UNIVERSITY INSTITUTE of COMPUTING
Asia's Fastest Growing University

NAAC GRADE A+
Accredited University

Statistical Techniques Using R Lab
Minor Project
Code - 24CAP-614

Submitted by:

Aman Sharma
UID:24MCI10254
Pragati Gupta
UID:24MCI10255

Submitted to:

Dr.Mausam
Ass.Professor
Emp.ID:E17645

Task: Choose a dataset from a repository like Kaggle or UCI Machine Learning Repository and perform exploratory data analysis using R. Explore the distribution of variables, identify outliers, and visualize relationships between variables using plots like histograms, scatter plots, and boxplots.

Code:

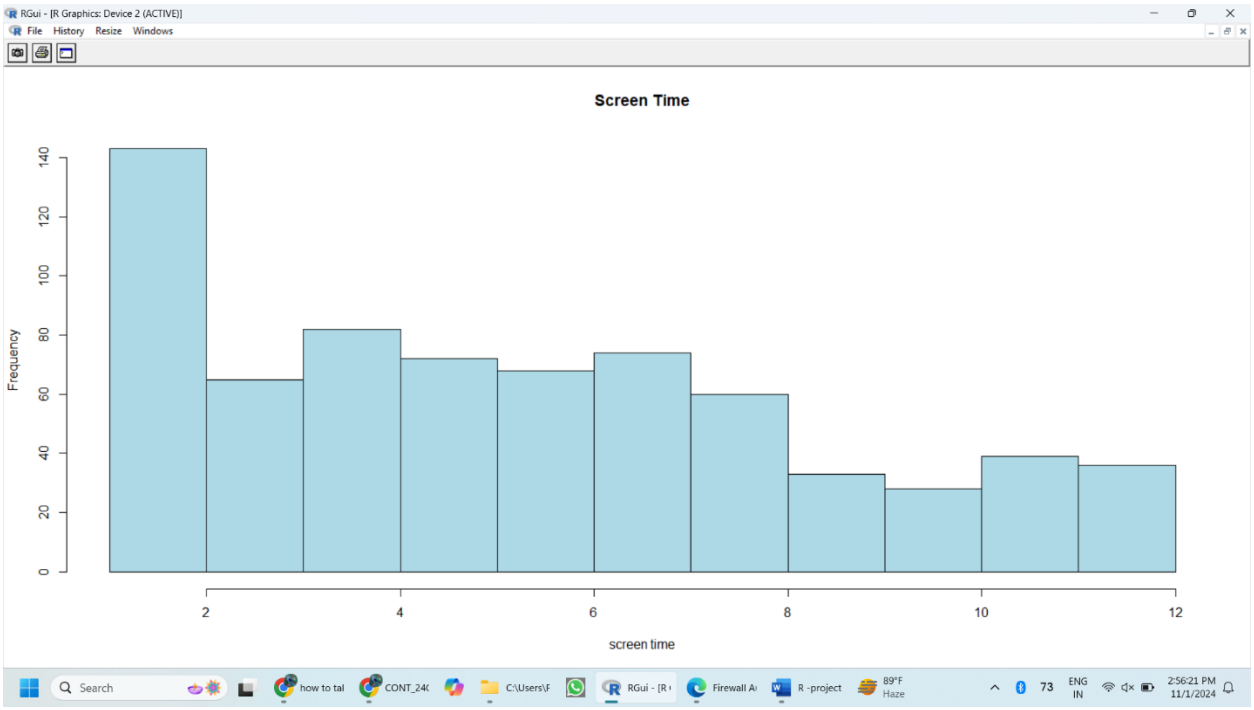
```
unzip("C:\\Users\\PRANSHUL GUPTA\\Downloads\\archive.zip",exdir="C:\\Users\\PRANSHUL GUPTA\\Downloads\\archive")
files<-list.files("C:\\Users\\PRANSHUL GUPTA\\Downloads\\archive",full.names=TRUE)
files
D<-read.csv("C:\\Users\\PRANSHUL GUPTA\\Downloads\\archive/user_behavior_dataset.csv")
D

#understand its structure
head(D)          #it display the first few rows
str(D)           #check the structure of dataset
summary(D)                  #summary of dataset
install.packages("dypler")
is.na(D$Age)
View(D)              #use to display the dataset
names(D)             #use to display the col names

#aggreation of dataset
```

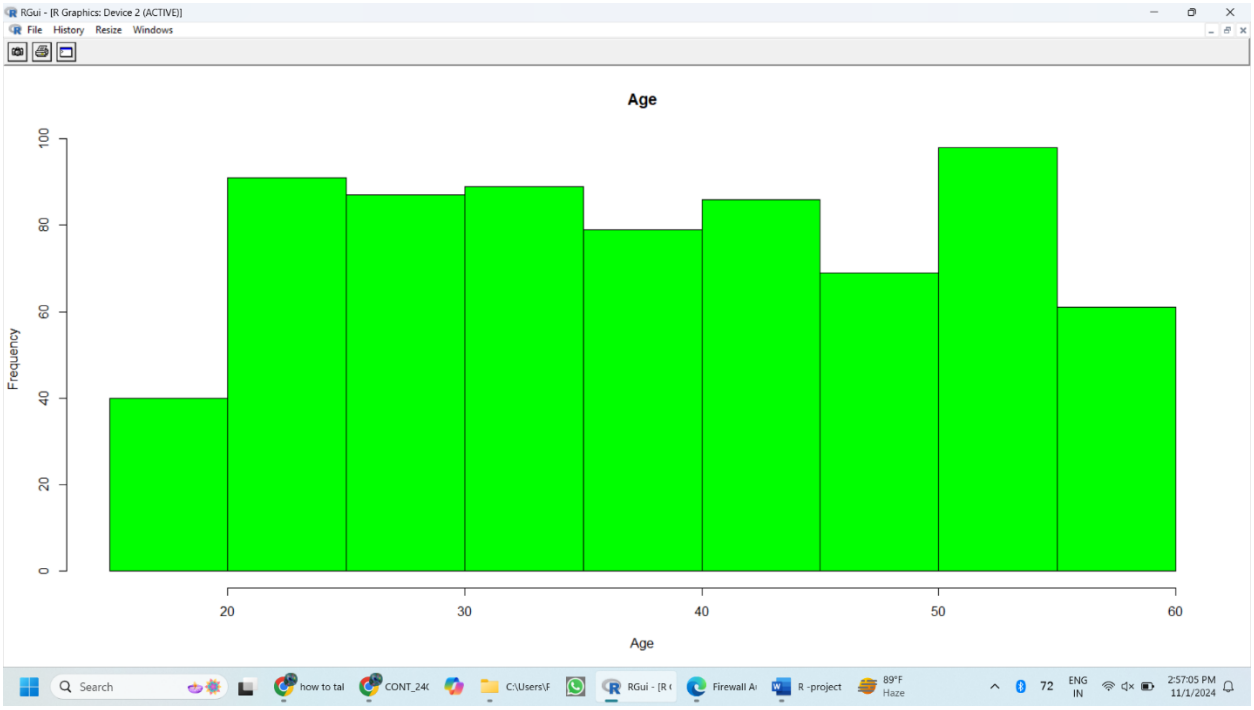
```
 #aggregate(dependent~independent,dataname,function)

aggregate(Screen.On.Time..hours.day.~ Age, D,mean)
aggregate(Screen.On.Time..hours.day.~ Gender, D,mean)
aggregate(Number.of.Apps.Installed~Age, D,mean)
aggregate(Screen.On.Time..hours.day.~ Age, D,sum)
aggregate(Screen.On.Time..hours.day.~ Gender, D,sum)

# Plotting histograms for numerical variable

hist(D$Screen.On.Time..hours.day., main = "Screen Time",
      xlab = "screen time",
      col = "lightblue",
      border = "black")
hist(D$Age, main = "Age",
      xlab = "Age",
      col = "green",
      border = "black")
hist(D$Number.of.Apps.Installed, main = "No. of Apps",
       xlab = "No. of app installed",
       col = "lightcoral",
       border = "black")

# Boxplots to identify outliers in the numerical variables

boxplot(D$Age, main = "Boxplot of Age",
        ylab = "Age",
        col = "lightblue",horizontal = TRUE )

boxplot(D$Data.Usage..MB.day., main = "Boxplot of Battery Usage",
        ylab = "Battery Usage",
```

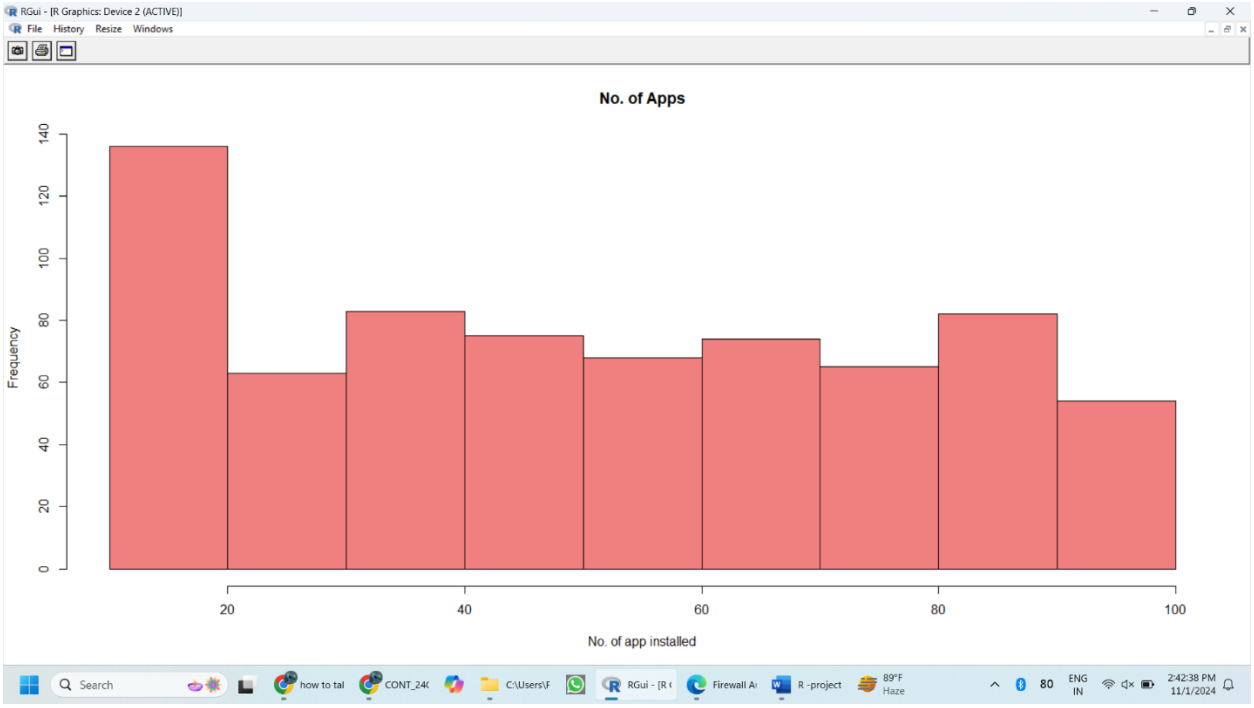
```
        col = "green", horizontal = TRUE)

boxplot(D$Screen.On.Time..hours.day.,main= " Boxplot of Screen Time",
        ylab = "Screen Time",
        col = "lightcoral", horizontal = TRUE)

boxplot(D$Number.of.Apps.Installed, main = "Boxplot of App Installed",
        ylab = "Apps",
        col = "pink", horizontal = TRUE)

install.packages("ggplot2")          #install the ggplot2 library for scatter plot
library(ggplot2)                     # Load the ggplot2 library
lengths(D)
D<-as.data.frame(D)
is.data.frame(D)                     #to check if D as a data frame

# Scatter plot of  Number of Apps Installed vs Battery Drain mAh day, colored by Age

ggplot(D, aes(x = Number.of.Apps.Installed, y = Battery.Drain..mAh.day., color = Age)) +
  geom_point(size = 2) +
  labs(title = "Scatter Plot of Number of Apps Installed vs Battery Drain mAh day",
       x = "Number of Apps Installed",
       y = "Battery Drain (mAh/day)") +
  theme_minimal()
```

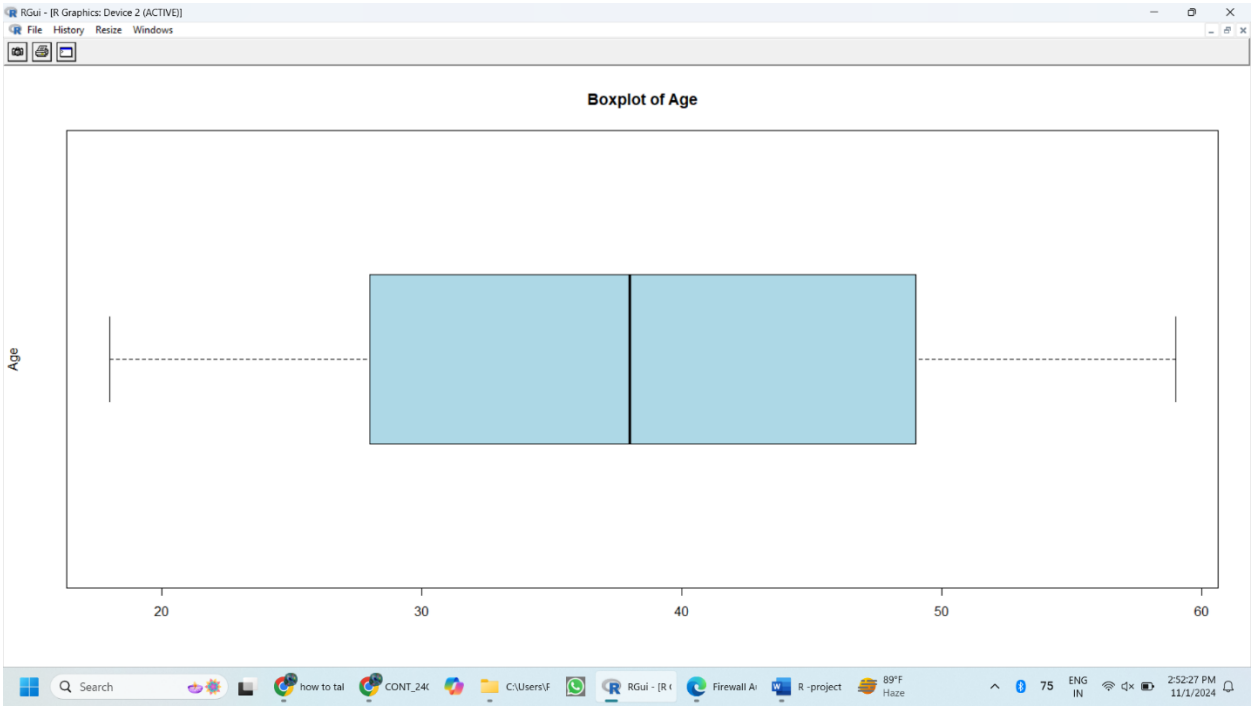
```
# Scatter plot of App Usage Time min day vs Number of Apps Installed, colored by Age

ggplot(D, aes(x = App.Usage.Time..min.day., y = Number.of.Apps.Installed, color = Age)) +
  geom_point(size = 2) +
  labs(title = "Scatter Plot  App Usage Time min day of vs Number of Apps Installed", x = " App Usage Time min day", y = "Number of Apps Installed") +
  theme_minimal()

#box plot: Number.of.Apps.Installed for different Gender

D <- boxplot( Number.of.Apps.Installed~ Gender,
              data=D,
              main="BOX-PLOT Graph",
              xlab="Age",
              ylab="Number of Apps Installed",
              col=c("lightblue", "purple", "lightgreen"),
              border="black",
              pch=16)
```

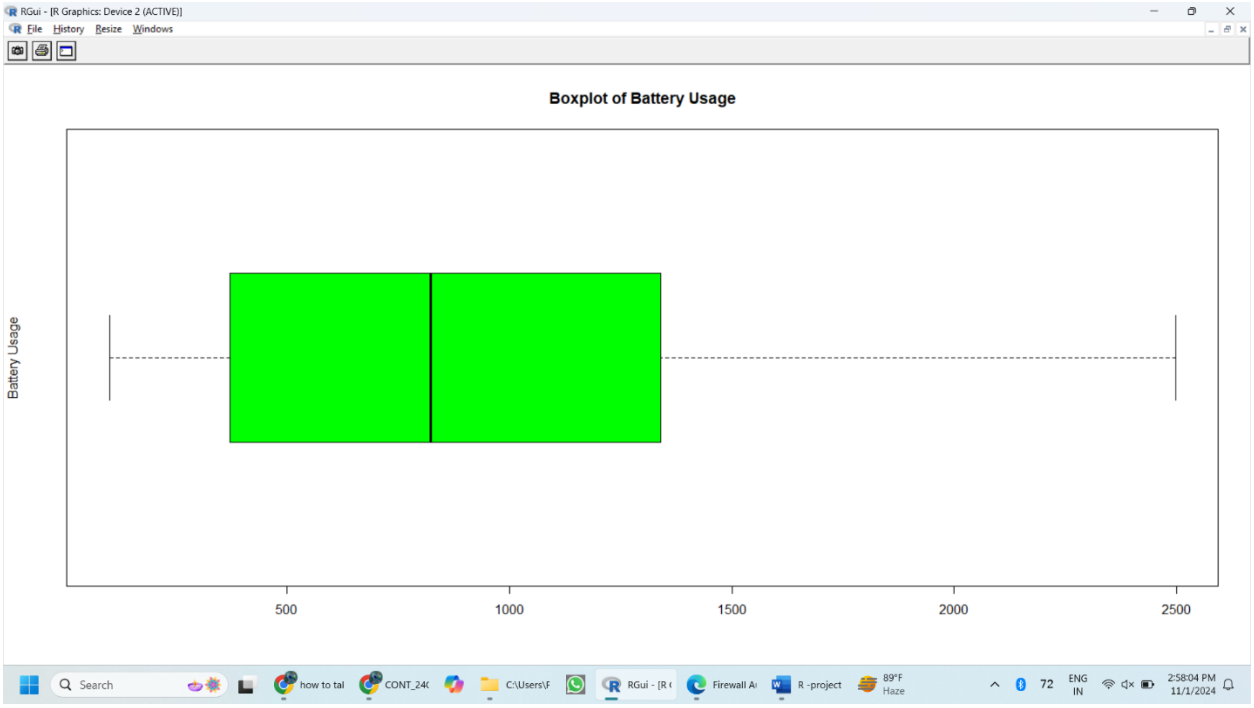
Outputs :

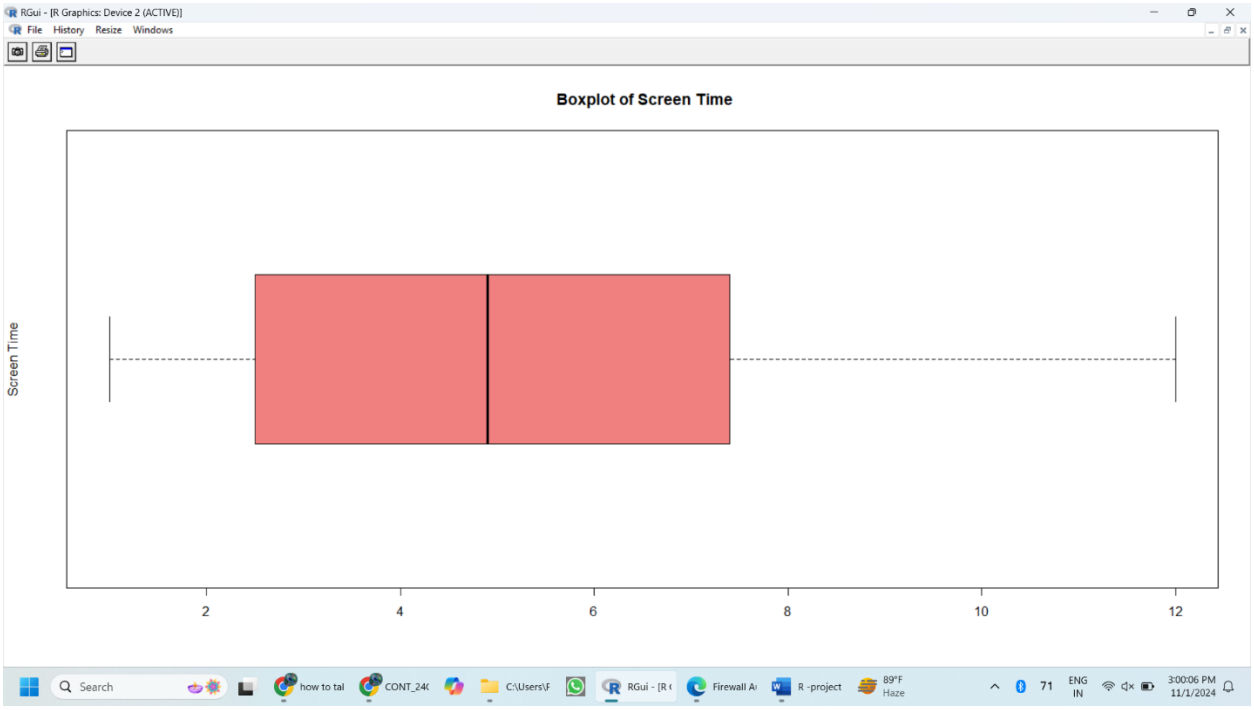

RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Screen Time
Frequency
0
20
40
60
80
100
120
140
2
4
6
8
10
12
screen time
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
73
ENG
IN
2:56:21 PM
11/1/2024

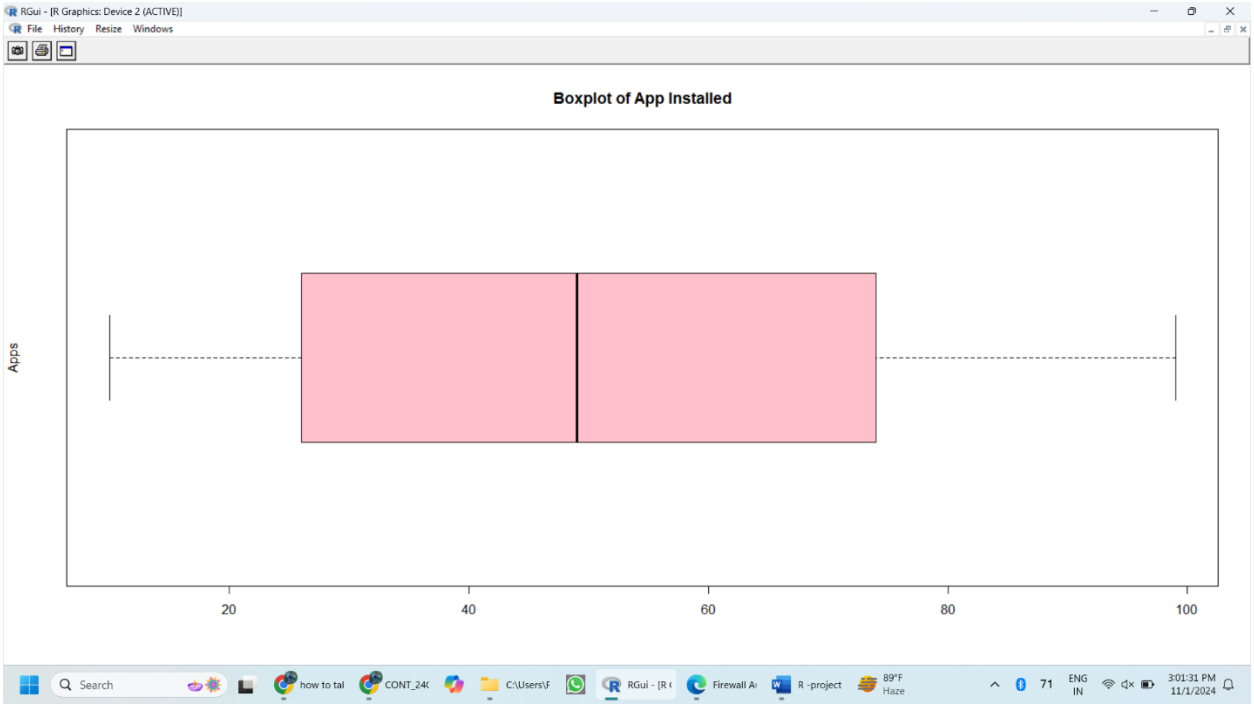

RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Age
Frequency
0
20
40
60
80
100
20
30
40
50
60
Age
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
72
ENG
IN
2:57:05 PM
11/1/2024

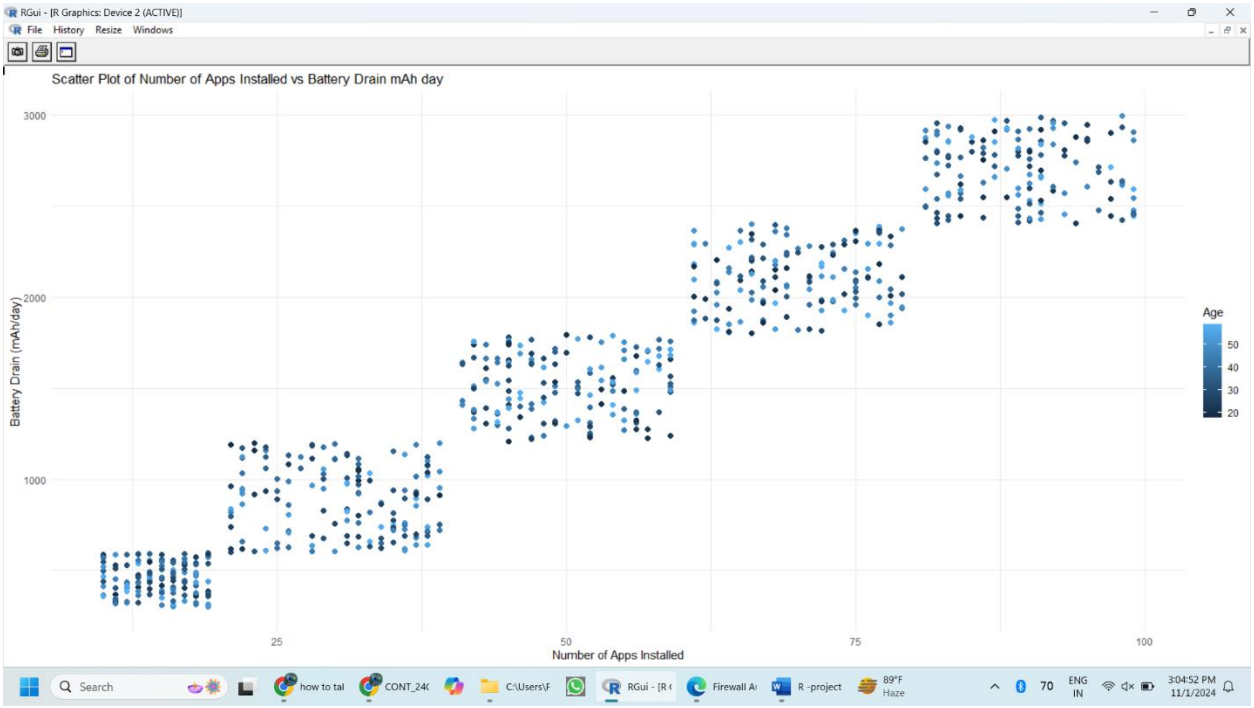

RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
No. of Apps
Frequency
0
20
40
60
80
100
120
140
20
40
60
80
100
No. of app installed
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
80
ENG
IN
2:42:38 PM
11/1/2024

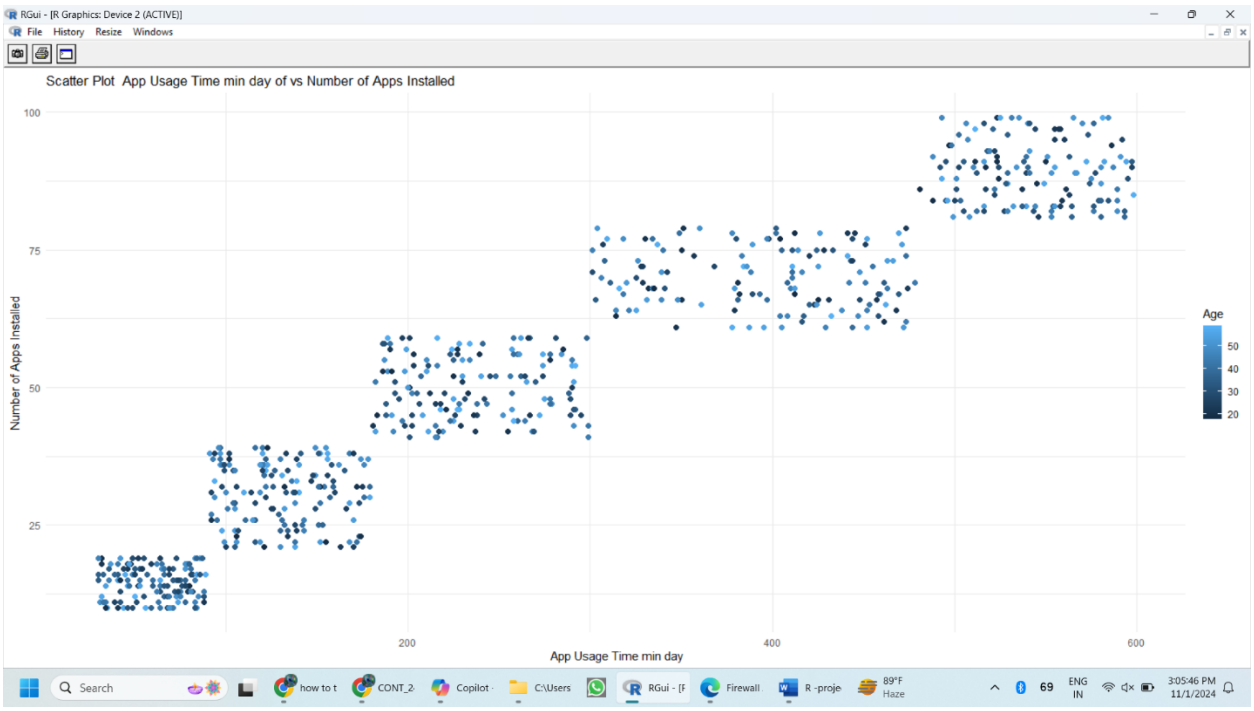

RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Boxplot of Age
Age
20
30
40
50
60
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
75
ENG
IN
2:52:27 PM
11/1/2024


RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Boxplot of Battery Usage
Battery Usage
500
1000
1500
2000
2500
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
72
ENG
IN
2:58:04 PM
11/1/2024


RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Boxplot of Screen Time
Screen Time
2
4
6
8
10
12
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
71
ENG
IN
3:00:06 PM
11/1/2024


RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Boxplot of App Installed
Apps
20
40
60
80
100
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
71
ENG
IN
3:01:31 PM
11/1/2024


RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Scatter Plot of Number of Apps Installed vs Battery Drain mAh day
Battery Drain (mAh/day)
3000
2000
1000
25
50
75
100
Number of Apps Installed
Age
50
40
30
20
Search
how to tal
CONT_24(
C:\Users\F
RGui - [R (
Firewall A(
R -project
89°F
Haze
70
ENG
IN
3:04:52 PM
11/1/2024


RGui - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Scatter Plot App Usage Time min day of vs Number of Apps Installed
Number of Apps Installed
App Usage Time min day
Age
50
40
30
20
100
75
50
25
200
400
600
Search
how to t
CONT_2
Copilot
C:\Users
RGui - [F
Firewall
R -proje
89°F
Haze
69
ENG
IN
3:05:46 PM
11/1/2024