# PEC 1 Análisis de datos ómicos

Ricardo Zamarreño Martins

2024-10-29

## Contents



### Abstract

En este documento se recoge la creación de un contenedor de tipo SummarizedExperiment que contenga los datos y metadas acerca del dataset elegido. En este caso, el dataset que se ha elegido ha sido: **human_cachexia.csv**. En este apartado incluimos el archivo description.md que se incluía en el dataset:

**This is the famous cachexia dataset used in several MetaboAnalyst tutorials Available from:**

https://rest.xialab.ca/api/download/metaboanalyst/human_cachexia.csv

- Successfully passed sanity check!
- Samples are not paired.
- 2 groups were detected in samples.
- All data values are numeric.
- A total of 0 (0%) missing values were detected.

Este dataset incluye información relacionada con la enfermedad humana **Caquexia**, aportando datos sobre la pérdida de músculo y las diferentes concentraciones de metabolitos.

Lo primero que haremos será descargar los datos, visualizarlos, revisarlos y crear el contenedor. Para ello nos vamos a valer de las bibliotecas SummarizedExperiment y readr. Posteriormente se realizará la creación de un repositorio en GitHub donde se incluirá toda la información y los archivos creados.

## Configuración Inicial

Primero, cargamos las librería necesarias. Estas librería serán imprescindibles para poder realizar el trabajo, ya que las sentencias utilizadas para la carga de datos y la creación del contenedor pertenecen a dichas librerías.

```
library(SummarizedExperiment)
library(readr)
```

## Carga de Datos

Descargamos y leemos el archivo CSV que contiene los datos metabolómicos, este caso, "human_cachexia.csv".

```
# Descargar los datos del repositorio
# Se realiza la descarga del dataset del repositorio

#"https://github.com/nutrimetabolomics/metaboData/blob/main/Datasets/2024-Cachexia/human_cachexia.csv"

# Leer los datos
data <- read_csv("human_cachexia.csv")
```

```
## Rows: 77 Columns: 65
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): Patient ID, Muscle loss
## dbl (63): 1,6-Anhydro-beta-D-glucose, 1-Methylnicotinamide, 2-Aminobutyrate,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Exploración de Datos

En este apartado vamos a realizar una exploración, un vistazo incial, de que datos se incluyen en el dataset. De esta forma podremos tener una visión general de cómo se estructuran los datos y que información nos proporcionan.

```
# Primeras filas del dataset
head(data)
```

```
## # A tibble: 6 x 65
##   `Patient ID` `Muscle loss` `1,6-Anhydro-beta-D-glucose` `1-Methylnicotinamide`
##   <chr>        <chr>                               <dbl>                  <dbl>
## 1 PIF_178      cachexic                             40.8                   65.4
## 2 PIF_087      cachexic                             62.2                  340.
## 3 PIF_090      cachexic                            270.                    64.7
## 4 NETL_005_V1  cachexic                            154.                    53.0
## 5 PIF_115      cachexic                             22.2                   73.7
## 6 PIF_110      cachexic                            213.                    31.8
## # i 61 more variables: `2-Aminobutyrate` <dbl>, `2-Hydroxyisobutyrate` <dbl>,
## #   `2-Oxoglutarate` <dbl>, `3-Aminoisobutyrate` <dbl>,
```

```
## #   `3-Hydroxybutyrate` <dbl>, `3-Hydroxyisovalerate` <dbl>,
## #   `3-Indoxylsulfate` <dbl>, `4-Hydroxyphenylacetate` <dbl>, Acetate <dbl>,
## #   Acetone <dbl>, Adipate <dbl>, Alanine <dbl>, Asparagine <dbl>,
## #   Betaine <dbl>, Carnitine <dbl>, Citrate <dbl>, Creatine <dbl>,
## #   Creatinine <dbl>, Dimethylamine <dbl>, Ethanolamine <dbl>, ...
```

Cpn el comando head() podemos ver las primeras filas del dataset, permitiéndonos hacernos una idea de cómo están organizados los datos y qué se va a incluir en la información.

```
# Estructura del dataset
str(data)
```

```
## spc_tbl_ [77 x 65] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Patient ID                : chr [1:77] "PIF_178" "PIF_087" "PIF_090" "NETL_005_V1" ...
##  $ Muscle loss               : chr [1:77] "cachexic" "cachexic" "cachexic" "cachexic" ...
##  $ 1,6-Anhydro-beta-D-glucose: num [1:77] 40.9 62.2 270.4 154.5 22.2 ...
##  $ 1-Methylnicotinamide      : num [1:77] 65.4 340.4 64.7 53 73.7 ...
##  $ 2-Aminobutyrate           : num [1:77] 18.7 24.3 12.2 172.4 15.6 ...
##  $ 2-Hydroxyisobutyrate      : num [1:77] 26.1 41.7 65.4 74.4 83.9 ...
##  $ 2-Oxoglutarate            : num [1:77] 71.5 67.4 23.8 1199.9 33.1 ...
##  $ 3-Aminoisobutyrate        : num [1:77] 1480.3 116.8 14.3 555.6 29.7 ...
##  $ 3-Hydroxybutyrate         : num [1:77] 56.83 43.82 5.64 175.91 76.71 ...
##  $ 3-Hydroxyisovalerate      : num [1:77] 10.1 79.8 23.3 25 69.4 ...
##  $ 3-Indoxylsulfate          : num [1:77] 567 369 665 412 166 ...
##  $ 4-Hydroxyphenylacetate    : num [1:77] 120.3 432.7 292.9 214.9 97.5 ...
##  $ Acetate                   : num [1:77] 126.5 212.7 314.2 37.3 407.5 ...
##  $ Acetone                   : num [1:77] 9.49 11.82 4.44 206.44 44.26 ...
##  $ Adipate                   : num [1:77] 38.1 327 131.6 144 15 ...
##  $ Alanine                   : num [1:77] 314 871 464 590 1119 ...
##  $ Asparagine                : num [1:77] 159.2 157.6 89.1 273.1 42.5 ...
##  $ Betaine                   : num [1:77] 110 245 117 279 392 ...
##  $ Carnitine                 : num [1:77] 265.1 120.3 25 200.3 84.8 ...
##  $ Citrate                   : num [1:77] 3714 2618 863 13630 854 ...
##  $ Creatine                  : num [1:77] 196.4 212.7 221.4 85.6 105.6 ...
##  $ Creatinine                : num [1:77] 16482 15835 24588 20952 6768 ...
##  $ Dimethylamine             : num [1:77] 633 608 735 1064 242 ...
##  $ Ethanolamine              : num [1:77] 645 488 407 821 365 ...
##  $ Formate                   : num [1:77] 441 252 250 469 114 ...
##  $ Fucose                    : num [1:77] 337 198.3 186.8 407.5 26.1 ...
##  $ Fumarate                  : num [1:77] 7.69 18.92 7.1 96.54 19.69 ...
##  $ Glucose                   : num [1:77] 395 8691 1353 863 6836 ...
##  $ Glutamine                 : num [1:77] 871 602 302 1686 433 ...
##  $ Glycine                   : num [1:77] 2039 1108 620 5064 395 ...
##  $ Glycolate                 : num [1:77] 685.4 652 141.2 70.8 26.6 ...
##  $ Guanidoacetate            : num [1:77] 154 110 183 103 53 ...
##  $ Hippurate                 : num [1:77] 4582 1737 4316 757 1153 ...
##  $ Histidine                 : num [1:77] 925 846 284 1043 327 ...
##  $ Hypoxanthine              : num [1:77] 97.5 82.3 114.4 223.6 66.7 ...
##  $ Isoleucine                : num [1:77] 5.58 8.17 9.3 37.71 40.04 ...
##  $ Lactate                   : num [1:77] 107 369 750 369 3641 ...
##  $ Leucine                   : num [1:77] 42.1 77.5 31.5 103.5 101.5 ...
##  $ Lysine                    : num [1:77] 146.9 284.3 97.5 290 122.7 ...
##  $ Methylamine               : num [1:77] 52.5 23.6 18.7 48.9 27.9 ...
##  $ Methylguanidine           : num [1:77] 9.97 7.69 4.66 141.17 5.31 ...
```

```
##  $ N,N-Dimethylglycine       : num [1:77] 23.3 87.4 24.5 40 46.1 ...
##  $ O-Acetylcarnitine         : num [1:77] 52.98 50.4 5.58 254.68 45.6 ...
##  $ Pantothenate              : num [1:77] 25.8 186.8 145.5 42.5 74.4 ...
##  $ Pyroglutamate             : num [1:77] 437 437 713 567 185 ...
##  $ Pyruvate                  : num [1:77] 21.1 37 29.4 64.1 12.3 ...
##  $ Quinolinate               : num [1:77] 165.7 73 192.5 86.5 38.1 ...
##  $ Serine                    : num [1:77] 284 392 296 1249 206 ...
##  $ Succinate                 : num [1:77] 154.5 244.7 142.6 144 68.7 ...
##  $ Sucrose                   : num [1:77] 45.1 459.4 160.8 111 75.2 ...
##  $ Tartrate                  : num [1:77] 97.51 32.79 16.28 837.15 4.53 ...
##  $ Taurine                   : num [1:77] 1920 1261 4273 1525 469 ...
##  $ Threonine                 : num [1:77] 184.9 198.3 110 376.1 64.1 ...
##  $ Trigonelline              : num [1:77] 943.9 208.5 192.5 992.3 86.5 ...
##  $ Trimethylamine N-oxide    : num [1:77] 2122 639 1153 1451 172 ...
##  $ Tryptophan                : num [1:77] 259.8 83.1 82.3 235.1 103.5 ...
##  $ Tyrosine                  : num [1:77] 290 167.3 60.3 323.8 142.6 ...
##  $ Uracil                    : num [1:77] 111 47 31.5 30.6 44.3 ...
##  $ Valine                    : num [1:77] 86.5 110 59.1 102.5 160.8 ...
##  $ Xylose                    : num [1:77] 72.2 192.5 2164.6 125.2 186.8 ...
##  $ cis-Aconitate             : num [1:77] 237 334 330 1863 101 ...
##  $ myo-Inositol              : num [1:77] 135.6 376.1 86.5 247.2 750 ...
##  $ trans-Aconitate           : num [1:77] 51.9 217 58.6 75.9 98.5 ...
##  $ pi-Methylhistidine        : num [1:77] 157.6 308 145.5 249.6 84.8 ...
##  $ tau-Methylhistidine       : num [1:77] 160.8 130.3 83.9 254.7 79.8 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   `Patient ID` = col_character(),
##   ..   `Muscle loss` = col_character(),
##   ..   `1,6-Anhydro-beta-D-glucose` = col_double(),
##   ..   `1-Methylnicotinamide` = col_double(),
##   ..   `2-Aminobutyrate` = col_double(),
##   ..   `2-Hydroxyisobutyrate` = col_double(),
##   ..   `2-Oxoglutarate` = col_double(),
##   ..   `3-Aminoisobutyrate` = col_double(),
##   ..   `3-Hydroxybutyrate` = col_double(),
##   ..   `3-Hydroxyisovalerate` = col_double(),
##   ..   `3-Indoxylsulfate` = col_double(),
##   ..   `4-Hydroxyphenylacetate` = col_double(),
##   ..   Acetate = col_double(),
##   ..   Acetone = col_double(),
##   ..   Adipate = col_double(),
##   ..   Alanine = col_double(),
##   ..   Asparagine = col_double(),
##   ..   Betaine = col_double(),
##   ..   Carnitine = col_double(),
##   ..   Citrate = col_double(),
##   ..   Creatine = col_double(),
##   ..   Creatinine = col_double(),
##   ..   Dimethylamine = col_double(),
##   ..   Ethanolamine = col_double(),
##   ..   Formate = col_double(),
##   ..   Fucose = col_double(),
##   ..   Fumarate = col_double(),
##   ..   Glucose = col_double(),
```

```
##   ..   Glutamine = col_double(),
##   ..   Glycine = col_double(),
##   ..   Glycolate = col_double(),
##   ..   Guanidoacetate = col_double(),
##   ..   Hippurate = col_double(),
##   ..   Histidine = col_double(),
##   ..   Hypoxanthine = col_double(),
##   ..   Isoleucine = col_double(),
##   ..   Lactate = col_double(),
##   ..   Leucine = col_double(),
##   ..   Lysine = col_double(),
##   ..   Methylamine = col_double(),
##   ..   Methylguanidine = col_double(),
##   ..   `N,N-Dimethylglycine` = col_double(),
##   ..   `O-Acetylcarnitine` = col_double(),
##   ..   Pantothenate = col_double(),
##   ..   Pyroglutamate = col_double(),
##   ..   Pyruvate = col_double(),
##   ..   Quinolinate = col_double(),
##   ..   Serine = col_double(),
##   ..   Succinate = col_double(),
##   ..   Sucrose = col_double(),
##   ..   Tartrate = col_double(),
##   ..   Taurine = col_double(),
##   ..   Threonine = col_double(),
##   ..   Trigonelline = col_double(),
##   ..   `Trimethylamine N-oxide` = col_double(),
##   ..   Tryptophan = col_double(),
##   ..   Tyrosine = col_double(),
##   ..   Uracil = col_double(),
##   ..   Valine = col_double(),
##   ..   Xylose = col_double(),
##   ..   `cis-Aconitate` = col_double(),
##   ..   `myo-Inositol` = col_double(),
##   ..   `trans-Aconitate` = col_double(),
##   ..   `pi-Methylhistidine` = col_double(),
##   ..   `tau-Methylhistidine` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Con el comando str(), podremos ver la estructura de los datos. Vemos que se trata de una tabla de 77 filas por 65 columnas, o lo que es lo mismo, 77 casos para 65 variables.

```
# Nombres de las columnas
colnames(data)
```

```
##  [1] "Patient ID"                 "Muscle loss"
##  [3] "1,6-Anhydro-beta-D-glucose" "1-Methylnicotinamide"
##  [5] "2-Aminobutyrate"            "2-Hydroxyisobutyrate"
##  [7] "2-Oxoglutarate"             "3-Aminoisobutyrate"
##  [9] "3-Hydroxybutyrate"          "3-Hydroxyisovalerate"
## [11] "3-Indoxylsulfate"           "4-Hydroxyphenylacetate"
## [13] "Acetate"                    "Acetone"
## [15] "Adipate"                    "Alanine"
```

```
## [17] "Asparagine"                "Betaine"
## [19] "Carnitine"                 "Citrate"
## [21] "Creatine"                  "Creatinine"
## [23] "Dimethylamine"             "Ethanolamine"
## [25] "Formate"                   "Fucose"
## [27] "Fumarate"                  "Glucose"
## [29] "Glutamine"                 "Glycine"
## [31] "Glycolate"                 "Guanidoacetate"
## [33] "Hippurate"                 "Histidine"
## [35] "Hypoxanthine"              "Isoleucine"
## [37] "Lactate"                   "Leucine"
## [39] "Lysine"                    "Methylamine"
## [41] "Methylguanidine"           "N,N-Dimethylglycine"
## [43] "O-Acetylcarnitine"         "Pantothenate"
## [45] "Pyroglutamate"             "Pyruvate"
## [47] "Quinolinate"               "Serine"
## [49] "Succinate"                 "Sucrose"
## [51] "Tartrate"                  "Taurine"
## [53] "Threonine"                 "Trigonelline"
## [55] "Trimethylamine N-oxide"    "Tryptophan"
## [57] "Tyrosine"                  "Uracil"
## [59] "Valine"                    "Xylose"
## [61] "cis-Aconitate"             "myo-Inositol"
## [63] "trans-Aconitate"           "pi-Methylhistidine"
## [65] "tau-Methylhistidine"
```

El comando colnames() nos aportará datos sobre el nombre de las columnas. En este trabajo, los nombres de las columnas responden a un número de identificación "Patient ID", a una variable de pérdida múscular en función de si presenta pérdida o no "Muscle loss" y a 63 tipos de metabolitos diferentes.

```
# Resumen estadístico de los datos
summary(data)
```

```
##    Patient ID        Muscle loss        1,6-Anhydro-beta-D-glucose
##  Length:77          Length:77          Min.   :  4.71
##  Class :character   Class :character   1st Qu.: 28.79
##  Mode  :character   Mode  :character   Median : 45.60
##                                        Mean   :105.63
##                                        3rd Qu.:141.17
##                                        Max.   :685.40
##  1-Methylnicotinamide 2-Aminobutyrate  2-Hydroxyisobutyrate 2-Oxoglutarate
##  Min.   :   6.42      Min.   :  1.28   Min.   : 4.85        Min.   :   5.53
##  1st Qu.:  15.80      1st Qu.:  5.26   1st Qu.:15.80        1st Qu.:  22.42
##  Median :  36.60      Median : 10.49   Median :32.46        Median :  55.15
##  Mean   :  71.57      Mean   : 18.16   Mean   :37.25        Mean   : 145.09
##  3rd Qu.:  73.70      3rd Qu.: 19.49   3rd Qu.:54.60        3rd Qu.:  92.76
##  Max.   :1032.77      Max.   :172.43   Max.   :93.69        Max.   :2465.13
##  3-Aminoisobutyrate 3-Hydroxybutyrate 3-Hydroxyisovalerate 3-Indoxylsulfate
##  Min.   :   2.61    Min.   :  1.70    Min.   :  0.92       Min.   :  27.66
##  1st Qu.:  11.70    1st Qu.:  5.99    1st Qu.:  5.26       1st Qu.:  82.27
##  Median :  22.65    Median : 11.70    Median : 12.55       Median : 144.03
##  Mean   :  76.76    Mean   : 21.72    Mean   : 21.65       Mean   : 218.88
##  3rd Qu.:  56.26    3rd Qu.: 29.96    3rd Qu.: 30.27       3rd Qu.: 333.62
```

```
##  Max.   :1480.30    Max.   :175.91    Max.   :164.02       Max.   :1043.15
##  4-Hydroxyphenylacetate    Acetate          Acetone          Adipate
##  Min.   : 15.49         Min.   :  3.49   Min.   :  2.29   Min.   :  1.55
##  1st Qu.: 41.68         1st Qu.: 16.28   1st Qu.:  4.95   1st Qu.:  6.11
##  Median : 70.11         Median : 39.65   Median :  7.10   Median : 10.18
##  Mean   :112.02         Mean   : 66.14   Mean   : 11.43   Mean   : 24.76
##  3rd Qu.:145.47         3rd Qu.: 86.49   3rd Qu.: 10.49   3rd Qu.: 19.11
##  Max.   :796.32         Max.   :411.58   Max.   :206.44   Max.   :327.01
##     Alanine          Asparagine        Betaine         Carnitine
##  Min.   :  16.78   Min.   :  6.69   Min.   :  2.29   Min.   :  2.18
##  1st Qu.:  78.26   1st Qu.: 20.49   1st Qu.: 28.79   1st Qu.: 14.44
##  Median : 194.42   Median : 42.10   Median : 64.72   Median : 23.81
##  Mean   : 273.56   Mean   : 62.28   Mean   : 90.32   Mean   : 52.09
##  3rd Qu.: 399.41   3rd Qu.: 89.12   3rd Qu.:127.74   3rd Qu.: 60.95
##  Max.   :1312.91   Max.   :273.14   Max.   :391.51   Max.   :487.85
##     Citrate           Creatine         Creatinine    Dimethylamine
##  Min.   :   59.74   Min.   :   2.75   Min.   : 1002   Min.   :  41.26
##  1st Qu.:  788.40   1st Qu.:  17.64   1st Qu.: 3498   1st Qu.: 142.59
##  Median : 1790.05   Median :  44.26   Median : 7631   Median : 304.90
##  Mean   : 2235.35   Mean   : 126.83   Mean   : 8734   Mean   : 358.17
##  3rd Qu.: 3071.74   3rd Qu.: 117.92   3rd Qu.:12333   3rd Qu.: 454.86
##  Max.   :13629.61   Max.   :1863.11   Max.   :33860   Max.   :1556.20
##   Ethanolamine        Formate           Fucose          Fumarate
##  Min.   :  16.12   Min.   :   6.42   Min.   :  5.70   Min.   : 0.79
##  1st Qu.:  86.49   1st Qu.:  53.52   1st Qu.: 29.37   1st Qu.: 2.23
##  Median : 204.38   Median :  95.58   Median : 61.56   Median : 4.10
##  Mean   : 276.26   Mean   : 147.40   Mean   : 88.67   Mean   : 8.44
##  3rd Qu.: 407.48   3rd Qu.: 167.34   3rd Qu.:123.97   3rd Qu.: 7.85
##  Max.   :1436.55   Max.   :1480.30   Max.   :407.48   Max.   :96.54
##     Glucose          Glutamine          Glycine          Glycolate
##  Min.   :  26.84   Min.   :  23.34   Min.   :  38.09   Min.   :  5.42
##  1st Qu.:  80.64   1st Qu.: 113.30   1st Qu.: 262.43   1st Qu.: 50.91
##  Median : 210.61   Median : 225.88   Median : 528.48   Median :130.32
##  Mean   : 559.85   Mean   : 306.87   Mean   : 880.72   Mean   :187.99
##  3rd Qu.: 407.48   3rd Qu.: 445.86   3rd Qu.:1096.63   3rd Qu.:267.74
##  Max.   :8690.62   Max.   :1685.81   Max.   :5064.45   Max.   :720.54
##  Guanidoacetate     Hippurate          Histidine        Hypoxanthine
##  Min.   :  7.03   Min.   :   92.76   Min.   :  14.15   Min.   :  3.78
##  1st Qu.: 33.78   1st Qu.:  492.75   1st Qu.:  66.69   1st Qu.: 20.70
##  Median : 64.72   Median : 1224.15   Median : 174.16   Median : 40.04
##  Mean   : 86.37   Mean   : 2286.84   Mean   : 292.64   Mean   : 61.10
##  3rd Qu.:108.85   3rd Qu.: 2921.93   3rd Qu.: 419.89   3rd Qu.: 83.93
##  Max.   :561.16   Max.   :19341.34   Max.   :1863.11   Max.   :265.07
##    Isoleucine        Lactate           Leucine           Lysine
##  Min.   : 1.790   Min.   :   7.32   Min.   :  2.51   Min.   : 10.49
##  1st Qu.: 3.900   1st Qu.:  35.52   1st Qu.:  9.12   1st Qu.: 30.27
##  Median : 7.170   Median :  81.45   Median : 19.11   Median : 69.41
##  Mean   : 8.709   Mean   : 158.46   Mean   : 24.36   Mean   :108.79
##  3rd Qu.:11.250   3rd Qu.: 139.77   3rd Qu.: 31.19   3rd Qu.:121.51
##  Max.   :40.040   Max.   :3640.95   Max.   :103.54   Max.   :788.40
##   Methylamine    Methylguanidine  N,N-Dimethylglycine O-Acetylcarnitine
##  Min.   : 1.51   Min.   :  1.70   Min.   :  0.79      Min.   :  1.23
##  1st Qu.: 5.26   1st Qu.:  4.26   1st Qu.:  7.03      1st Qu.:  3.94
##  Median :14.73   Median :  7.85   Median : 21.98      Median : 11.47
```

```
##  Mean   :17.38   Mean   : 15.32   Mean   : 26.35      Mean   : 19.73
##  3rd Qu.:24.05   3rd Qu.: 19.30   3rd Qu.: 40.04      3rd Qu.: 20.91
##  Max.   :52.46   Max.   :141.17   Max.   :120.30      Max.   :254.68
##   Pantothenate    Pyroglutamate        Pyruvate       Quinolinate
##  Min.   :  2.59   Min.   :  21.33   Min.   :  0.90   Min.   :  5.21
##  1st Qu.: 11.13   1st Qu.:  68.72   1st Qu.:  4.85   1st Qu.: 26.58
##  Median : 22.65   Median : 157.59   Median : 13.46   Median : 51.42
##  Mean   : 44.88   Mean   : 211.45   Mean   : 21.29   Mean   : 66.44
##  3rd Qu.: 41.26   3rd Qu.: 301.87   3rd Qu.: 29.08   3rd Qu.: 87.36
##  Max.   :692.29   Max.   :1064.22   Max.   :184.93   Max.   :259.82
##      Serine          Succinate         Sucrose          Tartrate
##  Min.   :  16.12   Min.   :  1.72   Min.   :   6.49   Min.   :  2.20
##  1st Qu.:  83.10   1st Qu.:  8.58   1st Qu.:  19.30   1st Qu.:  6.89
##  Median : 142.59   Median : 30.88   Median :  40.85   Median : 12.94
##  Mean   : 197.69   Mean   : 60.23   Mean   : 113.23   Mean   : 40.00
##  3rd Qu.: 270.43   3rd Qu.: 74.44   3rd Qu.:  94.63   3rd Qu.: 25.79
##  Max.   :1248.88   Max.   :589.93   Max.   :2079.74   Max.   :837.15
##     Taurine          Threonine       Trigonelline     Trimethylamine N-oxide
##  Min.   :  17.81   Min.   :  8.25   Min.   :  10.07   Min.   :  55.7
##  1st Qu.:  99.48   1st Qu.: 31.82   1st Qu.:  53.52   1st Qu.: 175.9
##  Median : 249.64   Median : 64.07   Median : 114.43   Median : 383.8
##  Mean   : 525.12   Mean   : 95.36   Mean   : 270.44   Mean   : 652.2
##  3rd Qu.: 665.14   3rd Qu.:137.00   3rd Qu.: 340.36   3rd Qu.: 735.1
##  Max.   :4272.69   Max.   :450.34   Max.   :2252.96   Max.   :5486.2
##    Tryptophan        Tyrosine          Uracil           Valine
##  Min.   :  8.67   Min.   :  4.22   Min.   :  3.10   Min.   :  4.10
##  1st Qu.: 21.33   1st Qu.: 23.57   1st Qu.: 11.94   1st Qu.: 12.18
##  Median : 46.99   Median : 60.34   Median : 27.39   Median : 33.12
##  Mean   : 66.24   Mean   : 81.76   Mean   : 35.56   Mean   : 35.67
##  3rd Qu.: 96.54   3rd Qu.:113.30   3rd Qu.: 44.26   3rd Qu.: 50.40
##  Max.   :259.82   Max.   :539.15   Max.   :179.47   Max.   :160.77
##      Xylose        cis-Aconitate      myo-Inositol    trans-Aconitate
##  Min.   :  10.07   Min.   :  12.94   Min.   : 11.59   Min.   :  4.90
##  1st Qu.:  29.96   1st Qu.:  36.23   1st Qu.: 30.27   1st Qu.: 12.43
##  Median :  50.40   Median : 129.02   Median : 78.26   Median : 26.84
##  Mean   : 100.93   Mean   : 204.22   Mean   :135.40   Mean   : 40.63
##  3rd Qu.:  89.12   3rd Qu.: 254.68   3rd Qu.:167.34   3rd Qu.: 57.40
##  Max.   :2164.62   Max.   :1863.11   Max.   :854.06   Max.   :217.02
##  pi-Methylhistidine tau-Methylhistidine
##  Min.   :  11.36    Min.   :  8.00
##  1st Qu.:  67.36    1st Qu.: 27.39
##  Median : 162.39    Median : 68.72
##  Mean   : 370.29    Mean   : 89.69
##  3rd Qu.: 387.61    3rd Qu.:130.32
##  Max.   :2697.28    Max.   :317.35
```

Por último, el comando summary(), nos arroja información estadística sobre los datos, facilitando la compresión de cómo varían los datos a lo largo del estudio.

## Preparación de Datos

Separamos los datos en sus componentes principales: IDs de paciente, información sobre pérdida muscular y matriz de datos metabolómicos. De esta forma podremos crear una matriz de datos y dos variables ante

las cuales responden los datos.

```
# Separar los IDs de paciente y los datos
patient_ids <- data$`Patient ID`
muscle_loss <- data$`Muscle loss`
data_matrix <- as.matrix(data[,-(1:2)])  # No incluye las dos primeras columnas
```

## Creación de Metadatos

En este apartado nos centramos en la creación de los metadatos para las columnas (metabolitos) y las filas (pacientes).

```
# Crear los metadatos de las columnas (metabolitos)
colData <- data.frame(
  metabolite = colnames(data_matrix),
  row.names = colnames(data_matrix)
)

# Crear los metadatos de las filas (pacientes)
rowData <- data.frame(
  patient_id = patient_ids,
  muscle_loss = muscle_loss,
  condition = ifelse(grepl("cachexic", patient_ids), "cachexic", "control"),
  row.names = patient_ids
)
```

## Creación del contenedor SummarizedExperiment

Construimos el contenedor SummarizedExperiment con los datos y metadatos preparados en los apartados anteriores.

```
se <- SummarizedExperiment(
  assays = list(abundance = data_matrix),
  colData = colData,
  rowData = rowData
)
```

## Adición de Metadatos del Experimento

Añadimos información adicional sobre el experimento como metadatos. En esta sección añadiremos datos relevantes sobre el estudio.

```
metadata(se) <- list(
    # Información general
    description = "Metabolomic data for cachexic and control patients",
    date = Sys.Date(),
    researcher = "MetaboAnalyst tutorials",
    experiment_type = "Human Cachexia Metabolomics",

    # Información técnica
```

```
    sample_type = "Blood plasma",
    technology = "NMR spectroscopy",

    # Estadísticas básicas
    total_metabolites = ncol(data_matrix),
    total_samples = nrow(data_matrix),
    missing_values = sum(is.na(data_matrix))
)
```

## Resumen Final

Por último, mostraremos el contenedor ya creado, así como sus metadatos. Además, crearemos el archivo contenedor .Rda solicitado en el ejercicio.

```
# Resumen del contenedor SummarizedExperiment
print(se)
```

```
## class: SummarizedExperiment
## dim: 77 63
## metadata(9): description date ... total_samples missing_values
## assays(1): abundance
## rownames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## rowData names(3): patient_id muscle_loss condition
## colnames(63): 1,6-Anhydro-beta-D-glucose 1-Methylnicotinamide ...
##   pi-Methylhistidine tau-Methylhistidine
## colData names(1): metabolite
```

```
metadata(se)  # Metadatos del experimento
```

```
## $description
## [1] "Metabolomic data for cachexic and control patients"
##
## $date
## [1] "2024-11-04"
##
## $researcher
## [1] "MetaboAnalyst tutorials"
##
## $experiment_type
## [1] "Human Cachexia Metabolomics"
##
## $sample_type
## [1] "Blood plasma"
##
## $technology
## [1] "NMR spectroscopy"
##
## $total_metabolites
## [1] 63
##
## $total_samples
```

```
## [1] 77
##
## $missing_values
## [1] 0
```

```
# Creación del archivo se
save(se, file = "summarized_experiment.Rda")
```

## Conclusiones

De esta forma hemos podido crear un contenedor del tipo SummaridezExperiment y observar un resumen de su estructura y sus metadatos. Si bien el dataset utilizado no contiene demasiados pacientes, si que podemos hacernos una idea de cómo de presente está la enfermedad en la población del estudio y qué aminoácidos se presentan con mayor proporción en el estudio (gracias a las medidas estadísticas calculadas anteriormente),

Para finalizar, realizamos la creación del repositorio en Github y así incluir todas las partes y archivos demandados en el trabajo.

## Repositorio Github

El código, los datos y el informe se pueden encontrar en el siguiente repositorio de GitHub:

Repositorio en GitHub