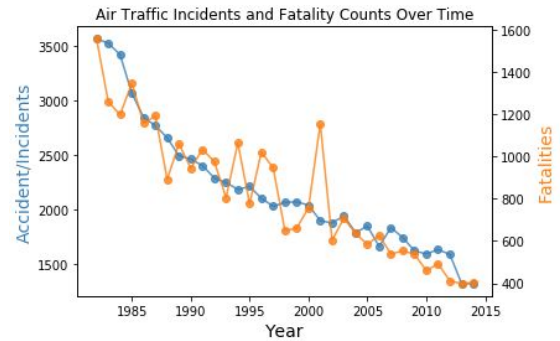


NTSB Air Incidents Data Exploration

The National Transportation Safety Board (NTSB) collects data about air traffic incidents and accidents in the United States and provides a written report for each occurrence. From the NTSB data inferences about the relative risk of air travel can be made. As shown in the figure to the right, both the number of incidents and fatalities have been decreasing over time, suggesting that air travel is becoming safer over time.



To learn from the over 70,000 flights recorded in the data, we determined common topics in the probable cause reports using Latent Dirichlet Allocation (LDA). LDA is a machine learning technique that determines topics in text based on the occurrences of words in each report. LDA distributes the topics across the reports, giving each report a score for how well they fit each topic. By examining the words that are most associated with each topic, a general sense of the circumstances that cause aircraft accidents can be determined. For example, Topic 3 included top terms such as “student”, “instructor”, “loss”, and “control”, perhaps suggesting an instructional flight in which control of the plane was lost. The themes of each topic in the NTSB probable cause reports are listed below. From these topics we can see that weather plays a large role in incidents (Topics 2, 7, & 8), as well as mechanical issues (Topic 6, Topic 1).

- Topic 1: Loss of power, lack of maintenance
- Topic 2: Dark/night conditions, weather
- Topic 3: Instructional flight, loss of control
- Topic 4: Inadequate fuel, pre-flight planning
- Topic 5: Collision, failure, maintain altitude
- Topic 6: Landing gear
- Topic 7: Snow, take-off, runway
- Topic 8: Gusts, wind, crosswinds

The NTSB data also includes characteristics of the flight including the air carrier, engine type and number, location data, as well as the purpose of the flight. To assess which flights possess the most risk the characteristics that made up the highest proportion of aircraft in the data was examined. Overall, the NTSB data suggests that smaller aircraft are the most risky flights. The most common makes of planes in the data, such as Cessna, Piper, and Beech primarily produce small aircraft for personal use. Additionally, of the 72,198 aircraft accidents and incidents documented in the NTSB data, 84.7% of incidents involve planes that have just one engine and 86.5% involved reciprocating engine aircraft, which are typically seen in small planes. These facts are summarized in the visualization below. These facts combined suggest that small planes make up the majority of aircraft incidents in the data, suggesting they are the most risky flights.

