INFO 5100

Cornell University

Pujaa Rajan (pr338)

Luis Plaz (lfp26)

Yuning Yang (yy693)

Project 2: Interactive Data

1. **Work Done by Each Team Member**

**Pujaa Rajan** - We all worked together on looking for data sets and drawing designs. Then, I helped create and clean two additional datasets to the main flight data set, the weather dataset and the delayed and cancelled flight data set. Using the weather data set, I implemented weather icons I found online that change based on which line you click on. I wrote the introduction, instructions, and conclusion part to create a storyline for our data visualization. I also worked on the color scheme with Yuning.

**Luis Plaz** – We all had part on every aspect of the project, refactoring code we might think could be improved and adding new aspects that we could explore with the data set. My main responsibility was code refactoring, managing the data through means of class implementations -trying to merge different data set inputs into one route flight object - as well as the visualization and animation of said objects. In terms of animation the implementation of the plane animation as well as the route arc representations was my main deliverable. Additionally, I was in charge of connecting the object dictionary with the other methods created by my team members and making the changes required to provide the data said methods required into the class attributes.

**Yuning Yang** - We worked on choosing the dataset and brainstorming some interesting ways of visualization. I mainly created the slider control to update the chosen date and the scatter plot to show how delayed time is correlated to several factors. Besides, I also worked on creating legends for each visualization and finalizing the color scheme of our project.

1. **Description of The Data**

We used two datasets

1. 2015 Flight Delays and Cancellations Data
   1. <https://www.kaggle.com/usdot/flight-delays>
   2. “The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations.”
2. January 2015 NYC Weather Data
   1. <https://www.weather.gov/climate/index.php?wfo=okx>

We had to reformat both data sets. First, we filtered the flight data to show flights within the continental US because that’s the map we were using. Even after filtering this, there were too many flight routes to plot. Because of this we decided to plot the month of January day by day. We did this because we wanted to specifically study the effect of weather on flight delays and cancellations. The month of January was a good subset of the data to explore because it had a variety of weather conditions including sunny days, rain, snow, and fog. There was also the winter storm Juno at the end of January.

The flight dataset was extremely detailed. We decided to only keep information that helped us identify flights and their delay details, which includes day and time of departure, flight number, flight carrier, distance, departure airport and its long and lat, arrival airport and its long and lat, arrival time, total delay, the different types of delays and their times and if it was cancelled or not. These were stored in a route object we created for each.

Next, we were able to use the weather.gov website to get January 2015’s data. In this data set, we also had some extra information like precipitation amount, and the high, average, and low temperatures for the day, which we ignored. We parsed through the data to one hot encode whether it was sunny, snowy, rainy, or foggy. Our end goal was to see how these weather conditions affected air travel, so we ignored the other more detailed numerical reports.

Finally, we merged the above two datasets on the date. This way for each flight we knew the weather when it took off, so we could analyze the relationship between weather and flight delays and cancellations better.

1. **Mapping from data to visual elements Description**

We tried to visualize delayed and cancelled flights and some information related to these flights in out project.

For the map, we plot an animated line from NYC to each destination if there is at least one delay/cancellation on that route. Color of the line indicates how many flights are delayed & cancelled on that specific day. We used threshold scale for this color mapping. Using only three ranges can nicely convey the information and avoid overwhelming the viewers with too much information. Mapping higher values to brighter color can also make the viewers focus on the routes that have more delays/cancellations. If the viewers are interested in specific route, they may click on the lines and detailed information will appear to the right of the map. Since there is few delayed flights for each route per day, and using complicated graph may be redundant, we used plain text to show detailed information of the delayed flights.

Besides the lines, we used plane animations to simulate the flights and it is used to show how long was the average delay for that route. There would be a plane if there is at least one delay/cancellation on that day. Duration of the animation is related to the average delay time, and routes with longer delays have longer animation duration. The average delay is scattered in a wide range. Linear mapping starting from 0 will make routes with shorter delays to fly really fast off the map but log transformation will also erase the difference and make it less interesting. So we finally used a linear mapping starting from 5000 ms to show the animation.

In addition, a slider is used to change the date of our visualization. As we choose from different dates, the flight routes information, an icon for the weather, and the points in scatter plot change. Two lines above the slider are for the total number of delayed/cancelled flights from Jan 1st to 31st. What’s interesting to note here is during Winter Storm Juno, the delayed line nosedives and the cancelled line skyrockets. It helps us compare delays & cancellations over the time interval.

For the scatter plot, each circle represents one specific flight. Its color represents the carrier of that flight. An ordinal scale is used here. (In d3 it’s called point scale) Y axis for the points are the flight delay time with a linear scale. There are four possible X feature. For scheduled departure time, we first transform the time in base 60 to base 10 (e.g. 14:30 - > 14.5), and then map it to a linear scale whose domain is [0, 24]. For delay type and carrier, ordinal scale from string to color is used. For distance, we used linear scale. When viewers chooses another X axis, the points will move back and forth to its new coordinate. This makes the visualization more interesting and also make viewers possible to compare the different dimensions of data within one plot.

1. **The Story**

On January 26th, 2015, the January 2015 North American blizzard, nicknamed Blizzard of 2015 and Winter Storm Juno, affected Canada, Central and Eastern United States, and even Southern Greenland and Western Europe. The snow disrupted transportation in Connecticut, New Jersey, Massachusetts, Rhode Island, and especially New York City. Because of the weather emergency, thousands of flights were delayed or cancelled.

In this case, the flight delays and cancellations are uncontrollable by the airlines. However, this is not always the case. More often than not, flight delays and cancellations are not weather related. In this visualization, we show all the flight routes out of NYC that people flew on each day in January 2015 and the weather on that day.

Here are some of the interesting discoveries. Flights with shorter distance tend to delay longer. Though extreme weather does affect flight delays and cancellations, most flights delays on usual days are not caused by weather. (But when taking a plane, they always tell you that there is a delay due to the weather.) Most delays are indeed caused by air system problems or airlines themselves.

References

Draw arc paths: http://bl.ocks.org/mhkeller/f41cceac3e7ed969eaeb

Animation: http://www.tnoda.com/blog/2014-04-02

Slider in d3: https://bl.ocks.org/mbostock/6452972

Weather icons: https://darkskyapp.github.io/skycons/

D3 legend: http://d3-legend.susielu.com/