Final project

2025-07-22

R Markdown

Identify one gene, one continuous covariate, and two categorical covariates in the provided dataset. Generate the following three plots using ggplot2 for your covariates of choice: charlson score, disease status, sex

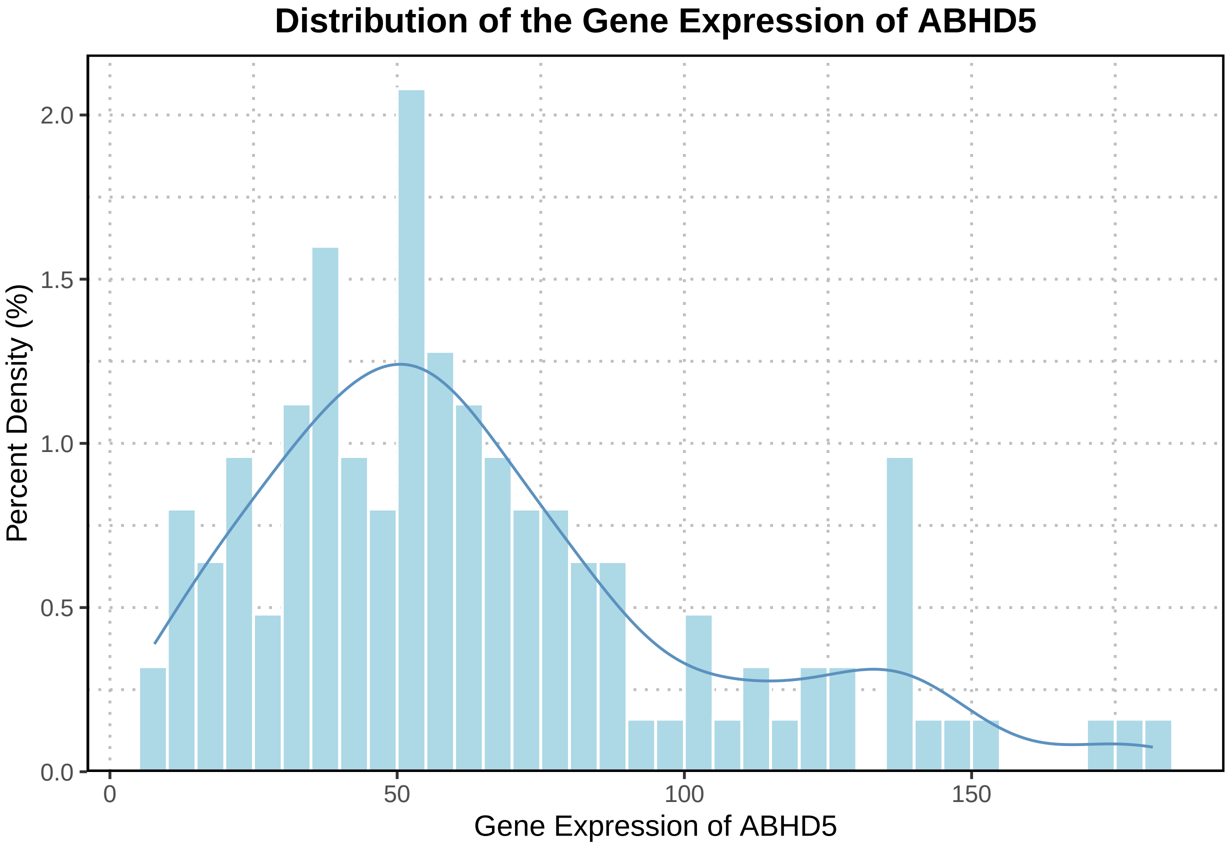
combine gene expression data and demographic data

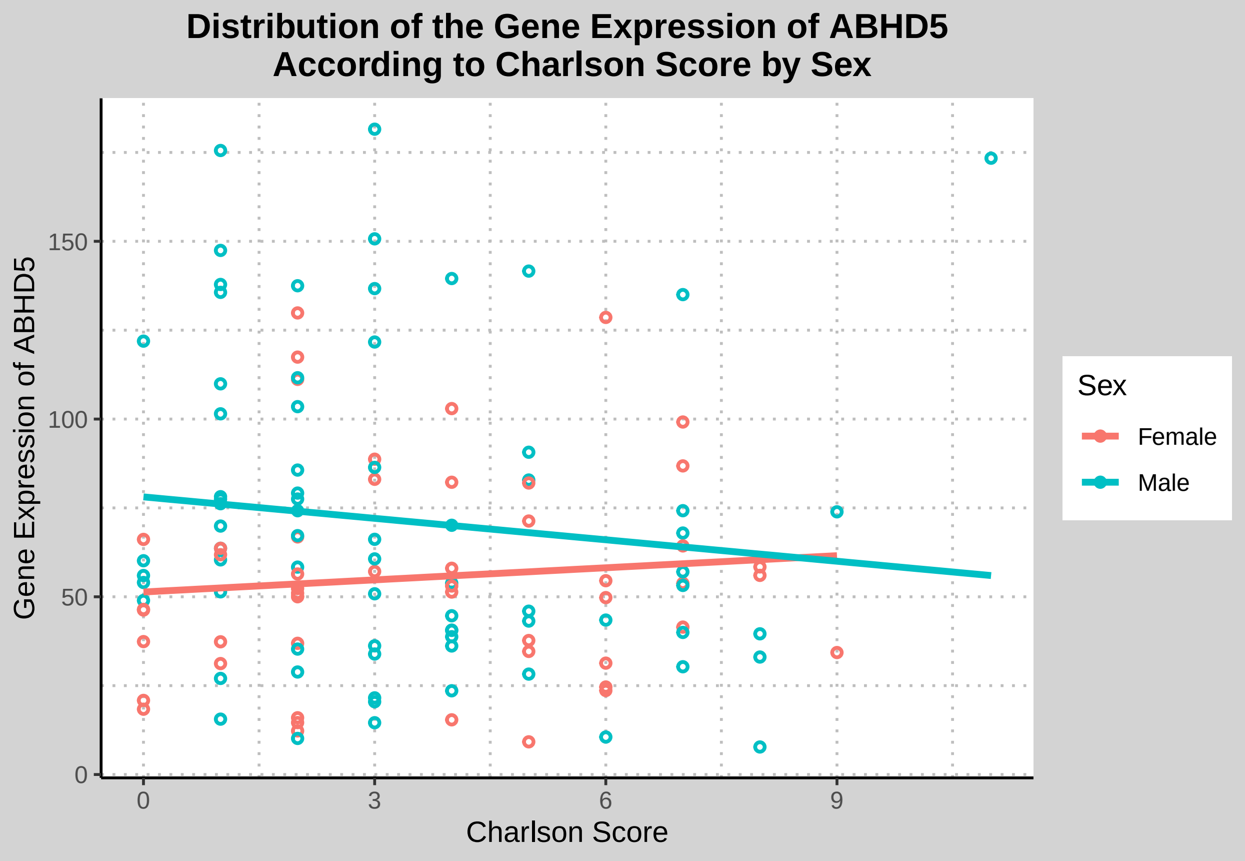
Function for three plots

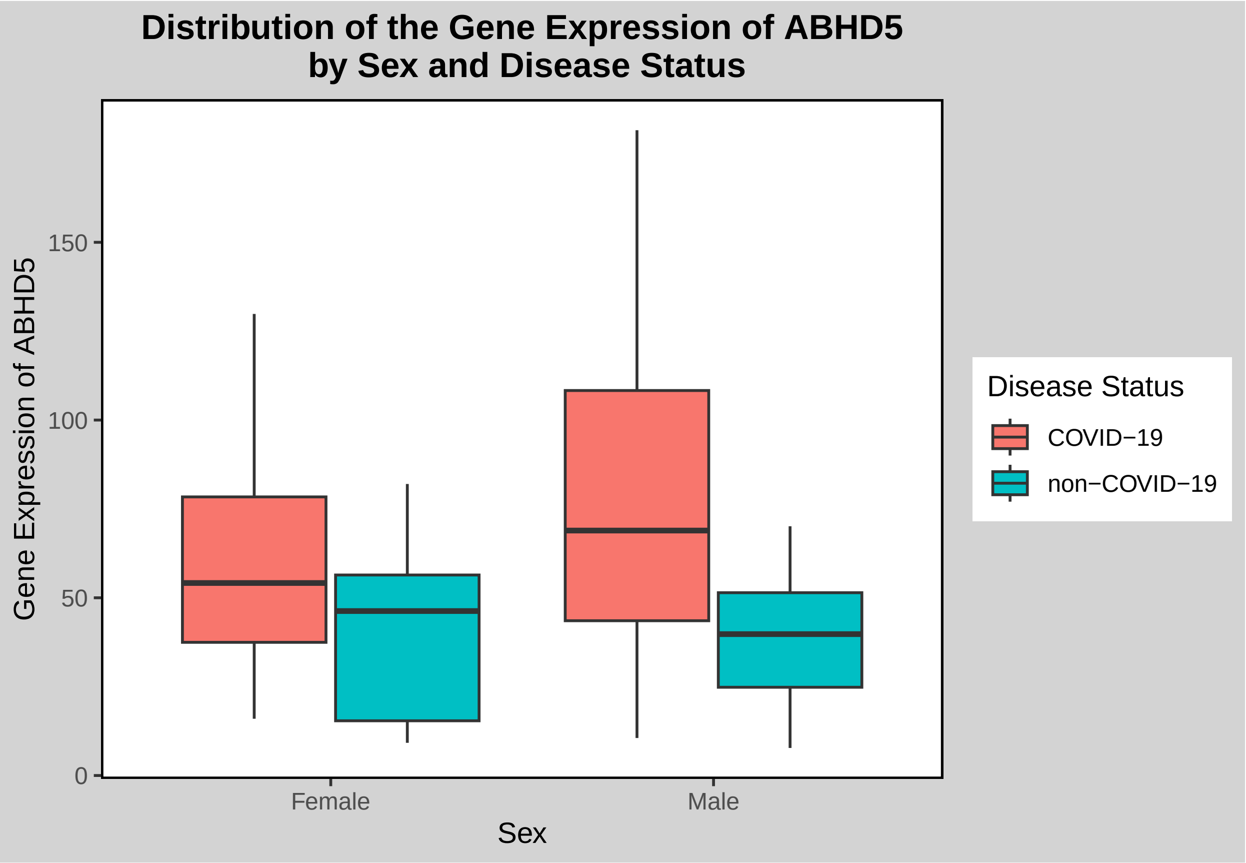
Your functions should take the following input: (1) the name of the data frame, (2) a list of 1 or more gene names, (3) 1 continuous covariate, and (4) two categorical covariates (10 pts) Select 2 additional genes (for a total of 3 genes) to look at and implement a loop to generate your figures using the function you created (10 pts)

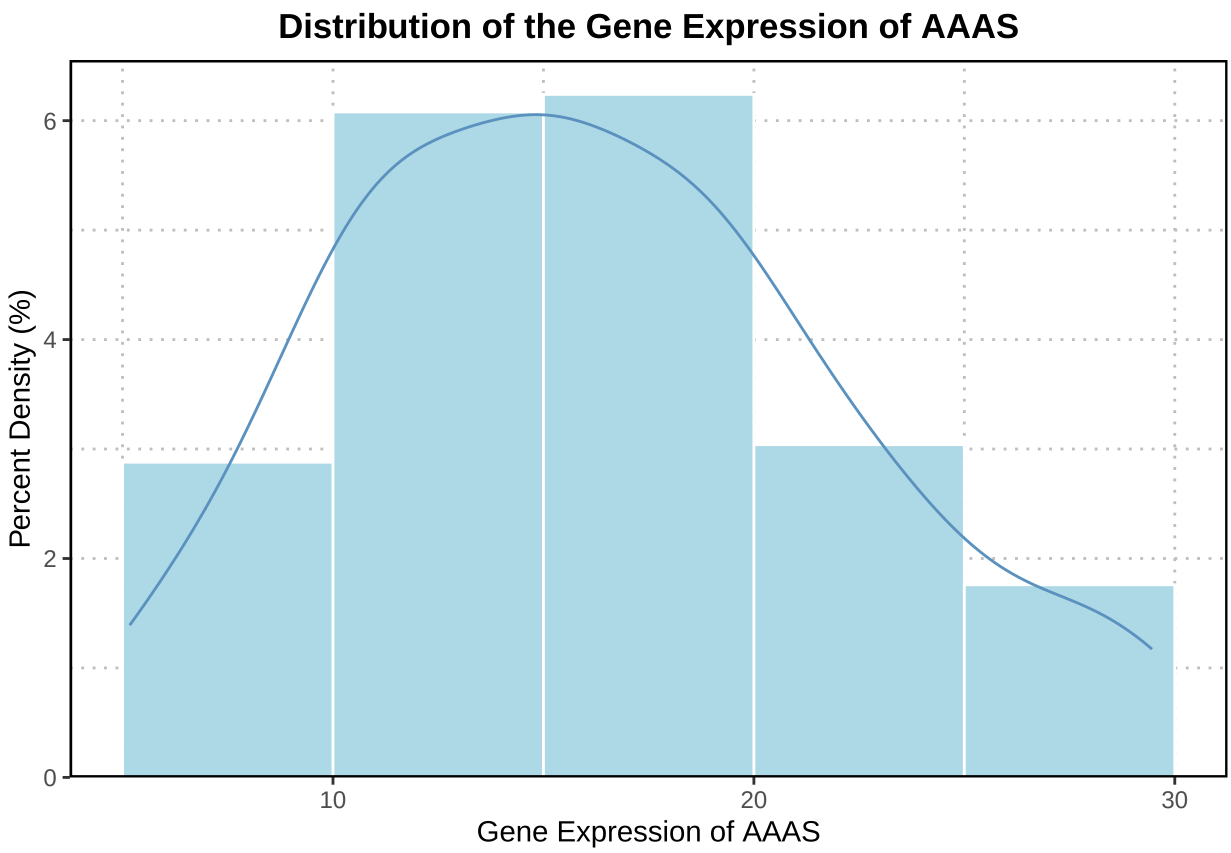
```r
library(ggplot2) #to load ggplot
library(ggpubr) #to load ggarrange

plot_mx <- function(dat, gene, cont, cat1, cat2){

 h <- ggplot(dat,
             aes(x = .data[[gene]])) +
    geom_histogram(aes(y = after_stat(density)*100),
                 binwidth = 5,
                 fill = "lightblue",
                 color = "white",
                 boundary = 0) +
  geom_smooth(aes(y = after_stat(density)*100),
              stat = "density",
              color = "#5B91BE" ,
              linewidth = 0.5) +
  labs(title = paste0("Distribution of the Gene Expression of ", gene),
       x = paste0("Gene Expression of ", gene),
       y = "Percent Density (%)") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05))) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    panel.grid.major = element_line(color = "gray",
                                    linewidth = 0.5,
                                    linetype = "dotted"),
    panel.grid.minor = element_line(color = "gray",
                                    linewidth = 0.5,
                                    linetype = "dotted"),
    panel.border = element_rect(color = "black", fill = NA, linewidth = 0.8),
    panel.background = element_rect(fill = "white"),
    axis.line = element_blank()
  )

 plot(h)


  s <- ggplot(dat,
              aes(x = .data[[cont]],
                  y = .data[[gene]],
                  color = .data[[cat1]])) +
    geom_point(shape = 1, size = 1, stroke = 1)+
    geom_smooth(method = "lm", linewidth = 1.2, fill = NA)+
  labs(title = paste0("Distribution of the Gene Expression of ",
                      gene, "\n According to ", cont, " by ", cat1),
       x = paste0(cont),
       y = paste0("Gene Expression of ", gene),
       color = cat1) +
  theme(
    plot.title = element_text(hjust = 0.5, face="bold"),
    panel.background = element_rect(fill = "white"),
    plot.background = element_rect(fill = "lightgray"),
    panel.grid.major = element_line(color = "gray",
                                    linewidth = 0.5,
                                    linetype = "dotted"),
    panel.grid.minor = element_line(color = "gray",
                                    linewidth = 0.5,
                                    linetype = "dotted"),
    axis.line = element_line(color = "black"),
    legend.position = "right")

  plot(s)


b <- ggplot(dat,
            aes(x = .data[[cat1]],
                y = .data[[gene]],
                fill = .data[[cat2]])) +
  geom_boxplot(position = position_dodge(width = 0.8)) +
  labs(title = paste0("Distribution of the Gene Expression of ",
                      gene, "\n by ", cat1, " and ", cat2),
       x = cat1,
       y = paste0("Gene Expression of ", gene),
       fill = cat2) +
  theme(
    plot.title = element_text(hjust = 0.5, face="bold"),
    panel.background = element_rect(fill = "white"),
    plot.background = element_rect(fill = "lightgray"),
    panel.border = element_rect(color = "black",
                                fill = NA,
                                linewidth = 0.8),
    panel.grid = element_blank(),
    axis.line = element_blank(),
    legend.position = "right")
```

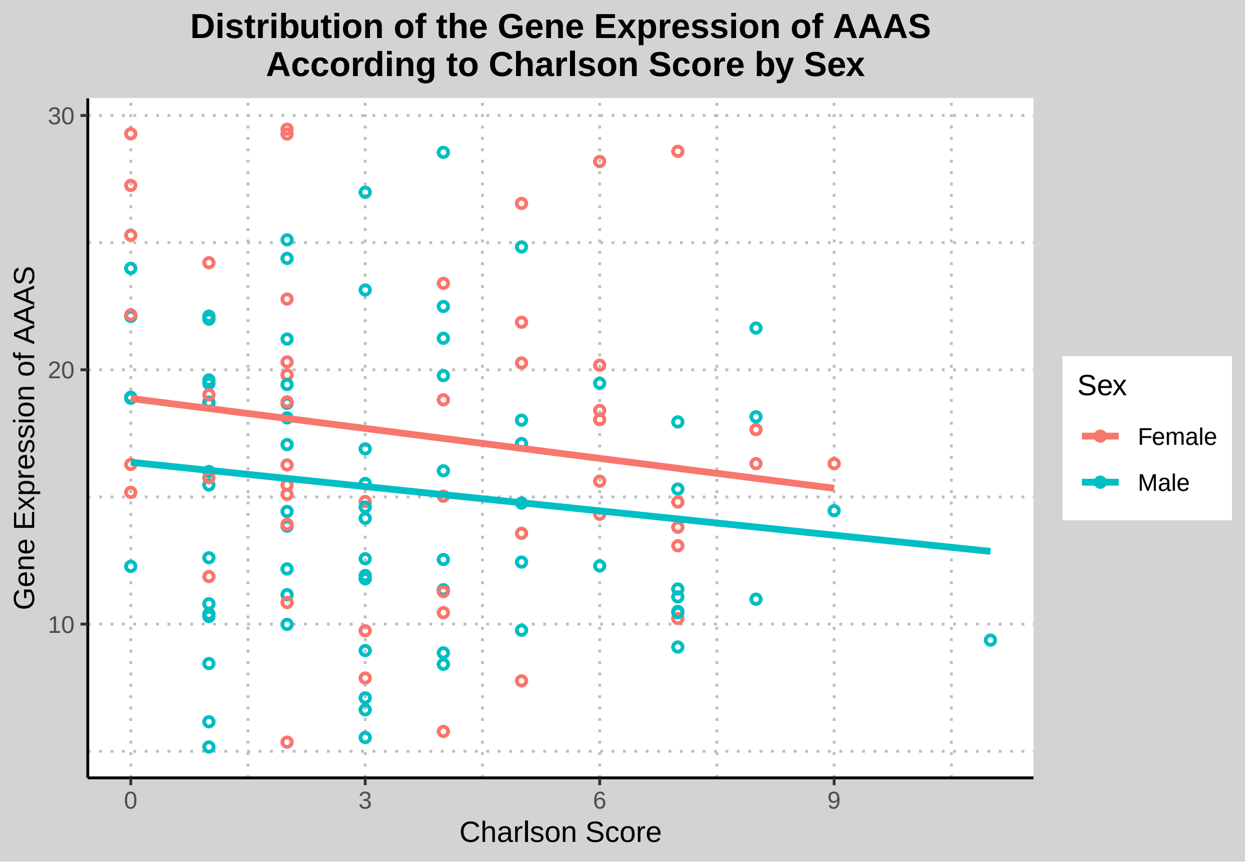
```
plot(b)

}
```

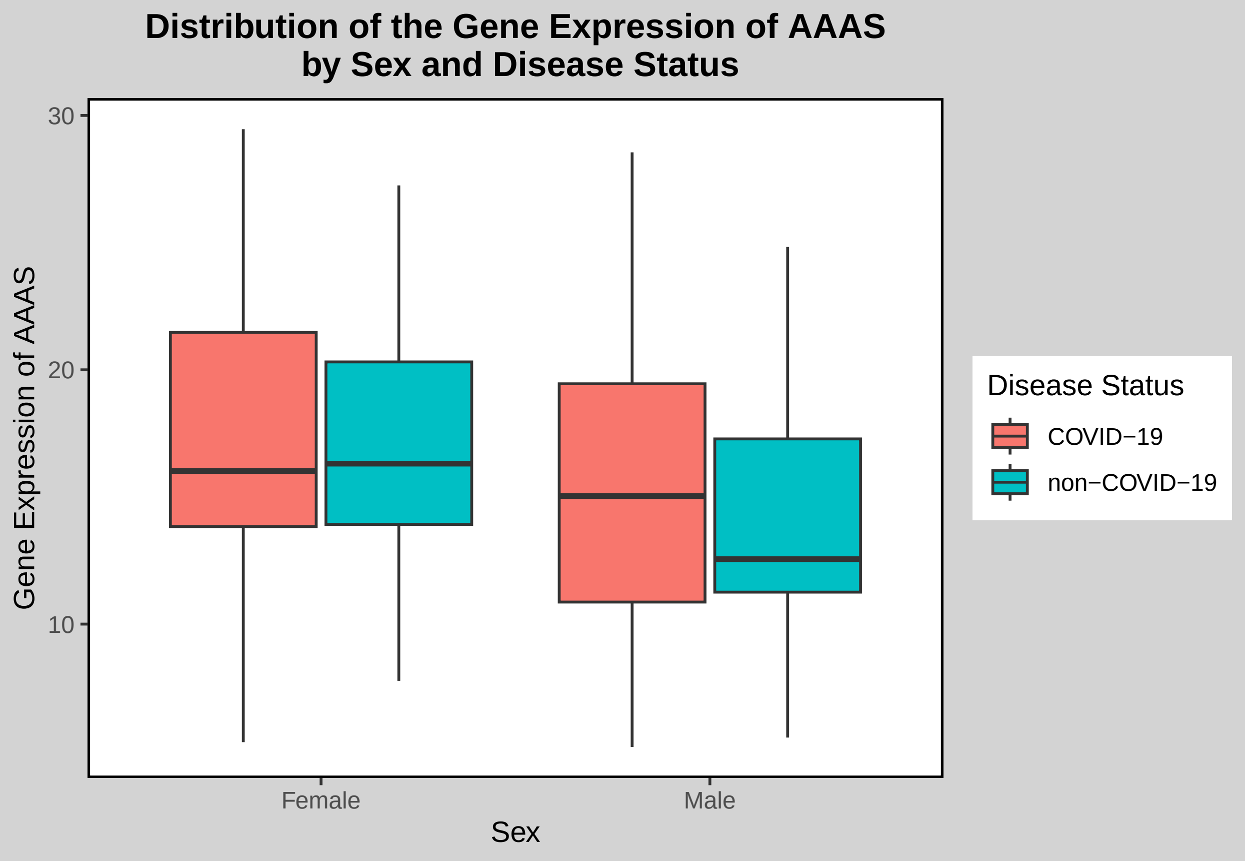

Distribution of the Gene Expression of ABHD5
Percent Density (%)
2.0
1.5
1.0
0.5
0.0
0
50
100
150
Gene Expression of ABHD5

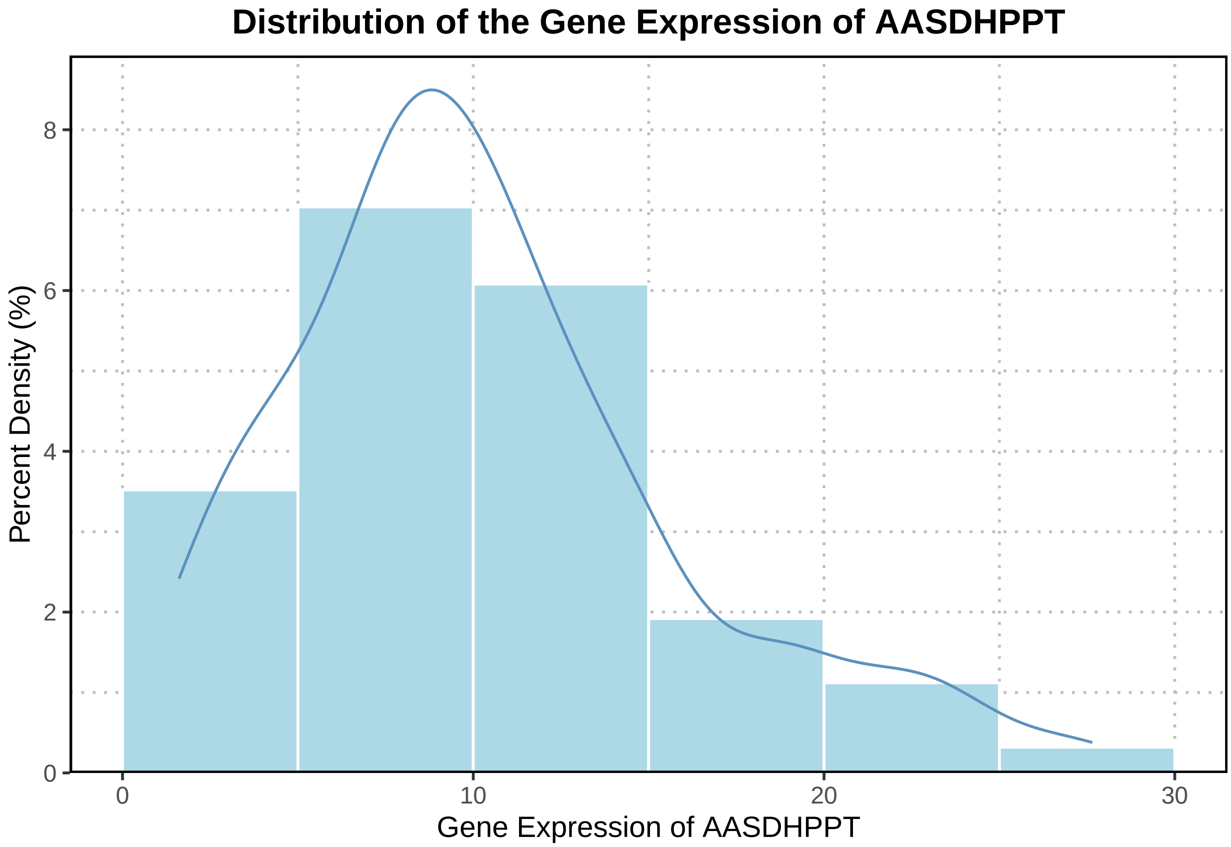

Distribution of the Gene Expression of ABHD5
According to Charlson Score by Sex
Gene Expression of ABHD5
0
50
100
150
Charlson Score
0
3
6
9
Sex
Female
Male

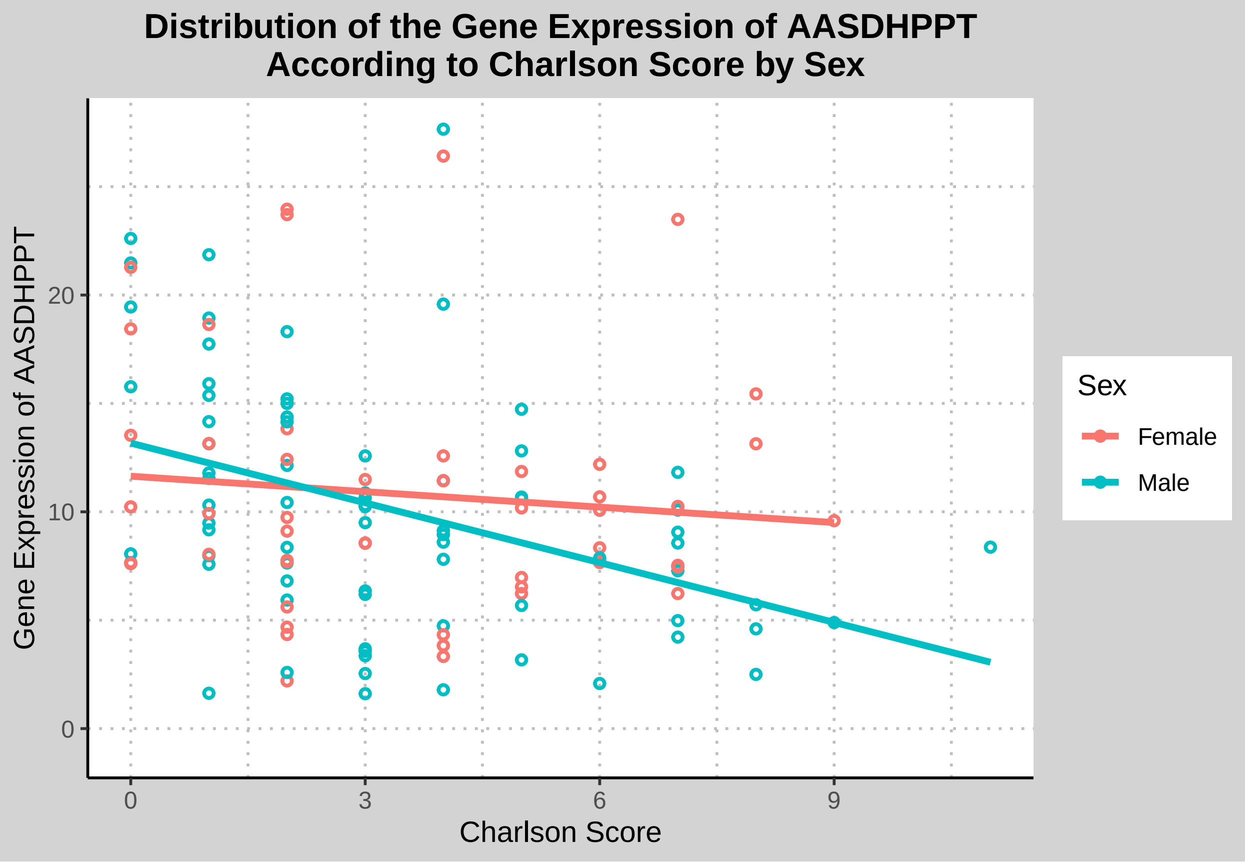

Distribution of the Gene Expression of ABHD5
by Sex and Disease Status
Gene Expression of ABHD5
0
50
100
150
Female
Male
Sex
Disease Status
COVID−19
non−COVID−19

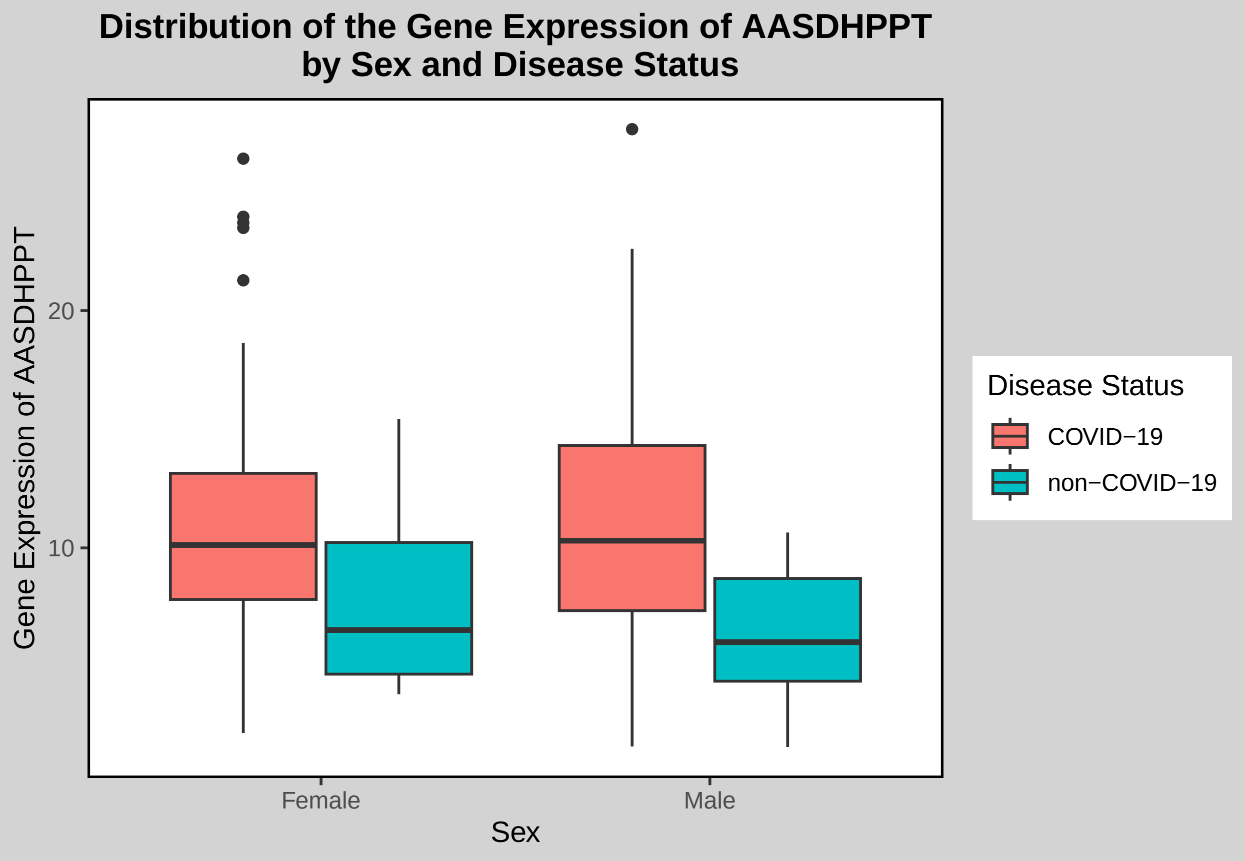

Distribution of the Gene Expression of AAAS
Percent Density (%)
0
2
4
6
10
20
30
Gene Expression of AAAS


Distribution of the Gene Expression of AAAS
According to Charlson Score by Sex
Gene Expression of AAAS
30
20
10
0
3
6
9
Charlson Score
Sex
Female
Male


Distribution of the Gene Expression of AAAS
by Sex and Disease Status
Gene Expression of AAAS
30
20
10
Female
Male
Sex
Disease Status
COVID−19
non−COVID−19


Distribution of the Gene Expression of AASDHPPT
Percent Density (%)
0
2
4
6
8
0
10
20
30
Gene Expression of AASDHPPT


Distribution of the Gene Expression of AASDHPPT
According to Charlson Score by Sex
Gene Expression of AASDHPPT
0
10
20
0
3
6
9
Charlson Score
Sex
Female
Male


Distribution of the Gene Expression of AASDHPPT
by Sex and Disease Status
Gene Expression of AASDHPPT
20
10
Female
Male
Sex
Disease Status
COVID−19
non−COVID−19