Data visualization dashboard (Kaggle flights dataset)

Insight 1

Research question

What types of delays are experienced throughout the year and which month is the most affected by delays?

Link

[Data visualization dashboard (Kaggle flights dataset) | Tableau Public](#)

Summary

Before adding the month division, I first determined that the annual predominant type of delay is departure. The first question I aimed to answer with this viz is how delays change throughout the year. Delays are the most abundant in June, and seem to coincide with holidays. By selecting different states correlations can be found between climate and population with delays. All states follow a similar pattern in terms of severity of the delays, but not every state experiences delays year-round; whereas populous states like California and New York experience delays every month, whereas Alaska and Washington experience negative delays during one or several months per year. These negative delays are exclusively arrival delays.

Design

1. I selected all measures that included the word "*delay*" into a histogram; there are 7 variables that match this requirement: weather, air system, airline, late aircraft, arrival and departure.
2. I segregated the cumulative delays by month and included a legend to explain the different colors. I decided to keep the multi-colored design, because different shades of the same color suggested a hierarchy.
3. I added a filter for state.

Resources

Masterschool Live Session - Project Walkthrough - Build your Data Visualization (Emiliano Argnani)
[Measure Values and Measure Names - Tableau](#)

[List of U.S. states by population density - Simple English Wikipedia, the free encyclopedia](#)

Insight 2

Research question

Which state has the most cancellations per year/month?

Link

[Data visualization dashboard (Kaggle flights dataset) | Tableau Public](#)

Summary

In this visualization I used a color filter on a map of the United State to indicate how many happen per month in a specific state. Over the whole year, Texas has the most canceled flights. To investigate whether this effect is consistent throughout the year, I added a filter on month. Texas os predominantly - but not exclusively - the worst performing state, although closely followed by California, Illinois, and New York.

Design

1. Created a map using the Longitude and Latitude measures
2. Used the SUM of cancellations to create a color-coded map based on the amount of cancellations
3. Added a filter based on month

Resources

[Build a Simple Map - Tableau](#)

Insight 3

Research question

To what extent does cancellations per airport correlate with cancellations per state and is this effect consistent between airlines?

Link

[Data visualization dashboard (Kaggle flights dataset) | Tableau Public](#)

Summary

After learning from insight one that most cancellations occurred in Texas, I thought it might be interesting to know which airports had the most cancellations and to what extent this correlates with the total amount of cancellations per state. Interestingly, most cancellations occurred at O'Hare International Airport in Chicago, Illinois. Although Illinois had a significant amount of cancellations throughout the year, it was never the worst state (which was almost unanimously Texas). To determine which airline had the cancellations from which airport, I added a filter. The origin airport was surprisingly diverse, with ORD being with less than half the airlines being the most canceled airport of origin. The first explanation that comes to mind is that the harsh weather conditions (especially in winter) would be to blame; however, this does not explain why Texas has the most cancellations. When investigating the amount of delays based on the airline it becomes clear that the amount of cancellations is the result of unequal distribution. First of all, the size of the airline: Alaska Airlines' highest number of cancellations is 5, whereas American Eagle's Airlines' is 129. Airlines have a limited number of airplanes which are stored in a *'hub'* at an airport of their choosing, and most of their departures will occur from there. Secondly, each airport naturally develops a "specialization", with some airports being a transfer airport (e.g., Frankfurt in Germany), which will clearly receive more traffic than a destination airport (e.g., Alaska).

To further investigate this phenomenon (future directions) would require data on the airlines themselves: what is their main airport, how many airplanes do they have, how many employees, who is their target audience in terms of budget, etc.

Design

1. Added "*Origin Airport*" to Columns and SUM(Canceled) to Rows
2. Changed the design from Table to a vertical bar chart
3. Colored the bar based on the sum of the cancellations per specific airport
4. Added a Single Value (List) filter based on Airline

Resources

Masterschool Live Session - Project Walkthrough - Build your Data Visualization (Emiliano Argnani)
Where are all the unused planes right now? - BBC Worklife