

Project 4 - "Create a Tableau Story" by Lakshmi Rajasekhar

I. Summary:

I choose the Flight Delay dataset to do the Tableau visualization. The dataset has many attributes which describe various characteristics like the airlines, airports, arrival and departure delays etc. Although multiple years' data are available, I choose 2007 as each year was big enough and took time to load into Tableau. Initially, I thought of creating visualizations from three perspectives:

- Delayed airports and causes,
- Delayed months and causes,
- Delayed airline carriers and causes.

Later, I realized that it would take many tableau sheets and dashboards to go in-depth into each of the above paths. Hence, I concentrated my story on the most delayed airlines (ExpressJet Airlines (EV)), the airports associated with this airline and investigated some causes of delays for EV.

II. Design:

Design choice of charts:

I started the story with exploring the percentage of total flights delayed/on-time/cancelled in 2007. I felt bar charts best depict the relationship between a categorical and numerical variable. I had the option of choosing bar charts or bubble charts but I finally went with bar charts. 'Bar charts' instantly gives us an idea about the gravity of a situation (here the % of delayed flights in 2007). I also decided to keep the 'FlightStatus' legend for most of the charts so that the viewer can check various combinations of on-time/delayed/cancelled flight statistics.

I added bubble chart for depicting 'FlightStatus' exclusively for Atlanta Airport and ExpressJet Airlines (EV). I chose this as in dashboards surrounded by other bar charts, bubble chart looked more compact, different and interpretable.

Whenever the charts dealt with geo-locations like flight origin/destination etc., I used maps so that the viewer can easily interpret the locations and relate.

For charts dealing with EV flight counts in Atlanta by month, I decided to add month as filter and add pages instead of adding multiple charts by month in the same page to reduce clutter.

When the data dealt with numeric only data like arrival delay and air time, I decided to go with scatter plots without aggregations as I felt this is more interpretable.

Initial version of tableau story:

https://public.tableau.com/views/Lakshmi_FlightDelay/2007FlightDelayAnalysisforExpressJetAirlinesEV?:embed=y&:display_count=yes&publish=yes

Final version of tableau story:

https://public.tableau.com/shared/QRYYYJPZD?:display_count=yes

III. Feedback:

Feedbacks gathered after initial version of visualization:

Feedback 1:

Your tableau story looks great. I really liked it. A few minor things that I think you should correct are the colors in case of bar graphs in viz 1 & viz 8

The x axis already mentions what bars are telling, it's not required to color the bars with different color

The second last slide (EV arrival and departure stats) is not quite easily interpretable. If you can add a bit more details in that slide then we can understand it well.

Overall your whole story looks great.

Changes made after feedback

I removed the redundant colors in the bar charts as suggested. But I wanted to highlight the delayed % (area of concentration) in the chart and hence annotated that separately.

I also changed last but two slides to denote ExpressJet Airlines (EV) flight arrivals and departures at Atlanta Airport. I agree that the earlier version of sum of taxi times did not make much sense about what I was trying to say (Prevalence of EV Flight delays and the gap in EV flight operations between 00:00-06:00hrs). I changed it to the count of flights arriving/departing by time and month. I also added the total arrivals/departures for all the flights in Atlanta airport by time and month for comparison. I also changed short forms like 'arr/dep' to full forms like 'arrivals/departures' in the charts.

Feedback 2:

- *What do you notice in the visualization?*

The data shown tries to make sense of the flight delays; which airlines and which airports have more delays and what could be the potential reasons for those delays. It identifies airlines with more delay, identifies EV as the airline with maximum delay and tries to investigate the reasons behind the delay.

- *What questions do you have about the data?*

Dec has the maximum flight delays. Is that because the volume of flights is higher and airports / airlines can't cope with that? Probably total number of flights, no of flights delayed and cancelled (actual numbers) also need to be shown on second graph.

- *What relationships do you notice?*

EV airlines has maximum delays at Atlanta airport both for incoming and outgoing flights. Delta airlines which is having maximum flights out of Atlanta airport, but their delay is less than EV.

- *What do you think is the main takeaway from this visualization?*

EV airlines has maximum delays at Atlanta airport even though they have lesser number of flights compared to Delta. The delays are primarily carrier related and can be solved by EV. If the departure is delayed, the arrival is also delayed, there is a definite correlation between them.

Best time to fly with EV from Atlanta is between 06:00 and 12:00.

- *Is there something you don't understand in the graphic?*

Not really.

Changes made after feedback

The main question asked in this feedback is about the increase in delays during December. Higher delays in December is most likely due to weather delays but investigating this with this project would make the story too huge to carve out. I initially started with investigating all the perspectives of delay (by Month, by Airline, by Airport) but quickly realized that it would lead to a lot of sheets.

As most of the points I wanted to make are already understood, I plan to investigate the "higher delay in December" point only later on as it would extend this story into another dimension. But I think this would be a good future work for this project.

Feedback 3:

- *What do you notice in the visualization?*

It is a performance analysis of Express Jet Airways.

- *What questions do you have about the data?*

- *Why the data of 2007? Will it be relevant at present? The conditions would have been changed now.*

- *Is there any special features for EV or at Atlanta Airport, which causes carrier delays, like long flight hrs, insufficient rest for the crew etc.*

Changes made after feedback

I added an 'Air time' versus arrival/departure delays for EV at Atlanta to the story but I did not find a clear relation between them. Although an interesting thought process, I did not find relevant attributes in the data to figure out if the carrier caused delays were due to insufficient rest for the crew. Using flight times and flight names, I can try to find the flights which run continuously. If a clear relation is found, assuming strict rules for flight crews' working hours, I can deduce that the inefficiency lies most probably with the intermediate crew. But, to keep the visualization less complex, I am keeping this as a future work.

IV. Resources:

1. Bureau of Transportation Statistics,
<https://www.rita.dot.gov/bts/help/aviation/index.html> (last accessed Jan 14, 2018)

V. Data Files:

1. Tableau story initial and final version links given above.
2. Dataset used: "2007.csv"