Analysis of Various Factors Causing Deaths in Aircraft Accidents

By Madhur Bansal, Sambit Das, Sonali Saren

Group 17

Under the guidance of Dr. Dootika Vats

Acknowledgement

We group-17 would like to express our profound gratitude towards *Dr. Dootika Vats*, our academic and project instructor for MTH208A(Data Science Lab), for her guidance and constant supervision throughout the process and providing creative ideas and necessary information regarding the project which led to the completion of this project. It has been a great learning experience and has also provided us with a practical insight of the theoretical knowledge gathered during the course lecture.

Abstract

This project deals with analysing data on aircraft accidents across various locations over a period of 1938 to 2022. The motive of this project is to understand the relation between the numbers of casualties for different parameters of aircraft accidents like type of the accident, phase of the accident, location of the accident, etc. We have also tried to figure out from the data whether we can filter out information and conclude anything on the progress of the aviation industry , by studying the trend of the total number of casualties with time and how they vary for various aircraft manufacturers.

Introduction

There are various types of vehicular accidents that can occur; aircraft accidents being a very serious one of them. While it is unfortunate that these accidents occur, air travel only gets more reliable as the previous accidents result in more crucial safety improvements. There are various reasons for such accidents. Defects in aircrafts can cause equipment failures leading to malfunctioning of some parts, thus causing accidents. Other reasons like human errors caused by mechanics, ground workers and sometimes crew members also lead to mishaps leading to casualties. Analyzing data on aircraft accidents can give us valuable insights and important information on various factors influencing the number of such casualties.

Data Description

Data Collection

The data was collected from Wikipedia using techniques of web scraping in R programming language. Libraries like tidyverse and rvest and functions like html_elements and html_table were used to scrape the data from the webpage.

Description

The final workable data we obtained after cleaning contains 544 observations ranging over a period from 1938 to 2022 and providing information about the following:

1. Number of casualties further divided into four parts:
    - Total number of deaths
    - Number of deaths of the crew members
    - Number of deaths of the passengers
    - Number of deaths of the ground workers
2. The type of aircraft accident
3. The name of the manufacturers of the aircrafts met with the accidents
4. The location of the accident which gives us information about the country in which the accident took place.
5. Phase of the accident
6. Year of the accident

There are six types of aircraft accidents in our data. They are explained as follows:

a) COM: It is the abbreviated form of commercial type of accidents. It takes into account all the accidents involving a commercial flight.
b) EXG: It refers to the accidents where the aircraft is attacked externally using ground based weapons like ground to air missiles.
c) EXG: It refers to the accidents where the aircraft is harmed using attacks originating in the sky. For example intentional downing by a military aircraft will fall under this category.
d) INB: It refers to accidents resulting from internal attacks involving a pre-planned bomb, without hijacking.
e) INH: It refers to the accidents resulting from internal attacks to commandeer an aircraft. Weapons used for this purpose including explosives is coded in this category.
f) MIL: It refers to the accidents involving military aircrafts.

There are nine phases of accidents considered in our data. They are explained as follows:

a) STD: This is the standing phase of the flight that is prior to pushback or after gate arrival or when parked.
b) TXI: This is the taxi phase which is prior to take-off. In this phase the flight moves under its own power.
c) TOF: This is the take-off phase starting from the initiation of take-off power through to 10m altitude.
d) ICL: ICL refers to the initial climb phase which marks the end of the take-off phase. During this period, the aircraft climbs to a pre-determined cruising altitude (approximately 300m)
e) ENR: This is the en-route phase which starts after the initial climb phase and continues till the initiation of the approach phase.
f) MNV: MNV stands for maneuvering. It is the act of flying an aircraft close to the ground involving steep turns and aerobatics.
g) APR: This is the approach phase where the aircraft gradually slows down and prepares for landing.
h) LDG: This is the landing phase of flight where the aircraft lands and exits the runway.
i) UNK: This is the abbreviated form for those accidents for which the phase of the flight could not be determined.

Key Questions

We aim to provide answer to the following questions from our analysis of the data:

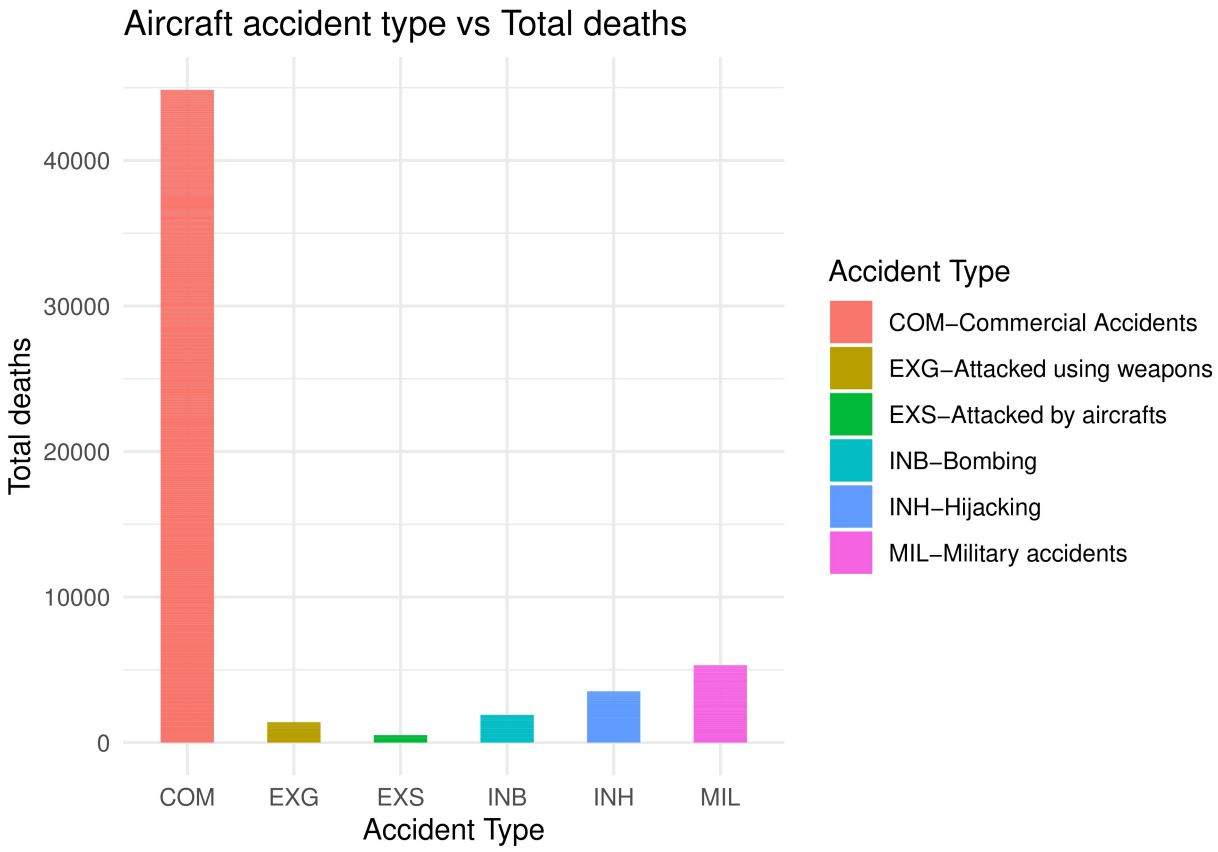
1. How is the number of deaths related to different parameters like type of accidents, phase of accidents and location of accidents?
2. What is the trend of the number of deaths with time?
3. Which airline company is the safest to travel with?
4. Which manufacturer have made the most improvement in terms of safety over the period of time?

Insightful Visualizations

Studying and analyzing data is tough without plotting since data visualization help us to notice every difference with respect to every variable and also helps us to study trends, quantifying relationships, or displaying proportions.Plots which are self explanatory are the **best plots**.

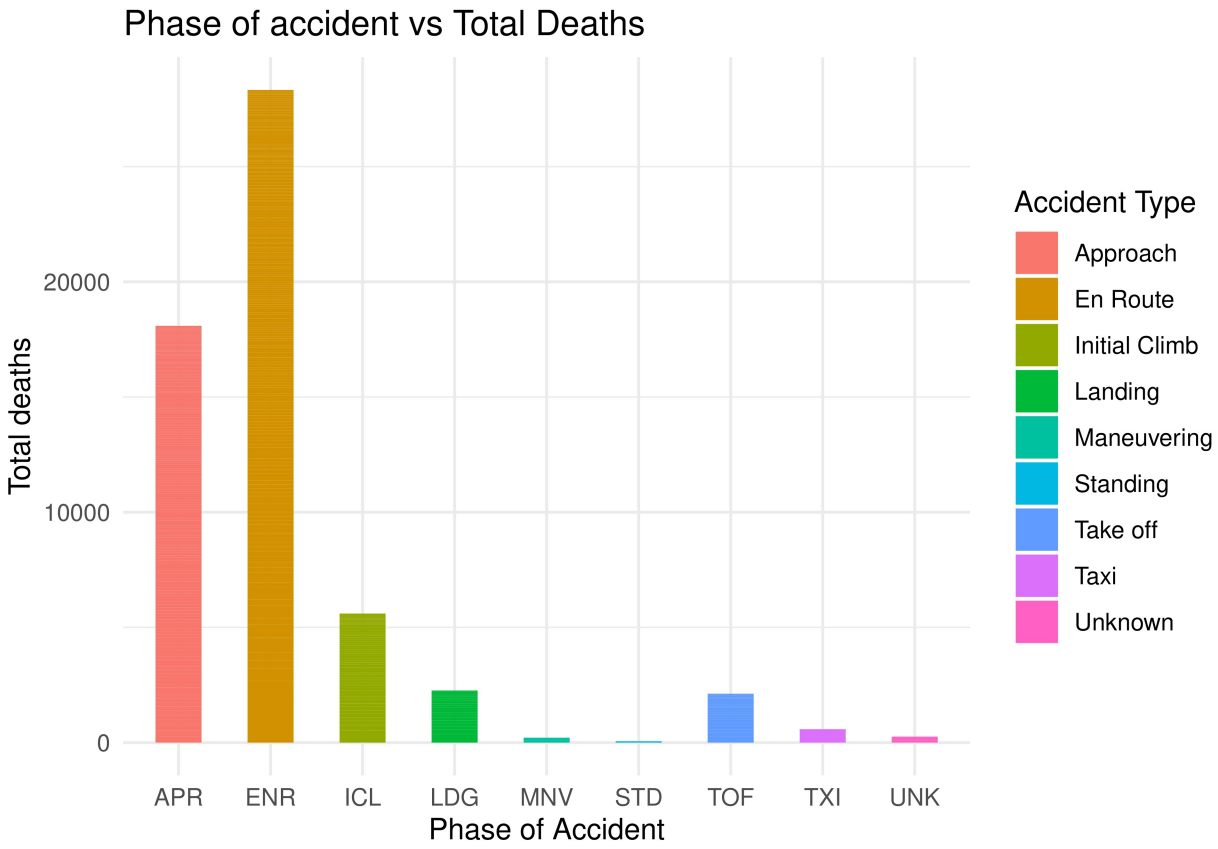
We started analyzing the relationship of the number of deaths with different factors by plotting it against them.

The **first plot** gives us an insight of the relationship between the number of deaths with the different types of aircraft accidents.



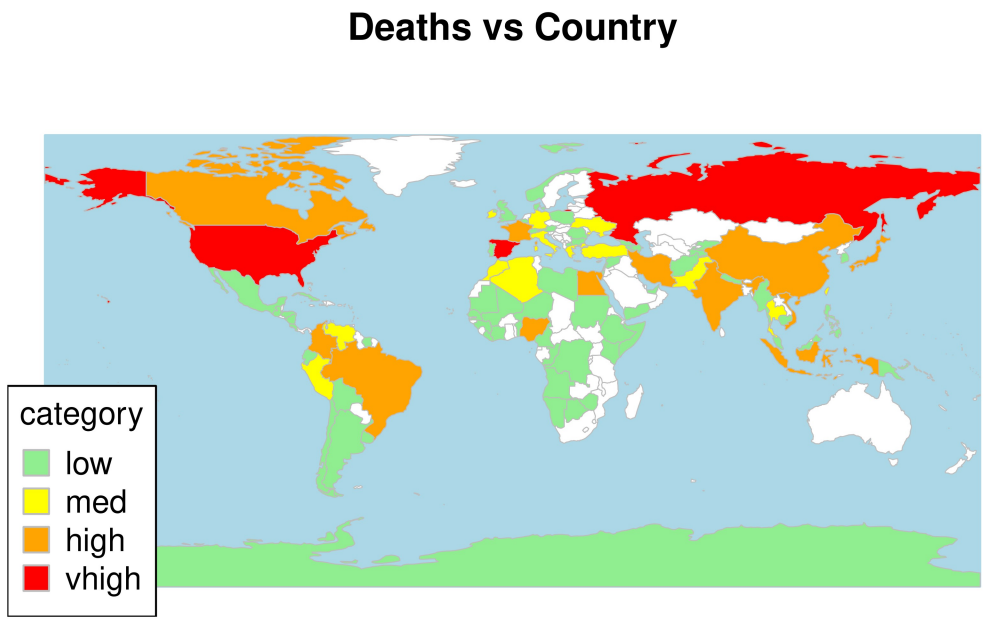
From the graph it is clear that commercial type of accidents contribute to the most number of deaths. One probable reason for this might be the fact that commercial flights are the most flown. Also it carries many passengers at a time, so an accident in a commercial flight is more likely to result in more number of deaths.We can see **Commercial** accidents caused nearly 45000 deaths, followed by **Military** accidents which caused about 5000 deaths, then by **Hijacking**, **Bombimg** and others. Hijacking and bombing kind of incidents are rare, hence contributing to a lower number of deaths.

The **second plot** gives us an insight on how the total number of deaths varies with the different phases of accidents.



From the graph it can be seen that accidents during the en-route phase have resulted in the maximum number of deaths followed by other phases like approach, initial climb and others. During the en-route phase, the flight is in mid air. Hence an accident during this point of no return is likely to be more fatal than accidents in other phases. Also we see that the number of deaths for the standing phase is the least which is obvious since during the standing phase the aircraft is not in motion, hence the chance of an accident is very low during this phase.

To study how **aircraft accidents** have caused deaths across different **locations** of the world, we have plotted a map for the total number of deaths corresponding to the different countries in our data.

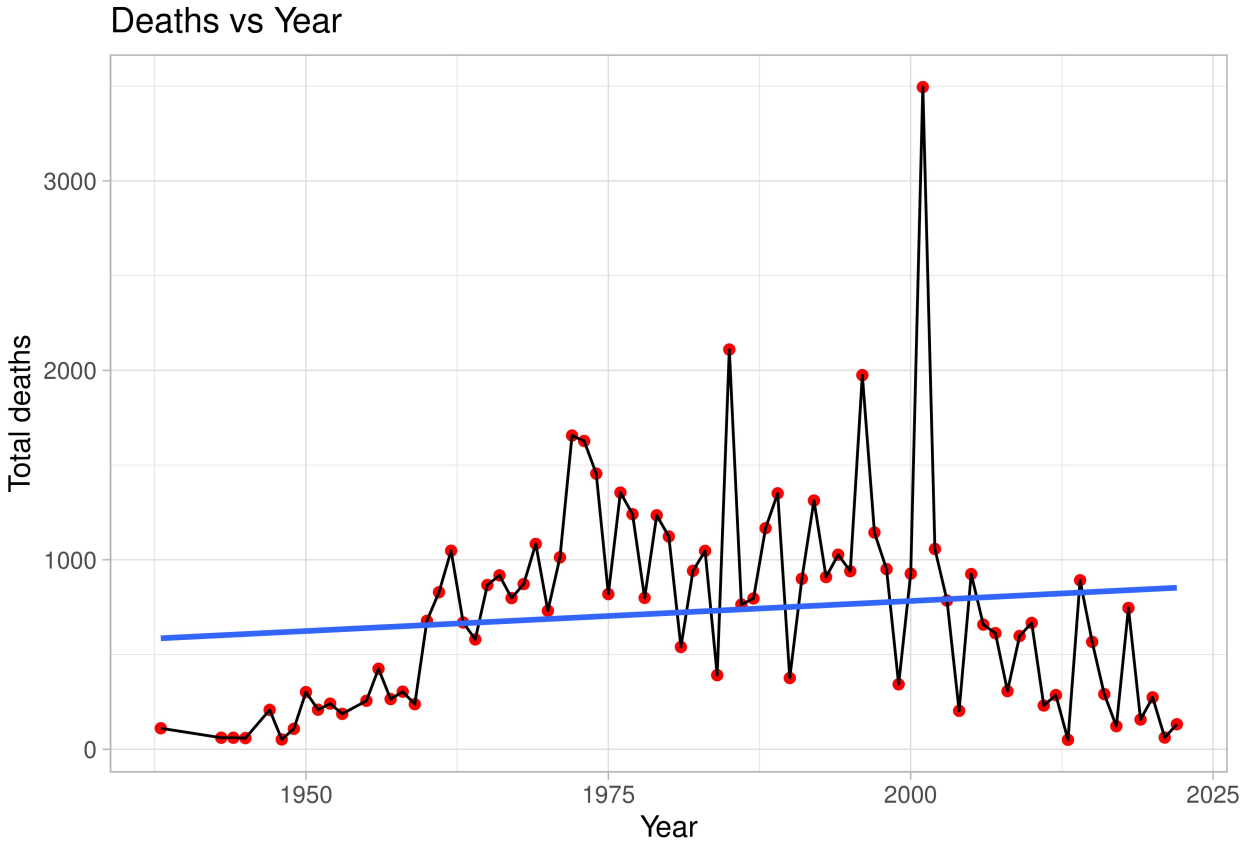


Here the red color denotes very high number of deaths, orange denotes high number of deaths, yellow denotes medium number of deaths and lastly green denotes low number of deaths. From the graph, we can see that countries like USA, Russia, India and parts of Canada, the number of casualties due to aircraft accidents is very high. On the contrary, in some countries of Africa and parts of South America, the number of deaths is on the lower side.

This could be explained by the fact that countries like USA, Russia and India are richer than those of Africa and parts of South America and thus have a bigger aircraft network. Flights in those countries are more frequent than the poorer countries and thus the number of deaths is distributed in such a way.

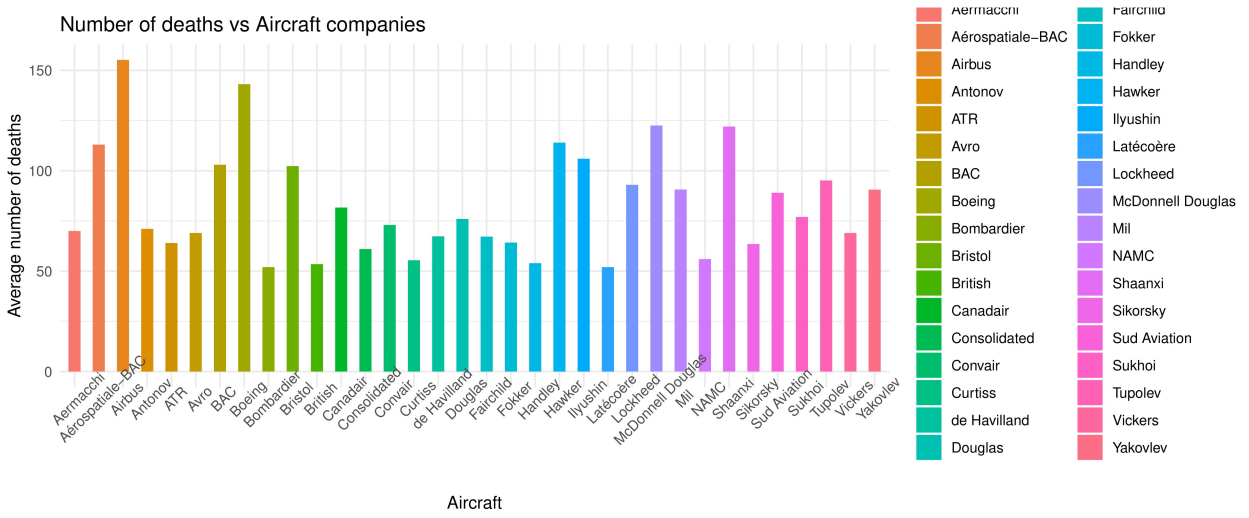
Unfortunately we could not obtain any data on the air traffic in different countries of the world to support our hypothesis. Hence we can guess at the best the reason for such a distribution of deaths in our plot.

Now to see the **trend of deaths over time** we have made the following scatter plot. The graph shows number of deaths every year from 1938-2022.



It can be observed that there is an **overall increase** in the number of deaths over time. But if we divide the time into smaller periods, a **gradual increase** can be observed from **1938-1970**. Then, there is an **extreme observation** around the year **2000**. After that, a **decreasing trend** in the number of deaths can be observed. This might be due to advancement in technology which has made airplanes safer, thus reducing the number of deaths.

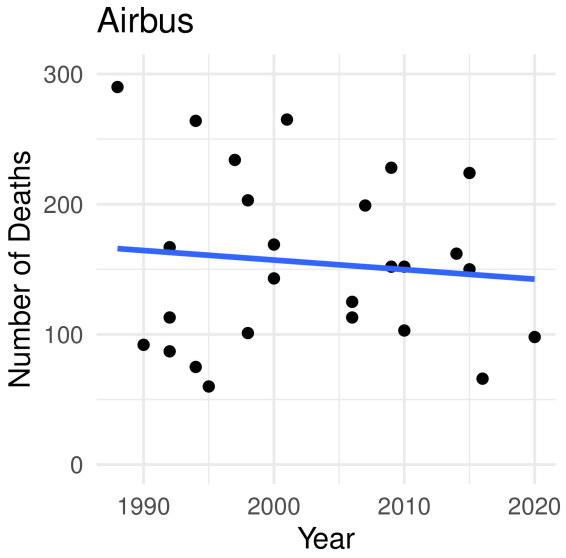
To observe which company was the safest to travel over the years, we made the following bar plot. Here the height of the bars represent the **average number of deaths per flight** for a particular company.

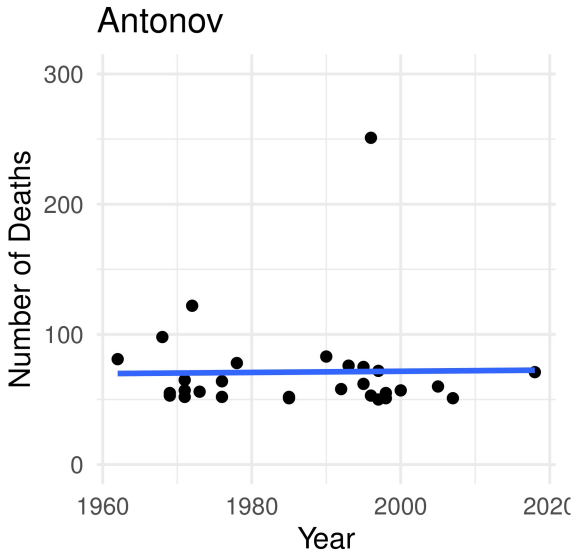


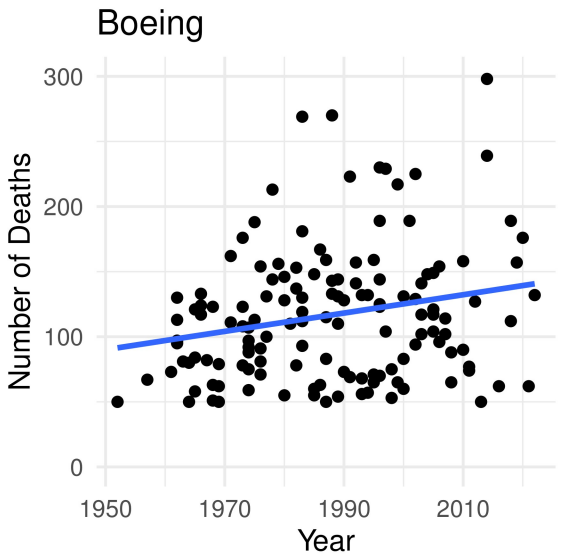
From the graph, we can easily observe that **Airbus** has the highest average deaths per flight, immediately followed by **Boeing**. **Latécoère** and **Bombardier** lie on the other end of the spectrum with the least number of average deaths per flight.

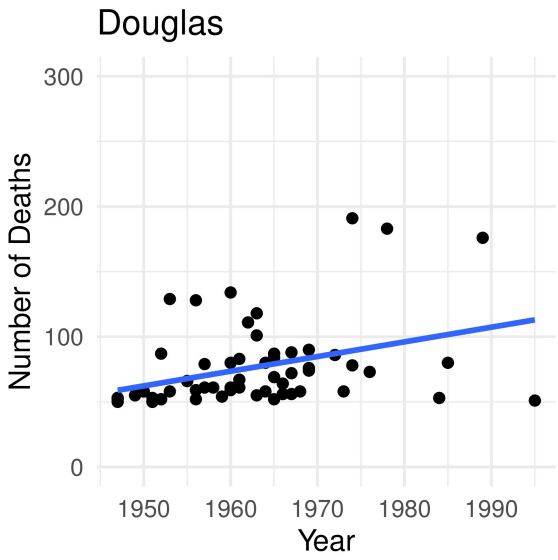
We can also make the plot of the **total number of deaths caused for each company**, but that will give us **wrong analysis**. The reason is simple, maybe a company has very large number of flights compared to others, hence resulting in more number of accidents. Thus studying the total number of deaths against various manufacturers would give us a biased result and would spread a false information.

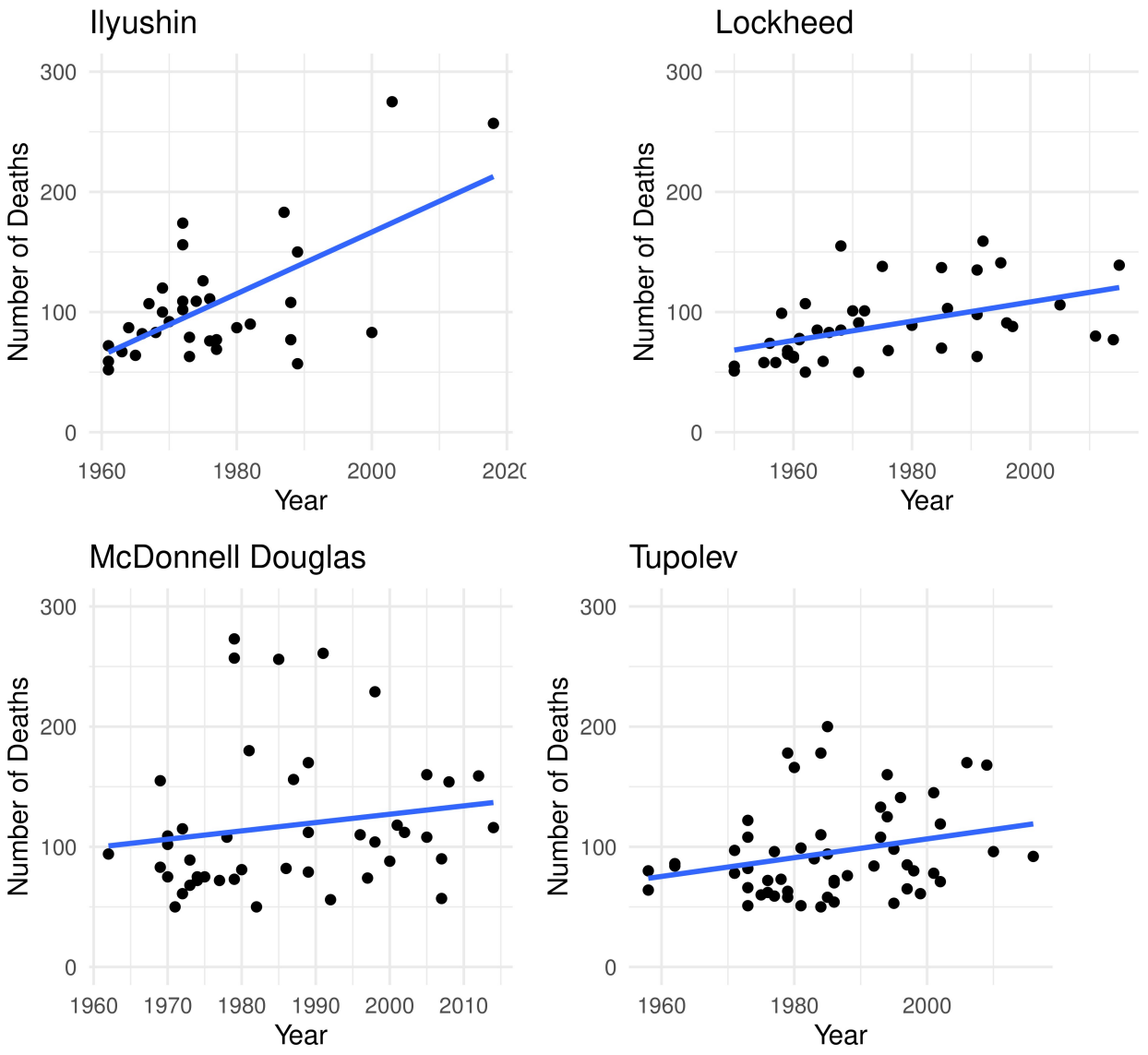
To answer our **final question**, we made scatter plots of total deaths over the years for each company. Here are few plots for some of the major companies such as Airbus, Boeing, Tupolev etc.











From the above graphs following oberservations can be made:

1. Here we can see that the number of deaths for Airbus flights have decreased over the period of time.
2. The number of deaths for Antonov has remained more or less unchanged with time.
3. **Boeing** and **Douglas** have recorded increasing number of deaths over the time period and so did the other companies.

So we can say that maybe **Airbus** has made more improvements in their air crafts as compared to others over time.

Conclusions

By analyzing our data set we have discovered some compelling patterns for airplane accidents. The results can be summarized in the following points:

1. Most of the deaths have occurred for **Commercial** aircrafts followed by **Military** aircrafts.
2. Accidents during the **En route** phase have resulted in the most number of deaths.
3. The number of deaths due to aircraft accidents have been on a **decrease** during the last few years.
4. **Airbus** and **Boeing** have the highest number of average deaths during an accident.
5. However, **Airbus** has made some quality improvements in their planes overtime.

The main objective of this project was to raise awareness of flight safety and understand the progress of the aviation industry. The aircraft industry has come a long way but it still has a lot to improve.

References

1. The data set was obtained from Wikipedia [List of deadliest aircraft accidents and incidents] https://en.wikipedia.org/wiki/List_of_deadliest_aircraft_accidents_and_incidents.
2. The reference for plots was taken from r-statistics [Top 50 ggplot2 Visualizations - The Master List] http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html.