Joseph Harvey, Flatiron Data Science, Phase 1 Project

Analysis of Aircraft Accidents in the NTSB Aviation Accident Dataset

The company is interested in becoming involved in the aviation industry. Specifically with owning and operating aircraft in a commercial endeavor, which would indicate an interest in airplanes or helicopters.

Overview of the Dataset

The NTSB aviation accident database that I used in this presentation contains data about civil aviation accidents that occurred from 1948 through 2022. The aircraft categories for these accidents include every conceivable kind of aircraft from airplanes and helicopters to hot-air balloons and powered parachutes.

There are 31 fields of data for each of the 88,889 accidents in the database. These fields are Event Id, Investigation Type, Accident Number, Event Date, Location, Country, Latitude, Longitude, Airport Code, Airport Name, Injury Severity, Aircraft damage, Aircraft Category, Registration Number, Make, Model, Amateur Built, Number of Engines, Engine Type, FAR Description, Schedule, Purpose of flight, Air carrier, Total Fatal Injuries, Total Serious Injuries, Total Minor Injuries, Total Uninjured, Weather Condition, Broad phase of flight, Report Status, and Publication Date.

Methodology and Cleaning the Dataset

Since a large part of this analysis concerns the type of aircraft, it was imperative that the Aircraft Category field be as complete as possible. After removing some duplicate records, I was left with 87,951 accident records. The category field contained about 32,000 values. This is only 37% complete. The Make field (which represents the manufacturer of the aircraft) was 99.9% complete, and I found it to be useful in helping to complete the Category field, since for makes such as Cessna, the resulting category was either airplane or empty (NaN). Therefore, I could fill in all the empty fields for Cessna with “airplane”. The make Sikorsky was exclusively helicopter, so its empty category fields were filled in as well.

After cleaning the Make field for misspellings or alternative versions of makes, and filling in the category fields for all the particular makes that could reasonably be filled in with one category, I was left with a category field that was 94% complete (82,608 records), with only about 5300 “Unknown” aircraft types.

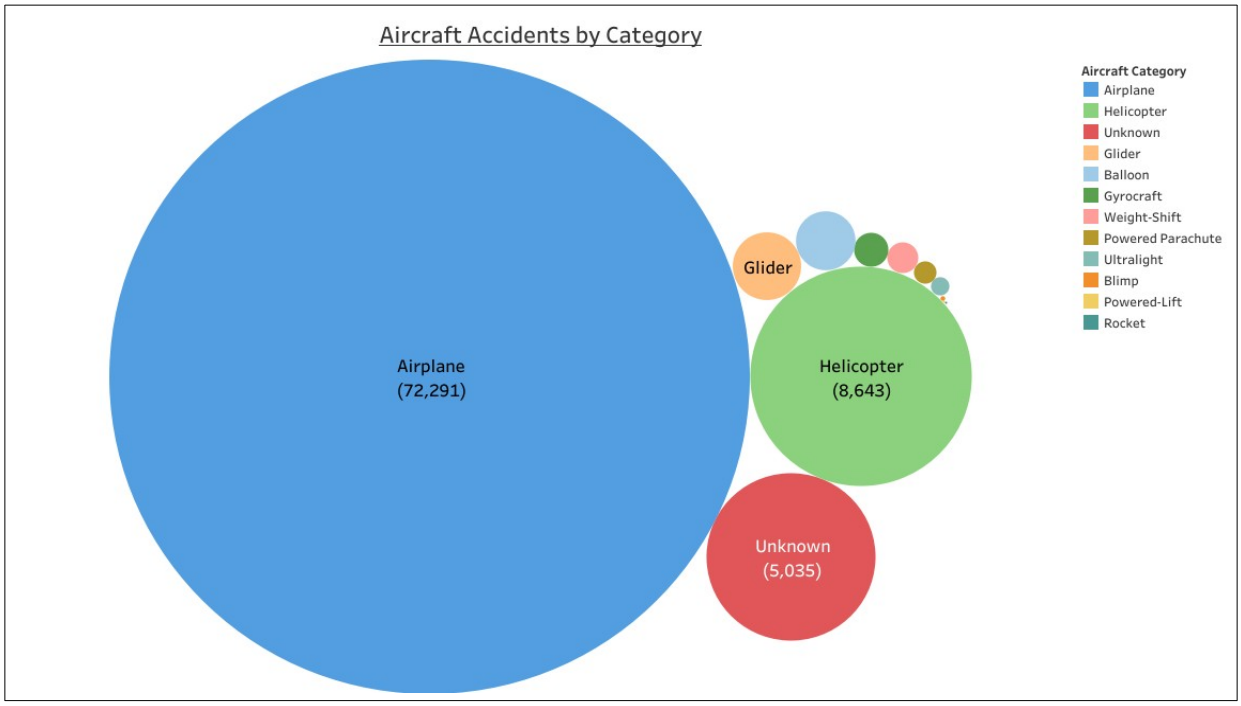
The fields that yielded the most useful information for this analysis are:

- Event Date: the date of the incident
- Aircraft Damage: the level of damage sustained, from destroyed to minor damage
- Aircraft Category: the type of aircraft involved in the incidents

- The 3 injuries fields (Fatal, Serious, Minor) and Total Uninjured: number of the injured and uninjured
- Broad phase of flight: from taxiing and takeoff to landing
- Report Status: the conclusion drawn from the investigation regarding the main cause of the incident

Aircraft Categories

Of the almost 88,000 accidents in the dataset, 72,291 are airplane accidents and helicopters account for over 8,600. As the below graph shows, over half of the remaining aircraft categories are unknown. Since over 90% of the records concern airplanes and helicopters, I will focus on those two aircraft categories.

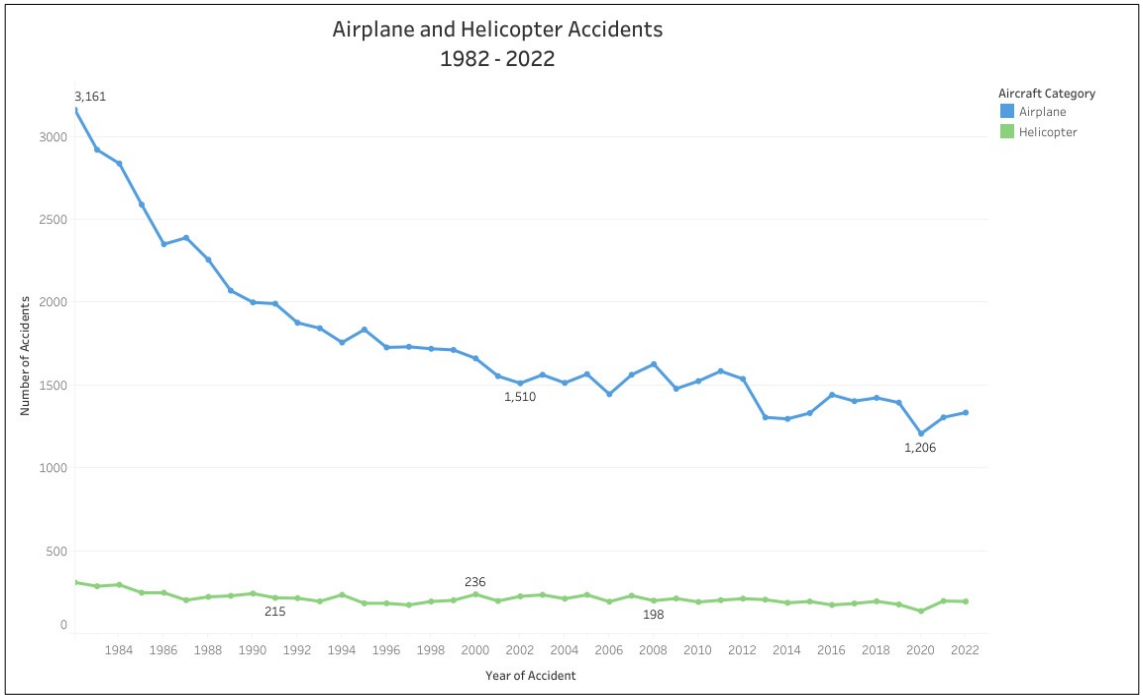


Timeline of Incidents

The dataset contains just a handful of entries prior to 1982, so I start my timeline at 1982 and continue through the end of 2022. The number of helicopter accidents is fairly constant over the 40 year period, hovering at around 200 accidents for most years.

The more interesting timeline concerns the airplane accidents. As you can see in the following chart, the number of incidents start at over 3,000 in 1982 with a definite downward trend over the next 40 years. This speaks well for airplanes as a possible venture, as the overall safety has drastically improved apparently.

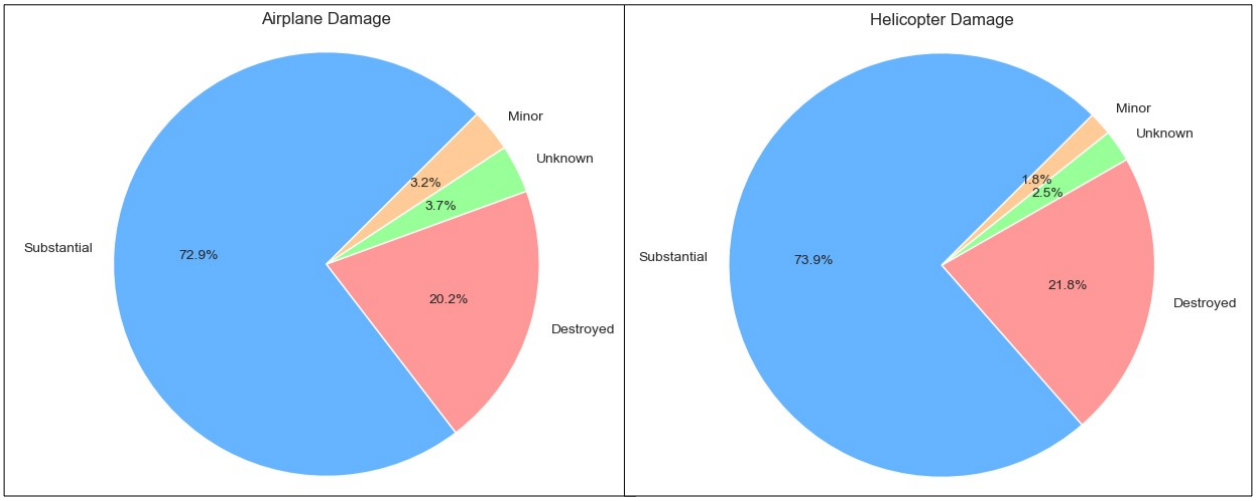
One interesting side note regarding this timeline: you can see that both helicopters and airplanes had their lowest number of accidents in 2020, the first year of the Covid-19 pandemic. This was most likely caused by the fact that there were far fewer aircraft flying during 2020. Those figures are not present in the NTSB dataset, but I feel it is a safe assumption.



Aircraft Damage

The dataset contains a field called Aircraft Damage. Damage is determined as Minor, Substantial, or Destroyed. No further details are provided concerning the damage levels.

The following charts show the percentage of damage levels for airplanes and helicopters in the dataset. At first glance, they look very similar and that is because they are. Substantial damage occurred in over 70% of both airplanes and helicopters, and over 20% in both categories were destroyed. The lesson here seems to be that if an accident during flight does occur, it is most likely going to be an expensive accident; therefore, a company operating aircraft needs to pay the utmost attention to safety in its operations in order to protect its investment.

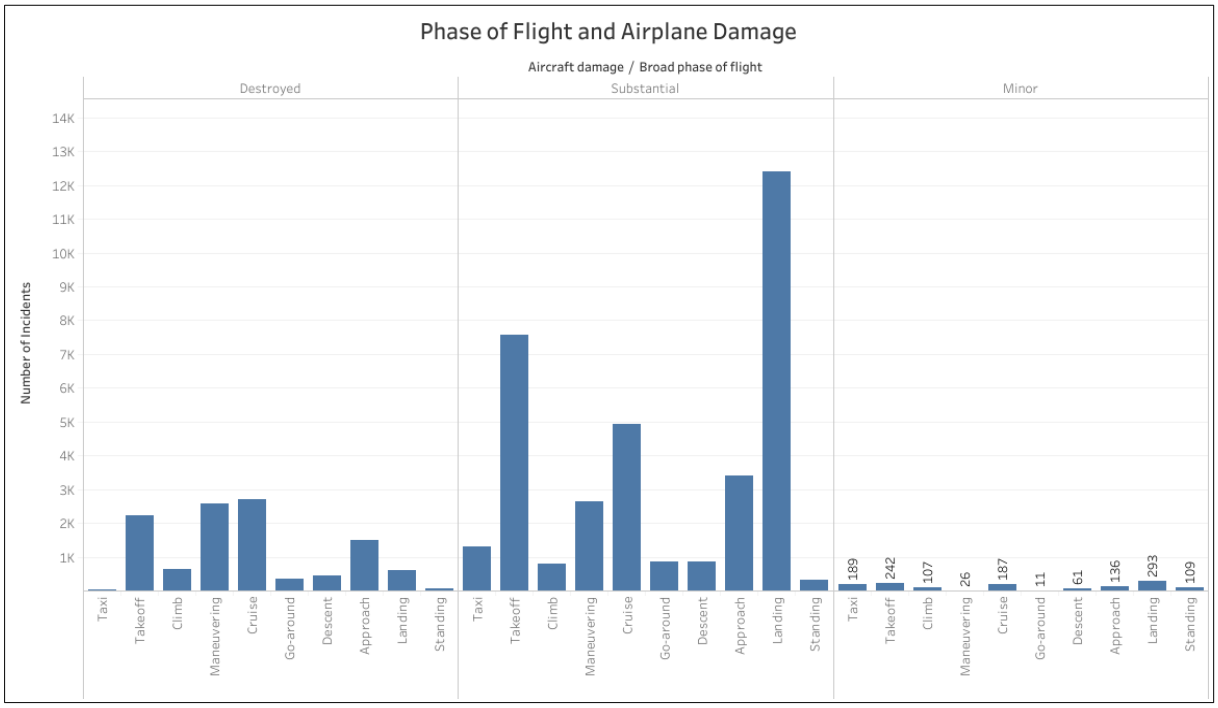

Airplane Damage
Minor
Unknown
Substantial
3.2%
3.7%
72.9%
20.2%
Destroyed
Helicopter Damage
Minor
Unknown
Substantial
1.8%
2.5%
73.9%
21.8%
Destroyed

Broad Phase of Flight and Aircraft Damage

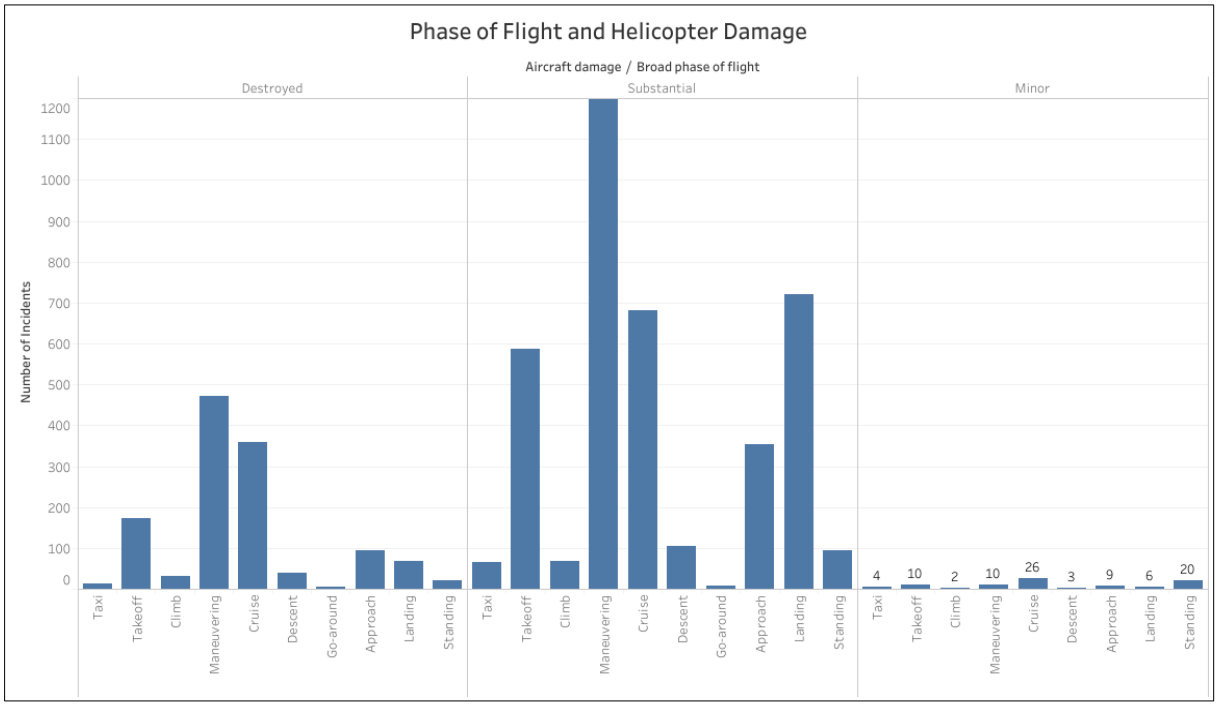
Broad Phase of Flight is a field for the accidents that indicate in what stage of flight the aircraft was in when the incident occurred. Since over 90% of the accidents result in substantial damage or destruction, it may be interesting to see when these incidents occur during the flight.

The following graph shows that both substantial and minor damage occurs in airplane accidents most often during the take off and landing stages of the flight, the beginning and end of the trip when the plane is lower in altitude and still relatively close to its origin or destination runway. The cruising portion of the flight (the longest portion for long-distance flights) shows a spike as well.

For the airplanes that were destroyed, we see that the end stage of the trip is not as much of a factor than the takeoff and middle portion of the flight. This makes sense for destruction since if a catastrophic event happens when the airplane is in its high-altitude cruising stage of the journey, the aircraft will be in a position farther from a safe destination (even an emergency landing) and the chances of destruction are much higher.



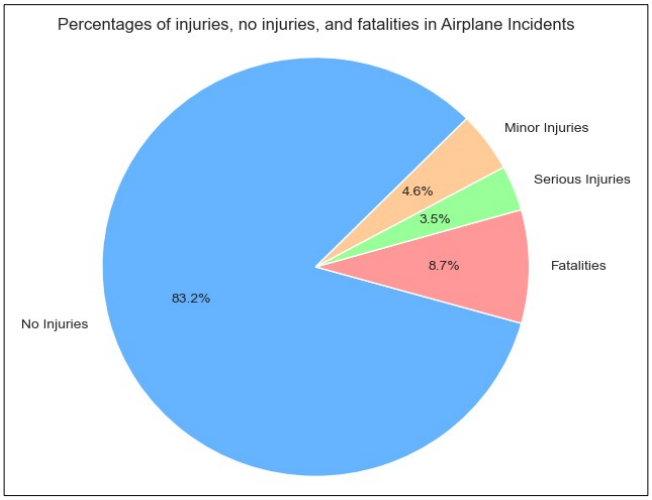
For helicopters, we see a different story in its phases of flight. For all 3 damage levels, the most common stage of the trip when the accident occurred is during the maneuvering and cruise portion, when the helicopter would be at its maximum altitude for the trip. But for substantial damage, take off and landing still play a significant but smaller role.

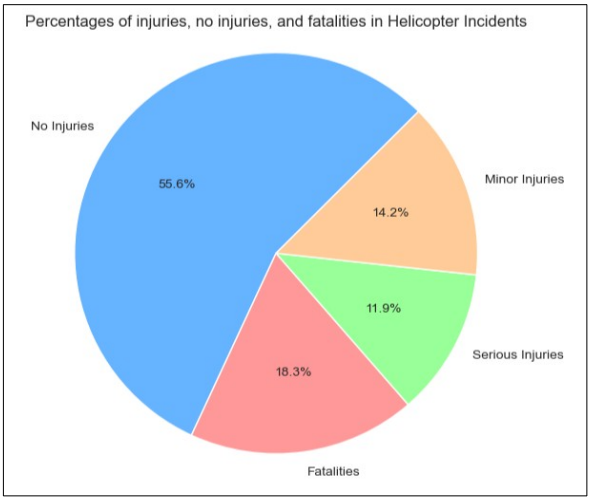


Injuries and Fatalities in Aircraft Accidents

Of even more importance for the owner/operator of aircraft is the safety of its crews and passengers. The NTSB accident dataset provides four fields for each incident concerning injuries and fatalities incurred during the accident. These fields are fatalities, minor injuries, serious injuries, and no injuries.

The two graphs below paint quite different stories for airplanes and helicopters in the percent of injuries and fatalities. As we see, the chances of experiencing no injuries in an airplane incident are about 83% and injuries and fatalities account for the remaining 17%. In helicopter accidents, however, the chance of injury or fatality is almost 45%.



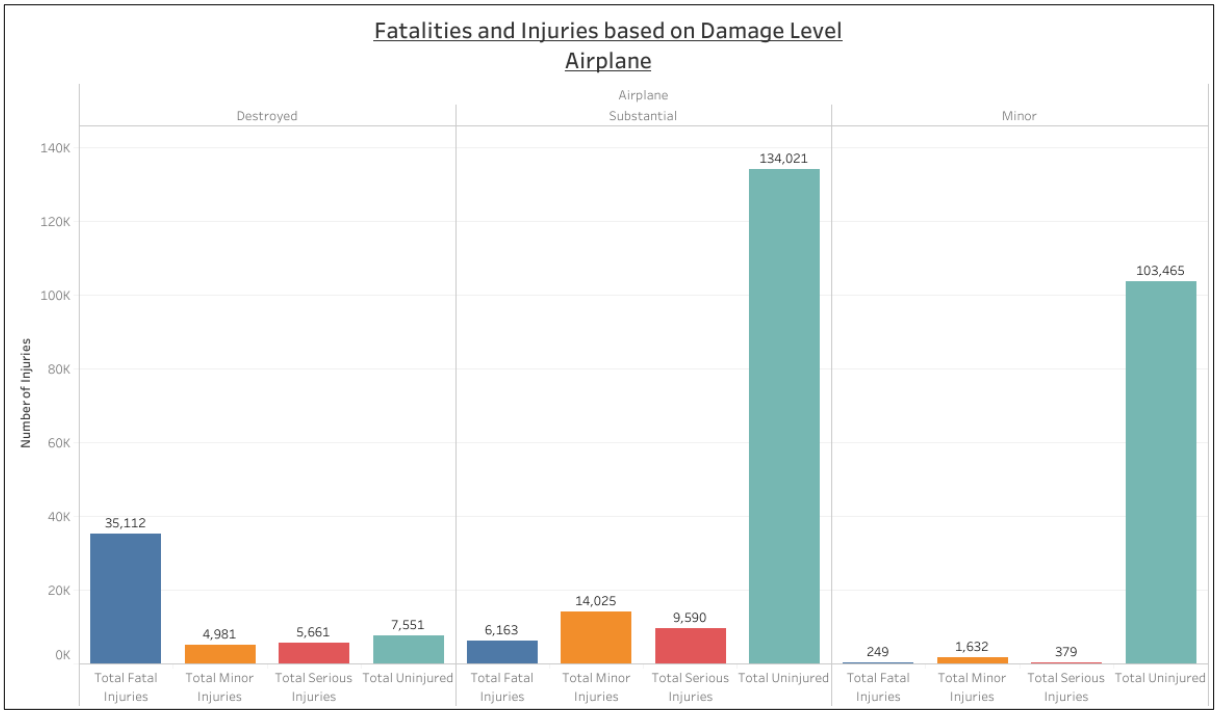


So we can conclude here that one is 3 times more likely to be injured and twice as likely to be killed in a helicopter accident than in an airplane accident.

What about the aircraft damage level as it relates to injuries or death?
The chart below shows the numbers of injuries, fatalities, and no injuries over the three different levels of aircraft damage: Destroyed, Substantial, and Minor damage.

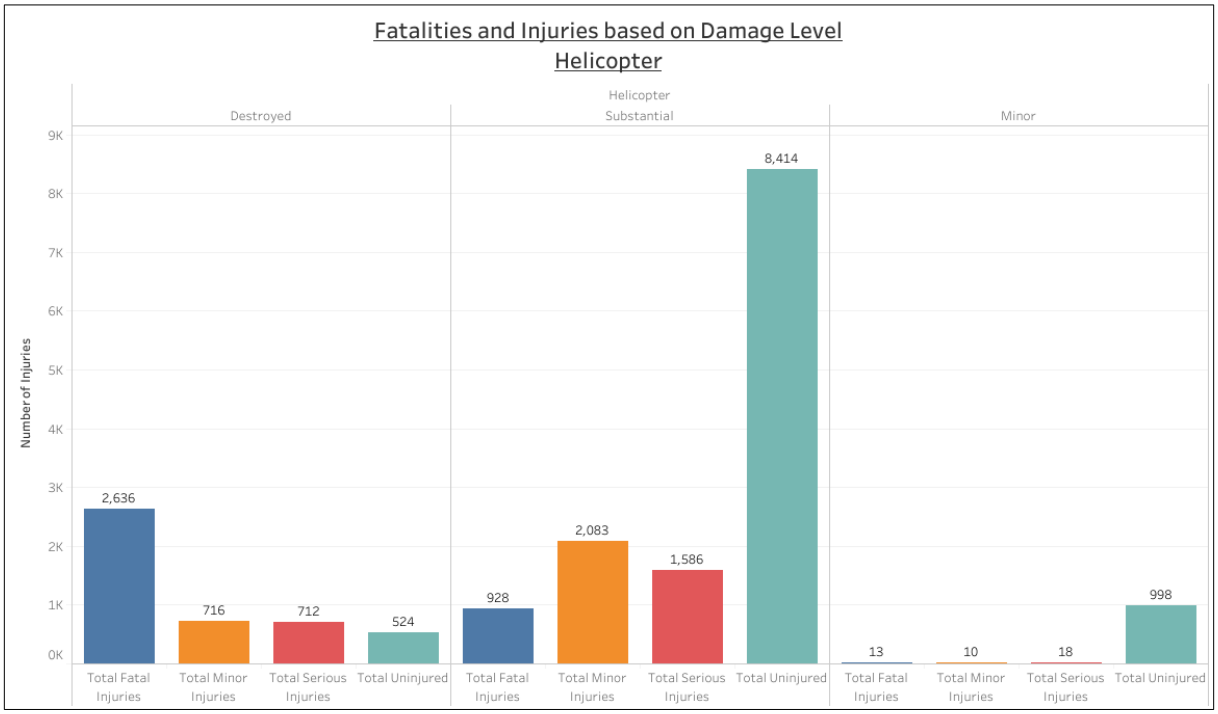
In airplane accidents that result in substantial or minor damage, the chance of there being no injuries is dramatically higher than the chance of injury or death. In incidents where the aircraft is destroyed, the chance of fatal injury is almost 5 times more likely than no injury. No surprise there.

For substantial and minor damage combined, the chance of fatal injury is only 2.5%. For destroyed airplanes, the chance of fatality is 66%.



As you can see in the following, in the case of helicopters, the story is pretty much the same. No injury is the most likely outcome in accidents resulting in substantial or minor damage, and fatal injury is most likely in aircraft that are destroyed.

The numbers do differ slightly when you consider the percentages. In airplane accidents resulting in substantial damage, 3.8% of the injuries are fatal. In helicopters that same calculation is 7.3% fatal. In destroyed helicopters, the fatality rate is 58%, a slightly smaller number (remember that fatalities in destroyed airplanes is 66%).

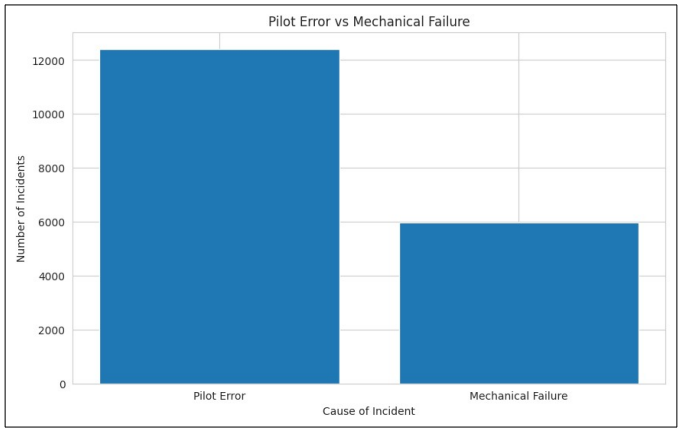


These comparisons lead me to conclude that a crash in a helicopter that results in damage (but not destruction) is a more dangerous situation for the people on board than the equivalent incident in an airplane. Destruction in both aircraft categories likely results in fatality.

Reason for Accident

The dataset provides a field called Report Status that details the reason for the aircraft accident. Unfortunately, this field is mostly non-informative, containing values such as Probable Cause, Unknown, Foreign, or Factual. In a textual sorting of keywords, I was able to identify that a number of values contained the words “pilot’s failure”, and these values constitute a human-error factor.

About 14% of the records (12,414) indicate pilot error as the main cause of the incident. Another 6.8% (5,966) contain a variety of causes for the incident, most of which point to mechanical or equipment issues.



So of these 18,380 informative values for Report Status, almost 68% are attributed to pilot error and about 32.5% attributed to various mechanical or equipment failures, many due to undetermined causes and some caused by human error in maintenance of equipment.

My takeaway from this section highlights the requirement of an aircraft operator to insure the proper training and continuing oversight and education of its pilots and maintenance staff in order to reduce the frequency of human error in the operation and maintenance of its aviation equipment.

<u>This analysis points out several useful takeaways:</u>

- The frequency of airplane accidents over the 40-year period from 1982 through 2022 dramatically decreased, whereas the frequency of helicopter incidents remained quite steady. It would be reasonable to chalk this up to advances in technology and safety infrastructure in the airplane travel industry. Therefore, airplane travel is more reliable and safer than it was in 1982, whereas travel by helicopter is relatively just as safe as it was 40 years ago.
- Regarding the safety of the humans on board aircraft, it is evident that the likelihood of injury or death in an airplane accident is much lower than in a helicopter accident.
- For destructive accidents in both airplanes and helicopters, the injury level is likely to be fatal.
- Helicopter accidents resulting in damage are more dangerous for the passengers than in the equivalent damage in an airplane accident.
- The importance of training, proper maintenance, and strict adherence to safety protocols cannot be stressed enough. Of the over 18,000 investigations that yielded an informative cause of the accidents, more than half were attributed to pilot error and about a third to equipment failure. These were accidents that most likely could have been avoided if not for human error in the operation and/or maintenance of the aircraft.
- The safety of air travel has significantly improved over the past half-century, but if an accident does occur during the operation of an aircraft, the likely result is

going to be substantial damage or destruction. There's rarely such a thing as "just a fender bender" in the airplane and helicopter industries.

Suggestions for further study

- It would be informative to aggregate some data concerning the overall number of aircraft in use in the industries over the same period as the accident dataset. This would give us the opportunity to evaluate the safety record of different makes and models of aircraft as we would have figures related to their overall representation in the industry. Then, recommendations could be made regarding what makes and models would be the best investment for the company from a reliability and safety standpoint.

- Completing the Report Status column of the dataset would be extremely helpful in providing more insights as to the safety protocols, training, and maintenance of the aircraft industries.