CS 3300 Project 2 Written Submission
Spring 2016

Ayah Dweik asd95
Matthew Teng mgt42
Charles Yu ccy27

Ayah assisted in creating cancellation graphic, and average airfares, styling the visualizations, and completed the written submission. Charles loaded the data into our index.html file with d3 and created necessary data objects to be used for the assignment. He created python scripts to filter the on time performance data. Additionally, he plotted the airports onto the map by coordinate location, created flight time's visualization, created cancellation graphic and delay graphic, and created clickable flight buttons. Finally he wrote the input function to map those buttons to input (and disable airport buttons that couldn't be reached by certain inputs) and some of the flight route input. Matthew incorporated custom css and a css framework (materialize css) to touch up on the interface of our visualization. The css framework was mainly used for to allow us to focus on presenting the actual data rather than focusing the UI too heavily. He also created interactivity to change the quarter according to the season selected, and other functional buttons. And finally, Matt wrote several bash scripts to compile data from our ontime.csv to figure out every airports possible flight paths, i.e., which airports each airport cannot fly to, and compiled that into a dictionary to be used for a validity check when we allow the user to select two paths for a flight.

The data used in our visualization was collected from datasets available on various websites. The airport coordinate data was found here; the flight on time performance data for 2015 was found here; the airfare data for 2015 was found here. We compiled the data about airline flight delays, cancellations, flight duration, and average airfare for routes from two of the data sets, and then the airport coordinates from the first data set. The variables are specific to airlines that fly between an origin and destination, so that the user is able to compare airlines. The datasets were incredibly large, so we decided to display information from routes between the top 30 busiest US airports. It was an important decision to write python scripts to filter the data such that only the top 30 busiest airports are used for on time performance. We also reformatted and consolidated flights in the filtered data by classifying flights into groups with the same unique origin, destination, quarter, and airline variables. Then, we found the averages of flight time, arrival delay, departure delay, and cancellation rate for each of these groups. Selectively choosing this subset allowed for clearer objectives and usability. Our data was broken down into four quarters, so we decided display our data through the lens of winter, spring, summer, and fall. In addition, we used the us.json file to create our map.

For the flight times visualization, we used a time scale to depict differences in airline flight durations. To do this, we found the maximum flight time for each set of flight paths, and we normalized this to a eight-second animation time. A flight animation other than the flight with the maximum time has an animation duration proportional to its duration over the maximum flight duration.

For the delays visualization, we aimed to show which airlines had the best on time arrival performance in comparison to its competitors that fly the same path in the same season. We achieved this with a linear scale based on the absolute value of the delay (as it could have negative delay if it arrives early) to show visual differences in delays.

The cancellation visualization aimed to give a quick-glance understanding of how frequently flights on a specific path are are cancelled during specific season. The graphic we generated displays 1000 arbitrary flights that an airline might fly for a given path and quarter, and based on the cancellation rate for that path, airline, and quarter, our graphic colors random flights that might be cancelled.

Finally, our project provides a clear evaluation of each airline based on their average airfares, likelihood of cancellation and delays, and their flight times. One clear trend we noticed was in the comparison between the path selected and season selected. The visualization showed that flight path metrics, delays and cancellations, increased as weather became more inclement (during the fall and reaching its peak in winter).. So for example, if you are flying in the northeast during the winter, you might expect longer flight duration, or increased likelihood for delay or cancellation. This is not surprising, but the accuracy definitely is! Our data visualization can also help flyers make informed decisions about the best times to fly, the best places to fly to, and which airlines to choose when flying a specific route during a

specific time of year (based on their own personal preferences for which flight metrics are most important — flight time, delays, cancellations, or fares).

Credits to:

1. http://materializecss.com/
    a. For its button, card, and icon classes.
2. http://stackoverflow.com/questions/13712697/set-background-color-in-hex
    a. For an rgb to hex conversion function.
3. http://stackoverflow.com/questions/13893127/how-to-draw-a-path-smoothly-from-start-point-to-end-point-in-d3-js
    a. For inspiration on drawing paths in various parts of our visualization.