Day 3 Lab Manual

UNIVARIATE ANALYSIS IN R - MEASURES OF CENTRAL TENDENCY

Exercise:

I. ARITHMETIC MEAN

a) Write suitable R code to compute the average of the following values.

12,7,3,4.2,18,2,54,-21,8,-5

b) Compute the mean after applying the trim option and removing 3 values from eachend.

c) Compute the mean of the following vector .

(12,7,3,4.2,18,2,54,-21,8,-5,NA)

#If there are missing values, then the mean function returns NA.

Find mean dropping NA values.

#To drop the missing values from the calculation use na.rm = TRUE

SOURCE CODE

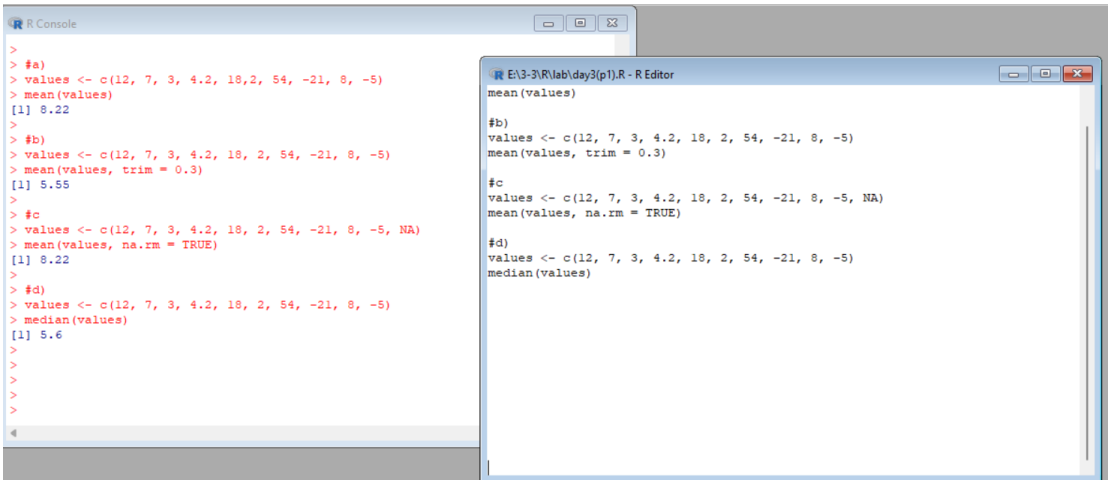
```
#a)

values <- c(12, 7, 3, 4.2, 18,2, 54, -21, 8, -5)

mean(values)

#b)

values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

mean(values, trim = 0.3)

#c

values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5, NA)

mean(values, na.rm = TRUE)
```

#d)

values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

median(values)

OUTPUT:



II.MEDIAN

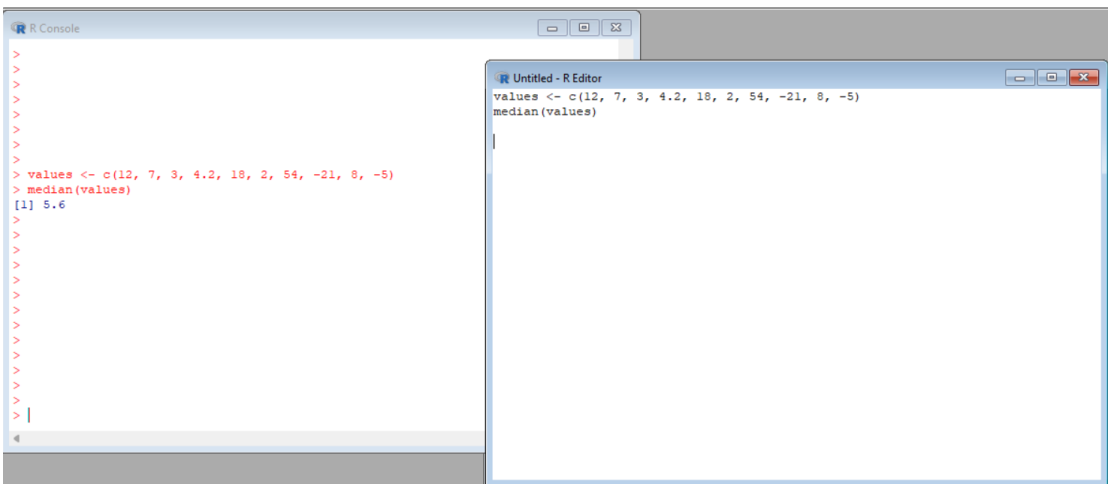
Write suitable R code to compute the median of the following values.

12,7,3,4.2,18,2,54,-21,8,-5

SOURCE CODE:

values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

median(values)



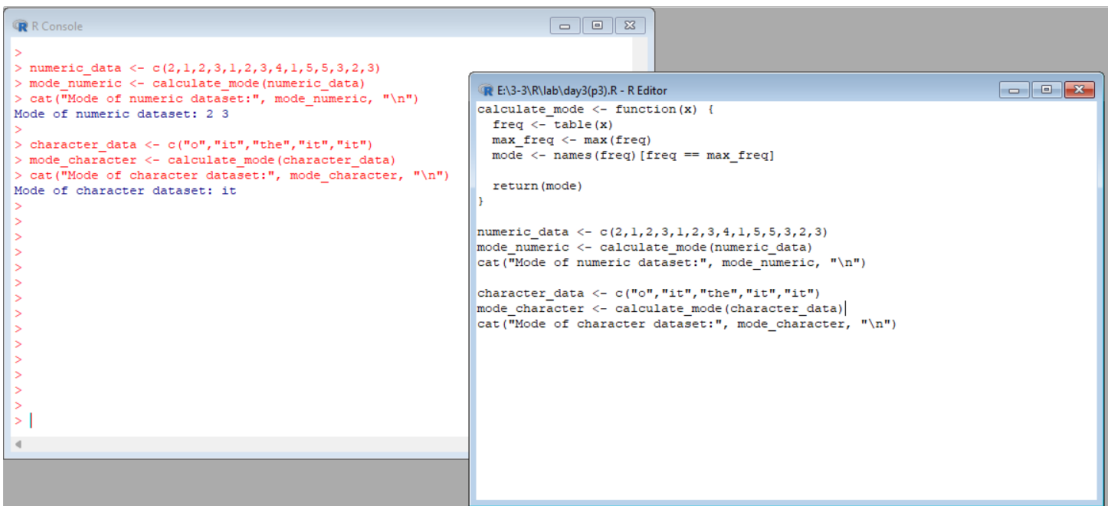
III. MODE

Calculate the mode for the following numeric as well as character data set in R.

(2,1,2,3,1,2,3,4,1,5,5,3,2,3) , ("o","it","the","it","it")

SOURCE CODE:

```
calculate_mode <- function(x) {
        freq <- table(x)
        max_freq <- max(freq)
        mode <- names(freq)[freq == max_freq]
        return(mode)
}
numeric_data <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)
mode_numeric <- calculate_mode(numeric_data)
cat("Mode of numeric dataset:", mode_numeric, "\n")
character_data <- c("o","it","the","it","it")
mode_character <- calculate_mode(character_data)
cat("Mode of character dataset:", mode_character, "\n")
```



UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION

Exercise: 4

Download mpg dataset which contains Fuel economy data from 1999 and 2008 for 38 popular models of car from the URL given below.

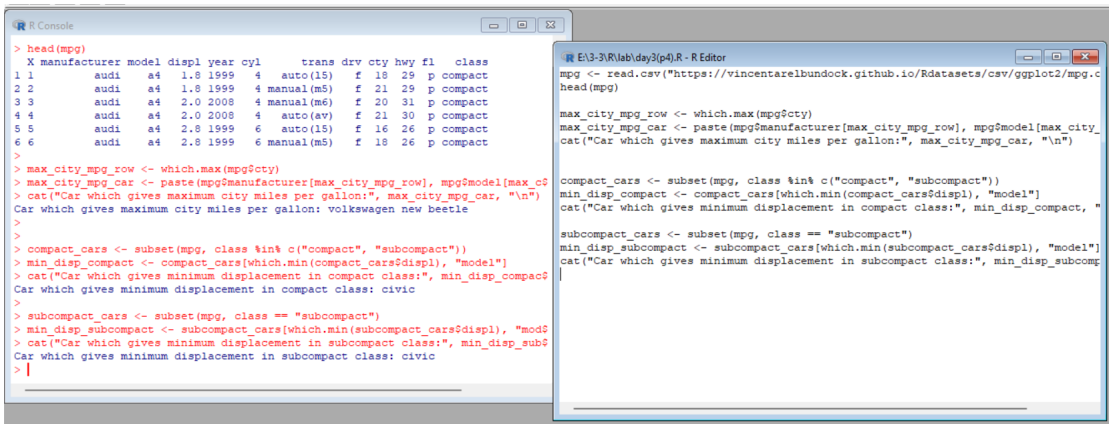
https://vincentarelbundock.github.io/Rdatasets/datasets.html

Answer the following queries

i) Find the car which gives maximum city miles per gallon
ii) Find the cars which gives minimum disp in compact and subcompact class

SOURCE CODE:

```
mpg <-
read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")

head(mpg)

max_city_mpg_row <- which.max(mpg$cty)

max_city_mpg_car <- paste(mpg$manufacturer[max_city_mpg_row],
mpg$model[max_city_mpg_row])

cat("Car which gives maximum city miles per gallon:", max_city_mpg_car, "\n")

compact_cars <- subset(mpg, class %in% c("compact", "subcompact"))

min_disp_compact <- compact_cars[which.min(compact_cars$displ), "model"]

cat("Car which gives minimum displacement in compact class:", min_disp_compact,
"\n")

subcompact_cars <- subset(mpg, class == "subcompact")

min_disp_subcompact <- subcompact_cars[which.min(subcompact_cars$displ),
"model"]

cat("Car which gives minimum displacement in subcompact class:",
min_disp_subcompact, "\n")
```



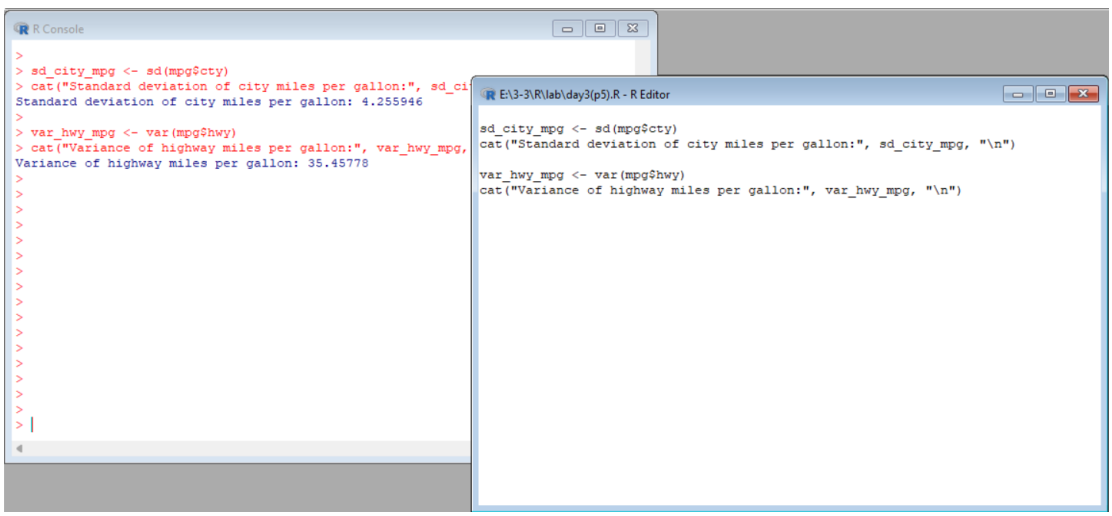
Exercise: 5

Use the same dataset as used in Exercise 4 and perform the following queries

i) Find the standard deviation of city milles per gallon ii) Find the variance of highway milles per gallon

SOURCE CODE:

```
sd_city_mpg <- sd(mpg$cty)

cat("Standard deviation of city miles per gallon:", sd_city_mpg, "\n")

var_hwy_mpg <- var(mpg$hwy)

cat("Variance of highway miles per gallon:", var_hwy_mpg, "\n")
```



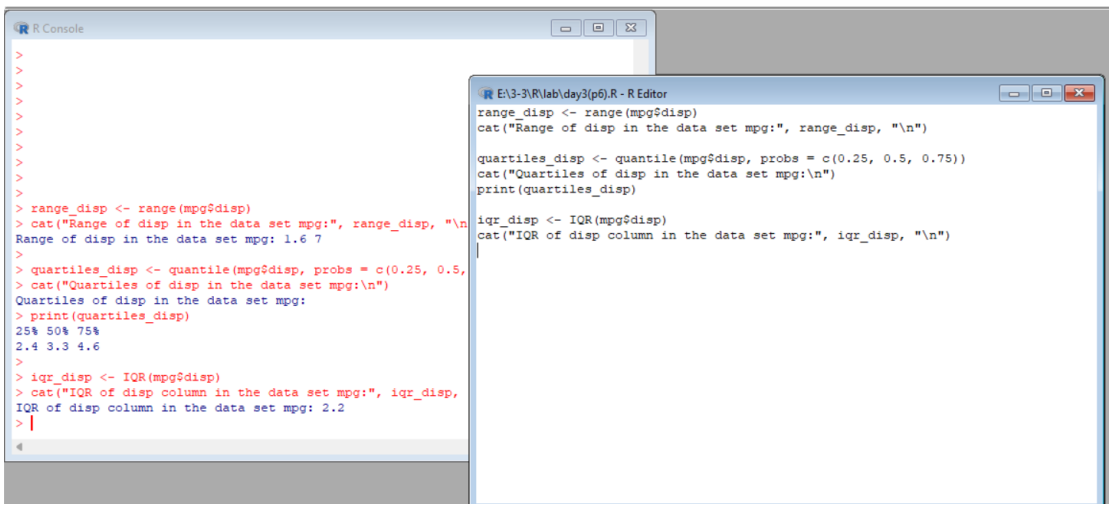
Exercise 6

Use the same dataset and perform the following queries

i) Find the range of the disp in the data set mpg
ii) Find the Quartile of the disp in the data set mpg
iii) Find the IQR of the disp column in the data set mpg

SOURCE CODE:

```
range_disp <- range(mpg$disp)

cat("Range of disp in the data set mpg:", range_disp, "\n")

quartiles_disp <- quantile(mpg$disp, probs = c(0.25, 0.5, 0.75))

cat("Quartiles of disp in the data set mpg:\n")

print(quartiles_disp)

iqr_disp <- IQR(mpg$disp)

cat("IQR of disp column in the data set mpg:", iqr_disp, "\n")
```



Exercise 7

#Install Library

library(e1071)

a. Find the skewness of city miles per mileage in the data set mpg ?

   Use qplot function and display the graph for the city miles per mileage column

b. Find the kurtosis of city miles per mileage in the data set mpg

SOURCE CODE:

```
library(e1071)

library(ggplot2)

skew_cty <- skewness(mpg$cty)

cat("Skewness of city miles per gallon in the data set mpg:", skew_cty, "\n")

qplot(mpg$cty, geom="histogram", binwidth=2, main="City Miles Per Gallon",
xlab="Miles Per Gallon")

kurt_cty <- kurtosis(mpg$cty)

cat("Kurtosis of city miles per gallon in the data set mpg:", kurt_cty, "\n")
```