

Tableau Write-up

Author: Amrut Deshpande

February 19, 2018

Visualising Flight Delay using Tableau

My Tableau workbook link:

a) First version- Version before review:

https://us-west-2b.online.tableau.com/#/site/amrutstableau/views/Flight_Delay1_0/Story1?:iid=6

b) Final version- Version after review:

https://us-west-2b.online.tableau.com/#/site/amrutstableau/views/Flight_DelayV2_0/Story1?:iid=7

Summary:

I obtained my dataset of Flight delays from RITA website which contained information on flight delays and the performances. The link to download the dataset is,
https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp

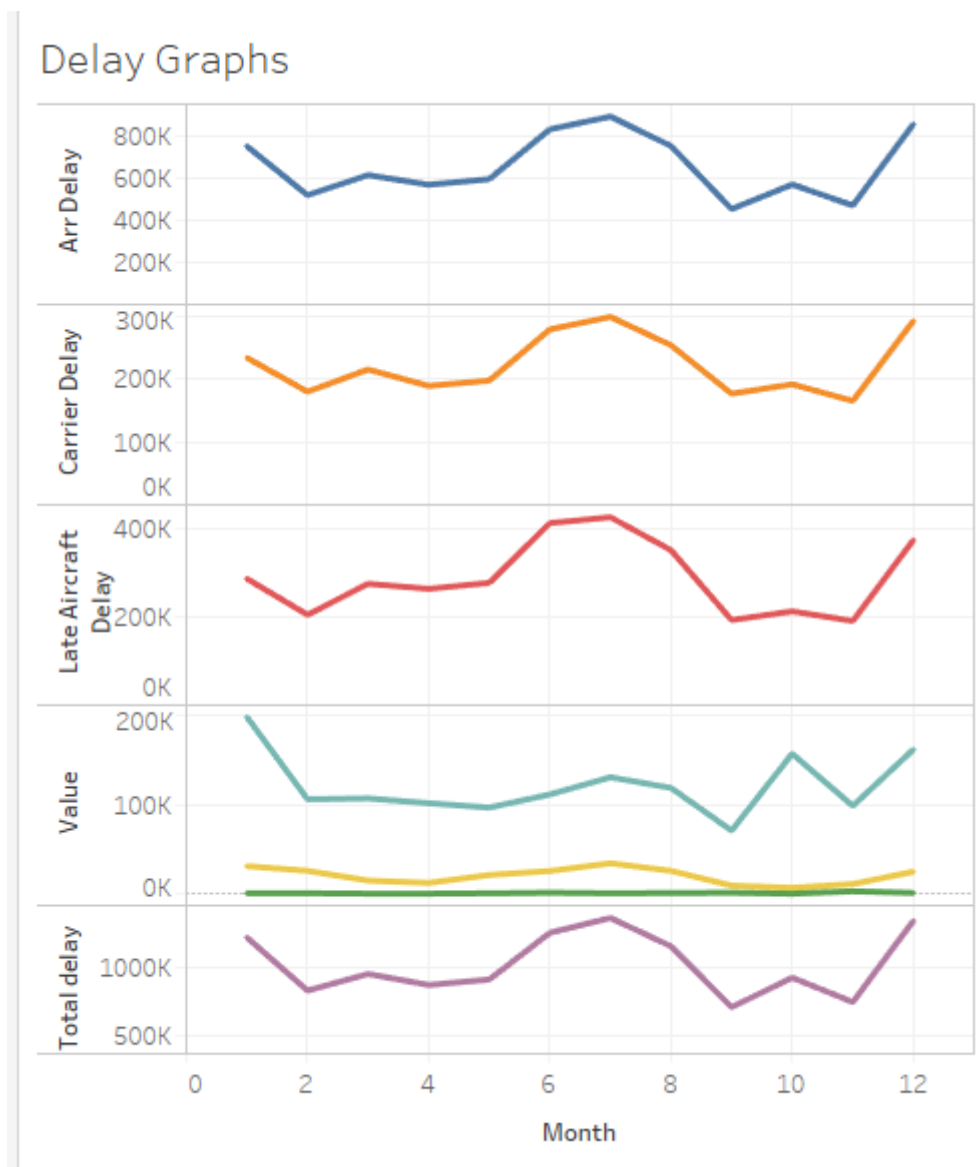
I downloaded the data from January 2012 to December 2017. Before starting my visualizations, I spent considerable time in understanding the variables of the dataset and its datatypes. I identified some of the interesting questions that I can explore such as:

- 1) Which month has the most number of total airline delays and how does the trend look like for the entire year?
- 2) Which are the reasons for flight delay and how does it behave over the year?
- 3) Geolocation plot of most number of arrival delays and the airports which have those delays?
- 4) What are the carriers that causes most delays when all airports are considered?
- 5) What are the top 5 airports and carriers contributing for most delays and their relationship?

Design

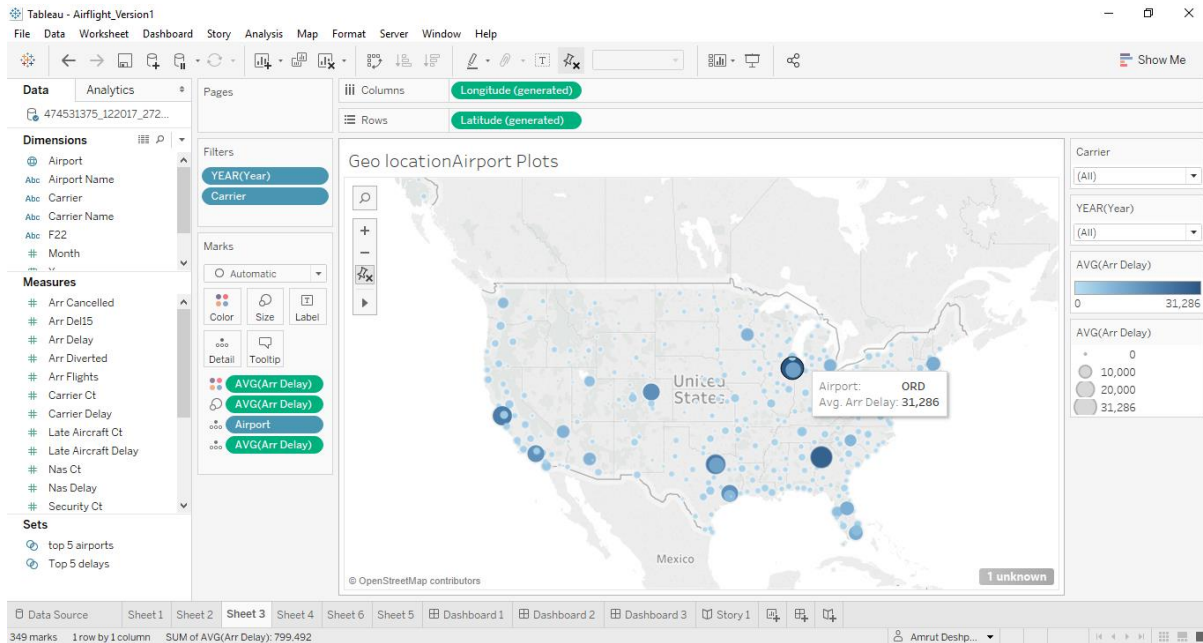
In my first design, I wanted to explore diverse types of delays to see which the major contributors are. Also, to know and answer my question of which month has the maximum number of delays over the last 5 years. I created a new dimension 'Total delays' by adding all other individual delays to see how they interact and behave over these years. I also added years, carrier name, airport name, Measure names with delay to my filters to investigate more on delays related to a bunch of carriers/airports or for a particular information was used for colour to enhance the visual.

So, my first design was a line plot of delays over months. Here is the image of the flight delays bar chart I first encountered. I received a feedback asked me to represent delays as a proportion of total delays. This was a meaningful feedback as it adds more clarity to the visuals. I measured values of all delays as a proportion of total delays and created a new flight delay bar chart. The update now looks as follows

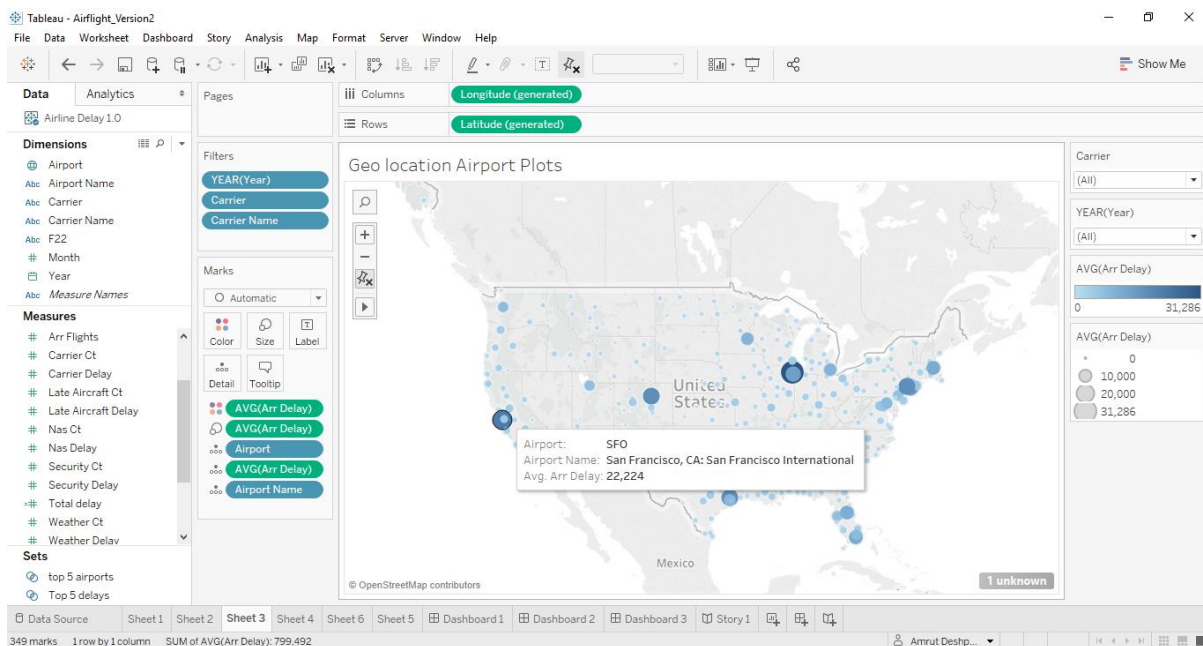


In my second design, I generated geolocation plot of airports. Initially there wasn't a variable to enable maps, I converted airports variable from string to geolocation to create maps. To highlight airports with high delays, I entered average total delay measure to size marks, also added it to colour marks. So top airport with highest delay would look something like dark orange colour with big circle separated from the rest. Here is the image of my first geolocation of airports I got a feedback "Enhance the visual in geolocation plot of airports by adding airport details".

I guess he meant by adding airport name details because that is what missing in my plot. I added airport name to the details section in marks to allow us to see which airports caused more delay. Here is the updated version



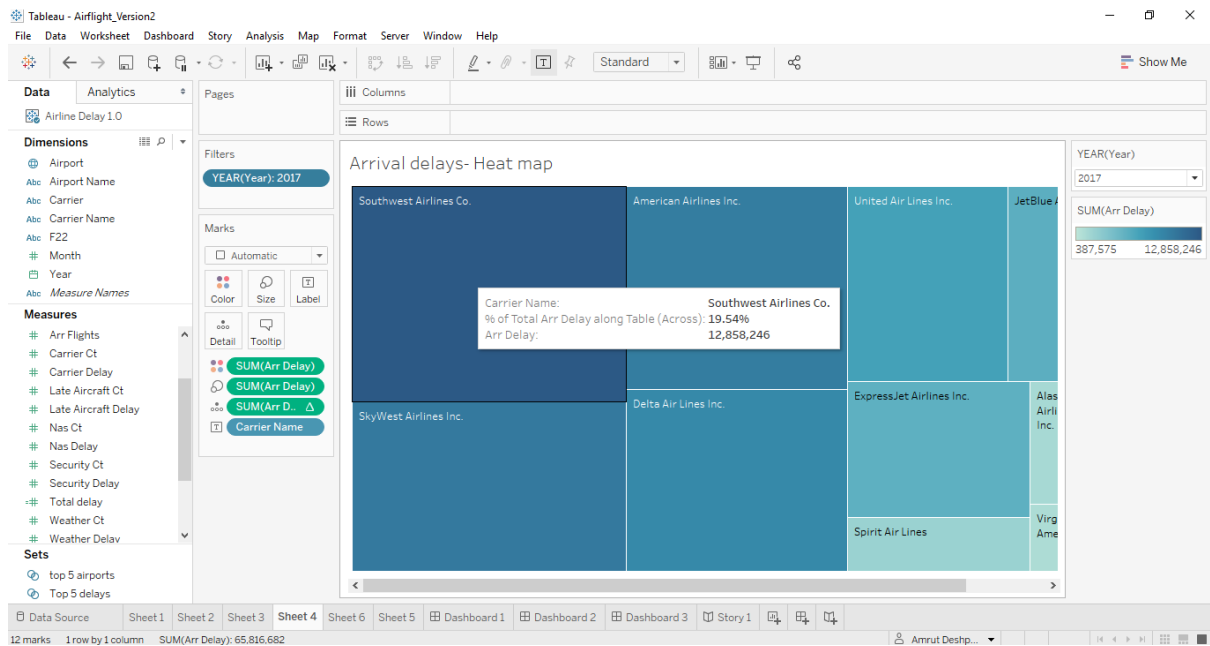
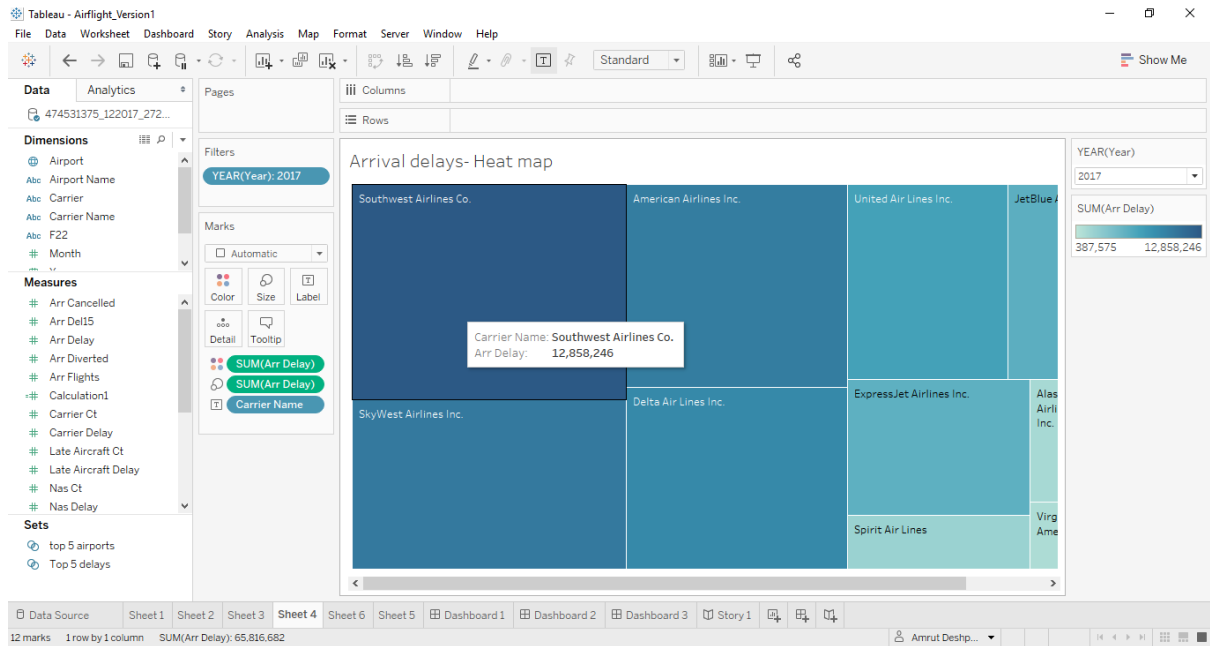
Version 1 image



Version 2 image

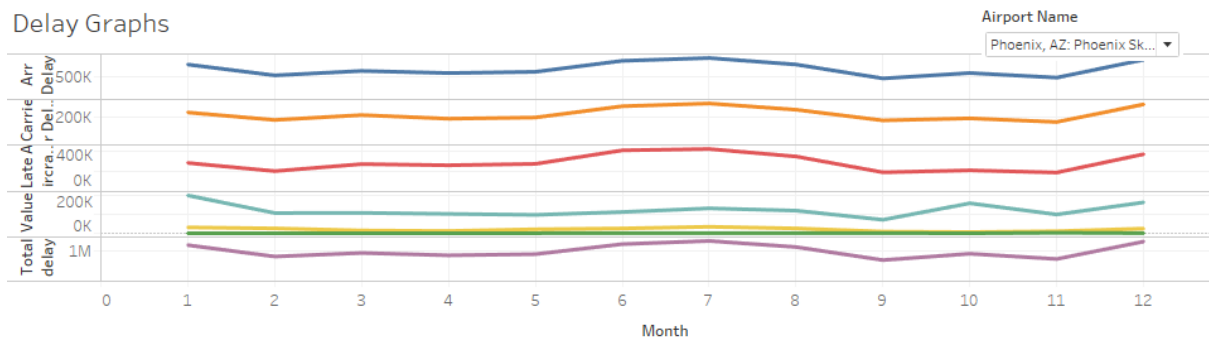
In my third design, I wanted to find which were the carriers that got delayed the most considering all airports and years. Also, for a particular carrier, which airport caused the highest delay. This would give me valuable information in a single plot. I created heatmap for carriers and mapped it to geolocation plot. I added delay information details so that when we hover over the map, we get delay information directly. Here is the image from my dashboard

A feedback for this plot was to represent percentage of delays across different carriers. I updated the same and here it is

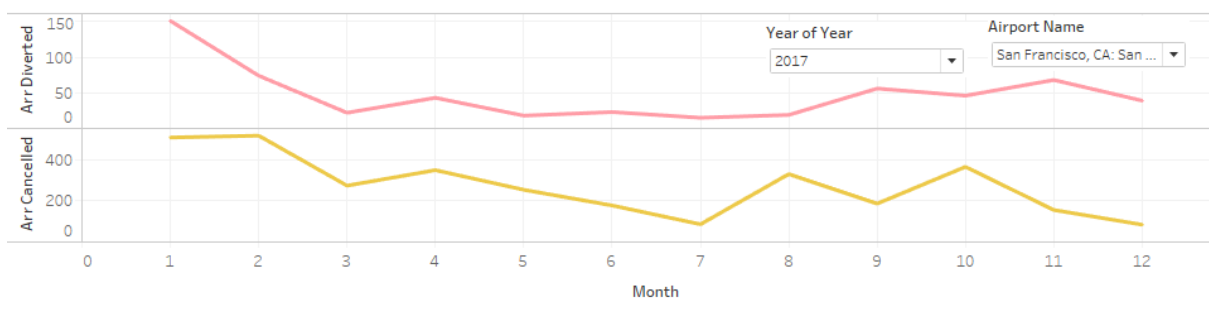


My final design involves creating plots for top 10 airports and carriers for most delays. I created two sets for the same. I created top 10 airports in geoplot and top 10 carriers in heatmap.

Delay Graphs



Arrival problems



Feedback:

My feedback link:

<https://discussions.udacity.com/t/feedback-needed-for-my-tableau-project/605005>

HiAmrut,

1. You can improve the inference by generating airport name to details.
2. Also, it's better to have a percentage metric of delays across different carriers
3. You can add total delays and compare rest of the delays with this over line chart/bar chart.
4. Make sure all filters are aligned perfectly and set to float.

Hope, this helps.

Resource:

Tableau documentation:

<https://www.tableau.com/support/help>