Submission 1_PBX

2025-07-13

Install packages & read data & local setting

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
#set path
setwd('/Users/stanpawn/Desktop/Data science 101/Submission 1')
#read data
metadata <- read.csv("QBS103_GSE157103_series_matrix-1.csv")
gene_exp_raw <- read.csv("QBS103_GSE157103_genes.csv")
#check data
glimpse(metadata)
```

```
## Rows: 126
## Columns: 25
## $ participant_id                 <chr> "COVID_01_39y_male_NonICU", "C~
## $ geo_accession                  <chr> "GSM4753021", "GSM4753022", "G~
## $ status                         <chr> "Public on Aug 29 2020", "Publ~
## $ X.Sample_submission_date       <chr> "Aug 28 2020", "Aug 28 2020", ~
## $ last_update_date               <chr> "Aug 29 2020", "Aug 29 2020", ~
## $ type                           <chr> "SRA", "SRA", "SRA", "SRA", "S~
## $ channel_count                  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ source_name_ch1                <chr> "Leukocytes from whole blood",~
## $ organism_ch1                   <chr> "Homo sapiens", "Homo sapiens"~
## $ disease_status                 <chr> "disease state: COVID-19", "di~
## $ age                            <chr> "39", "63", "33", "49", "49", ~
## $ sex                            <chr> " male", " male", " male", " m~
## $ icu_status                     <chr> " no", " no", " no", " no", " ~
## $ apacheii                       <chr> "15", " unknown", " unknown", ~
## $ charlson_score                 <int> 0, 2, 2, 1, 1, 1, 7, 7, 2, 1, ~
## $ mechanical_ventilation         <chr> " yes", " no", " no", " no", "~
## $ ventilator.free_days           <int> 0, 28, 28, 28, 23, 28, 28, 0, ~
## $ hospital.free_days_post_45_day_followup <int> 0, 39, 18, 39, 27, 36, 42, 0, ~
```

```
## $ ferritin.ng.ml.                        <chr> "946", "1060", "1335", "583", ~
## $ crp.mg.l.                               <chr> "73.1", " unknown", "53.2", "2~
## $ ddimer.mg.l_feu.                        <chr> "1.3", "1.03", "1.48", "1.32",~
## $ procalcitonin.ng.ml..                   <chr> "36", "0.37", "0.07", "0.98", ~
## $ lactate.mmol.l.                         <chr> "0.9", " unknown", " unknown",~
## $ fibrinogen                              <chr> "513", "unknown", "513", "949"~
## $ sofa                                    <chr> "8", " unknown", " unknown", "~
```

```
glimpse(gene_exp_raw)
```

```
## Rows: 100
## Columns: 127
## $ X                            <chr> "A1BG", "A1CF", "A2M", "A2ML1", "A3GALT2~
## $ COVID_01_39y_male_NonICU     <dbl> 0.49, 0.00, 0.21, 0.04, 0.07, 0.00, 0.03~
## $ COVID_02_63y_male_NonICU     <dbl> 0.29, 0.00, 0.14, 0.00, 0.00, 0.00, 0.05~
## $ COVID_03_33y_male_NonICU     <dbl> 0.26, 0.00, 0.03, 0.02, 0.00, 0.00, 0.07~
## $ COVID_04_49y_male_NonICU     <dbl> 0.45, 0.01, 0.09, 0.07, 0.00, 0.00, 0.00~
## $ COVID_05_49y_male_NonICU     <dbl> 0.17, 0.00, 0.00, 0.05, 0.07, 0.00, 0.00~
## $ COVID_06_.y_male_NonICU      <dbl> 0.21, 0.00, 0.08, 0.04, 0.00, 0.00, 0.03~
## $ COVID_07_38y_female_NonICU   <dbl> 0.49, 0.01, 0.23, 0.03, 0.07, 0.00, 0.00~
## $ COVID_08_78y_male_ICU        <dbl> 0.12, 0.00, 0.08, 0.01, 0.00, 0.00, 0.00~
## $ COVID_09_64y_female_ICU      <dbl> 0.51, 0.01, 0.88, 0.02, 0.79, 0.00, 0.00~
## $ COVID_10_62y_male_ICU        <dbl> 0.10, 0.00, 0.13, 0.01, 0.15, 0.00, 0.12~
## $ COVID_11_52y_female_NonICU   <dbl> 0.38, 0.02, 0.47, 0.03, 0.08, 0.00, 0.00~
## $ COVID_12_50y_male_ICU        <dbl> 0.45, 0.00, 0.16, 0.00, 1.75, 0.00, 0.00~
## $ COVID_13_37y_male_NonICU     <dbl> 0.18, 0.00, 0.07, 0.01, 0.00, 0.00, 0.00~
## $ COVID_14_55y_male_ICU        <dbl> 0.23, 0.00, 0.22, 0.04, 0.93, 0.00, 0.07~
## $ COVID_15_68y_male_ICU        <dbl> 0.42, 0.00, 0.07, 0.00, 0.15, 0.03, 0.00~
## $ COVID_16_48y_male_NonICU     <dbl> 0.41, 0.01, 0.58, 0.00, 0.19, 0.00, 0.00~
## $ COVID_17_54y_male_NonICU     <dbl> 0.63, 0.02, 0.15, 0.02, 0.00, 0.00, 0.05~
## $ COVID_18_70y_female_NonICU   <dbl> 0.47, 0.00, 0.30, 0.02, 0.06, 0.03, 0.03~
## $ COVID_19_51y_male_NonICU     <dbl> 0.33, 0.02, 0.11, 0.02, 0.00, 0.00, 0.04~
## $ COVID_20_62y_male_ICU        <dbl> 0.32, 0.00, 0.07, 0.00, 0.22, 0.00, 0.00~
## $ COVID_21_66y_male_ICU        <dbl> 0.18, 0.00, 0.00, 0.00, 0.37, 0.03, 0.00~
## $ COVID_22_43y_male_ICU        <dbl> 0.09, 0.00, 0.06, 0.00, 0.06, 0.00, 0.06~
## $ COVID_23_76y_male_ICU        <dbl> 0.18, 0.01, 0.03, 0.00, 0.07, 0.03, 0.04~
## $ COVID_24_55y_male_ICU        <dbl> 0.22, 0.01, 0.11, 0.02, 0.15, 0.00, 0.00~
## $ COVID_25_55y_male_ICU        <dbl> 0.29, 0.00, 0.09, 0.03, 0.00, 0.00, 0.06~
## $ COVID_26_41y_female_ICU      <dbl> 0.42, 0.00, 0.18, 0.00, 0.87, 0.00, 0.00~
## $ COVID_27_71y_female_ICU      <dbl> 0.16, 0.01, 0.23, 0.01, 0.18, 0.00, 0.00~
## $ COVID_28_63y_male_ICU        <dbl> 0.18, 0.00, 0.18, 0.05, 0.45, 0.00, 0.00~
## $ COVID_29_63y_female_ICU      <dbl> 0.35, 0.00, 0.03, 0.03, 0.15, 0.03, 0.08~
## $ COVID_30_54y_male_ICU        <dbl> 0.23, 0.00, 0.11, 0.01, 0.00, 0.00, 0.03~
## $ COVID_31_50y_male_ICU        <dbl> 0.15, 0.00, 0.47, 0.00, 0.00, 0.03, 0.00~
## $ COVID_32_72y_male_ICU        <dbl> 0.34, 0.01, 0.04, 0.00, 0.29, 0.00, 0.04~
## $ COVID_33_81y_male_NonICU     <dbl> 0.35, 0.00, 0.30, 0.06, 0.26, 0.00, 0.00~
## $ COVID_34_64y_female_NonICU   <dbl> 0.36, 0.00, 0.11, 0.00, 0.12, 0.00, 0.00~
## $ COVID_35_58y_female_NonICU   <dbl> 0.26, 0.00, 0.51, 0.02, 0.16, 0.00, 0.00~
## $ COVID_36_68y_male_NonICU     <dbl> 0.18, 0.01, 0.09, 0.00, 0.08, 0.00, 0.00~
## $ COVID_37_87y_male_NonICU     <dbl> 0.20, 0.00, 0.09, 0.07, 0.31, 0.00, 0.00~
## $ COVID_38_68y_male_ICU        <dbl> 0.29, 0.00, 0.10, 0.02, 0.35, 0.00, 0.00~
## $ COVID_39_80y_female_ICU      <dbl> 0.19, 0.00, 0.27, 0.00, 0.00, 0.07, 0.00~
## $ COVID_40_66y_male_ICU        <dbl> 0.22, 0.00, 0.17, 0.00, 0.08, 0.00, 0.00~
## $ COVID_41_74y_male_ICU        <dbl> 0.19, 0.00, 0.14, 0.00, 0.19, 0.00, 0.00~
```

```
## $ COVID_42_21y_female_ICU       <dbl> 0.24, 0.01, 0.33, 0.01, 0.39, 0.00, 0.00~
## $ COVID_43_83y_female_ICU       <dbl> 0.29, 0.00, 0.00, 0.00, 0.11, 0.00, 0.00~
## $ COVID_44_46y_male_ICU         <dbl> 0.22, 0.00, 0.14, 0.00, 0.00, 0.04, 0.00~
## $ COVID_45_62y_female_ICU       <dbl> 0.14, 0.00, 0.15, 0.03, 0.19, 0.00, 0.00~
## $ COVID_46_62y_male_ICU         <dbl> 0.53, 0.01, 0.10, 0.00, 0.06, 0.00, 0.00~
## $ COVID_47_78y_male_ICU         <dbl> 0.08, 0.01, 0.04, 0.03, 0.60, 0.00, 0.00~
## $ COVID_48_72y_female_ICU       <dbl> 0.19, 0.00, 0.06, 0.01, 0.23, 0.06, 0.02~
## $ COVID_49_73y_male_ICU         <dbl> 0.48, 0.00, 0.09, 0.03, 0.00, 0.00, 0.00~
## $ COVID_50_37y_male_ICU         <dbl> 0.08, 0.00, 0.01, 0.00, 0.00, 0.72, 0.00~
## $ COVID_51_58y_female_NonICU    <dbl> 0.21, 0.00, 0.13, 0.00, 0.00, 0.00, 0.00~
## $ COVID_52_71y_male_NonICU      <dbl> 0.25, 0.01, 0.00, 0.03, 0.00, 0.00, 0.00~
## $ COVID_53_35y_female_NonICU    <dbl> 0.25, 0.00, 0.64, 0.10, 0.00, 0.00, 0.00~
## $ COVID_55_62y_female_ICU       <dbl> 0.09, 0.00, 0.09, 0.01, 0.00, 0.00, 0.03~
## $ COVID_56_33y_female_NonICU    <dbl> 0.28, 0.00, 0.16, 0.09, 0.23, 0.00, 0.00~
## $ COVID_57_30y_female_NonICU    <dbl> 0.42, 0.00, 0.27, 0.01, 0.19, 0.05, 0.00~
## $ COVID_58_62y_male_NonICU      <dbl> 0.39, 0.00, 0.08, 0.00, 0.00, 0.00, 0.00~
## $ COVID_59_55y_male_NonICU      <dbl> 0.33, 0.00, 0.10, 0.00, 0.07, 0.00, 0.00~
## $ COVID_60_49y_male_NonICU      <dbl> 0.22, 0.00, 0.14, 0.00, 0.00, 0.02, 0.00~
## $ COVID_61_54y_female_NonICU    <dbl> 0.25, 0.00, 0.10, 0.03, 0.13, 0.00, 0.00~
## $ COVID_62_78y_female_ICU       <dbl> 0.21, 0.00, 0.04, 0.00, 0.05, 0.00, 0.00~
## $ COVID_63_39y_female_ICU       <dbl> 0.29, 0.00, 0.01, 0.00, 0.14, 0.00, 0.00~
## $ COVID_64_65y_male_ICU         <dbl> 0.38, 0.01, 0.04, 0.02, 0.56, 0.00, 0.04~
## $ COVID_65_84y_male_NonICU      <dbl> 0.40, 0.01, 0.07, 0.00, 0.58, 0.00, 0.00~
## $ COVID_66_66y_female_NonICU    <dbl> 0.64, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00~
## $ COVID_67_57y_male_ICU         <dbl> 0.37, 0.00, 0.35, 0.00, 0.00, 0.00, 0.00~
## $ COVID_68_79y_male_ICU         <dbl> 0.58, 0.00, 0.15, 0.01, 0.00, 0.05, 0.00~
## $ COVID_69_77y_female_NonICU    <dbl> 0.52, 0.00, 0.29, 0.02, 0.00, 0.00, 0.00~
## $ COVID_70_81y_male_NonICU      <dbl> 0.27, 0.00, 0.07, 0.00, 0.00, 0.06, 0.00~
## $ COVID_71_37y_male_ICU         <dbl> 0.07, 0.01, 0.12, 0.01, 0.00, 0.00, 0.19~
## $ COVID_72_50y_female_NonICU    <dbl> 0.52, 0.00, 0.10, 0.01, 0.00, 0.00, 0.00~
## $ COVID_73_82y_male_NonICU      <dbl> 0.46, 0.01, 0.02, 0.02, 0.17, 0.04, 0.04~
## $ COVID_74_55y_female_ICU       <dbl> 0.24, 0.00, 0.12, 0.02, 0.26, 0.00, 0.13~
## $ COVID_75_55y_male_NonICU      <dbl> 0.23, 0.01, 0.14, 0.00, 0.00, 0.00, 0.00~
## $ COVID_76_73y_female_ICU       <dbl> 0.17, 0.00, 0.09, 0.01, 0.04, 0.00, 0.04~
## $ COVID_77_55y_female_ICU       <dbl> 0.05, 0.00, 0.01, 0.00, 0.00, 0.00, 0.00~
## $ COVID_78_80y_male_NonICU      <dbl> 0.19, 0.00, 0.20, 0.00, 0.00, 0.00, 0.00~
## $ COVID_79_27y_male_NonICU      <dbl> 0.08, 0.01, 0.03, 0.00, 0.00, 0.00, 0.03~
## $ COVID_80_71y_male_ICU         <dbl> 0.28, 0.00, 0.05, 0.00, 0.05, 0.00, 0.00~
## $ COVID_82_67y_male_NonICU      <dbl> 0.39, 0.01, 0.10, 0.00, 0.00, 0.00, 0.00~
## $ COVID_83_85y_female_NonICU    <dbl> 0.47, 0.00, 0.18, 0.05, 0.00, 0.00, 0.00~
## $ COVID_84_75y_female_NonICU    <dbl> 0.35, 0.00, 0.03, 0.00, 0.17, 0.00, 0.03~
## $ COVID_85_62y_male_ICU         <dbl> 0.29, 0.00, 0.04, 0.00, 0.00, 0.00, 0.00~
## $ COVID_86_52y_female_NonICU    <dbl> 0.60, 0.00, 0.27, 0.02, 0.00, 0.00, 0.00~
## $ COVID_87_61y_male_ICU         <dbl> 0.65, 0.00, 0.15, 0.00, 0.00, 0.00, 0.00~
## $ COVID_89_90y_female_NonICU    <dbl> 0.20, 0.00, 0.07, 0.03, 0.14, 0.00, 0.00~
## $ COVID_90_86y_female_NonICU    <dbl> 0.40, 0.00, 0.05, 0.01, 0.31, 0.02, 0.00~
## $ COVID_91_29y_female_NonICU    <dbl> 0.60, 0.00, 0.03, 0.02, 0.05, 0.00, 0.05~
## $ COVID_92_82y_female_ICU       <dbl> 0.34, 0.00, 0.02, 0.04, 0.58, 0.00, 0.06~
## $ COVID_93_81y_female_ICU       <dbl> 0.37, 0.00, 0.11, 0.00, 0.05, 0.00, 0.00~
## $ COVID_94_24y_female_NonICU    <dbl> 0.81, 0.00, 0.17, 0.02, 0.00, 0.06, 0.00~
## $ COVID_95_49y_male_NonICU      <dbl> 0.37, 0.01, 0.20, 0.02, 0.15, 0.00, 0.00~
## $ COVID_96_51y_male_NonICU      <dbl> 1.61, 0.00, 0.02, 0.00, 0.00, 0.00, 0.00~
## $ COVID_97_76y_male_ICU         <dbl> 0.19, 0.00, 0.02, 0.05, 0.12, 0.03, 0.00~
## $ COVID_98_81y_male_NonICU      <dbl> 0.78, 0.00, 0.26, 0.00, 0.37, 0.00, 0.02~
```

```
## $ COVID_99_71y_male_ICU        <dbl> 0.33, 0.00, 0.02, 0.00, 0.04, 0.00, 0.00~
## $ COVID_100_74y_female_NonICU  <dbl> 0.30, 0.00, 0.09, 0.00, 0.04, 0.00, 0.02~
## $ COVID_101_58y_male_ICU       <dbl> 0.33, 0.00, 0.11, 0.03, 0.05, 0.00, 0.00~
## $ COVID_102_84y_male_NonICU    <dbl> 0.12, 0.00, 0.01, 0.01, 0.00, 0.07, 0.00~
## $ COVID_103_83y_male_NonICU    <dbl> 0.20, 0.00, 0.03, 0.03, 0.04, 0.00, 0.00~
## $ NONCOVID_01_54y_female_NonICU <dbl> 0.89, 0.00, 0.04, 0.00, 0.00, 0.00, 0.00~
## $ NONCOVID_02_65y_male_ICU      <dbl> 0.32, 0.00, 0.01, 0.00, 0.04, 0.00, 0.00~
## $ NONCOVID_03_65y_male_ICU      <dbl> 0.44, 0.00, 0.05, 0.02, 0.04, 0.00, 0.00~
## $ NONCOVID_04_90y_male_NonICU   <dbl> 0.21, 0.00, 0.05, 0.00, 0.21, 0.00, 0.00~
## $ NONCOVID_05_83y_female_NonICU <dbl> 0.31, 0.00, 0.01, 0.01, 0.00, 0.00, 0.00~
## $ NONCOVID_06_75y_female_ICU    <dbl> 0.89, 0.00, 0.14, 0.01, 0.00, 0.06, 0.00~
## $ NONCOVID_07_50y_male_ICU      <dbl> 0.45, 0.00, 0.07, 0.02, 0.00, 0.00, 0.00~
## $ NONCOVID_08_53y_female_ICU    <dbl> 0.47, 0.01, 0.04, 0.00, 0.15, 0.00, 0.00~
## $ NONCOVID_09_49y_female_NonICU <dbl> 0.40, 0.00, 0.04, 0.00, 0.00, 0.00, 0.05~
## $ NONCOVID_10_67y_male_ICU      <dbl> 0.33, 0.00, 0.05, 0.01, 0.23, 0.08, 0.00~
## $ NONCOVID_11_58y_female_NonICU <dbl> 0.58, 0.00, 0.03, 0.00, 0.00, 0.00, 0.00~
## $ NONCOVID_12_82y_male_ICU      <dbl> 0.12, 0.00, 0.02, 0.00, 0.00, 0.02, 0.00~
## $ NONCOVID_13_65y_male_ICU      <dbl> 0.31, 0.00, 0.04, 0.01, 0.32, 0.02, 0.00~
## $ NONCOVID_14_75y_female_ICU    <dbl> 0.16, 0.00, 0.08, 0.00, 0.05, 0.02, 0.00~
## $ NONCOVID_15_83y_unknown_ICU   <dbl> 0.59, 0.00, 0.03, 0.04, 0.00, 0.19, 0.00~
## $ NONCOVID_16_40y_female_ICU    <dbl> 0.34, 0.00, 0.07, 0.00, 0.13, 0.00, 0.00~
## $ NONCOVID_17_84y_female_ICU    <dbl> 0.37, 0.00, 0.07, 0.01, 0.18, 0.00, 0.00~
## $ NONCOVID_18_88y_male_ICU      <dbl> 0.33, 0.00, 0.06, 0.00, 0.00, 0.00, 0.00~
## $ NONCOVID_19_66y_female_ICU    <dbl> 0.25, 0.00, 0.11, 0.00, 0.04, 0.03, 0.00~
## $ NONCOVID_20_62y_female_ICU    <dbl> 0.20, 0.00, 0.01, 0.02, 0.00, 0.07, 0.03~
## $ NONCOVID_21_71y_male_NonICU   <dbl> 0.40, 0.00, 0.04, 0.02, 0.00, 0.00, 0.00~
## $ NONCOVID_22_63y_male_NonICU   <dbl> 0.30, 0.00, 0.02, 0.02, 0.00, 0.00, 0.00~
## $ NONCOVID_23_42y_female_NonICU <dbl> 0.70, 0.00, 0.02, 0.01, 0.00, 0.00, 0.00~
## $ NONCOVID_24_32y_female_NonICU <dbl> 0.75, 0.00, 0.27, 0.00, 0.06, 0.00, 0.00~
## $ NONCOVID_25_62y_male_NonICU   <dbl> 2.80, 0.00, 0.04, 0.00, 0.00, 0.00, 0.00~
## $ NONCOVID_26_36y_male_ICU      <dbl> 0.22, 0.00, 0.28, 0.00, 0.00, 0.00, 0.00~
```

Work with the gene expression data

```r
#working with rows & columns
gene_exp_t <- gene_exp_raw %>%
  column_to_rownames(var = "X") %>% #change the 1st columnname to rowname
  t() %>% #transpose the data frame
  as.data.frame()

gene_exp_t <- gene_exp_t %>%
  rownames_to_column(var="participant_id")

glimpse(gene_exp_t)
```

```
## Rows: 126
## Columns: 101
## $ participant_id <chr> "COVID_01_39y_male_NonICU", "COVID_02_63y_male_NonICU",~
## $ A1BG           <dbl> 0.49, 0.29, 0.26, 0.45, 0.17, 0.21, 0.49, 0.12, 0.51, 0~
## $ A1CF           <dbl> 0.00, 0.00, 0.00, 0.01, 0.00, 0.00, 0.01, 0.00, 0.01, 0~
## $ A2M            <dbl> 0.21, 0.14, 0.03, 0.09, 0.00, 0.08, 0.23, 0.08, 0.88, 0~
## $ A2ML1          <dbl> 0.04, 0.00, 0.02, 0.07, 0.05, 0.04, 0.03, 0.01, 0.02, 0~
## $ A3GALT2        <dbl> 0.07, 0.00, 0.00, 0.00, 0.07, 0.00, 0.07, 0.00, 0.79, 0~
```

```
## $ A4GALT      <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0~
## $ A4GNT       <dbl> 0.03, 0.05, 0.07, 0.00, 0.00, 0.03, 0.00, 0.00, 0.00, 0~
## $ AAAS        <dbl> 18.92, 18.68, 13.85, 22.11, 8.45, 19.60, 28.59, 10.50, ~
## $ AACS        <dbl> 4.07, 3.00, 1.83, 4.22, 1.17, 3.15, 4.24, 2.10, 4.86, 2~
## $ AADAC       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ AADACL2     <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.01, 0.00, 0.00, 0.00, 0~
## $ AADACL3     <dbl> 0.00, 0.06, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0~
## $ AADACL4     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ AADAT       <dbl> 0.00, 0.00, 0.00, 0.00, 0.03, 0.00, 0.03, 0.06, 0.00, 0~
## $ AAGAB       <dbl> 22.93, 21.69, 18.27, 26.69, 17.02, 17.50, 26.28, 13.91,~
## $ AAK1        <dbl> 7.12, 6.46, 3.92, 8.74, 7.26, 6.65, 12.78, 3.82, 6.23, ~
## $ AAMDC       <dbl> 17.19, 13.06, 13.35, 17.53, 10.14, 12.98, 12.62, 9.06, ~
## $ AAMP        <dbl> 61.08, 54.54, 25.19, 67.95, 18.29, 45.09, 51.35, 20.41,~
## $ AANAT       <dbl> 0.31, 0.00, 0.65, 0.08, 1.02, 0.12, 0.13, 0.13, 0.85, 0~
## $ AAR2        <dbl> 21.59, 19.22, 8.72, 20.83, 7.46, 13.60, 23.90, 8.18, 13~
## $ AARD        <dbl> 0.18, 0.08, 0.05, 0.03, 0.10, 0.09, 0.12, 0.06, 0.09, 0~
## $ AARS1       <dbl> 13.52, 13.91, 5.60, 15.34, 6.29, 13.70, 29.47, 4.97, 27~
## $ AARS2       <dbl> 2.43, 2.15, 1.08, 2.69, 0.55, 2.21, 4.35, 1.06, 2.72, 2~
## $ AARSD1      <dbl> 9.63, 11.80, 4.04, 14.61, 5.32, 11.30, 21.09, 4.32, 10.~
## $ AASDH       <dbl> 6.38, 4.83, 2.76, 8.23, 5.23, 4.91, 8.90, 4.57, 4.60, 4~
## $ AASDHPPT    <dbl> 19.45, 12.14, 7.64, 18.94, 11.54, 14.16, 23.49, 10.08, ~
## $ AASS        <dbl> 0.21, 0.42, 0.04, 0.41, 0.21, 0.26, 0.63, 0.09, 0.17, 0~
## $ AATF        <dbl> 45.83, 39.37, 42.35, 41.92, 30.56, 36.30, 44.16, 31.81,~
## $ AATK        <dbl> 5.60, 9.79, 5.34, 4.73, 1.50, 7.52, 4.16, 5.62, 12.09, ~
## $ ABAT        <dbl> 9.63, 10.36, 4.59, 11.02, 4.42, 7.04, 12.02, 7.26, 10.3~
## $ ABCA1       <dbl> 32.30, 15.84, 34.38, 14.24, 18.39, 3.64, 14.66, 5.34, 2~
## $ ABCA10      <dbl> 0.32, 0.37, 0.29, 0.31, 0.19, 0.22, 0.55, 0.12, 0.29, 0~
## $ ABCA12      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ABCA13      <dbl> 0.49, 3.36, 0.26, 0.13, 0.16, 0.97, 0.23, 6.62, 1.29, 8~
## $ ABCA2       <dbl> 8.47, 9.49, 14.24, 6.37, 5.90, 6.18, 9.16, 5.85, 15.72,~
## $ ABCA3       <dbl> 0.37, 0.71, 0.17, 0.94, 0.17, 0.43, 0.75, 0.17, 0.56, 0~
## $ ABCA4       <dbl> 0.01, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0~
## $ ABCA5       <dbl> 1.86, 2.81, 2.17, 2.94, 1.38, 1.89, 3.60, 2.78, 2.29, 2~
## $ ABCA6       <dbl> 0.19, 0.11, 0.07, 0.02, 0.03, 0.03, 0.05, 0.10, 0.03, 0~
## $ ABCA7       <dbl> 39.31, 30.42, 54.85, 18.91, 23.28, 23.43, 30.95, 27.47,~
## $ ABCA8       <dbl> 0.00, 0.00, 0.00, 0.01, 0.00, 0.00, 0.00, 0.00, 0.00, 0~
## $ ABCA9       <dbl> 0.27, 0.20, 0.33, 0.30, 0.21, 0.15, 0.23, 0.15, 0.20, 0~
## $ ABCB1       <dbl> 1.61, 1.68, 0.59, 3.14, 1.66, 1.92, 4.20, 0.33, 1.41, 1~
## $ ABCB10      <dbl> 15.59, 10.98, 4.04, 10.00, 5.69, 9.33, 15.09, 9.94, 9.0~
## $ ABCB11      <dbl> 0.38, 0.25, 0.14, 0.18, 0.07, 0.12, 0.18, 0.05, 0.25, 0~
## $ ABCB4       <dbl> 0.01, 0.19, 0.09, 0.74, 0.21, 0.26, 0.81, 0.07, 0.35, 0~
## $ ABCB5       <dbl> 0.04, 0.04, 0.15, 0.11, 0.04, 0.06, 0.11, 0.04, 0.07, 0~
## $ ABCB6       <dbl> 2.77, 2.52, 3.14, 2.11, 1.14, 2.08, 3.99, 2.57, 4.76, 2~
## $ ABCB7       <dbl> 6.42, 5.80, 2.59, 7.42, 3.40, 4.87, 7.33, 3.58, 5.55, 3~
## $ ABCB8       <dbl> 2.95, 2.35, 1.57, 2.71, 0.71, 2.41, 3.51, 0.72, 2.51, 1~
## $ ABCB9       <dbl> 0.20, 0.51, 0.05, 0.31, 0.12, 0.75, 1.11, 0.26, 4.04, 0~
## $ ABCC1       <dbl> 11.20, 9.39, 4.74, 13.05, 4.29, 8.45, 12.24, 4.61, 8.65~
## $ ABCC10      <dbl> 8.85, 4.46, 4.08, 5.75, 1.91, 3.88, 7.33, 3.21, 5.87, 4~
## $ ABCC11      <dbl> 0.04, 0.01, 0.09, 0.02, 0.02, 0.00, 0.02, 0.09, 0.42, 0~
## $ ABCC12      <dbl> 0.00, 0.00, 0.00, 0.01, 0.00, 0.00, 0.00, 0.00, 0.01, 0~
## $ ABCC2       <dbl> 1.65, 1.31, 1.72, 1.55, 2.56, 0.85, 1.57, 1.37, 3.28, 1~
## $ ABCC3       <dbl> 7.19, 16.21, 1.44, 11.11, 3.13, 4.94, 5.34, 3.87, 7.84,~
## $ ABCC4       <dbl> 5.96, 4.49, 0.58, 4.09, 7.37, 4.12, 8.10, 4.59, 13.45, ~
## $ ABCC5       <dbl> 14.83, 22.21, 13.02, 17.49, 5.53, 14.04, 23.46, 11.86, ~
```

```
## $ ABCC6         <dbl> 6.62, 5.87, 3.13, 3.37, 1.66, 3.96, 3.61, 6.19, 3.07, 3~
## $ ABCC8         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ABCC9         <dbl> 1.53, 1.89, 1.97, 2.82, 1.84, 1.37, 2.31, 1.24, 1.98, 1~
## $ ABCD1         <dbl> 11.26, 4.65, 5.83, 4.80, 1.93, 2.56, 3.96, 2.65, 4.35, ~
## $ ABCD2         <dbl> 1.55, 1.08, 0.39, 1.74, 0.59, 2.64, 5.25, 0.14, 1.55, 0~
## $ ABCD3         <dbl> 10.62, 11.23, 3.99, 12.38, 8.19, 8.52, 13.14, 10.82, 8.~
## $ ABCD4         <dbl> 11.19, 9.86, 7.83, 13.60, 5.43, 10.19, 12.96, 4.79, 8.9~
## $ ABCE1         <dbl> 14.09, 15.52, 4.50, 21.20, 7.16, 14.20, 28.46, 5.15, 19~
## $ ABCF1         <dbl> 24.40, 16.84, 7.89, 23.92, 12.33, 19.56, 28.37, 7.60, 1~
## $ ABCF2         <dbl> 8.77, 7.21, 3.12, 8.26, 2.76, 5.57, 11.28, 2.95, 6.56, ~
## $ `ABCF2-H2BE1` <dbl> 11.67, 8.84, 3.06, 12.51, 4.72, 6.47, 13.61, 3.78, 10.5~
## $ ABCF3         <dbl> 23.37, 18.63, 11.57, 23.62, 9.86, 16.81, 25.96, 9.26, 2~
## $ ABCG1         <dbl> 32.76, 23.76, 41.14, 15.72, 9.43, 13.84, 14.96, 7.55, 7~
## $ ABCG2         <dbl> 0.17, 0.00, 0.15, 0.14, 0.03, 0.13, 0.05, 0.07, 0.00, 0~
## $ ABCG4         <dbl> 0.01, 0.07, 0.00, 0.00, 0.00, 0.00, 0.05, 0.00, 0.00, 0~
## $ ABCG5         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ABCG8         <dbl> 0.02, 0.01, 0.00, 0.04, 0.04, 0.00, 0.02, 0.02, 0.03, 0~
## $ ABHD1         <dbl> 0.04, 0.14, 0.00, 0.31, 0.04, 0.21, 0.38, 0.00, 0.05, 0~
## $ ABHD10        <dbl> 14.74, 10.57, 6.84, 16.73, 6.93, 10.02, 18.21, 8.30, 10~
## $ ABHD11        <dbl> 8.50, 8.85, 8.22, 9.64, 4.02, 6.51, 8.27, 4.15, 6.66, 6~
## $ ABHD12        <dbl> 14.66, 9.02, 4.84, 13.22, 3.36, 9.33, 13.35, 4.17, 10.7~
## $ ABHD12B       <dbl> 0.55, 1.18, 0.88, 0.68, 0.68, 0.48, 0.32, 0.59, 1.43, 0~
## $ ABHD13        <dbl> 15.58, 14.33, 9.97, 16.70, 21.02, 11.72, 16.37, 17.79, ~
## $ ABHD14A       <dbl> 6.89, 7.08, 2.12, 9.21, 1.95, 8.84, 13.20, 1.69, 8.89, ~
## $ `ABHD14A-ACY1`<dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.17, 0.00, 0.00, 0~
## $ ABHD14B       <dbl> 18.53, 18.78, 7.63, 28.23, 6.49, 23.99, 36.75, 6.54, 15~
## $ ABHD15        <dbl> 6.82, 5.76, 2.41, 8.12, 2.26, 4.99, 8.38, 1.81, 4.43, 2~
## $ ABHD16A       <dbl> 77.48, 57.11, 77.73, 56.77, 63.66, 44.47, 57.29, 52.79,~
## $ ABHD16B       <dbl> 0.29, 0.19, 0.08, 0.05, 0.04, 0.12, 0.18, 0.04, 0.32, 0~
## $ ABHD17A       <dbl> 13.79, 11.89, 6.88, 12.28, 4.11, 11.68, 12.89, 3.71, 9.~
## $ ABHD17B       <dbl> 8.78, 7.38, 5.59, 9.89, 6.64, 7.10, 11.82, 6.96, 9.58, ~
## $ ABHD17C       <dbl> 1.37, 3.34, 1.64, 1.24, 1.17, 2.52, 2.60, 2.73, 2.56, 4~
## $ ABHD18        <dbl> 6.93, 5.10, 4.06, 6.57, 7.94, 3.93, 7.49, 4.65, 6.46, 3~
## $ ABHD2         <dbl> 46.50, 63.70, 36.05, 54.52, 72.44, 45.66, 44.91, 38.54,~
## $ ABHD3         <dbl> 73.51, 80.70, 113.96, 106.44, 190.95, 83.74, 121.66, 15~
## $ ABHD4         <dbl> 44.47, 44.43, 37.71, 41.75, 29.65, 34.00, 31.88, 35.22,~
## $ ABHD5         <dbl> 60.14, 79.17, 77.50, 69.89, 51.39, 60.38, 41.47, 135.01~
## $ ABHD6         <dbl> 3.47, 4.54, 2.15, 7.84, 1.45, 4.30, 5.11, 5.93, 4.96, 2~
## $ ABHD8         <dbl> 3.27, 0.99, 2.59, 1.86, 1.11, 1.27, 2.22, 0.84, 2.72, 2~
## $ ABI1          <dbl> 75.36, 60.41, 61.63, 66.52, 79.84, 54.23, 66.89, 64.55,~
## $ ABI2          <dbl> 3.22, 3.68, 0.67, 4.99, 2.12, 3.47, 5.82, 1.23, 2.57, 0~
```

```r
#combine gene expression & metadata
#common samples ID
common_ids <- intersect(metadata$participant_id, gene_exp_t$participant_id)
#filter data using these IDs
metadata_clean <- metadata %>% filter(participant_id %in% common_ids)
gene_expr_clean <- gene_exp_t %>% filter(participant_id %in% common_ids)

full_data <- left_join(metadata, gene_exp_t, by = "participant_id")
glimpse(full_data)
```

```
## Rows: 126
## Columns: 125
```

```
## $ participant_id                            <chr> "COVID_01_39y_male_NonICU", "C~
## $ geo_accession                             <chr> "GSM4753021", "GSM4753022", "G~
## $ status                                    <chr> "Public on Aug 29 2020", "Publ~
## $ X.Sample_submission_date                  <chr> "Aug 28 2020", "Aug 28 2020", ~
## $ last_update_date                          <chr> "Aug 29 2020", "Aug 29 2020", ~
## $ type                                      <chr> "SRA", "SRA", "SRA", "SRA", "S~
## $ channel_count                             <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ source_name_ch1                           <chr> "Leukocytes from whole blood",~
## $ organism_ch1                              <chr> "Homo sapiens", "Homo sapiens"~
## $ disease_status                            <chr> "disease state: COVID-19", "di~
## $ age                                       <chr> "39", "63", "33", "49", "49", ~
## $ sex                                       <chr> " male", " male", " male", " m~
## $ icu_status                                <chr> " no", " no", " no", " no", " ~
## $ apacheii                                  <chr> "15", " unknown", " unknown", ~
## $ charlson_score                            <int> 0, 2, 2, 1, 1, 1, 7, 7, 2, 1, ~
## $ mechanical_ventilation                    <chr> " yes", " no", " no", " no", "~
## $ ventilator.free_days                      <int> 0, 28, 28, 28, 23, 28, 28, 0, ~
## $ hospital.free_days_post_45_day_followup   <int> 0, 39, 18, 39, 27, 36, 42, 0, ~
## $ ferritin.ng.ml.                           <chr> "946", "1060", "1335", "583", ~
## $ crp.mg.l.                                 <chr> "73.1", " unknown", "53.2", "2~
## $ ddimer.mg.l_feu.                          <chr> "1.3", "1.03", "1.48", "1.32",~
## $ procalcitonin.ng.ml..                     <chr> "36", "0.37", "0.07", "0.98", ~
## $ lactate.mmol.l.                           <chr> "0.9", " unknown", " unknown",~
## $ fibrinogen                                <chr> "513", "unknown", "513", "949"~
## $ sofa                                      <chr> "8", " unknown", " unknown", "~
## $ A1BG                                      <dbl> 0.49, 0.29, 0.26, 0.45, 0.17, ~
## $ A1CF                                      <dbl> 0.00, 0.00, 0.00, 0.01, 0.00, ~
## $ A2M                                       <dbl> 0.21, 0.14, 0.03, 0.09, 0.00, ~
## $ A2ML1                                     <dbl> 0.04, 0.00, 0.02, 0.07, 0.05, ~
## $ A3GALT2                                   <dbl> 0.07, 0.00, 0.00, 0.00, 0.07, ~
## $ A4GALT                                    <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ A4GNT                                     <dbl> 0.03, 0.05, 0.07, 0.00, 0.00, ~
## $ AAAS                                      <dbl> 18.92, 18.68, 13.85, 22.11, 8.~
## $ AACS                                      <dbl> 4.07, 3.00, 1.83, 4.22, 1.17, ~
## $ AADAC                                     <dbl> 0, 0, 0, 0, 0, NA, 0, 0, 0, 0,~
## $ AADACL2                                   <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ AADACL3                                   <dbl> 0.00, 0.06, 0.00, 0.00, 0.00, ~
## $ AADACL4                                   <dbl> 0, 0, 0, 0, 0, NA, 0, 0, 0, 0,~
## $ AADAT                                     <dbl> 0.00, 0.00, 0.00, 0.00, 0.03, ~
## $ AAGAB                                     <dbl> 22.93, 21.69, 18.27, 26.69, 17~
## $ AAK1                                      <dbl> 7.12, 6.46, 3.92, 8.74, 7.26, ~
## $ AAMDC                                     <dbl> 17.19, 13.06, 13.35, 17.53, 10~
## $ AAMP                                      <dbl> 61.08, 54.54, 25.19, 67.95, 18~
## $ AANAT                                     <dbl> 0.31, 0.00, 0.65, 0.08, 1.02, ~
## $ AAR2                                      <dbl> 21.59, 19.22, 8.72, 20.83, 7.4~
## $ AARD                                      <dbl> 0.18, 0.08, 0.05, 0.03, 0.10, ~
## $ AARS1                                     <dbl> 13.52, 13.91, 5.60, 15.34, 6.2~
## $ AARS2                                     <dbl> 2.43, 2.15, 1.08, 2.69, 0.55, ~
## $ AARSD1                                    <dbl> 9.63, 11.80, 4.04, 14.61, 5.32~
## $ AASDH                                     <dbl> 6.38, 4.83, 2.76, 8.23, 5.23, ~
## $ AASDHPPT                                  <dbl> 19.45, 12.14, 7.64, 18.94, 11.~
## $ AASS                                      <dbl> 0.21, 0.42, 0.04, 0.41, 0.21, ~
## $ AATF                                      <dbl> 45.83, 39.37, 42.35, 41.92, 30~
## $ AATK                                      <dbl> 5.60, 9.79, 5.34, 4.73, 1.50, ~
```

```
## $ ABAT              <dbl> 9.63, 10.36, 4.59, 11.02, 4.42~
## $ ABCA1             <dbl> 32.30, 15.84, 34.38, 14.24, 18~
## $ ABCA10            <dbl> 0.32, 0.37, 0.29, 0.31, 0.19, ~
## $ ABCA12            <dbl> 0, 0, 0, 0, 0, NA, 0, 0, 0, 0,~
## $ ABCA13            <dbl> 0.49, 3.36, 0.26, 0.13, 0.16, ~
## $ ABCA2             <dbl> 8.47, 9.49, 14.24, 6.37, 5.90,~
## $ ABCA3             <dbl> 0.37, 0.71, 0.17, 0.94, 0.17, ~
## $ ABCA4             <dbl> 0.01, 0.00, 0.00, 0.00, 0.00, ~
## $ ABCA5             <dbl> 1.86, 2.81, 2.17, 2.94, 1.38, ~
## $ ABCA6             <dbl> 0.19, 0.11, 0.07, 0.02, 0.03, ~
## $ ABCA7             <dbl> 39.31, 30.42, 54.85, 18.91, 23~
## $ ABCA8             <dbl> 0.00, 0.00, 0.00, 0.01, 0.00, ~
## $ ABCA9             <dbl> 0.27, 0.20, 0.33, 0.30, 0.21, ~
## $ ABCB1             <dbl> 1.61, 1.68, 0.59, 3.14, 1.66, ~
## $ ABCB10            <dbl> 15.59, 10.98, 4.04, 10.00, 5.6~
## $ ABCB11            <dbl> 0.38, 0.25, 0.14, 0.18, 0.07, ~
## $ ABCB4             <dbl> 0.01, 0.19, 0.09, 0.74, 0.21, ~
## $ ABCB5             <dbl> 0.04, 0.04, 0.15, 0.11, 0.04, ~
## $ ABCB6             <dbl> 2.77, 2.52, 3.14, 2.11, 1.14, ~
## $ ABCB7             <dbl> 6.42, 5.80, 2.59, 7.42, 3.40, ~
## $ ABCB8             <dbl> 2.95, 2.35, 1.57, 2.71, 0.71, ~
## $ ABCB9             <dbl> 0.20, 0.51, 0.05, 0.31, 0.12, ~
## $ ABCC1             <dbl> 11.20, 9.39, 4.74, 13.05, 4.29~
## $ ABCC10            <dbl> 8.85, 4.46, 4.08, 5.75, 1.91, ~
## $ ABCC11            <dbl> 0.04, 0.01, 0.09, 0.02, 0.02, ~
## $ ABCC12            <dbl> 0.00, 0.00, 0.00, 0.01, 0.00, ~
## $ ABCC2             <dbl> 1.65, 1.31, 1.72, 1.55, 2.56, ~
## $ ABCC3             <dbl> 7.19, 16.21, 1.44, 11.11, 3.13~
## $ ABCC4             <dbl> 5.96, 4.49, 0.58, 4.09, 7.37, ~
## $ ABCC5             <dbl> 14.83, 22.21, 13.02, 17.49, 5.~
## $ ABCC6             <dbl> 6.62, 5.87, 3.13, 3.37, 1.66, ~
## $ ABCC8             <dbl> 0, 0, 0, 0, 0, NA, 0, 0, 0, 0,~
## $ ABCC9             <dbl> 1.53, 1.89, 1.97, 2.82, 1.84, ~
## $ ABCD1             <dbl> 11.26, 4.65, 5.83, 4.80, 1.93,~
## $ ABCD2             <dbl> 1.55, 1.08, 0.39, 1.74, 0.59, ~
## $ ABCD3             <dbl> 10.62, 11.23, 3.99, 12.38, 8.1~
## $ ABCD4             <dbl> 11.19, 9.86, 7.83, 13.60, 5.43~
## $ ABCE1             <dbl> 14.09, 15.52, 4.50, 21.20, 7.1~
## $ ABCF1             <dbl> 24.40, 16.84, 7.89, 23.92, 12.~
## $ ABCF2             <dbl> 8.77, 7.21, 3.12, 8.26, 2.76, ~
## $ 'ABCF2-H2BE1'     <dbl> 11.67, 8.84, 3.06, 12.51, 4.72~
## $ ABCF3             <dbl> 23.37, 18.63, 11.57, 23.62, 9.~
## $ ABCG1             <dbl> 32.76, 23.76, 41.14, 15.72, 9.~
## $ ABCG2             <dbl> 0.17, 0.00, 0.15, 0.14, 0.03, ~
## $ ABCG4             <dbl> 0.01, 0.07, 0.00, 0.00, 0.00, ~
## $ ABCG5             <dbl> 0, 0, 0, 0, 0, NA, 0, 0, 0, 0,~
## $ ABCG8             <dbl> 0.02, 0.01, 0.00, 0.04, 0.04, ~
## $ ABHD1             <dbl> 0.04, 0.14, 0.00, 0.31, 0.04, ~
## $ ABHD10            <dbl> 14.74, 10.57, 6.84, 16.73, 6.9~
## $ ABHD11            <dbl> 8.50, 8.85, 8.22, 9.64, 4.02, ~
## $ ABHD12            <dbl> 14.66, 9.02, 4.84, 13.22, 3.36~
## $ ABHD12B           <dbl> 0.55, 1.18, 0.88, 0.68, 0.68, ~
## $ ABHD13            <dbl> 15.58, 14.33, 9.97, 16.70, 21.~
## $ ABHD14A           <dbl> 6.89, 7.08, 2.12, 9.21, 1.95, ~
```

```
## $ `ABHD14A-ACY1`              <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, ~
## $ ABHD14B                      <dbl> 18.53, 18.78, 7.63, 28.23, 6.4~
## $ ABHD15                       <dbl> 6.82, 5.76, 2.41, 8.12, 2.26, ~
## $ ABHD16A                      <dbl> 77.48, 57.11, 77.73, 56.77, 63~
## $ ABHD16B                      <dbl> 0.29, 0.19, 0.08, 0.05, 0.04, ~
## $ ABHD17A                      <dbl> 13.79, 11.89, 6.88, 12.28, 4.1~
## $ ABHD17B                      <dbl> 8.78, 7.38, 5.59, 9.89, 6.64, ~
## $ ABHD17C                      <dbl> 1.37, 3.34, 1.64, 1.24, 1.17, ~
## $ ABHD18                       <dbl> 6.93, 5.10, 4.06, 6.57, 7.94, ~
## $ ABHD2                        <dbl> 46.50, 63.70, 36.05, 54.52, 72~
## $ ABHD3                        <dbl> 73.51, 80.70, 113.96, 106.44, ~
## $ ABHD4                        <dbl> 44.47, 44.43, 37.71, 41.75, 29~
## $ ABHD5                        <dbl> 60.14, 79.17, 77.50, 69.89, 51~
## $ ABHD6                        <dbl> 3.47, 4.54, 2.15, 7.84, 1.45, ~
## $ ABHD8                        <dbl> 3.27, 0.99, 2.59, 1.86, 1.11, ~
## $ ABI1                         <dbl> 75.36, 60.41, 61.63, 66.52, 79~
## $ ABI2                         <dbl> 3.22, 3.68, 0.67, 4.99, 2.12, ~
```

Choose variables

```
#check number of rows before start
nrow(full_data) == nrow(metadata)
```

```
## [1] TRUE
```

```
#check variables
names(full_data)
```

```
##   [1] "participant_id"
##   [2] "geo_accession"
##   [3] "status"
##   [4] "X.Sample_submission_date"
##   [5] "last_update_date"
##   [6] "type"
##   [7] "channel_count"
##   [8] "source_name_ch1"
##   [9] "organism_ch1"
##  [10] "disease_status"
##  [11] "age"
##  [12] "sex"
##  [13] "icu_status"
##  [14] "apacheii"
##  [15] "charlson_score"
##  [16] "mechanical_ventilation"
##  [17] "ventilator.free_days"
##  [18] "hospital.free_days_post_45_day_followup"
##  [19] "ferritin.ng.ml."
##  [20] "crp.mg.l."
##  [21] "ddimer.mg.l_feu."
##  [22] "procalcitonin.ng.ml.."
##  [23] "lactate.mmol.l."
##  [24] "fibrinogen"
##  [25] "sofa"
```

```
##  [26] "A1BG"
##  [27] "A1CF"
##  [28] "A2M"
##  [29] "A2ML1"
##  [30] "A3GALT2"
##  [31] "A4GALT"
##  [32] "A4GNT"
##  [33] "AAAS"
##  [34] "AACS"
##  [35] "AADAC"
##  [36] "AADACL2"
##  [37] "AADACL3"
##  [38] "AADACL4"
##  [39] "AADAT"
##  [40] "AAGAB"
##  [41] "AAK1"
##  [42] "AAMDC"
##  [43] "AAMP"
##  [44] "AANAT"
##  [45] "AAR2"
##  [46] "AARD"
##  [47] "AARS1"
##  [48] "AARS2"
##  [49] "AARSD1"
##  [50] "AASDH"
##  [51] "AASDHPPT"
##  [52] "AASS"
##  [53] "AATF"
##  [54] "AATK"
##  [55] "ABAT"
##  [56] "ABCA1"
##  [57] "ABCA10"
##  [58] "ABCA12"
##  [59] "ABCA13"
##  [60] "ABCA2"
##  [61] "ABCA3"
##  [62] "ABCA4"
##  [63] "ABCA5"
##  [64] "ABCA6"
##  [65] "ABCA7"
##  [66] "ABCA8"
##  [67] "ABCA9"
##  [68] "ABCB1"
##  [69] "ABCB10"
##  [70] "ABCB11"
##  [71] "ABCB4"
##  [72] "ABCB5"
##  [73] "ABCB6"
##  [74] "ABCB7"
##  [75] "ABCB8"
##  [76] "ABCB9"
##  [77] "ABCC1"
##  [78] "ABCC10"
##  [79] "ABCC11"
```

```
##  [80] "ABCC12"
##  [81] "ABCC2"
##  [82] "ABCC3"
##  [83] "ABCC4"
##  [84] "ABCC5"
##  [85] "ABCC6"
##  [86] "ABCC8"
##  [87] "ABCC9"
##  [88] "ABCD1"
##  [89] "ABCD2"
##  [90] "ABCD3"
##  [91] "ABCD4"
##  [92] "ABCE1"
##  [93] "ABCF1"
##  [94] "ABCF2"
##  [95] "ABCF2-H2BE1"
##  [96] "ABCF3"
##  [97] "ABCG1"
##  [98] "ABCG2"
##  [99] "ABCG4"
## [100] "ABCG5"
## [101] "ABCG8"
## [102] "ABHD1"
## [103] "ABHD10"
## [104] "ABHD11"
## [105] "ABHD12"
## [106] "ABHD12B"
## [107] "ABHD13"
## [108] "ABHD14A"
## [109] "ABHD14A-ACY1"
## [110] "ABHD14B"
## [111] "ABHD15"
## [112] "ABHD16A"
## [113] "ABHD16B"
## [114] "ABHD17A"
## [115] "ABHD17B"
## [116] "ABHD17C"
## [117] "ABHD18"
## [118] "ABHD2"
## [119] "ABHD3"
## [120] "ABHD4"
## [121] "ABHD5"
## [122] "ABHD6"
## [123] "ABHD8"
## [124] "ABI1"
## [125] "ABI2"
```

```r
'AAK1' %in% colnames(full_data)
```

```
## [1] TRUE
```

```r
summary(full_data$AAK1)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.840   3.790   5.090   5.403   6.460  13.890       1
```

```
# change ferritin from char to num
full_data <- full_data %>%
  mutate(ferritin = as.numeric(`ferritin.ng.ml.`))
```

```
## Warning: There was 1 warning in 'mutate()'.
## i In argument: 'ferritin = as.numeric(ferritin.ng.ml.)'.
## Caused by warning:
## ! NAs introduced by coercion
```

```
summary(full_data$ferritin)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     14.0   222.0   573.0   833.5  1091.5  5971.0      16
```

```
summary(full_data$AAK1)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.840   3.790   5.090   5.403   6.460  13.890       1
```

```
summary(full_data$ferritin.ng.ml.)
```

```
##     Length     Class      Mode
##        126 character character
```

```
#check distribution of categorical covarities
table(full_data$sex, useNA = "always")
```

```
##
##   female     male  unknown     <NA>
##       51       74        1        0
```

```
table(full_data$icu_status, useNA = "always")
```

```
##
##   no  yes <NA>
##   60   66    0
```

Graphs

```
#Histogram of AKK1 expression
ggplot(full_data, aes(x=AAK1))+
  geom_histogram(binwidth = 0.5, fill = 'steelblue', color = 'white', na.rm = TRUE) +
  labs(
    title = "Histogram of AAK1 Gene Expression",
    x = "AAK1 Expression Level",
    y = "Number of Samples"
  ) +
  theme_minimal()
```

Histogram of AAK1 Gene Expression
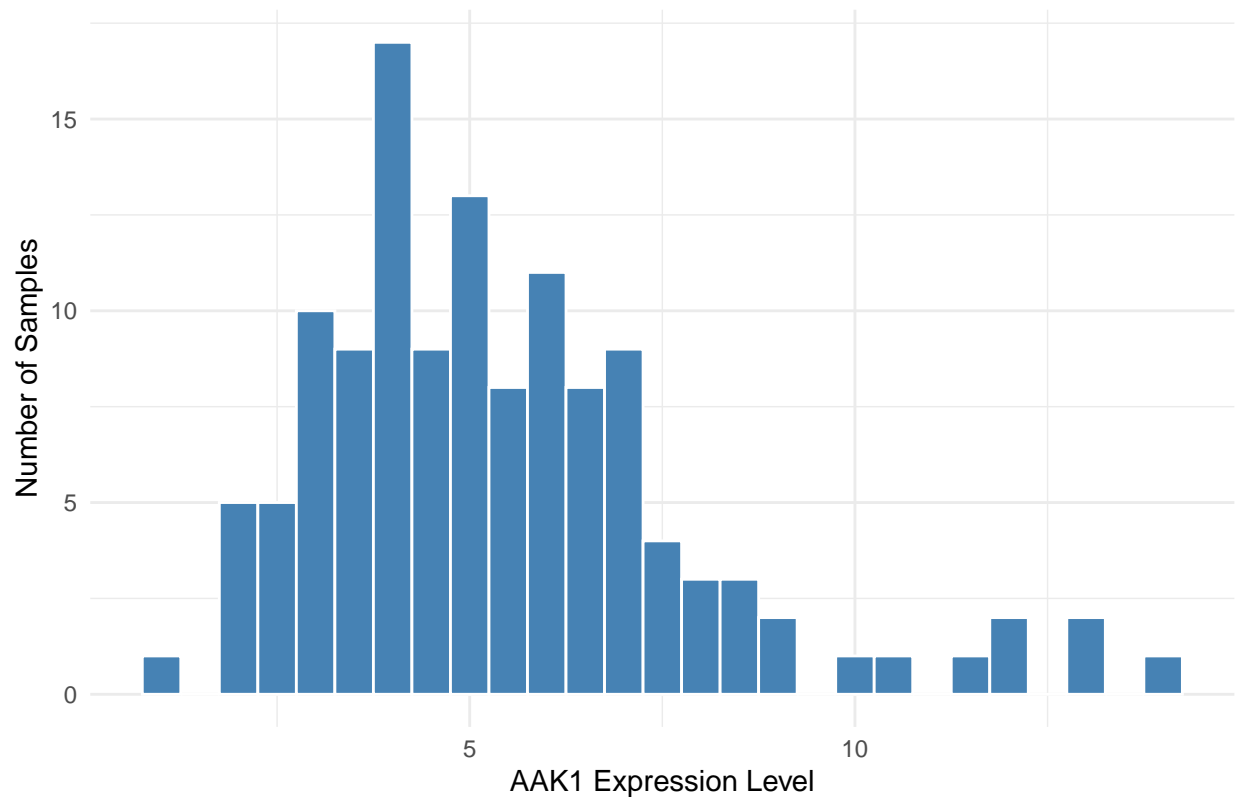

```
#Scattor plots of AKK1 and Ferritin
ggplot(full_data, aes(x=ferritin, y=AAK1, color=icu_status)) +
  geom_point(alpha = 0.7) +
  labs(
    title = "Scatterplot: AAK1 Expression vs. Ferritin Level",
    x = "Ferritin (ng/ml)",
    y = "AAK1 expression"
  ) +
  theme_minimal()
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## ('geom_point()').
```
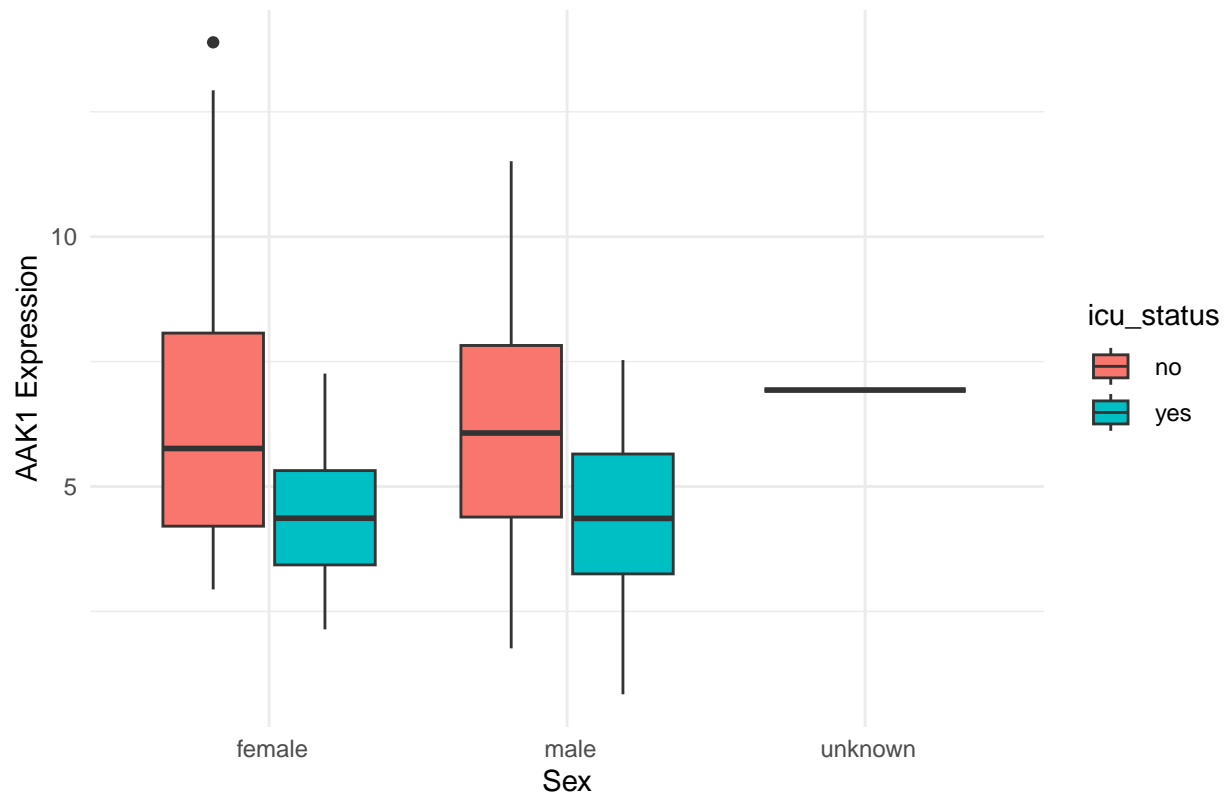
Scatterplot: AAK1 Expression vs. Ferritin Level
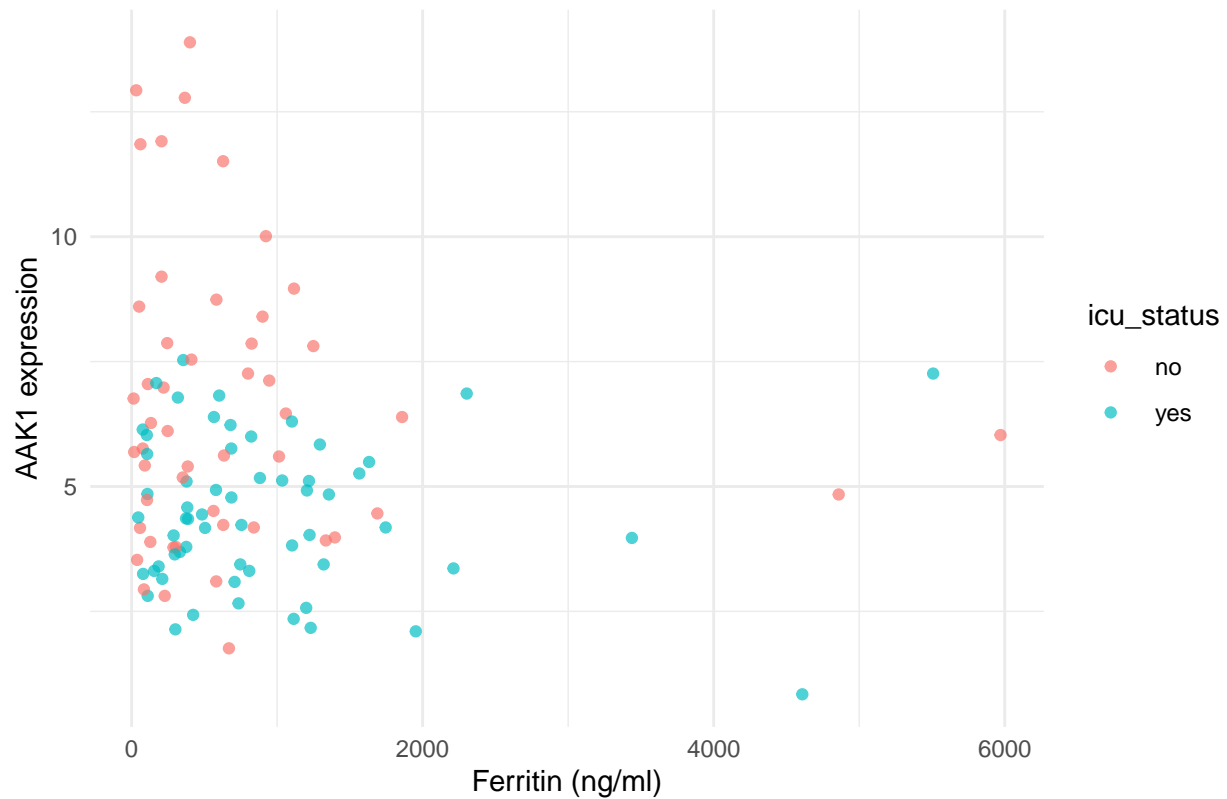
```
#Boxplot of AAK1 by Sex and ICU status
ggplot(full_data, aes(x = sex, y = AAK1, fill = icu_status)) +
  geom_boxplot() +
  labs(title = "AAK1 Expression by Sex and ICU Status",
       x = "Sex",
       y = "AAK1 Expression") +
  theme_minimal()
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## ('stat_boxplot()').
```

AAK1 Expression by Sex and ICU Status

submission2 continue:

```r
library(ggplot2) #import funtion
#generate gene plot
make_gene_plot <- function(df, gene, xvar, cat1, cat2) {
  df <- na.omit(df[, c(gene, cat1, cat2)])  #ignoring the warning of NA value
  ggplot(df, aes_string(x = cat1, y = gene, fill = cat2)) +
    geom_boxplot() +
    labs(
      title = paste(gene, "Expression by", cat1, "and", cat2),
      x = cat1,
      y = paste(gene, "Expression"),
      fill = cat2
    ) +
    theme_minimal()
}
make_gene_plot(full_data, "AAK1", "ferritin", "sex", "icu_status")
```
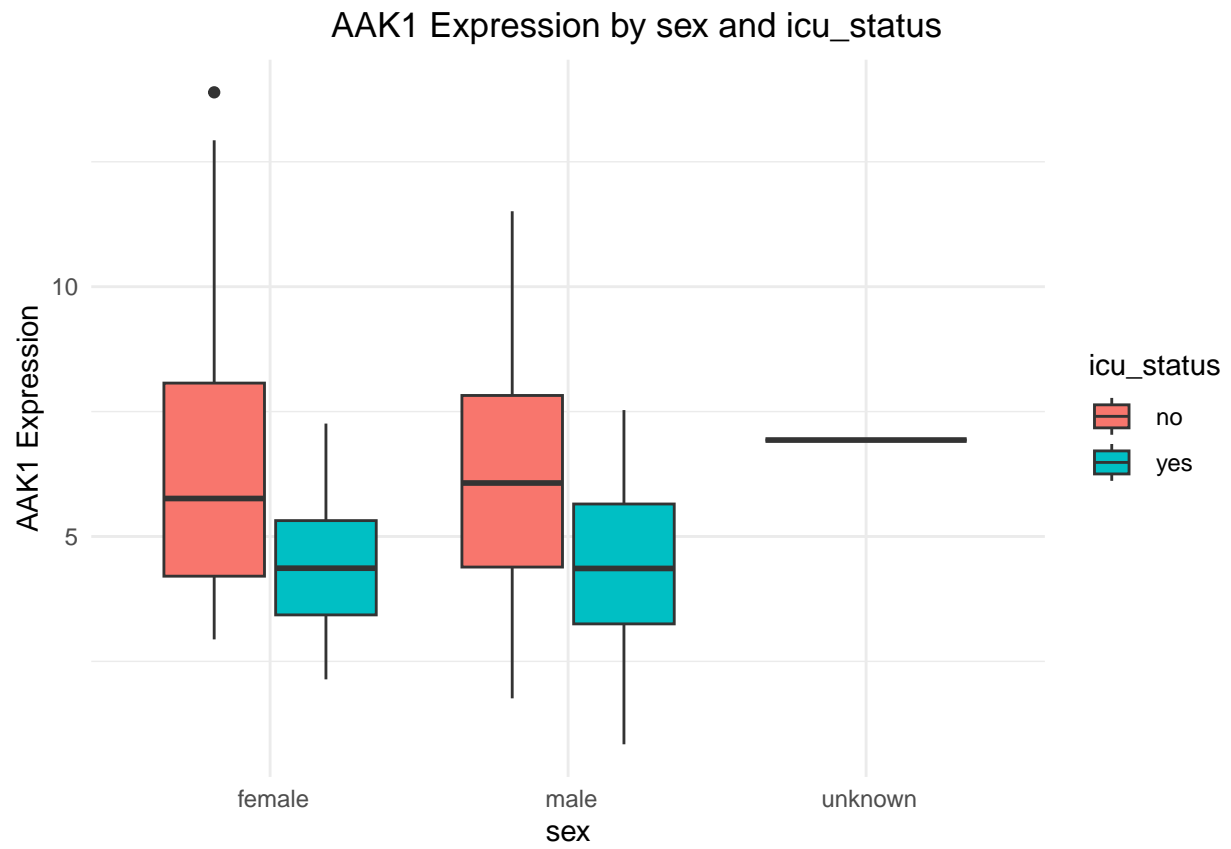
```
## Warning: 'aes_string()' was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with 'aes()'.
## i See also 'vignette("ggplot2-in-packages")' for more information.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

AAK1 Expression by sex and icu_status
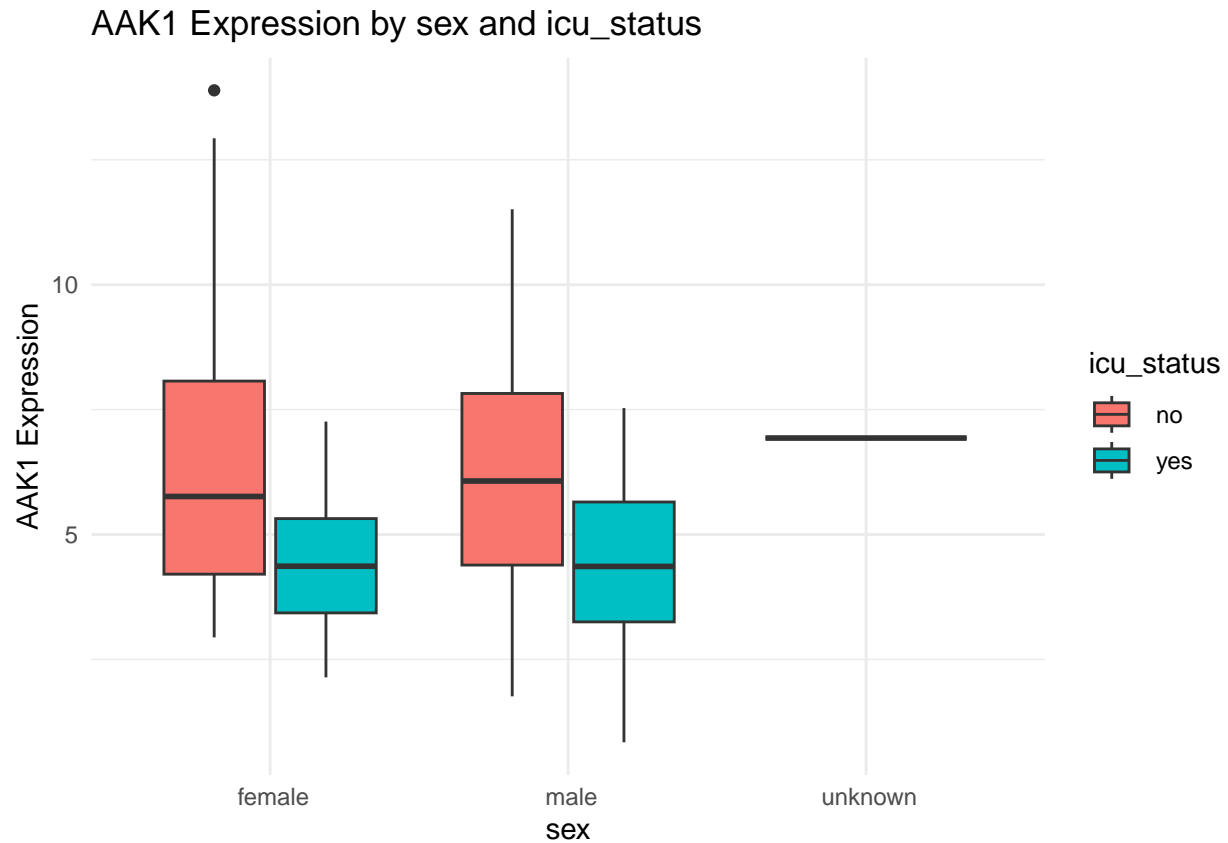
Use for loop to generate 3 different gene plots
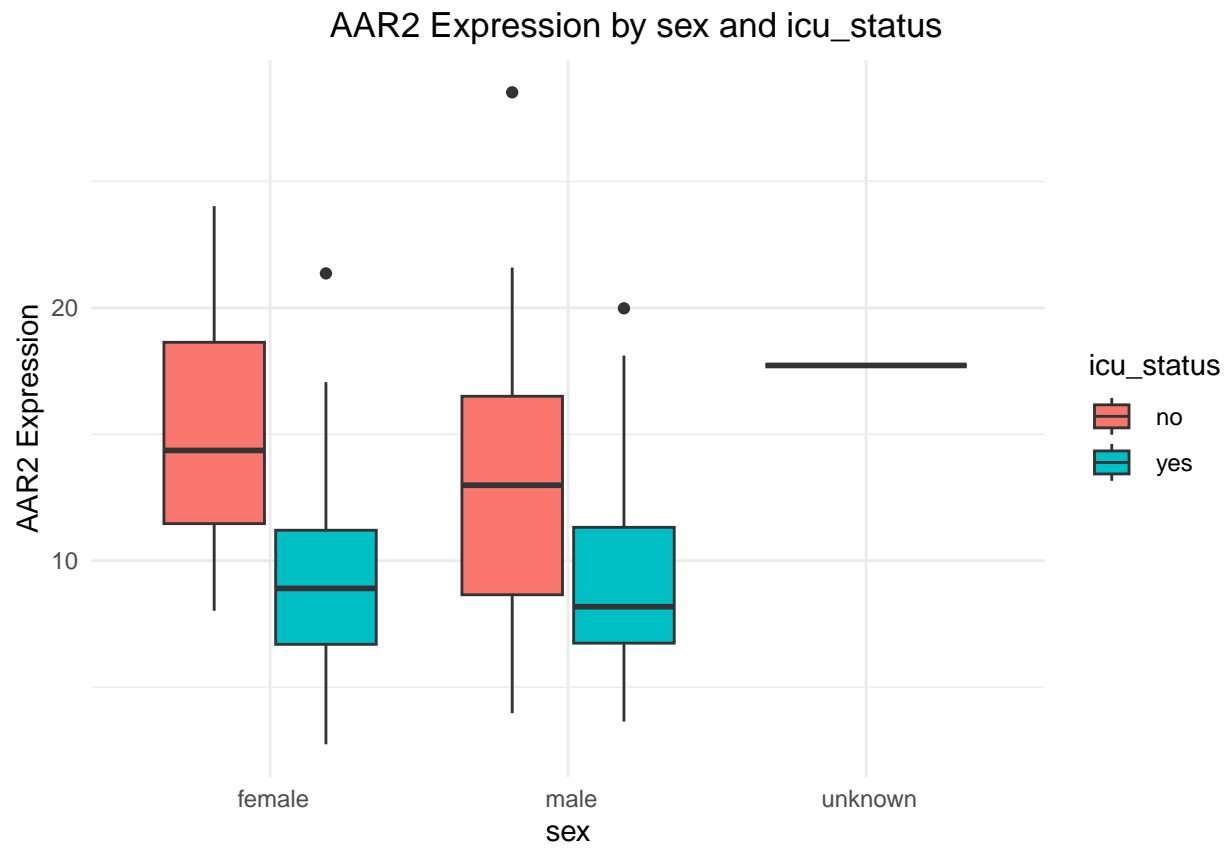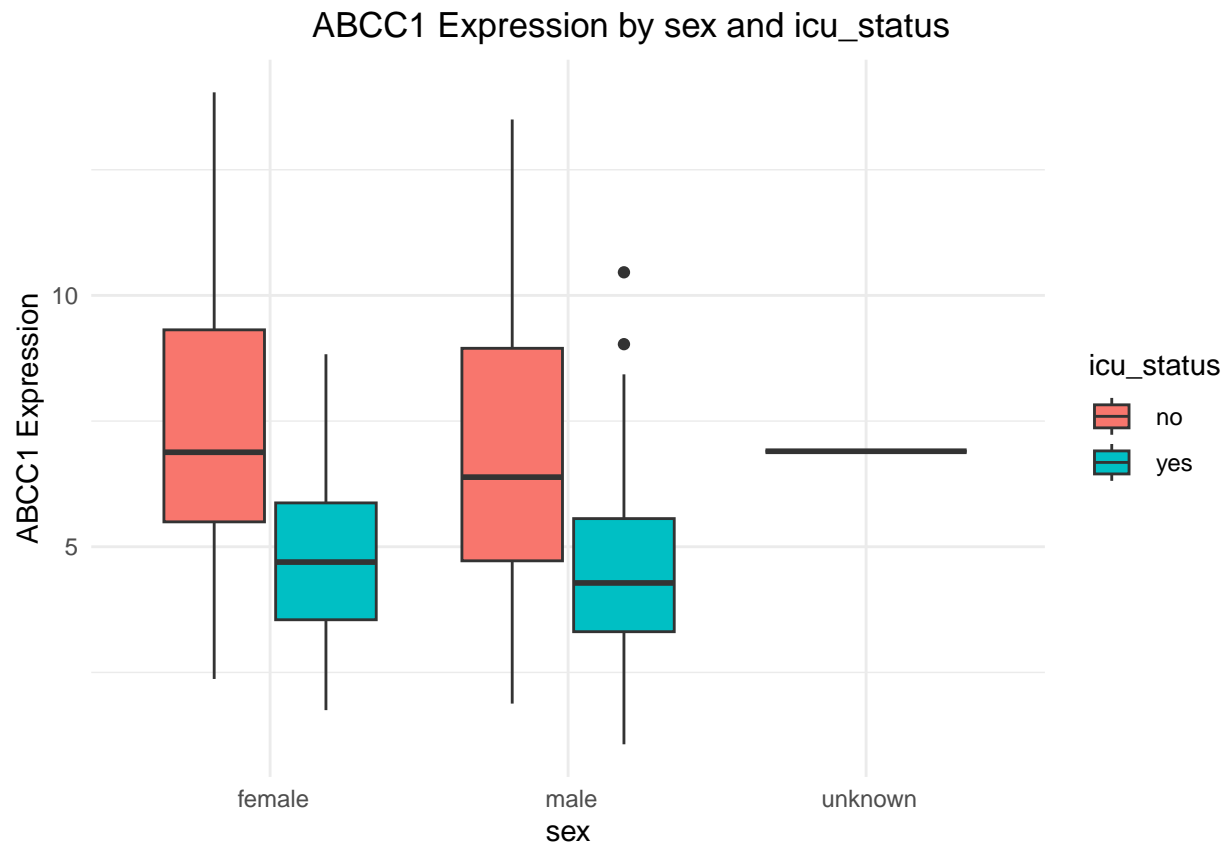
```
genes <- c("AAK1", "AAR2", "ABCC1")

make_gene_plot <- function(df, gene, xvar, cat1, cat2) {
  df <- na.omit(df[, c(gene, cat1, cat2)])
  ggplot(df, aes_string(x = cat1, y = gene, fill = cat2)) +
    geom_boxplot() +
    labs(
      title = paste(gene, "Expression by", cat1, "and", cat2),
      x = cat1,
      y = paste(gene, "Expression"),
      fill = cat2
    ) +
    theme_minimal() +
    theme(plot.title = element_text(hjust = 0.7))
}

for (g in genes) {
  print(make_gene_plot(full_data, g, "ferritin.ng.ml.", "sex", "icu_status"))
}
```

AAK1 Expression by sex and icu_status

AAR2 Expression by sex and icu_status

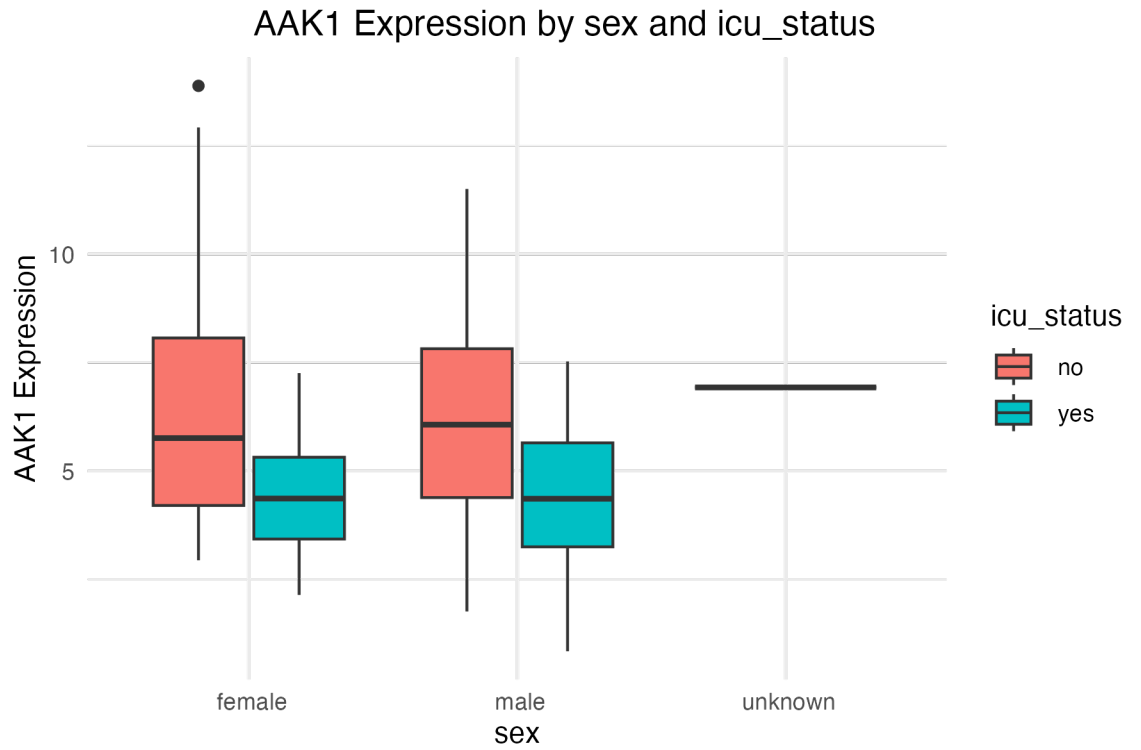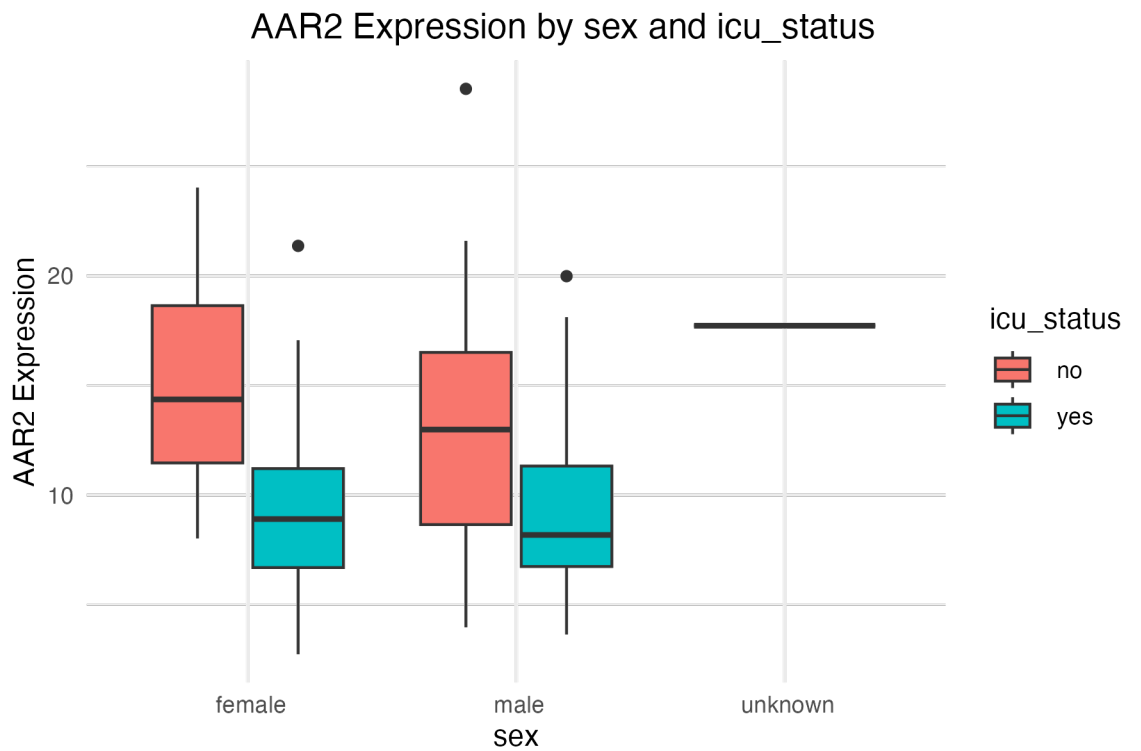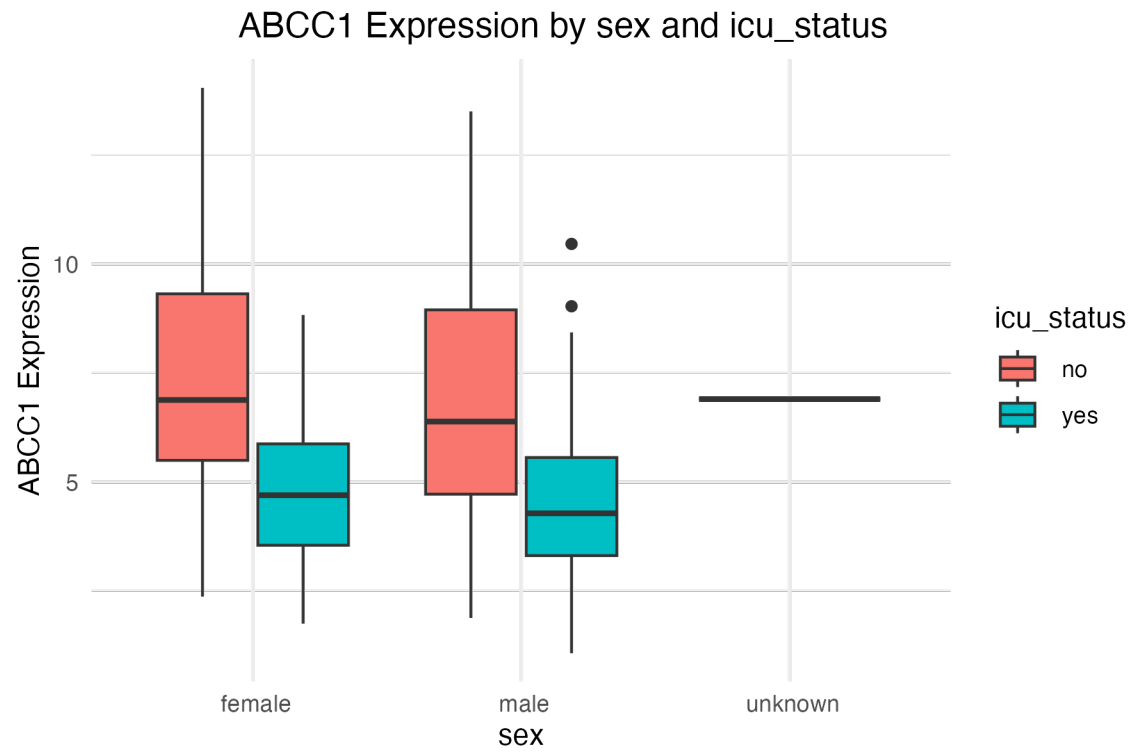## ABCC1 Expression by sex and icu_status

Saving plots

```
for (g in genes) {
  p <- make_gene_plot(full_data, g, "ferritin.ng.ml.", "sex", "icu_status")
  ggsave(filename = paste0(g, "_boxplot.png"), plot = p, width = 6, height = 4)
}
```

AAK1 Boxplot

AAR2 Boxplot

ABCC1 Boxplot