

# ML Assignment 1 By Arvind Chaurasia

2023-09-24

**1. Download a dataset from the web. You may use any source, but specify the source in your code. Also ensure that the data has a mix of quantitative and qualitative (categorical) variables.**

I have downloaded the Iris dataset directly from R because it's one of the built-in datasets that come with R. Here's how to access and load the Iris dataset in R:

## 2. Import the dataset into R

Since it's included as one of the built-in datasets in R. We can load it directly using the `data()` function in the R programming.

```
# In order to load the iris dataset.
data(iris)

# head command will import and open a data viewer window for first few rows (6) of Iris dataset
head(iris)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5         1.4         0.2   setosa
## 2           4.9         3.0         1.4         0.2   setosa
## 3           4.7         3.2         1.3         0.2   setosa
## 4           4.6         3.1         1.5         0.2   setosa
## 5           5.0         3.6         1.4         0.2   setosa
## 6           5.4         3.9         1.7         0.4   setosa
```

## 3. Print out descriptive statistics for a selection of quantitative and categorical variables

We will use the following code in order to print the descriptive statistics for a selection of quantitative and categorical variables.

```
# Summary statistics for quantitative variables in our dataset iris.

quantitative_vars <- iris[, c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width")]
summary(quantitative_vars)
```

```
##      Sepal.Length      Sepal.Width      Petal.Length      Petal.Width
## Min.      :4.300    Min.      :2.000    Min.      :1.000    Min.      :0.100
## 1st Qu.:5.100    1st Qu.:2.800    1st Qu.:1.600    1st Qu.:0.300
## Median :5.800    Median :3.000    Median :4.350    Median :1.300
## Mean     :5.843    Mean     :3.057    Mean     :3.758    Mean     :1.199
## 3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100    3rd Qu.:1.800
## Max.     :7.900    Max.     :4.400    Max.     :6.900    Max.     :2.500
```

```
# Now coming to the second part of the question, we will print the frequency table for the categorical variable "Species"
cat_var <- iris$Species
table(cat_var)
```

```
## cat_var
##      setosa versicolor  virginica
##           50           50           50
```

#### 4. Transform at least one variable. It doesn't matter what the transformation is.

For this we will reduce the value of "Sepal.Length" variables to half of its value.

```
# Half the values in the "Sepal.Length" variable and store the result in a new variable
iris$Half.Sepal.Length <- iris$Sepal.Length/2

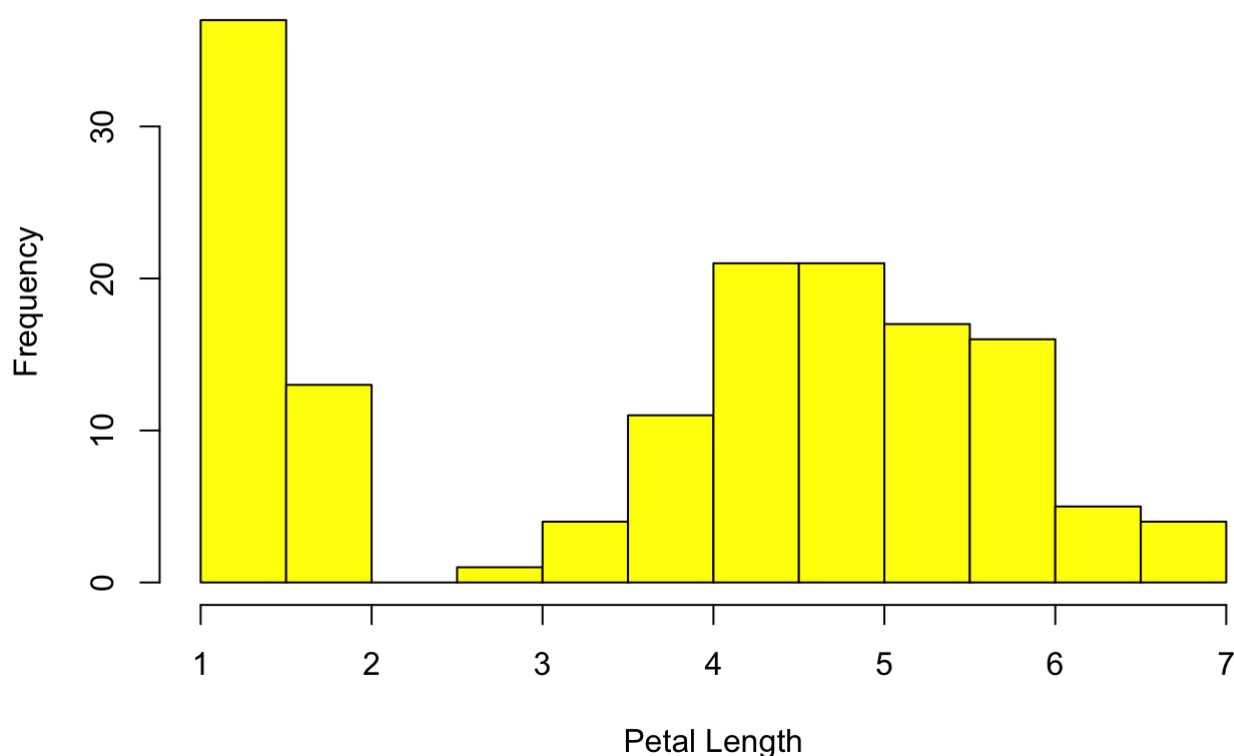
# View the first few rows of the modified dataset
head(iris)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species Half.Sepal.Length
## 1           5.1         3.5         1.4         0.2   setosa           2.55
## 2           4.9         3.0         1.4         0.2   setosa           2.45
## 3           4.7         3.2         1.3         0.2   setosa           2.35
## 4           4.6         3.1         1.5         0.2   setosa           2.30
## 5           5.0         3.6         1.4         0.2   setosa           2.50
## 6           5.4         3.9         1.7         0.4   setosa           2.70
```

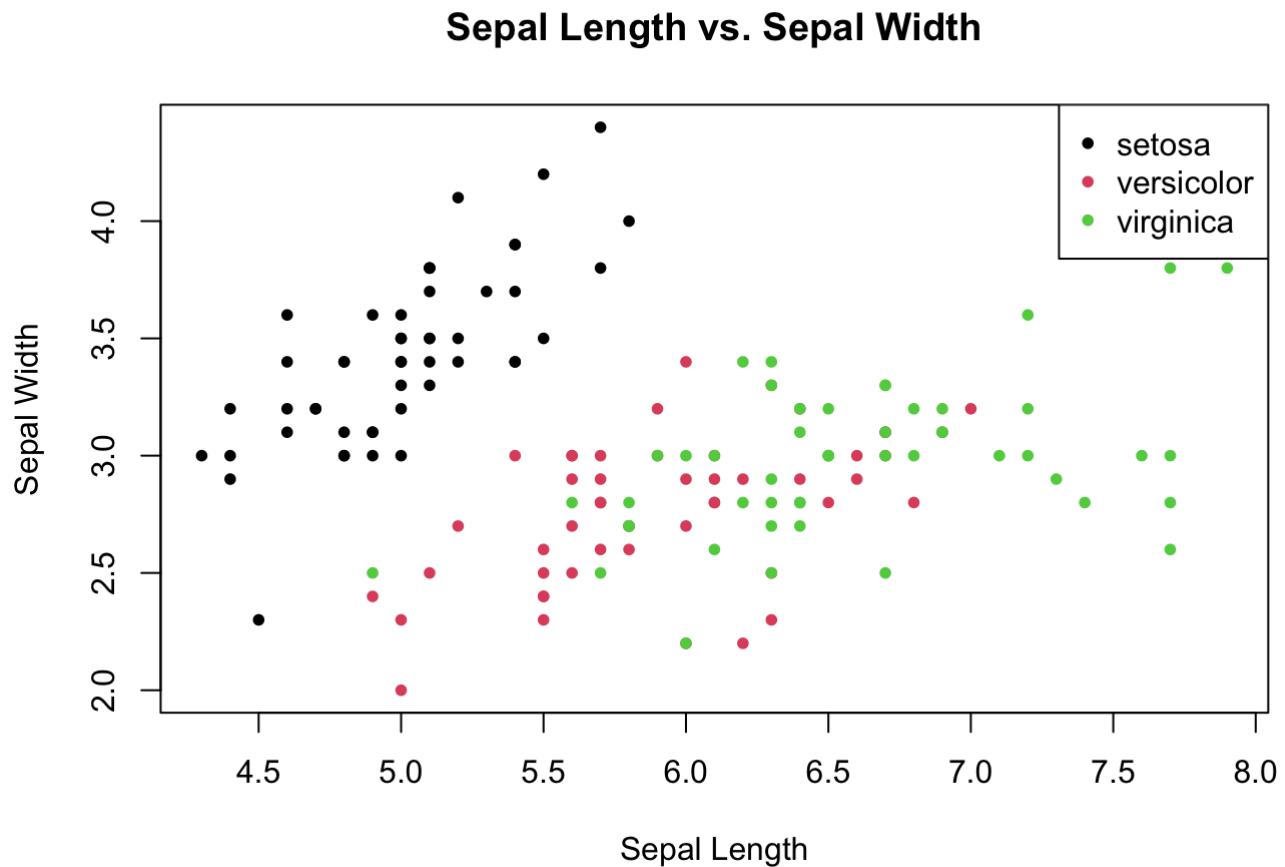
#### 5. Plot at least one quantitative variable, and one scatterplot.

```
# Plot a histogram of the "Petal.Length" variable
hist(iris$Petal.Length, col = "yellow", main = "Histogram of Petal Length", xlab = "Petal Length", ylab = "Frequency")
```

### Histogram of Petal Length



```
# Create a scatterplot of "Sepal.Length" against "Sepal.Width"
plot(iris$Sepal.Length, iris$Sepal.Width, col = iris$Species, pch = 20,
     main = "Sepal Length vs. Sepal Width", xlab = "Sepal Length", ylab = "Sepal Width")
legend("topright", legend = levels(iris$Species), col = 1:3, pch = 20)
```



**6. Upload your R program, and any associated datafiles to your git account. Remember to create a separate repository for this class.**

**7. Paste the address to your repository in the assignment submission box here in Canvas.**

The assignment was completed and uploaded on Github by Arvind Chaurasia. (Email: [acahauras@kent.edu](mailto:acahauras@kent.edu) (<mailto:acahauras@kent.edu>)) The coding and documentation was completed in Rmarkdown.