Flight Delay data visualization using Tableau

Project website:

https://public.tableau.com/profile/eyob4356#!/vizhome/projectsh1/Story1?publish=yes

Summary

In this project, I designed explanatory visualization of flight delay data downloaded from BTS website (https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp) using Tableau. In the visualization, I showed that the major delay cause is arrival delay in USA aviation system. Even though flight delay depends on carrier and airport, it is found that the overall number of delayed flights is decreasing over time and more than 95 % of the flights were on time flights. From the evaluation of historical data, it is found that AA (American airline) has the most diverted flights at DFW and MQ has the most canceled flights at WQ airport whereas HA airline has the best performance in doing on time flights.

Design

Even though the data has many features, I used few of these and calculated some that I believe are important for my visualization. The following are the features.

1. Numerical continuous variables: Year, Number of flights, Delay time.
2. Categorical variables: Airport, Carrier.
3. A categorical variable with continuous value: Delay causes.
4. Calculated continuous variables: Flight delay (%), on time Flight (%).

Initial design decisions:

My initial design decision was made based on the type of data I was dealing and the history I want to tell. In my effort to see the trend of number of flights, delay times, number of delays by different causes, flight performance (one time and delayed flights) over time, I choose line plots that effectively display the trend of a series of data points over points or bar charts, which emphasize the individual points they encode than data trends. Bar charts are superb to encode categorical variables especially if we want to focus primarily on individual values and support the comparison of one to another. Even though I considered bar charts to encode categorical variables in my initial design, I found that this design is not appropriate due to the nature of my data. One of the main reasons for the inappropriateness is a large number of categorical variables with an extended range of distributions often difficult to visualize in a single bar chart so that it fails to provide the complete information designed to provide. Hence, instead

of bar charts, I decided to use area and planar designs with appropriate color hues since they are dynamic and attractive design elements. Since area (bubbles) are less effective from an information design view at allowing an audience to make effective comparisons, I used a trick to overcome this problem by adding values as text, hover or tooltip so that my visualization blends both beauty and information to provide a complete information.

Design:

1. Line plots to display the trend of continuous variables over time (Number of flights, Delay time, Delay cause values, Flight delay (%), On time Flight (%)).
2. Colored legends to display categorical variable with continuous values.
3. I decided to use size and planar variables for presenting the number of flights by airport, where the x and y-axis represent the latitude and longitude of the airports in USA map, and the size of the circles on the map represent the number of flights.
4. I used packed bubble to plot Number of diverted and canceled flights by carrier and airport, where the size of the bubbles represents a number of canceled flights, the color hue number of diverted flights, airport and carrier are represented by text.
5. I decided to use size and color hue to display carrier performance where size represents the percentage of on time or delayed flights and color hue to display the difference in performance of carriers that have slightly similar performances.

Feedback

Here was the initial project
https://public.tableau.com/profile/eyob4356#!/vizhome/Initialdesign/Story3
Feedback for visualization:

1. I see that the total number of flights has decreased. This is clearly visible
2. I can see the concentration of the flights on the US map and get a general idea about the locations.
3. Arrival delay is the biggest reason for the delays in US aviation industry and leads by a big margin.
4. The Flights delay in minutes overtime plot has the filters in the middle of it. That's distracting.
5. The total delay times in minutes visualization" has too many colors and it's difficult to interpret what's going on. Some legend or the visualization of top 10 airlines/airports will help.

Thanks Ashutosh

Feedback for visualization project:

- Develop a more coherent storyline in the top boxes.
- Group similar graphs onto one slide, for example, flight situation on first slide, delay info on second slide etc.
- Resize map size to fit worksheet size.
- Avoid using the same type of graph twice Avoid using filters on the graph Optimize color scheme on last slide

George Liu

Corrections Based on feedback `

1. The Flights delay in minutes overtime plot has the filters in the middle of it. That's distracting. *The filters are moved out of the inside of the map.*
2. The total delay times in minutes visualization" has too many colors and it's difficult to interpret what's going on. Some legend or the visualization of top 10 airlines/airports will help. *This visualization is removed and replaced by another more informative feature and design (On time flights (%) and delayed flights (%))*

- Develop a more coherent storyline in the top boxes. *I added coherent storyline in the top boxes*
- Group similar graphs onto one slide, for example, flight situation on first slide, delay info on second slide etc. *Similar graphs are grouped onto one slide in a coherently fashion*
- Resize map size to fit worksheet size. *Map size are resized to fit worksheet size.*

In general I effectively addressed all the feedbacks given and executed a design change wherever necessary.

Resources

1. http://www.perceptualedge.com/articles/ie/the_right_graph.pdf
2. http://www.perceptualedge.com/articles/misc/Graph_Selection_Matrix.pdf
3. https://solomonmessing.wordpress.com/2014/10/11/when-to-use-stacked-barcharts/