Untitled

2025-07-15

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```r
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:
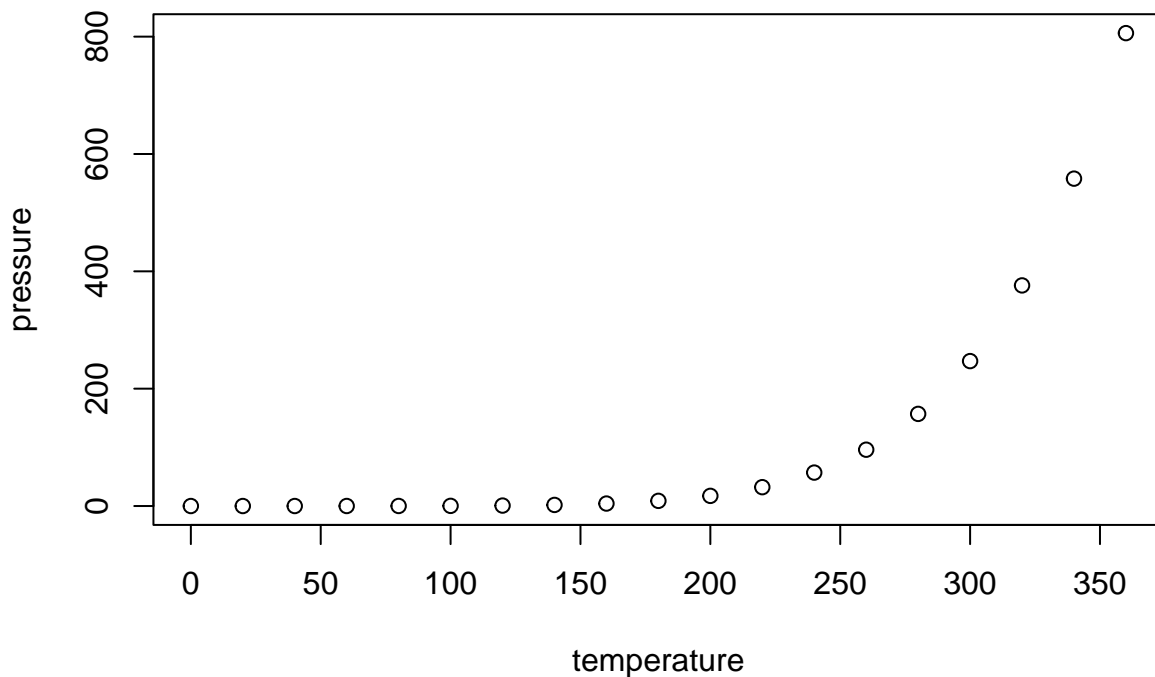
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```r
gene_expr <- read_csv("QBS103_GSE157103_genes.csv")
```

```
## New names:
## Rows: 100 Columns: 127
## -- Column specification
## ---------------------------------------------------------- Delimiter: "," chr
## (1): ...1 dbl (126): COVID_01_39y_male_NonICU, COVID_02_63y_male_NonICU,
## COVID_03_33y_...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```r
metadata <- read_csv("QBS103_GSE157103_series_matrix-1.csv")
```

```
## Rows: 126 Columns: 25
## -- Column specification ----------------------------------------------------------
## Delimiter: ","
## chr (21): participant_id, geo_accession, status, !Sample_submission_date, la...
## dbl  (4): channel_count, charlson_score, ventilator-free_days, hospital-free...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Transpose gene expression data so that genes are columns and samples are rows
gene_expr_t <- gene_expr %>%
  column_to_rownames(var = colnames(gene_expr)[1]) %>%
  t() %>%
  as.data.frame() %>%
  rownames_to_column("participant_id")

# Extract age, sex, and ICU status from participant_id
extract_info <- function(pid) {
  parts <- str_match(pid, ".*_(\\d+)y_(male|female)_(ICU|NonICU)")
  return(data.frame(
    age = as.numeric(parts[,2]),
    sex = parts[,3],
    ICU_status = parts[,4]
  ))
}
info <- extract_info(gene_expr_t$participant_id)
gene_expr_t <- cbind(gene_expr_t, info)

# Optionally join with metadata
full_data <- left_join(gene_expr_t, metadata, by = "participant_id")

# Select one gene (e.g., A1BG) and the covariates
plot_df <- full_data %>%
  select(
    participant_id,
    gene_expression = A1BG,
    age = age.x,
    sex = sex.x,
    ICU_status
  ) %>%
  drop_na()
```

```
print(plot_df)
```

```
##                participant_id gene_expression age     sex ICU_status
## 1        COVID_01_39y_male_NonICU            0.49  39    male     NonICU
## 2        COVID_02_63y_male_NonICU            0.29  63    male     NonICU
## 3        COVID_03_33y_male_NonICU            0.26  33    male     NonICU
## 4        COVID_04_49y_male_NonICU            0.45  49    male     NonICU
## 5        COVID_05_49y_male_NonICU            0.17  49    male     NonICU
## 6      COVID_07_38y_female_NonICU            0.49  38  female     NonICU
## 7          COVID_08_78y_male_ICU            0.12  78    male        ICU
## 8        COVID_09_64y_female_ICU            0.51  64  female        ICU
## 9          COVID_10_62y_male_ICU            0.10  62    male        ICU
## 10     COVID_11_52y_female_NonICU            0.38  52  female     NonICU
## 11         COVID_12_50y_male_ICU            0.45  50    male        ICU
## 12       COVID_13_37y_male_NonICU            0.18  37    male     NonICU
## 13         COVID_14_55y_male_ICU            0.23  55    male        ICU
## 14         COVID_15_68y_male_ICU            0.42  68    male        ICU
## 15       COVID_16_48y_male_NonICU            0.41  48    male     NonICU
## 16       COVID_17_54y_male_NonICU            0.63  54    male     NonICU
## 17     COVID_18_70y_female_NonICU            0.47  70  female     NonICU
## 18       COVID_19_51y_male_NonICU            0.33  51    male     NonICU
## 19         COVID_20_62y_male_ICU            0.32  62    male        ICU
## 20         COVID_21_66y_male_ICU            0.18  66    male        ICU
## 21         COVID_22_43y_male_ICU            0.09  43    male        ICU
## 22         COVID_23_76y_male_ICU            0.18  76    male        ICU
## 23         COVID_24_55y_male_ICU            0.22  55    male        ICU
## 24         COVID_25_55y_male_ICU            0.29  55    male        ICU
## 25       COVID_26_41y_female_ICU            0.42  41  female        ICU
## 26       COVID_27_71y_female_ICU            0.16  71  female        ICU
## 27         COVID_28_63y_male_ICU            0.18  63    male        ICU
## 28       COVID_29_63y_female_ICU            0.35  63  female        ICU
## 29         COVID_30_54y_male_ICU            0.23  54    male        ICU
## 30         COVID_31_50y_male_ICU            0.15  50    male        ICU
## 31         COVID_32_72y_male_ICU            0.34  72    male        ICU
## 32       COVID_33_81y_male_NonICU            0.35  81    male     NonICU
## 33     COVID_34_64y_female_NonICU            0.36  64  female     NonICU
## 34     COVID_35_58y_female_NonICU            0.26  58  female     NonICU
## 35       COVID_36_68y_male_NonICU            0.18  68    male     NonICU
## 36       COVID_37_87y_male_NonICU            0.20  87    male     NonICU
## 37         COVID_38_68y_male_ICU            0.29  68    male        ICU
## 38       COVID_39_80y_female_ICU            0.19  80  female        ICU
## 39         COVID_40_66y_male_ICU            0.22  66    male        ICU
## 40         COVID_41_74y_male_ICU            0.19  74    male        ICU
## 41       COVID_42_21y_female_ICU            0.24  21  female        ICU
## 42       COVID_43_83y_female_ICU            0.29  83  female        ICU
## 43         COVID_44_46y_male_ICU            0.22  46    male        ICU
## 44       COVID_45_62y_female_ICU            0.14  62  female        ICU
## 45         COVID_46_62y_male_ICU            0.53  62    male        ICU
## 46         COVID_47_78y_male_ICU            0.08  78    male        ICU
## 47       COVID_48_72y_female_ICU            0.19  72  female        ICU
## 48         COVID_49_73y_male_ICU            0.48  73    male        ICU
## 49         COVID_50_37y_male_ICU            0.08  37    male        ICU
## 50     COVID_51_58y_female_NonICU            0.21  58  female     NonICU
## 51       COVID_52_71y_male_NonICU            0.25  71    male     NonICU
```

```
## 52       COVID_53_35y_female_NonICU    0.25  35 female    NonICU
## 53          COVID_55_62y_female_ICU    0.09  62 female       ICU
## 54       COVID_56_33y_female_NonICU    0.28  33 female    NonICU
## 55       COVID_57_30y_female_NonICU    0.42  30 female    NonICU
## 56        COVID_58_62y_male_NonICU    0.39  62   male    NonICU
## 57        COVID_59_55y_male_NonICU    0.33  55   male    NonICU
## 58        COVID_60_49y_male_NonICU    0.22  49   male    NonICU
## 59       COVID_61_54y_female_NonICU    0.25  54 female    NonICU
## 60         COVID_62_78y_female_ICU    0.21  78 female       ICU
## 61         COVID_63_39y_female_ICU    0.29  39 female       ICU
## 62           COVID_64_65y_male_ICU    0.38  65   male       ICU
## 63        COVID_65_84y_male_NonICU    0.40  84   male    NonICU
## 64       COVID_66_66y_female_NonICU    0.64  66 female    NonICU
## 65           COVID_67_57y_male_ICU    0.37  57   male       ICU
## 66           COVID_68_79y_male_ICU    0.58  79   male       ICU
## 67       COVID_69_77y_female_NonICU    0.52  77 female    NonICU
## 68        COVID_70_81y_male_NonICU    0.27  81   male    NonICU
## 69           COVID_71_37y_male_ICU    0.07  37   male       ICU
## 70       COVID_72_50y_female_NonICU    0.52  50 female    NonICU
## 71        COVID_73_82y_male_NonICU    0.46  82   male    NonICU
## 72         COVID_74_55y_female_ICU    0.24  55 female       ICU
## 73        COVID_75_55y_male_NonICU    0.23  55   male    NonICU
## 74         COVID_76_73y_female_ICU    0.17  73 female       ICU
## 75         COVID_77_55y_female_ICU    0.05  55 female       ICU
## 76        COVID_78_80y_male_NonICU    0.19  80   male    NonICU
## 77        COVID_79_27y_male_NonICU    0.08  27   male    NonICU
## 78           COVID_80_71y_male_ICU    0.28  71   male       ICU
## 79        COVID_82_67y_male_NonICU    0.39  67   male    NonICU
## 80       COVID_83_85y_female_NonICU    0.47  85 female    NonICU
## 81       COVID_84_75y_female_NonICU    0.35  75 female    NonICU
## 82           COVID_85_62y_male_ICU    0.29  62   male       ICU
## 83       COVID_86_52y_female_NonICU    0.60  52 female    NonICU
## 84           COVID_87_61y_male_ICU    0.65  61   male       ICU
## 85       COVID_89_90y_female_NonICU    0.20  90 female    NonICU
## 86       COVID_90_86y_female_NonICU    0.40  86 female    NonICU
## 87       COVID_91_29y_female_NonICU    0.60  29 female    NonICU
## 88         COVID_92_82y_female_ICU    0.34  82 female       ICU
## 89         COVID_93_81y_female_ICU    0.37  81 female       ICU
## 90       COVID_94_24y_female_NonICU    0.81  24 female    NonICU
## 91        COVID_95_49y_male_NonICU    0.37  49   male    NonICU
## 92        COVID_96_51y_male_NonICU    1.61  51   male    NonICU
## 93           COVID_97_76y_male_ICU    0.19  76   male       ICU
## 94        COVID_98_81y_male_NonICU    0.78  81   male    NonICU
## 95           COVID_99_71y_male_ICU    0.33  71   male       ICU
## 96      COVID_100_74y_female_NonICU    0.30  74 female    NonICU
## 97          COVID_101_58y_male_ICU    0.33  58   male       ICU
## 98       COVID_102_84y_male_NonICU    0.12  84   male    NonICU
## 99       COVID_103_83y_male_NonICU    0.20  83   male    NonICU
## 100 NONCOVID_01_54y_female_NonICU    0.89  54 female    NonICU
## 101       NONCOVID_02_65y_male_ICU    0.32  65   male       ICU
## 102       NONCOVID_03_65y_male_ICU    0.44  65   male       ICU
## 103    NONCOVID_04_90y_male_NonICU    0.21  90   male    NonICU
## 104 NONCOVID_05_83y_female_NonICU    0.31  83 female    NonICU
## 105    NONCOVID_06_75y_female_ICU    0.89  75 female       ICU
```

```
## 106        NONCOVID_07_50y_male_ICU          0.45  50   male         ICU
## 107      NONCOVID_08_53y_female_ICU          0.47  53 female         ICU
## 108 NONCOVID_09_49y_female_NonICU          0.40  49 female      NonICU
## 109        NONCOVID_10_67y_male_ICU          0.33  67   male         ICU
## 110 NONCOVID_11_58y_female_NonICU          0.58  58 female      NonICU
## 111        NONCOVID_12_82y_male_ICU          0.12  82   male         ICU
## 112        NONCOVID_13_65y_male_ICU          0.31  65   male         ICU
## 113      NONCOVID_14_75y_female_ICU          0.16  75 female         ICU
## 114      NONCOVID_16_40y_female_ICU          0.34  40 female         ICU
## 115      NONCOVID_17_84y_female_ICU          0.37  84 female         ICU
## 116        NONCOVID_18_88y_male_ICU          0.33  88   male         ICU
## 117      NONCOVID_19_66y_female_ICU          0.25  66 female         ICU
## 118      NONCOVID_20_62y_female_ICU          0.20  62 female         ICU
## 119      NONCOVID_21_71y_male_NonICU          0.40  71   male      NonICU
## 120      NONCOVID_22_63y_male_NonICU          0.30  63   male      NonICU
## 121 NONCOVID_23_42y_female_NonICU          0.70  42 female      NonICU
## 122 NONCOVID_24_32y_female_NonICU          0.75  32 female      NonICU
## 123      NONCOVID_25_62y_male_NonICU          2.80  62   male      NonICU
## 124        NONCOVID_26_36y_male_ICU          0.22  36   male         ICU
```

```r
ggplot(plot_df, aes(x = gene_expression)) +
  geom_histogram(bins = 15, fill = "steelblue", color = "black") +
  labs(
    title = "Histogram of A1BG Expression",
    x = "A1BG Expression",
    y = "Count"
  ) +
  theme_minimal()
```

Histogram of A1BG Expression
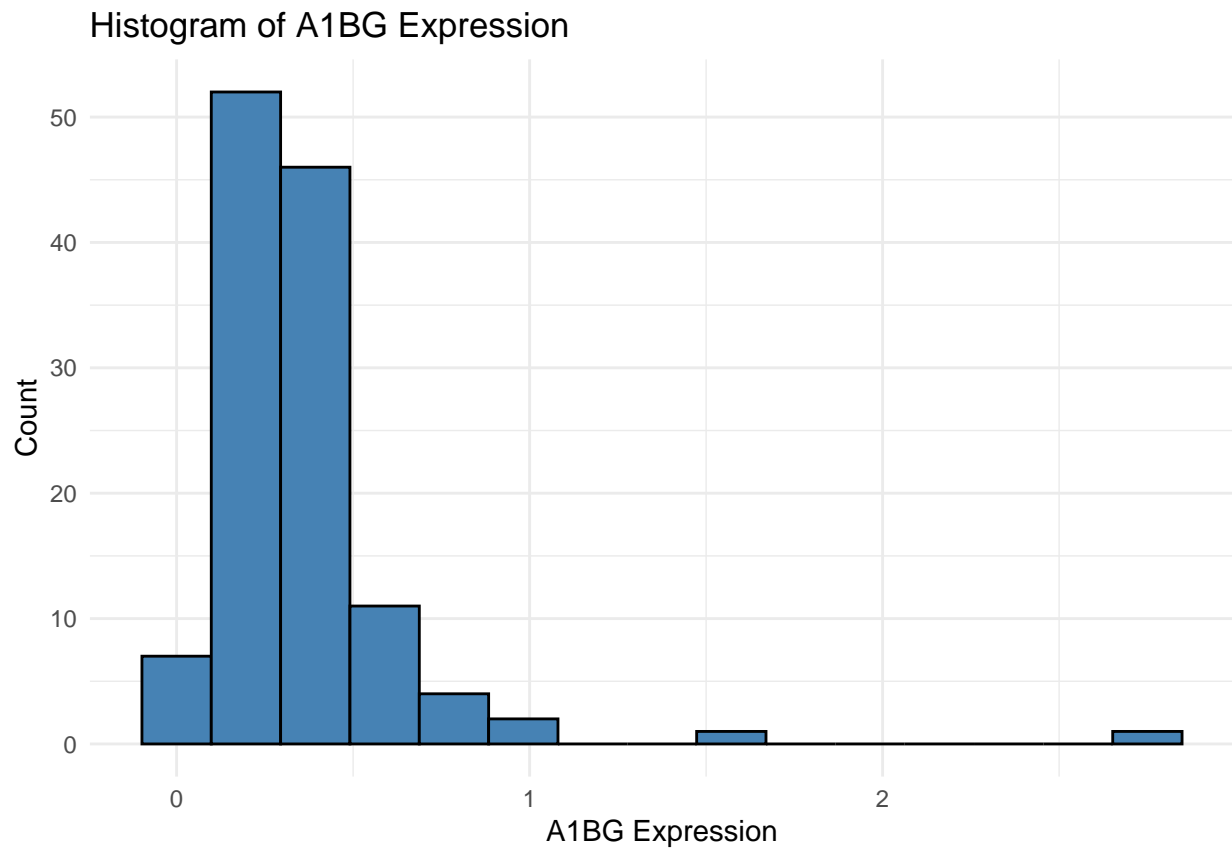

```
ggplot(plot_df, aes(x = age, y = gene_expression, color = sex, shape = ICU_status)) +
  geom_point(size = 2.5, alpha = 0.8) +
  labs(
    title = "A1BG Expression vs Age",
    x = "Age",
    y = "A1BG Expression"
  ) +
  theme_minimal()
```

A1BG Expression vs Age
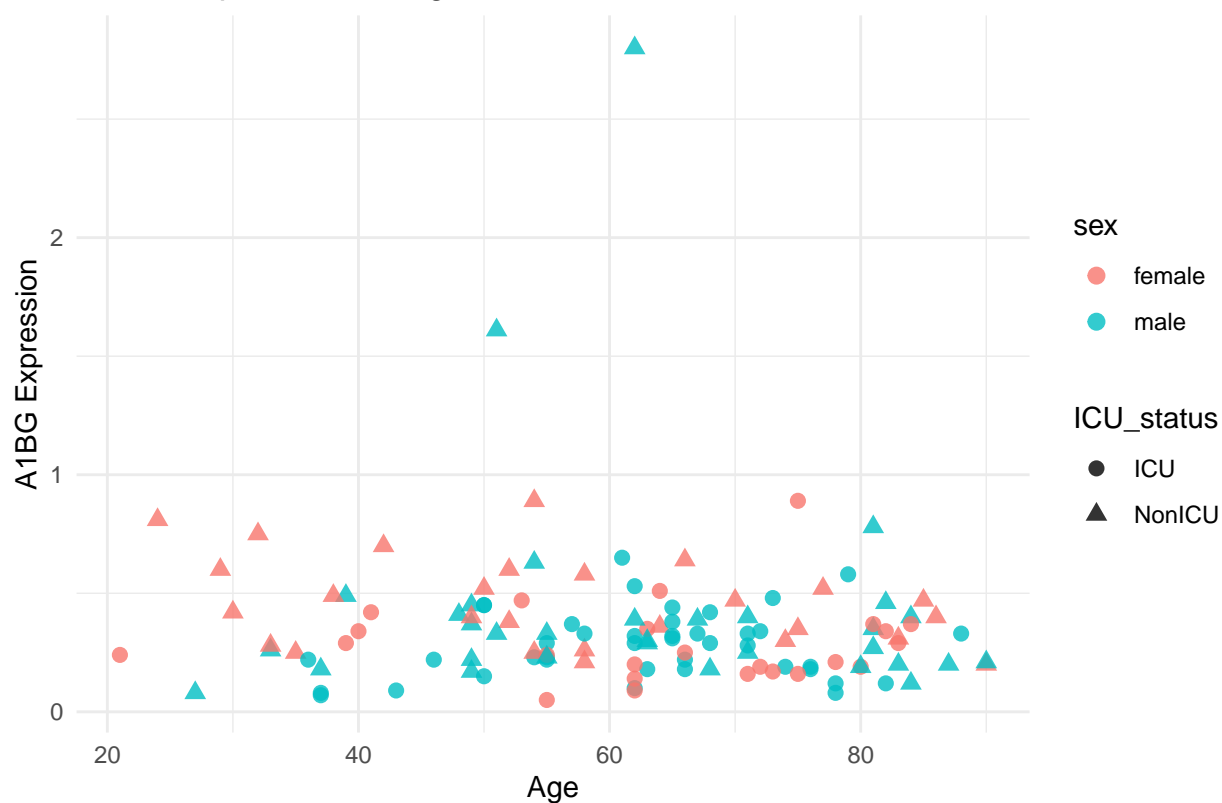

```
ggplot(plot_df, aes(x = sex, y = gene_expression, fill = ICU_status)) +
  geom_boxplot() +
  labs(
    title = "Boxplot of A1BG Expression by Sex and ICU Status",
    x = "Sex",
    y = "A1BG Expression"
  ) +
  theme_minimal()
```

Boxplot of A1BG Expression by Sex and ICU Status