US Flight Data 2017 - Project Notes

Author: Daniel Cummings

Email: daniel.j.cummings@me.com

Date: 7/14/2018

Final data visualization available at:

https://public.tableau.com/profile/daniel.cummings#!/vizhome/US_Flight_Data_2017/FlightDataStory

Initial draft visualization available at:

https://public.tableau.com/profile/daniel.cummings#!/vizhome/US_Flight_Data_2008_draft/FlightDataStory

Summary

The purpose of this project is to create data visualizations about United States flight data during 2017 using Tableau. The main areas of focus are related to airport traffic, flight delays, and reasons for cancellation by location, carrier, and time of year. The data set for this project can downloaded directly from the United States Department of Transportation (DOT), Bureau of Transportation Statistics website and is also included in the github repo for this project. More than 5.6 million flight records are used for this project.

Design

When I think of air travel I immediately start imagining exciting geographic locations I could go to for vacation. Then I think of the hassle of the check-in security process and (most of all) any past flights that have been severely delayed or cancelled. Lastly, I think of which airline carriers I've had good or bad experiences with. The features of geographic location, flight issues and carriers are what drove the visualization design process in this project.
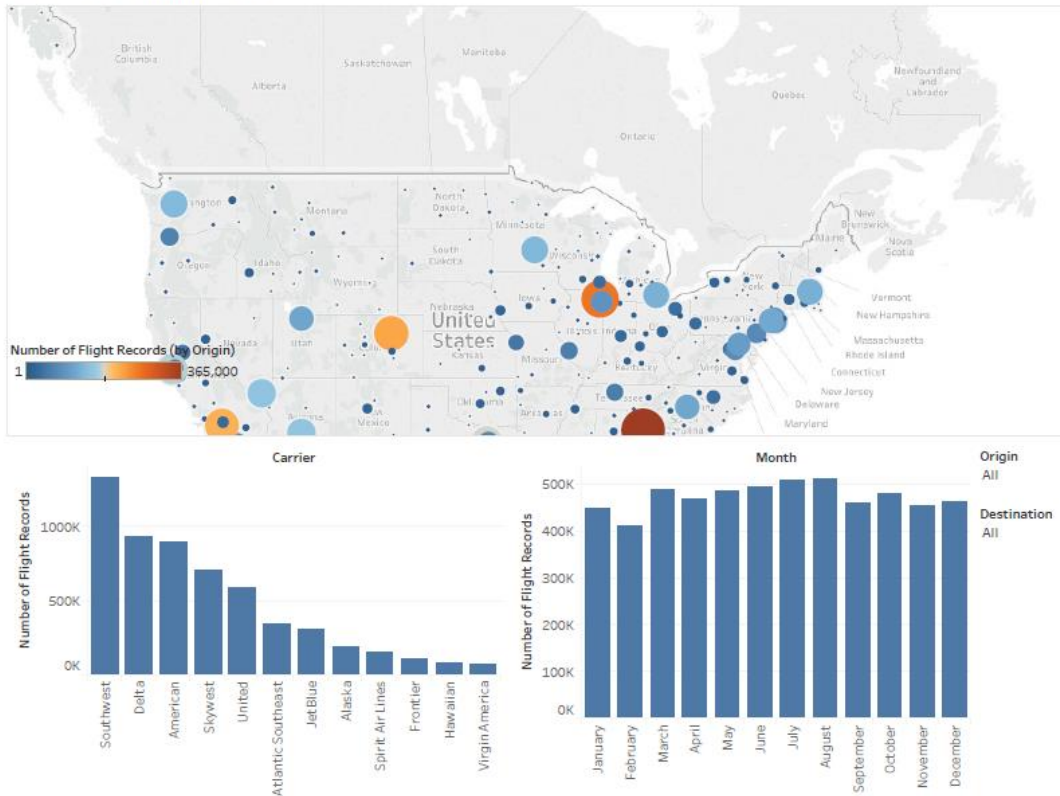
The Tableau story for this project was broken into four sections:

1. Airport traffic by geographic location
2. Departure delay by airport and route
3. Departure delay by carrier and month
4. Cancellations by condition, carrier and time of year

**Airport Traffic by Geographic Location (Story Dashboard 1)**

In the first dashboard I focused where the busiest airports were located. The 'maps' plotting option provided a great was to visually showcase the traffic (number of flight records) by airport. Bar plots were used to cleanly show the carrier and month features of the flight records. I chose a blue/orange color gradient where orange highlights the busier airports. After feedback, I highlighted the busiest airports by size and allowed for carrier/month bar plot sorting by origin and destination. The latter allows users to see what carriers are dominant in a particular airport.

Average Departure Delay by Airport and Route (Story Dashboard 2)

The purpose of this dashboard is to show the number of flights through airports does not always correlate to higher average delays. For the top plot, a log scale on the y-axis made it easier to see the lower traffic airports on the plot. However, to prevent the traffic feature from getting lost, I added traffic as a size marker on the top plot. The lower plot allows the user to view in detail the average delay in terms of origin(s) and destination(s). After feedback, I updated the lower 'route' plot to remove unnecessary colors and added a size marker for more common flight routes.
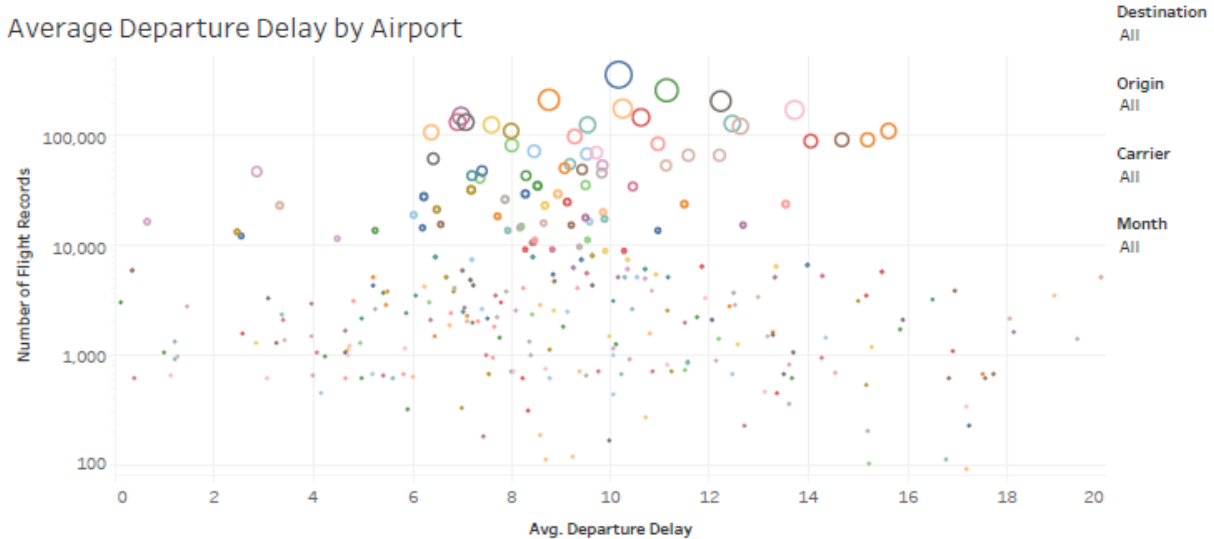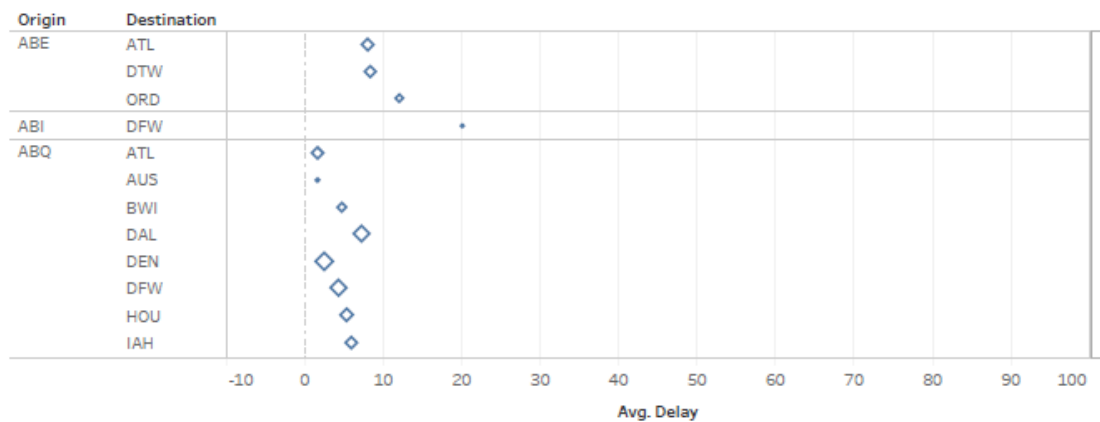
**Average Departure Delay by Carrier and Month (Story Dashboard 3)**

This dashboard complements the prior in that the user can explore average delay based on carrier and month. The map plot again seemed used for visualizing origin and destinations geographically. The most useful feature for this dashboard is the ability to filter by origin, destination and carrier. It is really interesting to see what carriers have the most average delay in terms of airport.
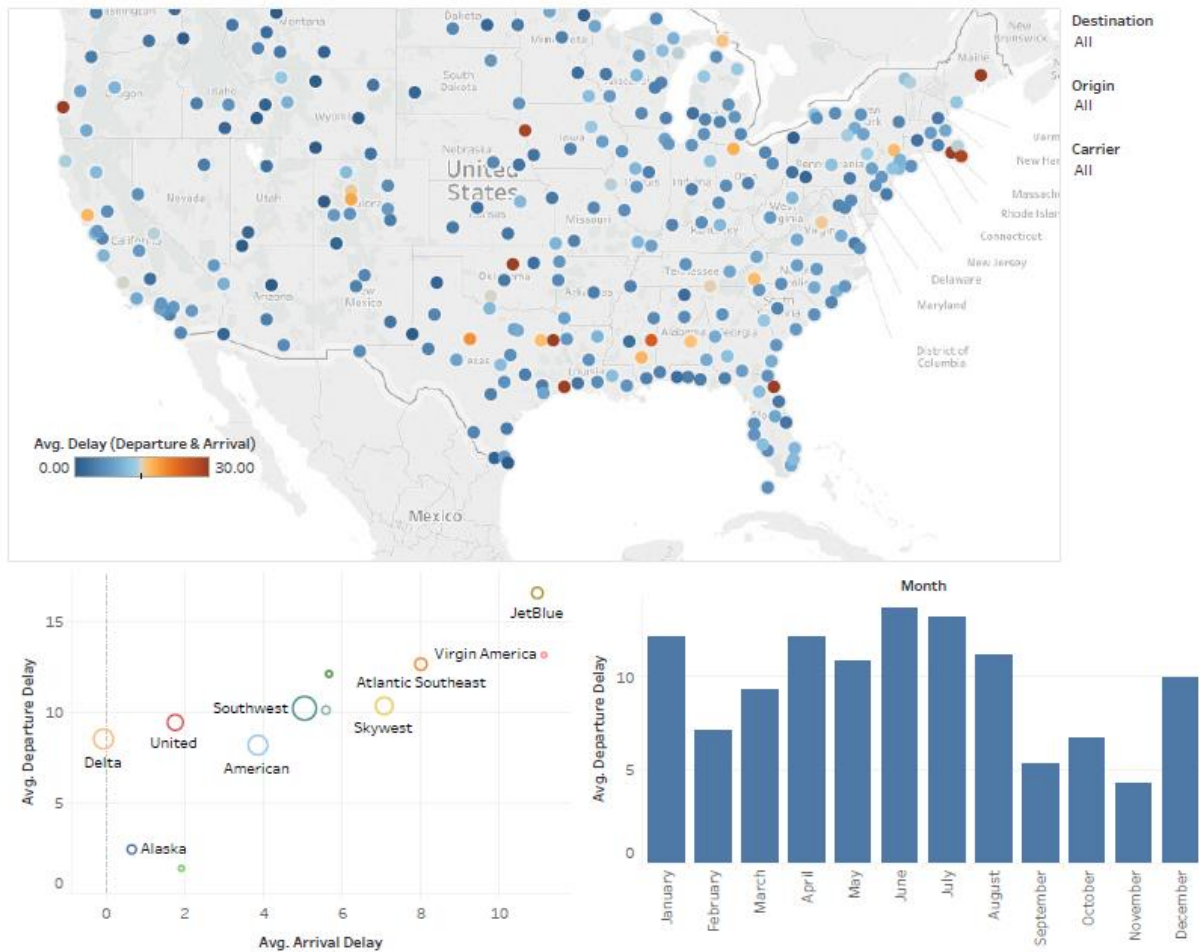
Cancellations by Carrier and Time of Year (Story Dashboard 4)

Cancellations are one of the most stressful problems travelers encounter. This dashboard examines the causes of cancellation over time. The lower right month seems especially simple and 'to-the-point' for highlighting that October and November were the best month to fly without worry of cancellation. Another feedback was to include the proportion of cancellation causes by carrier.
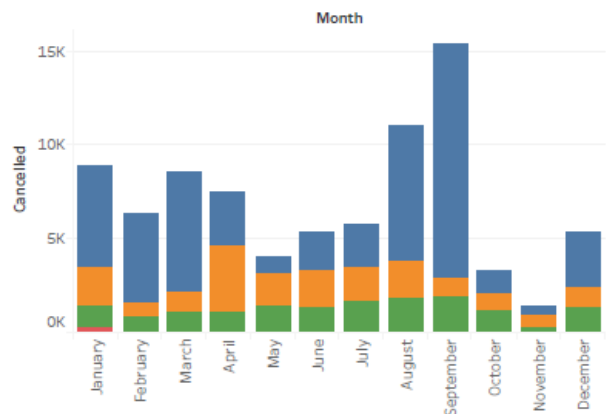

Cancellations throughout 2017
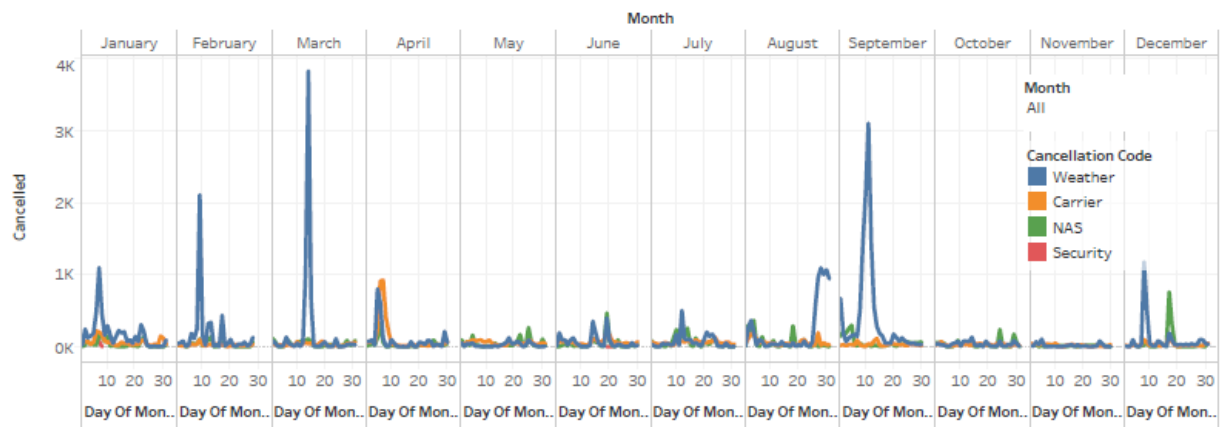
Cancellations by Carrier
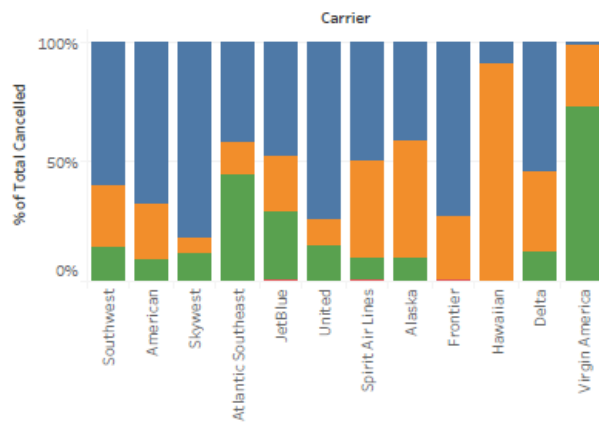
Total Cancellations by Month

Feedback:

The primary feedback I received on my early revisions was that I should use recent flight data. The initial data set I obtained (from another source) had data only as recent as 2008. After finding and implementing the 2017 data in the visualization, the other areas of feedback were:

- No red/green color schemes on first dashboard (due to colorblind issues)
- Focus/zoom the initial axis scales on few plots, but add the view zoom in/out options for the user
- Remove colors that conflict with categories between charts on the same dashboard

Resources

- The data set (among many others) can be directly obtained from the following Department of Transportation, Bureau of Transportation Statistics website:
  https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=
- Plotting geographic data:
  https://kb.tableau.com/articles/howto/plotting-geographic-data-using-custom-longitude-and-latitude
- Applying filters across multiple worksheets:
  https://onlinehelp.tableau.com/current/pro/desktop/en-us/filtering_global.html
- Final data visualization available at:
  https://public.tableau.com/profile/daniel.cummings#!/vizhome/US_Flight_Data_2017/FlightDataStory
- Initial draft visualization available at:
  https://public.tableau.com/profile/daniel.cummings#!/vizhome/US_Flight_Data_2008_draft/FlightDataStory