Flight Data Visualization Tableau Project

Tableau Workbook Link

Draft

https://public.tableau.com/profile/brandon.chung#!/vizhome/FlightDelayDataVisualization/Story1

Final

https://public.tableau.com/profile/brandon.chung#!/vizhome/FlightDelayDataVisualizationFinal/Story1

Final – Combined

https://public.tableau.com/profile/brandon.chung#!/vizhome/FlightDelayDataVisualizationFinalSingleDashboard/Story1?publish=yes

Summary

The data visualization choice I decided to go with is the flight data set from 2012 – 2017 which includes airlines and airport names as well as different types of delays and the delay times.

The visualizations created in Tableau will answer the following questions.

1. What are the top 10 carriers that are responsible for the most delays and which carrier had the highest average delay time over the course of 5 years?
2. What is the top airport that averaged the most 15 min delayed flights and which airline was it?
3. Among the five different causes of delays which airline had the highest average percentage delay, the type of delay it was and in which year did it take place?

Design - Draft

When I initially started creating my dashboard I designed five of them to be combined into a storyline and wanted to reduce it down to three plus an intro and a summary map for my conclusion as per my feedback from @Deac. My first design was to display the top delays for both carriers and airports because I wanted to capture the audience's attention to where in the U.S had the most delays and which airlines to avoid so it can help them prepare ahead of time. I chose these particular charts to begin with because they were very simplistic and easy to read at first glance to the reader
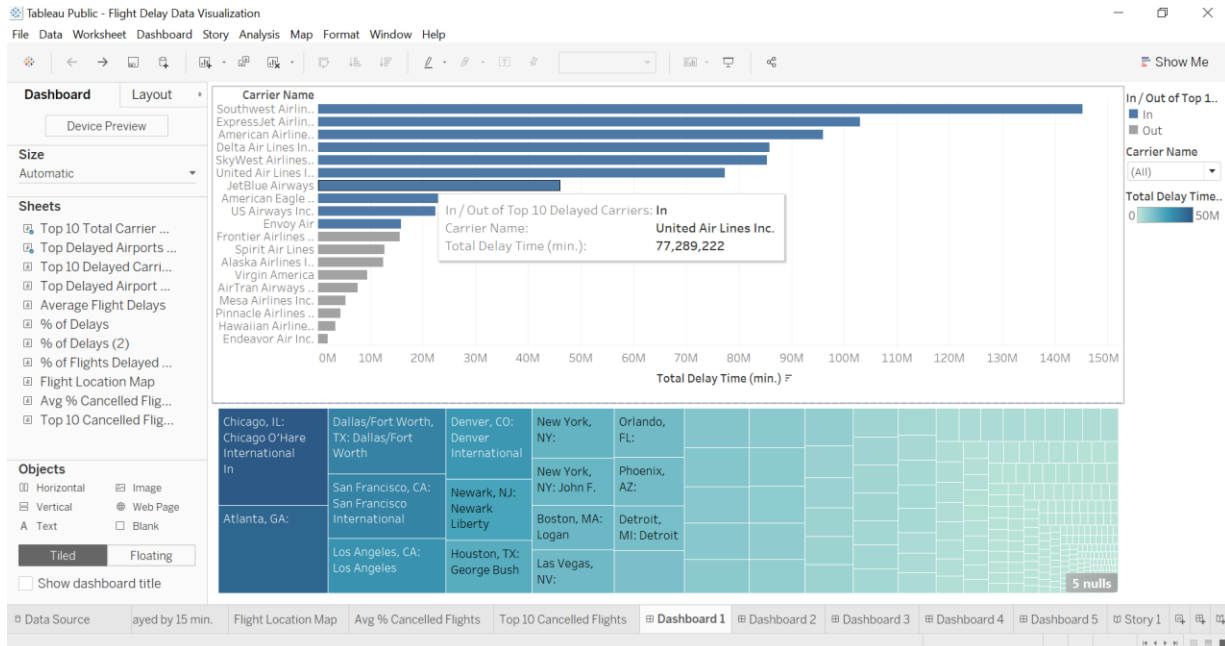
Draft dashboard 2 was designed to show the top 10 airports on the map with the top 10 delayed carriers at the top 10 delayed airports. At first I would try to wrap my head around this idea until I would figure out the drilldown and immediately I knew this would confuse the readers even more so I just deleted it. In the edited version I had to remove the top 10 carriers from the measures and replace it with a just a plain list of carriers. Adding a map early on was not a very good idea since it did not flow along with my data.
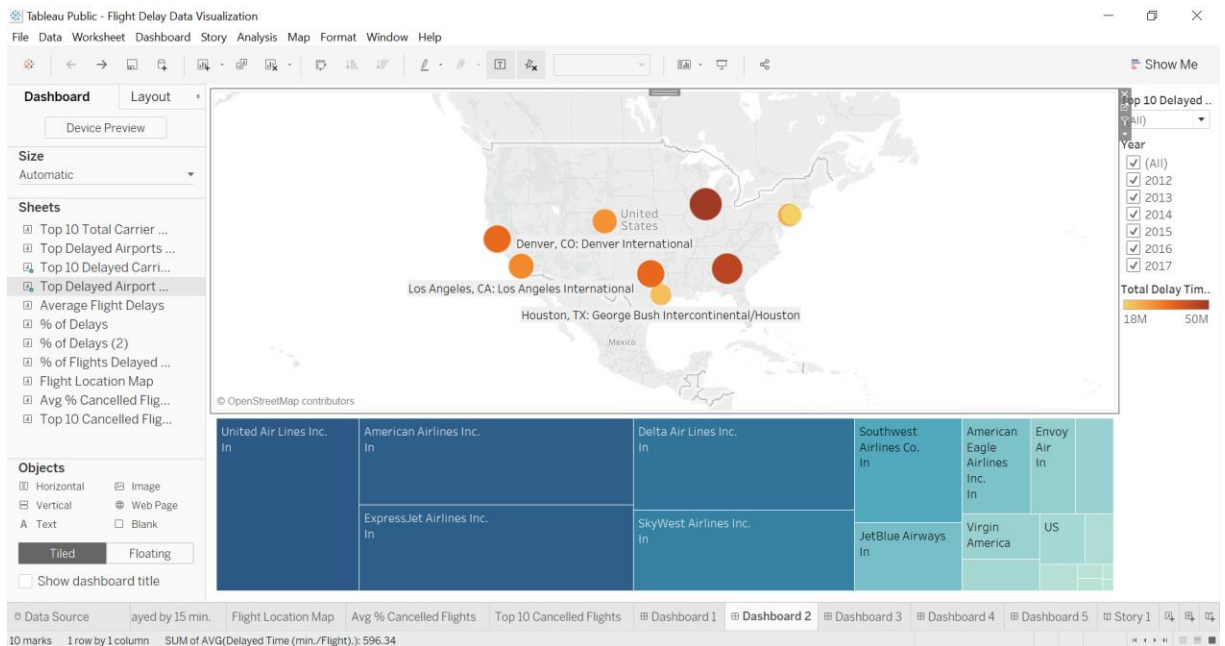

Draft dashboard 3 was first created to explain each cause of delay as a whole while comparing it to the total for the bar chart. Then I added the box plot to show the distributions for all the airports based on the carrier chosen but it would not be relevant to the readers. From the feedback, there was too much information to intake and was too much color that causes a lot distraction away from the data itself. Especially when viewing the mini data points on the boxplot for each airport the reader would not know if it really existed there and would miss it.
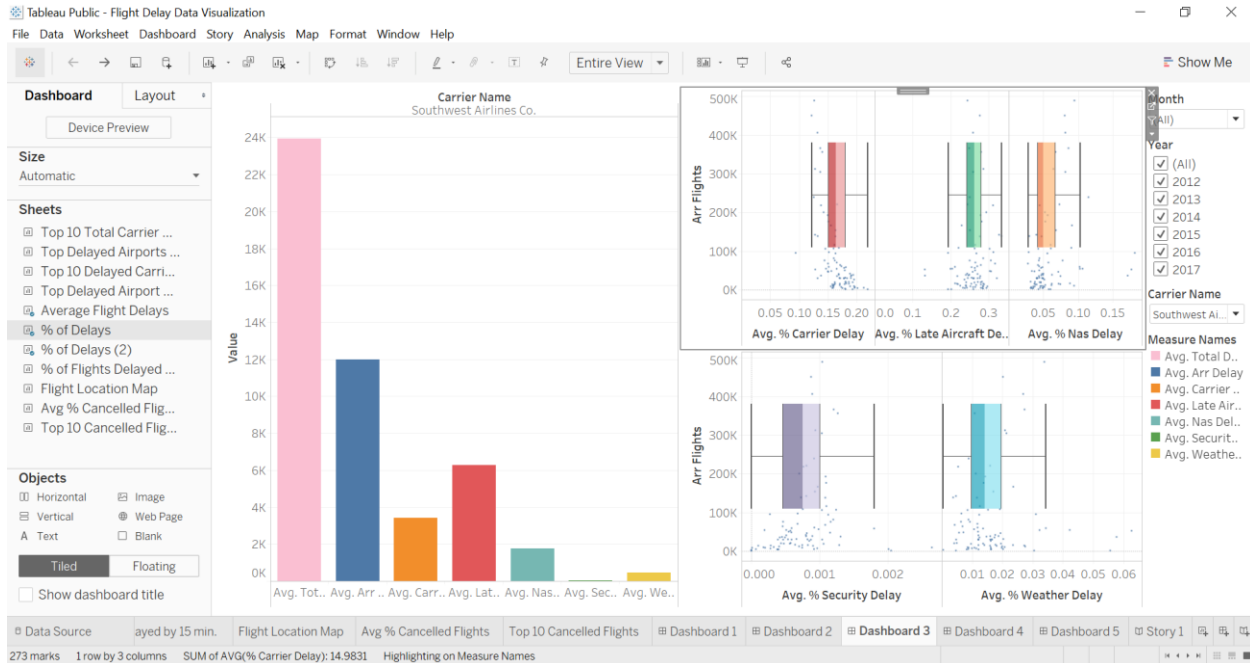
Draft dashboard 4 was displayed as a map that would correspond to each detail in the charts below but I knew having any more than 2 charts would make the dashboard more cluttered. Filtering by carrier name would then split up on the map at different locations in the U.S to where each airport was accommodating each delayed airline. It did not work out as I also found out the filters were not connected anyways.
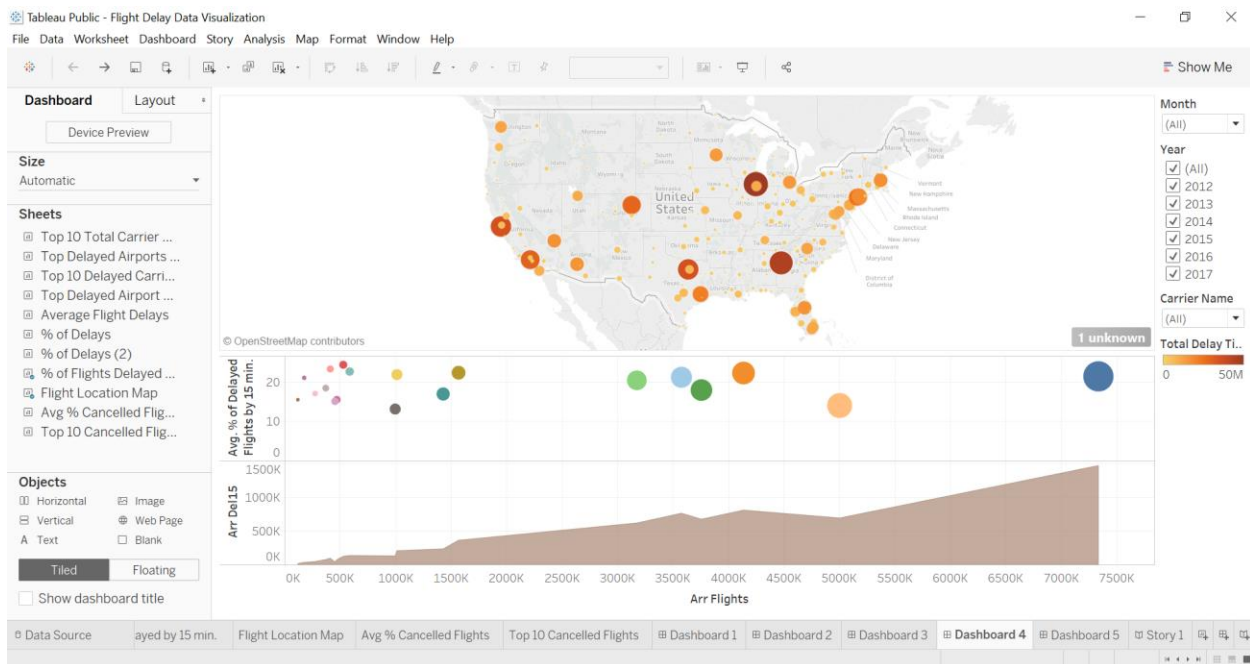

Draft dashboard 5 was completely deleted because the data did not flow well with the other dashboards as my story mostly talks about the delay times while this dashboard explains the top cancelled flights and their average percentages. This left a big gap in my storyline and could not connect the dots using this type of data unless I explained in my other dashboards about cancellation and correlated it with delayed times.
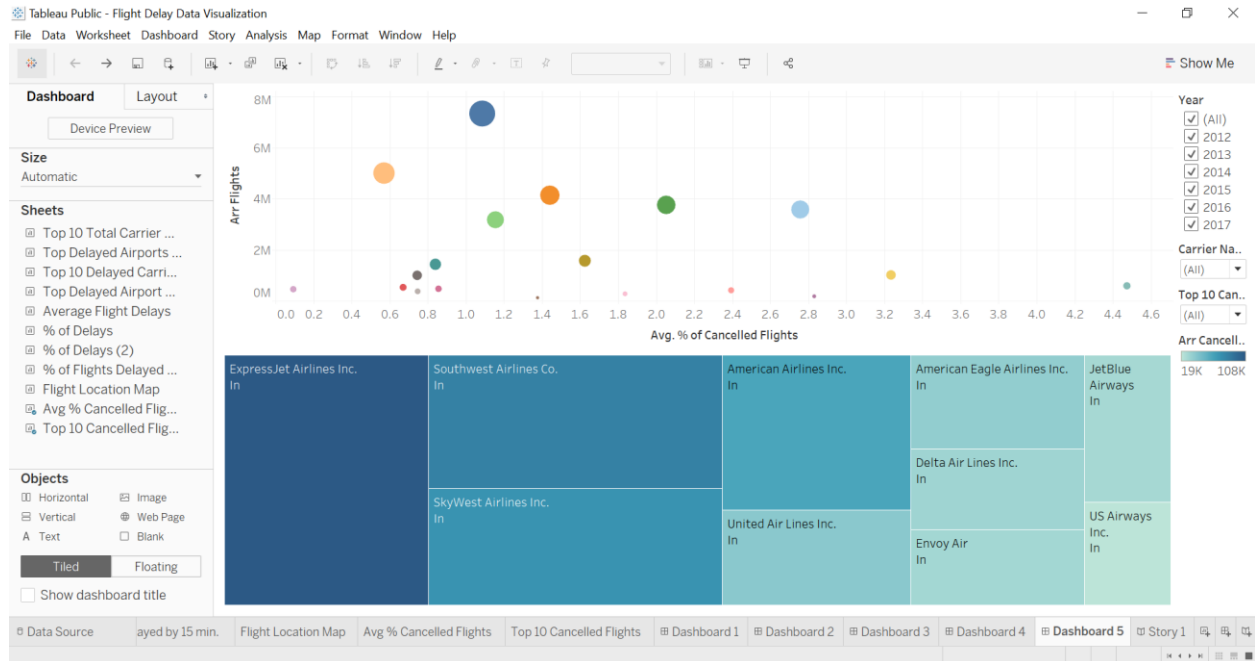
Design - Final Copy

For my introduction, I wanted to display a pie chart of the average percentage of all the flight actions an airline can take while attempting to arrive at the airport which includes arriving on time, delaying a flight, cancelling a flight and diverting a flight.
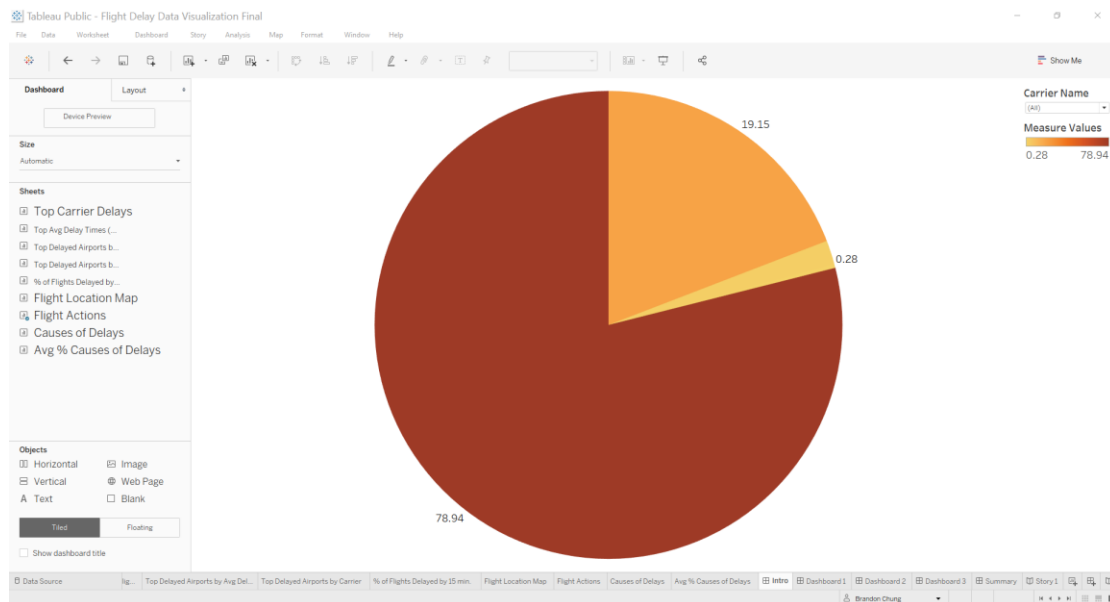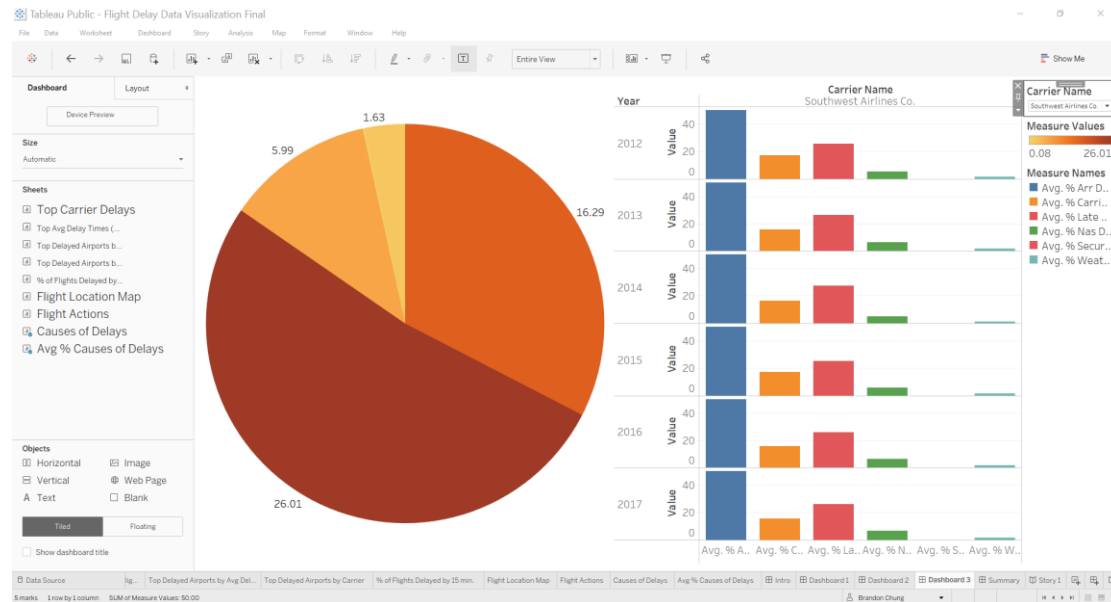

For the edited dashboard 1, I moved the bottom half of the "Top Delayed Airports by Carrier" sheet and moved it onto the edited dashboard 4 since I found it to flow with the map more. I added a more simplistic line chart segmented by year for the top average delay times in minutes per flight because it would give the readers a better understanding of the past trends. I excluded the filters for months, year and airport name reduced the amount of data overload at the start of the story.
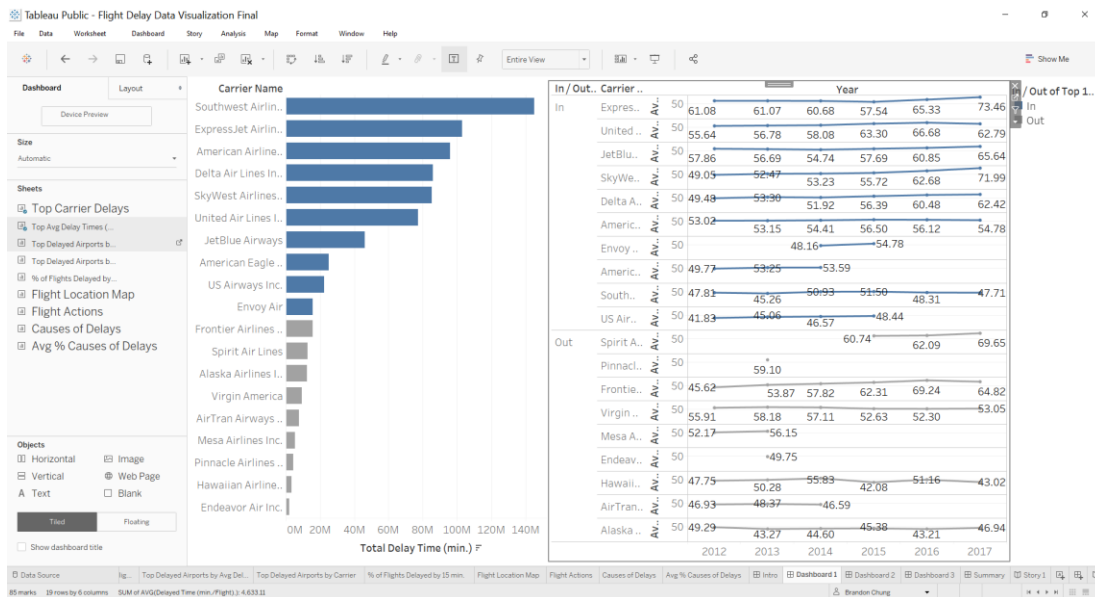
For the edited dashboard 2, I completely wiped out the content since I found they were similar in data to my other sheets but rearranged in a different view. Instead I added the top delayed airports by average delay times and moved the "% of flights Delayed By 15 min." sheet onto to the right of it as there was enough content to directly relate it back to average delay times from dashboard 1. This dashboard explains the notion of the number of delayed flights that arrived 15 minutes late amongst the total flights in the area bar below. Above the area bar is the section that would be shown as a percentage of how big or small the circles were indicating the magnitude of the delay count was by 15 min. I later added a filter in for the top 10 delayed airports as I wanted to show each individual airport when selected that would give me a rank of all airlines in order from smallest to largest. I chose these 2 particular charts because relaying the top airports on a treemap helps the reader understand which carriers actually had the highest average delayed flights and this will help make a better decision of whether to buy a ticket or not for the airliner.
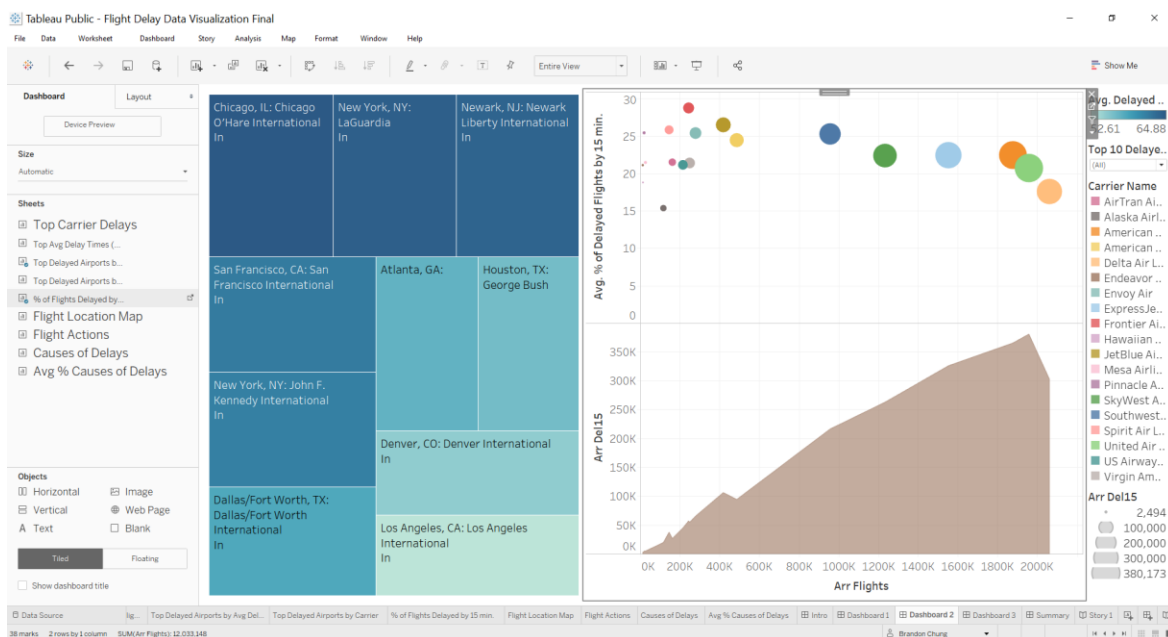
For the edited dashboard 3, I deleted out the box plot because although it was very impressive to show distribution it was not relevant to the reader. Instead I added a pie chart similar to the one in the introduction but this one was for all the causes of delays. I then added a year segment to the bar chart to show the trends overall for each carrier. These charts were chosen because it gives the reader an idea of what are causing most of the airline delays in the past and can forecast and prepare for these delays


For the edited dashboard 4, I removed the "% of flights Delayed By 15 min." sheet onto its very own tab as mentioned above in edited dashboard 2. I replaced it with the "Top Delayed Airports by Carrier" sheet as the data was closely aligned with what we wanted to visualize on the map as a summary. I matched the color pallets between the map and the treemap.
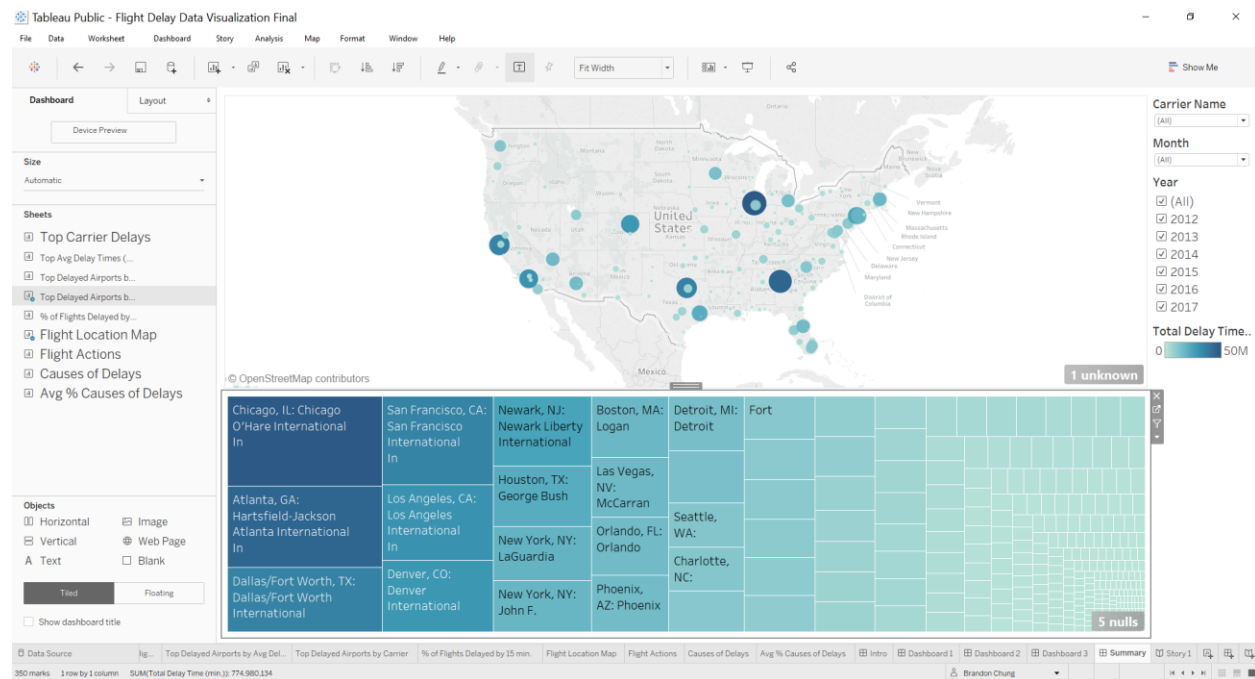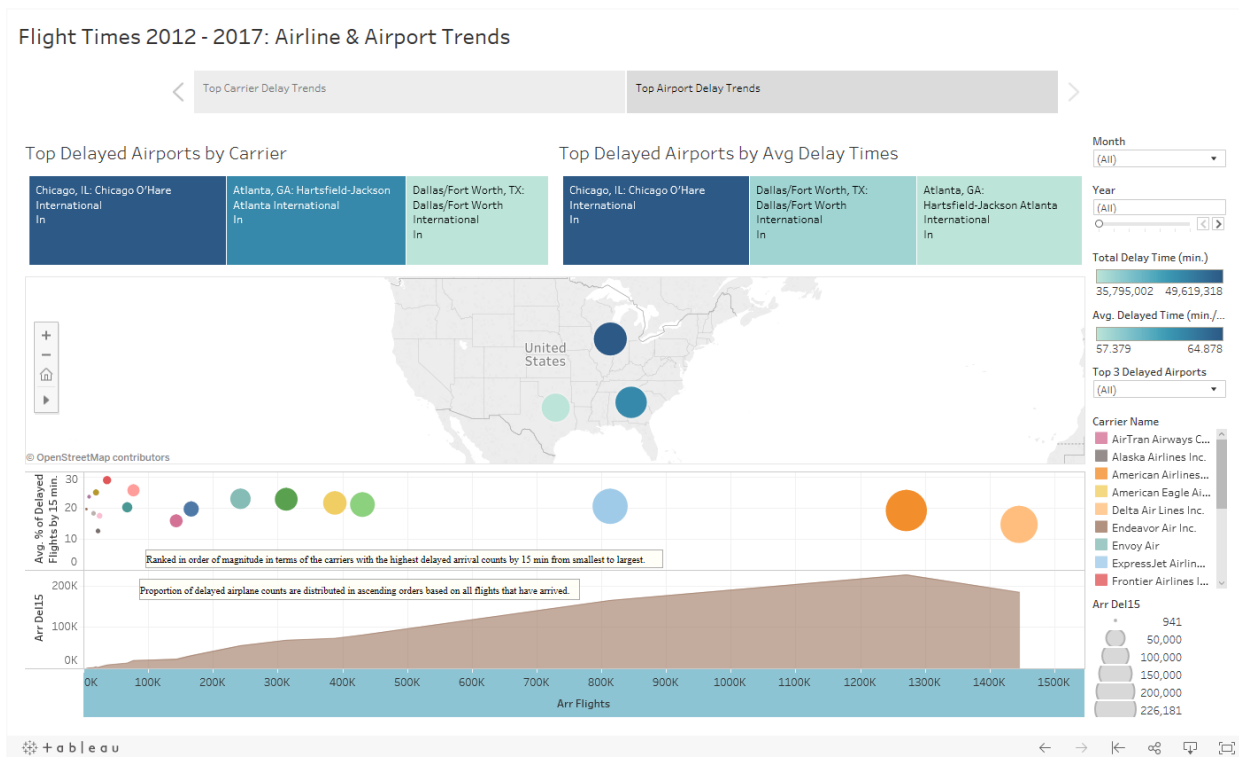
Design - Final Copy Combined Dashboard

With feedback from my reviewer as well, I decided to attempt to combine all the dashboards into a single one which I was only left with two which still provides readers with both sides of the top 3 carrier and airport trends
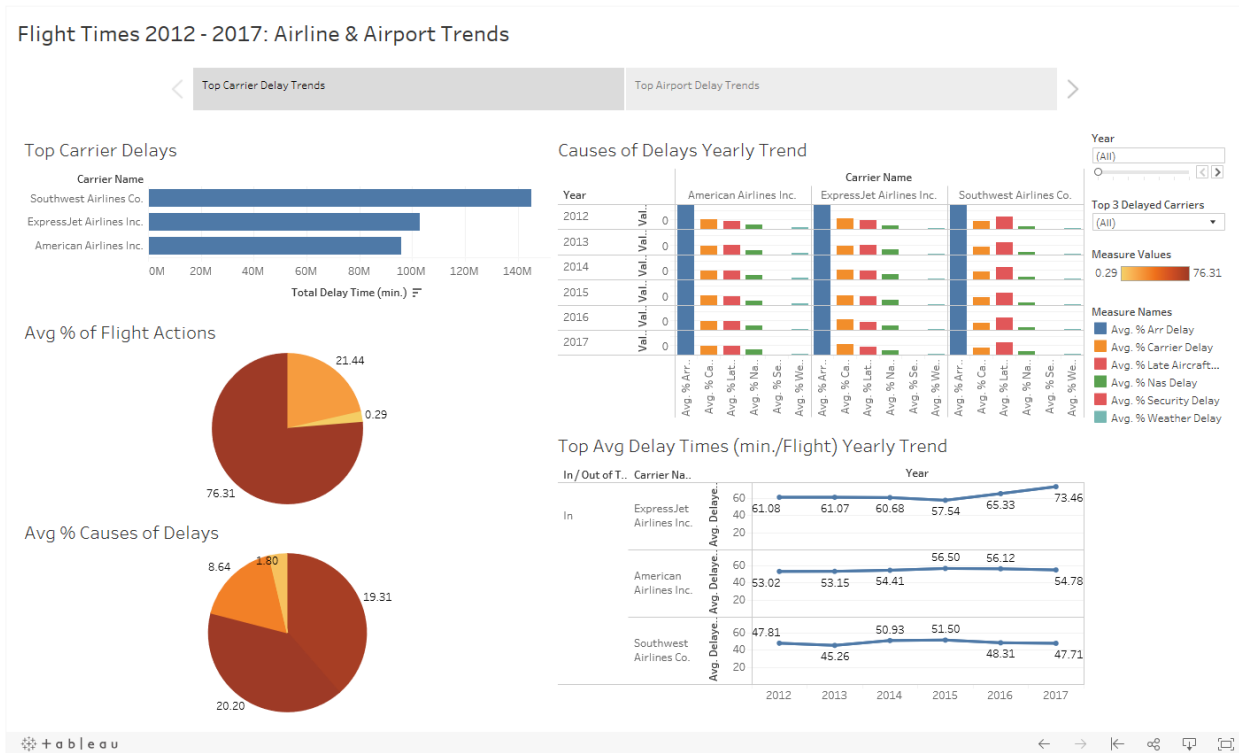
Feedback

All Threads
No new replies

brandonjchung 2 days ago
Hi everyone, I would like to get some feedback on my tableau project. What do you notice in the visualization?
What questions do you have about the data?
What relationships do you notice?
What do you think is the main takeaway from this visualization?
Is there something you don't understand in the graphic?
https://public.tableau.com/profile/brandon.chung#!/vizhome/FlightDelayDataVisualization/Story1 (edited)

Deac 🏠 18 hours ago
@brandonjchung What a terrific job you have done graphing so many combinations of data! You have really learned how to stretch the graphing ability of Tableau.

Having said that, each slide required quite a bit of time (at least for me) to understand each one. There is a top half and a bottom half of many slides and I didn't understand how the two correlated. Lots of color combinations as well. Maybe that's just my problem, nut OI do recall research describing the cognitive value of shape, color, text, etc. from our lessons.

The data visualization rubric talks about telling a story, or asking a question and answering it. Any one of your slides could be used to tell a story... all together they are trying to tell too much, at least for me. 🙂 There are too many colors and shapes that I try and correlate that don't seem to be related.

If you were to ask a question (or someone else were to ask a question) that could be answered by the data which could be answered by one or 2 of your slides instead of all of them (as great as they are), what would it be? What insight would did have doing all this work? Did you discover any unexpected relationships? How would you convey that discovery to a reader?

As I said, your work is a tour de force of graphing in Tableau, but it is a bit overwhelming and time consuming to understand, at least to me. Perhaps it could be simplified to a question and a 2 or 3 slides and a conclusion.