Cartography M.Sc.

Master thesis

3D interactive spatio-temporal and semantic visualization for discrete data associated with multiple categories and long historical events

Xiyu Wang

Technical University of Munich

TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology

TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITY OF TWENTE.

17.09.2021

3D interactive spatio-temporal and semantic visualization for discrete data associated with multiplecategories and long historical events

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

Author:	Xiyu Wang
Study course:	Cartography M.Sc.
Supervisor:	M.Sc. Chenyu Zuo
Reviewer:	Prof. Dr. Menno-Jan Kraak

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

Date of submission: 17.09.2021

Statement of Authorship

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

“3D interactive spatio-temporal and semantic visualization for discrete data associated with multiple categories and long historical events”

I have fully referenced the ideas and work of others, whether published or unpublished.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

Munich, 17.09.2021 Xiyu, Wang

Acknowledgement

The closure of my thesis marks the end of my Masters study, but there are so many precious and joyful memories in the past four semesters that I struggle to put together with just a few words.

Firstly, I am sincerely grateful for being accepted into the prestigious, international Erasmus Mundus Cartography M.Sc. Program (here after referred to as “the Program”), the only Erasmus Mundus Joint Master Program so far that primarily focuses on the study of cartography. The moment I received the admission letter started everything joyful and unforgettable.

Secondly, I owe an earnest thankfulness to M.Sc. Chenyu Zuo and Prof. Menno-Jan Kraak for being constantly patient and supportive when I faced considerable thought transitions during my thesis research. In fact, in the thesis topic selection phase, I saw "Can We See the Uncertainties in Calculation" under Chenyu’s supervision and not edit down as an option. This topic, from my perspective at that time, is quite statistics-based and hence more relevant to my undergraduate academic background. However, I ended up receiving my current topic under Chenyu’s supervision, and Prof. Menno-Jan Kraak was assigned as my reviewer. Once I kicked off my thesis research, Chenyu and Prof. Menno-Jan Kraak helped me stream my “broad” research identification into a specific research direction, enlightened me with the meaning of semantic distribution, encouraged me to explore cutting-edge spatio-temporal visualization prototypes, reminded me of seeing beyond the abilities of commercial GIS platforms and in the end, broadened my thoughts on future thesis research. Speaking of "semantic distribution", I am earnestly grateful to Ms. Madeleine Stapel from the TU Munich English Writing Center for discussing its philosophical domain with me, as well as offering constructive feedback on the word choice and sentence structure of my thesis. Anyway, my entire thesis research and writing process prove that I still have a lot to learn and to improve.

Thirdly, I deeply thank all kind hearted and dedicated coordinators, head professors and other teachers of the Program. In the first semester at TU Munich, M.Sc. Juliane Cron, being an informative and attentive Head Coordinator, crafted her own academic and career experiences into a series of heartwarming “mommy’s farewell words” during the very last lecture. During the thesis semester, not only did Juliane help to host master colloquiums, she also offered constructive answer to my academic question: "Why is it necessary to have an exclusive ‘Discussion’ chapter to answer the research questions, if the ‘Results’ chapter already shows some answers?" (Remark: she answered it together with Dr. Christian Murphy). Regarding the semester at TU Vienna, Prof. Georg Gartner fully understood my intention of supporting the academic atmosphere in class during the pandemic, and consoled every student by addressing the significance of living well in uncertainty. As for the semester at TU Dresden, Prof. Dirk Burghardt and Dr. Nikolas Prechtel persistently supported me in the first 4-5 days of the Alpine Field School, when I was physically down for the long excursions. M.Sc. Benjamin Schroeter’s wise words about dealing with issues in a large student group also reminded me of enhancing my “soft power” apart from academic capabilities. Moreover, Prof. Liqiu Meng, the Head Professor of the Program understood my persistent interest in the Program’s academic activities. For instance, I attended all students’ assignment or thesis-related presentations, joined all on line and on-

site meetings, and even read through all available thesis topics carefully. I was also keen on writing an academic abstract with one or more classmates for the 30th International Cartographic Conference (ICC) in parallel with my thesis writing. When discussing my potential contribution to the ICC with Prof. Liqiu Meng, considering my aforementioned interest, she modified the abstract topic as "how the Program thrived during the pandemic from a student's perspective" and set the targeting theme (to which the abstract is submitting) as "Education and Continuous Learning in Cartography". When writing the abstract, she and Juliane offered detailed feedback. Prof. Georg Gartner and Prof. Dirk Burghardt also encouraged me a great deal.

Fourthly, I find the first group picture that I had with my classmates on the TU Munich main campus very precious. Since the Covid-19 outbreak, the chances for everyone to meet up in the physical world have been strictly limited, which made me sincerely grateful to all valuable communications I had with each individual student. For example, Ozan Tuncbilek and I formed long-term academic cooperation since the Welcoming Session of the Program. In the thesis semester, Anouska Jasperson, Jakob Listabarth and Larissa Saad all answered my technical questions at least once. Ben McLeod raised an elaborated question in my midterm presentation rehearsal. Furthermore, Sifa Mawiyoo shared a video "Stories of Space, Time and Data" with my classmates and me last April, and it inspired my thesis research purpose. He also offered me his thesis poster and briefly summarized its creation for my reference. I also extended my academic discussions in addition to my classmates. As such, I heartily thank Jonathan King from the next intake for checking the English writing of my thesis, editing my ICC abstract and becoming my co-author (I always see us as mutual first authors of the abstract). We also frequently exchanged our ideas on the essence of cartography and our interpretation of the Program's education philosophy. Meanwhile, I deeply thank M.Sc. Soubarna Banik, my flat mate and a PhD candidate at the Department of Mathematics and Informatics, TU Munich for helping me with Latex document formatting. It made my first experience with Latex easier and more enjoyable.

Finally, I should thank my parents. Valuing my true growth more than academic achievements, they always encouraged me to support the teachers and help my classmates. They also never restricted my life behind my computer screen, as long as I follow the updated Covid-19 regulations. Besides ensuring my fruitful communication with others surround me, I occasionally introduced my new cultural experiences to them, such as some intriguing Russian customs. I also talked about the fascinating plotlines in Harry Potter, my favorite novel series to them, to demonstrate my gained knowledge of storytelling techniques. They in turn recommended The Ideal City, a popular Chinese TV series themed on "rules", "culture", "teamwork" and "young people who stay true to their aspiration", to name a few. Moreover, they supported me, a "water baby", on taking an intensive mountain climbing course, so that I can eventually visit Mount Ushba and explore Samatvro, Georgia, an archeological site where the senior lecturer of my undergraduate study Mr. Clifford Ogleby conducted cartographic surveying projects. (Remark: The Mount Ushba Mapping Campaign this year was announced as "everybody welcomed to attend" by Mariam Gambashidze, a student representative of my intake in the first and semesters.)

Abstract

There exists a large number of discrete, human-created, geo-related features under multiple categories and associated with long historical events around the world. The spatio-temporal distribution and semantic distribution of these features is worthy of renewed discussions while remaining a big challenge for cartographers and geodata analysts. Nevertheless, there are still some cutting-edge techniques for visualizing spatio-temporal distributions of discrete geodata, such as space-time cubes (STCs) for temporal geostatistics (ESRI, 2021). This thesis proposes three visualization prototypes that are inspired by STCs, but are capable of representing different categories of geographic data.

The core prototype is "space-time micro landscapes". To create it, the geographic surface with geo-related features distributed is tessellated via equal-sized grids. Afterwards, above each tessellated surface, space-time bins are created with aggregated 3D symbols employed to indicate both the numbers and categories of the geo-related features inside.

The research emphasizes that 3D geodata visualization prototypes such as the proposed ones give a simple and straightforward illustration of typical spatio-temporal and semantic patterns. Furthermore, since geovisual storytelling is powerful for gleaning discrete analysis results (maps, diagrams and other visual representations), as well as revealing typical spatio-temporal and semantic patterns of geodata, it is applied to represent the visual outcomes of the designed prototypes. The research also indicates that a multimedia story map with interactive parts is effective in conveying significant information from various perspectives.

The case study applied in the thesis' research is the spatio-temporal and semantic distribution of 14510 companies in Shanghai, China. Other cases to which the designed prototypes are applicable include, but are not limited to the spread of Covid-19 along with its variants across western Europe, and the evolution of cartography-related research areas around the world from the perspective of academic conferences.

Keywords: geovisualization, spatio-temporal distribution, semantic distribution, multiple categories, historical events, STC, spatial point aggregation, 3D symbol design, geovisual storytelling

Contents

List of Figures

List of Tables

List of Abbreviations

CBD	Central Business District
Covid-19	Corona Virus Disease 2019
CPC	the Communist Party of China
DSM	Digital Surface Model
GIS	Geographic Information System
GDP	Gross Domestic Product
HFRS	Hemorrhagic Fever with Renal Syndrome
KDE	Kernel Density Estimation
ICOMOS	International Council on Monuments and Sites
KMT	Kuo-Min Tang
Org_types	Organization types
Org_start_time	Organization start time (years of registration)
Firm_info	Firm’s Information
POI	Places of Interest
PRC	People’s Republic of China
STC	Space-Time Cube
UNHCR	United Nations Higher Commissioners of Refugees
USA	United States of America
VR	Virtual Reality
WGS84	World Geodetic System 1984
WWI	World War I
WWII	World War II

1 Introduction

1.1 Motivation and problem statement

The world is filled with tangible and intangible human-created, geo-related features, such as companies, schools, administrative complexes and theaters. Their ever-changing spatio-temporal distributions often reveal complex human thoughts and behaviors. These features are often categorized in detail by human beings, which evolve constantly and become essential semantic attributes, i.e., aspects that contribute to the “meaning” of the corresponding features (Google Arts and Culture, 2021). For instance, types, products and services are essential semantic attributes that determine what a group of companies registered in a city are. Throughout urban history, the number of newly registered companies belonging to various types tends to change based on significant events. The types and qualities of the products or the services companies can offer also evolve remarkably. These evolutions give people a better understanding of the size and complexity of real-world problems, and thus to “analyze, envision, reason and deliberate” (Andreinko et. al, 2006). As such, spatio-temporal and semantic distribution of geo-related features is worthy of renewed discussion by cartographers and geodata analysts.

Thanks to the advanced research of computer science, data analysis, information visualization, scientific visualization and other relevant disciplines, there has been a great deal of research on spatial-temporal visualization, particularly on pattern recognition (Andrienko, 2011). However, according to the academic literature, there is a lack of novel and minimalistic prototypes designed for visualizing discrete geo-related features that are created under multiple categories and are associated with long historical events. Here, “long historical events” refers to an event that lasts for months or years. For instance, the War against Japanese Aggression lasted from 1937 to 1945 in China. It also means an instant event that leaves a long-term effect since it takes place. For example, the foundation of the People’s Republic of China (PRC) was announced during the Founding Ceremony on 10.01.1949, but it had a decade-long influence on the economic development of Shanghai. The most advanced visual representation discovered that demonstrates spatio-temporal and semantic patterns of geo-related features together is the space time cube (STC) for temporal geostatistics (ESRI, 2021). It has been consistently applied in the analyses of increasing criminal cases in a town (Hashim, et. al., 2019), infected cases of a contagious virus across a country (Mo et. al., 2020), etc. Nonetheless, it is not yet capable of representing different types of criminal cases or variants of a virus at once. This absence of information therefore motivates this thesis’ research.

A research gap suggested by the literature investigation for this thesis is a precise definition of the term “semantic distribution”. In fact, various geoscientists conducted research that can be identified as “semantic distribution” without explicitly referring to this term. For instance, Ding, Yang and Meng (2015) semantically classified the destinations of taxi passengers reached from Shanghai Pudong International Airport in 47 days into “public”, “commercial”, “residential” and “industrial”. Afterwards, they designed a graphic user interface (GUI) to maximize the freedom of exploring the spatio-temporal and semantic information. In this GUI, the destinations are distributed in clusters, and colored in blue, yellow, green and magenta respectively, based on their semantic attributes. When a user switches from one timeslot to another, a different distribution of the destinations colored according to their semantic attributes will appear. Throughout the research, the term “semantic distribution” was not addressed, but a clear sign of it is demonstrated in the visualization outcome. As such, to extrapolate from the research conducted by Ding, Yang and Meng (2015), the term “semantic distribution” is defined for this thesis as “the spatio-temporal distribution of geo-related features’ semantic attributes”.

1.2 Research objectives

The primary focus of this thesis’ research is the intuitive and convincing visualization of the spatio-temporal and semantic distribution of geo-related features associated with multiple categories and long historical events. Furthermore, many researchers have concerned the storytelling potential of geodata visualization (Segel & Heer, 2010) throughout time. The old saying “A picture paints 1000 words” also holds on geovisual representations, and in recent years, more informative “data stories” are increasingly discovered by journalists, data scientists and cartographers (Segel & Heer, 2010) thanks to the complexity of social development. Geovisual storytelling, or story mapping focuses on "crafting continuous stories from discrete analysis results (maps, diagrams or other visual representations) to process unique attributes" (Segel and Heer, 2010). Hence the aim of this thesis’s research can be specified as how to visualize spatio-temporal and semantic distributions of discrete, multi-categorized geo-related features and display the outcomes via geovisual storytelling.

In order to achieve the aim, there are three underlying research objectives:

1. Propose novel proto types to visualize spatio-temporal and semantic distributions of a geodata set over various time periods;
2. Construct a “plotline” that crafts typical geospatial land semantic patterns of geodata into a story map;
3. Create a multimedia story map with interactivity to represent spatio-temporal and semantic patterns.

1.3 Research questions

There are four research questions associated with the research objectives:

1. What geovisualization techniques exist for investigating spatio-temporal distributions and/or semantic distribution of geographic features? How well do they perform?
2. How can novel spatio-temporal and semantic visualization prototypes for discrete,multi categorized, geo-related features be designed?
3. How can narrative visual storytelling narratives for representing spatio-temporal and semantic distributions be constructed?
4. Are there typical geospatial patterns and/or semantic patterns for geo-related features, in certain time periods?

1.4 Thesis structure

This thesis is structured into six chapters. Chapter one introduces the thesis topic, offers related research objectives and questions. Chapter two lists classic concepts and advanced techniques related to spatio-temporal distribution and semantic distribution, which serves as a foundation of further research and experiments. Geovisual storytelling is discussed in a separate section of this chapter. Chapter three, based on the efficiency and novelty of the techniques discussed in chapter two, proposes three spatio-temporal visualization prototypes. Meanwhile, a plotline for story mapping that addresses useful genres and tools for the story map is constructed. Chapter four introduces a case study to test the prototypes and demonstrates the results via three geospatial stories. Chapter five evaluates the story map and the prototypes. It also discusses important findings and addresses research limitations. Chapter six draws conclusions of the thesis, answers the research questions in chapter one and suggests further research ideas for spatio-temporal and semantic distribution.

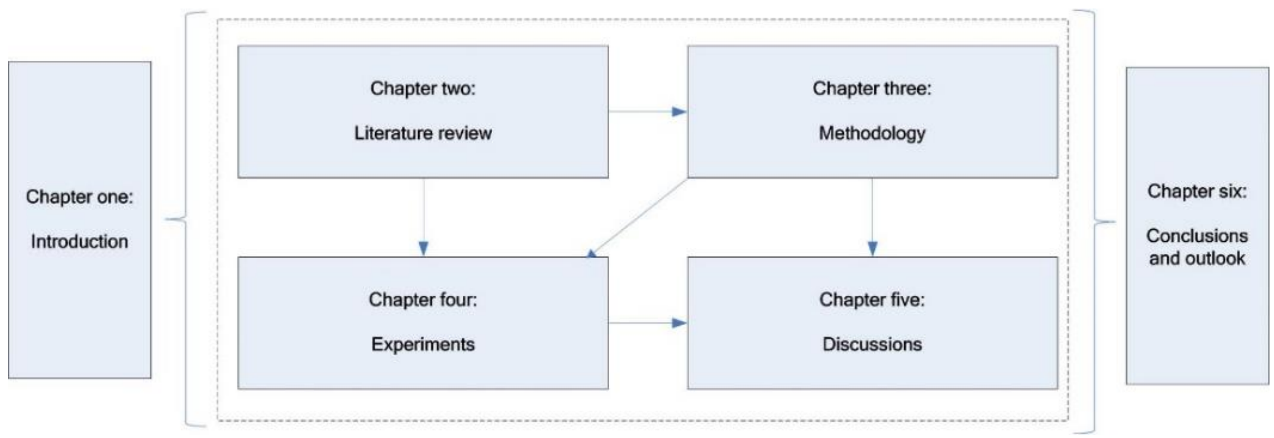


Figure 1.1 The diagram of thesis structure

2 Literature Review

This chapter summarizes theoretical backgrounds and existing techniques that are essential for the research of this thesis via a detailed literature review of conference papers, journal articles and webpages.

2.1 Geovisualization

Geovisualization is a loosely defined domain that addresses the visual exploration, analysis, synthesis and presentation of geospatial data (Andreinko et. al., 2008). Nowadays, geovisualization is largely involved inhuman movement, commodity flow, urban development and other problems that are related to geographical space and temporal changes or processes (Andrienko, 2011). Aiming at creating convincing visual representations such as thematic maps, time series and dashboards, geovisualization is often applied to assist analytical reasoning, a research domain of data exploration and in sight retrieval (Andrienko, 2011).

However, besides continuous geodata such as trajectories and commodity flow, there also exist discrete geodata created under multiple categories and during long historical events, such as companies, educational institutions and administrative complexes. The categories and historical events, as semantic attributes (the types, products and services of companies, as mentioned in chapter one, for example), increasingly make geodata multi-dimensional and heterogeneous, and hence persistently challenge geoscientists for precise visualization, in order to make people learn more efficiently from the past and make to-the-point spatial decisions (Andrienko, 2006).

Geovisualization did not become a research area until the early 1980s, thanks to Jacques Bertin, a French graphic theorist whose expertise is in cartographic design and information visualization (O'Sullivan & Unwin, 2010b). However, long before the 1980s, geoscientists had already begun to pursue the most intuitive and insightful representations of spatio-temporal and semantic distribution for geo-related features. A typical proof is the historic question raised by Jacques Barbeau Dubourg, a world-famous French physician, write, publisher and chronographer in the 1760s (Ferguson, 1991).

a. Temporal visualization **b.** The two Eyes of History

Figure 2.1 Jacques Barbeau Dubourg's historical question (Ferguson 1991, pictures from Info We Trust, 2021)

From Dubourg's perspective, cartography (or geography) and chronology are two eyes of history. One eye sees the physical world and reflects visible phenomena on a map, while the other sees time and reflects undergoing changes in data analysis (Ferguson, 1991). The real cooperation between these two eyes is vital for forming people's thorough perception of the world. As such, this thesis attempts to make the "eye of chronology" see clearer and reflect what it sees more effectively.

The following sections introduce some existing spatio-temporal visualization techniques, and briefly evaluate their performances.

2.2 Spatio-temporal visualization

2.2.1 2D and 3D mapping

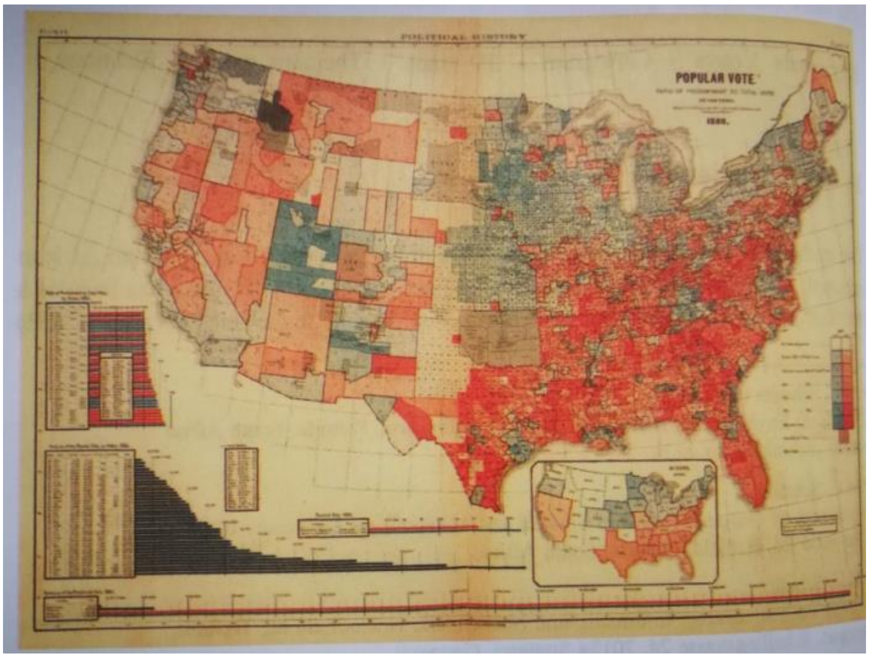
2.2.1.1 2D thematic mapping

2D thematic mapping has been frequently applied in representing multiple semantic attributes of geodata sets. When necessary, varied mapping styles can be combined in order to thoroughly convey the main information given by a map maker.

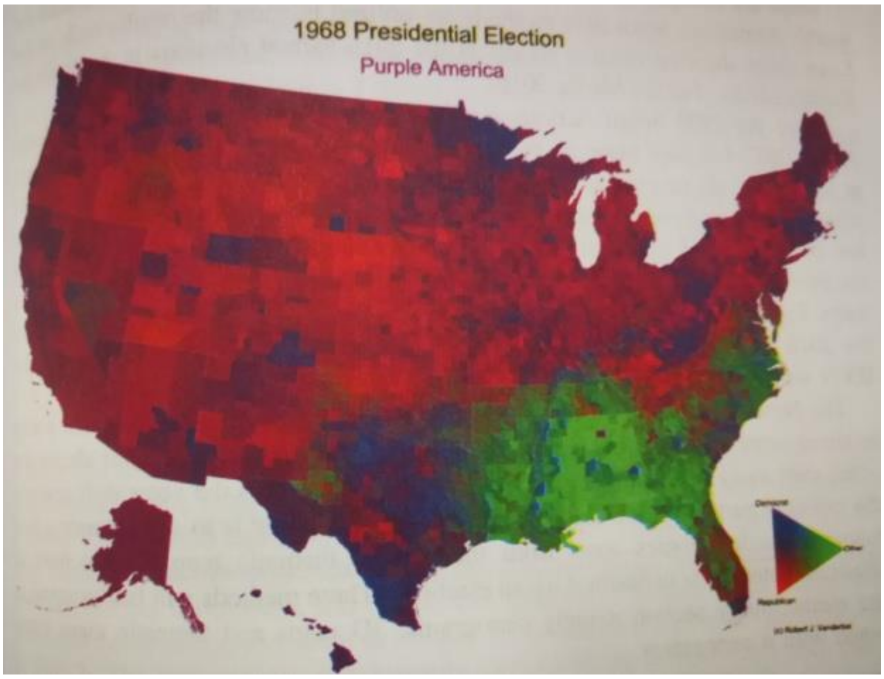
One frequently depicted topic for thematic mapping is political voting results, whose common semantic attributes are voters' preferred parties and expected candidates. The first map in history (**Figure 2.2**a) (Schmitz, 2017) that showed the county-level results of the US presidential elections represented counties with a blue or red hue, depending on whether a county saw more than 50% of its votes go to a Republican or a Democrat candidate. It also used different shades of hues to indicate percentage of votes per county.

To include voting results of two or more parties and improve map readability, other scholars proposed supplementary mapping techniques. Vanderbei (cited in Schmitz, 2017) implemented a color hue triangle to map US presidential voting patternsin1968. In the color hue triangle, red stands for the Republicans, blue for the Democrats and green for other minor parties in the USA. If the Republican and the Democrats each gain 50 percent of the votes in a county, the corresponding polygon of the county will appear in purple (**Figure 2.2**b). Vanderbei also introduced gradient fill of polygons, a technique for indicating the number of votes obtained in a county by proportionally shrinking the size of its corresponding polygon. In this way, two visual variables, color hue and shape, are harmonized to give map readers multiple information at once.

Dasymetric dot density mapping, a more up-to-date mapping technique uses color tones to represent voters' preferred parties (or candidates) and dots to represent individual votes. In other words, it ensures aggregation and individuality together in one. It refines a normal dot density map with ancillary information about the distribution (county boundaries, for example) of the variable being visualized. The 2016 US Presidential Election Map published by ESRI (2016) is a wonderful example (**Figure 2.3**). In the map, two color hues, navy blue and light pink, represent Hillary Clinton and Donald Trump respectively. Additionally, one dot represents one individual vote from the 130 million total votes across the US. However, instead of following the traditional dot density mapping technique and making the dots fill the corresponding collection counties uniformly, dasymetric dot density mapping aggregates them into more representative areas: the ones with residents. The aggregated result distinguishes residential areas and non-residential areas clearly. In addition, dots are set to 90% transparent to be prevented from occluding each other. To summarize, the map indicates the overall voting tendency concisely, and intuitively proves a common saying: every vote counts.



a. The1883 presidential election



b. The 1968 presidential election

Figure 2.2 Presidential Voting Map Origins (Schmitz, 2017)

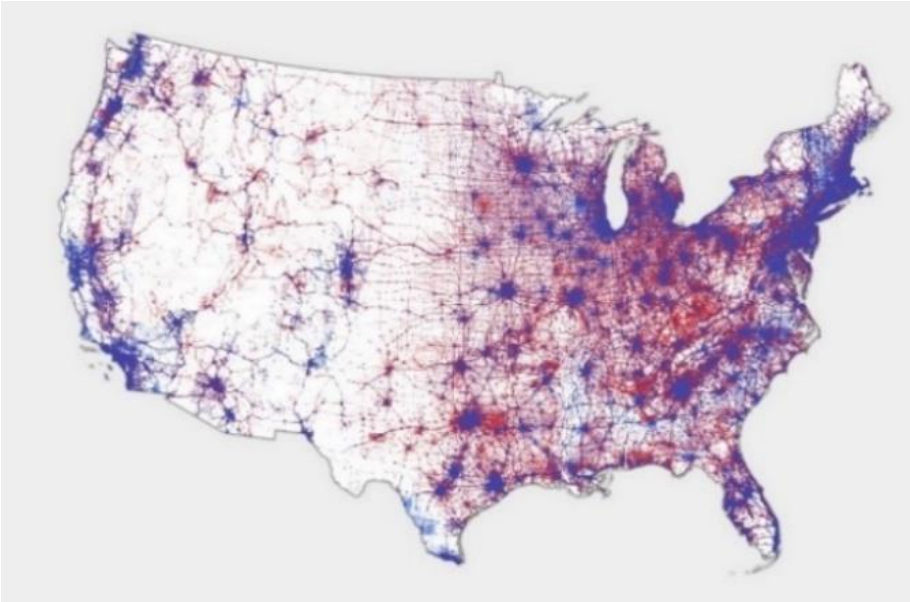
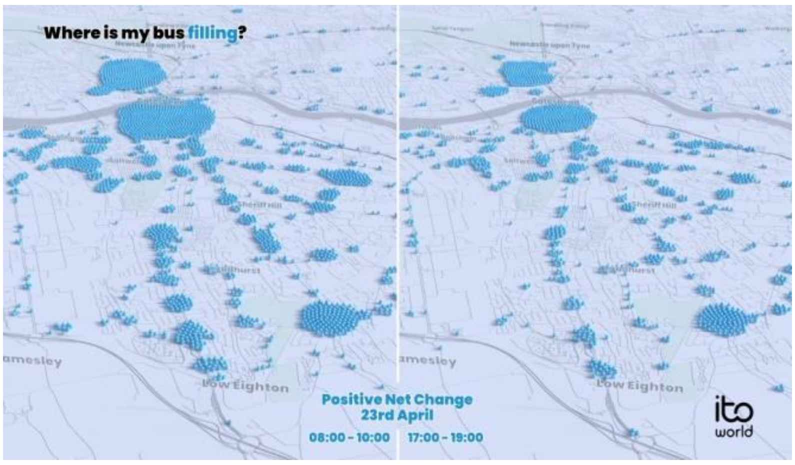


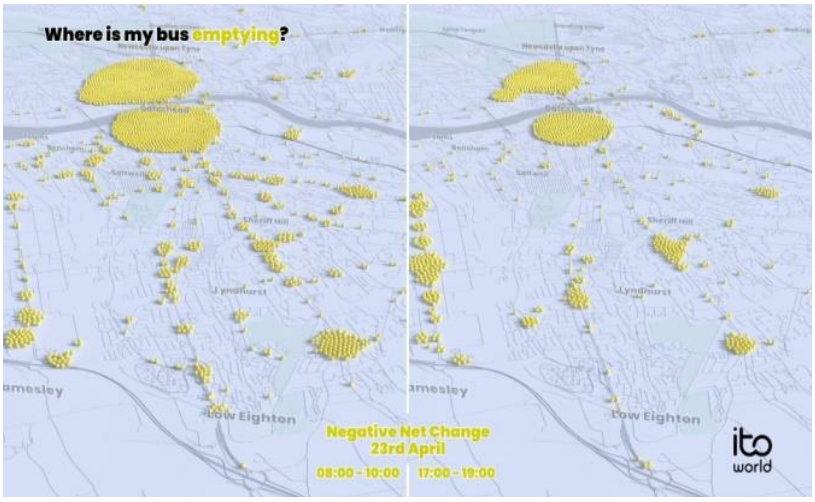
Figure 2.3 The 2016 presidential election (ESRI, 2016)

2.2.1.2 3D symbols in thematic maps

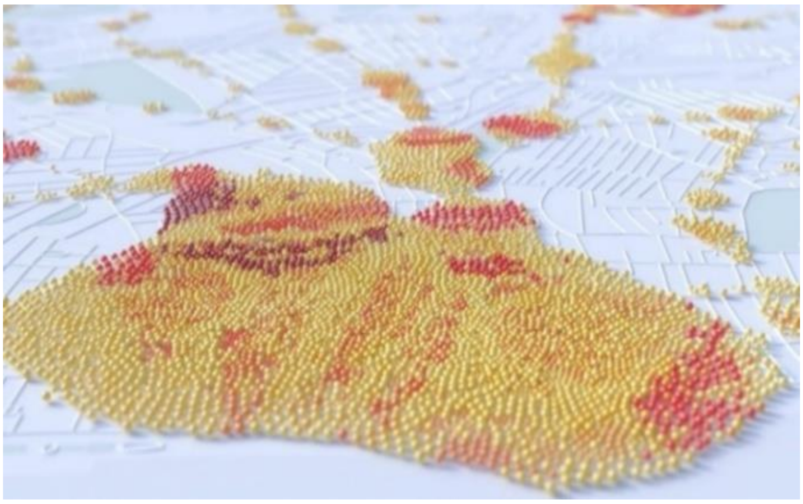
Apart from 2D mapping techniques, such as the aforementioned dasymetric dot density mapping, 3D symbol mapping, an innovative and effective technique has eventually caught people's attention. Craig Taylor (2021a), the ITO World Data Visualization Manager made the most of his expertise in 3D spatial analysis, informatics and geography to create new modes of communicating data. One of his projects uses marble-like symbols to depict individual passengers getting on and off buses in London. Blue and yellow represent passengers "getting on" and "getting off" respectively, and the bigger symbols in darker colors represent net capacity change. In other words, the more the passengers get on or off at a station, the bigger the corresponding symbols would become, and the darker their colors would be. As was the case of the aforementioned dasymetric dot density mapping, point occlusion is effectively avoided in Taylor's projects. However, each individual symbol looks much clearer in Taylor's projects, whereas in the dasymetric dot density map, the dots still appear being stuck together unless the map is zoomed in. It is also worth mentioning that aggregation is applied to demonstrate net capacity change by the size and color of the symbols.



a.



b.

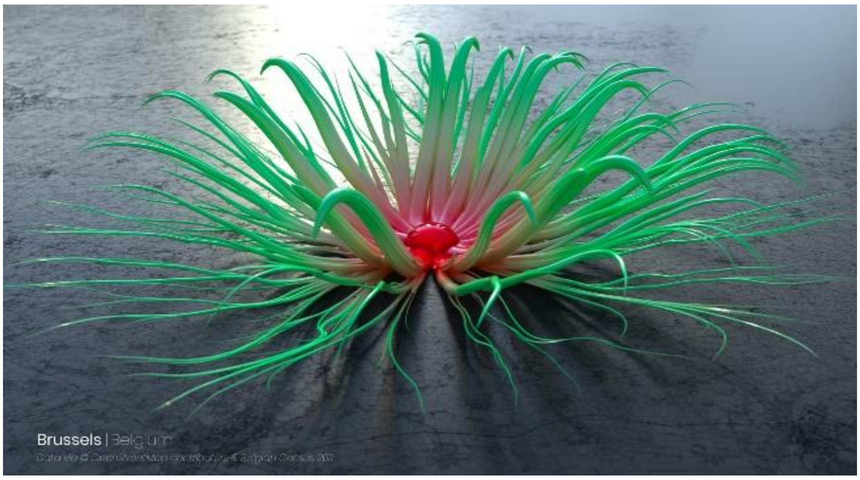


c.

Figure 2.4 Taylor's "Passengers" Design (Taylor, 2021a (a) getting on, (b) getting off, (c) net capacity change)

Commuter gardens is another Taylor's project (2021b) that incorporates 3D symbology. It includes 2D basemaps of major European cities with colorful, flower-like 3D symbols representing transportation networks attached to. The "flowers' stamen bundles" (centers) are placed onto important public transportation hubs of the corresponding cities. Their "petals" represent major roads or streets: they point towards their surrounding hubs, and their thickness indicate average number of daily commuters. In general, the flowers are not only eye-refreshing, but ideally avoid the overplotting issue of transportation networks and commuters' trajectories.

a. England’s commuter garden



b. A single “data flower”

Figure 2.5 Taylor’s "Data Flower" Design (Taylor, 2021b)

In fact, it is already well-applied in 2D thematic mapping to use aggregated symbols representing different categories. One typical example is pie chart mapping. Ding. et. al. (2016) used proportional pie chart symbols with a grey scale color scheme to depict the number of pick-up and drop-off events detected from the floating car data collected between Shanghai Pudong International Airport and various transportation hubs in Shanghai during different hours of a day. In addition, Choi (2019) visualized geo-temporal floating population data of tourists and residents of Jiju Island, South Korea via semi-transparent circles divided into arcs representing each month of a year. Even for non-geographic visualizations such as the proportion of scientific papers under each topic of a conference and the frequency of scientific papers from the same author being cited (Chen et. al., 2020), circular 2D symbols with colorized arcs or segments are implemented for visualization. However, 3D symbols have a definite advantage in its higher capacity and more outstanding visual effect in illustrating geo-related features and events (Choi, 2019).

2.2.2 Spatio-temporal visualization

2.2.2.1 Flow maps and space-time cubes

Flow maps represent linear movement between places over time. Arrows or lines with different thickness are used to represent the type, quantity and direction of movements. They can also be branched to indicate a group's dispersion. Throughout history, there have been many classic examples of flow maps, and one classic map that embeds thorough spatio-temporal information is "Figurative Map of the successive losses of men in the French army during the Russian Campaign 1812-1813" (in French: Carte figurative des pertes successives en hommes de l'Armée Francais dans la champagne de Russe 1812-1813) drawn by Charles-Joseph Minard. It incorporates colored branched arrows, geographic symbols, a timeline, a temperature curve and detailed annotations to illustrate how the army of Napoleon's campaign which, at the boarder to Russia at the beginning of June, 1812, comprised of 422000 soldiers across Europe, suffered great defeats and at the end of the campaign contained only 100000 soldiers. The march to Moscow is drawn in beige and the return in black. The annotations explain clearly the relationship between the soldier casualty and the inhospitable cold weather the army encountered in places it traversed during the campaign (Kveladze, Kraak & van Elzakker, 2017).

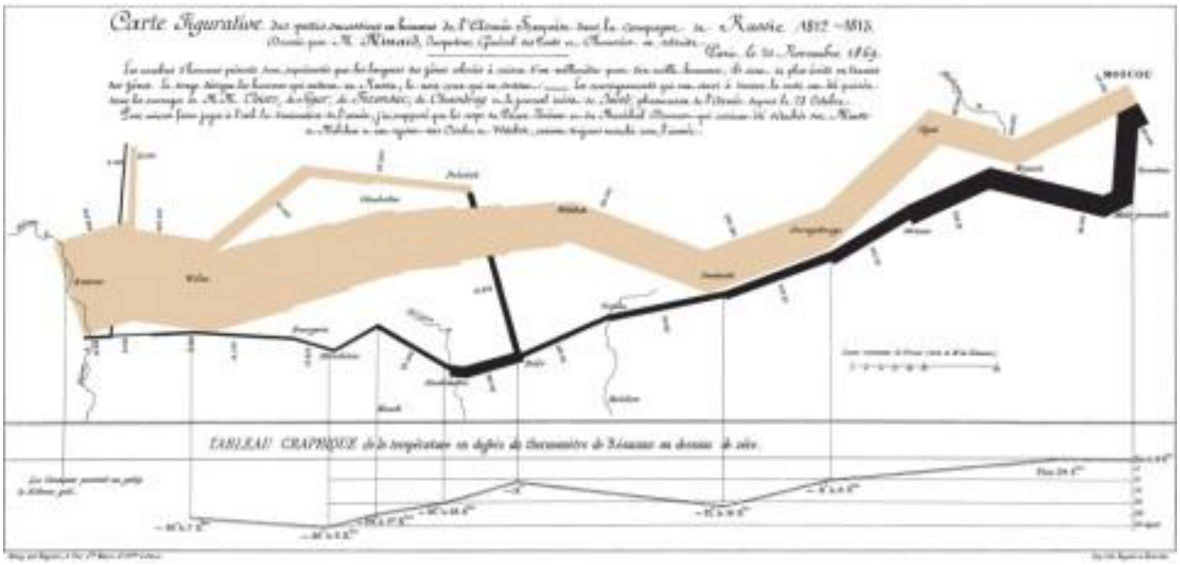


Figure 2.6 Carte figurative des pertes successives en hommes de l'Armée Francais dans la campagne de Russe1812-1813(Kveladze, Kraak & van Elzakker, 2017, picture from Brinch, 2019)

This flow map beautifully presents information about temporal change in a static form. It has inspired many other visualizations that attempt to effectively convey spatio-temporal variation. One such prominent successive visualization, as pointed out in chapter one, is space-time cube (STC), a three-dimensional, time-geography (a research area dealing with temporal variation in geographic events) visual representation come up by Haegerstrand at the end of 1960s (cited in Kraak, 2003). Before the STC become frequently applied, its potential benefits for analyzing complex spatio-temporal patterns were already discussed in academic literature such as Kristensson et. al. (2008). During the user test addressed in Kristensson et. al. (2008), 30 participants without previous experience of information visualization tools found the STC "cool", "intuitive" and "easy to understand". Eventually, the advancement of computer-based systems let it attract much more attention from geodata analysts (Filho et al., 2019). Generally speaking, the STC consists of a cube with a base representing geography (along the x- and y-axis) and the height representing time (the z-axis). Courtesy of computer-based systems, it is enabled to be viewed and interacted as a user's wish for insight retrieval (Kraak, 2003).

The STC is typically applied in two types of study. The first one is movement analysis, including individual movement, transport optimization, sports analysis and behavioral studies (Filho et al., 2019). In the STC for movement analysis, all nodes in one trajectory are illustrated three-dimensionally with the z-value as a timestamp, and the entire trajectory is referred as an "space-time path" (Kraak, 2003). In Kveladze, Kraak and van Elzakker (2017), Carte figurative des pertes successive en hommes de l'Armée Francais dans la champagne de Russe 1812-1813 is converted into an STC (**Figure 2.7**), in order to display more crucial information about the military campaign. Each branched trajectory of Napoleon's corp is represented along the temporal axis and is individually colored. The casualty that he corp experienced on each trajectory is represented by line thickness. A base map is attached at the bottom of the STC, offering a clear geographic context to the visualization. Comparing to the original 2D map, the STC indeed allows more information to be displayed, including movement duration and speeds (non-vertical lines' lengths and slopes), stop duration, and locations (vertical lines). In addition, during the usability test of the STC, most attendants found it easy to target time-related questions, thanks to the flexibility and interactivity of this visual representation.

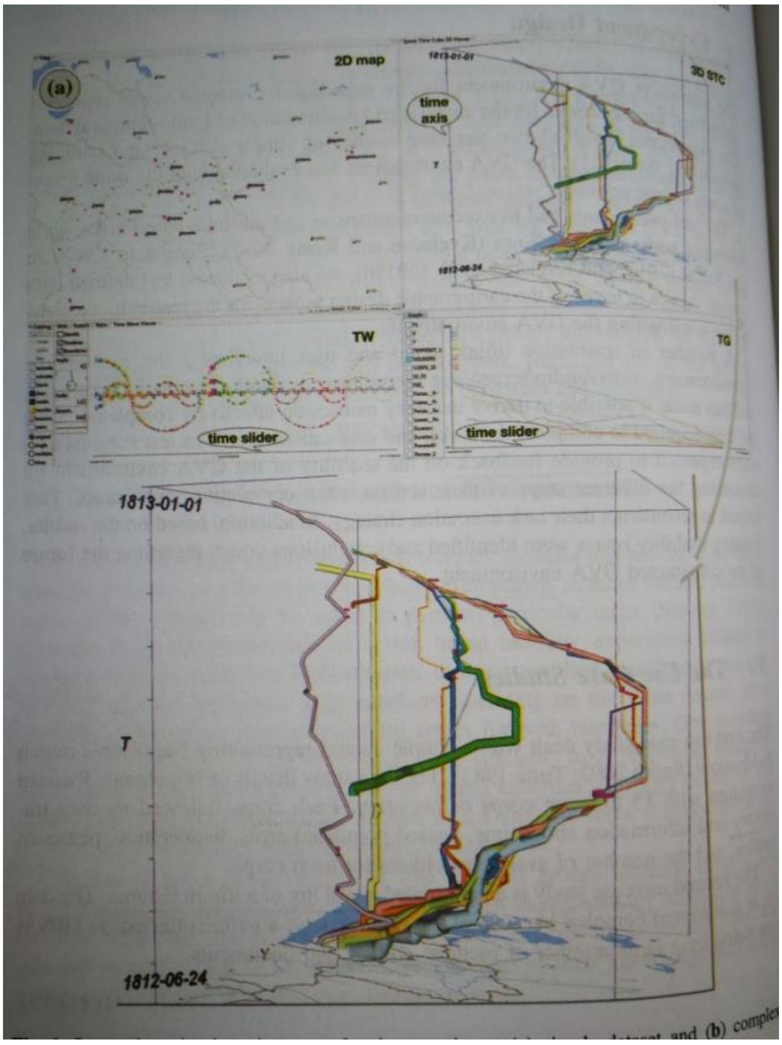


Figure 2.7 The STC for Minard's Classic Map (Kveladze, Kraak & van Elzakker, 2017)

Another remarkable research focuses on the STC for movement analysis in a Virtual Reality (VR) environment (Filho et al., 2019), in order to improve the classic, desktop-based STC in terms of estimating distance and visual depths from monocular depth cues alone, the mismatch of controlling between a 3D environment and a 2D interaction interface device (a mouse, for instance), as well as the challenges introduced by occlusion and visual clutter. It was addressed by Kveladze, Kraak and van Elzakkler (2017), as well as by Mayr and Windhager (2018) that visual clutter is a disturbing issue in a desktop-based STC. Immersive analytics, the core method applied in Filho et. al. (2019) combines stereoscopic displays and intuitive 3D interaction techniques to better explore 3D data representation. There is a virtual menu attached to the STC that allows users to thoroughly understand the spatial layout, properties of each trajectory, important times tamps, main directions of each trajectory, and soon. The STC can also be viewed in various scales and from different perspectives, which enable users to capture significant information on any preferred level.

Figure 2.8 STC in a VR environment (Filho et al., 2019)

Apart from physically transferring between places and forming trajectories, another type of movement studies the life span of "landscape objects", i.e., cultural heritages, including their creation, destruction, reconstruction and conservation. It is defined by the Charter of the Interpretation and Presentation of Cultural Heritage Sites issued by the International Council of Monuments and Sites (ICOMOS) as "temporal movement" or "landscape change". According to ICOMOS, it is particularly important to document and represent the overall process of heritage conservation and management, including all successive development phases and any conditions that had caused or influenced the site changes (Bogucka & Jahnke, 2018). In addition, all historical periods ought to be respected and incorporated into the visualization. Bogucka and Jahnke (2018) pioneeringly researched on representing the history of cultural heritage sites via an STC. By using appropriate imagery, descriptive materials, geodata base of cultural heritage sites and a georeferenced base map, different features of a cultural heritage site constructed in various time periods are illustrated as "space-time prisms" (a graphic representation that simplifies a feature's structure to prisms) and are depicted chronologically. The STC (**Figure 2.9** and **Figure 2.10**) allow people to view from various perspectives, and the time slider is useful for people to navigate to different space-time prisms. Similar to Kveladze, Kraak & Elzakker (2017), a user study was conducted by Bogucka and Jahnke for testing the STC for landscape change. The feedback from domain experts suggested that the STC excels for the completion rate and demonstrate the benefit of this novel visualization in demonstrating landscape changes. Previously, slider-based visualization which allow users to move a time slider to view landscapes changes over on only one layer (**Figure 2.9**a) was the most familiar to the domain experts and general audience. Layer superimposition, a spatio-temporal visualization method that merges multiple temporal positions into one integrated representation, while using graduated colors and transparency to depict all positions at once (Mayr & Windhager, 2018), is implemented in a slide-based visualization. Comparatively, the STC demonstrate the dynamic of cultural heritage conservation in a more intriguing and convenient way. Additionally, in the STC, the newly appeared reconstructed features no longer occlude the previously appeared ones.

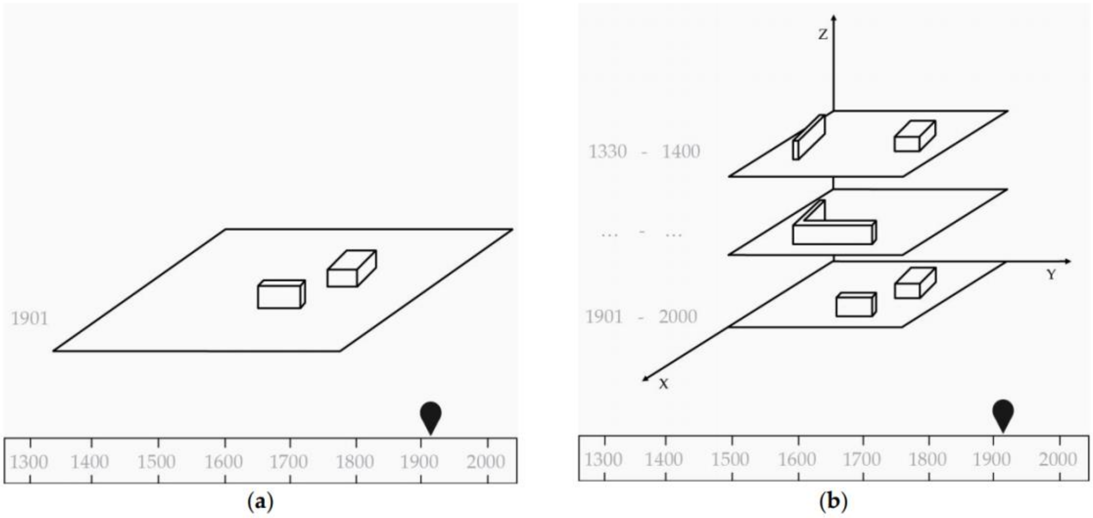


Figure 2.9 The construction of STC for cultural heritage conservation

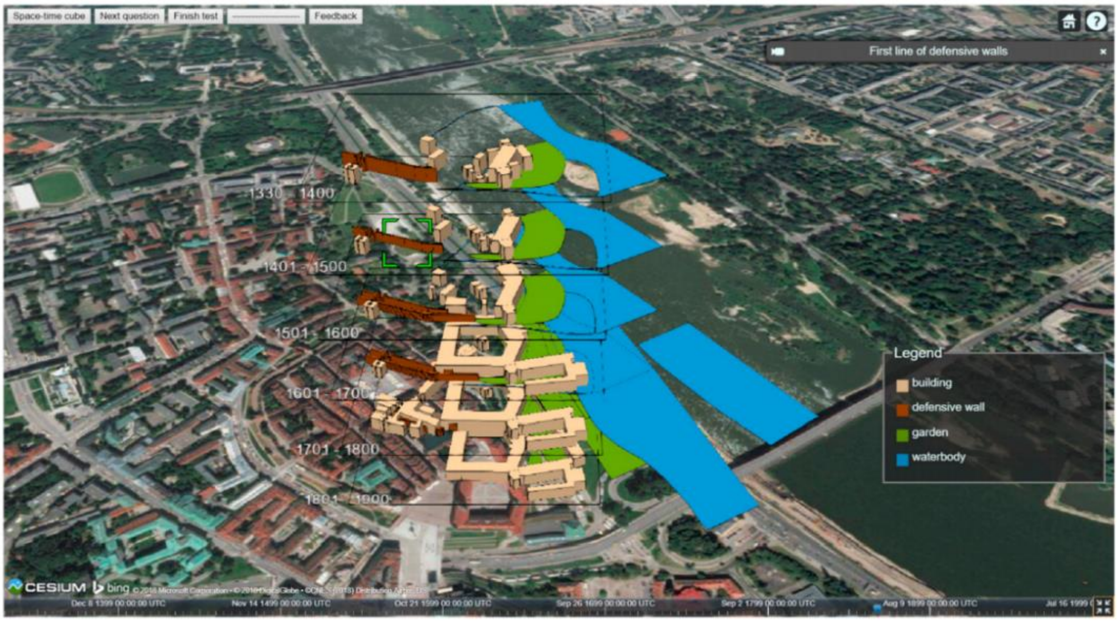
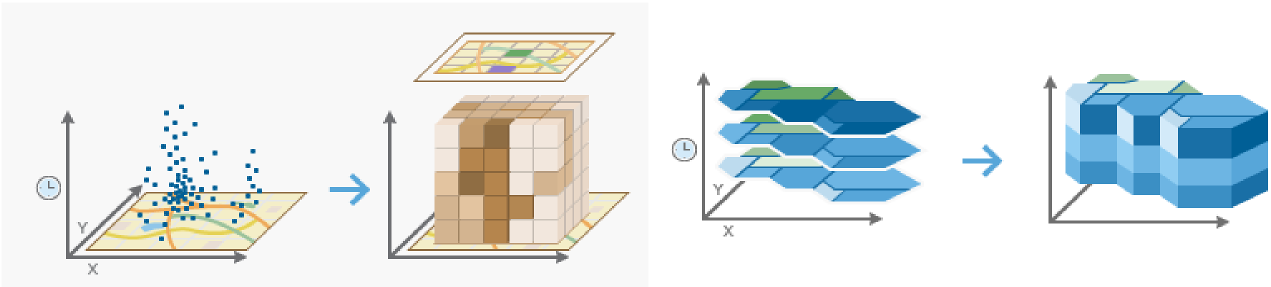


Figure 2.10 The finished STC for cultural heritage conservation

The study of STC for movement analysis, as seen from the previous three examples, focuses on continuous data. It is worth addressing that landscape objects are discrete geo-related features, but their "life spans" or "chronological movements" are continuous. The second type of STC-involved study that focuses on completely discrete data is temporal geostatistics, for which ArcGIS is a leading visualization platform. For the corresponding STC creation, the geographic surface is tessellated into fishnet grids or hexagonal grids (ESRI, 2021). Afterwards, space-time bins are created above each grid. The data points with the same temporal feature are aggregated (via sum, mean, median, max, min, etc.) into one space-time bin. In the end, aggregated values in different bins are visualized in a diverging color scale along the temporal axis, and all bins above one grid share the same location ID (which is stored in the attribute table of the input geodata) and are represented as a bin time series. It is worth mentioning that to create an STC, a projected coordinate system is required for measuring distance interval of the grids. In addition, a digital surface model (DSM) is required for setting a geographic reference.



a. Fishnet-grid STC **b.** Hexagonal-grid STC

Figure 2.11 The STC for spatio-temporal geostatistics (ESRI, 2021)

STCs for spatio-temporal geostatistics are frequently applied in the data mining process of many real-life problems. Public health issues, in particular the 2019 novel corona virus disease (Covid-19) outbreak, is one of the most-up-date examples. Mo et. al. (2020) sought to examine and analyze the spatio-temporal patterns of Covid-19 outbreaks in China from 23.01.2020 to 24.02.2020 based on STC. The test dataset, consisting of discrete confirmed cases, suspected cases and deaths were gathered online, collated in Microsoft Excel 2019 and imported into ArcGIS Pro. Afterwards, an STC based on a total of 868 spatial locations (grids) comprising 135*135 km squared was created on a vector base map of China. The values of each space-time bin in the STC record the corresponding morbidity via sum, and the time interval is set as one day. The entire STC is shown as **Figure 2.12**.

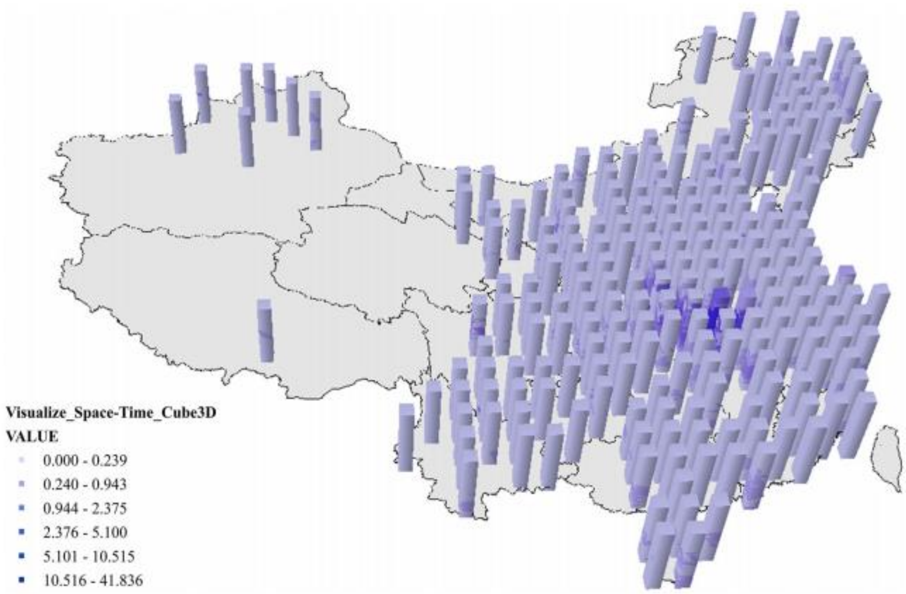


Figure 2.12 STC illustrating Covid-19 outbreak in China (Mo et. al., 2020)

It can be seen clearly that the STC highlights the areas with the most severe morbidity via space-time bins, and the temporal factor stands out independently from the geographic surface. The space-time bins on the same time step interval are illustrated on one temporal layer, indicating the severe cases across China in one day, and they are easy to access when the STC is zoomed in. The whole STC serves as a foundation of emerging hotspot analysis, a spatio-temporal data mining analysis focusing on highlighting the areas that eventually became Covid-19 prone. The STC, according to Mo et.al (2020), is a simple and powerful visualization that positions highly affected areas over time intuitively, and the overall visual effect is quite spectacular. However, it is also obvious that the space-time bins only represent the values of one data type: severe morbidity. The values of confirmed cases, suspected cases and deaths are not represented.

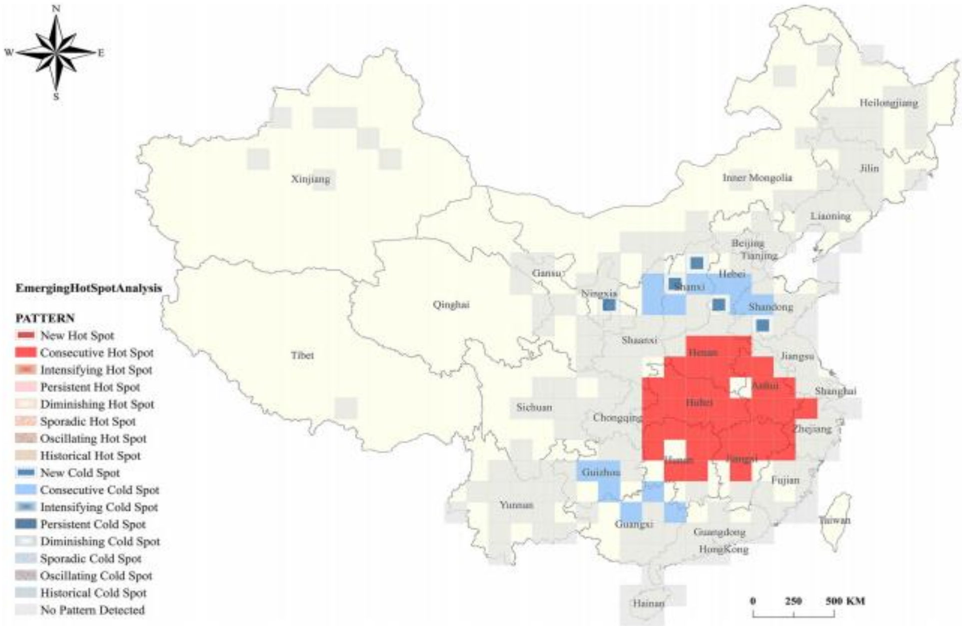


Figure 2.13 Emerging hot spot analysis of Covid-19 outbreak in China (Mo et. al., 2020)

Another covid-related scientific paper that selects an STC as the fundamental research tool is Purwanto et. al. (2021), for the pandemic outbreak in East Java, Indonesia. Instead of fishnet grids, hexagonal grids are selected for geographic surface tessellation. Apart from emerging hotspot analysis, spatio-temporal clustering was applied for the accurate detection of infected residents' movement trajectory.

Apart from public health issues, other real-life problems to which an STC is a simplistic and effective research tool include, but are not limited to crime cases (Hashim, et. al., 2019), heritage cultural metadata collection (Windhager et. al., 2019) and scattered bushfires (Turdukulov and Fazio, 2021). The main reason why STC is preferred, according to these scientific papers, is the impressive amount of insight contained in the minimalistic appearance of an STC.

2.2.2.2 Small multiples

Small multiples are a series of adjacent maps that demonstrate temporal change at the same place (Kraak, 2020). One of their apparent advantages is that by implementing the visualization method layer juxtaposition, a geographic event is depicted in more than one map in order to fragment the details of the spatio-temporal variation and therefore let readers absorb important information better (Mayr & Windhager, 2018). In Gaughan et al. (2016), dasymetric density mapping is applied to desegregate areal census data of Mainland China from 1990 to 2010 into smaller spatial units with the aid of remote sensing and GIS data, and an ensemble decision-tree classifier is applied to predict population density in the census-free years (census in Mainland China is organized in years whose number end with "0" and "5”). Based on the predicted density, it depicts the spatio-temporal distribution of major cities in China via small multiples (**Figure 2.14**). A graduated color scheme is used for illustrating the variation in population density across space and time. It is indeed a more convenient way of visualizing spatial-temporal distribution comparing to single static mapping, but in order to fully understand the relationships between the events and recognize significant patterns of information variation, readers need to cross compare the differences and mentally extrapolate spatial dynamics from the maps (Kraak, 2020). Comparing with an STC, which is capable of represent spatio-temporal patterns of a geographic event in one visual representation, small multiples are less efficient, and may require heavy visual workload when too many maps are created (Mayr & Windhager, 2018).

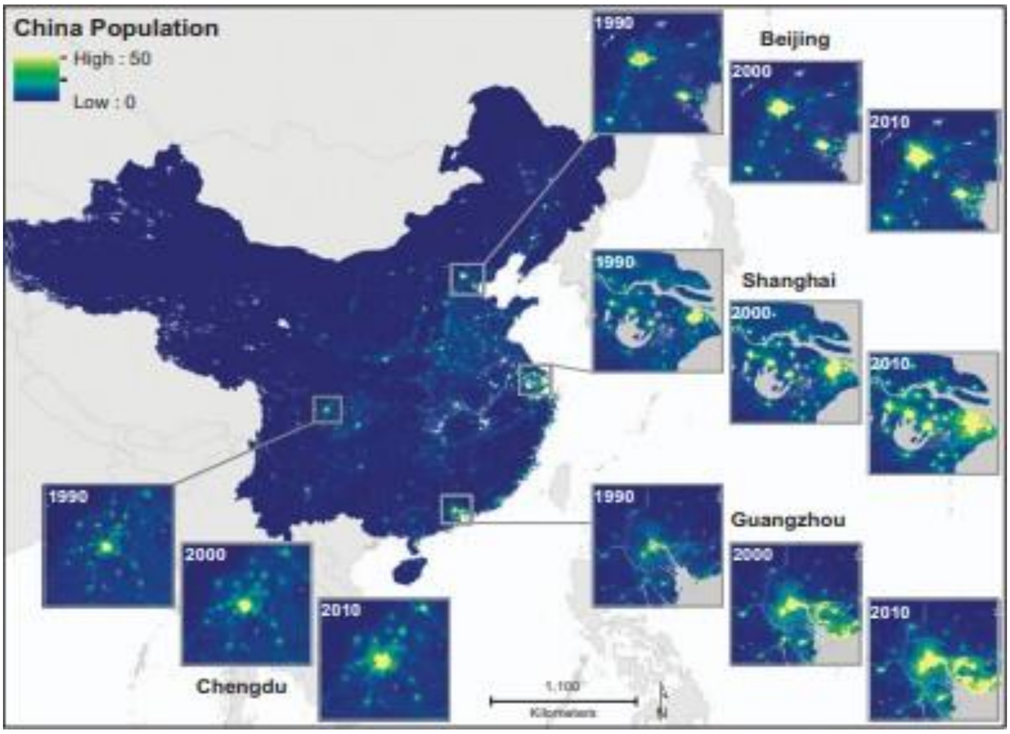


Figure 2.14 Small multiples illustrating Chinese population density (Gaughan et. al., 2016)

The following small multiples represent the concentrated regions of Covid-19 using Kernel Density Estimation (KDE), a non-parametric method to estimate the probability density of the spatial distribution of highly infected areas (Mo et. al., 2020). The temporal variation of the infected areas is visualized by a diverging color scheme (**Figure 2.15** (A), (B) and (C)), but a reader needs to read across all maps to discover the variation and form the spatial dynamic.

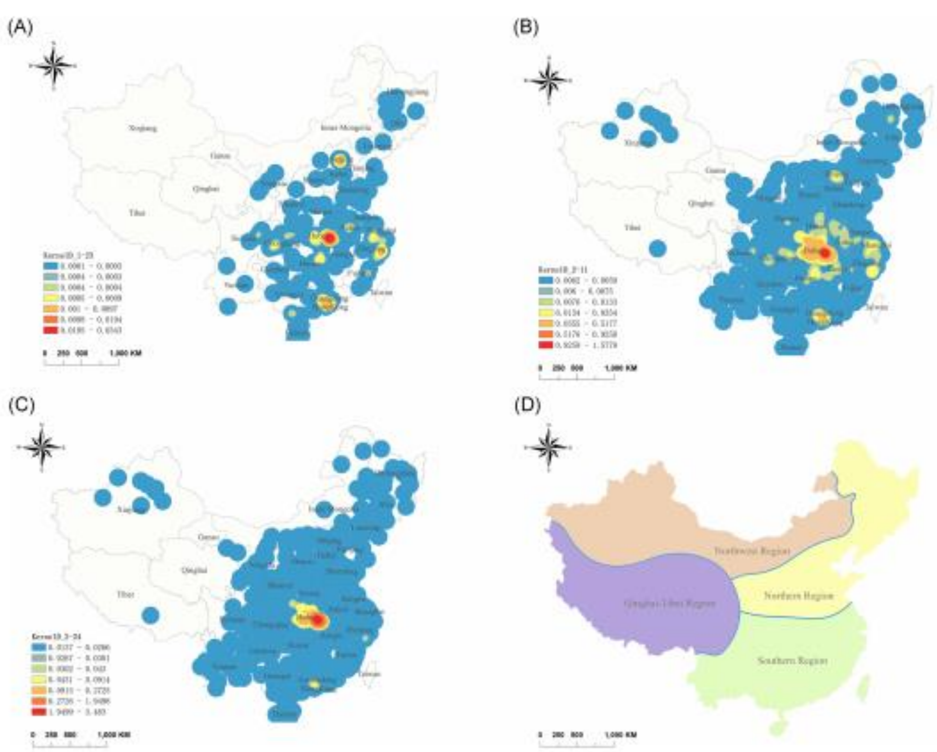


Figure 2.15 Small multiples showing KDE results of highly affected areas by Covid-19 (Mo et. al., 2020)

2.3 Geovisual Storytelling

2.3.1 General concepts

Before diving deep to the concept "geovisual storytelling", there are some fundamental terms worth being explained.

Story: an account of specific events, places and people (Roth, 2020). Notice that "people" may not necessarily be human beings, but can instead be personified features, scientific data, landmarks or any other items that are counted as "central characters". Human beings are sometimes regarded as "hidden characters" when using metaphors to convey ideas via the assigned "central characters" (Roth, 2020).

Storytelling: a technique of communicating a sequence of events as a way to document or explain the relationship between places and people or “central characters” (Roth, 2020). The communication of events in a story usually elaborates temporal changes and the "place-character" dynamic.

Visual storytelling: a specific storytelling strategy that implements visual aids including diagrams, images, videos, etc., to support the linguistic sign system (natural language) of a story (Segel & Heer, 2010). Thanks to the backdrop roles that visual aids play, the Economist once considered visualization designers as "melding the skills of computer science, statistics, artistic design and storytelling" (Segel & Heer, 2010).

As such, geovisual storytelling is a subset of visual storytelling that frequently applies geo-tagged media to more or less centrally illustrate a protagonist's movement (Mayr and Windhager. 2018). The most fundamental geo-tagged medium is maps. As many scholars such as Caquard, Cartwright (2014), Mayr and Windhager (2018) pointed out, maps are regularly used to study the geographic nature of stories. Robinson (1957, cited in Roth, 2020) also presented cartography in his studies as a story still "unraveling". The previously mentioned "Carte figurative des pertes successives en hommes de l'Armée Francais dans la champagne de Russe 1812-1813" is a convincing example of how strong narrative power a map has. Although the map did not explicitly clarify the exact details in each battle, the exact duration of each battle, the reason why the corp branched at different locations and other critical information, the simple an effective graphic representation is already powerful enough to cover the most important facts of the brutal campaign initiated by Napoleon (Brinch, 2019). This clearly explains why journalists, activists, lobbyists and individuals are all fond of using maps in articles, reports, digests and columns (Caquard & Cartwright, 2014). However, no matter how seamlessly and elegantly a map, especially a static one, can be designed, its temporal resolution is inevitably limited, and so is the accuracy of the story it tells. Fortunately, thanks to the ever-changing information technologies, various types of geolocated media such as photos, videos and animations are enabled to be combined into an online story map, a type of geodata-driven blog or journal. This brings the narrative capability of maps to a whole new level and even creates geovisual storytelling as a stand-alone research area.

Roth (2020) gave one of the most inclusive introductions to geovisual storytelling. Firstly, he addressed the multi-disciplinary characteristics of geovisual storytelling and its influence on the hybridization in cartography. Secondly, he listed ten repeated storytelling themes, with the most essential ones being "intuitive", "compelling", "memorable", "situated" and "fluid". Thirdly, he explained "linear three-act spatial narrative" (Cohn, 2013, cited in Roth, 2020), a popular spatial narrative consisting of setup, conflict and resolution, to help storytellers detect story elements from separate, inherently 2D maps and form a coherent plotline (**Figure 2.16**). Setup introduces essential background information, characters and problem context of a story. It also builds up the relationship between the specific place, time and social context of a story. Conflict describes the problem in more detail, emphasizes the tension between characters and incrementally advance the plotline of a story. In the end, resolution brings the narrative arc to its cumulating moments and present concluding scenes.

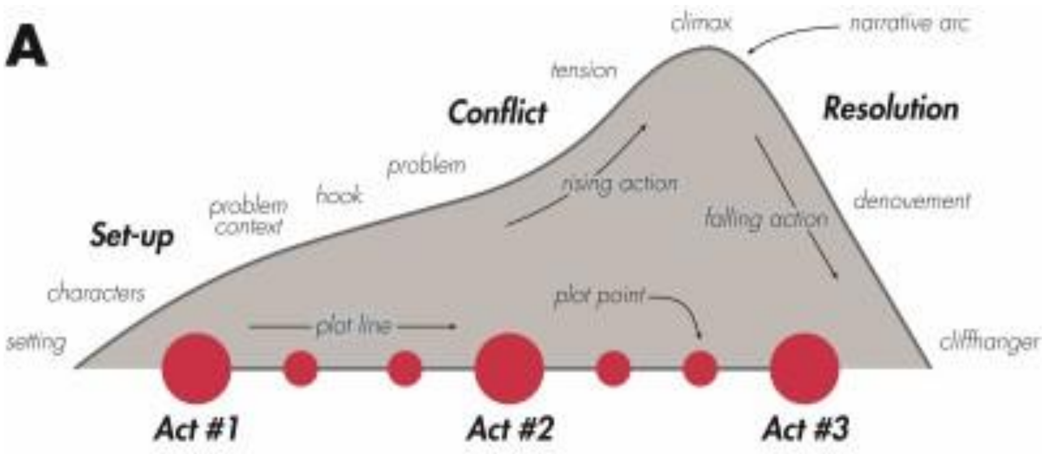


Figure 2.16 three-act spatial narrative (Roth, 2020)

Afterwards, Roth (2020) represented seven geovisual story telling genres for visually encoding a plotline, including static visual stories, longform info graphics, dynamic slide shows, narrated animations, multimedia visual experiences, personalized story maps and compilations. These genres have been well-adopted in non-geographic domains (Segel & Heer, 2010), but Roth (2020) modified them to let maps play a central or supporting role in them, and to embed more recent achievements of media and information technologies. Some genres such as static visual stories, long form info graphics and narrated animations are author-driven. An author-driven story can be narrated very appealingly, and while reading through it, people are able to mentally form the plotline and think critically. Nonetheless, they are not enabled to explore it with their own actions and unfold desired knowledge. Other genres encourage people to act, read and think more actively, and are therefore more reader-driven.

2.3.2 Examples of intriguing story maps

The following story maps combine natural language, maps and other geo-related medias in harmony. Author-driven and reader-driven genres cooperate flawlessly to make the entire story visually appealing.

The first one is "Exploring Iowa", created by Allen Carroll (2021) to introduce his anti-Covid-depression road trip. After some simple explanations of his reasons for travelling, the location of Iowa and major activities in the trip in texts, a large marked-up map with the travelling route highlighted in bright orange appears (**Figure 2.17**). On the route, POIs which Carroll's family stopped by are pinned. A reader can click on any pin to uncover the excitements of this road trip, but when the screen is simply scrolled down, POIs, along with annotations about their relevant history pop up one after the other at the moment they become part of the visual focus of the map displayed on the screen. This visualization strategy effectively raises a reader's interest to virtually follow the journey that the author took before, and acquire valuable insights during the journey.

Figure 2.17 Exploring Iowa (Caroll, 2021)

The second story map, "in Search of Home", is about a global political issue: refugee crisis (Lee, 2021). Apart from pictures and diagrams that illustrate the heart-wrenching situations of refugees, two marked-up maps are designed to construct an appealing geospatial narrative. One is a flow map showing the countries which refugees from Afghanistan, Myanmar, the Syrian Arab Republic, South Sudan and Venezuela frequently seek asylum. According to United Nations High Commissioner for Refugees (UNHCR), refugees from Afghanistan, Myanmar, the Syrian Arab Republic, South Sudan and Venezuela compose up to 67 percent of the world's refugee population (Lee, 2021). Therefore, understanding the actual situations of these countries is a crucial starting point to understand the pressure that major refugee-receiving countries are facing. The other is a dot map with 16 countries marked in order. When the story is scrolled down, similar to "Exploring Iowa", pop-ups of the corresponding countries will show up in order. There is a short video about the asylum-seeking emergencies the corresponding country faces in each pop-up. By hearing the narratives in the videos and interacting with the map, the audience can be totally taken into a virtual journey of the refugees' heartbreaking lives, and the sympathy towards the refugees can also be raised easily.

a. Flow map

b. Marked-up map with videos

Figure 2.18 Maps in "In Search of Home" (Lee, 2021)

Besides longform story maps that involve screen scrolling, there are also one-screen, dashboard-like story maps that balance author-driven and reader-driven genres properly. One example, as mentioned in chapter one, is the GUI designed by Ding, Yang and Meng (2015) that displays the spatio-temporal and semantic patterns of popular destinations that taxi passengers from Shanghai Pudong Airport reach in different days. There is also an interactive histogram in this GUI that indicates number of real-time passengers travelling from the airport to different destinations. Necessary background information about the average passengers traveling from the airport in different hours of a day is also explained in detail.

2.3.3 Up-to-date story telling gernes and immersive blocks

ArcGIS StoryMaps is one of the most commonly used online story mapping platforms enabling users to add a strong sense of place to their narratives, illustrate spatial relationships, and add visual appeal and credibility to their ideas (Caroll, 2021). This subsection summarizes some newly developed genres, sub-gernes and immersive blocks from ArcGIS StoryMaps which were applied in the making of a story map for this thesis.

2.3.3.1 Multimedia visual experiences

Multimedia visual experiences is a relatively novel, web-enabled gerne that enforces linearity (Roth, 2020). Hyperlinked text and geo-tagged multimedia materials make the internet a "web". In the meantime, multimedia visual experience uses hyper linking to activate an immersive array of images, graphics, audio and video that company a central textual narrative (Roth, 2020). Structure wise, it mashes up well with other genres and include slippy web maps to support drill-down or "martini glass narrative structure (Segel & Heer, 2010)". A martini glass narrative structure starts with an author-driven approach by questions, observations or text paragraphs to introduce the visualization. When the author’s intended narrative is complete, the visualization opens up to a reader-driven stage (Segel & Heer, 2010), which allows the audience to interactively control the pacing, and it helps develop a deep sense of place through rich integration of images, maps, videos and sounds (Roth, 2020).

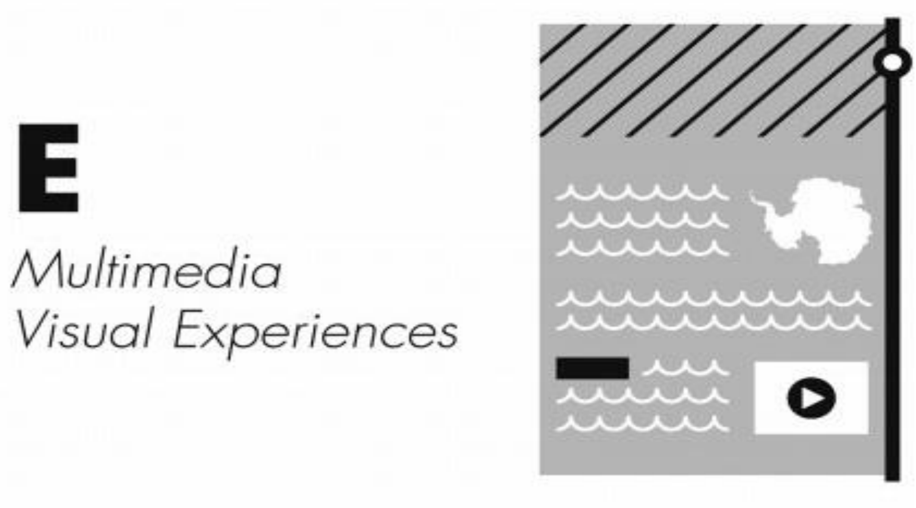


Figure 2.19 Multimedia visual experiences (Roth, 2020)

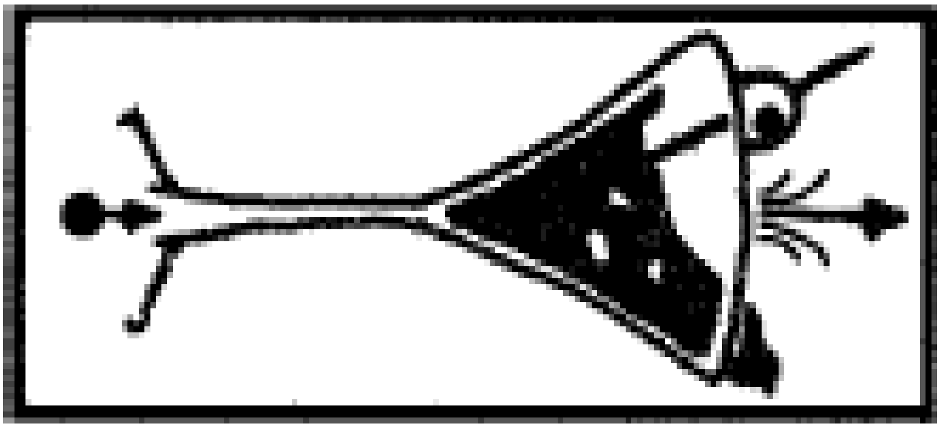


Figure 2.20 Martini glass visualization structure (Segel & Heer, 2010)

2.3.3.2 Sidecar

Sidecar (**Figure 2.21**) is a scrolling, slide-based immersive block which integrates all contents of a common sub-topic together. It is designed to break down a story map into concentrated fragments, so that users will not feel overwhelmed by information on a never-ending screen. This arrangement gives a neat and organized layout of an informative story map.

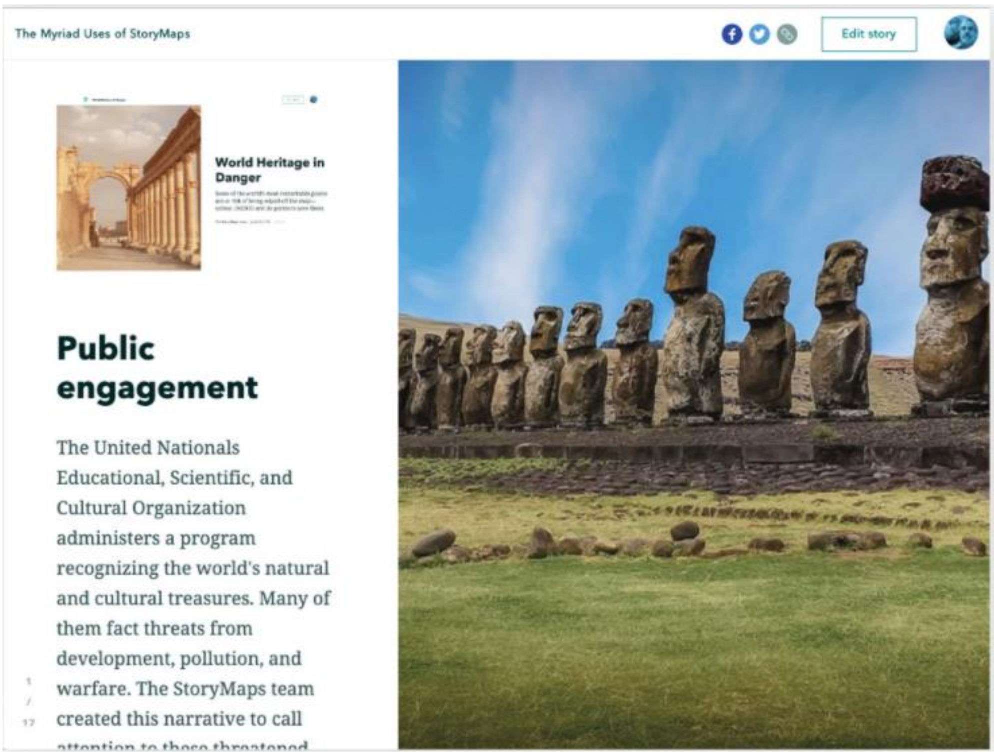


Figure 2.21 Sidecar (Caroll, 2021)

2.3.3.3 Swipe

Swipe (**Figure 2.22**) is a newly developed media block that enables users to compare two maps or images for one study are a simultaneously by using an interactive, drag able handle (Caroll, 2021). While absolute time is what many maps rely on, relative time is more widely employed in storytelling (Kraak et al., 2020). By simply moving the swiping handle, relative temporal change of natural landscapes, human settlements, population density, etc. can be demonstrated flawlessly for a user. The block offers an immediate, one-to-one relation of whatever contrast or similarity a storyteller tries to convey (Caroll, 2021). It is an advancement of ArcGIS StoryMap, because previously, temporal patterns among two maps, two images or two layers on the same map could only be examined by a user via constantly flipping slides, scrolling sidecars, reading multiple paragraphs and any other cumbersome methods (Caroll, 2021) which are very demanding for working memory and drastically increase interaction costs (Mayr & Windhager, 2018).

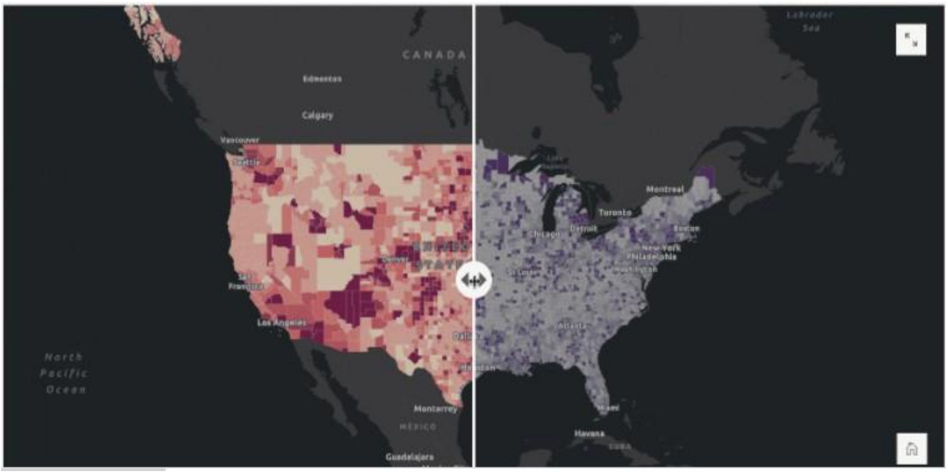


Figure 2.22 Swipe (Caroll, 2021)

2.3.3.4 Time-aware map

Time-aware map (**Figure 2.23**) is essentially an animation that consists of a base map and a published (to ArcGIS Online, for instance) time-enabled layer which directly enables a time slider for interactively browsing the entire dataset through time. Jasperson (2020) used a time-aware map to display the cumulative number of cases of cases per country from January 2020 to May 2020.

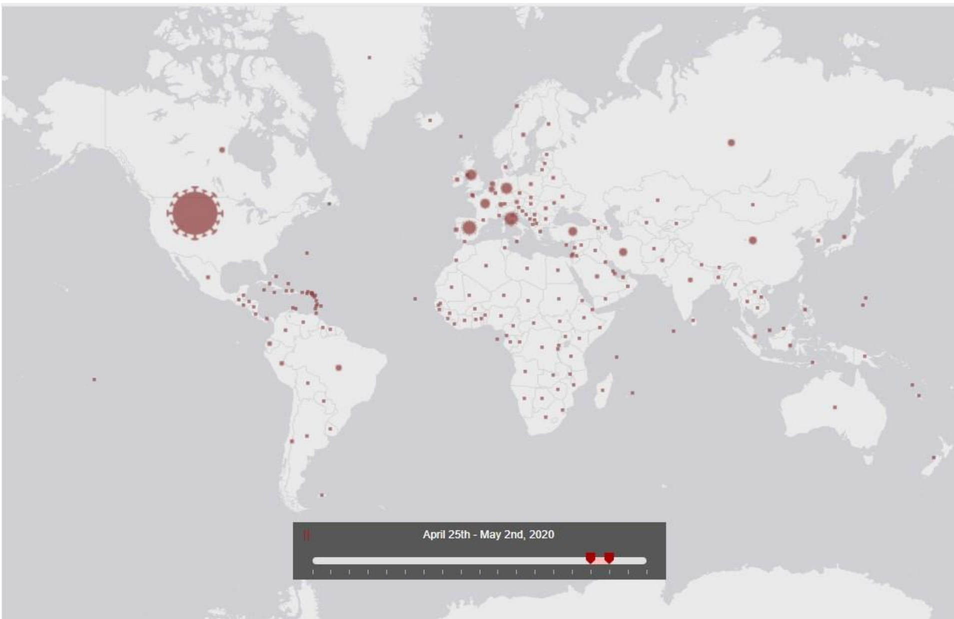


Figure 2.23 Swipe (Jasperson, 2020)

3 Methodology

This chapter introduces three visualization prototypes to represent the spatio-temporal and semantic distribution of human-created, discrete features associated with multiple categories. These features include, but are not limited to companies, schools and administrative complexes. Afterwards, a geovisual storytelling plot line for crafting the visual outcomes together is constructed.

3.1 Spatio-temporal visualization

3.1.1 Space-time micro landscapes

The following statements from Mayr and Windhager (2018), is the fundamental reason why the STC is selected as the basic form of the prototypes designed for this thesis:

Time helps people to establish casual and motivational links between discrete and incoherent events in the physical space via its linearity. By emphasizing the sequential and chronological aspects of these events, it benefits people to derive spatial dynamics a great deal. An STC not only provides a direct integration of spatio-temporal coordinates, but also depicts time to space, making the temporal and geographic information of an event similarly salient.

These statements, together with the research conducted by Kraak (2003), Bogucka and Jahnke (2018), Mo et. al. (2020), Purwanto et. al. (2021), Hashim et. al. (2019), as well as Turdukulov and Fazio (2021), justifies the prominence of STC in grouping spatio-temporal variations conveniently. Since one core topic of this thesis is the “spatio-temporal distribution” of geo-related features, a visual representation that makes temporal and geographic information similarly salient is an ideal choice.

In comparison, as addressed in Gaughan et. al. (2016) and Mo et. al. (2020, only the series of KDE maps) and Kraak et. al. (2020), layer superimposition and layer juxtaposition are risky of being cumbersome and redundant when representing spatio-temporal dynamics. Furthermore, according to Mayr and Windhager (2018), time is only encoded as ancillary information to arrange events in a sequential order in layer superimposition and layer juxtaposition. It is therefore not visually salient enough. Moreover, as mentioned in Taylor (2021, a & b), 3D symbol design is an emerging technique that uses intuitive and eye-refreshing graphics to illustrate significant characteristics of a geodataset. Hence this thesis proposes “space-time micro landscapes”, a prototype that adopts the basic design principle of the STC for spatio-temporal geostatistics (ESRI, 2021), but aims at representing the aggregated results of multiple data types (or categories) in each space-time bin.

The designing procedures of “space-time micro landscapes” is summarized as follows. The corresponding flow charts and are in the appendix.

A geodataset in XLSX, CSV, SQL, TXT or any other similar form and with discrete geo-related features recorded usually consists of the following fields. The specific test dataset provided for the case study of this thesis will be introduced in chapter four, “Experiments.”

Field	Property
ID	Numeric, string or a combination of both
Name	String
Time	Numeric, date or string
Address (Location)	String
Type (Category)	String
Coordinates	Numeric

Table 3.1 Common fields of a geodataset

Depending on how complicated the corresponding features are categorized into, there may be multiple fields of "type" (or "category") in a geodataset. For instance, a company can be categorized based on industry (manufacturing, transportation, architecture, commercial services, etc.), product (food, clothing, car spares, chemical, household, etc.), size (number of staff members), to name a few. In addition, the field "time" refers to the time when the corresponding feature showed in a geographic area, such as when a company was registered or a school was founded in a city.

After a geodataset gets imported into a data processing platform such as RStudio, Anaconda or D3 Library, the corresponding features can be plotted by using the "coordinates" columns and their general spatial distribution is observable. For this thesis, the data processing and visualization is completed in RStudio. As ArcGIS Pro is not yet enabled to let multiple categories of a geodataset illustrated in an STC, it will not be considered as an option.

A part of the geodataset within a certain time period can now be extracted, and the geographic surface where the corresponding features of the geodataset are located at is tessellated into equal-sized grids (**Figure 3.1**). It is worth mentioning that according to the spatial data analysis pitfall "non-uniformity of space", geo-related features are usually distributed with different densities, which sometimes lead to redundant or missing samples in different places (O'Sullivan & Unwin, 2010c). Therefore, an optimal tessellation should avoid two extreme situations: overcrowded points in one grid and too many empty grids elsewhere. A reasonable strategy is: when a time period with only a few features distributed across the geographic surface is chosen, more grids on the surface should be formed, so that the exact areas with features and the completely empty areas can be precisely distinguished. When a time period with a bigger number of features distributed across the geographic surface is selected, fewer grids should be formed, so that the features wouldn't be segregated redundantly. **Figure 3.1** and **Figure 3.2** demonstrate the difference between how a geographic surface with only a few sparsely distributed features is tessellated and how the same surface with more features distributed in obvious clusters is tessellated.

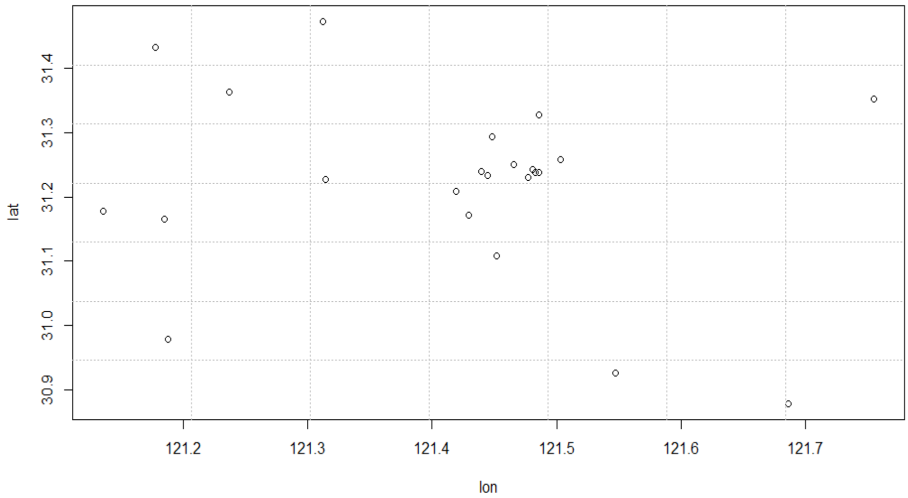


Figure 3.1 Tessellation (a few features, 7*7 grids)

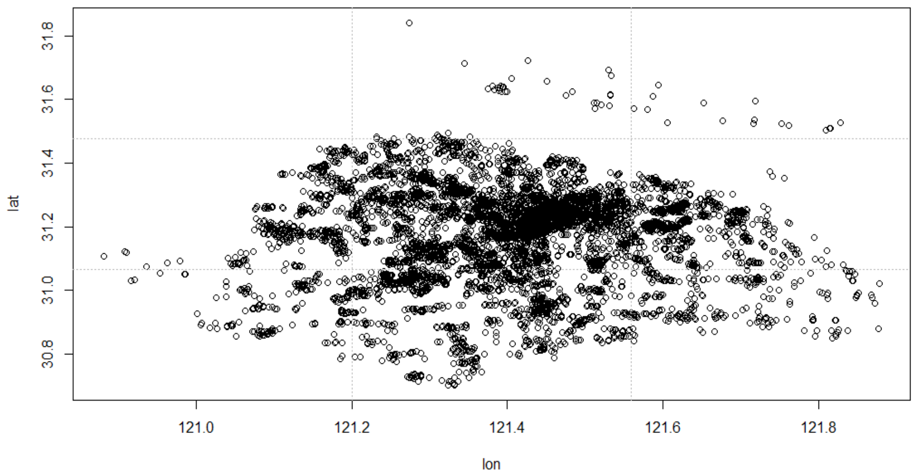


Figure 3.2 Tessellation (many more features, 3*3 grids)

The similar strategy also holds when the selected part of dataset is segregated along the temporal axis. There are some time step intervals with very few features, and others with hundreds and thousands of features. It is essential to choose a proper number of intervals to avoid overcrowded points in one time step interval and too many empty timestep intervals. A convenient method is: when the selected time period is a composite number (4 hours, 6 days or 10 years, for example), its factors can be used as the numbers of time step intervals, and the factor that makes the entire prototype appear less redundant will be chosen. If the selected time period is a prime number (5 hours, 7 days or 13 years, for example) the factors of its closest composite number can be the numbers of time step intervals, and the same steps will be executed for determining the well-suited time step interval.

After the spatial and temporal tessellations, space-time bins with 3D proportional symbols such as spheres (**Figure 3.3**) can be created. Every bin has a fixed position in space and time, much similar to the STC for spatio-temporal geostatistics (ESRI, 2021). In one bin, the size of each proportional symbol represents the number of its corresponding geo-related features, and its position indicates the centroid of these features. The z-coordinate (or t-coordinate) of the symbols are calculated by the average temporal values associated with the corresponding features. For instance, if there are 10 manufacturing companies founded from 1920 to 1930 and are located at the south-east corner of a city (one year, one company), the z-coordinate of their corresponding proportional symbol will be 1925.5. It only indicates the position of the symbol in the 3D space.

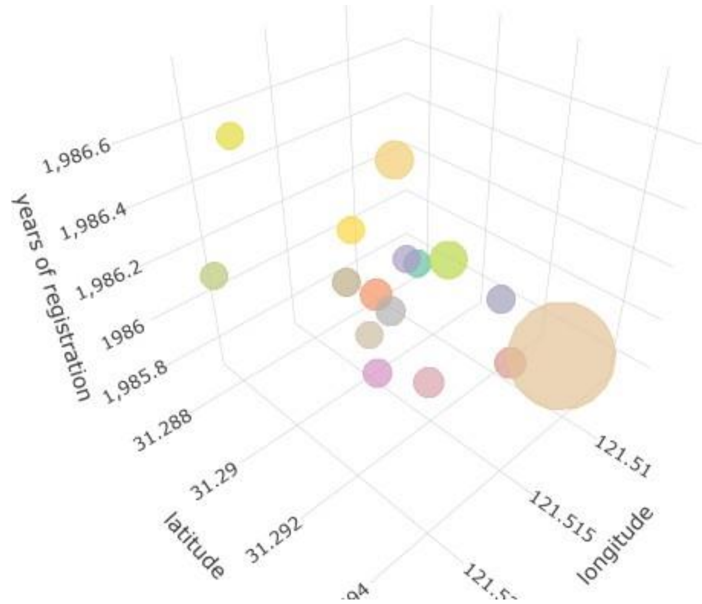


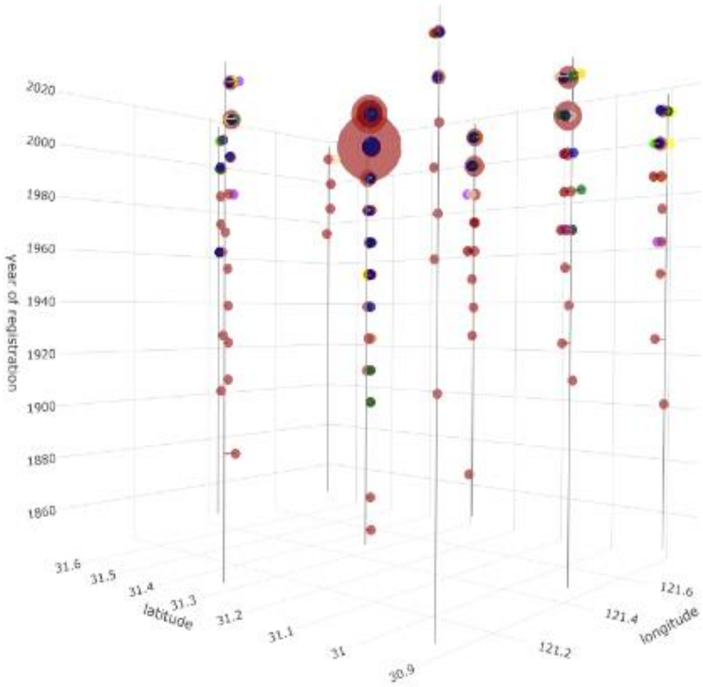
Figure 3.3 Aggregated symbols in one space-time bin

As seen from **Figure 3.3**, all scattered data points are replaced by their corresponding 3D proportional symbols. However, these symbols still appear rather scattered in the space-time bin. In order to make the entire prototype more resemble an STC, to strongly emphasize which bins contain what symbols, and to make all “bin time series” (as mentioned in chapter two) differentiated clearly from each other, a centralization factor lambda is applied to pull all proportional symbols diagonally towards the center of their corresponding space-time bins. The mathematical theory behind this step can be summarized as:

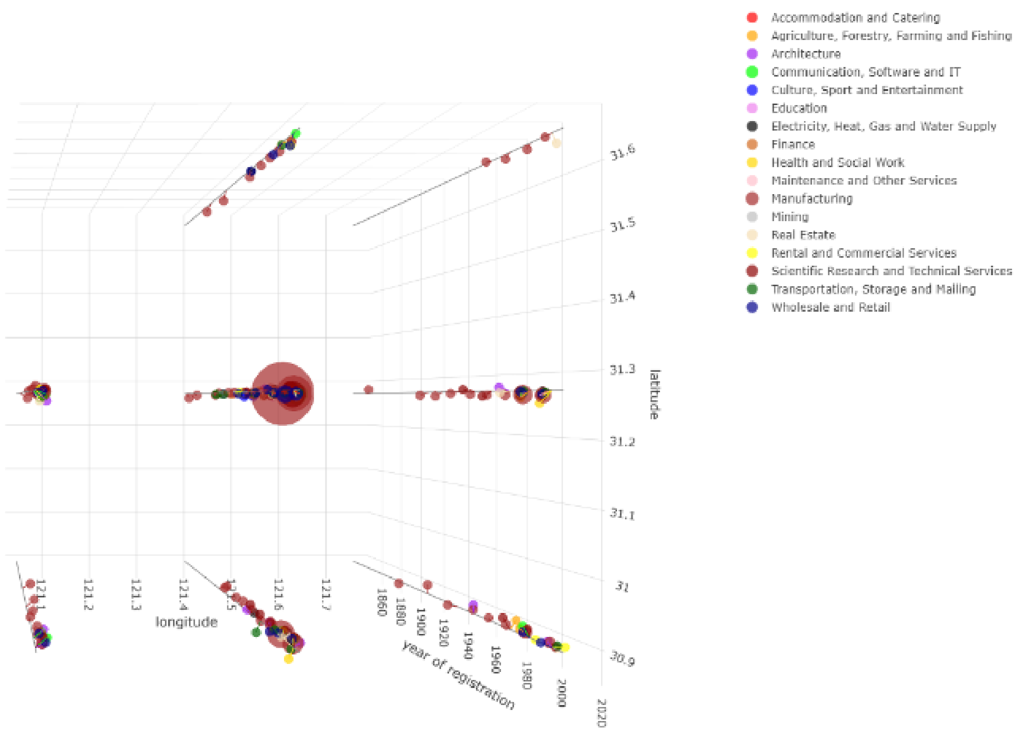
(lat1,lon1,tim1)=(lat,lon,tim)+[(lat0,lon0,tim0)-(lat,lon,tim)]*lambda,

In which the original position of each symbol is denoted as (lat0, lon0, tim0), the center of the corresponding space-time bin as (lat, lon, tim), the factor as lambda, and the new position of each symbol as (lat1, lon1, tim1). Lambda is set between 0 and 1. According to the formula above, all symbols in one space-time bin will be completely dragged to the center and coincide with each other when lambda is 0, and will all stay at their original positions (where their corresponding features are aggregated) when lambda is 1. Additionally, the lambda on the temporal dimension can be set smaller than the one on the spatial (lat and lon) dimension, in order to make the time step intervals more distinguishable from each other.

This way, rather than designing a single 3D symbol indicating all categories of and numbers of the corresponding data records and placing it at the center of the corresponding space-time bin, preserves the shapes of the original symbols, but only "micronizes" or “shrinks” their distribution by lambda. If one symbol, for instance, is located at the bottom left corner of a space-time bin, it will stay at the bottom left after lambda is applied, but only closer to the center of the bin. This therefore explains the term "micro landscapes".

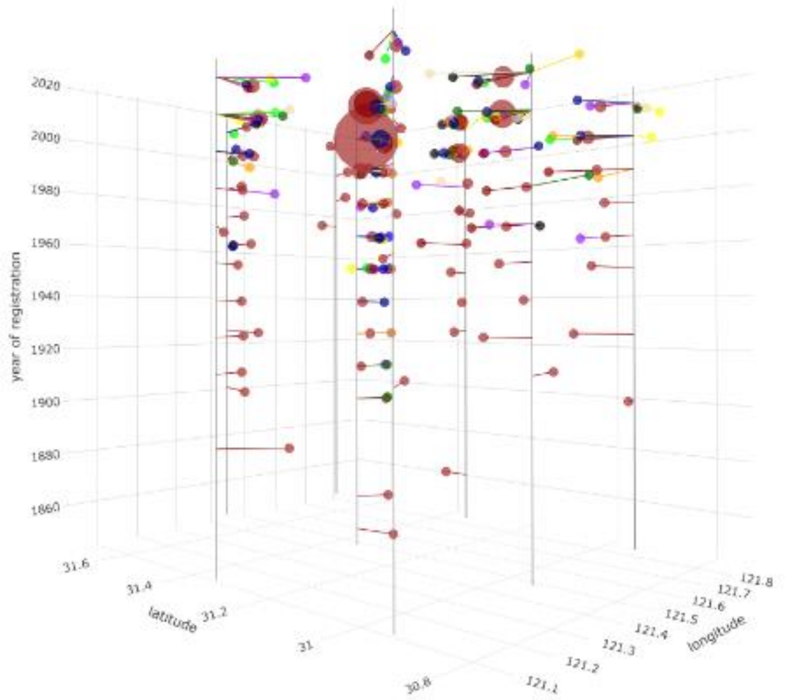


a. lambda=0.1 (perspective 1)

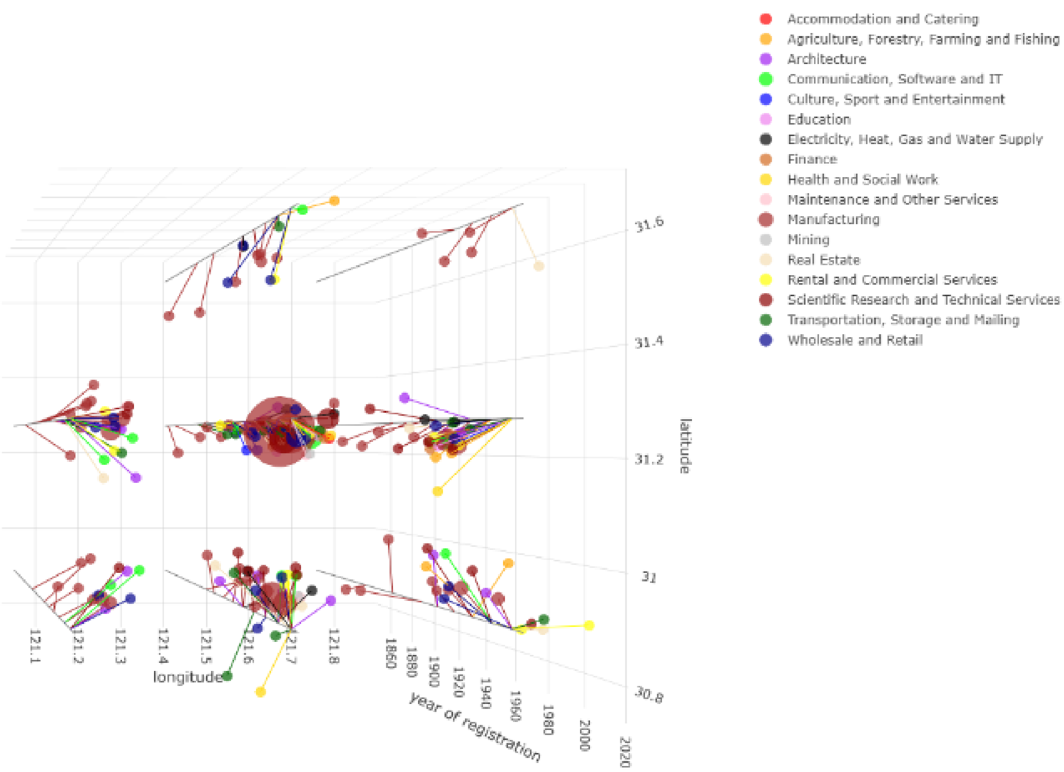


b. lambda=0.1 (perspective 2)

Figure 3.4 Space-time micro landscapes (grids, lambda = 0.1)



a. lambda=0.7 (perspective 1)

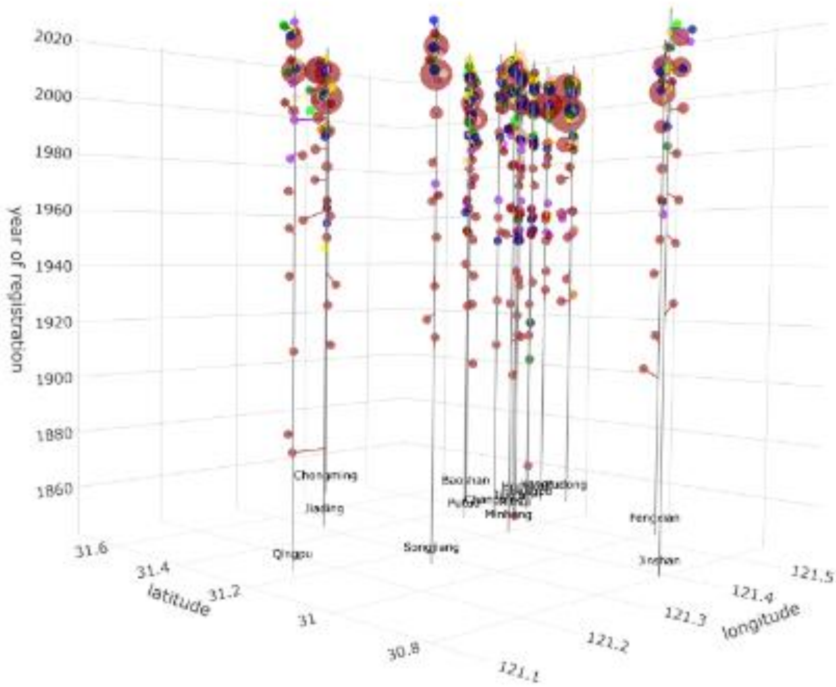


b. lambda=0.7 (perspective 2)

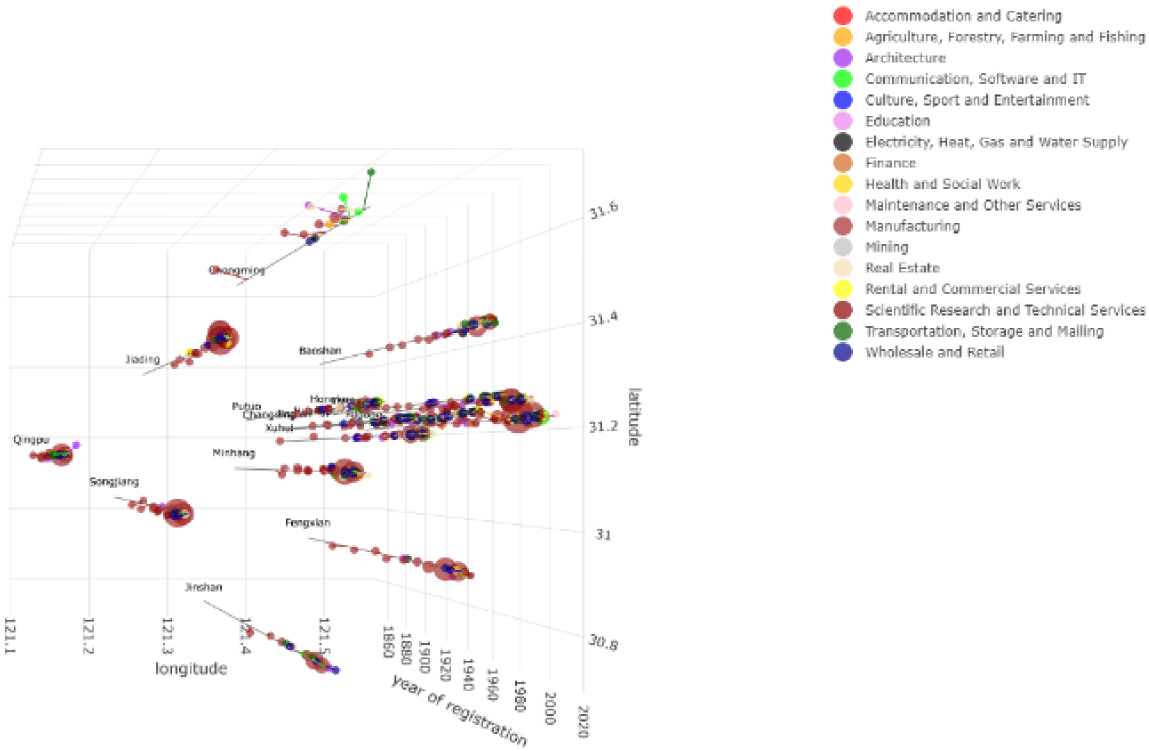
Figure 3.5 Space-time micro landscapes (front and top, lambda = 0.7)

From **Figure 3.4** and **Figure 3.5**, the effect of "micro landscapes" as 3D symbols can be sensed clearly. The numbers and categories of different features are represented by different colors and sizes of the proportional symbols (the individual spheres). The thin vertical lines indicate the centers of the space-time bins in which the symbols stay, and the “branches” or “petals” that connect the symbols with the centerlines indicate how far the symbols are from the centers of the corresponding space-time bins. Additionally, the effect of lambda is apparent: when lambda = 0.1 (closer to 0), the symbols are nearly concentrated at the centers of the space-time bins; and the temporal layers are quite distinguishable. When lambda=0.7 (closer to 1), on the other hand, the symbols become far away from the centers, making the temporal layers and the bin time series not as distinguishable. It is also obvious that a larger lambda would make the symbols more resemble “data flowers”, because the “branches” or “petals” become more visible. Furthermore, an alternative way of differentiating the temporal layers from each other is to colorize all segments of the centerlines and make the segments on one temporal layer shaded with the same hue. Users are able to immediately navigate to the desire interval they intend to explore based on the color coding of the line segments.

Apart from tessellating the geographic surface into grids, another type of aggregation highly resembles" creating an STC based on fixed locations" mentioned in chapter two. If there is information about "administrative units" of geographic features such as "districts", "community" or "street" in a dataset, it can be directly used for aggregation. There is no need to manually tessellate the study area. The corresponding space-time micro landscapes are shown in **Figure 3.4** and **Figure 3.5**, with lambda set as 0.1 and 0.7 respectively.

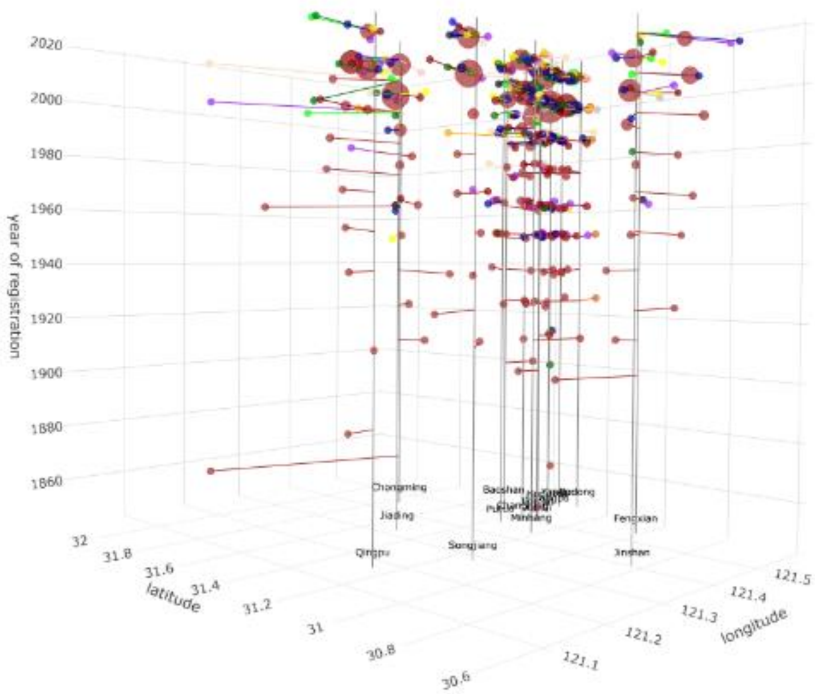


a. lambda=0.1 (perspective 1)

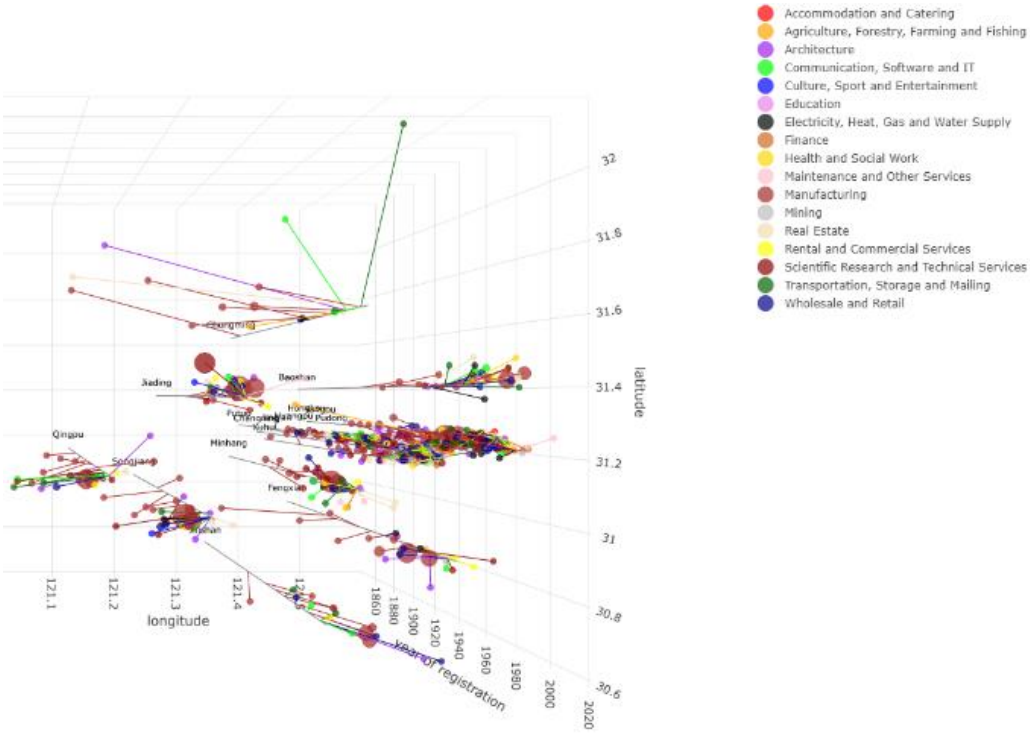


b. lambda=0.1 (perspective 2)

Figure 3.6 Space-time micro landscapes (administrative units, lambda = 0.1)



a. lambda=0.7 (perspective 1)



b. lambda=0.7 (perspective 2)

Figure 3.7 Space-time micro landscapes (administrative units, lambda = 0.7)

Once again, the effect of "micro landscapes" for the aggregated symbols can be sensed clearly. The names at the bottom of the centerlines are the names of the corresponding administrative units. It is also noticeable that a larger lambda let the proportional symbols spread even further away from the space-time bin, and thus make the bin time series even less undistinguishable from each other, comparing to the case of aggregation via grids. **This is because grids are set in uniform dimensions, which usually does not hold for real-world administrative units.** It proves further the purpose of lambda and the aforementioned formula.

Figure 3.4 – **Figure 3.7** demonstrate the space-time micro landscapes from two perspectives. Thanks to the interactivity of the prototypes, only one perspective of them will be demonstrated in the rest of the thesis. The source codes of the prototypes are linked in the appendix. Moreover, the legend of a space-time micro landscapes can be altered in R Studio to represent the numbers of features via different colors. **Figure 3.8** is a typical example.

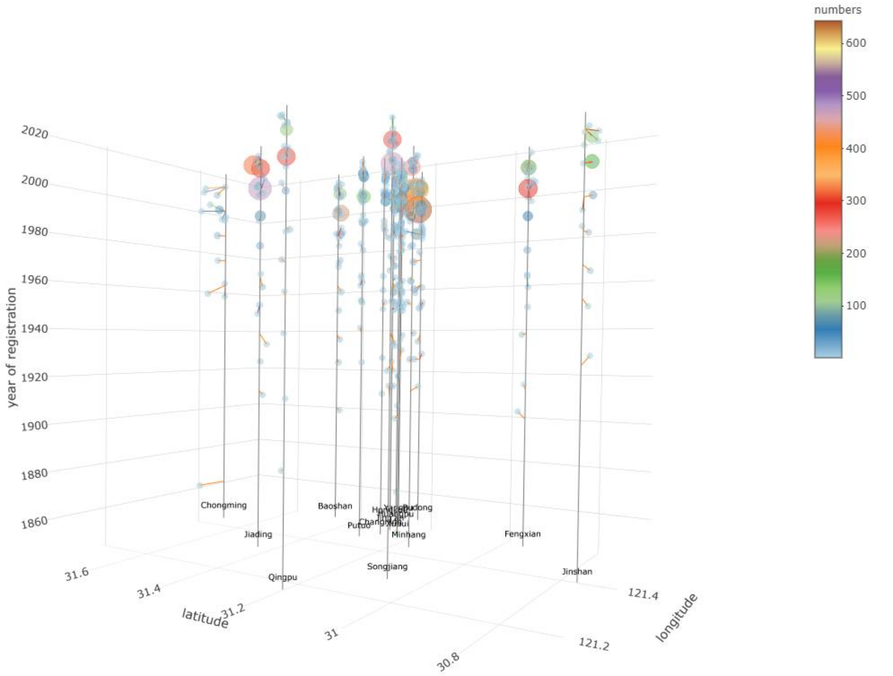


Figure 3.8 Legend representing number of features (administrative units)

In the appendix, a series of graphs demonstrate intuitively how different numbers of grids, numbers of time intervals, values of lambda and colorized centerlines vary the appearance of a space-time micro landscapes. The most optimal combination of these parameters preserves the shape of "data flower", differentiates the time step intervals from each other and minimalizes distractive visual appearance.

3.1.2 The "virgin land" problem

Virgin land, a concept that is originated in studies of agriculture, refers to a land parcel that is yet to be cultivated, as well as a person or an item is at yet underdeveloped (Collins Dictionary, 2021). In this thesis, virgin land refers to a region that is completely new for a certain geo-related feature, such as a type of company, a type of school, or infected cases of a contagious disease. The process of a type of feature started showing up and distributing around a place highly resembles the process of cultivation, which explains the name of the prototype. The study of the "virgin land" problem is hence important for understanding the characteristics of geodata, because it reveals which parts of a place (a city, a town, etc.) tend to be the "birthplace" of what type (category) of a geo-related feature, and hence critical for decision makers to issue new policies such as site planning (Sapena et. al., 2021).

The visualization process is simple and straight forward: Extract the spatial position soft he features with the earliest temporal factors (the first company under each type that was recorded in a geodataset, for instance), and then plot them in 3D. This makes it easier for a user to learn which categories of geo-related feature appear one after another, and which categories show up nearly spontaneously all at once. Thus, it is more convenient than an ArcGIS time-aware map which demonstrates the spatio-temporal variation in a 2D animation, and let the features appear on the basemap according to their temporal alignment, much similar to the "slide-based visualization" discussed in chapter two. The visual outcome of this prototype is in chapter four, “experiments”.

3.1.3 Maximal number of features

The foundation of geo-related features sometimes reflects the expansion of the corresponding geographic area. For instance, the construction of new residential communities in a city’s outskirts every year suggests urban expansion. Understanding the patterns of a geographic area’s expansion is also essential for spatial decision making such as site planning, and visualizing the sections with the maximum number of new geo-related features is an effective method of demonstrating the expansion.

The prototype creation process is similar to the one for "space-time micro landscapes", but with two exceptions. First, the features in each space-time bin are no longer categorized, but the number of them all (count) is illustrated by a proportional symbol right at the center of the corresponding space-time bin; second, on each temporal level, only the symbols representing the maximal number of features among all grids are visualized.

3.2 Story map design

3.2.1 Plotline

As discussed in chapter one, the thesis’s research aims at representing the spatio-temporal and semantic distribution of geo-related features effectively, and geovisual storytelling focuses on "crafting continuous stories from discrete analysis results (maps, diagrams or other visual representations) to process unique attributes" (Segel and Heer, 2010). Additionally, long historical events provide essential contexts for the spatio-temporal and semantic distribution of geo-related features, and according to Roth (2020), geovisual stories communicate the events as a way to document or explain the relationship between places and “central characters”, which refers to geodata in this case. As such, geovisual storytelling or story mapping is chosen for representing and explaining the visual outcomes of this thesis.

The story map design aims at using the 3D prototypes to aid the existing ArcGIS visual storytelling tools, such as static maps, interactive maps, swipe and time-aware maps. According to the "three-act" spatial narrative addressed in chapter two, the setup includes a brief introduction of the geographic area that the target geo-related features are located in, as well as a short description of the features and the associated historical background.

The main conflict is the dynamic between geo-related features and historical events of the corresponding geographic area. The positive and negative visual outcomes of geo-related features caused by the corresponding historical events are illustrated respectively. However, more often, some parts of a geodataset are of missing in formation, which may prevent the corresponding geo-related features from accurately reflecting the historical events. This is regarded as another core conflict and is included in the story map. Additionally, for the parts of a geodataset with complete information, traditional 2D mapping techniques may lead to unpleasant over plotting issues. It often becomes inevitable when a large number of features appear during a particular time period, and is addressed in the story map as well.

There are two resolutions, i.e., ending scenes associated with the story map. Firstly, the spatio-temporal and semantic distribution of human-created, geo-related features at a place is indeed influenced by significant historical events associated with that place. Secondly, instead of visualizing every single feature on their exact location and primarily relying on 2D mapping, the proposed visualizations present aggregated symbols in independent space-time bins and demonstrate the change of spatial data along the temporal axis.

3.2.2 Genres, tools and immersive blocks

The core visual storytelling genre is “multiple visual experiences”. It is, as discussed in chapter two, optimal for revealing the complete process of the temporal visualization design goals, as well as covering the history that a dataset may reflect on. Besides, the immersive blocks and other tools useful for the story map are sidecar for information fragmentation, swipe for demonstrating" before and after" and time-aware map for demonstrating the spatial distribution of data points overtime as an animation (before the aggregation). The exact definitions of these immersive blocks and tools are in chapter two.

4 Experiments

This chapter introduces a case study which the methodological frameworks proposed in chapter three is applied to, and narrates the spatio-temporal story derived from the test data set provided for this thesis.

4.1 Study area

The study area for this thesis is Shanghai, a 6,431 km squared metropolis located on the West Coast of the Pacific Ocean and the central section of the north-south coastline of China (Yang, 2002). It has served as one of the major trading ports and gateways to mainland China after the Second Opium War. Since the late 1970s when China began its economic reform, Shanghai has achieved rapid and sustained growth. Afterwards, the rapid development of finance, trade, transportation, communications, real estate and other tertiary industries have raised its proportion in Shanghai's gross domestic product (GDP) from 30% to 50% within a decade (Yang, 2002).

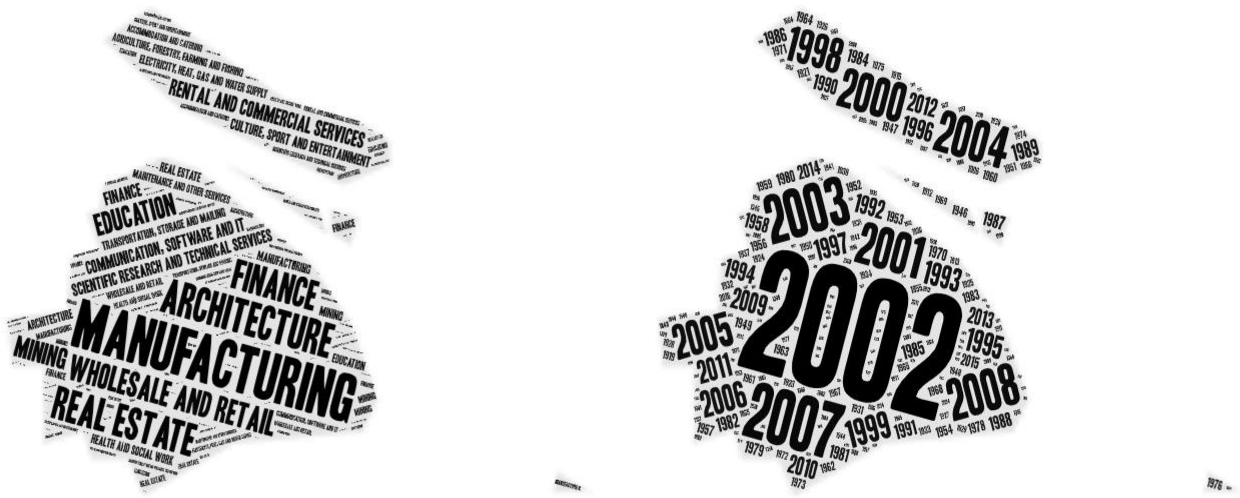
4.2 Data description

The test dataset is in the form of Excel spreadsheet, and contains information of 14510 companies in Shanghai. A part of the dataset is shown as follows. The parameter names are highlighted by the blue oval.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	_id	org_name	org_start_time	org_type	industry_type	location	org_city	wgs84_lat	wgs84_lon	firm_info	url	1986	1987
2	5da653752aeede1fc3db67e5	上海老凤祥有限公司	1848年	珠宝首饰及有关物品制造	制造业	上海市徐汇区漕溪路270号 （上	上海	31.17155173	121.429094	上海老凤祥有限公	http://inst	0	
3	5da653ce2aeede1fc3db67f7	上海周虎臣曹素功笔墨有限公司	1864年	笔的制造	制造业	上海市黄浦区金陵东路422号4—5	上海	31.2294968	121.476786	上海周虎臣曹素功	http://inst	0	
4	5da64f1e2aeede1fc3db670a	江南造船(集团)有限责任公司	1865年	船舶及相关装置制造	制造业	上海市崇明区长兴岛江南大道988	上海	31.35286322	121.75494	江南造船(集团)有	http://inst	0	
5	5da5d7102aeede1fc3db55fa	上海信谊万象药业股份有限公司	1890年	医药制造业	制造业	上海市青浦区赵重公路217号 （	上海	31.16504506	121.184484	上海信谊万象药业	http://inst	0	
6	5da657952aeede1fc3db68a6	上海东湖机械厂	1897年	通用设备制造业　专用设备制造业	制造业	上海市宝山区高逸路228号 （上	上海	31.32791433	121.485549	上海东湖机械厂基	http://inst	0	
7	5da68e762aeede1fc3db6f9e	上海大颜仪器设备有限公司	1900年	仪器仪表制造业	制造业	上海市徐汇区龙吴路2710号7号楼	上海	31.10853561	121.451574	上海大颜仪器设备	http://inst	0	
8	5da5e9302aeede1fc3db58de	上海盛伟慧科科技股份有限公司	1904年	气体压缩机械制造	制造业	上海市宝山区沪樊路53号1幢，8幢	上海	31.47328215	121.312461	上海盛伟慧科科技	http://inst	0	
9	5da6e7612aeede1fc3db7a7c	中交上海航道局有限公司	1905年	水上运输业	交通运输、仓储和邮政业	上海市黄浦区中山东一路13号6楼	上海	31.23867384	121.485466	中交上海航道局有	http://inst	0	
10	5da5d8f22aeede1fc3db5657	上海武马食品科技有限公司	1906年	食品制造业	制造业	上海市奉贤区海湾镇五四农场洪卫	上海	30.87789372	121.686235	上海武马食品科技	http://inst	0	
11	5da653b62aeede1fc3db67f2	上海铁路通信有限公司	1906年	通信设备制造	制造业	上海市西藏北路489号 （上海市	上海	31.25012382	121.465297	上海铁路通信有限	http://inst	0	
12	5da64f5d2aeede1fc3db6717	上海中华药业有限公司	1911年	医药制造业	制造业	上海市长宁区定西路685号2楼 （	上海	31.20896547	121.419208	上海中华药业有限	http://inst	0	
13	5da5d9342aeede1fc3db5665	上海华宝孔雀香精香料有限公司	1912年	香料、香精制造	制造业	上海市嘉定区叶城路1299号 （上	上海	31.36249083	121.237221	上海华宝孔雀香精	http://inst	0	
14	5da6a6112aeede1fc3db72cf	上海皮尔博格有色零部件有限公司	1912年	有色金属铸造	制造业	上海市嘉定工业区兴贤路1288号	上海	31.43258458	121.177226	上海皮尔博格有色	http://inst	0	
15	5da6503b2aeede1fc3db6746	上海减速机械厂有限公司	1913年	齿轮及齿轮减、变速箱制造	制造业	上海市虹口区昆明路100号 （上	上海	31.25744306	121.503422	上海减速机械厂有	http://inst	0	
16	5da6539d2aeede1fc3db67ed	上海电器股份有限公司人民电器厂	1914年	电气机械和器材制造业	制造业	上海市静安区共和新路3015号 （	上海	31.29322682	121.447842	上海电器股份有限	http://inst	0	
17	5da652bc2aeede1fc3db67c5	上海开林造漆厂	1915年	涂料制造	制造业	上海市青浦区工业园区崧泽大道83	上海	31.17699523	121.135929	上海开林造漆厂基	http://inst	0	
18	5da5d98f2aeede1fc3db5675	冠生园(集团)有限公司	1915年	食品制造业	制造业	上海市静安区新闸路1418号 （上	上海	31.23320334	121.444724	冠生园(集团)有限	http://inst	0	
19	5da64ebd2aeede1fc3db66fa	上海华线医用核子仪器有限公司	1916年	医疗诊断、监护及治疗设备制造	制造业	上海市闵行区华漕镇南华街165号	上海	31.22784619	121.314169	上海华线医用核子	http://inst	0	
20	5da6510d2aeede1fc3db6770	上海鹰牌衡器有限公司	1917年	衡器制造	制造业	上海市奉贤区青村镇林东路169弄1	上海	30.92558296	121.547451	上海鹰牌衡器有限	http://inst	0	
21	5da69aee2aeede1fc3db7167	上海华通开关厂有限公司	1919年	配电开关控制设备制造	制造业	上海市松江区长塔路418号 （上	上海	30.97886578	121.187677	上海华通开关厂有	http://inst	0	
22	5da6e6a52aeede1fc3db7a62	上海强生集团有限公司	1919年	公共电汽车客运	交通运输、仓储和邮政业	上海市黄浦区四川中路213号18楼	上海	31.23825847	121.482901	上海强生集团有限	http://inst	0	
23	5da653d82aeede1fc3db67f9	上海实业马利画材有限公司	1919年	颜料制造	制造业	上海市西康路850号 （上海市 邮	上海	31.24020084	121.439324	上海实业马利画材	http://inst	0	
24	5da6bb662aeede1fc3db75bd	上海市基础工程集团有限公司	1919年	房屋建筑业　土木工程建筑业	建筑业	上海市黄浦区江西中路406号 （	上海	31.24284548	121.48088	上海市基础工程集	http://inst	0	
25	5da654f82aeede1fc3db6827	上海造币有限公司	1920年	其他未列明金属制品制造	制造业	上海市普陀区光复西路17号 （上	上海	31.2518864	121.430634	上海造币有限公司	http://inst	0	
26	5da5d4c32aeede1fc3db559e	上海天原集团胜德塑料有限公司	1921年	初级形态塑料及合成树脂制造	制造业	上海市龙吴路4747号 （上海市	上海	31.07018782	121.45491	上海天原集团胜德	http://inst	0	
27	5da6507e2aeede1fc3db6754	上海冠生园食品有限公司	1921年	食品制造业	制造业	上海市静安区淮安路735号 （上	上海	31.24129292	121.446333	上海冠生园食品有	http://inst	0	
28	5da64eb42aeede1fc3db66f8	上海金泰工程机械有限公司	1921年	建筑工程用机械制造	制造业	上海市安亭洛浦路45号 （上海市	上海	31.30181285	121.163703	上海金泰工程机械	http://inst	0	
29	5da654a02aeede1fc3db681b	上海东方压缩机厂有限公司	1921年	气体压缩机械制造	制造业	上海市宝山区顾村工业园富联2路3	上海	31.37069722	121.401418	上海东方压缩机厂	http://inst	0	
30	5da664232aeede1fc3db6a81	上海一纺机械有限公司	1921年	纺织专用设备制造	制造业	上海市宝山区罗泾镇金石路881号	上海	31.46296312	121.356773	上海一纺机械有限	http://inst	0	

Figure 4.1 A snapshot of the original dataset

The companies recorded in the dataset are categorized by 17 "industry_type (industry types)". There are 2290 companies with missing "org_start_time (organization start time, i.e., year of registration)". For the companies with complete information, the earliest one was registered in 1848 and the latest in 2016. The following word clouds provide a general picture of the number of companies per industry type and the number of companies registered per year.



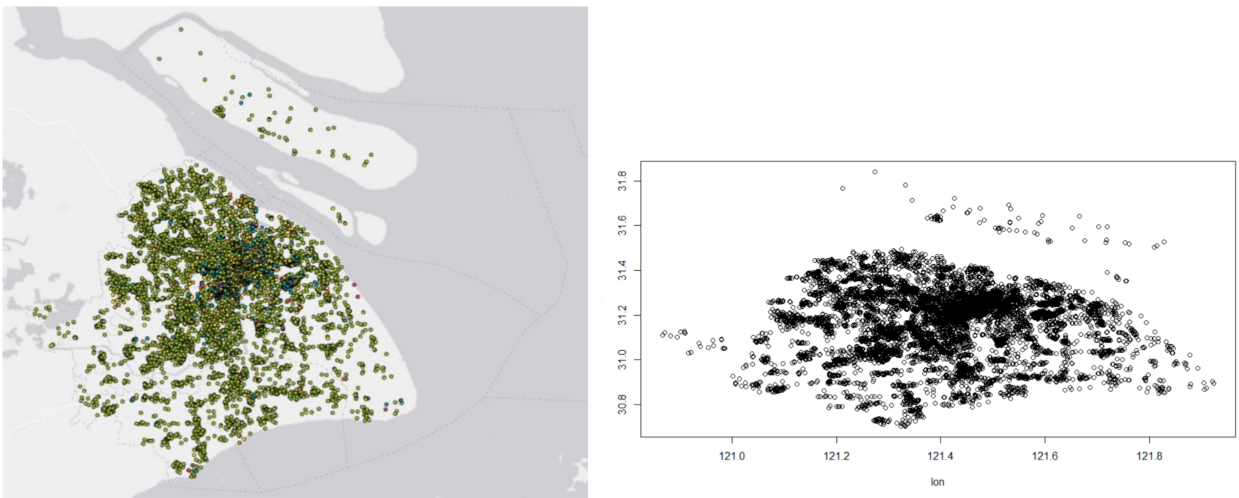
a. Word cloud for industry types

b. World cloud for years of registration

Figure 4.2 Word clouds for important semantic information

It is worth pointing out that there are only three companies in the dataset that were registered in 2016. Considering the rapid economic development of a metropolis in the 21st century, it is logically incorrect. One possible reason for this anomaly is that by the time the data set was made available, only three companies registered in 2016 had been recorded.

Figure 4.3 demonstrates the positions of all companies in the dataset, plotted via ArcGIS Pro and RStudio. The colors indicate the companies' industry types. In the 168 years of registration, these companies covered nearly the entire Shanghai, and were distributed the densest in the inner city (~121.4°-121.6° N, ~31.2°-31.4° E). Additionally, it can be seen that the latitude range and longitude range of Shanghai are both about one degree. It is therefore prudent to the set the number of distance intervals along the longitude and latitude equally.



a. The plot of companies (ArcGIS Pro)

b. The plot of companies (RStudio)

Figure 4.3 The plot of all companies recorded in the geodataset

An STC generated by ArcGIS Pro is created for illustrating the number of new companies per year (regardless of their industry types and other semantic attributes) and is shown in the appendix. The study area is tessellated into 10km-by-10km grids, because this grid dimension makes an STC produced fast and accurately. When a smaller dimension such as 5km-by-5km is chosen, the STC creation becomes much more time-consuming, and there are warning messages from ArcGIS Pro such as “too many temporal layer requests are sent” and “the statistical process is paused”. A larger grid dimension will remarkably speed up the STC creation and avoid these warning messages. In addition, the 10km-by-10km grid dimension covers the geographic surfaces nearly seamlessly, as seen from **Figure 0.3**. When a larger dimension such as 100km-by-100km is selected, the surface of Shanghai will no longer be precisely covered, as seen in **Figure 0.4**. The space-time bins are colored to demonstrate different counts, and the time step interval is set as one year.

Other columns that contain significant semantic attributes of the companies are org_name (organization names, i.e., names of companies), org_type (organization types, i.e., the secondary category of industry types. For instance, "medical equipment manufacturing", "food manufacturing", "textile manufacturing" as organization types are all secondary categories of "manufacturing", an industry type), location (address of companies), org_city (organization city, which is Shanghai for this case study), firm_info (firm information, i.e., a list of a company’s detailed information, including those in the aforementioned columns, along with products, ranking, contacted tails, etc.), and the number of patents the companies receive annually since 1986.

The number of organization types per industry type is listed in the following screenshot of a CSV file generated via RStudio. All companies belong to 649 organization types (org_types) and, the ones with complete information belong to 617. In addition, all industry types are manually translated into English before the file is written.

	A	B
1	industry_type	org_type
2	Mining	6
3	Electricity, Heat, Gas and Water Supply	34
4	Real Estate	211
5	Architecture	254
6	Transportation, Storage and Mailing	74
7	Education	12
8	Finance	76
9	Residential Services, Maintenance and Others	13
10	Scientific Research and Technical Services	1868
11	Agriculture, Forestry, Farming and Fishing	28
12	Wholesale and Retail	962
13	Health and Social Work	20
14	Culture, Sport and Entertainment	58
15	Communication, Software and IT	1300
16	Manufacturing	9360
17	Accommodation and Catering	12
18	Rental and Commercial Services	222

Figure 4.4 Number of organization types per industry type

4.3 Important historical events and governing policies

First of all, the rows of companies without year of registration are removed from the dataset to exclude noise. Afterwards, the remaining dataset was segregated according to the time line (created via Miro) in **Figure 4.5**. The length of the time periods is kept as 20-30 years, except the first one (1848-1919). In each time period, there is at least one historical event that influenced the spatial and semantic distributions of the then-registered companies in Shanghai.

The fundamental reason for selecting these events is their crucial influences on the political status, social status and economic status of China and therefore Shanghai. For instance, the signing of "Treaty of Nanking" exacerbated the colonial situation in China since the Second Opium War, triggered the massive industrialization in China and transformed Shanghai from an ordinary fishing town to a treaty port (Shanghai History Museum, Shanghai Revolution Museum, 2021). After over a century, the foundation of the Peoples republic of China (PRC) brought long-term peace, stability and unity to the national development, and Shanghai was undoubtedly benefited from it. Companies' growth was significantly boosted, and within one year such as 1958, there were 159 new companies registered in Shanghai, which was completely new for the local economic history. The detailed influence of other historical events, positive or negative, is reflected in section 4.4 of this chapter.

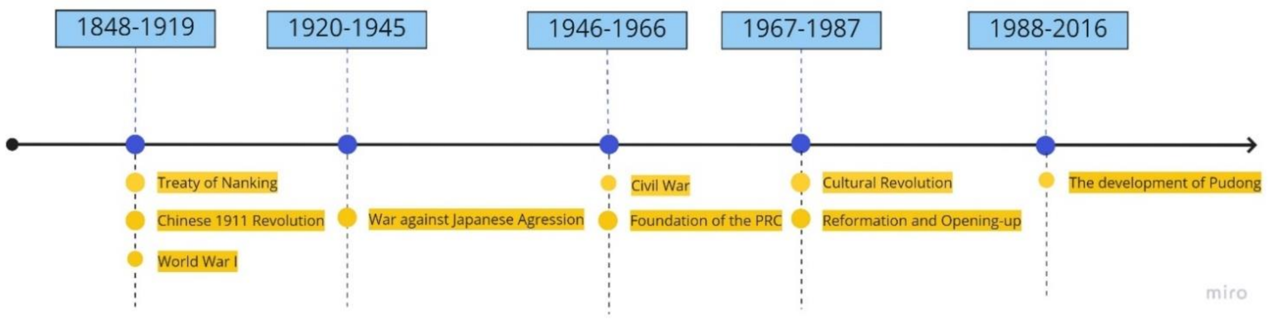


Figure 4.5 Data segregation

As such, the conflict of the story map made for this thesis is altered as "how Shanghai responded to all historical events indicated in the timeline, and thus altered the spatio-temporal and semantic distribution of companies".

4.4 Spatio-temporal stories

ArcGIS StoryMaps is the selected storytelling platform for presenting all visual outcomes, because it was introduced during a lecture of the previous semesters as a multi-functional, place-based storytelling platform (Caroll, 2021). The story map consists of three main sections: a brief introduction of the economic development in Shanghai and the general information of the test dataset; the evolution of geospatial locations and industry types (the semantic attribute this these mainly investigates on) of the companies in 2D visualizations; the 3D summaries of the distributions that are described in the previous section.

The whole story map is linked in the appendix, and its design principles will be discussed in chapter five in more details. The following subsections narrate three spatio-temporal stories included in the story map in a briefer way. The descriptions of historical events are based on the exhibition annotation in Shanghai History Museum, Shanghai Revolution Museum (2021) and Encyclopedia.com (2021), It is worth mentioning that all 2D maps were made in ArcGIS Pro and imported to ArcGIS StoryMaps, and all corresponding 3D spatio-temporal prototypes are created in RStudio. The symbol size, color coding, viewing perspectives and other properties of the prototypes will be discussed in chapter five.

The protagonists in this story are all companies with complete information in the test dataset. The antagonists are the historical events that hindered the growth of companies in terms of quantity or diversity. Other supporting characters are the historical events that fostered the growth of companies in both quantity and diversity.

The main takeaway of this story is: From the late 19th century to the early 2010s, Shanghai embarked on a distinct development path from a traditional Chinese cultural town to one of the southeast China’s most renowned economic centers (encyclopedia.com, 2021). Throughout time, significant historical events played indispensable roles in determining the numbers, shaping the geospatial distributions, influencing the industry types and organization types of companies in Shanghai. The study of temporal change of geospatial and semantic distributions would benefit decision makers for future business investments, urban planning, etc. (Sapena et. al., 2021) Meanwhile, the story map serves as a starting point for history and geography lovers to learn about economic development in Shanghai (similar to "economic development in Shanghai 101").

4.4.1 Companies vs. Historical backgrounds

After signing the Nanking Treaty in 1843, Shanghai was opened as a port by the Qing Dynasty in China, and began its history of colonization. The first recorded company in Shanghai was registered in 1848, and there were 11companies registered up until 1911.

Apart from four outliers, most companies are located in the inner city of Shanghai, and inside the border of the old foreign settlements (**Figure 4.6** Growth of companies up to 1911). According to Chinese modern history, foreign settlements were the concentrations of modern production mode, capital and advanced technologies that dominated the economic development of ports in China. Modern factories invested by foreigners were opened and industrial workers were trained in these settlements. Meanwhile, new ideology, knowledge and culture were disseminated also in these settlements. This includes “company”, a modern business association dealing with trading goods and services (Cambridge Dictionary, 2021). As such, the inner city of Shanghai, where the foreign settlements were established at, started to become the most significant company hub in Shanghai.

a. Companies 1848-1919 in 2D **b.** Foreign Settlements

Figure 4.6 Growth of companies up to 1911

The 1911 Revolution (also known as Xinhai Revolution) over threw the long-standing feudalism in China and began the era of republic. It contributed to the expansion of the foreign settlements in Shanghai, and thus generated new company hubs such as Jiading, Qingpu and Fengxian.

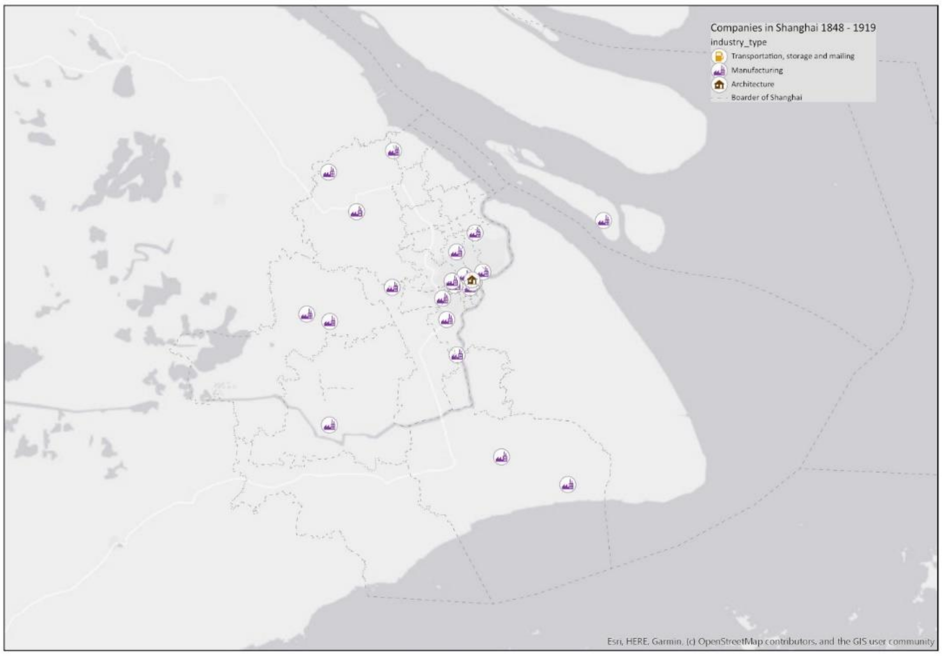


Figure 4.7 Companies registered up to 1919

The space-time micro-landscapes in **Figure 4.8** and **Figure 4.9** are the aggregated outcomes for the time period 1848-1919. The dominance of the inner city, the northwest and northeast suburbs as company hubs is emphasized by spatial data aggregation.

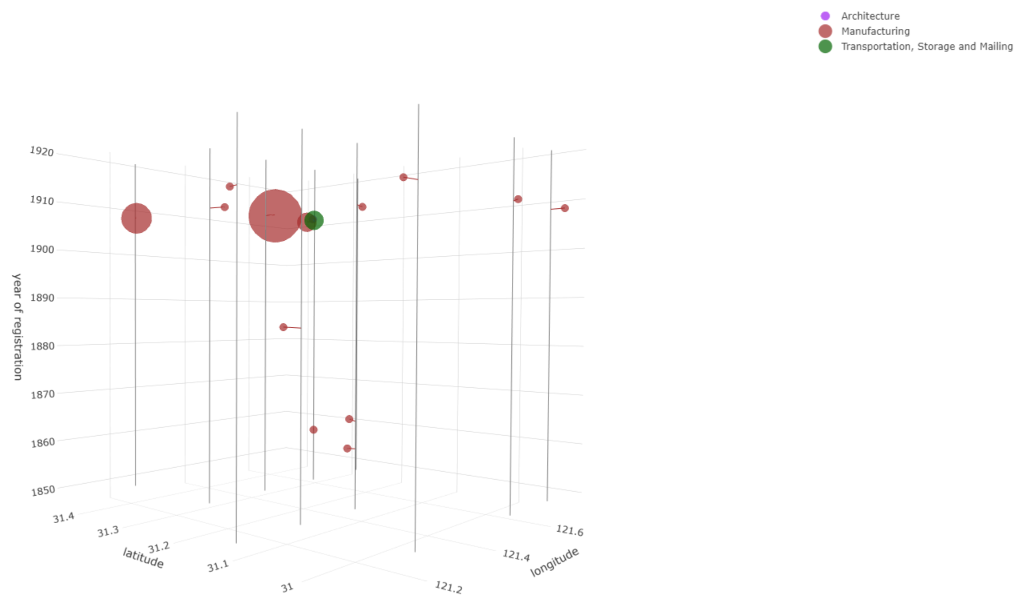


Figure 4.8 Space-time micro landscapes 1848-1919 (grids)

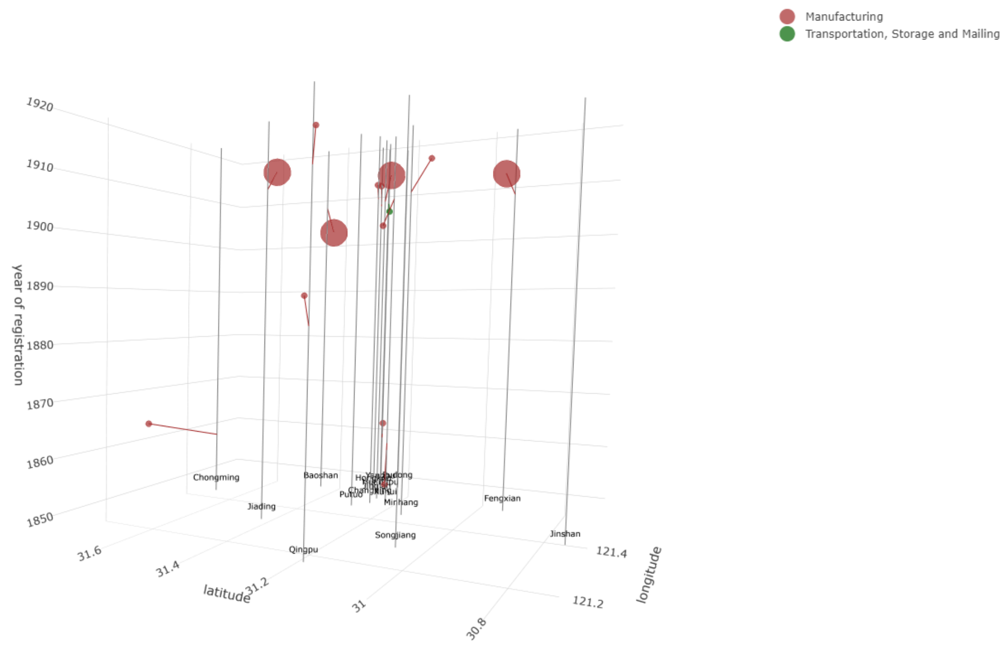


Figure 4.9 Space-time micro landscapes 1848-1919 (administrative units)

For **Figure 4.9**, "districts" were extracted from the "location" column of the test dataset as "administrative units".

The second event that fostered the economic development in Shanghai is WWI (1914-1918). Chinese laborers enhanced their practical skills on the battle fields, which led to a significant increase of companies in the city center, Jiading, Qingpu and Fengxian. Meanwhile, Pudong, the massive land in the east of Huangpu River started to become a new company hub for manufacturing. The swipe block in **Figure 4.10** demonstrates the massive company and urban development in Shanghai up to 1936.

New companies at this time were founded impressively in terms of numbers, but not yet in terms of industry types. Manufacturing was the most dominant industry type and its corresponding companies began to spread all over the territory of Shanghai. In contrast, the corresponding companies of "transportation, storage and mailing", "architecture", "wholesale and retail" and "finance" were only registered in limited numbers, and were concentrated in the inner city.

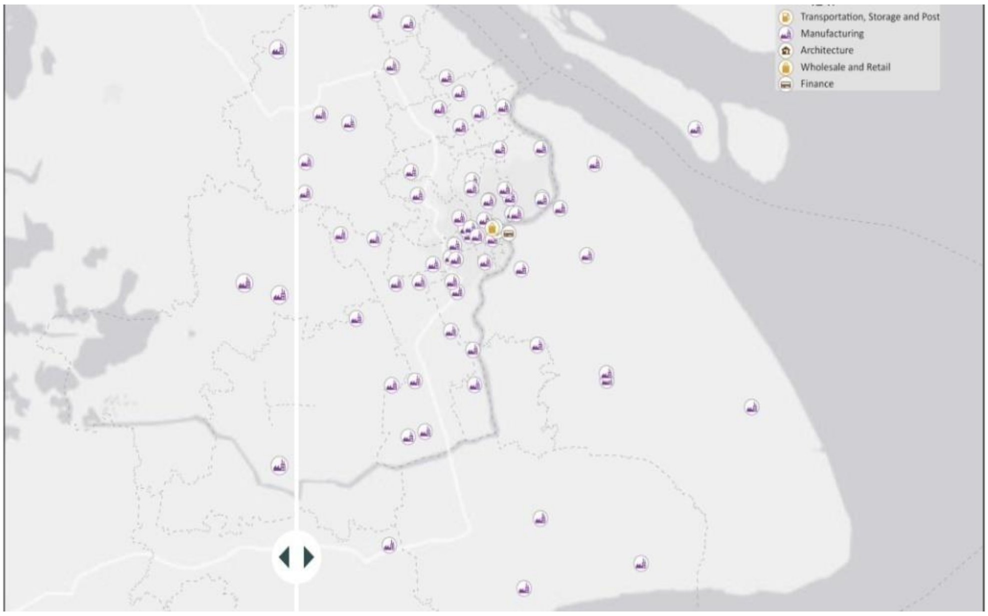


Figure 4.10 The swipe for 1848-1936 (left to right: 1848-1919, right to left: 1848-1936)

The space-time micro landscapes in **Figure 4.11** and **Figure 4.12** illustrate the spatial aggregated results. It can be seen that the 1910s and 1930s are when most companies were registered. Besides, the inner city and northwest out skirts are popular company hubs.

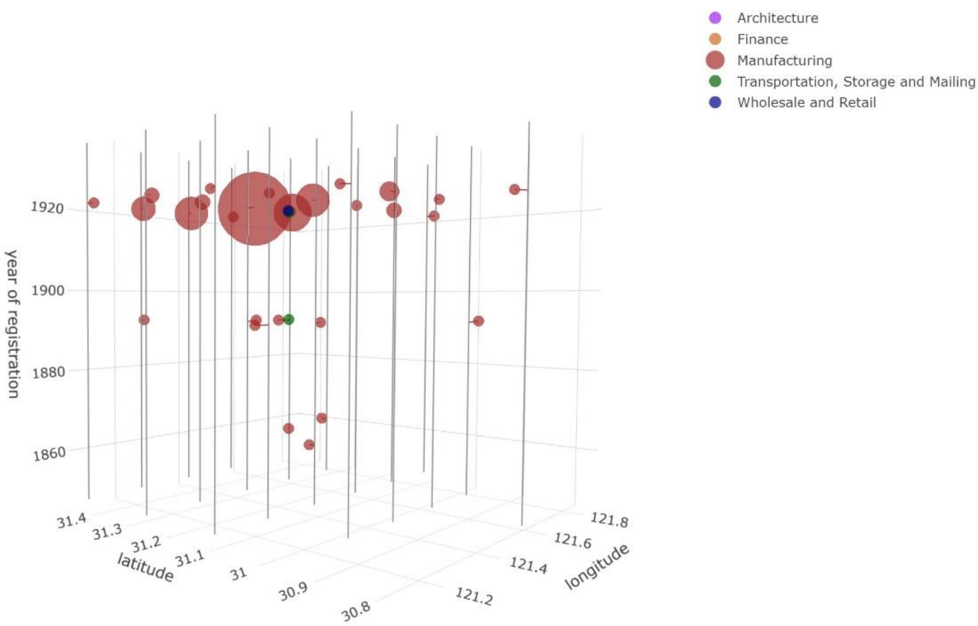


Figure 4.11 Space-time micro landscapes 1848-1936 (grids)

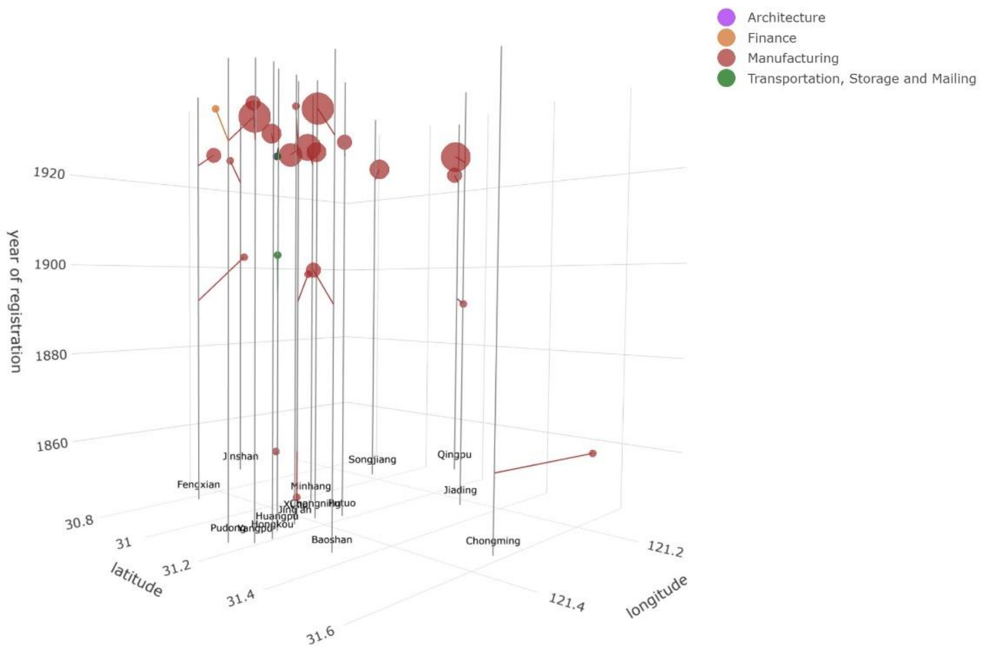


Figure 4.12 Space-time micro landscapes 1848-1936 (administrative units)

The positive effect brought by the 1911 Revolution and WWI, unfortunately, did not last long. The War against Japanese Aggression (1937-1945) caused considerable casualties in China. The economic development managed to move forward in Shanghai during the war, in order to support the national corps and maintain the function of city infrastructure. Manufacturing was the only industry type whose companies were increased, and the companies are located sparsely in Shanghai (**Figure 4.11**).

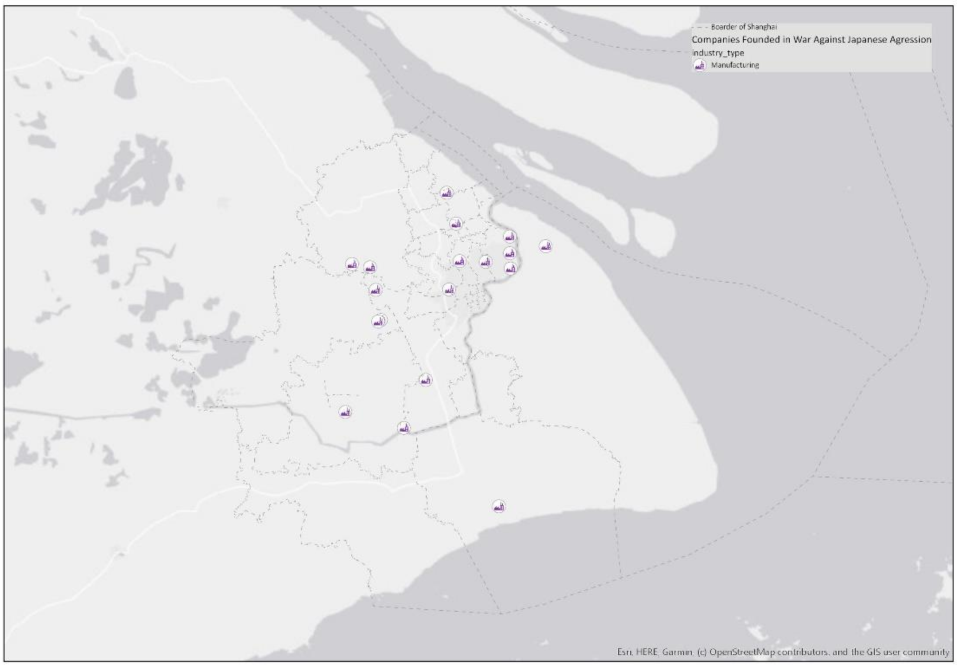


Figure 4.13 Companies in the War against Japanese Aggression

The Chinese Civil War unleashed by Kuo-Min Tang (KMT) took place immediately after the War against Japanese Aggression and lasted for four years (1946-1949). The depressing and disturbing war-induced effects continued impacting Shanghai's industrial development. The total number of companies did not increase impressively, but the economic development never ended, and even new industry types were introduced. The following swipe block illustrates the situation in the Chinese Civil War (**Figure 4.14**).

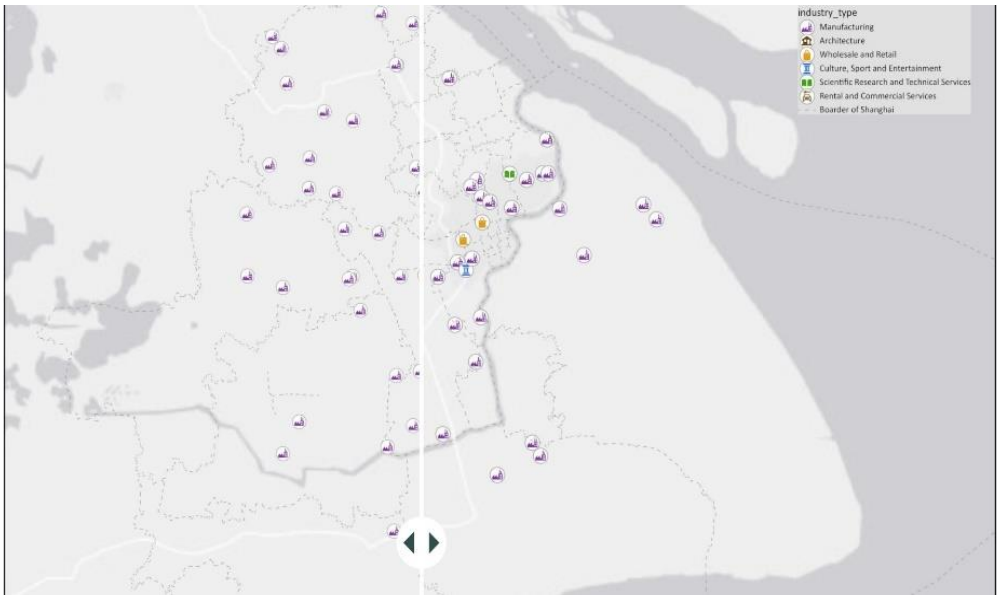


Figure 4.14 The swipe for1848-1949 (left to right: 1848-1936, right to left:1848-1949)

The space-time micro landscapes in **Figure 4.15** and **Figure 4.16** illustrate the slowed-down industrial development of Shanghai in wars in 3D. Apart from manufacturing, "rental and commercial services" and "scientific research and technical services" appeared to be new industry types. It means that third industry, particularly services started to thrive in Shanghai, apart from the industry types that ensured normal urban infrastructure, financial circulation and people's daily necessities.

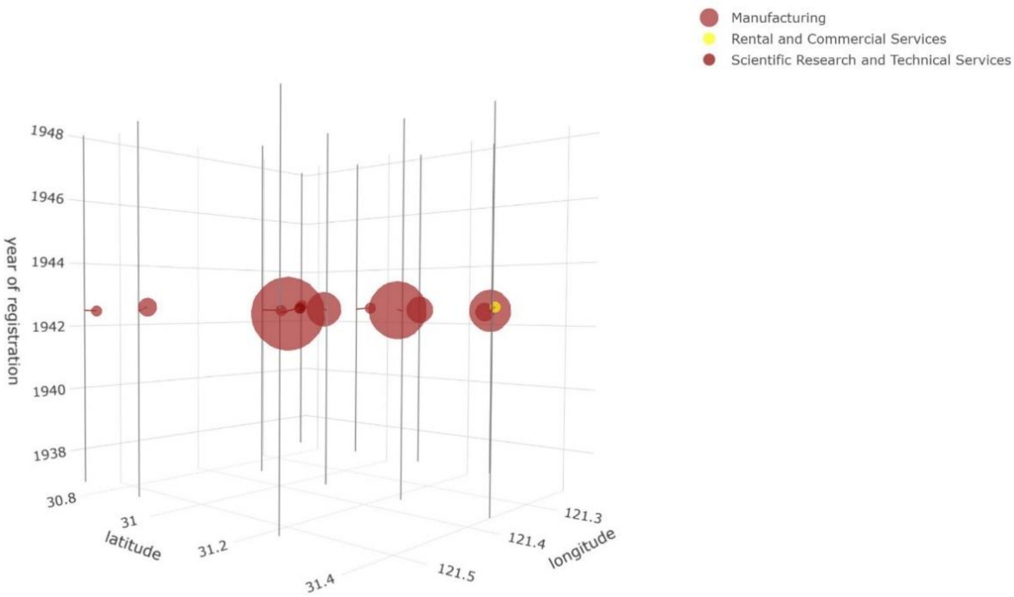


Figure 4.15 Space-time micro landscapes 1937-1948 (grids)

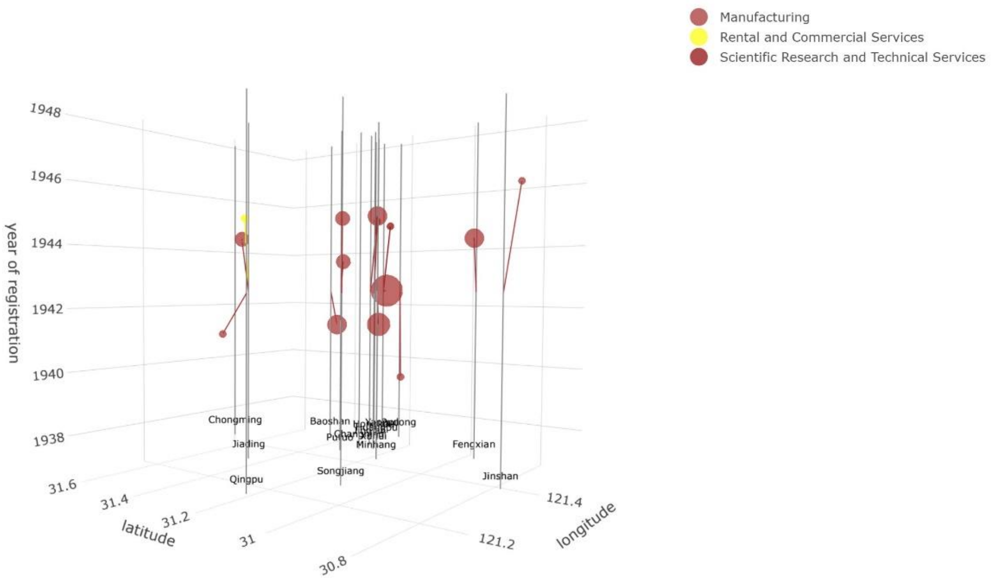


Figure 4.16 Space-time micro landscapes1937-1948 (administrative units)

"A rainbow always appears after a storm." This old saying precisely describes the situation in Shanghai when the wars finally ended and a brand-new PRC was finally founded in 1949. Within 10 years after the foundation of the PRC, the growth of companies in quantity and types, covering many more aspects of people's daily life, including culture, communication and health.

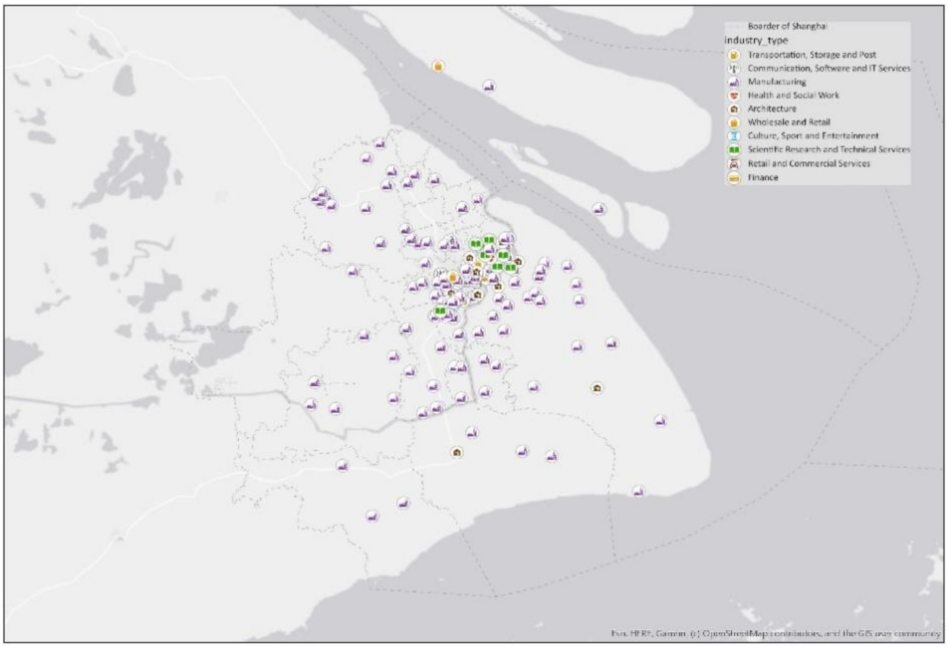


Figure 4.17 New companies from 1949 to 1959

Architecture became the second industry type whose corresponding companies started to increase in the outskirts of Shanghai. For other industry types whose companies remained increasing inside the inner city," scientific research and technical services" became the one whose companies increased the fastest. This is because after the foundation of the PRC, foreign power retreated from China with advanced science and technology. National-developed scientific research and up-to-date technology were therefore in high demand (encyclopedia.com, 2021). The companies under "scientific research and technical services" played an indispensable role in supporting the development of secondary industry in Shanghai, as well as restoring urban infrastructure.

The space-time micro landscapes in **Figure 4.18** and **Figure 4.19** illustrate this "golden era" from a 3D perspective. It can be seen that other than the inner city, the northwest outskirts remained popular in this decade. Moreover, southwest outskirts such as Jinshan and Fengxian gradually became hotspots for companies from around 1956. Pudong, over the decade, also had a good number of manufacturing companies registered.

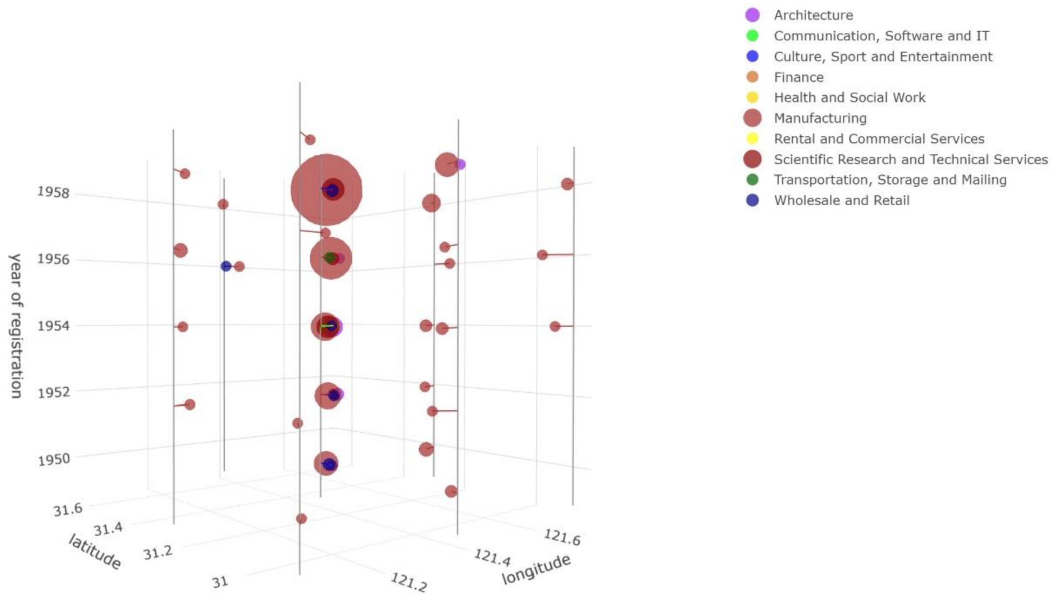


Figure 4.18 Space-time micro landscapes 1949-1959 (grids)

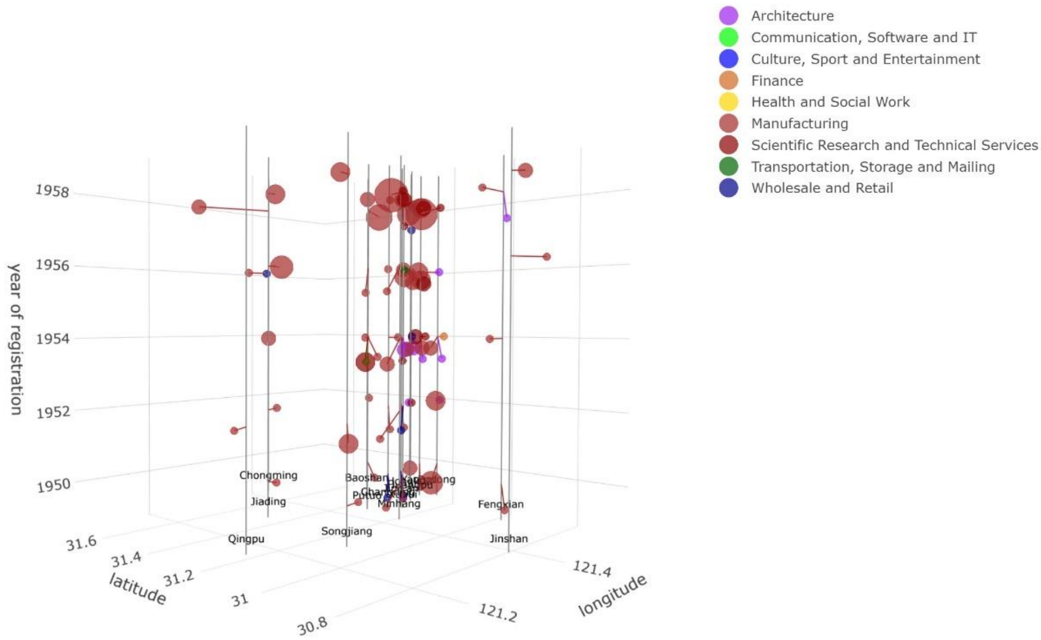


Figure 4.19 Space-time micro landscapes1949-1959

There were no longer wars in China after 1949, but the Culture Revolution, a tragic political and social chaos that took place from 1966 to 1976 crippled the business of the entirety of China, and Shanghai was no exception. Comparing to the previous time period (1949-1966), no new industry types were added. Technical equipment in many existing companies also became increasingly outdated due to the short of continued investment. Furthermore, apart from "manufacturing" being the most dominant type as always, "architecture", "scientific research and technical services" and "wholesale and retail" only saw a few corresponding companies increase.

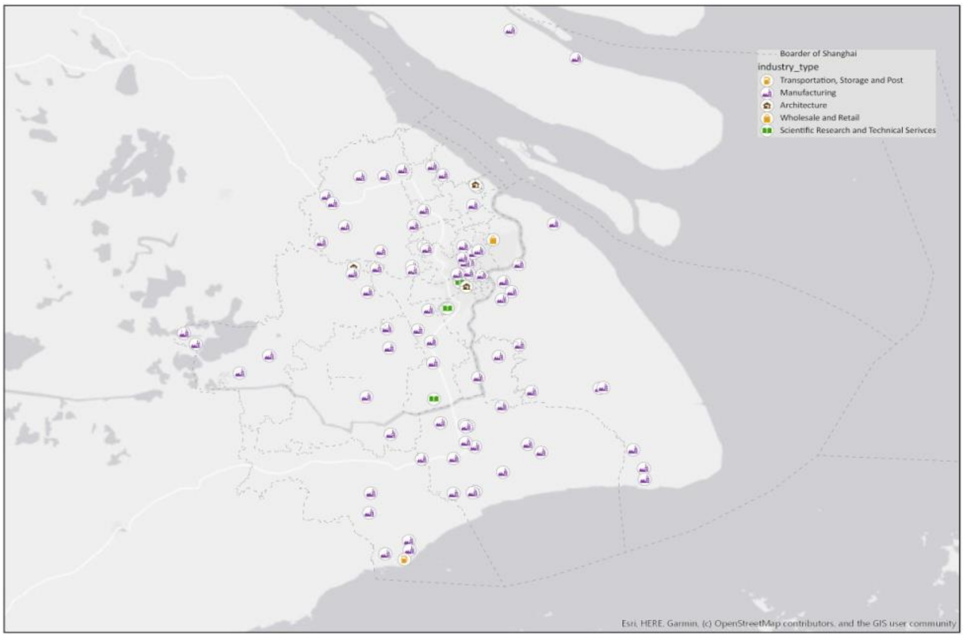


Figure 4.20 New companies in Culture Revolution

The following space-time micro landscapes emphasize the negative influence from Culture Revolution intuitively by spatial point aggregation. For example, even though manufacturing still remained dominant, the companies did not increase rapidly compared to the previous pictures, as the size of the 3D symbols did not increase obviously.

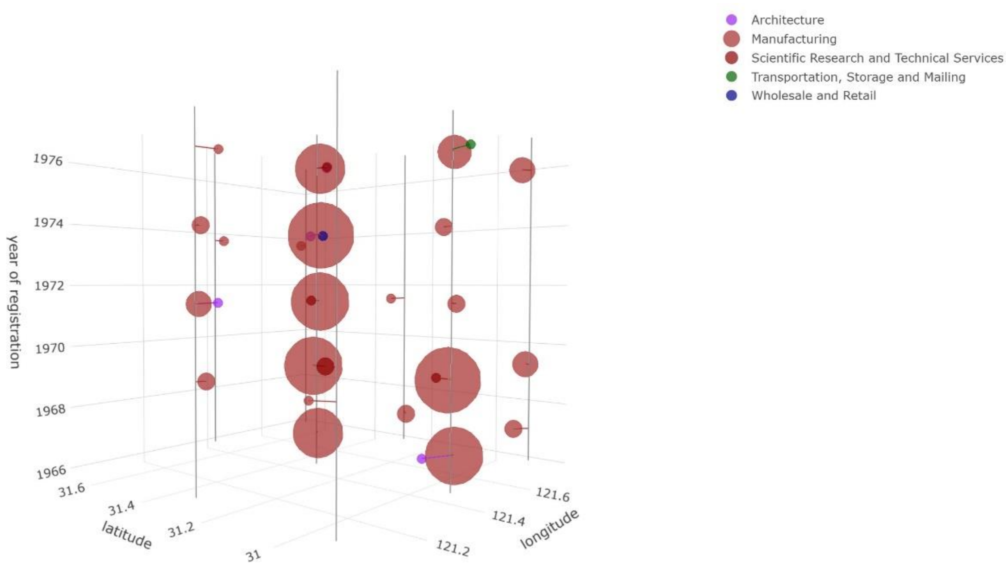


Figure 4.21 Space-time micro landscapes 1966-1976 (grids)

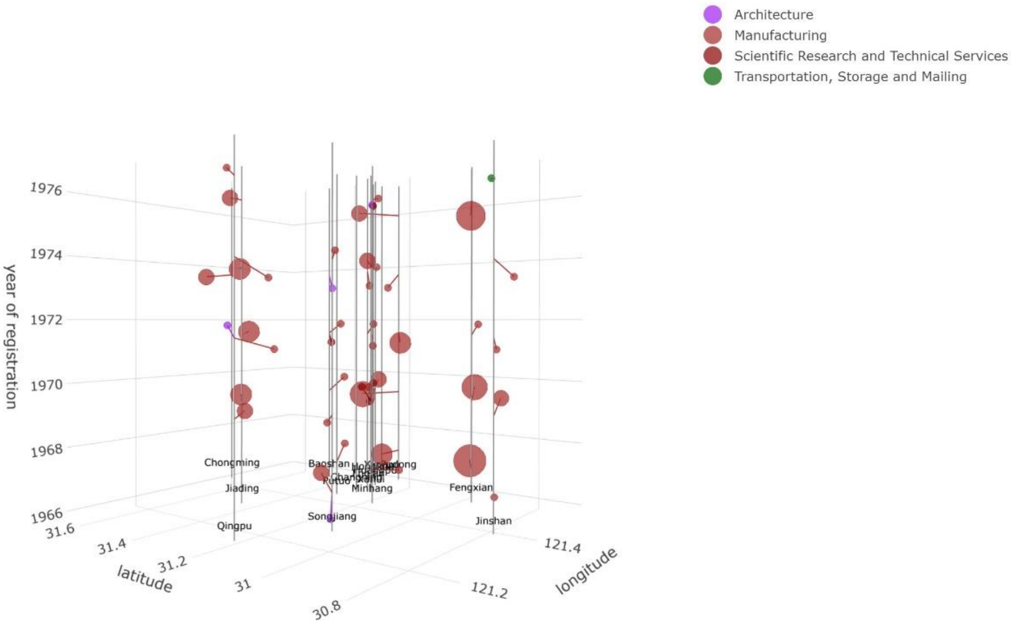


Figure 4.22 Space-time micro landscapes1966-1976 (administrative units)

On the bright side, however, the city center and nearly all outskirts were still significant company hubs in this hard time. Jiading, Baoshan, Qingpu and Fengxian performed particularly well.

4.4.2 The "virgin land" problem

The "virgin land" problem tells an interesting story by representing the spatio-temporal distributions of the first registered companies of each industry type.

Figure 4.23 convincingly illustrates the evolution or the “cultivation” of industry types in the land of Shanghai. Manufacturing was the first industry type that showed up. After a good 71-year gap, "transportation, storage and mailing" appeared as the second industry type in 1919. Afterwards, "architecture", "wholesale and retail", as well as "finance" appeared one after another, with nearly a decade’s gap between every two adjacent types. These companies, as mentioned in section 4.4.2, supplied fundamental requirements of people’s daily life and urban infrastructural engineering. "Scientific research and technical services", "culture, sport and entertainment", "electricity, heat, gas and water supply", "rental and commercial service", as well as "Communication, software and IT" appeared on almost the same temporal level from 1945 to 1960. Some of these industry types provide extra support on urban infrastructure and industrial development, while others focused on different aspects of human life. It is also noticeable that "service" started to play a decent role in the economic development to f Shanghai, thanks to the industry structural adjustment in the 1950s.

Another time period when new industry types blossomed in Shanghai was 1984-1992. "Reformation and Opening Up" was enacted as a revolutionary Chinese national policy in 1980, and Shanghai apparently benefited from it. "Accommodation and catering", "agriculture, forestry, farming and fishing", "mining", as well as "education" began to “cultivate” the land of Shanghai nearly spontaneously. This group of new industry types is a decent balance of first, second and third industries, showing appositive influence from a wise and open-minded policy to a city’s economic history.

The space-time skew (a 3D plot that projects points which are "lifted" by the temporal factor to the 2D surface via "heights", the thin vertical lines attached to the points. The heights and the data points together bring a skew-like visual effect) shown in **Figure 4.24** indicates that the inner city is the most popular "virgin land" for new industry types. From the 1990s to the start of the 21st century, Pudong became another popular choice as "virgin land" thanks to the national policy that is dedicated to its development.

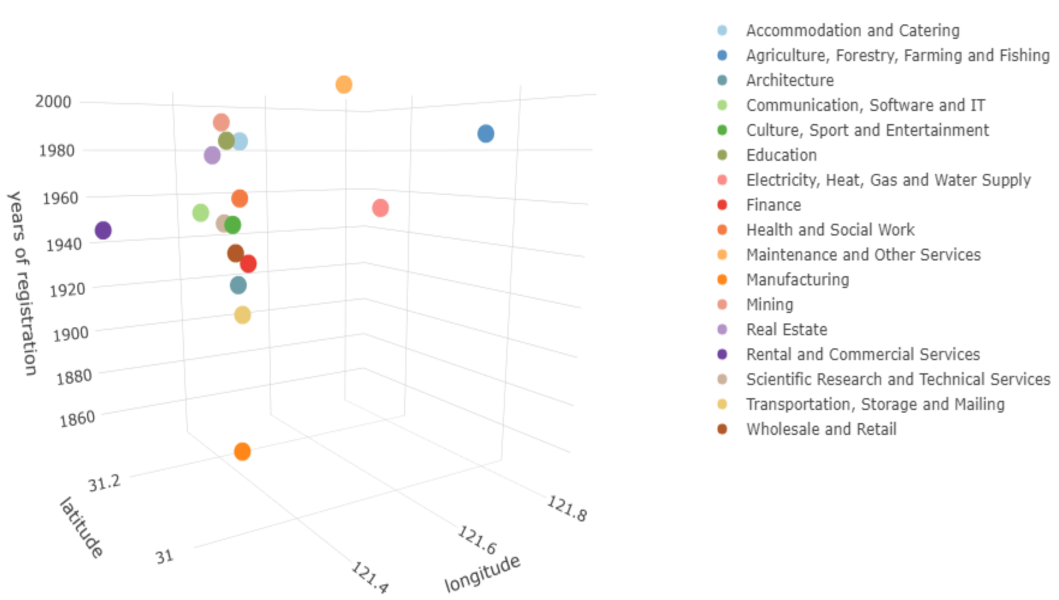


Figure 4.23 The “virgin land” problem in 3D

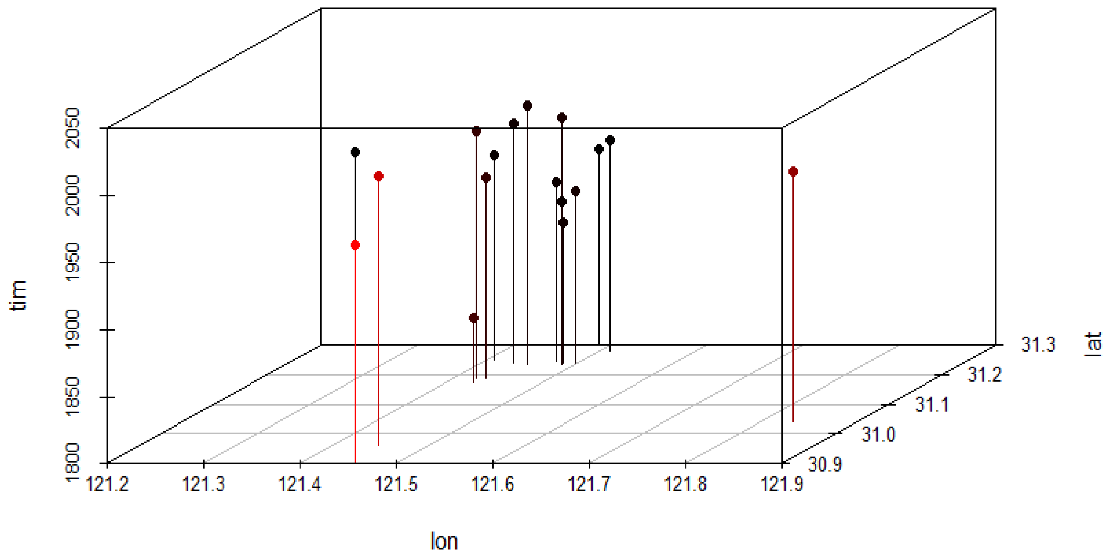


Figure 4.24 The “virgin land” problem space-time skew (projecting the points that are "lifted" by the temporal fact or onto the 2D surface, “tim” refers to “year of registration”)

The story of the "virgin land" problem can also be narrated from the perspective of "organization types". In order to minimize the effort of data translation, the 12 organization types under the industry type "transportation, storage and mailing" are applied as a simple example. Before the prototype creation, these organization types are translated manually into English.

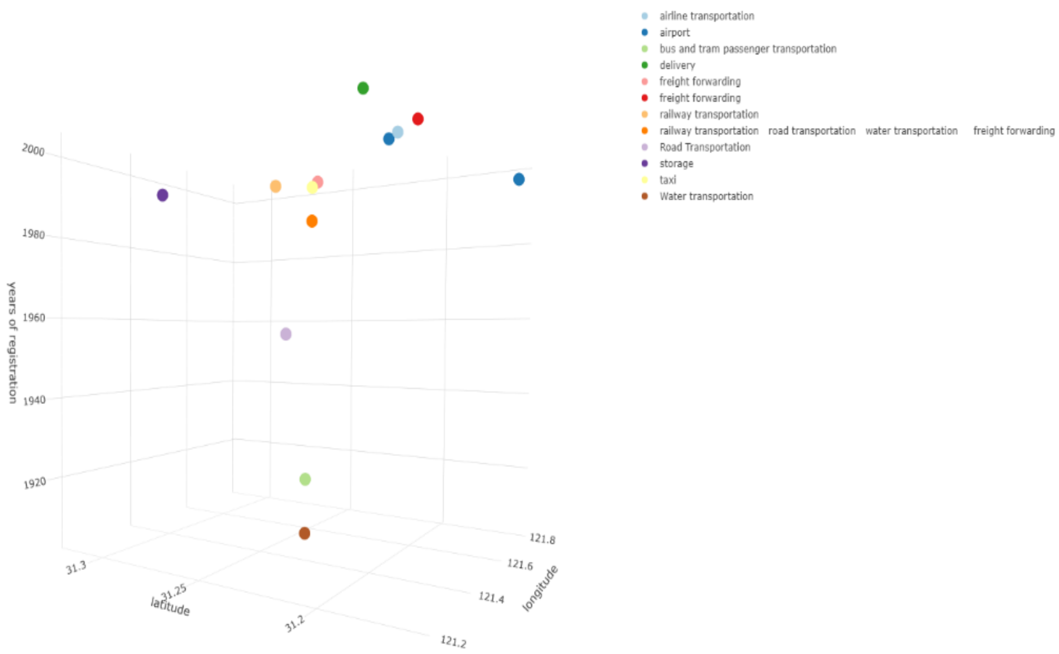


Figure 4.25 The “virgin land” problem from the perspective of “transportation, storage and mailing”

"Water transport" and "bus and tram passenger transport" are the first two organization types that started to “cultivate” the land of Shanghai, in 1919 and 1920 respectively. After an almost 40-year gap, "road transport” appeared as the third organization type, as a sign of more advanced economic development. According to Encyclopedia.com (2021), more raw materials and products were on-demand, especially for the development of heavy industry in the 1960s in Shanghai. After another good 30-year gap, "railway transport", "freight forwarder", "delivery service", "airport" and "airline transport" all showed upon nearly the same temporal level, from 1990 to 2000. Up to then, Shanghai had already become a modern metropolis, and diverse transportation, storage and logistics services definitely were in high demand.

One of the 12 companies plotted above interestingly belongs to five industry types. It was treated as a fact rather than a data anomaly in this story. Its full information is shown in the following screenshot:

5da6e7da2aeede1fc3db7a96	港中旅华贸国际物流股份有限公司	1984	Railway transport road transport water trans	交通运输、仓储和邮政业	上海市南	上海	31.2336	121.465	港中旅华贸国际物流股份有限公司 基本信息 曾用名：港中旅华贸国际货运有限公司 行业类名：铁路运输业 道路运输业 水上运输业 航空运输业 运输代理业 企业信息 成立时间：1984年 机构类型：经营性单位 企业性质：股份制	http://inst

Figure 4.26 The company belonging to five organization types (circled)

The following space-time skew indicates the location of these first-registered companies per organization type, i.e., the "virgin lands" intuitively. In the 1910s, the city center became a popular virgin land. In the late 20th century, besides the city center, a group of new virgin lands showed up. The most notable ones are the south and south-west outskirts of Shanghai and Pudong, “cultivated” by “airport”, “airplane”, “freight forwarding” and “delivery”. These companies are not able to thrive in the city center because of the massive free land they need for providing their services, and the outskirts are definitely ideal for them.

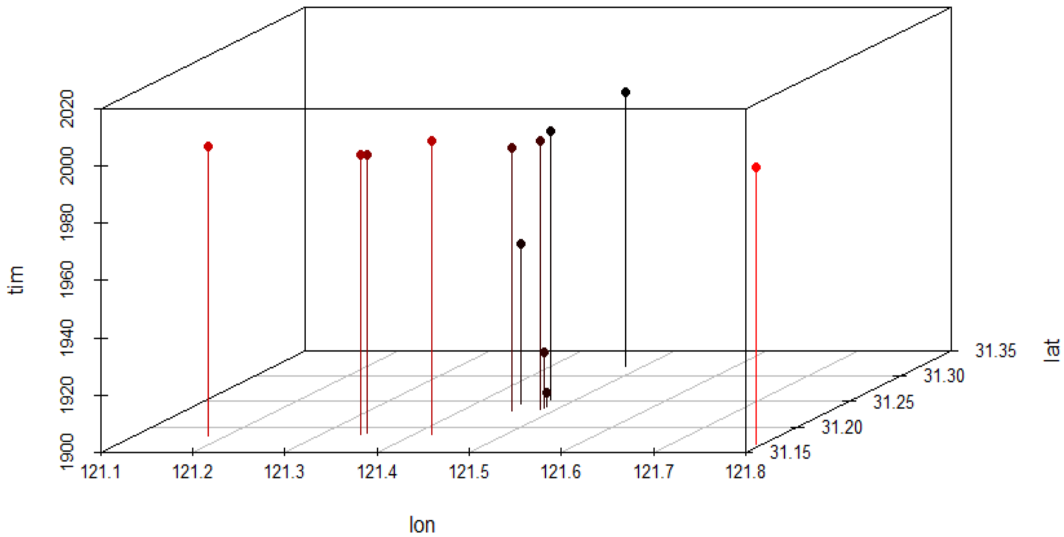


Figure 4.27 The space-time skew for "transportation, storage and mailing" (indicates the location of these points in the 2D surface; “tim” refers to “year of registration”)

4.4.3 Maximal number of features

As mentioned in chapter three, “maximum number of features” is designed for illustrating urban expansion. To create the prototype, the geographic area (Shanghai, in this case) is tessellated into fishnet grids, and the number of companies, regardless of any industry type, is counted in each space-time bin. Afterwards, only the space-time bins with the maximal number of companies each year will be illustrated. The geographic region in Shanghai with that maximal number of companies is therefore the new company hub, and indicates where the city was expanded to.

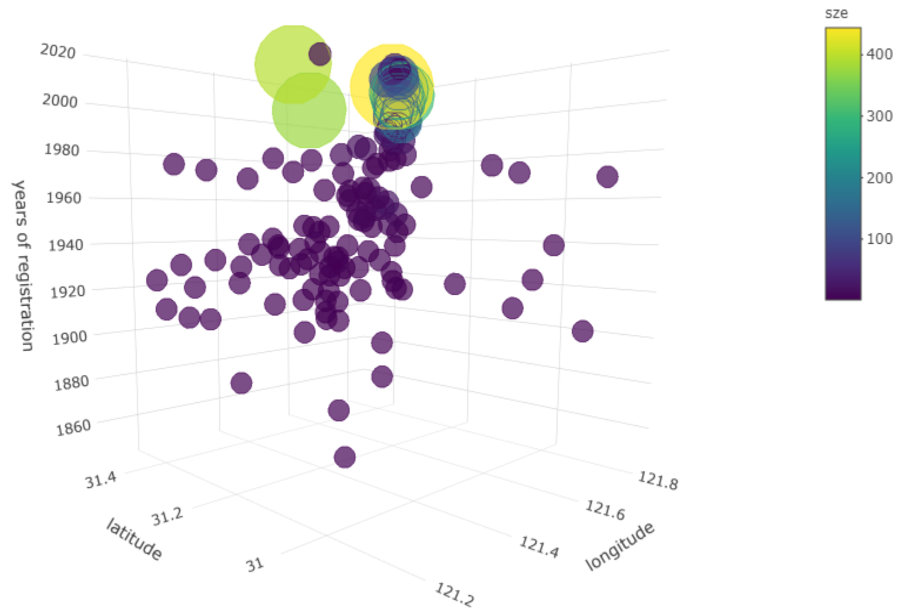


Figure 4.28 Maximal number of features in 3D

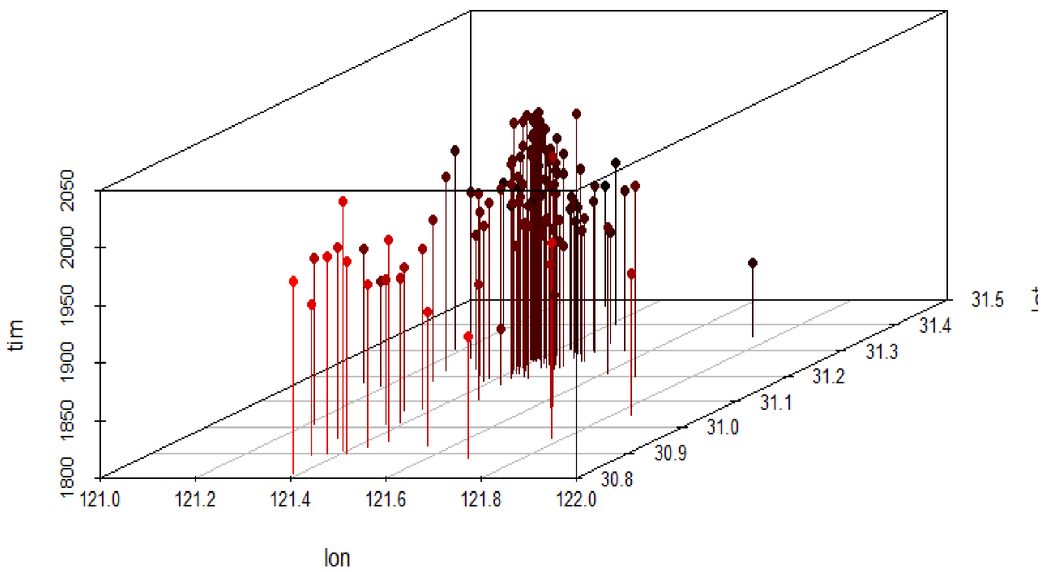


Figure 4.29 The space-time skew for "Maximal number of features" (indicates the location of these points in the 2D surface, “tim” refers to “year of registration”)

It can be detected from the 3D visualizations (**Figure 4.28** and **Figure 4.29**) that the urban expansion, or the generation of company hubs took place at first in the city center. In the course of time, it started in the north of the city, and eventually began in the south from the 1950s on. In the end, in the 1990s, the southeast outskirts (generally in the direction of Pudong) began to rise, and became another company hub. Meanwhile, no matter when and how the outer skirts started to develop, the inner city remained as the most significant company hubs. It does not suggest urban expansion, but further proves the popularity of the inner city for new companies, an important spatio-temporal pattern.

5 Discussions

Because the spatial-temporal stories in chapter four are extracted from the storymap designed for this thesis (linked in the appendix), and the prototypes not only aid the 2D visualizations, but also narrate new stories (the “virgin land” problem and the maximal value of features), this chapter summarizes general characteristics of the storymap and the prototypes. Meanwhile, it discusses the semantic meaning of some spatio-temporal stories, suggests other cases that are able to be visualized by the prototypes and lists the limitations of this thesis’ research.

5.1 The characteristics of the story map

Generally speaking, the storymap embeds a variety of geo-tagged or geo-related media, including static maps, time-aware maps, swipe blocks and videos, in order to thoroughly narrate the evolution of the spatio-temporal and semantic distribution of different companies in Shanghai within 168 years (1848-2016).

A neat and coherent visual narrative layout is designed for the story map. At the start, critical background information about Shanghai, as well as general information fall registered companies in the test geodata set is thoroughly given in linearity. Followed by it, sidecars with static maps inserted are applied to explain the spatio-temporal and semantic distributions of companies in each time period via 2D visual representations. In addition, special sections dedicated to significant historical events policies are created in the story map, in order to emphasize their strong influences on the spatio-temporal patterns of the then-registered companies. Typical examples are WWI, War against Japanese Aggression, the Foundation of the PRC and Cultural Revolution.

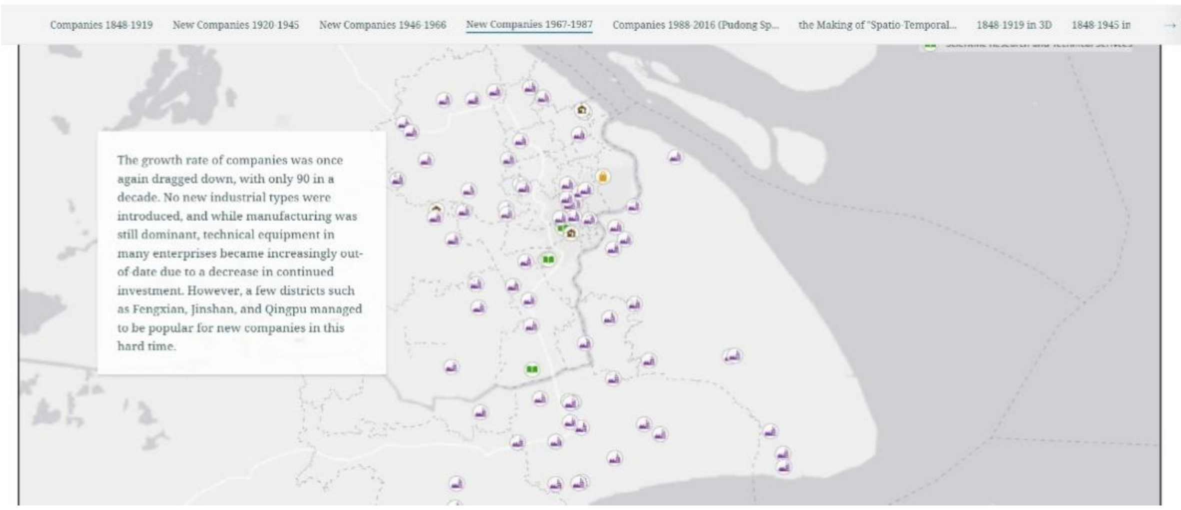


Figure 5.1 Special section in the storymap (Cultural Revolution)

The rest of the story map includes two other stories: the "virgin land problem" and "maximal value of features". In fact, the story of the “virgin land problem” about industry types, together with the space-time micro landscapes, not only reflects the industrial development in Shanghai to an extent, but also unveils some general industrial development patterns in a city. At the start, the main target of the local authority was usually to ensure the fundamental urban infrastructure, financial circulation and the residents’ daily necessities. In the course of time, the city withstood some stresses and strains (wars, economic depressions, to name a few), and the local authority would begin to transform the urban economy. “Manufacturing”, “architecture”, “transportation, storage and mailing” and “wholesale and retail”, as main stream companies, still play their part swell, and remain as essential semantic attributes of companies. However, companies dedicating to other aspects of people’s life such as “culture and sports”, “health and social work”, “communication and IT services”, “rental and commercial services”, “education” and “accommodation (hotel) and catering” began to contribute to the economics of this city. Meanwhile, specialized industry types such as" scientific research and technical services", "electricity, water, heat and gas supply" and "maintenance" that contribute to urban development and people’s daily life on a more detailed level also began to thrive. These companies apparently increased the semantic attributes of companies of a city, because people could eventually correlate “companies” with not only “urban infrastructure” and “daily necessities”, but also them.

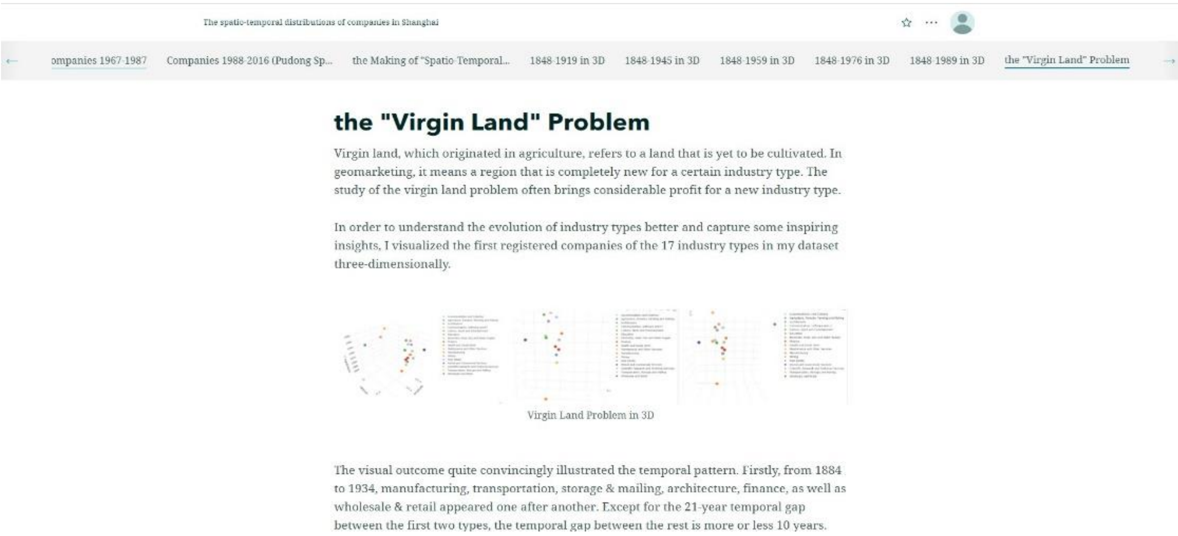
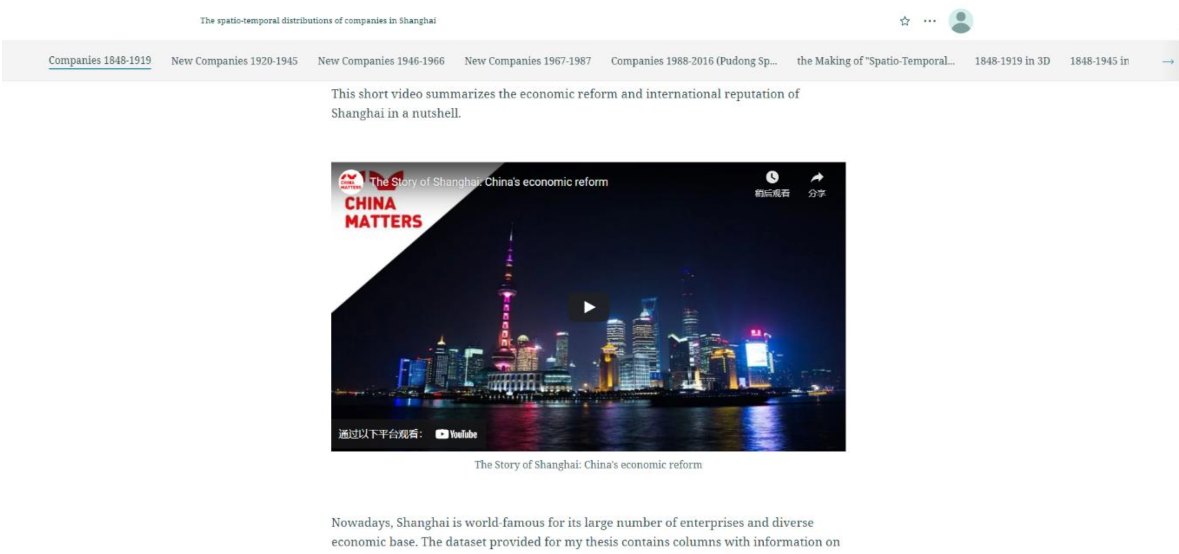


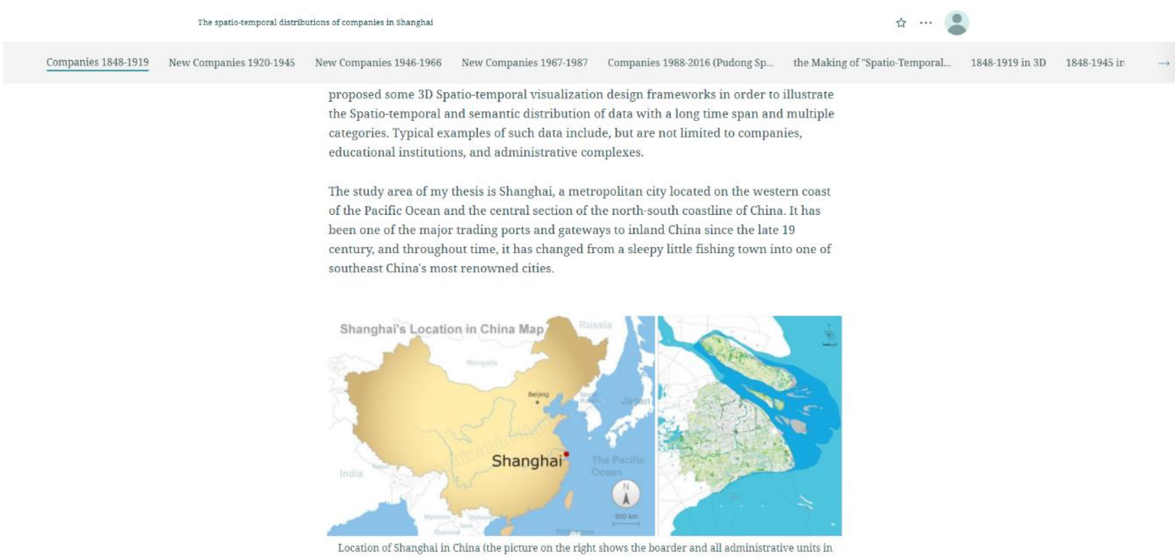
Figure 5.2 Story of the “virgin Land” problem

The "virgin land problem" story from the perspective of "transportation, storage and mailing" reflects the global transportation history and Chinese transportation history to some extent. From the late 19th century to the early 20th century, the beginning of industrialization allowed heavy vehicles including buses, trams, trains and ships, as well as massive transportation network including railway, waterway and road to develop in the cosmopolitan cities around the world, (Deutsches Museum Verkehrszentrum, 2021), and Shanghai was no exception. "Water transport", "bus and tram passenger transport" and "road transport" naturally became the first types of companies registered in Shanghai, and the typical semantic attributes that describe the methods of transportation at that time. The postal system also primarily relied on waterway and railway transports. Airplanes, however, most often served military purposes. After the 1950s, thanks to the rapid development of science and technology, the increased speed of transportation modes and the accelerated expansion of cosmopolitan cities such as Shanghai called for an adjustment in transportation behavior (Deutsches Museum Verkehrszentrum, 2021). Furthermore, the foundation of the PRC led to the reformation of civil aviation. As such, “taxi”, "airport", "airline transport", "storage" and "delivery service" not only became the newest types of companies in Shanghai, but also the up-to-date semantic attributes of the terms “transportation”, “storage” and “mailing”. In the 1920s and 1930s, it was hardly possible for people to correlate these terms with semantic attributes stemmed from automobile and airplane. This is no longer a problem in the modern society.

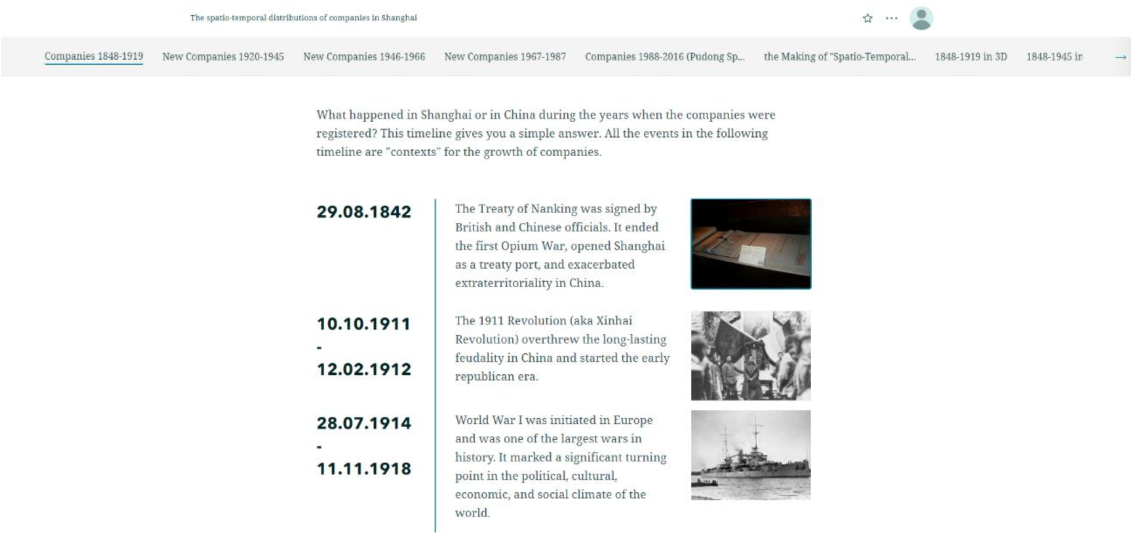
The story map also balances author-driven and reader-driven spatial narrative styles, as mentioned in chapter two. Author-driven styles are employed in the back ground information section (about the industrial development in Shanghai and the test dataset). The reader-driven styles include a time-aware map showing the spatio-temporal distributions of all companies in animation, and sidecars containing time-aware maps and swipe blocks that allow a user to grasp the "companies-events dynamic" flexibly. The 3D illustrations of companies’ distribution and the other two spatio-temporal stories are currently in author-driven styles, as there has not been discovered a way to publish the prototypes online and embed them into ArcGIS StoryMaps. ArcGIS Online is a community that allows a time-aware layer created by ArcGIS Pro uploaded to, but an equivalent community for RStudio is yet to be discovered. Videos accompanied by explanatory annotations are applied instead.



a. Author-driven 1

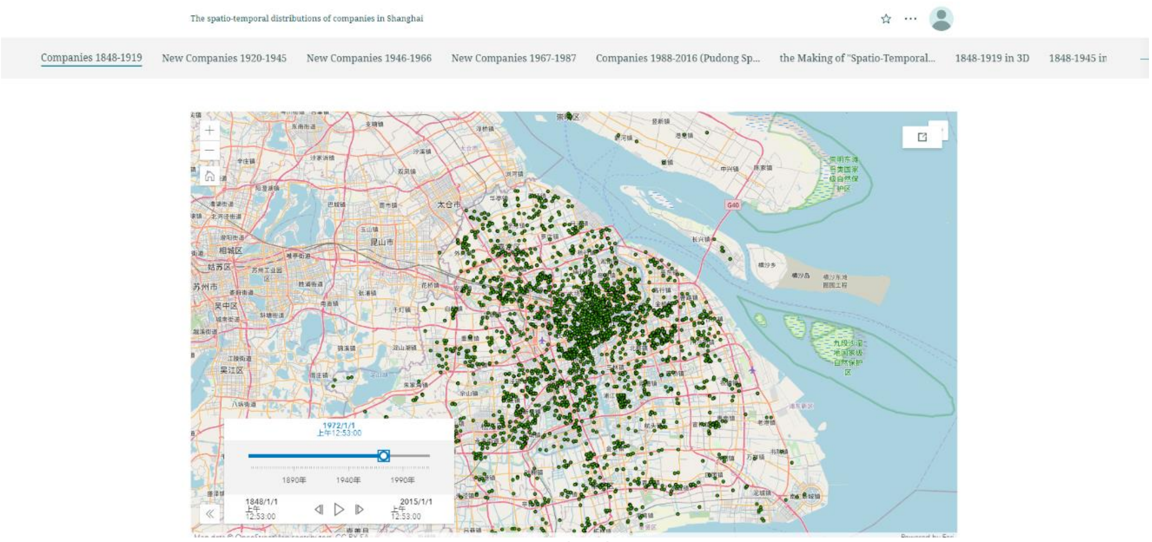


b. Author-driven 2



c. Author-driven3

Figure 5.3 Author-driven styles in the story map



a. Reader-driven 1

b. Reader-driven2

Figure 5.4 Reader-driven styles

The existing ArcGIS StoryMap tools, as demonstrated above, are already powerful enough for making insightful geovisual stories. However, in all static and time-aware maps, each company is depicted at its exact locations, and its industry type is indicated via an icon. The 3D prototypes, instead of displaying every single company's information, use aggregated symbols to summarize critical information. All 17 industry types and the number of companies per type are aggregated and visualized along the temporal ax is, to achieve a more concise visual outcome. Furthermore, in the 3D plots of the "virgin land" problem, individual companies are once again visualized. However, instead of letting their own information revealed, they are depicted to represented where the corresponding industry types started to develop in Shanghai. It is worth mentioning that "industry types" of companies is a typical example of "human-created categories" as mentioned in chapter one. Their evolution indicates the influence of historical events on human-created categories intuitively, and is emphasized by the 3D visual effects of the prototypes.

5.2 The performance of the 3D visualization prototypes

As mentioned in chapter three and chapter four, the prototypes proposed by this thesis' research are developed in RStudio via the Plotly package (CRAN, 2021). Just like the STCs designed in Kveladze, Kraak and Elzakker (2017), Bogucka and Jahnke (2018), ESRI (2021) and Mo et. al. (2020), the proto types can flexibly be zoomed, panned and rotated into different perspectives, which enable a user to navigate to the preferred temporal level for detailed observation. The symbol of the plot scan be switched on or off easily in RStudio so that a user is able to explore the spatio-temporal variation of the data under one category (one industry type in this case) easily. Unlike the STCs in the aforementioned scientific papers, a time slider is not designed for the space-time micro landscapes, but a keyboard prompt is programmed for inputting the desired time period, the number of distance intervals (when aggregation via grids is selected), the number of time step intervals and the centralization factor to easily customize the visual outcome. A method of directly importing a basemap into the prototypes is yet to be developed; however, to create a clearer geo-reference for the prototypes, a user can either plot a "space-time skew", as demonstrated in chapter four, to project all symbols onto the geographic surface, or re-import the data of the corresponding prototype into ArcGIS Pro and make a time-aware map.

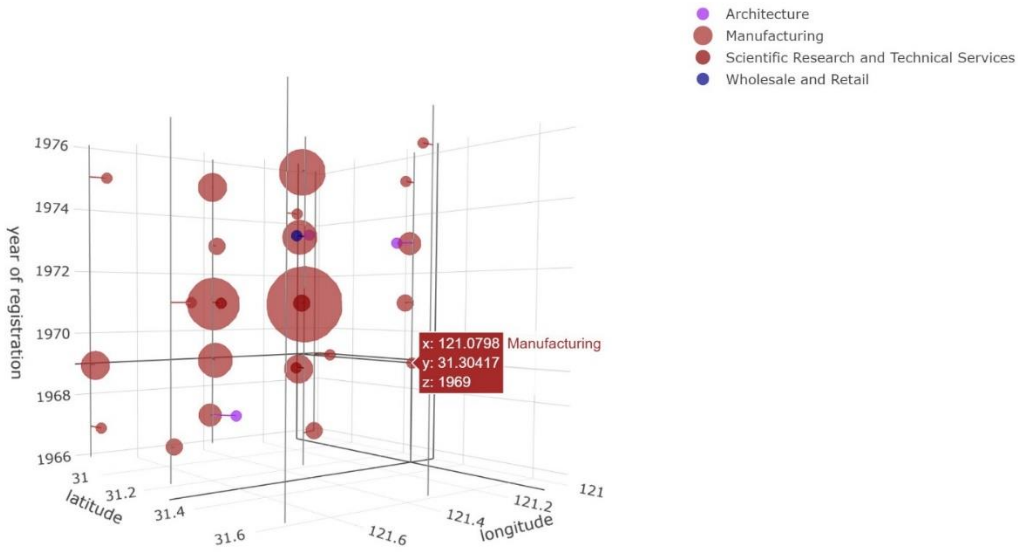


Figure 5.5 Navigating to the desired temporal level

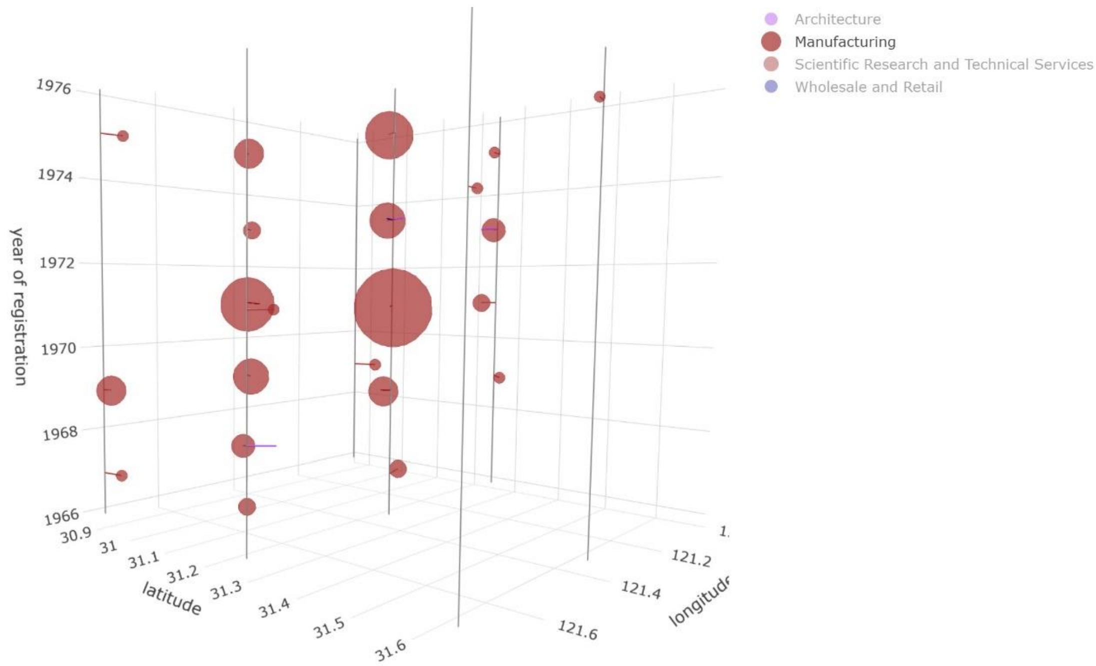


Figure 5.6 Distracting symbols can be turned off

By aggregating fundamental characteristic values such as the number of features and the cluster's centroids per category in space-time bins, the space-time micro landscapes make significant insights retrievable with less visual redundancy. It is not always feasible for traditional 2D mapping. Additionally, the layered structure of these prototypes demonstrates temporal change straight forwardly. The concentrated regions of companies per industry type are represented in every layer. Furthermore, the major difference between the space-time bins in the space-time micro landscapes and the ones in a traditional STC for temporal geostatistics (**Figure 0.1**-**Figure 0.4**) is that the former has an aggregated 3D symbol right at the center, indicating multiple categories, whereas the latter is filled with only one color shade indicating one aggregated feature value (count, sum, average, etc.).

The sizes of the proportional symbols are plotted according to the number of features per category in the specific time period selected. For example, there are in total 23 companies under only three types in the test dataset (see the last columns of **Figure 5.7**) that were registered in the time period 1848-1919, and 20 are under “manufacturing”. Therefore, in **Figure 4.9**a and **Figure 4.9**b, the proportional symbol of manufacturing is the biggest. It is the reason why a symbol already looks enormous in an early time period (before the WWII). The size of the symbols is decided by the same theory for any other time period chosen: the number of corresponding features under each category.

	X_id	org_name	org_start_time	org_type	industry_type
1	5da653752aeede1fc3db67e5	上海老凤祥有限公司	1848	珠宝首饰及有关物品制造	Manufacturing
2	5da653ce2aeede1fc3db67f7	上海周虎臣曹素功笔墨有限公司	1864	笔的制造	Manufacturing
3	5da64f1e2aeede1fc3db670a	江南造船(集团)有限责任公司	1865	船舶及相关装置制造	Manufacturing
4	5da5d7102aeede1fc3db55fa	上海信谊万象药业股份有限公司	1890	医药制造业	Manufacturing
5	5da657952aeede1fc3db68a6	上海东湖机械厂	1897	通用设备制造业 专用设备制造业	Manufacturing
6	5da68e762aeede1fc3db6f9e	上海大颛仪器设备有限公司	1900	仪器仪表制造业	Manufacturing
7	5da5e9302aeede1fc3db58de	上海盛伟薯科科技股份有限公司	1904	气体压缩机械制造	Manufacturing
8	5da6e7612aeede1fc3db7a7c	中交上海航道局有限公司	1905	水上运输业	Transportation, Storage and Mailing
9	5da5d8f22aeede1fc3db5657	上海武马食品科技有限公司	1906	食品制造业	Manufacturing
10	5da653b62aeede1fc3db67f2	上海铁路通信有限公司	1906	通信设备制造	Manufacturing
11	5da64f5d2aeede1fc3db6717	上海中华药业有限公司	1911	医药制造业	Manufacturing
12	5da5d9342aeede1fc3db5665	上海华宝孔雀香精香料有限公司	1912	香料、香精制造	Manufacturing
13	5da6a6112aeede1fc3db72cf	上海皮尔博格有色零部件有限公司	1912	有色金属铸造	Manufacturing
14	5da6503b2aeede1fc3db6746	上海减速机械厂有限公司	1913	齿轮及齿轮减、变速箱制造	Manufacturing
15	5da6539d2aeede1fc3db67ed	上海电器股份有限公司人民电器厂	1914	电气机械和器材制造业	Manufacturing
16	5da652bc2aeede1fc3db67c5	上海开林造漆厂	1915	涂料制造	Manufacturing
17	5da5d98f2aeede1fc3db5675	冠生园(集团)有限公司	1915	食品制造业	Manufacturing
18	5da64ebd2aeede1fc3db66fa	上海华线医用核子仪器有限公司	1916	医疗诊断、监护及治疗设备制造	Manufacturing
19	5da6510d2aeede1fc3db6770	上海厦牌衡器有限公司	1917	衡器制造	Manufacturing
20	5da69aee2aeede1fc3db7167	上海华通开关厂有限公司	1919	配电开关控制设备制造	Manufacturing
21	5da6e6a52aeede1fc3db7a62	上海强生集团有限公司	1919	公共电汽车客运	Transportation, Storage and Mailing
22	5da653d82aeede1fc3db67f9	上海实业马利画材有限公司	1919	颜料制造	Manufacturing
23	5da6bb662aeede1fc3db75bd	上海市基础工程集团有限公司	1919	房屋建筑业 土木工程建筑业	Architecture

Figure 5.7 Manufacturing dominating the industry types

The colors of the symbols are selected manually from the R color palette. **Figure 5.8** A part of the R color palette shows only around 210 out of 657 options in the palette:

```
> colors()
  [1] "white"                 "aliceblue"             "antiquewhite"          "antiquewhite1"
  [5] "antiquewhite2"         "antiquewhite3"         "antiquewhite4"         "aquamarine"
  [9] "aquamarine1"           "aquamarine2"           "aquamarine3"           "aquamarine4"
 [13] "azure"                 "azure1"                "azure2"                "azure3"
 [17] "azure4"                "beige"                 "bisque"                "bisque1"
 [21] "bisque2"               "bisque3"               "bisque4"               "black"
 [25] "blanchedalmond"        "blue"                  "blue1"                 "blue2"
 [29] "blue3"                 "blue4"                 "blueviolet"            "brown"
 [33] "brown1"                "brown2"                "brown3"                "brown4"
 [37] "burlywood"             "burlywood1"            "burlywood2"            "burlywood3"
 [41] "burlywood4"            "cadetblue"             "cadetblue1"            "cadetblue2"
 [45] "cadetblue3"            "cadetblue4"            "chartreuse"            "chartreuse1"
 [49] "chartreuse2"           "chartreuse3"           "chartreuse4"           "chocolate"
 [53] "chocolate1"            "chocolate2"            "chocolate3"            "chocolate4"
 [57] "coral"                 "coral1"                "coral2"                "coral3"
 [61] "coral4"                "cornflowerblue"        "cornsilk"              "cornsilk1"
 [65] "cornsilk2"             "cornsilk3"             "cornsilk4"             "cyan"
 [69] "cyan1"                 "cyan2"                 "cyan3"                 "cyan4"
 [73] "darkblue"              "darkcyan"              "darkgoldenrod"         "darkgoldenrod1"
 [77] "darkgoldenrod2"        "darkgoldenrod3"        "darkgoldenrod4"        "darkgray"
 [81] "darkgreen"             "darkgrey"              "darkkhaki"             "darkmagenta"
 [85] "darkolivegreen"        "darkolivegreen1"       "darkolivegreen2"       "darkolivegreen3"
 [89] "darkolivegreen4"       "darkorange"            "darkorange1"           "darkorange2"
 [93] "darkorange3"           "darkorange4"           "darkorchid"            "darkorchid1"
 [97] "darkorchid2"           "darkorchid3"           "darkorchid4"           "darkred"
```

Figure 5.8 A part of the R color palette

The 17 colors chosen from the palette to represent the industry types are listed the following table:

Industry type	**Color**
Accommodation and Catering	Red
Agriculture, Forestry, Farming and Fishing	Orange
Architecture	Purple
Communication, Software and IT	Green
Culture, Sport and Entertainment	Blue
Education	Violet
Electricity, Heat, Gas and Water Supply	Black
Finance	Chocolate
Health and Social Work	Gold
Maintenance and Other Services	Pink
Manufacturing	Brown
Mining	gray
Real Estate	wheat
Rental and Commercial Services	yellow
Scientific Research and Technical Services	Dark red
Transportation, Storage and Mailing	Dark green
Wholesale and Retail	Dark blue

Table 5.1 The color coding for the space-time micro landscapes

The similarity of some colors, such as gold and yellow, chocolate and brown, blue and dark blue, can be seen from the pictures cited in chapter three and chapter four. Additionally, the pale tone of some other colors such as gray and wheat reduce the visibility of the corresponding symbols to some degrees, especially when symbols with stronger color tones appear simultaneously. Therefore, a better color coding remains a future task.

The prototypes are designed to represent the spatio-temporal and semantic distribution of any geo-related features in any geographic area, although they are tested on the case of 14510 companies registered in Shanghai, China. The following two examples are explained as follows:

1. **The spread of Covid-19 and its variants in a city.** Notice the difference between this case and the case discussed in Mo et. al. (2020) is multiple types of viruses, hence multiple categories of a dataset are taken into account, rather than just one. The input dataset for this case is supposed to be similar to the test dataset for this thesis, and the equivalence of columns is represented in the following table. It is also worth emphasizing that the development of an epidemic normally requires time, from a few infected cases in a small region to a problem affecting a much larger area. This, in addition to the precautions and re-bounce, categorize an epidemic as a “long historical event”.

Column	**Equivalence**
ID	ID of an infected person
Name	Name of the infected person
Industry type	Type of the virus and its variant (for example: the virus itself is assigned as “type 0, the first variant as “type 1”, the second as “type 2”, etc.) that causes the infection
Address	The place where the person is infected (city, street, house, …)
Organization start time	The date or time when the person is infected
Coordinates	The geographic coordinates of the place where the person is infected

Table 5.2 The column equivalence (companies vs. Covid-19)

2.The spread of cartographic thoughts from the perspective of international academic conferences. The following table summarizes the equivalence of columns between the test dataset for this thesis and the input dataset of this case:

Column	Equivalence
ID	The ID of a conference paper (given it is published)
Name	The title of this paper
Industry type	The name of the conference to which the paper is submitted
Organization type	The theme (or conference session) to which the paper belongs
Address	The place (venue) where the conference takes place
Organization start time	The date or time when the paper is accepted by the conference
Coordinates	The geographic coordinates of the venue

Table 5.3 The column equivalence (companies vs. conference)

5.3 The unusual data behaviors

For the case study, the overall number of companies belonging to each industry type vary from type to type, so does the number of new companies per industry type per year. This results in oversized symbols in some space-time bins, especially those representing companies in the 2000s. However, apart from the dominance of and the roles played by different industry types in urban development, repetitive information is an unusual and in dispensable fact or to the oversized symbols.

	A	B	C	D	E	F	G	H	I	J	K	L
9517	5da7204a2aeede1fc3db82f8	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9518	5da721002aeede1fc3db82fd	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9519	5da7219f2aeede1fc3db8302	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9520	5da7223a2aeede1fc3db8307	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9521	5da722d02aeede1fc3db830c	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9522	5da723702aeede1fc3db8311	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9523	5da724112aeede1fc3db8316	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9524	5da724ca2aeede1fc3db831b	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9525	5da7256a2aeede1fc3db8320	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9526	5da726192aeede1fc3db8325	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9527	5da726bc2aeede1fc3db832a	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9528	5da7274b2aeede1fc3db832f	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9529	5da727f82aeede1fc3db8334	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9530	5da728a62aeede1fc3db8339	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9531	5da729462aeede1fc3db833e	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9532	5da72a092aeede1fc3db8343	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9533	5da72ab22aeede1fc3db8348	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9534	5da72b512aeede1fc3db834d	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9535	5da72c142aeede1fc3db8352	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9536	5da72cdb2aeede1fc3db8357	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9537	5da72d792aeede1fc3db835c	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9538	5da72e112aeede1fc3db8361	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9539	5da72e9b2aeede1fc3db8366	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9540	5da72f3b2aeede1fc3db836b	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9541	5da72fc52aeede1fc3db8370	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9542	5da730542aeede1fc3db8375	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9543	5da730dd2aeede1fc3db837a	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9544	5da731692aeede1fc3db837f	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9545	5da731f32aeede1fc3db8384	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0
9546	5da7327c2aeede1fc3db8389	顶柱检测技术（上海）股份有限公司	2008年	质检技术服务	科学研究和技术服务业	上海市嘉定区金沙江西路1555弄37	上海	31.23915812	121.326649	顶柱检测技术（上	http://inst	0

Figure 5.9 Repetitive information

In this screen shot, everything but ID (highlighted in bright yellow) for each company is the same, and there are over 1500 companies in the whole dataset with the same problem. Because IDs distinguish items from each other in a dataset, companies with different IDs are regarded as different ones. In the future, to reduce the extreme visual outcome of symbols, the logarithmic number of companies may replace the actual number of companies to achieve an improved visual outcome.

5.4 The missing data in "transportation, storage and mailing"

As seen from chapter four, the first organization type belonging to "transportation, storage and mailing" “cultivated” the land of Shanghai in 1905, and the last in 2003. Additionally, the temporal gap between the second and the third organization types is about 40 years. Similarly, the temporal gap between the third and the fourth is around 30 years. When viewing the 3D visual outcome, a reader with insufficient knowledge of Chinese economic history may mistakenly believe that the transportation industry developed much slower in Shanghai compared to the manufacturing or architectural industry. The large number of missing information in the test dataset explains the anomaly. Firstly, the group of 2290 companies with missing year of registration are a decent source of other organization types. Secondly, the latest year of registration of the companies in the dataset is 2016, rather than 2021 (present). The visual outcome would definitely be more convincing if the dataset was complete.

5.5 Unexplored columns and information

In this thesis, the only column containing semantic attributes in the test dataset that is thoroughly explored is "industry type". Considering the 649 categories, “organization types” definitely include much more intriguing semantic attributes. The 3D visualization prototypes are also able to be implemented to "organization types" to reveal significant spatio-temporal and semantic patterns. However, considering the time-consuming Chinese-English translation, this column is not yet vastly explored.

The possibly immense effort of data translation also applies to "product information" in the "firm info" column. There are hundreds and thousands of products for all recorded companies, and the product names are all standardized terms. Those listed in **Figure 5.10** is just a small example. Nonetheless, the visualization of spatio-temporal distribution of products in Shanghai is absolutely worth executing, to tell another story about "scattered data under multiple categories".

产品信息：18k白金镶钻翡翠戒指　翡翠手镯　翡翠镶嵌胸针　翡翠镶钻戒指　金条　墨翠罗汉手链　珍珠吊坠　珍珠耳环　珍珠挂件　珍珠戒指　珍珠胸针　钻石戒指　钻石项链

Figure 5.10 Product information (jewelry)

Furthermore, this thesis only includes one story about “maximal number of features": the regions with the maximum number of companies per year. There are many similar stories to be narrated from many other points of view, and one of which is "joint-investment companies". The 1347 joint-investment companies are extracted from the test dataset and are visualized in **Figure 5.11**. As such, the 3D visualization prototypes are applicable for illustrating the regions with the maximum number of joint-investment companies annually, with or without categories (industry types and/or organization types) specified. This will clearly reflect the character of Shanghai as an international metropolitan city.

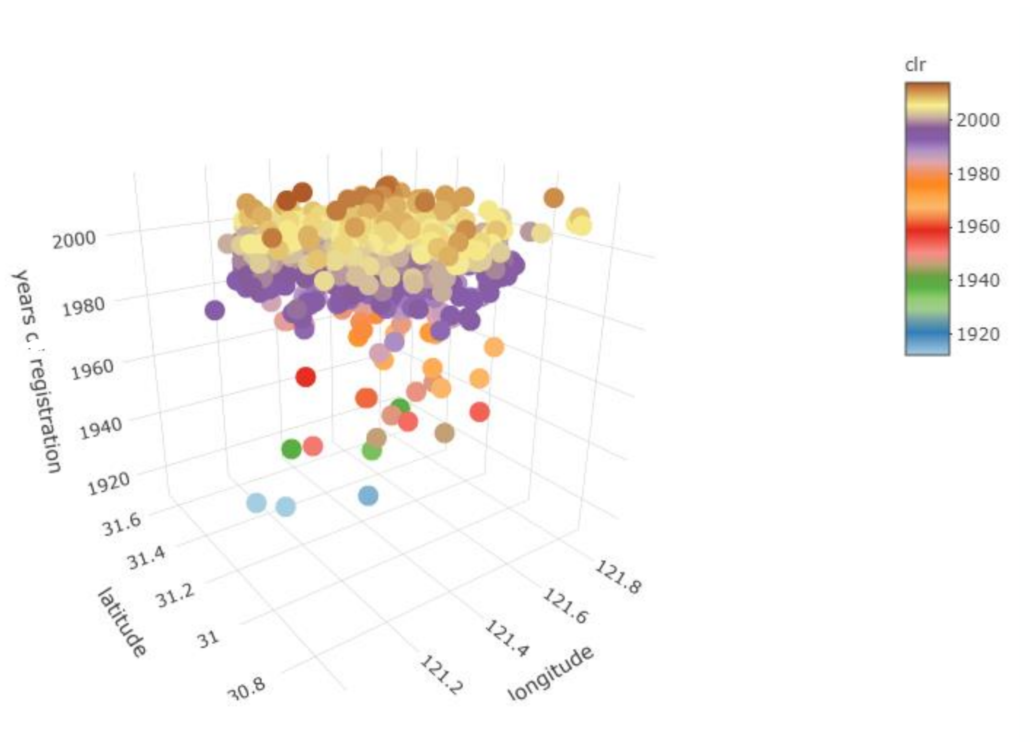


Figure 5.11 Joint-investment companies

6 Conclusions and Future work

The following sections lists important conclusions drawn from the thesis research, and discusses future work in order to better answer Dubourg's great historic question mentioned in chapter two.

6.1 Conclusions

This thesis proposed three prototypes for visualizing scattered spatio-temporal data, and they are represented via geovisual storytelling techniques. Meanwhile, as mentioned in chapter one, there has not been any scientific paper in the domain of cartography or any standard dictionary that gives a precise definition of the term "semantic distribution". Chen et. al. (2021), a paper in the scope of computer vision and machine learning is the only literature source found by the author with "semantic distribution" mentioned, even though just once and in the abstract. As such, this thesis attempted to conduct some research on the semantic distribution of geo-related features, i.e., the spatio-temporal distributions of semantic attributes.

There are several conclusions summarized from this thesis as follows.

1. An STC is advanced in representing spatio-temporal distribution of geo-related features concisely, and the over plotting issue of traditional 2D mapping is easily avoid. However, the proposed prototypes, particularly the space-time micro landscapes are capable of depicting multiple categories, hence the semantic attributes of geo-related features via aggregated 3D symbols. This answers research questions1 and 2.

2. The influences of historical events on the spatio-temporal and semantic distribution of geo-related features in a geographic area be reflected by their numbers under each category. Additionally, the categories of these features normally develop from basic to advanced. It can be a solid story line for cases such as enterprise evolution. Thus, research question 3 is answered.

3. In general, peaceful environments and revolutionary historical events foster the growth of geo-related features (not the case for an epidemic, though) in a geographic area, whereas wars and political chaos suppress their growth. A typical example is the growth of companies in a city such as Shanghai in terms of numbers and types. Meanwhile, for Shanghai, thanks to the historical influence of foreign settlements, the city center remained as a popular company hub in all time periods, while the outskirts eventually become new company hubs because of other influential historical events. Thus, research question 4 is answered.

4. The combination of 3D spatio-temporal visualization prototypes and classic ArcGIS StoryMap storytelling techniques are indeed effective in conveying important spatio-temporal and semantic information from various perspectives. This strengthens the answers to research question 1.

6.2 Future work

6.2.1 The "missing year of registration" problem

As addressed in chapter four and chapter five, there are 2290 companies whose "organization start time", i.e., year of registration is missing in the test dataset. A part of these companies' information is shown in the following screenshot (**Figure 6.1**).

In order to make the stories narrated in this thesis more reliable, the missing years of registration must be filled in. The missing years of registration may be roughly estimated by using the parts of the dataset with complete information. For instance, the missing year of registration of a toy manufacturing company can be estimated by searching for other toy manufacturing companies that are registered in the same time period, if they all produce the same type of toys (corresponding information lies in the "firm info" column). This rough estimation may serve as an empirical clue for accurate information acquisition online or offline.

	A	B	C	D	E	F	G	H	I	J	K
1	_id	org_name	org_start_tim	org_type	industry_type	location	org_city	wgs84_lat	wgs84_lon	firm_info	url
11192	5da5aded2aeede1fc3db5111	上海长光企业发展有限公司		纤维素纤维原料及纤维制造	制造业	上海市金山区金山卫镇老卫清路89	上海	30.72407981	121.29889	上海长光企业发展有	http://inst
11193	5da634f22aeede1fc3db6361	上海山阳化工设备厂		炼油、化工生产专用设备制造	制造业	上海市金山区石化海光路18号　（	上海	30.72467344	121.350825	上海山阳化工设备厂	http://inst
11194	5da5f6a32aeede1fc3db5a95	上海蓝欧化工科技有限公司		化学原料和化学制品制造业	制造业	上海市金山区秋实路688号A2　（_	上海	30.72772972	121.271853	上海蓝欧化工科技有	http://inst
11195	5da644ba2aeede1fc3db6587	上海求德生物化工有限公司		生物药品制造	制造业	上海市金山区金山卫镇秋实路688号	上海	30.72781006	121.271459	上海求德生物化工有	http://inst
11196	5da5f6dd2aeede1fc3db5a9b	上海炼升化工股份有限公司		化学原料和化学制品制造业	制造业	上海市金山区卫六路1259号　（上	上海	30.72935885	121.293556	上海炼升化工股份有	http://inst
11197	5da6ab862aeede1fc3db7389	上海爱米卡工贸有限公司		家用纺织制成品制造	制造业	上海市金山区金山大道3966号　（	上海	30.73288956	121.301693	上海爱米卡工贸有限	http://inst
11198	5da631522aeede1fc3db62ef	上海金狮化工有限公司		化学原料和化学制品制造业	制造业	上海市金山区第二工业区金瓯路88	上海	30.73403985	121.285167	上海金狮化工有限公	http://inst
11199	5da61b5e2aeede1fc3db5fde	上海富琪试验设备有限公司		试验机制造	制造业	上海市金山区枫泾工业园　（上海	上海	30.74416388	121.337438	上海富琪试验设备有	http://inst
11200	5da6394e2aeede1fc3db63f6	上海正强电器有限公司		电气机械和器材制造业	制造业	上海市金山区兴塔镇经济工业区	上海	30.74416388	121.337438	上海正强电器有限公	http://inst
11201	5da6b6f42aeede1fc3db7517	上海信博森化工有限公司		化学原料和化学制品制造业	制造业	上海市金山区吕港镇吕廊路32号	上海	30.74416388	121.337438	上海信博森化工有限	http://inst
11202	5da61f5a2aeede1fc3db6046	上海王泰电气有限公司		变压器、整流器和电感器制造	制造业	上海市金山区板桥东路428号　（_	上海	30.74730334	121.362609	上海王泰电气有限公	http://inst
11203	5da688f12aeede1fc3db6ef6	上海大花自动化科技股份有限公司		其他专用设备制造	制造业	上海市金山工业区卫昌路315号	上海	30.74873279	121.363989	上海大花自动化科技	http://inst
11204	5da666e52aeede1fc3db6ae7	上海十田实业有限公司		水资源专用机械制造	制造业	上海市金山区亭卫公路1066号　（	上海	30.75429283	121.353575	上海十田实业有限公	http://inst
11205	5da6d0e02aeede1fc3db77eb	上海英码克机电设备有限公司		机械设备、五金产品及电子产品批发	批发和零售业	上海市金山区龙皓路950弄3号写字	上海	30.75623609	121.330766	上海英码克机电设备	http://inst
11206	5da66a512aeede1fc3db6b46	上海侨世电气科技有限公司		电气机械和器材制造业	制造业	上海市金山区海丰路65号　（上海	上海	30.75973557	121.348281	上海侨世电气科技有	http://inst
11207	5da5b5422aeede1fc3db51ca	上海江浪流体机械制造有限公司		泵及真空设备制造	制造业	上海市金山区山阳镇南阳港东路35	上海	30.76985273	121.358117	上海江浪流体机械制	http://inst
11208	5da60a212aeede1fc3db5d36	上海成东科技有限公司		制药专用设备制造	制造业	上海市金山区山阳镇阳乐路367号	上海	30.77424365	121.342054	上海成东科技有限公	http://inst
11209	5da62f282aeede1fc3db62a1	上海东大聚氨酯有限公司		合成材料制造	制造业	上海市金山区山阳镇山宁路307号	上海	30.77461518	121.346787	上海东大聚氨酯有限	http://inst
11210	5da68ee02aeede1fc3db6fae	迪梦柯机械(上海)有限公司		通用设备制造业　专用设备制造业	制造业	上海市金山区山富西路289弄9号	上海	30.77660873	121.348189	迪梦柯机械(上海)有	http://inst
11211	5da5e2322aeede1fc3db5807	上海贝尼汽车科技有限公司		汽车零部件及配件制造	制造业	上海市金山区山阳镇山富东路181号	上海	30.77701435	121.356451	上海贝尼汽车科技有	http://inst
11212	5da71b632aeede1fc3db8243	上海四维数字图文有限公司		专业化设计服务	科学研究和技术服务业	上海市金山区亭卫公路2588号　（	上海	30.78513296	121.353838	上海四维数字图文有	http://inst
11213	5da692602aeede1fc3db702e	上海帆顺毛瑟包装有限公司		塑料包装箱及容器制造	制造业	上海市金山工业园区漕泾中一东路	上海	30.79587751	121.421923	上海帆顺毛瑟包装有	http://inst
11214	5da636482aeede1fc3db638d	上海通泉泵业制造有限公司		泵及真空设备制造	制造业	上海市金山区漕泾镇西部工业区	上海	30.79613494	121.406299	上海通泉泵业制造有	http://inst
11215	5da6e05c2aeede1fc3db7973	拜耳技术工程（上海）有限公司		机械设备、五金产品及电子产品批发	批发和零售业	上海市化学工业区目华路F3地块	上海	30.7981992	121.473496	拜耳技术工程（上海	http://inst
11216	5da5b9e72aeede1fc3db5240	上海中贵化工设备有限公司		炼油、化工生产专用设备制造	制造业	上海市金山区漕泾镇朱漕公路4333	上海	30.80564471	121.389096	上海中贵化工设备有	http://inst
11217	5da67f642aeede1fc3db6df3	上海帅翼驰铝合金新材料有限公司		有色金属合金制造	制造业	上海市金山区张堰镇振堰路699号	上海	30.80620654	121.266765	上海帅翼驰铝合金新	http://inst
11218	5da5ad282aeede1fc3db50f0	上海伊贝纳纺织品有限公司		纺织业	制造业	上海市金山区张堰镇金张公路258号	上海	30.80791869	121.279545	上海伊贝纳纺织品有	http://inst
11219	5da5b04c2aeede1fc3db514a	上海克兰传动设备有限公司		轴承、齿轮和传动部件制造	制造业	上海市金山区张堰镇金张公路609号	上海	30.80891055	121.263359	上海克兰传动设备有	http://inst
11220	5da680d62aeede1fc3db6e22	上海亿思特电气股份有限公司		电气机械和器材制造业	制造业	上海市金山工业区月工路777号	上海	30.810505	121.347113	上海亿思特电气股份	http://inst

Figure 6.1 Companies with missing year of registration (org_start_time circled)

6.2.2 The combination of 2.5D and 3D visualizations

As addressed in chapter five, an algorithm for directly inserting a basemap at the bottom of the prototypes is yet to be developed. Alternatively, a user is able to export the desired temporal layer from RStudio and re-import to ArcGIS Pro or QGIS, create a 2D or 2.5D map, and use it as an appropriate geographic reference. This, according to Mayr and Windhager (2018), proves that an STC can be switched to layer juxtaposition, to assist a user understand the exact situations on various temporal layers. However, apart from manual color scheme selection, a more efficient method in ArcGIS Pro that ensures the unity of color coding for the given industry types' symbols, regardless of the chosen temporal layer is also yet to be discovered. This issue is only resolved in RStudio. In addition, even though the efficient method ensuring the unity of color coding is discovered in ArcGIS, the same color hue may appear difference, i.e., with different shades in RStudio and ArcGIS Pro. This challenge is there to be overcome in the future.

6.2.3 A user test

Due to the time budget and the aim of this thesis' research: attempting to theoretically propose novel

3D visualization prototypes, which leans more to the scientific than the technical domain (in other words, the purpose of this thesis' research is not to design anything that goes straightly to the market), a user test has not yet conducted to examine the design concept and the usability of the prototypes. As such, as what was conducted in Kveladze, Kraak and Elzakker (2017), as well as in Bogucka and Jahnke (2018), a brief user test or a simple survey deserves to be arranged in the future.

6.2.4 More analyses on "spatio-temporal and semantic distribution"

6.2.4.1 The "first law of geography and chronology

The first law of geography "everything is related to everything else, but near things are more related than distant things (Tobler, cited in Dempsey, 2014)" is well-known for cartographers and geodata analysts. Will there be a temporal equivalent of this law, i.e., "things that appear more simultaneously are more related to things that appear in longer time intervals"? Or to extrapolate, will there be a "first law of geography and chronology", in another word "things that appear closer and more spontaneously are more related to things that appear distant and in longer time intervals"? It is worthwhile to propose a law or theory that more closely correlates geographical and chronological patterns.

The research of this thesis actually focuses more on spatio-temporal distributions than semantic distributions of geo-related features, as pointed out in chapter five. There are many more semantic analyses to be done, to enrich the content of the stories told in this thesis. In addition, the current prototypes maybe modified or redesigned for visualizing the results of these semantic analyses.

6.2.4.2 The "city center" phenomenon

One typical spatio-temporal pattern observed in this thesis regarding Shanghai is fact that human settlement in the outskirts gradually became crowded with companies from around the 1920s, especially Jiading, Minhang, Baoshan, Songjiang, Jinshan, Fengxian and Pudong. However, the city center never failed to attract new companies. For instance, it is the virgin land of many dominant industry types, including, but are not limited to manufacturing, whole sale and retail, architecture, scientific research and technical services and transportation.

Apart from the influence of old foreign settlements, one explanation for this pattern is the essential roles the Central Business District (CBD) plays in a city's prosperity and social environment. However, are there any other explanations? Is it possible that some companies that were registered in the CBD are the headquarters of other companies registered in the outskirts (Wei & Bai, 2008)? For instance, there are two companies with the brand name "冠生园 (pronounced as Guan Sheng Yuan)", as shown in **Figure 6.2** (the first and the last companies, highlighted in orange). What is the relationship between these two companies?

18	5da5d98f2aeede1fc3db5675	冠生园(集团)有限公司	1915年	食品制造业	制造业
19	5da64ebd2aeede1fc3db66fa	上海华线医用核子仪器有限公司	1916年	医疗诊断、监护及治疗设备制造	制造业
20	5da6510d2aeede1fc3db6770	上海鹰牌衡器有限公司	1917年	衡器制造	制造业
21	5da69aee2aeede1fc3db7167	上海华通开关厂有限公司	1919年	配电开关控制设备制造	制造业
22	5da6e6a52aeede1fc3db7a62	上海强生集团有限公司	1919年	公共电汽车客运	交通运输、仓储和邮政业
23	5da653d82aeede1fc3db67f9	上海实业马利画材有限公司	1919年	颜料制造	制造业
24	5da6bb662aeede1fc3db75bd	上海市基础工程集团有限公司	1919年	房屋建筑业　土木工程建筑业	建筑业
25	5da654f82aeede1fc3db6827	上海造币有限公司	1920年	其他未列明金属制品制造	制造业
26	5da5d4c32aeede1fc3db559e	上海天原集团胜德塑料有限公司	1921年	初级形态塑料及合成树脂制造	制造业
27	5da6507e2aeede1fc3db6754	上海冠生园食品有限公司	1921年	食品制造业	制造业

Figure 6.2 Companies with the same brand name

One possible answer to the questions above is that over time, the limited space in the CBD became no longer sufficient for new companies, but the vast space of a city's outskirts was more optimal for new industrial zones and branching companies. It is partially proved by the "virgin land problem" from the perspective of "transportation, storage and mailing", because the first companies belonging to "airport", "airline transport". "storage" and "delivery" were established in Pudong and other outskirts. However, in the test dataset, there is a lack of information about which companies in the outskirts are branching companies of which companies in the CBD. The test dataset can be regarded a "source material" to retrieve more valuable information via a web crawler, but the exact techniques (for instance, programming) are to be determined.

6.2.4.3 The "supply chain" relationship

Apart from the "headquarter-branch" relationship, another common relationship amongst companies is "supply chain". Even companies with different names, organization types and/or industry types can form such a relationship, to explain the "city center" phenomenon from a new point of view (Ellram, Tate & Billington, 2004). However, once again, there is a shortage of relevant information in the dataset to chain up the companies. This information deserves to be discovered online or offline in the future.

6.2.4.4 More historical events and governing policies as contexts

In this thesis, only some national historical events are investigated as contexts of potential correlation of the spatio-temporal and semantic distributions of companies. They all have vital influences on Shanghai's economic development, but to make the designed prototypes more robust for depicting semantic distributions, local historical events in Shanghai is worth being investigated.

6.2.4.5 The proposal of "virgin territory of words"

The virgin land problem addressed the geographic spread of semantic attributes briefly by discussing general urban development and global transportation history. Is it possible to map the change of language on a story map? How is it more effective to depict the first semantic attributes of a term in a geographic representation? Same as "the first law of geography and chronology", it is worthwhile to propose a concept of "virgin territory of words". This concept is to be proposed in the collaboration with a linguist and/or a historian.

Bibliography

Andrienko, G. (2011). Challenging problems of geospatial visual analytics. *Journal of Visual Languages and Computing*, 22: 251–256. https://doi.org/10.1016/j.jvlc.2011.04.001

Andrienko, G., Andrienko, N., Jankowski, P. Keim, D., Kraak, M. J., MacEachren, A. & Wrobel, S. (2007). Geovisual Analytics for Spatial Decision Support: Setting the research agenda. *International Journal of Geographical Information Science,* 839-857.https://doi.org/10.1080/13658810701349011

Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S., Wachowicz, M. (2008). Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization 7,*173–180.doi:10.1057/ivs.2008.23

Bogucka, E. & Jahnke, M. (2018). Feasibility of the Space-Time Cube in Temporal Cultural Landscape Visualization. International Journal of Geo-Information. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/13658810701349011

Brinch, S. (22, March 2019). Charles-Joseph Minard's map of napoleon's flawed russian campaign: An ever-current classic. [Web log post] Retrieved from https://sciencenorway.no/blog-blog-from-numbers-to-graphics-statistics/charles-joseph-minards-map-of-napoleons-flawed-russian-campaign-an-ever-current-classic/1618695

Cambridge Dictionary. (2021). Definition of "company". Retrieved from https://dictionary.cambridge.org/dictionary/english/company

Choi, I. (2019). GEO-TEMPORAL VISUALIZATION FOR TOURISM DATA USING COLOR CURVES. (Master Thesis). Department of Computer & Information Technology, Purdue University. Retrieved from http://hdl.handle.net/1805/18926

Chen, K., Wang, Y., Yu, M., Shen, H., Yu, X. & Shan, G. (2020). ConfVisExplorer: a literature-based visual analysis system for conference comparison. Journal of Visualization 24.381-395. Springer. https://doi.org/10.1007/s12650-020-00703-5

Chen, S., Xie, G., Liu, Y., Peng, Q., Sun, B., Li, H., You, X. and Shao, L. (2021). HSVA: Hierarchical Semantic-Visual Adaptation for Zero-Shot Learning. *35th Conference on Neural Information Processing Systems*. Retrieved from https://arxiv.org/abs/2109.15163

Chen, X., Zhang, X., Xie, H., Tao, X., Wang, F., Xie, N. & Hao, T. (2020). A bibliometric and visual analysis of artificial intelligence technologies-enhanced brain MRI research. *Multimedia Tools and Applications.* Springer. Retrieved from https://link.springer.com/article/10.1007%2Fs11042-020-09062-7

Collins Dictionary. (2021). Definition of "virgin land (soil)". Retrieved from https://www.collinsdictionary.com/dictionary/english/virgin-soil

Caquard, S. and Cartwright, W. (2014). Narrative cartography: From mapping stories to the narrative of maps and mapping. *Cartographic Journal, 51 (2)*,101-106. Retrieved from https://doi.org/10.1179/0008704114Z.000000000130

Caroll, A. (11, October 2021). Exploring Iowa. [Web log post] Retrieved from https://arcg.is/1GGPKn

Caroll, A. (7, October 2021). Getting started with ArGIS StoryMaps. [Web log post] Retrieved from https://arcg.is/aDqTP

CRAN. (2021). Plotly: Create interactive web graphics via'plotly.js'. Retrieved from https://cran.r-project.org/web/packages/plotly/index.html

Deutsches Museum Verkehrszentrum (2021). Exhibition Annotation. Retrieved from https://www.deutsches-museum.de/verkehrszentrum/besuch

Dempsey, C. (29, April 2014). Tobler's first law of geography. Geographic Basics. [Web log post] Retrieved from https://www.geographyrealm.com/toblers-first-law-geography/

Ding, L., Jahnke, M., Wang, S. & Karja, K. (2016). Understanding Spatiotemporal Mobility Patterns related to Transport Hubs from Floating Car Data. *Proceedings of the 13th International Conference on Location-Based Services.* LBS 2016, 14-16. Retrieved from https://www.semanticscholar.org/paper/Understanding-Spatiotemporal-Mobility-Patterns-to-Ding-Jahnke/3fb320b89cb494b25f50a73f54a43dd88c825adf

Ding, L., Yang, J. & Meng, L. (2015). Visual Analytics for Understanding Traffic Flows of Transport Hubs from Movement Data. *Proceedings of the 27th International Cartographic Conference, 16th General Assembly.* ISBN 978-85-88783-11-9. Retrieved from https://icaci.org/files/documents/ICC_proceedings/ICC2015/papers/2/785.html

Encyclopedia.com. (2021). History of Industries in Shanghai. Retrieved from https://www.encyclopedia.com/books/international-magazines/history-industries-shanghai

ESRI. (2016). Presidential election 2016: dasymetric dot density. Retrieved from https://carto.maps.arcgis.com/apps/webappviewer/index.html?id=8732c91ba7a14d818cd26b776250d2c3

ESRI. (2021). Create space time cube by aggregating points (space time pattern mining). Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/create-space-time-cube.htm

Ellram, L., Tate, W., Billington, C. (2004) Understanding and Managing the Service Supply Chain. *The Journal of Supply Chain Management: A Global Review of Purchasing and Supply.* The Institute of Supply Management, Inc.17-32. Retrieved fromhttps://doi.org/10.1111/j.1745-493X.2004.tb00176.x

Ferguson, S. (1991). The 1753 "Carte chronographique" of Jacques Barbaeu-Dubourg. *The Princeton University Library Chronicle, WINTER 1991, Vol. 52, No. 2 (WINTER 1991)*, 190-230. Retrieved from https://www.jstor.org/stable/26404421?seq=1&cid=pdf-reference#references_tab_contents

Filho, J., Stuezlinger, W. & Nedel. L (2019). Evaluating an immersive space-time cube geovisualization for intuitive trajectory data exploration.514–524. 10.1109/TVCG.2019.2934415

Yang, G. (2002). Shanghai's economic development: Its opportunities and challenges in the 21st century. *GLOBAL URBAN DEVELOPMENT METROPOLITAN ECONOMIC STRATEGY REPORT*,1–33. Retrieved from https://www.globalurban.org/GUD%20Shanghai%20MES%20Report.pdf

Gaughan, A., Stevens, F., Huang, Z., Nieves, J., Sorichetta, A., Lai, S., Ye, X., Linard, C., Hornby, G., Hay, S., Yu, H. & Tatem, A. (2016). Spatiotemporal patterns of population in mainland china, 1990 to 2010. DOI: 0.1038/sdata.2016.5

Google Culture and Art. (2021). Semantic Property. Retrieved from https://artsandculture.google.com/entity/semantic-property/m0c896?hl=en

Hashim, H., Mohd, W., Sadek, E. & Dimyati, K. (2019). MODELING URBAN CRIME PATTERNS USING SPATIAL SPACE TIME AND REGRESSION ANALYSIS. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-4/W16

Info We Trust. (2020). Stories of space, time and data. [Video] Retrieved from https://www.youtube.com/watch?v=rjl6IfSQZyw&t=2523s

Jasperson, A. (19, May 2020). From virus to pandemic. [Web log post] Retrieved from https://arcg.is/15WL5W0

Kraak, M-J. (2003) THE SPACE-TIME CUBE REVISITED FROM A GEOVISUALIZATION PERSPECTIVE. *Proceedings of the 21st International Cartographic Conference (ICC) Durban, South Africa*. Retrieved from https://icaci.org/files/documents/ICC_proceedings/ICC2003/Papers/255.pdf

Kraak, M-J., Roth. R, Ricker, B, Kagwa. A & Sourd, G. Le. (2020). Mapping for a sustainable world. The United Nations: New York, NY (USA). Retrieved from https://www.un-ilibrary.org/content/books/9789216040468

Kristensson, P., Dahlbaeck, N., Anundi, D., Bjoernstad, M., Gillberg, H., Haraldsson, J., Martensson, I, Nordvall, M. & Stahl, J. (2008). The Trade-offs with Space Time Cube Representation of Spatiotemporal Patterns. *arXiv:0707.1618v1 [cs.HC] 11 Jul 2007*, 1-16. Retrieved from https://www.researchgate.net/publication/220490099_The_Trade-offs_with_Space_Time_Cube_Representation_of_Spatiotemporal_Patterns

Kveladze, I., Kraak, M-J. & Elzakker, C.P. (2017). The usability of a geovisual analytics environment for the exploration and analysis of different datasets. *Advances in Cartography and GIScience*. Lecture Notes in Geoinformation and Cartography. Springer. 207–221.

Lee, S. (3, May 2021). In search of home. [Web log post] Retrieved from https://arcg.is/1954Sa

Mo, C., Tan, D., Mai., T., Bei., C., Qin, J., Pang., W., Zhang, Z. (2020). An analysis of spatiotemporal pattern for COIVD-19 in China based on space-time cube. *J Med Virol*. 2020; 92:1587–1595. DOI: 10.1002/jmv.25834.

Mayr, E. & Windhager, F. Once upon a Spacetime: Visual Storytelling in Cognitive and Geotemporal Information Spaces. *International Journal of Geo-Information*, 1-18. doi:10.3390/ijgi7030096

O’Sullivan, D & Unwin, D. J. (2010). Nonuniformity of Space and Edge Effects. *Geographic Information Analysis*. John Wiley & Sons, Inc. 40-41

O’Sullivan, D & Unwin, D. J. (2010). Geovisualization and Analysis. *Geographic Information Analysis*. John Wiley & Sons, Inc. 58-60

O'Sullivan, D & Unwin, D. J. (2010). The Graphic Variables of Jacques Bertin. *Geographic Information Analysis*. John Wiley & Sons, Inc. 60-61

Purwanto, P., Handoyo, S., Bachri, S., Astuti, I., Kresno, S., Utomo, B. & Alditano, Y. (2021). Spatiotemporal Analysis of COVID-19 Spread with Emerging Hotspot Analysis and Space–Time Cube Models in East Java, Indonesia. *International Journal of Geo-Information*, 1-22. https://doi.org/10.3390/ijgi10030133

Roth, R. (2020). Cartographic design as visual storytelling: Synthesis and review of map-based narratives, genres, and tropes: *Synthesis and Review of Map-Based Narratives, Genres, and Tropes, The Cartographic Journal*, DOI: 10.1080/00087041.2019.1633103 1–32.

Schmitz, P. (2017). Displaying voter gains and losses: Local government election in south Africa for 2011 and 2016. *Advances in Cartography and GIScience*. *Lecture Notes in Geoinformation and Cartography.* Springer. 309-323

Sapena, M., Wurm, M., Taubenboeck, H., Tuia, D. & Ruiz, L. (2020). Estimating quality of life dimensions from urban spatial pattern metrics. *Computers, Environment and Urban Systems 85*. Elsevier. Retrieved from https://www.sciencedirect.com/science/article/pii/S0198971520302829?via%3Dihub

Segel, E. & Heer, J. (2010). Narrative visualization: Telling stories with data. IEEE *TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, VOL. 16, NO. 6.1139–1148. Retrieved from https://www.researchgate.net/publication/47544586_Narrative_Visualization_Telling_Stories_with_Data

Shanghai History Museum, Shanghai Revolution Museum. (2021). Exhibition annotation. Retrieved from http://www.shh-shrhmuseum.org.cn/historymuseum/historymuseum/index.html

Taylor, C. (01., September 2021). Passengers getting on and off a bus. [Web log post] retrieved from https://twitter.com/CraigTaylorViz/status/1433020149037162498

Taylor, C. (17, March 2021). Data flower. [Web log post] Retrieved from https://twitter.com/CraigTaylorViz/status/1372125199852830720

Turdukulov, U. & Fazio, T. (2021). Exploring Kimberley Bushfires in Space and Time. Department of Spatial Science, Curtin University, Bentley, WA, Australia. *Proc. of the 3rd Annual Conference of Research@Locat.* 25-29. Retrieved from http://ceur-ws.org/Vol-1570/paper09.pdf

Windhager, F., Mayr, E., Schreder, G., Smuc, M., Federico, P. & Miksch, S. (2016) Reframing Cultural Heritage Collections in a Visualization Framework of Space-Time Cubes. *Proceedings of the 3rd HistoInformatics Workshop, Krakow, Poland, 11 July 2016.* Retrieved from http://ceur-ws.org/

Wei, H. & Bai, M. (2008). Zhongguo Shangshi Gongsi Zongbu Qianyi Xianzhuang Ji Tezheng Fenxi (The Actuality and Characteristics on Relocation of Listed Company Headquarters in China). *China Industrial Economy.* Institute of Industrial Economics, Chinese Academy of Social Sciences. 1-12. Retrieved from https://www.worldscientific.com/doi/10.1142/9789814452267_0009

Appendix

The complete story map:
https://storymaps.arcgis.com/stories/57e6c1bf8db4485fa311059b00e2304c

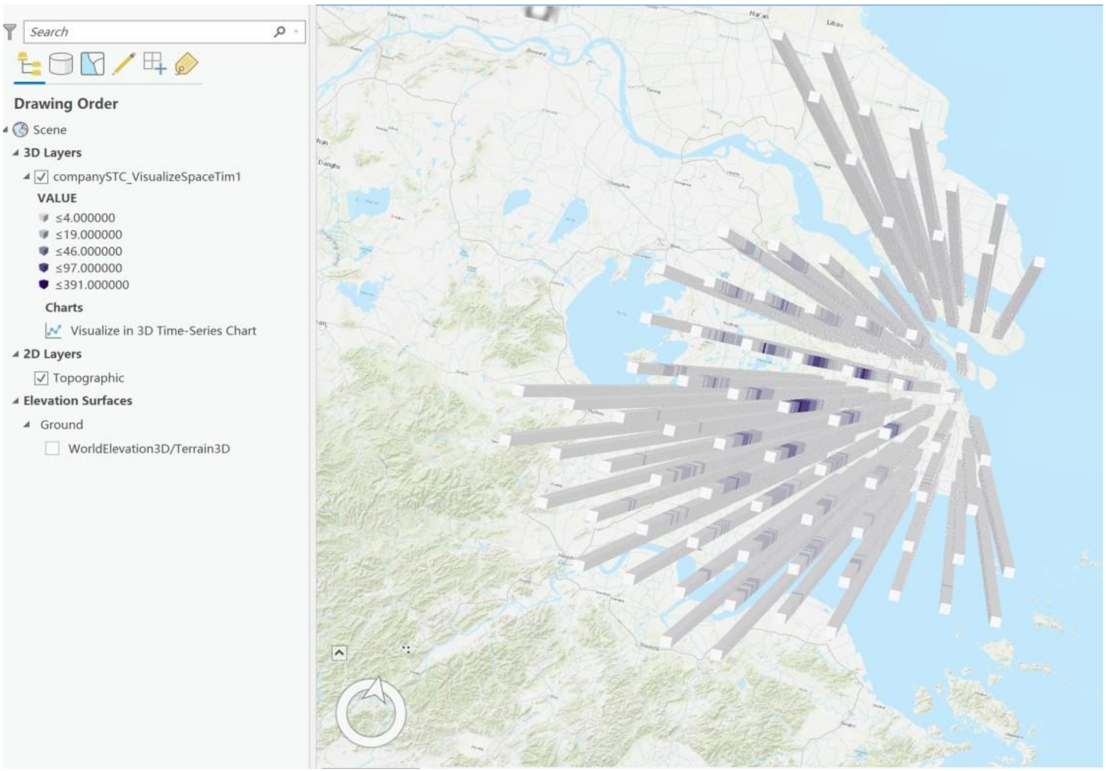


Figure 0.1 STC generated by ArcGIS Pro (count of companies per year, 10km-by-10km)

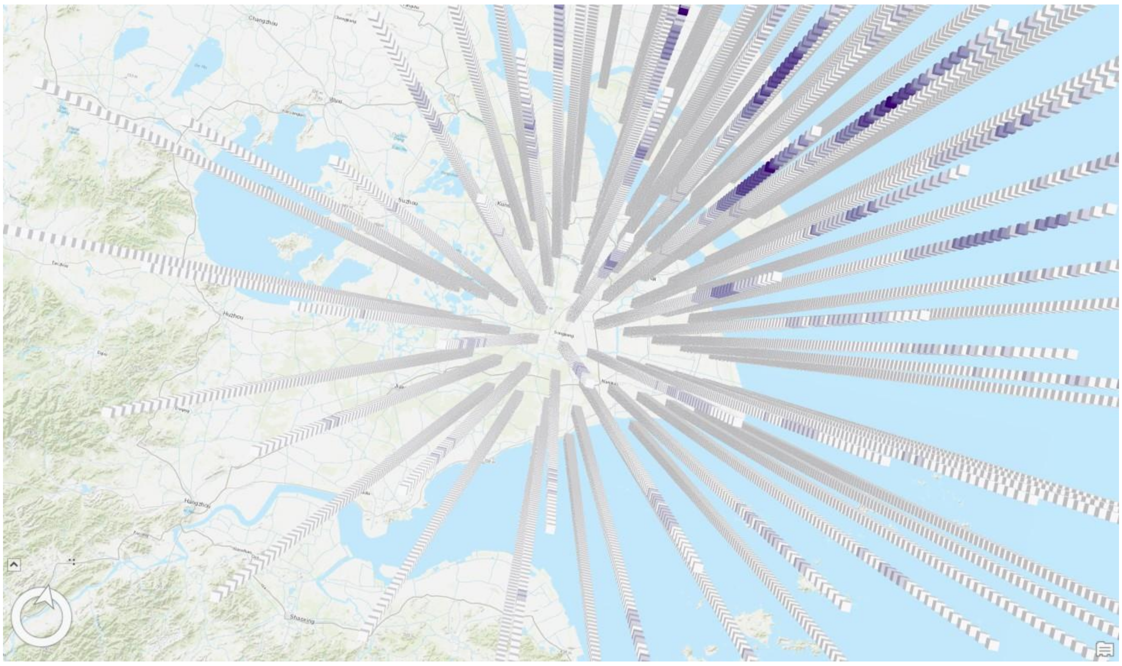


Figure 0.2 STC generated by ArcGIS Pro zoomed in (count of companies per year, 10km-by-10km)

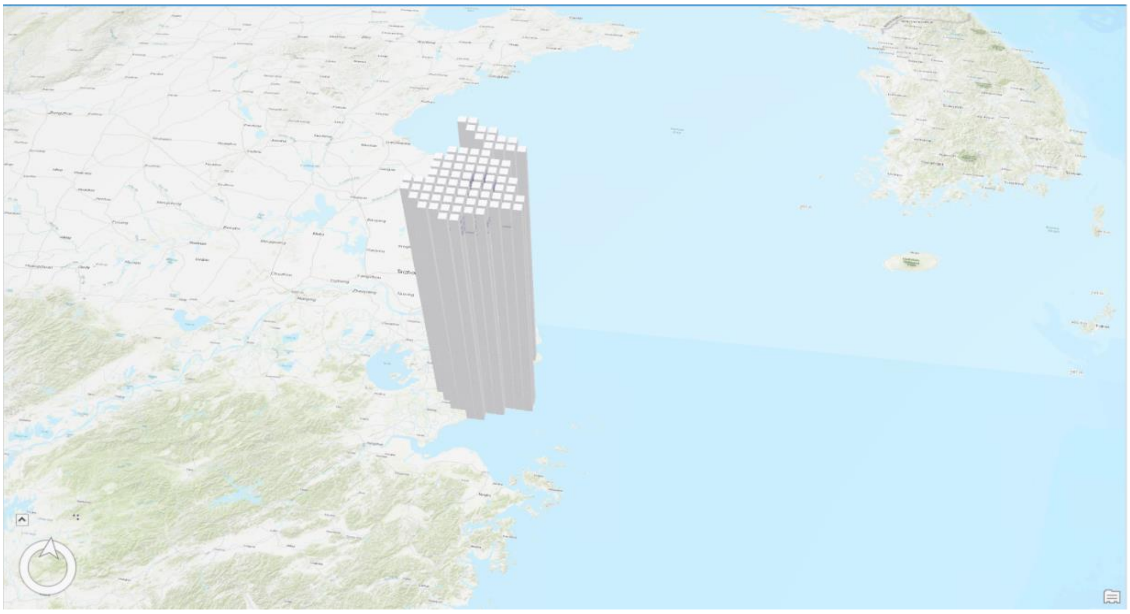


Figure 0.3 STC generated by ArcGIS Pro from the top (count of companies per year, 10km-by-10km)

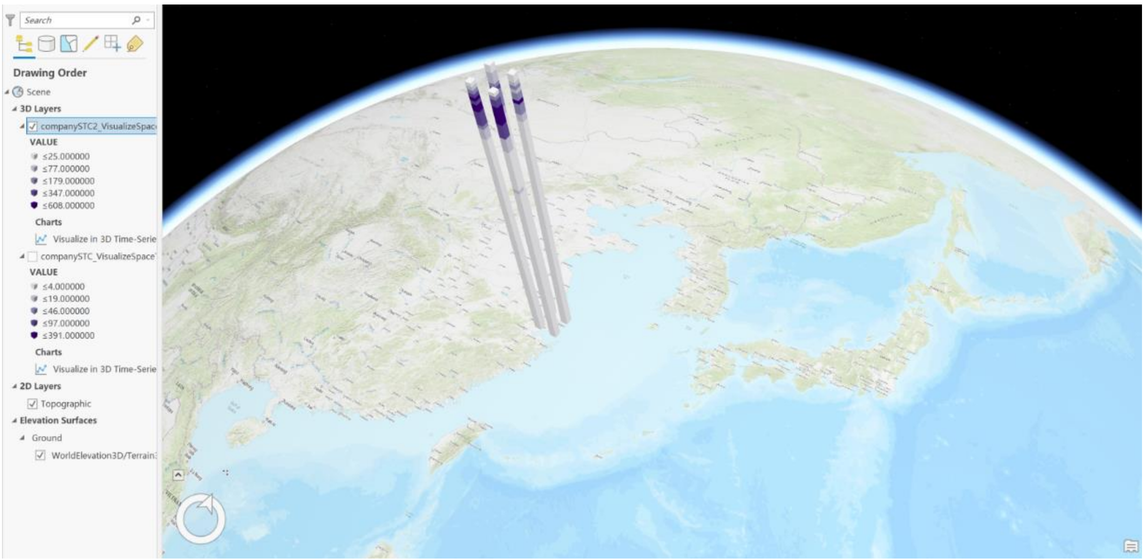


Figure 0.4 STC generated by ArcGIS Pro (count of companies per year, 100km-by-100km)

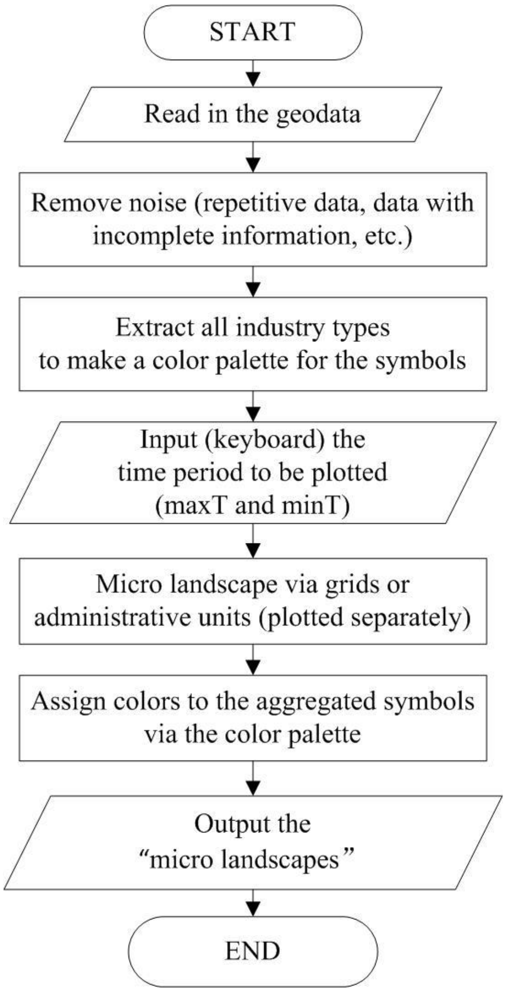


Figure 0.5 General flow chart of space-time micro landscapes

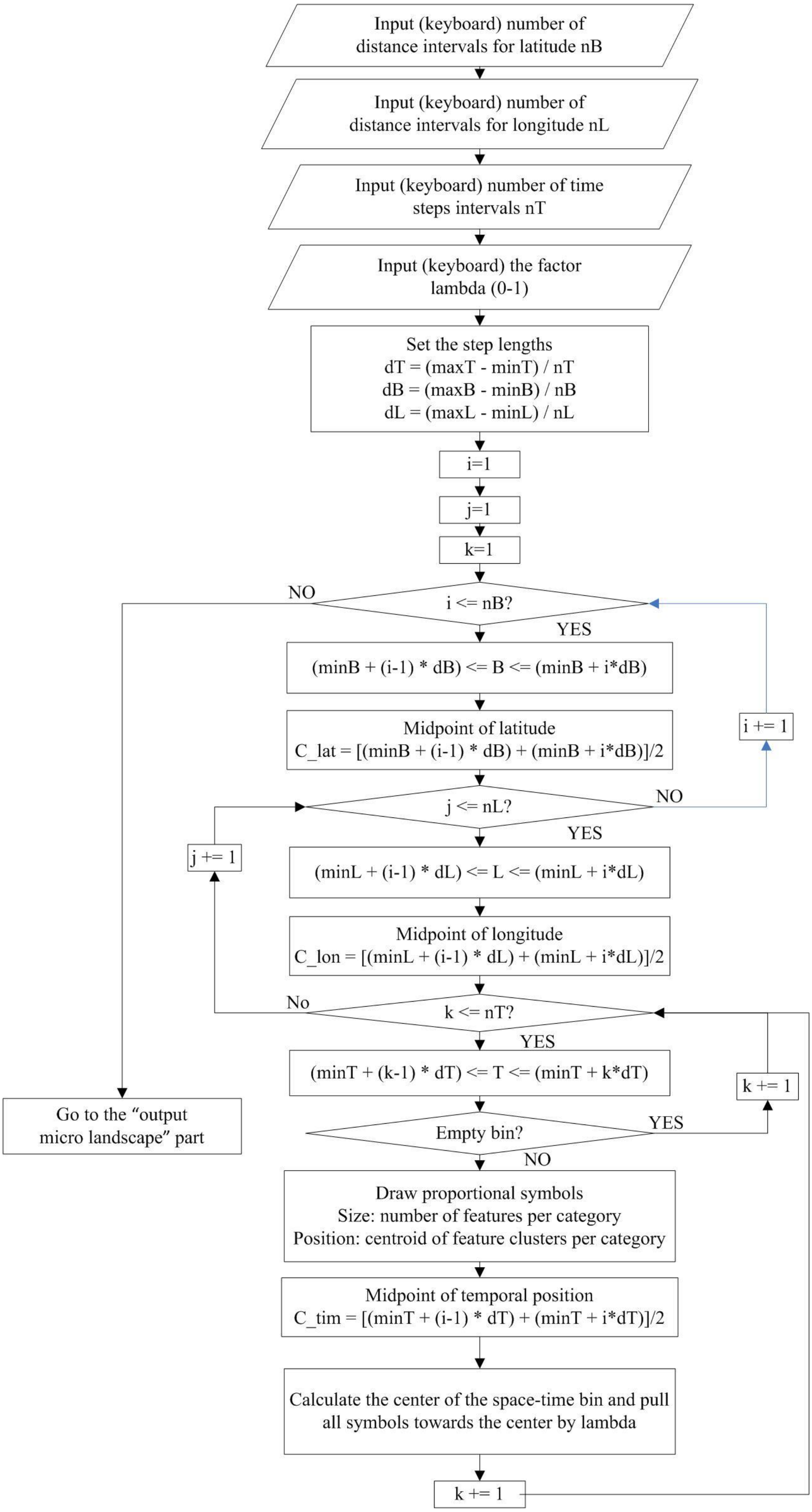


Figure 0.6 Flow chart of space-time micro landscapes via grids

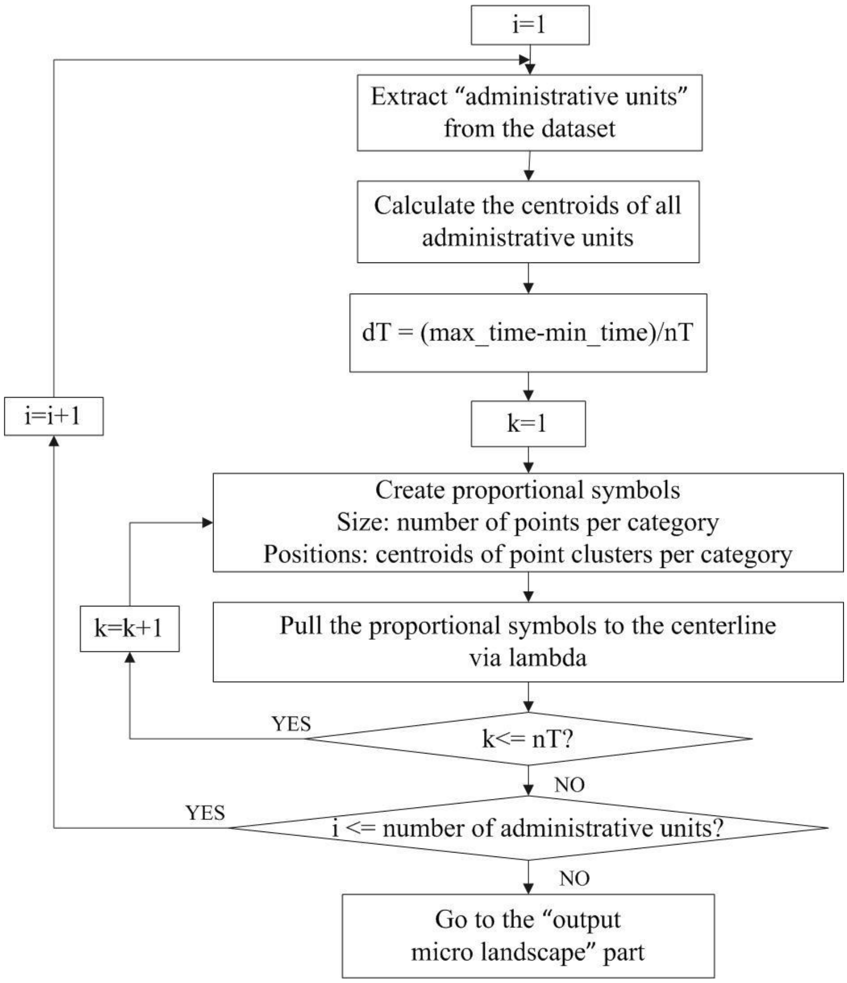


Figure 0.7 Flow chart of space-time micro landscapes via administrative units

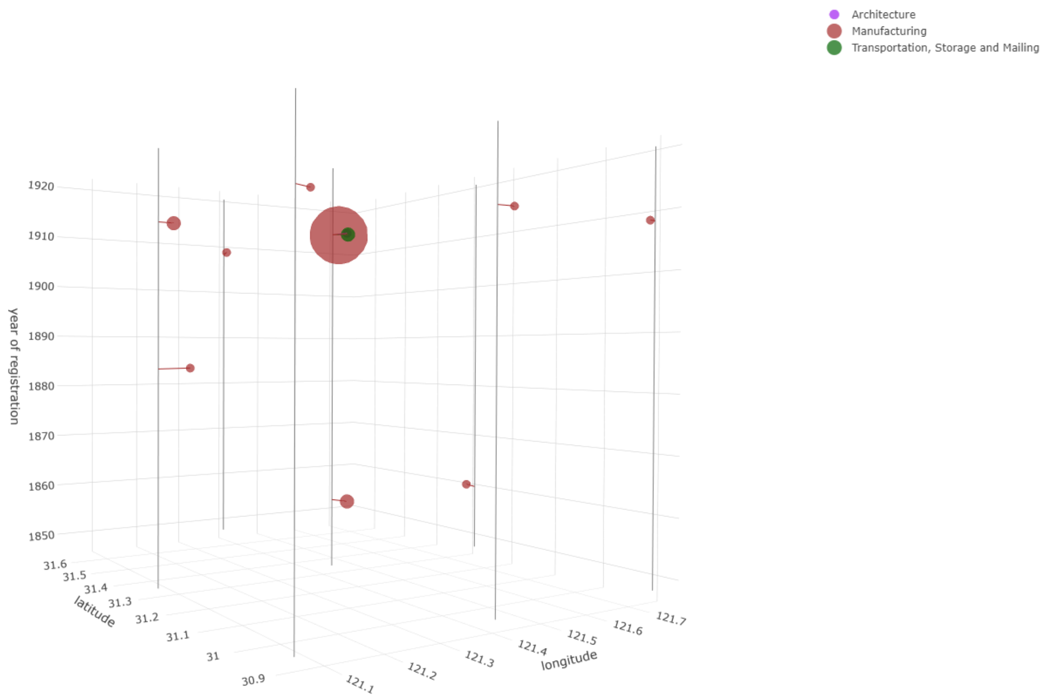


Figure 0.8 A few features, 3*3 grids, lambda = 0.3, 3 timesteps

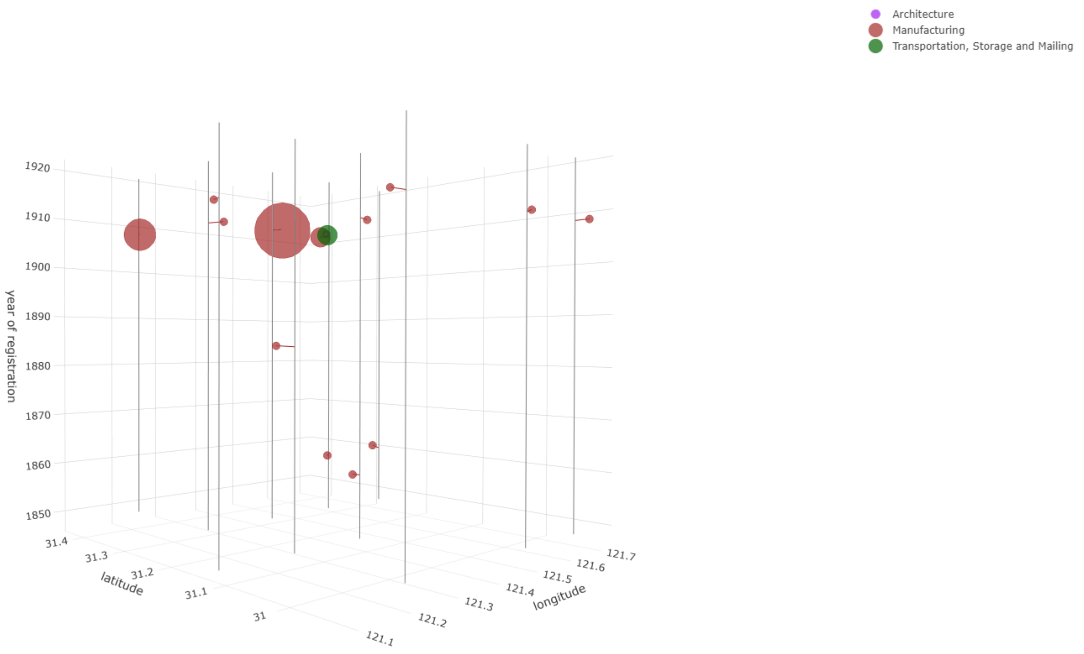


Figure 0.9 A few features, 7*7 grids, lambda = 0.3, 3 timesteps

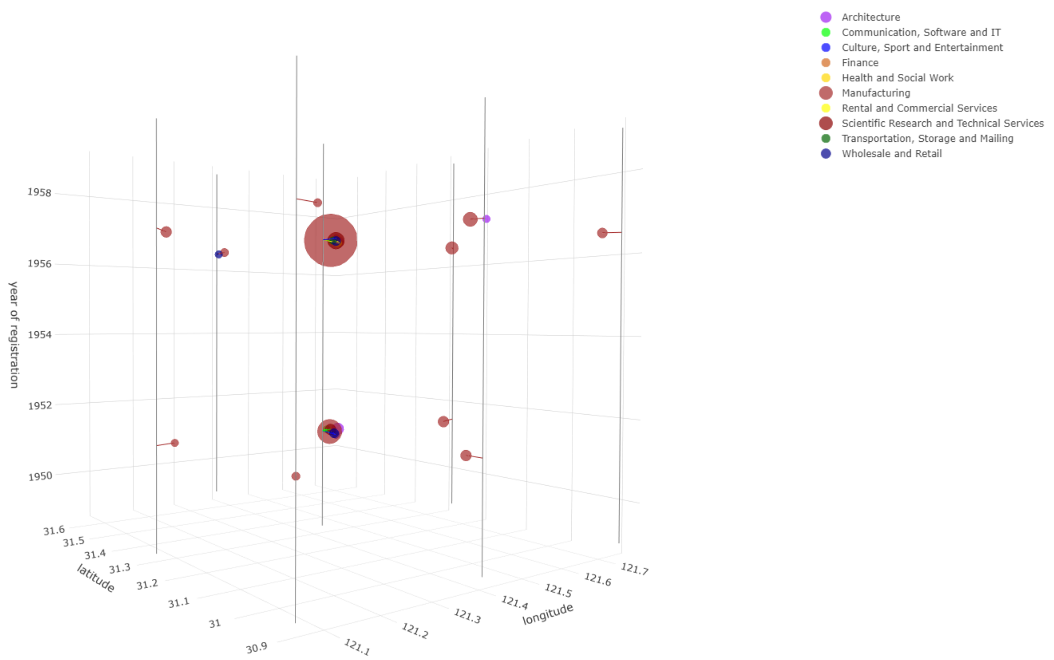


Figure 0.10 Many more features, 7*7 grids, lambda = 0.3, 2 timesteps

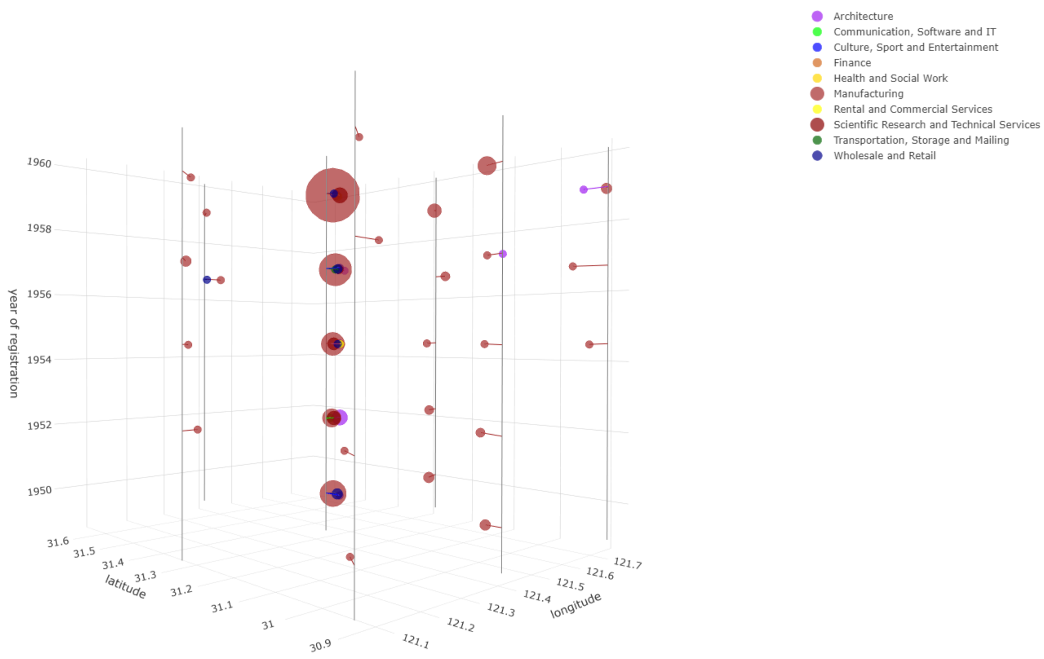


Figure 0.11 Many more features, 3*3 grids, lambda = 0.3, 5 timesteps

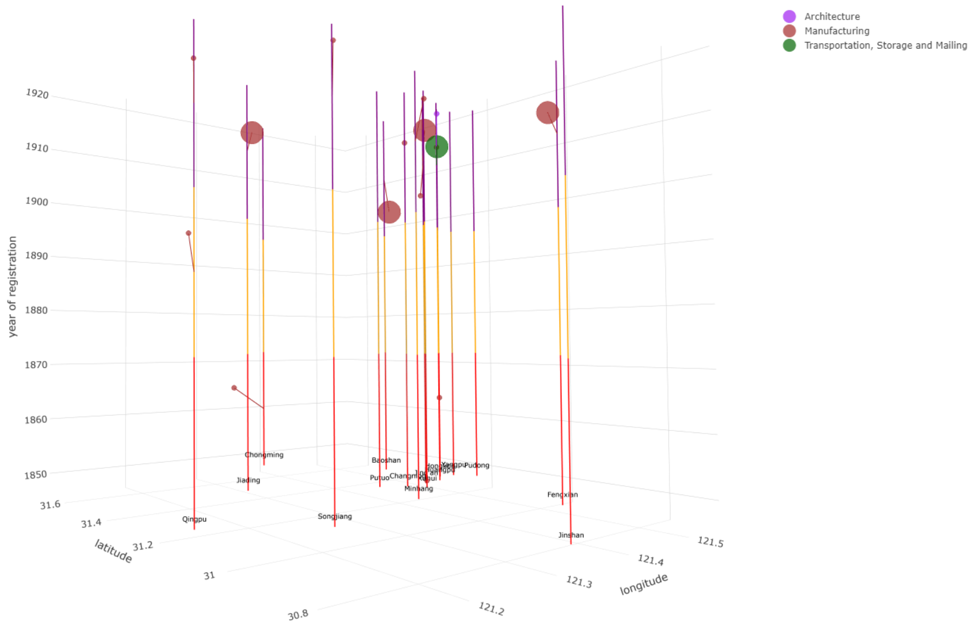


Figure 0.12 A few features (administrative units), lambda = 0.3, 2 timesteps, centerline colorized

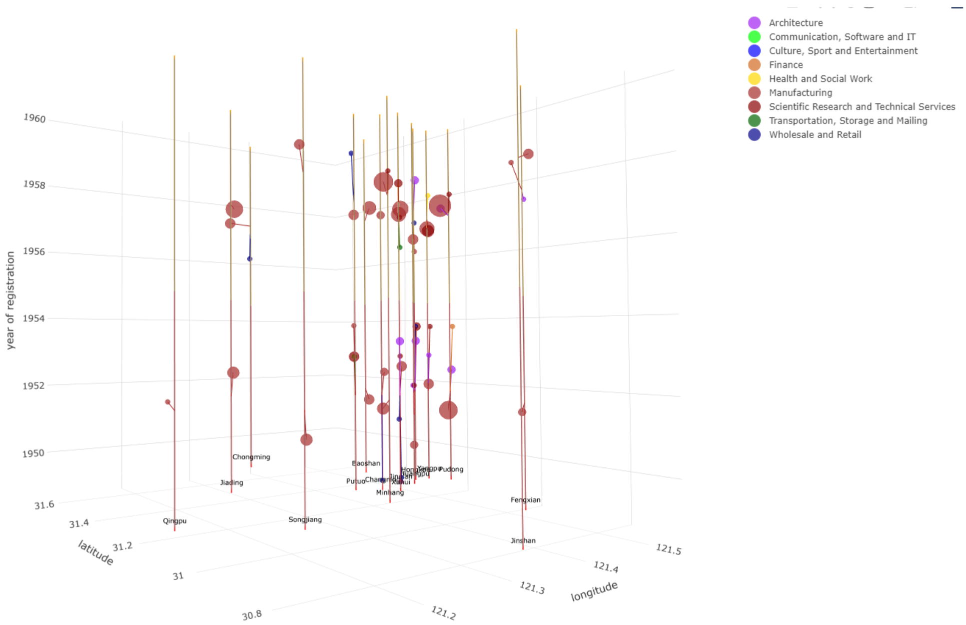


Figure 0.13 Many more features (administrative units), lambda = 0.3, 2 timesteps, centerline colorized

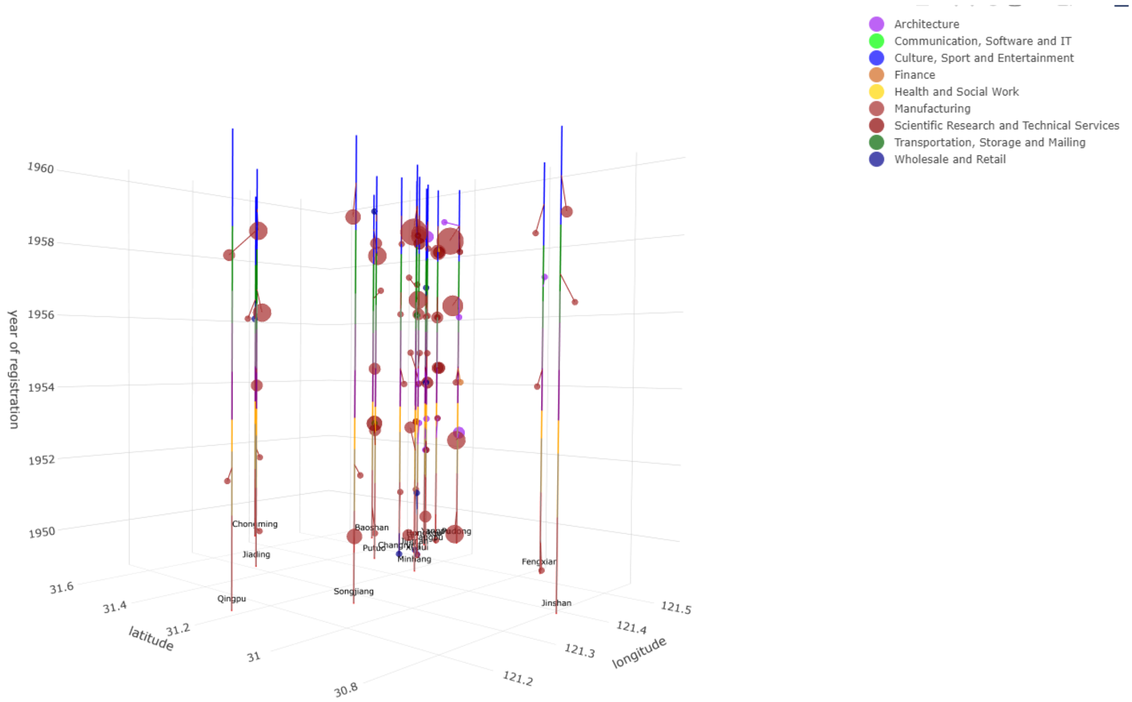


Figure 0.14 Many more features (administrative units), lambda = 0.3, 5 timesteps, centerline colorized

From the screenshots above, the ones in chapter three and chapter four, it can be seen that a small lambda (no less than 0.5) and colored centerlines makes space-time bin series and temporal layers the most distinguishable from each other.

Source codes and data: https://drive.google.com/file/d/1i9TZMpUv7ocoJ3-I2-Qy5q2p1fpN4g18/view?usp=sharing