DATA VISUALIZATIONS WITH TABLEAU, FINAL PROJECT

STUDENT: Guillermo Naranjo

DATE: Wednesday, Nov 22th, 2017.

SUMMARY

This visualization was created to explain key attributes related to delays and cancellation in U.S flights in 2008, those major events that go against availability and time performance while traveling by air. The visualization only uses U.S. local flights records from 2008, carrier and airports information as well

It starts with an availability overview over the year and behavior related to weekdays. Then, it introduces the major causes associated to cancellation and delays. Follows a presentation of three main delay issues: weather condition, NAS and carrier delays with information to prevent those delays either by deciding what carrier to use, where and when to go. Finally it suggests some airlines that have a good performance, delay wise.

Dataviz in Tableau Public:
https://public.tableau.com/views/DAND-DataVizProjectGuillermoNaranjo/FlightStories?:embed=y&:display_count=yes&publish=yes

DESIGN

Overview

To introduce the topic I used a line graph to show the changes over time in number of flight (availability). Initially I used a single line by day but it was not showing any interesting pattern. So I decided to try several dimensions to identify patterns and found that weekdays show interesting behavior. Before I had this:

FlightsByDay
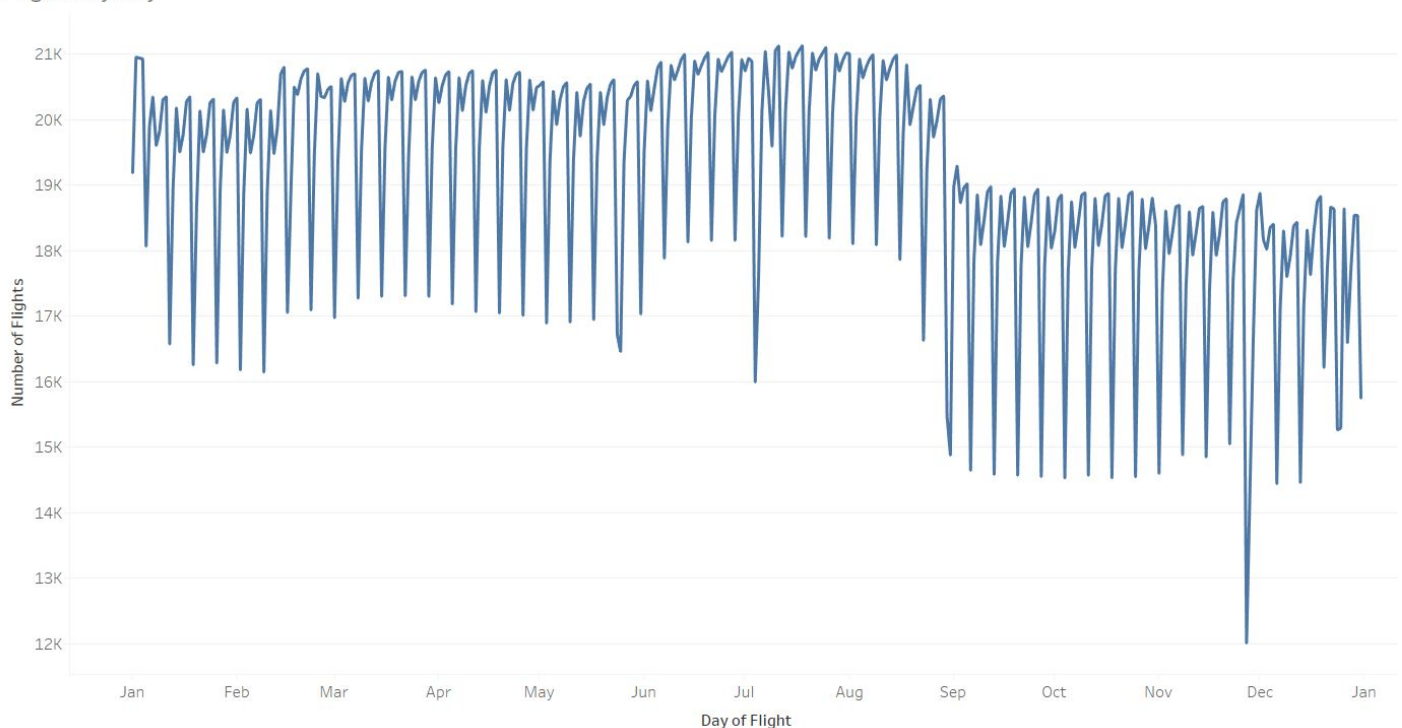
Once I showed this visualization at work several **people got confused by the line dropping consistently**, asking why. Then I found that it was a pattern with weekdays and changed it to this. Additionally, I had to explain some specific events over the year in the caption text to clarify:

U.S. flights departed in 2008 by day of the week.
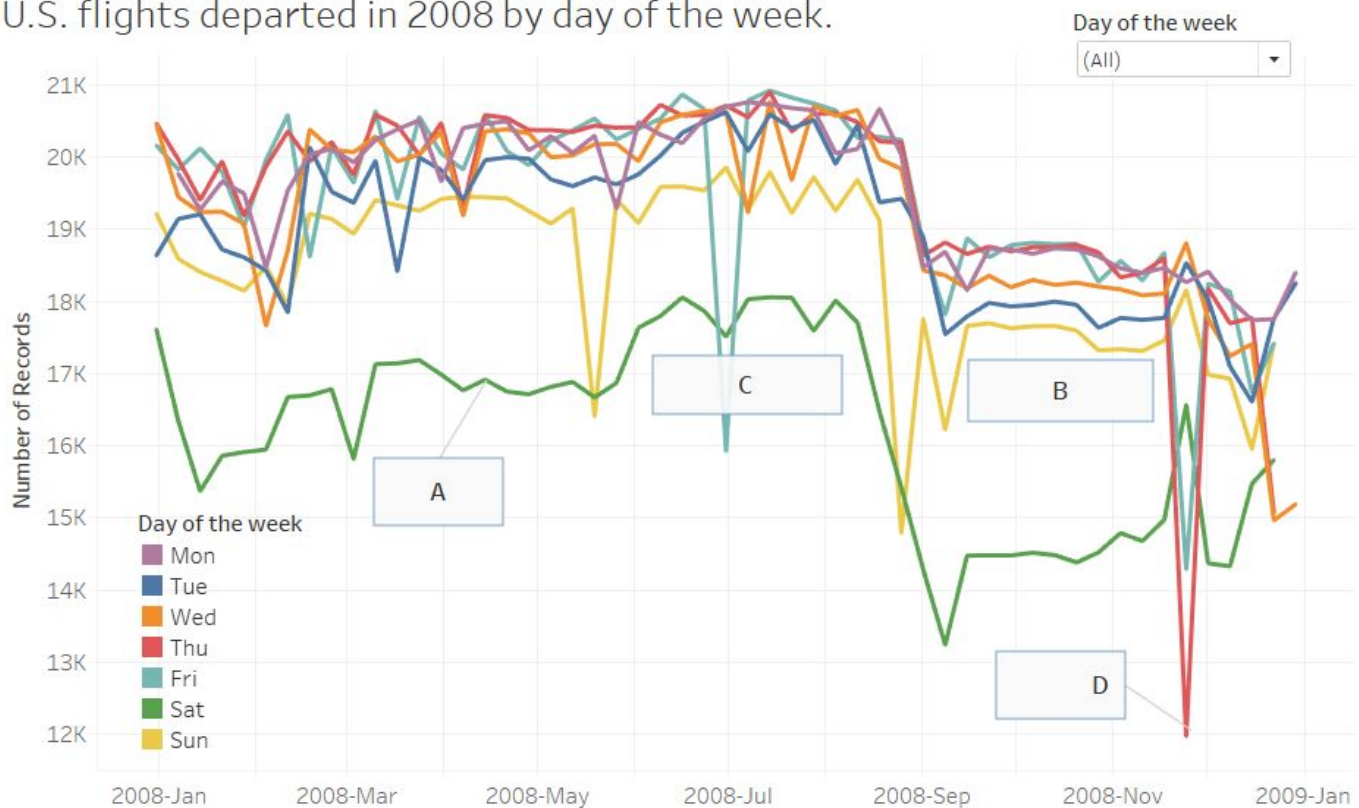
This chart shows some interesting patterns related to seasons and specific dates where the number of flights is not normal in U.S, (A) Saturday and Sundays have less flights compared to weekdays. (B). From September to December the number of flight decrease considerably. (C) On vacation season the number of flight increases.(D) In specific dates the volume of flight drops.

I used colored lines to prevent any confusion with the lines getting closed to each other.

Flights by weekdays
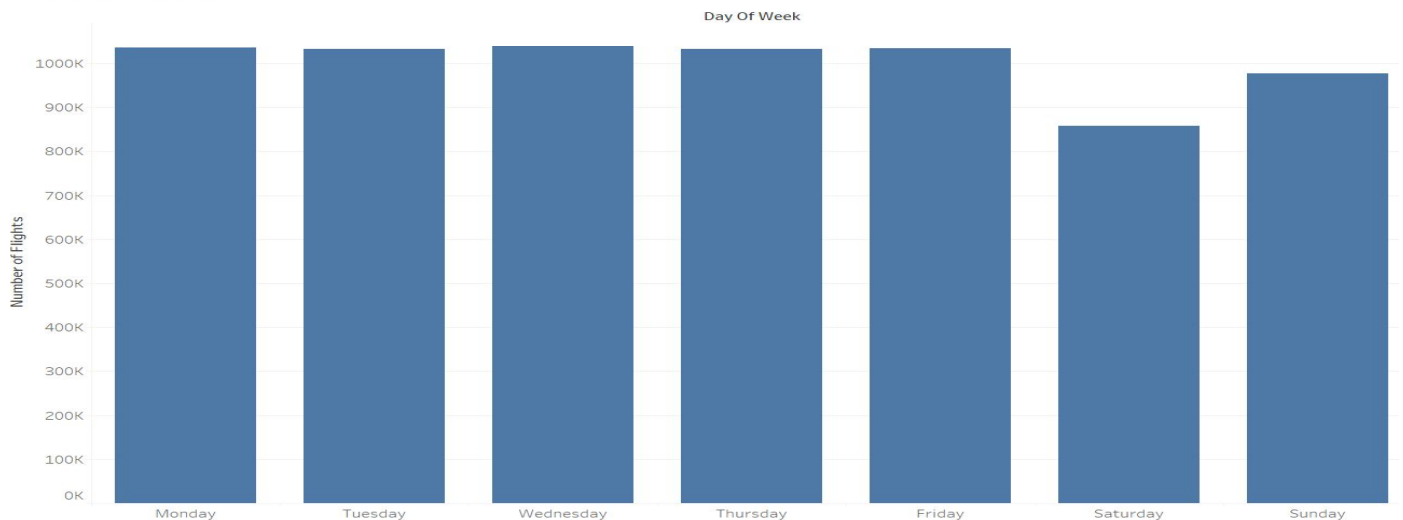
I used initially bar charts but even when **people understood** that most people travel on weekdays I loose an insight that I managed to visualize with box plots, since it also show the category distribution, possible outliers and (when added color … seasons):


Weekday flight distribution and special dates.
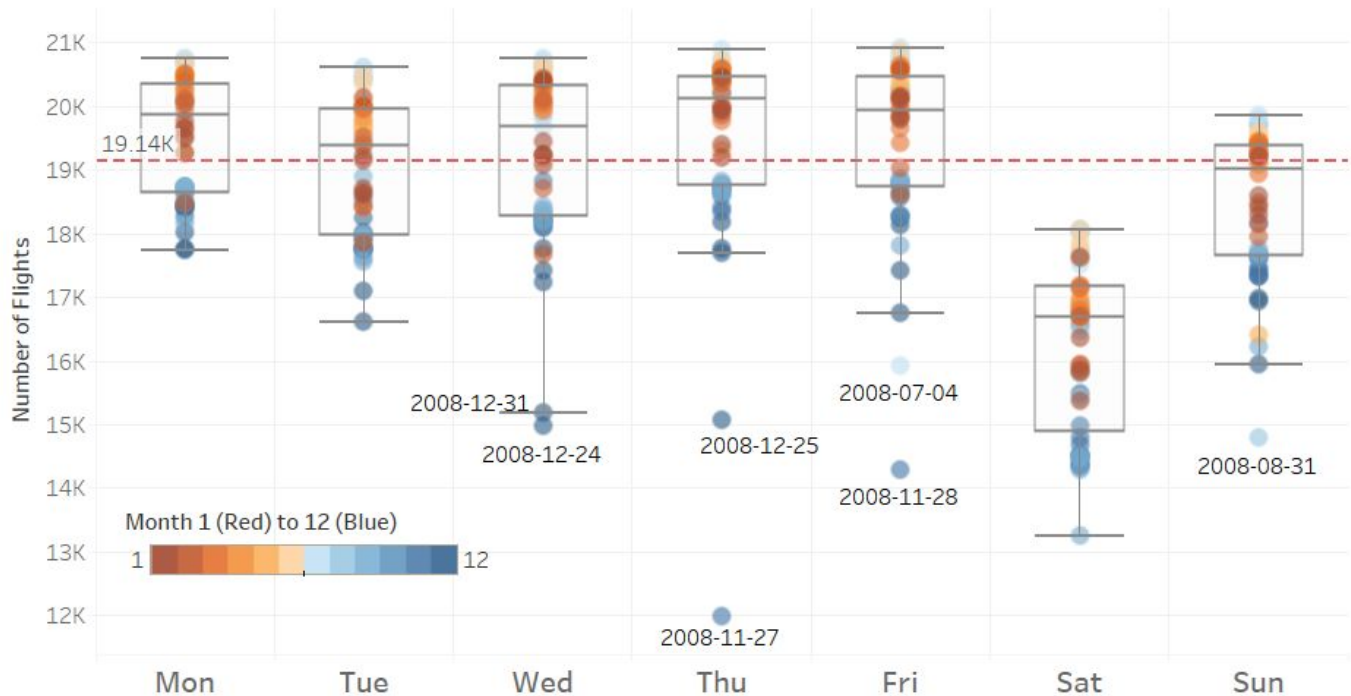
On **workdays** there is a median number of 19140 daily flights. With few exceptions been New Year, Christmas, Thanksgiving, Independence day and Hurracaine Gustav August 31th, 2008. **Saturdays** and Sunday are low on flights normally.

Colored are the months, starting January with read and moving to Dicember in blue. In this case we can see the average number of flights reduces at the end of the year.
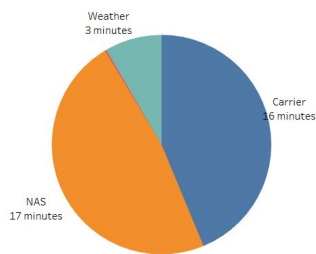
I thought the color would cause problem because I was using too many design elements at the same time but it was the opposite, **people understood that travelers flight more first half the year and less second half of the year or holidays.**

Cancellations and Delays
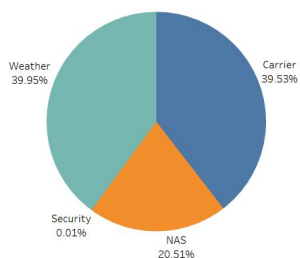
I went with two simple pie charts to explain issues share and it works, a simple chart so explain a simple concept. I used the same color for each category and **people saw the relations easily**.

Weather and Carriers usually cause flight cancellations but on the other hand National Air System and Carriers decisions are the reasons why flight get delayed. Security issues are not relevant.

Initially I tried to use line charts to explain the concept overtime with it was too noisy to draw any conclusion.

CancellationCode
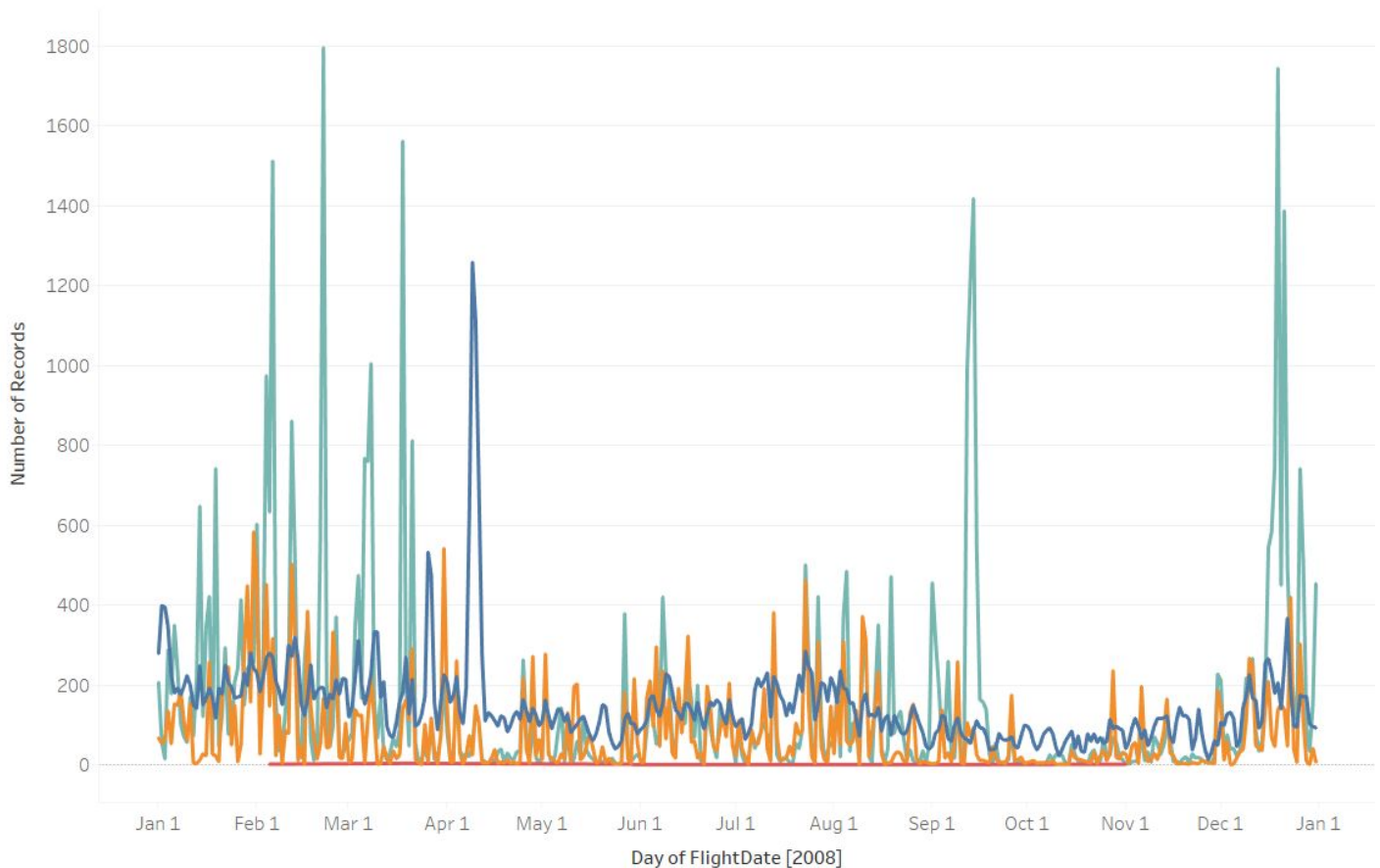
Then, I used maps to show with a red-yellow-green color code by state the impact on different categories like weather on delays for example. I also used Tableau page control to see changes overtime at once. **Maps were well received by my testers**. The same goes for NAS delays by state.

Average weather delay by state and month
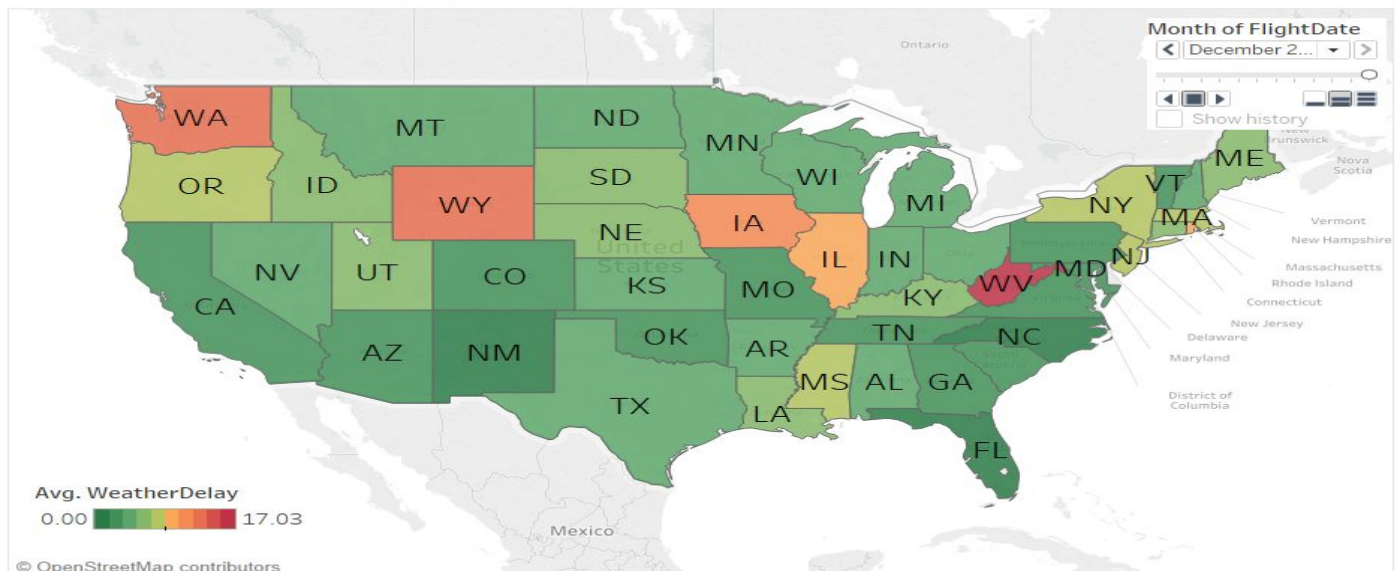
WA, WY , IA, IL and WV are states affected by delays related to weather conditions specially in Winter season.

While reviewing the DataViz a person asked if there was a relation between a late departure and a late arrival or if there is possible for the plain to go faster to reduce the impact, so I decide to use a scatter plot to how the relation. There is a clear relation and it was simple to explain. No colors or shapes were added to keep simple the relation. But a line of best fit was added to clarify even more the string relation. People took some time to see the relation.


Relation between average departure delay and arrival delay
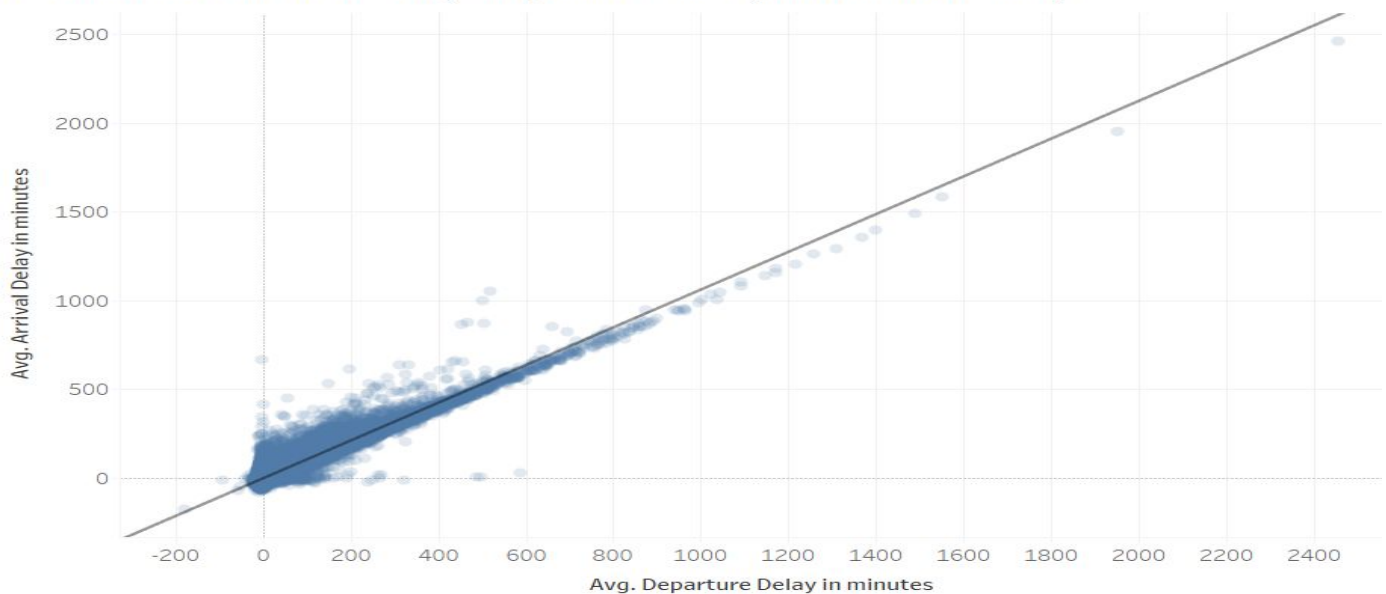
There is a direct linear relation between the flight departure delay and how late the flight arrives on destination.

Finally, I used again boxplots to show the distribution of each carrier against average carrier delay, to rank top and bottom performers.

Worst performing carriers by carrier delay. Avoid.
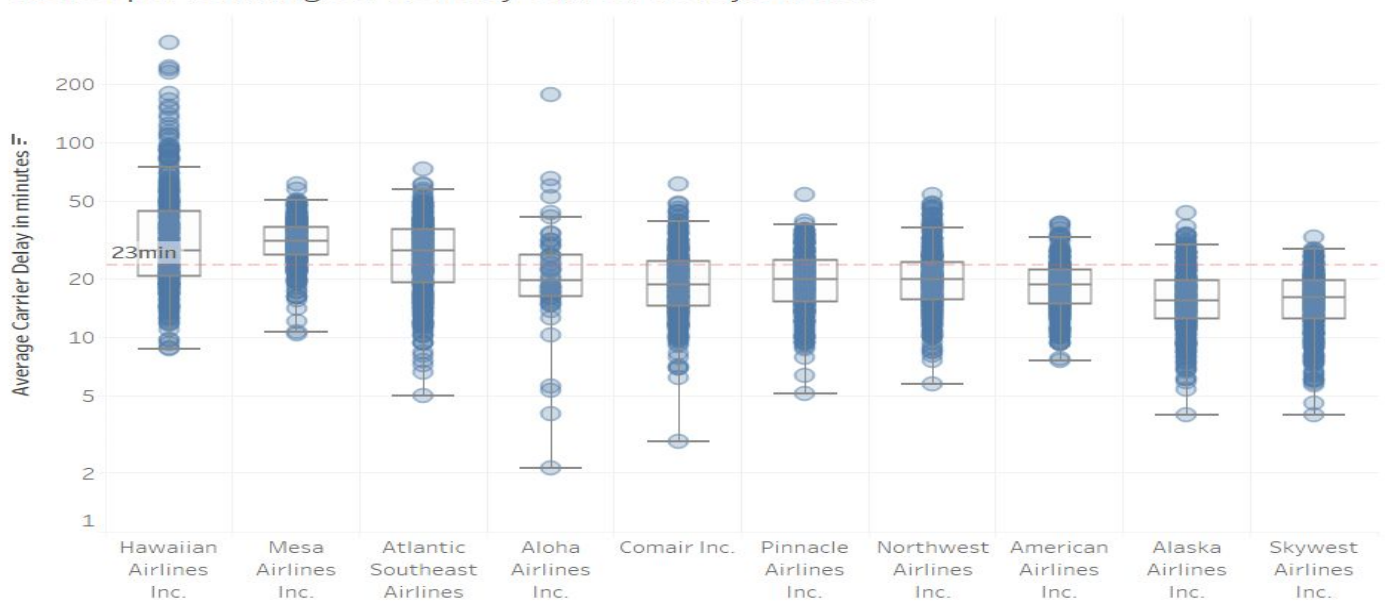
These are the worst performing carriers based on average delays in 2008. Interesting enough, there are two carriers from Hawaii in the list.

I found that I also could use a bar chart instead since there were no outliers of pattern. But I decided to go for boxplot and I did not have any issue explaining the concept to others.

I also tried to create network diagram maps to show the flight routes without success …. I then tried heatmaps and it was much data to display.

FEEDBACK

They were dashboards, showing several chats at once. You can find version 1 here:
https://public.tableau.com/views/DAND-DataVizVersion1/Dashboard1?:embed=y&:display_count=yes&publish=yes

Feedback received:
- "There is **no story telling**. Since you are showing everything it gets confusing."
- "There is a weird pattern in the "flight by day of the year" **line chart**, with lines dropping periodically, not showing why is it happening". See the graph on "overview" section above.
- "Flights by weekday" **bar chart** was simple enough to understand that there are less flights on saturday. But it´s **boring and plain** and doesn't provide more relevant information.
- **Pie chart** shows the number of cancelled flights by reason. "It was **easy to follow** and people suggested in using the same tool for delayed flights since both share same categories".
- People asked for **maps charts** as a usual way to understand patterns by region quickly.

Final version
- Once included both pie charts, **people understood the flight volume and categories easily**.
- For box plots weekdays chart, **people understood that travelers flight more often first half of the year and less second half of the year or holidays.** Moving from bar chart yo boxplot helped.
- **Departure delay vs arrival delay with the scatter** plot along with best fit line clarified even more the strong relation. People took some time to see the relation.
- Maps addition received positive comments. Nothing specific.

Additional resources
- http://kb.tableau.com/articles/howto/creating-sparklines
- https://www.transtats.bts.gov/Fields.asp?Table_ID=236
- https://onlinehelp.tableau.com/current/pro/desktop/en-us/maps_howto_origin_destination.html
- https://www.evolytics.com/blog/tableau-201-how-to-make-bullet-graphs/
- https://community.tableau.com/thread/230858