

Final Project

Annie Glenning

2024-07-23

```
getwd() # finding my working directory
```

```
## [1] "/Users/annieglenning/Documents/GitHub/QBS103"
```

```
setwd("/Users/annieglenning/Documents/Dartmouth/SU24/QBS_103/Data") # set my working directory to where  
original_gene_data <- read.table("QBS103_GSE157103_genes.csv", header = TRUE, sep = ",") # reading in t  
series_data <- read.table("QBS103_GSE157103_series_matrix.csv", header = TRUE, sep = ",") # reading in
```

```
# transposing the gene data  
original_gene_data <- as.data.frame(original_gene_data)  
if (is.character(original_gene_data[1, 1])) {  
  gene_data <- t(original_gene_data)  
  colnames(gene_data) <- gene_data[1, ]  
  gene_data <- gene_data[-1, ]  
  gene_data <- as.data.frame(gene_data)  
}
```

Identify one gene, one continuous covariate, and two categorical covariates

```
AAGAB <- as.numeric(gene_data$AAGAB) # one gene  
AAGAB
```

```
## [1] 22.93 21.69 18.27 26.69 17.02 17.50 26.28 13.91 13.57 12.54 30.25 17.97  
## [13] 28.07 17.66 12.84 26.42 21.98 17.65 27.72 14.54 14.36 10.10 13.21 12.63  
## [25] 19.75 16.05 17.29 14.19 10.18 14.40 14.90 12.41 17.23 13.30 18.66 21.90  
## [37] 19.03 11.82 8.11 18.47 6.29 19.75 8.36 21.71 13.75 13.21 17.74 6.42  
## [49] 8.70 4.55 18.97 22.00 30.85 18.12 18.17 24.32 9.09 20.36 18.09 27.68  
## [61] 15.57 25.64 10.35 27.70 16.26 11.61 6.83 28.83 27.37 15.33 21.83 30.10  
## [73] 11.88 28.81 11.65 15.04 24.31 17.14 14.70 14.86 22.38 17.84 9.97 18.84  
## [85] 17.95 10.35 12.82 14.99 12.69 15.17 16.91 17.93 11.85 7.37 7.84 6.05  
## [97] 14.77 16.32 15.23 11.55 12.03 10.13 6.43 13.69 10.22 11.82 14.58 9.00  
## [109] 16.94 14.67 16.96 17.49 13.35 8.18 16.21 10.70 17.78 9.43 8.33 11.42  
## [121] 12.57 7.48 10.35 14.06 9.49 5.24
```

```
series_data$age # one continuous covariate
```

```
## [1] "39" "63" "33" "49" "49" " : " "38" "78" "64" "62"
```

```
## [11] "52" "50" "37" "55" "68" "48" "54" "70" "51" "62"
## [21] "66" "43" "76" "55" "55" "41" "71" "63" "63" "54"
## [31] "50" "72" "81" "64" "58" "68" "87" "68" "80" "66"
## [41] "74" "21" "83" "46" "62" "62" "78" "72" "73" "37"
## [51] "58" "71" "35" "62" "33" "30" "62" "55" "49" "54"
## [61] "78" "39" "65" "84" "66" "57" "79" "77" "81" "37"
## [71] "50" "82" "55" "55" "73" "55" "80" "27" "71" "67"
## [81] "85" "75" "62" "52" "61" " " >89" "86" "29" "82" "81"
## [91] "24" "49" "51" "76" "81" "71" "74" "58" "84" "83"
## [101] "54" "65" "65" " " >89" "83" "75" "50" "53" "49" "67"
## [111] "58" "82" "65" "75" "83" "40" "84" "88" "66" "62"
## [121] "71" "63" "42" "32" "62" "36"
```

```
series_data$sex # first categorical covariate
```

```
## [1] " male" " male" " male" " male" " male" " male"
## [7] " female" " male" " female" " male" " female" " male"
## [13] " male" " male" " male" " male" " male" " female"
## [19] " male" " male" " male" " male" " male" " male"
## [25] " male" " female" " female" " male" " female" " male"
## [31] " male" " male" " male" " female" " female" " male"
## [37] " male" " male" " female" " male" " male" " female"
## [43] " female" " male" " female" " male" " male" " female"
## [49] " male" " male" " female" " male" " female" " female"
## [55] " female" " female" " male" " male" " male" " female"
## [61] " female" " female" " male" " male" " female" " male"
## [67] " male" " female" " male" " male" " female" " male"
## [73] " female" " male" " female" " female" " male" " male"
## [79] " male" " male" " female" " female" " male" " female"
## [85] " male" " female" " female" " female" " female" " female"
## [91] " female" " male" " male" " male" " male" " male"
## [97] " female" " male" " male" " male" " female" " male"
## [103] " male" " male" " female" " female" " male" " female"
## [109] " female" " male" " female" " male" " male" " female"
## [115] " unknown" " female" " female" " male" " female" " female"
## [121] " male" " male" " female" " female" " male" " male"
```

```
series_data$icu_status # second categorical covariate
```

```
## [1] " no" " no" " no" " no" " no" " no" " no" " no" " yes" " yes" " yes"
## [11] " no" " yes" " no" " yes" " yes" " no" " no" " no" " no" " no" " yes"
## [21] " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes"
## [31] " yes" " yes" " no" " no" " no" " no" " no" " yes" " yes" " yes" " yes"
## [41] " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes"
## [51] " no" " no" " no" " yes" " no" " no" " no" " no" " no" " no" " no"
## [61] " yes" " yes" " yes" " no" " no" " yes" " yes" " no" " no" " no" " yes"
## [71] " no" " no" " yes" " no" " yes" " yes" " no" " no" " no" " yes" " no"
## [81] " no" " no" " yes" " no" " yes" " no" " no" " no" " no" " yes" " yes"
## [91] " no" " no" " no" " yes" " no" " yes" " no" " yes" " no" " no" " no"
## [101] " no" " yes" " yes" " no" " no" " yes" " yes" " yes" " yes" " no" " yes"
## [111] " no" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes" " yes"
## [121] " no" " no" " no" " no" " no" " no" " yes"
```

Generate the following three plots using ggplot2

```
library(tidyverse)
```

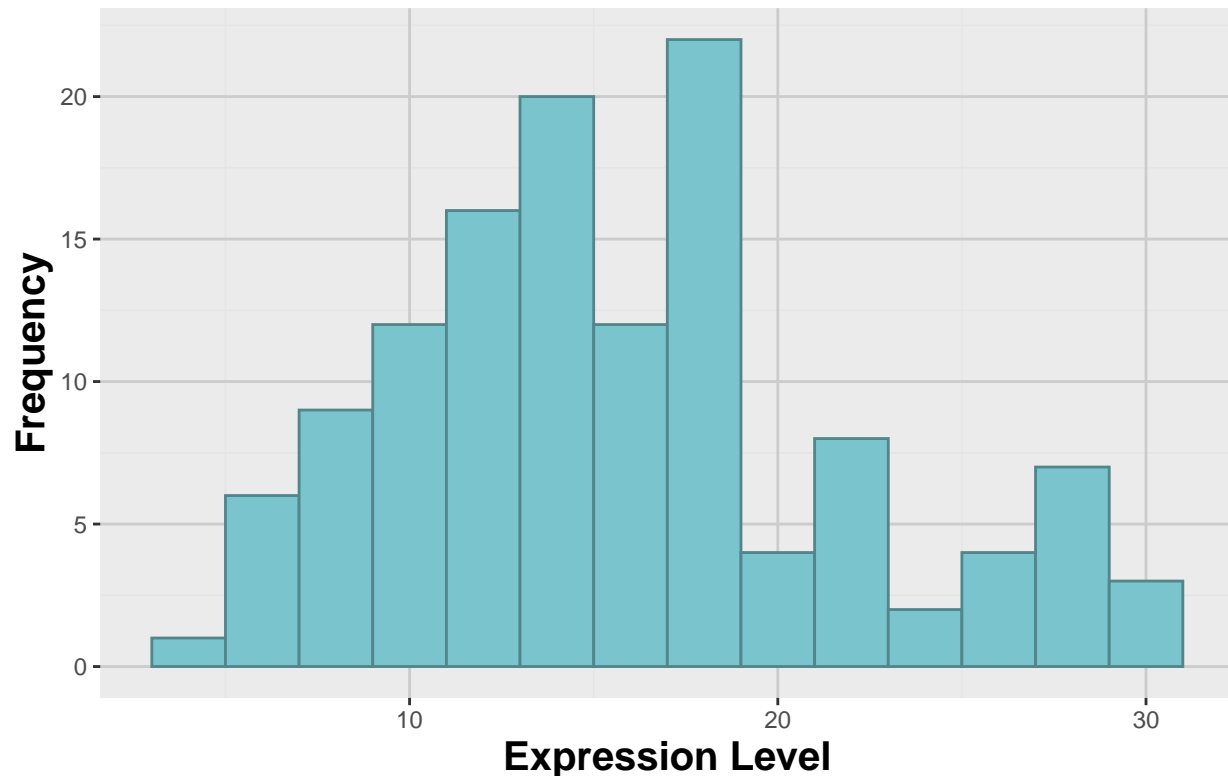
```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts()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

```
library(ggplot2)
```

Histogram for my Gene selected

```
ggplot(gene_data, aes(x = as.numeric(AAGAB))) +
  geom_histogram(binwidth = 2, fill = "cadetblue3", color = "cadetblue4") +
  labs( # labeling the title and axis
    title = "Histogram of AAGAB Gene Expression Levels",
    x = "Expression Level",
    y = "Frequency"
  ) +
  theme(
    plot.title = element_text(size = 20, face = "bold"), # title
    axis.title = element_text(size = 15, face = "bold"), # axis
    panel.grid.major = element_line(color = "grey80"), # background
    panel.grid.minor = element_line(color = "grey90")
  )
```

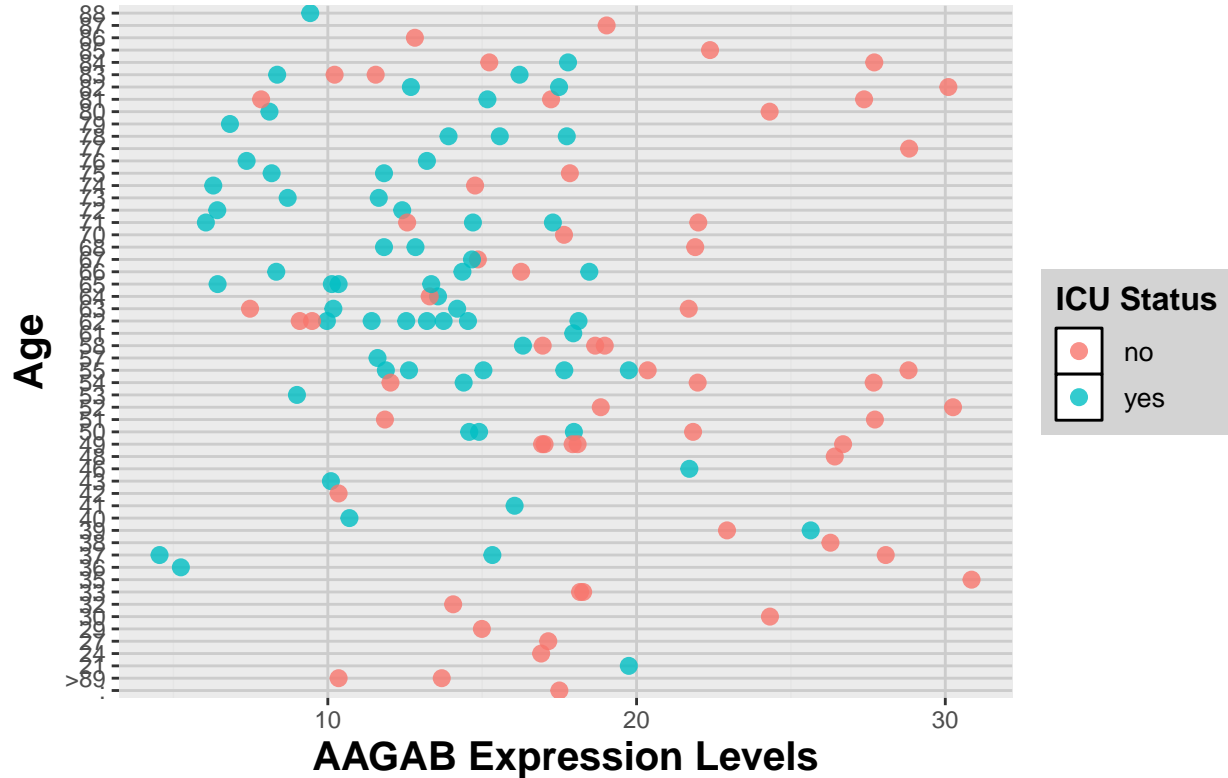
Histogram of AAGAB Gene Expression Levels



Scatterplot

```
ggplot(series_data, aes(x = as.numeric(AAGAB), y = age,
                        color = factor(icu_status))) + # seperating the points by icu status
  geom_point(size = 2.5, alpha = 0.8) + # size and transparency of the point
  labs( # labeling the title and axis
    title = "Scatterplot of the AAGAB Expression Levels vs. Age",
    x = "AAGAB Expression Levels",
    y = "Age",
    color = "ICU Status" # legend title
  ) +
  theme(
    plot.title = element_text(size = 19, face = "bold"),
    axis.title = element_text(size = 15, face = "bold"),
    legend.title = element_text(size = 12, face = "bold"), # editing the legend
    legend.text = element_text(size = 10),
    legend.background = element_rect(fill = "lightgray", color = NA),
    legend.key = element_rect(fill = "white", color = "black"),
    panel.grid.major = element_line(color = "grey80"), # background
    panel.grid.minor = element_line(color = "grey90")
  )
```

Scatterplot of the AAGAB Expression Levels vs



Box Plot

```
ggplot(series_data, aes(x = icu_status, y = as.numeric(AAGAB), color = sex)) +
  geom_boxplot() +
  labs( # labeling the title and axis
    title = "Box Plot of AAGAB Expression Levels by ICU Status and Sex",
    x = "ICU Status",
    y = "AAGAB Expression Levels"
  ) +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 12, face = "bold"),
    legend.title = element_text(size = 12, face = "bold"), # editing the legend
    legend.text = element_text(size = 10),
    legend.background = element_rect(fill = "lightgray", color = NA),
    legend.key = element_rect(fill = "white", color = "black"),
    panel.grid.major = element_line(color = "grey80"), # background
    panel.grid.minor = element_line(color = "grey90")
  )
```

Box Plot of AAGAB Expression Levels by ICU Status and

