David Larar, dal326
Sterling Kelly swk49
Zhilong Li zl242
INFO 3300
March 5th, 2019

Project 1 Writeup: Static Data Visualization

Our team focused on mortality statistics from the National Center for Health Statistics, a wing of the Center for Disease Control. Arias et al. published a research paper titled, "Deaths: Final Data for 2016," which includes in depth statistics regarding the causes of deaths for specific age groups from 1999 to 2016. In fact, the report was too specific. For example the report had statistics on deaths from "malignant neoplasm of bladder." While these deaths are certainly as important as the rest, we were more concerned with the most common and discussed types of death. We narrowed it down to deaths from alcohol, drugs, suicide, flu, accidents, and heart disease. Not only are these categories incredibly common, but also they are often preventable. Our group was intrigued to see if there was a correlation of spikes in these categories based on certain historical events, or if there was an age trend that society could become more cognizant of to prevent future deaths. To keep the relative perspective of total deaths, we grouped all other causes of death into a category titled "other."

In order to prepare for our visualization, we first had to manually scrape the data from the report into an excel spreadsheet. This process was tedious, however we were unable to find a prepared digital data set as detailed as the CDC report. We considered importing a CSV file, but elected to convert the data into JSON objects which would be easier to work with for the rest of the project.

As a team we ultimately decided to employ a stacked line chart as well as a heat map. Since the visualization had to be static, it kind of forced our hand in the design process. We considered a regular line chart with several lines but decided the graphic would be too cluttered and not effectively show the breakdown of deaths at each age. This decision made it a bit more difficult to decipher the change in deaths across ages for each cause. To illustrate, since drug deaths are stacked on top of suicide deaths, it can be difficult to see if a curve in the drug deaths is a product of it decreasing or a product of the suicide deaths below it decreasing. The colors chosen for the different causes of death were such that the difference causes could be clearly distinguished.

We decided that for three causes we found particularly interesting it was valuable to include heat maps to show in what years and at which ages the causes of death change.

In the heatmaps, we applied the knowledge of visual channels learnt in class to make sure that the colors are intuitive and easy to distinguish. The color scale runs from green to red, green representing a fall in percentage deaths due to the specific cause, and red representing an increase. White was used as a neutral color between the two colors, so that the color hues will never be confusing to the viewer. We believe that this

color scale is intuitive because people would naturally associate a darker red color with something negative, which in this case is that the percentage of deaths increased significantly, and vice versa for the green hues.

From here, we could tackle our main thesis of looking at these periods and speculating a cause. A correlation of an event and a change in deaths does not imply causality and is merely being offered for consideration.

From our visualization, the trends that we wished to communicate can be seen very clearly; the percentage of deaths due to heart disease fell significantly from 1999 to 2016, deaths by drug and suicide increased significantly, especially among the younger age groups. While most would accept these as facts in our society, we hope that our visualization can remind people of the significance of such trends. Medical advancements have gradually improved the survival rate of those suffering from heart diseases, even though the obesity rate in the US continues hitting new records. With the rise of social media and the increasingly stressful work and study environments, more people are resorting to suicide and drugs.

When beginning this project, we first wanted to compare the most common types of death people feared versus the actual most common types of death. However it was impossible to find this type of data for all age groups. But, the story we wish to convey is that many of the deaths we see are preventable and often at predictable ages. Society has an incredible fear of cancer which pushes millions of dollars of research into the field, yet for most young adult years more people are dying from drug related deaths than cancer. After viewing our data visualization we hope to leave our readers questioning how these deaths can be decreased and why they are occuring. We hope to drive a societal discussion of what truly matters. For example, why are there so many suicides in young adulthood? How can we prevent this? As long as we can deliver one bit of information regarding death rates that the reader was unaware of with, we feel that we have done our job. Our visualization is just a catalyst to more investigative research to improve the lives of everyone in our society.

Finally, in terms of work division, we sought to be as equitable as possible. As outlined in our milestones, we had weekly meetings in which we discussed what each team member was to do. We sought to divide the labor in a manner which each group member not only had the same amount of work, but also an equitable type of labor. In other words we had each group member complete tasks which involved coding, design and analysis. At the meetings each member volunteered to complete portions they felt most comfortable with and in the event we struggled to complete an item, we would work collaboratively often with the help of our mentor TA. This project took around ~75 work hours. Implementing the heat map provided us with the most challenge but we are happy with our results.