Data visualisation and story telling using tableau

By
Nasim Khadem
June 2017

Visualisation
Before: https://tabsoft.co/2tjraXY
After :   https://tabsoft.co/2MOyZNu

Summary
Have you ever been to an airport just to find out your flight has been delayed? If you are lucky it is not delayed for more than a couple of minutes. However if you are not so lucky your flight might have been delayed for hours or even worse cancelled. In this visualisation I show you the worst airlines, best month and days to fly and if you are the curious type like me I Show you what are the most common reason for flights being delayed or cancelled.

Design
I open the story with a map. Maps are very fun to work with and give a very general feeling of where I was going with the visualisation. However the initial design map , did not convey the message so beautifully. It was just a couple of points about airport IATA code and number of flights from that airport.
After the feedback round, I downloaded airport names and added the state and airport names to my data. This helped me to make a map and colour it according to each states number of flights. Also the map works as a filter to see airports in that state and the number of flights flown from there. A ranking was added to hoover so one could easily understand which airport rank it has.

In the second tab, I Show the airlines and what percent of the flights are done by which airline.I thought it would be interesting to see which airline has the most Flights, since one might get better prices, better hours or more frequent flies. However none of this information is available.
 I had to merge a table with the original data to get the airlines. Since the airlines were not included in the initial data file. I took the unique carrier field and matched it with the new data file to be able to get the airline names. After the feedback round I added the percent to the flights to understand which airline is responsible for what percent of domestic air traffic.

Third tab is all about cancellation. It first shows the percentage of cancelled flights. Looks like American eagle airline has 13 % of its flights cancelled. Don't want to be one of those guys ..
Then the fun begins with cancellation reasons. Seems like in cold months Dec, Jan , Feb and march most of the cancelations are due to bad weather conditions. In the rest of the year with the expectation of September, a carrier (either too late or faulty one) is the reason that the flights are cancelled.
Also seems like Friday have the highest chance of cancellation. However again this is an interactive plot and by clicking the airline you see changes in the behaviour.

After the feedback round I decided to remove the stacked Bar-plot and replace it with a side by side circle view. I felt this is easier to understand than stacked bar plots as It is always difficult to compare the results in stacked bar plot. In the side by side circle view

however one can see the very quickly that cancelation due to weather are highest in Q1. I decided to leave the lines separating different quarters as it was easier to see winter month and the trends was a bit easier to perceive. I also decide to change the "reason bar-plot " with average cancellation according to month. I felt this information was missing and in the feedback was mentioned that the "reason bar plot " does not add anything new to the story. Therefore I replaced it with cancelation per month plot.

Fourth tab is about delays. Which airline has the most delays. It is American Airline. Something to keep in mind next time you travel. It shows also both the arrival and departure delays. I showed both arrival and departure delays to show that in most cases the aircraft manages to speed up and arrival with less delays that the departure. I also showed what are the most causes for a delays. Here I choose a simple bar plot as it conveys the message very easy.
After feedback I removed lines that were cluttering the charts. I made better labels and I managed to keep the colour the same across tabs. I choose colour blind palette and miller stone.

In the fifth tab I choose a side by side circle plot to show different kind of delays. This plot had a very clear message and I liked it. Also I got good feedback about it. However I removed the grid lines. And made the colour consistent with tab 4. However I did not choose the colour blind palette for this visualisation since the values are also shown by shape so even if a person is colour blind can follow the plot. I also change the size from fixed size to automatic in order to get better views across different screens.

Overall I choose the simplicity of the plots over the complexity. I tried to convey one finding per plot as I find this making it easier for the audience to understand. I also removed any junk chart and managed high data ink Ratio. I used shape and colours only where it helped with conveying the message. On top I tried to kiss the plots (not literally I kept it simple stupid :))

FeedBack:

People want to be nice so it was a bit of challenge to get them to open up. what would you change? What does this plot convey? Can you explain the plot to me. Does the lines bother you? (This last one would not pass the mom test[1]) And then the real feed backs would flow.

1. Remove grid lines.
1. the map is pretty but it does not convey any real message.
2. There are still copy in the name of plot
3. A couple of typos.
4. Airport most traffic: I like the spatial viz on a map. But this alone makes it hard to answer the question: Which airport hast the most traffic? It would be good to additionally have a bar-plot or simple ranking of the airports by traffic including the traffic number (e.g ATL: 414,513, JFK118,104 ....) on the side or below the map. Bonus points if it updates with the selection on the map.
   (I designed a new map)

[1] https://www.amazon.com/Mom-Test-customers-business-everyone/dp/1492180742

5. Airline is Good. Maybe add percentages of total traffic to put it better into perspective: e.g. How much larger than the others are SW Airlines?
   (the percent was added to the plot)

6. Top Bar-plot : What is the unit on the x axis, percentage of flights cancelled? Please label it
   (was labeled)

7. Cancelled flights: Shows all the needed information i.e. cancellation due to weather decrease in summer months. Instead of stacked bar-plots you could try plotting it simply as lines (with dots at each month). But this is just a matter of personal taste (I don't like stacked bars). [Something similar to Tab5]
   (I also don't like stacked bars , they are not so easy to understand. I changed it to another visualisation similar to tab 5)

8. Number of flights: Not sure what this plot adds? also the red is labelled 'No Information' in the legend.
   (I removed the plot completely. No information means the flight was not cancelled. I excluded that from the cancellation plots.)

9. Cancellation by day of the week: Could be an interesting plot. It would be better to always fill all days (even if cancellation is 0), right now the labels are jumping between different airlines.
   (This I decided not to change. Since it is interesting to see the the break down for each of the airlines. And see if airlines have different reasons for cancellation. However I decided to add a filter so it makes it easier to select all the airlines. In the initial plot all the airlines are chosen.)

10. On the map tab , it makes it easier if you have the rank.
    (I added the rank to details, if you hoover over it you see the rank)

11. On Tab4 : x-axis are hours? Maybe add label to the plot.
    (Added the labels)

12. Tab 4: I like the Christmas effect, which is probably potentiated by the winter weather effect.
    (thank you)

13. Plot Cancellation reasons: does not add any new info
    (I replaced this plot by cancellation per month.)

14. Use different colours to the right plot. At first sight I thought it was delays because of arrivals and NSA.
    (made the colours across tabs consistent)

15. Also how come there are a lot of negative delays in October?
    (added annotation and text field to explain)

Struggles
The tableau public did not let me save my work book . Since I have changed fields, it needed an data extract and tableau public will not let you have data extracts. I had then to install tableau desktop free trial and redo all the visualisation, since there is no option to save locally on tableau public. Therefor I had to re-plot everything. Better to give all the students a license for the duration of the course. Or warn them from the tableau public :)

Future work
I hope to do another story with more details for year 2017 and 2018 with more field added.

Resources
http://stat-computing.org/dataexpo/2009/the-data.html
http://stat-computing.org/dataexpo/2009/supplemental-data.html
https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp
http://aspmhelp.faa.gov/index.php/ASPM:_Individual_Flights:_Definitions_of_Variables
https://onlinehelp.tableau.com/current/pro/desktop/en-us/
publish_workbooks_tableaupublic.html
https://community.tableau.com/thread/241919
https://kb.tableau.com/articles/howto/creating-combo-chart-that-shows-more-than-two-measures
https://community.tableau.com/thread/219666
https://onlinehelp.tableau.com/current/pro/desktop/en-us/
functions_functions_logical.html
https://public.tableau.com/en-us/s/blog/2014/07/story-points-sizing-tips