

Final Data Story Project Report

By Soumia Zohra El Mestari

Data Story Link :

https://public.tableau.com/profile/soumia8748#!/vizhome/USFlights_15525685068340/Story1?publish=yes

Summary :

Flights is a dataset that contains informations about United State flight delays and performance comes from RITA From 2017 to 2018 . It contains 31023 rows and 21 variables . In this short Explanatory report we will investigate the measures of performance : Delays , Diversion and Cancellations in more depth .

Design :

After multiple Iterations and based on multiple feedbacks I decided to present my vision based on a top down approach :

- The story line : I started by demonstrating the place of the airlines industry in the US economy and the number of flights per states from 2017 to the end of 2018 , then I presented my measures of performance which are cancellations , delays and Flights diversions : I explored those measures in multiple ways using other variables and ended up showing my findings .
- Colors : After multiple feedbacks I decided to avoid using so much colors unless it was necessary so in my visualizations two colors in maximum and for the visualizations to be color blindness friendly I used mostly blue - Orange degrees and avoided to use any green , yellow colors .
- Visualizations types : I used bar graphs , map , line graph , packed bubbles and tried to show trends using shapes as they appear to be more intuitive and easy to be interpreted by any human regardless of his background in order to touch a large public . Also I tried to keep the visualizations as simple as I can by not overloading the chart and keeping a high data ink ratio.
- Variables used in this investigation are clarified in the [variables](#) section.
- Findings : I also added some findings as notes in each section of the story to help the audience stay in track with me along with some motivations in the captions as questions to keep their curiosity along the way , to summarise these findings :
 - California , Texas and Florida scored the highest flights rates among all the US states from 2017 to 2018.
 - The biggest player that influenced negatively this sector is the delays in the first place, flights cancellations in the second place while flights diversion didn't record a high rate.
 - No improvements were really recorded as a significant decrease in delays from 2017 to 2018 the trend overall tend to sway up and down.
 - Carrier Delays and Late aircraft delays are the most frequent furthermore the Carrier with the most delays is also the same: Southwest Airlines for California and Texas .

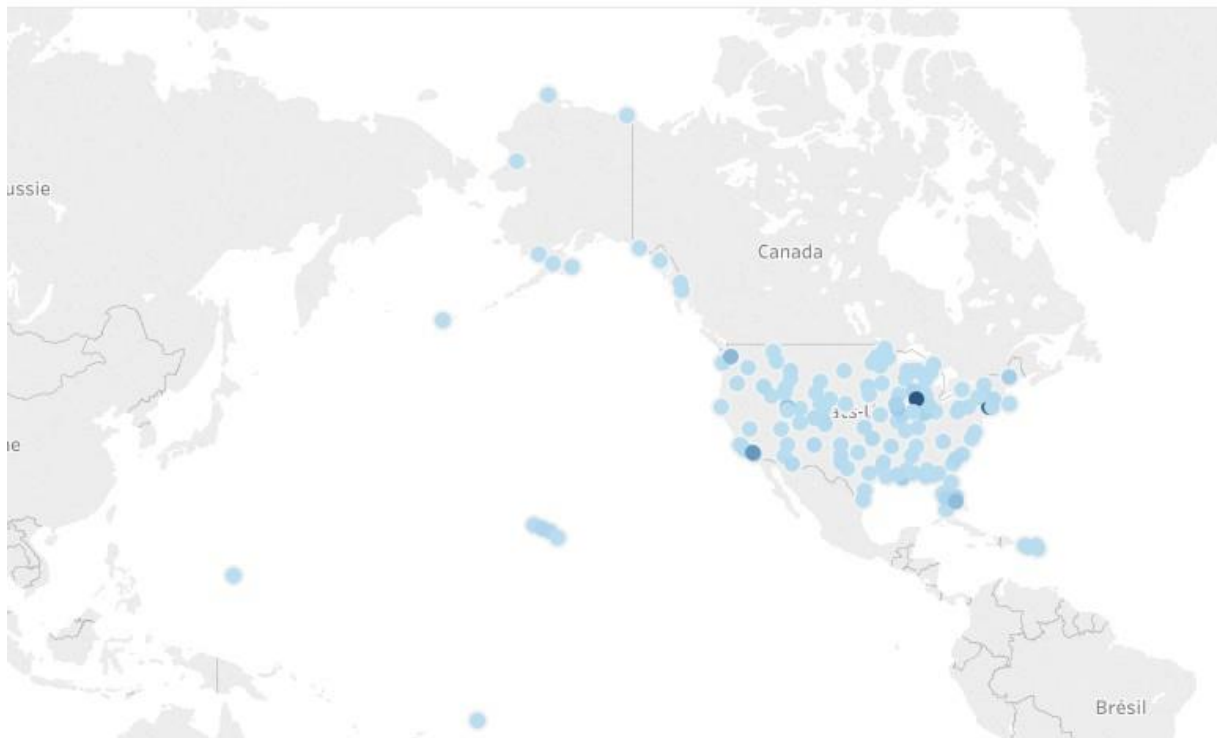
- Southwest Airlines , American Airlines and SkyWest Airlines scored the highest rates for cancellations and Diversion cases , those companies are scoring low service performance.
- Weather was a significant player in flights delays especially in Texas and California .
- Cancellations rates to be more frequent during Winter while Diversions in Summer .

Feedbacks :

In this section I tried to show the Data story to many persons : 15 years old teenagers , some of my data-analyst friends and other friends that have different backgrounds in order to collect diverse feedbacks and make my story touch the largest portion of audience regardless of their ages and sectors of activity .

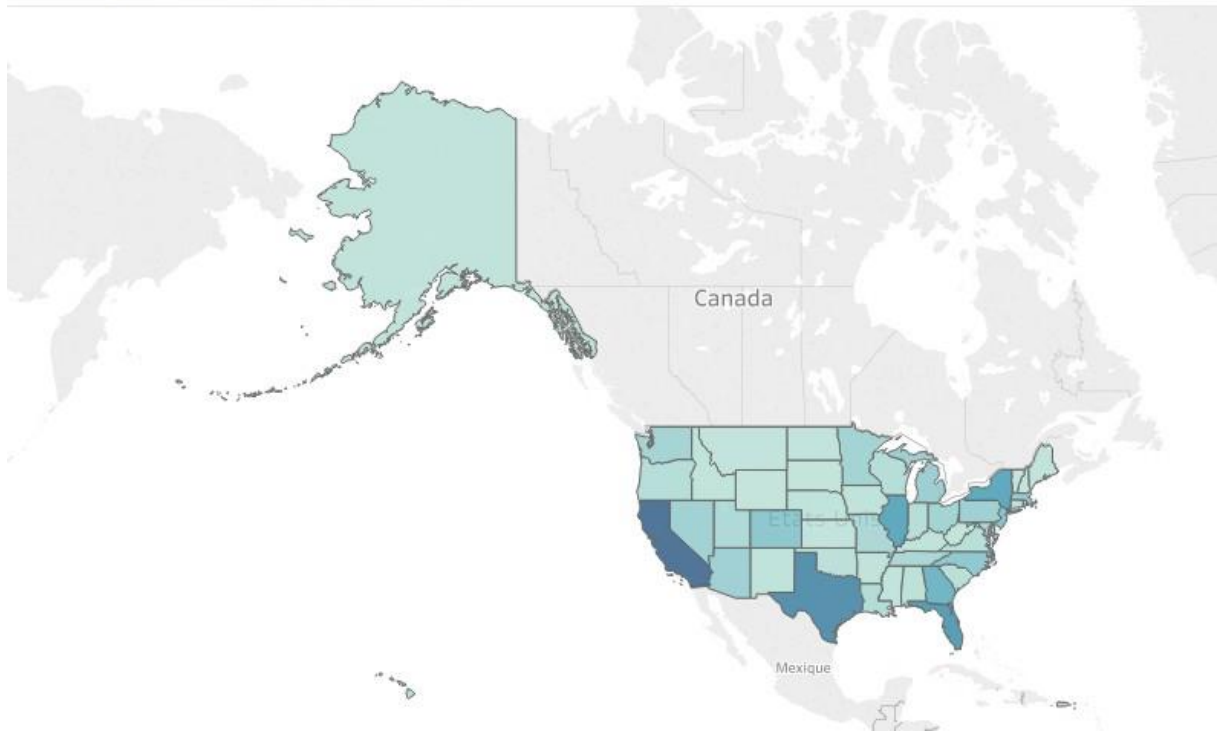
- First Feedback : The first version of all of my bar graphs contained a different color per bar , this led my audience think that some variables were more important than others => So I decided to keep just a blue color for all the bars in my charts.
- Second Feedback : Overplotting issue in the map The first map was filtered by cities in the US as shown below :

Delays per Location in United states

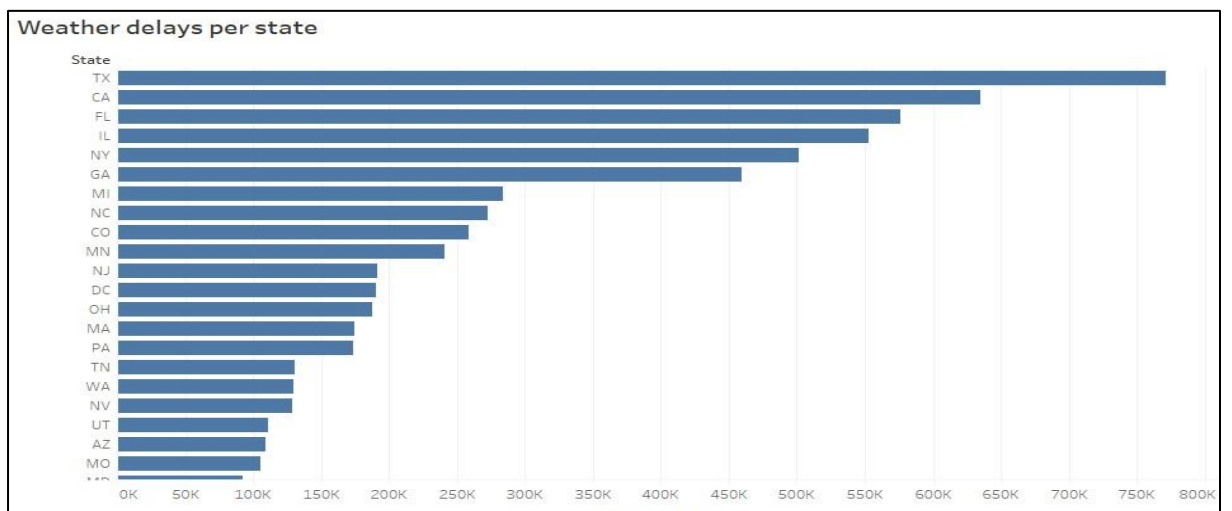


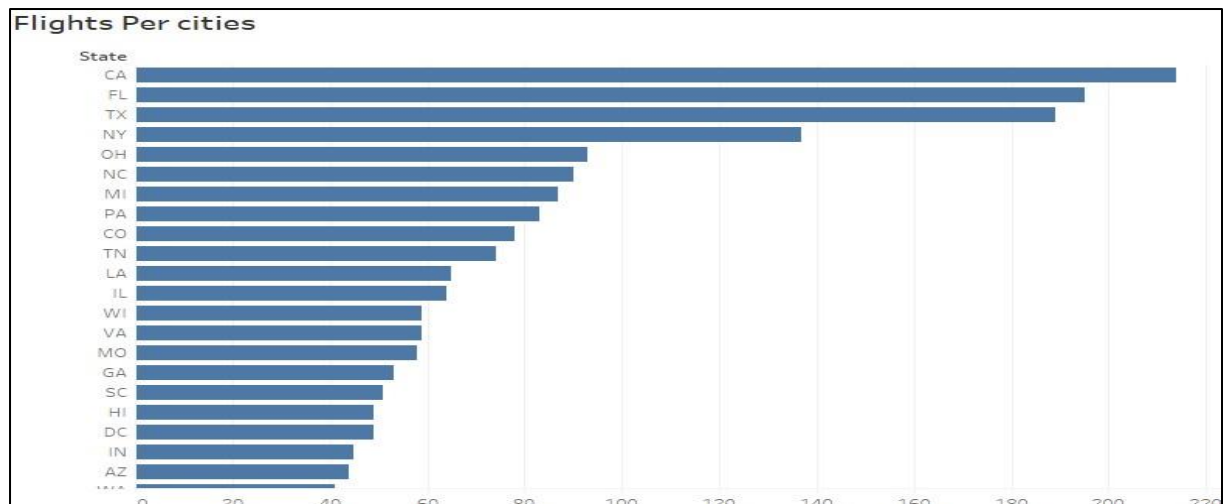
So this led to an overplotting in my visualization which was not clear at all , So instead I decided to aggregate by state to overcome this issue and the result was the one that you can find in the story , more clear and simple to understand. Here is the result :

Delays per Location in United states

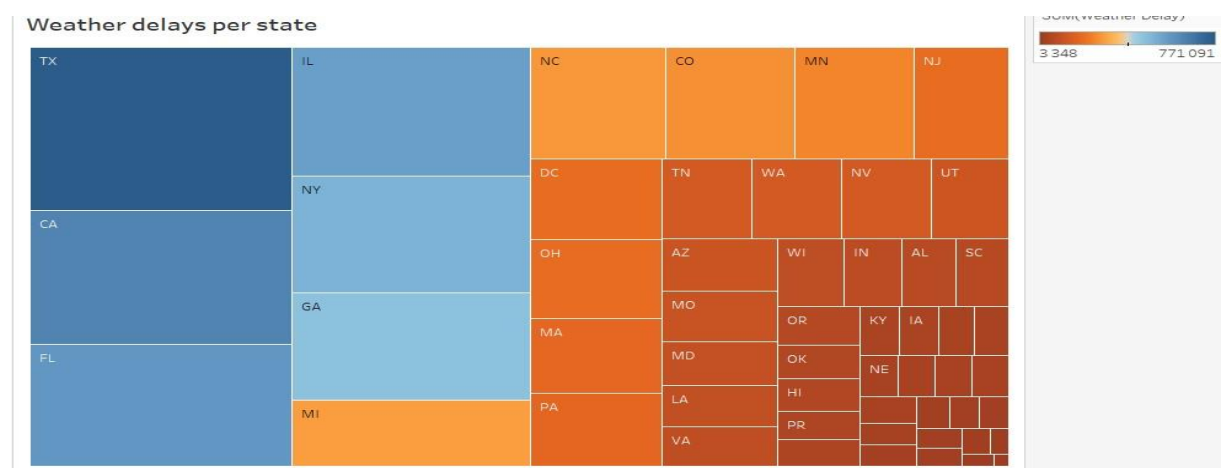


- Third feedback : weather delay per state as a bar graph was similar to a previous one(the flights per state) so this led my audience think that it was a duplicate :



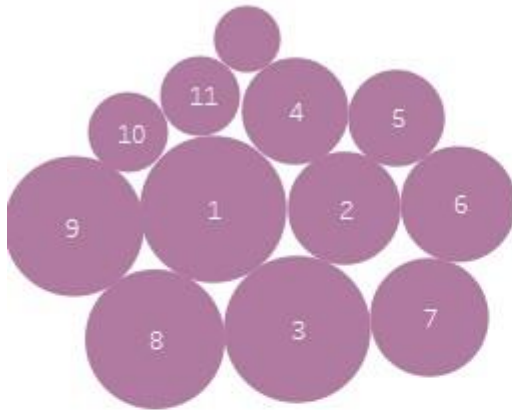


The two visualizations were similar in design so this led to a confusion with people thinking that it was a duplicate , to get rid of this problem I used a different type of visualizations : Treemaps in order to attach the reader attention and save him from thinking that it's the same plot . Here is the result :

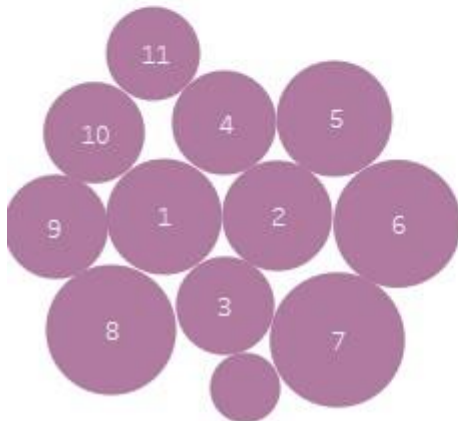


- Fourth Feedback : Diversions per month and cancellations per month being added to the same dashboard seemed to be hard to distinguish them :

Cancellations Per months

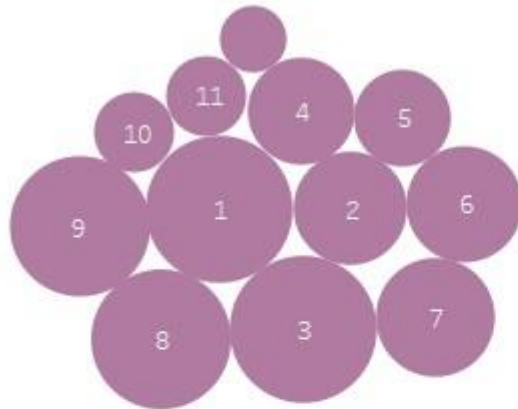


Flights diversions Per months

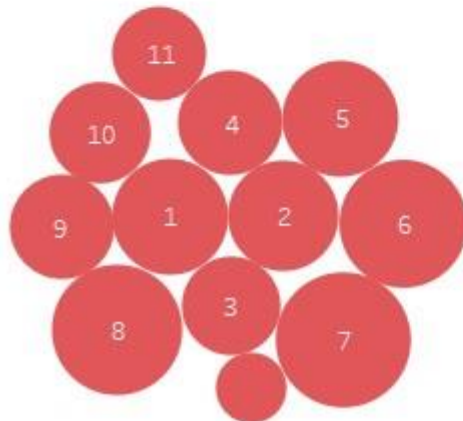


Two visualizations of the same type with the same colors in the same dashboards were not easy to separate So to overcome this I decided to play on colors and use different color for each one also to prevent the reader from thinking that one of them has more importance than the other I chosed colors that are close in degrees : purple and pink.

Cancellations Per months



Flights diversions Per months



Resources :

- to understand my variables I used the dataset source :
<https://www.bts.gov/topics/airlines-and-airports/understanding-reporting-causes-flight-delays-and-cancellations>
- To get help about using Tableau I used the Forum of Tableau
:<https://community.tableau.com/welcome>
- Used variables significations :
 - State : Abbreviations of the United States Of America names : further informations can be found here :
<https://abbreviations.yourdictionary.com/articles/state-abbrev.html>

- Categories of delays :
 - **Air Carrier:** The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
 - **Extreme Weather:** Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
 - **National Aviation System (NAS):** Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
 - **Late-arriving aircraft:** A previous flight with same aircraft arrived late, causing the present flight to depart late.
 - **Security:** Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.