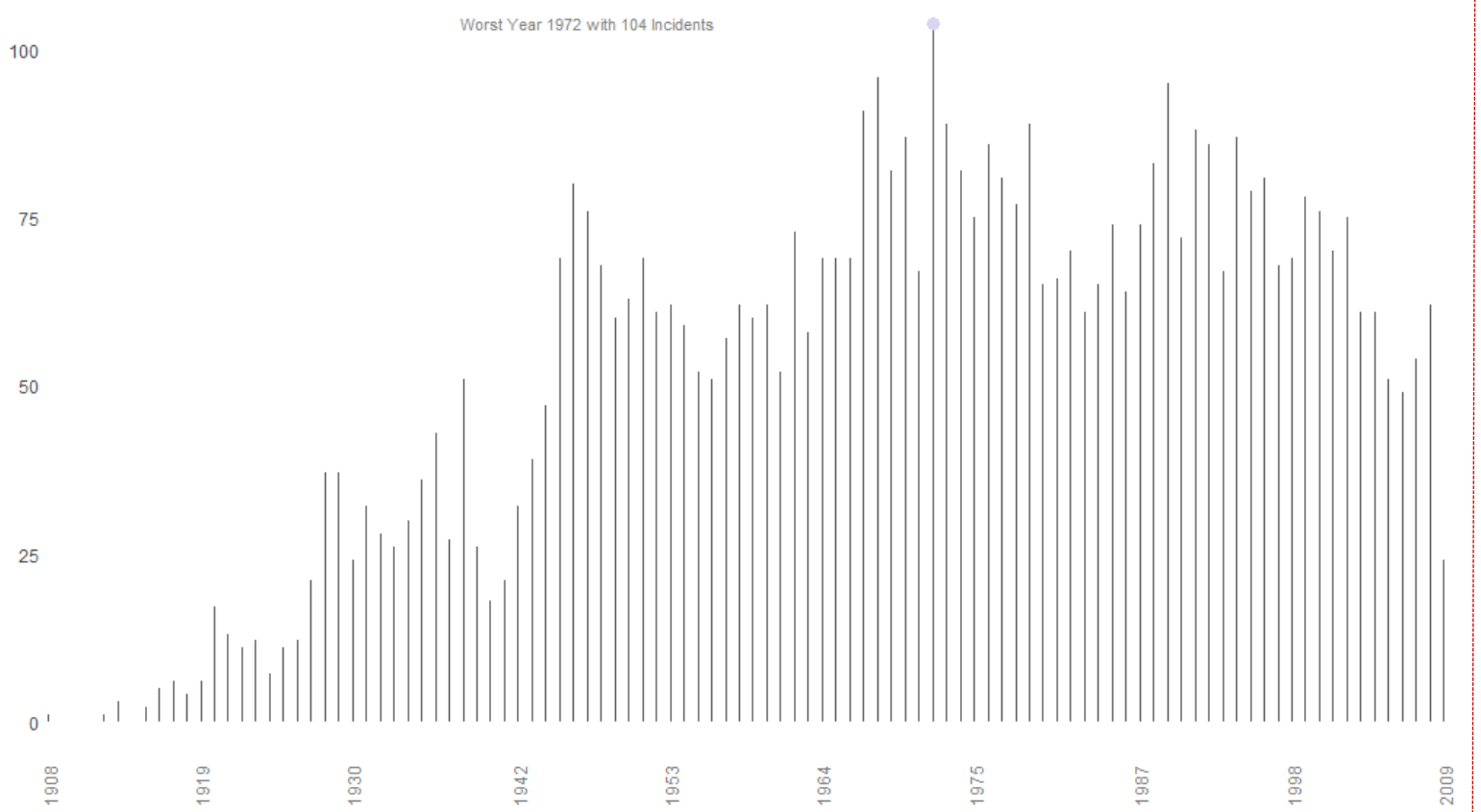
Plane Truth …

(Based on “Airplane Crashes” data set)

Presentation by

Burhan ul Haq

BASIC ANALYSIS



Best Year 1909, 1912 *

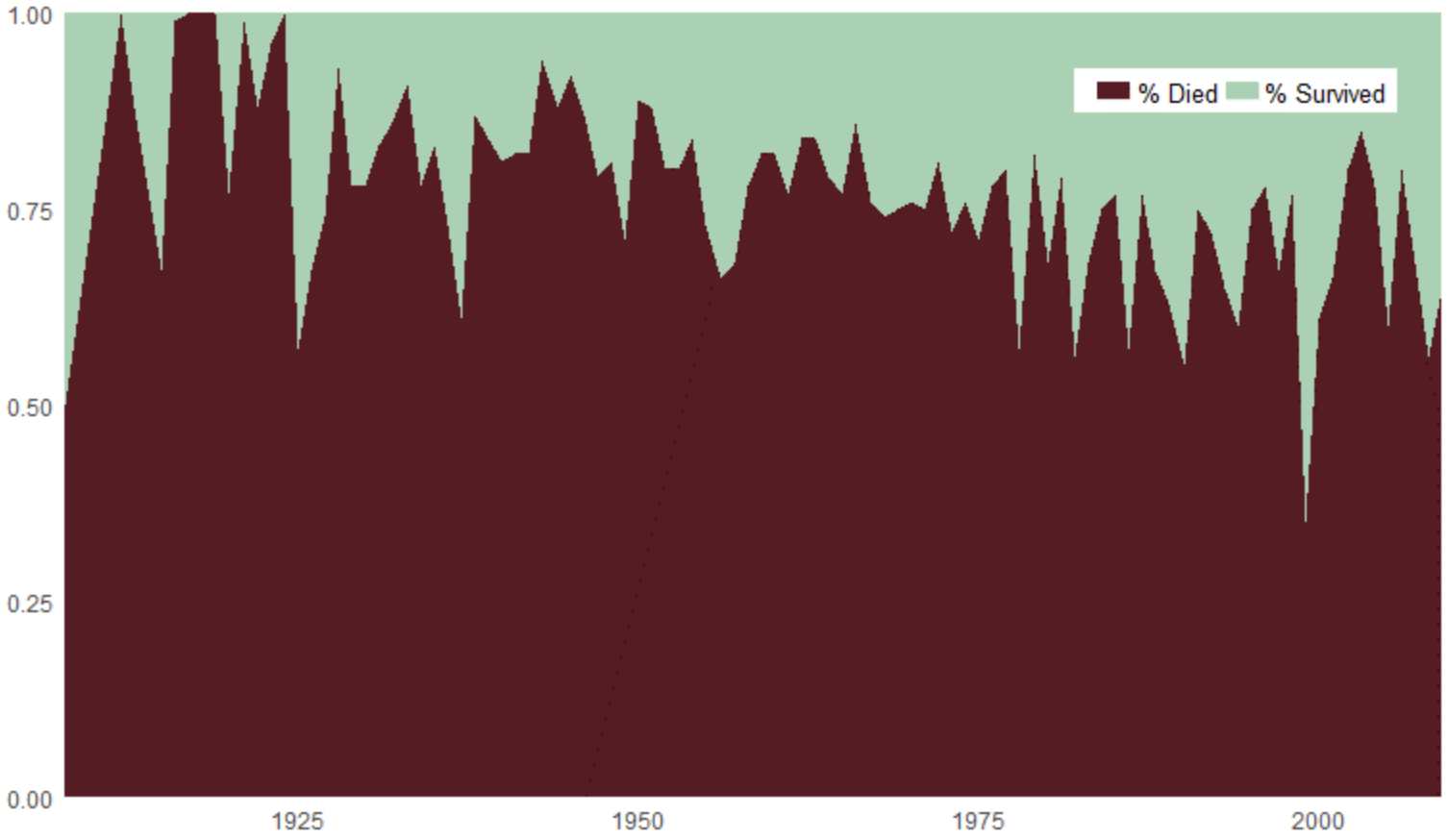
1 Incident

Worst Year 1972

104 Incidents

*: Not really, a better measure will be to use a reference measure such as Incidents / Total Passenger Miles

BASIC ANALYSIS



Best Year 1999

65 % or

1788 Survived / 2758 Aboard

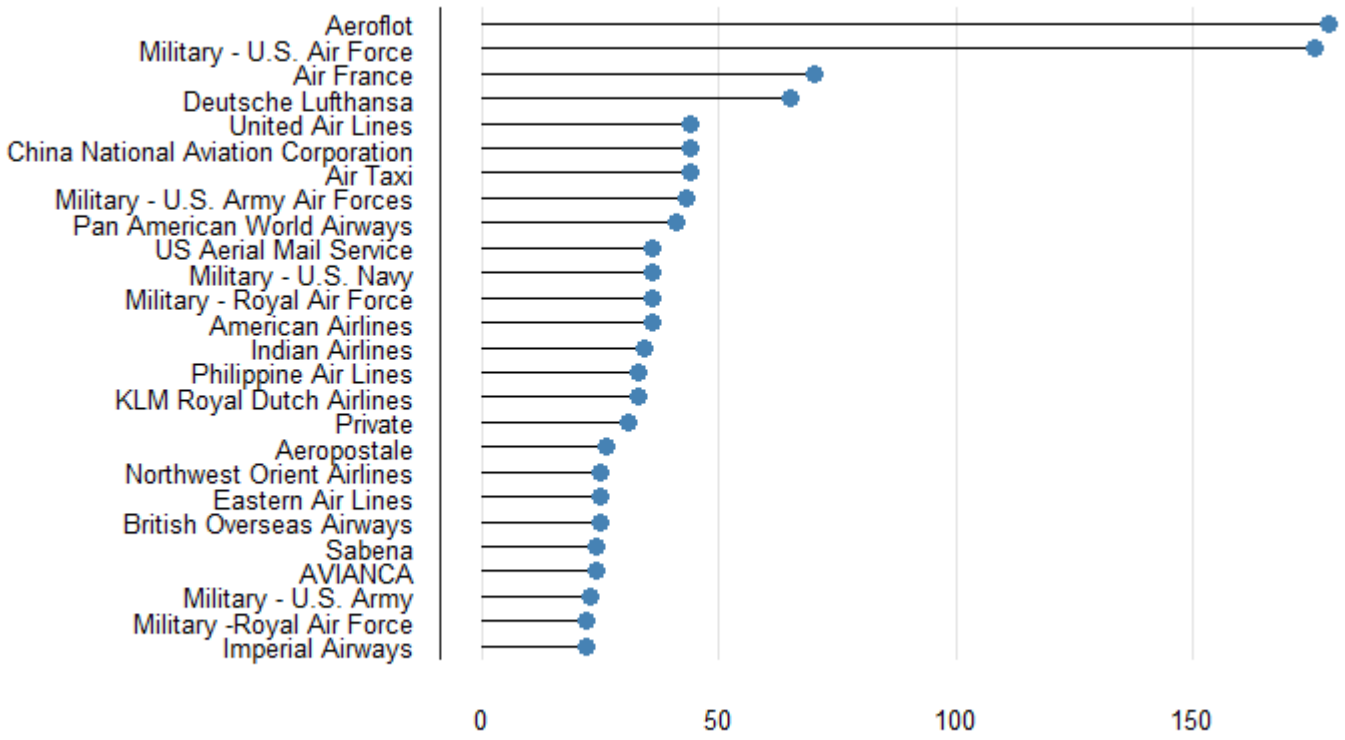
Worst Year 1917*

100% or

124 Died / 124 Aboard

*: 1912,1918, 1919 and 1924 also had 100% rate, but number of deaths were less

BASIC ANALYSIS



Best Operator *

1921 Operators with 1 Incident

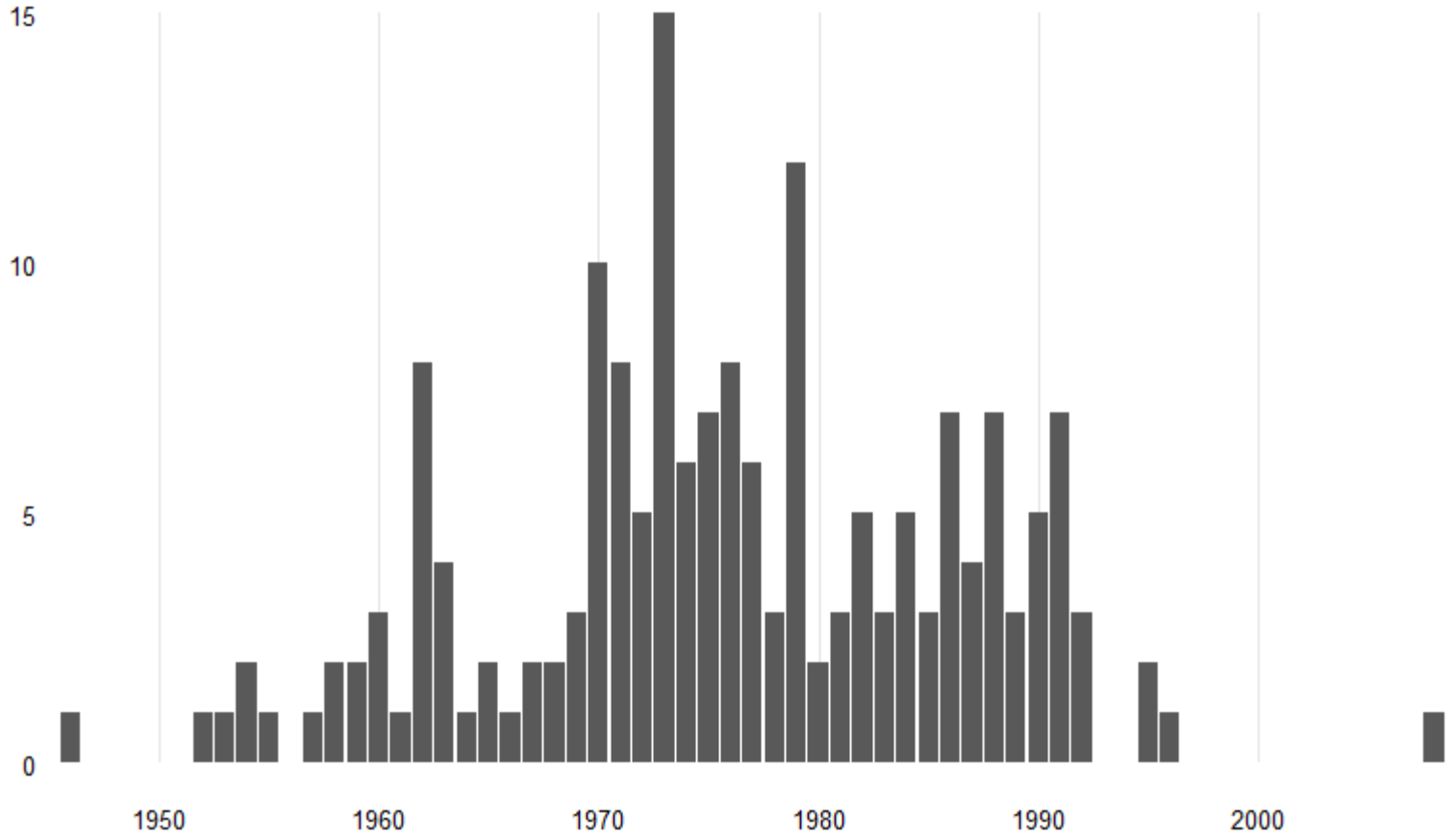
Worst Operator *

Aeroflot 179 Incidents

US Air Force 176 Incidents

*: Not really, a better measure will be to use a reference measure such as Incidents / Total Passenger Miles

BASIC ANALYSIS



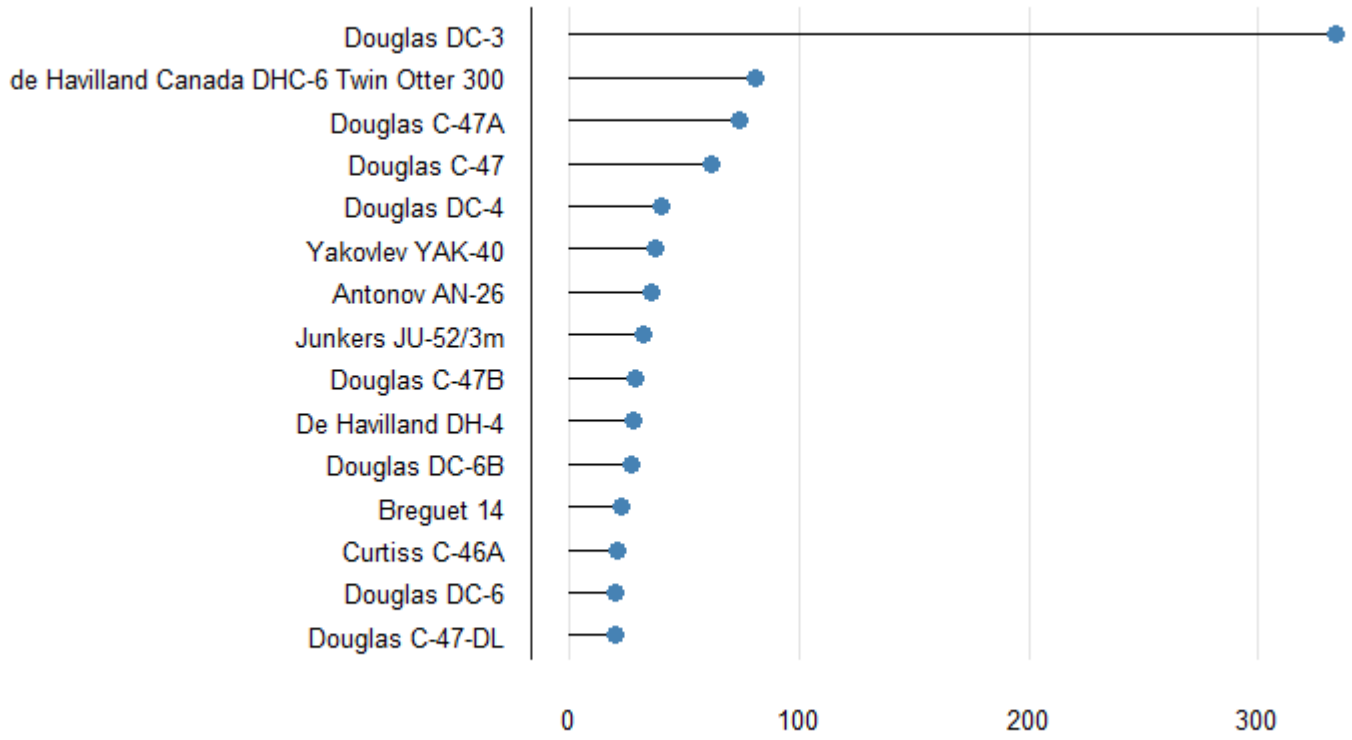
Aeroflot

has either improved

after 1990,

Or reduced operations

BASIC ANALYSIS



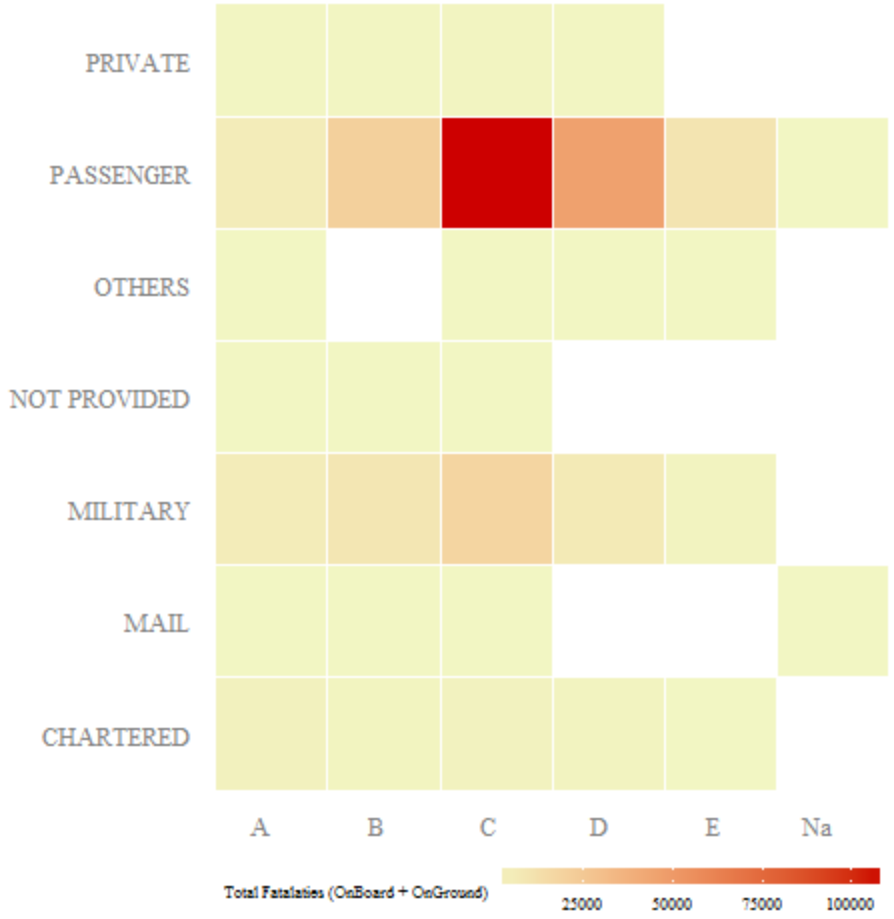
Worst Aircraft

Douglas DC-3 with 334

Incidents,

Havilland has 81 Incidents

INSIGHTS (1)



Aircraft Categories *

vis-à-vis

Operator Categories

Highest Fatalities

Medium Range Transport

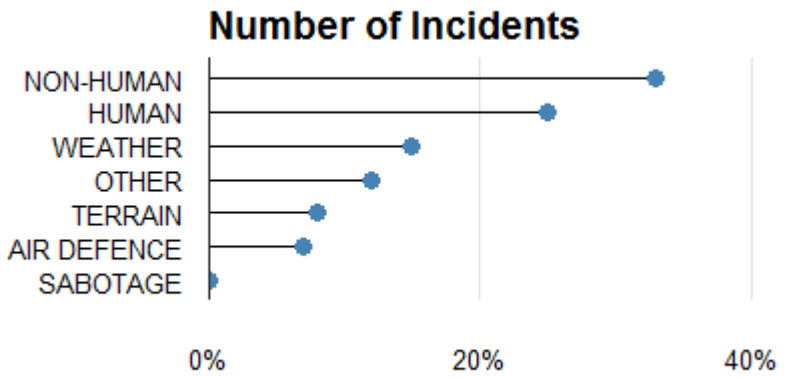
Aircrafts

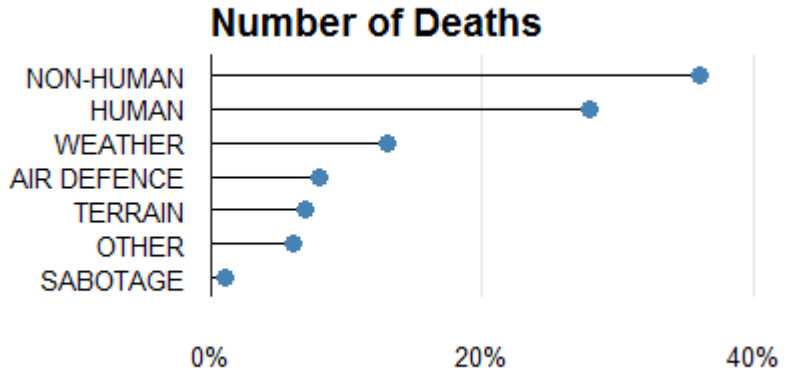
for

Passenger Airlines

*: Aircraft Categories based on wingspan, as per ICAO Classification

INSIGHTS (2)





Biggest Reason

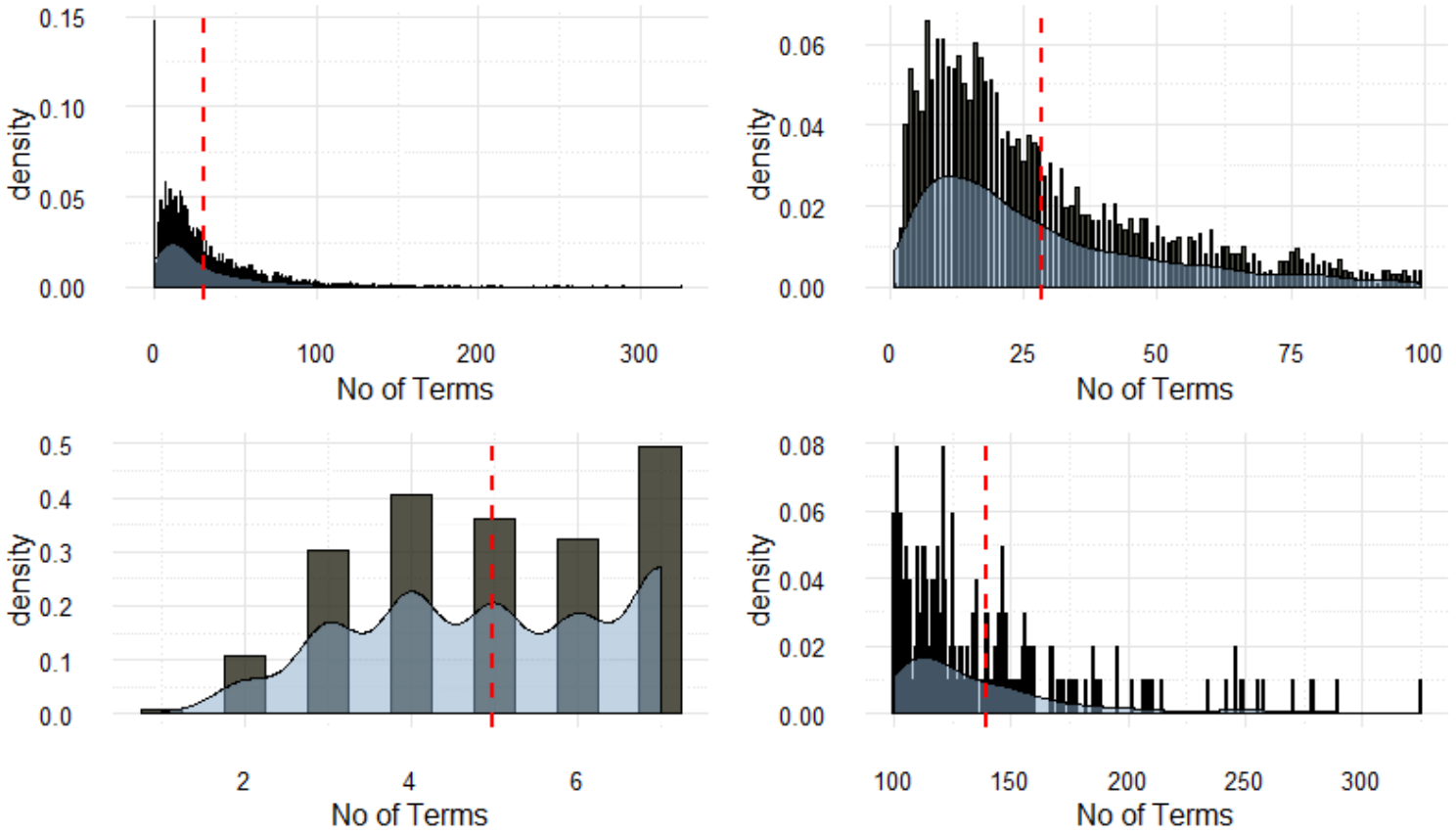
Non Human

Terms containing:

failure | engine* | wing* | fuel* | tail*

Root Cause: Poor maintenance, overuse or mishandling

INSIGHTS (3)



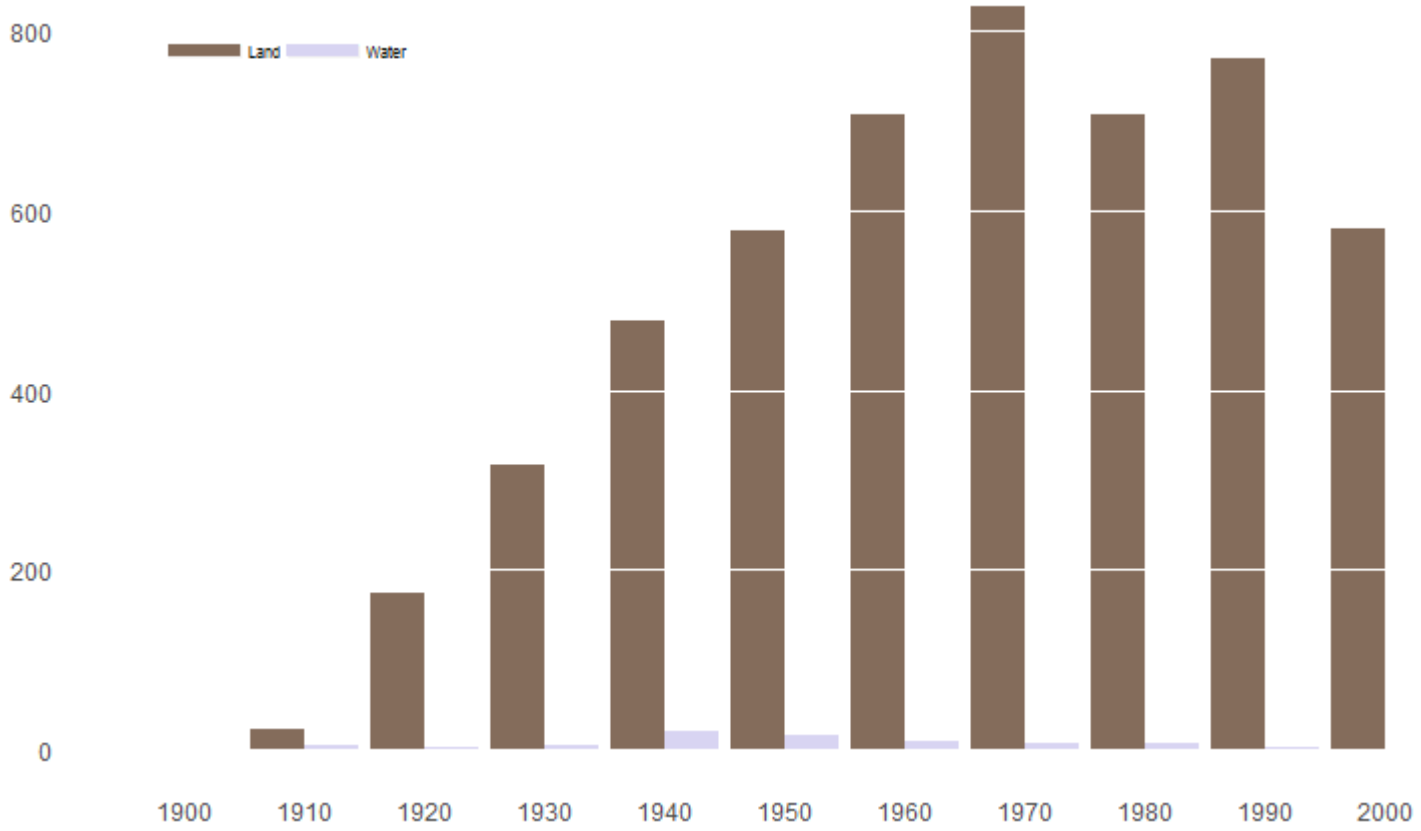
Text Mining Process

Number of Terms

vis-à-vis

Occurrence of Terms

INSIGHTS (4)

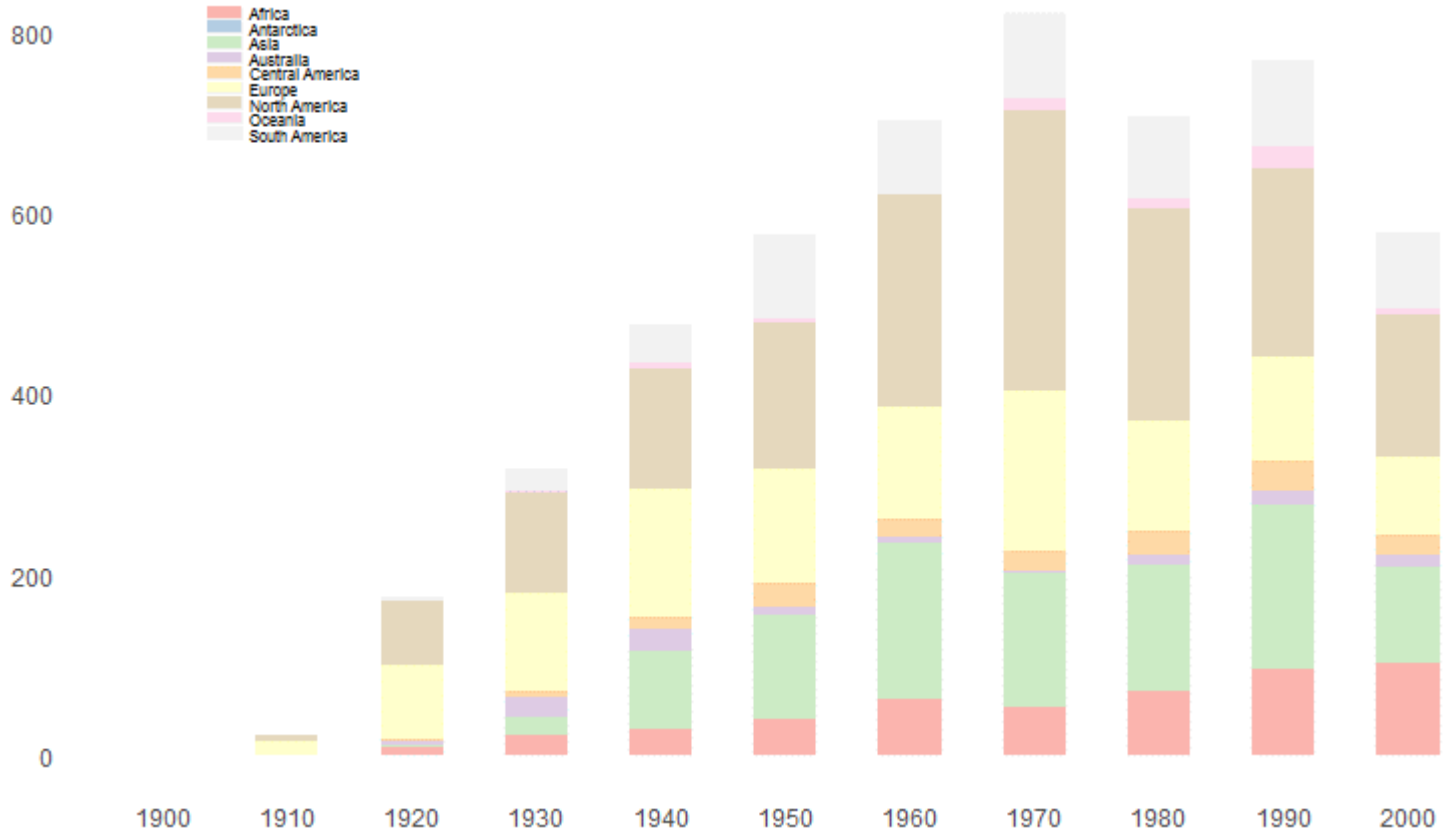


Zooming Out

Earth is ¾ Water

But Most of Crashes happen on Land

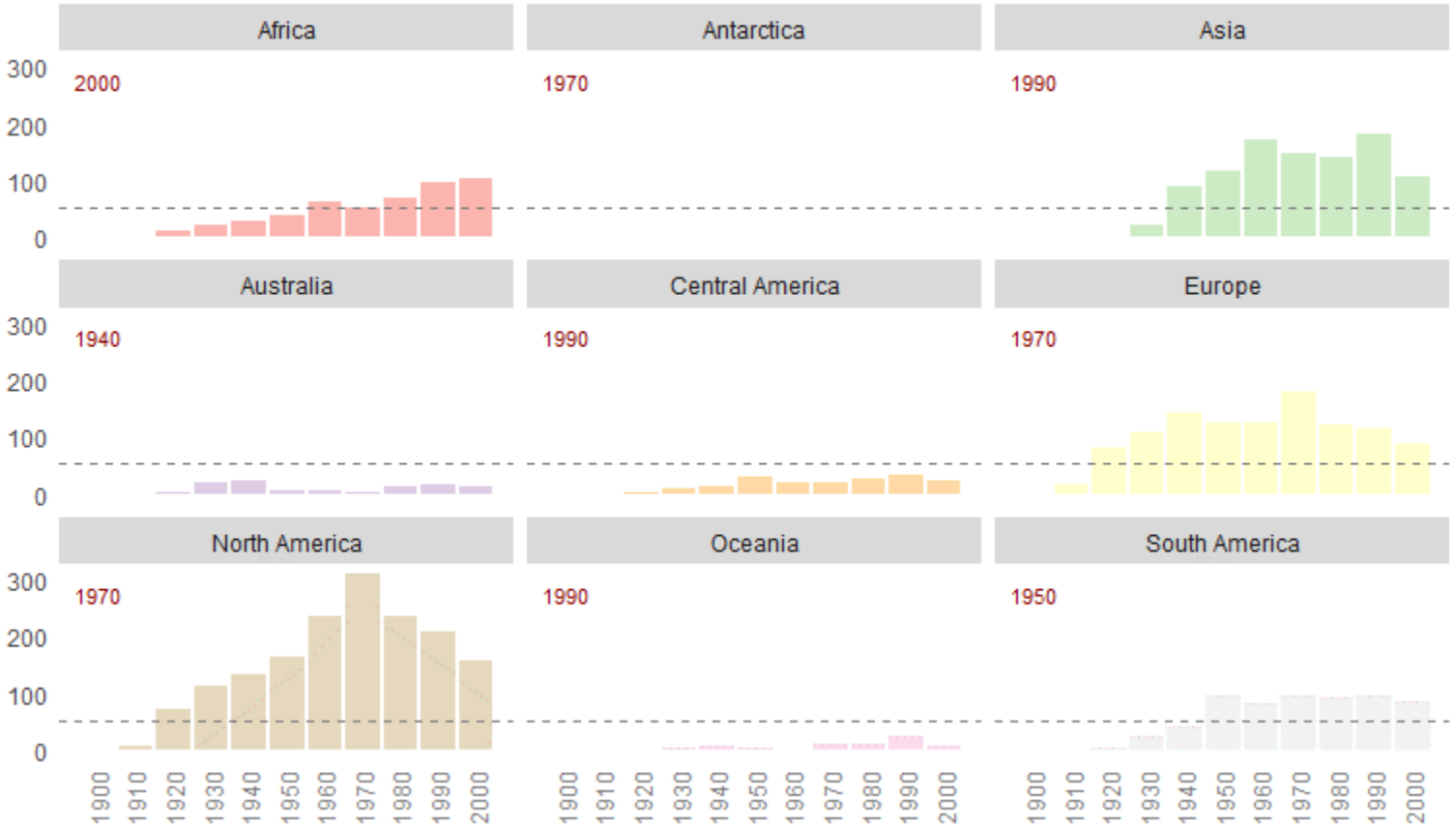
INSIGHTS (5)



Zooming In

Where on Land ?

INSIGHTS (6)

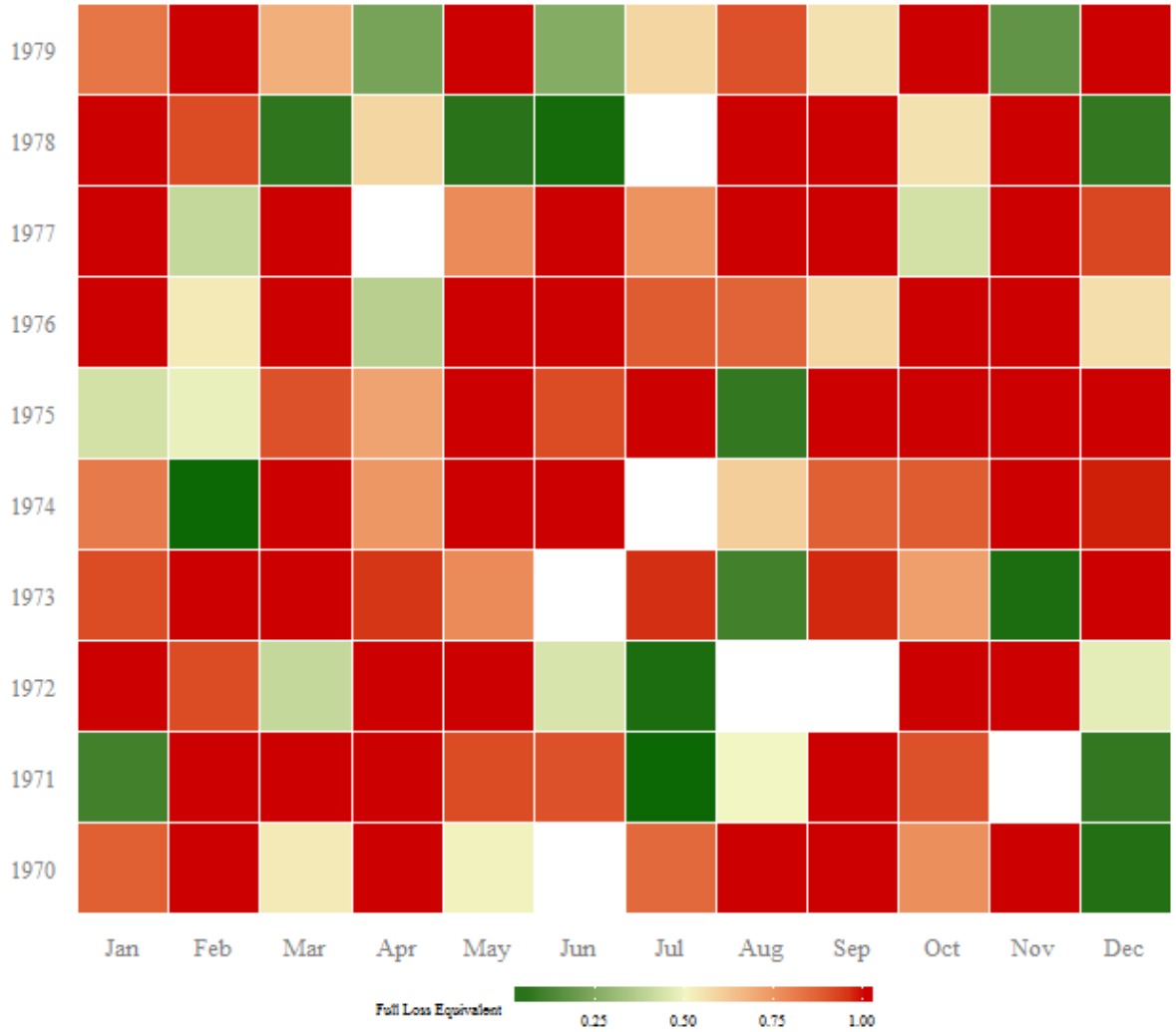


Zooming In

North America

peak in1970

INSIGHTS (7)



Zooming In

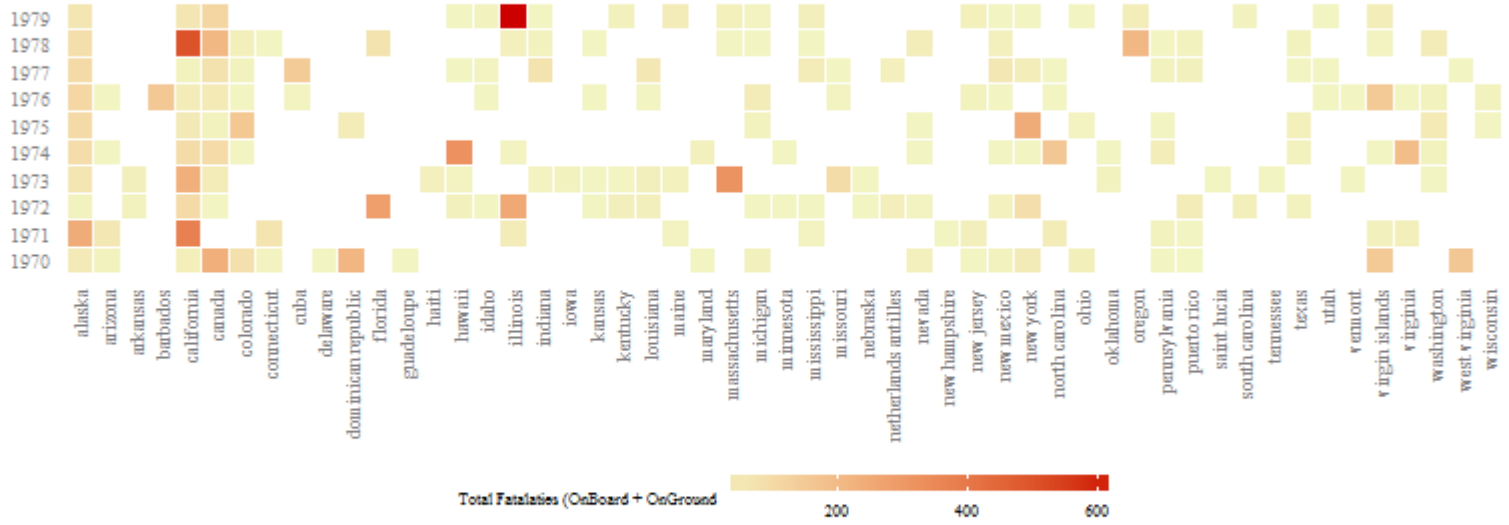
North America

peak in 1970

No visible difference across months or years, using FLE *

*: FLE or Full Loss Equivalent - Sum of the proportions of passengers killed. For example, 50 out of 100 passengers killed on a flight is an FLE of 0.50

INSIGHTS (8)



Zooming In

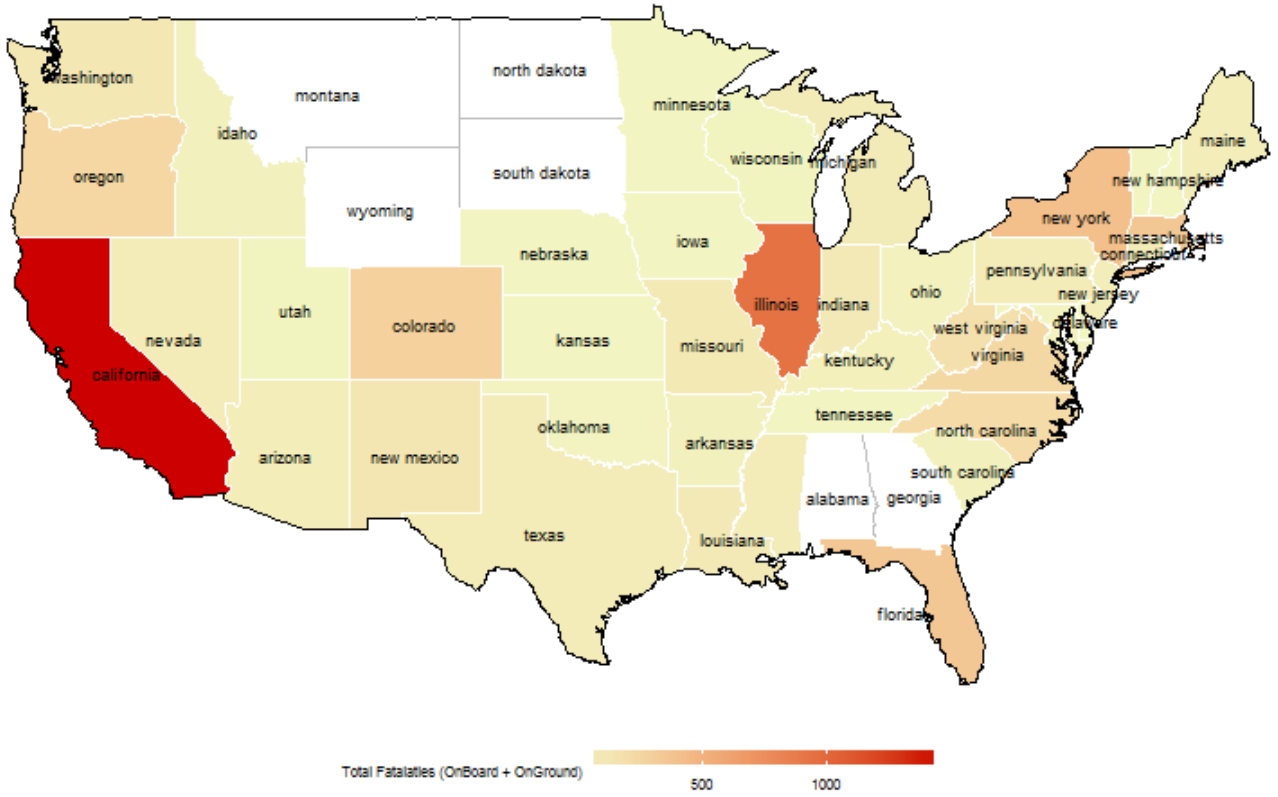
North America

peak in 1970

Illinois in 1979 and California

in 1978 stand out

INSIGHTS (9)



Zooming In

North America

peak in 1970

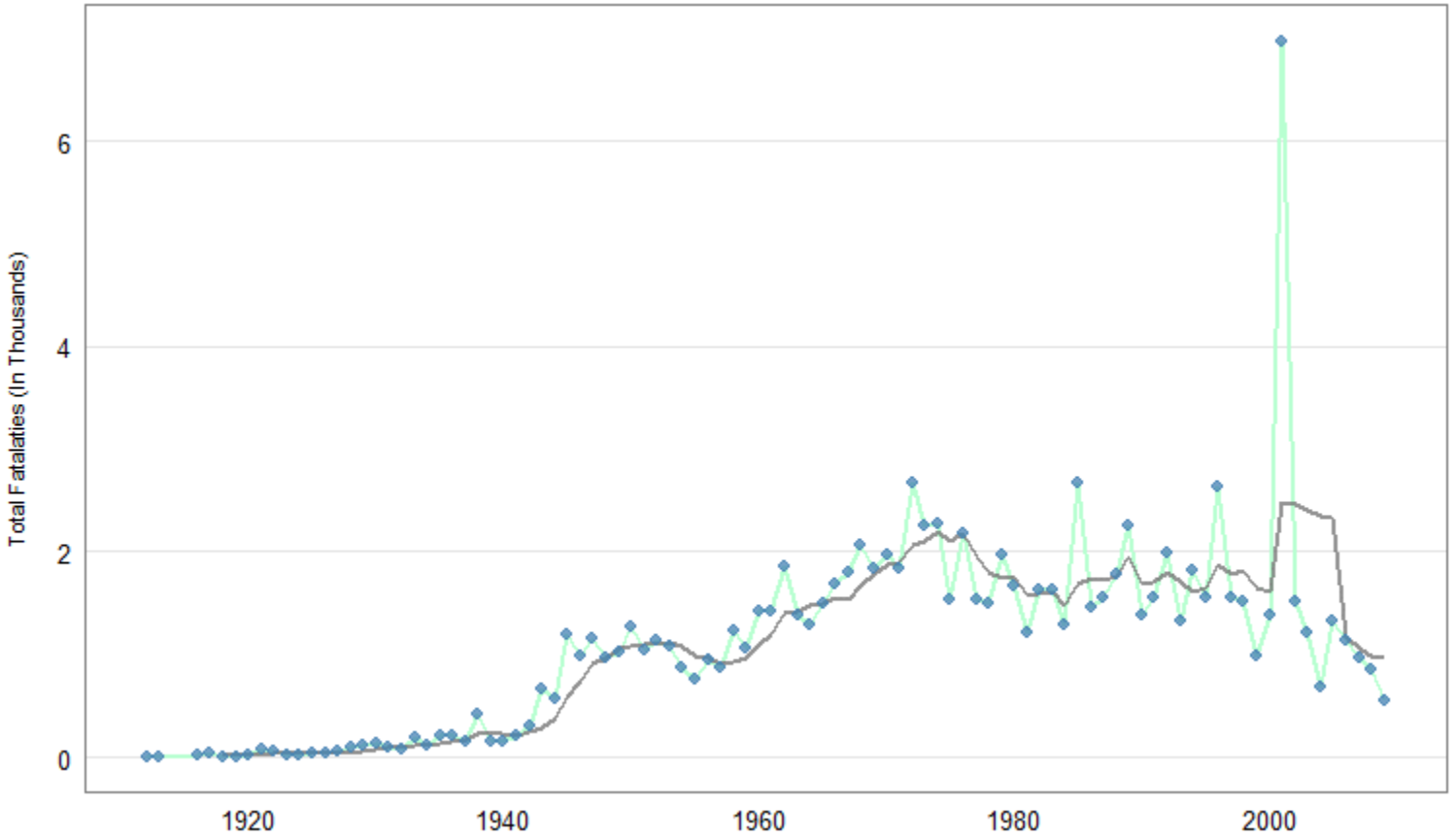
Illinois in 1979

&

California in 1978

stand out

INSIGHTS (10)



Total Fatalities are

declining ⬇

since mid 70s

barring some exceptions

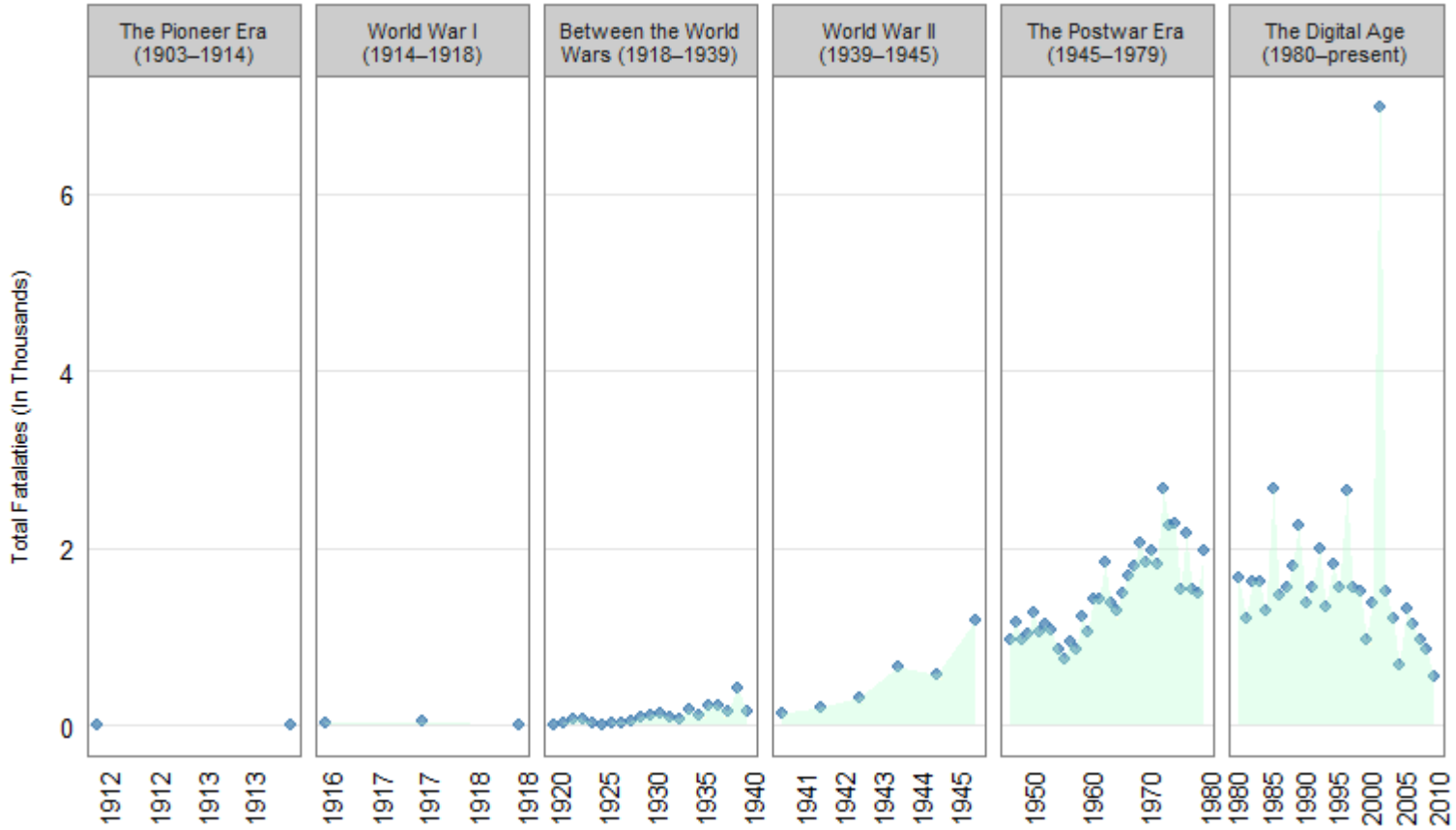
Grey line indicates

5 year rolling mean

or

moving average

INSIGHTS (11)

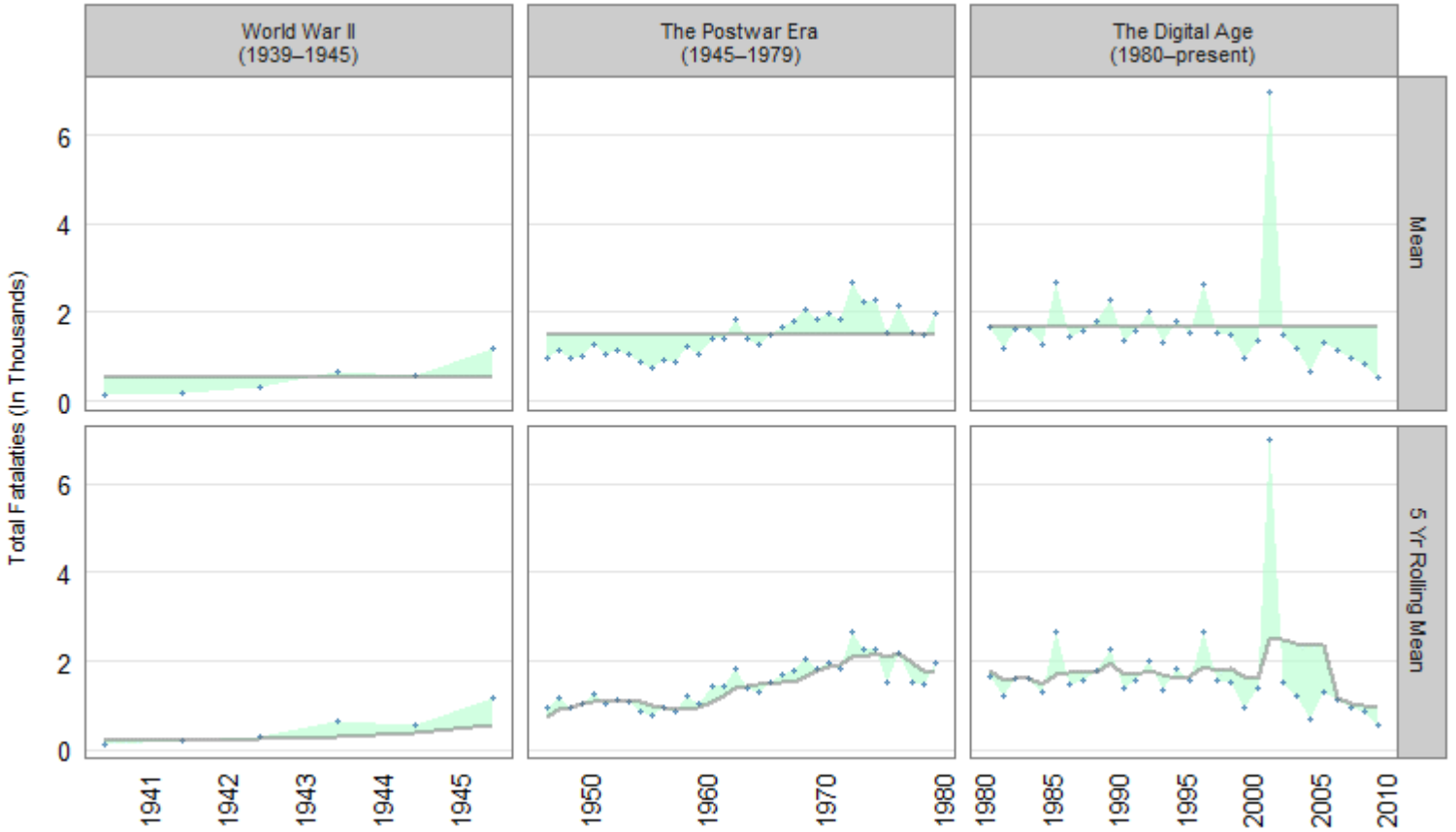


Trends across different ERA(s) of Aviation History

All action in last 3

lets dive in …

INSIGHTS (12)



Variation

across <u>last 3 Era(s)</u>

vis-a-vis

mean

&

5 year rolling mean

INSIGHTS (13)

Sabotage

has the largest variation

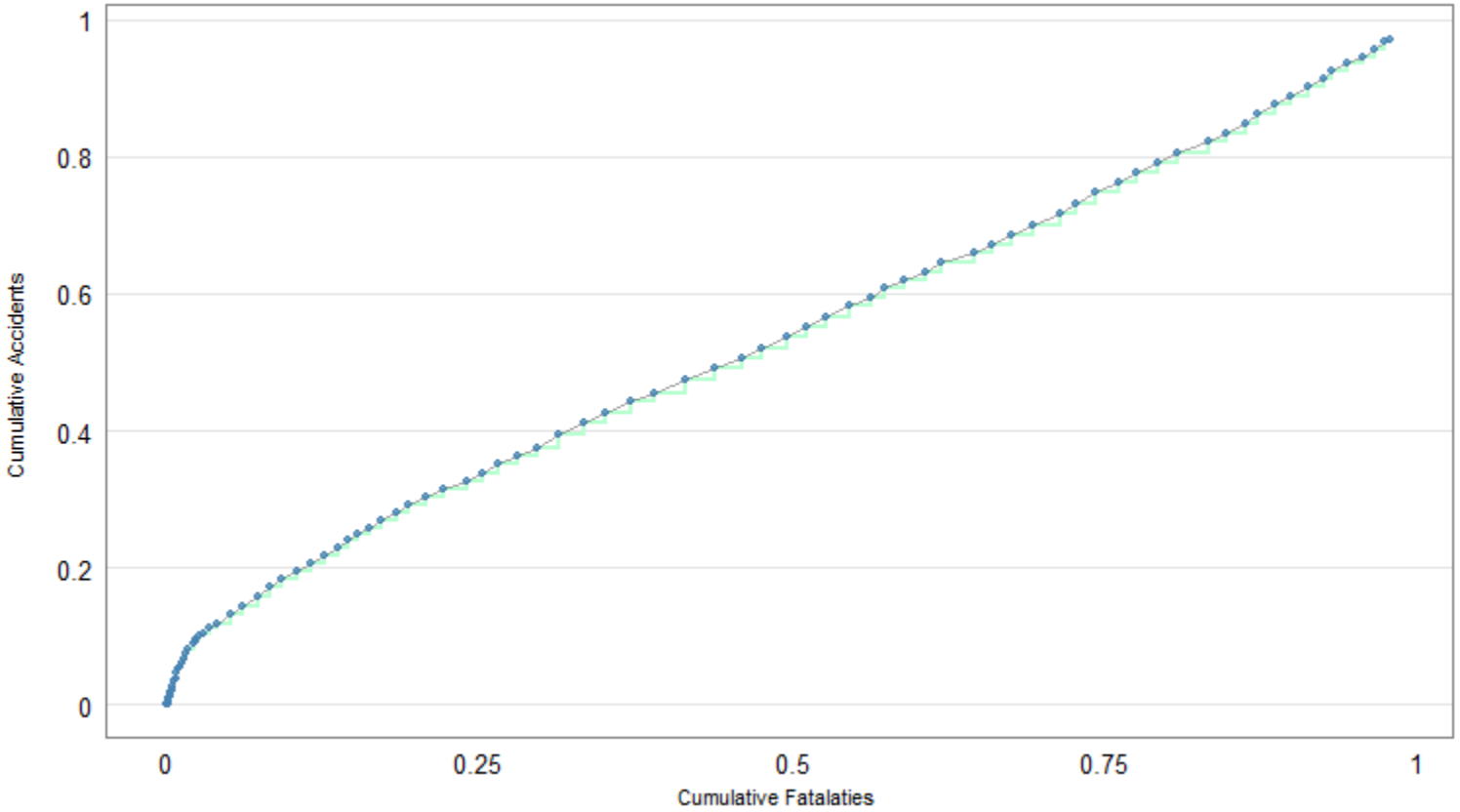
between

OnBoard Fatalities Per Accident

&

OnGround Fatalities Per Accident

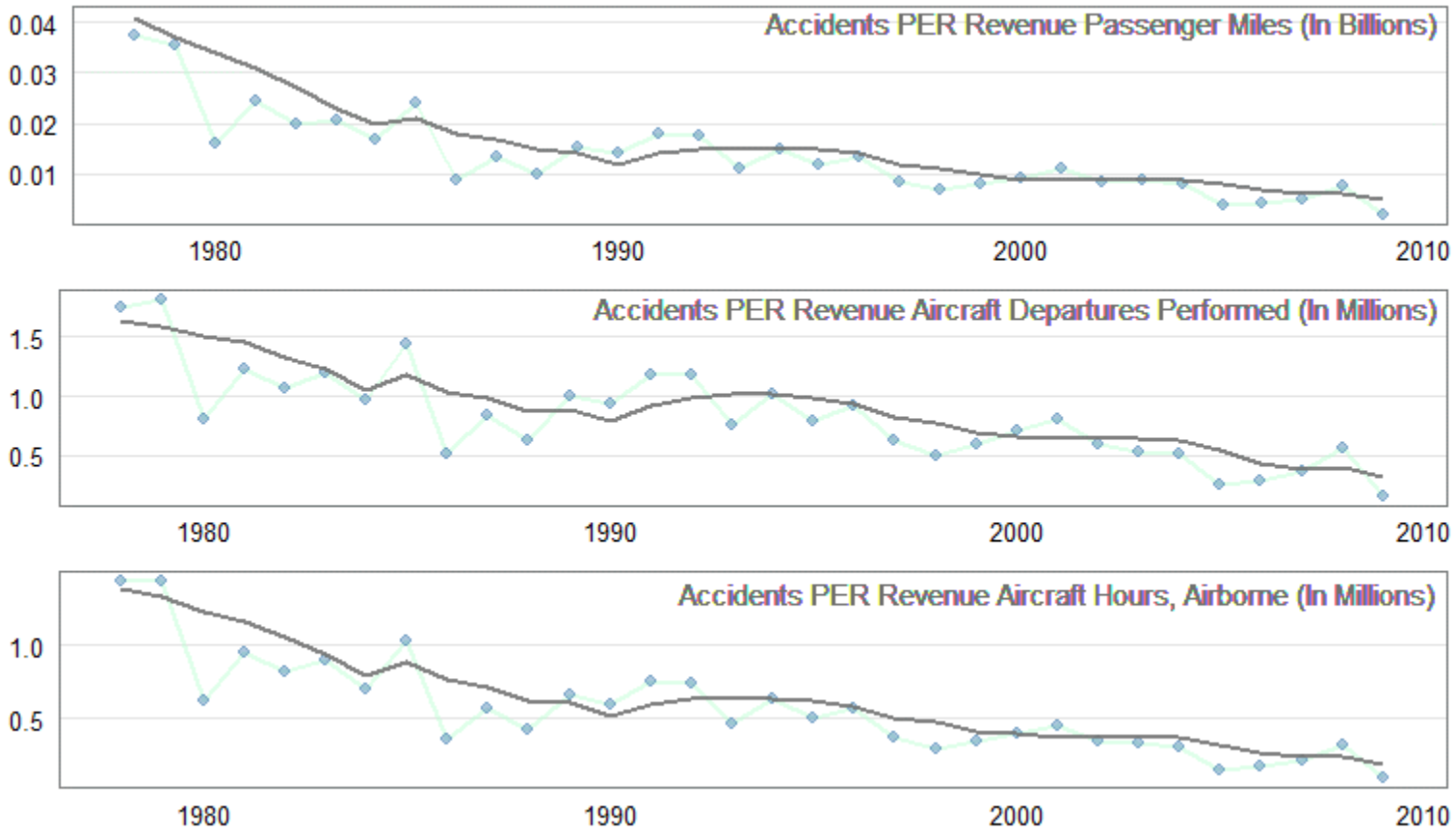
INSIGHTS (14)



Cumulative Frequency Graph –

how many accidents proportionately cause how many fatalities

INSIGHTS (15)



Accident Rate

has continuously dropped on unit basis *

Similar trend for all three:

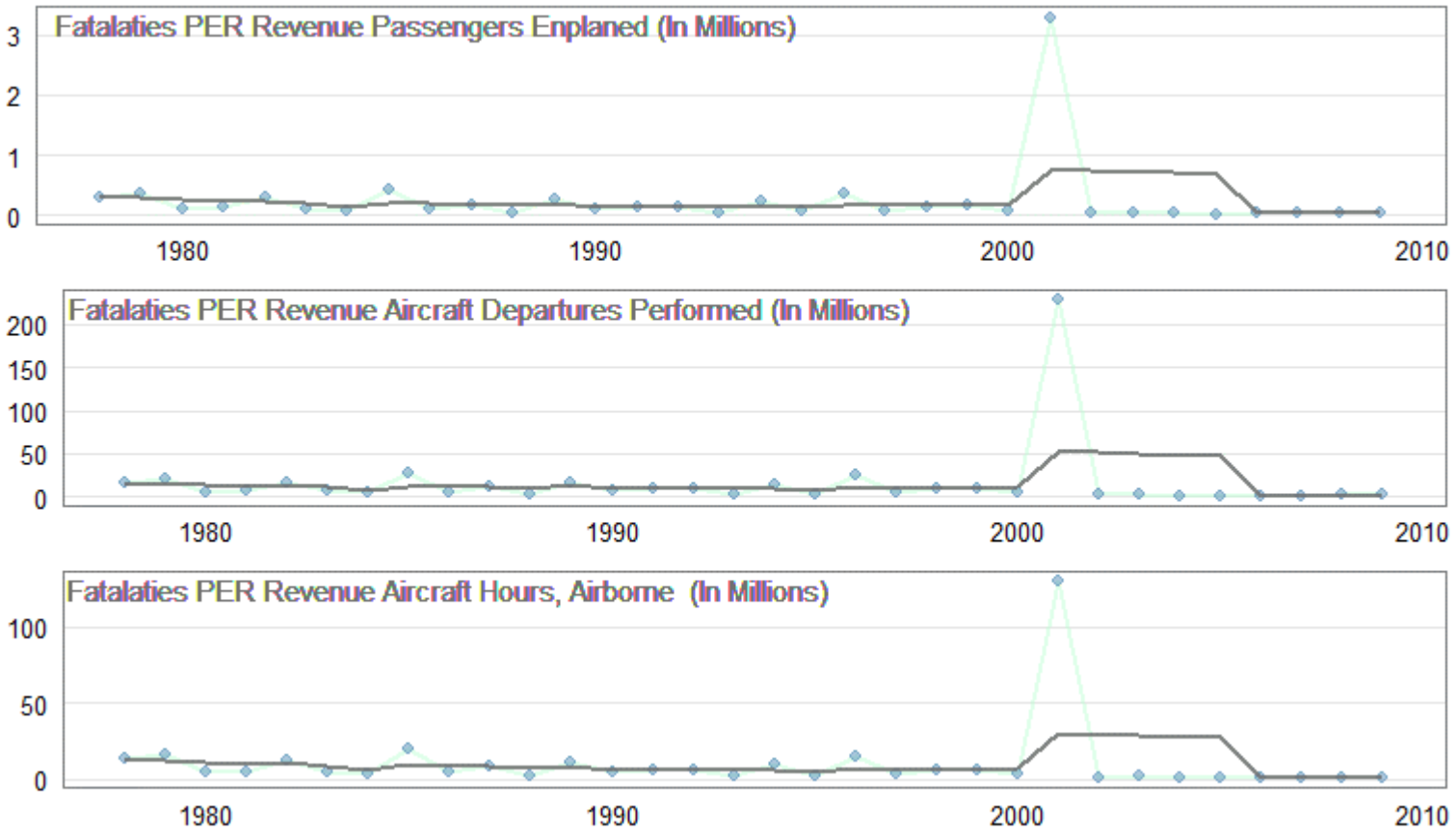
Per Passenger Miles

Per Departures

Per Aircraft Hours

*: US Traffic Data available post 1977, hence analysis includes only a subset of given data. Data Source: Bureau of Transportation Statistics

INSIGHTS (16)



Fatalities

has continuously dropped

on unit basis * analogously

Similar trend for all three:

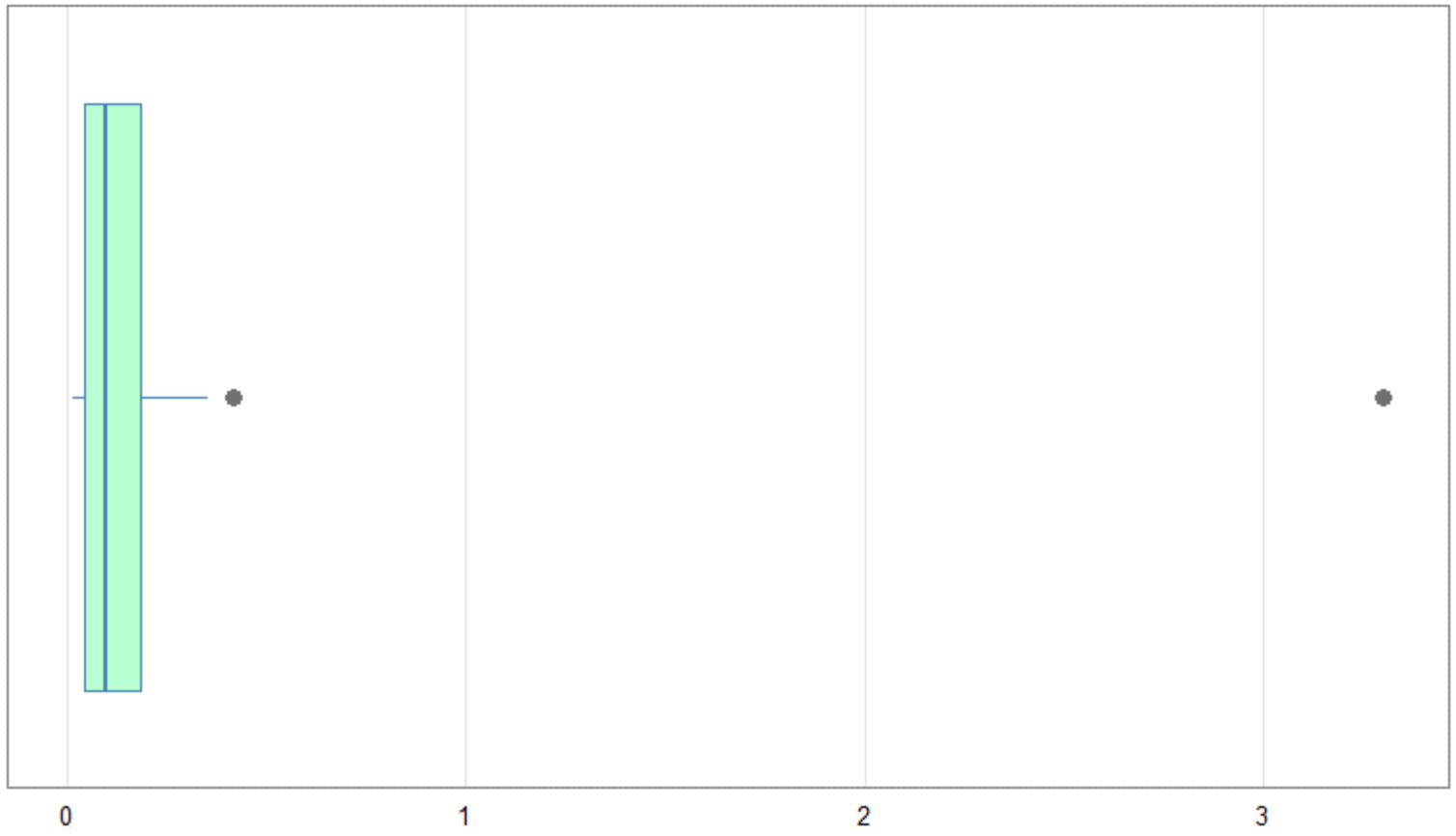
Per Enplaned Passenger

Per Departures

Per Aircraft Hours

*: US Traffic Data available post 1977, hence analysis includes only a subset of given data. Data Source: Bureau of Transportation Statistics

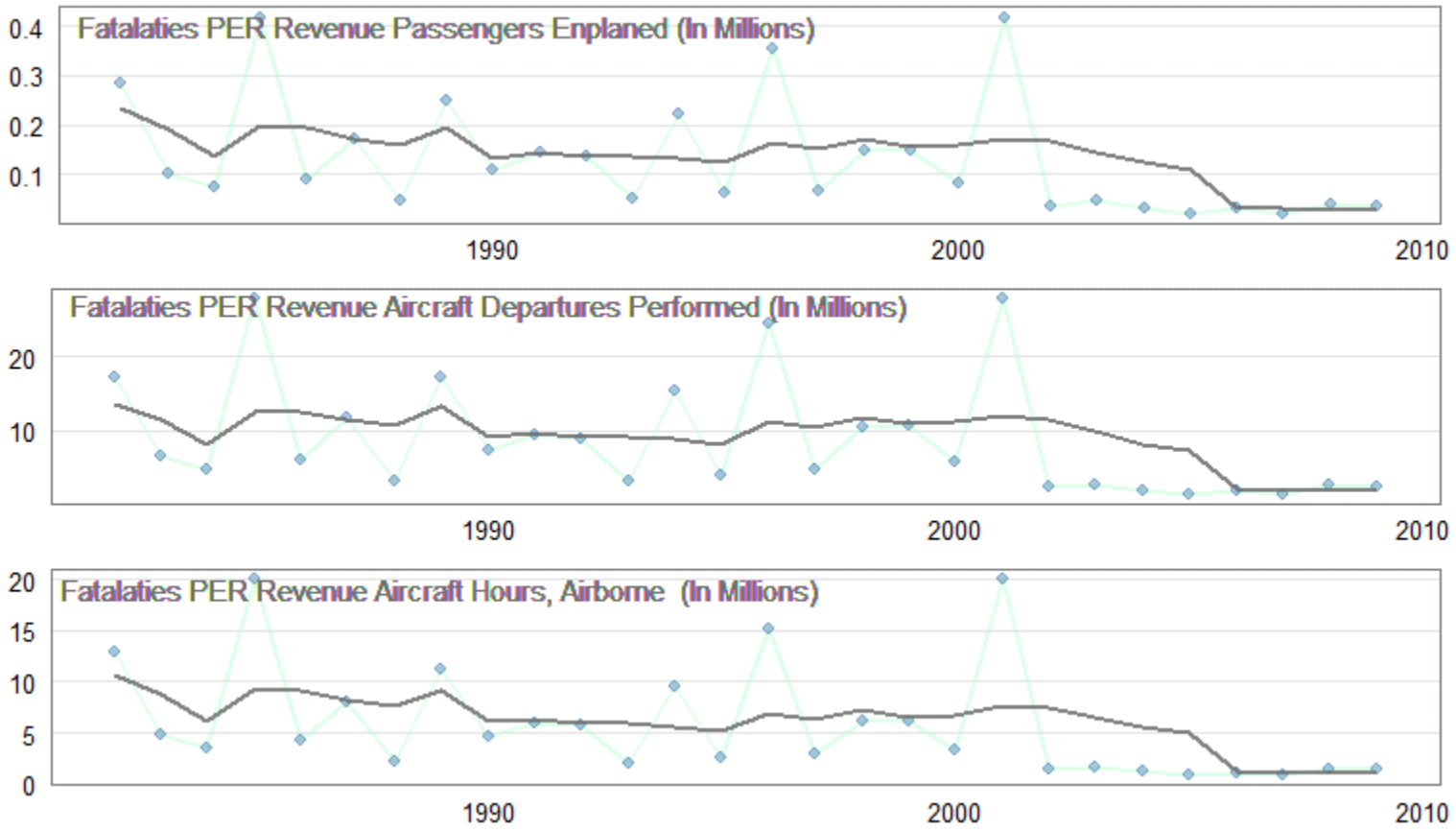
INSIGHTS (17)



Outlier Value of 3.296 occurs in year 2001 (9/11 debacle was a special event)

Winsorization, replacing 3.296 with second highest value 0.417. And similarly on the min end (0.017 with 0.0214)

INSIGHTS (18)



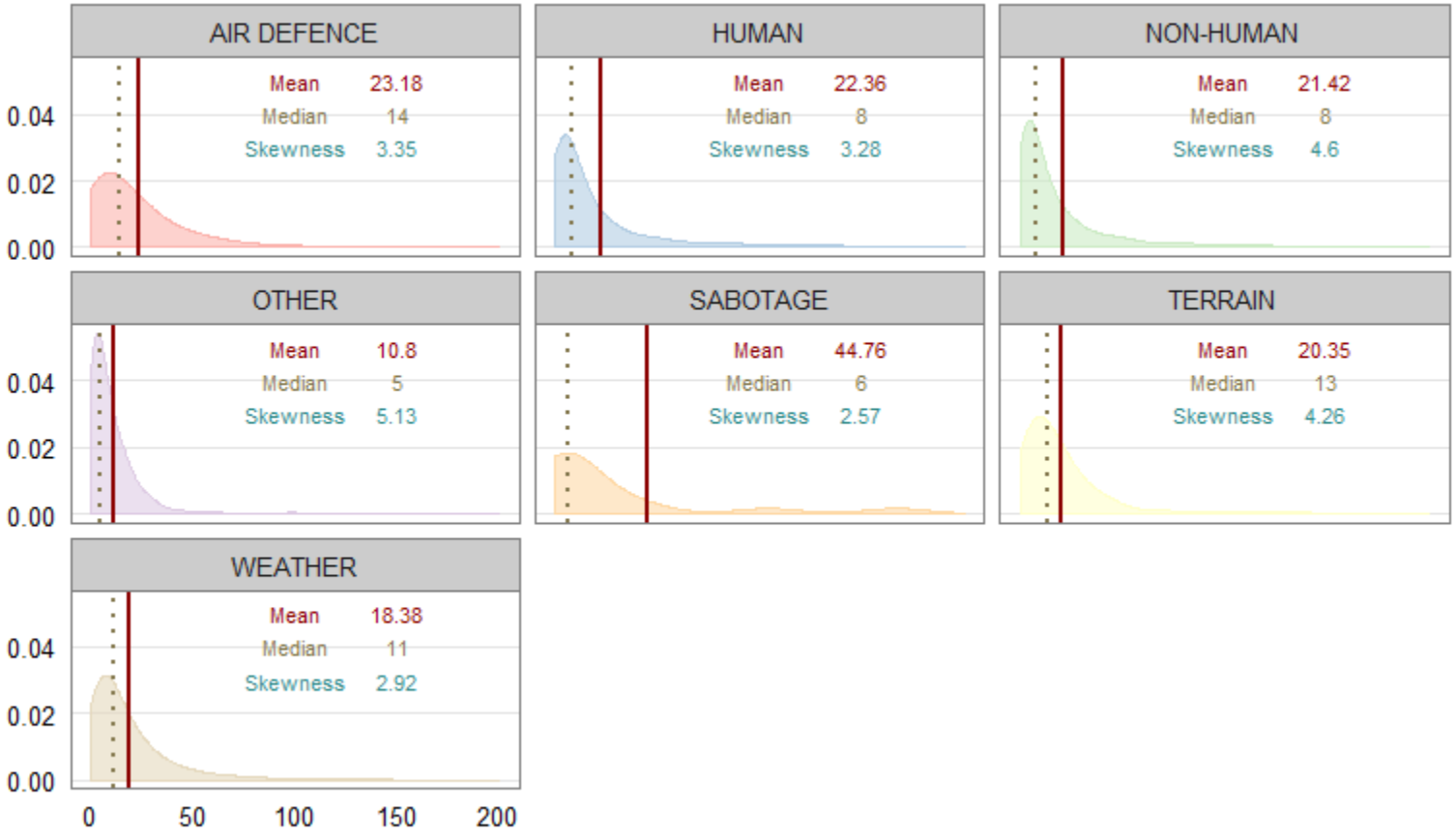
Fatalities

have continuously dropped on unit basis * analogously

Smoother rolling mean with much cleaner downward trend

*: US Traffic Data available post 1977, hence analysis includes only a subset of given data. Data Source: Bureau of Transportation Statistics

INSIGHTS (19)



Sabotage

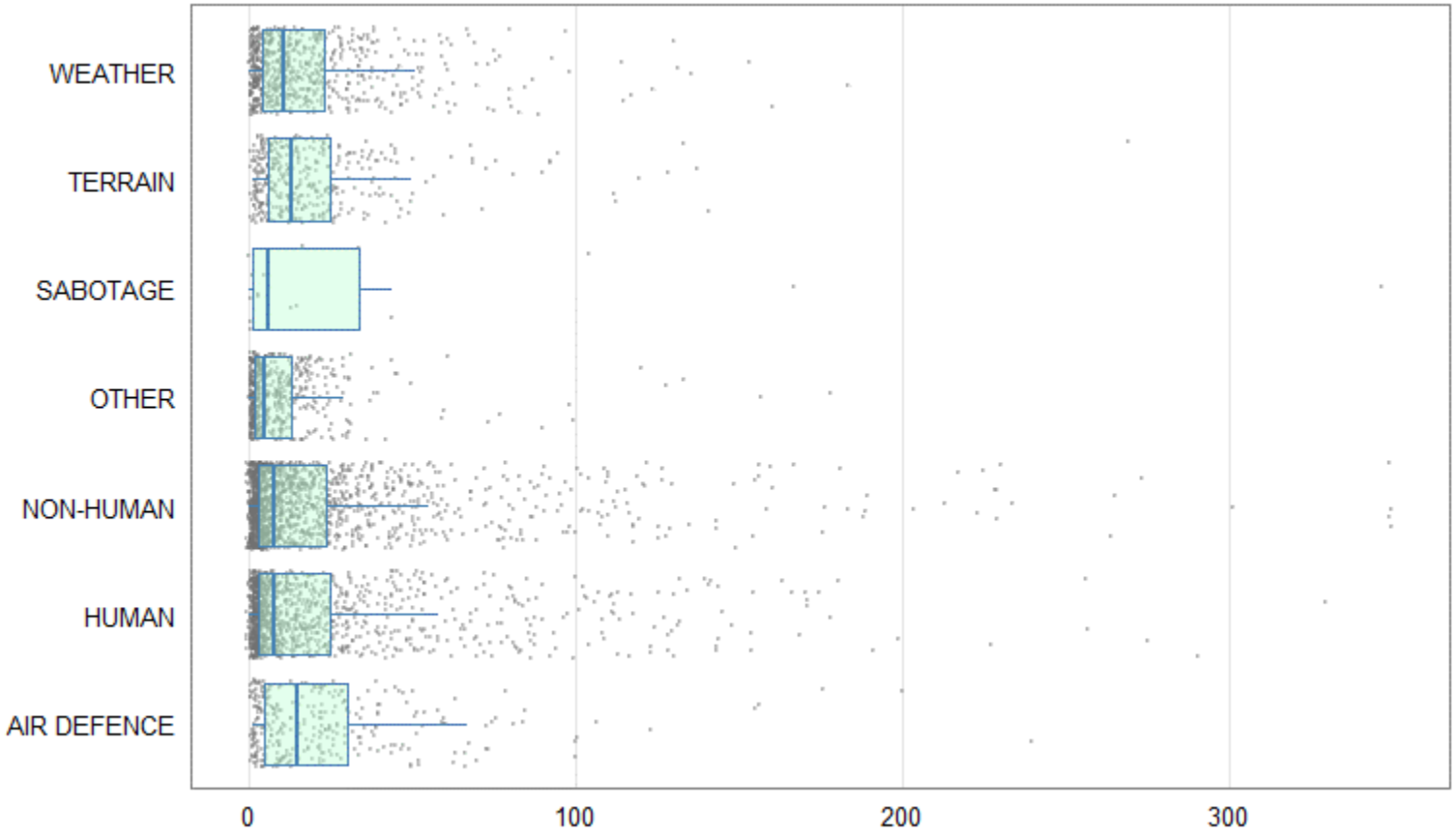
has lowest skew

&

highest delta between mean and median

Fatalities Distribution are positively or right skewed, with Mean > Median in all cases

INSIGHTS (20)



Distribution of

Total Fatalities*

across various

Reason categories

HIGH number of accidents

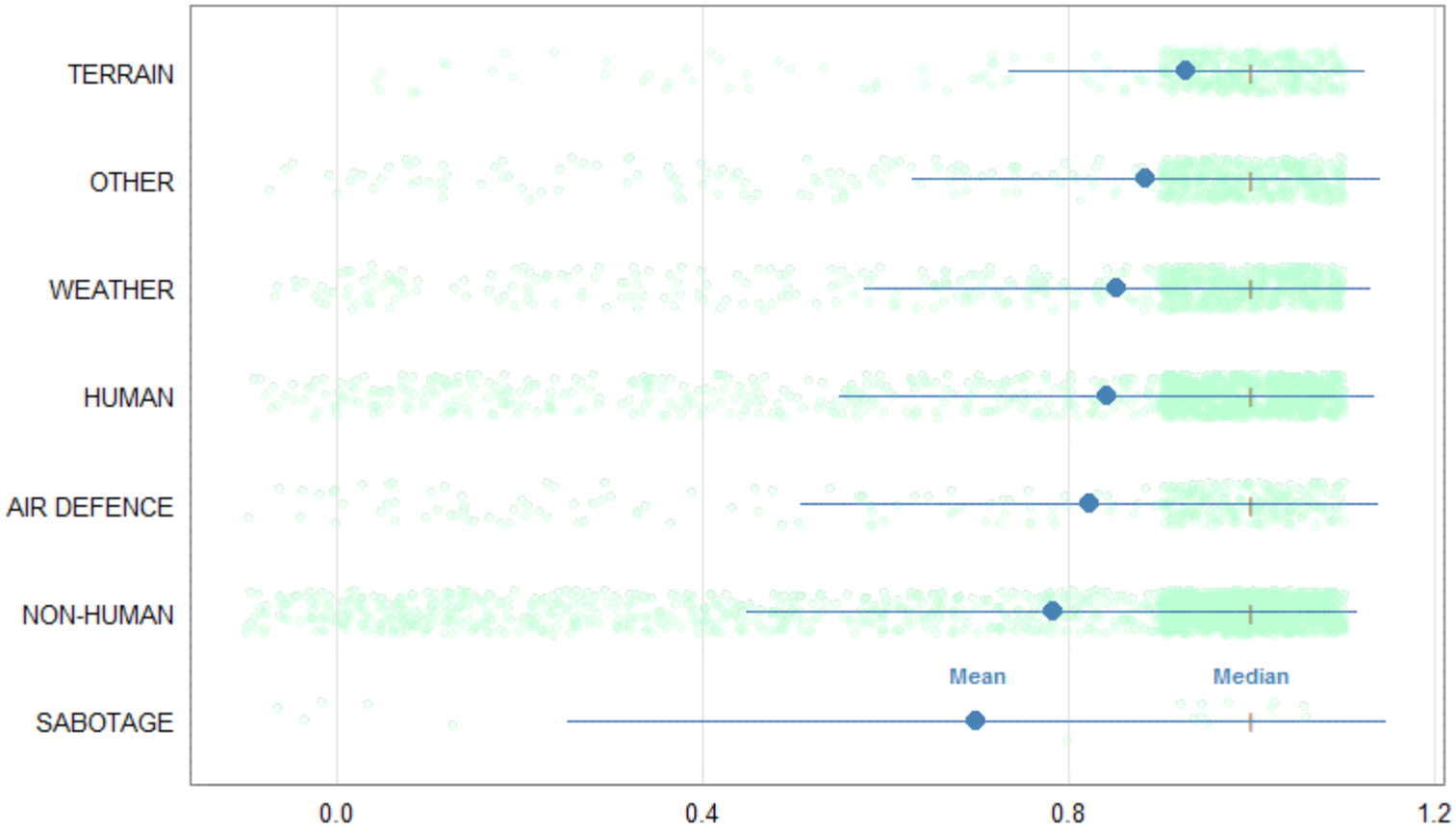
with FEW fatalities

Heavily +ve concentrated or

right skewed distribution

*: Winsorization, replacing top three and bottom three values for NON HUMAN to remove the outliers

INSIGHTS (21)



Distribution of

FLE*

across various

Reason categories

HIGH number of accidents

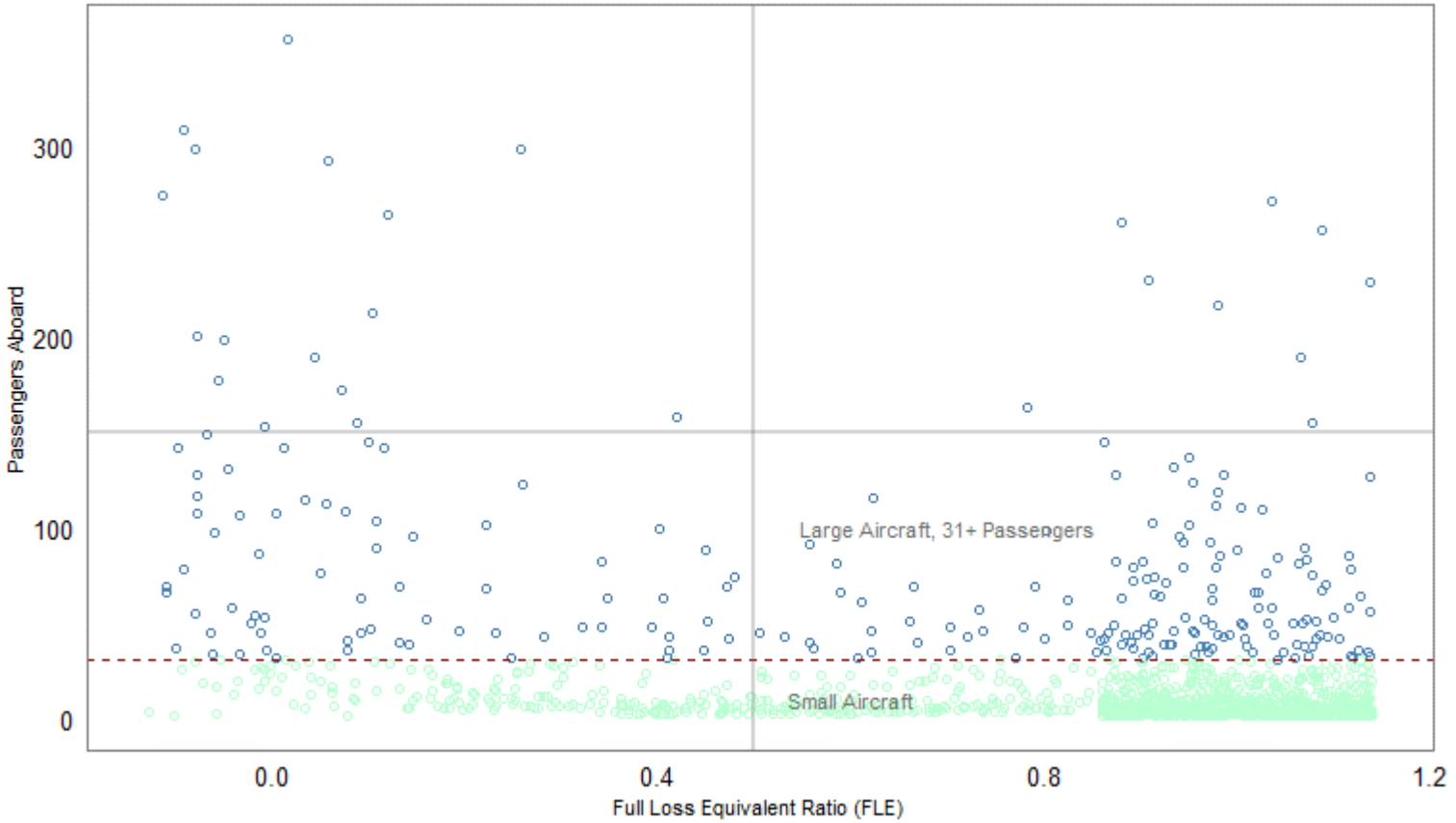
with HIGH FLE rate

Heavily –ve concentrated or

left skewed distribution

*: FLE or Full Loss Equivalent - Sum of the proportions of passengers killed. For example, 50 out of 100 passengers killed on a flight is an FLE of 0.50

INSIGHTS (22)



#20 and #21 indicate

MOST accidents with

FEW fatalities &

HIGH fatality rate (FLE =1)

Split

Passengers Aboard vs FLE

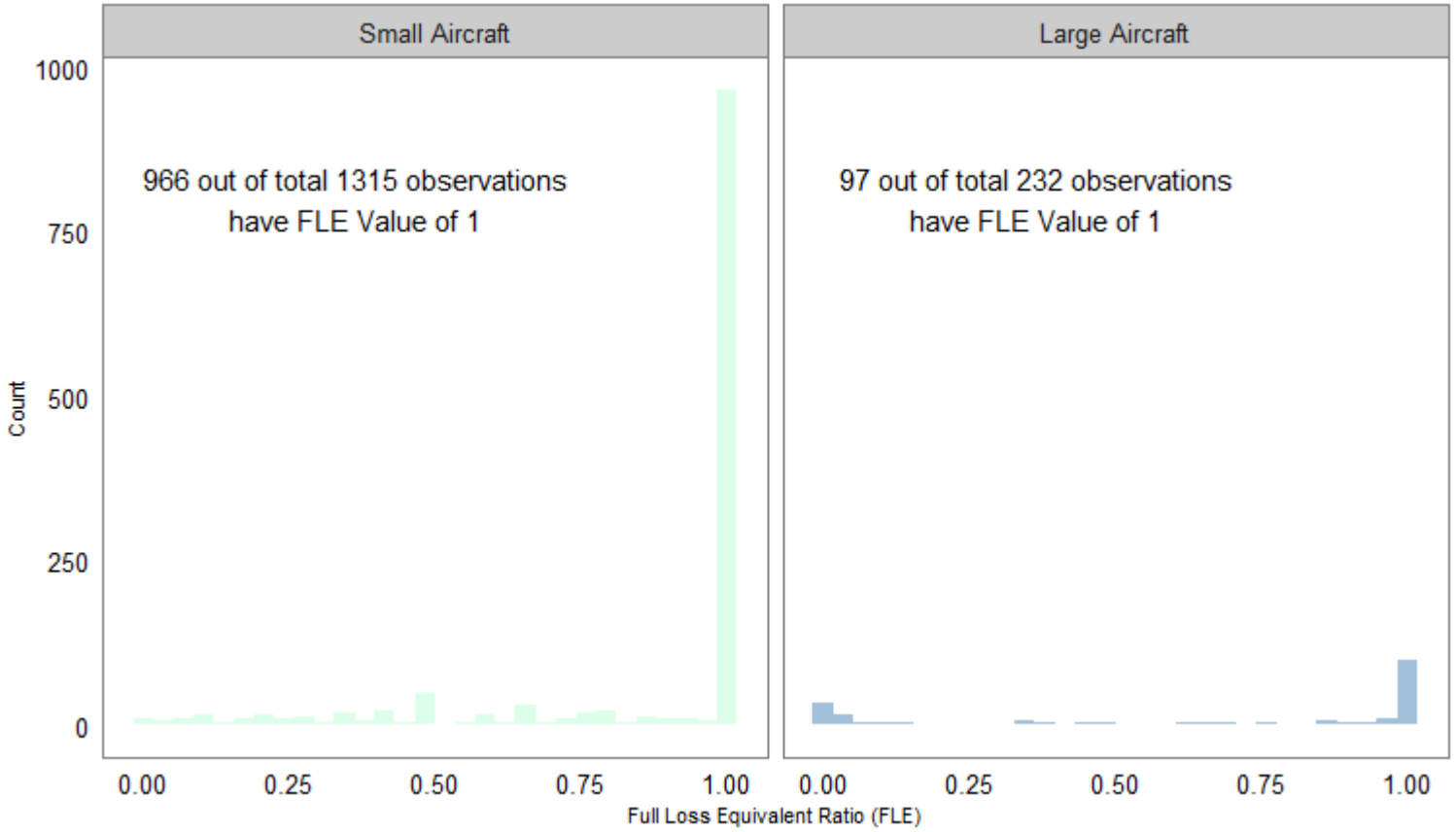
based on

Small and Large Aircraft *

criteria

*: FAA Definition of Small and Large Aircraft (at least 31 passenger seats) Based on 14 CFR Part 139

INSIGHTS (23)

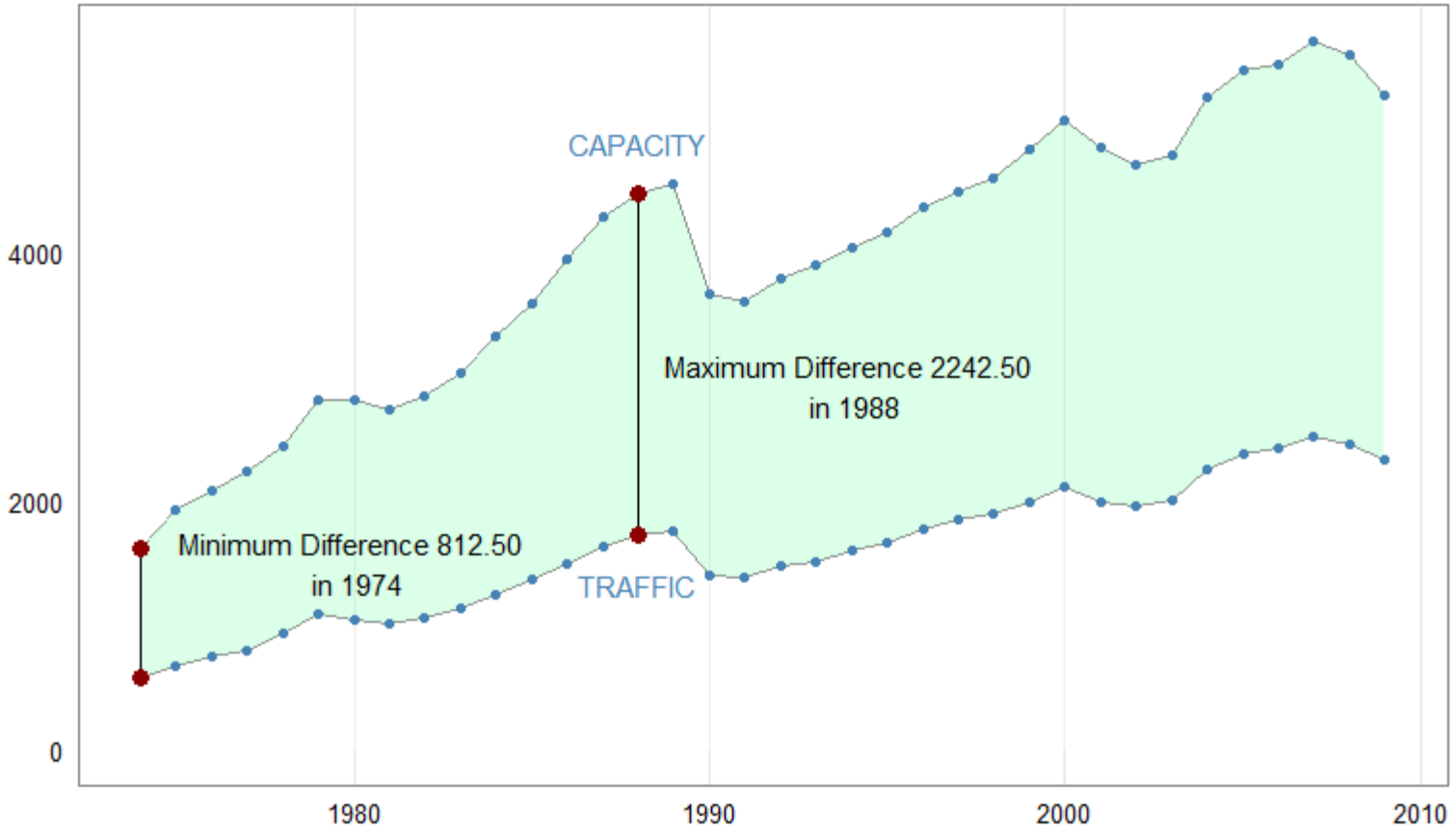


#23 validates

Most values concentrated in lower right corner

of #22

INSIGHTS (24)



Difference between

Total Capacity

(Available Seat Miles in Billions)

And

Total Traffic

(Revenue Passenger Miles in Billions)

*: Based on US Data from Bureau of Transportation Statistics