

Project 5 - Descriptive data analysis

Aleksandra Buchowicz, Michał Klimek, Marcin Ostrowski

2024-01-11

Loading the data

```
data <- read.csv('data_project_5_for_students.csv')
head(data)
```

```
##   Group Gender rs2476601 rs3087243 rs3807306 rs1800629 rs1799724
## 1   CFS      f         AG         GG         GG         AA         CC
## 2   HC      f         GG         AA         GT         GG         CC
## 3   HC      f         GG         GG         GG         GG         CC
## 4   HC      m         AG        <NA>         GT         GG         CC
## 5   CFS      f         AG         AG         TT         GG         CT
## 6   CFS      m         GG         AG         GT         GG         CC
```

Basic statistics

```
summary(data)
```

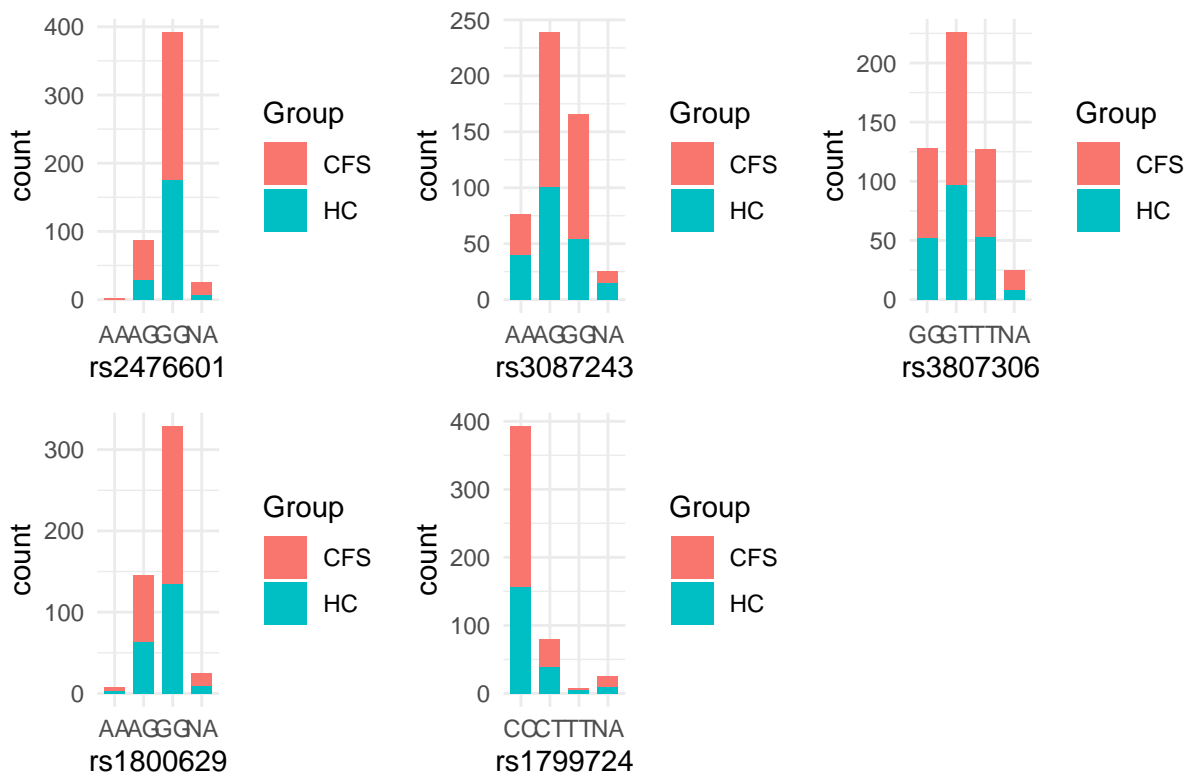
```
##      Group      Gender      rs2476601      rs3087243
## Length:506      Length:506      Length:506      Length:506
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##   rs3807306      rs1800629      rs1799724
## Length:506      Length:506      Length:506
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
```

```
##
##   CFS   HC <NA>
## 296  210    0
##
##    f    m <NA>
## 295  211    0
##
##   AA   AG   GG <NA>
##    2   87  392   25
##
```

```
##      AA      AG      GG <NA>
##      76     239    166     25
##
##      GG      GT      TT <NA>
##     128     226    127     25
##
##      AA      AG      GG <NA>
##       7     145    329     25
##
##      CC      CT      TT <NA>
##     393      80      8     25
```

Distributions of data among CFS and control groups

Distribution of genes among both groups



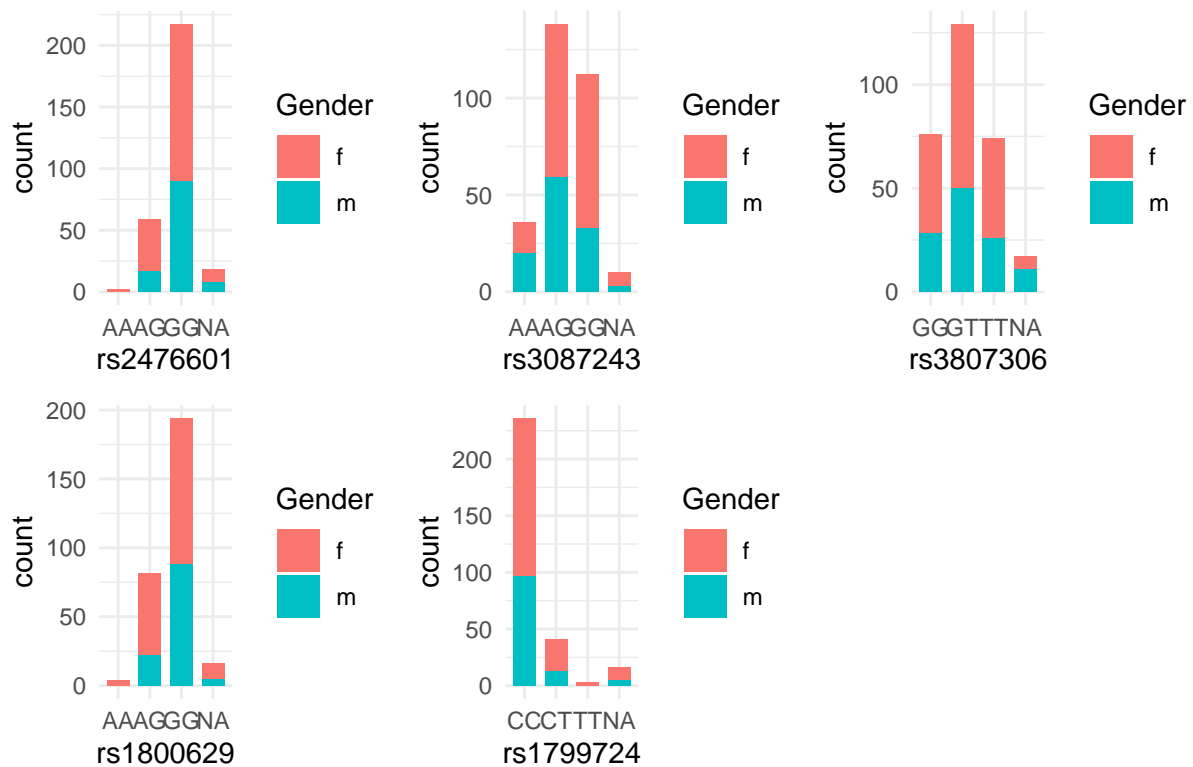
CFS group

```
## [1] "### Chronic Fatigue Syndrome ###"
```

```
##
##      f      m <NA>
##     181    115     0
##
##      AA      AG      GG <NA>
```

```
##      2    59   217    18
##
##      AA    AG    GG <NA>
##      36   138   112    10
##
##      GG    GT    TT <NA>
##      76   129    74    17
##
##      AA    AG    GG <NA>
##      4    82   194    16
##
##      CC    CT    TT <NA>
##     236    41     3    16
```

CFS Group



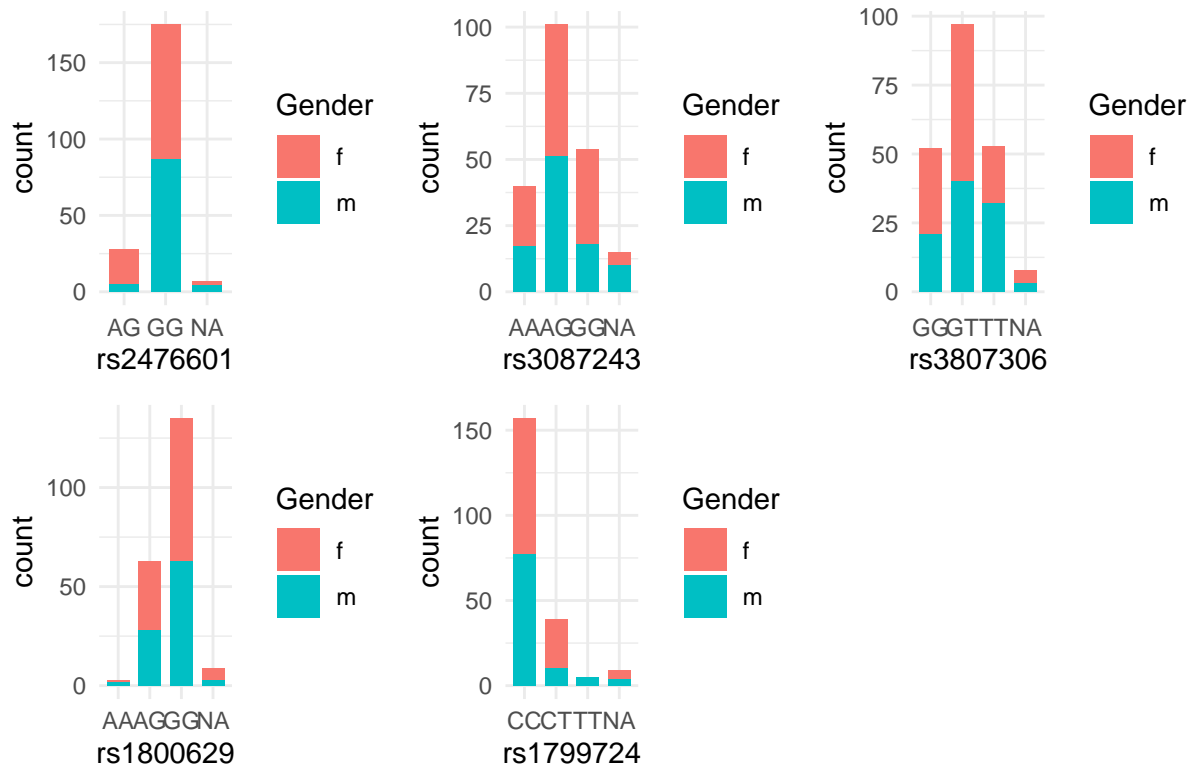
Healthy control group

```
## [1] "### Healthy Control ###"
```

```
##
##      f    m <NA>
##     114   96    0
##
##      AG    GG <NA>
##      28   175    7
##
```

```
## AA AG GG <NA>
## 40 101 54 15
##
## GG GT TT <NA>
## 52 97 53 8
##
## AA AG GG <NA>
## 3 63 135 9
##
## CC CT TT <NA>
## 157 39 5 9
```

Healthy Control



Distributions of data between genders

```
## [1] "### Females ###"
```

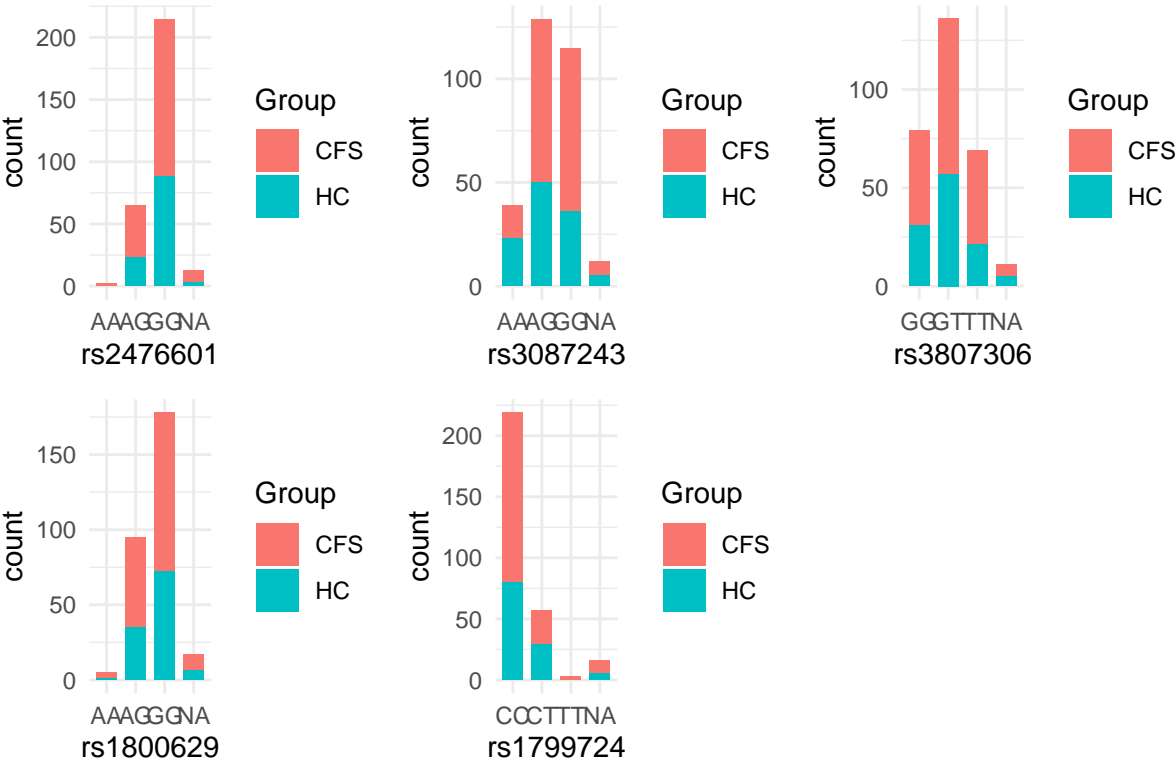
```
##
## CFS HC <NA>
## 181 114 0
##
## AA AG GG <NA>
## 2 65 215 13
##
## AA AG GG <NA>
```

```
##      39  129  115   12
##
##      GG    GT    TT <NA>
##      79  136   69   11
##
##      AA    AG    GG <NA>
##       5   95  178   17
##
##      CC    CT    TT <NA>
##     219   57    3   16
```

```
## [1] "### Males ###"
```

```
##
##      CFS    HC <NA>
##     115    96    0
##
##      AG    GG <NA>
##     22  177   12
##
##      AA    AG    GG <NA>
##     37  110   51   13
##
##      GG    GT    TT <NA>
##     49   90   58   14
##
##      AA    AG    GG <NA>
##       2   50  151    8
##
##      CC    CT    TT <NA>
##    174   23    5    9
```

Distribution of genes in females



Distribution of genes in males

