Day 3 Lab Manual Part 2

Rithik vasan.k
191921033

BIVARIATEANALYSIS IN R -COVARIANCE, CORRELATION, CROSSTAB

Exercise: 8

Reference	Status	Gender	TestNewOrFollowUp
KRXH	Accepted	Female	Test1 New
KRPT	Accepted	Male	Test1 New
FHRA	Rejected	Male	Test2 New
CZKK	Accepted	Female	Test3 New
CQTN	Rejected	Female	Test1 New
PZXW	Accepted	Female	Test4 Follow-up
SZRZ	Rejected	Male	Test4 New
RMZE	Rejected	Female	Test2 New
STNX	Accepted	Female	Test3 New
TMDW	Accepted	Female	Test1 New

- Load the dataset and Create a data frame and name it as dataframe1
- Load the function for crosstab

Note: Perform status+gender

Status	Gender	
	Female	Male
Accepted	5	1
Rejected	2	2

PROGRAM:

```
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("Test1", "Test1", "Test2", "Test3", "Test1",
                        "Test4", "Test4", "Test2", "Test3", "Test1")
)
dataframe1

crosstab1 <- xtabs(~ Status + Gender, data = dataframe1)
crosstab1
```

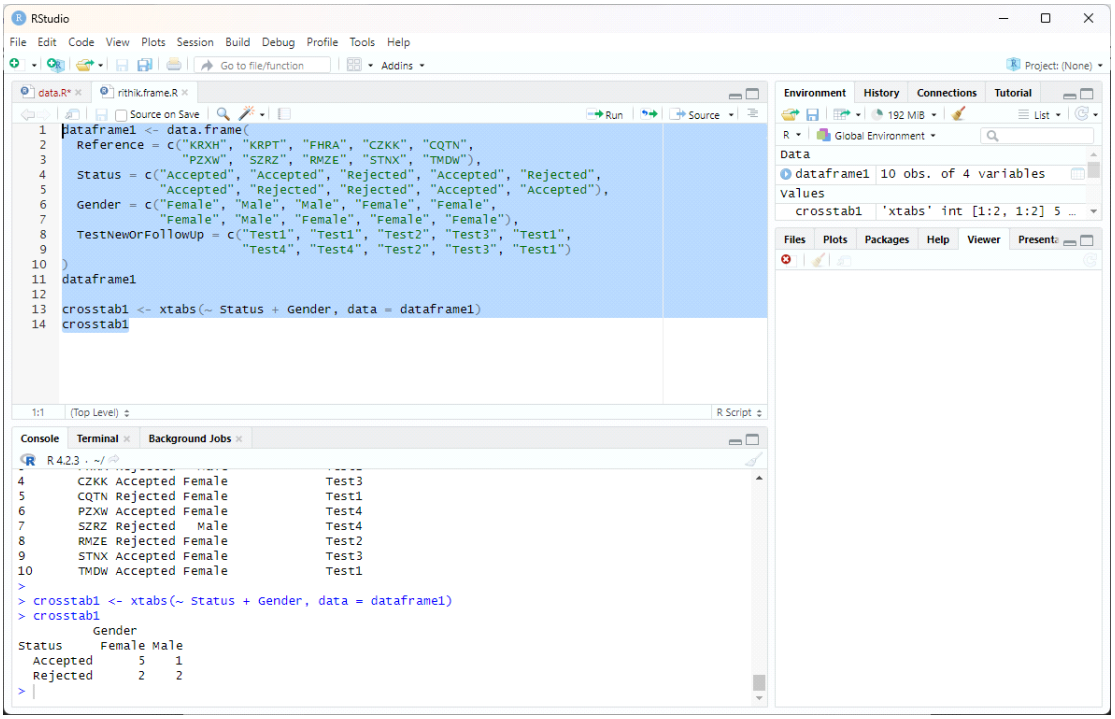
OUTPUT

```
> dataframe1
   Reference   Status Gender TestNewOrFollowUp
1       KRXH Accepted Female            Test1
2       KRPT Accepted   Male            Test1
3       FHRA Rejected   Male            Test2
4       CZKK Accepted Female            Test3
5       CQTN Rejected Female            Test1
6       PZXW Accepted Female            Test4
7       SZRZ Rejected   Male            Test4
8       RMZE Rejected Female            Test2
9       STNX Accepted Female            Test3
10      TMDW Accepted Female            Test1
>
> crosstab1 <- xtabs(~ Status + Gender, data = dataframe1)

> crosstab1
          Gender
Status     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



Exercise: 9

- Use Two Categorical Variables and Discover the relationships within a dataset
- 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.
- Save the file in the name of dataframe2

PROGRAM:

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

OUTPUT:

	TestNewOrFollowUp			
Status	Test1	Test2	Test3	Test4
Accepted	3	0	2	1
Rejected	1	2	0	1

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".

PROGRAM:

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

dataframe3
```

OUTPUT:

```
> dataframe3
TestNewOrFollowUp = Test1
          Gender
Status     Female Male
  Accepted      2    1
  Rejected      1    0

TestNewOrFollowUp = Test2
          Gender
Status     Female Male
  Accepted      0    0
  Rejected      1    1

TestNewOrFollowUp = Test3
          Gender
Status     Female Male
  Accepted      2    0
```

Rejected 0 0

TestNewOrFollowUp = Test4

Gender

Status Female Male

Accepted 1 0

Rejected 0 1

Exercise: 11

Row Percentages

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

- Create an xtabs() formula that cross-tabulates “Status“, and “Test“.
- Enclose the xtabs() formula in the tigerstats function, “rowPerc()” to display row percentages for “Status” by “Test“.

```
library(tigerstats)
install.packages(tigerstats)
cont_table <- xtabs(~ Status + TestNewOrFollowUp, data = dataframe1)
rowPerc(cont_table, col = "TestNewOrFollowUp")
```

Exercise 12

Column Percentages

- Create an xtabs() formula that cross-tabulates “Status“, and “Test“.
- Enclose the xtabs() formula in the tigerstats function, “colPerc()” to display row percentages for “Status” by “Test“.

PROGRAM:

```
library(tigerstats)
```

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

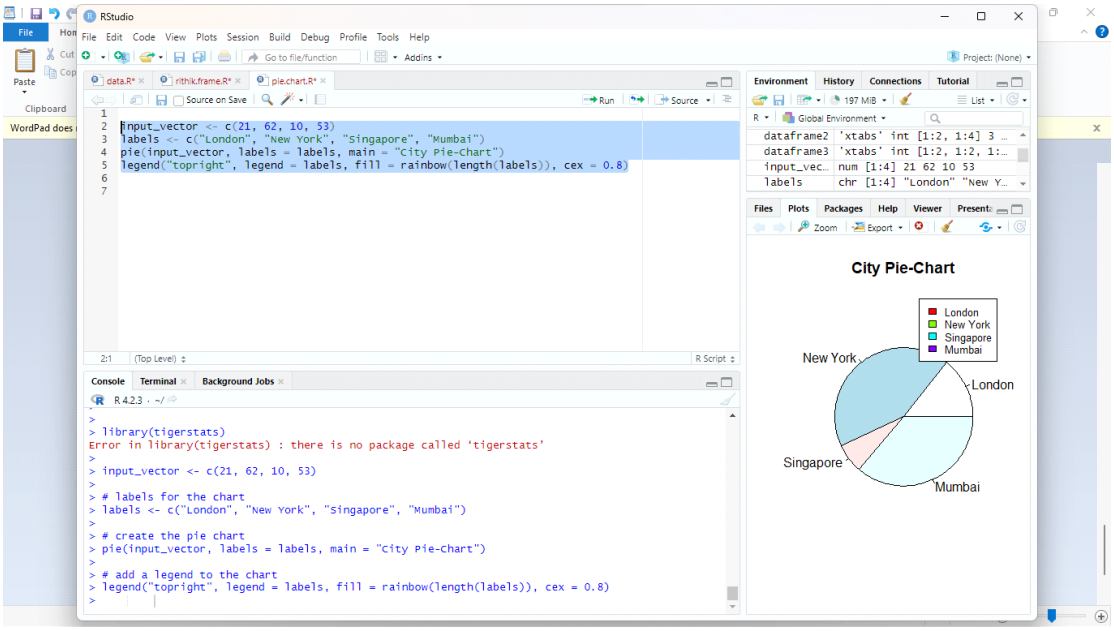
colPerc(cont_table, row = "Status")

VISUALIZATION IN R

- 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.

PROGRAM:

```
input_vector <- c(21, 62, 10, 53)
labels <- c("London", "New York", "Singapore", "Mumbai")
pie(input_vector, labels = labels, main = "City Pie-Chart")
legend("topright", legend = labels, fill = rainbow(length(labels)), cex = 0.8)
```



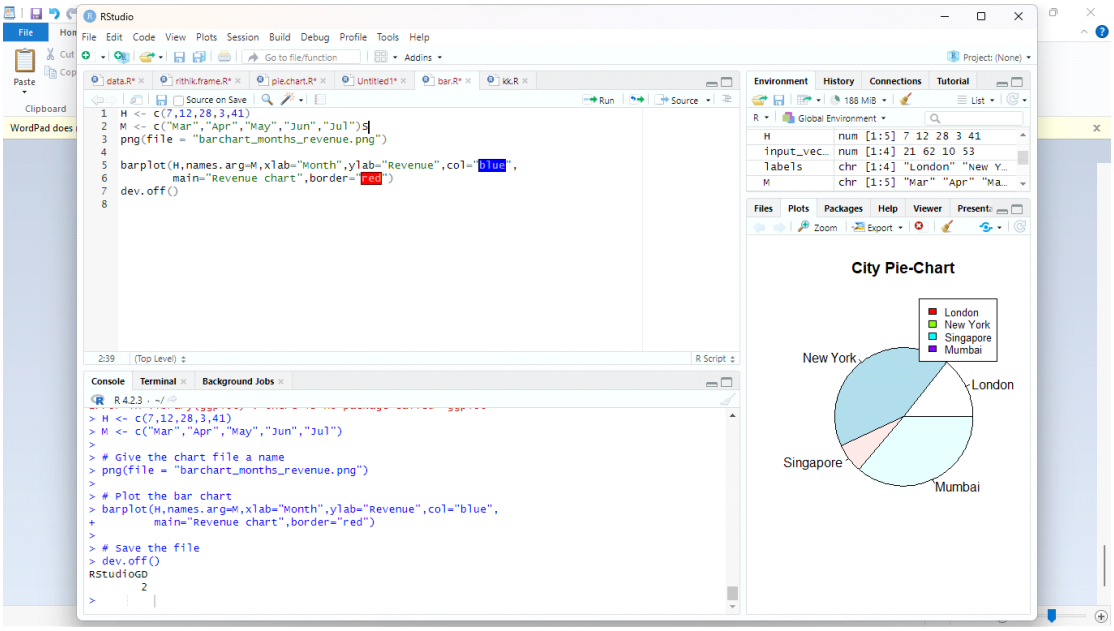
- Create a 3D Pie Chart for the dataset “political Knowledge” with suitable labels,colours and a legend at the top right corner of the chart.

```
political_knowledge <- c(35, 25, 20, 15, 5)
labels <- c("High", "Medium-High", "Medium", "Medium-Low", "Low")
colors <- c("darkgreen", "forestgreen", "olivedrab", "yellowgreen", "lightgreen")
pie3D(political_knowledge, labels = labels, explode = 0.05, col = colors,
    main = "Political Knowledge",
    radius = 1, depth = 0.5,
```

```
shade = 0.4, border = "black",
lwd = 2, explodePos = 3,
legend = c(1, 1), legend.cex = 0.8,
legend.labels = labels, legend.fill = colors,
labelcex = 0.8, labelcol = "black")
```

- 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”.

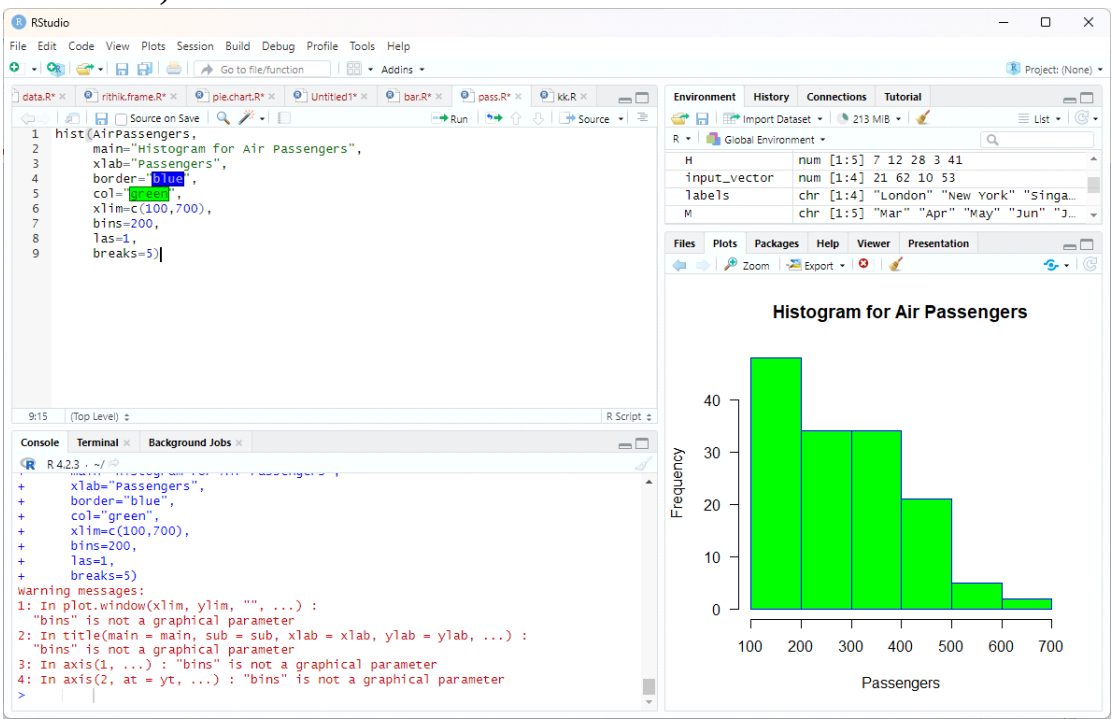
```
H <- c(7,12,28,3,41)
M <- c("Mar","Apr","May","Jun","Jul")
png(file = "barchart_months_revenue.png")
barplot(H,names.arg=M,xlab="Month",ylab="Revenue",col="blue",
      main="Revenue chart",border="red")
dev.off()
```



- 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

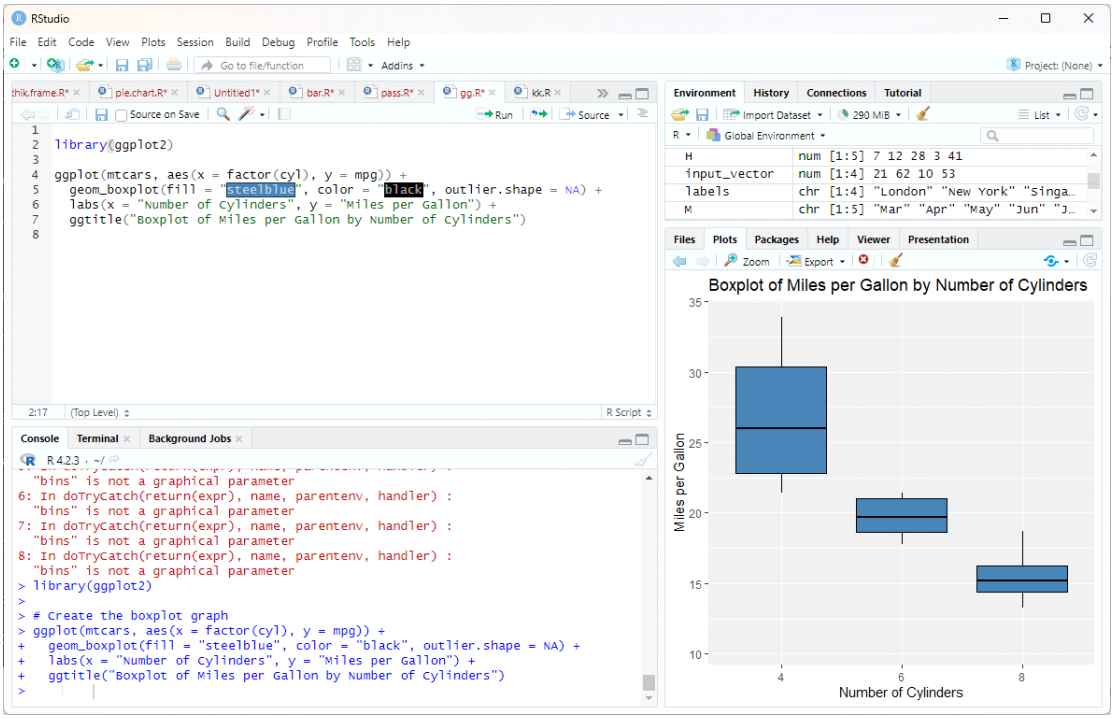
```
hist(AirPassengers,
   main="Histogram for Air Passengers",
   xlab="Passengers",
   border="blue",
   col="green",
```

```
    xlim=c(100,700),
    bins=200,
    las=1,
    breaks=5)
```



- 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.

```
library(ggplot2)
ggplot(mtcars, aes(x = factor(cyl), y = mpg)) +
  geom_boxplot(fill = "steelblue", color = "black", outlier.shape = NA) +
  labs(x = "Number of Cylinders", y = "Miles per Gallon") +
  ggtitle("Boxplot of Miles per Gallon by Number of Cylinders")
```


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
ggplot.R*
Source on Save
Run
Source
1 library(ggplot2)
2
3 ggplot(mtcars, aes(x = factor(cyl), y = mpg)) +
4 geom_boxplot(fill = "steelblue", color = "black", outlier.shape = NA) +
5 labs(x = "Number of Cylinders", y = "Miles per Gallon") +
6 ggtitle("Boxplot of Miles per Gallon by Number of Cylinders")
7
8
2:17 (Top Level)
R Script
Console
Terminal
Background Jobs
R 4.2.3 · ~/
"bins" is not a graphical parameter
6: In doTryCatch(return(expr), name, parentenv, handler) :
"bins" is not a graphical parameter
7: In doTryCatch(return(expr), name, parentenv, handler) :
"bins" is not a graphical parameter
8: In doTryCatch(return(expr), name, parentenv, handler) :
"bins" is not a graphical parameter
> library(ggplot2)
>
> # Create the boxplot graph
> ggplot(mtcars, aes(x = factor(cyl), y = mpg)) +
+ geom_boxplot(fill = "steelblue", color = "black", outlier.shape = NA) +
+ labs(x = "Number of Cylinders", y = "Miles per Gallon") +
+ ggtitle("Boxplot of Miles per Gallon by Number of Cylinders")
>
Environment
History
Connections
Tutorial
Import Dataset
290 MiB
List
R
Global Environment
H num [1:5] 7 12 28 3 41
input_vector num [1:4] 21 62 10 53
labels chr [1:4] "London" "New York" "Singa...
M chr [1:5] "Mar" "Apr" "May" "Jun" "J...
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Boxplot of Miles per Gallon by Number of Cylinders
Miles per Gallon
35
30
25
20
15
10
4
6
8
Number of Cylinders