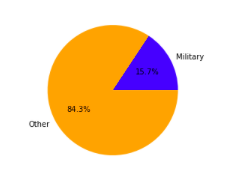
GROUP WRITE UP

Nahid Farzad, Abbie Kent, David Coy, Maxwell Moorman, Jessika Hall

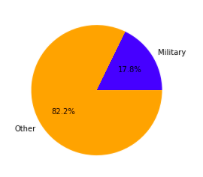
*Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots.*

1. **Do military operators have more crashes opposed to  other types of operators (ie: commercial, recreational)?**

We originally believed that because military operators conduct more dangerous flights, either test flights or mission-oriented flights, therefore inducing more plane crashes. However, once we isolated the operator variable; “military” and “other” we were able to determine that military- operator crashes actually accounted for the minority of crashes.



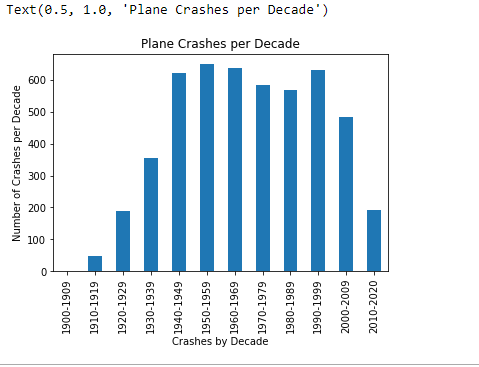
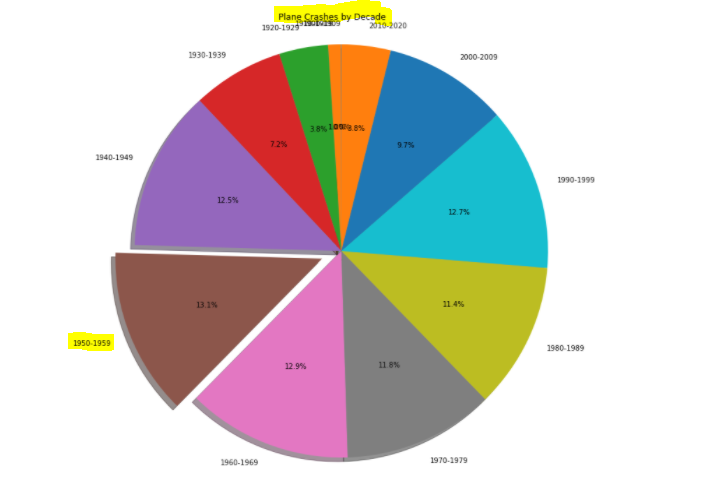
Next, in order to get a better view of present- day crashes, we looked at the last ten years of crashes. After pulling the necessary data for the past ten years, we were able to determine that military led flights still account for the minority of crashes. We thought there may have been a change in military vs other types of crashes due to the advent of technology or changes in world affairs but the past ten years show a similar percentage to the overall crash data.



After running this part of the analysis, we were able to determine a few conclusions. The particular analysis of comparing military to all-other operators is not necessarily always conducive to determining if a certain type of operator i.e. military, causes more crashes.  Additionally, our data set contained 2242 different types of operators; making it extremely difficult to evaluate operator- led / operator- caused crashes.

1. **Do certain decades have  a higher rate or global plane crashes?**

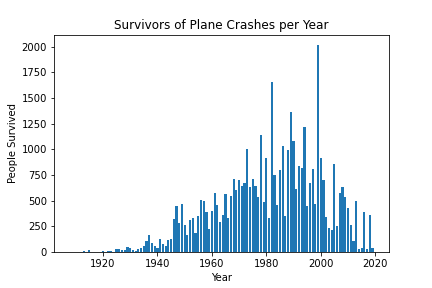
Prior to manipulating the data set, we believed that earlier decades would have higher rates of plane crashes due to less- advanced technology. We also believed that periods of war, i.e. the 1920s and 1940s may have higher rates of crashes due to new technology and dangerous conditions.

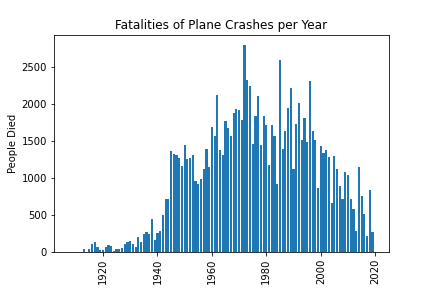


In order to visualize crashes per decade we extracted the “year” from the date.time format. Next we were able to sort each crash into the correct decade bin. Once sorted, we graphed our findings into a pie and bar chart. Once visualized, we were able to see that 1950 - 1959 contained the most crashes per decade. Additionally, when looking at the decades with the most amount of crashes in the pie chart, we determined that there was only a small percentage difference (1.7%) between 1940 and 2000. On the other hand, our bar chart seems to be a little less informative than the created pie chart.

When looking at crashes per decade, we determined that there were a few limitations. First, aviation, starting in 1908 was a very new industry. As the decades progressed, flight travel became more nuanced and a necessity for some. Therefore, it is difficult to draw conclusions from crashes per decade given the limited amount of data in the first few decades. It would have been helpful to have a larger data set full of flight data - not only crash data.

1. **How many people on average survive a flight?**

We suspected that more people would survive plane crashes as planes became more developed and used throughout history. In order to run this analysis, we first calculated the number of survivors from the aboard passengers and fatalities from the crash.With that data, the following figures were created:

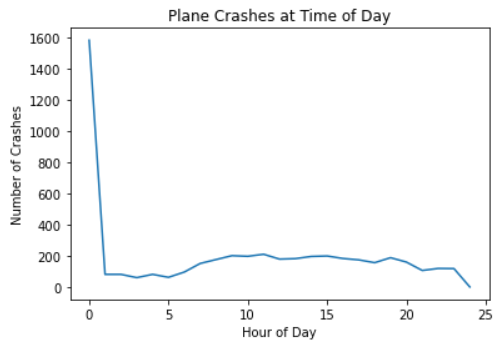


After visualization, it is hard to determine whether more people survived plane crashes later in history because our data set only contained plane crash data. Although the number of fatalities decreased in the last 20 years.

It appears that while the 1950’s had the highest number of plane crashes, the number of fatalities was less significant than other decades.

**4. Do planes crash at a higher rate based on time of day?**

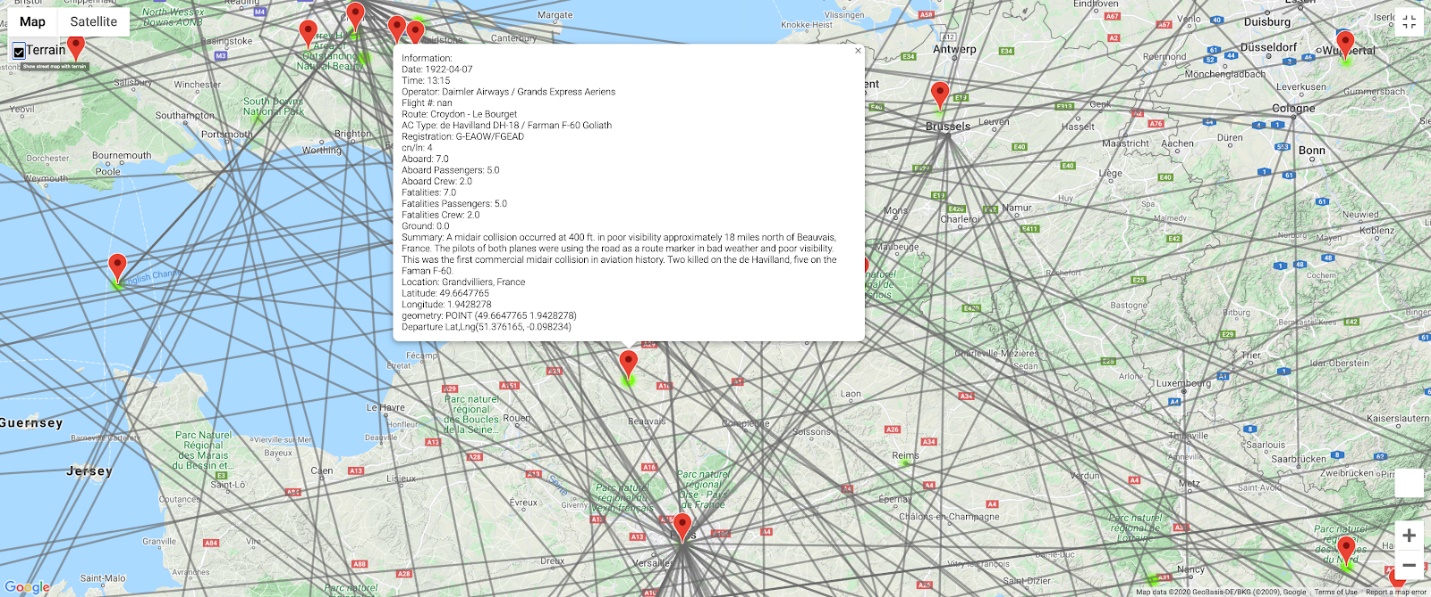
Initially, we thought that planes would crash at a higher rate at night. When filtering the data, we saw that not all plane crashes had a particular time associated with the crash. Therefore, each empty cell was filled with an empty hour minute timestamp (00:00). Unfortunately, this created problems when determining the number of plane crashes at midnight. The midnight timestamp was then removed to create a better visualization.

Our initial hypothesis was incorrect as the highest number of crashes occurs at around 10 AM.   
  


**5. Visualization of Plane Crashes on WorldWide Map**

Using Jupyter GMAPS import and thousands of google API calls for location information (latitude, longitude, country, city, vicinity), we were able to chart plane crash location worldwide and the intended route prior to crash.

There was a noted increase in plane crashes starting in the northeast United States as one would expect as the area was the birth of aviation and later, more people could afford to fly.



The number of US flights per last 50 years was also charted by decade. Lastly, flights with more than 100 fatalities were charted based on the site of plane crash and route taken prior to crash. There is a very noticeable increase in plane crashes not just in the United States, but the world throughout the years. Not until recently, does it appear that aviation has been made safer and we see a decrease in fatalities.’