**SAVEETHA SCHOOL OF ENGINEERING**

**SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES**

**ITA 0443 - STATISTICS WITH R PROGRAMMING FOR REAL TIME PROBLEM**

**DAY 4– LAB MANUAL Part 2**

**Reg No:** 192124154

**Name: K VENKATA ABHIRAM**

**LOGISTIC REGRESSION ANALYSIS IN R**

**Exercise**

**5. Create a logistic regression model using the “mtcars” data set with the information given**

**below.**

**The in-built data set &quot;mtcars&quot; describes different models of a car with their various engine**

**specifications. In &quot;mtcars&quot; data set, the transmission mode (automatic or manual) is described**

**by the column am which is a binary value (0 or 1). Create a logistic regression model**

**between the columns &quot;am&quot; and 3 other columns - hp, wt and cyl.**

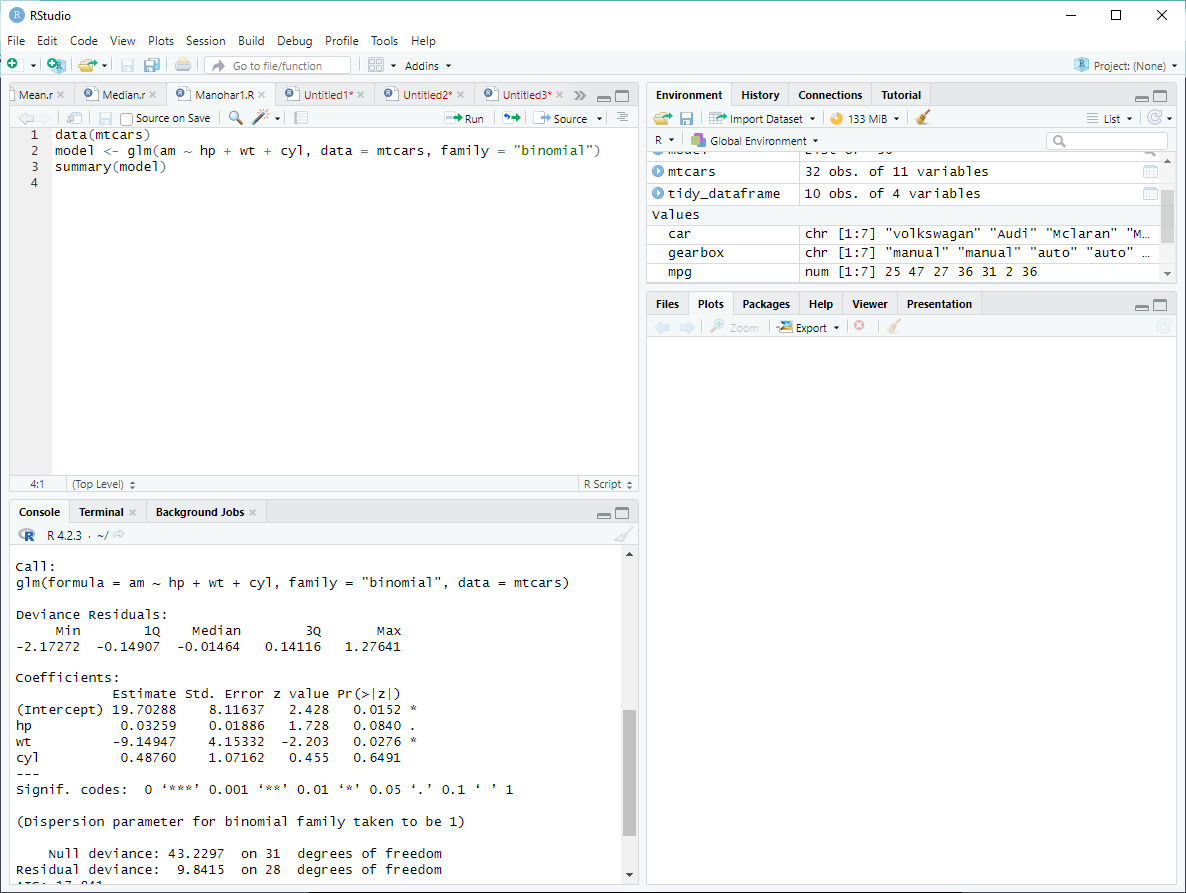
**Source Code:**

data(mtcars)

model <- glm(am ~ hp + wt + cyl, data = mtcars, family = "binomial")

summary(model)

**OUTPUT:**



**POISSON REGRESSION ANALYSIS IN R**

**Exercise :**

**6. Create a Poisson regression model using the in-built data set “warpbreaks” with**

**information given below.**

**In-built data set &quot;warpbreaks” describes the effect of wool type (A or B) and tension (low,**

**medium or high) on the number of warp breaks per loom. Consider &quot;breaks&quot; as the response**

**variable which is a count of number of breaks. The wool &quot;type&quot; and &quot;tension&quot; are taken as**

**predictor variables.**

**1.Randomly Sample the iris dataset such as 80% data for training and 20% for test and**

**create Logistics regression with train data, use species as target and petals width and**

**length as feature variables , Predict the probability of the model using test data,  Create**

**Confusion matrix for above test model**

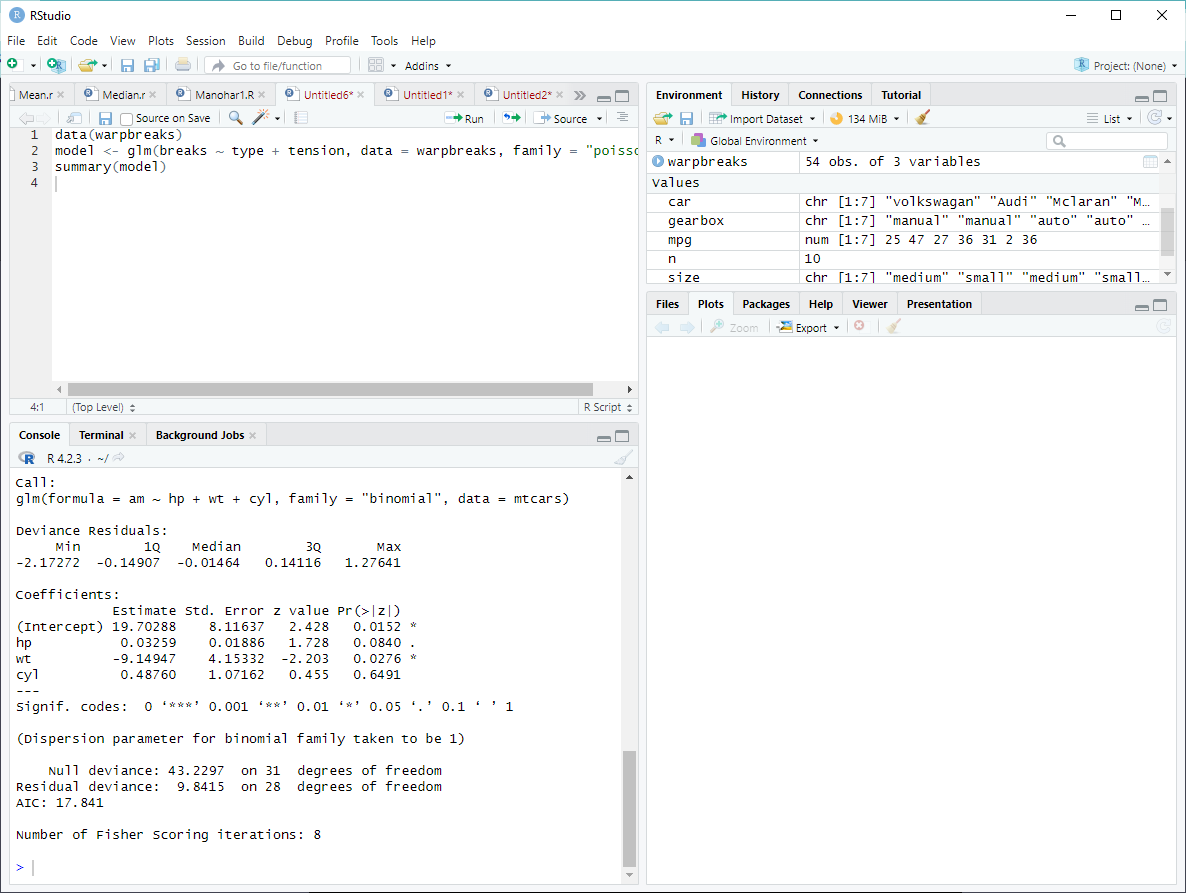
**Source Code:**

data(warpbreaks)

model <- glm(breaks ~ type + tension, data = warpbreaks, family = "poisson")

summary(model)

**OUTPUT:**



**2. (i)Write suitable R code to compute the mean, median ,mode of the following values**

**c(90, 50, 70, 80, 70, 60, 20, 30, 80, 90, 20)**

**Source Code:**

# Define the vector

values <- c(90, 50, 70, 80, 70, 60, 20, 30, 80, 90, 20)

# Compute the mean

mean(values)

# Compute the median

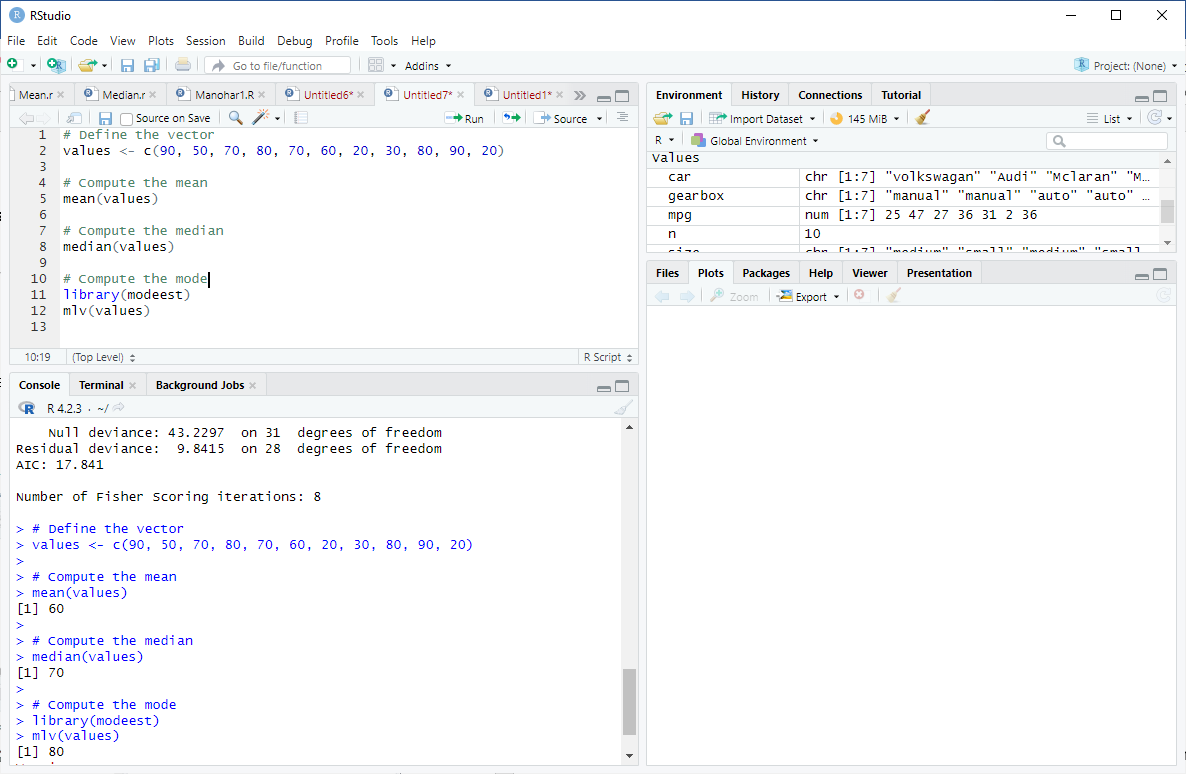
median(values)

# Compute the mode

library(modeest)

mlv(values)

**OUTPUT:**



**(ii) Write R code to find 2nd  highest and 3 rd Lowest value of above problem.**

**Source Code:**

# Define the vector

values <- c(90, 50, 70, 80, 70, 60, 20, 30, 80, 90, 20)

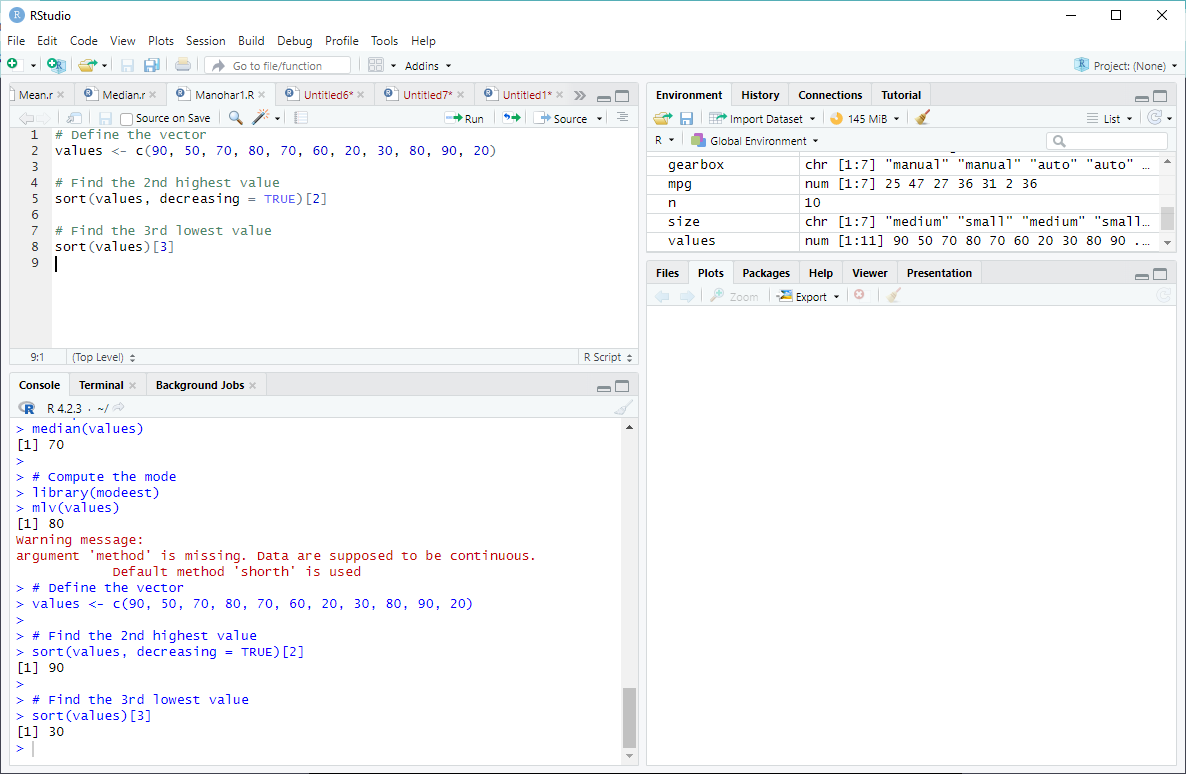
# Find the 2nd highest value

sort(values, decreasing = TRUE)[2]

# Find the 3rd lowest value

sort(values)[3]

**OUTPUT:**



**3. Explore the airquality dataset. It contains daily air quality measurements from New York**

**during a period of five months:**

**• Ozone: mean ozone concentration (ppb), • Solar.R: solar radiation (Langley),**

**• Wind: average wind speed (mph), • Temp: maximum daily temperature in degrees**

**Fahrenheit,**

**• Month: numeric month (May=5, June=6, and so on),• Day: numeric day of the month (1 -**

**4).**

**i. Compute the mean temperature(don’t use build in function)**

**Source Code:**

# Load the dataset

data(airquality)

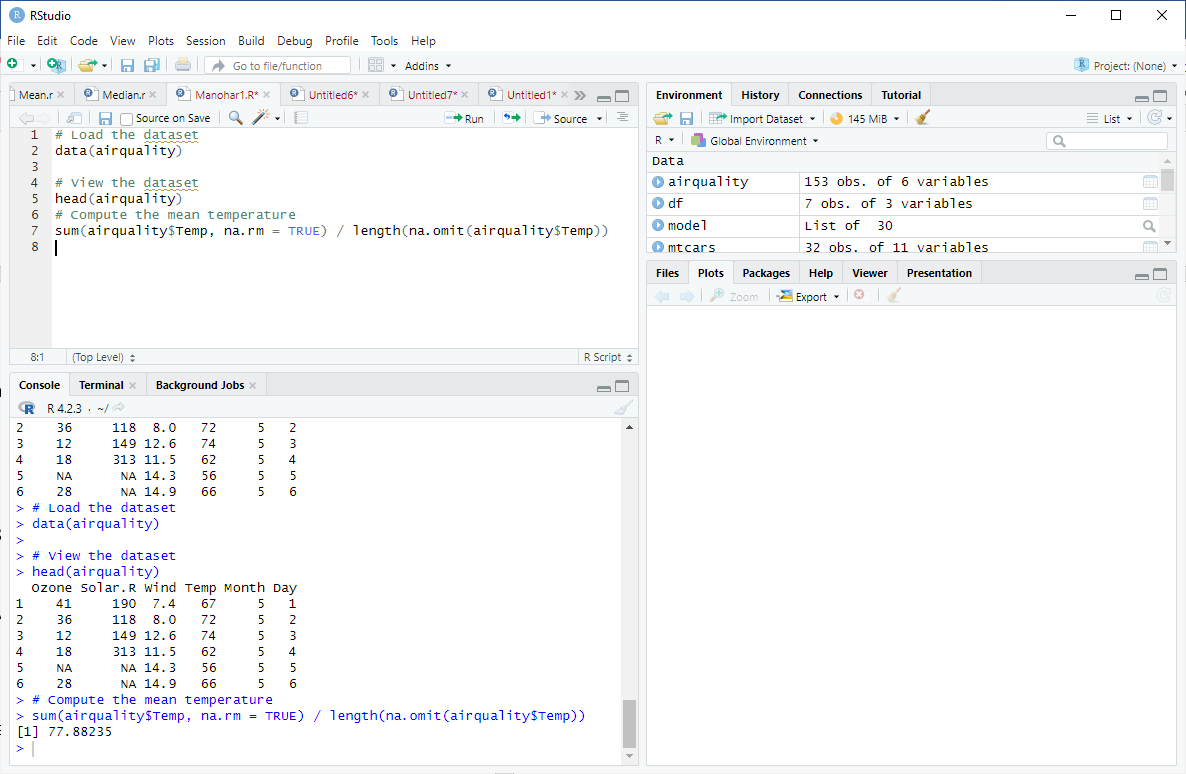
# View the dataset

head(airquality)

# Compute the mean temperature

sum(airquality$Temp, na.rm = TRUE) / length(na.omit(airquality$Temp))

**OUTPUT:**

****

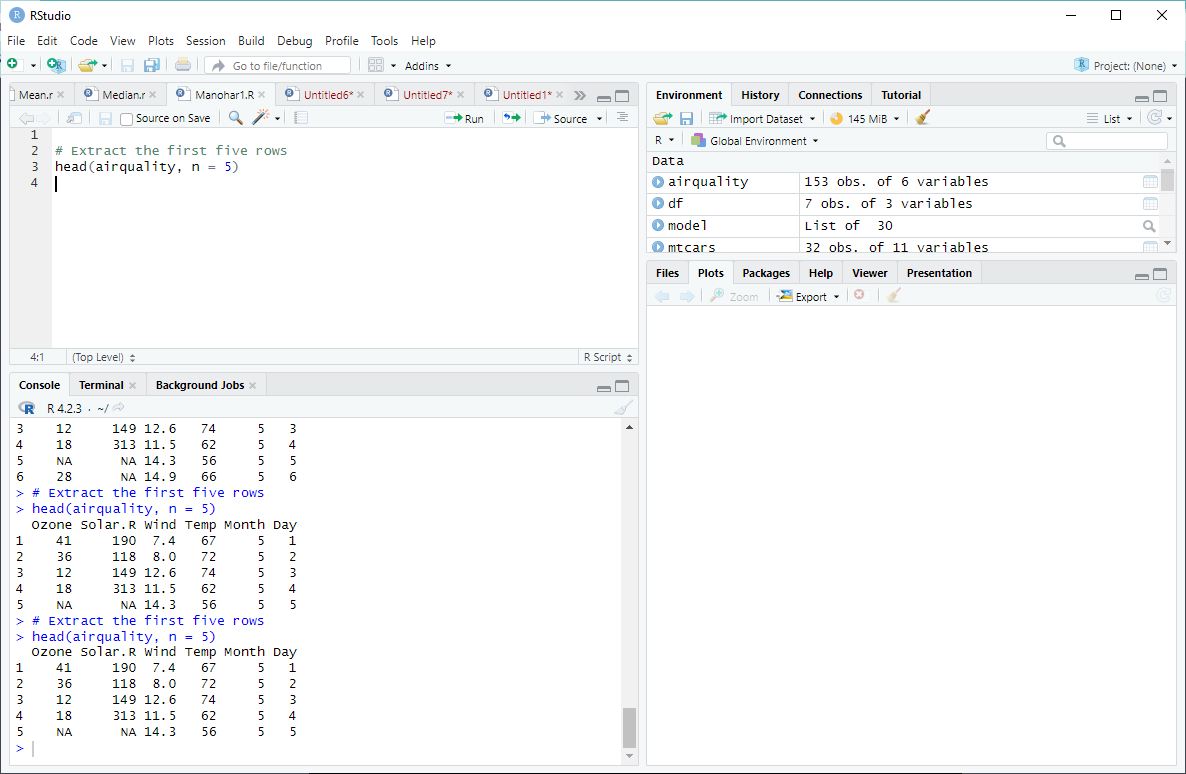
**ii. Extract the first five rows from air quality.**

**Source Code:**

# Extract the first five rows

head(airquality, n = 5)

**OUTPUT:**

****

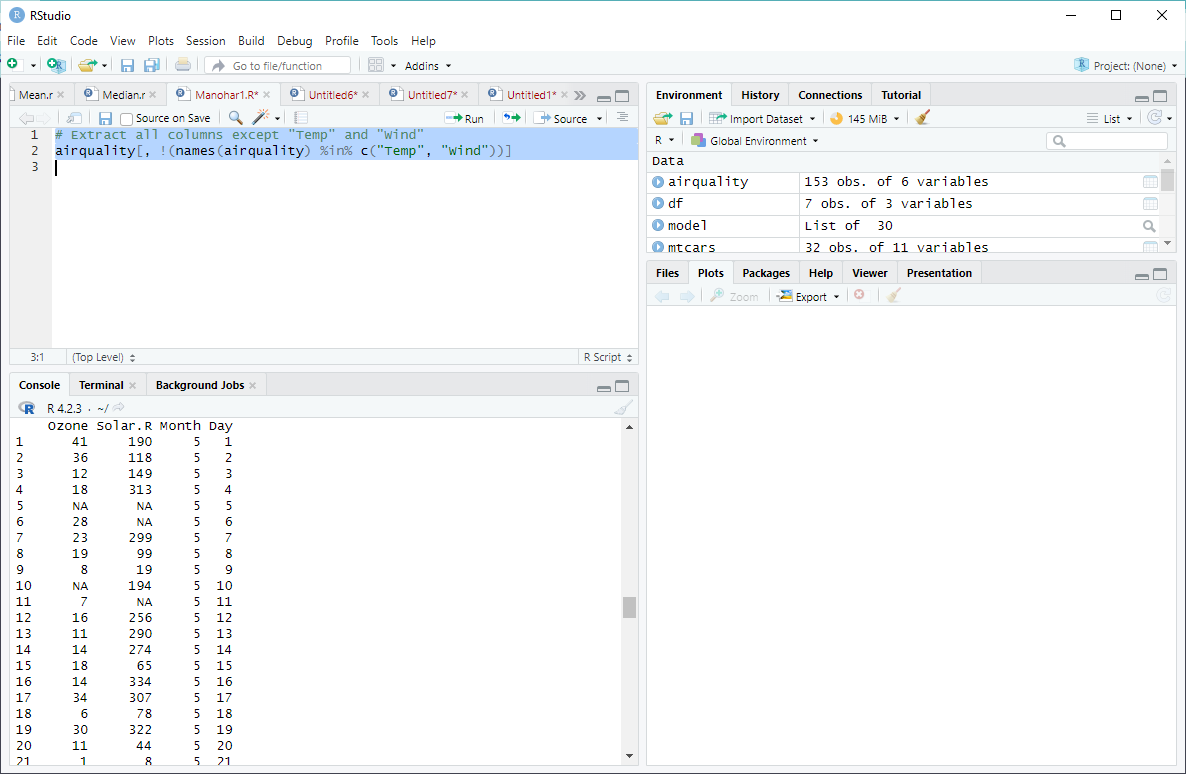
**iii. Extract all columns from air quality except Temp and Wind**

**Source Code:**

# Extract all columns except "Temp" and "Wind"

airquality[, !(names(airquality) %in% c("Temp", "Wind"))]

**OUTPUT:**

****

**iv. Which was the coldest day during the period?**

**Source Code:**

# Find the coldest day

airquality[which.min(airquality$Temp), c("Month", "Day")]

**OUTPUT:**

