Chong Ran (cr448)
Hao Xue (hx73)
Zichuan Zhou (zz278)

Where Do Americans Fly?

Mapping Non-Stop Air Routes Among Largest 30 U.S. Cities and Number of Passengers on Each Route

The image visualizes all the routes of non-stop flights among the largest 30 cities in the United States as well as the number of passengers on each route from January to November 2014.

Data Description

Map:

The map uses d3.geo.albersUsa projection. Coordinates of cities are from Wikipedia.

Population:

The population statistics are obtained from Wikipedia page of "List of United States cities by population". We use population estimates in 2013 as our data. Note that when we choose 30 largest U.S. cities, we skipped Fort Worth (17th) because it shares the same airport system with Dallas. Milwaukee, which ranks 31st in the country, is substituted for the spot.

We divide the population data by 10 thousands, round the numbers to nearest integers, and store them in an array.

Routes and Air Traffic:

We obtained our flight data from disclosed statistics provided by Department of Transportation Bureau of Transportation Statistics (http://www.transtats.bts.gov). The statistics include data for all public air traffic in the United States from January 2014 to November 2014. The excel table contains more than two million rows.

We only keep the column for passenger traffic data, deleting freight traffic column.

We use excel tools to filter destinations and origins, and sum over relevant entries to get total passenger traffic for a specific pair of origin and destination. We reformat the data in following ways:

1. We include Newark Liberty International Airport (EWR) in traffic calculation of New York City. Despite its location in Newark, the airport is one of the three major airports that serve New York and is United Airline's hub for the area. If we do not include Newark airport, New York's data would be incomplete.

2. We include Oakland International Airport in traffic calculation of San Francisco. Oakland airport mainly serves San Francisco and carries a lot of traffic for Southwest Airlines. We exclude San Jose International Airport. Although San Jose is often considered closely connected to San Francisco, it actually has a larger population than San Francisco and comprises a separate metropolitan area according to U.S. Census Bureau.

3. If there is no scheduled direct flight (there might be chartered flights, though) between two cities, we record the traffic as 0.

There are some other considerations in processing data:

1. We only take into account Los Angeles International Airport (LAX) for Los Angeles. Although LA area has several other airports in operation, these airports are much smaller in terms of traffic. Also, it is hard to determine which of these airports should be included.

2. Although Atlanta and Miami are important hubs and Atlanta airport is indeed the world's busiest airport, they are not included in the project since they both rank behind 40th on the city population list. They both comprise much larger metropolitan areas. It is worth noting that we did not use top 30 metropolitan areas in this project, which might seem more relevant, because a metropolitan area is likely to cover several airports. Given the limited time, we could not be able to accurately collect data for metropolitan areas. We hope to do a similar project for metro areas in future.

We then store the processed data in a 30*30 array.

Data Mapping

1. We map cities based on their coordinates. For example, projection([-73.9385, 40.6643]) returns the projection of New York City.

2. We wrote code of drawing route lines before drawing city dots, so that the routes are behind cities on map.

3. A nested for-loop is implemented to draw lines that represent routes and traffic. We used some data structures built with arrays to make sure we draw the lines starting from cities with least passenger traffic at first. In this way, the lines from the major traffic hubs, which are more likely to have heavy-traffic routes, will not be blocked by less important lines.

4. We split traffic data into 5 classes. Each class is represented by a type of line. Routes with heavier traffic are drawn with darker and thicker lines so that they are more visibly prominent. We chose 5 colors on a scale from yellow to green to dark red to correspond to the 5 data classes. We make sure these lines' darkness on a grayscale increases with the amount of traffic represented.

5. The population of cities are represented by the area of the dots. Since the range of population of the 30 cities is large (New York: Milwaukee = 841: 60) , we can't map the data linearly. We decide use logarithmic scale on the data.

Visualization Result

From the visualization, we are able to find some interesting facts.

1. Although some cities are large in population, they are less connected to other major cities. For example, Jacksonville has a larger population than San Francisco (thanks to its vast area), but it does not have any route with more than 500,000 passengers. Oklahoma City, Memphis, Louisville, El Paso are all characteristic of this trait. Explanations might be: 1. the reason these cities have a large population is due to their larger city area. The population density is actually lower than other metropolitan areas. 2. All of these cities are located in the South. They are not as developed as cities on East / West Coast or in Midwest.

2. Only five cities have scheduled non-stop flights to all other 29 cities. These cities are Chicago, Houston, Dallas, Denver, and Las Vegas. All of these cities have airports that are major hubs of major airlines. Some of them are not in top 20 population, but are important transportation hubs (for transfer passengers). It is interesting to notice that none of the cities is on East or West coast. One possible explanation is cities in the middle have easier access to other cities.

3. Generally, largest cities have largest traffic. The cities with most routes with more than 2 million passengers are New York (7), Chicago(6), Los Angeles(5). These cities are largest sources of passengers.

4. Chicago is the most important domestic hub in the United States. Even though New York has the busiest airport system in the country, New York gains the title due to its international gateway status. Domestically, As a major hub for three of the largest airlines in America (United, American, Southwest), Chicago carries more traffic. Actually, Chicago-New York is the country's busiest route.