Flight data visualization using Tableau

Author : Vaibhav Gaur

Date : 16/11/2018

Link to my Tableau Workbook

Initial Link (https://public.tableau.com/profile/thevaibhavgaur#!/vizhome/FlightDelayVer6Complete_1/Airports)

Final Link (https://public.tableau.com/profile/thevaibhavgaur#!/vizhome/FlightDelayVer9_3/Story1)

Summary

This dataset is obtained from the RITA website which contains information about flight delays and performance. The dataset I used ranges from 2012-2017. The dataset can be found in the following website:

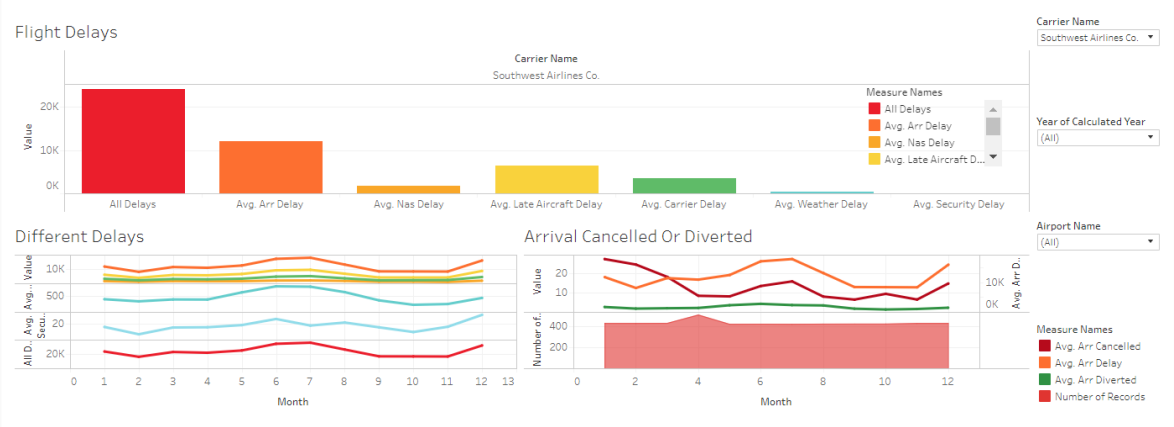
https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp (https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp)

I have created the visualizations in Tableau that answer the following questions:

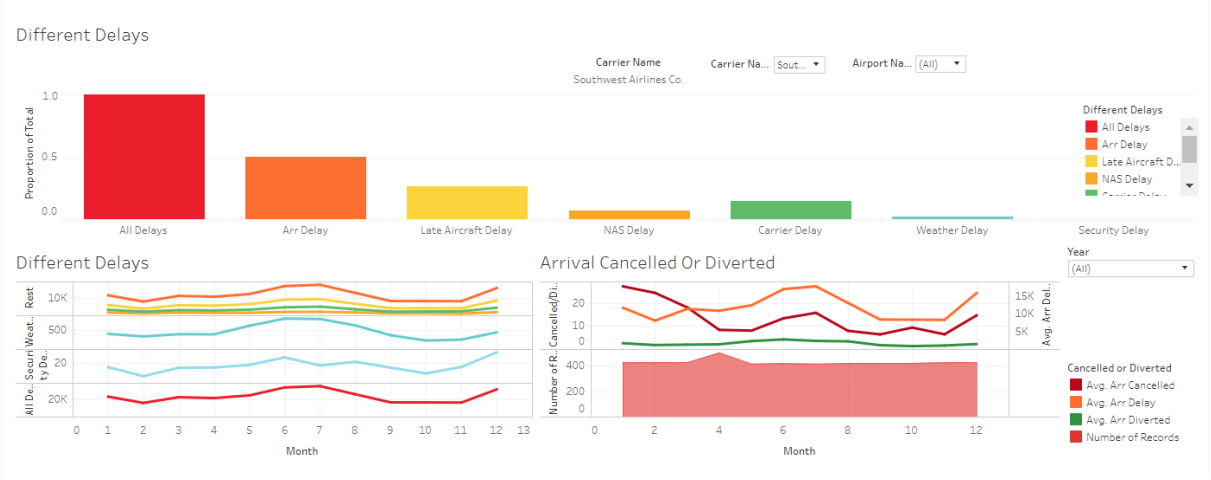
1. What are the percentages of different types of flight delays or cancelled/diverted arrivals and how they are distributed over months, carrier types and airports?
2. What are the carriers that cause most delays when all the airports are taken into account and what are the airports that contribute for these delays the most for different carriers?
3. What are the airports that cause the most delays when all the carriers are taken into account and also what are the airports that cause most delays for a particular carrier?
4. What are the top 10 airports and carriers that are responsible for the delays and when combined together how do they affect each other?

Design

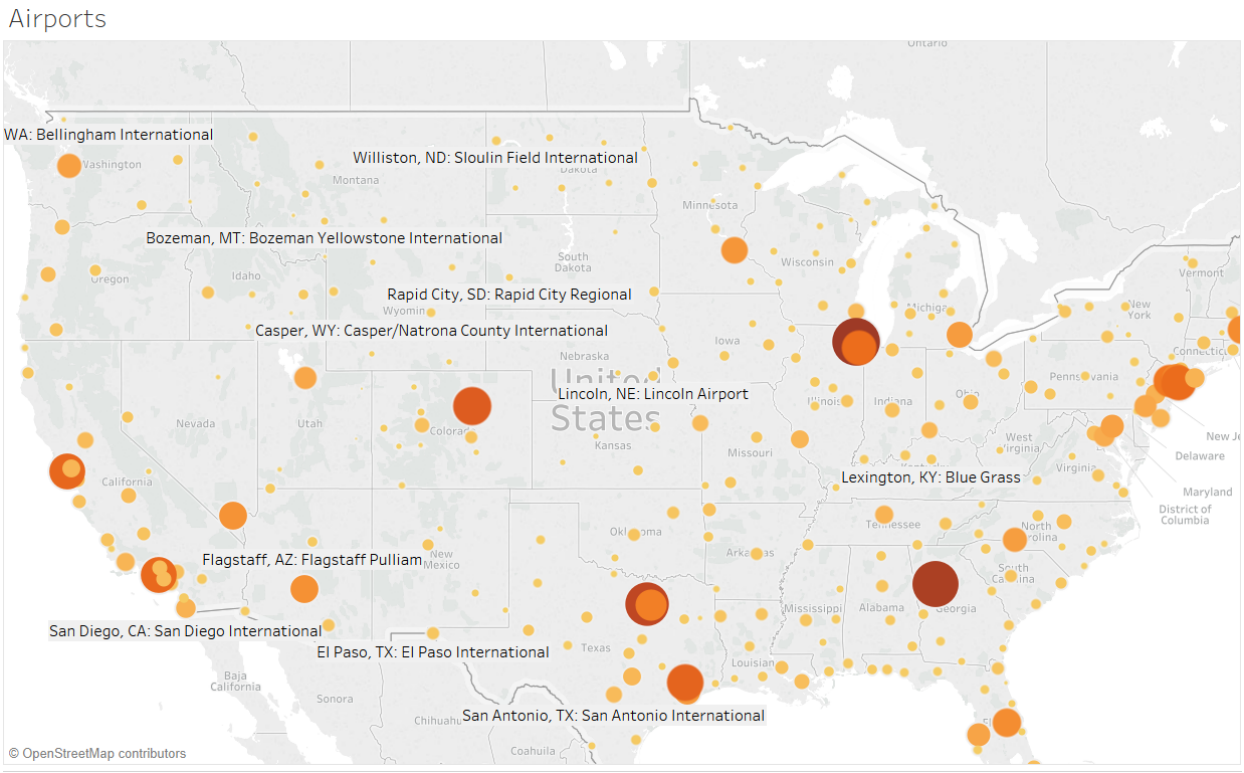
My initial design choice for the first point of the story was to break down different delay types to see a comparison between them and how much each was contributing towards the total delay. Also parallely I wanted to break down even further by months to see if the delays had a high and low point during a particular time of the year. For this reason, I created a derived column called 'All Delays' where I took every delay and added them up. And then in this first plot, I placed total delay also to make a point of reference. Here we can also filter by airport, carrier and year to drill down even further.



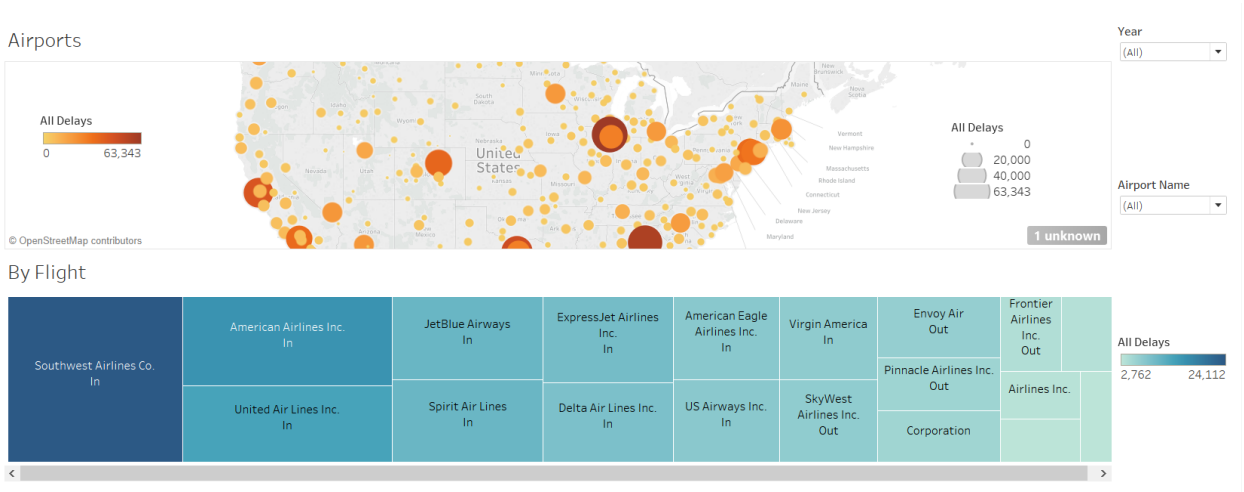
But after this I got my first feedback from the Udacity forum by **ankit450** where he stated "I'd like to see the percentage of whole that each bar represents." This feedback looked like a really good idea. So I converted each of the delays in the top plot to a proportion which tells us what percentage of the total delay did the different delays contribute. I also moved the dropdown box for airports to the top to decrease clutter. Now it looked like this.



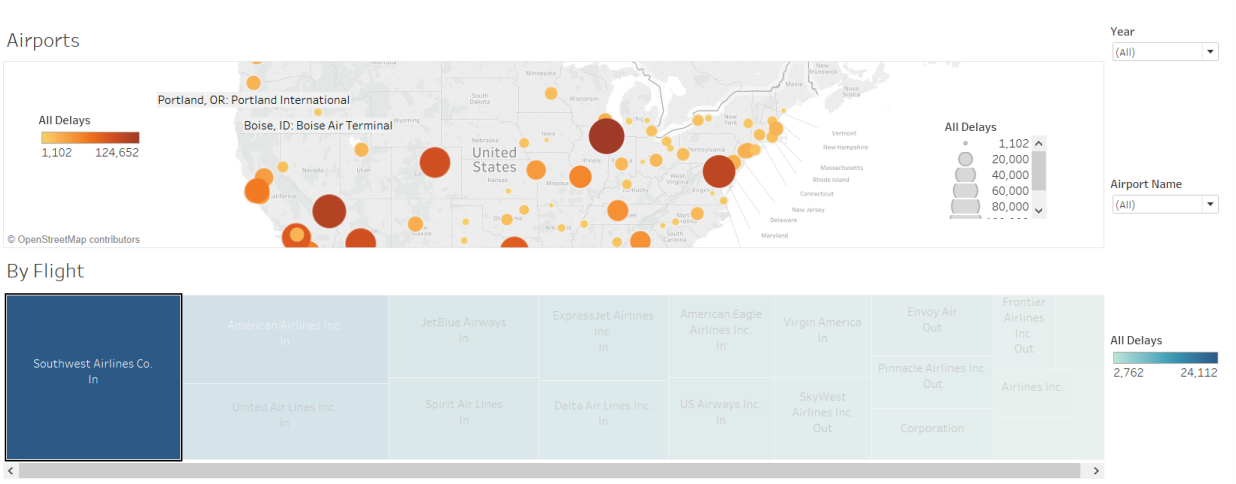
My second design consideration was that I was wondering if there's any way to make a geolocation plot for all the airports and then map the delays via a heatmap to that. I thought it would be really cool to do that. But in the original csv file, there was no field for the countries, cities or states. But I saw there was an airport name field which had all the abbreviations of the airports. So I changed the field type from string to geolocation and chose the option 'Airport' which actually made me a field which can be plotted in map! So in my 2nd plot I did just that. The size and also the darkness of the points showed a redundant plot of how much delay was happening in a particular airport. The more the delay, the bigger and darker the point is.



Now my next target was to see if there's a way to find out which were the carriers that got delayed the most when all the airports and all the years were taken into account. Also in the same pair of plot, I wanted to see that for a particular carrier, what are all the airports that contributes the most for its delay. So I made a heatmap for the carriers and mapped it to the above geolocation plot so that we can see for a particular year, which were the carriers that got delayed the most and what were the airports that contributed most for the delay. I also created an action here such that if only one particular flight was clicked in the heatmap, the correponding airports from which it got delayed were shown on the map.

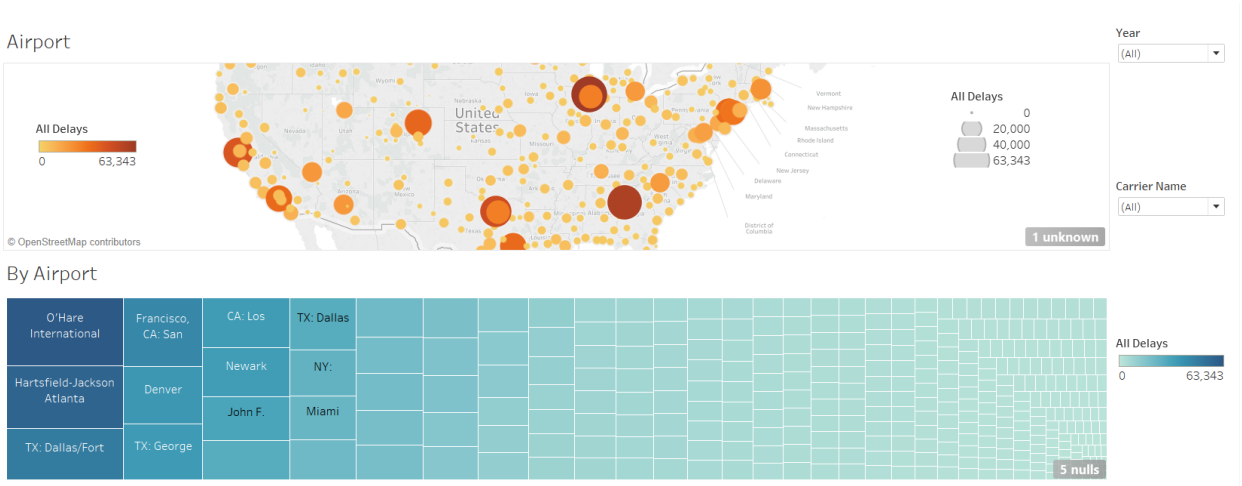


When a carrier is clicked:

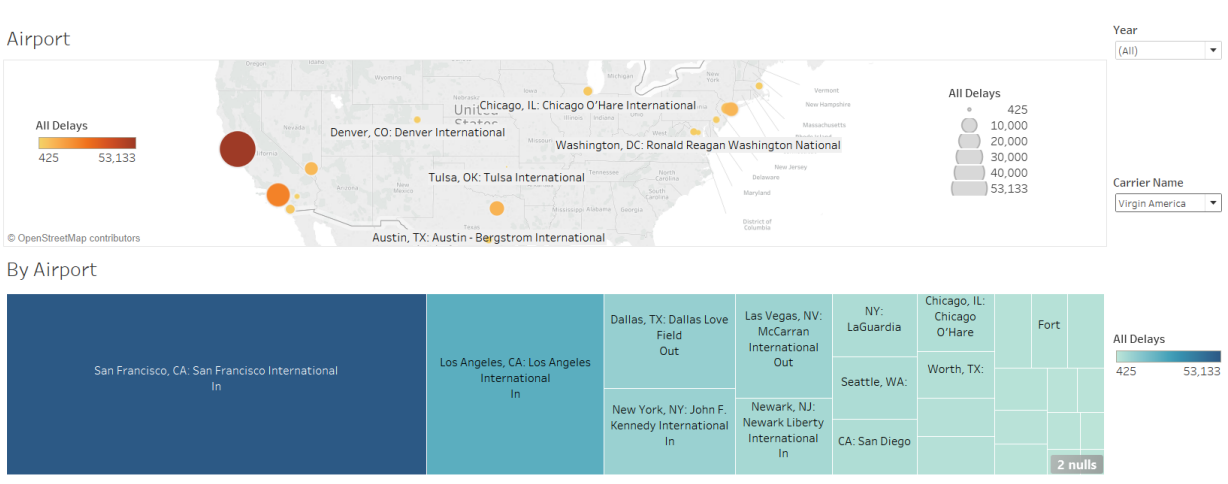


Similarly I wanted to create another plot where I can see which airports contributed for the most delays when all the carriers were taken into account or any one of the carriers is considered. I created a similar heatmap-geoplot mapping for this. When a particular carrier is selected from the dropdown, it shows all the airports that contributed the most for its delay. In this plot if you hover over an airport on the heatmap, you see its exact geolocation in the map.

For all Carriers



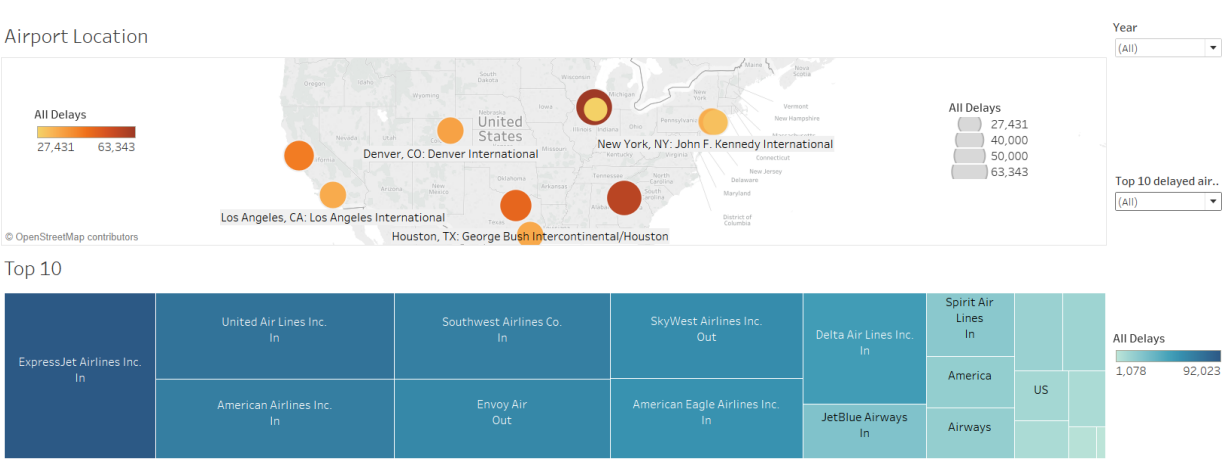
For Virgin America



The last thing I wanted to do was to find out the top 10 airports and carriers which caused the most delay and I wanted to see if these two sets are combined together, was there any new observations? So I created two sets:

1. Top 10 Delayed Airports
2. Top 10 Delayed Flights

These two sets only contained the top 10 airports and carriers respectively. So by using these two sets, I again created a similar plot. First I plotted all the top 10 airports in the geoplot. Then I plotted the top 10 carriers in a customizable heatmap. Whenever one of the top 10 airports is chosen, it will show us the carriers (which are also in the top 10) that got most delayed from that particular airport.



Feedbacks:

I received these set of feedback from **ankit450** in the Udacity forum:

1. I can appropriately add filters to see which airlines have delays and their causes - Thank you!

2. I'd like to see the the percentage of whole that each bar represents- I explained this change above and I have implemented it in my latest version of the story.

3. I'd like the caption section to be short. One line is enough to represent what each page is trying to say. The rest can be visualized by using filters of choice- This statement I didn't fully agree upon. So, I still haven't implemented it. Maybe I will, once I get more feedbacks. My reasoning behind this is even though some of us are trained in Tableau, many of us aren't. For them it might look a little bit unintuitive. That's why in the story, I explained how the geolocation will change if a point on the heatmap is clicked which made the storyboard a little big.

4. PLEASE CONSIDER EXPANDING THE GRID SIZE. OR THE SIZE OF THE PROJECT. It looks very cluttered and hard to make out- I floated some fields and placed them within the plots to avoid clutter. Also I got rid of borders and margins to achieve the same. Both are reflected in the latest version.

Link to the feedbacks

Link (https://discussions.udacity.com/t/feeback-for-create-a-tableau-story/911694)

Resources:

1. Tableau training videos: https://www.tableau.com/learn/training (https://www.tableau.com/learn/training)