A Web Application to Visualize Trends in Diabetes across the United States

Final Project Report

Team: New Bee

Team Members: Samyuktha Sridharan, Xuanyi Qi, Hanshu Lin

Introduction

This project develops a web application using data visualization to enable users to take informed decisions. The web application in this project visualizes data pertaining to diabetes patients across all states in the United States of America over several years. The application is developed as a predictive analysis tool for understanding the future trends in diabetes. The web application is designed to enable users to easily navigate between functionalities, visualize data for any given functionality using different visualization techniques, and also across different years.

Purpose of the Application

The purpose of this application is to provide information in the form of visualization for trends in Diabetic patients in the United States. This includes trends related to hospitalizations, medications and preventive care measures due to diabetes. This visualization application will help in Predictive analysis. Predictive analysis is a method of estimating what would happen in the future. Such an analysis could be achieved using the historical data visualized coupled with statistical methods such as trend line plotting. This will help users to be prepared for future outcomes.

Targeted Audience and their Goals

The targeted audiences of this application, also referred to as users, are doctors and analytical people. Their goal would be to understand the implications of diabetes, and their trend of several years. The graphical methods used in this application will also help the users to predict future outcomes and understand the correlation of different parameters affecting diabetic patients. The users also include the common people, such as the people living in the United States, who would like to know more about diabetic trends.

Methodology

The step-by-step methodology adopted for this project is outlined as follows.

1. Data Understanding: This step deals with familiarizing with the different variables of the data set, their data types and identifying missing data.

2. Data Structuring based on Tableau: The visualization software Tableau would be able to read input data in the Excel, Access or Text format. This project uses excel for data input. The first row should be the headers, and subsequent rows are considered as data.

3. Decision of Visualization Designs: This step includes deciding on the types of charts to plot, such as scatter plot, bar chart, Maps. The decision is based on mainly two factors: (i) selection of meaningful and useful data for the users, and (ii) type of possible plots based on selected variable's data type (nominal, categorical, etc.).

4. Prepare / Draw charts using Tableau.

5. Decision of analytical interactions within visualizations: There are several interaction techniques generally available for use. This project uses techniques such as filtering, highlighting, data tips, comparisons, zooming and panning.

6. Publish to Tableau server: This step generates a web link for each visualization, which will be embedded later in to the web application.

7. Create front-end design: This step creates an aesthetically appealing web page front end that the users interact with.

8. Embed to web application: The web links (for the different visualizations) generated in step 6 is embedded to the web application.

Assumption

The project assumes that users will use this web application in a desktop or laptop browser.

Project Accomplishments

The goals of this project were accomplished as follows.

- The team understood the data set thoroughly including the variables and the variable types.
- Data clean-up, such as removing garbage values, was performed.
- Different visualizations and trend charts were chosen to suit the data which would convey useful meaning. This resulted in a set of reasonably suitable visualizations.

Data

The data being studied and used in this project is related to diabetes patients, and is mostly timeline. Several years' of data across all states of the US are used for comparison. Such a comparison helps in understanding how diabetic patients' trends (data values) change over time, and with respect to geographic location. The data is obtained from Centers for Disease Control and Prevention (CDC) website [3].

Tools used

The technical tools used for this project are:

- NetBeans IDE 8.0.1 (html and Javascript): This has been used to develop the front-end design of the application. HTML and style sheets were used to implement design and navigation patterns, and the visualizations were embedded to the front-end using JavaScript API [1].
- Tableau: Tableau [2] was used to develop the different visualizations. It gives the flexibility of using geographical maps (for state data) for realizing good visuals. It is very powerful for plotting bar graphs, scatter plots, etc.
- Apache Tomcat Server: Tableau visualizations run in server hence, Apache TomCat is used for this purpose.
- Adobe Dreamweaver CS 6: To develop HTML pages.

Design Decisions

Designing of Web Application involved various design decisions to make. This application is targeted at viewing in a browser. Design decisions were involved in choosing of organization design patterns, navigational design patterns, visualization designs. Care has been taken in view of perception and color choices.

Organization Design Pattern

In order to design the overall layout of the application, it was suitable to adopt **Multiple Workspaces.** This is because, the users may want to view different graphs and trends in the application. Simultaneously, the users will also be provided with an overall view of the application at every point in time he/she spends with the application.

Navigation Design Pattern

Navigation is one of the key decision parameters when designing web application. The **Fully Connected** navigation pattern is used. This allows the users to navigate between pages and also return to the home page from each page.
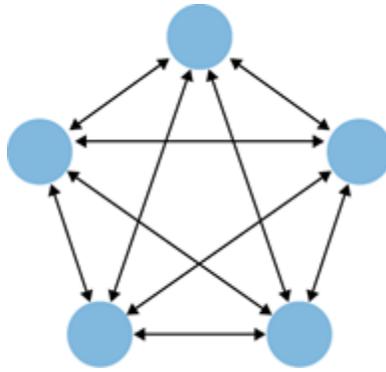
Visualizations

The visualizations in this web application show the following:

1. The 'Diabetes Prevalence' page shows the relationship among number of people diagnosed as diabetic, their adjusted ages and corresponding obesity percentage, within each geographic location for several years (year used as a filter field).
2. The 'Hospitalization' page shows the comparison of the number of discharges and the length of stay in the hospital for people diagnosed diabetic, as a first listed diagnosis or any listed diagnosis, across several years.
3. The 'Treatment' page shows the percentage of adults within diabetes by diabetes medication status, across several years.
4. The 'Self Reported Risk Factor' page shows trend pertaining to risk factors in diabetic patients.
5. The 'Preventive Care' page suggests the trend in people (as percentage) who took different preventive actions, and were also diagnosed diabetic.

Visualization Design Principles

The visualizations have been created using the following principles.

Scatter plot: This is a mathematical diagram that uses Cartesian coordinates to visualize the correlation between two variables. This helps in identifying patterns. The variables plotted in a scatter plot needs to be nominal values, and their values need not start from zero. With this understanding, the values (in percentages), to know the clustering and pattern of lower confidence and upper confidence values, have been plotted as a scatter plot.
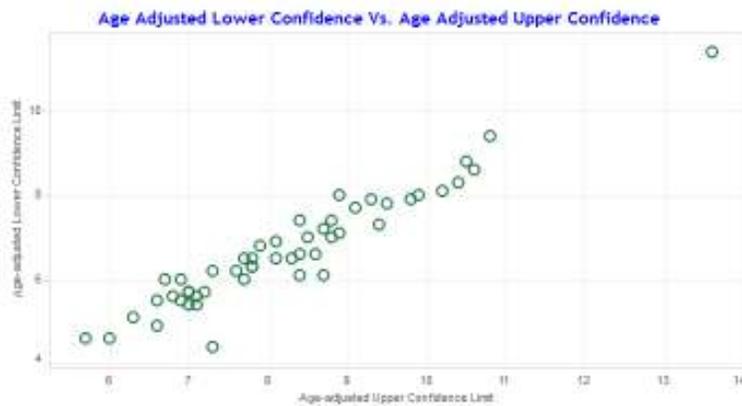

Line Graph with Trend Lines (to represent trends): The line graph is used to visualize how one variable varies with respect to another variable, such as time. Between two points on the line graph, a temporary trend can be observed, such as increasing, decreasing or flat. A trend line can also be included with the line graph to visualize and understand the overall trend.
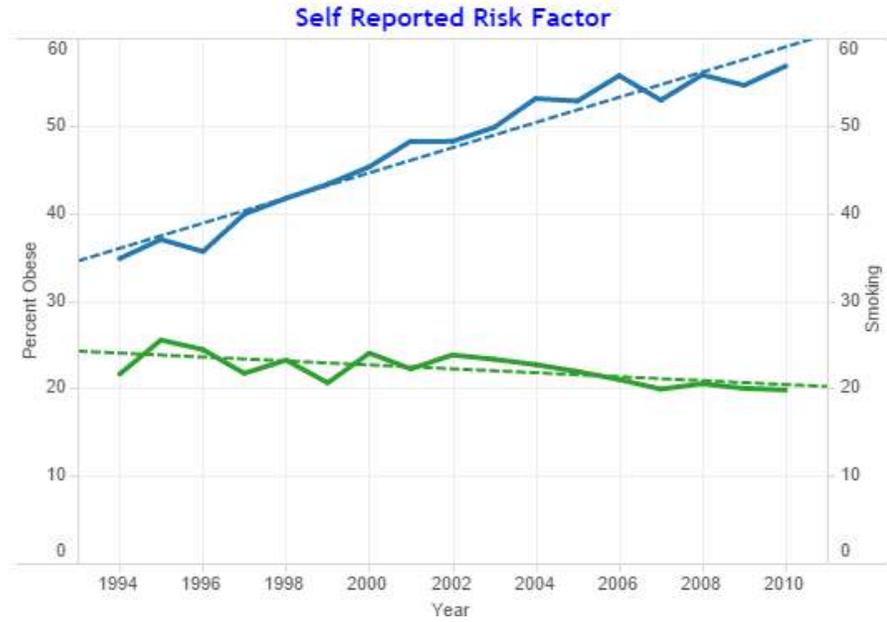
Bar chart: A bar chart is a method of representing variables where the length of the bar is proportional to the value represented by that variable. A bar chart always starts from zero. The specific categories being compared are shown on one of the x-axis (usually the x-axis for a vertical bar graph), while the discrete values are represented on the y-axis.
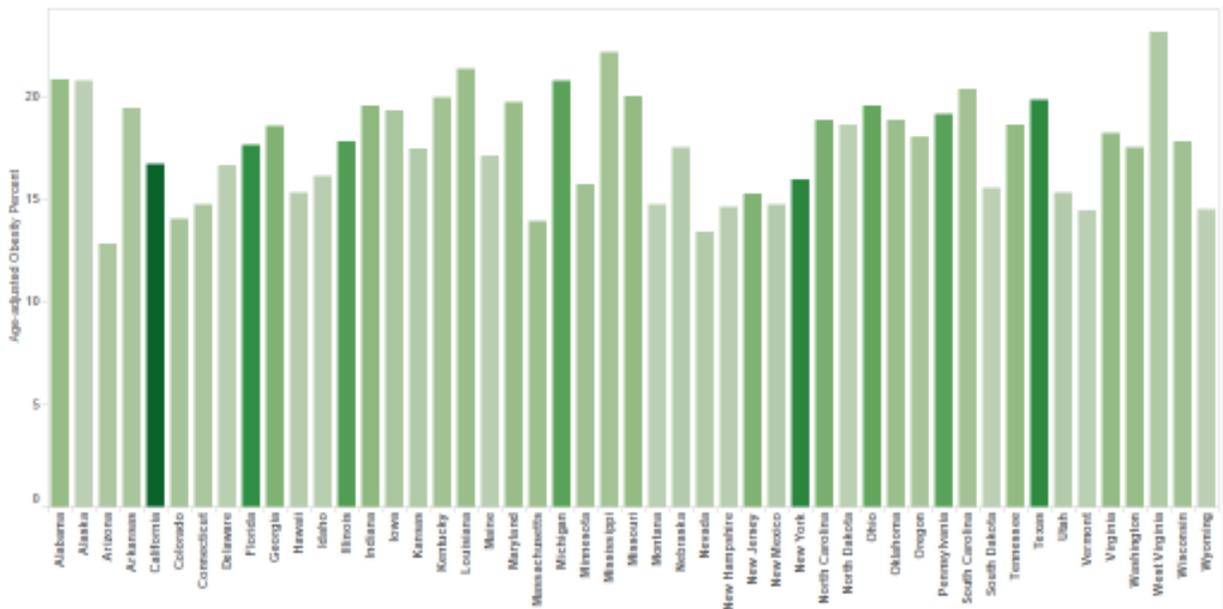
Maps: Maps are a method of incorporating geographical visualization. The US map is used in this project to visualize data corresponding to each state. It would provide the user with geographical perception and

visualization of the data being studied. The states within a map are colored using color gradations of a particular color.
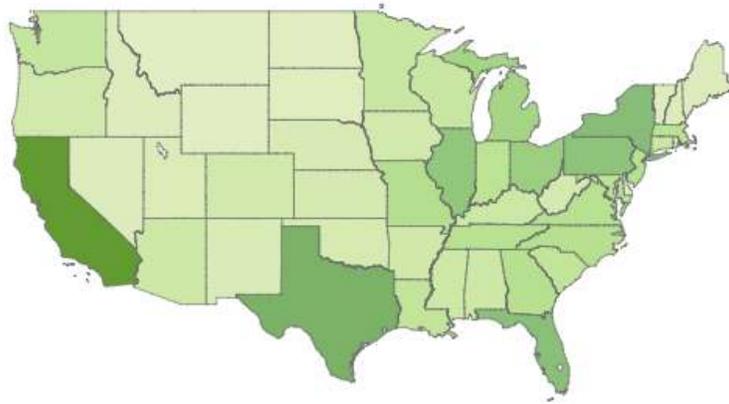
Dashboard Layout: The web application also has dashboards which were designed to represent state level data. The dashboard should achieve a single goal and in this application it achieves the goal of comparing the percentage of people obese compared to the number diagnosed diabetes.

The dashboard design follows Shneiderman's Mantra [4]. The overall structure of the design is displayed first and details will be provided on demand. The users will have the opportunity to drill down to the data they would be interested in for further analysis.

Note: The term dashboard used in this project refers to a dashboard-style layout. However, this dashboard is not refreshed on a daily basis, as a typical dashboard is intended to do. This is because of the type of data being handled (healthcare data monitored over years, unlike a manufacturing environment where data is monitored over days).
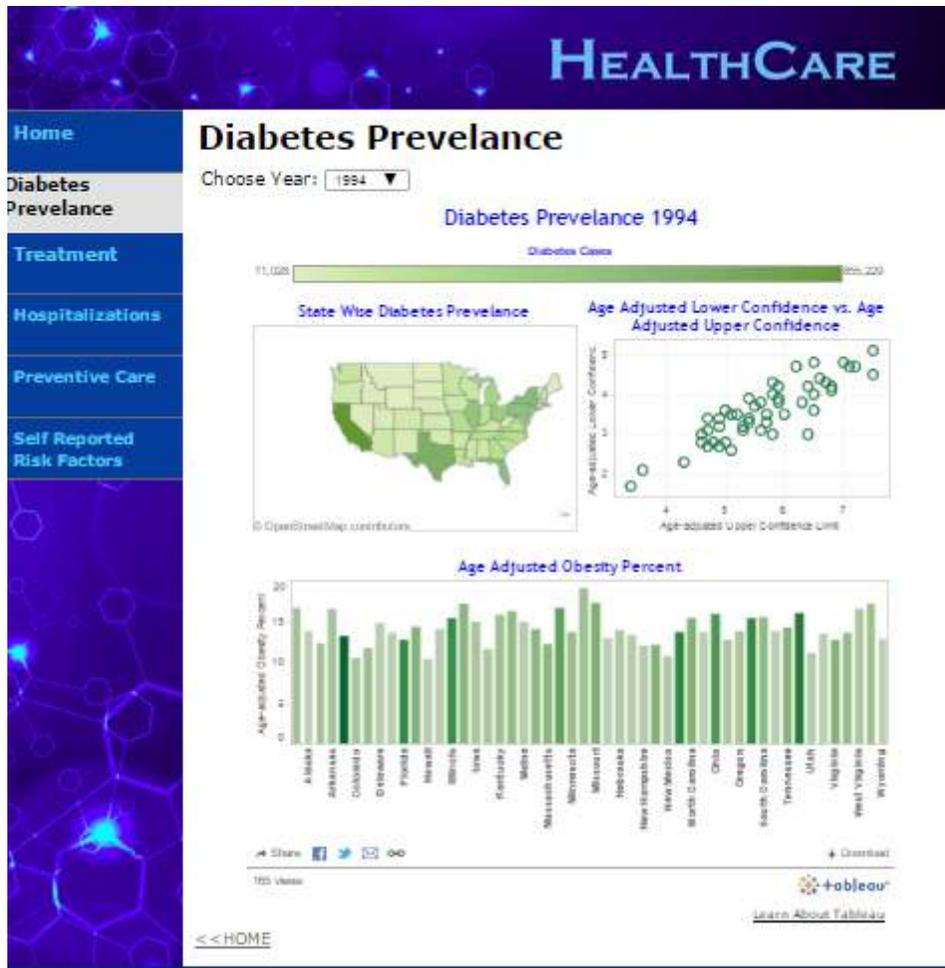
Analytical Interactions with Data

A good visualization provides meaningful interactions within the data. The visualizations developed in this project involve different interaction techniques. They are listed as follows.

1. **Comparison**

   An effective data analysis is to allow comparisons. The dataset chosen for this project had data for several years and comparison of data across years was thought suitable.

2. **Sorting**

   The visualization graphs are allowed to be sorted to analyze the same data in different perspectives.

3. **Filtering**

   The filtering option is provided for users to choose specific years of data. Users can choose year and obtain data and trend corresponding to that year.

# Diabetes Prevelance

Choose Year: 1994 ▾

4. Highlighting

Two levels of highlighting are used in the visualizations developed for this project. They are:

i. Brushing – Highlight the same data in multiple graphs

Brushing was used in the dashboards since users can click on each state and data corresponding to each state is highlighted.

ii. Data Spot light – Selected field is highlighted and the rest are dimmed
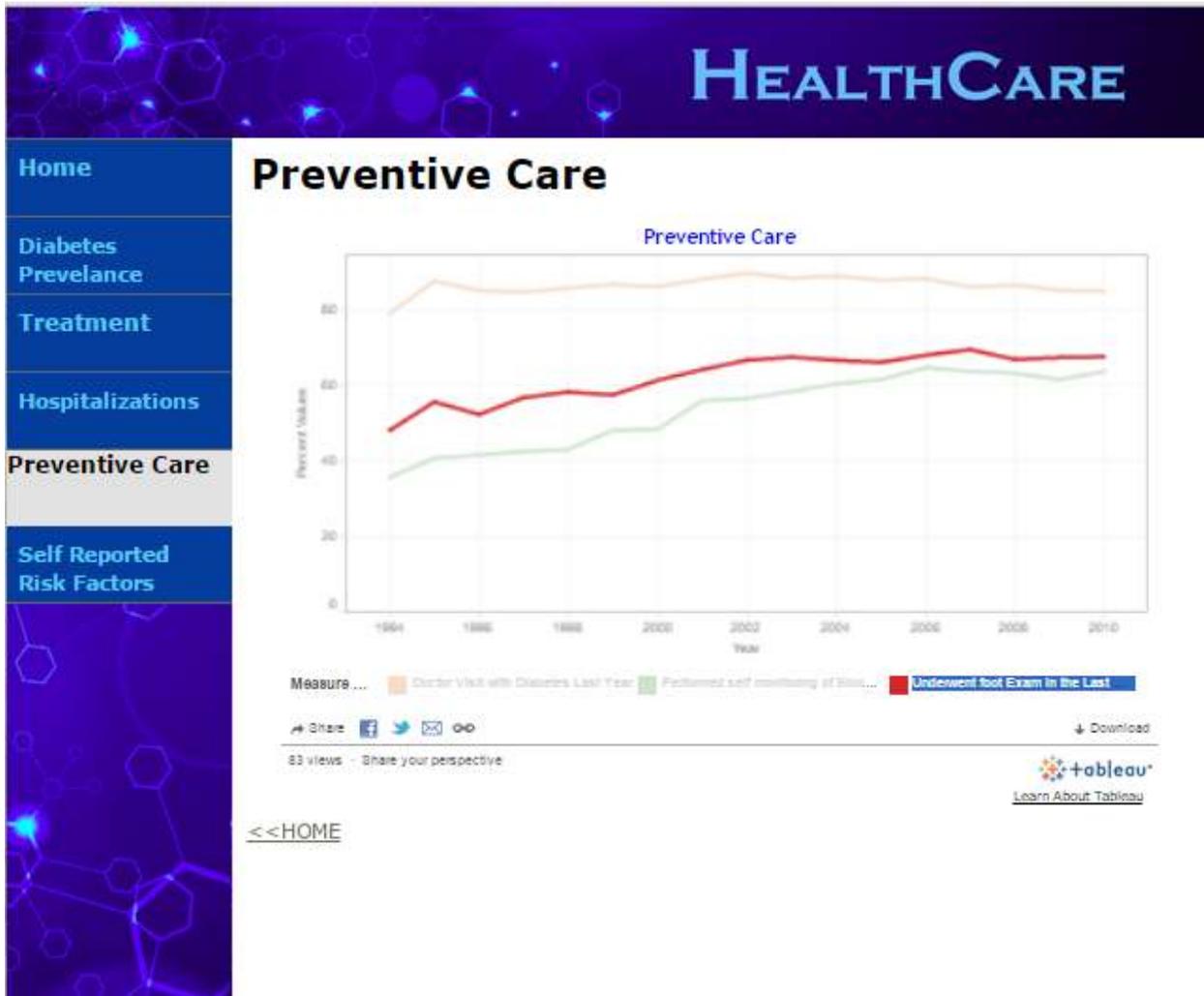
5. **Zooming and Panning**

   In the overall display of dashboards, the map is displayed as a small object. If the user wants to zoom to look for states in the east coast, the user would be able to do that by zooming and panning across the maps.

6. **Data Tips**

   Users can get more information by hovering over the mouse. A tool tip with more information regarding the data will be displayed to the user. This feature would aid the user to understand and analyze more data and to further drill down to the atomic level.
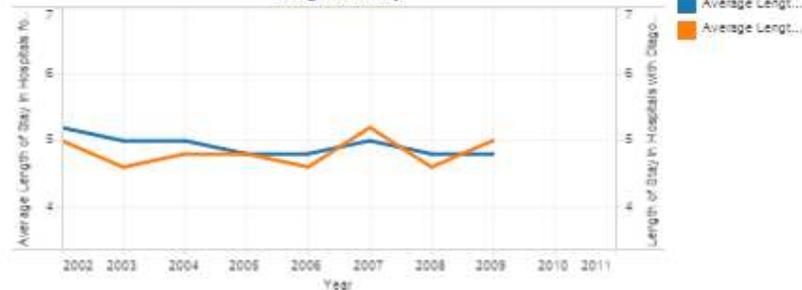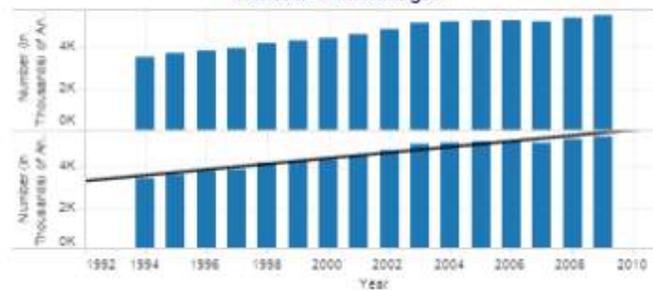
Conclusions

This project successfully developed a web application for interactively visualizing diabetic data. The meaningful plots using the available data were generated using Tableau. Interactions between different plots were also provided in the visualization. Data trends and comparison over several years were implemented. A front-end we application was generated giving importance to aesthetics and user interface. All the visualizations were embedded in to the web application.

Whenever data is updated to the data source for more recent years, the visualization can be refreshed (plots can be regenerated for updated data), and the same web application reused.

Future Work

This web application is currently designed for desktop and laptop browser screens. The same methodology can be applied with different design decisions for use in mobile browsers.

User Guide to Run the Application

The application users are requested to follow the below guideline to run and use this web application.

1. Unzip original project zip file.
2. Open NetBeans IDE, click "File" in tool bar and open original project folder.
3. If there is a warning of "Missing Java EE Server" under "Libraries", right click on project "HealthCare" and click "Resolve Missing Server Problem" to configure a server which is indispensable to support Tableau API.
4. Right click on project "HealthCare" and click "Run" to run this web application on local server.

References

[1] http://www.tableausoftware.com/new-features/javascript-api, Accessed on Nov 30, 2014.

[2] http://www.tableausoftware.com/public/, Accessed on Nov 15, 2014.

[3] http://www.cdc.gov/diabetes/data/national.html, Accessed on Nov 15, 2014.

[4] Ben Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336-343, Washington. IEEE Computer Society Press, 1996. http://citeseer.ist.psu.edu/409647.html