

## Experiment No. 8

**Title:** To Implement Principal Component Analysis in R

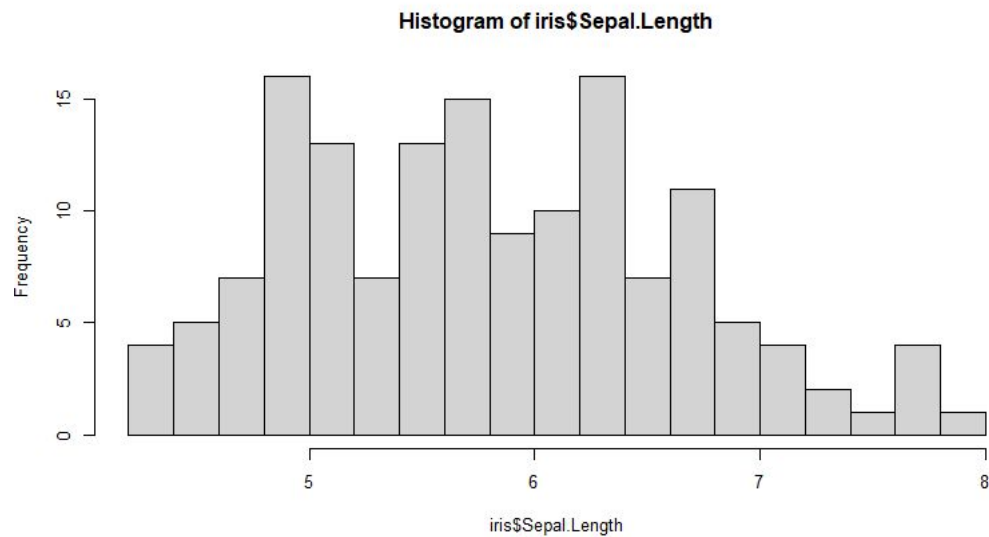
**Problem:**

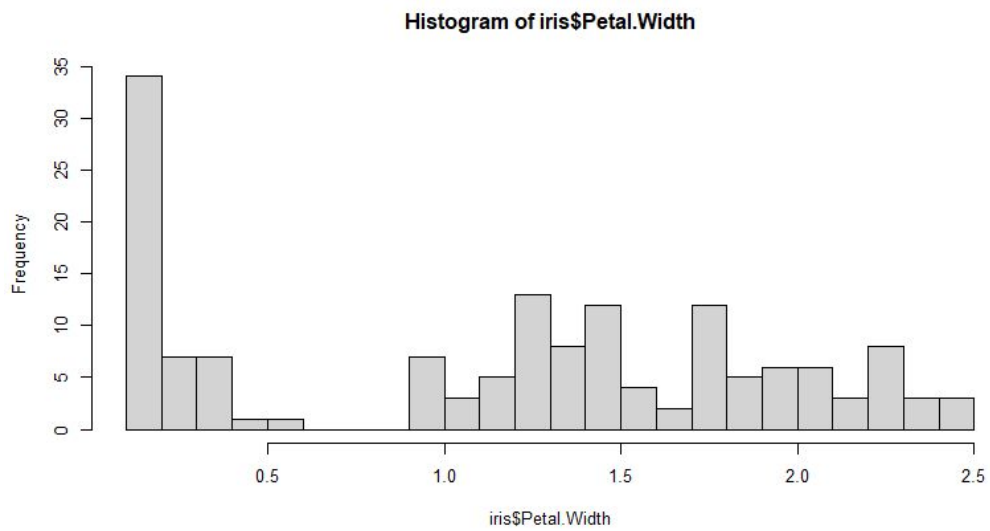
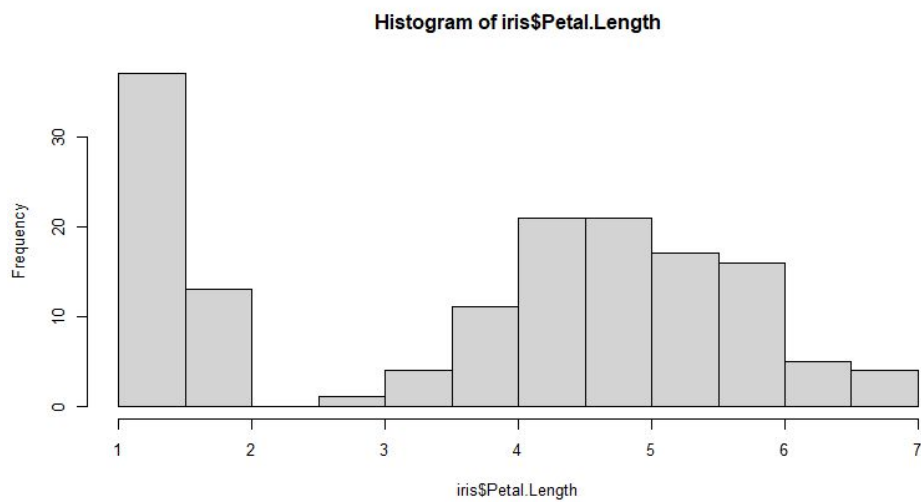
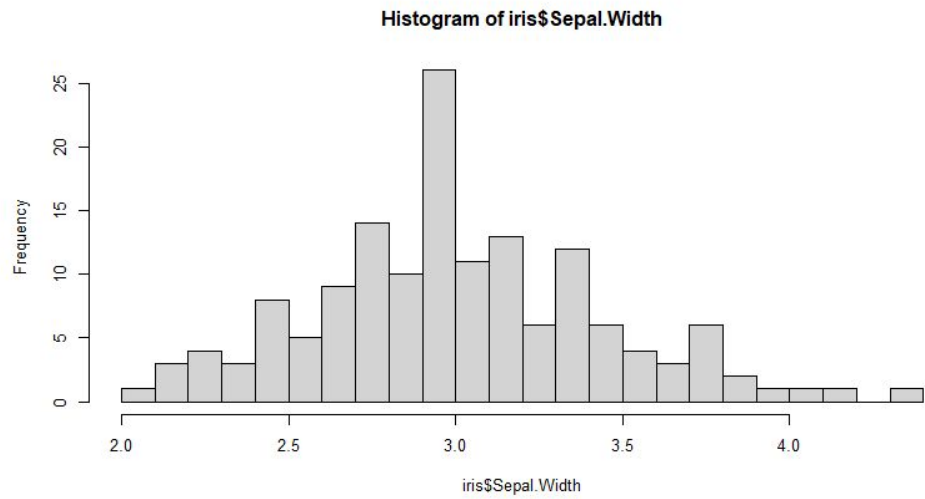
PCA on Iris Dataset

Steps are as follows:

- 1) Importing the dataset and plotting graph

```
data(iris)
head(iris)
par(mfrow = c(2, 2))
hist(iris$Sepal.Length, breaks = 20)
hist(iris$Sepal.Width, breaks = 20)
hist(iris$Petal.Length, breaks = 20)
hist(iris$Petal.Width, breaks = 20)
```





## 1) Log transform on the dataset

```
log.iris <- log(iris[, 1:4])
```

```
iris.species <- iris[, 5]
```

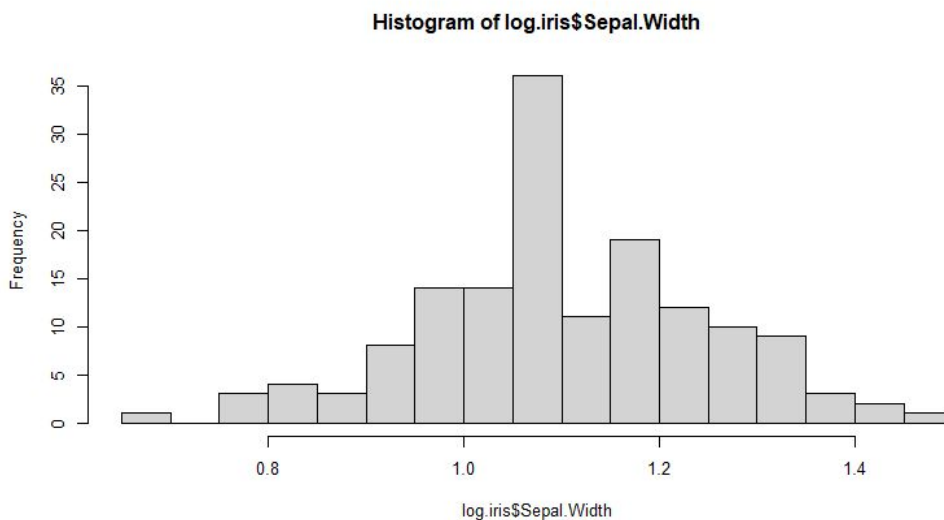
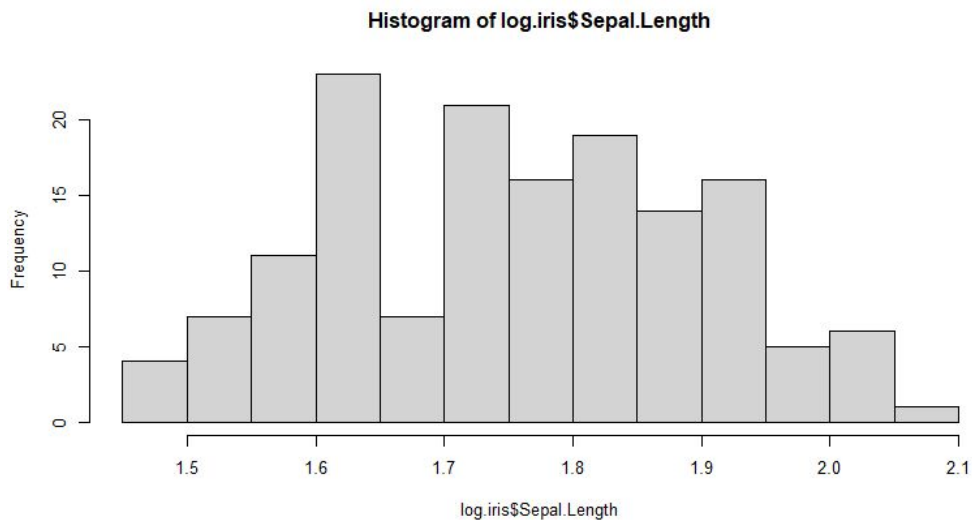
```
par(mfrow = c(2, 2))
```

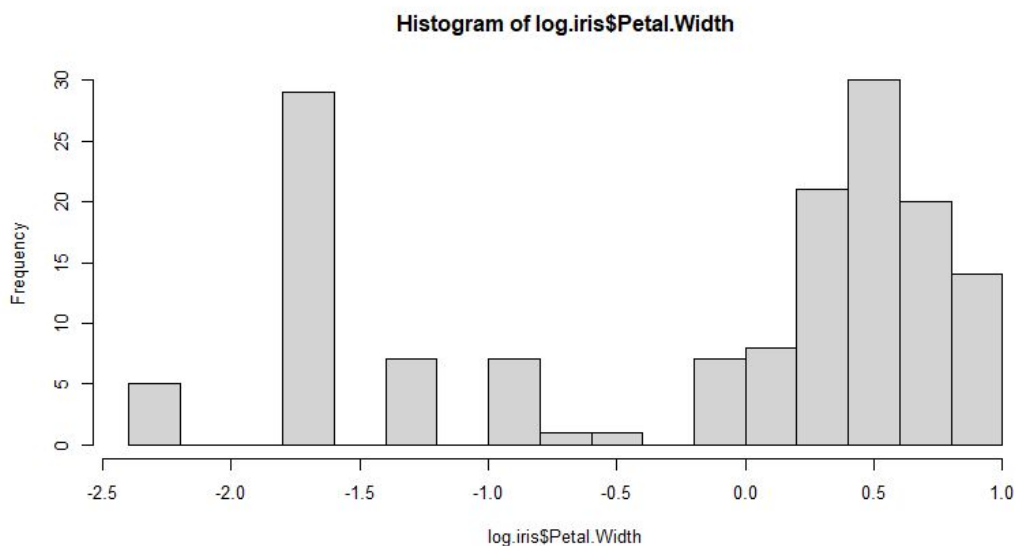
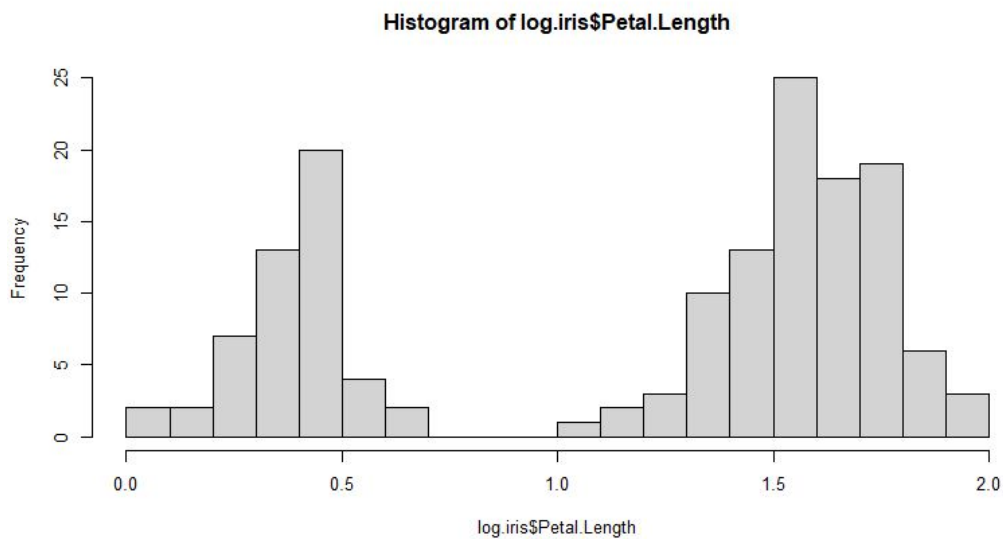
```
hist(log.iris$Sepal.Length, breaks = 20)
```

```
hist(log.iris$Sepal.Width, breaks = 20)
```

```
hist(log.iris$Petal.Length, breaks = 20)
```

```
hist(log.iris$Petal.Width, breaks = 20)
```





2) Since the data is still bimodal we need to use prcomp.

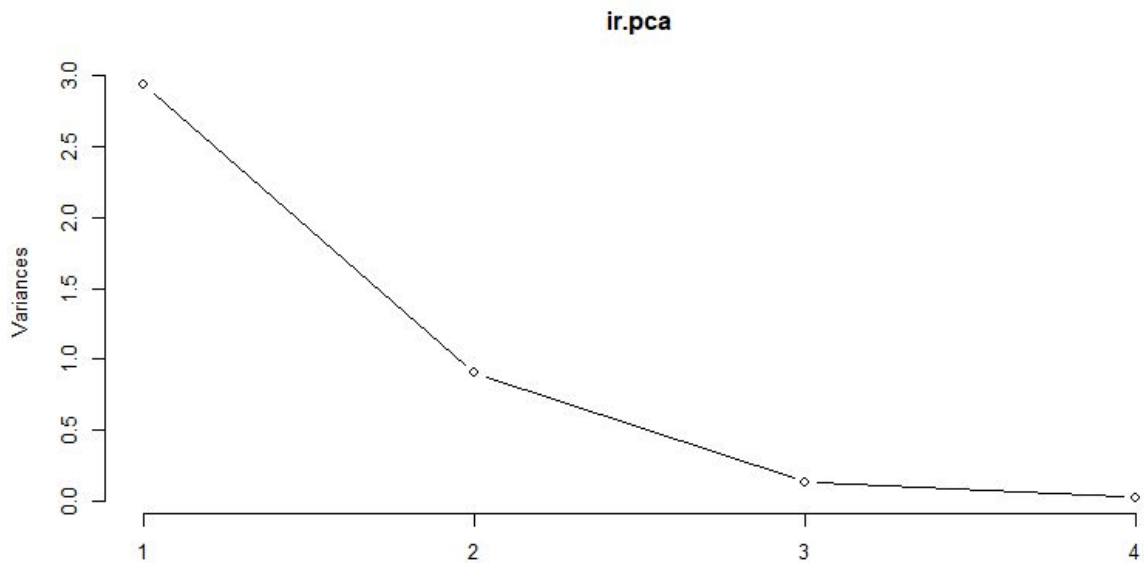
```
ir.pca <- prcomp(log.iris, center = TRUE, scale = TRUE)
print(ir.pca)
```

```
> ir.pca <- prcomp(log.iris, center = TRUE, scale = TRUE)

> print(ir.pca)
Standard deviations (1, ..., p=4):
[1] 1.7124583 0.9523797
[3] 0.3647029 0.1656840

Rotation (n x k) = (4 x 4):
      PC1
Sepal.Length  0.5038236
Sepal.Width   -0.3023682
Petal.Length  0.5767881
Petal.Width   0.5674952
      PC2
Sepal.Length -0.45499872
Sepal.Width  -0.88914419
Petal.Length -0.03378802
Petal.Width  -0.03545628
      PC3
Sepal.Length  0.7088547
Sepal.Width  -0.3311628
Petal.Length -0.2192793
Petal.Width  -0.5829003
      PC4
Sepal.Length  0.19147575
Sepal.Width  -0.09125405
Petal.Length -0.78618732
Petal.Width   0.58044745
```

3) Plotting iris principal component analysis graph.



4) Summary of the Iris PCA data

```
> summary(ir.pca)
Importance of components:
              PC1      PC2      PC3      PC4
Standard deviation  1.7125  0.9524  0.36470  0.16568
Proportion of Variance 0.7331 0.2268 0.03325 0.00686
Cumulative Proportion 0.7331 0.9599 0.99314 1.00000
```

5) Performing PCA again by diagonalizing the covariance matrix:

```
iris.mat <- as.matrix(log.iris)
cov.mat <- cor(iris.mat)
eigen(cov.mat)
```

```
> iris.mat <- as.matrix(log.iris)
> cov.mat <- cor(iris.mat)
> eigen(cov.mat)
eigen() decomposition
$values
[1] 2.9325135 0.9070271 0.1330082 0.0274512

$vectors
      [,1]      [,2]      [,3]      [,4]
[1,] 0.5038236 -0.45499872 0.7088547 0.19147575
[2,] -0.3023682 -0.88914419 -0.3311628 -0.09125405
[3,] 0.5767881 -0.03378802 -0.2192793 -0.78618732
[4,] 0.5674952 -0.03545628 -0.5829003 0.58044745
```