

Project

2024-07-30

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(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    3.5.1      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
```

```
genes <- read.csv("QBS103_GSE157103_genes.csv")
```

```
matrix <- read.csv("QBS103_GSE157103_series_matrix.csv")
```

```
num_sample = ncol(genes) - 1
```

```
num_gene = nrow(genes)
```

```
num_sample
```

```
## [1] 126
```

```
num_gene
```

```
## [1] 100
```

Select two datas

```
### Select the AADACL3 data, and I will try to convert it from row to column
```

```
genes_long <- genes %>%
```

```
  filter(X=="AAGAB") %>%
```

```
  gather(key = "participant_id")
```

```
genes_long <- genes_long[-1,]
```

```
### Selected the data, here I select charlson_score for continuous, icu_status and mechanical ventilation
```

```
matrix_selected <- matrix %>% select("participant_id", "charlson_score", "icu_status", "mechanical_ventilation")
```

```
plot_data <- merge(matrix_selected, genes_long, by = "participant_id")
```

```
plot_data$value <- as.numeric(as.character(plot_data$value))
```

```
plot_data$charlson_score <- as.numeric(as.character(plot_data$charlson_score))
```

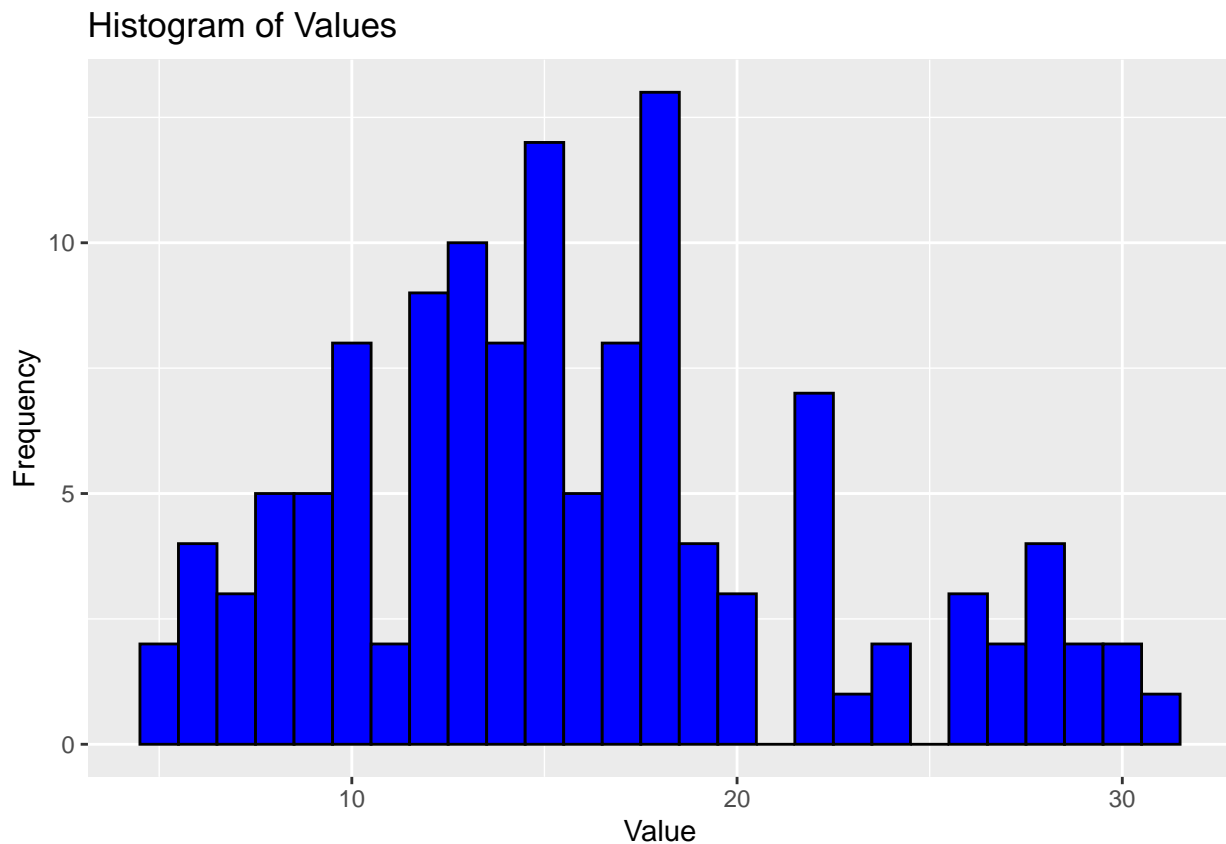
```
library(ggplot2)
# Define theme
newTheme <- theme(
  # Remove all borders and grid lines
  panel.border = element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
  # Set axis colors and size
  axis.line = element_line(colour = "white", linewidth = rel(2)),
  # Change all colors to light blue
  plot.background = element_rect(fill = "black"),
  panel.background = element_blank(),
  legend.key = element_rect(fill = 'black'),
  legend.background = element_rect(fill = 'black'),
  # Set font
  text = element_text(face="bold",colour = "white"),
  axis.text = element_text(face="bold",colour = "white"),
  legend.position = 'bottom')

```

Histogram

```
ggplot(plot_data, aes(x=value)) +
  geom_histogram(binwidth=1, fill="blue", color="black") +
  ggtitle("Histogram of Values") +
  xlab("Value") +
  ylab("Frequency")

```

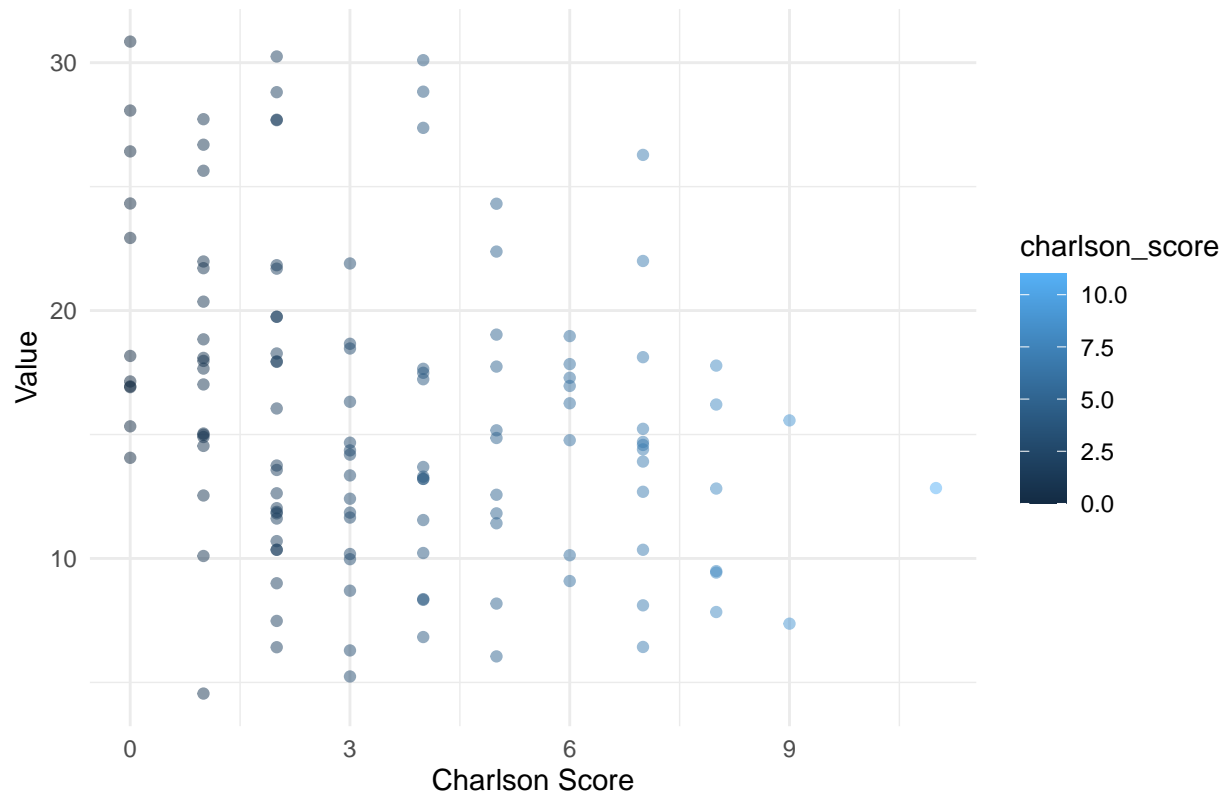


```
ggplot(plot_data, aes(x=charlson_score, y=value)) +
  geom_point(aes(color=charlson_score), alpha=0.5) +
  ggtitle("Scatterplot of AAGAB gene vs. Charlson Score") +

```

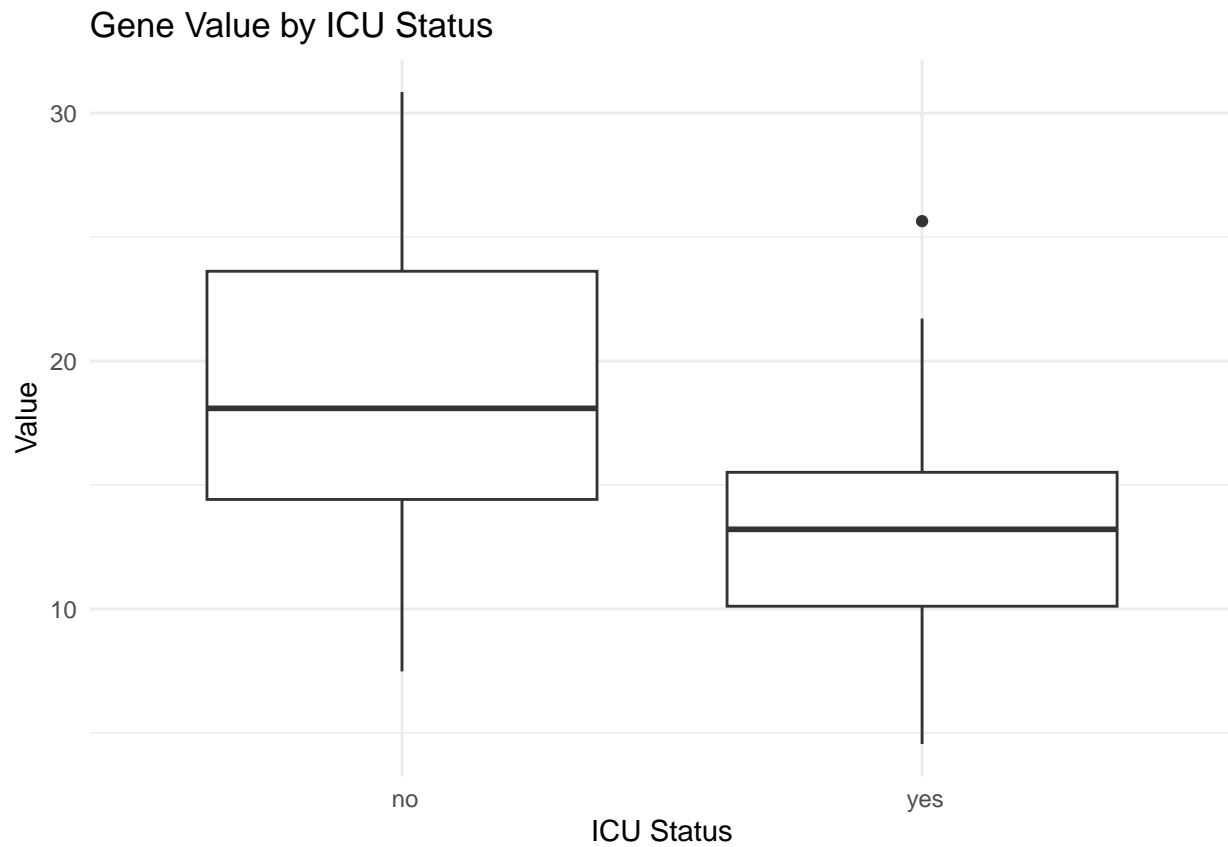
```
xlab("Charlson Score") +  
ylab("Value") +  
theme_minimal()
```

Scatterplot of AAGAB gene vs. Charlson Score



Box plot for icu_status and gene value

```
ggplot(plot_data, aes(x = icu_status, y = value)) +  
  geom_boxplot() +  
  ggtitle("Gene Value by ICU Status") +  
  xlab("ICU Status") +  
  ylab("Value") +  
  theme_minimal()
```



Box plot for mechanical_ventilation and gene value

```
ggplot(plot_data, aes(x = mechanical_ventilation, y = value)) +  
  geom_boxplot() +  
  ggtitle("Gene Value by mechanical_ventilation") +  
  xlab("mechanical_ventilation") +  
  ylab("Gene Value") +  
  theme_minimal()
```

