



SYMBIOSIS INSTITUTE OF TECHNOLOGY (SIT)

Constituent of Symbiosis International (Deemed University), Pune

(Established under Section 3 of the UGC Act of 1956 vide notification number F-9-12/2001-U-3 of the Government of India)

॥वसुषीव कुटुम्बकम्॥

Re-Accredited by NAAC with 'A' Grade

Assignment 05

Subject: Data Science Lab

Name of Student	Antriksh Sharma
PRN No.	20070122021
Branch	CS
Class	A
Academic Year & Semester	2023-24, VII
Date of Performance	21.08.23
Title of Lab Assignment	CORRELATION AND COVARIANCE

Theory: Put simply, both covariance and correlation measure the relationship and the dependency between two variables. Covariance indicates the direction of the linear relationship between variables while correlation measures both the strength and direction of the linear relationship between two variables. Correlation is a function of the covariance. What sets these two concepts apart is the fact that correlation values are standardized whereas covariance values are not.

Code & Output:

```
data <- iris[, c('Sepal.Length', 'Sepal.Width', 'Petal.Length', 'Petal.Width')]
cor_data = cor(data)
> print("Correlation matrix")
[1] "Correlation matrix"
> print(cor_data)
      Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000 -0.1175698   0.8717538   0.8179411
Sepal.Width     -0.1175698  1.0000000  -0.4284401  -0.3661259
Petal.Length     0.8717538  -0.4284401   1.0000000   0.9628654
Petal.Width      0.8179411  -0.3661259    0.9628654   1.0000000
print("Correlation matrix")
print(cor_data)
```

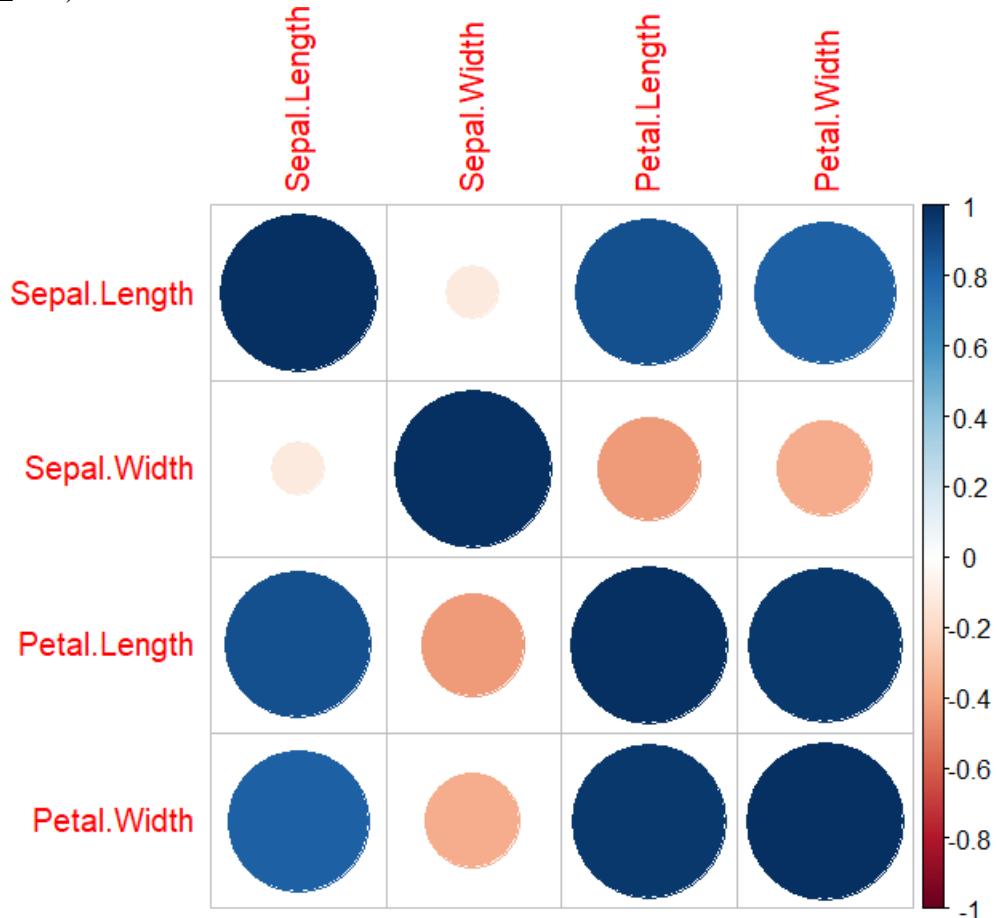


SYMBIOSIS INSTITUTE OF TECHNOLOGY (SIT)

Constituent of Symbiosis International (Deemed University), Pune

(Established under Section 3 of the UGC Act of 1956 vide notification number F-9-12/2001-U-3 of the Government of India)
Re-Accredited by NAAC with 'A' Grade

```
library(corrplot)
corrplot(cor_data)
```



A default correlation matrix plot (called a Correlogram) is generated. Positive correlations are displayed in a blue scale while negative correlations are displayed in a red scale.

```
library(car)
lm<-lm(Sepal.Width~Species,data=iris)
anova(lm)
```

```
> anova(lm)
Analysis of Variance Table

Response: Sepal.Width
          Df  Sum Sq Mean Sq F value    Pr(>F)
Species      2 11.345  5.6725   49.16 < 2.2e-16 ***
Residuals 147 16.962  0.1154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```