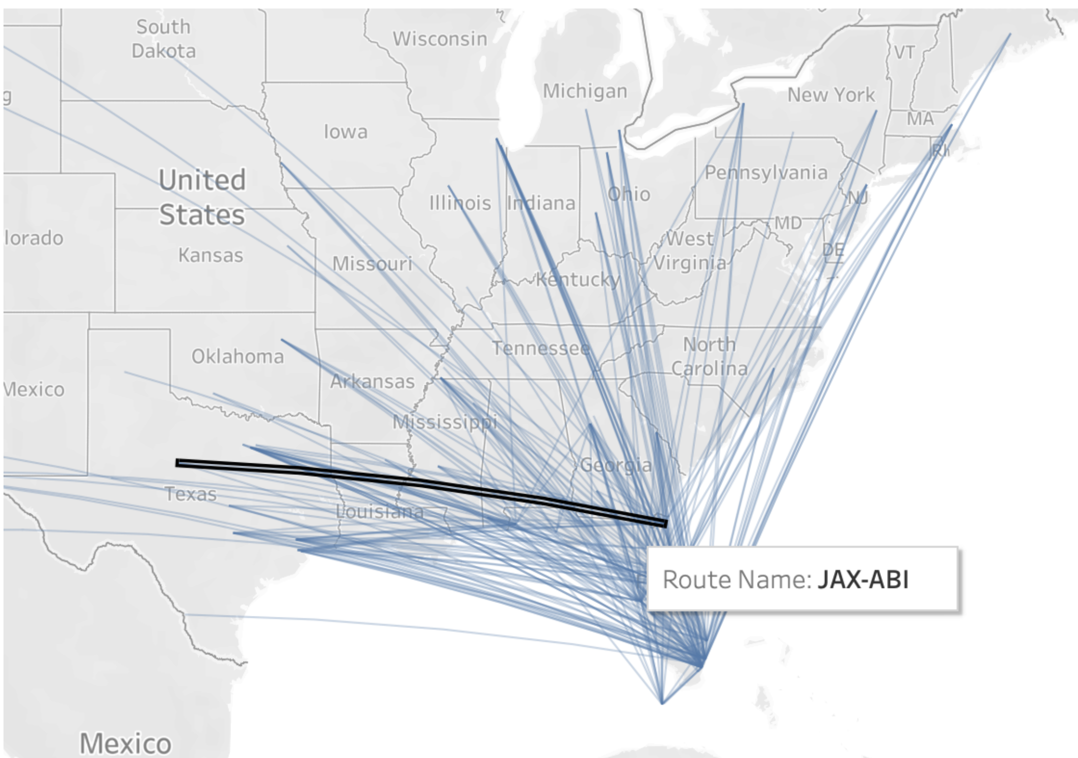
When looking for a dataset that fits the theme of tolerance, I came upon a US aviation accidents dataset which has crash records of airplanes, helicopters, balloons, gliders, and many more vehicles as well as how many people did and did not survive. Everytime people fly, there's always a risk of something going wrong. Some risks, like flying a commercial plane to a different state, are much more reasonable than gliding off a mountain. I believe that this perfectly encapsulates tolerance as US citizens realize that there are risks associated with aviation accidents that seem to be generally accepted in our society. Thus, my goal and question when creating these visualizations are what are my chances of dying given different crash conditions?

After first realizing the US had tens of thousands more crashes than other countries, I removed the other countries from the picture. Naturally, when trying to figure out how many flights ended up crashing, I started looking for a dataset with all domestic US flights that did not crash in order to create a proportion of the number of flights that crashed. Many hours were spent finding and joining this dataset to my accidents dataset, and even more time was spent joining more tables that had the latitude and longitude of the airport codes. I got to the point where I had a visualization of all flights coming from a chosen state, such as Florida in this example. At the time I felt that this view had a lot of potential to dive deeper into my data.



My vision was to add in the red lines for the flights that crashed. However, I unfortunately realized that although airplanes were a majority of my aviation accidents dataset, there were over 18,000 different makes of airplanes. After more investigation, most of the crashes came from smaller personal planes that were not in the total domestic flights database that I was trying to join my accidents dataset to.

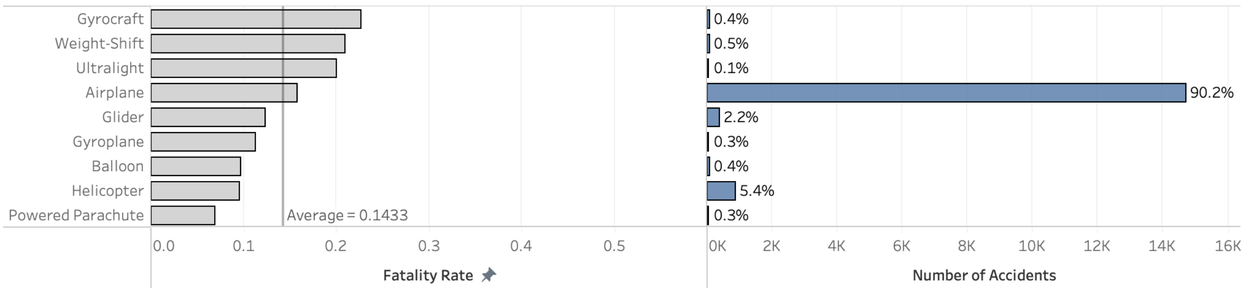
With a better understanding of my dataset and scrapping the idea of joining my accidents dataset to these new ones, I still did not have anything to create proportions out of. I did not want to use the count of accidents because the map would turn into a population density map as the number of people in an area increases, the number of people who fly aircrafts increases, which subsequently increases the number of accidents occurring.

In class discussion, someone recommended that I compare the number of plane crashes to the number of car crashes. While exploring car accident datasets was interesting and could have also turned into a great project, it dawned on me that instead of focusing on the proportions of crashes, I could create a fatality rate measure in my aircraft accidents dataset by taking the sum of fatalities divided by the total number of passengers on the aircraft. Now that I had this new value, I was able to explore deeper into my eda. My first step was excluding 2021 and the years before 1982 since these data were not complete. I then took a deeper look into months of crashes thinking that around the holidays the sum of fatalities or fatality rate could be higher as more people travel, but I found out that time of year had no effect on the number of crashes or how fatal they were. However, when looking at the weekday LOD, it became apparent that there were more fatal crashes on the weekend and Wednesday. The weekend makes sense because people have more free time to fly their vehicles, but I am really not sure (even after some research,) why Wednesday has a much higher fatality rate than the other weekdays.

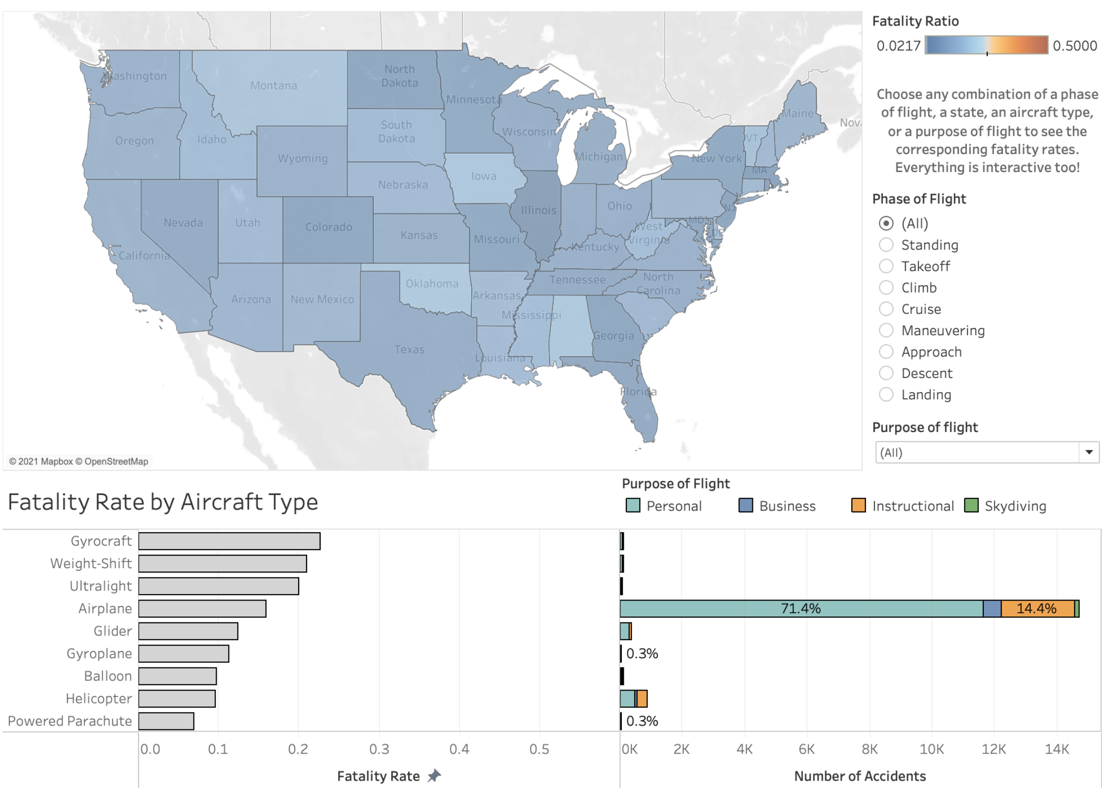
Additionally, one of the variables was Aircraft Damage with values of minor, substantial, and destroyed. As expected, destroyed vehicles had the highest fatality rate by far. It was at this point when I got to the phase of flight variable, where I found much more depth than most of the other variables in my dataset. Upon further investigation, different vehicles had different fatality rates based on which phase of flight the malfunction occurred. This was one of the first filters I put on my map for my interactive visualization.

When I added a year filter to my map and scrolled through the different year's fatality rate by state, I realized that there was not complete data each year for every state and aircraft. This makes sense because although aircraft accidents do occur, they are not super common. As a result, I decided that for my interactive visualization I would not put a year filter on my map and instead have the focus be on how the phase of flight and other factors affect the fatality rate.

This being said, I realized I needed to incorporate the fatality rate of each type of aircraft in my dashboard. For this, I used a simple bar graph with a fixed scale to make more accurate comparisons between phases of flights. Happy with what I had so far, I realized that I needed to account for the fact that airplanes make up a vast majority of aviation accidents. To help show this in my graph, I added another bar graph next to the first one that shows the total number of accidents per aircraft type. This is what it looked like:



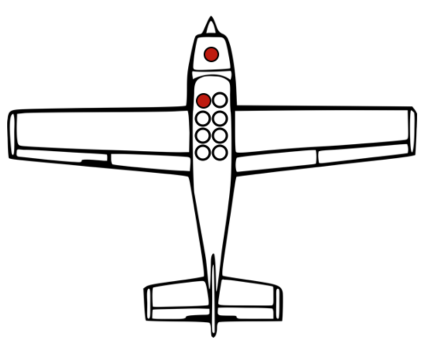
The percentages next to the bars represent the percent of total crashes each aircraft accounts for. I also put the average line in the left bar chart to help make comparisons between the average fatality rate for the selected phase of flight or state to other combinations of filters. To bring more depth to the right bar chart, I put the purpose of flight on color. This messed up the percentages as now each section of the bar has a percent of total, and some bars do not even have a percentage anymore since they are so small. I could not figure out how to turn the percentages back like they were in the above chart and decided that it was not too big of a deal to have the percentages on the inside as they were still putting in perspective how small a percentage of total crashes some of the aircraft categories have. Additionally, I think that the addition of a new layer of "purpose of flight" on color is a good trade off for the weirdly formatted percentages. This new color brings more depth to the sum of total fatalities, and allows the user to answer even more questions they have.



This is what the almost finished dashboard looks like. I added another filter for "purpose of flight" so that the user has more flexibility in exploring whatever questions they have. I also kept the "Fatality Rate by Aircraft Type" as grey bars to not get confused with the other colors going on in the visualization. To help explain what was going on and how to use the visualization, I added simple instructions above the filters. When choosing which color to use for the map, I decided to use a blue-white-orange scheme to differentiate the high and low fatality rates. These colors helped make the low and high fatality rates stand out more than a traditional white-red color scheme.

When playing around with the different filters, I found some interesting insights. I decided that the best way to convey my newfound information was through Tableau stories, as they allowed me to preset the dashboard for each "slide" and put text on the dashboard explaining what's going on. While the dashboard is the same for the first 3 slides, the fourth one has an average line in the bottom left barchart, emphasizing the takeaways from the slide. I will leave the Tableau story to the storytelling, but to say the least it is very interesting to see where a little research can take you in understanding the data.

Now that my interactive visualization was done, I was able to start on my static visualization. I had many ideas on how to communicate my data in a new way. I really wanted to take some of the most well known aircrafts and put people (circles) in them representing how many passengers could fit in them. Then, by shading some circles red, I could create a visual showing the fatality rate for the vehicle if it crashed. For some reason, the formatting of the circles was very weird. I think I accidentally created some sort of layers which did not allow me to move the circles individually, and thus I had to move every circle in the visualization all at once. This problem took me a very long time to solve, but after redoing the circles where I could move them individually I was back in business.



The above example shows that 2 out of 9 people on average will die from a personal plane crash. These images were much harder to find than originally planned. I spent a solid hour searching for these outlines. From terrible quality images to watermarked photos to outlines that were asking me to pay, I was very close to drawing my own images and importing them into Illustrator. In an earlier iteration I was using black silhouettes, but it was much harder to distinguish the passengers and fatalities. Also in an earlier version, I had 3d aircrafts flying on an arc. This idea also did not work as I could not find 3d outlines of many of the vehicles.

As I mentioned earlier, from deeper eda into the phase of flight, I discovered that different vehicles have different phases of flight where their fatality rate is the highest. I thought this would be really cool to represent by an arc underneath each aircraft, with the red area representing the deadliest part of the flight for an error to occur. When I asked for feedback from my friends on my visualization, they said that the blobs looked like blood drops which fit the theme of danger in my visualization. During presentations in class, this was also the part of my project that people said they really liked and thought was creative which gave me joy knowing how much work I've put into this project.

At the bottom of the graphic, I decided to bring the time variable into a scatter plot since time was not incorporated into my interactive visual. The scatter plot allows us to see that the number of crashes where many people died dropped drastically in 2001, most likely as a result of 9/11 and a 253 casualty air crash in November 2001 - one of the worst aircraft crashes in US history. Although this is not mentioned in the visualization, the fact that most large aircraft crashes happened during good weather conditions is pretty interesting as it went against my initial inclination. Upon further research, a majority of accidents occur when pilots "misread flight equipment or fail to properly address mechanical error" which does not depend on the weather conditions (Hedlund). To help better emphasize the large casualty aircraft crashes, I put the total fatal injuries on size. This might be a bit repetitive since it is also the y-axis, but it helps show the intensity and importance of the higher casualty crashes which is what I wanted to emphasize.

To the right of the time graph, I decided to add a bar graph that summarizes the arcs below the aircraft outlines. This graph breaks up the phases of flight by aircraft where the height of the bar represents the aircrafts fatality rate and the color represents the total fatal injuries. What is apparent from the average lines is that the higher the altitude an aircraft is when something goes wrong (you can look at takeoff vs cruising,) the greater the fatality rate is. Also, although airplanes do not always have the highest fatality rate, they account for the largest number of total fatal injuries. This is because there are more planes in the US than there are other aircrafts.

All in all, I think that both my static and interactive visualizations answer my question of what exactly the fatality rate is given various crash situations. Some takeaways from this project is that if a state has a much worse fatality rate than the other ones (such as helicopters in Va,) then that state can change their safety protocols on the particular aircraft and create new programs to better train their pilots. Additionally, if you had many options of methods of flight in your state, you could look at the interactive visualization to see which vehicle has the best fatality rate in the event of a crash. This is an important topic since it deals with preserving human life. Additionally, as mentioned before, many groups of people can benefit from this visualization as engineers can see during which stage their aircrafts tend to fail, pilots can see which stages of flight to be cautious of, and passengers can better choose their method of transportation given their purpose of flight, state, and other variables. There are many more insights to extract from this data, but I think that the visualization portrays which areas to improve on and the tolerance of the risks Americans take on aircrafts everyday.

Overall there was a lot I learned about Tableau and Illustrator through this project. I believe that this is the best work I have produced this semester, and I am very happy with the outcome. Through this class I have learned that everything takes many, many iterations, and that sometimes the best ideas come when you are not even working on the project. This plays into the theme that great visualizations take time, and that the more feedback you get the better the result will be. I've really enjoyed this class this semester - Thank you TA's for the helpful feedback on all of our work this semester, and thank you professor Fields for teaching us the elements of power visualizations and for patiently helping us improve our work in office hours.

Works Cited

https://www.baumhedlundlaw.com/aviation-accident/why-planes-crash/#:~:text=Many%20aviation%20accidents%20are%20caused,one%20reason%20why%20planes%20crash.

Data Source: https://www.kaggle.com/khsamaha/aviation-accident-database-synopses