



# Cartography M.Sc.

## Master thesis

User-centric geospatial analytics  
and visualization of catastrophe  
risk portfolio using the example  
of the insurance industry

Lin Zhu



2020

# **User-centric geospatial analytics and visualization of catastrophe risk portfolio using the example of the insurance industry**

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

Author: Lin, Zhu  
Study course: Cartography M.Sc.  
Supervisor: Dr. Mathias Jahnke (TUM)  
Reviewer: Dr. Paulo Raposo (ITC)  
Cooperation: Lars, Haider (Allianz Global Corporate & Speciality SE)  
Cosmin- Razvan, Tanasescu (Allianz Global Corporate & Speciality SE)

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

Date of submission: 18.09.2020

## **Statement of Authorship**

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

“User-centric geospatial analytics and visualization of catastrophe risk portfolio using the example of the insurance industry”

I have fully referenced the ideas and work of others, whether published or unpublished.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

Munich, 18. 09. 2020

Lin, Zhu

## Acknowledgment

---

The completion of this thesis cannot have been possible without the support, participation and encouragement of several people who I would like to express my sincere gratitude to.

First, I would like to thank my first supervisor Dr. Mathias, Jahnke from TU Munich who is always available for my questions and provides me valuable research guidance. My exterior supervisors Lars, Haider and Cosmin-Razvan, Tanasescu from Allianz Global Corporate & Speciality. Thank you for your great support, guidance and constructive suggestions during my thesis period. Especially the assistance on ensuring my thesis can progress smoothly under the COVID-19 pandemic.

I would like to thank all members from the Cartography MSc. Program for your companionship during these two years. Especially our coordinator Cron, Juliane from TU Munich. Thank you for your kindness and always be willing to provide your support for any problems. Additionally, I would like to thank Prof. Dr. Liqiu, Meng from TU Munich for her encouragement in the beginning of my thesis period when I was stuck in the topic. Thanks my reviewer Dr. Paulo. Raposo for providing your critical and valuable feedback and comments during my thesis period.

I would also like to thank my colleges from AGCS. Thanks Dr. Hannes, Roemer for your participation on discussions and meetings. Thanks other members from different departments for your valuable feedback and time.

In the end, I need to express my gratitude to my families who encouraged and supported me come to Europe to pursue a master degree. Thus, I got the chance to experience such a wonderful journey with wonderful people in last two years.

## Abstract

---

Nowadays, visualization techniques, especially with Geo-spatial extension are playing a significant role in communicating the information between the dataset and end users from multi scales and multi variate aspects. For supporting the spatial data exploration process efficiently, visualization strategies implemented in a Geo-visualization application need to be designed based on the dataset and targeted users. Thus, there is no universal solution for all visualization problem. Proposing a comprehensive visualization strategy is still a challenge work.

Point data is one of the typical spatial data types which indicates the certain location in the world. This thesis research focuses on the large point dataset and proposes a visualization strategy with a hierarchical visual prototype considering users' information appetites for exploring the dataset in different granularities. To verify the effectiveness of proposed visualization strategy, this thesis implemented the proposed visualization strategy into the insurance industry using the catastrophe (CAT) portfolio dataset provided by Allianz Global Corporate & Specialty (AGCS).

The workflow is guided by the concept of user-centred design which is started with assessing users' requirements and ended with user test. In this thesis, a visualization strategy with multi-resolution global hexagon layers is proposed for supporting users to explore the large location datasets and assisting their decision-making process. In the studied case, the proposed visualization strategy is applied for the CAT portfolio dataset and realized into a web-based visualization application. The application can be used for facilitating two main working scenarios for users from AGCS. In the end, the usability of the visualization strategy in this thesis also be evaluated via testing the application.

**Keywords:** Geo-visualization, tessellation, spatial binning, Data Aggregation, Risk Management

# Table of Contents

---

<b>1. Introduction</b>	<b>1</b>
1.1 Background and Research Motivation	1
1.2 Problem statement from studied case	2
1.3 Research Objectives and Questions	3
1.3.1 Research Objectives	3
1.3.2 Research Questions (RQ)	3
1.3 Hypothesis	4
1.4 Innovation	4
1.5 Thesis Structure	4
<b>2. Literature Review</b>	<b>6</b>
2.1 User-centred design	6
2.1.1 User-centred Design in Cartography	6
2.1.2 User requirement assessment	8
2.1.3 Usability testing	9
2.2 Multi-scale map visualization	10
2.2.1 Multi-scale Visualization based on grids	10
<b>3. Methodology</b>	<b>18</b>
3.1 Research plan	18
3.1.1 Workflow of thesis project	18
3.1.2 Visualization strategy	19
3.2 Conducting User-centred Design Loop	20
3.2.1 Methods for User Requirement Assessment	20
3.2.2 Methods for User Testing	21
3.3 Multi-scale Map Visualization	22
3.3.1 Hexagonal binning	23
3.3.2 Interaction	26
3.3.3 Visualization of hexagon layers	27

<b>4. Case study</b> .....	<b>29</b>
4.1 Background and Data.....	29
4.2 User-centred Design .....	30
4.2.1 User interview.....	31
4.2.2 User Requirement Assessments .....	32
4.2.3 User Test .....	34
4.3 Hexagon System.....	34
4.3.1 Hexagon creation .....	35
4.3.2 Visualization of Hexagon layer.....	37
4.3.3 Result Comparison .....	38
4.4 Platform Design.....	42
4.4.1 Software used .....	42
4.4.2 Platform layout .....	42
4.4.3 Widgets .....	43
4.4.4 Scenarios achieved in the application .....	44
<b>5. Result</b> .....	<b>48</b>
5.1 Thesis Application Result .....	48
5.2 Results of User test .....	48
<b>6. Discussion</b> .....	<b>54</b>
6.1 Answers to Research Questions.....	54
6.2 Hypothesis Verification .....	56
<b>7. Conclusion</b> .....	<b>57</b>
7.1 Outlook .....	57
7.2 Research work and findings.....	57
<b>Appendix</b> .....	<b>65</b>
Questionnaire.....	65
Questionnaire for the first round user test survey.....	65
Questionnaire for the second round user test survey.....	66
Code (take resolution in 10,000 km <sup>2</sup> as an example).....	67

Hexagon Creation .....	67
Hexagon Index .....	68
Assigning values .....	68
Hexagon Searching.....	69
Combining Values .....	70
Hexagon in Sphere (R script).....	70



## List of Figures

---

Figure 2.1-1: Usability engineering model .....	7
Figure 2.1-2: User Requirements Analysis(Maguire & Bevan, 2002) .....	8
Figure 2.2-1: Example of DGG in geographic coordinate system (Kevin Sahr, Denis White, and A. Jon Kimerling) .....	11
Figure 2.2-2: Example of Geodesic DGG (Kevin Sahr, Denis White, and A. Jon Kimerling)..	12
Figure 2.2-3:The shape of binning grid.....	12
Figure 3.1-1: The set-up of the project.....	19
Figure 3.1-2: the structure of visualization strategy .....	20
Figure 3.2-1: The workflow for user requirement assessment .....	21
Figure 3.2-2: The workflow for user test .....	22
Figure 3.3-1 :Comparison of shape distortion in Web Mercator projection coordinate system .....	26
Figure 3.3-2: Zooming for Multi-resolution hexagon layers .....	27
Figure 3.3-3: Differences caused by various classification methods ( <i>Data Classification</i> , 2018) .....	28
Figure 4.2-1:The procedure for user central design.....	31
Figure 4.2-2:Workflow of the thesis project .....	31
Figure 4.2-3: graphical scenarios .....	33
Figure 4.3-1: Neighbour hexagon layers' relationship .....	34
Figure 4.3-2: size and spacing of Hexagon .....	35
Figure 4.3-3: Hexagon geometric relationship .....	36
Figure 4.3-4: Hexagons along the world edge.....	36
Figure 4.3-5: Hexagon index system .....	36
Figure 4.3-6: Index matrix.....	37
Figure 4.3-7:The experimental range (1) and portfolio dataset located in the range (2).....	39
Figure 4.3-8: Experimental Hexagon layers.....	39
Figure 4.3-9: Result based on Hexagon Layer on Web Mercator Coordinate System.....	40

Figure 4.3-10: Result based on Hexagon Layer on the sphere.....	40
Figure 4.4-1: Layout design .....	43
Figure 4.4-2: Interface of Application.....	43
Figure 4.4-3: Functionalities of Widgets.....	44
Figure 4.4-4: Select interested metrics and layer scale.....	45
Figure 4.4-5: viewing risk accumulations on different levels .....	45
Figure 4.4-6 Detecting changes of metrics in quarters .....	46
Figure 4.4-7: Process on Manual Calculation .....	47
Figure 4.4-8: Process on Situation Awareness .....	47

## List of Tables

---

Table 2.1-1: Methods for user requirement analysis .....	9
Table 3.3-1: Hexagons Information created use ISEA43H(Sahr, 2018) .....	24
Table 3.3-2: comparison of different methods .....	25
Table 4.1-1: Main columns of location table.....	30
Table 4.2-1: User profile .....	32
Table 4.2-2: Interview Questions .....	32
Table 4.2-3: Summarization of scenarios .....	33
Table 4.3-1: Colour Scheme.....	38
Table 4.3-2: Layers' visible scale.....	38
Table 4.3-3:Result Comparison .....	41
Table 4.4-1: Processes and tools in the research .....	42
Table 5.2-1: Result of first round questionnaire .....	49
Table 5.2-2: Result of the second-round questionnaire.....	50
Table 5.2-3: Comparison between two applications.....	52

## Abbreviation

---

AGCS	Allianz Global Corporate & Specialty
AAL	Average Annual Loss
CAT	Catastrophe
CATlas	A Web-GIS platform for catastrophe analysis
CS	Consecutiv Storm
DGG	Discrete Global Grid
DW	Data Warehouse
EQ	Earthquake
FL	Flood
GIS	Geographical Information System
ISEA	Icosahedral Snyder Equal Area
ISEA43H	Icosahedral Snyder Equal Area hexagon system created by 4&3 Aperture method
KML	Keyhole Mark-up Language
LoB	Line of Business
MR-tree	Min/Max R-tree
NetExp	Net exposure
RICost	Reinsurance Cost
TIV	Total insured value
UCD	User-centred design
WGS84	World Geodetic System 1984
WS	Windstorm

# 1. Introduction

---

Geo-visualization techniques contribute to an improved understanding of large spatial datasets. It is expected to provide insights and solutions on revealing the hidden patterns and characteristics of the dataset. The mechanism for exploring large spatial dataset need to enable users to access, manipulate, view, communicate and detect the relationships in the data. However, due to the diversity of data sources and the complexity of end users' requirements, there is no unified mechanism suitable for all cases on spatial data analyzation and exploration. Proposing a comprehensive visualization strategy to assist the exploration process is still a challenge work and deserve more discussions and experiments. Point data is one of the typical spatial data types which indicates the certain location in the world. The point dataset is simple in the geometric shape but can be huge in the data volume. That requires a sophisticated method available for extracting and visualizing the information from massive point datasets and present them in a way that is easy to be understood. This thesis research focuses on the large point dataset and proposes a visualization strategy with a hierarchical visual prototype considering users' information appetites for exploring the dataset in different granularities. As the opening to the thesis, this chapter works on introducing the main research context in this thesis. The content can be divided into five sections:

1.1 Background and Research Motivation: this section states some basic concepts of the research problems and introduces the motivation of this thesis project.

1.2 Problem Statement from studied case: the practical work in this thesis research is conducted via a study case from the insurance industry. Contents in this section represents the concrete problem from an insurance company.

1.3 Research Objectives and Questions: the research problem mentioned above is broken down into several more specific sub research objectives and questions in this section.

1.3 Hypothesis: this section lists hypotheses corresponded to each research question before the research. These Hypotheses are also one of the main references for the research methodology (chapter 3).

1.4 Innovation: Innovations existed along with the research process are introduced in here.

1.5 Thesis Structure: the last section gives a brief description of the thesis structure and contents in each chapter.

## 1.1 Background and Research Motivation

Nowadays, visualization techniques, especially for visualizing geospatial dataset play a significant role in communicating the information between the dataset and end users from multi scales and multi variate aspects. To support the information communication, a

Geographic Information System (GIS) is required to provide interactive interfaces and support exploratory analysis on the map display (Beconytė & Kryžanauskas, 2010). Such interfaces are expected to be aligned with users' requirements for reaching the goal of *interface success* (Roth et al., 2015). Meanwhile, the explorative analyzation process has great demands on the data computation and interaction capabilities, especially with the increasing size of datasets and mass of information (Keim & Schneidewind, 2005).

The above facts highlight the need for research on exploring visualization strategies which can be implemented into the GIS system to convey the information more efficiently. Such strategies should be able to solve two main tasks:

- 1) presenting the important information with little distortion,
- 2) Assisting exploratory queries even for massive datasets (Keim & Schneidewind, 2005) .

Although there are some systems that provide different solutions, most of them are less flexible and dedicated to specific purposes with fixed scales and topics (Lobo et al., 2017; Wood & Dykes, 2008) or they are mostly ad-hoc to specific visualization techniques (Eppstein et al., 2015; Hahn & Döllner, 2017; Phan et al., 2005) . At the same time, users' requirements are rarely taken into consideration either though they play an important role in deciding the final communication efficiency.

Therefore, it is still a great challenge for cartographers to find out the best visualization solution on facilitating users with or without GIS background to explore and understand the massive amount spatial dataset efficiently.

## **1.2 Problem statement from studied case**

Among various spatial data types, point datasets which indicate certain locations in the space account for a large proportion. In many spatial analysis cases, handling large quantity of location datasets are required for interpreting and communicating the spatial information with different end users. To maximize the valuable information can be conveyed to users, methods on processing and visualizing not only need to take characteristics of the dataset but users' requirements and background into account. This induces the research needs on exploring an efficient visualization strategy for large point dataset which is also flexible to be customized based on user requirements.

Managing the financial risk is one of the main concerns for the global insurance corporation. As for an insurance company, its portfolio which consists of certain locations and multiple risk metrics has great affection on the dispersion and degree of potential risk. Hence, thoroughly understanding the composition and spatial-temporal pattern hidden on the portfolio dataset is crucial for insurance and reinsurance companies to dominate and eliminate the risk. For supporting different end-user groups to get deep insights of portfolio dataset, providing a comprehensive visualization solution to present and communicate the dataset gets a great demand.

Case from Allianz Global Corporate & Specialty (AGCS) provides an example that processing and visualizing the large amount of point dataset (Catastrophe (CAT) risk portfolio dataset) to users efficiently is challenge. In AGCS, there are different end-user groups, i.e. portfolio managers, underwriters, have the need on understanding the trend and distribution of exposure and potential loss via CAT risk portfolio dataset. Although, a Web-GIS platform called CATlas -portfolio explorer currently available online for users to explore the dataset at certain extent, it is still not sufficient to assist users in some working cases. Therefore, using AGCS as a study case by implementing the proposed visualization strategy to further optimize CATlas-portfolio explorer has important implication for the whole research.

### 1.3 Research Objectives and Questions

#### 1.3.1 Research Objectives

Based on the background and problem statement, this research aims at proposing a visualization strategy with flexible visual prototype considering users' information appetites for exploring large point dataset. Results from the research are expected to provide theoretical and practical insights which are beneficial for processing and visualizing the large spatial point dataset.

The case provided by AGCS serves as an example on implementing the proposed visualization strategy to support end users (e.g. portfolio managers and underwriters) understand and monitor the CAT risk portfolio more efficiently. For verifying the effectiveness, it is also expected to have an web-based application which integrates the studied dataset, visualization results and customized functions together and conduct the user test in the end.

Specifically, the main objective can be divided into four sub-objectives.

- (1) Designing the workflow for executing the concept of user centred design.
- (2) Identifying visualization demands from intended user groups
- (3) Proposing a visual strategy and implementing it into the case from AGCS with customized functions
- (4) Evaluating the developed visualization strategy on assessing the effectiveness

#### 1.3.2 Research Questions (RQ)

Furthermore, each research objective can be interpreted into the following research questions:

##### **RQ1: How to integrate the idea of User Centred Design (UCD) into work?**

- 1) What is the structure of UCD?
- 2) Which methods used for assessing user requirements?
- 3) Which methods used for conducting user test?

##### **RQ2: What is the visualization demand from intended user groups?**

- 1) What are the characteristics and use cases of different user groups?
- 2) At which point in the decision-making process will users need visualization support?

**RQ3: How is the visualization strategy structured?**

- 1) Which factors need to be considered for creating the visualization strategy?
- 2) Which visualization technique is suitable for the large point dataset?

**RQ4: How is the effectiveness of the proposed visual strategy?**

- 1) How is the usability of the Geo-visualization application based on proposed visual strategies?
- 2) What functions can be implemented based on the strategy?

### 1.3 Hypothesis

To ensure the final application can serve a function on users work, their requirements is one of the most important elements need to be always assessed and tracked during the development process. User-centred design provides the guide on conducting the design work to be always aligned with user requirements. Thus, this research plans to use the idea of user-centred design on guiding the design of visual strategy and expects:

- user centred design works as great function on supporting to grasp the user requirements and digging out key issues on communicating spatial-related information with users especially for whom without GIS-knowledge background.

The basic visual exploration procedure contains *Overview first, zoom and filter, then details-on-demand*, which covers the visual perspective from low resolution to details (Shneiderman, 1996). That means it is significant to access information in the different scales (Birch et al., 2007). Thus, the second Hypothesis for this research is:

- A visualization strategy consist of multi-scale map visualization method will be support users to understand large point datasets efficiently.

### 1.4 Innovation

Innovations of this thesis mainly come from the proposed visualization strategy. The final visualization strategy on solving problems collected from users is based on multi-resolution hexagonal layers to create a multi-scale visualization prototype. A python script is created to build up multi-resolution global hexagonal grids in the Web Mercator coordinate system. The hexagonal grid layers with designed index are extendable for multi variates and time stamps. They will be used for binning spatial points datasets and constructing the multi-scale map visualization on the Web application.

### 1.5 Thesis Structure

The thesis is divided into 7 chapters. Chapter 1 *Introduction* introduces the motivation of this thesis research and lists research problems that need to be solved ; Chapter 2 *Literature Review* focuses on related research works, existing applications and associated problems for



this project.; referencing the literature review work from Chapter 2, Chapter 3 *Methodology* illustrated the research plan and selected methodologies used for setting up the whole project and solving proposed research problems in the Chapter 1. Chapter 4 *Use Case* selected a use case from insurance industry and described how methods from chapter 3 are implemented in the real case; Chapter 5 *Results* gives answers for research problems introduced from Chapter 1; Chapter 6 and 7 make discussion and conclusion based on the whole thesis content.

## 2. Literature Review

---

This chapter outlines related research referenced in this thesis project. According to the first hypothesis, the research plans to follow the instruction of user centred design to execute the work for assuring the design/development results are always aligned with users' requirements. Therefore, in section 2.1, the work of literature review focuses on investigating the related theories and methods for implementing the idea of user centred design into practice especially in the field of Cartography. Specific methods introduced in this section are mainly focused on user requirements assessments and usability test.

The second hypothesis 2 mentioned implementing multi-scale map visualization method in the visualization strategy would be meaningful to get an efficient presentation. Thus, another part of literature review work introduced in section 2.2 illustrates core facets which constitute the multi-scale map visualization (i.e. generalization and map interaction) to provide theoretical support for the final visualization strategy design in this research.

### 2.1 User-centred design

#### 2.1.1 User-centred Design in Cartography

'User-centred design' is a term used for describing a development/design process in which users are involved and influence the final product (Hou,2004). The concept of 'user-centred design' was first proposed by Donald Norman (Norman & Draper, 1986) in his co-authored publication. Later on his book *The Psychology of Everyday Things* (Norman Donald, 1988) Norman pointed out the importance of usability of design further. He put forward seven design principles:

- (1) Building up conceptual model based on existing knowledge. Before conducting the design work, create a brief conceptual model which references knowledge widely grasped by the public.
- (2) Ensure tasks in the design are easy for users. Normally, users are able to memorize five items during the short period. Preventing the overload of users' memory, designers should limit the numbers and the complexity of tasks.
- (3) Making items easy to understand. Users should easily figure out the right way to execute the operations.
- (4) Making full use of graphics. Graphics help to convey the information in a more straightforward manner.
- (5) Being able to lead users start to use the design.
- (6) Preparing solutions for possible errors. Users should be informed about possible errors and be able to cope with errors.

(7) Make sure everything is at least aligned with standards which focus on guiding the design work always track with users' needs. (Hou,2004)

Noman's work draws designers and developer's attention to placing users into the centre when starting a new project. Usability engineering model(Figure 2.1-1) (Nielsen, 1992) proposed by Nielsen provides more concrete methodological instructions.

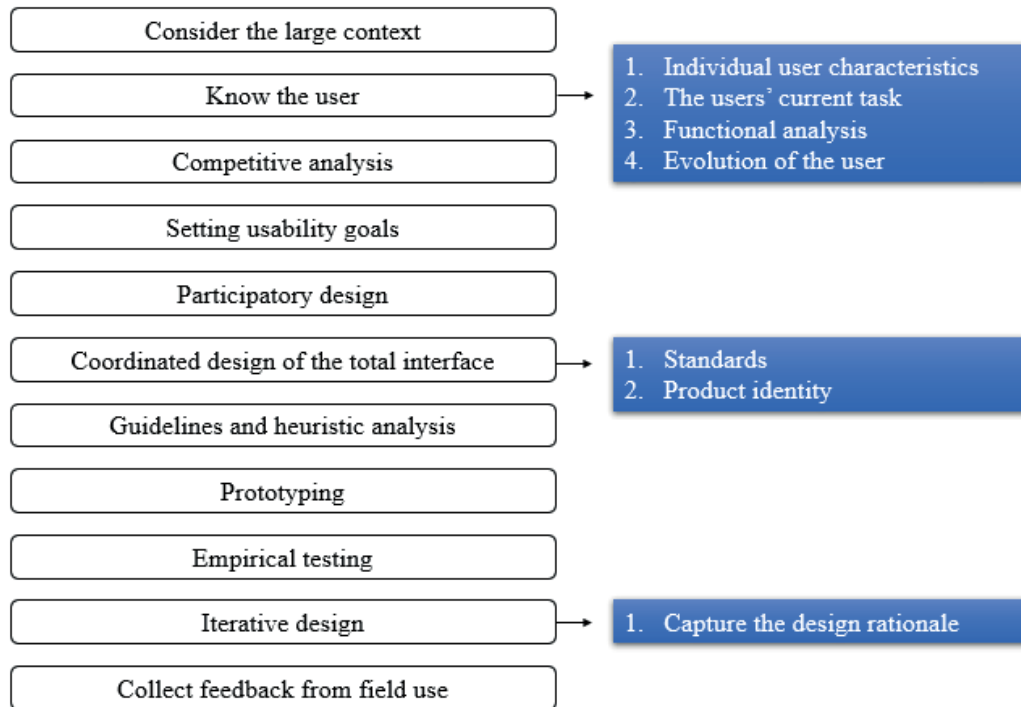


Figure 2.1-1: Usability engineering model

There are eleven elements in the usability engineering model. However, not all elements must be required to put into a product development process. The most essential elements are prototyping, empirical testing and iterative design. (Nielsen, 1992) A product cannot be developed perfectly at the first attempt, iterations in the user-centred design loop ensured the space and time for product revision during the development process.

With the developing demands on the web-based GIS application, the importance on implementing User-centred design technology into development work has aroused the great awareness among developers and scientists. It is apparent that without well-constructed applications and maps, users always hard to fully use functions in the application and understand the information behind the map. As a user-oriented science, integrating user-centred design method into cartography is not a new topic. However, it is always a great challenge to conduct user -centred design method in an application producing lifeline efficiently (Mao et al., 2005) . Main reasons for that come from two aspects: 1) Conducting user-centred design need sufficient time and money, 2) There is no systematic theoretical guidance on integrating User-centred design (UCD) into traditional application development work-flow especially in the field of Cartography. (Roth et al., 2015)

In practice, scientists in the field of cartography did not always use the same prototype of UCD process and had different stages during their work (Roth et al., 2015) . For example, in the project GeoVISTA CrimeViz scientists designed four stages on their UCD process: (1) Interviews on assessing users' needs: they performed comparative assessment interviews for both targeted a user group and a peer group, (2) Think Aloud study by experts: experienced experts are involved in the study to comb potential functions for the application, (3) developmental online survey: gathering comments and feedback from users on the beta release, the survey result on this stage will be the important reference for further revision and (4) additive online survey: the survey on this round aims on to verify the usability and utility of the full release (Roth et al., 2015) . Slocum et al. distilled the usability engineering model and set six steps for developing a water balance uncertainty visualization software. (1) developing the prototype (2) evaluating by domain experts (3) revising the software (4) evaluating by usability experts (5) revising the software (6) evaluating by the end users. In their project, they skip the users' involvement but develop the prototype by themselves at the beginning. End users only participated in the end for the final assessment, the feedback for the software revision mainly comes from related experts. (Slocum et al., 2003)

'User-centred design' is a flexible and multi-stage paradigm. There are no rigid rules on the number of stages that need to be conducted and when users should be involved during the process. But having users' input and product revision are most crucial requirements for getting a user-friendly result.

### 2.1.2 User requirement assessment

Assessing user requirements is an important section on developing human-centred systems or applications. It helps to enhance the productivity and reduce the training costs at the later stage (Maguire & Bevan, 2002) .Classical user requirement assessment methods include interviewing, field study, observation, diary keeping, storyboards and so on. Different methods in different situations and phases during the whole user requirements assessment procedure. Maguire & Bevan (2002) defined four sections through the whole user requirement analysis procedure (Shown as Figure 2.1-2). For these sections, they also selected typical techniques for each (summarized in Table 2.1-1).

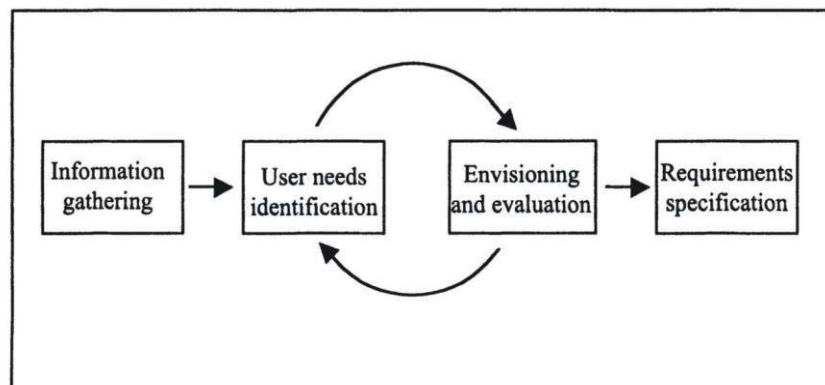


Figure 2.1-2: User Requirements Analysis(Maguire & Bevan, 2002)

**Table 2.1-1: Methods for user requirement analysis**

<b>Section</b>	<b>Method</b>
Information gathering	<ol style="list-style-type: none"> <li>1. Stakeholder Analysis</li> <li>2. Secondary Market Research</li> <li>3. Context of User Analysis</li> <li>4. Task Analysis</li> <li>5. Rich Pictures</li> <li>6. Field Study</li> <li>7. Diary Keeping</li> <li>8. Video Recording</li> </ol>
User needs identification	<ol style="list-style-type: none"> <li>1. User Surveys</li> <li>2. Focus Groups</li> <li>3. Interviewing</li> <li>4. Scenarios and Use Cases</li> <li>5. Feature Workshops</li> <li>6. Evaluating existing systems</li> </ol>
Envision and Evaluation	<ol style="list-style-type: none"> <li>1. Brainstorm</li> <li>2. Card Sorting</li> <li>3. Affinity Diagramming</li> <li>4. Storyboards</li> <li>5. Prototyping</li> <li>6. Allocation of function</li> <li>7. Design guidelines and standards</li> <li>8. Parallel Design</li> </ol>
Requirements Specification	<ol style="list-style-type: none"> <li>1. Task/Function Mapping</li> <li>2. Requirements Categorization</li> <li>3. Prioritization</li> <li>4. Criteria Setting</li> </ol>

Normally, a mix of methods is required. Lloyd.D et al (2011)proposed that supporting domain experts to understand the possibilities of visualization techniques when collecting their requirements serves a great function. They tried Volere method (Suzanne Robertson & James Robertson)and instructive lecture to introduce the basic concepts of possible visualization strategies to non-visualization experts to induce them to "cross the cognitive bridge" (Lloyd & Dykes, 2011) .In the project, conducted in cooperated with Leicestershire County Council (LCC), scientists calculated the word frequencies among emails communicated with LCC experts to summarize out key themes from their literature inquires (David Lloyd et al., 2007).

### **2.1.3 Usability testing**

The underlying objective of usability assessment goes to evaluation of the capacity of an application on satisfying the users' requirements. Specifically, the assessment result will grade an application on three aspects: effectiveness, usefulness and performance (Koua et al., 2006).The decision of assessing methods relies on setting visualization operations and knowledge-discovery tasks (Koua et al., 2006) .To test the usability comprehensively, it is meaningful to define and categorize the visualization operations and knowledge-discovery

tasks beforehand. For example, scientists from CommonGIS project set a series of task that corresponded with five interactive techniques and asked users to answer the questions about operation outcomes step by step (N. Andrienko et al., 2002) . Based on the task, testers can record the testing time, ask questions and calculate the accuracy of answers to quantify the usability. On the other hand, comments from post-hoc interviews and questionnaires also record the user's impression on the usefulness. In project NEFIS researchers assessed the usability of the geo-visualization techniques by getting the feedback from potential users about their impressions on the exploration procedure (G. Andrienko et al., 2006) .

## 2.2 Multi-scale map visualization

It is always easier to find a building in a city map than in a world map since the world map covers a larger searching range and more irrelevant information. Comparing to loading all details and information in one map scale, users can use multi-scale maps to narrow down the searching area on large scale maps and explore details among small scale maps. The multi-scale map visualization method displays a series of mutually inherited maps in different scales by interface interactions such as zooming. Zooming in or zooming out will trigger the changes of map display with different geographical information.

If we categorize existing multi-scale map visualization applications based on map contents, there are two main classes: applications for geographical objects and applications for geographical values. Considering a comprehensive multi-scale map visualization application, many factors like resolutions and relationships between different factors all need to be taken into account. In our research, we will build up a multi-scale map based on the geographical values and present them in a web-based GIS platform. Thus, literature introduced in the following sections will be surrounded with specific facets which are map prototype, generalization and interactions.

### 2.2.1 Multi-scale Visualization based on grids

#### ***Spatial Binning***

Spatial binning is a technique that helps to aggregate multi locations into polygonal areas. Such polygonal areas can be classified as irregular polygons and regular polygons. For irregular polygons, they are always constructed by political boundaries (e.g. countries and cities) and used to create thematic choropleth maps based on aggregated values such as population density or household incomes. As for regular polygons, they are the tessellation of regular polygons with different shapes such as hexagon or rectangle. (Battersby et al., 2017) In this section, we are mainly discussing the regular spatial binning including binning shapes, usages and existed problems.

#### 1. Advantages of spatial binning

One of the distinctive advantages of the regular spatial binning is *it hold shape and spacing in the symbols constant* (Battersby et al., 2017) which makes the visual spatial analysis

more effective. Meanwhile, in many cases, spatial binning can use its strengths to make up for the shortcomings of many other mapping methods. Take the dot map as an example. One big challenge of presenting point values in the dot map is handling the points' size and symbolizations. The dot map needs to present multi points symbolizes based on different points values for many instances. Like traffic fatality, places has more crash accidents cannot stand out because many fatalities are overlapped by others (john.nelson, 2013) .With the number of points growing, it becomes harder to interpret the congested dot map. Spatial binning helps to group points into different categories for better interpretation(GIS Lounge, 2014).

## 2. Visual design of spatial binning

The Discrete Global Grid (DGG) is the foundation of the spatial binning. A DGG is composed of a series of regions that segments the earth's surface into small partitions. These regions are also called as cells in the DGG (Sahr et al., 2003). Depending on the shape of the cell, the DGG can be divided into irregular DGG and regular DGG. Like water bodies, which are regarded as typical irregular regions on the earth's surface. And for regular DGG, it is already widely used in the field of geographical domain owing to its great advantages of unbiased shapes. A regular DGG is impartial for uncovering the spatial pattern and building spatial-related algorithms (Sahr et al., 2003). For the regular DGG, it can also further be divided into rectangular DGG and hexagonal DGG etc. according to the regular cell shapes. Besides, the DGG can be categorised based on the referenced coordinates system as well. One of the most used DGGs are created based on geographic coordinate system (showed as Figure 2.2-1) that are divided by latitudes and longitudes. Another is the grid system discretize the sphere called Geodesic Discrete Grids. For the Geodesic Discrete Grid system (showed as Figure 2.2-2), it can be further categorized by the design choices which are composed by the Base Polyhedron, Grid's Orientation, Partition and Transformation methods.

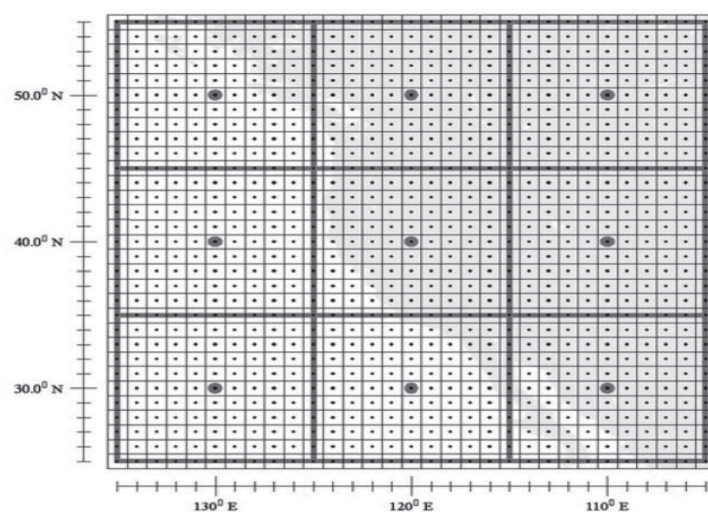
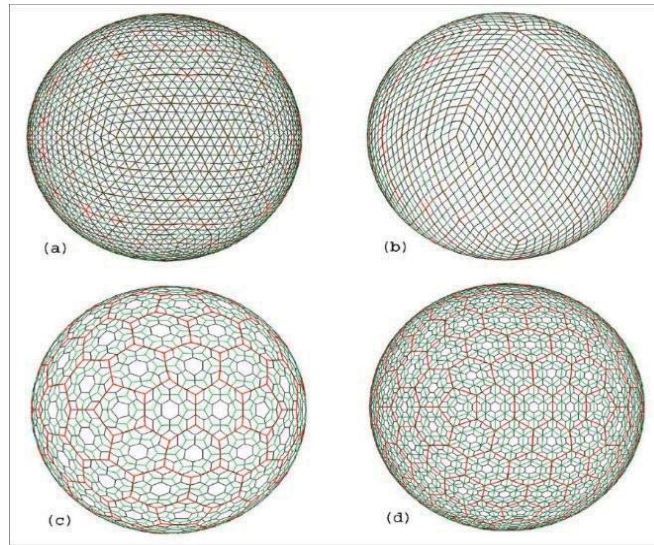


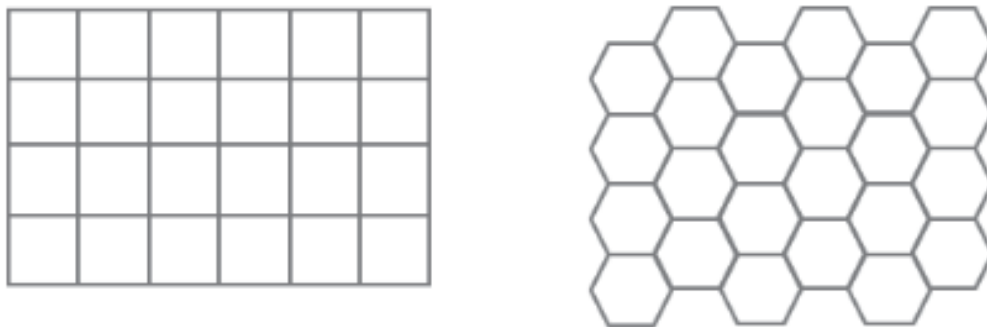
Figure 2.2-1: Example of DGG in geographic coordinate system (Kevin Sahr, Denis White, and A. Jon Kimerling)





**Figure 2.2-2: Example of Geodesic DGG (Kevin Sahr, Denis White, and A. Jon Kimerling)**

The shape of the DGG decides the spatial binning shapes. There are many regular shape choices and the most commonly used are rectangular and hexagon (Figure 2.2-3). The shape of hexagon was first implemented by Carr, Olsen and White (1992) in the spatial application. A rectangular grid is a relatively simple shape and easier to be created. However, in terms of accuracy, hexagon grids have better performance than the rectangular grids. (Birch et al., 2007) In our research, we also selected hexagon grids for the binning process. The detailed comparison of two shapes will be introduced in Chapter 3, Methodology.



**Figure 2.2-3: The shape of binning grid**

*(Why Hexagons?—ArcGIS Pro | Documentation, 2020)*

Besides the shape of the grids influences the visualization effect, other additional visual elements extend the possibility of binning map to present the more information. Figure 2.2-4 shows elements listed by Heimer et al. (2018) which could be added in the binning map. The map from john.nelson (2013) (showed as Figure 2.2-5) is able to present multivariate information in one grid. It was coloured based on the counts of traffic fatality and scaled to different size rely on the ratio of alcohol-related accidents.



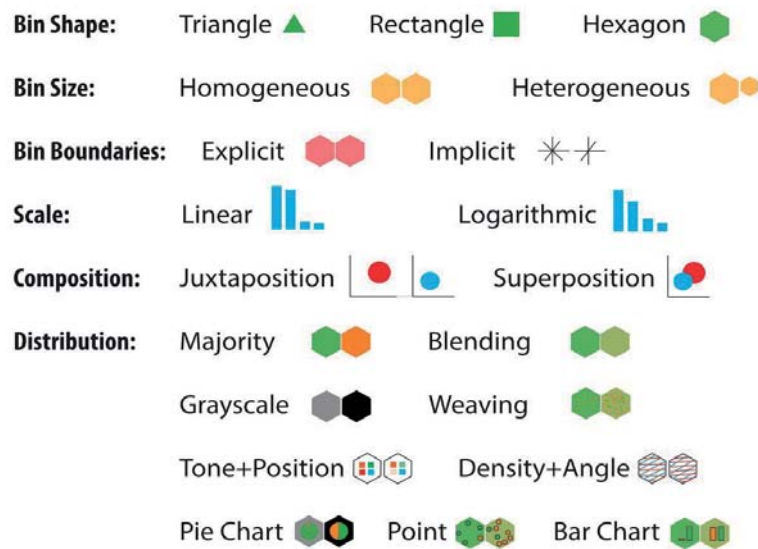


Figure 2.2-4: Visual Designs for Binned Aggregation.jpg (Heimerl et al., 2018)

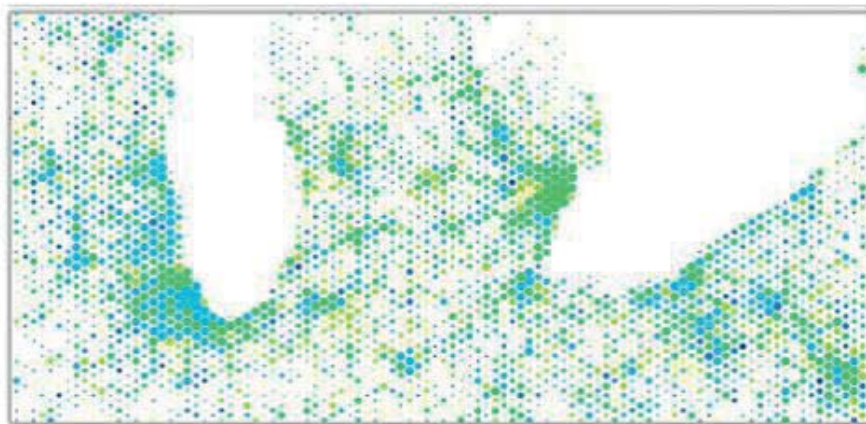


Figure 2.2-5: Hexagonal binning for presenting multivariate information (john.nelson, 2013)

### 3. Related problems

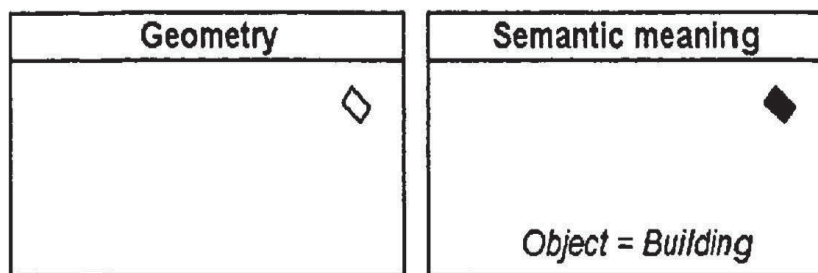
Due to the great advantages of spatial binning on benefiting the map to be easier for understanding, spatial binning has become a popular mapping method among cartographers. Therefore, there are more and more discussion on how to make spatial binning and propagate the benefits of spatial binning (Battersby et al., 2017). However, rare discussions are focusing on problems caused by projection coordinate along binning process. Since the geodesic areas are different for grids owns the same planar size in the projected coordinate system (Battersby et al., 2017). Most cases treat the spherical world as a flat surface and created the tessellation based on the projected plane. Errors caused by such method won't cause great influence on the large scale. But for place with larger latitude, the distortion on the size gets greater. Targeting to this issue, Battersby et al. gave the method to calculate the 'safe zone' for grids created on the web Mercator projection under the certain error tolerance (2017). In the chapter 3 methodology we will

give the further discussion on the regular hexagon grids created based on different reference system.

### **Generalization for Multi-scale map visualization**

We call the map presentations for small range areas with more details as large-scale maps and map presentations for large areas with less details as small-scale maps (Jones & Mark Ware, 2005). Changes of map in different scales are involved with modifications to the display, modifications to features' values, or both. These two modifications correspond with *cartographic generalization* and *statistical generalization* separately. They are always interdependent, which means the generalization method chosen for one type can always have the influence on the other one (Brewer & Buttenfield, 2007).

Inside the cartographic generalization, the activities can be further divided into *semantic generalization* and *geometric generalization*. semantic generalization concerns with choosing appropriate information to be presented, geometric generalization influenced spatial features' shape and structure (Jones & Mark Ware, 2005) . To be more specific, in the series of multi-scale maps, semantic generalization goes to create a classification system for map objects and decide which level of semantic details should be presented in which map scale. For example, a small-scale map may only require the representations of cities but a large-scale map would need to know the location of single building. On the other side, geometric generalization works on selecting suitable geometric shapes on constructing map symbols for each map scales. For example, cities are only required to be presented as points in small-scale maps but should have detailed city boundaries in large-scale maps (Jones & Mark Ware, 2005).



**Figure 2.2-6: Semantic and geometric meaning in the map (Ormsby & Mackaness, 1999)**

Therefore, it's vital for map generalization to find a way on building up the classification system. Existed researches gave diverse ideas on how to implement the map generation process. Zhou & Jones et al. thought factors influenced the generalized presentation can be called as generalization criteria. Even in the same scale, due to the different generalization criteria, the final representation of objects can be different (Shown as Figure 2.2-7). Therefore, They built Generalization Model which merges *resolution*, *generalization criteria* and *representation* together in and used the model to quantify the generalization process (Zhou & Jones, 2003). This model also helps to build up relationship between the multi-scale

database and web demonstration. Ormsby & Daniel et al. put forward the Phenomenological Approach for the map generation. They thought solutions for map generalization should take both object types and map types consideration and be very flexible. Therefore, they integrated three components 1) object's geometry 2) objects' semantics 3) inter-object relationship made for the spatial phenomenon into an object-based algorithm for the map generalization (Ormsby & Mackaness, 1999).

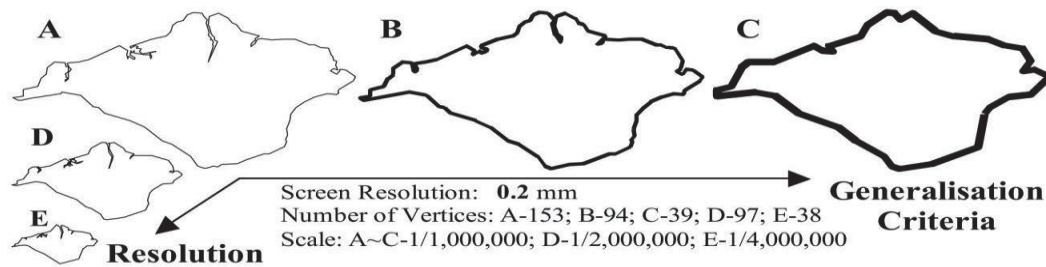


Figure 2.2-7: Resolution and Generalization criteria in Geometric generalization(Zhou & Jones, 2003)

(Spatial) data processing technique is the bridge between the raw data and visualization works. It works on digging out the knowledge potential of the datasets by organizing them more intelligent. Generally, (spatial) data processing technique can be categorized into four sub-divisions: dimensionality reduction technique, sub setting technique, segmentation technique and aggregation technique (Keim, 2007, 2005). Among them, the technique on aggregation has absorbed great research interests recently. An aggregation model works on summing up the uniform information from Tuples. As for spatial aggregation, According to Andrienko et al. (2006), the generic concept is grouping the areas whose characteristics can fit into the same division . Lopez et al. (2005) already did an extensive survey on spatial aggregation techniques. For a comprehensive dataset, not only does it possess a great volume of storage capacity but it can also sometimes be sensitive to privacy concerns. Thus, developing an efficient aggregation technique to summarize and explore the dominant patterns for better decision making is very critical. Scientists working on this field are keen on comparing and improving the performance of existing aggregation methods. Gomez et al. (2007)developed a framework for uniting GIS and OLAP (On Line Analytical Processing) and achieved aggregate queried hierarchically . Pedersen & Tryfona (2006)investigated the efficiency of traditional pre-aggregation method in the spatial DWs (data warehouses) and did extension on pre-aggregation techniques for spatial DWs . Zhang & Tsotras (2001)did experiment on improving the performance for MIN/MAX aggregation on spatial objects and introduced MR-tree method originated from R\*- tree as optimized result

### **Interaction Design**

Cartographic presentation gives the chance to communicate spatial-related information visually. However, it's not sufficient to explore the diversity of datasets just based on single map (Lobo et al., 2017). Compared to the standalone scale map, a pyramid of maps with multi scales enable the interaction can be applied in the Geo-visualization application

(Dumont et al., 2020). This is a great improvement, that multi scale maps could present different geographic information in different details via an inter-related map architecture. Thus, a good user interaction is regarded as an important complementary to the Geo-visualization application.

Regarding to ensure the quality of interaction design, Elmqvist, Moere et al. (2010) proposed a sets of interaction design guidelines for information visualization which can also give a good guidance on improving the interactions on the map visualization application

- 1) **Animation:** Using animation to transit views smoothly. The changes of views on the interface are always along with distortion. Animation could avoid the abrupt mode transitions in users' mind and give users the time to adjust to a new vision.
- 2) **Responds:** Providing instant feedbacks for individual interactions. Not only 'main' interactive event but also small clicks are all need instant responds in the platform.
- 3) **Simplification:** to minimize the number of interactions in the interface. It's recommended to integrate several interactions such as selection and filtering together in the visual presentation. It is also need to be prohibited to put all interaction separately in the control panel.
- 4) **Integration:** Integrating interactions into visual presentation. Interactions such as selection, zoom can naturally be put into visual presentation without interaction interfaces. For others such as sliders, it' s better to make the interface be put in the visual presentation seamlessly.
- 5) **Reward:** Rewarding users to use the interactions in the platform. The designed interactions should provide several 'Rewards' which can be regarded as concerned information such statistical charts, diagram for their exploration. This will encourage users to interact more and find more.
- 6) **Endless:** Ensuring the interactions are always existing. In other perspective, it gives requirements on the robustness of the system. Users can always have interactions in the platform for exploring.
- 7) **Structure:** Putting all interactions in a clear structure. Interactions for connecting different views and achieving different functions should be organized in a good structure. Thus, users could return back to the previous status.
- 8) **Consistent:** Avoiding changing modes instead containing different interactions in the same mode.

(Elmqvist et al., 2011)

Existed Geo-visualization applications did diverse explorations on the user interaction for Multi-scale maps. Efforts behind the interaction design are related with multi facets such as development tool, map contents, data structure etc. Thus, Innovations of interaction design can also be found from different aspects. 1) For getting better effects, sometimes, the interaction design needs to be cooperated with visual design at the same time. Slingsby, Dykes et al. (2014) selected five of interaction design rules above to be aligned with. Corresponding with preliminary user requirements, they designed five main interactions

*Selection, Filtering, Orientation, Reordering and Rescaling* and fitted them into the spatial analysis system. During orienting, the opacity of selected area will also be changed for presenting multi scales layers in the same selecting view. (Slingsby et al., 2014)



**Figure 2.2-8: Designing an exploratory visual interface (Slingsby et al., 2014)**

2) Zooming could be a very supportive interaction especially for exploring multi-scale maps. One of the classical cases comes from Duncan A. Smith's project on visualizing GHSL (Global Human Settlement Population Layer) in the interactive map. They embedded multi-scale raster layers with population datasets together and allows users to navigate the map from global to city-region scales by zooming method. Corresponding to different zoom levels, in their interactive map application, users are also able to get descriptive statistics by selection. (Smith, 2017) 3) Data format and structure influence the interaction design structure. Wood, J. et al. (2007) used mashup language to develop an *Encoding interaction* based on Google Map environment. It utilized the advantage of the KML data structure to enable users could query different interested contents based on items' hierarchical labels.

## 3. Methodology

---

This chapter will introduce the concrete research plan and methodologies implemented in this thesis project and give reasons for some specific selected methodologies. The chapter is divided into three sections:

3.1 research plan: There are no rigorous rules for steps to achieve the iterations in the user-centred design loop. This section will follow the general instruction of user centered design to propose a dedicated work flow plan for this thesis project.

3.2 conducting user-centred design loop: Based on the workflow introduced on the above section, this section will give out specific methods for conducting the user requirements assessments and user test

3.3 Multi-scale visualization: as the core of the whole thesis project, this section gives methods used in the visualization development work which includes the certain visualization techniques used, how to create the global hexagonal tessellation, interactions design and map symbolization.

### 3.1 Research plan

This thesis project follows the rule of user-centred design to plan the workflow, then finds suitable strategies for the visualization problem. The Research Plan section will give an overview on the project's set up which includes the procedure and components of workflow and visualization strategy. The plan of thesis set-up introduced in the section 3.1.1 will be used for conducting user-centred design when the project time is limited and design work is proceeding in the team. The visualization strategy introduced in the section 3.1.2 can be used as workflow prototype to achieve multi-scale map visualization for exploring large point datasets.

#### 3.1.1 Workflow of thesis project

The core of user-centred design is achieving the iteration in the working loop that ensures the application or design works meet the requirement of users. Such iterations aim at bringing opinions from related parties regularly to avoid massive changes later on in the development process. This research planned to conduct the iteration within the team (Shown as Figure 3.1-1)



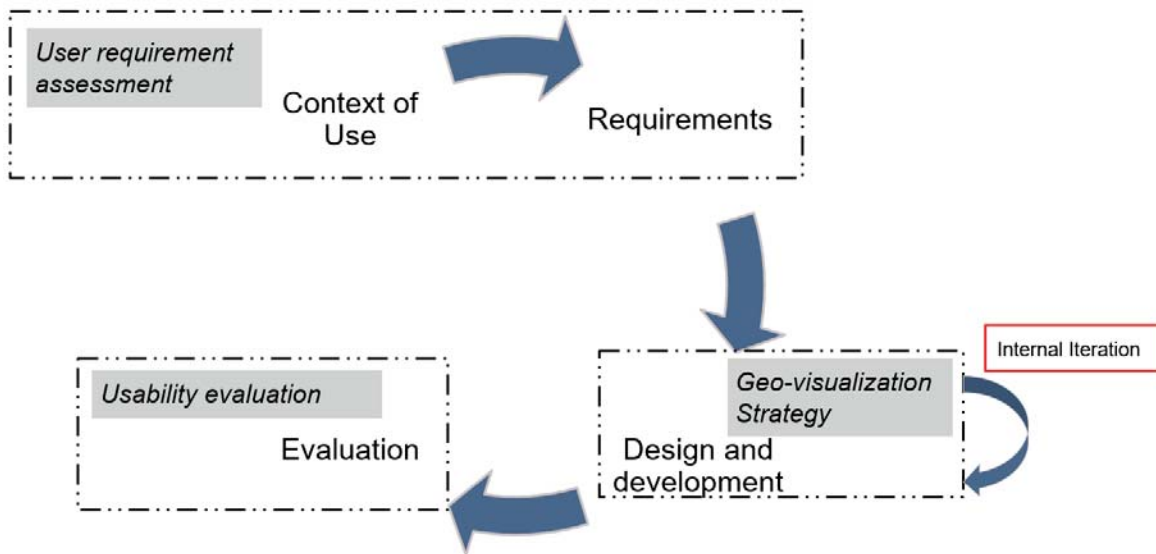


Figure 3.1-1: The set-up of the project

and get instant feedback from other team members who have long-term contact with end-users and have good understanding of users' needs. The whole procedure can be described as

- 1) Reaching out to users to collect all context of requirements,
- 2) Assessing collected requirements and translating requirements into concrete design structure and the application's functions,
- 3) Proposing a visualization strategy and conducting the design work and hosting regular discussion meetings with team members during the development procedure to request amendments,
- 4) Finishing the final application and return to end users to conduct the user test and evaluate the usability.

### 3.1.2 Visualization strategy

The visualization strategy came from the result of user requirement assessment. An interactive geo-visualization platform was proposed to be built for supporting multi-scale visualization of point location values. The total strategy (showed as Figure 3.1-2) can be divided into three parts:

- 1) Platform design: including the design of the platform's layout and functions to meet users' needs,
- 2) Multi-scale Map visualization: The global tessellation is chosen to build up multi-resolution layers for supporting the multi-scale visualization and spatial binning method is used to aggregate point locations' values into different layers; Zooming and panning are applied as main interactions for supporting multi-scale exploration,

3) Development: This stage goes on to realize the design ideas presented above; commercial GIS software or open source development tool can be used. Specific methods used for Multi-scale Map visualization will be introduced in the section 3.3. Chapter 4 will reference the conducted use case to introduce required layout's functions, platform layout and values to be mapped.

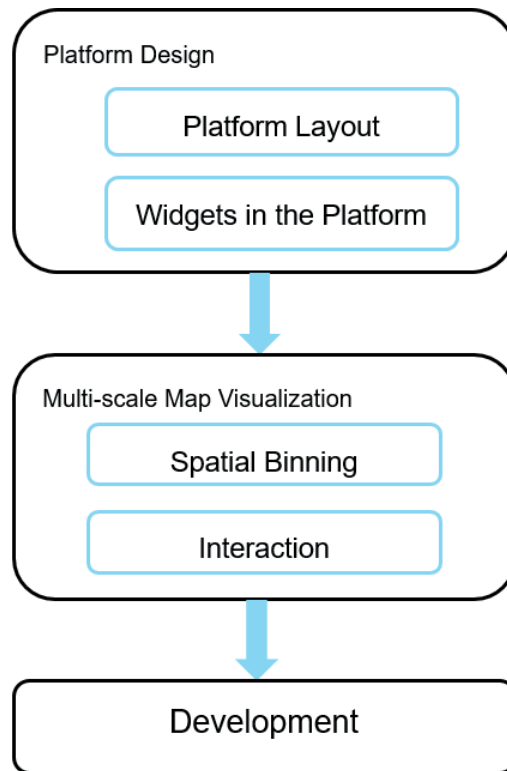


Figure 3.1-2: the structure of visualization strategy

## 3.2 Conducting User-centred Design Loop

In the loop of user-centred design (Figure 3.1-1), excluding the design and development work, all other steps concerned with users need to reference the science of user-related research to select suitable techniques for execution. This section will go on to introduce these certain methods and procedures on 1) collecting and assessing user requirements, and 2) User testing in this thesis project.

### 3.2.1 Methods for User Requirement Assessment

There are a lot methods available for the whole process of user requirement assessment, each of them has one or more suitable situations to be applied in. Based on previous experience, there were multiple methods combined for gathering and assessing the user requirements. In this thesis project, five users were targeted, each of them representing one department. Thus, it is feasible for us to have interviews with each of them in the beginning for gathering user requirements. Meanwhile, other team members who are responsible for supervising the project also have long-term contact with end users. Therefore, some



brainstorming sessions were also set up in the team to uncover valuable information from their previous experiences.

This research selected methods introduced by Martin Maguire & Nigel Bevan (2002) to plan a user requirement assessment workflow (showed as Figure 3.2-1) for this thesis project which starts with defining targeted users and ends with Task/Function Mapping.

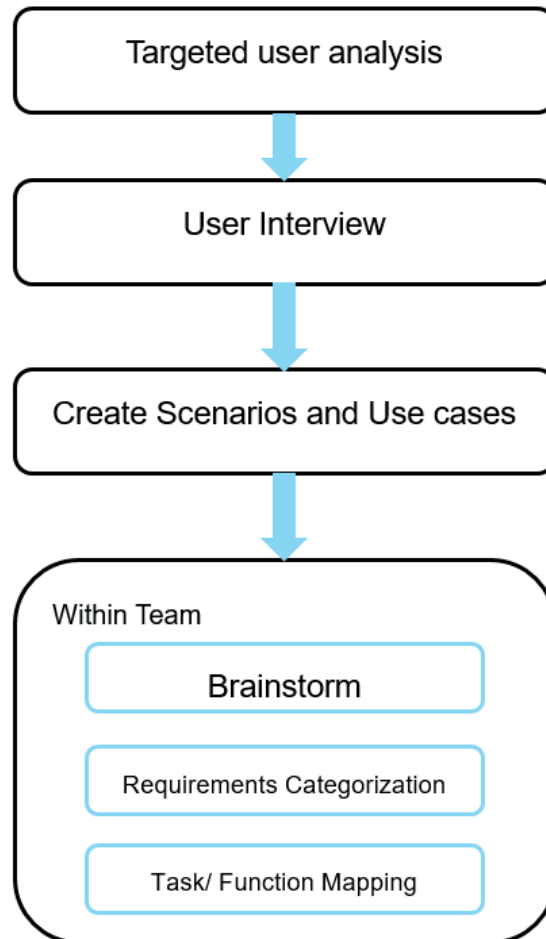


Figure 3.2-1: The workflow for user requirement assessment

### 3.2.2 Methods for User Testing

User testing as the last step on the workflow in Figure 3.1-1 works on evaluating how good the visual strategy is and the usability of the final visualization application. Before selecting suitable methods for the step of user test, the first thing to consider is what kind of test results we want to get. In this research project, it is hard to obtain convincing statistical results to quantify the application's usability from limited users. This research tends to understand users' perceptions of the application from their textual feedback in order to justify the results of the visualization strategy and extract ideas for future improvements. Interview is one of the most straightforward methods to collect the comments through interactive communication with users. Meanwhile, the questionnaire survey provides users more time to answer specific questions and give comments. Thus, this research integrated

these two methods into the user test procedure and divided them into two rounds (shown as Figure 3.2-2). During the first round, setting up an interview with users and collecting their first impression via the questionnaire survey. Next, users got the opportunity to operate the application by themselves and were asked to answer more specific questions provided in the second questionnaire round.

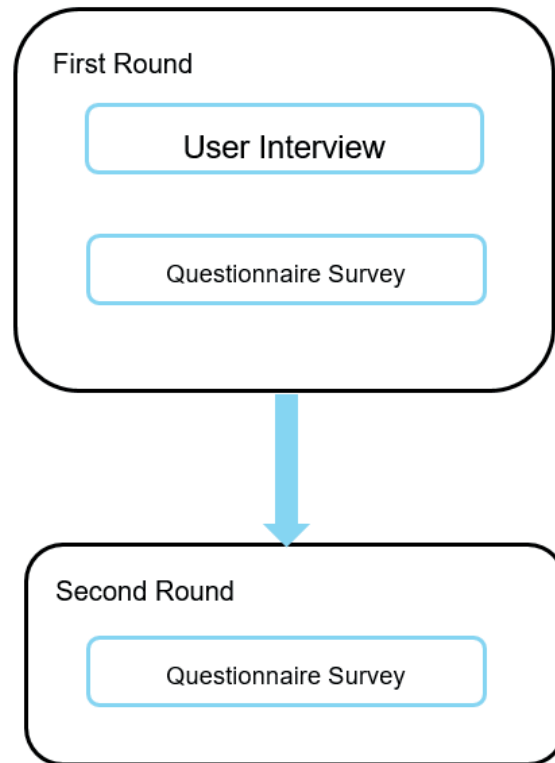


Figure 3.2-2: The workflow for user test

### 3.3 Multi-scale Map Visualization

For doing the multi-scale map visualization in this research, it is necessary to recall to the original data exploration problem that the thesis need to address. That is assisting users to explore the spatial distribution characteristics of large spatial point datasets and finding it is outliers. The proposed solution to this is using multi-scale maps to support exploration process and implement spatial binning method to aggregate massive amounts of point locations into simpler polygon layers. According to the Figure 3.1-2, the final visualization strategy is divided into three parts. Of them, the platform design changed with users 'preferences, required functions and development tools will vary depending on the development environment. Therefore, these two parts will be covered with more details in Chapter 4 based on the specific use case. This section only focuses on introducing methods for the multi-scale map visualization. It is divided into three parts:

1) Hexagonal binning: the basis of multi-scale map visualization is creating multi-resolution layers to support multi-scale demonstration. Based on the idea of spatial binning, creating

global hexagonal tessellation in different resolutions to aggregate location values at multiple map scales,

2) Interaction: Interaction techniques support users to switch map views on demand. This thesis research chose zooming as the main interaction method to help users' analysis the spatial properties of the dataset at different scales,

3) Visualization of hexagon layers: the problem of map visualization in this research is the same with the problem of choropleth map's visualization. It comes from two main aspects. Firstly, how to choose a suitable colour scheme; secondly, how to classify the value into different categories.

### **3.3.1 Hexagonal binning**

The advantages of regular spatial binning are distinguished. Firstly, compared to the shape of the political boundaries, regular polygons present the spatial phenomenon of changes smoother. Secondly, the regular polygon layer composed of equal-sized cells provided enumeration units for spatial analysis. Lastly, the size of the polygon can be adjusted according to the points' distribution (Battersby et al., 2017). Hexagonal binning is already a widely used method for spatial aggregation. Chapter 2, Literature Review already discussed existed applications using hexagonal binning. Which also means there are several different methods available for binning values in hexagonal tessellation. The difference depends mainly on the chosen coordinate system. This research decided to build up the hexagonal tessellation layers based on Web Mercator Coordinate System. Following contents will first identify the advantages of hexagon as the shape of tessellation then give the rationale for the choice of the coordinate system.

#### ***Why Hexagons?***

The choice of hexagons as the chosen tessellation shape, instead of the typical rectangle, requires some justification. Benefits of hexagonal tessellation which are mentioned frequently in different sources. Here, five main reasons are summarized:

1. Deriving the benefit from the hexagon's multilateral structure, there is less bias caused by grid's edges from hexagonal tessellation.
2. Hexagons approximate more closely to circles making them more suitable to represent naturally phenomena.
3. From a visual perspective, rectangular grids draw unbroken lines across the plan, such lines could hinder the real spatial pattern of the data. In contrast, the edges of hexagons tend to cut off the straight line and create curvilinear boundaries to present the pattern more clearly.
4. Hexagonal grids suffer less distortion compared to rectangular grids especially in the large research region.

5. Compared to four-edge rectangular grid, finding neighbourhoods among six-edge hexagon grids is more straightforward. (*Why Hexagons?—ArcGIS Pro | Documentation, 2020*)

### Coordinate System

Although it is unavoidable to have size distortion in the projected coordinate system, this research still decided to perform the tessellation in the Web Mercator coordinate system. In order to justify the feasibility of the chosen coordinate system, the contents below will first compare different results based on different reference systems, then elaborate on the visual benefits given by the projection coordinate system in the end. The error in the geodesic area of hexagonal cells created in the Web Mercator coordinate System be discussed the comparison of results in the Chapter 4, Section 4.4.3.

#### 1) Comparison of different reference system

A Discrete Global Grids (DGG) system is generated in the Icosahedral Snyder Equal Area (ISEA) Projection and partitions the world into equal geodesic grids. In the hierarchical DGG system, the sequential resolution of grid layers is determined by the cell topology and selected aperture. Once these two elements are decided, the correspond available resolutions are fixed in the meantime. Additionally, these resolutions are mostly irregular non-integers. Table 3.3-1 shows the information of hexagons created based on ISEA43H method. It is hard to find consistent hexagon layers with demanded resolutions in the same DGG system. Grids built in a projection coordinate system can be adjusted into different resolutions but at the cost of a decrease in size accuracy. It is plant to build up a series of hexagon layers having rounded and consistent resolutions (e.g. 1km<sup>2</sup>, 100 km<sup>2</sup>, 10000km<sup>2</sup>) to help users without GIS knowledge background comprehend layers more easily in this research. Table 3.3-2 gives more detailed comparison on hexagon grids created in different reference system.

**Table 3.3-1: Hexagons Information created use ISEA43H(Sahr, 2018)**

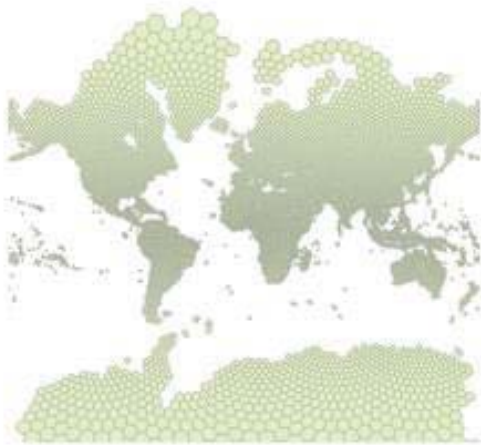
Mixed Aperture 4 and 3 (ISEA43H)							
# Ap 4	Res***	Number of Cells*	Hex Area** (km <sup>2</sup> )	Internode Spacing (km)			
				min	max	mean	std.dev.
1	2	1224	4,250,546.84770	2,050.27000	2,445.80000	2,210.50000	134.60200
1	3	3621	1,416,848.94923	1,143.10000	1,360.38000	1,280.77000	76.55180
2	3	4821	1,062,636.71193	1,021.81000	1,246.71000	1,111.74000	66.88040
1	4	1,082	472,282.98308	680.80200	835.86200	742.02700	44.62200
2	4	1,442	354,212.23731	565.39700	690.57000	642.04700	39.50560
3	4	1,922	265,659.17798	510.49600	628.59000	556.76100	33.52270
1	5	3,242	157,427.66103	374.89400	463.73500	428.31100	26.59520
2	5	4,322	118,070.74577	340.28000	420.15500	371.29600	22.39940
3	5	5,762	88,553.05933	280.26400	349.09400	321.32800	20.04290
4	5	7,682	66,414.79450	255.19700	315.51700	278.50600	16.82030
1	6	9,722	52,475.88701	226.83800	280.57700	247.56900	14.95740
2	6	12,962	39,356.91526	186.14700	233.60700	214.27900	13.42690
3	6	17,282	29,517.68644	170.12500	210.75100	185.68800	11.22680
4	6	23,042	22,138.26483	139.31500	175.54900	160.73100	10.09490
1	7	29,162	17,491.96234	123.74300	156.19900	142.87800	8.98060
5	6	30,722	16,603.69862	127.59200	158.48900	139.27000	8.42455
2	7	38,882	13,118.97175	113.41500	141.01800	123.79700	7.48970
3	7	51,842	9,839.22881	92.65450	117.43200	107.16800	6.74663
4	7	69,122	7,379.42161	85.06050	105.99400	92.84930	5.61894

**Table 3.3-2: comparison of different methods**

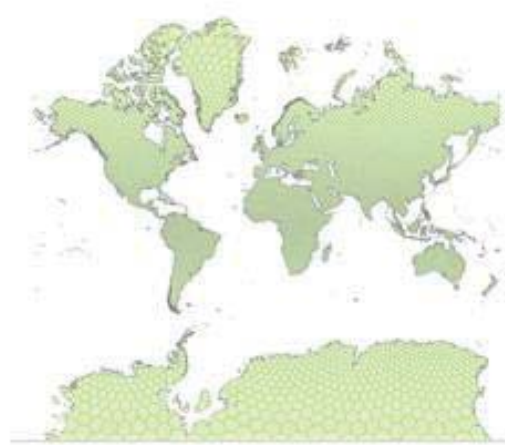
<b>index</b>	<b>Reference System</b>		<b>Advantages</b>	<b>Disadvantages</b>
1	WGS84 Mercator projected coordinate system	Web	1.Same coordinate system with web map 2.Resolution is adjustable	1. distortion in the distance, each cell doesn't have equal geodesic area size
2	WGS84 geographic coordinate system		1.Same coordinate system with other Map layers 2.Resolution is adjustable	1.The unit is degree 2.Each cell doesn't keep the uniform geodesic size
3	Icosahedral Snyder Equal Area Projection (ISEA43H)		1.Each cell keep the same geodesic area 2.Existed code for creating the grid in multi scale 3.Existed index method	1.The geodesic size of grid is stable 2.The size of the cell is not very flexible
4	Regional projected coordinate system		1.More complex 2.increase the precision in the geodesic area size	1.Divide world into different section with different projected coordinate system

## 2) Visual benefits

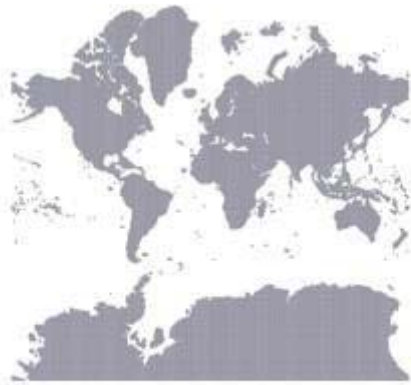
As most online maps are projected in Web Mercator coordinate system, it is important to take the visual benefits into account for hexagon grids created in the same reference system as the presentation platform. Even if the icosahedral hexagon grids enable each cell to keep the same area in reality, there are large shape distortion when grids projected into Web Mercator projection coordinate system in the presentation platform. The closer the grids location to earth's poles, the greater the distortion of grids visually (showed as Figure 3.3-1). This may influence how users interpret the map. To be more specific, if we assign the same colour for cells that have the same population density as an example, users may assume a greatly distorted cell has larger total population and vice versa (Battersby et al., 2017). Tessellating the world in the Web Mercator projection coordinate system presents the shape visually and lessens the chance of misinterpretation.



Hexagon based on Sphere system



Hexagon based on WGS 84 system



Hexagon based on Web Mercator system

Figure 3.3-1 :Comparison of shape distortion in Web Mercator projection coordinate system

### 3.3.2 Interaction

Following the guidance of the interaction design rule (Elmqvist et al., 2011) introduced in the chapter 2 Literature Review, Zooming was selected as the main interaction method on the map view for exploring multi-scale maps. Figure 3.3-2 illustrates how zooming helps to shift maps. On the small-scale map, users could first pan to the area interested then zoom using the mouse wheel to the large-scale map. This interaction is already available in the ArcGIS Web Map. But for organizing the sequence of appearance for multi-resolution layers, it is required to set correlated visible scales for each layer. The detailed information will be given in Chapter 4 Visualization of Hexagon Layer.

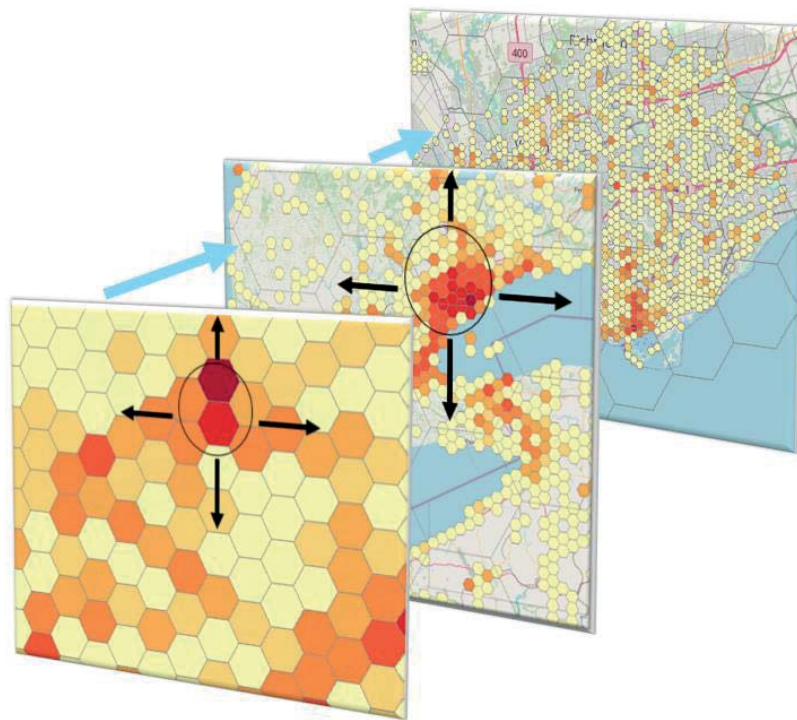


Figure 3.3-2: Zooming for Multi-resolution hexagon layers

### 3.3.3 Visualization of hexagon layers

The visualization ideas from choropleth map was brought into the hexagon layers' visualization, which requires taking both the classification of the map values and colour scheme selection into consideration. Meanwhile, the problem of choropleth maps are also problems for the hexagon layers 'visualization which are mainly caused by the classification method and chosen colour scheme. Different classification can lead to the different spatial patterns in the presentation. Figure 3.3-3 gives the example on the Country Population maps presented under different classification methods. Even though for the same dataset, Map classified by Equal Interval method shows more even distribution characterise than map classified by Quantiles Method.

Depends on the highlighted information need to be expressed via the presentation and the characteristic of the dataset, different cartographers may choose different classification methods for their maps. Detailed description of value classification and selected colour scheme for maps existed in our research are in Chapter 4 Use Case, Visualization of Hexagon layer.



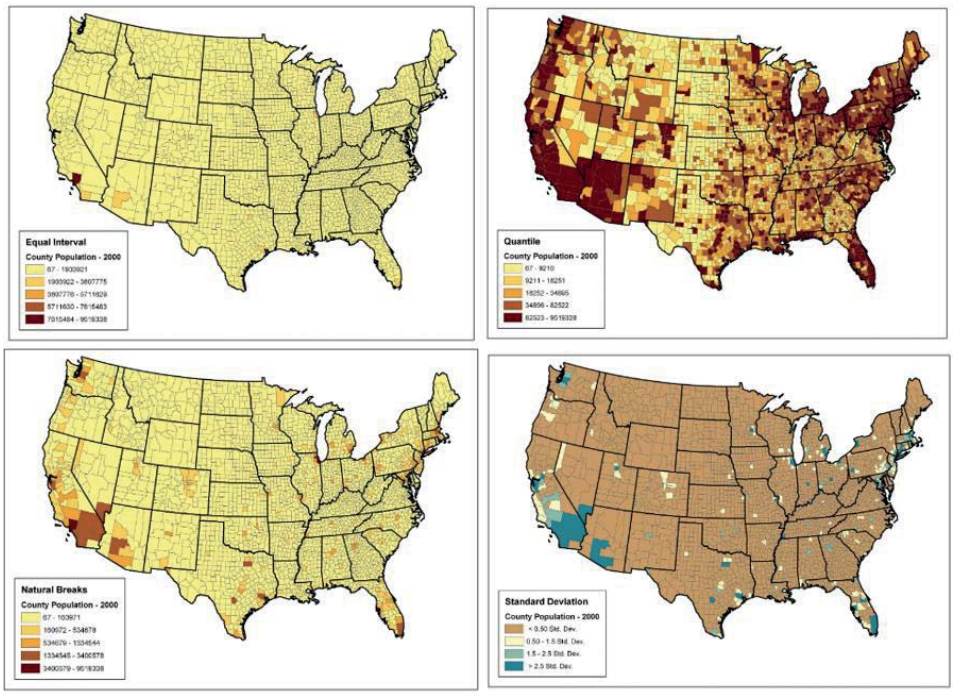


Figure 3.3-3: Differences caused by various classification methods (*Data Classification*, 2018)



## 4. Case study

---

This chapter focuses on implementing methodologies described in the Chapter 3 into a study case. The studied case comes from Catastrophe Management Team in Allianz Corporate and Speciality (AGCS). Contents of this chapter will be divided into four sections:

- 4.1 Background and Data: introduction of project's background and dataset used,
- 4.2 User Requirements Assessments: Methods on acquiring user requirements and assessment requirements which will be used as a reference on the visualizing process,
- 4.3 Visualization Strategy: Design of the application platform and functions in the platform,
- 4.4 Hexagon System: methods on creating the multi-resolution hexagon layers

### 4.1 Background and Data

The catastrophe risk portfolio dataset is composed of insured location information and catastrophe risk metrics. It is an important data source for underwriters and risk managers on steering the business in the insurance industry. Due to the great size of the data, it is impractical to sense the portfolio distribution only based on static tables. Providing a web-based geo-visualization platform to visualize the dataset and support users to interactively explore the dataset require a great demand.

There is a Geo-visualization web-based platform called CATlas- Portfolio Explorer developed by catastrophe risk management team which is currently used for exploring the portfolio dataset. This platform uses global equal-sized rectangular grids to visualize aggregated risk values in the web map. However, for obtaining better user experience, there is a great potential to further optimize the existed platform from different aspects include layout design, visualization methods and platform functions. Thus, this thesis implemented the methodologies from chapter three to develop a new portfolio explorer.

Specifically, the dataset used in this thesis project are Catastrophe risk portfolio dataset provided by catastrophe risk management team in AGCS, which contains pre-calculated catastrophe risk metrics cover four quarters from July. 2019 to April.2020 around world. The data stored as a table format in the Microsoft SQL server. Each record presents one insured location and the count of records per quarter is between 2 million and 2.5 million. Table 4.1-1 shows main attribute columns which are visualized in final maps.

Table 4.1-1: Main columns of location table

Index	Column Name	Description
1	Latitude	Latitude of insured location
2	Longitude	Longitude of insured location
3	TIV	Total insured value
4	NetExp_EQ	Net Exposure calculated for Earthquake
5	NetExp_FL	Net Exposure calculated for Flood
6	NetExp_WS	Net Exposure calculated for Windstorm
7	NetExp_CS	Net Exposure calculated for Convective Storm
8	AAL_PreCat_EQ	Average annual Loss (PreCat) calculated for Earthquake
9	AAL_PreCat_FL	Average annual Loss (PreCat) calculated for Flood
10	AAL_PreCat_WS	Average annual Loss (PreCat) calculated for Windstorm
11	AAL_PreCat_CS	Average annual Loss (PreCat) calculated for Convective Storm
12	RICost_EQ	Reinsurance Cost for Earthquake
13	RICost_FL	Reinsurance Cost for Flood
14	RICost_WS	Reinsurance Cost for Windstorm
15	RICost_CS	Reinsurance Cost for Convective Storm

\* Net Exposure = with all policies (all limits and deductibles, e.g. location and account level, special conditions...) and facultative reinsurance (on location level facultative reinsurance is allocated from policy/account level: Net Exposure Location = (Location TIV / Account TIV) \* Net Exposure Account

PreCat = Gross Loss + Facultative reinsurance (reinsurance for very specific locations and accounts) + per risk treaties (Applies per locations or policy or account – treaties for a whole LoB (line of business)/Portfolio that apply at the level of the risk (risk = locations or policy or account)) + quota share (Pro rate reinsurance - X% of all losses and X% of the premium is transferred to external insurer)

## 4.2 User-centred Design

Referring to the user-centred design loop introduced in the chapter 3, methodologies implemented in each step are specified here (showed as Figure 4.2-1). Figure 4.2-2 expands the loop from Figure 4.2-1 and gives out the general workflow of the thesis project. Although interview was utilized for both user requirement assessment and user testing, there are great differences in beforehand preparation. This section introduces contents of user requirement assessments and user testing, which are the start and end points of the user-centred design loop.

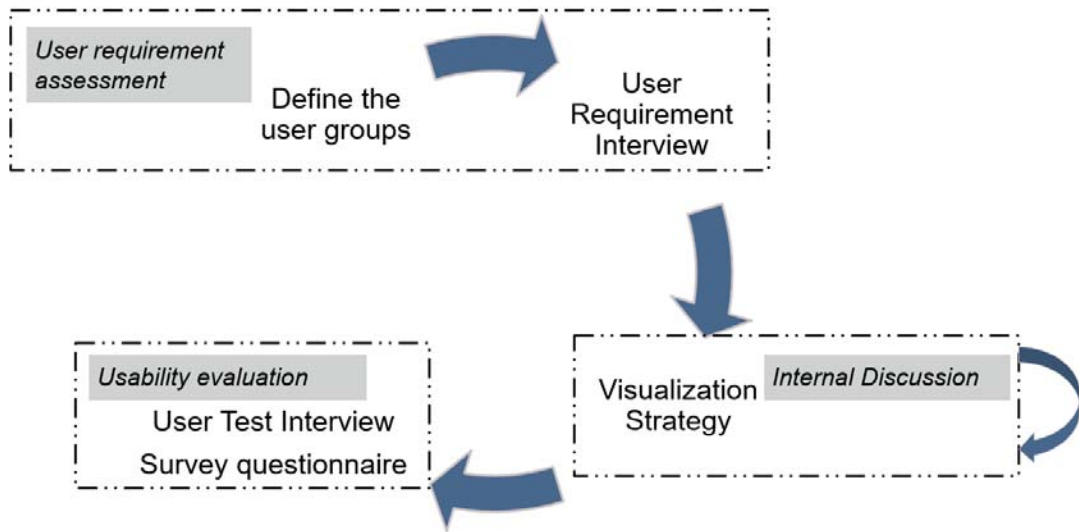


Figure 4.2-1: The procedure for user central design

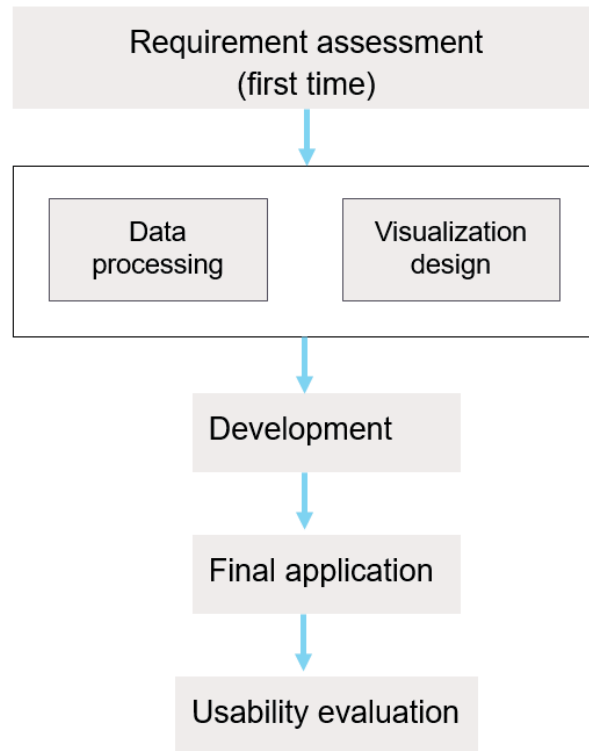


Figure 4.2-2: Workflow of the thesis project

#### 4.2.1 User interview

As the start point of user central design loop, user interview was selected as a way on collecting requirements. Since the final developed platform has limited targeted users who are from specific departments in AGCS, every user was invited to a user requirement meeting separately. Users selected in this research are from the underwriting and Risk

Management team, they also have prior experience using CATlas- Portfolio explorer. Description of users is showed in Table 4.2-1

**Table 4.2-1: User profile**

User Index	Department	Description
1	Underwriting-Property	1. Has experience on using CATlas 2. With Knowledge background on GIS
2	Underwriting -Energy& Construction	1. Has experience on using CATlas 2. Without Knowledge background on GIS
3	Risk Management	1. Has experience on using CATlas 2. Without Knowledge background on GIS
4	CAT risk steering	1. Has experience on using CATlas 2. With Knowledge background on GIS

### ***Preparation of the interview***

According to the research plan, it was expected to construct several classical scenarios after the interview. Thus, interviewees were allowed to have open discussion to share their needs or problems when handling tasks related with portfolio datasets during the interview. However, unlike the free brainstorming, it is also important to ensure the discussed contents are always on topic related and within our interested range. Several questions were prepared to guide the discussion before the meeting. Table 4.2-2 shows questions prepared.

**Table 4.2-2: Interview Questions**

Use case description	Goal / Benefit	Data / Resolution	Comments
What would you like to do? What is your key problem?	Why and where in your process is this beneficial?	What Cat portfolio data/metrics? Spatial and temporal resolution?	Additional comments

### ***During the Interview***

Generally, during an interview, there are three interviewers and one interviewee. Interviews were conducted in the form of online meeting. Among three interviewers (me and my two exterior supervisors from AGCS), there was at least one person that took responsibility for taking notes.

### **4.2.2 User Requirement Assessments**

After finishing interviews, interviewers were gathered together to share meeting notes and create scenarios on using portfolio dataset for each user. Figure 4.2-3 shows scenarios are depicted in graphs.

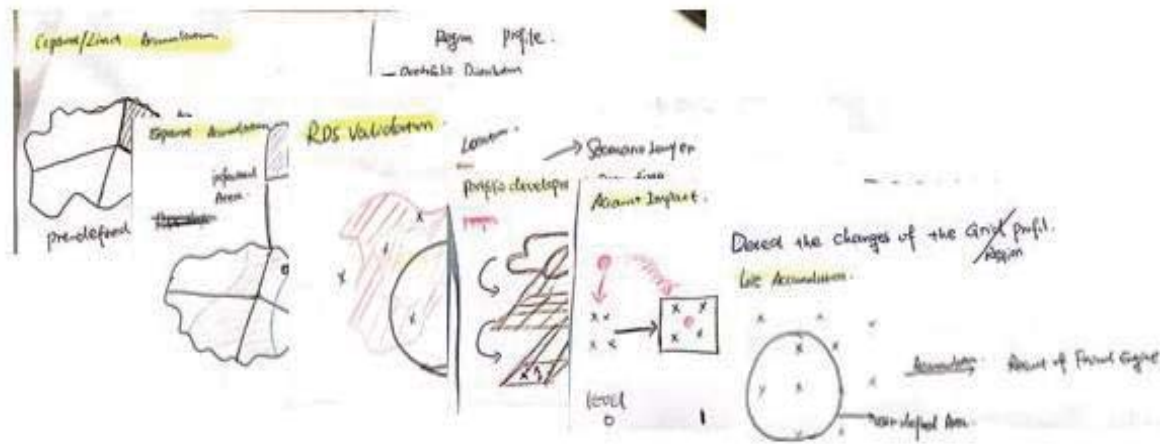


Figure 4.2-3: graphical scenarios

Even though targeted users are from different departments, there are still several overlapping scenarios which are mentioned by all of them. All scenarios are summarized in the table 4.2-3.

Table 4.2-3: Summarization of scenarios

Index	Scenarios	Descriptions
1	View their exposure and risk accumulations on different Levels over time	<ol style="list-style-type: none"> <li>1. Overview on Portfolio development (growth/diversity)</li> <li>2. Detect changes in key metrics (user defined)</li> <li>3. Help to understand changes</li> <li>4. Visualize/Communicate findings</li> </ol>
2	Create manual accumulation views based on a selection	<ol style="list-style-type: none"> <li>1. Refine question from Scenario 1 capacity steering</li> <li>2. Monitor business individually and dynamically in exposure hotspots</li> </ol>
3	Save and monitor an 'Accumulation layer'	<ol style="list-style-type: none"> <li>1. Persistent and precise evaluation of insurance related values at any given time</li> <li>2. Establishing workflows/ guidelines for e.g. Portfolio Management</li> </ol>
4	Estimate the impact of new locations to existing accumulation	<ol style="list-style-type: none"> <li>1. Capacity steering</li> <li>2. Monitoring new &amp; renewal business in exposure hot spots</li> <li>3. Impact estimation of new &amp; renewal business on exposure hot spots</li> </ol>

\*scenario 1 and 2 have been achieved in this thesis, Scenario 3 is built in the application, scenario 4 will be done in the future.

### Solution

Based on these scenarios, a visualization strategy which will be achieved during the visualization application development process was proposed. It contains two main parts:

- 1) Consistent multi-resolution global hexagon layers binning portfolio locations' multivariate values. Both of them are aimed of supporting the procedure of portfolio data exploration,
- 2) Application platform hosting cartographical and infographic presentation in the identical interface.

The application platform will be illustrated in details in section 4.4 and consistent multi-resolution global hexagon layers will be introduced in section 4.3.

### 4.2.3 User Test

User test is the end section for a user central design loop. Via certain user test methodologies, we could gather information for assessing the usability/quality of the application. Due to the limited user numbers in this thesis project, it is feasible in this research to invite users into a meeting and conduct deep discussion with them based on the application result.

This research divided the user test into two rounds.

First, hosting an application demonstration meeting and inviting all users to participate. Before the meeting, brief brochure introducing demonstrated portfolio explorer (test version) was prepared and sent to users. During the meeting, the application was demonstrated and solvable cases in the application also presented to users. Meanwhile, they received a survey questionnaire (showed in Appendix) to answer for gathering their first impression about the application.

Second, Survey on user experience. In this round, users can access to the application online and play it individually. Questionnaire used for this user test round showed as in Appendix.

## 4.3 Hexagon System

A consistent hexagon system works on carrying aggregated points' values in different resolutions. Hexagon layers in different resolution level also correspond to different visible scale ranges. It aims to supporting an explorative procedure that users can identify interested searching range based on lower resolution hexagon layer and dive into the interested region on the higher resolution hexagon layer by simply zooming (Figure 4.3-1 shows the relationship between neighbour hexagon layers).

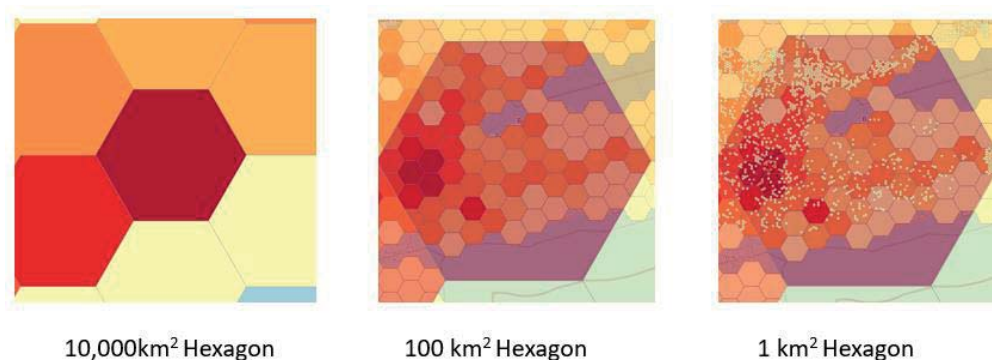


Figure 4.3-1: Neighbour hexagon layers' relationship

### 4.3.1 Hexagon creation

Equal-sized hexagon system works on comparing aggregated values spatially and revealing spatial patterns and clusters of data especially when the data is a great number of points. For this study case, the research created three resolution hexagon layers which are  $1\text{km}^2$ ,  $100\text{km}^2$  and  $10,000\text{km}^2$  as the prototype for binning aggregated location' attribute values. Figure 4.3-2 shows the size of the hexagon.

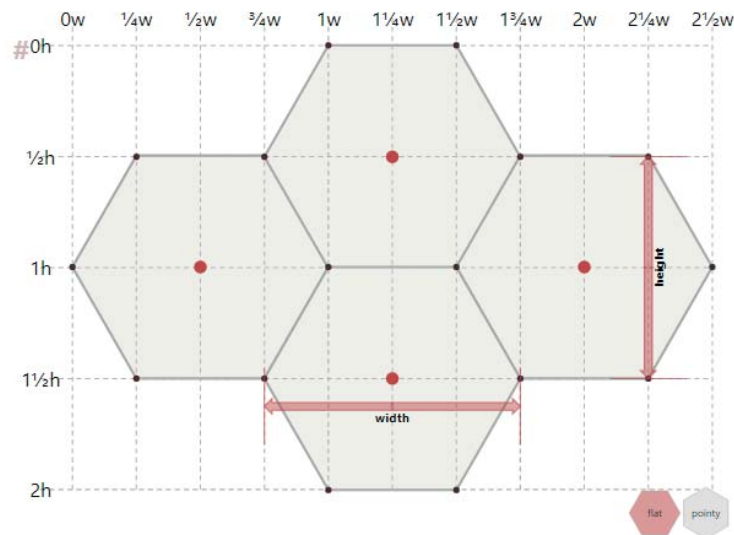


Figure 4.3-2: size and spacing of Hexagon

(Red Blob Games: Hexagonal Grids, 2020)

The hexagon layers reference the same projection coordinate system (Web\_Mercator\_Auxiliary\_Sphere). They are required to be consistent in the geometry and covered the world seamlessly. Figure 4.3-3 shows the geometrical relationship between two neighbour resolution layers. For example, one  $100\text{km}^2$  hexagon cell should intersect with one hundred  $1\text{km}^2$  hexagons. This research set the leftmost node in one hexagon as the start point when creating the hexagons (labelled in Figure 4.3-3). Therefore, to ensure each part of the world could be covered by hexagon layer, the start point for hexagons located along world left edge is out of world range. Afterwards, the exceed parts are clipped by the world shape. Figure 4.3-4 shows hexagons along the world edges. Afterwards, for creating hexagon layers meet above requirements, existed methods (e.g. ArcMap Tessellation tool) are not sufficient. They are limited by either computational memory space nor available hexagon resolution, Thus, based on ArcPy library, a python script was developed (attached in the Appendix) for customized hexagon creation.



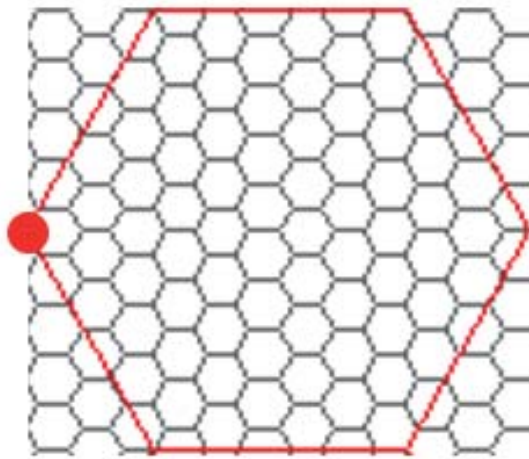


Figure 4.3-3: Hexagon geometric relationship

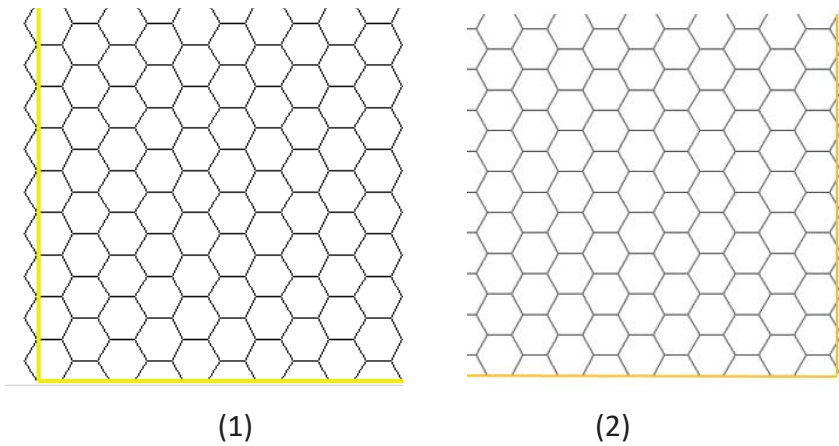


Figure 4.3-4: Hexagons along the world edge

The step after creating the hexagon layers is signing identical index for each hexagon cell. The index system is based on Cartesian Coordinate system (Figure 4.3-5).

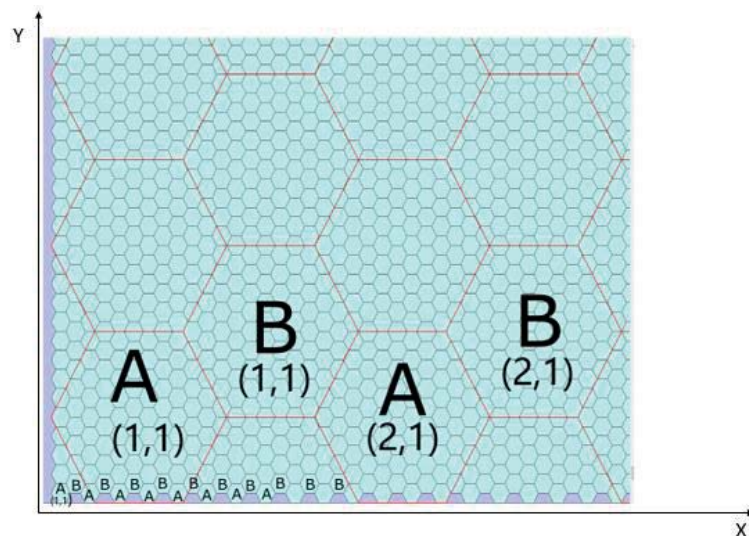


Figure 4.3-5: Hexagon index system



It divided the hexagon into A, B series, the start coordinate for both series are (1,1). When the hexagon index stored in to the table. Its format is structured as 'series-X-Y'. For example, the index for Hexagon A (1,1) will be formatted as 'A-1-1' in the database's table. Based on the index system, the relationship between neighbour hexagon layers can be built in to a matrix (Figure 4.4-6 (1) gives the example of kid hexagons for father hexagon A-1-1, Figure 4.4-6(2) gives the example of kid hexagons for father hexagon B-1-1). These matrixes help on searching interested cells among different hexagon layers faster (code for searching hexagon see Appendix).

			(1,10)	(1,10)	(2,10)	(2,10)	(3,10)	(3,10)	(4,10)			
		(0,9)	(1,9)	(1,9)	(2,9)	(2,9)	(3,9)	(3,9)	(4,9)	(4,9)		
	(0,8)	(0,8)	(1,8)	(1,8)	(2,8)	(2,8)	(3,8)	(3,8)	(4,8)	(4,8)	(5,8)	
	(0,7)	(0,7)	(1,7)	(1,7)	(2,7)	(2,7)	(3,7)	(3,7)	(4,7)	(4,7)	(5,7)	
(-1,6)	(0,6)	(0,6)	(1,6)	(1,6)	(2,6)	(2,6)	(3,6)	(3,6)	(4,6)	(4,6)	(5,6)	(5,6)
(-1,5)	(0,5)	(0,5)	(1,5)	(1,5)	(2,5)	(2,5)	(3,5)	(3,5)	(4,5)	(4,5)	(5,5)	(5,5)
(-1,4)	(0,4)	(0,4)	(1,4)	(1,4)	(2,4)	(2,4)	(3,4)	(3,4)	(4,4)	(4,4)	(5,4)	(5,4)
	(0,3)	(0,3)	(1,3)	(1,3)	(2,3)	(2,3)	(3,3)	(3,3)	(4,3)	(4,3)	(5,3)	
		(0,2)	(1,2)	(1,2)	(2,2)	(2,2)	(3,2)	(3,2)	(4,2)	(4,2)		
		(0,1)	(1,1)	(1,1)	(2,1)	(2,1)	(3,1)	(3,1)	(4,1)	(4,1)		
				(1,0)		(2,0)		(3,0)				
B	A	B	A	B	A	B	A	B	A	B	A	B

(1)

				(6,15)	(6,15)	(7,15)	(7,15)	(8,15)	(8,15)	(9,15)		
		(5,14)	(6,14)	(6,14)	(7,14)	(7,14)	(8,14)	(8,14)	(9,14)	(9,14)	(10,14)	
	(5,13)	(5,13)	(6,13)	(6,13)	(7,13)	(7,13)	(8,13)	(8,13)	(9,13)	(9,13)	(10,13)	
	(5,12)	(5,12)	(6,12)	(6,12)	(7,12)	(7,12)	(8,12)	(8,12)	(9,12)	(9,12)	(10,12)	
(4,11)	(5,11)	(5,11)	(6,11)	(6,11)	(7,11)	(7,11)	(8,11)	(8,11)	(9,11)	(9,11)	(10,11)	(10,11)
(4,10)	(5,10)	(5,10)	(6,10)	(6,10)	(7,10)	(7,10)	(8,10)	(8,10)	(9,10)	(9,10)	(10,10)	(10,10)
(4,9)	(5,9)	(5,9)	(6,9)	(6,9)	(7,9)	(7,9)	(8,9)	(8,9)	(9,9)	(9,9)	(10,9)	(10,9)
	(5,8)	(5,8)	(6,8)	(6,8)	(7,8)	(7,8)	(8,8)	(8,8)	(9,8)	(9,8)	(10,8)	
		(5,7)	(6,7)	(6,7)	(7,7)	(7,7)	(8,7)	(8,7)	(9,7)	(9,7)		
		(5,6)	(6,6)	(6,6)	(7,6)	(7,6)	(8,6)	(8,6)	(9,6)	(9,6)		
				(6,5)		(7,5)		(8,5)				
B	A	B	A	B	A	B	A	B	A	B	A	B

(2)








Figure 4.3-6: Index matrix

### 4.3.2 Visualization of Hexagon layer

There are four main factors influence the visualization effect of hexagon layers: 1) attribute values in the hexagon cell 2) Colour scheme 3) Classification method 4) The scale of layer's

visibility. Reference to the attribute table of the location table, there are twelve main key metrics need to be aggregated into the hexagon layers and visualized in the map. For these twelve variates, the research creates an identical colour scheme for all of them. Table 4.3-1 provides the colour scheme used in the layer. Number of legend's classes is uniformed to seven. The classification method mainly depends on the magnitude of the attribute values. Table 4.3-2 introduces the visible scale for layers in different resolution.

**Table 4.3-1: Colour Scheme**

Class	Colour	
1	#ffffb2	
2	#fed976	
3	#feb24c	
4	#fd8d3c	
5	#fc4e2a	
6	#e31a1c	
7	#b10026	

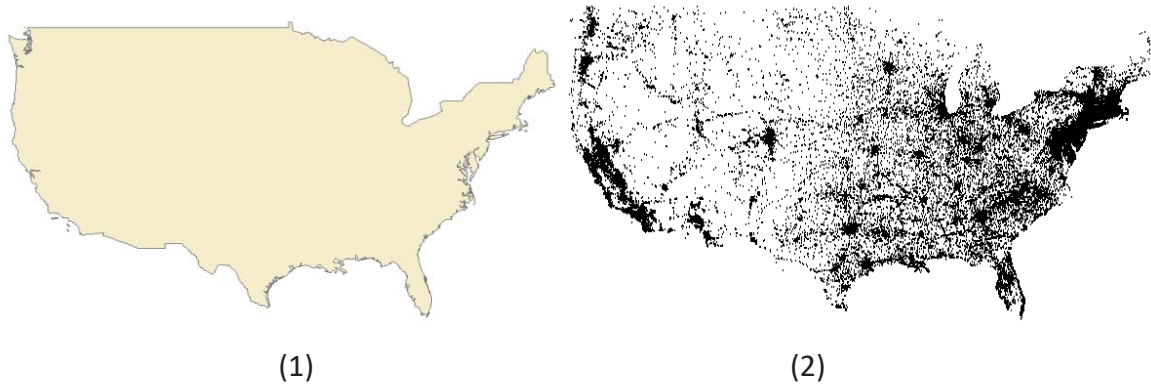
**Table 4.3-2: Layers' visible scale**

Layer	Visible Scale Range
Country	World (1:147,914,382) ----- Countries (1:25,000,000)
10,000 km <sup>2</sup>	Countries (1:25,000,000) ----- State (1:3,000,000)
100 km <sup>2</sup>	State (1:3,000,000) ----- Country (1: 750,000)
1 km <sup>2</sup>	Country (1: 750,000) ----- Town (1: 40,000)
Points	Town (1: 40,000) ----- Building (1: 1,250)

### 4.3.3 Result Comparison

Since the hexagon layer is created in the projection coordinate system, there is always been a distortion in the size. As the absolute value of the latitude increases, the degree of distortion gets greater as well. Before users come to the first hexagon layer, they see the country layer first and zoom into the interested country further. Although the empty hexagon layer is created to cover the world, users only compare hexagon values inside the country range. Considering this prerequisite, for calculating errors in the area size caused due to the planar coordinate system, this research selected a country as experimental region and compared the result between hexagon layers created in Web Mercator and created by Icosahedral method inside the country. The contiguous united states own close to half

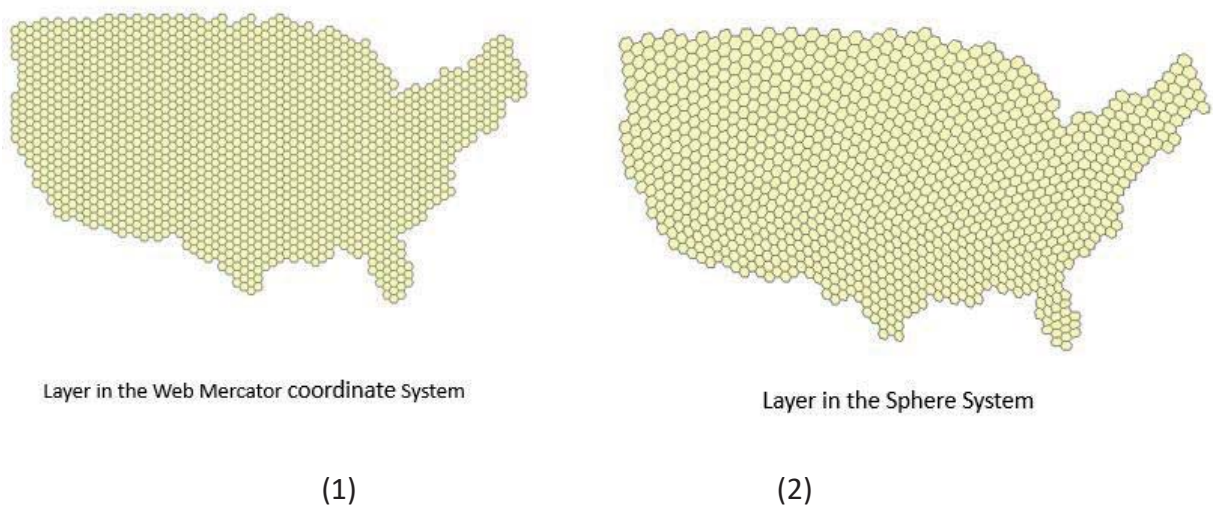
locations around the world and its territory spans more than 30 degrees in the direction of latitude. Thus, the contiguous united states were decided to be used as the experimental area (showed as Figure 4.3-7).



**Figure 4.3-7: The experimental range (1) and portfolio dataset located in the range (2)**

The aim of experiment is testing the capacity of hexagon grid layer in Web Mercator on detecting out high-value locations compared to hexagon grid layer in the sphere. Since the exploration ranges are based on the selection result from lower resolution layer, this experiment only do the comparison for layers with lowest resolution (10,000 km<sup>2</sup>) in the application. The experimenting steps are designed as:

- 1) Creating Hexagon layer for experimental area on Web Mercator coordinate System and the Sphere separately. (code attached in the Appendix, showed as Figure 4.3-8)
- 2) Binning locations' attributes into hexagon layers
- 3) Implementing the same classification method for two layers' attributes
- 4) Selecting areas belong to the first classification category (showed as Figure 4.3-9, 4.3-10)
- 5) Comparing selected results (result showed in Table 4.3-3)



**Figure 4.3-8: Experimental Hexagon layers**

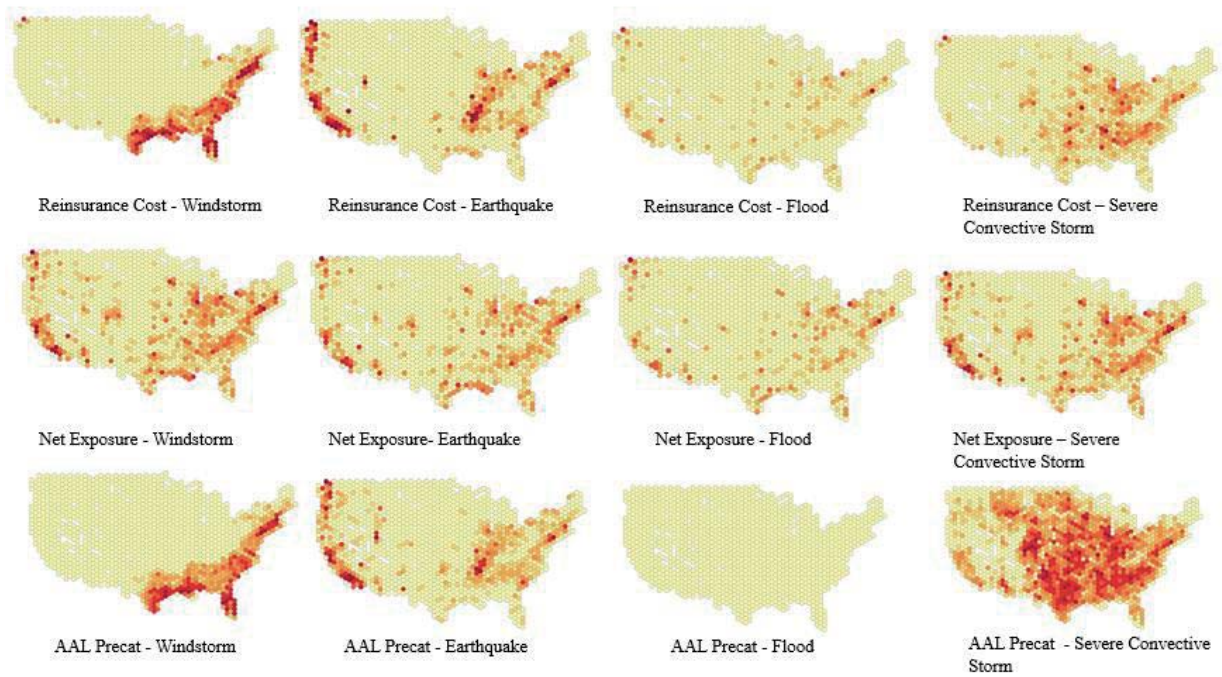


Figure 4.3-9: Result based on Hexagon Layer on Web Mercator Coordinate System

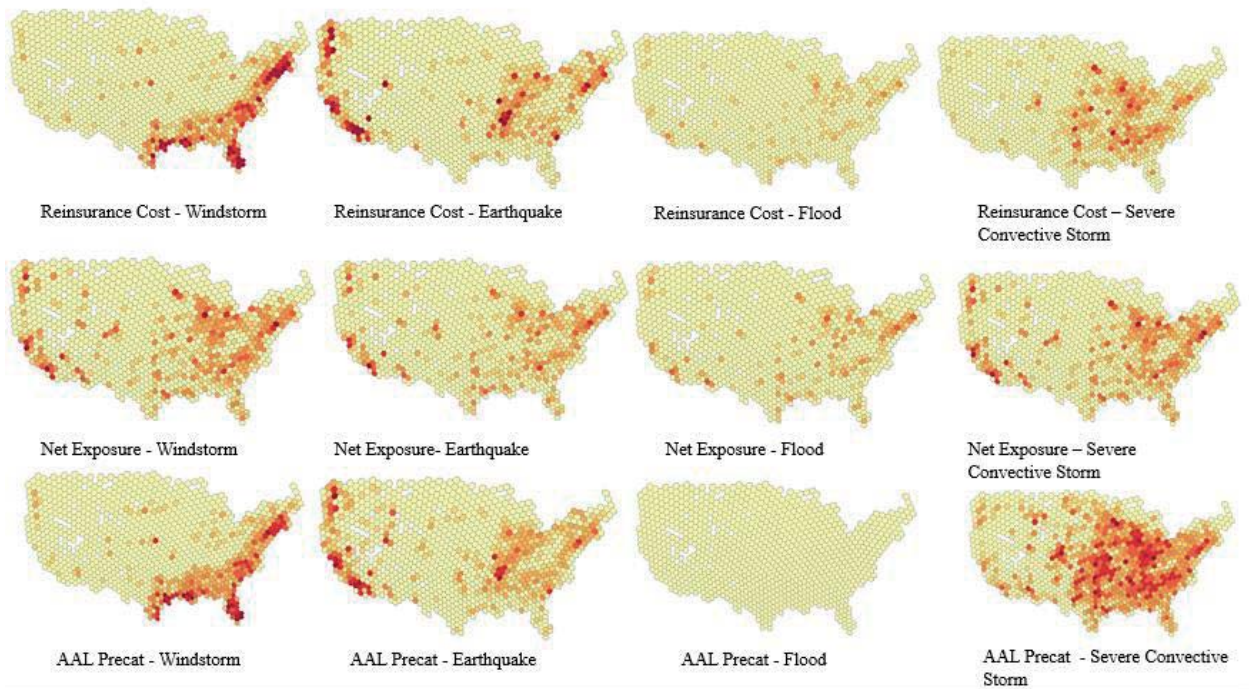


Figure 4.3-10: Result based on Hexagon Layer on the sphere



Table 4.3-3:Result Comparison

Attributes	Number of locations in the first category		Matched Locations
	Web Mercator	Sphere system	
Reinsurance Cost - Windstorm	102396	89067	78200
Reinsurance Cost - Earthquake	129627	175937	109137
Reinsurance Cost - Flood	15093	-	-
Reinsurance Cost – Severe Convective Storm	2863	1305	694
Net Exposure - Windstorm	123886	90670	85069
<b>Net Exposure- Earthquake</b>	<b>15109</b>	<b>43976</b>	<b>4568</b>
Net Exposure - Flood	15109	-	-
Net Exposure – Severe Convective Storm	131183	90670	86769
AAL Precat - Windstorm	61918	47106	44750
AAL Precat - Earthquake	142419	103604	97830
AAL Precat - Flood	-	-	-
AAL Precat - Severe Convective Storm	83903	51573	45642

'-': means there are no available cells have values belongs to the first classification category

Matched Locations: number of locations with the same Location ID appeared in cells in the first classification category

To ensure maps in two system are comparable, the contrast hexagon map (Figure 4.3-6 (2)) was created based on ISEA3H method. The layer’s resolution is in the level 6 which is close to the average hexagon area in layer based on Web Mercator (Figure 4.3-6 (1)). Referencing results from map presentations (Figure 4.3-7 & Figure 4.3-8), there are no great differences on the presented spatial patterns for each attribute on the map between two reference systems. Especially, for areas have relative high values are consistent in two systems. Users can easily identify the interested places possess high aggregated values. Furthermore, cells with values belongs to first classification category were selected for calculating the number of intersected locations. As Table 4.3-3 shows, except Reinsurance Cost-Earthquake and Net Exposure – Earthquake, map in Web Mercator detects more point locations than sphere system for further checking that reduces the probability of omitting locations interested. Except Net Exposure – Earthquake, most of locations (more than 50% in average) detected by layers in sphere system can also be found out in layers in Web Mercator. **To summarize, even though there are area distortions of hexagons in the web Mercator, it is efficient to**

explore large portfolio datasets and detect outliers by multi-resolution hexagon layers in this reference system. However, it is need to take care of detected locations for earthquake, double checking of the raw data in tables is required.

## 4.4 Platform Design

The designed application platform aims at enabling users to check values both in map and statistical charts interactively.

### 4.4.1 Software used

Recently, there are different commercial or open source software/API available for processing GIS tasks. Among them, ESRI products (including ArcMap, Map server, ArcGIS Online etc.) provide multiple solutions for users from data preparation to map publication. Especially, ArcGIS Web AppBuilder (developer version) enables users to configure a HTML app in the ArcGIS server and support developers to create customized widgets and themes. In the This research set up the development environment based on ArcGIS platform, uses ArcGIS Web AppBuilder (developer version) to create the web application and published the final application in the ArcGIS server. Table 4.4-1 gives detailed information on the tool used in each process.

Table 4.4-1: Processes and tools in the research

Process	Description	Tool
<b>Data Processing</b>	1. Build Hexagon System	Python
	2. Aggregate data	ArcGIS Desktop 10.5
	3.Mapping	
<b>Web Application</b>	1. Publish map layers	ArcGIS Online
	2. Create Web Maps	ArcGIS Web AppBuilder (Developer Edition)
	3. Develop the Web Application	ArcGIS Server Manager

### 4.4.2 Platform layout

The design of the application layout inherited the style of default Jewelry Box Theme (one type of web app theme which puts main widgets in the left of the app) from ArcGIS web app builder. Figure 4.4-1 is the prototype of application layout. Figure 4.4-2 is the final effect of the application's interface.

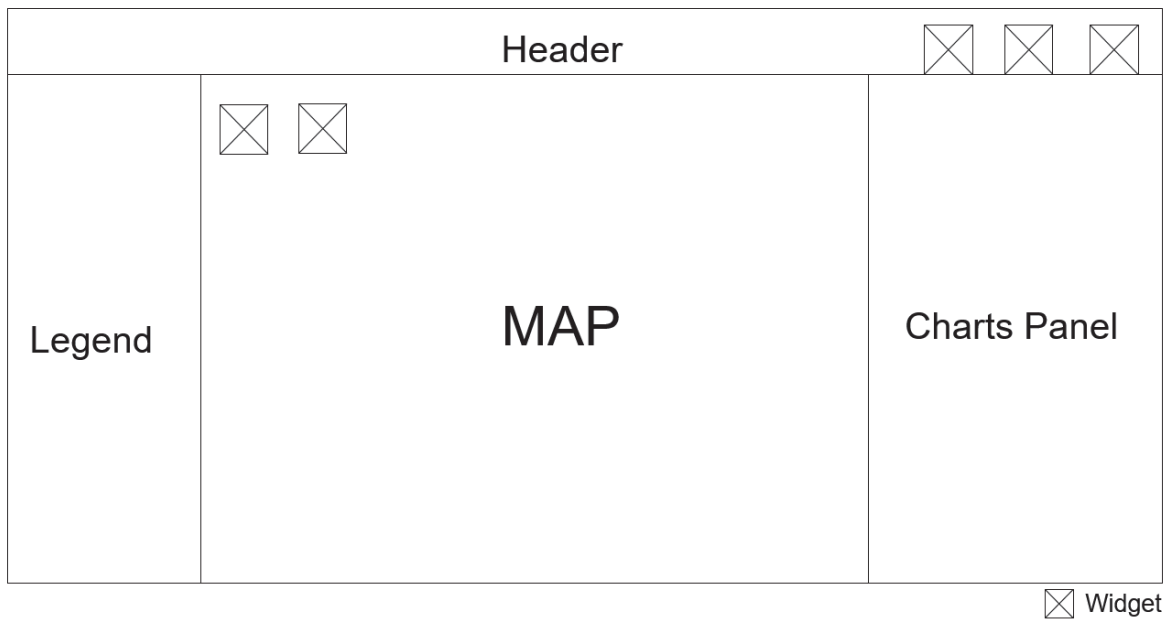


Figure 4.4-1: Layout design

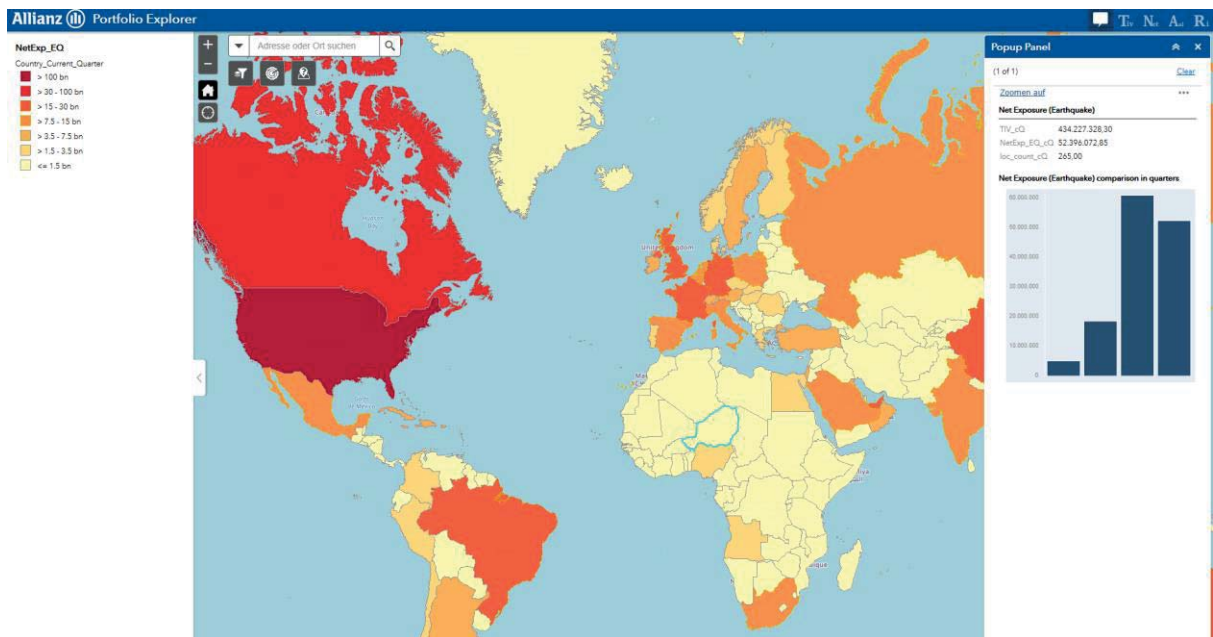


Figure 4.4-2: Interface of Application

#### 4.4.3 Widgets

ArcGIS Web AppBuilder (developer version) provides a wide range of widgets ready for configuration and also allows users to self-develop customized widgets. These widgets consist of functions and parameters can easily to be used in the Web APP. Corresponding to functions suggested by user requirements assessment section, there are four main widgets proposed to be added in the final application. They are Popup Panel, Automatic Legend, Customized Layer Selection and Manual Accumulation. Among them, Automatic Legend and

Manual Accumulation are default widgets provided by ArcGIS web APP builder. Figure 4.4-3 describes detailed functionalities for each widget.

### Layer Selection

By this widget, users can select layers they want to check based on interested attributes and perils



### Query Locations

This widget supports users get interested locations based on a series of selecting criteria



### Popup Panel

By this widget, users can select features they want to check based on interested attributes and perils, as well as development of values over time



### Manual Calculation



This widget supports users to manually define an area and come out summary charts for Cat metrics over the last four quarters



### Situation Awareness

This widget supports users to manually define an area and get back summary statistics for pre-defined attributes.



Figure 4.4-3: Functionalities of Widgets

#### 4.4.4 Scenarios achieved in the application

After achieving the multi-resolution hexagon layers and building them into the application which also integrates several customized widgets in, this section references back to scenarios summarized from user requirements in section 4.2.2 and demonstrates two of them which can be realized in the final application.

##### Scenario 1

##### Viewing the exposure and risk accumulations on different levels over time

This scenario is regarded as the most basic but important one to support users on grasping the distribution of the portfolio (based on interested metrics), detecting high-valued locations and steering how value changes with the time. Compared to requiring data from fixed tables, this scenario makes full use of the advantages from map presentation to help



users to get deep insights into the portfolio dataset over time and space in short time. Specifically, it can be accomplished in the application by the following process:

1. selecting interested metrics and desired layers using Layer Selection widget (Figure 4.4-4).
2. Navigating to high-valued polygons (in deep hue colour) via simply zooming and panning. By switching the map view from smaller scales to larger scales, users are able to focus on some candidate locations which are possible to have great influence in countrywide or even worldwide portfolio distribution (Figure 4.4-5).
3. In the zooming process, clicking interested hexagons or locations to compare the value in latest four quarters (Figure 4.4-6).

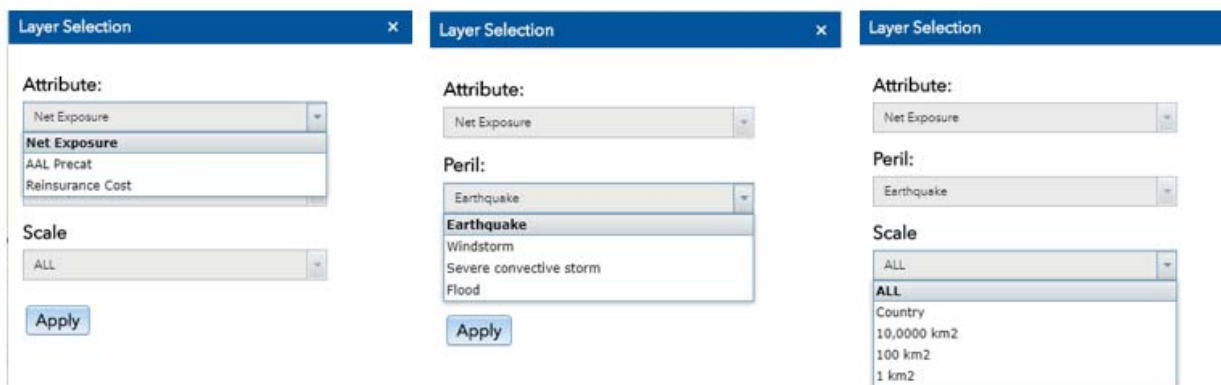


Figure 4.4-4: Select interested metrics and layer scale

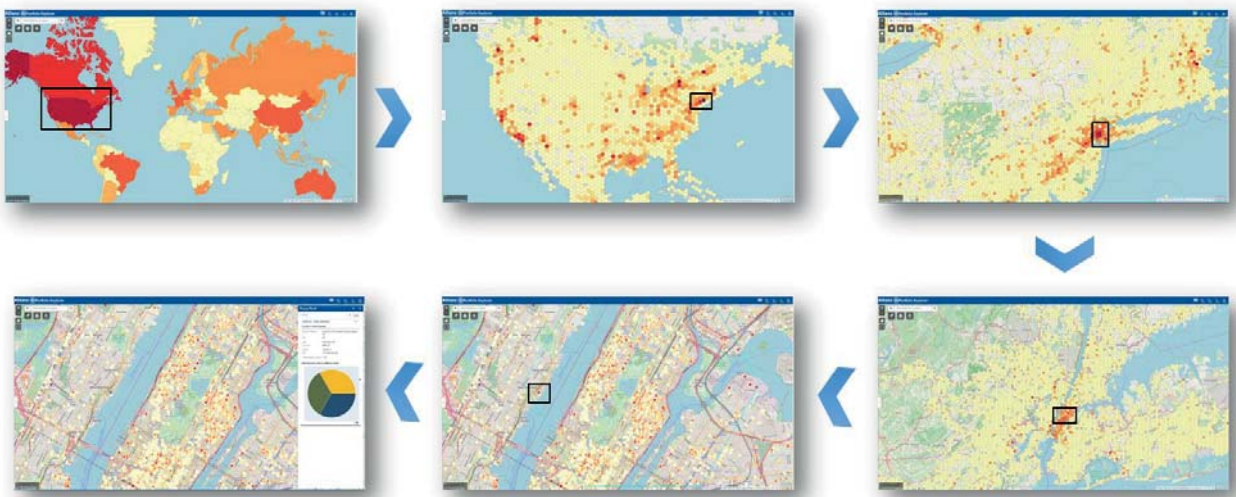


Figure 4.4-5: viewing risk accumulations on different levels

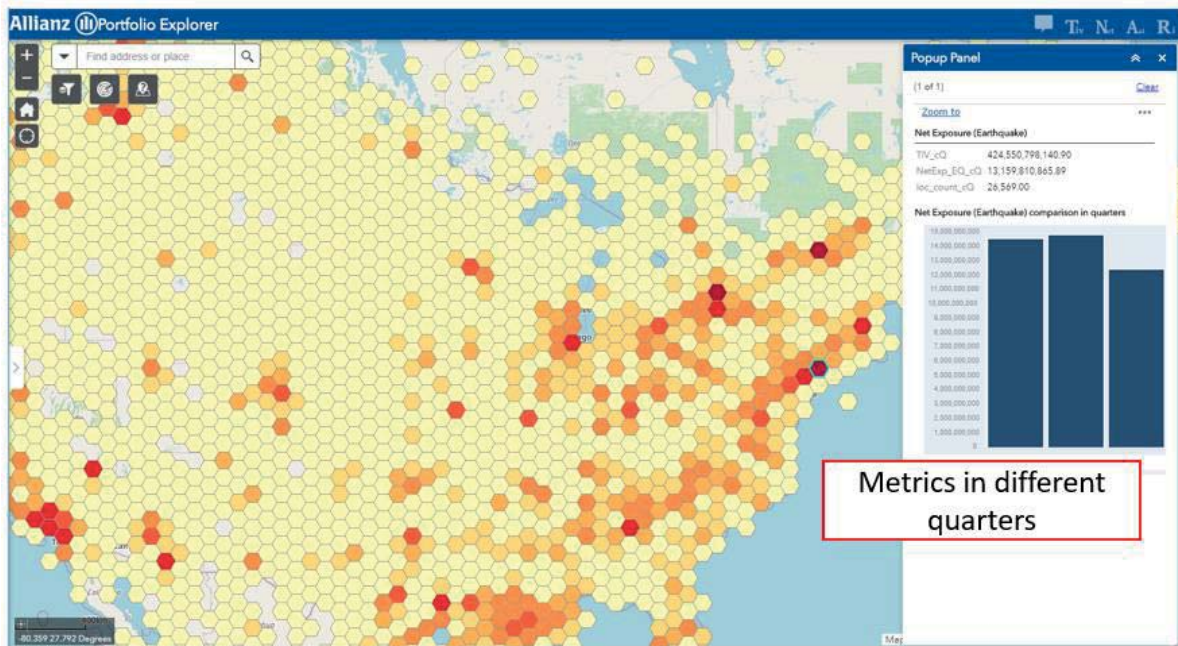


Figure 4.4-6 Detecting changes of metrics in quarters

## Scenario 2

### Create manual accumulation views based on a selection

The scenario above accomplishes the exploring process via checking values in individual hexagon or location which are pre-calculated. However, it is possible the area covered by one hexagon is smaller or bigger than the area that users want to check. Therefore, users are also expected to be able to define the area manually and check metrics including the trend inside the area. For this, there are two widgets available in the application.

1. Manual Calculation: manual calculation is divided by four key metrics and enables users to process the calculation in different resolution layers by simply drawing the shape. Line chart illustrates the trend of selected metric in different perils will be returned in the result panel. (Figure 4.4-7)
2. Situation Awareness: by using this widget, users are able to get the count of location, sum of all metrics in the self-defined area. The results are presented in the text and can also be exported as CSV file. (Figure 4.4-8)

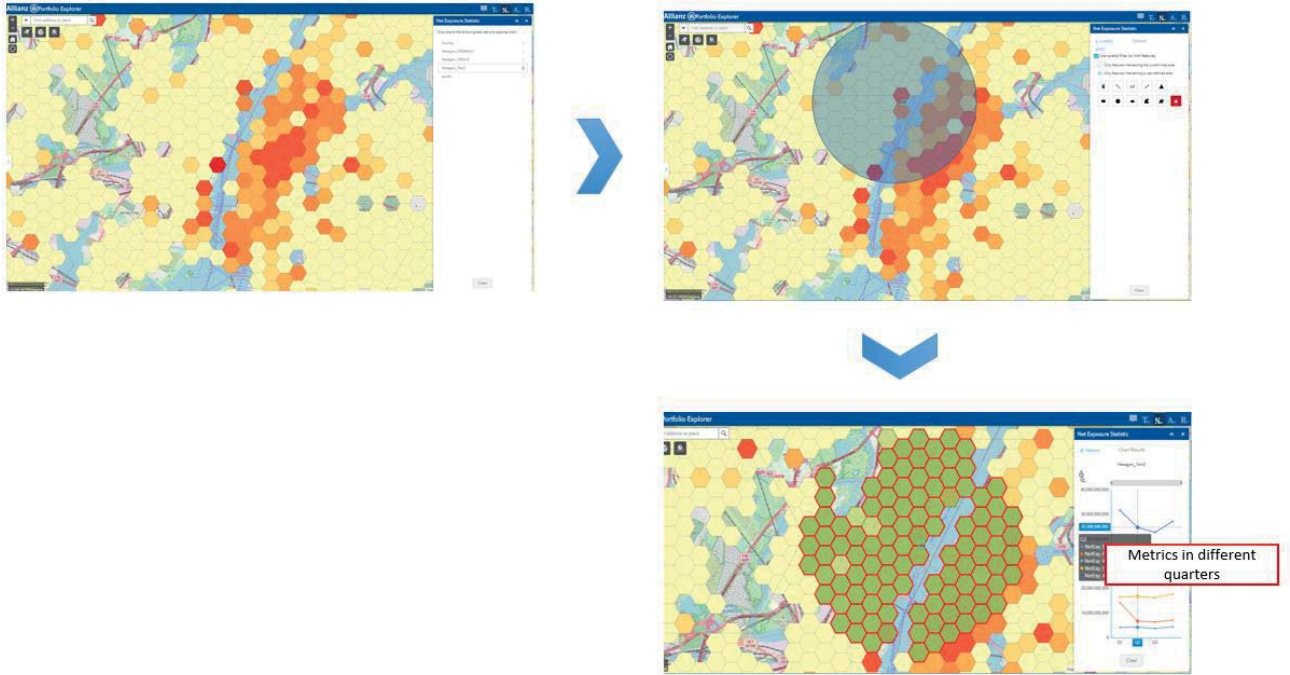


Figure 4.4-7: Process on Manual Calculation

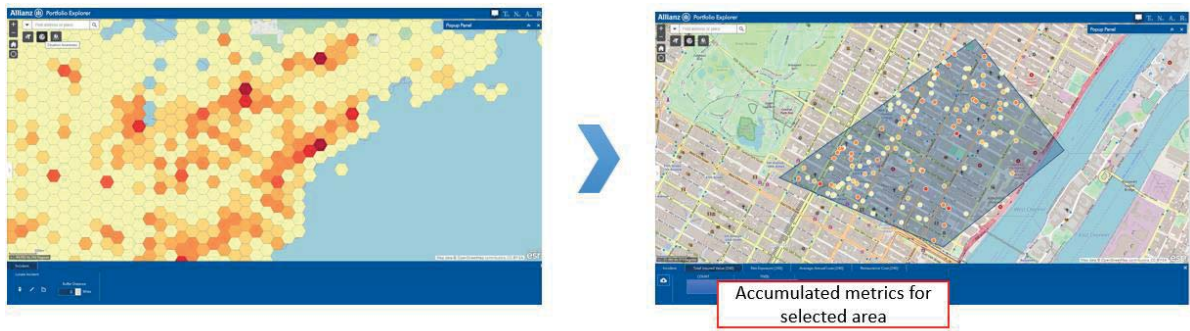


Figure 4.4-8: Process on Situation Awareness

## 5. Result

---

This chapter focus on summarizing the result gained from this thesis research. It includes two main parts:

5.1 Thesis application result which realized the theoretical research in the Web-based Geo-visualization platform,

5.2 Results of User test which are the important materials for evaluating the application result and guiding the future work.

### 5.1 Thesis Application Result

This thesis research developed a decision-aided Geo-visualization application aiming on helping users to better understand the large spatial datasets. The whole research process is based on the use-centred design concept and implemented multi-scale map visualization for supporting spatial data exploration process.

The proposed working flow on achieving user-centred design combing multi methods for user requirements assessment and user test that is dedicated for intra-company short-term project. On achieving multi-scale map visualization, this research developed a multi-resolution global hexagon layer system to aggregate attributes into different scales. The system is extendable for multi variates and time stamps.

### 5.2 Results of User test

Two rounds user test were executed for testing the application with targeted end users. The first round processed during the interview and aims at getting the feedback on the first impression, map design and function design. The second round went with the online questionnaire attached with available application in the server. This round of test provided sufficient time for users to explore the application and return the comments and feedback on using experience.

Regarding to the limited survey samples in the user test, it is only possible to summarize out the qualitative result from answered questionnaires and comments. However, they can still be supportive materials for us to analyse user experience and assess our final application in return. Table 5.2 -1 below is the result for the first-round user test.



Table 5.2-1: Result of first round questionnaire

<b>Question 1</b>	<b>Just based on the map, features in which colour arouse your interest ? (Multi -selection)</b>					
<b>Options</b>	Red	Green	Others	I don't care colour		
<b>Number of users</b>	1	0	1	1		
<b>Question 2</b>	<b>Which layer/layers will you spend more time to explore? (multi-selection)</b>					
<b>Options</b>	Country layer	10,000 km <sup>2</sup> layer	100 km <sup>2</sup> layer	1 km <sup>2</sup> layer	Points layer	
<b>Number of users</b>	1	1	1	1	0	
<b>Question 3</b>	<b>Do you think layer zooming will help you find out interested locations ?</b>					
<b>Options</b>	Definitely yes	Probably yes	Might or might not	Probably not	Definitely not	
<b>Number of users</b>	0	1	0	0	0	
<b>Question 4</b>	<b>Do you think manual calculation function will be useful for your work ?</b>					
<b>Options</b>	Definitely yes	Probably yes	Might or might not	Probably not	Definitely not	
<b>Number of users</b>	0	0	1	0	0	
<b>Question 5</b>	<b>On which layer/layers will you do manual calculation ? (Multi-selection)</b>					
<b>Options</b>	Country layer	10,000 km <sup>2</sup> layer	100 km <sup>2</sup> layer	1 km <sup>2</sup> layer	Points layer	I don't do manual calculation
<b>Numbers of users</b>	0	0	1	0	0	1
<b>Question 6</b>	<b>How much you care about the comparison of values in different quarters ?</b>					
<b>Options</b>	A great deal	A lot	A moderate amount	A little	Not at all	
<b>Number of users</b>	1	0	0	0	0	

Based on the above questionnaire result, it is possible to come out the summarization that the designed portfolio explorer (test version) gave a good impression among users. They are glad to see the provided in the platform especially the zooming and panning interaction to help on exploring interested locations.

After collecting all comments provided by users, the suggestions are mainly focus on two aspects:

- 1) Aggregating more attributes on the multi-resolution hexagon layers

2) Providing more filter criteria for selecting layers.

Table 5.2 -2 below is the result for the second-round user test, (1) for result on questions with options, (2) for result on questions requires comments.

**Table 5.2-2: Result of the second-round questionnaire**

(1)

<b>Question 1</b>	<b>Was the interface easy for you to understand ?</b>				
<b>Options</b>	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult
<b>Number of users</b>	1	2	1	0	0
<b>Question 2</b>	<b>Were hexagon layers easy for you to understand ?</b>				
<b>Options</b>	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult
<b>Number of users</b>	1	3	0	0	0
<b>Question 3</b>	<b>Was it easy for you to find some interested locations by zooming hexagon layers?</b>				
<b>Options</b>	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult
<b>Number of users</b>	1	1	0	2	0
<b>Question 5</b>	<b>How useful is the information in the pop up panel for you ?</b>				
<b>Options</b>	Extremely useful	Very useful	Moderately useful	Slightly useful	Not at all useful
<b>Number of users</b>	0	1	2	1	0
<b>Question 7</b>	<b>How do you think the rendering speed of the application ?</b>				
<b>Options</b>	Excellent	Good	Average	Poor	Terrible
<b>Number of users</b>	0	3	1	0	0

**Table 5.2-2: Result of the second-round questionnaire**

(2)

<b>Question 4</b>	<b>What other attributes you want to have in hexagon layers ?</b>
<b>Comments</b>	<ol style="list-style-type: none"> <li>1. Ability to filter and show By LoB, By Region and by LoB &amp; Region metrics</li> <li>2. Number of Locations and Accounts Average and Mean Account TIV. This allows me to quickly identify if I have an accumulation topic or not All values for my LOB only. AGCS totals are not relevant for me. Better would be putting my LOB's share of total in brackets after the LOB values. This would put it into a useful context</li> <li>3. It would be helpful to have the location export functionality from selected hexagons (and sub hexagons). Or more the in app. Maybe just user error but had a hard time viewing this without exporting to csv.</li> <li>4. Split by line of business!!! Number of accounts Mean and average location value</li> </ol>
<b>Question 6</b>	<b>What other information you want to have in the pop up panel?</b>
<b>Comments</b>	<ol style="list-style-type: none"> <li>1. Top Accounts, Top Locations</li> <li>2. As described in previous: number of locations and accounts, etc.</li> <li>3. Maybe not the pop-up panel but the table view of the selected hexagon is helpful. Or maybe descending Net exposure or whichever main attribute.</li> <li>4. as stated before</li> </ol>
<b>Question 8</b>	<b>What was the one thing you liked the most about the application?</b>
<b>Comments</b>	<ol style="list-style-type: none"> <li>1. Simplicity</li> <li>2. Ease of use compared to other CATlas application</li> <li>3. I do like the hexagon aggregation. The one concern with it might be at the edges and being fixed. A spider will move if the exposure moves, which in some cases is correct.</li> <li>4. seamless / automatic adjustment of resolution</li> </ol>
<b>Question 9</b>	<b>What was the one thing you disliked the most about the application ?</b>
<b>Comments</b>	<ol style="list-style-type: none"> <li>1. Not yet fully optimized wrong colours, symbol and "widgets", but it shows huge potential</li> <li>2. Too focused on CAT risk metrics. Ideally, we would bring in non-modelled information or AAL estimates too</li> <li>3. I did still experience some speed and rendering lag. This may be due to latency to the US.</li> <li>4. not (yet) personalized to my LOB - no breakdown by AGCS regions</li> </ol>

**Table 5.2-2: Result of the second-round questionnaire**

<b>Question 10</b>	<b>Apart from existed layers, do you have other layers expected to be added in the application ?</b>
<b>Comments</b>	<ol style="list-style-type: none"> <li>1. see by LoB and Region; market data to understand AGCS share (of course difficult to get); Premium data</li> <li>2. non-modelled AAL estimates</li> <li>3. Would be interesting to see aggregated and detailed hazard view within the hexagons. This might also be useful for fast/automated analysis of new accounts coming into the portfolio from a hazard and net exposure view.</li> <li>4. layer or at least option to drill down by AGCS Region</li> </ol>
<b>Question 11</b>	<b>Please add your further comments.</b>
<b>Comments</b>	<ol style="list-style-type: none"> <li>1. Great work on the data side; a bit more fine required to ensure the UI supports a smooth user journey while completing the workflows</li> <li>2. Correct the filter so I can select my LOB. Remember default setting (I don't want to select my LOB every time)</li> <li>3. Thank you for putting this together.</li> <li>4. Good contribution to the advancement of GIS at AGCS! Hopefully we can keep the momentum going and make something of it!</li> </ol>

During the second-round test, users were provided sufficient time on exploring the application and could provide more comprehensive feedback.

Regarding to the map and application itself, they are easy to be understood and used which achieved one of our preliminary goal that supporting users to use spatial data along their daily work via user friendly Geo-visualization application. According to the users' comments, they all agreed that current application is easier to be used compared to the original CATlas portfolio explorer. The reason can be easily recognized by the comparison between these two applications from three aspects: Application Layout, Map Presentation and Functionalities (Table 5.2-3).

**Table 5.2-3: Comparison between two applications**

	<b>CATlas portfolio explorer</b>	<b>Thesis Application</b>
<b>Application Layout</b>	Foldable Theme	Jewelry Box Theme
<b>Map Presentation</b>	Two static resolution rectangular grids	Dynamic Multi-resolution hexagon layers
<b>Functionalities</b>	<ol style="list-style-type: none"> <li>1. Layer Selection</li> <li>2. Manual calculation for one quarter</li> <li>3. Other default widgets from ArcGIS WebApp builder</li> </ol>	<ol style="list-style-type: none"> <li>1. Layer Selection</li> <li>2. Automatic attribute panel</li> <li>3. Manual calculation for four quarters</li> </ol>

\*Foldable theme: the default app theme in ArcGIS Web Appbuilder which enables all kinds of widgets can be added in the app.



Jewelry Box theme: another app theme in ArcGIS Web Appbuilder which is inherited from foldable theme. It places main widgets on the left of the app.

For these two applications, there are not great difference on the layout, both utilizes the foldable panels to present calculated results and legends. In the map presentation, new application replaces the rectangular grids by multi-resolution hexagon layers to support the dynamic exploration process. Such presentation takes the advantage of switching the details of map consistently in the exploration process. CATlas portfolio explorer provides more available functions than the new application. However, users still regard the new application is easier to use and support their work. This reflects the truth that more available functions do not mean the better usability. For some functions, users may do not know which scenarios are suitable to be applied in. It can also overwhelm users to understand the application. Thus, putting efforts on improving the capacity of one function and building the connection among different functions can be more necessary.

However, when go into detailed comments left by users, there are different opinions on how the application could help them for their work. Users working more with data in global vision treat the application be more practical for them compared to users working more with the regional data. In addition, different users have different concerns on variables presented in the map. Someone regards information presented in maps is usefully for them, but others suggest maps need to include more attributes. Even though the designed layer system and the application grant them more capacity to explore the dataset, they still cannot be put into use on assisting users' real work right now: Fortunately, this problem can be figured out by reorganizing the map data structure without changing the design from the scratch. This truth also gives the lesson for developing application targeted to multi-department users that ensuring the extensibility of the design in the beginning is necessary.

## 6. Discussion

---

### 6.1 Answers to Research Questions

Back to the research objectives introduced in the Chapter 1 introduction, the research aims on facilitating an effective information communication via multi-resolution visual layer-based visualization strategy for multivariate spatial temporal datasets. The main objective was divided into several sub-objectives and questions. In the end of the thesis, answers to these questions can be found one by one.

*RQ1: How to integrate the idea of User Centred Design (UCD) into work?*

- 1) *What is the structure of UCD?*
- 2) *Which methods used for assessing user requirements?*
- 3) *Which methods used for conducting user test?*

Iterating users' feedback are regarded as the core part in the loop of UCD. With the development of UCD practice, several complete and sustainable frameworks are available on guiding the design/development work to implement UCD. Even though, there are no rigorous requirements in the concept of UCD to rule concrete steps and methods of implementation. They are always diverse in different practices. In this research, there are four main steps set in the workflow which is started with user requirements assessment and ended with user test. Iterations on getting the instant feedback for current work are happened in the development process within the team. Many methods were used for assessing user requirements and conducting user test in existed applications. Based on the number of users, this research selected user interview as the main approach to collect users' requirements and feedback. In addition, brainstorming and questionnaire survey also played a role on the whole process. More concrete theoretical contents can be found in Chapter 2, literature Review and the sophisticated UCD plan in this research can be found in Chapter 3, Methodology.

*RQ2: What is the visualization demand from intended user groups?*

- 1) *What are the characteristics and use cases of different user groups?*
- 2) *At which point in the decision-making process will users need visualization support?*

In this research, the main objective is providing the visualization strategies for the large location dataset (AGCS CAT portfolio dataset as use case), thus the selected end users also come from AGCS correspondingly. Their works are about steering the portfolio risks from different perspectives. Of them, it is important to explore large portfolio dataset and find out high-valued location in the map application. A web-based Geo-visualization application is required to be used for supporting spatial data exploration. Most of them have limited and even non-GIS knowledge background, thus, interpreting complicated spatial pattern into easy understood map presentation also requires high demands. Different users have

different expectation on using the application. Someone hopes to have the visualization tool to assist them along the whole workflow, others may check it only after the great catastrophe event happens. Based on the experience from this research, it is challenge to come out a comprehensive solution meet all users demands. Alternatively, summarizing out a common scenario and problem on understanding the dataset is more practical on guiding the future work. The detailed answer for RQ2 can be found in Chapter 4 Case study section 2 user central design and Chapter 5, Result.

*RQ3: How is the visualization strategy structured?*

- 1) Which factors need to be considered for creating the visualization strategy?*
- 2) Which visualization technique is suitable for the large point dataset?*

Elementary factors influence the content of a visualization strategy include the characteristic of dataset and users' requirements. Other aspects such as development environment, number of users are taken into consideration when the development work starts. After assessing the dataset and user requirements, this research proposed a visualization strategy which is composed of a consistent multi-resolution global hexagon layer system and a web-based visualization application platform. The layer system works as a visual prototype for guiding the data procession and visualization. The application platform used for hosting the map presentation and all functions. One of the main problems on visualizing the massive amount point dataset is the overcrowding of points in the same map presentation. Although, some cartography techniques like Heat Map solved the problem through clustering. However, Heat Map is not available on indicating the attribute values behind locations. The multi-resolution hexagon layer system proposed in this research absorbs the idea from multi-scale map visualization and provides a visual prototype for the large point dataset. It can be used for aggregating the value of points in different levels and support users to explore the dataset in different detail level. Chapter 4, Case Study has more concrete description of the visualization strategy.

*RQ4: How is the effectiveness of the proposed visual strategy?*

- a) How is the usability of the Geo-visualization application based on proposed visual strategies?*
- b) What functions can be implemented based on the strategy?*

In the end of the research, two rounds of user test were executed to assess the usability of the final application for reflecting the effectiveness of the proposed visual strategy. According to the result of user test (Chapter 5), the application gets the positive feedback from end users especially for the map presentation, they thought aggregating large quantity of locations into multi-resolution layers help them to explore the portfolio dataset. One common feedback is comparing to the CATlas – Portfolio explorer, application developed in this thesis research is much easier for users to use and understand. Available functions in the application are designed based on the scenarios summarized from user requirements,

they are dedicated to this study case and serve for the CAT risk portfolio dataset. Information on functions can be found in Chapter 4, Platform Design section.

## 6.2 Hypothesis Verification

Through reviewing the research process and results obtained in this thesis, this section goes to discuss the validity of the initial hypothesis.

- *Hypothesis 1: user centred design works as great function on supporting to grasp the user requirements and digging out key issues on communicating spatial-related information with users especially for whom without GIS-knowledge background.*

The Hypothesis 1 can be easily verified by the user test result (chapter 5) which acquires most of positive feedback from end users on their using experience. Especially when comparing the current application with CATlas –portfolio explorer, the advantages of designing the visualization strategy and functions to be aligned with user requirements get more distinguished. Users have the same agreement on the current application is easier to be used and understood compared to the CATlas – portfolio explorer. At same time, if recalling the whole process, the design and development work got changed/adjusted for several times according to the instant feedback which also decreases the risk of changing everything from the scratch in the end. To summarize, visualization strategy and application designed based on users' requirements can make the function on supporting the information communication.

- *Hypothesis 2: A visualization strategy consist of multi-scale map visualization method will support users to understand large point datasets efficiently.*

Specifically, for achieving the multi-scale map visualization in the presentation, the visualization strategy in this research takes the method of building a multi-resolution hexagon layer system. Through it, users are able to access the information in different detail levels automatically. Compared to the static layers, the multi-resolution hexagon layers enable users to explore the information of the dataset in the presentation by simply zooming and panning. That is easier to be executed and does not interrupt users to switch the map view by themselves.

## 7. Conclusion

---

### 7.1 Outlook

Based on the achieved result and user's feedback in this research, the future work can focus on following aspects:

1) Embedding the map presentation into a Dashboard platform

Due to the time restriction, this research didn't achieve to integrate the map and functions into a Dashboard platform. However, since dashboard platform provides interactive interface space for presenting the statistical results more straight forward, it is still worth on designing a suitable dashboard platform for the application in the future.

2) Extending the available attribute in the hexagon layer

Even though, the visualization strategy proposed in this thesis serves the benefit on bridging the massive spatial data information with end users, there is a still a long road to put the idea into the real work. One of them is enabling more attributes into layers according to users' requirements and indicate the location numbers in the cell as well.

3) Organizing layers in better structure

The structure of the map data direct influences on:

1. The performance of the web map
2. Potential functions could be added into the platform
3. The extensibility of the map presentation.

Currently, there multi-resolution hexagon layer came with the research is only a visual prototype for the presenting multi variates. With the increasing numbers of variates to be loaded in the prototype in the future, a well-designed data structure is required to

4) Implementing more visual elements to extend the visible information in the single layer

The map currently only uses different colours to present the identical value category. As examples collected during the literature review, there are different methods on presenting more information in the same map view via changing different visual element.

### 7.2 Research work and findings

In this thesis, the concept of user-centred design was implemented through the whole research process. This research discussed and compared the pros and cons of global hexagonal tessellation on different system to make the final design of global multi-resolution hexagon layer system for the map presentation. The final project results were blended into a web-based Geo-visualization platform. Before the thesis project started, only the dataset to be used is defined, the concrete visualization solution was discussed and explored in the process. Research questions were proposed based on the understanding of the dataset's characteristics and the vision of the project outcomes in the beginning. It's challenging to find out research direction with so much uncertainty. The work of literature review collected relevant cases and identified the importance of implementing the user-

centred design into an explorative project process. Based on the guidance, this research set up the general working plan and conducted the user requirement assessment at the beginning. Based on the assessment result, an idea on presenting large point datasets into multi-scale maps to support users understanding the massive dataset came out. Such a decision requires the research to perform the literature review work again but the research fields are narrowed to the multi-scale map visualization. Multi-scale map visualization is more like a visualization concept which can be achieved by combining different facets but not a rigorously defined method. To achieve the multi-scale map visualization, different applications gave different ideas. Based on the available dataset, this research selected the grid-based spatial binning technique to compose a layer set for achieving the multi-scale map visualization.

The methodology is proposed for the project that is targeted to non-GIS background users on supporting them to understand large point datasets. This methodology covers the concrete working flow, all steps and methods required. Besides, discussion was made for deciding spatial tessellation shape and reference coordinate system.

Proposed methodology was realized in the case provided by the CAT management team from AGCS using the CAT risk portfolio dataset. ArcMap and Python were used for pre-processing procedure and the final application is hosted in the ArcGIS server. There are more concrete steps which define the specific required resolutions and visible scale ranges on building multi-resolution hexagon layer systems in the use case. In addition, to confirm hexagon layer build in the Web Mercator system is feasible to detect high-valued locations in our case, there is a result comparison with hexagon layers in the sphere in the end of the research. Final user test was conducted to reflect the final usability of the application and collect ideas for the future work.

To summarize, the research work can provide new insights on the following two aspects:

#### 1. Processing Geo-visualization project:

It's common that before a Geo-visualization project started, there are great uncertainties existed for making the final decision on the visualization solution. This also decides it would be impractical to specify a detailed development plan at the beginning of the project. User-centred design provides a flexible guideline to set up a dedicated working flow for different projects and explore out the solution during the process. For the user-centred design, the emphasis on the importance of involving users into the process is about getting even feedback and adjusting the work accordingly. Such feedback can come from end users directly but also from interior discussions with experienced members. Thus, for the project with limited users, it's also valuable to search opinions and set up discussions with different parties.

#### 2. Creating Hexagonal binning

The advantages of hexagonal binning for large point datasets were already well-recognized. However, many of related materials are online blogs which always neglect the differences

caused by the referenced coordinate system for building the tessellation. There are no agreement on which system is the best choice compared to others. They should be discussed and chosen after considering not only research regions, dataset used but also targeted users to balance the distortion between the shape and the size.

The visualization strategy consists of multi-resolution global hexagon layer system in this thesis has been verified its' effectiveness for visualizing the large point dataset in research's study case. The multi-resolution layer system has certain flexibility and can be served as a visual prototype for other point datasets. Combing with dedicated functions designed according to different users' requirements, it can finally promote into a comprehensive visualization solution for complexed cases.

## References

- Andrienko, G., Andrienko, N., Fischer, R., Mues, V., & Schuck, A. (2006). Reactions to geovisualization: an experience from a European project. *International Journal of Geographical Information Science*, 20(10), 1149–1171. <https://doi.org/10.1080/13658810600816524>
- Andrienko, N., Andrienko, G., Voss, H., Bernardo, F., Hipolito, J., & Kretchmer, U. (2002). Testing the Usability of Interactive Maps in CommonGIS. *Cartography and Geographic Information Science*, 29(4), 325–342. <https://doi.org/10.1559/152304002782008369>
- Battersby, S. E., Strebe, D. “d.”, & Finn, M. P. (2017). Shapes on a plane: evaluating the impact of projection distortion on spatial binning. *Cartography and Geographic Information Science*, 44(5), 410–421. <https://doi.org/10.1080/15230406.2016.1180263> (Cartography and Geographic Information Science, 44(5), 410-421).
- Beconytė, G., & Kryžanauskas, A. (2010). GEOGRAPHIC COMMUNICATION FOR SUSTAINABLE DECISIONS / GEOGRAFINĖS INFORMACIJOS KOMUNIKAVIMO PRIEMONĖS IR TVARIOJI PLĖTRA. *Technological and Economic Development of Economy*, 16(4), 603–612. <https://doi.org/10.3846/tede.2010.37>
- Birch, C. P.D., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, 206(3-4), 347–359. <https://doi.org/10.1016/j.ecolmodel.2007.03.041>
- Brewer, C. A., & Battenfield, B. P. (2007). Framing Guidelines for Multi-Scale Map Design Using Databases at Multiple Resolutions. *Cartography and Geographic Information Science*, 34(1), 3–15. <https://doi.org/10.1559/152304007780279078>
- Carr, D. B., Olsen, A. R., & White, D. (1992). Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. *Cartography and Geographic Information Systems*, 19(4), 228–236. <https://doi.org/10.1559/152304092783721231>
- Data Classification*. (2018, September 10). [https://saylordotorg.github.io/text\\_essentials-of-geographic-information-systems/s10-03-data-classification.html](https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s10-03-data-classification.html)
- David Lloyd, Jason Dykes, & Robert Radburn. (2007). *Understanding geovisualization users and their requirements—a user-centred approach*. [https://www.researchgate.net/publication/228336487\\_Understanding\\_geovisualization\\_users\\_and\\_their\\_requirements-a\\_user-centred\\_approach](https://www.researchgate.net/publication/228336487_Understanding_geovisualization_users_and_their_requirements-a_user-centred_approach)
- Dumont, M., Touya, G., & Duchêne, C. (2020). Designing multi-scale maps: lessons learned from existing practices. *International Journal of Cartography*, 6(1), 121–151. <https://doi.org/10.1080/23729333.2020.1717832>



- Elmqvist, N., & Fekete, J.-D. (2010). Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, 16(3), 439–454. <https://doi.org/10.1109/TVCG.2009.84>
- Elmqvist, N., Moere, A. V., Jetter, H.-C., Cernea, D., Reiterer, H., & Jankun-Kelly, T. J. (2011). Fluid interaction for information visualization. *Information Visualization*, 10(4), 327–340. <https://doi.org/10.1177/1473871611413180>
- Eppstein, D., van Kreveld, M., Speckmann, B., & Staals, F. (2015). Improved Grid Map Layout by Point Set Matching. *International Journal of Computational Geometry & Applications*, 25(02), 101–122. <https://doi.org/10.1142/S0218195915500077>
- GIS Lounge. (2014). *Binning in GIS*. <https://www.gislounge.com/binning-gis/>
- Gomez, L., Haesevoets, S., Kuijpers, B., & Vaisman, A. (2007, July 29). *Spatial Aggregation: Data Model and Implementation*. <https://arxiv.org/pdf/0707.4304>
- Hahn, S., & Döllner, J. (2017). *Hybrid-Treemap Layouting*. <https://doi.org/10.2312/EUROVISSHORT.20171137>
- Heimerl, F., Chang, C.-C., Sarikaya, A., & Gleicher, M. (2018). *Visual Designs for Binned Aggregation of Multi-Class Scatterplots*. <https://arxiv.org/pdf/1810.02445>
- Hou, W. User-centered design encyclopedia chapter.doc. [https://d1wqtxts1xzle7.cloudfront.net/6190316/10.1.1.94.381.pdf?response-content-disposition=inline%3B+filename%3DUser\\_centered\\_design.pdf&Expires=1597239568&Signature=JljojdZ~J~AvUVRmIhc9KneURwxLJgeJvty-QPH83FJzoTITCR5JqgQxs1aMSx8kt3zqYZPxbDMZpRijrnKgso7g~wDOgyAO58ZiL6l--hms1R7RG9zDyPHjqKJMjuMaJH5gpPdef62DAZJ2T4m7wDEVMMRH5TTw8gJ-xYHtyPMpdw-1wjKiCh8xvXBft~yxc4lrlAiuT6aug6fSoylmJC9sc00rAgO~6~yxa38mhGSgq6nDpMKEIYQCrlSM8KlJy-hEQ-2HHeeKLRmEnMe2XSP86nPyMtePuYwwwySUC7hllEWcRix5PgVB0EeyFE1DNjOauK0Zm5BkssjMF-AuQ\\_\\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](https://d1wqtxts1xzle7.cloudfront.net/6190316/10.1.1.94.381.pdf?response-content-disposition=inline%3B+filename%3DUser_centered_design.pdf&Expires=1597239568&Signature=JljojdZ~J~AvUVRmIhc9KneURwxLJgeJvty-QPH83FJzoTITCR5JqgQxs1aMSx8kt3zqYZPxbDMZpRijrnKgso7g~wDOgyAO58ZiL6l--hms1R7RG9zDyPHjqKJMjuMaJH5gpPdef62DAZJ2T4m7wDEVMMRH5TTw8gJ-xYHtyPMpdw-1wjKiCh8xvXBft~yxc4lrlAiuT6aug6fSoylmJC9sc00rAgO~6~yxa38mhGSgq6nDpMKEIYQCrlSM8KlJy-hEQ-2HHeeKLRmEnMe2XSP86nPyMtePuYwwwySUC7hllEWcRix5PgVB0EeyFE1DNjOauK0Zm5BkssjMF-AuQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA)
- john.nelson. (2013/07). *UXBlog | IDV Solutions' User Experience: A Hack for Scaling Your Hex Map*. <http://uxblog.idvsolutions.com/2013/07/a-hack-for-scaling-your-hex-map.html>
- Jones, C. B., & Mark Ware, J. (2005). Map generalization in the Web age. *International Journal of Geographical Information Science*, 19(8-9), 859–870. <https://doi.org/10.1080/13658810500161104>
- Keim, D. A. (2007, 2005). Information Visualization. In J. Dykes (Ed.), *Exploring geovisualization* (pp. 21–52). Elsevier. <https://doi.org/10.1016/B978-008044531-1/50420-6>
- Keim, D. A., & Schneidewind, J. (2005). Scalable Visual Data Exploration of Large Data Sets via MultiResolution. *Journal of Universal Computer Science*, 11(11), 1766-1779UR -

- <http://kops.uni-konstanz.de/handle/123456789/5449>. <https://doi.org/10.3217/jucs-011-11-1766>
- Kevin Sahr, Denis White, and A. Jon Kimerling. Geodesic Discrete Global Grid Systems. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.997&rep=rep1&type=pdf>
- Koua, E. L., Maceachren, A., & Kraak, M. -J. (2006). Evaluating the usability of visualization methods in an exploratory geovisualization environment. *International Journal of Geographical Information Science*, 20(4), 425–448. <https://doi.org/10.1080/13658810600607550>
- Lloyd, D., & Dykes, J [Jason] (2011). Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2498–2507. <https://doi.org/10.1109/TVCG.2011.209>
- Lobo, M.-J., Appert, C., & Pietriga, E. (2017). MapMosaic: dynamic layer compositing for interactive geovisualization. *International Journal of Geographical Information Science*, 31(9), 1818–1845. <https://doi.org/10.1080/13658816.2017.1325488>
- Lopez, I.F.V., Snodgrass, R. T., & Moon, B. (2005). Spatiotemporal aggregate computation: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 17(2), 271–286. <https://doi.org/10.1109/TKDE.2005.34>
- Maguire, M., & Bevan, N. (2002). User Requirements Analysis. In J. Hammond, T. Gross, & J. Wesson (Eds.), *IFIP - The International Federation for Information Processing: Vol. 99. Usability: Gaining a Competitive Edge* (pp. 133–148). Springer US. [https://doi.org/10.1007/978-0-387-35610-5\\_9](https://doi.org/10.1007/978-0-387-35610-5_9)
- Mao, J.-Y., Vredenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, 48(3), 105–109. <https://doi.org/10.1145/1047671.1047677>
- Nielsen, J. (1992). The usability engineering life cycle. *Computer*, 25(3), 12–22. <https://doi.org/10.1109/2.121503>
- Norman, D. A., & Draper, S. W. (1986). *User centered system design: New perspectives on human-computer interaction / edited by Donald A. Norman, Stephen W. Draper*. Lawrence Erlbaum Associates.
- Norman Donald. (1988). *The psychology of everyday things*. Basic Books.
- Ormsby, D., & Mackaness, W. (1999). The Development of Phenomenological Generalization Within an Object-oriented Paradigm. *Cartography and Geographic Information Science*, 26(1), 70–80. <https://doi.org/10.1559/152304099782424910>
- Pedersen, T. B., & Tryfona, N. (2006). Pre-aggregation in Spatial Data Warehouses. In C. S. Jensen, M. Schneider, B. Seeger, & V. J. Tsotras (Eds.), *Lecture Notes in Computer Science: Vol. 2121, Advances in Spatial and Temporal Databases: 7th International Symposium*,

- SSTD 2001 Redondo Beach, CA, USA, July 12-15, 2001 Proceedings* (pp. 460–478). Springer-Verlag Berlin Heidelberg.
- Phan, D., Xiao, L., Yeh, R., Hanrahan, P., & Winograd, T. (2005). Flow map layout. In J. Stasko & M. Ward (Eds.), *INFO VIS 05: IEEE symposium on information visualization / edited by John Stasko and Matt Ward* (pp. 219–224). IEEE.  
<https://doi.org/10.1109/INFVIS.2005.1532150>
- Red Blob Games: Hexagonal Grids*. (2020, August 10).  
<https://www.redblobgames.com/grids/hexagons/>
- Roth, R., Ross, K., & MacEachren, A. (2015). User-Centered Design for Interactive Maps: A Case Study in Crime Analysis. *ISPRS International Journal of Geo-Information*, 4(1), 262–301. <https://doi.org/10.3390/ijgi4010262>
- Sahr, K. (2018). User Documentation for Discrete Global Grid Generation Software.  
<https://www.discretglobalgrids.org/wp-content/uploads/2019/05/dggridManualV64.pdf>
- Sahr, K., White, D., & Kimerling, A. J. (2003). Geodesic Discrete Global Grid Systems. *Cartography and Geographic Information Science*, 30(2), 121–134.  
<https://doi.org/10.1559/152304003100011090>
- Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In *IEEE symposium on visual languages* (pp. 336–343). IEEE Comput. Soc. Press. <https://doi.org/10.1109/VL.1996.545307>
- Slingsby, A., Dykes, J [Jason], Wood, J., & Radburn, R. (2014). Designing an exploratory visual interface to the results of citizen surveys. *International Journal of Geographical Information Science*, 28(10), 2090–2125. <https://doi.org/10.1080/13658816.2014.920845>
- Slocum, T. A., Cliburn, D. C., Feddema, J. J., & Miller, J. R. (2003). Evaluating the Usability of a Tool for Visualizing the Uncertainty of the Future Global Water Balance. *Cartography and Geographic Information Science*, 30(4), 299–317.  
<https://doi.org/10.1559/152304003322606210>
- Smith, D. A. (2017). Visualising world population density as an interactive multi-scale map using the global human settlement population layer. *Journal of Maps*, 13(1), 117–123.  
<https://doi.org/10.1080/17445647.2017.1400476>
- Suzanne Robertson, & James Robertson. Mastering the Requirements Process: Getting Requirements Right.  
<http://ptgmedia.pearsoncmg.com/images/9780321815743/samplepages/0321815742.pdf>
- Why hexagons?—ArcGIS Pro | Documentation*. (2020, June 24).  
<https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-whyhexagons.htm>

- Wood, J., & Dykes, J [Jason] (2008). Spatially ordered treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1348–1355.  
<https://doi.org/10.1109/TVCG.2008.165>
- Wood, J., Dykes, J [Jason], Slingsby, A., & Clarke, K. (2007). Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1176–1183.  
<https://doi.org/10.1109/TVCG.2007.70570>
- Zhang, D., & Tsostras, V. J. (2001). Improving min/max aggregation over spatial objects. In W. G. Aref (Ed.), *ACM-GIS 2001: Proceedings of the 9th ACM International Symposium on Advances in Geographic Information Systems, November 9-10, 2001, Atlanta, GA, USA / editor, Walid G. Aref* (p. 88). ACM Press. <https://doi.org/10.1145/512161.512181>
- Zhou, S., & Jones, C. B. (2003). A Multi-representation Spatial Data Model. In G. Goos, J. Hartmanis, & J. van Leeuwen (Eds.), *Advances in Spatial and Temporal Databases: 8th International Symposium, SSTD 2003, Santorini Island, Greece, July 2003. Proceedings* (pp. 394–411). Springer Berlin Heidelberg.

# Appendix

---

## Questionnaire

### Questionnaire for the first round user test survey

<b>Question 1</b>	<b>Just based on the map, features in which colour arouse your interest? (Multi-selection)</b>						
<b>Options</b>	Red	Green	Others	I don't care colour			
<b>Question 2</b>	Which layer/layers will you spend more time to explore? (multi- selection)						
<b>Options</b>	Country layer	10,000 km <sup>2</sup> layer	100 km <sup>2</sup> layer	1 km <sup>2</sup> layer	Points layer		
<b>Question 3</b>	Do you think layer zooming will help you find out interested locations ?						
<b>Options</b>	Definitely yes	Probably yes	Might or might not	Probably not		Definitely not	
<b>Question 4</b>	Do you think manual calculation function will be useful for your work ?						
<b>Options</b>	Definitely yes	Probably yes	Might or might not	Probably not		Definitely not	
<b>Question 5</b>	On which layer/layers will you do manual calculation ? (Multi-selection)						
<b>Options</b>	Country layer	10,000 km <sup>2</sup> layer	100 km <sup>2</sup> layer	1 km <sup>2</sup> layer	Points layer	I do not do manual calculation	
<b>Question 6</b>	How much you care about the comparison of values in different quarters ?						
<b>Options</b>	A great deal	A lot	A moderate amount	A little		Not at all	
<b>Question 7</b>	Do you have other comments ?						
<b>Comments</b>							

## Questionnaire for the second round user test survey

<b>Question 1</b>	<b>Was the interface easy for you to understand ?</b>				
<b>Options</b>	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult
<b>Question 2</b>	Were hexagon layers easy for you to understand ?				
<b>Options</b>	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult
<b>Question 3</b>	Was it easy for you to find some interested locations by zooming hexagon layers?				
<b>Options</b>	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult
<b>Question 4</b>	What other attributes you want to have in hexagon layers ?				
<b>Comments</b>					
<b>Question 5</b>	How useful is the information in the pop up panel for you ?				
<b>Options</b>	Extremely useful	Very useful	Moderately useful	Slightly useful	Not at all useful
<b>Question 6</b>	What other information you want to have in the pop up panel ?				
<b>Comments</b>					
<b>Question 7</b>	How do you think the rendering speed of the application ?				
<b>Options</b>	Excellent	Good	Average	Poor	Terrible
<b>Question 8</b>	What was the one thing you liked the most about the application ?				
<b>Comments</b>					
<b>Question 9</b>	What was the one thing you disliked the most about the application ?				
<b>Comments</b>					
<b>Question 10</b>	Apart from existed layers, do you have other layers expected to be added in the application ?				
<b>Comments</b>					
<b>Question 11</b>	Please add your further comments.				
<b>Comments</b>					

## Code (take resolution in 10,000 km<sup>2</sup> as an example)

### Hexagon Creation

```
import arcpy
import numpy as np
import geopandas as gpd
from shapely.geometry import Polygon

def main():
    """Generate the grid using the following parameter"""
    SR = arcpy.SpatialReference(3857)
    dX = 62000
    dY = 62000 # X and Y cell widths
    cols = 216 #number of columns
    rows = 910 #number of rows
    y_l = -30240971 + 0.866*dY
    x_l = -20037507-dX/2
    corner = [x_l, y_l]
    angle = 0 # rotation angle, clockwise +ve
    # create the grid
    pnts = grid_array(corner, dX, dY, cols, rows, angle)
    output_shp = r'L:\AGCS\COO\CAI\CAI_MGMT\90_Team\Members\Lin Zhu\07_Master thesis\08_data\Hexagon_web_100km\100km_1407.shp"
    output_polygons(output_shp, SR, pnts)

    #clip the gridpyth
    hexagon = gpd.read_file(output_shp)
    polygon = Polygon([(-20037507, -30240972), (-20037507, 20120000), (20037507, 20120000), (20037507, -30240972), (-20037507, -30240972)])
    poly_gdf = gpd.GeoDataFrame([1], geometry=[polygon], crs=hexagon.crs)
    hexagon_clipped = gpd.clip(hexagon, polygon)
    hexagon_clipped.to_file('L:\AGCS\COO\CAI\CAI_MGMT\90_Team\Members\Lin Zhu\07_Master thesis\08_data\Hexagon_web_100km\100km_1407_cliped.shp', driver='ESRI Shapefile')

    print('finished !')

def grid_array(corner=[0, 0], dX=1, dY=1, cols=1, rows=1, angle=0):
    """create the array of pnts to pass on to arcpy using numpy magic"""
    X = [0.0, 0.5, 1.5, 2.0, 1.5, 0.5, 0.0]
    Y = [0.0, 0.866, 0.866, 0.0, -0.866, -0.866, 0.0]
    seed = np.column_stack((X, Y)) * [dX, dY] # first array corner values scaled
    u = [seed + [j * (dX * 3.0), i * (dY * 0.866)] for i in range(0, rows) for j in range(0, cols) if i % 2 == 0]
    t = [seed + [(j * (dX * 3.0)) + 1.5 * dX, i * (dY * 0.866)] for i in range(0, rows) for j in range(0, cols) if
        | not (i % 2 == 0)]
    u = u + t
    pnts = np.array(u) #
    x = [rotate(p, angle) for p in pnts] # rotate the scaled points
    pnts = [p + corner for p in x] # translate them
    return pnts

def rotate(pnts, angle=0):
    """rotate points about the origin in degrees, (+ve for clockwise) """
    angle = np.deg2rad(angle) # convert to radians
    s = np.sin(angle)
    c = np.cos(angle) # rotation terms
    aff_matrix = np.array([[c, -s], [s, c]]) # rotation matrix
    XY_r = np.dot(pnts, aff_matrix) # numpy magic to rotate pnts
    return XY_r

def output_polygons(output_shp, SR, pnts):
    """produce the output polygon shapefile"""
    msg = '\nRead the script header... A projected coordinate system required'
    assert (SR != None) and (SR.type == 'Projected'), msg
    polygons = []
    for pnt in pnts: # create the polygon geometry
        polygons.append(arcpy.Polygon(arcpy.Array([arcpy.Point(*xy) for xy in pnt]), SR))
    if arcpy.Exists(output_shp): # overwrite any existing versions
        arcpy.Delete_management(output_shp)
    arcpy.CopyFeatures_management(polygons, output_shp)

if __name__ == '__main__':
    """Generate the grid using the listed parameters"""
    main() # modify the parameters in demo to run
```



## Hexagon Index

```
# creating new column for the index
arcpy.AddField_management(myFeatureClass, 'TYPE', "TEXT", field_length=10,field_is_nullable="NULLABLE")
arcpy.AddField_management(myFeatureClass, 'HX', "LONG",field_is_nullable="NULLABLE")
arcpy.AddField_management(myFeatureClass, 'HY', "LONG", field_is_nullable="NULLABLE")
arcpy.AddField_management(myFeatureClass, 'index', "TEXT", field_length=10,field_is_nullable="NULLABLE")

# indexes are divided into A, B series
with arcpy.da.UpdateCursor(myFeatureClass, ["OID@", "TYPE", "HX", "HY", "index"]) as cursor:
    for row in cursor:
        if (row[0] < 98280 ):
            row[1] = 'A'
            n= (row[0])/216
            row[2] = row[0] - 216*n + 1
            row[3] = n+1
            row[4] = 'A-' + str(row[2])+ '-' +str(row[3])
            print (row)
            cursor.updateRow(row)
        elif (row[0] >= 98280):
            row[1] = 'B'
            n = (row[0]-98280) / 216
            row[2] = (row[0]-98280) - 216 * n + 1
            row[3] = n + 1
            row[4] = 'B-' + str(row[2]) + '-' + str(row[3])
            print (row)
            cursor.updateRow(row)
```

## Assigning values

```
#####
#spatial join with hexagon layer
hexagon = gpd.read_file("L:/AGCS/COO/CAT/CAT_MGMT/90_Team/Members/Lin Zhu/07_Master thesis/08_data/data_refreshed/10000km2.gdb",driver="FileGDB", layer='empty_10000km2')
#hexagon = gpd.read_file("Hexagon_web_10km/Hexagon_10km_1_cliped.shp")
join = gpd.sjoin(points,hexagon, how="left", op='intersects')
#join.to_csv('point2hexagon/points_100km_20191201_new.csv')
print(join.head())

#####
#calculation
point_value = join.filter(['TIV', 'NetExp_EQ', 'NetExp_FL', 'NetExp_WS', 'NetExp_CS', 'AAL_PreCat_EQ', 'AAL_PreCat_WS', 'AAL_PreCat_FL', 'AAL_PreCat_CS', 'AAL_PreCat_ALL', 'RiCost_EQ',
                           'RiCost_WS', 'RiCost_FL', 'RiCost_CS', 'RiCost_ALL', 'index'])
point_value_result = point_value.groupby(['index']).sum()
point_value['count'] = point_value.groupby(['index'])['TIV'].transform('size')
point_value_result['count'] = point_value_result.groupby(['index'])['count'].mean()
point_value_result['NetExp_ALL'] =point_value_result['NetExp_EQ']+point_value_result['NetExp_FL']+point_value_result['NetExp_WS']+point_value_result['NetExp_CS']
point_value_result = point_value_result.reset_index()
column_titles = ['TIV', 'NetExp_EQ', 'NetExp_FL', 'NetExp_WS', 'NetExp_CS', 'NetExp_ALL', 'AAL_PreCat_EQ', 'AAL_PreCat_WS', 'AAL_PreCat_FL', 'AAL_PreCat_CS', 'AAL_PreCat_ALL',
                 'RiCost_EQ', 'RiCost_WS', 'RiCost_FL', 'RiCost_CS', 'RiCost_ALL', 'count', 'index']
point_value_result = point_value_result.reindex(columns=column_titles)
print(point_value_result.head())
```



## Hexagon Searching

```
def search_child(code):
    """search the index of child hexagons"""
    A_series = []
    B_series = []
    code_list = code.split("-", 2)
    X_100km = int(code_list[1])
    Y_100km = int(code_list[2])
    Type = code_list[0]
    A_X = [0, 1, 2, 3, 4, 5]
    B_X = [-1, 0, 1, 2, 3, 4, 5]
    A_Y = [[3, 4, 5, 6, 7, 8], [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
            [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], [3, 4, 5, 6, 7, 8]]
    B_Y = [[4, 5, 6], [1, 2, 3, 4, 5, 6, 7, 8, 9], [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
            [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10], [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10], [1, 2, 3, 4, 5, 6, 7, 8, 9],
            [4, 5, 6]]
    for i in range(0, len(A_X)):
        for j in range(0, len(A_Y[i])):
            if (Type == 'A'):
                x = A_X[i] + 10 * (X_100km - 1)
                y = A_Y[i][j] + 10 * (Y_100km - 1)
                if (x > 0 and y > 0):
                    index = 'A-' + str(x) + '-' + str(y)

                    A_series.append(index)
            elif (Type == 'B'):
                x = A_X[i] + 5 + 10 * (X_100km - 1)
                y = A_Y[i][j] + 5 + 10 * (Y_100km - 1)
                if (x > 0 and y > 0):
                    index = 'A-' + str(x) + '-' + str(y)

                    A_series.append(index)

    for i in range(0, len(B_X)):
        for j in range(0, len(B_Y[i])):
            if (Type == 'A'):
                x = B_X[i] + 10 * (X_100km - 1)
                y = B_Y[i][j] + 10 * (Y_100km - 1)
                if (x > 0 and y > 0):
                    index = 'B-' + str(x) + '-' + str(y)

                    B_series.append(index)
                    #print ('B-' + str(x) + '-' + str(y))
            elif (Type == 'B'):
                x = B_X[i] + 5 + 10 * (X_100km - 1)
                y = B_Y[i][j] + 5 + 10 * (Y_100km - 1)
                if (x > 0 and y > 0):
                    index = 'B-' + str(x) + '-' + str(y)

                    B_series.append(index)
    return A_series, B_series
```

## Combining Values

```
import pandas as pd
from geopandas import gpd
from datetime import datetime

hexagon_100km_20190701 = gpd.read_file('Hexagon_web_100km_new/Hexagon_valued_20190701_new.shp')
hexagon_100km_20191001 = gpd.read_file('Hexagon_web_100km_new/Hexagon_valued_20191001_new.shp')
hexagon_100km_20191201 = gpd.read_file('Hexagon_web_100km_new/Hexagon_valued_20191201_new.shp')
hexagon_100km_20200401 = gpd.read_file('Hexagon_web_100km_new/Hexagon_valued_20200401_new.shp')
#hexagon = gpd.read_file("E:/AGCS/COO/CAT/MGMT/50_Team/Member/Lin Zhu/07_Master thesis/08_data/data_refreshed/10000km2.gdb", driver='FileGDB', layer='empty_10000km2')
hexagon = hexagon_100km_20200401.filter(['index', 'geometry'])

hexagon_100km_20190701 = hexagon_100km_20190701.filter(['TIV', 'NetExp_EQ', 'NetExp_FL', 'NetExp_WS', 'NetExp_CS', 'NetExp_ALL', 'AAL_PrecCat', 'AAL_Prec_1', 'AAL_Prec_2', 'AAL_Prec_3',
'AAL_Prec_4', 'RiCost_EQ', 'RiCost_WS', 'RiCost_FL', 'RiCost_CS', 'RiCost_ALL', 'index', 'count'])
hexagon_100km_20191001 = hexagon_100km_20191001.filter(['TIV', 'NetExp_EQ', 'NetExp_FL', 'NetExp_WS', 'NetExp_CS', 'NetExp_ALL', 'AAL_PrecCat', 'AAL_Prec_1', 'AAL_Prec_2', 'AAL_Prec_3',
'AAL_Prec_4', 'RiCost_EQ', 'RiCost_WS', 'RiCost_FL', 'RiCost_CS', 'RiCost_ALL', 'index', 'count'])
hexagon_100km_20191201 = hexagon_100km_20191201.filter(['TIV', 'NetExp_EQ', 'NetExp_FL', 'NetExp_WS', 'NetExp_CS', 'NetExp_ALL', 'AAL_PrecCat', 'AAL_Prec_1', 'AAL_Prec_2', 'AAL_Prec_3',
'AAL_Prec_4', 'RiCost_EQ', 'RiCost_WS', 'RiCost_FL', 'RiCost_CS', 'RiCost_ALL', 'index', 'count'])
hexagon_100km_20200401 = hexagon_100km_20200401.filter(['TIV', 'NetExp_EQ', 'NetExp_FL', 'NetExp_WS', 'NetExp_CS', 'NetExp_ALL', 'AAL_PrecCat', 'AAL_Prec_1', 'AAL_Prec_2', 'AAL_Prec_3',
'AAL_Prec_4', 'RiCost_EQ', 'RiCost_WS', 'RiCost_FL', 'RiCost_CS', 'RiCost_ALL', 'index', 'count'])

hexagon_100km_20190701 = hexagon_100km_20190701.rename(columns={'TIV': 'TIV_1', 'NetExp_EQ': 'NE_EQ_1', 'NetExp_FL': 'NE_FL_1', 'NetExp_WS': 'NE_WS_1', 'NetExp_CS': 'NE_CS_1', 'NetExp_ALL': 'NE_ALL_1',
'AAL_PrecCat': 'AAL_EQ_1', 'AAL_Prec_1': 'AAL_WS_1', 'AAL_Prec_2': 'AAL_FL_1', 'AAL_Prec_3': 'AAL_CS_1', 'AAL_Prec_4': 'AAL_ALL_1',
'RiCost_EQ': 'RI_EQ_1', 'RiCost_WS': 'RI_WS_1', 'RiCost_FL': 'RI_FL_1', 'RiCost_CS': 'RI_CS_1', 'RiCost_ALL': 'RI_ALL_1', 'count': 'count_1'})
hexagon_100km_20191001 = hexagon_100km_20191001.rename(columns={'TIV': 'TIV_2', 'NetExp_EQ': 'NE_EQ_2', 'NetExp_FL': 'NE_FL_2', 'NetExp_WS': 'NE_WS_2', 'NetExp_CS': 'NE_CS_2', 'NetExp_ALL': 'NE_ALL_2',
'AAL_PrecCat': 'AAL_EQ_2', 'AAL_Prec_1': 'AAL_WS_2', 'AAL_Prec_2': 'AAL_FL_2', 'AAL_Prec_3': 'AAL_CS_2', 'AAL_Prec_4': 'AAL_ALL_2',
'RiCost_EQ': 'RI_EQ_2', 'RiCost_WS': 'RI_WS_2', 'RiCost_FL': 'RI_FL_2', 'RiCost_CS': 'RI_CS_2', 'RiCost_ALL': 'RI_ALL_2', 'count': 'count_2'})
hexagon_100km_20191201 = hexagon_100km_20191201.rename(columns={'TIV': 'TIV_3', 'NetExp_EQ': 'NE_EQ_3', 'NetExp_FL': 'NE_FL_3', 'NetExp_WS': 'NE_WS_3', 'NetExp_CS': 'NE_CS_3', 'NetExp_ALL': 'NE_ALL_3',
'AAL_PrecCat': 'AAL_EQ_3', 'AAL_Prec_1': 'AAL_WS_3', 'AAL_Prec_2': 'AAL_FL_3', 'AAL_Prec_3': 'AAL_CS_3', 'AAL_Prec_4': 'AAL_ALL_3',
'RiCost_EQ': 'RI_EQ_3', 'RiCost_WS': 'RI_WS_3', 'RiCost_FL': 'RI_FL_3', 'RiCost_CS': 'RI_CS_3', 'RiCost_ALL': 'RI_ALL_3', 'count': 'count_3'})
hexagon_100km_20200401 = hexagon_100km_20200401.rename(columns={'TIV': 'TIV_4', 'NetExp_EQ': 'NE_EQ_4', 'NetExp_FL': 'NE_FL_4', 'NetExp_WS': 'NE_WS_4', 'NetExp_CS': 'NE_CS_4', 'NetExp_ALL': 'NE_ALL_4',
'AAL_PrecCat': 'AAL_EQ_4', 'AAL_Prec_1': 'AAL_WS_4', 'AAL_Prec_2': 'AAL_FL_4', 'AAL_Prec_3': 'AAL_CS_4', 'AAL_Prec_4': 'AAL_ALL_4',
'RiCost_EQ': 'RI_EQ_4', 'RiCost_WS': 'RI_WS_4', 'RiCost_FL': 'RI_FL_4', 'RiCost_CS': 'RI_CS_4', 'RiCost_ALL': 'RI_ALL_4', 'count': 'count_4'})

#hexagon_100km_20190701.join(hexagon_100km_20191001.set_index('index'), on='index')
left = pd.merge(hexagon_100km_20190701, hexagon_100km_20191001, on=['index'], how='outer')
left = pd.merge(left, hexagon_100km_20191201, on=['index'], how='outer')
left = pd.merge(left, hexagon_100km_20200401, on=['index'], how='outer')

hexagon_valued = hexagon.merge(left, on='index')
hexagon_valued.to_file('Hexagon_web_100km_new/new_Hexagon_100km_valued_combined_updated.shp', driver='ESRI Shapefile')
```

## Hexagon in Sphere (R script)

```
library(dggridR)
library(dplyr)
vignette('dggridR')
dggs <- dgconstruct(res=7, metric=FALSE)
dgg<-dgearthgrid(dggs, frame = TRUE, wrapcells = TRUE, savegrid="E:\\Master Thesis\\code\\Hexagon\\grid\\temp.shp")
countries <- map_data("world")
p<- ggplot() +
  geom_polygon(data=countries, aes(x=long, y=lat, group=group), fill=NA, color="black")+
  geom_polygon(data=dgg, aes(x=long, y=lat, group=group), fill="blue", alpha=0.4) +
  geom_path(data=dgg, aes(x=long, y=lat, group=group), alpha=0.4, color="white") +
  coord_equal()
p

dggs <- dgconstruct(res=6, metric=FALSE)
global <- dgearthgrid(dggs, frame=FALSE)
global.cell <- data.frame(cell=getSpPPolygonsIDSlots(global), row.names=getSpPPolygonsIDSlots(global))
global <- SpatialPolygonsDataFrame(global, global.cell)
for(i in 1:length(global@polygons)) {
  if(max(global@polygons[[i]]@Polygons[[1]]@coords[,1]) -
  min(global@polygons[[i]]@Polygons[[1]]@coords[,1]) > 180) {
    global@polygons[[i]]@Polygons[[1]]@coords[,1] <- (global@polygons[[i]]@Polygons[[1]]@coords[,1] +360) %% 360
  }
}
writeOGR(global, "E:\\Master Thesis\\code\\Hexagon\\grid\\grid_res_6.shp", "", "ESRI Shapefile")
```