Project Summary

Dashboard Links:

- **Dashboard 1:** https://public.tableau.com/profile/sai.saranya.gurram#!/vizhome/Flight_Delay_Analysis_1/Delays
- **Dashboard 2:** https://public.tableau.com/profile/sai.saranya.gurram#!/vizhome/Flight_Delay_Analysis_1/AverageDelaysByAirport
- **Dashboard 3:** https://public.tableau.com/profile/sai.saranya.gurram#!/vizhome/Flight_Delay_Analysis_1/DelaySpecifics
- **Months with Trend line:** https://public.tableau.com/profile/sai.saranya.gurram#!/vizhome/Flight_Delay_Analysis_1/AirlineDelayAvg_by_Month
- **Airline Delay Map:** https://public.tableau.com/profile/sai.saranya.gurram#!/vizhome/Flight_Delay_Analysis_1/Delay_Map
- **Airline Delay Count with Map:** https://public.tableau.com/profile/sai.saranya.gurram#!/vizhome/Flight_Delay_Analysis_1/AirlineDelay_Count_StateMap
- **Weather Delay by Month:** https://public.tableau.com/profile/sai.saranya.gurram#!/vizhome/Flight_Delay_Analysis_1/WeatherDelaybyMonth

Design Choices:

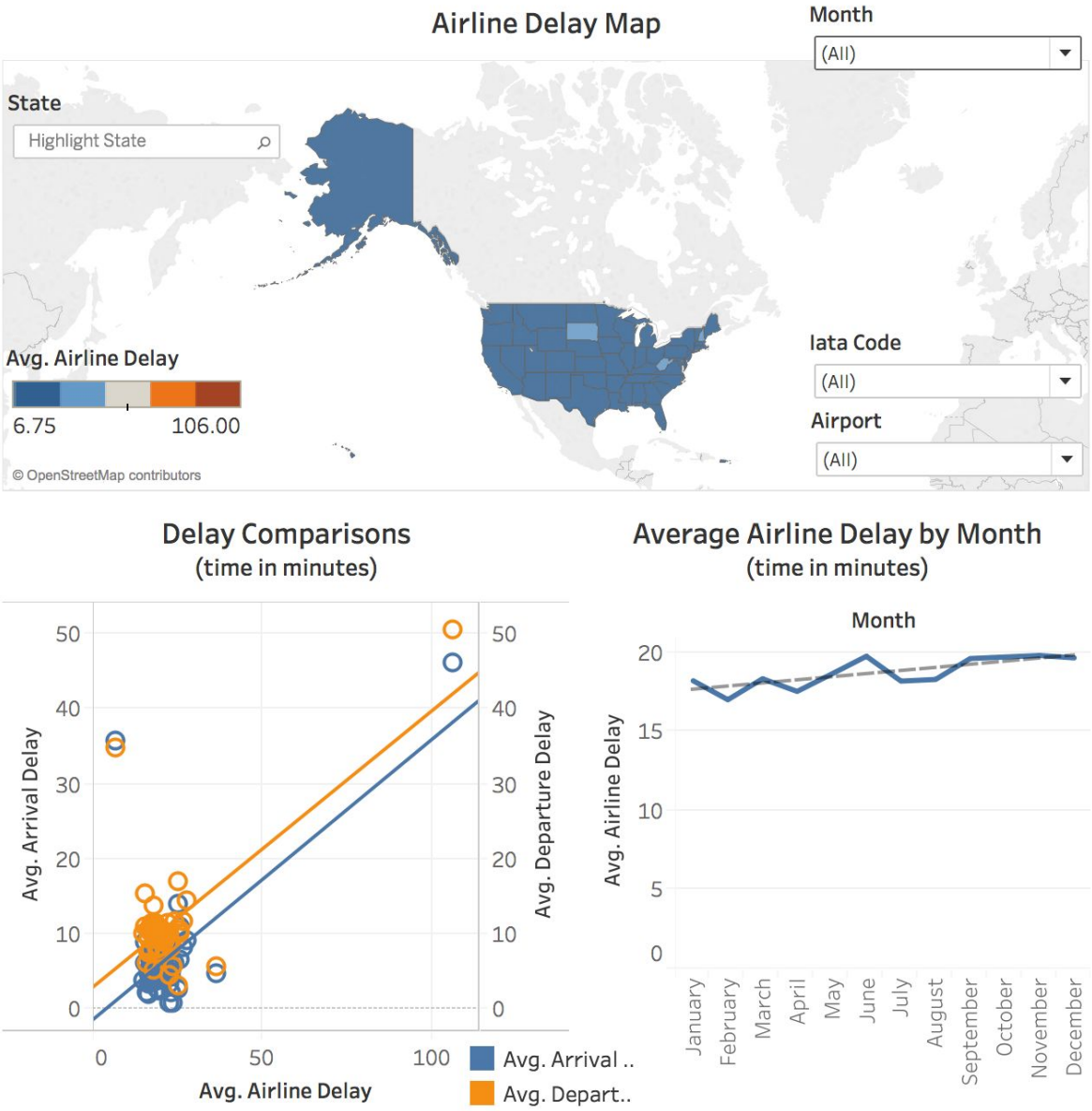
The choices for this design were primarily made to keep things simple and easy to use. When available, the Tableau palette specifically designed to be colorblind friendly has been utilized. When that palette was not an available choice, the colors chosen are oranges and blues, which are colorblind-friendly choices. Values that span a range of numbers are encoded in a range of colors, specifically orange:blue diverging. In order to make numerical differences more obvious in state maps, the colors were stepped in five steps, with the colors reversed so that higher numbers (in terms of delay minutes) were orange, while low delay numbers were blue. This is because people are generally familiar with colors closer to red being the less-desirable figures. Axis titles were made larger and rendered in bold for better legibility, while chart titles were removed from dashboards to reduce clutter. The most user-friendly design for month and airport

filters seemed to be single-value dropdown menus, while utilizing a highlighter for states wound up being the most user-friendly tool.

Summary:

This project looks at a variety of data in order to identify the data that may be correlated to overall airline delays. By comparing airline delays to arrival delays, departure delays, specific states, airports, and months in the year, as well as air time and travel distance, the project works to identify any characteristics that appear to have a strong relationship with airline delay within this data set. After analyzing the data, it becomes clear that a variety of factors appear to be correlated with airline delay, including arrival and departure delays, longer distances, and greater airtime. However, the factors with the largest influence on higher than average airline delays seem to be the month of the year and the location of flights. Although this insight cannot be explicitly drawn from the available data, it appears that the human factor (perhaps surrounding holidays and travel as well as location) may be the greatest influence on larger than average airline delays.

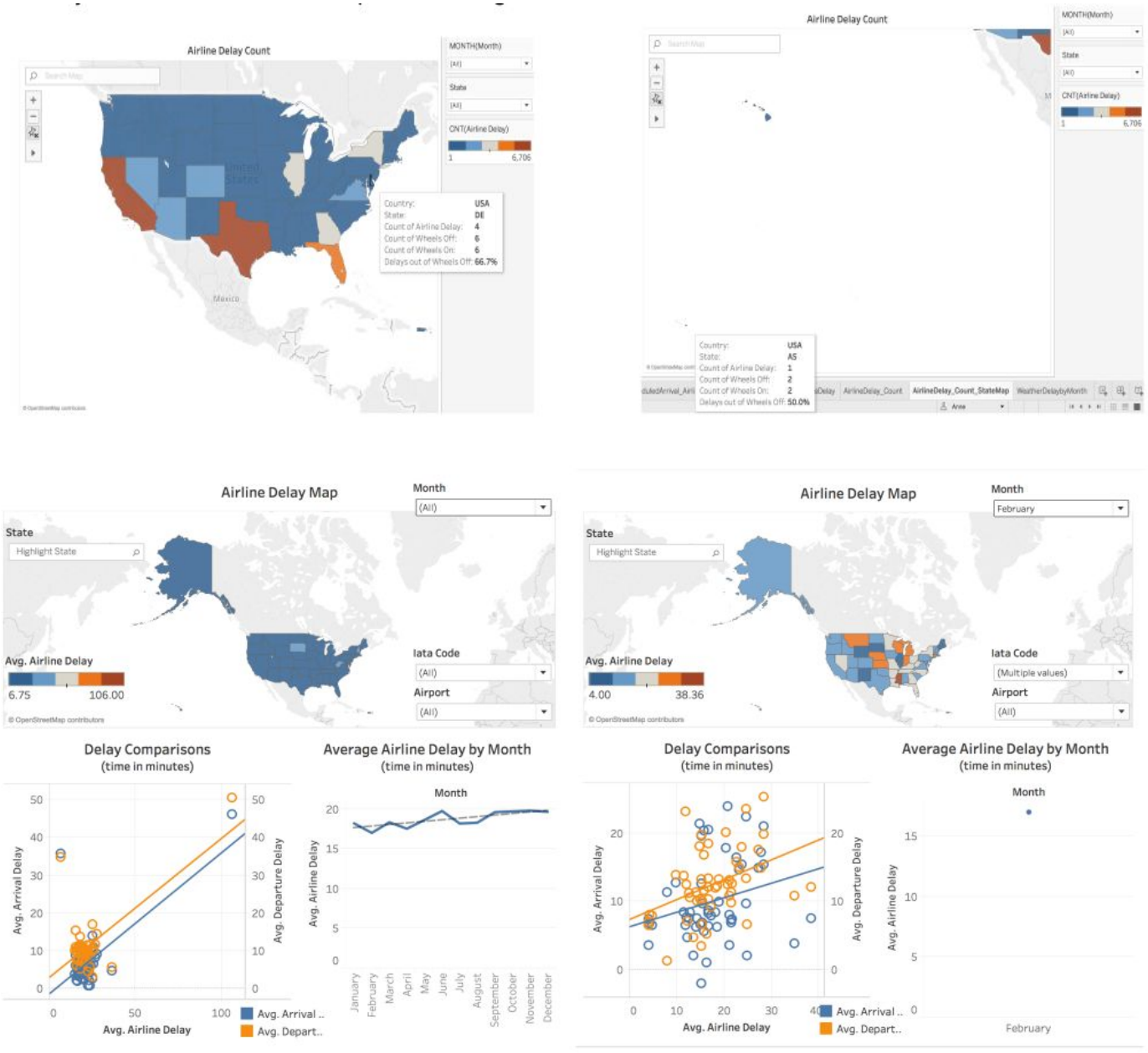
Description of Visualizations:



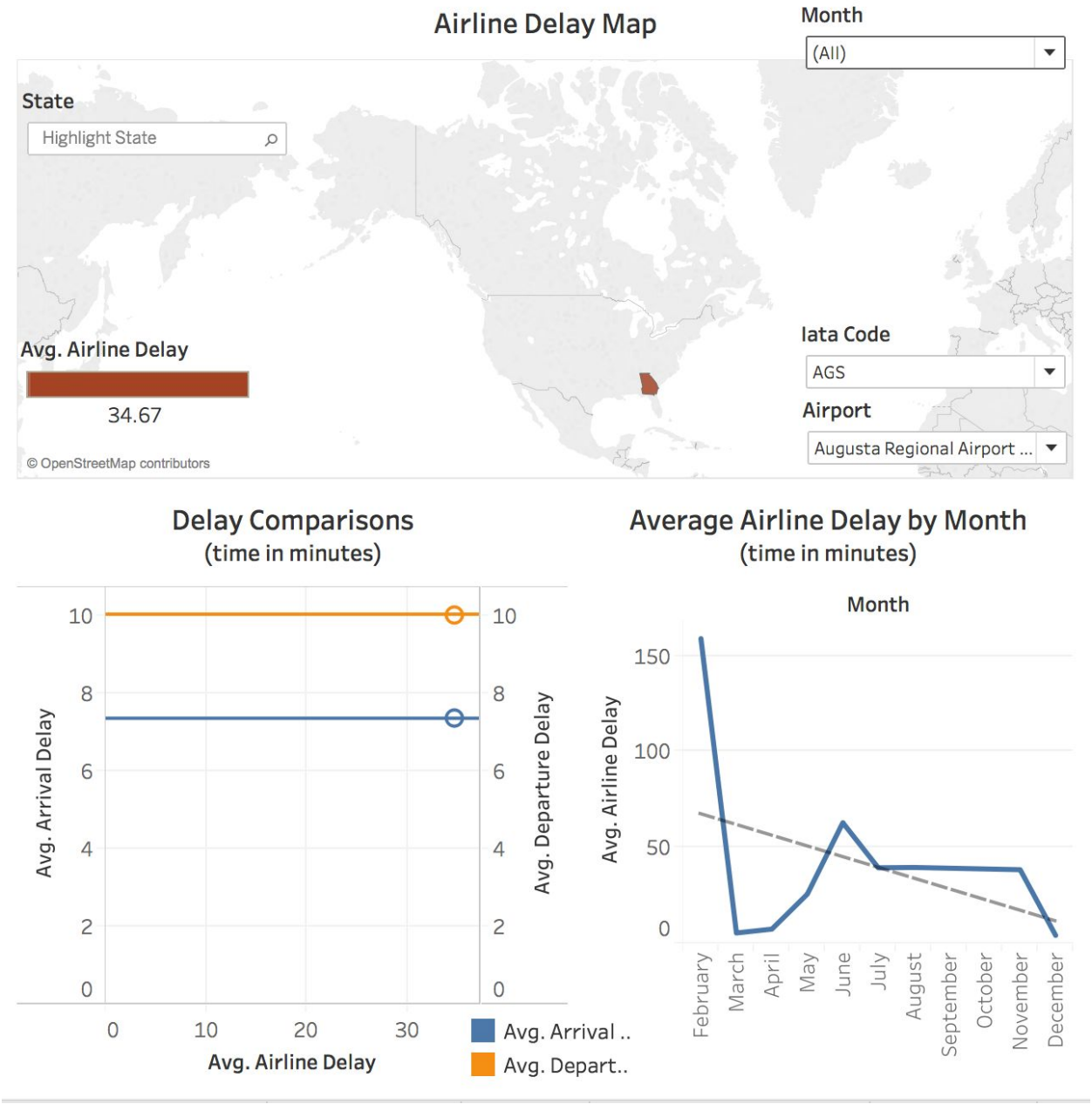
Average airline delay across all airports. (Time shown in minutes)(Dashboard 1)

The dashboard on the left examines the relationship between airline delay and arrival delay as well as airline delay and departure delay across all airports. There appears to be a strong relationship between low arrival delays, departure delays, and airline delays, with a few outliers, specifically Delaware and Pago Pago. (However, it should be noted that there is only a count of four airline delays out of a count of six wheels-off (66.7%) reported in Delaware, and only one

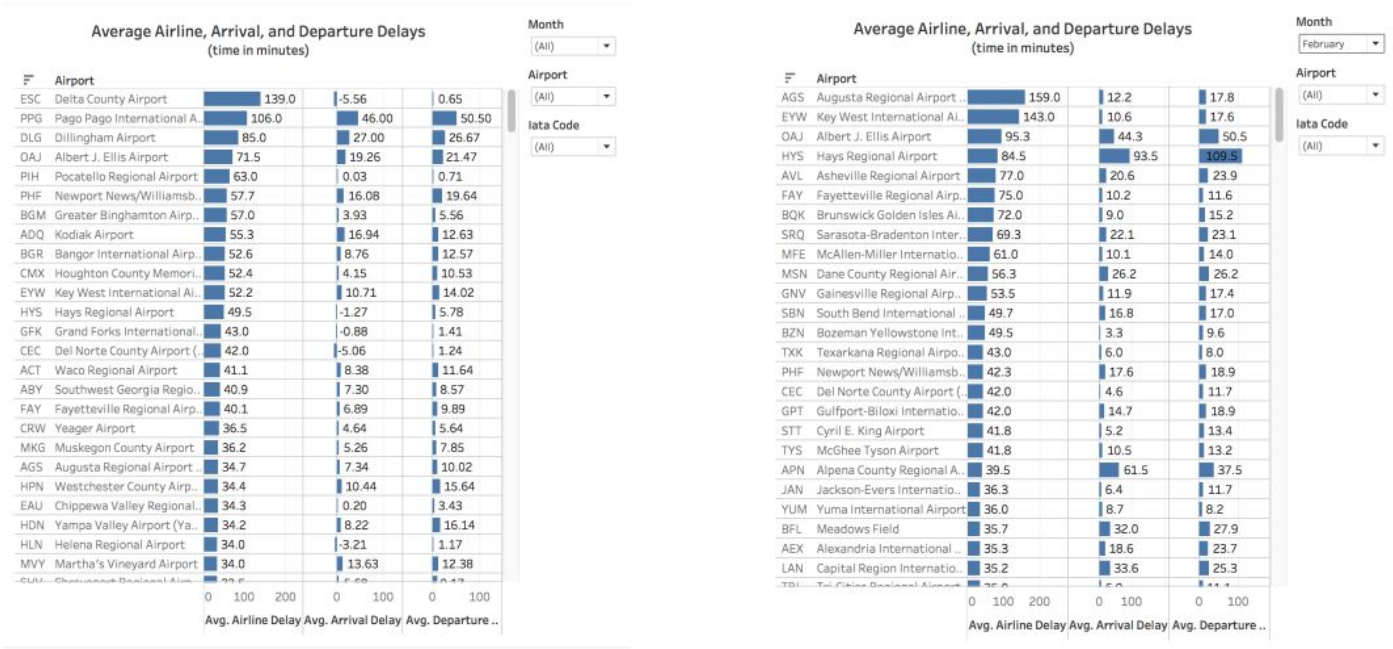
delay of the two wheels-off reported in Pago Pago (50%): see maps below.) Looking at the months in this year, while there is not a huge range of delays throughout the year, there are certain months that show higher and lower airline delays throughout the year, with peaks in June, November, and December. There is a clearly increasing trend of average airline delays as the year progresses, although the range of average delays is relatively small.



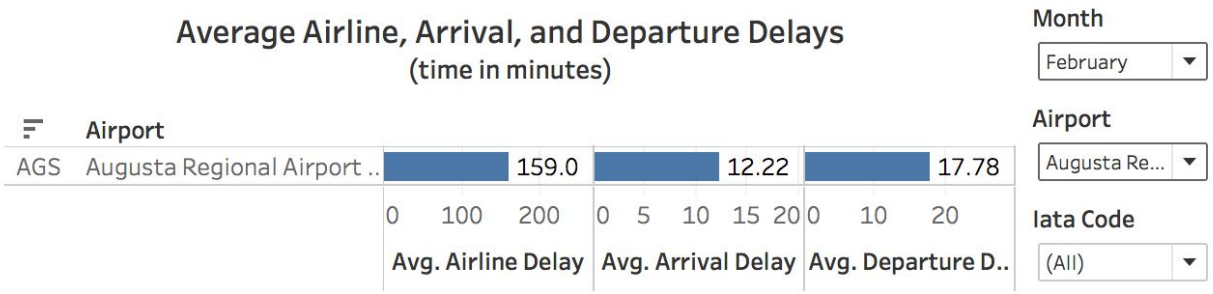
*Average airline delay plotted against average arrival delay and average departure delay. (Time shown in minutes.)(**Dashboard 1**)*



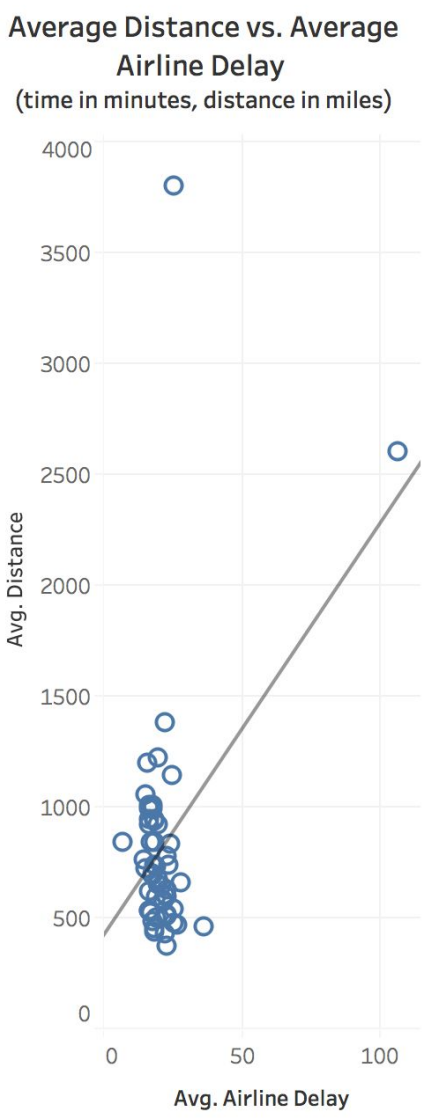
Looking more closely at the delays by month, it becomes clear that, although the appearance of a strong correlation between arrival, departure, and airline delays remains, the points become more spread out as the number of delays increases. The chart on the top right looks specifically at delays in the month of February, and it becomes clear that there are a number of states in this month with much higher than average delays as compared to the delays by state for the year as a whole. Looking specifically at Augusta Regional Airport (bottom left), February appears to be the month in which the number of airline, arrival, and departure delays are at an all-time high, followed by a drastic reduction in delays in March, with another increase in June, after which delays at Augusta decline.

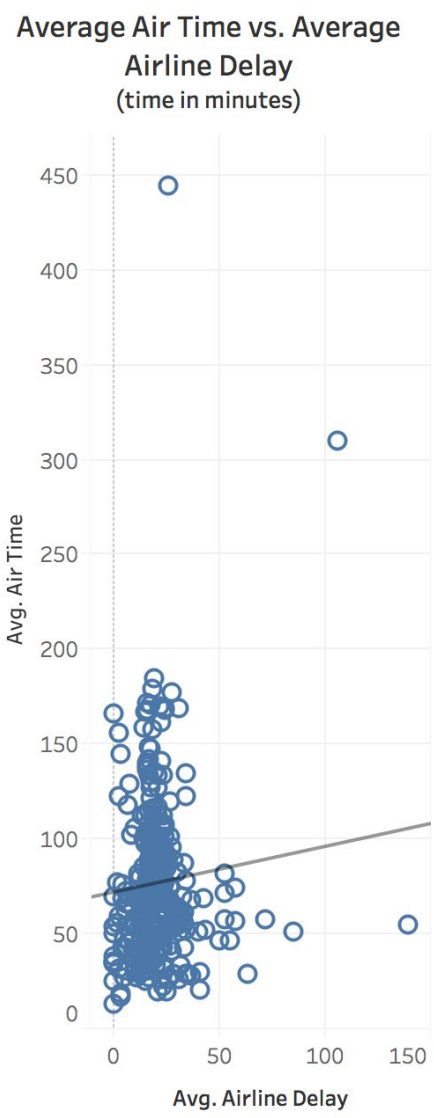


*Average airline delay compared to average arrival and departure delays in specific airports. (Time shown in minutes.)(**Dashboard 2**)*

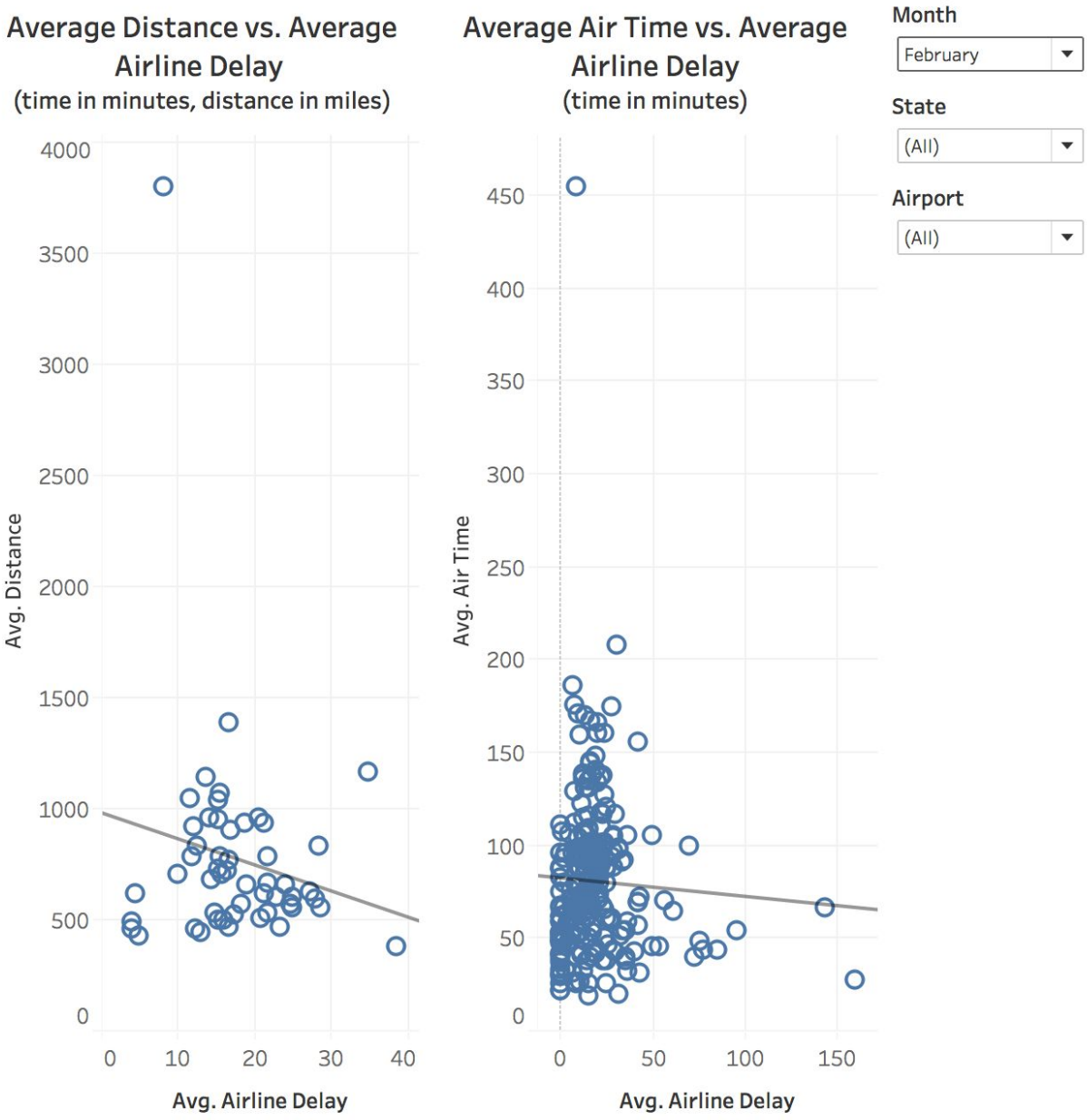


This dashboard looks at the average airline, arrival, and departure delays in 2015 by airport. Across the year, Delta County Airport and Pago Pago International airport have the highest average airline delays, but it appears that while Pago Pago does have both higher than average arrival and departure delays, Delta County Airport has the highest airline delays but actually has low departure delays and negative arrival delays (which means that they arrived, as an average, earlier than expected). In the month of February (top right), Augusta Regional Airport has the highest average airline delays, followed by Key West International Airport. The chart in the bottom corner looks only at Augusta Regional Airport, which has the highest average airline delays in the month of February.


Average Distance vs. Average Airline Delay
(time in minutes, distance in miles)
Avg. Distance
4000
3500
3000
2500
2000
1500
1000
500
0
0
50
100
Avg. Airline Delay


Average Air Time vs. Average Airline Delay
(time in minutes)
Avg. Air Time
450
400
350
300
250
200
150
100
50
0
0
50
100
150
Avg. Airline Delay

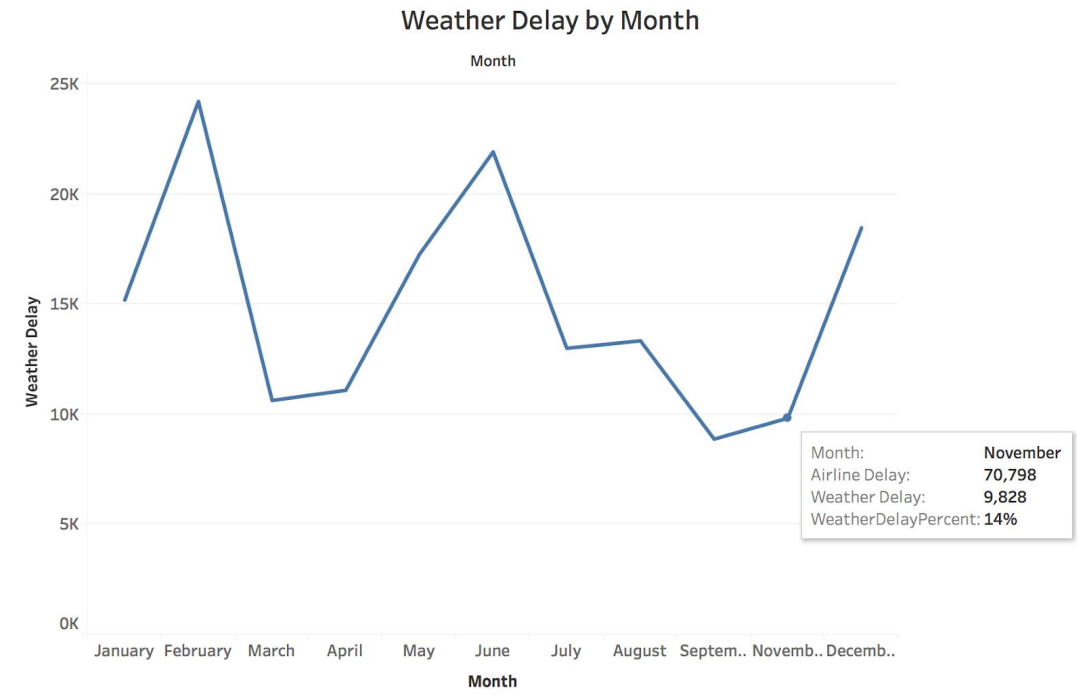

Month
(All)
State
(All)
Airport
(All)

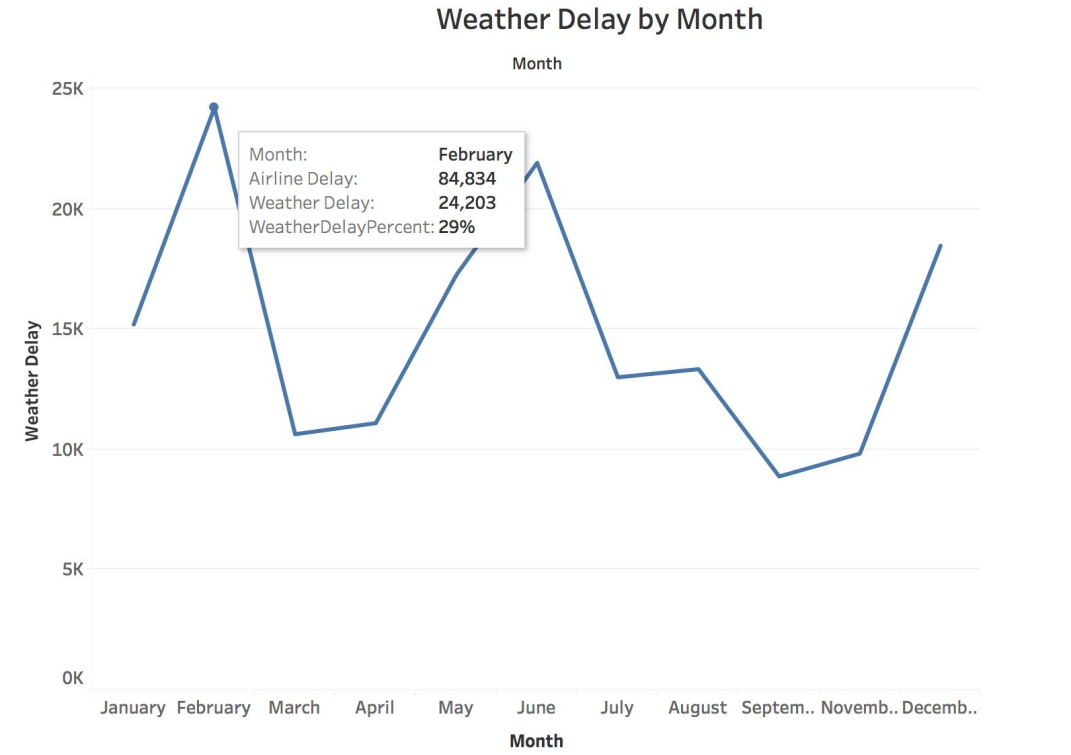


*Average airline delay as compared to average distance and airline delay compared to average air time. (Distance shown in miles. Time shown in minutes)(**Dashboard 3**)*

This dashboard examines the relationship between airline delay and average distance, and also airline delay and average air time. Though there are outliers, it does appear that a low average air time is often related to a low average airline delay and a low average distance appears related to a low average airline delay. Looking at the month of February, there is a slightly greater spread in average airline delay, even while average distance remains low, but with fewer outliers. Looking specifically at Augusta Regional Airport in the month of February, the

average distance is relatively low, but the average airline delay is very high. Average airtime is also extremely low while average airline delay is high.





Sum of airline and weather delays by month

While it could certainly be the case that weather delays are to blame for airline delays, this chart shows that weather delays are actually a relatively small percentage of airline delays and are nearly at their lowest in November, one of the months with the highest overall airline delays. Even at their peak, weather delays in February are less than 30% of airline delays.

After analyzing the data, it becomes clear that a variety of factors appear to be correlated with airline delay, including arrival and departure delays, longer distances, and greater airtime. However, the factors with the largest relationship to higher than average airline delays seem to be the month of the year and the location of flights. Although this insight cannot be explicitly drawn from the available data, it appears that the human factor (perhaps surrounding holidays and travel, as well as location and airport), may be the greatest influence on significantly higher than average airline delays.

Looking at the data as a whole, it appears that, though there are a number of factors at play, if you're looking for flights with fewer delays, you might want to stick to shorter flights with less air time, as well as trying to avoid flying in June and the end of the year (assuming that this data has remained consistent for the last few years and will continue to remain relatively consistent). You would also want to keep an eye on which airport you're flying out of, since delay figures in specific airports can be influenced by a number of factors.

Also, you might want to avoid flying out of Pago Pago…

Resources:

- flights.csv
- Flights - link here: https://www.kaggle.com/usdot/flight-delays/data
- US Demographic data - link here: https://www.kaggle.com/muonneutrino/us-census-demographic-data/data
- Youtube US data - link here: https://www.kaggle.com/datasnaek/youtube-new/data
- Video Walkthrough - link here: https://www.youtube.com/watch?v=9xqHA732LMA&feature=youtu.be