

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,]
```

The screenshot shows the RStudio interface with the following components:

- Source Panel:** Contains the R code used to load the 'WHO' dataset and create the 'Over60Population' column.
- Console Panel:** Displays the output of the R code, including the structure of the 'WHO' dataset and the first 10 rows of the 'Over60Population' column.
- Environment Panel:** Shows the objects in the global environment, including 'Over60Population' and 'WHO'.
- Files Panel:** Shows the files in the current project.
- Plots Panel:** Shows the plots in the current project.
- Packages Panel:** Shows the packages installed in the current project.
- Help Panel:** Shows the help documentation for the 'dplyr' package.

Console Output:

```
> library(tidyr)
> library(dplyr)
> WHO <- read.csv("D:/LPU/6th Semester/Practice/Programming in R/# 5 lec Data files/WHO.csv")
> str(WHO)
'data.frame': 194 obs. of 13 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 84
 64 ...
 $ 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 ...
 $ PrimarySchoolEnrollmentMale : num  NA NA 98.2 78.4 93.1 91.1 NA NA 96.9 NA ...
 $ PrimarySchoolEnrollmentFemale : num  NA NA 96.4 79.4 78.2 84.5 NA NA 97.5 NA ...
> library(dplyr)
> names(WHO)
[1] "country"      "Region"
[3] "population"   "Under15"
[5] "over60"       "FertilityRate"
[7] "LifeExpectancy" "ChildMortality"
[9] "CellularSubscribers" "LiteracyRate"
[11] "GNI"          "PrimarySchoolEnrollmentMale"
[13] "PrimarySchoolEnrollmentFemale"
> over60Population<-select(WHO,starts_with("Over"))
> str(over60Population)
'data.frame': 194 obs. of 1 variable:
 $ over60: num  3.82 14.93 7.17 22.86 3.84 ...
> over60Population[1:10,]
[1] 3.82 14.93 7.17 22.86 3.84 12.35 14.97 14.06 19.46 23.52
```

Environment Panel:

Object	Class	Attributes
Over60Population	data.frame	194 obs. of 1 variable
WHO	data.frame	194 obs. of 13 variables

Files Panel:

Name	Description	Version
assertthat	Easy Pre and Post Assertions	0.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.6
BH	Boost C++ Header Files	1.72.0-3
callr	Call R from R	3.4.3
cli	Helpers for Developing Command Line Interfaces	2.0.2
colorspace	A Toolbox for Manipulating and Assessing Colors and Palettes	1.4-1
crayon	Colored Terminal Output	1.3.4
desc	Manipulate DESCRIPTION Files	1.2.0
digest	Create Compact Hash Digests of R Objects	0.6.25
dplyr	A Grammar of Data Manipulation	0.8.5
ellipsis	Tools for Working with ...	0.3.0
evaluate	Parsing and Evaluation Tools that Provide More Details than the Default	0.14
fansi	ANSI Control Sequence Aware String Functions	0.4.1

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" ...
```

The screenshot shows the RStudio interface. The main window displays a data frame with 12 columns: FertilityRate, LifeExpectancy, ChildMortality, CellularSubscribers, LiteracyRate, GNI, and PrimarySchool. The 'PrimarySchool' column is highlighted with a red box. The 'Environment' pane on the right shows the 'who2' dataset with 194 observations and 12 variables. The 'User Library' pane on the right shows a list of installed packages.

FertilityRate	LifeExpectancy	ChildMortality	CellularSubscribers	LiteracyRate	GNI	PrimarySchool
5.40	60	98.5	54.26	NA	1140	NA,NA
1.75	74	16.7	96.39	NA	8820	NA,NA
2.83	73	20.0	98.99	NA	8310	98.2,96.4
NA	82	3.2	75.49	NA	NA	78.4,79.4
6.10	51	163.5	48.38	70.1	5230	93.1,78.2
2.12	75	9.9	196.41	99.0	17900	91.1,84.5
2.20	76	14.2	134.92	97.8	17130	NA,NA
1.74	71	16.4	103.57	99.6	6100	NA,NA
1.89	82	4.9	108.34	NA	38110	96.9,97.5
1.44	81	4.0	154.78	NA	42050	NA,NA
1.96	71	35.2	108.75	NA	8960	85.3,84.1
1.90	75	16.9	86.06	NA	NA	NA,NA
2.12	79	9.6	127.96	91.9	NA	NA,NA
2.24	70	40.9	56.06	56.8	1940	NA,NA
1.84	78	18.4	127.01	NA	NA	NA,NA
1.47	71	5.2	111.88	NA	14460	NA,NA
1.85	80	4.2	116.61	NA	39190	98.9,99.2
2.76	74	18.3	69.96	NA	6090	NA,NA
5.01	57	89.5	85.33	42.4	1620	NA,NA

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)
```

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Shows the R script with the following code:

```
> #Q3 Using 'WHO' dataset, filter a subset 'who_SEA' that contains countries from 'South-East Asia' region only.  
> who_SEA<-filter(WHO,Region=="South-East Asia" )  
> View(who_SEA)
```
- Environment Pane:** Lists the objects in the Global Environment:
 - over60Population: 194 obs. of 1 variable
 - WHO: 194 obs. of 13 variables
 - who_SEA: 11 obs. of 13 variables
 - who2: 194 obs. of 12 variables
- Viewer Pane:** Displays the filtered data for 'who_SEA' as a table with 11 rows and 7 columns.
- Console:** Shows the execution of the R commands.
- Package Manager:** Lists installed and available packages, with 'dplyr' checked.

Country	Region	Population	Under15	Over60	FertilityRate	LifeExpectancy
1 Bangladesh	South-East Asia	155000	30.57	6.89	2.24	70
2 Bhutan	South-East Asia	742	28.53	6.90	2.32	67
3 Democratic People's Republic of Korea	South-East Asia	24763	21.98	12.74	2.00	69
4 India	South-East Asia	1240000	29.43	8.10	2.53	65
5 Indonesia	South-East Asia	247000	29.27	7.86	2.40	69
6 Maldives	South-East Asia	338	29.03	6.65	2.31	77
7 Myanmar	South-East Asia	52797	25.28	8.15	1.98	65
8 Nepal	South-East Asia	27474	35.58	7.65	2.50	68
9 Sri Lanka	South-East Asia	21098	25.15	12.40	2.35	75

Q4. Using 'mtcars' dataset, 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)
```

The screenshot shows the RStudio interface. The main editor displays a data frame with 10 columns: mpg, cyl and gear, disp, hp, drat, wt, qsec, vs, am, and carb. The 'cyl and gear' column is highlighted with a red box, showing values like '6,4', '4,4', '6,3', etc. The Environment pane on the right shows the 'Data' section with the following objects:

Object	Size
mtcars2	32 obs. of 10 variables
Over60Population	194 obs. of 1 variable
WHO	194 obs. of 13 variables
who_SEA	11 obs. of 13 variables
who2	194 obs. of 12 variables

The User Library pane on the right shows a list of installed packages, including dplyr (checked), assertthat, backports, BH, callr, cli, colorspace, crayon, desc, digest, ellipsis, evaluate, and fansi.

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

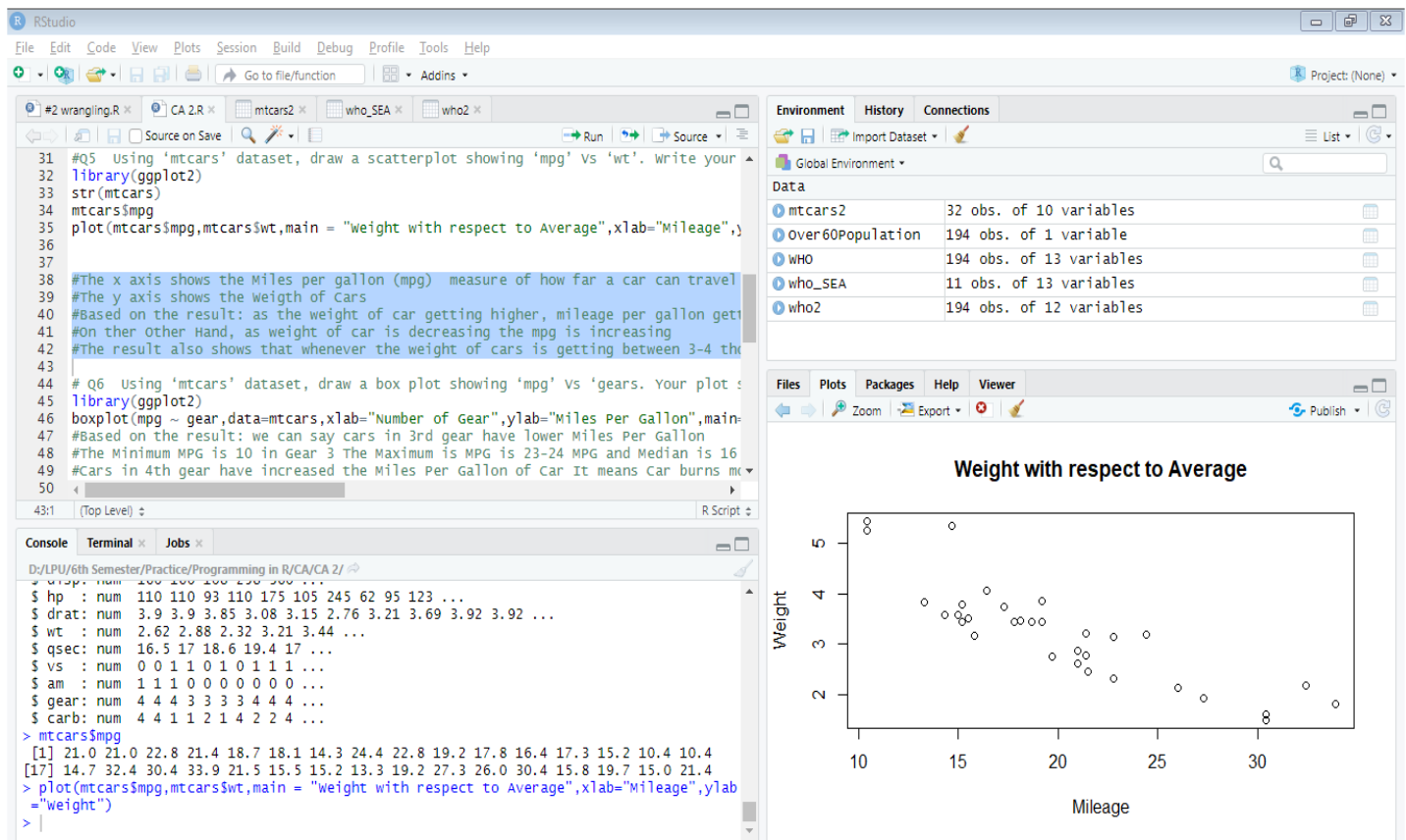
Answer5.

```
library(ggplot2)
```

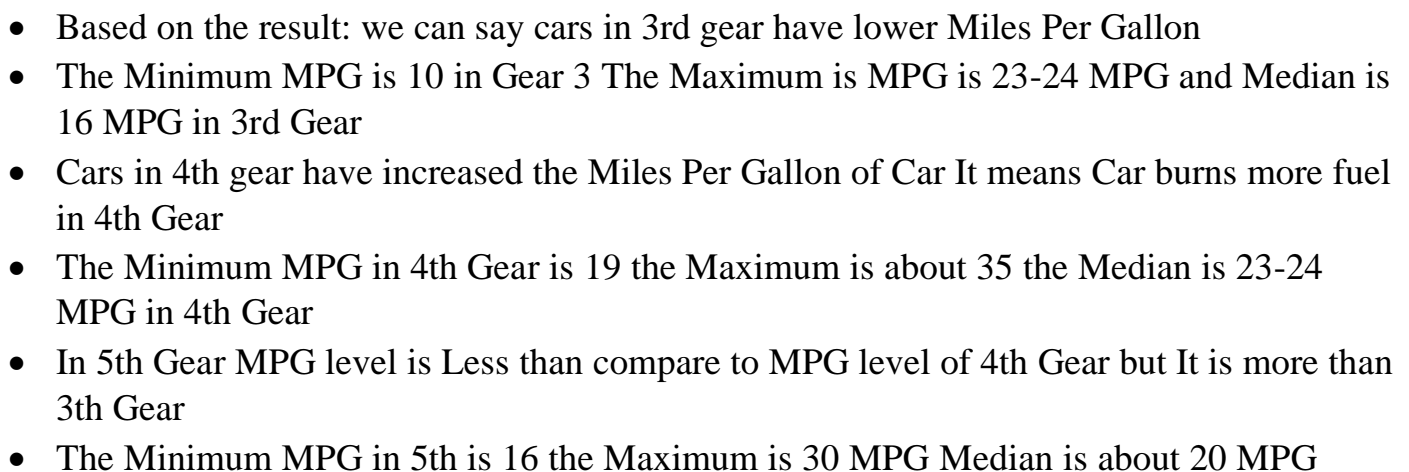
```
str(mtcars)
```

```
mtcars$mpg
```

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



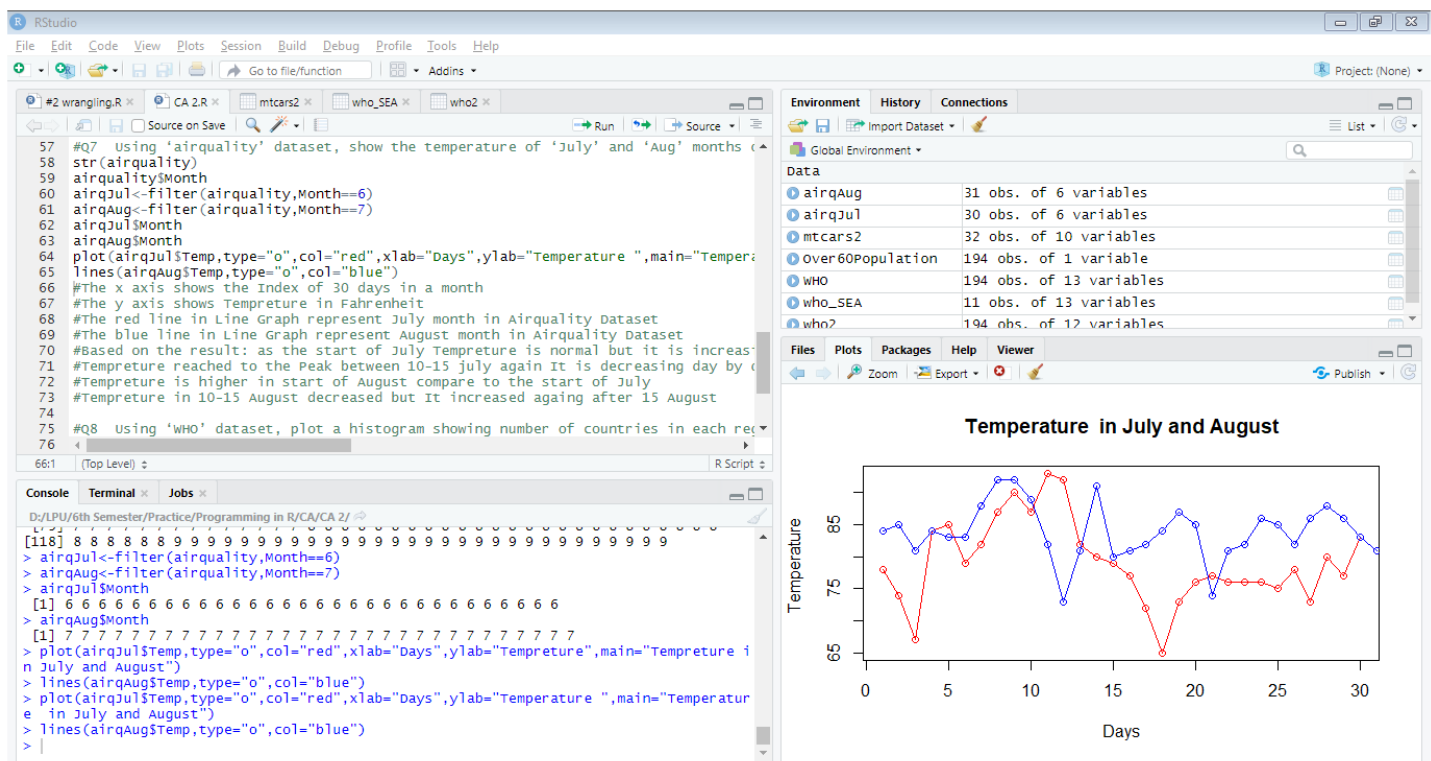
- The x axis shows the Miles per gallon (mpg) measure of how far a car can travel if you put just one gallon of petrol or diesel in its tank.
- The y axis shows the Weight of Cars
- Based on the result: as the weight of car getting higher, mileage per gallon getting lower
- On the Other Hand, as weight 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



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))
```

