Day 3 Lab Manual Part 2

BIVARIATEANALYSIS IN R -COVARIANCE, CORRELATION, CROSSTAB

Exercise: 8

	Reference	Status	Gender	TestNewOrFollowUp				
1	KRXH	Accepted	Female	Test1	New			
2	KRPT	Accepted	Male	Test1	New			
3	FHRA	Rejected	Male	Test2	New			
4	CZKK	Accepted	Female	Test3	New5	CQTN	Rejected	Female	Test1	New
6	PZXW	Accepted	Female	Test4	Follow-up			
7	SZRZ	Rejected	Male	Test4	New			
8	RMZE	Rejected	Female	Test2	New9	STNX	Accepted	Female	Test3	New
10	TMDW	Accepted	Female	Test1	New			

i) Load the dataset and Create a data frame and name it as dataframe1 ii) Load the function for crosstab

Note: Perform status+gender

Status	Gender	
	Female	Male
Accepted	5	1
Rejected	2	2

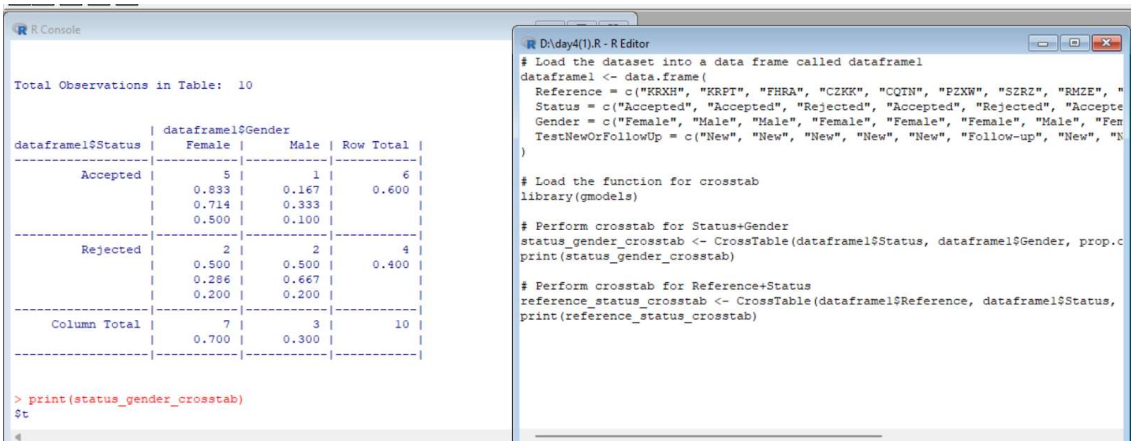
Note: Reference+Status

Reference	Status	
	Accepted	Rejected
CQTN	0	1
CZKK	1	0
FHRA	0	1
KRPT	1	0
KRXH	1	0
PZXW	1	0
RMZE	0	1
STNX	1	0
SZRZ	0	1
TMDW	1	0

SOURCE CODE:

```
# Load the dataset into a data frame called dataframe1

dataframe1 <- data.frame(

  Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ", "RMZE", "STNX",
"TMDW"),
```

```
  Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected", "Accepted", "Rejected",
"Rejected", "Accepted", "Accepted"),

  Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female",
"Female"),

  TestNewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New", "New", "New",
"New")

)


# Load the function for crosstab

library(gmodels)


# Perform crosstab for Status+Gender

status_gender_crosstab <- CrossTable(dataframe1$Status, dataframe1$Gender, prop.chisq = FALSE)

print(status_gender_crosstab)


# Perform crosstab for Reference+Status

reference_status_crosstab <- CrossTable(dataframe1$Reference, dataframe1$Status, prop.chisq =
FALSE)

print(reference_status_crosstab)
```



Exercise: 9

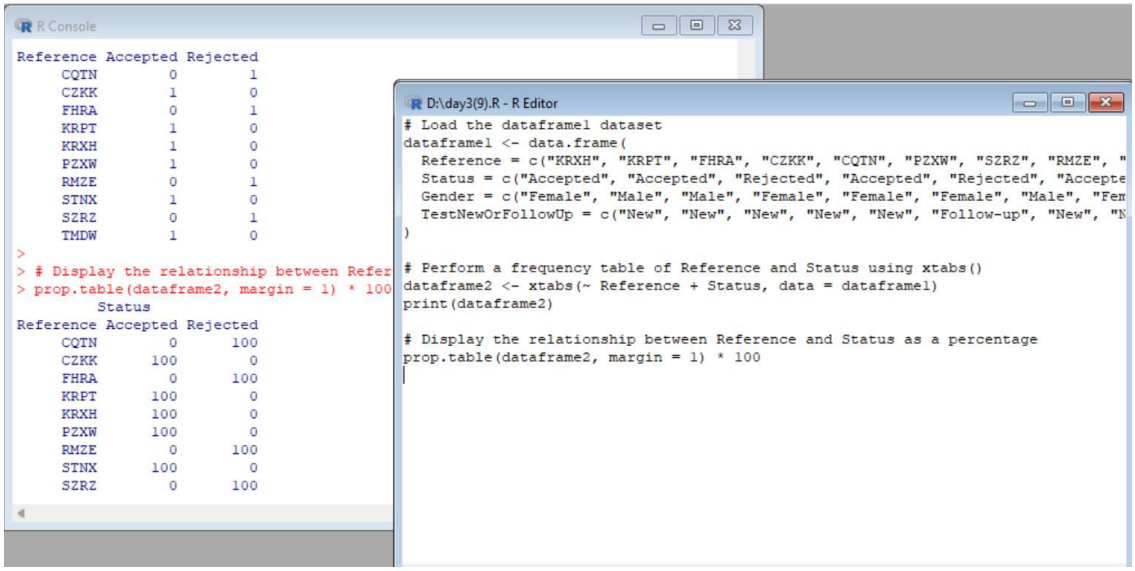
i) Use Two Categorical Variables and Discover the relationships within a dataset

ii) Next, using the xtabs() function, apply two variables from “dataframe1 “, to create a table delineating the relationship between the “Reference” category, and the “Status” category.

iii) Save the file in the name of dataframe2

SOURCE CODE:

```
# Load the dataframe1 dataset

dataframe1 <- data.frame(

  Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ", "RMZE", "STNX",
"TMDW"),

  Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected", "Accepted", "Rejected",
"Rejected", "Accepted", "Accepted"),

  Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female",
"Female"),

  TestNewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New", "New", "New",
"New")

)


# Perform a frequency table of Reference and Status using xtabs()

dataframe2 <- xtabs(~ Reference + Status, data = dataframe1)

print(dataframe2)


# Display the relationship between Reference and Status as a percentage

prop.table(dataframe2, margin = 1) * 100
```



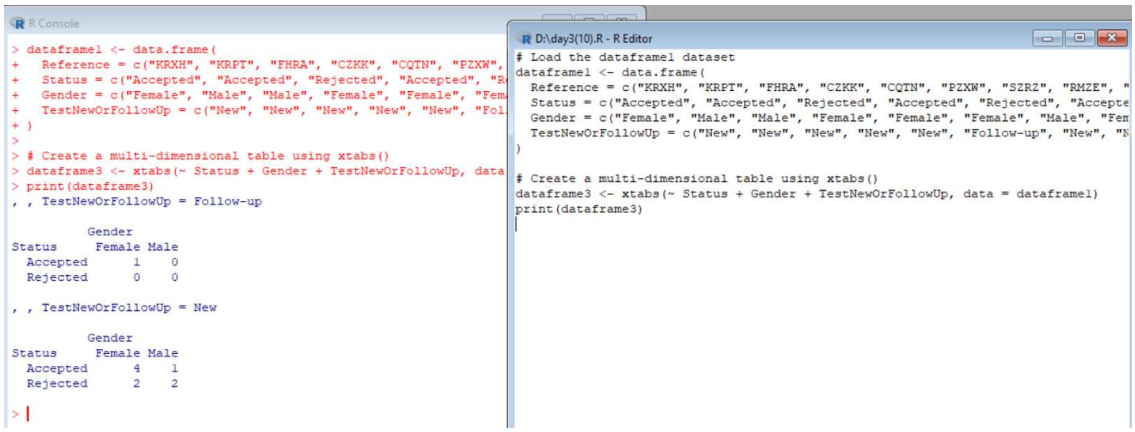
Exercise: 10

Use the same data frame using three Categorical Variables create a Multi-Dimensional Table

Apply three variables from “dataframe1” to create a Multi-Dimensional Cross-Tabulation of “Status“, “Gender“, and “Test“.

SOURCE CODE:

```
# Load the dataframe1 dataset

dataframe1 <- data.frame(

  Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ", "RMZE", "STNX",
"TMDW"),

  Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected", "Accepted", "Rejected",
"Rejected", "Accepted", "Accepted"),

  Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female",
"Female"),

  TestNewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New", "New", "New",
"New")

)


# Create a multi-dimensional table using xtabs()

dataframe3 <- xtabs(~ Status + Gender + TestNewOrFollowUp, data = dataframe1)

print(dataframe3)
```



Exercise: 11

Row Percentages

The R package “tigerstats” is required for the next two exercises.

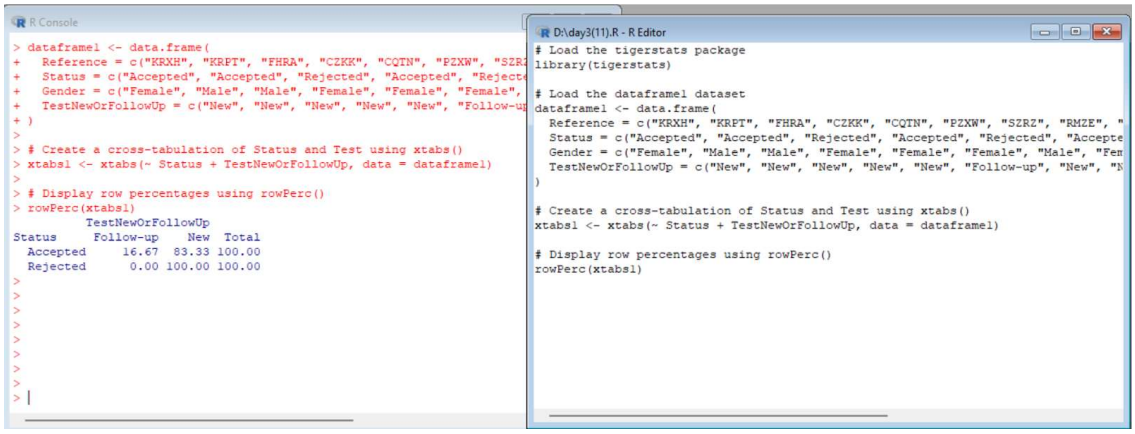
1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.

2) Enclose the xtabs() formula in the tigerstats function, “rowPerc()” to display row percentages for “Status” by “Test“.

SOURCE CODE:

```
# Load the tigerstats package

library(tigerstats)


# Load the dataframe1 dataset

dataframe1 <- data.frame(

  Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ", "RMZE", "STNX",
"TMDW"),

  Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected", "Accepted", "Rejected",
"Rejected", "Accepted", "Accepted"),

  Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female",
"Female"),

  TestNewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New", "New", "New",
"New")

)
```

```
# Create a cross-tabulation of Status and Test using xtabs()

xtabs1 <- xtabs(~ Status + TestNewOrFollowUp, data = dataframe1)


# Display row percentages using rowPerc()

rowPerc(xtabs1)
```



Exercise 12

Column Percentages

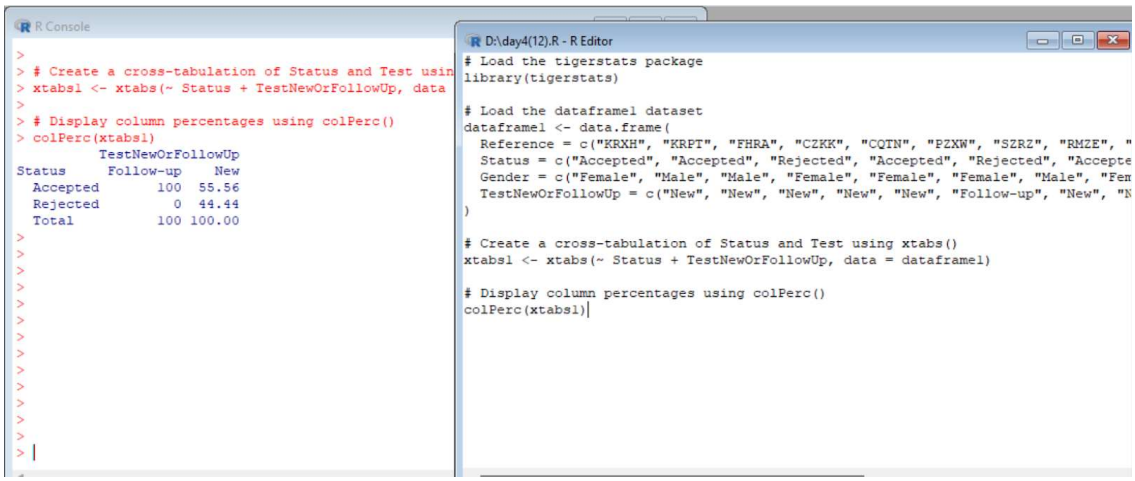
1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.

2) Enclose the xtabs() formula in the tigerstats function, “colPerc()” to display row percentages for “Status” by “Test“.

SOURCE CODE:

```
# Load the tigerstats package

library(tigerstats)


# Load the dataframe1 dataset

dataframe1 <- data.frame(

 Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ", "RMZE", "STNX",
"TMDW"),

 Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected", "Accepted", "Rejected",
"Rejected", "Accepted", "Accepted"),

 Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female",
"Female"),
```

```
  TestNewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New", "New", "New",
"New")

)


# Create a cross-tabulation of Status and Test using xtabs()

xtabs1 <- xtabs(~ Status + TestNewOrFollowUp, data = dataframe1)


# Display column percentages using colPerc()

colPerc(xtabs1)
```



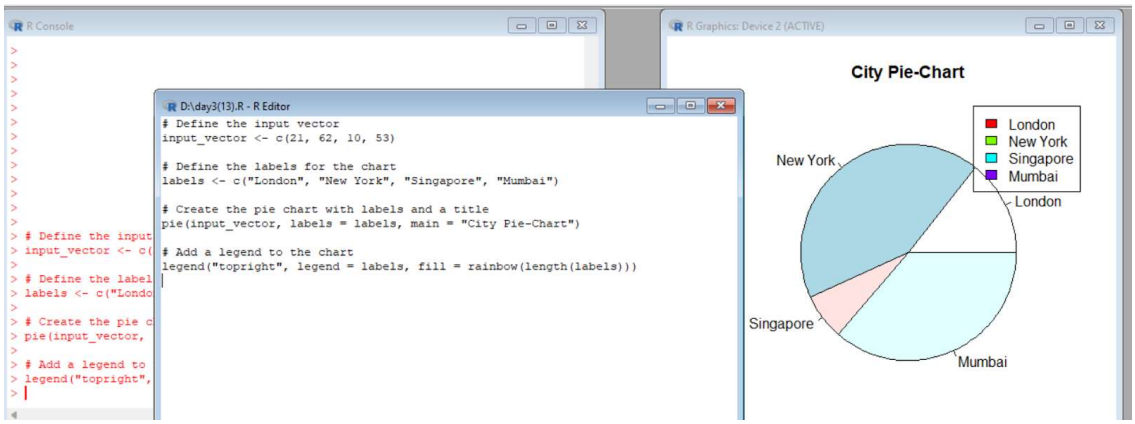
VISUALIZATION IN R

13. Write a program for creating a pie-chart in R using the input vector(21,62,10,53). Provide labels for the chart as ‘London’, ‘New York’, ‘Singapore’, ‘Mumbai’. Add a title to the chart as ‘city pie-chart’ and add a legend at the top right corner of the chart

SOURCE CODE:

```
# Define the input vector

input_vector <- c(21, 62, 10, 53)


# Define the labels for the chart

labels <- c("London", "New York", "Singapore", "Mumbai")
```

```
# Create the pie chart with labels and a title
pie(input_vector, labels = labels, main = "City Pie-Chart")

# Add a legend to the chart
legend("topright", legend = labels, fill = rainbow(length(labels)))
```



14. Create a 3D Pie Chart for the dataset "political Knowledge" with suitable labels,colours and a legend at the top right corner of the chart.

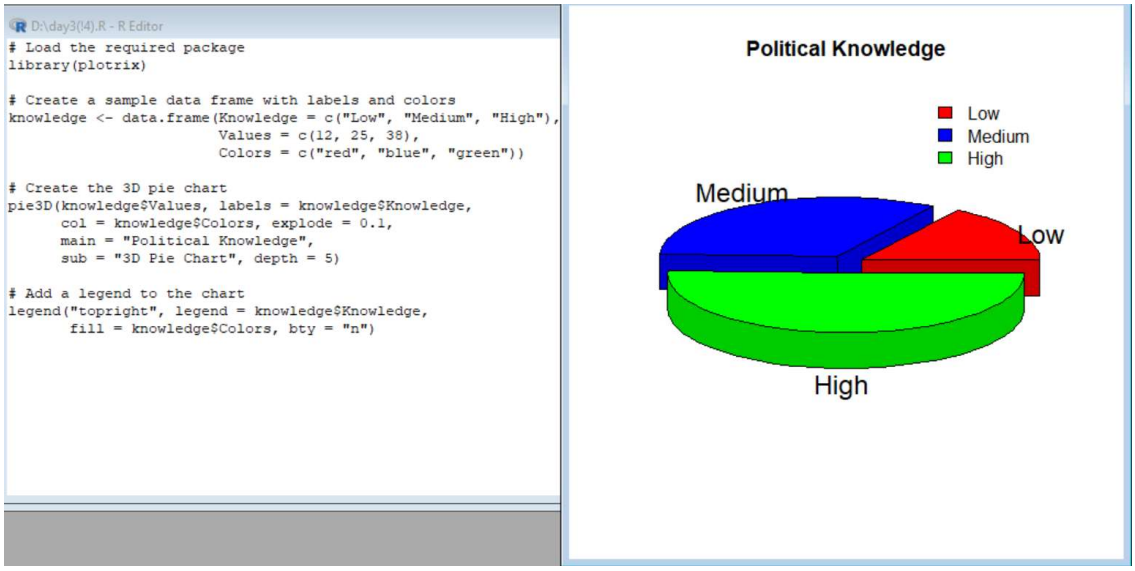
SOURCE CODE:

```
# Load the required package
library(plotrix)

# Create a sample data frame with labels and colors
knowledge <- data.frame(Knowledge = c("Low", "Medium", "High"),
                        Values = c(12, 25, 38),
                        Colors = c("red", "blue", "green"))

# Create the 3D pie chart
pie3D(knowledge$Values, labels = knowledge$Knowledge,
    col = knowledge$Colors, explode = 0.1,
    main = "Political Knowledge",
    sub = "3D Pie Chart", depth = 5)
```

Add a legend to the chart

legend("topright", legend = knowledge$Knowledge,

fill = knowledge$Colors, bty = "n")



15. Write a program for creating a bar chart using the vectors H=c(7,12,28,3,41) and M=c(“mar”, “apr”, “may”, “jun”, “jul”). Add a title to the chart as “Revenue chart”.

SOURCE CODE:

Define the data

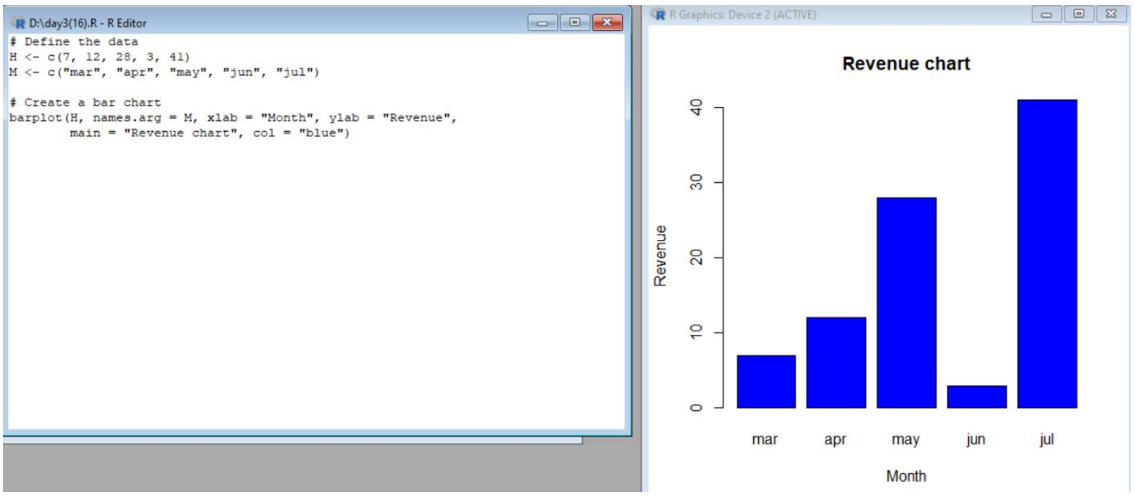
H <- c(7, 12, 28, 3, 41)

M <- c("mar", "apr", "may", "jun", "jul")

Create a bar chart

barplot(H, names.arg = M, xlab = "Month", ylab = "Revenue",

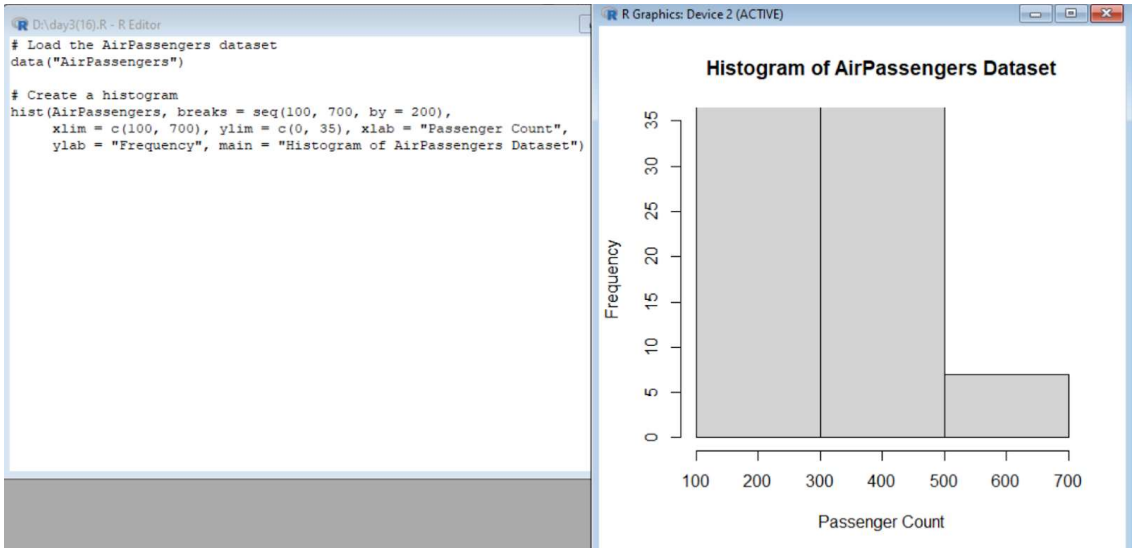
main = "Revenue chart", col = "blue")



16. Make a histogram for the "AirPassengers"dataset, start at 100 on the x-axis, and from values 200 to 700, make the bins 200 wide

SOURCE CODE:

```
# Load the AirPassengers dataset

data("AirPassengers")


# Create a histogram

hist(AirPassengers, breaks = seq(100, 700, by = 200),

    xlim = c(100, 700), ylim = c(0, 35), xlab = "Passenger Count",

    ylab = "Frequency", main = "Histogram of AirPassengers Dataset")
```



17. Create a Boxplot graph for the relation between "mpg"(miles per galloon) and "cyl"(number of Cylinders) for the dataset "mtcars" available in R Environment.

SOURCE CODE:

```
# Load the mtcars dataset

data(mtcars)


# Create a boxplot

boxplot(mpg ~ cyl, data = mtcars,

      xlab = "Number of Cylinders",

      ylab = "Miles per Gallon",

      main = "Boxplot of MPG vs. Number of Cylinders")
```