PROBABILITY & STATISTICS – LAB 2

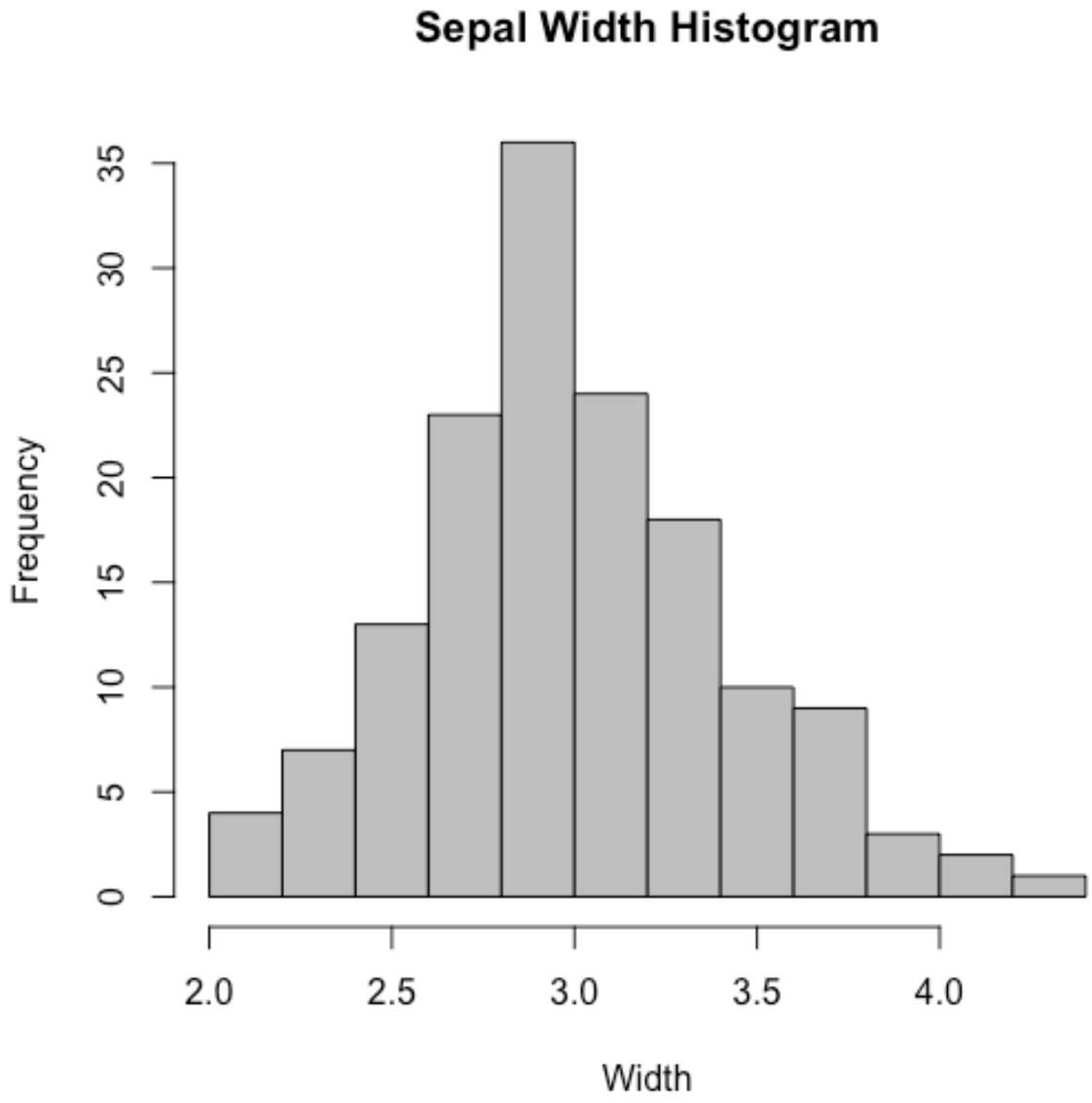
B.Tech. Computer Science and Engineering (Cybersecurity)

Name: Anish Sudhan Nair	Roll No.: K041
Batch: K2/A2	Date of performance: 06/01/2021

Aim: To work with charts and datasets in R.

1.Draw a histogram of sepal width from dataset iris. In Which range does the highest frequency occur?

```
data{"iris"}
iris
hist(iris$Sepal.Width, col ="grey" , main = "Sepal Width Histogram" , xlab ="Width", ylab="Frequency")
```



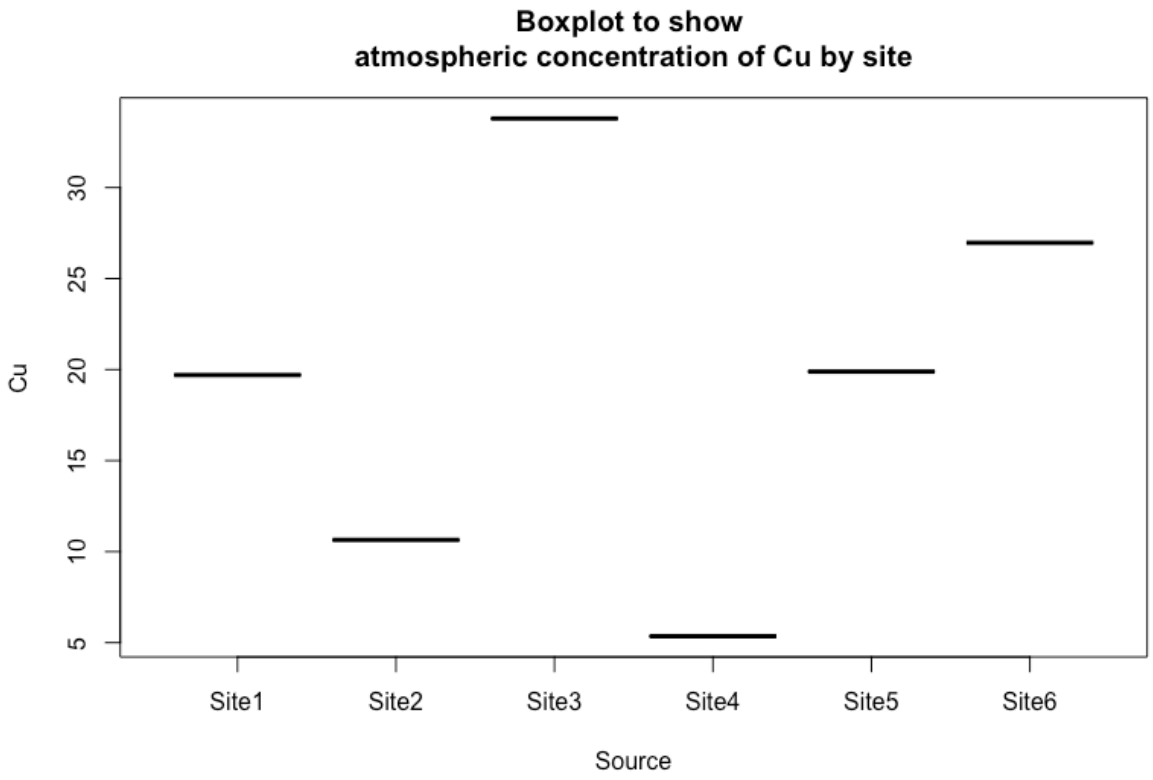
Range: 2.8-3.0 (Highest Frequency)

2. Create a csv file for the following data

Source	Cu
Site1	19.700
Site2	10.643
Site3	33.792
Site4	5.353
Site5	19.890
Site6	26.966

Draw a boxplot to show atmospheric copper concentration by sites.

```
data_cu<-read.csv("cu_atmos.csv")
data_cu
boxplot(data_cu$Cu ~ data_cu$Source, xlab = "Source",ylab = "Cu", main ="Boxplot to show
atmospheric concentration of Cu by site", col = "green",
        border = "black")
```

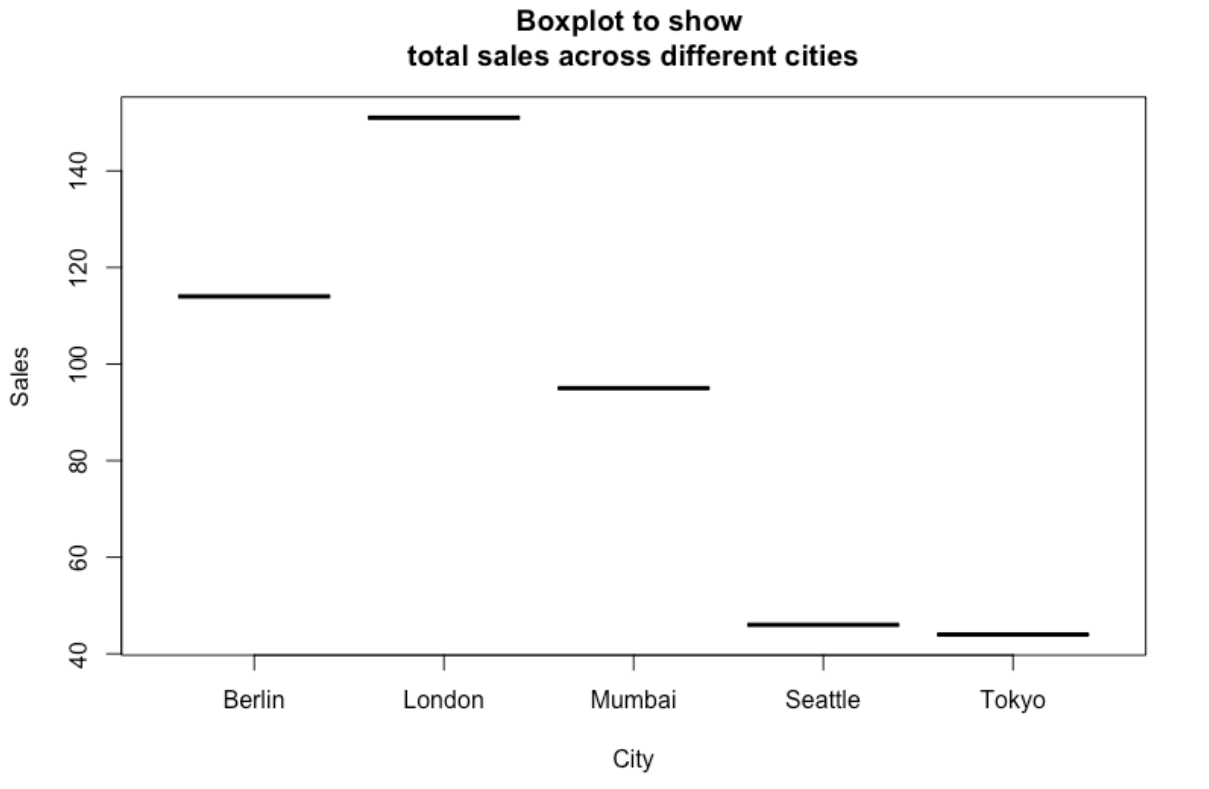


3. Create a csv file for the following data

##	City	ProductA	ProductB	ProductC
## 1	Seattle	23	11	12
## 2	London	89	6	56
## 3	Tokyo	24	7	13
## 4	Berlin	36	34	44
## 5	Mumbai	3	78	14

Draw a bar graph to show total sales across different cities

```
data_sales<-read.csv("city_sales.csv")
data_sales
boxplot(ProductA+ProductB+ProductC ~ City, data=data_sales, xlab = "City",ylab = "Sales", main ="Boxplot to show
total sales across different cities", col = "green",
        border = "black")
#same as data_sales$ProductA+data_sales$Productb+data_sales$ProductC ~ data_sales$City
```



4. Write R program to change the first level of a factor with another level of a given factor

CODE:

```
data = c(10,20,20,30,10,20,30,30,10,20,3,3,1)
fdata = factor(data)
print('The Original Levels:')
fdata
levels(fdata)[3]= 'a'
print('The Changed Levels:')
fdata
```

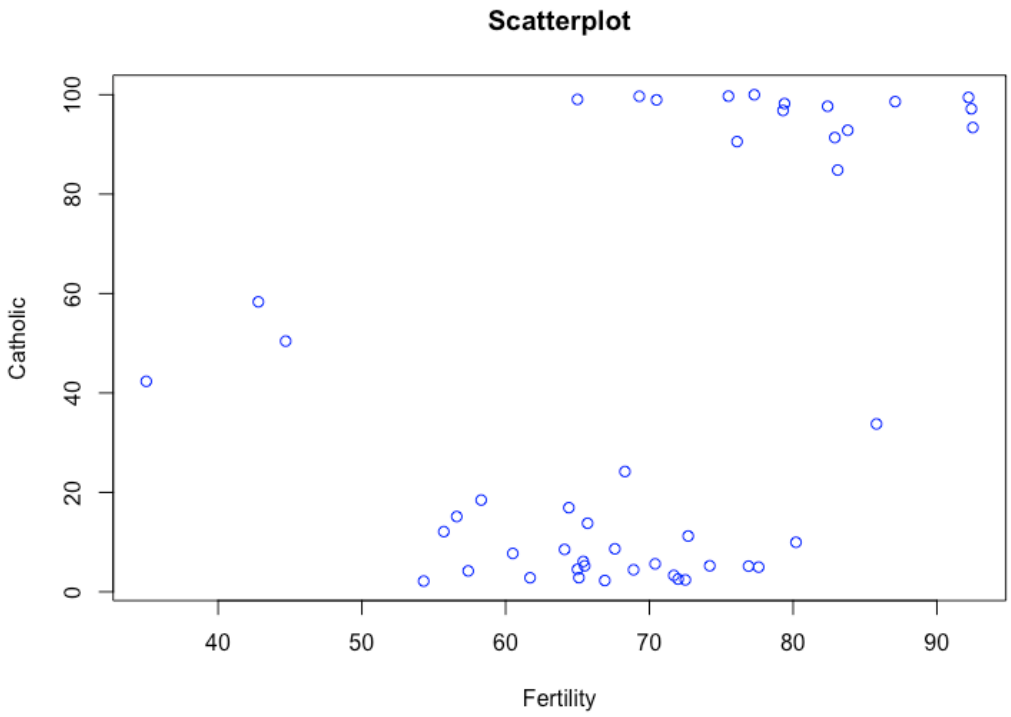
OUTPUT:

```
> data = c(10,20,20,30,10,20,30,30,10,20,3,3,1)
> fdata = factor(data)
> print('The Original Levels:')
[1] "The Original Levels:"
> fdata
 [1] 10 20 20 30 10 20 30 30 10 20 3  3  1
Levels: 1 3 10 20 30
> levels(fdata)[3]= 'a'
> print('The Changed Levels:')
[1] "The Changed Levels:"
> fdata
 [1] a  20 20 30 a  20 30 30 a  20 3  3  1
Levels: 1 3 a 20 30
```

5. The dataset swiss contains a standardized fertility measure and various socio-economic indicators for each of 47 French-speaking provinces of Switzerland in about 1888. a. Draw a scatterplot of Fertility against %Catholic. Which kind of areas have the lowest fertility rates? b. Discuss the relationship between the variables Education and Agriculture.

a.

```
data("swiss")
swiss
plot(swiss$Fertility, swiss$Catholic, col ="blue", main= "Scatterplot", xlab ="Fertility", ylab="Catholic")
```



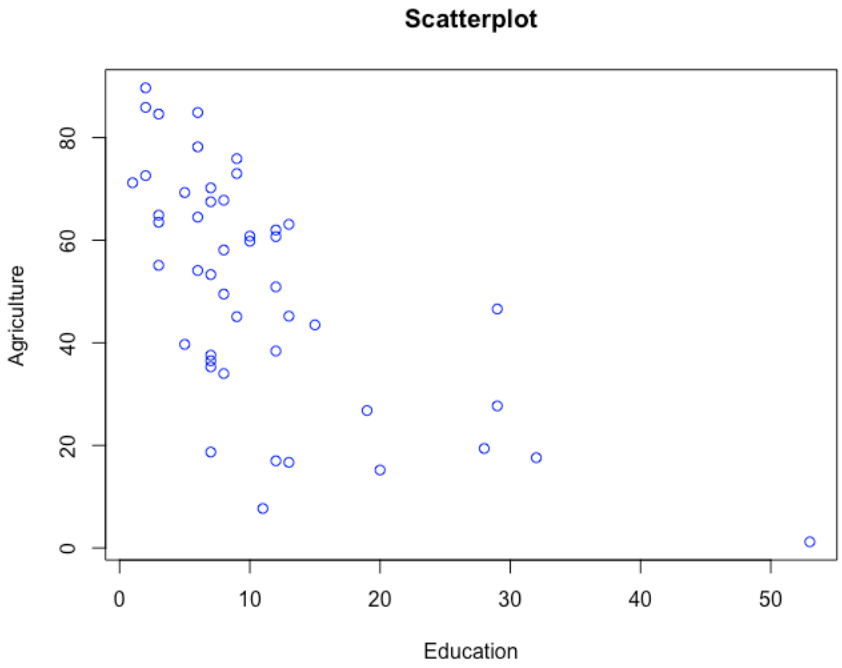
V. De Geneve has lowest fertility rate with 35.0%, having Catholic Population of about 42.34%.

The other regions with low fertility include Rive Gouche (42.8%) and Rive Droite (44.7%) having Catholic populations in percentages of 58.33% and 50.43% respectively.

The least fertility therefore lies in the 40-60% Catholic population range.

b.

Education and Agriculture variables have a direct yet negative correlation. The scatterplot below clearly indicates that persons with higher level of education are less likely to be involved in or with agriculture while those with lesser level of education are more likely to be involved with agriculture.


Scatterplot
Agriculture
Education
0
20
40
60
80
0
10
20
30
40
50