

Create a Tableau Story Write-Up

Summary

The main objective of the project was to create a data visualization using Tableau that tells a story or highlights trends or patterns in a data set. The main dataset used in the project contains information on the United States flight delays and performance in 2008 and comes from RITA. In the project, I have focused on highlighting general trends and patterns especially important from a passenger perspective, and factors minimizing risks related to delayed or cancelled flights.

The story is [published to Tableau Public](#).

Design

Dashboard1

The first view shows the carriers' landscape in the US. Three big numbers and a horizontal bars chart are used to show the 'market share' by carrier, by number of flights and number of hours spent in the air. Consistent, blue palette is used and different font sizes to make numbers more stand-out.

The horizontal bars chart allows easy comparison of different carriers' potential as the bars visually show proportions. The colors on bars add a second dimension to the picture showing that some airlines (like Skywest Airlines) have more shorter flights comparing to others (like United Airlines) – darker color. Combining the two makes the simple bar chart more interesting (and simple) to analyze.

Dashboard2

The view shows some patterns from two time-related perspectives (daily and monthly), with colors differentiating carriers (line charts for continuous data comparison). The view should be looked at with filters applied as displaying all information at once may not be much readable (the view is preset by default with some filters).

Using line charts allows showing trends over time but also allows comparing multiple analysis subjects at the same time and compare them in a quantitative way as well as regarding the shape of changes (over time and to each other).

Dashboard3

Three charts reveal information regarding delays and cancellations. Line charts show average and maximum delay trends and a cumulative bar chart shows clearly the distribution of cancellation reasons in each month.

The Average Delay chart compares arrival and departure delays and using a line chart allows overlying clearly the two types of delays and showing that their shapes are very similar, colors are used to easily distinguish between them. Maximum Delay line chart is a very simple yet powerful way of showing the delay trend over time and values at certain points of time. The cumulative bar chart is good for comparing quantities of cancelled flights and their distribution among different cancellation reasons. Again, visual clarity is key, and this chart allows elegant presentation of these both characteristics at the same time in an easy to comprehend manner.

Dashboard4

Here is another view on delays, state-based average distribution. For such kind of information (geographical) a map seems to be a good choice. Below a box plot chart where more in depth information can be revealed regarding each state and city/airport.

Using maps and color overlays allows viewers to quickly spot the most interesting areas and easily distinguish between different quantities (in a more categorical way (bucketing) – but it is sufficient in this case). Visuals are in general easier to comprehend than numbers. The box plots below show 5 summary statistics and are excellent at showing which states are outliers in terms of average delay. For example, although Chicago could have a bad fame with its outliers (arrival delays), the box plots allow to see broader picture easily/visually, that in majority of cases the median of delay times is around 10 minutes and there are even cases (quite a few) where planes arrived to Chicago ahead of schedule.

Dashboard5

The main objective of the view is to compare arrival and departure delays. To do this, side by side data is presented allowing to see with ease that there is not much difference between arrival and departure delays. Graphical (map, color, size of points) and numerical information (basic descriptive statistics and bubble size legend) is used to complete the picture.

The design decision behind this dashboard is the same as for dashboard4, a side by side composition of exactly the same components allows easy visual comparison and different ways of showing the similarities (visually on a map, numerically and on a colorful strip) help different viewers to use the easiest for them way of communication (some prefer visuals, other numbers). People are better in comparing things than quantifying them.

Dashboard6

To confirm correlation between arrival and departure delays and see the correlation in a 'by carrier' view scatter plots (and colors differentiating carriers) are used. An appropriate legend allows better understanding.

There is no better way to show correlation between two variables than by presenting them on a scatterplot chart.

Dashboard7

Shows flight paths with cancelled flights. As the information involve different airport locations and lines connecting them, a map view is the best choice for visualization.

The path view was a little bit challenging as to make it work in Tableau, the data had to be slightly transformed in a way that there should be origin and destination points defined in separate dataset rows and they should be connected and ordered together. The paths on a map revealed complexity behind all those flights connections and again made it easier for viewers to comprehend the data behind the flights by using visuals (connection lines on a map).

Whenever possible and makes sense I made the dashboards interactive, so it is possible to play with data and find other insights than presented in the summary boxes in the Tableau story. The same applies to filters which by default should support the summary description for each story screen.

Feedback

There were 2 people asked for feedback with the following result:

Person 1

Feedback	Action
On Dashboard2 change the filter name from 'Manufacturer' to 'Airplane Type' for better clarity.	Change done.
On Dashboard4 change the map title from 'Average Delay by State' to 'Average Arrival Delay by State'.	Change done.

Person 2

Feedback	Action
There should be less text in Tableau summary boxes.	Text reviewed and shortened whenever possible.
On Dashboard1 change colors of the big numbers to blue.	Change done.
On Dashboard2 change the chart title from 'Flights Monthly' to 'Flights Monthly per Manufacturer'.	Change done.
On Dashboard2 it would be nice to know why American Eagle Airlines stopped operating in September 2008.	I have checked and it seems the lines bankrupted and then emerged under

	American Airlines. I have added this information to the summary.
On Dashboard 3 add to summary a comment about Security (not influencing much flights cancellations).	Change done.
On Dashboard 6 add to summary a comment about Aloha Airlines arriving and departing ahead of schedule.	Change done.

Resources

1. [Combine year/month/day into a date column](#)
2. [Resolve encoding error for datasets after year 2000](#)
3. [Data blending in Tableau](#)
4. [Create maps that show paths](#) and [here](#)
5. [Calculated field of percent of total](#)
6. [Solution to opening a .twb file in Tableau 2018.3](#)

Data Files

Source	Comments
2008.csv	Main dataset downloaded from RITA, downloaded from http://stat-computing.org/dataexpo/2009/the-data.html
Airports.csv	Supplementary file with airport names from http://stat-computing.org/dataexpo/2009/supplemental-data.html
Carriers.csv	Supplementary file with carrier names from https://exploratory.io/data/kanaugust/9876993836892339
Plane-data.csv	Supplementary file with plane manufacturers data from http://stat-computing.org/dataexpo/2009/supplemental-data.html
FlighPaths2008.csv, FlighPaths2008_summary.csv	Transformed 2008.csv file to be rendered by Tableau in a way paths can be showed on a map (Dashboard7). Prepared using FlightPaths2008.ipynb.