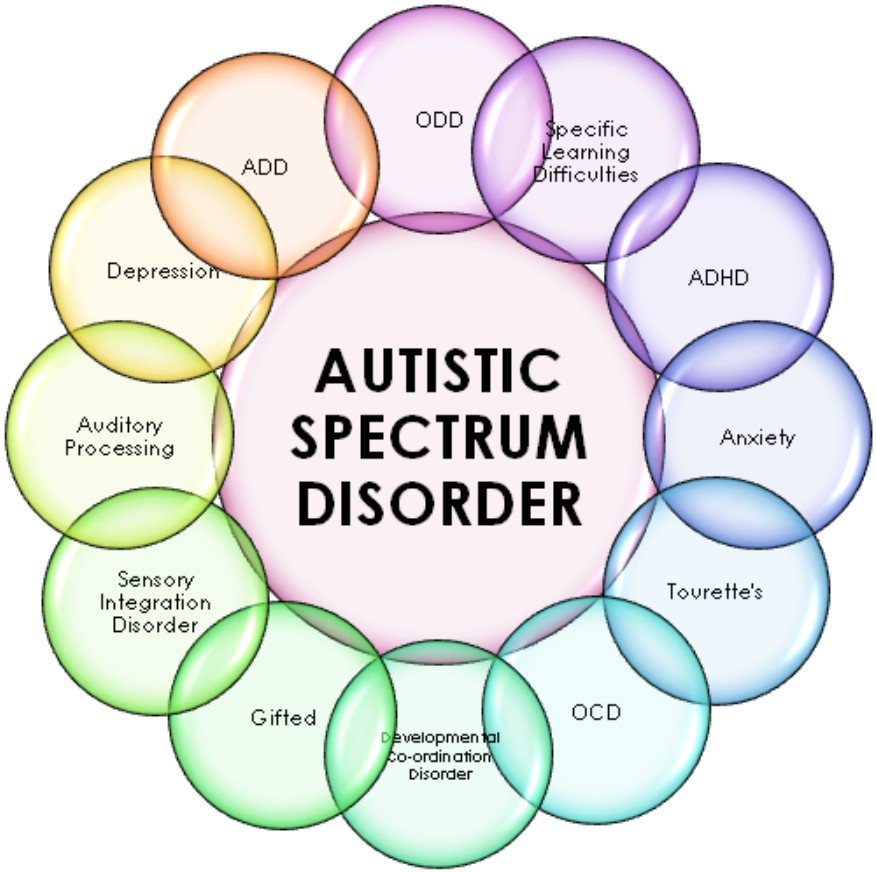
Final Project: Introduction to R

by Juliette Gudknecht

2022-06-29

Introduction

Autism affects 1 in 44 children in the United States (CDC 2020). The DSM categorizes autism as having repetitive restricted behaviors, and deficits in social communication (CDC 2021). Autistic people may behave, interact, communicate, and learn differently than most other people (CDC 2021). Frequently, their appearance does not distinguish them from others (CDC 2021). People with ASD might have a wide range of abilities and it varies widely across the spectrum (CDC 2021). My project seeks to look into a few research questions surrounding the condition. The condition is made up by many symptoms and co-occurring conditions, which can be seen in the photo below:



Citation for the photo is in the sources.

First, we will import libraries.

Libraries

```
library(tidyverse)
library(MASS)
library(psych)
library(EnvStats)
library(rstatix)
library(car)
library(prettydoc)
library(ggthemes)
library(formattable)
library(broom)
library(purrr)
library(corrplot)
library(patchwork)
library(ggplot2)
```

Data Sources

The data was from two sources.

1. The first source is a survey that I administered to autistic college students (past and prior) for a research course.

Variables:

Diagnosis_check: a check for autism diagnosis (later deleted)

ADHD_Screener: a check for co-occuring ADHD diagnosis

LD_Screener: a check for co-occuring Learning Disability diagnosis

Reading: a confidence scale of reading confidence by participants

Writing: a confidence scale of writing confidence by participants

Math: a confidence scale of math confidence by participants

Teacher_support: a qualitative response variable based on the question "Did teachers support you in college?" (later deleted)

Teacher_support_yn: a y/n type response to the Teacher Support question (No, Somewhat, Yes)

Struggle: a qualitative response variable based on the question "Did you struggle in college, and why?" (later deleted)

Struggle_yn: a y/n type response to the Struggle question (No, Somewhat, Yes)

Accommodations: a qualitative response variable based on the question "What could teachers have done for you in college?" (later deleted)

Extra_question: a qualitative response variable based on the question "Is there anything else you'd like us to know about you?" (later deleted)

2. The second source is from a paper on Plos One titled "Exploring the spatial working memory and visual perception in children with autism spectrum disorder and general population with high autism-like traits" by Manxue Zhang et al (also listed in sources).

Variables:

Group: group that the participants belong to (ASD, control, high asd traits, low asd traits)

Gender: gender (male = 1, female = 2)

Age: age of participant

AQ: autism quotient score

VMI-integration: visual motor integration score for integration

VMI-Visual perception: visual motor integration score for visual perception

VNM-Motor: visual motor integration motor scores

IQ: intelligence quotient score

```
dat <- read_csv("/Users/juliettegudknecht/Downloads/asdsurveyresponses.csv")
dat2 <-
        read_csv("/Users/juliettegudknecht/Downloads/columbia/introtor/pone.0235552.s001.csv",
         col_names = TRUE)
```

Data Cleaning for Data Set One

Deleting columns that aren't of interest to us.

```
dat <- dat %>% select(-c(Diagnosis_check, Teacher_support, Struggle, Accommodations,
        Extra_question))
```

Changing variables into factors or numerics with the appropriate order.

```
dat$Reading <- as.factor(dat$Reading)
dat$Math <- as.factor(dat$Math)
dat$Writing <- as.factor(dat$Writing)
levels(dat$Reading) <- list("Not Confident At All" = "Not Confident At All",
                  "Slightly Confident" = "Slightly Confident",
                  "Somewhat Confident" = "Somewhat Confident",
                  "Fairly Confident" = "Fairly Confident",
                  "Very Confident" = "Very Confident")
levels(dat$Math) <- list("Not Confident At All" = "Not Confident At All",
                  "Slightly Confident" = "Slightly Confident",
                  "Somewhat Confident" = "Somewhat Confident",
                  "Fairly Confident" = "Fairly Confident",
                  "Very Confident" = "Very Confident")
levels(dat$Writing) <- list("Not Confident At All" = "Not Confident At All",
                  "Slightly Confident" = "Slightly Confident",
                  "Somewhat Confident" = "Somewhat Confident",
                  "Fairly Confident" = "Fairly Confident",
                  "Very Confident" = "Very Confident")
dat$ADHD_screener <- as.factor(dat$ADHD_screener)
levels(dat$ADHD_screener) <- c("Autism", "Autism and ADHD")
dat$Teacher_support_yn <- as.factor(dat$Teacher_support_yn)
dat$Struggle_yn <- as.factor(dat$Struggle_yn)
dat$Age <- as.numeric(dat$Age)
```

Lets look into some descriptive statistics of the variables!

```
formattable(round(describe(dat, ranges=FALSE, skew=FALSE), 2))
```

	vars	n	mean	sd	se
User ID	1	30	15.50	8.80	1.61
ADHD_screener*	2	30	1.43	0.50	0.09
LD_screener	3	30	0.07	0.25	0.05
Age	4	30	35.03	11.76	2.15
Reading*	5	30	4.23	1.01	0.18
Writing*	6	30	4.13	1.04	0.19
Math*	7	30	3.27	1.14	0.21
Teacher_support_yn*	8	30	2.00	0.69	0.13
Struggle_yn*	9	30	2.80	0.55	0.10

This shows us that more people do not have ADHD in the sample, but not by much. This shows us that the math confidence score is the lowest of the 3, and that there is pretty high confidence overall in Reading and Writing. Teacher_support_yn is on a scale of 1-3 so it shows on average they felt somewhat supported. The struggle_yn is also on a 1-3 scale and shows they felt on average that they struggled. The sample size is 30 which is not the best sample size, but the survey was administered by me for a research course, so I think it is pretty reasonable.

Data Cleaning for Data Set Two

Deleting variables that aren't of interest to us.

```
dat2 <- dat2 %>% select(-c(`SWM Total errors`))
```

Turning the gender and group variable into a factor, and creating a factor group called ASD (autism spectrum disorder group) and TDC (typically developing children group). Let's take a look at the descriptive of the variables now in the dataset.

```
dat2$gender <- as.factor(dat2$gender)
levels(dat2$gender) <- c("Male", "Female")
dat2$group <- as.factor(dat2$group)
dat2$group2 <- as.factor(dat2$group)
levels(dat2$group2) <- list("ASD" = c("ASD"),
                 "TDC" = c("lower trait", "higher trait", "control"))
```

```
formattable(round(describe(dat2, ranges=FALSE, skew=FALSE)), 2)
```

	vars	n	mean	sd	se
group*	1	304	3	1	0
code*	2	304	124	83	5
gender*	3	304	1	0	0
age	4	304	9	2	0
AQ	5	304	16	5	0
VMI-integration	6	304	102	15	1
VMI-Visual Perception	7	304	107	14	1
VNM-motor	8	304	101	15	1
IQ	9	304	112	16	1
group2*	10	304	2	0	0

These descriptives show that the average age is 8.6 years old, average AQ is 16.2, average VMI-Integration is 101.9, average VMI-Visual Perception is 107.9, VNM-motor is 100.9, and average IQ is 112.3 (which is pretty high). This also shows that there are more typically developing children in the study than ASD students (group2). There is a pretty large sample size of 304 participants.

Research Questions for Data Set One

1. How do autistic people feel about their writing/reading/math abilities, and does co-occurring ADHD make a difference?

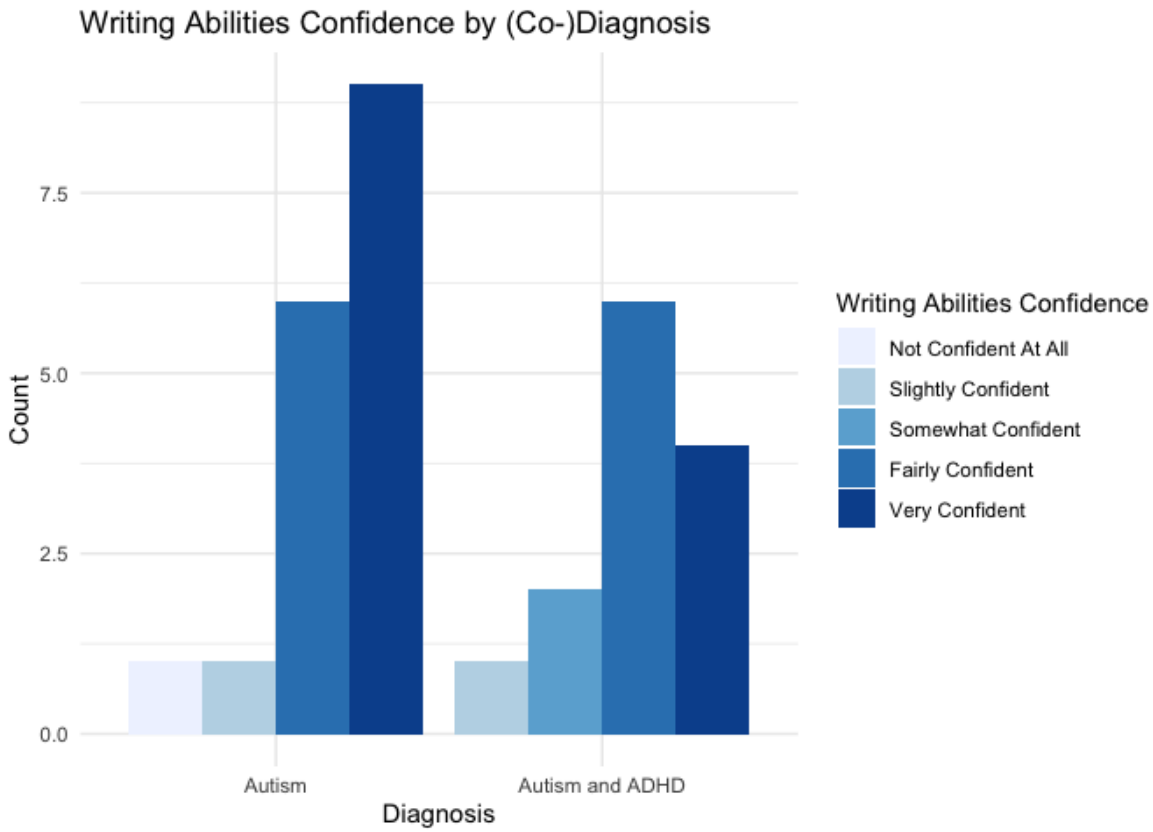
(For this question, I had to turn the scales into numerical scales for the t-test and calculating the mean.)

Writing:

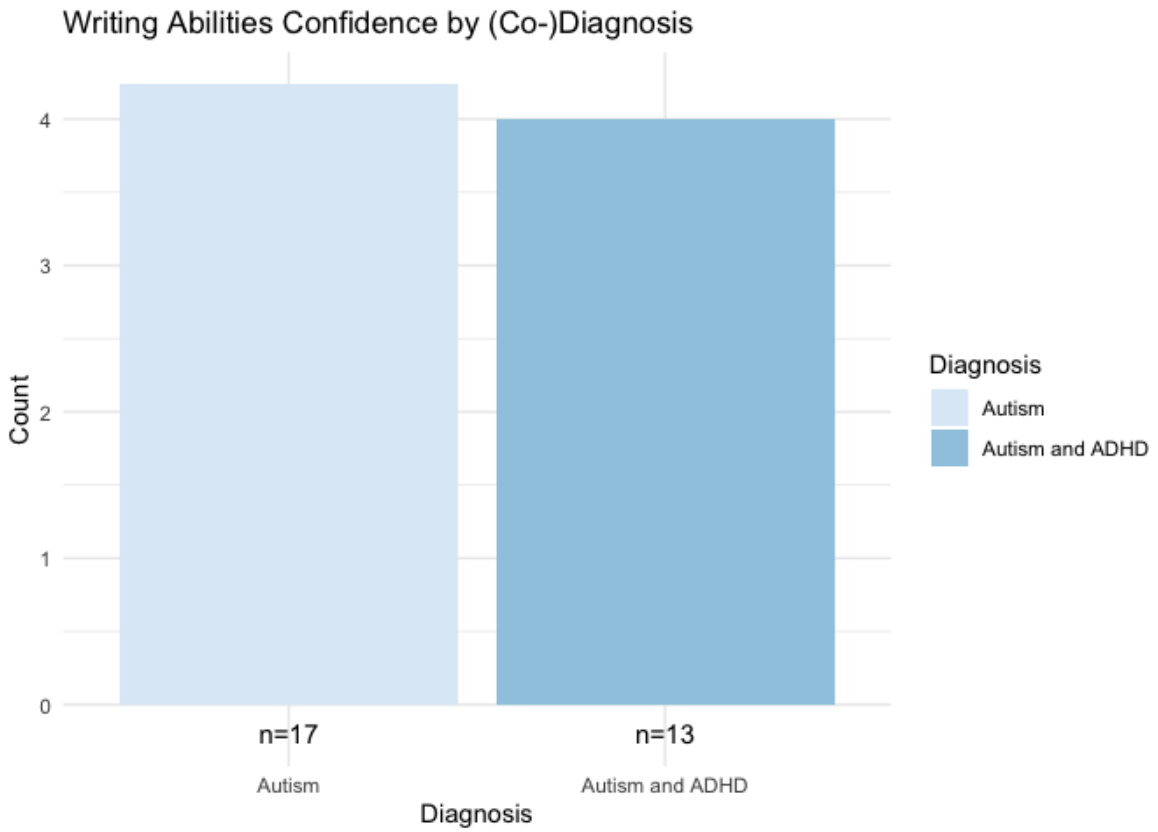
```
gg1 <- ggplot(data = dat, mapping = aes(x = ADHD_screener, fill = Writing)) +
geom_bar(position = "dodge") +
xlab("Diagnosis") +
ylab("Count") +
ggtitle("Writing Abilities Confidence by (Co-)Diagnosis") +
guides(fill = guide_legend(title = "Writing Abilities Confidence")) +
```

```
theme_minimal() +
scale_fill_brewer(palette="Blues")

gg1
```




```
dat$Writing_numeric <- as.numeric(dat$Writing)
ggplot(data = dat, mapping = aes(x = ADHD_screener, y = Writing_numeric, fill =
        ADHD_screener)) +
stat_summary(fun = mean, geom = "bar") +
stat_n_text(y.expand.factor = 0.3) +
xlab("Diagnosis") +
ylab("Count") +
ggtitle("Writing Abilities Confidence by (Co-)Diagnosis") +
guides(fill = guide_legend(title = "Diagnosis")) +
theme_minimal() +
scale_fill_brewer(palette="Blues")
```




```
dat %>% group_by(ADHD_screener) %>% summarize(mean(Writing_numeric), var(Writing_numeric))
```

```
## # A tibble: 2 × 3
##   ADHD_screener   `mean(Writing_numeric)` `var(Writing_numeric)`
##   <fct>                             <dbl>                  <dbl>
## 1 Autism                             4.24                  1.32
## 2 Autism and ADHD                    4                     0.833
```

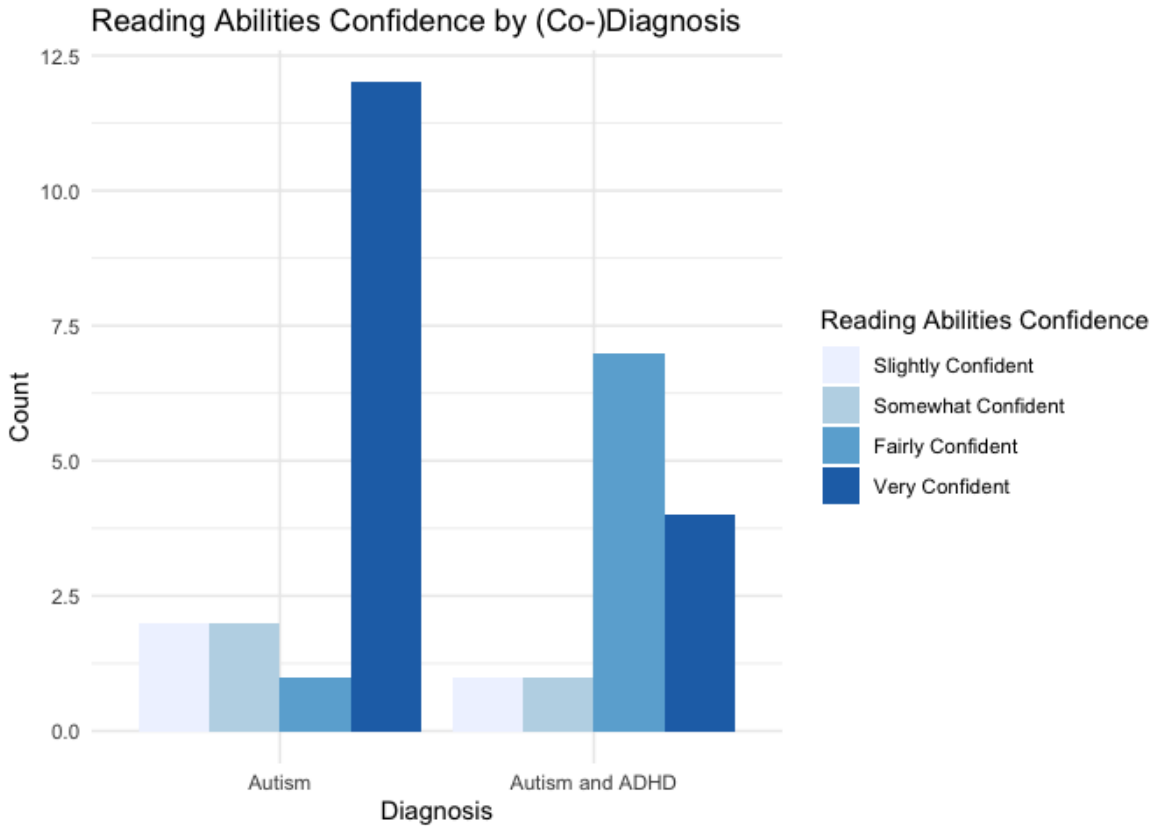
```
(t1 <- t.test(Writing_numeric ~ ADHD_screener, dat, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  Writing_numeric by ADHD_screener
## t = 0.60636, df = 28, p-value = 0.5492
## alternative hypothesis: true difference in means between group Autism and group Autism and
## 95 percent confidence interval:
##  -0.5595725  1.0301607
## sample estimates:
##          mean in group Autism mean in group Autism and ADHD
##                      4.235294                      4.000000
```

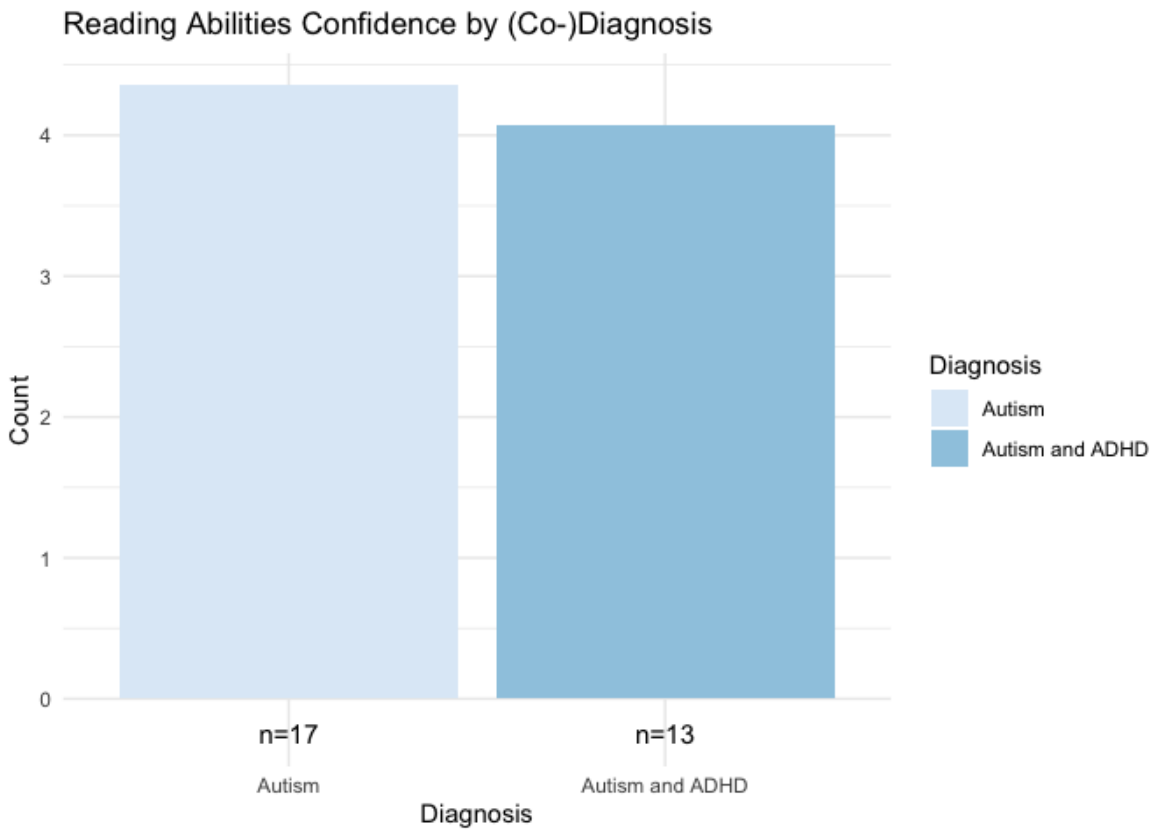
The p value for the t test shows that it is not statistically significant, meaning there is not a difference by diagnosis in writing confidence. Overall, the participants were pretty confident in their writing abilities with an average of fairly confident.

Reading:

```
gg2 <- ggplot(data = dat, mapping = aes(x = ADHD_screener, fill = Reading)) +
geom_bar(position = "dodge") +
xlab("Diagnosis") +
ylab("Count") +
ggtitle("Reading Abilities Confidence by (Co-)Diagnosis") +
guides(fill = guide_legend(title = "Reading Abilities Confidence"))+
theme_minimal() +
scale_fill_brewer(palette="Blues")
gg2
```




```
dat$Reading_numeric <- as.numeric(dat$Reading)
ggplot(data = dat, mapping = aes(x = ADHD_screener, y = Reading_numeric, fill =
        ADHD_screener)) +
stat_summary(fun = mean, geom = "bar") +
stat_n_text(y.expand.factor = 0.75) +
xlab("Diagnosis") +
ylab("Count") +
ggtitle("Reading Abilities Confidence by (Co-)Diagnosis") +
guides(fill = guide_legend(title = "Diagnosis")) +
theme_minimal() +
scale_fill_brewer(palette="Blues")
```




```
dat %>% group_by(ADHD_screener) %>% summarize(mean(Reading_numeric), var(Reading_numeric))
```

```
## # A tibble: 2 × 3
##   ADHD_screener   `mean(Reading_numeric)` `var(Reading_numeric)`
##   <fct>                             <dbl>                  <dbl>
## 1 Autism                             4.35                  1.24
## 2 Autism and ADHD                    4.08                  0.744
```

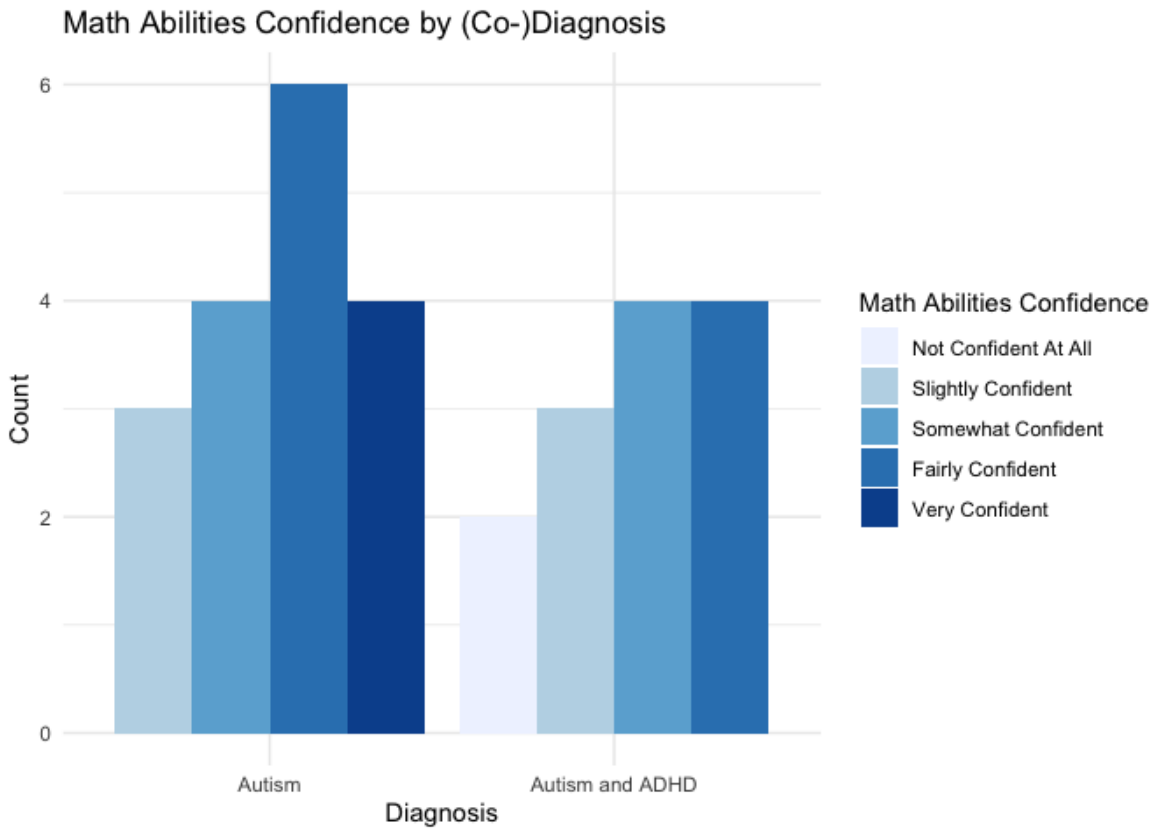
```
(t2 <- t.test(Reading_numeric ~ ADHD_screener, dat, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  Reading_numeric by ADHD_screener
## t = 0.73861, df = 28, p-value = 0.4663
## alternative hypothesis: true difference in means between group Autism and group Autism and
## 95 percent confidence interval:
##  -0.4894709  1.0415071
## sample estimates:
##          mean in group Autism mean in group Autism and ADHD
##                      4.352941                      4.076923
```

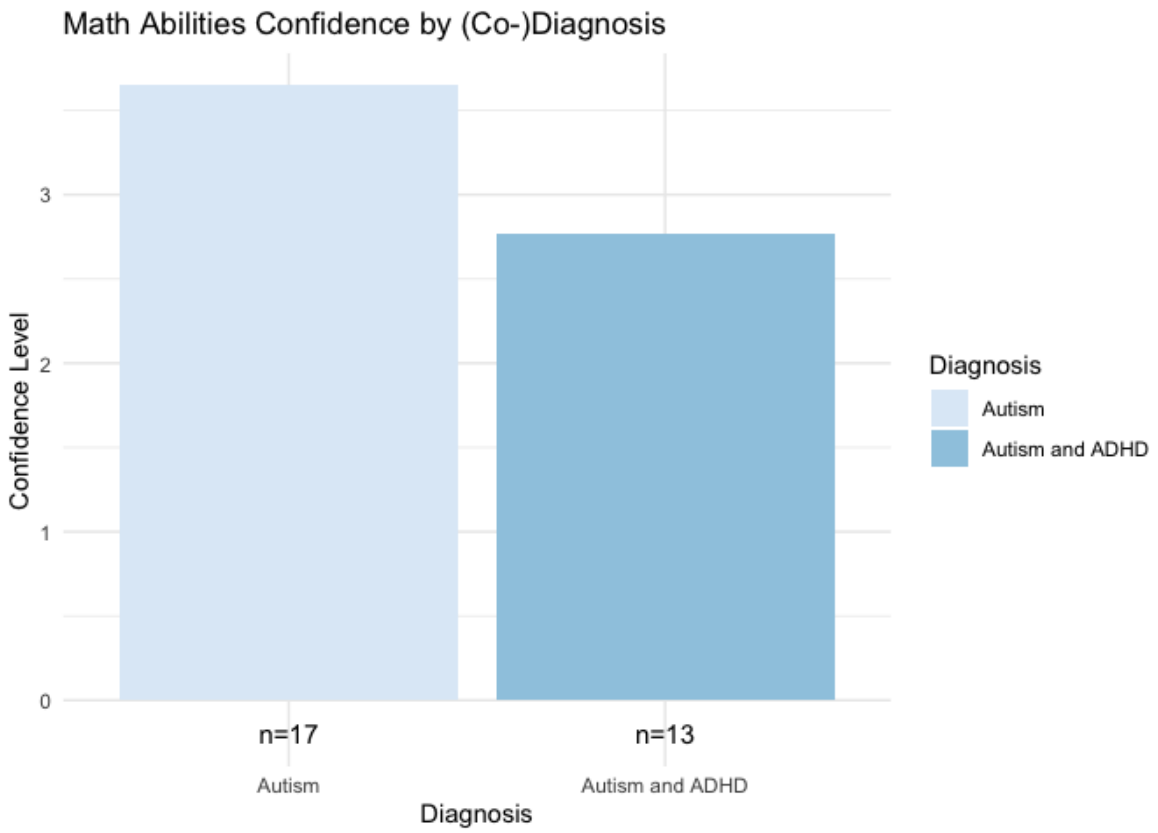
The p value for the t test shows that it is not statistically significant, meaning there is not a difference by diagnosis in reading confidence. Overall, the participants were pretty confident in their reading abilities with an average of fairly confident.

Math:

```
gg3 <- ggplot(data = dat, mapping = aes(x = ADHD_screener, fill = Math)) +
geom_bar(position = "dodge") +
xlab("Diagnosis") +
ylab("Count") +
ggtitle("Math Abilities Confidence by (Co-)Diagnosis") +
guides(fill = guide_legend(title = "Math Abilities Confidence")) +
theme_minimal() +
scale_fill_brewer(palette="Blues")
gg3
```




```
dat$Math_numeric <- as.numeric(dat$Math)
ggplot(data = dat, mapping = aes(x = ADHD_screener, y = Math_numeric, fill =
        ADHD_screener)) +
stat_summary(fun = mean, geom = "bar") +
stat_n_text(y.expand.factor = 0.3) +
xlab("Diagnosis") +
ylab("Confidence Level") +
ggtitle("Math Abilities Confidence by (Co-)Diagnosis")+
guides(fill = guide_legend(title = "Diagnosis")) +
theme_minimal() +
scale_fill_brewer(palette="Blues")
```




```
dat %>% group_by(ADHD_screener) %>% summarize(mean(Math_numeric), var(Math_numeric))
```

```
## # A tibble: 2 × 3
##   ADHD_screener   `mean(Math_numeric)` `var(Math_numeric)`
##   <fct>                          <dbl>               <dbl>
## 1 Autism                          3.65                1.12
## 2 Autism and ADHD                 2.77                1.19
```

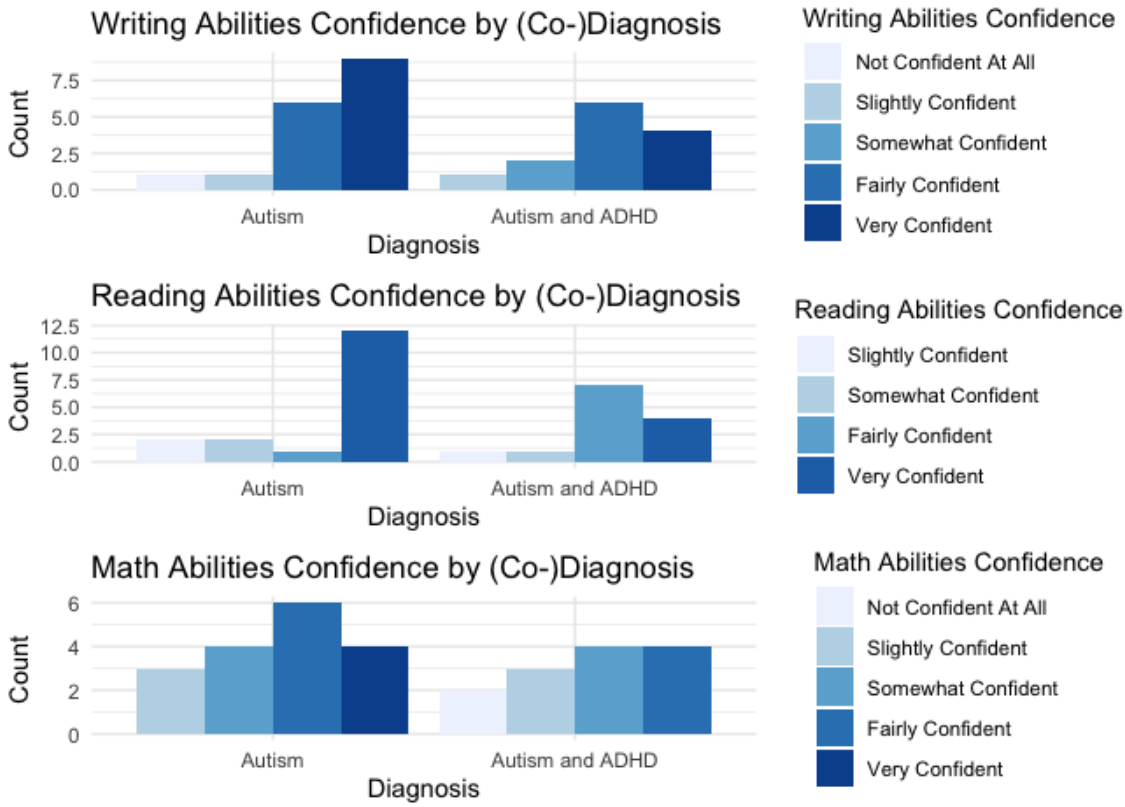
```
(t3 <- t.test(Math_numeric ~ ADHD_screener, dat, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  Math_numeric by ADHD_screener
## t = 2.2221, df = 28, p-value = 0.03453
## alternative hypothesis: true difference in means between group Autism and group Autism and
## 95 percent confidence interval:
##  0.06861562 1.68704049
## sample estimates:
##          mean in group Autism mean in group Autism and ADHD
##                      3.647059                      2.769231
```

We can see that the p value shows it is statistically significant that math is the only group that has a significant difference in confidence across diagnosis.

Lets see if math, writing, and math scores differ from each other and how much. Let's put all the graphs together! Let's also look and compare the t-test results again, next to each other. The order is writing, reading and then math. Math is the only significant one.

```
gg1 / gg2 / gg3
```




```
formattable(dat %>% group_by(ADHD_screener) %>% summarize(mean(Writing_numeric),
        mean(Reading_numeric), mean(Math_numeric)))
```

ADHD_screener	mean(Writing_numeric)	mean(Reading_numeric)	mean(Math_numeric)
Autism	4.235294	4.352941	3.647059
Autism and ADHD	4.000000	4.076923	2.769231

```
tab <- map_df(list(t1, t2, t3), tidy)
formattable(round(tab[c("estimate", "statistic", "p.value", "conf.low", "conf.high")], 2))
```

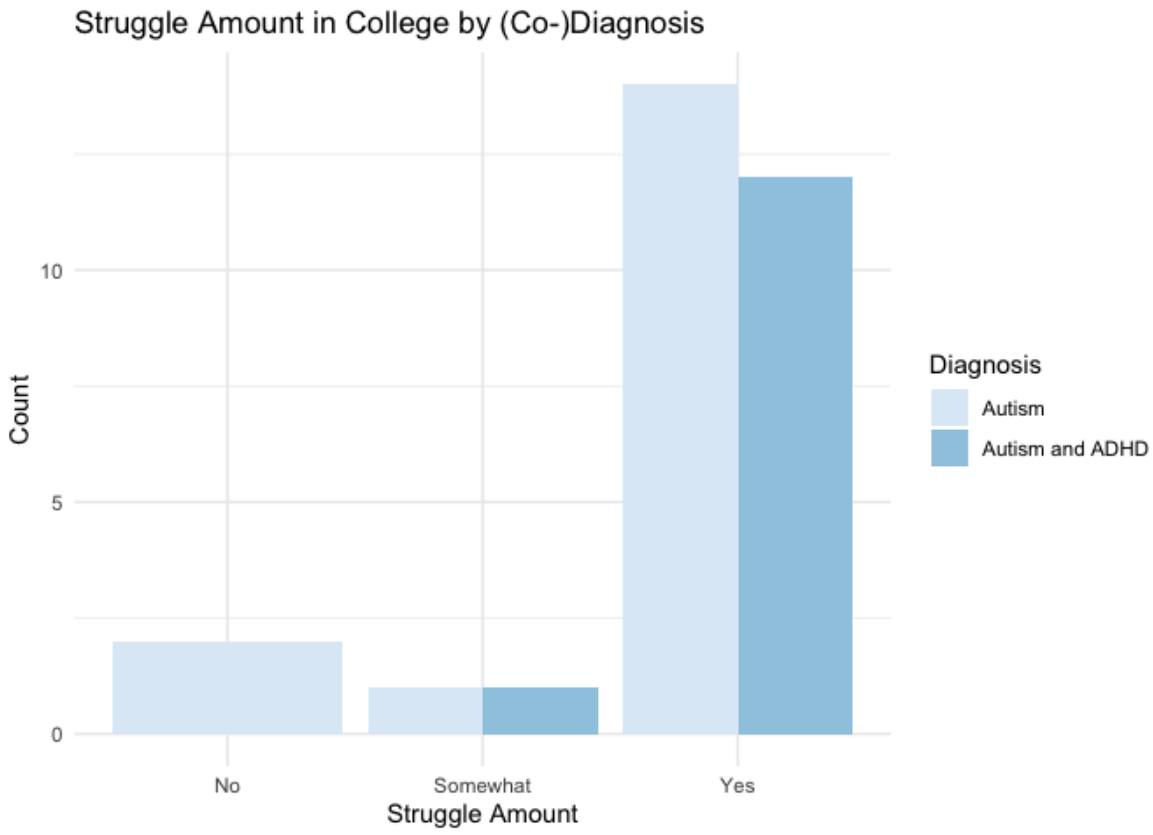
estimate	statistic	p.value	conf.low	conf.high

estimate	statistic	p.value	conf.low	conf.high
0.24	0.61	0.55	-0.56	1.03
0.28	0.74	0.47	-0.49	1.04
0.88	2.22	0.03	0.07	1.69

2.Do autistic people struggle in college, and does co-occurring ADHD have an effect on struggling in college?

Let's test this question in a plot. Struggle_yn will be turned into a numerical scale of 1-3 for statistical testing purposes.

```
ggplot(data = dat, mapping = aes(x = Struggle_yn, fill = ADHD_screener)) +
geom_bar(position = "dodge") +
ylab("Count") +
xlab("Struggle Amount") +
ggtitle("Struggle Amount in College by (Co-)Diagnosis")+
guides(fill = guide_legend(title = "Diagnosis")) +
theme_minimal() +
scale_fill_brewer(palette="Blues")
```



From looking at the graph, it would appear that the Autism and ADHD group struggle more due to no one saying they did not struggle. Let's look at their means and t-test results. First we will have to make a numeric

version of struggle so that we can numerically evaluate it.

```
dat$Struggle_yn_numeric <- as.numeric(dat$Struggle_yn)

formattable(dat %>% group_by(ADHD_screener)
            %>% summarize(mean(Struggle_yn_numeric),
                          var(Struggle_yn_numeric)))
```

ADHD_screener	mean(Struggle_yn_numeric)	var(Struggle_yn_numeric)
Autism	2.705882	0.47058824
Autism and ADHD	2.923077	0.07692308

```
(t4 <- t.test(Struggle_yn_numeric ~ ADHD_screener, dat, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  Struggle_yn_numeric by ADHD_screener
## t = -1.0729, df = 28, p-value = 0.2925
## alternative hypothesis: true difference in means between group Autism and group Autism and
## 95 percent confidence interval:
##  -0.6318564  0.1974673
## sample estimates:
##          mean in group Autism mean in group Autism and ADHD
##                      2.705882                      2.923077
```

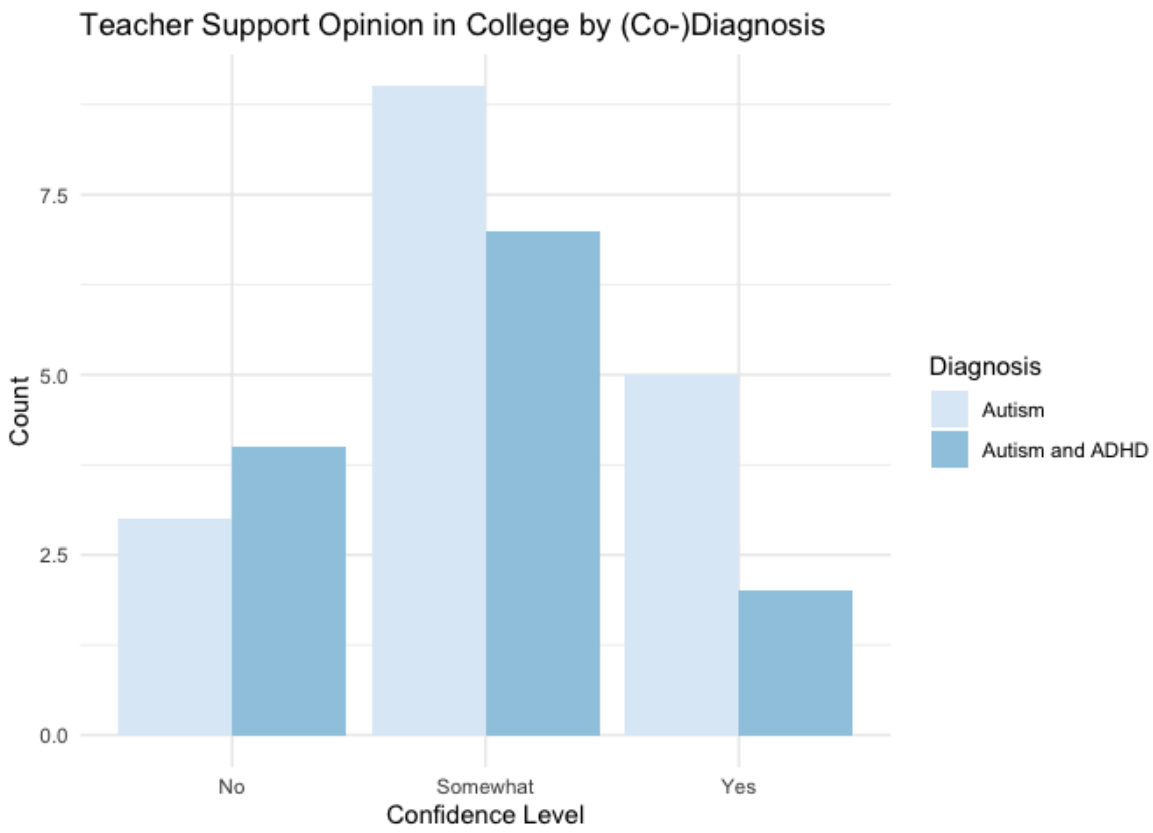
According to the means, it appears the Autism and ADHD group struggle more. But, according to the t-test results (the p-value is not significant), we can conclude that there is not a significant difference in their means and therefore a diagnosis of ADHD on top of autism does not cause students to struggle more in this cohort.

2.Do autistic people feel supported by their teachers, and does co-occurring ADHD have an effect on feeling supported?

Let's look into this with some plots. Teacher_support_yn will be turned into a numerical scale (1-3) for statistical testing purposes.

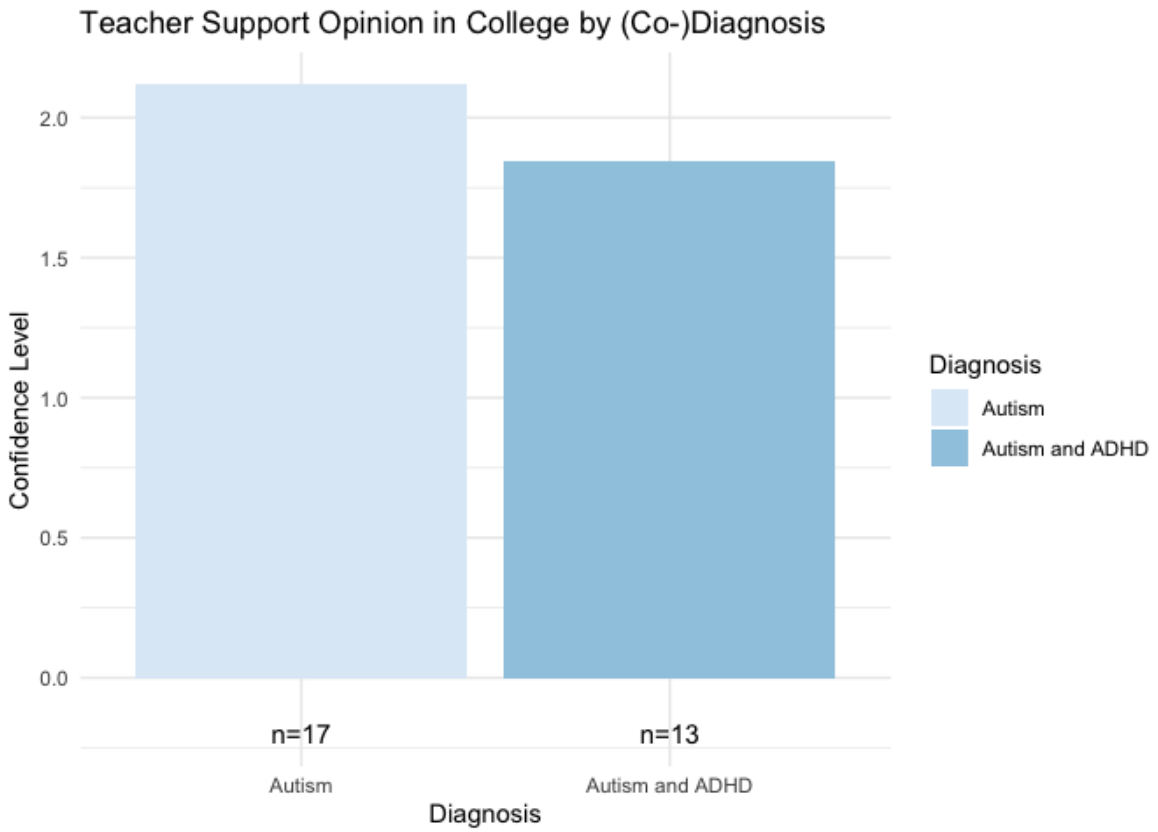
```
ggplot(data = dat, mapping = aes(x = Teacher_support_yn, fill = ADHD_screener)) +
geom_bar(position = "dodge") +
ylab("Count") +
xlab("Confidence Level") +
ggtitle("Teacher Support Opinion in College by (Co-)Diagnosis")+
guides(fill = guide_legend(title = "Diagnosis")) +
```

```
theme_minimal() +
scale_fill_brewer(palette="Blues")
```




```
dat$Teacher_yn_numeric <- as.numeric(dat$Teacher_support_yn)

ggplot(data = dat, mapping = aes(x = ADHD_screener, y = Teacher_yn_numeric, fill =
        ADHD_screener)) +
stat_summary(fun = mean, geom = "bar") +
stat_n_text(y.expand.factor = 0.6) +
xlab("Diagnosis") +
ylab("Confidence Level") +
ggtitle("Teacher Support Opinion in College by (Co-)Diagnosis") +
guides(fill = guide_legend(title = "Diagnosis")) +
theme_minimal() +
scale_fill_brewer(palette="Blues")
```




```
formattable(dat %>% group_by(ADHD_screener)
            %>% summarize(mean(Teacher_yn_numeric),
                          var(Teacher_yn_numeric)))
```

ADHD_screener	mean(Teacher_yn_numeric)	var(Teacher_yn_numeric)
Autism	2.117647	0.4852941
Autism and ADHD	1.846154	0.4743590

```
(t5 <- t.test(Teacher_yn_numeric ~ ADHD_screener, dat, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  Teacher_yn_numeric by ADHD_screener
## t = 1.0629, df = 28, p-value = 0.2969
## alternative hypothesis: true difference in means between group Autism and group Autism and
## 95 percent confidence interval:
##  -0.2517169  0.7947033
## sample estimates:
##          mean in group Autism mean in group Autism and ADHD
##                      2.117647                      1.846154
```

According to the t-test results (the p-value is not significant), we can conclude that there is not a significant difference in their means and therefore a diagnosis of ADHD on top of autism does not cause students to feel they have more support by their teachers in this cohort.

Research Questions for Data Set Two

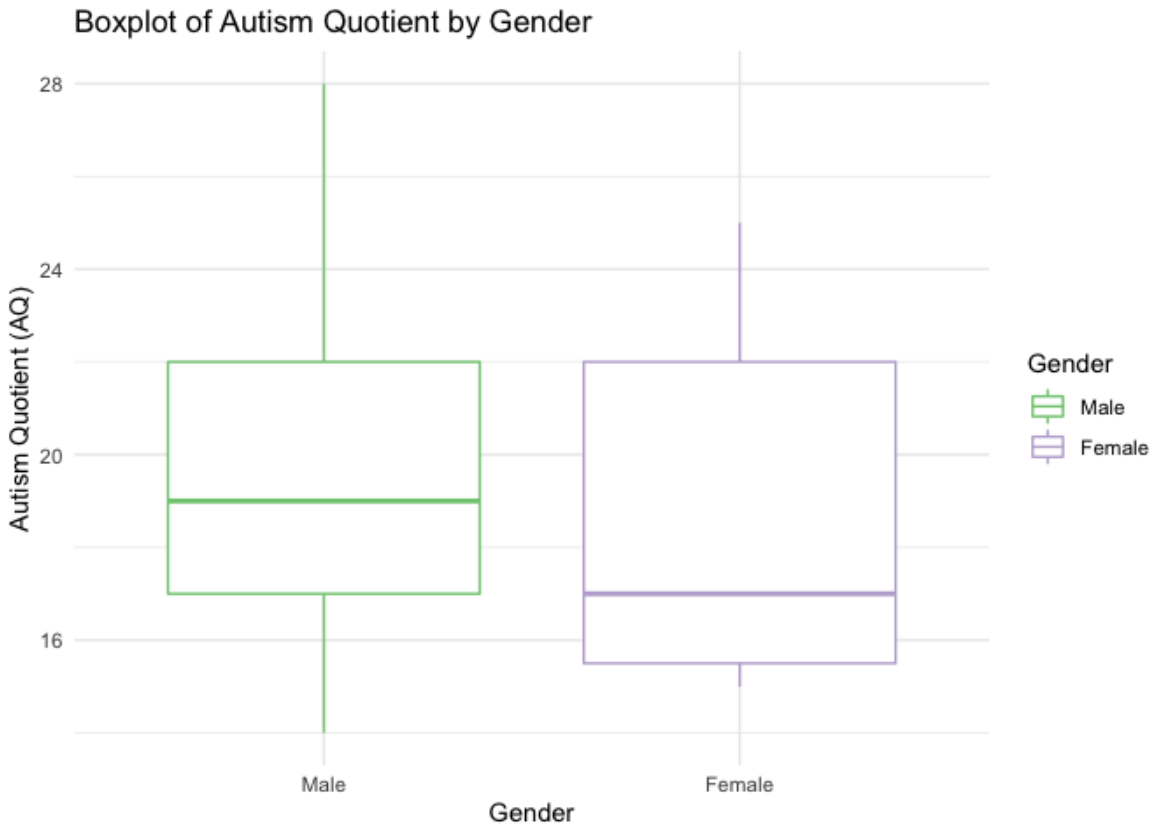
First, we will pull out only the autism group for our beginning research questions.

```
dat3 <- dat2 %>% filter(group == "ASD")
```

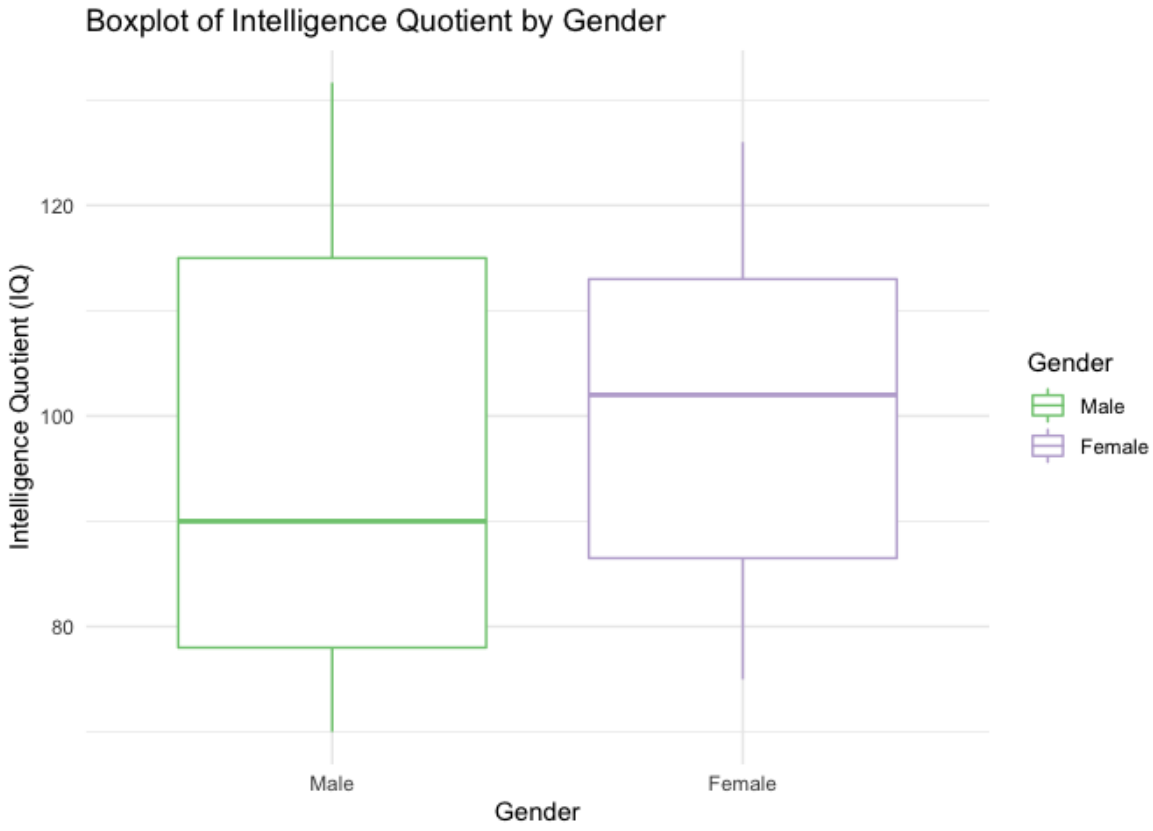
1. Does Gender have an effect on AQ, IQ, VMI-integration, VMI-Visual Perception or VNM-motor tests?

Let's test this with some plots!

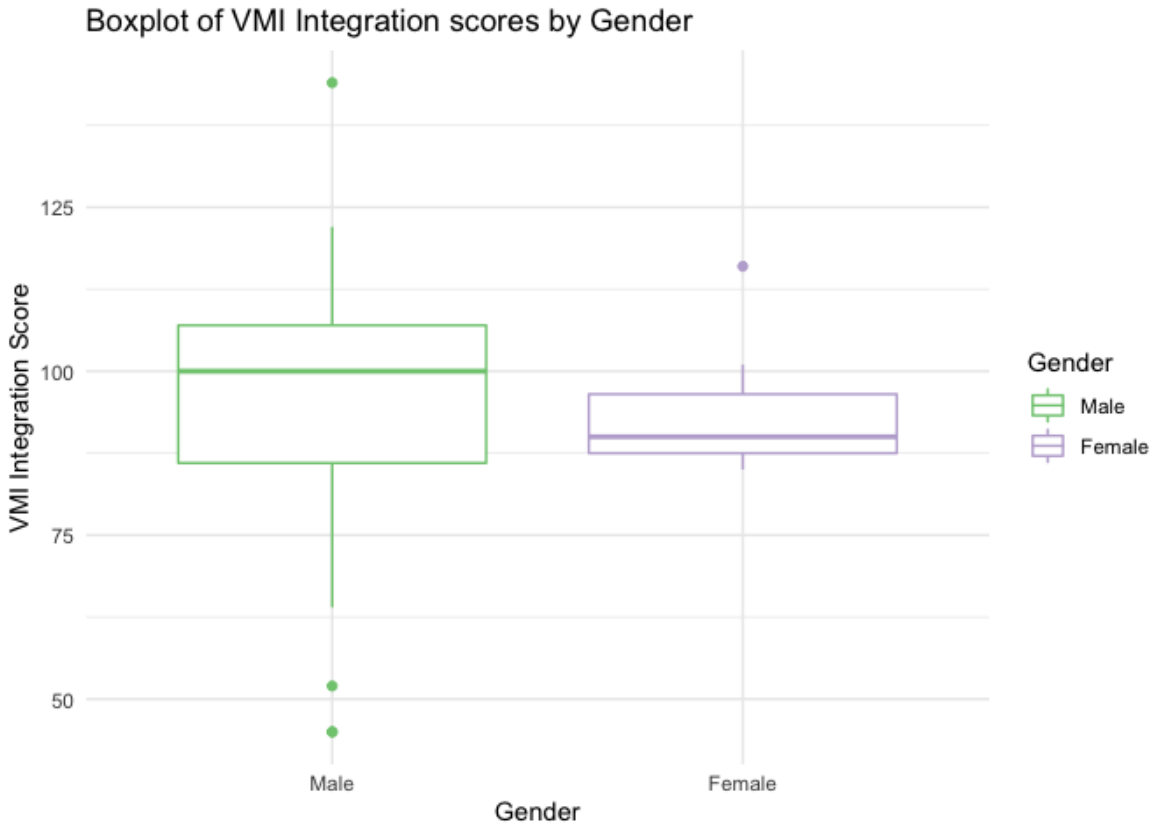
```
ggplot(dat3, aes(x = gender, y = AQ, color = gender)) +
  geom_boxplot() +
  labs(color = "Gender") +
  theme_minimal() +
  scale_color_brewer(palette="Accent") +
  xlab("Gender") +
  ylab("Autism Quotient (AQ)") +
  ggtitle("Boxplot of Autism Quotient by Gender")
```



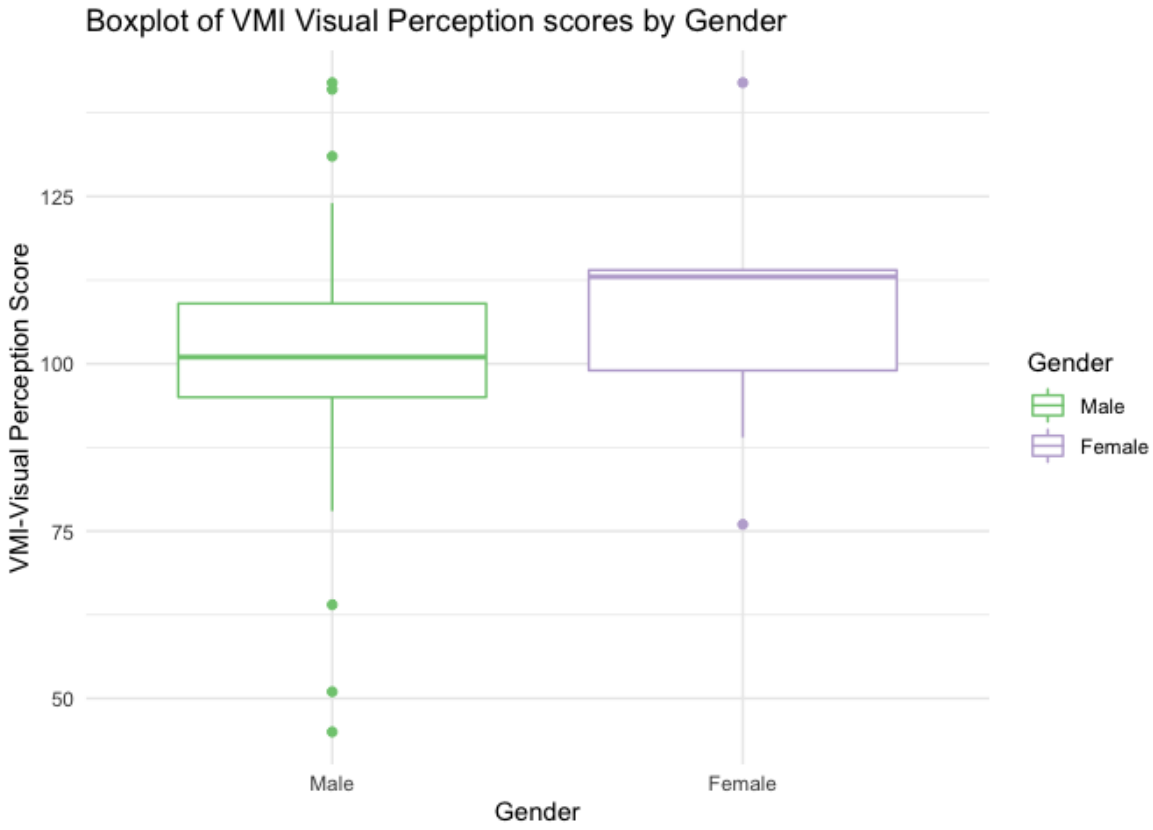
```
ggplot(dat3, aes(x = gender, y = IQ, color = gender)) +
  geom_boxplot() +
  labs(color = "Gender") +
  theme_minimal() +
  scale_color_brewer(palette="Accent") +
  xlab("Gender") +
  ylab("Intelligence Quotient (IQ)") +
  ggtitle("Boxplot of Intelligence Quotient by Gender")
```



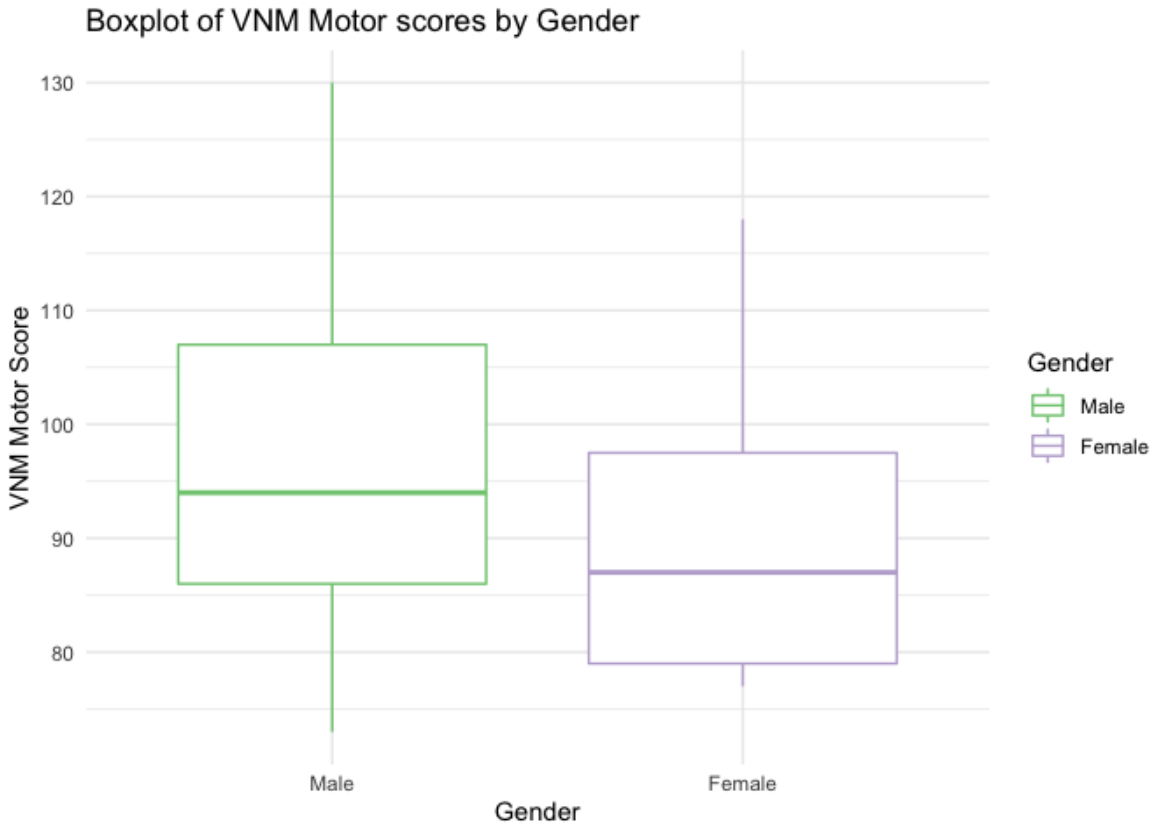

```
ggplot(dat3, aes(x = gender, y = `VMI-integration`, color = gender)) +
  geom_boxplot() +
  labs(color = "Gender") +
  theme_minimal() +
  scale_color_brewer(palette="Accent") +
  xlab("Gender") +
  ylab("VMI Integration Score") +
  ggtitle("Boxplot of VMI Integration scores by Gender")
```




```
ggplot(dat3, aes(x = gender, y = `VMI-Visual Perception`, color = gender)) +
  geom_boxplot() +
  labs(color = "Gender") +
  theme_minimal() +
  scale_color_brewer(palette="Accent") +
  xlab("Gender") +
  ylab("VMI-Visual Perception Score") +
  ggtitle("Boxplot of VMI Visual Perception scores by Gender")
```




```
ggplot(dat3, aes(x = gender, y = `VNM-motor`, color = gender)) +
  geom_boxplot() +
  labs(color = "Gender") +
  theme_minimal() +
  scale_color_brewer(palette="Accent") +
  xlab("Gender") +
  ylab("VNM Motor Score") +
  ggtitle("Boxplot of VNM Motor scores by Gender")
```



It is clear by looking at these graphs that gender looks like it makes a difference in the scores, but lets test that.

Let's statistically test to see if there is a difference by gender in the tests.

```
t.test(AQ ~ gender, dat3)
```

```
##
##  Welch Two Sample t-test
##
## data:  AQ by gender
## t = 0.55994, df = 7.5872, p-value = 0.5917
## alternative hypothesis: true difference in means between group Male and group Female is nc
## 95 percent confidence interval:
##  -2.90702  4.74829
## sample estimates:
##   mean in group Male mean in group Female
##             19.77778             18.85714
```

```
t.test(IQ ~ gender, dat3)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  IQ by gender
## t = -0.66914, df = 7.9986, p-value = 0.5222
## alternative hypothesis: true difference in means between group Male and group Female is no
## 95 percent confidence interval:
##  -23.25902  12.79693
## sample estimates:
##   mean in group Male mean in group Female
##             95.05467            100.28571
```

```
t.test(`VMI-integration` ~ gender, dat3)
```

```
##
##  Welch Two Sample t-test
##
## data:  VMI-integration by gender
## t = 0.56573, df = 13.637, p-value = 0.5808
## alternative hypothesis: true difference in means between group Male and group Female is no
## 95 percent confidence interval:
##  -8.126337 13.929512
## sample estimates:
##   mean in group Male mean in group Female
##             97.04444             94.14286
```

```
t.test(`VMI-Visual Perception` ~ gender, dat3)
```

```
##
##  Welch Two Sample t-test
##
## data:  VMI-Visual Perception by gender
## t = -0.84433, df = 7.5769, p-value = 0.4243
## alternative hypothesis: true difference in means between group Male and group Female is no
## 95 percent confidence interval:
##  -26.67542  12.47860
## sample estimates:
##   mean in group Male mean in group Female
##             101.0444             108.1429
```

```
t.test(`VNM-motor` ~ gender, dat3)
```

```
##
##  Welch Two Sample t-test
##
## data:  VNM-motor by gender
```

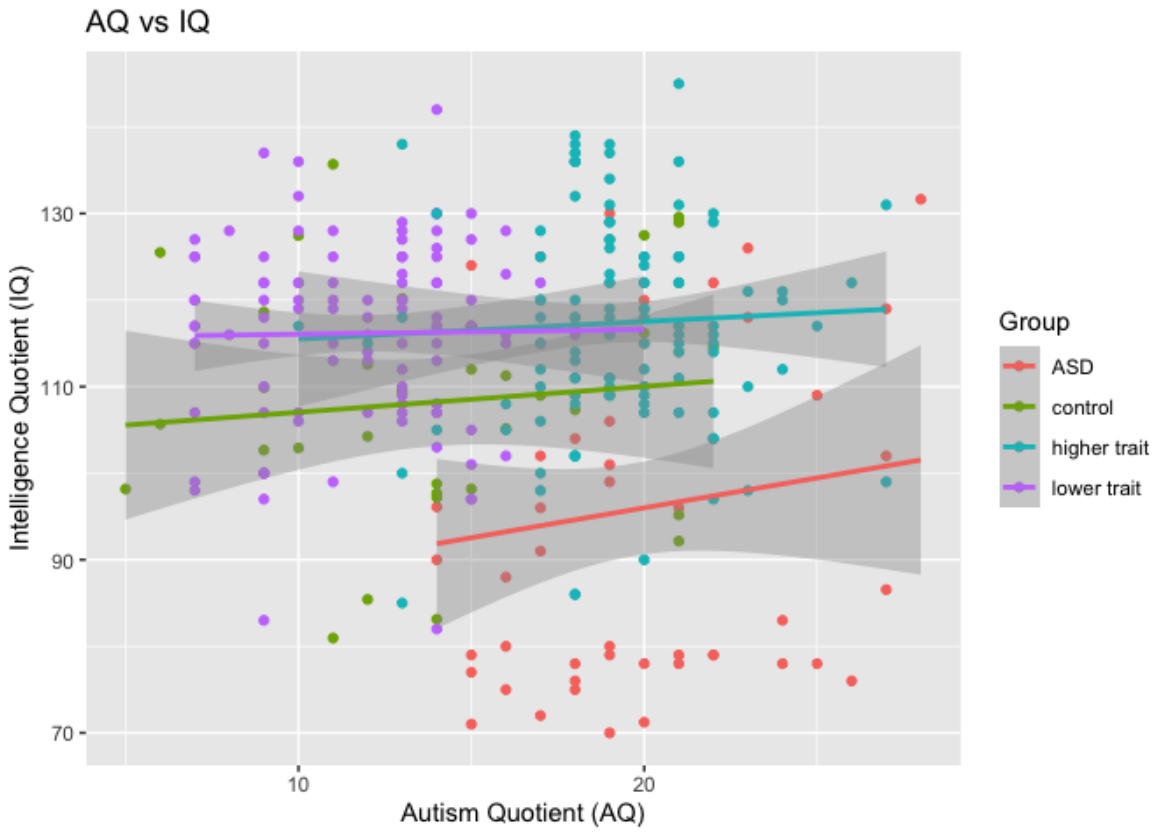
```
## t = 0.90207, df = 8.3284, p-value = 0.3924
## alternative hypothesis: true difference in means between group Male and group Female is nc
## 95 percent confidence interval:
##  -8.442059 19.413487
## sample estimates:
##   mean in group Male mean in group Female
##             96.20000             90.71429
```

All of the p values indicate that the effect of gender is not statistically significant, so therefore gender does not have an effect on the scores.

2. Does AQ have a relationship with any of the testing variables, and does it vary by group (ASD vs TDC)?

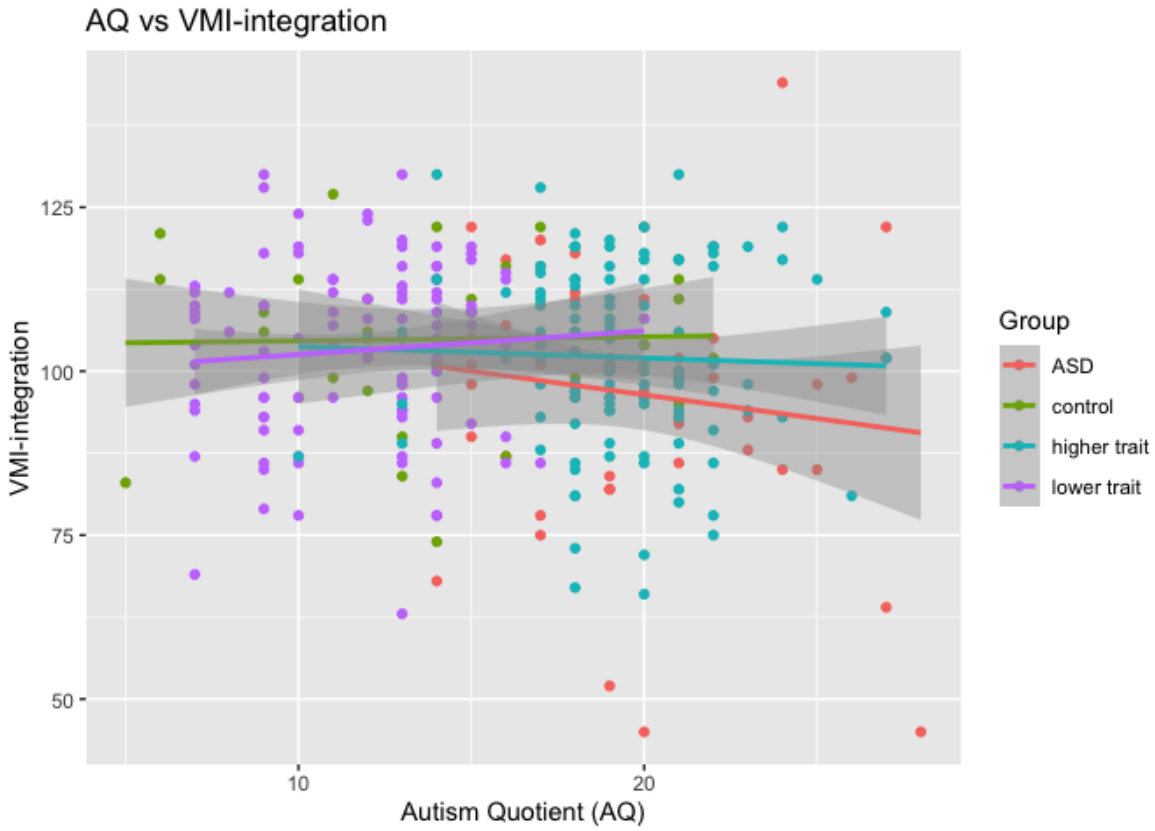
```
ggplot(dat2, aes(x = AQ, y = IQ, color = group)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(color = "Group") +
  ylab("Intelligence Quotient (IQ)") +
  xlab("Autism Quotient (AQ)") +
  ggtitle("AQ vs IQ")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



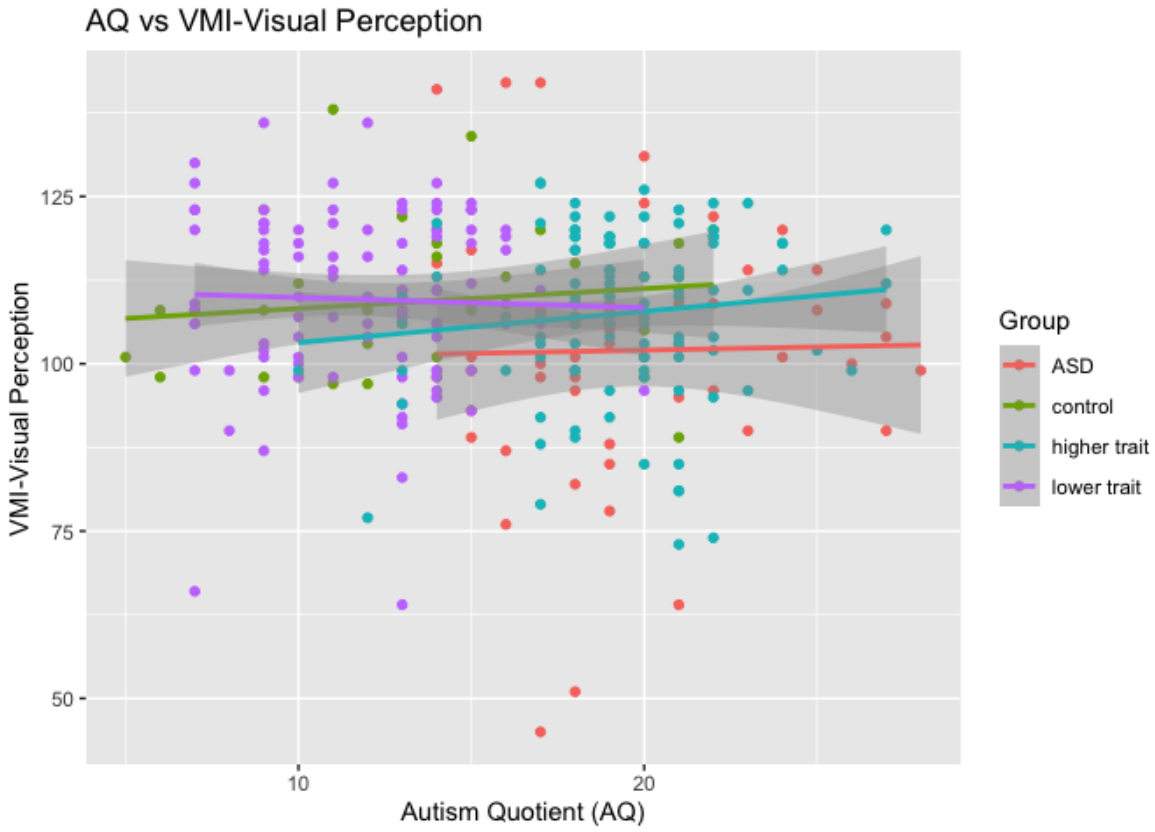
```
ggplot(dat2, aes(x = AQ, y = `VMI-integration`, color = group)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(color = "Group") +
  xlab("Autism Quotient (AQ)") +
  ylab("VMI-integration") +
  ggtitle("AQ vs VMI-integration")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



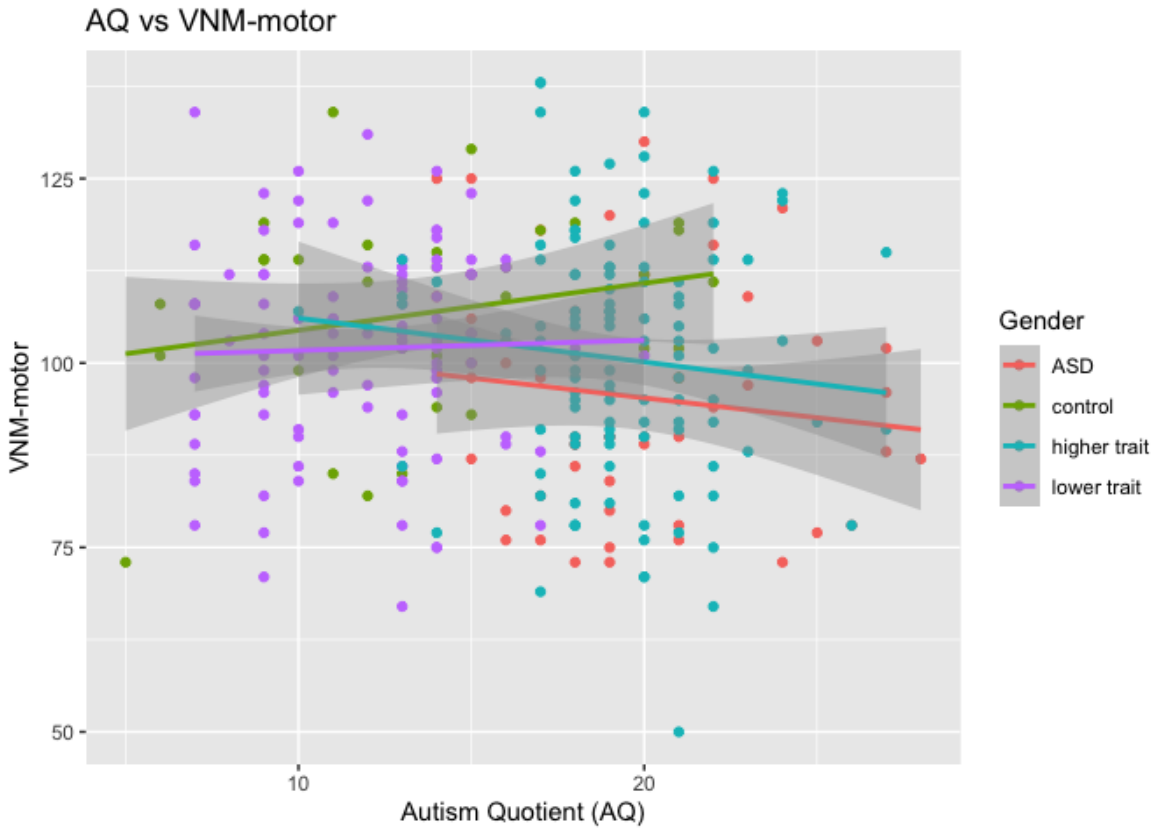

```
ggplot(dat2, aes(x = AQ, y = `VMI-Visual Perception`, color = group)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(color = "Group") +
  xlab("Autism Quotient (AQ)") +
  ylab("VMI-Visual Perception") +
  ggtitle("AQ vs VMI-Visual Perception")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



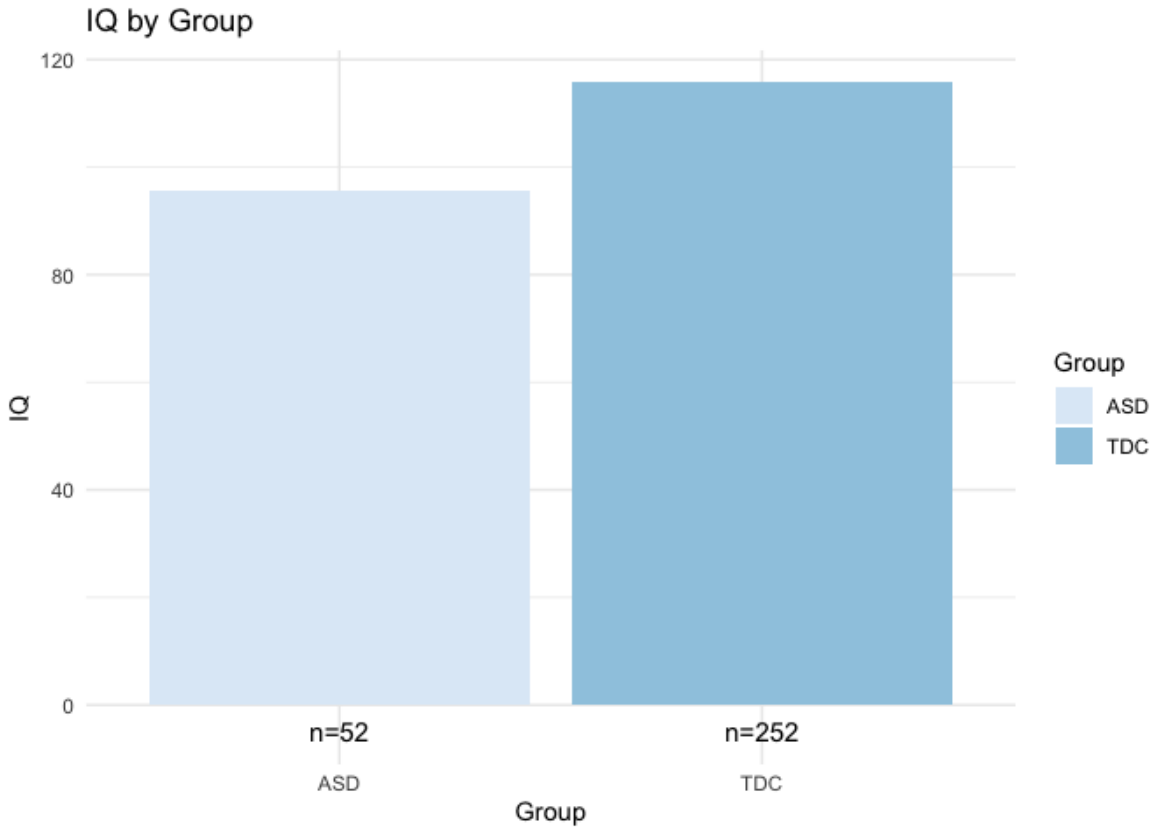

```
ggplot(dat2, aes(x = AQ, y = `VNM-motor`, color = group)) +
  geom_point() +
  geom_smooth(method="lm") +
  labs(color = "Gender") +
  xlab("Autism Quotient (AQ)") +
  ylab("VNM-motor") +
  ggtitle("AQ vs VNM-motor")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Based on the graphs, there is an obvious decrease in IQ in the autism group. There doesn't seem to be any other relationships worth noting here. Let's look further into the IQ differences in the two groups.

```
ggplot(data = dat2, mapping = aes(x = group2, y = IQ, fill = group2)) +
stat_summary(fun = mean, geom = "bar") +
stat_n_text(y.expand.factor = 1.0) +
xlab("Group") +
ylab("IQ") +
ggtitle("IQ by Group")+
theme_minimal() +
guides(fill = guide_legend(title = "Group")) +
scale_fill_brewer(palette="Blues")
```




```
formattable(dat2 %>% group_by(group2)
            %>% summarize(mean(IQ)))
```

group2	mean(IQ)
ASD	95.75885
TDC	115.70313

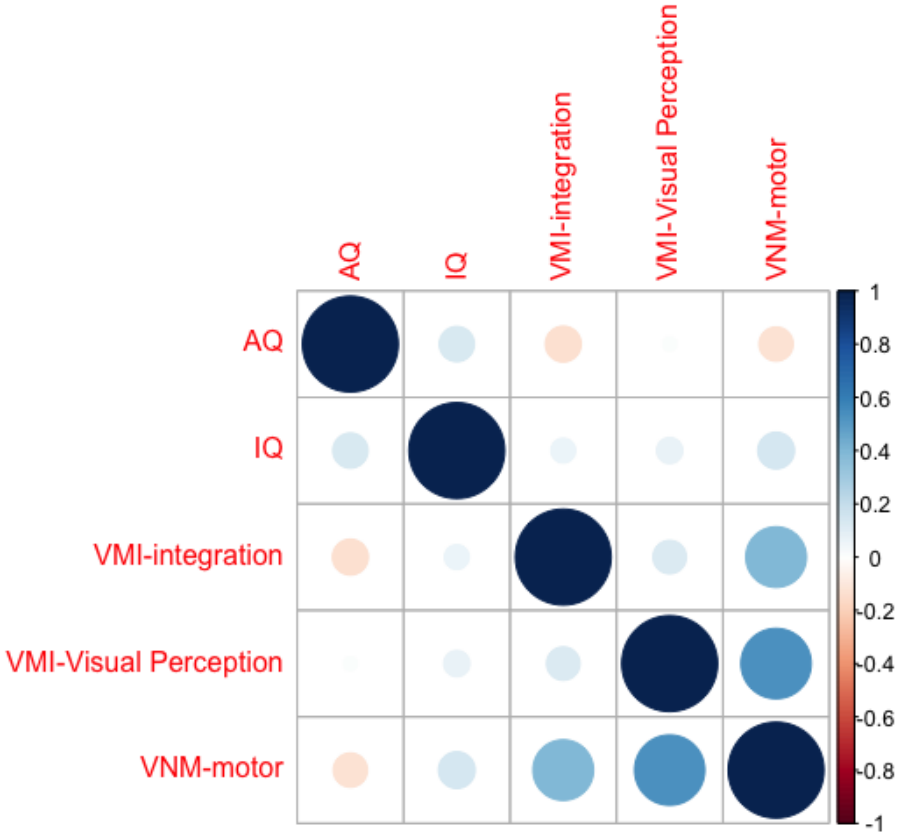
```
(t5 <- t.test(IQ ~ group2, dat2, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  IQ by group2
## t = -9.6161, df = 302, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group ASD and group TDC is not eq
## 95 percent confidence interval:
##  -24.02572 -15.86286
## sample estimates:
## mean in group ASD mean in group TDC
##          95.75885          115.70313
```

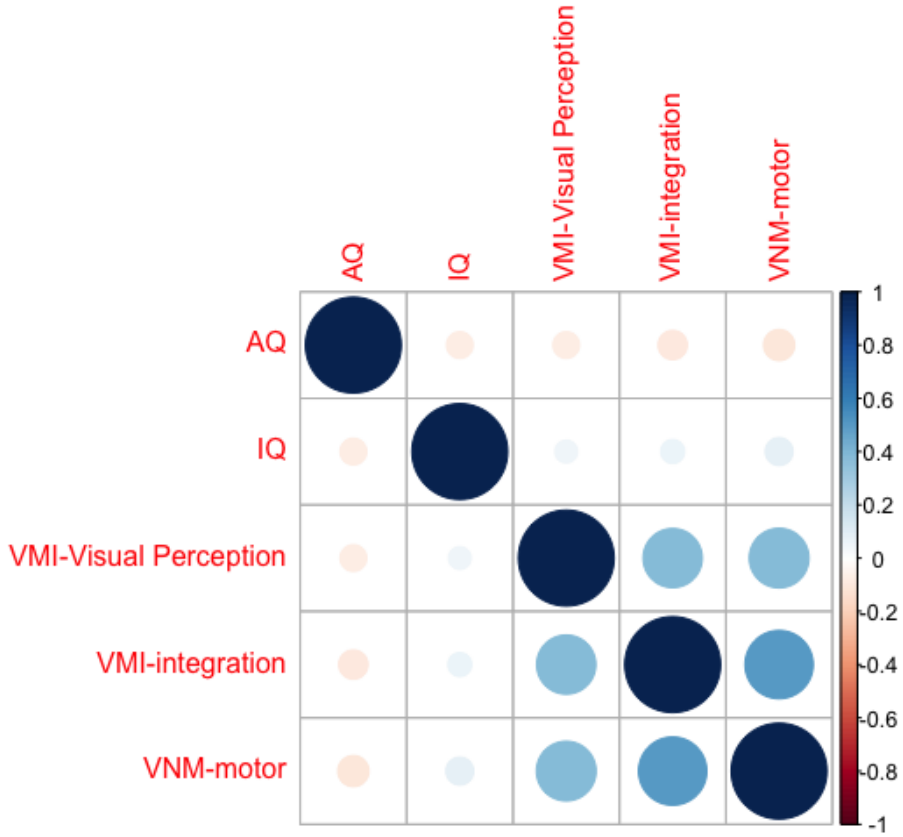
We can see that the typically developing group obviously has a higher IQ, which is then confirmed by the t-test significance in the p-value.

Let's look into correlations and see if any of the variables are correlated within the autism group, and then within all groups.

```
dat4 <- dat3 %>% select(c(AQ, IQ, `VMI-integration`, `VMI-Visual Perception`, `VNM-
        motor`))
corrplot(cor(dat4), order = "hclust")
```




```
dat5 <- dat2 %>% select(c(AQ, IQ, `VMI-integration`, `VMI-Visual Perception`, `VNM-
        motor`))
corrplot(cor(dat5), order = "hclust")
```



In both groups, there is a cluster of all of the VMI/VNM motor testing mechanisms being correlated to each other, which makes sense. The other tests do not correlate very well.

Further Considerations

This project looked at two data sources, one from a research course I took that I administered myself, and another from Plos One. Some limitations is the small sample size in Data Set One (n=30). Additionally, I would argue that the sample size was low in Data Set Two for the autism group as well (n=~50). Another limitation is the lack of standardized testing in Data Set One, which I plan to work on changing in the future with this study. Data Set Two showed some bias with higher IQs than average in the typically developing children's group and a slightly lower than average IQ for the autism group.

If I were to do this project again, I might have looked for a data set that I could preform multiple linear regressions on. Finding a data set was one set back as finding openly available autism data sets can be hard. I overall enjoyed this project.

Sources

CDC: 1 in 44 American Children Are Diagnosed With Autism. (2021, December 13). Verywell Health. https://www.verywellhealth.com/cdc-autism-children-5212627

Diagnostic Criteria | Autism Spectrum Disorder (ASD) | NCBDDD | CDC. (2020, June 29). Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/autism/hcp-dsm.html

Zhang, M., Jiao, J., Hu, X., Yang, P., Huang, Y., Situ, M., Guo, K., Cai, J., & Huang, Y. (2020). Exploring the spatial working memory and visual perception in children with autism spectrum disorder and general population with high autism-like traits. *PLOS ONE, 15(7)*, e0235552. https://doi.org/10.1371/journal.pone.0235552

Autism Spectrum Disorder. (2022). [Photograph]. https://www.dreamresearchgroup.com/2014/02/autism-spectrum-disorder.html