

# Boeing and Airbus Crashes Analysis 2000-2022

Nirawit Pongpisut

2024-04-24



## Introduction

Air travel has become easier and cheaper than ever before. This is partly due to the emergence of new airlines each year, which creates more competition and drives down prices. However, the industry remains largely dominated by the top two manufacturers, Airbus and Boeing. Due to recent negative news reports, there have been concerns that Boeing aircraft have been involved in more crashes in recent years, Let's find that Boeing's accident more than Airbus or not.

## Question

1. Which manufacturer made highest accident?
2. Which aircraft type is the highest accident?
3. What are the reason of accident?
4. What aircraft type is the highest accident cause by technical failure?

## Tools

1. Posit Cloud(R)

# Dataset

Historical Plane Crash Data(Kaggle.com) (<https://www.kaggle.com/datasets/abeperez/historical-plane-crash-data>)

## Analysis

### Download Library

```
library(tidyverse)
library(ggplot2)
library(dplyr)
library(lubridate)
library(ggthemes)
library(forcats)
library(RColorBrewer)
```

### Read CSV File and Clean-up The Data

```
# read csv file
df <- read_csv("Plane Crashes.csv")
# remove duplicate
df_clean <- df[!duplicated(df), ]
# create year column
df_clean$Year <- year(df_clean$Date)

#check row and column's name
nrow(df_clean)
```

```
## [1] 28457
```

```
colnames(df_clean)
```

```
## [1] "Date"          "Time"          "Aircraft"      "Operator"
## [5] "Registration" "Flight phase"  "Flight type"   "Survivors"
## [9] "Crash site"    "Schedule"      "MSN"           "YOM"
## [13] "Flight no."    "Crash location" "Country"       "Region"
## [17] "Crew on board" "Crew fatalities" "Pax on board"  "PAX fatalities"
## [21] "Other fatalities" "Total fatalities" "Circumstances" "Crash cause"
## [25] "Year"
```

## Find The Answer of The Questions

1.Who is the most accident (all accident)?

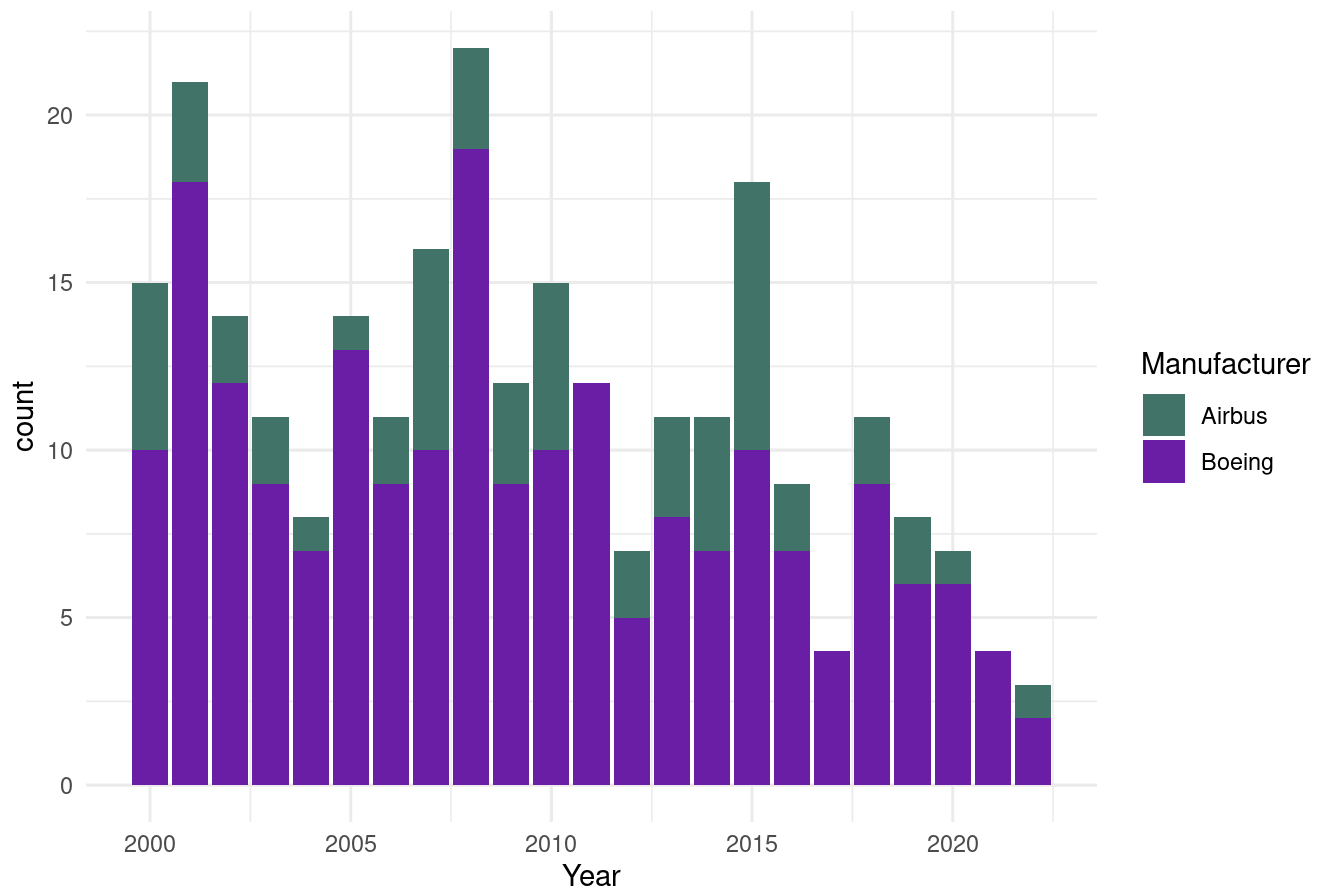
```
# filter for Boeing and Airbus crash between 2000-2022 (by all Reason)
# filter Airbus A3XX and Boeing 7XX because commercial aircraft of Airbus start with A3XX and Boeing start with 7XX, The others is for military aircraft
aircrashes_all_reason <- df_clean %>%
  filter(grepl("^Airbus A3", Aircraft) | grepl("^Boeing 7", Aircraft)) %>%
  mutate(Manufacturer = if_else(grepl("^Airbus", Aircraft), "Airbus", "Boeing")) %>%
  filter(Year >= 2000)

aircrashes_all_reason %>%
  count(Manufacturer)
```

```
## # A tibble: 2 × 2
##   Manufacturer      n
##   <chr>          <int>
## 1 Airbus           58
## 2 Boeing          206
```

```
# bar chart
ggplot(aircrashes_all_reason, aes(Year, fill=Manufacturer)) +
  geom_bar() +
  ggtitle("Airbus and Boeing Crashed Between 2000-2022") +
  theme_minimal() +
  scale_fill_manual(values=c('#43766C', '#6C22A6'))
```

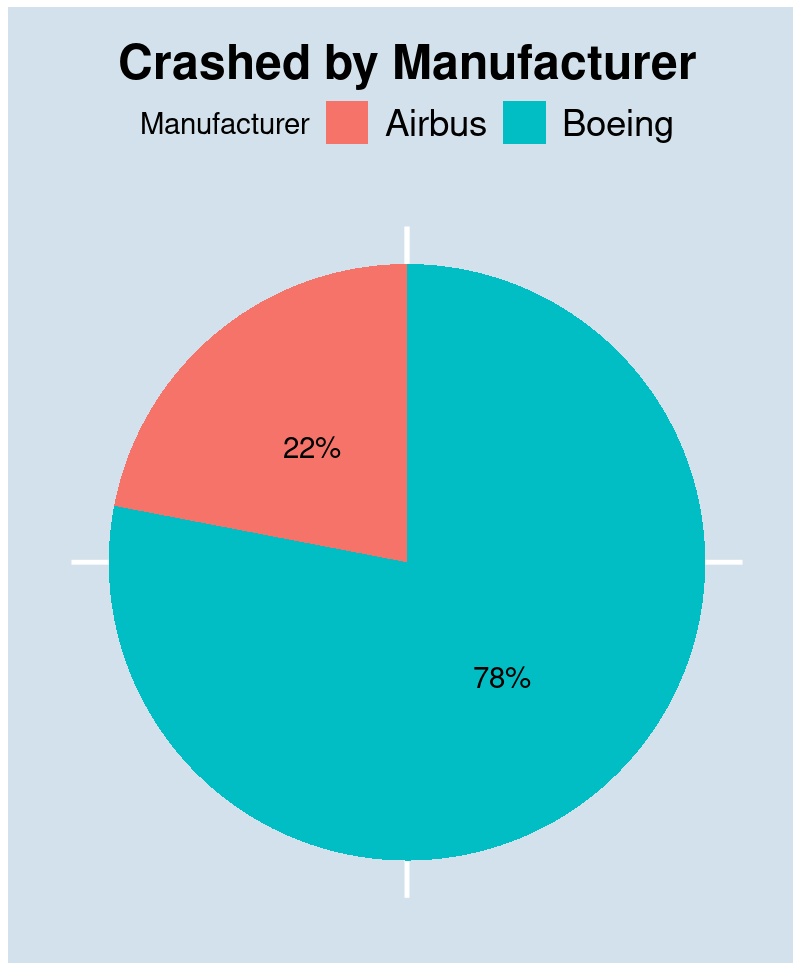
## Airbus and Boeing Crashed Between 2000-2022



You will see that Boeing have many aircraft crash from time to time.

```
# make a percentage for pie chart
percentage_manufacturer <- aircrashes_all_reason %>%
  group_by(Manufacturer) %>%
  summarize(counts = n(),
            percentage = n()/nrow(aircrashes_all_reason))

# Pie Chart : Crashed by Manufacturer
ggplot(percentage_manufacturer, aes(x="", y=percentage, fill=Manufacturer)) +
  theme_economist() +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0) +
  geom_text(aes(label = paste0(round(percentage*100), "%")),
            position = position_stack(vjust = 0.5)) +
  theme(panel.background = element_blank(),
        axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        axis.title = element_blank(),
        plot.title = element_text(hjust = 0.5, size = 18)) +
  ggtitle("Crashed by Manufacturer")
```

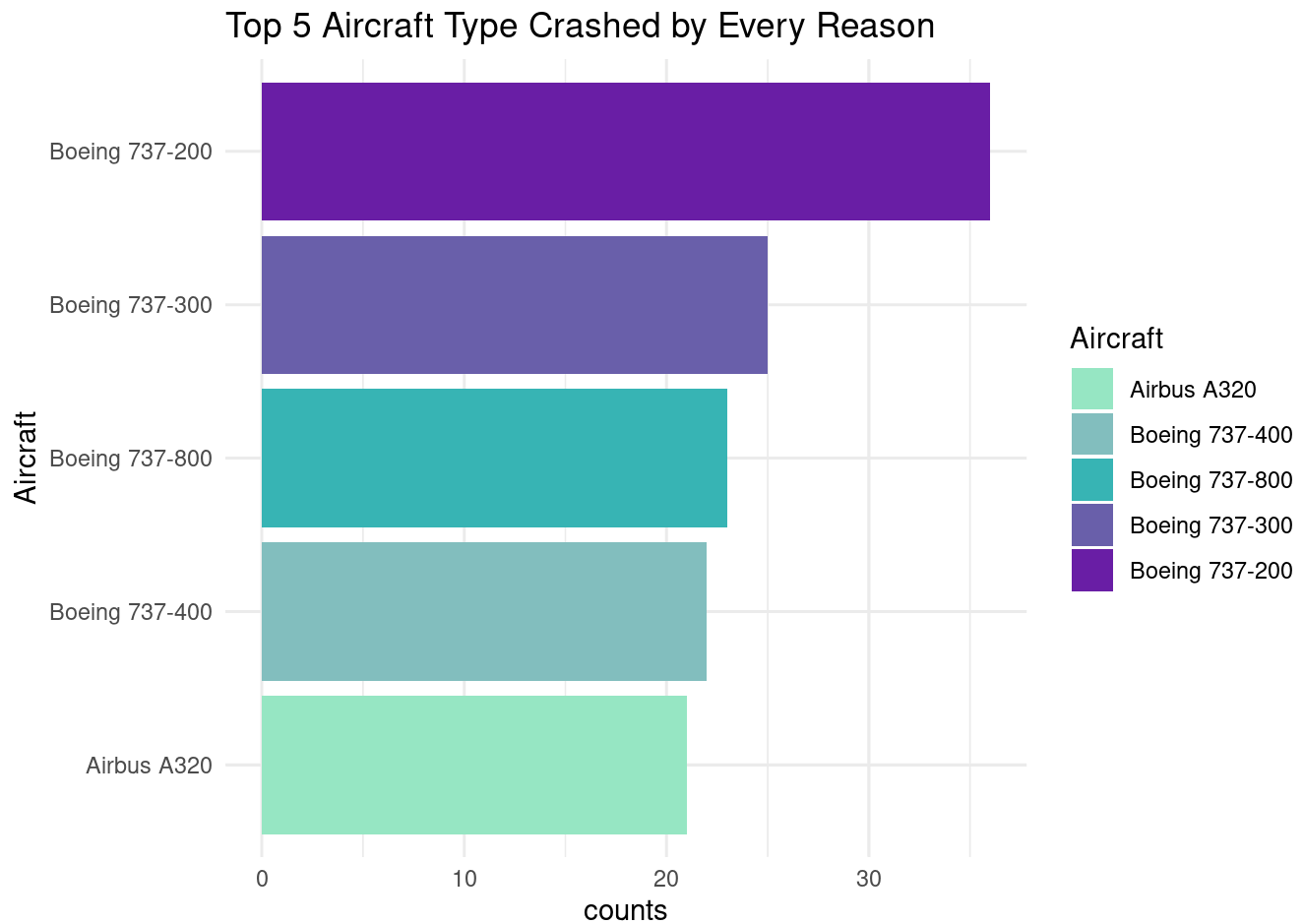


From 2000 to 2022 Boeing has an accident more a half of all event.

## 2. Which aircraft type is the highest accident?

```
count_type <- head(aircrashes_all_reason %>%
  select(Aircraft) %>%
  group_by(Aircraft) %>%
  summarise(counts = n()) %>%
  arrange(desc(counts)),5)

# bar chart : Top 5 Aircraft Type Crashed by Every Reason
count_type %>%
  select(Aircraft, counts) %>%
  mutate(Aircraft = fct_reorder(Aircraft, counts)) %>%
  ggplot(aes(Aircraft, counts, fill=Aircraft)) +
  geom_col() +
  theme_minimal() +
  coord_flip() +
  ggtitle("Top 5 Aircraft Type Crashed by Every Reason") +
  scale_fill_manual(values=c('#96E9C6', '#83C0C1', '#37B5B6', '#6962AD', '#6C22A6'))
```

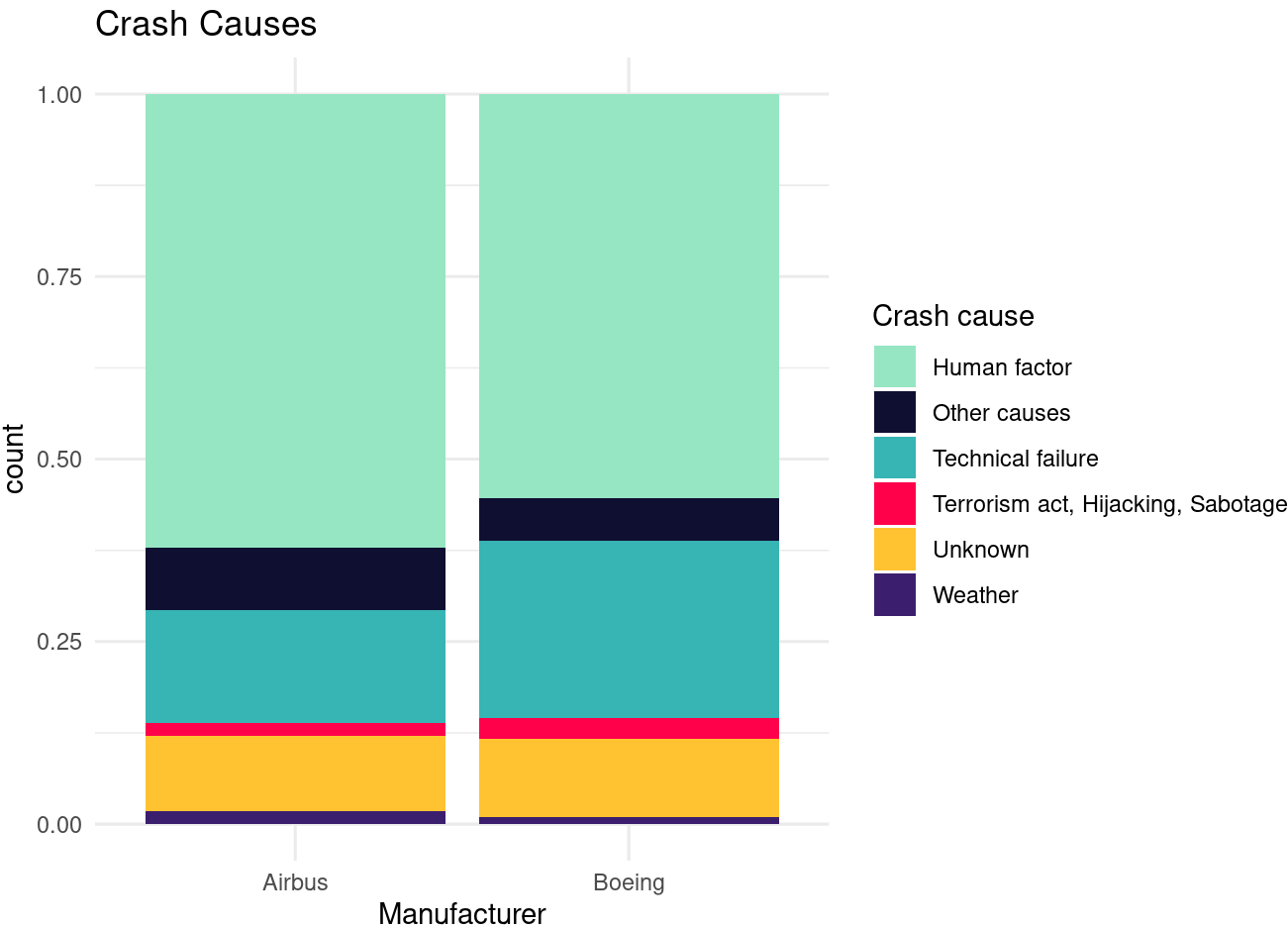


Looking at each type, Top 4 of the most accidents are from Boeing but we will find the crash cause in the next step. Maybe the accident caused by another reason that not by technical failure from manufacturer.

### 3. What are the reason of accident?

```
crash_cause <- select(aircrashes_all_reason, Manufacturer, `Crash cause`)

# plot1 : 100% bar stack chart
ggplot(crash_cause, aes(Manufacturer, fill=`Crash cause`)) +
  geom_bar(position = "fill") +
  theme_minimal() +
  ggtitle("Crash Causes") +
  scale_fill_manual(values=c('#96E9C6', '#0F1035', '#37B5B6', '#FF004D', '#FFC436', '#401F71'))
```



The most crash cause is from Human Factor but if we focus on the Technical Failure, Boeing has m any case than Airbus.

```
# Let's see how many case of accident that cause by technical failure?
# filter for Boeing and Airbus crash (Technical failure)
aircrashes_technical_reason <- aircrashes_all_reason %>%
  filter(grepl("Technical failure", `Crash cause`))

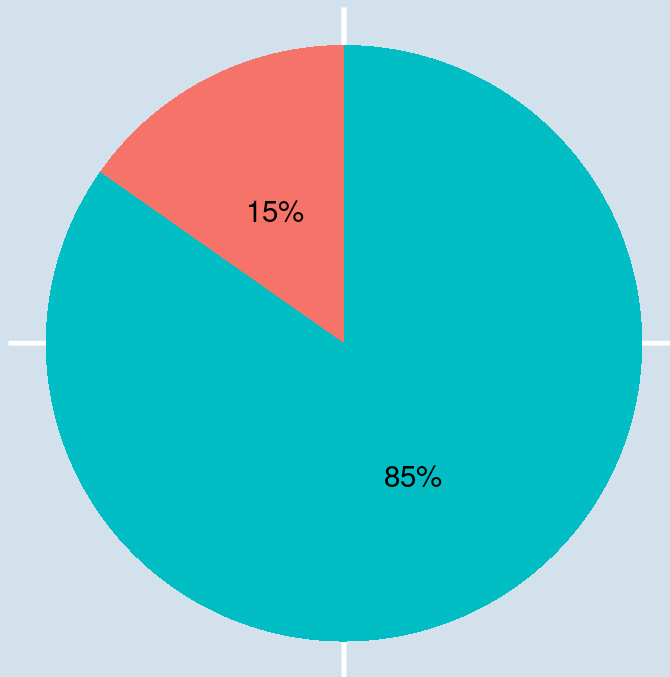
# percentage
percentage_thechnical_reason <-aircrashes_technical_reason %>%
  group_by(Manufacturer) %>%
  summarize(counts = n(),
            percentage = n()/nrow(aircrashes_technical_reason))

# Plot Pie Chart
ggplot(percentage_thechnical_reason, aes(x="", y=percentage, fill=Manufacturer)) +
  theme_economist() +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0) +
  geom_text(aes(label = paste0(round(percentage*100), "%")),
            position = position_stack(vjust = 0.5)) +
  theme(panel.background = element_blank(),
        axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        axis.title = element_blank(),
        plot.title = element_text(hjust = 0.5, size = 18)) +
  ggtitle("Crashed by Technical Failure")
```



## Crashed by Technical Failure

Manufacturer ■ Airbus ■ Boeing



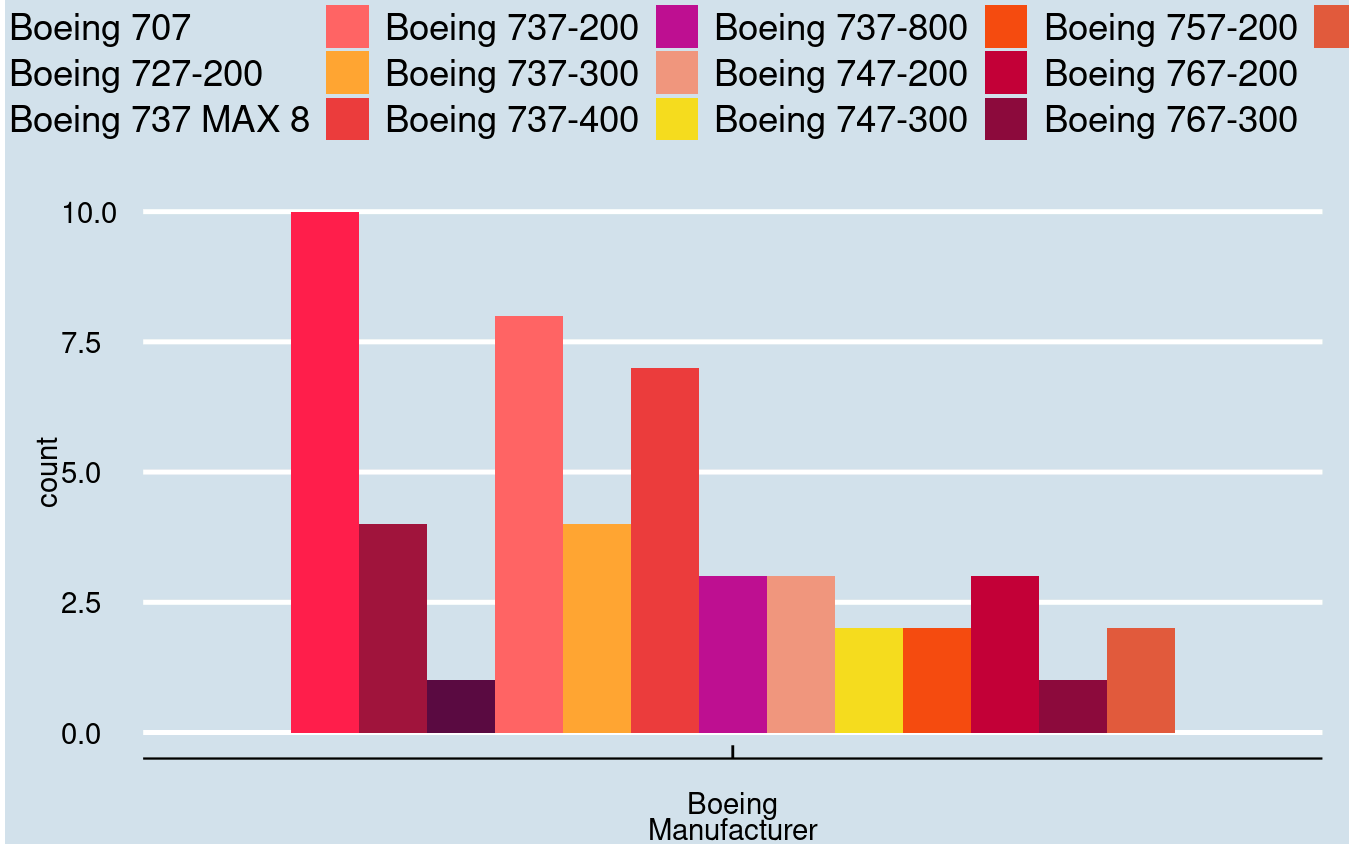
Focus on the Technical Failure, Boeing has up to 85% of all crash event from 2000-2022

4. What aircraft type is the highest accident cause by technical failure in each manufacturer?

```
## Boeing
boeing_tech_reason <- aircrashes_technical_reason %>%
  filter(grepl("^Boeing 7", Aircraft))

# plot boeing
ggplot(boeing_tech_reason, aes(Manufacturer, fill=Aircraft)) +
  geom_bar(position = "dodge")+
  theme_economist() +
  ggtitle("Boeing Crashed by Technical Failure in Each Aircraft Type") +
  scale_fill_manual(values=c('#FF204E', '#A0153E', '#5D0E41', '#FF6666',
                             '#FFA732', '#EF4040', '#C21292', '#F0997D',
                             '#F8DE22', '#F94C10', '#C70039', '#900C3F',
                             '#E25E3E'))
```

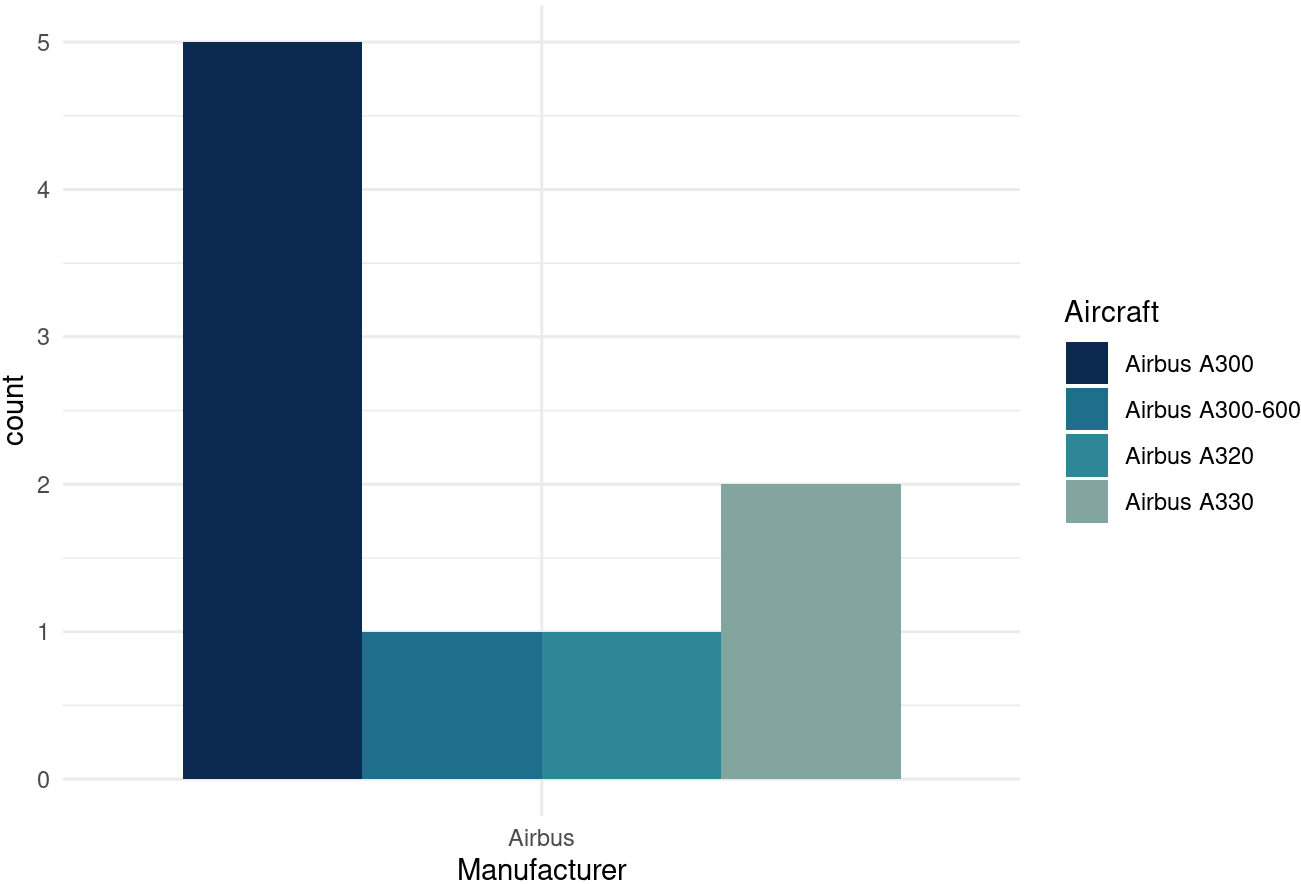
## Boeing Crashed by Technical Failure in Each Aircraft Type



```
## Airbus
airbus_tech_reason <- aircrashes_technical_reason %>%
  filter(grepl("^Airbus A3", Aircraft))

# plot airbus
ggplot(airbus_tech_reason, aes(Manufacturer, fill=Aircraft)) +
  geom_bar(position = "dodge")+
  theme_minimal() +
  ggtitle("Airbus Crashed by Technical Failure in Each Aircraft Type") +
  scale_fill_manual(values=c('#0E2954', '#1F6E8C', '#2E8A99', '#84A7A1'))
```

Airbus Crashed by Technical Failure in Each Aircraft Type



For Boeing, The highest accident aircraft type is Boeing 707.  
For Airbus, The highest accident aircraft type is Airbus A300.