the assessment of movement patterns on re-occurring but exceptional days, such as festivals, fairs and football matches. Such events tend to have a large impact on public transportation demand but their exceptional nature leads to very sparse counting data. The second main issue that was mentioned is the missing information on origin and destination of the passengers. The technical counting devices used can only register the number of passengers entering and exiting the vehicles but it is not possible to track individual passengers along their journey. The third main concern is about people moving individually, i.e. not using public transportation, as the internal counting mechanisms do not include these movements. As a public and commercial association, they want to reach as many potential public transportation users as possible and provide an adequate network to all citizens. Other mentioned issues in the first two interviews were data on people travelling underground where the coverage with counting devices is still sparse as well as information on which specific means of transport is being used.

As the planners from interview III are already working with mobile network data, several of the mentioned issues from interviews I and II do not apply here. However, one major possible improvement that was mentioned is the missing qualitative information on the movements. The planners are interested in both the means of transportation as well as the purpose of transportation, i.e. whether the people are moving for work-related reasons or for leisure. Currently, strategies to infer the latter are a comparison to official workplace statistics and the structural analysis of a simple bar-chart showing hourly movement quantities throughout a day (see the bar-chart on the top left in figure 4.4). A more pointed higher peak in the morning is associated with work-related travel while a wider peak in the morning in combination with a more pointed, higher peak in the afternoon is related to leisurely travel.

As for the usability of the visualizations of the used data, a mismatch between interview I and interview II can be observed. In interview I, the expert claims that using a spreadsheet with an included graphic representation (figures 4.2 and 4.3) as the main interface to the data is highly appreciated by the planners as it facilitates easy interactivity. A browser-based approach had been considered but deemed not useful. The planner in interview II, however, showed a great interest in a browser-based approach including a map showing current loads on the network-segments. They mentioned that this tool had been launched in conjunction with the current pandemic and that they would like to permanently incorporate such an interface on a broader basis. They mention the advantage of the spatial overview and context offered by a background map.

Based on the now established user needs, the development of a visualization concept (research objective 1.2) is described in the following section.

## 4.3 Data Exploration and Visualization

This section deals with research objective 1.2, aiming to answer the associated research questions (c) *How is the data at hand structured?*, (d) *What kind of visualizations are applicable in the specific case?* and (e) *How can those visualizations be generated, taking the intended level of interactivity into account?*

To explore the data and its structure and develop an idea of what kind of visualizations could represent it in an adequate way, the data was explored visually. Manually sketching visualization concepts allowed for an approach free from constraints imposed by technicalities. At the same time, different simple plots were generated with sampled subsets of the data to understand the grade of complexity and the variability present in the data. This also allowed for an exploration of different libraries to set the scope of what was technically feasible, given the temporal constraints of the thesis and a limited pre-existing knowledge of the programming environment. Taking into account the literature presented in sections 2.1 and 2.2 as well as the results of the expert user interviews (section 4.2), a visualization concept was developed with an idea of a possible user interface in mind. The data was then preprocessed and aggregated in ways that would facilitate the generation of the intended interactive visualizations. These visualizations were then generated using *Jupyter Notebooks*<sup>11</sup> and tested by the expert users that had been interviewed initially.

### 4.3.1 Visualization Concept

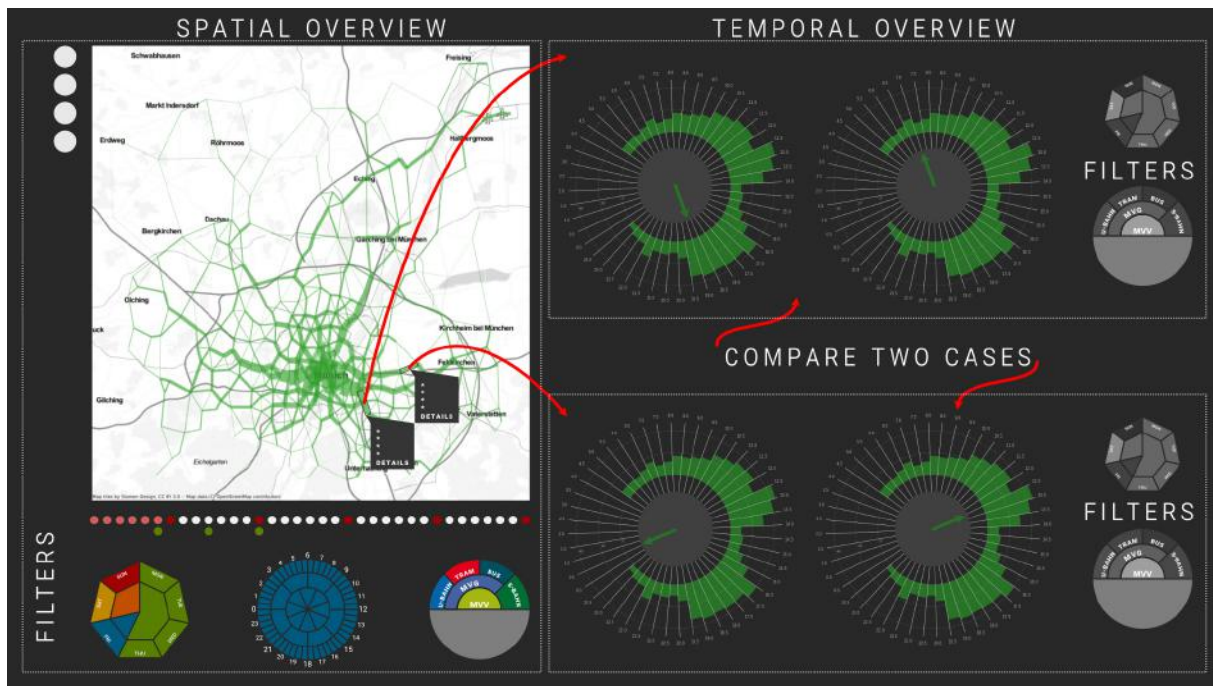
#### User Interface Concept

Based on information on the expert users' needs gathered in the interviews (see section 4.2.2), a basic user interface was conceptualized, shown in figure 4.6. Its main purpose is to facilitate the exploration of the dataset as an analytic tool to be used for planning. More specifically, it allows for the investigation of movement patterns according to different filter options. On the one hand, the analysis of spatial traffic patterns within different modes and at different times is of interest (see an explanation of the different modes in table 4.1 on page 18). On the other hand, investigating the evolution of these patterns for different modes throughout a day of a specific type, e.g. a typical workday, is of relevance. Furthermore, comparative tasks should be supported by the visualizations.

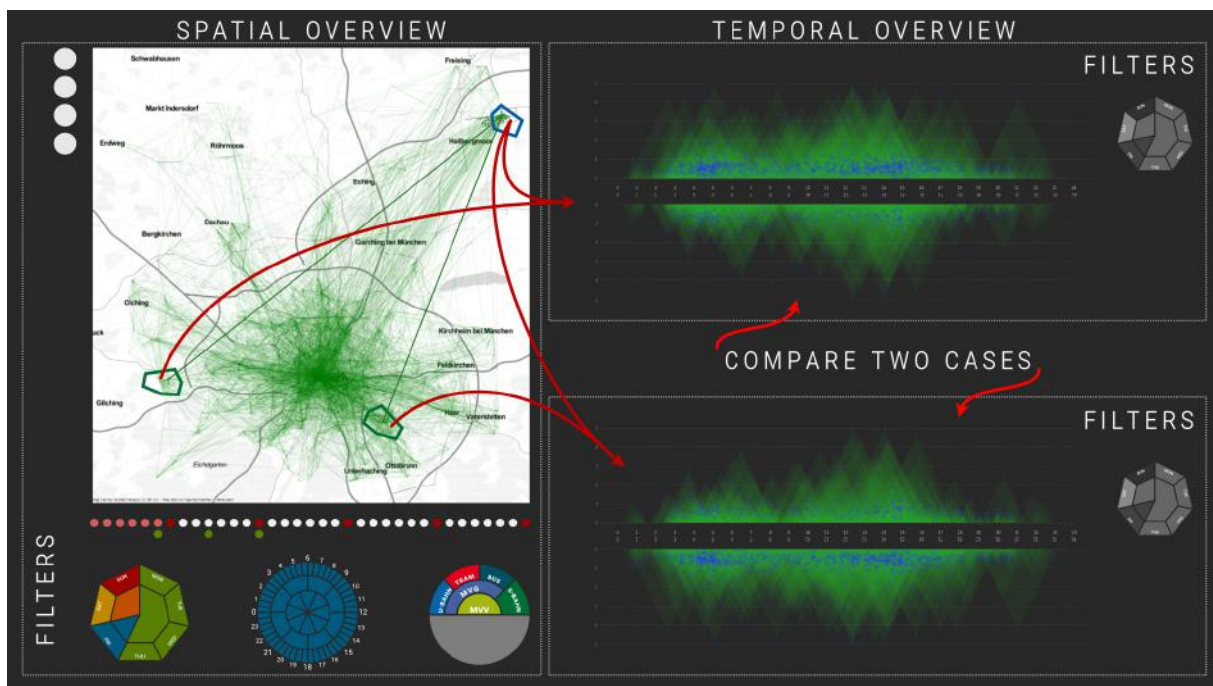
To reach this goal, a combination of different types of visualizations that were explored in section 2.2 (page 5 and following) was combined according to the literature presented in section 2.1 (page 4). A flow map with lines of variable width as first developed by Minard (see figures 2.2, 2.3 and 2.4 on page 7 and following) was complemented with temporal overviews building on rose diagrams (see figures 2.5 on page 9 and 2.6 on page 9) and the triangular representation of

---

<sup>11</sup><https://jupyter.org>



(a)



(b)

Figure 4.6: Basic user interface concept based on sample data and expert interviews, showing visualizations of sample flow data on the left and artificial data on the right (a) and sample OD data as lines on a map and in a triangular representation of the travel intervals (b). Red arrows indicate interactivity between the left and the right part of the interface.

intervals (see figure 2.1b on page 6). Flow maps offer a spatial overview for a given interval of time, with the lines encoding both the number of movements during that time via their width and a quality associated with the movement via the lines' colours. Radial bar charts derived from the idea of rose diagrams were chosen to offer a temporal overview of a day for a chosen spatial relation while emphasizing the cyclic character of the underlying temporal dimension within the data.

A triangular representation of time intervals was chosen to facilitate the investigation of trips of different temporal length. Potential temporal filter options are visualized in a cyclic manner for cyclic types of time (hours of a day and days of a week) and in a horizontal manner (days within the total dataset). To compare different subsets of data, a simple juxtaposition of multiple charts is adopted at this point.

The conceptualized user interface consists of a display split into two parts and with two versions. One version displays flow data (figure 4.6a) while the other version displays origin-destination data (OD data) (figure 4.6b). For a detailed description of both datasets, see section 4.3.2. On the left half, a flow map provides a spatial overview of the movements during times and within modes that can both be selected through filters below the map. On the right half of the display, a temporal overview is provided by means of radial bar charts and a triangular representation of intervals respectively. The temporal visualizations are either based on all data or on data spatially filtered through clicking on network segments or cells within the map in which case the data is visualized for each direction of movement individually (see red arrows). The charts can also be influenced by filtering the data they are based on through the filters to the right.

### Test Case Scenario

To test different visualizations, a scenario was developed, incorporating analytical questions that came up in the interviews and that matched the type of data at hand. The scenario is split into four different steps corresponding to four different sets of visualizations. Each of the four sets of visualizations consists of a spatial overview and a temporal overview in line with the concept described above.

Contrary to the user interface concept, the scenario does not feature juxtaposed temporal overviews of spatially different cases but the comparative part of the analytical process is contained in the analysis of a spatio-temporal anomaly. Comparison happens within single visualizations and is limited to comparing data from the day of the anomaly to data from otherwise comparable days within the dataset.

The scenario within the case study is constructed around the spatio-temporal anomaly in the form of a football event taking place on the Wednesday of the week of data. The four stages guide the user through the discovery of the anomaly and its analysis in space and time. The first two sets of visualizations (figures 4.7 and 4.8) feature the flow dataset while the third and fourth set (figures 4.9 and 4.11) display visualizations based on the OD dataset. Comparative tasks are encoded in the visualizations of stage 2 and 4. The associated tasks are described in section 4.4.1 (see table 4.2 on page 44).

### Spatial Overview

The spatial overview in each stage features a flow map showing lines of different width that represent flows between different cells of the mobile phone network. The base map is provided by *Mapbox Streets*<sup>12</sup> with a slight adjustment of the color for motorways in order to ensure their readability when combined with flow lines of different colours. It features a broad range of POIs and a detailed transit styling to support the analytic tasks through contextual information. The mobile network cells (see section 4.3.2) are shown as polygons in the background. The lines representing flows between different cells of the network connect the center points of the cells. In the first and third stages, these lines are colour-coded according to the chosen mode of transport. For the second and fourth stage, the mode is reflected in the colour of the cell polygons in the background, while the lines themselves are either green or black, depending on whether the corresponding anomaly is positive or negative, i.e. whether there is more or less movement on

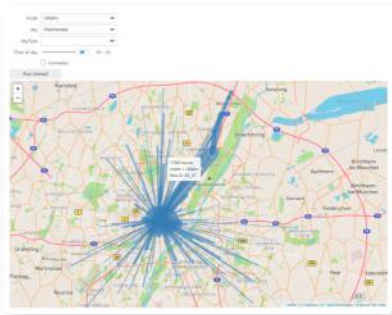
---

<sup>12</sup><https://www.mapbox.com/maps/streets>

## Stage 1: Flow Dataset, Overview

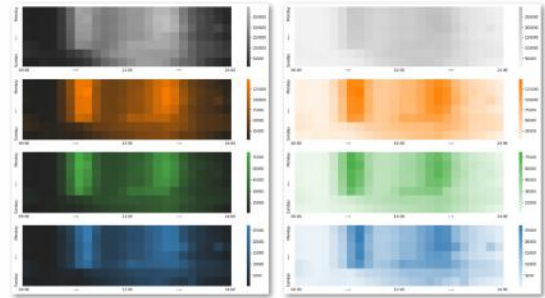
### Spatial Overview

#### FD flow map

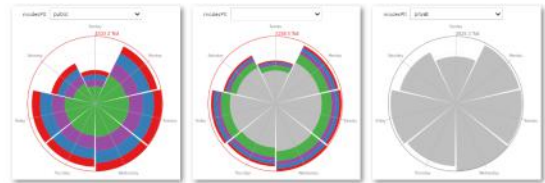


### Temporal Overview

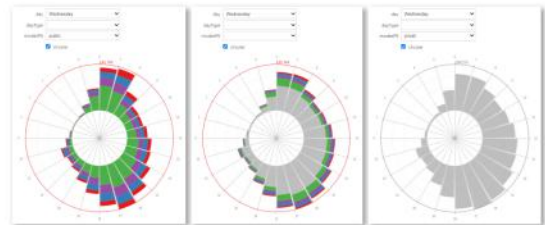
#### FD heatmap



#### FD coxcomb



#### FD radial bar chart



#### FD horizontal bar chart

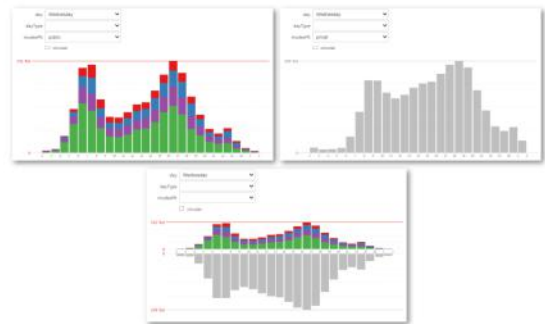
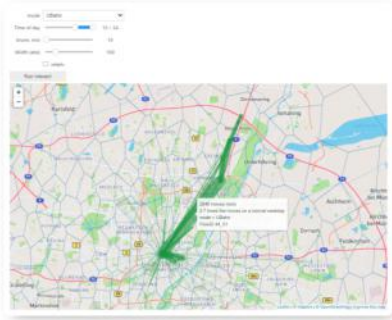


Figure 4.7: Overview of visualizations of the first stage within the scenario used for testing. Spatial overview on the left, temporal overview on the right. Larger figures of the individual visualizations can be found on pages 35, 36, 37 and 36.

## Stage 2: Flow Dataset, Anomalies

### Spatial Overview

#### FDanom flow map



### Temporal Overview

#### FDanom radial bar chart



#### FDanom horizontal bar chart



Figure 4.8: Overview of visualizations of the second stage within the scenario used for testing. Spatial overview on the left, temporal overview on the right. Larger figures of the individual visualizations can be found on pages 38 and 39.

the day of the football event compared to the other weekdays. Different filter options are offered, depending on the featured dataset and the specific analytic task.

### Temporal Overview

The temporal overview features horizontal and radial bar charts for all stages as well as a collection of heatmaps in the first stage. The colour-coding is consistent with that of the spatial overview, i.e. modes and the type of anomaly are reflected in the colours of the bars and the patches of the heat map.

The colour scheme (see figure 4.10) is based on a colorbrewer<sup>13</sup>-palette (Brewer, 2013) and roughly follows the colours used by the public transportation associations: green for trains (“Rail”), blue for the underground (“UBahn”) and red for tramways (“Tram”). As the official colours for bus

<sup>13</sup><https://colorbrewer2.org?type=qualitative,scheme=Set1,n=9>

## Stage 3: OD Dataset, Overview

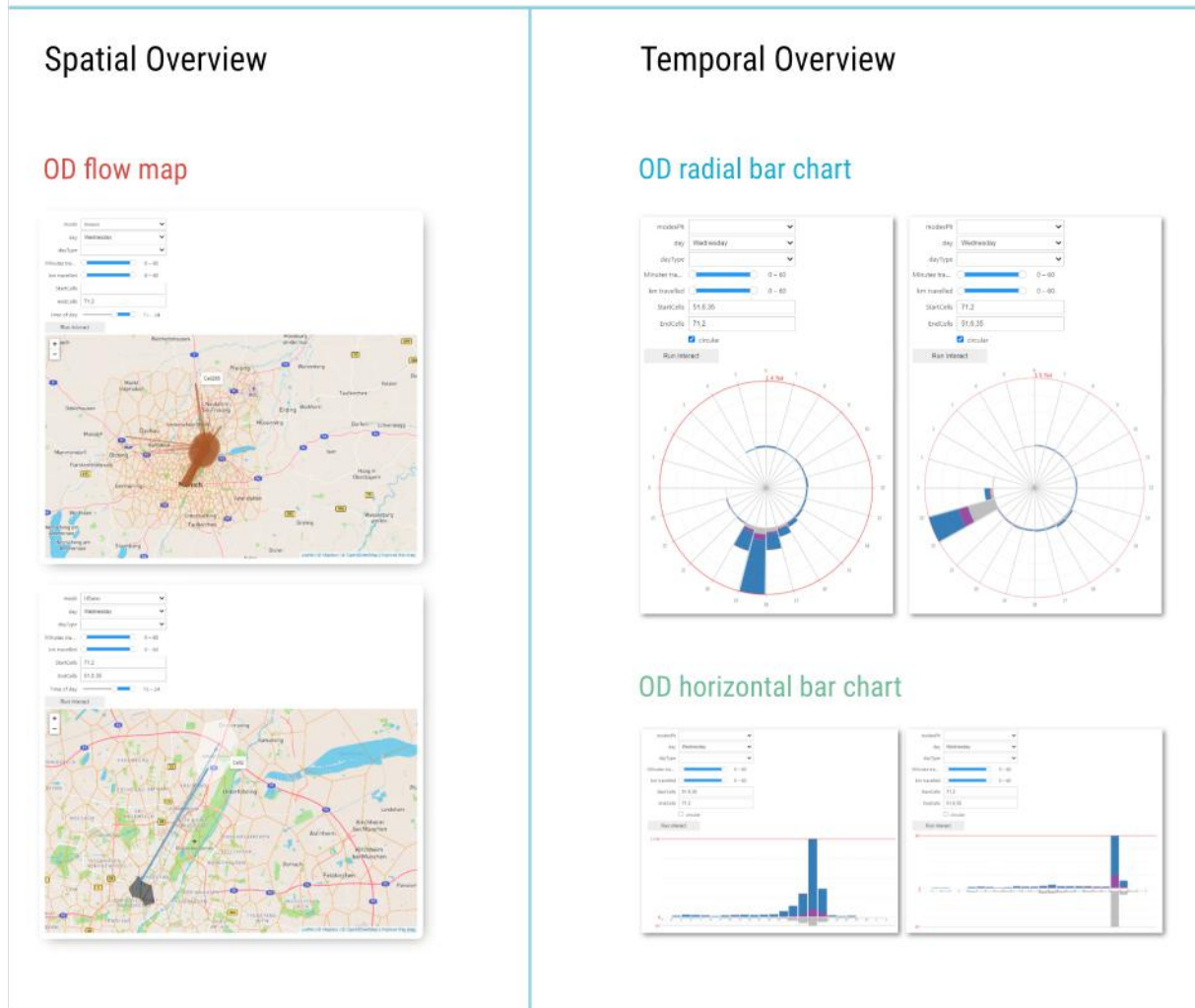


Figure 4.9: Overview of visualizations of the third stage within the scenario used for testing. Spatial overview on the left, temporal overview on the right. Larger figures of the individual visualizations can be found on pages 40 and 41.

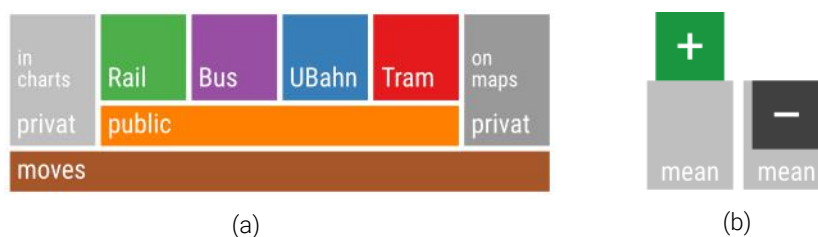


Figure 4.10: Colour scheme based on a palette by Colorbrewer.<sup>14</sup>

The colours that were used for different modes in maps and bar charts during the first and third stages of the user test scenario (a) and to represent anomalies in the second and fourth stages of the user test scenario (b).

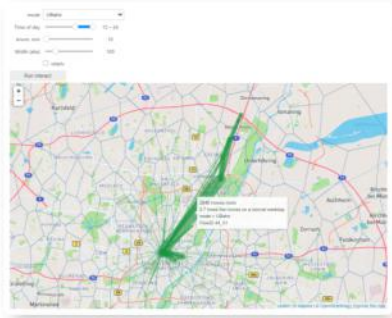
lines vary between blue and greenish blue, purple was chosen for the transport mode “Bus” to keep it distinguishable from the rest. The mode “privat” which is assigned to those movements that cannot be assigned to any of the modes above and therefore represents private transportation, is given a neutral grey colour with a lighter shade for the temporal overview plots on white

<sup>14</sup><https://colorbrewer2.org?type=qualitative,scheme=Set1,n=9>

## Stage 4: OD Dataset, Anomalies

### Spatial Overview

#### ODanom flow map



### Temporal Overview

#### ODanom radial bar chart



#### ODanom horizontal bar chart

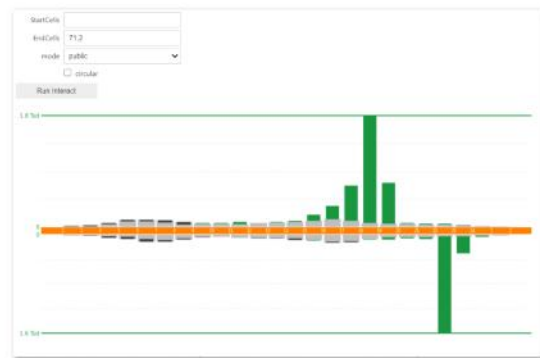


Figure 4.11: Overview of visualizations of the fourth stage within the scenario used for testing. Spatial overview on the left, temporal overview on the right. Larger figures of the individual visualizations can be found on pages 42 and 43.

background and a darker shade for the use on the colourful base map. The summed “public” movements are coloured orange and the total sum of all movements (“moves”) is given a brown colour. As for the colours that encode anomalies, a light grey is used to display the mean quantities, while green and charcoal grey represent positive and negative anomalies, metaphorically growing or burning average numbers.

The first plot for the temporal overview (“*FD heatmap*” in figure 4.7) is a collection of heatmaps of all data which spans one week. Each hour of the week is given a square of colour, with the colour intensity representing the amount of registered flows assigned to a certain mode during that hour. One heatmap is shown for each mode, in figure 4.7, it is shown for the modes “privat”, “public”, “Rail” and “UBahn”. The background is either dark or light. The purpose of this visualization is to gain an overview of the dataset and detect general patterns or trends within the data.

The next visualization (“*FD coxcomb*” in figure 4.7) gives an overview of all data in the form of a coxcomb chart, i.e. a stacked radial bar chart with the bars starting in the center of the plot. The bars show the number of trips assigned to the different modes and taken during each day of the



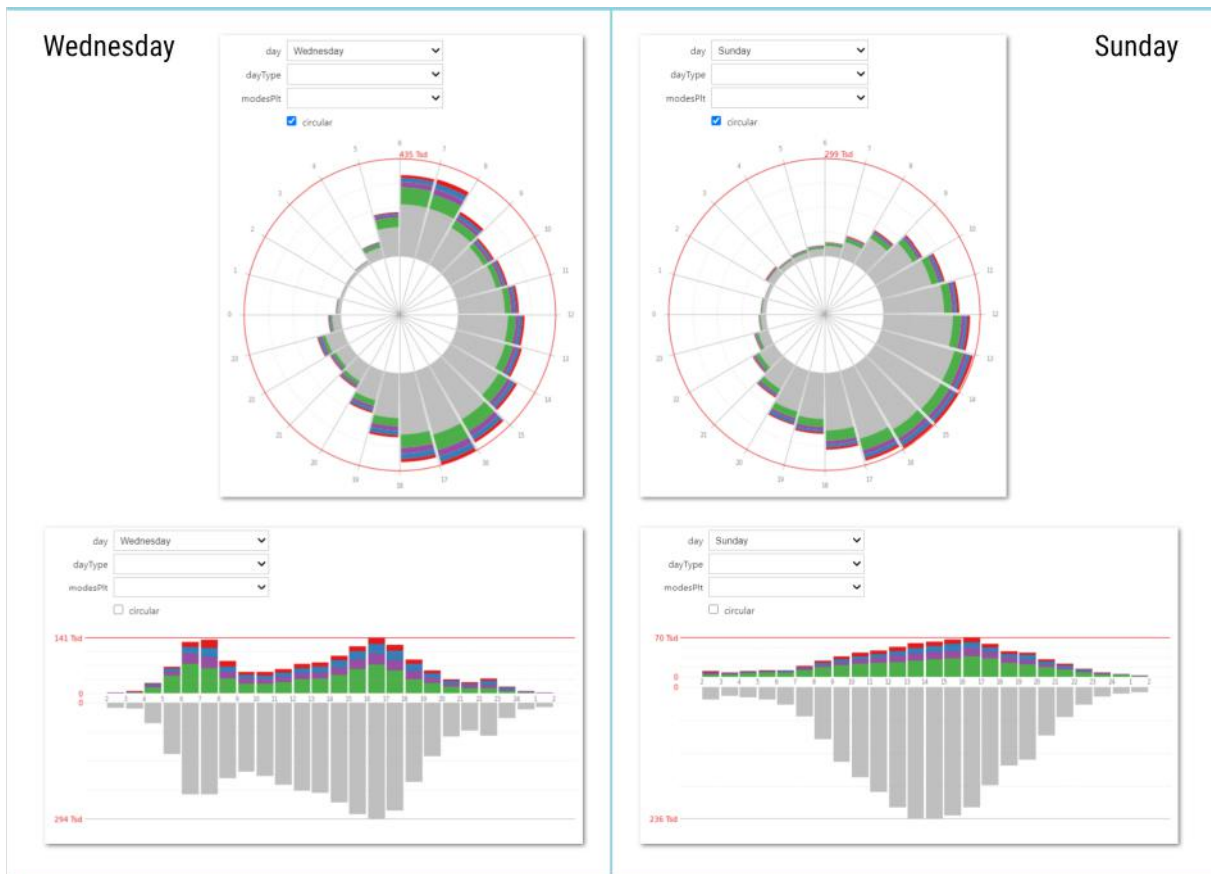


Figure 4.12: Movement patterns on two different days in radial and horizontal bar charts.

week. Through a drop-down menu, the plotted movements can be limited to those assigned to the private or public modes. This visualization differs from the preceding one as it reduces the temporal resolution from hours to days while offering quantitative information and allowing for a quantitative comparison between the different modes. Within the user test, it introduces the radial representation of temporally cyclic data and offers an occasion to discuss the topic of skewed perception of quantities when working with radial plots. As the bars in this visualization start in the center of the plot, the difference between the visualized quantity and the visually perceived quantity can be assumed to be too big for the plot to be used to reliably determine quantities visually. In the further course of the scenario, radial bar charts feature bars that are slimmer and start at about a third of the total radius to reduce visual skewing of quantities.

Radial and horizontal bar charts appear in all four stages. For the first and third stages which deal with a synoptic view of the data, they show the number of registered movements throughout one day with one bar for every hour. In the radial case, midnight is located at the left, the position corresponding to 9 o'clock on a conventional dial. This puts the morning hours at the top-right of the chart and the afternoon and evening hours at the bottom-right, both directions that are in line with the general direction of reading, meant to increase readability. In the horizontal version, the temporal axis starts and ends at 2 a.m. which is in line with the cut in the underlying data and reduces jumping between spots within the plot when analysing late night hours. The underlying data can be filtered for the weekday or type of weekday and the category of modes. In the case of all modes being plotted, the horizontal version plots the movements assigned to the mode "privat" downwards, below the horizontal axis. That way, it is possible to compare quantities for different modes and separately analyse the temporal evolution for the public and private modes within one chart. In the radial version, a similar splitting would come along with a greater skewing of quantities and was therefore not included (for a possible solution see figure 4.27 in section 4.5

on page 49). In both versions, the scale is indicated by the maximum number within the plot on a line and gridlines at quarters of that number. In the third stage (figure 4.9), the bar charts offer further filter options as they represent OD data which contains further information. Start- and end-cells of the trips can be chosen and the travelled minutes and kilometres can be limited.

For the second and fourth stages (figures 4.8 and 4.8) the radial and horizontal bar charts show anomalies in the data for Wednesday, the day a football event took place, with respect to averaged values for Monday, Tuesday and Thursday, the other three typical workdays within the dataset. Fridays are not treated as typical workdays as they generally demonstrate deviating movement patterns. The underlying data can be filtered for the assigned mode as well as for the start- and end-cells of the flows (second stage) or trips (fourth stage). The chosen mode is represented by a coloured line at the temporal axis. The averaged reference values are represented by grey bars out of which slightly slimmer green bars emerge and into which slightly slimmer charcoal grey bars submerge, representing positive and negative anomalies respectively. The scale is indicated analogous to the bar charts in the first and third stage. In cases where a start- and/or end-cell is chosen, two instances of the radial version are plotted, one displaying movement in the direction that was filtered for, the other displaying movement in the opposite direction. In the horizontal version, the two directions are plotted above and below the horizontal axis respectively.

Both horizontal and radial versions of the bar charts are provided to investigate which of these are preferred by the expert users. The hypothesis here is that while the horizontal version is likely much more familiar and quantities might be easier read than from the radial version, the radial bar chart might offer a representation that is closer to the cyclic nature of human movement patterns and thus could help detect patterns more intuitively and therefore more efficiently. The fact that certain times of the day, say the morning rush hour between six and eight o'clock, are always pointing in the same direction within the plot might facilitate a quicker orientation within each plot than searching for those times along a horizontal axis would. Comparing for example the data for Wednesday to that of Sunday using the radial version of the bar chart might give a more intuitive understanding of the different temporal patterns than when using the horizontal version (see figure 4.12).

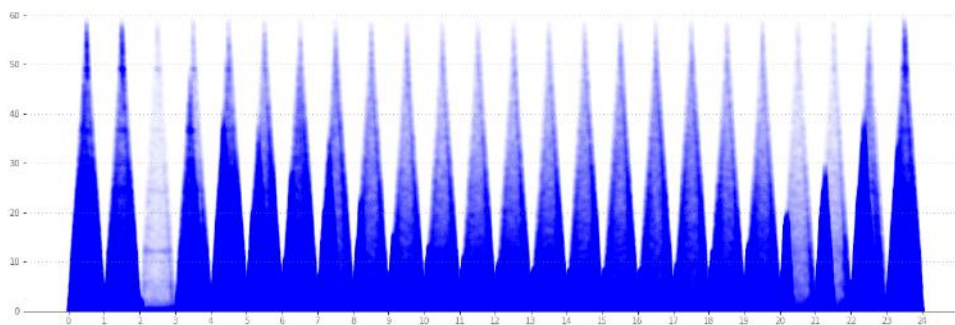


Figure 4.13: Triangular representation of a sample of the OD dataset. Each blue point corresponds to a trip between some origin and destination. Its horizontal position corresponds to the time of day the trip was taken, the vertical position to the time travelled. Because the underlying trip data was cut off at each full hour, the visualization fails to show meaningful patterns.

The fact that the final OD dataset was based on hourly data and thus shows cut-offs made it unfit to be visualized using the triangular representation of intervals as initially intended (see figure 4.13). For this reason, both the flow dataset and the OD dataset were visualized using one (stacked) bar for each hour which is structurally closest to the actual dataset. As both datasets provide data on intervals of time rather than discrete instances, interpolating the data in order to

visualize it as lines or stacked areas instead of bars would be less adequate. For the flow dataset, there is no information on when trips took place within each hour and for the OD dataset, trips that took place across hourly cut-offs are separated in the dataset which would need to be fixed within data preprocessing before it would make sense to display the data with a higher than hourly resolution.

### 4.3.2 Data Handling

To be able to generate the different visualizations described above, the preprocessed data was aggregated in different ways, within two Jupyter-Notebooks. This section will give a short overview of the data handling process; a detailed description is included in the notebooks which can be found online.<sup>15</sup>

#### Flow Data Structure

The preprocessed flow data consists of 168 separate csv-files, one for each hour of the week. Each file contains one row for every spatial relation and direction where movement was registered within the given hour and day. The columns provide the direction of movement, the IDs and geographical features of the network cells between which the movement occurred, counts of movement for each mode and their sum, as well as two timestamps for the given hour (see table A.1 in the appendix).

#### Origin-Destination Data Structure

The preprocessed origin-destination data consists of 168 separate csv-files, one for each hour of the week. Each file contains two rows for every trip registered within the given hour and day, one with data on the origin of the trip and one with data on the destination. The columns provide a timestamp, longitude and latitude, the travelled distance in kilometres, the travelled time in seconds, the trip-id, the number of cells that were travelled through, whether the row represents the start or the end of a trip and the modelled mode in which the majority of trip segments were taken (see table A.8).

#### Aggregation of Flow Data

A Jupyter-Notebook<sup>16</sup> was written to aggregate the preprocessed flow data into datasets usable for the different visualizations. The separate csv-files were combined into one data-frame (see table A.2) and the mode column names were simplified (see table 4.1 on page 18). A linestring was constructed for each row, connecting the centroids of start- and end-cell of the given spatial relation. Columns for the hour of the day, the weekday and a day-type were extracted from the timestamps. Movements registered between midnight and 02:00 am were assigned the preceding weekday to better represent human behaviour. Movement during those hours is typically part of the past day's activities rather than those of the coming day and public transport shuts down for the night around that time as well. A dataframe containing all unique network cell polygons and centroids (see table A.3) was extracted to be used as a contextual element in the map plots as well as for spatial aggregation of OD data (see below).

To prepare the flow data for plotting on a map, it was aggregated to unique spatial relations for each hour, taking the sum of movements in both directions since those were not represented on the flow maps. The resulting dataframe was extended by one additional column for each mode containing the mean amount of movers on the respective relation throughout the dataset (see table A.4). This value was used to normalize the line widths within the flow maps. Values

---

<sup>15</sup><https://github.com/klavere/xMNDvis>

<sup>16</sup>[https://github.com/klavere/xMNDvis/blob/master/D12\\_flows\\_DataPrep.ipynb](https://github.com/klavere/xMNDvis/blob/master/D12_flows_DataPrep.ipynb)

smaller than 5 were excluded from the mean as to prevent extremely wide lines popping up in cases where a relation shows a small movement during some few hours but usually has close to no movement registered. For example, a relation where only about two movements are registered during 20 hours of the week but during one hour, seven movements are registered, would show up as a line with a relative width of  $7/((7 + (20 * 2))/21) \approx 3$  instead of  $7/(7/1) = 1$ , showing that this movement is three times the mean amount. Since lines are only plotted if they represent at least five registered movements, excluding smaller movement from the calculation of the mean ensures that the plotted width actually represents the ratio to the width of all lines that would end up being plotted for single hours on this relation. An even stronger negative pop-out-effect would appear if the means were calculated over all 168 hours of the week, including those where no movement was registered. Preparing for the visualization of a temporal overview, a dataframe was extracted similarly to that used for maps but retaining the directional split while omitting columns containing geometries only necessary for plotting on maps. This dataframe thus contains one row for each spatial relation, direction of flow and hour of the week (see table A.6).

To visualize the movement anomalies of Wednesday with respect to the movement on Monday, Tuesday and Thursday, two separate dataframes were extracted from the first one. One dataframe containing all movements from Monday, Tuesday and Thursday and one containing all movements from Wednesday. The first dataframe was aggregated to unique spatial relations and hours, computing the mean of the registered movements in the different modes. Then, the two dataframes were merged and extended by columns containing the quantitative anomalies, i.e. the difference in the number of movements on Wednesday and the corresponding mean number of movements on the other three days (see table A.5). Another dataframe for visualizing the temporal evolution of the anomalies on Wednesday was extracted from the first, although again retaining the directional component of the movements (see table A.7).

### Aggregation of OD Data

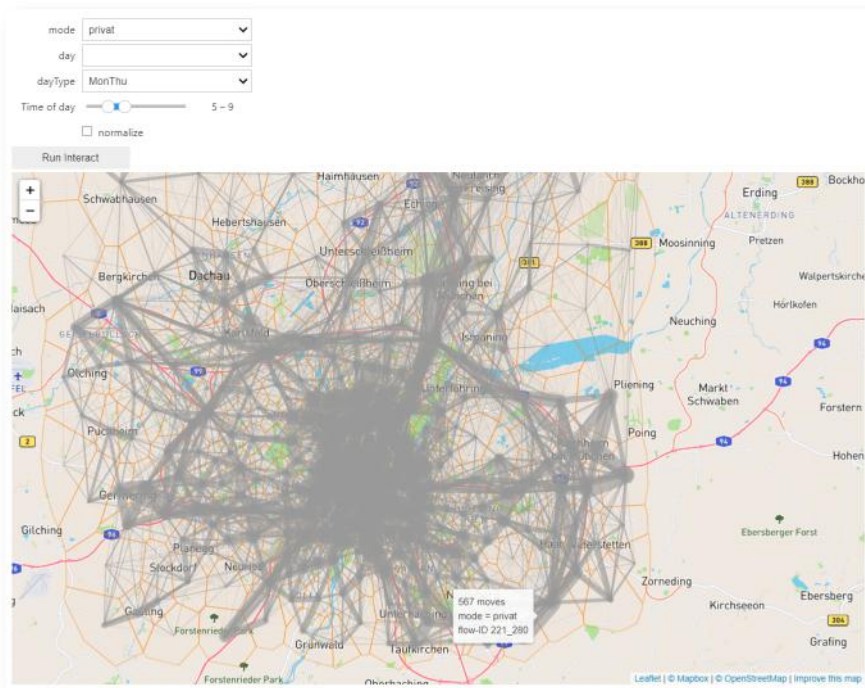
A second Jupyter-Notebook<sup>17</sup> was written for the aggregation of the preprocessed OD data into datasets to be used for the visualizations.

The data was spatially joined with the cell network (see table A.3) and the cell-IDs for the cells of origin and destination were added. The timestamps and modes were handled analogous to those in the flow dataset and the dataframe was aggregated to movements between different cells within one hour of the week. Travelled distances, times and the number of passed cells were averaged while the movements within each mode were counted (see table A.9). From the dataframe containing all movements, a pivot table was created, merging all rows containing data on the same spatial, temporal and directional entities into one row, thus adding columns for the mode-specific data. Movement counts were summed over while data on the travelled distance, time and number of network cells was averaged for each mode and spatio-temporal-relation. Finally, columns for the summed public modes and the sum of movements of all modes were added (see table A.10).

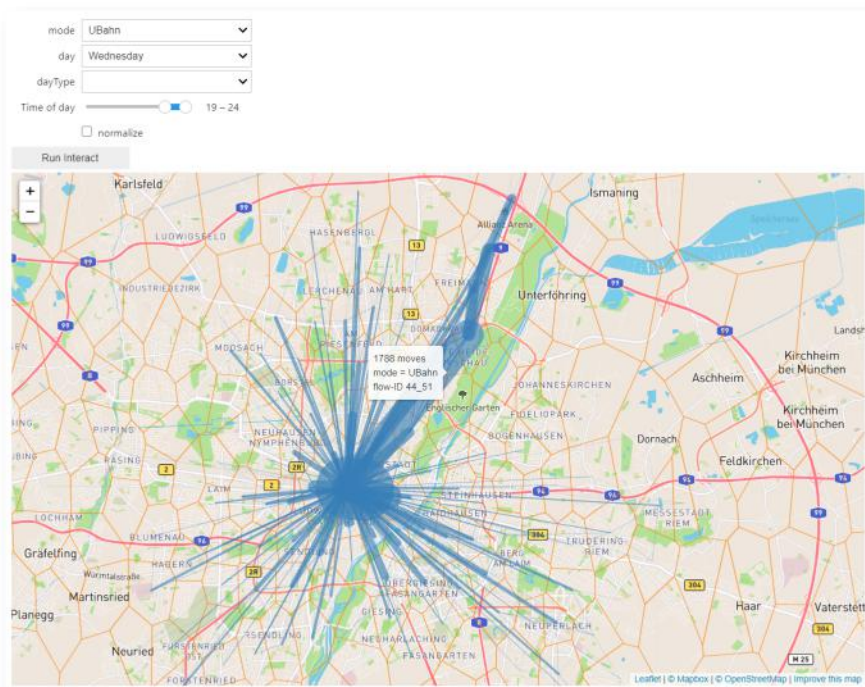
For the visualization of anomalies, a dataframe was constructed analogous to that of flow data anomalies, containing averaged movement counts of Monday, Tuesday and Thursday, the data of Wednesday and the respective anomalies (see table A.11).

---

<sup>17</sup>[https://github.com/klavere/xMNDvis/blob/master/D34\\_OD\\_DataPrep.ipynb](https://github.com/klavere/xMNDvis/blob/master/D34_OD_DataPrep.ipynb)



(a)



(b)

Figure 4.14: Flow maps generated using the flow dataset. (a): The averaged number of movements during 5 and 9 a.m. on a weekday (Monday through Thursday) that were assigned the mode “privat”. (b): The averaged number of movements registered between 7 p.m. and midnight on Wednesday that were assigned the mode “UBahn” (b).

### 4.3.3 Data Visualization

#### General Set-Up & Interactivity

The visualizations were generated within Jupyter-Notebooks through the *JupyterLab*<sup>18</sup> interface using different python packages. Interactivity was introduced via the *ipywidgets* and *folium* packages. With *ipywidgets*, custom widgets were created to run coded functions depending on parameters set interactively. Flows were plotted as lines on interactive maps using the *folium* package.<sup>19</sup> Other packages were tested but either did not feature all needed functionality yet (*mpld3*, *bokeh*) or were too slow with plotting the given amount of data (*plotly*, *jupyter dash*)

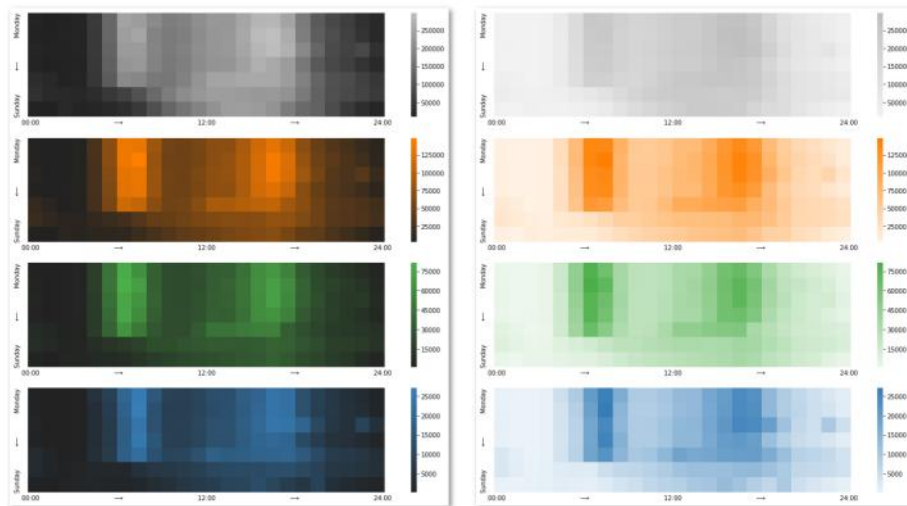


Figure 4.15: Heatmaps of the whole dataset according to different modes (“privat”, “public”, “Rail” and “UBahn”). Each hour is represented by one square. Colour intensity corresponds to the total number of registered movements.

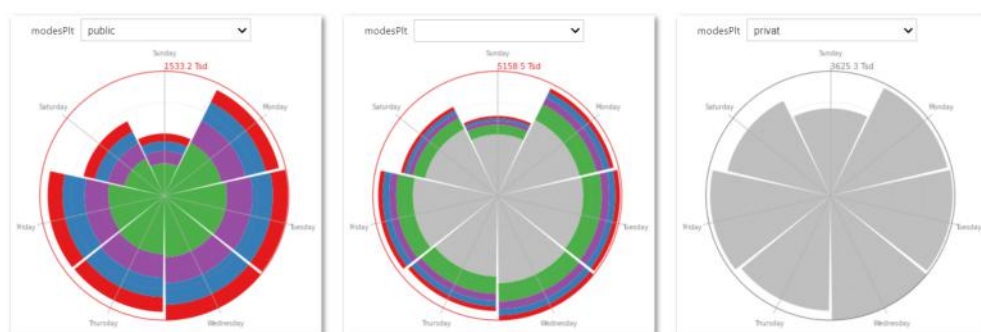
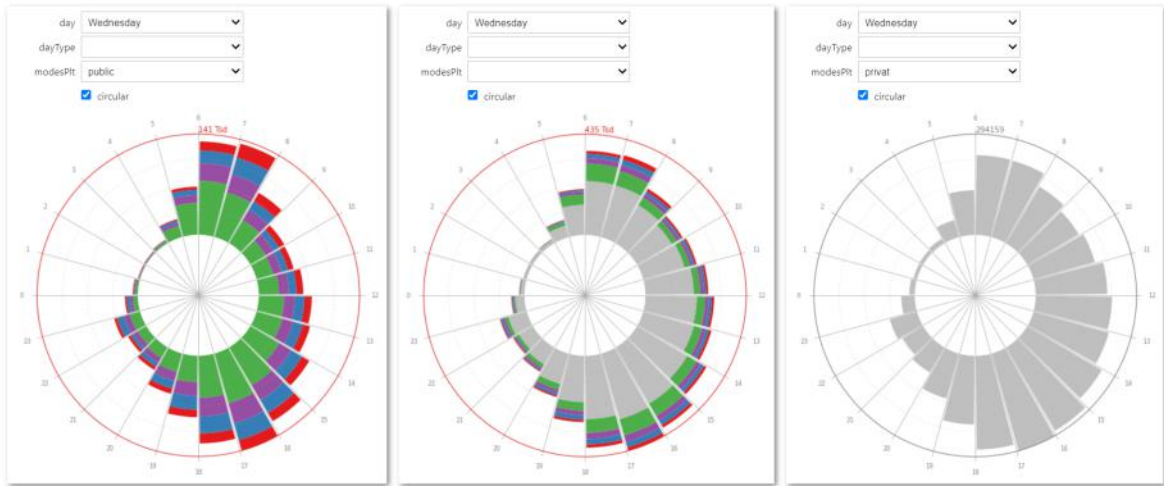


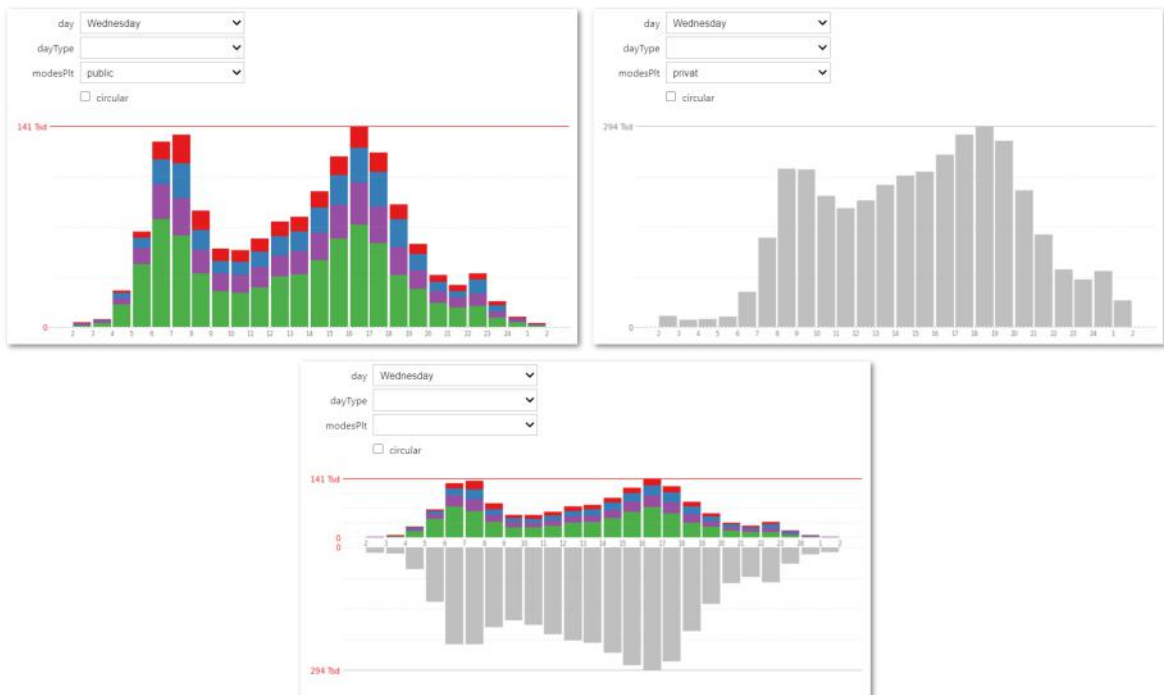
Figure 4.16: Coxcomb diagrams of the whole dataset. Total daily numbers of public (left), all (middle) or private (right) movements.

<sup>18</sup><https://jupyter.org>

<sup>19</sup>see the documentation in the notebooks for further details: <https://github.com/klavere/xMNDvis>



(a)



(b)

Figure 4.17: Bar charts for total numbers of public (left), all (middle) or private (right) movement on Wednesday. Each hour of the day is represented by one (stacked) bar. Colours indicate the assigned mode of transportation. (a): radial version, (b): horizontal version.

### Plotting Functions

The different plotting functions were developed by testing on sample datasets. To make sure that they would work for various chosen filters, some plots were done for all possible parameters, saved as png-files and checked for any irregularities or weaknesses that might occur for e.g.

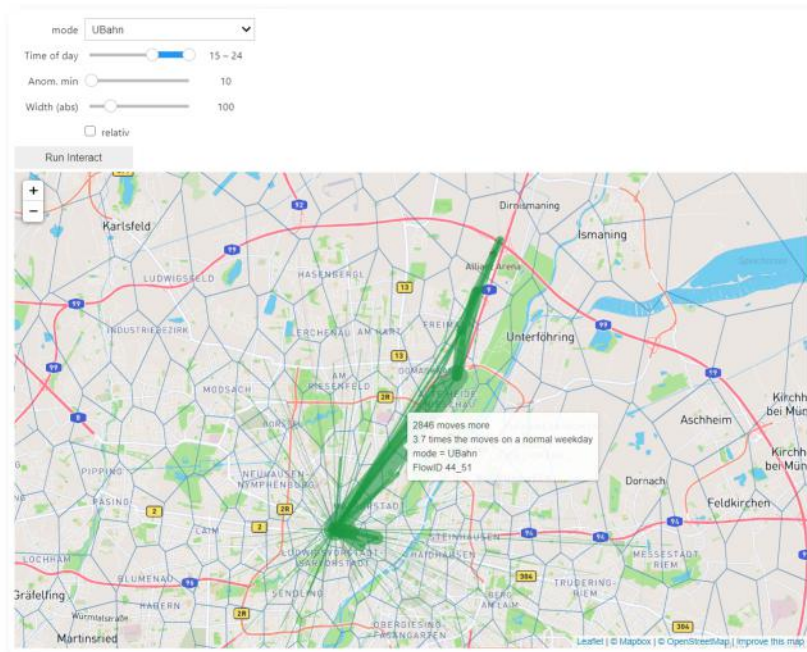


Figure 4.18: Flow map for the second stage showing anomalies within the flow dataset for Wednesday (3 p.m.-midnight) compared to the averaged numbers for Monday, Tuesday and Thursday. Lines are plotted for spatial relations where at least 10 people more or less travel on Wednesday.

incomplete subsets of the data. Eventually, two notebooks were written for each stage of the scenario, one for the spatial and one for the temporal overview. In the following, the different functions within these notebooks are summarized.

The mapping function<sup>20</sup> for the spatial overview in the first stage of the scenario (figure 4.14) first filters and aggregates the flow dataset (table A.4) according to the chosen temporal filters, averaging quantities over different days and summing over different hours. Flows with at least five movements are then plotted on a map on top of polygons representing the mobile network cells. The lines are coloured according to the chosen mode and their widths are calculated using the number of movements within the chosen mode and normalized to either the mean of all plotted lines' widths or that of each segment throughout the whole dataset in case "normalize" is checked in the filter options. Tooltips are added to the cell-polygons stating the cell-ID and to the flow lines containing the number of movements the line represents as well as the chosen mode and the flow-ID giving the two cell-IDs that are connected by the flow line.

In the code<sup>21</sup> for the visualizations in the temporal overview of stage one (figures 4.15, 4.16 and 4.17), separate functions are written for the horizontal and the radial bar charts and a third function to choose between the two, depending on the chosen parameters. Within the plotting functions, the datasets (see table A.6) are filtered and aggregated according to the chosen filter options.

The spatial overview in the third stage of the scenario deals with anomalies in the flow data (see table A.5). The code<sup>22</sup> generates a flow map (figure 4.18). Lines now show the anomalies of Wednesday with respect to the average movements of Monday, Tuesday and Thursday and are colour-coded according to the type of anomaly: green if there is more movement on Wednesday,

<sup>20</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/1S\\_flows\\_mapVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/1S_flows_mapVis.ipynb)

<sup>21</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/1T\\_flows\\_cycleVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/1T_flows_cycleVis.ipynb)

<sup>22</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/2S\\_flowsAnom\\_mapVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/2S_flowsAnom_mapVis.ipynb)



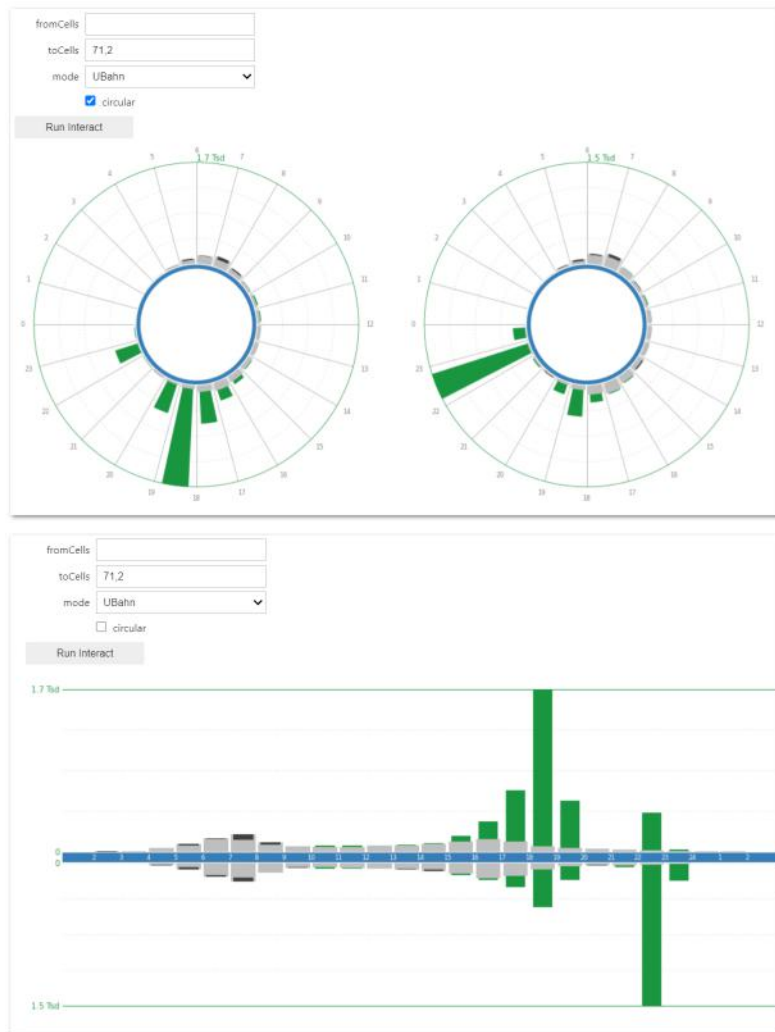
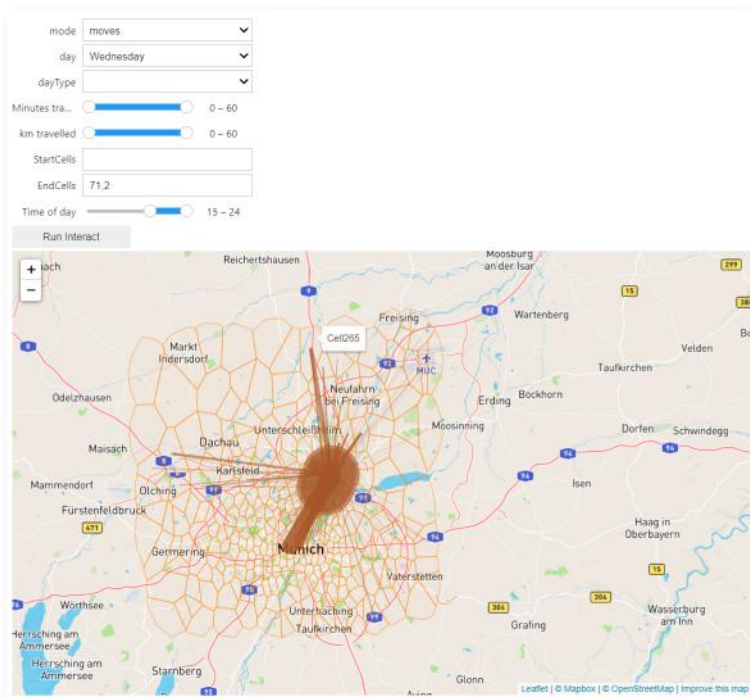


Figure 4.19: Bar charts showing Wednesday anomalies in the flow dataset. The dataset is filtered for movements that were assigned the mode “UBahn” and registered for spatial relations ending at cells 71 and 2 (left half in the radial version, top half in the horizontal version) or starting from these cells (right half in the radial version, bottom half in the horizontal version). In these cells is located Allianz Arena, the stadium that was hosting a football event on the given day. Each hour of the day is represented by one bar. Averaged movements on Monday, Tuesday and Thursday are shown as grey bars. Green bars growing from the grey bars represent positive anomalies on Wednesday, charcoal grey bars burning into the grey bars represent negative anomalies on Wednesday.

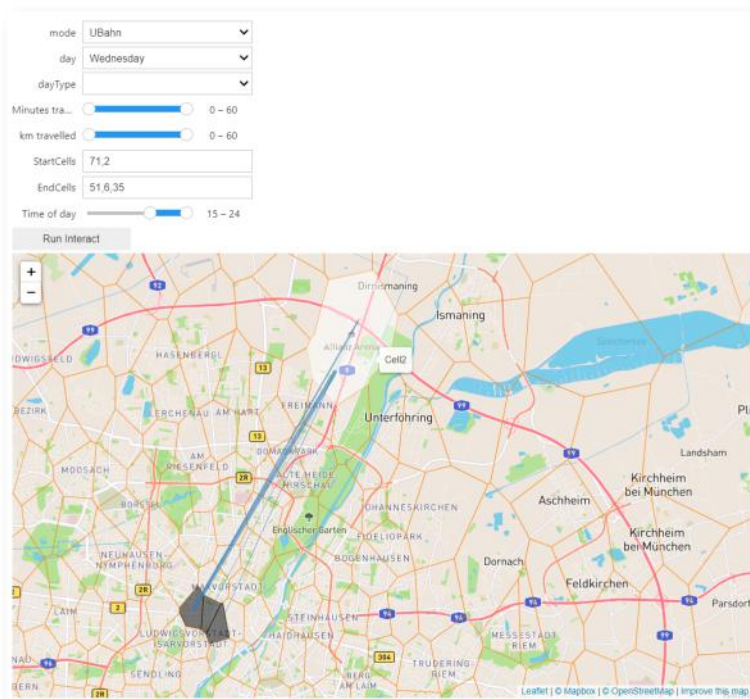
charcoal grey if there is less. The dataset is filtered and aggregated according to the chosen hours of the day and the minimal anomaly. The lines’ widths are calculated using the anomaly within the chosen mode and are again either normalized to the chosen width factor or the averaged values for Monday, Tuesday and Thursday in case “relativ” is checked in the widget. Tooltips for the cell polygons, now coloured according to the chosen mode, are added as for the first flow map while the tooltips for the flow lines additionally include the ratio of movements on Wednesday to the average movement numbers on the other three days.

The complementary temporal overview (see figure 4.19) is generated in a further notebook<sup>23</sup>. Again, two separate plotting functions for the horizontal and the vertical version are written while a third function filters and aggregates the dataset (table A.7) according to the cell filter options and decides which and how many plots to generate.

<sup>23</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/2T\\_flowsAnom\\_cycleVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/2T_flowsAnom_cycleVis.ipynb)

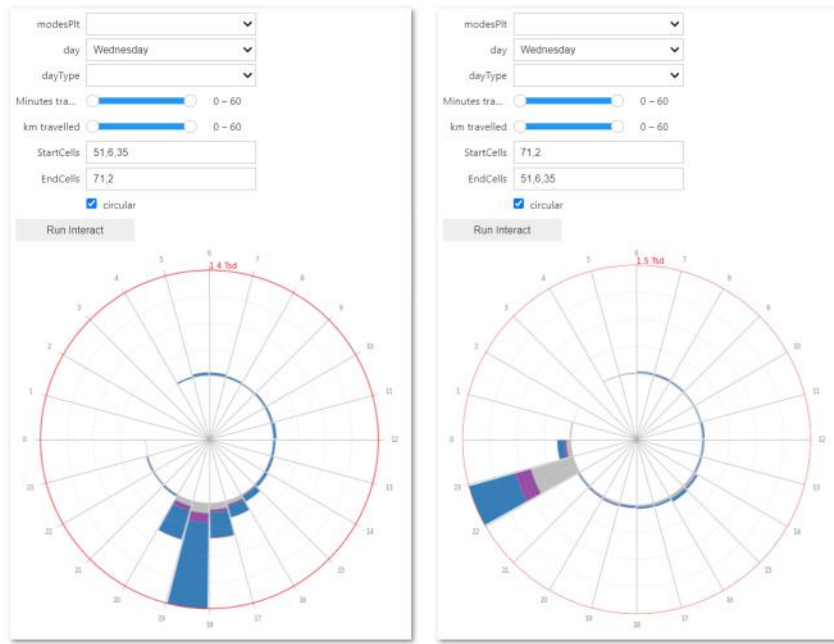


(a)

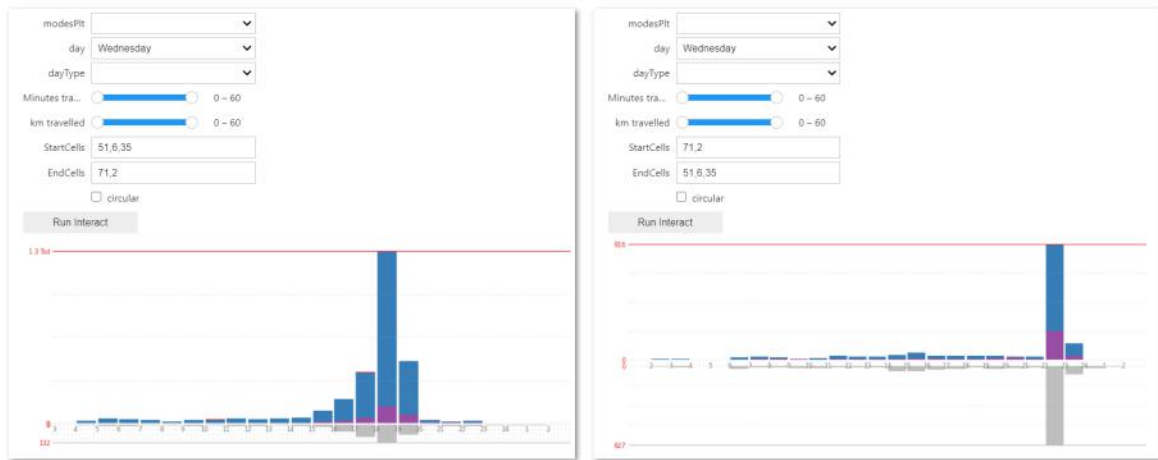


(b)

Figure 4.20: Flow maps showing OD data between 3 p.m. and midnight on Wednesday. (a): Movements of all modes ending at the network cells where Allianz Arena is located. (b): Movements of the mode “UBahn” starting at the network cells where Allianz Arena is located and ending at network cells containing three major underground stops in the city center.



(a)



(b)

Figure 4.21: Bar charts showing OD data for movements on Wednesday from the city center to the cells containing the stadium (left side) and back (right side). Most of these movements were assigned the modes “UBahn” (blue), “Bus” (purple) and “privat” (grey). (a): radial versions, (b): horizontal versions.

The mapping process<sup>24</sup> for the spatial overview of stage three is similar to those above but more filter options are given due to the richer underlying dataset (table A.10) and normalization is only done with respect to the plotted lines. The polygons corresponding to the chosen start- and end-cells are coloured white and black (see figure 4.20).

For the corresponding temporal overview (figure 4.21), two separate functions are written to plot the radial or the horizontal version and a third to filter and aggregate the data (table A.10) and

<sup>24</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/3S\\_OD\\_mapVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/3S_OD_mapVis.ipynb)

chose which version to plot.<sup>25</sup>

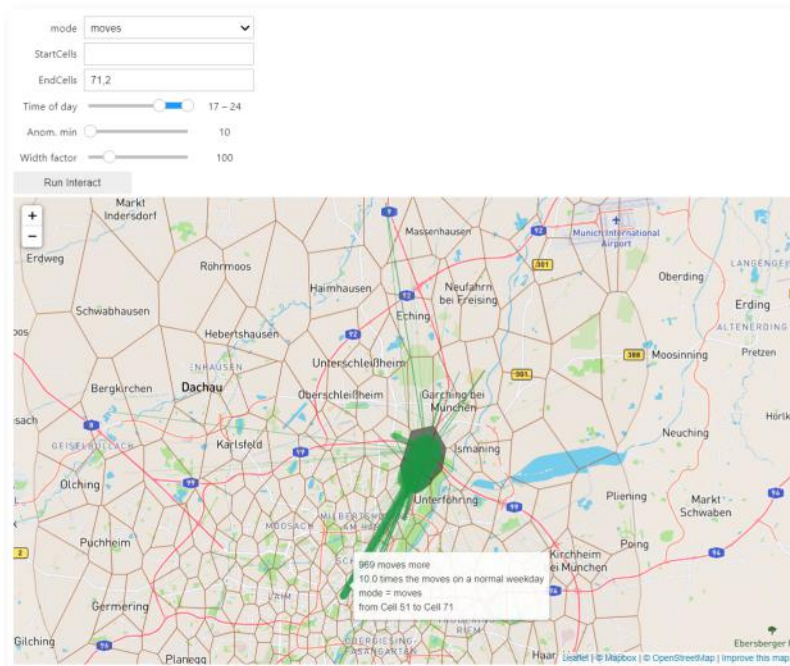


Figure 4.22: A flow map showing anomalies in the OD dataset for Wednesday with respect to the other three weekdays. Movements between 5 p.m. and midnight assigned any mode and ending at the cells containing the football stadium are plotted as green or charcoal grey lines if they differ from the averaged number by at least 10. No charcoal grey lines can be seen since there were no relations with less movement during the given times on Wednesday compared to the other days. The chosen end cells are coloured in black.

For the fourth stage of the scenario, the code<sup>26, 27</sup> is a combination of those for the other three stages but using the dataset of anomalies in the OD data (table A.11). The plotting function for the flow map (figure 4.22) is similar to that of the second stage (figure 4.18) but incorporating colours for the start- and end-cells as for the map in the third stage (figure 4.20). The temporal overview of stage four (figure 4.23), which is essentially generated equivalently to that for the second stage (figure 4.19).

<sup>25</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/3T\\_OD\\_CycleVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/3T_OD_CycleVis.ipynb)

<sup>26</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/4S\\_ODAnom\\_mapVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/4S_ODAnom_mapVis.ipynb)

<sup>27</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/4T\\_ODAnom\\_cycleVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/4T_ODAnom_cycleVis.ipynb)

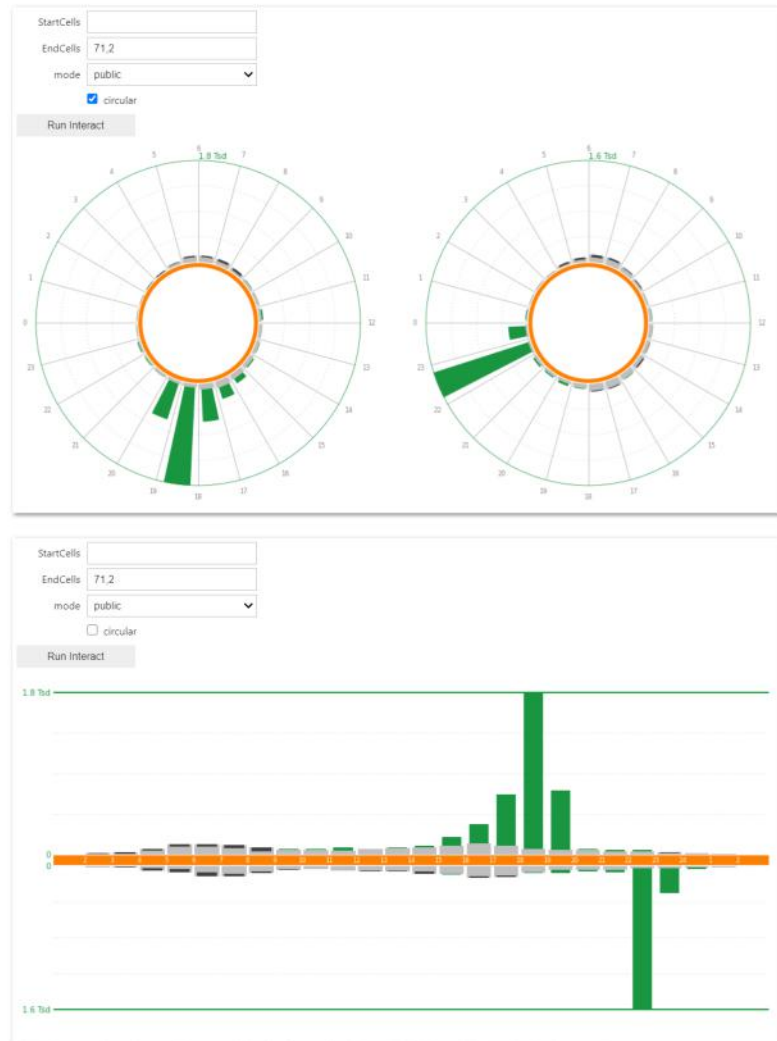


Figure 4.23: Bar charts showing Wednesday anomalies in the OD dataset for movements of any mode to (radial version: left, horizontal version: top) and from (radial version: right, horizontal version: bottom) the cells containing the stadium. Each hour is represented by one bar. Grey bars show the averaged numbers for Monday, Tuesday and Thursday. Green bars growing from the grey bars indicate positive anomalies (more movement on Wednesday), charcoal grey bars burning into the grey bars indicate negative anomalies (less movement on Wednesday).

## 4.4 User Testing

This section deals with research objective 2.2 and aims to answer the associated research questions (i) *Which methods for testing usefulness are appropriate in this specific case?* and (j) *How good is the usefulness of the visualizations?*. Research question (k) *How can the visualizations be improved to possibly increase usefulness?* will be dealt with thereafter.

### 4.4.1 The Test Set-Up

To evaluate the developed visualizations, a user test was conducted with four of the five expert users initially interviewed and a fifth substituting for one interviewee that was not available during the time of testing. A combination of a guided walk-through with sequences of free interaction during which the users were asked to think aloud, ensuing written questionnaires and a closing discussion to retrospectively summarize the individual impressions and reactions was applied.

Stage	Dataset & Vis.	Given Task	Intention
1	Flow-data, synoptic, see figure 4.7 on page 27	Play with the interface, look for any general patterns or conspicuities. Once the anomaly is detected: determine time and place of the anomaly	The users should become familiar with the interaction mechanisms and the symbolization, finding the anomaly (football event taking place on Wednesday evening)
2	Flow-data, comparative, see figure 4.8 on page 28	Choose filters, interact with visualizations, investigate the anomaly, question implications and conclusions	Test and ensure the users' understanding of the symbolization, find any misleading features of the visualizations
3	OD data, synoptic, see figure 4.9 on page 29	Figure out where people attending the football event come from and go to	Give the users the opportunity to interact further with the data after having gotten..
4	OD data, comparative, see figure 4.11 on page 30	Analyse how the anomaly is reflected in the OD dataset	..used to the interface and the types of visualizations, find any weaknesses in the visualizations, discuss quality of the dataset, understand the users' process of analysis and reasoning with the visualizations

Table 4.2: Stages within the practical part of the user test

During the verbal parts of the process, the audio was recorded, later transcribed and analysed (see Appendix: User Test Transcripts). The questionnaire was filled in individually by the users whereas the walk-through with the interactive part and the final discussion were done in two groups corresponding to the two different public transport associations. This was done to stimulate discussions and distribute the task of interaction among the individual users, so those who were not actively interacting with the visualizations at any given moment could focus on observation and reflection.

The technical set-up consisted of a computer, a projector, and a mouse. The display as seen in figure 4.24 was projected onto the wall and the users were asked to interact with the visualizations using the mouse. In four stages (see table 4.2), the test subjects were introduced to the different visualizations and the underlying dataset using the overview depicted in figure 4.25 which the participants also received a printed version of. After each introduction, the test users were asked to interact with the visualizations using the mouse while thinking aloud. Once the users felt like they had understood and explored the individual visualizations and any tasks were carried out, the next step was initiated.

At the beginning of this practical part, the users were handed the first page of the questionnaire (see Appendix: Questionnaire) to take notes and possibly start filling in the quantitative part on top of the page. While the number of test subjects is too small for any significant quantitative analysis, the quantitative questions were included in the questionnaire to encourage the users to critically reflect over the individual visualizations. After the practical part, the participants were asked to fill in the remainder of the questionnaire. Once they were done writing down their answers, a final discussion was held in which each participant was asked to summarize their thoughts, points of critique and suggestions for improvement.

After the tests, the audio recordings were transcribed and, alongside the answers to the

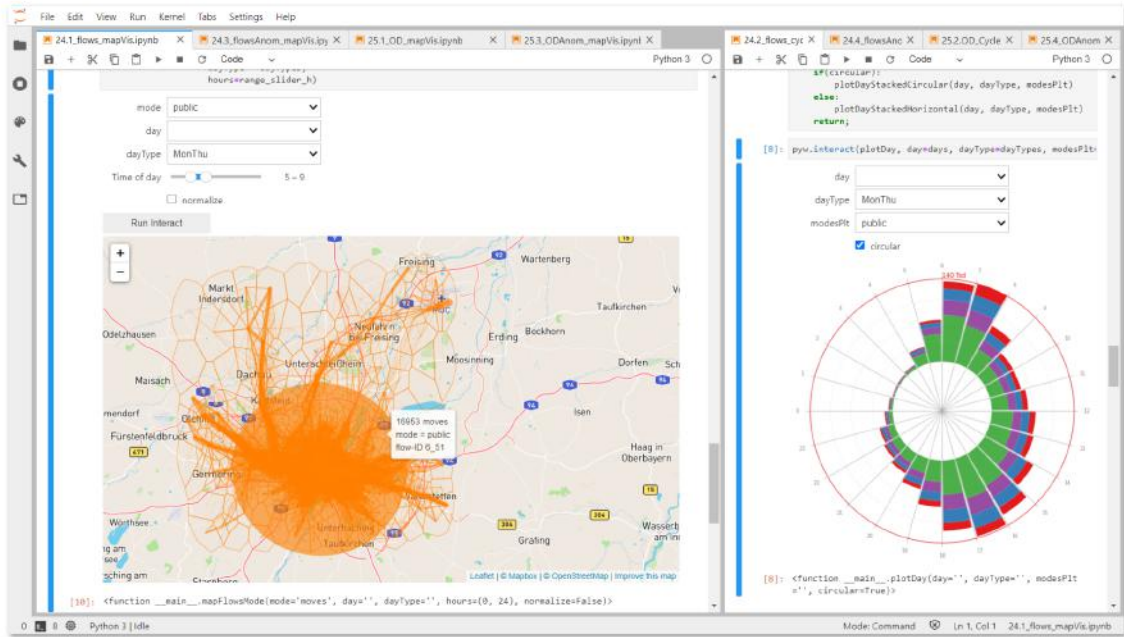


Figure 4.24: The *JupyterLab* interface used for testing. Notebooks with spatial overviews were opened on the left side, those containing temporal overviews were opened on the right. The notebook being tested at any moment was given more space by dragging the boarder sideways. For each new stage, the next two notebooks were switched to using tabs.

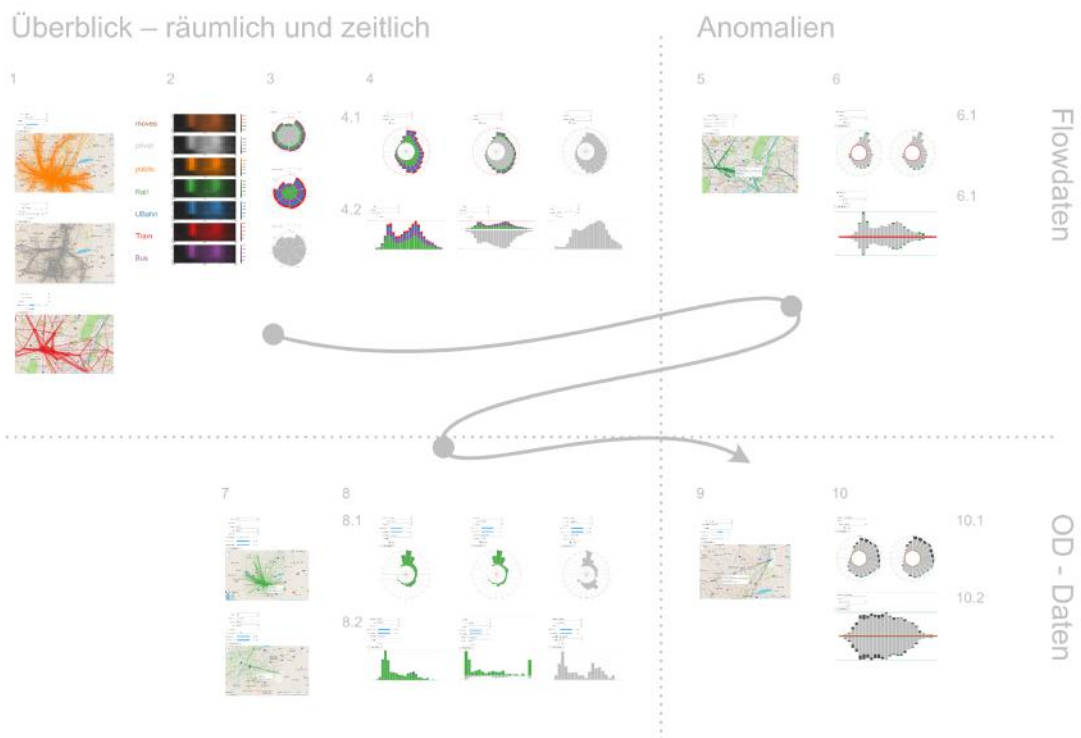


Figure 4.25: Overview of the visualizations used in the different stages. This overview was printed on A3 paper and handed to the participants during the test. A larger version can be seen in the appendix on page 165.

questionnaire analysed and summarized. The transcripts were not encoded since the number of participants was rather low, and thus a comparison between different user groups was not intended. The purpose of the analysis was rather to get expert user feedback on whether the general concept and the proposed visualizations of extended mobile network data would be of use within their field of urban public transportation planning.

#### **4.4.2 Results of the Evaluation**

In the following, the answers from the questionnaire (see Appendix: Questionnaire Results) are summarized and evaluated, following the structure of the questionnaire (see Appendix: Questionnaire). Statements from the audio recording are included to draw conclusions and implications for the usefulness of the tested visualization concept and the improvement thereof.

##### **What do you like about the visualizations?**

The expert users mentioned the beneficial temporal and spatial overview as well as an adequate level of complexity and information content. The layout and symbolization (colour scheme) was perceived as clean, clear and intuitive. The temporal and modal comparability was noted.

According to the given answers and the reactions during and after the practical part of the user test, the general concept was well appreciated.

##### **What are you missing? (functionality, filter parameters)**

A better resolution, both temporally in the data (10 minute-intervals were mentioned to be meaningful for planning purposes) as well as within the colours used for the heatmaps was requested. After switching to a light colour scheme (figure 4.15, right side), the users said they could discern the patterns better and the symbolization was more intuitive. This makes sense as the background of the interface was light which requires higher values to be represented by darker colours which is also what the users said they are used to. Improving the temporal resolution of the data is essentially a matter of ensuring a big enough amount of data to be able to aggregate to shorter intervals of time and offer a significant output while staying within privacy regulations.

The orientation of the radial bar charts was discussed; one user (A in Appendix: User Test I) proposed the option to change the orientation of the plot, another (D in Appendix: User Test II) said they liked the orientation the way it was. Adding an option to customize the orientation of the radial bar charts might be a possibility although this could lead to confusion and complicate the categorization of different temporal patterns.

Several users wished for a further modal breakdown, specifically breaking down the mode “privat” into modes for people moving on foot, bike or by car. A possibility to filter for the travelled distance even in the bar charts of the first stage of the scenario (figure 4.17) was suggested as a way to further modally break down the private movements. While a further breakdown of the private modes of transport is to be handled within the model used to assign the modes to the trips, adding aggregate information on the travelled distance or time to the flow data might be a quick fix to enhance filtering within this dataset. On the other hand, a possibly high variability of these values within data for a single spatial relation might require deeper filtering than by just one aggregate value.

The consideration of intermodality was asked for as presently each trip is assigned the mode corresponding to the majority of the trip segments. Whether it is possible to offer significant data while respecting privacy restrictions remains to be investigated.

Further processing of trips that were assigned to the “UBahn” mode was requested to improve the visualization of those movements and prevent the impression that all underground traffic actually starts or ends at the main station. This again is to be tackled in the preprocessing



of the data as on the aggregate level that was available, ID-information on the different trips was lost, which rules out reconnecting two trips to each other in order to find the true start- and end-cells of trips via the underground.

### **How would you want this visualization to be improved?**

One user (A in Appendix: User Test I) suggested visually differentiating between flow data and OD data, i.e. symbolize the movements in the flow maps for the third and fourth stages (figures 4.20 and 4.22) differently from those in the first and second stages (figures 4.14 and 4.18). While this was not requested by other users, investigating other possible visualization styles such as curved lines instead of straight ones might be beneficial.

The issue of visual clutter within the maps was brought up and discussed at several points during both tests. One of the suggested ways to improve it was the filtering for important connections, i.e. filter out flows with a small number of movements as done for the anomalies. Matching movements to the transport network was suggested by user D. A further possibility would be customizing the cells to which the data is being aggregated, which would also solve the problem of movements being seemingly misassigned. This was discussed when looking at private traffic in the OD dataset to and from the cell containing the stadium where the football match took place. Private movement took place even during the game which implied that, as the motorway crosses the same cell, the cell might have been registered as the destination of those trips due to a traffic jam in the area without the trips actually ending there. Dividing this cell into two separate ones or, more generally, introducing separate cells for the motorway could help resolve this type of misassignment.

All participants of the first test (A, B, C in Appendix: User Test I) agreed that the peak hour in the circular bar charts should be indicated more clearly and suggested moving the number indicating the maximum to the bar corresponding to it. User C suggested including the total daily number of plotted movements within the bar charts. All of these suggestions were implemented in the notebooks after the user tests.

User A asked for further supporting information in the form of tooltips or legends. Both a more extensive legend and further tooltips, especially for the bar charts, are possible within an actual user interface. Using the currently available packages within the given timeframe, tooltips on the bar charts were not possible. Several attempts were made, e.g. using the `mpld3` package which offers tooltips for general plots but does not support radial plots at this point.

The representation of intermodal travel chains, a higher temporal resolution and a bigger temporal extent of the data were mentioned. All of these points have to be tackled within the gathering and preprocessing of the data.

### **What kind of questions or problems within your field of work could this type of visualization help solve?**

The users named a number of possible areas of application including those that, within the initial expert interviews, were stated to be outside the scope of what is possible within the currently used framework. This indicates that the proposed concept would meet user needs that are not being met by the current framework.

### **Do you prefer the radial or the horizontal version of the bar charts or do you think both serve different purposes? Why? Which purposes come to mind?**

Several users said that they generally prefer the horizontal version of the bar charts due to its commonness and readability. Two users (C and D) said that they see different purposes for both versions and found the circular version to be very intuitive and concise after having gotten used to reading it. User D found the circular version useful for the analysis of individual cases while

they preferred the horizontal version for comparing movement in opposite directions or within different modes (see bottom parts of figures 4.17b, 4.19 and 4.23 as well as figure 4.21b). All participants of the first test (A, B, C in Appendix: User Test I) agreed that when plotting public modes along with the private mode in bar charts (middle part in figure 4.17a and bottom part in figure 4.17b), the horizontal version is clearly better than the radial one as it allows for both the quantitative comparison between two different subsets of the data as well as the temporal analysis of each of the parts. While the preference for the horizontal version was dominant, the possibility of switching between the two versions was welcomed. The feedback shows that, for comparative tasks, presenting the data within one chart is more effective than juxtaposition of two charts, even if the type of chart might offer a more intuitive representation when investigated individually.

### 4.5 Implementation of Test Results

This section deals with the implementation of results concerning research question (k) *How can the visualizations be improved to possibly increase usefulness?* The changes which have already been made in the notebooks will be demonstrated, followed by a description of possible further development that touch on both the processing of data and its visualization.

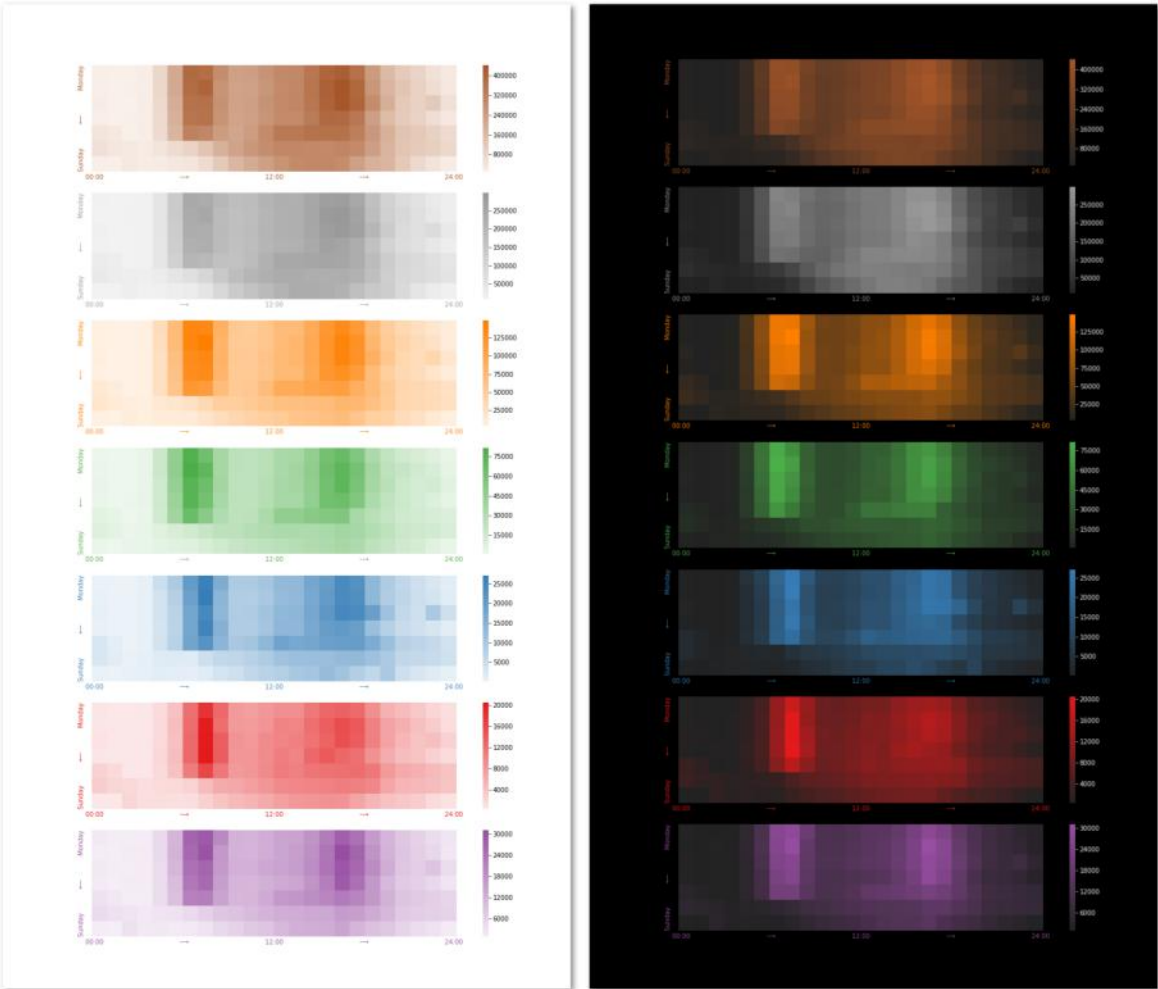


Figure 4.26: Heatmaps with a changed background-colour.

### 4.5.1 Changes in the Notebooks

The heatmaps in the first stage and the radial bar charts in the first and second stage have been modified<sup>28, 29</sup> to incorporate some of the results of the user tests. The radial bar charts in the third and fourth stage can be adapted accordingly.

For the heatmap, the choice between a dark and a light colour scheme can now be made interactively. A dark background as it would appear in a dark interface has been added (see figure 4.26).

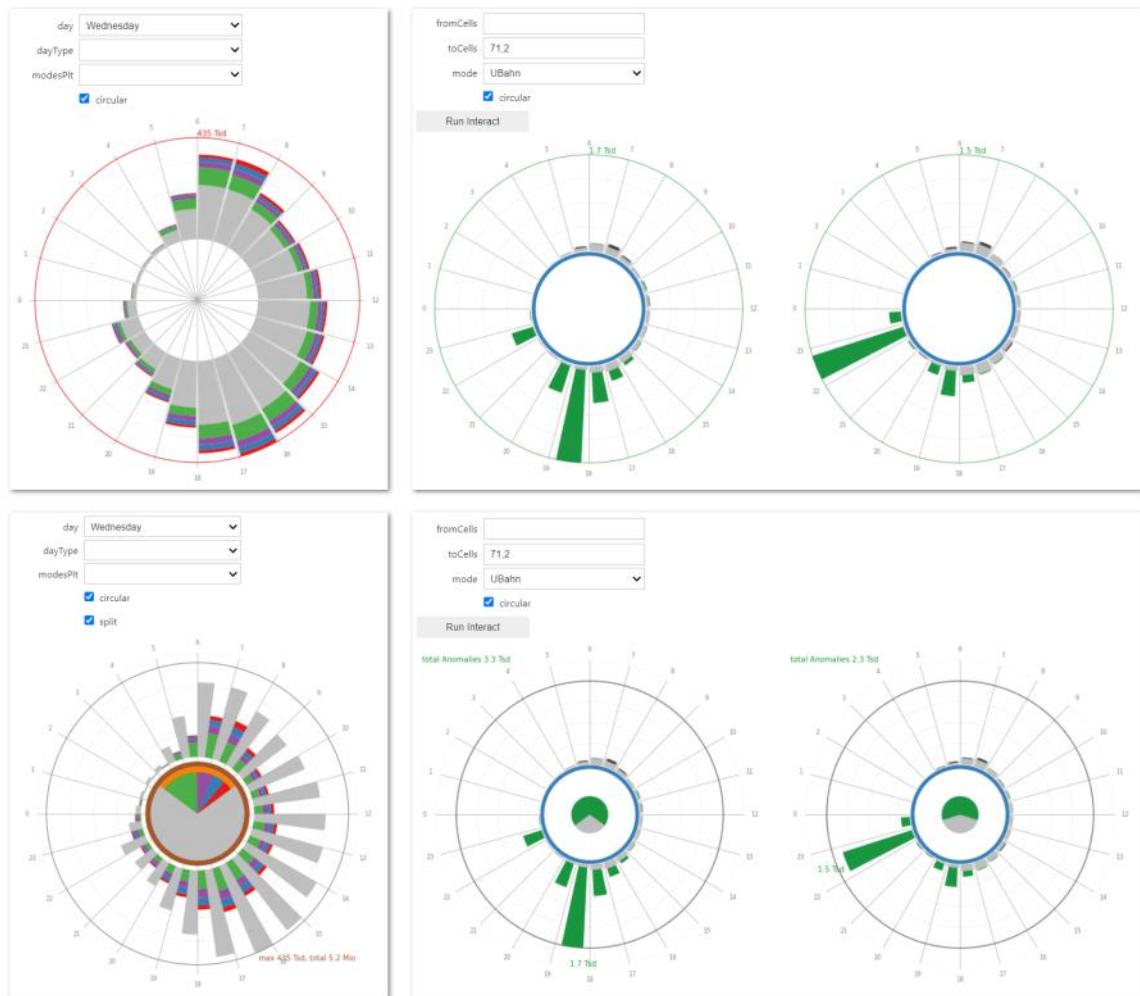


Figure 4.27: Radial bar for absolute numbers (left) and anomalies (right) charts for user testing (top) and after the incorporation of changes based on the test results (bottom).

A small pie chart was added to the radial bar charts. Its area represents the total number of movements during the plotted day, respecting Stevens's law (Stevens, 1957) as described by Munzner (2014) (2014, 5.5 and 10.4).

For the radial bar chart showing flow data (figure 4.27, left side), the pie chart is divided and coloured according to the parts that movements of the different modes occupy. The area of

<sup>28</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/1T\\_postUserTest\\_flows\\_cycleVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/1T_postUserTest_flows_cycleVis.ipynb)

<sup>29</sup>see the notebook at [https://github.com/klavere/xMNDvis/blob/master/2T\\_postUserTest\\_flowsAnom\\_cycleVis.ipynb](https://github.com/klavere/xMNDvis/blob/master/2T_postUserTest_flowsAnom_cycleVis.ipynb)

the pie chart depends on the total number with the highest possible number taking up all free space in the middle of the chart. The maximum number of movements is now plotted at the bar corresponding to that maximum. The total number of movements is plotted next to the maximum number in the colour corresponding to the plotted modes. When plotting both the public and private modes, it can be chosen that those are split similarly to the horizontal version. To keep the visually perceived quantity for both modes as equal as possible, the bars are reduced to half their width, starting at a common ground. It has to be kept in mind that even though the bars now visually only span half an hour, they still represent values for the full hour.

In the radial bar chart showing anomalies (figure 4.27, right side), the area of the pie chart depends on the dimension of the total daily number of anomalies. The light gray portion represents the average number of movements on the three weekdays taken for reference while a green or charcoal grey portion represents the summed anomalies throughout the day. The maximum is plotted next to the bar corresponding to it while the day's total number of anomalies is plotted on the top left of each plot. Plots for two directions now have the same scale, making it easier to see in which direction the numbers are higher.

#### 4.5.2 Further Possible Changes

An issue that should be tackled is the normalization of the flow line width within the flow maps (figures 4.14, 4.18, 4.20 and 4.22). It is important to keep the line widths dynamic, i.e. dependent on the values to be plotted, to account for subsets with comparatively low or high numbers and ensure the readability of those. Plotting such subsets with a constant function for the line width would render some lines very thin or too wide and thus negatively impact readability. On the other hand, comparability between plots with different chosen filter options has to be ensured. Incorporating even deeper interactivity here could both enhance the visualizations and make the user aware of the line width's dependence on certain pre-set values. It should also be considered whether to introduce a function for the line widths that keeps the line widths within certain constraints. This would ensure that all lines are well visible but it might impact the detection of outliers. When showing temporally relative flow numbers (e.g. the flow map in the first stage with checked "normalize"-option), a simple solution could be a discrete classification in a small number of classes. This would, of course, decrease the information content that can be gained visually. In the case of the flow maps showing OD data (figures 4.20 and 4.22), where filtering for specific spatial relations leads to the display of only a few lines, an improvement could be to keep all lines plotted and only highlight those corresponding to the chosen filter options (similar to figure A.6). While possibly increasing visual clutter, this would also give more contextual information.

The issue of visual clutter and overlapping lines could also be tackled earlier when processing the data. Clustering movements or matching them to the traffic network would lead to a cleaner and tidier visual output. However, more assumptions would have to be made and the output would be farther from the raw data-structure which could support drawing false conclusions.

As for the heatmaps (figure 4.26), an alternative that would facilitate quantitative reasoning and comparability while keeping the overview over all data intact could be horizon graphs as proposed by Heer, Kong und Agrawala (2009). However, horizon graphs might not be very meaningful given the current temporal resolution of one hour and would be used more effectively once the resolution can be increased.

Other issues to be tackled within the processing of the data are the reconnecting of split trips assigned to the "UBahn"-mode, the customization of aggregation cells and further differentiation within the "privat"-mode. Travel chains should be broken down and assigned multiple modes if feasible and possible within privacy restrictions. Information on the uncertainty connected to assigning a certain mode to a trip should be provided and visualized.

Finally, the visualizations have to be incorporated into a cohesive user interface, including complete supplementary information in the form of a legend and tooltips. Further interactivity and filters should be added to facilitate dynamic spatial aggregation and deeper filtering of OD data.

## 5 Discussion

In this chapter, the results from the case study presented above are discussed with regard to the scientific background. Implications for a further development are described along with the limitations of the proposed concept.

### 5.1 Results with Regard to the Scientific Background

#### Visualization of Spatiotemporal Data

Starting from insights from Visual Analytics today, taking a look at how (spatio-)temporal data has been visualized through time provided a valuable foundation for the development of the visualizations proposed in this study. Literature showed that the graphic representation of temporal data has been a rather modern development compared to the long history of cartography. However, it also became apparent that many ideas on how to visualize spatiotemporal data that are commonly used today have been around in some form for several centuries. Adapting those ideas to new interfaces and digital technologies remains a challenging but exciting endeavour. Two dominating factors are the automation of the generation of visualizations and the vast possibilities present when it comes to interactive elements. Incorporating these elements in a careful and effective way is being studied within the field of Visual Analytics which keeps delivering valuable insights. Putting those insights to use within different professions has been lacking. This study has been an attempt to help bridging the gap between Visual Analytics and the Public Transportation sector.

#### Visualizations of Mobile Network Data in Urban Public Transportation Planning

The expert user interviews conducted for this study showed that, as anticipated, few visualizations are currently being used in the process of urban public transportation planning within the study area and interactive elements are limited. However, the recent introduction of a dashboard solution based on mobile network data shows that there is a general interest for the incorporation of new kinds of data and visual applications. The results from the user tests and questionnaires show that incorporating extended mobile network data in the form of attributed flow data through interactive visualizations can help meet user needs that are being neglected in the current framework. Looking at mobile network data will in particular help gain *timely knowledge* on movements happening *beyond public transportation* and movement patterns on *days with unusual conditions* such as fairs or sport events which attract exceptionally high numbers of people and are therefore important to be considered in the planning process. Modelling the *mode of transportation* on mobile network data has been shown to substantially enhance the qualitative movement information that can be gained. Knowing *how* people move rather than only *where* lets planners understand the data at a deeper, more meaningful level which will help them make more appropriate changes in the transportation network.

In order for the data to be usable, applying lessons from user-centred design in the process of developing the visualizations of such data has been essential in this study. The valuable feedback gained from testing the proposed concept with expert users will help in further development.

## 5.2 Limitations of the Proposed Visualization Concept

The proposed concept is limited by both the amount and quality of the underlying data as well as its prototypic nature. On the data side, issues such as the availability of data in a useful temporal and spatial resolution have to be addressed. The uncertainty introduced along the data processing pipeline, specifically when modelling the different modes of transport, has to be determined and visually communicated. As for the visualization of the data, several things need to be improved and tested iteratively in order to reach a well-functioning product that is of high value for the planning process. A major issue is the dynamic adjustment of the width of lines in the flow maps. Different possibilities for improvement have been described in section 4.5.2 (page 50). The insights gained from user testing suggest that increasing the user's control over line width might be a practicable solution but good default values have to be found to facilitate readability and comparability in standard visualizations. A complementary quantitative study might help in finding pitfalls introduced by dynamic line widths. An alternative option for the visualization of OD data might be the use of diagram maps instead of flow maps as proposed by Andrienko, Andrienko, Fuchs und Wood (2017). Incorporating temporal overviews in the form of diagrams directly within the maps would further interconnect the temporal and spatial dimensions of the movement data and could help with the detection of anomalies or clusters.

## 6 Conclusion & Outlook

In this study, a visualization concept for attributed movement data stemming from mobile network data for the use within urban public transportation planning was developed and evaluated.

To this end, the development of a data visualization concept based on a review of relevant literature was embedded in a user-centred design cycle. Expert interviews were conducted to assess user requirements and user-tests were carried out in small groups of professionals whose individual feedback was additionally evaluated via a short questionnaire and afterwards discussed in the group. The qualitative data from the interviews acted, together with insights from the literature review, as a foundation for the development of the visualization concept. The tests, questionnaires and ensuing discussions were used to evaluate its usefulness, consisting of the usability of the visualization concept and the utility of the visualized dataset.

It was shown that the incorporation of attributed flow data stemming from mobile network data can help improve several issues of the current framework used for planning and would therefore be of great value. The developed visualization concept for such data was well received by the professionals. However, as only the first iteration of a user-centred design cycle could be realized within the time allocated for this research, several improvements need to be incorporated and evaluated in further cycles in order for a well working product to be developed.

This study took a step in connecting the fields of Visual Analytics and Public Transport Planning. Further collaboration is encouraged to make the application of recent findings within Visual Analytics fruitful for urban transportation and thus contribute to an efficient, climate-friendly and resilient future of urban regions.



## Bibliography

- Andrienko, G., Andrienko, N., Bak, P., Keim, D. & Wrobel, S. (2013). *Visual Analytics of Movement*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R. & Zhao, Y. (2017). Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions. *IEEE Transactions on Intelligent Transportation Systems*, 18 (8), 2232–2249.
- Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., ... Tominski, C. (2010). Space, time and visual analytics. *International Journal of Geographical Information Science*, 24 (10), 1577–1600.
- Andrienko, G., Andrienko, N., Fuchs, G. & Wood, J. (2017). Revealing Patterns and Trends of Mass Mobility Through Spatial and Temporal Abstraction of Origin-Destination Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, 23 (9), 2120–2136.
- Battersby, S. E., Finn, M. P., Usery, E. L. & Yamamoto, K. H. (2014). Implications of Web Mercator and Its Use in Online Mapping. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 49 (2), 85–101.
- Brewer, C. A. (2013). *ColorBrewer: Color Advice for Maps*. <https://colorbrewer2.org/#>.
- Büttner, B., Kinigadner, J., Ji, C., Wright, B. & Wulfhorst, G. (2018). The TUM Accessibility Atlas: Visualizing Spatial and Socioeconomic Disparities in Accessibility to Support Regional Land-Use and Transport Planning. *Networks and Spatial Economics*, 18 (2), 385–414.
- Cui, W. (2019). Visual Analytics: A Comprehensive Overview. *IEEE Access*, 7, 81555–81573.
- Field, K. (2020). *Mapping coronavirus coxcombs*.
- Great Britain. General Register Office, Farr, W., Simon, J., St. Thomas's Hospital. Medical School Library former owner & King's College London. (1852). *Report on the mortality of cholera in England, 1848-49 [electronic resource]*. London : Printed by W. Clowes.
- Heer, J., Kong, N. & Agrawala, M. (2009). Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (S. 1303). Boston, MA, USA: ACM Press.
- Jenny, B., Stephen, D. M., Muehlenhaus, I., Marston, B. E., Sharma, R., Zhang, E. & Jenny, H. (2018). Design principles for origin-destination flow maps. *Cartography and Geographic Information Science*, 45 (1), 62–75.
- Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J. & Ziegler, H. (2008). Visual Analytics: Scope and Challenges. In S. J. Simoff, M. H. Böhlen & A. Mazeika (Hrsg.), *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics* (S. 76–90). Berlin, Heidelberg: Springer.
- Kulpa, Z. (1997). Diagrammatic Representation of Interval Space in Proving Theorems about Interval Relations. , 9.
- Minard, C.-J. (1863). *Cartes figuratives*. Bibliothèque numérique patrimoniale des ponts et chaussées.
- Minard, C.-J. (1870). *Tableaux graphiques et cartes figuratives*. Bibliothèque numérique patrimoniale des ponts et chaussées.
- Munzner, T. (2014). *Visualization Analysis and Design*. A K Peters/CRC Press.
- Nielsen, J. (1993). *Usability engineering*. Boston, Mass. u.a.: AP Professional.
- Nielsen, J. (1994). Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies*, 41 (3), 385–397.
- Norman, D. A. (2013). *The design of everyday things* (Revised and expanded edition Aufl.). New York, New York: Basic Books.

- Phan, D., Xiao, L., Yeh, R. & Hanrahan, P. (2005). Flow map layout. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* (S. 219–224).
- Phillips, Y. F. (2018). *Analysis and visualisation of very-long-duration acoustic recordings of the natural environment* (Phd). Queensland University of Technology.
- Popova, M. (2017, Oktober). *W.E.B. Du Bois's Little-Known, Arresting Modernist Data Visualizations of Black Life for the World's Fair of 1900.*
- Priestley, J. (1772). *The History And Present State Of Discoveries Relating to Vision, Light, And Colours.* Johnson.
- Projekt xMND | Telefónica Deutschland. (2018). <https://www.telefonica.de/analytics/anonymisierte-daten/projekt-xmnd.html>.
- Projekt xMND: Wie Mobilfunkdaten die Verkehrsmodellierung revolutionieren könnten | civity Management Consultants. (2018). <https://civity.de/de/news/2018/12/projekt-xmnd-wie-mobilfunkdaten-die-verkehrsmodellierung-revolutionieren-k%C3%B6nnten/>.
- Qiang, Y., Delafontaine, M., Asmussen, K., Stichelbaut, B. & de Weghe, N. V. (2010). Modelling imperfect time intervals in a two-dimensional space. , 28.
- Rendgen, S. (2018). *The Minard System: The Complete Statistical Graphics of Charles-Joseph Minard.* New York: Princeton Architectural Press.
- Rosenberg, D. & Grafton, A. (2012). *Cartographies of Time: A History of the Timeline.* London: Abrams & Chronicle Books.
- Schubert, A. (2019). *Verkehr in München - MVG will Handydaten nutzen.* <https://www.sueddeutsche.de/muenchen/mvg-app-digitalisierung-1.4351183>.
- Schutz, B. (2009). *A First Course in General Relativity.* Cambridge.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64 (3), 153–181.
- te Brömmelstroet, M., Skou Nicolaisen, M., Büttner, B. & Ferreira, A. (2017). Experiences with transportation models: An international survey of planning practices. *Transport Policy*, 58, 10–18.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2. ed., 2. printing Aufl.). Cheshire, Conn.: Graphics Press.
- Willard, E. (1846). *The Temple of Time* [Timeline]. New York: A.S. Barnes & Co.
- Yang, Y., Dwyer, T., Jenny, B., Marriott, K., Cordeil, M. & Chen, H. (2019). Origin-Destination Flow Maps in Immersive Environments. *IEEE Transactions on Visualization and Computer Graphics*, 25 (1), 693–703.

# Appendix

Interview Transcriptions .....	58
Interview I .....	58
Interview II .....	69
Interview III .....	78
Structured Collection of Main Statements from Interviews .....	94
Screenshots from Interviews .....	98
Data Tables .....	104
User Test Transcripts .....	110
User Test I .....	110
User Test II .....	138
Overview of the Visualizations used for the User Tests .....	165
Questionnaire .....	166
Questionnaire Results .....	168

All Jupyter Notebooks written for Data Handling and Data Visualization can be found here:  
<https://github.com/klavere/xMNDvis>

The data used for this thesis will be available online at:  
<https://mcloud.de>

# Interview Transcripts

## Interview I

*Sie arbeiten mit der Datengewinnung und -analyse*

Genau richtig. Wir sind das Team Daten bei uns in der Mobilitätsforschung. Wir kümmern uns um die Datenerhebung, hauptsächlich Fahrgastzählungen mit technischen Zählungen aber auch manuelle Zählungen und auch Daten aus anderen verschiedenen Quellen. Wir haben schon versucht, Daten zu bekommen über Mobilfunkanbieter, haben da Daten ausgewertet oder Bewegungsdaten aus Apps für Mobilfunkgeräte, haben erste Kontakte jetzt auch zu Google und Apple um Standortdaten aus den GPS-Informationen der Smartphones zu bekommen, allerdings haben wir da noch überhaupt nichts gesehen, da sind wir noch am Anfang der Gespräche, das wird sich noch ein bisschen hinziehen.

the interviewee is working within the department for mobility research responsible for providing data used for planning and modelling. Different data sources have been tested, e.g. mobile network data

*Gut, also die Daten mit denen ich arbeite sind auch aus Mobilfunkdaten von dem xMND-Projekt*

Das sind ja die Telefonica-Daten

*Ja, genau. Was haben Sie noch für Daten außer den Bewegungsdaten?*

Also wir haben auch mit den Daten von der Telekom schon gearbeitet, das war damals Motion Logic. Diese Firma, weiß ich gar nicht obs die noch gibt, ist auf alle Fälle dabei, dass es die demächst nicht mehr gibt. Ob dieses Modell "Datenverkauf" bei der Telekom weitergeführt wird weiß ich nicht. Ob Vodafone sowas hat haben wir bisher nicht mitbekommen, ich glaube nicht.

*Ich wollte jetzt eher darauf hinaus, was unabhängig von Bewegungsdaten für Inputs kommen*

Also die anderen Inputs sind Daten, ich denke mal 95% aus technischen Zählgeräten, das sind Infrarotzählgeräte, die über den Türen angebracht sind, also Sensoren. Etwa 50% unserer Busse und 30% der Trambahnen haben solche Zählgeräte eingebaut und die erfassen einfach nur die Anzahl der Ein- und Aussteiger je Halt. Daraus berechnet sich relativ einfach, wie viele Leute sich gerade im Fahrzeug befinden, wir können aber nicht sagen, von wo nach wo die einzelne Person fährt, sondern wir wissen nur wie viele Leute zwischen zwei Haltestellen unterwegs sind. Um zu erfahren, wie viele Leute tatsächlich von wo nach wo fahren, kombinieren wir diese Zähl- und Befragungsdaten. Das heißt, wir befragen immer wieder mal, zum Teil in einem geographisch begrenzten Gebiet, zum Teil auch im Gesamtnetz, unsere Fahrgäste, wann sie mit welchem Verkehrsmittel, mit welcher Linie, mit welchem Vor- und Nachlauf, also inklusive aller Umstiege, von wo nach wo fahren.

Most of the used data stems from automatic counting devices mounted above the doors of about half of the busses and a third of the tramways. Further data which includes information on origin, lines used and destination is gathered manually via passenger surveys.

*Und das wird dann hochgerechnet*

Das wird dann hochgerechnet auf Basis der Zähl­daten. Das hat den Vor­teil, dass wir da noch Zusatz­infor­mationen bekommen: Ticketart, Alter, Ge­schlecht, um dann eben die Hoch­rechungen noch mal genauer machen zu können, oder auch um dann zum Beispiel im Marketing genauer auf die [...] zugehen zu können.

Data from passenger surveys is additionally used for marketing and information purposes

*Ok. Wie werden diese Informationen dann aufbereitet, bzw. was ist der Zugang zu den Informationen*

Also die Daten werden zum Teil in unsere Verkehrsmodelle gespielt. Das Verkehrsmodell ist ein VISUM-Modell, VISUM ist ein Softwareprodukt der Firma ptv. In diesem Modell kann mal einmal [stathetisch?] berechnen, wie sich die Fahrgäste verhalten, auf Basis von Annahmen und Algorithmen. Man kann aber auch tatsächlich erhobene Daten, also diese gezählten und hochgerechneten Befragungsdaten einfach in diesem Verkehrsmodell darstellen. Das nutzen wir auch, um da dann die Bewegungen in München für einen Tag darstellen zu können.

The data feeds into a VISUM traffic model which simulates traffic flows

*Ok, also es sind immer Daten von einem Tag?*

Es wird in der Regel auf einen Tag hochgerechnet, wir arbeiten in der Regel mit dem Wochentagstyp "Montag-Donnerstag-Schule", wir gehen davon aus, dass innerhalb dieser vier Tage einer Woche sich die Fahrgäste ungefähr gleich verhalten. Beim Montag ist es ein bisschen fraglich, weil es auch viele Wochenendpendler gibt, allerdings sind die eher im Regionalverkehr relevant als im innerstädtischen Verkehr. Im innerstädtischen Verkehr schlagen die trotzdem alle irgendwann am Arbeitsort, wo sie halt hinfahren wollen, auf. Und das machen sie halt Montag und Dienstag und jeden anderen Tag auch. Der Freitag ist immer eine Ausnahme, weil es tatsächlich frei haben, beziehungsweise Freitags früher Schluss machen. Deswegen klammern wir den einfach aus.

Temporal evaluation throughout a day is of interest, while the data is aggregated by different types of days. The most used type of day is "Monday-Thursday-School" as this is where the load is highest and relatively similar movement during those days can be assumed.

*Ok und das Wochenende wird auch nicht..*

Wochenende wird dann auch ausgeklammert. Wir haben aber auch Befragungsdaten fürs Wochenende, für Samstag und Sonntag, und auch Befragungsdaten für Ferientage, die auch wieder ganz anders sind, weil einfach der komplette Schülerverkehr fehlt. Aber die wichtigsten Tage sind für uns Montag-Donnerstag-Schule, weil das auch relevant ist für die Fahrzeugbeschaffung, weil das einfach der Maximalverkehr ist. Also wir müssen die Anzahl der Fahrzeuge und die Anzahl der Fahrer die wir haben am Maximum natürlich ausrichten.

data for fridays and weekends is gathered as well but are not as relevant as the load which also determines the needed number of vehicles is not as high on those days

*Es läuft also darauf hinaus, die korrekte bzw. passende Anzahl der Fahrer und Fahrzeuge zu ermitteln, als das ist das Hauptziel. Oder gibt es da noch mehr?*

Anzahl der Fahrzeuge und Fahrer aber auch natürlich dann runtergebrochen auf die Linienführung und das Streckennetz muss ja auch zur Nachfrage passen.

The main goals are to determine the number of needed vehicles and drivers as well as provide a network of lines that fits demand and travelled routes. This is more flexibly done within the bus network than within types of transport dependent on railways

*Also das heißt, welche Punkte müssen miteinander verbunden werden.*

Genau, richtig. Und da sind wir natürlich beim Schienenverkehr nicht sehr flexibel aber beim Busverkehr sind wir natürlich sehr flexibel. Da können wir relativ einfach Tangenten einführen oder ähnliches.

*Wie sieht es mit der zeitlichen Auflösung innerhalb eines Tages aus?*

Wir versuchen es fahrtenscharf. Die Befragungen klappen nicht ganz fahrtenscharf, da versuchen wir jede zweite Fahrt zu befragen. Das heißt also beim Fünf-Minuten-Takt haben wir dann ein zehn-Minuten-Raster, das aber dann für unser VISUM-Verkehrsmodell, wo der komplette Fahrplan hinterlegt ist, dann wieder auf die Fahren verschmiert wird. Das heißt also im Ergebnis haben wir wieder eine fahrtenscharfe Darstellung, die aber auf einer halb so großen Stichprobe beruht.

Temporal resolution is guided by the trip cycle which usually varies between five and ten minutes

*Ok. Also fahrtenscharf heißt jede einzelne Fahrt*

Genau. Zum Teil haben wir Zwei-Minuten-Takt, dann haben wir halt in jeder Fahrt Fahrgäste drin im Modell, auch wenn wir nicht jede Fahrt befragt haben.

*Haben Sie das Programm da, können Sie mir das zeigen?*

Das kann die [Praktikantin] zeigen nachher dann. Ich hab das nicht auf meinem Rechner installiert.

*Zu den räumlichen Auflösungen - da geht es wirklich um Stationen?*

Genau, das komplette Netz wird dargestellt.

Spatial resolution within the traffic model is determined by the route network.

*Was haben Sie da für Erfahrungen gemacht, mit den Mobilfunkdaten und deren Genauigkeit?*

Also wir haben vor zwei bis drei Jahren mit den Mobilfunkdaten der Telekom gearbeitet, die sind nicht stationsgenau, die gehen auf Funkzellen. Diese Funkzellen atmen, das heißt sie sind nicht immer gleich groß. Trotzdem hat man versucht auf relativ kleine Einheiten zu kommen. Wir haben damals gesagt, wir würden uns gerne an den sogenannten Verkehrszellen orientieren, die wir auch in unserem Verkehrsmodell hinterlegt haben. Da ist die Stadt München eingeteilt in 1000 Verkehrszellen, das Umland in weitere 1000.

The interviewee is aware of the rather coarse spatial resolution of typical mobile network data. In earlier attempts to work with such data, so-called traffic cells have been proposed as a spatial segmentation. In the traffic model, there are 2000 cells within the network of the region of Munich, 1000 of which are within the city. No more than one underground stop is within each of these cells.

*Die sind dann ein bisschen größer als Stationen nehm ich an?*

Ja, in einer Verkehrszelle sind auch mal zwei, drei Bushaltestellen, nie mehr als eine U-Bahn-Haltestelle. Das hat mittelmäßig funktioniert, vor allen Dingen auch weil natürlich viele Fahrgäste im U-Bahn-Netz unterwegs sind und da auch umsteigen, von einer U-Bahn in die andere, allerdings unser U-Bahn-Netz, ich glaube, in insgesamt nur vier oder fünf Funkzellen eingeteilt ist.

Das heißt also da kriegt ich kein sehr präzises Raster, was die Leute da unten machen.

*Verstehe. Waren die Daten dann anderweitig nützlich, also gab es andere Fragestellungen, die dann damit angegangen werden konnten?*

Ja, wir können uns einige Fragestellungen vorstellen, die wir mit Mobilfunkdaten beantworten können. Also uns würde sehr häufig schon genügen, so ungefähre Bewegungsdaten eines Tages zu haben, vor allen Dingen für Tage, die nicht normal sind. Sprich, Oktoberfest, Fußballspiele, ähnliche Sonderveranstaltungen, die zwar immer wiederkehren, die aber nur wenige Tage umfassen, sodass wir mit unseren technischen Zählgeräten eine sehr schwache Datenlage haben. Zudem kommt, dass wir mit unseren Zählgeräten natürlich nur die Fahrgäste, die bei uns in unseren Fahrzeugen unterwegs sind. Was uns fehlt ist der komplette motorisierte Individualverkehr, der Fahrradverkehr und auch die S-Bahnen.

*Das heißt potenzielle Fahrgäste*

Potenzielle Neukunden, oder auch tatsächlich Möglichkeiten - Also wir sind ja zwar natürlich auch ein Wirtschaftsunternehmen aber auch ein Unternehmen der Daseinsvorsorge, wo wir sagen würden, es geht auch um Umweltschutz und darum, Leute mitzunehmen, die sich kein Auto leisten können oder die aufgrund von irgendwelchen körperlichen oder geistigen Einschränkungen nicht in der Lage sind, Auto zu fahren, die dann eben auf den öffentlichen Verkehr angewiesen wären, und auch die sollen ein sehr gutes Angebot haben. Und da geht's natürlich darum, Strecken oder Verbindungen zu finden, die bisher nicht so gut bedient sind und die trotzdem eine Nachfrage hätten.

*Ok. Das heißt, um es nochmal zusammenzufassen: Die Fragestellungen, die im Moment nicht beantwortet werden können, und wo Sie hoffen, mit Mobilfunkdaten was erreichen zu können, sind zum einen besondere Tage, also Wochenende, besondere Festtage und so weiter, und zum anderen der nicht-hauseigene Verkehr sozusagen.*

Genau. Und bei der U-Bahn wäre es natürlich toll, weil wir da eben noch am Anfang der Ausrüstung mit technischen Zählgeräten sind, die haben bisher sehr sehr wenige technische Zählgeräte. Das wird noch ein zwei Jahre dauern, bis wir da eine gute Anzahl haben. Da wäre es natürlich auch schön, wenn wir da jetzt schon mehr über die Fahrgäste wissen würden. Aber da haben wir tatsächlich die Einschränkung, dass die Funkzellgröße im U-Bahn-Netz nicht sehr gut ist.

*Ok, das heißt, U-Bahn ist so ein Thema was Nachholbedarf hat sozusagen.*

Ja, ganz klar.

*Ok. Mit den gröberen Daten, also die räumlich auf jeden Fall eine gröbere Auflösung haben, die aus den Mobilfunkdaten kommen, welche zeitliche Auflösung innerhalb eines*

Questions that are hoped to be answered with mobile network data are: 1. Data on days with an abnormal use of transportation, such as days with fairs, festivals, football games and other special events that occur repeatedly but only cover a few days which limits the amount of data that can be gained from counting devices. 2. Data on the movement of people not using public transportation as the aim is to provide an adequate network for all people.

More knowledge on travellers using the underground would be helpful since there is presently only a limited amount of counting devices in these vehicles

*Tages wären da schon hilfreich, was wäre da wünschenswert?*

Also ein Halb-Stunden-Rhythmus ist das Minimum was wir brauchen. Gerade für solche Ereignisse wie Fußball, da kommen die Leute alle auf einmal und wenn wir das Raster da zu grob machen, dann nivelliert sich der Peak. Das nützt uns dann nichts, sondern da bräuchten wir es relativ genau. Besser wäre natürlich sowas wie ein fünf-Minuten-Raster aber das lässt die Datenmenge natürlich gleich ganz schnell explodieren.

*Ja und zum anderen hat man dann Probleme damit, dass es nicht aussagekräftig ist wenn man nicht genügend Daten hat.*

Genau, und da kommen wir dann irgendwann auch in die Datenschutzproblematik rein. Es ist wohl so, dass jedes Mobilfunkunternehmen da unterschiedliche Herangehensweisen hat; Soweit ich weiß, waren es bei Telefonica fünf Personen, die hatten dafür noch eine andere Einschränkung. Bei der Telekom waren es 30 Personen, die hatten dann eine andere Einschränkung dafür nicht. Ich weiß jetzt aber auch nicht mehr genau..

*Ja also ich arbeite aufgrund von Datenschutz eben auch auf einem relativ aggregierten Niveau*

Ja, das ist auch in Ordnung. Wir lassen ja keinen Bus wegen fünf Leuten fahren, in der Regel, oder versuchen es zumindest nicht. Also mit solchen Aggregaten kommen wir gut klar.

*Ok. Dann habe ich jetzt noch hauptsächlich Fragen zum Planungsprozess, was dann mit dem Programm später geklärt werden kann. Benutzen Sie außer dem Programm noch einen anderen Zugang zu den Daten?*

Nein. Also unsere technischen Zählraten landen in einer Datenbank und die ist mit einer Eigenentwicklung verknüpft, auf deren Basis wir unserer Angebotsplanungsabteilung Daten in einer sehr schönen Form zur Verfügung stellen können. Das sind Excel-Dateien, ich kanns Ihnen gerne kurz mal zeigen, wie sowas aussieht. [öffnet Datei] Das ist ein sogenannter Belastungsteppich. Man sieht, in dem Fall ist es die Linie 55, Putzbrunn nach Ostbahnhof. Hier ist der komplette Linienweg dargestellt, nach unten, und nach rechts alle Abfahrtszeiten, das geht hier nach rechts noch ein ganzes Stück weiter. Und ich sehe für jede Abfahrt an jeder Haltestelle, wie viele Fahrgäste sitzen im Fahrzeug. Also hier haben wir jetzt zum Beispiel die Fahrt um acht Uhr vierunddreißig und dreißig Sekunden, am Waldheimplatz Richtung Eulenspiegelstraße, sitzen an dem Wochentagstyp zehn Fahrgäste im Durchschnitt im Bus.

*Ok und das sind die Daten aus den Zählgeräten*

Genau. Das ist eine Darstellungsform, mit der unsere Angebotsplanung sehr gut arbeiten kann. Farblich hervorgehoben dann eben noch die Auslastung und rot umrandet die Spitzenstunde.

Temporal resolution is required to be at least within a 30 minute interval as special events such as just football games generate peaks of load that can otherwise be flattened too much

The interviewee is aware of privacy guidelines limiting the grade of detail that can be made available from mobile network data

The data gathered from counting devices is stored in a database which is connected to an in-house development based on a spreadsheet application. These data are visualized in two ways; a color-coded "load-carpet" showing typical numbers of passengers on a vehicle throughout the day along each segment of a given route and a line- and bar-chart showing the daily cycle of passenger loads between two consecutive stops along a given route. The planning department is said to be working very well with these visualizations. Hours with maximum load are marked red.



*..die aus den Zahlen kommt?*

Ja.

*Sind dann in der Stunde auch mehr Fahrzeuge unterwegs?*

Nein, es sind einfach nur mehr Fahrgäste. Also ich hab jetzt hier 179 Fahrgäste am Karl-Preis-Platz. Wenn ich die Fahrt davor nehme mit allen Fahrgästen in der Stunde danach, sinds auch noch 179, da ist es also noch gleich. Und hier sind es hundert.. Moment da ist irgendein Fehler drin.. Ah nee, haut schon hin, weil die Abfahrtszeiten ja, also es sind nicht immer sechs Fahrten je Stunde, sondern es sind auch mal nur fünf Fahrten, das heißt hier hab ich irgendwann mal einen Sprung, genau, hier hab ich mal einen zwanzig-Minuten-Takt drin, deswegen sind es dann bloß fünf Fahrten in der Stunde.

Temporal resolution is dependent on the specific interval between two trips.

*Wäre es möglich, davon einen Screenshot zu bekommen?*

Ja, kann ich machen.

*Wunderbar, das wäre super.*

Also das ist so eine Darstellungsform. Dann haben wir noch verschiedene andere Berichtsformen, die man ganz gut nutzen kann. Und für Sonderfragenstellungen, die sich mit diesen Standardberichten nicht darstellen lassen, gehen wir auch direkt in die Datenbank und können dort intensiv mit den Zahlen arbeiten.

For special enquiries, the database is directly consulted via SQL queries.

*Also einfach per SQL-Abfrage*

Genau. Also das ist, dadurch dass das eben eine Eigenentwicklung ist, ist das System sehr offen und sehr flexibel für uns.

*Und die anderen Darstellungen, stellen die auch Zahlen dar, oder inwiefern gibt es da grafische Elemente?*

Wir haben auch mal Grafiken, wir haben hier zum Beispiel eine sogenannte Tagesganglinie [öffnet Dokument], wo man an einem Querschnitt, das ist der Querschnitt Siglstraße - Lautensackstraße, oben in diese Richtung

*Querschnitt heißt eine Verbindung zwischen zwei Stationen?*

Zwischen zwei aufeinander folgenden Haltestellen, genau. Und da sieht man halt die Fahrgastverteilung, mit Darstellung, in rot, der Spitzenstunden wieder.

*Das heißt, man hat einen Tag?*

Man hat einen Tag, und zwar ist das in dem Fall Montag-Donnerstag-Schule,

und da sieht man halt dass die Trambahn hier dargestellt im zehn-Minuten-Raster, in diesem Zeitraum, das ist von April bis Ende Juli, das ist ein Zeitraum in dem nicht so viel los ist normalerweise, unsere Hauptzeiten sind natürlich im Winter, also November bis Februar ungefähr. Da haben wir jetzt im zehn-Minuten-Raster von sieben Uhr dreißig bis sieben Uhr vierzig vierzig Fahrgäste an diesem Querschnitt.

Data is aggregated over several weeks or months with similar demand, e.g. April to July

*Ok. Das heißt jede Fahrt hat eine Säule?*

Jedes zehn-Minuten-Intervall. Das heißt also, wenn wir einen fünf-Minuten-Takt haben, ist eine Säule zwei Fahrten, wenn wir einen zehn-Minuten-Takt haben, ist eine Säule eine Fahrt. Und wenn wir einen zwanzig-Minuten-Takt hätten, das haben wir hier, dann fehlt eine Säule.

In the bar-chart, one bar represents a ten-minute interval. This means each bar can represent a different number of trips, depending on the current interval and some bars disappear when the interval exceeds ten minutes.

*Ok also für alle zehn Minuten gibt es eine Säule, und darüber ist eine Kurve, die das mittelt?*

Nein, die Kurve ist der gleitende Stundenwert. Das heißt also ich hab hier die Säule für sieben Uhr dreißig und das hier, die rote Linie, ist dann die Stunde von sieben Uhr dreißig bis acht Uhr neunundzwanzig und neunundfünfzig Sekunden. Also das ist dann die rechte Skala, da haben wir dann, Moment, zweihundertzwanzig Fahrgäste in der Stunde.

*Ok. Also, wir haben ein Säulendiagramm, und dann den überlagerten gleitenden Stundenwert. Und die grauen Punkte, was sind das?*

Die grauen Punkte sind die Summen der Maxima der einzelnen Zählungen. Das heißt, wir haben hier, hab ich ja gesagt, aus den technischen Zählgeräten haben wir eine Vielzahl an Daten. Wir bilden dann so ungefähr drei-Monats-Zeiträume und in einem solchen drei-Monats-Zeitraum bilden die Säulen dann die Mittelwerte. Um aber auch zu sehen, ob es viele Ausreißer gibt, oder häufig vollere Fahrten, deswegen bilden wir auch die Summe der Maxima und das sind dann die grauen Punkte. Wir haben da zwanzig Zählungen, möglicherweise in diesen drei Monaten, und haben zwei Fahrten im zehn-Minuten-Raster und dann haben wir vielleicht, jetzt muss ich wieder auf diese Säule gehen, hier bei der Summer der Maxima einen Wert von 50. Dann hab ich vielleicht einen Maximalwert bei einer Fahrt von 30 und bei der darauffolgenden Fahrt von 20 Fahrgästen. Das gibt also eine Summe der Maxima von 50, der Mittelwert der beiden Fahrten liegt aber bei 40.

*Ja. Also die Summe bezieht sich auf die verschiedenen Fahrten innerhalb der..*

Der Mittelwert auch. Ich bilde einmal die Summe der Mittelwerte und einmal die Summe der Maxima.

*Und die blauen Linien sind wahrscheinlich der Takt..?*

Nein, das blaue ist die empfohlene Auslastung nach dem Nahverkehrsplan. Also in der Hauptverkehrszeit schreibt der Nahverkehrsplan vor, dass die

The data is put into relation with the recommended load.

Spitzenstundenstundenauslastung nicht über 65% liegen soll, und in den Nebenverkehrszeiten nicht über 50%. Das heißt also das ist ein Qualitätsmerkmal, ein Komfortmerkmal. 65% klingt halb leer, ist aber tatsächlich so.. 100% bedeutet jeder Sitzplatz, plus, in den freien Flächen pro Quadratmeter vier Personen. Das wären 100%.

*Ok, also da passt dann wirklich nicht mehr rein.*

Da passt noch mehr rein. Die 100%-Auslastung ist auch eine Rechengröße für Komfortmerkmale. Das ist noch immer weit unter der technischen Zulassungsgrenze, die die Busse haben, oder die Fahrzeuge. Das heißt, da dürften nach TÜV oder nach Zulassungsbehörde noch deutlich mehr Fahrgäste rein, man kann das noch wesentlich enger stapeln. Aber auch die 65% is ein Wert, wenn ich da vorm Bus stehe, .. das fühlt sich voll an. Da ist dann nicht mehr viel Platz, so gefühlt. Aber es gibt zum Beispiel im Fußballzubringerverkehr, in der U6 haben wir auch ab und zu mal 120%. Also da sind es dann auch mal fünf bis sechs Personen pro Quadratmeter, die da stehen.

*..und wenn man Fan genug ist, dann ist einem das egal.*

Genau. Zum Oktoberfest genauso. Das ist klar, da quetschen sich alle rein, die reinpassen. Und zum Teil auf einzelnen Abschnitten der U6 auch die die zur Uni fahren, zur LMU, beziehungsweise, auf der U1 zur TU natürlich das-selbe, da ist es dann schon mal richtig voll.

*Das heißt, die Kapazität von den einzelnen Fahrzeugen ist auch so eine Einheit, mit der sie dann arbeiten.*

Die ist da hinterlegt, genau.

*Ok. Aber zugrunde liegen natürlich die Zahlen der Personen die gezählt werden. Das heißt das bringt Ihnen was, wenn ich das in meinen Visualisierungen dann auch als Einheit nehme.*

Genau. Die Anzahl der Fahrgäste, oder der Personen, die sich von A nach B bewegen ist genau das was wir brauchen. Die Schwierigkeit bei den Mobilfunkdaten ist natürlich, festzustellen, mit welchem Verkehrsmittel die Leute unterwegs sind.

*Genau, darum geht es ja in dem Projekt auch.*

Das ist natürlich extrem schwierig. Also gerade einen nicht wahnsinnig vollen Bus von nem Auto zu unterscheiden, die gleich hintereinander fahren, ist natürlich kaum möglich. Tram geht vielleicht noch ein bisschen eher, weil die ein anderes Geschwindigkeitsprofil haben, aber auch nur da, wo sie eigene Gleiskörper haben. Genau also Fahrrad, das sind zum Teil die selben Strecken und Geschwindigkeiten. Also das wird nicht ganz einfach. Aber wenn wir insgesamt die Personen haben, die von A nach B fahren und auch von unseren Zählgeräten wissen, wie viele bei uns in unseren Fahrzeugen

The main interest lies in the absolute number of people travelling between different points. The interviewee mentions the difficulty of determining the means of transport from mobile network data and explains that the total number is already helpful as a comparison with the internal data can give the amount of people travelling outside public transportation

drin sind, dann können wir ja rückrechnen auf den Rest und können dann auch [...] ungefähr, wie viele mit dem Fahrrad und wie viele mit dem Auto unterwegs sind. Also die gesamte Zahl ist auch schon mal super.

*Das ist ja dann auch für die Zukunftsplanung wahrscheinlich das Relevantere.*

Genau

*Wenn ich davon auch noch einen Screenshot bekommen könnte, das wäre super. Gibt es noch andere Darstellungen, die Sie verwenden?*

Das sind unsere Standardberichte, die zwei, tatsächlich. Alles andere läuft dann über Tabellen, letztlich.

*Ok. Und das hier ist auch statisch, also man kann..*

Das ist statisch, ja. Also man kann, es gibt so ein Mouse-Over aber das ist ja..

The visualizations are static. The only interaction is a tooltip showing details for different elements

*Ok, aber sonst ist es an Interaktivität das..*

Ja also da gibt es noch ein paar Tabellen, die dahinter liegen, die aber auch keine andere Aussage haben als die Grafik selbst.

*Das heißt, wenn man was in den Tabellen ändert, hat das gleich eine Auswirkung auf die Grafik.*

Ja. Also wir haben lange überlegt, ob wir die Ausgabe browserbasiert machen oder irgendwie anders. Aber es hat sich rausgestellt, die Angebotsplanung arbeitet dann auch damit und die wollen auch gerne sehen was passiert, wenn ich die Kapazität ändere, wie wirkt sich das aus. Und da können sie jetzt in Excel einfach mit den Daten spielen..

Changing entries within the tables that feed the graphic changes the graphic. For that, the user has to switch between different sheets. A browser-based solution has been considered in the past but was deemed too complicated and not as flexible and interactive as spreadsheets which are said to have a high acceptance amongst planners as everybody knows how to use them.

*..und da immer hin- und herswitchen.*

Genau, und das ginge natürlich mit was browserbasiertem nicht.

*Das würde ich jetzt nicht ausschließen..*

Nicht so komfortabel. Also das wäre halt eine Krücke. Und das ist einfach ein vorhandenes System, das auch den Vorteil hat, dass jeder Excel bedienen kann.

*Ok.*

For special thematic questions, GIS is used to generate further visualizations but these are not part of the daily planning work.

Achso. GIS-Auswertungen machen wir natürlich noch. Also Darstellungen in GIS. Allerdings sind das keine Standardauswertungen, sondern da sitzen bei uns die Leute die halt mit GIS arbeiten dran und machen themenbasiert

Auswertungen und stellen es dann in GIS dar.

*Wissen Sie, welches Programm das ist?*

Wir arbeiten hauptsächlich mit QGIS, haben auch inzwischen einen QGIS-Server und können dann mit Web-GIS arbeiten. Im Haus haben wir auch ArcGIS, haben das früher hauptsächlich genutzt und gehen aber immer mehr auf QGIS, weil es open source ist, weil es die selben Möglichkeiten bietet und wir noch keine Nachteile gefunden haben. Also keine ernsthaften.

*Ok. Aber das sind dann nicht die Planer an sich, die damit arbeiten.*

Nein. Die arbeiten zum Teil ein bisschen mit GIS aber nicht so richtig. Wir haben einfach bei uns im Haus einen Kartographen und ein paar Geographen, die mit GIS gut arbeiten können und die machen dann diese Darstellungen.

*Und das ist mehr für die Berichterstattung.*

Das ist für die Berichterstattung aber auch zum Teil sind es inhaltliche Sachen, dass man sich anschaut, Einzugsgebiete zum Beispiel für Flyeraktionen aus dem Marketing oder für Informationen, wenn zum Beispiel eine Baustelle ist. Dass man sagt, ok, alle die vom Einzugsgebiet dieser Baustelle betroffen sind, die sollen einen Flyer im Briefkasten finden. Das macht man über GIS, um festzustellen, wie viele Flyer man drucken muss. Zum Beispiel. Oder auch, das machen wir zur Zeit nicht mehr, wird aber wieder kommen wahrscheinlich, die Umgebungspläne, die an den Bus- und Trambahn-Haltestellen aushängen. Die hat man mit einem GIS-Serienausdruck früher gemacht. Derzeit werden die extern vergeben, an eine externe Firma. Wahrscheinlich werden wir es aber wieder nach intern holen, weil die eigentlich besser waren.

*Ok, alles klar. Wie oft wird geplant, oder in welcher Frequenz werden Dinge verändert?*

Also es gibt den Jahresfahrplan, der wird immer im Herbst für das übernächste Jahr gemacht. Also mit einem guten Jahr Vorlauf. Das liegt nicht daran, dass wir so unglaublich langsam wären, sondern das liegt tatsächlich daran, dass der Fahrplan auch von der Stadt genehmigt werden muss, von der Genehmigungsbehörde, und die wollen einen sehr sehr großen Vorlauf, um dann auch Rückfragen stellen zu können, Widersprüche bearbeiten zu können und das alleine braucht schon ein Dreivierteljahr und vor diesem Dreivierteljahr haben wir noch mal ein halbes Jahr vorgeschaltet, also das sind Fristen, die sich aus bürokratischen Gründen kaum ändern lassen. Ansonsten gibt es Planungen für Baustellen, die gehen zum Teil sehr schnell. Also in der Regel wissen wir um Baustellen ja auch ein paar Monate im Voraus aber manchmal geht halt auch was kaputt und dann muss man sehr kurzfristig arbeiten. Und dann wird auch kurzfristig geplant, gerade in dem Bereich, indem es dann einen Schienenersatzverkehr geben muss.

The general planning process including administrative approval takes about one and a half years.

*Ok. Und das heißt, für den Jahresfahrplan, der beruht auf Zählungen, die ein bis zwei*

*Jahr zurückliegen im Endeffekt.*

In dem Moment wo er geplant wird sind sie aktuell, aber er wirkt sich halt mit einem Verzug von eineinhalb Jahren aus.

*Passiert es, dass sich Dinge verändern, unabhängig von Baustellen, die dann zu Planänderungen führen?*

Also so schnell ändern sich ja Dinge nicht in der Stadt. Wenn irgendwo ein neues Gewerbegebiet entsteht, kriegen wir das rechtzeitig mit, bei Wohngebieten genauso, die werden ja nicht innerhalb von ein paar Wochen gebaut sondern das dauert dann schon ein paar Jahre, das heißt, da haben wir eigentlich gar kein Problem. Was uns immer mal wieder kalt erwischt, ist wenn zum Beispiel die S-Bahn kurzfristig mitteilt, dass in drei Monaten die Stammstrecke gesperrt wird, da müssen wir dann überlegen, wie wir damit umgehen. Und zum Teil sind es große Veranstaltungen, aber die sind in der Regel auch ein paar Monate im Voraus bekannt, und da werden wir auch informiert. Also große Messen, die BAUMA ist eine der größten Messen weltweit. Das ist immer ein Riesenaufwand, die ist alle zwei Jahre, das ist ein Aufwand, der ist schon fast mit dem Oktoberfest vergleichbar.

*Ok. Dann fällt mir jetzt erst mal nichts mehr ein. Haben Sie noch was, was ich wissen sollte?*

Ja, vielleicht. Also wir haben ja eben schon mit den Telekomdaten, aber auch mit O2-Daten, also Telefonica-Daten, schon mal versucht zu arbeiten, haben zum Beispiel auch mal gesagt, wir würden gerne wissen, wo die Menschen herkommen, die zu einem Fußballspiel vom FC-Bayern gehen. Das hat für uns den großen Vorteil, wir wissen, wie viele Leute im Stadion sind und können dann schauen, wo sind die hergekommen. Und grad bei einem Bundesligaspiel, wo die Leute dann nur aus Deutschland kommen, als auch bei einem Championsleague-Spiel, wo dann vermutlich ganz viele Leute am Flughafen aufploppen und dann zum Stadion rausfahren. Da wollten wir dann die Herkunft, also den Startort der Fahrt wissen. Wir haben eigentlich gedacht, dass das eine sehr schöne, begrenzte Aufgabenstellung ist, die gut zu beantworten ist, und das hat nicht richtig toll geklappt. Also es wurden nur ungefähr 80% der Menschen gefunden, die im Stadion waren, auch nach Hochrechnung, was uns sehr gewundert hat, und die Startorte waren geographisch schlecht zugeordnet. Also man gesehen, es gab zum Beispiel in der Nähe von Rosenheim ganz ganz viele Startorte, die aber wahrscheinlich in Rosenheim gestartet sind, die aber irgendwie 20km verschoben waren.

*Ok, das heißt, die räumliche Auflösung war nicht hinreichend*

Ich glaube, da war einfach ein Datenfehler drin, ich weiß es nicht. Also, gut, was dahinter lag konnten wir uns natürlich nicht erklären, aber sowas stärkt nicht das Vertrauen in die Daten, das ist ganz klar, und mit 80% Genauigkeit können wir auch nicht arbeiten. Also wir arbeiten mit den technischen Zählraten mit 95% Genauigkeit, eigentlich sogar mehr, aber mit einer garantierten 95% und mit weniger gibt sich unsere Planungsabteilung auch nicht

A past experience with working with mobile network data was underwhelming as only about 80% of people visiting a football game could be modeled using the extrapolated data and the spatial distribution seemed unrealistic. An accuracy of 95% was mentioned to be the minimum.

zufrieden, ehrlicherweise.

*Ok, da muss ich mal gucken, ob da in den Daten die ich habe solche Veranstaltungen mit drin sind.*

Irgendsowas ist bestimmt mit drin. Sowas ist dann immer sehr spannend. Vor allen Dingen ist es ja dann zur eigenen Kontrolle ganz gut, wenn ich weiß, wie viele Leute müssen da sein und wie viele finde ich in meinen Daten. Und da ist so ein Fußballspiel toll, weil da krieg ich hinterher die sehr genaue Anzahl immer mitgeteilt vom Verein, die veröffentlichen, wie viele Zuschauer da waren. Eigentlich muss ich die dann fast alle finden.

*Ok. Wunderbar. Dann vielen Dank erstmal. Dann hören wir mal mit der Aufnahme auf.*

## Interview II

*Ich habe zwei große Fragenbereiche, wir können ziemlich frei reden und Du kannst mir alles erzählen, wovon du denkst, dass es relevant ist. Und zwar geht es mir einmal darum, wie der Planungsprozess aussieht und dann darum, was der Inhalt ist. Wir können gerne mit dem "Was" anfangen. Und zwar: Was ist das Ziel der Planung, wie würdest Du das beschreiben?*

Also das Ziel was man am besten fassen kann und was auch die stärkste Grundlage hat, nach der man sich richten kann, ist die Erfüllung des Nahverkehrsplans. Wir sind da auch bei der Erstellung mit beteiligt und das Dokument, kennst du den Nahverkehrsplan? In diesem Dokument werden eben für München, das gibt es auch für jeden Landkreis außerhalb von München, werden einfach gewisse Standards an die Erschließung mit dem Nahverkehr festgehalten. Das sind dann so Sachen wie, also so Qualitätsstandards, die sagen, innerhalb von einem Gebiet, das ist dann eingeteilt, zum Beispiel ist das jetzt ein städtisches Gebiet, das ist sehr dicht, und das andere ist dann eher ein weniger dichtes Gebiet oder ein Randgebiet, das ist dann eher ein bisschen zersiedelt und je nach dem, in welchem Gebiet ich mich befinde, gibt es Standards, innerhalb von welchem Radius ich eine Haltestelle zum Bus oder zur U-Bahn oder zur Tram oder zur S-Bahn erreichen muss und das variiert dann zum Beispiel auch - U-Bahn und S-Bahn werden als ein bisschen höherwertiges Verkehrsmittel gesehen, weil sie ja auch schneller sind, und man deswegen größere Distanzen zurücklegen kann. Da ist dann zum Beispiel der Radius, gerade im äußeren Bereich schon mal ein bisschen höher, mit 600m Radius oder sowas und dann heißt es aber auf der anderen Seite wieder, in der Innenstadt soll ich innerhalb von ner Luftlinie von 300m zu einer Bushaltestelle kommen oder so. Also das ist da festgelegt. Und das ist halt so dieses schwarz auf weiß, wo jemand sagen kann, hier ihr als Nahverkehrsunternehmen erfüllt nicht den Nahverkehrsplan. Und das ist unsere Aufgabe, das quasi zu erfüllen und das der Bevölkerung anzubieten. Und das ist so dieses Ziel, dass uns auch von städtischer Seite im Grunde schon bestätigt ist, auch wenn wir dann quasi mit einer Planung kommen, sagen, wir müssen das hier machen, um den Nahverkehrsplan zu erfüllen. Aber grundsätzlich haben wir schon auch so ein bisschen ein ideelles Ziel, mehr Leute in den ÖV zu bekommen und das kann man nur machen, indem man ein möglich attraktives Angebot hat und da geht es quasi zum einen drum, Sachen

Aim of planning: ensuring a certain quality of life for all people living in the region.

different means are graded differently according to speed and possible distances

legally binding document states quality standards

zu verdichten, mehr umsteigefreie Verbindungen anbieten zu können, mehr Zuverlässigkeit, mehr Pünktlichkeit, eine höhere Erreichbarkeit. Und gerade bei so Neubaugebieten, dass man von Anfang an eine sehr hohe Qualität an ÖPNV hat, dass man von Anfang an da ist und denen den Service bietet und dann auch so Sachen wie Nachtverkehr oder so was. Dass man nicht sagt, ah aber ihr wohnt jetzt da irgendwo am Arsch der Welt aber um acht Uhr kommt ihr da nicht mehr weg, sondern dass man da eben von Anfang an eine hohe Qualität hat um das attraktiv zu machen, um möglichst viele Leute in den ÖV zu bringen.

*Ja, ok. Welche konkreteren Fragestellungen ergeben sich dann daraus?*

Also bei den ganzen Bestandslinien.. Also ich bin ja jetzt von der Angebotsplanung Bus, bei der Tram und U-Bahn gibt es halt dann immer noch so Spezifika und da haben wir halt beim Bus die größte Freiheit, das muss man schon sagen, weil wir eben nicht auf einer Trasse laufen. Bei den Bestandslinien geht es in der Regel immer darum, haben wir die richtige Kapazität, also haben wir die richtige Busgröße auf den Gefäßen, haben wir die richtigen Takte. Das sind dann auch wirklich so Sachen, das passen wir dann manchmal um eine Stunde auch an, dass wir sagen, oh wir haben jetzt gesehen, vielleicht ist da eine Schule hingekommen oder was weiß ich, da ist jetzt schon immer ein großer Peak an Leuten um 13 Uhr statt wie sonst immer ab 16 Uhr und dann sagen wir, ok, da müssen wir den Takt Zehn ein paar Stunden früher anfangen lassen. Wie gesagt, in den letzten Jahren haben wir uns eigentlich immer nur mit Ausweitungen beschäftigt, nicht mit irgendwelchen Einsparungen, mal schauen, was da noch auf uns zu kommt. Genau das sind so die Bestandslinien. Bestandslinien ist auch immer, wie kann man manchmal auch neu verknüpfen um dann zum Beispiel so eine Verteilung an den Busbahnhöfen zu optimieren, weil wir da tatsächlich auch an unsere Kapazitätsgrenzen immer kommen mittlerweile, dass es halt alles so voll mit Bussen gestellt ist. Wenn man die aber gut verknüpft, dass die Wartezeiten irgendwie gut zusammenpassen, dann kann man halt aus so einer kleinen Fläche viel mehr an Linien rausholen, als wenn man jetzt sagt, jede Linie endet da und hat dann irgendwie zehn Minuten Wartezeit oder sowas. Das sind so die Optimierungen. Und dann die andere Aufgabe ist dann eben neue Linien. Also, da planen wir von Grund auf, sagen, ok, da fehlt irgendwie eine Verbindung, und da haben wir dann beim Bus eben die Frage, fahren wir rechts rum, fahren wir linksrum, zu welchem übergeordneten Knoten wollen die Menschen, in welchem Bereich arbeiten vielleicht da viele Menschen und wo haben sie die größtmöglichen Umsteigemöglichkeiten und da geht es dann tatsächlich wirklich mit, wo kann man gut mit dem Bus fahren, wo kann man vielleicht einen anderen Bus entlasten, indem man das dann noch parallel setzt oder schafft man eine ganz neue Verbindung und also das sind dann eben so Ansätze wo wir dann auch das Verkehrsmodell zum Beispiel befragen, also die neue Linie da eingeben, Varianten bilden und schauen, wo können wir die meisten Fahrgäste gewinnen, wo haben wir vielleicht auch den größten Gewinn in der Fahrzeit, dass jetzt jemand vielleicht nicht mehr durch die Innenstadt fahren muss und genau, das ist so, was auf Planungsebene eigentlich so untersucht werden muss.

*Ja. Wenn du sagst, dass man sieht, oh da brauchen wir eine neue Bushaltestelle, wie*

idealistic aim: get as many people as possible using public transportation by offering an attractive a product as possible by densifying the schedule and the network, offering routes that match demands, increase reliability, punctuality and accessibility.

Concrete aims within bus planning: offer adequate capacity, both spatially and temporally

distribution of buses between trips has to be optimized for efficient use of available parking space

Finding demand and optimal routes for new lines and connections



*kommt man zu dem .. woher kommt die Information, bzw. was ist die Grundlage für die Entscheidung oder die Idee, in dem Raum eine neue Station zu bauen?*

Grunderfahrung hat man glaube ich schon so das Gefühl, dass man, wenn man so auf eine Karte schaut, also die Haltestellen sind eigentlich schon immer relativ gleichmäßig verteilt. Also es gibt schon immer so Abstände, die variieren vielleicht von 300 bis 450 Metern oder sowas. Und die Abstände machen natürlich immer nur Sinn, wenn das durchgehend besiedelt ist. Wenn man jetzt da irgendwie durchs Nirgendwo fährt, dann halten wir da natürlich nicht und das ist dann eher so ein Erfahrungswert, wo man dann sagt, ok, spätestens nach so und soviel Metern bräuchten wir eigentlich mal wieder eine Bushaltestelle und dann orientieren wir uns daran, ob da zum Beispiel, dass wir uns an Zuwegungen aus irgendwelchen Siedlungsgebieten irgendwie positionieren, irgendwie, dass wir uns an Kreuzungsbereiche positionieren, wo zum Beispiel auch Fußgängerüberwege sind, weil das ja immer wichtig ist auch, wie man dann von der Bushaltestelle ins Gebiet kommt, oder ist da irgendwie so ein Point of Interest, irgendwie ein großer Einkaufsmarkt oder eine große Arbeitsstätte oder eine Schule oder so Sachen. Also da orientieren wir uns dann eigentlich direkt vor Ort immer.

spatial distribution of stops is determined by geographical distribution of residential areas, workplaces, points of interests and other infrastructure

*Ok, das heißt, die Info da kommt aus Karten - also man schaut auf eine Karte und guckt sich das an?*

Ja, die Info kommt tatsächlich vor allem auch aus Ortskenntnis. Also jeder hier, das wird immer intensiver, je länger man daran arbeitet, aber die, gerade in Gebieten in denen man plant, muss man eigentlich auch vor Ort ein bisschen hingehen und..

information fed mostly by local knowledge and individual local investigations

*..und dann fällt einem gleich auf..*

ja, beziehungsweise, also hier im Team ist einfach auch schon riesiges Grundwissen da, so ein Wissensschatz an diesen ganzen, wo ist eine Schule, wo wird was gebaut und so..

*Ok. Dann können wir vielleicht zu dem "Wie" übergehen. Also Du hast ja schon das Modell angesprochen, das ist in VISUM auch?*

Genau, ja.

*Magst Du mir das kurz zeigen?*

Ich hab jetzt tatsächlich die [?] nicht hier, weil ich das natürlich im Homeoffice.. Also ist das halt auch das pvt.. [I] meinte, dass du dir das auch oben schon mal angeschaut hast vorher bei der Verkehrsforschung?

*Ja, kurz.*

Ja also ich kann jetzt da gar nicht rein leider. Mit dem Modell arbeiten wir auch. Oft ist es so, dass wir wirklich nur denen eine Karte schicken und sa-

gen, das wäre die eine Variante, das wäre die andere Variante, das ist ein Entwurf für einen Fahrplan, den müssen sie auch ins Modell einpflegen, wo man dann schon Umstiege und sowas beachten muss, idealerweise, dass man sagt, ok, der Umstieg zur U-Bahn ist in dem Fall jetzt auch ganz gut, dass das natürlich auch attraktiv ist. Ich kann dir mal zeigen, was dabei rauskommt. Und das hat bis jetzt viel die Verkehrsforschung für uns gemacht, dass sie das dann ins Modell eingepflegt haben und berechnet haben, und genau, ich fang jetzt seit nicht so langer Zeit an, dass wir es auch ein bisschen selber anwenden, dass wir es selber einpflegen und dann die Verkehrsforschung das nur noch auswerten muss und kontrolliert, ob quasi alles richtig kalibriert ist und so Sachen. Genau was dann rauskommt, da schauen wir uns dann eben an, was wir für Mehrverkehr haben, was wir für Verlagerungen haben auch vom MIV, was wir vielleicht auch an Zugewinn an komplett neuen Fahrgästen haben, die vielleicht vorher irgendwie anders gefahren sind..

*..was meinst du mit MIV?*

Motorisierter Individualverkehr

*ah ok*

Genau also die, die Auto fahren. Entschuldigung, da ist man dann aber auch ein bisschen betriebsblind dann mit diesen ganzen Abkürzungen. Und bei uns werden sehr viele Abkürzungen verwendet. Jetzt muss ich es noch mal kurz suchen. [öffnet Dokument] Hier ist zum Beispiel, da haben wir eine Linienverlängerung von hier, also bis hier gibt es die Linie schon und wir haben jetzt geprüft, was das für einen Effekt hat, wenn wir sie bis zum Königsplatz verlängern. Genau, da haben wir quasi den Linienverlauf, den schicken wir dann an VM, an die Verkehrsforschung, mit einem Fahrplan. Das wird dann in das Modell eingepflegt und dann wird das berechnet. Und das ist jetzt das Ergebnis und das zeigt halt, wie viele Leute tatsächlich, in dem Modell, theoretisch mit dieser Linie mitfahren würden. Das ist einmal die komplette Last und dann gibt es einmal noch den Differenzplot, das ist quasi dann immer, das ohne der Verlängerung ist der Normalfall und das mit der Verlängerung ist der Planfall sozusagen. Das vergleicht man und setzt in Differenz und dann sieht man hier, wie viel man auf der Linie gewinnt, aber gleichzeitig auch, wo entlastet man zum Beispiel Linien. Und wenn wir jetzt sagen, gerade in der Innenstadt ist es im Moment unser Ziel, ein bisschen die U-Bahnen zu entlasten, weil die halt einfach mit dem Ausbau nicht hinterherkommen, und dann sehen wir hier ganz schön, ok, der Effekt ist sogar ziemlich gut, weil die, die hier normalerweise in die U5 einsteigen und sich dann am Hauptbahnhof in die U-Bahnen verteilen, wenn sie hier in die Außenäste müssen, und die machen das jetzt alle so, dass sie sagen, ok, ich fahr einfach mit dem Bus durch, steig dann hier in die schon viel leerere U-Bahn am Königsplatz oder am Stiglmaierplatz ein. Oder das gleiche natürlich umgekehrt, hier sind auch viele Arbeitsplätze, die von außerhalb kommen, die dann sagen, ok, ich muss mich dann nicht am Hauptbahnhof durchkämpfen, sondern ich steig da ganz entspannt um, wo noch nicht so viel los ist. Und hier haben wir jetzt zum Beispiel gar nicht so eine große Wahl, wie wir jetzt über die S-Bahngleise kommen, aber wenn wir uns jetzt nicht sicher wären könnte man sagen, ok, wir machen eine Linienführung so und eine so, und

Direct interaction with traffic model is limited, usually requested from other department. Results are displayed on map within model, showing additional traffic and transferred traffic usage, e.g. from individual traffic towards public transportation

Differences between current situation and situation after network changes are being visualized

different routing options are tested

dann schauen wir mal, welche im Modell theoretisch besser ankommt sozusagen, und das hilft schon auch tatsächlich stark für die Entscheidung, wie wir die Linienführung gestalten.

*Ok. Das heißt, ihr schlagt was vor und gebt das an die Modellierung weiter und kriegt dann das als Ergebnis. Und das ist auch alles an visuellem Output, also..*

Genau, die Karten sind das Visuelle und dann gibt es eben noch so Zahlen, wie Personenbeförderungsfälle und eben das was ich vorher gesagt hab, wie viele Personen von welchem Verkehrsmittel kommen. Also ist es jetzt nur eine Verlagerung innerhalb vom ÖV-System, also das heißt, alle die U-Bahn fahren, fahren jetzt Bus, oder wir hier zum Beispiel auch welche von den Autofahrern begeistern, den Bus zu nehmen, weil es billiger ist, weil es attraktiver ist, weil es genau die Verbindung ist, die sie vorher gebraucht hätten so ungefähr. Das kriegen wir auch noch aus dem Modell.

maps are accompanied by numbers

*Ok. Und auf der Grundlage entscheidet ihr dann.*

Genau. Unter Anderem. Also, die Ausführung ist tatsächlich dann noch mal komplizierter. Also das ist jetzt eine Planung aber dann gehen wir mit dieser Planung in die Bereisung, das ist ein sehr schönes altertümliches Wort, es gibt die Bereisungskommission, die besteht eben aus einem Mitarbeiter von uns, auch welche vom Betrieb, vom Busbetrieb, die dann entscheiden, ob das auch für den Bus fahrbar ist und auch für den Fahrer keine zu großen Risiken mit irgendwelchen, keine Ahnung, Zebrastreifen oder Einfahrten die man nicht einsehen kann, oder ist es zu schmal oder irgendwie solche Sachen, die das beurteilen. Die Polizei ist dabei, das Baureferat ist dabei, das KVR ist dabei, das Kreisverwaltungsreferat, die die ganzen verkehrlichen Anordnungen machen, dann sagen, ok, wenn da Halteverbote hinmüssen, dann sind die diejenigen, die das anordnen und die fahren dann tatsächlich da draußen rum und schauen sich das an und treffen dann die Entscheidung, eine Haltestelle - wir sagen, wir möchten gerne eine Haltestelle ungefähr in diesem Bereich und die fahren dann raus und sagen, ok, die Haltestelle ist hier geeignet. Die setzen das dann fest und so wird das dann festgesetzt. Also das Letzte ist quasi die Infrastrukturprüfung, die entscheidet, ob wir so fahren können, wie wir uns das überlegt haben oder nicht. Das ist so die letzte Instanz tatsächlich. So Infrastrukturprüfung und Verkehrssicherheit, so könnte man das nennen. Und bauliche Machbarkeit.

Execution of a new route within the network only after elaborate survey process possible

*Ok. Also das betrifft jetzt alles neue Linien. Wenn wir zurück gehen zu der Bestandsplanung, da ist die Auslastungsplanung auch ein Teil davon?*

Genau, für die Bestandslinien ist das Ausschlaggebende unsere Fahrgastzahlen, also jetzt für uns für die Planung, die Fahrgastzahlen und sowas wie die Pünktlichkeitswerte. Also wenn wir jetzt sehen, eine Linie hat extrem, also verliert irgendwie immer in einem gewissen Bereich extrem an Fahrzeit, dann gibt es dann noch mal die Abteilung, die sich um Infrastruktur kümmert, die dann da optimieren, ob man die Ampelschaltung noch optimieren kann, oder im Zweifel muss man einfach tatsächlich im Fahrplan die Fahrzeit anpassen, das ist dann auch sowas was dann noch kontrolliert wird,

planning within consisting routes based mostly on passenger counts and information on punctuality

wichtiger für uns sind die...

*..und die Pünktlichkeitszahlen kommen von den Fahrzeugen?*

Die kommen von den Fahrzeugen, genau, das ist alles digital. Und was für uns tatsächlich noch entscheidend ist, sind die Fahrgastzahlen, das sind für die ganzen Bestandslinien.. [öffnet Dokument] Das sind jetzt zum Beispiel Fahrgastzahlen von Montag bis Donnerstag, also quasi ein Werktag, ein normaler, der Freitag ist ja immer schon wieder ein bisschen anders, in einem Zeitraum vom 7.1. bis zum 20.3., das ist wahrscheinlich irgendwie sowas wie nach den Weihnachtsferien bis zu den Osterferien oder sowas müsste der Zeitraum ungefähr sein. Und das ist quasi die Auslastung in den Bussen. Hier sieht man immer die erste Haltestelle, und dann ist er vom HARRAS zu Am HARRAS gefahren und vom HARRAS zu Am HARRAS waren acht Leute, sind durchschnittlich an so einem Werktag im Januar im Bus. Und hier ist das fahrtengenau und hier können wir eben ganz gut einschätzen, ok, hier ist eine Überlastung, wie reagieren wir, oder wir wissen, das ist jetzt hier nicht der Fall, aber zum Beispiel, wir steigen erst hier mit dem Takt 10 ein und haben hier Takt 20, dann würde man bei so einem Teppich jetzt sagen, ok, dann müssen wir den Takt 10 wahrscheinlich schon hier oder hier einsetzen lassen. Also das ist so die Grundlage für die Bestandslinien die wir uns anschauen, und da gibt es tatsächlich ganz ganz unterschiedliche Bilder. Also hier ist so alles super, und dann dafür ist aber zum Beispiel hier dann schon relativ voll, und das ist dann immer die, da kann man dann immer Entscheiden, haben wir die Möglichkeit, einfach ein größeres Gefäß einzusetzen, dann haben die Leute mehr Platz in dieser einen Fahrt, oder verdichten wir eben den Takt, oder wenn das jetzt gezielt so Schülerfahrten sind, zum Beispiel ist hier alles im grünen Bereich, und plötzlich ist eine Fahrt die halt einfach voll ist, da gibt es dann auch extra Schülerverstärker, die dann auch wirklich ganz gezielt einfach noch mal nach der vollen Fahrt fahren. Und so können wir da auch reagieren.

description of load-carpet

detection of overload

visual impressions can vary a lot

*Ok. Gibt es außer der Tabelle noch andere Grundlagen, die ihr..*

Also die Tabelle ist eigentlich eine Zusammenfassung aus Daten, die in einer Datenbank liegen und wir greifen dann tatsächlich gezielt auch auf die Datenbank zu. Also wir können auch SQL-Abfragen machen und wenn wir jetzt zum Beispiel eine Fragestellung haben, wie schaut es an der und der Haltestelle aus, oder wir kriegen da irgendwie einen Input über Bürger, die sagen, das ist irgendwie immer voll, da können wir dann auch ganz gezielt uns einzelne Fahrten oder einzelne Haltestellen dann noch mal detaillierter anschauen, weil das ist ja immer ein Mittelwert aus diesen Fahrten im Januar. Und da können wir dann tatsächlich die Ist-Werte auch anschauen und das darauf dann nochmal ein bisschen besser beurteilen. Aber das machen wir dann eigentlich nur bei konkreten Fragestellungen.

targeted querying of database for specific questions, e.g. specific stops or following up input from customers

*Ok also in der Datenbank landen auch die ganz aktuellen Zahlen einfach direkt*

Genau die landen da direkt und beim [I] werden die dann aggregiert eben. Und sobald er die kontrolliert hat und eingepflegt hat, also das OK gegeben

hat, können wir auch drauf zugreifen.

*Das heißt, was die Auflösung angeht, sind wir jetzt zeitlich immer, also arbeitet ihr mit fahrtengenauen Daten*

Genau, fahrtengenaue Daten. Aber wir brauchen jetzt nicht zwingend die Ist-Daten, sondern wir brauchen eher immer so einen guten Überblick. Deswegen sind wir eben auch auf diese aggregierten Daten, die immer einen gewissen Zeitraum darstellen, der so zu fassen ist, wo man sagt, ok, in der Zeit verändert sich das Verhalten nicht so stark. Dann hast du wieder so einen Cut wie zum Beispiel Ferien oder Feiertage oder die Wiesn, müssen wir zum Beispiel auch immer extra behandeln, dass die Mittelwerte quasi nicht verfälscht werden, sondern dass es möglichst immer die Zeiträume fasst, in denen sich die Menschen wahrscheinlich relativ homogen verhalten, da hat mal der eine Urlaub, dann der Andere, aber nicht immer alle so auf einen Schlag, so.

temporal aggregation over similar periods, e.g. weeks between school holidays, and overview more important than very timely data

*Das heißt ihr habt so Tagestypen..*

..genau, Montag-Donnerstag-Schule, Montag-Donnerstag-Ferien, Freitag, Samstag, Sonntag und Feiertage sind auch immer das Gleiche.

*ja, und innerhalb, also für jede Gruppe sozusagen, schaut ihr euch dann den Ablauf innerhalb des Tages an auf diesem Niveau*

Genau. Grad bei, das macht man zum Beispiel bei der U-Bahn, wo man nicht so fahrtengenaue Zahlen hat, da gibt es eine Tagesganglinie, das ist dann nicht so ein Teppich sondern das ist dann wie so ein Diagramm mit Balken, wo man dann am Ende quasi so eine Linie rauskriegt. Das kriegt man theoretisch hier auch so ein bisschen raus, wenn man das jetzt hier drüberlegt, nur die können halt immer nur die Querschnitte betrachten und wir können tatsächlich hier auch sagen, ok das ist Münchner Freiheit, wir haben so und soviel Einsteiger, so und soviel Aussteiger und die Last so. Ja und halt einfach fahrtengenaue. Genau, das ist dann die Tagesganglinie in Zeitintervallen von zehn Minuten.

within each type of day, the daily cycle is analyzed, also using a bar-/line-chart ("Tagesganglinie") with a temporal resolution of ten minutes

*Ok und das sind so die zwei graphischen Dinge mit denen ihr arbeitet. Und dann Karten..*

Ja genau, was jetzt noch ein bisschen neuer ist, was wir jetzt auch vorantreiben wollen, weil wir das ziemlich gut finden, ist das das automatisiert wird, mal schauen ob es lädt, also quasi eine Karte, in der visuell dargestellt ist, welche Querschnitte überbelastet sind [opens document] ah jetzt, genau. Also zum Beispiel diese Karte, die ist jetzt, hat die Verkehrsforschung relativ neu, das kam jetzt mit der Coronazeit tatsächlich ein bisschen auf, das ist zum Beispiel ein tool, das fänden wir ganz cool, wenn das in Zukunft auch irgendwie zur Verfügung gestellt wird. Das greift auch auf die Datenbank zu und da kriegen wir quasi so einen Überblick, das ist jetzt der ganze Tag, wäre dann auch noch ganz cool, wenn man das auswählen kann, zum Beispiel zwischen sieben und neun Uhr, dass man sieht, wo sind denn Querschnitte, die überlastet sind, dann kann man hingehen, ok, ist jetzt auf dem Quer-

new development in conjunction with the pandemic: representation of load on network sections on a browser-based map-interface. Has been highly appreciated, intention to develop further automatisations of that kind of visualization

schnitt nur eine Linie, ok das ist hier jetzt nur eine Linie, dann wissen wir, ok, dann müssen wir uns die Linie mal genauer anschauen..

*..also man sieht hier eine Karte einfach..*

Genau, das ist jetzt OpenStreetMap denk ich mal und dann sind hier irgendwie unsere Linien hinterlegt und..

OSM background map with lines, data from database containing passenger counts

*..und das stellt die Daten aus der Datenbank direkt dar..*

Genau. Da haben wir halt dann quasi immer den Querschnitt, wahrscheinlich hier eine Haltestelle und hier eine Haltestelle, und dann zum Beispiel hier .. irgendwo gibt es ..

*..ok also man sieht die Linien, wo die Busse fahren und die sind eingefärbt..*

Genau, das sind auch nicht nur Busse, das ist tatsächlich auch Tram und .. nein, U-Bahn ist es nicht. Ohne U-Bahn, genau. Und dann..

data for tramways and buses

*..und gelb bedeutet, dass es nicht so sehr ausgelastet und rot, dass es überlastet..*

..überlastet ist, genau. Irgendwo ist zum Beispiel.. Hier ist dann ein Querschnitt, 130, aber da wissen wir, ok die Tram fährt hier auch parallel. Da muss man sich dann alle drei Sachen genauer anschauen, wenn jetzt mehrere Linien irgendwo fahren, aber das gibt schon mal ein gutes Bild, wo so die Problemschwerpunkte liegen. Das ist natürlich jetzt über den ganzen Tag verteilt vielleicht nicht gar so aussagekräftig, aber wenn man das jetzt noch aufteilen kann in .. Ah, hier zum Beispiel, sieben bis neun Uhr, da kann man das dann schon ein bisschen besser differenzieren, weil sich das Bild wahrscheinlich für Mittags dann schon ein bisschen unterscheiden wird. Genau, das sind jetzt immer relativ aktuelle Zahlen, und für uns wäre ja dann eher relevant, so einen ganzen Zeitraum sich dann wieder anzuschauen. Genau, das ist noch so eine visuelle Sache.

lines are color-coded according to load

different temporal aggregations, e.g. morning hours

*Sehr interessant, das geht in die Richtung was ich vielleicht auch dann machen kann. In welchen Einheiten denkt ihr? Ihr stellt die Anzahl der Personen, die unterwegs sind in Verbindung mit der Kapazität*

Nee, ja auch, wir haben immer die Auslastung, das ist hier zum Beispiel, das ist dann immer wie hoch die Auslastung ist, das kriegt man ja nur mit der Kapazität und grundsätzlich ist die Last so das entscheidende, Last und Auslastung, die dann gewisse Grenzwerte nicht überschreiten soll. Die Grenzwerte sind übrigens auch wieder im Nahverkehrsplan festgesetzt. Und genau, wenn wir eben jetzt so einzelne Haltestellen betrachten, dann schauen wir uns auch immer noch Einsteiger, Aussteiger an und was uns fehlt bei diesen ganzen Zahlen, was natürlich ein großer Vorteil der Handydaten wäre, sind eben Bewegungsprofile. Dass wir sagen können, wir haben eine Reisekette. Wir haben immer nur, eine anonyme Person steigt hier ein, und wir wissen, eine anonyme Person steigt hier aus, wir wissen aber .. wir können ir-

complete travels including changes between lines are missing - hope to get that out of mobile network data

gendwie so Reisezeiten schätzen und wir haben Befragungen wo wir sowas rauskriegen, wie weit die im Durchschnitt unterwegs sind und sowas, aber wir wissen nicht, ob jetzt der Großteil der Personen wirklich an der Münchner Freiheit aussteigt und alle wollen irgendwie zum Harras, sondern das ist ja ein reger Wechsel und so, das fehlt uns tatsächlich ein bisschen.

stations at which many people change lines are of interest

*Das heißt, würde es auch schon was bringen, zu wissen, wie lange Personen zu welchem Zeitpunkt unterwegs sind?*

Ja

*Weil mit Datenschutz ist es schwierig, solche genauen Trajektorien zu haben, deshalb arbeite ich auch auf einer relativ aggregierten Basis*

Ja genau, also diese Reiseketten kriegen wir ja immer aus so Umfragen, die jährlich gemacht werden aber bei diesen Reisezeiten ist dann natürlich immer die Frage, kann man ganz klar sagen, der fährt jetzt ÖV, kann man ihn dann ganz klar dem ÖV zuordnen? Wenn man das klar sagen kann und sagen kann, jetzt fährt der ÖV und das zu einer so langen Zeit, dann bringt uns das schon was.

travel interval information in connection with information on whether or not public transport is being used would be useful

*Ok. und zu wissen, von wo nach wo die Leute insgesamt reisen, das wäre auch was.*

Ja genau. Wir können sowas über das Verkehrsmodell ja anschauen lassen, aber das ist ja einfach nur ein Modell, das sind irgendwelche Schätzungen, die ja auf diesen Erhebungen basieren und das dann einfach hochrechnen. Die sind auch wirklich nicht verkehrt aber genau, grad mit so Veränderungen in der Stadt ist dann das Modell natürlich ein bisschen träge, weil es halt dann erst wieder aktualisiert wird, wenn das ganze Modell wieder aktualisiert wird und ja und auch grad so zu Großereignissen oder sowas, das ist dann halt irgendwie so ein Erfahrungswert aber.. ..ja also grad wenn es dann von diesem Standardtag abweicht halt.

Actual (not simulated) origin-destination information is desired, especially on days with abnormal travel patterns

*Ok. .. Wenn wir die räumliche Auflösung angucken, da arbeitet ihr ja mit Stationen hauptsächlich, was ja bei den Mobilfunkdaten schwierig ist, weil man die nicht so sehr genau verorten kann. Was wäre eine minimale Auflösung, die schon was bringt? Hilft es, wenn ich weiß, wie viel um die Stammstrecke rum passiert im Vergleich zum Außenraum, oder ..*

*Spatial resolution in the context of underground might be sufficient if line is identifiably and origin and destination can be identified up to a small number of stations. The bus network in the city is very fine, rough spatial information is not very useful.*

Also bei.. naja da müssen wir jetzt vielleicht ein bisschen unterscheiden, also bei der U-Bahn ist es vielleicht tatsächlich, wenn man sagen kann, welche Linie und ungefähr, wenn man sagt, ok, das ist im Bereich dieser drei Haltestellen. Oder so. Also wenn es jetzt wirklich nur so.. ..da ist unser Busnetz halt viel zu fein, dass wir sagen, ... dass wir daraus dann irgendwie was ableiten können für eine Planung. Das ist bei der U-Bahn dann eben vielleicht tatsächlich ein bisschen besser anwendbar, dass man es auf die Linie definieren kann und eben zumindest welchen Abschnitt, ... ja also dadurch dass wir beim Bus so genaue Zahlen haben mit den Zählgeräten, haben wir insofern tatsächlich eine sehr gute Datenlage. Das ist schon so eine feine Ebene, dass wir da wirklich auf die Fahrt genau sagen können, die ist meistens über-

füllt. Nur bei dieser einen Fahrt müssen wir irgendwas machen oder sowas.

*Also das was da relevant ist, sind wirklich diese..*

..Querschnitte. Genau und dass man sagt, dass man zum Beispiel sieht, auf welcher Linie der Effekt zu spüren ist. Eigentlich müsste es schon dann auch liniengenau sein, dass man auch vielleicht Parameter verändern kann, dass man auch sieht, auf was es sich auswirkt.

*Gut. Dann fällt mir jetzt grad nichts mehr ein. Wenn du noch was hast, was ich sehen soll oder von dem du denkst, dass es interessant sein könnte..*

Ja also tatsächlich sind das eigentlich unsere wichtigsten Grundlagen.. ...Ja also so Reisezeitvergleiche, wie komm ich jetzt zu dem Ort und wenn ich dann irgendwie eine neue Linie hab, eine neue Verbindung, wie groß ist mein Zeitvorteil, wenn ich jetzt da direkt hinfahren kann, wo ich da natürlich immer wissen muss, wie lang brauch ich denn dann von da nach da ungefähr.

*Das ist auch was, was ihr aus dem Modell kriegt..?*

Nein das ist händisch eigentlich. Aus dem Modell gibt es das grundsätzlich auch, diese Zeitersparnisse und sowas, das schon, aber manchmal ist die Linienführung auch klar und dann sagen wir, wir modellieren es jetzt auch gar nicht, und dann macht man das irgendwie händisch. Dann gibt man das einfach in dieses Fahrplantoolein und sagt, ok, ich möchte jetzt vom Harsras zur .. nach Feldmoching oder sowas und dann, wie leitet der mich jetzt und wie lange brauch ich, und dann hab ich ja quasi meine Planung von der Direktverbindung und sag, ok, mit den Durchschnittsgeschwindigkeiten brauch ich da so und solange und dass man das dann halt so aufschlüsselt.

Comparison of travel times gained from schedule information

*Und diese Durchschnittsgeschwindigkeiten beruhen auf..*

Also dadurch, dass wir schon über eine Vielzahl der Straßen fahren, wo man auch sagt, irgendwann sind das dann auch alle Straßen, die mit einem Bus überhaupt befahrbar sind, weil manche es einfach von der Breite oder den Höhen oder was weiß ich, nicht hergeben, haben wir bei dem Großteil natürlich auch dann schon Erfahrungswerte, wo wir sagen, da können wir die Reisezeiten von einer bestehenden Buslinie übernehmen und dann, wenn wir tatsächlich eine komplette Neuerschließung haben ist es so ein Durchschnittswert, den gibt es auf jeden Fall, aber der variiert zum Beispiel auch zwischen Schnellbus und Metrobus und also so ein Stadtbus, so ein normaler, ist ein bisschen langsamer unterwegs als so ein Schnellbus und so Sachen.

*Ja, spannend. Dann beende ich mal die Aufnahme.*

### **Interview III**

The start of the interview was not recorded due to abruptness, the interviewee was talking about different departments



A ... einen Tarifbereich, da geht es natürlich um die Weiterentwicklung des Tarifs, einen Verkehrsforschungsbereich, die machen hauptsächlich Fahrgastzählungen und -befragungen, das man sieht wo ist die Nachfrage, wie ist die Auslastung, wo muss man vielleicht Takte verdichten oder so. Was gibt es noch - einen Regionalbusplanungsbereich, die entwickeln halt das Regionalbusnetz weiter und eben uns als konzeptioneller Planungsbereich.

..und mit "konzeptioneller Planungsbereich", was..

A Ja wir machen nicht nur konzeptionelle Planung, wir machen natürlich auch sehr viel Fahrplan, das ist auch hier im Bereich angesiedelt. Fahrplanerstellung und -vertrieb, da geht es tatsächlich um den Druck der Fahrpläne und Fahrgastinformation im weiteren Sinne, das waren früher ausschließlich oder weitgehend Printmedien und dann kam vor .. Mitte/Ende der Neunzigerjahre kam die elektronische Fahrplanauskunft, also eigentlich schon Ende der Achtzigerjahre aber da war es noch Bildschirmtext und - alles nicht wichtig für Sie, die EFA in der Form wie wir sie jetzt kennen haben wir seit Ende der Neunzigerjahre, als Onlinetool, zuvor gabs das auf CD und in der letzten Fortschreibung ist die EFA noch sehr viel stärker Kartenbasiert geworden, also da finden Sie jetzt ja im Hintergrund immer eine Umgebungskarte wo die Fahrwege dargestellt werden, die Umstiege die Fußwege, wir haben seit auch schon über zwanzig Jahren eine Tür-zu-Tür-Auskunft, wo eben auch Fußwege beauskunftet werden, Park-und-Ride, Bike-und-Ride, also die ganze intermodale Wegeketten wird dargestellt und seit dem letzten Update von der EFA jetzt eben sehr viel stärker kartenbasiert, dass man im Hintergrund immer die Karte offen hat und da halt die Wege sieht. Karten sind nicht jedermanns Sache, manche Leute lesen lieber einen Fahrplan in der Tabellenform, andere orientieren sich lieber an der Karte, das ist ja individuell unterschiedlich, wir bieten da eben beide Wege an. Und mit die höchste Reichweite oder Wirkung hat natürlich unsere Fahrplanauskunft mit Millionen von Zugriffen zum Teil täglich sogar, oder Fahrtauskünften, die wir rechnen, also hat sich eigentlich immer weiterentwickelt, inzwischen ist ja die mobile Auskunft sogar noch wichtiger als die rechnergestützte Internetauskunft, also das verschiebt sich alles hin zum Smartphone. Aber das ist eben jetzt Kundeninformation und Fahrplaninformation. Was wir machen ist ja eher die konzeptionelle Seite, also Planung für die Zukunft, wie werden wir das Netz weiterentwickeln, wo werden wir den Takt verstärken müssen, wo müssen oder können wir neue Linien überlegen, und und und. Und dafür hatten wir eine Datenbasis, die war extern, die hat die Firma Intraplan immer wieder fortgeschrieben und da gibt es vom Bereich Verkehrsforschung gefütterte Ist-Daten, also Analysedaten, und in die Zukunft gerichtete Prognosedaten wo eben künftige Einwohner, Arbeitsplätze, Verkehrsmittelnutzung und so weiter da einfließt. Das war eigentlich analog. Das sind zwei dicke Ordner gewesen immer, ein Analyseteil mit den ganzen Tabellen und Verkehrszellen und ein Anlagenteil wo die ganzen Karten drauf waren, also Verkehrsbelastung und so weiter und so weiter. Wir haben dann vor ungefähr 15 Jahren dann versucht ein eigenes Planungstool zu nutzen, das hieß RegioPlan, das war von Intraplan im Prinzip das gleiche womit die die standardisierte Bewertung für große Verkehrsprojekte rechnen. Das war aber dann letztlich haben wir das nicht weiterverfolgt oder weiterentwickelt, das war zu komplex, zu einseitig auf München ausgerichtet und wurde wohl auch nicht

The interviewees are working within conceptual planning. Other departments are tariff planning, traffic research and regional bus planning

Conceptual planning is also responsible for passenger information which has lately been focusing on making maps the main interface for timetable information while also providing other interfaces like a classic timetable as passengers have a variety of preferences

As for conceptual planning, the main objective is to find out how and where to increase the network spatially or temporally densify journeys. To that end, a multitude of data sources are being used: analytical data, future residents, workplaces and transportation usage among others, in the form of tables and maps. Attempts have been made to digitize the process and today, they are mainly working with a VISUM traffic model which is shared among several stakeholders.

weiterentwickelt. Und jetzt haben wir ein anderes Verkehrsmodell, das wir mit der Landeshauptstadt München zusammen haben und der MVG, das in Kooperation entstanden ist und eben weiterentwickelt wurde.

*ist das das in VISUM läuft?*

**A** Genau, das VISUM-Verkehrsmodell. Und es gibt jetzt auch noch ein bayerisches Verkehrsmodell, das Bayernmodell, parallel dazu, das macht im Prinzip das Gleiche oder Ähnliche was vorher Intraplan oder Regioplan mit der standardisierten Bewertung oder mit ihren Planungstools gemacht haben, aber ist etwas abgespeckt, sodass wir halt kleinere Anfragen oder schmalere Untersuchungen im Rahmen von einem Consulting-Projekt, das kann man eben mit dem Verkehrsmodell berechnen. Wenn es dann in die politische Entscheidung geht, ob man eine Strecke wirklich weiter feinkonzipiert oder eben in die Feinplanung gibt, da muss dann wohl die standardisierte Bewertung nach wir vor durchgeführt werden, das macht dann weiterhin Intraplan extern, und da gehts ja darum, dass jedes Projekt, das öffentlich gefördert wird, einen höheren Nutzen haben muss als es Kosten hat. Also einen Nutzen-Kosten-Faktor über eins-komma-null, das war immer so die wichtige Kenngröße, ansonsten wird das Projekt halt zumindest nicht mit Bundesmitteln oder Landesmitteln gefördert. Aber für kleinere Anfragen nutzen wir eben jetzt unser Verkehrsmodell zunehmend, über das ich aber gar nicht so gut Bescheid weiß, da ist [B] der Fachmann dafür. Da kann man eben auch Netze kreieren, da ist auch das Wegenetz hinterlegt, das Straßenwegenetz, die Straßenklassen wohl auch Fußwegenetz und Radwege weiß ich nicht, und eben das heutige Verkehrsgerüst vom ÖPNV und man kann in dem Tool halt relativ einfach, bitte korrigier mich wenn ich Schmarrn erzähl, neue Trassen planen, also eine neue S-Bahn, ne neue Straßenbahn, was auch immer, und neue Verkehrsumlegungen machen, mit dem Ist-Zustand, und da sieht man dann halt, wie sich der Verkehr verlagern würde, wenn man jetzt eine zusätzliche Linie einrichten würde, was so einen ersten Eindruck gibt, ist da die Nachfrage da. Aber ich hab wahrscheinlich schon mehr gesagt über das Planungstool als ich weiß oder sagen sollte, deswegen müsste an der Stelle [B] weitermachen.

Two VISUM traffic models are used to test possible new connections and get insight into expected demands. Before realisation of new projects, a standardised evaluation has to show that the expected benefit is higher than the costs.

**B** Ja nee, du hast schon eine gute Zusammenfassung gegeben. Ich denke erstmal, das wäre quasi eine Frage an Sie: In welche Richtung, also haben Sie schon so einen Fragebogen für uns vorbereitet oder ist das eher ein informales Interview oder..

*Ja das ist eher ein qualitatives Interview jetzt, genau. Sie erzählen mir einfach..*

**B** Also keinen festen Fragen oder..

*Nein, also ich hab so einen Leitfaden aber sonst nicht nein*

**B** Ah ok, also weil Sie ja das Thema Mobilfunkdaten im Fokus haben, vielleicht kannst du (A) auch ein paar Sätze zu euren Teralyticsdaten kurz sagen, weil wir haben auch das seit ein paar Monaten neu gekauft und ich denke das geht eher in Richtung dem was für Sie interessant ist wahrscheinlich.

A Ja, gerne. Wir machen derzeit eine großangelegte Studie zur verkehrlichen Sinnhaftigkeit einer MVV-Verbundraumerweiterung. Die ist ja schon länger in Diskussion und von verschiedenen Landkreisen auch gefordert oder angestrebt, und jetzt untersuchen wir für acht Landkreise und zwei kreisfreie Städte eine MVV-Erweiterung und im ersten Schritt eben die Verkehrsbeziehungen, wie sie heute sind im Status Quo, und wo eben Verkehrsnachfrage ist. Dazu nutzen wir alle möglichen statistischen Daten die eben verfügbar sind, haben die zum Teil auch gekauft, Gemeindedaten, Pendlerdaten und so weiter. Worüber es relativ wenig Datenmaterial gibt, ist natürlich der Freizeitverkehr und auch für den Individualverkehr haben wir halt nur für bestimmte Straßenquerschnitte was. Also um den Verkehr ganzheitlich abbilden zu können, haben wir uns gedacht, sollten wir uns mal Mobilfunkdaten anschauen. Die wurden ja von verschiedenen Netzbetreibern angeboten, da haben wir uns informiert und haben letztlich .. die Firma heißt Teralytics und die nutzt Telefonica-Mobilfunkdaten. Da muss man natürlich wissen, wie sind die Marktanteile von Telefonica, ist das repräsentativ für die Gesamtbevölkerung, wie rechnen die das hoch und so weiter. Das schaut aber eigentlich alles sehr gut und sehr repräsentativ aus. Die Daten haben wir gekauft, jetzt zunächst mal für ein Jahr. Also das ist eigentlich ein Lizenzprodukt, was monatlich abgerechnet wird und das haben wir jetzt seit März und werden das dann halt auf jeden Fall mal bis nächstes Jahr Februar, dann ist die eine Jahreslizenz quasi aufgebraucht, und dann muss man sehen, wie und in welchem Umfang wir sie weiter beziehen. Und da haben die Kollegen von Teralytics eben Verkehrszellen geschnitten und können dann halt Auskünfte geben, wie viele Fahrten von einer Verkehrszelle zur nächsten stattfinden. Die gibt es in unterschiedlicher Größe, die werden je nach Zoomstufe aggregiert oder so wie es der Kunde halt braucht. Also es gibt verschiedene Datensätze sag ich jetzt mal. Im Wesentlichen sind das zwei - ein bundesweiter Datensatz insbesondere für die Langstrecke, das ist für uns jetzt im Verbundverkehr nicht so interessant, weil wir jetzt weniger auf Bahnfernverkehr oder Flugreisen oder sowas abzielen, sondern halt auf den Nahverkehr. Deswegen haben wir den Kurzstanz-Datensatz heißt der, gekauft. Der ist Länderweise verfügbar, wir haben da natürlich auch nicht ganz Bayern genommen sondern die Bereiche, die uns interessieren und wo wir halt die Verkehre abbilden wollen. Und da schauen wir halt uns jetzt an, wie sind zu bestimmten Tagen, Tageszeiten die Verkehrsbeziehungen von einer Zelle zur anderen, vergleichen das dann auch noch mit den Daten die wir über statistische Landesämter oder andere Stellen eben haben um zu schauen, in wie weit ist das repräsentativ, passt das zusammen, ist das plausibel.

*..wie groß sind die Zellen ungefähr, mit denen Sie arbeiten?*

A Gemeindeebene. Und das ist schon ziemlich groß..

*..das heißt, man weiß nur, wie die Leute sich zwischen Gemeinden bewegen*

A Ja. Und das ist eigentlich ungewöhnlich groß weil die meisten kommerziellen Einsatzzwecke die interessieren sich dafür, wie viele Leute gehen an meinem Schaufenster vorbei oder wo ist eine gute Lage für ein Geschäft oder, ja, wie ist das Publikumsaufkommen, verteilt sich das in den Straßen, und sind viel kleinräumiger orientiert. Im Prinzip können die die Daten runterbrechen

current objective is to determine sensibility of expansion of area connected the by the transport association. First step is to analyse the current traffic relations and demands

a dash board tool for the investigation of mobile network data has been acquired to complement official workplace statistics. After first comparisons, the quality of the mobile network data has been deemed good and representative

investigate data for specific days, times of day and traffic relations between different cells, comparison of different data sources for validation of plausibility

very coarse spatial resolution (municipal regions) is sufficient and adequate for current objective

bis auf Mobilfunkmastebene, das ist dann die relativ kleinste Einheit, es geht dann zum Teil noch feiner, wenn man GPS noch berücksichtigt oder Trigonometrie, also mit Empfangsstärkenfeldmessung, dann kommt man auch noch in kleinere Einheiten rein als nur Funkmastebene, wobei die ja im städtischen Bereich eh schon relativ dicht sind. Für unsere Zwecke ist das nicht erforderlich und auch nicht zielführend. Wie wollen jetzt erst mal eine Strukturuntersuchung machen, wo wir die groben Verkehrsströme untersuchen wollen.

*..da gehts ja wahrscheinlich auch mehr um Pendelverkehr und solche Muster..*

**A** Ja, Freizeitverkehr am Wochenende, Pendlerverkehr unter der Woche, aber die Telefondaten unterscheiden ja nicht nach Verkehrszweck. Also ich kann jetzt kein Häkchen setzen, ich möchte mir nur den Freizeitverkehr anschauen, das wäre schön, aber das geht leider nicht, das sind alle Verkehre. Und wir müssen dann praktisch Einschränkungen machen, zum Beispiel, dass man die Berufsspitze am Morgen nicht betrachtet und außen vor lässt, oder nur das Wochenende sich anschaut, dass wir jetzt bestimmte, ja, Verkehrszwecke herausfiltern können.

*..in was für zeitlichen Einheiten arbeiten Sie da, stündlich oder..*

**A** Ja. Wir kriegen Stundenwerte, theoretisch müsste es auch feiner gehen von der Datengrundlage, nur kriegt man dann halt keine Zahlen mehr, weil man dann sehr schnell in den Bereich kommt wo die Anonymisierungsgrenze unterschritten ist. Also alles unter fünf Fahrten wird da nicht dargestellt, also einfach ausgeixst der Wert, und deswegen muss man da natürlich zeitliche Ebenen zusammenfassen, wo man dann halt genügend Fahrten hat, dass sie statistisch relevant werden. Aber wie gesagt, das Tool haben wir jetzt seit März diesen Jahres und wir sind noch in der Erprobungsphase, wir haben noch nicht sehr viel damit gearbeitet und es werden auch noch Funktionen dafür entwickelt, die wir für unsere Fragestellungen brauchen. Zum Beispiel, dass wir als Anwender auch die Zellgröße selbst bestimmen können. Dass wir die Verschmelzung vornehmen und nicht das System automatisch. Da sind wir in Kontakt mit dem Anbieter oder mit den Kollegen eben von Teralytics und die sind, das ist kein fertiges Produkt, sondern das wird nach Kundenbedürfnissen eben weiterentwickelt.

*Ok. Das heißt eine bisschen größere Freiheit in der Interaktion ist momentan angestrebt.*

**A** Ja, jeder Kunde hat da sicher andere Bedürfnisse und ich glaube auch, dass wir da der erste Verbund sind, der die Daten jetzt nutzt. Und ansonsten sind das halt institutionelle Kunden, also ich weiß nicht, Ministerien zum Beispiel oder auch private, die halt das für Werbezwecke nutzen wollen oder für wirtschaftliche Zwecke. Da hat natürlich jeder andere Anforderungen. Und wir machen eine Verkehrsuntersuchung, da haben wir natürlich andere Bedürfnisse als jetzt jemand, der das für seinen Werbeverteiler braucht oder die Verbreitung von Covid untersuchen will oder was auch immer.

*Ok. Wie sind die Daten dann aufbereitet, also wie interagieren Sie mit den Daten dann*

purpose of transport is of interest but not directly represented in mobile network data. Specific temporal filtering is being used to identify the assumed purpose of transport

data based on hourly intervals

desire for more control over area-specific grade of aggregation

*momentan? Wie sieht das aus, ganz praktisch, auf dem Bildschirm?*

**B** Das ist so ein online-Dashboard, und jeder kriegt quasi dann ein Konto.

**C** Man kann sich das so vorstellen, das ist eigentlich im Hintergrund eine Google-Maps-Karte oder OpenStreetMap, man kann da bestimmte, da gibt man eben ein, ähnlich wie bei Google, wenn man eben eine Route plant, einfach von, entweder die Gemeinde direkt anklicken, oder man gibt ein von Ottobrunn nach Garmisch-Partenkirchen, dann kann man noch Parameter einstellen, ob das jetzt alle Wochentage anzeigen soll oder nur einen Tag oder auch beliebige Tage in der Woche, welchen Monat man sich anschauen will, also man hat immer die Daten vom aktuellen Monat und halt auch die vom vergangenen Jahr, man kann auch mehrere Monate auswählen, und dann kriegt man auf dem Bildschirm dann quasi so einen Pfeil angezeigt, wie viele Fahrten an diesem jeweiligen Tag, den man ausgewählt hat, von Ottobrunn nach Garmisch gehen und auch zurück. Genau und wenn man halt mehrere Gebiete auswählt, das geht auch, ich kann auch sagen, von Ottobrunn in den ganzen, in alle Gemeinden des Landkreises Garmisch-Partenkirchen zum Beispiel, genau, da kriegt ich halt einfach für jeden Verkehrsstrom nen Pfeil, der dann je nach dem, wie viele Fahrten da drauf sind, unterschiedlich dick ist. Ja und dann hat man halt so eine, wir nennen das immer "Spinnenkarte", weil wir das bei den Pendlern so bezeichnen, Pendlerströme, da hat man halt so einen groben Eindruck, wo viel los ist und wo wenig los ist, also so kann man sich das optisch vorstellen.

spatial filtering according to region of origin and/or destination, temporal filtering and aggregation according to days of week and months

number of travellers represented by thickness of line or arrow

*Mhm, könnte ich mir das mal anschauen, vielleicht später nach dem Interview dann.*

**C** Ja, genau, am Schluss dann.

**A** Das können wir uns gern bei mir am Rechner oder auch bei den Kollegen anschauen, das ist ein bisschen andersrum aufgezümt wie wir es sonst machen. Sonst sammelt man ja ganz viele Daten und zur Visualisierung referenziert man die oder bringt die auf eine Karte. Bei dem Dashboard, da sieht man eigentlich erst man alles auf der Karte und es gibt aber die Möglichkeit, die Rohdaten in csv zu exportieren oder auch in ein Shapefile, wo man es dann mit ArcGIS weiterverarbeiten kann. Aber den ersten Eindruck bekommt man visuell eben über die Karte oder über die Spinnen.

method of working with a visual interface as a first introduction to the data is new but seems to be appreciated. Different export-options for further use of filtered data

*Ok, das heißt, mit ArcGIS arbeiten Sie dann auch im Anschluss.*

**C** Ja. Also wir haben auch, hat ja [B] ja schon angesprochen gerade, mit den Pendlerdaten, die kommen jetzt nicht nur von Teralytics oder den Mobilfunkdaten, sondern es gibt auch von der Bundesagentur für Arbeit Statistiken für jede Gemeinde, wo Sozialversicherungsbeschäftigte eben ihren Arbeitsplatz haben und da kann man eben ablesen, aus jeder Gemeinde pendeln so und so viele Leute in eine andere Gemeinde, und die Daten haben wir eben für unsere Zwecke auch aufbereitet und dann aber, ich sag jetzt mal händisch, dann auch visualisiert, sodass man dann für einen Landkreis sieht, aus welchen Nachbarlandkreisen oder Gemeinden die Leute dann auch pendeln. Das hat jetzt mit den Mobilfunkdaten nichts zu tun sondern da hat man

processing of workplace statistics and manual visualization for investigation of specific spatial relations

dann eben auch eine andere Datenbasis, wo dann wirklich eine genaue Statistik vorhanden ist, wie viele Leute aus Gemeinde XY arbeiten im Nachbarlandkreis. Genau, und das ist dann eben der umgekehrte Weg wie von der Karte auf die Tabelle, gehts quasi da von den Daten ins Grafische und da haben wir dann Zeichenprogramme, wo man das dann auch individuell einfach einzeichnen kann. Geht natürlich auch mit ArcGIS, da ist der Kollege (B) der Experte, wir haben das immer manuell dann noch gezeichnet, aber das ist dann auch manchmal fast schneller, wie wenn man jetzt dann in ArcGIS das erst einprogrammieren muss.

*Das heißt, wenn man die beiden Datenquellen vergleicht, dann..*

**C** ..sollten dann hoffentlich ähnliche Strukturen rauskommen, das kommt auch immer darauf an, welchen Tag man nimmt. Wenn man jetzt einen Wochenendtag auswählt bei Teralytics, dann sind natürlich andere Verkehrsbeziehungen zu erwarten wie jetzt unter der Woche. Und unter der Woche sollten dann im Prinzip dieselben Strukturen rauskommen wie jetzt auch bei den Pendlerzahlen. Wobei natürlich die Pendlerzahlen auch nicht alle, da sind jetzt keine Studenten dabei, und Beamte auch nicht, das sind wirklich sozialversicherungspflichtig Beschäftigte. Aber es ist schon eine repräsentative Menge.

knowledge about different completeness characteristics of different data sources, comparison is done for validation and better understanding

**A** Das sind keine Beamten, keine Minijobber, keine Selbstständigen vor allen Dingen auch. Natürlich keine Schüler, Studenten, die ganzen Versorgungsfahrten.. Also was Teralytics liefert sind ja alle Verkehre, und Pendler ist halt ein Teil des Berufsverkehrs also das ist natürlich nur eine Teilmenge des Ganzen.

Knowledge about limitation of representativity of official workplace statistics

**C** Also ich weiß nicht, ob Sie das interessiert, wie das dann auf dem Plakat aussieht, wie wir das jetzt machen, sonst hol ich noch schnell die Präsentation von letztem Dienstag, da ist es ja alles drauf. Dann hol ich das mal schnell.

**A** Das macht es sicher anschaulicher, wenn man sich eine Karte direkt anschaut, als nur so theoretisch darüber spricht.

*Ja.. Das heißt der Hauptgrund, warum sie auf die Mobilfunkdaten jetzt zugreifen, ist, dass da die Gesamtheit jetzt repräsentiert ist.*

**A** Ja, und

*..gibt es noch andere..*

**A** ..und die Aktualität natürlich. Andere Daten sind immer mindestens ein Jahr alt oder zwei oder werden nur alle paar Jahre fortgeschrieben. Und da kriegen wir halt jetzt im Juli die Daten für Juni. Das ist natürlich schon großartig also aktueller gehts eigentlich nicht. Selbst wenn wir sie selbst erheben, die Daten, müssen die erst eingegeben, aufbereitet, plausibilisiert werden, da vergehen natürlich auch Monate, bis man dann die eigens erhobenen Daten

incentives to use mobile network data are completeness (compared to workplace statistics) and timeliness of the data. Internal counting data has to be processed for several months before usage

auch nutzen kann und bei den Mobilfunkdaten sind die im nächsten Monat verfügbar, das ist natürlich von der Aktualität nicht zu schlagen.

*Ja. .. Und die Hauptfrage ist dann immer, von wo nach wo die Leute gehen oder..*

**A** Ja, und wie es sich über den Tag verteilt, und über die Woche. Aber natürlich Quelle-Ziel ist immer die wichtigste Frage.

two main analytic questions are origin-destination information and temporal evolution within a day

**C** [comes back with a printed presentation] Ja also wir haben das dann immer, das sind jetzt nur die Pendleranalysen, aber da gehts eben immer darum, es geht jetzt um diese Verbunderweiterungsstudie und da haben wir uns eben jeden Landkreis einzeln angeguckt, immer den Landkreis der untersucht wurde ist eben gelb dargestellt und dann aus jeder Gemeinde die Pendlerzahlen in die Nachbarlandkreise, das ist hier dargestellt. Und das sind aber die sogenannten Ströme, sowohl die Pendler, was haben wir jetzt da für ein Beispiel.. Das ist Landshut, Stadt Landshut, in den Landkreis Freising, und auch zurück. Und der Pfeil quasi ist dann immer der Wert für die Pendler quasi die rausgehen und reinkommen also Ein- und Auspendler.

**A** Wir haben zwei Karten gemacht, weil in einer wäre das ein bisschen zu unübersichtlich oder überladen gewesen. Die eine zeigt die Beziehungen mit den Nachbarlandkreisen, die andere eigentlich nur die Beziehungen nach München. Und damit die Karten dann trotzdem für sich selber stehen haben wir dann nachträglich hier trotzdem noch die Werte von den Nachbarlandkreisen aufgenommen und hier auch den Wert für München. Die Strahlen zeigen immer die Verkehre in beiden Richtungen in dem Fall. Also wenn man da jetzt noch Richtungspfeile rein.. dann wirds auch wieder unübersichtlich. Da geht's ja auch nur drum zu sehen, von wo nach wo, was ist das Hauptquellgebiet, und wo gehts hin.

for static maps, several maps are generated with visually reduced figures for reference in order to reduce complexity or visual overload

*..und dann schauen Sie sich den zeitlichen Verlauf extra an im Prinzip oder..*

**A** Ja, das sind jetzt keine Mobilfunkdaten, das sind die Daten der Bundesagentur für Arbeit, also die Pendlerdaten eben.

*Da geht man einfach davon aus, dass es morgens in die eine Richtung geht, und abends in die andere.*

**B** Das sind Ein- und Auspendler oder? Also die sind in beide Richtungen.

**C** Ja, aber da gehts in erster Linie darum, wo sind überhaupt Verkehre, wann die Leute da fahren ist jetzt im Endeffekt gar nicht so wichtig, sondern dass man halt zeigt, bei Landshut ist das ganz schön, dass man da auch sieht, es geht auch viel jetzt zumindest in den Osten in den Landkreis Dingolfing, und hier wollen wir eben auch zeigen, ja, macht das denn für den Landkreis überhaupt Sinn, dem MVV beizutreten, sind da überhaupt ausreichend Verkehrsbeziehungen da. Und deshalb diese Darstellungen. Und man sieht eben auch bei Landshut, die absolut größte Zahl geht nach Dingolfing und Landau. Aber natürlich sind auch genug Verflechtungen mit den Nachbarlandkreisen, die

main interest in amount, not direction of travel

current aim is the investigation of the sensibility of an expansion to further regions

eben jetzt schon im MVV sind oder dann zukünftig Mitglieder im MVV wären.

*Ja. Und die Karten sind aber, also die zeichnen Sie und die sind aber dann statisch, also..*

**A** Die zeichnen wir, also in dem Fall sind das ArcGIS-Karten eben.

*..die aber auch nicht interaktiv sind..*

**A** Die auch nicht interaktiv sind, die einfach nur Daten visualisieren sollen. Und das ist jetzt auch.. da bin ich mir jetzt gar nicht sicher, wahrscheinlich wird die auch gezeichnet sein.

non-interactive maps used for presentation purposes

**C** Die ist auch gezeichnet, ja.

**A** Das ist auch keine ArcGIS-Karte..

**C** Also der Hintergrund schon - die Hintergrundkarten, die Layer mit den Gemeindegrenzen, das ist natürlich alles in ArcGIS vorbereitet, und die roten Pfeile sind dann mit einem Zeichenprogramm dann drübergelegt, wie gesagt, das könnte man natürlich alles auch mit ArcGIS machen aber wir, bisher, praktizieren das meistens anders, genau, da muss man halt die Daten erstmal aufbereiten aus den Statistiken und das alles dann dementsprechend visualisieren.

**A** Das sind jetzt Teralytics-Karten, die sehen so aus.

*mhm, da ist ein bisschen mehr los..*

**A** Ja, das kommt ja auch drauf an, was man alles auswählt. In der Karte werden jetzt praktisch alle Verkehre von München in den Alpenraum gezeigt und im Alpenraum ist halt hier jetzt man exemplarisch Garmisch angeklickt und deswegen sieht man hier das jetzt noch mal in dick, das andere ist ein bisschen zurückgefahren, aber hier sieht man halt alle Zielorte quasi, also was wird aus München alles angefahren, hier sieht man die Verteilung über die Woche, das ist die Tagesganglinie, sehr zweigipfelig. Was ist das da wieder .. das ist die Distanz, ja also die Entfernung. Also ich mein, das ist jetzt alles eher weniger aussagekräftig, weil das sind ja alles Verkehre von München in den Alpenraum, da sollte sich die Distanz, natürlich verteilt die sich auch weiter, je nachdem ob ich an den Tegernsee fahre oder den Chiemsee oder die Allgäuer Alpen, aber das sind so die Standardausgabeparameter, das sind die Wochentage, die Stundenintervalle, die Distanz, und das sind die wichtigsten Verknüpfungen, an der Stelle sieht man eben die Top-Ziele und die Top-Ausgangspunkte. Und das ist jetzt im Vergleich, das sind die Verkehre zwischen den Kreisstädten der umliegenden Landkreise und der Landeshauptstadt. Die haben einfach ein anderes Muster, also das soll einfach nur ein Vergleichswert sein, der zeigt, ah ok, schau mal, hier hauptsächlich unter der Woche. Weil es nämlich Wirtschaftsverkehr ist. Da sehr viel am Wochenende, weil in die Alpen fährt man halt mal in der Freizeit. Oder auch hier die

structure of evolution throughout the day

day of the week, hour of the day and travelled distance as standard parameters. Visual identification of extremal datapoints from ordered bar charts



Tagesganglinie, steile Kurve am Vormittag, weil halt ungefähr alle bis zu einer bestimmten Zeit anfangen müssen zu arbeiten. Am Nachmittag die ist viel flacher, weil es gibt ja viele Halbtagskräfte, manche gehen früher, andere bleiben länger, da verteilt es sich mehr, während beim Ausflug, da fahren am Abend die meisten zurück und nicht mittags oder zwischendurch. Also da sieht man einfach strukturelle Unterschiede, die wir dann interpretieren müssen. Das Programm sagt uns einfach nur "Anzahl der Fahrten von A nach B". Und was das dann für Verkehrszwecke sind, das ergibt sich dann daraus, wie ich die Anfrage zugeschnitten hab, und dafür sind diese Hilfsgrößen ganz hilfreich, wo ich dann sehe, ok, an der Stelle ist es jetzt Berufsverkehr oder Wirtschaftsverkehr, an der Stelle ist es Freizeitverkehr..

interpretation of different graphs with context knowledge, analysis of curvature and distribution in temporal graphs

*..aus der Tagesverteilung..*

**A** Natürlich immer hauptsächlich und nicht nur, weil das ja grundsätzlich alle Verkehre sind bei Teralytics. Da haben wir dann halt noch verglichen, wie sich die unterscheiden. Der Berufsverkehr ist eigentlich immer gleich stark, da gibt es nur geringe Schwankungen, beim Freizeitverkehr natürlich mehr, weil da fährt keiner, wenn das Wetter schlecht ist, dafür fahren in den Ferien sehr viele.. Und da ist es eigentlich andersrum, da fahren am wenigsten in den Ferien. Logisch. Aber das ist dann die Interpretation der Daten. Erst mal geht es drum, Anfragen zu stellen, die zu visualisieren, und dann muss man die Ergebnisse eben interpretieren.

clear understanding of data visualization and interpretation workflow

*Ok. Ja, also das das ich gerade konzipiere, wo es da hin gehen soll, ist auch so eine Art Dashboard*

**A** Sie wollen selber eins entwickeln?

*Ja, also, je nachdem wie viel Zeit bleibt ist das dann halb Mock-up, halb funktional, aber die Visualisierungen laufen schon größtenteils.*

**B** Aber woher kriegen Sie die Daten, wenn ich das fragen darf? Oder haben Sie die schon, die Mobilfunkdaten.

*Das ist aus diesem Projekt, xMND, und eben über Fraunhofer, also die modellieren die Daten. Ein großer Punkt bei dem Projekt ist auch, dass versucht wird, eben das Transportmittel zu ermitteln. Also die haben auch GPS-Daten, die sie damit verschränken wollen im Prinzip, mit denen ich natürlich nicht arbeiten kann, weil die zu Personenbezogen sind.*

**B** Kann man, A, kann man in den Teralyticsdaten auch nach dem Verkehrsmittel unterscheiden?

**A** Ja, aber sehr eingeschränkt. Das Verkehrsmittel Flugzeug kann man gut isolieren, weil man da in einer Zelle einfach verschwindet und in einer ganz anderen Zelle wieder auftaucht und dazwischen kein Weg ist. [...] Also das funktioniert technisch natürlich wunderbar. Bahnverkehr funktioniert angeblich auch aber ich bin nicht überzeugt davon. Es heißt, ab 30km Entfernung

könne man auch.. Weil die Bahn läuft ja zum Teil auch auf anderen Trassen oder Wegen, oder es ist nicht immer neben den Gleisen auch noch eine Straße, und auch aus dem Bewegungsprofil heraus kann man halt sehen, ist das ein Fahrplan oder steht der im Stau oder bewegt der sich eben anders. Das würde angeblich funktionieren, ab 30km. Ich sag mal jaa.. eingeschränkt. Oder wir haben einfach nicht die ausreichenden Entfernungen im Verbund, auch nicht in einem erweiterten Verbund, der wäre ja schon im Durchmesser 100km oder so, also da müsste man da eigentlich Zugfahrten von Autofahrten theoretisch unterscheiden können. Aber es ist sehr eingeschränkt. Der größte Balken bei den Abfragen ist immer "keine Aussage" oder "keine Unterscheidung möglich".

very limited mode detection within the currently used products

**B** Aber zwischen zum Beispiel zu Fuß und Fahrrad und Auto kann man auch nicht unterscheiden..

**A** Technisch müsste das eigentlich gehen. Und andere Anbieter können das angeblich auch.

**B** Und sie arbeiten so mit diesen Beschleunigungsprofilen und so weiter.

**A** Wir arbeiten da auch da dran. Wir machen einen e-Tarif grad, wo ich mit Check-In, Check-Out einfach nur ins Fahrzeug gehe, und beim Aussteigen wisch ich wieder zurück, Fahrt beenden, und falls ich das vergesse, dann muss ja das System mich eigenständig ausloggen können, das funktioniert derzeit noch nicht aber da dran wird gearbeitet, dass man einfach anhand des Bewegungsprofils unterscheiden kann, ob einer mit dem Fahrrad fährt oder mit der Straßenbahn, was ja ähnlich schnell ist, sag ich mal. Also ich mein, Fußgänger kann ich immer unterscheiden, weil der sich einfach viel langsamer bewegt. Aber ein Auto das im zähfließenden Stadtverkehr ist oder auch ein Radfahrer, das könnte auch eine Straßenbahn sein, die Bewegungsprofile zu unterscheiden, daran arbeitet man halt jetzt. Jetzt muss man den Leuten sagen "Hast du ausgecheckt, Fahrt beendet?" und die müssen sich aktiv auschecken. Und in einer späteren Version soll das automatisch gehen, dass einfach das Programm erkennt, so wie der sich jetzt bewegt, oder er bewegt sich nicht mehr, das kann ja auch heißen, es fährt grad kein Zug, oder er ist schon angekommen.

experience with mode detection process within the development of an automated payment tool within a mobile phone app

*Das heißt, man zahlt dann genau die Strecke, die man..*

**A** Genau, das ist dann streckenbezogen, die Strecke wird modelliert, und dann zahlt man einen kilometerabhängigen Tarif. Das ist grad in der Erprobung, in der Pilotphase, und wird sicher in Zukunft da auch noch dahingehend weiterentwickelt, dass man eben Bewegungsprofile genauer unterscheiden kann. Da arbeitet man dann eben mit GPS-Daten oder mindestens mit der Triangulation, das ist ja hier bei unserer Betrachtungsebene nicht notwendig.

*Also ich hab mich bisher hauptsächlich auf die Stadt München konzentriert, was die Daten angeht, und es ist wohl wahrscheinlich so ähnlich wie bei den Flugreisen, also wenn jemand U-Bahn fährt, kann man das wohl auch ziemlich gut erkennen. [...] Haben*

*Sie auch Fahrgastzähldaten aus Zählgeräten?*

**B** Die haben wir noch nicht oder nur begrenzt oder? Die kriegen wir jetzt dieses Jahr..

**A** Die haben wir zum Teil auf verschiedenen Linienfahrzeugen und Betreibern, die sind aber noch in der Erprobungsphase. Also sie zählen noch nicht zuverlässig genug, dass man sie wirklich automatisch zählen lassen könnte. Die heißen zwar so aber man muss jetzt schon noch kalibrieren und überprüfen und das Finetuning eben machen. Theoretisch können sie es aber in der Praxis gibt es doch noch viele Abweichungen und da sind die Kollegen auch ziemlich dran und beschäftigt, die automatischen Zählssysteme zu kalibrieren.

counting devices are still being configured

**B** Aber es ist auch geplant, dass alle MVV-Linien in Zukunft diese automatischen Fahrgastzählssysteme haben oder?

**A** Ich glaub es nicht, ich weiß es aber nicht.

**B** Das ist bei der Verkehrsforschung und nicht bei uns und deswegen sind wir auch nicht so informiert.

**A** Ich denke die Schienenverkehrsmittel werden es alle kriegen, die Straßenbahnen wohl auch. Ob es alle Buslinien kriegen..

*Soweit ich weiß, haben die Hälfte der Stadtbusse..*

**B** Ja, die sind deutlich voran, also im Vergleich zu den Regionalbussen ist die MVG da deutlich voran. Weil die machen das auch seit Jahren, und wir haben da in den letzten Jahren erst angefangen, mit den automatischen Fahrgastzählssystemen im Regionalbus.

**A** Ja.

**B** Ja aber die MVG, soweit ich das weiß, die machen das seit einiger Zeit. Weil auch in dem Verkehrsmodell für die Kalibrierung und so, verwenden wir auch die, also innerhalb des MVG-Gebiets, verwenden wir auch die Fahrgastzahlen, die aus diesen automatischen Fahrgastzählssystemen kommen. Deswegen denke ich, dass sie das schon seit ein paar Jahren machen und nicht erst jetzt.

**A** Also was es schon lange gibt, das sind wohl, ja das ist im Grunde eine große Waage im Fahrzeug, und aus dem Durchschnittsgewicht der Personen wird dann im Grunde die Anzahl der Reisenden grob ermittelt. Das gibt es schon lange wie gesagt, aber die Zählssysteme an den Türen, am Zustieg und so, weiß ich nicht, ob das bei der MVG auch schon so lange im Einsatz ist. Aber die sind da auf jeden Fall den Anderen voraus.

rough estimations of number of passengers through weighing

*Ja das sind wohl die Daten mit denen sie momentan am meisten arbeiten... Ja gut,*

*dann glaub ich fühlt mir jetzt hier erst mal nichts mehr ein. Dann würde ich mir das gerne noch anschauen.*

A Ja natürlich. Dann schauen wir das bei mir an. [öffnet Browser] Das Ganze ist einfach ein online-Tool, das heißt man geht über das Internet rein .. und meldet sich an .. und dann muss man den Datensatz auswählen. Das wäre eben jetzt dieser Langdistanzdatensatz, der eben wie gesagt für die längeren Reisen gedacht ist und das sind die Kurzstanzreisen, die einfach anders verarbeitet werden, die haben wir lizenziert. Und dann sieht man auch gleich, welchen Bereich man lizenziert hat und kann dann hier auswählen, ob man den Rechner mehr belasten will oder weniger, also je nachdem was man als Hintergrundkarte halt einblendet. Das wären die Kartenstile .. am kräftigsten ist wohl der hier, mit der Straßenkarte im Hintergrund und das ist eben der Umgriff unserer Verbunderweiterung, den wir brauchen. Und man sieht, dass die Zellen unterschiedlich konfektioniert sind. Das ist zum Beispiel jetzt eine sehr große Zelle, jetzt hat's umgeschaltet.

*Ah ok, also wenn man näher hin zoomt, dann werden die Zellen kleiner.*

A Ja. Das sieht man vor allem in München schön.

*..und wenn man dann eine auswählt und weiter rauszoomt..*

A ..dann besteht München nur noch aus ein paar groben Einheiten und wenn ich davon eine auswähle, dann steht hier "München Bezirksteil Altmosach und 17 andere", die dann zusammengefasst sind. Wenn ich jetzt aber hergeh und sag, da muss ich die Auswahl erstmal aufheben .. Wenn ich sag "kein Clustering" .. dann wird eigentlich von Anfang an die kleinste Größe, die wir zur Verfügung haben ausgewählt und wenn ich jetzt hier einen Bezirksteil anklicke, dann steht hier nur der Bezirksteil "Schönfeld Vorstadt" und keine anderen. [klickt]

*..und wenn man einen von denen anklickt, dann werden die..*

A ..und dann hab ich hier eine Auswahl, also ich weiß jetzt ehrlich gesagt nicht, wie man mehrere Monate zusammenfassen kann. Aber hier kann man auf jeden Fall die einzelnen Monate auswählen, wir haben von Januar 2019 bis Mai 2020 mit einem kleinen Loch das wir haben von Dezember bis März, leider, wenn wir uns alle Monate anschauen, und auch miteinander vergleichen. Also hier, Mai 2020 vergleichen mit Mai 2019 zum Beispiel. Und dann schauen wir uns alle Reisen jetzt mal an von Erding nach Weilheim, das sollten eigentlich nicht so viele sein, und zwar die Reisen von - nach. Dann haben wir da .. das ist alles witzlos, weil Ausgangspunkt ist nur Erding und Ziel ist nur Weilheim, und das ist die Gesamtzahl von stolzen 1300 .. nein .. 105 Unterschied .. plus 135, das heißt das sind eigentlich nur ..

filter for months

change detection,  
compare different months,  
intuitive color-coding  
important

*..das sind mehr geworden einfach oder?*

A Nein das sind .. es ist blau also mehr Reisende im Mai 2019, das ist ein biss-

chen verwirrend, man sollte eigentlich den älteren Monat nach oben packen, damit das sie Farbgebung ist die man sich erwartet. .. Ein Rückgang durch Corona eben.

**C** Was man vielleicht noch sagen muss, aber das ist auch schon klar, dass das jetzt hochgerechnete Werte sind, also jetzt nicht .. oder? so hab ich das verstanden, das sind jetzt nicht alle Telefonica-Daten, die, was warn's 1300-irgendwas, sondern die rechnen das halt hoch auf die, ich weiß nicht, ungefähr ein Drittel der User nutzt halt Telefonica und dann wird das eben mit einem Faktor gewichtet, sodass das dann möglichst die Summe ist, die halt wirklich fährt und jetzt nicht nur die Zahl, die da erfasst wurde, sondern das wird eben dann hochgerechnet. Wie da die genauen Umrechnungsfaktoren sind weiß ich jetzt aber nicht.

**A** So, wenn man sich das jetzt anschaut, dann sieht man natürlich, dass Corona-bedingt natürlich fast in alle Richtungen die Reisen abgenommen haben. Es gibt ein paar Verbindungen, wo aber sowieso von Haus aus nicht viel los ist, wo es jetzt mehr ist als im Vorjahr. Das sind aber eigentlich die Ausreißer. Die wichtigen Verbindungen haben stark abgenommen. Der größte Pfeil ist jetzt hier, ich mein, das meiste wird ausgeblendet, damit man überhaupt noch was erkennen kann, wenn ich aber direkt draufgeh, dann stellt er mich alles dar, dann sieht die Karte so aus. Und das ist doch überwiegend rot. Die wichtigsten Beziehungen sind zum Beispiel von Erding zum Flughafen und von Erding nach München eben. Und was wir jetzt hier wollen ist zum Beispiel, dass wir München immer als Gesamtheit darstellen können, weil so wird München auf die 118 oder was es sind, Stadtbezirksteile aufgeteilt und dann teilen sich natürlich auch die Zahlen auf und die müssen wir dann hinterher wieder zusammenfassen, dass wir die gesamte Stadt abbilden können. Weil hier bei den einzelnen Stadtbezirksvierteln natürlich München dann nicht unbedingt ganz vorne auftaucht bei den Top-Zielen. Also schon einzelne, hier "Ludwigvorstadt Kliniken" .. was ist das .. ah! Da haben wir wieder eine Clusterung, die hab ich scheinbar wieder eingeschaltet, hier, wo die zusammengefasst werden. Also da besteht München eben nicht aus 111 sondern nur aus 10 oder 15 Zellen. Wenn ich die Clusterung ausschalte .. dann ist München aus den Top-Zielen ganz raus, weil jetzt teilt es sich eben auf so viele verschiedene Zellen auf, dass München dann in der Liste nicht mehr auftaucht. Und das ist zum Beispiel was, was noch geändert werden soll, an dem Produkt was wir jetzt lizenziert haben. Genau. Top-Verbindungen von-nach und dann eben ..

dynamic clustering, more control is desired

*..kann man da auch drauf klicken, auf die Topverbindungen?*

**A** Ja, wenn ich da drauf klick, dann wird die hier auch noch stärker hervorgehoben. Das sieht man jetzt an dem .. also wenn ich da drüber geh, dann seh ich's auf der Karte auch. Wenn ich jetzt die Daten weiterverarbeiten will, dann hab ich die Möglichkeit, das hier zu exportieren und kann da sagen, ich möchte es als geojson, als shapefile, oder halt nur als csv-datei. Dann kann ich's halt in der Tabellenkalkulation weiterverarbeiten.

linked views

*Ok. Das heißt, es wird auch hauptsächlich verwendet, um einen Überblick zu kriegen*

*und dann, spezielle Fälle zu erkennen.*

A Ja.. erst einmal ein Überblick und dann geh ich rein in die speziellen Fragen, die mich jetzt konkret interessieren.

overview & detail

*Mhm, also zwei spezielle Landkreise oder so, die irgendwie .. untrsucht werden sollen..*

A Ja, genau. Verkehre von-nach zum Beispiel, oder jetzt schau ich mir erst mal alles an und dann schau ich mir die stärksten Relationen noch mal genauer an, oder wie auch immer.

investigate for maxima

*..und da ist auch die Tages ..*

A Die Tageszeit? Das ist hier, also die Tagesganglinie..

*..und da kann man dann auch filtern?*

A ..da bin ich jetzt fast ein bisschen überfragt, wie man die interpretieren soll. Dass ich am Nachmittag mehr Verkehre hab wir morgens. Naja .. da überlagert sich halt Wirtschaftsverkehr, Freizeitverkehr, Versorgungsverkehr und alles und in Summe hab ich dann am Nachmittag das höchste Aufkommen.

*ja .. und wenn man das da einschränkt, dann schränkt es sich dort auch ein ..?*

A Ja. Das kann ich ja hier über den Schieberegler, da kann ich einfach bestimmte Tageszeiten ausblenden und dann wird dementsprechend auch der Zielbaum angepasst. Also natürlich die Gesamtzahl der Fahrten, das ist ja klar, aber dann auch die Ziele zum Teil. Oder ich nehm jetzt mal von der Distanz her alle Fahrten unter 10km raus. Dann schaut's auch schon wieder anders aus. Also den Nahverkehr blend ich jetzt aus, weil ich sag, ok, Nahverkehr, das ist vielleicht überwiegend Versorgungsfahrten oder so, ich möchte einfach nur den überörtlichen Verkehr, dann kann ich hier die Distanz eben noch einschränken. Ich hab die Möglichkeit, hier auch Ansichten, heißt das, zu speichern. Bestimmte Abfragen, die wir schon mal gemacht haben, die kann man dann da eben wieder herstellen, zum Beispiel die Alpenkarte, die wir gerade schon gesehen haben, kann ich hier mit einem Klick sozusagen wieder herstellen. Sonst müsste ich ja hier erstmal wieder alle Münchner Stadtbezirke auswählen und dann eben den Zielraum, den man betrachten will. Und jetzt hab ich quasi die Anfrage wieder hergestellt und kann jetzt zum Beispiel einen neuen Zeitraum nehmen. Einen anderen Monat oder eine andere Fragestellung mit den gleichen Quell- und Ziel-Zellen. Ja.. sonst müsste man es immer wieder händisch zusammen schneiden und das ist eben auch eine Funktion, die noch in Zukunft kommen wird, dass man halt mit einem Klick sagen kann "München, alle Stadtteile bitte" und nicht die einzelnen. .. [C], magst du mich ergänzen?

interactive filtering

save and restore filter options

C Du hast das vollumfassend erklärt, ich glaub da wär der X fast noch .. der hat schon mehr mit Teralytics gemacht wie ich aber ich glaub so das wichtigste hast du jetzt alles erwähnt, mir fällt jetzt da nichts mehr ein.

A Also wenn Sie da einen Screenshot haben wollen dann .. ist der schnell gemacht.

*Sehr gerne.*

## Structured Collection of Statements from the Interviews

In the following, results from the interviews are listed. The list is structured by topics suggested in the interview guide. I, II and III indicate the interviews, where the respective points were mentioned.

### Aim of planning

- I, II Provide correct amount and size of vehicles at the right times
- I, II, III Provide (new) connections that are useful for the citizens (I,II: especially flexible with buses, less flexible with rail-dependent traffic like trams or underground)
- II, III Determine how and where to densify journeys temporally or increase the network spatially
- I, II, III Attract people who are currently not using public transportation
- I, II Fulfil the “Nahverkehrsplan”, a document that legally states the desired levels of accessibility and availability of public transport for different areas
- I, II Provide a comfortable and sustainable infrastructure for all citizens in the region
- III Conceptual planning: determine sensibility of expansion of the traffic association network area. First step: analyse current traffic relations and demands between different municipalities (DE: Kommunen, Bezirke)

### Planning schedule

- I, II Yearly plans done in autumn for the year after next
- I, II Construction site planning, sometimes at short-notice in case of urgent repair work etc., usually some months ahead
- III Project-related planning schedule

### Data sources that have been / are being used:

- I, II Passenger counts acquired with technical instruments above the doors of vehicles: 50% of buses, 30% of trams are equipped. About 95% of the used data used stems from these instruments.
- I, II Manual passenger counts
- I, II Manual survey data on “Reiseketten”: origin-destination information for a small number of passengers, including supplementary information on age, gender, etc.
- I, II Mobile network data
- I, II Movement data from different apps
- I, II Movement data from smartphones (contacts are starting to be established)
- I, II Some disappointing experience has been gained with working with mobile network data. Data accuracy of 80% was insufficient (“only 80% of spectators at a football game could be found in the data, even after extrapolation”) and strange origin of a big number of trajectories led to distrust in the data.
- III Analytical data and forecast data
- III Number of future residents and workplaces
- III Transportation usage information



- III Mobile network data: data has been deemed representative and of good quality after some comparisons with other data sources, e.g. from statistical offices.
- III Incentives to make use of mobile network data are its timeliness (compared to manually gathered and processed data and official statistics) and its completeness in representation (compared to workplace statistics)
- III Different completeness characteristics of different data sources is being taken into account
- III Limitation of representativeness of official workplace statistics
- III Data from counting devices is still rare, devices are still being configured
- III A rough estimation of the number of passengers is gained through weighing of full vehicles

### Missing information

- I, II Origin-destination information with a sufficient coverage: the technical devices count only entering and exiting passengers but do not give information on from and where to the passengers are travelling
- I, II More (dense) data for special days, e.g. football matches, Oktoberfest, fairs, etc.
- III Information on how people travel, i.e. which means of transport they are using is very limited in the currently used dashboard solution

### Analytical Questions

- I, II, III Generally: How well does the provided network and journey density match the demand and where does it need adjustment / improvement?
- I, II, III From where to where do people (regularly) travel?
- II, III How do different times of the day / weekdays / months compare?
- I, II At what times during a day are the maximum amount of people travelling in a certain segment of the network on a generic day, e.g. Mon-Thu? How many people are travelling then?
- I, II From where to where do how many people travel that go to a special event (fair, festival, football game)? When do they travel?
- I, II Where do people change lines / mode of transport?
- I, II From where to where do people travel that use the subway?
- III When do how many people travel throughout a day's cycle between specific areas?
- III Which spatial connections are more travelled on than others?
- III What is the purpose of travel (commute / leisure)? (Currently, this is derived from shape of the daily hydrograph. Specific temporal filtering, e.g. omitting rush hours, is being used to assume purpose of transport.)

### Questions that cannot be answered within the current framework and would be of interest

- I, II Origin-Destination information. The current framework offers only sparse data of this kind which is collected manually by interviewing passengers.
- I, II Movement data and patterns, even of rough spatial resolution, for re-occurring but exceptional days such as big football events, festivals and fairs. Thin data due to irregularity of such days.
- I, II Movement data and patterns for people using motorised individual traffic, other trains or

the bicycle. To attract new customers but also to plan for a more sustainable and inclusive future

- I, II More information on underground-traffic, since those trains are not yet well equipped with counting devices.
- I, II Travel-time in combination with information on whether or not the person was using public transportation would be useful.
- III Mode of transportation detection is very limited within the currently used dashboard. There is some experience with the process of mode detection in conjunction with the development of an automated payment toll within a mobile phone application.

## Interfaces

- I, II transport model (in VISUM by ptv) combining assumptions with measurement data
- I, II Database containing passenger counts
- I, II Colour-coded table (“Belastungsteppich”, see fig.4.2) showing the counts of passengers between stations of one specific line. Based on DB, one column per trip (temporal resolution along the horizontal axis varies depending on specific trip interval)
- I, II Course of a day for connection between two subsequent stations (“Tagesganglinie”, see fig.4.3) showing the load of passengers. Irregular aggregation (sometimes two trips summed in one bar, sometimes missing bars because of lower trip frequency). Comparison to maximum load suggested in the “Nahverkehrsplan”, taking into account the vehicles’ passenger space. Mouse-over shows numbers / legend for each symbol. Interactivity by changing entries in further table sheets and switching back to the graphic.
  - I Browser-based visualizations were deemed too difficult to interact with. “Everyone can use Excel”, it is “easy to play with the data in Excel”.
- I, II GIS visualizations done with QGIS including a QGIS-Server for Web-GIS for special evaluations or analyses for a certain topic, e.g. geomarketing, information on construction site plans, maps at the stations. Not as a standard interface.
  - II A Web-GIS(?)-Application showing the route network as lines, colored (yellow-red) according to passenger counts. Was implemented due to the corona-pandemic. Aggregated w.r.t. days / morning hours / evening hours. Highly appreciated by the planner. Connects load information to location on a map and hence gives better overview.
- III Two VISUM traffic models, one for the region of Munich and one for Bavaria that are shared among different stakeholders. Used to test possible new connections and get insight into expected demands.
- III Manually processed and visualized workplace statistics for investigation of specific spatial relations in static maps for reports. See figures A.3 and A.4
  - Thickness of lines between municipalities indicate number of commuters
  - color-coding of quantitative balance.
  - To reduce cluttering and visual overload, multiple maps with visual elements of reduced complexity are generated
- III A dashboard tool by Teralytics, displaying movements extracted from mobile network data. See figures A.5, A.6 and A.7.
  - The method of working with a visual interface as a first introduction to the data is new but seems to be appreciated. A clear understanding of data visualization and

interpretation workflow is shown.

- number of travellers represented by thickness of arrow-line on a map
- Map and different supplementary charts are linked, filtering impacts all graphs. Ordered bar charts show extremal data points
- Data exploration process follows overview-and-detail sequence and is supported by the tool
- Spatial filtering for region of origin and/or destination
- Dynamic spatial aggregation, depending on zoom-level
- Temporal filtering for days of the week
- Temporal filtering for month and year
- Comparison of different months
- Standard parameters are day of the week, hour of the day and travelled distance
- Visual identification of extremal data points via ordered bar charts
- Investigation of maxima through interactive filtering
- supplementary graphs are interpreted with context knowledge, analysis of curvature and distribution in temporal graphs
- Change detection is color-coded. Comparing different months shows arrows that represent the difference between the different time spans and are coloured red or blue.
- Filter combination or views can be saved and returned to.
- export functionality of filtered data to different file types for further use, e.g. spreadsheet calculation

### Temporal resolutions

I, II course within one day of main interest

I, II types of days:

- Monday-Thursday-School (main type because highest passenger load)
- Monday-Thursday-Holidays
- Friday
- Saturday
- Sunday/Holiday

I, II Data for each trip, usually every 10 minutes ("fahrtscharf") within the model

I, II Temporal resolution should not be less than half-hourly, to be able to observe peaks, e.g. after football games

I, II Aggregated data for different day-types, within several weeks or months, usually between school holidays. Winter months (November - February) show highest load.

III Hourly aggregation of data

### Spatial resolutions

I, II Every station is registered in the model

I, II Data for each trip between two stations

- I, II Possible to work with slightly bigger traffic cells (“Verkehrszellen”) which divide the city into 1000 entities and the surrounding region into another 1000 entities. Each cell can contain two to three bus-stops but no more than one underground-stop.
- III Municipalities and districts as smallest entities. Rather coarse but sufficient and adequate for current objective
- III More control over default spatial aggregation is being asked for, e.g. the city of Munich should always be treated as a whole instead of a number of individual districts (DE: Bezirke).

## Screenshots from Interviews

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	<b>Linie 55, Putzbrunn, M.-Haslbeck-Straße - Ostbahnhof, Richtung: Rück, Last</b>																	
2	Stand 2019-09-25																	
3																		
4	Datenquelle: ZDMS, AggregationsID A 07 190415 190728 BT alle, 19.04.2019 - 28.07.2019																	
5																		
6	Bearbeiter: obermaier.sebastia																	
	<b>Last Mittelwert</b>																	
7		<b>Erste Abfahrtszeit</b>	05:03:30	05:23:30	05:43:00	06:03:00	06:23:00	06:43:00	07:03:00	07:23:00	07:43:00	08:03:00	08:23:00	08:34:30	09:03:00	09:23:00	09:34:30	09:55:30
8		<b>Linie</b>	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
9		<b>Kurs</b>	4	6	8	4	6	8	4	6	8	4	5	6	8	4	5	6
10	Putzbrunn, M.-Haslbeck-Straße													2			2	
11	Putzbrunn, W.-v.-Braun-Straße													2			3	
12	Putzbrunn, Kirchenzentrum													3			4	
13	Putzbrunn, Brennerei													5			4	
14	Putzbrunn, Waldkolonie													5			4	
15	Putzbrunn, Seemannsiedlung													6			5	
16	Am Hain													6			6	
17	Waldheimplatz													6			6	
18	Waldperlach																	
19	Waldperlach		1	1	1	2	3	1	2	5	2	2	2	7	7	5	7	4
20	Waldheimplatz		2	1	2	3	6	3	4	9	5	4	5	10	10	9	10	4
21	Eulenspiegelstraße		2	1	3	3	7	3	4	9	6	5	6	11	12	11	12	4
22	Im Gefilde		3	2	5	4	8	3	5	15	8	7	8	14	14	12	13	4
23	Tribulaunstraße		4	6	6	5	8	4	6	15	8	8	9	14	14	13	14	4
24	Klinikum Neuperlach		11	17	10	10	10	8	8	18	11	12	14	18	20	18	20	4
25	Sudermannallee		14	18	12	11	11	9	9	19	13	13	18	21	21	21	23	4
26	Kafkastraße		14	20	13	12	11	11	11	19	14	14	19	23	22	23	25	4
27	Neuperlach Zentrum		4	6	5	4	3	5	3	7	4	4	7	7	7	11	10	4
28	Thomas-Dehler-Straße		5	6	5	4	4	6	4	7	4	5	7	7	8	12	11	4
29	Friedhof Perlach		5	7	6	4	4	6	4	7	4	5	7	7	7	11	11	4
30	Wilhelm-Hoegner-Straße		5	7	6	5	3	5	4	7	4	5	7	8	7	12	11	4
31	Pfanzeltplatz		8	9	7	7	5	6	7	8	7	8	8	10	11	17	14	7
32	Weddigenstraße		11	13	9	9	11	9	10	14	13	13	13	17	18	24	21	12
33	Adilostraße		12	13	10	10	11	10	10	14	14	13	14	20	19	27	22	15
34	Woferlstraße		13	13	11	11	14	11	12	15	18	17	20	25	26	32	28	18
35	Diakon-Kerolt-Weg		15	13	11	13	14	13	16	17	20	21	25	32	34	38	40	27
36	Ramersdorf		16	12	11	13	15	12	16	17	20	21	25	32	38	41	43	26
37	Karl-Preis-Platz		12	9	8	11	13	10	13	17	16	18	21	28	31	36	36	25
38	Anzinger Straße		10	9	9	12	15	10	13	17	16	19	20	27	30	34	37	29
39	Orleansstraße		9	9	8	11	15	10	12	17	16	18	20	27	29	33	34	28
40	Ostbahnhof		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure A.1: Load carpet

Tagesganglinie Linie 18, Querschnitt Siglstraße - Lautensackstraße, Auswertung der Kennzahl Last  
 AggregationsID A 01 190415 190728 BT alle, Fahrten wurden der Soll-Abfahrtszeit zugeordnet

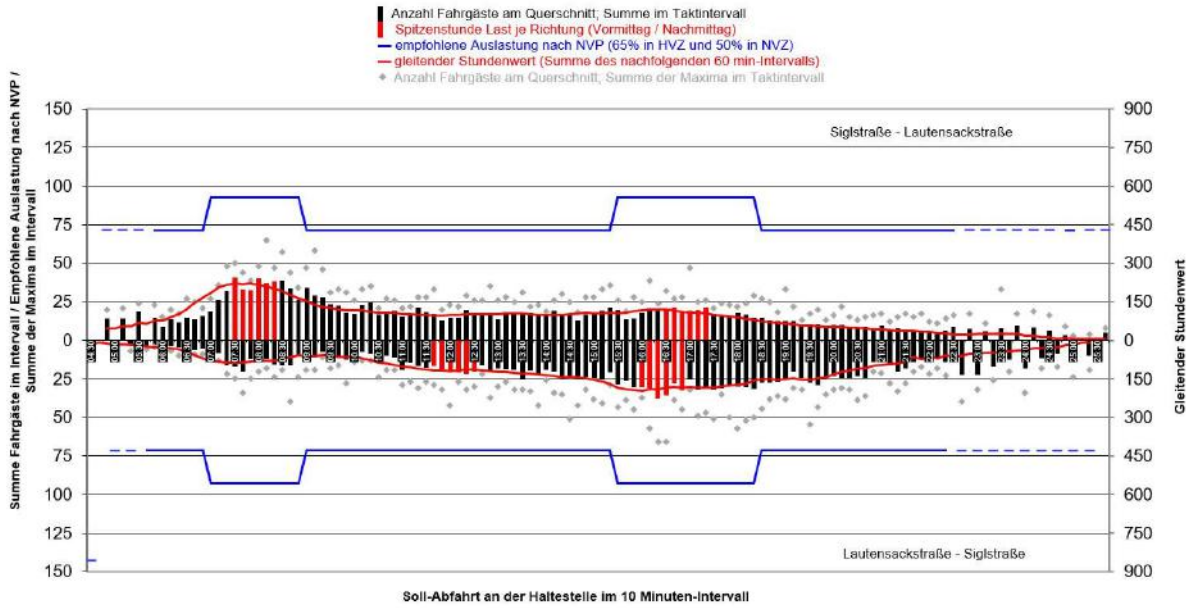


Figure A.2: Daily graph

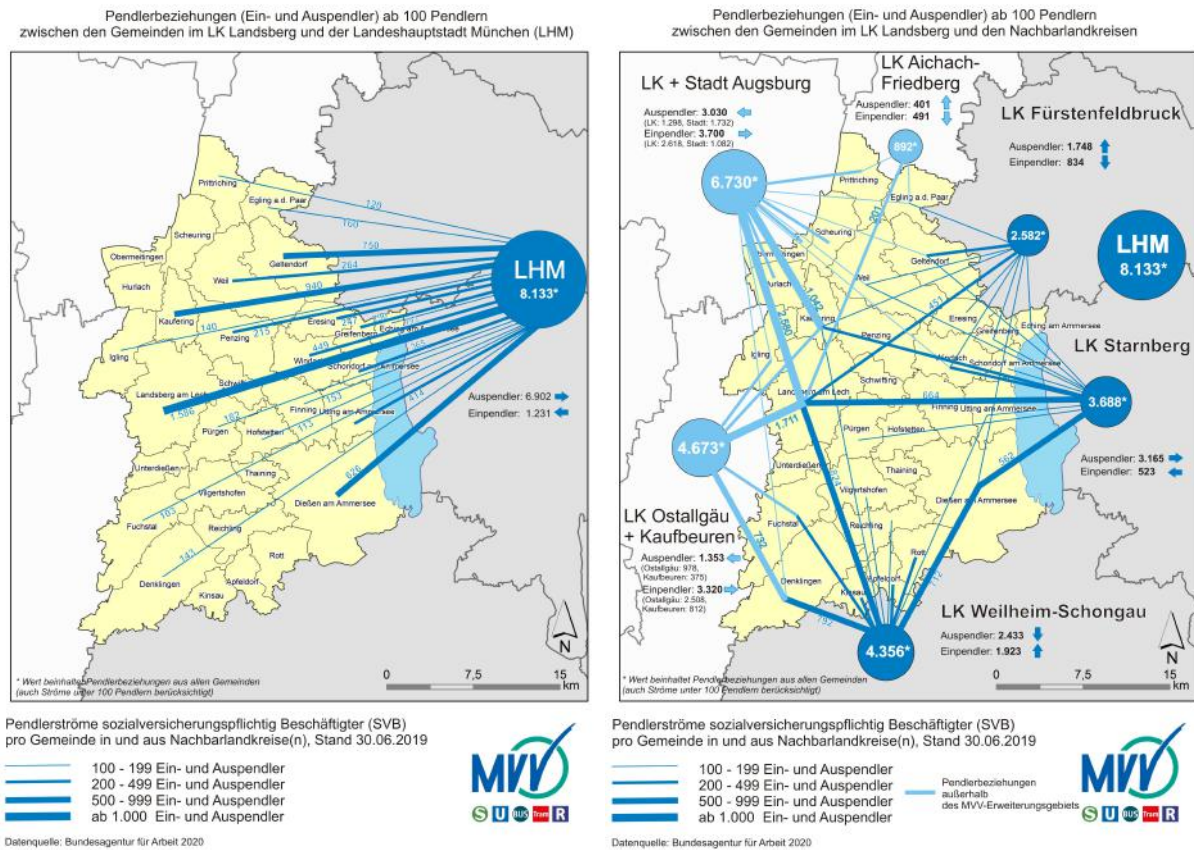
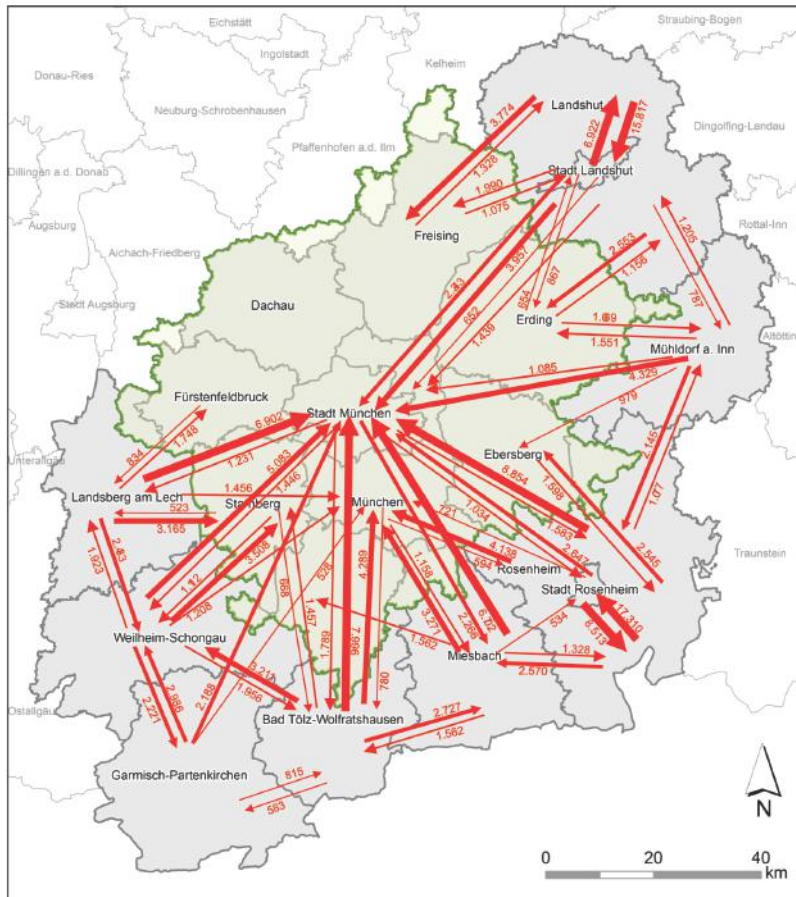


Figure A.3: Commuter flows Landsberg

Pendlerbeziehungen ab 500 Pendlern  
zwischen den Verbunderweiterungsländkreisen und dem MVV-Verbundraum



Pendlerbeziehungen (Auspendler) sozialversicherungspflichtig Beschäftigter (SVB) zwischen den Landkreisen,  
Stand: 30.06.2019

- von 500 bis 1.000
- über 1.000 bis 2.000
- über 2.000 bis 3.000
- über 3.000 bis 6.000
- über 6.000



Datenquelle: Bundesagentur für Arbeit 2020

Figure A.4: Commuter flows between districts

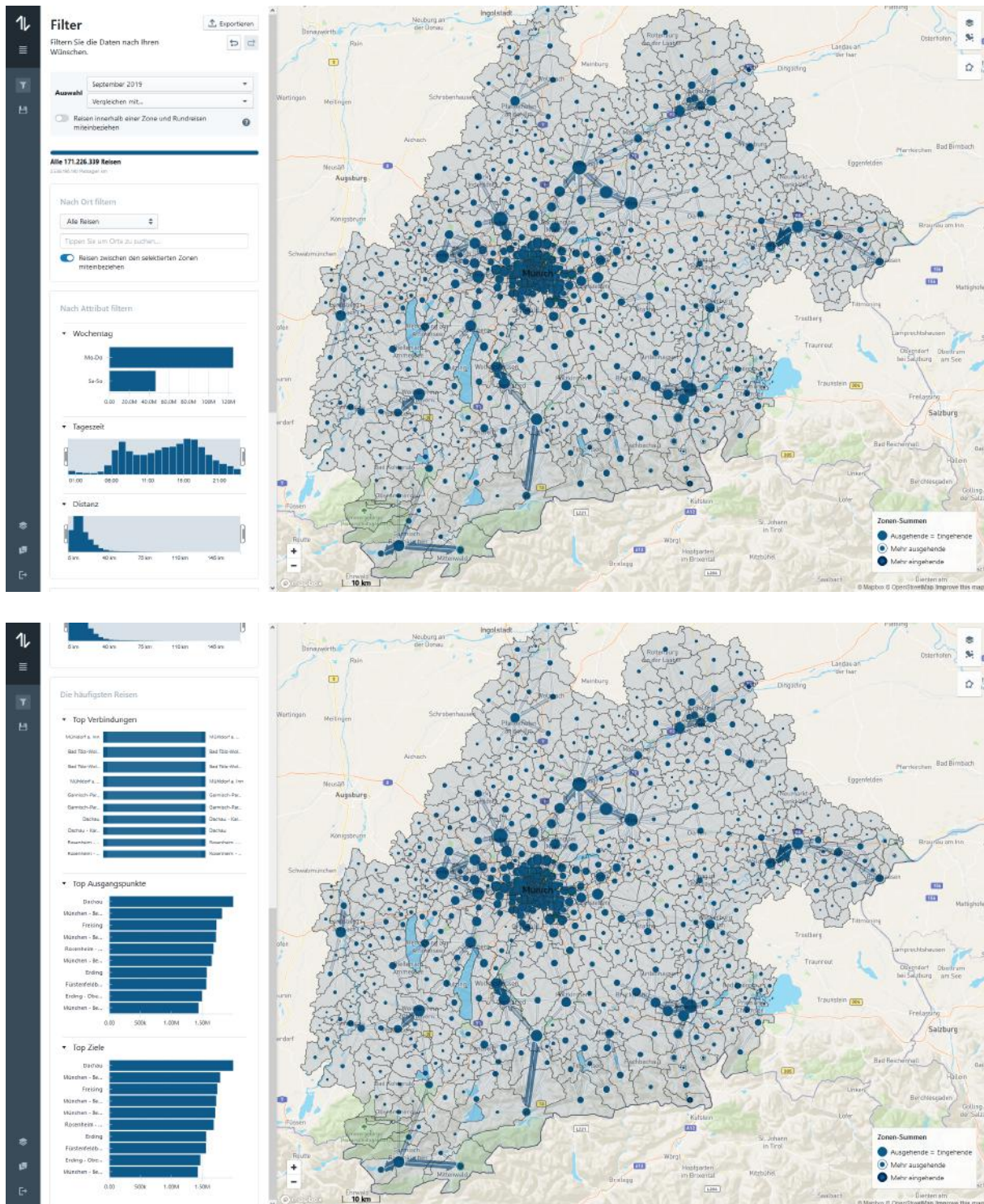


Figure A.5: Dashboard interface by Teralytics

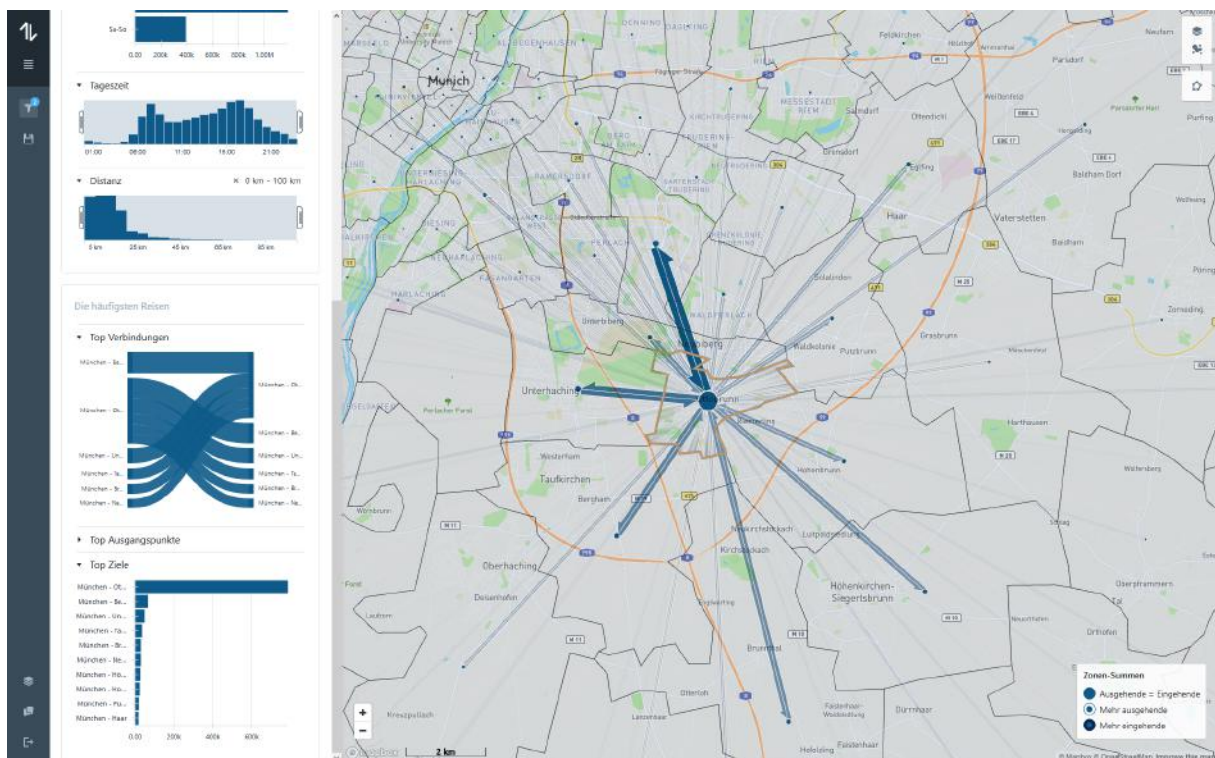
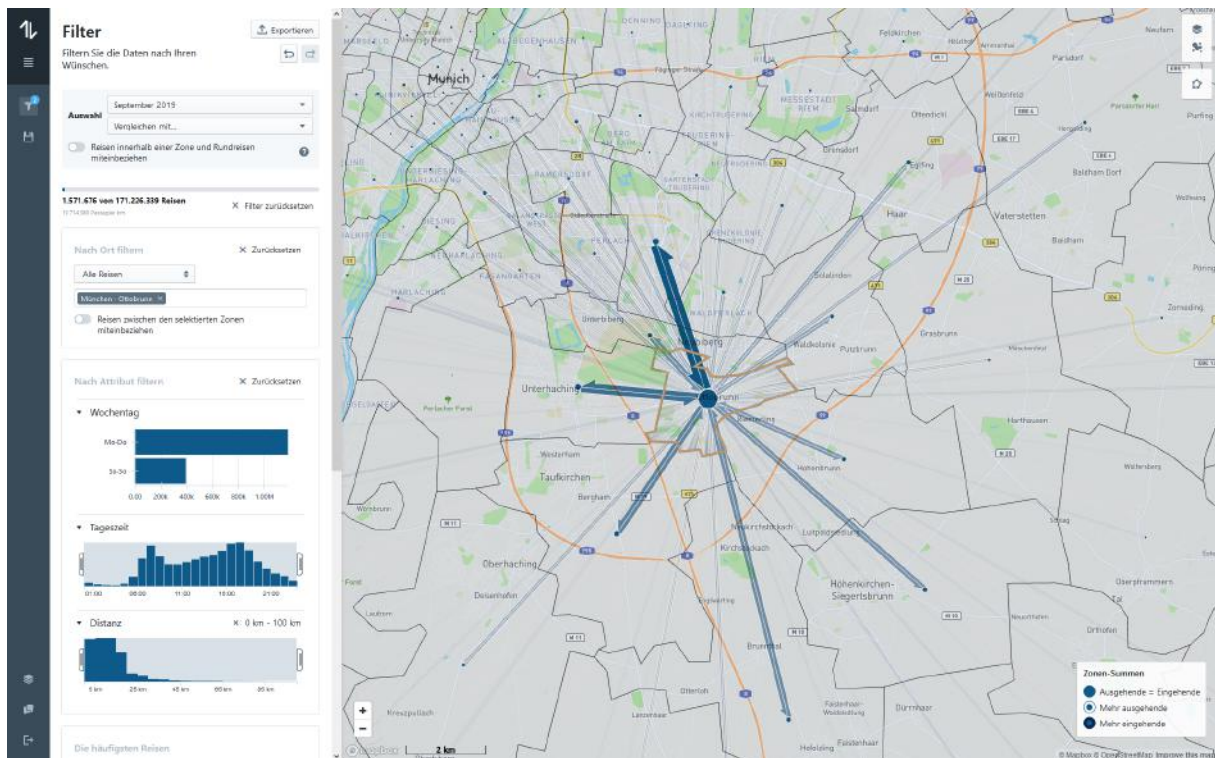


Figure A.6: Dashboard interface by Teralytics



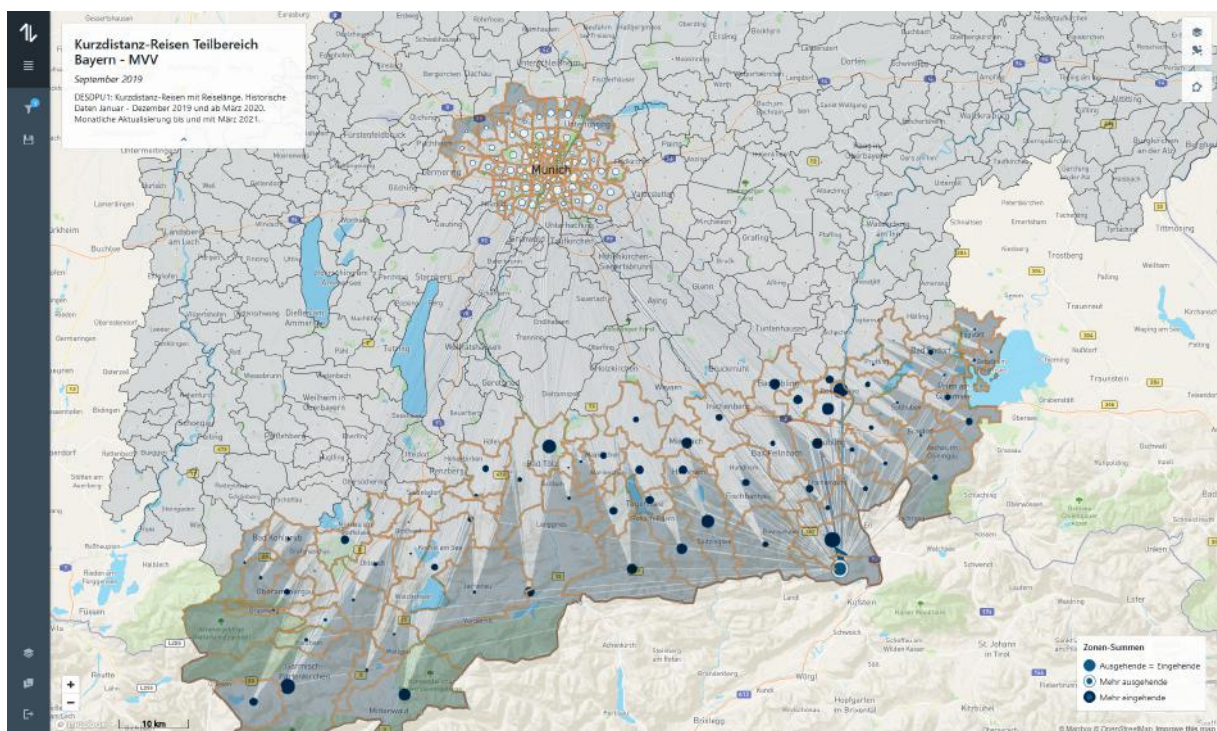


Figure A.7: Dashboard interface by Teralytics

## Data Tables

Unnamed: 0	flow_direction	from_cell	to_cell	privat	Mode::Bus	Rail	Mode::Subway	Mode::Tram	moves	geometry_from	centroid_from	geometry_to	centroid_to	start	end
89	51->12	51	12	30.0	5.0	3.0	45.0	0.0	83	POLYGON ((11.551072 48.14077400000001, 11.5510... 11.5510... 48.14077400000001, 11.5510...))	POINT (11.55698349465301 48.1400832102274)	POLYGON ((11.5278 48.11866400000002, 11.526281... 11.526281... 48.11866400000002, 11.5278...))	POINT (11.51948297268458 48.12314298472728)	1.552648e+09	1.552651e+09
178	51->127	51	127	2.0	0.0	21.0	9.0	0.0	32	POLYGON ((11.551072 48.14077400000001, 11.5510... 11.5510... 48.14077400000001, 11.5510...))	POINT (11.55698349465301 48.1400832102274)	POLYGON ((11.484245 48.13346999999998, 11.4815... 11.4815... 48.13346999999998, 11.484245...))	POINT (11.48918562422034 48.14301958141964)	1.552648e+09	1.552651e+09
235	51->138	51	138	12.0	1.0	2.0	34.0	0.0	49	POLYGON ((11.551072 48.14077400000001, 11.5510... 11.5510... 48.14077400000001, 11.5510...))	POINT (11.55698349465301 48.1400832102274)	POLYGON ((11.550843 48.19328999999998, 11.5486... 11.5486... 48.19328999999998, 11.550843...))	POINT (11.54324825280848 48.1874409720002)	1.552648e+09	1.552651e+09
268	51->151	51	151	7.0	0.0	0.0	11.0	0.0	18	POLYGON ((11.551072 48.14077400000001, 11.5510... 11.5510... 48.14077400000001, 11.5510...))	POINT (11.55698349465301 48.1400832102274)	POLYGON ((11.517623 48.153397, 11.513544 48.16... 11.513544 48.16... 48.153397, 11.517623...))	POINT (11.5209077502935 48.16024859038502)	1.552648e+09	1.552651e+09

Table A.1: Preprocessed flow data.



flowID	flowCoords	→									
		hour	moves	privat	public	Rail	UBahn	Tram	Bus	Tram	Bus
100_104	[[48.22013108151531, 11.52767333862167], [48.2...	4	15.0	13.0	2.0	1.0	0.0	0.0	0.0	1.0	
100_104	[[48.22013108151531, 11.52767333862167], [48.2...	5	54.0	54.0	0.0	0.0	0.0	0.0	0.0	0.0	
100_104	[[48.22013108151531, 11.52767333862167], [48.2...	6	53.0	52.0	1.0	1.0	0.0	0.0	0.0	0.0	
100_104	[[48.22013108151531, 11.52767333862167], [48.2...	7	36.0	36.0	0.0	0.0	0.0	0.0	0.0	0.0	

movesMDD	privatMDD	publicMDD	RailMDD	UBahnMDD	TramMDD	BusMDD	movesAnom	privatAnom	publicAnom	RailAnom	UBahnAnom	TramAnom	BusAnom
17.333333	17.333333	0.000000	0.000000	0.0	0.0	0.0	-2.333333	-4.333333	2.000000	1.000000	0.0	0.0	1.0
37.666667	37.000000	0.666667	0.666667	0.0	0.0	0.0	16.333333	17.000000	-0.666667	-0.666667	0.0	0.0	0.0
44.333333	44.000000	0.333333	0.333333	0.0	0.0	0.0	8.666667	8.000000	0.666667	0.666667	0.0	0.0	0.0
52.666667	52.666667	0.000000	0.000000	0.0	0.0	0.0	-16.666667	-16.666667	0.000000	0.000000	0.0	0.0	0.0

Table A.5: Aggregated flow anomalies for plotting on maps.

flowID	flowIDDir	from_cell	to_cell	dayInt	weekday	dayType	hour	moves	privat	public	Rail	UBahn	Tram	Bus
100_138	1	100	138	3	Thursday	MonThu	13	16	14.0	2.0	2.0	0.0	0.0	0.0
100_139	1	100	139	4	Friday	Fri	15	15	10.0	5.0	4.0	1.0	0.0	0.0
100_142	-1	142	100	3	Thursday	MonThu	8	16	16.0	0.0	0.0	0.0	0.0	0.0
100_158	1	100	158	5	Saturday	Weekend	15	15	15.0	0.0	0.0	0.0	0.0	0.0

Table A.6: Aggregated flows for plotting on bar charts.

flowID	flowIDDir	from_cell	to_cell	hour	moves	privat	public	Rail	UBahn	Tram	Bus
100_104	-1	104	100	5	15.0	15.0	0.0	0.0	0.0	0.0	0.0
100_104	-1	104	100	6	25.0	25.0	0.0	0.0	0.0	0.0	0.0
100_104	-1	104	100	10	17.0	17.0	0.0	0.0	0.0	0.0	0.0
100_104	-1	104	100	13	20.0	20.0	0.0	0.0	0.0	0.0	0.0

movesMDD	privatMDD	publicMDD	RailMDD	UBahnMDD	TramMDD	BusMDD	movesAnom	privatAnom	publicAnom	RailAnom	UBahnAnom	TramAnom	BusAnom
15.000000	15.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
20.333333	20.333333	0.0	0.0	0.0	0.0	0.0	4.666667	4.666667	0.0	0.0	0.0	0.0	0.0
15.333333	15.333333	0.0	0.0	0.0	0.0	0.0	1.666667	1.666667	0.0	0.0	0.0	0.0	0.0
19.666667	19.666667	0.0	0.0	0.0	0.0	0.0	0.333333	0.333333	0.0	0.0	0.0	0.0	0.0

Table A.7: Aggregated flow anomalies for plotting on bar charts.

first_seen	avg_lon	avg_lat	dist	time	trip_id	trip_length	point_type	prediction_majority
1552559042	11.415	48.108	21.894	1442	1	11	start	Rail
1552561168	11.610	48.128	21.894	1442	1	11	end	Rail
1552557895	11.489	48.122	0.587	712	2	6	start	privat
1552558620	11.484	48.123	0.587	712	2	6	end	privat
1552558327	11.581	48.159	15.505	991	3	7	start	privat
1552560267	11.675	48.264	15.505	991	3	7	end	privat

Table A.8: Preprocessed OD data.

startCellID	endCellID	ODCoords	dayInt	hour	weekday	dayType	mode_prediction_majority	N	dist	cellsN	traveltime
6	6	[[[48.13924065791094, 11.56371285141089], [4...	6	0	Sunday	Weekend	UBahn	24	0.068708	2.541667	371.708333
6	6	[[[48.13924065791094, 11.56371285141089], [4...	6	0	Sunday	Weekend	Rail	52	0.056827	2.634615	316.211538
6	6	[[[48.13924065791094, 11.56371285141089], [4...	6	0	Sunday	Weekend	privat	23	0.164391	2.652174	329.304348
7	7	[[[48.13734108741345, 11.5748150980363], [48...	6	0	Sunday	Weekend	privat	49	0.121653	3.326531	654.632653
8	8	[[[48.17678500109166, 11.53594499206671], [4...	6	0	Sunday	Weekend	privat	43	0.123744	2.813953	398.976744
10	10	[[[48.13784596493853, 11.63653054310525], [4...	6	0	Sunday	Weekend	privat	23	0.140696	2.956522	448.695652

Table A.9: All OD data merged with cell polygons in one data frame.

startCellID	endCellID	ODCoords	dayInt	hour	weekday	dayType	Bus	Rail	Tram	UBahn	privat	cellsN_mean_Bus																																																																	
1	1	[[[48.15396752884293, 11.51032914199828], [4...	0	0	Monday	MonThu	2.0	0.0	0.0	0.0	14.0	3.5																																																																	
1	1	[[[48.15396752884293, 11.51032914199828], [4...	0	2	Monday	MonThu	0.0	0.0	0.0	0.0	31.0	0.0																																																																	
1	1	[[[48.15396752884293, 11.51032914199828], [4...	0	4	Monday	MonThu	0.0	0.0	0.0	0.0	27.0	0.0																																																																	
1	1	[[[48.15396752884293, 11.51032914199828], [4...	0	5	Monday	MonThu	0.0	0.0	0.0	0.0	54.0	0.0																																																																	
<table border="1"> <thead> <tr> <th>cellsN_mean_Rail</th> <th>cellsN_mean_Train</th> <th>cellsN_mean_UBahn</th> <th>cellsN_mean_privat</th> <th>dist_mean_Bus</th> <th>dist_mean_Rail</th> <th>dist_mean_Train</th> <th>dist_mean_UBahn</th> <th>dist_mean_privat</th> <th>traveltime_mean_Bus</th> <th>traveltime_mean_Train</th> <th>traveltime_mean_UBahn</th> <th>traveltime_mean_privat</th> </tr> </thead> <tbody> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>3.357143</td> <td>0.263</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.067643</td> <td>1005.0</td> <td>0.0</td> <td>0.128065</td> </tr> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>2.838710</td> <td>0.000</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.192296</td> <td>0.0</td> <td>0.0</td> <td>0.217852</td> </tr> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>3.629630</td> <td>0.000</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> </tr> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>2.592593</td> <td>0.000</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>0.0</td> </tr> </tbody> </table>													cellsN_mean_Rail	cellsN_mean_Train	cellsN_mean_UBahn	cellsN_mean_privat	dist_mean_Bus	dist_mean_Rail	dist_mean_Train	dist_mean_UBahn	dist_mean_privat	traveltime_mean_Bus	traveltime_mean_Train	traveltime_mean_UBahn	traveltime_mean_privat	0.0	0.0	0.0	3.357143	0.263	0.0	0.0	0.0	0.0	0.067643	1005.0	0.0	0.128065	0.0	0.0	0.0	2.838710	0.000	0.0	0.0	0.0	0.0	0.192296	0.0	0.0	0.217852	0.0	0.0	0.0	3.629630	0.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.592593	0.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
cellsN_mean_Rail	cellsN_mean_Train	cellsN_mean_UBahn	cellsN_mean_privat	dist_mean_Bus	dist_mean_Rail	dist_mean_Train	dist_mean_UBahn	dist_mean_privat	traveltime_mean_Bus	traveltime_mean_Train	traveltime_mean_UBahn	traveltime_mean_privat																																																																	
0.0	0.0	0.0	3.357143	0.263	0.0	0.0	0.0	0.0	0.067643	1005.0	0.0	0.128065																																																																	
0.0	0.0	0.0	2.838710	0.000	0.0	0.0	0.0	0.0	0.192296	0.0	0.0	0.217852																																																																	
0.0	0.0	0.0	3.629630	0.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0																																																																	
0.0	0.0	0.0	2.592593	0.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0																																																																	
<table border="1"> <thead> <tr> <th>traveltime_mean_Rail</th> <th>traveltime_mean_Train</th> <th>traveltime_mean_UBahn</th> <th>traveltime_mean_privat</th> <th>public</th> <th>moves</th> <th>traveltime_mean_public</th> <th>dist_mean_public</th> <th>traveltime_mean_moves</th> <th>dist_mean_moves</th> </tr> </thead> <tbody> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>656.071429</td> <td>2.0</td> <td>16.0</td> <td>251.25</td> <td>0.06575</td> <td>453.660714</td> <td>0.066696</td> </tr> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>440.225806</td> <td>0.0</td> <td>31.0</td> <td>0.00000</td> <td>0.00000</td> <td>220.112903</td> <td>0.064032</td> </tr> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>598.962963</td> <td>0.0</td> <td>27.0</td> <td>0.00000</td> <td>0.00000</td> <td>299.481481</td> <td>0.096148</td> </tr> <tr> <td>0.0</td> <td>0.0</td> <td>0.0</td> <td>326.629630</td> <td>0.0</td> <td>54.0</td> <td>0.00000</td> <td>0.00000</td> <td>163.314815</td> <td>0.108926</td> </tr> </tbody> </table>													traveltime_mean_Rail	traveltime_mean_Train	traveltime_mean_UBahn	traveltime_mean_privat	public	moves	traveltime_mean_public	dist_mean_public	traveltime_mean_moves	dist_mean_moves	0.0	0.0	0.0	656.071429	2.0	16.0	251.25	0.06575	453.660714	0.066696	0.0	0.0	0.0	440.225806	0.0	31.0	0.00000	0.00000	220.112903	0.064032	0.0	0.0	0.0	598.962963	0.0	27.0	0.00000	0.00000	299.481481	0.096148	0.0	0.0	0.0	326.629630	0.0	54.0	0.00000	0.00000	163.314815	0.108926															
traveltime_mean_Rail	traveltime_mean_Train	traveltime_mean_UBahn	traveltime_mean_privat	public	moves	traveltime_mean_public	dist_mean_public	traveltime_mean_moves	dist_mean_moves																																																																				
0.0	0.0	0.0	656.071429	2.0	16.0	251.25	0.06575	453.660714	0.066696																																																																				
0.0	0.0	0.0	440.225806	0.0	31.0	0.00000	0.00000	220.112903	0.064032																																																																				
0.0	0.0	0.0	598.962963	0.0	27.0	0.00000	0.00000	299.481481	0.096148																																																																				
0.0	0.0	0.0	326.629630	0.0	54.0	0.00000	0.00000	163.314815	0.108926																																																																				

Table A.10: All OD flows for plotting on maps and bar charts.

startCellID	endCellID	ODCoords	hour	moves	privat	public	Rail	UBahn	Tram	Bus
1	1	[[[48.15396752884293, 11.51032914199828], [4...	4	19.0	17.0	2.0	1.0	0.0	0.0	1.0
1	1	[[[48.15396752884293, 11.51032914199828], [4...	5	46.0	45.0	1.0	0.0	0.0	1.0	0.0
1	1	[[[48.15396752884293, 11.51032914199828], [4...	6	115.0	113.0	2.0	1.0	0.0	0.0	1.0
1	1	[[[48.15396752884293, 11.51032914199828], [4...	7	175.0	167.0	8.0	0.0	0.0	1.0	7.0

moves_MDD	privat_MDD	public_MDD	Rail_MDD	UBahn_MDD	Tram_MDD	Bus_MDD
20.333333	20.000000	0.333333	0.000000	0.0	0.000000	0.333333
44.000000	43.666667	0.333333	0.333333	0.0	0.000000	0.000000
158.000000	155.000000	3.000000	0.000000	0.0	0.000000	3.000000
199.333333	194.333333	5.000000	0.666667	0.0	0.666667	3.666667

dist_mean_moves	dist_mean_privat	dist_mean_public	dist_mean_Rail	dist_mean_UBahn	dist_mean_Train	dist_mean_Bus
0.105169	0.144588	0.065750	0.0	0.0	0.0	0.263000
0.091733	0.183467	0.000000	0.0	0.0	0.0	0.000000
0.165316	0.232133	0.098500	0.0	0.0	0.0	0.394000
0.175835	0.289599	0.062071	0.0	0.0	0.0	0.248286

traveltime_mean_moves	traveltime_mean_privat	traveltime_mean_public	traveltime_mean_Rail	traveltime_mean_UBahn	traveltime_mean_Train	traveltime_mean_Bus
431.073529	519.647059	342.500000	1112.0	0.0	0.0	258.000000
185.919444	273.088889	98.750000	0.0	0.0	395.0	0.000000
304.515487	252.530973	356.500000	157.0	0.0	0.0	1269.000000
363.873824	343.604790	384.142857	0.0	0.0	1373.0	163.571429

movesAnom	privatAnom	publicAnom	RailAnom	UBahnAnom	TramAnom	BusAnom
-1.333333	-3.000000	1.666667	1.000000	0.0	0.000000	0.666667
2.000000	1.333333	0.666667	-0.333333	0.0	1.000000	0.000000
-43.000000	-42.000000	-1.000000	1.000000	0.0	0.000000	-2.000000
-24.333333	-27.333333	3.000000	-0.666667	0.0	0.333333	3.333333

Table A.11: All OD anomalies for plotting on maps and bar charts.

# User Test Transcripts

## User Test I

Stage 1

*Also ich habe verschiedene Visualisierungen vorbereitet, es gibt immer eine Kartenansicht und eine andere Ansicht. Wir gehen da in vier Schritten durch und das Szenario sozusagen ist, dass wir uns erst die Daten an sich anschauen und ein bisschen die Art der Visualisierung kennenlernen und dann auf eine bestimmte Anomalie eingehen und die genauer untersuchen. Und jetzt geht's am Anfang ... werden Sie so eine Kartenansicht sehen mit verschiedenen Filtern, da gibt es einmal einen Filter "Mode", da kann man auswählen zwischen "Moves", das sind alle Bewegungen, also ich habe, ich sollte vielleicht damit anfangen, was für Daten das sind, wir haben zwei verschiedene Datentypen und wir fangen mit Flowdaten an, das heißt das sind Daten, die zeigen, wann sich Handys zwischen welchen Funkzellen bewegt haben. Und diesen Bewegungen wird jeweils ein Mode zugeschrieben aufgrund eines Modells, das außerhalb meiner Arbeit liegt. Genau, also "Moves" sind alle Modes, einfach nachfragen, wenn ich irgendwas nicht deutlich erkläre, Sie sind die ersten. Genau. Und "Privat" sind alle Bewegungen, denen kein anderer Typ zugeschrieben wurde. Und dann gibt es "Rail", das ist im Prinzip S-Bahn, aber auch eben Regionalzüge, also alle oberirdischen Gleisbewegungen*

**A** Und "public" ist die Summe der unteren vier oder?

*Richtig, genau.*

**A** Wie kann die Verkehrsmittelzuordnung passieren?

*Soweit ich weiß, passiert das hauptsächlich über Bewegungsprofile eben, also wie sich die Leute bewegen. Die Daten die da zugrunde liegen sind ein bisschen genauer als die die ich hab, aber ich muss halt auf dieser aggregierten Ebene arbeiten und zu den Daten gab es auch noch ein kleines Sample an GPS-Daten, die damit verknüpft waren.*

**C** Und "Privat" heißt jetzt nicht, dass der Fahrzweck quasi privat ist, sondern dass ich Autofahrer, zu Fuß, ..

*Genau, alles was nicht zu dem anderen zugeordnet werden kann.*

**C** Ok also privates Verkehrsmittel quasi, kein privater Fahrzweck.

*Genau, ja. Sie sehen auch hier schon mal, oder wir können gleich mal gucken, also hier hab ich einmal "public" ausgewählt, einmal "privat"*

**A** Man könnte ja private auch weiter untergliedern in Fahrrad, zu Fuß und Auto..

*Potenziell, ja.*

**A** Das ginge ja vom Bewegungsprofil viel besser, als tatsächlich das Auto vom Bus zu unterscheiden.

*proposal: further differentiation within mode "privat": biking, walking and car*



*Ja, wahrscheinlich. Genau, und einmal "Tram", das schauen wir uns dann gleich an. Dann gibt's diese Darstellung, das nennt sich "Heatmaps", da sehen Sie von oben nach unten die Wochentage - also ich hab Daten von einer Woche im März letzten Jahres, der elfte bis 17. März - und nach rechts die Zeit, also Stunde null bis Stunde 23. Und je nachdem, wie stark das eingefärbt ist, ist die Anzahl an Bewegungen, in dem ganzen Raum.*

**C** Auf welchen Raum bezieht sich das, auf die Stadt München oder auf eine bestimmte Strecke?

*Im Prinzip auf die Stadt München, also bei dieser Visualisierung jetzt auf alles und später können wir es dann auch auswählen.*

**C** Ok, also das ganze Stadtgebiet..

*Genau, ich zeig Ihnen gleich, wie genau die Ausdehnung ist.*

**B** Also die y-Achse ist die Wochentage, und x ist Zeit.

*Zeit, genau. Jede Stunde hat eine Zelle. Die gleichen Daten sieht man hier, allerdings in einem Kreis abgebildet, das ist im Prinzip wie ein Säulendiagramm, nur als Kreis. Und die Höhe der jeweiligen Farbe zeigt, wie viele mit dem jeweiligen Transportmittel unterwegs waren. Da können wir dann auch auswählen, ob wir nur die öffentlichen sehen wollen oder nur die privaten. Hier oben sieht man immer das Maximum.*

**B** ..Maximum Moves oder?

*Ja, genau, und da ist es eben für den ganzen Tag. Und dann gehen wir ein bisschen ins Detail und schauen uns die einzelnen Tage an. Und hier haben wir jetzt ein Säulendiagramm, wie sie es auch kennen. Der Cut ist jetzt hier von zwei bis zwei Uhr. Zum Einen, weil die IDs alle zwischen zwei und drei Uhr gewechselt werden alle zwei Tage, das heißt, da ist es ein bisschen unregelmäßig, und zum Anderen endet der Tag ja..*

**A** Das hab ich jetzt nicht verstanden, warum endet der Tag nicht um Mitternacht?

*Weil die Bewegungsmuster ja auch, also wenn man nachts unterwegs ist, dann kommt man - wenn man zwischen ein und zwei Uhr unterwegs ist, dann ist das eher eine Bewegung, die zum vorherigen Tag gehört, als..*

**A** Grad beim öffentlichen Verkehr, oder S-Bahn überhaupt, die fährt bis eins oder zwei und dann hat sie Betriebspause.

*Genau, ja. Das spielt da auch eine Rolle.*

**A** Wobei ich mich wunder, dass nachts eigentlich gar keine Bewegungen stattfinden, auch nicht am Wochenende. Oder ist das nur so skaliert, dass da diese Heatmaps, die waren ja stockfinster aufd' Nacht.

*Ja, das ist die Skalierung hauptsächlich.*

**A** Da sieht man ja nichts.. wenigstens am Wochenende müsste ja dann durchgehend..

*Ja, das ist vielleicht am Beamer auch noch schwerer zu erkennen, hier sieht man schon, dass es da ein bisschen heller ist, als da, aber ja, das ist die Skalierung, deswegen sieht man das dann hier auch besser.*

**B** Aber nur weil es dunkel ist, heißt es ja nicht, dass es da keine Bewegungen gibt, weil auf der Skala sieht man, dass es immer noch zwischen null und 80 000

*Genau. Wenn wir uns die Tagesganglinie anschauen im Prinzip, das kennen Sie ja als Säulendiagramm auch, da haben wir auch wieder den Maximalwert angezeichnet und dann gibt's eine Darstellung, wo nur die privaten, also nur die unzugeordneten Verkehre dargestellt werden und eine wo beide dargestellt werden und das ganze gibt's auch als Kreisdiagramm. Und hier fängt Mitternacht, also null Uhr ist hier links, wo man auf einer uhr neun Uhr gewöhnt ist, da muss man ein bisschen drauf achten vielleicht, weil wir ja 24 Stunden haben und nicht zwölf. Genau. Die Morgenstunden sieht man rechts oben dann und die Nachmittagsstunden rechts unten. Gibt es zu den allgemeinen Darstellungsformen schon Fragen oder sollen wir beginnen?*

**C** Nee, also für mich ist jetzt auch soweit alles klar.

*Ok. Genau, und die Farbkodierung bleibt auch so erstmal.*

**A** Eine Frage vielleicht noch zu dem 4.2-Diagramm, dem mittleren, was zeigt das graue nach unten?

*Die privaten Verkehre.*

**A** Achso. Ok, danke. Ist klar.

**C** Und die 117 Tausend ist quasi der Höchstwert, der erreicht wird in einer Stunde. Also nicht die Summe von allen Fahrten..

*Nee, genau. Das ist der Höchstwert, der hier ist.*

**C** Und wo kommen die Daten her, das sind Handydaten auch einfach, also ist das Google oder..

*Ja, genau, Telefonica auch.*

**A** Ja, wir haben die gleichen Daten aber leider gibt's da überhaupt keine Unterscheidung..

modes are appreciated /  
missing in data they have  
been using

*Ja, genau, die Unterscheidung wird eben in dem xMND-Projekt gemacht, Fraunhofer*

*macht das, also von denen krieg ich die Daten vorbearbeitet.*

**A** Ach die bereiten die noch nach.

*Genau, ja. Also das Hauptziel in dem Projekt ist eben, die nutzbar zu machen für den öffentlichen Verkehr, für die Planung. Ok, dann fangen wir jetzt hier an - Einer von Ihnen darf anfangen, mit der Maushoheit. Sie dürfen jetzt hier Filter einstellen... ..Sie können versuchen, auch zu gucken, ob irgendwas auffällig ist, was wir uns dann genauer anschauen sollen.*

Vis 1

**C** Und wenn ich hier Tag schon eingebe, dann..

*..dann wird der Tagestyp ignoriert.*

**C** Ok. Da brauch ich also nichts mehr..

*Genau. .. Ja, das dauert ein bisschen. Ich kann das auch noch größer machen, moment, so. Also hier sieht man die Ausdehnung, das sind die Zellen, die verwendet wurden, da ist jetzt im Süden ein bisschen mehr abgeschnitten, das liegt einfach daran, dass der Flughafen noch dabei ist, damit es nicht zu viele Daten werden.. Genau und jetzt können Sie auch zoomen und ein bisschen damit spielen.*

**C** Also ich hab jetzt ausgewählt "Zug", das heißt, das ist jetzt S-Bahn und Regionalverkehr, Fernverkehr nicht, oder?

*Doch, auch. Also alles..*

**C** Gut, da fährt ja wahrscheinlich innerhalb von München keiner. Man sieht halt die S-Bahn..

slight irritation / confusion  
because of misassigned  
modes

**A** Aber da sind doch auch da Bewegungsmuster, Pfeile, wo keine Schienenstrecken sind.

*Genau, also was die Pfeile.. die gehen immer von der Mitte der Zelle aus, das heißt aber nicht, dass man in der Mitte angefangen hat, sondern halt von der Zelle in..*

**B** Das ist ja Quellzelle und Zielzelle..

**C** Also wenn da jetzt einer von Schwabhausen nach Ebertshausen fährt und die S-Bahn fährt halt die Kurve hier, das wird dann aber als Verbindung dann trotzdem direkt gezeichnet oder? Der wird jetzt quasi nicht entlang der S-Bahn..

*Der ist, also in der Darstellung ist es so genau wie möglich. Also das ist nicht von-nach, sondern das ist wirklich der Weg durch die Funkzellen im Prinzip.*

**C** Ah, ok.

*Das Modell ist natürlich nicht perfekt, also ..*

A Geh mal da hin wo ich mich auskenn..

C Das musst du mir sagen aber ungefähr kann ichs.. Noch weiter rüber oder? Aber das ist wahrscheinlich schon nicht mehr dabei oder? Was sind die orangenen Linien jetzt, weil ich hab ja eigentlich nur..

*Die dünnen? Das sind die Begrenzungen von den Zellen. Wenn man da drüber geht, dann sieht man auch, welche Zelle das ist, genau.*

C Aha. Und diese ganz dünnen Linien, das sind auch.. gehen da auch Schienen? Eigentlich nicht oder?

cross referencing with background knowledge about network

*Nein, das sind halt Bewegungen, die dem Modus "Schienen" zugeordnet wurden*

C Ah ok. Ja klar, Innenstadt ist natürlich am Meisten los.. Aha.. Ja interessant.

*Wenn wir jetzt sehen, also ich wollte jetzt grad sagen, wie wir sehen ist in der Innenstadt immer viel los, wenn man hier auf "normalize" geht, dann wird das normalisiert gegenüber dem Mittelwert über die ganze Zeit auf diesem Segment.*

A Was wird dann gemacht?

*Die Dicke der Linie wird normalisiert. Also jetzt sieht man, da ist der Hauptbahnhof kleiner, weil da einfach immer viel los ist, jetzt sieht man praktisch, Samstag zwischen sechs und 18 Uhr - wo ist da mehr los als sonst, oder ja, wie viel ist da los im Vergleich zu dem kompletten Durchschnitt.*

C Und wie seh ich jetzt, ob mehr oder weniger los ist, einfach, wenn die Linie dicker ist ist es mehr..

tooltip is being used

B Da steht das sogar, null Komma vier times the normal amount.

C Und wie muss ich hier, ist da jede Bewegung eine Linie oder ist das irgendwie..

*Nee, das ist hier jetzt zum Beispiel, die Linie sind 43 Bewegungen. Also die Linien gehen immer von .. zwischen zwei Zellen gibt es immer nur eine Linie.*

C Aber das sagt mir jetzt nicht von wo nach wo, das muss ich dann immer in der Karte gucken, von wo nach wo die Linie geht? Ich seh halt 43 Fahrten..

*..mhm, und die FlowID ist die, also das sind die Zellen. Zwischen den Zellen 112 und 164, wobei die Richtungen da keine Rolle spielen.*

C Und das ist jetzt halb so viel wie es im Durchschnitt ist.

Genau.

**C** Ja, ist ja logisch eigentlich.

*Deswegen ist sie auch so dünn die Linie.*

**B** Der wöchentliche Durchschnitt oder welcher Durchschnitt?

*Der komplette Durchschnitt von den ganzen Daten in dem Fall.*

**B** Von den ganzen Daten. Auf dieser Beziehung.

Ja.

[...]

*Ich hätte jetzt gesagt, wir können noch mal die U-Bahn anschauen, weil die ist ein bisschen..*

**C** da.. oder wählen sie gleich aus.

**B** Vielleicht erstmal ohne normalize.

*Da sieht man auch wieder, dass am Hauptbahnhof ganz viel los ist.. Da ist es so, dass die.. das U-Bahnnetz hat Verstärker für die Funktürme, in den Tunneln drin, die aber alle über den Hauptbahnhof laufen. Das heißt, wenn man irgendwo in die U-Bahn einsteigt, dann springt das Telefon sofort zum Hauptbahnhof. Das führt dazu, dass man die gut erkennen kann, aber zum Anderen eben auch dazu, dass alle erstmal zum Hauptbahnhof fahren, wenn man das nicht noch genauer analysiert und schaut, wo sie dann wieder aussteigen.*

**C** Gibt's da auch irgendeinen Wert für die, ich sag jetzt mal in Anführungszeichen Fehlerwahrscheinlichkeit, dass er mir eine U-Bahnfahrt ausspuckt aber die gar nicht als U-Bahnfahrt stattgefunden hat, also nur falsch zugeordnet wurde?

uncertainty information is asked for

*Ja, den gibt es noch nicht. Genau. Also das Projekt ist halt auch noch am Laufen*

**C** Dass man halt sagt, mit 90% ist es auch wirklich eine .. oder keine Ahnung wie hoch die Sicherheit ist, aber..

Ja, genau.

**C** Bei Martinsried gibt's glaub ich noch keine U-Bahn, die wird erst gebaut, oder? Da wo jetzt die.. oder ist das Großhadern schon?

**A** Das ist Großhadern.

C Ja, ok. Dann wirs stimmen ja.

A Da geht aber noch ein ganz dünner Pfeil nach Martinsried.

C Die sind dann vielleicht mit dem Bus weitergefahren

A ..oder mit dem Auto..

*Es kann auch sein, dass das Handy sich halt in beide Zellen einwählt, und dann die andere gewählt wird.. Also da sind noch Unsicherheiten drin auf jeden Fall.*

B Diese Darstellung ist dann ein bisschen .. suboptimal im Vergleich zu der Darstellung mit der S-Bahn zum Beispiel, weil da hat man dieses Problem nicht, dass jede Fahrt dem Hauptbahnhof zugeordnet wird. Weil hier sieht es jetzt so aus, als ob alle in Richtung Hauptbahnhof fahren würden, was nicht immer stimmt..

display of underground movements "suboptimal" due to assignment to cell 51

Ja.

B Und wofür ist dann diese Darstellung gut sozusagen? Weil hier kann ich die .. kann ich immer noch die Quell- und Zielzellen sehen von dem Movement sozusagen? Weil wenn die immer zum Hauptbahnhof zugeordnet werden, dann steht das immer, Quellzelle, keine Ahnung, Zelle 100 und Zielzelle als Hauptbahnhofzelle?

*Ja. in dem Fall schon noch. Wir kommen noch zu ein bisschen anderen Daten, aber da ist das Problem immer noch da. Also.. ja. Das ist einfach noch ein Schritt in der Modellierung gelöst werden muss. Wir können noch mal kurz auf den Privatverkehr gucken, weil das ein bisschen anders aussieht.*

[ recording stopped and resumed ]

*Ja, also hier sieht man auf jeden Fall schon die Autobahn..*

A Aber man sieht auch, dass im Dachauer Hinterland so und so nichts los ist, also auch bei den privaten Verkehrsmitteln.

*Ja, also das sind hier 1000 Bewegungen. Wenn wir da auf normalisieren gehen ... genau, dann sieht man extrem wenig, weil im Privatverkehr halt so ein irgendwie gutes Grundrauschen hat, denk ich. .. Dann würde ich sagen, schauen wir uns die zeitlichen Verläufe an. Hier sind nochmal die Heatmaps. Gibt es da schon was, was Ihnen auffällt?*

B Man kann die Hauptverkehrszeiten sehr gut sehen, vor allem am Vormittag, im ÖV. Also die zwei Hauptverkehrszeiten.

A Und dass das Muster halt sehr ähnlich ist über alle Modes.

Vis 2  
general patterns are being recognised, comparisons between modes → scales and colours should be consistent to facilitate comparison

**B** Außer die, was ist rot nochmal?

*Tram.*

**B** Tram. Und blau ist?

*U-Bahn.*

**B** U-Bahn. Da ist ein bisschen mehr ausgebildet, am Nachmittag vor allem. Die am Vormittag sind sehr präsent und die am Nachmittag sind ein bisschen .. im Vergleich zum Beispiel zu, was ist lila oder rot?

*Bus. .. Genau, dann haben wir diese Darstellung, da können Sie jetzt wieder auswählen.*

Vis 3

**C** Willst du, B, oder..

**B** Dieser Wert hier bedeutet, das ist der höchste Wert von einem Tag..

*Genau, der höchste Wochentag.*

**B** Da wissen wir jetzt aber nicht, ob das Dienstag ist oder Mittwoch ist.

→ indicate maximum amount  
at max. hour

*Ja, das ist ein bisschen klein.. Also es ist der Mittwoch.*

**B** Kann man das sehen, welcher Tag das ist?

**A** Reinzoomen.

*Ja.. Aber wir können auch gleich übergehen zu den Tagen. Da haben wir jetzt, da können Sie jetzt auch vergleichen, also wenn wir Dienstag und Mittwoch anschauen, genau. Dienstag sind 149 Tausend*

Vis 4

**A** ..und jetzt mach Mittwoch.. 141.

*Ah das ist weniger. Genau, ja und unten waren wir bei privat und öffentlich zusammen oder?*

**B** Nee, das waren nur..

*Was ist dann.. moment.. .. Achso, ja, klar. Hier [Vis 4] ist das Maximum die maximale Stunde und die Summe ist natürlich eine andere dann. Also hier [Vis 3] ist das Maximum der Tagessummen und oben [Vis 4] ist das Maximum der Stundensumme. Das heißt, Mittwoch, hier sieht man jetzt am Laptop jedenfalls noch einen kleinen weißen Streifen, den sehen Sie jetzt da an der Wand natürlich nicht.. Also in Summe ist es am Mittwoch mehr aber am Dienstag gibt es eine größere Maximalstunde.*

**B** Wenn ich was vorschlagen darf, würde ich vorschlagen, dass dieser, der

proposal: indicate  
maximum at maximum  
day or annotate maximum  
day on top

Wert hier den man hier sieht, kann man das auch so darstellen, dass er immer da auftaucht, wo es tatsächlich der Wert ist? Weil hier zum Beispiel ist es relativ klar, dass da zwischen sieben und acht der höchste Wert ist aber hier war das eben nicht der Fall, deswegen wäre es sehr hilfreich, wenn dieser Wert dann hier am starken Tag steht. Oder oben, dann kann man den höchsten Tag noch als Textfeld hinzufügen und sagen "Wednesday: .." oder irgendwie sowas halt, das wäre der Vorschlag.

*Ja, gute Idee, danke. .. Wenn Sie "circular" nicht wählen, also wenn Sie das abwählen, dann kommt das normale Säulendiagramm.*

**B** Da ist es dann eindeutig. Hier sieht man, welche Stunde den höchsten Wert hat, im Vergleich zu dem Kreis. Hier [Vis 3] kann man das nicht so einstellen, wie eben oder?

identification of maximum hour is unambiguous when looking at the horizontal version of Vis 4

*Nee, aber jetzt werden wir auch mit den Stundendarstellungen weiterarbeiten.*

**A** Haben Sie die Oberfläche programmiert oder wie?

*Genau, das läuft in Jupyter Notebooks, das sind so, ja, Notizbücher quasi, wo man den Code reinschreiben kann und dann gleich auch ausführen kann und man kann es halt reproduzieren sozusagen, also das ist ganz angenehm. Aber es ist halt noch kein fertiges Userinterface, also das..*

*proposal: show the sum of all movements of the day*

**C** Was ich auch noch cool fände, wenn man einen Tag auswählt, dass er die ganze Tagessumme auch noch angibt. Also jetzt nicht nur die Stunde, wo quasi am meisten ist, sondern immer irgendwo noch so-und-so-viel Fahrten an einem Dienstag oder einem Mittwoch..

*Ja. Ok. Dann um zum nächsten zu kommen schauen wir uns jetzt nochmal den Mittwoch an.*

**A** Da ist die Spitze am Nachmittag.

*Genau. Und wenn man das vergleicht mit einem normalen Montag bis Donnerstag, oder mit allen.. Hier. Da ist die Spitze am Vormittag.*

**B** Noch eine Frage - Den DayType haben sie jetzt als Montag bis Donnerstag - ist das Montag bis Donnerstag oder ist das der Durchschnitt..

*Das ist der Durchschnitt der vier Tage, genau. Also ein typischer Wochentag sozusagen. Wenn man jetzt noch mal vergleicht, also hier haben wir den Durchschnitt und da am Mittwoch, ist eben die Spitze am Nachmittag. Fällt sonst noch was auf? .. Ich geh noch mal zurück .. und nochmal zum Mittwoch ..*

**B** Ja, dass der höchste Wert ein bisschen höher ist als der Durchschnitt.. Und allgemein ein bisschen mehr Verkehr in der Nacht, sozusagen, im Vergleich zu dem Durchschnitt.



C Und was ist das Lilane, U-Bahn, nee, U-Bahn ist blau..

*Genau, lila ist Bus*

C Ja, halt in späten Abendstunden mehr Bus oder? Zwischen 10 und 11. Aber gut das ist..

*Vielleicht gehen wir noch mal zu der Kartenansicht zurück und schauen uns den Mittwoch an.. ..wollen Sie einfach einstellen..?*

Vis 1

B Ja ..

A [lacht] Also wir müssen die normalisiert darstellen, für unsere Verbund-raumerweiterung. Sind das die Zellen, die dann Teralytics zur Verfügung gestellt hat, oder wo sind die her.. die Daten, haben die Teralytics zur Verfügung gestellt für die Masterarbeit oder..

*Für das Projekt insgesamt. Genau, wir schauen uns noch mal Mittwoch Abend an, weil da ja ein bisschen was anders war.. fällt da was auf?*

B Das hier, diese Beziehung..

A Da war eine Veranstaltung, wahrscheinlich ein Champions-League-Spiel

Anomaly recognized with some guidance

*Bingo! Genau. Wir können auch noch mal auf die U-Bahn gehen, da sieht man deutlich, dass da mehr los ist, ne? Da ist die Allianzarena*

A Und die kommen auch von Garching, also die Linie geht ja über die Arena hinaus.

*Ja, wobei Garching ist da..*

C Ah, Garching ist weiter draußen.

*Die Arena ist in der Zelle und das kommt aus der Zelle, also ob das jetzt von da kommt oder..*

reminder necessary that lines connect cell centroids and do not depict the actual path of movement → can this be improved visually?

A Ja, also es kommt von außen wenig, es kommt alles aus München raus.

*Ja. Jedenfalls mit der U-Bahn. Wir können nochmal die komplette..*

A Achso das war nur U-Bahn jetzt.

*Das war nur U-Bahn, genau. Wir können noch mal alle zusammen anschauen. .. Ja, also man sieht dass da auf jeden Fall mehr los ist, ne?*

A, B, C Ja, mhm.

Gut, dann ist der nächste Schritt jetzt

[ recording stopped and resumed ]

Stage 2

*Also der nächste Schritt ist der Vergleich von Mittwoch mit den anderen drei Wochentagen. Und da haben wir jetzt auch wieder eine Kartenansicht und jetzt sind die Farben ein bisschen anders. Und zwar sehen wir in grün, wenn es am Mittwoch mehr ist und in schwarz, wenn es am Mittwoch weniger ist.*

**B** ..im Vergleich zu dem Durchschnitt.

*..zu dem Durchschnitt von Montag, Dienstag und Donnerstag, ja. Man kann wieder den Modus einstellen, die Tageszeit, was die Mindestabweichung sein soll, das sind absolute Zahlen, also mindestens 10 Leute mehr oder weniger. Und die Zahlen sind alle noch nicht hochgerechnet, also das sind wirklich nur die Telefonica-Zahlen, das hätte ich vielleicht am Anfang sagen sollen. Man kann vielleicht mal vier denken.. und da kann man noch die Dicke einstellen, wenn es nicht passt. Und relativ bedeutet hier, dass dann die Dicke abhängig von der relativen Anomalie ist.*

**B** Das passt sich an, quasi.

*Genau, wenn relativ angewählt ist, dann heißt das, wenn ich auf einem Segment 110 statt 100 Leute habe, dann wird die Dicke berechnet, indem ich zehn durch 100 teile. Also ein Zehntel mehr. Und wenn es nicht angewählt ist, dann sind das die absoluten Anomalien. Und in der zeitlichen Darstellung ist es ähnlich. Welcher Modus angezeigt wird, das sieht man hier [Vis 5] an den Zellfarben, die sind dann anders, bei der Tram sind sie jetzt zum Beispiel rot. Hier [Vis 6.1] ist auch die Farbe in der Mitte, die den Modus angibt. Dann sehen wir in grau den Durchschnitt von Montag, Dienstag und Donnerstag für jede Stunde, und dann in grün oder schwarz das was am Mittwoch mehr oder weniger ist. Und hier hab ich zum Beispiel ausgewählt von Zelle 51, das ist der Hauptbahnhof, mit der Tram irgendwoandershin.*

**B** Das sind dann alle ausgehenden moves aus der Zelle

*Genau*

**B** Mit Tram.

*Mit Tram, ja.*

*Und das Gleiche haben wir auch noch mal als normales Säulendiagramm, mit der gleichen Auswahl.*

**B** Können Sie nochmal reinzoomen in die Abbildung [Vis 6.2]? Was ist oben und unten in dem Fall?

*Ah, genau, oben und unten ist das gleiche wie hier [Vis 6.1] links und rechts, das ist Hin und Zurück. Das heißt, hier ist es von Zelle 51 und hier ist es zur Zelle 51. Also rechts oder unten ist immer das Gegenteil sozusagen.*

**B** Und die Skalierung hier ist auch.. also was ist das drei Komma acht..

*3,8 Tausend sind die, dieser ganze Balken zum Beispiel.*

**B** Das heißt, die Skalierung ist einheitlich oder? Also die oberen Balken haben keine andere Skalierung als die unteren.

*Ja. Also die Mittellinie passt sich an.*

**C** Das sieht jetzt sehr symmetrisch aus. Ist das immer so oder ist das Zufall? Oder, versteh ich das richtig, um acht Uhr, sieben Uhr, haben wir quasi zum Hauptbahnhof mehr Fahrten als im Durchschnitt, oder vom Hauptbahnhof weg, und zum Hauptbahnhof weniger oder?

*Genau, also es kommen ein bisschen mehr rein als rausgehen, moment, andersrum, hier: vom Hauptbahnhof nach draußen ist oben und unten ist von draußen zum Hauptbahnhof. Genau, das heißt also in dem Fall könnte man sagen, mit der Tram fahren Leute vom Hauptbahnhof weg aber die kommen halt wahrscheinlich anders.*

**A** Und am Nachmittag müsste es dann eigentlich vom Verhältnis her andersrum sein. Wenn die Leute heimfahren.

**B** Ja nicht unbedingt, weil die Nachmittags-Spitzenstunde ist ein bisschen schwammig, die ist nicht so präsent wie am Vormittag, also da verteilt es sich ein bisschen.

**C** Aber der grüne Balken heißt doch immer, dass es mehr sind, oder, haben Sie gesagt?

*Genau, am Mittwoch gegenüber der anderen..*

**C** Ok, da steht aber minus 2750 am grünen Balken, was heißt das dann?

**B** Das ist nur die andere Richtung.

*Ja, genau.*

**C** Ah ok.

*Aber guter Punkt, das Minus ist da falsch.*

**C** Das heißt, es sind in beiden Richtungen mehr als im Durchschnitt.

*Zu der Stunde, ja. Und hier sind es zum Beispiel weniger. Genau, dann schauen wir uns das an. Sie dürfen wieder auswählen, jetzt sind wir wie gesagt bei den Anomalien vom Mittwoch.*

**B** Diese Anomalien min. Das hat wahrscheinlich einen Einfluss auf die Re-

question about consistency of scale → maybe indicate consistent scale with consistent grid, apply consistent scale for both directions in circular versions, too!

finds sign mistake in plot → implies careful exploration of graphs

Vis 5 shows interest in code execution process

chenzeit oder?

*Ja ja, je mehr, also je feiner der Filter ist, desto schneller, also desto weniger Linien zeichnet das natürlich und desto schneller geht das. Genau.*

**B** Also je gröber, desto schneller oder? Und je feiner, desto mehr...

*Hmm, der Filter wird ja immer angewandt und ich glaube die Hauptrechenzeit ist da die Visualisierung der einzelnen Linien. Also das heißt, je mehr man runter filtert, desto schneller geht's.*

**A** Das ist ein Prozentwert?

*Nee, das sind absolute Abweichungen, also da sind 55, mindestens 55 Leute mehr oder weniger unterwegs.*

**A** Als während dem Rest der Woche zu diesen Zeiten.

*Genau. Man kann auch da nochmal drüber gehen mit der Maus, über die Linie, dann sieht man das noch mal genauer.*

**B** Ja, 5,3 mal das normale.

*Ja, hier sind es acht mal so viele.*

**A** ..dann dürfte da nichts sein.. ..wo sind wir denn hier.. ..Schloss Nymphenburg.. Das sind die, die vor dem Spiel noch München anschauen.. ..Da ist von Haus aus kein Verkehr und deswegen auch keine Anomalie.

*Genau, wenn man das relativ abwählt, dann würde man da auch was sehen, das funktioniert genauso wie vorher, da sieht man dann auch Abweichungen, also wenn zum Beispiel eine Abweichung ist von zehn Leuten, da aber normalerweise niemand unterwegs ist, dann..*

**A** Aber das sind natürlich 100 mehr pro Zelle, also.. ..ok.. nach dem Spiel.. ..und das ist die Abweichung gegenüber dem Wochenschnitt, Monatsschnitt, durchschnittlichen Mittwoch.. oder?

*Durchschnittlicher Montag, Dienstag, Donnerstag. Also ich hab nur Daten von einer Woche, das heißt ich hab den Durchschnitt von Montag, Dienstag und Donnerstag berechnet.*

**A** Und das sind hier zwei Zellen oder? Das eine ist Allianz Arena und das andere ist Kieferngarten oder so..

*Genau, ja, das sind jetzt hier zwei. Es kann sein, dass die U-Bahn hier auch noch oberirdisch fährt oder?*

*“Nymphenburg Palace .. that's are the ones that want to explore the city before the game” → reasoning based on context knowledge and base map*

C Ja, ab Studentenstadt.

*..und dass das dann..*

C Das heißt, dann wir es zu Rail gezählt oder, nicht U-Bahn?

*Das ist eine gute Frage, können wir mal gucken, wenn man Rail auswählt.*

C .. Nein, da sieht man nichts.

A Naja, die werden das Netzwerk ja schon einpflegen.. Und da wo die U-Bahn ist kann man ja gar kein Rail finden..

*Ja, wie das genau funktioniert, weiß ich leider nicht.*

A ..die Tram profitiert nicht wirklich vom Spiel.. ..aber die haben da die Bewegungen so fein, dass sie da eine Bewegung nachvollziehen können. Wie ist das vom Bewegungsprofil mit Stops und Durchschnittsgeschwindigkeit und so weiter..

*Soweit ich weiß, ja.*

A Wer bereitet das auf?

*Fraunhofer macht das Modell und da ist glaub ich noch eine andere Firma dabei, das kann ich noch mal nachschauen. Dann können wir uns auch noch den zeitlichen Verlauf anschauen. Da können wir jetzt zum Beispiel mal gucken, welche Zellen interessant sind links [Vis 5]. Also .. ja, wir können mal alle nehmen. Und wenn wir jetzt auswählen, die Allianzarena ist in Zelle zwei.. und die daneben ist die 71, die können wir noch dazu nehmen.*

shows interest in who is responsible for data processing

Vis 6

A Was sind denn das für Zellen?

*Um die Funktürme*

A Sind das Funkzellen.

*Ja. .. Dann kann man hier auch den öffentlichen Verkehr anschauen..*

C Das heißt links ist dann wieder von Zelle 2 und 71 in alle anderen und recht ist dann wieder..

*Genau, ja. Wenn man jetzt schaut, wer zum Hauptbahnhof fährt, dann wird das auch ein bisschen deutlicher.*

C Ja, schon cool.

impressed by result for start cells 71,2 and end cell 51 in Vis 6 (anomalies)

A Die Leute fahren um halb sieben hin und um halb elf zurück.

C Das kommt gut hin.

A Das kommt hin, ja.

*Man sieht auch, wenn man den privaten Verkehr nimmt..*

C Da ist das ein bisschen ausgewogener.

*Nicht ganz so spitz, ja.*

[ recording stopped and resumed ]

Stage 3

*Dann wäre der nächste Schritt, also wenn Sie da nicht mehr rumspielen wollen .. dass wir jetzt OD-Daten anschauen, also Origin-Destination, sprich von wo nach wo die Leute fahren, um zu schauen, wo die Leute herkommen, die zum Spiel fahren. Und da sind die Ansichten wieder ähnlich. Genau, eigentlich unterscheidet sich das kaum von diesen hier [Vis 1 & 4], nur dass es mehr Filteroptionen gibt.*

A Aber die Karte müsste sich doch unterscheiden oder?

*Jaja, klar. Ich meine die Art der Darstellung im Moment. .. Genau. Dann können Sie wieder einstellen, was Sie einstellen wollen.*

Vis 7

A So. Jetzt hab ich von überall nach überall einfach.

*Ja. Wobei ..*

A Nur die weiten Entfernungen. Das ist dann eigentlich nur Flughafen nach Eching.

C Aber die Kilometer werden nur innerhalb des Raumes gerechnet, also wenn eine von Ingolstadt nach München fährt, das ist dann wurscht oder?

*Ja. Das wird nicht aufgenommen. Und es wird auch nur, aus Rechenkapazitätsgründen, werden eben auch nur die Bewegungen innerhalb jeder Stunde aufgezeichnet. Das wäre auch der nächste Schritt eben.*

C Und 60km ist dann das weiteste was im Untersuchungsgebiet geht.

*Genau, also das Maximum war irgendwie 58 oder so.*

A Kann ich das auch irgendwie normalisieren?

asks about possibility of normalisation to show only large values

*Moment. Nein, das hab ich da rausgelassen.*

A Dass man irgendwie nur große Werte darstellt oder so.

*Mhm, also die Zeit, wie lange jemand gefahren ist, kann man erhöhen.*

**B** Das nützt wahrscheinlich nicht viel, weil..

**A** Wenn ich jetzt die Kilometer runternehm dann hab ich jetzt nur noch alles grau.

**B** ..nur alle Fahrten die mehr als 30 Minuten dauern.

*Ja.. Vielleicht gehen wir nochmal zu dem Mittwoch zurück.*

**A** Mir fällt's grad nicht ein, wer hat das gesagt, will wirklich niemand nach Dachau..

**C** Ich glaub der L. hat das gesagt.. ..nee irgendeiner hat's gefragt.. der Herr M. Also nach Dachau schon aber halt nur in den Süden.

**A** Die Verkehrsbeziehungen der Beitrittslandkreise zum MVV. Da grenzt eben keiner an Dachau an, deswegen. Und dann haben wir auch nur Werte über 500 reingeschrieben, und dann war kein Pfeil nach Dachau. Das hat irgendeiner komisch gefunden.

tell about an example where a high threshold led to misjudgement ("nobody wants to go there")

*Also ich hab jetzt nochmal Mittwoch Nachmittag oder Abend eingestellt...und da kann man jetzt eine Start- oder Zielzelle auswählen*

**B** Dann vielleicht Start 51 und dann Ziel 2

*2? Ok.. Genau*

**B** Das ist jetzt sehr wenig..

*Das ist weil nur eine Verbindung angezeigt wird, dann wir das normalisiert.*

**B** Das sind jetzt also insgesamt 1080 Fahrten vom Hauptbahnhof zur Allianzarena.

*Genau. Wir können auch schauen..*

**A** Aber wo ist jetzt da der Balken, also..?

**B** Der ist da ganz dünn

*..weil eben nur einer angezeigt wird, dann hat der die...aber wenn wir jetzt die Startzelle weglassen..*

can't find line → normalisation of line thickness is problematic when number of plotted lines is small. → either normalise dynamically / manually or highlight chosen connections and make rest more transparent

**A** ist das jetzt U-Bahn wieder oder was?

*Das sind alle jetzt im Moment.*

**A** ..wieso ist dann nur zwischen Hauptbahnhof und Arena was los?

**B** da ist auch Richtung Westen..

**A** Ja aber das ist nicht viel.

*Ja, also der Großteil kommt eben aus der Innenstadt.*

**C** Die aus Osten sind halt wahrscheinlich nicht dabei, weil die halt von woanders herkommen, aus Salzburg zum Beispiel, da fährst du ja schon außerhalb von München los.

*Ja, wobei der Start ist ja innen*

**C** Stimmt, ja, der wählt sich ja ein..

*Also wir können jetzt auch mal die privaten anschauen*

**A** Aber nach Vaterstätten raus fährt kein Mensch..

**A** Der Dachaubalken ist in Wirklichkeit die eine Zelle für alle die über die A8 kommen, oder? Und sich dann da halt erstmals anmelden im System

*Genau, also hier sieht man eben auch sehr gut, dass die Autobahnen..*

**A** Da sind die zwei Autobahnen

**C** Aber dass aus dem Osten nichts kommt ist ein bisschen.. ..vielleicht gibt's da keine Telefonica-Nutzer

*Ja, das hört halt hier auf, die Daten die wir haben. Wenn man die öffentlichen Verkehre nimmt..*

**A** ..das ist auch einigermaßen klar [lacht]

*Ja. .. So, dann können wir noch den Tagesverlauf auch hier anschauen.. ..wer möchte das auswählen?.. Genau das ist jetzt wieder wie ganz am Anfang, weil keine Start- und Zielzellen ausgewählt sind..*

**A** Ja, ich hab keine Tastatur..

*..ja, ich geb ein, wieder die gleichen?*

**A** ja.. nein, machma mal nur Ziel Allianz

asks about  
misassignment of origin  
due to limited  
geographical extent of the  
data → maybe define custom  
cells at the border to  
differentiate between cur-off  
movement and actual  
O-/D-cells

Vis 8



Ok. ..

A Aha. was ist jetzt da violett?

Bus.

[...]

A ÖV ist nicht alles.. was ist das dann?

B Das ist privat.

A Jaja, aber das heißt mit der U-Bahn..

*Das ist jetzt nur die eine Richtung.*

A Das ist die Zielzelle..

Genau

A Also mit dem Auto kommen die während dem Spiel an..

*Ja, da fährt halt auch der Autobahnring durch die Zelle, ne. Wenn man links guckt..*

A Ja, aber was durchfährt wird ja nicht berücksichtigt oder? Sondern nur..

B Vielleicht ist da eine Raststätte oder irgendwas..?

*Kommt drauf an.. ..ja oder wenn da Stau ist, denk ich auch, dass.. also wenn man lang genug sich in der Zelle aufhält, dann wird das als Endzelle gewertet.*

A Also die fahren nicht wirklich zum Spiel sondern..

C ..hoffentlich nicht, sonst stehns da im Stau.

*ja, 14 Tausend Zuschauer..*

C Aber ist das hochgerechnet, nee das ist jedes Auto oder jede Fahrt halt gezählt, ob da jetzt 5 drin sitzen im Auto oder einer..

*Naja, jedes Handy wird gezählt.*

C Achso, jedes Handy. Schon.

*Ja. Aber halt auch nur die Telefoncadaten*

A Und nicht nur jedes Handy sondern auch hochgerechnet auf alle..

*“so they arrive during the game by car..” not possible to differentiate between movers on Autobahn and in the stadium → define custom cells that are semantically meaningful, e.g. stadium, or facilitate filtering for how long the movers stay in the end cell to enhance analysis.*

*Neenee, noch nicht hochgerechnet*

**C** Bei uns ist das schon so aber hier halt noch nicht.

**A** Achso, das sind tatsächliche Rohdaten.

*Ja, ja, genau. Wenn man die andere Richtung anschaut, also von der Zelle weg..*

**B** Was ist der Marktanteil von Telefonica?

*Ich glaube man kann so mit nem Viertel rechnen ungefähr, aber das ist ja auch unterschiedlich..*

**A** ..angeblich so dreißig Prozent aber..

**C** Ja das kann schon sein, sind ja nur noch drei. [...]

*..gut, sollen wir zum nächsten, letzten Punkt gehen?*

**C** Ja

*Dann machen wir jetzt das gleiche wie vorher auch und schauen die Anomalien an, aber eben diesmal von den Start-Ziel-Daten. Also der Unterschied von dem Mittwoch zu den anderen drei Werktagen. ... Ja. Dann darf wieder jemand wählen.*

Stage 4

Vis 9

**C** Ach, ich bin wieder dran

**A** Die ersten werden die letzten sein

**C** Ja, das können Sie wieder eingeben

*Wieder die gleichen oder?*

**C** Ja. Und was sehen wir jetzt hier anderes als gerade eben?

*Die Abweichungen. Also jetzt ist es wieder grün und schwarz, kennzeichnet ob es mehr oder weniger sind als an den anderen drei Tagen, am Mittwoch. Also 12 mal so viele in dem Fall..*

**C** Was, siebzehn Komma acht, wollen da hin. Aber das ist jetzt wieder nur Start und Ziel oder?

*Genau.*

**C** Das heißt es gibt ja unter der Woche an Spieltagen auch immer den Bus zur Arena als Verstärker. Das heißt, wenn der irgendwo hier losfährt, an der Donnersberger Brücke glaub ich fährt der los, dann zeichnet er aber den Strich

mentions extra bus line on game days → use of context knowledge

trotzdem von A nach B an und nicht jetzt..

*..in dem Fall schon, ja. Wir können nochmal zurückgehen zu den anderen Daten..*

**C** Nee ja, das.. ..aber Tram sollte vielleicht nicht zu viel mehr sein.. ..da ist auch nichts.. ..Mhm.. ..wenn man jetzt nichts angewählt hat, dann zeigt er wieder von allen nach allen oder?

*Ja. ..das ist der andere, der linke.. ..genau, das ist jetzt auch, wenn man da dann die Zielzellen eingibt vielleicht..*

**C** Ahja. ..Mhm..

*Ja, und auch hier können wir wieder den Tagesverlauf anschauen. Da können wir die gleichen Zellen eingeben.. ..genau, da sieht es so aus, als wäre viel los, aber das Maximum ist auch nur 35 oder 39*

**A** Ja, der Kreis ist immer gleich groß, die Zahlen da oben halt nicht.

*Ja. Wenn man den Maßstab einheitlich halten würde, dann würde man halt bei ganz vielen gar nichts sehen und dann..*

**C** Hier sind's 35 bloß oder was?

*Ja.*

**C** Was sind Rail zu diesen Zellen.. Das ist einfach die U-Bahn oder?

*..von den Zellen weg.. Ja, das sind halt 39 Leute, die..*

**C** ..die als Zugfahrer wieder..

*Genau.*

**C** Mhm.. ..ja, interessant.

**A** Als Zugfahrer mit Zielzelle Stadion.

**C** Na, die fahren mit Startzelle, quasi jetzt. Also die Gegenrichtung ist der rechte Kreis.

**A** Wieso, Endzelle ist doch Stadion?

*Genau, auf der linken Seite ist die Richtung und auf der rechten Seite ist die Rückrichtung. Das heißt da ist die Startzelle Stadion.*

**C** Und das sind halt dann irgendwie Fehler .. entweder sind das welche die

Vis 10

mentions different maxima on same sized circular plots → visual scale indication is missing, add to plot somehow, e.g. as a bubble representing total plot maximum in the middle

identification of misassigned moves

mit der U-Bahn fahren und als Zug gezählt werden, vielleicht weil die U-Bahn oben startet. Wenn man jetzt U-Bahn eingibt ist es sozusagen wesentlich mehr oder?

*Genau, ja, mehr als Tausend. Eins Komma drei Tausend und hier eins Komma eins.*

**A** Achso

**C** Das scheint mir aber trotzdem eigentlich fast zu wenig.

**A** Das ist ja nur die Spitzenstunde.

**C** Ja. Aber in den restlichen Stunden ist ja fast nichts.

**B** Können wir nochmal Rail auswählen?

*Mhm.*

**B** Jetzt am Vormittag.. wieso ist das dann deutlich weniger zwischen sieben und acht? Das bedeutet, dass..

*..dass am Mittwoch deutlich weniger als im Durchschnitt von Montag, Dienstag und Donnerstag..*

**B** Ja, woran kann das liegen, das sind ja schon deutlich weniger?

*Ja. Wobei es halt insgesamt auch nur 35 Leute sind.*

**B** Achso, ja.

**C** Eigentlich dürfte da gar kein Rail vorkommen.

*Wir können mal alle Zellen angucken, da ist einfach am Mittwoch ein bisschen weniger los als an den anderen Tagen, was Schienenverkehr angeht.*

**B** Mhm, da war dann einfach die Fallzahl nur sehr gering.

*Ja, genau. Also ich denke, dass da der Montag Vormittag halt schwer wiegt, dass da irgendwie mehr los ist als sonst, allgemein. .. Genau und was, achso, was bei den Start-Ziel-Daten noch zu sagen ist vielleicht, ist, dass der Modus der zugeteilt wird, ist der Hauptmodus, also die Mehrheit des Trips wurde mit dem Verkehrsmittel gefahren.*

**B** Wie in der [...], da rechnen die das auch so.

**A** Jetzt nochmal. Was bedeutet der dunkelgraue..

*..Mittwoch sind soviel weniger als an den anderen drei Werktagen, im Durchschnitt.*

A Mhm. Weil ich dachte, der grüne Ring wäre der Tageshöchstwert.

B Also dieser Außenring oder?

A Ja.

*Mhm, ja, achso. In dem Fall ist das der Tageshöchstwert vom Durchschnitt.*

A Achso, der graue geht nach unten und wenn es mehr wären dann würde es nach oben gehen.

*Genau, ja. Und hier ist es jetzt tatsächlich der Höchstwert von Montag, Dienstag, Donnerstag, weil das, ja. Weil es halt insgesamt, die Spitzenstunde ist weniger als an den anderen Tagen.*

C Der Durchschnittswert ist das von Montag, Dienstag, Donnerstag oder?

*Ja, genau,*

A Aber wir haben schon den Mittwoch ausgewählt jetzt oder?

*Jaja. Das sind die Anomalien vom Mittwoch. Aber das graue ist eben der Durchschnitt von Montag, Dienstag, Donnerstag..*

A Achso. Das ist der Gesamtverkehr, wir haben keine Start- und Zielzellen.

*Genau, ja.*

A Weil am Stadion müssten es ja mehr sein.

*Ja.*

A Genau. Da sollte es so ausschauen. Jetza.

*Genau. .. Gut, ich weiß nicht wie viel Sie da [Fragebogen] schon eingetragen haben?*

A ..die ganze Zeit mitgeschrieben

*Sehr gut [lacht]. Da gibt es auch noch eine Rückseite, da würde ich Sie bitten, das jetzt auszufüllen und dann können wir auch noch mal drüber reden, abschließend.*

[ recording stopped - participants fill in questionnaire - recording resumed ]

Discussion

*Genau, also wir können jetzt auch nochmal allgemein drüber reden, vielleicht zu den Fragen die hinten stehen, so ein bisschen. Ja. Sie können einfach anfangen zu erzählen, was Ihr Eindruck ist und wie sie das bewerten und so weiter. .. Was Ihnen aufgefallen ist oder ..*

**C** Also mir gings jetzt so dass es in meiner Sicht sehr interessant, auf den ersten Blick find ichs schon auch komplex, also wenn man jetzt die Diagramme zum ersten Mal sieht, also mir ist es so gegangen, dass ich am Anfang dachte, oh, das Kreisdiagramm, was sagt mir das jetzt alles, und wenn mans halt dann öfter sieht, dann ist es ja klar, was die Sachen zeigen. Ich finde, man muss sich halt ein bisschen mit dem Thema beschäftigen. Also wenn man jetzt nur zwei Minuten Zeit hätte und das Diagramm sieht, dann ist es wahrscheinlich schon, also dann wär es für mich kompliziert, aber wenn man sich so ein bisschen beschäftigt mit dem Thema, dann kommt man auch rein. Und ich find auch vor allem die Karten recht interessant, wo man das auch wirklich noch auf dem Hintergrund sieht, wo die Verkehrsströme laufen, dass man nicht da..

*..dass man den geographischen Bezug hat..*

**C** ..ja und jetzt nicht nur mehr oder weniger sondern auch sehen kann, wo wollen die Leute hin. Ja. Aber ich hab jetzt nicht viel vermisst, also ich fand das schon sehr allumfassend, dass man alles einstellen konnte und so. Also die Sachen, die wir halt schon gesagt haben, dass man die Zahl am Kreisdiagramm da hin schreibt wo der höchste Punkt ist und dann auch noch den kumulierten Wert hat. Aber ansonsten find ichs schon sehr ausführlich.

*Gibt es was wo Sie denken, dass es zu viele Einstellmöglichkeiten gibt, also zu..*

**C** Ja, das vielleicht eher, aber das ist jetzt mein persönliches Ding, ich bin dann immer leicht überfordert, von dem was man da alles einstellen kann, und die Uhrzeit noch vom ganzen Tag und noch die Kilometer und noch die Kilometer, weiß nicht, was man da noch alles einstellen kann, aber das ist eigentlich auch gut so, dass man halt so feine Einstellungen vornehmen kann. Man muss ja nicht alles verstellen. Also besser, mehr Möglichkeiten und man lässt halt die meisten Sachen gleich. Ich glaube man sollte halt nicht alles verdrehen, sonst kann mans ja nicht vergleichen. Also immer nur einen Parameter vielleicht verändern. Aber sonst..

**A** Ideal ist ja ein System, wo ich viel einstellen kann, aber nichts einstellen muss. Also, was quasi schon voreingestellt ist, ich kann dann aber die Karten gut anpassen, dazu muss ich mich aber auch halt noch ein bisschen reinschauen, das ist nicht alles sofort intuitiv. Ich find die Visualisierungen gut und brauchbar, würde mir zum Beispiel bei den Heatmaps eine andere Skalierung wünschen.

*Ja.*

**A** ..dass man die Bandbreite noch besser darstellen kann, wo dann einfach nicht alles was nicht wahnsinnig hohe Werte sind gleich schwarz wird oder so. Die Uhr ist gewöhnungsbedürftig, man kennt die eher als horizontales System, aber wenn man sich drauf eingestellt hat, kein Problem. Natürlich ist normal die zwölf oben und da ist sie quasi im Westen, aber da muss man halt dann [...] wenn wir Morgen haben dann zeigt er nach oben, abends

finds it very interesting, mentions complexity, needs some getting used to, especially the circular versions.

stresses importance of base map to facilitate geographical reasoning

qualitative information (where) better than just quantitative information (how many), filter options seem complete, daily total should be added, indication of highest value should be placed at bin with highest value

an ideal system is one where much can be specified but nothing has to be specified. Thinks, the visualizations are good and useful, would like a different scale for heatmaps → refine colour scheme for heatmaps, make scale consistent

The circular version need some getting used to but is no problem once one is attuned to it. Once you've grasped it, it is intuitively readable. More legend, e.g. on tooltips would be helpful.

nach unten und wenn mans mal begriffen hat, ist das auch intuitiv lesbar. Ja gut, auch die 24 Stunden, das Ziffernblatt ist vielleicht nicht jedem geläufig aber das muss man halt einmal verstanden haben und dazu ist halt eine Legende sehr hilfreich, also vielleicht an manchen Stellen noch mehr Legende. Oder, was man jetzt heutzutage macht, sind ja dann so Mouse-Over-Texte, dass ich jetzt drüber fahr und dann kommt dann gleich irgendwie eine Erklärungstafel, was seh ich da, was wird dargestellt. Ich würde die OD-Daten grundsätzlich anders darstellen als die Flowdaten, damit man gleich in der Karte immer schon sieht, was das für Daten sind. So muss ich halt im Kleingedruckten schauen oder an der Überschrift. Schaut eigentlich gleich aus von der Darstellung her und zeigen mir doch was anderes [noise] ..noch dazu, wenn sie auch noch gleich dargestellt werden.

*Ja, also was der Unterschied war, den man vielleicht nicht so auffällig gesehen hat, wenn man die Zellen auswählt, dann werden die weiß und schwarz, je nachdem, also die Startzellen werden weiß und die Zielzellen werden schwarz.*

A Mhm, ok, ja. Wenn man die Zellen auswählt.

*Ja, genau. Wenn man sie nicht auswählt, dann..*

A ..ansonsten schauts gleich aus, ja. Ja, dann ansonsten, die Zellen sollte man natürlich, aber kann man ja glaub ich auch, mit der Maus auswählen können. Oder muss man nur reinschreiben?

*Jetzt muss mans reinschreiben, aber ja, das Konzept..*

A ..die Nummer kennt natürlich keiner, also, beim Teralytics Dashboard kann ichs mit der Maus auswählen natürlich auch..

*Genau, ja. Die Idee ist natürlich, das in eine Dashboard zu haben und die interaktiv zu verbinden, aber das ist halt nochmal mehr Programmieraufwand*

A Dann ist natürlich je nach [...] und Räumlichkeit der Darstellung auch eine unterschiedliche Zellengröße sinnvoll, also das ist jetzt hier immer eine Mobilfunkzelle, aber vielleicht gibt's ja auch Betrachtungen, wo man da fünf oder zehn aneinandergrenzende zusammenfassen kann, dass die Karte aufgeräumt wird oder übersichtlicher, dass nur die wichtigen Verbindungen dargestellt werden, also das find ich auch gut, dass man das parametrisieren kann, da war ja der Punkt, wie viel mehr müssens denn sein, bei den Anomalien, dass es dargestellt wird. Aber grundsätzlich fand ich das sehr interessant und schön visualisiert auch. Gefällt mir schon.

*Schön, danke.*

B Ja, ich würde auch noch ergänzen. Also ich fand das auch sehr gut und vor allem die Darstellungen, die Abbildungen fand ich sehr "clean" und intuitiv. Und die beiden haben schon viele Dinge, die ich sagen wollte, erwähnt, ich fand auch die Darstellung mit der Uhr ein bisschen gewöhnungsbedürftig

A visually differing representation of OD-data compared to flow data would be welcomed → *black and white colored cells are apparently not enough, maybe find a different linestyle or try curved lines*

*proposal: interlinking the plots by making it possible to choose cells by clicking on them in the map. → dashboard concept*

*proposal: variable cell size, i.e. different levels of aggregation would make sense, show only the most important relations on the map. → clustering of lines, aggregating major connections into fewer lines, or matching movements to network relations*

generally very interesting, nicely visualized.

tig, also die horizontalen sind in der Regel einfacher zu verstehen, weil jeder kennt das. Aber ich fand die auch interessant, weil ich denke, das ist auch ein bisschen so ein Eyecatcher, wo man so eine Darstellung sieht, und dann liest man das ein bisschen genauer und dann ist das schon, das erreicht das Ziel, was man erreichen will mit der Darstellung finde ich. Und ja, das mit der Heatmap, ja, das ist eine schöne Darstellung, aber daraus kann man nicht viel rauslesen außer so sehr grobe Trends. Ja, interessant aber wenn es um die, wie sagt man, Brauchbarkeit geht, ist es eher weniger im Vergleich zu den anderen. Ja, genau, und das mit den U-Bahn, hatte ich schon erwähnt, dass die Zielzelle immer als Hauptbahnhof zugeordnet wird, das fand ich nicht ideal, aber das liegt halt an dem System.

A Aber warum ist das so?

B Ich denke weil die Tunnel haben die Signalverstärker, die sind immer nur mit der Hauptbahnhofzelle verbunden, das sind nur Verstärker und nicht komplett neu.

A Im ganzen U-Bahnnetz?

B Ja, sieht so aus. In den Tunneln.

A Also alles telefoniert über einen Mobilfunkmasten der beim Hauptbahnhof steht.

*Genau, ja, wenn man in der U-Bahn sitzt.*

A Ok. Wusst ich auch nicht.

B Und was wollte ich noch sagen .. ah genau, dass die Zuordnung von so Quell- und Zielzellen natürlich auch manchmal ein bisschen problematisch ist, vor allem die Fahrten die von außerhalb des Untersuchungsgebietes reinkommen, die werden in der Regel zu den ersten Randzellen quasi als Quellzelle zugeordnet, und das kann manchmal verwirrende Ergebnisse ausgeben. Und andersrum, quasi auf der anderen Seite, das Problem mit Zielzellen ist auch das Gleiche, es kommt drauf an, wie wir gesehen haben, mit der Allianzarena-Zelle, da war auch die Autobahn und wenn man Stau hat oder eine Raststätte hat, dann werden die Fahrten die da nicht enden, werden trotzdem als beendet .. und als Zielzelle zugeordnet und das find ich auch manchmal ein bisschen problematisch. Aber genau, das könnte man theoretisch ein bisschen einstellen, also die.. aber das mit dem einbrechenden Verkehr ist ein bisschen schwieriger denk ich, weil da, normalerweise die erste Zelle ist wahrscheinlich..

*Ja, da müsste man halt das Gebiet größer machen.*

A Ja, das Problem haben wir ja auch beim grenzüberschreitenden Verkehr, wo die Handys halt ins Roaming gehen, dann haben dann alle Ziel Oberaudorf oder Kiefersfelden und dabei [...] fahren.

Finds it very good, mentions clean design and intuitiveness. Circular plot need some getting used to but are an eye-catcher and reach the goal that is to be reached. Heatmaps support only rough analysis of trends but are not as useful → *useful for data overview and general quality check?*

representation of U-Bahn mode (destination is always cell 51) not ideal but due to system.

mentions that assignment of origin and destination cells is problematic in border regions of data extent and with traffic jam situations → *add filter options for OD-data or customize cells in specific areas and match to public transport network and Autobahn*



**B** Genau, ja. Genau, und ich denke das.. ja, und in der Regel, mit diesen Anomalien-Daten-Darstellungen fand ich immer die horizontale Balkendarstellungen einfacher und aussagekräftiger als die mit der Karte.

*Als die mit der Karte?*

**B** Genau, als die mit der Karte. Weil, da waren einfach zu viele Linien und die waren zu groß, zu klein, und dann das mit dem farblichen Unterschied war auch manchmal nicht so intuitiv, aber wenn man die exakten Zellen auswählt und dann die horizontale Balkendarstellung sich anschaut, oder auch diese runde Darstellung, dann wird das sehr schnell sehr klar, also man sieht quasi die Menge, dann sieht man auch schwarz oder grün, das fand ich besser.

*Ja. Da ist ja der Sinn, also der Karten und der Tagesübersicht ist ja, sind ja unterschiedliche Dinge. Also, bei der Karte, auf der anderen, ja, bei der Tagesübersicht sieht man nicht, wo Maxima sind oder so, also die kann man nur auf der Karte natürlich entdecken, aber den zeitlichen Verlauf..*

**B** Ja, in der Praxis würde ich die Kartendarstellung quasi als ersten Ausgangspunkt nehmen und dann mir quasi genauere Analysen, die Zellen aussuchen, und dann diese Balkendarstellung anschauen. Aber ja, das wars von meiner Seite. Aber wie gesagt, ich fand auch super, dass man so viele Sachen auswählen konnte und einstellen konnte, das fand ich sehr gut. Also da, ich hab auch geschrieben, dass ich nichts theoretisch vermisst hab, alles was man braucht war drin.

*Schön, ja. Also dann würde ich vielleicht noch was zu den runden und den horizontalen gleich sagen und noch vorher kurz zur Heatmap. Das ist natürlich, wenn man mehr Daten hat, also wenn man einen ganzen Monat an Daten hat, dann kann man da einfacher Muster auch erkennen. Also wenn es ein Wochenende gibt, das ganz anders ist als die anderen Wochenenden, dann wäre das, also die Darstellung an sich ist einfach für die geringe Anzahl an Daten, die wir haben, also eben nur die eine Woche, natürlich nicht ideal, deswegen, haben wir auch nur am Anfang mal.. ..was man schon sieht, wenn man es sehen will, ist eben dieser Peak am Mittwoch Abend, aber es ist nicht was was einem sofort auffällt. Ja, genau, und zu den runden und den horizontalen, da war meine Überlegung eben, dass der Tagesablauf ja nicht aufhört um null Uhr oder Mitternacht und man hat diesen zyklischen Aspekt von Zeit also von einem Tag, das heißt also eine zyklische Darstellung, also eine Kreisdarstellung, finde ich, liegt näher an den Daten an sich. Was man natürlich nicht so leicht sehen kann, wie Sie auch gesagt haben, ist, ob der jetzt größer ist als der Balken, oder der und der, das ist ganz klar, dass..*

**B** Aber ich finde die Kombination von beiden sehr gut, also.

*Ja, genau. Also..*

**B** ..wenn sie so automatisch generiert werden, dann gerne beide.

*Ja, ok. Was mich halt auch interessieren würde ist, wenn man kurz auf die Darstellung*

Finds bar charts (horizontal or circular) to be simpler and more informative than maps for the display of anomalies. Too many lines, too thick or too thin → *normalization on maps needs to be refined, introduce clustering of lines or matching to network to tidy up map*

Sees map as starting point and would do more detailed analysis with temporal overviews for specific cells.

shares thoughts behind horizontal and circular version

combination of horizontal and circular version is welcomed, keep both if possible

*schaut, entweder die runde oder die horizontale, und sagt, was passiert zwischen sieben und neun. Dann ist mein Gedanke, dass das bei den kreisförmigen, dass man das schneller findet, also wenn man die kennt weiß man, acht Uhr ist immer da oben irgendwie, und horizontal hat man ja nicht so den Ankerpunkt. Also beim Kreis die Richtung ist immer die gleiche, wenn ich hier [horizontale Darstellung] sag, zwischen acht und sieben ist halt die siebte und achte Säule, dann muss ich erstmal mich orientieren und sagen, ok, hier ist null und muss dann rüber gehen. Ich weiß nicht, ob Sie verstehen, was ich meine.*

**A** Jaja, klar, sonst müssen da [horizontale Darstellung] im Prinzip auch Hilfslinien wieder eingezeichnet werden, im drei- oder mindestens sechs-Stunden-Rhythmus, dass man sich besser orientieren kann. Vor allem wenn der Nullpunkt nicht null ist, sondern jetzt zwei hier..

*Ja. Also wo genau der Peak ist, denk ich, kann man vielleicht an dem Kreisdiagramm schneller erkennen. .. Genau, das war so die Überlegung, deswegen wollte ich die beiden mit einbringen.*

**A** Ja. Die haben auch beide ihre Berechtigung, das ist sicher. Ich glaube, die Mehrheit wird die horizontalen bevorzugen, weil man die halt kennt. Das heißt aber nicht, dass die andere Version schlechter ist oder so, sondern einfach unbekannter. .. Das kennt man von Himmelsrichtungen eher oder von qualitativ anderen Diagrammen, als zeitliche Verteilung, auch mit dem 24-Stunden-Muster, aber wie gesagt, man muss es halt einmal erfasst haben, dann kommt man damit auch zurecht.

**B** Aber vor allem in diesem, in der mittleren Abbildung [Vis 4, middle version] zum Beispiel, da finde ich, da sieht man meiner Meinung nach..

**A** Ja, in der mittleren ist die horizontale klar besser.

*Ja, da wäre natürlich auch, also eine andere Möglichkeit wäre gewesen, den Kreis innen größer zu machen und das dann auch nach innen laufen zu lassen. Aber dann kriegt man Probleme, also dann ist der Kreis nicht immer an der gleichen Stelle, und das wollte ich jetzt vermeiden in dem Fall.*

**A** Naja, man kriegt auch von der Fläche dann eine falsche Vorstellung, weil das "private" macht doch drei oder vier mal so viel aus wie "public" und bei dem Kreis fahr ich ja flächenmäßig zusammen [...] und da kommt dann eine falsche Vorstellung rüber.

*Ja, das ist auch das Problem ein bisschen vielleicht bei der Nummer drei [Vis 3], das sind eben keine Flächen, die die Zahlen wiedergeben, sondern die..*

**A** Was ich mit dem Auge als erstes erfasse sind ja Flächen, bewusst oder unbewusst, und da ist der Kreis halt nicht so gut wie die Säule.

*Ja, genau.*

both the horizontal and the circular version are relevant, expects majority to prefer horizontal version because it is better-known

horizontal version clearly better in the case of public and private modes (Vis 4, middle version) → find a way to incorporate split into circular version without skewing ratio due to changing height-area ratios

A Links [Vis 4, left side] find ich haben die beide, sind die gleichwertig und hauptsächlich Geschmackssache, aber in der Mitte finde ich auch, dass das horizontale klar besser ist.

*Ja. Und dann kommt es natürlich drauf an, was man rauskriegen will. Also ob man das allgemeine Muster sehen will zwischen, ich kann das noch mal holen [Vis 4, interactive], da find ich eben ganz schön am Wochenende..*

A Oh! Ja.

*..ja, sieht's eben so aus. Oder wir können auch einfach den Sonntag nehmen. Da ist so ein Muster, und am Montag ist das Muster eben extrem anders. Das heißt, wenn man das schnell sehen will, dann ist das glaub ich, könnte es intuitiver sein als die horizontale Version und wenn man halt genaue Werte sehen will und wirklich, ja, was genaueres sehen will, dann .. man sieht das schon auch hier aber es ist..*

A Ein bisschen problematisch ist halt, dass sie immer auf 100% also immer, egal welche Fallzahlen ich hab, es ist immer gleich hoch das Diagramm. Ob das jetzt drei Fehlerfälle sind oder dreizehn Tausend .. da muss man halt auf die absolute Zahl schauen..

Finds it problematic that all plots have equal size and therefore different scales. → *visually indicate total amount of movements shown in a given plot*

*..ja, das könnte man natürlich auch noch..*

A ..und mit dem Kreisdiagramm, das nicht so bekannt ist, da hab ich erst mal zu tun, mich zu orientieren, wo find ich welche Stunde, was ist alles dargestellt, dann geht die absolute Zahl der Fälle halt ein bisschen unter.

*Ja, das könnte man im Prinzip auch noch in die Mitte irgendwie legen, dass man da einen Kreis mit der absoluten, ja.. genau. Ja, aber wie gesagt, wenn man das halt festlegt, dann verschwindet das meiste einfach und manche Sachen werden groß... Genau, dann hab ich eigentlich alles gesagt, was ich sagen wollte. Wenn Sie nichts mehr haben, dann..*

C ja schön.

[...]

## User Test II

*Ja, also wenn es auf Englisch, wenn es die lieber ist oder euch, dann*

E Ja, gefällt mir ehrlich auf Englisch aber du brauchst nicht..

*Ja, ok, wir können auch einfach mischen, und du sagst, wie du möchtest.*

E Ja, wie du willst.

*Ok, genau. Also ich hab ja nach den Interviews Visualisierungen vorbereitet mit den Daten dich ich von dem Projeet habe, von dem ich erzählt habe, dem xMND-Projekt, da geht es eben darum, dass man Mobilfunkdaten nutzbar machen will für urban public transportation, für den öffentlichen Verkehr, für die Planung, genau. Und jetzt hab ich ein bisschen was vorbereitet, also so eine Art Szenario, wo wir in vier Schritten durchgehen. Das hier ist ein Überblick über die verschiedenen Visualisierungen, den habt ihr auch vor euch. Dann habt ihr so einen Fragebogen, da könnt ihr wenn ihr wollt gerne währenddessen schon anfangen, das obere auszufüllen und euch Notizen machen, aber auch einfach reden. Den könnt ihr euch kurz anschauen. .. Und wenn es Fragen gibt, dann einfach melden.*

E Du hast mit [R-Studio?] diese Graphen gemacht oder?

*Mit python in Jupyter Notebooks. Aber das seht ihr gleich. Grundsätzlich hab ich zwei verschiedene Typen von Daten, einmal Flow-Daten, das sind die auf der oberen Hälfte. Das heißt das sind Bewegungen zwischen Zellen. Also immer wenn sich ein Handy von einer Zelle in eine andere bewegt, also ummeldet sozusagen, dann wird das registriert. Und die sind vorbereitet und nach einem Modell werden denen verschiedene Modi zugeschrieben. Und die Modi sind eigentlich hier privat, Rail U-Bahn, Tram, Bus. Dann ist die public die Summe der vier unteren und moves sind alle zusammen. Das heißt Rail ist Schienenverkehr, also S-Bahn, Fernverkehr, Regionalverkehr..*

D Ok, also Oberfläche?

*Oberfläche, genau, wobei Tram nochmal extra erkannt wird. U-Bahn ist U-Bahn, Bus ist Bus und privat ist alles was da nicht reinfällt. Also Auto, zu Fuß, Fahrrad, alles was nicht erkannt werden kann. Das Modell ist nicht perfekt natürlich. Ihr werdet gleich sehen, dass es das schon relativ gut erkennt, also hier sieht man den Privatverkehr, da sieht man die Autobahn ganz gut. Das hier ist der öffentliche Verkehr, aber das ist nicht 100% natürlich.*

D Also die Zuordnung.

*Ja, genau*

D Da hat man eine Fehlerquote einfach

*Ja, genau. Wobei die Fehlerquote halt auch noch nicht bekannt ist. Also das Modell ist noch in Bearbeitung. Wir haben bei jedem Schritt immer eine Kartenansicht. Oben gibt*

es immer Filter, die verändern sich ein bisschen. Wobei man den Modus auswählen kann, den Tag oder den Tagestyp, Montag bis Donnerstag, Freitag und Wochenende. Die Tageszeit kann man auswählen und ob es normalisiert werden soll oder nicht. Normalisiert heißt in dem Fall, dass.. ..also hier ist es nicht normalisiert, da sieht man einfach die absoluten Zahlen, also die Dicke der Linien entspricht der absoluten Zahl. Und normalisiert bedeutet, dass ich die Dicke davon abhängig mache, wie die Zahl zu den gewählten Filtern von dem Durchschnitt über die ganzen Daten abweicht. Genau. Und die ganzen Daten heißt, das ist eine Woche im März 2019, der 11. bis 17. März, das heißt das sind vergleichsweise wenig Daten aber mehr war schwierig, auch wegen Corona, zu handeln. Genau, dann haben wir im ersten Abschnitt jetzt diese Heatmaps als eine weitere Visualisierung. Da seht ihr die Wochentage von oben nach unten die Stunden von links nach rechts. Das heißt jedes Kästchen ist eine Stunde und je stärker das eingefärbt ist, desto mehr Bewegungen wurden da registriert. Und das ist auf den ganzen Bereich..

D Wie groß sind diese Zellen, sind die unterschiedlich groß oder..?

Also die Zellen sind Voronoipolygone also ..

D mhm, ja..

.. um die Funkzellen rum.

D Ahja, ok.

Das heißt auch hier, da sieht man die Zelle und dann eine Linie immer vom Mittelpunkt der Zelle. Das heißt nicht, dass die alle vom Mittelpunkt kommen, sondern dass die halt aus der Zelle kommen.

D Ja genau, ok, mhm.

Hier hab ich die Bewegungen über die Tage summiert, also diese ganze Woche, und ihr seht die verschiedenen Modi in verschiedenen Farben. Grau ist wieder privat, dann grün ist Schienenverkehr, wir nennen das jetzt einfach mal S-Bahn. Lila ist Bus, Blau ist U-Bahn und rot ist Tram. Da kann man auch auswählen, ob man nur die öffentlichen sehen will oder nur die privaten oder alle zusammen. Und dann gehen wir weiter und schauen uns den Tagesverlauf an. Und da gibt es immer eine kreisförmige Darstellung und eine horizontale. Bei der kreisförmigen ist es so.. ..oder fangen wir mit der horizontalen an, die kennt man wahrscheinlich besser. Genau, da fangen wir um zwei Uhr nachts an und hören um zwei Uhr nachts auf, weil der Tagesablauf ja nicht um Mitternacht endet. Und im Prinzip ist das das gleiche wie bei dem hier [Vis 3], die sind einfach übereinander gestapelt. Man kann den Tag und den Tagestyp auswählen und ob man beides sehen will oder nur öffentlich oder nur privat. Und hier sind die öffentlichen nach unten angezeichnet und die privaten nach oben. Und bei dem Kreisförmigen ist null Uhr links. Also nur zur Orientierung, weil wir ja 24 Stunden haben und nicht zwölf, also es ist nicht wie bei der Uhr. Genau, das heißt die Morgenstunden sind hier oben und die Nachmittagsstunden hier unten. Und sonst ist es das gleiche wie bei dem horizontalen. Die Zahl hier oben ist das Maximum vom Diagramm. Das Diagramm bleibt immer gleich groß, einfach weil die Unterschiede sonst zu groß wären. Und die gleichen Zahlen gibt es hier [horizontale Version] auch. Gibt es zu den Darstellungen schon Fragen? Sonst

würden wir anfangen.

E So I couldn't understand exactly how can I ... how can I find my data with this kind of graph [Vis 2]?

Ok, so the stronger the colour, the more movements.

E Ah. Yeah but..

*It is more for an overview, so it's not.. ..or what is the question?.. So if it's very light and very colourful, then there is a lot of people moving in that means of transport.*

E Ok, thank you.

Vis 2

*Ok. .. Dann könnt ihr anfangen und hier.. oder wir fangen rechts an. Das sind die heat-maps eben, das ist im Prinzip das gleiche, was ihr gerade gesehen habt. Da könnt ihr schauen, ob euch schon irgendwas auffällt, und einfach erzählen.*

D Also was wir daraus lesen?

*Mhm. Ihr könnt auch hoch und runterscrollen.*

weekend clearly visible

D Ja, ok. Also quasi, was man immer deutlich sieht ist das Wochenende, hier.. .. Das waren alle oder?

*Ja, das waren alle, das ist ein bisschen dunkel jetzt auf dem Beamer.*

D Man sieht immer dass die Morgenspitze, also die Morgenspitze sieht man ganz gut und dass eben, das ist die Beobachtung die wir auch immer haben, dass die Nachmittagsspitze ein bisschen breiter ist, die Hauptzeit. .. Auch dass es sich beim Bus ein bisschen länger verteilt.. Da haben wir auch viele Schüler drin und so Sachen, das sieht man dann auch..

identifies rush hours on mornings and afternoons, compares their qualities to own observations

*Gibt es sonst irgendwelche Auffälligkeiten, die man nicht erwarten würde vielleicht?*

D Hmm.. Also das heißt, hier sieht man jetzt nicht, also, da ist Verkehr aber man erkennt nicht zwingend.. also gar nichts ist ja wahrscheinlich nicht oder?

*Ja, genau, das ist ein bisschen..*

D Ja der Unterschied zum Privatverkehr, gerade eben hier in den Zwischenstunden und am Wochenende..

*Dass da ein bisschen mehr ist..*

tries to compare different modes → colours are not catered to comparison, would have to be adjusted, e.g. same colour palette within plot, indication of mode outside the plot. Scale should be consistent for comparison.

D Also dass quasi, wenn in den öffentlichen nicht mehr viel los ist, dass halt dann genau zu den Zeiten aber dann privat mehr los ist. Aber auch trotzdem

ist auffällig, dass also wenn beim ÖV die Spitzenstunden sind, dass auch im Privatverkehr die Spitzen zu sehen sind.

Ja

D Ja, also es ist nicht nur.. also privat heißt ja nicht Freizeit oder? Ja aber genau, beim privaten sieht man auf jeden Fall noch mehr Freizeitverkehr oder die Zeiten zwischen den Spitzenstunden, während beim ÖV die Spitzenstunden sehr deutlich werden. .. Ok something that you can read out of this, E? Do you want to add somethign?

E No..

Ok. Sollen wir auf Englisch switchen einfach?

D Ja, wenns geht, gerne glaub ich oder?

E Ja.

Vis 1

Ok. Alright, then we'll take a look at the maps. Which means you can choose some filters here and then run interact. And then you'll see a map.

D Ok. Do you want to do?

E No, you do it.

D Ok, let's see.

So once you choose a day, the daytype doesn't play a role.

D Ah ok. And if I don't choose a day, then.. ..it's the average.

Yes.

D Ok.

So here you can also see the cells and the extent of the data

D Ok, and if I.. ..ok so the difference is.. ..not too big, I assume.

E Can I check some things for Bus Nacht? Because I know that now we have, for [Tram?], problems in this area. I wanted to check in the night there.

uses map to check very specific area → resolution good enough for large-scale analysis?

D Yeah, from 22..

E ..and..

D weekend or..?

E No, on the weekend, we have a U-Bahn.

D But not in this.. this is the [..?]

E Yes, I wanted to check how many people use I mean nowadays the Busline.

D During the week or on the weekend?

E During the week.

D I think Monday or Tuesday as day would be good then.

E And day, I should change it or?

D Leave it clear, yes.

E Ok.. ..

*You might want to uncheck "normalize", sometimes, it becomes very thin..*

D And if you move to the line with the mouse, is it.. yes. 29 moves

E Ok so here we have problems or..?

D This doesn't show.. this shows where people use the bus between.. so we can't read it.. so it doesn't tell us if we miss something because now people are just using what we offer. And you see that there are a couple of people. 13, 6, 15.

*I should mention that these numbers are not extrapolated, so it's only the Telefonica-users. So it's probably, like a third of the people..*

D Ok.

*You can also look at the private movers, if you want to see what you're missing.*

E Ah ok.

D Ok, still a lot of people using..[..?]

E And here also..

D Yeah, that's not like all of it but some of them could be potential public transport users. ... I think, in the inner circle, it could also be bike.



*Yes, bike, or even walking.*

D This is interesting.. ..do you know what happened, was there a football match maybe?

*Yes! Very nice, that's what I wanted to hear [laughs] So right now you have the data from Monday through Thursday. Do you think you can find out, when the match happened, at what day? .. Ok so now you chose Monday for the day.*

*Finds anomaly without any guidance! May be supported by great familiarity with the city's structure.*

D Ah, do you mean the day or the..

*The day, yes.*

*finds time of anomaly*

D Yes, you always see the airport, too.. .. and something. Oh!

*Yup.*

D And you see, even.. Ok I think then you could also see the U-Bahn, right?

E I'm sorry, actually, I don't know what's happening in this area?

D Allianzarena, the football match.

E Ah ok. And when do we usually have a football match? On Wednesday or..?

*This was a Championsleague game, so, I don't know..*

D They can be on any day..

E But it's interesting for me because usually, when I talk to people in, yeah in Deutschland, and then actually, I found for such an evening, they prefer to use public transportation because they want to drink alcohol and something like that but.. I'm surprised.

D I think it starts at nine o'clock, something like that, right?

*Could be, yes. You can also check "normalize" and then you see the difference to the mean of all week. It's a little less but you can still see it..*

D Yeah, ok. Yes, this is what football does to our U6 [laughs]. .. Ok interesting. And this is just between seven and..

E But this data ist based on which year exactly?

*2019, that is just one week in March.*

D One week in March, ok.

*Alright, so that's wonderful that you found it on your own, I didn't need to say anything.*

E Yes, interesting!

Vis 2

*Now, maybe we go back to these. Can you find it in the heatmaps, too, the match?*

D Hmm.. ..it could be this one.

E Honestly, I can't recognize it. Or I can't find the relation between.. So the first one is clear for me, about public and not public graph. But when we talked about Buslinie and at the same time I wanted to look at this graph, but I couldn't find different.. I don't know..

*differences within the graph?*

E Mhm, I recognize, ok, between twelve and 24 we have more demand..

*dissatisfying to not be able to tell quantities but only rough patterns*

*..But it's hard to see quantities.*

E yes.. I don't know if it's possible to change, whether there are such options or not. But usually, when we have more high demanding and so, then we use dark colour, ok? But in this graph, it is different. When we have more demand. And so.. Do you understand?

*Yes, so there is also this option [light colour palette], where it's the opposite.*

D Ahja.

E Yeah. I think it actually makes sense, at least for experts, because they are used to see when the colour is a little bit darker and then ok..

*finds it easier to identify a darker colour with a greater quantity. → use light colour palette on light background and change colours so they work well on a light palette.*

D We have more demand.. And also you see what I mentioned earlier, that when it is black, or dark, then you have the impression that there is no traffic at all. And here you see a little bit.

*Yes. I mean, it basically depends on the background and now we have a white background, so this might make more sense.*

D Yes, I think that maybe, it's like here this is the peak from people going back.

*finds small peak on Wednesday night*

*..people going back, yes.*

D And so they are somewhere here where they go to the Allianzarena. This is darker than on the days before.

E That's so interesting!

Vis 3

Ok. so here is the weekly overview, I'm not sure if we have to.. I mean it's just a quick plot. Can you relate that to the football game in some way or..?

D Yes.. ..so normally I think Tuesday is the day with the highest movements and I was actually, as I saw it, I was wondering, why here we have Wednesday with the most movements, so it makes sense. And especially also that Thursday is smaller than Friday. I was wondering about that.

*switching between different visualizations (temporal and spatial overview) helped with sense making*

Vis 4

Ok, yes. Maybe we can check that, too. Ok then we have the daily overview. So this is all the days but you can change the filters a bit and see if you can recognize the event..

E That's interesting because, usually, between seven and nine, we keep high demanding [..?] but here..

D It's public and private, right?

Yes, exactly.

D Ok, maybe it's clearer if you see, private is higher in the afternoon, probably, and yes, you see a peak until nine..

*compares result to expectations, notes disproval of expectation*

E Yes.. ..but usually we consider seven until nine. But it shows us, people prefer to use public transportation between six until eight.

D Yeah, because a large amount are the pupils to the school and they are before eight o'clock. So the traffic is still a lot also between eight and ten actually with a lot of people who are working. But the school, like all the, you see the bus is really high here, that's all the, I think it's more the students going to school. I don't know if it's.. Ok, yeah, ok here you see also the..

E Mhm.. Monday?

*So the Monday was also the first day after a school break, so maybe that's what makes a difference, too..*

D So everybody, that.. yeah.. ..this is very cool

E Yeah, so impressive. Like you can track, where there are people and..

*And you can also, if you uncheck the circular option, you can also see the horizontal version, you might be more used too.*

D Ahja, ok, mhm. Ok, I think it's.. if you want to check both [public and private], I think this is actually quite interesting, because that you already have like a modal split, something like that. Of course, in this case it is a little bit different categories, because you have private there and stuff like that and if you just look at the same ones..

*finds split visualization interesting (Vis 4, middle version)*

E And how do you get both options?

*The first thing that doesn't say anything. Yes, I could have called it "both"..*

D I don't know if you.. what would interest me as well is the kilometers, how long do people travel. Do you have something like this?

*filters for travelled distance wanted (→ OD-data)*

*Mhm, yeah, we're getting to that later.*

D Ahja, ok. Because this is, ahja, ok, this is when they change from one cell to another, right?

*Yes, exactly.*

D That is one count. Like, we.. here, the number says that it counted..

*..363 thousand people..*

D ..moving from one cell to another..

*..moving somewhere in the network, in the extent..*

D It's the cellphones that are counted

*Yes. So it's cellphones moving, exactly. .. Right, I forgot one thing about U-Bahn, that looks a bit different than the others, the other modes. Do you have any idea why? That everything goes to Hauptbahnhof? ... This is actually due to the fact that the network in the U-Bahn-tunnels is all going through the cell at Hauptbahnhof. So you have.. Verstärker, or, what do you call that.. like if your phone logs into the network and you are in the U-Bahn, it's all through the cell at Hauptbahnhof. So once you step into the U-Bahn and go underground, your phone "is" at Hauptbahnhof so to speak.*

D Just in the city center or everywhere?

*I think everywhere where tunnels are. So.. ..I mean that makes it easy to identify U-Bahn, because you jump a long distance, but it also means that people going from here to here might not actually go all the way to Hauptbahnhof but might get off here, but the phone goes to Hauptbahnhof and then back here. So that's something that the model will tackle later.*

D Ok so you have to like clean it..

*U-Bahn Problem: further data-processing necessary*

*..yes, you have to reconnect the movements. But the way the phone takes is through Hauptbahnhof. .. Alright. Then maybe we can move to the next step.*

[ recording stopped and resumed ]

Stage 2

*Alright. The next stage is that now that we have identified that on Wednesday there is*

something happening, we take a look at the anomalies of Wednesday. And now we have two or three different kinds of visualizations. So we also have a map view again, now the colours are a bit different, we have green and black. Green means that there is more movement on Wednesday and black means there is less movement on Wednesday with respect to the other three weekdays, so I took the mean from Monday, Tuesday and Thursday and the value for Wednesday, for each hour, and then calculated the difference. And now in the filters you can again choose the mode, and the colour of the mode will show up in the cells, so now here the cells are red, which means it's Tram. And you can choose the time of day and the minimum anomaly. So if you only want to see movements where there is at least 50 people more or less than on the other days, then you can check that here. This is something to control the width if it gets to crazy. And then if you choose relative, it's the relative anomaly. So if there is a hundred people usually moving on Monday, Tuesday and Thursday, and now we have a hundred and ten, if you don't check it, it shows a thickness of 10 and if you check it, it shows that it's a tenth more. You'll see that if you don't check it, sometimes movements in the outer parts pop up, just because there are no movements at all before and then there is a lot more. And then on the time-side, we also have the horizontal and the circular version, and now you can choose the cells that you want to look at. So right now, cell 51, it's the one at Hauptbahnhof, you can see that if you hover over the cell [on the map], it gives you the number. You don't have to choose them, then it sums up all the cells, but if you want to, then you can look at traffics between different cells. So you can also choose more than one cell. So this is right now people coming from Hauptbahnhof and going anywhere else. And on the left, you can see that direction and on the right is the other direction. So here it's on the right people going to Hauptbahnhof from anywhere else. And on the horizontal plot, you can see the same thing. And one direction is the upper one, so the one from Hauptbahnhof is on the upper half and the one to Hauptbahnhof is on the lower half. Yes, and if you don't choose any cells, it's just one plot or one half, because it sums over the directions. ... Is that clear, more or less?

D Mhm

Ok, then we can take a look at that. Do you want to start with the map or the others?

E Of course.

The map? Ok. Then you can choose the filters and..

E [..?]

Alright, so now you chose all the lines where at least 10 people more or less travel by bus on Wednesdays compared to Monday, Tuesday and Thursday, throughout the whole day.

E .. Ok interesting. This line is..

No, that's the cell.

D It's the cell. So if there is no line, it means that the traffic is the same as the rest of the week.

Vis 5

chooses map as starting point

Exactly.

D .. but, it's not, it is always Wednesday right now, right?

Yes.

D Ok.

*So now relative might be.. ..and now you can also change the minimal anomaly, to..*

D Ok, so even, also the private moves are..

*..much higher, yeah.*

D ..that's interesting. A lot of people from the west of Munich are going to see the game, right?

*Yes, there might be a different reason, that they are coming from there.. do you have an idea? .. So if you zoom out a bit.. ..yes, the network, or the extent of the data stops, right, so..*

D ok, so they come from the Autobahn, they enter..

*..exactly. So they register, once they enter the network.. yeah. So for the private moves you can see the streams coming from the Autobahn pretty well.*

D Ahja, here as well..

*Ah yes, so the width thing only works if you don't choose relative.*

D Ahja, ok.

*But you can do that.*

D Even, so from the other U-Bahns you have more on Wednesday. So that will never be smaller. .. But you don't have any black lines as well. So the whole, here, the whole network has a little bit more..

E [..?] in VISUM?

D VISUM is always a simulation and this is real movement data. Yeah. Ok.

*Alright, then we can look at the time. Or maybe you want to choose some cells that you want to look at?*

D Ok, yeah, maybe we choose Münchner Freiheit, I think that's interesting. Cell 14. Because we have a U-Bahn and the.. Tram. Ok and this is from cell

*ambiguity of start cells leads to misjudgement → custom cell configuration / network matching / refined border cell treatment has to be introduced*

Vis 6

to, .. so this is to 14 and this is..

*..from 14, exactly.*

D Even if you don't enter 14 here.

*Yes. That's, the left side is always to 14, so if you put, I don't know, Hauptbahnhof, you have 14 to 51 on the left, and 51 to 14 on the right.*

D Ok.

*..and if you don't put anything here, it's 14 to anywhere and anywhere to 14.*

D Ok. To 14, from 14.

*You want.. what?*

D So this is from 14 and this is to 14.

*Yes. Exactly.*

D Ahja, ok. you can see more Tram to Münchner Freiheit..

*Right. So I'm not sure, I explained the plot properly. So the grey values are the ones, the means from Monday, Tuesday and Thursday..*

D Ja.

*green and black bars in anomaly plots seem to be intuitively readable*

*..did I say that? I don't know. And the green is the anomaly, so..*

D Yeah. and here we have more traffic on Wednesday..

*..yeah. And the black thing means less.*

D ..to Münchner Freiheit than..

*Ja.*

D And if we do the same for the U-Bahn, it should be somehow.. ..ja. Ok. That's interesting. .. Ok, all moves in the morning on Wednesday are a little bit less than..

*ja.. ..it might be different when you compare it to other Wednesdays, because now, probably the Monday is more.. or weighs down the other.. ..and if you want to look at the horizontal plot you can do that, too.*

D Mhm.

*Right. And maybe to go back to our "case" and look at the Allianz Arena, so cell 2 and cell 71 maybe. And then at U-Bahn, you can see the peaks.*

D Ja. You can see when they came and when they went back as well.

*Exactly. So now we have, left is going to Allianz Arena and right is coming from there. ... If you want to check the movement between Hauptbahnhof and there..*

D And Münchner Freiheit again, I think that was 14?

*Between there and the..*

D the Allianz Arena

*So that is only eight or eleven total, so there is.. ..ja. I mean you can see there is a lot going on but..*

E Yeah, so actually between eight nine and ten when they go to..

D Mhm.. ..what's the total? The difference or..

*The total is the highest entry. So here it's on Wednesdays a peak between six and seven. So eight people [laughs] travel between there.*

D Ok, yeah. There are obviously some..

*So I guess that's.. ..where is Münchner Freiheit.. Here, yeah.. Alright, so here is no line connecting them.. so for this specific case it's not many people.. so I'm not sure..*

D Mhm.. Ah ok.

*Maybe if you take that instead, Giselastraße.. ..yeah, it's also not many people.. ..here it's more than a thousand, so.. ..Alright, is there anything more in these two graphs you want to explore or should we go to the next? Ok.*

[ recording stopped and resumed ]

Stage 3

*Alright, so the next step is, now we look at OD-Data, so origin-destination-data. Which means, it tries to track a person from where to where they are going. The origin and destination of the trip. I have to say that the data is still cropped into hours, so it's only trips between each full hour, due to limited capacity at the model side. But yes, so now we can look at the map again, and we are back to the overview, so it's not anomalies at the moment, we're going to look at that in the end. But you can choose start and end cells and the mode, the time of day. Now you can also choose the minutes travelled and the kilometres travelled. And I checked, the maximum was like 58 or something, so that's why the maximum is 60km here. Alright, and then for the temporal overview, we again have circular and horizontal versions and you can basically choose the same filters. And it's stacked again. So here the example is Rail, so that's why it's green, but*



*it's the same colour coding as here. Alright. Here we go. Map again first or..?*

Vis 7

E Yeah.

*Ok.*

E Days and this..?

D Yeah, you can choose whatever you want. If you choose day, then the day-type is already..

E Ah yeah, ok.

*Yeah. Maybe leave those filters in the beginning, because it might get very thin. The minutes and the kilometres I mean. And just look at the total first. .. Time of day you can change.*

E Ok. .. Startcell..

D ..do you know where the start is and where the destination is?

*When you type it into startcells and endcells, you can choose a start cell.*

E I want to take a look at..

*..81? Ok. .. so now you can see all the traffic coming from that cell.*

E Mhm..

D Can you write 63? ..

E Can you choose [..?]

*Alright, so now we have the small or short movements, up to 10km for private in the mornings.*

D Ok so these would also be potential.. some of them

*Potential bus users?*

D Yeah, or.. ..maybe not bike users..

→ further splitting within mode "privat" might help with analysis

*Yeah. Feel free to talk and tell us what you're doing.*

D Ok so, because until five kilometres, you can walk or bike easily. And so people who use the bike for longer distances or the car for distances that are, I don't know .. ah ok .. I think 15 is already a lot.

E With bicycle, yeah.

D With bicycle? I don't know..

*Already takes an hour I guess almost.*

D So people still bike or use the car. They could also be using public transport.

E But let's take less, until eight kilometres..

D Yeah, the difference is not so high. But you see there is still a lot of people moving even in the city center. Ok. This is something where I think they go by car. Ok. If you enter start cell, this as a start cell I think. .. Ok most likely, they are on the car already from the Autobahn, right?

*Yes, that's what I would guess, too.*

D So a lot have.. ..Sendling. Mhm.

*So, the fact that they stop in that cell means that they stay in that cell for a long time, I don't know, maybe half an hour, I don't actually know the exact threshold. But can you think of some reason that they stop there other than that they actually want to go there? When they are coming from the Autobahn onto the Mittlerer Ring?*

*ambiguity of destination cell  
→ further filtering / custom  
cells / network matching*

D That's a traffic jam..

*..a traffic jam, yeah.. I think so. It could be an explanation to why they all stop there.*

D Because it's not famous for all the office space here, that's all living, housing.. I think here, somewhere in the south.. here is a lot of office buildings and industrial..

*destination cell for rush hour  
traffic doesn't line up with  
context knowledge*

*Yeah. Do you want to look at that cell? So movements that end in 58. Now I think we still have the private movements, right?*

D Yeah. .. Yes that's from the main station

*so the biggest line..*

D So they come from somewhere.. so you could also say they interchange here and then..

*Yeah, so since it's private, I'm not sure if that's mostly unrecognised public transportation users, because if you start at Hauptbahnhof, I'm not sure how much sense it makes to move privately to Siemenswerke..*

D Ja.. ja.

*But you can compare it to the public transportation maybe..*

D Mhm..

*Hm ok.. no..*

[laughing]

E Ok interesting..

D Because there is a S-Bahn, maybe we can try the S-Bahn..?

*Yeah, public has S-Bahn, too, though. So public is the sum altogether*

D Ah, it's the sum of the others, ok then it doesn't make much sense.

*It should be... hm that's interesting, what does it say? 53 moves.. hm, ok. Maybe I wasn't careful enough. Alright, do you want to take a look again at the football game example and see if you can figure out where people are coming from there?*

D You know the cell, right?

*Yes, so Allianz Arena is two, and then 71 is next to it and then you might want to choose Wednesday. Which mode do you want to look at?*

E Public? Oder Privat.

D Let's do public.

*Ok. I'll switch those back on. And then maybe the evening, right? .. Ah, and now it's startcells, so people coming from the stadium publicly.. ..and if you put it into endcells.. ..yes.*

D Ah but.. ..why are there two lines that are meeting in..

*..at Hauptbahnhof? Because I chose two end cells. So one is going here and one is going here..*

D Ah ok ok. These are two. Ok.

*You can also I think .. ah, nevermind. Now you can look at the private movements, too.*

D So now the movements from here are not in it..

*Ja...let's see.*

D Ah maybe they arrived earlier.

*→ clustering close lines for a clearer picture or match movements to network*

*Ja. Yeah, now you see a few of them. And if you hover, maybe they are actually pretty thick but the ones at Allianz Arena are so huge..*

**D** But now they are [...] so they are a little bit earlier than...and here also.

**E** Yeah. And what is here?

**D** [...]? So that's when they come on the Autobahn, then that's the cell where their phones log in, so they come somewhere from the north. Ja

*Ja*

**D** Ok.

Vis 8

*Alright, then maybe we can take a look at the distribution throughout the day again, so we choose the same end cells, 2 and 71. And once they..*

**D** Ah ok. So it's private, bus and U-Bahn

*Yeah.*

**D** Because there is no S-Bahn.

*And it's 15 thousand maximum. So if you want to look at only the public ones, maybe it gets a bit clearer.*

**D** So this should, if we just set end cells, it should be the same as before, number four, right?

*yeah... well, movements that go through these cells and have a different start and end point don't show up here.*

**D** Ah ok. It's just the ending point.

*If it actually ends there, yes.*

**D** Ok, ja.

*And the other way around.. so coming from the stadium looks like this.*

**D** So this has a little bit more information, not more information, but it's a little bit cleaner. Because the other just say how much movement is in the cell, but if you want to know how many are actually going to the Allianz Arena, or to the cell the Allianz Arena is in, and ending there, this is a little bit more exact.

*evaluation of data types  
OD-data deemed a little bit  
more exact*

*Yeah, you filter out the ones that just pass through, yes. And here you also have the horizontal version.*

D Mhm.. hm.

*Alright. Shall we go to the last stage?*

D Mhm!

[ recording stopped and resumed ]

Stage 4

*Alright. So for the last one we do a similar thing as before, we look at the anomalies of Wednesday, but for OD-Data. So again, origin-destination, and it's basically the same as before. So black means it's less than on the other three weekdays and green means it's more. And you can choose start and end cells and the time of day and the minimal anomaly. Ja. And you can choose between circular and horizontal. Do you want to start with the map again, or..? Ok.*

Vis 9

D Ok. Yeah, I think let's enter the Allianz Arena again.

*Ok. At start or end?*

E End?

*Do you want to change the time of day?*

E Yes, let's look at the evening.

*Mhm. And now we have all moves. .. so what do you see?*

D Hauptbahnhof [laughs] This is also.. It's a parking space.. ah, a hotel.. ok..

E So maybe you should put privat. Then we can understand this..

D It's just in private. So.. ..maybe that's also a parking space..

*Yeah, might be. Or again a traffic jam. So it's recognised as their starting point.*

D Ah ok, yeah. Public is coming from the city center.. Mhm.. [laughs]

E [laughs] [..?]

D [..?] ..the U-Bahn .. That's interesting because, in the flow data, we have a change in Tram as well and..

*..even at those cells?*

D Like in the whole city there was more traffic in Tram..

*Yeah, so if you look at all cells again, that might be.. because now we only had the*

*Allianz Arena as end cells*

D Ah, right, right, true.

*Yeah, so there they are*

D Ah ja, ok, yeah.

E Can you change [..?]

*Mhm.. yes, so you can have points with less.. ..yeah..*

*looks for specific case*

D yeah.. interesting

E [..?] ..interesting.. The Tram 15, I'm not exactly sure about the line, but we have not so.. ..yeah, here.

D What was it there?

E Do you know which line? 15?

D So what's the point?

E Yeah but so we don't have high moving between here and..

D It's the difference from Wednesday to the other days.

E Ah ok.

D So if there is no line, it basically means that there is the same traffic as on the other weekdays.

*You can look at the absolute again here [Vis 7] so weekdays and.. the evening or?*

D I think it was the evening. .. ahja. But you see here, that's the normal traffic.

E Mhm..

D And this is just the difference. .. But we don't have information about if we end here, they interchange to U-Bahn?

*No, that would be, so for the OD-Data, the mode is the majority mode. So the mode that they use at most of the time. So it might be that they use the Tram for a bit and then the U-Bahn for the rest, so then..*

D So then they get U-Bahn

*Exactly. So the main mode is the mode that they are assigned to.*

D So you have the travel chain in one transportation mode. And not from walking to Tram, to U-Bahn..

*Yeah, no, that would require more work in the model I guess.*

D Yeah and I think that's always the difficult part how you get movement data that are go along with Datenschutz..

mentions complication  
due to data privacy issues

*..the privacy and stuff, yeah. Once you only have one person moving in that way it's hard to show it.. Right. Then we have the temporal overview as a last thing. Maybe we go back to the Allianz Arena case. And you can play with the modes and circular and stuff if you want to. So on the left you see starting at the Allianz Arena and on the right you see going to Allianz Arena.. ..and again, it's the differences between Wednesday and the other weekdays.*

Vis 10

D Mhm. Public is really clear but private is actually quite similar. We have like between seven and nine, coming and going, here it's a little higher.

*So again, now the grey values are the normal ones and the green is the anomaly so it is clear that there is a lot more on Wednesday, but yeah, it's not as pointed.*

D But also in both directions..

*Yeah.*

D Ok, ja, interesting. .. Something you want to try, E?

E So..

D Yeah, we don't have Tram in the start cell.

*Yeah, that's very small numbers so there is just one more [laughs]*

E And what is this?

*This means, there is one more movement during the night.. I mean it might just be..*

D But it's strange because we, I think we don't have a Tram in the..

*Yeah. So I think it might be a misassignment*

D Ah, ja, ja.

*The other movements are much bigger.. Alright, anything else you want to go back to, or..*

E no..

D no, I'm fine.

*Ok. Then I would ask you to fill out the questionnaire. You can do it in English or German, whatever you want. On the front page there is the numbers of the visualizations, just make crosses where you want them. And on the back there is some more questions. And then when you're done, we can talk about it a bit more.*

[participants fill in questionnaire]

[ recording stopped and resumed ]

*Then we can have a short discussion, just you telling me your overall thoughts, anything you want to mention..*

Discussion

off topic

E So actually can I start with the questionnaire?

*Yeah, I mean, we don't have to go through it all, because I'll read it..*

E No, no I mean this kind of questionnaire..

*Oh, yeah, ok.*

E About your.. your researching. So I think this kind of questionnaire it's like a personal view and maybe, it should be a bit.. ..so because.. .."do you feel you can trust the data", something like that.. so actually I, basically, I trust you and I trust this kind of information, ok? But .. I have a little bit problem with this kind of question, do you understand what I mean? Ok, if you want to publish your paper, for I don't know, some high quality and something like that, I'm pretty sure, this kind of question, it's, they don't.. ..so because, we don't analyse based on quantity, quality. Usually we want to talk about quantity and something like that. Do you understand what I mean? But because I mean, I have researcher experience, and when I propose a questionnaire based on this kind of word exactly, so they say no, we don't approve such a questionnaire, you know? It's just a tip that if you want to invite other.. or.. and maybe, I think it's better you change this, because..

D What would you suggest?

E Yeah, because, do you feel, you can trust the data and something like that.. If I want to say, ok, do you use this kind of visualizations for your paper, so, do you know, it's a different kind..?

*Yeah. So for quantitative and qualitative analysis, we have, I only have you and three other people, so there is no way I can make a quantitative analysis. So the questions in the beginning are more to get you thinking about the visualizations, and..*



E ..because, do you know, this kind of graph, is usually used for paper. And it's so approve your analysis and your data, then you want to publish, based on my experience, but.. ..and.. another ..[..?] and so and maybe it's actually as I asked you about too simple and too complex, maybe you should, I don't know, add some sentence about what you think, because.. do you understand what I mean or you don't..

*Like how to understand complex and simple.. yeah.*

E Because maybe complex is sometimes for paper good but sometimes it's not..

*It's not well defined, you mean..*

E Yeah. And but .. ah ok.. and about your.. and also about .. so .. I know, when you want to offer some new [kind?] or new planning or something like that, and usually, we take, we have a look about Wednesday or at pm and am. And so actually, we do also use data based on day and .. but so.. ok so maybe.. two days are some special examples, where people change their behaviour, and so, as you say, ok now, maybe Monday people prefer to go, I don't know, university or school or something like that and during week, they stay at home. I'm not sure, because actually last year, last week, I read articles about how we should prepare our city based on such a happening in the world and something like that. And say, ok, actually, it's ok, usually, we analyse based on what we have always about I don't know, about planning, about determining strategy and something like that. But with such a huge problems, everything changes and we couldn't track people, how their behaviour, and something like that. Maybe we should change or behave, or yeah.. ..it was interesting for me because I thought we would actually get to see, ok, when people, for example, we consider between six until nine, there is a high demanding for people to take the bus line and something like that but and also in March when I checked the topics..

D But it's 2019, it's last year, in march.

E Yeah, I know, but because we don't know how we go in the future based on this corona virus and so on and maybe some people say ok maybe we should consider in two years this same situation.. ..but yeah, maybe.. so.. and it is.. I can't understand because I think this kind of data, do you know R Studio or not?

*Know what? R Studio? Yeah.*

E So this kind of information and actually is based.. ..I thought that it's based on this R Studio, ok? But I think.. limitation data, is a bit difficult to analyse, do you know?

*The fact that it's not a lot of data? Yes, yes, yes, of course.*

E It has an effect actually.

*Yeah, yeah.*

E But we can't do anything, also I know, but.. maybe you can just mention, I know this kind of limitations in .. research and something like that..

*So, the data preparation is done at Fraunhofer, that I do the project with. Or, they are in the project and I am doing my Master Thesis with them. And they have to aggregate the data because I am not allowed to use it in a raw form, because of privacy issues. So I don't have, like I can have requests and tell them what I need but it was not possible to process more data at the moment.*

E But so for paper, I mean you usually mention in the end of paper, ok, I know I have such problems and after that you just approve yourself, ok I know such a [..?] data..

*Ok. Do you want to say anything about the visualizations?*

E Ok, I wanted to say, maybe if after two months when I use the circular version of the visualizations, it will be clear for me but I'm used to the horizontal ones, do you know? But because of this, I think, how can I judge such a [..?] because now..

*..you need more time to get used to..*

E yeah, I'm used to show my data or understand other data horizontally.

*Yes. Do you see any reason to get used to, would you choose to try to get used to the circular one or would you rather stick to.. like is there anything you think might be easier to read from the circular one and that would be a reason to get used to it? Or do you think the horizontal one is fine and you'd want to stick to it?*

E Good question. So I think it really depends on the person because sometimes you [..?] ..some new things but some people like me they like to stick to what they are used to seeing, yeah, I don't know, I mean it's a personal.. do you understand what I mean?

*So you think it's a personal.. yes.*

D I think you could get used to the circular if you look at just for example one traffic mode. Because it's like a clock where you have a feel for the day. You have a feel for the time of the day and you very quickly see where is the bubble, like where do people accumulate and on the other hand, the horizontal is like, ok the day ends at this point and here you get the feeling, that it's a time continuum, especially if you have an average, you see for example the step from two o'clock to three o'clock, like where you have to jump back, and so I think that's very convenient, that you see the whole day as a continuum. But if you want to compare, either the directions, what we had with the de-

expects to need two months to get used to circular versions

thinks that choice of horizontal or circular option depends on personal taste  
Proposes reduced complexity for getting used to circular version.  
Feels that circular version facilitates intuitiveness ("feel for the day") and understanding of temporal distribution. Likes that it represents the continuity of a day's cycle. Thinks the horizontal plot better facilitates comparative tasks (directions in Vis 6 and 10 and modes in Vis 4 and 8, middle version)

stinations, or compare the mode, I think it's more practical, if you have it on one, on a horizontal plot.

*Yeah.*

**E** So, but, basically I agree with you because, ok, when we see some circle, it's the first things that we imagine, it's some hours and something like that, ok, and so when, and firstly I saw ok, this circle divide means that we always have to be always [..?] and when I pay attention and then I see ok no, you divide the circle based on..

*..24 hours.*

**D** Yeah, you could also like turn it and make the zero on top to that it's like this half is day and this half the other half of the day, so like, that it's.. ..but I don't have a problem with reading, I find it really good. You see the afternoon and the morning and, yeah.

no problem reading  
circular version

*Yeah. .. So maybe I share what the idea or..*

**E** But

*Once you're done*

**E** But because this one, the horizontal, it's, you show until two am but this one..

**D** Also. The two is..

*So the fact that it shows between two and two is because I.. so for the hours between midnight and two o'clock in the morning, I attributed the day before. So what happens on what is officially Thursday..*

**D** ..because the U-Bahn [..?] one o'clock and then..

*Yeah. And also the IDs, the trip-IDs, get reset between two and three every second day or something, so.. and also of course the public transportation system doesn't stop at midnight, but a bit later. So that's why I also show it in that way in the horizontal one. But yes.*

**D** No, I think, both have some information and one is clearer to compare and the other one is clearer to read the development of the day.

relevance of both plot  
versions, cater to different  
tasks (comparative and  
analytical)

*Yeah. nice. So that was also one reason why I wanted to see what you have to say about both of them because I feel like if you look at a bunch of different ones, as you said, in the circular one, it's very easy to spot eight o'clock, once you get used to it, you know exactly where it is. If you have the horizontal one, you always have to count. Unless you have like a.. you could put a help line or something in, of course. And the general*

*shares thoughts behind  
horizontal and circular  
versions of plots*

*pattern I also feel like is more clear in the circular one, but then if you actually want to measure differences, the horizontal one might be more helpful.*

*proposal: match movements to network*

D Yeah. .. And I think what is very interesting is that it's a first step to get movement data. And I, ok yeah, it's maybe one.. I don't know if in the future it's maybe also a possibility to try to match them to a U-Bahn-Line. Because you know you can see it but then you don't say, they are moving between these lines, but you also put them on the right U-Bahn. You know what I mean?

*You mean on the map?*

D Yeah.

*So you..*

D ..if you say, ok, they are travelling U6.

*Yeah, ok. So you match it to your network. Mhm.*

*filter options are important*

D And.. ..so I think it's really important, that you can choose the day and the daytype and also the hours. For planning reasons, especially the graphics [Vis 4], I think we need higher..

*..resolution?*

*higher temporal resolution would be needed for planning*

D higher resolution. Because for the, I think at least ten minutes, that would be useful, because then you can, we have to react like for a single U-Bahn for example, you know, ok, this U-Bahn is too full, we have to react with this one. And for the Bus, we already have as I told you, the data for every single line. So I think to use it like really for planning, we need a higher resolution. But what I think is also very interesting is that you can take a look at the anomalies. And not in the model but how are people moving like in reality because we always have events that are coming again, like football, like fairs and that.. ..we always have a network, a plan, for how we drive in these times but here you could compare it and say, ok, is the plan we have, does it match reality, and also what you said, with the corona-virus, we could track the developments. Is there really for example, you compare March to April, to May, and then see, is there a slow development and in which direction the development goes.

*anomalies from "real" movement data are valuable for movement analysis on special days (fairs, football events, Christmas, New Years) and to understand trends, e.g. in conjunction with a pandemic*

E So what do you think, do you think it is useful for short time or.. ..long time.. I think

D ..for short events that are coming again or that you want to analyse, for example we always want to analyse Christmas and New Years because it's like completely different and on the other hand it's hard to find people that are doing the countings on these days, because it's holidays. So this is an example I would think of and say, ok, let's see how people really move on

these days that are like so unique that it doesn't make sense to.. offer the same as on other days, because people use it completely different. So I think that would be very convenient to have these comparisons to like normality and the anomalies. [..?]

E And is it possible to add an option based on month?

*If you have the data, of course, yes. But the whole project has data for four weeks I think, the whole of March, four or five weeks. And for my work I only have one week. Because the basically, so because of corona, they all work from home as well, so they don't have the same computing capacities at home and they can't.. manage larger amounts of data. And also it takes a lot.. I mean, even the data I have now took some time to process. For the visualizations.*

E So this programme, I couldn't understand, what is the name of the programme?

*The programme? I'm just using Jupyter Notebooks, so basically..*

D Python is the language.

*Yeah. So I write it in python and I can change stuff and execute it here. But it's, of course it's not a dashboard. But I know that X is using a product from Teralytics, and they are using Telefonica data as well. And they provide a dashboard with a map and some graphs, not as complicated, they are a bit simple, but, yeah. .. So the concept of course would be that if you click here, something changes here, and you have them interactive in like the different visualizations interact with each other basically. But yeah, this is the "raw" version*

E In general, yeah, it's so interesting.

D But that's also the next problem, if you would have all the data, all the time, how and where do you process it. Do you keep it and is it manageable, so that you get the information you need in the amount of time that makes sense. So you always need like a month to analyse the data from two months ago..

*So I think for the [Teralytics product], they have always the data from, at least month. So I'm not sure they have data for August already but for July, they already have data. So I mean, one you automate it, it works quickly. .. So that would be also something, you'd like to see, so very up-to-date data? Or..*

D ..It's hard to say because we don't have up-to-date data but we have I think also with a delay of.. ..ok right now we have up-to-date data because of the corona.. we have, we actually track every day. But usually I think it's also always a month later that they checked all the data for mistakes and errors and stuff like that and the clean it and then we have the.. ..so on our server we have the realtime data but it has to be checked. .. I think the up-to-date data, we would just need if there is a very special occasion, like an anomaly, like a football game or what's also interesting, if there is a huge construction

mentions data management challenge

up-to-date data useful for very special occasions like construction sites.

site or something like that, where we have a change in movement and then to..

*..like are the measures working that you put in place.*

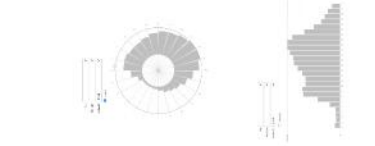
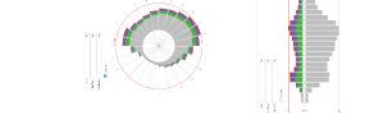
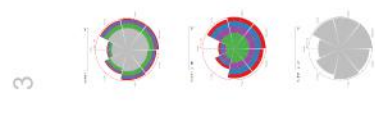
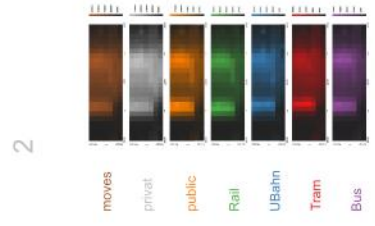
**D** Yeah.

*Alright, then if you don't have anything else to say or to add..*

**E** It's so cool [laughs]

*I'm glad. Ok.*

# Überblick – räumlich und zeitlich



# Anomalien

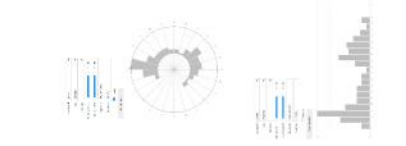


6.1 6.1

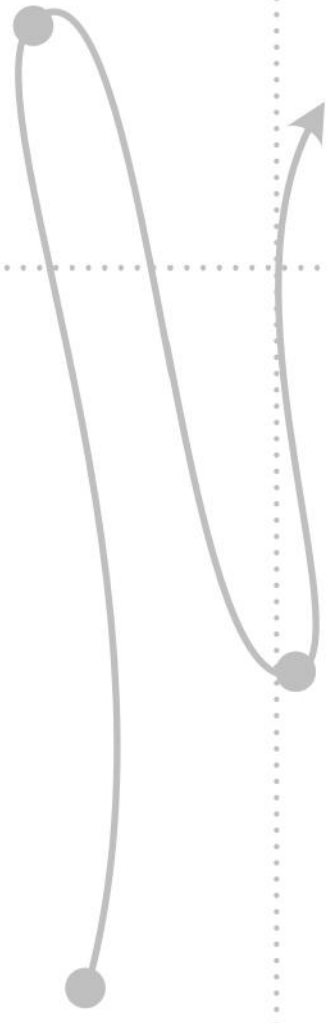
# OD - Daten



8.1 8.2



10.1 10.2



# Fragebogen

## Questionnaire

Visualisierung Nr. Visualization nr.	1	2	3	4.1	4.2	5	6.1	6.2	7	8.1	8.2	9	10.1	10.2
<b>Wie brauchbar finden Sie diese Visualisierung?</b> How useful do you deem this visualization?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Wie bewerten die Komplexität dieser Visualisierung?</b> How do you rate the level of complexity of this visualization?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Wie intuitiv bewerten Sie diese Visualisierung?</b> How intuitive do you think this visualization is?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Haben sie den Eindruck, dass sie den gezeigten Daten vertrauen können?</b> Do you feel you can trust the data shown?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Notizen

### Notes



## Fragebogen Questionnaire

Was gefällt Ihnen an den Visualisierungen?  
What do you like about the visualizations?

Was vermissen Sie? (Funktionalität, Filterparameter)  
What are you missing? (functionality, filter parameters)

Welche Verbesserungen würden Sie sich wünschen?  
How would you want this visualization to be improved?

Beim Lösen welcher Aufgaben oder Fragestellungen aus Ihrem Arbeitsbereich könnten diese Art der Visualisierung helfen?  
What kind of questions or problems within your field of work could this type of visualization help solving?

Bevorzugen Sie die runde oder horizontale Versionen der Visualisierungen 4, 6, 8 und 10 oder denken Sie, dass die beiden Versionen unterschiedlichen Zwecken dienen? Warum? Welche Zwecke fallen Ihnen ein?

Do you prefer the circular or the horizontal version of the visualizations 4, 6, 8 and 10 or do you think both serve different purposes? Why? Which purposes come to mind?

## Questionnaire Results

### Was gefällt Ihnen an den Visualisierungen? What do you like about the visualizations?

- A Räumliche und zeitliche Verteilungsmuster werden gut aufgezeigt Spatial and temporal patterns of distribution are well demonstrated
- B Sehr "clean" und intuitiv very "clean" and intuitive
- C guter Überblick der Verkehrsströme good overview of traffic flows
- C Vergleichbarkeit von Verkehrsmittelnutzung und Tagen / Uhrzeiten Comparability of transport use and days / times
- D Klarer Farbencode, übersichtlich, gutes Maß an Informationsgehalt Clear colour code, clearly laid out, good level of information content
- E I find it interesting and it is kind of visualizations use for paper and it will be use to appove your date to analysen I find it interesting and it is kind of visualizations use for paper and it will be use to appove your date to analysen
- E It will be useful for determining strategy and developing public transportation in short time. It will be useful for determining strategy and developing public transportation in short time.

### Was vermissen Sie? (Funktionalität, Filterparameter) What are you missing? (functionality, filter parameters)

- A Unterliederung der privat-Verkehre nach Verkehrsmittel (Auto, Rad, Fußweg) Breakdown of private traffic by means of transport (car, bicycle, foot)
- A Skalierbarkeit, zB der HeatMaps, Ausrichtung der Uhr Scalability, e.g. of the HeatMaps, alignment of the clock
- B siehe Notizen see notes
- C - -
- D Unterscheidung Fuß / Rad / MIV Distinction foot / bike / motorized individual transportation
- D Intermodalität, Kleinere Intervalle Intermodality, smaller intervals
- E I think we should consider paramter km in graph 4.1 and we can recognize how many km people prefer to take public transportation or private I think we should consider paramter km in graph 4.1 and we can recognize how many km people prefer to take public transportation or private
- E I am not sure, focusing on data just based on Wednesday or special day. It will be useful or not. I am not sure, focusing on data just based on Wednesday or special day. It will be useful or not.

### Welche Verbesserungen würden Sie sich wünschen? How would you want this visualization to be improved?

- A OD-Daten sollten nicht auf die gleiche Weise dargestellt werden → Unterscheidbarkeit OD data should not be presented in the same way → distinguishability
- A andere Darstellung der Spitzenstunde (Zahl an der h) different representation of the peak hour (number on the hour)
- A noch mehr Legende oder Erklärungen über Mouse-over-Texte even more legend or explanations via mouse-over texts
- B Vorschlag: Der höchste Wert soll mit der entsprechenden "Zeit" oder "Tag" in der Abbildung

dargestellt werden Proposal: The highest value should be shown next to the corresponding time or day within the plot

- C Bei Diagrammen: Wert für gesamten Tag anzeigen, nicht nur "Spitzenstunde" With charts: Display value for entire day, not just for the peak hour
- D Darstellung von Wegeketten Display of travel chains
- D Stundenintervall ist ok, für Planung wäre mind. 10 min-Intervall brauchbar hourly interval is ok, for planning at least 10 min interval would be useful
- E Adding option based on month Adding option based on month

**Beim Lösen welcher Aufgaben oder Fragestellungen aus Ihrem Arbeitsbereich könnten diese Art der Visualisierung helfen? What kind of questions or problems within your field of work could this type of visualization help solving?**

- A Verbundraumerweiterung → Modal Split, Verkehrsmittel, SW-Analysen Network expansion → modal split, means of transport, SW analyses
- B Pendler Verflechtungen, Nachfrage Abschätzungen commuter interrelations, demand estimations
- C Untersuchung von Verkehrsbeziehungen für bestimmte Tage: Vergleich der Verkehrsmittelwahl, Bestimmung wichtiger und stark nachgefragter Relationen (zB für Kapazitäten in Bus & Bahn) Investigation of traffic relations for certain days: Comparison of the choice of means of transport, determination of important and highly demanded relations (e.g. for capacities in bus & train)
- D Analyse von Verhalten bei Großveranstaltungen / Sperrungen / ... Analysis of behaviour during major events / closures / ...
- D Potenzialanalyse für ÖV Potential analysis for public transport
- D OD-Analysen OD-Analysis
- E Finding which area has potential to increase our service in future Finding which area has potential to increase our service in future
- E Based on private function, we can realize which are used to increase service based on time Based on private function, we can realize which are used to increase service based on time
- E determining strategy for short time determining strategy for short time

**Bevorzugen Sie die runde oder horizontale Versionen der Visualisierungen 4, 6, 8 und 10 oder denken Sie, dass die beiden Versionen unterschiedlichen Zwecken dienen? Warum? Welche Zwecke fallen Ihnen ein? Do you prefer the circular or the horizontal version of the visualizations 4, 6, 8 and 10 or do you think both serve different purposes? Why? Which purposes come to mind?**

- A Dient dem gleichen Zweck und ist Geschmackssache. Ich bevorzuge die horizontale Darstellung Serves the same purpose and is a matter of taste. I prefer the horizontal representation
- B Eher die horizontale Darstellungen, sind einfacher zu verstehen I prefer the horizontal version, it is easier to understand
- C Auf den ersten Blick sind für mich die horizontalen Visualisierungen leichter zu verstehen. Runde Diagramme sind jedoch "kompakter" und liefern - nach etwas Einlesezeit - auf einen Blick aussagekräftige Infos. Evtl. für Fachzeitschriften, für Leute, die mit der Materie vertraut sind. At first sight, the horizontal visualizations are easier to understand for me. Round diagrams, however, are more "compact" and provide - after some getting used to - meaningful information at a glance. Possibly for professional journals, for people who are familiar

with the subject.

- D Für Analyse einzelner Modis oder Modal Split → runde Version. Sehr übersichtlich. Horizontale Version für Vergleiche zwischen Modis gut geeignet, Vergleich Hin-/Rückrichtung ... For analysis of individual modes or modal split → round version. Very clearly arranged. Horizontal version for comparisons between modes well suited, comparison of both directions ...
- E I prefer the horizontal version of the visualizations, I think it was more clear to compare at a same time and I get more information regarding time and I think circular version has limitation to show service or moving between 12 am until 2am because basically it should be clock. After our discussion I understood that you divide your circulare 2am until 2 pm  
I prefer the horizontal version of the visualizations, I think it was more clear to compare at a same time and I get more information regarding time and I think circular version has limitation to show service or moving between 12 am until 2am because basically it should be clock. After our discussion I understood that you divide your circulare 2am until 2 pm

### Notizen Notes

- A Probleme am Rand des Umgriffs mit ein-/ausbrechendem Verkehr Problems on the edge of the extent of the data dealing with inbound/outbound traffic
- B Flow Daten mit der U-Bahn → Wegen der Zuordnung der Hbf-Zelle immer als Zielzelle, ist die Darstellung eher unbrauchbar bzw. verwirrend. Flow data with U-Bahn → Because of the assignment of the main station cell as target cell always, the display is rather useless or confusing.
- B Generelles Problem mit der Feststellung von Quell-Zelle oder Ziel-Zelle. zB die Fahrten, die außerhalb des Untersuchungsgebiets rein kommen werden zu den Randzellen zugeordnet. Andererseits die Fahrten die auf der Autobahn stattfinden werden teilweise zu der Allianz Arena Zelle als "Ziel" zugeordnet. General problem with the determination of source cell or destination cell, e.g. the trips that come in from outside the investigation area are assigned to the border cells. Also, the trips taking place on the highway are partly assigned to the Allianz Arena cell as a "destination".
- C --
- D zu Vis. Nr. 2: sehr intuitiv, wenn hell zu dunkel (weißer Hintergrund). Bei höherer Auflösung Anomalien leichter erkennbar. about Vis. No. 2: very intuitive with a light background. Anomalies are easier to detect at higher resolutions.
- E --

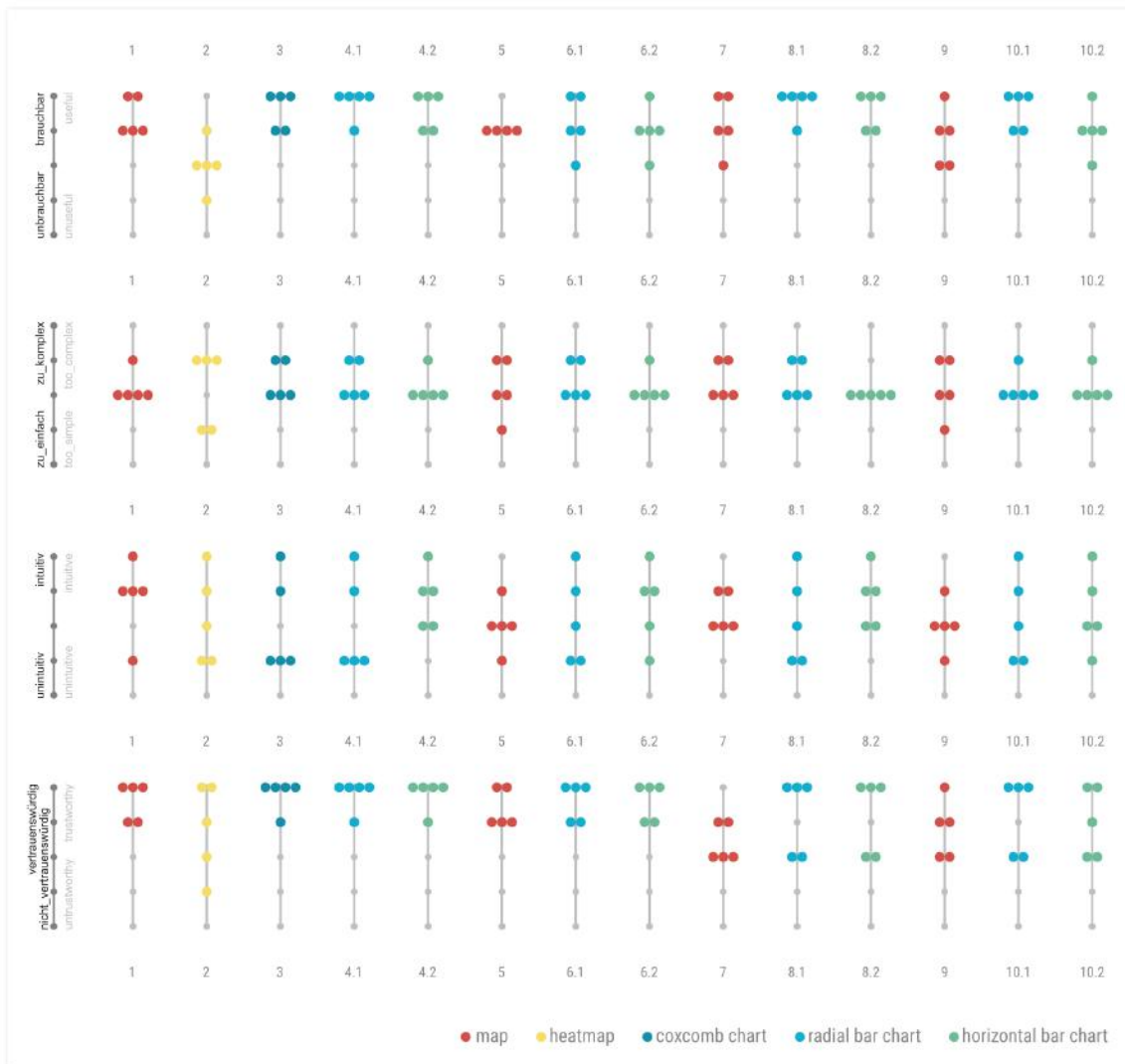


Figure A.8: Answers to the first part of the questionnaire.