**Day 3 Lab Manual**

**NAME:**BALA MANOHAR.J

**REG NO:**192124133

**UNIVARIATE ANALYSIS IN R - MEASURES OF CENTRAL TENDENCY**

**Exercise: I. ARITHMETIC MEAN**

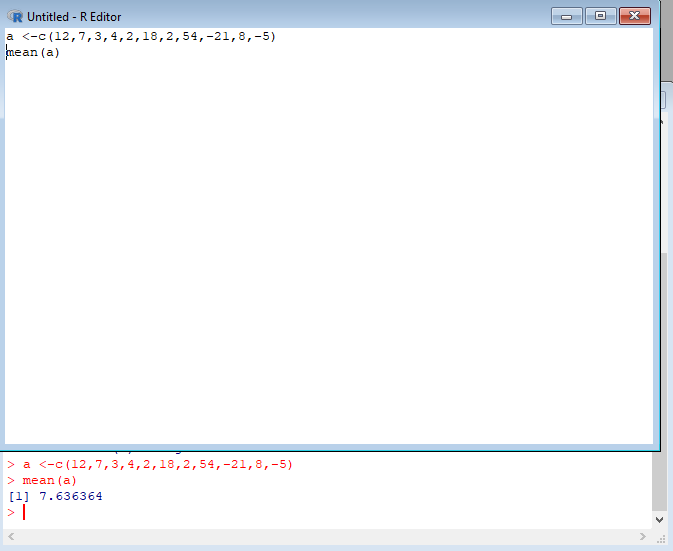
1. Write suitable R code to compute the average of the following values. 12,7,3,4.2,18,2,54,-21,8,-5

**Source Code:**

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

mean(values)

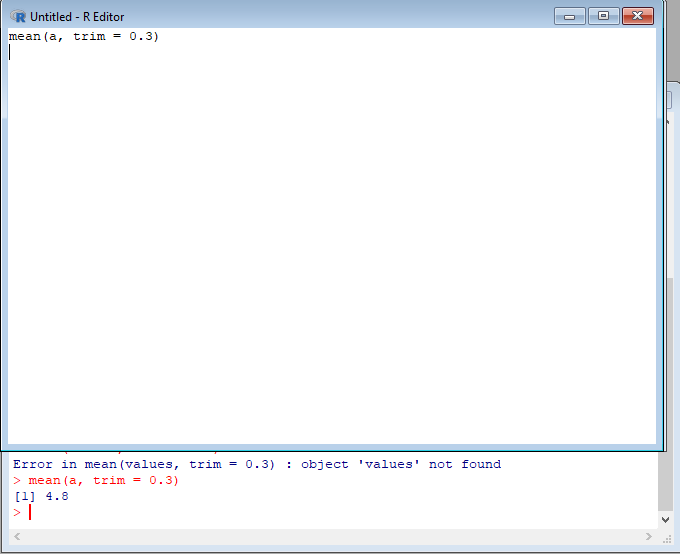
**OUTPUT:**

****

1. Compute the mean after applying the trim option and removing 3 values from each end.

**Source Code:**

mean(values, trim = 0.3)

**OUTPUT: **

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:**

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

mean(values, na.rm = TRUE)

**OUTPUT:**

[1] 6.55

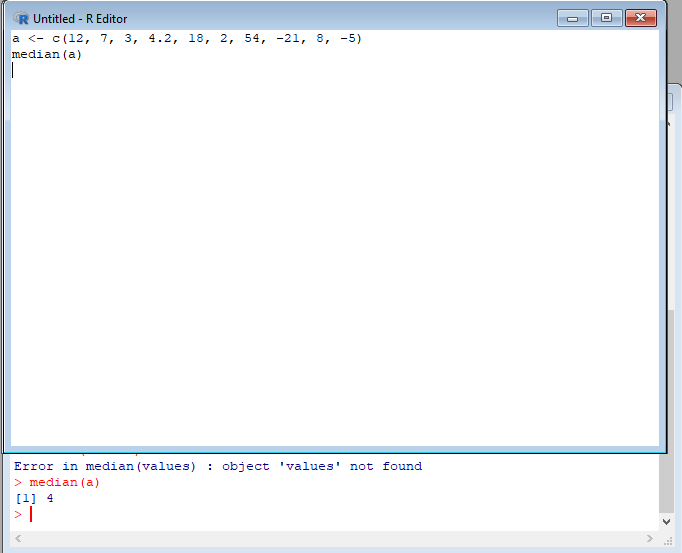
**II.MEDIAN**

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)

**OUTPUT:**

**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) , (&quot;o&quot;,&quot;it&quot;,&quot;the&quot;,&quot;it&quot;,&quot;it&quot;)

**Source Code:**

find\_mode <- function(x) {

table\_x <- table(x)

mode\_x <- names(table\_x)[which.max(table\_x)]

return(mode\_x)

}

**OUTPUT:**

# Numeric data set

numeric\_values <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

find\_mode(numeric\_values)

# Output

[1] "2"

# Character data set

character\_values <- c("o","it","the","it","it")

find\_mode(character\_values)

# Output

[1] "it"

**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:

1. **Find the car which gives maximum city miles per gallon**

**Source Code:**

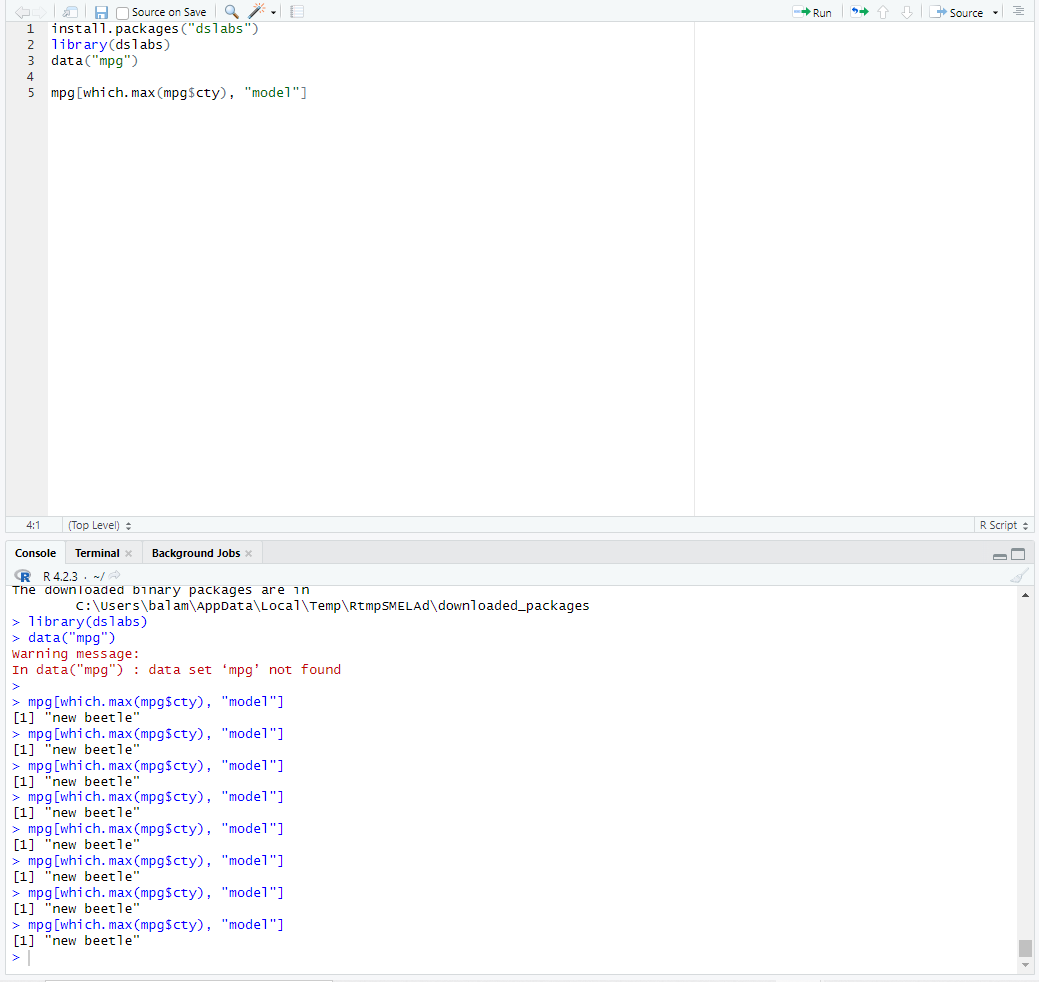
install.packages("dslabs")

library(dslabs)

data("mpg")

mpg[which.max(mpg$cty), "model"]

**OUTPUT:**

****

1. **Find the cars which gives minimum disp in compact and subcompact class**

**Source Code:**

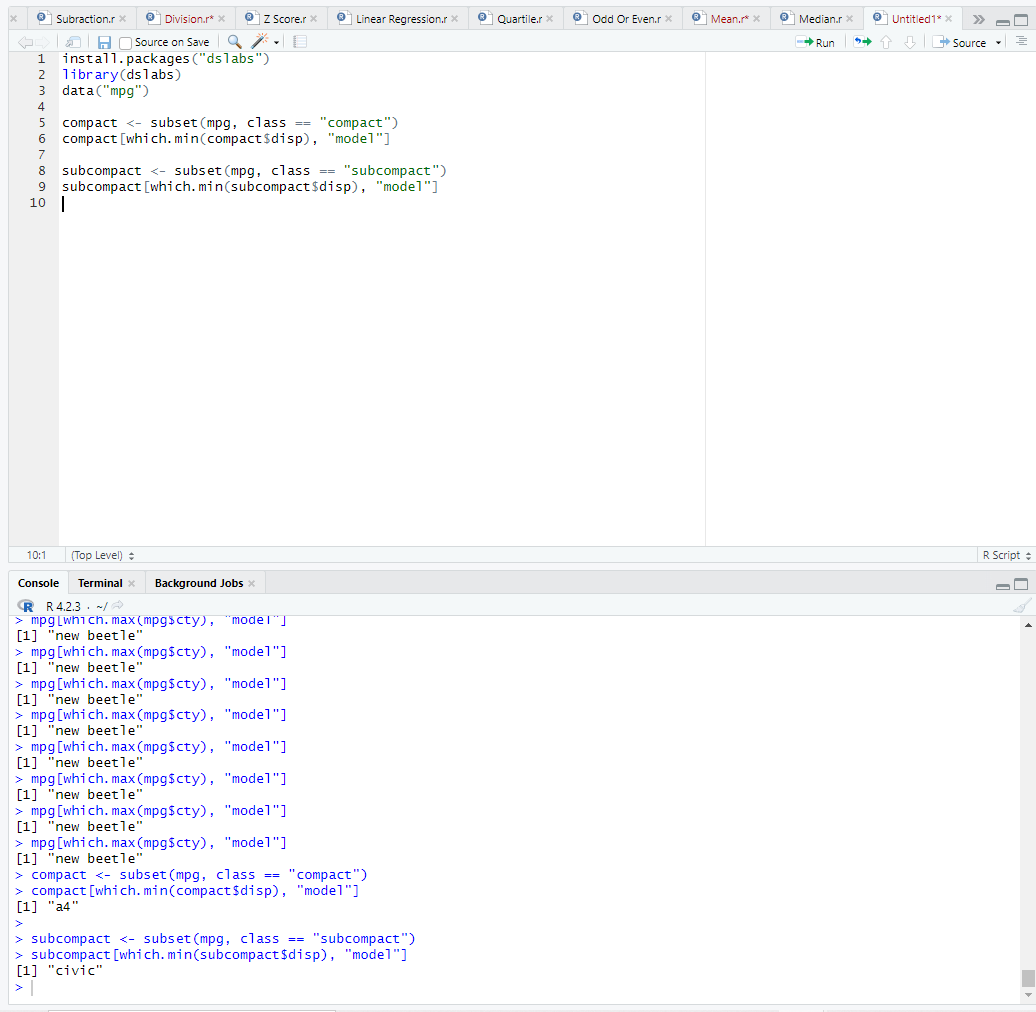
compact <- subset(mpg, class == "compact")

compact[which.min(compact$disp), "model"]

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

subcompact[which.min(subcompact$disp), "model"]

**OUTPUT:**

****

**Exercise: 5**

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

1. **Find the standard deviation of city milles per gallon**

**Source Code:**

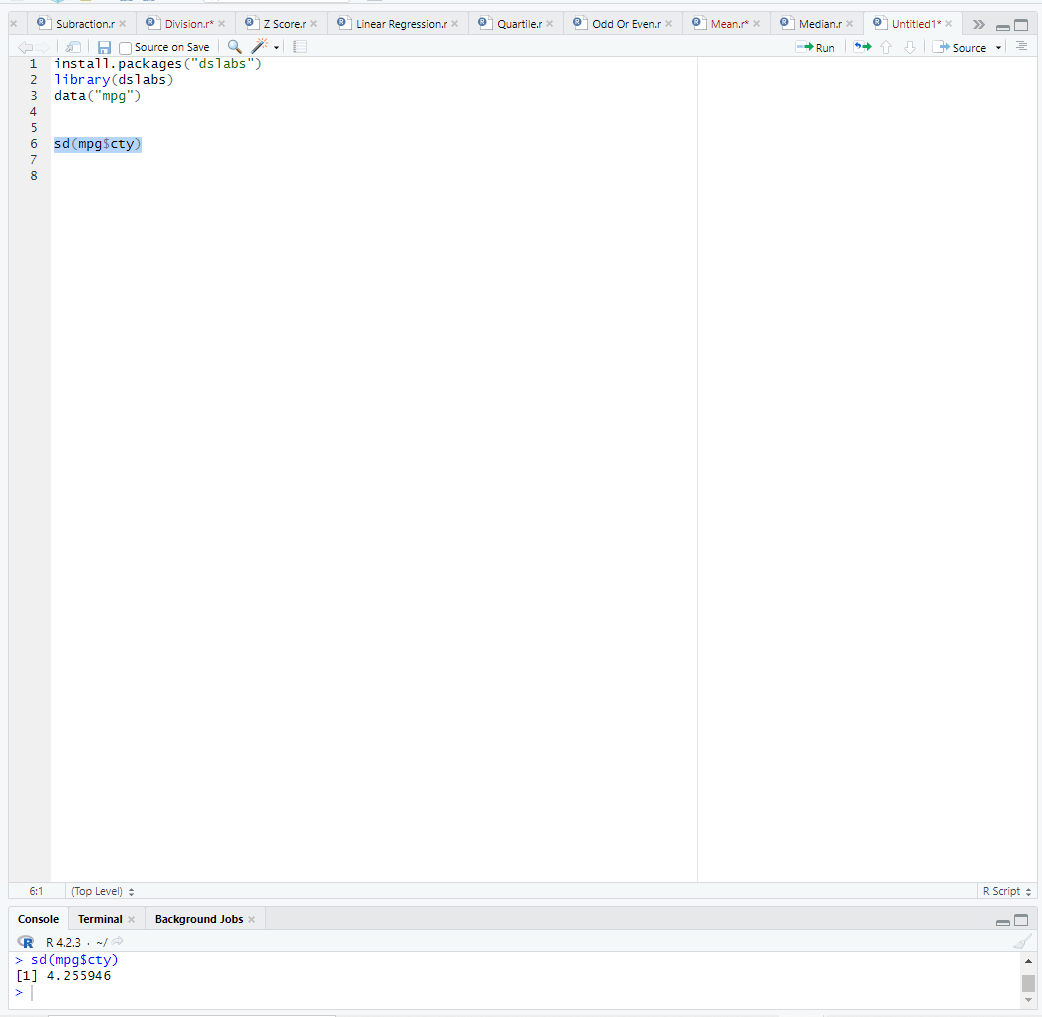
install.packages("dslabs")

library(dslabs)

data("mpg")

sd(mpg$cty)

**OUTPUT:**

****

1. **Find the variance of highway milles per gallon**

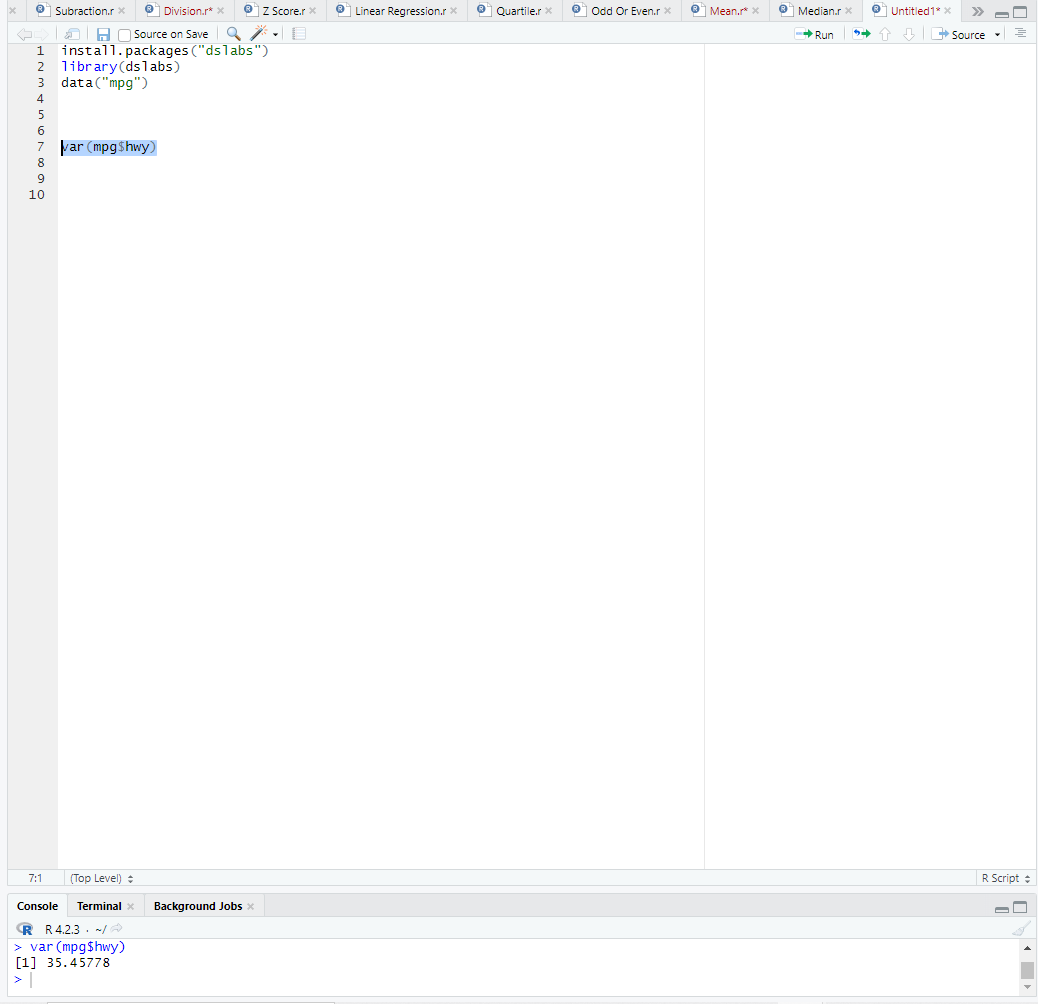
**Source Code:**

install.packages("dslabs")

library(dslabs)

data("mpg")

var(mpg$hwy)

**OUTPUT:**

**Exercise 6**

Use the same dataset and perform the following queries

1. Find the range of the disp in the data set mpg

**Source Code:**

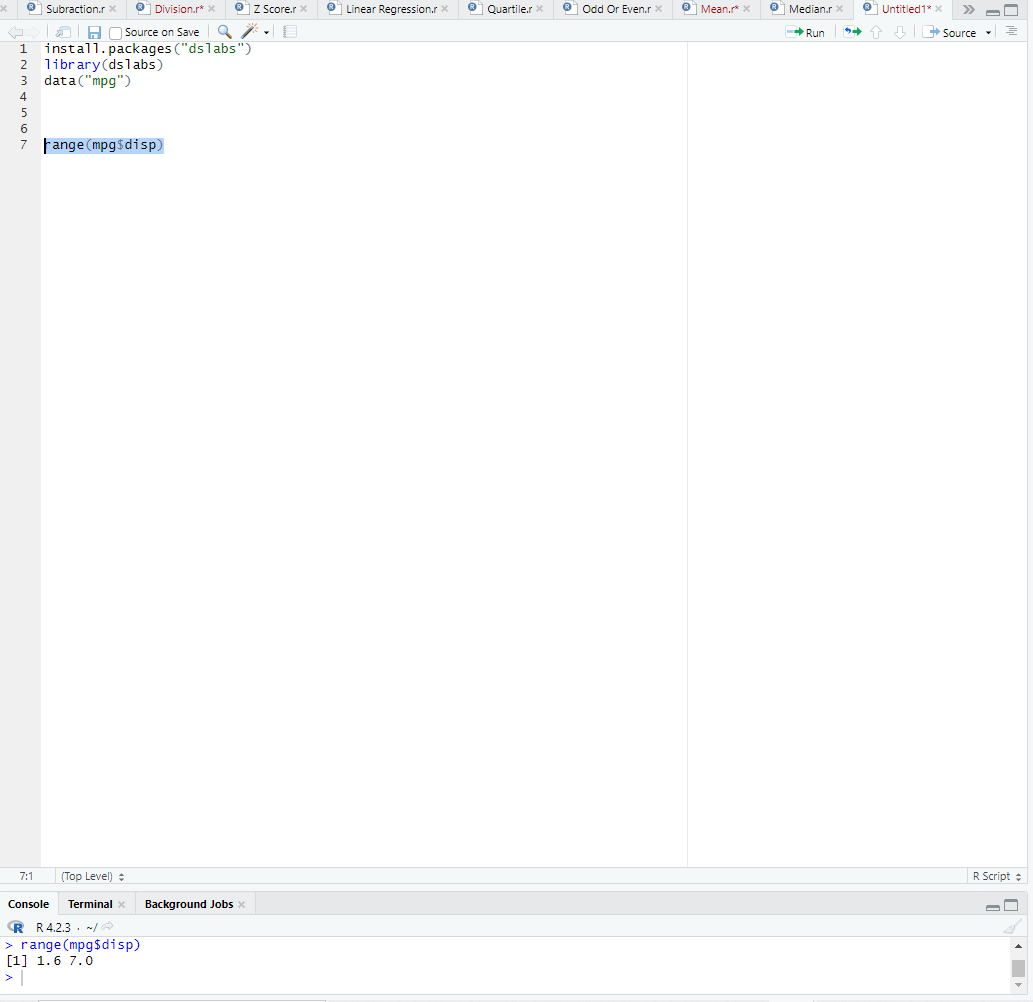
install.packages("dslabs")

library(dslabs)

data("mpg")

range(mpg$disp)

**OUTPUT:**

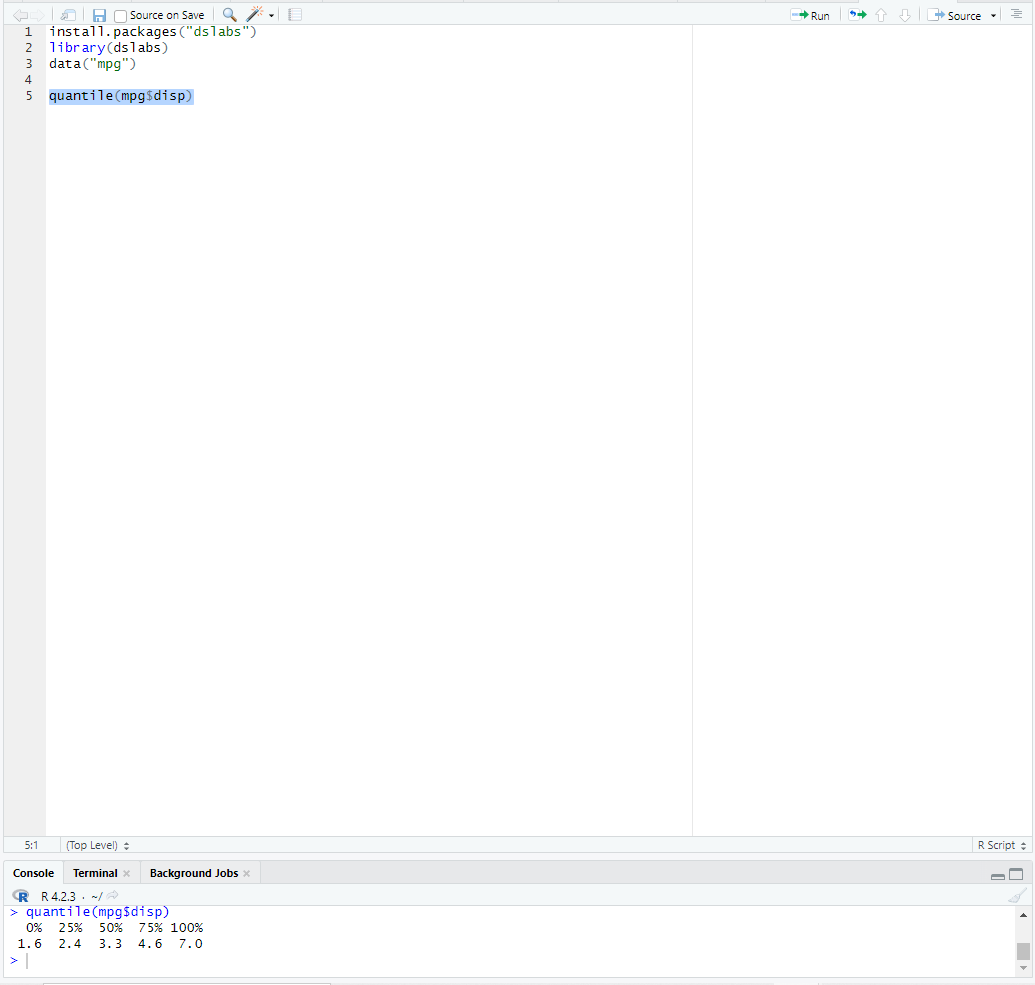
****

1. **Find the Quartile of the disp in the data set mpg**

**Source Code:**

quantile(mpg$disp)

**OUTPUT:**



**iii) Find the IQR of the disp column in the data set mpg**

**Source Code:**

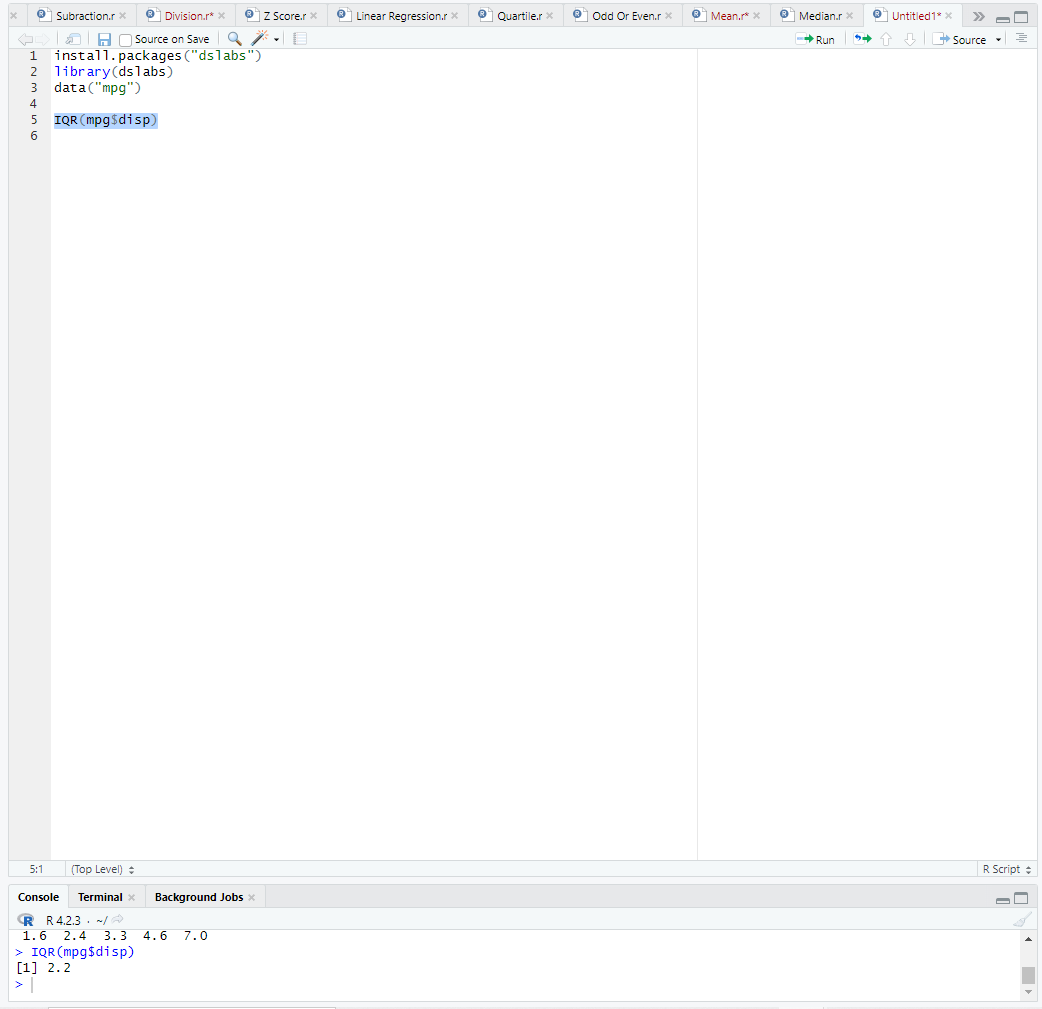
install.packages("dslabs")

library(dslabs)

data("mpg")

IQR(mpg$disp)

**OUTPUT:**

****

**Source Code:**

install.packages("dslabs")

library(dslabs)

data("mpg")

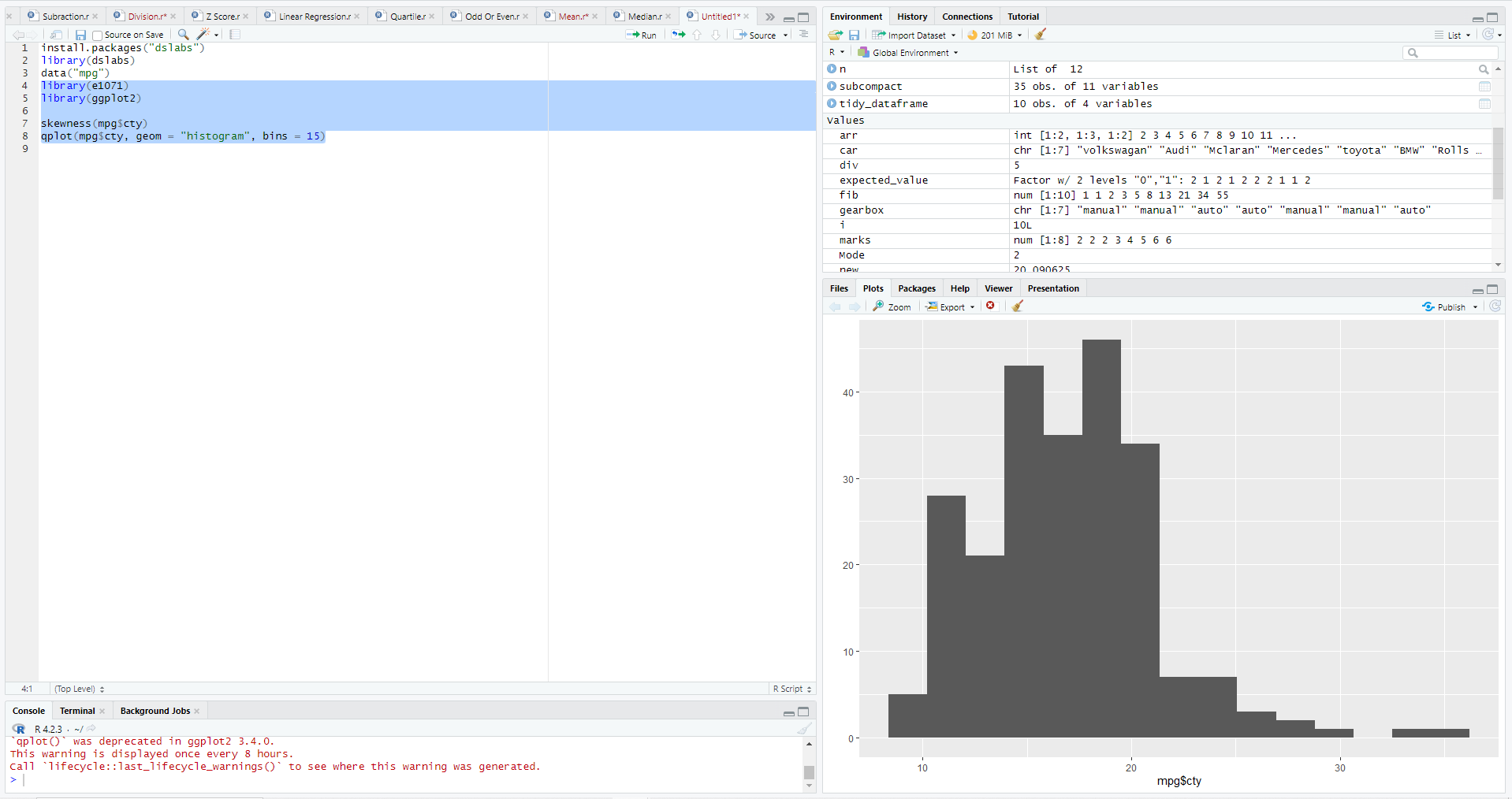
library(e1071)

library(ggplot2)

skewness(mpg$cty)

qplot(mpg$cty, geom = "histogram", bins = 15)

**OUTPUT:**

****

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

**Source Code:**

install.packages("dslabs")

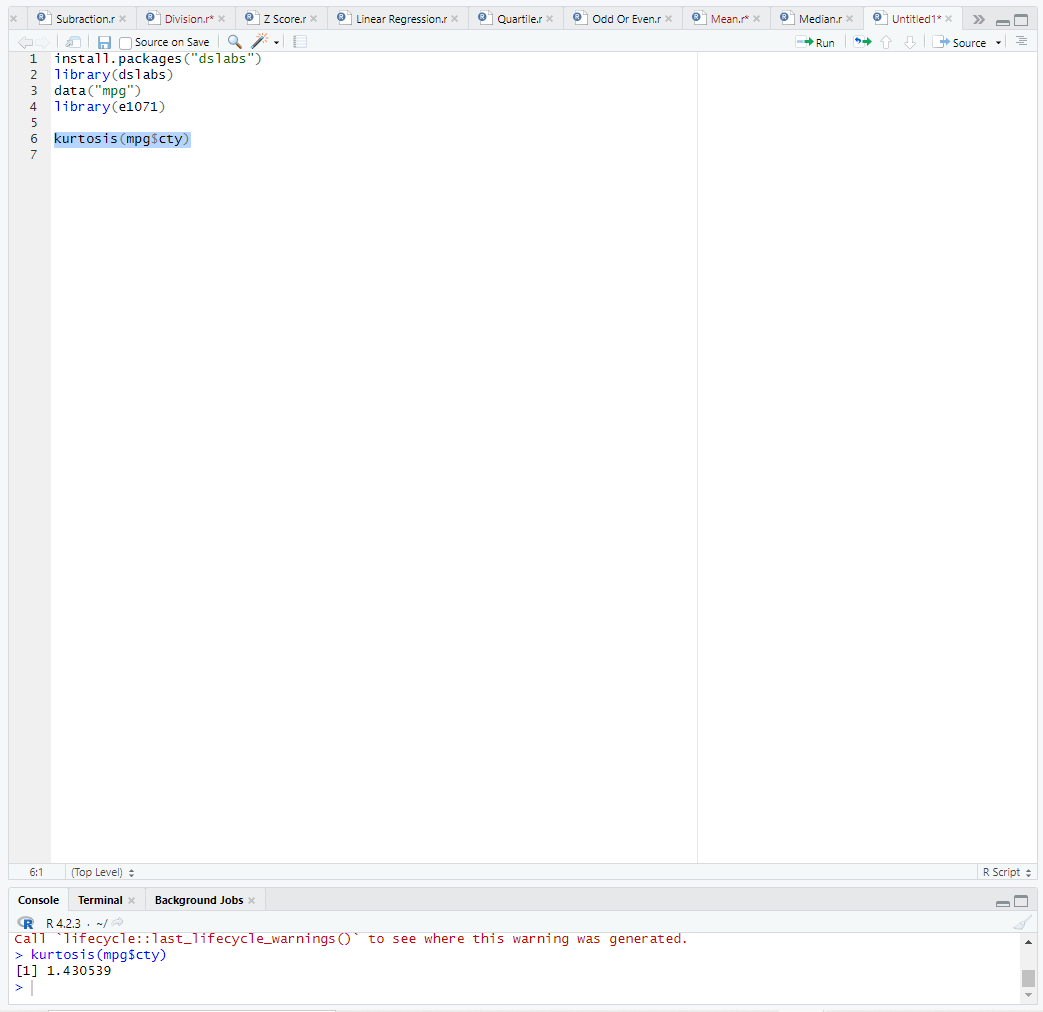
library(dslabs)

data("mpg")

library(e1071)

kurtosis(mpg$cty)

**OUTPUT:**

****