

INTRODUCING LEVEL OF DETAIL TO 3D THEMATIC MAPS

By

HATİCE BİLLUR ENGİN

Submitted to the Graduate School of Engineering and Natural Sciences

in partial fulfillment of the requirements for the degree of

Master of Science

SABANCI UNIVERSITY

Spring 2008

INTRODUCING LEVEL OF DETAIL TO 3D THEMATIC MAPS

**APPROVED BY:**

Assist. Prof Selim Balcısoy  
(Dissertation Adviser)

\_\_\_\_\_

Assist. Prof Burçin Bozkaya

\_\_\_\_\_

Assoc. Prof Erkay Savaş

\_\_\_\_\_

Assist. Prof Gürdal Ertek

\_\_\_\_\_

Assist. Prof Yücel Saygın

\_\_\_\_\_

**DATE OF APPROVAL:**

30.07.2006

© Hatice Billur Engin 2008

ALL RIGHTS RESERVED

# INTRODUCING LEVEL OF DETAIL TO 3D THEMATIC MAPS

Hatice Billur Engin

EECS, M.Sc. Thesis, 2008

Thesis Supervisor : Asist. Prof. Selim Balcısoy

Keywords: Geographical Information Visualization, Thematic Maps, Level of Detail, Cartography

## **ABSTRACT**

This thesis investigates the three dimensional visualization of geography related statistical data, organized in different abstraction levels considering their distance to camera. If the aim is to tell a story about place (linked with geography), thematic maps are said to be one of the most generic methods. With the help of texturing technology, two dimensional thematic maps are generated in real time and projected on a predefined terrain. Introducing level of detail for data abstraction with respect to camera movements advanced the system into a multiscale visualization.

# ÜÇ BOYUTLU TEMATİK HARİTALARDA FARKLI DETAY SEVİYELERİNİN UYGULANMASI

Hatice Billur Engin

EECS, Yüksek Lisans Tezi, 2008

Tez Danışmanı: Yar. Doç. Selim Balcısoy

Anahtar Kelimeler: Coğrafi Bilgi Görselleştirme, Tematik Haritalar, Detay Seviyeleri, Kartografya.

## ÖZET

Bu tezin amacı coğrafyayla ilintili istatistiki verilerin, kameraya olan uzaklıkları dikkat alınarak, farklı detay seviyelerinde üç boyutlu olarak görselleştirilmesidir. Yer kavramıyla ilgili bilgi görselleştirmek konusunda en sık kullanılan metodlardan biri tematik haritalardır. Kaplama teknolojisinden yararlanılarak gerçek zamanda yaratılan iki boyutlu tematik haritalar, önceden tanımlanmış bir yeryüzü parçasının üzerine projekte edilir. Kamera hareketlerine göre değişkenlik gösteren detay seviyeleri tekniğinin, verilere uygulanmasıyla birlikte birden fazla ölçekli bir görselleştirme elde edilmiştir .

**To my parents...**

## ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my supervisor Selim Balcısoy for his valuable advice and guidance of this work. I am grateful to him not only for the completion of this thesis, but also for his unconditional support from the beginning. I feel myself privileged as his student.

I would like to thank all my friends in Computer Graphics Lab, particularly to Ceren Kayalar and Selçuk Sümengen for their friendship and assistance through out the last two years.

I am grateful to my thesis committee members Burçin BOZKAYA, Erkay SAVAŞ, Gürdal ERTEK, Yücel SAYGIN for their valuable review and comments on the dissertation.

I would like to thank to my precious family for their unconditional love and belief in me that made everything real.

My sincere thanks to İrfan and Semra Damgacı for encouraging me at all my times of hesitation.

Last but not least, special thanks to my little nephew Pelin Damgacı for all the joy she brought to my life with her birth.

This research is supported by TÜBİTAK (Research Grant 105K165).

## Table of Contents

1. INTRODUCTION .....	1
1.1. Visualization: An Overview.....	1
1.2. Problem Definition.....	2
1.3. Summary of Contributions.....	3
1.4. Thesis Outline .....	4
2. MOTIVATION AND RELATED WORK .....	6
2.1. Cartography and a Short Overview of the Historical Development .....	6
2.2. Overview on Visualization.....	9
2.2.1. Scientific visualization .....	9
2.2.2. Information visualization .....	9
2.2.3. Maps and Information Visualization.....	10
Thematic Maps.....	12
2.2.4. Multiscale visualization .....	15
2.2.5. Geographic visualization.....	17
3. THE SYSTEM.....	22
3.1. Overview .....	22
3.2. Input Data.....	23
3.3. Statistical Foundations .....	24
3.3.1. Exploratory data analysis .....	24
3.3.2. Equal intervals.....	25
3.3.3. Box-plot .....	26
3.3.4. Normalization.....	27
3.4. Visualization .....	27
3.4.1. Subdivisions .....	28
3.4.2. Colorization.....	31
3.3.4 Hatching .....	34
3.3. Terrain Visualizer.....	35
3.4. User Interface.....	37
3.4.1. Navigation.....	37
3.4.2. Legend.....	37
4. CASE STUDY .....	38



4.1. Population Data of San Francisco .....	38
4.1.1. Input Data.....	38
4.1.2. Exploratory Data Analysis .....	39
4.1.3. Colorization.....	42
4.1.4. Road Network .....	44
4.2. Multiple Data Visualization .....	44
4.2.1. Symbology .....	44
4.2.3. Road Network .....	45
4.3. Comparison with ArcScene.....	46
5. RESULTS, DISCUSSIONS and FURTHER STUDY .....	48
5.1. Results.....	48
5.2. Discussions and Further Study.....	49
REFERENCES.....	51

## LIST OF FIGURES

Figure 1.1 Proposed geographic visualization system’s snapshots in two different levels of detail with different unit subdivision areas.....	3
Figure 2.1 City plan of Çatal Höyük. Image courtesy of Ali Turan in “Turkey in maps”.....	6
Figure 2.2 City plan of Çatal Höyük. Recreation of the original plan. In cartographical study, this wall painting is considered as the first illustrations of terrain.....	6
Figure 2.3 Map of Tuscany and the Chiana Valley (1502).....	7
Figure 2.4. Heidelberg Castle and gardens (1650).....	7
Figure 2.5. An example of elevation contour map.....	8
Figure 2.6. London underground railway map, a famous example of information visualization.....	10
Figure 2.7. A portion of Snow’s map.....	12
Figure2.8. Choropleth map of Africa’s population distribution in 1990[26].....	13
Figure2.9. Isarithmic map of high temperatures for America [25].....	13
Figure2.10. Dot mapping of America’s population distribution data.[26].....	14
Figure 2.11. Choropleth mapping of Hispanic distribution over states[26].....	15
Figure 2.12. 3D model of Legible Cities (a) viewing clustered neighborhood, (b) individual building viewing.....	16
Figure 2.13. Matrix view of Legible Cities.....	17
Figure 2.14. (a) Traditional map (b) New York Times Cartogram of Electoral College Votes.....	18
Figure 2.15.Adoption of non-photorealistic techniques of computer graphics to geovisualization [36].....	19

Figure 2.16. Turkish population distribution of London [39].....	20
Figure 2.17. LOD design rules. The exact boundaries of forest areas are visualized if the camera comes close to the terrain. The textures are computed on the fly [41].....	21
Figure 3.1. Program Flow Diagram.....	23
Figure 3.2. Unit subdivision area for highest resolution.....	28
Figure 3.3. Unit subdivision area for medium resolution, consisting of 4 grid points.....	29
Figure 3.4. Unit subdivision areas for (b) highest and (a) medium resolution.....	29
Figure 3.5. Unit subdivision area for lowest resolution consisting of 16 grid points, where $P_{ij}$ is a grid point and $a$ is $1/8$ of unit square's edge.....	30
Figure 3.6. Shading of nonuniform subdivisions.....	30
Figure 3.7. Illustration of how to determine whether a point is inside a polygon or not [44].....	31
Figure 3.8. Legend for visualization of unclassified data.....	32
Figure 3.9. Scale generated for the visualization of data classified with equal intervals method.....	32
Figure 3.10 scale generated for the visualization of data classified with box-plot method.....	34
Figure 3.11. Hatching method for different level of details.....	34
Figure 3.12. wire model of the terrain.....	36
Figure 3.13. Newell Method for calculating surface normals.....	36
Figure 3.14. The legend.....	37
Figure 4.1. Pie Chart of San Francisco population data.....	40

Figure 4.2. Box plot of San Francisco population data.....41

Figure 4.3.(a) High-rise buildings of San Francisco, located on Google Earth,  
 (b) Population density pattern of this visualization system.....42

Figure 4.4. legend for box-plot method.....42

Figure 4.5. legend for equal intervals method.....43

Figure 4.6. Hatching styles for different level of details.....43

Figure 4.7. Road network of San Francisco reinforces the population density transitions.....44

Figure 4.8. Visualization of multiple data layers.....45

Figure 4.9. Visualization of two layers of statistical data and road network together.....45

Figure 4.10. Via using level of detail information density of the depiction is kept constant, so observing the data from a distant point is possible in the proposed system.....46

Figure 4.11. Due to unvarying size of squares, information density increases as camera gets away from the representation region.....46

Figure 4.12. Clustering effect of borders with unvarying sizes(a) and comparison of landscape illustrations from a distant point (a) illustration made via ArcScene (b) illustration made via proposed system.....47

Figure 4.13. Visualization of geographical data with ArcScene(a) and with the proposed system.....47

Figure 5.1. The user has the freedom to visualize her dataset in three different distribution models.....49

## TABLE OF ABBREVIATIONS

KML	Keyhole Markup Language
XML	eXtensible Mark-up Language
GIS	Geographical Information System

# 1. INTRODUCTION

## 1.1. Visualization: An Overview

*Despite what color your eyes or your eyebrows are,  
your color is what is seen by the one facing you...<sup>1</sup>*

Can Yücel [1]

The smartest person ever lived wouldn't be known to be smart unless she was capable of communicating with others in a proper way. No matter how great an idea is, without being submitted to the benefit of other people, it wouldn't lead to a new course. Donald Norman [2] states the importance of external aids for enhancing cognitive abilities as he says "*the power of unaided mind is highly overrated. Without external aids, memory, thought, and reasoning are all constrained.*"

When the focus is on communication, graphical inventions have always been among strongest external aids. A graphical representation may help us to put an idea into words, as they say a picture is worth a thousand words, or it may facilitate to figure out the idea itself, as Bertin [3] describes it; "*using vision to think*".

The term visualization is originated from a special issue of Computer Graphics authored by Bruce McCormick and his colleagues [4] and the purpose of visualization is to help human mind understand and interact with large volumes of data. Showing the relationships in a meaningful way, may help people get a different perspective about the data and represented facts may be recognized in a shorter time.

---

<sup>1</sup> *Ne renk olursa olsun kaşın gözün,  
Karşındakinin gördüğüdür rengin..*

Claiming that visualization is one of the key concepts in multidisciplinary environments won't be overrating the phenomenon. Especially in the last several years, with the evolution of computer technology and the rise of World Wide Web, visualization has turned out to be a vital communication and analysis tool.

The proposed visualization system is a multi-scale geographical visualization, designed to display statistical data in 3D.

## **1.2. Problem Definition**

Visualizing statistical data in relationship with geography is a complex task. This complexity lies in the expanded information density due to the statistical data introduced.

Defining a visual language for depiction of multiple data layers in 3D, which will keep the complexity of created representation at optimum, is the starting point of the problem handled in this thesis. Keeping the complexity in optimum means balancing the information represented with the area shown in the scene at that particular moment. This is the point where level of detail gets into act.

Since a geographic visualization system tends to contain a variety of data layers, such as traffic load, demographics and geographic features, each layer's symbology must be simple and in contrast with each other in order to be recognizable. Thus for each additional layer a distinct symbology should be designed.

3D terrain visualization is a topic studied in depth, in the last decade [5, 6, 7, 8]. Now it is possible to visualize a large landscape and interact with it in real time, using an average notebook. Level of detail algorithms are produced in order gain speed, while maintaining the most detailed view for the places nearest to the camera, showing less features for regions away from camera and not drawing at all the regions placed outside the scene [9, 10].

In this study an algorithm is developed for the visualization of spatial statistical data in 3D. The approach is to construct a thematic map from input data, then to assign detail levels on it and to change those detail levels due to the position of camera, in order to preserve a constant information density. As the displayed region gets distant to camera, representation of data becomes increasingly simplified. Simplifying means enlarging the unit subdivision area of the region automatically, ensuring that the information is still readable. As the distance between a region and camera decreases, unit subdivision area gets smaller, leading to a more detailed depiction.

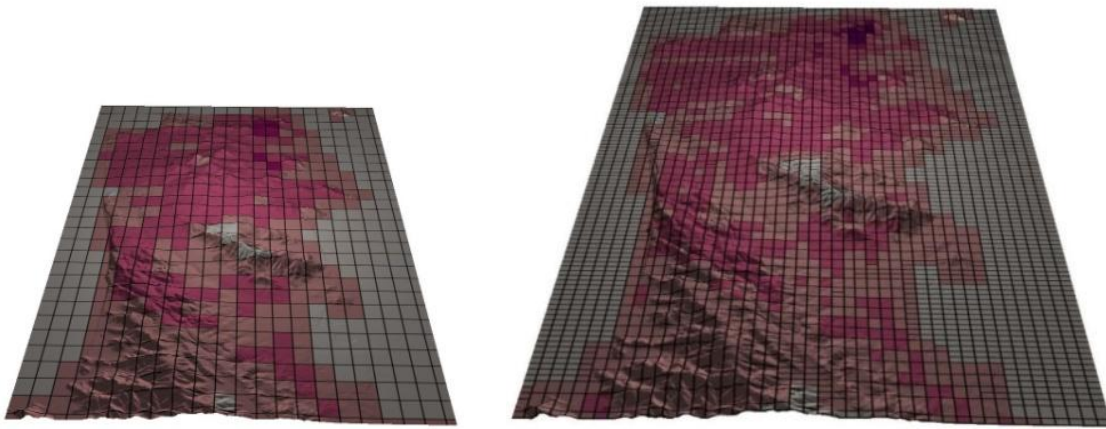


Figure 1.1 Proposed geographic visualization system's snapshots in two different levels of detail with different unit subdivision areas.

After the production process, 2D thematic map is wrapped on the 3D terrain as a texture. With the aid of advanced computer technology, namely using frame buffer object, doing flexible off-screen rendering is possible, during run time. By this system, the geography of the landscape is preserved, as the statistical data is visualized.

### 1.3. Summary of Contributions

In the age of Second Life [11], games with an enormous modeling effort, Google Earth [12], Google Maps [13], super computers with multiple cores and fast internet connections, the concept of visualization is strictly bounded with 3D models and high interactivity. Realizing the need for a 3D visualization system in which user can observe



statistical data as she navigates through the scene, a flexible 3D thematic map generator is proposed in this thesis. The main contributions of this research are:

- The possibility of observing data in a 3 dimensional environment. Since we live in a 3 dimensional world, our natural way of looking at things corresponds to 3D visualization. Visualizing spatial data in its original 3D geography will lead to a much faster understanding and avoiding confusions.
- User has the opportunity to choose between different statistical visualizations and decide which one of them best fits the distribution of input data.
- Displaying of the relationship between statistical data and geography in an intuitive way.
- Thanks to its texture based technical architecture, level of detail is introduced to thematic mapping. Since dynamical texturing is possible with the technological advances, data abstraction process is done in real time.
- Introducing automated details on demand to thematic maps, where details are automatically visualized when the viewpoint gets closer to the terrain.

#### **1.4. Thesis Outline**

This chapter briefly indicates why visualization is important and draws attention to increasing importance of 3D graphical representations in our everyday life. It also gives an abstract about the visualization system developed and the main aspects of this approach. Finally it gives a list of contributions. Subsequent chapters contribute to the thesis as the following:

*Motivation and Related Work:* The second chapter gives a brief historical background of cartography and provides definitions about geographical visualization field. Some examples of present geographical visualization systems are presented and their comparisons with the proposed system are made.

*System:* The third chapter describes details of the proposed system and explains its technical aspects.

*Case Study:* The fourth chapter demonstrates the visualization of demographic data of San Francisco city with the developed method and a multiscale visualization is performed.

*Results, Discussions and Further Study:* The fifth chapter denotes the accomplishments and limitation of the work. It also refers to further works and possible improvements for the method.

## 2. MOTIVATION AND RELATED WORK

### 2.1. Cartography and a Short Overview of the Historical Development

*Cartography is defined as the art and science of making maps to simplify and represent real world features (Monmonier 1996).*

The oldest city plan discovered to date is Çatal Höyük. In 1963 during excavations in Çatal Höyük, Konya, cartographical depictions of landscape were exposed. The wall painting is 3 meters long and dates aback to 6200 B.C. Aside from buildings, the twin peaks of the volcano Hasan Dağ are also illustrated. In cartographical study, these “molehills” are considered as the first illustrations of terrain [14].



Figure 2.1 City plan of Çatal Höyük. Image courtesy of Ali Turan in “Turkey in maps”.



Figure 2.2 City plan of Çatal Höyük. Recreation of the original plan. In cartographical study, this wall painting is considered as the first illustrations of terrain.

One of the precious earnings of Renaissance art is *perspective* (from Latin *perspicere*, meaning “see through”). Before perspective, the objects were drawn in their life-size or in a size according to their spiritual importance. With the use of perspective the sense of distance is taken into account. Including perspective in depiction of landscapes led to more realistic 3-dimensional representations of shapes. The first example of terrain representation from a bird’s eye perspective was Leonardo da Vinci’s maps of Tuscany.



Figure 2.3 Map of Tuscany and the Chiana Valley (1502).

It wasn't until 19<sup>th</sup> century, a totally top-view demonstration of maps devoid of any explanation sights took place. Since then, various techniques are improved to illustrate characteristics of topography, such as coloring of height layers, which aid the comprehension of terrain visualizations.

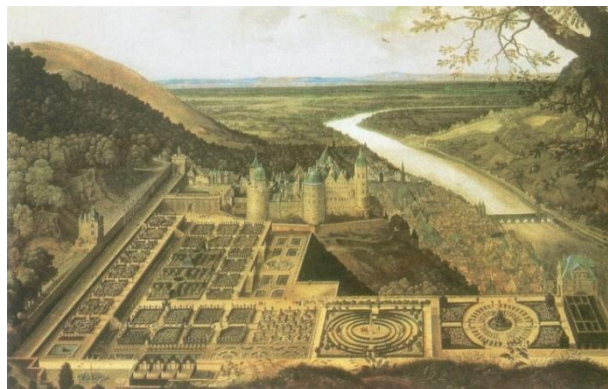


Figure 2.4. Heidelberg Castle and gardens (1650).

By the time Pieter Bruinss used contour lines for the first time in 1584, maps had reached a clear form of representation already. Contour lines acted as an opportunity of representing information about the altitudes.

Contour lines are considered as the cartographic depictions of terrain. They are lines formed from connected points of the same elevation.

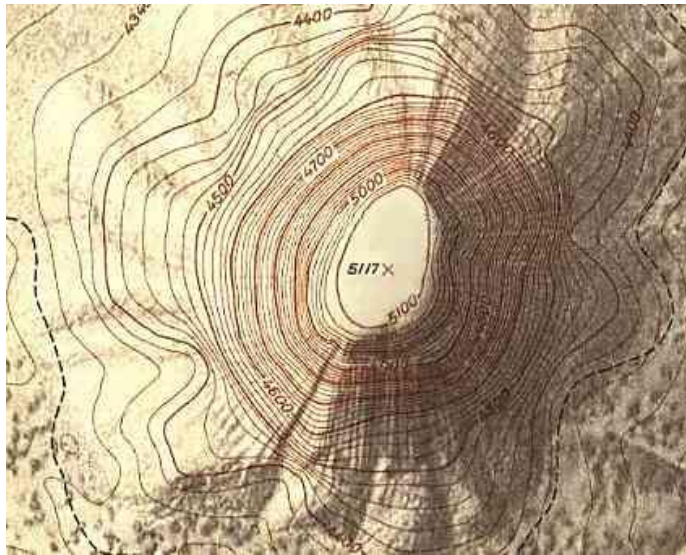


Figure 2.5. An example of elevation contour map.

In 1958 the first elevation models were created by Miller and Laflamme in the Photogrammetry Laboratory of the Civil Engineering Department of M.I.T. [15].

With the evolution of technology, in the last decades computers took their share in many different disciplines. This led scientists to improve new approaches and fresh research areas have emerged as a result. Likewise cartographers got a hold of new tools and methods. Cartography was a discipline based on pen and ink; it has now become dependent on computers. Computer scientists and cartographers paraphrased static maps into interactive visualizations. Combining static maps with multimedia and interactivity, they improved paper-based maps into easy to use graphical interfaces, available for making analysis on geospatial data. These technological changes in cartography had many consequences. First of all, more realistic demonstrations of earth are achieved. Besides, map making has become possible for anyone who owns a personal computer.

Consequently, as there was only one general map of a place, designed by professionals, now there are many different representations of a place, designed with diverse motivations.

## **2.2. Overview on Visualization**

### **2.2.1. Scientific visualization**

The objective of scientific visualization is described by McCormick as “*to leverage existing scientific methods by providing insight through visual methods*”. Scientific visualization deals with generating images from numerical or figurative data via computers. This field of science is based on methods originated from areas such as; computer vision, computer graphics, signals processing. Scientific visualization includes a variety of topics such as medical imaging and visualization of molecular structure and fluid flows.

### **2.2.2. Information visualization**

*Information visualization is the use of computer-supported, interactive, visual representations of abstract data to amplify cognition.* [16] The notion of information visualization differs from scientific visualization in the data they represent. Information visualization focuses on depiction of abstracted data to intensify cognition.

In many cases the data to be visualized are available in large amounts. In order to extent information from these data, putting some human insight into visualization process is vital. To exemplify this idea, London underground railway map will be helpful. This is an example of abstracted data visualization, which is about picking the information from raw data.

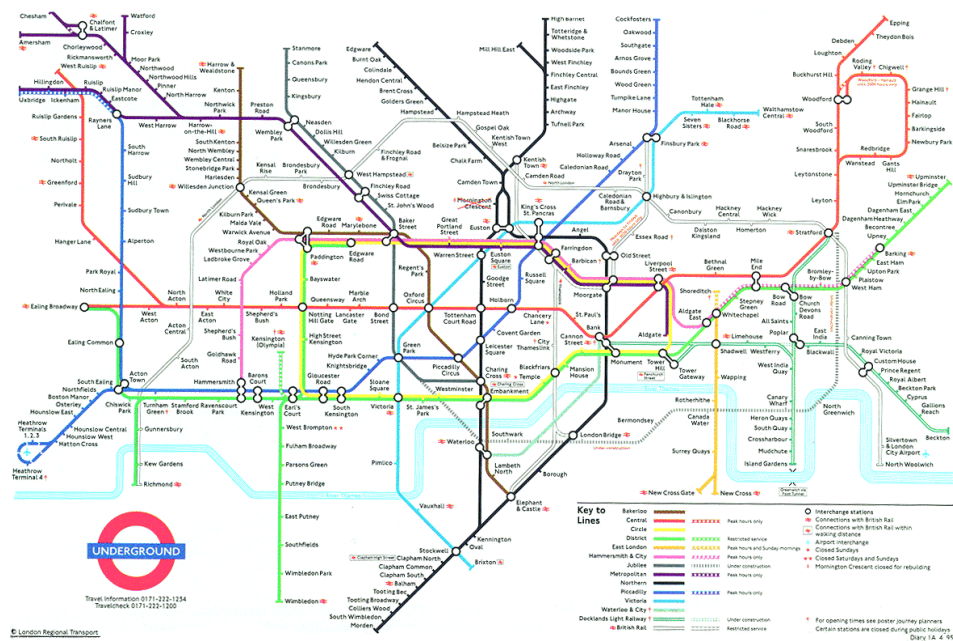


Figure 2.6. London underground railway map, a famous example of information visualization.

In the recent years World Wide Web has become an effective visualization platform. There are various examples of internet-based information visualization such as Musicoverly[17], Faces of the Dead[18], Tracking the Tread[19], etc. In addition to these, web sites more recently have begun serving to non-experts for producing visualizations. These web tools such as ManyEyes[20] - where users can upload data, construct visualizations and leave their comment on either data or visualization – ended the monopoly of expert generated depictions. ManyEyes project is not just constituting a research platform, it is an effort to “democratize” visualization technology .

### 2.2.3. Maps and Information Visualization

*A graphic statement that locates facts. (Krygiar & Wood)*

*A symbolized image of geographic reality, representing selected features or characteristics, resulting from the creative efforts of cartographers and designed for use when spatial relationships are of special relevance. (ICA, 1995)*

*An image of a place at a particular point in time, but that place has been intentionally reduced in size, and its contents have been selectively distilled to focus on one or two particular items. The results of this reduction and distillation are then encoded into a symbolic representation of the place. Finally, this encoded, symbolic image of a place has to be decoded and understood by a map reader who may live in a different time period and culture. [21]*

Some map definitions are given above. In most cases maps reveal facts; they are symbolic illustrations of attributes or relations of things and the elements of maps are usually smaller than their actual sizes. In addition to geography related subjects, a mapmaker can design maps of anything by using pictorial descriptions, such as anatomy, galaxies, social events, etc. Using properly verified data, she can highlight the subject with the help of simplification and generalization methods.

*Maps give their makers the power to define the territory in their terms and write a singular vision onto the landscape [22].* Mapmakers have the power to highlight any detail through a consistently designed model. Maps are created for many reasons; to demonstrate a cause for a public problem, to show the best possible location for a new store, to justify a reason, to show a route to home and etc. A map may personify ideas, may give information about a place we have never seen before, may contain patterns that lead to explanation of events and make the reader recognize truth that is there but doesn't meet the eye. Maps are one of the oldest techniques of information visualization, abstracting space for diverse forms of reasoning.

Maps branches into two types; *reference maps* which give information about the location and geographic features of a place and *thematic maps* that tell about the spatial pattern of a geographical distribution.

Since the focus of this study is representing the statistical pattern of a theme in 3D, thematic maps are essential for us.



## Thematic Maps

*A thematic map (or statistical map) is used to display the spatial pattern of a theme or attribute. They are used to emphasize the spatial pattern of one or more geographic attributes (or variables), such as population density, family income, and daily temperature maximums. [23]*

By means of displaying statistical data about a geographic area, a thematic map displays an enormous amount of data in a single image. This property of thematic maps fits the statement of Tufte et. al (1983) mentioning that “*Visualization can often represent a large amount of data in a small space*”. One of the earliest and the most famous examples of thematic maps is undoubtedly Snow’s cholera map in 1855, which is also emphasized by Tufte [24]. It is a good example for the usage of maps in analysis.

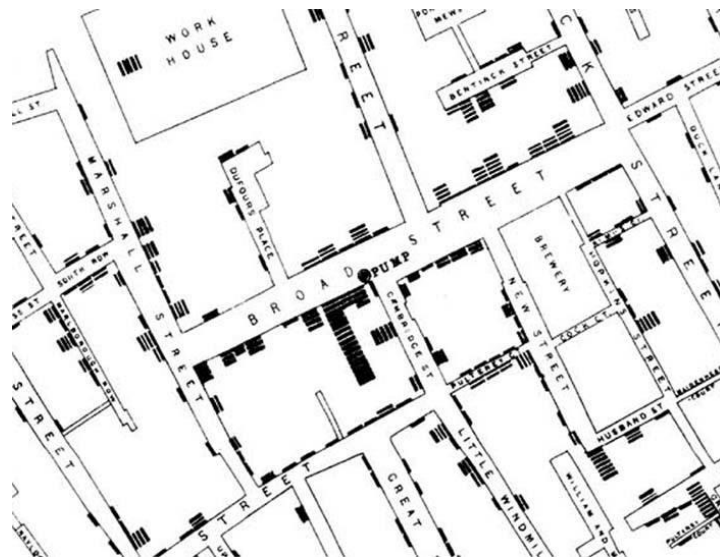


Figure 2.7. A portion of Snow’s map.

Snow’s map is a plan of London city streets and pump positions. It is a simple representation, in which cholera victims are shown with a scratch. The pattern of deaths mostly centered around one pump and with the removal of that pump (which was near a sewer a line), maybe by a coincidence, new cholera cases ceased almost at once.

## Types of thematic maps

- *Proportional Symbol Maps:* This type of thematic maps is produced with the principle of scaling symbols corresponding to magnitude of data. These maps are used to visualize data related with point locations, such as cities or counties.

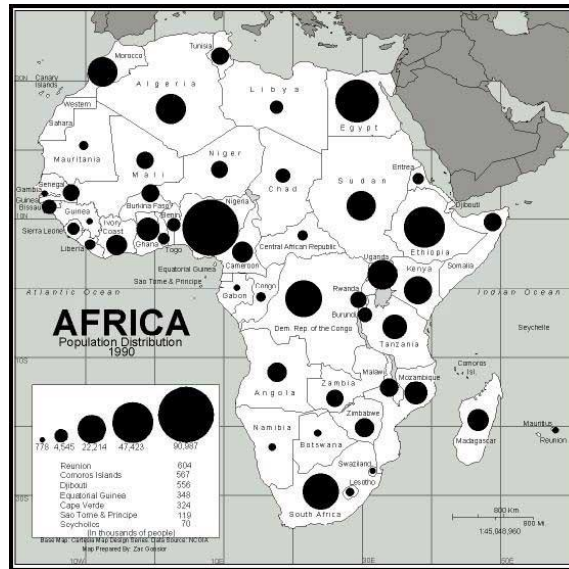


Figure 2.8. Choropleth map of Africa's population distribution in 1990 [26].

- *Isarithmic (Contour) Maps:* These maps are representation of contour lines produced by interpolating some known points. They are suitable for displaying smooth continuous data.

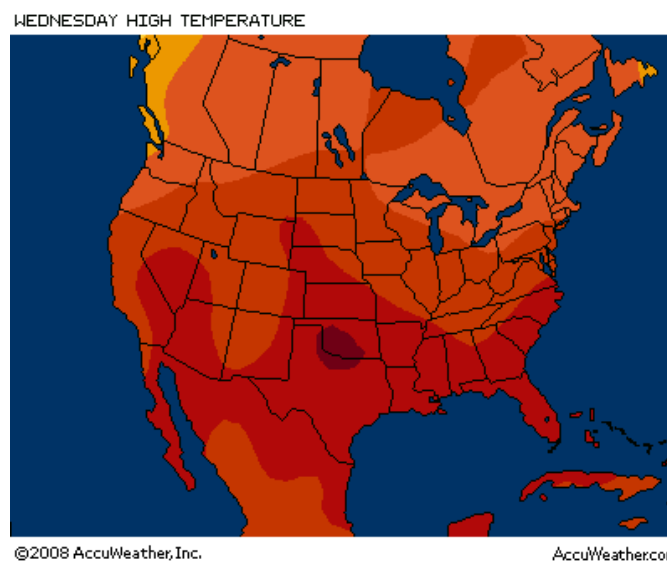


Figure 2.9. Isarithmic map of high temperatures for America [25]

- *Dot Mapping*: This is the kind of thematic map Dr. Snow used to represent cholera deaths. In these maps, one dot is set equal to a constant quantity of the depiction's subject matter, and those dots are located to places they occur.

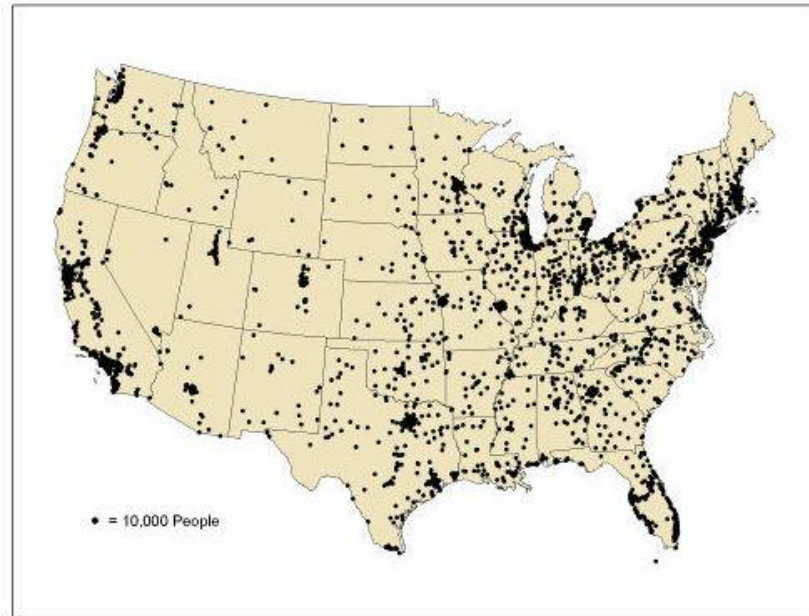


Figure 2.10. Dot mapping of America's population distribution data.[26]

- *Choropleth (Statistical) Maps*: Choropleth maps are commonly used to represent the spatial attributes of a statistical data. They divide the area into smaller enumeration units and shade them according to the measure of statistical variable. This is the type of thematic maps used in this study.

The first modern choropleth map was published by Charles Dupin in 1826. It was a choropleth map with shading from black to white. Subject of thematic map was the distribution of illiteracy in France.

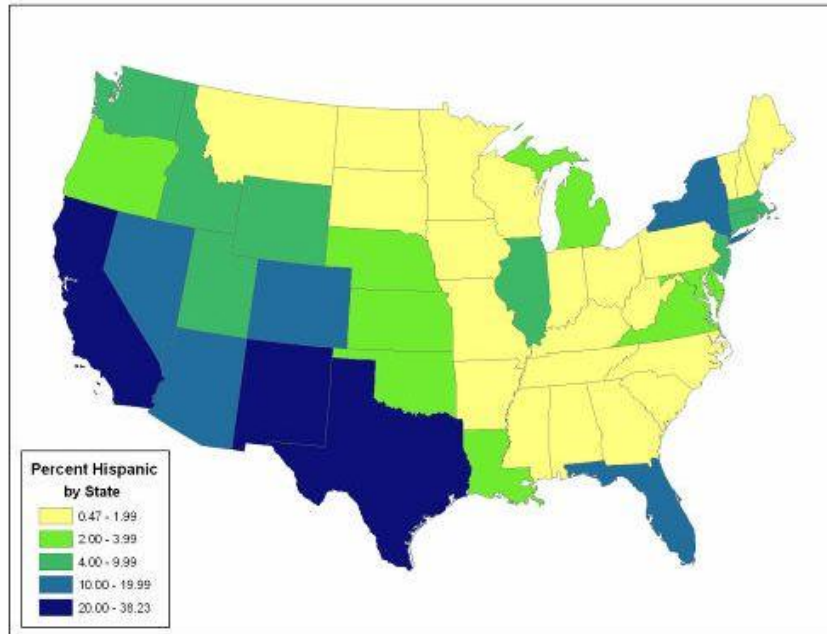


Figure 2.11. Choropleth mapping of Hispanic distribution over states [26]

#### 2.2.4. Multiscale visualization

When exploring large datasets, analysts often work through a process of “*Overview first, zoom and filter, then details-on-demand*” [27]. This principle is the key motivation for multi-scale visualizations.

As the user navigates through the scene, system switches between symbolizations and data frequencies in order to keep the density of information constant. In “overview” state whole data must be visualized and to avoid overwhelming the reader with an unrecognizable amount of information, details must be vanished, data must be highly abstracted. Too much detail will hinder the overview.

In the proceeding levels (as the user zooms in the scene) while the area displayed gets closer to camera, density of information will get lower. In these levels more detail must be represented to adjust the information density.

In multi-scale visualizations changes in information density may be done in two ways. One way is processing data (filter , aggregate, etc.) before visualization process. The other way is leaving the data untouched and changing the symbology, such as showing a

city in the overview level with a polygon and as the user zooms in letting the labels (city name) appear.

An example of processing data for the visualization changes in data density is Legible Cities[28]. Geographical data has a multiresolution character, since it is structured from blocks, tracts, counties, states and so on. Multiscale systems are suitable with their flexibility to make observations in different scales without breaking the interrelations. Legible Cities is an urban visualization system making benefit of this concept. Users have both opportunities to observe relationships of neighborhoods and to look at individual buildings. Abstraction of data is done via clustering algorithms.

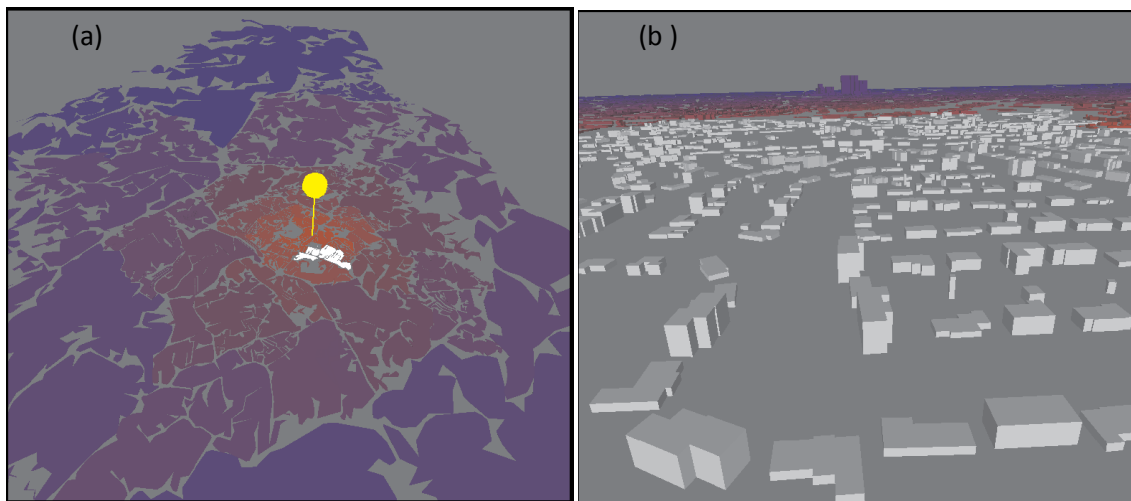


Figure 2.12. 3D model of Legible Cities (a) viewing clustered neighborhood, (b) individual building viewing.

There are two views available in Legible Cities: a 3D model view and a matrix of multidimensional data which is displayed in a separate window. Although the interrelation between buildings and geographical regions are visualized in a self explaining way, the data window of the application is rather complicated and needs some extra effort.

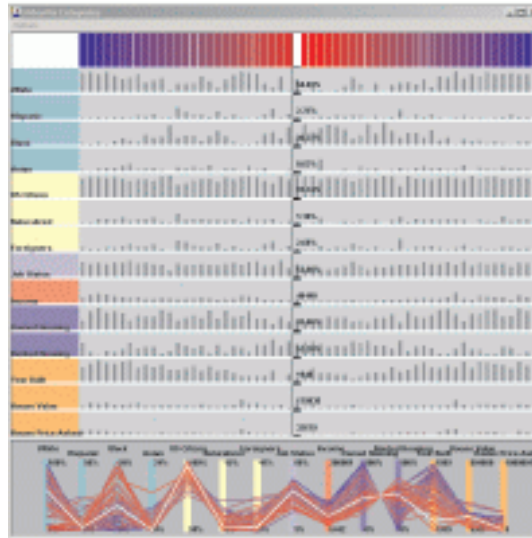


Figure 2.13. Matrix view of Legible Cities.

### 2.2.5. Geographic visualization

*Geographic Visualization: The use of concrete visual representations whether on paper or through computer displays or other media- to make spatial contexts and problems visible, so as to engage the most powerful human information-processing abilities, those associated with vision. (Allan MacEachren, 1992)*

Geographic visualization is one of the subdivisions of information visualization. Geography, space, has such an influence in cognition of information that in most of representations, terrain is included as a reference point for the data. When geographic mapping of the data is achievable, visualizing data in relation to its spatial values will guide the system to an intuitive depiction.

Cartography is one of the most suitable application areas for multi-scale information visualizations. It has a character which tends to be scale-specific and inter-related between scales. Since there are many attributes, relations and details in a map, mapmaker decides for each layer what to include and not to include in the representation to highlight the underlying pattern of the subject.

In the last years with the vast spread of Google Maps and Google Earth usage, geographical visualization systems with easy to use interfaces are increasing. Some of these systems are mentioned below.

Cartograms are geographical data visualizations, produced by the principle of distorting a map according to statistical factor represented. Although its regions are resized, objective of a cartogram is to resemble the original geography.

Before computers were available for reproducing maps, cartograms were drawn by hand. Since drawing them by hand is a very complicated task, algorithms that construct cartograms are frequently studied.

There are three types of a cartogram:

- The contiguous area cartogram aims to deform regions while preserving the adjacencies There are many papers published dealing with the problem of making continuous cartograms that strictly retain the original topology of given geography.[29,30,31,32]
- The non-contiguous area cartogram [33] scales down the regions to obtain desired sizes but generally the adjacencies are lost.
- The rectangular cartograms[34,35] represent each area by a rectangle.

Besides the usual appliance in information visualization, cartograms maintain a special representation of geographical data. They lay emphasis on the raw data instead of the area involved. For example in a population-based choropleth map densely populated areas may be less than the low populated areas, thus the general pattern of the corresponding map will be drawing attention to the lower values. Since the cartograms demonstrate the areas in relation with a parameter, the cartogram of the same data will reveal a completely different impression.

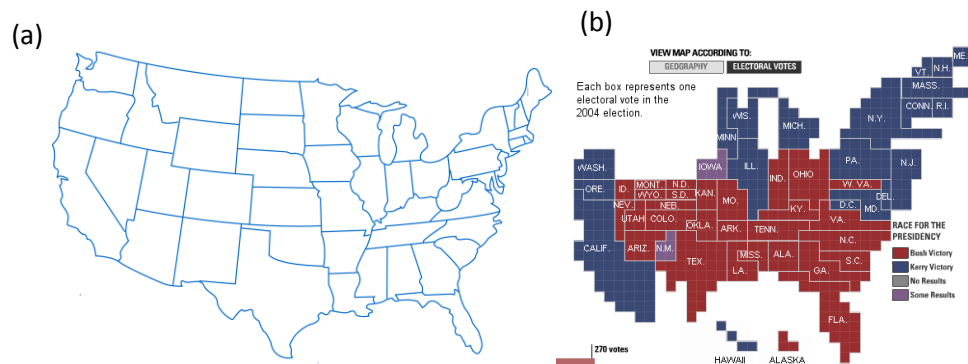


Figure 2.14. (a) Traditional map (b) New York Times Cartogram of Electoral College Votes.

Adoption of non-photorealistic techniques of computer graphics, to geographical visualization results in depictions which are familiar from paper based cartography. Buchin et. al. [36] and his colleagues improved a technique for computer generated reproduction of traditional terrain illustration. Terrain surface is visualized effectively with tonal variations and slope lines. Using a texture based approach, they developed a system which computes the surface measures and slope lines of the terrain, given a digital elevation model. This approach is suitable for producing reference maps more than thematic maps.

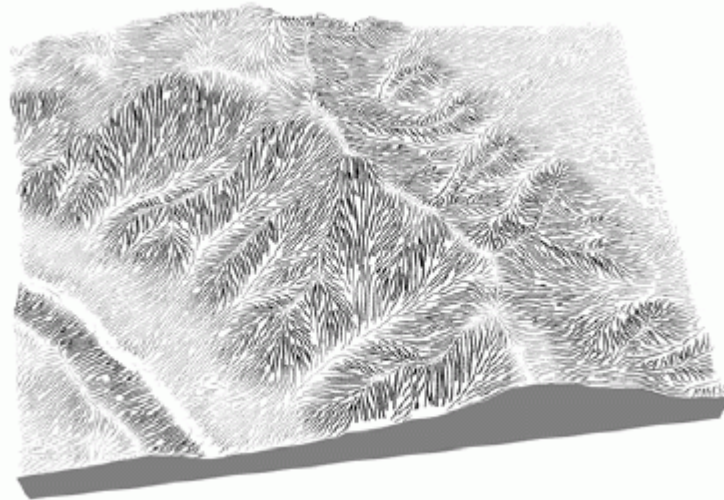


Figure 2.15. Adoption of non-photorealistic techniques of computer graphics to geovisualization [36]

After Google released their Google Maps Application Programming interface, which enables users to develop their own application, called *mash-ups* (Purvis et. al, 2006.) , feeding from Google's streamed data, combining geographical data from other sources, doing analysis and serving their outcome as a layer through Google map interface.

One of the mash-up examples is GMapCreator1 [37], a freeware application developed for 2D thematic mapping in Google Maps. It can read shapefiles [38] and generate thematic maps based on a field in its attribute table. These thematic maps are rendered as a series of raster image data and for different zoom levels these raster images are stored in a quadtree.



To sum up, this application produces raster images from shapefiles and displays them on Google Maps as an additional layer. An example of thematic mapping through GMapCreator1 can be tested online [39] .

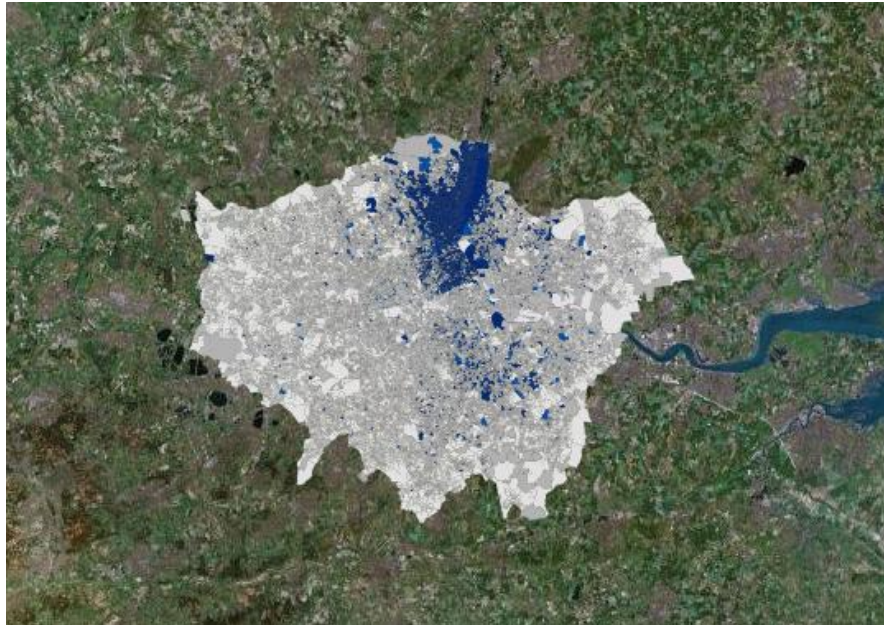


Figure 2.16. Turkish population distribution of London [39]

As the use of geobrowsers (such as Google Earth, Google Maps, Microsoft Virtual Earth) outspreads, online exploratory geographic visualization opportunities enlarges. Thematic mapping via Google Earth using KML [40], similar to given example above, is also possible. But since geobrowsers capability of making analysis on data is limited nowadays, additional GIS tools are required for generating a thematic map.

Jürgen Döllner, combining multiresolution texture models with geographical visualization, improved many innovative methods. He uses image pyramid and texture tree structures for storage and organization of texture layers [41]. To name some examples of his work :

- With the addition of a luminance texture on a cartographic or topographic texture a system for highlighting a region of interest is maintained in one of his studies. [42]

- For a level of detail terrain as the layer resolutions get lower, details get lost accordingly. In order to prevent this side effect shading is based on a topographic texture. [42]
- For visualizing thematic data, a 2D texture of thematic data is constructed on projected on 3D terrain (Figure 2.17). Multiple layers are produced with this approach and they can be turned on or off. 3D objects are included in thematic maps to visualize data in a different way.[41]

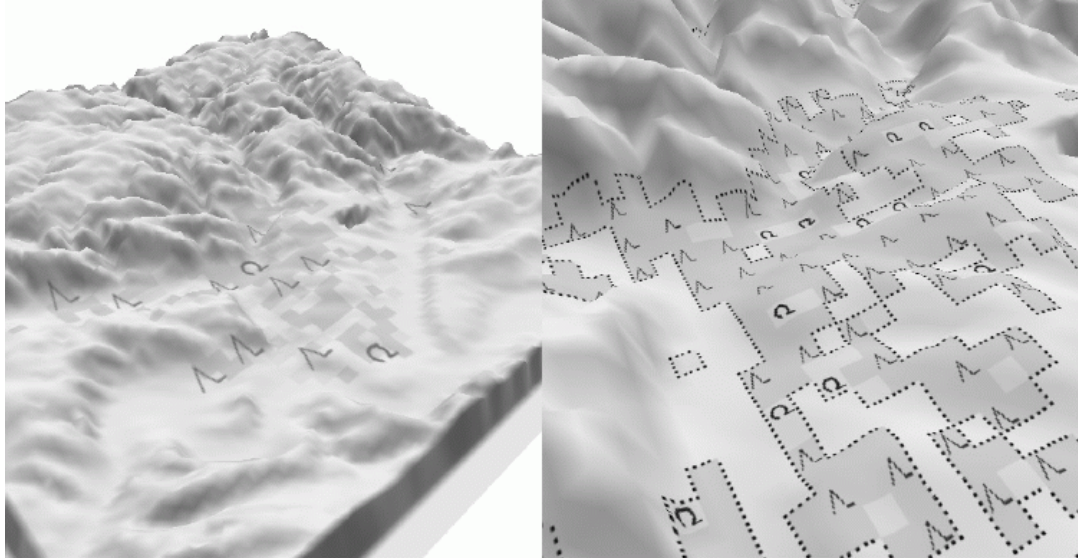


Figure 2.17. LOD design rules. The exact boundaries of forest areas are visualized if the camera comes close to the terrain. The textures are computed on the fly [41].

### **3. THE SYSTEM**

#### **3.1. Overview**

The objective of this study is to develop a self-explaining 3D statistical data visualization based on a terrain and to advance it as a multi-scale system which will perform level of detail according to camera locations.

If the aim is to illustrate spatial data, thematic maps are said to be one of the most generic methods. Consequently, the corresponding statistical data are chosen to be presented as a thematic map in our study. We improved a method with the intention of constituting thematic maps, automatically from the input data.

The program flow starts with reading the inputs and storing them. Then visualization process begins with partitioning the area into smaller subdivisions. These subdivisions are shaded according to their distance to camera and the resulting screen image is saved as a texture. Respectively terrain is constructed from the elevation grid and previously generated texture is wrapped on it. As the camera moves, the texture is modified and patched on terrain. There is a loop between generating texture and projecting it on terrain. This continues until the termination of the program.

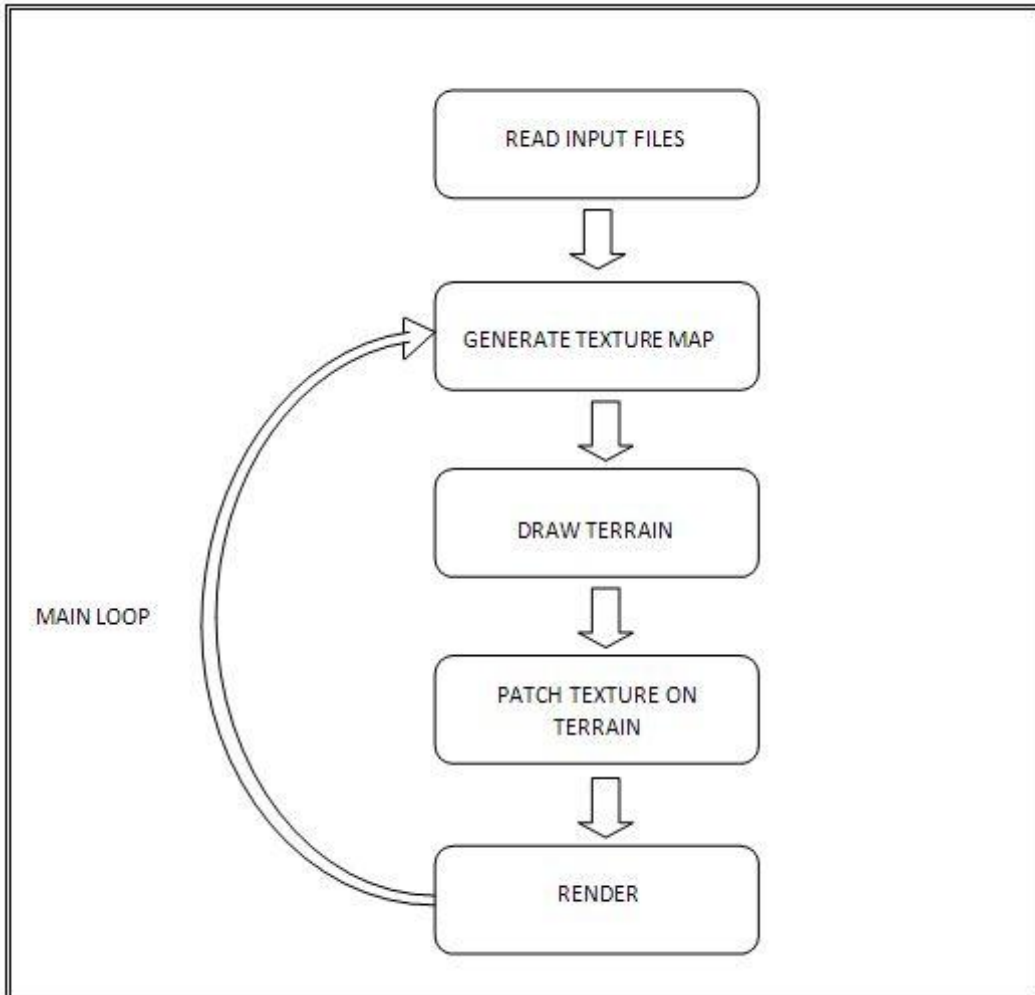


Figure 3.1. Program Flow Diagram.

### 3.2. Input Data

- Height Map:

This data is assumed to be elevation points, measured in equal distances in both X and Y directions. These points form a square shaped regular grid. Via making use of surface construction methods, terrain is created from this data layer.

- Statistical Data Sets:

Like the elevation data layer, this layer is also supposed to be consisting of data points gauged in constant distances on XY plane. Associated with each point, there is a statistical value. As a result this level's elements are placed in a square-like manner, too.

- Road Data :

In our software, roads are embodied as lines. These lines are nothing more than bounded points, which have X, Y and Z values. Thus the outcome is a 3D line network. Accordingly, road data layer is nothing more than a point cluster, which is paired into groups so as to make up lines.

- Polygon Data:

Likewise other data layers explained above, this level is a point collection as well. Point bunch is separated into groups, each representing a polygon. Aforementioned polygons are the boundaries of terrain's subdivisions.

### **3.3. Statistical Foundations**

#### **3.3.1. Exploratory data analysis**

When you don't have any hypothesis about the statistical data, rather than trying to fit it into standard forms, data should be explored in a manner to develop new hypotheses. That is what John Tukey (1977) introduced as Exploratory Data Analysis (EDA), and believed to be one of the most essential progresses in statistical analysis in the last 25 years.

In an effort to find out the characteristics of data, descriptive statistical methods are chosen to be used in this thesis. First of all, a raw table, a tabular display in which the measurements are listed from lowest to highest, is selected as the starting point for the investigation. Given that raw table consists of sorted values, specific information about data can be discovered with bare eye exploration. It provides the minimum and maximum values and additional information may be gained with assistance of simple mathematic operations. Such as:

*Range: Length of the smallest interval containing all the data. It is calculated by subtracting the smallest observations from the greatest. [43]*

*Mode: The most frequently occurring value, and is thus generally useful for only nominal data, such as land cover map. [23]*

*Median: The middle value in an ordered set of data or, alternatively, the 50th percentile, because 50 percent of the data are below it.*[23]

*Mean: It is often referred to as “the average” of the data and is calculated by summing all values and dividing by the number of values.*[23]

Moreover, when examined carefully duplication of values and outliers (values that are quite unusual) may be observed.

Even if raw tables are useful for having a summary of data, they do not provide information about the distribution of values.

After getting an overview of the data, a need of finding out the distribution pattern emerges. As an initial attack, producing a pie chart and box plot of values is suitable. They are good visual displays for detecting the data in predetermined intervals.

Lastly data is prepared for the visualization process. There are two types of data arrangements used in our system, classed and unclassed mapping. Classifying the raw data as combining them into classes or groups, with each class represented by a unique symbol results in a *classed map*; in contrast, if each raw data value is depicted by a unique symbol, an *unclassed map* results.[23]

There are major advantages for both classed and unclassed maps. While unclassed maps portray the data distribution more precisely, classed maps, having narrow number of categories, makes the depiction easier to understand.

Two options are available in our study for data classification; equal intervals method and quintiles method, each suitable for different purposes. Unclassed maps are abstracted via normalization process.

### **3.3.2. Equal intervals**

Equal intervals (or equal steps) method is about forming up classes which occupy same width along the number line.

After deciding the number of classes that the data will be separated into, through dividing the range of data by the number of classes, a class interval along the number line will be established. Adding up this value to lowest value of data recursively, the upper limit of each class may be determined. Besides lower limits of classes are the values coming after the highest value of the prior class.

After the upper and lower limits of classes are calculated, we can now use them in our pie chart and observe which classes are empty, which are overcrowded and get a sense about the distribution of data.

In equal intervals method, some ranges may be blank areas and some ranges may get overcrowded. Those limits will be meaningless. Pie chart of data will reveal the utility of a legend prepared for equal interval limits.

### **3.3.3. Box-plot**

One technique representative of Tukey's work is the *box plot*. Here, a rectangular box represents the interquartile range, and the middle line within the box represents the median, or 50th percentile. The position of the median, relative to the 75th (upper quartile) and 25th (lower quartile) percentiles, is an indicator of whether the distribution is symmetric or skewed.[23]

Giving further information about the technique, the interquartile range is the absolute difference between the 75th and 25th percentiles, or where the middle 50 percent of the data. An important characteristic of the interquartile range is that it, like the median, is unaffected by outliers in the data.[23]

The following quantities (called *fences*) are needed for identifying extreme values in the tails of the distribution: [43]

lower inner fence: lower quartile - 1.5 \* inter quartile

upper inner fence: upper quartile + 1.5 \* inter quartile

lower outer fence: lower quartile - 3 \* inter quartile

upper outer fence: upper quartile + 3 \* inter quartile

A point beyond an inner fence on either side is considered a *mild outlier*. A point beyond an outer fence is considered an *extreme outlier*. [43]

We used box-plot approach to visualize the extreme and mild outliers in data and to envision the distribution characteristics of it.

### 3.3.4. Normalization

*Normalization refers to the division of multiple sets of data by a common variable in order to negate that variable's effect on the data, thus allowing underlying characteristics of the data sets to be compared.* [43]

In order to diminish the effect of dispersity in data, instead of using raw values, data points' distance to minimum of input set, is divided into the range of data.

$$\text{Normalized Value} = \frac{\text{raw data} - \text{min.of data}}{\text{max of data} - \text{min of data}}$$

This way data's relativity to its range is visualized.

This technique also has some inefficiency, such as if there are a bunch of extreme values in a data set and remaining members of the set is distributed in a narrow range, differentiation of values will be difficult.

### 3.4. Visualization

Visualization process of our study is simply generating a thematic map from input statistical data set\s according to level of detail. System is composed of three steps. Namely:

- Partitioning the geography into smaller subdivisions.
- Filtering the information according to subdivisions' size.
- Colorization of these subdivisions.



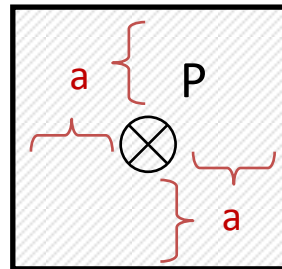
### 3.4.1. Subdivisions

#### Uniform subdivisions

In this mode subdivisions are equal squares. Switching between different resolution levels is maintained by changing the size of squares. For each detail level precalculated abstract data is used. Here is the 3 resolution levels of data and how they are estimated:

##### *Highest Resolution:*

Terrain is partitioned into small, square-shaped enumeration units. Each one corresponds to a unique point in data grid. Squares are assumed to have the value of the corresponding grid point.



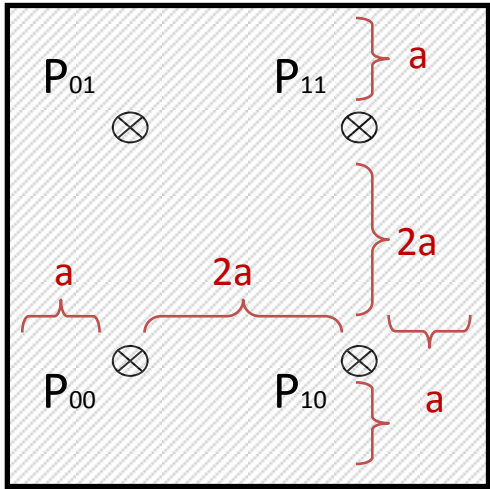
P is the grid point

a is the half size of unit square's edge

Figure 3.2. Unit subdivision area for highest resolution.

##### *Medium Resolution :*

As the distance between a point and the camera gets larger, details become less recognizable and number of points displayed increases so as the density of information. To avoid a crowded visualization, clusters are formed from grid points. For medium resolution, terrain is separated into squares containing four grid points. Each square's value is the average of corresponding four data points' values.



$P_{ij}$  is a grid point  
 $a$  is 1/4 of unit square's edge

Figure 3.3. Unit subdivision area for medium resolution, consisting of 4 grid points.

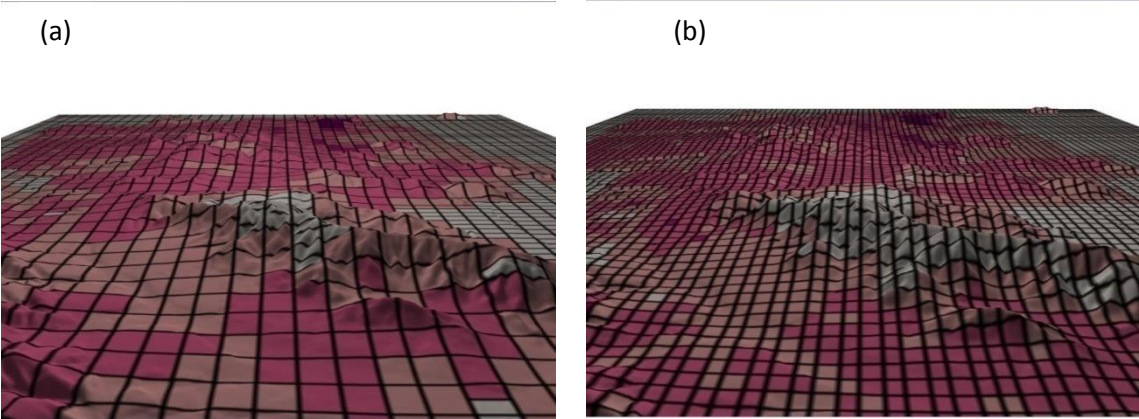


Figure 3.4. Unit subdivision areas for (b) highest and (a) medium resolution.

*Lowest Resolution:*

This phase is the lowest resolution state. This time terrain is separated into relatively bigger squares which covers 16 grid points. Average of each 16 grid points is the value of correspondingsubdivision area.

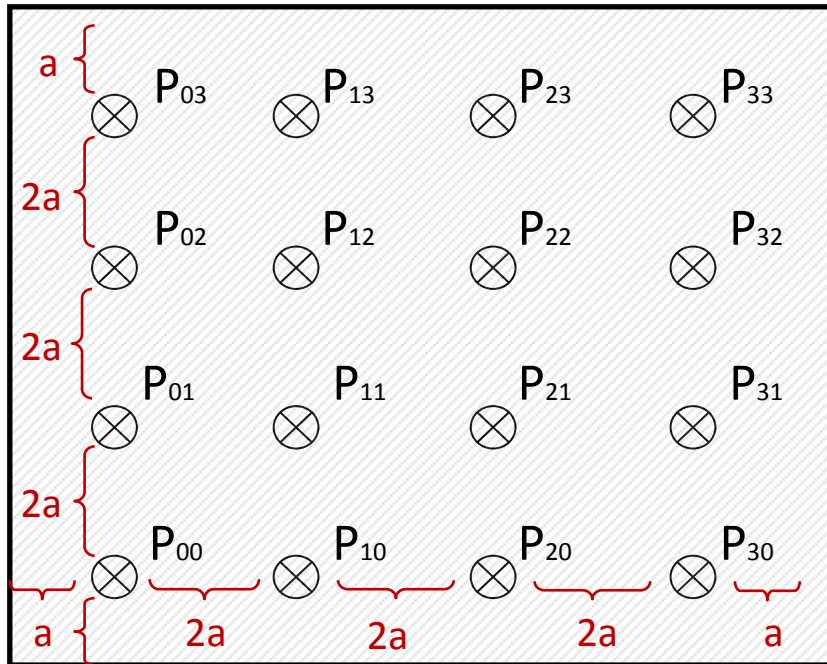


Figure 3.5. Unit subdivision area for lowest resolution consisting of 16 grid points, where  $P_{ij}$  is a grid point and  $a$  is  $1/8$  of unit square's edge

### Non-Uniform (Vectorial) Subdivisions

Natural boundaries of geographical subdivisions (blocks, tracts) are more likely to be irregular, than having regular shapes which enable mathematical predictions. Accordingly, determining the value of an irregular subdivision requires some mathematical effort.

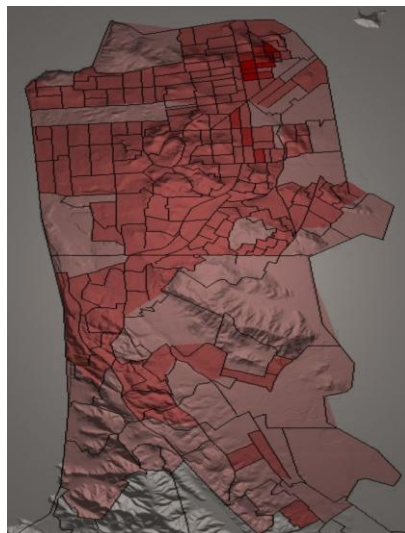


Figure 3.6. Shading of nonuniform subdivisions.

First thing to do is to find how many population points are inside the subdivision, which is in fact a polygon. Testing whether a point is inside polygon or not is a very common problem encountered in computer graphics.

Consider a polygon made up of  $N$  vertices  $(x_i, y_i)$  where  $i$  ranges from 0 to  $N-1$ . The last vertex  $(x_N, y_N)$  is assumed to be the same as the first vertex  $(x_0, y_0)$ , that is, the polygon is closed. To determine the status of a point  $(x_p, y_p)$  consider a horizontal ray emanating from  $(x_p, y_p)$  and to the right. If the number of times this ray intersects the line segments making up the polygon is even, then the point is outside the polygon. Whereas if the number of intersections is odd, then the point  $(x_p, y_p)$  lies inside the polygon (Figure 3.7). [44]

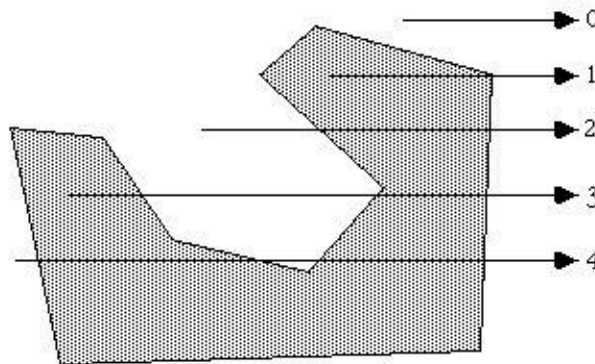


Figure 3.7. Illustration of how to determine whether a point is inside a polygon or not [44].

Value of each subdivision area is evaluated by taking the average of points inside it.

### 3.4.2. Colorization

The primary aim of producing a choropleth texture is to give a sense about the data density of a place. In the means of representing another dimension, color was used. There are two different methods we used to colorize produced choropleth maps; shading and hatching.

### *Shading*

It is known that objects are perceived in comparison to or in relation to other objects or the surrounding areas of the composition and the weight of an object, the impression of heaviness or lightness, depends on its brightness or luminosity. Consideration of the interdependencies between light and dark is an important aid in creating an accomplished scene. [45]

In light of this information, data were mapped to a legend of colors, which varies from lighter to darker. Areas where information density is high are shaded with darker colors and the less dense areas are shaded with lighter colors.

The *RGB color model* is a color model in which red, green, and blue are combined together in diverse ways to come up with a wide range of colors. RGB name comes from the initials of the three primary colors, red, green, and blue. Each color's intensity is shown with a number between 0 and 255, from least intensity to full intensity respectively. Blending of red, green and blue in full intensity form white color, while a combination of them each having intensity 0 creates black color.

Subdivisions are shaded based on their values calculated in the previous step. There are three choices that the user can switch between, during runtime.

First method uses unclassed maps:

RGB values of each region are assigned in relation to the normalized value mentioned in section 3.3.4. All regions have full intensity of red while intensities of green and blue changes due to normalized values of regions.

R : 255

G :  $255 * (1 - \text{normalized value})$

B :  $255 * (1 - \text{normalized value})$

The outcome is a color-ramp from red through white.

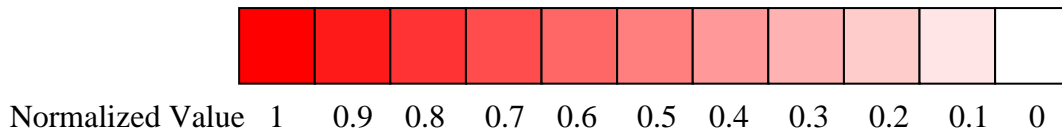
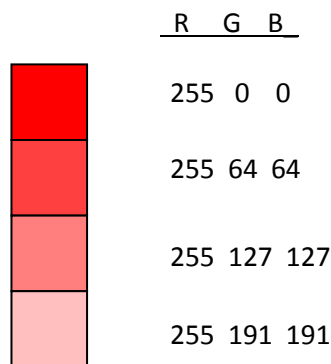


Figure 3.8. Legend for visualization of unclassified data.

Second and third methods produce classed maps:

In these two methods, before shading subdivisions, data are classified into groups using equal intervals and box-plot methods respectively. Subsequent to classification step, each group is mapped to a color. While seeking for the appropriate color set for a choropleth map, Color Brewer website may be very helpful [46].

*Equal Intervals Method:*



Data are classified into four equal intervals. A pack of four colors, from red to a very light pink, are generated with a similar calculation mentioned in unclassified mapping and mapped to each interval.

Figure3.9. Scale generated for the visualization of data classified with equal intervals method.

*Box-plot Method:*

Data are classified into 5 equal intervals according to parameters of box-plot method. As mentioned in the previous section Color Brewer is a useful tool for determining suitable colors for a thematic map. Those sequential colors and their RGB values are obtained from Color Brewer [47].

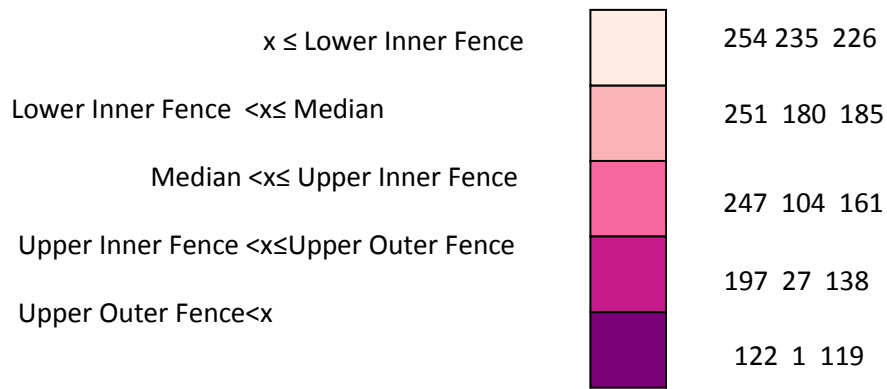


Figure 3.10 scale generated for the visualization of data classified with box-plot method.

### 3.3.4 Hatching

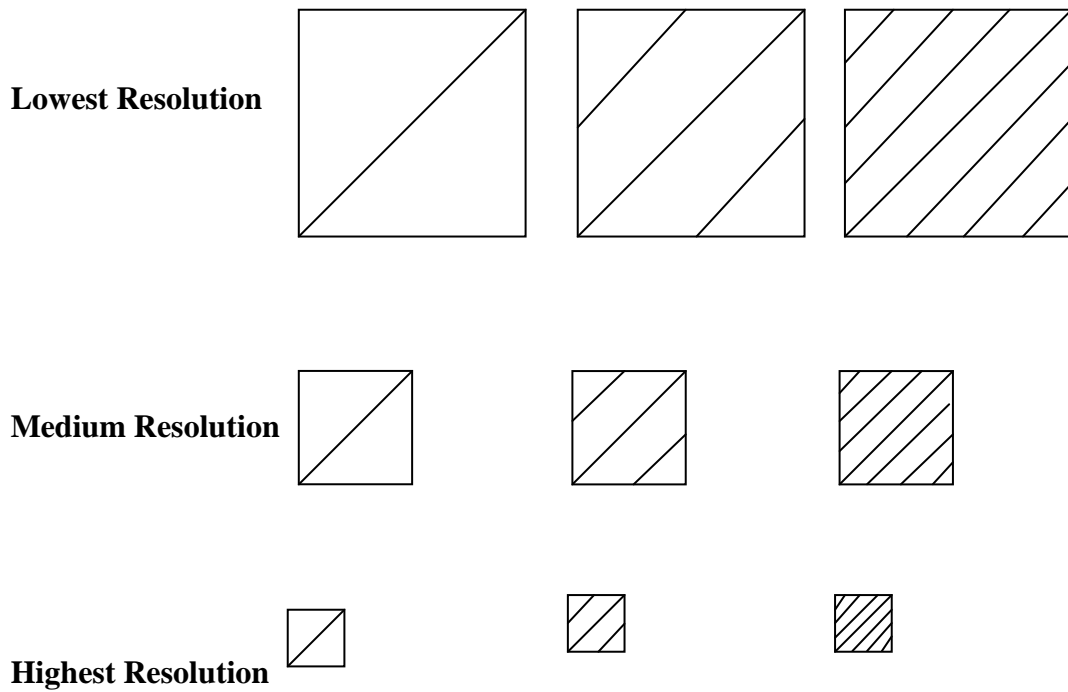


Figure 3.11. Hatching method for different level of details.

This mode is an attempt to visualize the data density with hatches, which is a method frequently used in paper based maps.

Differentiation of different classes is the key idea of this concept. While continuity is preserved with the parallel alignment of the lines, contrast is maintained with the additional strokes.

More densely populated squares are hatched with more lines and there are 3 population classes. The same method of hatching is applied to all sizes of squares, in different level of details.

Since dredging is applied to classes of points, unclassified data can't be represented using this method.

### **3.3. Terrain Visualizer**

While modeling a geography-related information visualization system, drawing the corresponding landscape is vital. The terrain forms a base for the structure and acts as a reference point for displayed spatial data. Generally speaking, including the landscape improves the comprehension of representations and provides useful insights.

The influence of geographical and spatial metaphors is so strong that they can be found in most information visualization systems. [48]

The structure of digital terrain model used in this study is based on regular rectangular grid coordinates. The data consists of elevation values measured in equal distances, through X and Y directions. Thus each grid point has X, Y and Z coordinates.

For the purpose of forming a continuous surface, triangulation technique was performed. Each point, except the ones on the edges, is shared by 6 triangle shaped around it. Each triangle shares vertices and edges with its neighbors, thus the continuity of surface is maintained and possible crack formations are prevented.



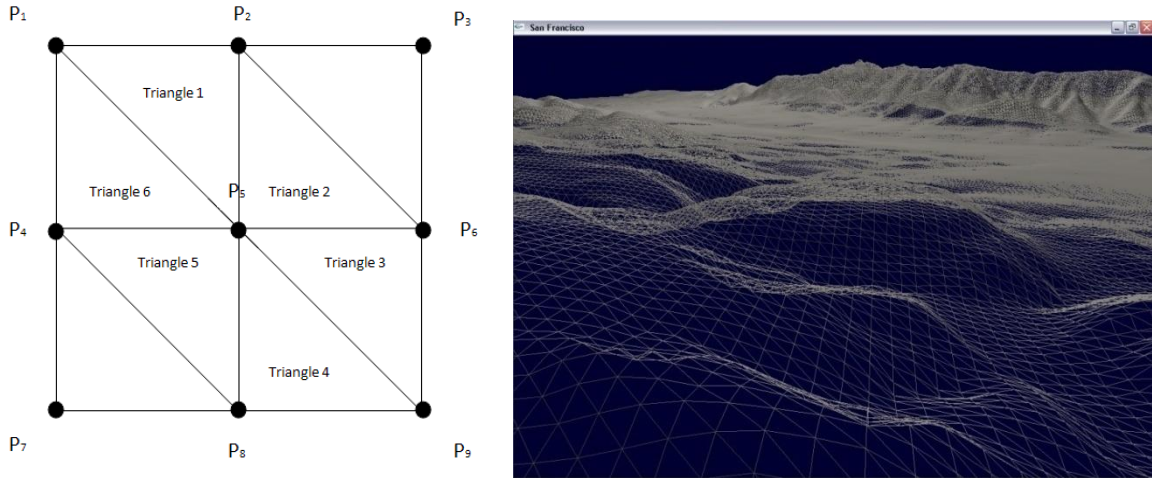


Figure 3.12. wire model of the terrain.

As shown above in Figure 3.12. most of the grid points (except for the ones on edges) are the common vertices (intersection point) of 6 triangles surrounding. With the intention of preventing discrete lighting effects, average of 6 triangles' normal are assigned as the normal of corresponding point.

While calculating triangle normals, Newell Method[49] was used. It is suitable for finding normal of planar polygons. The method computes the components  $m_x$ ,  $m_y$  and  $m_z$  of the normal  $\mathbf{m}$  according to the formulas:

$$m_x = \sum_{i=0}^{N-1} (y_i - y(\text{next } i))(z_i + z(\text{next } i))$$

$$m_y = \sum_{i=0}^{N-1} (z_i - z(\text{next } i))(x_i + x(\text{next } i))$$

$$m_z = \sum_{i=0}^{N-1} (x_i - x(\text{next } i))(y_i + y(\text{next } i))$$

Figure 3.13. Newell Method for calculating surface normals.

where  $N$  is the number of vertices in the face,  $(x_i, y_i, z_i)$  is the position of  $i^{\text{th}}$  vertex, and “next  $i$ ” is the vertex is the index of the next vertex around the face after  $i$ .

### 3.4. User Interface

#### 3.4.1. Navigation

Navigation controls of the software are via keyboard and mouse. Rotation and movement along three axes is possible while terrain exploration is done through first person narration.

User can also choose between different statistical methods, for the visualization of the input data, again using the keyboard.

#### 3.4.2. Legend

With multiple statistical methods present, it is necessary to provide a legend that indicates the limits of each interval and the corresponding color they are mapped to. Legend bar is placed at the bottom of the display window. As the user switches between representations, scale color and limits are updated.

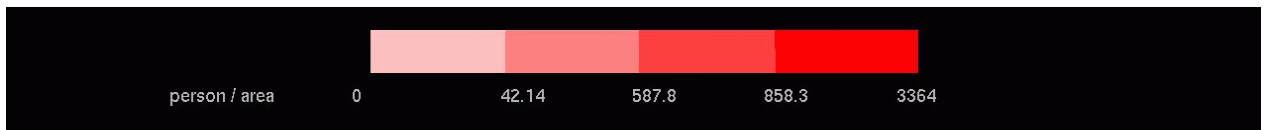


Figure 3.14. The legend.

## 4. CASE STUDY

### 4.1. Population Data of San Francisco

While implementing the program, as a case study, the city of San Francisco is selected. Since San Francisco is a city with hills and seaboard, it has a wide elevation range. Thus it is suitable for a 3D visualization system.

#### 4.1.1. Input Data

Height Map:

Height map of San Francisco is downloaded from USGS (United States Geological Survey) website [50]. Data maintained was in NED format, which is a raster file with 1 arc second resolution. Boundaries of input were latitudes -122.52 and -122.35, longitudes 37.59 and 37.82.

Statistical Data Set:

Demographics of California State are downloaded from U.S. Census Bureau website [50]. This data is based on year 2000 U.S. census and has a resolution of 7.5 arc seconds. Since downloaded files covered the whole state, the region of interest was extracted with the help of ArcGis [51].

Road Data:

Road data is maintained from ESRI resources. [52]

Polygon Data:

Block and tract subdivision data, which are based on year 2000 U.S. census, are maintained again from U.S. Census Bureau website [50].

#### 4.1.2. Exploratory Data Analysis

Since the characteristics of data aren't known in order to develop an hypothesis about data, population data of San Francisco are investigated carefully.

- *Raw table*

From this initial attack minimum, maximum, median, mean and range of data are obtained:

Minimum population = 0

Maximum population = 3363.54

Range = 3363.54

Median = 42.14

Mean = 201.96

- *Equal Intervals*

After observing raw table, we found out that the input population data consists of 5163 points and there are outliers in the data which causes the range to be wide.

With the intention of getting an overview about the distribution of data, information should be partitioned into many intervals. The density of each interval is assumed to reveal facts about the distribution of data around number line.

34 intervals with a constant range of 100 are established. Class limits are calculated and each point is placed into one of these intervals. The outcome is shown below as a pie chart,

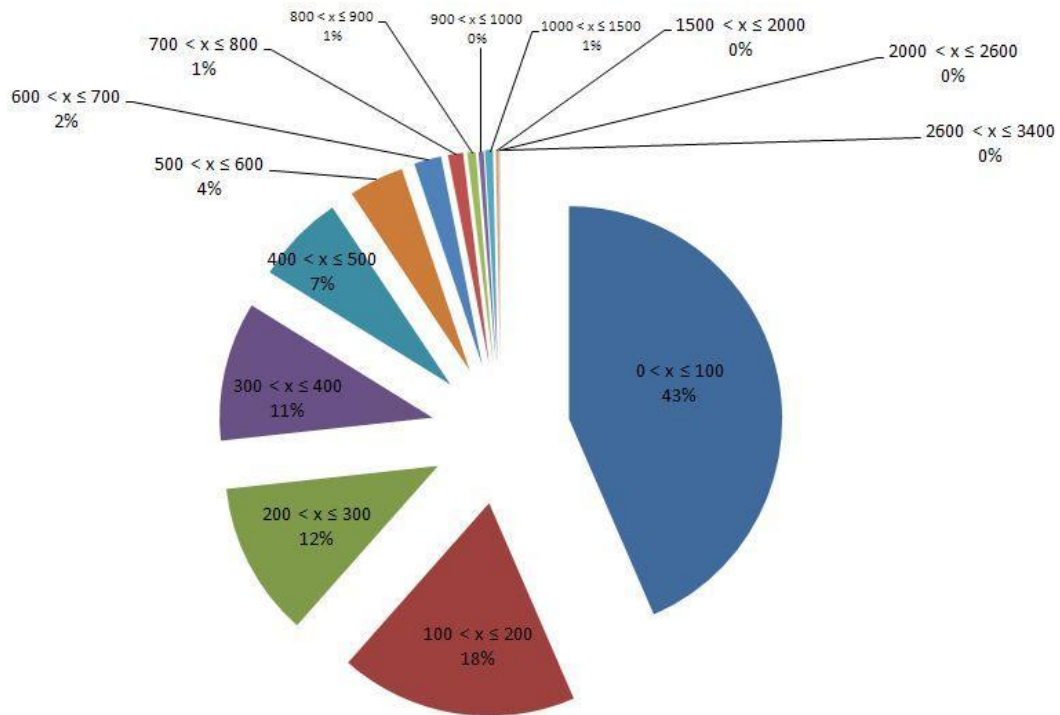


Figure 4.1. Pie Chart of San Francisco population data.

As one can see from above pie chart and table, this data is gathered between 0 and 1000. The elements between 3400 and 1000 are insignificant compared to other cluster. It is apparent that using a legend with equal interval limits won't be appropriate since some intervals will overlap with empty spaces.

Another conclusion that may be drawn via observing those visual displays is that the higher the population interval gets, the smaller the number of elements in relevant interval gets. This may lead to difficulty in differentiation of values between 0 and 1000.

- *Box-Plot of Data*

Drawing the box-plot of data will help us gain further information about the distribution of data. A box-plot consists of a box representing the interquartile range and a line, representing the median, passing through this box. Sticking with this definition, first things to calculate are median and interquartile range:

*Lower Quartile* : 17.82

*Median* : 42.14

*Upper Quartile* : 317.35

*Interquartile Range* = Upper Quartile – Lower Quartile = 299.53

Following these, fences are calculated in order to track mild and extreme outliers:

Lower Inner Fence: Lower Quartile - 1.5 \* Interquartile Range = 858.31

Upper Inner Fence: Upper Quartile + 1.5 \* Interquartile Range = 587.83

Lower Outer Fence: Lower Quartile - 3 \* Interquartile Range = -133.44

Upper Outer Fence: Upper Quartile + 3 \* Interquartile Range = -523.14

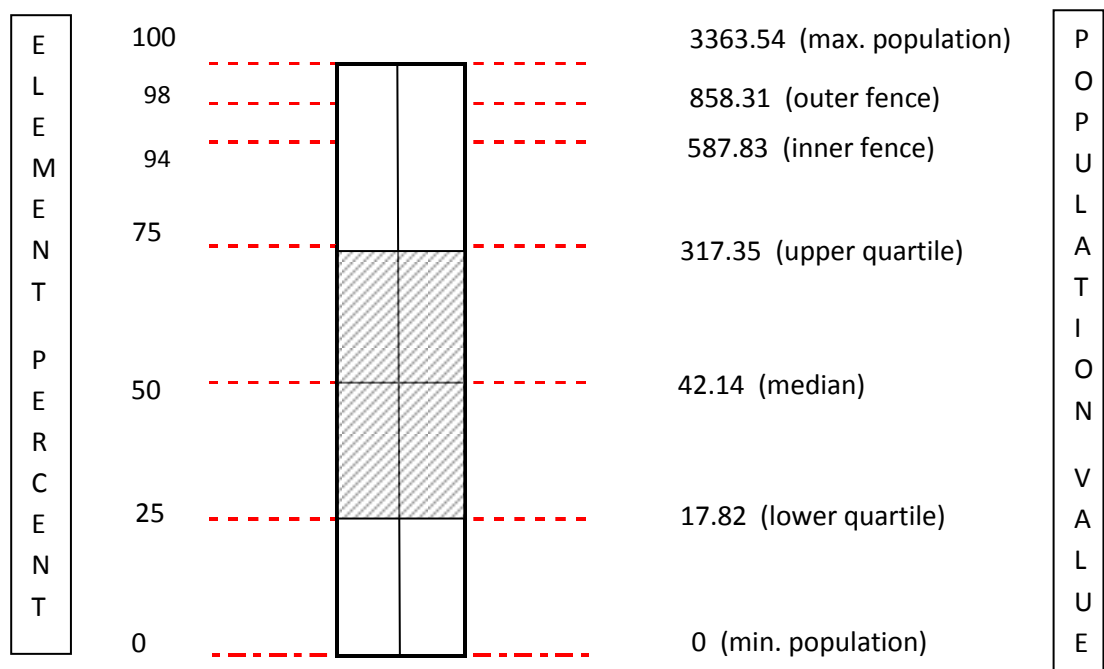


Figure 4.2. Box plot of San Francisco population data.

By applying box-plot method, 78 extreme and 210 mild outliers are detected. Knowing that the data are strongly reliable, none of those outliers are discarded from data. Besides, population data of San Francisco vary significantly, since traditional two or three floored buildings and recently constructed high-rise structures both exist in this city. With

the intention of checking this hypothesis, locations of high-rises of San Francisco[53] are marked on Google Earth[12], and a comparison is done with population densities we use.



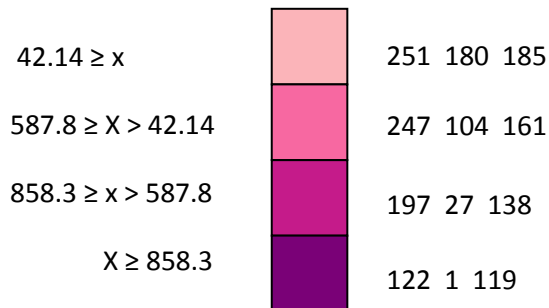
Figure 4.3.(a) High-rise buildings of San Francisco, located on Google Earth,  
 (b) Population density pattern of this visualization system.

The cluster of information at the north east of the city is a pattern that matches in both of the maps.

### 4.1.3 Colorization

- *Shading*

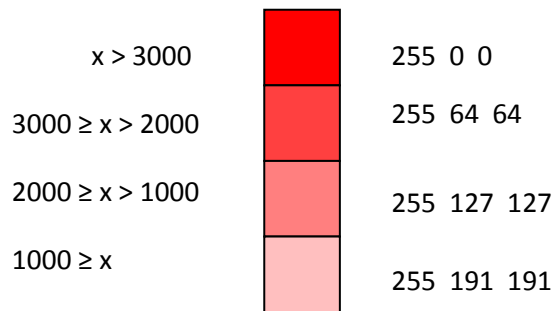
#### *Uniform Subdivisions*



After intervals are decided via box-plot method, this color legend is mapped to relevant classes. Since there aren't any elements in the first interval (Lower Inner Fence  $< x \leq$  Lower Quartile), it isn't included in this legend.

Figure 4.4. legend for box-plot method.

### *Non-Uniform Subdivisions*



In the non-uniform subdivisions mode, interval limits are decided by equal intervals method or quintiles method, too. After class limits are determined, color legend of the map is generated manually.

RGB values of the legend is given on the left.

Figure 4.5. legend for equal intervals method.

### *Hatching*

In the light of box-plot and equal intervals methods, data is separated into 3 different population density classes. Each density class is mapped with a hatching style. The user can switch between two methods used in this study for defining classes (equal intervals, quintiles), while unclassified data can't be visualized.

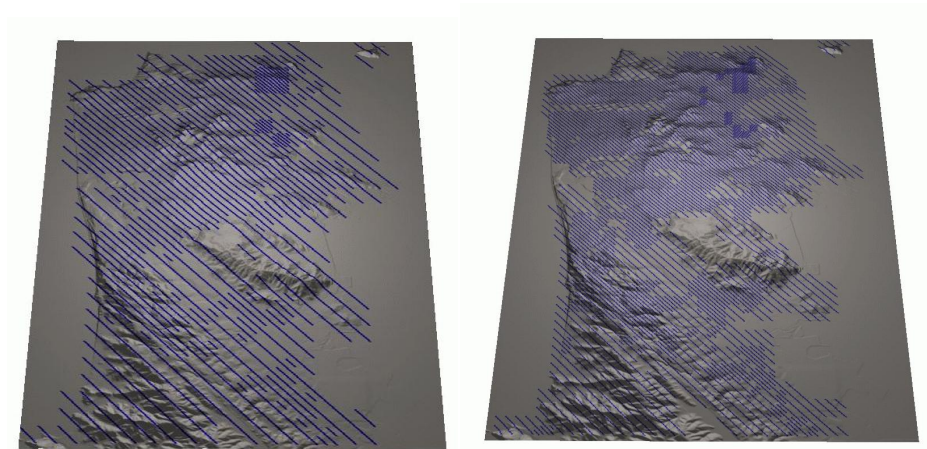


Figure 4.6. Hatching styles for different level of details.



#### 4.1.4. Road Network

As mentioned in the input data section, roads are represented as lines, which consist of bounded 3 dimensional points. Since road coordinates are in accordance with the landscape coordinates, drawing the road network is as simple as drawing a line loop.

Road network of San Francisco reinforces the population density transitions. As a general pattern in highly populated regions, road network is denser, while in regions with less population, road network is less crowded.

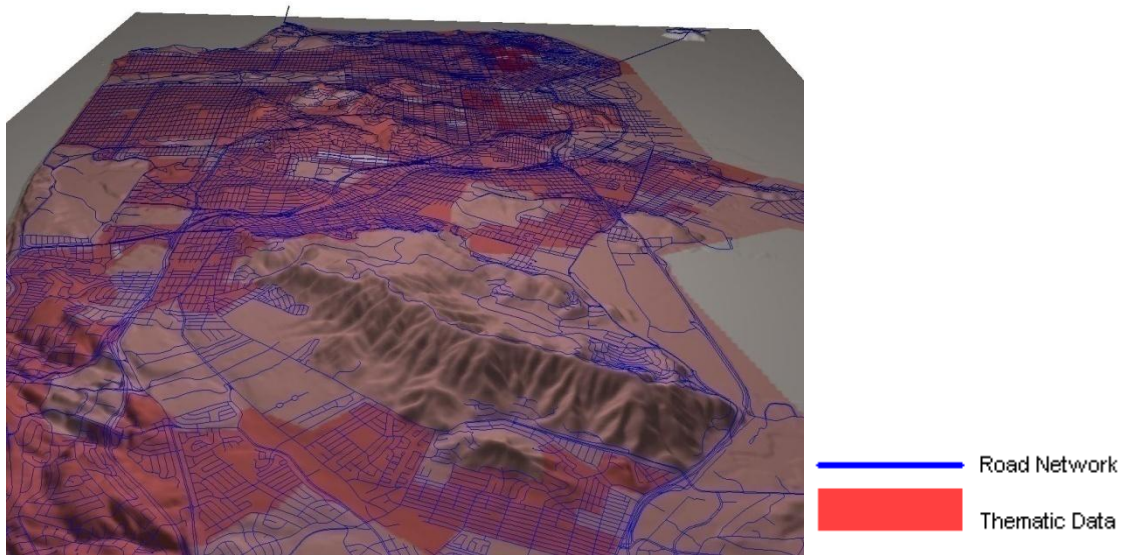


Figure 4.7. Road network of San Francisco reinforces the population density transitions.

## 4.2. Multiple Data Visualization

In order to test the performance of our system, in multivariate data visualization, a synthetic data layer is added to the existing San Francisco case study. All inputs are identical to the previous case study's inputs, except an additional randomly generated statistical data set.

### 4.2.1. Symbology

This time instead of using both hatching and shading options for visualization of one data set, hatching technique is used to visualize San Francisco's demographic data while the additional data set is classified and shaded due to both equal intervals and box-plot legends.

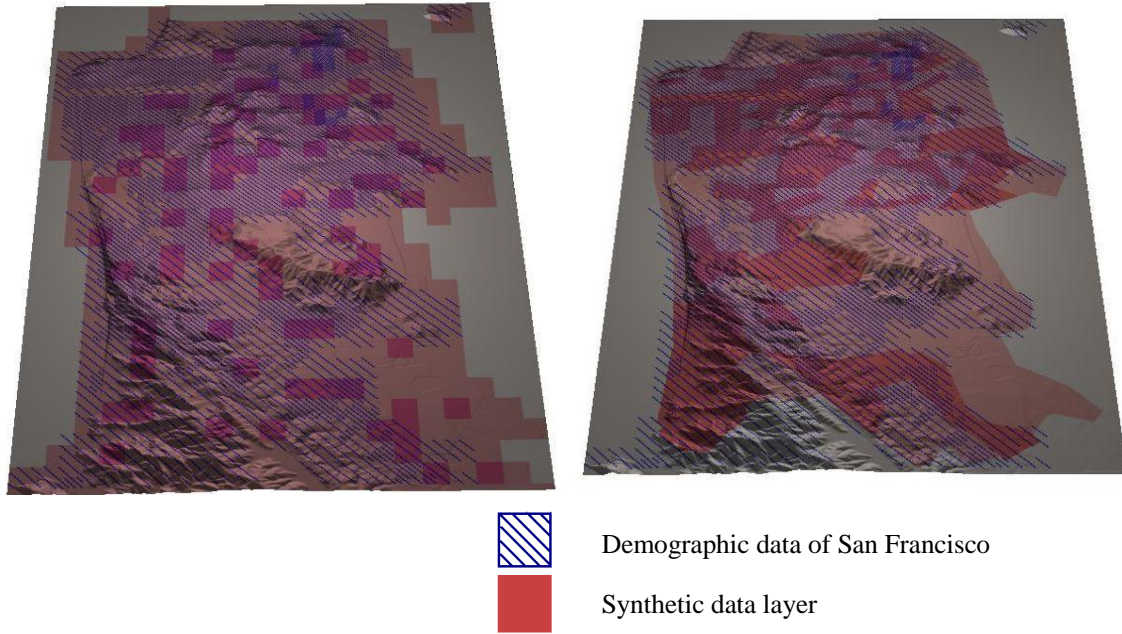


Figure 4.8. Visualization of multiple data layers.

The output is an efficient multivariate data representation. Since the colorization of two layers created a contrast, information of different layers are easily distinguished.

#### 4.2.3. Road Network

When road network layer is turned on, visualization becomes too crowded and complicated. This is because hatches and roads both have lines in their symbolization. Therefore the dimension of data layers should be kept up to 2; any layers more than that make it complicated, in this state.

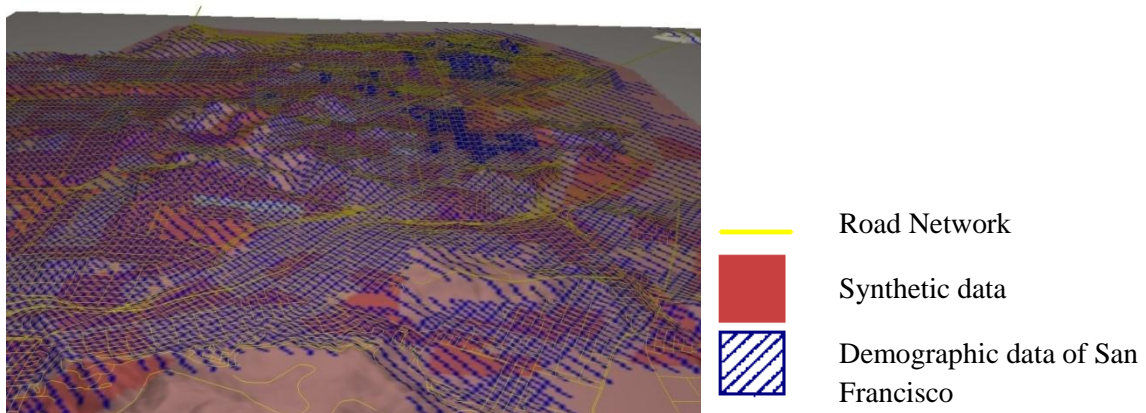


Figure 4.9. Visualization of two layers of statistical data and road network together.

### 4.3. Comparison with ArcScene

*ArcScene is a 3D visualization application that allows you to view your GIS data in three dimensions* [54]. In order to compare our study with the state of art in geographical information visualization systems, we chose to visualize our data in ArcScene. Some major advantages of using our software are observed during this comparison:

Since our study uses level of detail approach while visualizing thematic data, as the camera zooms out of the scene depiction of information is simplified and system switches into a less detailed visualization. ArcScene uses the same detail level regardless of camera position, thus as the camera gets away from representation region, information density increases.

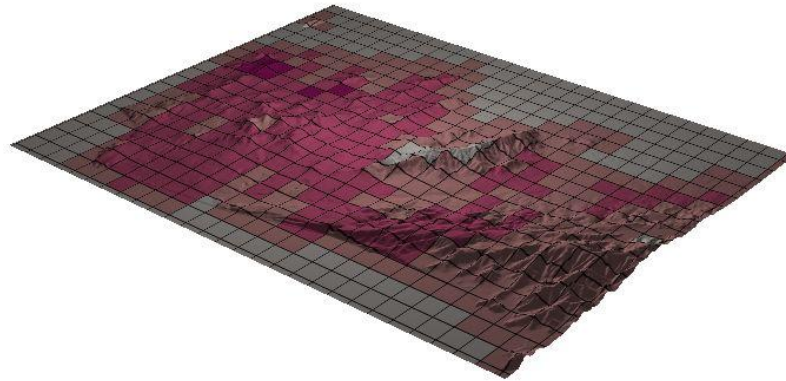


Figure 4.10. Via using level of detail information density of the depiction is kept constant, so observing the data from a distant point is possible in the proposed system.



Figure 4.11. Due to unvarying size of squares, information density increases as camera gets away from the representation region.

Besides, observation of thematic data from a distant point is nearly impossible with ArcScene due to clustering of unvarying size of squares (Figure 4.12) and those squares' borders creates anti aliasing as a side effect (Figure 4.13).

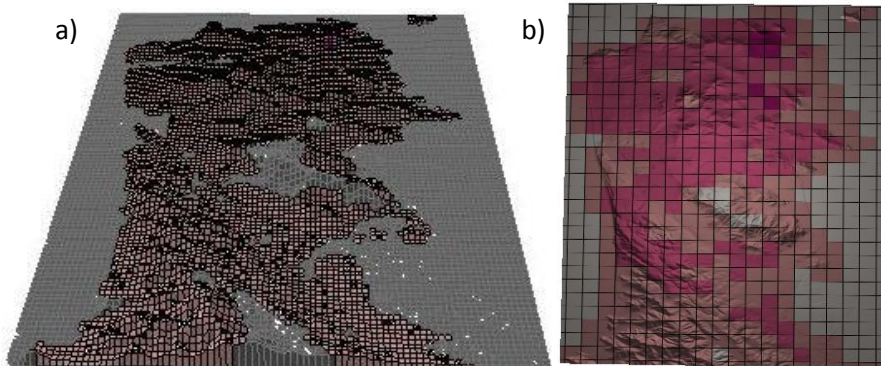


Figure 4.12. Clustering effect of borders with unvarying sizes(a) and comparison of landscape illustrations from a distant point (a) illustration made via ArcScene (b) illustration made via proposed system.

Finally, our program rendered more realistic illustrations of the landscape, when compared to ArcScene. We chose to create 2D textures of thematic data and wrap it on a more detailed 3D terrain, that's why height changes are smoother. Via making use of light, 3D effects are highlighted.

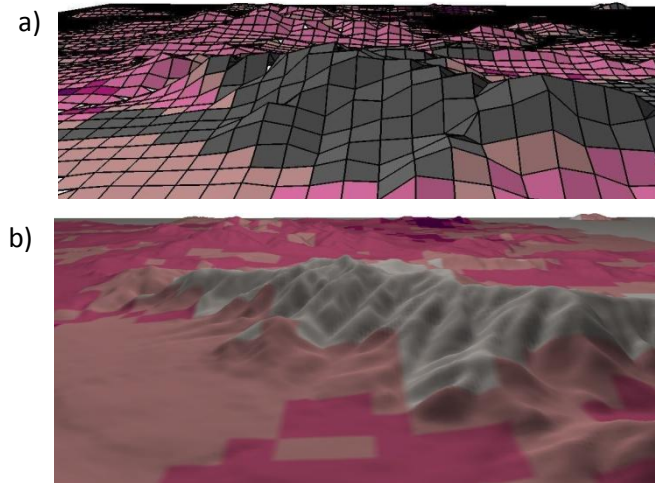


Figure 4.13. Visualization of geographical data with ArcScene(a) and with the proposed system (b).

## **5. RESULTS, DISCUSSIONS and FURTHER STUDY**

### **5.1. Results**

Visualization of geographical data is vital for conveying the information patterns underlying. Since geographical databases consist of great amount of data interrelated with geography, without attaching these data to some spatial references, it is impossible for human mind to pick the essence of information among a bunch of meaningless numerical data.

Creation of models is what our minds do to deal with the tremendous amount of stimulus coming from external world. Similar to our brain's stimulus filtering process for focusing us on a task, geographical visualization system presented deals with the statistical data and converts the massive numerical input into one consistent representation.

The proposed system produces 3D environments in which up to two data sets are intuitively embedded into, in real time. As the user navigates through the scene, she has the opportunity to observe data in different scales.

Introduction of automated details on demand to thematic maps is one of the significant contributions of this work. The details are automatically visualized when the viewpoint gets closer to the terrain. Real time modification of produced thematic map enables great user interactivity when combined with 3D the navigation. Abstraction of data, due to the distance between observer and data prevents visualization of unreadable details and clusters the data through larger unit representation regions (lower resolutions). Both individual data and the overall pattern of information are depicted with the introduction of level of detail technique to 3D thematic maps.

For investigating and developing hypothesis from the input data, user has the freedom to visualize her dataset in three different distribution models. Representations may be generated with classed or unclassed data. Two different interval determining methods are implemented, which is a fine tool for understanding the divisions of data over number line.

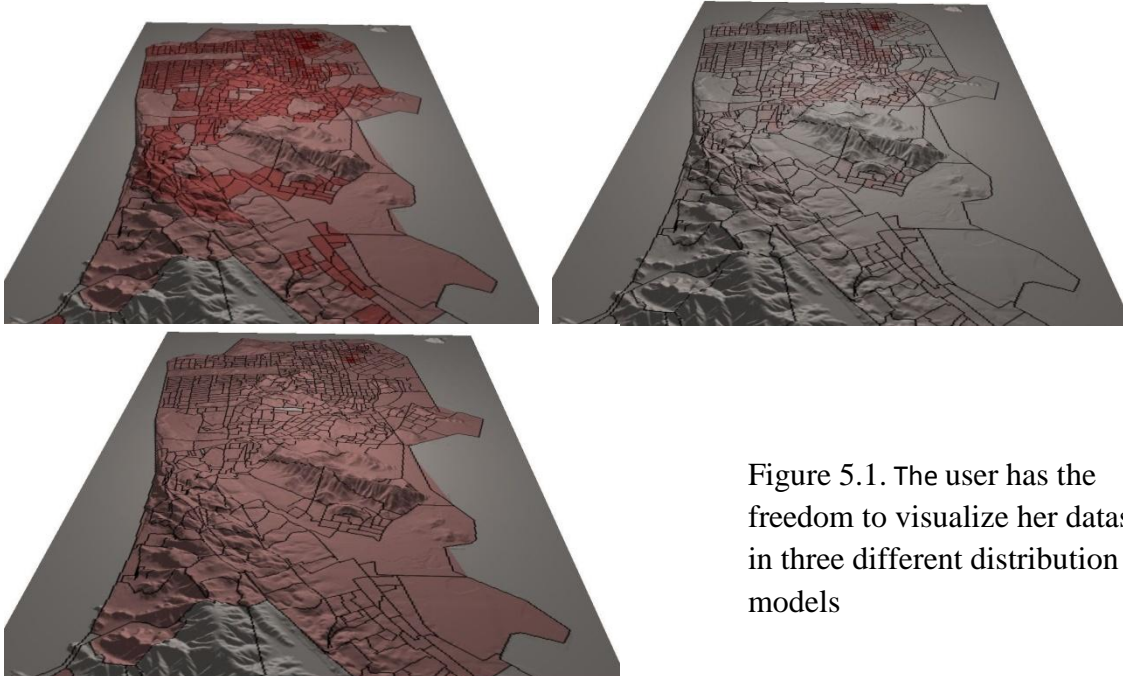


Figure 5.1. The user has the freedom to visualize her dataset in three different distribution models

## 5.2. Discussions and Further Study

There are some issues in this system which needs further study, in order to improve this project into a more flexible and user friendly application.

Some data have to be input by the user, which in the future can be decided by the system itself, such as interval limits of data. Even the most suitable statistical classification method for input data, can be determined via some intelligent algorithms, by the computer itself. Decreasing the manual work done by implementing more elaborate analytic tools will increase the usability of software.

Although our system has a consistent visual language of its own, it isn't efficient for visualizing more than two data sets, which are represented as separate layers. Increasing the displayed data layers will be the subject to be investigated next, in the near future.

While creating visualizations that appeal to the eye, letting the user choose their own color set and generating their texture patterns instead of hatching will help the diversity of depictions produced. Thus a valid user interface should be designed with additional options.

Last but not least; even though this system is found to be “good-looking” and easy to understand by users, with the intention of finding out its success about being informative and naming the strong and weak sides, a complete usability study is required.

## REFERENCES

- [1] Yücel, C. , "Herşey Sende Gizli", İstanbul.
- [2] D. A. Norman, "Things That Make Us Smart: Defending Human Attributes in the Age of the Machine", 1993.
- [3] J. Bertin, "La Graphique et le Traitement Graphique de L'Information.(English version: Graphics and Graphic Information Processing)", Translated by Walter de Gruyter, Berlin, pp. 24-31 , Flammarion, Paris, 1981.
- [4] T. A. DeFanti, M. D. Brown, and B. H. McCormick, ."Visualization-Expanding Scientific and Engineering Research Opporunities", IEEE Computer, vol. 22, no.8 , pp. 12-25, 1989.
- [5] D. Cohen-Or, and Y. Levanoni, "Temporal continuity of levels of detail in delaunay triangulated terrain" Proceedings of the 7th conference on Visualization '96, ,pp. 37–42, San Francisco, California, Oct. 1996.
- [6] F. Losasso, and H. Hoppe, "Geometry Clipmaps: terrain rendering using nested regular grids", ACM Transactions on Graphics, ACM Press, vol. 23, 2004.
- [7] M. Clasen and H. C. Hege, "Terrain Rendering using Spherical Clipmaps" , EuroVis 2006 – Proc. Eurographics / IEEE VGTC Symposium on Visualization, pp. 91-98, 2006.
- [8] J. Schneider and R. Westermann, "GPU-Friendly High-Quality Terrain Rendering," Journal of WSCG, vol. 14, pp. 49-56, 2006.
- [9] A. Asirvatham and H. Hoppe, "Terrain rendering using GPU-based geometry clipmaps," GPU Gems, vol. 2, pp. 109–122, 2005.
- [10] A. Brodersen, "Real-time visualization of large textured terrains," in Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia Dunedin, ACM Press, New Zeland, 2005.
- [11] Lindon Lab. , Second Life., Availabla from : [secondlife.com/](http://secondlife.com/), (Accessed 20.07.2008)
- [12] Google Inc., Google Earth: Explore, Search and Discover [online], Available from: [earth.google.com/](http://earth.google.com/) (Accessed 2007-08-20).
- [13] Google Inc., Google Maps, Available from: [maps.google.com/](http://maps.google.com/) (Accessed 20.07.2008).
- [14] E. Imhof, "Cartographic Relief Presentation", De Gruyter, Berlin, 1982.



- [15] C. L. Miller, R. A. Laflamme, "The Digital Terrain Model- Theory and Application. Photogrammetric Engineering", 1958.
- [16] S. K. Card, J. D. Mackinlay, B. Shneiderman, "Readings in Information Visualization Using Vision to Think", Morgan Kaufmann, 1999 .
- [17] Musicoverly, Available from: [www.musicoverly.com](http://www.musicoverly.com) , retrieved 03-07-2008.
- [18] New York Times: Faces of the Dead, Available from: [www.nytimes.com/ref/us/20061228\\_3000FACES\\_TAB1.html](http://www.nytimes.com/ref/us/20061228_3000FACES_TAB1.html), retrieved 03.07.2008.
- [19] Tracking the Threat, Available from: [www.trackingthethreat.com/flash/nav.jsp](http://www.trackingthethreat.com/flash/nav.jsp), retrieved 03.07.2008.
- [20] F. B. Viégas, M. Wattenberg, F. Ham, J. Kriss, M. McKeon, "Many Eyes: A Site for Visualization at Internet Scale", IEEE Transactions on Visualization and Computer Graphics, vol. 13, No. 6, 2007.
- [21] D. Stephens, "Making Sense of Evidence series on History Matters: The U.S. Survey on the Web".
- [23] R. Fawcett-Tang, W. Owen, "Mapping – an illustrative guide to graphic navigational systems", Rockport Publishers, 2002.
- [23] T. A. Slocum, R. B. McMaster, F. C. Kessler, H. H. Howard, "Thematic Cartography and Geographic Visualization", Pearson/PrenticeHall, 2004.
- [24] E. R. Tufte, "Visual and Statistical Thinking: Displays of Evidence for Making Decisions", Graphics Press, 1997.
- [25] 2008 Accuweather, Inc. Available from: [www.accuweather.com](http://www.accuweather.com), retrieved: 23.07.08.
- [26] B. McManus, "Thematic Mapping", GIS Resources Document 05-70, 2005.
- [27] Human Computer Interaction Lab , University of Maryland, Available from: [www.cs.umd.edu/hcil/research/visualization.shtml](http://www.cs.umd.edu/hcil/research/visualization.shtml), retrieved 20.07.2008.
- [28] R. Chang, G. Wessel, R. Kosara, E. Sauda, W. Ribarsky, "Legible Cities: Focus-Dependent Multi-Resolution Visualization of Urban Relationships", IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 6, 2007.
- [29] D. A. Keim , S. C. North, C. Panse, J. Schneidewind, "Visualizing geographic information: VisualPoints vs CartoDraw", Information Visualization, Palgrave Macmillan, vol. 2, no. 1, pp. 58-67(10) March 2003 .

- [30] D. A. Keim, S. C. North, C. Panse, J. Schneidewind, "Efficient cartogram generation: a comparison", Proceedings of the IEEE Information Visualization 2002, pp. 33-36, Boston, Massachusetts, 2002.
- [31] D. A. Keim, S. C. North, C. Panse, "Cartodraw: a fast algorithm for generating contiguous cartograms", IEEE TVCG, 2003.
- [32] S. Florisson, M. Kreveld, B. Speckmann, "Rectangular Cartograms: Construction & Animation", Proceedings of the twenty-first annual symposium on Computational geometry, Pisa, Italy, 2005.
- [33] J. Olson., "Noncontiguous area cartograms", The Professional Geographer, no. 28, pp.371–380, 1976.
- [34] E. Raisz, "The rectangular statistical cartogram", Geographical Review, no. 24, pp. 292–296, 1934.
- [35] M. van Kreveld and B. Speckmann, "On rectangular cartograms", In Proc. 12th Europ. Symp. Algorithms, pp. 724–735, Springer, 2004.
- [36] K. Buchin, M. C. Sousa, J. Döllner, F. Samavati, M. Walther, "Illustrating Terrains using Direction of Slope and Lighting", 4th ICA Mountain Cartography Workshop, Vall de Nuria, Spain, 2004.
- [39] London Profiler, Accessed from: [www.londonprofiler.org](http://www.londonprofiler.org) , retrieved 20.07.2008.
- [37] M. Gibin, A. Singleton, R. Milton, P. Mateos, P. Longley, "An Exploratory Cartographic Visualisation of London through the Google Maps API", Applied Spatial Analysis and Policy, vol. 1, no. 2, Springer, 2008.
- [38] ESRI ShapeFile, Available from: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, retrieved 20.08.2008.
- [40] J. Dykes and J. Wood "Visualization with Google Earth and GIS", tutorial from InfoVis 2007, Sacramento, California, 2007.
- [41] J. Döllner, K. Hinrichs, "Dynamic 3D Maps and Their Texture-Based Design", Proceedings of Computer Graphics International, 2000.
- [42] K. Baumann, J. Döllner, K. Hinrichs, "Integrated Multiresolution Geometry and Texture Models for Terrain Visualization", Joint Eurographics-IEEE TCVG Symposium on Visualization, 2000.
- [43] The Free Encyclopedia, Available from: [www.wikipedia.org/](http://www.wikipedia.org/), retrieved 20.07.2008.

- [47] Engineering Statistics Handbook, Available from :  
[www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm](http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm), retrieved 20.07.2008.
- [44] P. Bourke, 1987, Available from:  
[local.wasp.uwa.edu.au/~pbourke/geometry/insidepoly/](http://local.wasp.uwa.edu.au/~pbourke/geometry/insidepoly/) , retrieved 20.07.2008.
- [45] R. Mach, P. Petschek, "Visualization of Digital Terrain and Landscape Data - A Manual" , Berlin, Springer, 2006.
- [46] C. A. Brewer," ColorBrewer - Selecting Good Color Schemes for Maps [online]", 2002. Available from: [www.personal.psu.edu/cab38/ColorBrewer/ColorBrewer.html](http://www.personal.psu.edu/cab38/ColorBrewer/ColorBrewer.html), retrieved Accessed 20.07.2008.
- [48] C. Chen, "Information Visualization : Beyond the horizon", Springer, 2002.
- [49] F. S. Hill, "Computer Graphics using OpenGL" , pp. 292, Prentice Hall, 2001.
- [50] U.S. Census Bureau, Available from: [www.usgs.gov/](http://www.usgs.gov/) retrieved 20.08.2008.
- [51] ESRI' s ArcGIS, Available from: [www.esri.com/products.html](http://www.esri.com/products.html), retrieved 20.08.2008.
- [52] ESRI Data & Maps library
- [53] List of buildings in San Francisco, Available from:  
[http://en.wikipedia.org/wiki/List\\_of\\_tallest\\_buildings\\_in\\_San\\_Francisco](http://en.wikipedia.org/wiki/List_of_tallest_buildings_in_San_Francisco), retrieved 20.07.2008.
- [54] ESRI' s ArcScene, 3D display environment Available from:  
[http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=ArcScene\\_3D\\_display\\_environment](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=ArcScene_3D_display_environment), retrieved 20.08.2008.