

R_final_v2

2025-07-19

import

```
library(ggplot2) #import lib
genes <- read.csv("~/Desktop/QBS103_GSE157103_genes.csv", row.names = 1)
metadata <- read.csv("~/Desktop/QBS103_GSE157103_series_matrix-1.csv")
```

to make 3 plots

```
plot_all_figures <- function(metadata, gene_name, cont_var, cat_var1, cat_var2) {

  temp <- metadata # make a copy
  temp[[gene_name]] <- as.numeric(genes[gene_name, ])

  # remove row where age is : and convert >89 to 90
  #trim whitespace to be able to find :
  #from https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/trimws
  metadata <- metadata[trimws(metadata[[cont_var]]) != ":", ]
  temp[[cont_var]][temp[[cont_var]] == ">89"] <- "90"

  # convert age to numeric
  temp[[cont_var]] <- as.numeric(temp[[cont_var]])

  # filter out rows with missing values
  metadata_clean <- temp[!(is.na(temp[[cont_var]]) | is.na(temp[[gene_name]])), ]
  metadata_clean <- metadata_clean[metadata_clean[[cat_var1]] != " unknown", ]

  #histogram for gene expression
  print(
    ggplot(metadata_clean, aes_string(x = gene_name)) +
      geom_histogram(bins = 50, fill = "pink", color = "black") +
      scale_x_continuous(breaks = seq(0, 190, by = 10)) +
      scale_y_continuous(breaks = seq(0, 10, by = 1)) +
      labs(title = paste("Histogram of", gene_name, "Expression"),
           x = paste(gene_name, "Expression Level"),
           y = "Frequency")
  )

  #scatterplot
  print(
    ggplot(metadata_clean, aes_string(x = cont_var, y = gene_name)) +
      scale_x_continuous(breaks = seq(10, 100, by = 10)) +
      scale_y_continuous(breaks = seq(0, 200, by = 10)) +
```

```

    geom_point(size = 2, colour = "pink", alpha = 0.8) +
    labs(title = paste(gene_name, "Expression vs.", cont_var),
         x = "Age (years)",
         y = paste(gene_name, "Expression Level"))
  )

  #boxplot
  print(
    ggplot(metadata_clean, aes_string(x = cat_var2, y = gene_name, fill = cat_var1)) +
    geom_boxplot() +
    scale_fill_manual(values = c("female" = "pink", "male" = "deeppink")) +
    labs(title = paste(gene_name, "Expression by", cat_var2, "and", cat_var1),
         x = "ICU Status",
         y = paste(gene_name, "Expression Level"),
         fill = "Sex") +
    scale_y_continuous(breaks = seq(0, 200, by = 20))
  )
}

```

plot genes

```

genes_to_plot <- c("ABHD4", "ABHD5", "ABHD6")

for (gene in genes_to_plot) {
  plot_all_figures(metadata, gene_name = gene, cont_var = "age", cat_var1 = "sex", cat_var2 = "icu_status")
}

```

```

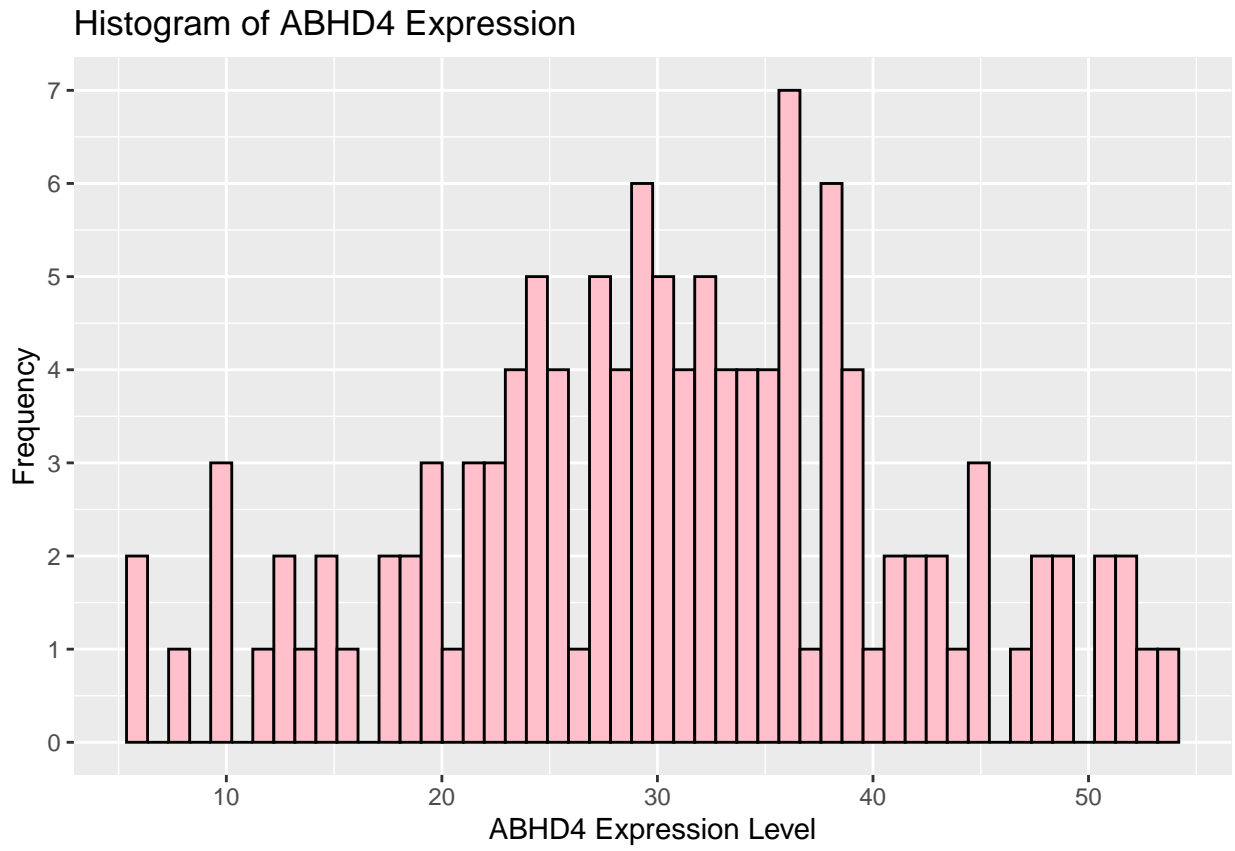
## Warning in plot_all_figures(metadata, gene_name = gene, cont_var = "age", : NAs
## introduced by coercion

```

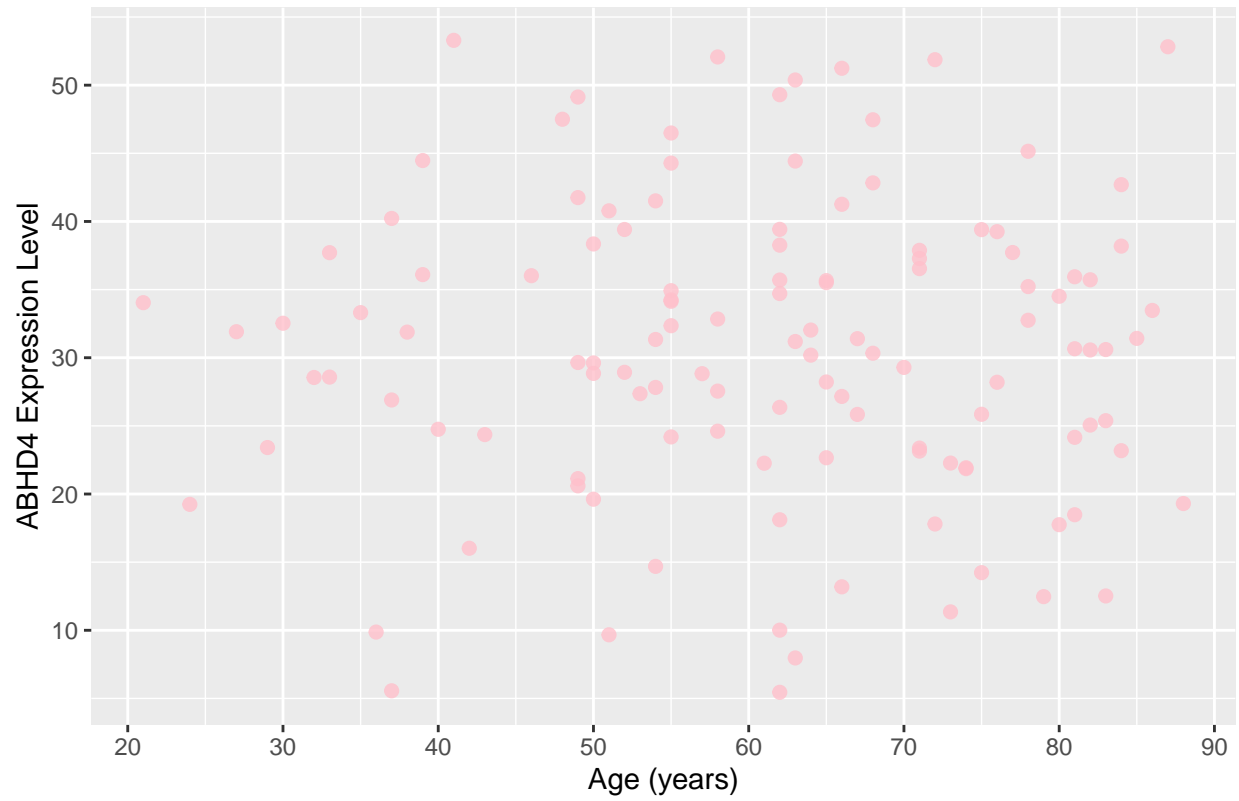
```

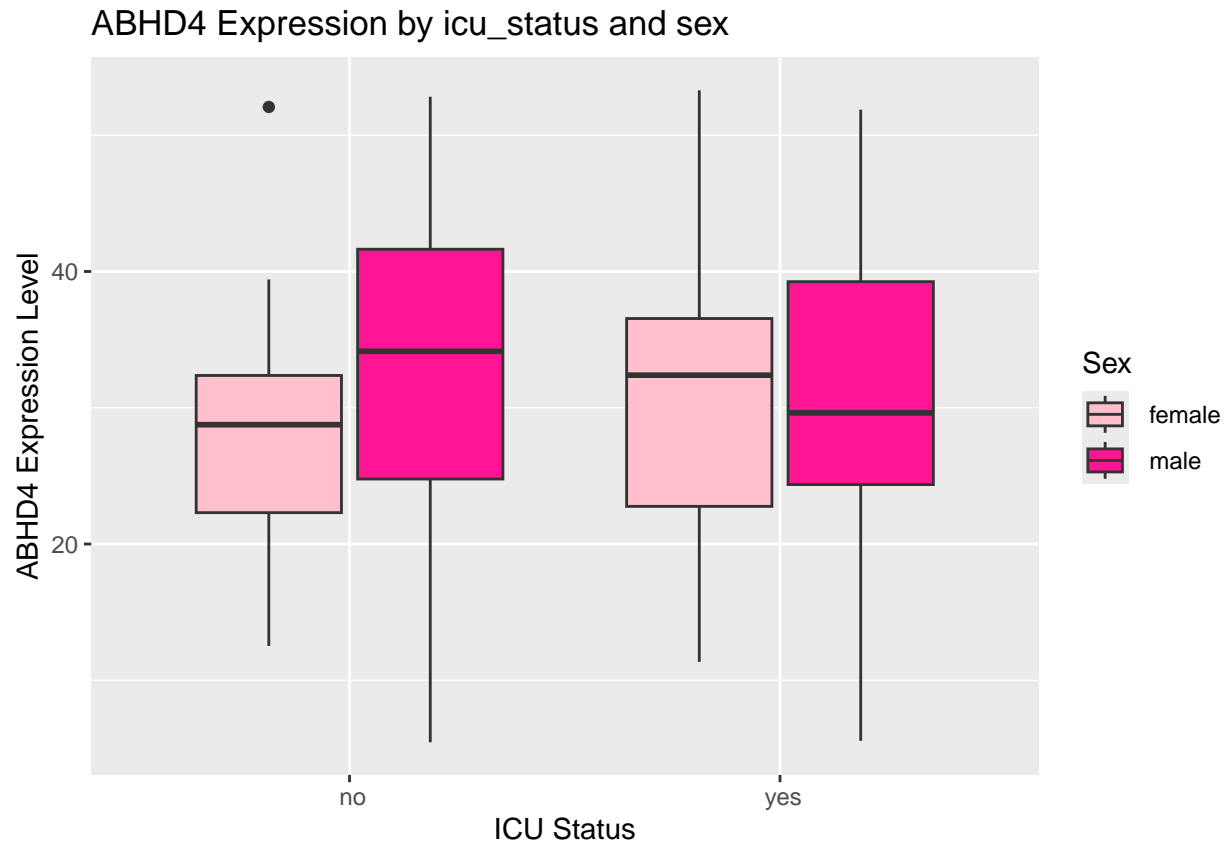
## Warning: 'aes_string()' was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with 'aes()'.
## i See also 'vignette("ggplot2-in-packages")' for more information.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning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

```

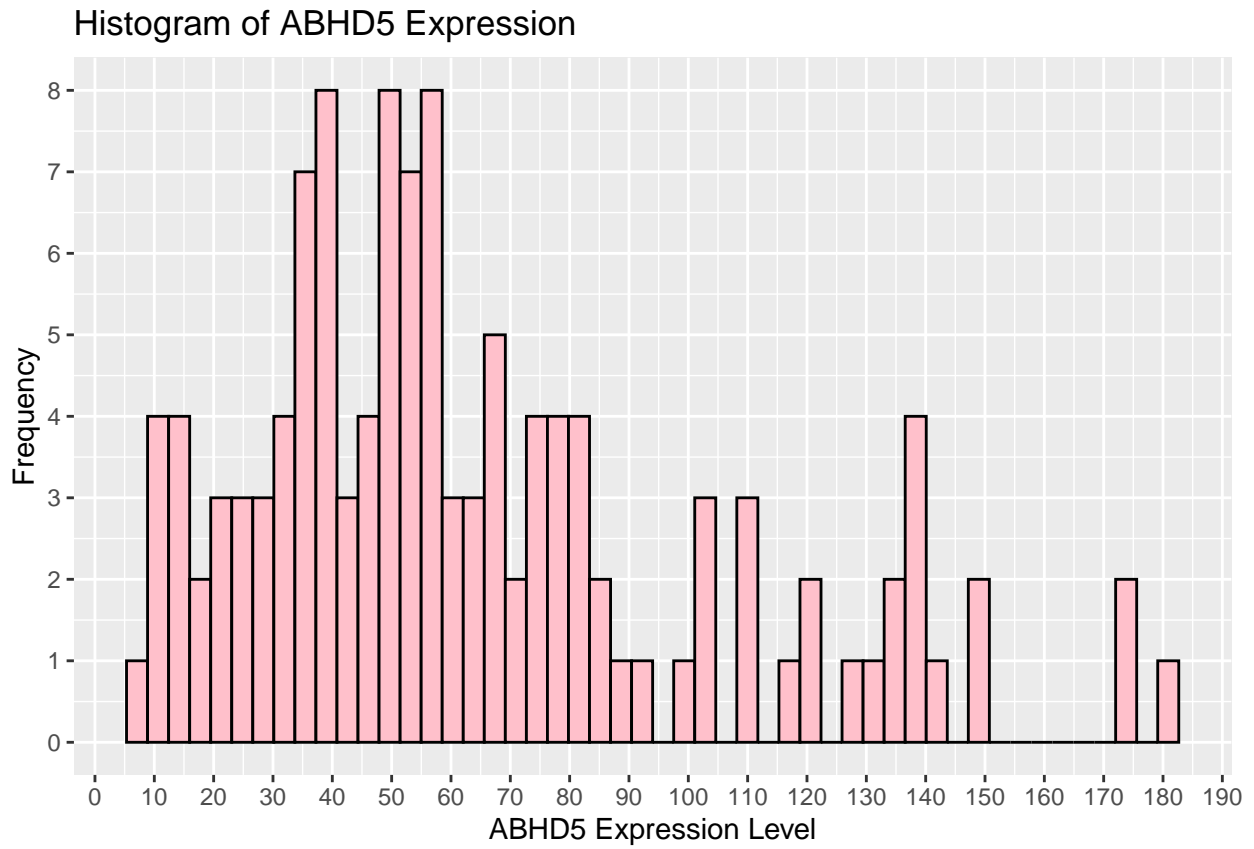


ABHD4 Expression vs. age

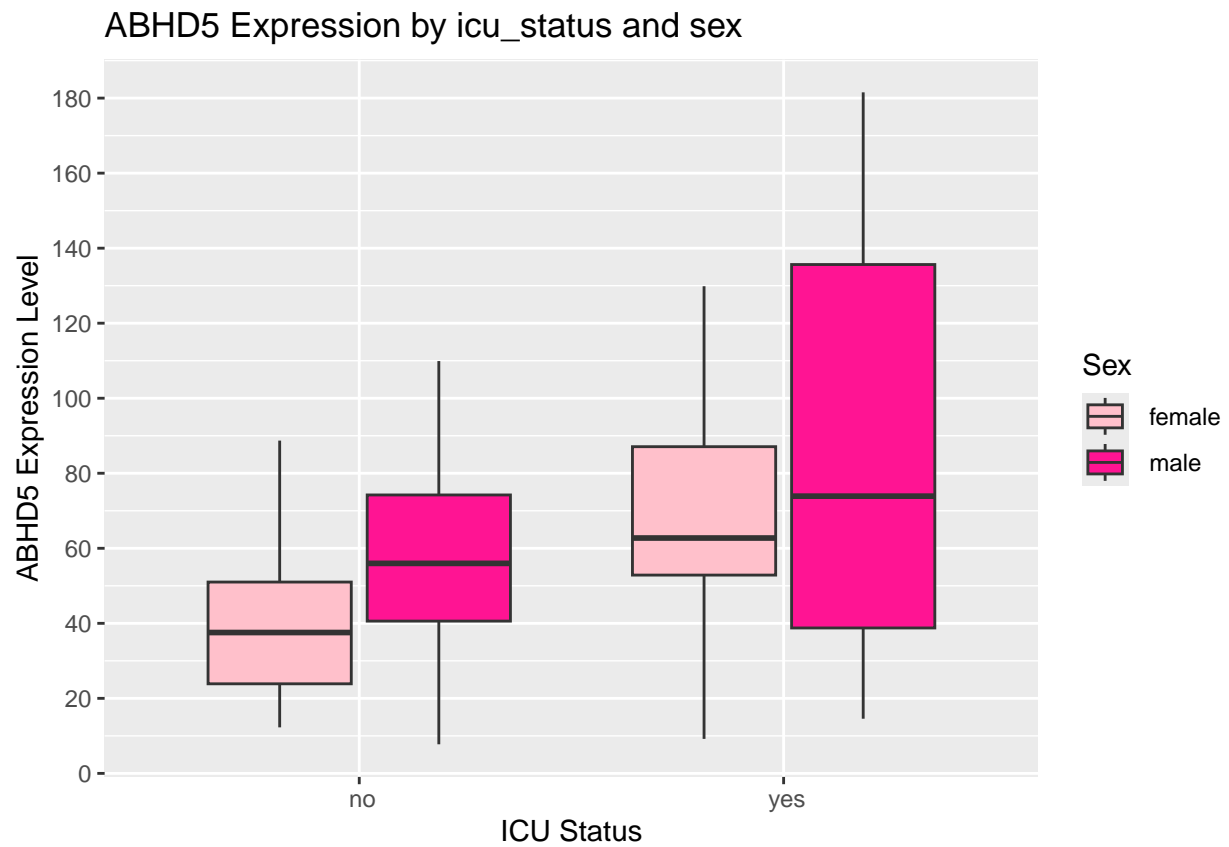




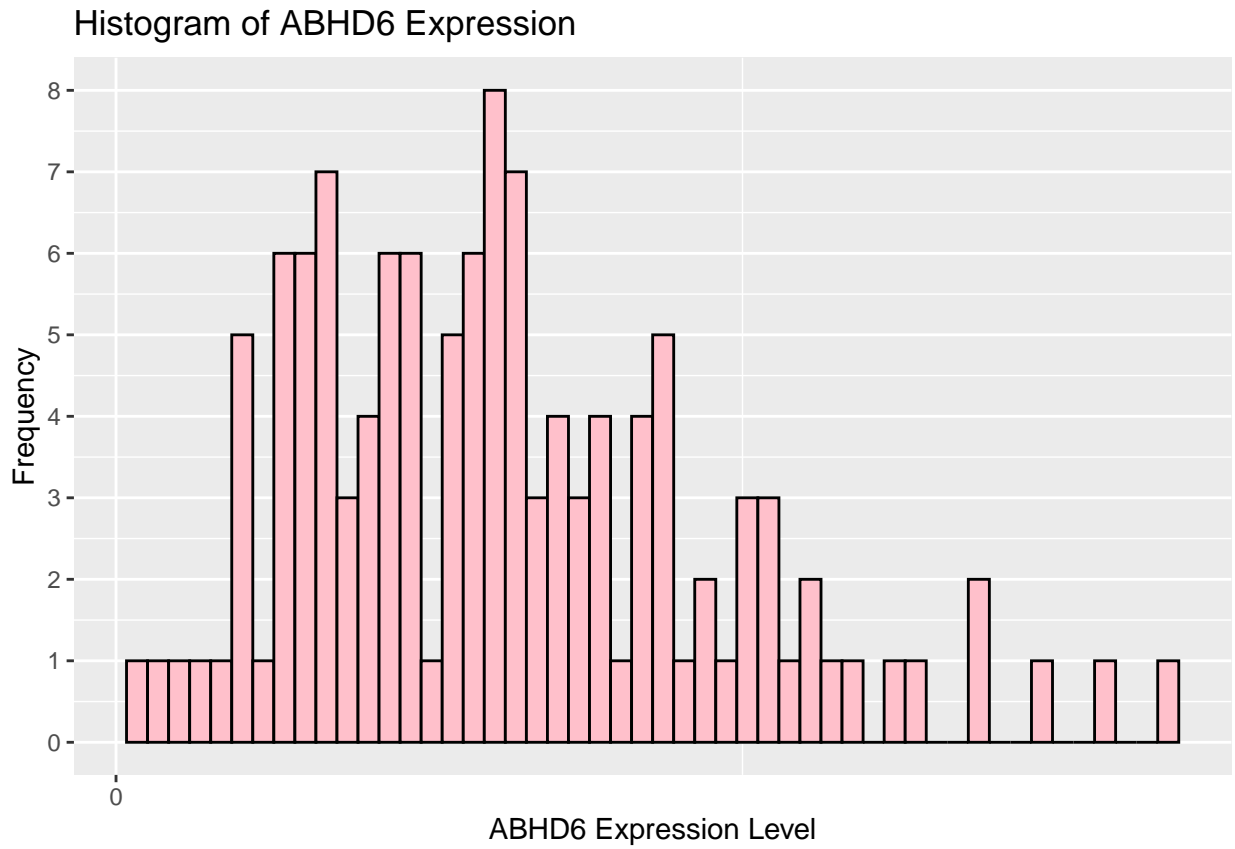
```
## Warning in plot_all_figures(metadata, gene_name = gene, cont_var = "age", : NAs  
## introduced by coercion
```



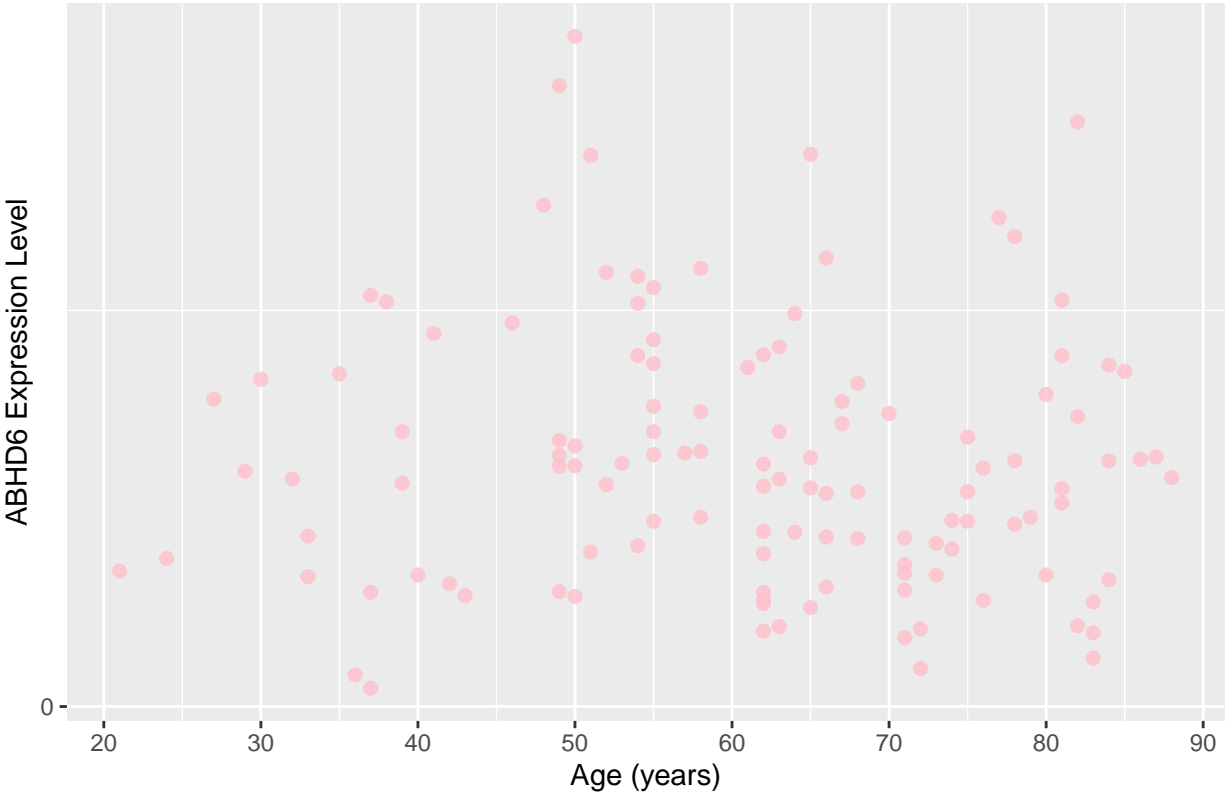




```
## Warning in plot_all_figures(metadata, gene_name = gene, cont_var = "age", : NAs  
## introduced by coercion
```

ABHD6 Expression vs. age



ABHD6 Expression by icu_status and sex

