CityData: A Geographical Data Visualization System For Better City Life

Honglie Hu

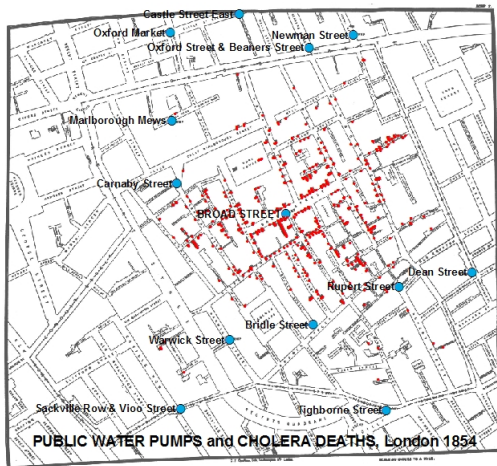December 2017

1 Introduction

Figure 1: Dr. John Snow's map [2]

In 1831, when cholera broke out in London, thousands of people died because of it. Dr. John Snow labelled each death case on Broad Street and found out most death cases happened around the public well in Figure 1. This map reveals the source of cholera and had saved lots of lives at that time. This is one famous and effective example of geographical data visualization's powerful functionality happened at early stage.

At present, human have already developed an solid method to prevent and overcome cholera.

But human's exploration on geographical data visualization is still on the way. Now is an era of information explosion, in daily lives, the quantity of data each person needs to deal with is much bigger than it before. Because of that, data visualization, as an efficient way for people to access information among data, grabs more and more attentions. As for geographical data visualization, besides offering the data a good looking, it can display the spatial distribution pattern in a comparatively direct way, which can tremendously cut down the time of making relative decisions. Furthermore, with geographical data visualization, it's more convenient to dig out deeper further information buried in tremendous geographical data.

1.1 Aim of the Project

After knowing the functionality of geographical data visualization, refer to Dr. John Snow's cholera map, which saved lots of lives at that time, it's significant and necessary to apply such powerful technique to actually benefit humankind and people's daily lives. However, due to social and economic causes, nowadays, there are much more geographical data visualization systems that are applied on scientific research and business rather than on people's daily

1

lives. For example, Business Explorer Map (BEx Map)[3] is a system that deals with business data and solve business problems; and the temporal map system[4] is designed for scientific research purpose and can oversee the energy output from power plants. In contrast, there are a few systems whose main attention is on people's daily lives. Hence, the main aim of the project is to create a system, CityData, that can visualize a city's data, such as the energy consumption data, transportation data and population distribution data, and can support and boost a city's resource integration, thus benefiting people's daily lives in city. But this project only builds a basic version of CityData rather than a full fledged version, and more future works will be needed.

2 Description and Designing

Due to time and resource causes, this project will only build a basic version of CityData rather than a full fledged version, and more future works will be needed. Furthermore, initially, only one kind of data such as population distribution data will be applied to test the system. In order to build the system that can fulfill my goals, the project will be divided into the following four main components:

2.1 Data Processing

The most basic and important part of the project is data, which is the first challenge while making geographical data visualization since datasets can come from heterogeneous sources with various data qualities and formats[1]. Then, the first component is data processing, which includes spatialization of data, data cleaning and data transformation. The basic idea here is to clearly and orderly combine geographical information with corresponding data point. ArcKarma[1] can be a great example and reference of my data processing component. And it employs a training-by-example strategy to generate transformation programs and automatically transforms attribute data to a uniform format. Similar algorithms and programs will be applied in this component.

2.2 User Interface

The motif of the system's user interface is minimalism, and there is no excess information besides necessary select boxes. Also, user-friendliness is always one significant point while doing designing. In short, there are three select boxes: Box A for selecting geographical areas; Box B for selecting topics such as population, energy and transportation; and Box C for selecting time range of the data.

2.3 Data and System

After getting the data and finishing the system's basic structure, next step is to connect these two things together. PostgreSQL, which is an object-relational database management system (ORDBMS), will be used to store all data. Also, in the database, geographical information is always the primary key, and corresponding informations such as energy and population will be stored in the same record. Then, Google Maps JavaScript API will be used to fetch necessary geographical maps and to connect the database with geographical maps.

2.4 Data Visualization

The last step is to do geographical data visualization on maps. So far, since population distribution data will be used to test the system

Figure 2: Estimated Project Timeline

initially, bubble maps is the preferred data visualization implementation. And the size of the bubble chart can display the density of population distribution. The geographical area with a big bubble chart means there is a high density of population, and vice versa. Then, for other data, different but efficient implementations will be applied. For example, structure map will be used for transportation information; and heat map will be used for energy consumption information. And all these geographical data visualization implementations will be built based on Google Maps API.

3 Timeline

In order to streamline the project, a Gantt Chart (See Figure 2) is created to display how the work is arranged in next semester.

References

[1] CHIANG, Y.-Y., WU, B., ANAND, A., AKADE, K., AND KNOBLOCK, C. A. A system for efficient cleaning and transformation of geospatial data attributes. In *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2014), SIGSPATIAL '14, ACM, pp. 577–580.

[2] JOHN, S. Gis analyses of snow's map. https://www1.udel.edu/johnmack/frec682/cholera/cholera2.html. [Online; accessed 3-Dec-2017].

[3] MOHRAZ, K. Geographical analysis in sap business information warehouse. In *Proceedings of the 8th ACM International Symposium on Advances in Geographic Information Systems* (New York, NY, USA, 2000), GIS '00, ACM, pp. 191–193.

[4] RODRIGUES, N., NETZEL, R., ULLAH, K. R., BURCH, M., SCHULTZ, A., BURGER, B., AND WEISKOPF, D. Visualization of time series data with spatial context: Communicating the energy production of power plants. In *Proceedings of the 10th International Symposium on Visual Information*

3

Communication and Interaction (New York, NY, USA, 2017), VINCI '17, ACM, pp. 37–44.