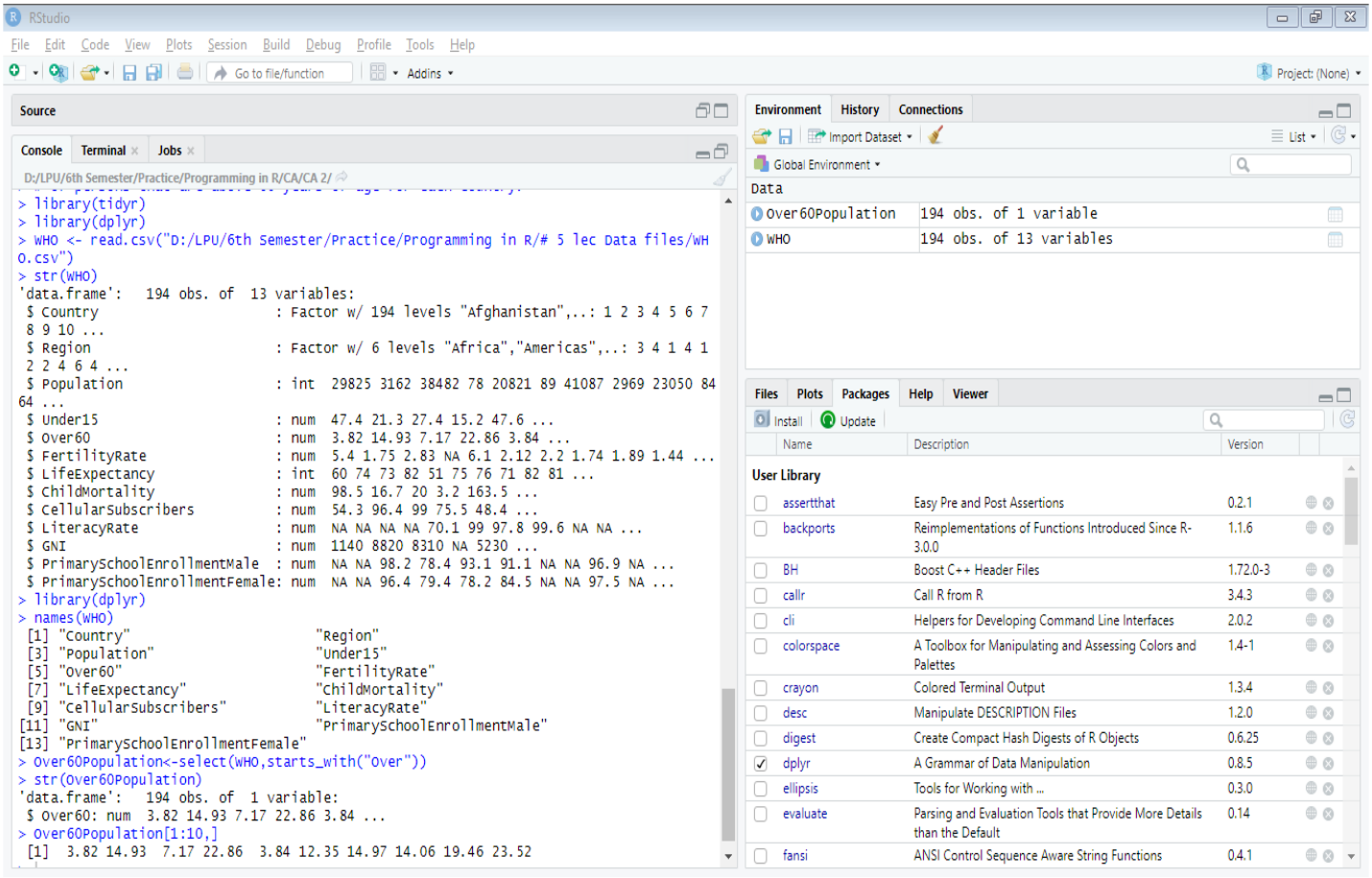
Q1. Using 'WHO' dataset, derive a new column 'Over60Population' that contains number of persons that are above 60 years of age for each country.

Answer1.

```
library(tidyr)
library(dplyr)
WHO <- read.csv("D:/LPU/6th Semester/Practice/Programming in R/# 5 lec Data files/WHO.csv")
str(WHO)
names(WHO)
Over60Population<-select(WHO,starts_with("Over"))
str(Over60Population)
Over60Population[1:10,]
```



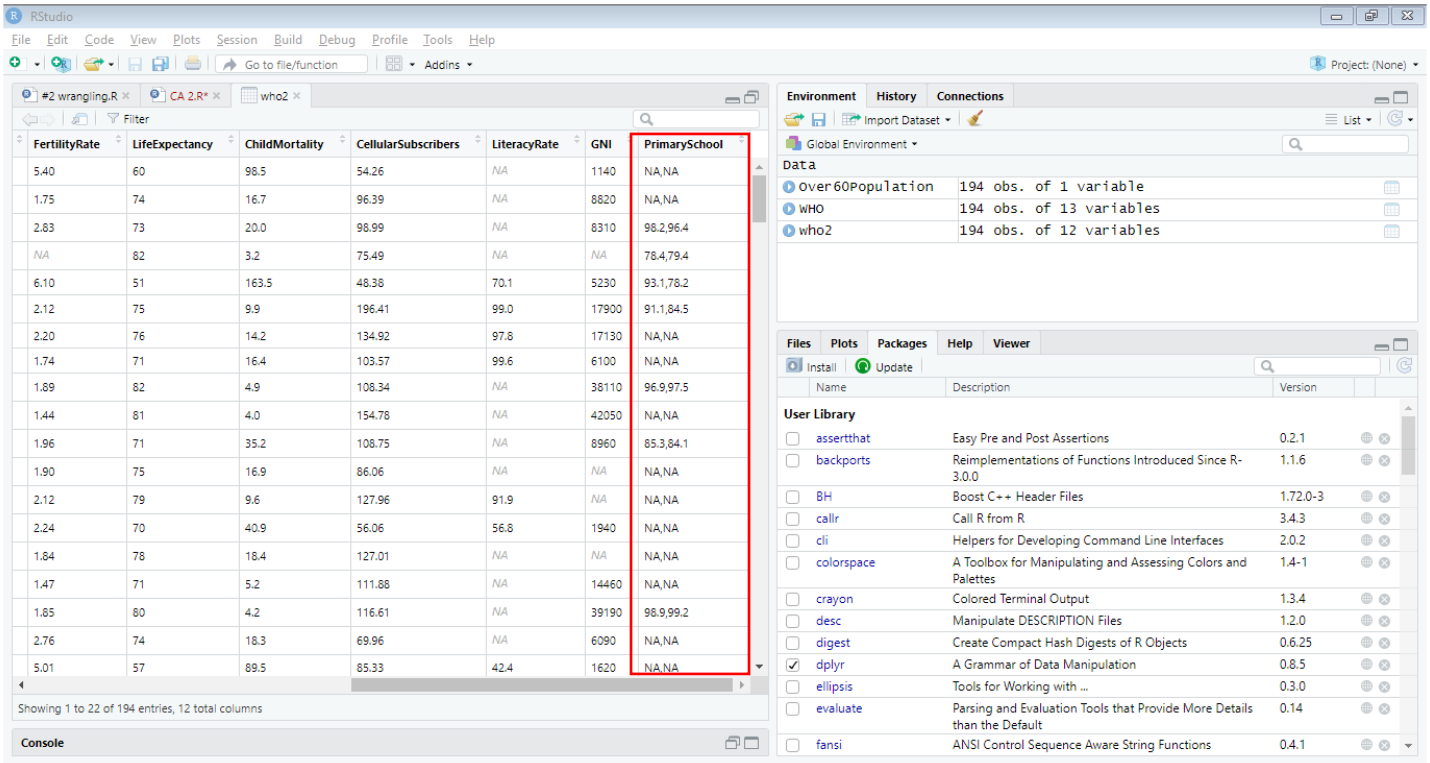
Q2. Using ‘WHO’ dataset, derive a new dataset who2, where ‘PrimarySchoolEnrollmentMale’ and ‘PrimarySchoolEnrollmentFemale’ are united to give a new column ‘PrimarySchool’ with comma as a separator.

Answer2.

```
library(tidyr)

who2<-unite(WHO,"PrimarySchool", PrimarySchoolEnrollmentMale,
            PrimarySchoolEnrollmentFemale,sep=",")

str(who2)

View(who2)
```

Output:

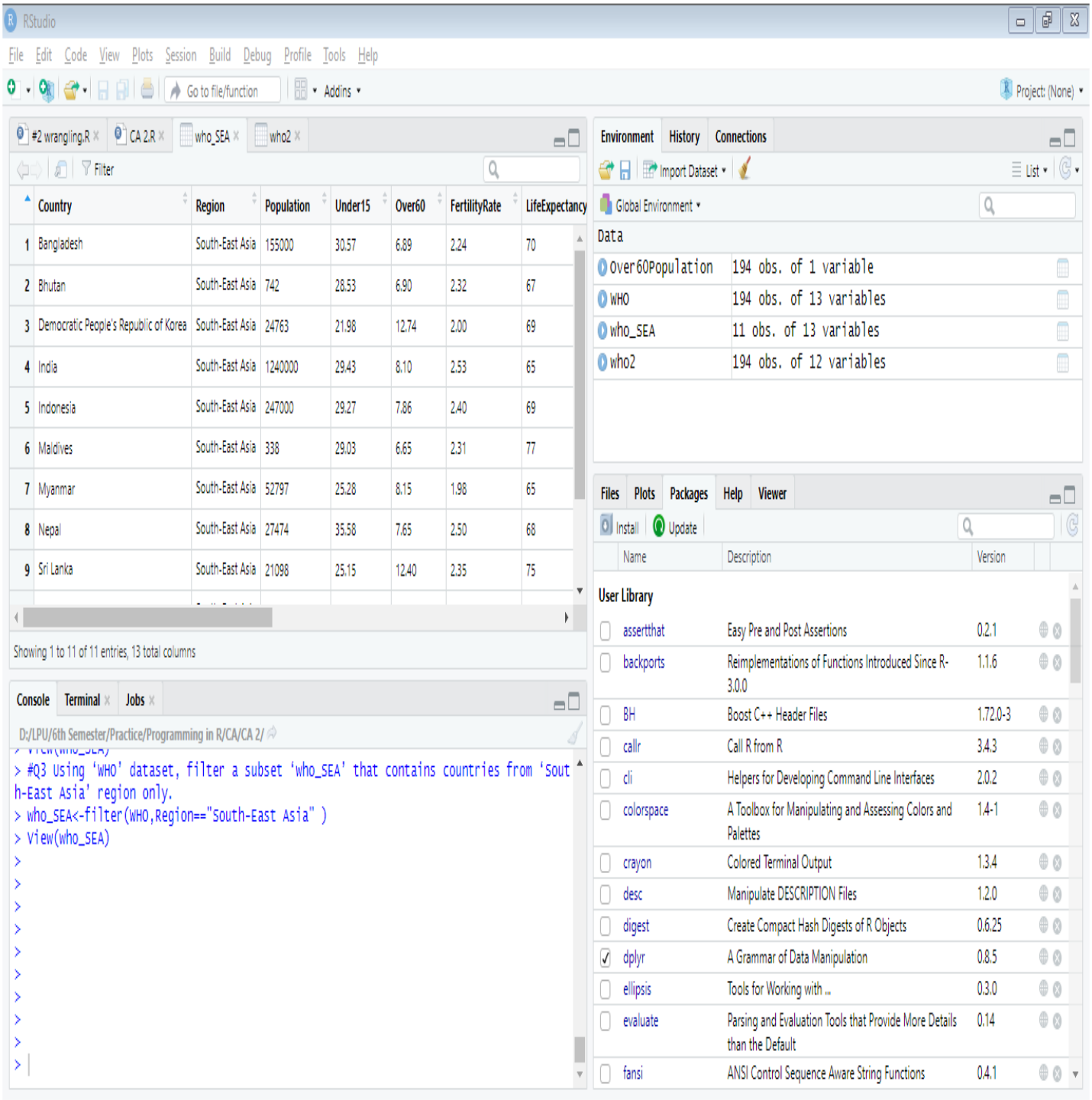
```
'data.frame':  194 obs. of  12 variables:
 $ Country             : Factor w/ 194 levels "Afghanistan",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Region              : Factor w/ 6 levels "Africa","Americas",..: 3 4 1 4 1 2 2 4 6 4 ..
.
 $ Population          : int  29825 3162 38482 78 20821 89 41087 2969 23050 8464 ...
 $ Under15             : num  47.4 21.3 27.4 15.2 47.6 ...
 $ Over60              : num  3.82 14.93 7.17 22.86 3.84 ...
 $ FertilityRate       : num  5.4 1.75 2.83 NA 6.1 2.12 2.2 1.74 1.89 1.44 ...
 $ LifeExpectancy      : int  60 74 73 82 51 75 76 71 82 81 ...
 $ ChildMortality      : num  98.5 16.7 20 3.2 163.5 ...
 $ CellularSubscribers: num  54.3 96.4 99 75.5 48.4 ...
 $ LiteracyRate        : num  NA NA NA NA 70.1 99 97.8 99.6 NA NA ...
 $ GNI                 : num  1140 8820 8310 NA 5230 ...
 $ PrimarySchool       : chr  "NA,NA" "NA,NA" "98.2,96.4" "78.4,79.4" ...
```



Q3. Using ‘WHO’ dataset, filter a subset ‘who_SEA’ that contains countries from ‘South-East Asia’ region only.

Answer3.

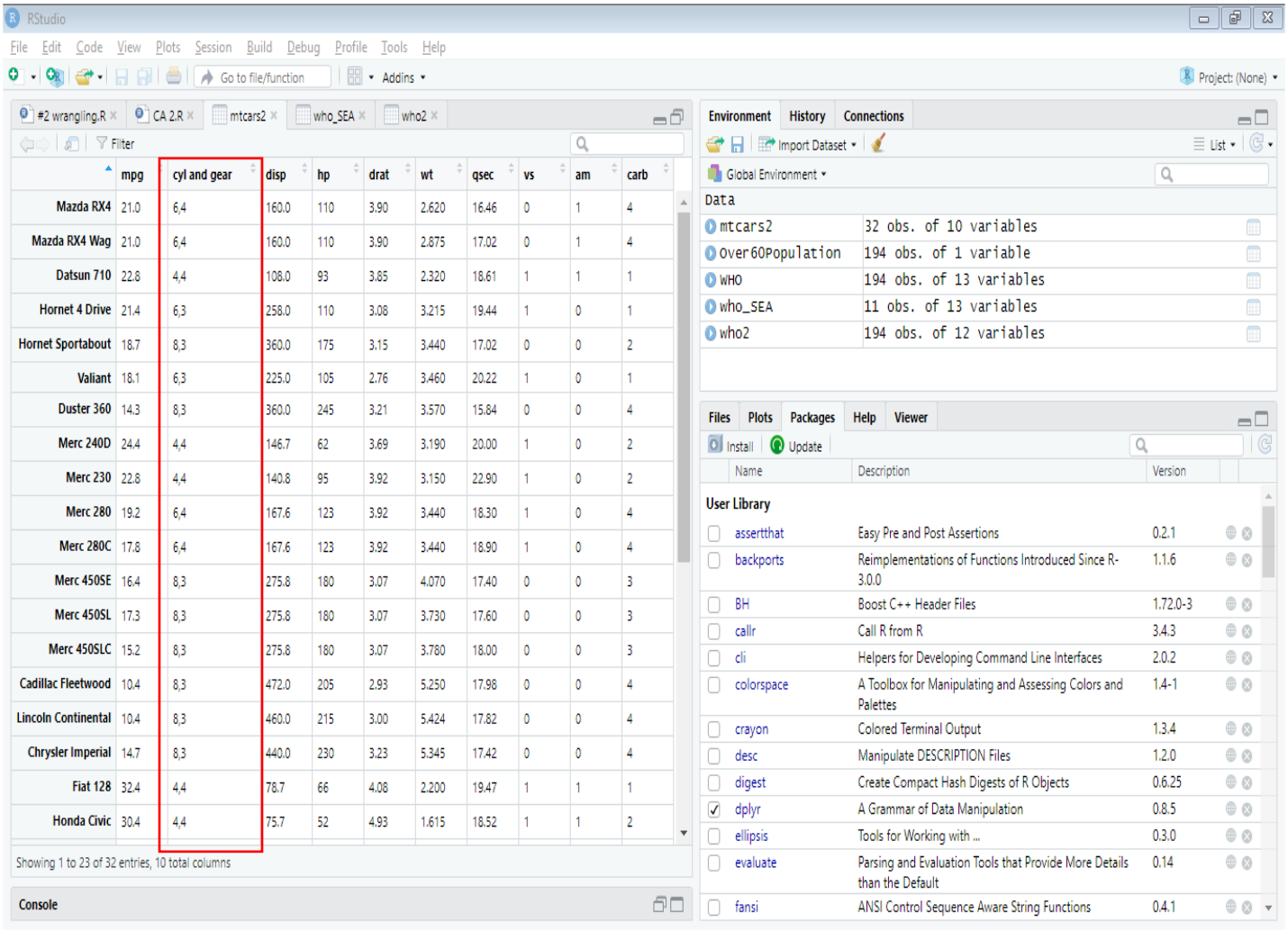
```
who_SEA<-filter(WHO,Region=="South-East Asia" )

View(who_SEA)
```



Q4. Using ‘mtcars’ datset, where ‘cyl’ and ‘gear’ are united to give a new column named as ‘cyl and gear’ with comma as a separator between two values.

Answer4.

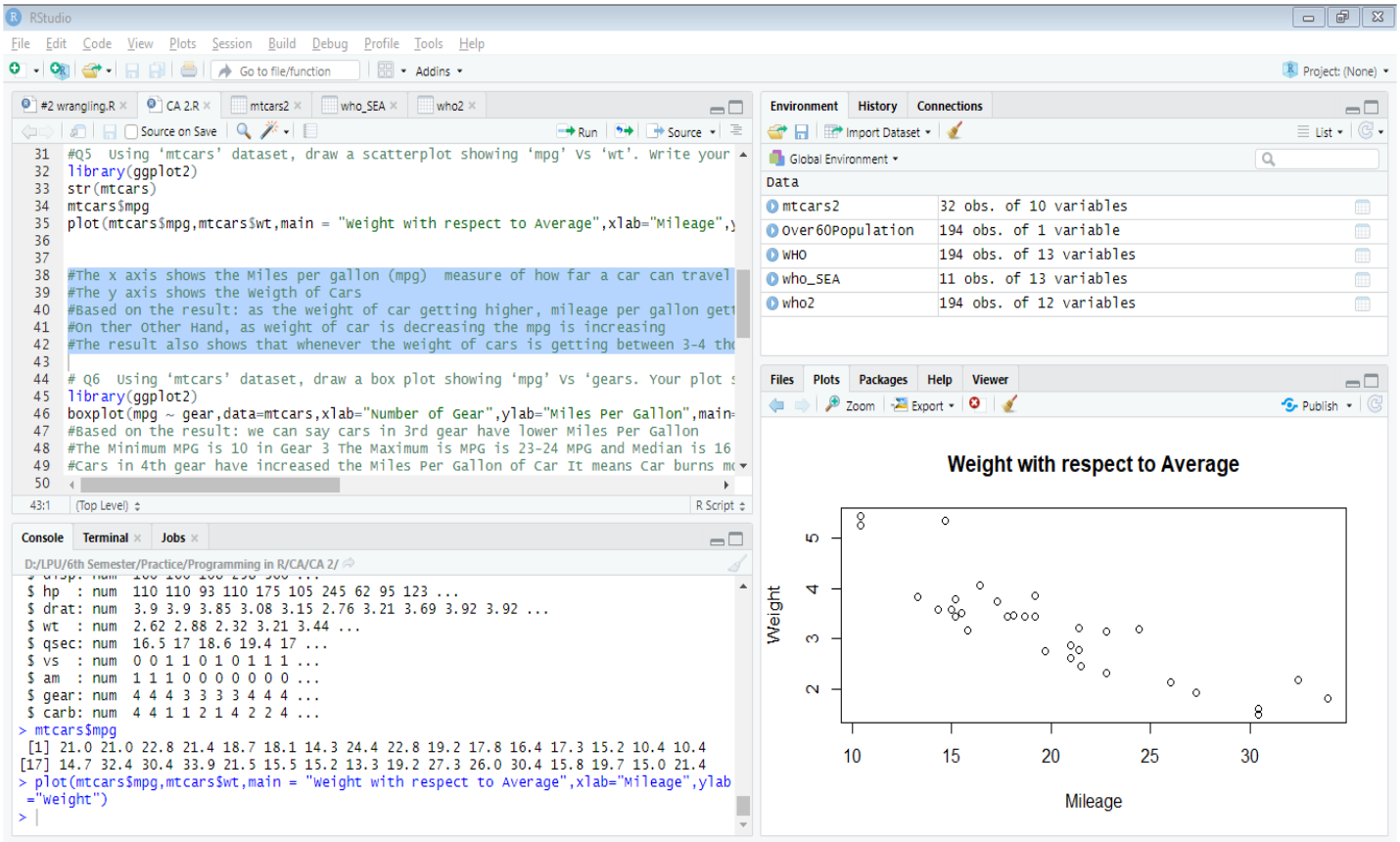
```
who_SEA<-filter(WHO,Region=="South-East Asia" )

library(tidyr)

str(mtcars)

mtcars2<-unite(mtcars,"cyl and gear",cyl,gear,sep=",")

View(mtcars2)
```



Q5. Using ‘mtcars’ dataset, draw a scatterplot showing ‘mpg’ Vs ‘wt’. Write your interpretation about the graph.

Answer5.

```
library(ggplot2)

str(mtcars)

mtcars$mpg

plot(mtcars$mpg,mtcars$wt,main = "Weight with respect to Average",xlab="Mileage",ylab
="Weight")
```



- The x axis shows the Miles per gallon (mpg) measure of how far a car can travel if you put just one gallon of petrol or diesel in its tank.
- The y axis shows the Weight of Cars
- Based on the result: as the weight of car getting higher, mileage per gallon getting lower
- On the Other Hand, as weight of car is decreasing the mpg is increasing
- The result also shows that whenever the weight of cars is getting between 3-4 thousand lbs. the MPG increases

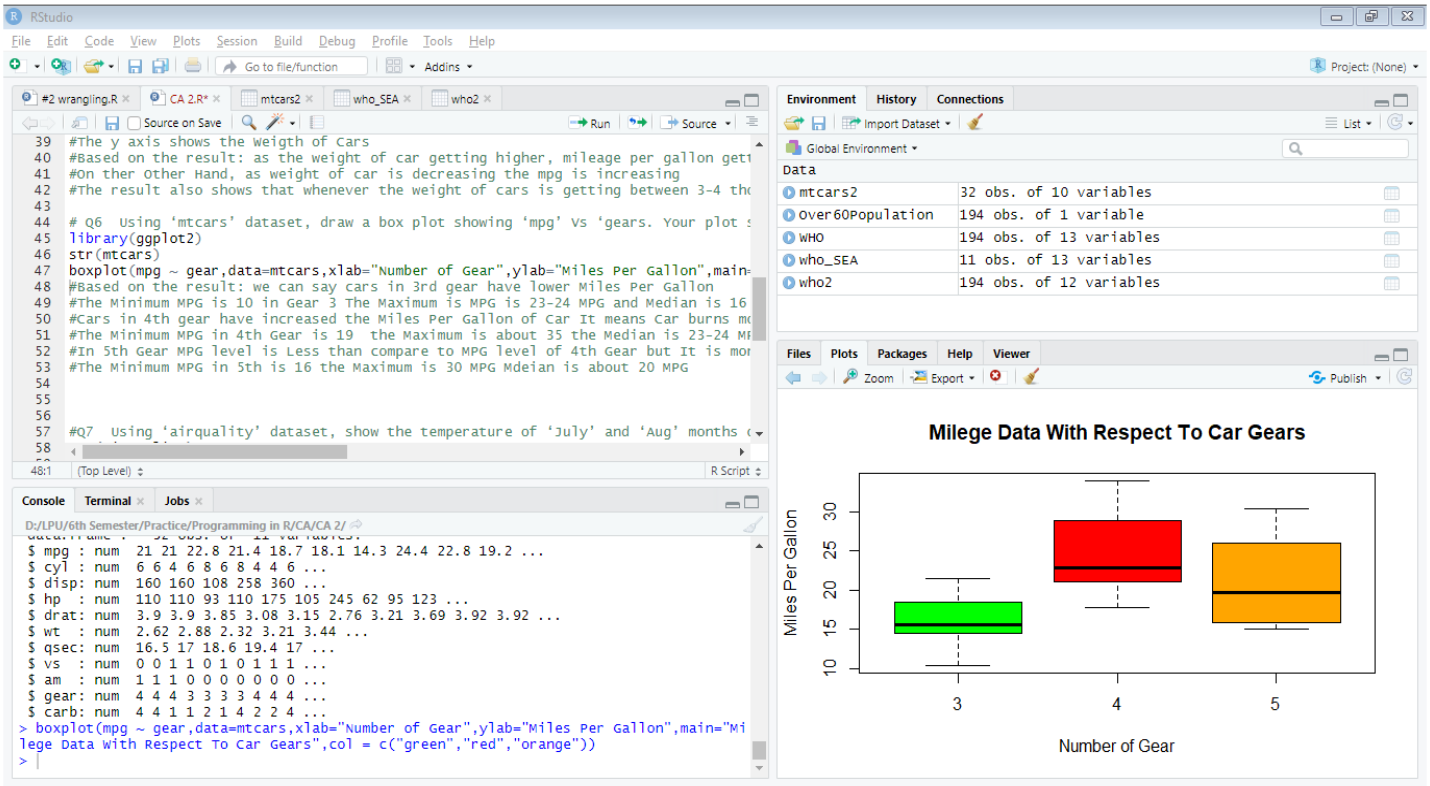
Q6. Using ‘mtcars’ dataset, draw a box plot showing ‘mpg’ Vs ‘gears. Your plot should have three box-plots separate for cars with 3, 4 & 5 gears. Write your interpretation of the plot.

Answer6.

library(ggplot2)

str(mtcars)

boxplot(mpg ~ gear, data=mtcars, xlab="Number of Gear", ylab="Miles Per Gallon", main="Milege Data With Respect To Car Gears", col = c("green","red","orange"))

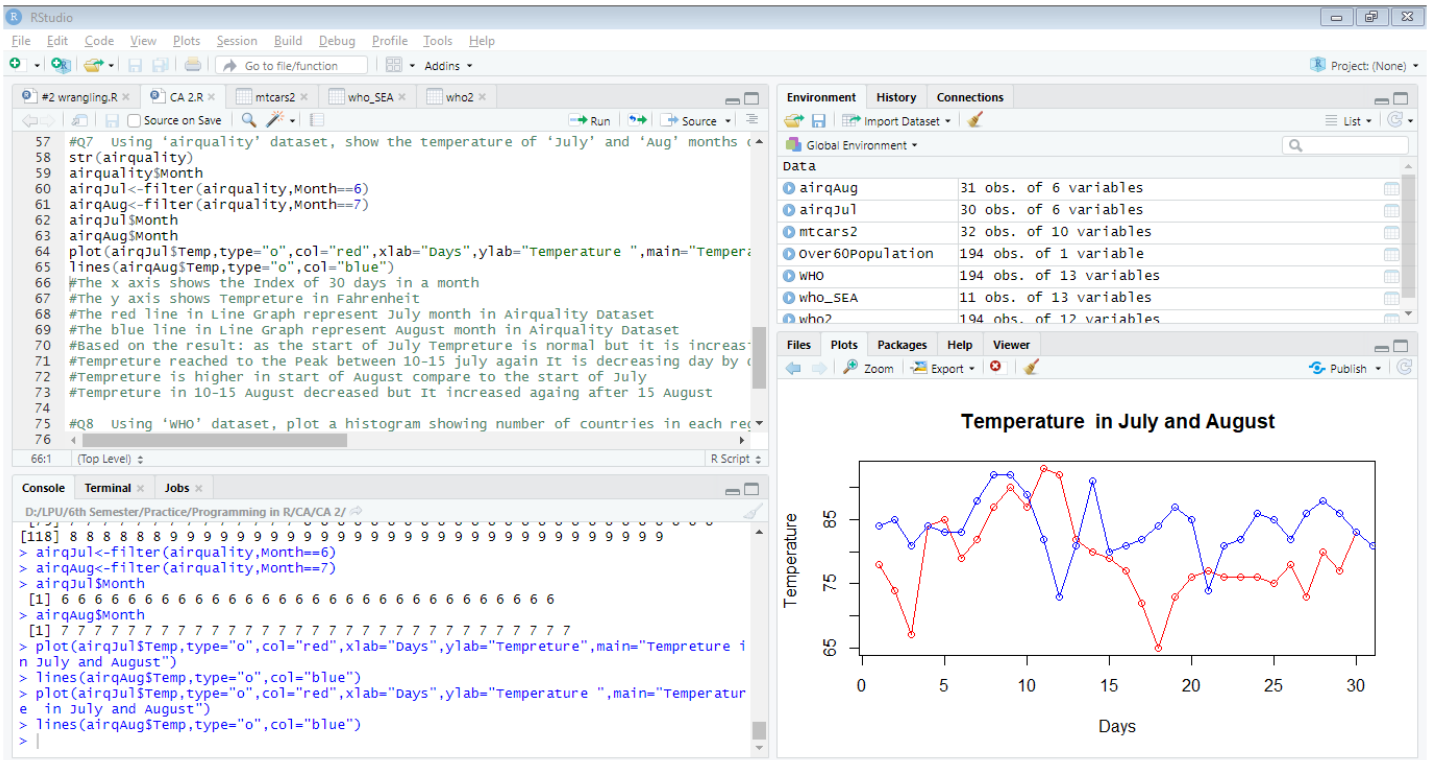


- Based on the result: we can say cars in 3rd gear have lower Miles Per Gallon
- The Minimum MPG is 10 in Gear 3 The Maximum is MPG is 23-24 MPG and Median is 16 MPG in 3rd Gear
- Cars in 4th gear have increased the Miles Per Gallon of Car It means Car burns more fuel in 4th Gear
- The Minimum MPG in 4th Gear is 19 the Maximum is about 35 the Median is 23-24 MPG in 4th Gear
- In 5th Gear MPG level is Less than compare to MPG level of 4th Gear but It is more than 3th Gear
- The Minimum MPG in 5th is 16 the Maximum is 30 MPG Median is about 20 MPG

Q7. Using ‘airquality’ dataset, show the temperature of ‘July’ and ‘Aug’ months on the same line graph. Use separate color for both the lines. Write your interpretation about the plot.

Answer7.

```
str(airquality)
airquality$Month
airqJul<-filter(airquality,Month==6)
airqAug<-filter(airquality,Month==7)
airqJul$Month
airqAug$Month
plot(airqJul$Temp,type="o", col="red", xlab="Days", ylab=" Temperature ",
main=" Temperature in July and August")
lines(airqAug$Temp,type="o",col="blue")
```



- The x axis shows the Index of 30 days in a month
- The y axis shows Temperature in Fahrenheit
- The red line in Line Graph represent July month in Airquality Dataset
- The blue line in Line Graph represent August month in Airquality Dataset
- Based on the result: as the start of July Temperature is normal but it is increasing day by day
- Temperature reached to the Peak between 10-15 July again It is decreasing day by day
- Temperature is higher in start of August compare to the start of July
- Temperature in 10-15 August decreased but It increased again after 15 August

Q8. Using ‘WHO’ dataset, plot a histogram showing number of countries in each region.
Answer8.

```
library(ggplot2)

WHO <- read.csv("D:/LPU/6th Semester/Practice/Programming in R/# 5 lec Data files/WHO.csv")

Regions<-as.numeric(factor(WHO$Region))

hist(Regions,main=" Histogram of Countries in Each Region", xlab="Regions",ylab="Frequency",col=rainbow(6),border="blue")

legend("topright",c("Africa","Americas","Eastern Mediterranean","Europe","South-East Asia","Western Pacific"),cex =0.65,fill=rainbow(6))
```