Tutorial 2 sample answer

Question 1

a)

Command/code:

```
x <- c(30,30,31,31,32,32,33,33,34,34,35,35,36,36)
y <- c(29.5,30.2,32.1,34.5,36.3,35,38.2,37.6,37.7,
       36.1,33.6,34.2,26.8,27.4)
TIRES <- data.frame(x=x, y=y)
```

b)

Command/code:

```
names(TIRES) <- c("pressure", "mileage")
```

c)

Command/code:

```
TIRES[1:4,]
#or
head(TIRES, n=4)
```

Output:

```
  pressure mileage
1       30    29.5
2       30    30.2
3       31    32.1
4       31    34.5
```

d)

Command/code:

```
subset(TIRES, mileage>35)
#or
TIRES[TIRES$mileage>35,]
```

Output:

```
   pressure mileage
5        32    36.3
7        33    38.2
8        33    37.6
9        34    37.7
10       34    36.1
```

e)

Command/code:

```
getwd()
```

Output:

```
[1] "C:/Users/Hilmi Majid/Documents"
```

f)

Command/code:

```
write.csv(TIRES, file="Question1e.csv",
          row.names=FALSE)
```

(The file will be created in the location provided. In the above command, the file is created inside the working directory.)

Question 2

a)

Command/code:

```
TestScore <- read.csv("TestScore.csv")
```

(The file location depends on where the csv file is located. In the command above, the file is located inside the working directory.)

b)

Command/code:

```
names(TestScore)
```

Output:

```
[1] "Student"    "Score"      "IQ"
[4] "StudyHours" "GPA"        "Class"
```

c)

Command/code:

```
TestScore[8,]
```

Output:

```
  Student Score IQ StudyHours GPA Class
8       8    75 95         10 2.1     B
```

d)

Command/code:

```
library(psych)
describe(TestScore)
```

Output:

```
             vars  n   mean    sd median trimmed
Student         1 10   5.50  3.03    5.5    5.50
Score           2 10  83.20 11.10   82.5   83.38
IQ              3 10 100.90 11.22  100.0   99.88
StudyHours      4 10  18.50 11.80   20.0   18.12
GPA             5 10   2.43  0.68    2.3    2.36
Class*          6 10   1.50  0.53    1.5    1.50
              mad  min   max range  skew
Student      3.71  1.0  10.0   9.0  0.00
Score       12.60 65.0 100.0  35.0 -0.06
IQ           7.41 85.0 125.0  40.0  0.65
StudyHours  11.12  0.0  40.0  40.0  0.14
GPA          0.52  1.5   3.9   2.4  0.73
Class*       0.74  1.0   2.0   1.0  0.00
             kurtosis   se
Student         -1.56 0.96
Score           -1.43 3.51
IQ              -0.32 3.55
StudyHours      -0.99 3.73
GPA             -0.27 0.21
Class*          -2.19 0.17
```

e)	Command/code:
	`mean(TestScore$GPA)`
	`median(TestScore$GPA)`
	`var(TestScore$GPA)`
	
	Output:
	`> mean(TestScore$GPA)`
	`[1] 2.43`
	`> median(TestScore$GPA)`
	`[1] 2.3`
	`> var(TestScore$GPA)`
	`[1] 0.4578889`
	
	Explanation:
	The mean for the GPA is 2.43, the median for GPA is 2.3, and the variance for GPA is 0.457889.

f)	Command/code:
	`quantile(TestScore$GPA, c(0.25,0.75))`
	`IQR(TestScore$GPA)`
	
	Output:
	`> quantile(TestScore$GPA, c(0.25,0.75))`
	`  25%   75%`
	`2.100 2.675`
	`> IQR(TestScore$GPA)`
	`[1] 0.575`
	
	Explanation:
	The first quartile, Q_1 is 2.1 and the third quartile, Q_3 is 2.675. The interquartile range is 0.575.

g)	Command/code:
	`library(e1071)`
	`skewness(TestScore$GPA)`
	`kurtosis(TestScore$GPA)`
	
	Output:
	`> skewness(TestScore$GPA)`
	`[1] 0.7254797`
	`> kurtosis(TestScore$GPA)`
	`[1] -0.2749991`
	
	Explanation:
	Skewed to the right (but only slightly), since the skewness is positive. Platykurtic since the kurtosis is negative.

h)	Command/code:
	`subset(TestScore, Class=="B")`
	`# or`
	`TestScore[TestScore$Class=="B",]`
	`classB <- subset(TestScore, Class=="B")`
	
	Output:
	`  Student Score  IQ StudyHours GPA Class`
	`2       2    95 104         40 2.6     B`
	`5       5    85 100         20 2.4     B`
	`6       6    80 100         20 2.2     B`
	`8       8    75  95         10 2.1     B`
	`9       9    72  85          0 1.5     B`

i)	Command/code:
	`mean(classB$IQ)`
	`median(classB$IQ)`
	`var(classB$IQ)`
	
	Output:
	`> mean(classB$IQ)`
	`[1] 96.8`
	`> median(classB$IQ)`
	`[1] 100`
	`> var(classB$IQ)`
	`[1] 53.7`
	
	Explanation:
	The mean IQ for students in class B is 96.8. The median IQ for students in class B is 100. And the variance for the IQ of students in class B is 53.7.

j)	Command/code:
	`max(classB$IQ) - min(classB$IQ)`
	
	Output:
	`[1] 19`
	
	Explanation:
	The range for the IQ of students in class B is 19.

k)	Command/code:
	`quantile(TestScore$StudyHours, probs=0.85)`
	
	Output:
	`  85%`
	`28.25`
	
	Explanation:
	85% students' study hours are less than 28.25. In other words, the minimum value for the top 15% of students' study hours is 28.25. The value 28.25 separates the lower 85% and top 15% students' study hours.

l)	Command/code:
	`table(TestScore$StudyHours)`
	`names(table(TestScore$StudyHours))[which.max(`
	`table(TestScore$StudyHours))]`
	
	Output:
	`> table(TestScore$StudyHours)`
	
	`0 5 10 15 20 25 30 40`
	`1 1 1 1 3 1 1 1`
	`> names(table(TestScore$StudyHours))[which.max(`
	`table(TestScore$StudyHours))]`
	`[1] "20"`
	
	Explanation:
	The mode for the study hours is 20, as it is the value with the highest frequency. The second line of the command/code is not necessary as it is obvious from the table of frequency that 20 has the highest frequency.

Question 3

a)

Command/code:

```
property_sales <- read.csv(file.choose())
```

b)

Command/Output:

```
> names(property_sales)
[1] "Sales"             "Land.value"
[3] "Improvement.value" "Neighbourhood"
```

c)

Command/code:

```
head(property_sales, 6)

#OR

property_sales[1:6,]
```

Output:

```
> property_sales[1:6,]
  Sales Land.value Improvement.value Neighbourhood
1 378.0      81.84            243.30        Cheval
2 273.0      60.48            134.47        Cheval
3 321.2     115.58            255.42        Cheval
4 395.0     119.61            202.05        Cheval
5 272.0      84.69            133.58        Cheval
6 350.0      78.69            154.70        Cheva
```

d)

Command/output:

```
> table(property_sales$Neighbourhood)

      Cheval   DavisIsles HuntersGreen     HydePark
          44           42           56           34
```

e)

Command/output:

```
> max(property_sales$Sales)
[1] 3200
> min(property_sales$Sales)
[1] 140
> max(property_sales$Sales)-min(property_sales$Sales)
[1] 3060
```

Comment:

Highest sales price is 3200, and lowest sales price is 140. The range is 3060.

f)	Command/output:
	`> mean(property_sales$Sales)`
	`[1] 549.8932`
	`> median(property_sales$Sales)`
	`[1] 408.5`
	`> var(property_sales$Sales)`
	`[1] 166285.8`
	
	Comment:
	The mean is 549.89, median is 408.5, and variance is 166285.8

g)	Command:
	`Davis <- subset(property_sales,`
	`Neighbourhood == "DavisIsles")`
	`mean(Davis$Sales)`
	`median(Davis$Sales)`
	`var(Davis$Sales)`
	
	`#OR`
	
	`mean(property_sales$Sales[property_sales$Neighbourhood == "DavisIsles"])`
	
	Output:
	`> mean(Davis$Sales)`
	`[1] 818.2357`
	`> median(Davis$Sales)`
	`[1] 707.5`
	`> var(Davis$Sales)`
	`[1] 268296.2`
	
	Comment:
	The mean is 818.24, median is 707.5, and variance is 268296.2.

h)	Command:
	`library(e1071)`
	`skewness(Davis$Sales)`
	`kurtosis(Davis$Sales)`
	
	Output:
	`> skewness(Davis$Sales)`
	`[1] 2.443209`
	`> kurtosis(Davis$Sales)`
	`[1] 8.312356`
	
	Comment:
	Skewed to the right because the skewness is positive. Leptokurtic because kurtosis is positive.

i)	Command:
	`Cheval <- subset(property_sales,`
	`Neighbourhood=="Cheval")`
	`quantile(Cheval$Land.value, c(0.25,0.5,0.75))`
	`IQR(Cheval$Land.value)`
	
	Output:
	`> quantile(Cheval$Land.value, c(0.25,0.5,0.75))`
	`  25%  50%  75%`
	`63.6050 81.5350 118.1925`
	`> IQR(Cheval$Land.value)`
	`[1] 54.5875`
	
	Comment:
	The first quartile is 63.605, and the third quartile is 118.19. The interquartile range is 54.5875.