**Tableau visualization for flights**

**Initial story:** [**https://public.tableau.com/profile/ilaria2704#!/vizhome/Flightanalysis\_0/Story1?publish=yes**](https://public.tableau.com/profile/ilaria2704#!/vizhome/Flightanalysis_0/Story1?publish=yes)

**Final version:**

[**https://public.tableau.com/profile/ilaria2704#!/vizhome/Flightanalysisv2/Story1**](https://public.tableau.com/profile/ilaria2704#!/vizhome/Flightanalysisv2/Story1)

Summary

The data used in this exercise comes from [RITA](http://stat-computing.org/dataexpo/2009/the-data.html) and contains information about 2008 flights delays and performances mostly in the US. The dataset is almost 700MB and definitely has some interesting facts. Before starting, I decided to manually look for other data to extend the current dataset. In fact, airports were mentioned just using the IATA code, making it complex to understand it quickly. Merging the code with the city names makes the data more understandable.

Design

In my work I decided to keep things simple and do not use too complex visualizations. In particular, I chose to use **geo map** to give the user a clear sense of where those cities are located in the US. At the same time I used **histograms** to clearly identify top values and generic trends for different measures. Data over different months were visualized using line graphs.

In one visualization I used a **stacked plot** to emphasize the amount of cancellations in different cities. During the feedback session, though, I figure out that the person who analyzed my report didn’t really understand this kind of plot, creating a wrong perception of results. To avoid this issue, I decided to replace the stacked plot with a **stacked histogram**. I was also suggested to look at percentages of cancellations instead of row numbers. To achieve this, I created a new calculated field and created a set with the top 10 cities with the highest percentage of cancelled flights to use later.

Regarding what I wanted to extract from this data, I focused mostly on cancellations and delays and several facts and figures around those.

Feedback

Visualizations are usually clear, hereby an analysis slide by slide:

Slide 2: graphs show the same measure about cancellations number, why not considering a percentage of cancellation depending on the number of total flights?

Slide 3: the first stacked plot is not very clear. At first sight it looks like Atlanta is the city with the highest cancellation in February, but actually it’s Chicago! Better to use a different kind of plot to give away a clear information. The second graph about day of the week is not very informative, can be left out.

Slide 4: Regarding the data, what does it mean NAS?

Slide 5: the fact that a filter it’s present for the second graph is not easy to notice because of its position.

The main takeaway from this visualization is cancellations and delays. The first happens a lot on big airports and it’s influenced by the season, the second is present a lot on different airport and carriers.

Resources

RITA dataset [http://stat-computing.org/dataexpo/2009/2008.csv.bz2]

[Airports information](https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat) [https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat]

Tableau website