12 July 2018

US Flight Delays 2008

Tableau Stories:
Version 1: https://tabsoft.co/2zpRnJX
Final Version: https://tabsoft.co/2udTko0

Summary:

The visualization created using the Tableau story linked above depicts flight delays (both departure delays and arrival delays) in the United States in 2008. Airport hubs were defined as airports with more than 100,000 flight departures in 2008. The takeaways from the visualization are that some carriers, such as United and American, have longer delays on average than others. Also, delays are generally longer in the summer and winter months, and generally worsen as the day progresses.

Design:

I knew from the start that I wanted to create a visualization to delve into flight delays with this dataset. Specifically, I wanted to focus on factors that a passenger could control such as carrier, time of year for travel, scheduled flight departure and arrival times, and the day of the week for flying. I first wanted to create a map to draw in the audience, and the map nicely displays airports all over the U.S. Sizing the airports based on the number of records (flights) was the best option as number of flights is a continuous variable.

In the airport hub charts, grouping to only include the airports with more than 100,000 departing flights for the year made the visualizations more digestible for the viewer as it avoided the issue of overplotting. Size was again used to display the continuous variable of departure delay and the map to display airport location since I wanted to make a point about airport geography and delays.

I chose to represent the average delay and number of departing flights for each airport (and in a later plot, for each carrier) in a scatter plot as a scatter plot nicely represent the trend between two continuous variables for multiple data points. The line plot was used in multiple places to display delays over time because line plots nicely show trends over time, and are a visual that allows the viewer to quickly be able to see the trend being portrayed.

To represent delays by carrier and delays by day of the week I used bar charts, as bar charts are great tools to display data for a continuous variable over a categorical variable. The sorted delays by carrier plot makes it easy for the viewer to distinguish which carriers have higher delays, and the stacked delays by day of week bar charts help make it clearer which bars are taller.

Based on the feedback I received, I changed labels to tooltips where the labels made the visualization more confusing. I also used a different color palette for the flight delays by type visualization to make each type of delay more clear. Originally, I thought it might be best to use one color for the three causes of most significant delays but it ended up being confusing. I also altered captions and added visualizations with the number of flights over each month and over

each day of the week. For flights over each month and day of the week, I decided to represent the data with the same types of charts as the data for average delay over each month (line chart) and day of the week (bar chart). This way, the viewer could easily see the differences in trends for average delay versus number of flights when placed side by side. Lastly, I changed the story layout to a single dashboard layout design to make the comments more readable and the visualization flow better.

Feedback:

The feedback I received on the first draft of the visualization was:

- The labels with the number of flights for each airport on the Airport Hub Delays plot make the visualization confusing since they don't correlate with the size
- The caption for the slide with airport hubs is confusing and should better portray that the focus is on the hubs in the plots
- The caption on the flight delays by type shouldn't be first-person
- The colors on the flight delays by type plot make it confusing to determine what is what
- Should remove the labels with the number of flights for each carrier on the delays by carrier plots, and remove the mention of "smaller airlines" in the caption as these confuse the viewer as to what the finding is
- If possible, add a plot with the number of flights over each month for 2008 to see if more people fly in the summer and winter
- Alter the dashboard so comments are readable without having to scroll

Resources:

- http://www.vizwiz.com/2016/07/tableau-tip-tuesday-layout-tips-for.html