

Summary:

There are very few things as universally despised as airport delays. Using data on U.S. Airport domestic flight delays from 2016-2017 this set of visualizations explores the prevalence and source of delays. There are two times of the year (December-January and May-August) that length of delay and number of 15-plus minute delays is higher. Combined with geographic visualization and delay classifications it is clear that weather is not a prominent source of delays.

Design:[Final Visualization](#)[Initial Visualization](#)

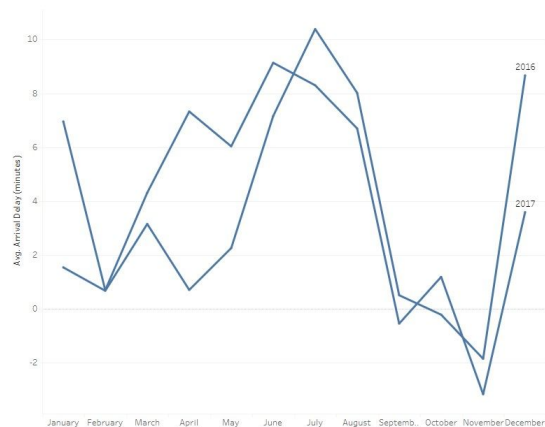
Designing the visualizations it was important to choose ones that told the clear story of airline delays, so many were created, the paired down to the best tale. Initially I built this using the last five years of data, but unfortunately this is too many lines of data to upload to Tableau Public. In reflecting on what I found and the story I was telling it seemed fewer years rather than few locations was a better direction since my message was not about a change in the industry with delays, it was about the delays and findings were consistent in the shorter and longer term datasets. In meeting with both of my reviewers for feedback they suggested ways to simplify and at the same time amplify the message trying to be shared.

For the first visualization I looked at average delays by month of the year. I chose to use a line chart to show the monthly delay averages because line charts are useful for showing trends over time series data. Initially I only looked at average length and had the two years graphed as separate lines to show the similar pattern. This created some confusion with reviewers. In working to make it more clear, I also realized that average length could be significantly different from total number. I chose to switch to dual axis to show the similarity in the patterns and to show the average of the two years with a filter to allow seeing them individually.

The second visualization looks at geographic distribution of the average amount of arrival delays over 15 minutes by state. I chose to map this and use maps rather than symbol maps because I was only looking at one variable, so color rather than color and size would be used. Originally I considered mapping each airport's location, so using the symbol map, however mapping state averages told a clearer story. Initially I did this by average length of delay, but decided to switch because how often seemed more important and would avoid some really long individual events impacting the story. The greatest feedback I received was about the colors because I had a gradient and it made it hard to compare the states. Switching to a stepped gradient told a much, much clearer story.

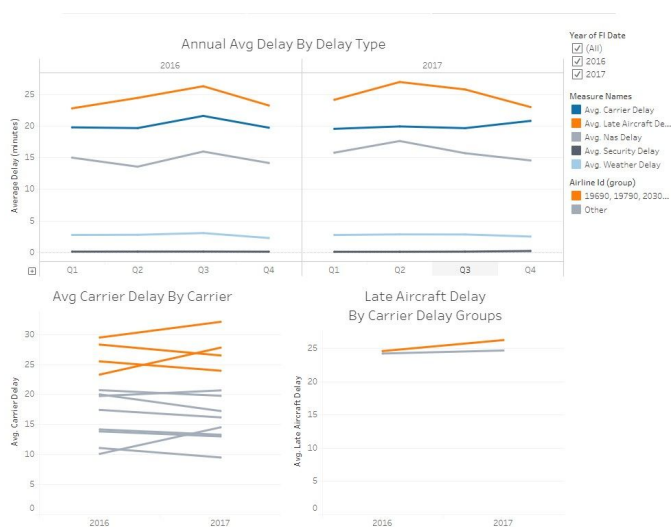
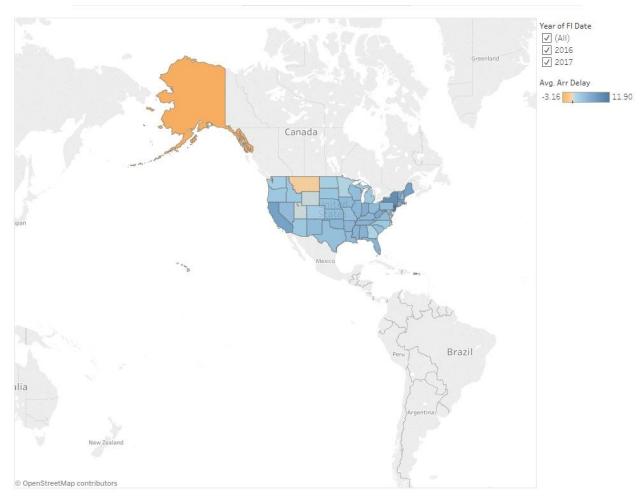
The third and final visualization shows the average delay length by delay classification. I chose to use a line chart to show the monthly delay averages because line charts are useful for showing trends over time series data. It shows the two years side-by-side to show the patterns and consistency. Initially I had another pair of graphs that divided by carrier down below and groups them to explore carrier delay's connection to late aircraft, but the feedback I got was this muddled the story. I agree it did because it was hard to explain why I was doing it.

Feedback:



I had two different people review the visualizations and provide feedback. The first thing one noted was to use different colors for the two years in this visualization. The second was to clean up the x-axis labels. I hadn't noticed they were crunched and cut off.

For the second visualization there was some confusion caused by the Filter title (not cleanup from YEAR(FI DATE)) and the caption provided at the top. The color key also with the change in color at zero made the map less telling, questioning if I could enlarge it and add more midpoints to see the difference in colors. Both viewers did like ability to click states and see more and change between years.



For the third and final visualization there was confusion with the two lower graphs I created. They were not clear in the information ("On the avg carrier delay by carrier, what's the difference between the various orange lines and the various gray lines? Why does one orange line trend up and the other three trend down? What does that signify?") and were muddying the message that the rest of the story was telling.

Resources:

Data Source Field Explanations: https://www.transtats.bts.gov/Fields.asp?Table_ID=236

Explanation of the Delay categories:

<https://www.bts.gov/topics/airlines-and-airports/understanding-reporting-causes-flight-delays-and-cancellations>