# Flight Delays and Cancellations

## INTRODUCTION

This project aims to reveal insights from a data set of flight delays and cancellations which comes from a Kaggle dataset. Only the performance of US domestic flights operated by large carriers in 2015 were included.

There are three visualizations in the report and they were used to answer the following questions:

1. Which airports are the busiest in terms of origin flights? What was the busiest time of the year? Which airlines travelled the longest? Was there a difference between weekdays and weekends?

2. Which airlines and airports are the worst in terms of delays? If they were delayed, what was their average time of delay? What was the rate of flight cancellation per airline?

3. What was the number one cause of delay per airline? Do the number of delays vary by the time of the year? If I wanted to travel from one city to another, which airline should I choose to fly me there the fastest?

The story is hosted in Tableau Public under the following link:

https://public.tableau.com/views/FlightDelays_20/Story1?:embed=y&:display_count=yes&publish=yes

# FLIGHT COUNT

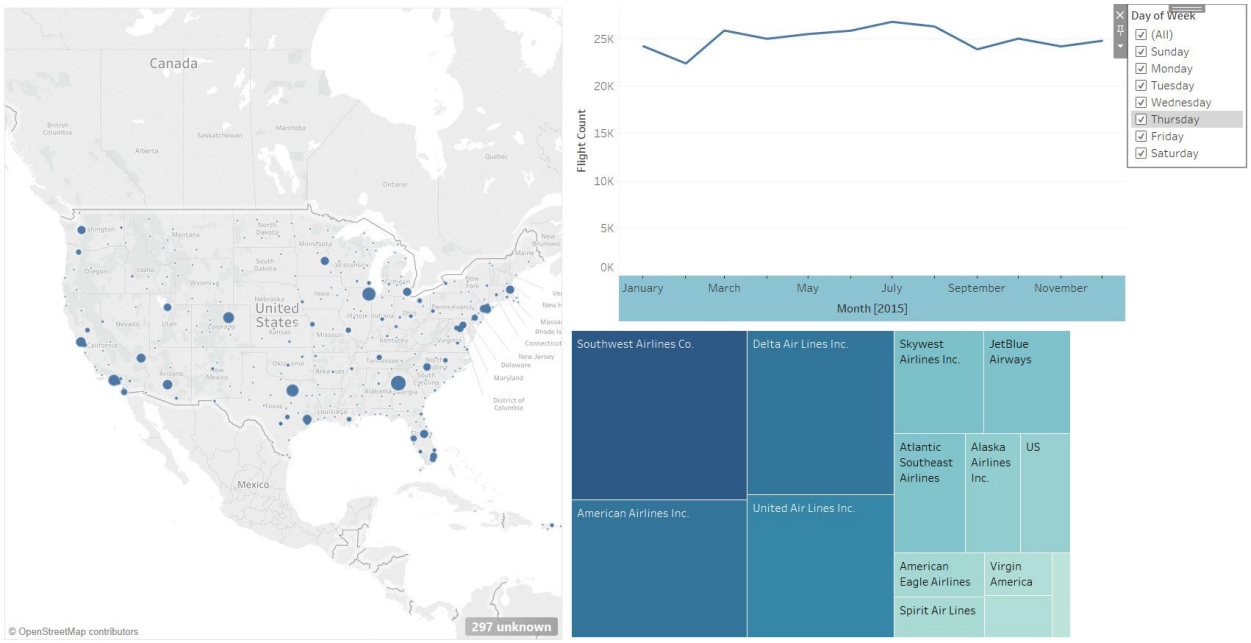


*Figure 1. First Dashboard Showing Flight Counts*

First, I wanted to examine which airport was the busiest in terms of the count of origin flights. A column named Flight Count was added in the Data Source tab with a value of 1. There are 300,000 flights in the dataset so summing the Flight Count should result in the same number. Then, the IATA code for the origin was renamed as Origin Code and Tableau recognized the geographic data assigning a geographic role in that field. The airports are reflected in the Tableau map server which is shown in the left side of the first dashboard. The size of the blue circles represents the number of flights. The busiest airport is Hartsfield-Jackson Atlanta International Airport (ATL) with 18,056 departing flights. This is followed by Chicago O'Hare International Airport followed by Dallas/Forth Worth International Airport with 14,684 and 12,281 flights, respectively. Airport Council International (ACI) also lists ATL as the busiest airport but in terms of total passengers.[1]

The original data separated the days, months and years into separate columns. Combining their string data types into one column and converting them into date format yielded a more useful column for data analysis. The date dimension was then converted to continuous type and separated into months to show the line graph in the upper right-hand side. There is a relatively constant number of flights throughout the year with a lowest count in February because it has only 28 days compared to other months. One would infer that the flights would spike during the yearend holiday seasons (November/December) but that is not the case here. The highest number of flights was in July which may be attributed to the Fourth of July festivities.

Using a treemap instead of the usual column chart to compare the distance travelled per airlines yields the visual data on the lower right-hand side. The largest square belongs to Southwest Airlines which has

[1] http://www.aci-na.org/content/airport-traffic-reports

a combined 47.7 million miles of distance travelled. American Airlines Inc. and Delta Airlines Inc. follow closely with around 38+ million miles. Business Insider ranks Delta as #1 followed by Southwest and as #2 and #3, respectively as the three largest US airlines.[2] Southwest may have longer flight routes than either Delta and American and this accounts for the largest distance travelled.

A filter was added on the rightmost side of the dashboard to examine if there was a noticeable difference in the number of flights between weekdays and weekends. The original data contained numeric data which was converted into a string. The number 1 was equivalent to Monday, 2 for Tuesday, etc. This was verified against the date. As shown in Figure 2 below, the flights for weekends exhibit the same pattern across the three visuals when compared to whole data.

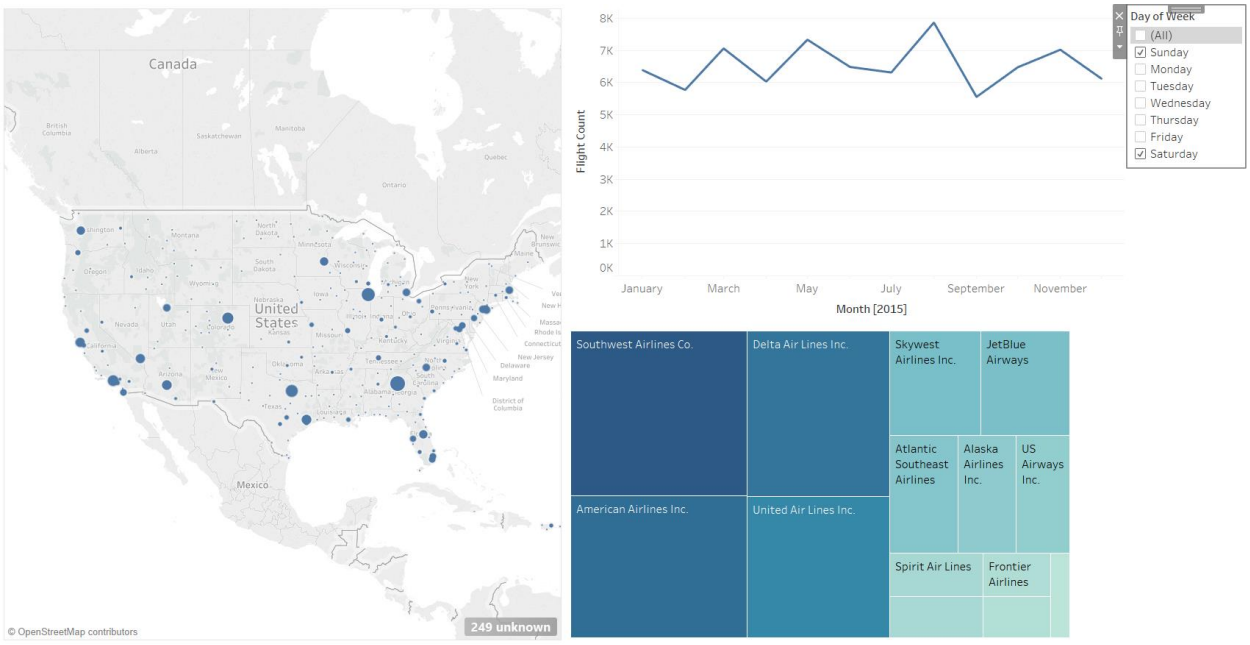


*Figure 2. Weekend Flights*

Since weekends have almost the same pattern, it can be inferred that weekdays should also behave similarly as shown in Figure 3. Finally, Thursday seems to have the most number of flights followed closely by Monday and Friday. The least popular day is Saturday. Thursday may have the most number of flights because it is the cheapest day to fly. Most travelers prefer to come back from vacation on Sunday to maximize their time away which explains the low number. The most popular days for business travelers, meanwhile, are Monday and Friday.[3]

[2] http://www.businessinsider.com/these-are-the-7-biggest-us-airlines-2016-4/#2-southwest-6
[3] https://www.smartertravel.com/2017/06/19/best-worst-days-fly/

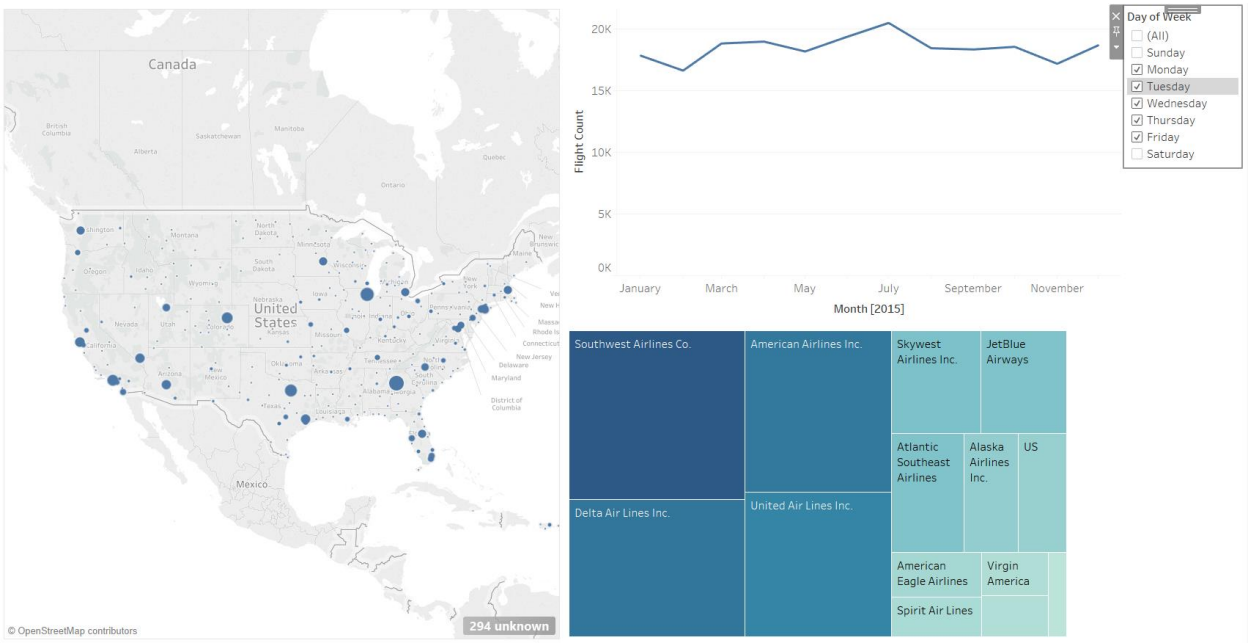


*Figure 3. Weekday Flights*

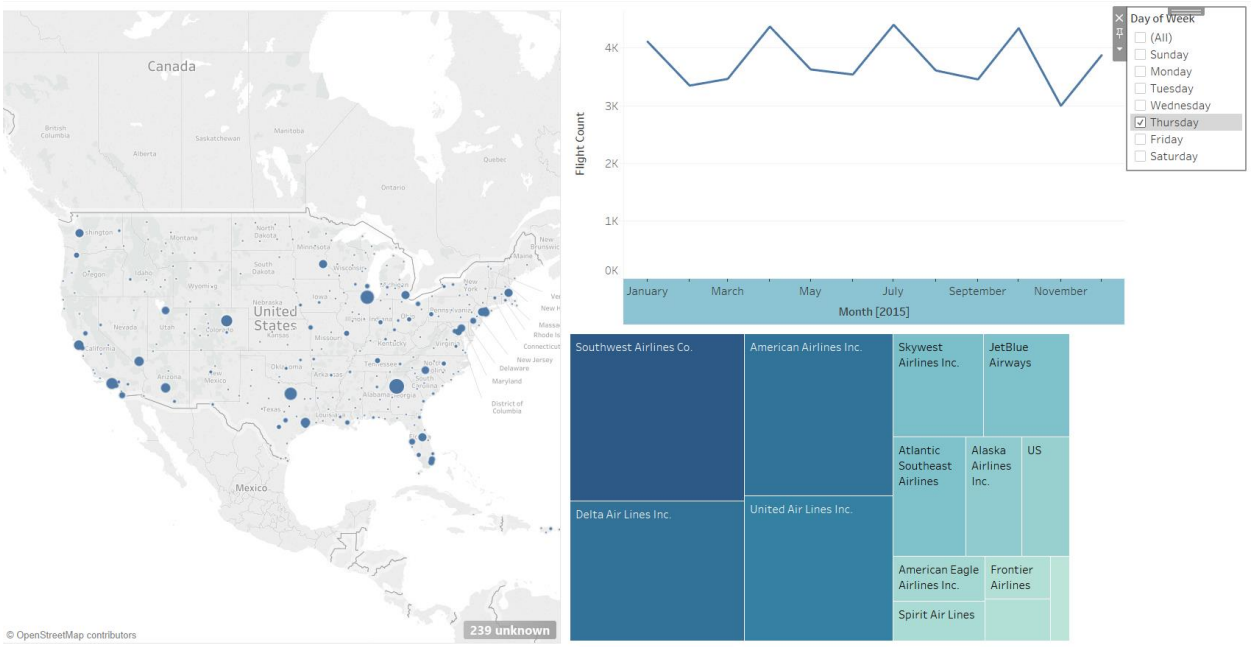


*Figure 4. Thursday Flights*

# FLIGHT CANCELLATIONS AND DELAYS

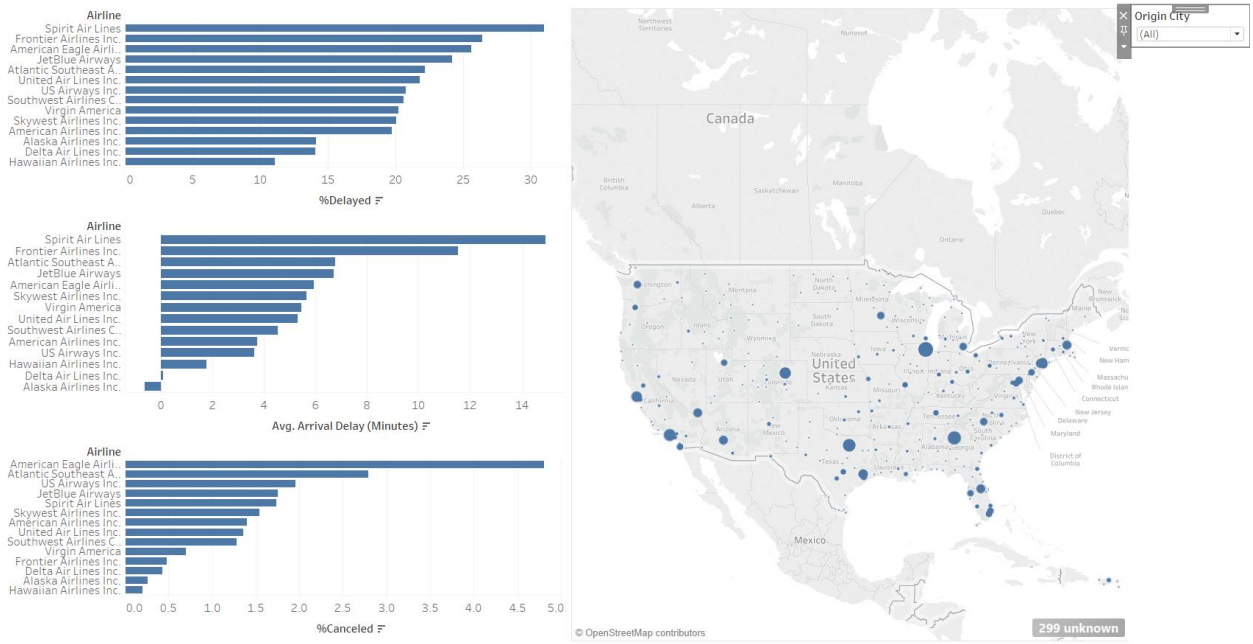


*Figure 5. Second Dashboard Showing Delays of Airlines*

Now that the flight counts have been revealed in the previous dashboard, the next logical step is to determine which airports and airlines are the worst in terms of delays and cancellations. The Federal Aviation Administration (FAA) considers a flight to be delayed when it is 15 minutes later than its scheduled time. A cancellation occurs when the airline does not operate the flight at all for a certain reason.[4] The column Arrival Delay is the difference between the two columns Arrival Time and Scheduled Arrived. Using the criteria set by FAA, a new column was created to determine if a flight can be considered as delay. A value of 1 was given when the Arrival Delay was greater than 15 minutes. Another calculated field was created to divide the sum of the delayed flights over the sum of the total flights multiplied by 100 to get the percent delay of each airline. There is already a column to indicate if a flight was cancelled (1=cancelled) and the percent cancellation was calculated for each airline. Like the first dashboard, the airports are reflected in the Tableau map server which is shown in the right side. However, the Destination Code was used instead of the Origin Code to be more consistent with the three column charts on the left.

As shown in Figure 5, the worst airline in terms of delays is Spirit Air Lines which usually gets delayed in 3 out of 10 flights. Frontier Airline Inc. and American Eagle Airlines Inc. round up the 2nd and 3rd spots. It is worth noting that Southwest, American and Delta sit in the middle of the pack which is nice for customers who travel with them often. When flights get delayed, Spirit Airlines also has the worst waiting time with an average arrival delay of around 15 minutes. Frontier has an average of 12 minutes while American Eagle has 6 minutes. In terms of cancellations, there is nearly a 1 in 20 chance that an American Eagle flight will be cancelled. It is miles ahead of its 2nd worst competitor, Atlantic Southeast Airlines which has a value of 2.8%. Spirit, Frontier, and American Eagle are in the bottom 5 in terms of flight distance which may mean fewer flights resulting to a greater impact on delay and cancellation percentage. From the

[4] https://en.wikipedia.org/wiki/Flight_cancellation_and_delay#cite_note-1

period between September 2014 and September 2015, Popular Mechanics ranked Frontier, American Eagle and United Airlines as the three worst performing airlines in terms of delays.[5] There are several reasons of the cause of delay and these will be further investigated on the 3rd dashboard.

If Spirit, Frontier, and American Eagle are the worst, then the best airlines in terms of the least amount of delays are Hawaiian Airlines Inc., Delta Airlines Inc., and Alaska Airlines Inc. Hawaiian and Alaska are small airlines in terms of distance travelled which is even more impressive. Delta's sheer amount of flights may have offset the delay percentage. The Travel and Leisure Website agrees with the findings ranking Hawaiian, Alaska and Delta as the best with high percentages of on time flights.[6]

Meanwhile, looking at the map on the right side, the airport with the most number of arrival delays is the Chicago O'Hare International Airport (ORD) with over 3,500 delays followed by Hartsfield-Jackson Atlanta International Airport (ATL) and Dallas/Forth Worth International Airport (DFW) with around 2900 and 2600 delays, respectively. Coincidentally, these three airports have the most number or flights, so they would naturally have the most delays. Fortune's report from the Bureau of Transportation Statistics show that nine out of the most delayed travel routes in 2014 involved Chicago.[7] If you have a flight from Chicago, there is a high chance that you will not get to your destination on time.

To further investigate the claim and see which airline from Chicago will most likely cause the delay, the filter on the right side is selected as Chicago being the Origin Airport. There are plenty of delayed flights going to the East Coast as shown by the large circle sizes in Figure 6. The flights responsible for most of these delays belong to Spirit, Frontier and Skywest. The worst airline and airports combo for a passenger would be riding Spirit from Chicago and going to LaGuardia Airport (Marine Air Terminal) in New York.

The filter allows the user of the dashboard to investigate other cities as points of origin to determine which airlines would be the best for flying. If you are travelling from the busiest airport in the US, your best airline bet would be Delta because it has the least chance of getting delayed (Figure 7). This of course assumes that it has an available route to your destination.

[5] http://www.popularmechanics.com/flight/g2303/airlines-with-most-delays/
[6] http://www.travelandleisure.com/slideshows/best-and-worst-airlines-for-delays-2014#no-1-hawaiian
[7] http://fortune.com/2015/06/17/delayed-flights-chicago/

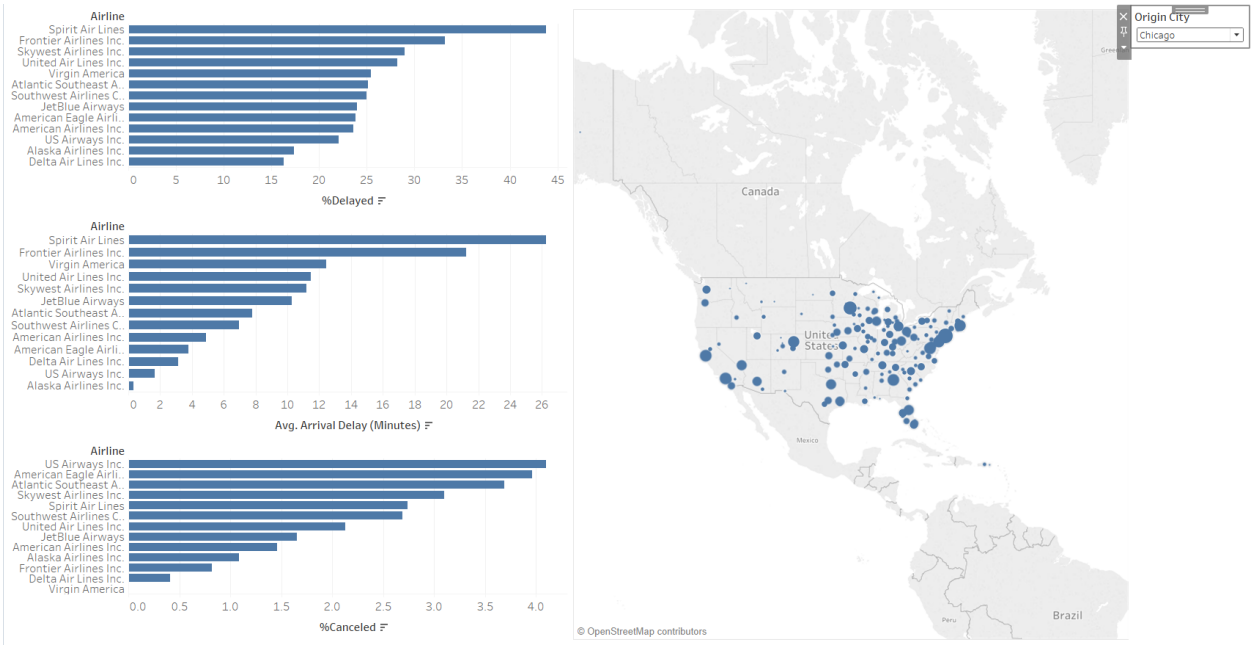


*Figure 6. Flights from Chicago O'Hare International Airport*

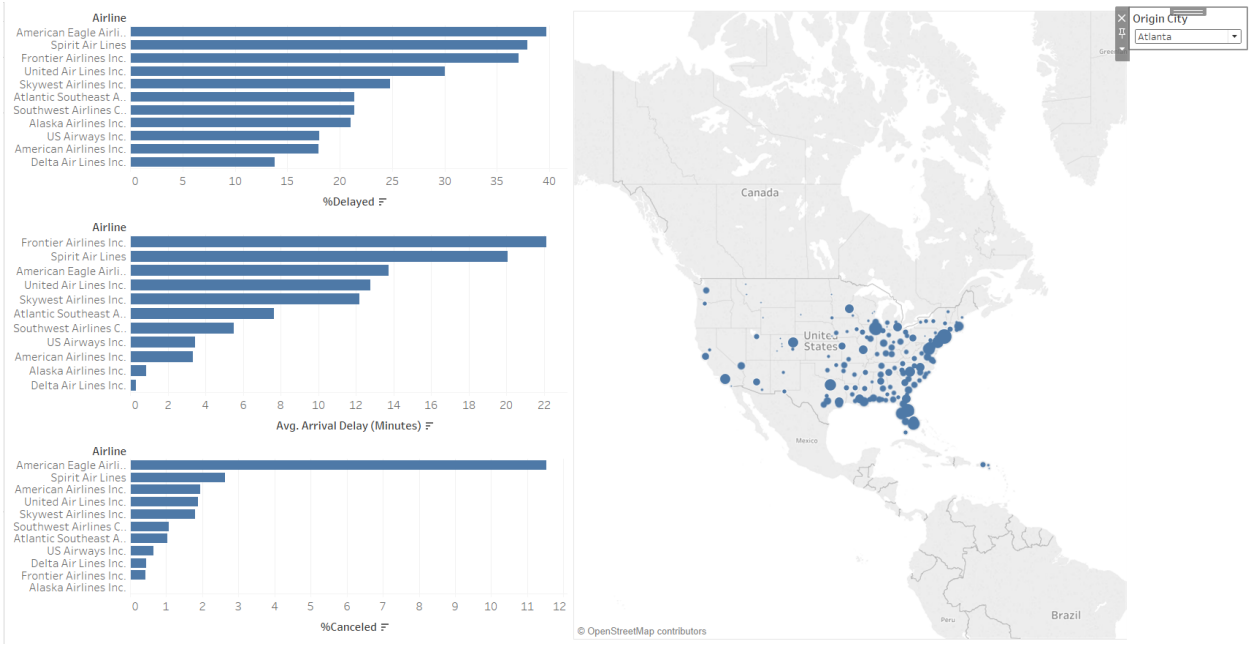


*Figure 7. Flights from Hartsfield-Jackson Atlanta International Airport*

# CAUSES OF DELAYS

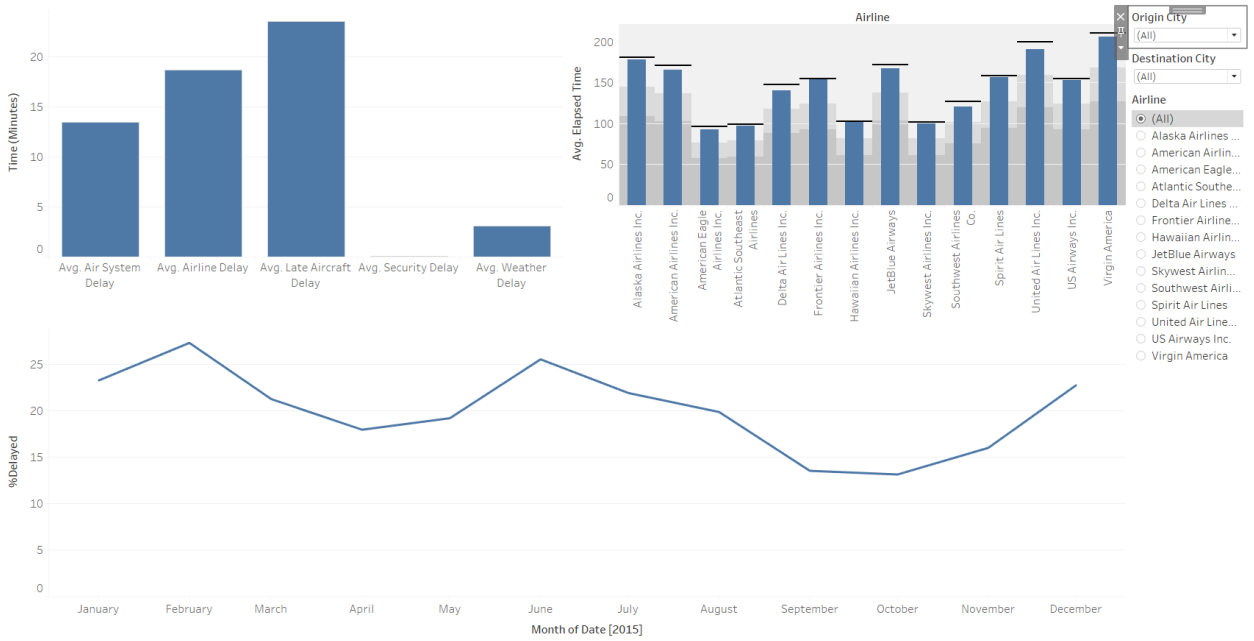


*Figure 8. Third Dashboard Showing Causes of Delays*

The data set already identified five causes of delays of airlines and airports which are (1) Air System, (2) Airline, (3) Late Aircraft, (4) Security, and (5) Weather. The Bureau of Transportation Statistics (BTS) provides a clear definition of these causes:[8]

- Air Carrier: The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
- Extreme Weather: Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
- National Aviation System (NAS): Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
- Late-arriving aircraft: A previous flight with same aircraft arrived late, causing the present flight to depart late.
- Security: Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines more than 29 minutes at screening areas.

As shown in Figure 8 above, the number one cause of delay is late aircraft delay which results to almost 24 minutes of delay followed by airline and air system delays. The data set exactly follows the trend shown by BTS with security delay accounting for nearly nothing.

[8] https://www.rita.dot.gov/bts/help/aviation/html/understanding.html

The line chart above shows that February 2015 had the highest percentage of delays followed by June, December, and January. These findings can be easily explained by the winter weather in February, summer season in June and the holiday seasons in December and January. Hovering the mouse through the chart would reveal that June had more delayed flights than February but the latter had a higher percentage due to the lower count of flights as revealed earlier. Surprisingly, November has relatively low percentage of delayed flights even though it is Thanksgiving. This may be attributed to the airline's anticipation of the high demand, so they added more personnel to support the passenger rush.

The bullet graph on the right was inspired by FiveThirtyEight's method of finding the fastest airline on a particular route.[9] The chart aims to compare an airline's scheduled flight time against the elapsed time. Scheduled time is defined as the planned time needed for the flight trip while elapsed time is the air time plus taxi in and taxi off. Taxi on and off is just the time duration elapsed between the time point that the aircraft's wheels touch/leave the ground and the gate entry/exit from airport. All airlines except Hawaiian Airlines have elapsed time less than scheduled time which means they all arrive earlier than expected. This is consistent with the delay percentage which is still a minority (topping out at 30%). Hawaiian's data may have been caused by its travel distance from its islands to the mainland. The magnitude of each airline's values cannot be compared to each other since they have different flight routes and distances. This chart will be more useful when we go deeper into the analysis.

The usefulness of the bullet graph would be harnessed using the filters on the right. If I wanted to see which flight is the fastest from one airport to another, I would be able to find out through the result of the chart. For example in Figure 9 , there are nine flights available from the worst airport (ORD) to the busiest airport (ATL). The lowest elapsed time among the airlines is Southwest Airlines Co. with an average of 106 minutes against its schedule time of 110 minutes. It guarantees the fastest flight between the two cities and it actually delivers. If travelling from ORD to ATL, choose Southwest and avoid Frontier.

Now that Southwest Airlines was the aircraft of choice from ATL to ORD, the analysis can be further drilled down to find out the % of delays in each month and their causes. If travelling on December, there would be a 1 in 2 chance that the flight would be delayed which is caused by the late arrival of the aircraft.

[9] https://fivethirtyeight.com/features/how-we-found-the-fastest-flights/

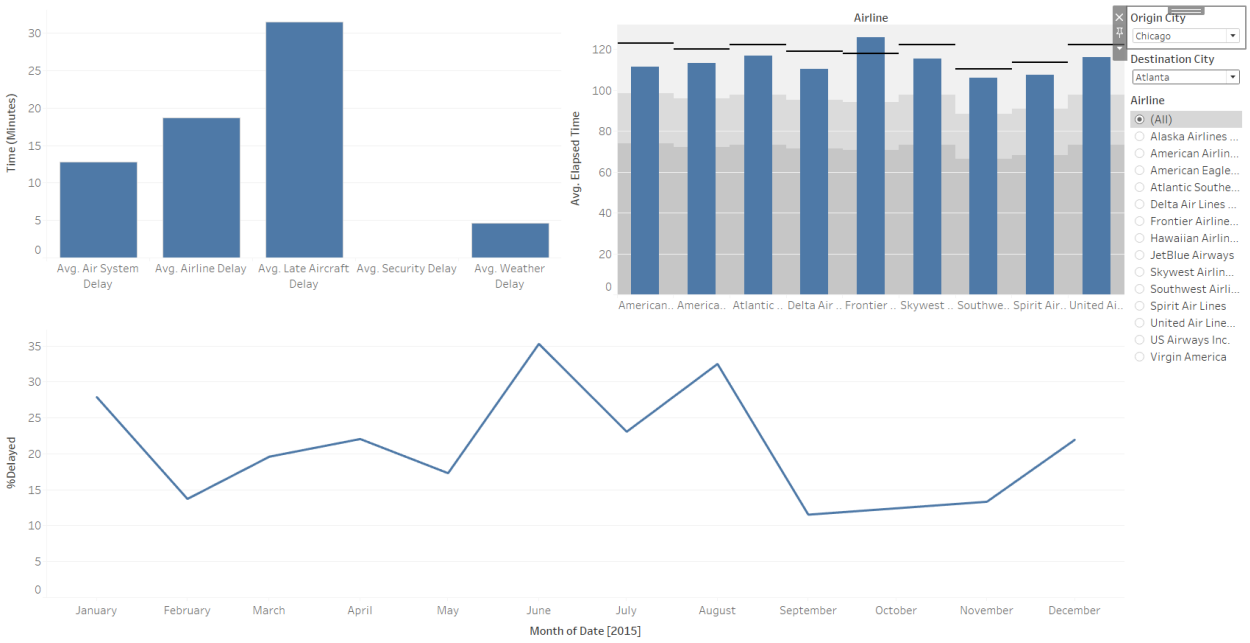


*Figure 9. Flights from Chicago to Atlanta*

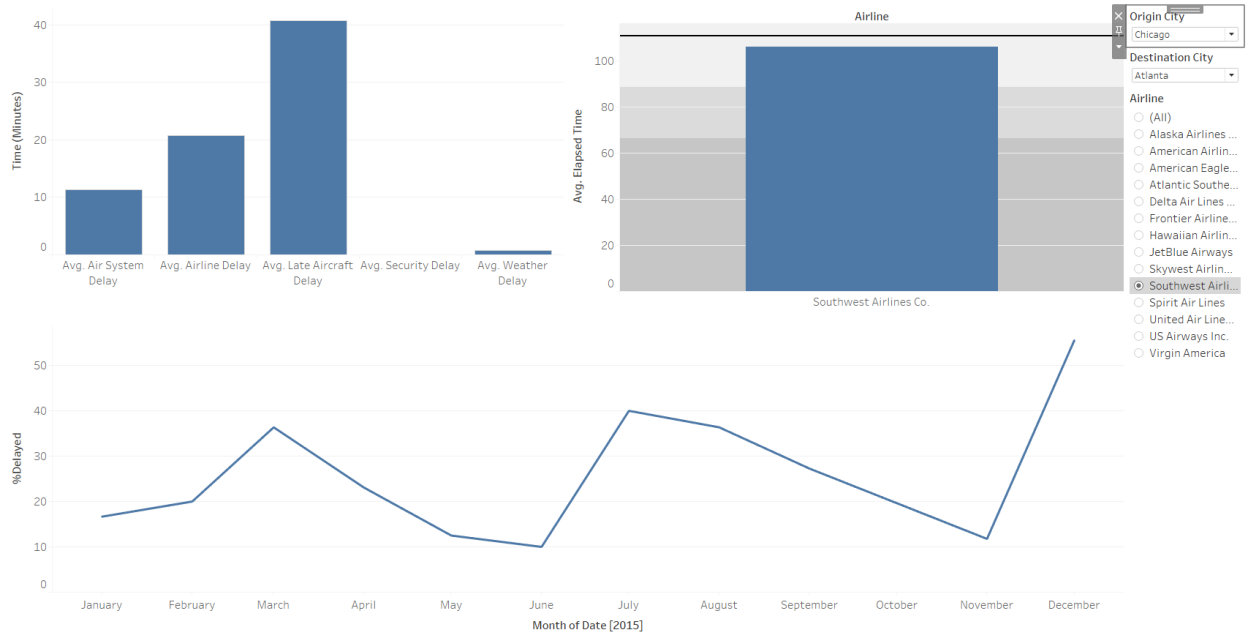


*Figure 10. Flights from Chicago to Atlanta with Southwest*

# CONCLUSION

This report was able to answer the questions in the introduction. The key findings are illustrated here:

1. Hartsfield-Jackson Atlanta International Airport (ATL) is the busiest airport in terms of origin flights. The busiest time of the year was July. Southwest Airlines travelled the longest distance. There was no significant difference between weekend and weekday flights.

2. Spirit Airlines is the worst airline in terms of delay percentage (31%). It also had the longest delay with an average time of around fifteen minutes. The most notorious airline in terms of flight cancellations is American Airlines with a rate of 4.8%.

3. The number one cause of delay per airline is late-arriving aircraft. The number of delays vary by the time of the year with February topping in percentage delay and June topping the total number of delays. It is possible to choose the fastest airline from one route to another by comparing their elapsed times and scheduled times.