PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Samuel W. Bloomquist

Entitled
WEB-BASED GEOTEMPORAL VISUALIZATION OF HEALTHCARE DATA

For the degree of Master of Science

Is approved by the final examining committee:

Shiaofen Fang

Mihran Tuceryan

Yuni Xia

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Shiaofen Fang

Approved by: Shiaofen Fang 10/09/2014

Head of the Graduate Program Date

WEB-BASED GEOTEMPORAL VISUALIZATION OF HEALTHCARE DATA

A Thesis

Submitted to the Faculty

of

Purdue University

by

Samuel W. Bloomquist

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

December 2014

Purdue University

Indianapolis, Indiana

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my advisor Dr. Shiaofen Fang for his support of my MS research. His lectures in the classroom were the spark that inspired me to pursue research in the field of information visualization. He has continuously pushed me to refine and improve upon the techniques covered in this thesis, and I am grateful for his leadership, inspiration, and extensive knowledge.

In addition to my advisor, I would like to thank the other members of my research committee for their encouragement and collaboration: Dr. Yuni Xia, Dr. Mathew Palakal, Jeremy Keiper, Anand Krishnan, Thanh Minh Nguyen, Weizhi Li, Shenghui Jiang, and Dr. Shaun Grannis.

Finally, for their endless patience and encouragement of my academic pursuits in the midst of professional and family life, I would like to thank my wife Shannon Bloomquist and our daughter Henrietta.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Bloomquist, Samuel W. M.S., Purdue University, December 2014. Web-based Geotemporal Visualization of Healthcare Data. Major Professor: Shiaofen Fang.

Healthcare data visualization presents challenges due to its non-standard organizational structure and disparate record formats. Epidemiologists and clinicians currently lack the tools to discern patterns in large-scale data that would reveal valuable healthcare information at the granular level of individual patients and populations. Integrating geospatial and temporal healthcare data within a common visual context provides a twofold benefit: it allows clinicians to synthesize large-scale healthcare data to provide a context for local patient care decisions, and it better informs epidemiologists in making public health recommendations.

Advanced implementations of the Scalable Vector Graphic (SVG), HyperText Markup Language version 5 (HTML5), and Cascading Style Sheets version 3 (CSS3) specifications in the latest versions of most major Web browsers brought hardware-accelerated graphics to the Web and opened the door for more intricate and interactive visualization techniques than have previously been possible. We developed a series of new geotemporal visualization techniques under a general healthcare data visualization framework in order to provide a real-time dashboard for analysis and exploration of complex healthcare data. This visualization framework, HealthTerrain, is a concept space constructed using text and data mining techniques, extracted concepts, and attributes associated with geographical locations.

HealthTerrain's association graph serves two purposes. First, it is a powerful interactive visualization of the relationships among concept terms, allowing users to explore the concept space, discover correlations, and generate novel hypotheses. Second, it functions as a user interface, allowing selection of concept terms for further visual analysis.

In addition to the association graph, concept terms can be compared across time and location using several new visualization techniques. A spatial-temporal choropleth map projection embeds rich textures to generate an integrated, two-dimensional visualization. Its key feature is a new offset contour method to visualize multidimensional and time-series data associated with different geographical regions. Additionally, a ring graph reveals patterns at the fine granularity of patient occurrences using a new radial coordinate-based time-series visualization technique.

CHAPTER 1. INTRODUCTION

Personal health records, point of care testing systems, and other ancillary digital devices have the potential to generate massive amounts of digital health-related information. Consequently, clinical providers and health care administrators must discern meaningful patterns from increasingly large datasets to make informed clinical decisions for individual patients and develop broad strategies for patient populations.[1] Variations in data types and the diversity of data consumers increase the complexity of the challenge. Health information must be leveraged to support timely decision-making and enable effective trend/pattern detection; this requires increasingly innovative and novel methods for visualizing healthcare data using pragmatic and easily understood frameworks. Several challenges must be addressed to achieve this goal:

1) Health data is a data-rich but information-poor domain. In Electronic Health Record (EHR) systems, data are almost always heterogeneous, unstructured, hierarchical, and longitudinal.
2) EHR systems are often extremely large. While it is possible to visualize an EHR system on small scales and with a focused scope, high impact knowledge discoveries more likely come from global scale (population-wide) visualization and knowledge mining.

3) Visualizing population-level health data often involves the difficult task presenting geospatial data that changes over time in a common visual context without using animation.

To overcome these challenges, feature extraction of healthcare data through mining provides critical organization and structure to what is currently a disconnected, freeform body of healthcare data. This feature space often consists of healthcare terms (ontology) and their relationships. Therefore, the effective integration of data processing, data mining, and text mining is necessary in healthcare data visualization. Although the scope of healthcare data is vast, the visualization of aggregated features combined with patient-level visualization can be very effective in revealing the patterns and trends of population health. Developing multiple visualization methods that show different perspectives of the data is a necessary means to that end.

HealthTerrain, a prototype visualization system, was created to visualize aggregate features and patient-level healthcare information by overcoming the challenges presented by synthesizing unstructured, large, geotemporal healthcare data sets. Developed through a collaborative research initiative between the Indiana University – Purdue University, Indianapolis (IUPUI) Department of Computer Science, the IUPUI School of Informatics and Computing, and the Regenstrief Institute, HealthTerrain makes available new geotemporal visualization techniques under a general healthcare data visualization framework in order to provide a real-time dashboard for analysis and exploration of complex healthcare data.

CHAPTER 2. OVERVIEW

2.1 Related Work

Prior to the research for HealthTerrain, health data visualization of large-scale datasets had not been extensively studied. In most existing works and visualization systems, the use of EHR data is a secondary concern or the scope is very limited.

LifeLines[2] uses a traditional 2D timeline visualization technique to visualize specific patient medical and health history. It emphasizes the visualization of temporal ordering of events, but the scope is limited to individual patient history rather than effects and patterns in aggregate data. LifeLines2,[3] an extension of LifeLines, enables multiple patient comparisons for analysis; however, the individualized focus of LifeLines2's visualization design limits its scalability. A similar system called TimeLine[4] re-organizes and re-groups multiple EHR content types in a Y-axis layout to track multiple events along the same timeline. As with LifeLines2, Timeline is focused on a series of events pertaining to a handful of individual patients rather than on aggregate patterns from immense collections of healthcare data. "A Novel Approach to Viewing Medical Data"[5] describes another set of tools for visualizing a patient’s EHR to aid physician diagnosis and decision-making; however, its visualization techniques are limited to parallel coordinates and a traditional matrix view, and its scope is restricted to analysis of individual medical conditions. CLEF[6] is a system enabling visual navigation through a patient’s medical record using semantically and temporally organized networks to represent medical history events. CLEF supports limited text processing capabilities for generating individual patient report summaries; however, the text processing does not drive the visualization techniques, and advanced mining algorithms are not implemented. Similarly to LifeLines, LifeLines2, and Timeline, CLEF focuses on individual patient timelines as opposed to aggregate data patterns and hypotheses.

A literature review of articles focusing on visualization tools for infectious diseases is given in "Visualization and Analytics Tools for Infectious Disease Epidemiology."[7] Unlike the tools described previously, many of the tools featured in this literature review focus on visualizing aggregate and geospatial healthcare data; those tools that make use of geospatial data, however, use standard dot map, choropleth map, and isopleth map techniques. Novel geospatial and geotemporal visualization techniques are not cited.

Visual analytics systems making use of both geographic and temporal data for syndromic surveillance activities were discussed in "Reviewing and Managing Syndromic Surveillance SaTScan Datasets Using an Open Source Data Visualization Tool"[8] and "Understanding Syndromic Hotspots."[9] The resulting visualizations split images into map-based views for geographic data and linear graphs for visualizing time series data rather than the unified geotemporal images present in HealthTerrain.

Many techniques exist for geospatial visualization of time series data including time color-coding,[12] connected timelines,[13] and time curves,[14] but they often result in visual clutter and occlusion. A well-known technique is space-time-cube. [15, 16, 17] Space-time-cube integrates time and spatial information in a 3D visual representation; unfortunately, the visual representation provided is flawed because the geospatial map is only given at time zero. This causes the sense of space-time embedding to diminish as the data moves along the time axis; visual clutter also becomes an issue with large datasets.

2.2 Objective

HealthTerrain is a web-based interactive healthcare data visualization system created to integrate geospatial and temporal healthcare data within a common visual context. HealthTerrain's concept space approach combines data mining, data processing, and text mining techniques to effectively transform heterogeneous, unstructured, hierarchical, and longitudinal healthcare data to a uniform space of controlled ontology and associations. Multiple toolkits with data filters visualize data on different scales, providing both fine- and coarse-grained visual analytics tools.

One of the unique challenges in healthcare data visualization is visualizing time-series data with associated geospatial information. Adding an additional time dimension is very difficult to implement in a geospatial context without employing animated visualizations, which do not allow for the kind of in-depth user analysis that static images do. HealthTerrain overcomes this challenge by breaking up a large geospatial representation (e.g. the state of Indiana) into smaller representations (e.g. counties or zip codes) and embedding the time variable within the smaller representations. On this smaller scale, time variable representation replaces spatial resolution while the larger representation maintains spatial significance. This strategy can also be useful for multidimensional data visualization within a geospatial context.

2.3 Dataset Source

The HealthTerrain visualization system uses a public health disease reporting system from the Regenstrief Institute as its prototype dataset. The Institute implemented and continues to maintain an unparalleled Health Information Exchange (HIE) along with an HIE-based automated electronic lab reporting (ELR) and case-notification system for over ten years in the State of Indiana. Its Notifiable Condition Detector (NCD) system uses a standards-based messaging and vocabulary infrastructure that includes Health Level Seven (HL7) and Logical Observation Identifier Names and Codes (LOINC).[18] The NCD receives real-time HL7 version 2 clinical transactions daily, including diagnoses and laboratory studies as well as transcriptions from hospitals, national labs, and local ancillary service organizations.[19] The system automatically detects positive cases of notifiable conditions and forwards alerts to local and state health departments for review and follow up. These alerts enable more effective and efficient public health population health monitoring and case management.

CHAPTER 3. METHODS

3.1 System Design

HealthTerrain is a browser-based web application using leading-edge graphics capabilities implemented in the latest versions of most major Web browsers, including Chrome, Firefox and Safari. The core technology is WebGL in an HTML5 canvas. The system's architecture pattern employs the Ruby on Rails (RoR) framework for delivering Web applications with AJAX (Advanced Javascript and XML) services and a classic Model-View-Controller architecture.

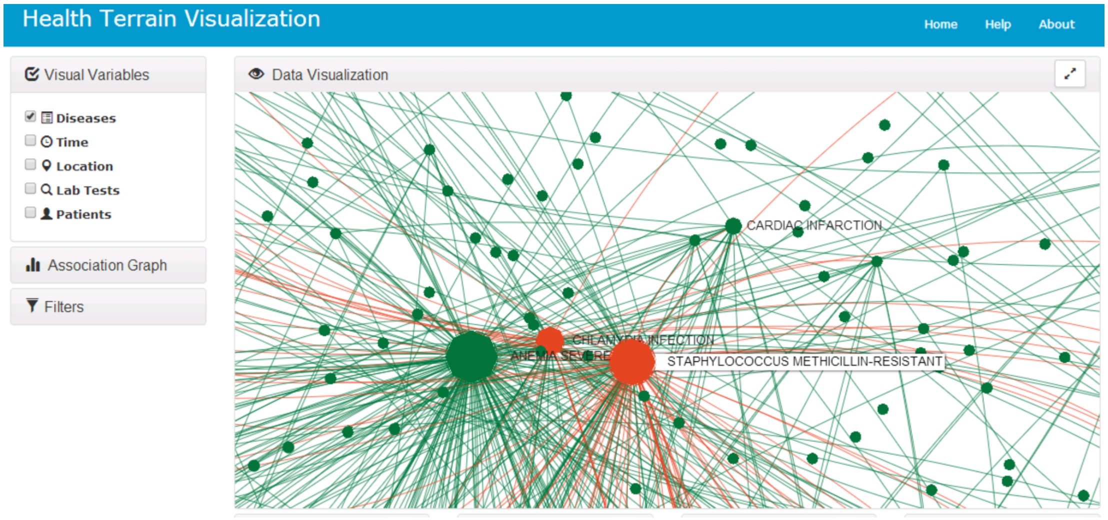
The user interface is a modern Web application built on a combination of traditional HTML form submission and REST-ful service calls to query and retrieve data from a MySQL relational database in various data delivery formats such as Extensible Markup Language (XML) and JavaScript Object Notation (JSON). The visualizations use HTML, CSS, SVG, and WebGL technologies with a number of open source Javascript libraries such as sigma.js, d3.js, jquery.js and three.js for drawing, displaying and interacting with data and graphics. The interactive user interface is designed to support data exploration and hypothesis generation based on concept associations.

The process is as follows:

1) The user chooses concept terms that apply to the visualization objectives from the association map (Figure 2.1.1). Various data filters such as time, gender, race, and age are then selected using the filter interface (Figure 2.1.2).

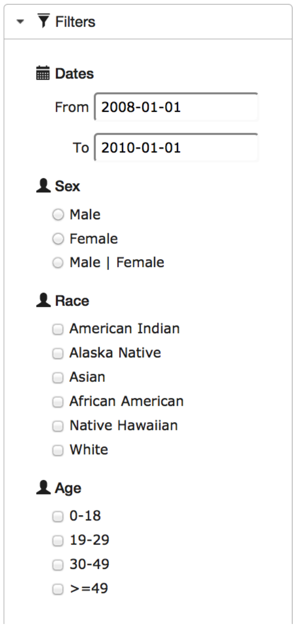
2) Based on the categorization and combination of the concept terms, the appropriate tools will be activated to visualize the filtered dataset. To achieve various visualization effects, the user can apply interactive operations such as zooming, mouseover, and picking. For example, if a user wishes to apply multiscale visualization, he or she simply zooms in on a geospatial region and switches the visualization from a county-based visual representation to a zip code-based visual representation.

3) After exploring a visualization method, the user may want to generate another visualization for comparison. The current visualization in the display window can be minimized to the sidebar column of the primary system window (Figure 2.1.3), which can later be activated again as needed.



☑ Diseases
□ Time
□ Location
□ Lab Tests
□ Patients

Figure 2.1.1 Association Map - Disease Selection



Filters

Dates

From 2008-01-01

To 2010-01-01

Sex

○ Male
○ Female
○ Male | Female

Race

□ American Indian
□ Alaska Native
□ Asian
□ African American
□ Native Hawaiian
□ White

Age

□ 0-18
□ 19-29
□ 30-49
□ >=49

Figure 2.1.2 Visualization Filters

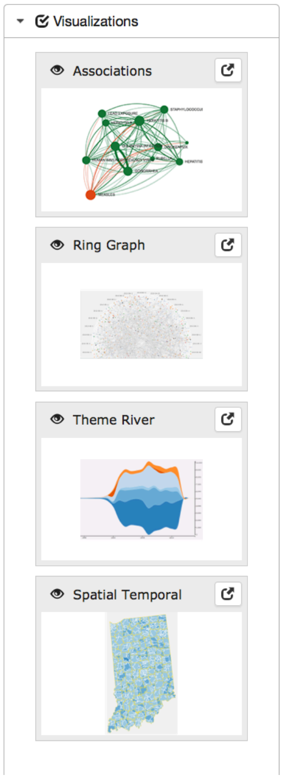


Figure 2.1.3 Visualization Sidebar

3.2 Data and Text Mining

While a detailed description of the algorithms used to mine data into the concept space is irrelevant to this thesis, a high-level explanation of the results of that process provides necessary background information on the visualizations we developed. Mining processes were used to identify diseases, symptoms, mental behaviors, risky behaviors, and medication terms from 325,791 textual patient visit reports. The total number of terms extracted for each category is given in Table 2.2.1.

Table 3.2.1 Total Concept Space Terms by Category

Term Type	**Number of terms extracted**
Diseases	7988
Symptoms	10803
Mental Behavior	712
Risky Behavior	244
Medications	5721

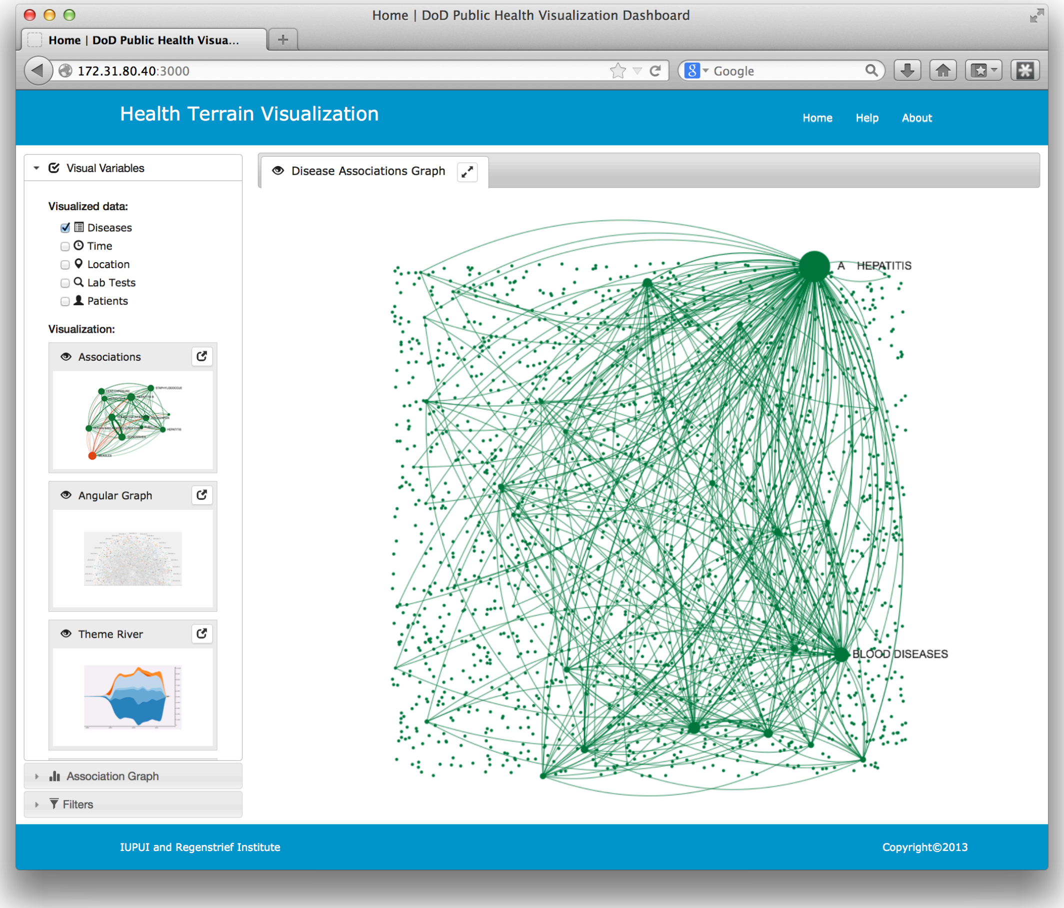
Further mining and analysis of the data identified comorbidity of the diseases across the 325,791 reports. This was accomplished by computing the pair-wise significance of each disease with all the corresponding conditions (i.e., the symptoms, mental behavior, risky behavior, and medications). Table 2.2.2 shows the comorbidity of conditions relating to the top 10 most occurring diseases.

Table 3.2.2 Comorbid Conditions With Top 10 Most Occurring Diseases

	Diseases	Symptoms	Mental Behavior	Risky Behavior	Medications
hypertension	diabetes, renal disease, pulmonary hypertension, artery disease,	chest pain, nausea, vomiting, dyspnea, abdominal pain, weakness,	abuse, depression, dementia, anxiety, altered mental status, drug use,	smoking, tobacco use, smokes, compliance, impression, drinking, lying,	insulin, hepatitis, tobacco, oxygen, glucose, lasix,
diabetes	diabetes mellitus, hypertension, type 2, artery disease, renal disease,	nausea, vomiting, chest pain, abdominal pain, diarrhea,	abuse, depression, altered mental status, drug use,	smoking, compliance, tobacco use, compliant, impression, drinking,	insulin, glucose, tobacco, hepatitis, humulin,
pneumonia	lower lobe pneumonia, aspiration, aspiration pneumonia, copd,	shortness of breath, chest pain, dyspnea, chills, vomiting,	abuse, dementia, aggressive, confusion, altered mental status,	smoking, impression, drinking, smokes, tobacco, compliant, compliance,	oxygen, avelox, albuterol, prednisone, levaquin,
hepatitis	hepatitis c, hepatitis b, cirrhosis, liver disease, encephalopathy,	nausea, abdominal pain, vomiting, diarrhea, chills,	abuse, dependence, confusion, drug use, opiate, depression,	smoking, drinking, smokes, tobacco use, illicit drug use,	hepatitis, hepatitis b, prograf, lactulose, ammonia, antibody,
svd	gbs, pcc, ofc, strep, hep, external genitalia,	prn pain, constipation, cramping, headache, breakthrough pain,	abuse, drug use, depression, substance abuse,	smokes, illicit drug use, smoking, tobacco use,	micronor, vitamin, antibody, ibuprofen, stool softener,
anemia	renal diabetes, hypertension, renal disease, hepatitis, heart failure,	nausea, abdominal pain, vomiting, chest pain, fatigue, weakness,	abuse, depression, anxiety, dementia, confusion, altered mental status,	smoking, drinking, impression, tobacco use, compliance,	iron, vitamin, hepatitis, coumadin, oxygen, prednisone,
renal disease	end-stage renal disease, end stage renal disease, diabetes, hypertension, artery disease	nausea, vomiting, chest pain, abdominal pain, chills, shortness of breath,	altered mental status, abuse, confusion, dementia, depression, confused	smoking, compliance, impression, tobacco use, illicit drug use, drinking	calcium, insulin, glucose, coumadin, hepatitis, bicarbonate,
asthma	pneumonia, diabetes, copd, hypertension, airway disease,	wheezing, shortness of breath, wheezes, coughing, dyspnea,	abuse, depression, mdi, anxiety, drug use, aggressive,	smoking, impression, drinking, tobacco use, crying,	albuterol, prednisone, medrol, oxygen, atrovent, advair,
hiv	aids, pneumonia, hepatitis, infection, infectious disease, herpes, meningitis,	nausea, vomiting, diarrhea, abdominal pain, headache, weakness,	abuse, depression, schizophrenia, drug use, dementia, dependence,	compliance, smoking, drinking, impression, lying, tobacco use	hepatitis, bactrim, vitamin, cocaine, acetaminophen, hepatitis b,
diabetes mellitus	diabetes, hypertension, artery disease, renal disease,	vomiting, nausea, chest pain, abdominal pain, diarrhea,	abuse, depression, altered mental status, dementia,	smoking, tobacco use, compliance, illicit drug use,	insulin, glucose, humulin, tobacco, hepatitis,

3.3 Visualizing Concept Space Associations

Association map (Figure 2.3.1) is a graph visualization of the relationships among the diseases and other terms in the concept space. It serves as a platform supporting interactive selection of concepts to dynamically visualize data using a variety of system tools. Edge thickness indicates the strength of association; node size can reflect the number of other nodes to which a given node has a significant association or the total occurrence of a term (e.g. disease) in the dataset. The graph highlights related nodes and the edges that have significant associations to the selected index node.


Home | DoD Public Health Visualization Dashboard
Home | DoD Public Health Visua...
172.31.80.40:3000
Google
Health Terrain Visualization
Home
Help
About
Visual Variables
Disease Associations Graph
A HEPATITIS
BLOOD DISEASES
Associations
Angular Graph
Theme River
Association Graph
Filters
IUPUI and Regenstrief Institute
Copyright©2013

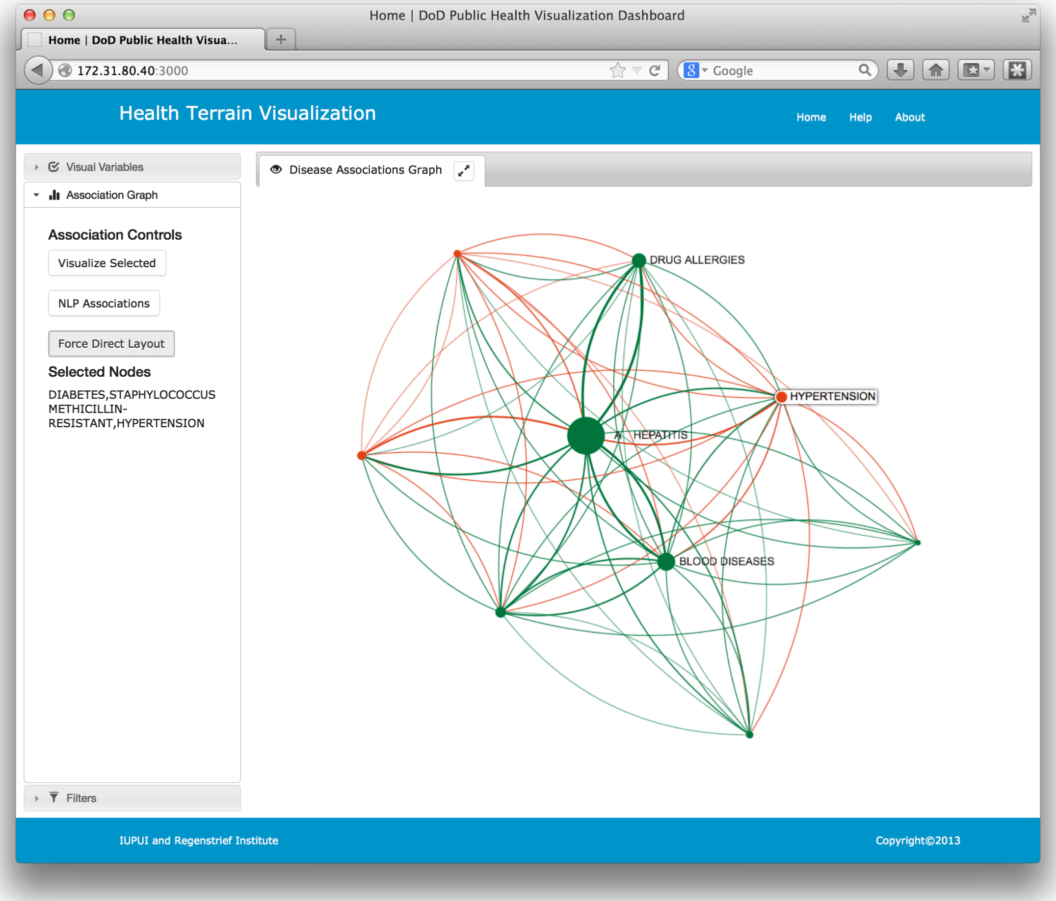
Visualized data:

- ☑ Diseases
- ☐ Time
- ☐ Location
- ☐ Lab Tests
- ☐ Patients

Visualization:

Figure 2.3.1 Association Map

After individual terms are selected in the association map, the visualization can be redrawn as shown in Figure 2.3.2 using the spring embedder[20] algorithm described in Figure 2.3.3 to display only terms highly associated with the selection and arranged into more comprehensible layouts.



Selected Nodes

DIABETES,STAPHYLOCOCCUS METHICILLIN-RESISTANT,HYPERTENSION

Figure 2.3.2 Selected Nodes in Association Map

$$E_s = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}k(d(i,j)-s(i,j))^2$$

Figure 2.3.3 Spring-embedder Algorithm

In Figure 2.3.3, *d(i,j)* represents the 2D Euclidean distance of two nodes and *s(i,j)* is a similarity metric of two nodes representing the heuristic of the layout.

A final variation on the association graph visualization shows the concept space categories (diseases, mental behaviors, risky behaviors, and symptoms) and terms related significantly to the selected diseases (Figure 2.3.4).

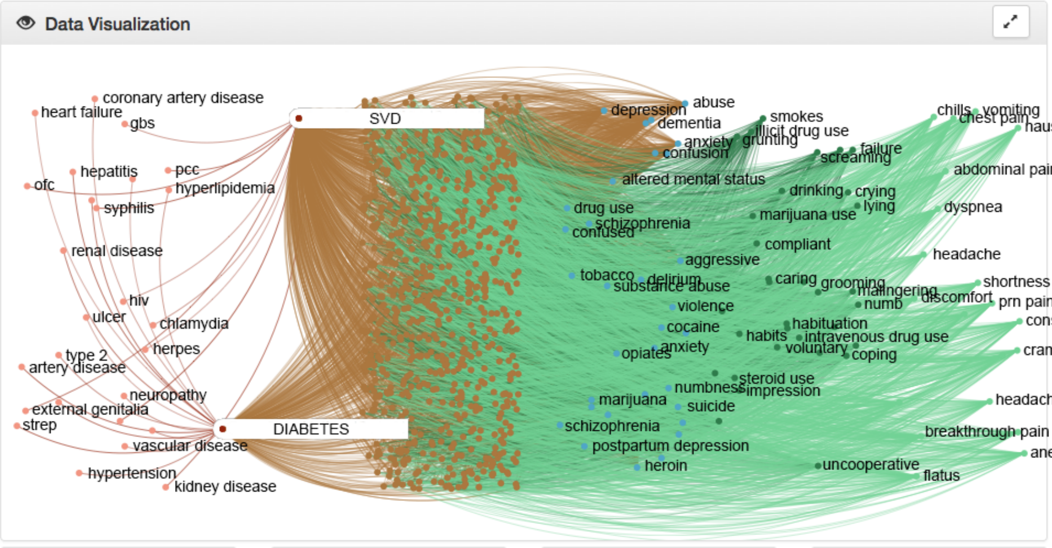
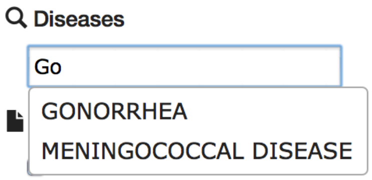


Figure 2.3.4 Association Graph of Concept Space Relationships

The large number of nodes and edges in the unfiltered association map can make it difficult for users to find specific diseases (Figure 2.3.1). A search box added to the visualization filters with autocomplete capabilities eases that process (Figure 2.3.5).



Diseases

Go

Figure 2.3.5 Disease Search Autocomplete

3.4 Choropleth Map Visualizations

In order to visualize occurrences of diseases or symptoms in a geospatial context, a simple choropleth (heat) map visualization was developed. It uses the Albers equal-area conic projection[21] to project the 3D curved surface of the Earth onto a 2D map while maintaining accurate relative surface areas of various comparable subregions (e.g. states, counties, zip codes).

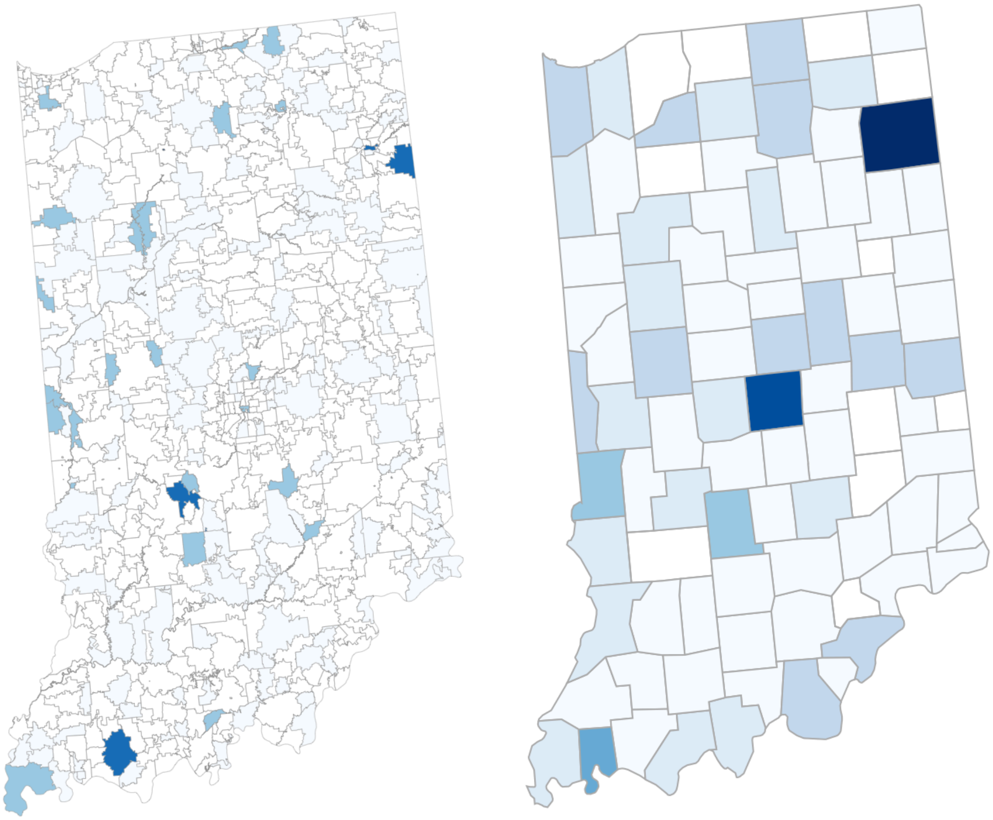


Figure 2.4.1 Simple Choropleth Map of Disease Occurrences

Standard single-attribute choropleth visualizations are useful but very limited in scope. The richness of the term associations in the mined concept space is one of the key attributes to the HealthTerrain system, making it necessary to develop a way to visualize multiple attributes for a disease (e.g. the associated diseases, symptoms, and behaviors) in each subregion on a terrain surface. Previous work in this area has focused on paint-inspired blending of colors to mimic the visual cues to which people are accustomed from early childhood.[22] This turned out to be a poor starting point; early research for the HealthTerrain system duplicated those methods in an HTML5 SVG canvas with mixed results. Figure 2.4.2 shows success at converting the red-green-blue (RGB) color standard of the World Wide Web to a red-yellow-blue (RYB) representation and then implementing a subtractive blending scheme. The middle column of colors represents a blend between the blue hues on the left and the red hues on the right. The resulting purplish tones mimic the hue that would be produced by mixing paint samples. The blended hues remain consistent across the spectrum of brightness from light to dark.

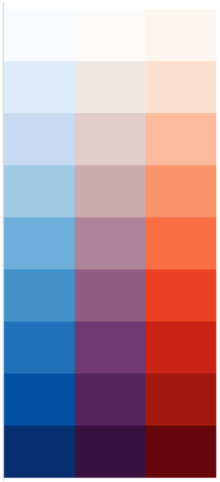


Figure 2.4.2 Paint-Inspired Color Blending

Unfortunately, this approach was unsuccessful when applied to smaller geographic regions such as zip codes and counties. When a dark hue of one color was blended with a light hue of another, viewers experienced difficulty mentally separating the disparate attribute strengths represented by the single blended color. Further attempts to refine the technique by applying paint-like translucent alpha patterns to the blend were stymied by inconsistencies and defects in Web browser implementations of the SVG specification. The HealthTerrain research team ultimately abandoned this approach to multidimensional choropleth map visualizations; however, the success achieved in mimicking paint-inspired color blending may prove useful in future work.

3.4.1 Offset Contours

The HealthTerrain research team's second attempt at solving the problem of how to visualize multiple attributes in a choropleth map resulted in the development of a new offset contour technique. In this approach, a larger geographic area is divided into smaller units such as counties and zip codes. We subdivided the interior of the unit into colored regions representing each attribute by offsetting the boundary curve toward the interior as illustrated in Figure 2.4.3. The boundary shape of the geographic unit remains intact.

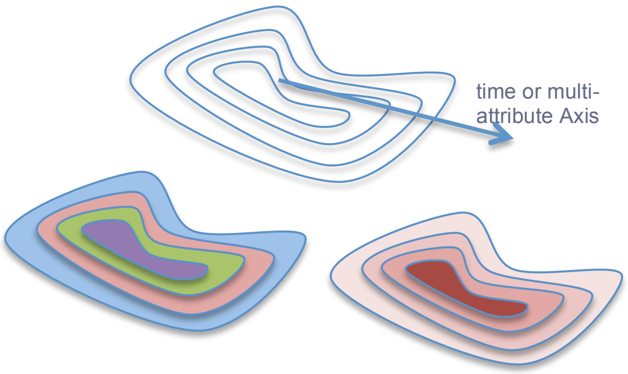
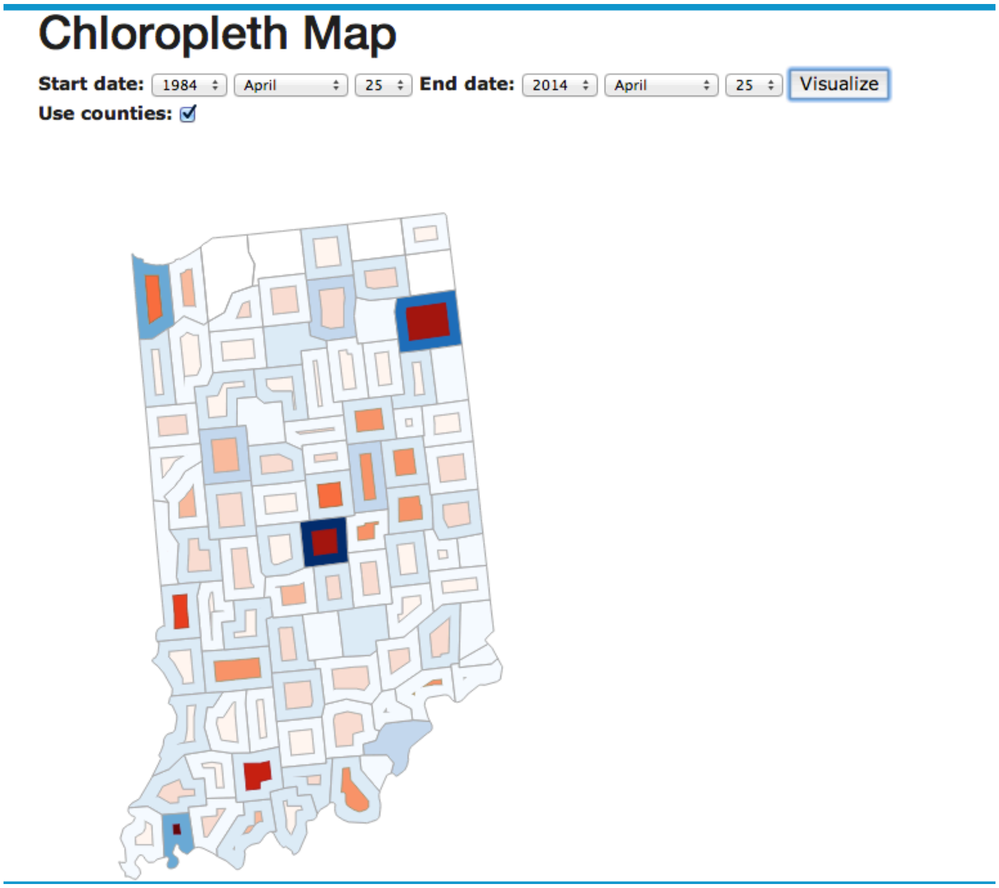


Figure 2.4.3 Offset Contour Illustration

We considered using one of several known geometric algorithms to generate offset curves; ultimately, SVG's built-in image erosion operator[23] achieved the desired effect with minimal code and efficient in-browser rendering times. Figure 2.4.4 demonstrate this new technique with two comparable attributes; Figure 2.4.5 displays four comparable attributes. In contrast to the paint-inspired color blending approach described in section 2.4, the individual hue of the color representing each attribute is easily discernable as evidenced in Figure 2.4.4.

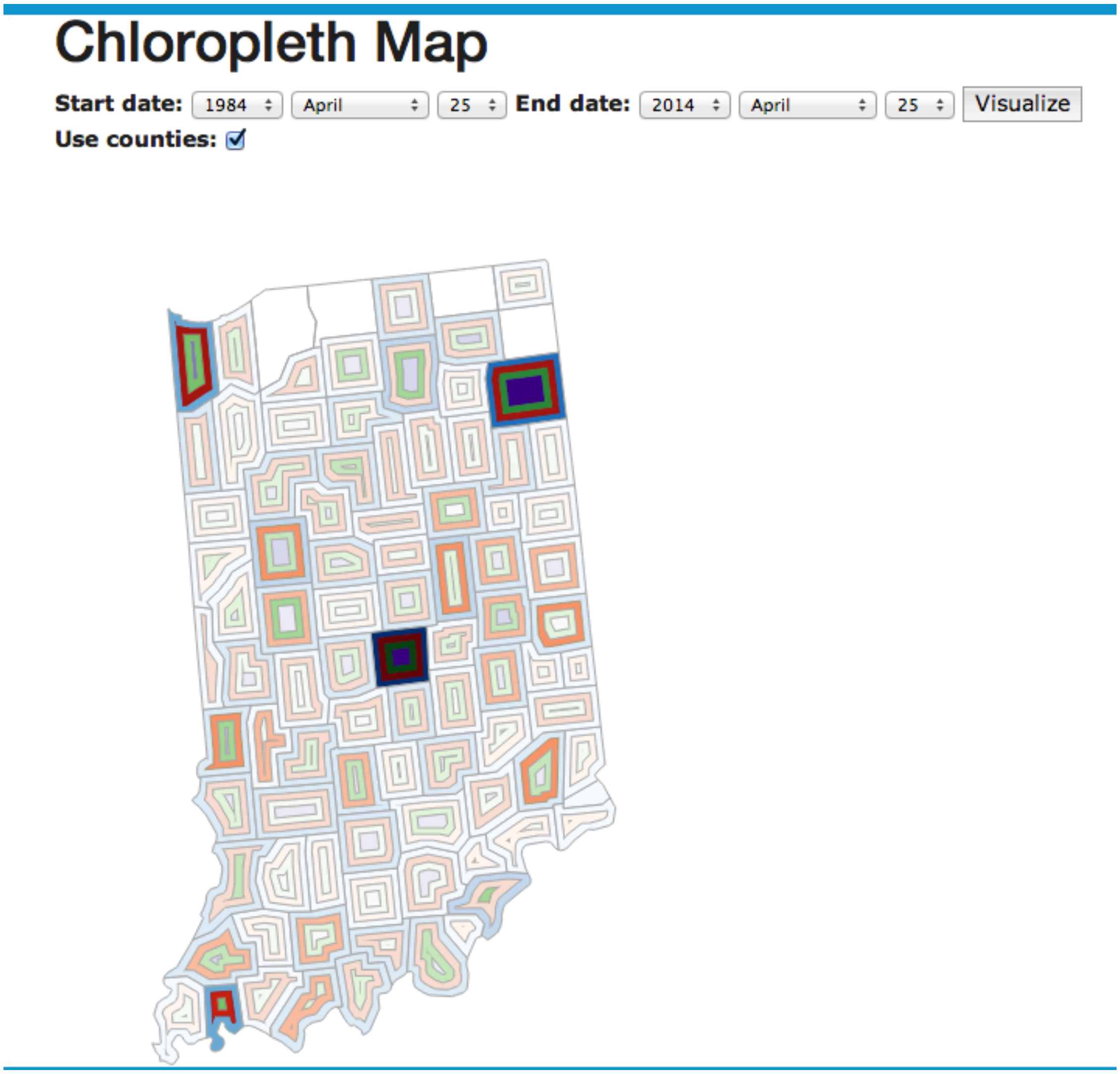


Chloropleth Map

Start date: 1984 April 25 End date: 2014 April 25

Use counties: ☑

Figure 2.4.4 Multidimensional Offset Contours Comparing Two Attributes



Chloropleth Map

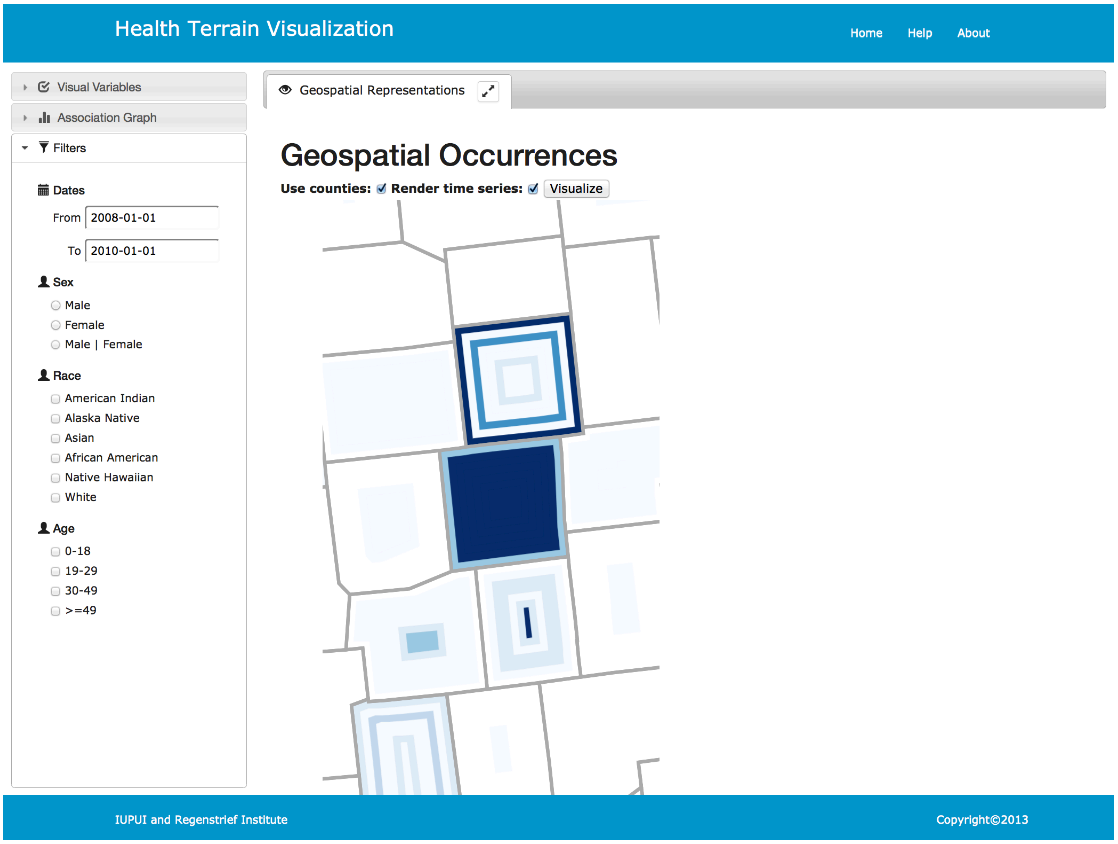
Start date: 1984 April 25 End date: 2014 April 25

Use counties: ☑

Figure 2.4.5 Multidimensional Offset Contours Comparing Four Attributes

This offset contour technique can also be applied to time-series data with similar geographic units (Figure 2.4.6). The series timeline is divided into equal time intervals represented by the offset contours. Varying hues can be used to represent the attribute changes (e.g. occurrence of a disease) over time. Figure 2.4.6 demonstrates this effect with a zoomed-in view of Lyme disease occurrences in Marion and surrounding counties from January 2008 to December

2009. This approach is particularly suitable for population-level healthcare data, which typically have attributes defined for fixed geographic units such as counties or zip codes.



Dates

From 2008-01-01

To 2010-01-01

Sex

○ Male
○ Female
○ Male | Female

Race

□ American Indian
□ Alaska Native
□ Asian
□ African American
□ Native Hawaiian
□ White

Age

□ 0-18
□ 19-29
□ 30-49
□ >=49

Use counties: ☑ Render time series: ☑ Visualize

Figure 2.4.6 Time Series Offset Contours

3.5 Ring Graph

The ring graph is a new visualization method designed specifically for patient-centric data visualization. Each patient is modeled as a point in a radial coordinate system. The radial space is subdivided into multiple rings, each of which represents one visualization term that was selected from the association map. These terms are typically disease names but can also be other associated terms such as symptoms and risky behaviors. The circumference of this radial space represents the time-axis. Thus, time is encoded as the radial angle of the points. The number of patient occurrences determines the order of the gray and white rings; the term with the least number of occurrences will occupy the inner ring, and the term with the most occurrences will occupy the outer ring. Since the inner ring has the least amount of surface area, ordering outward from minimal to maximal occurrences reduces the clutter and overlap of the individual data points and reveals intricate patterns.

The ring graph shows the distribution of patient-level data over a time-attribute space. One significant attribute, typically “age,” is represented by radius. Other patient attributes, such as race and gender, are represented by both the color and shape of the dots.

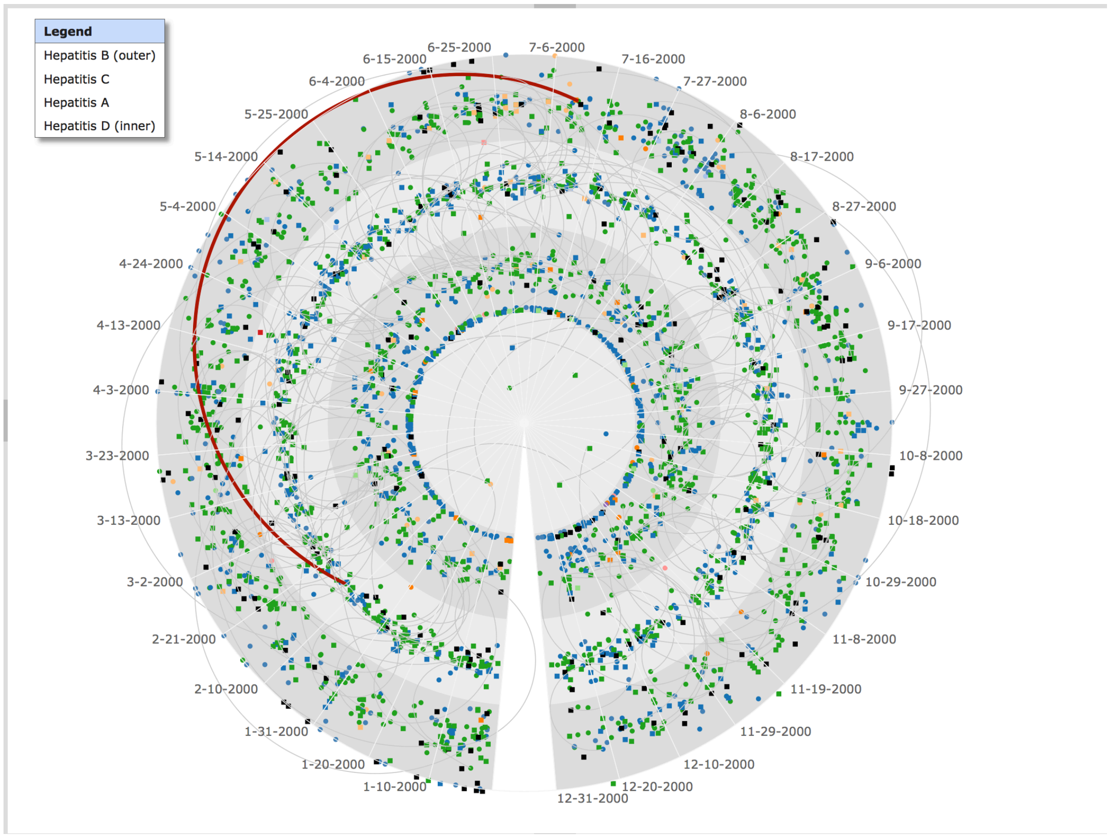
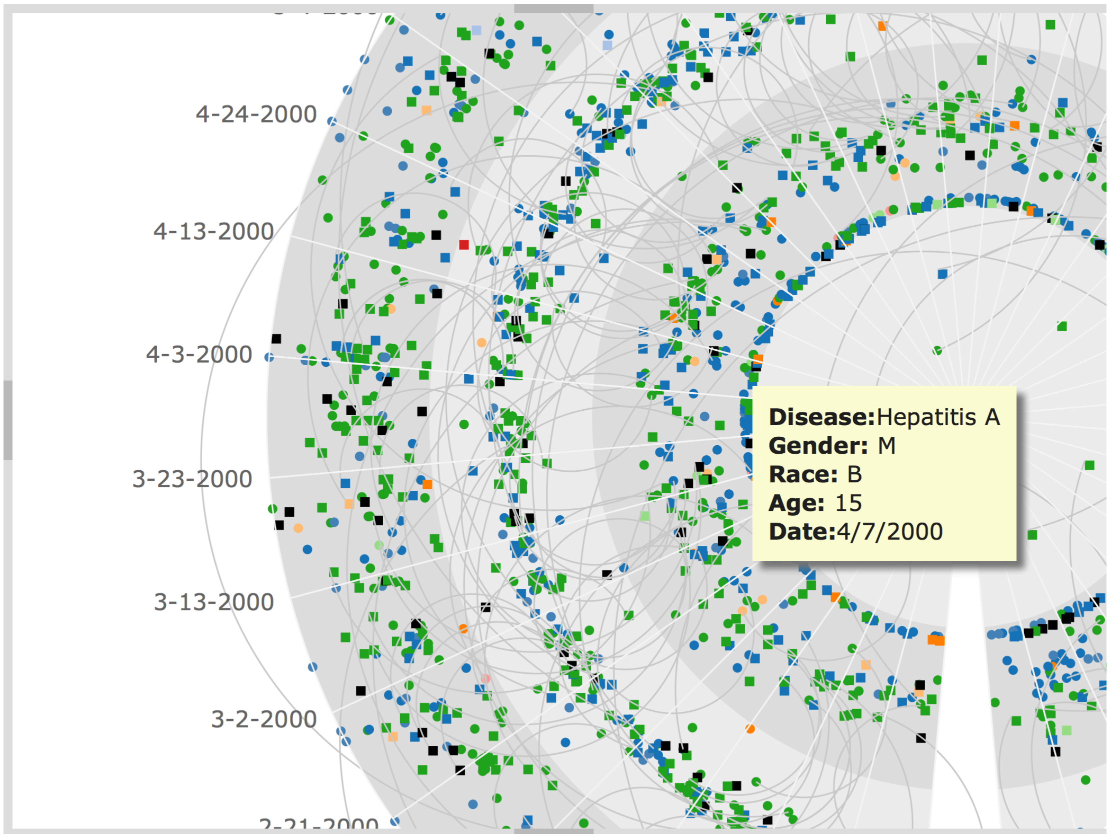


Figure 2.5.1 Ring Graph of Hepatitis A, B, C, and D

Occurrences of the same patient associated with multiple terms (e.g. diagnosed with multiple diseases) are connected with curves across the graph. A connecting curve is highlighted when users mouse over the patient dot or connecting curve. To avoid clutter, the connecting curves are drawn with adjustable semi-transparent lines. Lowering the transparency more clearly reveals the associations between terms. Hovering over the individual patient points displays patient record details as shown in Figure 2.5.2.



Disease: Hepatitis A
Gender: M
Race: B
Age: 15
Date:4/7/2000

Figure 2.5.2 Ring Graph Patient Details

CHAPTER 4. RESULTS

4.1 Association Map Results

The association map is an introductory view into the relationships between terms in the concept space that provides an exploratory interaction for selecting specific terms to use in HealthTerrain's other visualizations. An adjustable threshold value determines whether the strength of the relationship between any two terms in the concept space warrants a visible edge in the graph. If the threshold is set to a high value, only the most significant relationships will be drawn and displayed. As a user lowers the threshold, broader and more loosely related terms will be connected by edges. Users with expert knowledge can determine a threshold value that provides the desired balance between a statistically noisy graph and one with potentially valid novel connections between diseases, symptoms, behaviors, and medications.

4.2 Offset Contours Results

A time-series example of the offset contours choropleth mapping technique is shown in Figure 2.4.6. It includes all the cases of Lyme disease from January 2008 to December 2009. The time period is divided into six equal intervals. Shades of blue are used to represent different levels of the occurrence in each county. Alternating contours of white hue provide an obvious visual cue to the seasonal nature of Lyme disease occurrences since Lyme disease is most commonly transmitted by parasites in the tick family who are dormant during winter months.

Patterns of interest for epidemiologists and clinicians are immediately apparent in the counties directly above and below the dark blue Marion County. For example, the steadily darkening hue from the center outward in Hamilton County to the north shows a significant increase in occurrences from 2008 to 2009. Johnson County to the south, on the other hand, shows a significant lightening of hue from the center of the county outward over the same time period. In this, HealthTerrain achieves one of its main goals by providing an opportunity for hypothesis generation and/or validation.

Suppose, for example, that Johnson County implemented a pesticide program to curb tick populations in the spring of 2008. A government clinician evaluating the program might notice the significant decrease in occurrences of Lyme disease in Johnson County in the midst of rising neighboring county incidences. The interactivity of the system allows such a hypothetical user to narrow and broaden the timeline to quickly and efficiently validate whether these trends coincide with the timing of the program or instead are part of some bigger trend with other causes.

4.3 Ring Graph Results

Figure 2.5.1 shows an example of the ring graph for Hepatitis A, B, C and D over time. The innermost circular region represents Hepatitis D, which has very few cases. Users will see an obvious pattern in the next circular band representing Hepatitis A; while Hepatitis A occurs fairly evenly across the age spectrum, there exists an obvious visual exception to that distribution in a concentration of young black patients. The next band, Hepatitis C, shows a heavy concentration of middle-aged patients. Cross-arcs in the ring graph represent the same patient being diagnosed with multiple diseases at different times.

As with the other visualizations, much of the usefulness of the ring graph comes from the interactive explorations that are possible with this real-time system. Patterns emerge when users view larger numbers of terms and greater spans of time. Users may adjust and filter the results included in the visualization or hover over patients involved in the patterns in order to drill down and solicit clarifying data at a finer-grained resolution.

4.4 Conclusions

HealthTerrain provides an interactive browser-based platform for data exploration and visual analytics of sizeable healthcare data collections. It contributes meaningfully to the body of technical visualization knowledge as one of the first systems to successfully organize and synthesize large-scale clinical data sets in an interactive environment. This prototype system augments users' ability to discern patterns in disparate healthcare data, ultimately leading to better-informed decisions by clinicians and managers for individual patients and populations.

HealthTerrain's information-rich concept space effectively compresses large, heterogeneous, and historical patient and public health data into a unified, intuitive, and comprehensive representation. Any patient-based dataset can be easily converted into a concept space using a set of standard text- and data-mining tools, which can then be visualized by the system. Future work seeks to provide more customizable interface features so that the system can be adapted to different healthcare applications. This will allow the visualization system to operate independently of specific data formats and will also help facilitate interoperability among multiple EHR systems.

HealthTerrain's visualization techniques are specifically designed for the interactive exploration of healthcare data. The association map provides an initial platform to explore relationships of various health concepts within a controlled ontology. The ring graph provides a means to explore temporal patterns within patient populations. The offset contour map approach is an innovative technique for geotemporal data visualization, providing a way for users to filter different color bands within the context of a spatial landscape. This is particularly important for healthcare data, which is often composed of geographic information and population-level disease information. The use of offset contours for time-series data provides a unique non-animation based spatiotemporal visualization with a rich geographic context.

With increasingly large collections of clinical and notifiable disease data, the possibilities to explore many potential correlations between particular diseases and other clinical features (such as clinical concepts ground in discharge summaries) continue to grow. Such correlations may reveal predictors of clinical outcomes and suggest potential future interventions to reduce disease burden. For example, in previous studies the HealthTerrain research team identified specific communicable diseases that were associated with other rates of co-morbid communicable disease.[23] Using the ring graph and association map visualizations, we explored potential correlations among communicable diseases (e.g., HIV and syphilis), clinical concepts related to communicable diseases (e.g., alcohol use and sexually transmitted disease), and temporal correlations among diseases.

To ensure usability, HealthTerrain's current visualization framework was designed with valuable input from public health and clinical stakeholders. As part of an initiative currently funded by the Department of Defense, it leverages usability guidance[24, 25] to assess workflow, alerts, navigation, and layout as well as visualization effectiveness. In addition, assessments involving hospital group managers, regulatory entities such as the Centers for Disease Control, and medical researchers will inform future revisions of the framework.

The challenges involved in creating visualizations for unstructured, large, geotemporal healthcare data sets have barred epidemiologists and clinicians from exploring large-scale healthcare data sets and discovering the wealth of valuable information embedded at the individual patient and population levels. HealthTerrain addresses these challenges, uniquely integrating geospatial and temporal healthcare data within a common visual context in a way that other visualization systems do not. With continued stakeholder support and feedback, the prototype techniques made available through the HealthTerrain system will be refined and many new visualizations will be created.

HealthTerrain is poised to positively impact the working lives of clinicians and epidemiologists by making possible an unprecedented large-scale synthesis of healthcare data. In the hands of public and clinical healthcare professionals, the result of the HealthTerrain research team's efforts can provide a rich context for local patient care decisions and better inform public health recommendations.

REFERENCES

REFERENCES

[1] Grossman C, Powers B, McGinnis JM (Ed). Digital infrastructure for the learning health care system: the foundation for continuous improvement in health and health care. The National Academies Press, 2011

[2] Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B., Lifeline: Visualizing Personal Histories, CHI, 1996, pp. 221-227.

[3] Wang, T.D., Plaisant, C., Quinn, A.J., Stanchak, R., Murphy, S., Shneiderman, B. Aligning Temporal Data by Sentinel Events: Discovering Patterns in Electronic Health Records, CHI'08, 2008, pp. 457-466.

[4] Bui, A., Aberle, D.R., Kangarloo, H. Timeline: Visualizing Integrated Patient Records. IEEE Trans. Information Technology in Biomedicine 11(4):462-473.

[5] Mane, K., Borner, K. Computational Diagnostics: A Novel Approach to Viewing Medical Data. Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization, CMV '07, 2007, pp. 27-34.

[6] Hallett, C. Multi-Modal Presentation of Medical Histories. IUI'08: 13th International Conference on Intelligent User Interfaces. 2008, pp. 80-89.

[7] Carroll LN et al. Visualization and analytics tools for infectious disease epidemiology: A systematic review. J Biomed Inform (2014), http://dx.doi.org/10.1016/j.jbi.2014.04.006.

[8] Grannis SJ, Egg J, Overhage JM. Reviewing and managing syndromic surveillance SaTScan datasets using an open source data visualization tool. AMIA Annu Symp Proc. 2005:967. PubMed PMID: 16779254.

[9] Maciejewski R, Rudolph S, Hafen R, Abusalah A, Yakout M, Ouzzani M, Cleveland WS, Grannis SJ, Wade M, Ebert DS. Understanding Syndromic Hotspots - A Visual Analytics Approach. IEEE Symposium on Visual Analytics Science and Technology, pp. 35-42, 2008.

[10] Duke JD, Li X, Grannis SJ. Data visualization speeds review of potential adverse drug events in patients on multiple medications. J Biomed Inform. 2010 Apr;43(2):326-331. PubMed PMID: 19995616.

[11] Maciejewski R, Hafen R, Rudolph S, Tebbetts G, Cleveland WS, Ebert DS, Grannis SJ. Generating synthetic syndromic-surveillance data for evaluating visual-analytics techniques. IEEE Comput Graph Appl. 2009 May-Jun;29(3):18-28. PubMed PMID: 19642612.

[12] THE NEW YORK TIMES COMPANY: Openpaths, Retrieved February 01, 2013, from https://openpaths.cc.

[13] GOOGLE: Latitude, Retrieved February 01, 2013, from http://www.google.com/latitude/.

[14] ECCLES R., KAPLER T., HARPER R., WRIGHT W.: Stories in GeoTime. In VAST (Oct. 2007), Ieee, pp. 19–26.

[15] Kraak, Menno-Jan, and P. F. Madzudzo. "Space time visualization for epidemiological research." ICC 2007: Proceedings of the 23nd international cartographic conference ICC: Cartography for everyone and for you. 2007.

[16] Kraak, M. J. and A. Kousoulakou (2004). A visualization environment for the space-time-cube. Developments in spatial data handling 11th International Symposium on Spatial Data Handling. P. F. Fisher. Berlin, Springer Verlag: 189-200.

[17] Andrienko, N., G. L. Andrienko, et al. (2003). Visual data exploration using space-time cube. 21st International Cartographic Conference, Durban, South Africa.

[18] Overhage JM, Grannis SJ, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. Am J Public Health. 2008 Feb;98(2):344-50. PubMed PMID: 18172157.

[19] Fighting disease outbreaks with two-way health information exchange, last retrieved from http://newsinfo.iu.edu/news/page/normal/11948.html

[20] Stephen G. Kobourov. Spring Embedders and Force Directed Graph Drawing Algorithms. arXiv: 1201.3011.

[21] Bostock, M. D3 Geo Projections. Retrieved September 24, 2014, from https://github.com/mbostock/d3/wiki/Geo-Projections

[22] Gossett, N., Chen, B.: Paint inspired color mixing and compositing for visualization. In: INFOVIS ’04: Proceedings of the IEEE Symposium on Information Visualization, pp. 113–118 (2004)

[23] Gichoya J, Gamache RE, Vreeman DJ, Dixon BE, Finnell JT, Grannis S. An evaluation of the rates of repeat notifiable disease reporting and patient crossover using a health information exchange-based automated electronic laboratory reporting system. AMIA Annu Symp Proc. 2012;2012:1229-36.

[24] NIST Interagency/Internal Report - 7432. (2010) Common Industry Specification for Usability — Requirements. Retrieved July 10, 2014 from http://www.nist.gov/manuscript-publication-search.cfm?pub_id=51179.

[25] Zhang J, Walji M. TURF: Toward a unified framework of EHR usability. Journal of Biomedical Informatics, 2011; 44 (6):1056-1067.

PUBLICATION

PUBLICATION

Many of the ideas and concepts presented in this thesis have been submitted and accepted for publication in the Journal of the American Medical Informatics Association. The date and issue of publication has not yet been determined:

Fang, S., Palakal, M., Xia, Y., Bloomquist, S., Nguyen, T.M., Krishnan, A., Jiang, S., Li, W., Keiper, J., Grannis, S. Health-Terrain: A Visual Analytics System for Health Data. J. Am. Med. Inform. Assoc.