

my_first_Rhomework

Shaista Fathima

r Sys.Date()‘

R Markdown Homework 1

```
library(dplyr)
library(datasets)
install.packages("Hmisc", repos = "http://cran.us.r-project.org")
```

The following packages were installed to create this data sheet

```
##
## The downloaded binary packages are in
## /var/folders/_h/km41059x4g71gbfyr_y54qx80000gn/T//Rtmpi07Cwz/downloaded_packages
```

```
install.packages("ggpubr", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
## /var/folders/_h/km41059x4g71gbfyr_y54qx80000gn/T//Rtmpi07Cwz/downloaded_packages
```

```
install.packages("ggplot2", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
## /var/folders/_h/km41059x4g71gbfyr_y54qx80000gn/T//Rtmpi07Cwz/downloaded_packages
```

```
install.packages("contribution", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
## /var/folders/_h/km41059x4g71gbfyr_y54qx80000gn/T//Rtmpi07Cwz/downloaded_packages
```

```
# tinytex::install_tinytex(force = TRUE)
```

```
library(psych)
```

```
pacman::p_load(pacman, dplyr, GGally, ggplot, ggthemes, gvis, httr,
lubridate, plotly, rio, rmarkdown, shiny, stringr, tidyr)
```

```
## Warning: package 'ggplot' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages

## Warning in p_install(package, character.only = TRUE, ...):

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'ggplot'

## Warning: package 'gvis' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages

## Warning in p_install(package, character.only = TRUE, ...):

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'gvis'

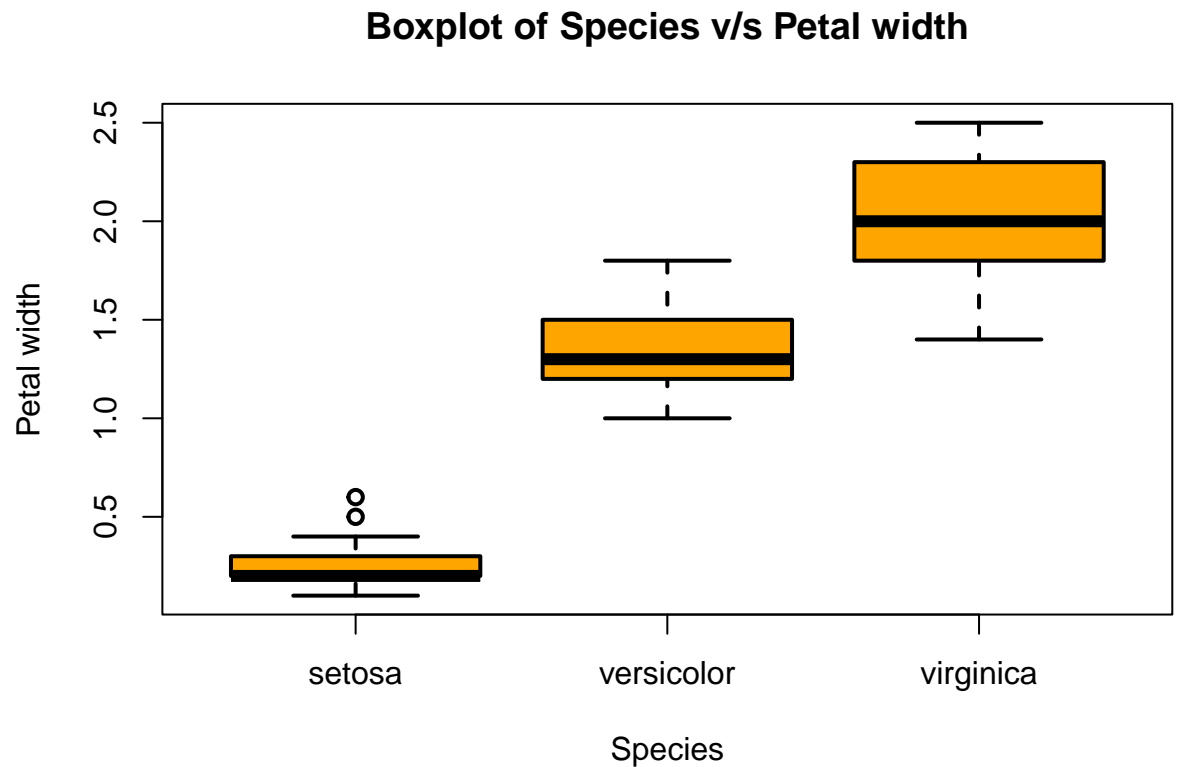
## Warning in pacman::p_load(pacman, dplyr, GGally, ggplot, ggthemes, gvis, : Failed to install/load:
## ggplot, gvis
```

Packages used for coding: tools, tcltk, stats4, splines, parallel, grid, compiler, stats, graphics, grDevices, utils, datasets, methods, base,

Performing checks to determine quality of data (missing values, outliers)

```
plot(iris$Species, iris$Petal.Width, col="orange", xlab= "Species",
ylab= "Petal width", main= "Boxplot of Species v/s Petal width", lwd =
2, )
```

Adding a boxplot to identify Outliers which is shows the relationship btw Numerical and Cate-



gorical data.

#####—How to identify missing values—(Note: iris data set din't originally have a missing value, so one value was created by duplicating iris data set to iris 2).

```
iris1 <- iris
iris1[1,1] <- NA
iris1[7,4] <-NA
```

Creating Missing values

```
iris1[!complete.cases(iris1),]
```

Identifying and locating the Missing Variables

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          NA          3.5          1.4          0.2 setosa
## 7          4.6          3.4          1.4          NA setosa
```

#####Checking for duplicate rows

Choosing a dataset and analyzing the type of data: individual elements (no of observations, variables and its type)

```
data()
data("iris")
str(iris)
```

```
## 'data.frame': 150 obs. of 5 variables:
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

#Viewing top and bottom rows of data frame

```
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
```

```
tail(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 145 6.7 3.3 5.7 2.5 virginica
## 146 6.7 3.0 5.2 2.3 virginica
## 147 6.3 2.5 5.0 1.9 virginica
## 148 6.5 3.0 5.2 2.0 virginica
## 149 6.2 3.4 5.4 2.3 virginica
## 150 5.9 3.0 5.1 1.8 virginica
```

###Shape of the data frame

```
dim(iris)
```

```
## [1] 150 5
```

###Number of rows and columns in the data frame

```
# data(iris)
nrow(iris)
```

```
## [1] 150
```

```
ncol(iris)
```

```
## [1] 5
```

Duplicated rows in the iris dataset

```
duplicated(iris$Sepal.Length)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE TRUE TRUE FALSE
## [13] TRUE FALSE FALSE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [25] TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE
## [37] TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE
## [49] FALSE TRUE FALSE FALSE FALSE TRUE FALSE TRUE FALSE TRUE FALSE TRUE
## [61] TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE TRUE TRUE TRUE
## [73] TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [85] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [97] TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE FALSE TRUE FALSE
## [109] TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE
## [121] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE
## [133] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [145] TRUE TRUE TRUE TRUE TRUE TRUE
```

```
iris$Sepal.Length[duplicated(iris$Sepal.Length)]
```

```
## [1] 4.6 5.0 4.9 5.4 4.8 5.4 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 4.7
## [19] 4.8 5.4 5.2 4.9 5.0 5.5 4.9 4.4 5.1 5.0 4.4 5.0 5.1 4.8 5.1 4.6 5.0 5.5
## [37] 5.7 4.9 5.2 5.0 5.6 5.8 5.6 5.9 6.1 6.3 6.1 6.4 6.6 6.7 6.0 5.7 5.5 5.5
## [55] 5.8 6.0 5.4 6.0 6.7 6.3 5.6 5.5 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7
## [73] 6.3 5.8 6.3 6.5 4.9 6.7 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 6.0 6.9 5.6 7.7
## [91] 6.3 6.7 7.2 6.2 6.1 6.4 7.2 6.4 6.3 6.1 7.7 6.3 6.4 6.0 6.9 6.7 6.9 5.8
## [109] 6.8 6.7 6.7 6.3 6.5 6.2 5.9
```

```
b <- iris$Sepal.Length[!duplicated(iris$Sepal.Length)]
```

```
duplicated(iris)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE
```

```
which(duplicated(iris))
```

```
## [1] 143
```

```
subset(iris, Species == "setosa")[1:50,]
```

Counting the number of categorical variables in each subset. Filtering variables and observations in the dataset

##	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
## 7	4.6	3.4	1.4	0.3	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 9	4.4	2.9	1.4	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa
## 11	5.4	3.7	1.5	0.2	setosa
## 12	4.8	3.4	1.6	0.2	setosa
## 13	4.8	3.0	1.4	0.1	setosa
## 14	4.3	3.0	1.1	0.1	setosa
## 15	5.8	4.0	1.2	0.2	setosa
## 16	5.7	4.4	1.5	0.4	setosa
## 17	5.4	3.9	1.3	0.4	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 19	5.7	3.8	1.7	0.3	setosa
## 20	5.1	3.8	1.5	0.3	setosa
## 21	5.4	3.4	1.7	0.2	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 23	4.6	3.6	1.0	0.2	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 34	5.5	4.2	1.4	0.2	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 39	4.4	3.0	1.3	0.2	setosa
## 40	5.1	3.4	1.5	0.2	setosa

```
## 41      5.0      3.5      1.3      0.3 setosa
## 42      4.5      2.3      1.3      0.3 setosa
## 43      4.4      3.2      1.3      0.2 setosa
## 44      5.0      3.5      1.6      0.6 setosa
## 45      5.1      3.8      1.9      0.4 setosa
## 46      4.8      3.0      1.4      0.3 setosa
## 47      5.1      3.8      1.6      0.2 setosa
## 48      4.6      3.2      1.4      0.2 setosa
## 49      5.3      3.7      1.5      0.2 setosa
## 50      5.0      3.3      1.4      0.2 setosa
```

```
subset(iris, Species == "versicolor")[1:50,]
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
## 51          7.0         3.2         4.7         1.4 versicolor
## 52          6.4         3.2         4.5         1.5 versicolor
## 53          6.9         3.1         4.9         1.5 versicolor
## 54          5.5         2.3         4.0         1.3 versicolor
## 55          6.5         2.8         4.6         1.5 versicolor
## 56          5.7         2.8         4.5         1.3 versicolor
## 57          6.3         3.3         4.7         1.6 versicolor
## 58          4.9         2.4         3.3         1.0 versicolor
## 59          6.6         2.9         4.6         1.3 versicolor
## 60          5.2         2.7         3.9         1.4 versicolor
## 61          5.0         2.0         3.5         1.0 versicolor
## 62          5.9         3.0         4.2         1.5 versicolor
## 63          6.0         2.2         4.0         1.0 versicolor
## 64          6.1         2.9         4.7         1.4 versicolor
## 65          5.6         2.9         3.6         1.3 versicolor
## 66          6.7         3.1         4.4         1.4 versicolor
## 67          5.6         3.0         4.5         1.5 versicolor
## 68          5.8         2.7         4.1         1.0 versicolor
## 69          6.2         2.2         4.5         1.5 versicolor
## 70          5.6         2.5         3.9         1.1 versicolor
## 71          5.9         3.2         4.8         1.8 versicolor
## 72          6.1         2.8         4.0         1.3 versicolor
## 73          6.3         2.5         4.9         1.5 versicolor
## 74          6.1         2.8         4.7         1.2 versicolor
## 75          6.4         2.9         4.3         1.3 versicolor
## 76          6.6         3.0         4.4         1.4 versicolor
## 77          6.8         2.8         4.8         1.4 versicolor
## 78          6.7         3.0         5.0         1.7 versicolor
## 79          6.0         2.9         4.5         1.5 versicolor
## 80          5.7         2.6         3.5         1.0 versicolor
## 81          5.5         2.4         3.8         1.1 versicolor
## 82          5.5         2.4         3.7         1.0 versicolor
## 83          5.8         2.7         3.9         1.2 versicolor
## 84          6.0         2.7         5.1         1.6 versicolor
## 85          5.4         3.0         4.5         1.5 versicolor
## 86          6.0         3.4         4.5         1.6 versicolor
## 87          6.7         3.1         4.7         1.5 versicolor
## 88          6.3         2.3         4.4         1.3 versicolor
## 89          5.6         3.0         4.1         1.3 versicolor
## 90          5.5         2.5         4.0         1.3 versicolor
```

## 91	5.5	2.6	4.4	1.2	versicolor
## 92	6.1	3.0	4.6	1.4	versicolor
## 93	5.8	2.6	4.0	1.2	versicolor
## 94	5.0	2.3	3.3	1.0	versicolor
## 95	5.6	2.7	4.2	1.3	versicolor
## 96	5.7	3.0	4.2	1.2	versicolor
## 97	5.7	2.9	4.2	1.3	versicolor
## 98	6.2	2.9	4.3	1.3	versicolor
## 99	5.1	2.5	3.0	1.1	versicolor
## 100	5.7	2.8	4.1	1.3	versicolor

```
subset(iris, Species == "virginica")[1:50,]
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 101	6.3	3.3	6.0	2.5	virginica
## 102	5.8	2.7	5.1	1.9	virginica
## 103	7.1	3.0	5.9	2.1	virginica
## 104	6.3	2.9	5.6	1.8	virginica
## 105	6.5	3.0	5.8	2.2	virginica
## 106	7.6	3.0	6.6	2.1	virginica
## 107	4.9	2.5	4.5	1.7	virginica
## 108	7.3	2.9	6.3	1.8	virginica
## 109	6.7	2.5	5.8	1.8	virginica
## 110	7.2	3.6	6.1	2.5	virginica
## 111	6.5	3.2	5.1	2.0	virginica
## 112	6.4	2.7	5.3	1.9	virginica
## 113	6.8	3.0	5.5	2.1	virginica
## 114	5.7	2.5	5.0	2.0	virginica
## 115	5.8	2.8	5.1	2.4	virginica
## 116	6.4	3.2	5.3	2.3	virginica
## 117	6.5	3.0	5.5	1.8	virginica
## 118	7.7	3.8	6.7	2.2	virginica
## 119	7.7	2.6	6.9	2.3	virginica
## 120	6.0	2.2	5.0	1.5	virginica
## 121	6.9	3.2	5.7	2.3	virginica
## 122	5.6	2.8	4.9	2.0	virginica
## 123	7.7	2.8	6.7	2.0	virginica
## 124	6.3	2.7	4.9	1.8	virginica
## 125	6.7	3.3	5.7	2.1	virginica
## 126	7.2	3.2	6.0	1.8	virginica
## 127	6.2	2.8	4.8	1.8	virginica
## 128	6.1	3.0	4.9	1.8	virginica
## 129	6.4	2.8	5.6	2.1	virginica
## 130	7.2	3.0	5.8	1.6	virginica
## 131	7.4	2.8	6.1	1.9	virginica
## 132	7.9	3.8	6.4	2.0	virginica
## 133	6.4	2.8	5.6	2.2	virginica
## 134	6.3	2.8	5.1	1.5	virginica
## 135	6.1	2.6	5.6	1.4	virginica
## 136	7.7	3.0	6.1	2.3	virginica
## 137	6.3	3.4	5.6	2.4	virginica
## 138	6.4	3.1	5.5	1.8	virginica
## 139	6.0	3.0	4.8	1.8	virginica
## 140	6.9	3.1	5.4	2.1	virginica


```
## 141      6.7      3.1      5.6      2.4 virginica
## 142      6.9      3.1      5.1      2.3 virginica
## 143      5.8      2.7      5.1      1.9 virginica
## 144      6.8      3.2      5.9      2.3 virginica
## 145      6.7      3.3      5.7      2.5 virginica
## 146      6.7      3.0      5.2      2.3 virginica
## 147      6.3      2.5      5.0      1.9 virginica
## 148      6.5      3.0      5.2      2.0 virginica
## 149      6.2      3.4      5.4      2.3 virginica
## 150      5.9      3.0      5.1      1.8 virginica
```

```
iris$"Sepal.Length"
```

Counting the number of numeric variables

```
## [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7 5.4 5.1
## [19] 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5 4.9 5.0
## [37] 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0 6.4 6.9 5.5
## [55] 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6 6.7 5.6 5.8 6.2 5.6 5.9 6.1
## [73] 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7 5.5 5.5 5.8 6.0 5.4 6.0 6.7 6.3 5.6 5.5
## [91] 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3 5.8 7.1 6.3 6.5 7.6 4.9 7.3
## [109] 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 7.7 6.0 6.9 5.6 7.7 6.3 6.7 7.2
## [127] 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7 6.3 6.4 6.0 6.9 6.7 6.9 5.8 6.8
## [145] 6.7 6.7 6.3 6.5 6.2 5.9
```

```
iris$"Sepal.Width"
```

```
## [1] 3.5 3.0 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 3.7 3.4 3.0 3.0 4.0 4.4 3.9 3.5
## [19] 3.8 3.8 3.4 3.7 3.6 3.3 3.4 3.0 3.4 3.5 3.4 3.2 3.1 3.4 4.1 4.2 3.1 3.2
## [37] 3.5 3.6 3.0 3.4 3.5 2.3 3.2 3.5 3.8 3.0 3.8 3.2 3.7 3.3 3.2 3.2 3.1 2.3
## [55] 2.8 2.8 3.3 2.4 2.9 2.7 2.0 3.0 2.2 2.9 2.9 3.1 3.0 2.7 2.2 2.5 3.2 2.8
## [73] 2.5 2.8 2.9 3.0 2.8 3.0 2.9 2.6 2.4 2.4 2.7 2.7 3.0 3.4 3.1 2.3 3.0 2.5
## [91] 2.6 3.0 2.6 2.3 2.7 3.0 2.9 2.9 2.5 2.8 3.3 2.7 3.0 2.9 3.0 3.0 2.5 2.9
## [109] 2.5 3.6 3.2 2.7 3.0 2.5 2.8 3.2 3.0 3.8 2.6 2.2 3.2 2.8 2.8 2.7 3.3 3.2
## [127] 2.8 3.0 2.8 3.0 2.8 3.8 2.8 2.8 2.6 3.0 3.4 3.1 3.0 3.1 3.1 3.1 2.7 3.2
## [145] 3.3 3.0 2.5 3.0 3.4 3.0
```

```
iris$"Petal.Length"
```

```
## [1] 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 1.5 1.6 1.4 1.1 1.2 1.5 1.3 1.4
## [19] 1.7 1.5 1.7 1.5 1.0 1.7 1.9 1.6 1.6 1.5 1.4 1.6 1.6 1.5 1.5 1.4 1.5 1.2
## [37] 1.3 1.4 1.3 1.5 1.3 1.3 1.3 1.6 1.9 1.4 1.6 1.4 1.5 1.4 4.7 4.5 4.9 4.0
## [55] 4.6 4.5 4.7 3.3 4.6 3.9 3.5 4.2 4.0 4.7 3.6 4.4 4.5 4.1 4.5 3.9 4.8 4.0
## [73] 4.9 4.7 4.3 4.4 4.8 5.0 4.5 3.5 3.8 3.7 3.9 5.1 4.5 4.5 4.7 4.4 4.1 4.0
## [91] 4.4 4.6 4.0 3.3 4.2 4.2 4.2 4.3 3.0 4.1 6.0 5.1 5.9 5.6 5.8 6.6 4.5 6.3
## [109] 5.8 6.1 5.1 5.3 5.5 5.0 5.1 5.3 5.5 6.7 6.9 5.0 5.7 4.9 6.7 4.9 5.7 6.0
## [127] 4.8 4.9 5.6 5.8 6.1 6.4 5.6 5.1 5.6 6.1 5.6 5.5 4.8 5.4 5.6 5.1 5.1 5.9
## [145] 5.7 5.2 5.0 5.2 5.4 5.1
```

```
iris$"Petal.Width"
```

```
##      [1] 0.2 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 0.2 0.2 0.1 0.1 0.2 0.4 0.4 0.3
##     [19] 0.3 0.3 0.2 0.4 0.2 0.5 0.2 0.2 0.4 0.2 0.2 0.2 0.2 0.4 0.1 0.2 0.2 0.2 0.2
##     [37] 0.2 0.1 0.2 0.2 0.3 0.3 0.2 0.6 0.4 0.3 0.2 0.2 0.2 0.2 1.4 1.5 1.5 1.3
##     [55] 1.5 1.3 1.6 1.0 1.3 1.4 1.0 1.5 1.0 1.4 1.3 1.4 1.5 1.0 1.5 1.1 1.8 1.3
##     [73] 1.5 1.2 1.3 1.4 1.4 1.7 1.5 1.0 1.1 1.0 1.2 1.6 1.5 1.6 1.5 1.3 1.3 1.3
##     [91] 1.2 1.4 1.2 1.0 1.3 1.2 1.3 1.3 1.1 1.3 2.5 1.9 2.1 1.8 2.2 2.1 1.7 1.8
##    [109] 1.8 2.5 2.0 1.9 2.1 2.0 2.4 2.3 1.8 2.2 2.3 1.5 2.3 2.0 2.0 1.8 2.1 1.8
##   [127] 1.8 1.8 2.1 1.6 1.9 2.0 2.2 1.5 1.4 2.3 2.4 1.8 1.8 2.1 2.4 2.3 1.9 2.3
##   [145] 2.5 2.3 1.9 2.0 2.3 1.8
```

#####Summary of the data showing mix, max values, mean median and quartiles, Variance, standard deviation, skew

```
summary(iris)
```

1. Mode for each numerical Data column

```
##      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
##           Species
##   setosa      :50
##   versicolor:50
##   virginica   :50
##
##
##
```

```
summary (iris$Species)
```

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

```
mode <- function(x)
{unique_val <- unique(x)
counts <- vector()
for (i in 1:length(unique_val)){
counts[i] <- length(which(x== unique_val[i]))}

position <- c(which(counts == max(counts)))
if (length(unique_val)== length(x))
mode_x <- 'Mode does not exist'
else
mode_x <- unique_val[position]
return(mode_x)}
```

```
attach(iris)
summary(Sepal.Length)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  4.300   5.100   5.800   5.843   6.400   7.900
```

```
mode(Sepal.Length)
```

```
## [1] 5
```

```
mode(Sepal.Width)
```

```
## [1] 3
```

```
mode(Petal.Length)
```

```
## [1] 1.4 1.5
```

```
mode(Petal.Length)
```

```
## [1] 1.4 1.5
```

```
library(tidyverse)
library(psych)
```

```
describe(iris$Petal.Width)
```

2. Mean, Median and Standard Deviation for each data column

```
##      vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 150  1.2 0.76    1.3    1.18 1.04 0.1 2.5   2.4 -0.1    -1.36 0.06
```

```
describe(iris$Sepal.Length)
```

```
##      vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 150 5.84 0.83    5.8    5.81 1.04 4.3 7.9   3.6 0.31    -0.61 0.07
```

```
describe(iris$Petal.Length)
```

```
##      vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 150 3.76 1.77    4.35    3.76 1.85   1 6.9   5.9 -0.27    -1.42 0.14
```

```
describe(iris$Sepal.Width)
```

```
##      vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 150 3.06 0.44      3    3.04 0.44   2 4.4   2.4 0.31    0.14 0.04
```

Converting a numerical data to categorical variable.

Identifying similar rows

```
class(iris$Sepal.Length)
```

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

```
Sepal.Length.Category <- cut(iris$Sepal.Length,
breaks= c(0,4,5,6,7,8),
labels = c("First","Second","Third", "Fourth", "Fifth"))
iris$Sepal.Length.Category <- cut(iris$Sepal.Length,
breaks= c(0, 4,5,6,7,8),
labels = c("First","Second","Third", "Fourth", "Fifth"))
View(iris)
head(iris, 40)
```

```
##      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
## 7           4.6         3.4         1.4         0.3   setosa
## 8           5.0         3.4         1.5         0.2   setosa
## 9           4.4         2.9         1.4         0.2   setosa
## 10          4.9         3.1         1.5         0.1   setosa
## 11          5.4         3.7         1.5         0.2   setosa
## 12          4.8         3.4         1.6         0.2   setosa
## 13          4.8         3.0         1.4         0.1   setosa
## 14          4.3         3.0         1.1         0.1   setosa
## 15          5.8         4.0         1.2         0.2   setosa
## 16          5.7         4.4         1.5         0.4   setosa
## 17          5.4         3.9         1.3         0.4   setosa
## 18          5.1         3.5         1.4         0.3   setosa
## 19          5.7         3.8         1.7         0.3   setosa
## 20          5.1         3.8         1.5         0.3   setosa
## 21          5.4         3.4         1.7         0.2   setosa
## 22          5.1         3.7         1.5         0.4   setosa
## 23          4.6         3.6         1.0         0.2   setosa
## 24          5.1         3.3         1.7         0.5   setosa
## 25          4.8         3.4         1.9         0.2   setosa
## 26          5.0         3.0         1.6         0.2   setosa
```

## 27	5.0	3.4	1.6	0.4	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 34	5.5	4.2	1.4	0.2	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 39	4.4	3.0	1.3	0.2	setosa
## 40	5.1	3.4	1.5	0.2	setosa
##	Sepal.Length.Category				
## 1	Third				
## 2	Second				
## 3	Second				
## 4	Second				
## 5	Second				
## 6	Third				
## 7	Second				
## 8	Second				
## 9	Second				
## 10	Second				
## 11	Third				
## 12	Second				
## 13	Second				
## 14	Second				
## 15	Third				
## 16	Third				
## 17	Third				
## 18	Third				
## 19	Third				
## 20	Third				
## 21	Third				
## 22	Third				
## 23	Second				
## 24	Third				
## 25	Second				
## 26	Second				
## 27	Second				
## 28	Third				
## 29	Third				
## 30	Second				
## 31	Second				
## 32	Third				
## 33	Third				
## 34	Third				
## 35	Second				
## 36	Second				
## 37	Third				
## 38	Second				
## 39	Second				

40

Third

```
library(tidyverse)
```

Visualizations to illustrate relationship between variables, trend, distribution of variables.

Scatterplot, for visualizing the relationship between two numerical variables Petal length v/s Petal width

```
plot(iris$Petal.Length, iris$Petal.Width,  
     col = "maroon", pch = 19,  
     main = "Scatter plot of Petal Length v/s Petal Width",  
     xlab = "Petal Length", ylab = "Petal Width", lwd = 1)
```

Scatter plot of Petal Length v/s Petal Width



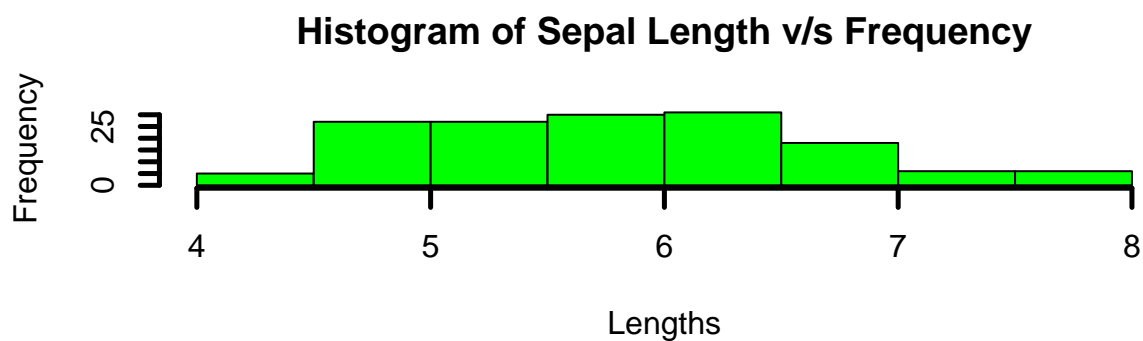
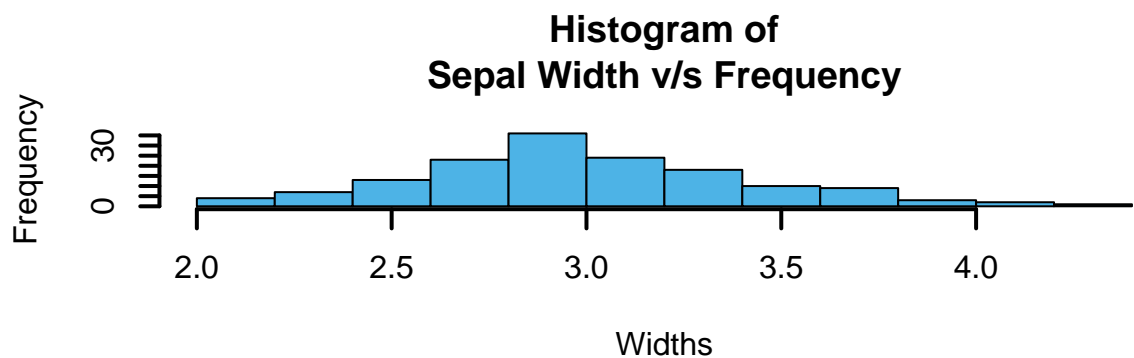
####Conclusion: #### Summary: Analyzing the scatter plot we can conclude that the Petal length is directly proportional to petal width, and greater the increase in its width and lengths, variation is and scatter is more pronounced. There is a larger variability in data when petal length exceeds 4cm mark. We can conclude that in this data set the maximum species of iris flowers fall above the 4cm mark in length and 1cm mark in width

Histogram for determining the relationship between each Sepal and Petal v/s their frequencies which represents an interesting trend.

```
par(mfrow = c(2,1))

hist(iris$Sepal.Width, xlab = "Widths", lwd = 2,
     main = "Histogram of
     Sepal Width v/s Frequency", col= rgb(0.3, 0.7, 0.9))

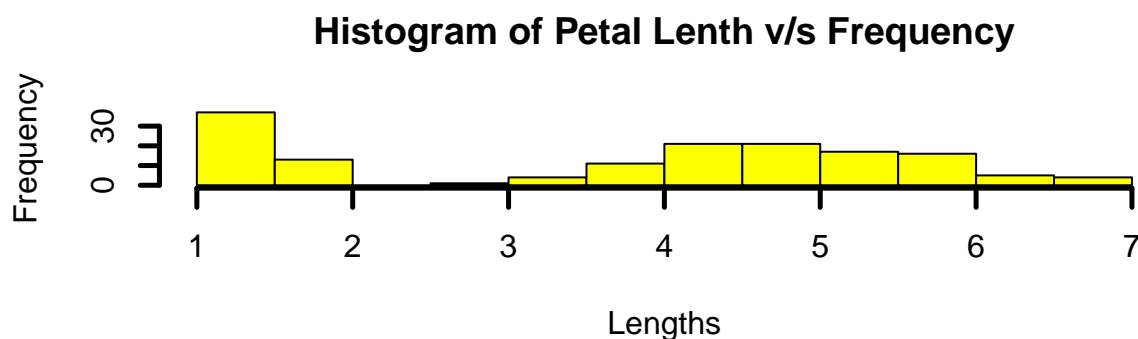
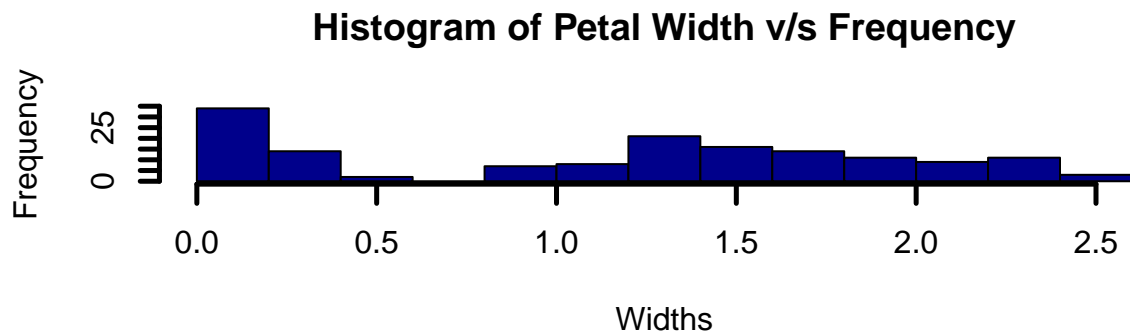
hist(iris$Sepal.Length, col = rgb(0, 1, 0),
     xlab = "Lengths", lwd = 2.5, main = "Histogram of Sepal Length v/s Frequency")
```



```
par(mfrow = c(2,1))

hist(iris$Petal.Width, col = rgb(0, 0, 139, max=255), xlab =
     "Widths", lwd = 2.5, main = "Histogram of Petal Width v/s Frequency")

hist(iris$Petal.Length, col= rgb(255, 255, 0, max=255), xlab =
     "Lengths", lwd = 2.5, main = "Histogram of Petal Lenth v/s Frequency")
```



Conclusion: **Summary:** From the histogram we can visualize that Sepal widths and Sepal lengths are maximum at 3cm and 6cm respectively. Hence, these sizes concatenate the maximum volume of iris species. Another point to note is that, the highest frequency of flowers fall in the range where Sepal length is twice that of Sepal width. The trend shapes into the form of a bell curve which shows an increasing frequency of parameters (length and width respectively) followed by a smooth decline. But when considering the petal widths and lengths, variability is high..and it does not correspond to the situation of a bell curve. The highest frequency of petal lengths is between 1-2 cm range, and petal width is maximum between 0 to 0.5

####Grouped Barchart showing the distribution of variables.

```
library(tidyverse)
library(ggsci)
library(ggpubr)
library(dplyr)
```

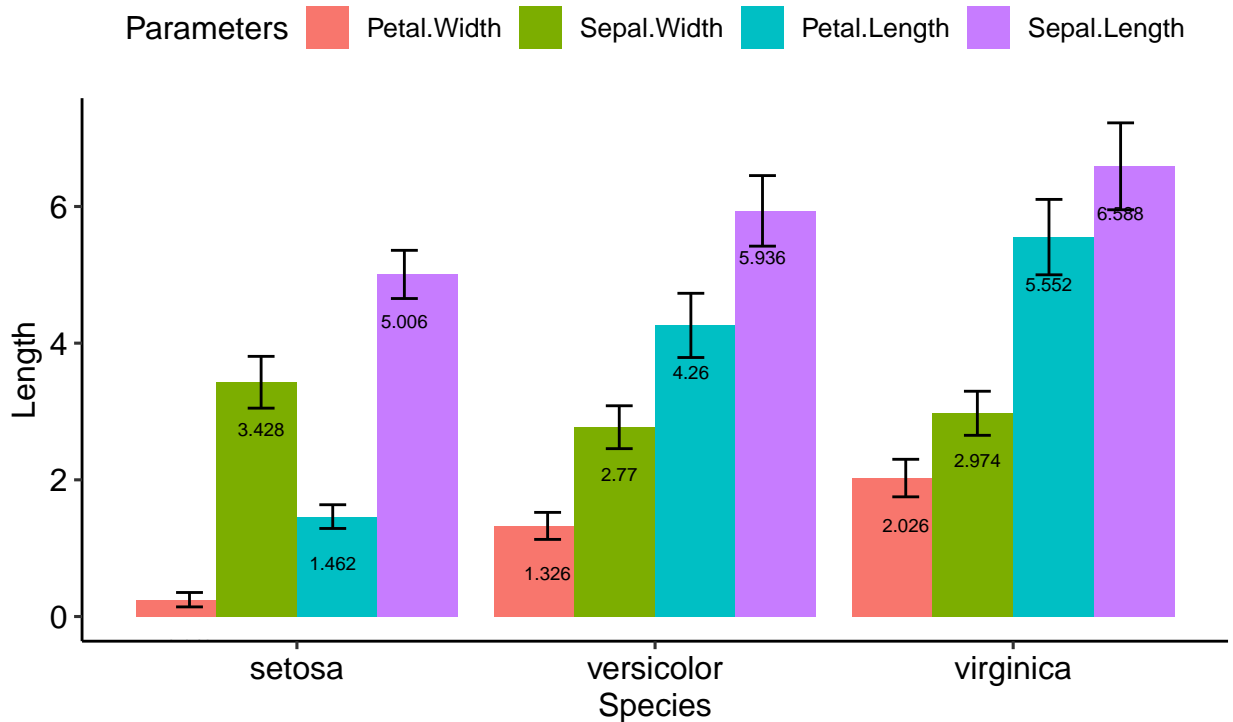
```
iris_plot <- iris %>%
  gather(FlPrt, Length, 1:4)%>%
  group_by(Species,FlPrt)%>%
  summarise( mnL=mean(Length),sdL=sd(Length))%>%
  ggplot(aes(reorder(Species,mnL), mnL ,fill= reorder(FlPrt,mnL)))+
  geom_bar(stat="identity", position="dodge")+
  geom_text(mapping=aes(label=mnL),
            position = position_dodge(width=0.8), cex=2.4, vjust=4)+
  labs(title="Distribution of the iris flower parameters across three species",
       subtitle = "Sepal and Petal lengths", x="Species",
       y = "Length", fill= "Parameters")+
```



```
geom_errorbar(mapping=aes(ymin=mnL-sdL, ymax=mnL+sdL),
              width=0.3, position=position_dodge(width=0.8))+
theme_pubr()
```

```
iris_plot
```

Distribution of the iris flower parameters across three species Sepal and Petal lengths



####Conclusion: ####Summary of Bar Plot: The x axis is distributed into three species of flowers, versus the y axis which shows different categories of lengths. The data corresponds that **virginica** is the largest iris species in size since its petal and sepal lengths are greater than **Setosa** and **versicolor**. The flower setosa seems to have a fairly small size (smaller lengths and widths). Hence, the graph shows a noticeable differentiation between the three categorical species. The margin of random sampling error shown in the iris flower population is indicated. The chances of error in sampling is greater for Petal and Sepal lengths. Lengths show greater variation in data.