# Homework 4

## Q1) Importing CSV file

Ans:

dataset = read.csv("/Users/Raja/Desktop/FAA\_r.csv", header = TRUE,sep=",")

Returns the dataset of type dataframe.

To find the instruction on how to use the functions

We use

>?read.scv - to find the instrunctions

>??read.csv – lookup for finding matching function name

**Q2. How many variables in the data set ? what are their names?**

Ans:

*length(colnames(dataset))*

*7*

length of colnames gives variables in the data set

*colnames(dataset)*

*"type" "duration" "no\_psng" "speed\_ground" "speed\_air" "height" "pitch"*

colnames prints the names of those column names

**Q3.How many Observations in total? How many observations for Airbus?**

*str(dataset)*

*'data.frame': 800 obs. of 7 variables:*

*$ type : Factor w/ 2 levels "Airbus","Boeing": 1 2 1 2 1 2 1 2 1 1 ...*

*$ duration : num 121 152 131 273 168 ...*

*$ no\_psng : int 58 68 45 69 66 60 64 64 64 65 ...*

*$ speed\_ground: num 85.3 73.2 112.1 57.1 80.3 ...*

*$ speed\_air : num NA NA 111 NA NA ...*

*$ height : num 26.6 14.4 18.1 44.5 37.9 ...*

*$ pitch : num 3.65 3.89 4.01 4.03 4.33 ...*

str gives the no. of observations and no. of variables and their details

there are 800 observations in the given dataset.

*length(dat[dat$type=='Airbus',1])*

400

there are 400 observations for Airbus

**4. Calculate the mean for each of the flight parameters (measures). Please also report the corresponding standard deviation.**

Ans:

**Duration:**

> mean(dat$duration)

[1] 149.174

> sd(dat$duration)

[1] 50.38649

**No\_psng:**

> mean(dat$no\_psng)

[1] 60.04

> sd(dat$no\_psng)

[1] 7.840615

**speed\_ground**

> mean(dat$speed\_ground)

[1] 79.35776

> sd(dat$speed\_ground)

[1] 20.19064

**speed\_air**

speed\_air has some NA values, used na.rm=t to remove not availables

> mean(dat$ speed\_air,na.rm=T)

[1] 102.0379

> sd(dat$ speed\_air,na.rm=T)

[1] 10.3253

**height**

> mean(dat$height)

[1] 29.75949

> sd(dat$height)

[1] 10.04644

**pitch**

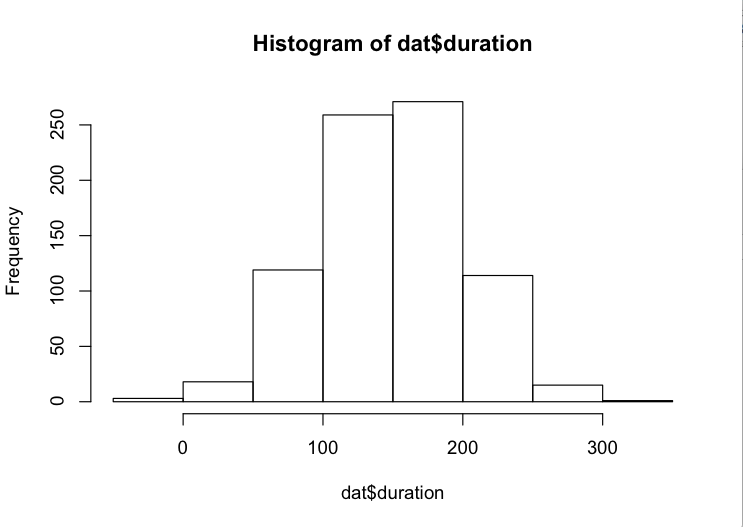
> mean(dat$pitch)

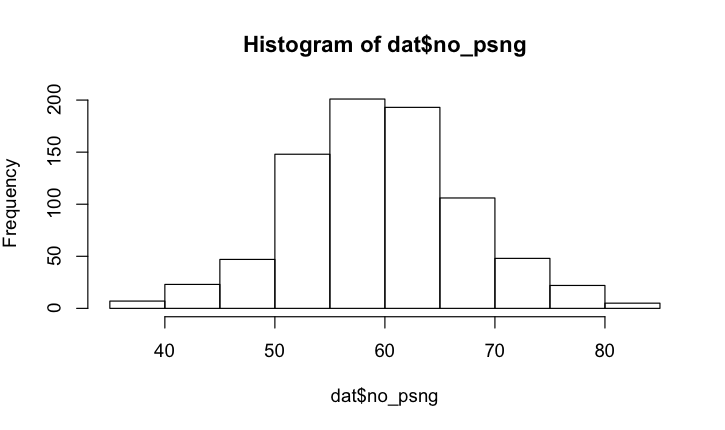
[1] 4.170417

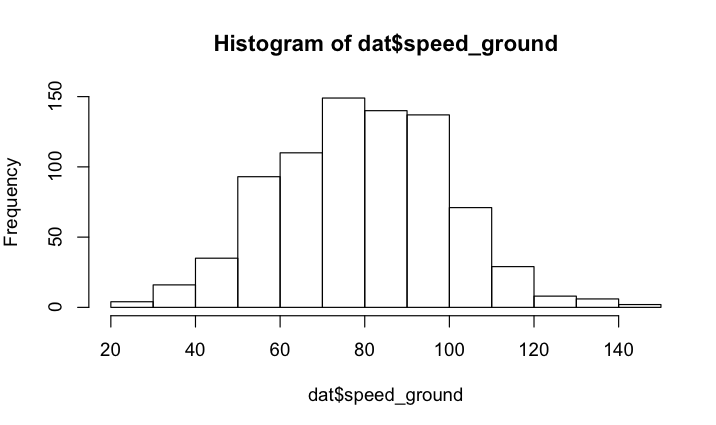
> sd(dat$pitch)

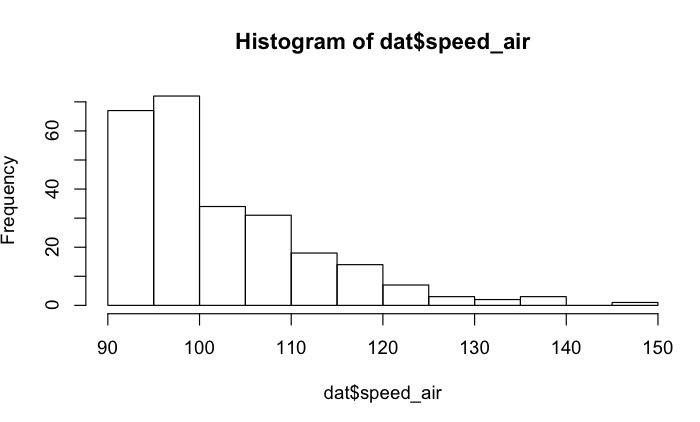
[1] 0.49442

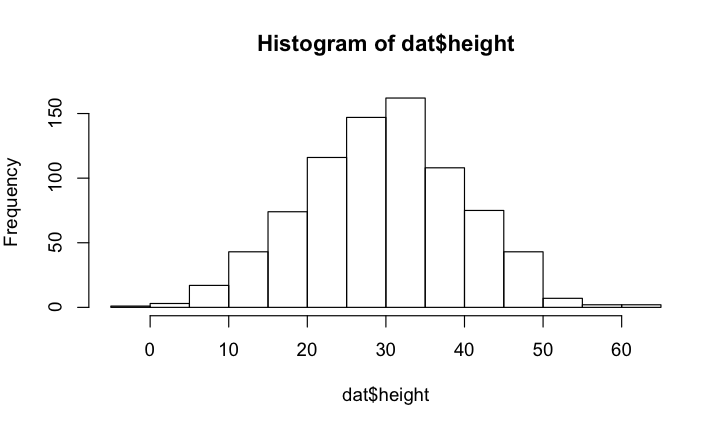
**5. Histograms**

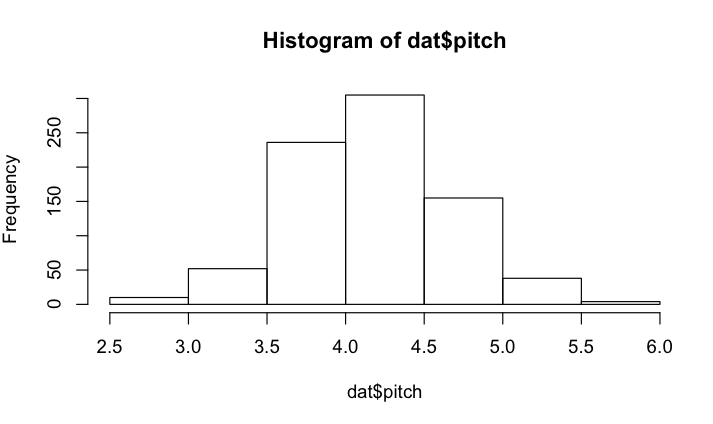












**6.Missing Values**

summary(dat)

|  |
| --- |
| type duration no\_psng speed\_ground speed\_air height pitch |
| Airbus:400 Min. :-21.39 Min. :37.00 Min. : 23.40 Min. : 90.13 Min. :-1.366 Min. :2.654 |
| Boeing:400 1st Qu.:115.80 1st Qu.:55.00 1st Qu.: 65.55 1st Qu.: 94.72 1st Qu.:23.143 1st Qu.:3.853 |
| Median :150.40 Median :60.00 Median : 79.60 Median : 98.36 Median :29.944 Median :4.151 |
| Mean :149.17 Mean :60.04 Mean : 79.36 Mean :102.04 Mean :29.759 Mean :4.170 |
| 3rd Qu.:182.81 3rd Qu.:65.00 3rd Qu.: 93.61 3rd Qu.:107.42 3rd Qu.:36.367 3rd Qu.:4.494 |
| Max. :314.35 Max. :84.00 Max. :149.42 Max. :148.73 Max. :61.206 Max. :5.911 |
| NA's :548 |

From the above table, it shows that there 548 not available values for speed\_air. To explore the data, We use summary to display the results.

is.na(dat$speed\_air) gives the missing values as True and remaining false.

**Q7**

Ans: speed of the air flow (defined as the difference between speed\_ground and speed\_air).

>airflow=(dat$speed\_air-speed\_ground);

>mean(airflow,na.rm=T)

[1] 0.08395205

**Q8)**

Ans:

sum(dat$duration<40)

[1] 9

totally there are 9 flights whose duration less than 40.

dat[dat$duration<40,]

type duration no\_psng speed\_ground speed\_air height pitch

26 Boeing -3.630527 57 69.70764 NA 32.184000 3.737644

130 Airbus -21.389093 57 75.04745 NA 24.644355 3.635539

318 Airbus 18.017157 53 98.20148 98.42889 27.065475 3.962545

363 Boeing -1.225362 62 83.25273 NA 7.218443 3.898639

364 Airbus 32.559910 60 58.54371 NA 35.432061 4.215479

377 Boeing 9.590482 70 97.80262 97.57404 45.108303 3.969965

598 Airbus 24.525263 52 70.48631 NA 23.919686 4.440839

640 Airbus 16.582249 60 106.29541 104.96263 40.915652 4.784315

725 Airbus 28.487486 48 32.03438 NA 25.354042 4.487389

There are 6 flights whose heights is less than 6

sum(dat$height<6)

[1] 6

Removing observations whose duration <40 and height <6

> temp=dat[!dat$duration<40,]

> temp=temp[!temp$height<6,]

summary(temp)

type duration no\_psng speed\_ground speed\_air height pitch

Airbus:391 Min. : 40.38 Min. :37.00 Min. : 23.40 Min. : 90.13 Min. : 6.20 Min. :2.654

Boeing:394 1st Qu.:116.96 1st Qu.:55.00 1st Qu.: 65.08 1st Qu.: 94.59 1st Qu.:23.21 1st Qu.:3.852

Median :151.80 Median :60.00 Median : 79.56 Median : 98.36 Median :30.16 Median :4.152

Mean :151.00 Mean :60.08 Mean : 79.33 Mean :102.10 Mean :29.97 Mean :4.171

3rd Qu.:183.25 3rd Qu.:65.00 3rd Qu.: 93.56 3rd Qu.:107.46 3rd Qu.:36.49 3rd Qu.:4.497

Max. :314.35 Max. :84.00 Max. :149.42 Max. :148.73 Max. :61.21 Max. :5.911

NA's :539

Q9) Divide the cleaned data set (as obtained in Step 8) into two subsets: Airbus and Boeing.

>airbusdataset=temp[temp$type=="Airbus",]

>summary(airbusdataset)

type duration no\_psng speed\_ground speed\_air height pitch

Airbus:391 Min. : 40.85 Min. :37.0 Min. : 26.64 Min. : 90.17 Min. : 6.20 Min. :2.700

Boeing: 0 1st Qu.:121.43 1st Qu.:55.0 1st Qu.: 66.57 1st Qu.: 94.98 1st Qu.:22.43 1st Qu.:3.841

Median :149.23 Median :60.0 Median : 79.56 Median : 98.97 Median :29.74 Median :4.146

Mean :152.09 Mean :60.4 Mean : 79.86 Mean :102.37 Mean :29.65 Mean :4.177

3rd Qu.:184.09 3rd Qu.:65.0 3rd Qu.: 93.57 3rd Qu.:107.95 3rd Qu.:36.56 3rd Qu.:4.516

Max. :314.35 Max. :84.0 Max. :149.42 Max. :148.73 Max. :61.21 Max. :5.911

NA's :268

> boeingdataset=temp[temp$type=="Boeing",]

> summary(boeingdataset)

type duration no\_psng speed\_ground speed\_air height pitch

Airbus: 0 Min. : 40.38 Min. :38.00 Min. : 23.40 Min. : 90.13 Min. : 6.306 Min. :2.654

Boeing:394 1st Qu.:114.26 1st Qu.:55.00 1st Qu.: 63.49 1st Qu.: 93.85 1st Qu.:24.044 1st Qu.:3.873

Median :153.13 Median :60.00 Median : 79.57 Median : 98.02 Median :30.353 Median :4.162

Mean :149.91 Mean :59.76 Mean : 78.80 Mean :101.83 Mean :30.287 Mean :4.164

3rd Qu.:182.28 3rd Qu.:65.00 3rd Qu.: 93.52 3rd Qu.:106.80 3rd Qu.:36.438 3rd Qu.:4.472

Max. :278.46 Max. :83.00 Max. :142.55 Max. :139.67 Max. :58.696 Max. :5.537

NA's :271

Q10)

There are 391 observations in Airbus and 394 observations in Boeing

Mean height of Airbus and Boeing are 29.65 and 30.287

Mean pitch of Airbus and Boeing are 4.146 and 4.164

Mean speed\_ground of Airbus and Boeing are 79.56 and 79.57

Mean speed\_air of Airbus and Boeing are 102.37 and 101.83

From the above observation, there is no major difference between these two aircraft makes.