Final Project Submission 2

Vui Doan

2025-07-14

```r
#############################################
# Reading gene expression data with read.csv2. #
#############################################

# The field separator is set to comma.

# The argument row.names is set to 1 so that when the file is read the first column as row.names of the

genes_data <- read.csv2(file = paste0("C:/Users/Jade/Desktop/",
"Dartmouth Classes/R QBS 103/GitRepo/Repository_One/",
"Gene_expression_QBS103_GSE157103_genes.csv"),
sep= ",",
row.names = 1,dec = ".")

# read.scv2 will load the data into a data.frame

covariates_data <- read.csv2(file = paste0("C:/Users/Jade/Desktop/",
"Dartmouth Classes/R QBS 103/GitRepo/Repository_One/",
"Covariates_QBS103_GSE157103_series_matrix-1.csv"),
sep= ",",
row.names = 1)

# Selecting covariates

slected_covariates <- c("hospital.free_days_post_45_day_followup",
"disease_status","sex")

covariates_data_filt <- covariates_data[,
slected_covariates]

selected_gene_tran <- data.frame(t(as.matrix(genes_data)))

data <- data.frame(covariates_data_filt,selected_gene_tran)

colnames(selected_gene_tran)[c(1:3)]
```

```
## [1] "A1BG" "A1CF" "A2M"
```

```r
selected_genes <- c("A1BG","A1CF","A2M" )

covariate_cont <- "hospital.free_days_post_45_day_followup"
```

```r
covariates_cat <- c("disease_status","sex")

function_vui <- function(d_f,list_of_genes,cont_cov,list_cat_cov){

  for (gene in list_of_genes){

    # Plotting histograms

    plot <- ggplot(d_f, aes(x = !!sym(gene))) + geom_histogram() +
    labs(title = paste("Histogram for gene: ", gene, sep =""))
    print(plot)

    # Plotting scatterplot.

    plot_2 <- ggplot(data, aes(y = !!sym(gene),
    x = hospital.free_days_post_45_day_followup)) +
    geom_point() +
    ggtitle(paste("Expression of ",
    gene,"by Hospital Free Days", sep = ""))

    print(plot_2)

    # Plotting boxplots.

    for (cat_cov in list_cat_cov){

        plot_3 <- ggplot(data, aes(x = !!sym(cat_cov) , y = !!sym(gene))) +
        geom_boxplot() + ggtitle(paste("Expression of ",
        gene,"by ", cat_cov, sep = ""))

        print(plot_3)

    }

  }

}

library(ggplot2)

function_vui(data,selected_genes,covariate_cont,covariates_cat)
```
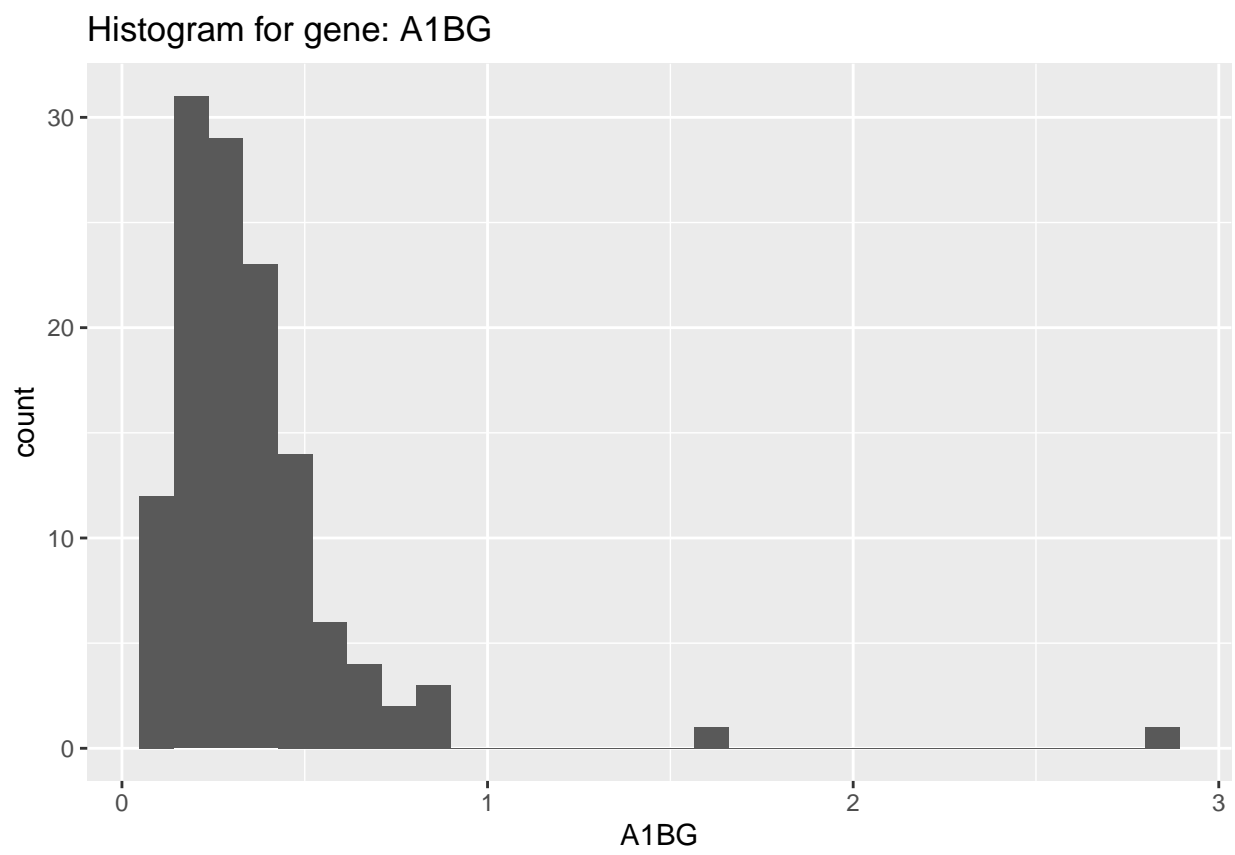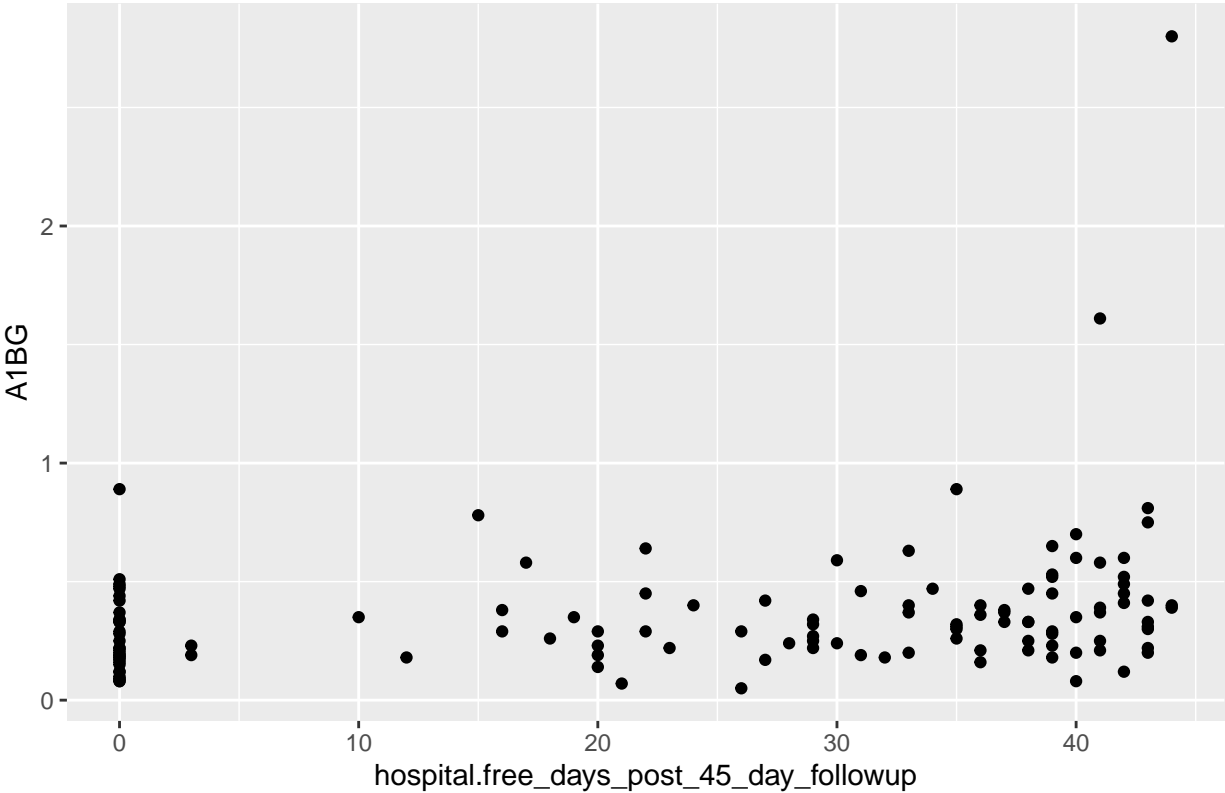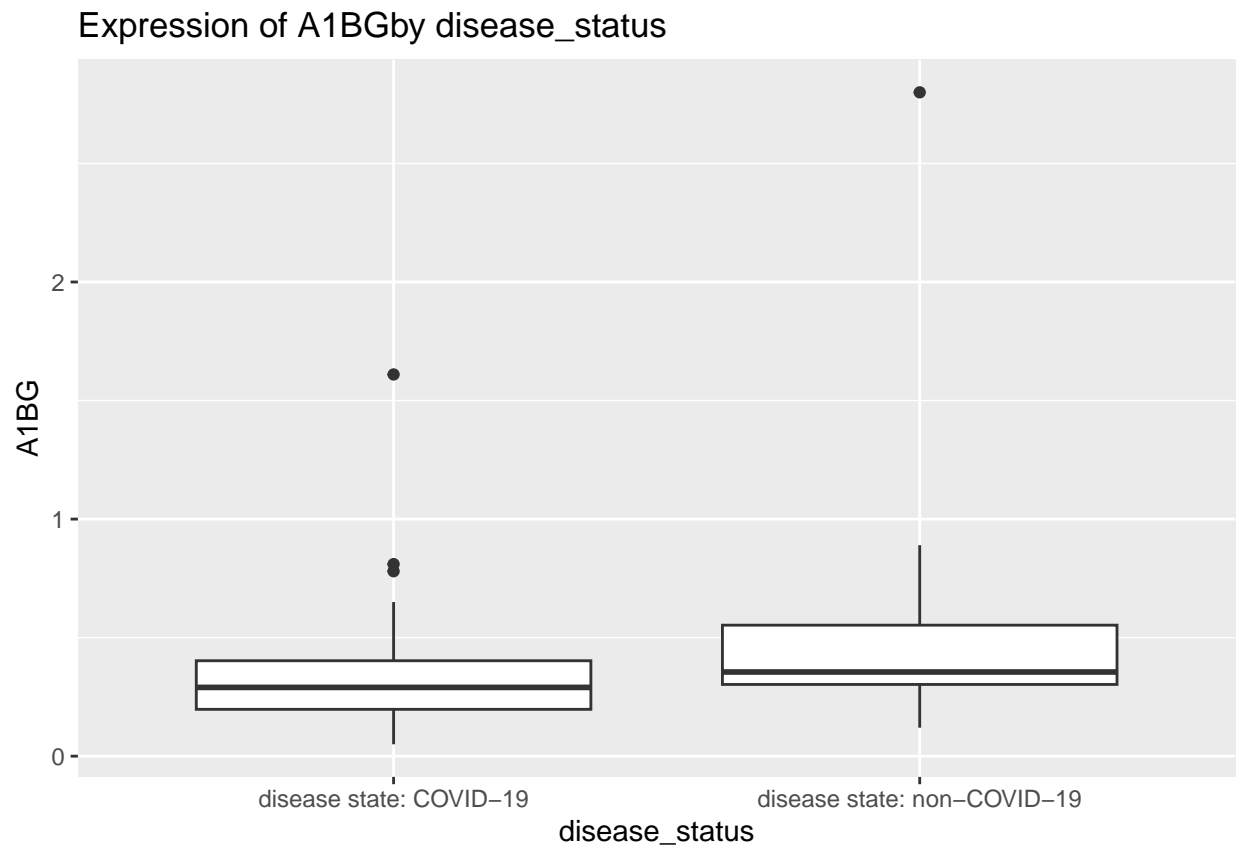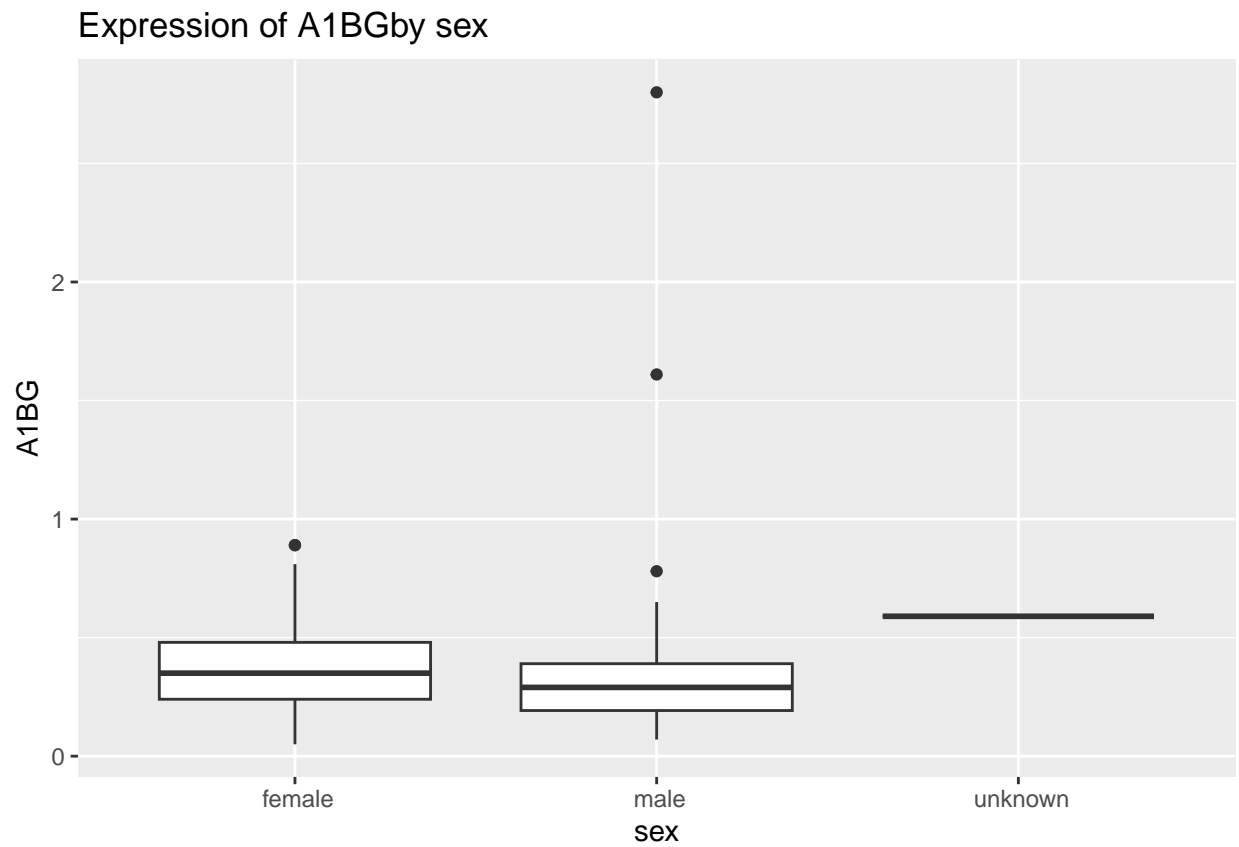
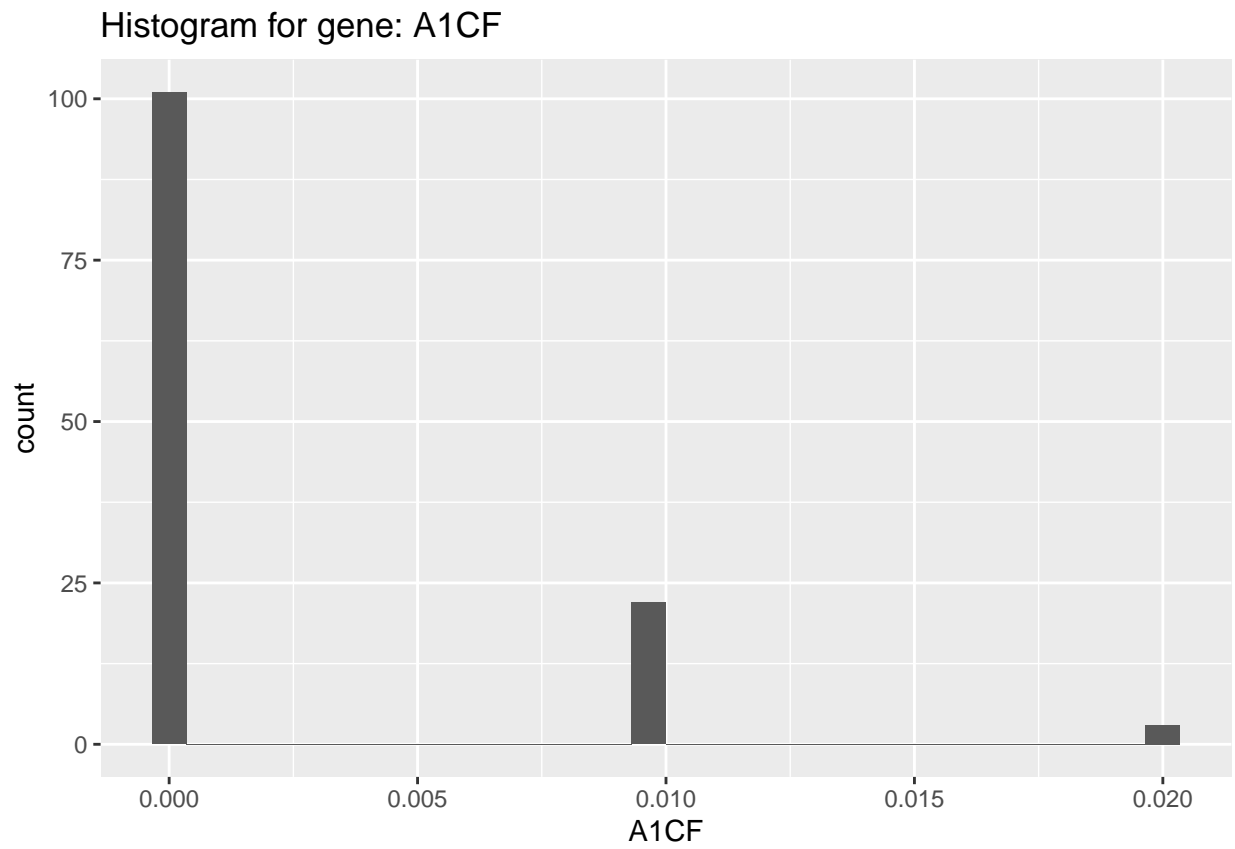## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Histogram for gene: A1BG

Expression of A1BGby Hospital Free Days

Expression of A1BGby disease_status

Expression of A1BGby sex

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Histogram for gene: A1CF

Expression of A1CFby Hospital Free Days

Expression of A1CFby disease_status
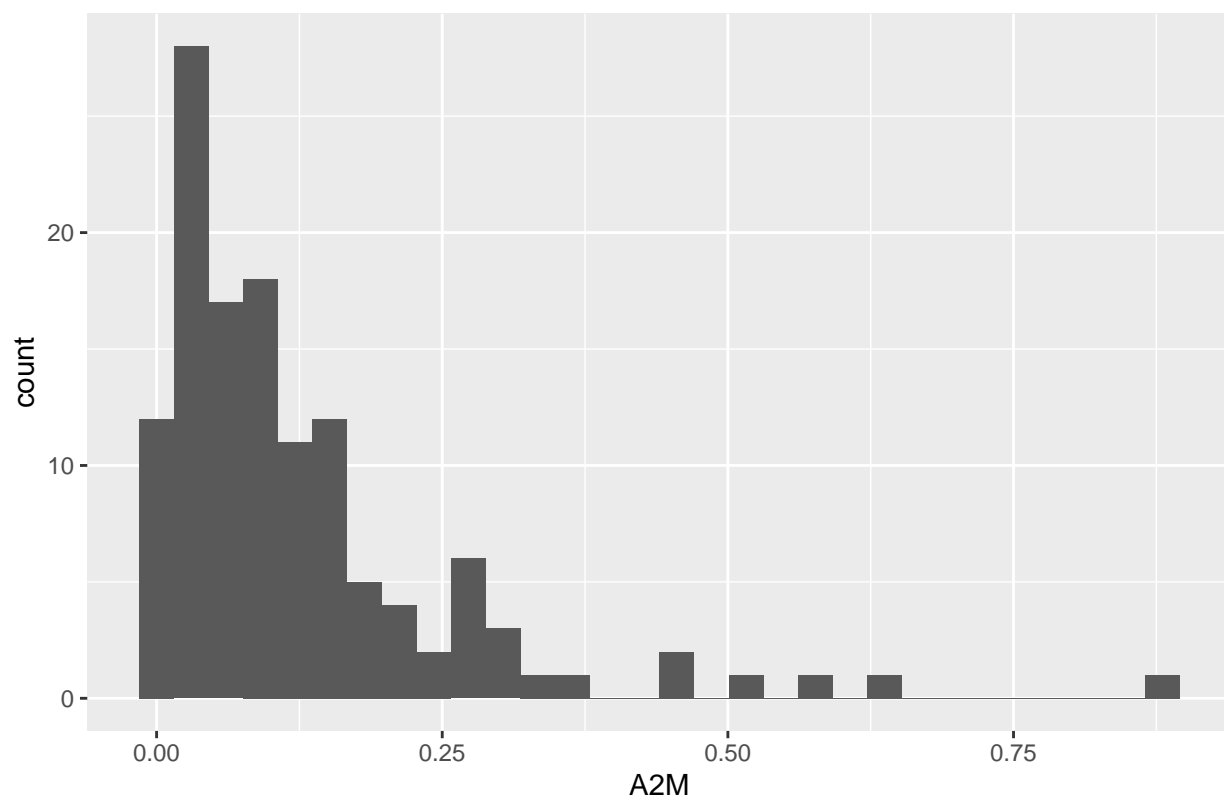
Expression of A1CFby sex

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
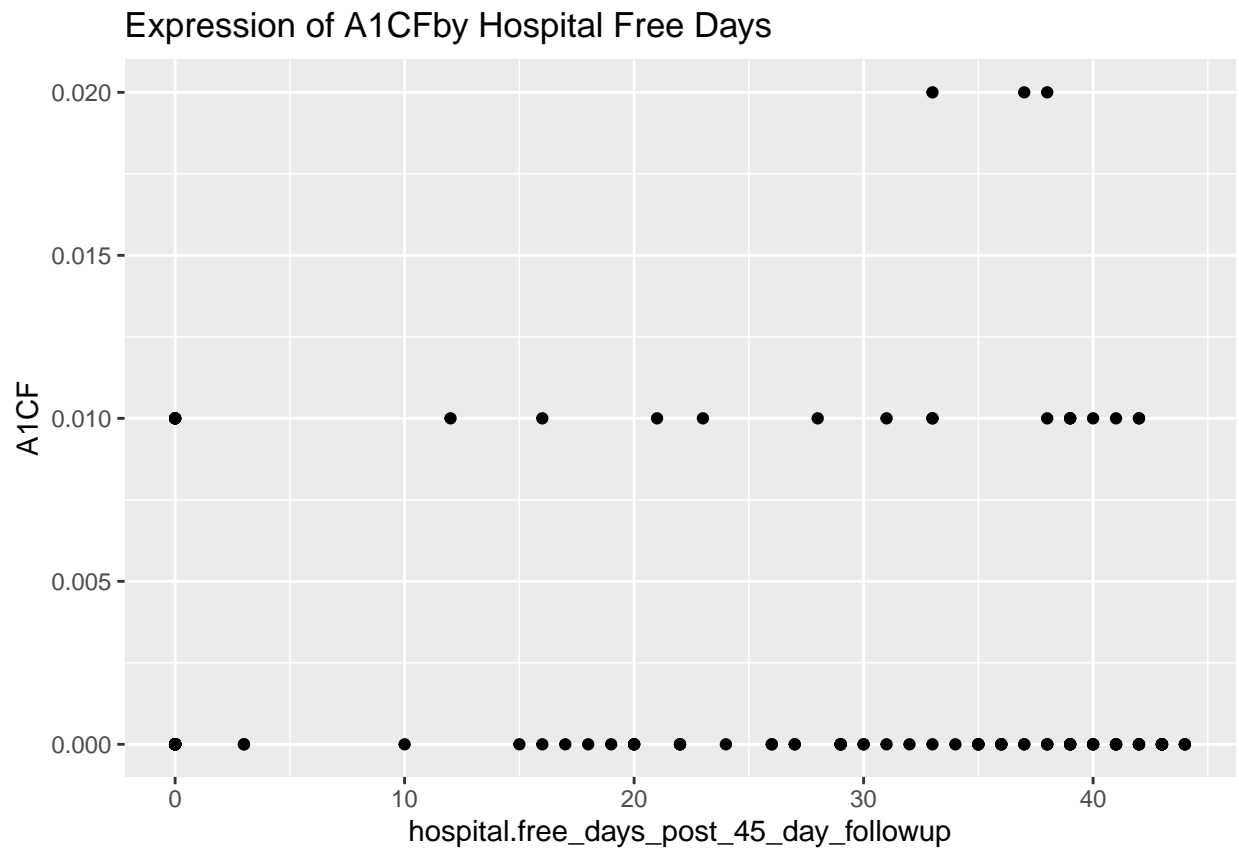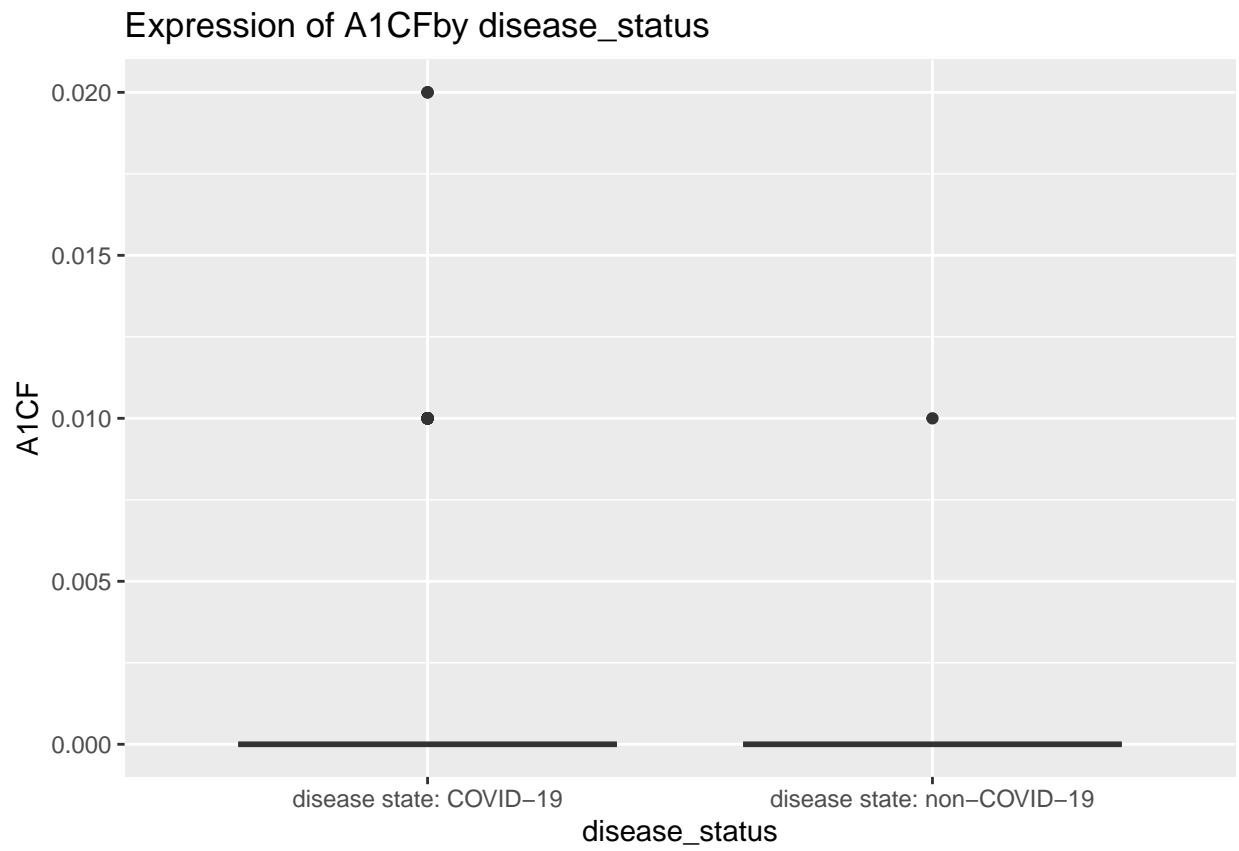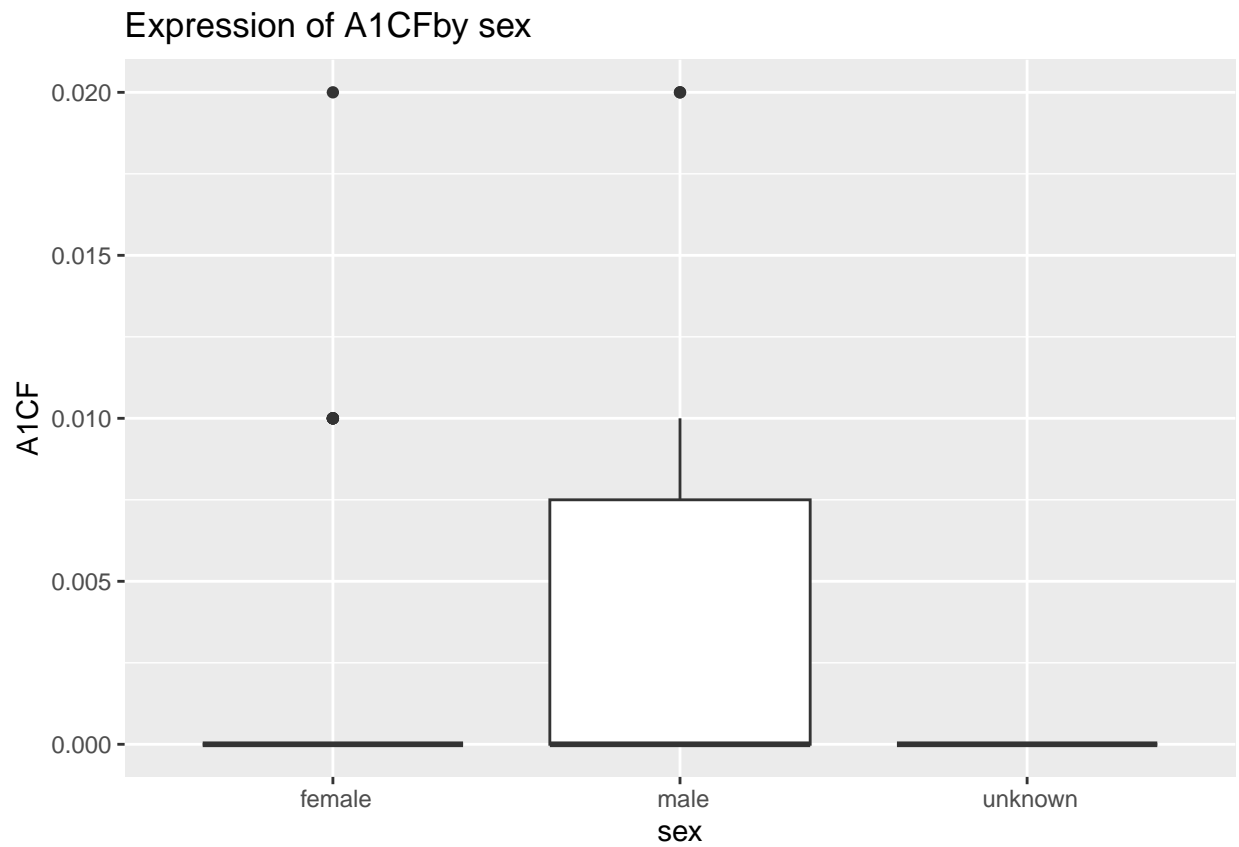
Histogram for gene: A2M

Expression of A2Mby Hospital Free Days

Expression of A2Mby disease_status

Expression of A2Mby sex