**Data Ware Housing and Data Mining.**

**Vikas – 192311353**

**22-2-2025**

**5 . .Make a histogram for the “AirPassengers “dataset, start at 100 on the x-axis, and from values 200 to 700, make the bins 150 wide.**

**CODE:**

data("AirPassengers")

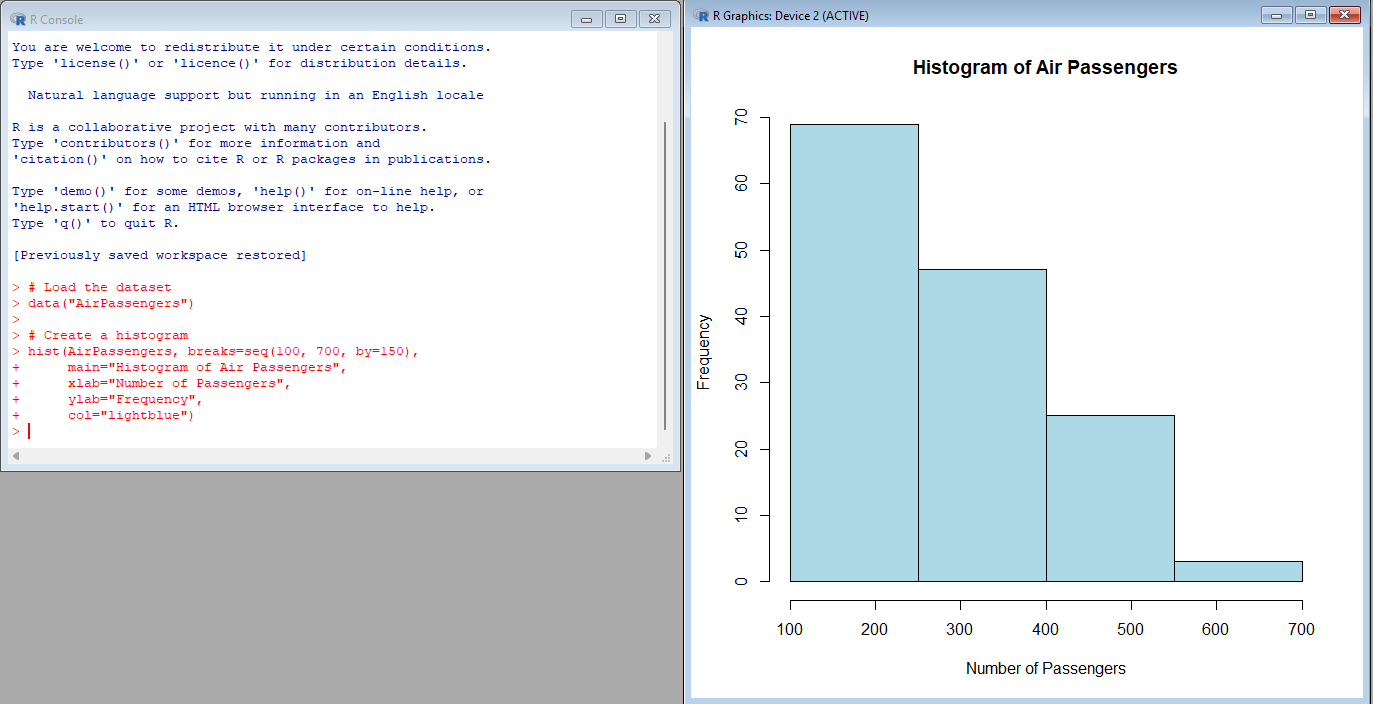
hist(AirPassengers, breaks=seq(100, 700, by=150),

main="Histogram of Air Passengers",

xlab="Number of Passengers",

ylab="Frequency",

col="lightblue")

**OUTPUT: **

**6. Obtain Multiple Lines in Line Chart using a single Plot Function in R.Use attributes“mpg”and“qsec”of the dataset “mtcars”.**

**CODE:**

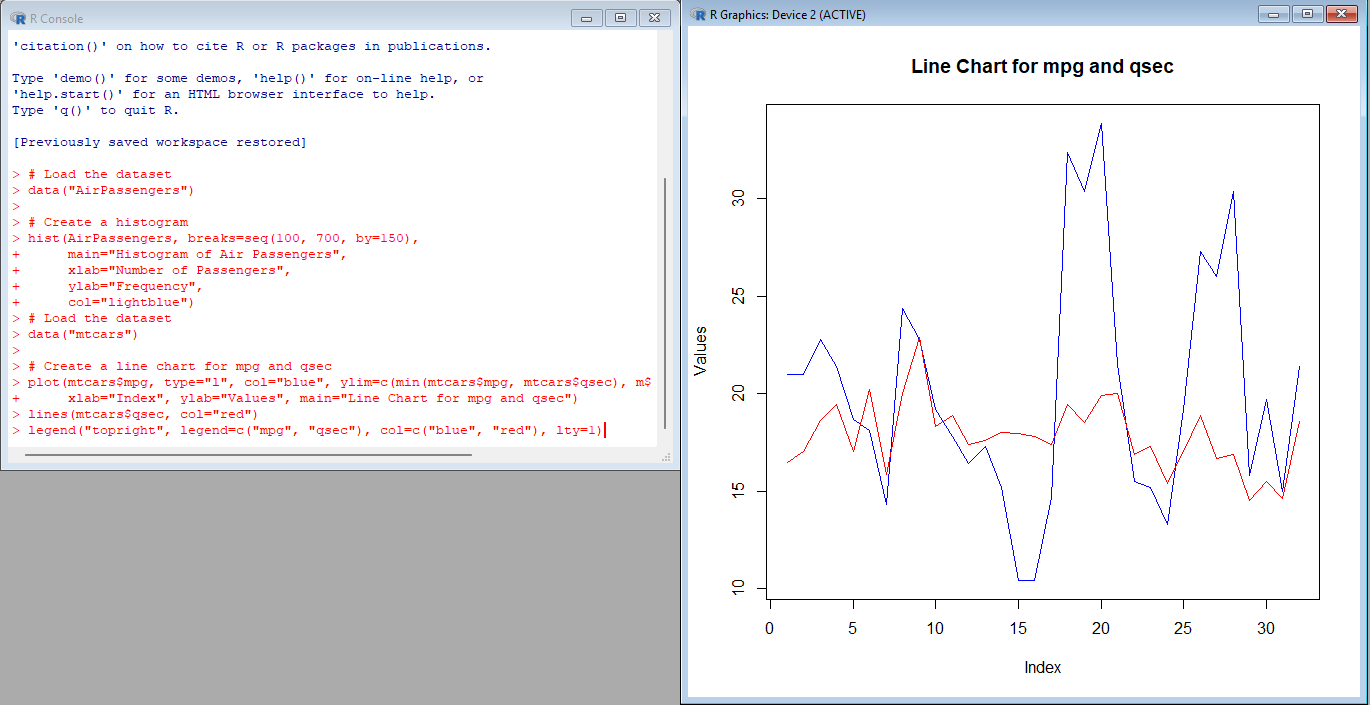
data("mtcars")

plot(mtcars$mpg, type="l", col="blue", ylim=c(min(mtcars$mpg, mtcars$qsec), max(mtcars$mpg, mtcars$qsec)),

xlab="Index", ylab="Values", main="Line Chart for mpg and qsec")

lines(mtcars$qsec, col="red")

legend("topright", legend=c("mpg", "qsec"), col=c("blue", "red"), lty=1)

**OUTPUT:**

**8. .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.**

**CODE:**

**data("mtcars")**

**boxplot(mpg ~ factor(cyl, levels=unique(mtcars$cyl)),**

**data=mtcars,**

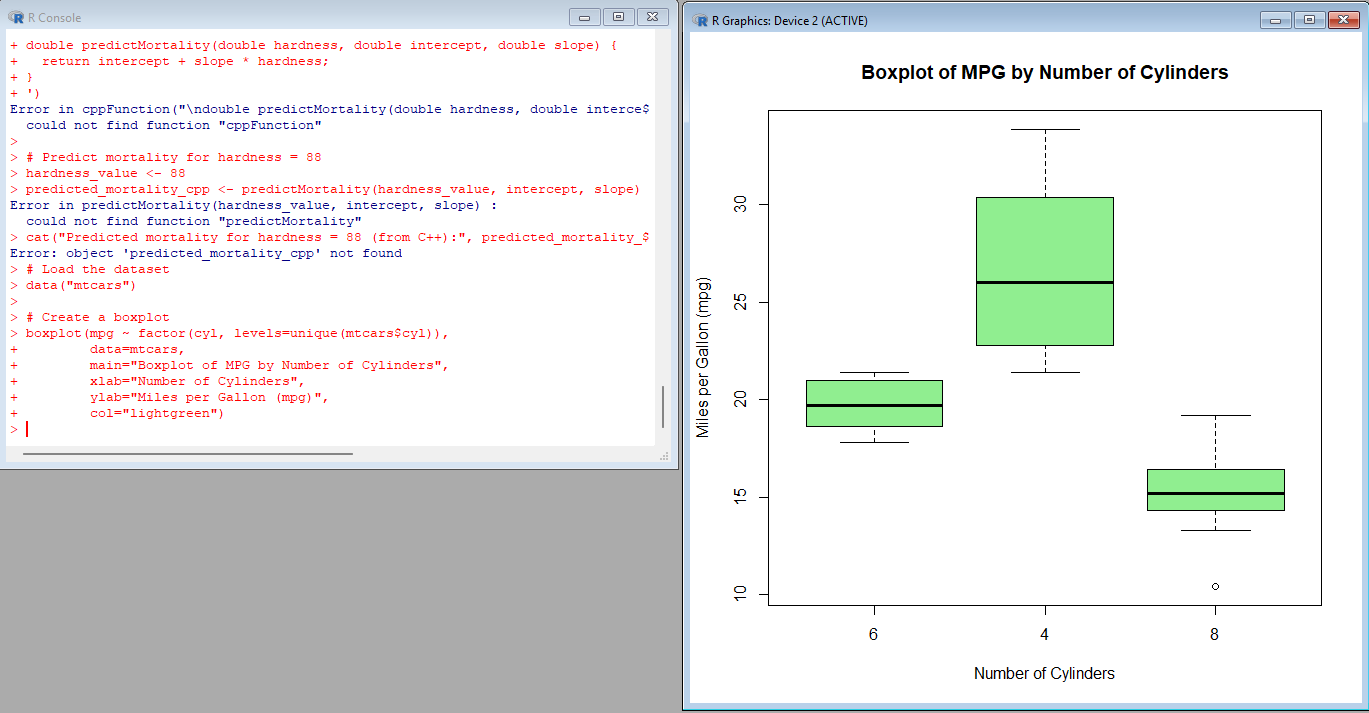
**main="Boxplot of MPG by Number of Cylinders",**

**xlab="Number of Cylinders",**

**ylab="Miles per Gallon (mpg)",**

**col="lightgreen")**

**OUTPUT:**

****

**9. Assume the Tennis coach wants to determine if any of his team players are scoring**

**outliers. To visualize the distribution of points scored by his players, then how can he**

**decide to develop the box plot? Give suitable example using Boxplot visualization**

**technique.**

**CODE:**

**scores <- c(10, 12, 15, 14, 13, 20, 25, 30, 100) # 100 is an outlier**

**boxplot(scores,**

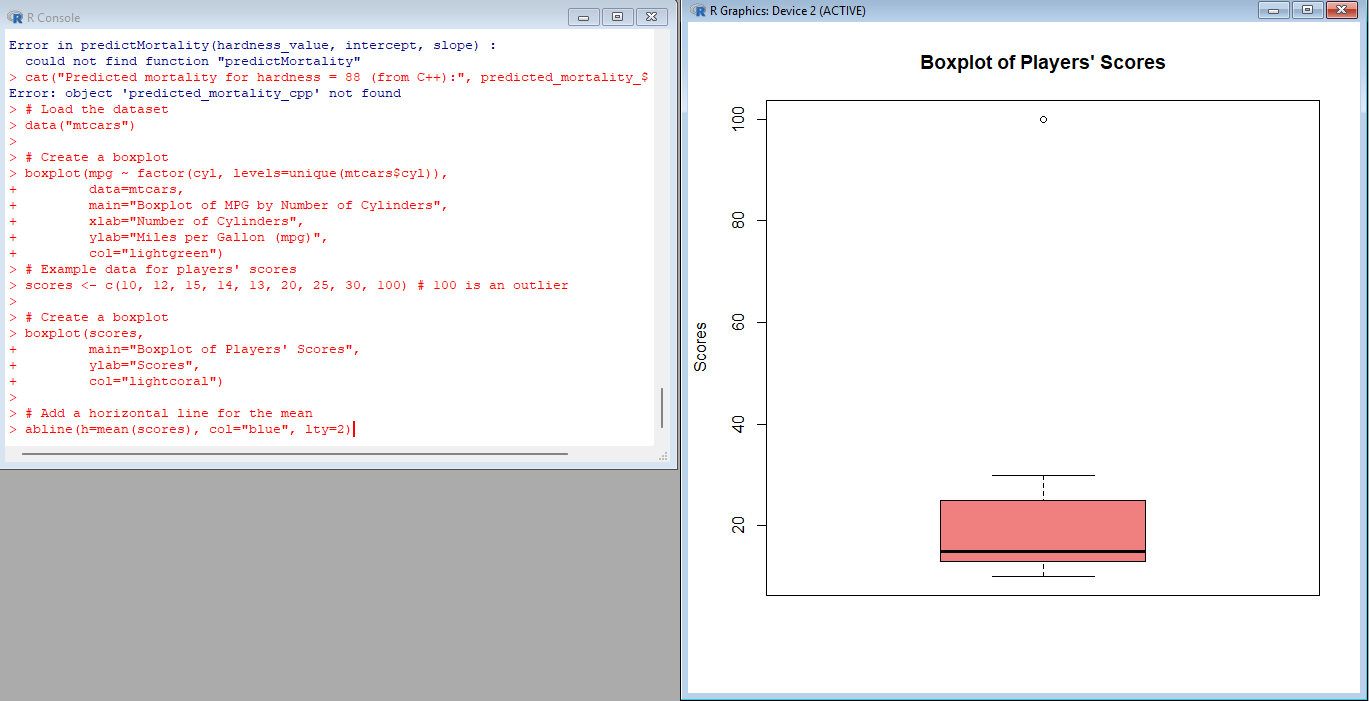
**main="Boxplot of Players' Scores",**

**ylab="Scores",**

**col="lightcoral")**

**abline(h=mean(scores), col="blue", lty=2)**

**OUTPUT:**

****

**10. Implement using R language in which age group of people are affected byblood pressure based on the diabetes dataset show it using scatterplot and bar chart (that is BloodPressure vs Age using dataset “diabetes.csv”)**

**CODE:**

**diabetes <- read.csv("diabetes.csv")**

**plot(diabetes$Age, diabetes$BloodPressure,**

**main="Blood Pressure vs Age",**

**xlab="Age",**

**ylab="Blood Pressure",**

**pch=19, col="purple")age\_groups <- cut(diabetes$Age, breaks=seq(0, 100, by=10))**

**avg\_bp <- tapply(diabetes$BloodPressure, age\_groups, mean)**

**barplot(avg\_bp,**

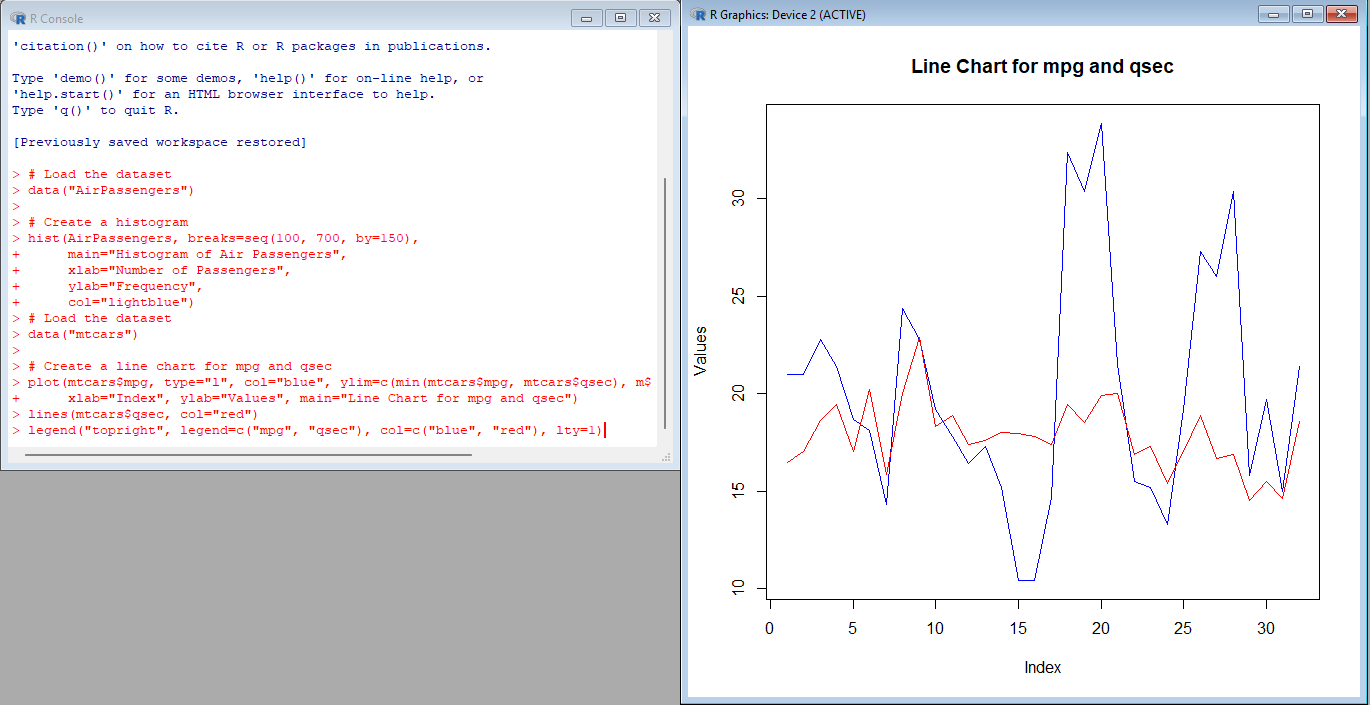
**main="Average Blood Pressure by Age Group",**

**xlab="Age Group",**

**ylab="Average Blood Pressure",**

**col="lightblue")**

**OUTPUT:**

****