LSC541 Module 2: Assignment

Daniel Bihnam

Read in Data

```
getwd()
```

```
[1] "C:/Users/Daniel/Desktop/BioCoding/LSC541/Module 2"
```

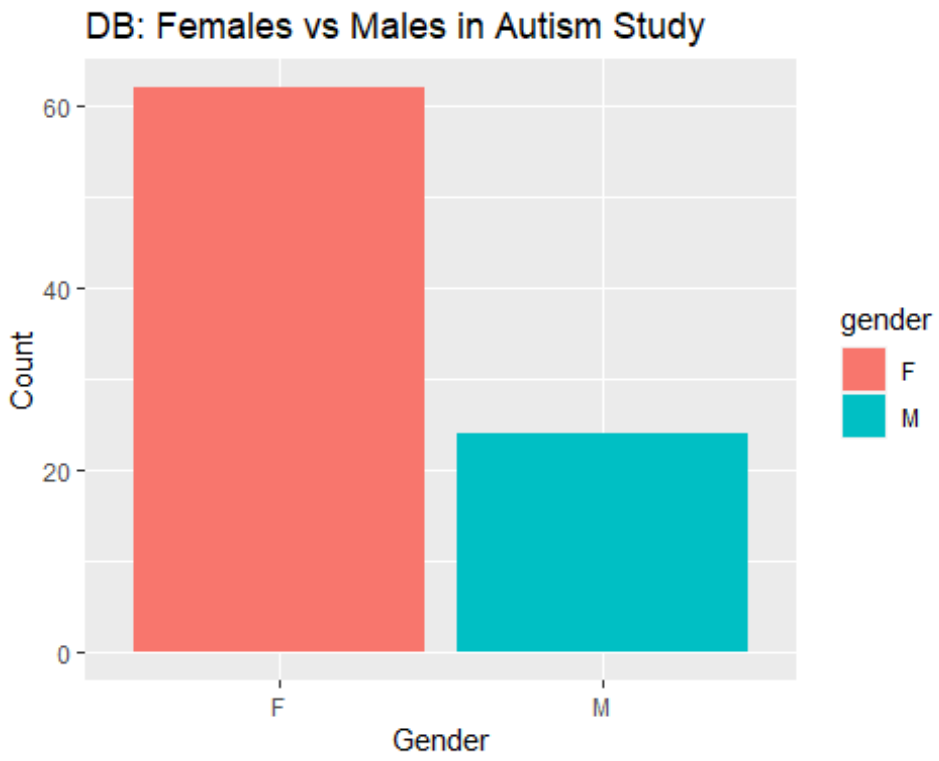
```
setwd("C:/Users/Daniel/Documents")
autism0 <- read.table('C:/Users/Daniel/Downloads/data1_LSC598.txt', header =
T)
autism <- na.omit(autism0)
#Reading in data and omitting a N/A value from the data
```

Qualitative Data

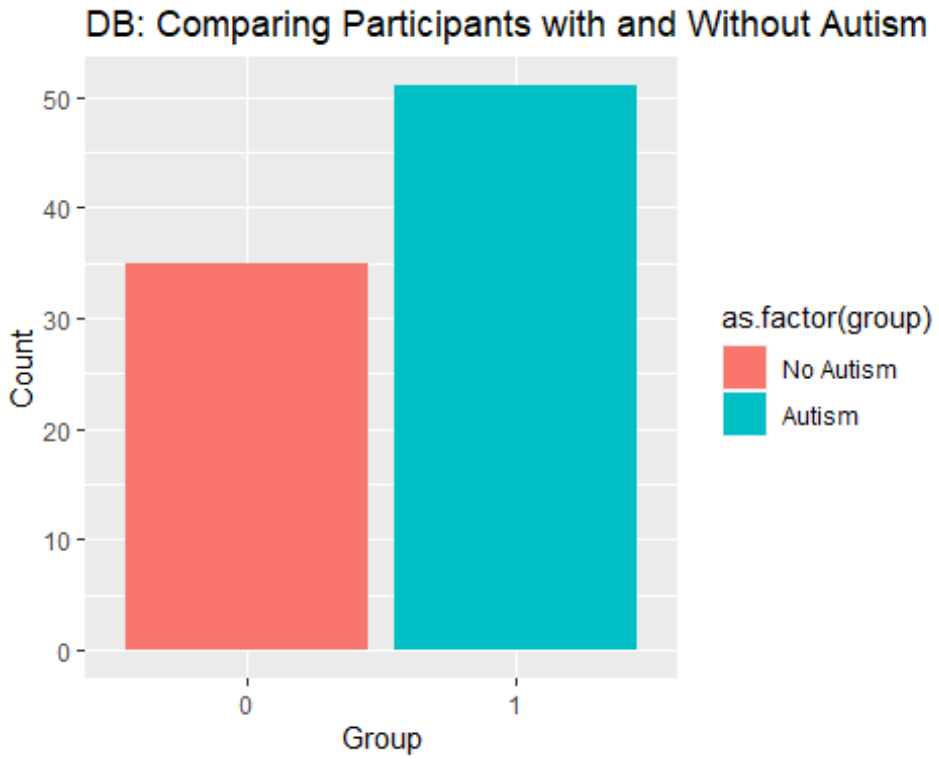
```
library(ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.3.1
```

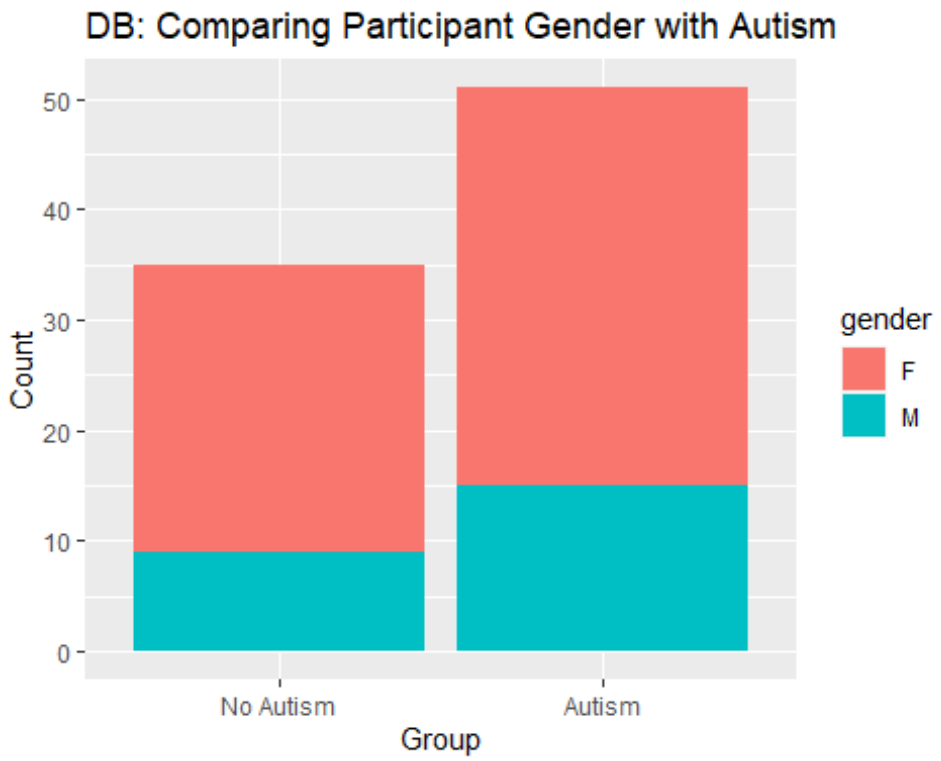
```
ggplot(autism,aes(x=gender,fill=gender))+
  geom_bar()+
  xlab('Gender')+
  ylab('Count')+
  ggtitle('DB: Females vs Males in Autism Study')
```



```r
ggplot(autism,aes(x=as.factor(group),fill=as.factor(group)))+
  geom_bar()+
  xlab('Group')+
  ylab('Count')+
  ggtitle('DB: Comparing Participants with and Without Autism')+
  scale_fill_discrete(labels=c('No Autism','Autism'))
```




```r
ggplot(autism,aes(x=as.factor(group),fill=gender))+
  geom_bar()+
  xlab('Group')+
  ylab('Count')+
  ggtitle('DB: Comparing Participant Gender with Autism')+
  scale_x_discrete(labels=c('0'='No Autism','1'='Autism'))
```



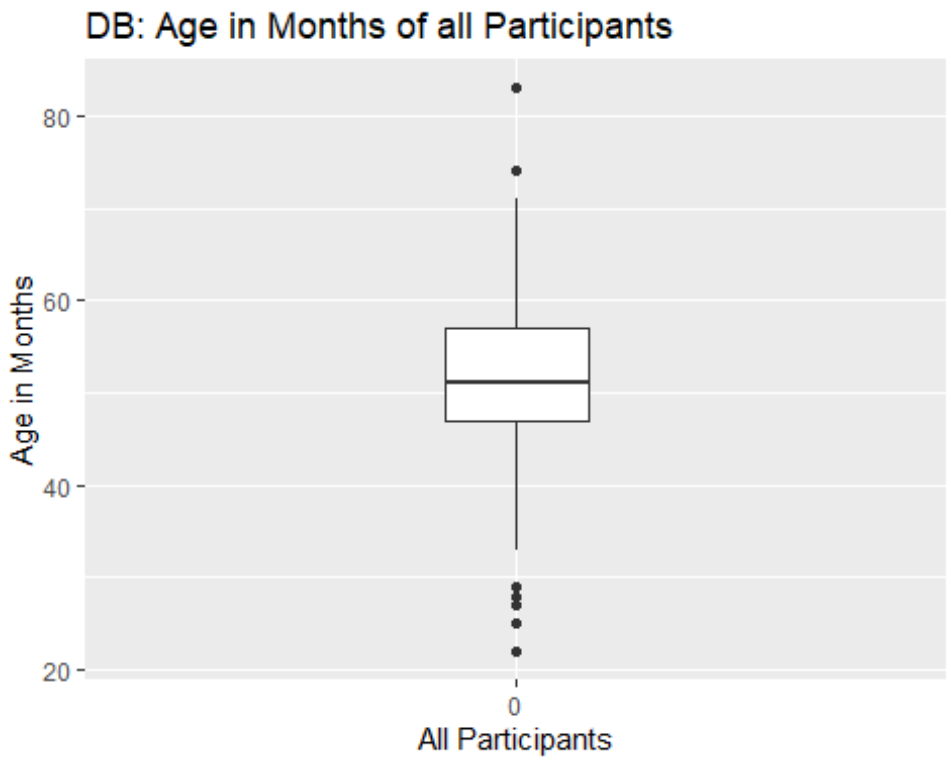
1. Barplot of the count of females and males in the data set. There are more than twice the amount of females in the data set as there are males.

2. Barplot comparing amount of people with and without autism in this study. There are more people included in this study that have autism than those that do not have autism.

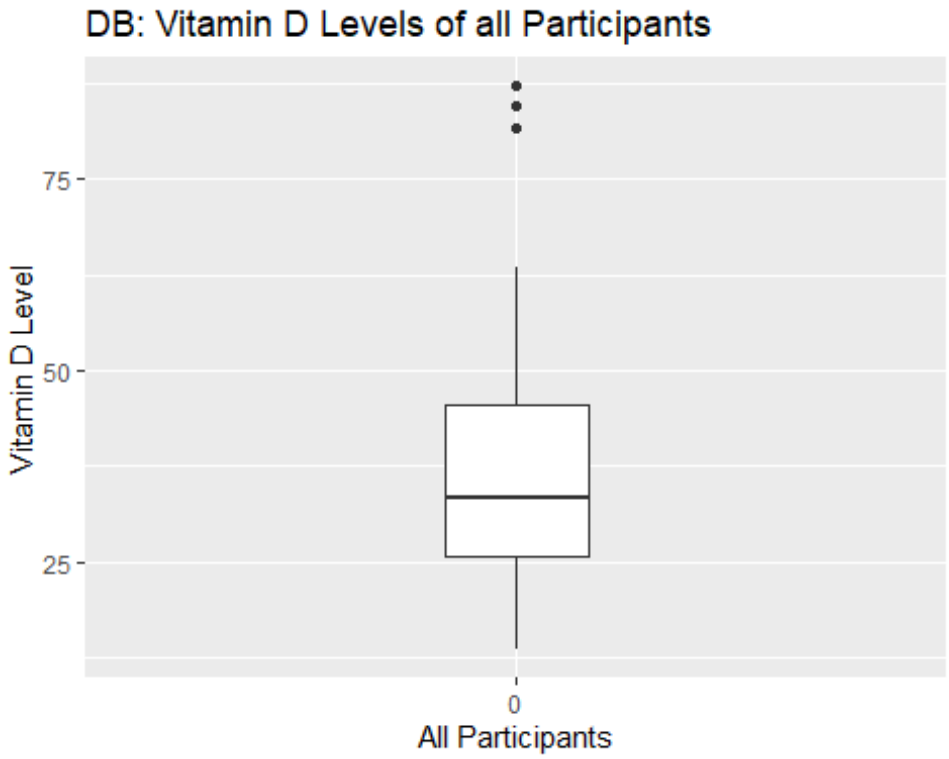
3. These barplots can be combined to display both variables at the same time. Visually, there looks to be a similar proportion of males and females with and without autism. It is important to note that this data set may not be a strong predictor for a general population. The data is made up predominantly of females, and the prevalence of autism among the participants is higher than that of the general population.

Quantitative Data

```
ggplot(autism,aes(x=as.factor(0),y=age_month))+
  geom_boxplot(width=0.2)+
  ylab('Age in Months')+
  xlab('All Participants')+
  ggtitle('DB: Age in Months of all Participants')
```




```
ggplot(autism,aes(x=as.factor(0),y=vitD_level))+
  geom_boxplot(width=0.2)+
  ylab('Vitamin D Level')+
  xlab('All Participants')+
  ggtitle('DB: Vitamin D Levels of all Participants')
```



1. Boxplot of participant ages in months. There was no second variable listed for us to compare to, so I just plotted the age variable against 0 as a placeholder. This data is slightly skewed towards younger participants.

2. Boxplot of participant vitamin D levels. Similar to the previous plot, there was no variable we were told to compare to, so I just plotted it against x=0 as a placeholder. This data is slightly skewed towards those with lower vitamin D levels, but we can see many outliers with very high vitamin D levels.

Hypothesis Testing of Continuous Variables

```
t.test(autism$vitD_level)


    One Sample t-test

data:  autism$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

t.test(autism$age_month)
```

```
    One Sample t-test

data:  autism$age_month
t = 43.315, df = 85, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 48.72556 53.41398
sample estimates:
mean of x
 51.06977
```

1. One-sample t-test for the vitD_level variable. The null hypothesis is that the vitamin D level of the participants is equal to zero, and the alternative hypothesis is that it is not equal to zero. Our sample mean was found to be equal to 35.96, and our t-test yielded a significant p-value (<0.05). We can report with 95% confidence that the true mean vitamin D level lies between 32.63 and 39.30. Thus, we reject the null hypothesis that the average vitamin D level of the participants was zero.

2. One-sample t-test for the age_month variable. The null hypothesis is that the age in months of the participants is equal to zero, and the alternative hypothesis is that it is not equal to zero. Our sample mean was found to be equal to 51.07, and our t-test yielded a significant p-value (<0.05). We can report with 95% confidence that the true mean age in months lies between 48.73 and 53.41. Thus, we reject the null hypothesis that the average age in months of our participants was zero. It is important to note that age was grouped into full months in the collection of this data. However, age is still a continuous variable even though it may appear to be discrete.