QBS103 Project Submission 1 - Elodie Richard

2024-07-25

```
project1.data <- list.files(path = "/Users/elodierichard/Documents/QBS103/Project Submission 1 Data", pa
print(project1.data) #I first moved both data files into one folder on my laptop to retrieve it
```

```
## [1] "QBS103_GSE157103_genes.csv"         "QBS103_GSE157103_series_matrix.csv"
```

```
setwd("/Users/elodierichard/Documents/QBS103/Project Submission 1 Data") #this is to set the working di

genes <- read.csv("QBS103_GSE157103_genes.csv") #this is to rename and retrieve the first gene data fil
the_matrix <- read.csv("QBS103_GSE157103_series_matrix.csv") #this is to rename and retrieve the second

head(genes) #this is used to visualize the data and only for 6 rows
```

```
##         X COVID_01_39y_male_NonICU COVID_02_63y_male_NonICU
## 1    A1BG                     0.49                     0.29
## 2    A1CF                     0.00                     0.00
## 3     A2M                     0.21                     0.14
## 4   A2ML1                     0.04                     0.00
## 5 A3GALT2                     0.07                     0.00
## 6  A4GALT                     0.00                     0.00
##   COVID_03_33y_male_NonICU COVID_04_49y_male_NonICU COVID_05_49y_male_NonICU
## 1                     0.26                     0.45                     0.17
## 2                     0.00                     0.01                     0.00
## 3                     0.03                     0.09                     0.00
## 4                     0.02                     0.07                     0.05
## 5                     0.00                     0.00                     0.07
## 6                     0.00                     0.00                     0.00
##   COVID_06_.y_male_NonICU COVID_07_38y_female_NonICU COVID_08_78y_male_ICU
## 1                    0.21                      0.49                  0.12
## 2                    0.00                      0.01                  0.00
## 3                    0.08                      0.23                  0.08
## 4                    0.04                      0.03                  0.01
## 5                    0.00                      0.07                  0.00
## 6                    0.00                      0.00                  0.00
##   COVID_09_64y_female_ICU COVID_10_62y_male_ICU COVID_11_52y_female_NonICU
## 1                    0.51                 0.10                      0.38
## 2                    0.01                 0.00                      0.02
## 3                    0.88                 0.13                      0.47
## 4                    0.02                 0.01                      0.03
## 5                    0.79                 0.15                      0.08
## 6                    0.00                 0.00                      0.00
##   COVID_12_50y_male_ICU COVID_13_37y_male_NonICU COVID_14_55y_male_ICU
## 1                  0.45                    0.18                 0.23
## 2                  0.00                    0.00                 0.00
## 3                  0.16                    0.07                 0.22
## 4                  0.00                    0.01                 0.04
## 5                  1.75                    0.00                 0.93
```

```
## 6                  0.00                     0.00                   0.00
##   COVID_15_68y_male_ICU COVID_16_48y_male_NonICU COVID_17_54y_male_NonICU
## 1                  0.42                     0.41                    0.63
## 2                  0.00                     0.01                    0.02
## 3                  0.07                     0.58                    0.15
## 4                  0.00                     0.00                    0.02
## 5                  0.15                     0.19                    0.00
## 6                  0.03                     0.00                    0.00
##   COVID_18_70y_female_NonICU COVID_19_51y_male_NonICU COVID_20_62y_male_ICU
## 1                       0.47                     0.33                  0.32
## 2                       0.00                     0.02                  0.00
## 3                       0.30                     0.11                  0.07
## 4                       0.02                     0.02                  0.00
## 5                       0.06                     0.00                  0.22
## 6                       0.03                     0.00                  0.00
##   COVID_21_66y_male_ICU COVID_22_43y_male_ICU COVID_23_76y_male_ICU
## 1                  0.18                  0.09                 0.18
## 2                  0.00                  0.00                 0.01
## 3                  0.00                  0.06                 0.03
## 4                  0.00                  0.00                 0.00
## 5                  0.37                  0.06                 0.07
## 6                  0.03                  0.00                 0.03
##   COVID_24_55y_male_ICU COVID_25_55y_male_ICU COVID_26_41y_female_ICU
## 1                  0.22                  0.29                   0.42
## 2                  0.01                  0.00                   0.00
## 3                  0.11                  0.09                   0.18
## 4                  0.02                  0.03                   0.00
## 5                  0.15                  0.00                   0.87
## 6                  0.00                  0.00                   0.00
##   COVID_27_71y_female_ICU COVID_28_63y_male_ICU COVID_29_63y_female_ICU
## 1                    0.16                  0.18                    0.35
## 2                    0.01                  0.00                    0.00
## 3                    0.23                  0.18                    0.03
## 4                    0.01                  0.05                    0.03
## 5                    0.18                  0.45                    0.15
## 6                    0.00                  0.00                    0.03
##   COVID_30_54y_male_ICU COVID_31_50y_male_ICU COVID_32_72y_male_ICU
## 1                  0.23                  0.15                 0.34
## 2                  0.00                  0.00                 0.01
## 3                  0.11                  0.47                 0.04
## 4                  0.01                  0.00                 0.00
## 5                  0.00                  0.00                 0.29
## 6                  0.00                  0.03                 0.00
##   COVID_33_81y_male_NonICU COVID_34_64y_female_NonICU
## 1                     0.35                      0.36
## 2                     0.00                      0.00
## 3                     0.30                      0.11
## 4                     0.06                      0.00
## 5                     0.26                      0.12
## 6                     0.00                      0.00
##   COVID_35_58y_female_NonICU COVID_36_68y_male_NonICU COVID_37_87y_male_NonICU
## 1                       0.26                     0.18                    0.20
## 2                       0.00                     0.01                    0.00
## 3                       0.51                     0.09                    0.09
```

```
## 4                      0.02                    0.00                    0.07
## 5                      0.16                    0.08                    0.31
## 6                      0.00                    0.00                    0.00
##   COVID_38_68y_male_ICU COVID_39_80y_female_ICU COVID_40_66y_male_ICU
## 1                  0.29                    0.19                  0.22
## 2                  0.00                    0.00                  0.00
## 3                  0.10                    0.27                  0.17
## 4                  0.02                    0.00                  0.00
## 5                  0.35                    0.00                  0.08
## 6                  0.00                    0.07                  0.00
##   COVID_41_74y_male_ICU COVID_42_21y_female_ICU COVID_43_83y_female_ICU
## 1                  0.19                    0.24                    0.29
## 2                  0.00                    0.01                    0.00
## 3                  0.14                    0.33                    0.00
## 4                  0.00                    0.01                    0.00
## 5                  0.19                    0.39                    0.11
## 6                  0.00                    0.00                    0.00
##   COVID_44_46y_male_ICU COVID_45_62y_female_ICU COVID_46_62y_male_ICU
## 1                  0.22                    0.14                  0.53
## 2                  0.00                    0.00                  0.01
## 3                  0.14                    0.15                  0.10
## 4                  0.00                    0.03                  0.00
## 5                  0.00                    0.19                  0.06
## 6                  0.04                    0.00                  0.00
##   COVID_47_78y_male_ICU COVID_48_72y_female_ICU COVID_49_73y_male_ICU
## 1                  0.08                    0.19                  0.48
## 2                  0.01                    0.00                  0.00
## 3                  0.04                    0.06                  0.09
## 4                  0.03                    0.01                  0.03
## 5                  0.60                    0.23                  0.00
## 6                  0.00                    0.06                  0.00
##   COVID_50_37y_male_ICU COVID_51_58y_female_NonICU COVID_52_71y_male_NonICU
## 1                  0.08                      0.21                    0.25
## 2                  0.00                      0.00                    0.01
## 3                  0.01                      0.13                    0.00
## 4                  0.00                      0.00                    0.03
## 5                  0.00                      0.00                    0.00
## 6                  0.72                      0.00                    0.00
##   COVID_53_35y_female_NonICU COVID_55_62y_female_ICU COVID_56_33y_female_NonICU
## 1                       0.25                    0.09                      0.28
## 2                       0.00                    0.00                      0.00
## 3                       0.64                    0.09                      0.16
## 4                       0.10                    0.01                      0.09
## 5                       0.00                    0.00                      0.23
## 6                       0.00                    0.00                      0.00
##   COVID_57_30y_female_NonICU COVID_58_62y_male_NonICU COVID_59_55y_male_NonICU
## 1                       0.42                    0.39                    0.33
## 2                       0.00                    0.00                    0.00
## 3                       0.27                    0.08                    0.10
## 4                       0.01                    0.00                    0.00
## 5                       0.19                    0.00                    0.07
## 6                       0.05                    0.00                    0.00
##   COVID_60_49y_male_NonICU COVID_61_54y_female_NonICU COVID_62_78y_female_ICU
## 1                     0.22                      0.25                    0.21
```

```
## 2                    0.00                    0.00                    0.00
## 3                    0.14                    0.10                    0.04
## 4                    0.00                    0.03                    0.00
## 5                    0.00                    0.13                    0.05
## 6                    0.02                    0.00                    0.00
##   COVID_63_39y_female_ICU COVID_64_65y_male_ICU COVID_65_84y_male_NonICU
## 1                    0.29                  0.38                    0.40
## 2                    0.00                  0.01                    0.01
## 3                    0.01                  0.04                    0.07
## 4                    0.00                  0.02                    0.00
## 5                    0.14                  0.56                    0.58
## 6                    0.00                  0.00                    0.00
##   COVID_66_66y_female_NonICU COVID_67_57y_male_ICU COVID_68_79y_male_ICU
## 1                       0.64                 0.37                 0.58
## 2                       0.00                 0.00                 0.00
## 3                       0.00                 0.35                 0.15
## 4                       0.00                 0.00                 0.01
## 5                       0.00                 0.00                 0.00
## 6                       0.00                 0.00                 0.05
##   COVID_69_77y_female_NonICU COVID_70_81y_male_NonICU COVID_71_37y_male_ICU
## 1                       0.52                    0.27                 0.07
## 2                       0.00                    0.00                 0.01
## 3                       0.29                    0.07                 0.12
## 4                       0.02                    0.00                 0.01
## 5                       0.00                    0.00                 0.00
## 6                       0.00                    0.06                 0.00
##   COVID_72_50y_female_NonICU COVID_73_82y_male_NonICU COVID_74_55y_female_ICU
## 1                       0.52                    0.46                    0.24
## 2                       0.00                    0.01                    0.00
## 3                       0.10                    0.02                    0.12
## 4                       0.01                    0.02                    0.02
## 5                       0.00                    0.17                    0.26
## 6                       0.00                    0.04                    0.00
##   COVID_75_55y_male_NonICU COVID_76_73y_female_ICU COVID_77_55y_female_ICU
## 1                     0.23                   0.17                   0.05
## 2                     0.01                   0.00                   0.00
## 3                     0.14                   0.09                   0.01
## 4                     0.00                   0.01                   0.00
## 5                     0.00                   0.04                   0.00
## 6                     0.00                   0.00                   0.00
##   COVID_78_80y_male_NonICU COVID_79_27y_male_NonICU COVID_80_71y_male_ICU
## 1                     0.19                    0.08                 0.28
## 2                     0.00                    0.01                 0.00
## 3                     0.20                    0.03                 0.05
## 4                     0.00                    0.00                 0.00
## 5                     0.00                    0.00                 0.05
## 6                     0.00                    0.00                 0.00
##   COVID_82_67y_male_NonICU COVID_83_85y_female_NonICU
## 1                     0.39                      0.47
## 2                     0.01                      0.00
## 3                     0.10                      0.18
## 4                     0.00                      0.05
## 5                     0.00                      0.00
## 6                     0.00                      0.00
```

```
##   COVID_84_75y_female_NonICU COVID_85_62y_male_ICU COVID_86_52y_female_NonICU
## 1                       0.35                  0.29                       0.60
## 2                       0.00                  0.00                       0.00
## 3                       0.03                  0.04                       0.27
## 4                       0.00                  0.00                       0.02
## 5                       0.17                  0.00                       0.00
## 6                       0.00                  0.00                       0.00
##   COVID_87_61y_male_ICU COVID_89_90y_female_NonICU COVID_90_86y_female_NonICU
## 1                  0.65                      0.20                       0.40
## 2                  0.00                      0.00                       0.00
## 3                  0.15                      0.07                       0.05
## 4                  0.00                      0.03                       0.01
## 5                  0.00                      0.14                       0.31
## 6                  0.00                      0.00                       0.02
##   COVID_91_29y_female_NonICU COVID_92_82y_female_ICU COVID_93_81y_female_ICU
## 1                       0.60                    0.34                    0.37
## 2                       0.00                    0.00                    0.00
## 3                       0.03                    0.02                    0.11
## 4                       0.02                    0.04                    0.00
## 5                       0.05                    0.58                    0.05
## 6                       0.00                    0.00                    0.00
##   COVID_94_24y_female_NonICU COVID_95_49y_male_NonICU COVID_96_51y_male_NonICU
## 1                       0.81                     0.37                     1.61
## 2                       0.00                     0.01                     0.00
## 3                       0.17                     0.20                     0.02
## 4                       0.02                     0.02                     0.00
## 5                       0.00                     0.15                     0.00
## 6                       0.06                     0.00                     0.00
##   COVID_97_76y_male_ICU COVID_98_81y_male_NonICU COVID_99_71y_male_ICU
## 1                  0.19                    0.78                 0.33
## 2                  0.00                    0.00                 0.00
## 3                  0.02                    0.26                 0.02
## 4                  0.05                    0.00                 0.00
## 5                  0.12                    0.37                 0.04
## 6                  0.03                    0.00                 0.00
##   COVID_100_74y_female_NonICU COVID_101_58y_male_ICU COVID_102_84y_male_NonICU
## 1                        0.30                   0.33                      0.12
## 2                        0.00                   0.00                      0.00
## 3                        0.09                   0.11                      0.01
## 4                        0.00                   0.03                      0.01
## 5                        0.04                   0.05                      0.00
## 6                        0.00                   0.00                      0.07
##   COVID_103_83y_male_NonICU NONCOVID_01_54y_female_NonICU
## 1                      0.20                         0.89
## 2                      0.00                         0.00
## 3                      0.03                         0.04
## 4                      0.03                         0.00
## 5                      0.04                         0.00
## 6                      0.00                         0.00
##   NONCOVID_02_65y_male_ICU NONCOVID_03_65y_male_ICU NONCOVID_04_90y_male_NonICU
## 1                     0.32                    0.44                       0.21
## 2                     0.00                    0.00                       0.00
## 3                     0.01                    0.05                       0.05
## 4                     0.00                    0.02                       0.00
```

```
## 5                    0.04                    0.04                         0.21
## 6                    0.00                    0.00                         0.00
##   NONCOVID_05_83y_female_NonICU NONCOVID_06_75y_female_ICU
## 1                          0.31                       0.89
## 2                          0.00                       0.00
## 3                          0.01                       0.14
## 4                          0.01                       0.01
## 5                          0.00                       0.00
## 6                          0.00                       0.06
##   NONCOVID_07_50y_male_ICU NONCOVID_08_53y_female_ICU
## 1                     0.45                       0.47
## 2                     0.00                       0.01
## 3                     0.07                       0.04
## 4                     0.02                       0.00
## 5                     0.00                       0.15
## 6                     0.00                       0.00
##   NONCOVID_09_49y_female_NonICU NONCOVID_10_67y_male_ICU
## 1                          0.40                      0.33
## 2                          0.00                      0.00
## 3                          0.04                      0.05
## 4                          0.00                      0.01
## 5                          0.00                      0.23
## 6                          0.00                      0.08
##   NONCOVID_11_58y_female_NonICU NONCOVID_12_82y_male_ICU
## 1                          0.58                      0.12
## 2                          0.00                      0.00
## 3                          0.03                      0.02
## 4                          0.00                      0.00
## 5                          0.00                      0.00
## 6                          0.00                      0.02
##   NONCOVID_13_65y_male_ICU NONCOVID_14_75y_female_ICU
## 1                     0.31                       0.16
## 2                     0.00                       0.00
## 3                     0.04                       0.08
## 4                     0.01                       0.00
## 5                     0.32                       0.05
## 6                     0.02                       0.02
##   NONCOVID_15_83y_unknown_ICU NONCOVID_16_40y_female_ICU
## 1                        0.59                       0.34
## 2                        0.00                       0.00
## 3                        0.03                       0.07
## 4                        0.04                       0.00
## 5                        0.00                       0.13
## 6                        0.19                       0.00
##   NONCOVID_17_84y_female_ICU NONCOVID_18_88y_male_ICU
## 1                       0.37                      0.33
## 2                       0.00                      0.00
## 3                       0.07                      0.06
## 4                       0.01                      0.00
## 5                       0.18                      0.00
## 6                       0.00                      0.00
##   NONCOVID_19_66y_female_ICU NONCOVID_20_62y_female_ICU
## 1                       0.25                       0.20
## 2                       0.00                       0.00
```

```
## 3                      0.11                       0.01
## 4                      0.00                       0.02
## 5                      0.04                       0.00
## 6                      0.03                       0.07
##   NONCOVID_21_71y_male_NonICU NONCOVID_22_63y_male_NonICU
## 1                        0.40                        0.30
## 2                        0.00                        0.00
## 3                        0.04                        0.02
## 4                        0.02                        0.02
## 5                        0.00                        0.00
## 6                        0.00                        0.00
##   NONCOVID_23_42y_female_NonICU NONCOVID_24_32y_female_NonICU
## 1                          0.70                          0.75
## 2                          0.00                          0.00
## 3                          0.02                          0.27
## 4                          0.01                          0.00
## 5                          0.00                          0.06
## 6                          0.00                          0.00
##   NONCOVID_25_62y_male_NonICU NONCOVID_26_36y_male_ICU
## 1                        2.80                     0.22
## 2                        0.00                     0.00
## 3                        0.04                     0.28
## 4                        0.00                     0.00
## 5                        0.00                     0.00
## 6                        0.00                     0.00
```

```
head(the_matrix)
```

```
##            participant_id geo_accession               status
## 1 COVID_01_39y_male_NonICU    GSM4753021 Public on Aug 29 2020
## 2 COVID_02_63y_male_NonICU    GSM4753022 Public on Aug 29 2020
## 3 COVID_03_33y_male_NonICU    GSM4753023 Public on Aug 29 2020
## 4 COVID_04_49y_male_NonICU    GSM4753024 Public on Aug 29 2020
## 5 COVID_05_49y_male_NonICU    GSM4753025 Public on Aug 29 2020
## 6  COVID_06_:y_male_NonICU    GSM4753026 Public on Aug 29 2020
##   X.Sample_submission_date last_update_date type channel_count
## 1              Aug 28 2020      Aug 29 2020  SRA             1
## 2              Aug 28 2020      Aug 29 2020  SRA             1
## 3              Aug 28 2020      Aug 29 2020  SRA             1
## 4              Aug 28 2020      Aug 29 2020  SRA             1
## 5              Aug 28 2020      Aug 29 2020  SRA             1
## 6              Aug 28 2020      Aug 29 2020  SRA             1
##               source_name_ch1 organism_ch1          disease_status age   sex
## 1 Leukocytes from whole blood Homo sapiens disease state: COVID-19  39  male
## 2 Leukocytes from whole blood Homo sapiens disease state: COVID-19  63  male
## 3 Leukocytes from whole blood Homo sapiens disease state: COVID-19  33  male
## 4 Leukocytes from whole blood Homo sapiens disease state: COVID-19  49  male
## 5 Leukocytes from whole blood Homo sapiens disease state: COVID-19  49  male
## 6 Leukocytes from whole blood Homo sapiens disease state: COVID-19   :  male
##   icu_status apacheii charlson_score mechanical_ventilation
## 1         no       15              0                   yes
## 2         no  unknown              2                    no
## 3         no  unknown              2                    no
## 4         no  unknown              1                    no
## 5         no       19              1                   yes
```

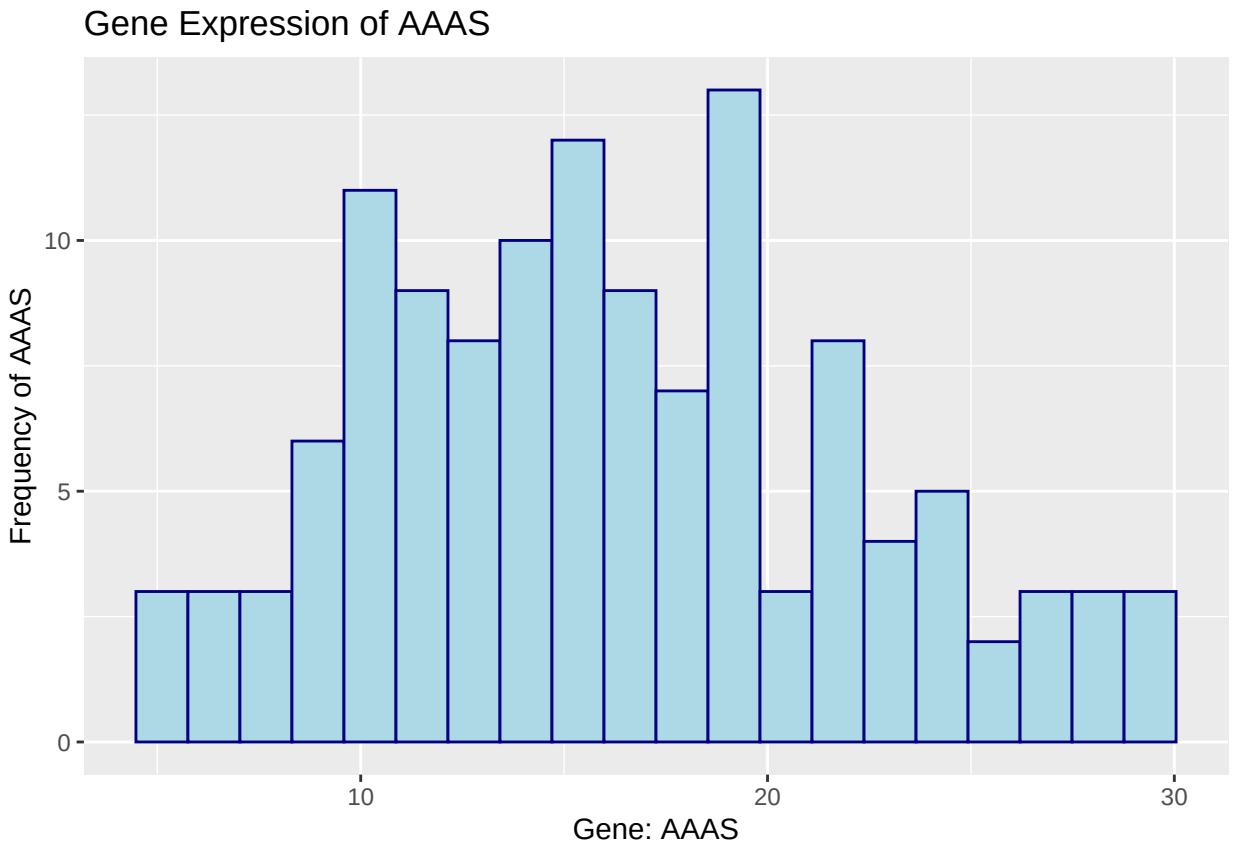
```
## 6         no  unknown             1                     no
##   ventilator.free_days hospital.free_days_post_45_day_followup ferritin.ng.ml.
## 1                    0                                       0             946
## 2                   28                                      39            1060
## 3                   28                                      18            1335
## 4                   28                                      39             583
## 5                   23                                      27             800
## 6                   28                                      36             563
##   crp.mg.l. ddimer.mg.l_feu. procalcitonin.ng.ml.. lactate.mmol.l. fibrinogen
## 1      73.1             1.3                    36            0.9        513
## 2   unknown            1.03                  0.37        unknown    unknown
## 3      53.2            1.48                  0.07        unknown        513
## 4     251.1            1.32                  0.98           0.87        949
## 5     355.8            0.69                  4.92           1.48        929
## 6     129.1         unknown                  0.67           0.86        769
##       sofa
## 1        8
## 2  unknown
## 3  unknown
## 4  unknown
## 5        7
## 6  unknown
```

```
#creating a genes data frame for the genes file
test_genes <- as.data.frame(t(genes))
names(test_genes) <- test_genes[1,] #this allows the genes table to be organized according to names by
test_genes <- test_genes[-1,] #this removes the first row containing "x" in the genes file so that it c
```

```
test_genes$participant_id <- row.names(test_genes) #this will move the participant id into it's own col
combined <- merge(test_genes, the_matrix, by = 'participant_id') #this combines the genes file and matr
```

```
#Histogram using the gene AAAS
library(ggplot2)
setwd("/Users/elodierichard/Documents/QBS103/Project Submission 1 Data")

combined$AAAS <- as.numeric(combined$AAAS) #this is so that the plot can pull just the gene AAAS from t

histogram <- ggplot(combined, aes(x=combined$AAAS)) + #this called on ggplot to use the file "combined"
  geom_histogram(bins = 20, color = 'navy', fill = 'lightblue') +  #this generated the histogram with t
  labs(title = 'Gene Expression of AAAS', #this labeled the title and the axis
        x= 'Gene: AAAS' ,
        y= 'Frequency of AAAS' )
plot(histogram)
```

```
## Warning: Use of `combined$AAAS` is discouraged.
## i Use `AAAS` instead.
```




```
#Scatterplot of the gene expression of AAAS compared to ferritin levels
library(ggplot2)
combined$ferritin.ng.ml. <- as.numeric(combined$ferritin.ng.ml.) #used to pull out ferritin to plot
```

```
## Warning: NAs introduced by coercion
```

```
#comments mostly the same as for histogram except for a few changes
scatterplot <- ggplot(combined, aes(x= combined$ferritin.ng.ml., y = combined$AAAS)) + #need to specifi
  geom_point(bins = 10, color = 'violet') +  #use geom_point for a scatter plot to be generated
  labs(title = 'Gene Expression of AAAS vs. Ferritin Levels' ,
       x= 'Ferritin Levels (ng/mL)',
       y= 'Gene Expression of AAAS')
```

```
## Warning in geom_point(bins = 10, color = "violet"): Ignoring unknown
## parameters: `bins`
```

```
plot(scatterplot)
```

```
## Warning: Use of `combined$ferritin.ng.ml.` is discouraged.
## i Use `ferritin.ng.ml.` instead.
```

```
## Warning: Use of `combined$AAAS` is discouraged.
## i Use `AAAS` instead.
```

```
## Warning: Removed 16 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

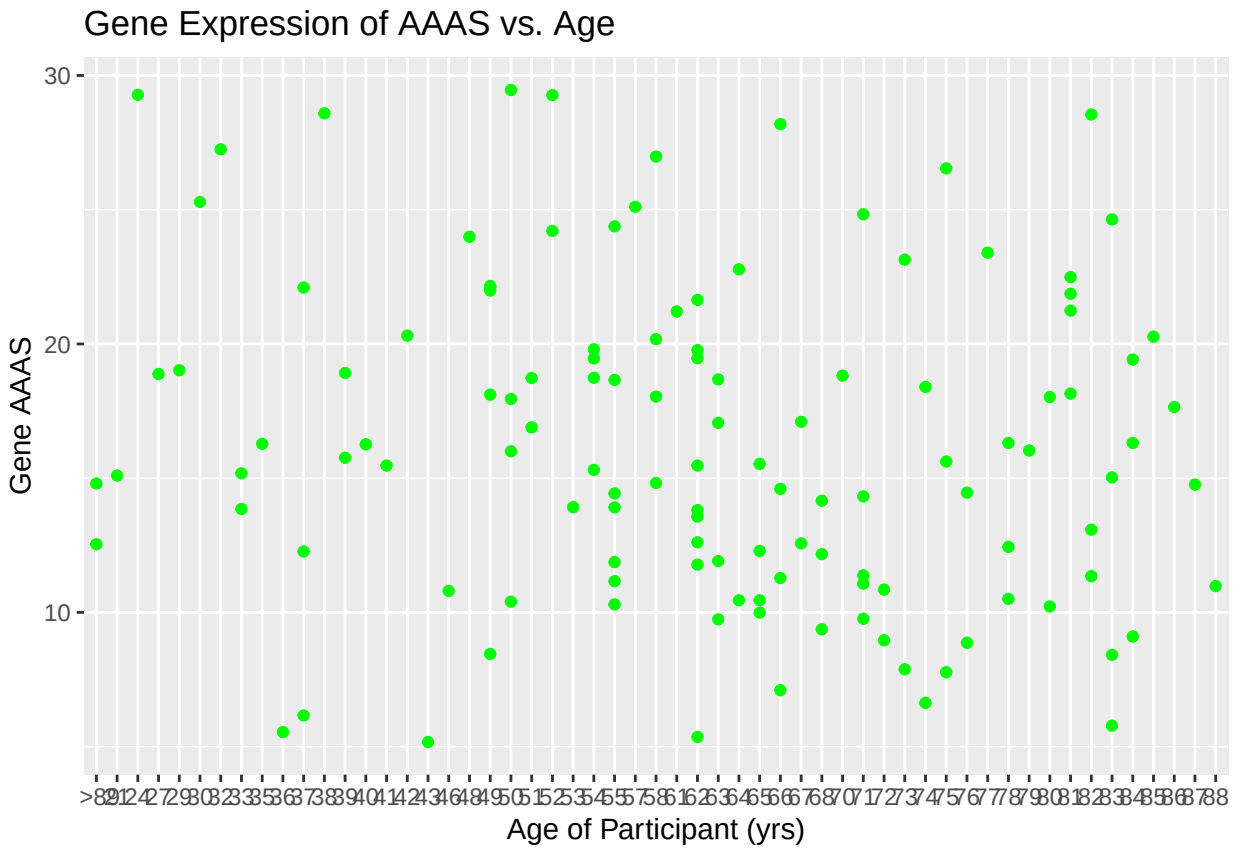

```
#Scatterplot for Gene Expression vs Age (this was run to compare different data to see differences) not
library(ggplot2)

scatterplot_practice<- ggplot(combined, aes(x= combined$age, y = combined$AAAS)) +
  geom_point(bins = 10, color = 'green') +
  labs(title = 'Gene Expression of AAAS vs. Age' , x= 'Age of Participant (yrs)', y= 'Gene AAAS')
```

```
## Warning in geom_point(bins = 10, color = "green"): Ignoring unknown parameters:
## `bins`
```

```
plot(scatterplot_practice)
```

```
## Warning: Use of `combined$age` is discouraged.
## i Use `age` instead.
```

```
## Warning: Use of `combined$AAAS` is discouraged.
## i Use `AAAS` instead.
```




```
#Boxplot comparing gene expression of AAAS related to ICU status depending on Age
library(ggplot2)
#similar process to histogram and scatterplot with a few adjustments
boxplot <- ggplot(combined, aes(x=icu_status, y = AAAS, fill = age)) + #need to add a fill to demonstra
  geom_boxplot(bins = 10, fill = 'maroon') + #to generate a box plot use geom_boxplot
  labs(title = 'Gene Expression of AAAS vs ICU Status and Participant Age',  #to label each attribute o
       x= 'ICU Status of Participant' ,
       y= 'Gene Expression of AAAS',
       fill= 'Age of Participant (yrs)')
```

```
## Warning in geom_boxplot(bins = 10, fill = "maroon"): Ignoring unknown
## parameters: `bins`
```

```
plot(boxplot)
```


Gene Expression of AAAS vs ICU Status and Participant Age
Gene Expression of AAAS
30
20
10
no
yes
ICU Status of Participant