

submission2

Xinqiao Li

2025-07-28

```
#task1
library(ggplot2)
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ lubridate  1.9.4      ✓ tibble     3.3.0
## ✓ purrr      1.0.4      ✓ tidyr      1.3.1
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```

plot_gene_expression <- function(data, gene_list, cont_cov, cat_cov1, cat_cov2) {
  for (gene in gene_list) {
    if (!gene %in% colnames(data)) {
      message(paste("Gene", gene, "not found. Skipping."))
      next
    }

    # histogram
    hist_plot <- ggplot(data, aes_string(x = gene)) +
      geom_histogram(binwidth = 0.1, fill = "lightblue", color = "black", alpha = 0.
7) +
      labs(
        title = paste("Histogram of", gene, "Expression"),
        x = paste(gene, "Expression"),
        y = "Count"
      ) +
      theme_minimal()
    print(hist_plot)

    # scatterplot
    scatter <- ggplot(data, aes_string(x = cont_cov, y = gene)) +
      geom_point(color = "darkgreen", alpha = 0.7) +
      labs(
        title = paste("Scatterplot of", gene, "vs", cont_cov),
        x = cont_cov,
        y = paste(gene, "Expression")
      ) +
      theme_minimal()
    print(scatter)

    # boxplot
    combined_box <- ggplot(data, aes_string(x = cat_cov1, y = gene, fill = cat_cov2))
+
      geom_boxplot(alpha = 0.7) +
      labs(
        title = paste("Boxplot of", gene, "by", cat_cov1, "and", cat_cov2),
        x = cat_cov1,
        y = paste(gene, "Expression")
      ) +
      theme_minimal()
    print(combined_box)
  }
}

```

```
# task2
library(ggplot2)
library(tidyverse)

setwd("/Users/zaozao/Desktop/Xinqiao")
gene_expression <- read.csv("QBS103_GSE157103_genes.csv", row.names = 1, check.names
= FALSE)
meta <- read.csv("QBS103_GSE157103_series_matrix-1.csv", check.names = FALSE)

selected_genes <- c("AASS", "ABCD1", "ABI2")
expr_data <- as.data.frame(t(gene_expression[selected_genes, ]))
colnames(expr_data) <- selected_genes
expr_data$participant_id <- rownames(expr_data)

meta$participant_id <- as.character(meta[[1]])
full_data <- merge(meta, expr_data, by = "participant_id")
full_data$age <- as.numeric(as.character(full_data$age))
```

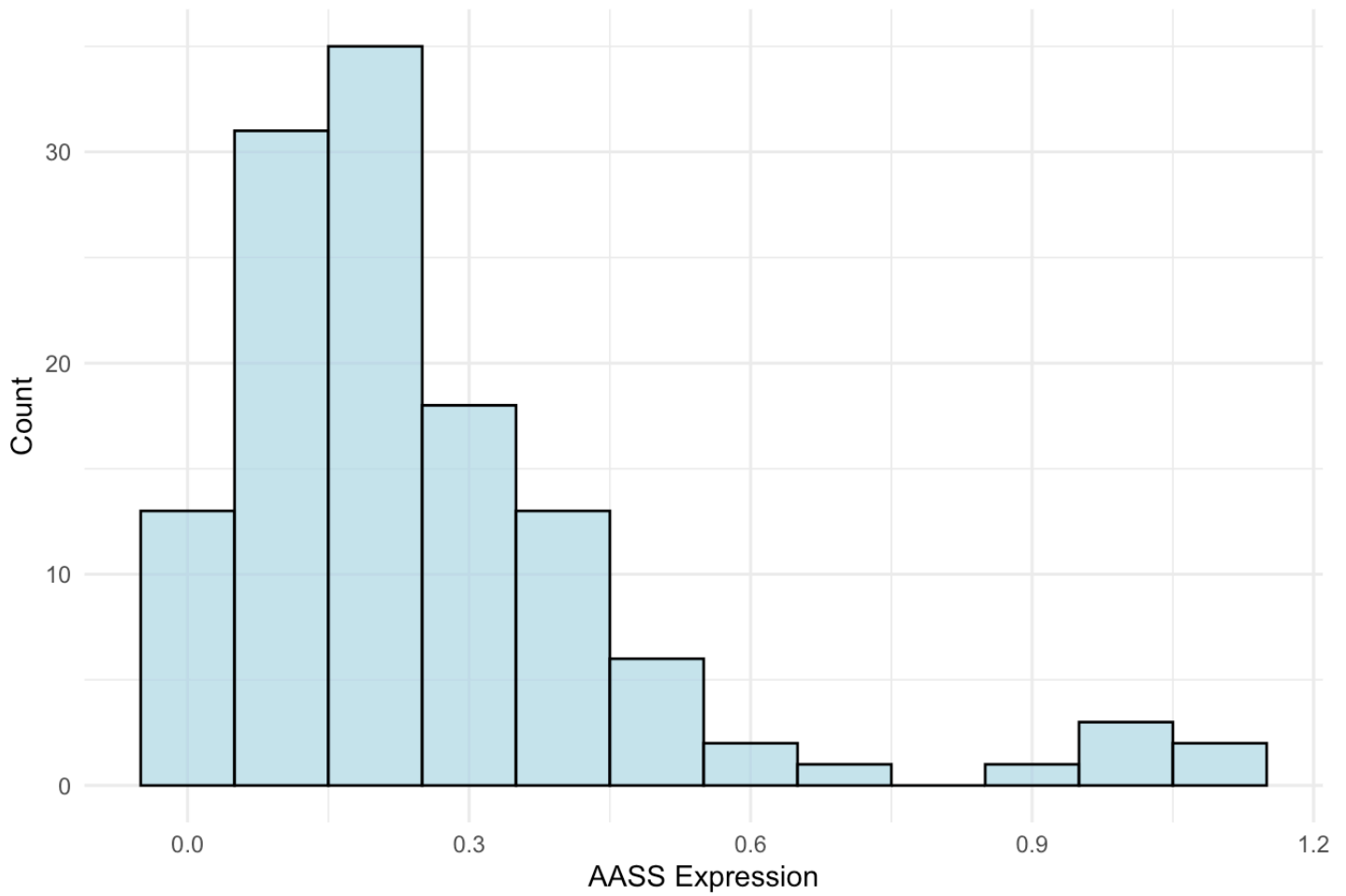
```
## Warning: NAs introduced by coercion
```

```
full_data <- full_data[full_data$icu_status != "_id", ]

# making plot
plot_gene_expression(
  data = full_data,
  gene_list = selected_genes,
  cont_cov = "age",
  cat_cov1 = "sex",
  cat_cov2 = "icu_status"
)
```

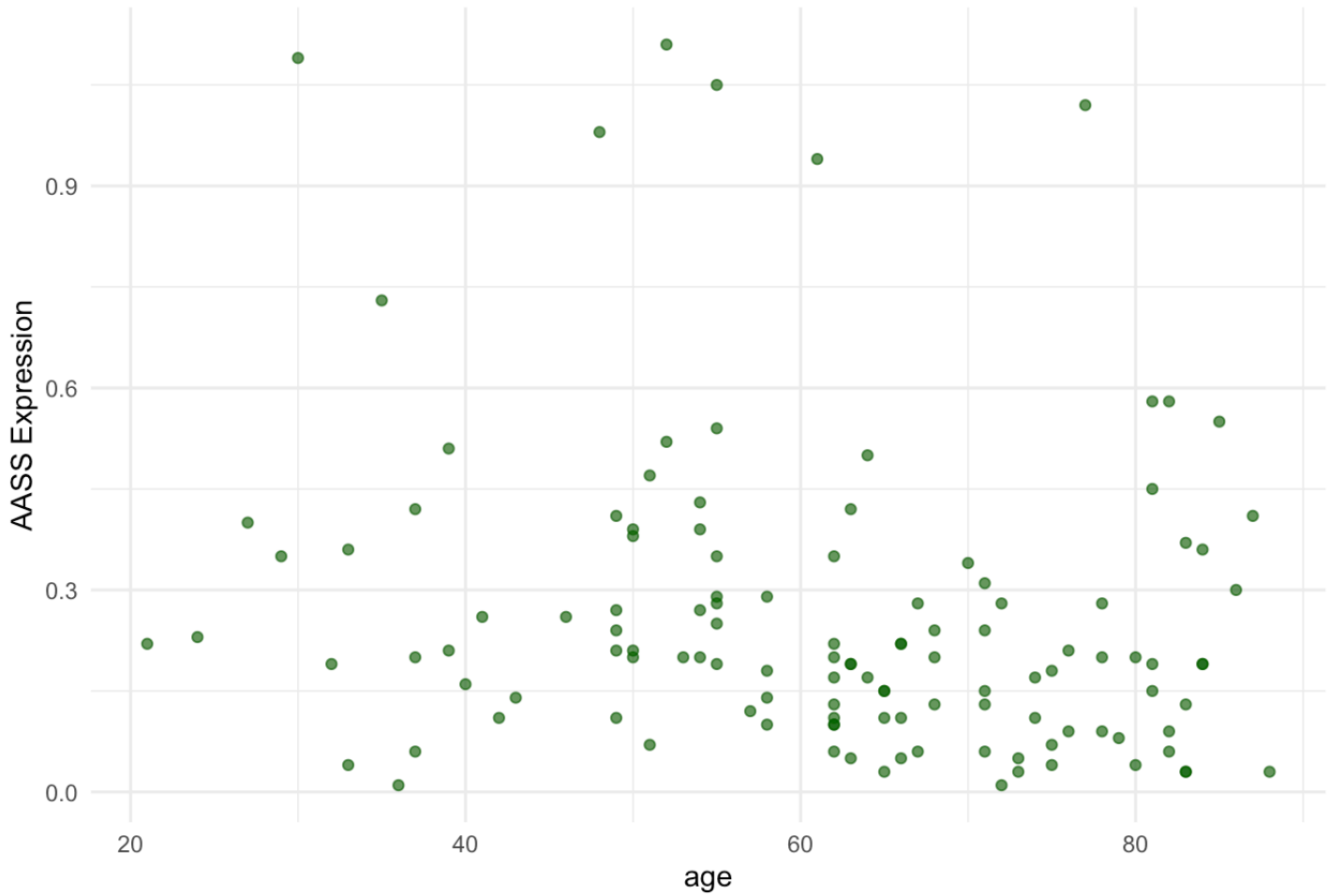
```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Histogram of AASS Expression

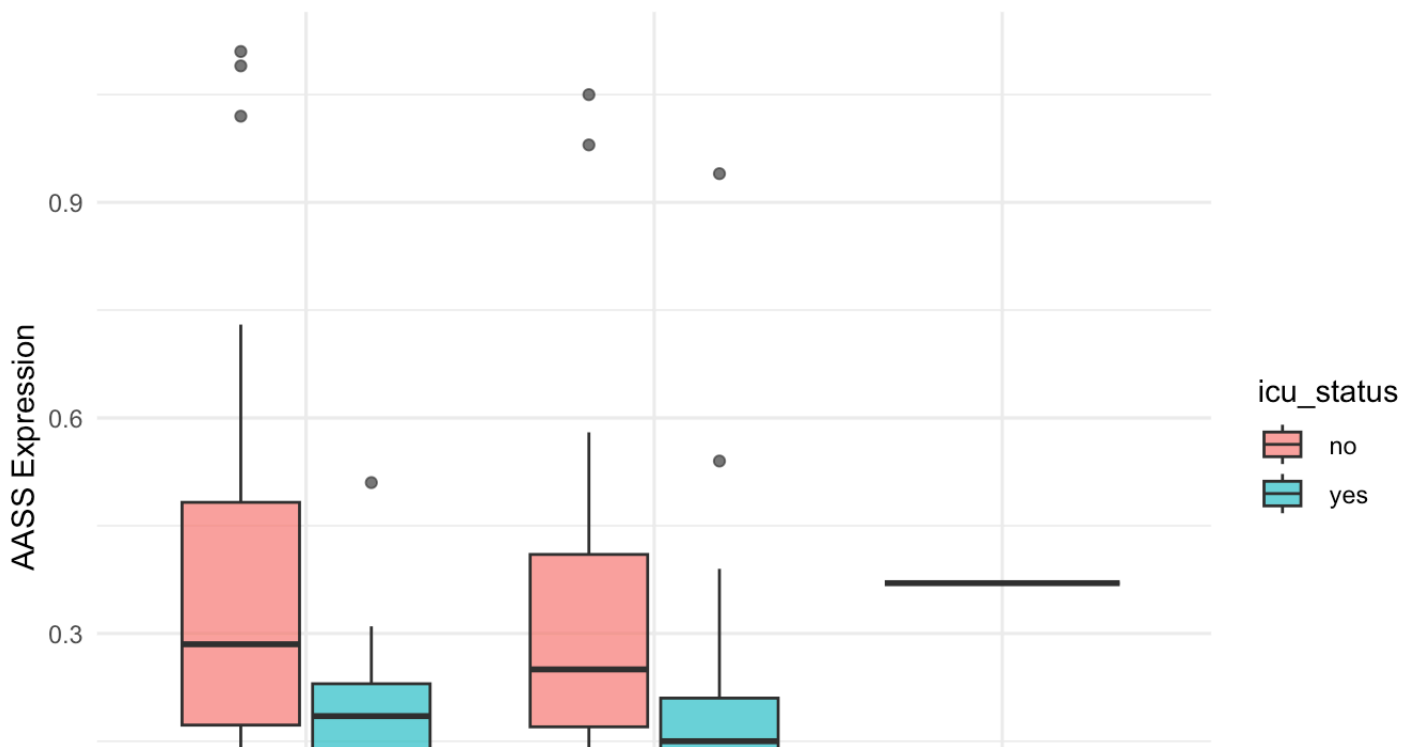


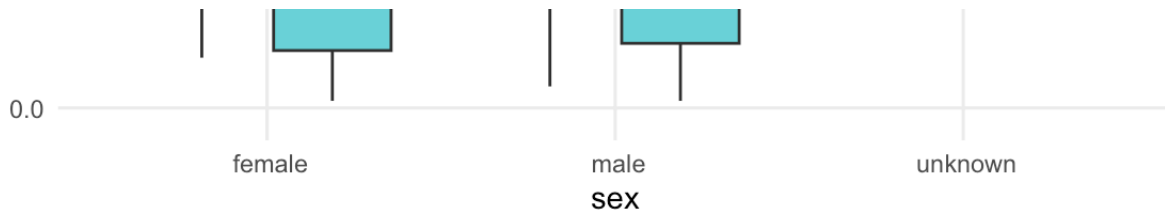
```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```


Scatterplot of AASS vs age

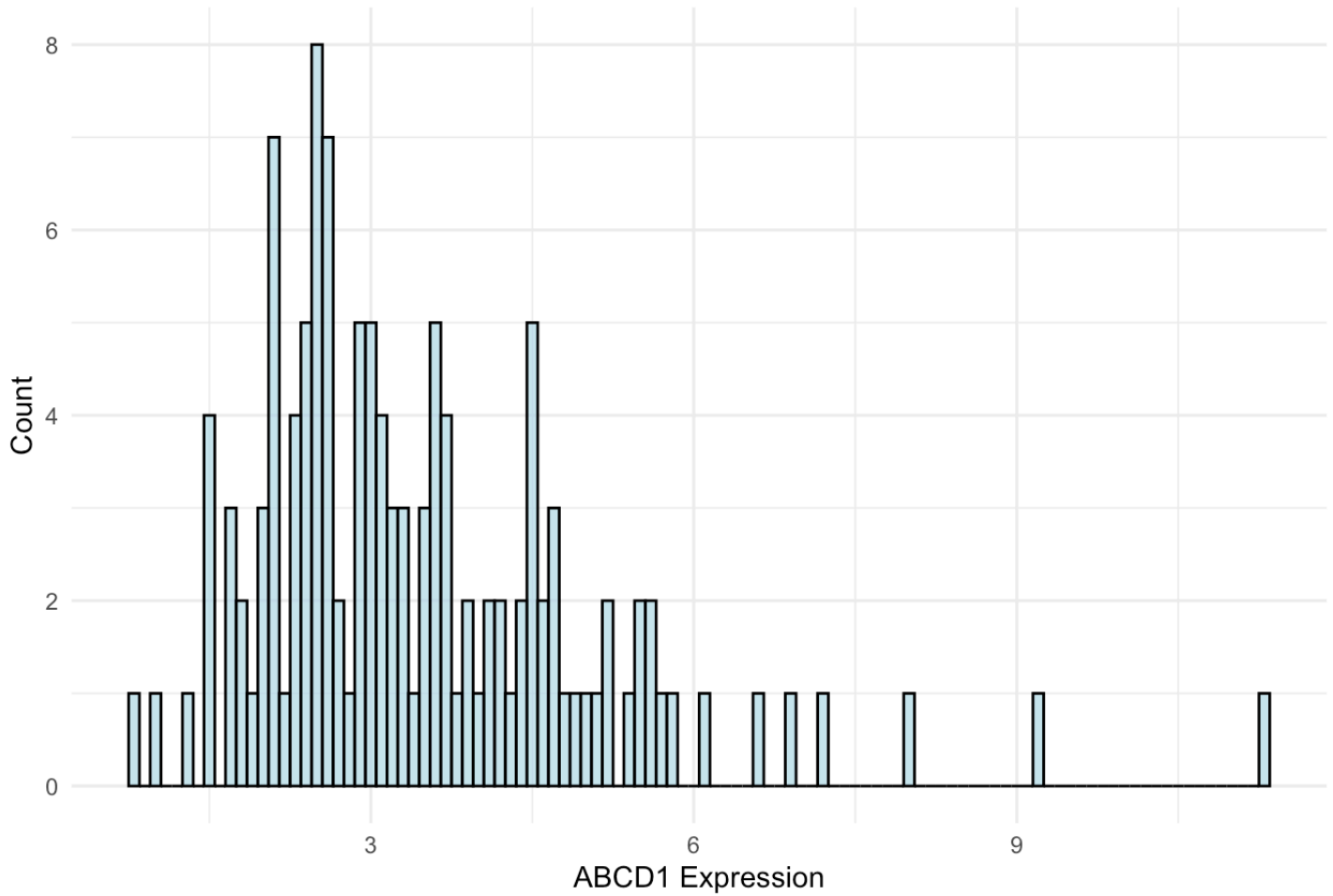


Boxplot of AASS by sex and icu_status



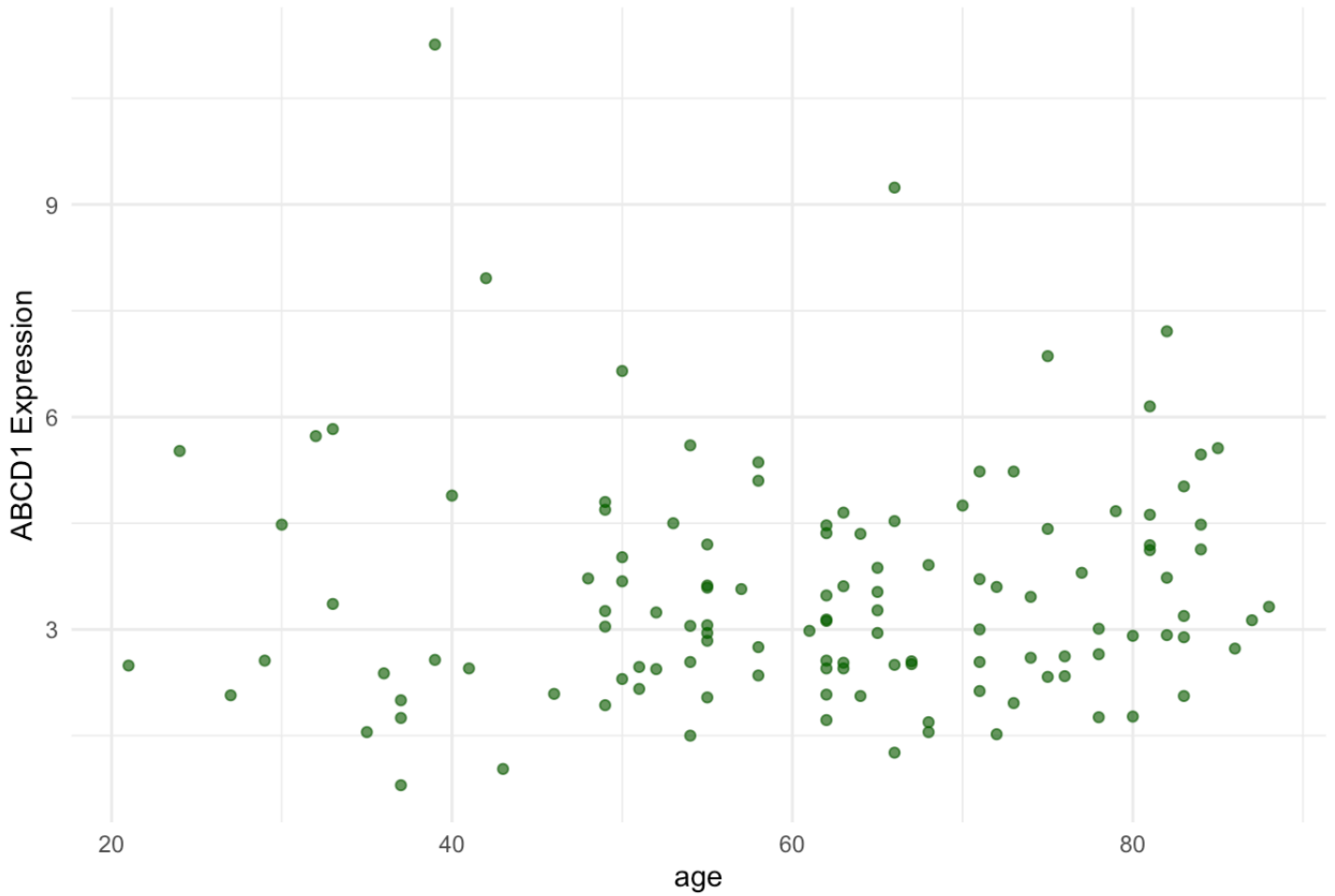


Histogram of ABCD1 Expression

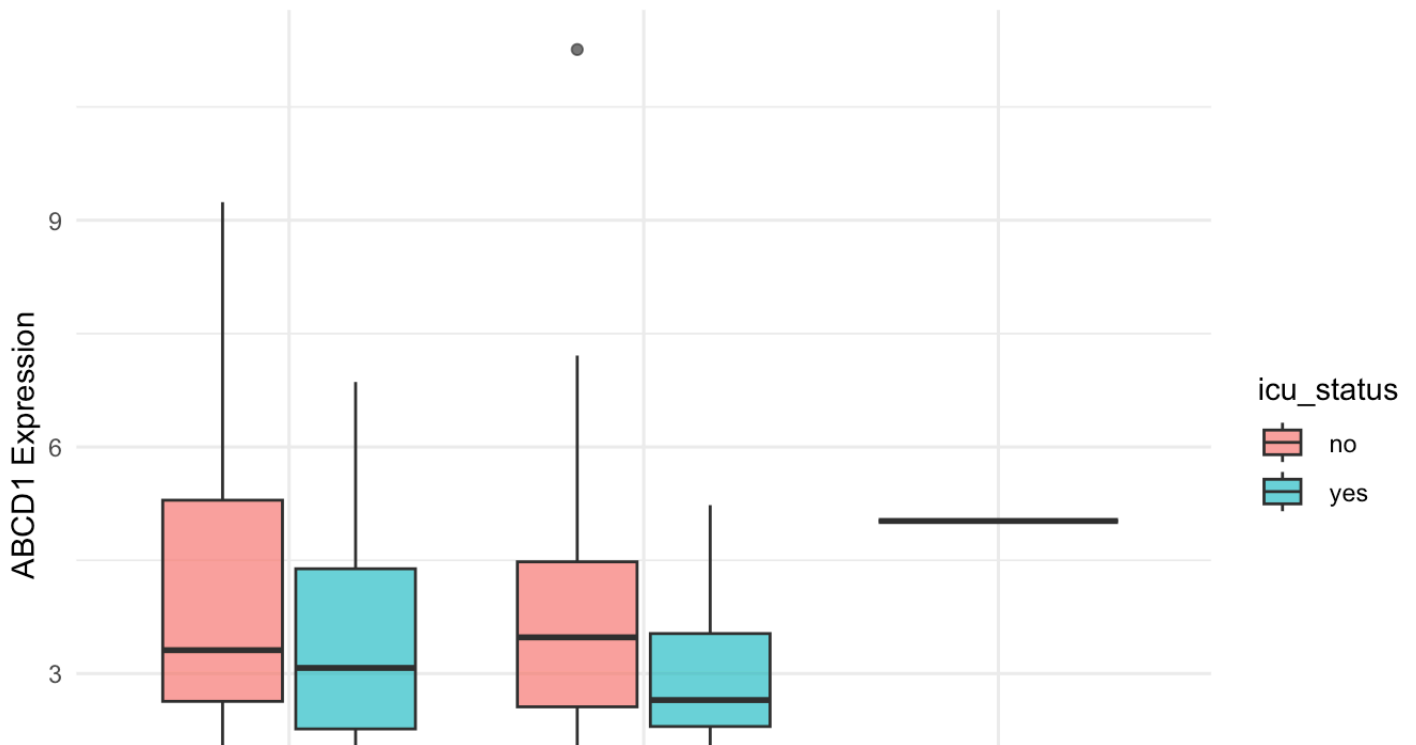


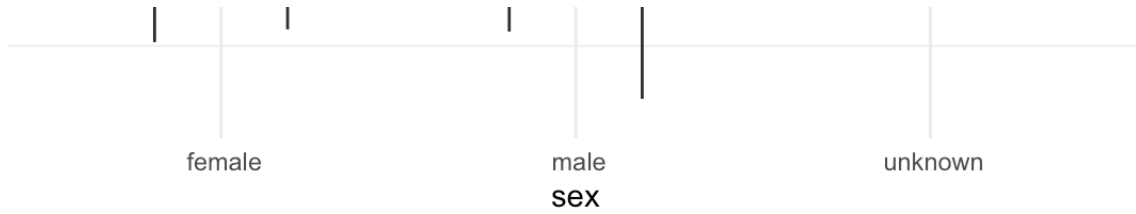
```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```


Scatterplot of ABCD1 vs age

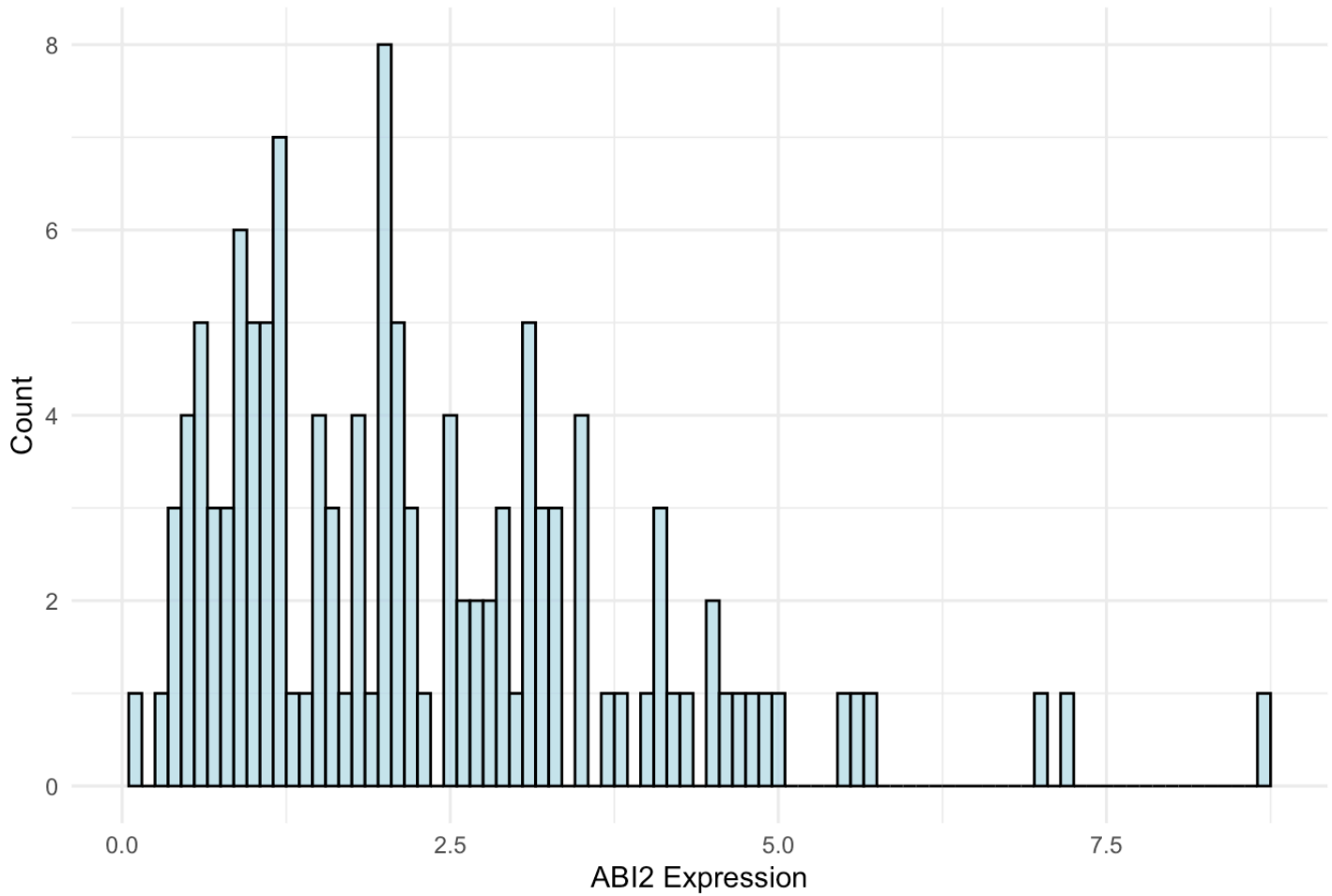


Boxplot of ABCD1 by sex and icu_status



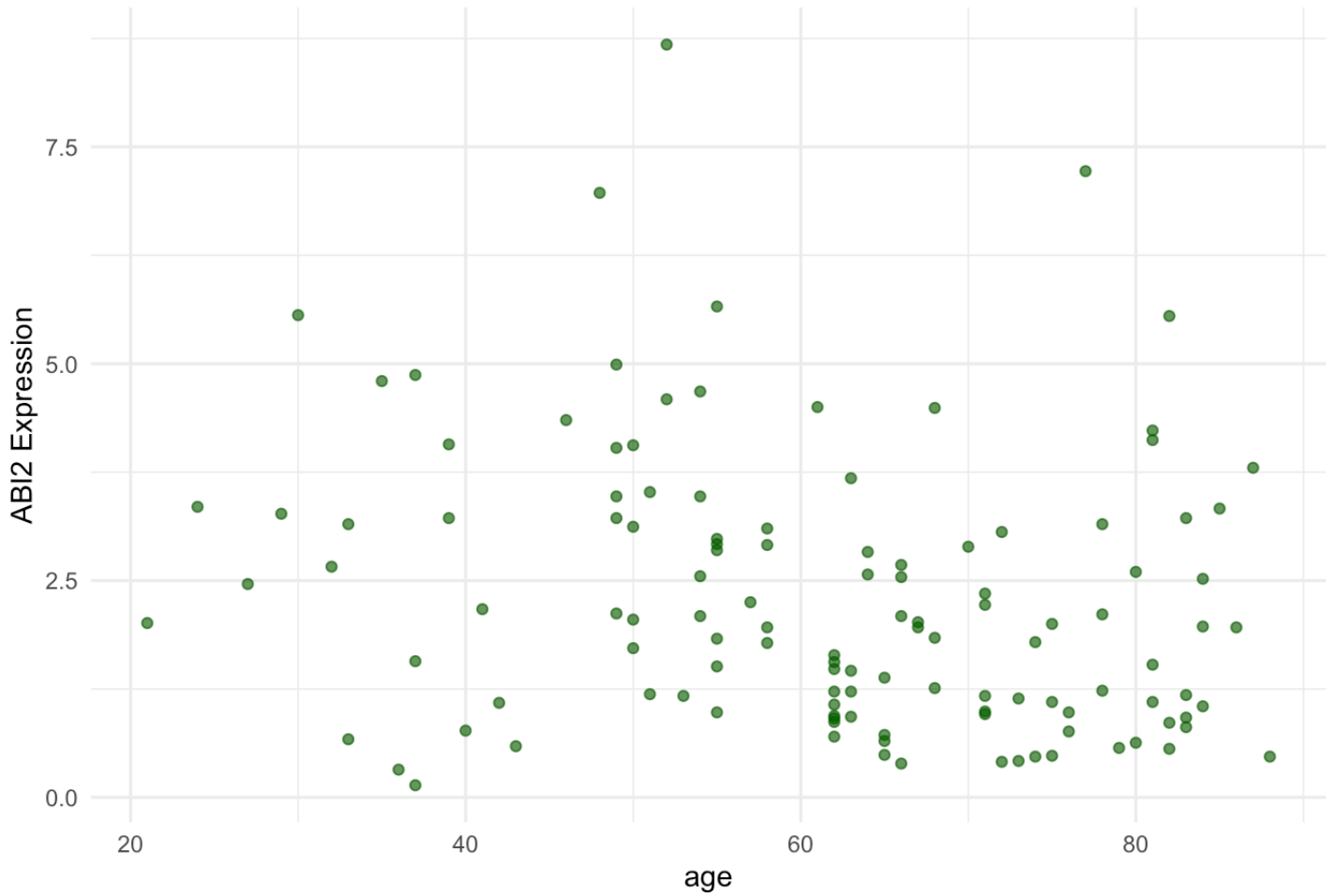


Histogram of ABI2 Expression



```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Scatterplot of ABI2 vs age



Boxplot of ABI2 by sex and icu_status

