Assignment 2

Iryna Barvi

4/3/2022

# Section 1: Description of the data. (10 points)  
# The dataset I am using is the data behind the story Should Travelers Avoid Flying Airlines That Have Had Crashes in the Past?. The dataset contains information about airlines, the available seat kilometers flown every week, and the number of incidents and fatalities. This data was collected by the Aviation Safety Network between 1985 and 2014. From this dataset, I want to find the correlation between fatalities and the airline involved. This is .csv file with comma delimiters.

# Section 2: Reading the data into R. (5 points)  
airlines <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/airline-safety/airline-safety.csv"  
airline\_safety <- read.csv(airlines)

# Section 3: Clean the data. (10 points)   
  
colnames(airline\_safety)

## [1] "airline" "avail\_seat\_km\_per\_week" "incidents\_85\_99"   
## [4] "fatal\_accidents\_85\_99" "fatalities\_85\_99" "incidents\_00\_14"   
## [7] "fatal\_accidents\_00\_14" "fatalities\_00\_14"

f1 <- rename(airline\_safety, Airline = airline)  
f2 <- rename(f1, Available\_Seat\_Km\_Per\_Week = avail\_seat\_km\_per\_week)  
f3 <- rename(f2, Incidents\_1985\_1999 = incidents\_85\_99)  
f4 <- rename(f3, Fatal\_Accidents\_1985\_1999 = fatal\_accidents\_85\_99)  
f5 <- rename(f4, Fatalities\_1985\_1999 = fatalities\_85\_99)  
f6 <- rename(f5, Incidents\_2000\_2012 = incidents\_00\_14)  
f7 <- rename(f6, Fatal\_Accidents\_2000\_2012 = fatal\_accidents\_00\_14)  
f8 <- rename(f7, Fatalities\_2000\_2012 = fatalities\_00\_14)  
colnames(f8)

## [1] "Airline" "Available\_Seat\_Km\_Per\_Week"  
## [3] "Incidents\_1985\_1999" "Fatal\_Accidents\_1985\_1999"   
## [5] "Fatalities\_1985\_1999" "Incidents\_2000\_2012"   
## [7] "Fatal\_Accidents\_2000\_2012" "Fatalities\_2000\_2012"

# Section 4: Characteristics of the data. (20 points)

This dataframe has 56 rows and 8 columns. The names of the columns and a brief description of each are in the table below:

airline\_table <- matrix(c("Airline","Airline (asterisk indicates that regional subsidiaries are included)","Available Seat Km Per Week","Available seat kilometers flown every week","Incidents 1985 1999","Total number of incidents, 1985–1999","Fatal Accidents 1985 1999","Total number of fatal accidents, 1985–1999","Fatalities 1985 1999","Total number of fatalities, 1985–1999","Incidents 2000 2012","Total number of incidents, 2000–2014","Fatal Accidents 2000 2012","Total number of fatal accidents, 2000–2014","Fatalities 2000 2012","Total number of fatalities, 2000–2014"),ncol = 2, byrow = TRUE)  
colnames(airline\_table) <- c('Header','Definition')  
airline\_table <-as.table(airline\_table)  
airline\_table

## Header   
## A Airline   
## B Available Seat Km Per Week  
## C Incidents 1985 1999   
## D Fatal Accidents 1985 1999   
## E Fatalities 1985 1999   
## F Incidents 2000 2012   
## G Fatal Accidents 2000 2012   
## H Fatalities 2000 2012   
## Definition   
## A Airline (asterisk indicates that regional subsidiaries are included)  
## B Available seat kilometers flown every week   
## C Total number of incidents, 1985–1999   
## D Total number of fatal accidents, 1985–1999   
## E Total number of fatalities, 1985–1999   
## F Total number of incidents, 2000–2014   
## G Total number of fatal accidents, 2000–2014   
## H Total number of fatalities, 2000–2014

# Section 5: Summary statistics. (20 points)  
airline\_four <- subset(f8, select=c(Airline,Available\_Seat\_Km\_Per\_Week,Fatalities\_2000\_2012))  
airline\_summary <- summary(airline\_four)  
print(airline\_summary)

## Airline Available\_Seat\_Km\_Per\_Week Fatalities\_2000\_2012  
## Length:56 Min. :2.594e+08 Min. : 0.00   
## Class :character 1st Qu.:4.740e+08 1st Qu.: 0.00   
## Mode :character Median :8.029e+08 Median : 0.00   
## Mean :1.385e+09 Mean : 55.52   
## 3rd Qu.:1.847e+09 3rd Qu.: 83.25   
## Max. :7.139e+09 Max. :537.00