

Final Project

2023-07-24

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:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(ggpubr)
gene_expression <- read.csv(file = "QBS103_finalProject_geneExpression.csv")
meta_data <- read.csv(file = "QBS103_finalProject_metadata.csv")
# Transpose the dataset in order to link the 2 sets of data
gene_transposed <- data.frame(cbind(names(gene_expression), t(gene_expression)))
rownames(gene_transposed) <- NULL
colnames(gene_transposed) <- gene_transposed[1,]
# delete the first redundant row to match the entries
gene_final <- gene_transposed[-1,]
# Create a data frame with the participant ID, gene number 5 (PRTN3), one continuous covariate (age) and
# two categorical covariants (sex and ICU status)
linked_data <- data.frame("Participant_ID" = meta_data$participant_id,
                          "Age" = meta_data$age,
                          "Sex" = meta_data$sex,
                          "ICU_Status" = meta_data$icu_status,
                          "BPI_Gene" = as.numeric(gene_final$BPI))
linked_data
```

```
##           Participant_ID Age Sex ICU_Status BPI_Gene
## 1 COVID_01_39y_male_NonICU 39 male         no    19.05
## 2 COVID_02_63y_male_NonICU 63 male         no   113.34
```

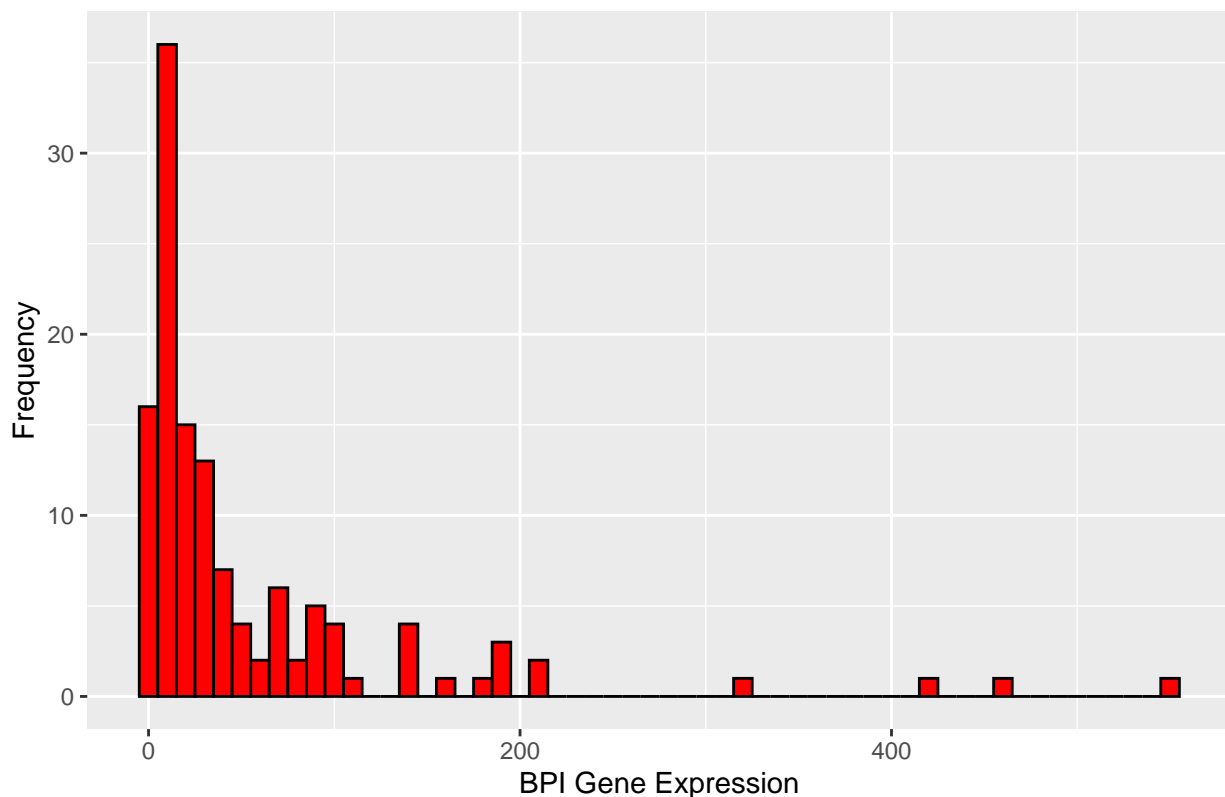
## 3	COVID_03_33y_male_NonICU	33	male	no	30.39
## 4	COVID_04_49y_male_NonICU	49	male	no	14.03
## 5	COVID_05_49y_male_NonICU	49	male	no	8.49
## 6	COVID_06_:y_male_NonICU	:	male	no	97.92
## 7	COVID_07_38y_female_NonICU	38	female	no	24.10
## 8	COVID_08_78y_male_ICU	78	male	yes	456.31
## 9	COVID_09_64y_female_ICU	64	female	yes	33.23
## 10	COVID_10_62y_male_ICU	62	male	yes	417.69
## 11	COVID_11_52y_female_NonICU	52	female	no	1.83
## 12	COVID_12_50y_male_ICU	50	male	yes	104.31
## 13	COVID_13_37y_male_NonICU	37	male	no	6.45
## 14	COVID_14_55y_male_ICU	55	male	yes	33.73
## 15	COVID_15_68y_male_ICU	68	male	yes	65.63
## 16	COVID_16_48y_male_NonICU	48	male	no	44.14
## 17	COVID_17_54y_male_NonICU	54	male	no	53.92
## 18	COVID_18_70y_female_NonICU	70	female	no	75.88
## 19	COVID_19_51y_male_NonICU	51	male	no	16.55
## 20	COVID_20_62y_male_ICU	62	male	yes	73.00
## 21	COVID_21_66y_male_ICU	66	male	yes	10.10
## 22	COVID_22_43y_male_ICU	43	male	yes	190.51
## 23	COVID_23_76y_male_ICU	76	male	yes	54.36
## 24	COVID_24_55y_male_ICU	55	male	yes	30.74
## 25	COVID_25_55y_male_ICU	55	male	yes	553.29
## 26	COVID_26_41y_female_ICU	41	female	yes	9.00
## 27	COVID_27_71y_female_ICU	71	female	yes	206.25
## 28	COVID_28_63y_male_ICU	63	male	yes	22.36
## 29	COVID_29_63y_female_ICU	63	female	yes	3.75
## 30	COVID_30_54y_male_ICU	54	male	yes	9.06
## 31	COVID_31_50y_male_ICU	50	male	yes	83.40
## 32	COVID_32_72y_male_ICU	72	male	yes	214.47
## 33	COVID_33_81y_male_NonICU	81	male	no	29.98
## 34	COVID_34_64y_female_NonICU	64	female	no	24.31
## 35	COVID_35_58y_female_NonICU	58	female	no	62.58
## 36	COVID_36_68y_male_NonICU	68	male	no	86.40
## 37	COVID_37_87y_male_NonICU	87	male	no	8.35
## 38	COVID_38_68y_male_ICU	68	male	yes	3.72
## 39	COVID_39_80y_female_ICU	80	female	yes	63.50
## 40	COVID_40_66y_male_ICU	66	male	yes	27.03
## 41	COVID_41_74y_male_ICU	74	male	yes	7.33
## 42	COVID_42_21y_female_ICU	21	female	yes	65.84
## 43	COVID_43_83y_female_ICU	83	female	yes	185.08
## 44	COVID_44_46y_male_ICU	46	male	yes	27.62
## 45	COVID_45_62y_female_ICU	62	female	yes	9.09
## 46	COVID_46_62y_male_ICU	62	male	yes	5.41
## 47	COVID_47_78y_male_ICU	78	male	yes	27.87
## 48	COVID_48_72y_female_ICU	72	female	yes	187.36
## 49	COVID_49_73y_male_ICU	73	male	yes	88.72
## 50	COVID_50_37y_male_ICU	37	male	yes	139.29
## 51	COVID_51_58y_female_NonICU	58	female	no	163.95
## 52	COVID_52_71y_male_NonICU	71	male	no	9.09
## 53	COVID_53_35y_female_NonICU	35	female	no	69.93
## 54	COVID_55_62y_female_ICU	62	female	yes	92.26
## 55	COVID_56_33y_female_NonICU	33	female	no	22.91
## 56	COVID_57_30y_female_NonICU	30	female	no	35.75

## 57	COVID_58_62y_male_NonICU	62	male	no	36.09
## 58	COVID_59_55y_male_NonICU	55	male	no	4.63
## 59	COVID_60_49y_male_NonICU	49	male	no	43.53
## 60	COVID_61_54y_female_NonICU	54	female	no	7.62
## 61	COVID_62_78y_female_ICU	78	female	yes	85.81
## 62	COVID_63_39y_female_ICU	39	female	yes	21.36
## 63	COVID_64_65y_male_ICU	65	male	yes	50.82
## 64	COVID_65_84y_male_NonICU	84	male	no	10.75
## 65	COVID_66_66y_female_NonICU	66	female	no	16.47
## 66	COVID_67_57y_male_ICU	57	male	yes	15.83
## 67	COVID_68_79y_male_ICU	79	male	yes	140.00
## 68	COVID_69_77y_female_NonICU	77	female	no	15.35
## 69	COVID_70_81y_male_NonICU	81	male	no	7.22
## 70	COVID_71_37y_male_ICU	37	male	yes	316.72
## 71	COVID_72_50y_female_NonICU	50	female	no	8.46
## 72	COVID_73_82y_male_NonICU	82	male	no	10.64
## 73	COVID_74_55y_female_ICU	55	female	yes	144.28
## 74	COVID_75_55y_male_NonICU	55	male	no	9.36
## 75	COVID_76_73y_female_ICU	73	female	yes	72.87
## 76	COVID_77_55y_female_ICU	55	female	yes	28.84
## 77	COVID_78_80y_male_NonICU	80	male	no	3.64
## 78	COVID_79_27y_male_NonICU	27	male	no	12.69
## 79	COVID_80_71y_male_ICU	71	male	yes	54.66
## 80	COVID_82_67y_male_NonICU	67	male	no	7.01
## 81	COVID_83_85y_female_NonICU	85	female	no	36.33
## 82	COVID_84_75y_female_NonICU	75	female	no	38.29
## 83	COVID_85_62y_male_ICU	62	male	yes	0.85
## 84	COVID_86_52y_female_NonICU	52	female	no	12.21
## 85	COVID_87_61y_male_ICU	61	male	yes	36.77
## 86	COVID_89_90y_female_NonICU	>89	female	no	21.28
## 87	COVID_90_86y_female_NonICU	86	female	no	16.03
## 88	COVID_91_29y_female_NonICU	29	female	no	8.75
## 89	COVID_92_82y_female_ICU	82	female	yes	177.09
## 90	COVID_93_81y_female_ICU	81	female	yes	5.63
## 91	COVID_94_24y_female_NonICU	24	female	no	1.33
## 92	COVID_95_49y_male_NonICU	49	male	no	34.68
## 93	COVID_96_51y_male_NonICU	51	male	no	2.69
## 94	COVID_97_76y_male_ICU	76	male	yes	10.88
## 95	COVID_98_81y_male_NonICU	81	male	no	135.96
## 96	COVID_99_71y_male_ICU	71	male	yes	2.47
## 97	COVID_100_74y_female_NonICU	74	female	no	73.98
## 98	COVID_101_58y_male_ICU	58	male	yes	96.72
## 99	COVID_102_84y_male_NonICU	84	male	no	88.54
## 100	COVID_103_83y_male_NonICU	83	male	no	30.79
## 101	NONCOVID_01_54y_female_NonICU	54	female	no	5.43
## 102	NONCOVID_02_65y_male_ICU	65	male	yes	25.15
## 103	NONCOVID_03_65y_male_ICU	65	male	yes	13.71
## 104	NONCOVID_04_90y_male_NonICU	>89	male	no	10.04
## 105	NONCOVID_05_83y_female_NonICU	83	female	no	9.55
## 106	NONCOVID_06_75y_female_ICU	75	female	yes	2.95
## 107	NONCOVID_07_50y_male_ICU	50	male	yes	5.17
## 108	NONCOVID_08_53y_female_ICU	53	female	yes	20.94
## 109	NONCOVID_09_49y_female_NonICU	49	female	no	8.09
## 110	NONCOVID_10_67y_male_ICU	67	male	yes	20.08

```
## 111 NONCOVID_11_58y_female_NonICU 58 female no 14.95
## 112 NONCOVID_12_82y_male_ICU 82 male yes 8.61
## 113 NONCOVID_13_65y_male_ICU 65 male yes 5.93
## 114 NONCOVID_14_75y_female_ICU 75 female yes 2.15
## 115 NONCOVID_15_83y_unknown_ICU 83 unknown yes 4.98
## 116 NONCOVID_16_40y_female_ICU 40 female yes 99.79
## 117 NONCOVID_17_84y_female_ICU 84 female yes 8.16
## 118 NONCOVID_18_88y_male_ICU 88 male yes 18.56
## 119 NONCOVID_19_66y_female_ICU 66 female yes 4.65
## 120 NONCOVID_20_62y_female_ICU 62 female yes 11.13
## 121 NONCOVID_21_71y_male_NonICU 71 male no 6.69
## 122 NONCOVID_22_63y_male_NonICU 63 male no 3.84
## 123 NONCOVID_23_42y_female_NonICU 42 female no 2.12
## 124 NONCOVID_24_32y_female_NonICU 32 female no 6.09
## 125 NONCOVID_25_62y_male_NonICU 62 male no 1.49
## 126 NONCOVID_26_36y_male_ICU 36 male yes 32.10
```

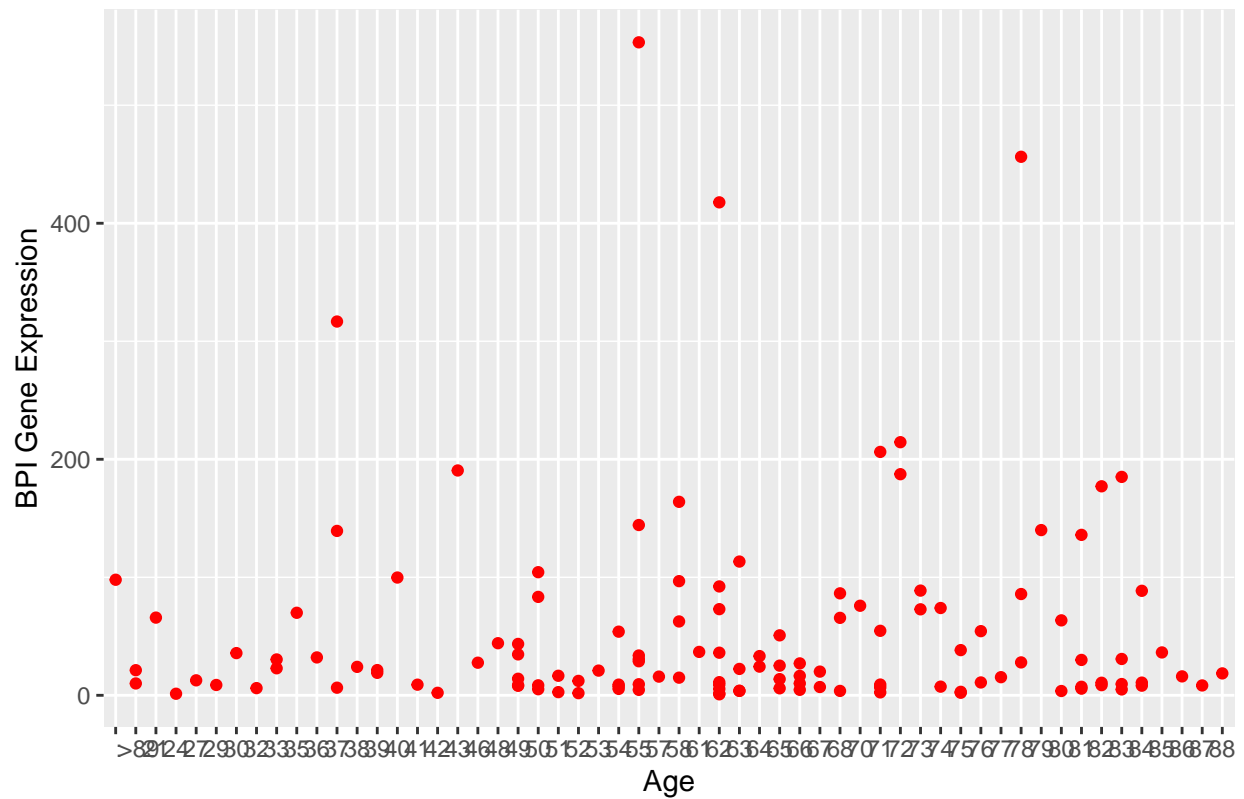
```
# Histogram for gene expression
ggplot(linked_data, aes(x = BPI_Gene)) +
  geom_histogram(binwidth = 10, fill = "red", color = "black") +
  labs(title = "Histogram for Gene Expression", x = "BPI Gene Expression", y = "Frequency")
```

Histogram for Gene Expression



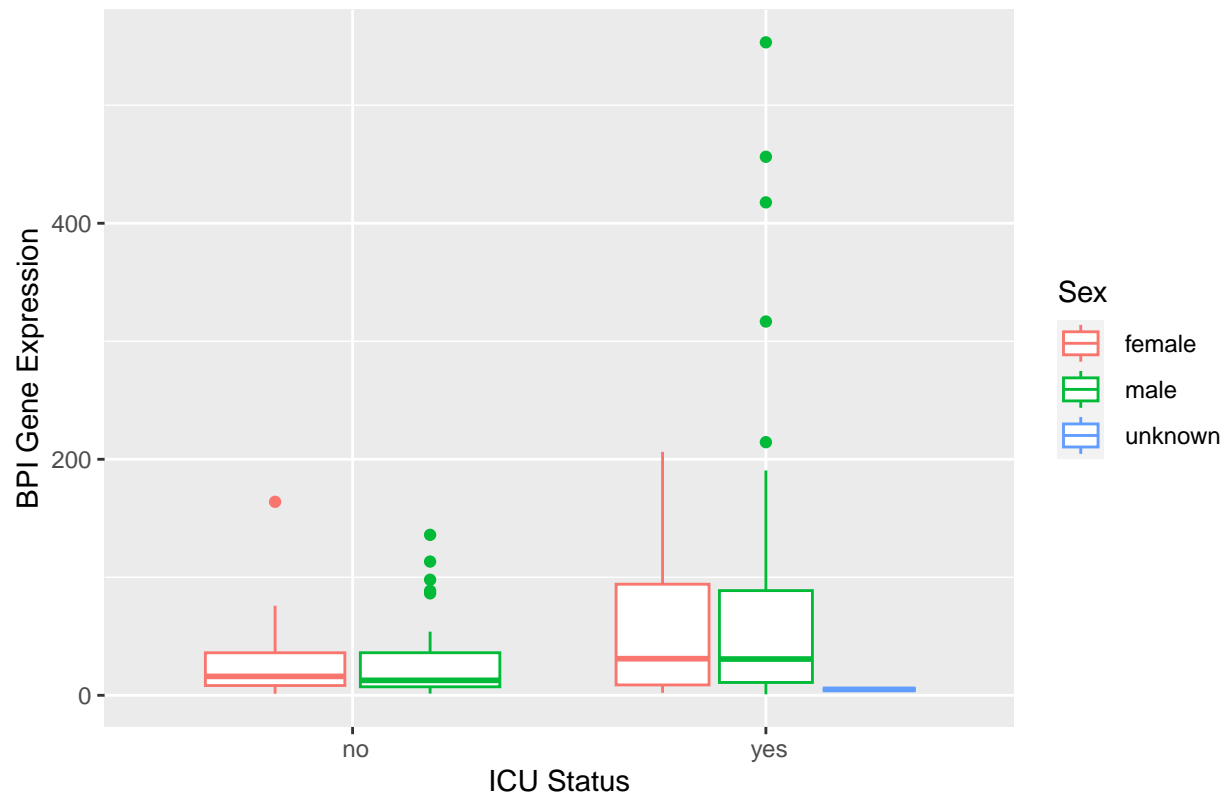
```
# Scatter plot
ggplot(linked_data, aes(x = Age, y = BPI_Gene)) +
  geom_point(color = "red") +
  labs(title = "Scatterplot for Gene Expression and Continuous Covariat", y = "BPI Gene Expression")
```

Scatterplot for Gene Expression and Continuous Covariat



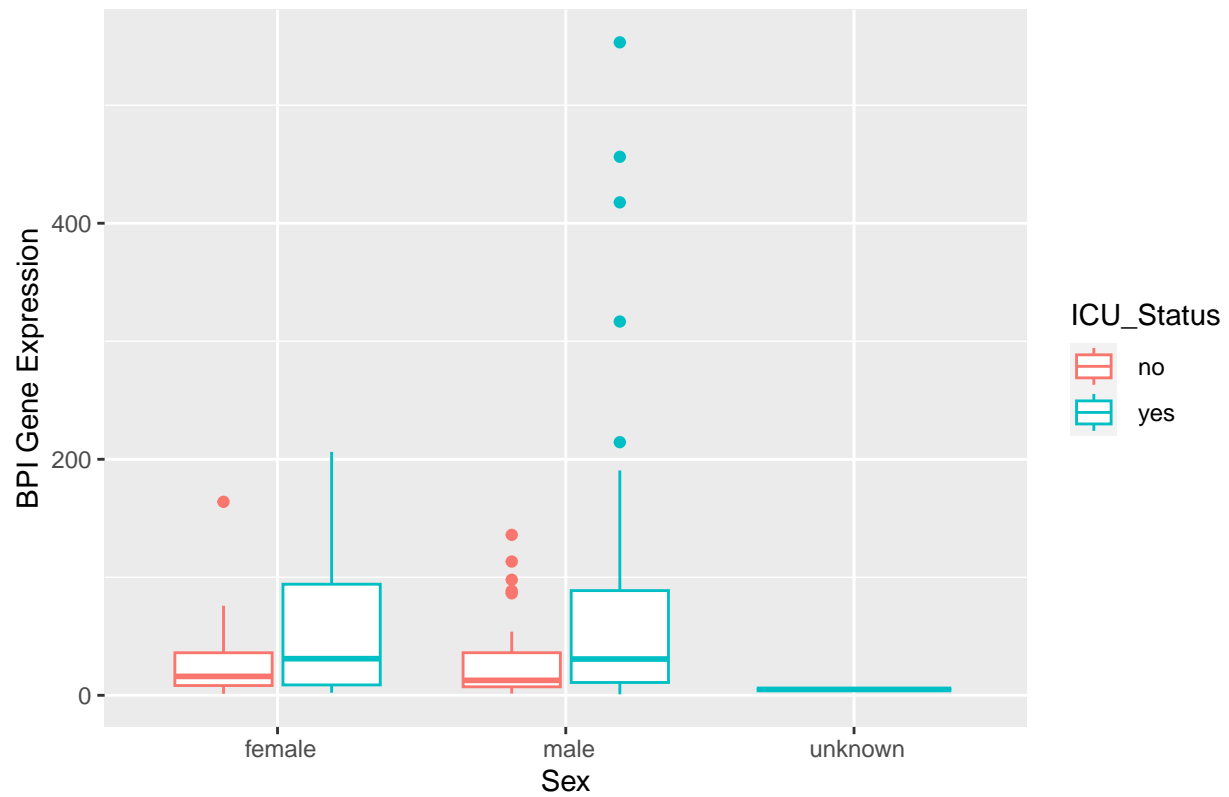
```
# Box plot
icuplot <- ggplot(linked_data, aes(x = ICU_Status, y = BPI_Gene, color = Sex)) +
  geom_boxplot() +
  labs(title = "Gene Separated by ICU Status", x = "ICU Status", y = "BPI Gene Expression")
sexplot <- ggplot(linked_data, aes(x = Sex, y = BPI_Gene, color = ICU_Status)) +
  geom_boxplot() +
  labs(title = "Gene Separated by Biological Sex", y = "BPI Gene Expression")
icuplot
```

Gene Separated by ICU Status



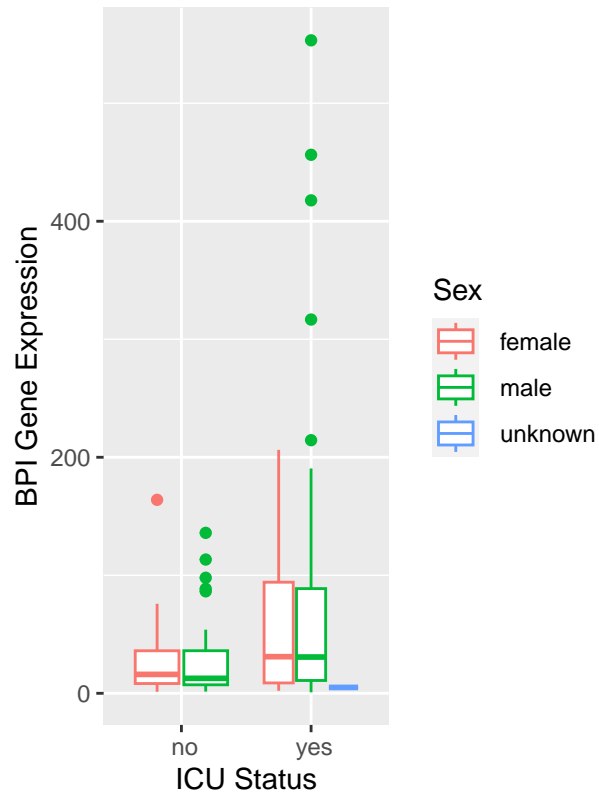
sexplot

Gene Separated by Biological Sex



```
ggarrange(icuplot, sexplot, labels = c("A", "B"))
```

A Gene Separated by ICU Status



B Gene Separated by Biological Sex

