

WIPRO

Tableau Case Study 2

DST



Sayani Maiti
7-25-2018
Employee id - 390004

Contents

Problem 1	2
Problem 2	4

Problem 1

Aim: To determine the correlation coefficient in Tableau by calling cor function in R:

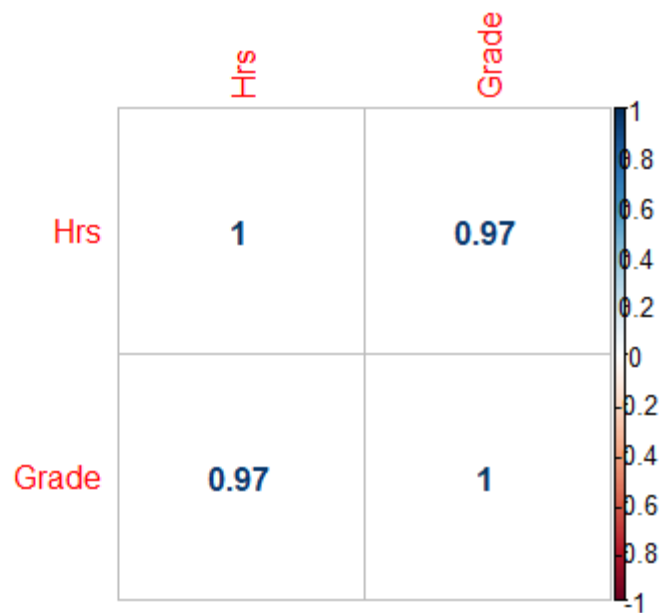
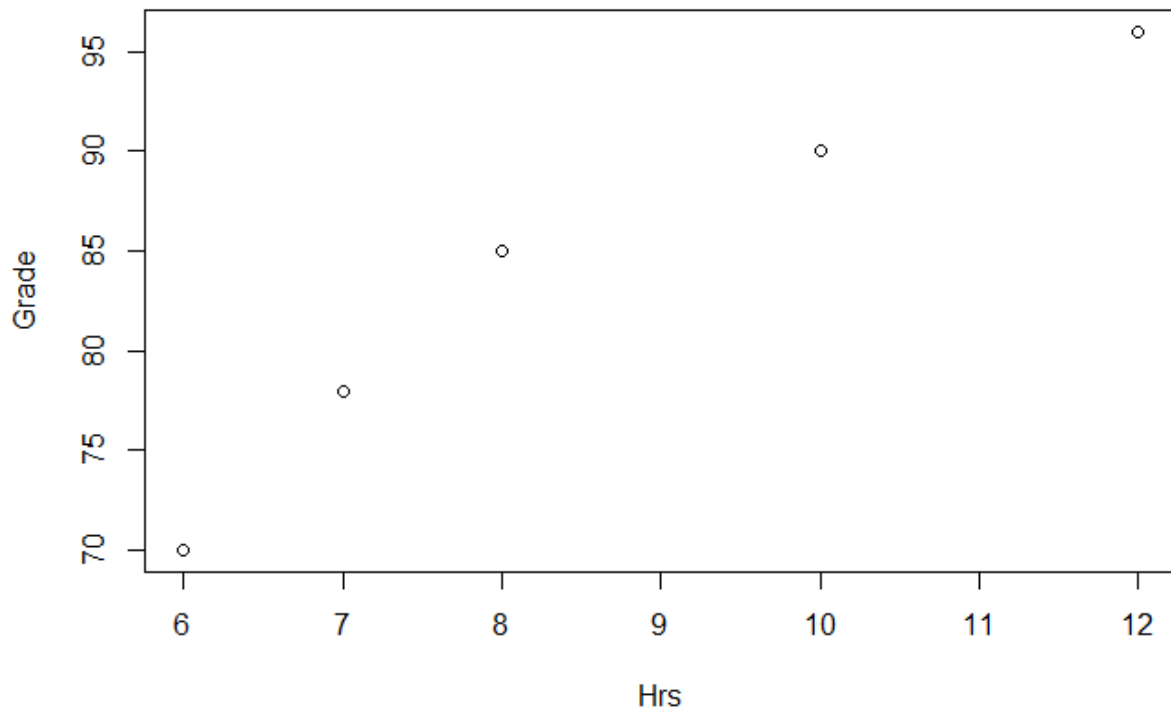
Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation is one when the values of variables increase or decrease in parallel; a negative correlation is when the value of one variable increases as the value of the other variable decreases.

	A	B
1	No. of hours studied	Grades
2	6	70
3	7	78
4	8	85
5	10	90
6	12	96

Solution: Rcode

```
1 setwd("C:\\Users\\Sayani\\Documents\\wipro\\Tableau")
2 cor = read.csv("Tab_CS2.csv",1)
3 cor
4 attach(cor)
5 #timeseries = ts(cost, frequency = 12, start = c(1))
6 plot(Hrs,Grade)
7 newdatacor <- cor(cor)
8 #is.corr=FALSE
9 corplot::corplot(newdatacor, method = "number")
10 corplot::corplot(newdatacor, method = "circle")
11
12 |
```

Scatter plot:



Correlation plot:

Conclusion: from the graph, it can be seen that the hrs to grade correlation coefficient is 0.97 which is almost equal to 1. Therefore, it is positive correlation

Problem 2

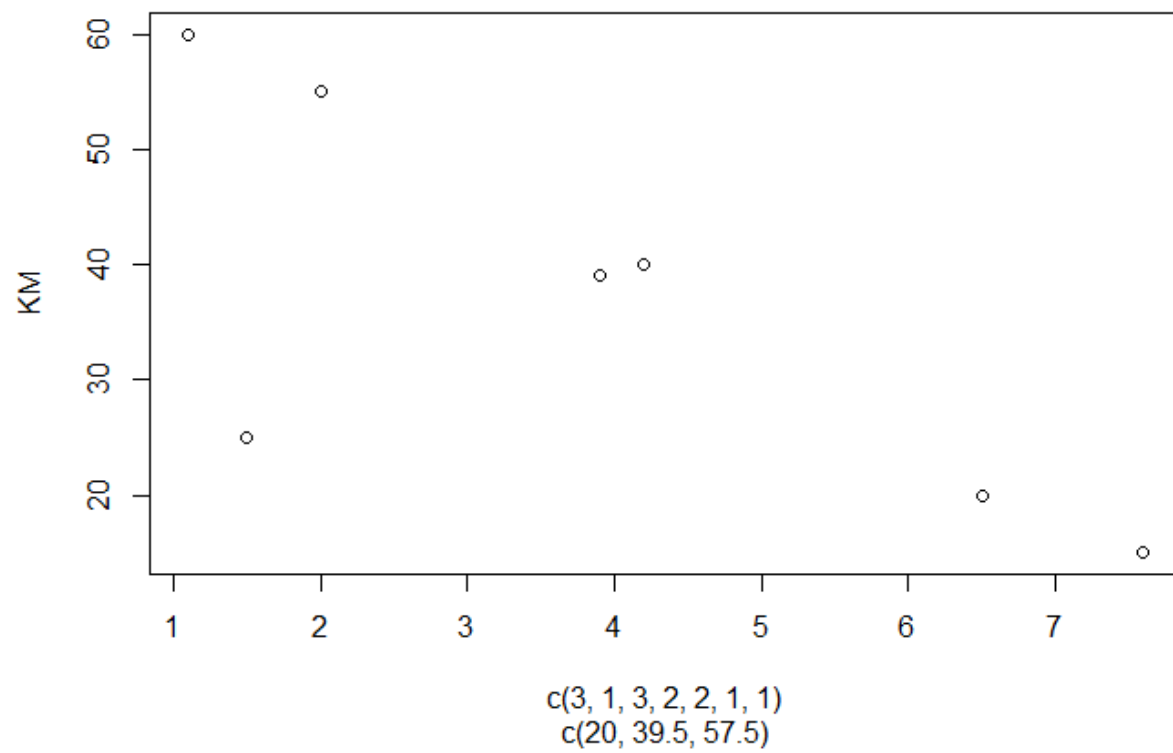
To split the given data in "Cars.xlsx" into three clusters by using kmeans clustering. PI provide step by step solution in form of document which will split car.xlsx into 3 clusters and also in document show scatter plot for petrol vs kms with cluster nos as shape.

Input: "Cars.xlsx". Data set as shown in Fig. below

	A	B
1	Petrol	Kilometers
2	1.1	60
3	6.5	20
4	4.2	40
5	1.5	25
6	7.6	15
7	2	55
8	3.9	39

R code:

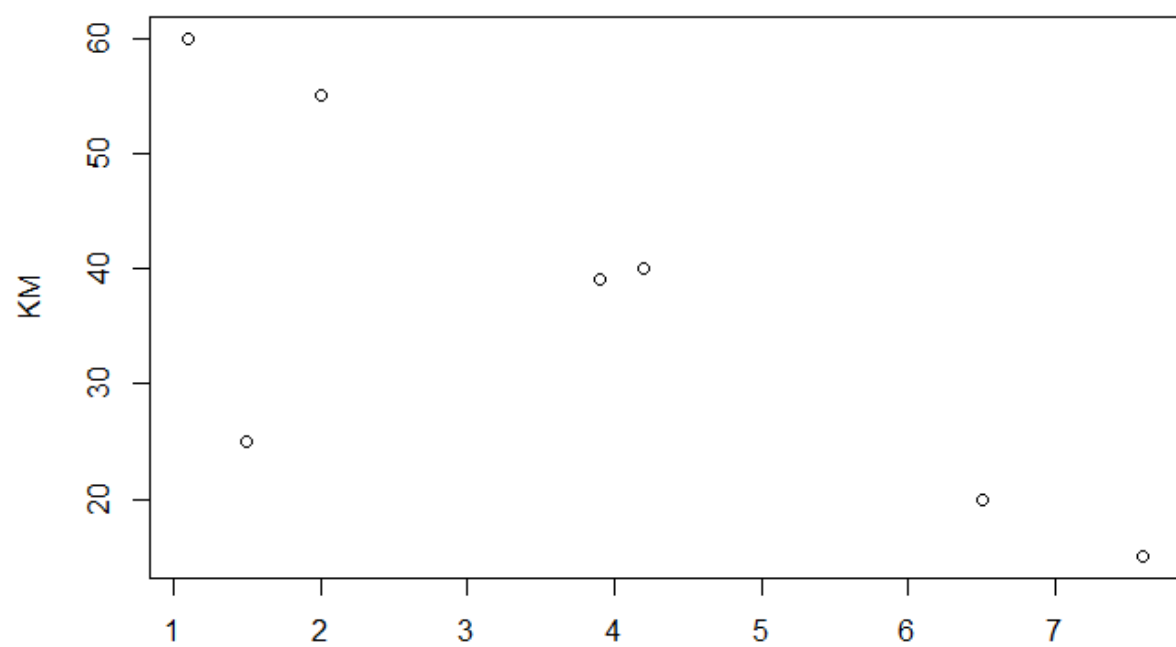
```
setwd("C:\\Users\\Sayani\\Documents\\wipro\\Tableau")
cor1 = read.csv("Cars.csv",1)
cor1
attach(cor1)
#timeseries = ts(cost, frequency = 12, start = c(1))
plot(Petrol,KM)
newdatacor <- cor(cor1)
#is.corr=FALSE
corrplot::corrplot(newdatacor, method = "number")
corrplot::corrplot(newdatacor, method = "circle")
kmeans(Petrol,3,iter.max=5,nstart=5)
```



Conclusion:

The above scatter plot is based on the KM

Data is scattered around 20, 40 and 57.



c(3, 1, 3, 2, 2, 1, 1)
c(5.2, 4.05, 1.55, 20, 39.5, 57.5)