Disease Control and Prevention Analysis

Shraddha Kulkarni
shraddhk@usc.edu

Seher Khan
seherkha@usc.edu

December 4, 2019

Abstract

The paper presents an analysis of Mortality rates, Disease Control and Prevention techniques across various States of the United States using advanced visualizations created with the help of d3.js. The main goal of our project summarizes Mortality Statistics.

Disease Control, Visualization, Analysis, Mortality Rate, Healthcare Expenditure, Clustering

1 Introduction

Some of the leading causes of death in the United States include heart disease, cancer, alcoholism and injuries caused by accidents. Many of these are preventable, and being able to visualize the numbers could help us develop improved strategies to avoid and alleviate their occurrence. We expect our application to be useful for the general public. It will make them more aware of the diseases they are susceptible to based on their location and other demographic traits. Medical students could also benefit from knowing which fields require their interest the most. Finally, policymakers would greatly benefit from a tool that would help identify major health problems in their region and so they can draft effective policies. The detailed study of Historical data on Mortality Rates and Causes of Deaths from the year 1950-2016 show interesting trends.

Using various visualizations that we've learned this semester through the course Information Visualization taught by Dr. Luciano Nocera, we try to represent various statistics related to Mortality Rate, and the causes of Deaths. We've used the Pandas library in Python to read data of multiple years and combine them to find their average values and store them as a CSV file. We've also used tools such as Microsoft Excel to read the data and get the required values from it using its pivot table functionality. The visualizations that we've created are using d3.js and the website is created in Vue.js. We've used different node modules such as topojson, bootstrap, particles js to make the website more interactive and visually appealing.

2 Data

We've used data that has been collected over the years by the Centers for Disease Control and Prevention (CDC). CDC is one of the major operating components of the Department of Health and Human Services. It is the nation's protection agency, working 24/7 to protect America from health and safety threats, both foreign and domestic. CDC increases the health security of our nation.
The data could be found at the Url: *https* : *//www.cdc.gov/nchs/*
The data is available in pdf, excel/CSV, and ppt format for download. The datasets are created based on particular topics such as Population, Poverty, Disease Control, Births, Abortion, Life Expectancy, etc. There are tables which have a summary of data over the years, which were the main reference for this project.

3 Information about the dataset

- Age-adjusted rates are calculated using the year 2000 standard population. Before 2001, age-adjusted rates were calculated using standard million proportions based on rounded

population numbers.

- Starting with 1999 data, the rules for selecting CLRD and Pneumonia as the underlying cause of death changed, increasing the number of deaths for CLRD and a decrease in the number of deaths for Pneumonia. Therefore, trend data for these two causes of death should be interpreted with caution.
- Starting with 2011 data, the rules for selecting Renal failure as the underlying cause of death were changed. The result is a decrease in the number of deaths for Nephritis, nephrotic syndrome, and nephrosis and an increase in the number of deaths for Diabetes mellitus. Therefore, trend data for these two causes of death should be interpreted with caution.
- Rates for 2001 include September 11-related deaths for which death certificates were filed as of October 24, 2002. Death rates for Hispanic, American Indian or Alaska Native, and Asian or Pacific Islander persons should be interpreted with caution because of inconsistencies in reporting Hispanic origin or race on the death certificate compared with population figures. The net effect of misclassification is an underestimation of deaths and death rates for races other than white and black.

4 Approach

4.1 Design Consideration

Our design exhibits an equal balance between unidimensionality and multidimensionality. Various factors such as gender, race, origin, age-group, and expenditure are displayed using Maps, graphs, and charts. Interactivity ensures multidimensionality. For our audience's understanding, we have used familiar charts and avoided redundancy. The visualization is more abstract than figurative because most of the references applied are conceptual tools such as bar, line, and pie rather than vivid physical realities. Sticking to functionality has been our forte.

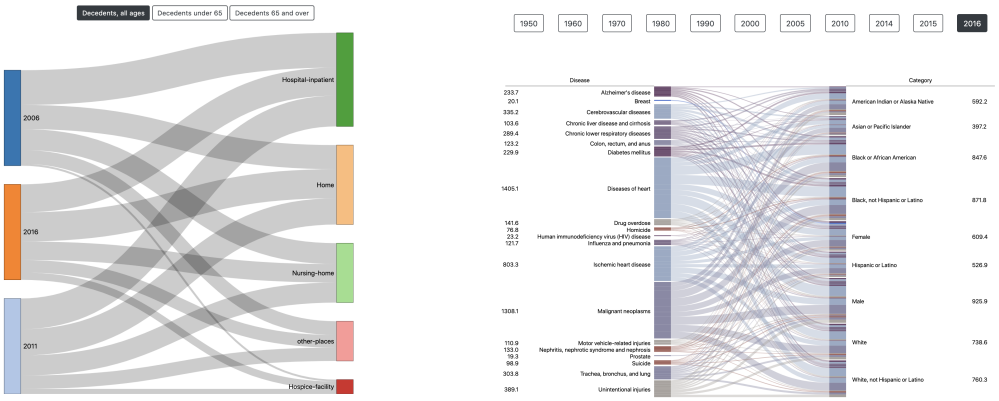
4.2 Technical Considerations and Development Plan

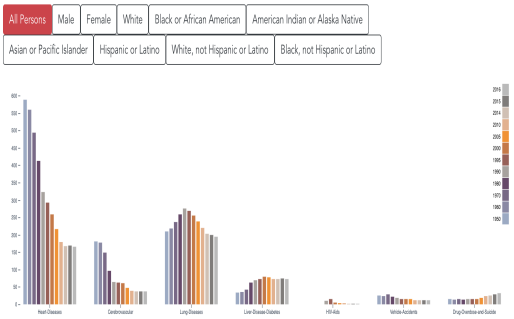
This project has been developed on the Vue.js framework. Separate components have been created for each of the charts created in Vue. D3.js is a JavaScript library for manipulating documents based on data, which we have used to construct our charts for visualizing. We used bootstrap for grid layout and CSS for styling and creating buttons.

5 System

5.1 Disease-Category Correspondence

This visualization consists of two parts i.e. Sankey Diagram and Grouped Bar Chart. The Sankey Diagram represents the place of Death of the residents, which can be one of Home, Hospital inpatient, Nursing home/Long-term care facility, Hospice facility or other places.

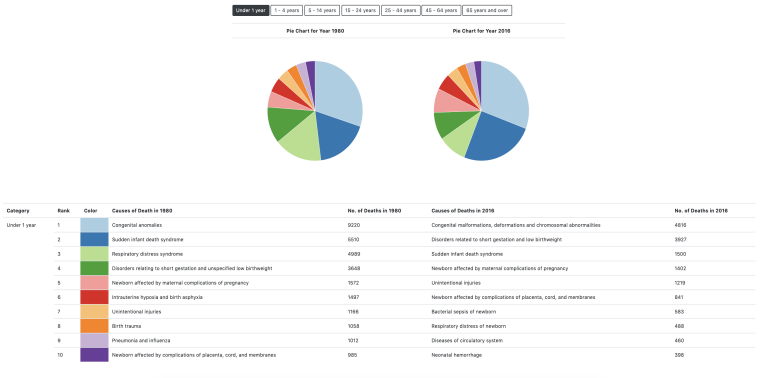




We have also made a Bipartite Graph which displays one-to-one correspondence between the Diseases and Category. This figure helps us to analyze the major reasons for Death for particular sex, race or origin. Since it's a two-way chart, we can explore the breakdown from both sides i.e. selecting a Disease from the left side will give us a breakdown of deaths due to that disease. Selecting a Category from the right side displays a breakdown of diseases that have led to deaths for a particular community of people.

Grouped Bar chart helps us to analyze the trends over the years for a particular category. Selecting sex, race, or origin, the screen dynamically displays a vertical bar chart for major Diseases and their height is the Death rate over the years. It starts in 1950 and goes on until 2016, with a gap of 5 years. The major diseases covered included Hearth Diseases, Cerebrovascular, Lung and Liver Diseases, Diabetes, HIV Aids, Vehicle Accidents, Drug Overdose and Suicide. The Heart and Cerebrovascular diseases have shown a steady dip, whereas Lung and Liver Diseases have a fluctuating trend. There has been an increase in deaths due to drug overdose and Suicide, which is a clear indicator of the dire Mental health conditions across the globe.

5.2 Age Group Mortality Exploration - I

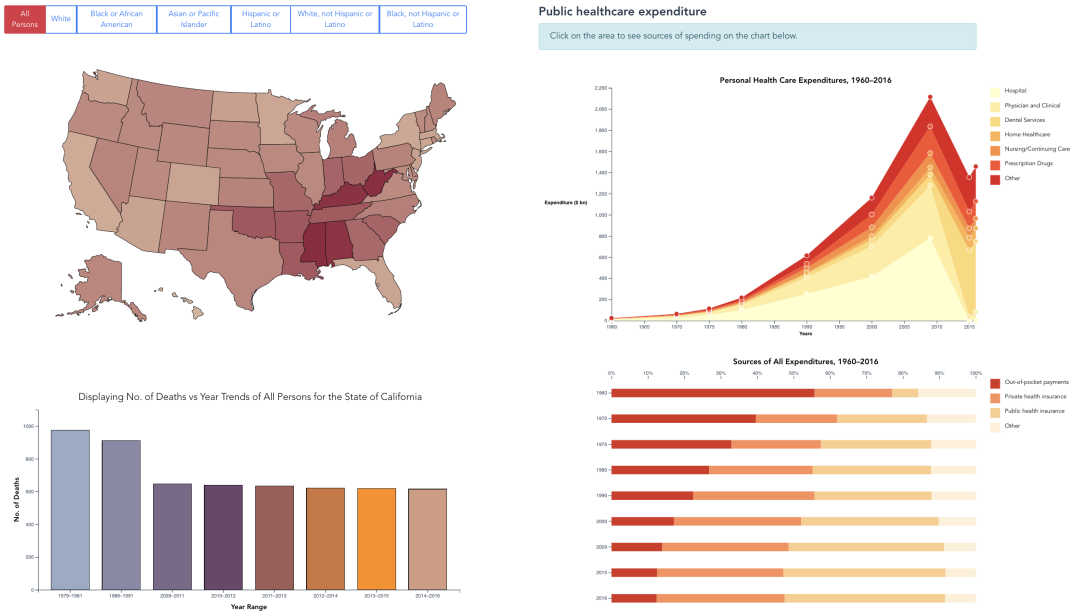


The visualization consists of two pie charts and a table, for the years 1980 and 2016. There has been an attempt to compare the leading causes of death in the two respective years. Considering there are more than 35 years of difference, there is an expectation that the causes which were more relevant back then have taken been brought under control. Unintentional injuries, Congenital malformations, chromosomal abnormalities, and Malignant neoplasms are major reasons for death for kids aged 4 years or less. Whereas, Suicide and Homicide seem to be more significant for children between the age of 5-24 years. While 40+ people mostly have suffered from Heart Diseases, Diabetes, Liver and Respiratory Diseases.

5.3 Age Group Mortality Exploration - II

The Heatmap serves to highlight the trend in death rate across time and age-groups. The default "overall" view colors each cell concerning its number. This is to highlight any concerning number immediately. We can then either select "Age" or "Years", through which the cells color range can be altered. For example, by selecting "Year" we find that distribution of deaths across age groups has remained roughly the same over the decade. On the other hand, by selecting "Age", we can observe that only in the 1-14 and 65+ years categories there is a clear and consistent decline in deaths. The Stacked/Grouped bar chart provides the same data as the heatmap but with greater granularity, allowing the user to spot subtle trends.

5.4 Geographic Mortality Exploration and Healthcare Expenditure



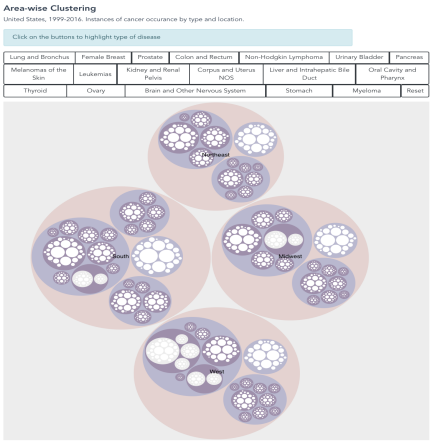
This visualization consists of a Choropleth and a Bar Chart. It represents the data from the year 1970 to 2016. Clicking a particular race, or origin, the choropleth changes dynamically. The page also enables the user to click on particular States to look at the individual Death trends over the years in a particular State.

The Stacked Area Chart depicts the trend in public health expenditure in the US by category since the 1960s. We can see that health care expenditure shot up since 1980 from a little over 200bn to 2100bn dollars in 2010, since then it has fallen and stabilized around 1400bn dollars. In the future, it would be interested to see if this is a result of large population growth or other factors.

The complete Stacked Bar Chart shows the source of these expenditures over the same period. In default view (which is for total healthcare expenditure) we can see that over half of the spending was sourced from out of pocket payments in 1960. However, this rapidly declined and was replaced by both public and private insurance.Clicking on a specific category in the area chart (or the

legend), update the second chart. This allows one to see trends in the source of funding for a particular category. Clicking on the white chart area resets the second chart.

5.5 Area-wise Clustering



The first circle packing chart displays data of the top five causes of death by region (division, subdivision, state, and city) over the period 1999-2016. Clicking on the button of cause colors the circle of that type in all bubbles. This can be used to focus on the data of one or a few causes of interest. Similarly, the Circle Packing chart for cancer displays data of cancer cases by type and by region (division, subdivision, state, and city) over the period 1999-2016. Clicking on the button of a cancer type colors the circle of that type in all bubbles. This can be used to focus on the data of one or a few cancers of interest. For example, by selecting Lung/Bronchus, Breast, Prostate, and Colon we find that these are the most common cancer types across all regions in the US.

6 Conclusion

The main goal of our project summarizes Mortality Statistics. We aimed at creating visualizations that will be useful for anyone who wants to get insights about the Health data. We wanted to create a site that is user-friendly, interactive, and knowledgeable. There were more than 100 datasets on the CDC website, deciding on the most important information, merging and preprocessing the datasets was a huge part of our project. We also particularly focused on creating an aesthetically pleasing site. The fonts and colors have been used accordingly. Overall, a lot of thought has gone over this project, and we believe it will be helpful for the general audience.

7 References

1. Michael Bostock, Vadim Ogievetsky and Jeffrey Heer.: D3: Data-Driven Documents.
4. Murray, S.: Interactive Data Visualization for the Web.O'Reilly Media (2013)
5. Ware, C.: Visual Thinking for Design. Morgan Kaufmann (2008)
6. D3.js, https://d3js.org/
7. The National Center for Health Statistics. 2018.
Available from: https://www.cdc.gov/nchs/products/nvsr.htm