source: Visualizing geodetic information with JEQL, http://lin-ear-th-inking.blogspot.com/

Project Proposal

Prepared for: CS171 Final Project

Prepared by: Masahiro Kusunoki, Nhu Nguyen, Zeno Ziemke

GitHub Repository: https://github.com/nhunhu9/cs171-pr-flightdata

April 3, 2015

OVERVIEW

Background and Motivation

In 2010, the number of commercial departures topped 30 million annually. This number calculates to over 80,000 departures per day, and with the number ever increasing the figure is expected to have reached double that amount by 2030[1]. This is an incredibly resourceful dataset considering that each one of these flights have been carefully recorded and documented. Hence we felt that there would be great potential in visualizing this data.

Trends in the frequencies of commercial flights are also linked to other important factors such as economic standing, cultural attraction or political affinity of two countries. They may also be linked to specific events (such as the FIFA World Cup or the Olympic Games) that would bring new trends across particular a time span. We would be interested in seeing whether we can find other factors that correlate with the phenomenon of global transportation and if so, whether we will be able to predict new trends off of different variables in the future.

However, perhaps most importantly, what drew our attention to this topic the most is the fact that we are all interested in traveling and flying and would like to learn more about how global transportation has evolved to look like what we currently experience today.

Primary Objectives

The primary objective of this project is to visualize how different cities in the world have been connected over time. What have been the global trends for the establishment of flight routes and which routes have been the most (or least) popular? To answer these questions, we will plot the data of all the commercial flight routes from “OpenFlights” database, and map these routes onto a map of the globe. If we add a feature for filtering depending on the year, then this can show us over time how new flight routes have been established (or terminated) and how the world has become more connected through commercial flights over time.

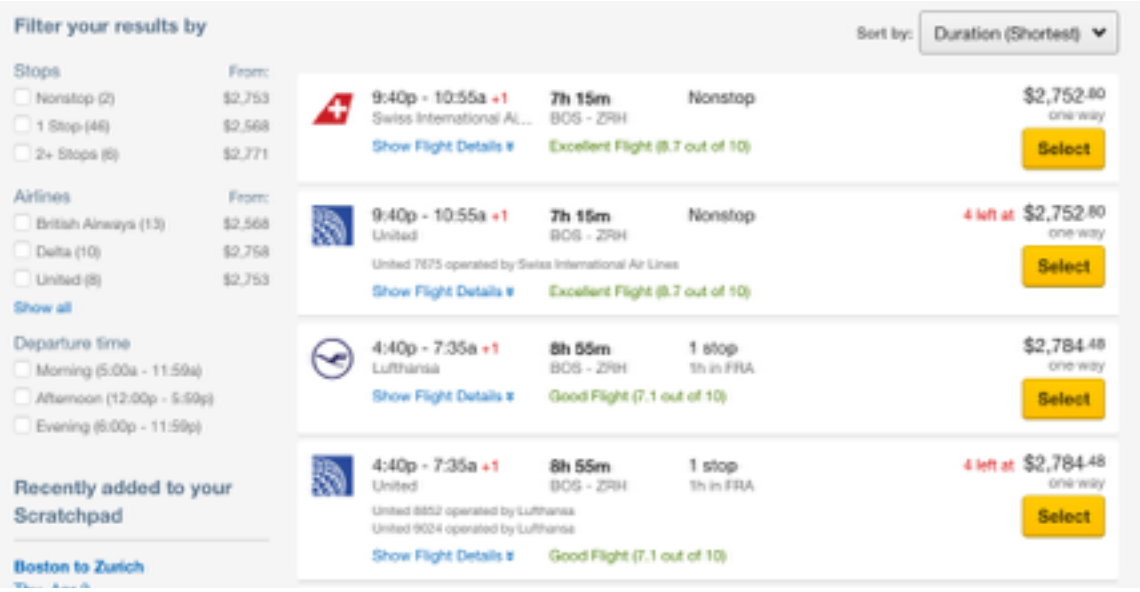
Other questions we may consider are:

- Are there any correlations with other variables (e.g. GDP, population)?
- What are the prices of the different routes at different times of the year?

The first question is beneficial because it will allow us to understand the phenomenon of global transportation fly in more depth. We can imagine with relative ease that the countries’ population would be correlated to the number of routes that exist to the country, but perhaps there may be exceptions to this rule and if so, we can explore other variables that explain this.

FLIGHT VISUALIZATION

The second question is beneficial from a more practical standpoint. When we try to search online for our own flights to book, currently, many of the search websites provide a very static and stale looking response. These search results could be improved greatly if we did not simply represent each option independently but rather compared these different options to each other adding in visualization for factors such as geographic locations of transits, times of departure and daytime/nighttime analysis. This would be an optional feature for the project, but hopefully we can integrate Google's flight search API to return some flight price comparisons integrated into the visualization.



source: Search results from http://www.expedia.com/

source: Search results from https://www.hipmunk.com/

DATA

OpenFlights Database

Link: http://openflights.org/data.html

This database is in csv format and contains hand collected information about flight routes, airports and airlines. It contains 58000 flight routes with the following fields:

> airport departure | city departure | country departure | long. departure (decimal) | lat. departure (decimal) | airport arrival | city arrival | long. arrival (decimal) | lat. arrival (decimal) | airline name | airline country based | route ID | number of stops | distances (km.) | domestic

Furthermore it contains information about the passenger volumes between cities (3000 routes) with the following fields:

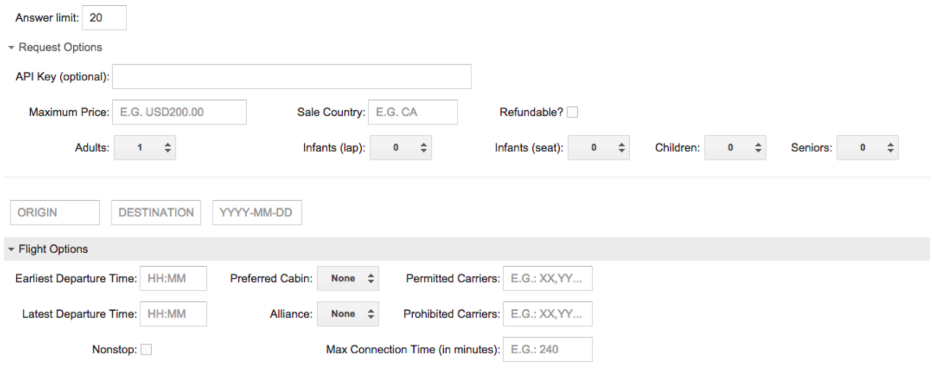
> departure city name | arriving city name | number passengers 1990 | number passengers 1995 | number passengers 2000 | number passengers 2005 | number passengers 2010

The passenger volumes between cities are available for the years 1990, 1995, 2000, 2005 and 2010. The other data is not available over time.

Flight Fares from QPX Express API/ (50 free requests/day)

Link: https://developers.google.com/qpx-express

This Google API provides flight fares. The JSON based API offers the following options for searching flight rates:



It returns routes including up-to-date fare information in JSON format. The API is restricted to 50 free requests/day. Additional requests costs $0.035 US per request. We expect that the 50 free requests/day will be sufficient so that at least 1-2 users can try out our visualization on the demo day.

General Economic Data

Link: http://data.worldbank.org/

This dataset contains detailed economic data like GDP, Import/Export, Currency Rates, etc. for all major countries of the world. The data is available on yearly and quarterly basis.

Airplane Crashes and Fatalities Since 1908

Link: https://opendata.socrata.com/Government/Airplane-Crashes-and-Fatalities-Since-1908/q2te-8cvq

This database is available in csv format and contains airplane fatalities from 1908 to 2009 with the following fields (~5000 incidents):

departure airport, departure city, departure country, long. departure (decimal), lat. departure (decimal), arrival airport, arrival city, arrival country, long. departure (decimal), lat. departure (decimal), airline name, airline country, active airline, distance between airports (km),domestic flight

DATA PROCESSING

Most of our information is available in clean CSV formats. We expect to do some basic format conversions. For example converting date formats or number formats for all datasets.

To use the flight fare API we will need to integrate the API calls into our application.

For the General Economic Data and Openflights database, we only expect to filter and aggregate data depending on the calculations and visualizations we intend to perform.

The Airplane Crashes and Fatalities will require additional cleanup:

One issue is that this dataset does not provide geolocation data for the crash locations. If we have time we will try to automatically get the longitude/latitude for the locations of the crashes (they are saved in text format in the data set, for example: “Fort Myer, Virginia”) using the Google Maps API. This would allow us to plot the crashes on a world map.

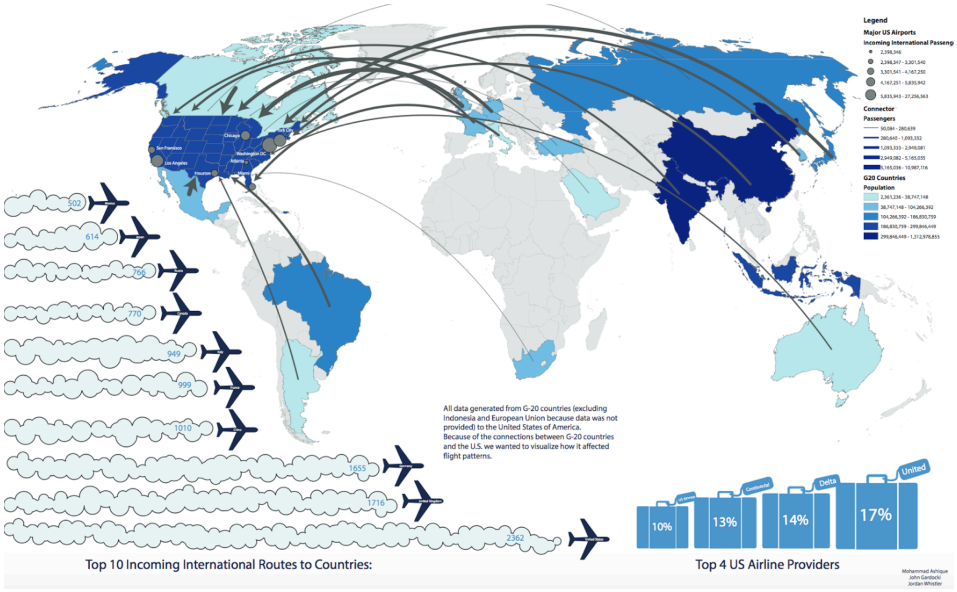
Furthermore this dataset contains incidents for all kind of incidents including military plane incidents. Thus we will need to filter out the commercial airline incidents.

VISUALIZATION

Linked geographic map:

We want to use a geographic map to let the user see an overview of the connectivity of the world and how it changes over time. A linked geographic map will use links to represents the flight routes between cities. Links that are in the same category (for example, routes from Europe to Asia and vice versa) would have the same color. This way, we can clearly which cities are more connected to each other based on the density of the colored links. Below are two examples of linked geographic map:

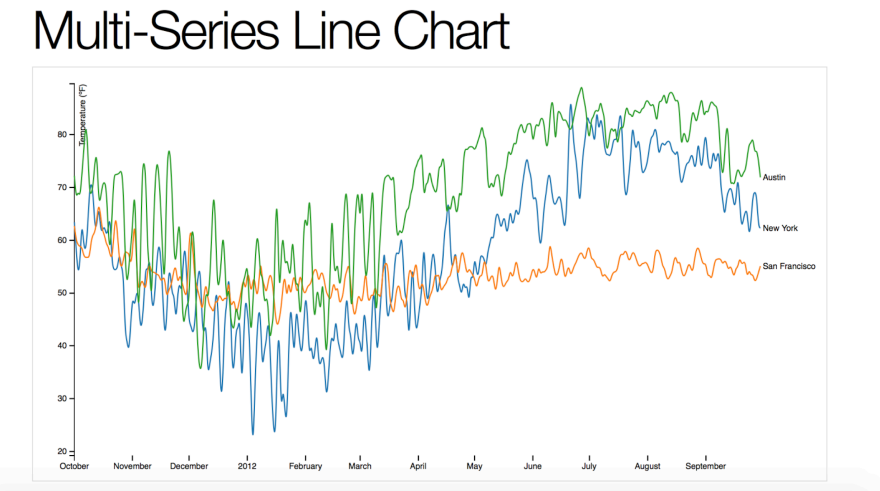
source: Visualization Analysis and Resource,
http://spatialanalysis.co.uk/2012/06/mapping-worlds-biggest-airlines/



source: International Passengers Traveling to the United States of America,
http://www.visualizing.org/visualizations/international-passengers-traveling-united-states-america

Multi-series line chart

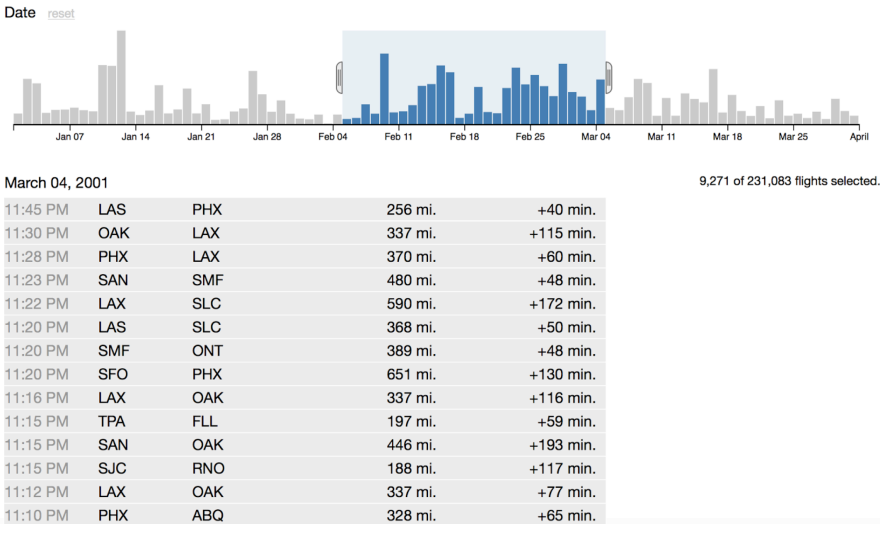
A multi-series line chart can be used to show the correlation of the changes in the number of commercial flights with the changes in variables such as population, GDP, seasons, special events, etc. The benefit of using a line graph is that multiple lines can be plotted on the same graph, but doesn't make the information presented more confusing. An example of a line graph is below:



source: Multi-Series Line Chart, Mike Bostock
http://bl.ocks.org/mbostock/3884955

Bar chart with brush:

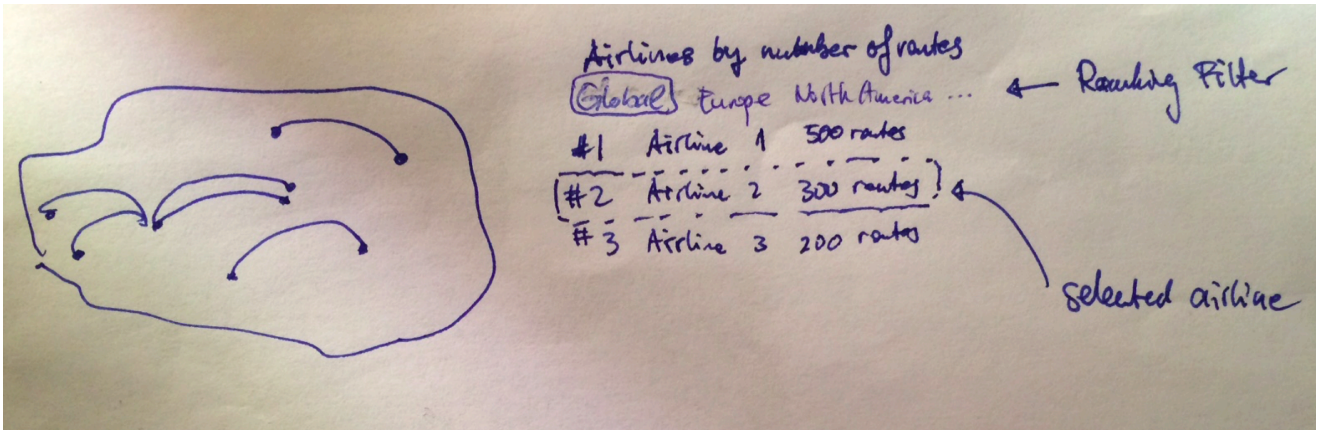
Using a bar chart with brush, we can see the total number of people using commercial flights for each time unit, and when the user select a time frame using a brush, a table will appears which shows specific routes and the average number of people using flying on that route during that time period. The user can slide the brush to see how the average number of people flying changes over time for all routes. An example of a bar chart with brush and a table is on the right:



source: Crossfilter, Fast Multidimensional Filtering for Coordinated
http://square.github.io/crossfilter/

Airline Ranking:

Airlines will be ranked in a table by the number of routes they offer. It is possible to filter the ranking by continent. Clicking on an airline in the ranking will highlight the routes offered by that airline on the world map.



source: Nhu Nguyen, 2015

FEATURES

Must-Have Features

- Feature for mapping the routes of flights onto a global map
- Filtering for airline, for route, or for continent/county selection
- Comparing the number of people flying on a specific route across time

Optional Features

- Visualizations that can improve booking process: (displaying/filtering) the (cheapest/direct/shortest) flights
- Show prices of flights departing from that airport on a specific date on click of a country/airport
- Visualize relationships between economic data (GDP, buying power etc.) and passenger numbers/number of flight routes between countries/continents
- Visualization for relationships between major international events (FIFA World Cup, Olympics) and the number of passengers flying/number of routes offered to the country
- Visualization for airplane crashes in the last century

PROJECT SCHEDULE.

Week 0: Proposal Submission (April 3rd) (8%)

- Sketches of visualizations
- Clean up/convert data in JSON
- Divide tasks for each member

Week 1: Milestone 0 (April 10th)

- Design prototype of visualization
- Begin map of flight routes visualization
- Experiment with Google's flight search API

Week 2: Milestone 1 (April 17th) (15%)

- Complete map of flight routes
- Complete filtering, selection functions for map
- Update process book

Week 3: Milestone 2 (April 24th)

- Begin optional features
- Implement flight route and GDP relationship chart
- Implement flight price search and display feature

Week 4: Milestone 3 (May 1st)

- Complete flight price search visualization
- Complete GDP relationship chart
- Complete screencast
- Complete process book

Week 5: Project Deadline (May 5th) (35%)

- Confirm that all features are functional and fully interacting
- Take a break, good job!

May 7th: Project Presentation

- Be Awesome and kick butt