

Submission #1

Eric Wang

2024-07-27

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- 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
```

```
df1 <- read.csv("/Users/wangdeyao/Downloads/QBS103/QBS103_GSE157103_genes.csv")
df2 <- read.csv("/Users/wangdeyao/Downloads/QBS103/QBS103_GSE157103_series_matrix.csv")

data_long <- df1 %>%
  filter(X == "ABCA1")%>%
  gather(key = "Sample", value = "Expression",
         COVID_01_39y_male_NonICU:NONCOVID_26_36y_male_ICU)

new_merged <- merge(data_long, df2, by.x = "Sample", by.y = "participant_id")%>%
  select("Sample", "Expression", "age", "sex", "icu_status")

head(new_merged)
```

```
##           Sample Expression age    sex icu_status
## 1 COVID_01_39y_male_NonICU   32.30 39   male       no
## 2 COVID_02_63y_male_NonICU   15.84 63   male       no
## 3 COVID_03_33y_male_NonICU   34.38 33   male       no
## 4 COVID_04_49y_male_NonICU   14.24 49   male       no
## 5 COVID_05_49y_male_NonICU   18.39 49   male       no
## 6 COVID_07_38y_female_NonICU  14.66 38 female       no
```

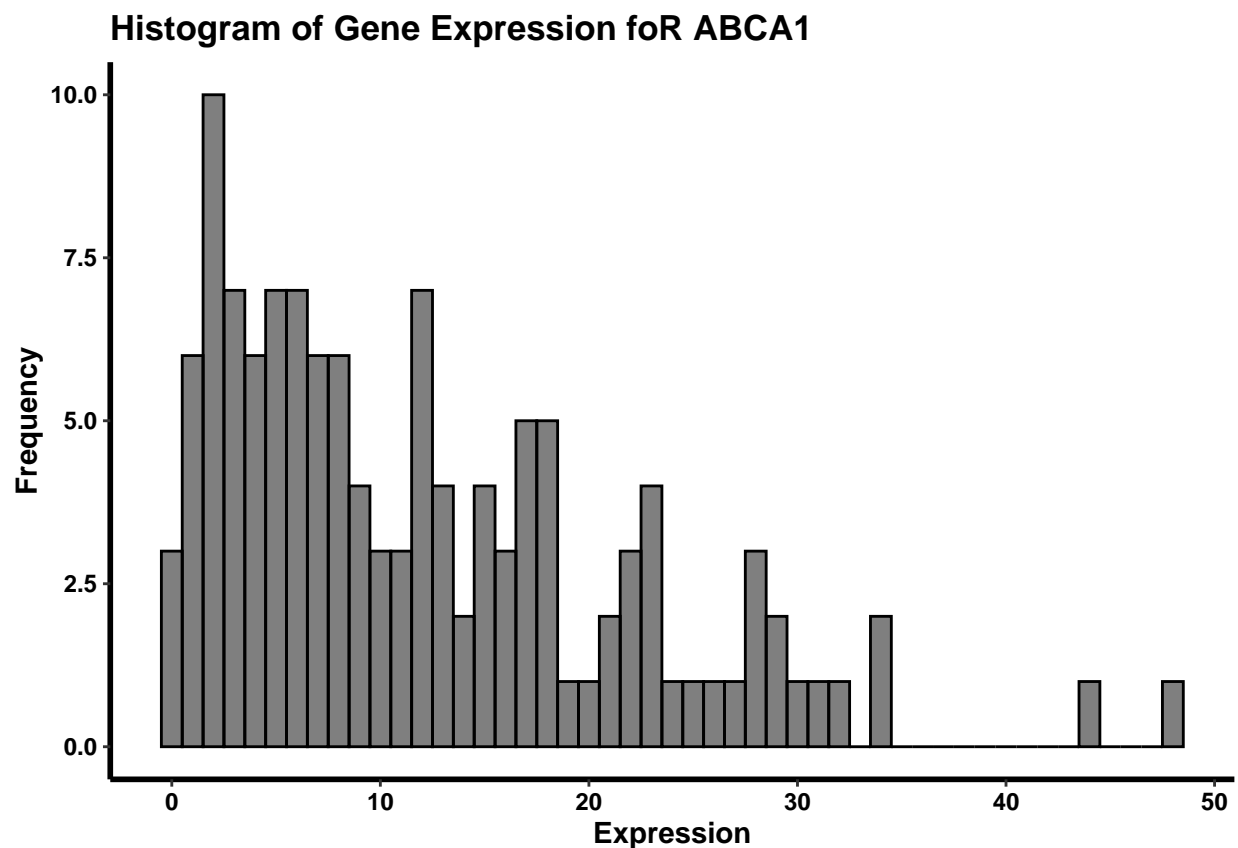
```
newTheme <- theme(
  panel.border = element_blank(), panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(),
  axis.line = element_line(colour = "black", linewidth = rel(2)),
  plot.background = element_rect(fill = "white"),
  panel.background = element_blank(),
```

```

legend.key = element_rect(fill = 'grey',color = "grey"),
legend.background = element_rect(fill = 'grey'),
legend.box.background = element_rect(color = "black"),
text = element_text(face="bold",colour = "black"),
axis.text = element_text(face="bold",colour = "black"),
  legend.position = 'bottom')

ggplot(new_merged, aes(x = Expression)) +
  geom_histogram(binwidth = 1, fill = "black", color = "black", alpha = 0.5) +
  labs(title = "Histogram of Gene Expression for ABCA1",
       x = "Expression",
       y = "Frequency") +
  newTheme

```



```
new_merged$age <- as.numeric(new_merged$age)
```

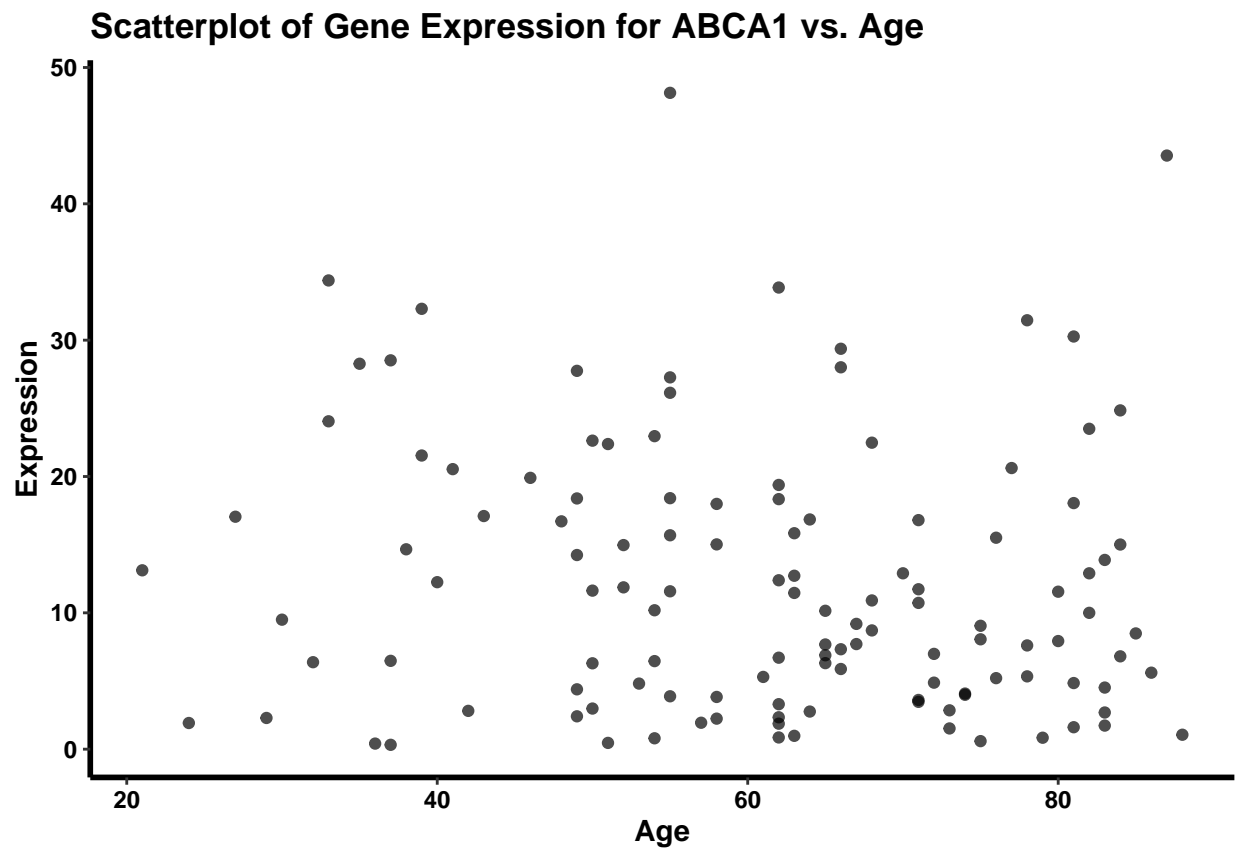
Warning: NAs introduced by coercion

```

ggplot(new_merged, aes(x = age, y = Expression)) +
  geom_point(color = "black", alpha = 0.7) +
  scale_y_continuous(breaks = seq(0, 100, by = 10)) +
  labs(title = "Scatterplot of Gene Expression for ABCA1 vs. Age",
       y = "Expression", x = "Age") +
  newTheme

```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```



```
colorpalette <- c("red", "blue", "yellow")
ggplot(new_merged, aes(x = sex, y = Expression, fill = icu_status))+
  geom_boxplot(alpha = 0.7)+
  labs(title = "Boxplot of Gene Expression for ABCA1 by Sex and ICU Status",
       x = "Sex", y = "Expression")+
  scale_fill_manual(values = colorpalette)+
  newTheme
```

