CS 171 Data Visualization
Final Project Proposal

Wenshuai Ye[1], Danyang Chen[1], and Wenwan Yang[1]

[1]Department of Computational Science and Engineering

March 28, 2015

1 Background and Motivation

With the significant increase of flights each year, more and more passengers get stuck at the airport. According the investigation in 2014, the 6 million domestic flights in US required an extra 80 million minutes to arrive at their destinations.

2 Project Objectives

The first goal of this project is to find the fastest airline on any particular route. For example, after the calculation of past years' data, the interactive may tell the passenger the average time that Delta took from New York to Seattle is 20 minutes less than the United. The second goal is to find the best and worst performing airports. For instance, the result shows that all flights flying out of New York had longer delay than flights from Joplin Regional(JLN). Taking this into consideration, we found tat which airlines are fastest relative to the distance they travel and the airports they fly into and out of.

3 Data

3.1 Data Processing

This data comes from the Bureau of Transportaion statistics(BTS).The data contain every flight flown by a major airline within the US. There are seven datasets included in this project. The first one is the ontime dataset which includes the origin_airport_id,origin_airport_seq_id, origin_city_market_id, dest_airport_id, dest_airport_seq_id, best_city_market_id. The other datasets are individual files with origin_airport_id, origin_airport_seq_id, origin_city_market_id, dest_airport_id, dest_airport_seq_id, best_city_market_id. The Airport ID is an identification number assigned by US department of Transportation to identify a unique airport.The OriginAirportSeqID is an identification number to identify a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time. OriginCityMarketID is an identification number to idenfity a city market. Use this field to consolidate airports serving the same city market. Similar for DestiAirportID, DestAirportSeqID and DestCityMarketID.

3.2 Data Visualization

The first data visualization shows the US map and each airport. By clicking on the airport, Details about this airport will appear on the right which contains the information about the average delay of this airport, the airport that has connection with this airport and the flight volume of this airport. We will build a search system that allows interactive selection of airport. For the selected airport, the visualization will show the information about the airport. A barchart will be implemented in the second visualisation for the delay time distribution of each airport. The third visualisation will be a vertical area chart that shows the distribution of the time it shaved off a typical flight relative to other airlines over a given period. a table ranked by the time if shaved off a typical flight relative to other airlines over a given period. Event handler and the aggregation function will be integrated with this visualization so that it could adapt to changes in the brush selection.

4 Features

4.1 Must-Have Features

There data visualizations described above are features that must be implemented in this project. We would consider our project as a failure without these features.

4.2 Optional Features

In the second data visualization, users have the option to visualize the data by different states.For the third visualization, we consider to add the data about the economic status of each airline and visualize the relationship between the economic status and the average delayed time.

5 Project Schedule

April 3rd: Project Proposal Due
April 4th to April 10th: Data cleaning and Processing
April 11th to April 16th: First Data Visualization
April 17th: Milestone 1 Due
April 20th to April 26th: Project Reviews with TFS
April 18th to April 25th: Second and Third Data Visualization
April 26th to April 30th: Optional Features Visualization
May 1st to May 4th:Project Website and Screen-Cast
May 5th: Final Project Due
May 7th: Final Project Presentation