Submisson1

Ethan Zhang

2024-07-27

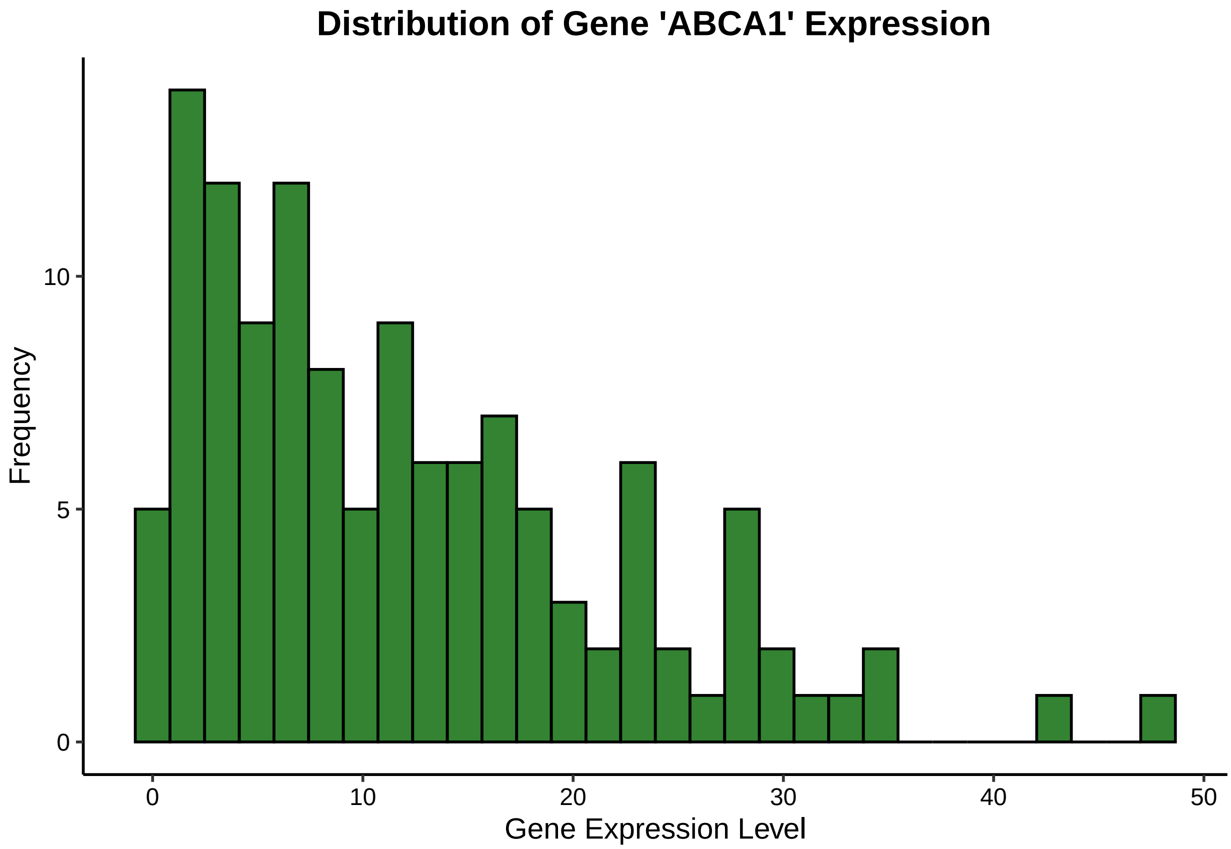
R Markdown

```
new_gene <- read.csv("/Users/zhyihan/Documents/Dartmouth Life/courses/QBS 103-Data Science/final projec
ser_m <-read.csv("/Users/zhyihan/Documents/Dartmouth Life/courses/QBS 103-Data Science/final project/QBS

#### I select "ABCA1" gene, and then convert wide to long
new_gene_long <- new_gene %>% filter(X == 'ABCA1') %>% gather(key = participant_id, value = expression,

#### I select "charlson_score" as continuous covariate, "sex" and "icu status" as categorical covariate
new_ser_m <- ser_m %>% select(participant_id, sex, icu_status,age)

#### merge two two tables with selected variables
new_data <- merge(new_ser_m,new_gene_long, by = "participant_id")
head(new_data)
```

```
##                participant_id    sex icu_status age     X expression
## 1   COVID_01_39y_male_NonICU    male         no  39 ABCA1      32.30
## 2   COVID_02_63y_male_NonICU    male         no  63 ABCA1      15.84
## 3   COVID_03_33y_male_NonICU    male         no  33 ABCA1      34.38
## 4   COVID_04_49y_male_NonICU    male         no  49 ABCA1      14.24
## 5   COVID_05_49y_male_NonICU    male         no  49 ABCA1      18.39
## 6 COVID_07_38y_female_NonICU  female         no  38 ABCA1      14.66
```

```
#define a theme for all plots
New_theme <- theme(
  panel.border = element_blank(),
  panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(),
  # Set plot background
  plot.background = element_rect(fill = "white"),
  panel.background = element_blank(),
  legend.background = element_rect(fill = 'snow2'),
  legend.text = element_text(color = "black", size = 8),
  legend.title = element_text(color = "black", face = "bold", size = 8),
  legend.key = element_rect(fill = "snow2",color = "snow2"),
  legend.box.background  = element_rect(color = "black"),
  ##make the title center
  plot.title = element_text(hjust = 0.5, size = 13, face = "bold"),
  plot.subtitle = element_text(hjust = 0.5, size = 12, face = "italic"),
  title = element_text(color = "black"),
  axis.line = element_line(color = "black"),
  axis.text = element_text(color = "black"),
  legend.position = 'right'
)
```

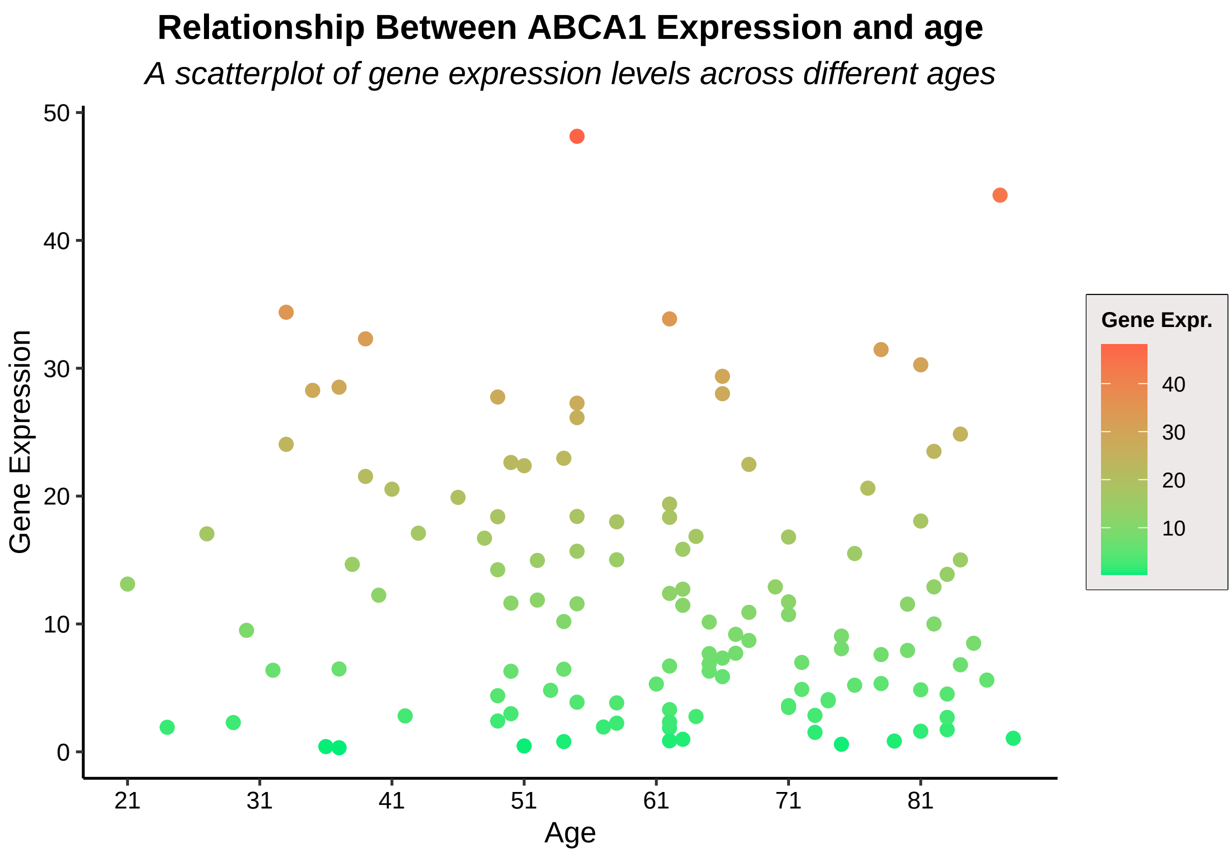
```
####Histogram for gene expression
ggplot(new_data, aes(x=expression)) + geom_histogram(bins =30, fill = "darkgreen", color = "black", alph
       x = "Gene Expression Level",
       y = "Frequency") + New_theme
```



```
####Scatterplot for gene expression and continuous covariate
##I make the color fade from green to tomato
class(new_data$expression)
```

```
## [1] "numeric"
```

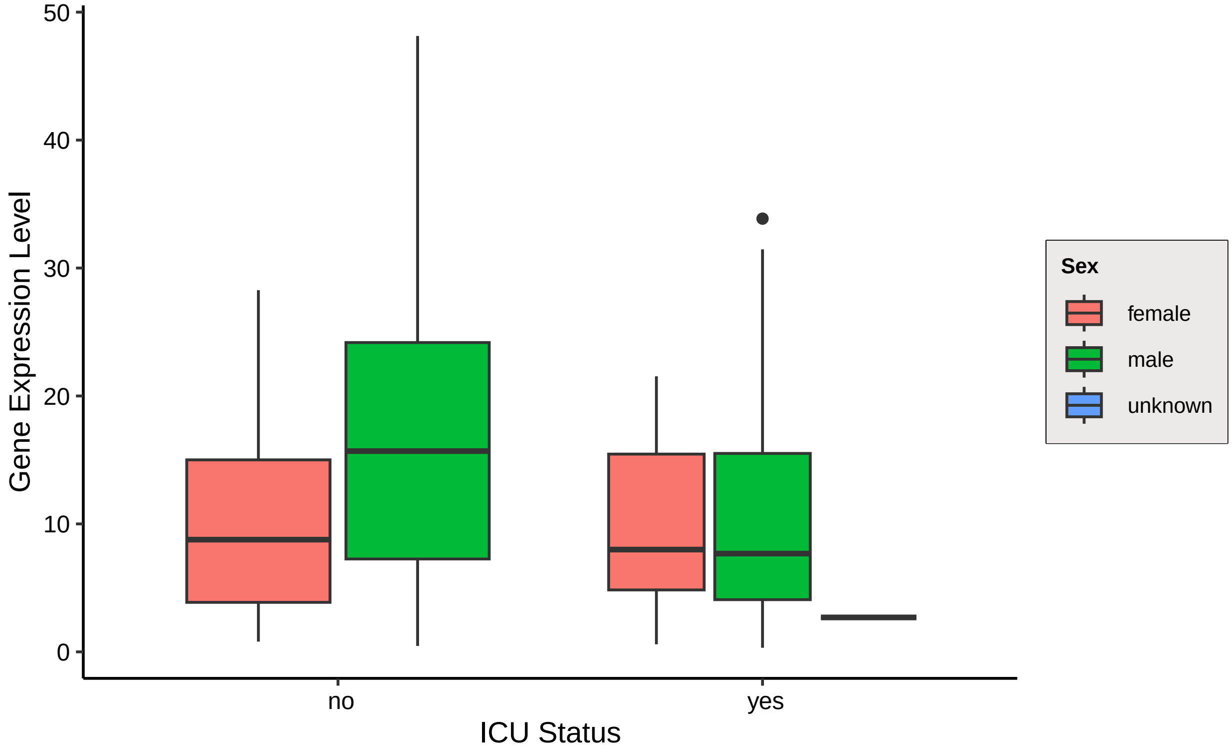
```
new_data$age[!grepl("^[0-9]+$", new_data$age)] <- NA
new_data<-new_data%>%drop_na()
new_data$age <- as.character(new_data$age)
new_data$age <- as.numeric(new_data$age)
ggplot(new_data,
       aes(x = age, y = expression, color = expression)) + geom_point(size = 2) + labs(
         title = "Relationship Between ABCA1 Expression and age",
         subtitle = "A scatterplot of gene expression levels across different ages",
         y = 'Gene Expression',
         x = 'Age',
         color = 'Gene Expr.'
       ) + scale_color_gradient(low = "springgreen2", high = "tomato1") + New_theme +scale_x_continuous
```



```
####Boxplot of gene expression separated by both categorical covariates
ggplot(new_data,aes(x = icu_status ,y = expression,fill = sex)) +geom_boxplot()+ New_theme +labs(
    title = "Distribution of Gene Expression by ICU Status and Sex",
    subtitle = "Boxplots showing variation in gene expression across ICU status and sex",
    x = "ICU Status",
    y = "Gene Expression Level",
    fill = "Sex"
  )
```

Distribution of Gene Expression by ICU Status and Sex

Boxplots showing variation in gene expression across ICU status and sex



```
#ggsave("/Users/zhyihan/Documents/Dartmouth Life/courses/QBS 103- Data Science/final project/scat.png",
```