



American International University-Bangladesh (AIUB)

Department of Computer Science and Engineering (CSE)

Faculty of Science & Technology (FST)

Spring 2022-23

Course: Introduction to Data Science

Course Code: CSC 4180

Section: E

Mid Term Project

Submitted by:

Fahim Mahmud Bhuiyan

ID: 20-42970-1

Department: CSE

Submitted To:

Tohedul Islam

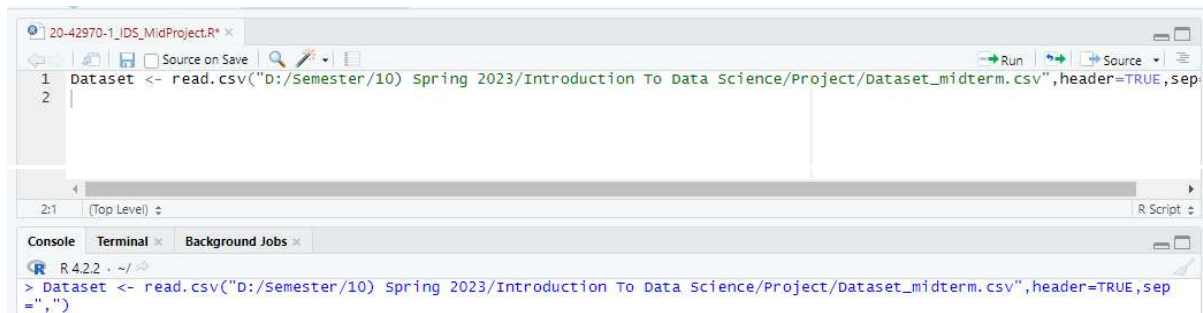
Assistant Professor

Department of Computer Science

Submitted Date: 13th March 2023

1) Import the Data Set

```
Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep=",")
```



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

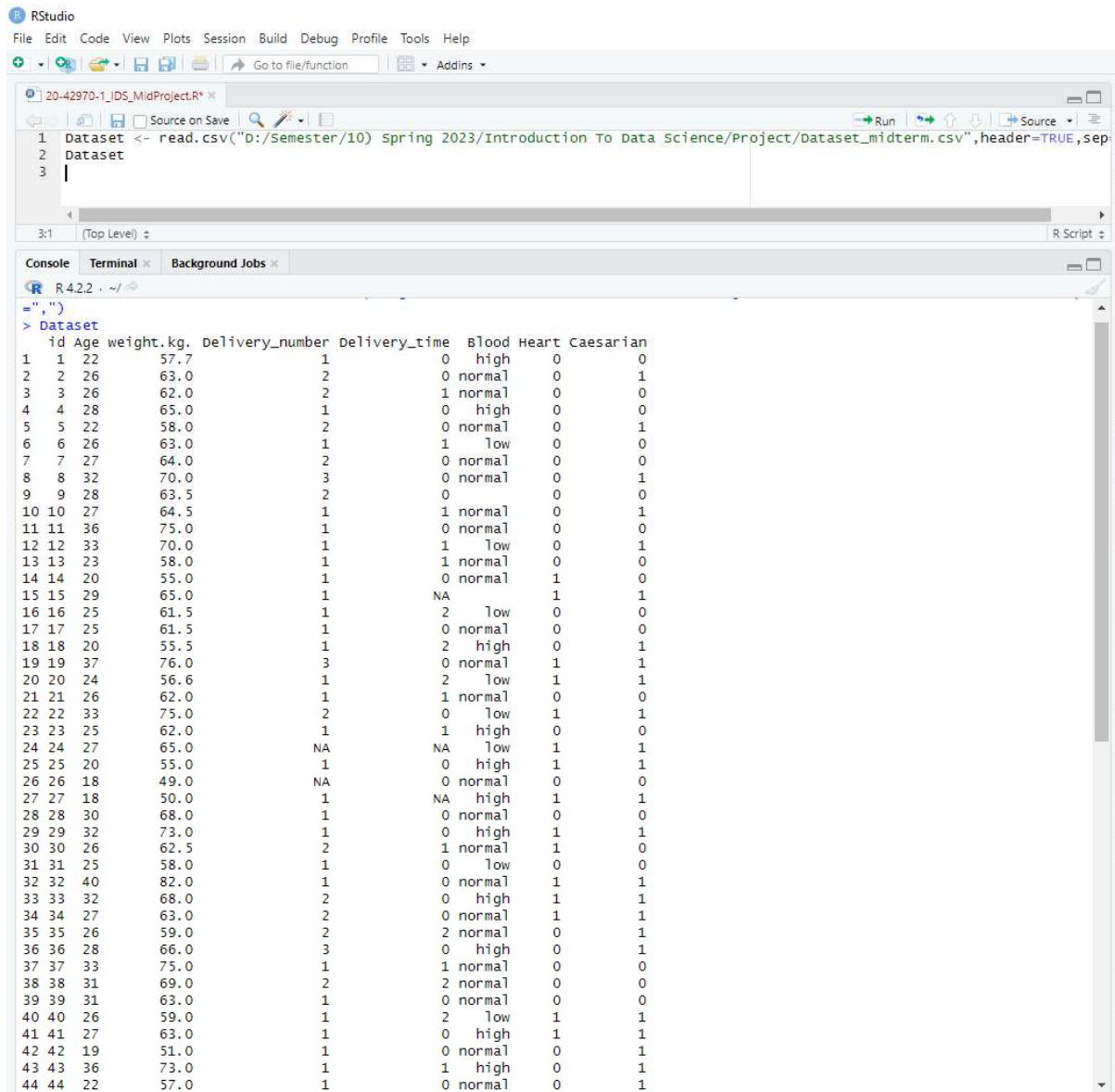
```
1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep=",")
2
```

The console shows the execution of the code:

```
> Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep=",")
```

2) Print the Data Set

Dataset



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

```
1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep=",")
2 Dataset
3
```

The console shows the output of the `Dataset` command:

```
> Dataset
  id Age weight.kg. Delivery_number Delivery_time Blood Heart Caesarian
1  1  22    57.7         1             0 high    0         0
2  2  26    63.0         2             0 normal  0         1
3  3  26    62.0         2             1 normal  0         0
4  4  28    65.0         1             0 high    0         0
5  5  22    58.0         2             0 normal  0         1
6  6  26    63.0         1             1 low     0         0
7  7  27    64.0         2             0 normal  0         0
8  8  32    70.0         3             0 normal  0         1
9  9  28    63.5         2             0         0         0
10 10 27    64.5         1             1 normal  0         1
11 11 36    75.0         1             0 normal  0         0
12 12 33    70.0         1             1 low     0         1
13 13 23    58.0         1             1 normal  0         0
14 14 20    55.0         1             0 normal  1         0
15 15 29    65.0         1             NA        1         1
16 16 25    61.5         1             2 low     0         0
17 17 25    61.5         1             0 normal  0         0
18 18 20    55.5         1             2 high    0         1
19 19 37    76.0         3             0 normal  1         1
20 20 24    56.6         1             2 low     1         1
21 21 26    62.0         1             1 normal  0         0
22 22 33    75.0         2             0 low     1         1
23 23 25    62.0         1             1 high    0         0
24 24 27    65.0         NA            NA low     1         1
25 25 20    55.0         1             0 high    1         1
26 26 18    49.0         NA            0 normal  0         0
27 27 18    50.0         1             NA high    1         1
28 28 30    68.0         1             0 normal  0         0
29 29 32    73.0         1             0 high    1         1
30 30 26    62.5         2             1 normal  1         0
31 31 25    58.0         1             0 low     0         0
32 32 40    82.0         1             0 normal  1         1
33 33 32    68.0         2             0 high    1         1
34 34 27    63.0         2             0 normal  1         1
35 35 26    59.0         2             2 normal  0         1
36 36 28    66.0         3             0 high    0         1
37 37 33    75.0         1             1 normal  0         0
38 38 31    69.0         2             2 normal  0         0
39 39 31    63.0         1             0 normal  0         0
40 40 26    59.0         1             2 low     1         1
41 41 27    63.0         1             0 high    1         1
42 42 19    51.0         1             0 normal  0         1
43 43 36    73.0         1             1 high    0         1
44 44 22    57.0         1             0 normal  0         1
```

3) Show The Attributes Name from the Data Set

names(Dataset)

```
20-42970-1_IDS_MidProject.R
1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep
2 Dataset
3 names(Dataset)

> names(Dataset)
[1] "id"          "Age"          "weight.kg."   "delivery_number" "delivery_time" "blood"
[7] "Heart"       "Caesarian"
```

4) Summary of every attribute from the Data Set

summary(Dataset)

```
20-42970-1_IDS_MidProject.R
1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep
2 Dataset
3 names(Dataset)
4 summary(Dataset)

4:1 (Top Level)
R Script

Console Terminal Background Jobs
R 4.2.2 ~ /
> summary(Dataset)
      id      Age      weight.kg.  delivery_number delivery_time      blood      Heart
Min.   :1.00   Min.   :18.00   Min.   : 49.00   Min.   :1.000   Min.   :0.0000   Length:80   Min.   :0.000
1st Qu.:20.75  1st Qu.:25.00   1st Qu.: 61.00   1st Qu.:1.000   1st Qu.:0.0000   class :character 1st Qu.:0.000
Median :40.50  Median :28.00   Median : 63.50   Median :1.500   Median :0.0000   Mode  :character Median :0.000
Mean   :40.50  Mean   :29.68   Mean   : 65.13   Mean   :1.679   Mean   :0.6234   Mean   :0.375
3rd Qu.:60.25  3rd Qu.:32.00   3rd Qu.: 68.00   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:1.000
Max.   :80.00  Max.   :95.00   Max.   :110.00   Max.   :4.000   Max.   :2.0000   Max.   :1.000

      Caesarian
Min.   :0.0000
1st Qu.:0.0000
Median :1.0000
Mean   :0.5641
3rd Qu.:1.0000
Max.   :1.0000
NA's   :2
```

5) Find all types of data from a Data Set

str(Dataset)

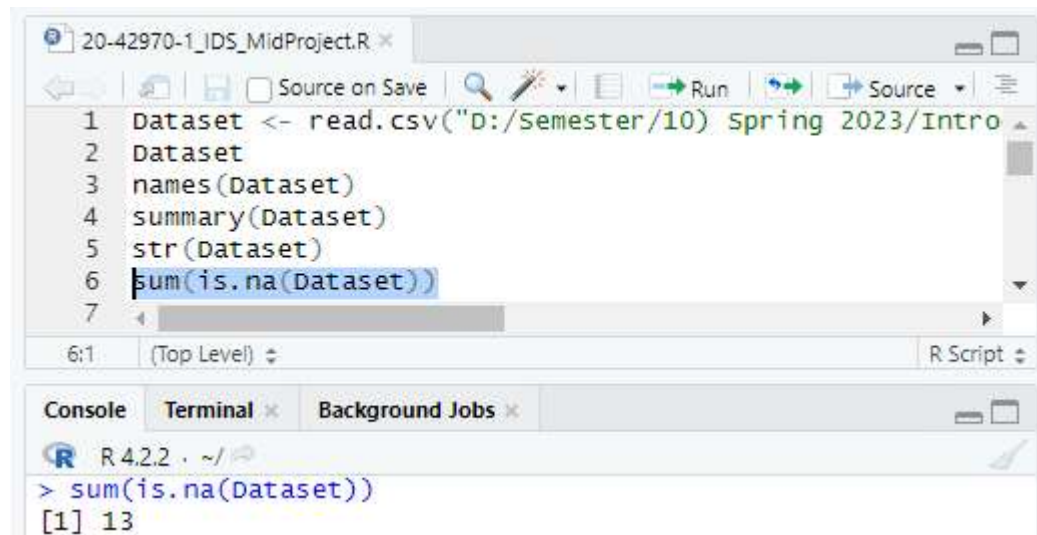
```
20-42970-1_IDS_MidProject.R
1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep
2 Dataset
3 names(Dataset)
4 summary(Dataset)
5 str(Dataset)

5:1 (Top Level)
R Script

Console Terminal Background Jobs
R 4.2.2 ~ /
      Caesarian      NA's :3      NA's :3      NA's :2      NA's :3
Min.   :0.0000
1st Qu.:0.0000
Median :1.0000
Mean   :0.5641
3rd Qu.:1.0000
Max.   :1.0000
NA's   :2
> str(Dataset)
'data.frame': 80 obs. of 8 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Age     : int  22 26 26 28 22 26 27 32 28 27 ...
 $ weight.kg.: num  57.7 63 62 65 58 63 64 70 63.5 64.5 ...
 $ delivery_number: int  1 2 2 1 2 1 2 3 2 1 ...
 $ delivery_time : int  0 0 1 0 0 1 0 0 0 1 ...
 $ blood      : chr  "high" "normal" "normal" "high" ...
 $ Heart      : int  0 0 0 0 0 0 0 0 0 ...
 $ Caesarian   : int  0 1 0 0 1 0 0 1 0 1 ...
```

6) Show the Sum of The Data Set

`sum(is.na(Dataset))`



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

```
1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Intro
2 Dataset
3 names(Dataset)
4 summary(Dataset)
5 str(Dataset)
6 sum(is.na(Dataset))
7
```

The console output shows the result of the command:

```
> sum(is.na(Dataset))
[1] 13
```

7) Show the Types of Data for All Attributes.

`mode(Dataset$id)`

`mode(Dataset$Age)`

`mode(Dataset$weight.kg.)`

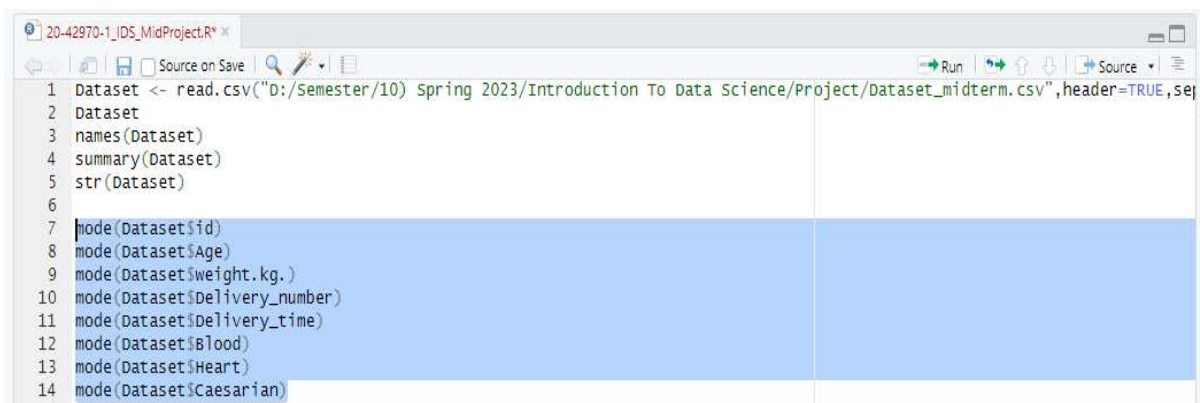
`mode(Dataset$Delivery_number)`

`mode(Dataset$Delivery_time)`

`mode(Dataset$Blood)`

`mode(Dataset$Heart)`

`mode(Dataset$Caesarian)`



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

```
1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep
2 Dataset
3 names(Dataset)
4 summary(Dataset)
5 str(Dataset)
6
7 mode(Dataset$id)
8 mode(Dataset$Age)
9 mode(Dataset$weight.kg.)
10 mode(Dataset$Delivery_number)
11 mode(Dataset$Delivery_time)
12 mode(Dataset$Blood)
13 mode(Dataset$Heart)
14 mode(Dataset$Caesarian)
```

```
R 4.2.2 . ~/
7:1 (Top Level)
R Script

Console Terminal Background Jobs

$ Blood      : chr  "high" "normal" "normal" "high" ...
$ Heart      : int   0 0 0 0 0 0 0 0 0 ...
$ Caesarian  : int   0 1 0 0 1 0 0 1 0 1 ...

> mode(Dataset$id)
[1] "numeric"
> mode(Dataset$Age)
[1] "numeric"
> mode(Dataset$weight.kg.)
[1] "numeric"
> mode(Dataset$Delivery_number)
[1] "numeric"
> mode(Dataset$Delivery_time)
[1] "numeric"
> mode(Dataset$Blood)
[1] "character"
> mode(Dataset$Heart)
[1] "numeric"
> mode(Dataset$Caesarian)
[1] "numeric"
```

8) Compute the mean and median value of the Heart

```
meanHeart = mean(Dataset$Heart)
```

```
print(meanHeart)
```

```
medianHeart = median(Dataset$Heart)
```

```
print(medianHeart)
```

```
20-42970-1_IDS_MidProject.R*
Source on Save

1 Dataset <- read.csv("D:/Semester/10")
2 Dataset
3 names(Dataset)
4 summary(Dataset)
5 str(Dataset)
6
7 mode(Dataset$id)
8 mode(Dataset$Age)
9 mode(Dataset$weight.kg.)
10 mode(Dataset$Delivery_number)
11 mode(Dataset$Delivery_time)
12 mode(Dataset$Blood)
13 mode(Dataset$Heart)
14 mode(Dataset$Caesarian)
15
16 meanHeart = mean(Dataset$Heart)
17 print(meanHeart)
18
19 medianHeart = median(Dataset$Heart)
20 print(medianHeart)
```

```
19:1 (Top Level)
R 4.2.2 . ~/

> meanHeart = mean(Dataset$Heart)
> print(meanHeart)
[1] 0.375
> medianHeart = median(Dataset$Heart)
> print(medianHeart)
[1] 0
```



```
names(sort(table(Dataset$weight.kg)))
```

```
names(sort(table(Dataset$Blood)))
```

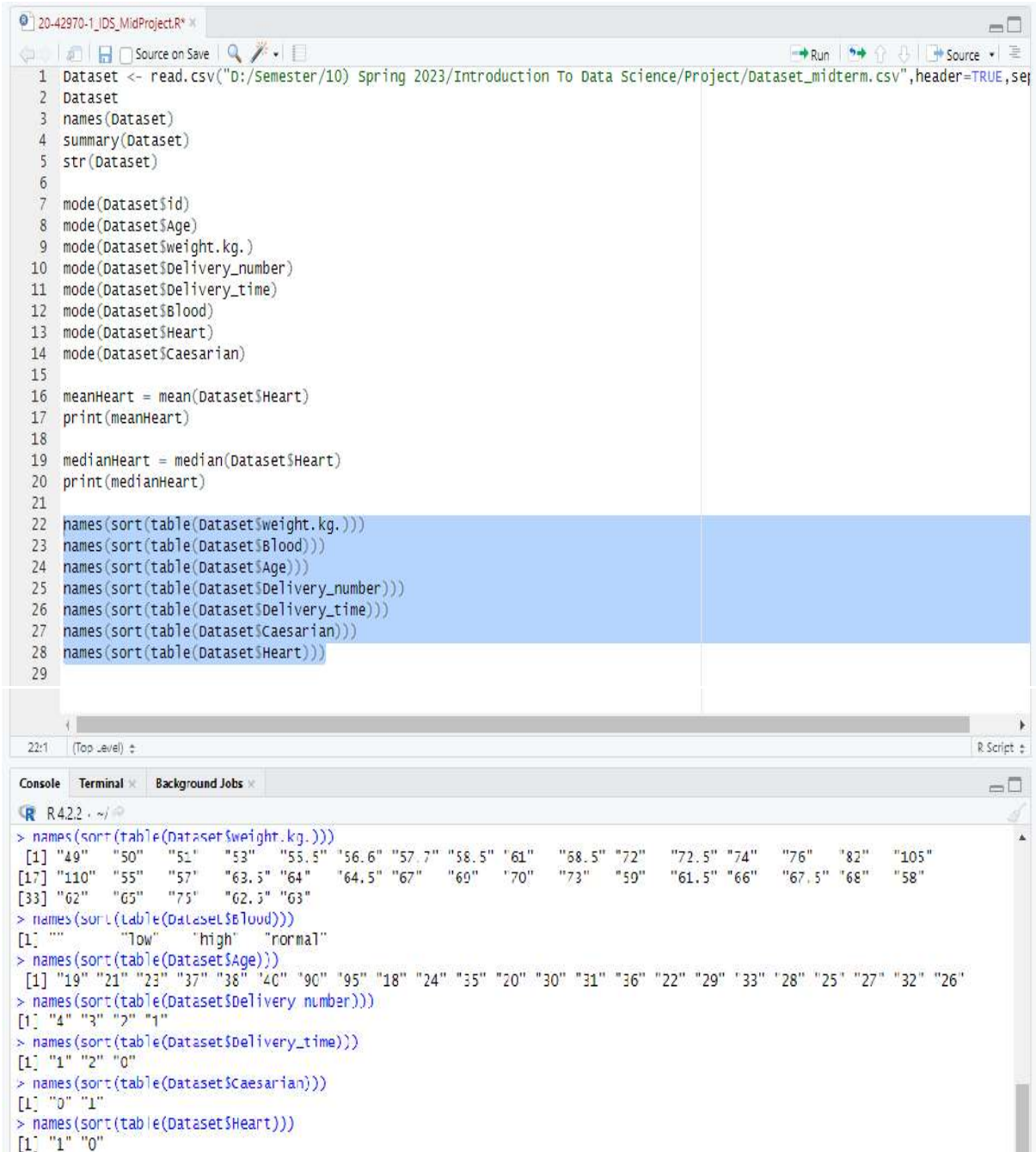
```
names(sort(table(Dataset$Age)))
```

```
names(sort(table(Dataset$Delivery_number)))
```

```
names(sort(table(Dataset$Delivery_time)))
```

```
names(sort(table(Dataset$Caesarian)))
```

```
names(sort(table(Dataset$Heart)))
```



The screenshot shows an RStudio window with a script editor and a console. The script editor contains R code for reading a CSV file and performing various statistical operations. The console shows the output of these operations.

```
20-42970-1_IDS_MidProject.R
Source on Save
Run
Source

1 Dataset <- read.csv("D:/Semester/10) Spring 2023/Introduction To Data Science/Project/Dataset_midterm.csv",header=TRUE,sep
2 Dataset
3 names(Dataset)
4 summary(Dataset)
5 str(Dataset)
6
7 mode(Dataset$id)
8 mode(Dataset$Age)
9 mode(Dataset$weight.kg.)
10 mode(Dataset$Delivery_number)
11 mode(Dataset$Delivery_time)
12 mode(Dataset$Blood)
13 mode(Dataset$Heart)
14 mode(Dataset$Caesarian)
15
16 meanHeart = mean(Dataset$Heart)
17 print(meanHeart)
18
19 medianHeart = median(Dataset$Heart)
20 print(medianHeart)
21
22 names(sort(table(Dataset$weight.kg.)))
23 names(sort(table(Dataset$Blood)))
24 names(sort(table(Dataset$Age)))
25 names(sort(table(Dataset$Delivery_number)))
26 names(sort(table(Dataset$Delivery_time)))
27 names(sort(table(Dataset$Caesarian)))
28 names(sort(table(Dataset$Heart)))
29
```

Console Output:

```
R 4.2.2 ~
> names(sort(table(Dataset$weight.kg.)))
[1] "49" "50" "51" "53" "55.5" "56.6" "57.7" "58.5" "61" "68.5" "72" "72.5" "74" "76" "82" "105"
[17] "110" "55" "57" "63.5" "64" "64.5" "67" "69" "70" "73" "59" "61.5" "66" "67.5" "68" "58"
[33] "62" "65" "75" "62.5" "63"
> names(sort(table(Dataset$Blood)))
[1] "" "low" "high" "normal"
> names(sort(table(Dataset$Age)))
[1] "19" "21" "23" "37" "38" "40" "90" "95" "18" "24" "35" "20" "30" "31" "36" "22" "29" "33" "28" "25" "27" "32" "26"
> names(sort(table(Dataset$Delivery_number)))
[1] "4" "3" "2" "1"
> names(sort(table(Dataset$Delivery_time)))
[1] "1" "2" "0"
> names(sort(table(Dataset$Caesarian)))
[1] "0" "1"
> names(sort(table(Dataset$Heart)))
[1] "1" "0"
```

9) Show the Standard Deviation of all attributes.

```
sdweight= sd(Dataset$weight.kg.)
print(sdweight)

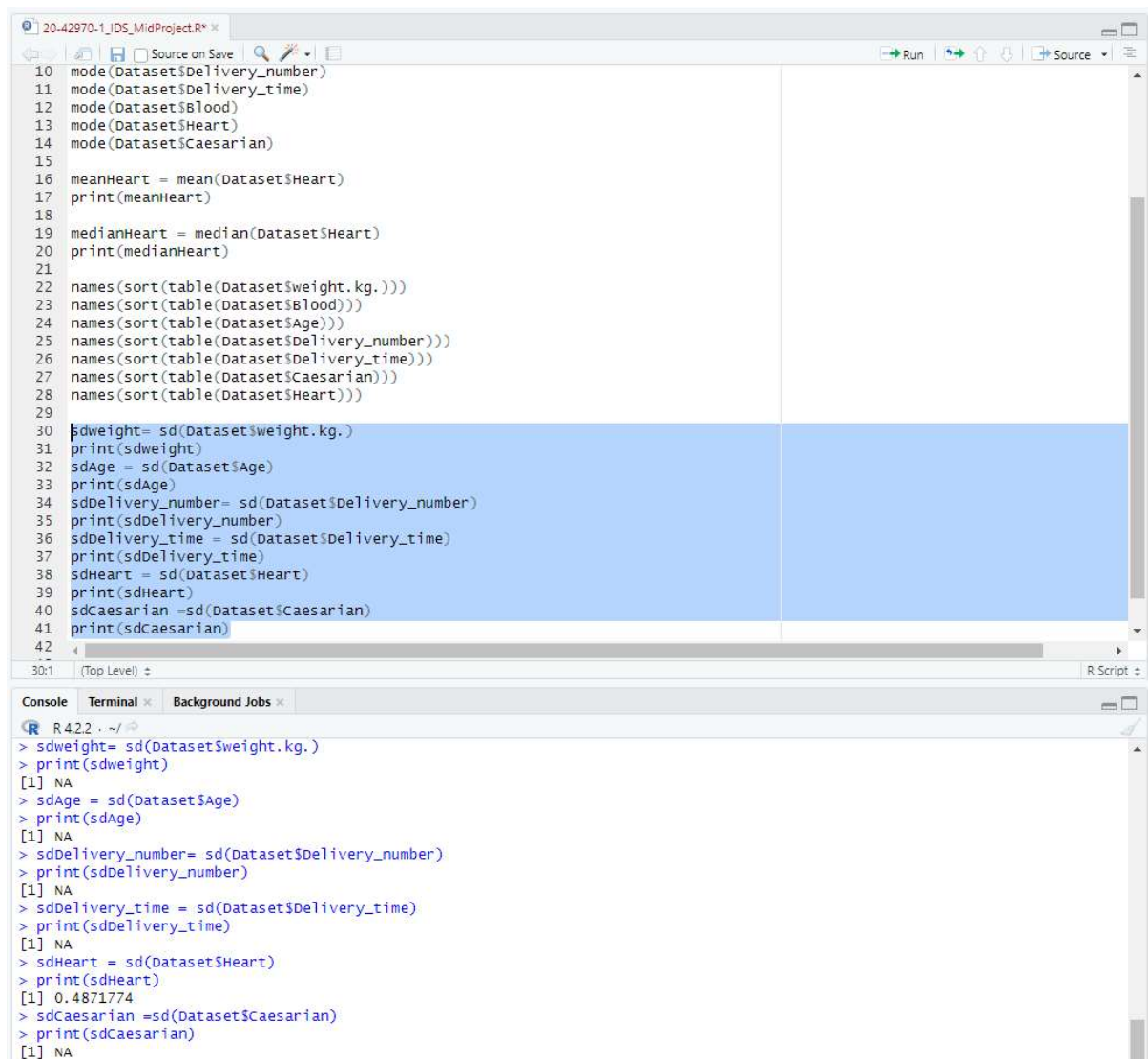
sdAge = sd(Dataset$Age)
print(sdAge)

sdDelivery_number= sd(Dataset$Delivery_number)
print(sdDelivery_number)

sdDelivery_time = sd(Dataset$Delivery_time)
print(sdDelivery_time)

sdHeart = sd(Dataset$Heart)
print(sdHeart)

sdCaesarian =sd(Dataset$Caesarian)
print(sdCaesarian)
```



The screenshot displays the R Studio interface. The script editor on the left contains R code for calculating standard deviations for various attributes of a dataset. The console at the bottom shows the execution of this code, with most standard deviation calculations resulting in NA, except for 'sdHeart' which returns 0.4871774.

```
10 mode(Dataset$Delivery_number)
11 mode(Dataset$Delivery_time)
12 mode(Dataset$Blood)
13 mode(Dataset$Heart)
14 mode(Dataset$Caesarian)
15
16 meanHeart = mean(Dataset$Heart)
17 print(meanHeart)
18
19 medianHeart = median(Dataset$Heart)
20 print(medianHeart)
21
22 names(sort(table(Dataset$weight.kg.)))
23 names(sort(table(Dataset$Blood)))
24 names(sort(table(Dataset$Age)))
25 names(sort(table(Dataset$Delivery_number)))
26 names(sort(table(Dataset$Delivery_time)))
27 names(sort(table(Dataset$Caesarian)))
28 names(sort(table(Dataset$Heart)))
29
30 sdweight= sd(Dataset$weight.kg.)
31 print(sdweight)
32 sdAge = sd(Dataset$Age)
33 print(sdAge)
34 sdDelivery_number= sd(Dataset$Delivery_number)
35 print(sdDelivery_number)
36 sdDelivery_time = sd(Dataset$Delivery_time)
37 print(sdDelivery_time)
38 sdHeart = sd(Dataset$Heart)
39 print(sdHeart)
40 sdCaesarian =sd(Dataset$Caesarian)
41 print(sdCaesarian)
42
```

Console Output:

```
> sdweight= sd(Dataset$weight.kg.)
> print(sdweight)
[1] NA
> sdAge = sd(Dataset$Age)
> print(sdAge)
[1] NA
> sdDelivery_number= sd(Dataset$Delivery_number)
> print(sdDelivery_number)
[1] NA
> sdDelivery_time = sd(Dataset$Delivery_time)
> print(sdDelivery_time)
[1] NA
> sdHeart = sd(Dataset$Heart)
> print(sdHeart)
[1] 0.4871774
> sdCaesarian =sd(Dataset$Caesarian)
> print(sdCaesarian)
[1] NA
```

10) Show the range of all attributes.

```
idRange=max(Dataset$id, na.rm=TRUE)-min(Dataset$id, na.rm=TRUE)

print(idRange)

AgeRange=max(Dataset$Age, na.rm=TRUE)-min(Dataset$Age, na.rm=TRUE)

print(AgeRange)

weightRange=max(Dataset$weight.kg., na.rm=TRUE)-min(Dataset$weight.kg., na.rm=TRUE)

print(weightRange)

Delivery_numberRange=max(Dataset$Delivery_number, na.rm=TRUE)-
min(Dataset$Delivery_number, na.rm=TRUE)

print(Delivery_numberRange)

Delivery_timeRange=max(Dataset$Delivery_time, na.rm=TRUE)-min(Dataset$Delivery_time,
na.rm=TRUE)

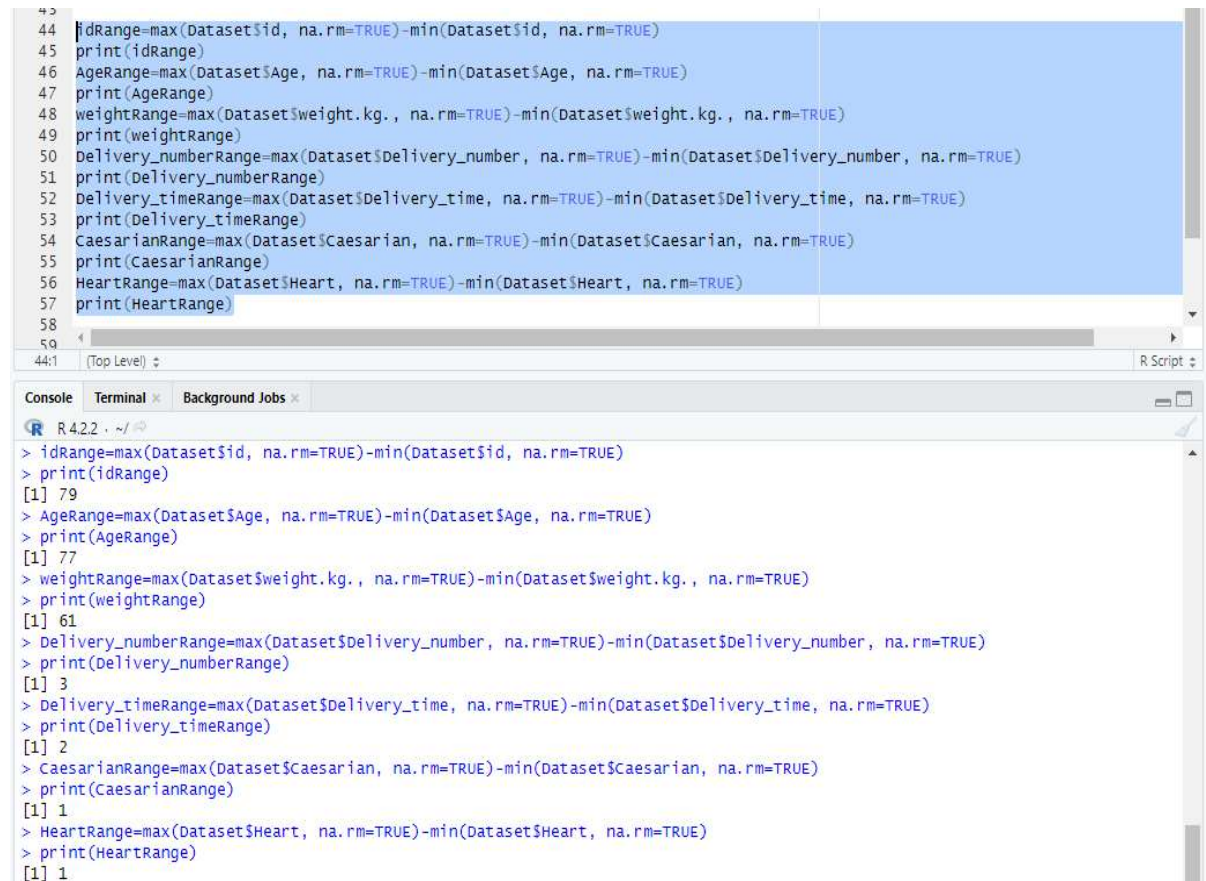
print(Delivery_timeRange)

CaesarianRange=max(Dataset$Caesarian, na.rm=TRUE)-min(Dataset$Caesarian, na.rm=TRUE)

print(CaesarianRange)

HeartRange=max(Dataset$Heart, na.rm=TRUE)-min(Dataset$Heart, na.rm=TRUE)

print(HeartRange)
```



The screenshot displays the R Studio environment. The top pane shows the R script editor with the following code:

```
44 idRange=max(Dataset$id, na.rm=TRUE)-min(Dataset$id, na.rm=TRUE)
45 print(idRange)
46 AgeRange=max(Dataset$Age, na.rm=TRUE)-min(Dataset$Age, na.rm=TRUE)
47 print(AgeRange)
48 weightRange=max(Dataset$weight.kg., na.rm=TRUE)-min(Dataset$weight.kg., na.rm=TRUE)
49 print(weightRange)
50 Delivery_numberRange=max(Dataset$Delivery_number, na.rm=TRUE)-min(Dataset$Delivery_number, na.rm=TRUE)
51 print(Delivery_numberRange)
52 Delivery_timeRange=max(Dataset$Delivery_time, na.rm=TRUE)-min(Dataset$Delivery_time, na.rm=TRUE)
53 print(Delivery_timeRange)
54 CaesarianRange=max(Dataset$Caesarian, na.rm=TRUE)-min(Dataset$Caesarian, na.rm=TRUE)
55 print(CaesarianRange)
56 HeartRange=max(Dataset$Heart, na.rm=TRUE)-min(Dataset$Heart, na.rm=TRUE)
57 print(HeartRange)
58
59
```

The bottom pane shows the Console with the following output:

```
R 4.2.2 ~ /
> idRange=max(Dataset$id, na.rm=TRUE)-min(Dataset$id, na.rm=TRUE)
> print(idRange)
[1] 79
> AgeRange=max(Dataset$Age, na.rm=TRUE)-min(Dataset$Age, na.rm=TRUE)
> print(AgeRange)
[1] 77
> weightRange=max(Dataset$weight.kg., na.rm=TRUE)-min(Dataset$weight.kg., na.rm=TRUE)
> print(weightRange)
[1] 61
> Delivery_numberRange=max(Dataset$Delivery_number, na.rm=TRUE)-min(Dataset$Delivery_number, na.rm=TRUE)
> print(Delivery_numberRange)
[1] 3
> Delivery_timeRange=max(Dataset$Delivery_time, na.rm=TRUE)-min(Dataset$Delivery_time, na.rm=TRUE)
> print(Delivery_timeRange)
[1] 2
> CaesarianRange=max(Dataset$Caesarian, na.rm=TRUE)-min(Dataset$Caesarian, na.rm=TRUE)
> print(CaesarianRange)
[1] 1
> HeartRange=max(Dataset$Heart, na.rm=TRUE)-min(Dataset$Heart, na.rm=TRUE)
> print(HeartRange)
[1] 1
```


11) Annotate high as 1, normal as 2, and low as 3 from the Blood attribute.

```
Dataset$Blood <- factor(Dataset$Blood,
  levels = c("high","normal","low"),
  labels = c(1,2,3))
```

Dataset

20-42970-1_IDS_MidProject.R x Dataset x

Source on Save

```
59 dataset$Blood <- factor(dataset$Blood,
60   levels = c("high","normal","low"),
61   labels = c(1,2,3))
62 dataset
63
```

59:1 (Top Level) ±

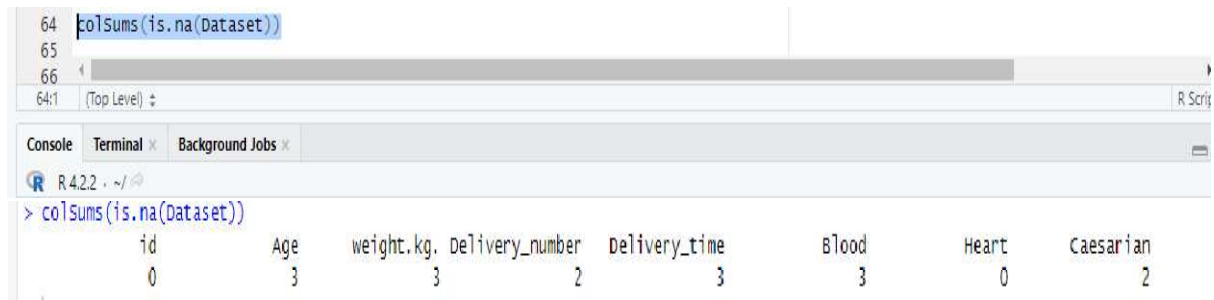
Console Terminal x Background Jobs x

R 4.2.2 · D:/Semester/10) Spring 2023/Introduction To Data Science/Project/

20	20	24	56.6	1	2	3	1	1
21	21	26	62.0	1	1	2	0	0
22	22	33	75.0	2	0	3	1	1
23	23	25	62.0	1	1	1	0	0
24	24	27	65.0	NA	NA	3	1	1
25	25	20	55.0	1	0	1	1	1
26	26	18	49.0	NA	0	2	0	0
27	27	18	50.0	1	NA	1	1	1
28	28	30	68.0	1	0	2	0	0
29	29	32	73.0	1	0	1	1	1
30	30	26	62.5	2	1	2	1	0
31	31	25	58.0	1	0	3	0	0
32	32	40	82.0	1	0	2	1	1
33	33	32	68.0	2	0	1	1	1
34	34	27	63.0	2	0	2	1	1
35	35	26	59.0	2	2	2	0	1
36	36	28	66.0	3	0	1	0	1
37	37	33	75.0	1	1	2	0	0
38	38	31	69.0	2	2	2	0	0
39	39	31	63.0	1	0	2	0	0
40	40	26	59.0	1	2	3	1	1
41	41	27	63.0	1	0	1	1	1
42	42	19	51.0	1	0	2	0	1
43	43	36	73.0	1	1	1	0	1
44	44	22	57.0	1	0	2	0	1
45	45	36	72.5	4	0	1	1	1
46	46	28	62.5	3	0	2	1	1
47	47	26	NA	1	0	2	0	0
48	48	32	67.5	2	0	1	1	1
49	49	26	62.5	2	2	2	0	0
50	50	NA	NA	2	0	3	1	1
51	51	33	68.5	3	2	2	1	0
52	52	21	53.0	2	1	3	1	1
53	53	30	68.0	3	2	1	0	0
54	54	35	74.0	1	1	3	0	0
55	55	29	63.5	2	0	2	1	1
56	56	25	59.0	2	0	2	0	0
57	57	32	67.5	3	1	3	1	1
58	58	95	110.0	1	0	3	0	1
59	59	26	61.5	1	0	1	0	1
60	60	30	67.5	2	1	1	1	NA
61	61	22	58.5	1	2	1	0	0
62	62	NA	NA	1	0	2	0	1
63	63	32	67.0	2	0	3	0	1
64	64	32	67.0	2	0	2	1	1
65	65	31	66.0	1	2	1	1	0
66	66	35	72.0	2	0	2	0	1

12) Find the Missing Value for All Attributes.

```
colSums(is.na(Dataset))
```



The screenshot shows an R Studio window with the command `colSums(is.na(Dataset))` entered in the console. The output is a table showing the number of missing values for each column in the Dataset.

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian
	0	3	3	2	3	3	0	2

13) Find the Specific Row Number of Null Values of Age, Weight, Delivery Number, Delivery Time and Caesarian.

```
which(is.na(Dataset$Age))
```

```
which(is.na(Dataset$weight.kg.))
```

```
which(is.na(Dataset$Delivery_number))
```

```
which(is.na(Dataset$Delivery_time))
```

```
which(is.na(Dataset$Blood))
```

```
which(is.na(Dataset$Caesarian))
```



The screenshot shows an R Studio window with the following commands entered in the console:

```
66 which(is.na(Dataset$Age))
67 which(is.na(Dataset$weight.kg.))
68 which(is.na(Dataset$Delivery_number))
69 which(is.na(Dataset$Delivery_time))
70 which(is.na(Dataset$Blood))
71 which(is.na(Dataset$Caesarian))
72
73
```

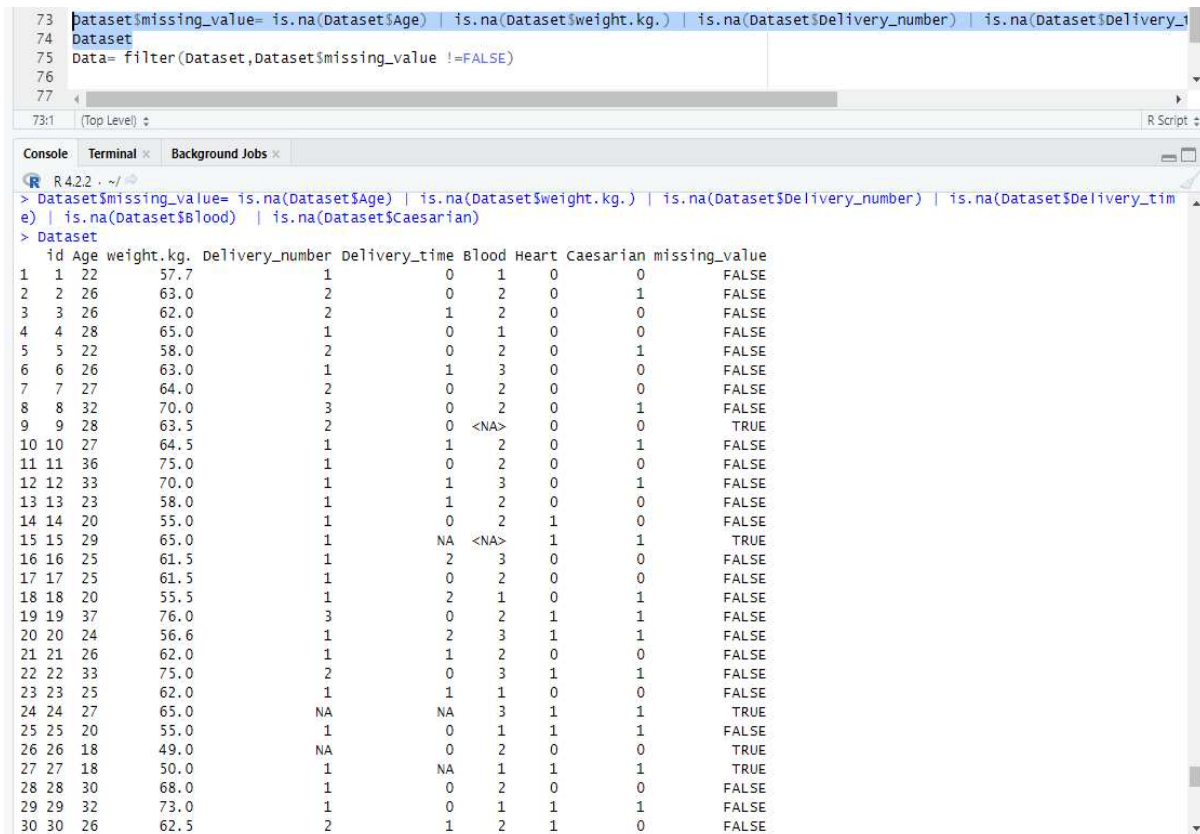
The output of these commands is shown below:

```
> which(is.na(Dataset$Age))
[1] 50 62 78
> which(is.na(Dataset$weight.kg.))
[1] 47 50 62
> which(is.na(Dataset$Delivery_number))
[1] 24 26
> which(is.na(Dataset$Delivery_time))
[1] 15 24 27
> which(is.na(Dataset$Blood))
[1] 9 15 72
> which(is.na(Dataset$Caesarian))
[1] 60 77
```

14) Detect the Outlier as a Missing Value.

```
Dataset$missing_value= is.na(Dataset$Age) | is.na(Dataset$weight.kg.) |  
is.na(Dataset$Delivery_number) | is.na(Dataset$Delivery_time) | is.na(Dataset$Blood) |  
is.na(Dataset$Caesarian)
```

Dataset



```
73 Dataset$missing_value= is.na(Dataset$Age) | is.na(Dataset$weight.kg.) | is.na(Dataset$Delivery_number) | is.na(Dataset$Delivery_t  
74 Dataset  
75 Data= filter(Dataset, Dataset$missing_value !=FALSE)  
76  
77
```

Console

```
> Dataset$missing_value= is.na(Dataset$Age) | is.na(Dataset$weight.kg.) | is.na(Dataset$Delivery_number) | is.na(Dataset$Delivery_tim  
> | is.na(Dataset$Blood) | is.na(Dataset$Caesarian)  
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1	0	1	0	0	FALSE
2	2	26	63.0	2	0	2	0	1	FALSE
3	3	26	62.0	2	1	2	0	0	FALSE
4	4	28	65.0	1	0	1	0	0	FALSE
5	5	22	58.0	2	0	2	0	1	FALSE
6	6	26	63.0	1	1	3	0	0	FALSE
7	7	27	64.0	2	0	2	0	0	FALSE
8	8	32	70.0	3	0	2	0	1	FALSE
9	9	28	63.5	2	0	<NA>	0	0	TRUE
10	10	27	64.5	1	1	2	0	1	FALSE
11	11	36	75.0	1	0	2	0	0	FALSE
12	12	33	70.0	1	1	3	0	1	FALSE
13	13	23	58.0	1	1	2	0	0	FALSE
14	14	20	55.0	1	0	2	1	0	FALSE
15	15	29	65.0	1	NA	<NA>	1	1	TRUE
16	16	25	61.5	1	2	3	0	0	FALSE
17	17	25	61.5	1	0	2	0	0	FALSE
18	18	20	55.5	1	2	1	0	1	FALSE
19	19	37	76.0	3	0	2	1	1	FALSE
20	20	24	56.6	1	2	3	1	1	FALSE
21	21	26	62.0	1	1	2	0	0	FALSE
22	22	33	75.0	2	0	3	1	1	FALSE
23	23	25	62.0	1	1	1	0	0	FALSE
24	24	27	65.0	NA	NA	3	1	1	TRUE
25	25	20	55.0	1	0	1	1	1	FALSE
26	26	18	49.0	NA	0	2	0	0	TRUE
27	27	18	50.0	1	NA	1	1	1	TRUE
28	28	30	68.0	1	0	2	0	0	FALSE
29	29	32	73.0	1	0	1	1	1	FALSE
30	30	26	62.5	2	1	2	1	0	FALSE

```
Data= filter(Dataset, Dataset$missing_value !=FALSE)
```

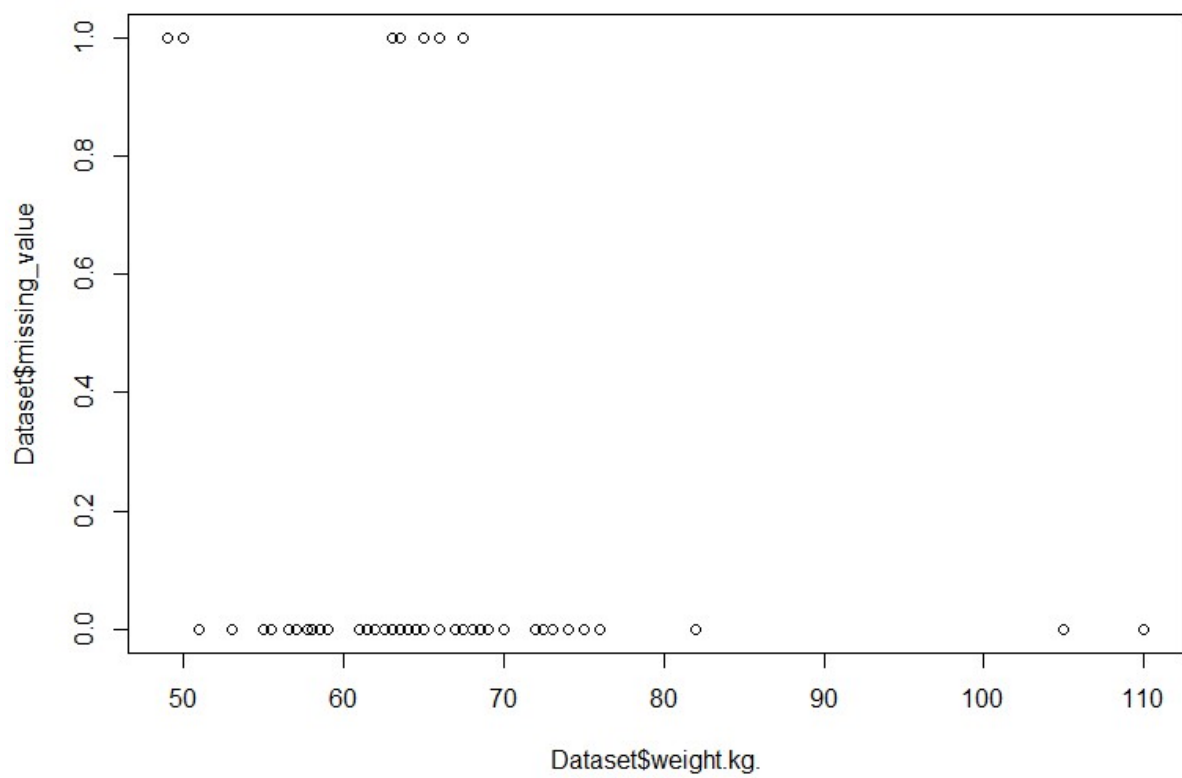
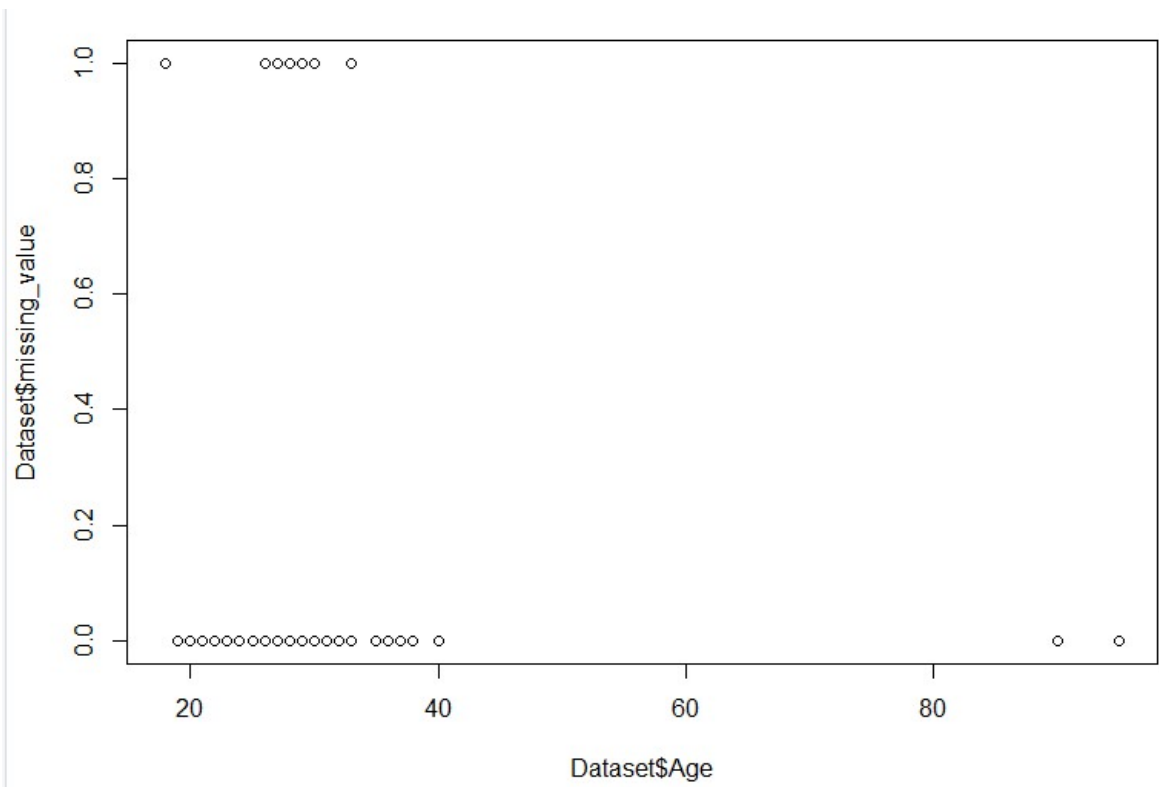
```
plot(Dataset$Age, Dataset$missing_value)
```

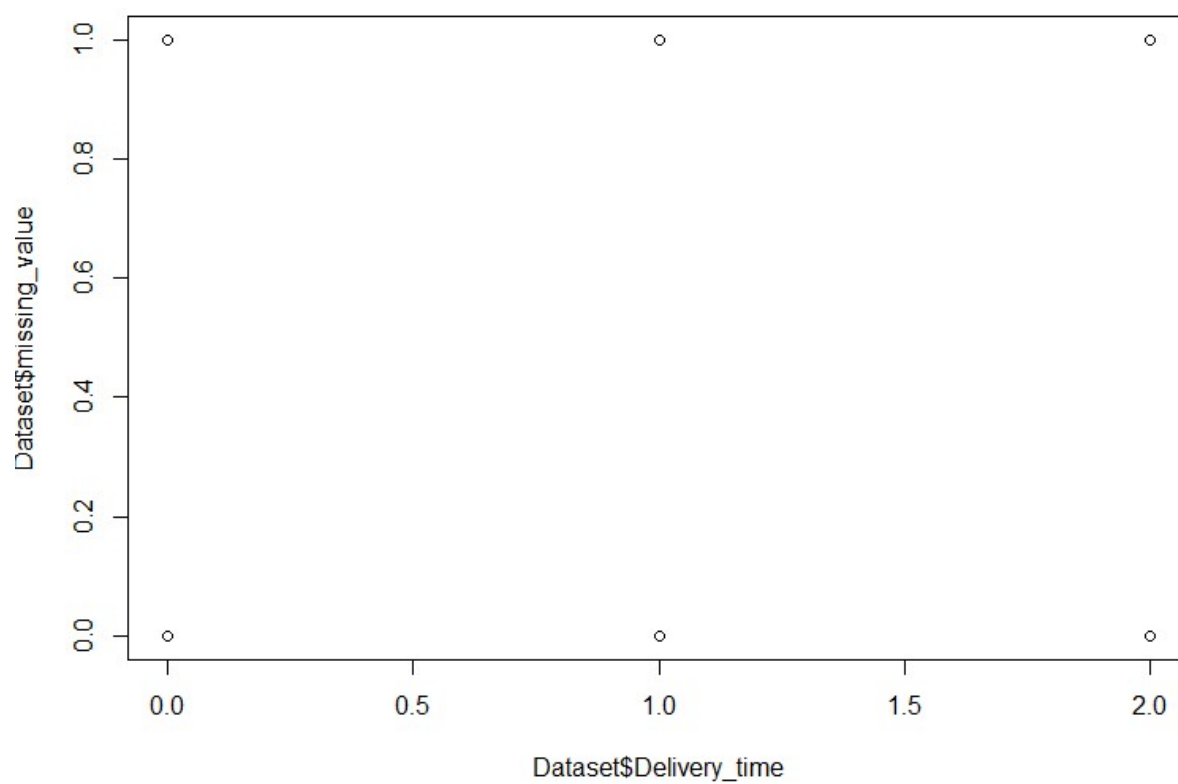
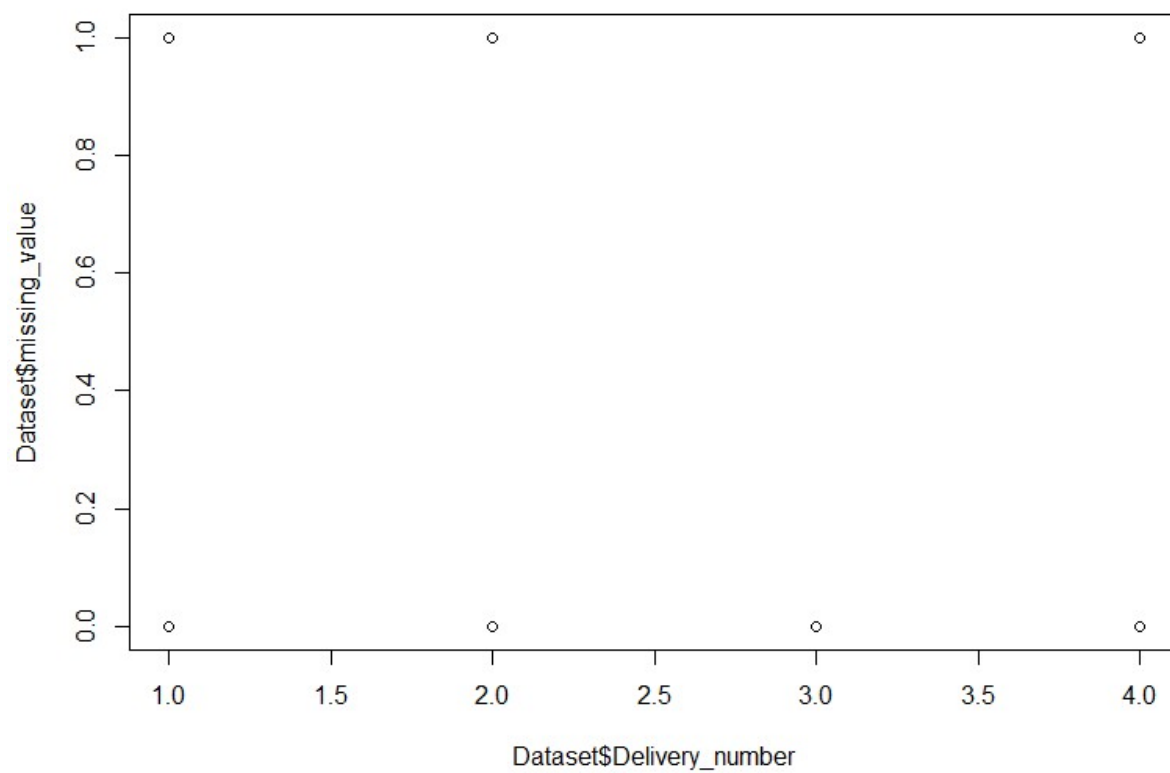
```
plot(Dataset$weight.kg., Dataset$missing_value)
```

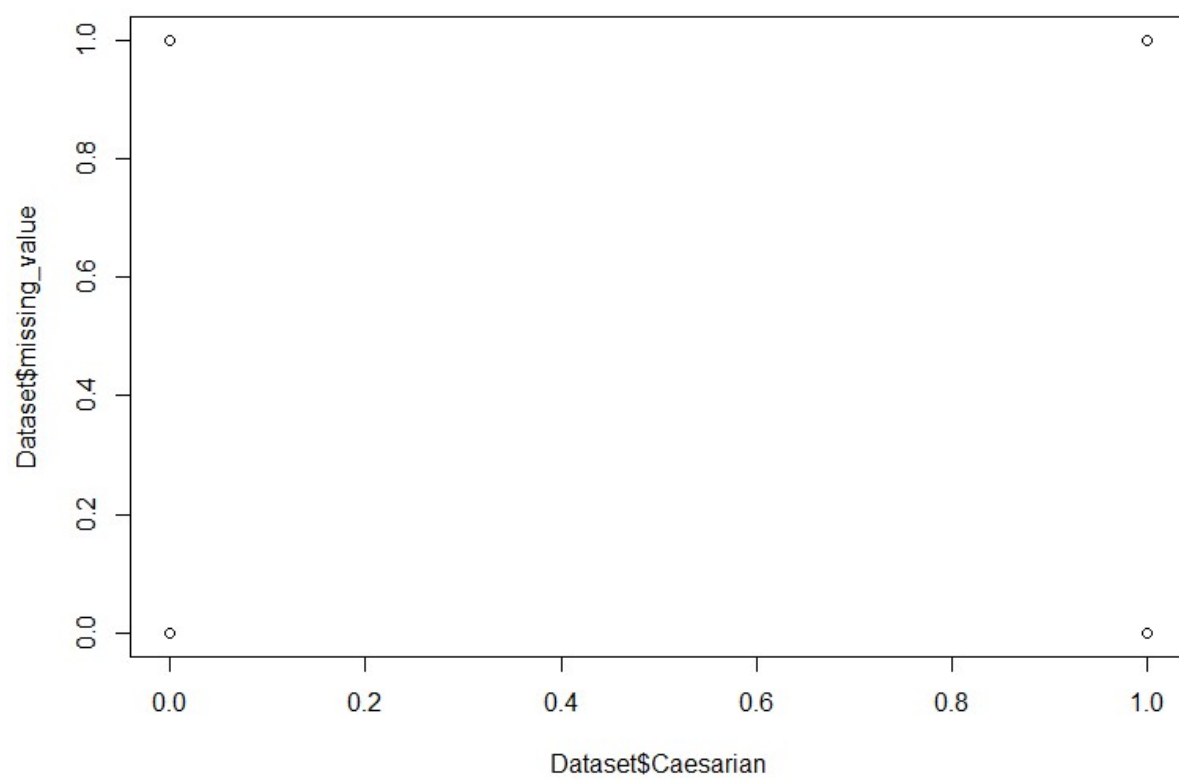
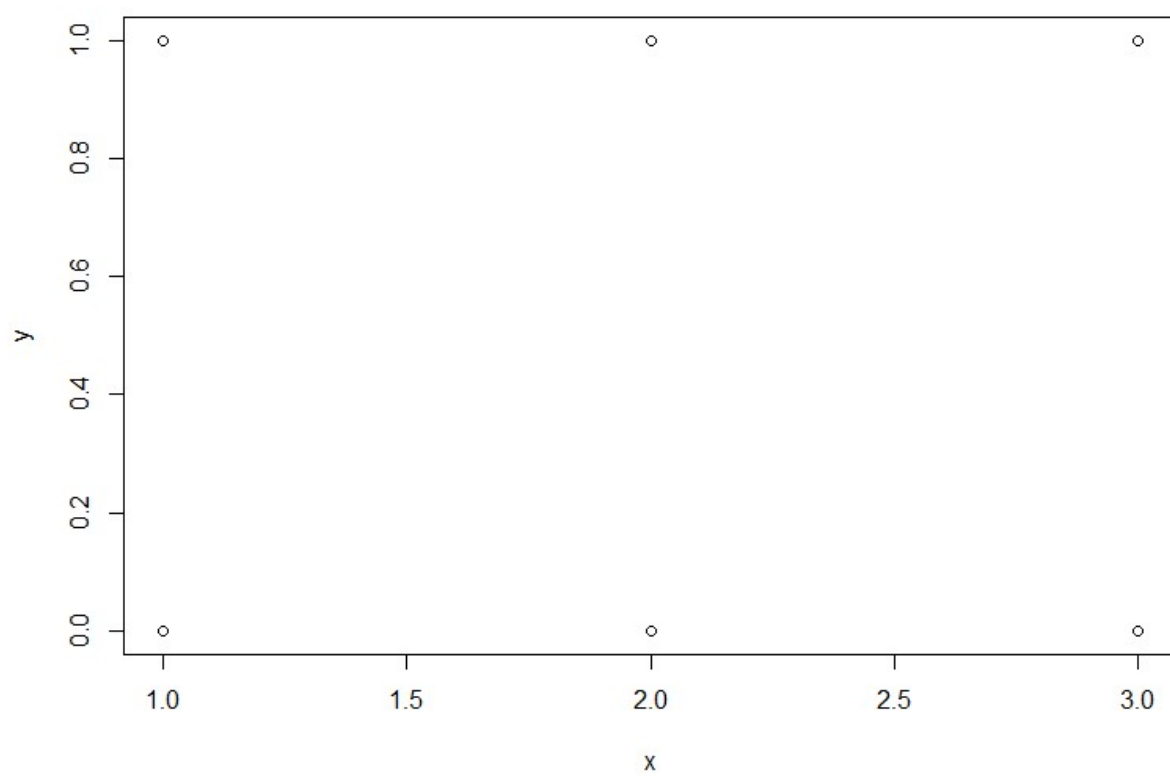
```
plot(Dataset$Delivery_number, Dataset$missing_value)
```

```
plot(Dataset$Delivery_number, Dataset$Blood)
```

```
plot(Dataset$Caesarian, Dataset$missing_value)
```







15) Recover missing values by the following strategies for Age, Weight, Delivery Number, Delivery Time, Blood, and Caesarian attributes.

I. Recover missing values with the mean value.

```
meanvalueAge <-mean(Dataset$Age,na.rm=TRUE)
```

```
meanvalueAge
```

```
Dataset[is.na(Dataset$Age), "Age"] <-meanvalueAge
```

```
Dataset
```

93	meanvalueAge <-mean(Dataset\$Age,na.rm=TRUE)	
94	meanvalueAge	
95	Dataset[is.na(Dataset\$Age), "Age"] <-meanvalueAge	
96	Dataset	
97		
98		
99		

93:1	(Top Level) ⌵
------	---------------

Console	Terminal x	Background Jobs x
---------	------------	-------------------

R 4.2.2 · ~/

```
> meanvalueAge <-mean(Dataset$Age,na.rm=TRUE)
> meanvalueAge
[1] 29.67532
> Dataset[is.na(Dataset$Age), "Age"] <-meanvalueAge
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22.00000	57.7	1	0	1	0	0	FALSE
2	2	26.00000	63.0	2	0	2	0	1	FALSE
3	3	26.00000	62.0	2	1	2	0	0	FALSE
4	4	28.00000	65.0	1	0	1	0	0	FALSE
5	5	22.00000	58.0	2	0	2	0	1	FALSE
6	6	26.00000	63.0	1	1	3	0	0	FALSE
7	7	27.00000	64.0	2	0	2	0	0	FALSE
8	8	32.00000	70.0	3	0	2	0	1	FALSE
9	9	28.00000	63.5	2	0	<NA>	0	0	TRUE
10	10	27.00000	64.5	1	1	2	0	1	FALSE
11	11	36.00000	75.0	1	0	2	0	0	FALSE
12	12	33.00000	70.0	1	1	3	0	1	FALSE
13	13	23.00000	58.0	1	1	2	0	0	FALSE
14	14	20.00000	55.0	1	0	2	1	0	FALSE
15	15	29.00000	65.0	1	NA	<NA>	1	1	TRUE
16	16	25.00000	61.5	1	2	3	0	0	FALSE
17	17	25.00000	61.5	1	0	2	0	0	FALSE
18	18	20.00000	55.5	1	2	1	0	1	FALSE
19	19	37.00000	76.0	3	0	2	1	1	FALSE
20	20	24.00000	56.6	1	2	3	1	1	FALSE
21	21	26.00000	62.0	1	1	2	0	0	FALSE
22	22	33.00000	75.0	2	0	3	1	1	FALSE
23	23	25.00000	62.0	1	1	1	0	0	FALSE
24	24	27.00000	65.0	NA	NA	3	1	1	TRUE
25	25	20.00000	55.0	1	0	1	1	1	FALSE
26	26	18.00000	49.0	NA	0	2	0	0	TRUE
27	27	18.00000	50.0	1	NA	1	1	1	TRUE
28	28	30.00000	68.0	1	0	2	0	0	FALSE

```
meanvalueWeight <-mean(Dataset$weight.kg.,na.rm=TRUE)
```

```
meanvalueWeight
```

```
Dataset[is.na(Dataset$weight.kg.), "weight.kg."] <-meanvalueWeight
```

```
Dataset
```

The screenshot shows an R Studio window with the following components:

- Source Editor:** Contains R code for installing packages, removing missing values, and calculating the mean weight. Lines 98-101 are highlighted in blue.
- Console:** Shows the execution of the code, including the calculation of meanvalueweight and the update of the Dataset.
- Table:** Displays the Dataset as a table with columns: id, Age, weight.kg., Delivery_number, Delivery_time, Blood, Heart, Caesarian, and missing_value.

```
85 install.packages("matrixStats")
86 library(matrixStats)
87 install.packages("dplyr")
88 library(dplyr)
89
90 remove_missingvalue<- na.omit(Dataset)
91 remove_missingvalue
92
93 meanvalueAge <-mean(Dataset$Age,na.rm=TRUE)
94 meanvalueAge
95 Dataset[is.na(Dataset$Age), "Age"] <-meanvalueAge
96 Dataset
97
98 meanvalueweight <-mean(Dataset$weight.kg.,na.rm=TRUE)
99 meanvalueweight
100 Dataset[is.na(Dataset$weight.kg.), "weight.kg."] <-meanvalueweight
101 Dataset
102
```

```
> meanvalueweight
[1] 65.12727
> Dataset[is.na(Dataset$weight.kg.), "weight.kg."] <-meanvalueweight
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22.00000	57.70000	1	0	1	0	0	FALSE
2	2	26.00000	63.00000	2	0	2	0	1	FALSE
3	3	26.00000	62.00000	2	1	2	0	0	FALSE
4	4	28.00000	65.00000	1	0	1	0	0	FALSE
5	5	22.00000	58.00000	2	0	2	0	1	FALSE
6	6	26.00000	63.00000	1	1	3	0	0	FALSE
7	7	27.00000	64.00000	2	0	2	0	0	FALSE
8	8	32.00000	70.00000	3	0	2	0	1	FALSE
9	9	28.00000	63.50000	2	0	<NA>	0	0	TRUE
10	10	27.00000	64.50000	1	1	2	0	1	FALSE
11	11	36.00000	75.00000	1	0	2	0	0	FALSE
12	12	33.00000	70.00000	1	1	3	0	1	FALSE
13	13	23.00000	58.00000	1	1	2	0	0	FALSE
14	14	20.00000	55.00000	1	0	2	1	0	FALSE
15	15	29.00000	65.00000	1	NA	<NA>	1	1	TRUE
16	16	25.00000	61.50000	1	2	3	0	0	FALSE
17	17	25.00000	61.50000	1	0	2	0	0	FALSE
18	18	20.00000	55.50000	1	2	1	0	1	FALSE
19	19	37.00000	76.00000	3	0	2	1	1	FALSE
20	20	24.00000	56.60000	1	2	3	1	1	FALSE
21	21	26.00000	62.00000	1	1	2	0	0	FALSE
22	22	33.00000	75.00000	2	0	3	1	1	FALSE
23	23	25.00000	62.00000	1	1	1	0	0	FALSE
24	24	27.00000	65.00000	NA	NA	3	1	1	TRUE
25	25	20.00000	55.00000	1	0	1	1	1	FALSE
26	26	18.00000	49.00000	NA	0	2	0	0	TRUE
27	27	18.00000	50.00000	1	NA	1	1	1	TRUE
28	28	30.00000	68.00000	1	0	2	0	0	FALSE
29	29	32.00000	73.00000	1	0	1	1	1	FALSE

```
meanvalueDelivery_number <-mean(Dataset$Delivery_number,na.rm=TRUE)
```

```
meanvalueDelivery_number
```

```
Dataset[is.na(Dataset$Delivery_number), "Delivery_number"] <-meanvalueDelivery_number
```

```
Dataset
```

103	meanvalueDelivery_number <-mean(Dataset\$Delivery_number,na.rm=TRUE)	
104	meanvalueDelivery_number	
105	Dataset[is.na(Dataset\$Delivery_number), "Delivery_number"] <-meanvalueDelivery_number	
106	Dataset	
107		

103:1 (Top Level) ⚡

Console	Terminal	Background Jobs
R 4.2.2 . ~/		
80 80 24.00000 57.00000 2 2 2 0 0 FALSE		
> meanvalueDelivery_number <-mean(Dataset\$Delivery_number,na.rm=TRUE)		
> meanvalueDelivery_number		
[1] 1.679487		
> Dataset[is.na(Dataset\$Delivery_number), "Delivery_number"] <-meanvalueDelivery_number		
> Dataset		
	id	Age weight.kg. Delivery_number Delivery_time Blood Heart Caesarian missing_value
1	1	22.00000 57.70000 1.000000 0 1 0 0 FALSE
2	2	26.00000 63.00000 2.000000 0 2 0 1 FALSE
3	3	26.00000 62.00000 2.000000 1 2 0 0 FALSE
4	4	28.00000 65.00000 1.000000 0 1 0 0 FALSE
5	5	22.00000 58.00000 2.000000 0 2 0 1 FALSE
6	6	26.00000 63.00000 1.000000 1 3 0 0 FALSE
7	7	27.00000 64.00000 2.000000 0 2 0 0 FALSE
8	8	32.00000 70.00000 3.000000 0 2 0 1 FALSE
9	9	28.00000 63.50000 2.000000 0 <NA> 0 0 TRUE
10	10	27.00000 64.50000 1.000000 1 2 0 1 FALSE
11	11	36.00000 75.00000 1.000000 0 2 0 0 FALSE
12	12	33.00000 70.00000 1.000000 1 3 0 1 FALSE
13	13	23.00000 58.00000 1.000000 1 2 0 0 FALSE
14	14	20.00000 55.00000 1.000000 0 2 1 0 FALSE
15	15	29.00000 65.00000 1.000000 NA <NA> 1 1 TRUE
16	16	25.00000 61.50000 1.000000 2 3 0 0 FALSE
17	17	25.00000 61.50000 1.000000 0 2 0 0 FALSE
18	18	20.00000 55.50000 1.000000 2 1 0 1 FALSE
19	19	37.00000 76.00000 3.000000 0 2 1 1 FALSE
20	20	24.00000 56.60000 1.000000 2 3 1 1 FALSE
21	21	26.00000 62.00000 1.000000 1 2 0 0 FALSE
22	22	33.00000 75.00000 2.000000 0 3 1 1 FALSE
23	23	25.00000 62.00000 1.000000 1 1 0 0 FALSE
24	24	27.00000 65.00000 1.679487 NA 3 1 1 TRUE
25	25	20.00000 55.00000 1.000000 0 1 1 1 FALSE
26	26	18.00000 49.00000 1.679487 0 2 0 0 TRUE
27	27	18.00000 50.00000 1.000000 NA 1 1 1 TRUE


```
meanvalueDelivery_time <-mean(Dataset$Delivery_time,na.rm=TRUE)
```

```
meanvalueDelivery_time
```

```
Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-meanvalueDelivery_time
```

```
Dataset
```

```
108 meanvalueDelivery_time <-mean(Dataset$Delivery_time,na.rm=TRUE)
109 meanvalueDelivery_time
110 Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-meanvalueDelivery_time
111 Dataset
112
113
```

108:1 (Top Level) ▾

Console Terminal Background Jobs

R 4.2.2 . ~/

```
> meanvalueDelivery_time
[1] 0.6233766
> Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-meanvalueDelivery_time
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22.00000	57.70000	1.000000	0.0000000	1	0	0	FALSE
2	2	26.00000	63.00000	2.000000	0.0000000	2	0	1	FALSE
3	3	26.00000	62.00000	2.000000	1.0000000	2	0	0	FALSE
4	4	28.00000	65.00000	1.000000	0.0000000	1	0	0	FALSE
5	5	22.00000	58.00000	2.000000	0.0000000	2	0	1	FALSE
6	6	26.00000	63.00000	1.000000	1.0000000	3	0	0	FALSE
7	7	27.00000	64.00000	2.000000	0.0000000	2	0	0	FALSE
8	8	32.00000	70.00000	3.000000	0.0000000	2	0	1	FALSE
9	9	28.00000	63.50000	2.000000	0.0000000	<NA>	0	0	TRUE
10	10	27.00000	64.50000	1.000000	1.0000000	2	0	1	FALSE
11	11	36.00000	75.00000	1.000000	0.0000000	2	0	0	FALSE
12	12	33.00000	70.00000	1.000000	1.0000000	3	0	1	FALSE
13	13	23.00000	58.00000	1.000000	1.0000000	2	0	0	FALSE
14	14	20.00000	55.00000	1.000000	0.0000000	2	1	0	FALSE
15	15	29.00000	65.00000	1.000000	0.6233766	<NA>	1	1	TRUE
16	16	25.00000	61.50000	1.000000	2.0000000	3	0	0	FALSE
17	17	25.00000	61.50000	1.000000	0.0000000	2	0	0	FALSE
18	18	20.00000	55.50000	1.000000	2.0000000	1	0	1	FALSE
19	19	37.00000	76.00000	3.000000	0.0000000	2	1	1	FALSE
20	20	24.00000	56.60000	1.000000	2.0000000	3	1	1	FALSE
21	21	26.00000	62.00000	1.000000	1.0000000	2	0	0	FALSE
22	22	33.00000	75.00000	2.000000	0.0000000	3	1	1	FALSE
23	23	25.00000	62.00000	1.000000	1.0000000	1	0	0	FALSE
24	24	27.00000	65.00000	1.679487	0.6233766	3	1	1	TRUE
25	25	20.00000	55.00000	1.000000	0.0000000	1	1	1	FALSE
26	26	18.00000	49.00000	1.679487	0.0000000	2	0	0	TRUE
27	27	18.00000	50.00000	1.000000	0.6233766	1	1	1	TRUE
28	28	30.00000	68.00000	1.000000	0.0000000	2	0	0	FALSE
29	29	32.00000	73.00000	1.000000	0.0000000	1	1	1	FALSE

```
meanvalueCaesarian <-mean(Dataset$Caesarian,na.rm=TRUE)
```

```
meanvalueCaesarian
```

```
Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-meanvalueCaesarian
```

```
Dataset
```


20-42970-1_IDS_MidProject.R
Dataset
Source on Save

```

112
113 meanvalueCaesarian <-mean(Dataset$Caesarian,na.rm=TRUE)
114 meanvalueCaesarian
115 Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-meanvalueCaesarian
116 Dataset
117

```

93:1 (Top Level)

Console Terminal Background Jobs

R 4.2.2 . ~/

```

> meanvalueCaesarian
[1] 0.5641026
> Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-meanvalueCaesarian
> Dataset

```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22.00000	57.70000	1.000000	0.0000000	1	0	0.0000000	FALSE
2	2	26.00000	63.00000	2.000000	0.0000000	2	0	1.0000000	FALSE
3	3	26.00000	62.00000	2.000000	1.0000000	2	0	0.0000000	FALSE
4	4	28.00000	65.00000	1.000000	0.0000000	1	0	0.0000000	FALSE
5	5	22.00000	58.00000	2.000000	0.0000000	2	0	1.0000000	FALSE
6	6	26.00000	63.00000	1.000000	1.0000000	3	0	0.0000000	FALSE
7	7	27.00000	64.00000	2.000000	0.0000000	2	0	0.0000000	FALSE
8	8	32.00000	70.00000	3.000000	0.0000000	2	0	1.0000000	FALSE
9	9	28.00000	63.50000	2.000000	0.0000000	<NA>	0	0.0000000	TRUE
10	10	27.00000	64.50000	1.000000	1.0000000	2	0	1.0000000	FALSE
11	11	36.00000	75.00000	1.000000	0.0000000	2	0	0.0000000	FALSE
12	12	33.00000	70.00000	1.000000	1.0000000	3	0	1.0000000	FALSE
13	13	23.00000	58.00000	1.000000	1.0000000	2	0	0.0000000	FALSE
14	14	20.00000	55.00000	1.000000	0.0000000	2	1	0.0000000	FALSE
15	15	29.00000	65.00000	1.000000	0.6233766	<NA>	1	1.0000000	TRUE
16	16	25.00000	61.50000	1.000000	2.0000000	3	0	0.0000000	FALSE
17	17	25.00000	61.50000	1.000000	0.0000000	2	0	0.0000000	FALSE
18	18	20.00000	55.50000	1.000000	2.0000000	1	0	1.0000000	FALSE
19	19	37.00000	76.00000	3.000000	0.0000000	2	1	1.0000000	FALSE
20	20	24.00000	56.60000	1.000000	2.0000000	3	1	1.0000000	FALSE
21	21	26.00000	62.00000	1.000000	1.0000000	2	0	0.0000000	FALSE
22	22	33.00000	75.00000	2.000000	0.0000000	3	1	1.0000000	FALSE
23	23	25.00000	62.00000	1.000000	1.0000000	1	0	0.0000000	FALSE
24	24	27.00000	65.00000	1.679487	0.6233766	3	1	1.0000000	TRUE
25	25	20.00000	55.00000	1.000000	0.0000000	1	1	1.0000000	FALSE
26	26	18.00000	49.00000	1.679487	0.0000000	2	0	0.0000000	TRUE
27	27	18.00000	50.00000	1.000000	0.6233766	1	1	1.0000000	TRUE
28	28	30.00000	68.00000	1.000000	0.0000000	2	0	0.0000000	FALSE
29	29	32.00000	73.00000	1.000000	0.0000000	1	1	1.0000000	FALSE
30	30	26.00000	62.50000	2.000000	1.0000000	2	1	0.0000000	FALSE
31	31	25.00000	58.00000	1.000000	0.0000000	3	0	0.0000000	FALSE
32	32	40.00000	82.00000	1.000000	0.0000000	2	1	1.0000000	FALSE
33	33	32.00000	68.00000	2.000000	0.0000000	1	1	1.0000000	FALSE
34	34	27.00000	63.00000	2.000000	0.0000000	2	1	1.0000000	FALSE
35	35	26.00000	59.00000	2.000000	2.0000000	2	0	1.0000000	FALSE
36	36	28.00000	66.00000	3.000000	0.0000000	1	0	1.0000000	FALSE
37	37	33.00000	75.00000	1.000000	1.0000000	2	0	0.0000000	FALSE
38	38	31.00000	69.00000	2.000000	2.0000000	2	0	0.0000000	FALSE
39	39	31.00000	63.00000	1.000000	0.0000000	2	0	0.0000000	FALSE
40	40	26.00000	59.00000	1.000000	2.0000000	3	1	1.0000000	FALSE
41	41	27.00000	63.00000	1.000000	0.0000000	1	1	1.0000000	FALSE
42	42	19.00000	51.00000	1.000000	0.0000000	2	0	1.0000000	FALSE

43	43	36.00000	73.00000	1.000000	1.0000000	1	0	1.0000000	FALSE
44	44	22.00000	57.00000	1.000000	0.0000000	2	0	1.0000000	FALSE
45	45	36.00000	72.50000	4.000000	0.0000000	1	1	1.0000000	FALSE
46	46	28.00000	62.50000	3.000000	0.0000000	2	1	1.0000000	FALSE
47	47	26.00000	65.12727	1.000000	0.0000000	2	0	0.0000000	TRUE
48	48	32.00000	67.50000	2.000000	0.0000000	1	1	1.0000000	FALSE
49	49	26.00000	62.50000	2.000000	2.0000000	2	0	0.0000000	FALSE
50	50	29.67532	65.12727	2.000000	0.0000000	3	1	1.0000000	TRUE
51	51	33.00000	68.50000	3.000000	2.0000000	2	1	0.0000000	FALSE
52	52	21.00000	53.00000	2.000000	1.0000000	3	1	1.0000000	FALSE
53	53	30.00000	68.00000	3.000000	2.0000000	1	0	0.0000000	FALSE
54	54	35.00000	74.00000	1.000000	1.0000000	3	0	0.0000000	FALSE
55	55	29.00000	63.50000	2.000000	0.0000000	2	1	1.0000000	FALSE
56	56	25.00000	59.00000	2.000000	0.0000000	2	0	0.0000000	FALSE
57	57	32.00000	67.50000	3.000000	1.0000000	3	1	1.0000000	FALSE
58	58	95.00000	110.00000	1.000000	0.0000000	3	0	1.0000000	FALSE
59	59	26.00000	61.50000	1.000000	0.0000000	1	0	1.0000000	FALSE
60	60	30.00000	67.50000	2.000000	1.0000000	1	1	0.5641026	TRUE
61	61	22.00000	58.50000	1.000000	2.0000000	1	0	0.0000000	FALSE
62	62	29.67532	65.12727	1.000000	0.0000000	2	0	1.0000000	TRUE
63	63	32.00000	67.00000	2.000000	0.0000000	3	0	1.0000000	FALSE
64	64	32.00000	67.00000	2.000000	0.0000000	2	1	1.0000000	FALSE
65	65	31.00000	66.00000	1.000000	2.0000000	1	1	0.0000000	FALSE
66	66	35.00000	72.00000	2.000000	0.0000000	2	0	1.0000000	FALSE
67	67	28.00000	62.50000	3.000000	0.0000000	2	0	1.0000000	FALSE
68	68	29.00000	64.50000	2.000000	0.0000000	2	1	0.0000000	FALSE
69	69	25.00000	62.00000	1.000000	0.0000000	3	0	1.0000000	FALSE
70	70	27.00000	61.00000	2.000000	2.0000000	3	0	0.0000000	FALSE
71	71	90.00000	105.00000	1.000000	0.0000000	3	0	1.0000000	FALSE
72	72	29.00000	65.00000	1.000000	2.0000000	<NA>	1	1.0000000	TRUE
73	73	28.00000	64.00000	2.000000	0.0000000	2	0	0.0000000	FALSE
74	74	32.00000	69.00000	3.000000	0.0000000	2	1	0.0000000	FALSE
75	75	38.00000	75.00000	3.000000	2.0000000	1	1	1.0000000	FALSE
76	76	27.00000	62.50000	2.000000	1.0000000	2	0	0.0000000	FALSE
77	77	33.00000	66.00000	4.000000	0.0000000	2	0	0.5641026	TRUE
78	78	29.67532	63.00000	2.000000	1.0000000	1	0	1.0000000	TRUE
79	79	25.00000	58.00000	1.000000	2.0000000	3	0	1.0000000	FALSE
80	80	24.00000	57.00000	2.000000	2.0000000	2	0	0.0000000	FALSE

II. Recover missing values with the median value.

```
medianvalueAge <-median(Dataset$Age,na.rm=TRUE)
```

```
medianvalueAge
```

```
Dataset[is.na(Dataset$Age), "Age"] <-medianvalueAge
```

```
Dataset
```

```
120 medianvalueAge <-median(Dataset$Age,na.rm=TRUE)
121 medianvalueAge
122 Dataset[is.na(Dataset$Age), "Age"] <-medianvalueAge
123 Dataset
```

123:8 (Top Level) ⚙

Console	Terminal x	Background Jobs x
R 4.2.2 . ~/		
80 80 24 57.0 2 2 2 0 0 FALSE		
>		
> Data= filter(Dataset, Dataset\$missing_value !=FALSE)		
> medianvalueAge <-median(Dataset\$Age,na.rm=TRUE)		
> medianvalueAge		
[1] 28		
> Dataset[is.na(Dataset\$Age), "Age"] <-medianvalueAge		
> Dataset		
	id	Age weight.kg. Delivery_number Delivery_time Blood Heart Caesarian missing_value
1	1	22 57.7 1 0 1 0 0 FALSE
2	2	26 63.0 2 0 2 0 1 FALSE
3	3	26 62.0 2 1 2 0 0 FALSE
4	4	28 65.0 1 0 1 0 0 FALSE
5	5	22 58.0 2 0 2 0 1 FALSE
6	6	26 63.0 1 1 3 0 0 FALSE
7	7	27 64.0 2 0 2 0 0 FALSE
8	8	32 70.0 3 0 2 0 1 FALSE
9	9	28 63.5 2 0 <NA> 0 0 TRUE
10	10	27 64.5 1 1 2 0 1 FALSE
11	11	36 75.0 1 0 2 0 0 FALSE
12	12	33 70.0 1 1 3 0 1 FALSE
13	13	23 58.0 1 1 2 0 0 FALSE
14	14	20 55.0 1 0 2 1 0 FALSE
15	15	29 65.0 1 NA <NA> 1 1 TRUE
16	16	25 61.5 1 2 3 0 0 FALSE
17	17	25 61.5 1 0 2 0 0 FALSE
18	18	20 55.5 1 2 1 0 1 FALSE
19	19	37 76.0 3 0 2 1 1 FALSE
20	20	24 56.6 1 2 3 1 1 FALSE
21	21	26 62.0 1 1 2 0 0 FALSE
22	22	33 75.0 2 0 3 1 1 FALSE
23	23	25 62.0 1 1 1 0 0 FALSE
24	24	27 65.0 NA NA 3 1 1 TRUE
25	25	20 55.0 1 0 1 1 1 FALSE

```
medianvalueweight <-median(Dataset$weight.kg.,na.rm=TRUE)
```

```
medianvalueweight
```

```
Dataset[is.na(Dataset$weight.kg.), "weight.kg."] <-medianvalueweight
```

```
Dataset
```

125	medianvalueweight <-median(Dataset\$weight.kg.,na.rm=TRUE)
126	medianvalueweight
127	Dataset[is.na(Dataset\$weight.kg.), "weight.kg."] <-medianvalueweight
128	Dataset

125:1 (Top Level) ⚙

Console	Terminal ×	Background Jobs ×
R 4.2.2 . ~/		
78	78	28 63.0 2 1 1 0 1 TRUE
79	79	25 58.0 1 2 3 0 1 FALSE
80	80	24 57.0 2 2 2 0 0 FALSE
> medianvalueweight <-median(Dataset\$weight.kg.,na.rm=TRUE)		
> medianvalueweight		
[1] 63.5		
> Dataset[is.na(Dataset\$weight.kg.), "weight.kg."] <-medianvalueweight		
> Dataset		
	id	Age weight.kg. Delivery_number Delivery_time Blood Heart Caesarian missing_value
1	1	22 57.7 1 0 1 0 0 FALSE
2	2	26 63.0 2 0 2 0 1 FALSE
3	3	26 62.0 2 1 2 0 0 FALSE
4	4	28 65.0 1 0 1 0 0 FALSE
5	5	22 58.0 2 0 2 0 1 FALSE
6	6	26 63.0 1 1 3 0 0 FALSE
7	7	27 64.0 2 0 2 0 0 FALSE
8	8	32 70.0 3 0 2 0 1 FALSE
9	9	28 63.5 2 0 <NA> 0 0 TRUE
10	10	27 64.5 1 1 2 0 1 FALSE
11	11	36 75.0 1 0 2 0 0 FALSE
12	12	33 70.0 1 1 3 0 1 FALSE
13	13	23 58.0 1 1 2 0 0 FALSE
14	14	20 55.0 1 0 2 1 0 FALSE
15	15	29 65.0 1 NA <NA> 1 1 TRUE
16	16	25 61.5 1 2 3 0 0 FALSE
17	17	25 61.5 1 0 2 0 0 FALSE
18	18	20 55.5 1 2 1 0 1 FALSE
19	19	37 76.0 3 0 2 1 1 FALSE
20	20	24 56.6 1 2 3 1 1 FALSE
21	21	26 62.0 1 1 2 0 0 FALSE
22	22	33 75.0 2 0 3 1 1 FALSE
23	23	25 62.0 1 1 1 0 0 FALSE
24	24	27 65.0 NA NA 3 1 1 TRUE
25	25	20 55.0 1 0 1 1 1 FALSE

```
medianvalueDelivery_number <-median(Dataset$Delivery_number,na.rm=TRUE)
```

```
medianvalueDelivery_number
```

```
Dataset[is.na(Dataset$Delivery_number), "Delivery_number"] <-medianvalueDelivery_number
```

```
Dataset
```

```
130 medianvalueDelivery_number <-median(Dataset$Delivery_number,na.rm=TRUE)
131 medianvalueDelivery_number
132 Dataset[is.na(Dataset$Delivery_number), "Delivery_number"] <-medianvalueDelivery_number
133 Dataset
```

130:1 (Top Level) ↕

Console Terminal × Background Jobs ×

R 4.2.2 . ~/

```
80 80 24 57.0 2 2 2 0 0 FALSE
> medianvalueDelivery_number <-median(Dataset$Delivery_number,na.rm=TRUE)
> medianvalueDelivery_number
[1] 1.5
> Dataset[is.na(Dataset$Delivery_number), "Delivery_number"] <-medianvalueDelivery_number
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1.0	0	1	0	0	FALSE
2	2	26	63.0	2.0	0	2	0	1	FALSE
3	3	26	62.0	2.0	1	2	0	0	FALSE
4	4	28	65.0	1.0	0	1	0	0	FALSE
5	5	22	58.0	2.0	0	2	0	1	FALSE
6	6	26	63.0	1.0	1	3	0	0	FALSE
7	7	27	64.0	2.0	0	2	0	0	FALSE
8	8	32	70.0	3.0	0	2	0	1	FALSE
9	9	28	63.5	2.0	0	<NA>	0	0	TRUE
10	10	27	64.5	1.0	1	2	0	1	FALSE
11	11	36	75.0	1.0	0	2	0	0	FALSE
12	12	33	70.0	1.0	1	3	0	1	FALSE
13	13	23	58.0	1.0	1	2	0	0	FALSE
14	14	20	55.0	1.0	0	2	1	0	FALSE
15	15	29	65.0	1.0	NA	<NA>	1	1	TRUE
16	16	25	61.5	1.0	2	3	0	0	FALSE
17	17	25	61.5	1.0	0	2	0	0	FALSE
18	18	20	55.5	1.0	2	1	0	1	FALSE
19	19	37	76.0	3.0	0	2	1	1	FALSE
20	20	24	56.6	1.0	2	3	1	1	FALSE
21	21	26	62.0	1.0	1	2	0	0	FALSE
22	22	33	75.0	2.0	0	3	1	1	FALSE
23	23	25	62.0	1.0	1	1	0	0	FALSE
24	24	27	65.0	1.5	NA	3	1	1	TRUE
25	25	20	55.0	1.0	0	1	1	1	FALSE
26	26	18	49.0	1.5	0	2	0	0	TRUE
27	27	18	50.0	1.0	NA	1	1	1	TRUE


```
medianvalueDelivery_time <-median(Dataset$Delivery_time,na.rm=TRUE)
```

```
medianvalueDelivery_time
```

```
Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-medianvalueDelivery_time
```

```
Dataset
```

```
135 medianvalueDelivery_time <-median(Dataset$Delivery_time,na.rm=TRUE)
136 medianvalueDelivery_time
137 Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-medianvalueDelivery_time
138 Dataset
```

135:1 (Top Level) ⚡

Console Terminal Background Jobs

R 4.2.2 ~ /

```
> medianvalueDelivery_time <-median(Dataset$Delivery_time,na.rm=TRUE)
> medianvalueDelivery_time
[1] 0
> Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-medianvalueDelivery_time
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1.0	0	1	0	0	FALSE
2	2	26	63.0	2.0	0	2	0	1	FALSE
3	3	26	62.0	2.0	1	2	0	0	FALSE
4	4	28	65.0	1.0	0	1	0	0	FALSE
5	5	22	58.0	2.0	0	2	0	1	FALSE
6	6	26	63.0	1.0	1	3	0	0	FALSE
7	7	27	64.0	2.0	0	2	0	0	FALSE
8	8	32	70.0	3.0	0	2	0	1	FALSE
9	9	28	63.5	2.0	0	<NA>	0	0	TRUE
10	10	27	64.5	1.0	1	2	0	1	FALSE
11	11	36	75.0	1.0	0	2	0	0	FALSE
12	12	33	70.0	1.0	1	3	0	1	FALSE
13	13	23	58.0	1.0	1	2	0	0	FALSE
14	14	20	55.0	1.0	0	2	1	0	FALSE
15	15	29	65.0	1.0	0	<NA>	1	1	TRUE
16	16	25	61.5	1.0	2	3	0	0	FALSE
17	17	25	61.5	1.0	0	2	0	0	FALSE
18	18	20	55.5	1.0	2	1	0	1	FALSE
19	19	37	76.0	3.0	0	2	1	1	FALSE
20	20	24	56.6	1.0	2	3	1	1	FALSE
21	21	26	62.0	1.0	1	2	0	0	FALSE
22	22	33	75.0	2.0	0	3	1	1	FALSE
23	23	25	62.0	1.0	1	1	0	0	FALSE
24	24	27	65.0	1.5	0	3	1	1	TRUE
25	25	20	55.0	1.0	0	1	1	1	FALSE
26	26	18	49.0	1.5	0	2	0	0	TRUE
27	27	18	50.0	1.0	0	1	1	1	TRUE

```
medianvalueCaesarian <-median(Dataset$Caesarian,na.rm=TRUE)
```

```
medianvalueCaesarian
```

```
Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-medianvalueCaesarian
```

```
Dataset
```

```

142 medianvalueCaesarian <-median(Dataset$Caesarian,na.rm=TRUE)
143 medianvalueCaesarian
144 Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-medianvalueCaesarian
145 Dataset
146

```

142:1 (Top Level) ↕

Console Terminal × Background Jobs ×

R 4.2.2 · D:/Semester/10) Spring 2023/Introduction To Data Science/Project/ ↗

```
> medianvalueCaesarian <-median(Dataset$Caesarian,na.rm=TRUE)
```

```
> medianvalueCaesarian
```

```
[1] 1
```

```
> Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-medianvalueCaesarian
```

```
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1.0	0	1	0	0	FALSE
2	2	26	63.0	2.0	0	2	0	1	FALSE
3	3	26	62.0	2.0	1	2	0	0	FALSE
4	4	28	65.0	1.0	0	1	0	0	FALSE
5	5	22	58.0	2.0	0	2	0	1	FALSE
6	6	26	63.0	1.0	1	3	0	0	FALSE
7	7	27	64.0	2.0	0	2	0	0	FALSE
8	8	32	70.0	3.0	0	2	0	1	FALSE
9	9	28	63.5	2.0	0	<NA>	0	0	TRUE
10	10	27	64.5	1.0	1	2	0	1	FALSE
11	11	36	75.0	1.0	0	2	0	0	FALSE
12	12	33	70.0	1.0	1	3	0	1	FALSE
13	13	23	58.0	1.0	1	2	0	0	FALSE
14	14	20	55.0	1.0	0	2	1	0	FALSE
15	15	29	65.0	1.0	0	<NA>	1	1	TRUE
16	16	25	61.5	1.0	2	3	0	0	FALSE
17	17	25	61.5	1.0	0	2	0	0	FALSE
18	18	20	55.5	1.0	2	1	0	1	FALSE
19	19	37	76.0	3.0	0	2	1	1	FALSE
20	20	24	56.6	1.0	2	3	1	1	FALSE
21	21	26	62.0	1.0	1	2	0	0	FALSE
22	22	33	75.0	2.0	0	3	1	1	FALSE
23	23	25	62.0	1.0	1	1	0	0	FALSE
24	24	27	65.0	1.5	0	3	1	1	TRUE
25	25	20	55.0	1.0	0	1	1	1	FALSE
26	26	18	49.0	1.5	0	2	0	0	TRUE
27	27	18	50.0	1.0	0	1	1	1	TRUE
28	28	30	68.0	1.0	0	2	0	0	FALSE
29	29	32	73.0	1.0	0	1	1	1	FALSE
30	30	26	62.5	2.0	1	2	1	0	FALSE
31	31	25	58.0	1.0	0	3	0	0	FALSE
32	32	40	82.0	1.0	0	2	1	1	FALSE
33	33	32	68.0	2.0	0	1	1	1	FALSE
34	34	27	63.0	2.0	0	2	1	1	FALSE
35	35	26	59.0	2.0	2	2	0	1	FALSE
36	36	28	66.0	3.0	0	1	0	1	FALSE
37	37	33	75.0	1.0	1	2	0	0	FALSE
38	38	31	69.0	2.0	2	2	0	0	FALSE
39	39	31	63.0	1.0	0	2	0	0	FALSE
40	40	26	59.0	1.0	2	3	1	1	FALSE
41	41	27	63.0	1.0	0	1	1	1	FALSE

42	42	19	51.0	1.0	0	2	0	1	FALSE
43	43	36	73.0	1.0	1	1	0	1	FALSE
44	44	22	57.0	1.0	0	2	0	1	FALSE
45	45	36	72.5	4.0	0	1	1	1	FALSE
46	46	28	62.5	3.0	0	2	1	1	FALSE
47	47	26	63.5	1.0	0	2	0	0	TRUE
48	48	32	67.5	2.0	0	1	1	1	FALSE
49	49	26	62.5	2.0	2	2	0	0	FALSE
50	50	28	63.5	2.0	0	3	1	1	TRUE
51	51	33	68.5	3.0	2	2	1	0	FALSE
52	52	21	53.0	2.0	1	3	1	1	FALSE
53	53	30	68.0	3.0	2	1	0	0	FALSE
54	54	35	74.0	1.0	1	3	0	0	FALSE
55	55	29	63.5	2.0	0	2	1	1	FALSE
56	56	25	59.0	2.0	0	2	0	0	FALSE
57	57	32	67.5	3.0	1	3	1	1	FALSE
58	58	95	110.0	1.0	0	3	0	1	FALSE
59	59	26	61.5	1.0	0	1	0	1	FALSE
60	60	30	67.5	2.0	1	1	1	1	TRUE
61	61	22	58.5	1.0	2	1	0	0	FALSE
62	62	28	63.5	1.0	0	2	0	1	TRUE
63	63	32	67.0	2.0	0	3	0	1	FALSE
64	64	32	67.0	2.0	0	2	1	1	FALSE
65	65	31	66.0	1.0	2	1	1	0	FALSE
66	66	35	72.0	2.0	0	2	0	1	FALSE
67	67	28	62.5	3.0	0	2	0	1	FALSE
68	68	29	64.5	2.0	0	2	1	0	FALSE
69	69	25	62.0	1.0	0	3	0	1	FALSE
70	70	27	61.0	2.0	2	3	0	0	FALSE
71	71	90	105.0	1.0	0	3	0	1	FALSE
72	72	29	65.0	1.0	2	<NA>	1	1	TRUE
73	73	28	64.0	2.0	0	2	0	0	FALSE
74	74	32	69.0	3.0	0	2	1	0	FALSE
75	75	38	75.0	3.0	2	1	1	1	FALSE
76	76	27	62.5	2.0	1	2	0	0	FALSE
77	77	33	66.0	4.0	0	2	0	1	TRUE
78	78	28	63.0	2.0	1	1	0	1	TRUE
79	79	25	58.0	1.0	2	3	0	1	FALSE
80	80	24	57.0	2.0	2	2	0	0	FALSE

III. Recover missing values with the mode value.

```
modeAge=names(sort(table(Dataset$Age)))[1]
```

```
modeAge
```

```
Dataset[is.na(Dataset$Age), "Age"] <-modeAge
```

```
Dataset
```

```
145 modeAge=names(sort(table(Dataset$Age)))[1]
146 modeAge
147 Dataset[is.na(Dataset$Age), "Age"] <-modeAge
148 Dataset
149
150
151
```

145:1 (Top Level) ↕

Console Terminal x Background Jobs x

R 4.2.2 · ~/

```
> modeAge
[1] "19"
> Dataset[is.na(Dataset$Age), "Age"] <-modeAge
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1	0	1	0	0	FALSE
2	2	26	63.0	2	0	2	0	1	FALSE
3	3	26	62.0	2	1	2	0	0	FALSE
4	4	28	65.0	1	0	1	0	0	FALSE
5	5	22	58.0	2	0	2	0	1	FALSE
6	6	26	63.0	1	1	3	0	0	FALSE
7	7	27	64.0	2	0	2	0	0	FALSE
8	8	32	70.0	3	0	2	0	1	FALSE
9	9	28	63.5	2	0	<NA>	0	0	TRUE
10	10	27	64.5	1	1	2	0	1	FALSE
11	11	36	75.0	1	0	2	0	0	FALSE
12	12	33	70.0	1	1	3	0	1	FALSE
13	13	23	58.0	1	1	2	0	0	FALSE
14	14	20	55.0	1	0	2	1	0	FALSE
15	15	29	65.0	1	NA	<NA>	1	1	TRUE
16	16	25	61.5	1	2	3	0	0	FALSE
17	17	25	61.5	1	0	2	0	0	FALSE
18	18	20	55.5	1	2	1	0	1	FALSE
19	19	37	76.0	3	0	2	1	1	FALSE
20	20	24	56.6	1	2	3	1	1	FALSE
21	21	26	62.0	1	1	2	0	0	FALSE
22	22	33	75.0	2	0	3	1	1	FALSE
23	23	25	62.0	1	1	1	0	0	FALSE
24	24	27	65.0	NA	NA	3	1	1	TRUE
25	25	20	55.0	1	0	1	1	1	FALSE
26	26	18	49.0	NA	0	2	0	0	TRUE
27	27	18	50.0	1	NA	1	1	1	TRUE
28	28	30	68.0	1	0	2	0	0	FALSE
29	29	32	73.0	1	0	1	1	1	FALSE


```
modeweight=names(sort(table(Dataset$weight.kg.)))[1]
```

```
modeweight
```

```
Dataset[is.na(Dataset$weight.kg.), "weight.kg."] <-modeweight
```

```
Dataset
```

```
152 modeweight=names(sort(table(Dataset$weight.kg.)))[1]
153 modeweight
154 Dataset[is.na(Dataset$weight.kg.), "weight.kg."] <-modeweight
155 Dataset
```

152:1 (Top Level) ↕

Console Terminal Background Jobs

R 4.2.2 · ~/

80	80	24	57.0	2	2	2	0	0	FALSE
> modeweight=names(sort(table(Dataset\$weight.kg.)))[1]									
> modeweight									
[1] "49"									
> Dataset[is.na(Dataset\$weight.kg.), "weight.kg."] <-modeweight									
> Dataset									
	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1	0	1	0	0	FALSE
2	2	26	63	2	0	2	0	1	FALSE
3	3	26	62	2	1	2	0	0	FALSE
4	4	28	65	1	0	1	0	0	FALSE
5	5	22	58	2	0	2	0	1	FALSE
6	6	26	63	1	1	3	0	0	FALSE
7	7	27	64	2	0	2	0	0	FALSE
8	8	32	70	3	0	2	0	1	FALSE
9	9	28	63.5	2	0	<NA>	0	0	TRUE
10	10	27	64.5	1	1	2	0	1	FALSE
11	11	36	75	1	0	2	0	0	FALSE
12	12	33	70	1	1	3	0	1	FALSE
13	13	23	58	1	1	2	0	0	FALSE
14	14	20	55	1	0	2	1	0	FALSE
15	15	29	65	1	NA	<NA>	1	1	TRUE
16	16	25	61.5	1	2	3	0	0	FALSE
17	17	25	61.5	1	0	2	0	0	FALSE
18	18	20	55.5	1	2	1	0	1	FALSE
19	19	37	76	3	0	2	1	1	FALSE
20	20	24	56.6	1	2	3	1	1	FALSE
21	21	26	62	1	1	2	0	0	FALSE
22	22	33	75	2	0	3	1	1	FALSE
23	23	25	62	1	1	1	0	0	FALSE
24	24	27	65	NA	NA	3	1	1	TRUE
25	25	20	55	1	0	1	1	1	FALSE
26	26	18	49	NA	0	2	0	0	TRUE
27	27	18	50	1	NA	1	1	1	TRUE


```
modeDelivery_number=names(sort(table(Dataset$Delivery_number)))[1]
```

```
modeDelivery_number
```

```
Dataset[is.na(Dataset$Delivery_number), "Delivery_number"] <-modeDelivery_number
```

```
Dataset
```

```
157 modeDelivery_number=names(sort(table(Dataset$Delivery_number)))[1]
158 modeDelivery_number
159 Dataset[is.na(Dataset$Delivery_number), "Delivery_number"] <-modeDelivery_number
160 Dataset
```

157:1 (Top Level) ⚡

Console Terminal Background Jobs

R 4.2.2 . ~/

```
80 80 24 57 2 2 2 0 0 FALSE
> modeDelivery_number=names(sort(table(Dataset$Delivery_number)))[1]
> modeDelivery_number
[1] "4"
> Dataset[is.na(Dataset$Delivery_number), "Delivery_number"] <-modeDelivery_number
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1	0	1	0	0	FALSE
2	2	26	63	2	0	2	0	1	FALSE
3	3	26	62	2	1	2	0	0	FALSE
4	4	28	65	1	0	1	0	0	FALSE
5	5	22	58	2	0	2	0	1	FALSE
6	6	26	63	1	1	3	0	0	FALSE
7	7	27	64	2	0	2	0	0	FALSE
8	8	32	70	3	0	2	0	1	FALSE
9	9	28	63.5	2	0	<NA>	0	0	TRUE
10	10	27	64.5	1	1	2	0	1	FALSE
11	11	36	75	1	0	2	0	0	FALSE
12	12	33	70	1	1	3	0	1	FALSE
13	13	23	58	1	1	2	0	0	FALSE
14	14	20	55	1	0	2	1	0	FALSE
15	15	29	65	1	NA	<NA>	1	1	TRUE
16	16	25	61.5	1	2	3	0	0	FALSE
17	17	25	61.5	1	0	2	0	0	FALSE
18	18	20	55.5	1	2	1	0	1	FALSE
19	19	37	76	3	0	2	1	1	FALSE
20	20	24	56.6	1	2	3	1	1	FALSE
21	21	26	62	1	1	2	0	0	FALSE
22	22	33	75	2	0	3	1	1	FALSE
23	23	25	62	1	1	1	0	0	FALSE
24	24	27	65	4	NA	3	1	1	TRUE
25	25	20	55	1	0	1	1	1	FALSE
26	26	18	49	4	0	2	0	0	TRUE
27	27	18	50	1	NA	1	1	1	TRUE

```
modeDelivery_time=names(sort(table(Dataset$Delivery_time)))[1]
```

```
modeDelivery_time
```

```
Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-modeDelivery_time
```

```
Dataset
```

```
162 modeDelivery_time=names(sort(table(Dataset$Delivery_time)))[1]
163 modeDelivery_time
164 Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-modeDelivery_time
165 Dataset
166
167
```

162:1 (Top Level) ↕

Console Terminal Background Jobs

R 4.2.2 ~ /

```
> modeDelivery_time=names(sort(table(Dataset$Delivery_time)))[1]
> modeDelivery_time
[1] "1"
> Dataset[is.na(Dataset$Delivery_time), "Delivery_time"] <-modeDelivery_time
> Dataset
```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1	0	1	0	0	FALSE
2	2	26	63	2	0	2	0	1	FALSE
3	3	26	62	2	1	2	0	0	FALSE
4	4	28	65	1	0	1	0	0	FALSE
5	5	22	58	2	0	2	0	1	FALSE
6	6	26	63	1	1	3	0	0	FALSE
7	7	27	64	2	0	2	0	0	FALSE
8	8	32	70	3	0	2	0	1	FALSE
9	9	28	63.5	2	0	<NA>	0	0	TRUE
10	10	27	64.5	1	1	2	0	1	FALSE
11	11	36	75	1	0	2	0	0	FALSE
12	12	33	70	1	1	3	0	1	FALSE
13	13	23	58	1	1	2	0	0	FALSE
14	14	20	55	1	0	2	1	0	FALSE
15	15	29	65	1	1	<NA>	1	1	TRUE
16	16	25	61.5	1	2	3	0	0	FALSE
17	17	25	61.5	1	0	2	0	0	FALSE
18	18	20	55.5	1	2	1	0	1	FALSE
19	19	37	76	3	0	2	1	1	FALSE
20	20	24	56.6	1	2	3	1	1	FALSE
21	21	26	62	1	1	2	0	0	FALSE
22	22	33	75	2	0	3	1	1	FALSE
23	23	25	62	1	1	1	0	0	FALSE
24	24	27	65	4	1	3	1	1	TRUE
25	25	20	55	1	0	1	1	1	FALSE
26	26	18	49	4	0	2	0	0	TRUE
27	27	18	50	1	1	1	1	1	TRUE
28	28	30	68	1	0	2	0	0	FALSE

```
modeCaesarian=names(sort(table(Dataset$Caesarian)))[1]
```

```
modeCaesarian
```

```
Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-modeCaesarian
```

```
Dataset
```

```

167 modeCaesarian=names(sort(table(Dataset$Caesarian)))[1]
168 modeCaesarian
169 Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-modeCaesarian
170 Dataset
171
172

```

167:1 (Top Level) ↕

Console Terminal × Background Jobs ×

R 4.2.2 · ~/

80 80 24 57 2 2 2 0 0 FALSE

```

> modeCaesarian=names(sort(table(Dataset$Caesarian)))[1]
> modeCaesarian
[1] "0"
> Dataset[is.na(Dataset$Caesarian), "Caesarian"] <-modeCaesarian
> Dataset

```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1	0	1	0	0	FALSE
2	2	26	63	2	0	2	0	1	FALSE
3	3	26	62	2	1	2	0	0	FALSE
4	4	28	65	1	0	1	0	0	FALSE
5	5	22	58	2	0	2	0	1	FALSE
6	6	26	63	1	1	3	0	0	FALSE
7	7	27	64	2	0	2	0	0	FALSE
8	8	32	70	3	0	2	0	1	FALSE
9	9	28	63.5	2	0	<NA>	0	0	TRUE
10	10	27	64.5	1	1	2	0	1	FALSE
11	11	36	75	1	0	2	0	0	FALSE
12	12	33	70	1	1	3	0	1	FALSE
13	13	23	58	1	1	2	0	0	FALSE
14	14	20	55	1	0	2	1	0	FALSE
15	15	29	65	1	1	<NA>	1	1	TRUE
16	16	25	61.5	1	2	3	0	0	FALSE
17	17	25	61.5	1	0	2	0	0	FALSE
18	18	20	55.5	1	2	1	0	1	FALSE
19	19	37	76	3	0	2	1	1	FALSE
20	20	24	56.6	1	2	3	1	1	FALSE
21	21	26	62	1	1	2	0	0	FALSE
22	22	33	75	2	0	3	1	1	FALSE
23	23	25	62	1	1	1	0	0	FALSE
24	24	27	65	4	1	3	1	1	TRUE
25	25	20	55	1	0	1	1	1	FALSE
26	26	18	49	4	0	2	0	0	TRUE
27	27	18	50	1	1	1	1	1	TRUE
28	28	30	68	1	0	2	0	0	FALSE
29	29	32	73	1	0	1	1	1	FALSE
30	30	26	62.5	2	1	2	1	0	FALSE
31	31	25	58	1	0	3	0	0	FALSE
32	32	40	82	1	0	2	1	1	FALSE
33	33	32	68	2	0	1	1	1	FALSE
34	34	27	63	2	0	2	1	1	FALSE
35	35	26	59	2	2	2	0	1	FALSE
36	36	28	66	3	0	1	0	1	FALSE
37	37	33	75	1	1	2	0	0	FALSE
38	38	31	69	2	2	2	0	0	FALSE
39	39	31	63	1	0	2	0	0	FALSE
40	40	26	59	1	2	3	1	1	FALSE
41	41	27	63	1	0	1	1	1	FALSE
42	42	19	51	1	0	2	0	1	FALSE
43	43	36	73	1	1	1	0	1	FALSE
44	44	22	57	1	0	2	0	1	FALSE
45	45	36	72.5	4	0	1	1	1	FALSE
46	46	28	62.5	3	0	2	1	1	FALSE
47	47	26	49	1	0	2	0	0	TRUE
48	48	32	67.5	2	0	1	1	1	FALSE
49	49	26	62.5	2	2	2	0	0	FALSE
50	50	19	49	2	0	3	1	1	TRUE

51	51	33	68.5	3	2	2	1	0	FALSE
52	52	21	53	2	1	3	1	1	FALSE
53	53	30	68	3	2	1	0	0	FALSE
54	54	35	74	1	1	3	0	0	FALSE
55	55	29	63.5	2	0	2	1	1	FALSE
56	56	25	59	2	0	2	0	0	FALSE
57	57	32	67.5	3	1	3	1	1	FALSE
58	58	95	110	1	0	3	0	1	FALSE
59	59	26	61.5	1	0	1	0	1	FALSE
60	60	30	67.5	2	1	1	1	0	TRUE
61	61	22	58.5	1	2	1	0	0	FALSE
62	62	19	49	1	0	2	0	1	TRUE
63	63	32	67	2	0	3	0	1	FALSE
64	64	32	67	2	0	2	1	1	FALSE
65	65	31	66	1	2	1	1	0	FALSE
66	66	35	72	2	0	2	0	1	FALSE
67	67	28	62.5	3	0	2	0	1	FALSE
68	68	29	64.5	2	0	2	1	0	FALSE
69	69	25	62	1	0	3	0	1	FALSE
70	70	27	61	2	2	3	0	0	FALSE
71	71	90	105	1	0	3	0	1	FALSE
72	72	29	65	1	2	<NA>	1	1	TRUE
73	73	28	64	2	0	2	0	0	FALSE
74	74	32	69	3	0	2	1	0	FALSE
75	75	38	75	3	2	1	1	1	FALSE
76	76	27	62.5	2	1	2	0	0	FALSE
77	77	33	66	4	0	2	0	0	TRUE
78	78	19	63	2	1	1	0	1	TRUE
79	79	25	58	1	2	3	0	1	FALSE
80	80	24	57	2	2	2	0	0	FALSE

IV. Remove the rows that have missing values from the data set.

```
remove_missingvalue<- na.omit(Dataset)
```

```
remove_missingvalue
```



```

90 remove_missingvalue<- na.omit(Dataset)
91 remove_missingvalue
92

```

90:1 (Top Level) ↕

Console Terminal Background Jobs

R 4.2.2 ~/

```

> remove_missingvalue<- na.omit(Dataset)
> remove_missingvalue

```

	id	Age	weight.kg.	Delivery_number	Delivery_time	Blood	Heart	Caesarian	missing_value
1	1	22	57.7	1	0	1	0	0	FALSE
2	2	26	63.0	2	0	2	0	1	FALSE
3	3	26	62.0	2	1	2	0	0	FALSE
4	4	28	65.0	1	0	1	0	0	FALSE
5	5	22	58.0	2	0	2	0	1	FALSE
6	6	26	63.0	1	1	3	0	0	FALSE
7	7	27	64.0	2	0	2	0	0	FALSE
8	8	32	70.0	3	0	2	0	1	FALSE
10	10	27	64.5	1	1	2	0	1	FALSE
11	11	36	75.0	1	0	2	0	0	FALSE
12	12	33	70.0	1	1	3	0	1	FALSE
13	13	23	58.0	1	1	2	0	0	FALSE
14	14	20	55.0	1	0	2	1	0	FALSE
16	16	25	61.5	1	2	3	0	0	FALSE
17	17	25	61.5	1	0	2	0	0	FALSE
18	18	20	55.5	1	2	1	0	1	FALSE
19	19	37	76.0	3	0	2	1	1	FALSE
20	20	24	56.6	1	2	3	1	1	FALSE
21	21	26	62.0	1	1	2	0	0	FALSE
22	22	33	75.0	2	0	3	1	1	FALSE
23	23	25	62.0	1	1	1	0	0	FALSE
25	25	20	55.0	1	0	1	1	1	FALSE
28	28	30	68.0	1	0	2	0	0	FALSE
29	29	32	73.0	1	0	1	1	1	FALSE
30	30	26	62.5	2	1	2	1	0	FALSE
31	31	25	58.0	1	0	3	0	0	FALSE
32	32	40	82.0	1	0	2	1	1	FALSE
33	33	32	68.0	2	0	1	1	1	FALSE
34	34	27	63.0	2	0	2	1	1	FALSE
35	35	26	59.0	2	2	2	0	1	FALSE
36	36	28	66.0	3	0	1	0	1	FALSE
37	37	33	75.0	1	1	2	0	0	FALSE
38	38	31	69.0	2	2	2	0	0	FALSE
39	39	31	63.0	1	0	2	0	0	FALSE
40	40	26	59.0	1	2	3	1	1	FALSE
41	41	27	63.0	1	0	1	1	1	FALSE
42	42	19	51.0	1	0	2	0	1	FALSE
43	43	36	73.0	1	1	1	0	1	FALSE
44	44	22	57.0	1	0	2	0	1	FALSE
45	45	36	72.5	4	0	1	1	1	FALSE
46	46	28	62.5	3	0	2	1	1	FALSE
48	48	32	67.5	2	0	1	1	1	FALSE
49	49	26	62.5	2	2	2	0	0	FALSE
51	51	33	68.5	3	2	2	1	0	FALSE
52	52	21	53.0	2	1	3	1	1	FALSE
53	53	30	68.0	3	2	1	0	0	FALSE
54	54	35	74.0	1	1	3	0	0	FALSE
55	55	29	63.5	2	0	2	1	1	FALSE
56	56	25	59.0	2	0	2	0	0	FALSE
57	57	32	67.5	3	1	3	1	1	FALSE
58	58	95	110.0	1	0	3	0	1	FALSE
59	59	26	61.5	1	0	1	0	1	FALSE
61	61	22	58.5	1	2	1	0	0	FALSE
63	63	32	67.0	2	0	3	0	1	FALSE
64	64	32	67.0	2	0	2	1	1	FALSE
65	65	31	66.0	1	2	1	1	0	FALSE
66	66	35	72.0	2	0	2	0	1	FALSE
67	67	28	62.5	3	0	2	0	1	FALSE
68	68	29	64.5	2	0	2	1	0	FALSE
69	69	25	62.0	1	0	3	0	1	FALSE
70	70	27	61.0	2	2	3	0	0	FALSE
71	71	90	105.0	1	0	3	0	1	FALSE
73	73	28	64.0	2	0	2	0	0	FALSE
74	74	32	69.0	3	0	2	1	0	FALSE
75	75	38	75.0	3	2	1	1	1	FALSE
76	76	27	62.5	2	1	2	0	0	FALSE
79	79	25	58.0	1	2	3	0	1	FALSE
80	80	24	57.0	2	2	2	0	0	FALSE