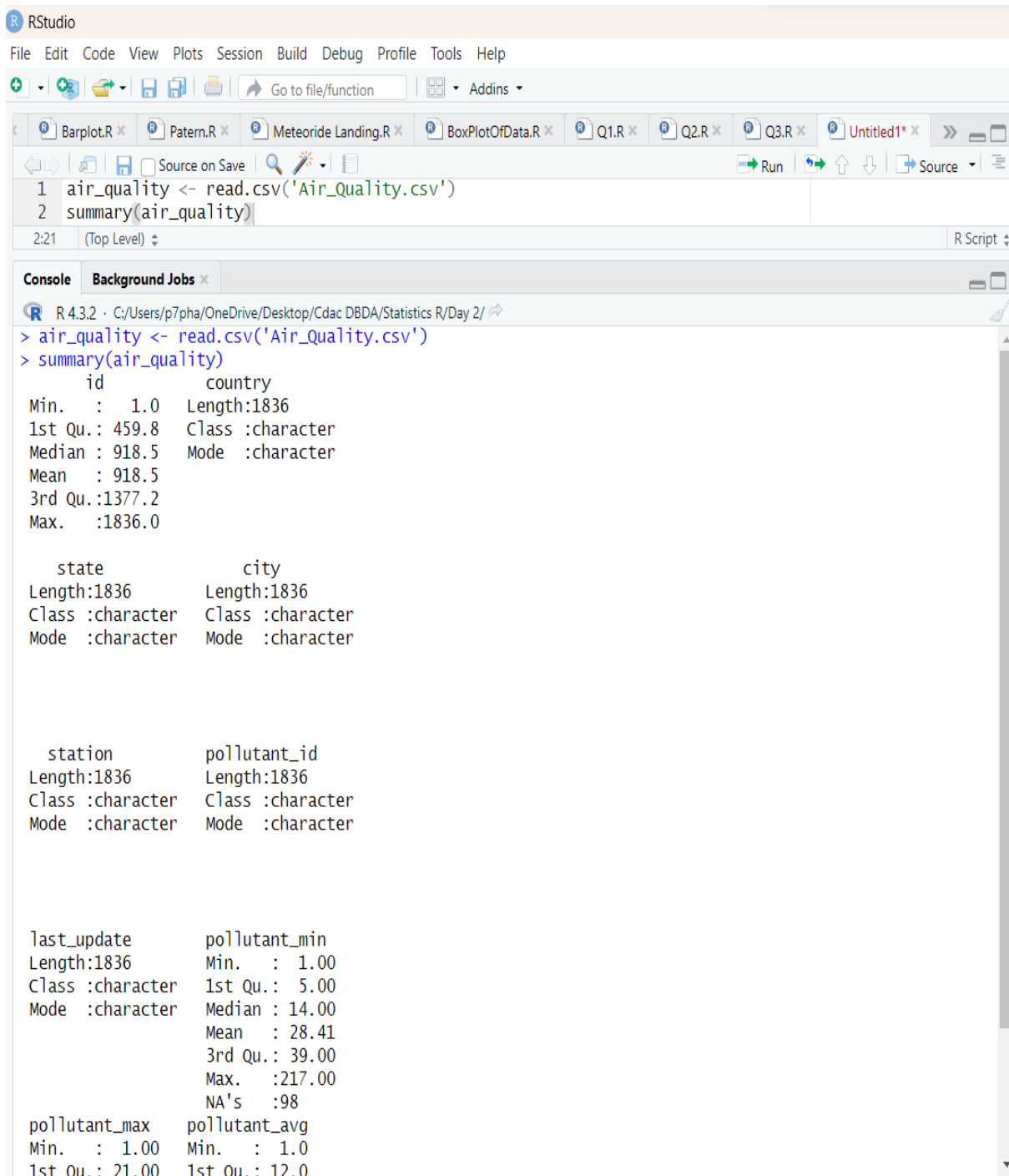


ADVANCED ANALYTICS USING STATISTICS– LAB 2

1. Load the dataset of Air_Quality.

Ans=



The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 air_quality <- read.csv('Air_Quality.csv')
2 summary(air_quality)
```

The console output shows the summary of the 'air_quality' dataset:

```
R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
> air_quality <- read.csv('Air_Quality.csv')
> summary(air_quality)
      id          country
Min.   : 1.0      Length:1836
1st Qu.: 459.8    Class  :character
Median : 918.5    Mode   :character
Mean    : 918.5
3rd Qu.:1377.2
Max.    :1836.0

      state        city
Length:1836      Length:1836
Class  :character Class  :character
Mode   :character Mode   :character

      station      pollutant_id
Length:1836      Length:1836
Class  :character Class  :character
Mode   :character Mode   :character

      last_update      pollutant_min
Length:1836          Min.   : 1.00
Class  :character    1st Qu.: 5.00
Mode   :character    Median : 14.00
                        Mean    : 28.41
                        3rd Qu.: 39.00
                        Max.    :217.00
                        NA's    :98

      pollutant_max      pollutant_avg
Min.   : 1.00          Min.   : 1.0
1st Qu.: 21.00         1st Qu.: 12.0
```

2. Find the measures of central tendency for each numerical column.

Ans= i)pollutant_min:-

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Barplot.R x Patern.R x Meteoride Landing.R x BoxPlotOfData.R x Q1.R x Q2.R x Q3.R x Untitled1* x
Source on Save Run
Breakpoints cannot be set until the file is saved.
1 air_quality <- read.csv('Air_Quality.csv')
2 summary(air_quality)
3
4 aq <- air_quality$pollutant_min
5 aq
6 mean_value<- mean(aq,na.rm = TRUE)
7 print(mean_value)
8 median_value <- median(aq,na.rm = TRUE)
9 print(median_value)
10
11 a<- table(aq)
12 names(a)[which(a==max(a))]
13
> mean_value<- mean(aq,na.rm = TRUE)
> print(mean_value)
[1] 28.41427
> median_value <- median(aq,na.rm = TRUE)
> print(median_value)
[1] 14
> a<- table(aq)
> names(a)[which(a==max(a))]
[1] "1"
```

ii)pollutant_max:-

```
14 aq1 <- air_quality$pollutant_max
15 aq1
16 mean_value1 <- mean(aq1, na.rm = TRUE)
17 print(mean_value1)
18 median_value1 <- median(aq1, na.rm = TRUE)
19 print(median_value1)
20 b <- table(aq1)
21 names(b)[which(b==max(b))]
[1] "6"
[1] 96.87342
[1] 63
Error in name(b) : could not find function "name"
> names(b)[which(b==max(b))]
[1] "6"
>
```

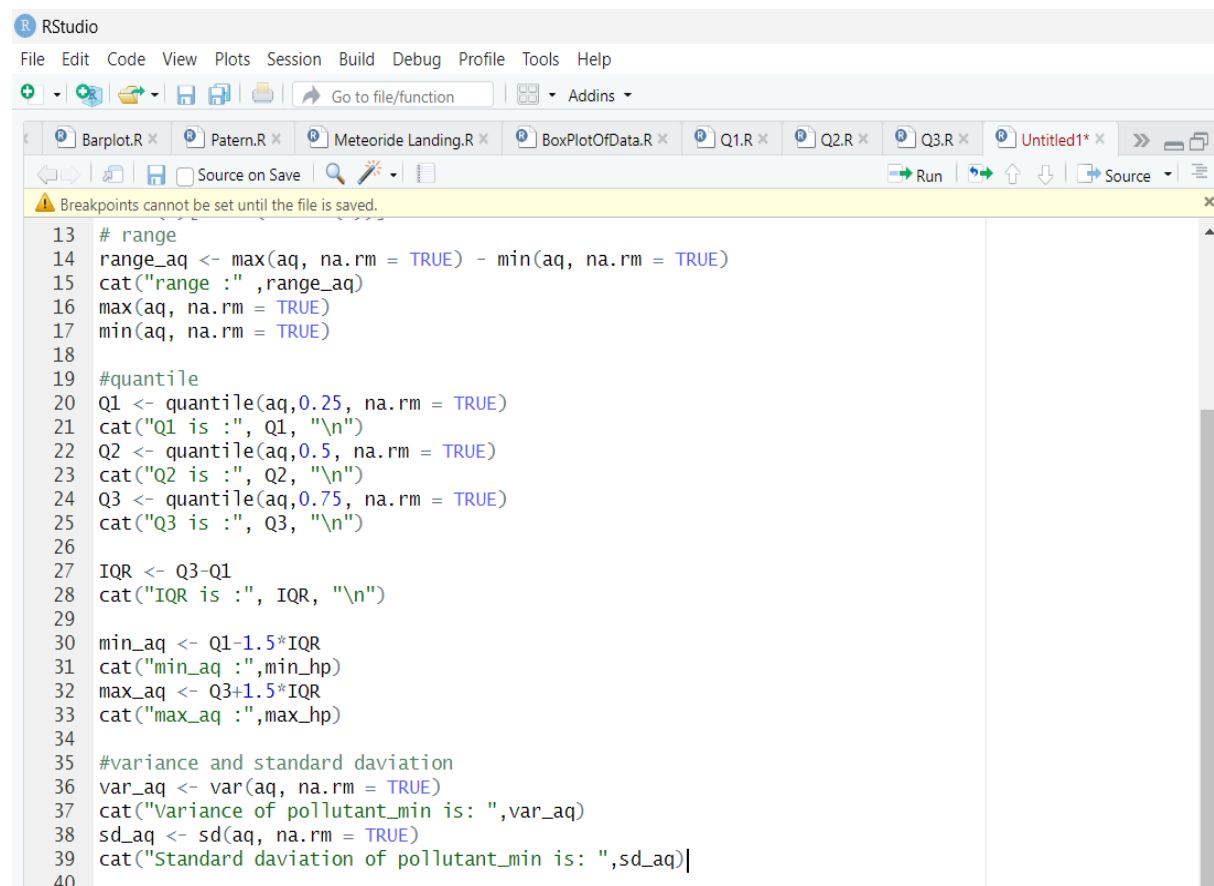
iii)pollutant_avg:-

```
23 aq2 <- air_quality$pollutant_avg
24 aq2
25 mean_value2 <- mean(aq2, na.rm = TRUE)
26 print(mean_value2)
27 median_value2 <- median(aq2, na.rm = TRUE)
28 print(median_value2)
29 c <- table(aq2)
30 names(c)[which(c==max(c))]
```

```
> mean_value2 <- mean(aq2, na.rm = TRUE)
> print(mean_value2)
[1] 54.10069
> median_value2 <- median(aq2, na.rm = TRUE)
> print(median_value2)
[1] 54.10069
> c <- table(aq2)
> names(c)[which(c==max(c))]
[1] "5"
> |
```

3. Find the measures of dispersion for each numerical column.

Ans= i)pollutant_min:-



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Barplot.R x Patern.R x Meteoride Landing.R x BoxPlotOfData.R x Q1.R x Q2.R x Q3.R x Untitled1* x
Source on Save Run Up Down Source
Breakpoints cannot be set until the file is saved.
13 # range
14 range_aq <- max(aq, na.rm = TRUE) - min(aq, na.rm = TRUE)
15 cat("range :",range_aq)
16 max(aq, na.rm = TRUE)
17 min(aq, na.rm = TRUE)
18
19 #quantile
20 Q1 <- quantile(aq,0.25, na.rm = TRUE)
21 cat("Q1 is :", Q1, "\n")
22 Q2 <- quantile(aq,0.5, na.rm = TRUE)
23 cat("Q2 is :", Q2, "\n")
24 Q3 <- quantile(aq,0.75, na.rm = TRUE)
25 cat("Q3 is :", Q3, "\n")
26
27 IQR <- Q3-Q1
28 cat("IQR is :", IQR, "\n")
29
30 min_aq <- Q1-1.5*IQR
31 cat("min_aq :",min_hp)
32 max_aq <- Q3+1.5*IQR
33 cat("max_aq :",max_hp)
34
35 #variance and standard daviation
36 var_aq <- var(aq, na.rm = TRUE)
37 cat("Variance of pollutant_min is: ",var_aq)
38 sd_aq <- sd(aq, na.rm = TRUE)
39 cat("Standard daviation of pollutant_min is: ",sd_aq)
40
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source

Console Background Jobs
R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
> range_aq <- max(aq, na.rm = TRUE) - min(aq, na.rm = TRUE)
> cat("range :", range_aq)
range : 216> max(aq, na.rm = TRUE)
[1] 217
> min(aq, na.rm = TRUE)
[1] 1
> Q1 <- quantile(aq, 0.25, na.rm = TRUE)
> cat("Q1 is :", Q1, "\n")
Q1 is : 5
> Q2 <- quantile(aq, 0.5, na.rm = TRUE)
> cat("Q2 is :", Q2, "\n")
Q2 is : 14
> Q3 <- quantile(aq, 0.75, na.rm = TRUE)
> cat("Q3 is :", Q3, "\n")
Q3 is : 39
> IQR <- Q3 - Q1
> cat("IQR is :", IQR, "\n")
IQR is : 34
> min_aq <- Q1 - 1.5 * IQR
> cat("min_aq :", min_hp)
min_aq : 305.25> max_aq <- Q3 + 1.5 * IQR
> cat("max_aq :", max_hp)
max_aq : 305.25
> var_aq <- var(aq, na.rm = TRUE)
> cat("Variance of pollutant_min is: ", var_aq)
Variance of pollutant_min is: 1183.622> sd_aq <- sd(aq, na.rm = TRUE)
> cat("Standard deviation of pollutant_min is: ", sd_aq)
Standard deviation of pollutant_min is: 34.40381
> |
```

ii) pollutant_max:-

```
50
51 # range
52 range_aq1 <- max(aq1, na.rm = TRUE) - min(aq1, na.rm = TRUE)
53 cat("range :", range_aq1)
54 max(aq1, na.rm = TRUE)
55 min(aq1, na.rm = TRUE)
56
57 #quantile
58 Q1 <- quantile(aq1, 0.25, na.rm = TRUE)
59 cat("Q1 is :", Q1, "\n")
60 Q2 <- quantile(aq1, 0.5, na.rm = TRUE)
61 cat("Q2 is :", Q2, "\n")
62 Q3 <- quantile(aq1, 0.75, na.rm = TRUE)
63 cat("Q3 is :", Q3, "\n")
64
65 IQR <- Q3 - Q1
66 cat("IQR is :", IQR, "\n")
67
68 min_aq1 <- Q1 - 1.5 * IQR
69 cat("min_aq1 :", min_hp)
70 max_aq1 <- Q3 + 1.5 * IQR
71 cat("max_aq1 :", max_hp)
72
73 #variance and standard deviation
74 var_aq1 <- var(aq1, na.rm = TRUE)
75 cat("Variance of pollutant_max is: ", var_aq1)
76 sd_aq1 <- sd(aq1, na.rm = TRUE)
77 cat("Standard deviation of pollutant_max is: ", sd_aq1)
78
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source

Console Background Jobs
R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
> range_aq1 <- max(aq1, na.rm = TRUE) - min(aq1, na.rm = TRUE)
> cat("range :", range_aq1)
range : 499> max(aq1, na.rm = TRUE)
[1] 500
> min(aq1, na.rm = TRUE)
[1] 1
> Q1 <- quantile(aq1, 0.25, na.rm = TRUE)
> cat("Q1 is :", Q1, "\n")
Q1 is : 21
> Q2 <- quantile(aq1, 0.5, na.rm = TRUE)
> cat("Q2 is :", Q2, "\n")
Q2 is : 63
> Q3 <- quantile(aq1, 0.75, na.rm = TRUE)
> cat("Q3 is :", Q3, "\n")
Q3 is : 124
> IQR <- Q3 - Q1
> cat("IQR is :", IQR, "\n")
IQR is : 103
> min_aq1 <- Q1 - 1.5 * IQR
> cat("min_aq1 :", min_hp)
min_aq1 : 305.25> max_aq1 <- Q3 + 1.5 * IQR
> cat("max_aq1 :", max_hp)
max_aq1 : 305.25
> var_aq1 <- var(aq1, na.rm = TRUE)
> cat("Variance of pollutant_max is: ", var_aq1)
Variance of pollutant_max is: 10975.72> sd_aq1 <- sd(aq1, na.rm = TRUE)
> cat("Standard deviation of pollutant_max is: ", sd_aq1)
Standard deviation of pollutant_max is: 104.7651
> |
```

iii) pollutant_avg:-

```
88
89 # range
90 range_aq2 <- max(aq2, na.rm = TRUE) - min(aq2, na.rm = TRUE)
91 cat("range :", range_aq2)
92 max(aq2, na.rm = TRUE)
93 min(aq2, na.rm = TRUE)
94
95 #quantile
96 Q1 <- quantile(aq2, 0.25, na.rm = TRUE)
97 cat("Q1 is :", Q1, "\n")
98 Q2 <- quantile(aq2, 0.5, na.rm = TRUE)
99 cat("Q2 is :", Q2, "\n")
100 Q3 <- quantile(aq2, 0.75, na.rm = TRUE)
101 cat("Q3 is :", Q3, "\n")
102
103 IQR <- Q3 - Q1
104 cat("IQR is :", IQR, "\n")
105
106 min_aq2 <- Q1 - 1.5 * IQR
107 cat("min_aq2 :", min_hp)
108 max_aq2 <- Q3 + 1.5 * IQR
109 cat("max_aq2 :", max_hp)
110
111 #variance and standard daviation
112 var_aq2 <- var(aq2, na.rm = TRUE)
113 cat("Variance of pollutant_avg is: ", var_aq2)
114 sd_aq2 <- sd(aq2, na.rm = TRUE)
115 cat("Standard daviation of pollutant_avg is: ", sd_aq2)

115:55 (Top Level) R Script
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

Console Background Jobs

```
R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
> range_aq2 <- max(aq2, na.rm = TRUE) - min(aq2, na.rm = TRUE)
> cat("range :", range_aq2)
range : 313> max(aq2, na.rm = TRUE)
[1] 314
> min(aq2, na.rm = TRUE)
[1] 1
> Q1 <- quantile(aq2, 0.25, na.rm = TRUE)
> cat("Q1 is :", Q1, "\n")
Q1 is : 12
> Q2 <- quantile(aq2, 0.5, na.rm = TRUE)
> cat("Q2 is :", Q2, "\n")
Q2 is : 31
> Q3 <- quantile(aq2, 0.75, na.rm = TRUE)
> cat("Q3 is :", Q3, "\n")
Q3 is : 70
> IQR <- Q3-Q1
> cat("IQR is :", IQR, "\n")
IQR is : 58
> min_aq2 <- Q1-1.5*IQR
> cat("min_aq2 :", min_aq2)
min_aq2 : 305.25> max_aq2 <- Q3+1.5*IQR
> cat("max_aq2 :", max_aq2)
max_aq2 : 305.25
> var_aq2 <- var(aq2, na.rm = TRUE)
> cat("Variance of pollutant_avg is: ", var_aq2)
Variance of pollutant_avg is: 3699.578> sd_aq2 <- sd(aq2, na.rm = TRUE)
> cat("Standard deviation of pollutant_avg is: ", sd_aq2)
Standard deviation of pollutant_avg is: 60.82416
> |
```

4. Find the mode of city column.

Ans=

```
110
117 #mode of city column
118 aq3 <- air_quality$city
119 aq3
120 d <- table(aq3)
121 names(d)[which(d==max(d))]
```

121:27 (Top Level) R Script

Console Background Jobs

R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/

[949]	"Jabalpur"	"Jabalpur"	"Katni"	"Katni"
[953]	"Katni"	"Katni"	"Katni"	"Katni"
[957]	"Katni"	"Mandideep"	"Mandideep"	"Mandideep"
[961]	"Mandideep"	"Mandideep"	"Mandideep"	"Mandideep"
[965]	"Pithampur"	"Pithampur"	"Pithampur"	"Pithampur"
[969]	"Pithampur"	"Pithampur"	"Pithampur"	"Ratlam"
[973]	"Ratlam"	"Ratlam"	"Ratlam"	"Sagar"
[977]	"Sagar"	"Sagar"	"Sagar"	"Sagar"
[981]	"Satna"	"Satna"	"Satna"	"Satna"
[985]	"Singrauli"	"Singrauli"	"Singrauli"	"Singrauli"
[989]	"Singrauli"	"Singrauli"	"Singrauli"	"Ujjain"
[993]	"Ujjain"	"Ujjain"	"Ujjain"	"Ujjain"
[997]	"Ujjain"	"Ujjain"	"Aurangabad"	"Aurangabad"

[reached getOption("max.print") -- omitted 836 entries]

```
> d <- table(aq3)
> names(d)[which(d==max(d))]
[1] "Delhi"
> |
```

5. Find if there are outliers in the numerical columns using boxplot and if there is, display the value.

Ans= i)pollutant_min:-

```

34
35 #boxplot and outliers
36 boxplot(aq, na.rm = TRUE)
37 outliers <- aq[aq<min_aq | aq > max_aq]
38 cat("outliers of pollutant_min:", outliers)

```

37:40 (Top Level) R Script

Console Background Jobs

R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/

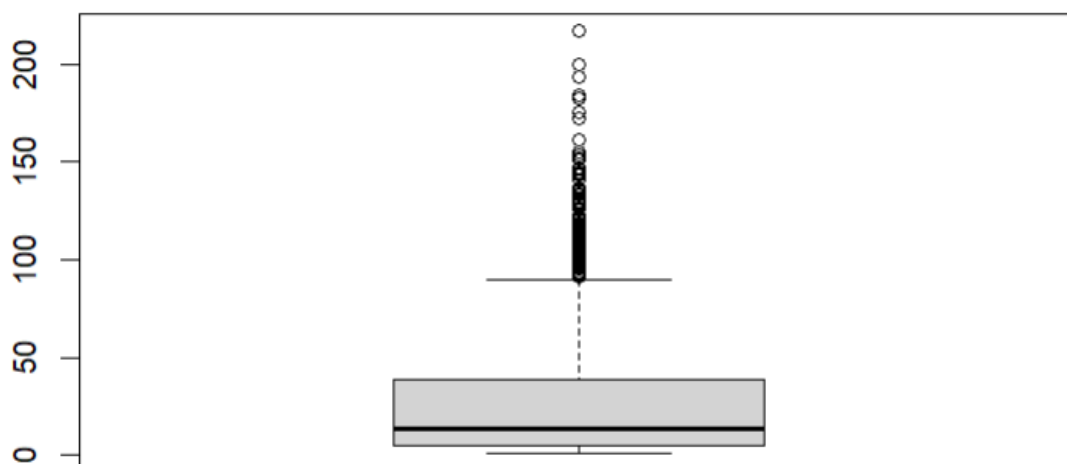
```

> boxplot(aq, na.rm = TRUE)
> outliers <- aq[aq<min_aq | aq > max_aq]
> cat("outliers of pollutant_min:", outliers)
outliers of pollutant_min: 95 105 NA 113 NA NA 129 111 133 104 102 111 113 131 175 120 133 NA NA NA 127 106
137 108 144 193 136 101 127 123 106 135 120 109 95 NA 110 120 119 161 119 93 121 97 NA 133 NA 172 NA NA 104
126 106 117 113 105 117 96 91 113 99 91 217 100 103 117 129 147 200 144 NA NA NA 102 111 123 146 131 109 NA
126 135 155 120 95 155 NA 184 NA NA NA NA NA NA NA 94 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA 99 91 NA NA 107 91 126 93 93 99 NA NA NA NA 129 NA NA NA NA NA NA NA 107 109 NA NA NA
NA NA NA NA NA NA NA 126 NA NA NA NA NA 104 151 122 110 133 117 103 141 NA 118 127 145 102 152 NA 147 141 1
08 92 107 110 102 109 121 112 128 117 115 146 105 136 153 143 182 130 93 123 136 131 152 134 110 NA NA NA N
A NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
> |

```

Files Plots Packages Help Viewer Presentation

Zoom Export Publish



ii)pollutant_max:-

```
78 #boxplot and outliers
79 boxplot(aq1, na.rm = TRUE)
80 outliers1 <- aq1[aq1<min_aq1 | aq1 > max_aq1]
81 cat("outliers of pollutant_mmax:", outliers1)
```

81:46

(Top Level) ↕

R Script ↕

Console

Background Jobs x

R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/ ↗

```
> boxplot(aq1, na.rm = TRUE)
> outliers1 <- aq1[aq1<min_aq1 | aq1 > max_aq1]
> cat("outliers of pollutant_mmax:", outliers1)
outliers of pollutant_mmax: 312 282 474 295 315 403 NA 344 NA NA 281 348 423 321 408 336 303 281 333 500 36
7 304 NA NA NA 315 304 308 348 362 489 396 458 345 500 500 308 312 341 321 348 324 NA 470 318 422 306 473 3
09 342 323 413 NA NA 419 339 NA 347 436 NA 318 368 314 330 313 288 404 306 377 304 403 500 283 422 408 320
304 406 454 361 442 316 500 500 314 301 NA NA NA 309 344 329 NA 350 305 327 326 500 NA 311 298 341 330 467
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA 282 NA NA 335 417 374 NA NA NA NA NA 302 NA 320
NA NA NA NA 310 NA NA 347 NA NA NA NA 301 NA NA NA NA NA NA NA 286 341 320 371 408 327 360 416 328 332 NA N
A NA 291 281 NA NA 321 NA NA NA NA NA NA NA NA NA NA 346 302 386 464 459 477 314 347 304 310 316 353 346 39
3 487 NA 433 422 363 466 335 406 NA 500 308 295 361 303 378 384 317 287 288 340 281 330 285 362 392 352 469
495 324 352 313 368 322 377 315 358 301 339 450 NA NA NA NA NA NA NA 300 284 NA NA NA NA NA NA NA NA NA
NA NA NA 372
> |
```

Files

Plots

Packages

Help

Viewer

Presentation



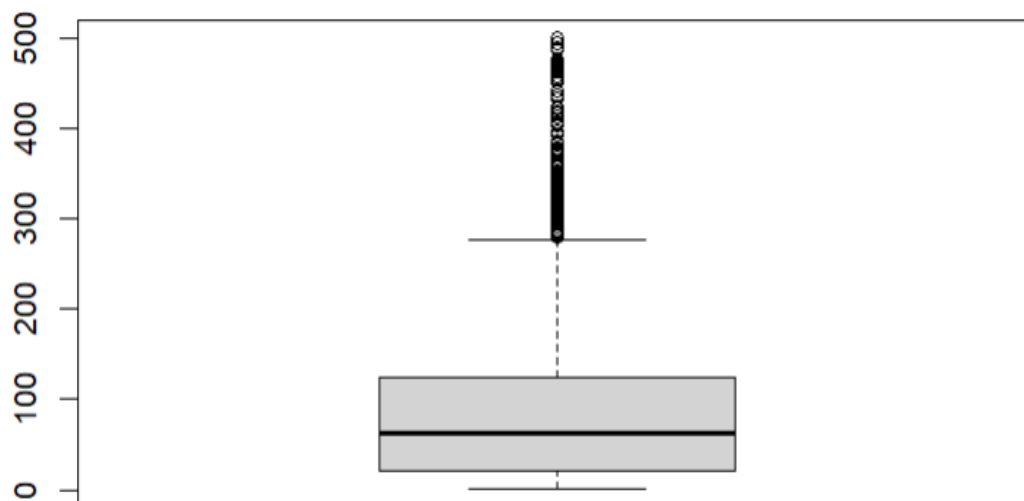
Zoom



Export ▾



Publish ▾



iii) pollutant_avg:-

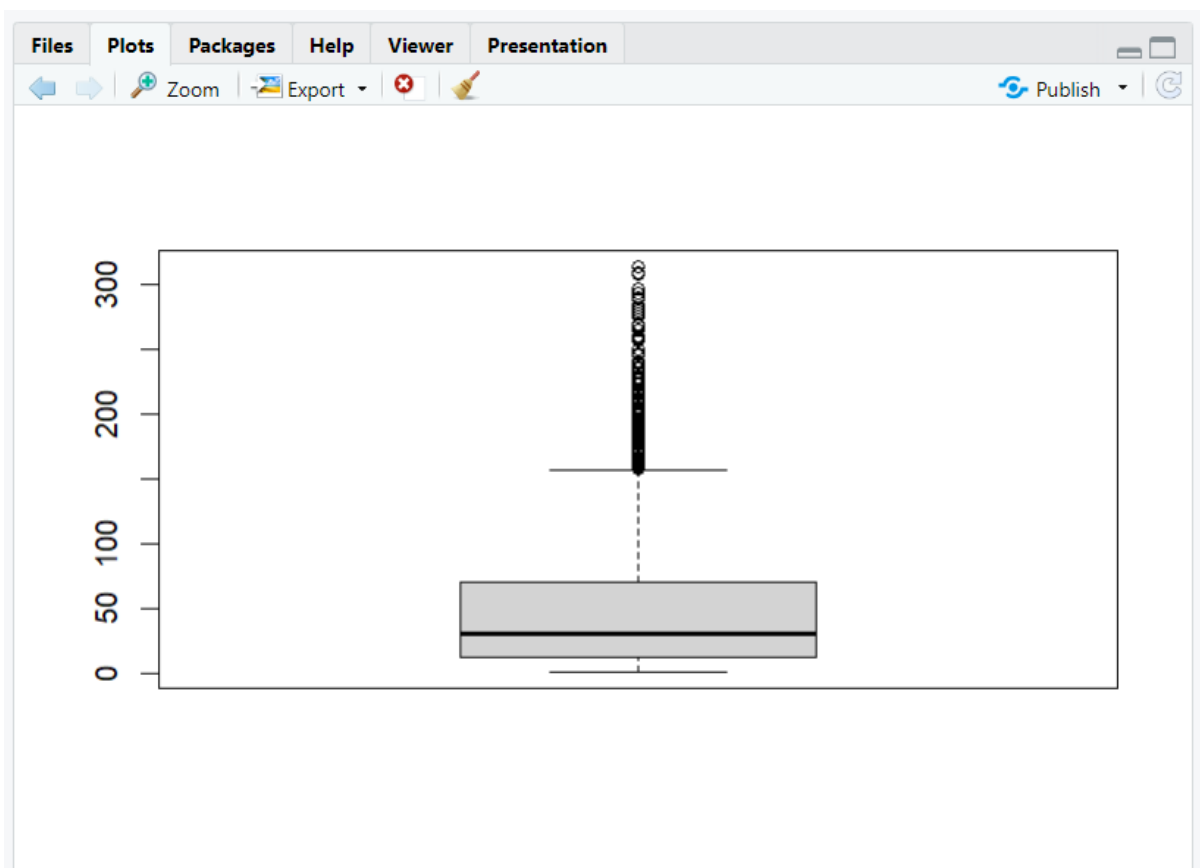
```
120
121 #boxplot and outliers
122 boxplot(aq2, na.rm = TRUE)
123 outliers2 <- aq2[aq2<min_aq2 | aq2 > max_aq2]
124 cat("outliers of pollutant_mmax:", outliers2)
125
```

124:46 (Top Level) R Script

Console Background Jobs

R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/

```
> boxplot(aq2, na.rm = TRUE)
> outliers2 <- aq2[aq2<min_aq2 | aq2 > max_aq2]
> cat("outliers of pollutant_mmax:", outliers2)
outliers of pollutant_mmax: 202 200 223 199 NA 235 NA NA 190 240 205 177 220 199 164 174 173 208 293 280 15
9 NA NA NA 173 160 170 184 185 212 268 297 213 216 178 190 183 165 164 196 182 170 201 165 201 NA 194 195 2
20 275 165 191 205 240 NA 178 NA 260 179 NA 184 NA 250 189 195 179 204 228 309 240 282 188 201 314 250 210
182 174 162 246 265 206 NA NA NA 210 159 215 166 182 NA 277 205 247 167 240 NA 212 257 226 167 NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA 161 178 161 NA NA NA NA NA NA NA NA NA NA 161 NA NA 1
87 NA NA NA NA 173 NA NA NA NA NA NA NA 162 206 241 164 163 223 219 192 162 NA NA NA NA NA 166 NA NA NA NA
NA NA NA NA NA NA 180 199 247 240 224 189 165 179 204 184 218 165 181 173 195 232 NA 158 270 250 259 290 21
5 239 NA 229 178 202 236 194 229 191 240 169 179 250 204 258 223 206 280 194 308 229 284 289 176 210 158 20
2 176 215 170 194 208 219 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
> |
```



6. Check if there are missing values in the columns. Replace the missing values.

Ans= i)pollutant_min:-

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Barplot.R x Patern.R x Meteorite Landing.R x BoxPlotOfData.R x Q1.R x Q2.R x Q3.R x Air_Quality.R* x OddInteger.R* x Median.R x Mode.R* x CommuteTimes.F x
Source on Save Run
14 #replace
15 air_quality$pollutant_min= ifelse(is.na(air_quality$pollutant_min),median(air_quality$pollutant_min,na.rm = TRUE),air_quality$pollutant_min)
16 print(air_quality$pollutant_min)
17
14:9 (Top Level) z R Script
Console Background Jobs x
R 4.3.2 C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
[32] 82 1 13 2 6 5 1 37 10 18 52 5 13 55 10 18 30 15 1 1 38 1 35 35 9 1 2 36 3 34 43
[63] 10 3 1 13 2 10 2 4 32 9 35 76 1 2 4 25 37 32 32 10 2 8 16 17 21 57 12 4 5 57 37
[94] 51 50 14 3 1 22 5 61 41 4 4 2 19 5 87 95 26 4 7 34 11 51 78 6 8 2 36 3 31 43 15
[125] 9 7 88 105 8 14 6 30 2 113 14 75 3 15 14 4 85 129 10 3 12 10 8 60 111 16 2 15 83 133 3
[156] 5 4 18 3 54 104 21 102 15 43 111 4 7 6 75 1 86 113 28 10 6 29 20 70 131 74 3 5 22 4 67
[187] 175 11 1 1 5 1 14 120 49 90 9 66 133 65 7 5 22 9 68 22 7 6 34 10 14 14 14 22 3 73 127
[218] 31 4 1 5 8 82 71 30 10 3 58 15 70 106 10 18 15 73 137 18 7 4 20 2 35 108 60 3 4 38 7
[249] 87 144 26 7 27 23 15 193 136 21 6 1 18 20 55 101 8 3 1 10 1 80 127 24 7 8 55 6 63 123 38
[280] 6 4 38 3 63 106 6 24 31 60 135 62 6 4 5 3 77 120 20 5 6 5 4 85 109 36 3 19 13 11 95
[311] 14 9 10 1 8 2 82 110 27 51 18 23 120 23 2 1 5 10 87 119 6 7 2 42 3 55 161 34 4 2 23
[342] 6 70 119 1 6 11 36 21 93 121 12 4 4 20 3 67 97 14 2 4 17 5 14 133 22 6 1 52 7 14 172
[373] 24 6 1 5 4 76 86 14 3 1 8 40 82 20 46 3 1 63 78 4 32 54 57 4 2 11 14 7 41 51 11
[404] 6 6 10 6 71 79 6 2 3 13 12 104 126 4 2 30 54 42 30 32 9 2 4 58 1 73 83 11 4 1 2
[435] 4 44 57 9 2 7 8 8 2 32 29 2 1 50 89 106 2 21 57 38 52 74 3 7 22 9 70 117 1 3 2 39
[466] 1 113 105 29 17 8 14 3 53 117 32 5 3 33 4 69 96 21 6 15 14 52 91 113 13 2 4 23 1 99 74
[497] 34 6 7 2 1 38 91 4 2 6 20 1 217 100 1 6 1 1 16 9 88 103 7 4 7 51 1 60 60 19 22 19
[528] 74 117 11 2 12 16 19 81 129 14 35 5 3 147 6 6 1 2 1 200 144 10 7 2 29 18 14 14 30 2
[559] 70 102 59 15 1 14 19 111 123 22 9 4 36 8 146 131 19 5 2 4 35 109 14 31 7 11 34 6 126 135 28
[590] 4 9 24 29 77 57 16 11 5 21 18 155 120 2 1 4 11 13 32 31 40 6 5 6 1 49 58 18 6 1 5
[621] 2 95 20 1 1 40 85 155 11 2 12 14 38 85 34 2 44 29 54 88 36 12 9 15 1 67 184 16 8 3 14
[652] 29 40 37 11 6 1 18 6 4 27 14 14 17 1 14 14 19 42 39 10 1 6 16 24 28 58 44 5 1 7 1
[683] 42 46 13 10 4 22 2 40 23 10 50 36 58 16 1 7 18 19 30 38 14 3 4 18 13 20 29 2 2 3 6
[714] 15 38 50 1 2 45 7 18 22 4 19 18 34 20 2 5 16 31 17 39 28 4 2 14 12 23 27 25 1 2 28
[745] 9 14 23 21 4 1 16 16 37 31 21 1 6 23 11 15 10 8 1 19 10 8 53 68 6 2 4 28 9 70 4
[776] 1 16 42 17 18 5 1 24 42 1 16 22 5 5 15 42 19 7 14 10 8 37 57 8 3 14 2 14 3 20 52
[807] 11 30 8 1 1 2 13 49 69 14 2 7 8 9 17 28 15 1 17 9 20 21 32 9 2 2 17 7 1 8 1
[838] 4 94 34 30 48 13 1 5 37 3 38 65 7 1 11 1 13 14 14 14 14 14 44 27 14 14 14 14 43 4
[869] 31 16 2 1 26 23 17 24 45 4 14 21 1 14 14 18 1 85 6 26 36 6 1 5 2 6 22 38 11 1 1
[900] 15 38 2 10 10 3 4 14 13 39 78 22 2 9 29 27 17 13 17 14 36 51 14 3 4 12 5 37 76 16 2
[931] 14 16 10 14 14 4 31 57 39 3 10 19 8 42 73 24 1 5 16 15 17 70 23 3 5 24 9 14 14 14 14
[962] 14 12 11 41 65 11 5 17 25 52 11 46 11 5 11 14 20 8 1 29 47 25 8 29 50 27 8 8 28 6 22
[993] 61 8 2 4 15 8 9 56
[reached getoption("max.print")] -- omitted 836 entries ]
>
```

ii)pollutant_max:-

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Barplot.R x Patern.R x Meteorite Landing.R x BoxPlotOfData.R x Q1.R x Q2.R x Q3.R x Air_Quality.R* x OddInteger.R* x Median.R x Mode.R* x CommuteTimes.F x
Source on Save Run
61 #replace
62 air_quality$pollutant_max= ifelse(is.na(air_quality$pollutant_max),median(air_quality$pollutant_max,na.rm = TRUE),air_quality$pollutant_max)
63 print(air_quality$pollutant_max)
64
64:1 (Top Level) z R Script
Console Background Jobs x
R 4.3.2 C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
> air_quality$pollutant_max= ifelse(is.na(air_quality$pollutant_max),median(air_quality$pollutant_max,na.rm = TRUE),air_quality$pollutant_max)
> print(air_quality$pollutant_max)
[1] 109 138 42 5 42 45 82 111 120 44 6 9 52 34 131 120 88 3 10 68 49 312 276 187 9 15 154 46 31 13 4 101 28 24 38 32
[37] 5 1 40 19 21 282 7 17 67 11 70 60 42 4 6 54 49 131 85 11 3 11 80 59 95 123 18 4 106 58 103 474 15 12 72 15
[73] 76 148 295 54 5 100 42 97 108 38 3 24 56 49 64 136 69 32 10 112 207 144 127 61 4 6 81 61 166 114 16 28 13 36 68 260
[109] 173 134 7 35 104 219 86 125 86 23 11 86 67 162 121 20 15 75 315 403 109 63 15 83 69 344 63 171 9 33 63 47 233 281 51 5
[145] 38 130 147 274 206 45 122 25 348 423 71 6 16 95 62 321 237 33 161 36 247 408 84 8 22 123 32 336 261 138 35 8 122 86 303 281
[181] 141 6 11 119 168 33 500 69 13 8 124 16 96 220 110 108 21 367 180 125 29 21 110 57 304 64 9 19 178 39 63 63 60 21 315
[217] 213 120 6 23 122 9 223 137 107 23 3 70 22 304 192 28 75 154 308 274 72 11 8 125 80 348 216 82 3 38 105 73 362 489 155 13
[253] 60 103 97 396 458 81 8 28 123 108 273 159 39 9 6 48 91 345 235 108 12 58 70 121 500 500 166 8 19 148 162 308 194 7 118 99
[289] 312 341 143 8 17 128 110 321 274 127 8 25 143 163 348 244 114 5 40 104 97 324 63 135 18 11 126 52 470 210 166 104 72 147 221 114
[325] 5 36 105 103 318 422 71 10 21 113 77 306 473 102 6 16 141 43 309 342 10 8 18 134 100 323 413 116 8 17 103 130 217 185 57 4
[361] 8 63 128 63 250 47 10 3 125 56 63 419 68 9 7 112 69 339 242 63 14 55 14 347 436 259 76 133 117 174 137 9 72 248 138 6
[397] 2 12 63 13 239 154 12 6 11 105 37 318 247 73 10 3 92 36 368 220 8 25 95 118 45 204 148 10 2 8 100 59 314 145 12 4
[433] 3 36 17 88 130 30 5 12 69 38 187 114 2 2 100 330 313 2 21 71 288 248 404 13 32 78 83 306 377 14 9 74 84 54 304 180
[469] 33 19 23 68 205 403 500 81 9 27 148 56 195 283 76 26 36 102 96 422 408 33 10 6 68 12 320 153 36 6 25 73 107 304 168 64
[505] 5 7 104 75 406 454 4 1 4 98 13 361 442 8 4 11 120 90 316 182 39 117 238 209 167 82 5 23 68 124 236 179 43 74 123 500
[541] 500 6 6 7 129 140 314 301 11 11 5 55 30 63 63 63 108 43 266 185 72 19 16 104 21 309 227 41 12 86 60 71 344 250 111 11
[577] 53 86 121 329 63 32 7 61 117 35 350 305 77 7 36 93 171 218 116 22 14 20 84 23 327 203 12 3 10 101 15 114 102 44 6 6
[613] 22 8 51 62 20 7 9 46 66 219 21 4 6 59 326 500 16 2 28 63 40 311 83 34 106 87 298 251 40 13 33 73 49 164 341 19
[649] 9 7 33 31 330 236 34 14 15 57 87 53 467 63 63 74 59 63 63 90 86 74 11 2 31 26 58 34 116 46 5 43 12 61 198 144
[685] 44 21 6 101 7 105 24 15 124 108 123 58 2 9 33 29 66 111 50 4 8 58 81 97 120 79 2 7 53 75 234 78 1 4 100 34
[721] 22 23 31 48 87 143 66 9 6 49 50 78 62 31 4 8 28 43 83 100 38 1 5 31 43 64 55 28 6 5 36 20 83 108 44 1
[757] 8 34 20 92 122 84 8 26 106 40 98 131 68 4 5 58 67 159 97 1 60 84 73 125 50 4 168 70 49 52 64 22 8 25 54 54
[793] 65 63 34 17 39 68 34 32 63 6 63 15 30 95 53 71 60 10 19 84 72 118 218 45 2 11 49 20 78 103 55 2 26 29 91 38
[829] 46 23 6 7 61 44 24 57 2 52 114 148 70 51 14 2 7 113 43 77 102 20 2 121 44 27 63 63 63 63 53 31 63 63 63
[865] 63 63 45 54 64 127 29 2 3 60 47 55 53 54 5 63 72 1 63 63 101 6 92 41 78 111 36 1 9 47 7 77 119 32 6 6 49
[901] 42 80 104 56 3 15 88 34 132 173 106 3 26 111 124 58 41 19 15 282 189 87 6 82 108 48 228 184 108 4 19 112 97 63 63 111
[937] 335 417 176 8 33 168 114 220 214 113 4 12 108 57 192 374 81 5 109 122 87 63 63 63 63 63 82 44 302 133 38 6 46 53 102 155
[973] 208 38 9 67 63 21 13 28 53 90 28 11 268 320 40 14 26 104 116 150 216 93 6 11 100 107 79 107
[reached getoption("max.print")] -- omitted 836 entries ]
>
```

iii) pollutant_avg:-

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Barplot.R Patern.R Meteorite Landing.R BoxPlotOfData.R Q1.R Q2.R Q3.R Air_Quality.R OddInteger.R Median.R Mode.R CommuteTimes.R
Source on Save Run Source
108 #replace
109 air_quality$pollutant_max= ifelse(is.na(air_quality$pollutant_max),median(air_quality$pollutant_max,na.rm = TRUE),air_quality$pollutant_max)
110 print(air_quality$pollutant_max)
110:33 (Top Level) R Script

Console Background Jobs
R 4.3.2 - C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
> air_quality$pollutant_max= ifelse(is.na(air_quality$pollutant_max),median(air_quality$pollutant_max,na.rm = TRUE),air_quality$pollutant_max)
> print(air_quality$pollutant_max)
[1] 109 138 42 5 42 45 82 111 120 44 6 9 52 34 131 120 88 3 10 68 49 312 276 187 9 15 154 46 31 13 4 101 28 24 38 32
[37] 5 1 40 19 21 282 7 17 67 11 70 60 42 4 6 54 49 131 85 11 3 11 80 59 95 123 18 4 106 58 103 474 15 12 72 15
[73] 76 148 295 54 5 100 42 97 108 38 3 24 56 49 64 136 69 32 10 112 207 144 127 61 4 6 81 61 166 114 16 28 13 36 68 260
[109] 173 134 7 35 104 219 86 125 86 23 11 86 67 162 121 20 15 75 315 403 109 63 15 83 69 344 63 171 9 33 63 47 233 281 51 5
[145] 38 130 147 274 206 45 122 25 348 423 71 6 16 95 62 321 237 33 161 36 247 408 84 8 22 123 32 336 261 138 35 8 122 86 303 281
[181] 141 6 11 119 168 333 500 69 13 8 124 16 96 220 110 108 21 367 180 125 29 21 110 57 304 64 9 19 178 39 63 63 63 60 21 315
[217] 213 120 6 23 122 9 223 137 107 23 3 70 22 304 192 28 75 154 308 274 72 11 8 125 80 348 216 82 3 38 105 73 362 489 155 13
[253] 60 103 97 396 458 81 8 28 123 108 273 159 39 9 6 48 91 345 235 108 12 58 70 121 500 500 166 8 19 148 162 308 194 7 118 99
[289] 312 341 143 8 17 128 110 321 274 127 8 25 143 163 348 244 114 5 40 104 97 324 63 135 18 11 126 52 470 210 166 104 72 147 221 114
[325] 5 36 105 103 318 422 71 10 21 113 77 306 473 102 6 16 141 43 309 342 10 8 18 134 100 323 413 116 8 17 103 130 217 185 57 4
[361] 8 63 128 63 250 47 10 3 125 56 63 419 68 9 7 112 69 339 242 63 14 55 14 347 436 259 76 133 117 174 137 9 72 248 138 6
[397] 2 12 63 13 239 154 12 6 11 105 37 318 247 73 10 3 92 36 368 220 8 25 95 118 45 204 148 10 2 8 100 59 314 145 12 4
[433] 3 36 17 88 130 30 5 12 69 38 187 114 2 2 100 330 313 2 21 71 288 248 404 13 32 78 83 306 377 14 9 74 84 54 304 180
[469] 33 19 23 68 205 403 500 81 9 27 148 56 195 283 76 26 36 102 96 422 408 33 10 6 68 12 320 153 36 6 25 73 107 304 168 64
[505] 5 7 104 75 406 454 4 1 4 98 13 361 442 8 4 11 120 90 316 182 39 117 238 209 167 82 5 23 68 124 236 179 43 74 123 500
[541] 500 6 6 7 129 140 314 301 11 11 5 55 30 63 63 63 108 43 266 185 72 19 16 104 21 309 227 41 12 86 60 71 344 250 111 11
[577] 53 86 121 329 63 32 7 61 117 35 350 305 77 7 36 93 171 218 116 22 14 20 84 23 327 203 12 3 10 101 15 114 102 44 6 6
[613] 22 8 51 62 20 7 9 46 66 219 21 4 6 59 326 500 16 2 28 63 40 311 83 34 106 87 298 251 40 13 33 73 49 164 341 19
[649] 9 7 33 31 330 236 34 14 15 57 87 53 467 63 63 74 59 63 63 90 86 74 11 2 31 26 58 34 116 46 5 43 12 61 198 144
[685] 44 21 6 101 7 105 24 15 124 108 123 58 2 9 33 29 66 111 50 4 8 58 81 97 120 79 2 7 53 75 234 78 1 4 100 34
[721] 22 23 31 48 87 143 66 9 6 49 50 78 62 31 4 8 28 43 83 100 38 1 5 31 43 64 55 28 6 5 36 20 83 108 44 1
[757] 8 34 20 92 122 84 8 26 106 40 98 131 68 4 5 58 67 159 97 1 60 84 73 125 50 4 168 70 49 52 64 22 8 25 54 54
[793] 65 63 34 17 39 68 34 32 63 6 63 15 30 95 53 71 60 10 19 84 72 118 218 45 2 11 49 20 78 103 55 2 26 29 91 38
[829] 46 23 6 7 61 44 24 57 2 52 114 148 70 51 14 2 7 113 43 77 102 20 2 121 44 27 63 63 63 63 53 31 63 63 63
[865] 63 63 45 64 127 29 2 3 60 47 55 53 54 5 63 72 1 63 63 101 6 92 41 78 111 36 1 9 47 7 77 119 32 6 6 49
[901] 42 80 104 56 3 15 88 34 132 173 106 3 26 111 124 58 41 19 15 282 189 87 6 82 108 48 228 184 108 4 19 112 97 63 63 111
[937] 335 417 176 8 33 168 114 220 214 113 4 12 108 57 192 374 81 5 109 122 87 63 63 63 63 63 82 44 302 133 38 6 46 53 102 155
[973] 208 38 9 67 63 21 13 28 53 90 28 11 268 320 40 14 26 104 116 150 216 93 6 11 100 107 79 107
[ reached getOption("max.print") -- omitted 836 entries ]
>
```

7. Find the type of skewness and kurtosis of pollutant_avg.

Ans=

```
148 skewness_pollutant_avg <- skewness(aq2, na.rm = TRUE)
149 cat("Skewness of pollutant_avg:", skewness_pollutant_avg, "\n")
150 kurtosis_pollutant_avg <- kurtosis(aq2, na.rm = TRUE)
151 cat("Kurtosis of pollutant_avg:", kurtosis_pollutant_avg, "\n")
152
```

151:64 (Top Level) ↕

Console Background Jobs x

R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/ ↗

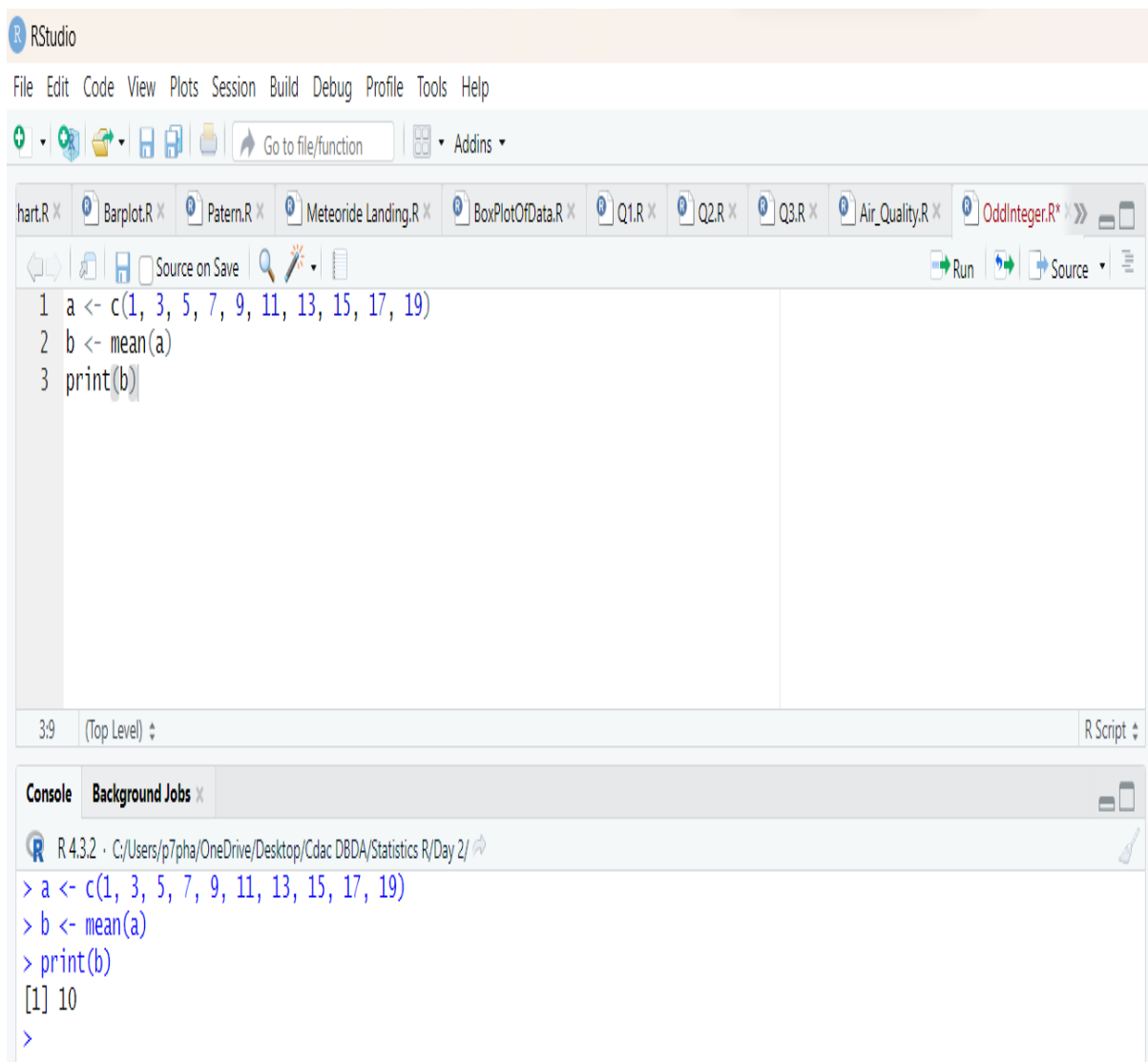
```
> skewness_pollutant_avg <- skewness(aq2, na.rm = TRUE)
> cat("Skewness of pollutant_avg:", skewness_pollutant_avg, "\n")
Skewness of pollutant_avg: 1.723487
> kurtosis_pollutant_avg <- kurtosis(aq2, na.rm = TRUE)
> cat("Kurtosis of pollutant_avg:", kurtosis_pollutant_avg, "\n")
Kurtosis of pollutant_avg: 5.529456
> |
```

Ans=

[illegible]

9. Find the mean of the first 10 odd integers.

Ans=



The screenshot shows the RStudio interface. The script editor contains the following R code:

```
1 a <- c(1, 3, 5, 7, 9, 11, 13, 15, 17, 19)
2 b <- mean(a)
3 print(b)
```

The console at the bottom shows the execution of this code:

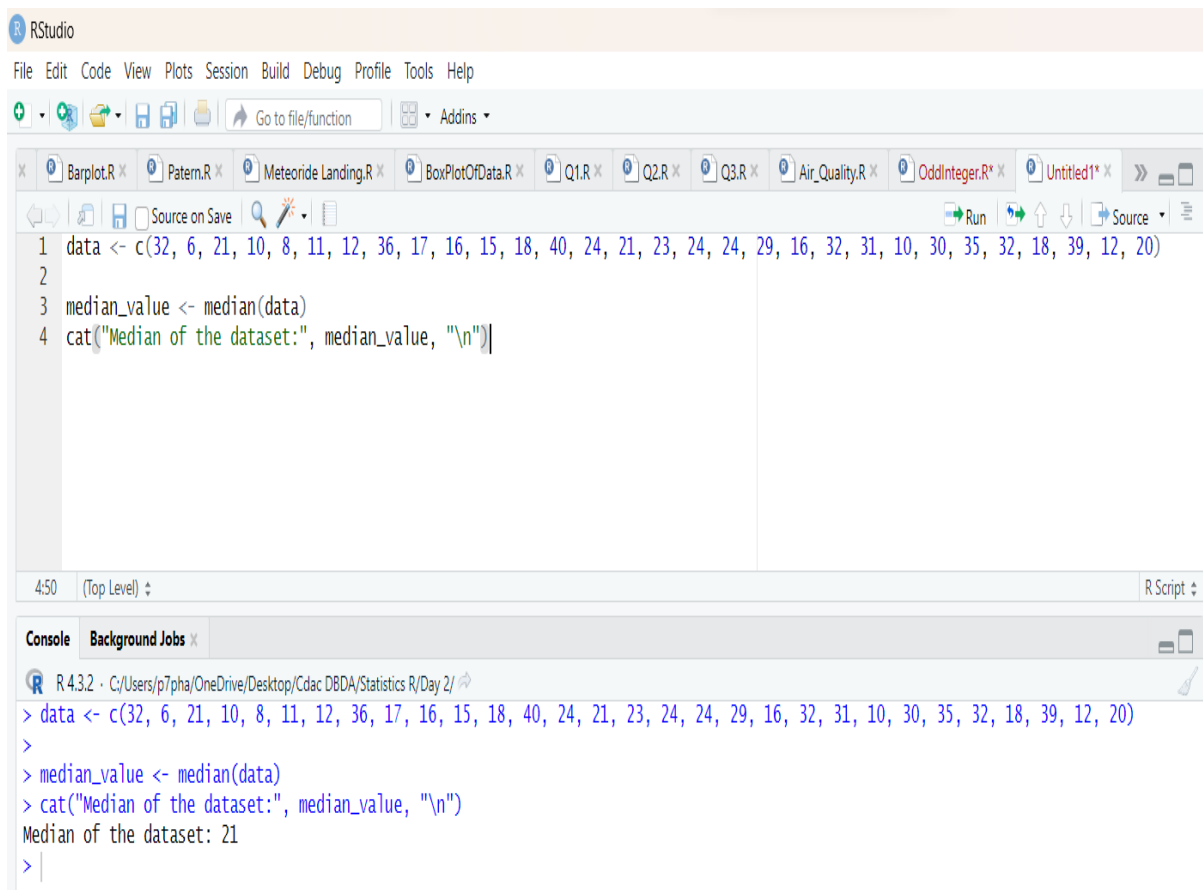
```
> a <- c(1, 3, 5, 7, 9, 11, 13, 15, 17, 19)
> b <- mean(a)
> print(b)
[1] 10
>
```

The status bar at the bottom indicates the current line is 3:9 at the Top Level.

10. What is the median of the following dataset?

32, 6, 21, 10, 8, 11, 12, 36, 17, 16, 15, 18, 40, 24, 21, 23, 24, 24, 29, 16, 32,
31, 10, 30, 35, 32, 18, 39, 12, 20.

Ans=



The screenshot shows the RStudio interface. The script editor contains the following R code:

```
1 data <- c(32, 6, 21, 10, 8, 11, 12, 36, 17, 16, 15, 18, 40, 24, 21, 23, 24, 24, 29, 16, 32, 31, 10, 30, 35, 32, 18, 39, 12, 20)
2
3 median_value <- median(data)
4 cat("Median of the dataset:", median_value, "\n")
```

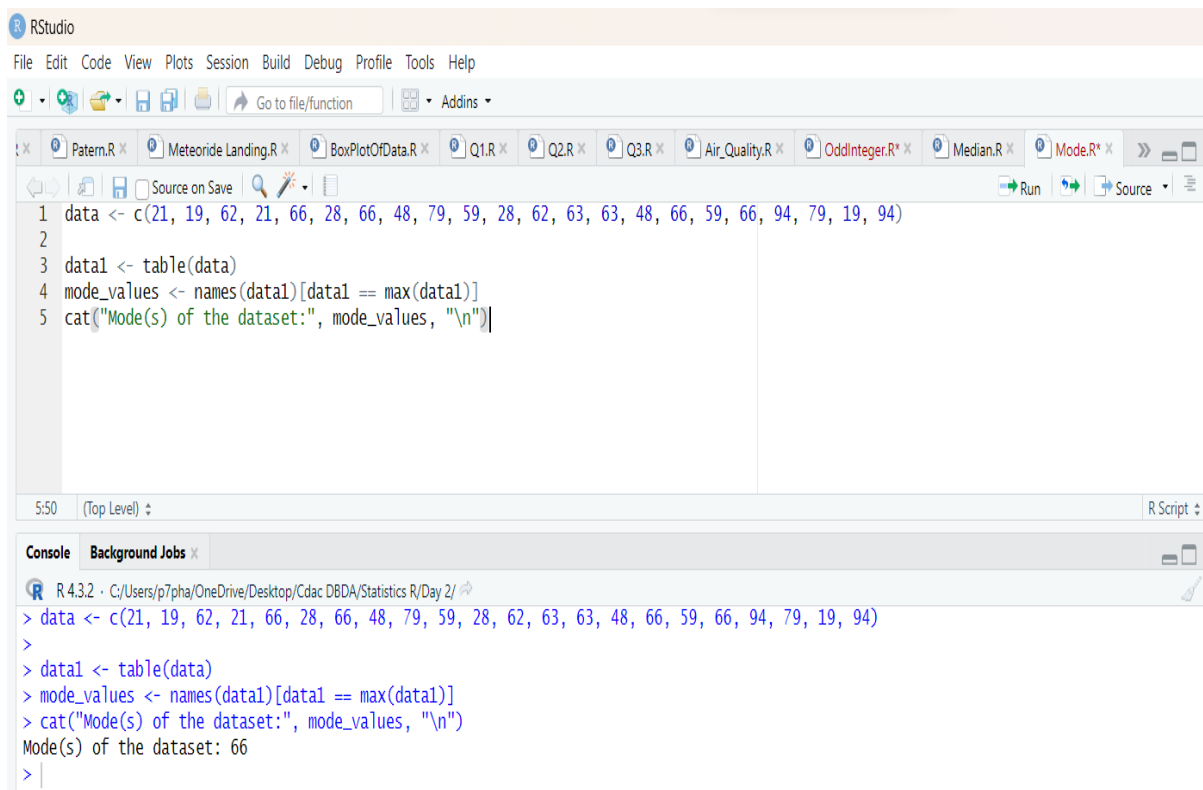
The console output shows the execution of the code:

```
> data <- c(32, 6, 21, 10, 8, 11, 12, 36, 17, 16, 15, 18, 40, 24, 21, 23, 24, 24, 29, 16, 32, 31, 10, 30, 35, 32, 18, 39, 12, 20)
>
> median_value <- median(data)
> cat("Median of the dataset:", median_value, "\n")
Median of the dataset: 21
> |
```


11. Identify the mode for the following dataset

21, 19, 62, 21, 66, 28, 66, 48, 79, 59, 28, 62, 63, 63, 48, 66, 59, 66, 94, 79, 19, 94.

Ans=



The screenshot shows the RStudio interface. The script editor contains the following R code:

```
1 data <- c(21, 19, 62, 21, 66, 28, 66, 48, 79, 59, 28, 62, 63, 63, 48, 66, 59, 66, 94, 79, 19, 94)
2
3 data1 <- table(data)
4 mode_values <- names(data1)[data1 == max(data1)]
5 cat("Mode(s) of the dataset:", mode_values, "\n")
```

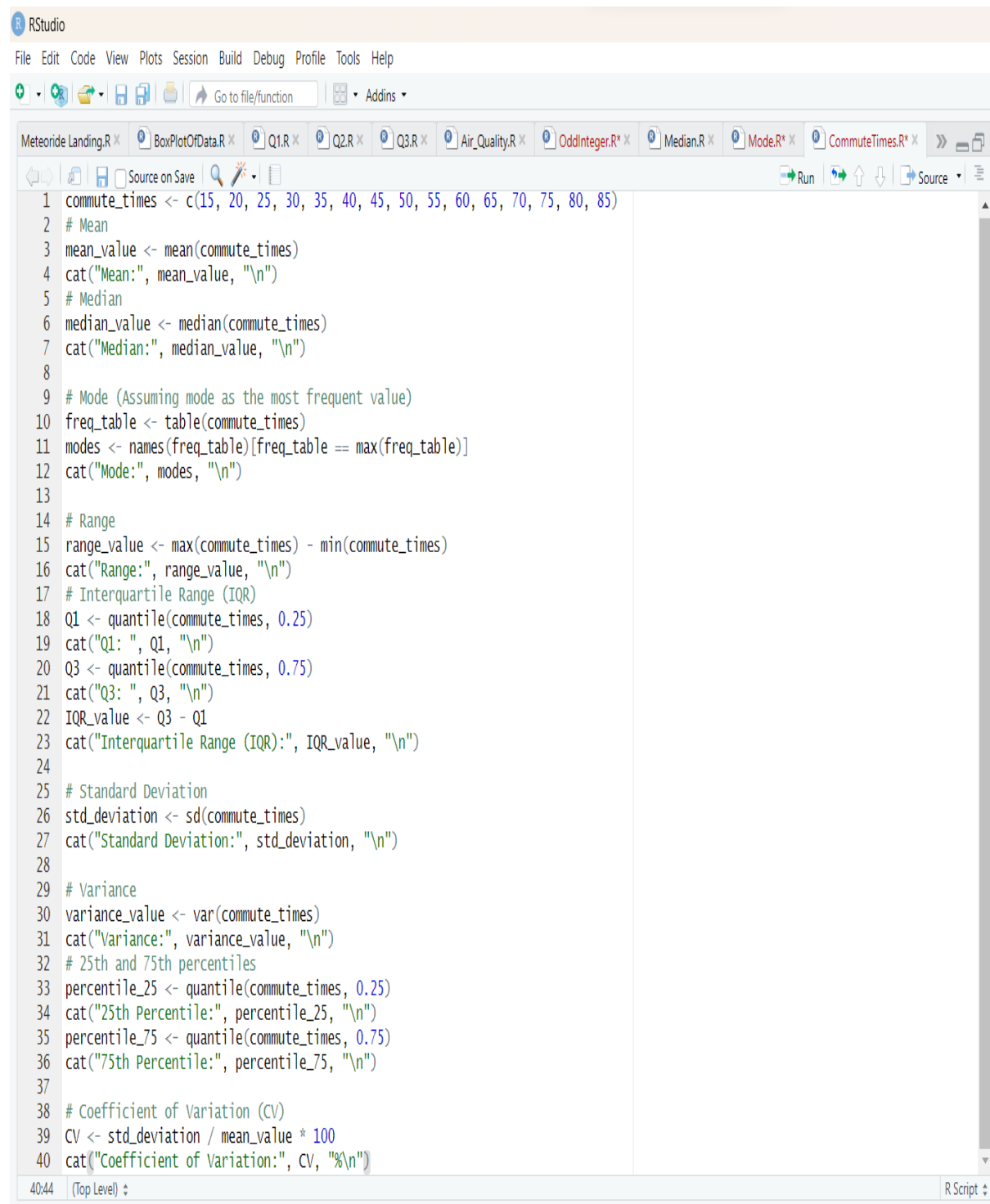
The console output shows the execution of the code:

```
> data <- c(21, 19, 62, 21, 66, 28, 66, 48, 79, 59, 28, 62, 63, 63, 48, 66, 59, 66, 94, 79, 19, 94)
>
> data1 <- table(data)
> mode_values <- names(data1)[data1 == max(data1)]
> cat("Mode(s) of the dataset:", mode_values, "\n")
Mode(s) of the dataset: 66
>
```

12. You are provided with a dataset representing the daily commute times (in minutes) of a group of employees. Your task is to perform a comprehensive analysis using various statistical measures. The dataset is as follows:

`commute_times <- c(15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85).`

Ans=



```
1 commute_times <- c(15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85)
2 # Mean
3 mean_value <- mean(commute_times)
4 cat("Mean:", mean_value, "\n")
5 # Median
6 median_value <- median(commute_times)
7 cat("Median:", median_value, "\n")
8
9 # Mode (Assuming mode as the most frequent value)
10 freq_table <- table(commute_times)
11 modes <- names(freq_table)[freq_table == max(freq_table)]
12 cat("Mode:", modes, "\n")
13
14 # Range
15 range_value <- max(commute_times) - min(commute_times)
16 cat("Range:", range_value, "\n")
17 # Interquartile Range (IQR)
18 Q1 <- quantile(commute_times, 0.25)
19 cat("Q1: ", Q1, "\n")
20 Q3 <- quantile(commute_times, 0.75)
21 cat("Q3: ", Q3, "\n")
22 IQR_value <- Q3 - Q1
23 cat("Interquartile Range (IQR):", IQR_value, "\n")
24
25 # Standard Deviation
26 std_deviation <- sd(commute_times)
27 cat("Standard Deviation:", std_deviation, "\n")
28
29 # Variance
30 variance_value <- var(commute_times)
31 cat("Variance:", variance_value, "\n")
32 # 25th and 75th percentiles
33 percentile_25 <- quantile(commute_times, 0.25)
34 cat("25th Percentile:", percentile_25, "\n")
35 percentile_75 <- quantile(commute_times, 0.75)
36 cat("75th Percentile:", percentile_75, "\n")
37
38 # Coefficient of Variation (CV)
39 CV <- std_deviation / mean_value * 100
40 cat("Coefficient of Variation:", CV, "%\n")
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Background Jobs
R 4.3.2 · C:/Users/p7pha/OneDrive/Desktop/Cdac DBDA/Statistics R/Day 2/
> mean_value <- mean(commute_times)
> cat("Mean:", mean_value, "\n")
Mean: 50
> median_value <- median(commute_times)
> cat("Median:", median_value, "\n")
Median: 50
> freq_table <- table(commute_times)
> modes <- names(freq_table)[freq_table == max(freq_table)]
> cat("Mode:", modes, "\n")
Mode: 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
> range_value <- max(commute_times) - min(commute_times)
> cat("Range:", range_value, "\n")
Range: 70
> Q1 <- quantile(commute_times, 0.25)
> cat("Q1: ", Q1, "\n")
Q1: 32.5
> Q3 <- quantile(commute_times, 0.75)
> cat("Q3: ", Q3, "\n")
Q3: 67.5
> IQR_value <- Q3 - Q1
> cat("Interquartile Range (IQR):", IQR_value, "\n")
Interquartile Range (IQR): 35
> std_deviation <- sd(commute_times)
> cat("Standard Deviation:", std_deviation, "\n")
Standard Deviation: 22.36068
> variance_value <- var(commute_times)
> cat("Variance:", variance_value, "\n")
Variance: 500
> percentile_25 <- quantile(commute_times, 0.25)
> cat("25th Percentile:", percentile_25, "\n")
25th Percentile: 32.5
> percentile_75 <- quantile(commute_times, 0.75)
> cat("75th Percentile:", percentile_75, "\n")
75th Percentile: 67.5
> CV <- std_deviation / mean_value * 100
> cat("Coefficient of Variation:", CV, "%\n")
Coefficient of Variation: 44.72136 %
>
```