Bon Voyage! A graphical display of airline destinations

Andrew Enfield

Mike Browne

Todd Schultz

# Concept

Our planned visualization is a map display that starts with points representing all major airports in the US that offer commercial airline service. Selecting a point triggers a popup box providing the name of the airport and its ICAO code, and displays connecting lines to all the destinations that are serviced by direct flights. Hovering the selector over one of the destination airports will cause a popup box to appear that contains the airport name and ICAO code, and a listing of all the airlines that service that route and their overall on-time percentages. All data will be based on on-time performance data for the 2017 calendar year data from US Department of Transportation’s Bureau of Transportation Statistics.

# Purpose

The purpose of the visualization is for users gain a better overall understanding of commercial airline services offered in the US and the on-time performance of various airline carries for specified routes.

# Intended User

The intended users are anyone interested in learning more about commercial airline service within the US. This could include anyone who is preparing to plan a trip that might require air travel. The knowledge can be used to help the user select the best airline service for themselves understanding the available routes and on-time performance of each airline servicing the route, which would be provide by this visualization, and the ticket pricing, which would be provided by any multitude of travel booking options.

# Related Work

Related work includes all travel search sites; any map web sites that offer transportation searches, such as Google Maps; sites offering airline on-time performance, such as major news carriers like USA Today and Forbes; and data-driven news carriers, like FiveThirtyEight.

# Roles

Different project roles will be assumed as needed. There will be a need for general oversight to make sure deadlines are not missed, analysis of the 2017 dataset into summary statistics, and visualization development and testing.

# Data

The data for this project is from the US Department of Transportation’s Bureau of Transportation Statistics (www.bts.gov) and will be limited to the data collected for the calendar year 2017. Limiting the data to the 2017 calendar year provides a manageable sized dataset for this project but also provides the timeliest dataset that contains a complete year of data that should capture any seasonality effects.

The data is provided in monthly files that contain the on-time data for non-stop domestic flights by major air carriers. The data includes, but is not limited to, the date, flight number, airline name and carrier code, tail number, origination airport, destination airport, scheduled departure and arrival times, the actual departure and arrival times, actual delay (in minutes), and a logical field denoting if the flight was officially considered delayed by the US Department of Transportation. We can process the data to derive a list of all airports serviced by major airlines and all the direct destinations. The files can also be processed to estimate an average percent of on-time arrivals for each airline for each departure and arrival city pair.

Even with only a single year, we'll still need to work with ~7M rows of data, as each month of data contains approximately 600K rows/flights. Depending on Tableau's performance and our specific needs, we'll pre-process the data as required.

# Anticipated Results

The anticipated result is a map with all the major US airports displayed as dots and as a user clicks an airport, the name and ICAO code are shown in a text box and lines connecting the airport to all the other airports that are serviced by direct flights. The names and ICAO codes of the other airports should be shown in pop-up text boxes as well. A stretch goal would be to have the on-time performance for all the airlines that service a route shown in a mouse-over pop-up box when the user hovers over a destination airport. Additional analysis could be performed on the on-time performance to break out more segments such as seasons, day of week, and time of day.

**Use case**

I am Ryan Bingham, a traveler seeking a direct flight from City X to City Y in the US.

For a given flight day of week, II need to decide which flight to take that:

* Has the lowest risk of departure/arrival delay
* The shortest flight duration