Creating a Tableau Story:

United States Flight Delay and Performance

Introduction

An investigation was done on U.S domestic flight performance over a two year period, from January 2007 to December 2008. Visualizations were created using Tableau Public to showcase various finding. The data was obtained and downloaded from ASA Sections on: Statistical Computing at http://stat-computing.org/dataexpo/2009/the-data.html

The variables that were mainly focused on were:

1. Cancelled flights across the different airlines in the U.S.
2. Delayed flights across the different airlines in the U.S.

Design:

My final draft of the Tableau story can be accessed here: https://public.tableau.com/profile/khrystyne2616#!/vizhome/CreatingaTableauStoryUnitedStatesFlightDelayandPerformance/Story1?publish=yes

The main design choices to display the data was bar graphs and line charts. Multiple new variables (in the form of calculations from existing variables) were created. For example, percentage of flights cancelled was a calculation created in order to show the range of differences across airlines, in addition to the sum of the number of flights cancelled across airlines. In the first Visualization, percentages were also created to showcase the percentage of overall flights each airline had in relation to each other. A pie chart was also included to demonstrate the ratios of flights across airlines. Flight delay causes were visualized using line charts in order to show what caused flight delays over time (during months and days of the week).

Bar graphs were used to demonstrate which individual airlines were mainly responsible for each cause of delay, and then a line chart was used in conjunction with said bar graphs to show which top 6 airlines had the most delays by each cause of delay. Also, an average line was incorporated in order to indentify which airlines were above the average line. This was done to show readers which airlines were performing poorly, and during which months, in comparison to other airlines.

After posting my draft in the student hub and recieving feedback from my peers, I made the following changes to my visualizations:

1. I changed the colour pallete for each visualization as I failed to apply one of the most important data visualazation rules; and that is to ensure that all vizualizations are clour-blind friendly. My initial draft included the colours red and green together on the same graphs, I then changed the colour pallete to the built in color blind friendly pallete: Color Blind.

2. I had initially created a graph for months where the x-axis displayed months from 0 to 13 and my graphs for days displayed days from 0 to 8. I edited the x-axis on both graghs to correctly display months and days from 1 to 12 and 1 to 7, respectively.

3. On the Average delay graph, the airline is identified on the x-axis, while various colours (on each individual bar) also represented each airline, this was not necessary. Therefore, I changed the colours to one standard color throughout the bars instead of having each bar represent an airline by colour. This was done for readability.

4. On the "Reasons for flight delays" graph, I changed the chart to appear in a descending order so that the main reason for flight delay appears first.

Feedback:

My first draft of the Tableau story can be found here:
https://public.tableau.com/profile/khrystyne2616#!/vizhome/Storyforflightcancelationsdelays/Story1?publish=yes

1. We should avoid using the colors green and red together because colorblind people will have difficulty analyzing the graph. See this article on this: https://www.tableau.com/about/blog/2016/4/examining-data-viz-rules-dont-use-red-green-together-53463

2. The chart that uses months in x-axes shows 0 and 13 although these numbers are not valid months. Please remove them. The same thing happens with days of week.

3. On my side I think it is a really complete analysis. Here are additional recommendations: 1) I have been quite confused with the "average departure and arrival delay" graphs. On the first 2 charts, there are a lot of colors to identify the airline companies. But are they really necessary? (Especially if the name of the company is written on the x axis). I preferred the third graph at the bottom, much easier to read.

4. On the "Reasons for flight delays" it would have been easier to present the bar chart with a descending order so that the main reason for flight delay appears first.

Summary

Southwest Airlines co. has a large number of flights in comparison to the other airlines. This can clearly be seen in the first visualization where flights are broken down into overall percentages of flighs by airline. Southwest Airlines has 16.39% of all flights, whereas the airline with the second largest number of flights only has 8.56% of flights.

Airline carrier is the main factor responsible for the most flight cancellations across all airlines. Weather is the second cause for most flight cancellations, NAS is the third, and security is last. We found that American Eagle Airlines Inc. had the largest number of flight cancellations over a two year period and even though airline carrier was the main cause of cancellations across all airlines, in this visualization we can see that the main cause for American Eagle Airlines Inc.s' cancellations were caused by the weather (overall 48.98% of flights). The graph shows that the airline carrier (overall 21.21% of flights) was least responisible for this airlines' flight cancellations, that is, after security. Skywest Airlines Inc. had the highest number of no cancelations, and therefore, the least number of flights that were cancelled.

The main reason for flight delay across all airlines was Late Aircraft Delay and security delay causing the least flight delays. It seems as though overall airline flight delays did decrease from 2007 to 2008. Looking at the overall average flight delays, Atlantic Southeast Airlines had the longest delay in minutes that the other airlines. The most flight delays accured in December and the least delays accured in September across all airlines. Flight delays peak on a Friday, and the least delays happen on a Saturday, with the exeption of airline carrier that displayed the least delays accuring on a Tuesday.

We looked at the different reasons for flight delays individually to distinguish which airlines had the longest delays by reason. We found that the airline with the most flight delays caused by weather was PSA Airlines Inc. Alaska Airlines inc had the longest delays caused by security. JetBlue Airways had the longest delays caused by their air system, JetBlue Airways also had the longest flight delays caused by Late Aircraft. Atlantic Southeast Airlines had the longest delays caused by the airline carrier.

In conclusion, the visualizations show us which days of the week, and in which months, would be best to fly to avoid delays. It is also helpful to know which airlines have the most delays and cancellations, in order to choose an alternative. It is important to note that the smaller airlines have the lowest number of overall flights and, subsequently, the least number of delays and cancelations. However, this does not mean that airlines with a larger number of flights will have the most delays and cancellations. If we take a look at Southwest Airlines co. for example, this

airline has the largest number of flights across all airlines, but did not have the highst number of flight cancellations. It had 6.28% less flight cancellations than the airline with the most cancellations at 13.79%. Also, Southwest Airlines co. was the third airline with the least cancelled flights. However, They do have the highest number of flight cancellations due to security, this may be problematic for some given todays climate of security threats.

Resources:

Calculated percentages: https://community.tableau.com/thread/195678

Types of calculations in Tableau: https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_calculatedfields_understand_types.htm

Creating a "Percent Of Total" Calculated Field Without Using Table Calculations: https://kb.tableau.com/articles/howto/creating-a-percent-of-total-calculated-field-without-using-table-calculations