

Project Presentation

Sai Lakkireddy

2023-22-08

Chosen Gene: AAMP

AAMP stands for “Angio-Associated Migratory Cell Protein” gene. This gene is responsible for making the AAMP protein, which is involved in cell movement and angiogenesis, contributing to processes like wound healing and tissue repair. Genes and their corresponding proteins are crucial for the proper functioning of our bodies.

In this analysis, we study the association between AAMP’s gene expression, age, COVID Status and ICU Status

Importing and combining the Data from two csv files

Steps:

1. Import both the csv files
2. Convert the gene expression into long format
3. Inner join it with meta data

```
#set current working directory to the previous folder
setwd("../")

#import the geneExpression and metaData csv from the data folder
data_gene_Expression <- read.csv("data/QBS103_finalProject_geneExpression.csv", header=TRUE)
data_meta <- read.csv("data/QBS103_finalProject_metadata.csv", header=TRUE)

#We convert the geneExpression data from wide form into long form
data_gene_Expression.longForm <- data_gene_Expression %>%
  pivot_longer(cols = starts_with(c("COVID_", "NONCOVID_")),
    names_to = "participant_id",
    values_to = "gene_expression_value"
  )

#make a final data frame by combining the two data sets and making it a data frame
final_df <- as.data.frame(data_gene_Expression.longForm %>% inner_join( data_meta,
  by=c('participant_id'))

#head(final_df)
```

Pre-processing the data

Steps:

1. Remove “unknown” strings and prefixes
2. Convert the class the columns to their appropriate type

```
#rename with x column with gene
final_df <- rename(final_df, gene = X)

#remove all unknown strings and substitute it with NAs
final_df[, 16:27][final_df[, 16:27] == ' unknown' | final_df[, 16:27] == 'unknown'] <- NA

#format the disease status column to just include the status
final_df$disease_status <- sub('disease state: ', '', final_df$disease_status)

#convert the column type of disease_status, sex, icu_status and mechanical_ventilation to factor
final_df <- final_df %>%
  mutate_at(vars(disease_status, sex, icu_status, mechanical_ventilation), as.factor)

#convert the class of age, charlson_score
final_df <- final_df %>%
  mutate_at(vars(age, apacheii, ferritin.ng.ml.,
                  crp.mg.l., ddimer.mg.l.fe.,
                  procalcitonin.ng.ml., lactate.mmol.l., fibrinogen, sofa), as.integer)

#head(final_df)
```

Optional - handle missing values

```
# check all the numeric columns
num_cols <- names(select_if(final_df, is.numeric))

# Create an imputation model
imputation_model <- mice(final_df[num_cols], method = "pmm", printFlag = FALSE)

# Perform the imputation
imputed_data_final <- complete(imputation_model)

final_df[num_cols] <- imputed_data_final[num_cols]

#head(final_df)
```

Create a subset the AAMP Gene and the chosen covariates

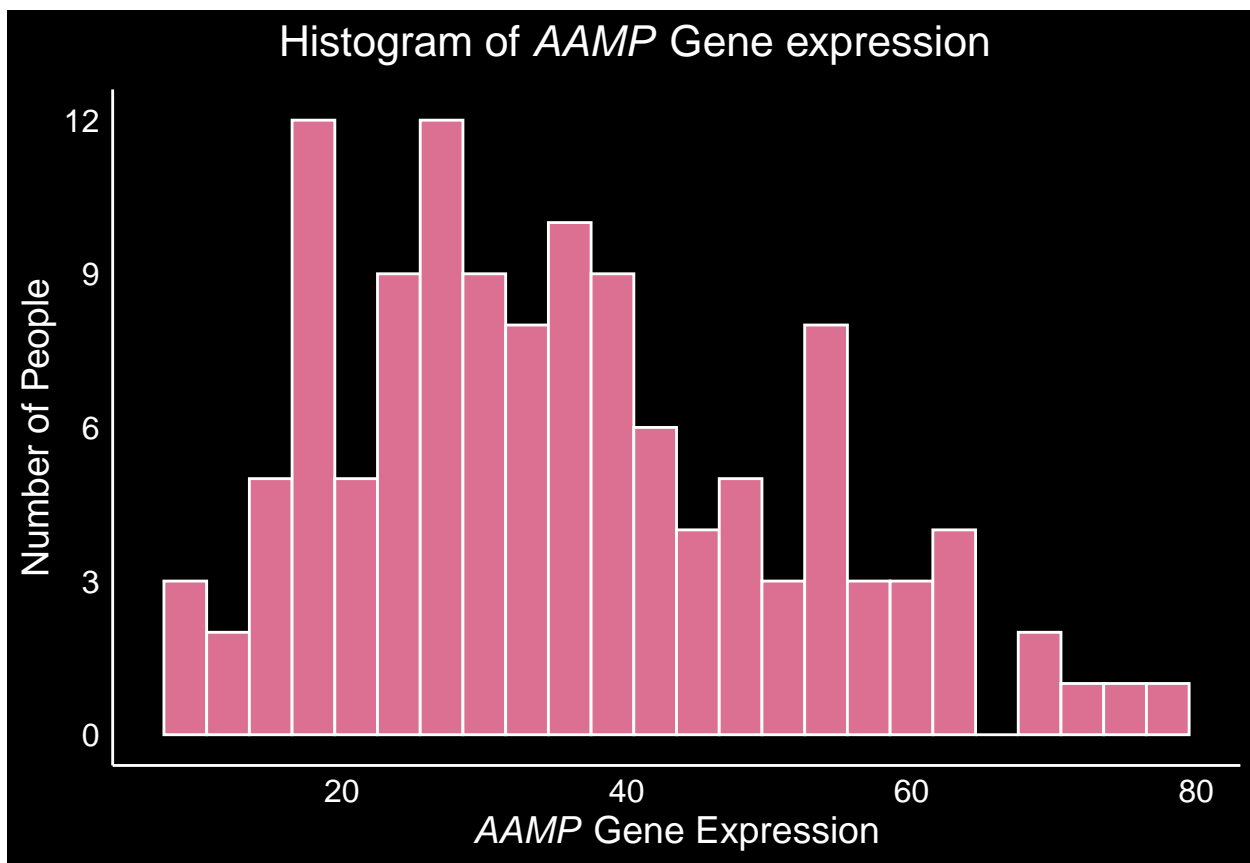
```
final_subset <- final_df[final_df$gene == 'AAMP',
  c('gene',
    'gene_expression_value', 'age',
    'icu_status', 'disease_status')]
```

```
#head(final_subset)
```

Histogram for Gene Expression

```
breaks <- seq(0, 15, by = 3)

# Create the histogram with integer bins
ggplot(final_subset, aes(x = gene_expression_value)) +
  geom_histogram(binwidth = 3, color = "white", fill = "#DB7093") +
  scale_y_continuous(breaks = breaks) +
  theme_minimal() +
  theme(
    panel.background = element_rect(fill = "black"),
    plot.background = element_rect(fill = "black"),
    axis.line = element_line(color = "white"),
    axis.text = element_text(color = "white", size = 12),
    axis.title = element_text(color = "white", size = 14),
    panel.grid = element_blank(),
    plot.title = element_text(color = "white", size = 16, face = "bold", hjust = 0.4),
  ) +
  ggtitle(expression(paste("Histogram of ", italic("AAMP"), " Gene expression")))) +
  xlab(expression(paste(italic("AAMP"), " Gene Expression")))) +
  ylab("Number of People")
```

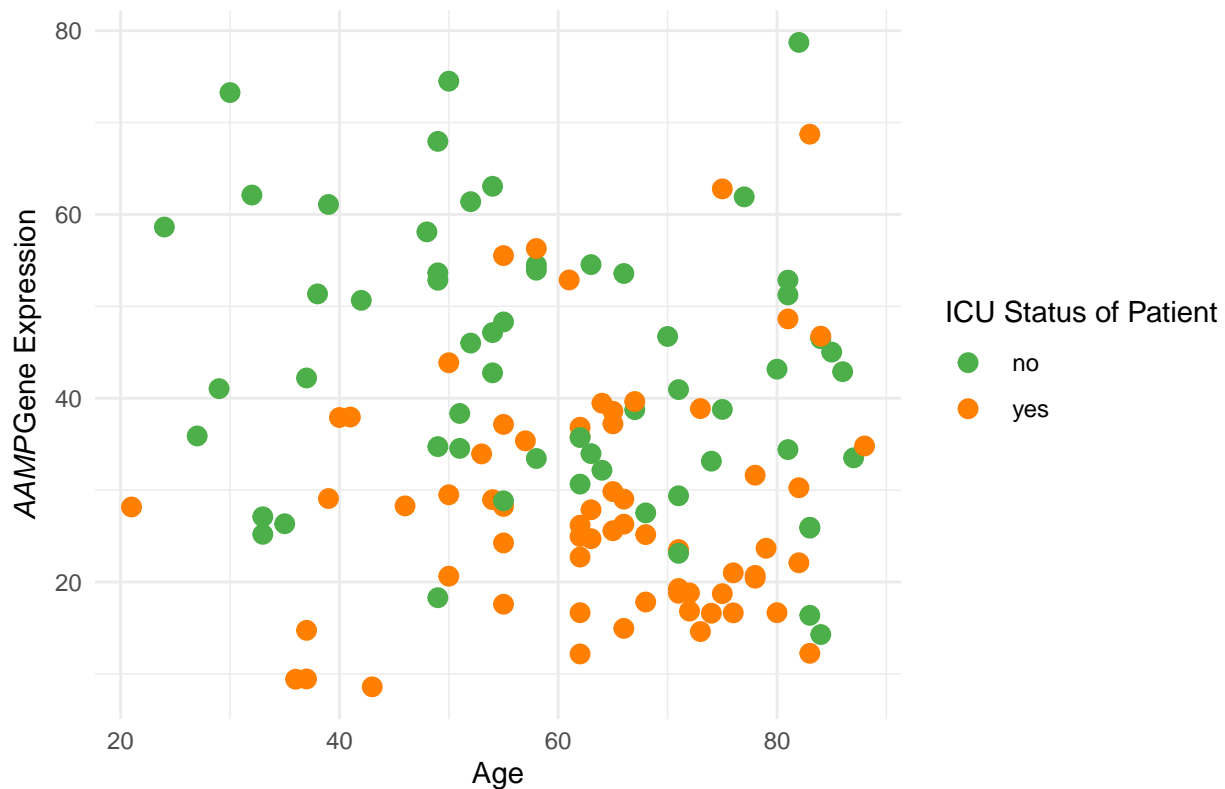


Scatter plot: Age vs Gene Expression factoring for ICU status

```
my_colors_1 <- c("#4DAF4A", "#FF7F00")

# Create the scatter plot with custom color scheme
ggplot(final_subset, aes(x = age, y = gene_expression_value, color = icu_status)) +
  geom_point(size = 3) +
  scale_color_manual(values = my_colors_1, name = "ICU Status of Patient") +
  theme_minimal() +
  theme(
    plot.title = element_text(color = "navy", size = 13, face = "bold", hjust = 0.4)
  ) +
  ggtitle(expression(paste("Scatter Plot: ", italic("AAMP"), " Gene Expression vs Age and ICU Status"))) +
  xlab("Age") +
  ylab(expression(paste(italic("AAMP"), "Gene Expression")))
```

Scatter Plot: AAMP Gene Expression vs Age and ICU Status



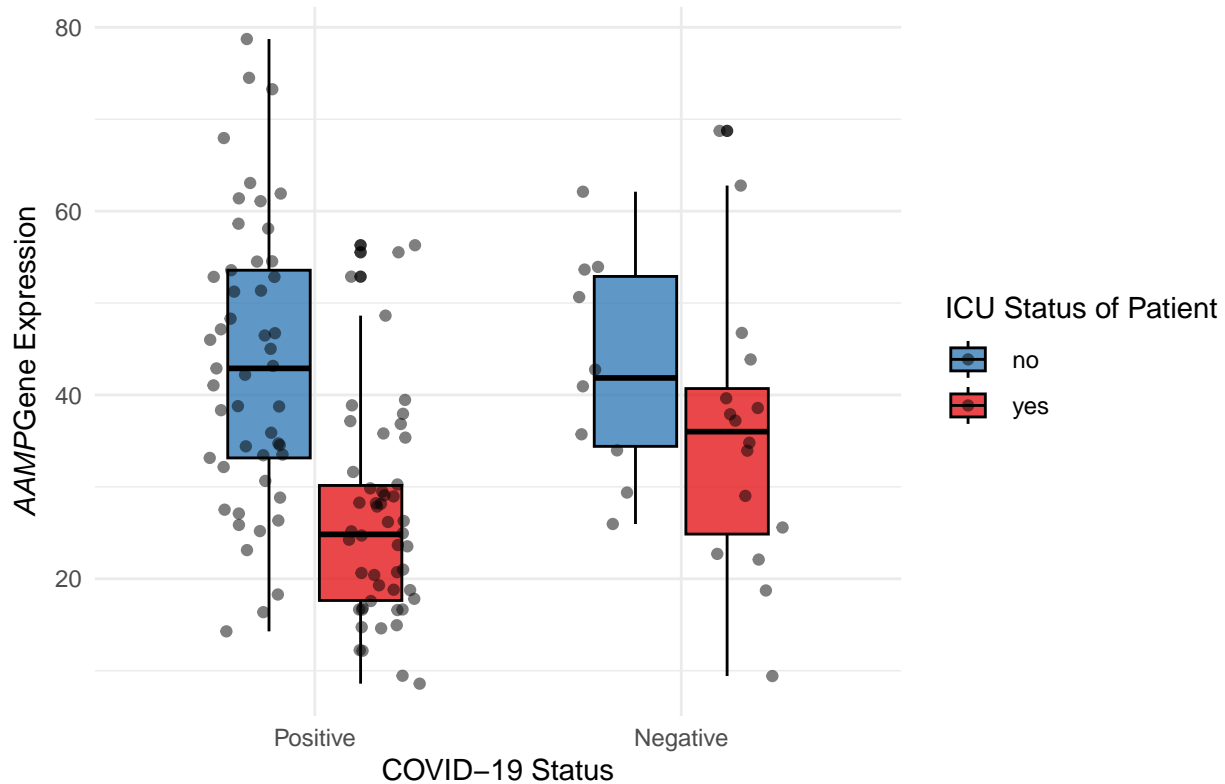
```
ggplot(final_subset, aes(x = age, y = gene_expression_value)) +
  geom_hex(bins = 20) +
  scale_fill_viridis_c() +
  labs(title = "Hexbin Scatter Plot: Gene Expression Value vs Age",
       x = "Age",
       y = "Gene Expression Value") +
  ggtitle(expression(paste("Hex Bin Plot: ", italic("AAMP"), " Gene Expression vs Age"))) +
  theme_minimal()
```



Box plot: Gene Expression by COVID and ICU Status

```
my_colors_2 <- c("#377eb8", "#e41a1c")
ggplot(final_subset, aes(x = disease_status, y = gene_expression_value, fill = icu_status)) +
  geom_boxplot(color = "black", width = 0.5, alpha = 0.8) +
  scale_fill_manual(values = my_colors_2, name = "ICU Status of Patient") +
  geom_jitter(position = position_jitterdodge(), alpha = 0.5) +
  scale_x_discrete(labels = c("COVID-19" = "Positive", "non-COVID-19" = "Negative")) +
  theme_minimal() +
  theme(
    plot.title = element_text(color = "darkgreen", size = 13, face = "bold", hjust = 0.4)
  ) +
  ggtitle(expression(paste("Box Plot: ", italic("AAMP"), " Gene Expression by COVID and ICU Status"))) +
  xlab("COVID-19 Status") +
  ylab(expression(paste(italic("AAMP"), "Gene Expression")))
```

Box Plot: AAMP Gene Expression by COVID and ICU Status



Plots generated by a function for genes ABHD18, AAMP and ABHD17C

```
my_plots_function <- function(dataFrame, genes.list, cont.covariate, cat.covariates) {
  #breaks <- seq(min(freque), 15, by = 3)
  my_colors_1 <- c("#4DAF4A", "#FF7F00")
  my_colors_2 <- c("#377eb8", "#e41a1c")
  all.plots.list <- list()

  for (gene in genes.list) {
    gene_subset <- final_df[final_df$gene == gene,
                           c('gene', 'gene_expression_value', paste(cont.covariate), paste(cat.covariates))

    histogram <- ggplot(gene_subset, aes(x = gene_expression_value)) +
      geom_histogram(binwidth = 1, color = "white", fill = "#DB7093") +
      #scale_y_continuous(breaks = breaks) +
      theme_minimal() +
      theme(
        panel.background = element_rect(fill = "black"),
        plot.background = element_rect(fill = "black"),
        axis.line = element_line(color = "white"),
        axis.text = element_text(color = "white", size = 12),
        axis.title = element_text(color = "white", size = 14),
        panel.grid = element_blank(),
```

```

    plot.title = element_text(color = "white", size = 16, face = "bold", hjust = 0.4),
  ) +
  ggtitle(substitute(Histogram ~ of ~ italic(gene) ~ Gene ~ expression, list(gene = gene))) +
  xlab(substitute(italic(gene) ~ Gene ~ Expression, list(gene = gene))) +
  ylab("Number of People")

scatter.plot <- ggplot(gene_subset, aes(y = gene_expression_value, x = gene_subset[[cont.covariate]],
geom_point(size = 3) +
scale_color_manual(values = my_colors_1, name = "ICU Status of Patient") +
theme_minimal() +
  theme(
    plot.title = element_text(color = "navy", size = 13, face = "bold", hjust = 0.4)
  ) +
  ggtitle(substitute(Scatter ~ Plot ~ italic(gene) ~ Gene ~ Expression ~ vs ~ Age ~ and ~ ICU ~ Status)) +
  xlab(substitute(cont.covariate)) +
  ylab(substitute(italic(gene) ~ Gene ~ Expression))

box.plot <- ggplot(gene_subset, aes(x = gene_subset[[cat.covariates[1]]], y = gene_expression_value, fill =
geom_boxplot(color = "black", width = 0.5, alpha = 0.8) +
scale_fill_manual(values = my_colors_2, name = "ICU Status of Patient") +
geom_jitter(position = position_jitterdodge(), alpha = 0.5) +
scale_x_discrete(labels = c("COVID-19" = "Positive", "non-COVID-19" = "Negative")) +
theme_minimal() +
  theme(
    plot.title = element_text(color = "darkgreen", size = 13, face = "bold", hjust = 0.4)
  ) +
  ggtitle(substitute(Box ~ plot ~ of ~ italic(gene) ~ Gene ~ Expression ~ by ~ COVID ~ and ~ ICU ~ Status)) +
  xlab("COVID-19 Status") +
  ylab(substitute(italic(gene) ~ Gene ~ Expression))

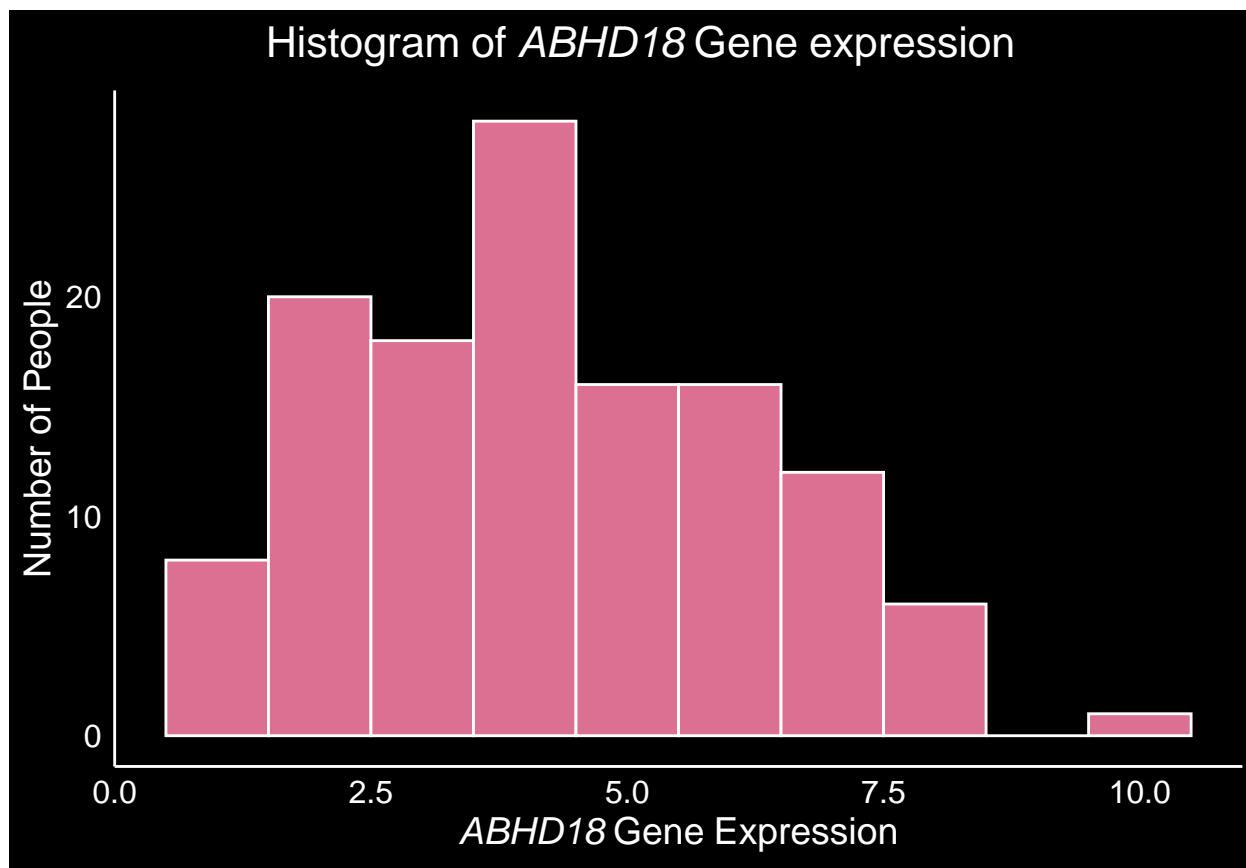
  all.plots.list[[gene]] <- list(histogram = histogram, scatter.plot = scatter.plot, box.plot = box.plot)
}

return(all.plots.list)
}

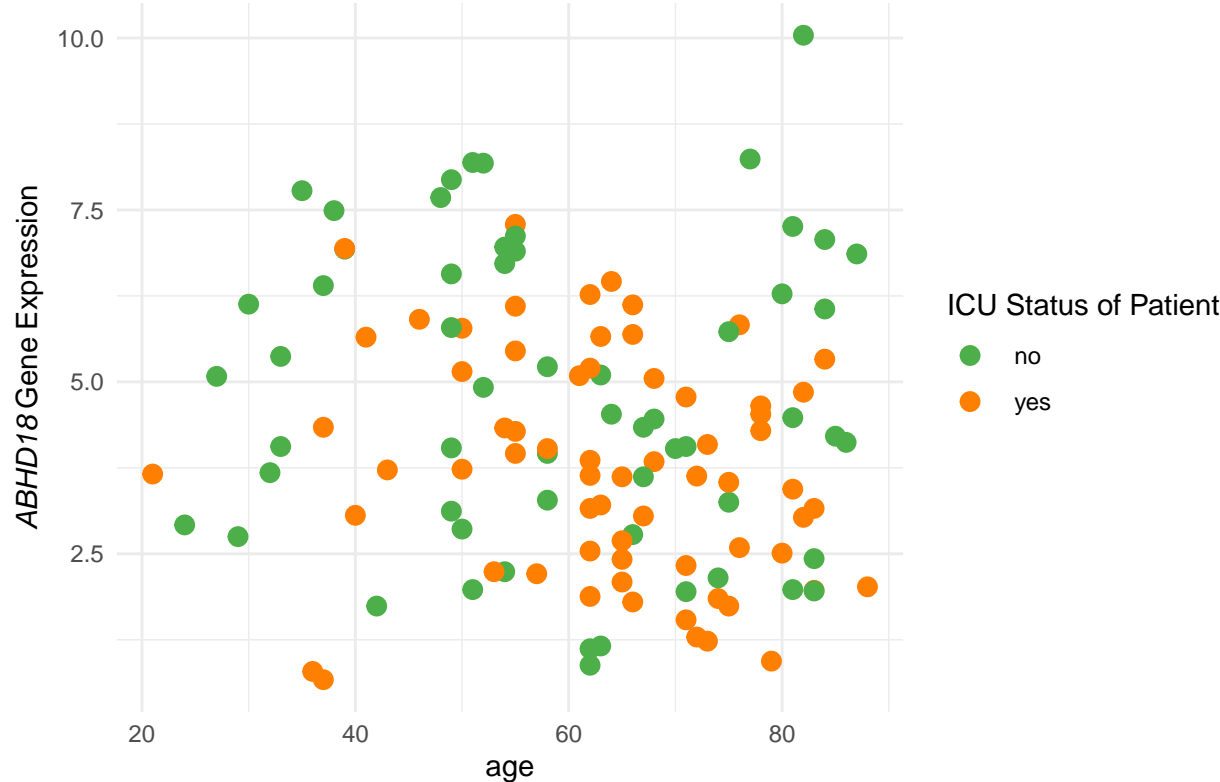
all.plots.list <- my_plots_function(final_subset, c('ABHD18', 'AAMP', 'ABHD17C'), 'age', c('disease_status', 'sex'))

for (gene in names(all.plots.list)) {
  print(all.plots.list[[gene]]$histogram)
  print(all.plots.list[[gene]]$scatter.plot)
  print(all.plots.list[[gene]]$box.plot)
}

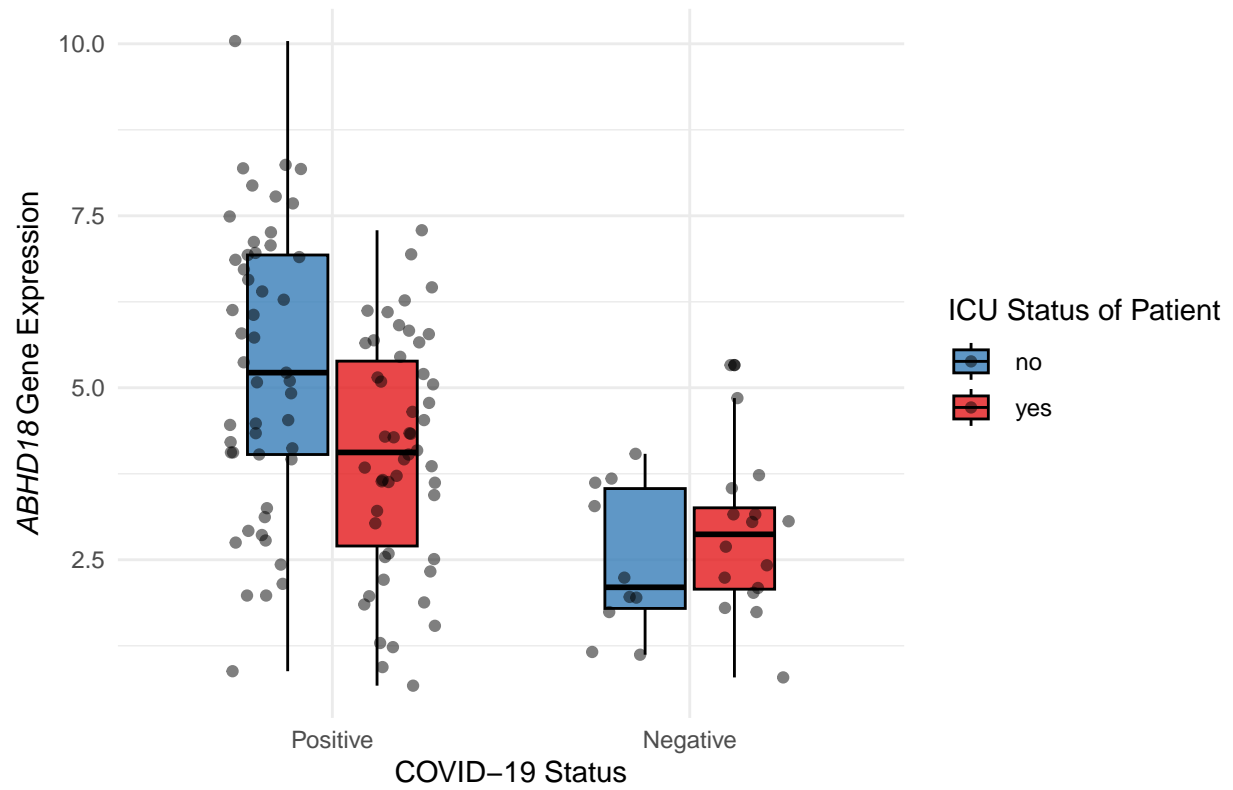
```

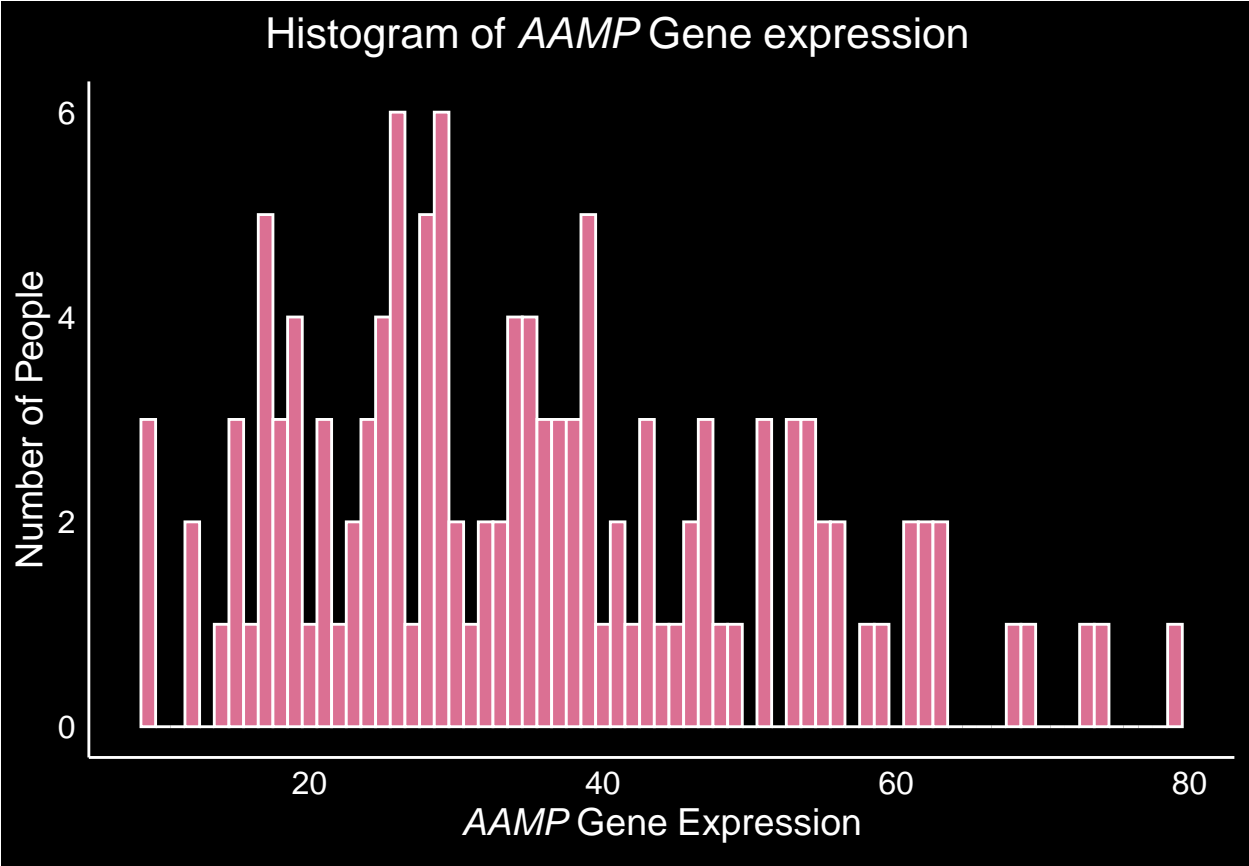


Scatter Plot *ABHD18* Gene Expression vs Age and ICU Status

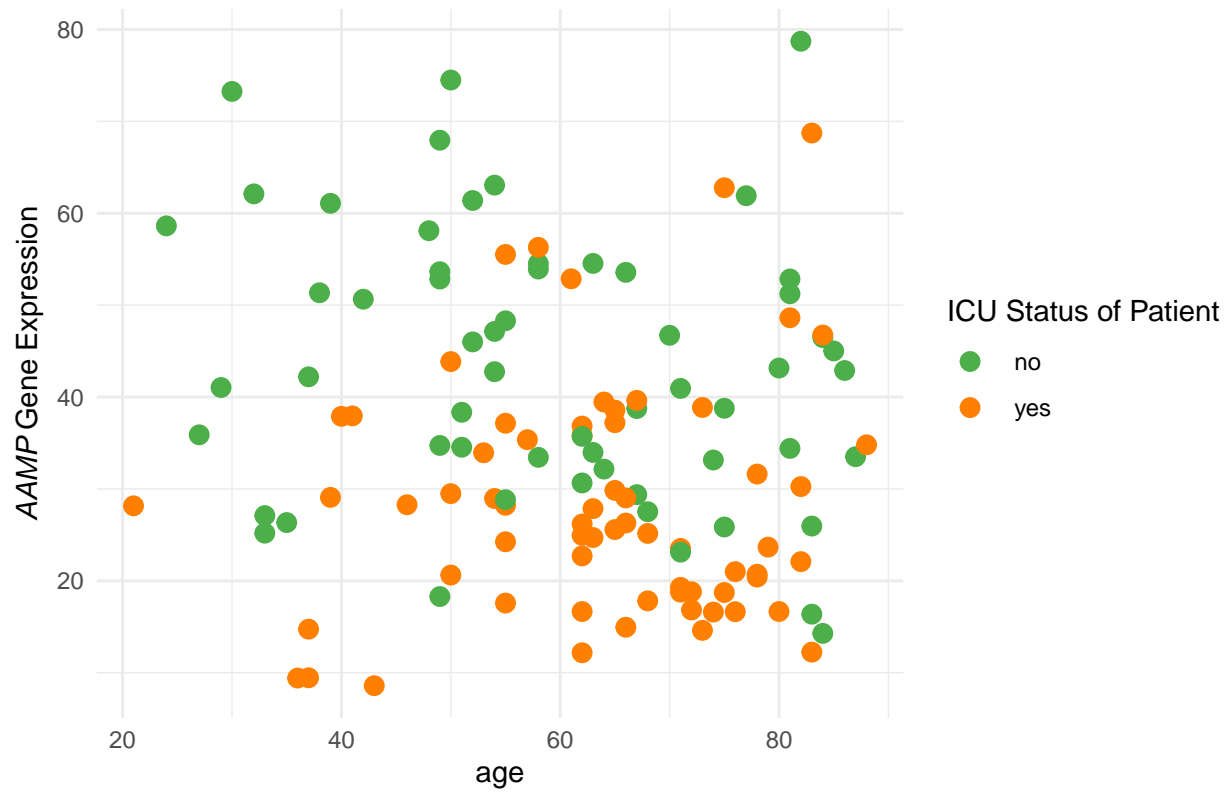


Box plot of *ABHD18* Gene Expression by COVID and ICU Status

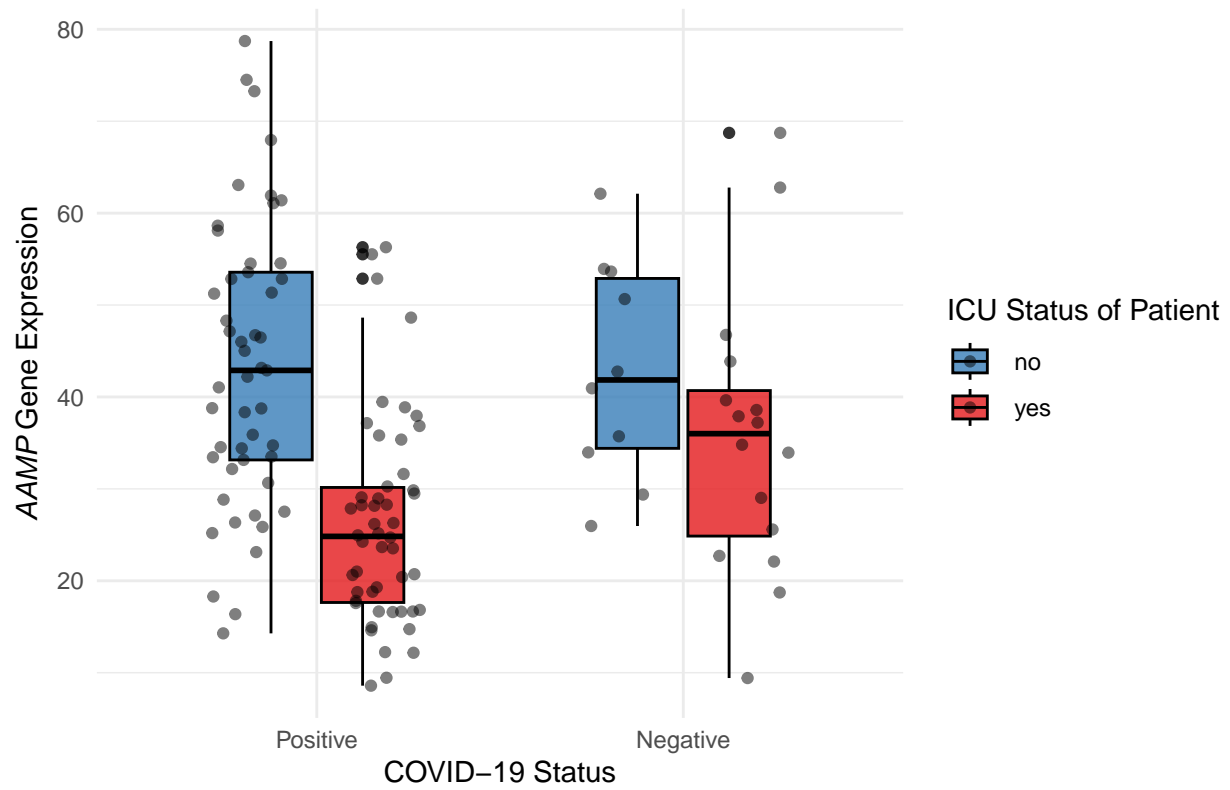




Scatter Plot AAMP Gene Expression vs Age and ICU Status

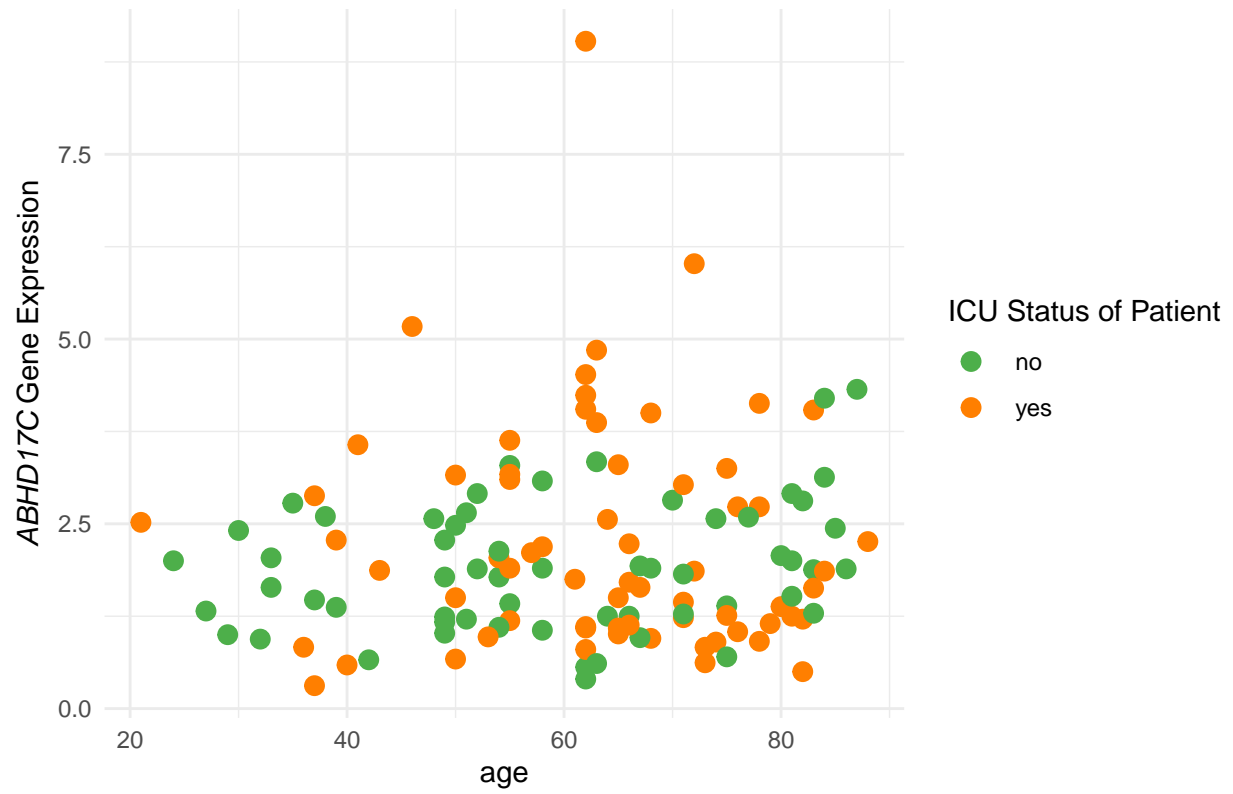


Box plot of *AAMP* Gene Expression by COVID and ICU Status

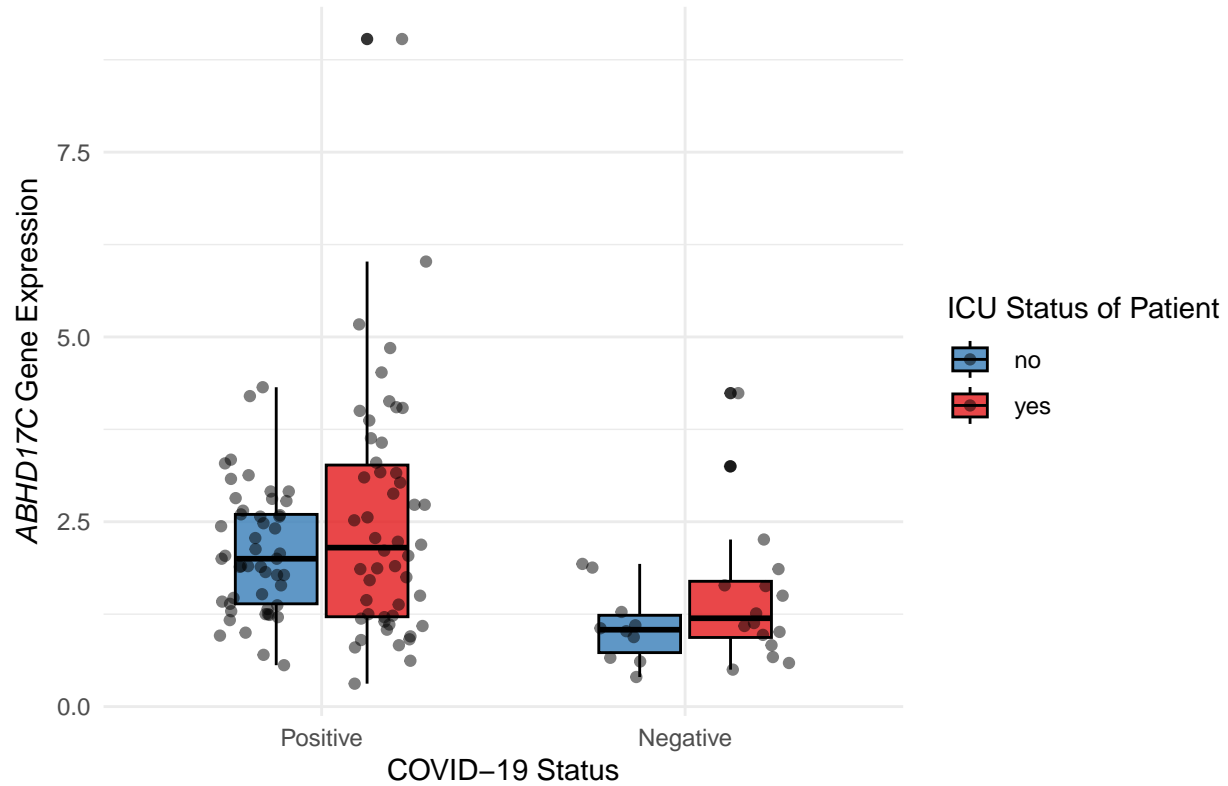




Scatter Plot *ABHD17C* Gene Expression vs Age and ICU Status



Box plot of *ABHD17C* Gene Expression by COVID and ICU Status



##Final presentation subset with 3 (1+2) Continuous co-variables and 3(2+1) co-variables

```
final_df_subset <- final_df[, c('participant_id', 'gene', 'gene_expression_value', 'age', 'charlson_score',
                                'fibrinogen', 'icu_status', 'disease_status', 'mechanical_ventilation')]

head(final_df_subset)
```

```
##      participant_id  gene gene_expression_value age charlson_score
## 1 COVID_01_39y_male_NonICU APOA1             0.00 39             0
## 2 COVID_02_63y_male_NonICU APOA1             0.12 63             2
## 3 COVID_03_33y_male_NonICU APOA1             0.00 33             2
## 4 COVID_04_49y_male_NonICU APOA1             0.09 49             1
## 5 COVID_05_49y_male_NonICU APOA1             0.08 49             1
## 6 COVID_07_38y_female_NonICU APOA1            0.00 38             7
##      fibrinogen icu_status disease_status mechanical_ventilation
## 1          513      no      COVID-19      yes
## 2          304      no      COVID-19      no
## 3          513      no      COVID-19      no
## 4          949      no      COVID-19      no
## 5          929      no      COVID-19      yes
## 6          478      no      COVID-19      no
```

##Used the code from the class to get the statistics table ##Put the table in the latex file (cannot knit it)


```

#library(knitr)
#library(kableExtra)

table1 <- data.frame(
  'Variable' = c('Gene Expression Value mean (sd)',
                'Age mean (sd)',
                'Charlson Score (sd)',
                'Fibrinogen (sd)',
                'Disease Status n (%)',
                'COVID-19',
                'non-COVID-19',
                'Mechanical Ventilation n (%)',
                'yes',
                'no'),
  'Value' = c(
    paste(round(mean(final_df_subset$gene_expression_value), 2),
          ' (',
          round(sd(final_df_subset$gene_expression_value), 2),
          ')'),
    paste(round(mean(final_df_subset$age), 2),
          ' (',
          round(sd(final_df_subset$age), 2),
          ')'),
    paste(round(mean(final_df_subset$charlson_score), 2),
          ' (',
          round(sd(final_df_subset$charlson_score), 2),
          ')'),
    paste(round(mean(final_df_subset$fibrinogen), 2),
          ' (',
          round(sd(final_df_subset$fibrinogen), 2),
          ')'),
    ' ',
    paste(round(nrow(final_df_subset[final_df_subset$disease_status == "COVID-19", ]) * 100 / nrow(final_df_subset)),
          '%'),
    paste(round(nrow(final_df_subset[final_df_subset$disease_status == "non-COVID-19", ]) * 100 / nrow(final_df_subset)),
          '%'), ' ',
    paste(round(nrow(final_df_subset[final_df_subset$mechanical_ventilation == " yes", ]) * 100 / nrow(final_df_subset)),
          '%'),
    paste(round(nrow(final_df_subset[final_df_subset$mechanical_ventilation == " no", ]) * 100 / nrow(final_df_subset)),
          '%')
  )
)

#kable(x = table1, caption = 'Summary Statistics Table 1',
      #col.names = c("Variable", "Value"),
      #align = c('l','c'), escape = FALSE) %>%
#add_indent(positions = c(6,7,9,10)) %>%
#kable_styling(bootstrap_options = "striped", full_width = FALSE)

```

```
##Convert long format to wide format and drop NA rows
```

```
library(tidyr)
```

```

pivoted_final_df <- reshape(
  data = final_df_subset,
  idvar = c("participant_id", "age", "charlson_score", "fibrinogen", "icu_status", "disease_status", "mechanical_ventilation"),
  timevar = "gene",
  direction = "wide"
)

head(pivoted_final_df)

```

```

##           participant_id age charlson_score fibrinogen icu_status
## 1 COVID_01_39y_male_NonICU 39              0         513       no
## 2 COVID_02_63y_male_NonICU 63              2         304       no
## 3 COVID_03_33y_male_NonICU 33              2         513       no
## 4 COVID_04_49y_male_NonICU 49              1         949       no
## 5 COVID_05_49y_male_NonICU 49              1         929       no
## 6 COVID_07_38y_female_NonICU 38             7         478       no
## disease_status mechanical_ventilation gene_expression_value.APOA1
## 1 COVID-19              yes              0.00
## 2 COVID-19              no              0.12
## 3 COVID-19              no              0.00
## 4 COVID-19              no              0.09
## 5 COVID-19              yes              0.08
## 6 COVID-19              no              0.00
## gene_expression_value.APOA2 gene_expression_value.APOM
## 1              0.20              1.97
## 2              NA              NA
## 3              0.00              1.15
## 4              0.00              3.79
## 5              0.21              1.39
## 6              0.00              5.90
## gene_expression_value.PRTN3 gene_expression_value.LCN2
## 1              1.31              87.71
## 2              NA              NA
## 3              1.30             121.86
## 4              0.08              31.73
## 5              0.00              16.42
## 6              0.30              69.86
## gene_expression_value.CD24 gene_expression_value.BPI
## 1              9.51             19.05
## 2              NA              NA
## 3             14.29             30.39
## 4             31.38             14.03
## 5              7.95              8.49
## 6             30.54             24.10
## gene_expression_value.CTSG gene_expression_value.DEFA1
## 1              3.51             124.42
## 2             77.92              NA
## 3              4.47             173.99
## 4              8.65             312.46
## 5              0.46              71.83
## 6              6.01             424.39
## gene_expression_value.DEFA4 gene_expression_value.MMP8
## 1             24.85             32.68

```

## 2	NA	NA
## 3	45.57	26.22
## 4	42.34	7.07
## 5	3.49	14.84
## 6	54.09	17.41
##	gene_expression_value.MPO	gene_expression_value.AGT
## 1	15.38	0.00
## 2	NA	NA
## 3	2.29	0.03
## 4	4.48	0.03
## 5	0.76	0.00
## 6	4.82	0.00
##	gene_expression_value.FBLN5	gene_expression_value.NID1
## 1	1.19	10.92
## 2	NA	NA
## 3	0.22	0.71
## 4	3.05	4.07
## 5	0.53	0.61
## 6	3.57	3.66
##	gene_expression_value.SERPINB1	gene_expression_value.GPLD1
## 1	279.25	0.11
## 2	NA	NA
## 3	237.35	0.13
## 4	245.49	0.30
## 5	599.93	0.19
## 6	180.03	0.31
##	gene_expression_value.CLEC3B	gene_expression_value.VWF
## 1	1.41	0.87
## 2	NA	NA
## 3	0.50	0.26
## 4	0.70	1.23
## 5	0.10	0.52
## 6	0.33	0.94
##	gene_expression_value.A1BG	gene_expression_value.A1CF
## 1	0.49	0.00
## 2	NA	NA
## 3	0.26	0.00
## 4	0.45	0.01
## 5	0.17	0.00
## 6	0.49	0.01
##	gene_expression_value.A2M	gene_expression_value.A2ML1
## 1	0.21	0.04
## 2	NA	NA
## 3	0.03	0.02
## 4	0.09	0.07
## 5	0.00	0.05
## 6	0.23	0.03
##	gene_expression_value.A3GALT2	gene_expression_value.A4GALT
## 1	0.07	0
## 2	0.00	0
## 3	0.00	0
## 4	0.00	0
## 5	0.07	0
## 6	0.07	0

##	gene_expression_value.A4GNT	gene_expression_value.AAAS
## 1	0.03	18.92
## 2	NA	NA
## 3	0.07	13.85
## 4	0.00	22.11
## 5	0.00	8.45
## 6	0.00	28.59
##	gene_expression_value.AACS	gene_expression_value.AADAC
## 1	4.07	0
## 2	NA	NA
## 3	1.83	0
## 4	4.22	0
## 5	1.17	0
## 6	4.24	0
##	gene_expression_value.AADACL2	gene_expression_value.AADACL3
## 1	0	0
## 2	NA	NA
## 3	0	0
## 4	0	0
## 5	0	0
## 6	0	0
##	gene_expression_value.AADACL4	gene_expression_value.AADAT
## 1	0	0.00
## 2	NA	NA
## 3	0	0.00
## 4	0	0.00
## 5	0	0.03
## 6	0	0.03
##	gene_expression_value.AAGAB	gene_expression_value.AAK1
## 1	22.93	7.12
## 2	NA	NA
## 3	18.27	3.92
## 4	26.69	8.74
## 5	17.02	7.26
## 6	26.28	12.78
##	gene_expression_value.AAMDC	gene_expression_value.AAMP
## 1	17.19	61.08
## 2	NA	54.54
## 3	13.35	25.19
## 4	17.53	67.95
## 5	10.14	18.29
## 6	12.62	51.35
##	gene_expression_value.AANAT	gene_expression_value.AAR2
## 1	0.31	21.59
## 2	NA	NA
## 3	0.65	8.72
## 4	0.08	20.83
## 5	1.02	7.46
## 6	0.13	23.90
##	gene_expression_value.AARD	gene_expression_value.AARS1
## 1	0.18	13.52
## 2	NA	NA
## 3	0.05	5.60
## 4	0.03	15.34

## 5	0.10	6.29
## 6	0.12	29.47
##	gene_expression_value.AARS2	gene_expression_value.AARSD1
## 1	2.43	9.63
## 2	NA	11.80
## 3	1.08	4.04
## 4	2.69	14.61
## 5	0.55	5.32
## 6	4.35	21.09
##	gene_expression_value.AASDH	gene_expression_value.AASDHPPT
## 1	6.38	19.45
## 2	NA	NA
## 3	2.76	7.64
## 4	8.23	18.94
## 5	5.23	11.54
## 6	8.90	23.49
##	gene_expression_value.AASS	gene_expression_value.AATF
## 1	0.21	45.83
## 2	NA	NA
## 3	0.04	42.35
## 4	0.41	41.92
## 5	0.21	30.56
## 6	0.63	44.16
##	gene_expression_value.AATK	gene_expression_value.ABAT
## 1	5.60	9.63
## 2	NA	NA
## 3	5.34	4.59
## 4	4.73	11.02
## 5	1.50	4.42
## 6	4.16	12.02
##	gene_expression_value.ABCA1	gene_expression_value.ABCA10
## 1	32.30	0.32
## 2	NA	NA
## 3	34.38	0.29
## 4	14.24	0.31
## 5	18.39	0.19
## 6	14.66	0.55
##	gene_expression_value.ABCA12	gene_expression_value.ABCA13
## 1	0	0.49
## 2	NA	NA
## 3	0	0.26
## 4	0	0.13
## 5	0	0.16
## 6	0	0.23
##	gene_expression_value.ABCA2	gene_expression_value.ABCA3
## 1	8.47	0.37
## 2	NA	NA
## 3	14.24	0.17
## 4	6.37	0.94
## 5	5.90	0.17
## 6	9.16	0.75
##	gene_expression_value.ABCA4	gene_expression_value.ABCA5
## 1	0.01	1.86
## 2	NA	NA

## 3	0.00	2.17
## 4	0.00	2.94
## 5	0.00	1.38
## 6	0.00	3.60
## gene_expression_value.ABCA6	gene_expression_value.ABCA7	
## 1	0.19	39.31
## 2	NA	NA
## 3	0.07	54.85
## 4	0.02	18.91
## 5	0.03	23.28
## 6	0.05	30.95
## gene_expression_value.ABCA8	gene_expression_value.ABCA9	
## 1	0.00	0.27
## 2	NA	NA
## 3	0.00	0.33
## 4	0.01	0.30
## 5	0.00	0.21
## 6	0.00	0.23
## gene_expression_value.ABCB1	gene_expression_value.ABCB10	
## 1	1.61	15.59
## 2	NA	NA
## 3	0.59	4.04
## 4	3.14	10.00
## 5	1.66	5.69
## 6	4.20	15.09
## gene_expression_value.ABCB11	gene_expression_value.ABCB4	
## 1	0.38	0.01
## 2	NA	NA
## 3	0.14	0.09
## 4	0.18	0.74
## 5	0.07	0.21
## 6	0.18	0.81
## gene_expression_value.ABCB5	gene_expression_value.ABCB6	
## 1	0.04	2.77
## 2	NA	NA
## 3	0.15	3.14
## 4	0.11	2.11
## 5	0.04	1.14
## 6	0.11	3.99
## gene_expression_value.ABCB7	gene_expression_value.ABCB8	
## 1	6.42	2.95
## 2	NA	NA
## 3	2.59	1.57
## 4	7.42	2.71
## 5	3.40	0.71
## 6	7.33	3.51
## gene_expression_value.ABCB9	gene_expression_value.ABCC1	
## 1	0.20	11.20
## 2	NA	NA
## 3	0.05	4.74
## 4	0.31	13.05
## 5	0.12	4.29
## 6	1.11	12.24
## gene_expression_value.ABCC10	gene_expression_value.ABCC11	

## 1	8.85	0.04
## 2	NA	NA
## 3	4.08	0.09
## 4	5.75	0.02
## 5	1.91	0.02
## 6	7.33	0.02
##	gene_expression_value.ABCC12	gene_expression_value.ABCC2
## 1	0.00	1.65
## 2	NA	1.31
## 3	0.00	1.72
## 4	0.01	1.55
## 5	0.00	2.56
## 6	0.00	1.57
##	gene_expression_value.ABCC3	gene_expression_value.ABCC4
## 1	7.19	5.96
## 2	NA	NA
## 3	1.44	0.58
## 4	11.11	4.09
## 5	3.13	7.37
## 6	5.34	8.10
##	gene_expression_value.ABCC5	gene_expression_value.ABCC6
## 1	14.83	6.62
## 2	NA	5.87
## 3	13.02	3.13
## 4	17.49	3.37
## 5	5.53	1.66
## 6	23.46	3.61
##	gene_expression_value.ABCC8	gene_expression_value.ABCC9
## 1	0	1.53
## 2	NA	NA
## 3	0	1.97
## 4	0	2.82
## 5	0	1.84
## 6	0	2.31
##	gene_expression_value.ABCD1	gene_expression_value.ABCD2
## 1	11.26	1.55
## 2	NA	NA
## 3	5.83	0.39
## 4	4.80	1.74
## 5	1.93	0.59
## 6	3.96	5.25
##	gene_expression_value.ABCD3	gene_expression_value.ABCD4
## 1	10.62	11.19
## 2	11.23	NA
## 3	3.99	7.83
## 4	12.38	13.60
## 5	8.19	5.43
## 6	13.14	12.96
##	gene_expression_value.ABCE1	gene_expression_value.ABCF1
## 1	14.09	24.40
## 2	NA	NA
## 3	4.50	7.89
## 4	21.20	23.92
## 5	7.16	12.33

## 6	28.46	28.37
##	gene_expression_value.ABCF2	gene_expression_value.ABCF2-H2BE1
## 1	8.77	11.67
## 2	NA	NA
## 3	3.12	3.06
## 4	8.26	12.51
## 5	2.76	4.72
## 6	11.28	13.61
##	gene_expression_value.ABCF3	gene_expression_value.ABCG1
## 1	23.37	32.76
## 2	NA	NA
## 3	11.57	41.14
## 4	23.62	15.72
## 5	9.86	9.43
## 6	25.96	14.96
##	gene_expression_value.ABCG2	gene_expression_value.ABCG4
## 1	0.17	0.01
## 2	NA	NA
## 3	0.15	0.00
## 4	0.14	0.00
## 5	0.03	0.00
## 6	0.05	0.05
##	gene_expression_value.ABCG5	gene_expression_value.ABCG8
## 1	0	0.02
## 2	NA	NA
## 3	0	0.00
## 4	0	0.04
## 5	0	0.04
## 6	0	0.02
##	gene_expression_value.ABHD1	gene_expression_value.ABHD10
## 1	0.04	14.74
## 2	NA	NA
## 3	0.00	6.84
## 4	0.31	16.73
## 5	0.04	6.93
## 6	0.38	18.21
##	gene_expression_value.ABHD11	gene_expression_value.ABHD12
## 1	8.50	14.66
## 2	NA	NA
## 3	8.22	4.84
## 4	9.64	13.22
## 5	4.02	3.36
## 6	8.27	13.35
##	gene_expression_value.ABHD12B	gene_expression_value.ABHD13
## 1	0.55	15.58
## 2	NA	NA
## 3	0.88	9.97
## 4	0.68	16.70
## 5	0.68	21.02
## 6	0.32	16.37
##	gene_expression_value.ABHD14A	gene_expression_value.ABHD14A-ACY1
## 1	6.89	0.00
## 2	NA	NA
## 3	2.12	0.00

## 4	9.21	0.00
## 5	1.95	0.00
## 6	13.20	0.17
##	gene_expression_value.ABHD14B	gene_expression_value.ABHD15
## 1	18.53	6.82
## 2	NA	NA
## 3	7.63	2.41
## 4	28.23	8.12
## 5	6.49	2.26
## 6	36.75	8.38
##	gene_expression_value.ABHD16A	gene_expression_value.ABHD16B
## 1	77.48	0.29
## 2	NA	NA
## 3	77.73	0.08
## 4	56.77	0.05
## 5	63.66	0.04
## 6	57.29	0.18
##	gene_expression_value.ABHD17A	gene_expression_value.ABHD17B
## 1	13.79	8.78
## 2	NA	NA
## 3	6.88	5.59
## 4	12.28	9.89
## 5	4.11	6.64
## 6	12.89	11.82
##	gene_expression_value.ABHD17C	gene_expression_value.ABHD18
## 1	1.37	6.93
## 2	NA	NA
## 3	1.64	4.06
## 4	1.24	6.57
## 5	1.17	7.94
## 6	2.60	7.49
##	gene_expression_value.ABHD2	gene_expression_value.ABHD3
## 1	46.50	73.51
## 2	NA	NA
## 3	36.05	113.96
## 4	54.52	106.44
## 5	72.44	190.95
## 6	44.91	121.66
##	gene_expression_value.ABHD4	gene_expression_value.ABHD5
## 1	44.47	60.14
## 2	NA	NA
## 3	37.71	77.50
## 4	41.75	69.89
## 5	29.65	51.39
## 6	31.88	41.47
##	gene_expression_value.ABHD6	gene_expression_value.ABHD8
## 1	3.47	3.27
## 2	NA	NA
## 3	2.15	2.59
## 4	7.84	1.86
## 5	1.45	1.11
## 6	5.11	2.22
##	gene_expression_value.ABI1	gene_expression_value.ABI2
## 1	75.36	3.22

```
## 2          NA          NA
## 3        61.63        0.67
## 4        66.52        4.99
## 5        79.84        2.12
## 6        66.89        5.82
```

```
pivoted_final_df <- na.omit(pivoted_final_df)
```

```
library(pheatmap)
library(RColorBrewer)

set.seed(100)

genes.names <- c("gene_expression_value.A2ML1", "gene_expression_value.A3GALT2",
                 "gene_expression_value.A4GALT", "gene_expression_value.A4GNT",
                 "gene_expression_value.AAAS", "gene_expression_value.AACS",
                 "gene_expression_value.ABHD5", "gene_expression_value.ABHD6",
                 "gene_expression_value.ABHD8", "gene_expression_value.ABI1")
tengene.data <- pivoted_final_df[,genes.names]

head(tengene.data)
```

```
## gene_expression_value.A2ML1 gene_expression_value.A3GALT2
## 1          0.04          0.07
## 3          0.02          0.00
## 4          0.07          0.00
## 5          0.05          0.07
## 6          0.03          0.07
## 7          0.01          0.00
## gene_expression_value.A4GALT gene_expression_value.A4GNT
## 1          0          0.03
## 3          0          0.07
## 4          0          0.00
## 5          0          0.00
## 6          0          0.00
## 7          0          0.00
## gene_expression_value.AAAS gene_expression_value.AACS
## 1        18.92          4.07
## 3        13.85          1.83
## 4        22.11          4.22
## 5         8.45          1.17
## 6        28.59          4.24
## 7        10.50          2.10
## gene_expression_value.ABHD5 gene_expression_value.ABHD6
## 1        60.14          3.47
## 3        77.50          2.15
## 4        69.89          7.84
## 5        51.39          1.45
## 6        41.47          5.11
## 7       135.01          5.93
## gene_expression_value.ABHD8 gene_expression_value.ABI1
## 1         3.27        75.36
## 3         2.59        61.63
```

## 4	1.86	66.52
## 5	1.11	79.84
## 6	2.22	66.89
## 7	0.84	64.55

```
heatmap.data <- pivoted_final_df[c("icu_status","disease_status","gene_expression_value.A2ML1","gene_ex
    "gene_expression_value.A4GALT","gene_expression_value.A4GNT",
    "gene_expression_value.AAAS" ,"gene_expression_value.AACS",
    "gene_expression_value.ABHD5","gene_expression_value.ABHD6",
    "gene_expression_value.ABHD8","gene_expression_value.ABI1")]

colnames(heatmap.data) <- c("ICU Status","Disease Status","A2ML1","A3GALT2",
    "A4GALT","A4GNT",
    "AAAS" ,"AACS",
    "ABHD5","ABHD6",
    "ABHD8","ABI1")
ann.data <- heatmap.data[c("ICU Status", "Disease Status")]

brewercolors <- brewer.pal(8, "Pastel2")

ann.colors <- list(
  "ICU Status" = c(' no' = 'coral',
    ' yes' = 'turquoise'),
  "Disease Status" = c('COVID-19' = 'pink',
    'non-COVID-19' = 'lightgreen')
)

pheatmap(
  as.matrix(scale(heatmap.data[,3:12])),
  cluster_rows = TRUE,
  cluster_cols = TRUE,
  show_rownames = FALSE,
  fontsize = 8,
  main = "Heat map with 10 Genes",
  annotation_row = ann.data,
  annotation_colors = ann.colors,
  color = brewercolors
)
```

