Airlines Safety Visualizaton

Shaik

2022-09-21

# Reading Data

#read data using read.csv  
airline\_data <- read.csv(("https://raw.githubusercontent.com/fivethirtyeight/data/master/airline-safety/airline-safety.csv"))  
#clean data (rename columns)  
head(airline\_data)

## airline avail\_seat\_km\_per\_week incidents\_85\_99  
## 1 Aer Lingus 320906734 2  
## 2 Aeroflot\* 1197672318 76  
## 3 Aerolineas Argentinas 385803648 6  
## 4 Aeromexico\* 596871813 3  
## 5 Air Canada 1865253802 2  
## 6 Air France 3004002661 14  
## fatal\_accidents\_85\_99 fatalities\_85\_99 incidents\_00\_14 fatal\_accidents\_00\_14  
## 1 0 0 0 0  
## 2 14 128 6 1  
## 3 0 0 1 0  
## 4 1 64 5 0  
## 5 0 0 2 0  
## 6 4 79 6 2  
## fatalities\_00\_14  
## 1 0  
## 2 88  
## 3 0  
## 4 0  
## 5 0  
## 6 337

# Data Structure Checking and Data Cleaning

#Checking Structures  
str(airline\_data)

## 'data.frame': 56 obs. of 8 variables:  
## $ airline : chr "Aer Lingus" "Aeroflot\*" "Aerolineas Argentinas" "Aeromexico\*" ...  
## $ avail\_seat\_km\_per\_week: num 3.21e+08 1.20e+09 3.86e+08 5.97e+08 1.87e+09 ...  
## $ incidents\_85\_99 : int 2 76 6 3 2 14 2 3 5 7 ...  
## $ fatal\_accidents\_85\_99 : int 0 14 0 1 0 4 1 0 0 2 ...  
## $ fatalities\_85\_99 : int 0 128 0 64 0 79 329 0 0 50 ...  
## $ incidents\_00\_14 : int 0 6 1 5 2 6 4 5 5 4 ...  
## $ fatal\_accidents\_00\_14 : int 0 1 0 0 0 2 1 1 1 0 ...  
## $ fatalities\_00\_14 : int 0 88 0 0 0 337 158 7 88 0 ...

#Loading tidyverse package   
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --  
## v ggplot2 3.3.6 v purrr 0.3.4   
## v tibble 3.1.8 v dplyr 1.0.10  
## v tidyr 1.2.1 v stringr 1.4.1   
## v readr 2.1.2 v forcats 0.5.2   
## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

#Renaming Variables  
airline\_data <- airline\_data %>%  
 rename("incidents(1985-1999)"= "incidents\_85\_99",  
 "fatal\_accidents(1985-1999)" = "fatal\_accidents\_85\_99",  
 "fatalities(1985-1999)" = "fatalities\_85\_99",  
 "incidents(2000-2014)" = "incidents\_00\_14",  
 "fatal\_accidents(2000-2014)" = "fatal\_accidents\_00\_14",  
 "fatalities(2000-2014)" = "fatalities\_00\_14")  
  
#Variable Selection  
Summary\_data <- subset(airline\_data, SELECT = -C(incidents(2000-2014), fatal\_accidents(2000-2014),fatalities(2000-2014)))   
Summary\_data

## airline avail\_seat\_km\_per\_week incidents(1985-1999)  
## 1 Aer Lingus 320906734 2  
## 2 Aeroflot\* 1197672318 76  
## 3 Aerolineas Argentinas 385803648 6  
## 4 Aeromexico\* 596871813 3  
## 5 Air Canada 1865253802 2  
## 6 Air France 3004002661 14  
## 7 Air India\* 869253552 2  
## 8 Air New Zealand\* 710174817 3  
## 9 Alaska Airlines\* 965346773 5  
## 10 Alitalia 698012498 7  
## 11 All Nippon Airways 1841234177 3  
## 12 American\* 5228357340 21  
## 13 Austrian Airlines 358239823 1  
## 14 Avianca 396922563 5  
## 15 British Airways\* 3179760952 4  
## 16 Cathay Pacific\* 2582459303 0  
## 17 China Airlines 813216487 12  
## 18 Condor 417982610 2  
## 19 COPA 550491507 3  
## 20 Delta / Northwest\* 6525658894 24  
## 21 Egyptair 557699891 8  
## 22 El Al 335448023 1  
## 23 Ethiopian Airlines 488560643 25  
## 24 Finnair 506464950 1  
## 25 Garuda Indonesia 613356665 10  
## 26 Gulf Air 301379762 1  
## 27 Hawaiian Airlines 493877795 0  
## 28 Iberia 1173203126 4  
## 29 Japan Airlines 1574217531 3  
## 30 Kenya Airways 277414794 2  
## 31 KLM\* 1874561773 7  
## 32 Korean Air 1734522605 12  
## 33 LAN Airlines 1001965891 3  
## 34 Lufthansa\* 3426529504 6  
## 35 Malaysia Airlines 1039171244 3  
## 36 Pakistan International 348563137 8  
## 37 Philippine Airlines 413007158 7  
## 38 Qantas\* 1917428984 1  
## 39 Royal Air Maroc 295705339 5  
## 40 SAS\* 682971852 5  
## 41 Saudi Arabian 859673901 7  
## 42 Singapore Airlines 2376857805 2  
## 43 South African 651502442 2  
## 44 Southwest Airlines 3276525770 1  
## 45 Sri Lankan / AirLanka 325582976 2  
## 46 SWISS\* 792601299 2  
## 47 TACA 259373346 3  
## 48 TAM 1509195646 8  
## 49 TAP - Air Portugal 619130754 0  
## 50 Thai Airways 1702802250 8  
## 51 Turkish Airlines 1946098294 8  
## 52 United / Continental\* 7139291291 19  
## 53 US Airways / America West\* 2455687887 16  
## 54 Vietnam Airlines 625084918 7  
## 55 Virgin Atlantic 1005248585 1  
## 56 Xiamen Airlines 430462962 9  
## fatal\_accidents(1985-1999) fatalities(1985-1999) incidents(2000-2014)  
## 1 0 0 0  
## 2 14 128 6  
## 3 0 0 1  
## 4 1 64 5  
## 5 0 0 2  
## 6 4 79 6  
## 7 1 329 4  
## 8 0 0 5  
## 9 0 0 5  
## 10 2 50 4  
## 11 1 1 7  
## 12 5 101 17  
## 13 0 0 1  
## 14 3 323 0  
## 15 0 0 6  
## 16 0 0 2  
## 17 6 535 2  
## 18 1 16 0  
## 19 1 47 0  
## 20 12 407 24  
## 21 3 282 4  
## 22 1 4 1  
## 23 5 167 5  
## 24 0 0 0  
## 25 3 260 4  
## 26 0 0 3  
## 27 0 0 1  
## 28 1 148 5  
## 29 1 520 0  
## 30 0 0 2  
## 31 1 3 1  
## 32 5 425 1  
## 33 2 21 0  
## 34 1 2 3  
## 35 1 34 3  
## 36 3 234 10  
## 37 4 74 2  
## 38 0 0 5  
## 39 3 51 3  
## 40 0 0 6  
## 41 2 313 11  
## 42 2 6 2  
## 43 1 159 1  
## 44 0 0 8  
## 45 1 14 4  
## 46 1 229 3  
## 47 1 3 1  
## 48 3 98 7  
## 49 0 0 0  
## 50 4 308 2  
## 51 3 64 8  
## 52 8 319 14  
## 53 7 224 11  
## 54 3 171 1  
## 55 0 0 0  
## 56 1 82 2  
## fatal\_accidents(2000-2014) fatalities(2000-2014)  
## 1 0 0  
## 2 1 88  
## 3 0 0  
## 4 0 0  
## 5 0 0  
## 6 2 337  
## 7 1 158  
## 8 1 7  
## 9 1 88  
## 10 0 0  
## 11 0 0  
## 12 3 416  
## 13 0 0  
## 14 0 0  
## 15 0 0  
## 16 0 0  
## 17 1 225  
## 18 0 0  
## 19 0 0  
## 20 2 51  
## 21 1 14  
## 22 0 0  
## 23 2 92  
## 24 0 0  
## 25 2 22  
## 26 1 143  
## 27 0 0  
## 28 0 0  
## 29 0 0  
## 30 2 283  
## 31 0 0  
## 32 0 0  
## 33 0 0  
## 34 0 0  
## 35 2 537  
## 36 2 46  
## 37 1 1  
## 38 0 0  
## 39 0 0  
## 40 1 110  
## 41 0 0  
## 42 1 83  
## 43 0 0  
## 44 0 0  
## 45 0 0  
## 46 0 0  
## 47 1 3  
## 48 2 188  
## 49 0 0  
## 50 1 1  
## 51 2 84  
## 52 2 109  
## 53 2 23  
## 54 0 0  
## 55 0 0  
## 56 0 0

# Data Summary

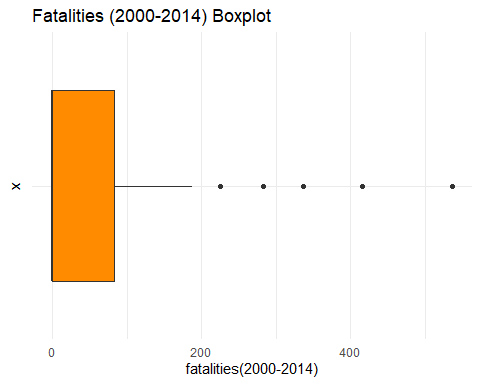
summary(Summary\_data)

## airline avail\_seat\_km\_per\_week incidents(1985-1999)  
## Length:56 Min. :2.594e+08 Min. : 0.000   
## Class :character 1st Qu.:4.740e+08 1st Qu.: 2.000   
## Mode :character Median :8.029e+08 Median : 4.000   
## Mean :1.385e+09 Mean : 7.179   
## 3rd Qu.:1.847e+09 3rd Qu.: 8.000   
## Max. :7.139e+09 Max. :76.000   
## fatal\_accidents(1985-1999) fatalities(1985-1999) incidents(2000-2014)  
## Min. : 0.000 Min. : 0.0 Min. : 0.000   
## 1st Qu.: 0.000 1st Qu.: 0.0 1st Qu.: 1.000   
## Median : 1.000 Median : 48.5 Median : 3.000   
## Mean : 2.179 Mean :112.4 Mean : 4.125   
## 3rd Qu.: 3.000 3rd Qu.:184.2 3rd Qu.: 5.250   
## Max. :14.000 Max. :535.0 Max. :24.000   
## fatal\_accidents(2000-2014) fatalities(2000-2014)  
## Min. :0.0000 Min. : 0.00   
## 1st Qu.:0.0000 1st Qu.: 0.00   
## Median :0.0000 Median : 0.00   
## Mean :0.6607 Mean : 55.52   
## 3rd Qu.:1.0000 3rd Qu.: 83.25   
## Max. :3.0000 Max. :537.00

# Data Visualization 1: Boxplot

The first visualization is a boxplot for the fatalities from 2000 to 2014 plotted below. The boxplot indicates that there are outliers in the number of fatalities with the distribution in the fatalities being skewed to the right.

ggplot(Summary\_data) +  
 aes(x = "", y = `fatalities(2000-2014)`) +  
 geom\_boxplot(fill = "#FF8C00") +  
 labs(title = "Fatalities (2000-2014) Boxplot") +  
 coord\_flip() +  
 theme\_minimal()



# Data Visualization 1: Scatterplot

The second visualization is a scatterplot for the incidents and fatalities from 2000 to 2014 plotted below. The scatterplot indicates that as the number of incidents increase the number fatalities also increase.

ggplot(Summary\_data) +  
 aes(x = `incidents(2000-2014)`, y = `fatalities(2000-2014)`) +  
 geom\_point(shape = "circle", size = 1.5, colour = "#112446") +  
 labs(  
 title = "Incidents and Fatalities 2000-2014 Scatterplot"  
 ) +  
 theme\_minimal()

