submission1

Antara Sengupta

2024-07-27

Loading in necessary packages

```
library(readr)
library(dplyr)
```

```
##
## 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
```

```
library(ggplot2)
library(tibble)
library(tidyr)
library(dplyr)
```

Loading Data

```
# loading in the two separate datasets provided to us (metadata & gene exppression data)
series <- read.csv("data/QBS103_GSE157103_series_matrix.csv")
genes <- read.csv("data/QBS103_GSE157103_genes.csv")

# displaying first few rows of each dataframe to get familiarized with the data
head(series)
```

```
##             participant_id geo_accession                status
## 1 COVID_01_39y_male_NonICU    GSM4753021 Public on Aug 29 2020
## 2 COVID_02_63y_male_NonICU    GSM4753022 Public on Aug 29 2020
## 3 COVID_03_33y_male_NonICU    GSM4753023 Public on Aug 29 2020
## 4 COVID_04_49y_male_NonICU    GSM4753024 Public on Aug 29 2020
## 5 COVID_05_49y_male_NonICU    GSM4753025 Public on Aug 29 2020
## 6  COVID_06_:y_male_NonICU    GSM4753026 Public on Aug 29 2020
##   X.Sample_submission_date last_update_date type channel_count
## 1              Aug 28 2020      Aug 29 2020  SRA             1
## 2              Aug 28 2020      Aug 29 2020  SRA             1
## 3              Aug 28 2020      Aug 29 2020  SRA             1
## 4              Aug 28 2020      Aug 29 2020  SRA             1
## 5              Aug 28 2020      Aug 29 2020  SRA             1
## 6              Aug 28 2020      Aug 29 2020  SRA             1
##                source_name_ch1 organism_ch1          disease_status age   sex
## 1 Leukocytes from whole blood Homo sapiens disease state: COVID-19  39  male
## 2 Leukocytes from whole blood Homo sapiens disease state: COVID-19  63  male
## 3 Leukocytes from whole blood Homo sapiens disease state: COVID-19  33  male
## 4 Leukocytes from whole blood Homo sapiens disease state: COVID-19  49  male
## 5 Leukocytes from whole blood Homo sapiens disease state: COVID-19  49  male
## 6 Leukocytes from whole blood Homo sapiens disease state: COVID-19   :  male
##   icu_status apacheii charlson_score mechanical_ventilation
## 1         no       15              0                    yes
## 2         no  unknown              2                     no
## 3         no  unknown              2                     no
## 4         no  unknown              1                     no
## 5         no       19              1                    yes
## 6         no  unknown              1                     no
##   ventilator.free_days hospital.free_days_post_45_day_followup ferritin.ng.ml.
## 1                    0                                       0             946
## 2                   28                                      39            1060
## 3                   28                                      18            1335
## 4                   28                                      39             583
## 5                   23                                      27             800
## 6                   28                                      36             563
##   crp.mg.l. ddimer.mg.l_feu. procalcitonin.ng.ml.. lactate.mmol.l. fibrinogen
## 1      73.1              1.3                    36             0.9        513
## 2   unknown             1.03                  0.37         unknown    unknown
## 3      53.2             1.48                  0.07         unknown        513
## 4     251.1             1.32                  0.98            0.87        949
## 5     355.8             0.69                  4.92            1.48        929
## 6     129.1          unknown                  0.67            0.86        769
##      sofa
## 1       8
## 2 unknown
## 3 unknown
## 4 unknown
## 5       7
## 6 unknown
```

```
head(genes)
```

```
##         X COVID_01_39y_male_NonICU COVID_02_63y_male_NonICU
## 1    A1BG                     0.49                     0.29
## 2    A1CF                     0.00                     0.00
## 3     A2M                     0.21                     0.14
## 4   A2ML1                     0.04                     0.00
## 5 A3GALT2                     0.07                     0.00
## 6  A4GALT                     0.00                     0.00
##   COVID_03_33y_male_NonICU COVID_04_49y_male_NonICU COVID_05_49y_male_NonICU
## 1                     0.26                     0.45                     0.17
## 2                     0.00                     0.01                     0.00
## 3                     0.03                     0.09                     0.00
## 4                     0.02                     0.07                     0.05
## 5                     0.00                     0.00                     0.07
## 6                     0.00                     0.00                     0.00
##   COVID_06_.y_male_NonICU COVID_07_38y_female_NonICU COVID_08_78y_male_ICU
## 1                    0.21                       0.49                  0.12
## 2                    0.00                       0.01                  0.00
## 3                    0.08                       0.23                  0.08
## 4                    0.04                       0.03                  0.01
## 5                    0.00                       0.07                  0.00
## 6                    0.00                       0.00                  0.00
##   COVID_09_64y_female_ICU COVID_10_62y_male_ICU COVID_11_52y_female_NonICU
## 1                    0.51                  0.10                       0.38
## 2                    0.01                  0.00                       0.02
## 3                    0.88                  0.13                       0.47
## 4                    0.02                  0.01                       0.03
## 5                    0.79                  0.15                       0.08
## 6                    0.00                  0.00                       0.00
##   COVID_12_50y_male_ICU COVID_13_37y_male_NonICU COVID_14_55y_male_ICU
## 1                  0.45                     0.18                  0.23
## 2                  0.00                     0.00                  0.00
## 3                  0.16                     0.07                  0.22
## 4                  0.00                     0.01                  0.04
## 5                  1.75                     0.00                  0.93
## 6                  0.00                     0.00                  0.00
##   COVID_15_68y_male_ICU COVID_16_48y_male_NonICU COVID_17_54y_male_NonICU
## 1                  0.42                     0.41                     0.63
## 2                  0.00                     0.01                     0.02
## 3                  0.07                     0.58                     0.15
## 4                  0.00                     0.00                     0.02
## 5                  0.15                     0.19                     0.00
## 6                  0.03                     0.00                     0.00
##   COVID_18_70y_female_NonICU COVID_19_51y_male_NonICU COVID_20_62y_male_ICU
## 1                       0.47                     0.33                  0.32
## 2                       0.00                     0.02                  0.00
## 3                       0.30                     0.11                  0.07
## 4                       0.02                     0.02                  0.00
## 5                       0.06                     0.00                  0.22
## 6                       0.03                     0.00                  0.00
##   COVID_21_66y_male_ICU COVID_22_43y_male_ICU COVID_23_76y_male_ICU
## 1                  0.18                  0.09                  0.18
## 2                  0.00                  0.00                  0.01
## 3                  0.00                  0.06                  0.03
## 4                  0.00                  0.00                  0.00
## 5                  0.37                  0.06                  0.07
## 6                  0.03                  0.00                  0.03
##   COVID_24_55y_male_ICU COVID_25_55y_male_ICU COVID_26_41y_female_ICU
## 1                  0.22                  0.29                    0.42
## 2                  0.01                  0.00                    0.00
## 3                  0.11                  0.09                    0.18
## 4                  0.02                  0.03                    0.00
## 5                  0.15                  0.00                    0.87
## 6                  0.00                  0.00                    0.00
##   COVID_27_71y_female_ICU COVID_28_63y_male_ICU COVID_29_63y_female_ICU
## 1                    0.16                  0.18                    0.35
## 2                    0.01                  0.00                    0.00
## 3                    0.23                  0.18                    0.03
## 4                    0.01                  0.05                    0.03
## 5                    0.18                  0.45                    0.15
## 6                    0.00                  0.00                    0.03
##   COVID_30_54y_male_ICU COVID_31_50y_male_ICU COVID_32_72y_male_ICU
## 1                  0.23                  0.15                  0.34
## 2                  0.00                  0.00                  0.01
## 3                  0.11                  0.47                  0.04
## 4                  0.01                  0.00                  0.00
## 5                  0.00                  0.00                  0.29
## 6                  0.00                  0.03                  0.00
##   COVID_33_81y_male_NonICU COVID_34_64y_female_NonICU
## 1                     0.35                       0.36
## 2                     0.00                       0.00
## 3                     0.30                       0.11
## 4                     0.06                       0.00
## 5                     0.26                       0.12
## 6                     0.00                       0.00
##   COVID_35_58y_female_NonICU COVID_36_68y_male_NonICU COVID_37_87y_male_NonICU
## 1                       0.26                     0.18                     0.20
## 2                       0.00                     0.01                     0.00
## 3                       0.51                     0.09                     0.09
## 4                       0.02                     0.00                     0.07
## 5                       0.16                     0.08                     0.31
## 6                       0.00                     0.00                     0.00
##   COVID_38_68y_male_ICU COVID_39_80y_female_ICU COVID_40_66y_male_ICU
## 1                  0.29                    0.19                  0.22
## 2                  0.00                    0.00                  0.00
## 3                  0.10                    0.27                  0.17
## 4                  0.02                    0.00                  0.00
## 5                  0.35                    0.00                  0.08
## 6                  0.00                    0.07                  0.00
##   COVID_41_74y_male_ICU COVID_42_21y_female_ICU COVID_43_83y_female_ICU
## 1                  0.19                    0.24                    0.29
## 2                  0.00                    0.01                    0.00
## 3                  0.14                    0.33                    0.00
## 4                  0.00                    0.01                    0.00
## 5                  0.19                    0.39                    0.11
## 6                  0.00                    0.00                    0.00
##   COVID_44_46y_male_ICU COVID_45_62y_female_ICU COVID_46_62y_male_ICU
## 1                  0.22                    0.14                  0.53
## 2                  0.00                    0.00                  0.01
## 3                  0.14                    0.15                  0.10
## 4                  0.00                    0.03                  0.00
## 5                  0.00                    0.19                  0.06
## 6                  0.04                    0.00                  0.00
##   COVID_47_78y_male_ICU COVID_48_72y_female_ICU COVID_49_73y_male_ICU
## 1                  0.08                    0.19                  0.48
## 2                  0.01                    0.00                  0.00
## 3                  0.04                    0.06                  0.09
## 4                  0.03                    0.01                  0.03
## 5                  0.60                    0.23                  0.00
## 6                  0.00                    0.06                  0.00
##   COVID_50_37y_male_ICU COVID_51_58y_female_NonICU COVID_52_71y_male_NonICU
## 1                  0.08                       0.21                     0.25
## 2                  0.00                       0.00                     0.01
## 3                  0.01                       0.13                     0.00
## 4                  0.00                       0.00                     0.03
## 5                  0.00                       0.00                     0.00
## 6                  0.72                       0.00                     0.00
##   COVID_53_35y_female_NonICU COVID_55_62y_female_ICU COVID_56_33y_female_NonICU
## 1                       0.25                    0.09                       0.28
## 2                       0.00                    0.00                       0.00
## 3                       0.64                    0.09                       0.16
## 4                       0.10                    0.01                       0.09
## 5                       0.00                    0.00                       0.23
## 6                       0.00                    0.00                       0.00
##   COVID_57_30y_female_NonICU COVID_58_62y_male_NonICU COVID_59_55y_male_NonICU
## 1                       0.42                     0.39                     0.33
## 2                       0.00                     0.00                     0.00
## 3                       0.27                     0.08                     0.10
## 4                       0.01                     0.00                     0.00
## 5                       0.19                     0.00                     0.07
## 6                       0.05                     0.00                     0.00
##   COVID_60_49y_male_NonICU COVID_61_54y_female_NonICU COVID_62_78y_female_ICU
## 1                     0.22                       0.25                    0.21
## 2                     0.00                       0.00                    0.00
## 3                     0.14                       0.10                    0.04
## 4                     0.00                       0.03                    0.00
## 5                     0.00                       0.13                    0.05
## 6                     0.02                       0.00                    0.00
##   COVID_63_39y_female_ICU COVID_64_65y_male_ICU COVID_65_84y_male_NonICU
## 1                    0.29                  0.38                     0.40
## 2                    0.00                  0.01                     0.01
## 3                    0.01                  0.04                     0.07
## 4                    0.00                  0.02                     0.00
## 5                    0.14                  0.56                     0.58
## 6                    0.00                  0.00                     0.00
##   COVID_66_66y_female_NonICU COVID_67_57y_male_ICU COVID_68_79y_male_ICU
## 1                       0.64                  0.37                  0.58
## 2                       0.00                  0.00                  0.00
## 3                       0.00                  0.35                  0.15
## 4                       0.00                  0.00                  0.01
## 5                       0.00                  0.00                  0.00
## 6                       0.00                  0.00                  0.05
##   COVID_69_77y_female_NonICU COVID_70_81y_male_NonICU COVID_71_37y_male_ICU
## 1                       0.52                     0.27                  0.07
## 2                       0.00                     0.00                  0.01
## 3                       0.29                     0.07                  0.12
## 4                       0.02                     0.00                  0.01
## 5                       0.00                     0.00                  0.00
## 6                       0.00                     0.06                  0.00
##   COVID_72_50y_female_NonICU COVID_73_82y_male_NonICU COVID_74_55y_female_ICU
## 1                       0.52                     0.46                    0.24
## 2                       0.00                     0.01                    0.00
## 3                       0.10                     0.02                    0.12
## 4                       0.01                     0.02                    0.02
## 5                       0.00                     0.17                    0.26
## 6                       0.00                     0.04                    0.00
##   COVID_75_55y_male_NonICU COVID_76_73y_female_ICU COVID_77_55y_female_ICU
## 1                     0.23                    0.17                    0.05
## 2                     0.01                    0.00                    0.00
## 3                     0.14                    0.09                    0.01
## 4                     0.00                    0.01                    0.00
## 5                     0.00                    0.04                    0.00
## 6                     0.00                    0.00                    0.00
##   COVID_78_80y_male_NonICU COVID_79_27y_male_NonICU COVID_80_71y_male_ICU
## 1                     0.19                     0.08                  0.28
## 2                     0.00                     0.01                  0.00
## 3                     0.20                     0.03                  0.05
## 4                     0.00                     0.00                  0.00
## 5                     0.00                     0.00                  0.05
## 6                     0.00                     0.00                  0.00
##   COVID_82_67y_male_NonICU COVID_83_85y_female_NonICU
## 1                     0.39                       0.47
## 2                     0.01                       0.00
## 3                     0.10                       0.18
## 4                     0.00                       0.05
## 5                     0.00                       0.00
## 6                     0.00                       0.00
##   COVID_84_75y_female_NonICU COVID_85_62y_male_ICU COVID_86_52y_female_NonICU
## 1                       0.35                  0.29                       0.60
## 2                       0.00                  0.00                       0.00
## 3                       0.03                  0.04                       0.27
## 4                       0.00                  0.00                       0.02
## 5                       0.17                  0.00                       0.00
## 6                       0.00                  0.00                       0.00
##   COVID_87_61y_male_ICU COVID_89_90y_female_NonICU COVID_90_86y_female_NonICU
## 1                  0.65                       0.20                       0.40
## 2                  0.00                       0.00                       0.00
## 3                  0.15                       0.07                       0.05
## 4                  0.00                       0.03                       0.01
## 5                  0.00                       0.14                       0.31
## 6                  0.00                       0.00                       0.02
##   COVID_91_29y_female_NonICU COVID_92_82y_female_ICU COVID_93_81y_female_ICU
## 1                       0.60                    0.34                    0.37
## 2                       0.00                    0.00                    0.00
## 3                       0.03                    0.02                    0.11
## 4                       0.02                    0.04                    0.00
## 5                       0.05                    0.58                    0.05
## 6                       0.00                    0.00                    0.00
##   COVID_94_24y_female_NonICU COVID_95_49y_male_NonICU COVID_96_51y_male_NonICU
## 1                       0.81                     0.37                     1.61
## 2                       0.00                     0.01                     0.00
## 3                       0.17                     0.20                     0.02
## 4                       0.02                     0.02                     0.00
## 5                       0.00                     0.15                     0.00
## 6                       0.06                     0.00                     0.00
##   COVID_97_76y_male_ICU COVID_98_81y_male_NonICU COVID_99_71y_male_ICU
## 1                  0.19                     0.78                  0.33
## 2                  0.00                     0.00                  0.00
## 3                  0.02                     0.26                  0.02
## 4                  0.05                     0.00                  0.00
## 5                  0.12                     0.37                  0.04
## 6                  0.03                     0.00                  0.00
##   COVID_100_74y_female_NonICU COVID_101_58y_male_ICU COVID_102_84y_male_NonICU
## 1                        0.30                   0.33                      0.12
## 2                        0.00                   0.00                      0.00
## 3                        0.09                   0.11                      0.01
## 4                        0.00                   0.03                      0.01
## 5                        0.04                   0.05                      0.00
## 6                        0.00                   0.00                      0.07
##   COVID_103_83y_male_NonICU NONCOVID_01_54y_female_NonICU
## 1                      0.20                          0.89
## 2                      0.00                          0.00
## 3                      0.03                          0.04
## 4                      0.03                          0.00
## 5                      0.04                          0.00
## 6                      0.00                          0.00
##   NONCOVID_02_65y_male_ICU NONCOVID_03_65y_male_ICU NONCOVID_04_90y_male_NonICU
## 1                     0.32                     0.44                        0.21
## 2                     0.00                     0.00                        0.00
## 3                     0.01                     0.05                        0.05
## 4                     0.00                     0.02                        0.00
## 5                     0.04                     0.04                        0.21
## 6                     0.00                     0.00                        0.00
##   NONCOVID_05_83y_female_NonICU NONCOVID_06_75y_female_ICU
## 1                          0.31                       0.89
## 2                          0.00                       0.00
## 3                          0.01                       0.14
## 4                          0.01                       0.01
## 5                          0.00                       0.00
## 6                          0.00                       0.06
##   NONCOVID_07_50y_male_ICU NONCOVID_08_53y_female_ICU
## 1                     0.45                       0.47
## 2                     0.00                       0.01
## 3                     0.07                       0.04
## 4                     0.02                       0.00
## 5                     0.00                       0.15
## 6                     0.00                       0.00
##   NONCOVID_09_49y_female_NonICU NONCOVID_10_67y_male_ICU
## 1                          0.40                     0.33
## 2                          0.00                     0.00
## 3                          0.04                     0.05
## 4                          0.00                     0.01
## 5                          0.00                     0.23
## 6                          0.00                     0.08
##   NONCOVID_11_58y_female_NonICU NONCOVID_12_82y_male_ICU
## 1                          0.58                     0.12
## 2                          0.00                     0.00
## 3                          0.03                     0.02
## 4                          0.00                     0.00
## 5                          0.00                     0.00
## 6                          0.00                     0.02
##   NONCOVID_13_65y_male_ICU NONCOVID_14_75y_female_ICU
## 1                     0.31                       0.16
## 2                     0.00                       0.00
## 3                     0.04                       0.08
## 4                     0.01                       0.00
## 5                     0.32                       0.05
## 6                     0.02                       0.02
##   NONCOVID_15_83y_unknown_ICU NONCOVID_16_40y_female_ICU
## 1                        0.59                       0.34
## 2                        0.00                       0.00
## 3                        0.03                       0.07
## 4                        0.04                       0.00
## 5                        0.00                       0.13
## 6                        0.19                       0.00
##   NONCOVID_17_84y_female_ICU NONCOVID_18_88y_male_ICU
## 1                       0.37                     0.33
## 2                       0.00                     0.00
## 3                       0.07                     0.06
## 4                       0.01                     0.00
## 5                       0.18                     0.00
## 6                       0.00                     0.00
##   NONCOVID_19_66y_female_ICU NONCOVID_20_62y_female_ICU
## 1                       0.25                       0.20
## 2                       0.00                       0.00
## 3                       0.11                       0.01
## 4                       0.00                       0.02
## 5                       0.04                       0.00
## 6                       0.03                       0.07
##   NONCOVID_21_71y_male_NonICU NONCOVID_22_63y_male_NonICU
## 1                        0.40                        0.30
## 2                        0.00                        0.00
## 3                        0.04                        0.02
## 4                        0.02                        0.02
## 5                        0.00                        0.00
## 6                        0.00                        0.00
##   NONCOVID_23_42y_female_NonICU NONCOVID_24_32y_female_NonICU
## 1                          0.70                          0.75
## 2                          0.00                          0.00
## 3                          0.02                          0.27
## 4                          0.01                          0.00
## 5                          0.00                          0.06
## 6                          0.00                          0.00
##   NONCOVID_25_62y_male_NonICU NONCOVID_26_36y_male_ICU
## 1                        2.80                     0.22
## 2                        0.00                     0.00
## 3                        0.04                     0.28
## 4                        0.00                     0.00
## 5                        0.00                     0.00
## 6                        0.00                     0.00
```

Formatting and cleaning data

```
# goal is to merge both dataframes - we can merge them based on participant ID

# transposing the genes df and turning it back to a df so it can have 'participant_id' as rows (the sam
genes_transposed <- as.data.frame(t(genes))

# assigning row x with gene name to be the column headers
colnames(genes_transposed) <- as.character(genes_transposed[1, ])

# getting rid of the x row because it has redundant info (same values as column names)
```

```r
genes_transposed <- genes_transposed[-1, ]

# setting the row name to participant id so we can use this column to merge with metadata
# used this source to find the rownames_to_column function https://forum.posit.co/t/rstudio-rownames-to
genes_transposed <- rownames_to_column(genes_transposed, var = "participant_id")

# reshaping the gene so that genes and expression can be their own columns using pivot_longer
# used this source to learn how to implement pivot_longer: https://tidyr.tidyverse.org/reference/pivot_
genes_long <- genes_transposed %>%
   pivot_longer(
     cols = -participant_id,  #accessing all columns with the exception of participant_id because it is
     names_to = "gene", #setting name column to gene
     values_to = "expression"  #setting values column to expression
   )

# merging long gene data and series metadata to become one dataframe
all_data <- merge(genes_long, series, by = "participant_id") #merging it on 'participant_id' column

# taking a look at the new comprehensive dataframe
head(all_data)
```

```
##             participant_id    gene expression geo_accession
## 1 COVID_01_39y_male_NonICU    A1CF       0.00    GSM4753021
## 2 COVID_01_39y_male_NonICU    A1BG       0.49    GSM4753021
## 3 COVID_01_39y_male_NonICU   AADAC       0.00    GSM4753021
## 4 COVID_01_39y_male_NonICU AADACL2       0.00    GSM4753021
## 5 COVID_01_39y_male_NonICU AADACL3       0.00    GSM4753021
## 6 COVID_01_39y_male_NonICU AADACL4       0.00    GSM4753021
##                  status X.Sample_submission_date last_update_date type
## 1 Public on Aug 29 2020              Aug 28 2020      Aug 29 2020  SRA
## 2 Public on Aug 29 2020              Aug 28 2020      Aug 29 2020  SRA
## 3 Public on Aug 29 2020              Aug 28 2020      Aug 29 2020  SRA
## 4 Public on Aug 29 2020              Aug 28 2020      Aug 29 2020  SRA
## 5 Public on Aug 29 2020              Aug 28 2020      Aug 29 2020  SRA
## 6 Public on Aug 29 2020              Aug 28 2020      Aug 29 2020  SRA
##   channel_count             source_name_ch1 organism_ch1
## 1             1 Leukocytes from whole blood Homo sapiens
## 2             1 Leukocytes from whole blood Homo sapiens
## 3             1 Leukocytes from whole blood Homo sapiens
## 4             1 Leukocytes from whole blood Homo sapiens
## 5             1 Leukocytes from whole blood Homo sapiens
## 6             1 Leukocytes from whole blood Homo sapiens
##            disease_status age   sex icu_status apacheii charlson_score
## 1 disease state: COVID-19  39  male         no       15              0
## 2 disease state: COVID-19  39  male         no       15              0
## 3 disease state: COVID-19  39  male         no       15              0
## 4 disease state: COVID-19  39  male         no       15              0
## 5 disease state: COVID-19  39  male         no       15              0
## 6 disease state: COVID-19  39  male         no       15              0
##   mechanical_ventilation ventilator.free_days
## 1                    yes                    0
## 2                    yes                    0
## 3                    yes                    0
## 4                    yes                    0
## 5                    yes                    0
## 6                    yes                    0
##   hospital.free_days_post_45_day_followup ferritin.ng.ml. crp.mg.l.
## 1                                       0             946      73.1
## 2                                       0             946      73.1
## 3                                       0             946      73.1
## 4                                       0             946      73.1
## 5                                       0             946      73.1
## 6                                       0             946      73.1
##   ddimer.mg.l_feu. procalcitonin.ng.ml.. lactate.mmol.l. fibrinogen sofa
## 1              1.3                    36             0.9        513    8
## 2              1.3                    36             0.9        513    8
## 3              1.3                    36             0.9        513    8
## 4              1.3                    36             0.9        513    8
## 5              1.3                    36             0.9        513    8
## 6              1.3                    36             0.9        513    8
```

Choosing focuses of interests: genes, continuous covariates and categorical covariates

gene: AASDHPPT continuous covariate: ferritin levels categorical covariate: disease status, gender

```
#subsetting the dataframe so that it only contains columns that I want to perform further analysis on


#getting rid of all unknown values in the data
#unique(all_data$ferritin.ng.ml.)
#unique(all_data$Age)
#used the above lines to look at unique values, and saw that unknown is formatted as " unknown", so wil
all_data[all_data == " unknown"] <- NA

# Drop rows with any NA values
all_data <- na.omit(all_data)

#using select to subset https://www.educative.io/answers/what-is-the-select-function-in-r
covid_data <- all_data %>% select(participant_id, gene,expression, ferritin.ng.ml., sex, disease_status

#taking a look at the new subsetted data frame
head(covid_data)
```

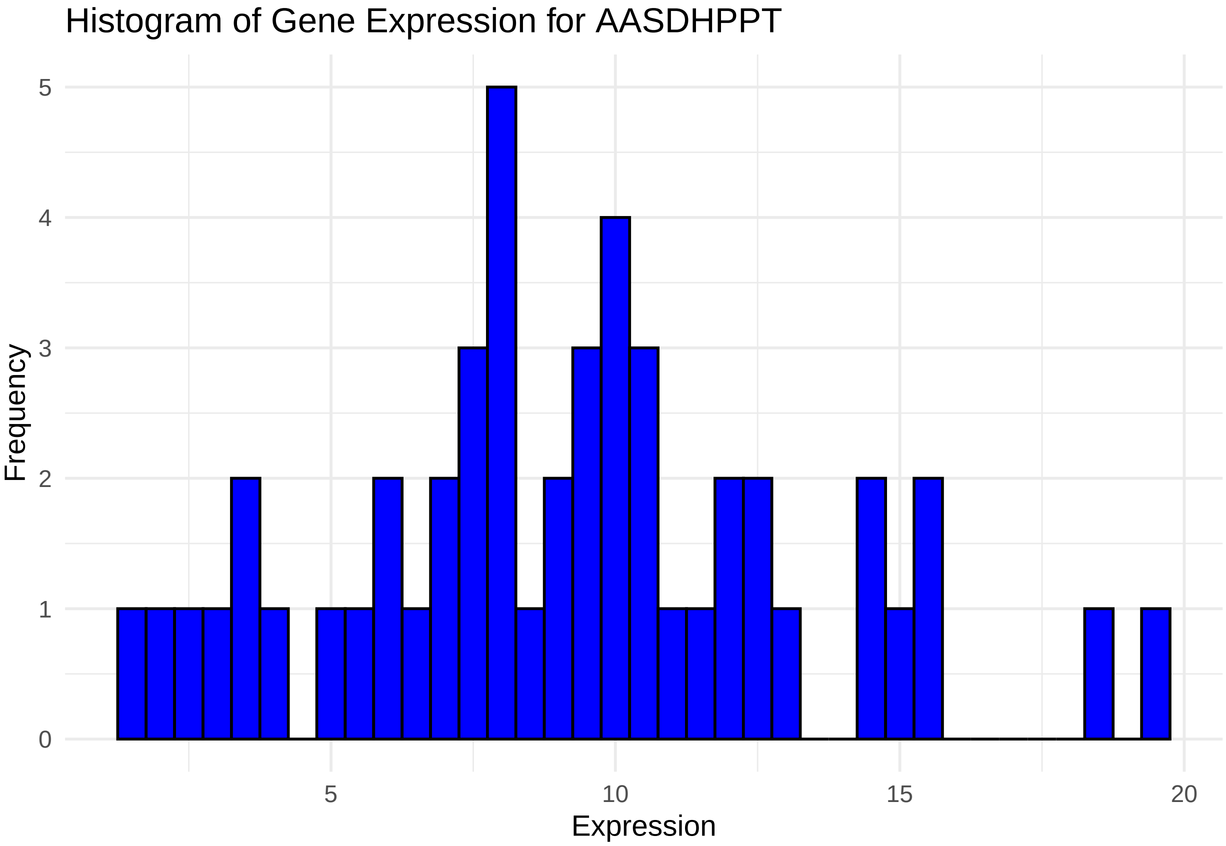
```
##             participant_id    gene expression ferritin.ng.ml.   sex
## 1 COVID_01_39y_male_NonICU    A1CF       0.00             946  male
## 2 COVID_01_39y_male_NonICU    A1BG       0.49             946  male
## 3 COVID_01_39y_male_NonICU   AADAC       0.00             946  male
## 4 COVID_01_39y_male_NonICU AADACL2       0.00             946  male
## 5 COVID_01_39y_male_NonICU AADACL3       0.00             946  male
## 6 COVID_01_39y_male_NonICU AADACL4       0.00             946  male
##            disease_status
## 1 disease state: COVID-19
## 2 disease state: COVID-19
## 3 disease state: COVID-19
## 4 disease state: COVID-19
## 5 disease state: COVID-19
## 6 disease state: COVID-19
```

Plotting gene expression (histogram)

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v purrr     1.0.2
## v lubridate 1.9.2     v stringr   1.5.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
AASDHPPT_data <- covid_data %>%
  filter(gene == "AASDHPPT") %>%
  select(expression) %>%
  mutate(expression = as.numeric(expression))  # Ensure expression is numeric

# Plot histogram
ggplot(AASDHPPT_data, aes(x =as.numeric(expression))) +
  geom_histogram(binwidth = 0.5, color = "black", fill = "blue") +
  labs(title = "Histogram of Gene Expression for AASDHPPT",
       x = "Expression",
       y = "Frequency") +
  theme_minimal() #choosing preferred theme
```



Plotting gene expression vs. ferritin (scatterplot)

```r
# referred to these sites when plotting
# https://r-graph-gallery.com/interactive-charts.html
# https://r-graph-gallery.com/scatterplot.html
# http://www.sthda.com/english/wiki/ggplot2-scatter-plots-quick-start-guide-r-software-and-data-visuali

# using plotly to add interactive element to the graph
#install.packages("plotly")
library(plotly)
```

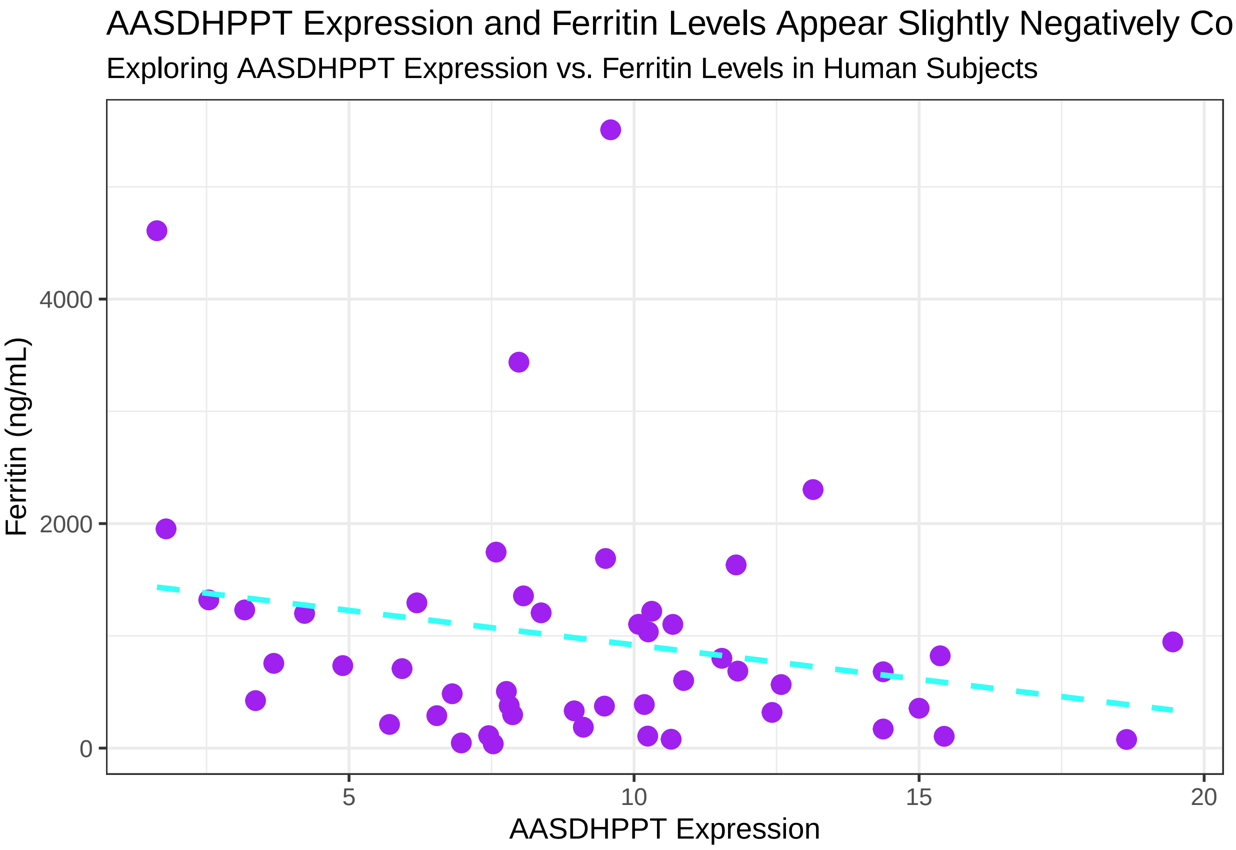
```
##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

```r
# subsetting the data so that it only contains rows where gene = AASDHPPT and also the sex variable so
data <- covid_data %>%
  filter(gene == "AASDHPPT")

# certain columns in the dataset with numeric values are stored in character types, setting them to num
data$expression <- as.numeric(data$expression)
data$ferritin.ng.ml. <- as.numeric(data$ferritin.ng.ml. )

scatter_plot <- ggplot(data, aes(x = expression, y = ferritin.ng.ml.)) +
    geom_point(color = "purple",size = 3) + #increasing size of the data points on the graoh
    geom_smooth(method = "lm", se = FALSE, color = "#33FFF7",linetype = "dashed") +
  labs(title = "AASDHPPT Expression and Ferritin Levels Appear Slightly Negatively Correlated",
      subtitle = "Exploring AASDHPPT Expression vs. Ferritin Levels in Human Subjects",
       x= "AASDHPPT Expression" ,
       y = "Ferritin (ng/mL)") +
  theme_bw() # changing to preferred theme +

interactive_scatter <- ggplotly(scatter_plot)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```r
# displaying regular graph and interactive
scatter_plot
```

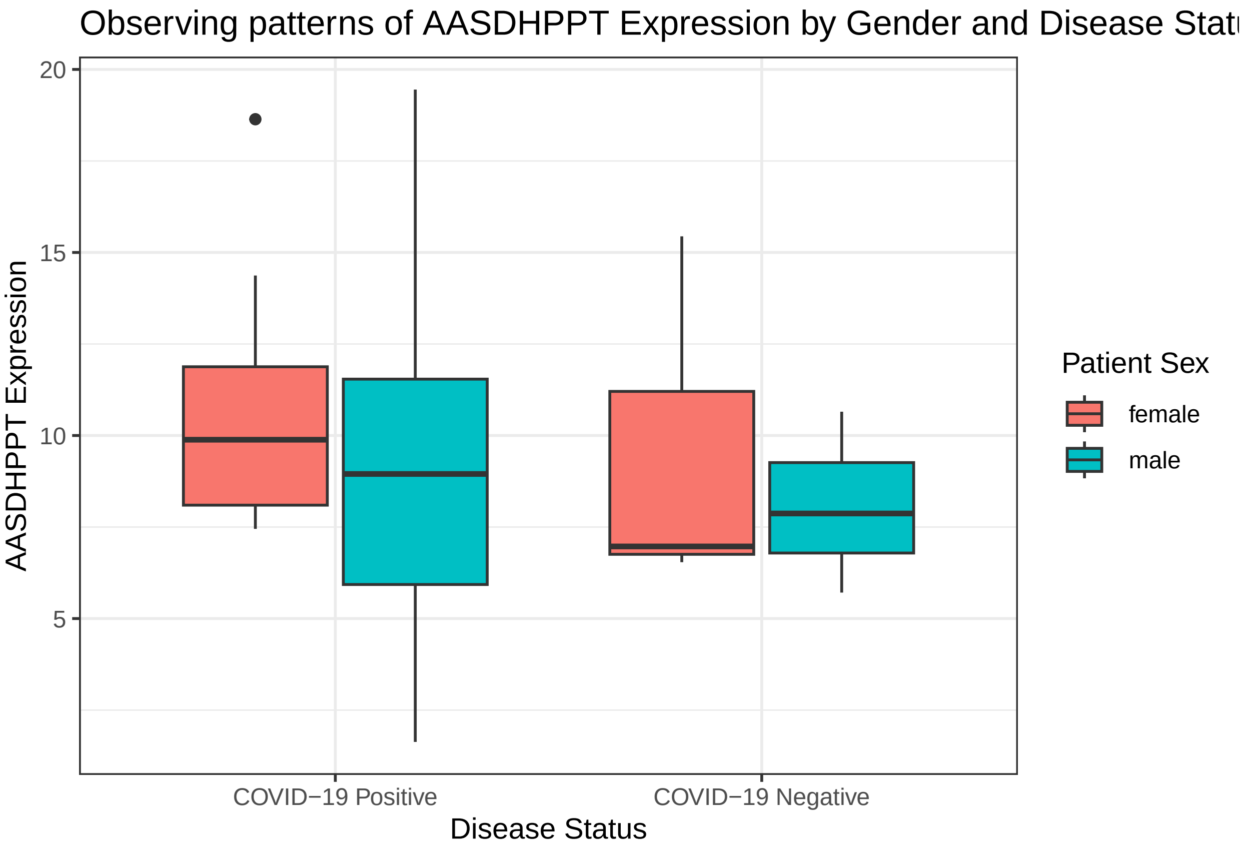
```
## `geom_smooth()` using formula = 'y ~ x'
```



```
interactive_scatter
```

Plotting gene expression vs.disease status by sex (boxplot)

```
# creating boxplot
ggplot(data, aes(x = disease_status, y = as.numeric(expression), fill = sex)) +
  geom_boxplot() +
  scale_x_discrete(labels = c("disease state: COVID-19" = "COVID-19 Positive", "disease state: non-COVI
  labs(title = "Observing patterns of AASDHPPT Expression by Gender and Disease Status", x= "Disease Sta
  theme_bw() # changing to preferred theme
```


Observing patterns of AASDHPPT Expression by Gender and Disease Statu
20
15
10
5
AASDHPPT Expression
COVID−19 Positive
COVID−19 Negative
Disease Status
Patient Sex
female
male