Literature Survey

Tae Hyon Lee

December 14, 2018

Abstract

With the exponential growth in the number of mobile devices and apps to help collect user data such as location and time, there has been growing opportunities to analyze these user data and understand the user's behavior by performing an in-depth analysis with statistical methods and visual techniques such as spatio-temporal visualization. The development of new software and customizable libraries has provided us the means and capacity to easily display user data such as spatio-temporal data with a more sophisticated, customizable visualization technique. In this survey, we examine previously researched techniques for visualizing spatio-temporal data and identify some of the weaknesses of the techniques and suggest how we can make an improvement with today's technology. Lastly, we will explore some of the designs of visual and statistic analysis that can be used to understand and explain patterns and anomalies that can be found from spatio-temporal visualizations.

1 Introduction

Spatio-temporal visualization, also known as geo-temporal visualization, is a visual representation of geographical data such as coordinates that is portrayed over time. Automatic and anonymous collection of a large quantity of detailed spatio-temporal data was not common as it requires accurate recording of both location and time for a long period of time and for a large number of participants. There is often a limit to manual collection of spatio-temporal data as it can be prone to human error and bias. Consequently, it can be difficult to explore certain fields of the area where we need to collect a large quantity of detailed data without any bias involved. However, in the recent years, with the growing dominance of peer-to-peer riding services such as Uber and Lyft, mobile devices can now easily perform an automatic, wide collection of spatial and temporal data of its users. And with the increasing availability of these large quantities of the data, there is a growing opportunity for analysts to visualize them with new techniques that can fully take advantage of new types of variables collected with today's technology. However, it is important to note that when researching these new techniques, one should always take readability into account as it has become increasingly more important to focus on readability for the audience with different backgrounds when presenting a visualization in a business environment [12]. For example, a visualization has to be simple and informative enough so that key decision makers who often do not have a technical background can understand and grasp the key findings of the analysis in a reasonable amount of time. Especially with the growing popularity of user-friendly business intelligence tools such as Tableau, analysts have to compete with the readability of such tools to create aesthetic visualizations that deliver useful insights from complex data.

1.1 Background

With the growth of mobile devices and applications, many companies such as Amazon and Google are able to collect massive user information for a wide range of purposes. One of the purposes is to use these data to understand the users' behavior so that they can provide a better user experience. This practice can be seen by the two major internet advertisement companies such as Facebook and Google that collect user information and interactions on a daily basis to display ads to the users that are most relevant to them. The same applies to peer-to-peer rideshare companies such as Uber and transportation rental companies like LimeBike and Scoot as they collect user information and try to understand their users' behaviors and patterns to provide better services such as faster availability of drivers or bikes.

1.2 Scope of Work

The survey will be limited to recent spatio-temporal visualization techniques that use maps or similar visualizations to display traffic related spatio-temporal data.

1.3 Outline

First, we will go over some of the previously well-known visualization methods, such as 3D GIS and heat map, that can visualize spatio-temporal data over time. Next, We will identify some of the key weaknesses of visualizing spatio-temporal data in a 3D format and explain how using a variation of heat maps can overcome those weaknesses. We will then discuss some of the designs and methodologies that can be used to identify patterns and anomalies through visual analysis. Lastly, we will explain how statistical analysis and machine learning algorithms can be used to explain the patterns and anomalies found from the visual analysis.

2 Spatio-temporal Visualization Techniques

2.1 Visualizing Traffic Flow

Traffic flow data are complex in that it contains both the geographical and temporal components in each recorded movement. As a result, it can be challenging to effectively visualize traffic flow. As an attempt to tackle this challenge, a simple and expressive visual technique was introduced to visualize spatial data on a geovisualization and separate dot plot with time filters as shown in Figure 1 [7]. The strength of this approach is that you can see the change in the duration of trips on two different days but it is hard to compare the duration of trips for two different destinations where there is a lot of overlap and because red generally dominates blue color. Moreover, although this approach can help identify traffic movement patterns for one origin and two or three destinations, it fails to provide an overall change of traffic flow in all parts of the city.

Another approach to visualizing spatio-temporal data is to focus on displaying traffic flow properties such as location, direction, and intensity as shown in Figure 2 [13]. This approach can be useful to get the sense of traffic density in major parts of the cities and general direction of traffic but it fails to capture the change in density and the flow of traffic over time to be able to identify any kind of anomalies or patterns.

Zoning of similar areas and comparing traffic flow and density by color and thickness of the line is shown in Figure 3 [5]. This approach can be useful when it comes to which streets are dominated by the traffic flow from which zones of areas over time. But it is not possible to compare by how much the streets are occupied by the flows from different zones. The line graph shown in Figure 3 simply replaces the zone dominating each group of streets.

All three of these approaches in attempting to visualize spatio-temporal data have its strength and weaknesses but for a simple but effective spatio-temporal visualization to be created it must have the following properties:

1. The visualization must capture the overall goal of the visualization: display change of overall traffic over time at a selected area

2. Both the spatial and temporal component of data must be presented on one graph as it might be hard for the audience to understand the correlation between the two or more graphs.

3. There must not be any overlap of lines or dots to compare two types of data as shown in Figure 1 unless you can filter out the overlapping areas.

4. The visualization must be presented in a way that the different zones or areas of the data must be comparable to all other zones or areas in the map.

2.2 3D Geographic Information Systems (3D GIS)

Three-dimensional geographic information system is a visualization system that displays territorial data that have x,y coordinates and as well as z value for height on an x,y,z map. There are a few 3D GIS approaches researched and discussed for visualizing spatio-temporal data [9, 20, 21]. 3D GIS is useful for exploratory analysis of activity-travel patterns as it can uncover patterns and

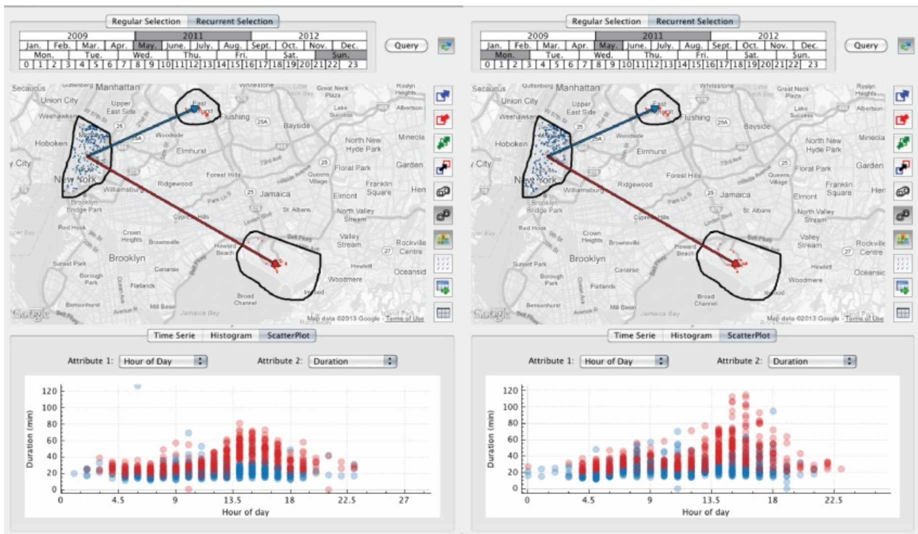

Figure 1: This spatio-temporal visualization shows two separate visualizations. A map that shows the origin and destination of a taxi trip and a dot plot that compares the duration of a trip with a different destination from the same origin. As a whole, the figure compares the two visualizations between the Mondays and Sundays of May in 2011 [7].

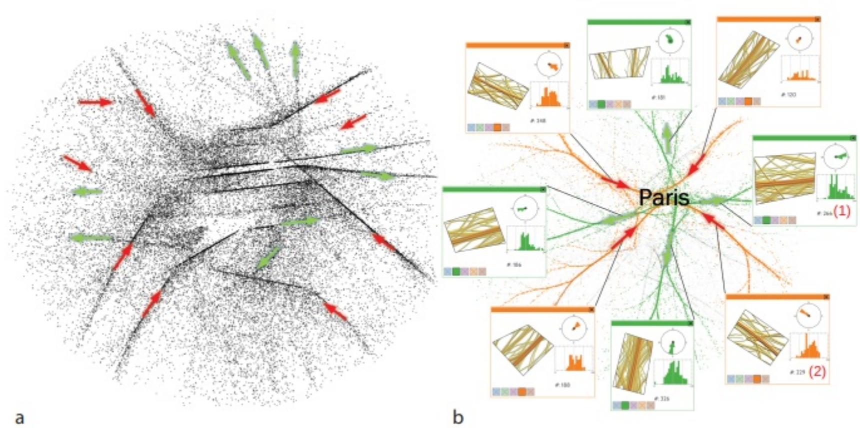

Figure 2: This spatio-temporal visualization attempts to visualize spatial portion (density of traffic) with colors and temporal portion with arrows [13].

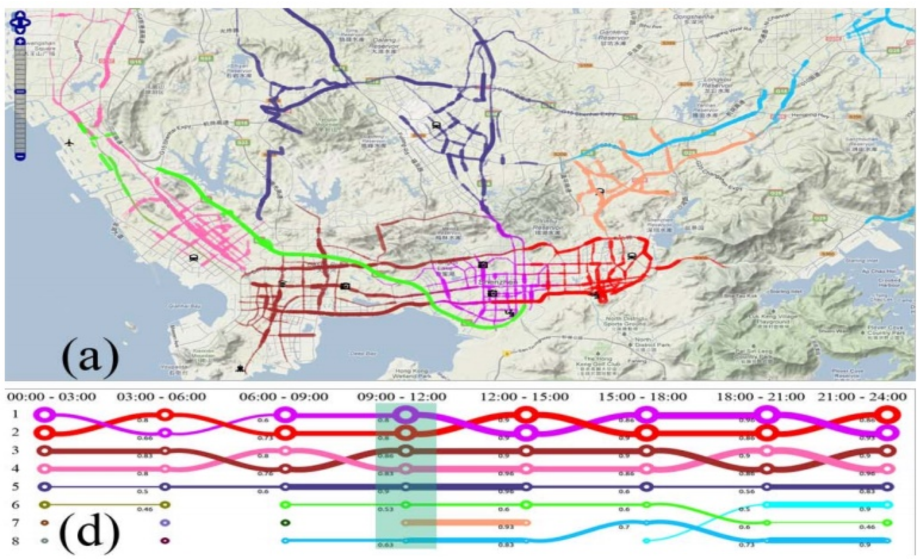

Figure 3: This spatio-temporal visualization utilizes zones to understand the flow of traffic from each zone with the change in time. The number on the left of (d) is a number for a group of street name that is close to each other. The thickness of the line represents the density of the traffic [5].

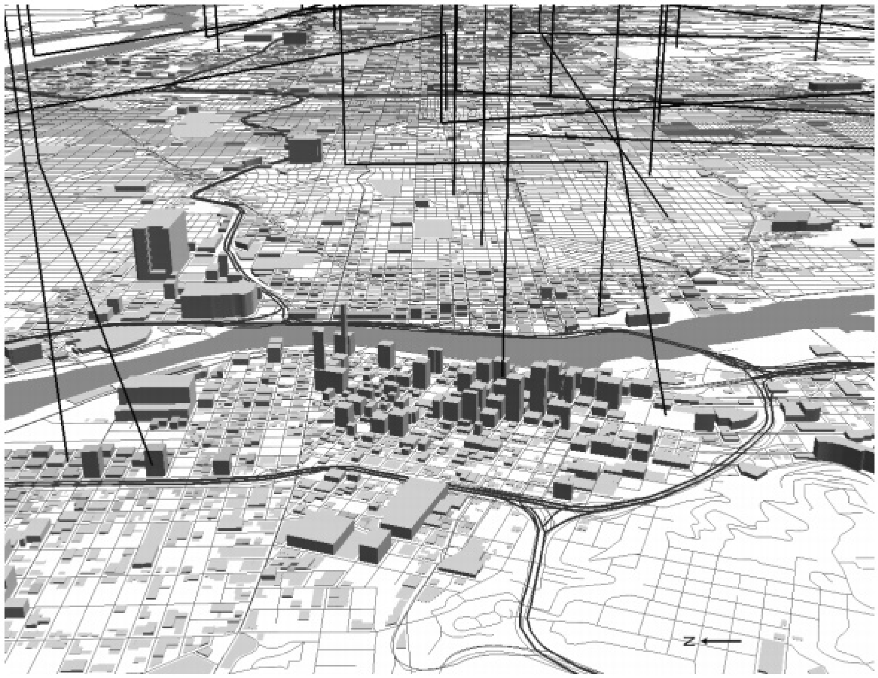

Figure 4: A close-up view of downtown Portland show in 3D GIS. The 3D component of this visualization provides a strong sense about the geographical context through virtual reality-like view [9].

hidden insights on characteristics of space-time activities of different population subgroups. The GIS-based 3D visualization method can also capture the difference and similarities of human activity patterns such as migration among different gender/ ethnic groups as shown in Figure 4.

However, there are many shortcomings with using 3D visualization in both the hardware performance side and the visual understanding side. First, when it comes to large datasets, rendering 3D data can be even more taxing on the machine it is running on than rendering 2D data. And because 3D GIS is an interactive visual tool, the user might experience slow performance. Second, due to the limitation in the human ability to identify patterns when it comes to looking at visualization with multiple dimensions, 3D visualizations might not be as easy to interpret and analyze compared to 2D visualizations.

2.3 Heat Map

The solution to the problem presented in the 3D GIS is using heat maps. A *heat map* is a 2D visualization of data such as pollution, crime, etc that are

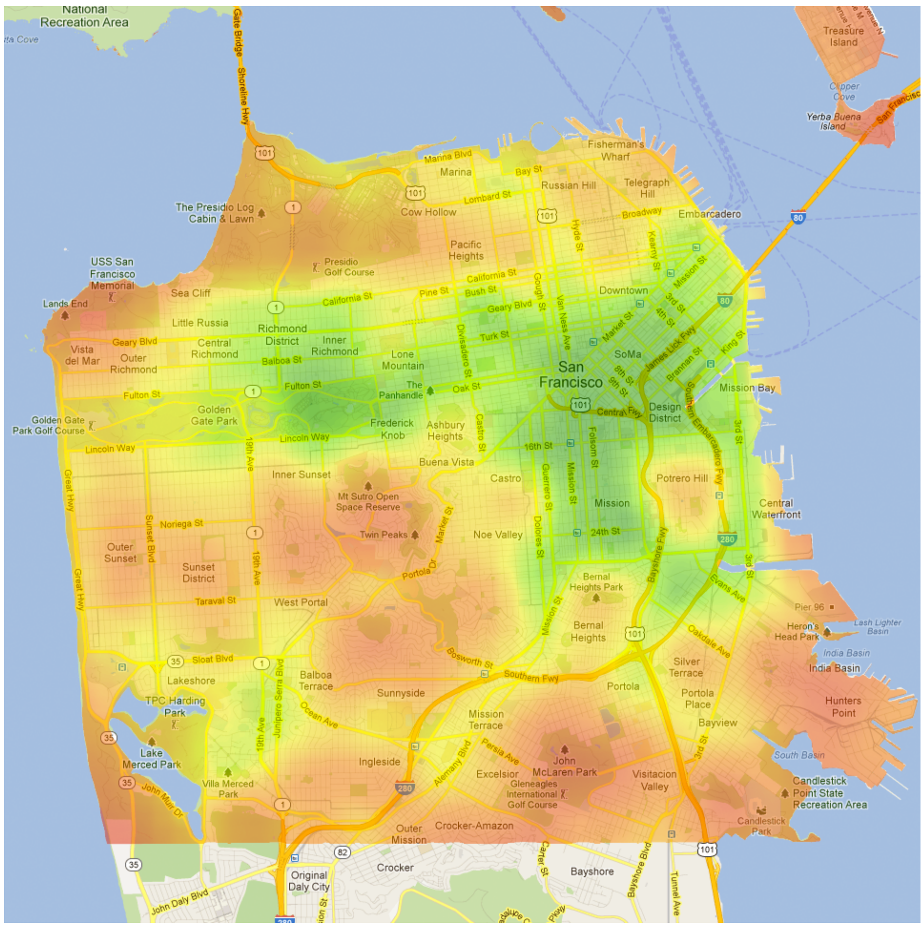

Figure 5: The heat map of ratings of bike routes in the city of San Francisco. The green area has higher ratings of bike friendly routes. Source: https://www.movoto.com/blog/opinions/maps-of-san-francisco/

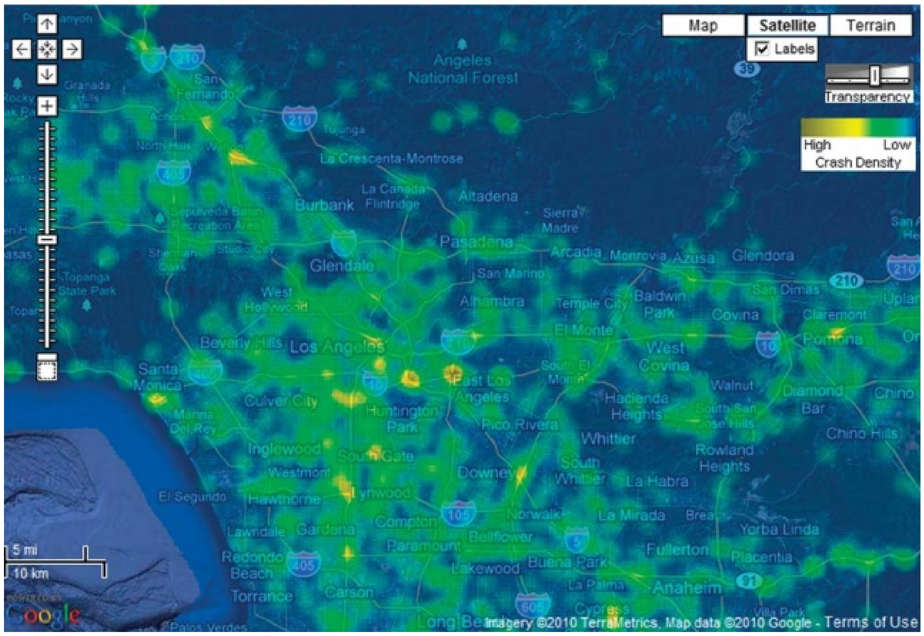

Figure 6: A heat map in SafeRoadMaps that shows the crash density in the surroundings of Los Angeles [8].

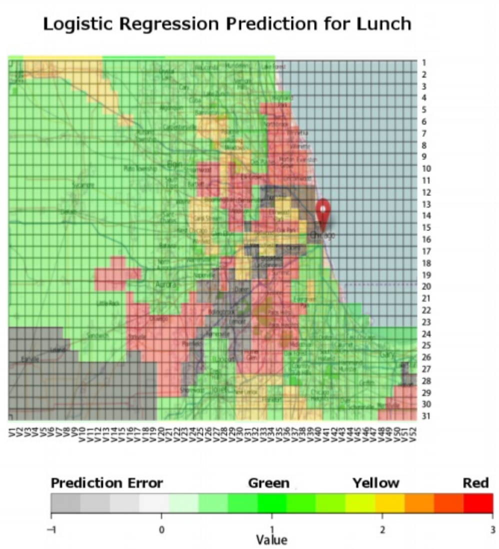

Figure 7: A heat map of traffic density in a certain city [16].

represented by varying colors. An example of heat map is shown in Figure 5. Although they cannot deliver as detailed information as 3D GIS can, heat maps are useful for easily identifying patterns and anomalies, and effectively conveying important information for any type of audience. It is especially useful when you want to compare one variable for different regions in a map or an image. They are often very easy to understand as they have one scale depicted by changes in the color in a color scheme. For example, in figure 5, red and green are used to depict the minimum and maximum value of the crash density in the scale while yellow is used to indicate the middle value. The followings are some of the examples that demonstrate the usefulness of a heat map: SafeRoadMaps [8] and Bing map [16].

In the figure 6, you can see the city view of SafeRoadMaps. In this program, the heat map is used to represent the crash density of a city from high to low indicated by the yellow-green color scheme. Another example of a well-implemented heat map is the heat map shown in figure 7. This heat map uses Bing map and adds grid-like zones to display traffic density of each zone. By looking at the colors of each zone, the audience can easily tell which areas in the map have heavy traffic during lunchtime. In both examples of the heat map, the visualizations utilize two or more distinct colors to depict which areas have high or low values.

3 Human Activity and Patterns

It is important to note that visualizations are not only used for conveying your message to the audience but for analyzing data by identifying visual patterns or anomalies. In this section, we will discuss how we can perform visual analysis to identify human activity patterns and anomalies. Then, we will discuss how we can back up the reasoning behind those patterns and anomalies through machine learning algorithms.

3.1 Visual Analysis

To thoroughly analyze visualizations, analysts need to implement a systematic design that helps us identify patterns and anomalies [3]. First, we need to define types of patterns that can be detected in movement data and data about other phenomena [14]. Then, the analysts need to apply appropriate transformations, computations, and visualizations that are needed to identify the patterns types in spatio-temporal data. Next, they need to identify the spatial and temporal similarity of movement between users to identify movement patterns [2]. For example, analysts can aggregate data with similar traits to help identify and explain obvious patterns within traffic movements. If we cluster data that contains trips of drivers who live in within a certain area and we see that they have similar movements throughout the day, we can generalize about their lives what kind of group they are depending on what kind of movement they make.

Similarly, we can use these patterns in movements to classify geographic space into industrial or any other zones. Finally, there needs to be a measure to detect temporal changes in any variable to uncover seasonal patterns. Through various methods mentioned above such as data transformation, clustering, and measuring temporal changes, we can visually analyze patterns that may not be apparent by looking at the data.

3.2 Human-activity and Machine Learning

The patterns and anomalies found from visualization techniques can be sometimes explained through general knowledge of the data. For instance, if there is a high traffic during rush hour for areas that have a freeway, it can be obvious to say that people are using the freeway to drive to work or to go home. However, if there is a certain area where there is a gradual increase in traffic over a few years, it can be harder to explain since we do not know the exact reasoning. To understand human-pattern and activities, we need other incorporate other information out there such as demographics, number of restaurants and office buildings, or even average housing price in certain areas. Often, it can be a combination of these variables that make up the reasoning. Although visualization tools such as table heat map can be used to compare the combination of two variables, it can be challenging to visualize and analyze multiple variables altogether for advanced analysis. This is where statistical analysis and machine learning comes in to help understand the data better.

One example of analyzing human movements and behaviors through taxi services is shown in "Study on spatial and temporal mobility pattern of urban taxi services" [4]. In this paper, the author discusses some of the statistical methods that can be used for understanding human-activity by analyzing taxi service data. Their findings reflect and give hints about the level of human activity and some properties of social networks. With the help of statistical and machine learning methodologies such as linear regression, logistic regression, Bayesian classifier and support vector regression with the addition to other types of related data, it can be easy to identify some of the significant variables that impact the density of flow of traffic [1, 5]. With the availability of open source libraries such as Scikit-learn and pandas, it has become possible to analyze large sets of data with just a few lines of code [11].

4 Conclusion

Overall, we think that this methodology will be very useful not only for identifying pattern in visual analysis but also for walking an audience through the analysis in a business setting. The heatmap designed to be simple yet insightful for a non-technical audience and can be very fitting for helping out in key decision making. We explored some of the techniques researched in the past including 3D GIS but we think that heat map is the most fitting for the specific goal for a simple and informative chart. Also, with the combination of statistical analysis such as a machine learning model can add more details to the analysis and explain anomalies and patterns found from the visual analysis.

4.1 Future Research Challenges

It might be challenging to acquire recent, quality spatio-temporal dataset and any other related data such as census data to perform a full analysis on human-activity. Another challenge might be finding the appropriate libraries or tools that can help the researchers experiment visualization techniques and analysis designs that take full advantage of the new types of data and meet the current expectation of visualization in the business world.

References

[1] Jinyoung Ahn, Eunjeong Ko, and Eun Yi Kim. Highway traffic flow prediction using support vector regression and bayesian classifier. In *Big Data and Smart Computing (BigComp), 2016 International Conference on*, pages 239–244. IEEE, 2016.

[2] Gennady Andrienko and Natalia Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*, pages 51–58. IEEE, 2008.

[3] Natalia Andrienko and Gennady Andrienko. Designing visual analytics methods for massive collections of movement data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 42(2):117–138, 2007.

[4] Genlang Chen, Xiaogang Jin, and Jiangang Yang. Study on spatial and temporal mobility pattern of urban taxi services. In *Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on*, pages 422–425. IEEE, 2010.

[5] Ding Chu, David A Sheets, Ye Zhao, Yingyu Wu, Jing Yang, Maogong Zheng, and George Chen. Visualizing hidden themes of taxi movement with semantic transformation. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific*, pages 137–144. IEEE, 2014.

[6] Titus Irma Damaiyanti, Ardi Imawan, and Joonho Kwon. Extracting trends of traffic congestion using a nosql database. In *Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on*, pages 209–213. IEEE, 2014.

[7] Nivan Ferreira, Jorge Poco, Huy T Vo, Juliana Freire, and Cláudio T Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, 2013.

[8] Brian N Hilton, Thomas A Horan, Richard Burkhard, and Benjamin Schooley. Saferoadmaps: Communication of location and density of traffic fatalities through spatial visualization and heat map analysis. *Information Visualization*, 10(1):82–96, 2011.

[9] Mei-Po Kwan and Jiyeong Lee. Geovisualization of human activity patterns using 3d gis: a time-geographic approach. *Spatially integrated social science*, 27, 2004.

[10] Nasser M Nasrabadi. Pattern recognition and machine learning. *Journal of electronic imaging*, 16(4):049901, 2007.

[11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[12] Jon Salm. The science of infographics: The surprising way the brain processes visuals. *buffer*.

[13] Roeland Scheepens, Christophe Hurter, Huub Van De Wetering, and Jarke J Van Wijk. Visualization, selection, and analysis of traffic flows. *IEEE transactions on visualization and computer graphics*, 22(1):379–388, 2016.

[14] Hansi Senaratne, Manuel Mueller, Michael Behrisch, Felipe Lalanne, Javier Bustos-Jiménez, Jörn Schneidewind, Daniel Keim, and Tobias Schreck. Urban mobility analysis with mobile network data: A visual analytics approach. *IEEE Transactions on Intelligent Transportation Systems*, 19(5):1537–1546, 2018.

[15] Zeqian Shen and Kwan-Liu Ma. Mobivis: A visualization system for exploring mobile data. In *Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific*, pages 175–182. IEEE, 2008.

[16] Anna Izabel J Tostes, Fátima de LP Duarte-Figueiredo, Renato Assunção, Juliana Salles, and Antonio AF Loureiro. From data to knowledge: city-wide traffic flows analysis and prediction using bing maps. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, page 12. ACM, 2013.

[17] Ari Wibisono, Wisnu Jatmiko, Hanief Arief Wisesa, Benny Hardjono, and Petrus Mursanto. Traffic big data prediction and visualization using fast incremental model trees-drift detection (fimt-dd). *Knowledge-Based Systems*, 93:33–46, 2016.

[18] Leland Wilkinson and Michael Friendly. The history of the cluster heat map. *The American Statistician*, 63(2):179–184, 2009.

[19] Chang Yu and Zhao-Cheng He. Analysing the spatial-temporal characteristics of bus travel demand using the heat map. *Journal of Transport Geography*, 58:247–255, 2017.

[20] Hongbo Yu. Spatio-temporal gis design for exploring interactions of human activities. *Cartography and Geographic Information Science*, 33(1):3–19, 2006.

[21] Hongbo Yu and Shih-Lung Shaw. Exploring potential human activities in physical and virtual spaces: a spatio-temporal gis approach. *International Journal of Geographical Information Science*, 22(4):409–430, 2008.