Comparing U.S. flight data in 1988 and 2008

Eileen Hertwig

March 15, 2019

Links to Tableau Story

First Version

https://public.tableau.com/shared/6SC8P2JQJ?:display_count=yes

Final Version

https://public.tableau.com/shared/C4RC2SXXW?:display_count=yes

1 Summary

In this Tableau Story I compared U.S. flight data from 1988 to 2008 and found the following main result: The absolute number of flights increased in this period as did the absolute and average time of delay. The proportion of cancelled flights has increased over time, especially in the summer months. Less flights get cancelled on weekends and delays are the worst on Fridays.

2 Design

Slide 1: Total number of flights

I started out with looking at the total number of flights. I used barcharts for both years and show data for each month individually as well as a line for each years monthly average. The different heights of the bars clearly show that the number of flights have increased. At first I used linecharts, but after removing a graph showing the total number of flights by day of the week (it didn't really offer any insight since there wasn't much fluctuation), the slide appeared too empty, so barcharts were the better option.

The individual month show that this is not due to a few outliers but a stable feature over the entire year. A map shows the airports with the most flights leaving. A filter can be applied to look at the two years individually.

Slide 2: Delays by location

With this slide I at first just wanted to show that it hardly matters if we look at arrival or departure delay time (because they are very similar). I did this with side-by-side bar charts again monthly for each year. So from this point on, I just used the arrival delay for showing delays.

In the beginning I had two maps showing the total time (summed) of arrival and departure delay. However, they looked basically the same and I had made my point already with the barcharts. It was pointed out to me that the total delay time can be misleading since some airports have much more flights than others. I substituted the departure delay map with a map showing the average delay time (now just arrival delay for both). So, average and total can be compared.

Slide 3: Differences in delays (2008-1988)

In this slide I show differences between the years 2008 and 1988 for the delay time. The first plot is a barchart for each month, so we can see that while delays have increased overall, this is not true for every month (to make this statement stronger we would need multi-year averages, at the moment we just compare two years).

The next plot shows differences in average delay time per airport. While for most airports this time has increased, for some it has decreased. It was pointed out to me, that my first visualization did show this clearly, so I color coded positive and negative values in both plots.

Slide 4: Cancelled flights

Here I look at the flights that were cancelled. The first plot is a side-by-side barchart showing the total number of flights and the number of cancelled flights for each year next to each other just to get a visual for the proportions.

Below I show the proportion of cancelled flights per month. This visualization clearly shows that much more flights are cancelled in winter (probably due to weather), however, the difference between summer and winter was much greater in 1988 than in 2008 (to get a deeper insight more years and weather data would be needed).

A map showing the proportion of cancelled flights completes this slide.

This slide just got some layout changes after the feedback that some legends are not really readable (I changed everything to a fixed size).

Slide 5: Statistics by day of the week

As a last slide two stacked bar charts show the proportion of cancelled flights and the average delay time for each day of the week. The years are again color coded.

After the feedback I changed the total delay time to average delay time.

3 Feedback

These are the points that I received as feedback:

- Improve titles, axes labels and legends to make the plots better understandable.
- In many cases it could be better to use an average delay time and not the total sum, because there were more flights in 2008 than in 1988 and some airports have more flights than others.
- Change the size of the story, so that it displays well on all screens.
- Slide 3: indicate more clearly if there are negative differences in the delay times.

4 Resources

The data is explained on the following page:
http://stat-computing.org/dataexpo/2009/the-data.html