VISUALIZATION FOR EXPLORATORY ANALYSIS OF SPATIO-TEMPORAL DATA

by

HASAN SERDAR ADALI

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabancı University

August 2012

© HASAN SERDAR ADALI 2012
All Rights Reserved

VISUALIZATION FOR EXPLORATORY ANALYSIS OF SPATIO-TEMPORAL DATA

Hasan Serdar Adalı

EECS, M.Sc. Thesis, 2012

Thesis Advisor: Assist.Prof.Dr.Selim Balcısoy

Keywords: Heat Map, Data Visualization, Spatio-Temporal Data, Visual Analytics

Abstract

Analysis of spatio-temporal data has become critical with the emerge of ubiquitous location sensor technologies and applications keeping track of such data. Especially with the widespread availability of low cost GPS devices, it is possible to record data about the location of people and objects at a large scale. Data visualization plays a key role in the successful analysis of these kind of data. Due to the complex nature of this analysis process, current approaches and analytical tools fail to help spatio-temporal thinking and they are not effective when solving large range of problems.

In this work, we propose an interactive visualization tool to support human analyst understand user behaviors by analyzing location patterns and anomalies in massive collections of spatio-temporal data. The tool that we developed within this work combines a geovisualization framework with 3D visualizations and histograms. Tool's effectiveness in exploratory analysis is tested by trend analysis and anomaly detection in a real mobile service dataset with almost 1.5 million rows.

UZAM-ZAMANSAL VERİLERİN KEŞİFSEL ANALİZİ İÇİN GÖRSELLEŞTİRME

Hasan Serdar Adalı

EECS, Yüksek Lisans Tezi, 2012

Tez Danışmanı: Yrd. Doç. Selim Balcısoy

Anahtar Kelimeler: Isı haritası, Veri görselleştirmesi, Uzam-zamansal veriler, Görsel Analitik

Özet

Konum sensörü teknolojilerinin gelişmesi ve bu sensörlerden elde edilen verilerin kullanımının yaygınlaşması ile birlikte, uzam-zamansal veri adı verilen, hem konum hem de zaman bilgisi içeren veri setlerinin analizi çok daha kritik bir hale gelmiştir. Özellikle ucuza maledilebilen GPS sensörlü cihazlarının kolay erişilebilirliği sayesinde artık büyük çaptaki insan topluluklarının ve objelerin pozisyonlarını kayıt altında tutabilmek kolaylaşmış ve analiz amacıyla depolanan bu uzam-zamansal verilerin miktarını çok yüksek boyutlara ulaştırmıştır. Veri görselleştirmesi, depolanan bu verilerin etkili analizi için gereken yardımı sağlamada kilit role sahiptir. Ancak uzam-zamansal verinin analiz sürecinin karmaşık yapısından dolayı, günümüzdeki görselleştirme yaklaşımları ve araçları ile istenilen düzeyde hızlı bir analiz yapmak, her durumda mümkün olmamaktadır. Bu çalışmada, çok büyük kapasitedeki uzam-zamansal verilerin analizine katkıda bulunabilmek ve verinin içerdiği patern ve anormalliklerin tespiti amacıyla, çeşitli görselleştirme tekniklerinin bir arada bulunduran interaktif bir araç sunmaktayız. Coğrafi görselleştirme, histogram ve üç boyutlu teknikler içeren bu aracın efektifliğini ölçmek amacıyla yaptığımız çalışmalarda, Türkiye'de hala kullanılmakta olan bir mobil servis uygulamasının verilerinden faydalandık.

Acknowledgements

First and foremost, I wish to express my sincere gratitude to my supervisor Asst. Prof. Dr. Selim Balcısoy for his guidance and advices. He inspired and motivated me to work in this project. Without his encouragment, I would have never managed to finish my work.

I am honored to have Berrin Yanıkoğlu, Cemal Yılmaz, Erkay Savaş and Burin Bozkaya as members of my thesis committee. I am grateful for their valuable review and comments on the thesis.

I also would like to thank all my lab colleagues in Computer Graphics Laboratory, especially Dr. Ekrem Serin, for his great efforts and cooperation in all the studies we made together. I would like to thank Ceren Kayalar, Seluk Smengen, Murat elik Cansoy and Candemir Dğer for their great support and friendship. Last but not least, thanks to Mustafa Tolga Eren who made this project much better with his great vision and expertise.

Finally, I wish to thank my family for always loving and supporting me all the way.

TABLE OF CONTENTS

List of Figures

1 INTRODUCTION

1.1 Overview

"*Exploratory data analysis is detective work–numerical detective work–or counting detective work–or graphical detective work. A detective investigating a crime needs both tools and understanding. If he has no fingerprint powder, he will fail to find fingerprints on most surfaces. If he does not understand where the criminal is likely to have put his fingers, he will will not look in the right places. Equally, the analyst of data needs both tools and understanding (p 1: Tukey (1977))*"

When John Tukey[6] came up with the idea of visually exploring data sets to discover their main characteristics, graphical representation of data had become crucial for data analysis. Nowadays, as the amount of data that we produce gets massive, data visualization field plays more important role in analyzing massive data sets to derive information from them. Considering the type of data and the purpose of analysis, appropriate use of visualization techniques aids users to understand their massive data without using a statistical model or having formulated a hypothesis. In this thesis, we propose an interactive geographic data visualization framework which contains several visualization techniques to help users detect patterns and anomalies in their spatio-temporal data.

Over the past few years, visualization community have worked in closely related problems with the cartographic and geographic information system(GIS) communities. The crossdisciplinary connection between these fields facilitates the visual display and interactive exploration of geospatial data and the information derived from it. Analysis of this geographic information in time and space is becoming a very important subject with the increasing use of location data. Recently, researchers are looking for approaches to deal with the complexities of the current data and problems. In their extensive work, Andrienko and Andrienko[7] emphasize the need of visualization techniques and analytical tools that will support spatio-temporal thinking and contribute to solving a large range of problems. However, due to the sophisticated nature of spatio-temporal data analysis, current visualization techniques and analytical tools[8] are not fully effective and they need to be improved.

To address issues in spatio-temporal data analysis, we experimented several different visualization techniques and proposed an interactive tool to support exploratory analysis of spatio-temporal data. Most of the spatio-temporal visualization techniques and tools in the litareture are designed for the analysis of trajectory data. On the other hand, geographic data that we will use in our visualizations do not hold any information to identify individual objects. In another words, we do not have any trajectories and our analysis of spatio-temporal data solely depends on the spatial correlation between regions of consecutive heat map representations. Our methods and the tool are designed considering this characteristic of our data.

1.2 Thesis Outline

This thesis propose a geovisualization framework, series of visualization techniques and an interactive tool to help exploratory analysis of spatio-temporal data. Our main purpose is to understand how our geographic data changes over time and which visualization techniques help data analyst find patterns and relationships in data.

Chapter 2 of the thesis reviews the literature of visualization in related fields. Since we work with geographic data , the first part of our literature mentions studies focused on the visualization of geographic data. In the second part, we review spatio-temporal data visualization techniques and tools.

In Chapter 3, we propose a geographic heat map visualization framework. From the representation of geographic coordinates on map to the visualization of data density using heat maps, this chapter clearly describes each step requires to create an interactive framework which will be the basis of our future spatio-temporal data analysis.

Visualization techniques that aim to help users detect the change in data are provided in Chapter 4. We integrated noval visualization techniques as well as traditional approaches into our geovisualization framework so that users will be able to see the change in data more easily.

Chapter 5 presents a tool that combines our spatio-temporal data visualizaiton techniques with other complementary methods for fast and quick exploratory analysis. It explains how users should interact with our tool in couple of use case scenerios from stastical and exploratory data analysis.

Finally, we provide conclusive remarks on the studies and results in Chapter 6. In this chapter, possible future study directions are discussed.

2 RELATED WORK

In this chapter, we will review in detail the visualization concepts and methods of visual analysis that we based our work.

2.1 Visualization

Although visualization has emerged as a new research discipline during the last two decades, it has been an effective way to tell the story of information since the dawn of man. As Tufte explained in his timeless classic studies[9] [10], visualization can be an efficient, coherent and effective way of presenting any kind of information. Visualization is generally classified into two main topics, Scientific and Information Visualization. In the first part of our visualization review, we will show some important examples from both topics that relates our work. Later on, we will be more focused on the visualization of specific data types such as spatial, temporal and spatio-temporal.

2.1.1 Overview

During the last decade, there has been many methods developed in *Information Visualization* to visualize the abstract data without explicit spatial references[11][12]. This type of data can be related to business, demographics or network graphs and it usually consist hundreds of dimensions which makes it almost impossible to be mapped to a display space naturally. Rendering high dimensional data with standart visualization techniques like plots, line graphs and bar charts produce ineffective results. Therefore, novel visualization techniques need to be developed by employing different approaches[13][14][15][16]. The value of these visualization is measured based on their effectiveness and efficiency[17].

In *Scientific Visualization*, the data sets to be visualized are generally 3D geometries or can be understood as scalar, vectorial, or tensorial fields with explicit references to time and space. A survey of current visualization techniques can be found in [18][19]. Often, 3D scalar fields can be visualized by isosurfaces or semi-transparent point clouds rendered with volume rendering[20]. To this end, methods based on optical emission- or absorption models are used which visualize the volume by ray-tracing or projection. Also, in the recent years significant work focused on the topology-based visualization of complex 3-dimensional flow data in aerospace engineering[21]. While current research has focused mainly on efficiency of the visualization techniques to enable interactive exploration, more and more methods to automatically derive relevant visualization parameters come into focus of research. Also, interaction techniques such as focus&context [22] gain importance in scientific visualization.

2.1.2 Time-Series Data Visualization

Another type of data is temporal data in which the elements can be regarded as a function of time. A wide repertoire of interactive techniques that focus on visualising temporal components of data sets is available. Important analysis tasks here include the identification of patterns(irregular or periodical), trends and correlations of the data elements over time, and application-dependent analysis functions and similarity metrics have been proposed in fields such as finance, science, engineering, etc. Visualization of time-related data is important to arrive at good analysis results [23]. Some visualizations represent series of data values along a calendar division to arrange data values according to different temporal granularities[24](See 2.1). It is also possible to comprise different levels of granularity and aggregation to explore patterns at different temporal levels[25].

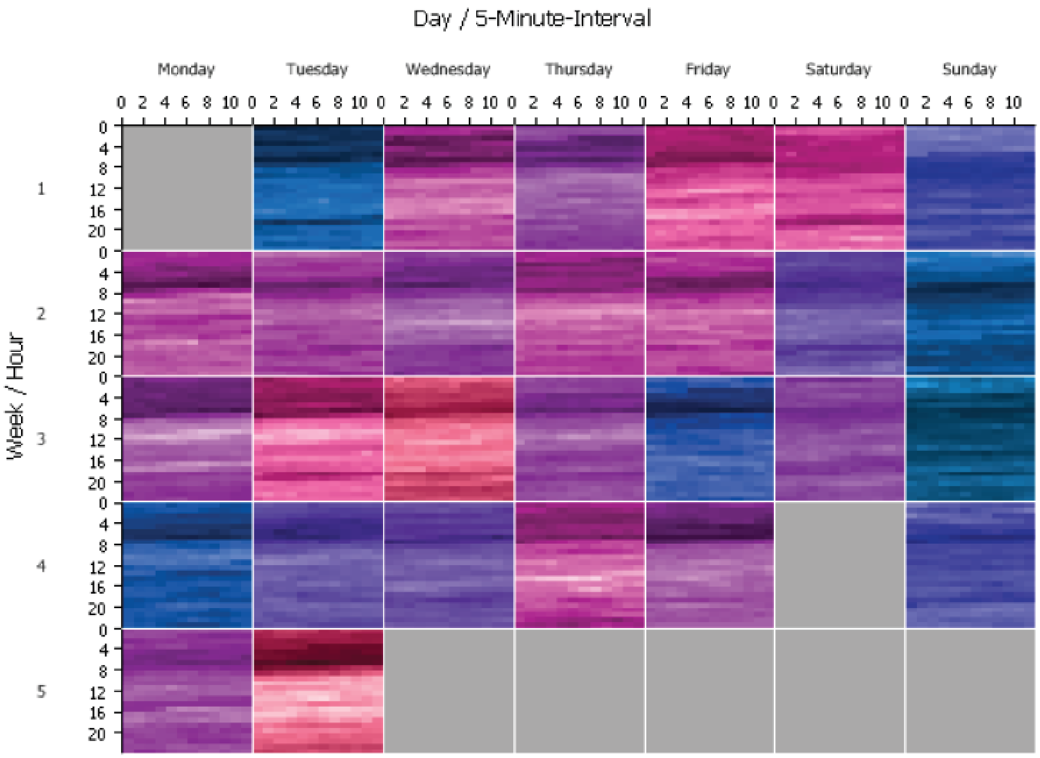


Figure 2.1: Police assignments data that shows the number of deployed units for intervals that are five minutes long. Each block represents one day. Inside the blocks, the hours are shown in rows. Each row has one pixel for every five-minute-interval[1].

2.1.3 Geovisualization

When spatial component of data comes from geographic measurements such as GPS position data, it is classified as *geospatial* data. Finding spatial relationships and patterns among this data is of special interest, requiring the development of appropriate management, representation and analysis functions. Visualization of geospatial data (*geovisualization*) often plays a key role in the successful analysis . Back in the old days, when computers were not so fast and powerful, geographic data can only be displayed as a result of seamless efforts by experts in the geography and cartography. Thanks to recent improvements in the computer graphics and geographic information science(GIS) field, there have been notable studies to visualize and analyze massive geographic data over the past few years. The common goal of geovisualization studies is to speed up the computational processing of geographic data to support understanding by means of novel maps.

In their extensive work, Maceachren and Kraak[26] explores different techniques and the research challenges in geovisualization field. They address the important points of geovisualization such as representation of geospatial information, integration of visual with computational methods, interface design for geovisualization environments and cognitive/usability aspects of geovisualization.

Different cartographic techniques have been used to represent geospatial information . Among many other visualizations that use geographic maps[27], thematic mapping[28] techniques are designed to show a particular theme connected with a specific geographic area. Heat map[29] technique(a.k.a heat density map) is also adopted for geographic data visualization and data analysis in several important examples. Fisher proposes an interactive heat map system that visualizes the geographic areas with high user attention in order to understand the use of online maps(See 2.2). Mehler et. al also uses a geographic visualization technique similar to heat map[30] in which they geographically analyze the news sources. Another interactive framework that takes advantage of heat maps is introduced by Scheepens et al. [31]in which they aim to visualize the trajectory data of vessels.

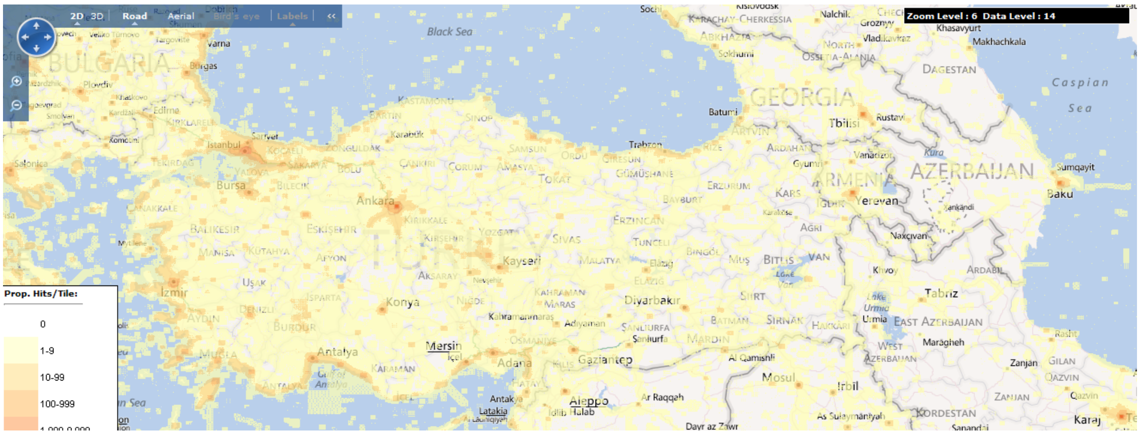


Figure 2.2: A good example of using color to visualize the data density[2]

2.1.4 Spatio-temporal Data Visualization

Perhaps one of the most important and ubiquitous data types is data with references to both time and space. This type of data is often referred as *spatio-temporal data.*

The concept of spatio-temporal data is defined in both GIS[32],data minning[33] and visualization[34]. Visualization of spatio-temporal data involves the direct depiction of each record in a data set so as to allow the analyst to extract noteworthy patterns by looking at the displays and interacting with them. Increasing number of studies on management[35] and analysis[7][36] of spatio-temporal data in the last decade indicate the importance of this data type.

In their analytic review, Andrienko[34] et al. discuss various visualisation techniques for spatio-temporatal data, with a focus on exploration. They categorize the techniques by what data they can be used for and the kinds of exploration questions can be asked of them. For visualizing the spatial change over time in data, Scheepens et. al propose an interactive visualization framework(See 2.4)which analysis the trajectory data of vessels to understand their behavior and risks[4]. Several time-oriented visualization methods are also presented [37][3] to analyze and support effective allocation of resources in a spatio-temporal context.(See 2.3)

After the space-time cube method has been revisited for the analysis of geographic data in many works[38][39], it has been used frequently in visualizing spatio-temporal data[40][41]. The space-time cube approach made the idea of using third dimension for representing the dimension of time popular and 3D visualization techniques have been used on visualizing hierarchies that change over time in a geo-spatial context [42].

When the spatio-temporal data sets are very large and complex, existing techniques may not be effective to allow the analyst to extract important patterns. Users may also have difficulty perceiving, tracking and comprehending numerous visual elements that change simultaneously. One way to deal with this problem would be the aggregatation or summarization of data prior to graphical representation and visualization[43][44]. Another idea can be applying data minning computational techniques such as self-organizing map(SOM) to extract semi or fully automatically specific types of feature or pattern from data prior to visualization[45][46]. In these approaches, SOM can group and arrange the regions according to the similarity of

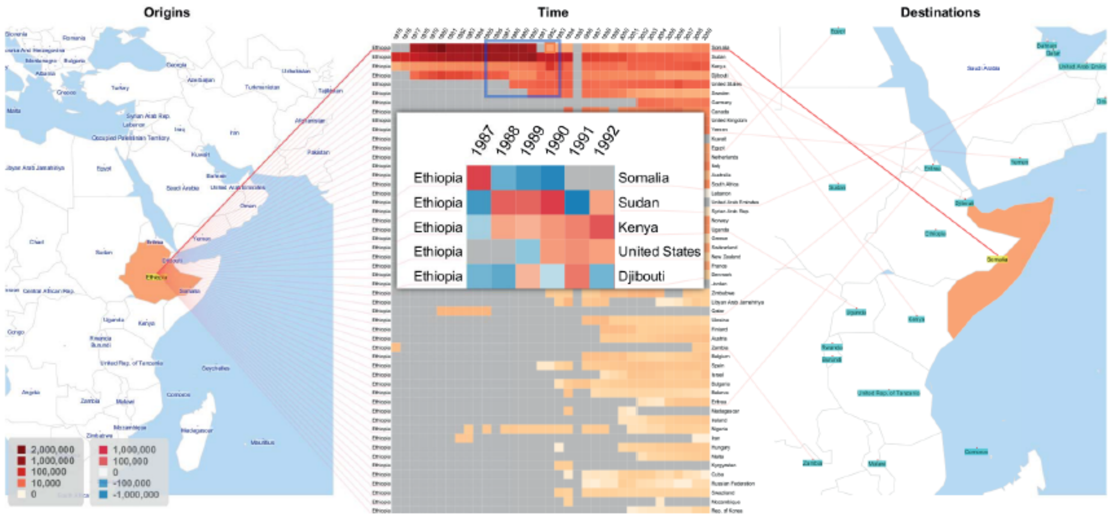


Figure 2.3: In this visualization[3], the origins and the destinations of the flows are displayed in two separate maps, and the changes over time of the flow magnitudes are represented in a separate heatmap view in the middle.

the temporal patterns or the intervals according to the similarity of the spatial distribution patterns. It is also possible to develop projections of large and complex data which move items away from their geographic locations to fill the graphic space more efficiently. Some techniques combine methods from information visualisation and cartography to develop semi-spatial views of large numbers of features[47].

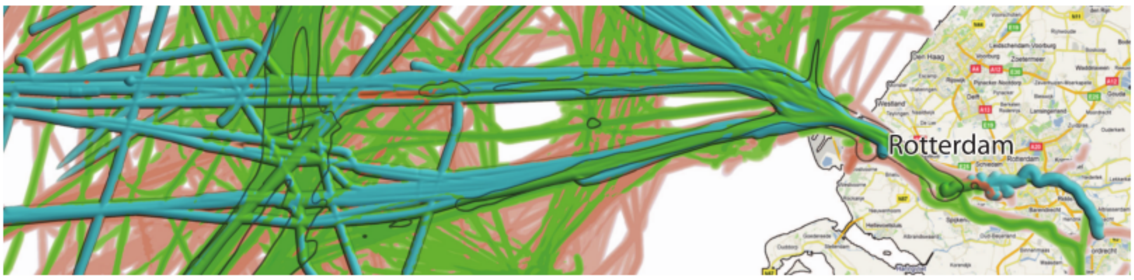


Figure 2.4: Spatio-temporal visualization of vessel trajectories[4]

2.2 Analysis of Spatio-temporal Data

2.2.1 Visual Analytics

Visual analytics is a relatively new term which has the main idea of developing knowledge, methods, technologies and practice that exploit and combine the strengths of human and electronic data processing[5]. Utilizing visual analytics methods can be useful to explore and understand the temporal variation of spatial situations[48]. New approaches in this field can support spatio-temporal thinking and contribute to solving a large range of problems[49].

The analysis of data with references both in space and in time is a challenging research topic. Major research challenges include [50]: scale, as it is often necessary to consider spatio-temporal data at different spatio-temporal scales; the uncertainty of the data as data are often incomplete, interpolated, collected at different times, or based upon different assumptions; complexity of geographical space and time, since in addition to metric properties of space and time and topological/temporal relations between objects, it is necessary to take into account the heterogeneity of the space and structure of time; and complexity of spatial decision making processes, because a decision process may involve actors with different roles, interests, levels of knowledge of the problem domain and the territory.

Combination of analytical approaches together with advanced visualization techniques is the key to success when designing a tool that will support the exploratory analysis of spatio-temporal data. The role of visualization in the knowledge discovery and exploratory analysis process can be seen in Figure 2.5.

2.2.2 Applications and Tools

There have been numerous applications and tools developed that aims to simplify the analysis of spatio-temporal data. Most of them focused on visualizing the movement data(GPS data) to detect locational trends and to analyze different human behaviors. They develop novel visualization techniques for displaying and tracking events, objects and activities within combined temporal and geospatial display [40][51][52](See Figure 2.6(a)). In some cases, systems integrate computational, visual, and cartographic methods for visual analysis and exploration of multivariate spatio-temporal data [53]. Furthermore, combining traditional graphical representations of data such as histograms and scatterplots with novel visualization techniques in an interactive environment help analyst to explore spatial and temporal aspects

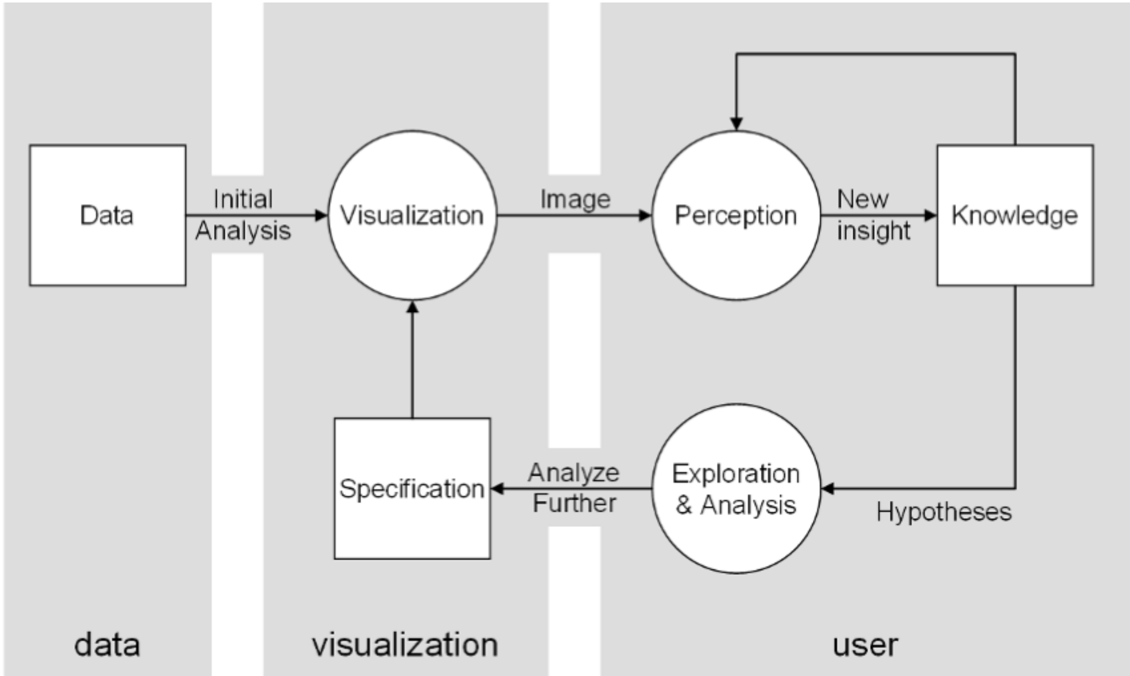
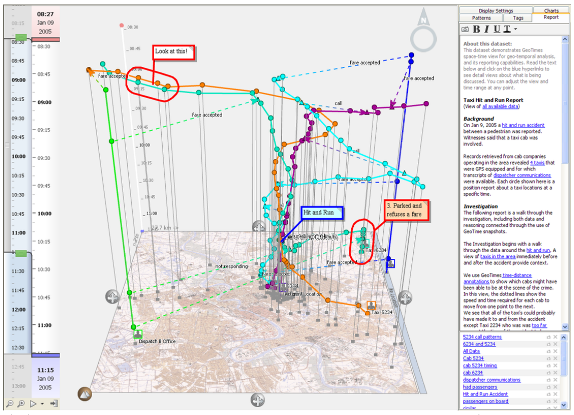
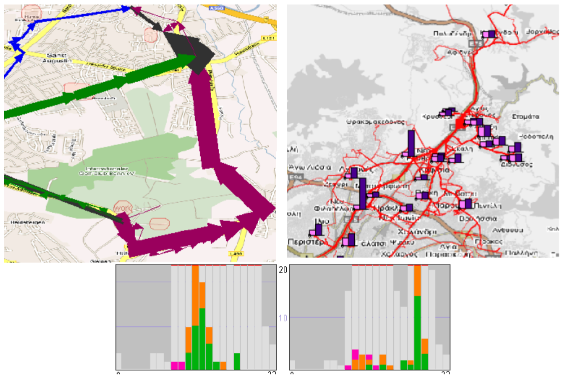


Figure 2.5: The sense making loop for Visual Analytics [5]

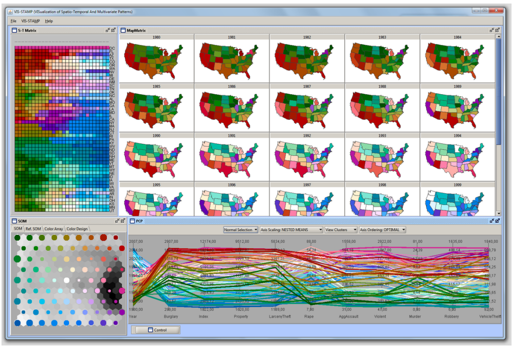
of data quickly[54](Figure 2.6(b)). When there is no movement or any trajectory in spatio-temporal data, it is possible to observe the regional changes.(See Figure 2.6(d) and 2.6(c))



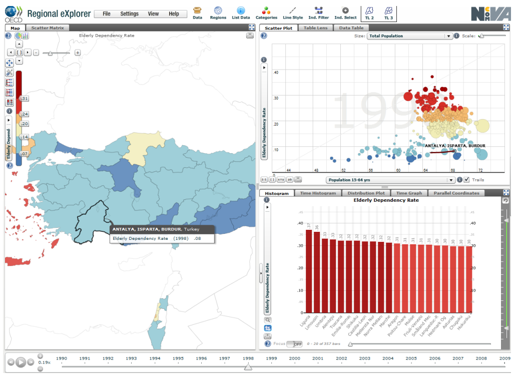
(a) A system to visualize data items (e.g., objects, events, transactions, flows) in their spatial and temporal context[52]. It provides a dynamic, interactive version of the space-time cube concept, where a map plane illustrates the spatial context and time is mapped vertically along the third display dimension.



(b) With the help of histograms and interactive maps, this tool shows the routes between the significant places[54]



(c) The tool[55] based on self organizing map technique to help analysts investigate complex patterns across multivariate, spatial, and temporal dimensions via clustering, sorting, and visualization.



(d) A Web-based interactive visual system OECD eXplorer allows specialists and general public to explore regional statistics data from OECD (Organisation for Economic Cooperation and Development, http://www.oecd.org/home/)[56].)

Figure 2.6: Tools for analysis of spatio-temporal data

3 GEOVISUALIZATION FRAMEWORK USING HEAT MAPS

In this section, we will propose a geovisualization framework which will be used in our future spatio-temporal data analysis. First of all, we will take a look the characteristics of geographic data that we used in our visualizations. Later on, we will explain in detail how we visualize these geographic data using heat maps. After we visualize geographic locations as a dot density map, we will show how clustering them improves our visualizations. Finally, we will talk about the ways to create intensity maps from points which will produce the final heat map representation after colorization step. The main structure of our geovisualization framework can be seen in the 3.1 below.

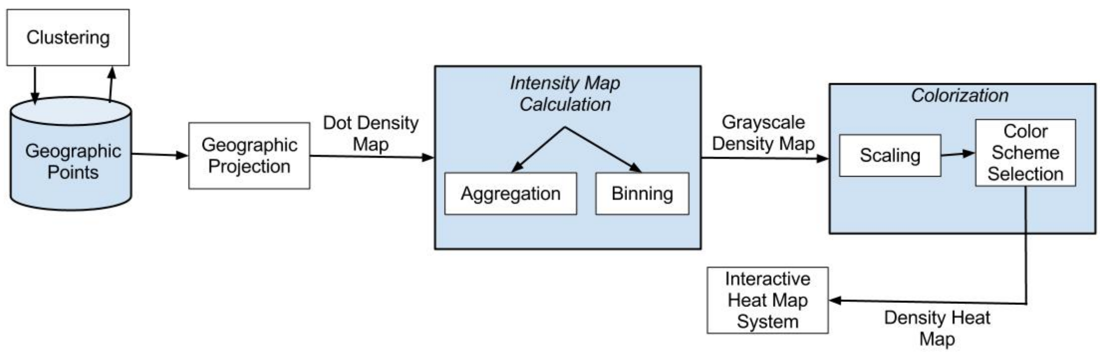


Figure 3.1: Overview of Heat Map System

3.1 Input Data Characteristics

Spatio-temporal data that we visualized in this project come from a real dataset produced by a well-known mobile service. Through this service, users query geographic locations of the other users. Each query returns the time of query as well as geographic coordinates of the queried user. Geographic coordinates in these data are represented by longitudes and latitudes in decimal degrees. We can represent each row of our geographic data as a triple:

Q = (lon, lat, time)

An example row from our actual data would be:

(29.17753056,40.91841389,'2011-07-13 00:00:00.99032')

We classify our data as spatio-temporal data or more specifically, geographically referenced discrete time-series data because it contains time references along with geospatial information. For simplicity in our work, we did not deal with spatio-temporal databases[57]. Instead, we stored results of all queries in a standart MySQL database and let user choose data that fall in any time interval or extent. Our database is composed of the results of 2,404,526 queries that has been done between 02-02-2011 and 01-04-2012.

3.2 Dot Density Maps

One of the simplest ways to represent geographic data is to visualize them as points on the map. This type of geographic visualization is called dot density map. For our geovisualization framework, we implemented a dot density map that works on OpenStreetMap[58](3.2). Integrating dot density map with OSM provides us an interactive framework in which we can zoom in and focus on specific parts of data density as we desired. These type of interactions help data analyst to visually explore data better and faster.

There are some important issues that need to be considered while implementing interactive dot density maps. One is the alignment of geographic points on the map accurately .This issue can be achieved only when the projection of geographic data coordinates and the projection of underlying map system are the same. However, in our case, they are not the same so we need to transform our geographic coordinates.

3.2.1 Coordinate Transformation

The type of projection that is used by OSM (also used by Google Maps[59],Microsoft Virtual Earth[60] and other commercial API providers) is called Spherical Transverse Mercator[61] . This term is used to refer to the fact that these providers use the Mercator projection which treats the earth as a sphere, rather than a projection which treats the earth as an ellipsoid. This affects calculations done based on treating the map as a flat plane, and is therefore important to be aware of when working with these map providers. In order to properly overlay our geographic data on top of the maps provided by the any API providers, it is neccesary to use this projection.

Projections in GIS are commonly referred to by their “EPSG” codes, identifiers managed by the European Petroleum Survey Group. The identifier of our geographic coordinates is “EPSG:4326”, which describes maps where latitude and longitude are treated as X/Y values and identifier of Spherical Mercator is “EPSG:900913” which describes coordinates in meters in X/Y.

To transform our data from lat/lon to meters, we first convert our decimal degrees to radian:

lat = (lat / 180) * Π

lon = (lon / 180) * Π

Then using ellipsoid model constant(sm_a = 6378137.0), we can describe map coordinates that represent our geographic data as meters:

X_m = sm_a * (lon)

Y_m = sm_a * $\log((\sin(lat) + 1)/\cos(lat))$

As a result of this transformation, we sucessfully reproject our geographic data so that they share the same projection with OSM coordinate system. There is one more step before we visualize map coordinates on the screen , that is to convert them to screen coordinates. Convertion process is described as following formula:

res: resolution of OSM

ext: bounds(top,bottom,left,right) of the current visible portion of OSM in meters

$X_p = 1/\text{res}*(X_m\text{-ext.left})$

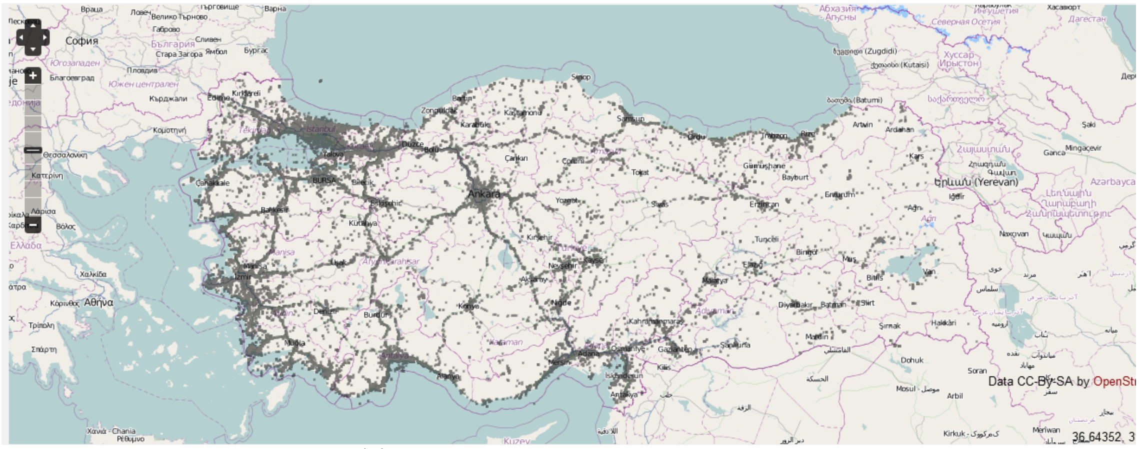
$Y_p = 1/\text{res}*(\text{ext.top-}Y_m)$

The resulting raster image that visualizes the density of a random day is illustrated in Figure 2.

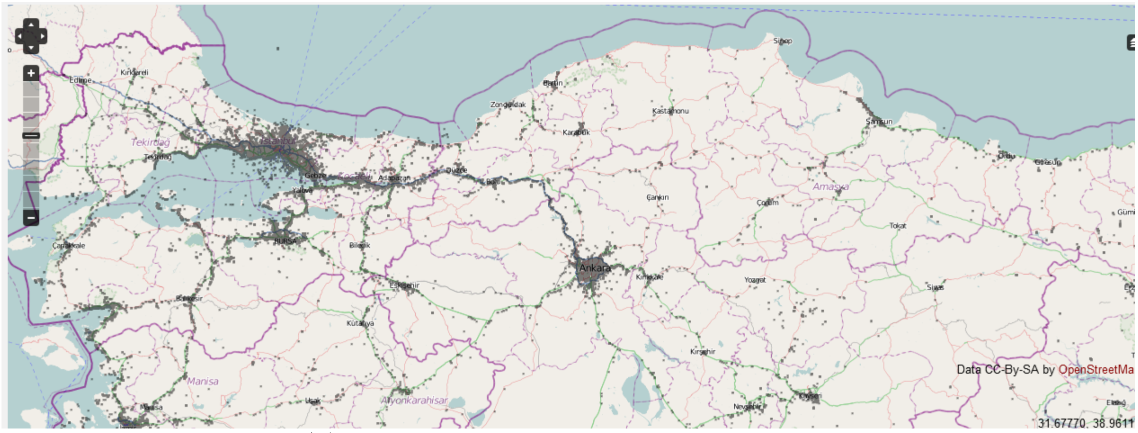
3.2.2 Interaction

Another important feauture in our dot density map visualization is interaction because for a faster and better data analysis, it is necessary to let the analyst interact with their data[ref]. Let's think about the situation in which we want to focus only on data density of a certain area. In our interactive visualization framework, this can be done very quickly by panning and zooming in area of interest. To give our system panning and zooming abilities, we implemented our system so that it automatically reprojects geographic coordinates to the screen coordinates each time when the zoom level or extent of map changes. Figure 3a shows examples of the same daily data with different zoom levels. In addition to panning and zooming, we also let users choose any time period of interest to help them focus on exploring certain time events. Figure 3c visualizes the data density of another day on our system.

Even though geographic coordinates in our data are calculated by multilateration[1] and there might be some errors in the accuracy, our spatio-temporal data analysis will not be so sensitive to the exact location so we can tolerate some error.



(a) Dot density map of monthly data



(b) Same data with more focus on dense areas

Figure 3.2: Dot density map on OpenStreetMap

3.3 Spatial Clustering

Keeping in mind that we are dealing with a large amount of data, almost 2.5 million points, visualizing them using only dots is an insufficient representation of data. Due to the fact that the same points highly likely to occupy the same pixels, dot density map visualization fails to differentiate a pixel with 1000 points between another with only 1 point. Therefore, we looked for ways to improve the dot density map so that it represents data more quantitatively and we come up with spatial clustering. In the literature, clustering has been used for data visualization of large scale data[62]. For our visualization purposes, we clustered independent points on the map and represent the size of each cluster with a circle containing a scaled radius(See 3.3). To group points according to their spatial correlations, we adapted a popular clustering method, K-means ++ clustering[63] for our geographic data. Even though K-means++ works slower than K-means algorithm[64], real time visualization is not our primary goal and super-polynomial complexity of K-means++ has not been so crucial for our system. We can always preprocess data, cluster them offline and save the result in the database for further use.

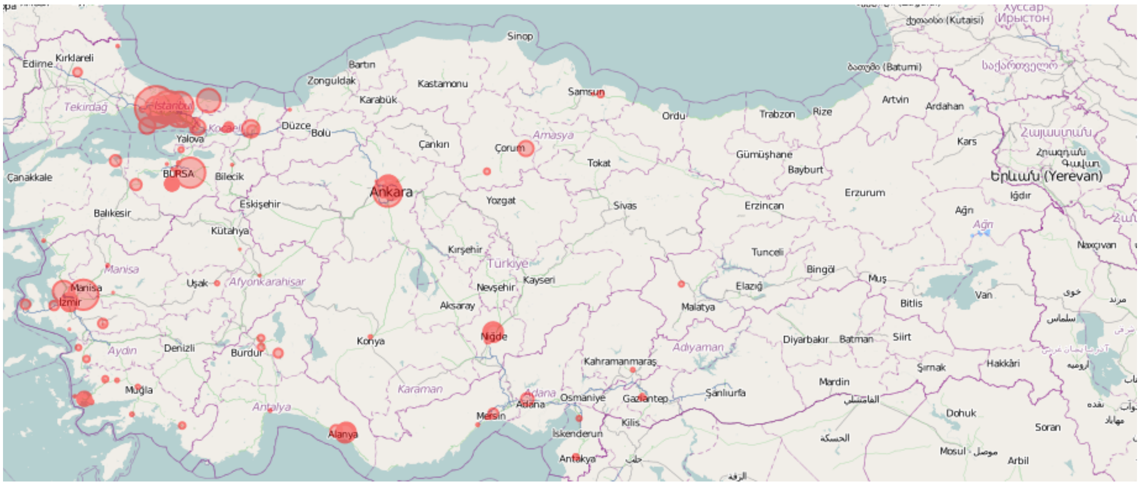


Figure 3.3: Spatial Cluster Map of daily data

Accuracy is more important in our analysis because as we will see in the following chapters, we will try to accurately detect anomalies and patterns. K-means++ algorithm produces more accurate results than K-means by finding the best possible initial set of clusters and maintaining the Euclidean distance among cluster centers as much as possible.

With the help of spatial clustering, we were able to simplify our visualization and represent data densities by the size of clusters. Errorenous data points were filtered and not included in the clusters. However, clustering fails to achieve a continous representation of geographic data on map. Some outlier areas that have few points may not be included in any cluster in the final visualization. For a smooth and continous representation of data, we improved our system in the following sections.

3.4 Intensity Map

The intensity map creation is the core of our geographic heat map visualization. As we know, heat maps use colors to represent geographic data density on the map. Before we map density values to colors, we need to create intensity map of our data which is basically a raster image that we produce by calculating the density value of each pixel on the screen. Calculated density values of pixels are scaled into 0-255 to produce a grayscale intensity map image(3.6(c)) (3.5(b)).

Calculation of intensity values is done by either aggregating data points in vector grids[2] or by kernel density mapping(blending)[31]. We implemented both techniques in our visualization framework and compared them in 3.4.

BINNING	BLENDING
Accuracy depends on the size of grid	Accuracy depends on the radius of radial gradient
Efficient rendering	Slow rendering time
Harder to implement (Requires different structures such as vector grid)	Blending pixels is easy to implement
Vector grid does not work with OpenLayers	Works on limited browsers due to HTML5

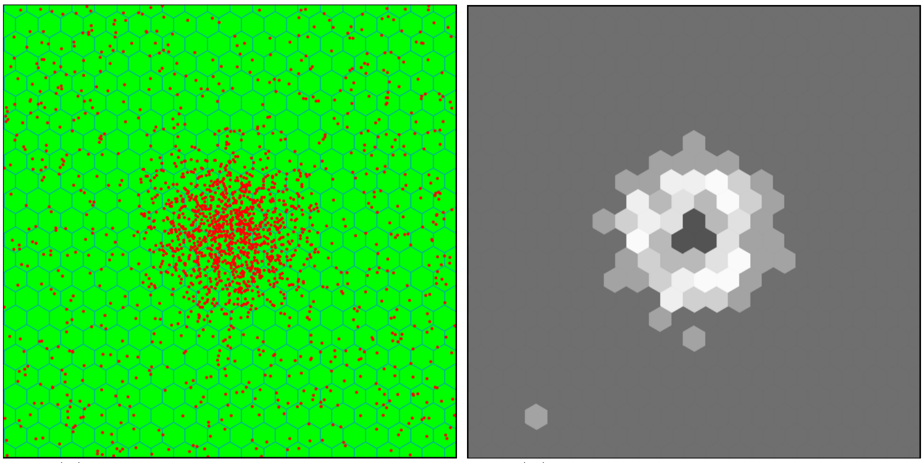
Figure 3.4: Comparison between intensity calculation methods.

3.4.1 Vector Grids for Binning Data

The technique of using vector grids to aggregate 2-dimensional was first described by Clevelend and McGill[65]. They specified that squares be used to bin the data,with each bin then being tranformed into a "sunflower", with each "petal" representing a datum within the bin. Binning is a general term for grouping dataset of N values into less than N discrete groups. These groups/bins may be spatial, temporal or any other attribute-based. In our geographic visualization framework, we adopted spatial binning technique in which we group geographical coordinates (lat/lon) in rectangular or hexagonal grids/bins. Because of their computational simplicity, rectangular bins are usually chosen over hexagonal bins. However we did the opposite for several reasons.

As we mentioned before, we do not need to draw every geographic points on the map in real time, we can render them offline. Instead of computational complexity, we care about smooth and accurate representation of our geographic data. D.B. Carr et al suggested in their paper[66] that hexagonal bins represent data better than rectangular ones and they cite various reasons for this advantage. It has been also observed by Adler[67] that hexagonal grids produce smoother representation of data because they look rounder than squares. Indeed, a regular tessellation of a 2D surface is not possible with polygons of greater than six sides, making the hexagonal tessellation the most efficient and compact division of 2D data space. Considering these observations, we prefer using hexagonal grids to aggregate geographic data. Our intenstity map creation algorithm that uses hexagonal grids is as follows:

- Create hexagonal grids of a specific size on the map
- Construct hexagonal grid data structure by determining corresponding hex in the grid stores hexes for each geographic coordinate in the dataset to be binned.
- Calculate the density of points for each grid and scale them into 0-255 for a grayscale image



(a) Geographic points in a grid (b) Intensity map after binning

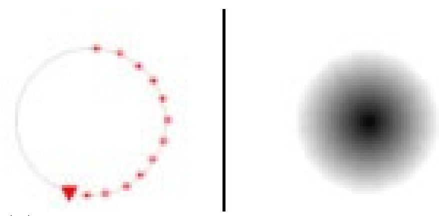
Figure 3.5: Intensity calculation with hexagonal binning

Binning can be good for both the users and developers of interactive thematic maps or other visualizations. As we have seen in the dot density map, showing every single point can lead to cognitive overload for the user and it may even be inaccurate, as overlapping points lead to a misreading of density. On the developer side, binning presents an advantage to the system in efficiency via reducing the number of points drawed on the screen. Additionally, a binned representation may reveal data patterns not readily seen in the raw point representation of the data.

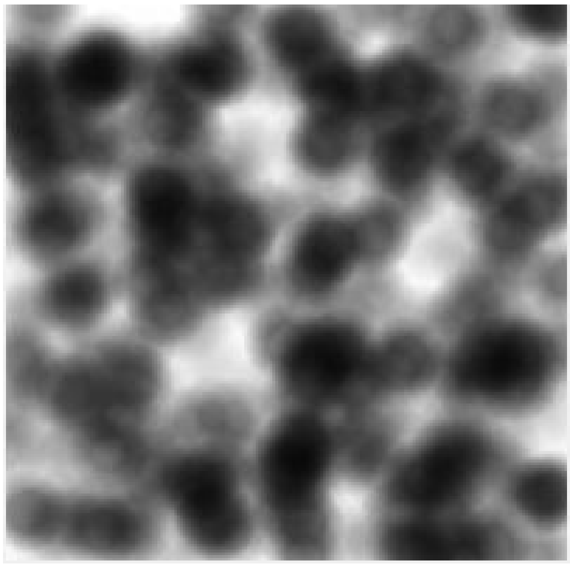
3.4.2 Radial Gradient Blending

Another way of calculating intensity values of a geographic data would be using a fundamental data smoothing problem called kernel density estimation[68]. In our work, we will implement a similar data smoothing technique in which we represent each geographic location as a radial gradient, a filled circle which has full intensity in its center but its value decreases out of the center according to a specific routine. As circle reaches outer radius, intensity value becomes zero for that point. Choice of radius is up to the user, small radius produces more detailed visualization but it reduces the efficiency of system.

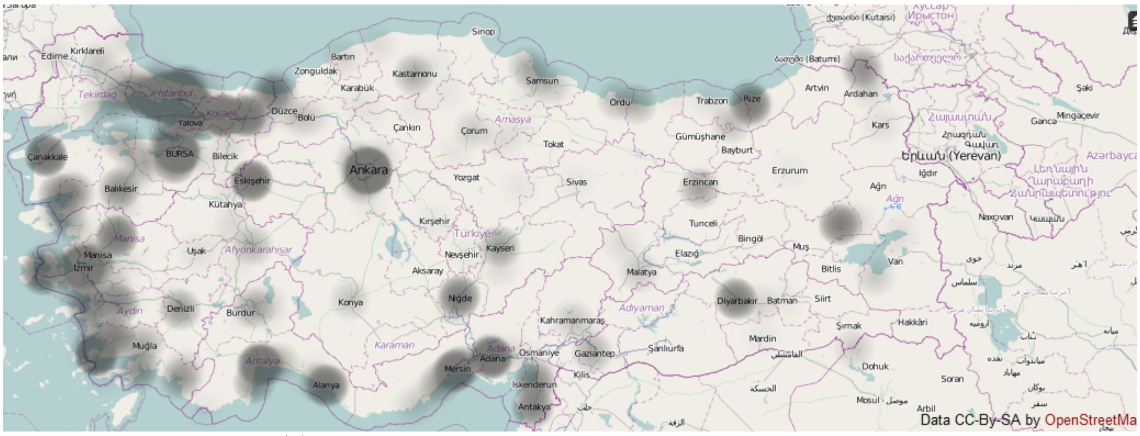
To calculate the intensity of every pixel with this method, we use additive blending technique in which we sum intensity values of geographical points that occupy the same pixel and scale the intensity result to a continuous interval. In order to visualize the intensity map as a grayscale image, we scaled the intensity value of each pixel between 0-255.



(a) Radial gradient representation of a point



(b) Blending radial gradients



(c) Grayscale intensity map of our geographic data

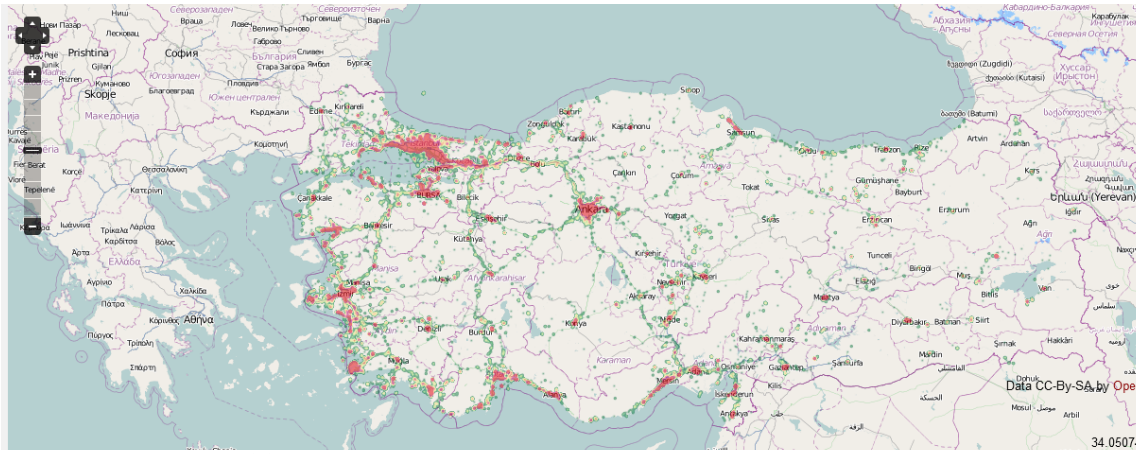
Figure 3.6: Intensity map calculation with radial gradient

3.5 Colorization

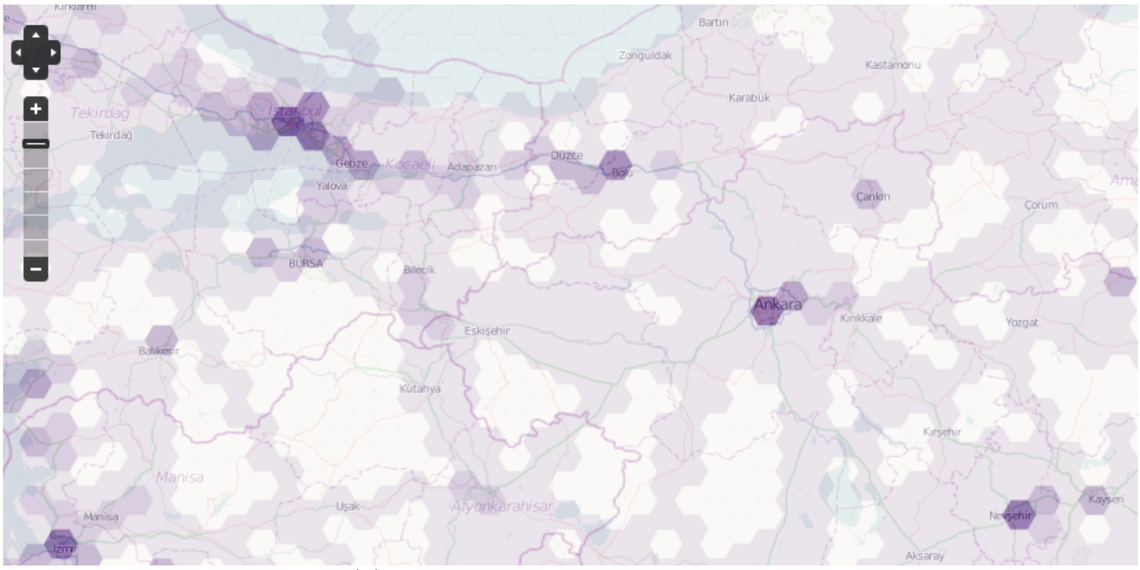
After we calculate the intensity value for each pixel on the screen, we finalize heat map visualization by mapping these values to a certain color scheme for a better visual representation. This colorization process involves choosing the appropriate mapping function to assign different colors to different density values. The mapping of density values to color values is arranged in our framework through the use quantitative scaling techniques[69]. Depending on the distribution of data, limitations of the system and user's desire, any one of these scaling techniques can be used.

Quantitative scales are functions that map continous intensity values to a specific color range. Linear scales[ref] are the most common example of quantitative scales and a good default choice if user has no knowledge about the spatial distribution of geographic data. Mapping is linear in a sense that color range value C_r can be expressed as a linear function of intensity value domain I_d as $C_r = I_d + b$. However, we observed that linear scale is not a good choice for our geographic data because of its spatial distribution. The difference between intesity values of dense areas and other areas are so much that linear scaling is able to assign a few color for these areas.

While colorizing heat maps, it is also important to choose a smooth color gradient to percieve different data densities fast and clear. It may be a good idea to use softwares that gives color advices for cartography[70]. Depending on the number of data classes that we choose to visualize, we can pick a color scheme and a color system. We used a color scheme that slelects the number of color classes and puts them in a gradient.



(a) Colored heat map of daily data using radial gradient



(b) Colorized hexagonal grid heat map

Figure 3.7: Colorization

4 VISUALIZATION OF CHANGE OVER TIME IN GEOGRAPHIC DATA

Having successfully visualized geographic data with heat maps, we now turn our attention to the concept of time in our data. In this chapter, we propose series of visualization techniques that might help users analyze geographic data over time for detecting anomalies and outliers. Different visualization approaches such as animating time snapshots of data chronogically, overlaying different kind of maps and representing data in 3D are integrated into the geographic visualization framework that we developed earlier in the previous chapter for a better analysis of change in geographic data.

4.1 Heat Map Raster Animation

Animation is one of the straightforward methods when the task is visualization of change over time[71]. If there is enough number of raster images that represents different time instants, it is possible to show them in series to the users and let them use their human pattern recognition skills to detect any changes in data. Even though, several researchments in the visualization field discuss that animating spatio-temporal data would usually lead to worse results than simply showing all the small multiples, we wanted to experiment usefulness of animation in our spatio-temporal analysis. We add an animation feature into our heat map framework to let it show consecutive heat map images that represents a time period in a smooth way.

In our interactive system, we let users select the time period of interest and the fixed time difference between two consecutive heat map images. Before animation process, the system renders specific number of small multiples (heat map images) for selected time period and frequency. These images are then used as key frames of animation. In case there is not enough key frames for smooth animation due to users choice of time parameters, image morphing[72] technique is applied between density map frames.

As a result of our experiments with animating data, we observed that our cognitive skills do not work fast enough to visually explore entire data when there are too many data points or when change in data occurs in different areas simultaneously.

4.2 Heat Map of Change

In the previous chapter, we used heat maps to visualize the geographic data at any time instant. Since we are now interested in how these data change over time, it might be a good idea to visualize only the change in data using heat maps so that users can observe trends or outliers more quickly. In this section, we will visualize the spatial difference between geographic data of average and another time instant using heat maps.

This requires the calculation of change over time between the density map visualizations that we will compare. After this calculation, we map positive and negative change values to different color schemes for a clear visualization. We used two different methods to compute change in our spatial data. First one is simply finding the difference of each geographical point between two selected density maps. Resulting heat density map representation of this straightforward calculation can be seen on Figure X. Since negative and positive changes can sometimes be overlapped in this method, different color schemes occlude each other which makes reading any information very hard. The other change calculation method would be similar to the use of bins or grids as we did in previous sections. Instead of taking the difference of each individual points in two images, we find the difference of points that lie the same grid in two images. Although this gives a rough result, it is a better approach in a way that it gets rid of overlapping effect. However, grids close to each other can still have positive and negative values. Considering that they will be represented with different color schemes, sharp changes are observed in this (figure X) type of not so smooth visualization. Calculating the difference between any data and average of whole data may also provide interesting results (Figure X). This technique can be used to detect extreme changes.

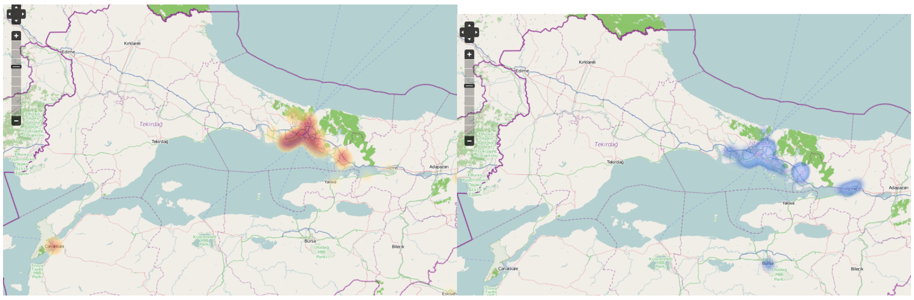


Figure 4.1: Heat map representation of change in data. Red means increase in the areas while blue inditaces negative change.

4.3 Overlaying Maps

So far, we have seen that visually comparing small multiples or animating consecutive density map images are not good enough for extracting information from spatio-temporal data. By superimposing different kind of maps, we aim to produce multivariate visualizations that will allow us to compare our geographic data at different time events. In our method, instead of comparing geographic visualizations seperately, we overlay them on top of each other to help users perceive change faster.

We need consider some important issues before we start overlaying operation. Obviously, we must first choose the time events that will be compared. At this point, comparing two any consecutive representations is the straightforward approach but randomness causes meaning of analyis to be heavily relied on the choice of these two time events. For a more meaningful comparison, we chose one of the overlays as the heat map visualization that represents the average of data and then compare it with the others. However, superimposing two heat map images produce complex and unmeaningful visualizations due to the mix of colors on the intersectional areas. Therefore, we need to represent one of the time events with a different type of visualization which will help users to distinguish and compare two time events clearly.For this purpose, we experimented with visualizations that use bump mapping and isoline techniques.

4.3.1 Average of Geographic Data

Change in data over time can also be spotted by users if they compare data representations of different time events. However, comparing random two consecutive geographic data representation would not be succesful at detecting anomalies or outliers unless users choose significant time events. At this point, making one of these representations more meaningful might improve the comparison. For this purpose, it is necessary to look into the mathemetical representation of spatial data to provide a solid background of representations. Here, we will explain the concept of average in our geographic data and how we visualize it for our comparisons. We will adopt a grid technique similar to STING[73]:

R **:** rectangular region of interest with boundries top-left(x_1, y_1) and bottom-right (x_2, y_2)

G : mxn grid representation of R (See 4.2(a))

$G_{m,n}$: grid at mth row and nth column

t_i **:** start time of analysis

t_e **:** end time of analysis

t_d **:** temporal difference between two consecutive time instants

T : set of time instants for visualization where $t_i + k \times t_d \leq T \leq t_e$ and $k\epsilon$

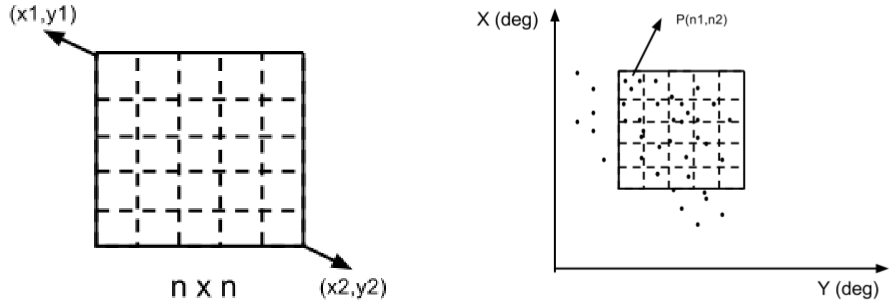
$P_{m,n,t}$: total number of points in G$_{m,n}$at time t (See 4.2(b))

$A_{m,n}$: average of geographic data in grid G$_{m,n}$

Using above definitions, we can define average of geographic data in a grid as:

$A_{m,n} = \frac{1}{T}\sum_{0}^{T}(P_{m,n,t})$

Defining average of geographic data in a grid lets us visualize it using our previous intensity calculation techniques in Chapter 3. With our heat map technique, resulting raster image of a geographic data for a randomly selected time period can



(a) Geographic extent window with boundries (x_1, y_1) , (x_2, y_2) and nxn grids. Here n depends on the resolution of system or in other words zoom level of interactive map.

(b) Scatterplot of total points in time period t1-t2. Extent window is located on the area of interest for average calculation.

Figure 4.2: Geographic data average with grid method

be seen in 4.3.

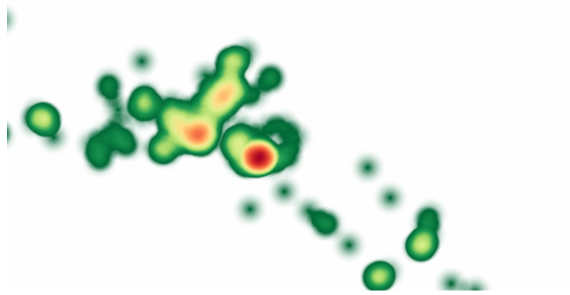


Figure 4.3: Heat map representation of average data from Istanbul

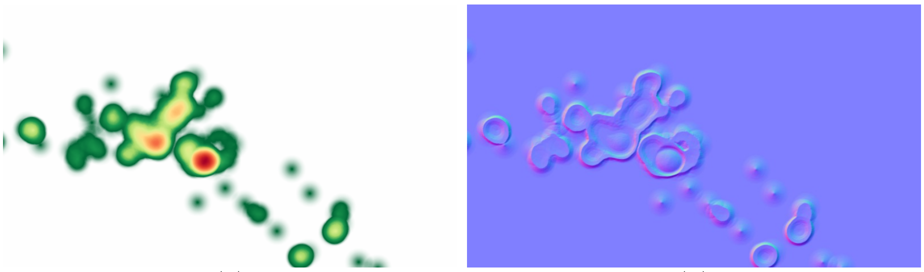
4.3.2 Bump Mapping

For our framework, we implemented a popular technique for rendering more realistic surfaces in Computer Graphics field called Bump mapping[74] to represent geographic data. It gives a surface bumpy look by modifying its surface normal that is used in lighting calculations. We used this technique in our geographic visualizations to see if overlaying bump map with heat map would help us to compare two time events successfully.

Algorithm that we used to visualize our geographic data with bump mapping technique is as following:

1. Take heat map representation of desired time event

2. Store pixel coordinate and intensity value of each pixel of the heat map image in a height map
3. Calculate the modified surface normal map from the height map (See 4.4(b))
4. Combine surface normal of the height map with the surface normal of the true geometry
5. Recalculate the intensity value of each pixel using the Phong Shading Model.



(a) (b)

Figure 4.4: Normal map representation of geographic data

After visualizing geographic data with bump mapping technique, we overlaid it on top of the heat map representation of average data to make a comparison. We also add an interactive point light to the scene to let users discover the changes on any area. As you can see in 4.5, overlaying makes detecting outliers easier without adding complexity to the visualization. On the other hand, this type of visualizaiton is also not suitable for a quantitative comparison between average and the other time event.

Even though bump mapping reduces the complexity of visualization and allows us to discover some outliers in data, we find it not useful to detect any patterns or anomalies because it does not give any quantitative information by itself.

4.3.3 Contour Map

Contour map is cartographic map that consists isolines (also known as contour line) in which the value of any variable remains constant on the same isoline[75]. It is a

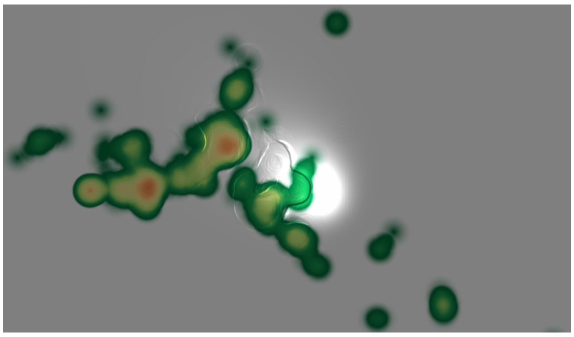
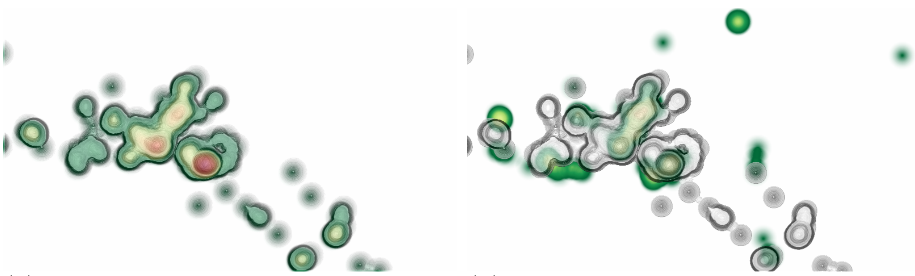


Figure 4.5: Overlaying Bump map with Heat Map

good way for grouping points of equal values, usually height. In most of the cases, isolines can be extracted from triangulated network model of height map representation of data[76]. As we mentioned in the previous section, we already represented our geographic data as height map by replacing pixel density with height value. Therefore, extracting isolines are straightforward using this method. On the other hand, extracting isolines in this manner would result inconsistent representations of the same data with our heat map technique.

In our heat map representation, we grouped pixels with the same(or close) density values using the same color and represent them in a similar way to isolines. By doing so, we will be able to compare heat map and contour map representations of our geographic data to see the difference between the areas in the same group(See 4.6(b)).



(a) When isolines are overlayed on the heat map representation of same data, color transitions become more obvious.

(b) Detecting outliers by overlaying isolines over heat map

Figure 4.6: Overlaying isolines on heat maps

4.4 HeatCube

In previous sections, we only dealt with 2D geographical visualizations for our exploratory analysis purposes. What would happen if we add another dimension(time) to our visualizations and use 3D representation of spatio-temporal data? What would we see in this 3D visualization different than 2D? How much complexity would it bring to our visualization? To answer these questions, we implemented a technique which improves our 2D heat map method in a 3D fashion by using volume rendering. We called it HeatCube.

For the generation of HeatCube, our framework takes advantage of volume rendering which is a useful technique to visualize spatio-temporal data [43]. Using GPU based ray tracing algorithm, we built a system that quickly and automatically constructs volume representation of the heat map images generated in the previous step. These images are stacked on top of each other so that each slice represents the data density of a specific time. As a result of volume rendering, interpolation between slices of heat map help users to perceive the change of geographic data over time in a smooth way. By walking through HeatCube 3D environment with the camera, users view their spatio-temporal data with different angles to detect patterns and anomalies.

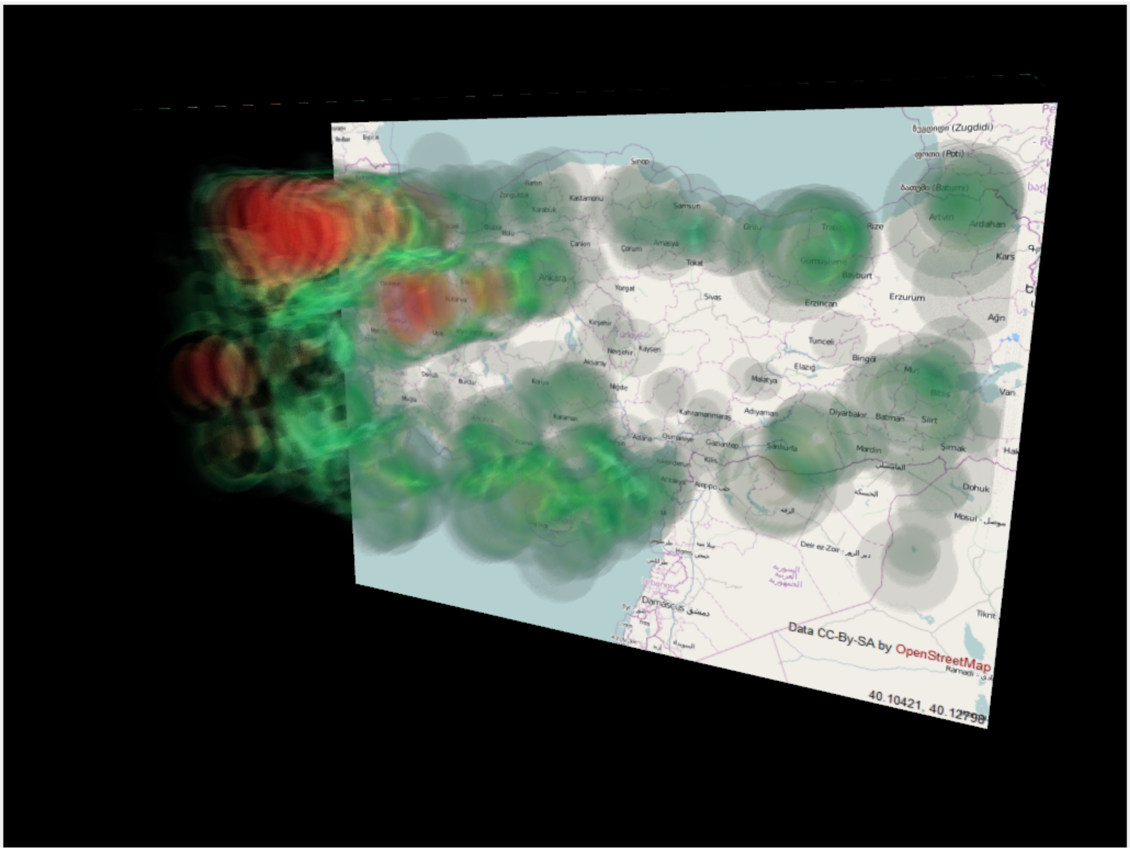


Figure 4.7: An example view from our interactive HeatCube visualization

5 SPATIO-TEMPORAL DATA ANALYSIS TOOL

Visualization systems designed for different exploratory tasks accelarete the analysis of spatio-temporal data[55]. To aid the statistical and exploratory analysis of our geographic spatio-temporal data , we designed a visualization tool that will combine all the visualization techniques that we developed in previous sections. In this chapter, we will describe each component of our tool and we will talk about the possible use cases of our tool in our spatio-temporal data analysis. With use case scenerios, we will try to show the power of such a tool in trend analysis and error detection. The main structure and the possible interactions in our tool can be seen in 5.1.

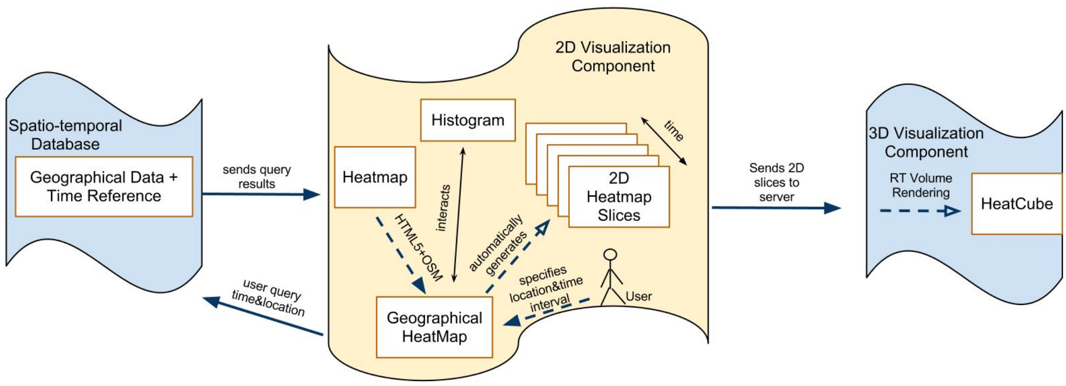


Figure 5.1: Main structure of our tool

5.1 Design

Design process of our tool starts with deciding the necessary operations to be performed on our data. After all, what we want to achieve by our tool is that it should let us interact with our data to perfom cognitive tasks like locating, identifying and distinguishing. An important mantra that we kept in mind when designing our tool is “Analyze first, Show the Important, Zoom, filter and analyze further, Details on demand”. The tool that we designed consists of 3 main compenents which interact with each other to provide faster data analysis(See 5.2). In the following subsections, I will briefly explain the feautures and interactions of these components and the role that they play in exploratory analysis tasks.

5.1.1 Temporal Component

This component is designed only for temporal visualization of data. For consistency with other visualization components, we implemented it in a similar manner to heat map in which we map color density to data density of certain time periods. This visualization is composed of rectangles which corresponds to the minimum time period of analysis that is selected by user. For example, if user desires to observe data density at every hour, we show 24(hour)*30(day) rectangles with appropriate color densities in our visualization.

There are two important features of this component that play critical roles in our tool. One is the ability to show the average of data density. In this component, as user selects the data period for average calculation, we sum the amount of data density for each rectangle in that time period. In order to achieve better results, average value for each time period is calculated by considering density values at the same hour of the same day of week within three months period. This is mainly because each day of week has different usage characteristics in our input data. For example, comparing a Sunday’s data density value with a Friday’s would produce misleading results. After analyst spots the anomalies that needs further attention, he can examine these time periods in more detail using our tool.

Another important feature of this component is its interactive abilities in the system. Picking any rectangle(s) would set the time period of whole system to the period that rectangle(s) represents. Of course, users can also select a continous range with more than one rectangles to analyze different time periods.

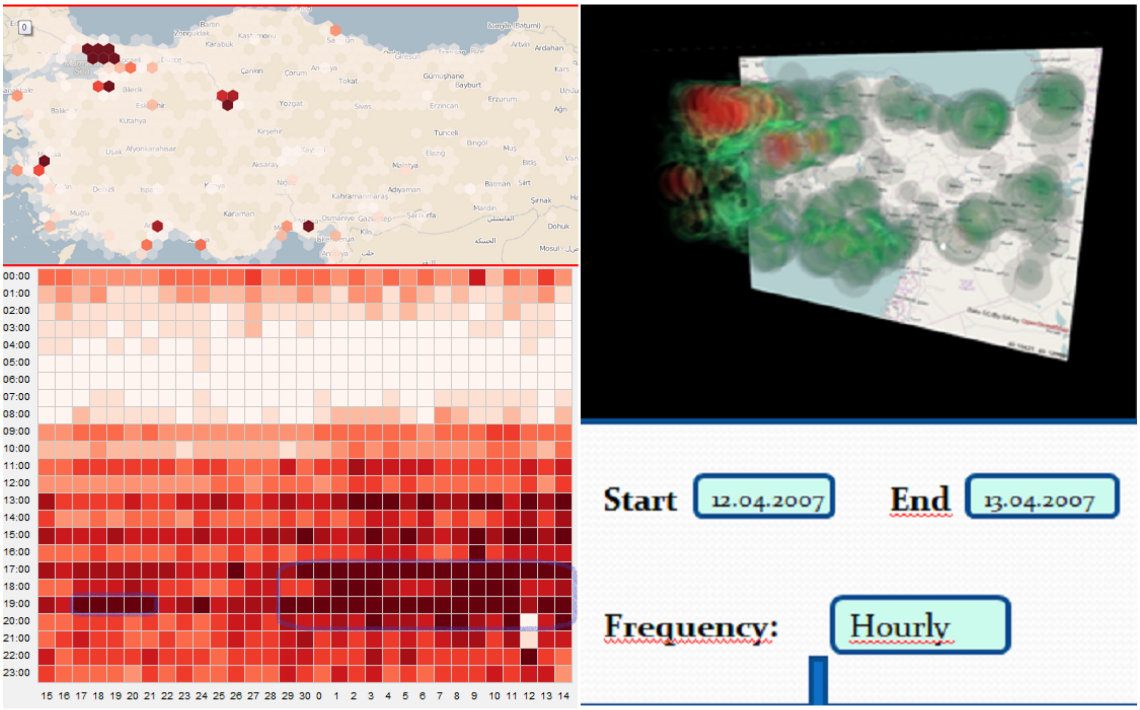


Figure 5.2: Our tool for spatio-temporal data analysis.

5.1.2 Geographic Component

For the geographic visualization in our tool, we used the geographic heat map with hexagonal bins that we described in Chapter 3. This component is able to visualize the average of data in the chosen time period, using the average calculation technique from Chapter 4. This interactive geographic visualization let users locate areas that needs further attention according to their analysis purposes. Interactive hexagonal bins in the visualization improves the efficiency of analysis in a sense that it narrows down the search area for a faster analysis of spatio-temporal data. Selecting these bins will update other components so that they will visualize the data for only chosen geographical area.

5.2 Exploratory Data Analysis Tasks

Exploratory data analysis tasks are focused on finding the answer to 3 questions, "when, what, where", for 3 different components of data[Ref]. In accord with our focusing on the temporal component of our geographic spatio-temporal data, it is appropriate to distinct questions related to time from the others. With this in mind, tasks can be categorized into two cognitive operation, identify and compare.[34].

In this section, we will consider our tool from the perspective of its suitability for the identification and comparison of different objects and time moments. We will describe how each tasks can be accomplished with our tool by keeping "Overview first, zoom/filter, details on demand" mantra in mind.

5.2.1 Identify

Identification tasks have the goal to determine the characteristics of spatial objects or locations at a given moment. For this kind of task, system must be able to visualize data referring to a particular user-selected time moment. A data subset referring to a single time moment does not involve temporal variation and, hence, it can be represented and operated as ordinary time-irrelevant spatial data. Since the histogram view shows the overview of data in temporal aspect, users can use it to decide which particular time moments are more likely to contain the events they are looking for. These time periods may consist outliers, periodic events or even regular patterns. After the time period is selected, users look at the data representation of a particular time moment. At this point, our heat map framework helps users to narrow down the area of interest by choosing a region which will let them observe the data density at particular locations. All of these interactions update every visualization components accordingly.

5.2.2 Compare

When comparing two time moments of spatio-temporal data, an analyst either wants to detect or measure the change in data.

There are many ways to compare time moments and *detect changes* or anomalies by interacting with the components of our tool. As we have mentioned in Chapter 4, one of the most popular techniques for change detection is animating/iterating the map(conseqtutively displaying maps representing situations at different moments) which we implemented in our framework. However, in order to find places where changes occur, one needs to visually scan each map and compare fragments of the maps which is not easy with animation. The overlaying techniques that we defined in the previous chapter can support this process. However, overlaying is only possible for limited number of maps. Therefore, if comparing two different time snap-shots of data is not enough, users of our tool can observe their data in 3D to understand the change in it.

For *measuring changes* of spatial characteristics of our data, a suitable technique could be overlaying. Location changes can be measured when regions of locations are represented on a map via different visualizations such as bump maps or isolines. When there are many regions that change simultenously, it is appropriate to apply filtering so that the map shows only the changes that occurred during a selected time interval. Change heat maps can also be good for estimating changes in thematic properties expressed by numeric attributes. For accurate evaluation of the amounts of change, a change map may be combined with the direct lookup technique: the user points on an object/ location on the map with the mouse cursor, and the corresponding amounts are displayed on the screen.

Trend Analysis

Trend means long term movement,tendency or rate of change in a time series[Ref]. In statistics, trend analysis often refers to techniques for extracting underlying pattern of behavior in well-ordered dataset which would otherwise be partly hidden by noise data. On the other hand, spotting spatial trends with stastistical methods is not trivial when dealing with geographic time-series data. Data visualization techniques allow analyst to see the trend in data before stastistically analyze it and they do it by taking advantage of human eye ability recognize trends quicker than any other methods.

Most common way of analyzing the trend in spatio-temporal data is taking averages over a certain period and seeing if these averages change with time. Users can spot spatial/temporal trends and patterns in data with the following interactions between different components of our tool:

1. User selects the initial time period of trend analysis. This will initiliaze the visualization for heat map framework.

2. With the help of histogram visualization, user detects a possible temporal trend in time-series data. The other components will be updated so that time period that includes a trend will be the new time period for all visualizations.

3. User sets the time difference between each consecutive time event. As a result, heat map framework produces series of images that represents each time event in the new time period.

4. User has an option to animate the produced heat map raster images for detecting spatial trend visually.

5. If visual exploration fails using animation, user take advantage of our heat map framework's ability to display heat map of change in data for the current time period. Dense areas in new heat map representation will indicate an area with possible trend.

6. If user wants to see the areas with possible trends in detail, he/she can 3D spatio-temporal data visualization. When the area with possible trend is detected, user can use 3D visualization

Additionally, a flow diagram for detecting trends using our tool is in Figure.

Outlier and Anomaly Detection

In statistics, if an obsevation appear to distinguish numerically from other members of data, it is an outlier[77]. Outlier values in a data set can be easily detected using mathematical formulas. For our data, we are more interested in spatial outliers, geographical referenced objects whose non-spatial attribute values are significantly

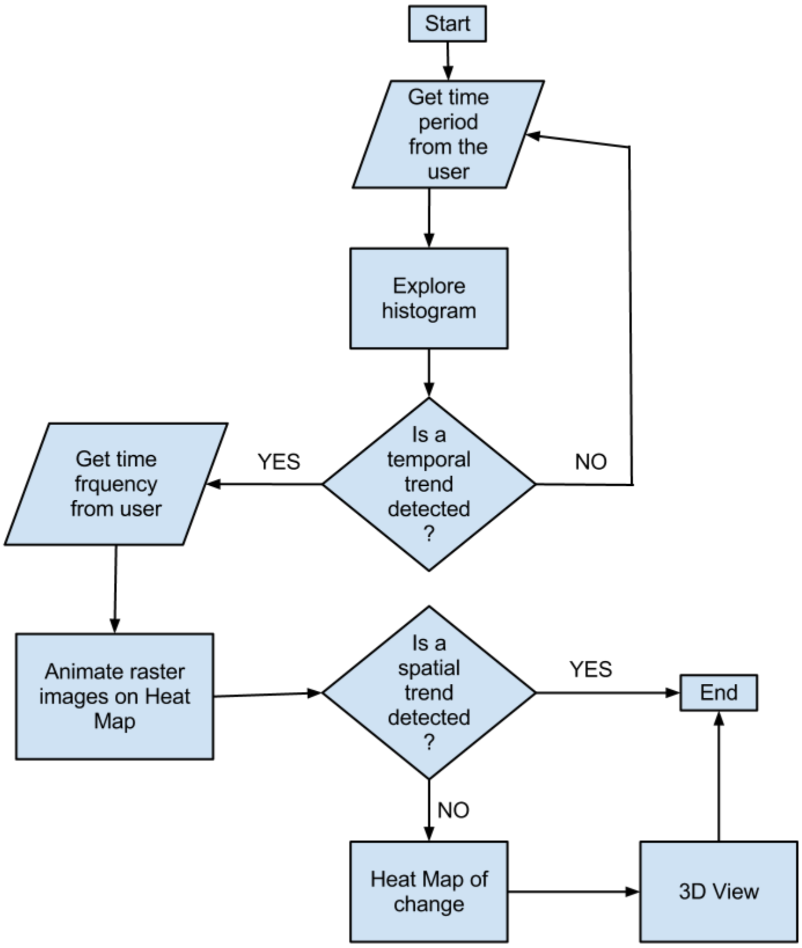


Figure 5.3: Trend analysis routine using our tool

differerent from those of other spatially referenced objects in their spatial neighborhoods. In another words, a spatial outlier is a local instability, or an extreme observation with respect to its neighboring values.

There are several algorithms to detect spatial outliers in spatio-temporal data[78] but when the analysis is too complex annd too many parameters have to be considered, none of them can be as fast as the ability of human brain to detect outliers. Therefore, if we aid our brain with appropriate visualization techniques, it can discover any outlier efficiently and effectively. In the previous chapter, we mentioned that animating heat map raster images is not of the appropriate techniques for outlier detection because of cognitive overload. Instead, using overlaying maps or exploiting 3D visualization of our framework would be more suitable for this kind of task. It naturally allows you to spot seperate areas or volumes which indicates an outlying observation.

6 RESULTS

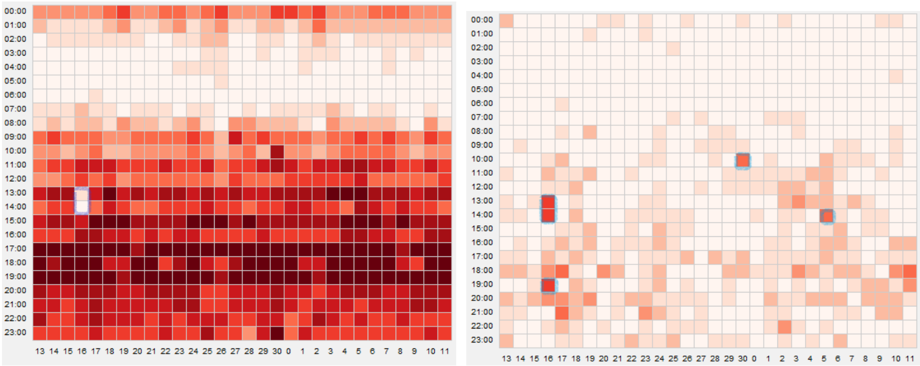
We run two different types of data analysis scenerios with our input data for evaluating the efficiency of our tool. We begin with analysis of anomalies in data, followed by trend analysis.

6.1 Anomaly Detection Scenerio

To test the effectiveness of our tool, we will follow the anomaly detection routine that we described earlier. Before we go through the routine, let us explain the purpose of anomaly detection in this scenerio. With our tool, we will try to help analysts find time periods with system shutdowns as fast and accurate as we can. On the other hand, shutdowns can be local or general which means temporal analysis alone may not be enough for the complete analysis of anomalies. Therefore, to locate areas where shutdowns occured, we need spatial analysis of time varying data.

Temporal visualization using calendar heat map helps us to spot some anomalies in some time periods. As you can see in 6.1, we were able to spot unusual activities in our data between a random time period. We are more interested in time periods with unusual low activity(data density) or no activity at all because they can give us clues about shutdowns. Rectangles with unusual low activities are highlighted in 6.1(a). However, this low/no activity could be a normal behavior for that time period. In our input data, density gets very low everyday between 02:00-06:00 due to the fact that most of the users of service sleep. So, switching to the average view could be useful for spotting anomalies(shutdowns) in the system. In this view(See 6.1(b)), analyst is able to see the difference between average density value and actual value for each time period. After analyzing density values in more detailed time period(e.g. minute), if there's still some activity in the time period, this might mean the shutdown was local and further analysis needs to be done to locate the areas where anomalies occured.

Using geographic visualization in our tool, we were able to examine the data density of the time period with possible anomalies. However, visualizing the actual data may itself does not help finding any anomalies(See 6.2(a)). Average view is more useful in this regard. As we can see in 6.2(b), we detect an unusual behavior around Istanbul at that time period. For the detailed spatio-temporal analysis of data at the selected time period and area, analyst can use the 3D visualization which shows the data density change over time.



(a) Data density between 13.06.2011 - 13.07.2011. Possible anomaly is highlighted on the 16.06 between 13:00-15:00

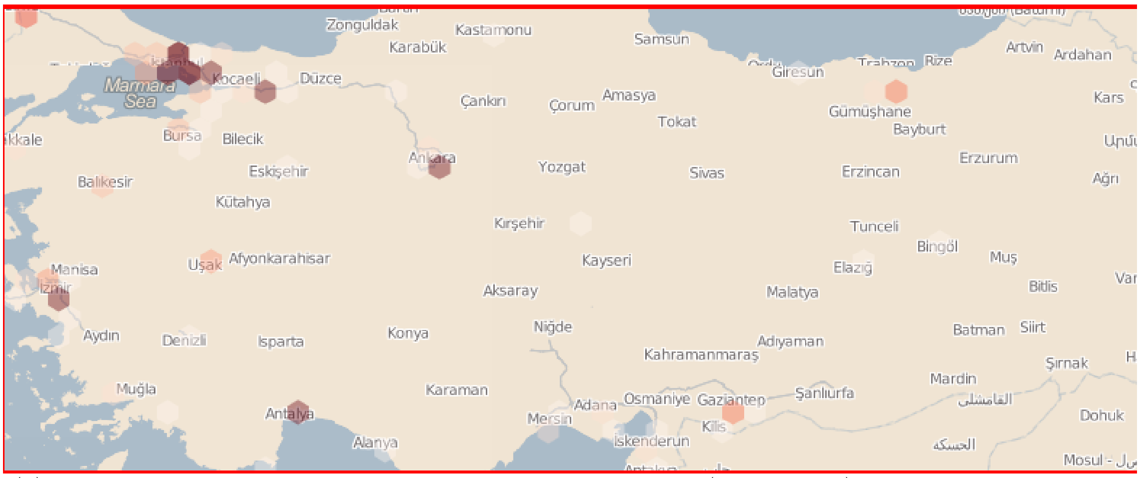
(b) Average view of data at the same time period reveals some other anomalies

Figure 6.1: Temporal anomaly detection

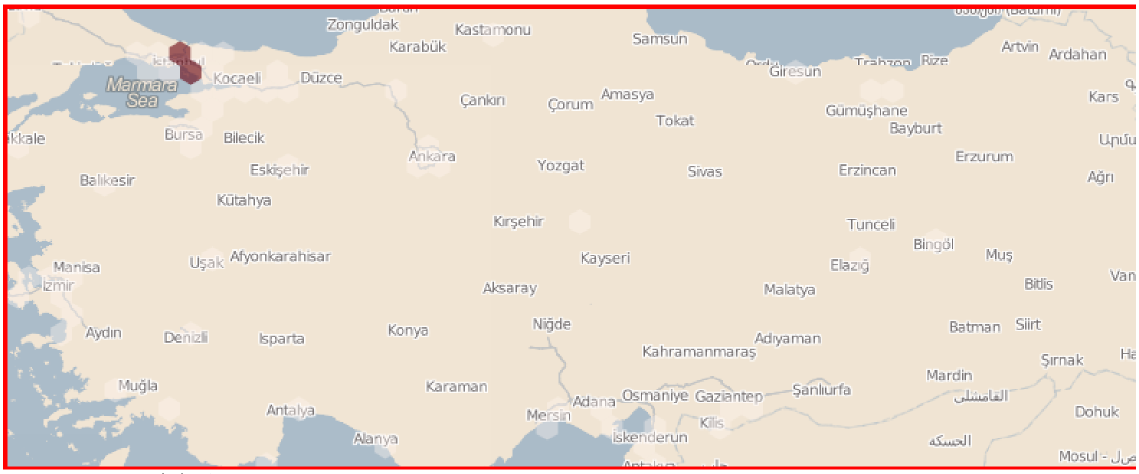
6.2 Trend Analysis Scenerio

In this scenerio, we will test the performance of our tool when analyzing the behavior of users after an important event. More specifically for our input data, we want to our tool aid the analysis of customer behavior after the system makes some advertisements on certain days. Througout this scenerio, we will be looking for answers of such questions:

- When does the advertisement starts to effect the usage of system?
- Which areas have shown increase in usage after advertisement?



(a) Data density of input data for the selected time periods(13:00-15:00) with possible anomalies



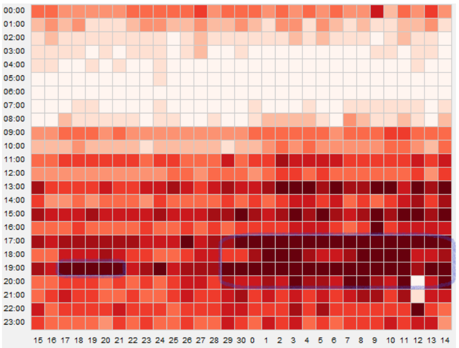
(b) Average view implies that unusual activity is arised around Istanbul

Figure 6.2: Spatial anomaly detection

- Which areas are the most effected areas after advertisement and why?

To see if our tool helps analyst to spot a pattern or a trend in data after these advertisements, we will follow the trend analysis routine that we described in the previous chapter.

First thing for analyst to do is again focusing a time period with the help of calendar heat map visualization. Since we know the advertisement dates, we know the starting date for analysis at this scenerio. For example, to see the effect of advertisement on 25.08.2011, we will start from that date and look for increasing amount of usage(See 6.3(a)). As we spot some increasing trends couple of days after the advertisement, this answers our first analysis question. In order to find the areas that gets more activity after the advertisement, analyst needs to examine the average view of geographic component. This brings out some interesting areas around Samsun,anakkale etc.(See 6.3(b)). However, this 2D view alone does not tell us the complete story about how user behavior changed after the advertisement. For that, we need to examine the 3D visualization component that will aid us in the spatio-temporal data analysis.



(a) Possible increase of usage in the dates between 29.08 - 14.09 after service advertisement.

(b) For 29.08 - 14.09, areas with unusual high activities can be spotted in the average view of our geographic visualization.

Figure 6.3: Trend Analysis

7 CONCLUSION AND FUTURE WORK

7.1 Conclusion

Here in this project, we successfully developed a geovisualization framework to represent the density of our geographic data. We believe that this framework considers all of the design aspects to correctly represent the geographic data using heat maps. Combination of our proposed visualization techniques of change with our heat map framework could be useful for those who are interested in detecting anomalies and trends in geographic time-series data. With the help of our framework, users can animate their data, compare it with the average or overlay different snapshots to have an initial understanding of the change in data. Using the tool that we designed, they can solve several exploratory analysis tasks by interacting with different visualization components.

7.2 Future work

For our future work, we will be more focused on the 3D representation of our spatio-temporal data. We want to several interactions in our HeatCube system to support data analysis. These interactions can be filtering data to reduce the complexity or cutting slices of it to look from a different perspective. Currently, we are working on isosurfaces to represent our spatio-temporal data. Preliminary results of our future work can be seen in 7.1. We believe that with appropriate improvements in the future, our methods can be very useful in the analysis of spatio-temporal data.

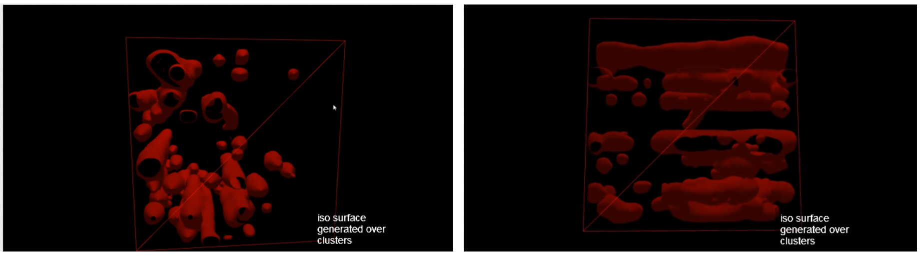


Figure 7.1: Isosurface experiments with data

Bibliography

[1] T. Lammarsch, W. Aigner, A. Bertone, J. Gartner, E. Mayr, S. Miksch, and M. Smuc, “Hierarchical temporal patterns and interactive aggregated views for pixel-based visualizations,” in *Information Visualisation, 2009 13th International Conference*, july 2009, pp. 44 –50.

[2] D. Fisher, “Hotmap: Looking at geographic attention.” *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1184–1191, 2007.

[3] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne, “Flowstrates: An approach for visual exploration of temporal origin-destination data,” *Eurographics/IEEE-VGTC Symposium on Visualization 2011, Computer Graphics Forum, International Journal of the Eurographics Association*, vol. 30, no. 3, p. 971–980, 2011. [Online]. Available: http://diuf.unifr.ch/people/boyandii/papers/flowstrates-eurovis11.pdf

[4] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk, “Interactive visualization of multivariate trajectory data with density maps,” Mar. 2011, pp. 147–154.

[5] D. Keim, G. Andrienko, J.-D. Fekete, C. Grg, J. Kohlhammer, and G. Melanon, “Visual analytics: Definition, process, and challenges,” in *Information Visualization*, ser. Lecture Notes in Computer Science, A. Kerren, J. Stasko, J.-D. Fekete, and C. North, Eds. Springer Berlin / Heidelberg, 2008, vol. 4950, pp. 154–175.

[6] J. W. Tukey, *Exploratory Data Analysis.* Addison-Wesley, 1977.

[7] N. Andrienko and G. Andrienko, *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*, 1st ed., Springer-Verlag, Ed. Springer, 2005. [Online]. Available: http://www.amazon.com/Exploratory-Analysis-Spatial-Temporal-Data/dp/3540259945

[8] G. L. Andrienko, N. V. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M.-J. Kraak, H. Schumann, and C. Tominski, "Space, time and visual analytics." *International Journal of Geographical Information Science*, vol. 24, no. 10, pp. 1577–1600, 2010.

[9] E. R. Tufte, *Envisioning Information*, 4th ed. Graphics Pr, May 1990.

[10] ——, *The Visual Display of Quantitative Information*, 2nd ed. Graphics Pr, May 2001. [Online]. Available: http://www.worldcat.org/isbn/0961392142

[11] R. Spence, *Information visualization*, ser. ACM Press books. Harlow, England: Addison-Wesley, 2001.

[12] A. Unwin, M. Theus, and H. Hofmann, *Graphics of Large Datasets: Visualizing a Million (Statistics and Computing)*, 1st ed. Springer, Jul. 2006. [Online]. Available: http://www.worldcat.org/isbn/0387329064

[13] A. Inselberg, "Visualizing high dimensional datasets and multivariate relations (tutorial am-2)," in *Tutorial notes of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '00. New York, NY, USA: ACM, 2000, pp. 33–94. [Online]. Available: http://doi.acm.org/10.1145/349093.349102

[14] C. Forsell, S. Seipel, and M. Lind, "Simple 3d glyphs for spatial multivariate data," *Information Visualization, IEEE Symposium on*, vol. 0, p. 16, 2005.

[15] D. Keim, H.-P. Kriegel, and M. Ankerst, "Recursive pattern: a technique for visualizing very large amounts of data," in *Visualization, 1995. Visualization '95. Proceedings., IEEE Conference on*, oct-3 nov 1995, pp. 279 –286, 463.

[16] B. Shneiderman, "Tree visualization with tree-maps: A 2-d space-filling approach," *ACM Transactions on Graphics*, vol. 11, pp. 92–99, 1991.

[17] J. J. V. Wijk, "The value of visualization," 2005, pp. 79–86.

[18] C. Johnson and C. Hansen, *Visualization Handbook*. Orlando, FL, USA: Academic Press, Inc., 2004.

[19] R. Fuchs and H. Hauser, "Visualization of multi-variate scientific data," *Computer Graphics Forum*, vol. 28, no. 6, pp. 1670–1690, 2009. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2009.01429.x

[20] K. Engel, M. Hadwiger, J. M. Kniss, and C. Rezk-Salama, "Real-Time Volume Graphics ," pp. 595–748. [Online]. Available: http://www.eg.org/EG/DL/Conf/EG2006/tutorials/0595-0748.pdf

[21] R. Laramee, H. Hauser, L. Zhao, and F. Post, "Topology-based flow visualization, the state of the art," in *Topology-based Methods in Visualization*, ser. Mathematics and Visualization, H. Hauser, H. Hagen, H. Theisel, G. Farin, H.-C. Hege, D. Hoffman, C. R. Johnson, K. Polthier, and M. Rumpf, Eds. Springer Berlin Heidelberg, 2007, pp. 1–19.

[22] J. Kruger, J. Schneider, and R. Westermann, "Clearview: An interactive context preserving hotspot visualization technique," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, pp. 941–948, 2006.

[23] W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski, "Visual methods for analyzing time-oriented data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 1, pp. 47 –60, jan.-feb. 2008.

[24] T. Mintz D., Fitz-Simons and M. Wayland, "Tracking air quality trends with sas/graph," 1997, pp. 807–812.

[25] M. H. Shimabukuro, E. F. Flores, M. C. F. de Oliveira, and H. Levkowitz, "Coordinated views to assist exploration of spatio-temporal data: A case study," in *Proceedings of the Second International Conference on Coordinated & Multiple Views in Exploratory Visualization*, ser. CMV '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 107–117. [Online]. Available: http://dx.doi.org/10.1109/CMV.2004.6

[26] A. M. MacEachren and M. J. Kraak, "Research challenges in geovisualization," *Cartography and Geographic Information Science (CaGIS)*, vol. 28, pp. 3–12, 2001.

[27] A. MacEachren, *How Maps Work: Representation, Visualization, and Design*, 1st ed. The Guilford Press, Jun. 2004.

[28] T. A. Slocum, R. B. Mcmaster, F. C. Kessler, and H. H. Howard, *Thematic Cartography and Geovisualization (3rd Edition) (Prentice Hall Series in Geographic Information Science)*, 3rd ed. Prentice Hall, Apr. 2008.

[29] L. Wilkinson and M. Friendly, "The History of the Cluster Heat Map," *The American Statistician*, vol. 63, no. 2, pp. 179–184, May 2009.

[30] A. Mehler, Y. Bao, X. Li, Y. Wang, and S. Skiena, "Spatial analysis of news sources." *IEEE transactions on visualization and computer graphics*, vol. 12, no. 5, pp. 765–71, 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17080798

[31] N. Willems, H. van de Wetering, and J. J. van Wijk, "Visualization of vessel movements," *Computer Graphics Forum*, vol. 28, no. 3, pp. 959–966, 2009.

[32] "Temporal GIS and Spatio-Temporal Modeling."

[33] J. Roddick, K. Hornsby, and M. Spiliopoulou, "An Updated Bibliography of Temporal, Spatial, and Spatio-temporal Data Mining Research," 2001, pp. 147–163.

[34] N. Andrienko, "Exploratory spatio-temporal visualization: an analytical review," *Journal of Visual Languages & Computing*, vol. 14, no. 6, pp. 503–541, Dec. 2003. [Online]. Available: http://dx.doi.org/10.1016/S1045-926X(03)00046-6

[35] T. Abraham and J. F. Roddick, "Survey of spatio-temporal databases," *Geoinformatica*, vol. 3, no. 1, pp. 61–99, Mar. 1999. [Online]. Available: http://dx.doi.org/10.1023/A:1009800916313

[36] T.-M. Rhyne, A. M. MacEachren, and J. Dykes, "Guest editors' introduction: Exploring geovisualization." *IEEE Computer Graphics and Applications*, vol. 26, no. 4, pp. 20–21, 2006.

[37] P. Shanbhag, P. Rheingans, and M. desJardins, "Temporal visualization of planning polygons for efficient partitioning of geo-spatial data." in *INFOVIS*, J. T. Stasko and M. O. Ward, Eds. IEEE Computer Society, 2005, p. 28.

[38] M. J. Kraak, "The space-time cube revisited from a geo-visualization perspective," *Proceedings of the 21st International Cartographic Conference*, vol. 1995, 1988. [Online]. Available: http://www.itc.nl/library/Papers_2003/art_proc/kraak.pdf

[39] P. Fisher, M.-J. Kraak, and A. Koussoulakou, "A Visualization Environment for the Space-Time-Cube," in *Developments in Spatial Data Handling*. Berlin/Heidelberg: Springer Berlin Heidelberg, 2005, ch. 15, pp. 189–200. [Online]. Available: http://dx.doi.org/10.1007/3-540-26772-7_15

[40] U. D. Turdukulov, M.-J. Kraak, and C. A. Blok, "Designing a visual environment for exploration of time series of remote sensing data: In search for convective clouds," *Computers & Graphics*, vol. 31, no. 3, pp. 370–379, Jun. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.cag.2007.01.028

[41] P. Gatalsky, N. V. Andrienko, and G. L. Andrienko, "Interactive analysis of event data using space-time cube." in *IV*. IEEE Computer Society, 2004, pp. 145–152.

[42] S. Hadlak, C. Tominski, H.-J. Schulz, and H. Schumann, "Visualization of attributed hierarchical structures in a spatio-temporal context," *International*

Journal of Geographical Information Science, vol. 24, no. 10, pp. 1497–1513, 2010.

[43] U. Demsar and K. Virrantaus, "Space-time density of trajectories: exploring spatio-temporal patterns in movement data," *Int. J. Geogr. Inf. Sci.*, vol. 24, no. 10, pp. 1527–1542, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1080/13658816.2010.511223

[44] R. Scheepens, N. Willems, H. van de Wetering, G. Andrienko, N. Andrienko, and J. J. van Wijk, "Composite density maps for multivariate trajectories." *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2518–27, Dec. 2011.

[45] G. Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. Von Landesberger, P. Bak, and D. Keim, "Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns," *Computer Graphics Forum*, vol. 29, no. 3, pp. 913–922, 2010. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2009.01664.x

[46] P. Agarwal and A. Skupin, Eds., *Self-Organising Maps: Applications in Geographic Information Science*, 1st ed. Wiley, Sep. 2008. [Online]. Available: http://www.worldcat.org/isbn/0470021675

[47] J. Wood and J. Dykes, "Spatially ordered treemaps," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, pp. 1348–1355, 2008.

[48] N. Andrienko, G. Andrienko, H. Stange, T. Liebig, and D. Hecker, "Visual Analytics for Understanding Spatial Situations from Episodic Movement Data," *KI - Künstliche Intelligenz*, pp. 1–19, Mar. 2012. [Online]. Available: http://www.springerlink.com/index/10.1007/s13218-012-0177-4

[49] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M.-J. Kraak, H. Schumann, and C. Tominski, "Space,

time and visual analytics," *International Journal of Geographical Information Science*, vol. 24, no. 10, pp. 1577–1600, Oct. 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/13658816.2010.508043

[50] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M. J. Kraak, A. MacEachren, and S. Wrobel, "Geovisual analytics for spatial decision support: Setting the research agenda," *Int. J. Geogr. Inf. Sci.*, vol. 21, no. 8, pp. 839–857, Jan. 2007. [Online]. Available: http://dx.doi.org/10.1080/13658810701349011

[51] P. Gatalsky, N. Andrienko, and G. Andrienko, "Interactive analysis of event data using space-time cube," *Information Visualisation, International Conference on*, vol. 0, pp. 145–152, 2004.

[52] T. Kapler and W. Wright, "Geo time information visualization," *Information Visualization*, vol. 4, no. 2, pp. 136–146, Jul. 2005. [Online]. Available: http://dx.doi.org/10.1145/1103520.1103526

[53] C. Forlines and K. Wittenburg, "Wakame: sense making of multi-dimensional spatial-temporal data," in *Proceedings of the International Conference on Advanced Visual Interfaces*, ser. AVI '10. New York, NY, USA: ACM, 2010, pp. 33–40. [Online]. Available: http://doi.acm.org/10.1145/1842993.1843000

[54] G. Andrienko, N. Andrienko, and S. Wrobel, "Visual analytics tools for analysis of movement data," *SIGKDD Explor. Newsl.*, vol. 9, no. 2, pp. 38–46, Dec. 2007. [Online]. Available: http://doi.acm.org/10.1145/1345448.1345455

[55] D. Guo, J. Chen, A. M. MacEachren, and K. Liao, "A visualization system for space-time and multivariate patterns (vis-stamp)." *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 6, pp. 1461–1474, 2006.

[56] M. Jern, "Collaborative web-enabled geoanalytics applied to oecd regional data," in *Proceedings of the 6th international conference on Cooperative design, visualization, and engineering*, ser. CDVE'09. Berlin,

Heidelberg: Springer-Verlag, 2009, pp. 32–43. [Online]. Available: http://dl.acm.org/citation.cfm?id=1812983.1812988

[57] N. Pelekis, B. Theodoulidis, I. Kopanakis, and Y. Theodoridis, "Literature review of spatio-temporal database models," *Knowl. Eng. Rev.*, vol. 19, pp. 235–274, Sep. 2004. [Online]. Available: http://portal.acm.org/citation.cfm?id=1068516

[58] M. M. Haklay and P. Weber, "OpenStreetMap: User-Generated Street Maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

[59] M. Purvis, J. Sambells, and C. Turner, *Google Maps Application Development.* Apress, Aug. 2006. [Online]. Available: http://www.worldcat.org/isbn/1590597079

[60] G. Kimchi, "History and lessons from Microsoft Virtual Earth," in *GIS '09: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.* New York, NY, USA: ACM, 2009, p. 2. [Online]. Available: http://dx.doi.org/10.1145/1653771.1653774

[61] J. Snyder, *Map Projections-A Working Manual.* United States Government Printing, Feb. 1983. [Online]. Available: http://onlinepubs.er.usgs.gov/djvu/PP/PP_1395.pdf

[62] G. S. Michaels, D. B. Carr, M. Askenazi, S. Fuhrman, X. Wen, and R. Somogyi, "Cluster analysis and data visualization of large-scale gene expression data." *Pac Symp Biocomput*, pp. 42–53, 1998. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/9697170

[63] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[64] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979. [Online]. Available: http://dx.doi.org/10.2307/2346830

[65] W. S. Cleveland and R. McGill, "The many faces of a scatterplot," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. pp. 807–822, 1984.

[66] D. B. Carr, R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield, "Scatterplot Matrix Techniques for Large N," *Journal of the American Statistical Association*, vol. 82, no. 398, 1987. [Online]. Available: http://dx.doi.org/10.2307/2289444

[67] J. Adler, *Baseball Hacks*. O'Reilly Media, Inc., 2006.

[68] M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, Sep. 1956. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177728190

[69] M. Bostock, V. Ogievetsky, and J. Heer, "Data-driven documents," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2301 –2309, dec. 2011.

[70] M. Harrower and C. Brewer, "Colorbrewer.org: An online tool for selecting colour schemes for maps," *Cartographic Journal, The*, vol. 40, no. 1, pp. 27–37, 2003-06-01T00:00:00. [Online]. Available: http://www.ingentaconnect.com/content/maney/caj/2003/00000040/00000001/art00004

[71] C. R. Ehlschlaeger, A. M. Shortridge, and M. F. Goodchild, "Visualizing spatial data uncertainty using animation," *Computers and Geosciences*, vol. 23, no. 4, pp. 387 – 395, 1997, ¡ce:title¿Exploratory Cartograpic Visualisation¡/ce:title¿. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098300497000058

[72] G. Wolberg, "Image morphing: a survey," *The Visual Computer*, vol. 14, pp. 360–372, 1998, 10.1007/s003710050148. [Online]. Available: http://dx.doi.org/10.1007/s003710050148

[73] W. Wang, J. Yang, and R. R. Muntz, "Sting: A statistical information grid approach to spatial data mining," in *Twenty-Third International Conference on Very Large Data Bases*, M. Jarke, M. J. Carey, K. R. Dittrich, F. H. Lochovsky, P. Loucopoulos, and M. A. Jeusfeld, Eds. Athens, Greece: Morgan Kaufmann, 1997, pp. 186–195. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.5117

[74] N. Max and B. Becker, "Bump shading for volume textures," *Computer Graphics and Applications, IEEE*, vol. 14, no. 4, pp. 18 –20, july 1994.

[75] G. F. Panza, S. Mueller, and G. Calcagnile, "The gross features of the lithosphere-asthenosphere system in europe from seismic surface waves and body waves," *Pure and Applied Geophysics*, vol. 118, pp. 1209–1213, 1980, 10.1007/BF01593061. [Online]. Available: http://dx.doi.org/10.1007/BF01593061

[76] M. Van Kreveld, "Efficient methods for isoline extraction from a tin," *International journal of geographical information systems*, vol. 10, no. 5, pp. 523–540, 1996. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/02693799608902095

[77] P. Vankeerberghen, J. Smeyers-Verbeke, R. Leardi, C. Karr, and D. Massart, "Robust regression and outlier detection for non-linear models using genetic algorithms," *Chemometrics and Intelligent Laboratory Systems*, vol. 28, no. 1, pp. 73 – 87, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0169743995800417

[78] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, pp. 85–

126, 2004, 10.1023/B:AIRE.0000045502.10941.a9. [Online]. Available: http://dx.doi.org/10.1023/B:AIRE.0000045502.10941.a9