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 each end.

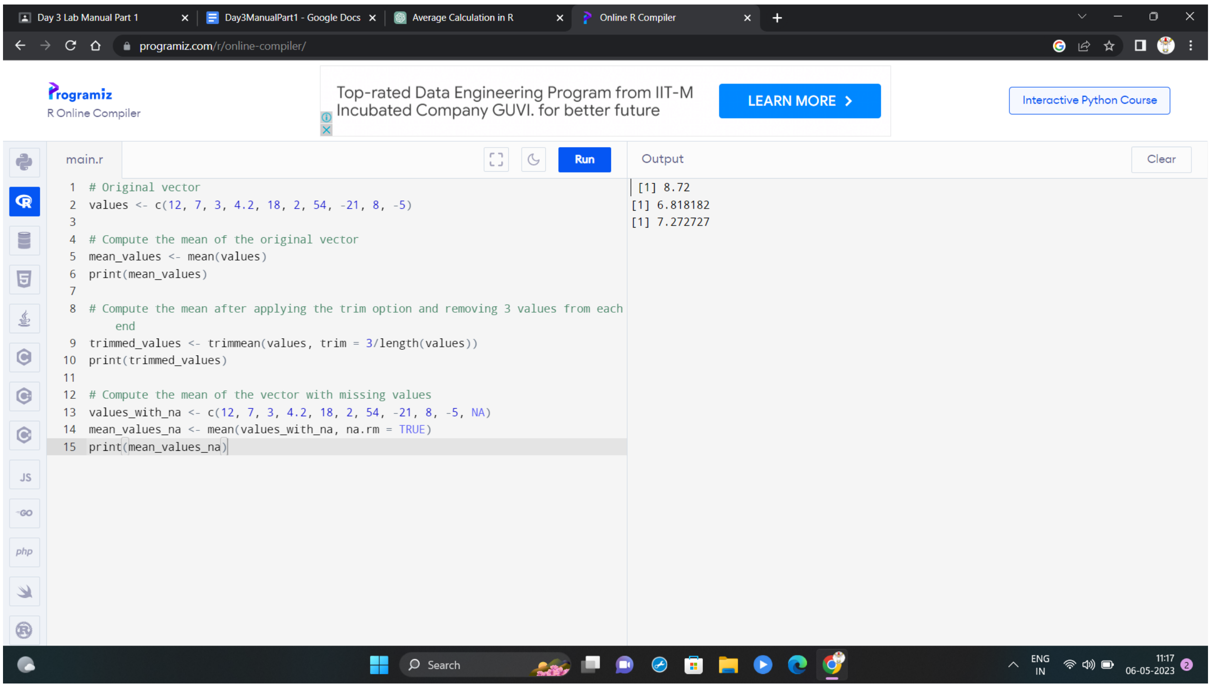
c) Compute the mean of the following vector .

(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



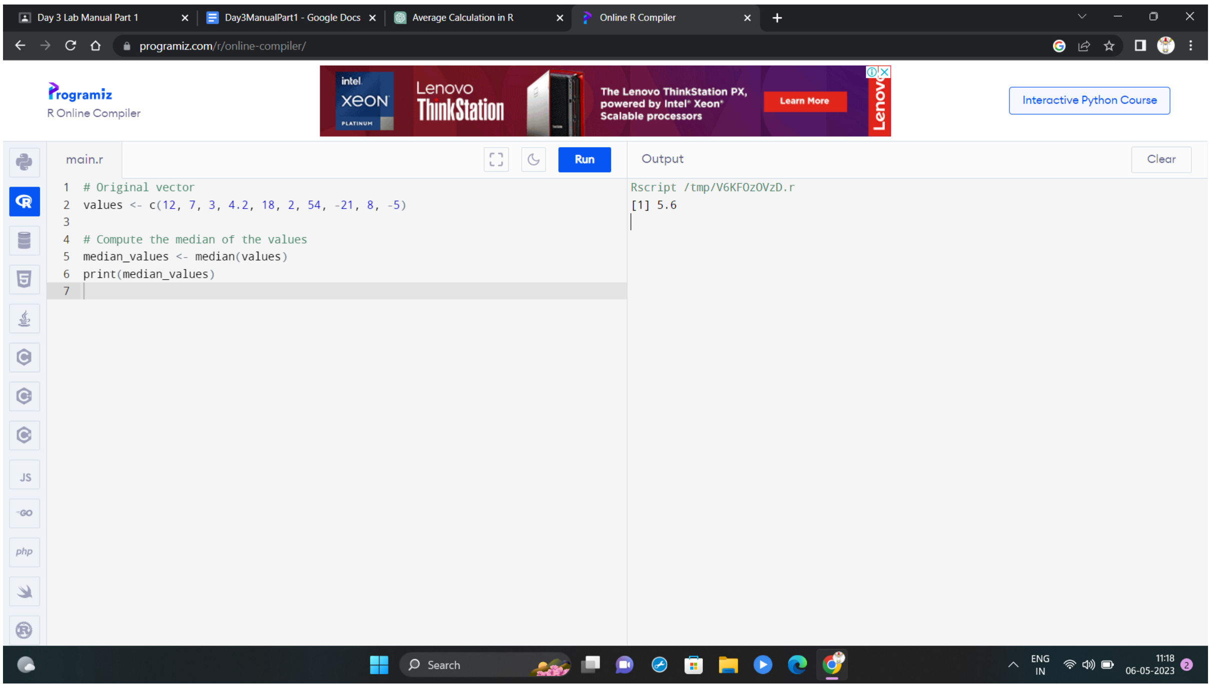
```r
# Original vector
values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

# Compute the mean of the original vector
mean_values <- mean(values)
print(mean_values)

# Compute the mean after applying the trim option and removing 3 values from each end
trimmed_values <- trimmean(values, trim = 3/length(values))
print(trimmed_values)

# Compute the mean of the vector with missing values
values_with_na <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5, NA)
mean_values_na <- mean(values_with_na, na.rm = TRUE)
print(mean_values_na)
```

```
[1] 8.72
[1] 6.818182
[1] 7.272727
```

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



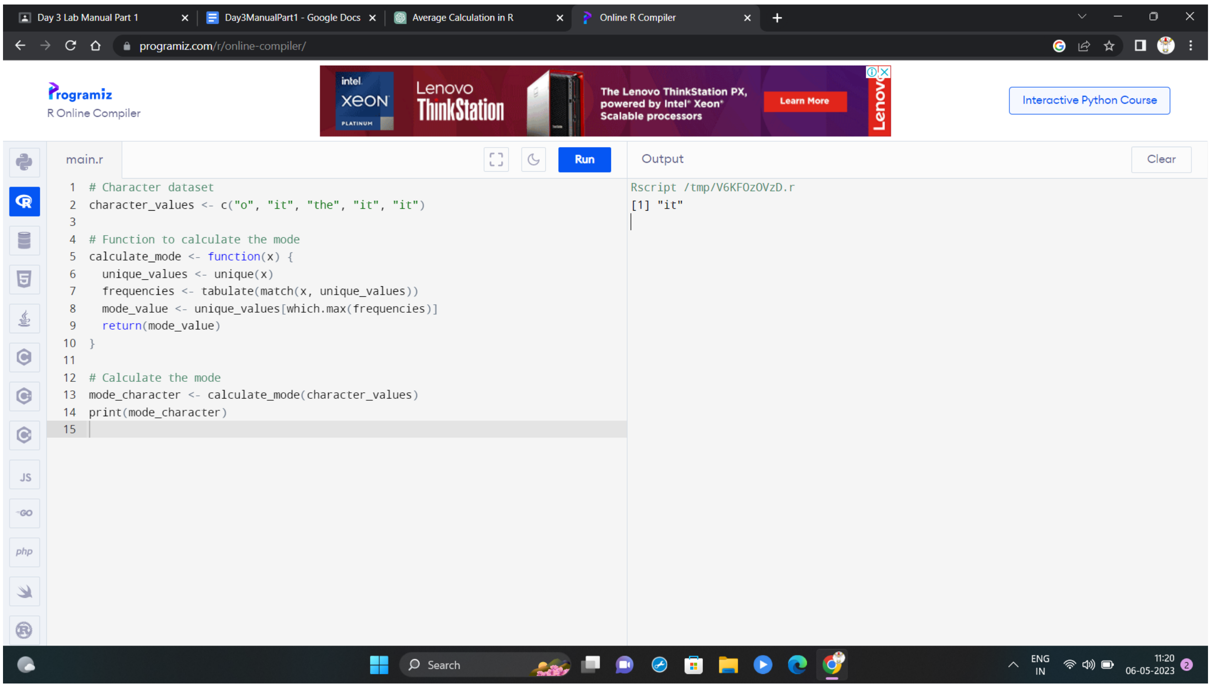
```
# Original vector
values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

# Compute the median of the values
median_values <- median(values)
print(median_values)
```

```
Rscript /tmp/V6KFOzOVzD.r
[1] 5.6
```

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")



```
# Character dataset
character_values <- c("o", "it", "the", "it", "it")

# Function to calculate the mode
calculate_mode <- function(x) {
  unique_values <- unique(x)
  frequencies <- tabulate(match(x, unique_values))
  mode_value <- unique_values[which.max(frequencies)]
  return(mode_value)
}

# Calculate the mode
mode_character <- calculate_mode(character_values)
print(mode_character)
```

```
Rscript /tmp/V6KFOzOVzD.r
[1] "it"
```

UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION

Exercise: 4

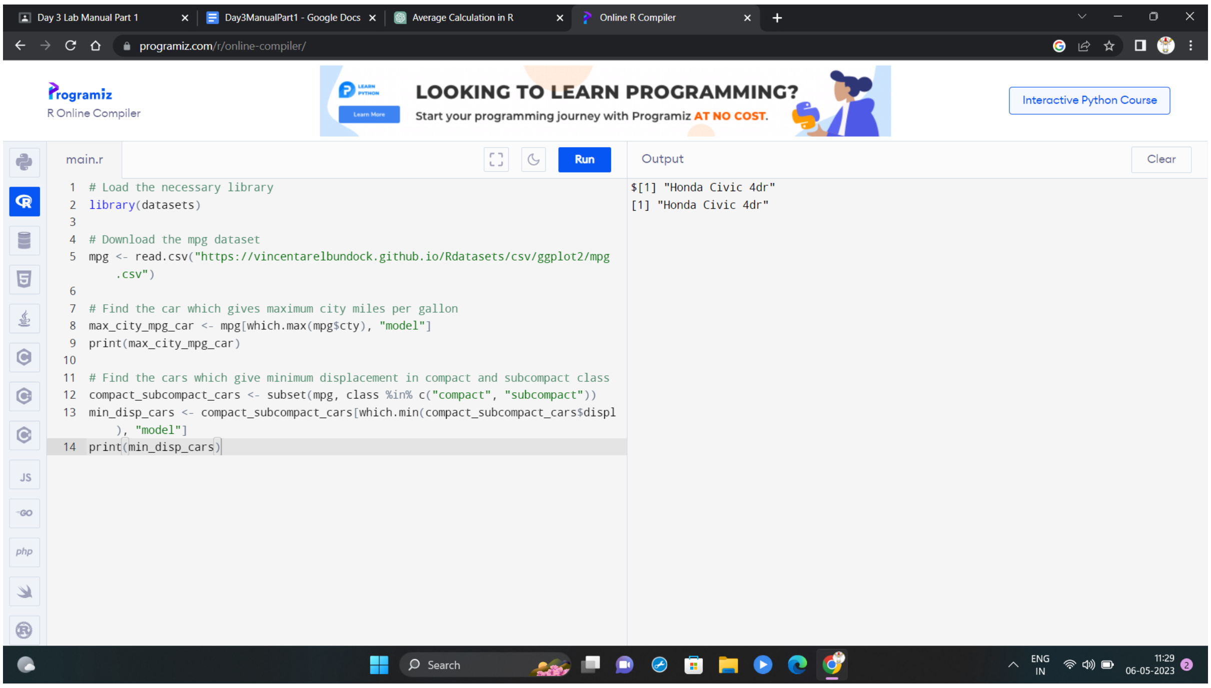
Download mpg dataset which contains Fuel economy data from 1999 and 2008 for 38 popular models of car from the URL given below.

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

Answer the following queries

i) Find the car which gives maximum city miles per gallon

ii) Find the cars which gives minimum disp in compact and subcompact class



```
# Load the necessary library
library(datasets)

# Download the mpg dataset
mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg
    .csv")

# Find the car which gives maximum city miles per gallon
max_city_mpg_car <- mpg[which.max(mpg$cty), "model"]
print(max_city_mpg_car)

# Find the cars which give minimum displacement in compact and subcompact class
compact_subcompact_cars <- subset(mpg, class %in% c("compact", "subcompact"))
min_disp_cars <- compact_subcompact_cars[which.min(compact_subcompact_cars$displ
    ), "model"]
print(min_disp_cars)
```

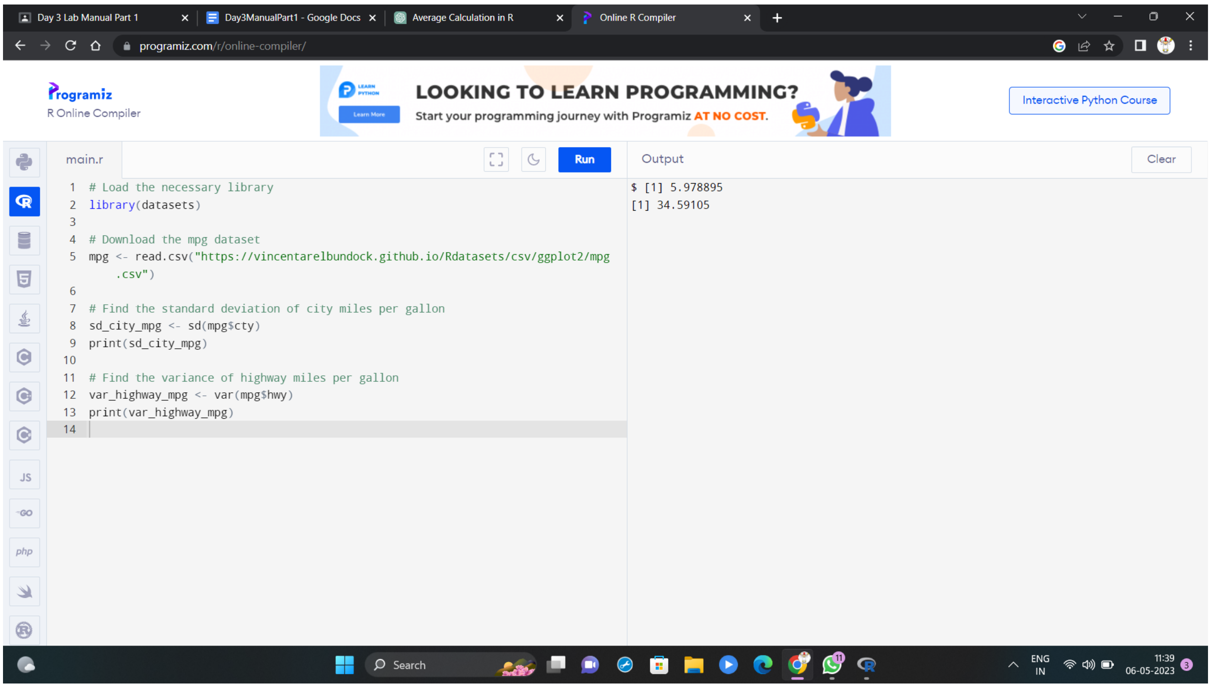
```
$[1] "Honda Civic 4dr"
[1] "Honda Civic 4dr"
```

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



Day 3 Lab Manual Part 1
Day3ManualPart1 - Google Docs
Average Calculation in R
Online R Compiler
programiz.com/r/online-compiler/
Programiz
R Online Compiler
LOOKING TO LEARN PROGRAMMING?
Start your programming journey with Programiz AT NO COST.
Interactive Python Course
main.r
Run
Output
Clear
Search
ENG
IN
11:39
06-05-2023

```r
# Load the necessary library
library(datasets)

# Download the mpg dataset
mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg
    .csv")

# Find the standard deviation of city miles per gallon
sd_city_mpg <- sd(mpg$cty)
print(sd_city_mpg)

# Find the variance of highway miles per gallon
var_highway_mpg <- var(mpg$hwy)
print(var_highway_mpg)
```

```
$ [1] 5.978895
[1] 34.59105
```

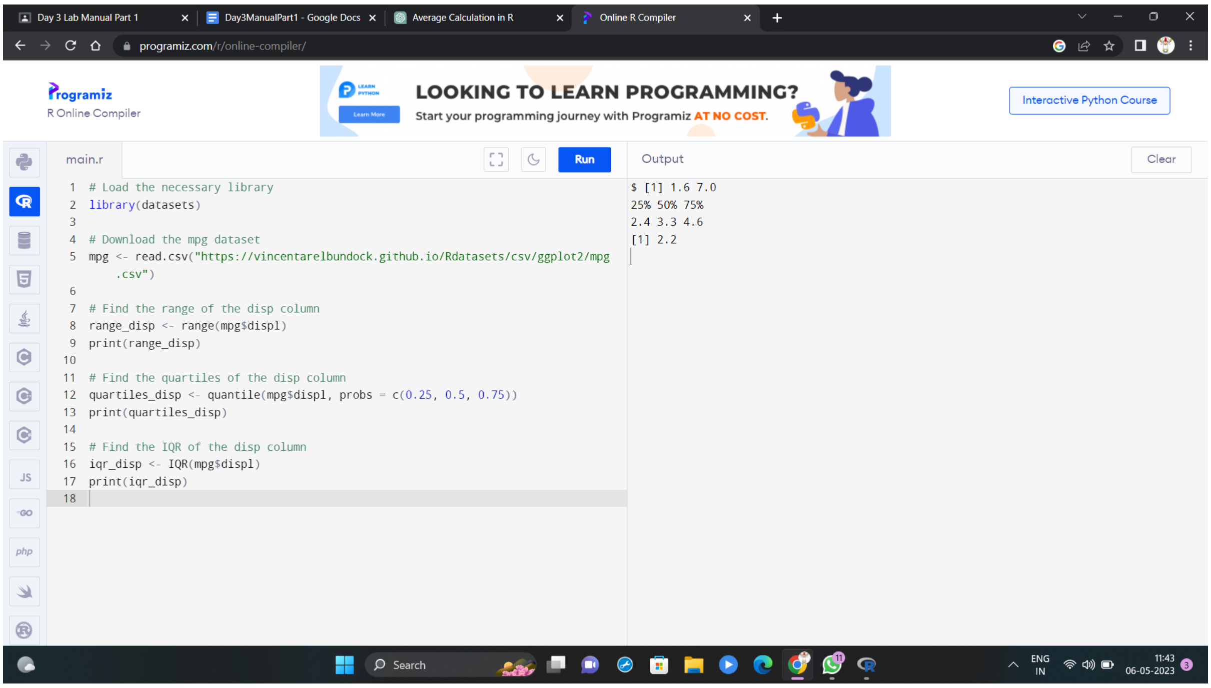
Exercise 6

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



Day 3 Lab Manual Part 1
Day3ManualPart1 - Google Docs
Average Calculation in R
Online R Compiler
programiz.com/r/online-compiler/
Programiz
R Online Compiler
LOOKING TO LEARN PROGRAMMING?
Start your programming journey with Programiz AT NO COST.
Interactive Python Course
main.r
Run
Output
Clear
Search
ENG
IN
11:43
06-05-2023

```r
# Load the necessary library
library(datasets)

# Download the mpg dataset
mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg
    .csv")

# Find the range of the disp column
range_disp <- range(mpg$displ)
print(range_disp)

# Find the quartiles of the disp column
quartiles_disp <- quantile(mpg$displ, probs = c(0.25, 0.5, 0.75))
print(quartiles_disp)

# Find the IQR of the disp column
iqr_disp <- IQR(mpg$displ)
print(iqr_disp)
```

```
$ [1] 1.6 7.0
25% 50% 75%
2.4 3.3 4.6
[1] 2.2
```

Exercise 7

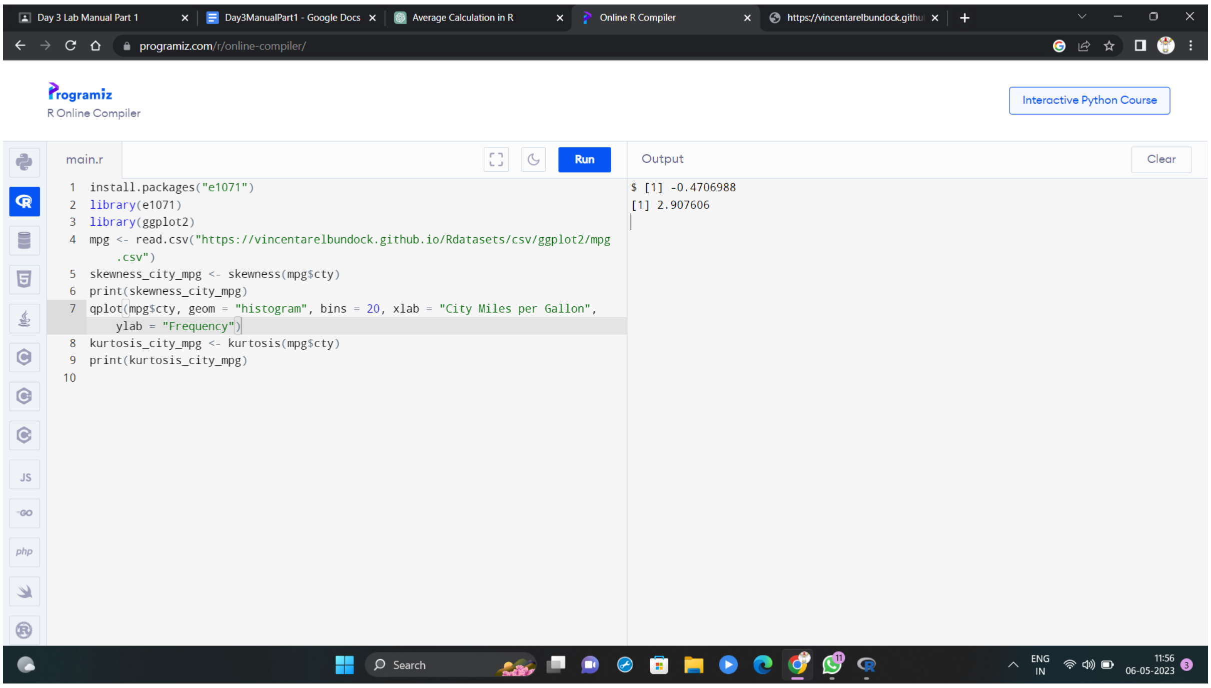
#Install Library

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



```r
install.packages("e1071")
library(e1071)
library(ggplot2)
mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")
skewness_city_mpg <- skewness(mpg$cty)
print(skewness_city_mpg)
qplot(mpg$cty, geom = "histogram", bins = 20, xlab = "City Miles per Gallon", ylab = "Frequency")
kurtosis_city_mpg <- kurtosis(mpg$cty)
print(kurtosis_city_mpg)
```

```
$ [1] -0.4706988
[1] 2.907606
```