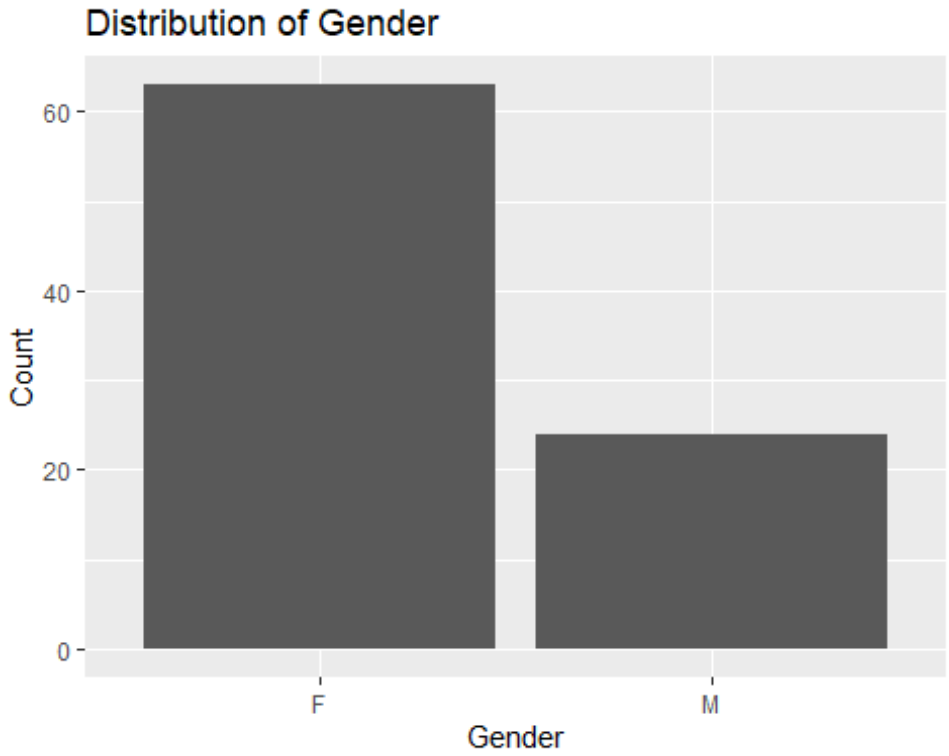
Greg Finkelberg: HW02

R Packages

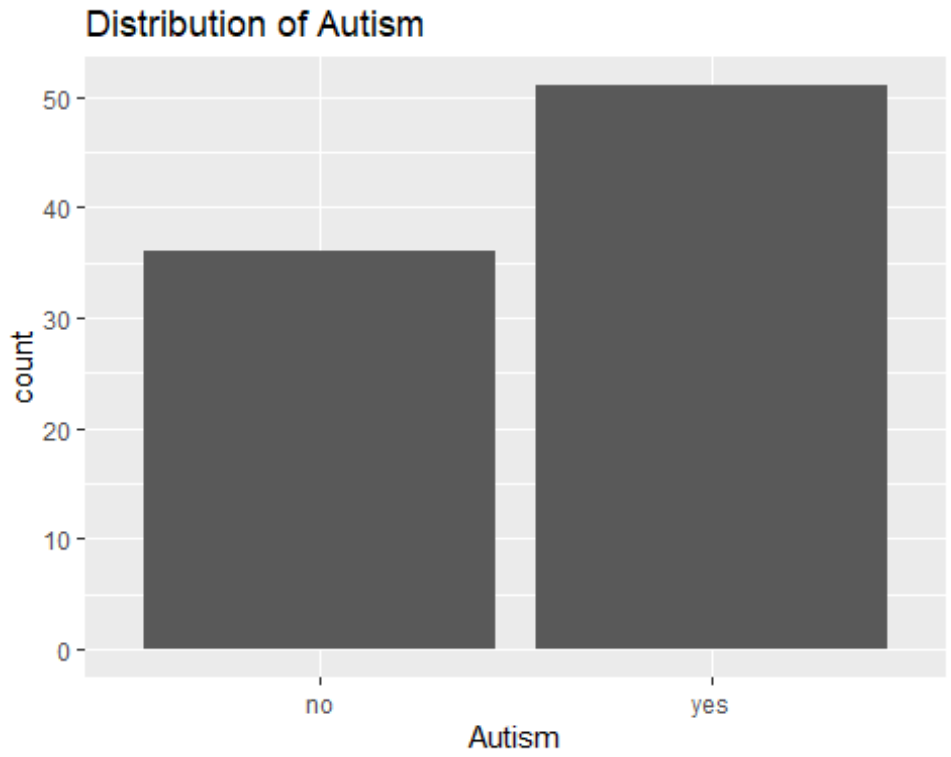
```
library(tidyverse)
library(mosaic)
library(quarto)
```

1. Bar plots of qualitative variables

```
data1_LSC598 <- read.csv("C:/Users/Greg/Downloads/data1_LSC598.txt", sep="")

df <- data1_LSC598 %>% mutate(autism = case_when(group == '0' ~ 'no', group
== '1' ~ "yes"))

ggplot(df, aes(x = gender)) + geom_bar(stat = "count") + labs(title =
"Distribution of Gender", x = "Gender", y = "Count")
```



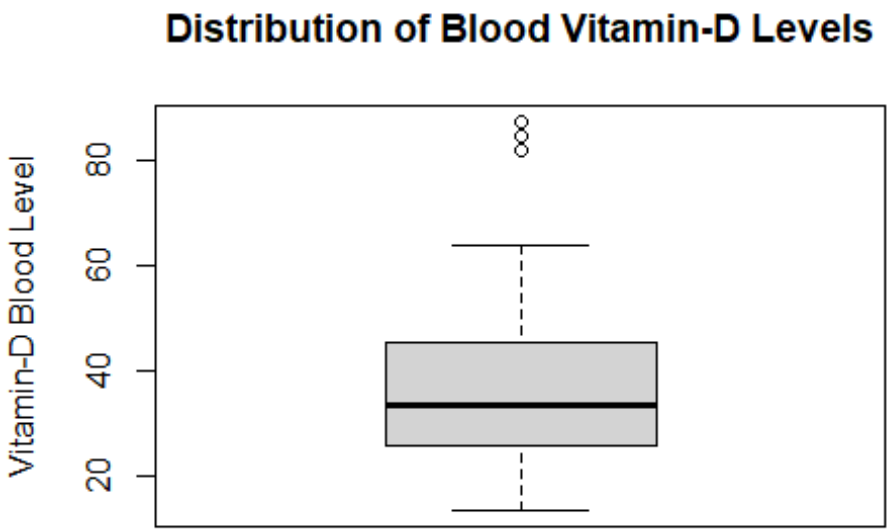

```
ggplot(df, aes(x = autism)) + geom_bar(stat = "count") + labs(title =
"Distribution of Autism", x = "Autism")
```



2. Box plots of quantitative variables

```
boxplot(df$age_month, ylab = "Age (months)", main = "Distribution of Age")
```

```
boxplot(df$vitD_level, ylab = "Vitamin-D Blood Level", main = "Distribution
of Blood Vitamin-D Levels")
```



3. One-sample t-test for continuous variable

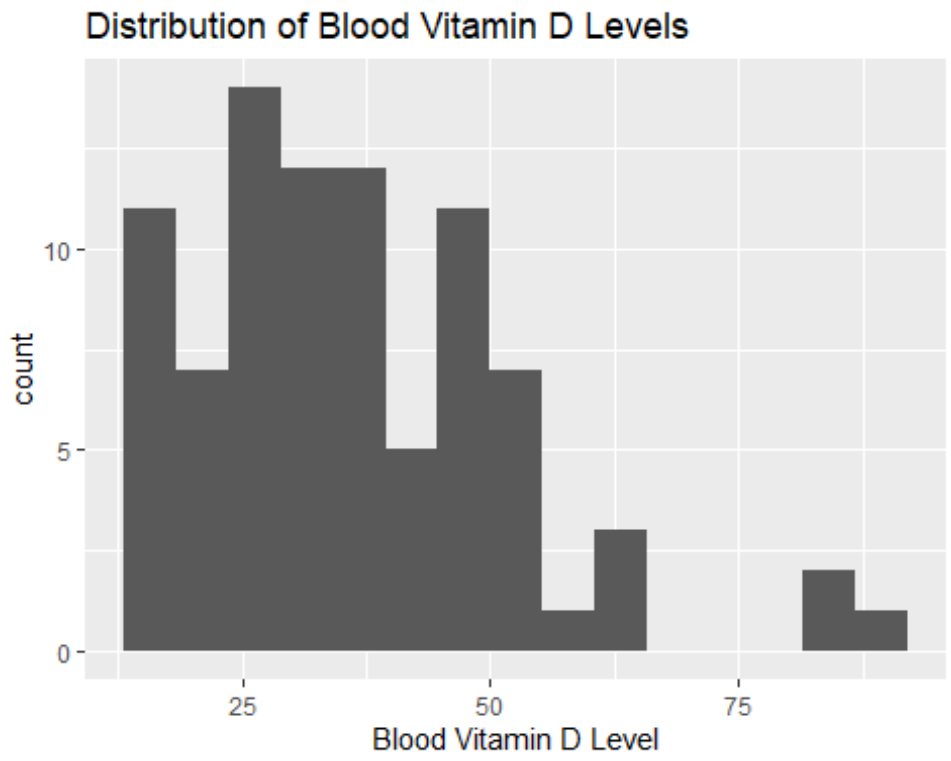
Null hypothesis: The true mean of blood vitamin-D level is equal to 0.

Alternative hypothesis: True mean is not equal to 0.

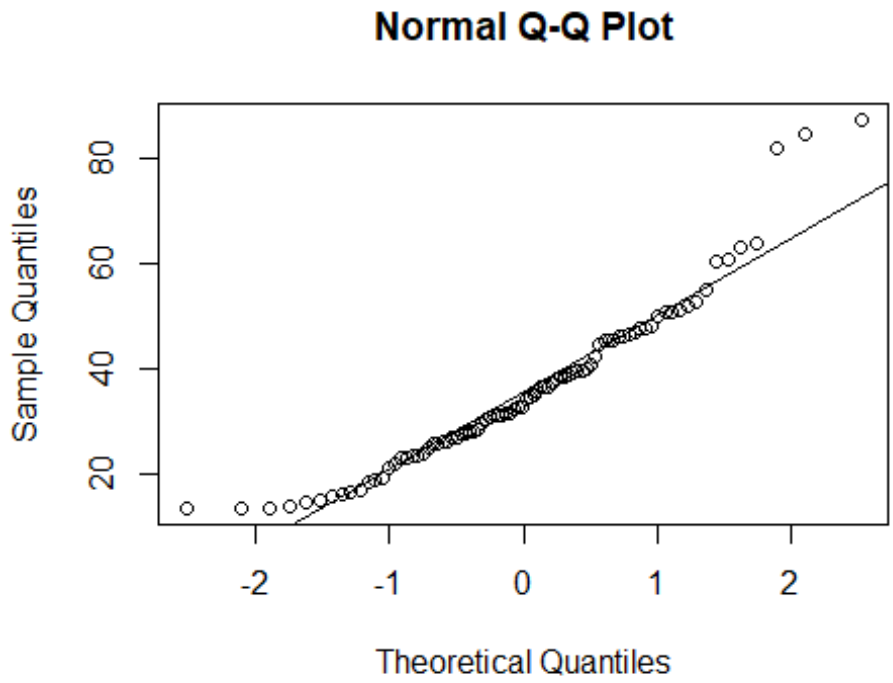
```
vitD <- df %>% drop_na(vitD_level)

fav_stats(vitD$vitD_level)
```

```
  min     Q1 median     Q3  max     mean       sd  n missing
 13.6 25.775   33.4 45.475 87.2 35.96395 15.57167 86       0
```

```
ggplot(vitD, aes(x = vitD_level)) + geom_histogram(bins = 15) + labs(title =
"Distribution of Blood Vitamin D Levels", x = "Blood Vitamin D Level")
```




```
qqnorm(vitD$vitD_level)
qqline(vitD$vitD_level)
```



Data is not normally distributed, but the sample is greater than 30 so we can go ahead with the t-test.

```
t.test(vitD$vitD_level, mu = 0)
```

```
    One Sample t-test

data:  vitD$vitD_level
t = 21.418, df = 85, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 32.62538 39.30253
sample estimates:
mean of x
 35.96395
```

The very small p-value means we reject the null hypothesis as the mean vitamin-D blood level does not equal 0.