

JFAITH_QBS103_final_project

2025-07-12

Comparing Biomarkers to Covid-19

Data Wrangling

Read in files

```
gene_data <- read.csv("data/QBS103_GSE157103_genes.csv")
meta_data <- read.csv('data/QBS103_GSE157103_series_matrix-1.csv')
```

Function To Create Gene Specific Dataframe

```
#gene_name = 'ABCF1'
process_dataframe <- function(gene_name) {
  # Select gene
  gene <- gene_data[which(gene_data$X==gene_name), ]

  # Get all column names except X
  column_names <- names(gene)[-1]

  #Melt dataframe
  #referenced: https://tidyr.tidyverse.org/reference/pivot_longer.html
  gene_long <- pivot_longer(
    gene,
    cols = all_of(column_names),
    names_to = "participant_id",
    values_to = "expression")

  gene_dataset <- inner_join(gene_long, meta_data, by="participant_id")
  # referenced: https://datascienceplus.com/merging-datasets-with-tidyverse/

  gene_dataset$HospitalDaysGroup <- cut(
    gene_dataset$hospital.free_days_post_45_day_followup,
    breaks = c(0, 10, 20, 30, 40, 50),
    labels = c('Under 10', '10-20', '20-30', '30-40', 'Over 40'),
    right = FALSE
  )
  return (gene_dataset)
}
```

Visualizations

```
#Custom Color Palette
colorPalette <- c('#004777', '#F05D5E', '#A8D0DB', '#136F63', '#FFC857', '#9DD9D2')

myTheme <- theme(
  panel.border = element_blank(),
  panel.grid.major = element_line(colour="grey", linewidth = rel(.1)),
  panel.grid.minor= element_line(colour="grey", linewidth = rel(.1)),
  #panel.grid.minor = element_blank(),
  # Define my axis
  plot.title = element_text(colour = "black", hjust = .5, face='bold'),
  axis.line = element_line(colour = '#004777', linewidth = rel(1)),
  axis.title.x = element_text(color='black'),
  #axis.title.y = element_text(color='black'),
  axis.text = element_text(color='#004777'),
  # Set plot background
  panel.background = element_blank()
)
```

Build a function to create the plots you made for Presentation 1, incorporating any feedback you received on your submission. 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)

```
## documentation on working with .data[[]]
##https://dplyr.tidyverse.org/articles/programming.html

create_graphs <- function(
  dataframe_name,
  gene_name,
  continuous_covariate,
  categorical_covariates
) {

  #https://stackoverflow.com/questions/69338331/r-markdown-create-section-headers-in-r-and-print-r-outp
  cat("\n\nnewpage\n")
  cat(paste0("# Visuals for ", gene_name, "\n\n"))

  histogram_title <- paste(gene_name, "x Count of Participants")
  histogram <- ggplot(dataframe_name, aes(x = expression)) +
    geom_histogram(binwidth=1, color="white", fill='#004777') +
    labs(title=histogram_title, x = 'Gene Expression', y = 'Count') +
    myTheme

  print(histogram)

  cleaned_continuous_covariate <- gsub("[.]", " ", continuous_covariate)
  scatterplot_title <- paste("Gene Expression vs", cleaned_continuous_covariate)

  scatterplot <- ggplot(dataframe_name, aes(x = expression, y = .data[[continuous_covariate]])) +
    geom_point() +
    labs(title=scatterplot_title, x = 'Gene Expression', y = cleaned_continuous_covariate) +
    scale_color_manual(values = colorPalette)+
}
```

```

myTheme
print(scatterplot)

boxplot_title <- paste(gene_name, 'vs', categorical_covariates[1])
boxplot <- ggplot(
  dataframe_name,
  aes(
    x = .data[[categorical_covariates[1]]],
    y = expression,
    color = .data[[categorical_covariates[2]]]
  )) +
  geom_boxplot() +
  labs(
    title = boxplot_title,
    x = categorical_covariates[1],
    y = 'Gene Expression'
  ) +
  scale_color_manual(values = colorPalette) +
  myTheme
print(boxplot)
}

```

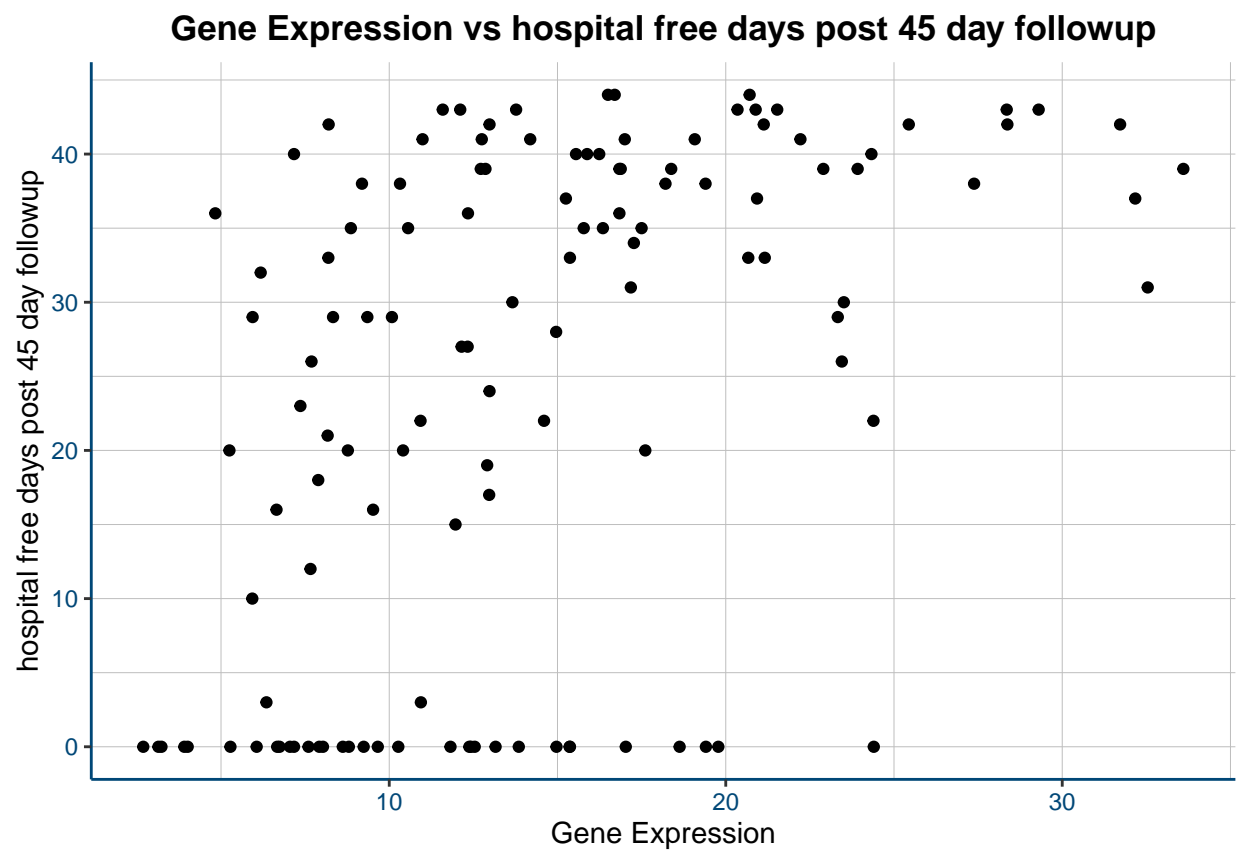
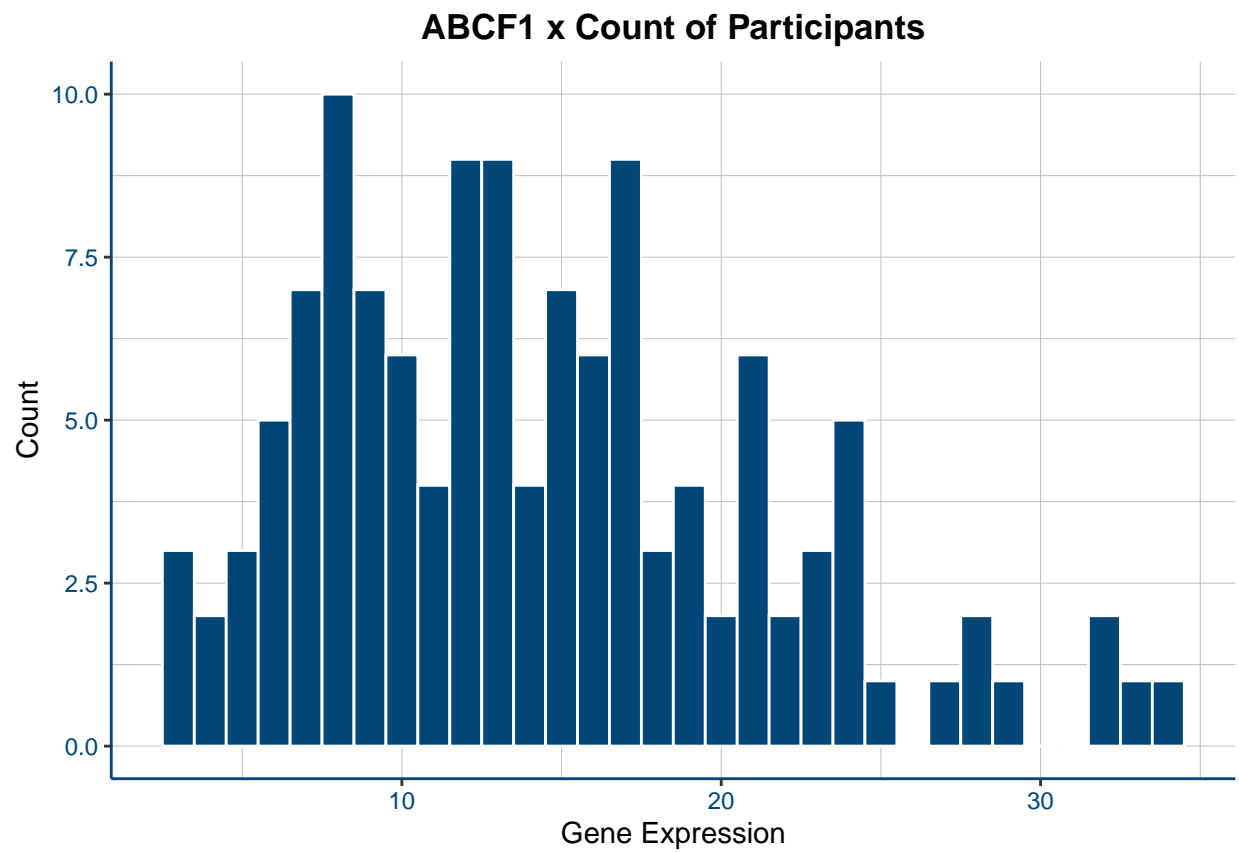
Displaying All Graphs

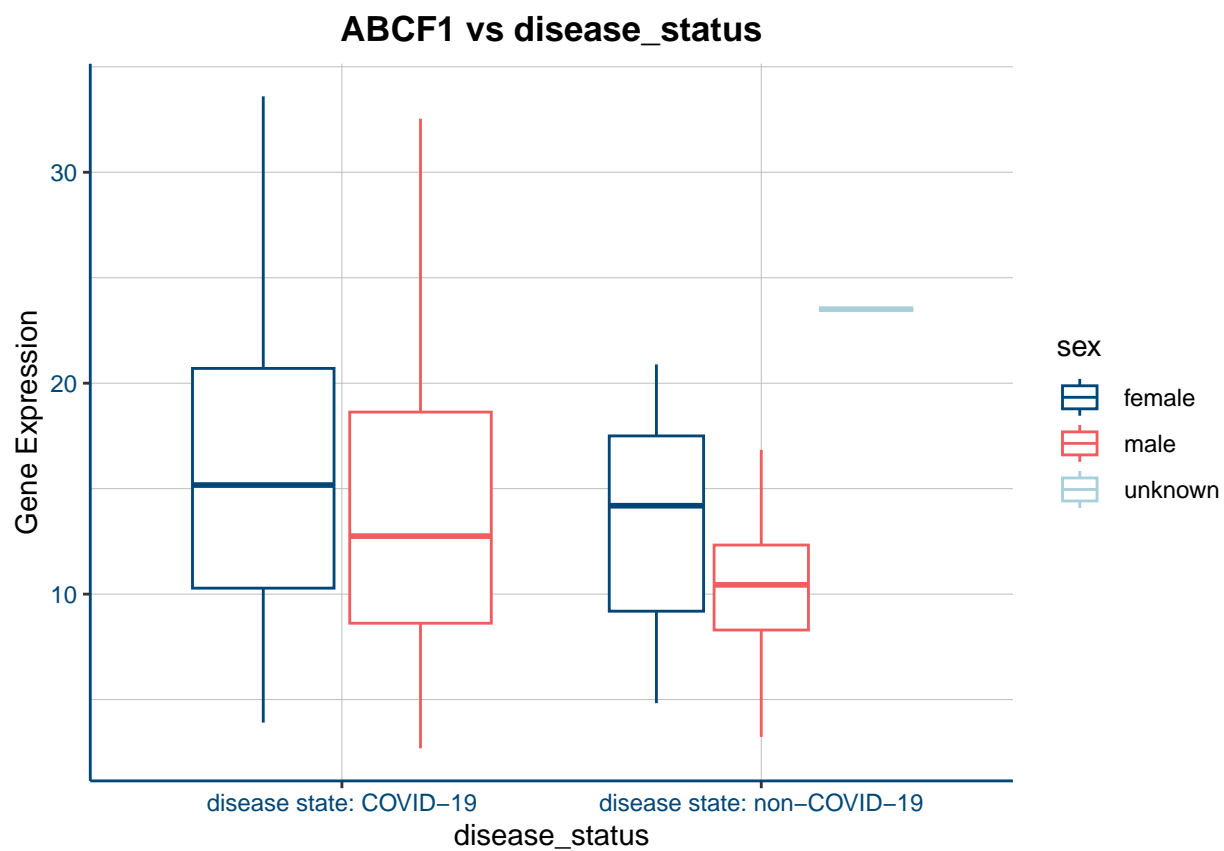
```

for (gene in c('ABCF1', 'AAGAB', 'ABAT')){
  gene_df <- process_dataframe(gene)
  create_graphs(gene_df, gene, 'hospital.free_days_post_45_day_followup', c('disease_status', 'sex'))
}

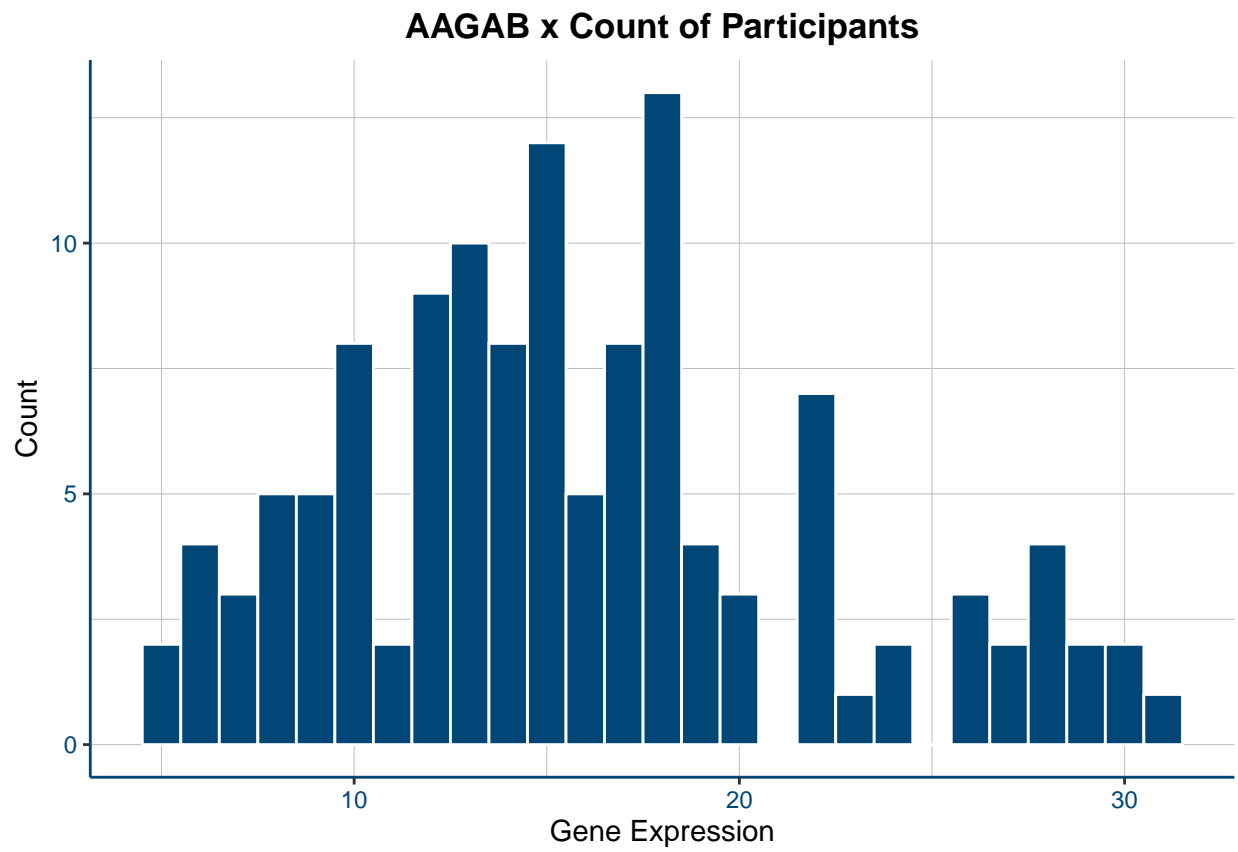
```

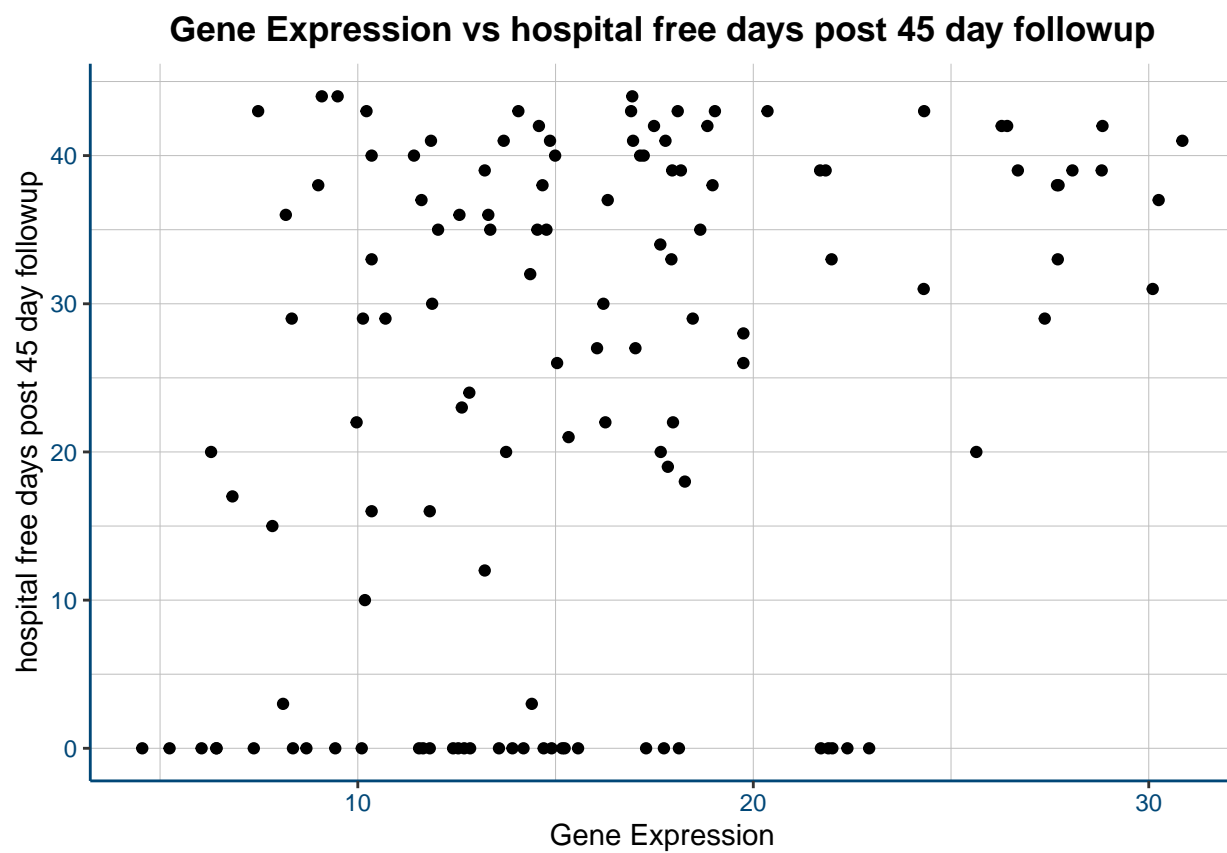
Visuals for ABCF1

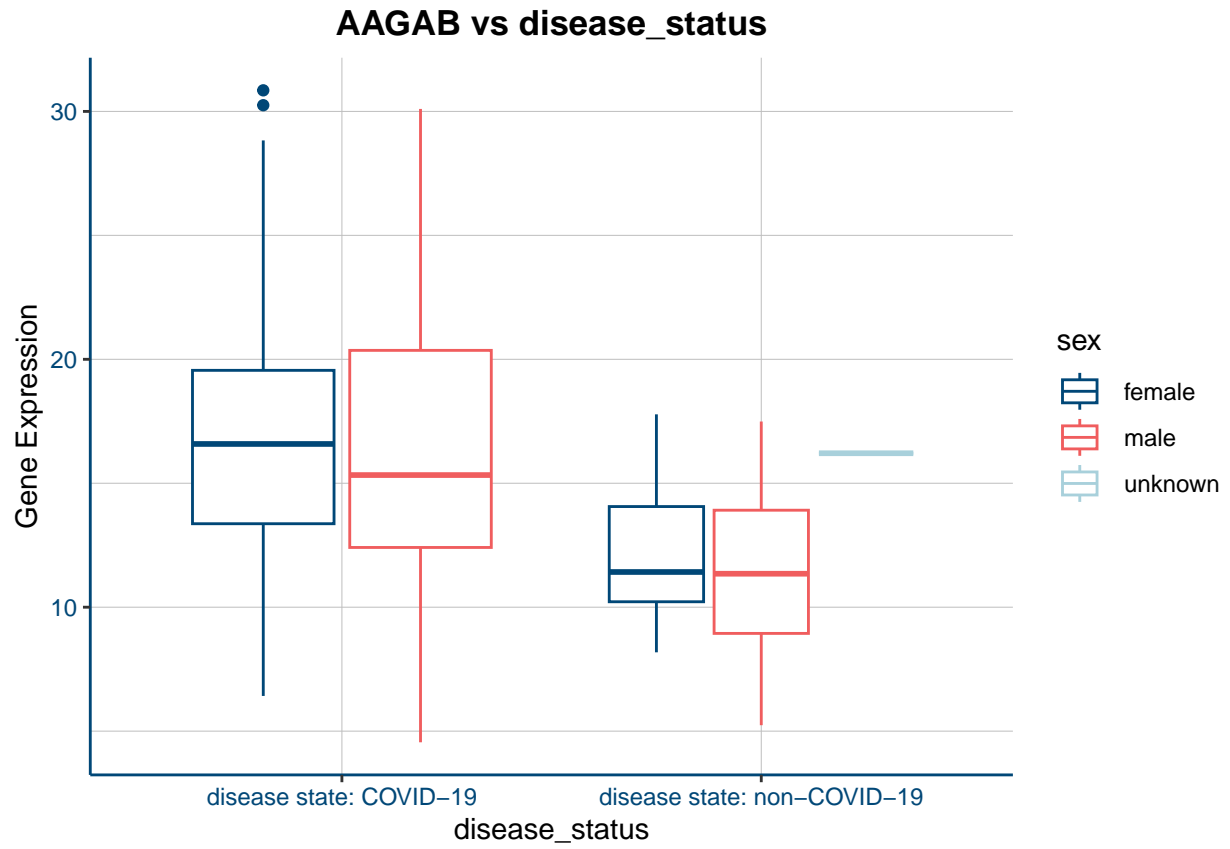




Visuals for AAGAB







Visuals for ABAT

