Assignment: Data Analysis

Task 1: Data Aggregation Using `aggregate` and `dplyr`

Question 1:

Run the following lines and study how they work. Then state what they do and output for us.

```r
# Creating the dataframe
df1 <- data.frame(Name =
c('James','Paul','Richards','Marico','Samantha','Ravi','Raghu',
'Richards','George','Ema','Samantha','Catherine'),
                  State = c('Alaska','California','Texas','North
Carolina','California','Texas','Alaska','Texas','North
Carolina','Alaska','California','Texas'),
                  Sales = c(14,24,31,12,13,7,9,31,18,16,18,14))

# Aggregating using aggregate
agg_result <- aggregate(df1$Sales, by = list(df1$State), FUN = sum)
print(agg_result)

##          Group.1   x
## 1         Alaska  39
## 2     California  55
## 3 North Carolina  30
## 4          Texas  83

# Aggregating using dplyr
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.5.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

dplyr_result <- df1 %>%
  group_by(State) %>%
  summarise(sum_sales = sum(Sales))
print(dplyr_result)
```

```
## # A tibble: 4 × 2
##   State          sum_sales
##   <chr>              <dbl>
## 1 Alaska                39
## 2 California            55
## 3 North Carolina        30
## 4 Texas                 83
```

Explanation:

- The above code creates a data frame df1 containing Name, State, and Sales.
- Then it calculates the sum of sales grouped by State using both aggregate() and dplyr package.

Output: - The output will show the sum of Sales for each state using both methods (aggregate() and dplyr).

Task 2: World Cup Matches Data Analysis

Question 2a:

Use R to read the WorldCupMatches.csv from the DATA folder on Google Drive. Then perform the following:

(a) Find the size of the data frame. How many rows, how many columns?

```
# Read the WorldCupMatches dataset
wc_data <- read.csv("C:\\Users\\Lenovo\\Downloads\\WorldCupMatches.csv")

# Find the size of the data frame
data_size <- dim(wc_data)
data_size

## [1] 852  20
```

Explanation:

- The dim() function returns the number of rows and columns in the data frame.

Output: - The number of rows and columns in the WorldCupMatches dataset.

(b) Use summary function to report the statistical summary of your data.

```
# Get the statistical summary of the dataset
summary(wc_data)

##       Year        Datetime              Stage              Stadium
##  Min.   :1930   Length:852         Length:852         Length:852
##  1st Qu.:1970   Class :character   Class :character   Class :character
##  Median :1990   Mode  :character   Mode  :character   Mode  :character
##  Mean   :1985
```

```
##   3rd Qu.:2002
##   Max.   :2014
##
##      City           Home.Team.Name      Home.Team.Goals   Away.Team.Goals
##   Length:852        Length:852          Min.   : 0.000    Min.   :0.000
##   Class :character  Class :character    1st Qu.: 1.000    1st Qu.:0.000
##   Mode  :character  Mode  :character    Median : 2.000    Median :1.000
##                                         Mean   : 1.811    Mean   :1.022
##                                         3rd Qu.: 3.000    3rd Qu.:2.000
##                                         Max.   :10.000    Max.   :7.000
##
##   Away.Team.Name      Win.conditions       Attendance
Half.time.Home.Goals
##   Length:852        Length:852          Min.   :  2000    Min.   :0.0000
##   Class :character  Class :character    1st Qu.: 30000    1st Qu.:0.0000
##   Mode  :character  Mode  :character    Median : 41580    Median :0.0000
##                                         Mean   : 45165    Mean   :0.7089
##                                         3rd Qu.: 61375    3rd Qu.:1.0000
##                                         Max.   :173850    Max.   :6.0000
##                                         NA's   :2
##   Half.time.Away.Goals   Referee        Assistant.1        Assistant.2
##   Min.   :0.0000       Length:852       Length:852         Length:852
##   1st Qu.:0.0000        Class :character  Class :character  Class
:character
##   Median :0.0000        Mode  :character  Mode  :character  Mode
:character
##   Mean   :0.4284
##   3rd Qu.:1.0000
##   Max.   :5.0000
##
##      RoundID            MatchID          Home.Team.Initials
Away.Team.Initials
##   Min.   :     201  Min.   :      25  Length:852          Length:852
##   1st Qu.:     262  1st Qu.:    1189  Class :character    Class
:character
##   Median :     337  Median :    2191  Mode  :character    Mode
:character
##   Mean   :10661773  Mean   : 61346868
##   3rd Qu.:  249722  3rd Qu.: 43950059
##   Max.   :97410600  Max.   :300186515
##
```

Output: - A summary of each column, including mean, min, max, and other statistics.

(c) Find how many unique locations Olympics were held at.

```
# Find unique locations
unique_locations <- length(unique(wc_data$Location))
unique_locations

## [1] 0
```

Output: - The number of unique locations where World Cup matches were held.

(d) Find the average attendance.

```
# Find average attendance
avg_attendance <- mean(wc_data$Attendance, na.rm = TRUE)
avg_attendance

## [1] 45164.8
```

Output: - The average attendance at the World Cup matches.

(e) For each Home Team, what is the total number of goals scored?

```
# Aggregating goals scored by Home Team using dplyr
home_goals <- wc_data %>%
  group_by(`Home.Team.Name`) %>%
  summarise(total_goals = sum(`Home.Team.Goals`, na.rm = TRUE))
print(home_goals)

## # A tibble: 78 × 2
##    Home.Team.Name total_goals
##    <chr>                <int>
##  1 Algeria                  5
##  2 Angola                   0
##  3 Argentina              111
##  4 Australia                7
##  5 Austria                 31
##  6 Belgium                 27
##  7 Bolivia                  1
##  8 Brazil                 180
##  9 Bulgaria                11
## 10 Cameroon                11
## # i 68 more rows
```

Output: - The total number of goals scored by each Home Team.

(f) What is the average number of attendees for each year?

```
# Average attendance for each year
avg_attendance_year <- wc_data %>%
  group_by(Year) %>%
  summarise(avg_attendance = mean(Attendance, na.rm = TRUE))
avg_attendance_year

## # A tibble: 20 × 2
##     Year avg_attendance
##    <int>          <dbl>
##  1  1930         32808.
##  2  1934         21353.
##  3  1938         20872.
##  4  1950         47511.
##  5  1954         29562.
##  6  1958         23423.
```

```
##  7  1962        27912.
##  8  1966        48848.
##  9  1970        50124.
## 10  1974        49099.
## 11  1978        40679.
## 12  1982        40572.
## 13  1986        46039.
## 14  1990        48389.
## 15  1994        68991.
## 16  1998        43517.
## 17  2002        42269.
## 18  2006        52491.
## 19  2010        49670.
## 20  2014        55375.
```

Output: - The average attendance for each year, along with any trends or patterns.

Task 3: Metabolites Data Analysis

Question 3:

Use R to read the metabolites.csv from the DATA folder on Google Drive. Then perform the following:

(a) Find how many Alzheimer's patients there are in the data set.
```r
# Read the metabolites dataset
metabolites_data <- read.csv("C:\\Users\\Lenovo\\Downloads\\metabolite.csv")

# Find number of Alzheimer's patients
alzheimers_patients <- sum(metabolites_data$Diagnosis == 'Alzheimer', na.rm =
TRUE)
alzheimers_patients

## [1] 0
```

Explanation:

- Counts the number of Alzheimer's patients in the dataset by checking the `Diagnosis` column.

Output: - The count of Alzheimer's patients in the dataset.

(b) Determine the number of missing values for each column.
```r
# Find missing values for each column
missing_values <- colSums(is.na(metabolites_data))
missing_values
```

```
##       Label         Phe         Pro
Ser
##           0           0           0
0
##         Thr        ADMA   alpha.AAA
c4.OH.Pro
##           0           0           0
20
##   Carnosine  Creatinine        DOPA
Dopamine
##           1           0           0
20
##   Histamine   Kynurenine      Met.SO
Nitro.Tyr
##           0           0           1
62
##         PEA   Putrescine   Sarcosine
Serotonin
##          69           0           0
0
##  Spermidine    Spermine    t4.OH.Pro
Taurine
##           0          60           0
2
##        SDMA          C0         C10
C10.1
##           0           0           0
0
##       C10.2         C12      C12.DC
C12.1
##           0           0           1
0
##         C14        C14.1    C14.1.OH
C14.2
##           0           0           1
0
##    C14.2.OH         C16      C16.OH
C16.1
##           2           0           1
0
##    C16.1.OH       C16.2    C16.2.OH
C18
##           2           2           1
0
##       C18.1    C18.1.OH      C18.2
C2
##           0           7           0
0
##          C3       C3.OH        C3.1
C4
```

```
##                    0               8               2
0
##      C3.DC..C4.OH.             C4.1              C5
C5.M.DC
##                    0               0               0
1
##      C5.OH..C3.DC.M.           C5.1          C5.1.DC
C6..C4.1.DC.
##                    0               5               2
0
##      C5.DC..C6.OH.             C6.1           C7.DC
C8
##                    4               2               1
0
##                   C9    lysoPC.a.C14.0   lysoPC.a.C16.0
lysoPC.a.C16.1
##                    1               0               0
0
##      lysoPC.a.C17.0   lysoPC.a.C18.0   lysoPC.a.C18.1
lysoPC.a.C18.2
##                    0               0               0
0
##      lysoPC.a.C20.3   lysoPC.a.C20.4   lysoPC.a.C24.0
lysoPC.a.C26.0
##                    0               0               0
0
##      lysoPC.a.C26.1   lysoPC.a.C28.0   lysoPC.a.C28.1
PC.aa.C24.0
##                    0               0               0
0
##         PC.aa.C26.0      PC.aa.C28.1      PC.aa.C30.0
PC.aa.C32.0
##                    0               0               0
0
##         PC.aa.C32.1      PC.aa.C32.2      PC.aa.C32.3
PC.aa.C34.1
##                    0              47               0
0
##         PC.aa.C34.2      PC.aa.C34.3      PC.aa.C34.4
PC.aa.C36.0
##                    0               0               0
0
##         PC.aa.C36.1      PC.aa.C36.2      PC.aa.C36.3
PC.aa.C36.4
##                    0               0               0
0
##         PC.aa.C36.5      PC.aa.C36.6      PC.aa.C38.0
PC.aa.C38.3
##                    0               0               0
0
```

```
##        PC.aa.C38.4       PC.aa.C38.5       PC.aa.C38.6
PC.aa.C40.1
##                  0                 0                 0
0
##        PC.aa.C40.2       PC.aa.C40.3       PC.aa.C40.4
PC.aa.C40.5
##                  0                 0                 0
0
##        PC.aa.C40.6       PC.aa.C42.0       PC.aa.C42.1
PC.aa.C42.2
##                  0                 0                 0
0
##        PC.aa.C42.4       PC.aa.C42.5       PC.aa.C42.6
PC.ae.C30.0
##                  0                 0                 0
0
##        PC.ae.C30.1       PC.ae.C30.2       PC.ae.C32.1
PC.ae.C32.2
##                 10                 0                 0
0
##        PC.ae.C34.0       PC.ae.C34.1       PC.ae.C34.2
PC.ae.C34.3
##                  0                 0                 0
0
##        PC.ae.C36.0       PC.ae.C36.1       PC.ae.C36.2
PC.ae.C36.3
##                  0                 0                 0
0
##        PC.ae.C36.4       PC.ae.C36.5       PC.ae.C38.0
PC.ae.C38.1
##                  0                 0                 0
52
##        PC.ae.C38.2       PC.ae.C38.3       PC.ae.C38.4
PC.ae.C38.5
##                 19                 0                 0
0
##        PC.ae.C38.6       PC.ae.C40.1       PC.ae.C40.2
PC.ae.C40.3
##                  0                 0                 0
0
##        PC.ae.C40.4       PC.ae.C40.5       PC.ae.C40.6
PC.ae.C42.0
##                  0                 0                 0
0
##        PC.ae.C42.1       PC.ae.C42.2       PC.ae.C42.3
PC.ae.C42.4
##                  0                 1                 0
0
##        PC.ae.C42.5       PC.ae.C44.3       PC.ae.C44.4
PC.ae.C44.5
```

```
##                       0                  0                  0
0
##          PC.ae.C44.6        SM..OH..C14.1        SM..OH..C16.1
SM..OH..C22.1
##                       0                  0                  0
0
##        SM..OH..C22.2        SM..OH..C24.1           SM.C16.0
SM.C16.1
##                       0                  0                  0
0
##             SM.C18.0            SM.C18.1           SM.C20.2
SM.C24.0
##                       0                  0                  0
0
##             SM.C24.1            SM.C26.0           SM.C26.1
H1_1
##                       0                  0                  0
0
##                  H1             Urea_N         L.Arginine_N
L.Leucine_N
##                       0                  1                  1
1
##          EDTAca_N X2.Hydroxybutyrate X3.Hydroxybutyrate
Acetate
##                       1                  1                  1
1
##       Acetoacetate            Acetone           Betaine
Carnitine
##                       1                  1                  1
1
##             Choline           Creatine    Dimethyl.sulfone
Ethanol
##                       1                  1                  1
2
##             Formate            Glucose           Glycerol
Hypoxanthine
##                       2                  1                  1
1
##          Isobutyrate         Isopropanol           Lactate
Malonate
##                       1                  1                  1
1
```

Output: - The number of missing values for each column.

(c) Remove the rows which have missing values for the Dopamine column.

```r
# Remove rows where Dopamine column has missing values
metabolites_no_na_dopamine <-
metabolites_data[!is.na(metabolites_data$Dopamine), ]
head(metabolites_no_na_dopamine)
```

```
##          Label  Phe Pro Ser Thr ADMA alpha.AAA c4.OH.Pro Carnosine
Creatinine
## 1  Alzheimer 72.8 166 170 282 1.15     0.760     0.236     1.270
49.9
## 4  Alzheimer 94.1 129 162 201 1.10     0.795        NA     0.675
80.1
## 5  Alzheimer 79.8 126 115 199 1.24     1.360        NA     1.280
60.5
## 8    Healthy 83.6 119 135 268 1.18     0.779     0.215     0.647
30.6
## 9    Healthy 73.7 124 145 307 1.17     0.785     0.186     0.590
39.8
## 10   Healthy 81.5 151 174 269 1.30     0.767     0.185     1.130
44.0
##      DOPA Dopamine Histamine Kynurenine Met.SO Nitro.Tyr PEA Putrescine
## 1  0.265    0.233     0.225       5.21  0.526     0.027  NA      0.068
## 4  0.264    0.234     0.209       5.80  0.389        NA  NA      0.110
## 5  0.271    0.231     0.210       4.46  0.466        NA  NA      0.118
## 8  0.275    0.244     0.214       5.66  0.245     0.002  NA      0.161
## 9  0.259    0.233     0.210       6.36  0.413        NA  NA      0.121
## 10 0.257    0.225     0.210       5.69  0.178        NA  NA      0.183
##    Sarcosine Serotonin Spermidine Spermine t4.OH.Pro Taurine SDMA   C0
C10
## 1      17.8     0.147      0.188       NA      24.0     125 1.13 18.2
0.059
## 4      18.7     0.255      0.353       NA      23.1     159 1.34 23.5
0.071
## 5      22.5     0.390      0.473       NA      26.9     149 1.24 13.6
0.139
## 8      23.3     0.215      0.276       NA      10.7     133 1.04 13.3
0.051
## 9      22.1     0.166      0.327       NA      16.0     215 1.24 15.8
0.061
## 10     18.5     0.320      0.236       NA      27.9     140 1.84 15.6
0.068
##    C10.1 C10.2   C12 C12.DC C12.1   C14 C14.1 C14.1.OH C14.2 C14.2.OH
C16
## 1  0.312 0.038 0.030  0.042 0.290 0.023 0.019    0.008 0.008    0.006
0.046
## 4  0.317 0.040 0.045  0.048 0.275 0.026 0.028    0.010 0.013    0.011
0.074
## 5  0.472 0.074 0.056  0.079 0.394 0.034 0.043    0.016 0.025    0.017
0.062
## 8  0.217 0.030 0.041  0.035 0.174 0.024 0.017    0.007 0.006    0.007
0.060
## 9  0.258 0.036 0.037  0.038 0.228 0.022 0.018    0.007 0.007    0.007
0.054
## 10 0.319 0.041 0.039  0.037 0.284 0.026 0.022    0.008 0.013    0.007
0.076
##    C16.OH C16.1 C16.1.OH C16.2 C16.2.OH   C18 C18.1 C18.1.OH C18.2    C2
```

```
C3
## 1   0.008 0.009    0.007 0.005    0.013 0.013 0.024    0.003 0.016 1.97
0.354
## 4   0.011 0.015    0.008 0.006    0.009 0.020 0.035    0.004 0.033 2.10
0.278
## 5      NA 0.024    0.014 0.012    0.025 0.031 0.034    0.012 0.017 5.62
0.436
## 8   0.006 0.010    0.005 0.004    0.008 0.020 0.025    0.004 0.019 1.66
0.258
## 9   0.005 0.012    0.005 0.005    0.009 0.014 0.026    0.003 0.016 2.21
0.233
## 10  0.009 0.012    0.007 0.006    0.014 0.024 0.031    0.004 0.028 2.58
0.364
##     C3.OH  C3.1    C4 C3.DC..C4.OH.  C4.1    C5 C5.M.DC C5.OH..C3.DC.M.
C5.1
## 1  0.008 0.015 0.082          0.045 0.025 0.094   0.023           0.026
0.030
## 4  0.010 0.017 0.110          0.077 0.031 0.145   0.034           0.041
0.035
## 5  0.029 0.035 0.106          0.099 0.069 0.141   0.094           0.058
0.073
## 8  0.008 0.012 0.082          0.047 0.021 0.107   0.023           0.023
0.021
## 9  0.008 0.014 0.088          0.029 0.024 0.127   0.024           0.024
0.025
## 10 0.010 0.015 0.100          0.038 0.025 0.292   0.024           0.023
NA
##     C5.1.DC C6..C4.1.DC. C5.DC..C6.OH.  C6.1 C7.DC    C8    C9
lysoPC.a.C14.0
## 1    0.020        0.022         0.014 0.018 0.011 0.062 0.016
2.23
## 4    0.016        0.029         0.016 0.027 0.017 0.091 0.018
2.19
## 5    0.049        0.052         0.040 0.040 0.036 0.192 0.041
1.88
## 8    0.017        0.036         0.011    NA 0.009 0.062 0.011
2.13
## 9    0.016        0.026         0.018 0.015 0.013 0.064 0.014
2.10
## 10   0.014        0.044         0.017 0.018 0.014 0.067 0.013
2.06
##    lysoPC.a.C16.0 lysoPC.a.C16.1 lysoPC.a.C17.0 lysoPC.a.C18.0
lysoPC.a.C18.1
## 1            37.9           2.66          0.446           9.00
8.58
## 4            32.8           2.39          0.323           7.21
7.22
## 5            24.5           1.27          0.382           6.66
5.39
## 8            33.7           3.09          0.455           6.96
```

```
7.31
## 9             36.0            3.46           0.435            7.27
8.11
## 10            32.7            2.48           0.368            6.49
7.86
##    lysoPC.a.C18.2 lysoPC.a.C20.3 lysoPC.a.C20.4 lysoPC.a.C24.0
lysoPC.a.C26.0
## 1            7.27           1.83            8.25           0.079
0.113
## 4            7.62           1.64            6.75           0.066
0.086
## 5            3.60           0.97            6.26           0.084
0.118
## 8            7.53           2.35            8.73           0.061
0.083
## 9            6.75           2.08            7.82           0.068
0.083
## 10          10.10           1.86           12.40           0.061
0.086
##    lysoPC.a.C26.1 lysoPC.a.C28.0 lysoPC.a.C28.1 PC.aa.C24.0 PC.aa.C26.0
## 1           0.053          0.108          0.072       0.082       0.438
## 4           0.045          0.076          0.076       0.076       0.486
## 5           0.053          0.092          0.072       0.069       0.401
## 8           0.047          0.124          0.078       0.082       0.424
## 9           0.037          0.096          0.061       0.064       0.430
## 10          0.052          0.088          0.074       0.065       0.453
##    PC.aa.C28.1 PC.aa.C30.0 PC.aa.C32.0 PC.aa.C32.1 PC.aa.C32.2 PC.aa.C32.3
## 1        0.571        2.35        11.4        9.22          NA       0.092
## 4        0.685        3.33        18.6       13.30       0.053       0.079
## 5        0.513        1.78        13.8        5.03          NA       0.102
## 8        0.605        2.32        12.1       11.50          NA       0.093
## 9        0.486        2.05        10.4       11.40          NA       0.082
## 10       0.578        2.28        13.8        8.10          NA       0.111
##    PC.aa.C34.1 PC.aa.C34.2 PC.aa.C34.3 PC.aa.C34.4 PC.aa.C36.0 PC.aa.C36.1
## 1        109.0        71.0       1.430       0.200        2.38        21.7
## 4        106.0        93.6       1.590       0.190        2.57        20.9
## 5         83.4        35.9       0.709       0.135        1.83        20.5
## 8         83.6        60.6       1.580       0.251        2.16        18.4
## 9         89.6        55.1       1.390       0.221        1.62        18.3
## 10        80.7        82.6       1.640       0.244        1.75        14.6
##    PC.aa.C36.2 PC.aa.C36.3 PC.aa.C36.4 PC.aa.C36.5 PC.aa.C36.6 PC.aa.C38.0
## 1         42.4        42.7       120.0        1.86       0.084       1.230
## 4         48.8        41.2       122.0        1.76       0.070       1.160
## 5         28.5        21.9        98.1        1.70       0.048       1.100
## 8         34.4        41.5       110.0        2.03       0.106       1.150
## 9         32.0        41.8       102.0        1.77       0.081       0.902
## 10        38.1        31.2       138.0        1.80       0.092       0.904
##    PC.aa.C38.3 PC.aa.C38.4 PC.aa.C38.5 PC.aa.C38.6 PC.aa.C40.1 PC.aa.C40.2
## 1         32.1        95.1        16.8        41.6       0.195       0.074
## 4         28.7        92.7        14.3        29.9       0.220       0.097
```

```
## 5          23.3        101.0         13.8         36.2        0.165        0.044
## 8          31.3         81.7         14.5         42.8        0.225        0.099
## 9          31.1         75.9         13.6         37.6        0.181        0.100
## 10         22.2         98.2         12.7         31.5        0.189        0.130
##      PC.aa.C40.3 PC.aa.C40.4 PC.aa.C40.5 PC.aa.C40.6 PC.aa.C42.0 PC.aa.C42.1
## 1        0.491        3.48         5.66         21.8        0.364        0.226
## 4        0.433        3.59         5.06         14.0        0.427        0.223
## 5        0.525        3.37         5.29         22.5        0.125        0.095
## 8        0.384        4.02         5.49         20.6        0.365        0.193
## 9        0.298        2.88         4.88         17.5        0.396        0.216
## 10       0.315        2.87         3.32         12.6        0.301        0.186
##      PC.aa.C42.2 PC.aa.C42.4 PC.aa.C42.5 PC.aa.C42.6 PC.ae.C30.0 PC.ae.C30.1
## 1        0.108       0.272        0.272        0.291        0.173        0.027
## 4        0.119       0.268        0.267        0.254        0.223        0.049
## 5        0.083       0.206        0.205        0.280        0.095        0.082
## 8        0.089       0.266        0.237        0.288        0.212        0.028
## 9        0.086       0.215        0.206        0.223        0.179        0.012
## 10       0.092       0.228        0.199        0.224        0.164        0.017
##      PC.ae.C30.2 PC.ae.C32.1 PC.ae.C32.2 PC.ae.C34.0 PC.ae.C34.1 PC.ae.C34.2
## 1        0.022        1.65        0.371        0.880         3.66         2.48
## 4        0.023        2.47        0.459        0.964         4.06         3.09
## 5        0.023        1.72        0.316        1.060         3.28         1.70
## 8        0.033        1.68        0.335        1.140         3.51         2.28
## 9        0.022        1.53        0.291        0.683         3.20         2.03
## 10       0.024        1.70        0.341        0.800         3.56         2.52
##      PC.ae.C34.3 PC.ae.C36.0 PC.ae.C36.1 PC.ae.C36.2 PC.ae.C36.3 PC.ae.C36.4
## 1        0.813        0.498         5.64         1.90         1.17         6.96
## 4        1.020        0.552         4.75         2.01         1.35         8.36
## 5        0.722        0.553         5.95         1.47         0.76         4.78
## 8        0.868        0.479         5.65         1.97         1.25         7.15
## 9        0.663        0.384         4.79         1.54         1.12         6.43
## 10       0.983        0.498         4.85         1.93         1.04         7.08
##      PC.ae.C36.5 PC.ae.C38.0 PC.ae.C38.1 PC.ae.C38.2 PC.ae.C38.3 PC.ae.C38.4
## 1        4.79        0.474        0.287        0.538         2.66         6.33
## 4        5.97        0.397        0.022        0.144         1.97         5.99
## 5        4.00        0.430        0.271        0.246         1.80         5.45
## 8        4.04        0.550        0.078        0.559         2.90         5.73
## 9        3.32        0.442           NA        0.224         2.53         5.03
## 10       4.52        0.458           NA        0.053         1.99         5.79
##      PC.ae.C38.5 PC.ae.C38.6 PC.ae.C40.1 PC.ae.C40.2 PC.ae.C40.3 PC.ae.C40.4
## 1        5.51        1.95        0.574        0.575        0.940         1.76
## 4        5.63        1.97        0.425        0.540        0.742         1.45
## 5        4.34        1.51        0.430        0.432        0.632         1.10
## 8        4.53        1.71        0.584        0.499        0.991         1.84
## 9        3.82        1.28        0.472        0.555        0.737         1.35
## 10       4.84        1.58        0.670        0.572        0.757         1.16
##      PC.ae.C40.5 PC.ae.C40.6 PC.ae.C42.0 PC.ae.C42.1 PC.ae.C42.2 PC.ae.C42.3
## 1        1.77        1.59        0.629        0.316        0.192        0.277
## 4        1.62        1.25        0.637        0.299        0.159        0.208
## 5        1.25        1.47        0.660        0.355        0.138        0.174
```

```
## 8          1.53         1.46         0.631         0.271         0.186         0.282
## 9          1.34         1.21         0.645         0.264         0.168         0.229
## 10         1.30         1.18         0.670         0.294         0.186         0.222
##     PC.ae.C42.4 PC.ae.C42.5 PC.ae.C44.3 PC.ae.C44.4 PC.ae.C44.5 PC.ae.C44.6
## 1         0.264       0.888       0.065       0.168       0.536       0.494
## 4         0.392       0.863       0.069       0.237       0.517       0.611
## 5         0.162       0.513       0.081       0.154       0.178       0.134
## 8         0.333       0.861       0.085       0.183       0.534       0.429
## 9         0.319       0.925       0.069       0.184       0.671       0.460
## 10        0.266       0.660       0.080       0.156       0.420       0.399
##     SM..OH..C14.1 SM..OH..C16.1 SM..OH..C22.1 SM..OH..C22.2 SM..OH..C24.1
## 1          1.420         1.330          2.07          1.86         0.597
## 4          1.720         1.480          2.97          2.84         0.682
## 5          0.987         1.480          1.96          1.74         0.478
## 8          1.500         1.160          2.98          2.59         0.631
## 9          1.230         0.984          2.21          2.07         0.507
## 10         1.760         1.450          2.60          2.45         0.586
##     SM.C16.0 SM.C16.1 SM.C18.0 SM.C18.1 SM.C20.2 SM.C24.0 SM.C24.1 SM.C26.0
## 1       44.9     7.99     14.5    10.40    0.290    12.20     27.3    0.147
## 4       52.4     8.61     17.2    11.50    0.261    11.80     27.9    0.138
## 5       40.6     5.86     13.0     8.34    0.196     9.29     20.5    0.111
## 8       37.9     7.92     11.9     9.59    0.199     9.36     18.8    0.117
## 9       37.3     6.65     12.0     8.48    0.183    10.30     23.6    0.090
## 10      48.3     9.66     15.4    11.90    0.295    10.50     26.7    0.097
##     SM.C26.1 H1_1   H1 Urea_N L.Arginine_N L.Leucine_N EDTAca_N
## 1      0.337 3356 3356     NA           NA          NA       NA
## 4      0.353 2652 2652  500.8         16.0        27.1      2.5
## 5      0.283 2258 2258  132.5         13.2        57.9      2.5
## 8      0.227 2464 2464  182.3         49.8        63.9      0.0
## 9      0.268 2725 2725  255.8         39.0        64.6      2.7
## 10     0.305 3583 3583  184.4         66.0        54.6      2.3
##     X2.Hydroxybutyrate X3.Hydroxybutyrate Acetate Acetoacetate Acetone
Betaine
## 1                  NA                 NA      NA           NA      NA
NA
## 4                12.7                7.2     9.8          4.8     4.0
13.9
## 5                35.2               44.7    20.2         18.9    18.9
33.9
## 8                21.7               20.9    19.5         15.4     6.6
35.3
## 9                17.1               37.9    20.0         22.0     9.3
12.0
## 10               21.3               46.2    22.5         26.2     9.4
29.5
##     Carnitine Choline Creatine Dimethyl.sulfone Ethanol Formate Glucose
Glycerol
## 1         NA      NA       NA               NA      NA      NA      NA
NA
## 4        7.7    11.8     14.7              1.3     6.4    14.4   629.5
```

```
322.0
## 5      18.5    27.7    35.4            5.5    13.0    40.0  1618.0
271.6
## 8      13.0    21.5    25.9            5.4    10.2    23.2  2120.5
406.5
## 9      14.4    15.4    25.3            3.5     5.1    25.0  1644.2
479.9
## 10     15.7    28.1    25.6            2.2     4.4    29.4  2875.0
722.2
##      Hypoxanthine Isobutyrate Isopropanol Lactate Malonate
## 1             NA          NA          NA      NA       NA
## 4            8.6         2.5         4.4  1037.7      7.6
## 5            0.0         6.1        11.2  2199.9     11.7
## 8            5.8         4.5         4.4  1661.4     11.0
## 9            5.6         5.9         6.7  2048.5      8.0
## 10           5.3         5.5         2.7  1801.7     13.2
```

Output: - The first few rows of the data frame after removing rows with missing `Dopamine` values.

(d) Replace missing values in the c4-OH-Pro column with the median value.

```
# Replace missing values in c4-OH-Pro with the median
median_c4OHPro <- median(metabolites_no_na_dopamine$c4.OH.Pro, na.rm = TRUE)
metabolites_no_na_dopamine$c4.OH.Pro[is.na(metabolites_no_na_dopamine$c4.OH.P
ro)] <- median_c4OHPro
head(metabolites_no_na_dopamine)
```

```
##         Label  Phe Pro Ser Thr ADMA alpha.AAA c4.OH.Pro Carnosine
Creatinine
## 1  Alzheimer 72.8 166 170 282 1.15     0.760     0.236     1.270
49.9
## 4  Alzheimer 94.1 129 162 201 1.10     0.795     0.199     0.675
80.1
## 5  Alzheimer 79.8 126 115 199 1.24     1.360     0.199     1.280
60.5
## 8    Healthy 83.6 119 135 268 1.18     0.779     0.215     0.647
30.6
## 9    Healthy 73.7 124 145 307 1.17     0.785     0.186     0.590
39.8
## 10   Healthy 81.5 151 174 269 1.30     0.767     0.185     1.130
44.0
##      DOPA Dopamine Histamine Kynurenine Met.SO Nitro.Tyr PEA Putrescine
## 1  0.265    0.233     0.225       5.21  0.526     0.027  NA      0.068
## 4  0.264    0.234     0.209       5.80  0.389        NA  NA      0.110
## 5  0.271    0.231     0.210       4.46  0.466        NA  NA      0.118
## 8  0.275    0.244     0.214       5.66  0.245     0.002  NA      0.161
## 9  0.259    0.233     0.210       6.36  0.413        NA  NA      0.121
## 10 0.257    0.225     0.210       5.69  0.178        NA  NA      0.183
##      Sarcosine Serotonin Spermidine Spermine t4.OH.Pro Taurine SDMA   C0
C10
```

```
## 1        17.8       0.147       0.188        NA       24.0      125 1.13 18.2
0.059
## 4        18.7       0.255       0.353        NA       23.1      159 1.34 23.5
0.071
## 5        22.5       0.390       0.473        NA       26.9      149 1.24 13.6
0.139
## 8        23.3       0.215       0.276        NA       10.7      133 1.04 13.3
0.051
## 9        22.1       0.166       0.327        NA       16.0      215 1.24 15.8
0.061
## 10       18.5       0.320       0.236        NA       27.9      140 1.84 15.6
0.068
##    C10.1 C10.2   C12 C12.DC C12.1   C14 C14.1 C14.1.OH C14.2 C14.2.OH
C16
## 1  0.312 0.038 0.030  0.042 0.290 0.023 0.019    0.008 0.008    0.006
0.046
## 4  0.317 0.040 0.045  0.048 0.275 0.026 0.028    0.010 0.013    0.011
0.074
## 5  0.472 0.074 0.056  0.079 0.394 0.034 0.043    0.016 0.025    0.017
0.062
## 8  0.217 0.030 0.041  0.035 0.174 0.024 0.017    0.007 0.006    0.007
0.060
## 9  0.258 0.036 0.037  0.038 0.228 0.022 0.018    0.007 0.007    0.007
0.054
## 10 0.319 0.041 0.039  0.037 0.284 0.026 0.022    0.008 0.013    0.007
0.076
##    C16.OH C16.1 C16.1.OH C16.2 C16.2.OH   C18 C18.1 C18.1.OH C18.2   C2
C3
## 1   0.008 0.009    0.007 0.005    0.013 0.013 0.024    0.003 0.016 1.97
0.354
## 4   0.011 0.015    0.008 0.006    0.009 0.020 0.035    0.004 0.033 2.10
0.278
## 5      NA 0.024    0.014 0.012    0.025 0.031 0.034    0.012 0.017 5.62
0.436
## 8   0.006 0.010    0.005 0.004    0.008 0.020 0.025    0.004 0.019 1.66
0.258
## 9   0.005 0.012    0.005 0.005    0.009 0.014 0.026    0.003 0.016 2.21
0.233
## 10  0.009 0.012    0.007 0.006    0.014 0.024 0.031    0.004 0.028 2.58
0.364
##    C3.OH  C3.1    C4 C3.DC..C4.OH.  C4.1    C5 C5.M.DC C5.OH..C3.DC.M.
C5.1
## 1  0.008 0.015 0.082         0.045 0.025 0.094   0.023           0.026
0.030
## 4  0.010 0.017 0.110         0.077 0.031 0.145   0.034           0.041
0.035
## 5  0.029 0.035 0.106         0.099 0.069 0.141   0.094           0.058
0.073
## 8  0.008 0.012 0.082         0.047 0.021 0.107   0.023           0.023
0.021
```

```
## 9  0.008 0.014 0.088        0.029 0.024 0.127   0.024         0.024
0.025
## 10 0.010 0.015 0.100        0.038 0.025 0.292   0.024         0.023
NA
##    C5.1.DC C6..C4.1.DC. C5.DC..C6.OH.  C6.1 C7.DC    C8    C9
lysoPC.a.C14.0
## 1    0.020        0.022        0.014 0.018 0.011 0.062 0.016
2.23
## 4    0.016        0.029        0.016 0.027 0.017 0.091 0.018
2.19
## 5    0.049        0.052        0.040 0.040 0.036 0.192 0.041
1.88
## 8    0.017        0.036        0.011    NA 0.009 0.062 0.011
2.13
## 9    0.016        0.026        0.018 0.015 0.013 0.064 0.014
2.10
## 10   0.014        0.044        0.017 0.018 0.014 0.067 0.013
2.06
##    lysoPC.a.C16.0 lysoPC.a.C16.1 lysoPC.a.C17.0 lysoPC.a.C18.0
lysoPC.a.C18.1
## 1            37.9           2.66          0.446           9.00
8.58
## 4            32.8           2.39          0.323           7.21
7.22
## 5            24.5           1.27          0.382           6.66
5.39
## 8            33.7           3.09          0.455           6.96
7.31
## 9            36.0           3.46          0.435           7.27
8.11
## 10           32.7           2.48          0.368           6.49
7.86
##    lysoPC.a.C18.2 lysoPC.a.C20.3 lysoPC.a.C20.4 lysoPC.a.C24.0
lysoPC.a.C26.0
## 1            7.27           1.83           8.25          0.079
0.113
## 4            7.62           1.64           6.75          0.066
0.086
## 5            3.60           0.97           6.26          0.084
0.118
## 8            7.53           2.35           8.73          0.061
0.083
## 9            6.75           2.08           7.82          0.068
0.083
## 10          10.10           1.86          12.40          0.061
0.086
##    lysoPC.a.C26.1 lysoPC.a.C28.0 lysoPC.a.C28.1 PC.aa.C24.0 PC.aa.C26.0
## 1           0.053          0.108          0.072       0.082       0.438
## 4           0.045          0.076          0.076       0.076       0.486
## 5           0.053          0.092          0.072       0.069       0.401
```

```
## 8            0.047          0.124          0.078          0.082          0.424
## 9            0.037          0.096          0.061          0.064          0.430
## 10           0.052          0.088          0.074          0.065          0.453
##      PC.aa.C28.1 PC.aa.C30.0 PC.aa.C32.0 PC.aa.C32.1 PC.aa.C32.2 PC.aa.C32.3
## 1          0.571        2.35        11.4        9.22          NA       0.092
## 4          0.685        3.33        18.6       13.30       0.053       0.079
## 5          0.513        1.78        13.8        5.03          NA       0.102
## 8          0.605        2.32        12.1       11.50          NA       0.093
## 9          0.486        2.05        10.4       11.40          NA       0.082
## 10         0.578        2.28        13.8        8.10          NA       0.111
##      PC.aa.C34.1 PC.aa.C34.2 PC.aa.C34.3 PC.aa.C34.4 PC.aa.C36.0 PC.aa.C36.1
## 1          109.0        71.0       1.430       0.200        2.38        21.7
## 4          106.0        93.6       1.590       0.190        2.57        20.9
## 5           83.4        35.9       0.709       0.135        1.83        20.5
## 8           83.6        60.6       1.580       0.251        2.16        18.4
## 9           89.6        55.1       1.390       0.221        1.62        18.3
## 10          80.7        82.6       1.640       0.244        1.75        14.6
##      PC.aa.C36.2 PC.aa.C36.3 PC.aa.C36.4 PC.aa.C36.5 PC.aa.C36.6 PC.aa.C38.0
## 1           42.4        42.7       120.0        1.86       0.084       1.230
## 4           48.8        41.2       122.0        1.76       0.070       1.160
## 5           28.5        21.9        98.1        1.70       0.048       1.100
## 8           34.4        41.5       110.0        2.03       0.106       1.150
## 9           32.0        41.8       102.0        1.77       0.081       0.902
## 10          38.1        31.2       138.0        1.80       0.092       0.904
##      PC.aa.C38.3 PC.aa.C38.4 PC.aa.C38.5 PC.aa.C38.6 PC.aa.C40.1 PC.aa.C40.2
## 1           32.1        95.1        16.8        41.6       0.195       0.074
## 4           28.7        92.7        14.3        29.9       0.220       0.097
## 5           23.3       101.0        13.8        36.2       0.165       0.044
## 8           31.3        81.7        14.5        42.8       0.225       0.099
## 9           31.1        75.9        13.6        37.6       0.181       0.100
## 10          22.2        98.2        12.7        31.5       0.189       0.130
##      PC.aa.C40.3 PC.aa.C40.4 PC.aa.C40.5 PC.aa.C40.6 PC.aa.C42.0 PC.aa.C42.1
## 1          0.491        3.48        5.66        21.8       0.364       0.226
## 4          0.433        3.59        5.06        14.0       0.427       0.223
## 5          0.525        3.37        5.29        22.5       0.125       0.095
## 8          0.384        4.02        5.49        20.6       0.365       0.193
## 9          0.298        2.88        4.88        17.5       0.396       0.216
## 10         0.315        2.87        3.32        12.6       0.301       0.186
##      PC.aa.C42.2 PC.aa.C42.4 PC.aa.C42.5 PC.aa.C42.6 PC.ae.C30.0 PC.ae.C30.1
## 1          0.108       0.272       0.272       0.291       0.173       0.027
## 4          0.119       0.268       0.267       0.254       0.223       0.049
## 5          0.083       0.206       0.205       0.280       0.095       0.082
## 8          0.089       0.266       0.237       0.288       0.212       0.028
## 9          0.086       0.215       0.206       0.223       0.179       0.012
## 10         0.092       0.228       0.199       0.224       0.164       0.017
##      PC.ae.C30.2 PC.ae.C32.1 PC.ae.C32.2 PC.ae.C34.0 PC.ae.C34.1 PC.ae.C34.2
## 1          0.022        1.65       0.371       0.880        3.66        2.48
## 4          0.023        2.47       0.459       0.964        4.06        3.09
## 5          0.023        1.72       0.316       1.060        3.28        1.70
## 8          0.033        1.68       0.335       1.140        3.51        2.28
```

```
## 9        0.022          1.53          0.291          0.683          3.20          2.03
## 10       0.024          1.70          0.341          0.800          3.56          2.52
##      PC.ae.C34.3 PC.ae.C36.0 PC.ae.C36.1 PC.ae.C36.2 PC.ae.C36.3 PC.ae.C36.4
## 1        0.813       0.498        5.64          1.90        1.17          6.96
## 4        1.020       0.552        4.75          2.01        1.35          8.36
## 5        0.722       0.553        5.95          1.47        0.76          4.78
## 8        0.868       0.479        5.65          1.97        1.25          7.15
## 9        0.663       0.384        4.79          1.54        1.12          6.43
## 10       0.983       0.498        4.85          1.93        1.04          7.08
##      PC.ae.C36.5 PC.ae.C38.0 PC.ae.C38.1 PC.ae.C38.2 PC.ae.C38.3 PC.ae.C38.4
## 1        4.79        0.474        0.287        0.538        2.66          6.33
## 4        5.97        0.397        0.022        0.144        1.97          5.99
## 5        4.00        0.430        0.271        0.246        1.80          5.45
## 8        4.04        0.550        0.078        0.559        2.90          5.73
## 9        3.32        0.442          NA         0.224        2.53          5.03
## 10       4.52        0.458          NA         0.053        1.99          5.79
##      PC.ae.C38.5 PC.ae.C38.6 PC.ae.C40.1 PC.ae.C40.2 PC.ae.C40.3 PC.ae.C40.4
## 1        5.51        1.95        0.574        0.575        0.940          1.76
## 4        5.63        1.97        0.425        0.540        0.742          1.45
## 5        4.34        1.51        0.430        0.432        0.632          1.10
## 8        4.53        1.71        0.584        0.499        0.991          1.84
## 9        3.82        1.28        0.472        0.555        0.737          1.35
## 10       4.84        1.58        0.670        0.572        0.757          1.16
##      PC.ae.C40.5 PC.ae.C40.6 PC.ae.C42.0 PC.ae.C42.1 PC.ae.C42.2 PC.ae.C42.3
## 1        1.77        1.59        0.629        0.316        0.192        0.277
## 4        1.62        1.25        0.637        0.299        0.159        0.208
## 5        1.25        1.47        0.660        0.355        0.138        0.174
## 8        1.53        1.46        0.631        0.271        0.186        0.282
## 9        1.34        1.21        0.645        0.264        0.168        0.229
## 10       1.30        1.18        0.670        0.294        0.186        0.222
##      PC.ae.C42.4 PC.ae.C42.5 PC.ae.C44.3 PC.ae.C44.4 PC.ae.C44.5 PC.ae.C44.6
## 1        0.264       0.888       0.065        0.168        0.536        0.494
## 4        0.392       0.863       0.069        0.237        0.517        0.611
## 5        0.162       0.513       0.081        0.154        0.178        0.134
## 8        0.333       0.861       0.085        0.183        0.534        0.429
## 9        0.319       0.925       0.069        0.184        0.671        0.460
## 10       0.266       0.660       0.080        0.156        0.420        0.399
##      SM..OH..C14.1 SM..OH..C16.1 SM..OH..C22.1 SM..OH..C22.2 SM..OH..C24.1
## 1        1.420         1.330         2.07          1.86          0.597
## 4        1.720         1.480         2.97          2.84          0.682
## 5        0.987         1.480         1.96          1.74          0.478
## 8        1.500         1.160         2.98          2.59          0.631
## 9        1.230         0.984         2.21          2.07          0.507
## 10       1.760         1.450         2.60          2.45          0.586
##      SM.C16.0 SM.C16.1 SM.C18.0 SM.C18.1 SM.C20.2 SM.C24.0 SM.C24.1 SM.C26.0
## 1      44.9     7.99     14.5     10.40    0.290    12.20     27.3     0.147
## 4      52.4     8.61     17.2     11.50    0.261    11.80     27.9     0.138
## 5      40.6     5.86     13.0      8.34    0.196     9.29     20.5     0.111
## 8      37.9     7.92     11.9      9.59    0.199     9.36     18.8     0.117
## 9      37.3     6.65     12.0      8.48    0.183    10.30     23.6     0.090
```

```
## 10      48.3      9.66      15.4     11.90     0.295     10.50      26.7     0.097
##     SM.C26.1 H1_1   H1 Urea_N L.Arginine_N L.Leucine_N EDTAca_N
## 1     0.337 3356 3356    NA          NA          NA        NA
## 4     0.353 2652 2652 500.8        16.0        27.1       2.5
## 5     0.283 2258 2258 132.5        13.2        57.9       2.5
## 8     0.227 2464 2464 182.3        49.8        63.9       0.0
## 9     0.268 2725 2725 255.8        39.0        64.6       2.7
## 10    0.305 3583 3583 184.4        66.0        54.6       2.3
##    X2.Hydroxybutyrate X3.Hydroxybutyrate Acetate Acetoacetate Acetone
Betaine
## 1                  NA                 NA      NA           NA      NA
NA
## 4                12.7                7.2     9.8          4.8     4.0
13.9
## 5                35.2               44.7    20.2         18.9    18.9
33.9
## 8                21.7               20.9    19.5         15.4     6.6
35.3
## 9                17.1               37.9    20.0         22.0     9.3
12.0
## 10               21.3               46.2    22.5         26.2     9.4
29.5
##    Carnitine Choline Creatine Dimethyl.sulfone Ethanol Formate Glucose
Glycerol
## 1        NA      NA       NA               NA      NA      NA      NA
NA
## 4       7.7    11.8     14.7              1.3     6.4    14.4   629.5
322.0
## 5      18.5    27.7     35.4              5.5    13.0    40.0  1618.0
271.6
## 8      13.0    21.5     25.9              5.4    10.2    23.2  2120.5
406.5
## 9      14.4    15.4     25.3              3.5     5.1    25.0  1644.2
479.9
## 10     15.7    28.1     25.6              2.2     4.4    29.4  2875.0
722.2
##    Hypoxanthine Isobutyrate Isopropanol Lactate Malonate
## 1            NA          NA          NA      NA       NA
## 4           8.6         2.5         4.4  1037.7      7.6
## 5           0.0         6.1        11.2  2199.9     11.7
## 8           5.8         4.5         4.4  1661.4     11.0
## 9           5.6         5.9         6.7  2048.5      8.0
## 10          5.3         5.5         2.7  1801.7     13.2
```

Output: - The first few rows after replacing missing values in the c4-OH-Pro column with the median.

(e) (Optional) Drop columns with more than 25% missing values.

```
# Drop columns with more than 25% missing values
threshold <- 0.25 * nrow(metabolites_data)
```

```
metabolites_clean <- metabolites_data[, colSums(is.na(metabolites_data)) <=
threshold]
head(metabolites_clean)

##        Label  Phe Pro Ser Thr ADMA alpha.AAA Carnosine Creatinine  DOPA
## 1 Alzheimer 72.8 166 170 282 1.15     0.760     1.270       49.9 0.265
## 2 Alzheimer 93.4 138 142 217 1.05     0.929     1.350       48.8 0.252
## 3 Alzheimer 68.6 161 158 208 1.00     0.620     0.998       30.4 0.268
## 4 Alzheimer 94.1 129 162 201 1.10     0.795     0.675       80.1 0.264
## 5 Alzheimer 79.8 126 115 199 1.24     1.360     1.280       60.5 0.271
## 6 Alzheimer 82.5 167 173 333 1.35     1.150     1.010       24.0 0.275
##   Histamine Kynurenine Met.SO Putrescine Sarcosine Serotonin Spermidine
## 1     0.225       5.21  0.526      0.068      17.8     0.147      0.188
## 2     0.211       5.44  0.387      0.087      20.2     0.231      0.233
## 3     0.217       5.20  0.651      0.260      14.4     0.196      0.384
## 4     0.209       5.80  0.389      0.110      18.7     0.255      0.353
## 5     0.210       4.46  0.466      0.118      22.5     0.390      0.473
## 6     0.212       7.01  0.417      0.262      30.8     0.140      0.856
##   t4.OH.Pro Taurine SDMA   C0   C10 C10.1 C10.2   C12 C12.DC C12.1   C14
C14.1
## 1      24.0     125 1.13 18.2 0.059 0.312 0.038 0.030  0.042 0.290 0.023
0.019
## 2      29.3     120 1.65 17.0 0.051 0.288 0.039 0.038  0.038 0.265 0.026
0.017
## 3      20.9     139 1.57 12.6 0.083 0.357 0.054 0.032  0.048 0.302 0.021
0.031
## 4      23.1     159 1.34 23.5 0.071 0.317 0.040 0.045  0.048 0.275 0.026
0.028
## 5      26.9     149 1.24 13.6 0.139 0.472 0.074 0.056  0.079 0.394 0.034
0.043
## 6      26.0     379 1.44 26.7 0.058 0.238 0.042 0.039  0.035 0.196 0.029
0.023
##   C14.1.OH C14.2 C14.2.OH   C16 C16.OH C16.1 C16.1.OH C16.2 C16.2.OH   C18
## 1    0.008 0.008    0.006 0.046  0.008 0.009    0.007 0.005    0.013 0.013
## 2    0.008 0.009    0.009 0.070  0.009 0.013    0.006 0.006    0.012 0.014
## 3    0.010 0.010    0.009 0.076  0.011 0.019    0.010 0.005    0.013 0.016
## 4    0.010 0.013    0.011 0.074  0.011 0.015    0.008 0.006    0.009 0.020
## 5    0.016 0.025    0.017 0.062     NA 0.024    0.014 0.012    0.025 0.031
## 6    0.009 0.010    0.007 0.081  0.006 0.012    0.005 0.007    0.015 0.017
##   C18.1 C18.1.OH C18.2   C2    C3 C3.OH  C3.1    C4 C3.DC..C4.OH.  C4.1
C5
## 1 0.024    0.003 0.016 1.97 0.354 0.008 0.015 0.082         0.045 0.025
0.094
## 2 0.025    0.003 0.028 1.95 0.184 0.009 0.013 0.108         0.080 0.025
0.077
## 3 0.025       NA 0.018 1.70 0.371    NA 0.012 0.057         0.035 0.039
0.096
## 4 0.035    0.004 0.033 2.10 0.278 0.010 0.017 0.110         0.077 0.031
0.145
## 5 0.034    0.012 0.017 5.62 0.436 0.029 0.035 0.106         0.099 0.069
```

```
0.141
## 6 0.035    0.004 0.029 3.49 0.461 0.008 0.014 0.123        0.068 0.026
0.090
##    C5.M.DC C5.OH..C3.DC.M.  C5.1 C5.1.DC C6..C4.1.DC. C5.DC..C6.OH.  C6.1
C7.DC
## 1   0.023            0.026 0.030   0.020        0.022         0.014 0.018
0.011
## 2   0.032            0.026 0.024   0.021        0.030         0.018 0.015
0.010
## 3   0.045            0.024 0.037   0.018        0.022         0.029 0.031
0.021
## 4   0.034            0.041 0.035   0.016        0.029         0.016 0.027
0.017
## 5   0.094            0.058 0.073   0.049        0.052         0.040 0.040
0.036
## 6   0.019            0.037 0.022   0.016        0.063         0.016 0.019
0.014
##        C8    C9 lysoPC.a.C14.0 lysoPC.a.C16.0 lysoPC.a.C16.1 lysoPC.a.C17.0
## 1 0.062 0.016           2.23           37.9           2.66          0.446
## 2 0.058 0.014           1.97           22.1           1.31          0.270
## 3 0.090 0.017           2.12           33.7           2.53          0.399
## 4 0.091 0.018           2.19           32.8           2.39          0.323
## 5 0.192 0.041           1.88           24.5           1.27          0.382
## 6 0.073 0.014           2.11           29.1           2.09          0.348
##    lysoPC.a.C18.0 lysoPC.a.C18.1 lysoPC.a.C18.2 lysoPC.a.C20.3
lysoPC.a.C20.4
## 1           9.00           8.58           7.27          1.830
8.25
## 2           5.35           3.94           4.42          0.958
4.60
## 3           7.51           7.73           8.02          2.050
9.84
## 4           7.21           7.22           7.62          1.640
6.75
## 5           6.66           5.39           3.60          0.970
6.26
## 6           5.84           6.30           8.10          1.970
7.04
##    lysoPC.a.C24.0 lysoPC.a.C26.0 lysoPC.a.C26.1 lysoPC.a.C28.0
lysoPC.a.C28.1
## 1           0.079          0.113          0.053          0.108
0.072
## 2           0.059          0.066          0.042          0.076
0.058
## 3           0.075          0.126          0.049          0.078
0.092
## 4           0.066          0.086          0.045          0.076
0.076
## 5           0.084          0.118          0.053          0.092
0.072
```

```
## 6        0.083           0.112           0.050           0.099
0.083
##    PC.aa.C24.0 PC.aa.C26.0 PC.aa.C28.1 PC.aa.C30.0 PC.aa.C32.0 PC.aa.C32.1
## 1        0.082       0.438       0.571        2.35        11.4        9.22
## 2        0.065       0.409       0.521        1.99        12.7        5.40
## 3        0.099       0.458       0.605        2.69        16.6       11.60
## 4        0.076       0.486       0.685        3.33        18.6       13.30
## 5        0.069       0.401       0.513        1.78        13.8        5.03
## 6        0.073       0.450       0.620        2.61        14.7        8.98
##    PC.aa.C32.3 PC.aa.C34.1 PC.aa.C34.2 PC.aa.C34.3 PC.aa.C34.4 PC.aa.C36.0
## 1        0.092       109.0        71.0       1.430       0.200        2.38
## 2        0.067        64.2        60.5       0.879       0.127        2.05
## 3        0.105       108.0        83.1       1.930       0.210        2.30
## 4        0.079       106.0        93.6       1.590       0.190        2.57
## 5        0.102        83.4        35.9       0.709       0.135        1.83
## 6        0.107        90.2        85.6       1.790       0.213        2.48
##    PC.aa.C36.1 PC.aa.C36.2 PC.aa.C36.3 PC.aa.C36.4 PC.aa.C36.5 PC.aa.C36.6
## 1        21.7        42.4        42.7       120.0        1.86       0.084
## 2        14.3        35.6        24.3        83.7        1.05       0.046
## 3        19.9        44.9        43.9       146.0        2.09       0.057
## 4        20.9        48.8        41.2       122.0        1.76       0.070
## 5        20.5        28.5        21.9        98.1        1.70       0.048
## 6        15.5        43.2        46.0       114.0        3.47       0.103
##    PC.aa.C38.0 PC.aa.C38.3 PC.aa.C38.4 PC.aa.C38.5 PC.aa.C38.6 PC.aa.C40.1
## 1        1.230        32.1        95.1       16.80        41.6       0.195
## 2        0.946        21.9        78.9        9.91        25.1       0.211
## 3        1.210        34.5       107.0       17.50        36.6       0.212
## 4        1.160        28.7        92.7       14.30        29.9       0.220
## 5        1.100        23.3       101.0       13.80        36.2       0.165
## 6        1.390        28.9        78.0       13.10        48.4       0.205
##    PC.aa.C40.2 PC.aa.C40.3 PC.aa.C40.4 PC.aa.C40.5 PC.aa.C40.6 PC.aa.C42.0
## 1        0.074       0.491        3.48        5.66        21.8       0.364
## 2        0.057       0.358        3.39        4.08        14.2       0.419
## 3        0.118       0.395        3.56        5.34        16.7       0.476
## 4        0.097       0.433        3.59        5.06        14.0       0.427
## 5        0.044       0.525        3.37        5.29        22.5       0.125
## 6        0.120       0.346        2.63        3.25        18.9       0.451
##    PC.aa.C42.1 PC.aa.C42.2 PC.aa.C42.4 PC.aa.C42.5 PC.aa.C42.6 PC.ae.C30.0
## 1        0.226       0.108       0.272       0.272       0.291       0.173
## 2        0.216       0.109       0.336       0.317       0.248       0.147
## 3        0.281       0.118       0.300       0.206       0.267       0.209
## 4        0.223       0.119       0.268       0.267       0.254       0.223
## 5        0.095       0.083       0.206       0.205       0.280       0.095
## 6        0.233       0.135       0.228       0.254       0.271       0.221
##    PC.ae.C30.1 PC.ae.C30.2 PC.ae.C32.1 PC.ae.C32.2 PC.ae.C34.0 PC.ae.C34.1
## 1        0.027       0.022        1.65       0.371       0.880        3.66
## 2        0.024       0.020        2.01       0.360       0.763        2.68
## 3        0.046       0.030        2.40       0.477       0.938        4.04
## 4        0.049       0.023        2.47       0.459       0.964        4.06
## 5        0.082       0.023        1.72       0.316       1.060        3.28
```

```
## 6       0.039        0.029        2.01        0.397        0.920        3.26
##    PC.ae.C34.2 PC.ae.C34.3 PC.ae.C36.0 PC.ae.C36.1 PC.ae.C36.2 PC.ae.C36.3
## 1        2.48       0.813       0.498        5.64        1.90       1.170
## 2        2.32       0.905       0.398        3.89        1.54       0.873
## 3        2.95       1.030       0.554        5.95        2.29       1.240
## 4        3.09       1.020       0.552        4.75        2.01       1.350
## 5        1.70       0.722       0.553        5.95        1.47       0.760
## 6        2.58       1.000       0.443        4.95        2.05       1.170
##    PC.ae.C36.4 PC.ae.C36.5 PC.ae.C38.0 PC.ae.C38.3 PC.ae.C38.4 PC.ae.C38.5
## 1        6.96        4.79       0.474        2.66        6.33        5.51
## 2        6.40        5.36       0.325        1.80        5.37        4.49
## 3        9.05        6.63       0.478        2.87        7.06        5.64
## 4        8.36        5.97       0.397        1.97        5.99        5.63
## 5        4.78        4.00       0.430        1.80        5.45        4.34
## 6        7.04        4.47       0.590        2.46        5.55        4.60
##    PC.ae.C38.6 PC.ae.C40.1 PC.ae.C40.2 PC.ae.C40.3 PC.ae.C40.4 PC.ae.C40.5
## 1        1.95       0.574       0.575       0.940        1.76        1.77
## 2        1.63       0.281       0.491       0.702        1.43        1.55
## 3        1.98       0.759       0.654       0.817        1.51        1.64
## 4        1.97       0.425       0.540       0.742        1.45        1.62
## 5        1.51       0.430       0.432       0.632        1.10        1.25
## 6        1.80       0.481       0.598       0.826        1.25        1.38
##    PC.ae.C40.6 PC.ae.C42.0 PC.ae.C42.1 PC.ae.C42.2 PC.ae.C42.3 PC.ae.C42.4
## 1        1.59       0.629       0.316       0.192       0.277       0.264
## 2        1.20       0.616       0.260       0.157       0.200       0.311
## 3        1.49       0.686       0.356       0.241       0.288       0.319
## 4        1.25       0.637       0.299       0.159       0.208       0.392
## 5        1.47       0.660       0.355       0.138       0.174       0.162
## 6        1.61       0.669       0.265       0.195       0.253       0.316
##    PC.ae.C42.5 PC.ae.C44.3 PC.ae.C44.4 PC.ae.C44.5 PC.ae.C44.6
SM..OH..C14.1
## 1       0.888       0.065       0.168       0.536       0.494
1.420
## 2       0.840       0.071       0.220       0.470       0.515
1.390
## 3       0.957       0.065       0.228       0.565       0.603
1.840
## 4       0.863       0.069       0.237       0.517       0.611
1.720
## 5       0.513       0.081       0.154       0.178       0.134
0.987
## 6       0.814       0.085       0.232       0.554       0.539
1.320
##    SM..OH..C16.1 SM..OH..C22.1 SM..OH..C22.2 SM..OH..C24.1 SM.C16.0
SM.C16.1
## 1         1.33          2.07          1.86         0.597     44.9
7.99
## 2         1.25          2.47          2.20         0.640     42.1
6.88
## 3         1.58          2.69          2.63         0.665     44.8
```

```
8.91
## 4           1.48            2.97            2.84           0.682      52.4
8.61
## 5           1.48            1.96            1.74           0.478      40.6
5.86
## 6           1.12            2.51            2.16           0.640      42.6
8.49
##    SM.C18.0 SM.C18.1 SM.C20.2 SM.C24.0 SM.C24.1 SM.C26.0 SM.C26.1 H1_1    H1
## 1     14.5    10.40    0.290    12.20     27.3    0.147    0.337 3356 3356
## 2     12.7     8.52    0.211    10.40     25.6    0.130    0.317 2509 2509
## 3     14.6    11.60    0.304    11.50     28.8    0.163    0.364 2661 2661
## 4     17.2    11.50    0.261    11.80     27.9    0.138    0.353 2652 2652
## 5     13.0     8.34    0.196     9.29     20.5    0.111    0.283 2258 2258
## 6     13.0    10.60    0.270     9.58     23.7    0.135    0.316 3031 3031
##    Urea_N L.Arginine_N L.Leucine_N EDTAca_N X2.Hydroxybutyrate
## 1     NA           NA          NA       NA                 NA
## 2  201.9         22.5        35.3      2.0              12.40
## 3  193.3         21.0        25.4      1.8              11.33
## 4  500.8         16.0        27.1      2.5              12.70
## 5  132.5         13.2        57.9      2.5              35.20
## 6  193.3         32.2        26.5      0.0              17.20
##    X3.Hydroxybutyrate Acetate Acetoacetate Acetone Betaine Carnitine
Choline
## 1                 NA      NA           NA      NA      NA        NA
NA
## 2                8.5    13.2          5.7     5.1    22.0       8.7
14.2
## 3               11.7     5.8          9.3     5.6    19.1      15.3
14.5
## 4                7.2     9.8          4.8     4.0    13.9       7.7
11.8
## 5               44.7    20.2         18.9    18.9    33.9      18.5
27.7
## 6               16.0    23.6          7.8     5.5    16.9      16.7
25.9
##    Creatine Dimethyl.sulfone Ethanol Formate Glucose Glycerol Hypoxanthine
## 1       NA               NA      NA      NA      NA       NA           NA
## 2     14.5              4.7    16.6    24.6  1489.7    324.6          6.3
## 3     17.8              2.1     8.1    27.4  1343.9    201.3          6.0
## 4     14.7              1.3     6.4    14.4   629.5    322.0          8.6
## 5     35.4              5.5    13.0    40.0  1618.0    271.6          0.0
## 6     18.6              3.4     5.0    35.5  1791.8    274.2          8.8
##    Isobutyrate Isopropanol Lactate Malonate
## 1          NA          NA      NA       NA
## 2         3.6         1.9  1171.6     10.4
## 3         2.5         2.5  1938.1     13.1
## 4         2.5         4.4  1037.7      7.6
## 5         6.1        11.2  2199.9     11.7
## 6         2.3         2.4  1486.7     11.8
```

Output: - The first few rows after dropping columns with more than 25% missing values.