Cartography M.Sc.

Master thesis

An Interactive Visual Interface for Geodata Query with Space-time Cube

Mengyao Gao

Technical University of Munich

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITY OF TWENTE.

2021

An Interactive Visual Interface for Geodata Query with Space-time Cube

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

Author: Mengyao, Gao
Study course: Cartography M.Sc.
Supervisor: Chenyu Zuo (TUM)
Reviewer: Florian Ledermann (TUW)

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

Date of submission: 12.07.2021

Statement of Authorship

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

“An Interactive Visual Interface for Geodata Query with Space-time Cube”

I have fully referenced the ideas and work of others, whether published or unpublished.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

Munich, 12.07.2021 Gao, Mengyao

Acknowledgment

First of all, I would like to express my gratitude to my supervisor, Chenyu Zuo, who has given so many valuable suggestions and been so supportive during the whole research. Without her guidance, it is impossible to finish this master thesis.

I would also like to thank my reviewer, Florian Ledermann for his advises in the proposal defense and the midterm presentation. Also, I want to express my thanks to all the staff of Cartography M.Sc. especially all coordinators at the four universities. I appreciate all the efforts they have made to organize everything and care about our lives especially in this epidemic situation.

Besides, I would like to thank all the participants who have been engaged in the evaluation stage and gave valuable feedbacks.

Finally, I want to express my deepest appreciations to my family, my classmates, and all friends, who have supported me so much all the time.

Abstract

Nowadays, much effort has been made in developing data portals to enable various types of data available to public so that more useful information can be generated. However, the open data is still not reaching the expected reuse level (Safarov et al., 2017). The visual interface has a great influence on users' search behaviors and the success rate of finding needed datasets (Xiao et al., 2020). Therefore, some visualization platforms for data portals have been built up to improve efficiency of data query. Most of them focus on visualizing data content directly and only use time sliders to visualize spatiotemporal datasets. Some important factors of geodata query such as data quality are not visualized and users cannot have a good overview of all datasets in a database.

Therefore, this research wants to incorporate an innovative time visualization method and tries to visualize aggregated metadata which are important in data querying instead of data content to provide a good overview of related information and help users find needed datasets effectively. It also tries to improve the user interface in visualizing aggregated metadata of all datasets. Data quality, captured time of dataset, data size, and data category of all available datasets are visualized to improve efficiency of data querying.

In order to achieve that, this research proposes a visualization method that combines space time cube, which is a popular time visualization method recently, and 3D symbols, which use different visual variables to represent different metadata, namely data quality, data category, and data size. Based on the method, a visual interface for geodata query with space time cube is built up. Jiangsu province, one of the most economically active provinces in China, is used as the test area. Four types of datasets, namely geographic data, social data, economic data, and nighttime light data, are used as test datasets. Unlike traditional 2D maps, a 3D scene is created to establish the whole space time cube and all datasets are visualized using 3D symbols. Then, the effectiveness and satisfaction of this visual interface are evaluated.

Benchmark tasks are carried out to evaluate the effectiveness. The overall performance is good, which suggests that this visual interface could assist users in identifying needed datasets and providing a good overview of important metadata. As for the possibility of exploring spatial distribution, more evaluation needs to be carried out since the problem of poor understanding of data content affected the result of identifying spatial distribution. The satisfaction of this visual interface is assessed by the responses in the user attitude investigation. The result is quite diverse. Almost half of the participants thought the space time cube was suitable to visualize spatiotemporal data in this visual interface and that it was simple to complete usual tasks with it. Other participants reported that it was difficult to understand the 3D map. Concluded from participants' feedbacks, user-friendly

interaction with 3D scene is a critical factor of user’s attitude towards this visual interface.

Future work could concentrate on comparing the efficiency of using space time cube to visualize datasets and using time slider directly to display datasets.

Keywords: web map, space-time cube, 3D visualization, geodata query

Contents

List of Figures

List of Tables

Abbreviations

2D	Two-dimension
3D	Three-dimension
ACMDL	Association for Computing Machinery Digital Library
BDTD	Brazilian Digital Library of Theses and Dissertations
DG	Design Goal
DN	Digital Number
KB	Kilobit
OGD	Open Government Data
SDTS	Spatial Data Transfer Standard
STC	Space-time Cube
VGI	Volunteered Geographic Information
VIIRS	visible infrared imaging radiometer suite

1 Introduction

1.1 Motivation and Problem Statement

Current times have led us to an era of data explosion. With the presence of the ubiquitous sensors, data are generated and collected automatically every day. This includes satellite images, trajectories, meteorological data and so on. Additionally, the volunteered geographic information (VGI) has emerged, since users can generate and publish their own content including their geographic information (Goodchild, 2007). Moreover, online social media platforms have allowed users to share data about themselves and their surroundings. These social media data provide a new perspective on the physical world, termed as social sensing (Wang et al., 2019). Due to the presence of all these data, there is an emphasis on publishing various datasets as open government data (OGD). Therefore an increasing number of data portals are providing access to economic, social, and geographic datasets (Xiao et al., 2020). By reusing these various types of datasets, researchers and citizens have the opportunity to generate more information and deliver more services (Davies, 2010). Moreover, researchers can also issue warnings about serious events by analyzing social media datasets (Goodchild, 2013). In summary, enabling data reuse can lead to numerous findings.

However, open data is far from reaching their expected level of reuse (Safarov et al., 2017). One reason reported by the users is that it is difficult to find the needed datasets even if they are published on the internet (Gebka et al., 2020). In order to make it easier for users to find the data and stimulate the reuse of open data, the construction of open data platforms is the necessary. A research has found that visual interface has a great influence on the users' behavior of searching and the success rate of finding data (Xiao et al., 2020). There are some platforms using visualization methods to enable fast and effective visual query. For instance, Facet, a web-based application for querying and exploration of datasets, uses maps to display spatial information of datasets (Stadler et al., 2014). Map4rdf, which is a faceted browser for geospatial data, displays spatial features using points, lines, and polygons in a map (de León et al., 2012).

However, there are still some defects of those visualization platforms. The first one is that they focus on the visualization of data content, but the description of data, i.e., the metadata, is missing. For instance, the data quality is hardly accessible to users. Knowing the data quality information is the prerequisite for a scientific use of data. High quality datasets can provide basic guarantees for the reliability of studies, and low quality data help form hypotheses in the early stages of research (Goodchild, 2013). Few data portals allow users to easily access data quality information. Users may find the data quality really poor after downloading them or experiment with poor quality data without noticing it. For example, users may use remote sensing images directly without knowing the uncertainty caused by the preprocessing steps. It may also take users a long time to download high quality data with a large size in the early stage of research, when it is not necessary. These misconducts may lead to inefficiency and inaccuracy. Moreover, those visualization platforms often use traditional 2D maps. It is not clear whether novel visualization methods, such as space time cube and visualization of metadata, would give a good overview of geodata and enable users to find their needed datasets effectively.

Therefore, this project proposes a visualization method for geodata portals based on the defects of existing visualization platforms. It also builds up a visual interface to visualize data quality, spatial information, temporal information, data size, and data category of spatiotemporal datasets, using space time cube to make it easier for geodata query. In addition, the effectiveness and satisfaction of the visual interface are evaluated.

1.2 Research Identification

1.2.1 Research Questions

- How to apply the space time cube visualization method to illustrate metadata of spatiotemporal data to enable users to get an overview and find needed datasets effectively?

a) Which visual variables should be chosen to visualize aggregated metadata and applied to 3D symbols in the whole space time cube?

b) Which functionalities should be designed in this space time cube to enable users to get a good overview and find needed datasets effectively?

- How to evaluate the proposed visualization method?

a) How is the effectiveness of the proposed visualization method using space time cube?

b) How is the satisfaction of users towards the proposed visualization method using space time cube?

1.2.2 Research Objectives

The overall objective of this project is to develop a visualization method of aggregated metadata using space time cube to enable users to get good overview and find needed datasets effectively.

It can be divided into five sub-objectives:

1) To understand the components of data quality.

2) To gain an overview of related works in visualizing semantic information of spatiotemporal data using space time cube.

3) To develop the conceptual design for visualization of metadata.

4) To implement a prototype based on the conceptual design.

5) To evaluate the effectiveness and satisfaction of this visual interface.

1.2.3 Innovation

The innovation of this project is to visualize the aggregated metadata of the datasets, which gives users a good overview of important information of all datasets. Especially, we visualize the quality of geodata, which aims at giving intuitively users data quality information to improve efficiency and accuracy in the query process. Additionally, the use of space time cube, with third dimension representing time information, may give good overview of temporal distribution and provide possibility of data exploration.

1.3 Thesis Structure

This thesis is composed of six chapters. The first chapter is the introduction, demonstrating the problem and research identification. Then related work is listed in the second chapter, including existing visualization platforms of data portals, visualization methods of semantic information of spatiotemporal data especially space time cube, and data quality. The third chapter describes the data and the methodology used here. A conceptual design is developed in this chapter. Then the prototype is implemented based on the conceptual design in Chapter 4. The fifth chapter illustrates the evaluation process of the effectiveness and satisfaction. The results are also displayed and discussed. Finally, the sixth chapter summarizes this thesis and gives recommendations for future work.

2 State of Art

This chapter reviews related literatures to this research. It first introduces some commonly accepted spatiotemporal visualization methods and discusses space time cube in detail. Then, since the utilization of space time cube visualization method needs to establish a 3D scene, the definition, classification, issues, and visual variables of 3D visualization are discussed. This chapter also discusses existing visualization platforms of data portals and their defects. With one of those defects is the lack of visualization of important metadata information, data quality, one of the important metadata in data query, would be discussed and one common component would be selected to represent data quality information.

2.1 Spatiotemporal Visualization

With large amounts of data generated every day, monitoring changes of phenomena has become a hot topic, which requires processing of data in different time periods. Visualization does not only focus on displaying information in a specific time but also dedicates to show changes of objects. This section reviews some commonly accepted time visualization methods and discusses space time cube in detail.

There are mainly five methods to visualize time (Ormeling & Kraak, 2010). The first method is single static map, which uses one map with symbols designed for different time periods to show changes. Using different color values to indicate objects in different time and using arrows to suggest the movement of objects are two common examples. The single static map is clear and easy for users to detect changes at the first glance. However, features may block each other. The second method is small multiple of maps, which uses individual maps covering the same spatial range but captured at different times to show changes. It is challenging for users to detect changes if many individual maps are displayed. Figure 2.1 and Figure 2.2 visualizes urban growth of the city of Maastricht using the single map and small multiple of maps separately. The third method is animated map. This method is similar to the second one but with more frames. Interaction is important for users to get useful information. Important information can be highlighted using several visual variables. The fourth method is cartogram. It is often applied in visualization of traveling time. Space in time cartogram is distorted according to the traveling time from the given origin.

Figure 2. 1. Using a single map to show urban growth of the city of Maastricht, the Netherlands (Ormeling & Kraak, 2010)

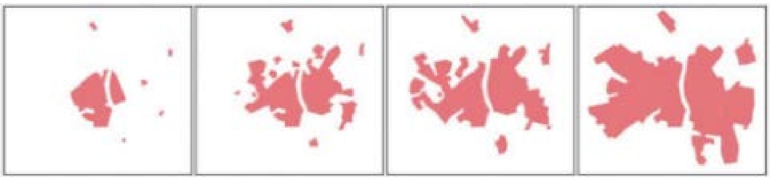


Figure 2. 2. Using small multiple of maps to show urban growth of the city of Maastricht, the Netherlands (Ormeling & Kraak, 2010)

Space-time cube, proposed in the end of last century (Hägerstraand, 1970), has been popular these days with the development of computer science. In space-time cube, time is represented by height. There are three kinds of spatiotemporal information that can be visualized by space-time cube (Kveladze, 2015). The first one is spatiotemporal path which depicts the path of moving objects. The well-known Minard map: ‘Napoleone’s March on Moscow’, in which the path of the army is a typical spatiotemporal path, has been adapted using space-time cube (Kraak, 2003). Figure 2.3 shows the new version of the Minard map. Space-time cube is popular in visualizing trajectories since movement patterns can be easily detected within this cube (Hägerstraand, 1970). The second one is station, which refers to static objects with varying attribute values change over time. For example, a researcher uses it to visualize crime data of each census in different years to detect the change of crime hot spots (Hashim et al., 2019). Figure 2.4 shows changes of hot spots pattern using space time cube. The last one is spatiotemporal prism, which indicates the available spatial range for a moving object. For example, space-time cube has been applied to visualize isovist of moving objects, which is shown in Figure 2.5 (Leduc et al., 2019).

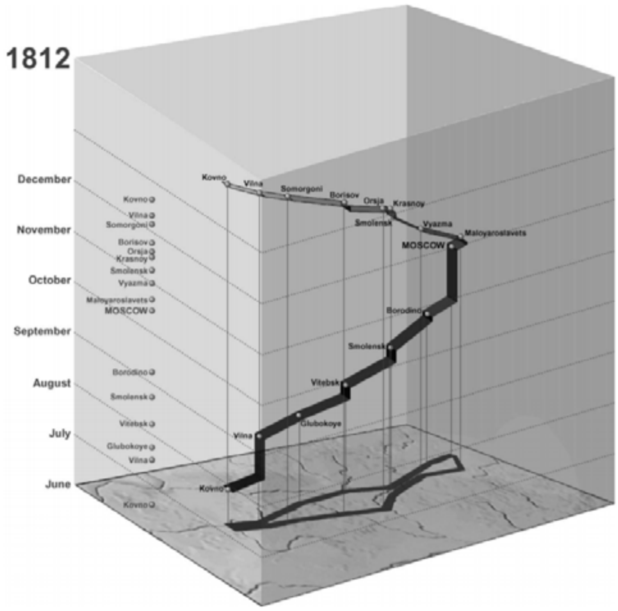


Figure 2. 3. Using space-time cube to show Napoleone’s march on Moscow (Kraak, 2003)

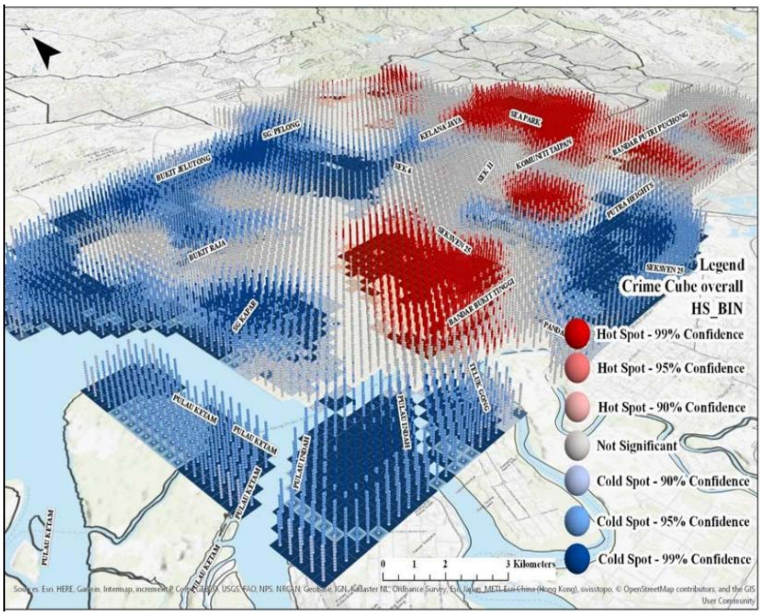


Figure 2. 4. Using space time cube to show crime hot spots pattern in 2017 (Hashim et al., 2019)

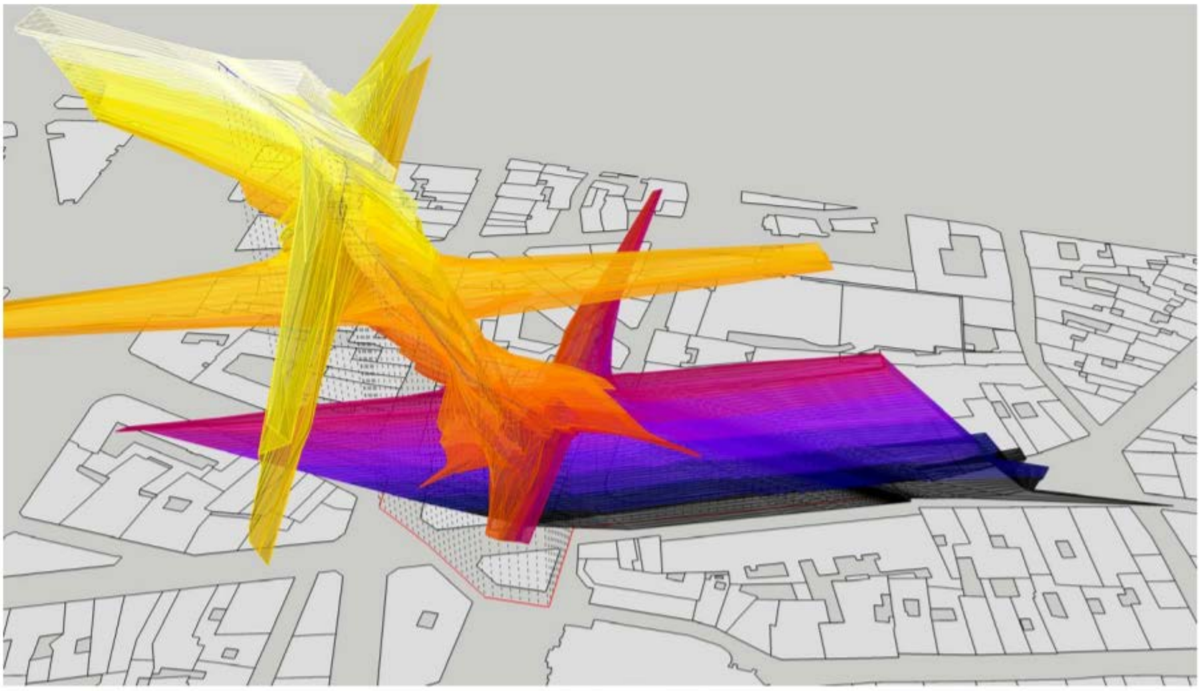


Figure 2. 5. Isovist visualization along one road using space-time cube (Leduc et al., 2019)

Nowadays, with the development of computation ability, space time cube has been utilized in many scenarios and incorporated with some novel techniques such as augmented reality (Turchenko, n.d.). It has not been applied in the visualization of data portals. In this project, space-time cube is accepted as the time visualization method of datasets to test if it is suitable here to help users find needed datasets.

2.2 3D Visualization Methods

Since space-time cube is accepted in this research, the visualization platform needs to establish a 3D scene, which requires more insights into 3D visualization. This section would discuss the definition, classification, issues, and visual variables of 3D visualization.

3D visualization has gained more attention with the development of computation ability of computers. Some studies have built up 3D platforms to aid in education (Sieber et al., 2016), and phenomena simulation (Kloeckl et al., 2012). 3D visualization usually uses perspective projection, transforms spatial information, and adds extra height dimension to represent more information (Petrovic, 2003).

3D visualization can be classified into photorealistic visualization and symbolic visualization (Pegg, 2012). The photorealistic visualization tries to simulate the real world. Symbols are always generalized real objects such as buildings and they try to leave a real-world impression on users (Semmo et al., 2015). Many platforms belong to this kind of 3D visualization and the well-known one is the Google Earth. The symbolic visualization pays more attention to designing 3D symbols.

Compared with traditional 2D visualization, 3D visualization has some unique characteristics such as use of light and observing view. There are no clear results showing that 3D maps are better than 2D maps. It depends on the application and the users' proficiency (Panchaud et al., 2012). Although there are many researchers studying the design and applications of 3D visualization, there are no official standards and systematic design guidelines to instruct cartographers in generating user-friendly 3D maps (Haeberling, 2004).

There are some issues which are critical in 3D visualization such as occlusion problem, data clutter, depth perception, and 3D annotation (Pegg, 2012). Moreover, since 3D visualization uses perspective projection, the scale is not fixed, which leads to different levels of detail regarding the distance from the position of object to the user's view (Pegg, 2012). For 3D visualization, interaction plays critical role in solving aforementioned problems. For example, 3D visualization can utilize brushing, filtering, and navigation techniques to solve occlusion problem (Bleisch et al., 2008).

Based on the studies mentioned above, it is difficult to know whether the application of 3D visualization in this research would help users perform query task effectively or not. Sufficient evaluation and user study needs to be carried out.

Although there are no standards of generating 3D maps, some researchers have proposed several design rules based on usual practice. Compared with traditional visualization, 3D visualization has more visual variables related to 3D rendering. Based on the steps of 3D maps design process, it can be divided into three design aspects including modelling of objects, symbolization, and visualization (Haeberling et al., 2008). In the modeling of objects aspect, it contains several sub-aspects such as digital terrain model object, models of map, models of orientating objects, and models of label objects. The visualization aspect is related to the camera setting, lighting, shadow, perspective, and atmospheric effects. These two aspects are more concerned with 3D scene generation, while the symbology aspect is more related to the 3D symbol design.

The symbolization aspect includes graphic appearance, texture, special aspects, and animations (Haeberling et al., 2008). This aspect is similar to the visual variables in 2D visualization such as symbol size, texture, and shape, while the visual guidance of these variables is not the same. For example, point symbol in 3D visualization can be rendered as translation, billboards, solid charts or custom symbols with different shapes. Figure 2.6 displays mountain cableways using billboards. Line symbol can be visualized as 3D curved lines with varying sizes representing different transport volumes in the flow map (Sieber et al., 2016).

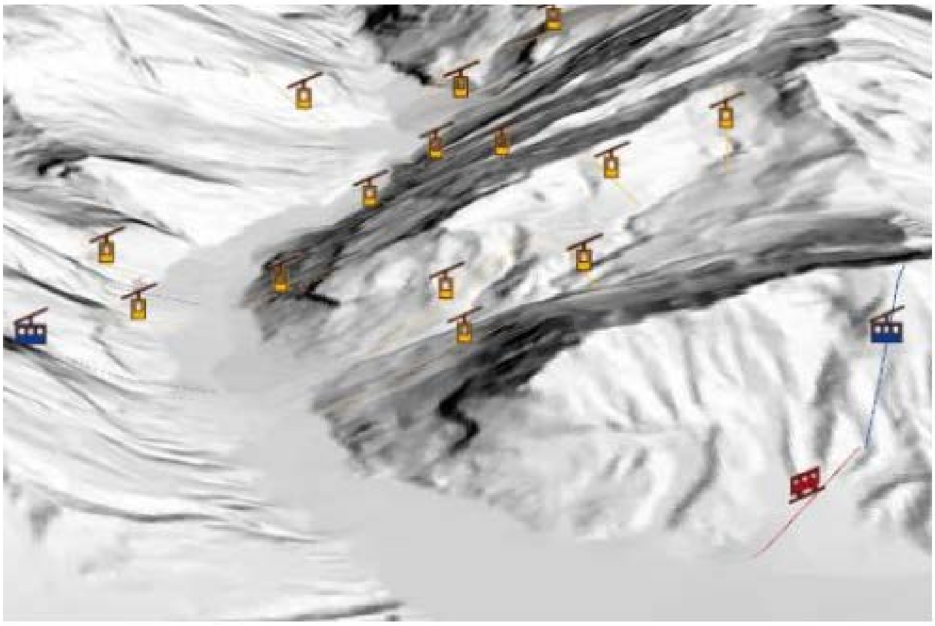


Figure 2. 6. Using shape variable to render point symbol using billboard (Haeberling et al., 2008)

2.3 Visual Data Query Platforms

In order to form the basic understanding of what the data query platform in this research should present, this section reviews the two main branches of data portal visualization. It also discusses the defects of existing visualization platforms and the trend of development in this area.

There are mainly two branches of the development of visual data query platforms. The first one focuses on improving the query of interlinked datasets and using maps to show the results. The second branch concentrates on how to design the visualization to give better overview of useful information.

The first branch is developed to help users better explore relations between interlinked datasets. It tries to visualize datasets and the relations between them so that it could be easier for users to explore the web of datasets. Considering the large number of datasets, these platforms tend to focus on the query process and then display the result. For example, Map4rdf is a faceted tool to display various types of datasets such as energy resources using point, line, and polygon symbols after executing queries (de León et al., 2012). The screenshot of the website that deployed this tool is shown in Figure 2.7 (Ontology Engineering Group, n.d.). When clicking features on the map, related information would pop up. Facete is another web-based platform which executes complex queries on interlinked datasets and uses point symbol to display results on a map (Stadler et al., 2014). If there is a large amount of data in a certain area, a special rectangular symbol is used to instruct users to zoom in. Another example is Sextant (Bereta et al., 2013). It is a tool that enables exploration of linked geodata and thematic mapping. It also visualizes data content directly.

Figure 2. 7. Screenshot from Map4RDF using point symbols to show locations of hills (Ontology Engineering Group, n.d.)

Gradually, visualizing time information become an important characteristic of the visualization platforms in the first branch. Sextant, as mentioned before, has developed into SexTant with capitalized t emphasizing time (Nikolaou et al., 2015). Time is visualized here by means of a time slider. Figure 2.8 is the screenshot of SexTant showing the change of land cover. Considering using time line to display temporal information, there is another example: Spacetime (Valsecchi & Ronchetti, 2014). It aims at helping users execute queries upon the DBpedia dataset and displaying results using point symbol on the map. Users can view datasets at different time by sliding the time slider.

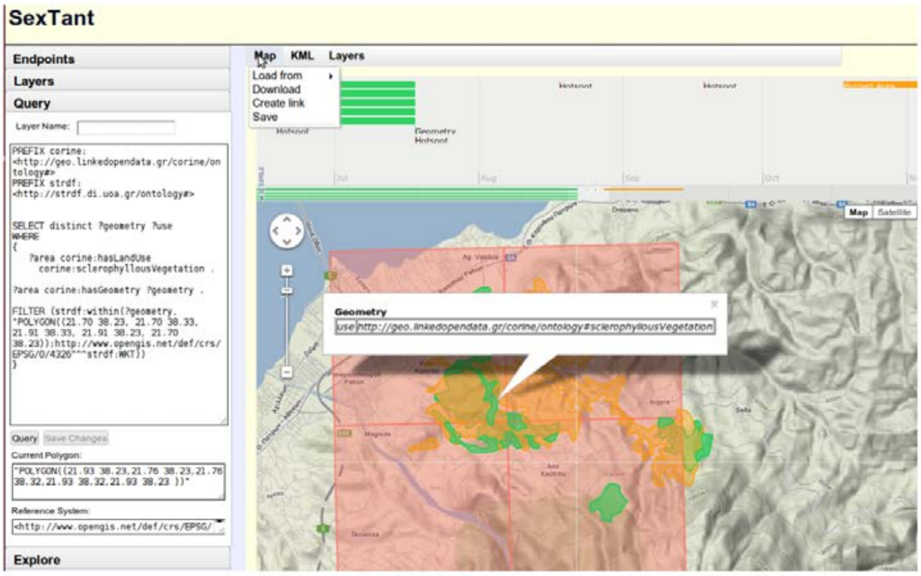


Figure 2. 8. Screenshot of SexTant showing the change of land cover (Nikolaou et al., 2015)

The second branch focuses on the visualization. For example, there is a platform for visualizing Brazilian Digital Library of Theses and Dissertations (Ramalho & Segundo, 2020). Instead of visualizing the data content directly, it visualizes statistic information with map and various kinds of charts to provide an overview of datasets in this data portal. Figure 2.9 shows the choropleth map that displays density of publications by geographic distribution, which gives a good overview of publication density.

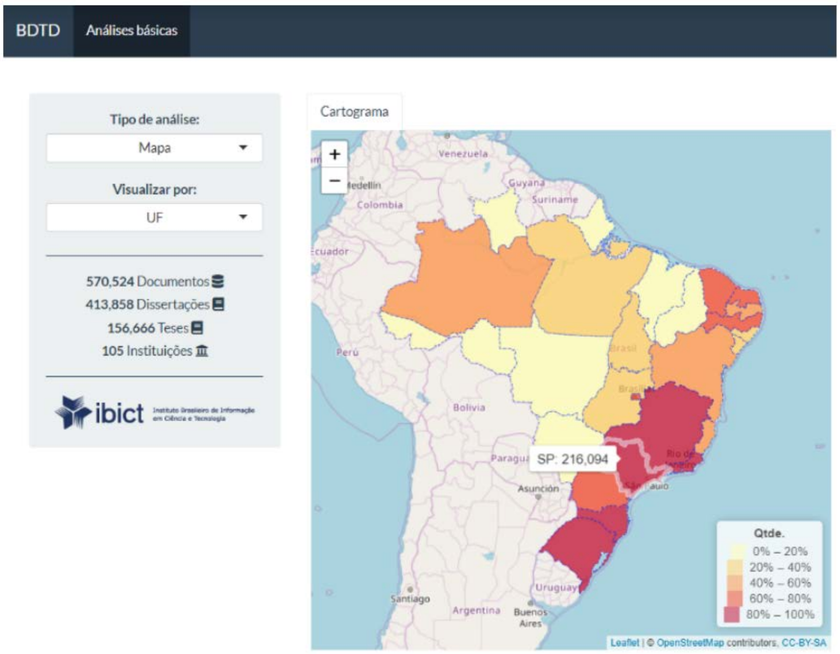


Figure 2. 9. Choropleth map in BDTD platform visualizing density of publications (Ramalho & Segundo, 2020)

However, there are still some shortness of these visualization platforms. Visualizing either all datasets in the database or the query result only, those platforms usually visualize the data content directly which would not give a good overview of important metadata. Some information is not shown, such as data quality, an important factor when users search for data. Moreover, considering the trend of incorporating time visualization, time slider is still the popular way and there are not many studies on applying some novel visualization methods.

Considering shortness mentioned above, visualizing aggregated metadata such as data quality may be better than visualizing data content directly. There are different components to assess data quality. Next section would discuss the components of data quality and which one should be selected.

2.4 Data Quality

There are various components of data quality in different standards and scenarios, which means it is difficult to visualize all components for different types of data. There are some important components that exist in almost all data transfer standards. In order to find out those components and select an appropriate one for this research, this section reviews traditional components of data quality and components in the Spatial Data Transfer Standard (SDTS). Then, completeness is chosen.

2.4.1 Components of Data Quality

Studies focusing on data quality can be dated back to 1990s. In the era of big data, data are generated at a surprising rate every day, but most data are not produced under systematic quality check. Whether data quality is good or not is not only associated with the data itself but also concerned with their applications. During these years, many studies have proposed different components of data quality and assessment standards. Commonly mentioned dimensions are listed in the Table 2.1.

Table 2. 1. Commonly mentioned components of data quality

Dimensions	Definitions
Accessibility	The difficulty level of access to data, which is usually concerned with openness of data (Cai & Zhu, 2015).
Accuracy	Attribute accuracy is related to the correctness of qualitive attributes or the accuracy of the quantitative attributes. For geodata, spatial accuracy refers to the difference between the position recorded within data and the real

	position (Veregin, 1999).
Completeness	Completeness refers to the extent to which all values and all attributes of all features are recorded (Veregin, 1999).
Consistency	Consistency refers to the apparent contradictions of datasets in the database (Veregin, 1999).
Readability	Readability illustrates whether the data can be interpreted (Cai & Zhu, 2015).
Timeliness	Timeliness refers to whether the data are up-to-date (Pipino et al., 2002).

2.4.2 Spatial Data Transfer Standard

With the emphasizing of data integration and data sharing, recording data quality within the metadata has become a requirement. Many institutions would develop data quality standard while publishing data standard. One of those is the Spatial Data Transfer Standard, which is well-known in the US (Veregin, 1999). The Spatial Data Transfer Standard aims at data exchanging. Data quality components in SDTS are listed in Table 2.2 (Veregin, 1999).

Table 2. 2. Data quality components in SDTS (Veregin, 1999)

Component	Description
Lineage	Refers to source materials, methods of derivation and transformations applied to a database. ● Includes temporal information (date that the information refers to on the ground). ● Intended to be precise enough to identify the sources of individual objects (i.e., if a database was derived from different sources, lineage information is to be assigned as an additional attribute of objects or as a spatial overlay).
Positional accuracy	Refers to the accuracy of the spatial component. ● Subdivided into horizontal and vertical accuracy elements. ● Assessment methods are based on comparison to source, comparison to a standard of higher accuracy,

	deductive estimates or internal evidence. ● Variations in accuracy can be reported as quality overlays or additional attributes.
Attribute accuracy	Refers to the accuracy of the thematic component. ● Specific tests vary as a function of measurement scale. ● Assessment methods are based on deductive estimates, sampling or map overlay.
Logical consistency	Refers to the fidelity of the relationships encoded in the database. ● Includes tests of valid values for attributes, and identification of topological inconsistencies based on graphical or specific topological tests.
Completeness	Refers to the relationship between database objects and the abstract universe of all such objects. ● Includes selection criteria, definitions and other mapping rules used to create the database.

Besides those assessment standards, there are mainly three functional forms to calculate those components (Pipino et al., 2002): The first one is simple ratio, defined by the ratio of the number of desired or undesired outcomes to the total number of outcomes. For example, completeness can be calculated by dividing the total number of values by the number of missing values. The second one is weighted average, which uses the weighted average of different variables to get the value of one indicator. Last ones are minimum or maximum functions. They assign the minimum or maximum of indicators as the measurements of one dimension.

Among all mentioned components, completeness plays an important role and is independent of data application (Pipino et al., 2002). This research would visualize completeness.

After deciding the visualized component of specific metadata, the visualization methods would be reviewed in the following sections.

2.5 Summary

This chapter begins with a review of current visualization platforms. After analyzing the defects of these platforms, this research decides to visualize aggregated metadata of all datasets to give a good overview of important information. Then, based on the basic idea of visualizing metadata, data quality, an important metadata, is reviewed. The components used to assess data quality are discussed and completeness is selected to be visualized. Additionally, Space time cube, together with 3D visualization, is reviewed and accepted as the theory foundations of this research's proposed visualization method.

3 Data and Methodology

In this chapter, a concept framework for visualization of multivariable spatiotemporal data is adapted. Then, three components of the concept framework, namely data framework, user task analysis, and the visualization framework, are discussed in detail.

Nowadays, visualization plays an important role in helping solving various kinds of problems. For example, there are many visualization platforms after the pandemic broke out, such as visualization of confirmed cases in different countries and trajectories of patients. However, whether users can get useful and needed information from those visualization results easily is a critical problem. Sometimes users find it really difficult to understand it. Previous works on user task analysis and knowledge of perception are needed to design user-friendly visualization methods. In order to instruct in the visualization design process, a concept framework for visualization of multivariable spatiotemporal data was proposed (Li & Kraak, 2010). In this concept framework, three components, namely user tasks, data framework, and visualization framework, influence each other and help in addressing visualization problems. Since the problem this concept framework solves is consistent with the visualization problem studied in this thesis, it is adapted here.

Three components in the concept framework adapted to this research, namely user tasks, data framework, and visualization framework, are illustrated in Figure 3.1.

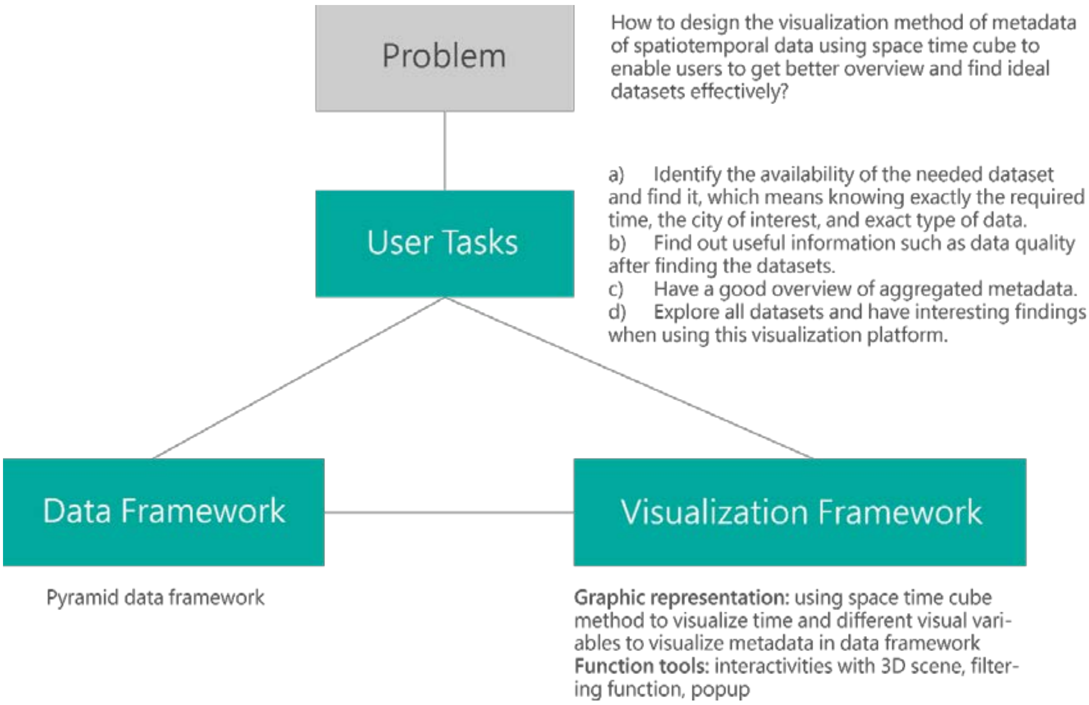


Figure 3. 1. The approach to visual problem-solving involving user tasks, data framework, and visualization framework

With various types of data generated every day, it is important to build up a data framework to abstract and organize data for later use. In order to do that, many data frameworks are proposed. Bertin has proposed his data framework, in which each variable has its length depicting the unit or possible values, its level of organization such as quantitative and qualitative, and the visual variable used to visualize it (Bertin, 1983). Later, a pyramid data framework was proposed with hierarchical structure and it combined the data framework of Bertin (Mennis et al., 2000). In this data framework, object is at the highest level with time, space, and attribute at the lower level. Each dimension can be divided into more variables. For example, there are different kinds of attributes with their own length and level of organization. This information is quite important in deciding visual variables. In this research, the pyramid data framework is accepted.

Table 3.1 shows how the datasets used here structure. Each dataset in the data framework has its own captured time information, spatial information, and attributes. In the space dimension, geographic coordinates are chosen since they can be easily transformed into other projected coordinates when necessary. In the time dimension, day, week, month, and year are possible components since most data are available daily, weekly, monthly, or yearly. As for the attribute dimension, data size, data category, and data quality are selected as components. Among those components, data size and data category are commonly used filters in big data portals. Since data quality is quite important in data reuse as discussed in previous chapters, it is also chosen. As reviewed in Chapter 2, completeness is one of the most common and important components in all data quality standards and is easily calculated. It is selected here to represent the data quality of datasets.

Table 3. 1. The pyramid data framework in the concept framework of this visualization

1st level	2nd level	3rd level	Length[1]	Level of organization[2]
Dataset	Space	Latitude	-	Quantitative
		Longitude	-	Quantitative
	Time	Day	Day	Quantitative
		Week	Week	Quantitative
		Month	Month	Quantitative

[1] Length refers to the units of quantitative attributes or possible values of qualitative attributes.

[2] Level of organization describes whether the component is qualitative or quantitative.

		Year	Year	Quantitative
	Attribute	Data size	Kilobit (KB)	Quantitative
		Data category	Geographic data, social data, economic data, nighttime light	Qualitative
		Completeness	none	Quantitative

3.1 Data Description

In this section, descriptions and resources of four types of datasets selected in this research, namely geographic data, social data, economic data, and nighttime light data, are given.

The first type of dataset is geographic data, which is a commonly used dataset in spatial analysis and mapping. In this research, a dataset which records the boundaries of each city in Jiangsu province is used as the test geographic dataset. It is downloaded from National Catalogue Service for Geographic Information on June, 2018. It records the zip code and name of each city. Table 3.2 shows the description of geographic data in detail.

Table 3. 2. Description of the geographic dataset used in this research

Dataset	Attribute name	Description
Geographic data	Latitude	These two attributes record sets of geographic coordinates of each feature
	Longitude	
	name	It records the name of each city
	Zip code	It records the postcode of each city

The second type is economic data which records the commonly used economic indicator values that could reflect the development of first industry, second industry, tertiary industry, medical and health services, residents' living standards, and public infrastructure construction. It is quite important in studying the economic development of a region. It is also downloaded frequently from data portals. Here the economic datasets in Jiangsu province are chosen as the test datasets. The datasets are collected from National Earth System Science Data Center on March 17th, 2019. Table 3.3 shows the description of this dataset in detail.

Table 3. 3. Description of the economic dataset used in this research

Dataset	Attribute name	Description
Economic data	Economic_id	It is used to identify different records
	Factor	It records the name of corresponding economic indicator
	Year	The time of this record
	Value	It records the value of corresponding value
	Postcode	It records the postcode of the city

The third kind of data is the social data. There are various types of social data, such as data from social media platforms, data of papers, and so on. Since social data are very important in researches studying social behaviors, social events, and society, they are really popular in many data portals. This research uses one type of social data as the test dataset that records the geographic locations of the institutions where the authors of published paper are located. It is obtained from Association for Computing Machinery Digital Library (ACMDL). Table 3.4 shows the description of this paper dataset.

Table 3. 4. Description of the social dataset used in this research

Dataset	Attribute name	Description
Social data	Paper_id	It is used to identify different papers
	Lat	They record geographic location of each institution where the authors of this corresponding paper are located
	lon	
	Date	It records the date of publication
	Author_order	It records the author's order of this paper
	Country_code	It records the country code of the institution where the authors of this corresponding paper are located

	Province_code	It records the province code of the institution where the authors of this corresponding paper are located
	County_code	It records the county code of the institution where the authors of this corresponding paper are located

The last one is nighttime light data. Actually, it belongs to the category of secondary data calculated from remote sensing images. There are also various types of data got from remote sensing, such as the DEM, satellite cloud images and so on. They are really important in geography, hydrology, forestry, and many other related fields. In this research, nighttime light data is chosen as the test dataset. It records the digital number (DN) value of each city that is calculated from the remote sensing image obtained by visible infrared imaging radiometer suite (VIIRS). This value could reflect the economic activity level of each city at night. The original remote sensing images are published in EARTHDATA. Table 3.5. shows the description of this nighttime light data.

Table 3. 5. Description of nighttime light dataset used in this research

Dataset	Attribute name	Description
Nighttime light data	Province	It indicates the name of the corresponding province
	Prefecture	It indicates the name of the city
	County	It indicates the name of the county
	Year	It records the year of this DN value
	Month	It records the month of this DN value
	DNvalue	It records the specific digital number value

As discussed in Chapter 2, components of data quality are not the same in different data standards. While among them, completeness exists in all data quality assessment. It is an important factor that indicates whether the data contains all necessary records. Knowing this information when downloading datasets can reduce the possibility of downloading useless datasets and improve the efficiency.

Completeness refers to the extent to which all values and all attributes of all features are recorded (Cai & Zhu, 2015). With this definition, it is appropriate to use simple ration functional form to calculate completeness. Here the completeness of each dataset is equal to the ratio of the number of available records to the number of all required records.

3.2 User Tasks Analysis

According to the aforementioned literatures in the Section 2.3 about visualization of data portals and interviews with open data portal users with different backgrounds, the user tasks are analyzed and summarized in this section. Knowing users and usual tasks well would help increase the usability of the visualization platform and it is also a key component in the concept framework adapted in this research.

As for the user groups, this visualization platform is not only designed for practitioners in geo-related fields, but also for all users who perform queries on data portals, hoping to find needed datasets for research, marketing, or exploration.

There are typically four kinds of user tasks. The first one is to identify the availability of the needed dataset and find it, which means knowing exactly the required time, the city of interest, and exact type of data. This one is the most common scenario. For example, users may want to find the economic dataset in Nanjing in 2009. The second one is to find out useful information such as data quality after finding the datasets, which would help users help users better decide whether to use it. The next one is related to having a good overview of aggregated metadata. When users only know parts of their queries, they need to get an overview of other attributes to make better decisions. For example, if users want to study the development of a city, they would like to know what types of datasets are available in that city. The last scenario describes higher level of overview such as spatial distribution. Users may want to explore all datasets and have interesting findings when using this visualization platform, as a good visualization platform is not only able to assist users in fulfilling basic tasks but also able to provide possibilities for data exploration. For example, by comparing data quality of the same type of dataset in each city, users may find out the spatial distribution of that and deduce more useful hypothesizes.

In summary, the visualization framework proposed in this research should be able to address questions related to the three forementioned user tasks.

3.3 Design Goals

Based on the user tasks analyzed in the previous section, the design goals（DG） of this visual interface are set and listed as follow:

DG1: Visualize spatiotemporal distribution of datasets with space time cube. Datasets usually have geographic reference and time coverage of datasets plays a critical role in their applications. Visualizing their spatial and temporal information at the same time would give a better overview of datasets coverage and allow users to quickly find datasets within the desired time range. Additionally, comparing changes of data over time could provide more possibility of data exploration.

DG2: Incorporate visualization of data quality with space time cube. Here appropriate visual variable would be applied to 3D symbols which represent datasets to visualize data quality. It aims at helping users quickly obtain the information of data quality and improve the efficiency of finding needed datasets for different purposes of usage.

DG3: Incorporate visualization of data category of datasets with space time cube. Having a good overview of semantic meaning of the spatiotemporal datasets is important for users. In this thesis, we take the category of the dataset as an example. By using 3D symbols in the established space time cube to visualize data category, this visualization model could give an overview of which types of data are available in the study area.

DG4: Incorporate visualization of data size of datasets with space time cube. Size of datasets matters when users want to download data. In this thesis, data size would be visualized by an appropriate visual variable applied to 3D symbols in the whole space time cube, aiming at giving an overview of the general volume of all datasets.

3.4 Visual Interface Design Methods

This section would illustrate the visual design of the visual interface. Firstly, all components that should be presented in this visual interface are decided and laid out. Then, the main symbol, which uses different visual variables and represents dataset, is shown. After that, the base map which shows the number of datasets available in each city and the interactivities of the visual interface are discussed.

3.4.1 Layout Design

Before laying out all components of this interface, what should be included in the visual interface needs to be decided first. The major component is the 3D symbol map. The whole 3D scene is a large space time cube containing 3D symbols which represent available datasets. Saturation, color, and size are applied to the symbols to visualize data quality, data category, and data size respectively. Map controls would be added to help users better understand the scene. Legend, as an important component of map in helping people understand the map content, is the second component. Since the symbolization of the base map and the 3D symbols are different, it would be clearer to separate them and create two legends. Another important component is the title of this visual interface. It should be prominent and describe the main idea of this visual interface. The fourth component is the selection panel. Filtering function can not only help users find datasets quickly but also help remove unnecessary information to give better overview. After checking the filter functions of some data portals, two frequently used filters, location and category, are selected. Two selection panels used for city filter and data category filter separately are included. The last one is the information panel. Since some usual tasks require users to know more details of datasets, it is important to include that and indicate resources in the information panel. The acknowledgement is also included in this information panel. But since it contains big blocks of text and acts as supplementary information, it is hidden in the main interface.

In summary, there are generally five components in this visualization interface, namely title, information panel of data and contributors, map, legend, and filter panels. These components are laid out in one page. Figure 3.2 illustrates the basic layout design.

Figure 3. 2. Layout of the visual interface of geodata browsing using space time cube

3.4.2 Symbol Design

Since the administrative levels that different datasets cover are not always the same, this project would take the city level as an example. Other administrative levels can be visualized with the same method and linked to each other.

As discussed in Section 2.1, space-time cube is used to visualize the capture time of datasets here. By applying the space time cube method, the whole space is established with z-axis representing time. In this large space, each dataset of each city would be visualized using a symbol, with height representing the time range. Since there are many datasets and the shape is a visual variable with high visual prominence, it is not suitable here to use a symbol with complicated shape. Cube, which is a quite simple and commonly used symbol in thematic 3D visualization, is used here to represent one dataset.

There are several ways to put those cube symbols on the map, such as the geometric center of the city polygon and the location of administrative center. Since the position of cube symbols indicating which city those datasets record, it is better to put them in the geometric center, or else it would leave a false impression that those datasets only records data in specific county. If there are some datasets of the same category indicating the same city but with different captured times, they would be visualized as cube symbols stacked on top of each other along the z-axis. Figure 3.3 shows those cube symbols.

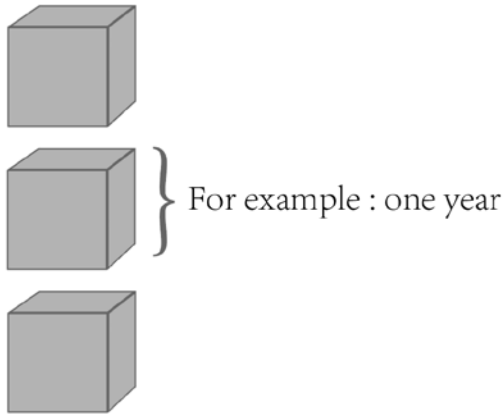


Figure 3. 3. Cube symbols representing datasets in the whole space time cube with size, color hue, color saturation representing data size, data category, and data quality separately

After deciding the shape of the symbol, visual variables need to be chosen to visualize selected metadata according to their level of organization which has been discussed in the data framework. Size is a powerful visual variable in reflecting quantitative difference. Since it is easy for users to link the size of dataset to the size of the symbol that represents dataset, the size of the cube is used here to visualize the size of dataset. Besides, color hue is a very common visual variable to represent qualitative attribute. It is applied to visualize data category, the only qualitative metadata in this research. Considering the occlusion problem and the visual burden of the 3D scene, color saturation is the appropriate one among the rest visual variables to visualize completeness, which is also a quantitative metadata.

3.4.3 Base Map Design

Except the 3D symbols, base map is necessary to show basic geographic information and can be used to show supplementary information. Since the number of available datasets in each city is useful in addressing problems related to the second type of user task, base map is designed to visualize that information to give a good overview and help users to compare the open data work in each city. There are several ways to visualize that information, such as cartogram, choropleth map and charts. Using chart symbols in each city could address more information but the map would be less clear. As for the cartogram, since base map should be combined with data cubes, it would be very confusing to distort the space. In this case, choropleth map is chosen. The number of available datasets would be visualized by different color values.

3.4.4 Interaction Design

Considering the aforementioned literatures about 3D visualization and interviews with targets users who have previous experience with data portals and web-based 3D visualization platforms, three types of functionalities are designed.

The first set of functionalities is interactivities with the 3D scene. Some people reported that poor interactivities with 3D scene would make it difficult to understand the map. The internal defects of 3D scene such as occlusion problems would cause visual cognition difficulties and it would take users longer time to finish usual tasks. In this 3D scene, typical interactivities are designed including zooming in, zooming out, and rotation.

The second set of functionalities is the filter function, which is designed to help users better perform queries and explore the interface. Filtering functions are provided as a common practice in data portals. The location and data category are quite common filter conditions that exist in almost all data portals. So, in this visual interface, users can filter those datasets by city and category.

The last set of functionalities aims to design pop-up windows or tooltips when users exploring this interface. Since the visualization could not explain every characteristic of data to users, supplementary information should be shown when users query specific datasets. In this case, when users hover a data cube, a popup will be shown indicating information about data category, data size, data quality, and time.

3.5 Summary

This chapter describes the details of datasets and methodology used here. It begins with a concept framework for visualization of multivariable spatiotemporal data. Then, three components of this framework are discussed and applied to the visual interface. The application of this design is elaborated in the next chapter, using Jiangsu province as the study area.

4 Case Study and Results

In order to evaluate the proposed visualization method, a prototype was implemented in this chapter. Jiangsu province, which is one of the most economically active provinces in China, was selected as the study area. In the first section, completeness of each dataset would be calculated. Then, based on the methodology discussed in the previous chapter, steps to build up the prototype and the final results are shown in the following sections.

4.1 Test Area and Data Processing

Jiangsu province, one of the most economically active provinces in China with many available datasets, is chosen as the study area. It is located in the Yangtze River Delta and borders Shanghai. The landform of Jiangsu is composed of plain, water area, low mountains, and hills. It is also one of the birthplaces of ancient Chinese civilization, with a total of 13 national historical and cultural cities. So, there are various kinds of data available in Jiangsu for this research. Some of them are chosen and collected.

In this section, four categories of data, namely geographic data, economic data, social data, and nighttime light data in Jiangsu province, are collected and processed. In order to visualize the data quality of these datasets, the completeness needs to be calculated. As discussed in Chapter 3, the completeness would be equal to the ratio of the number of available records to the total number of records.

For the geographic data, each record of each city has all attribute values and polygons of all cities are available in every available year, which means the completeness of this type of test datasets are all 1. For the social data, all papers have been collected from the original dissertation database and all records have all corresponding attribute values, which means the completeness of this type of datasets are also 1.

For the economic data, since there are totally 129 economic factors according to the standard of Bureau of Statistics, the completeness value is equal to the total number of available factors divided by 129. The result is shown as in Table 4.1.

Table 4. 1. Completeness of each economic dataset.

City	Postcode	2009	2010	2011	2012	2013
Nanjing	3201	0.89	0.97	0.99	0.87	0.64
Wuxi	3202	0.89	0.97	0.99	0.87	0.64
Xuzhou	3203	0.89	0.97	0.99	0.87	0.64
Changzhou	3204	0.00	0.00	0.00	0.00	0.00
Suzhou	3205	0.89	0.97	0.99	0.87	0.64
Nantong	3206	0.89	0.97	0.99	0.87	0.63

Lianyungang	3207	0.89	0.97	0.99	0.87	0.63
Huai'an	3208	0.00	0.00	0.00	0.00	0.00
Yancheng	3209	0.89	0.97	0.99	0.87	0.63
Yangzhou	3210	0.89	0.97	0.99	0.87	0.63
Zhenjiang	3211	0.89	0.97	0.99	0.87	0.63
Taizhou	3212	0.89	0.97	0.99	0.87	0.63
Suqian	3213	0.89	0.97	0.99	0.87	0.63

For the nighttime light data, since not all remote sensing images of all months in the visualized time range are available, the completeness of each year is calculated as the number of available months divided by 12, which is the total number of months in a year. Table 4.2 shows the result.

Table 4. 2. Completeness of each nighttime light dataset

City	Postcode	2013	2014	2015	2016	2017	2018	2019	2020
Nanjing	3201	0.83	0.83	0.67	0.92	0.58	0.92	0.92	0.08
Wuxi	3202	0.83	0.83	0.67	0.92	0.58	0.92	0.92	0.08
Xuzhou	3203	0.83	0.83	0.67	0.92	0.58	0.92	0.92	0.08
Changzhou	3204	0.83	0.83	0.67	0.92	0.58	0.92	0.92	0.08
Suzhou	3205	0.83	0.83	0.67	0.92	0.58	0.92	0.92	0.08
Nantong	3206	0.83	0.83	0.67	0.92	0.58	0.92	0.92	0.08
Lianyungang	3207	0.75	0.75	0.58	0.92	0.67	0.92	0.92	0.08
Huai'an	3208	0.75	0.75	0.58	0.92	0.67	0.92	0.92	0.08
Yancheng	3209	0.75	0.75	0.58	0.92	0.67	0.92	0.92	0.08
Yangzhou	3210	0.75	0.75	0.58	0.92	0.67	0.92	0.92	0.08
Zhenjiang	3211	0.75	0.75	0.58	0.92	0.67	0.92	0.92	0.08
Taizhou	3212	0.75	0.75	0.58	0.92	0.67	0.92	0.92	0.08
Suqian	3213	0.75	0.75	0.58	0.92	0.67	0.92	0.92	0.08

4.2 Graphic interface

After preprocessing needed datasets, this section describes some details about building up the prototype. In general, this platform would focus on Jiangsu province and each dataset of each city would be represented by a cube symbol. It is designed to allow users find needed datasets easily and give good overview of important aggregated metadata.

As mentioned in Chapter 3, the base map is a choropleth map using different color to indicate the number of available datasets in each city to give a good overview. The number of datasets in each city ranges from 0 to 28. Since the distribution of that is non-uniform distribution after examining those numeric values, natural breaks method is used here for classification with the number of classes 4. In order to give a clear base map without influencing the cube symbols which use different color to indicate different data categories, four different gray values are chosen to indicate the number of datasets. Additionally, in order to highlight the study area and not to leave the impression that only Jiangsu province has datasets, two buffer layers are added. The label of cities in Jiangsu province are preserved while those in cities of other provinces are deleted. The custom map is shown as Figure 4.1. Map controls such as compass, scale, and legend are created via programming.

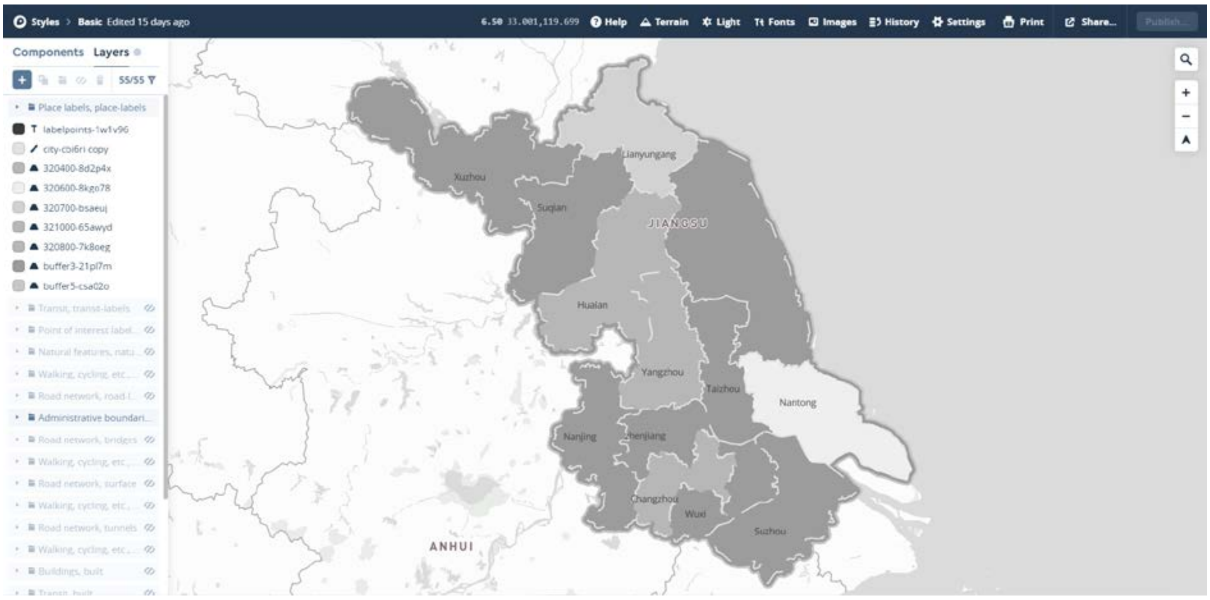


Figure 4. 1. Custom base map using different gray values indicating different number of datasets in Mapbox Studio

According to methodology in Chapter 3, color is applied to visualize data category. Four colors are applied to indicate four types of data with green, yellow, red, and dark blue representing geographic data, social data, economic data, and nighttime light data respectively. Additionally, color saturation is used to visualize completeness. Since color with low saturation may leave an uncertainty impression, the saturation of each cube is equal to the corresponding color with full saturation multiply the corresponding completeness value of each dataset. As for the data size, it is intuitive to use the size of cube to visualize it. The smallest data size is 9kb, while the largest one is 227kb. In order to give an intuitive impression and reduce the possibility of overlapping between different types of datasets, the linear function is chosen with adapted coefficients. After comparing several linear functions, the linear function between the real data size and the size of cube shown as follow gives the best visualization result.

$$\text{Size of cube} = \text{data size} * 0.9 + 50$$

Since the time of most datasets uses the unit year, the height of each cube would represent one year.

Figure 4.2 shows the visualization result of all cube symbols and the base map generated in previous section.

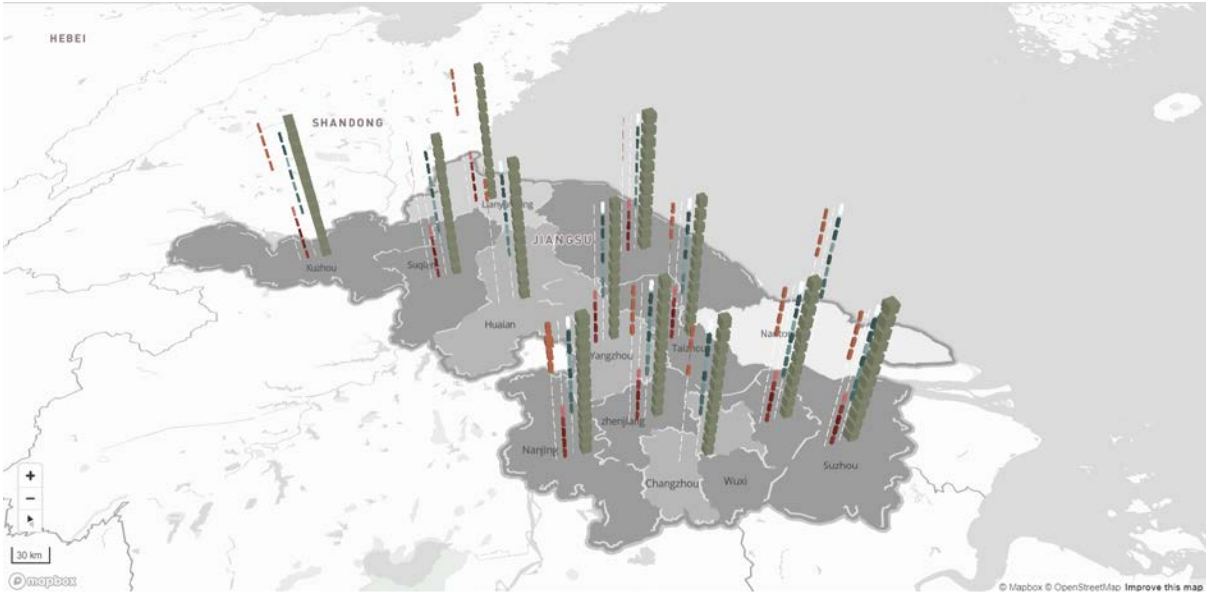
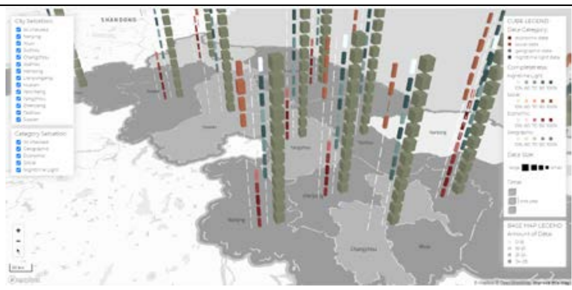
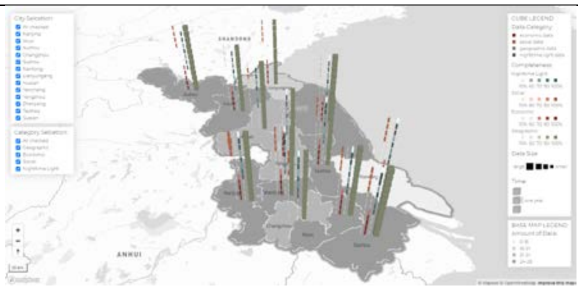
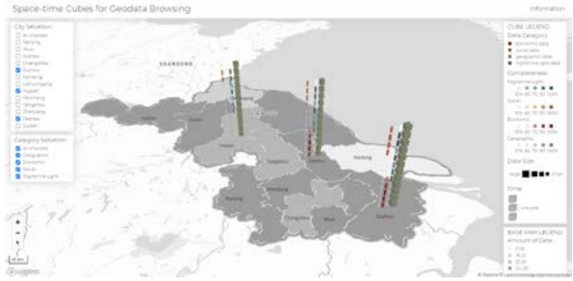
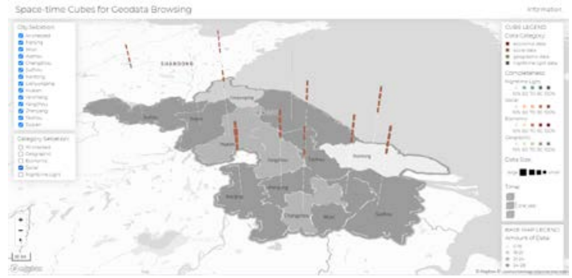
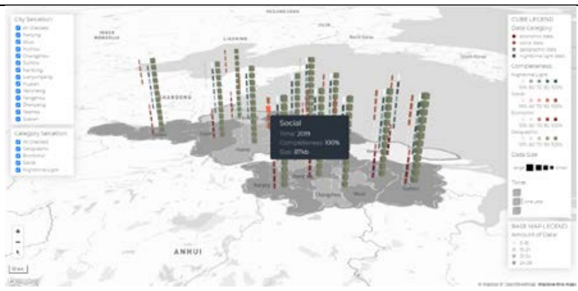


Figure 4. 2. Custom base map with cube symbols in this visual interface

As mentioned in Chapter 3, there are three sets of functionalities need to be implemented in this prototype. The first one is interactivities with 3D scene. In this set, zooming in would allow users to have a detailed look of each data cube. Zooming out function would allow users to have a good overview of all cubes. By scrolling the mouse wheel, users could zoom in and zoom out. Rotation is another quite important interactivity. When users press the left and right mouse buttons at the same time and move the mouse, the whole scene would rotate, which could solve the occlusion problem to some extent. The second set of functionalities is the filter function. Users can check or uncheck the checkboxes of city selection panel and data category selection panel. Only cubes of the checked cities and checked data categories would be displayed. This function really helps users find needed datasets especially when there is a large amount of data. The last set is to show related information of each dataset cube. When users hover a dataset cube, a popup would be displayed. Table 4.3 summarizes the functionalities implemented in this prototype.

Table 4. 3. Functionalities in the prototype

Category	Functionality	Implementation	What does it do?
Interactivity with 3D scene	Zoom in and zoom out		Zooming in allows users to have a detailed look at datasets. Zooming out allows users to have a good

			overview.
	Rotation		Rotation could solve the occlusion problem to some extent.
Filter	Area filtering		Filter function could assist users in finding needed datasets.
	Category filtering		
Hover	Popup		The popup of each cube could give more detailed information about the hovered dataset.

There are many map providers delivering web map services such as Mapbox, HERE, leaflet, and so on. This prototype selects Mapbox. Mapbox GL JS is a JavaScript library, published and maintained by Mapbox, which uses WebGL to render interactive maps. Base map is highly customized in Mapbox studio and published as tiles which can be rendered and shown by Mapbox GL JS. Cube symbols could be created using a plugin named Threebox.js, which provides a convenient way to manage 3D objects with geographic coordinates and to synchronize the map and 3D scene cameras.

The website is available at https://unruffled-dijkstra-9403b9.netlify.app.

Figure 4.3 is the cover page of the visual interface. It introduces the aim of this research, basic information of test area, and some instructions. It is really useful to give a rough impression of the research background before users enter the main visual interface. Moreover, it can tell users that the detailed description of each dataset is shown in the hidden information panel. Additionally, some notifications for participants in the evaluation stage can be displayed here.

Figure 4. 3. Cover page of the visual interface: space time cube for geodata browsing

Figure 4.4 is the main visual interface which displays the 3D map. Users can use mouse to navigate through this 3D scene and query for further information of each dataset by hovering the corresponding data cube. Legends are displayed to help users understand the visualization. What's more, users could check and uncheck cities and data categories to show parts of data. When users click information link in the top right corner, the information panel would show up.

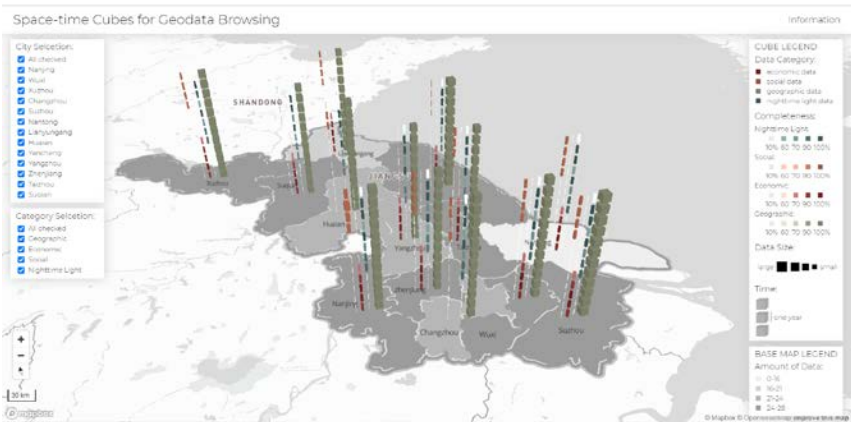


Figure 4. 4. The visual interface to show available datasets of each city in Jiangsu province

Figure 4.5 is the information panel. It starts with the description of each type of dataset, including their meaning, resources, and collecting time. Then, detailed explanation of each visual variable utilized here is given, followed by contributor information.

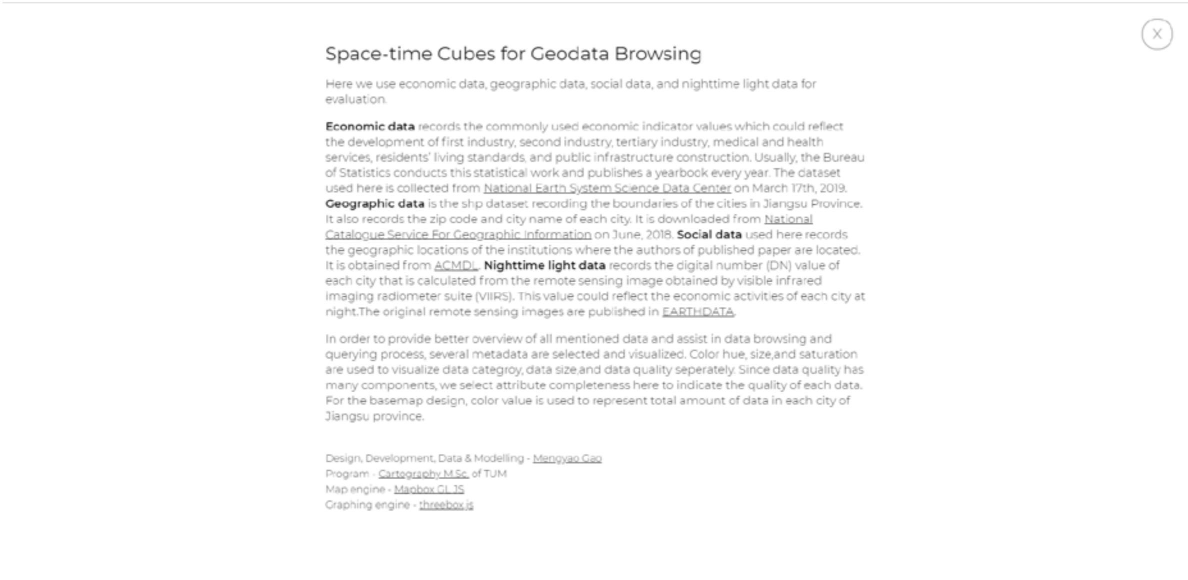


Figure 4. 5. The Information panel including description and resources of datasets and contributors

This visual interface assists users in performing usual tasks, as discussed in section 3.2. For the first task, if the user wants to find the economic dataset of Nanjing in 2009, he could use the city and data category filter functions. By only checking economic data and Nanjing, he could find all available economic datasets in Nanjing. Then he can get the time information by hovering some cubes to find out the needed dataset. For the second task, if the user wants to do economic related studies on Suqian, he can check the city Suqian only. Then he could find there are four types of datasets available. He could use economic datasets and nighttime light datasets in available years. If he doesn't know what the nighttime light data records, he can check the information panel. As for the last task, it is easy to get a good overview. For example, when users only check the social data, he could find that the number of papers published in the northern part of Jiangsu province is less than that in the southern part. Then he may develop a hypothesis such as the research development level in the southern part is higher than that in the northern part.

4.3 Summary

This chapter describes how the visualization method proposed in Chapter 3 being implemented into a prototype using suitable techniques. The detailed steps of generating each component are elaborated. The evaluation of this prototype is going to be discussed in the next chapter.

5 Evaluation

In this chapter, evaluation methods are reviewed and suitable methods for evaluating the effectiveness and satisfaction of the prototype are chosen. The details and procedures of those methods applied here are discussed later. After finishing this evaluation process by collecting questionnaires' responses from participants, the results are shown and explained.

5.1 Evaluation Methods

This section illustrates the dimensions of usability and lists common methods to evaluate that. Then, appropriate methods are selected. How those methods are applied in the evaluation stage and why specific questions are chosen here are discussed in detail.

According to the ISO, the usability refers to effectiveness, efficiency, and satisfaction (International Organization for Standardization [ISO], 2015). In this research, the effectiveness means the accuracy and completeness users could finish usual tasks mentioned in Chapter 3 using this prototype. The efficiency is related to resources such as time needed to achieve that accuracy and completeness when users finish user tasks such as finding datasets, having a good overview, discovering spatial distribution, and so on. The last aspect, satisfaction, refers to the comfort and satisfaction when users using this visual tool. Since this research only wants to evaluate whether space time cube, this novel visualization method, could assist users in finishing usual tasks, only effectiveness and satisfaction of this visualization method were evaluated.

Many methods has been proposed these years for usability evaluation and there are mainly two categories, namely quantitative method and qualitative method (van Elzakker & Wealands, 2007). Commonly used quantitative methods include user survey, eye-tracking, remote evaluation, and so on. Qualitive methods includes interview, task analysis, user observation, and so on. To evaluate the effectiveness, the common method would set some benchmark tasks for users to finish and check the accuracy and completeness of each task. In this study, it is adapted and the correctness of users' answer of each task would be used to assess the effectiveness. To evaluate the satisfaction, qualitative methods such as interview would be appropriate since attitude towards a visual interface is quite subjective. Traditional way to do the interview is to ask interviewers some well-structured questions face to face. However, because of social distance restriction, online questionnaire would be created here with well-structured questions and distributed to participants.

5.1.1 Profile of Participants

The participants of this experiment included not only Europeans but also Chinese, since Chinese are more familiar with the test area and tend to use the data portal of Jiangsu province. There were 32 participants in total with 19 females and 13 males. Twenty-eight participants were between 20 and 30 years old. Nighty-four percent of them had used data portals before. Twenty-three participants had used web-based 3D visualization several times before while seven participants had no experience at all. Two participants had declared that they used 3D visualization platforms a lot.

5.1.2 Benchmark Tasks

There are mainly two types of tasks: finding needed datasets and getting information overview. For the first type, one task is designed to test if users could find the needed dataset effectively. For the second type, three tasks of overview, namely the overview of time, data category, and data quality, are designed. The last task is related to spatial distribution, which refers to higher level of overview.

Therefore, there are totally 5 benchmark tasks with each of them having different user scenarios, search aims, search area, search time, search attribute, cognitive operations, and different levels of data availability. Table 5.1 summarizes all tasks with the aforementioned aspects. Task 1 aims to test if users could find desired dataset when knowing the required time, city of interest, and the exact type of data. For example, the user may want to get the economic data in 2009 in Nanjing. Since the data size information would be shown in the popup after finding the corresponding cube, this value is used here to indicate whether the user finds the cube/data correctly. Task 2 aims to test if users could have an overview of available time. For example, when users decide the field of their studies and want to find related data, users need to find out when the data are available. Task 3 aims to find out if users could have a good overview of available data categories. For example, when users want to study the development of a specific city, what types of data are available is an important factor of deciding which aspect to evaluate. Task 4 aims to test if users can identify and compare data quality, which is one of this research's innovation. For example, when users find the desired type of datasets, users need to compare the data quality, which is important for further use. Task 5 aims to test if users could find out spatial distribution, as a good interface should not only make it easy for users to find data, but also provide possibility for data exploration. Users may want to dig into the spatial distribution of data and propose original hypothesizes.

Table 5. 1. Benchmark tasks for evaluating effectiveness of this visualization platform

Tasks	T1	T2	T3	T4	T5

aim	To test if this platform can enable users find needed datasets	To test if this platform can give an effective overview of time	To test if this platform can give an effective overview of data category	To test if users can identify and compare data quality, which is one of this project's innovation.	To test if this platform can enable users explore spatial distribution.
Search Area	Single	Single	Single	Single	Multiple
Search Time	Single	Multiple	Multiple	Multiple	Multiple
Search Attribute	Single	Single	Single	Single	Single
Cognitive Operation	Identification	Summary	Identification, summary	Identification, comparison	Comparison, summary, deduction
Data Availability	High	High	High	High	Middle

There are several ways to structure those tasks, such as multiple choice, filling blanks, true or false, and so on. Since multiple choice only lists some options which may give hints to users, it is difficult to detect wired responses from users. As for filling blanks, users may write some answers randomly if they cannot find related information. So, presenting those tasks as statements and let participants judge if it is true or false would be more suitable here. Three options, namely true, false, and not sure, are given. By setting options true and false, it cannot give any hint to users and could detect if the users really find right information. Moreover, by setting the option not sure, it is possible to know why users cannot find the right answer. For example, if the user chooses the option false, he may not find the data cube correctly. If the user chooses not sure, he may not find any related information. Table 5.2 lists all the statements and their answers.

Table 5. 2. Statements and correct answers in benchmark tasks used for evaluating the effectiveness of the visual interface

Task	Statement	Correct answer
T1	In 2009, the size of economic data in Nanjing is 13kb.	True

T2	Nighttime light data from 2012 to 2020 in Wuxi is available.	False
T3	There are two kinds of data available in Lianyungang.	False
T4	The completeness of nighttime light data in Suzhou in 2019 is better than that in 2017.	True
T5	Institutions in the southern part of Jiangsu province published more paper than that in the northern part.	True

5.1.3 User Attitude Investigation

After users finished benchmark tasks, they were asked to answer some questions related to their attitude towards this visualization platform. These questions are created to evaluate the satisfaction of this proposed visualization method. There are five questions in total. The first and second questions are concerned with the usability of this visual interface, with the first one asking about the general feelings when finishing benchmark tasks and the second one focusing on if the visual interface gives all needed information effectively. The third question is about users' attitude towards space time cube. The fourth question asks for users' suggestions, which could help in the future work. The last question is an open question allowing users to add any comment. All questions are listed in Table 5.3.

Table 5. 3. Questions in user attitude investigation used for evaluating the satisfaction of the visual interface

Number	Questions
Q1	How do you rank the overall ease of usability of this visualization interface?
Q2	Do you think the map elements are sufficient enough to assist you in fulfilling all tasks? If not, what else would you like to add?
Q3	Do you think the space time cube is suitable to visualize spatiotemporal data in this platform? If not, why?
Q4	What would you suggest to improve this visualization interface?
Q5	Additional comments?

5.1.4 Procedure

The tool used for creating the questionnaire is Qualtrics. The whole questionnaire is divided into three parts. The first part collects users' basic information and the second part lists all the benchmark tasks used to evaluate the effectiveness of this visualization platform. The last part has some questions related to users' feedback and subjective attitude towards this visual interface. The whole evaluation process lasted for one month. Participants were told at the beginning of the questionnaire that their responses would only be used for this research. Appendix 1 is the questionnaire.

5.2 Evaluation Results Analysis

In this section, effectiveness of this visualization platform is estimated by the correctness of benchmark tasks' answers. Answers of questions in the user attitude investigation would reflect the satisfaction of this visualization interface.

5.2.1 Effectiveness

Effectiveness of this visualization platforms refers to if users could finish usual tasks, such as finding the datasets and getting good overview of important information. Table 5.4 summarizes the correctness of each task.

Table 5. 4. Correctness of each task in benchmark task stage

Task	Correctness	Rate of not sure
T1	0.78	0.06
T2	0.88	0.03
T3	0.75	0.06
T4	0.91	0.06
T5	0.38	0.59

Judging from this result, the overall performance is good. The first four tasks received relatively high correctness and only a few users reported not sure, which means that participants could identify needed data cube among all these cubes and this platform could help users gain a good overview of time, data category, and data quality. However, the last task got really low correctness and 19 of participants reported not sure instead of giving the wrong answer.

The correctness of the first task is 0.78 and the rate of people giving not sure is 0.06. There were 5 participants choosing false. Among them, three of them had never used web-based 3D visualization before. The only one male participant who reported using data portals usually and using web-based 3D visualization a lot chose the false option and he gave three wrong answers out of five tasks. Moreover, he didn't give any comment. In this case, one possible reason maybe that he didn't take this questionnaire seriously and the other possible reason maybe that the 3D visualization in this visual interface is quite different from those he had used before, which made him quite difficult to get used to it. There is one participant said that she found that the overlap between cubes brought difficulties to users to select wanted data cube. As for the other 3 participants, the reason why they didn’t give the right answer is may because of their lack of 3D visualization experience. In general, most participants chose the right answer.

The correctness of the second task is 0.88 and the rate of users reporting not sure is quite low, which is a quite good result. It shows that the range of available time is quite clear for most participants.

The correctness of the third task is 0.75. Most participants who reported false gave the feedback that the occlusion problem made it difficult to have a good overview, which had a bad impact on the result. Those feedbacks are quite interesting. The possible reason for that is that they didn’t know how to rotate the 3D scene and had problem with using the filter function. Lack of experience of using 3D visualization before and less clear instruction of how to interact with the 3D scene may explain that. In general, most participants gave the correct answer.

The correctness of the fourth task is the highest. With the correctness of 0.91, it shows that it is really effective to compare the data quality between different datasets in this visual interface.

The result of the last task is the worst with the correctness only 0.38. However, instead of choosing the false option, most participants reported not sure. One reason for that could be that to judge the statement, users need to understand the meaning of social data first which is shown in the hidden information panel. Some users didn’t read the introduction page and the hidden information panel carefully and had no idea what the last task asks about.

In summary, the overall performance is good. This interface can assist users in identifying datasets and having a good overview of important metadata. However, for the possibility of exploring spatial distribution, more evaluation should be carried out.

5.2.2 Satisfaction

The satisfaction of this visualization platform is evaluated from the answers of questions in the user attitude investigation.

For the first question *“How do you rank the overall ease of usability of this visualization interface?”*, Figure 5.1 shows the result. 12 participants reported neutral while among the remaining participants, 9 of them reported somewhat difficult. It is quite interesting that the total number of participants thinking it was easy or somewhat easy is almost the same with that who thought it was somewhat difficult or difficult. According to the following questions' answers of participants who thought it was difficult or somewhat difficult, the reasons may be as follow. The first one is that the interface is not adaptive. Participants' computers having higher resolution or larger screen size would get smaller font size and large spaces between data cubes. While, if their computers have lower resolution or smaller screen size, the legend would be cut off and the hover function may fail. Another reason is that some participants thought the overlap of those cube symbols make it difficult to find needed datasets.

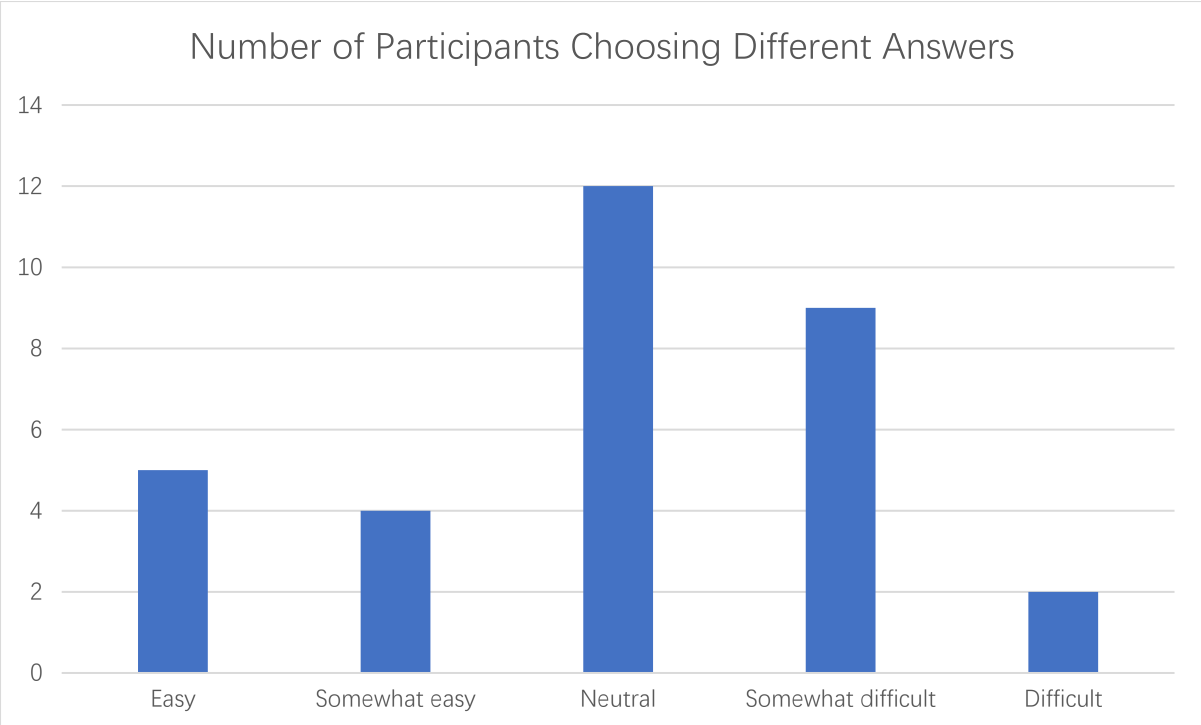


Figure 5. 1. The result of question "How do you rank the overall ease of usability of this visualization interface?"

For the second question "*Do you think the map elements are sufficient enough to assist you in fulfilling all tasks? If not, what else would you like to add?*", 21 participants reported yes and Figure 5.3 shows the result. There are some suggestions from participants who reported no. Base map, which indicates the number of datasets, was not easily understandable for some of them and one participant said that it would be clear to give number directly instead of using color. One other comment is that the z-axis could be created with time labeled.

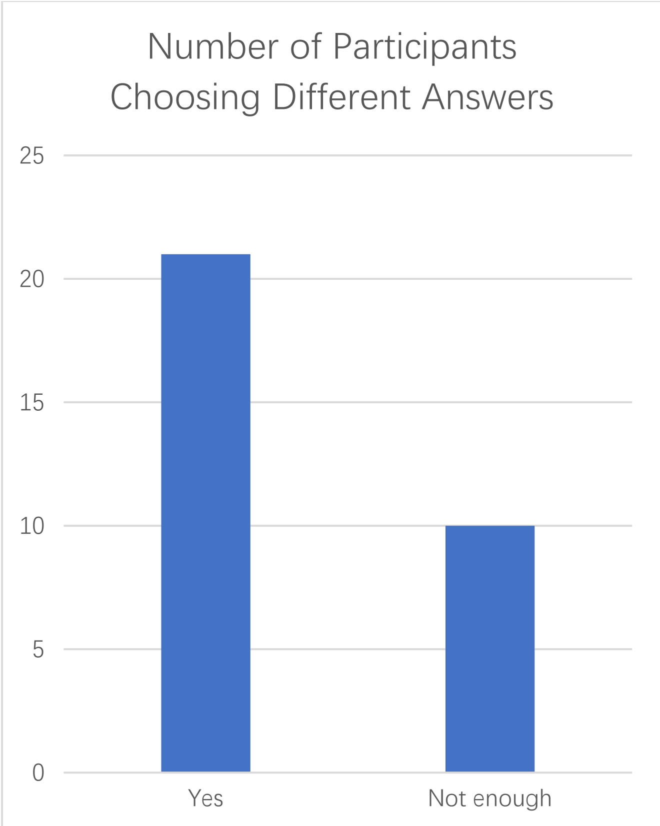


Figure 5. 2. The result of question "Do you think the map elements are sufficient enough to assist you in fulfilling all tasks?"

For the third question "*Do you think the space time cube is suitable to visualize spatiotemporal data in this platform? If not, why?*", Figure 5.3 shows the result which is quite diverse. 21 participants reported space time cube is suitable in this case. While there are several users thought the space time cube is more suitable to visualize process over time than for comparative tasks. Specially, several participants reported that the occlusion problem in 3D scene made it difficult to have a good overview and rotation would help. This comment indicates that for participants who were not familiar with 3D interface had trouble with interaction with 3D scene, which had a negative impact on the overall experience of this 3D visualization platform.

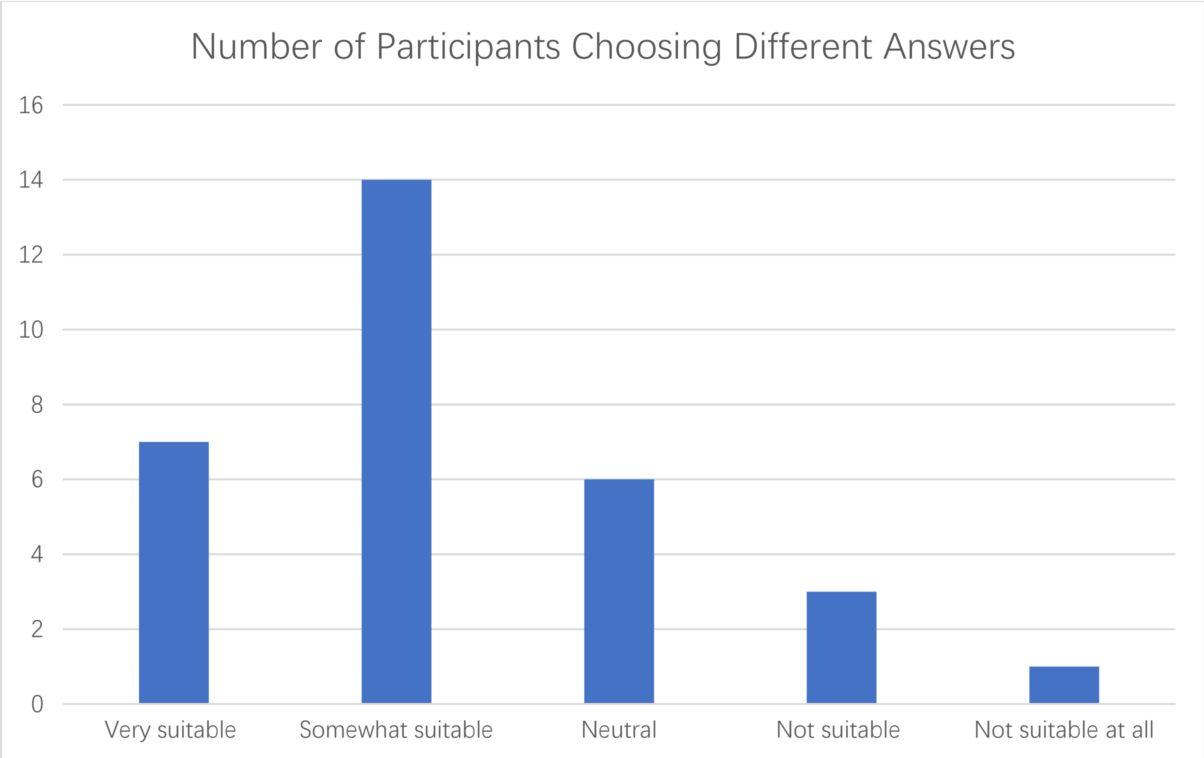


Figure 5. 3. The result of question "Do you think the space time cube is suitable to visualize spatiotemporal data in this platform?"

For the fourth question "*What would you suggest to improve this visualization interface?*", some constructive feedbacks are listed in Table 5.5.

Table 5. 5. Constructive feedbacks from participants

Number of participants	Feedback
Five	It would be better to allow users to rotate the scene, which could solve the occlusion problem and make it easier to navigate.
Two	It would be better to implement adaptive design.
Two	The color scheme could change since the colors used to represent economic data and social data are quite similar.
One	2D visualization of time such as time line may be more suitable here.
One	It would be better to add more information in the popup such as city, detail description of data.

One	It would be clearer to add labels to different years.
One	In the city selection panel, it would be better if these cities are ordered alphabetically.

From the suggestions of those participants, several aspects could be improved. The first one is to give detailed instructions of how the interact with the 3D scene. The second improvement could be adaptive design. Additionally, changing the color scheme, giving more information in the popup, adding labels of time, arranging city order alphabetically in the city selection panel could be helpful.

As for the additional comments, only 8 participants gave their answers. Four of them gave positive responses, thinking this interface was quite interesting and useful. Two of them gave negative responses because of poor interactivities with 3D scene. One participant thought the number of datasets used here was not large enough. The last participant thought the base map needed to be improved.

5.3 Summary

This chapter illustrates the evaluation of the prototype proposed in Chapter 4 and discusses the result. Effectiveness and satisfaction are evaluated using benchmark tasks and user attitude investigation. The result of benchmark tasks is good except for the one concerned with spatial distribution, which means users could find desired datasets and have a good overview of visualized metadata but whether it is useful for exploring spatial distribution is not sure. The reason for the poor result of last task related to spatial distribution identification may be that users have trouble with understanding the content of data, which is reported frequently by users. As for the satisfaction of this prototype, the result is quite diverse. Participants have different opinions towards this visualization platform and the lack of experience with interaction with 3D scene before has a great negative impact on the users' attitude towards this visual interface.

6 Conclusion and future work

6.1 Conclusion

This research designed a visual interface for geodata query which combines the novel time visualization method, space time cube, and 3D symbols using different visual variables to represent different aggregated metadata. The utilization of space time cube and visualization of aggregated metadata give users good overview of time and other important information of all datasets, which is one of the innovations of this research. It visualizes three important metadata, namely data size, data category, and data quality. Especially, by giving the data quality information intuitively, the efficiency and accuracy in the query process can be improved. Then, the effectiveness and satisfaction of the visual interface are evaluated. The result of effectiveness evaluation shows that this visual interface could assist users in finding needed datasets and have a good overview of important information. As for the satisfaction of the visual interface, the result is quite diverse. Almost half participants had positive attitude towards it. While the others reported it was difficult to navigate through the 3D scene and understand the map.

6.2 Achievement of the Research Objective

The objective of this research is to develop a visualization method of aggregated metadata using space time cube to enable users to get a good overview and find needed datasets effectively. In order to achieve this overall objective, several sub-objectives are listed and fulfilled.

Data quality, an important metadata for data query and an innovation of this research, needs to be visualized. In order to fulfill this, the components of data quality are reviewed in Section 2.4. There are different components of data quality when assessing different datasets and in different data standards. Among those components, completeness, as one commonly used and easily calculated component, is selected to be visualized here to represent the data quality of datasets.

Together with completeness, data category and data size are also selected here. After deciding all metadata needed to be visualized, related works of visualizing semantic information of spatiotemporal data is reviewed and discussed in Section 2.1 and Section 2.2. Space time cube, an innovative visualization method of time is discussed in detail. Moreover, the classification, the development, the advantages, and problems of 3D visualization are reviewed. 3D visual variables, which are important in this research, are discussed and chosen to visualize selected metadata.

After finishing the literature review, the conceptual design is developed in Chapter 3. A concept framework for visualization of multivariable spatiotemporal data is adapted. Three components, namely data framework, user tasks, and visualization framework are elaborated.

Then a prototype using web-related technologies is implemented based on the conceptual design. Economic, geographic, social, and nighttime light data of Jiangsu province are used as test datasets. This process is elaborated and the website is built up in Chapter 4.

To evaluate the effectiveness and satisfaction of this proposed visualization method using space time cube, benchmark tasks and user attitude investigation were carried out using the online questionnaire in Chapter 5. Responses from 32 participants were collected and evaluated. Result is shown and discussed also in Chapter 5.

6.3 Answers of the Research Questions

As for the research questions, they are answered during this research.

- How to design the visualization method of metadata of spatiotemporal data using space time cube to enable users to get better overview and find needed datasets effectively?

a) Which visual variables should be chosen to visualize aggregated metadata and applied to 3D symbols in the whole space time cube?

As discussed in Chapter 3, size, color, and saturation of each symbol represent data size, data category, and completeness respectively. Size is a powerful visual variable in reflecting quantitative difference. Since it is easy for users to link the size of dataset to the size of the symbol that represents dataset, the size of the cube is used to visualize the size of dataset. Besides, color hue is applied to visualize data category. Considering the occlusion problem and the visual burden of the 3D scene, color saturation is the appropriate one to visualize completeness.

b) Which functionalities should be designed to enable users to get a good overview and find needed datasets effectively?

According to Chapter 3 and Chapter 5, interactivities with 3D scene play most important roles in 3D visualization. It can deal with some typical problems of 3D scene such as occlusion problem. User-friendly interactivities could leave positive impression on users. Additionally, instructions of how to interact with 3D scene are also important especially for users who are not familiar with it.

The second functionality is to show popups when hovering data cubes. Based on the evaluation result, information such as city information, time, data content, data size, data quality should be included in the popup. Among that information, description of each dataset is quite important since showing it in the popup is the easiest way for users to find out what this dataset about. At the same time, the corresponding city can be highlighted to help users link datasets and their corresponded cities.

The last set of functionalities is filter functions. Performing filter by category and city could assist users in finding datasets quickly. Additionally, filtering datasets by time so that users could have an overview of datasets in a specific year is also helpful.

- How is the usability of this proposed visualization method?

a) How is the effectiveness of this proposed visualization method using space time cube?

As discussed in section 5.2.1, users could find needed datasets and have a good overview of related metadata easily using this visualization platform. However, users cannot explore the spatial distribution of data. The reasons of that maybe that some users find it difficult to understand the content of the dataset and they are not familiar with interactivities with 3D scene.

b) How is the satisfaction of this proposed visualization method using space time cube?

The user experience of this proposed visualization method is quite diverse. Almost half of participants think it is quite interesting and intuitive using space time cube to display datasets, while some other participants think time slider could be a better way to visualize time information and 3D visualization makes it difficult to interact compared with traditional 2D maps.

6.4 Recommendations and future work

In summary, although the diversity of users' feedback, there are still many users think using space time cube and 3D visualization to visualize aggregated metadata and the time of datasets are quite interesting and intuitive. However, there are still some problems to be solved. The most commonly reported difficulties are occlusion problem and difficulty to find out the content of data directly from the data category visualization. Additionally, it is not clear whether 3D scene with space time cubes is more efficient than 2D map with time slider in geodata browsing.

Based on that, the recommendations for future works can be:

- Add more data content description in the popup so that user could have a rough impression on the data content when hovering data cubes, which may help explore spatial distribution.

- Add instructions of interaction with 3D scene right before users begin to use this visual interface. It is especially useful for the people who are not familiar with web-based 3D visualization platform.

- Evaluate and compare the efficiency of using space time cube to visualize datasets and using time slider to display datasets.

- Include more heterogenous datasets to test the usability.

- Using adaptive design to adept different screens and devices.

7 References

Bereta, K., Nikolaou, C., Karpathiotakis, M., Kyzirakos, K., & Koubarakis, M. (2013). SexTant: Visualizing Time-Evolving Linked Geospatial Data. International Semantic Web Conference (Posters & Demos), 177–180.

Bertin, J. (1983). Semiology of graphics; diagrams networks maps.

Bleisch, S., Nebiker, S., & others. (2008). Connected 2D and 3D visualizations for the interactive exploration of spatial information. XXI ISPRS Congress. Beijing, China.

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. Data Science Journal, 14.

Davies, T. (2010). Open data, democracy and public sector reform. A Look at Open Government Data Use from Data. Gov. Uk, 1–47.

de León, A., Wisniewki, F., Villazón-Terrazas, B., & Corcho, O. (2012). Map4rdf-faceted browser for geospatial datasets. Proceedings of the First Workshop on USING OPEN DATA. W3C.

Gebka, E., Crusoe, J., & Ahlin, K. (2020). Open data reuse and information needs satisfaction: A method to bridge the gap. EGOV-CeDEM-EPart 2020, 41.

Girres, J.-F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. Transactions in GIS, 14(4), 435–459.

Goodchild, M. F. (2007). Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. International Journal of Spatial Data Infrastructures Research, 2(2), 24–32.

Goodchild, M. F. (2013). The quality of big (geo) data. Dialogues in Human Geography, 3(3), 280–284.

Haeberling, C. (2004). A Cartographic Design Principle for 3D Maps. Retrieved July 12, 2013.

Haeberling, C., Bär, H., & Hurni, L. (2008). Proposed cartographic design principles for 3D maps: A contribution to an extended cartographic theory. Cartographica: The International Journal for Geographic Information and Geovisualization, 43(3), 175–188.

Hägerstraand, T. (1970). What about people in regional science? Papers in Regional Science, 24(1), 7–24.

Hashim, H., Wan Mohd, W., Sadek, E., & Dimyati, K. (2019). MODELING URBAN CRIME PATTERNS USING SPATIAL SPACE TIME AND REGRESSION ANALYSIS. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

International Organization for Standardization. (2015). ISO 9241-11: 1998—Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability.

Kloeckl, K., Senn, O., & Ratti, C. (2012). Enabling the real-time city: LIVE Singapore! Journal of Urban Technology, 19(2), 89–112.

Kraak, M.-J. (2003). Geovisualization illustrated. ISPRS Journal of Photogrammetry and Remote Sensing, 57(5–6), 390–399.

Kveladze, I. (2015). Space-time cube, design and usability [PhD Thesis]. University of Twente.

Leduc, T., Tourre, V., & Servières, M. (2019). The space-time cube as an effective way of representing and analysing the streetscape along a pedestrian route in an urban environment. SHS Web of Conferences, 64, 03005.

Li, X., & Kraak, M.-J. (2010). A temporal visualization concept: A new theoretical analytical approach for the visualization of multivariable spatio-temporal data. 2010 18th International Conference on Geoinformatics, 1–6.

Mennis, J. L., Peuquet, D. J., & Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. International Journal of Geographical Information Science, 14(6), 501–520.

Nikolaou, C., Dogani, K., Bereta, K., Garbis, G., Karpathiotakis, M., Kyzirakos, K., & Koubarakis, M. (2015). Sextant: Visualizing time-evolving linked geospatial data. Journal of Web Semantics, 35, 35–52.

Ormeling, F., & Kraak, M.-J. (2010). Cartography: Visualization of spatial data. Guilford Press.

Ontology Engineering Group. (n.d.). *Map4RDF.* http://certidatos.ign.es/map

Panchaud, N., Iosifescu-Enescu, I., Eichenberger, R., Sieber, R., & Hurni, L. (2012). Service-Driven 3D Atlas Cartography [PhD Thesis]. Master Thesis, ETH Zurich, Switzerland.

Pegg, D. (2012). Design issues with 3D maps and the need for 3D cartographic design principles.

Petrovic, D. (2003). Cartographic design in 3D maps. Civil Engineering, 10–16.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, 45(4), 211–218.

Ramalho, L. de F., & Segundo, W. R. (2020). R-Shiny as an Interface for Data Visualization and Data Analysis on the Brazilian Digital Library of Theses and Dissertations (BDTD). Publications, 8(2), 24.

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. Information Polity, 22(1), 1–24.

Semmo, A., Trapp, M., Jobst, M., & Döllner, J. (2015). Cartography-oriented design of 3D geospatial information visualization–overview and techniques. The Cartographic Journal, 52(2), 95–106.

Sieber, R., Serebryakova, M., Schnürer, R., & Hurni, L. (2016). Atlas of Switzerland goes online and 3D—Concept, architecture and visualization methods. In Progress in Cartography (pp. 171–184). Springer.

Stadler, C., Martin, M., & Auer, S. (2014). Exploring the web of spatial data with facete. Proceedings of the 23rd International Conference on World Wide Web, 175–178.

Thakur, S., & Hanson, A. J. (2010). A 3D visualization of multiple time series on maps. 2010 14th International Conference Information Visualisation, 336–343.

Turchenko, M. (n.d.). Space-Time Cube Visualization in a Mixed Reality Environment.

Valsecchi, F., & Ronchetti, M. (2014). Spacetime: A two dimensions search and visualisation engine based on linked data. Conference on Advances in Semantic Processing (SEMAPRO).

van Elzakker, C. P., & Wealands, K. (2007). Use and users of multimedia cartography. In Multimedia cartography (pp. 487–504). Springer.

Veregin, H. (1999). Data quality parameters. Geographical Information Systems, 1, 177–189.

Wagner Filho, J. A., Stuerzlinger, W., & Nedel, L. (2019). Evaluating an immersive space-time cube geovisualization for intuitive trajectory data exploration. IEEE Transactions on Visualization and Computer Graphics, 26(1), 514–524.

Wang, D., Szymanski, B. K., Abdelzaher, T., Ji, H., & Kaplan, L. (2019). The age of social sensing. Computer, 52(1), 36–45.

Xiao, F., Wang, Z., & He, D. (2020). Understanding users' accessing behaviors to local Open Government Data via transaction log analysis. Proceedings of the Association for Information Science and Technology, 57(1), e278.

Appendix 1: Questionnaire

A Graphic Visual Interface for Geodata Query with Space-time Cube

We are interested in evaluating the usability of the visualization platform for geodata query using space time cube. For this study, you will be presented with the prototype. Then, you will be asked to answer some questions about it. This is an anonymous questionnaire and your responses will only be used for this evaluation.

This study should take you around 10 minutes to complete. Your participation in this research is voluntary. You have the right to withdraw at any point during the study. The principal investigator of this study can be contacted at zbxyz567@outlook.com.

Please finish this questionnaire with your computer.

By clicking the button below, you acknowledge:

- Your participation in the study is voluntary.
- Your information would be gathered only for the evaluation of this visualization platform.
- You are aware that you may choose to terminate your participation at any time for any reason.

o I consent, begin the study

o I do not consent, I do not wish to participate

Start of Block: User Information

The following questions would gather some basic information about yourself. Your data would be only used for evaluating this platform.

Q1 What is your gender?

o Male

o Female

o Non-binary / third gender

o Prefer not to say

Q2 How old are you?

o <20

o 20-30

o 30-40

o >40

Q3 How often do you use data portals?

o Never

o Sometimes

o Usually

o Often

o always

Q4 Have you worked with web-based 3D visualization before, like Google Earth?

o Never

o Several times

o A lot

End of Block: User Information

Start of Block: Benchmark Tasks

To determine whether the following statements are correct or not, checking this website: https://unruffled-dijkstra-9403b9.netlify.app

Please open this website only with your computer.

You may need to read the introduction and explore the website on your own first.

Q1 In 2009, the size of economic data in Nanjing is 13kb.

o True

o False

o Not sure

Q2 Nighttime light data from 2012 to 2020 in Wuxi is available.

o True

o False

o Not sure

Q3 There are two kinds of data available in Lianyungang.

o True

o False

o Not sure

Q4 The completeness of nighttime light data in Suzhou in 2019 is better than that in 2017.

o True

o False

o Not sure

Q5 Institutions in the southern part of Jiangsu province published more paper than that in the northern part.

o True

o False

- o Not sure

Q6 Do you have some interesting findings?

____

End of Block: Benchmark Tasks

Start of Block: User Feedback

Your feedback is really important!

Q1 How do you rank the overall ease of usability of this visualization interface?

- o Difficult
- o Somewhat difficult
- o Neutral
- o Somewhat easy
- o Easy

Q2 Do you think the map elements are sufficient enough to assist you in fulfilling all tasks? if not, what else would you like to add?

- o yes
- o not enough ____

Q3 Do you think the space time cube is suitable to visualize spatiotemporal data in this platform? if not, why?

- o Very suitable
- o Somewhat suitable
- o Neutral
- o Not suitable ____
- o Not suitable at all ____

Q4 What would you suggest to improve this visualization interface?

Q5 Additional comments

End of Block: User Feedback