Visualization of Domestic Flight Performance in the United States
Final Project Proposal

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

[1]Department of Computational Science and Engineering

April 4, 2015

1 Background and Motivation

With the significant increase in the number of flights each year, more and more passengers get stuck at the airport due to delay. According to the investigation conducted in 2014, the aggregate minutes of delay in reaching final destinations amount to over 80 millions. Inspired by this situation, we desire to identify airports of best and worst performance associated with certain routes and airlines.

2 Project Objectives

Our first goal of this project is to find the most on-time airlines on any particular route. For example, after the calculation of past years' data, the interaction may tell the passenger the average time it takes for Delta Airline from New York to Seattle turns out 20 minutes less than United Airline. Our 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 can identify which airlines are fastest relative to the distance they travel and the airports they fly into and out of. Our goals are consistent with the principles of data visualization because visualization allows us to gain better intuitions on locations and routes than plain texts.

3 Data

3.1 Data Processing

This data comes from the Bureau of Transportaion statistics(BTS)[1]. The data contains every flight flown by a major carrier located in the US. Seven datasets are included in this project. The fundamental one is the *ontime dataset* that contains the *origin_airport_id*, *origin_city_market_id*, *dest_airport_id*, *dest_city_market_id*, etc. The other datasets are individual files with additional detailed information that tells you the departure delay, arrival delay, taxi time, etc. Notice that *AirportID* is an identification number assigned by the 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 used to identify a city market. This field is used to consolidate airports serving the same city market. Other keys such as *DestiAirportID*, *DestAirportSeqID* and *DestCityMarketID* have similar functions. The features included in these datasets are so rich that we will not explain them in detail.

3.2 Data Visualization

The first data visualization shows the US map with each dot representing an airport. By clicking on the airport, details on this airport will appear on the right which contains the information about the airport name, average delay hours and flight volume. The bar chart on the right of the US map shows the delay time of each airport. Users can choose to sort it by delay hours, flight volumes or airport names. Adding a time slider will not be that challenging. However, smooth transitions are hard to achieve and therefore,

[1]http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236DB_Short_Name=On-Time

this will be an optional feature. The second visualisation is a line chart of major airlines' average delay time, followed by the optional feature that specifies the reasons for each delay. Besides, we will have a search function to search for the information on the departure airports and arrival airports based on users' inputs. Event handler and the aggregation function will be integrated with this visualization so that it can adapt to changes in the brush selection.

4 Features

4.1 Must-Have Features

a **US map with each dot representing an airport**
When the user moves mouse over or brushes the airport(s), flight routes connected to the selected airport(s) will be shown. When the user clicks on one airport, the information on that airport such as name, flight volume and average delay hours will be displayed.

b **Bar chart of each airport's average delay time**
The user can sort it by airport names, delay hours, and flight volumes. When an airport is selected, the corresponding bar will be highlighted. When the user click on one bar, it will have the same effect as selecting that airport on the map.

c **Line chart of major airline's average delay time**
When the user moves mouse over these lines, the one with the mouse over it will be highlighted. The line chart will be integrated with the map and the bar chart to adapt changes in the selection of airports.

4.2 Optional Features

a Show the information of that specific route when clicking on any particular route.

b Implement search function to search for airports.

c Select data of different years.

d Show bar chart informing the cause of delay.

e Visualize data by different states.

f Add the data of the economic status of each airline and visualize the relationship between the economic status and the average delay 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

6 Mockup

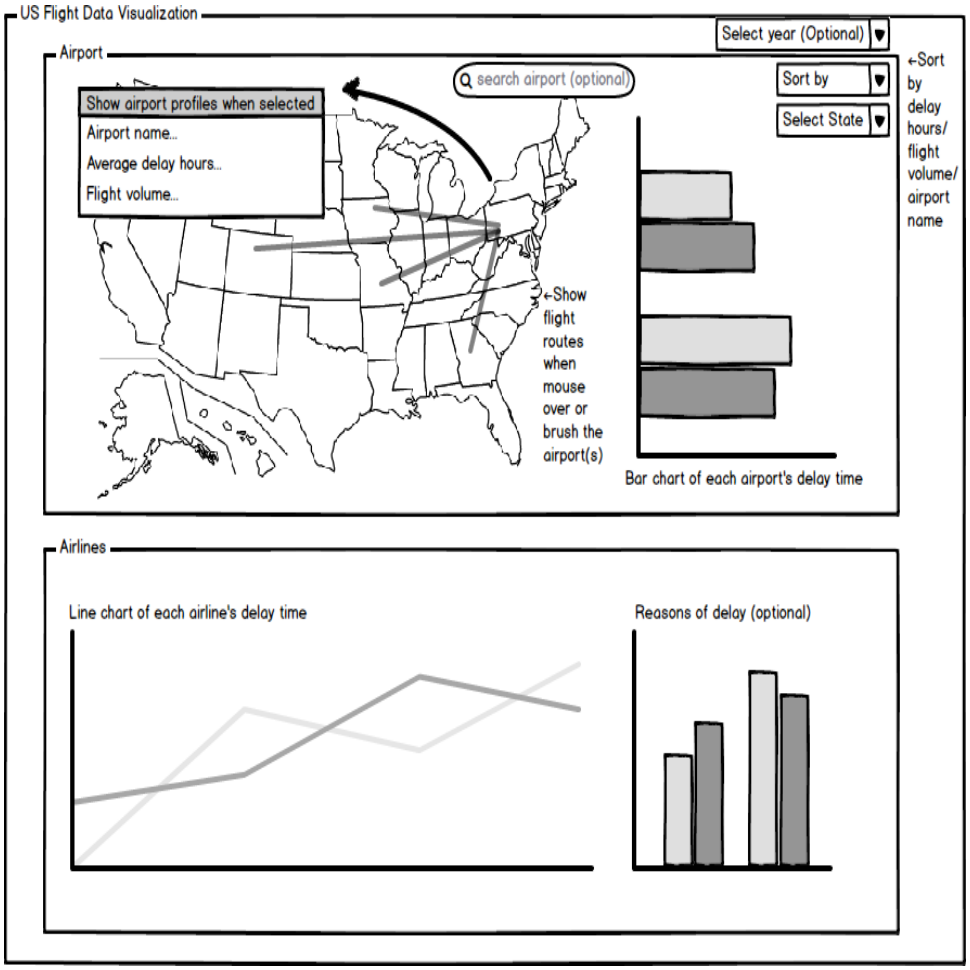

Figure 1: This is a sketch created by mockups.