

1. Covariance and correlation

Children of three ages are asked to indicate their preference for three photographs of adults. Do the data suggest that there is a significant relationship between age and photograph preference? What is wrong with this study?

		Photograph:		
Age of child		A	B	C
5-6 years:		18	22	20
7-8 years:		2	28	40
9-10 years:		20	10	40

1. Use `cov()` to calculate the sample covariance between B and C.
2. Use another call to `cov()` to calculate the sample covariance matrix for the preferences.
3. Use `cor()` to calculate the sample correlation between B and C.
4. Use another call to `cor()` to calculate the sample correlation matrix for the preferences.

OUTPUT

```
1 Age=c(5:10)
2 A=c(18,2,20)
3 B=c(22,28,10)
4 C=c(20,40,40)
5 Age
6 A
7 B
8 C
9 a=cov(B,C)
10 a
11 b=data.frame(
12   Age=c(5:10),
13   A=c(18,2,20),
14   B=c(22,28,10),
15   C=c(20,40,40)
16 )
```

Console output:

```
R 4.3.0 ~ /~>
> Age=c(5:10)
> A=c(18,2,20)
> B=c(22,28,10)
> C=c(20,40,40)
> Age
[1] 5 6 7 8 9 10
> A
[1] 18 2 20
> B
[1] 22 28 10
> C
[1] 20 40 40
> a=cov(B,C)
> a
```

Environment pane shows:

- Data: 6 obs. of 4 variables
- b: num [1:4, 1:4] 3.5 0.8 -4.8 8 0.8 ...
- d: num [1:4, 1:4] 1 0.0485 -0.313 0.414 0.0485 ...
- f: num [1:4, 1:4] 1 0.0485 -0.313 0.414 0.0485 ...
- Values: a -20

Files pane shows:

- Home
- Name
- Size
- Modified
- .Rhistory 35 B May 1, 2023, 11:11 AM
- desktop.ini 402 B May 1, 2023, 6:27 PM
- hello world.cpp 73 B May 1, 2023, 11:02 AM
- hello world.exe 322.6 KB May 1, 2023, 11:02 AM
- My Music
- My Pictures
- My Videos
- sum.cpp 177 B May 1, 2023, 11:13 AM
- sum.exe 3 MB May 1, 2023, 11:13 AM

dataset using an equal-frequency partitioning method with bin equal to 3 (ii) apply data smoothing using bin means and bin boundary.
(iii) Plot Histogram for the above frequency division

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation:


```
d=c(1,1,5,5,5,5,5,8,8,10,10,10,10,12,14,14,14,15,15,15,15,15,18,18,18,18,18,20,20,20,20,20,21,21,21,21,25,25,25,25,28,28,30)
d
breaks<-quantile(d,seq(0,1,1/3))
partitioned_data=cut(d,breaks,include.lowest=TRUE,labels=FALSE)
partitioned_data
mean_smoothing=function(d,bin_size)
7-{
  breaks=seq(min(d),max(d),by=3)
  partitioned_data=cut(d,breaks,include.lowest=TRUE,labels=FALSE)
  bin_means=tapply(d,partitioned_data,mean)
  smoothed_data=bin_means[partitioned_data]
  return(smoothed_data)
13-}
14 smoothed_data <- mean_smoothing(d,3)
15 smoothed_data
16 hist(data, breaks = 3, col = "lightblue", xlab = "Prices", main = "Histogram")
17
965 mean_smoothing(d,bin_size) :
```
- Console:** Shows the execution output:


```
R 4.3.0 ~ /
> d=c(1,1,5,5,5,5,5,8,8,10,10,10,10,12,14,14,14,15,15,15,15,15,18,18,18,18,18,20,20,20,20,20,21,21,21,21,25,25,25,25,28,28,30)
> d
[1] 1 1 5 5 5 5 5 8 8 10 10 10 10 12 14 14 14 15 15 15 15 15 18 18 18 18 18
[25] 20 20 20 20 20 21 21 21 25 25 25 25 28 28 30
> breaks<-quantile(d,seq(0,1,1/3))
> partitioned_data=cut(d,breaks,include.lowest=TRUE,labels=FALSE)
> partitioned_data
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3
[38] 3 3 3 3 3 3 3 3
> mean_smoothing=function(d,bin_size)
+ {
  breaks=seq(min(d),max(d),by=3)
  partitioned_data=cut(d,breaks,include.lowest=TRUE,labels=FALSE)
  bin_means=tapply(d,partitioned_data,mean)
  smoothed_data=bin_means[partitioned_data]
  return(smoothed_data)
}
```
- Environment:** Lists variables in the Global Environment:

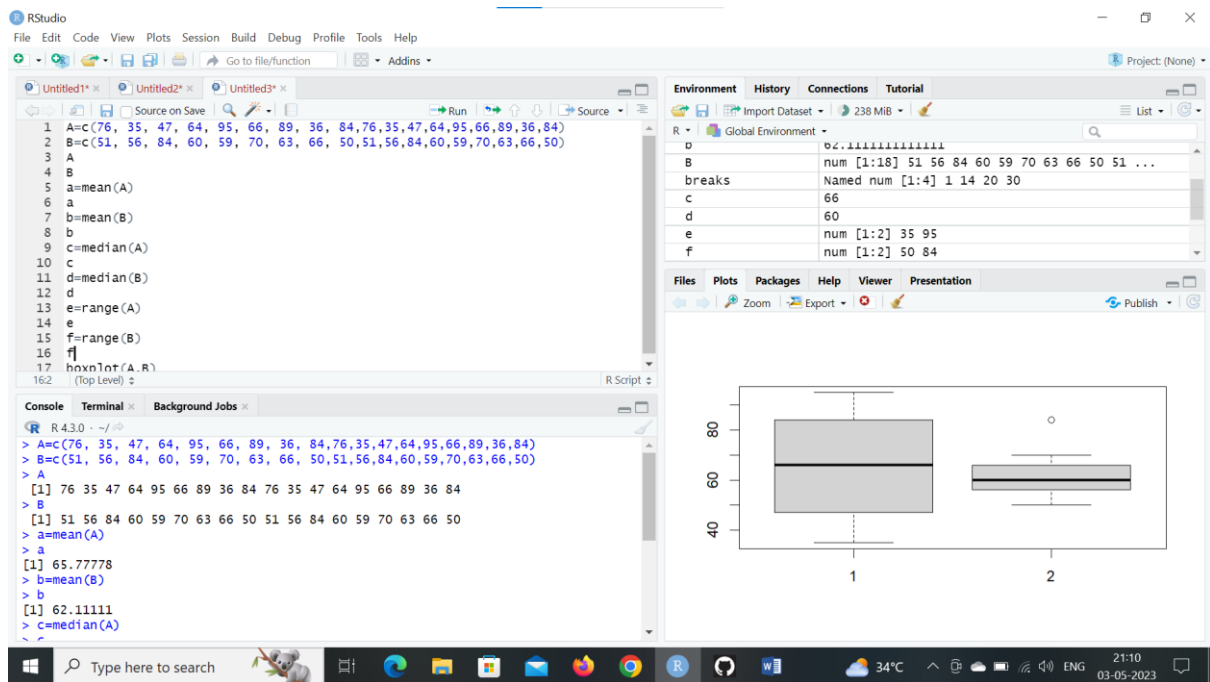
Variable	Value
breaks	Named num [1:4]: 1 14 20 30
d	num [1:46]: 1 1 5 5 5 5 8 8 10 10 12 14 14 14 15 15 15 15 15 18 18 18 18 18 20 20 20 20 20 21 21 21 25 25 25 25 28 28 30
partitioned_data	int [1:46]: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3
smoothed_data	num [1:46]: 1 1 5 5 5 5 8 8 10 10 12 14 14 14 15 15 15 15 15 18 18 18 18 18 20 20 20 20 20 21 21 21 25 25 25 25 28 28 30
- Files:** Lists files in the project:

Name	Size	Modified
.Rhistory	35 B	May 1, 2023, 11:11 AM
desktop.ini	402 B	May 1, 2023, 6:27 PM
hello world.cpp	73 B	May 1, 2023, 11:02 AM
hello world.exe	322.6 KB	May 1, 2023, 11:02 AM
My Music		
My Pictures		
My Videos		
sum.cpp	177 B	May 1, 2023, 11:13 AM
sum.exe	3 MB	May 1, 2023, 11:13 AM

Class A: 76, 35, 47, 64, 95, 66, 89, 36, 84, 76, 35, 47, 64, 95, 66, 89, 36, 84

- Find which class had scored higher mean, median and range.
- Plot above in boxplot and give the inferences

OUTPUT

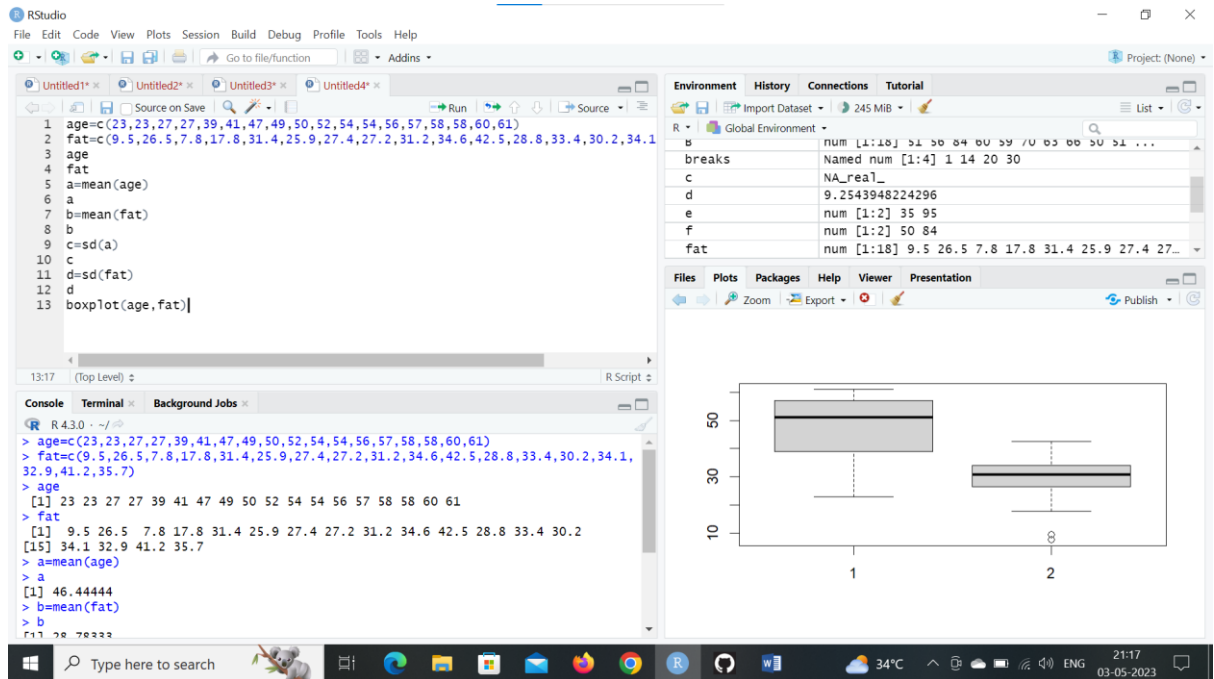


4. Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

<i>age</i>	23	23	27	27	39	41	47	49	50
<i>%fat</i>	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
<i>age</i>	52	54	54	56	57	58	58	60	61
<i>%fat</i>	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- (a) Calculate the mean, median, and standard deviation of age and %fat. (b) Draw the boxplots for age and %fat.

OUTPUT



5. suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- Use min-max normalization to transform the value 39 for age onto the range [0.0, 1.0].
- Use z-score normalization to transform the value 39 for age, where the standard deviation of age is 12.94 years.
- Use normalization by decimal scaling to transform the value 35 for age. Perform the above functions using R – tool

OUTPUT

```

1 age=c(23,23,27,27,39,41,47,49,50,52,54,54,56,57,58,60,61)
2 fat=c(9.5,26.5,7.8,17.8,31.4,25.9,27.4,27.2,31.2,34.6,42.5,28.8,33.4,30.2,34.1,32.9,41.2,35.7)
3 age
4 fat
5 age_norm=c((39-min(age))/(max(age)-min(age)))
6 age_norm
7 age_znorm <- (39 - mean(age)) / sd(age)
8 age_znorm
9 max_val <- max(abs(age))
10 n <- nchar(as.character(max_val)) - 1
11 age_dec_norm <- 35 / 10^n
12 age_dec_norm

```

Environment:

Variable	Value
age	num [1:16] 23 23 27 27 39 41 47 49 50 52 54 54 56 57 58 60 61
fat	num [1:16] 9.5 26.5 7.8 17.8 31.4 25.9 27.4 27.2 31.2 34.6 42.5 28.8 33.4 30.2 34.1 32.9 41.2 35.7
age_norm	num [1:16] 0.4210526
age_znorm	num [1:16] -0.9486833 -0.6324555 -0.3162278 0.3162278 1.5811388
max_val	61
n	1

6. Use following group of data: 200, 300, 400, 600, 1000
- (a) min-max normalization by setting min = 0 and max = 1 (b)
 - (b) z-score normalization
 - (c) (c) z-score normalization using the mean absolute deviation instead of standard deviation (d) normalization by decimal scaling

OUTPUT

```

1 a=c(200, 300, 400, 600, 1000)
2 a
3 min_max_norm <- (a-min(a))/(max(a)-min(a))
4 min_max_norm
5 z_score_norm <- (a-mean(a))/sd(a)
6 z_score_norm
7 mad <- mean(abs(a-median(a)))
8 z_score_mad_norm <- (a-median(a))/mad
9 z_score_mad_norm
10 max_val <- max(abs(a))
11 n <- nchar(as.character(max_val)) - 1
12 a_dec_norm <- a/1000^n
13 a_dec_norm

```

Environment:

Variable	Value
a	num [1:5] 200 300 400 600 1000
min_max_norm	num [1:5] 0 0.125 0.25 0.5 1
z_score_norm	num [1:5] -0.9486833 -0.6324555 -0.3162278 0.3162278 1.5811388
z_score_mad_norm	num [1:5] -0.9090909 -0.4545455 0.0000000 0.9090909 2.7272727
max_val	1000
n	1

7. Two Maths teachers are comparing how their Year 9 classes performed in the end of year exams. Their results are as follows:

Class A: 76, 35, 47, 64, 95, 66, 89, 36, 84, 76, 35, 47, 64, 95, 66, 89, 36, 84

Class B: 51, 56, 84, 60, 59, 70, 63, 66, 50, 51, 56, 84, 60, 59, 70, 63, 66, 50

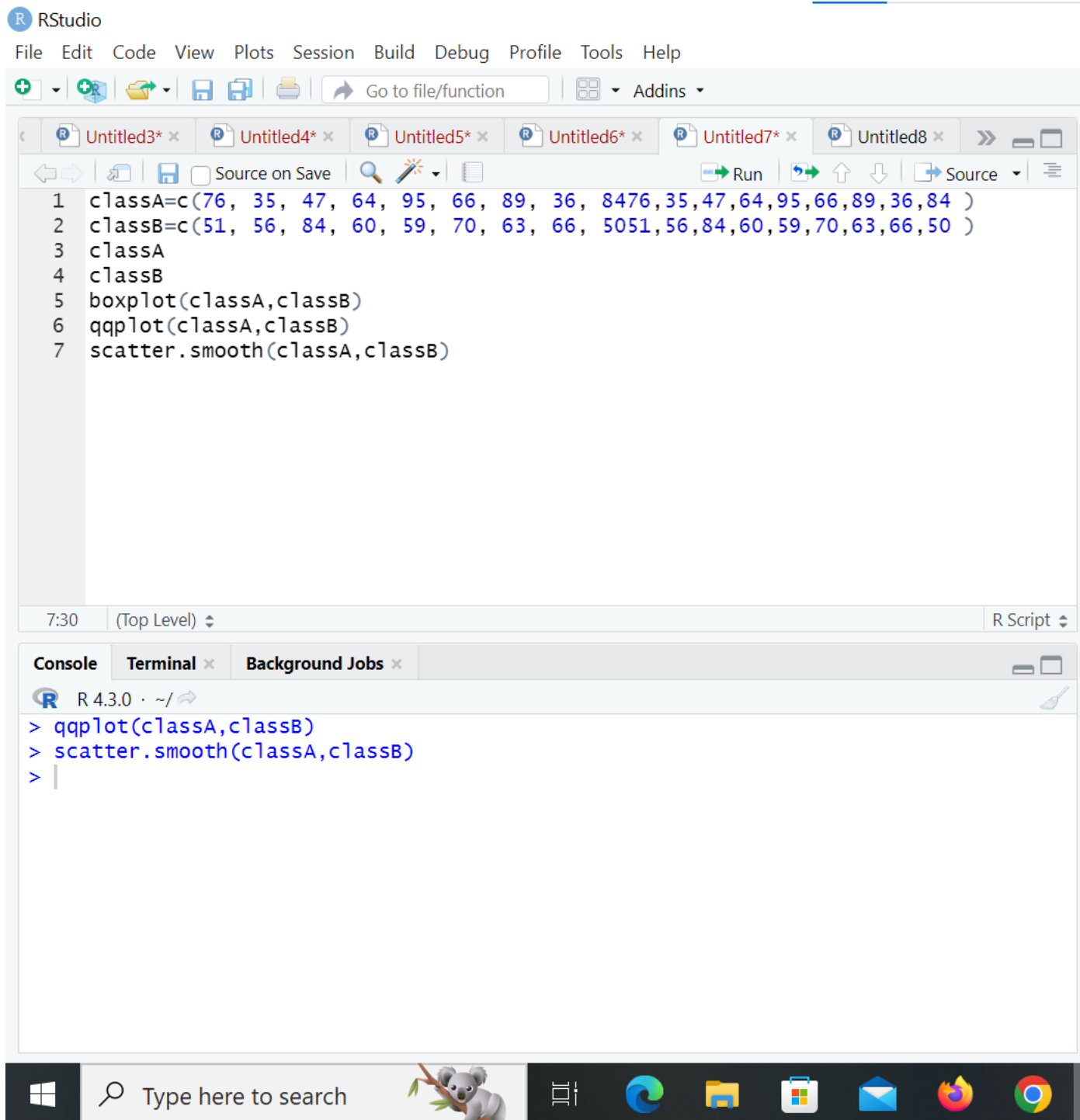
Draw the boxplot, scatter plot, qq plot for class a and b

The screenshot shows the RStudio interface with the following components:

- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations (new, open, save, print), navigation (back, forward), and execution (run, source).
- Source Editor:** Contains the following R code:

```
1 classA=c(76, 35, 47, 64, 95, 66, 89, 36, 84, 76, 35, 47, 64, 95, 66, 89, 36, 84 )
2 classB=c(51, 56, 84, 60, 59, 70, 63, 66, 50, 51, 56, 84, 60, 59, 70, 63, 66, 50 )
3 classA
4 classB
5 boxplot(classA,classB)
6 qqplot(classA,classB)
7 scatter.smooth(classA,classB)|
```
- Console:** Shows the execution of the following commands:

```
> qqplot(classA,classB)
> scatter.smooth(classA,classB)
> qqplot(classA,classB)
> boxplot(classA,classB)
> |
```
- Status Bar:** Displays '7:30' and '(Top Level)'.
- Taskbar:** At the bottom, showing the Windows Start button, search bar, and various application icons (RStudio, File Explorer, Calendar, Mail, Firefox, Chrome).



8. . suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

<i>age</i>	23	23	27	27	39	41	47	49	50
<i>%fat</i>	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
<i>age</i>	52	54	54	56	57	58	58	60	61
<i>%fat</i>	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

Draw the boxplot,scatter plot,qq plot

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled3* x Untitled4* x Untitled5* x Untitled6* x Untitled7* x Untitled8* x

Source on Save Run Source

```
1 age=c(23,23,27,27,39,41,47,49,50,52,54,54,56,57,58,58,60,61)
2 fat=c(9.5,26.5,7.8,17.8,31.4,25.9,27.4,27.2,31.2,34.6,42.5,28.8,33.4,30.2,34.1,
3 age
4 fat
5 boxplot(age,fat)
6 scatter.smooth(age,fat)
7 qqplot(age,fat)
```

7:1 (Top Level) R Script

Console Terminal x Background Jobs x

R 4.3.0 · ~/

```
> scatter.smooth(age,fat)
> qqplot(age,fat)
> |
```



Type here to search



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source on Save Run Source

```
1 age=c(23,23,27,27,39,41,47,49,50,52,54,54,56,57,58,58,60,61)
2 fat=c(9.5,26.5,7.8,17.8,31.4,25.9,27.4,27.2,31.2,34.6,42.5,28.8,33.4,30.2,34.1,
3 age
4 fat
5 boxplot(age,fat)
```

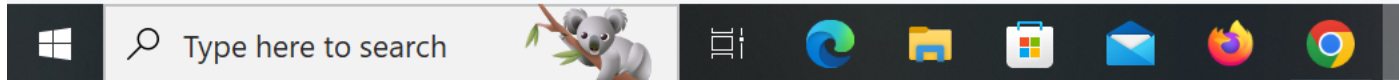
5:17 (Top Level) R Script

Console Terminal Background Jobs

R 4.3.0 · ~/

```
> age=c(23,23,27,27,39,41,47,49,50,52,54,54,56,57,58,58,60,61)
> fat=c(9.5,26.5,7.8,17.8,31.4,25.9,27.4,27.2,31.2,34.6,42.5,28.8,33.4,30.2,34.1,3
2.9,41.2,35.7)
> age
[1] 23 23 27 27 39 41 47 49 50 52 54 54 56 57 58 58 60 61
> fat
[1] 9.5 26.5 7.8 17.8 31.4 25.9 27.4 27.2 31.2 34.6 42.5 28.8 33.4 30.2
[15] 34.1 32.9 41.2 35.7
> boxplot(age,fat)
>
```

Type here to search



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled3* x Untitled4* x Untitled5* x Untitled6* x Untitled7* x Untitled8* x

Source on Save Run Source

```
1 age=c(23,23,27,27,39,41,47,49,50,52,54,54,56,57,58,58,60,61)
2 fat=c(9.5,26.5,7.8,17.8,31.4,25.9,27.4,27.2,31.2,34.6,42.5,28.8,33.4,30.2,34.1,
3 age
4 fat
5 boxplot(age,fat)
6 scatter.smooth(age,fat)
```

6:24 (Top Level) R Script

Console Terminal x Background Jobs x

R 4.3.0 · ~/

```
> scatter.smooth(age,fat)
> |
```



Type here to search

