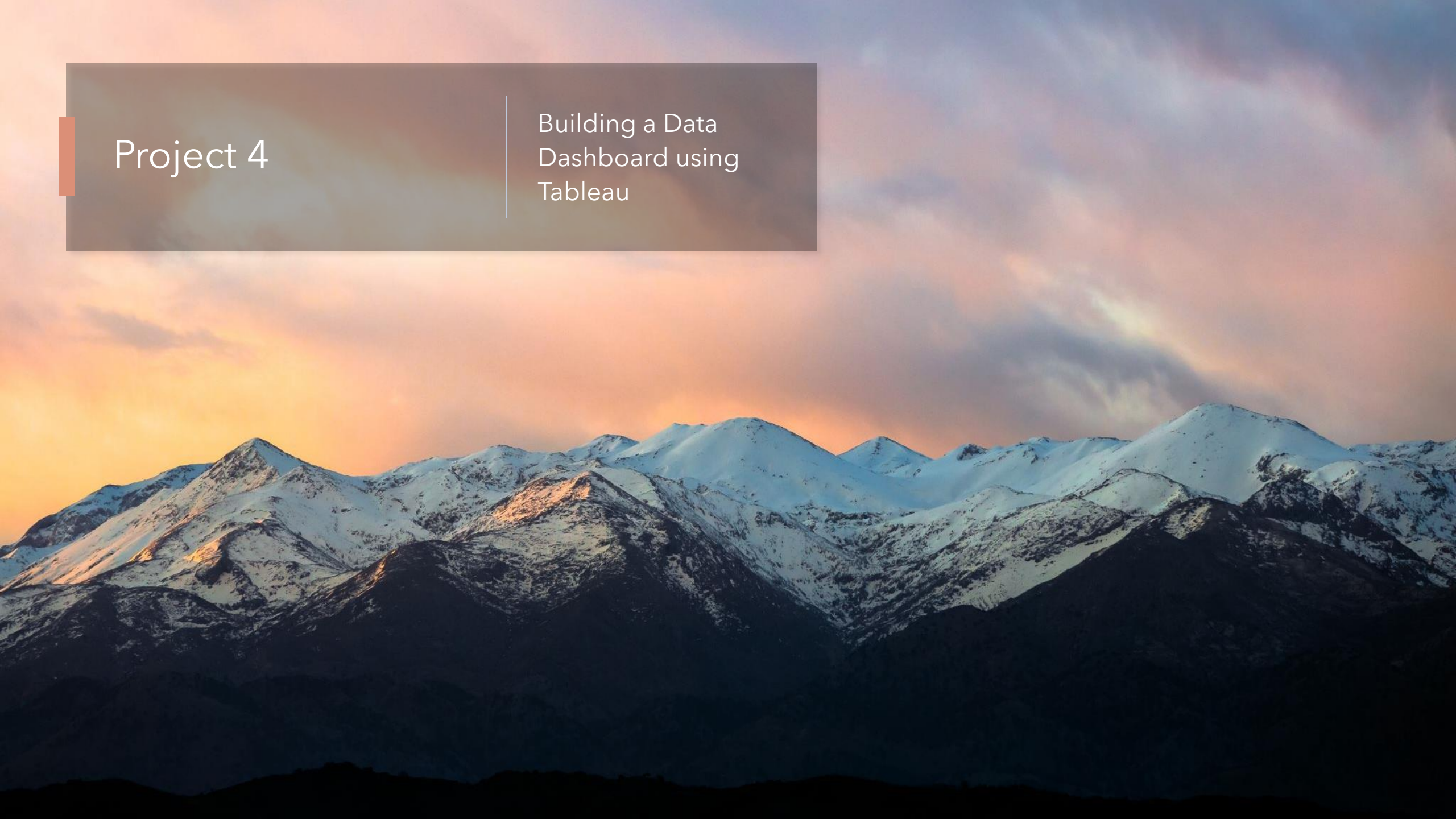
Project 4

Building a Data Dashboard using Tableau



Summary

- This visualization pertains to US domestic flights dataset, with data for the year 2015, displaying trends in Cancelation Ratios in different months, states, airports and airlines.

Notes

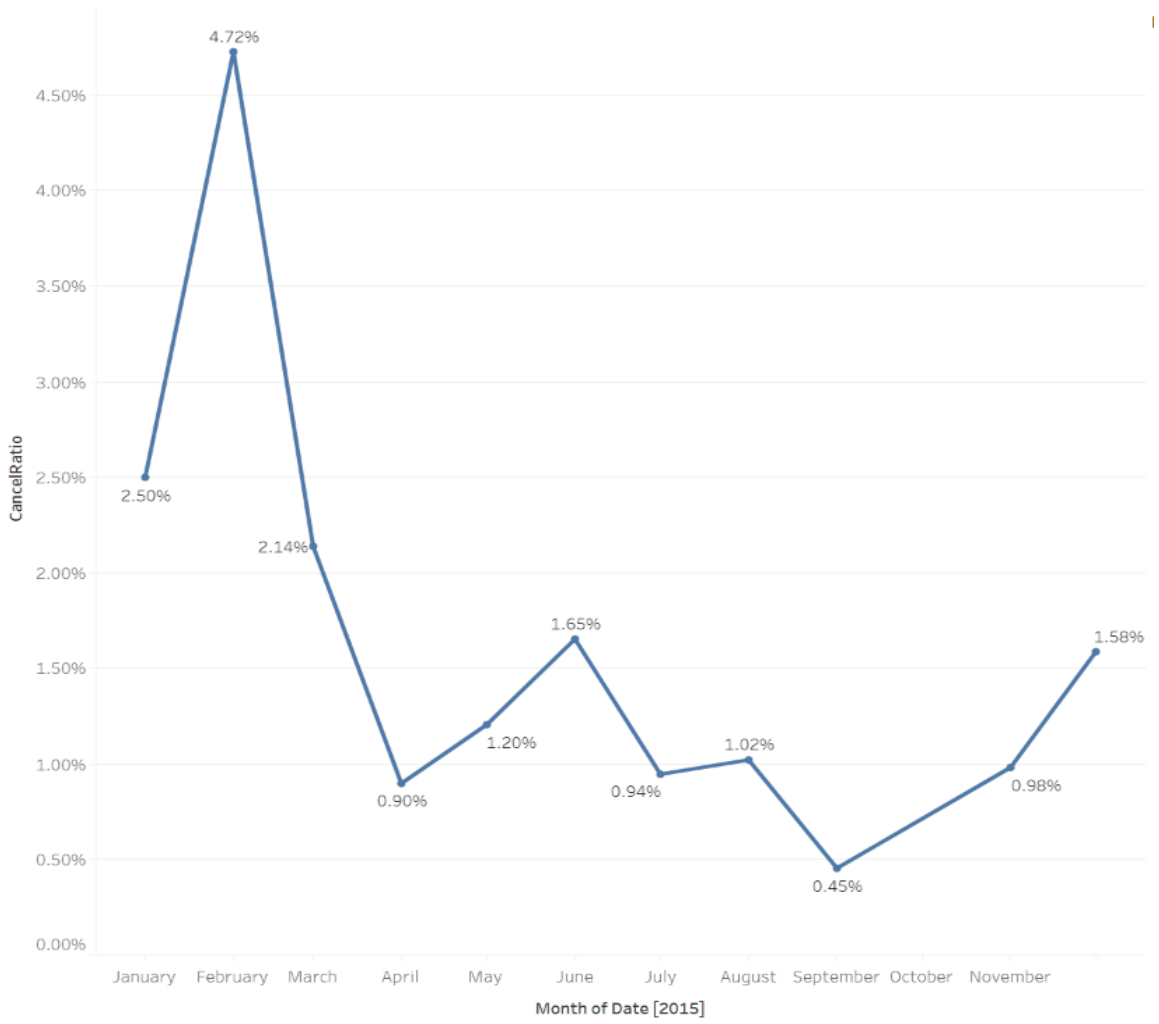
- Joined flights.csv to airports.csv, using Destination Airport as a foreign key and IATA CODE as a primary key, in order to include the airport name. This was achieved using na inner join, thus combining tables on values matching both tables.
- A map wasn’t used, as it was considered that the results could more easily be seen using a treemap.
- Negative values were included in Orange, while positive values are encoded as Blue, working for colorblindness.
- A story was created to better communicate insights and narrative.
- Cancel ratio is a calculated field: sum([Cancelled])/COUNT([Cancelled]).
- Date is a calculated field: DATE(STR([Month]) +'/'+ STR([DAY])+'/' + STR([Year])).
- A hierarchy was created with Airport and AIRLINE, making use of the join.

Research questions

- Are there specific months for the year of 2015 that stand out in terms of cancelations?
- Does geographic dispersion influence the cancelation ratio?
- Is there a correlation between Arrival Delays and cancelations?
- How do airports and Airlines contribute to cancelations?

Insights

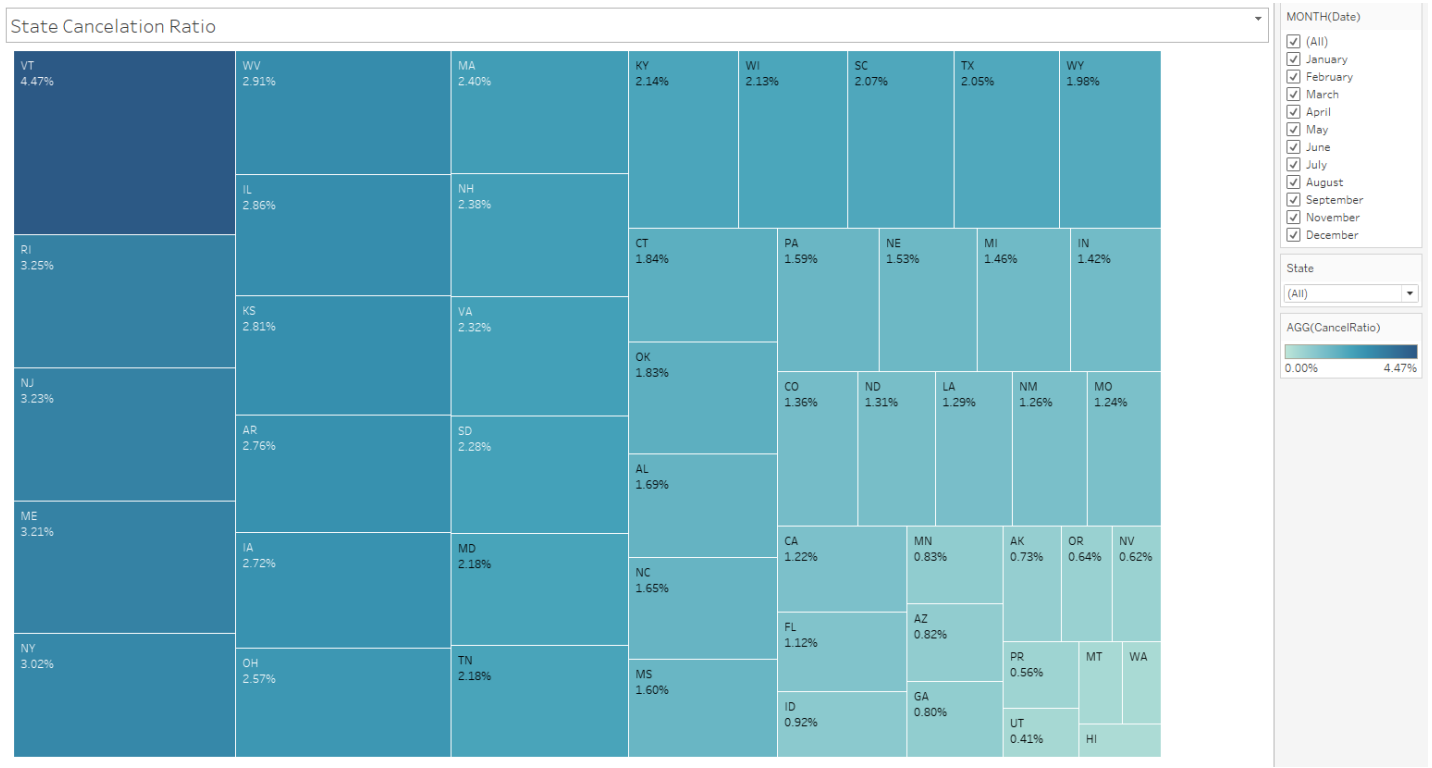
Cancelation ratios were significantly higher in February than the rest of the months.
It hits its lowest point on September.



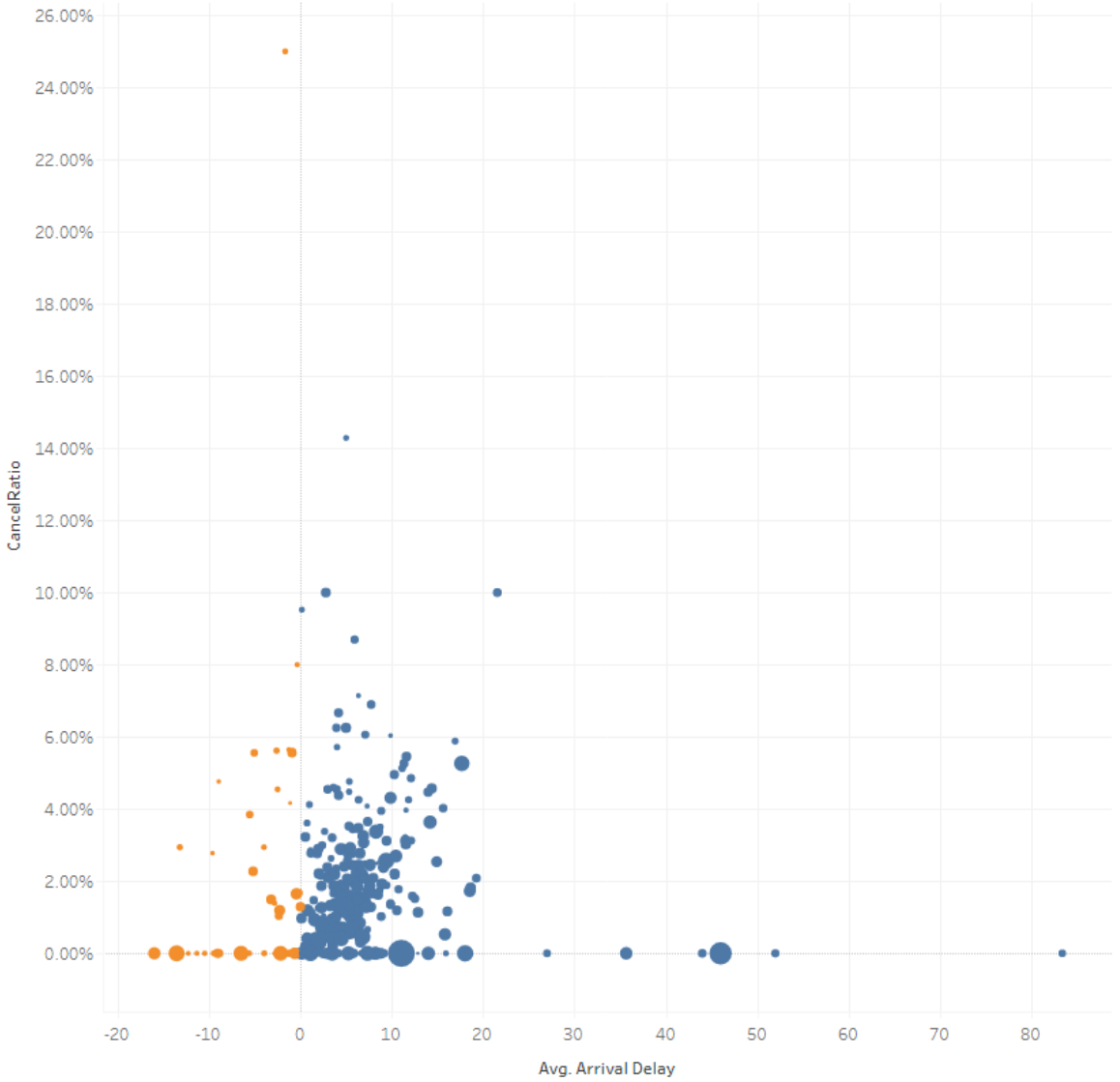
Vermont was the single most contributing state to cancelation ratio (4.47%).

In February (the highest point), New Hampshire was the biggest contributor (23.53%).

Over the year of 2015, Hawaii had the lowest ratio (0.30%).

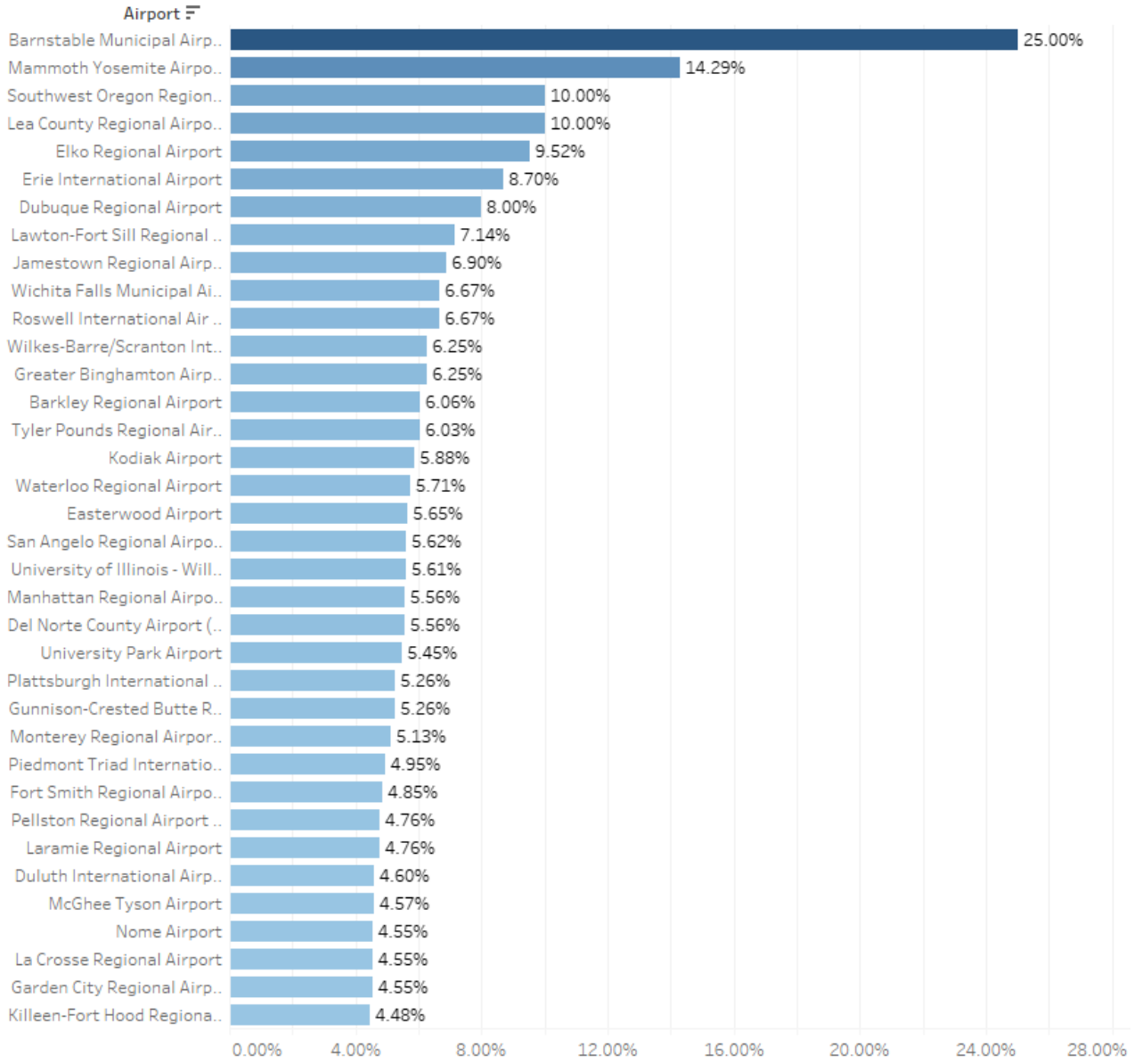


There can be identified a positive correlation between Cancelation Ratio and Average Arrival Detail, comparing airlines.



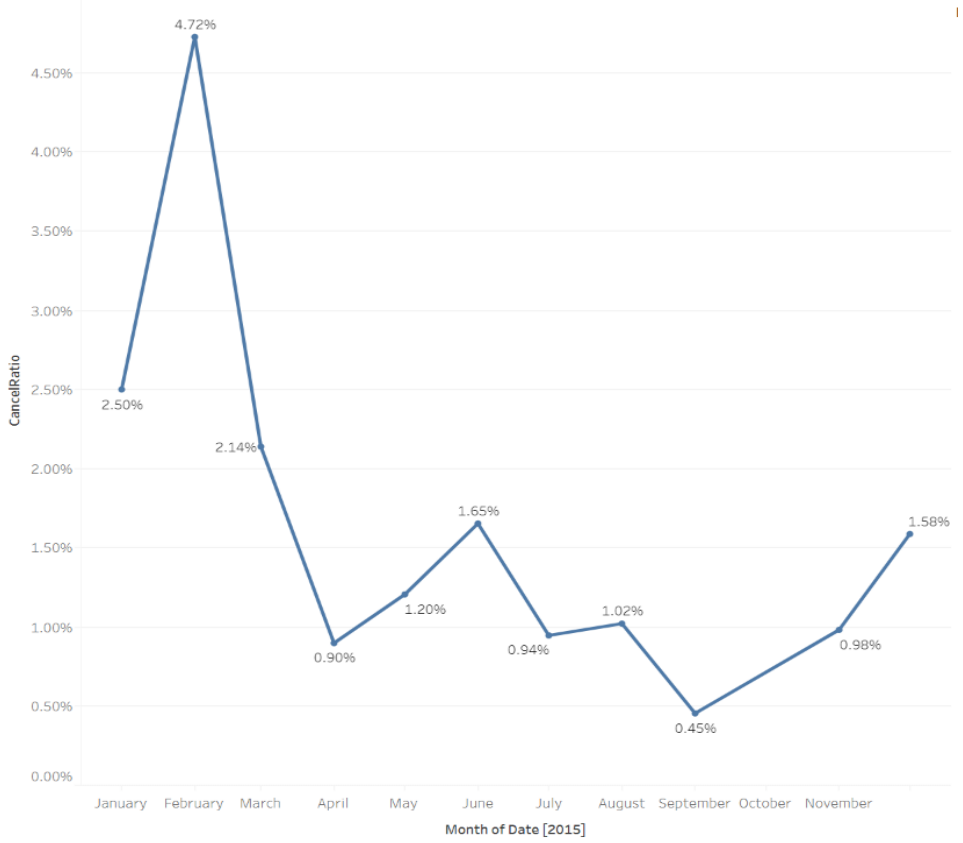
We may see that Barnstable Municipal is the biggest contributor.

We can then filter by a count of canceled flights (to exclude flights airports with low counts), where we can find different airports that are more significant.



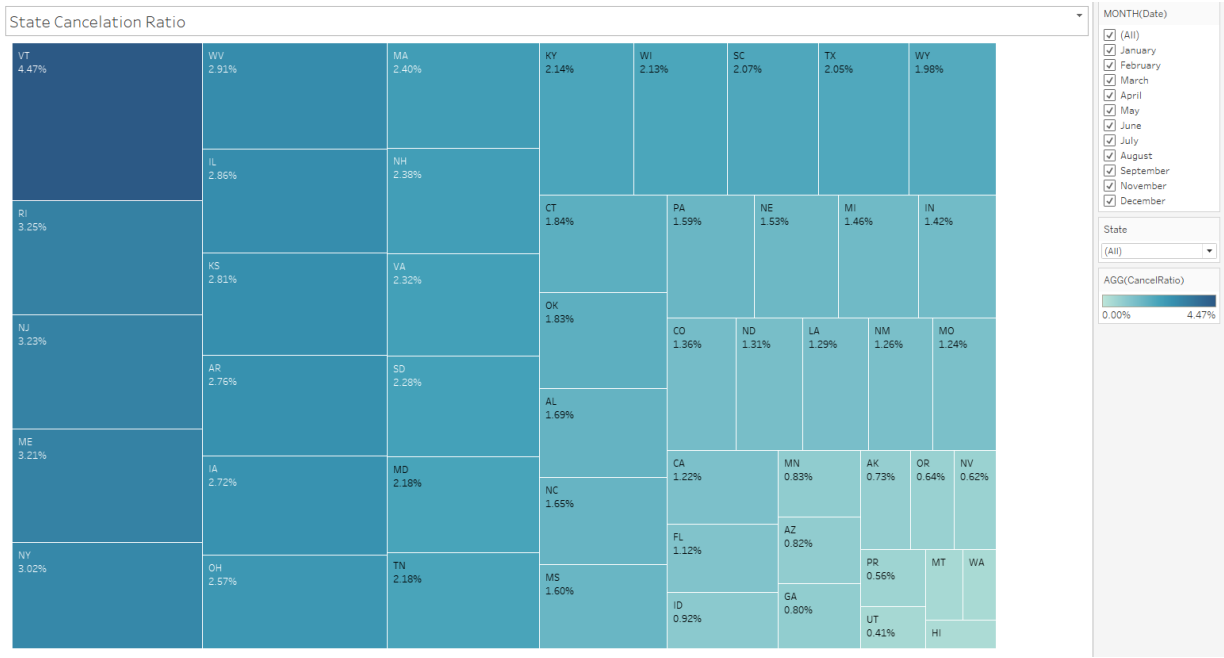
Design Choices

- Built a line chart in blue color, which shows trends over the months, while adding data labels to facilitate reading.



Design Choices

Opted for a treemap instead of a bar chart or a map to display this information, as we can very easily and quickly identify the most or less contributing states at a glance, with no chart junk.



Design Choices

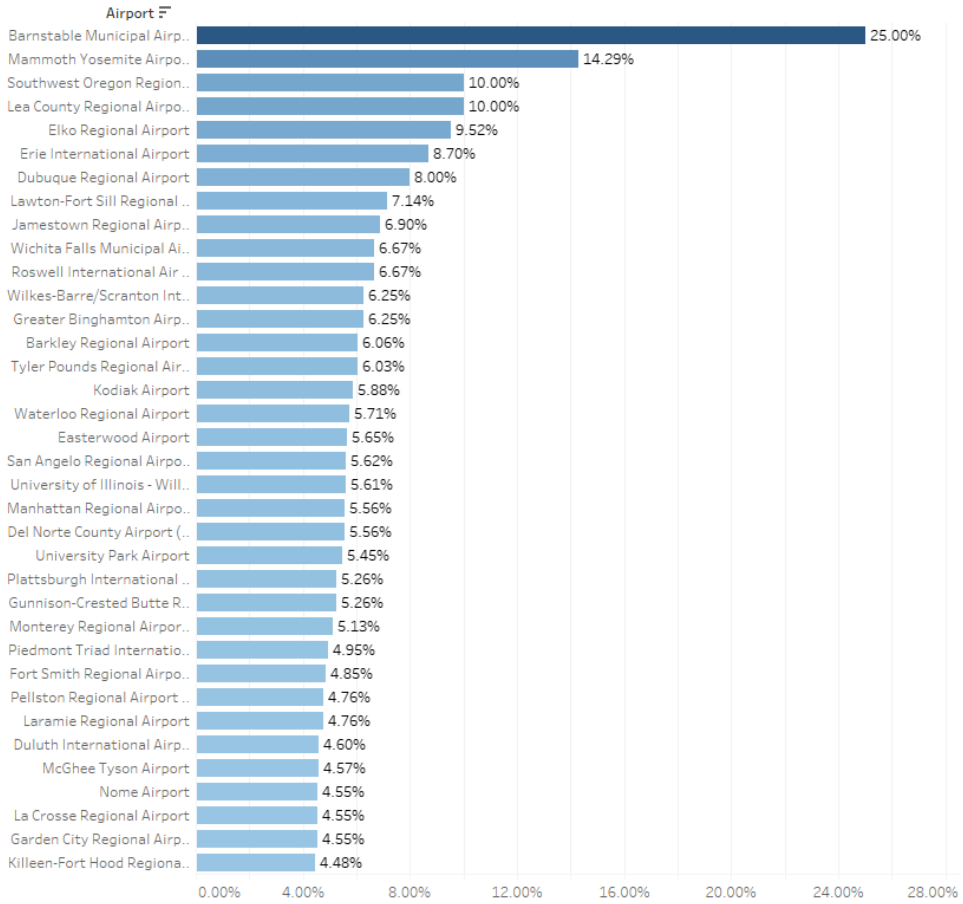
Used a scatter plot to plot Average Arrival Delay and Cancel Ratio based on Airports.

Used a Blue color for positive values and Orange for negative values.

Design Choices

The bar chart can display the information in a very visual way that is helpful for making comparisons. Data labels were added to the visualization to more easily compare.

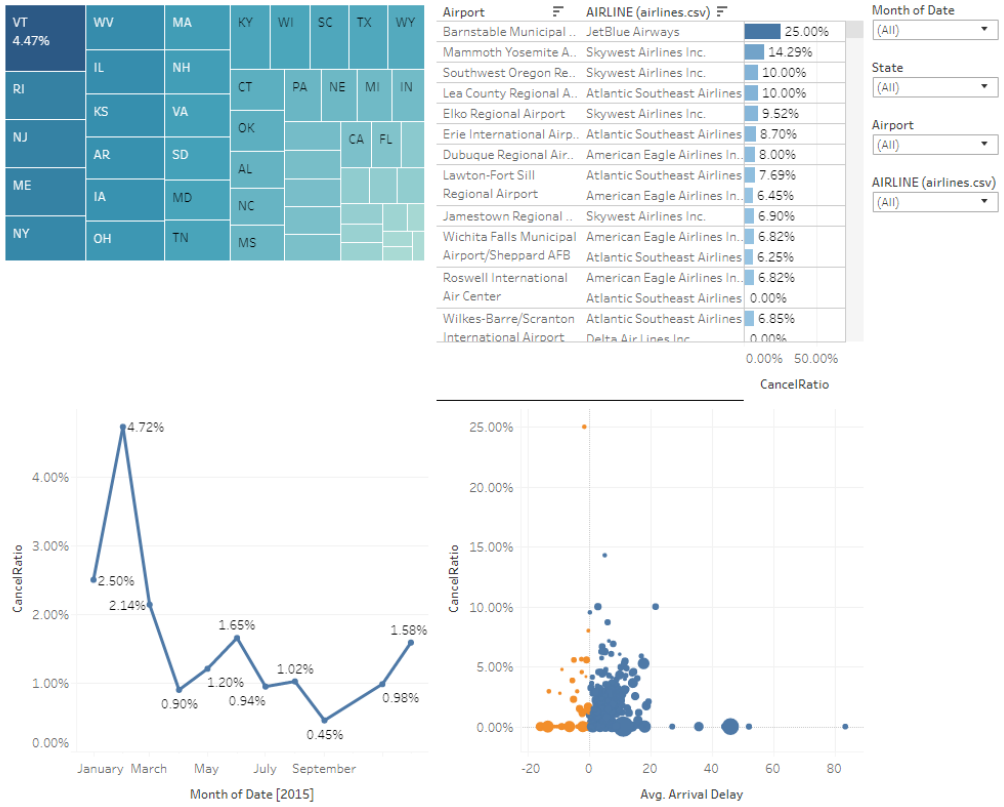
By using a hierarchy, the user is able to toggle to airline to give more information for a particular airport.



Design Choices

On the dashboard, by clicking any of the charts, it interacts with others.

Despite that, I decided to include filter as they might be more intuitive for some of the users who are not used to it.



Sources

- https://community.tableau.com/thread/204537 - Create date field column
- https://community.tableau.com/thread/147188 - Change measures number formats
- https://help.tableau.com/current/pro/desktop/en-us/joining_tables.htm - Joins

Link to Story

- https://public.tableau.com/profile/gustavo.pacheco#!/vizhome/DataVizProjectDAFFinal/MyStory?publish=yes