Final Project Submission 1

Ryan Pontius

2025-07-14

Histogram

```r
# In addition to the inclass lectures/assignments, I referenced the following
# Youtube video's to further my knowledge and understanding of how to use
# ggplot2 and tidyverse: https://www.youtube.com/watch?v=Bg4qxVNaDck
# https://www.youtube.com/watch?v=Rp6Xwy4BfxI In addition,
# https://chat.dartmouth.edu/ was an additional resource I used for parts of
# this assignment

library(tidyr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4      v purrr     1.0.4
## v forcats   1.0.0      v readr     2.1.5
## v ggplot2   3.5.2      v stringr   1.5.1
## v lubridate 1.9.4      v tibble    3.3.0
## -- 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 errors
```

```r
library(ggplot2)
library(colorspace)

# getwd()

df_genes <- read.csv("~/Desktop/QBS_103/Final_Project/GSE157103_genes.csv")
df_matrix <- read.csv("~/Desktop/QBS_103/Final_Project/GSE157103_series_matrix-1.csv")

# head(rownames(df_genes)) head(df_genes[,1])


base_green <- "#ADCCA3"   # chat.dartmout.edu taught me how to assign colors using colorspace.
pastel_olive <- darken(base_green, amount = 0.2)

colnames(df_genes)[1] <- "Genes"   # Names first column of df_genes 'Genes'
gene_ABCD3 <- filter(df_genes, df_genes["Genes"] == "ABCD3")   # Pulls the row for gene ABCD3 into a dat
```

```
## Warning: Using one column matrices in `filter()` was deprecated in dplyr 1.1.0.
```

```
## i Please use one dimensional logical vectors instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```r
hist_values <- as.numeric(gene_ABCD3[, -1])  # I asked chatgpt for help here. I could not figure out ho

hist(hist_values, breaks = 15, main = "ABCD3 Expression", xlab = "Expression Values",
    ylim = c(0, 20), xlim = c(0, 20), col = pastel_olive)
```
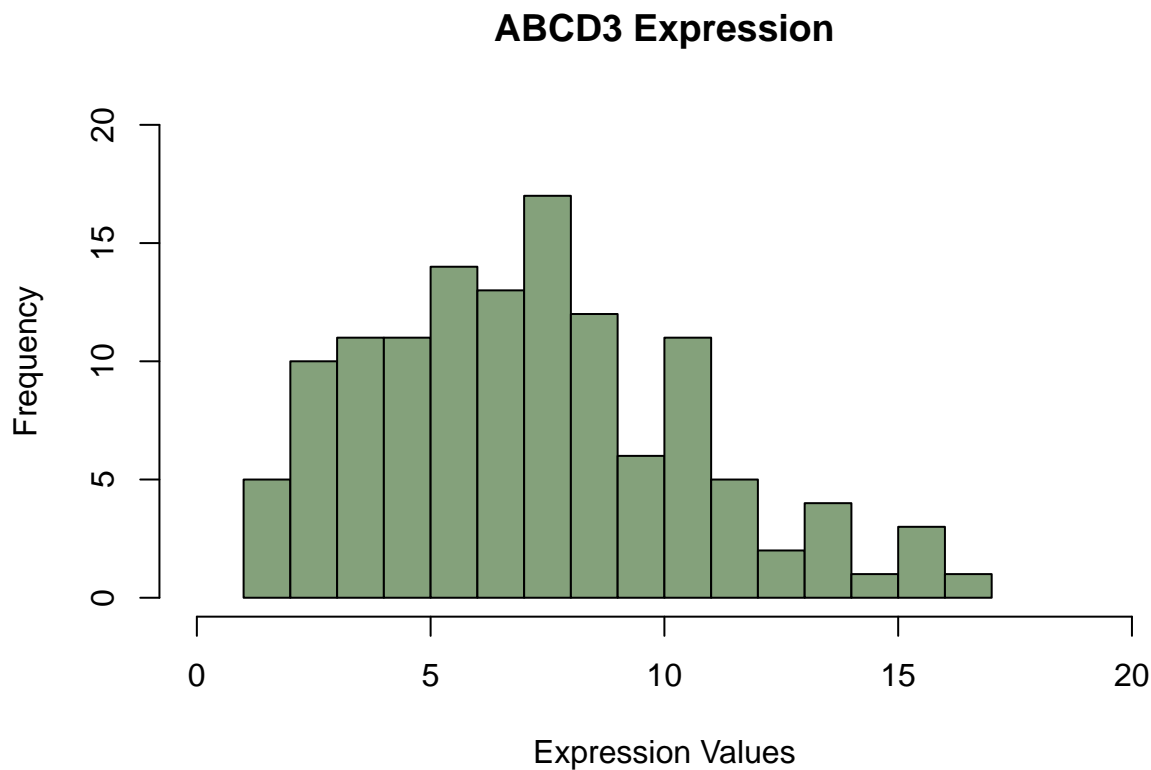
ABCD3 Expression

Scatter Plot

```r
samples <- colnames(df_genes)[-1]  # Pull column names into a data frame.
df_ABCD3 <- data.frame(participant_id = samples, Expression = hist_values)  # Initialize new data frame
# I used chat.dartmout.edu for help on merging & flipping row to column for the
# data frames. It suggested the merge() function and explained how to approach
# it.
merged_df <- merge(df_ABCD3, df_matrix, by = "participant_id")  # Merge Data frames by 'participant_id'
# head(merged_df)

## Scatter Plot
ggplot(merged_df, aes(Expression, charlson_score)) + geom_point() + theme_classic() +
    labs(title = "ABCD3 Gene Expression vs. Charlson Score", x = "Expression", y = "Charlson Score")
```

ABCD3 Gene Expression vs. Charlson Score
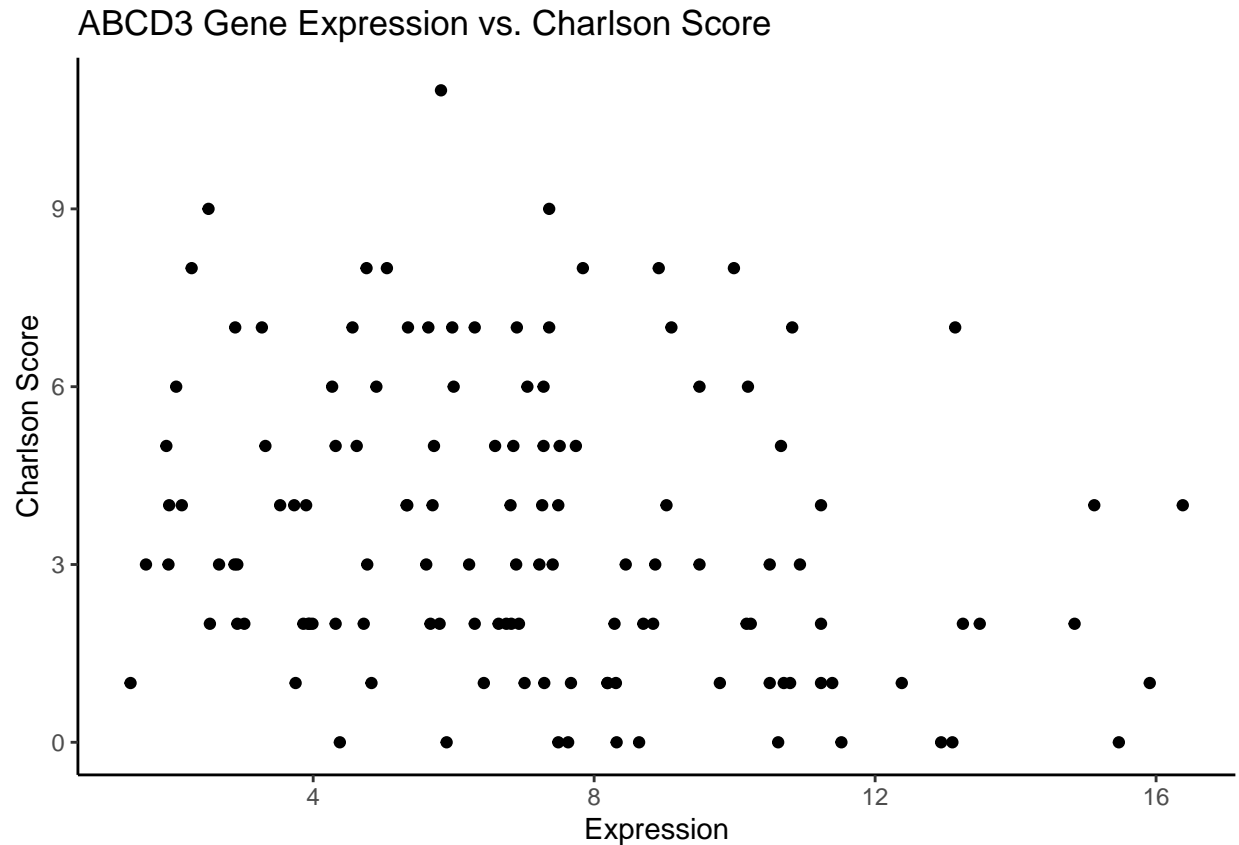
Box Plots

```
# I used chat.dartmout.edu to help explain to me what was meant by 'boxplot of
# gene expression separated by both categorical covariates' in the rubric, and
# it told me to approach it like I would a normal boxplot but use the 'fill'
# function for the second variable, which is what I did.  chat.dartmout.edu
# told me to use the 'fill' feature for sex variable and it told me how to
# adjust the width of my box plots, and also how to center my title.
ggplot(merged_df, aes(x = icu_status, y = Expression, fill = sex)) + geom_boxplot(width = 0.8) +
    theme_classic() + labs(title = "ABCD3 Gene Expression by Biological Sex and ICU Status",
    x = "ICU Status", y = "Gene Expression")
```

ABCD3 Gene Expression by Biological Sex and ICU Status