Solution_HW1

Wei-Rou,Xu

2023-11-15

1. What are the types of variables?

```
SL=iris$Sepal.Length
SW=iris$Sepal.Width
PL=iris$Petal.Length
PW=iris$Petal.Width
Species=iris$Species

class(SL)
```

```
## [1] "numeric"
```

```
class(SW)
```

```
## [1] "numeric"
```

```
class(PL)
```

```
## [1] "numeric"
```

```
class(PW)
```

```
## [1] "numeric"
```

```
class(Species)
```

```
## [1] "factor"
```

**2. For each variables, give the histogram/barplot and boxplot.**

- histogram

```
par(mfrow = c(2,2))
hist(SL,main="Histogram of Sepal Length")
hist(SW,main="Histogram of Sepal Width")
hist(PL,main="Histogram of Petal Length")
hist(PW,main="Histogram of Petal Width")
```

Histogram of Sepal Length

Histogram of Sepal Width

Histogram of Petal Length

Histogram of Petal Width
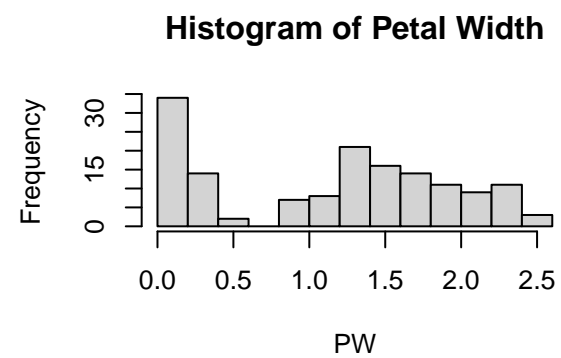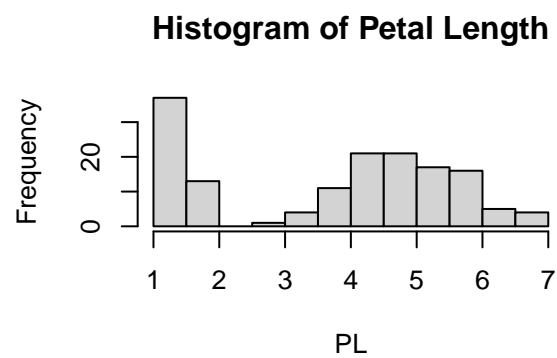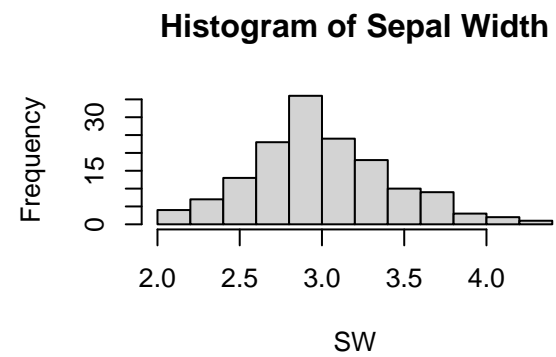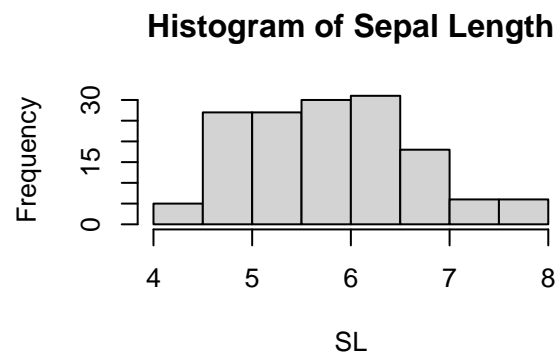
- barplot

```
barplot(table(Species),main="barplot of Species")
```
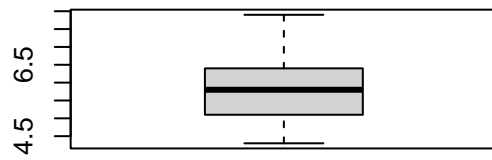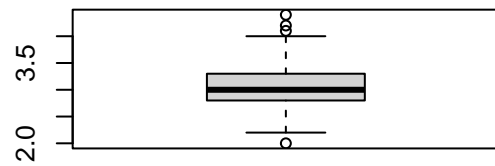
barplot of Species
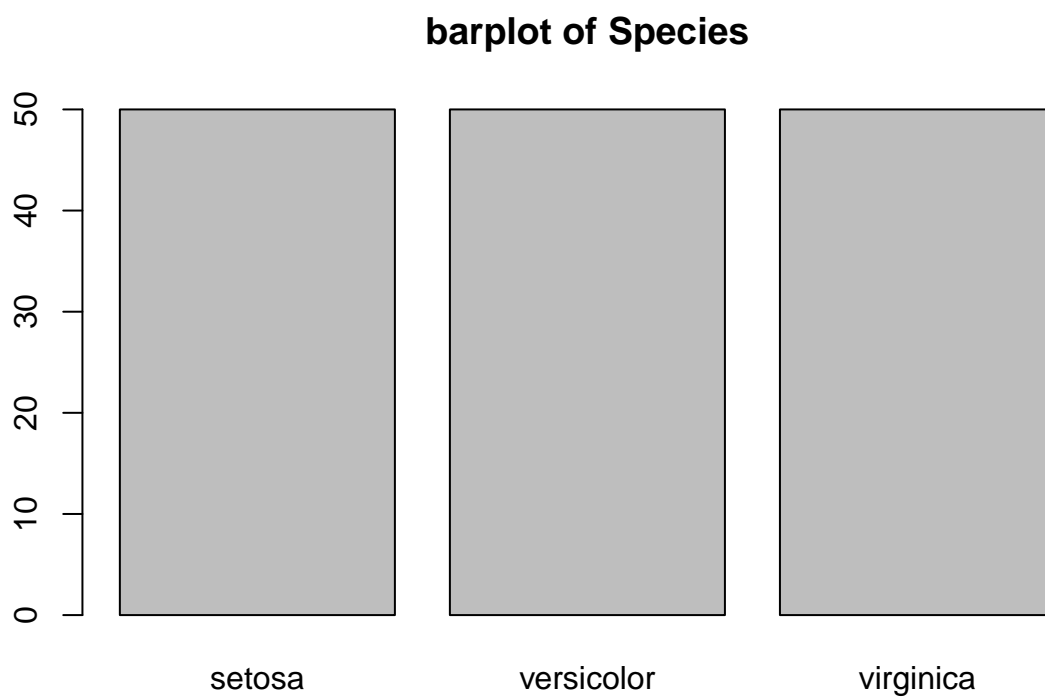
- boxplot

```r
par(mfrow = c(2,2))
boxplot(SL,main="boxplot of Sepal Length")
boxplot(SW,main="boxplot of Sepal Width")
boxplot(PL,main="boxplot of Petal Length")
boxplot(PW,main="boxplot of Petal Width")
```

boxplot of Sepal Length

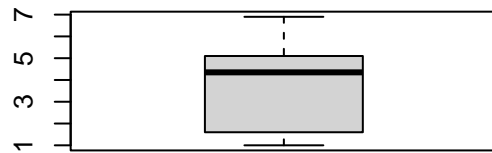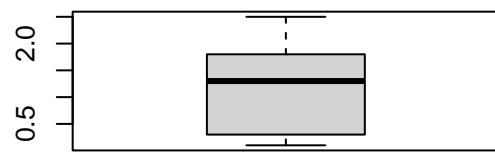## boxplot of Sepal Width

boxplot of Petal Length

boxplot of Petal Width

```r
par(mfrow = c(2,2))
boxplot(SL ~ Species,ylab="Sepal.Length",main="boxplot of Sepal Length")
boxplot(SW ~ Species,ylab="Sepal.Width",main="boxplot of Sepal Width")
boxplot(PL ~ Species,ylab="Petal.Length",main="boxplot of Petal Length")
boxplot(PW ~ Species,ylab="Petal.Width",main="boxplot of Petal Width")
```

boxplot of Sepal Length

boxplot of Sepal Width

boxplot of Petal Length

boxplot of Petal Width
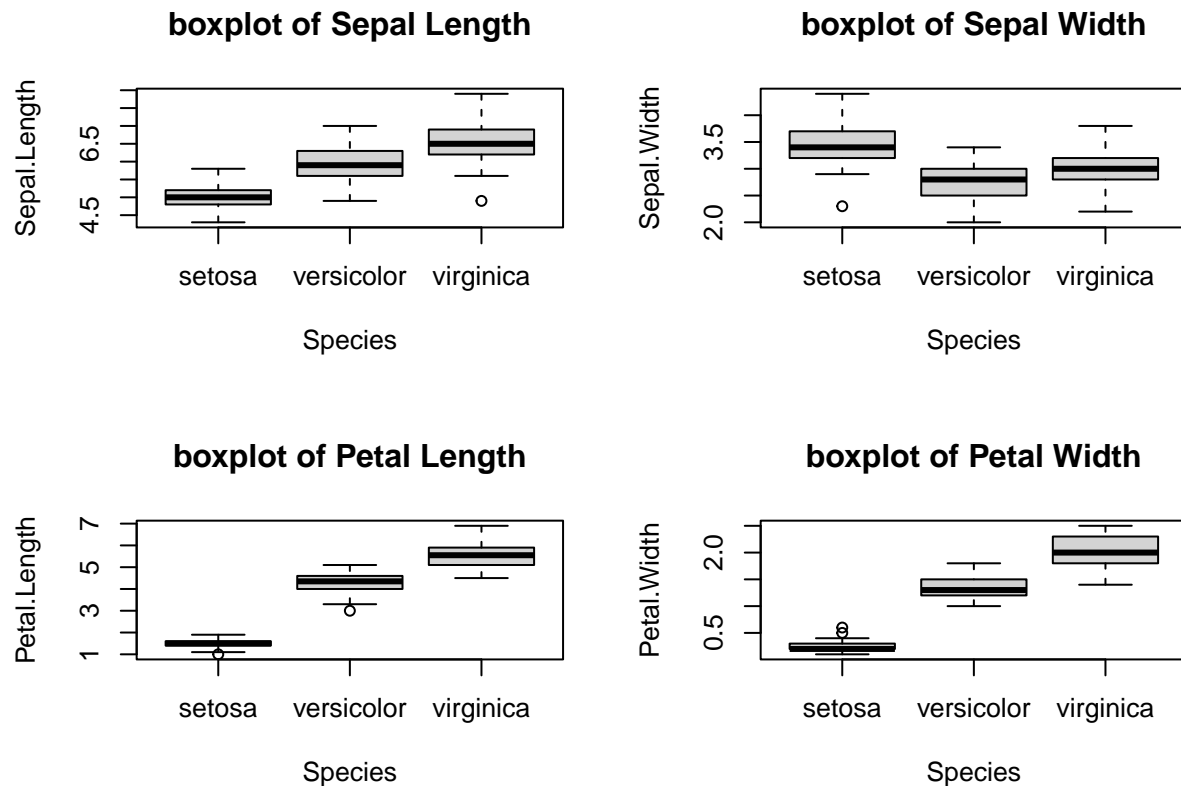
3. Use a table to summarize variables with the descriptive statistics including mean, median, variance, standard deviation, range, **IQR**, skewness, and kurtosis.

```r
library(fBasics)
sum=round(cbind(basicStats(SL),basicStats(SW),basicStats(PL),basicStats(PW)),4)
colnames(sum)=c("Sepal Length","Sepal Width","Petal Length","Petal Width")

#range
range=sum[4,]-sum[3,] #max-min
rownames(range)="range"

#IQR
IQR=sum[6,]-sum[5,] #Q3-Q1
rownames(IQR)="IQR"

summarize=rbind(sum[7,],sum[8,],sum[13,],sum[14,],
                range,sum[5,],sum[6,],IQR,sum[15,],sum[16,])
summarize
```

```
##            Sepal Length Sepal Width Petal Length Petal Width
## Mean             5.8433      3.0573       3.7580      1.1993
## Median           5.8000      3.0000       4.3500      1.3000
## Variance         0.6857      0.1900       3.1163      0.5810
## Stdev            0.8281      0.4359       1.7653      0.7622
```

```
## range                3.6000       2.4000       5.9000       2.4000
## 1. Quartile          5.1000       2.8000       1.6000       0.3000
## 3. Quartile          6.4000       3.3000       5.1000       1.8000
## IQR                  1.3000       0.5000       3.5000       1.5000
## Skewness             0.3086       0.3126      -0.2694      -0.1009
## Kurtosis            -0.6058       0.1387      -1.4169      -1.3582
```

Kurtosis can also be represented by the following answer:

```
library(moments)
round(cbind(kurtosis(SL),kurtosis(SW),kurtosis(PL),kurtosis(PW)),4)
```

```
##        [,1]  [,2]   [,3]   [,4]
## [1,] 2.4264 3.181 1.6045 1.6639
```

4. Are all cardinal variables symmetric?

Observing the figures of question 2, we find that Petal.Length and Petal.Width are obviously asymmetric and have double peaks. According to the skewness results in the third question, it is found that the skewness of Sepal.Length and Sepal Width are both positive, indicating a slight right skewness, while the skewness of Petal.Length and Petal.Width are both negative, indicating a slight left skewness.

5. Is there any outlier?

Sepal.Width has four outliers with values 4.4, 4.1, 4.2, and 2.0.

```
boxplot(SL ,plot=F)$out
```

```
## numeric(0)
```

```
boxplot(SW ,plot=F)$out
```

```
## [1] 4.4 4.1 4.2 2.0
```

```
boxplot(PL ,plot=F)$out
```

```
## numeric(0)
```

```
boxplot(PW ,plot=F)$out
```

```
## numeric(0)
```