

Exercise-7

c. Reading XML dataset in R.

Description: **sudo apt-get install r-cran-xml**

#Run this command if the library not found while executing

```
if (!requireNamespace("XML", quietly = TRUE)) {  
  install.packages("XML")  
}  
# Set the working directory (replace with your desired directory)  
library(XML)  
# Specify the XML file  
setwd("/home/lab8")  
xml_file <- "samplexml.xml"  
xml_data <- xmlTreeParse(xml_file, useInternalNodes = TRUE)  
print("Structure of the XML data:")  
print(xml_data)  
root_node <- xmlRoot(xml_data)  
child_nodes <- xmlChildren(root_node)  
cat("Contents of the XML data:\n")  
for (child in child_nodes) {  
  cat(xmlName(child), ": ", xmlValue(child), "\n")  
}
```

Output:

Structure of the XML data:

```
<?xml version="1.0"?>  
<RECORDS>  
  <STUDENT>  
    <ID>1</ID>  
    <NAME>Alia</NAME>
```

```
<MARKS>620</MARKS>
<BRANCH>IT</BRANCH>
</STUDENT>
<STUDENT>
  <ID>2</ID>
  <NAME>Brijesh</NAME>
  <MARKS>440</MARKS>
  <BRANCH>Commerce</BRANCH>
</STUDENT>
<STUDENT>
  <ID>3</ID>
  <NAME>Yash</NAME>
  <MARKS>600</MARKS>
  <BRANCH>Humanities</BRANCH>
</STUDENT>
<STUDENT>
  <ID>4</ID>
  <NAME>Mallika</NAME>
  <MARKS>660</MARKS>
  <BRANCH>IT</BRANCH>
</STUDENT>
</RECORDS>
```

Contents of the XML data:

```
STUDENT : 1Alia620IT
STUDENT : 2Brijesh440Commerce
STUDENT : 3Yash600Humanities
STUDENT : 4Mallika660IT
```

Exercise-8

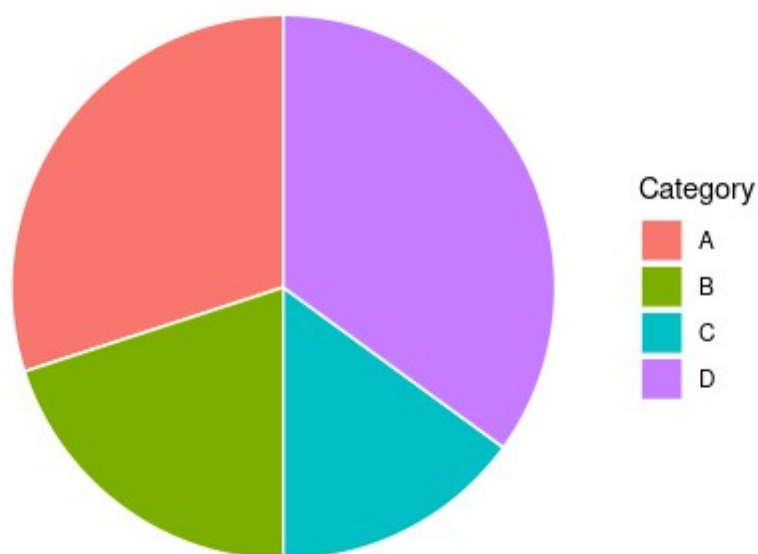
a. Implement R Script to create a Pie chart, Bar Chart, scatter plot and Histogram (Introduction to ggplot0 graphics)

```
if (!requireNamespace("ggplot2", quietly = TRUE)) {  
  install.packages("ggplot2")  
}  
library(ggplot2)  
# Create a sample dataset  
data <- data.frame(  
  Category = c("A", "B", "C", "D"),  
  Value = c(30, 20, 15, 35)  
)  
# Pie Chart  
pie_chart <- ggplot(data, aes(x = "", y = Value, fill = Category)) +  
  geom_bar(stat = "identity", width = 1, color = "white") +  
  coord_polar("y") +  
  ggtitle("Pie Chart") +  
  theme_void()  
# Bar Chart  
bar_chart <- ggplot(data, aes(x = Category, y = Value, fill = Category)) +  
  geom_bar(stat = "identity", color = "white") +  
  ggtitle("Bar Chart") +  
  theme_minimal()  
# Scatter Plot  
scatter_plot_data <- data.frame(  
  X = c(1, 2, 3, 4, 5),  
  Y = c(10, 8, 12, 6, 14)  
)  
scatter_plot <- ggplot(scatter_plot_data, aes(x = X, y = Y)) +  
  geom_point(size = 3, color = "blue") +  
  ggtitle("Scatter Plot") +  
  theme_minimal()  
# Histogram  
histogram_data <- data.frame(  
  Values = c(25, 30, 40, 35, 50, 45, 55, 60, 70, 65)  
)  
histogram <- ggplot(histogram_data, aes(x = Values)) +
```

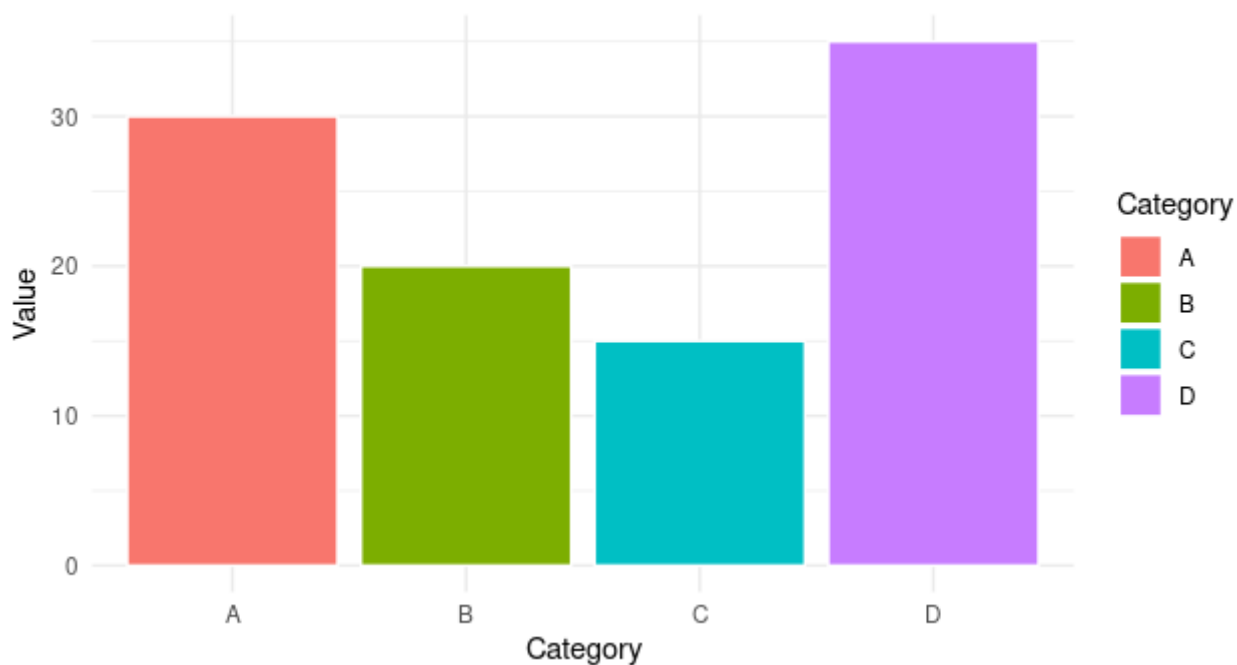
```
geom_histogram(binwidth = 5, fill = "orange", color = "black", alpha = 0.7) +  
ggtitle("Histogram")  
print(pie_chart)  
print(bar_chart)  
print(scatter_plot)  
print(histogram)
```

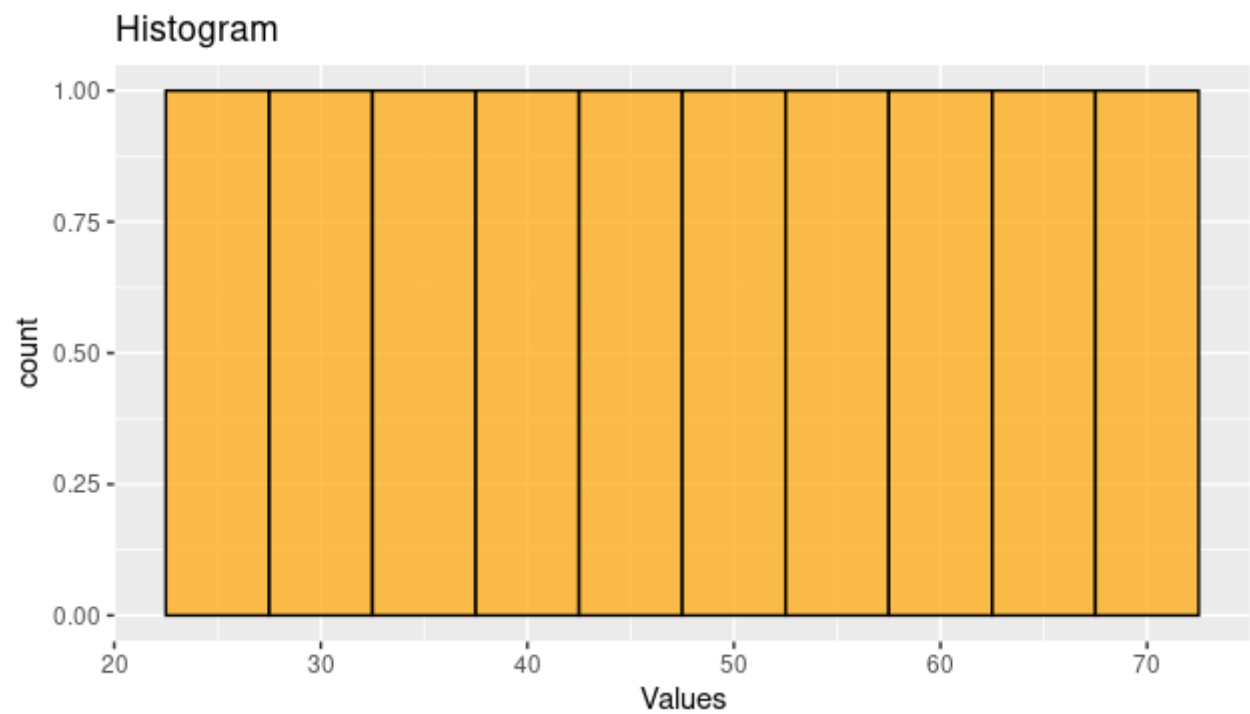
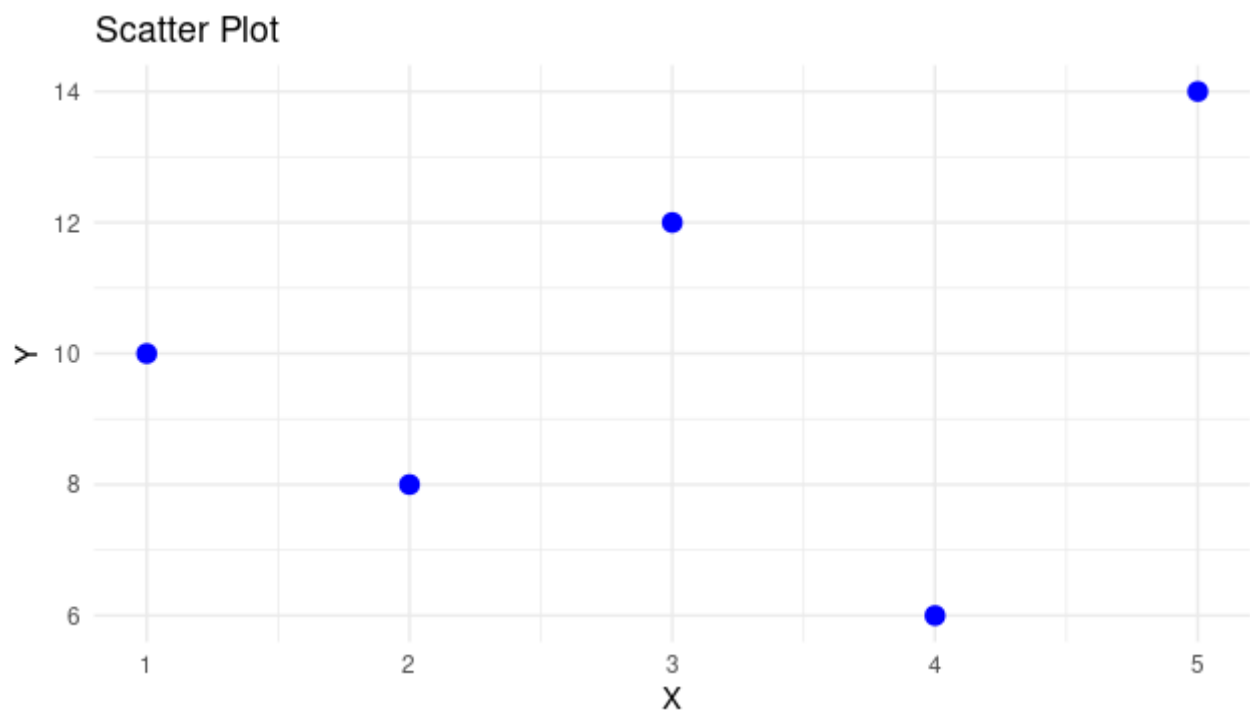
Output:

Pie Chart



Bar Chart





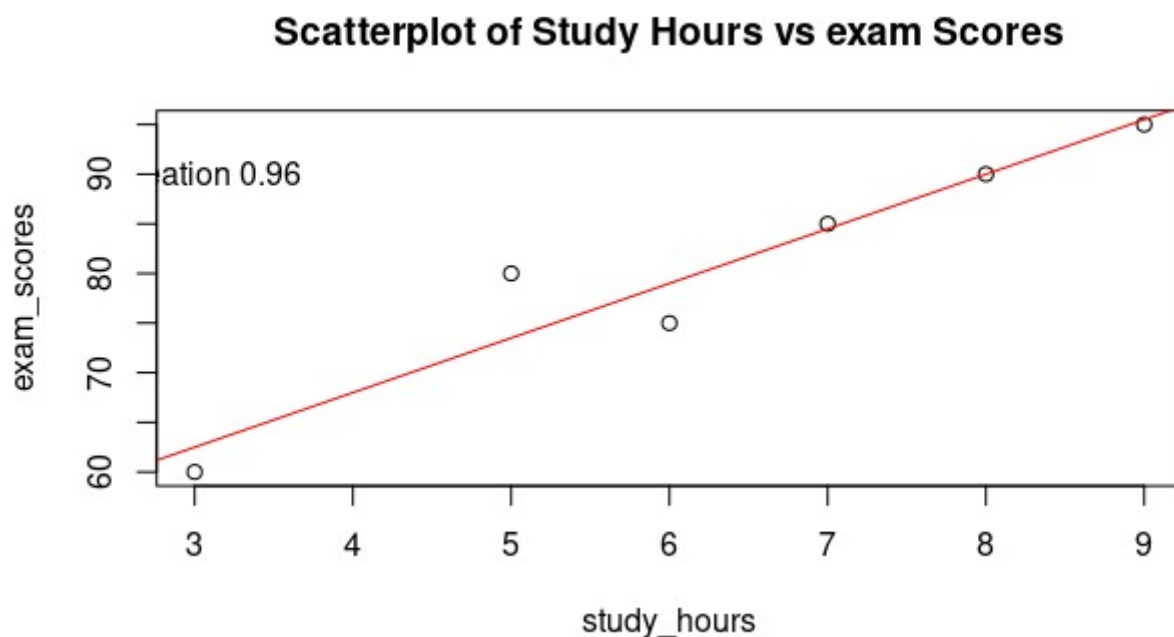
Exercise-9

*Correlation and linear regression example

```
study_hours<-c(5,7,3,8,6,9)
exam_scores<-c(80,85,60,90,75,95)
correlation<-cor(study_hours,exam_scores)
correlation
plot(study_hours,exam_scores,main="Scatterplot of Study Hours vs exam Scores")
abline(lm(exam_scores~study_hours),col="red")
text(3,90,paste("correlation",round(correlation,2)))
```

Output:

```
[1] 0.9569094
```



Linear and Multiple regression examples:

1>

```
x<-c(151,174,138,186,128,136,179,163,152,131)
```

```
y<-c(63,81,56,91,47,57,76,72,62,48)
```

```
relation<-lm(y~x)
```

```
print(relation)
```

Output:

Call:

```
lm(formula = y ~ x)
```

Coefficients:

(Intercept)	x
-38.4551	0.6746

2>

```
x<-c(151,174,138,186,128,136,179,163,152,131)
```

```
y<-c(63,81,56,91,47,57,76,72,62,48)
```

```
relation<-lm(y~x)
```

```
print(summary(relation))
```

Output:

Call:

```
lm(formula = y ~ x)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-6.3002	-1.6629	0.0412	1.8944	3.9775

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-38.45509	8.04901	-4.778	0.00139 **
x	0.67461	0.05191	12.997	1.16e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.253 on 8 degrees of freedom

Multiple R-squared: 0.9548, Adjusted R-squared: 0.9491

F-statistic: 168.9 on 1 and 8 DF, p-value: 1.164e-06