CS 3300 Project 2: Write-up

Kelly Fam, Jasmine Chen, Salief Lewis
Kjf63, jc2535, sl2367

Our initial dataset contains information related to wildfires throughout the United States from 1992 to 2015. The data includes a variety of characteristics ranging from the name of the fire to the year it took place. The data was retrieved from Kaggle at the following link, https://www.kaggle.com/rtatman/188-million-us-wildfires. For our visualizations we simplified the dataset to focus strictly on California wildfires, the state with the most incidences. Ultimately we chose to include the year of the fire, the cause of the fire, the size of the fire, the coordinates of the fire, as well as the name of each fire. For the map of California, we used this dataset: https://github.com/scottpham/california-counties, which separated California into its counties. From that dataset, we used the name of the counties as well as the location of each county.

Our first depiction incorporates a map of California from the TopoJSON projections, on which we overlaid the geographical data related to each county. Fire size was the main characteristic used to convey the relationship between each occurrence so we color coded the instances. The larger the fire, the lighter each circle appears. Additionally the circles are intuitively proportioned to reflect their size. The number of circles on the map is limited to 40,000, because the website generally freezes and crashes if it is trying to deal with too many data points at once.

Our second visualization represents how and when each reported fire was created. It also reflects what percentages of the fires occurred during which year and what factors contributed the most to their causation. It is a bipartite graph which is used to capture a relationship between two types of objects where the distinction between the types of objects is important. In this particular graph, the two variables it is focused on is the year and the cause of each fire. Now, with this in mind, this bipartite graph highlighted which causes contributed the most to starting a wildfire, and captured the percentages of each cause and year. What is noteworthy about this graph is that despite lightening individually being the highest of the causes of wildfires, combining the other causes still is more than just lightning caused wildfires alone. We used this code as reference for using viz and implementing a bipartite graph: http://bl.ocks.org/NPashaP/5a31486cea3f246c2e7dfd56937aeee0.

The overall website is stylized to reflect wildfires, so we went with a darker background with a fire color scheme for the elements--different shades of red, orange, and yellow. Each circle that marks a wildfire is mapped from the latitude and longitude coordinates.

The map was hard to read with so many points on it, so we decided to implement a filtering system where you can click on the causes and the map will update accordingly. In addition, we also decided to implement zooming in, so users can see each individual point more clearly. Once zoomed in, the user can hover over each fire to see the name of the fire as well as the year it occured. This feature is not present in the zoomed out map because we thought that it was much too hard to navigate through so many tiny circles at once, and decided to only keep the county name hovering in the zoomed out map. We used this code as reference for zooming in: https://bl.ocks.org/mbostock/4699541.

From a geographical standpoint, our visualization tells us that the majority of wildfires occur within the northern part of California. Although these fires are more frequent, they appear to be smaller in comparison to the occasional large fires that affect the southern part of the state.

The visualization also tells us that although wildfires are a profound threat in the state of California, they are largely a result of human development in fire-prone wildlands. The data suggests that lightning is the number one cause of wildfires in California, but it ignores the fact that much of the dryness and humidity in these areas can be attributed to global warming. Perhaps in a future analysis of this dataset we could compare global temperatures to the frequency of fires over the years.

We were surprised to see how dense these incidences were in some areas as opposed to others. This can be explained with a topographic map of the state showing that the areas least commonly affect are those with the highest elevation.