

# A Century of Aviation Accidents: Trends, Risk Factors, and Insights

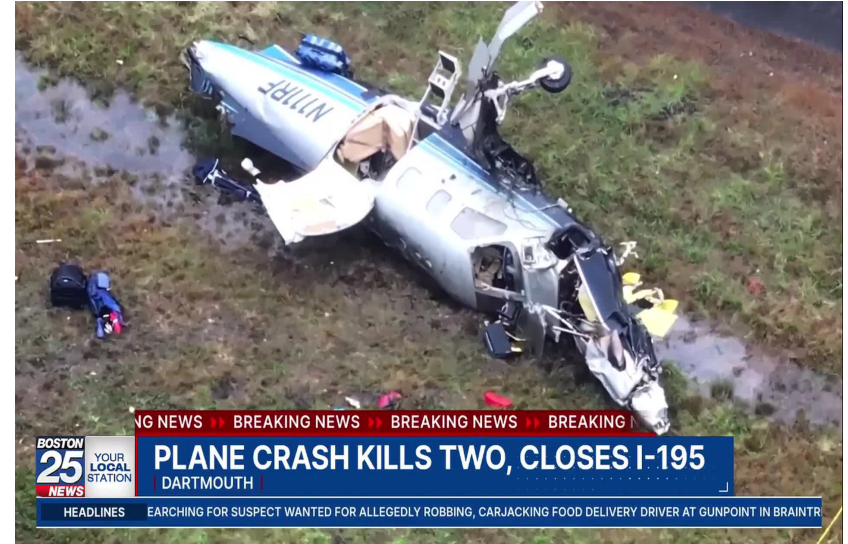
ECE 143 Final Project Presentation

Brandon (A17002478)  
Chengguo (A69041855)  
Ho-Yen (A69044575)  
Shweta (A69032860)  
Puneet (A69030311)



# Motivation

- Even rare aviation accidents carry enormous human, economic, and regulatory consequences.
- Uncover long-term risk patterns that are impossible to see in isolated reports.
- Goal:
  - Identify which aircraft types, flight phases, and time periods exhibit elevated crash severity
  - Explore what these patterns reveal about changes in aviation safety over time.



# Dataset

- **4,962** accident records spanning **1920–2024**
- Includes aircraft type, operator, location, fatalities, time, and summary text
- Contains both structured fields (year, operator, weather keywords) and unstructured text (summaries)
- Requires cleaning and parsing (dates, times, location fields, fatality extraction)



Data columns (total 25 columns):				
#	Column	Non-Null Count		Dtype
0	aboard	4952	non-null	object
1	aircraft_type	4951	non-null	object
2	cn_ln	4951	non-null	object
3	date	4951	non-null	object
4	detail_url	4961	non-null	object
5	fatalities	4952	non-null	object
6	flight_no	4951	non-null	object
7	ground_fatalities	4907	non-null	float64
8	location	4951	non-null	object
9	operator	4951	non-null	object
10	raw_text	10	non-null	object
11	registration	4951	non-null	object
12	route	4950	non-null	object
13	summary	4951	non-null	object
14	time	4951	non-null	object
15	year_page_url	4960	non-null	object
16	date_parsed	4951	non-null	object
17	time_raw	4951	non-null	object
18	time_hhmm	3498	non-null	object
19	fatalities_total	4943	non-null	float64
20	fatalities_passengers	4737	non-null	float64
21	fatalities_crew	4737	non-null	float64
22	location_city	4928	non-null	object
23	location_state	3331	non-null	object
24	location_country	2777	non-null	object

# Data Processing

## Scraping

- Python script pulled accident records from PlaneCrashInfo.com
- Parsed yearly pages + individual accident pages
- Extracted key fields (date, aircraft type, fatalities, summary)
- Combined everything into a single CSV

## Cleaning

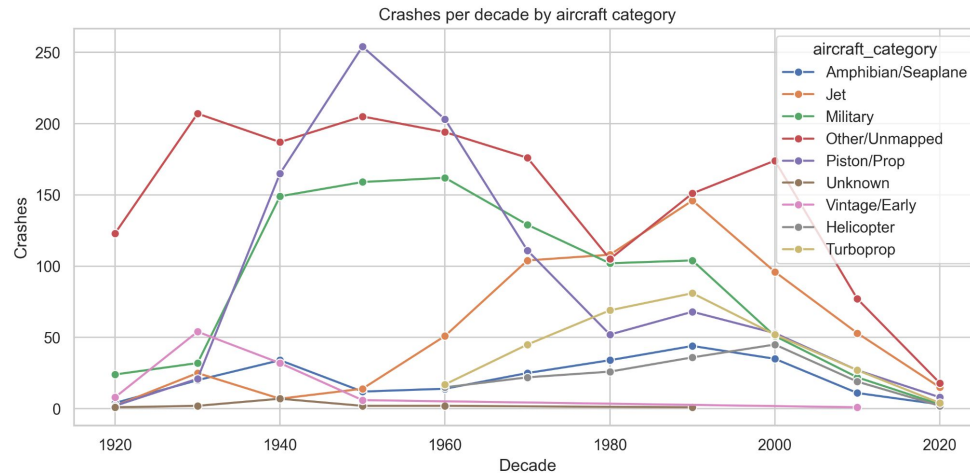
- Standardized aircraft types and flight phases
- Fixed inconsistent dates and missing numeric fields
- Removed unusable or duplicate entries

## Feature Engineering

- Added decade, hour, and fatality ratio
- Categorized aircraft (Jet, Prop, Turboprop, Helicopter, Military)

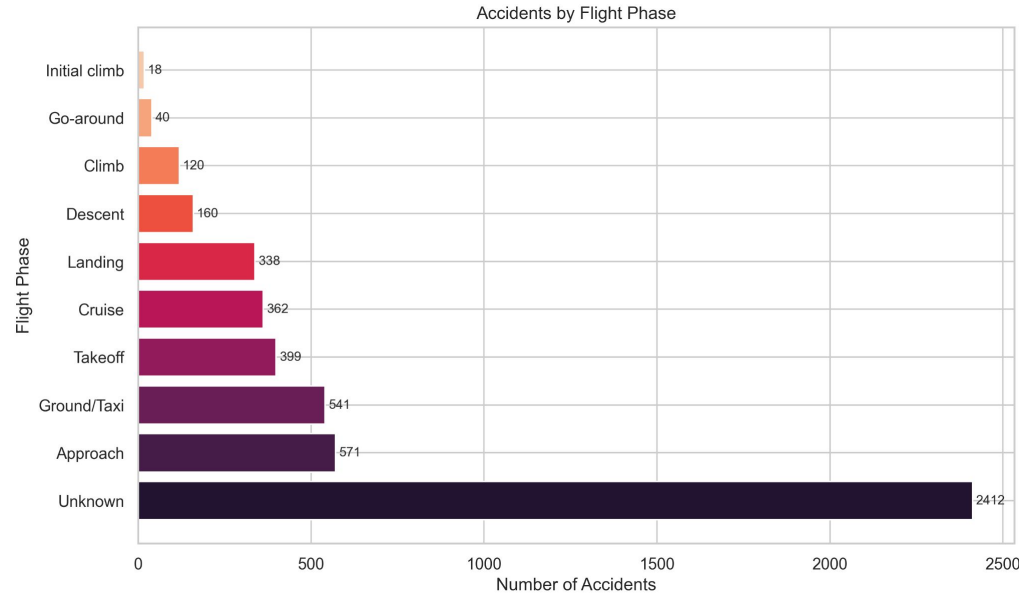
# Aircraft Category

- **Early decades:** Piston/prop dominate due to immature tech & minimal regulation.
- **Jet era (1950s+):** Crash counts rise simply because global jet traffic expands (*exposure effect*).
- **Mid-century peak:** Military & piston aircraft spike during WWII / Cold War activity.
- **Recent decades:** Helicopter & turboprop accidents rise as their operational use grows.



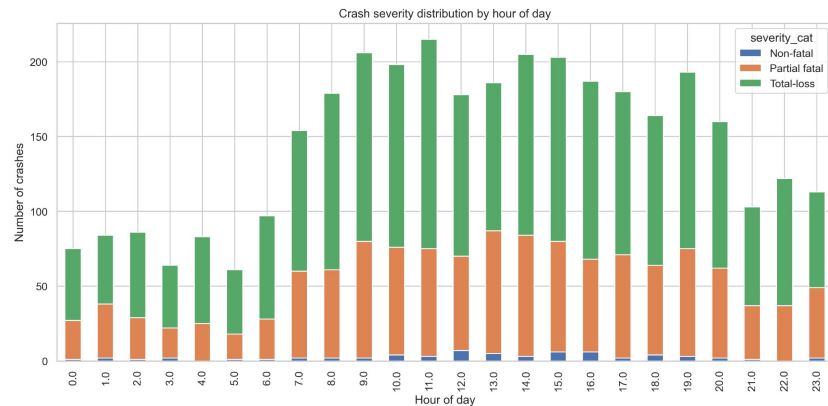
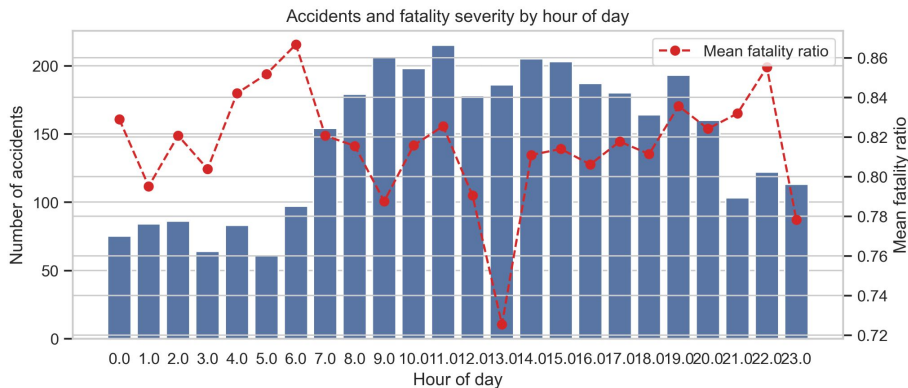
# Flight Phase

- Most crashes occur during takeoff, approach, and landing → highest workload + least margin for error.
- Cruise accidents are fewer but often more catastrophic due to altitude and speed.
- Large “Unknown” category comes from older reports lacking detailed phase information.



# Hours-of-Day

- Crash frequency strongly follows global flight schedules
- Severity is not highest when frequency is highest
- Nighttime accidents are deadlier because “total-loss” crashes are more common

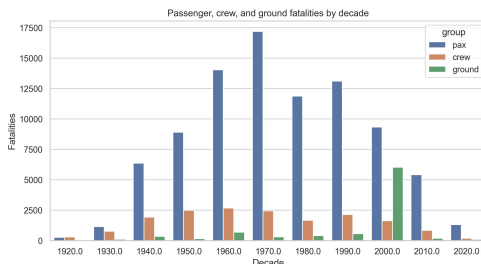
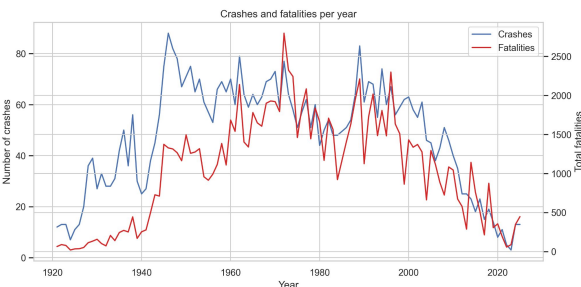
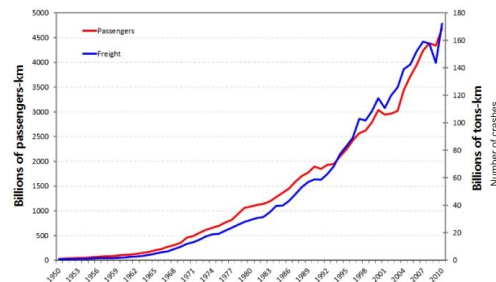


# Safety Over Time

Despite exponential growth in flight volume from the early 1900s

- Total crashes increased from 1910s and peaked in 1970s
- Decreased since 1970s until modern day
- **Modern aviation significantly safer**

Interesting datapoint in 2000s that breaks this trend: isolated 9/11 incident



## ACCIDENT DETAILS

Date: September 11, 2001

Time: 0847

Location: New York City, New York

Operator: American Airlines

Flight #: 11

Route: Boston - Los Angeles

AC Type: Boeing 767-223ER

Registration: N334AA

cn / ln: 22332/169

Aboard: 92 (passengers:81 crew:11)

Fatalities: 92 (passengers:81 crew:11)

Ground: 2750

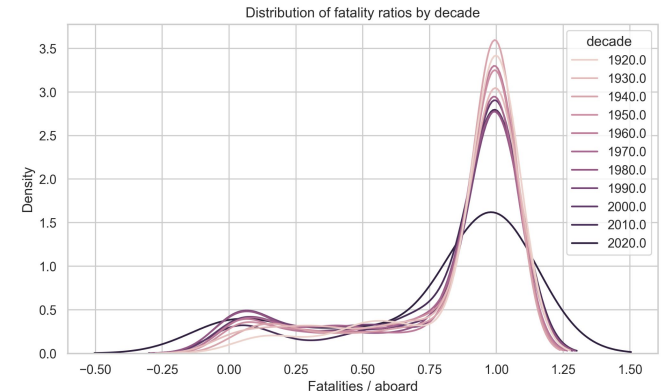
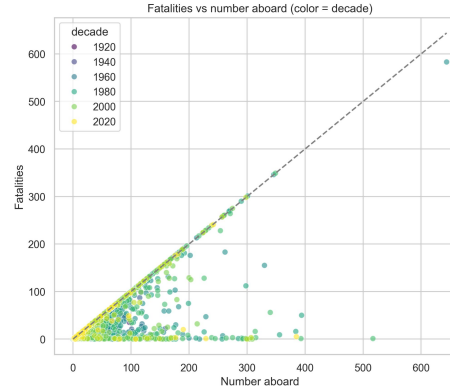
Summary: The aircraft was hijacked shortly after it left Logan International Airport in Boston. The hijackers took control of the aircraft and deliberately crashed it into the north tower of the World Trade Center between the 94th and 99th floors at approximately 450 mph. After 102 minutes, the building collapsed. It was one of four planes that were hijacked the same day.

# Safety Over Time (continued)

Fatality ratio is measured as total fatalities divided by total passengers aboard.

Observations include

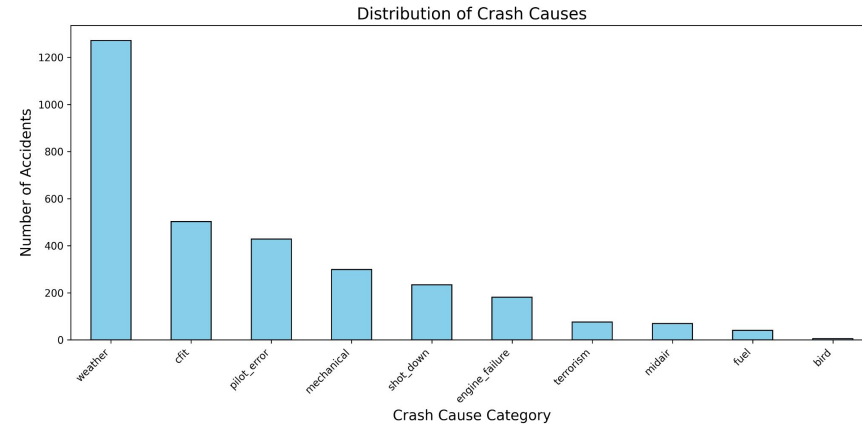
- Latest data points (yellow dots and dark line) show how in recent years, airplanes have become much safer
- This is likely a downstream effect of tighter regulations, increased safety standards, innovation in aircraft technology, and commercialization of passenger aircraft.



# Distribution of Crash Causes

Crash causes are grouped into major categories based on accident summaries. Observations include

- Weather is by far the most common cause (>1200 cases), showing its persistent impact on aviation risk.
- CFIT and pilot error together form a large share of preventable accidents, while mechanical failures decline in later decades as aircraft engineering and maintenance improve.

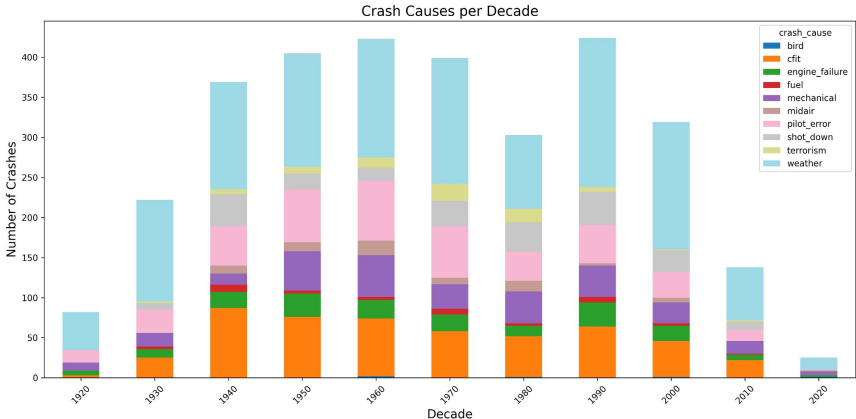


## Crash Causes Over the Decades

Crash causes are aggregated by decade to show how dominant risk factors change over time.

## Observations include:

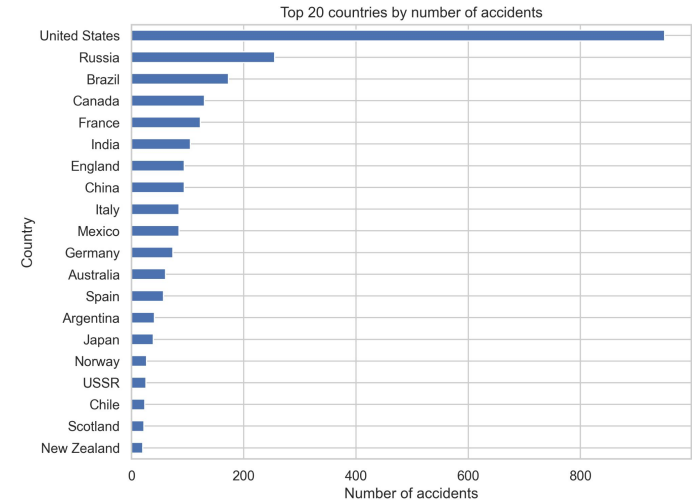
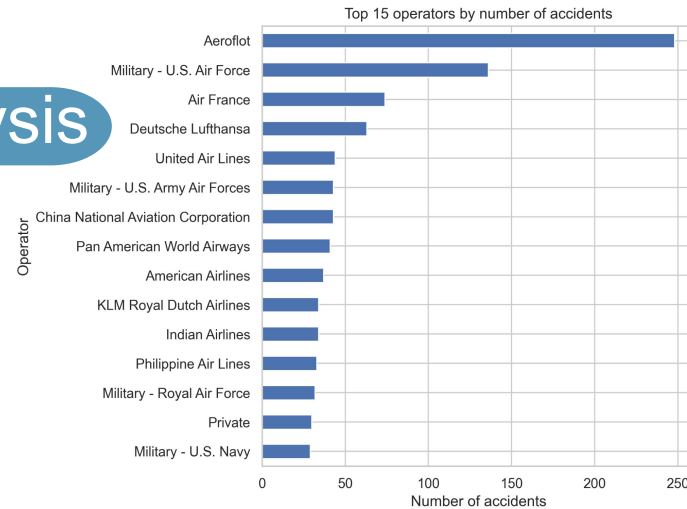
- Weather-related crashes remain the leading cause in every decade, while CFIT and pilot error peak around the mid-20th century and decline with better navigation aids
- Spikes in shot-down and mid-air events around the 1940–1970 period align with WWII/Cold War activity, followed by a post-1990 decline across most causes as regulations, automation, and forecasting improve



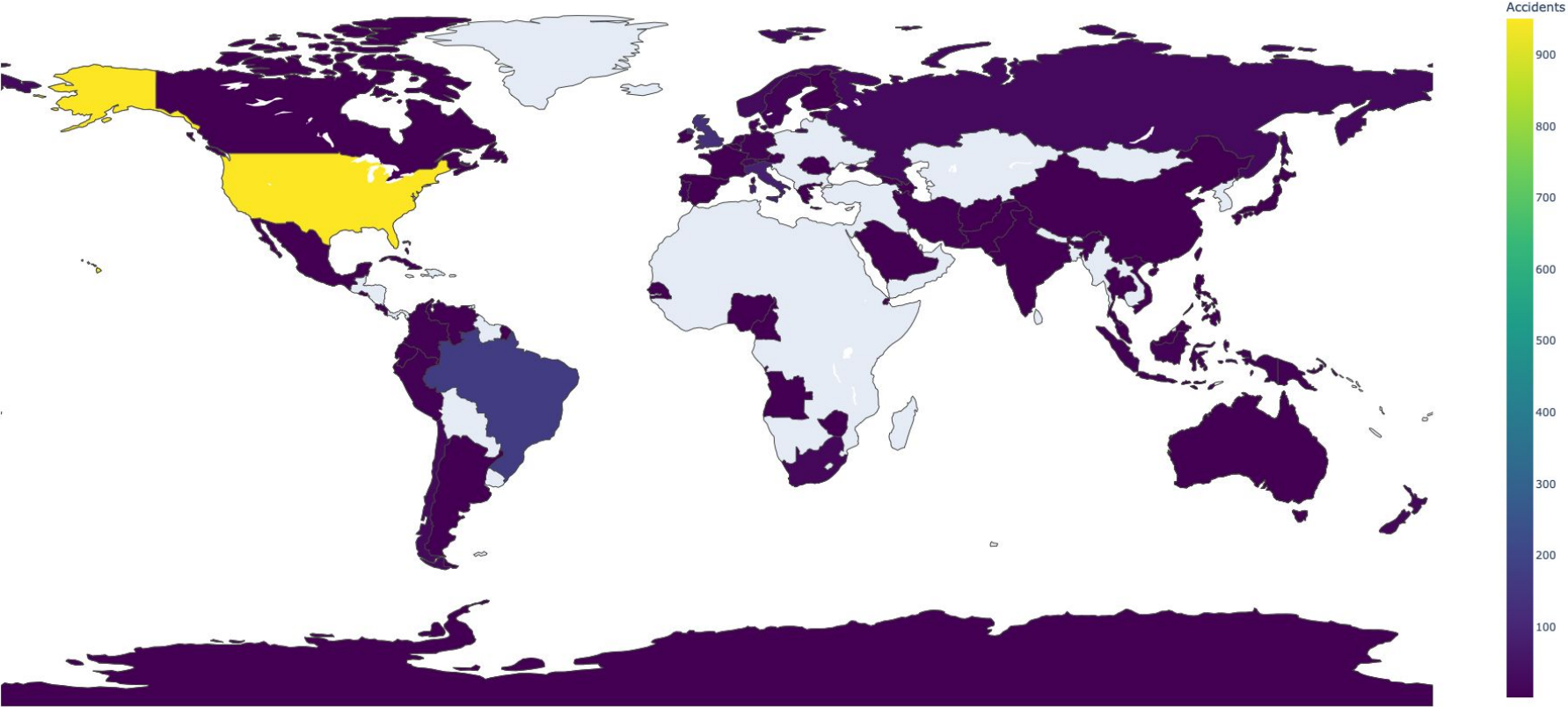
# Airline/Country of Origin Analysis

Data clearly dominated/skewed towards the United States. Likely for several reasons

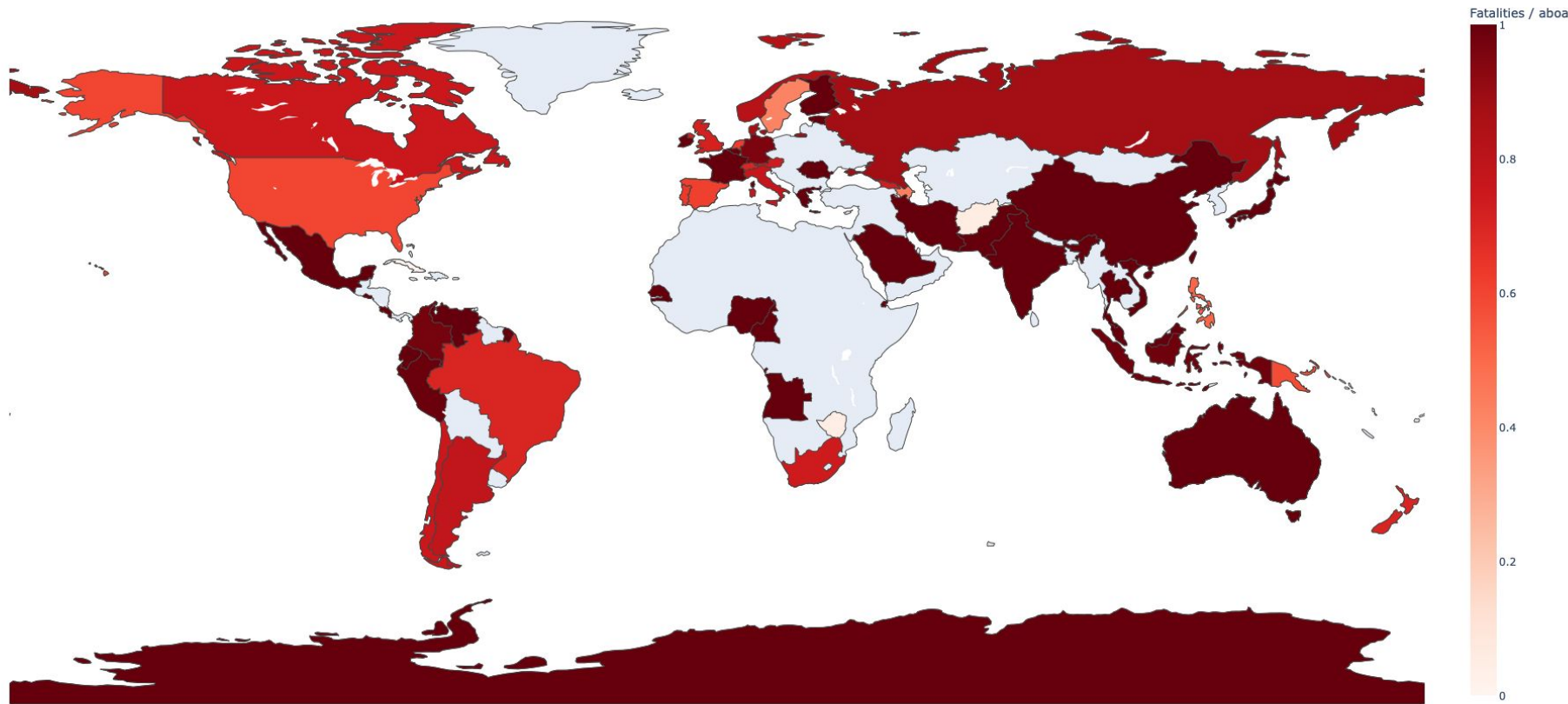
- The U.S. were leaders in aircraft innovation, leading to larger number of casualties as the technology matured. Other countries likely adopted aircraft after maturity.
- Aeroflot (Russian) also experienced many accidents. This is likely an effect of the Cold War where the U.S. and Soviet Union were competing in technology and sacrificing safety standards in the process. This agrees with our previous data where most accidents happened around ~1970s which is around the peak of the Cold War.



Total recorded accidents by country



Country-level fatality ratio ( $\Sigma$  fatalities /  $\Sigma$  aboard)



# Conclusion

- **Safety Trends**

- Accident counts peak mid-century
- Fatality severity steadily declines after the 1980s

- **Risk Contexts**

Certain conditions consistently elevated risk:

- Approach/Landing phases,
- Night operations,
- Military and early piston/prop aircraft,

- **Severity Predictors**

- Aircraft category
- Flight phase
- Visibility-related hours of day

Aviation safety is improving  
But specific risk factors remain consistently dangerous across history



**THANK YOU FOR ATTENTION**