Exploratory Data Analysis

Gabriel Lahman
Rowan Lavelle
Jordan Majoros
Matthew Zelenin

11/11/2019

```r
library(dplyr)
library(tidyr)
library(alr4)

kki_demographics = read.csv('data/KKI_demographicInfo.csv')
kki_handedness = read.csv('data/KKI_handedness.csv')
kki_mABC = read.csv('data/KKI_movementAssessmentBatteryforChildren.csv')
kki_SRS = read.csv('data/KKI_SocialResponsivenessScaleQuestionnaire.csv')
kki_WISC = read.csv('data/KKI_WechslerIntelligenceScaleforChildren.csv')


full_dat = full_join(kki_demographics, kki_mABC)
full_dat = full_join(full_dat, kki_SRS)
full_dat = full_join(full_dat, kki_WISC)
full_dat = full_join(full_dat, kki_handedness)
full_dat = full_dat %>% drop_na(c('SecondaryDiagnosis', 'CurrentlyNotTakingMeds', 'CurrentlyTakingAtomo
full_dat = subset(full_dat, !(is.na(SRS_VERSION) & is.na(WISC_VERSION)))
WISC4_full_dat = subset(full_dat, WISC_VERSION == 4)
WISC5_full_dat = subset(full_dat, WISC_VERSION == 5)
```
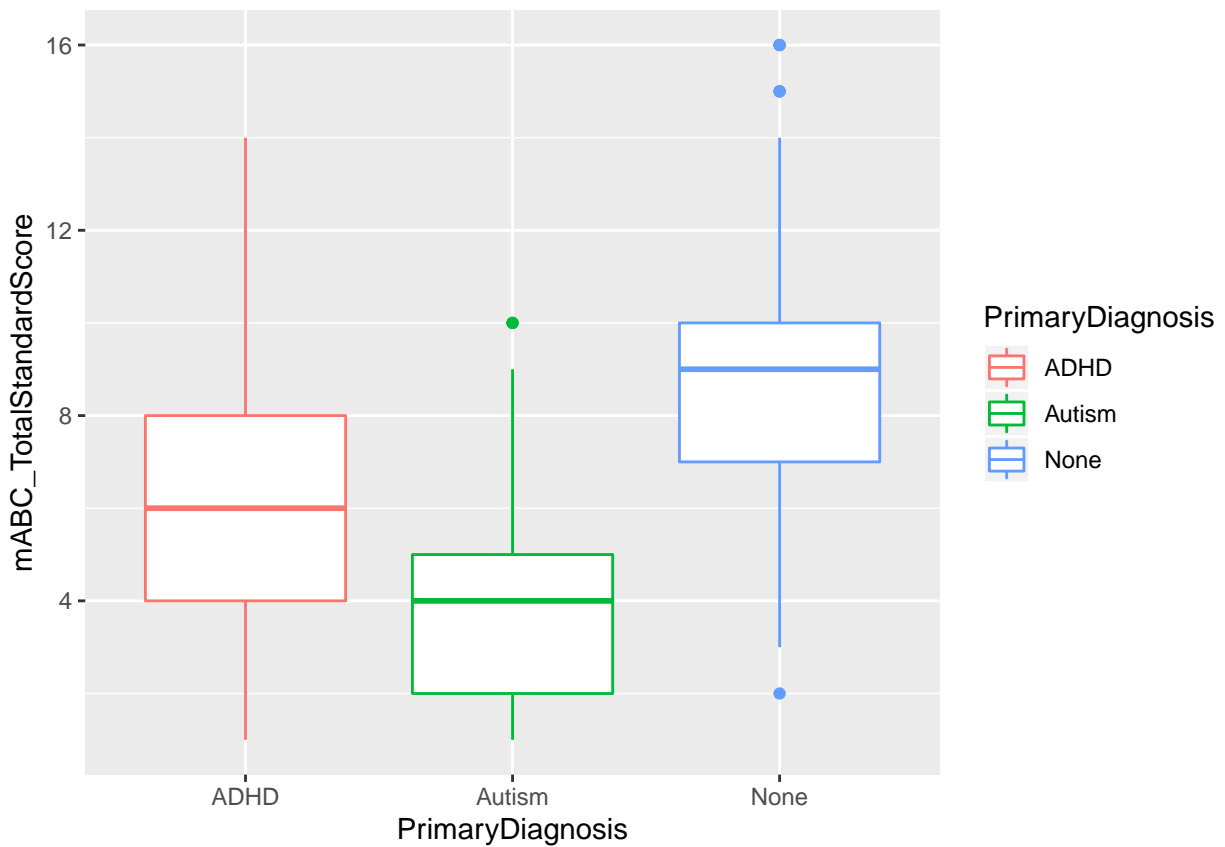
Research Question

Is there a statistically significant difference in predictive power of the motor skills vs. social responsivness
model when we include intelligence metrics as predictors?

```r
library(ggplot2)
ggplot(full_dat, aes(x=PrimaryDiagnosis, y=mABC_TotalStandardScore, color=PrimaryDiagnosis)) +
  geom_boxplot()
```
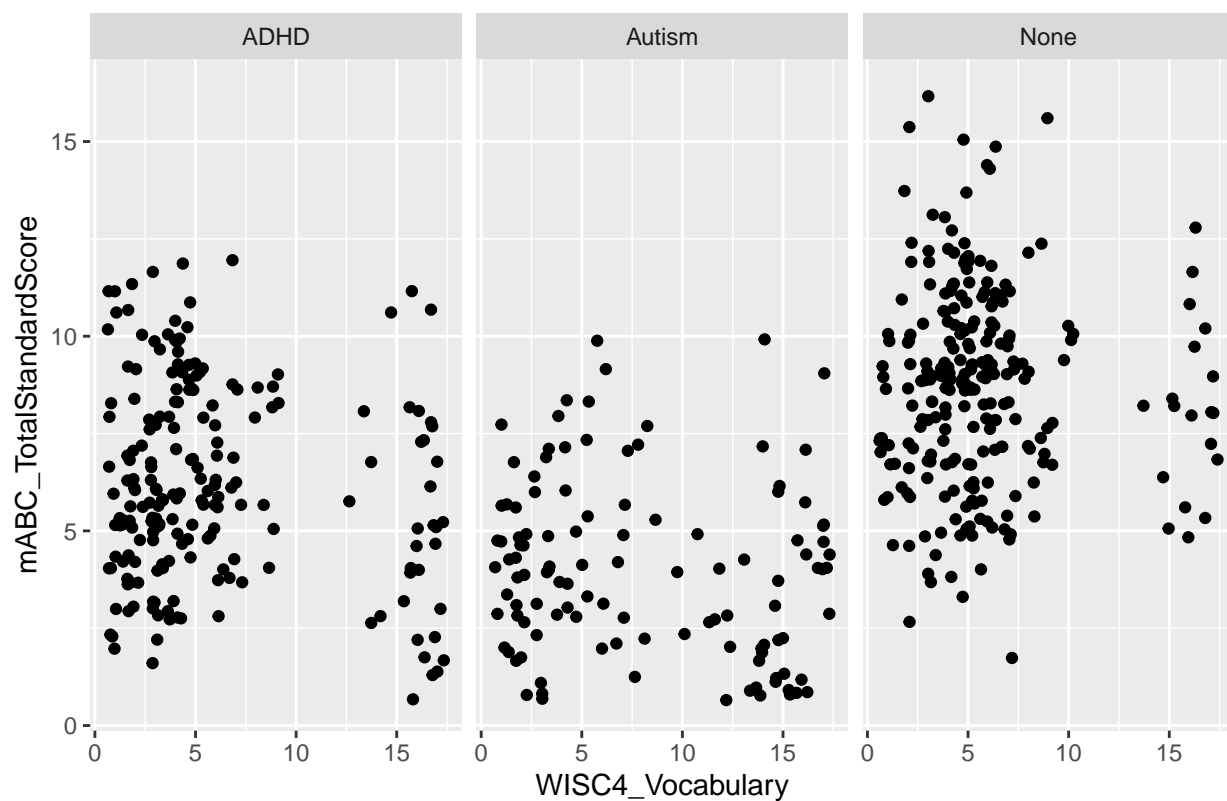
```
ggplot(full_dat, aes(x=mABC_AGE, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  ggtitle('Age vs. mABC Total Standard Score')
```

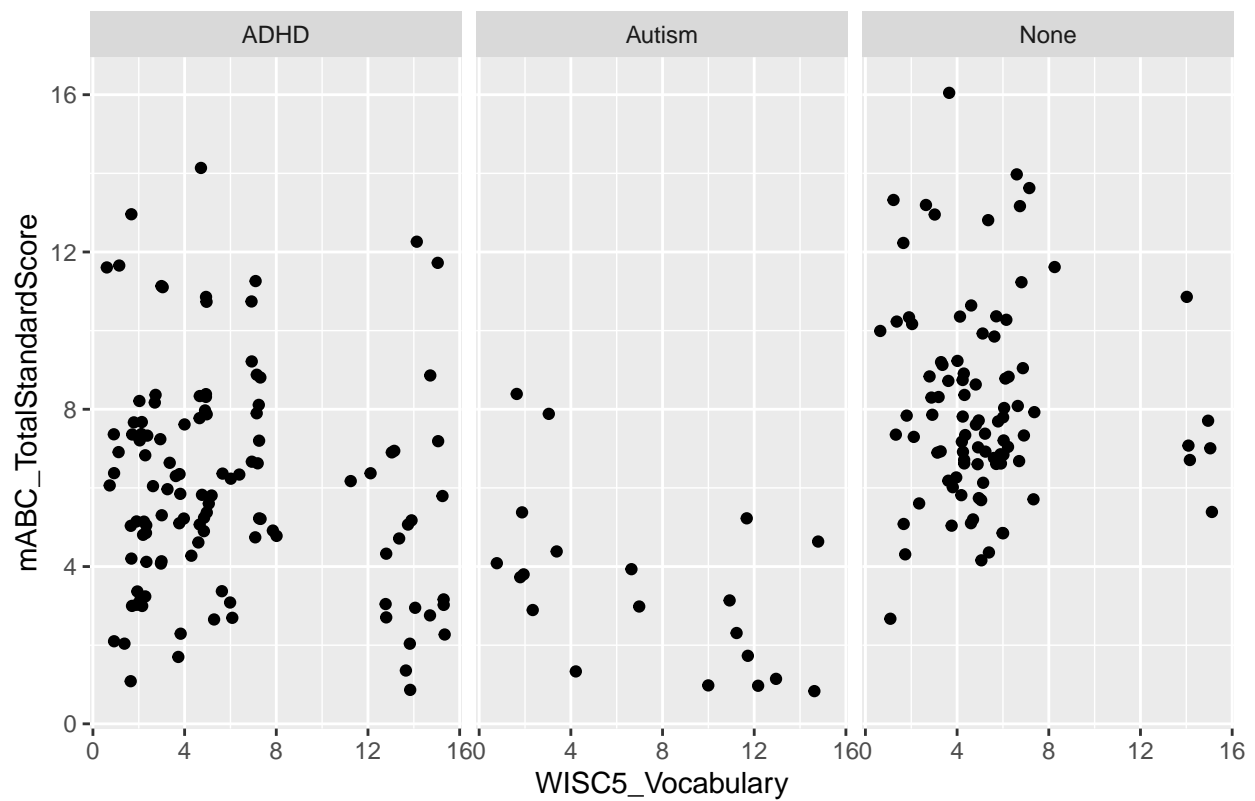## Age vs. mABC Total Standard Score
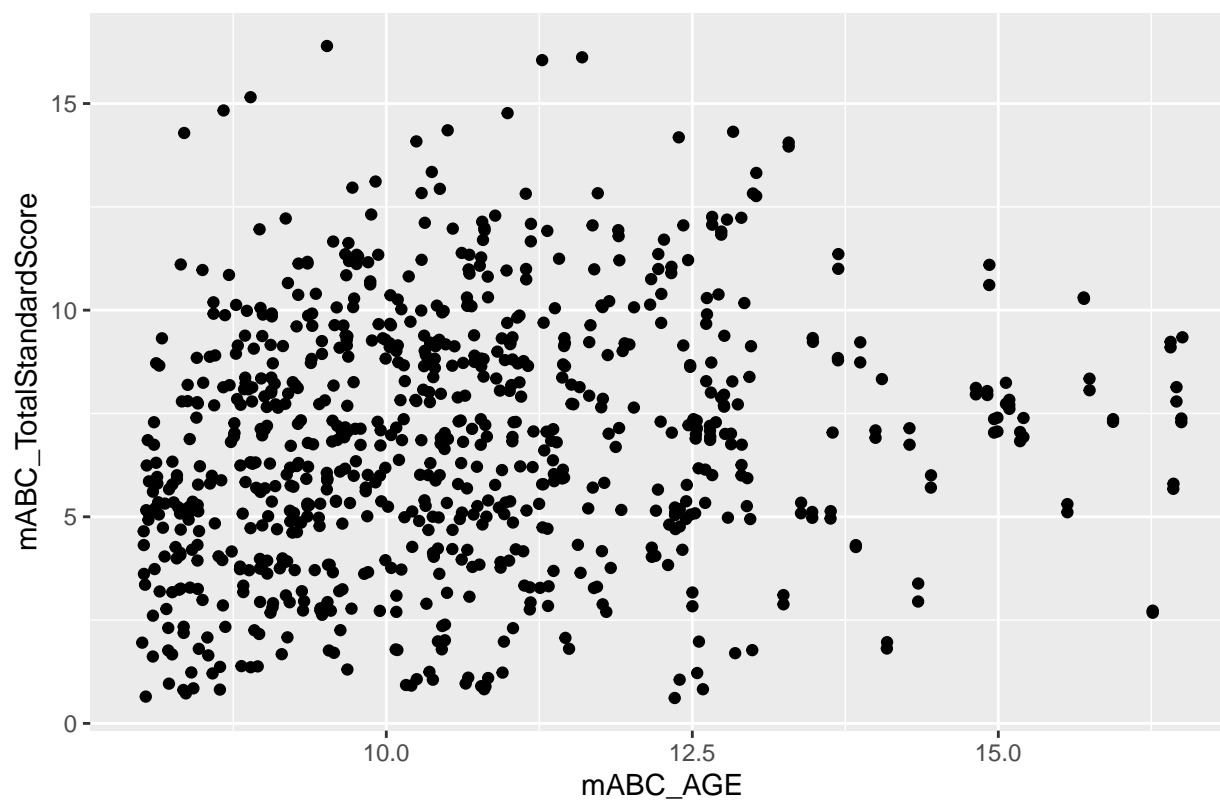

```
ggplot(full_dat, aes(x= WISC4_Vocabulary, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_Vocabulary vs. mABC Total Standard Score')
```

WISC4_Vocabulary vs. mABC Total Standard Score
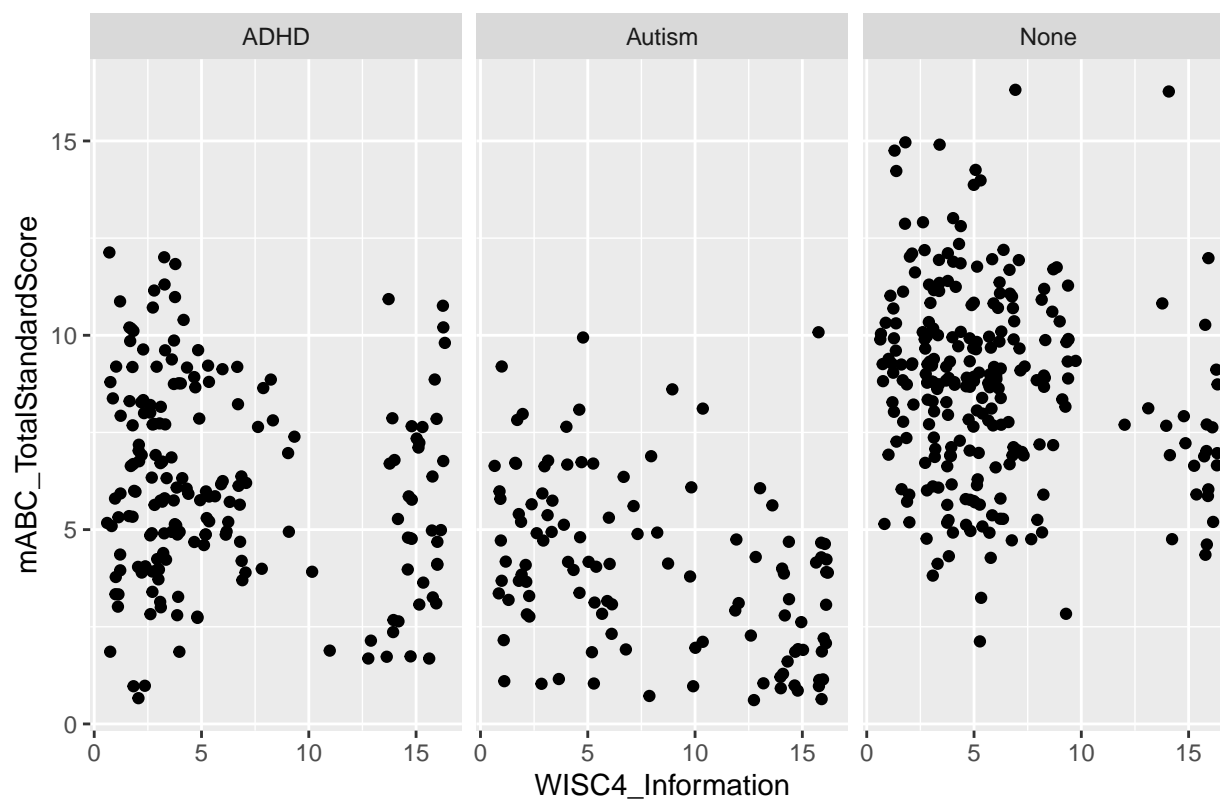


```
ggplot(full_dat, aes(x= WISC5_Vocabulary, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC5_Vocabulary vs. mABC Total Standard Score')
```

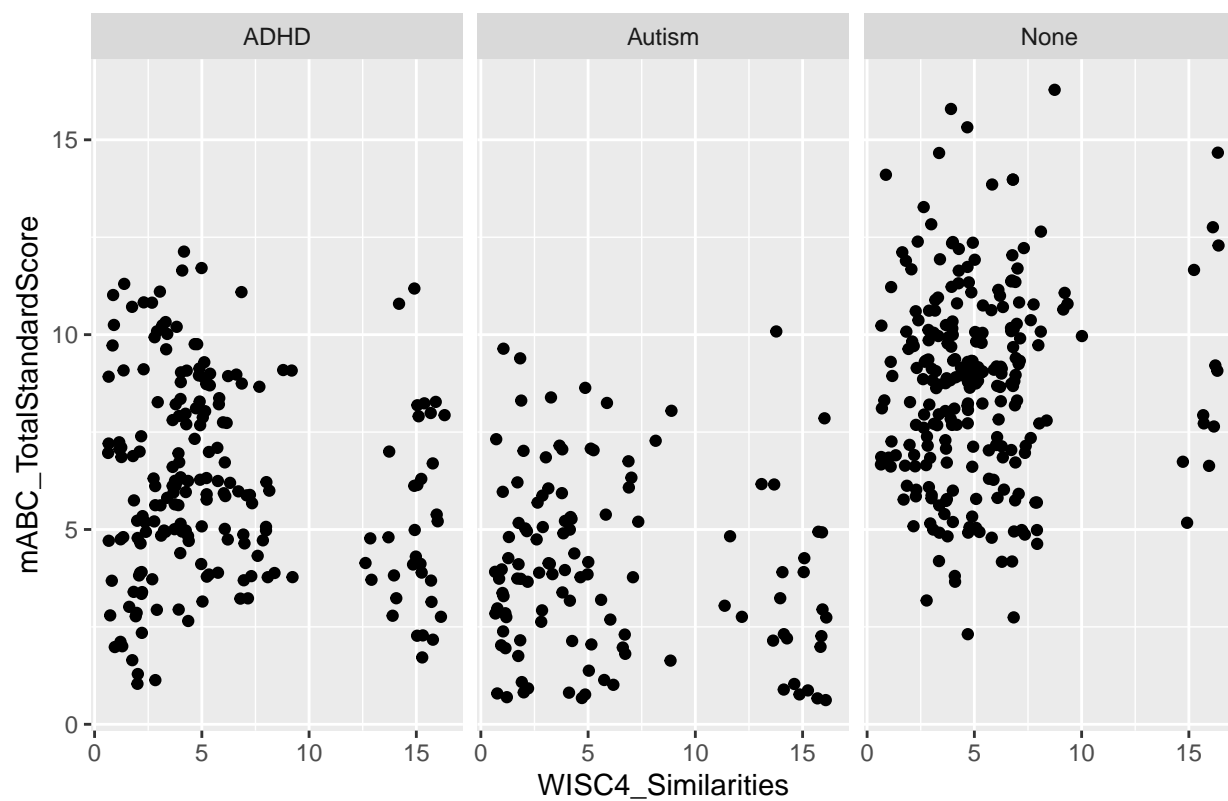# WISC5_Vocabulary vs. mABC Total Standard Score


```
ggplot(full_dat, aes(x= WISC4_Information, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_Information vs. mABC Total Standard Score')
```

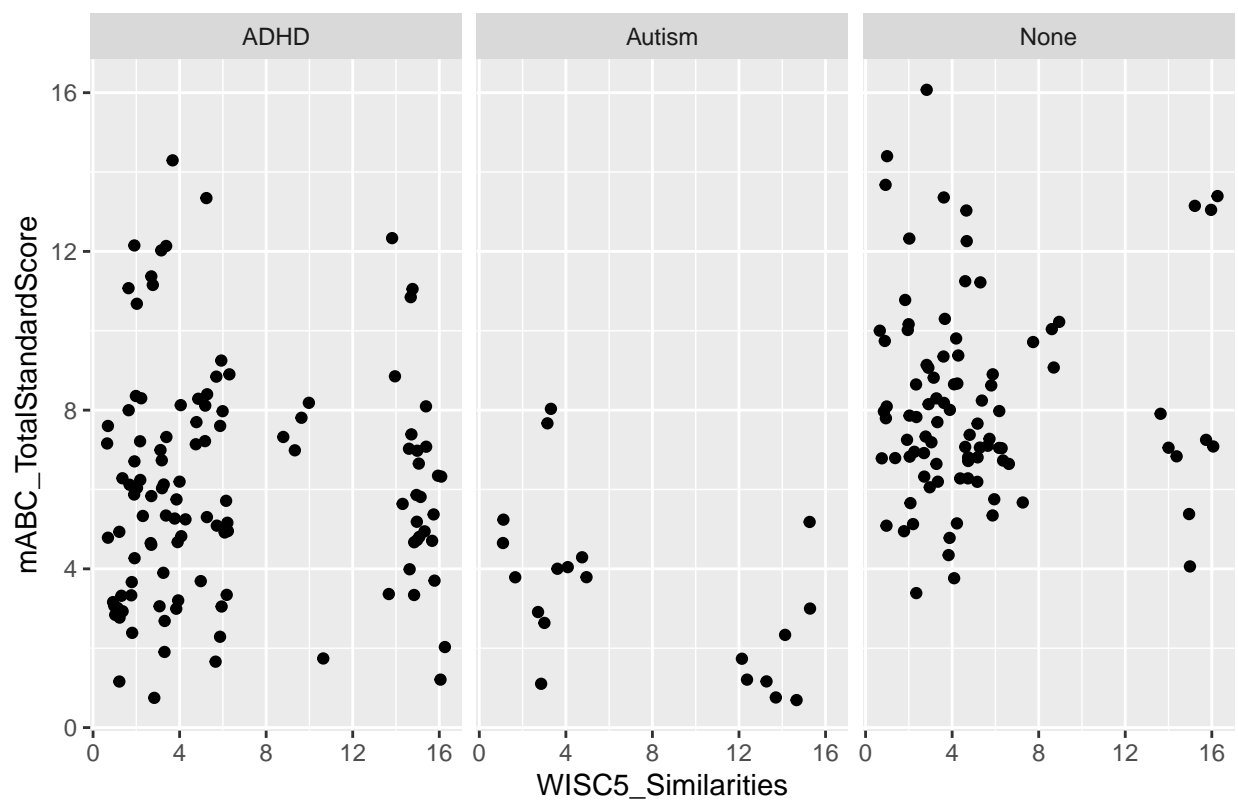## WISC4_Information vs. mABC Total Standard Score


```
ggplot(full_dat, aes(x= WISC4_Similarities, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_Similarities vs. mABC Total Standard Score')
```

WISC4_Similarities vs. mABC Total Standard Score


```r
ggplot(full_dat, aes(x= WISC5_Similarities, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC5_Similarities vs. mABC Total Standard Score')
```
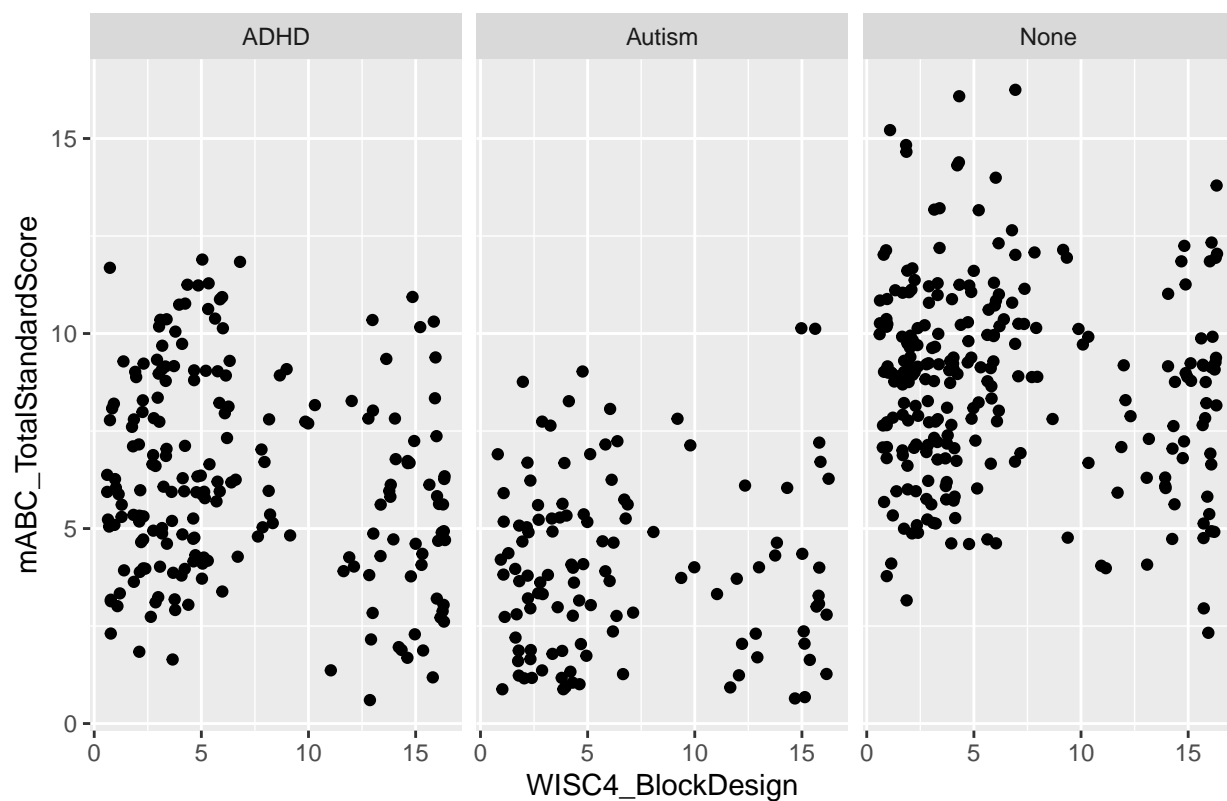
WISC5_Similarities vs. mABC Total Standard Score


```
ggplot(full_dat, aes(x= WISC4_BlockDesign, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_Design vs. mABC Total Standard Score')
```
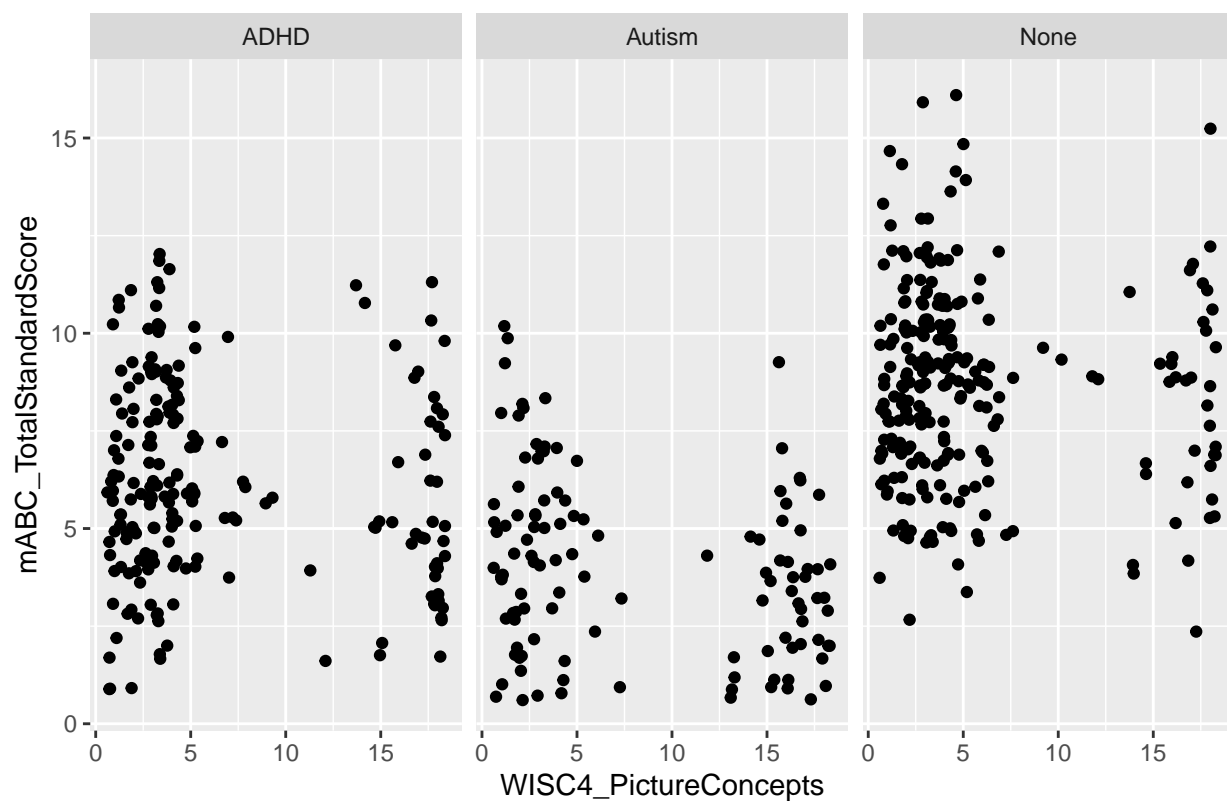
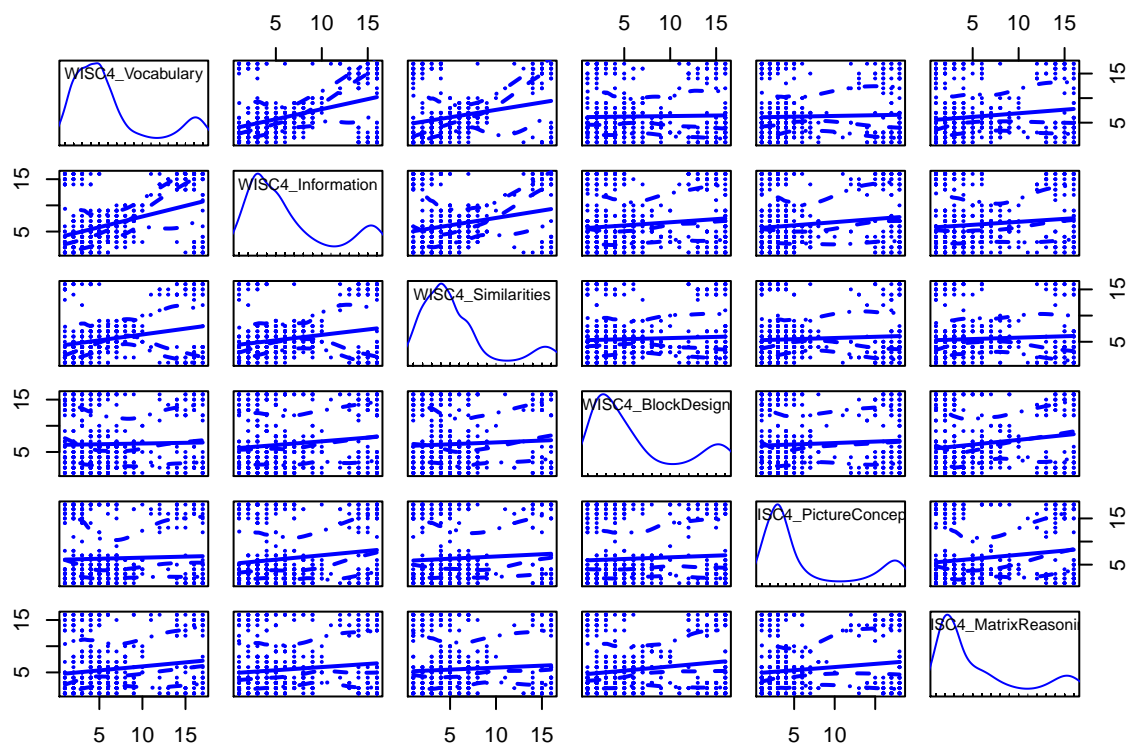# WISC4_Design vs. mABC Total Standard Score


```
ggplot(full_dat, aes(x= WISC4_PictureConcepts, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_PictureConcepts vs. mABC Total Standard Score')
```
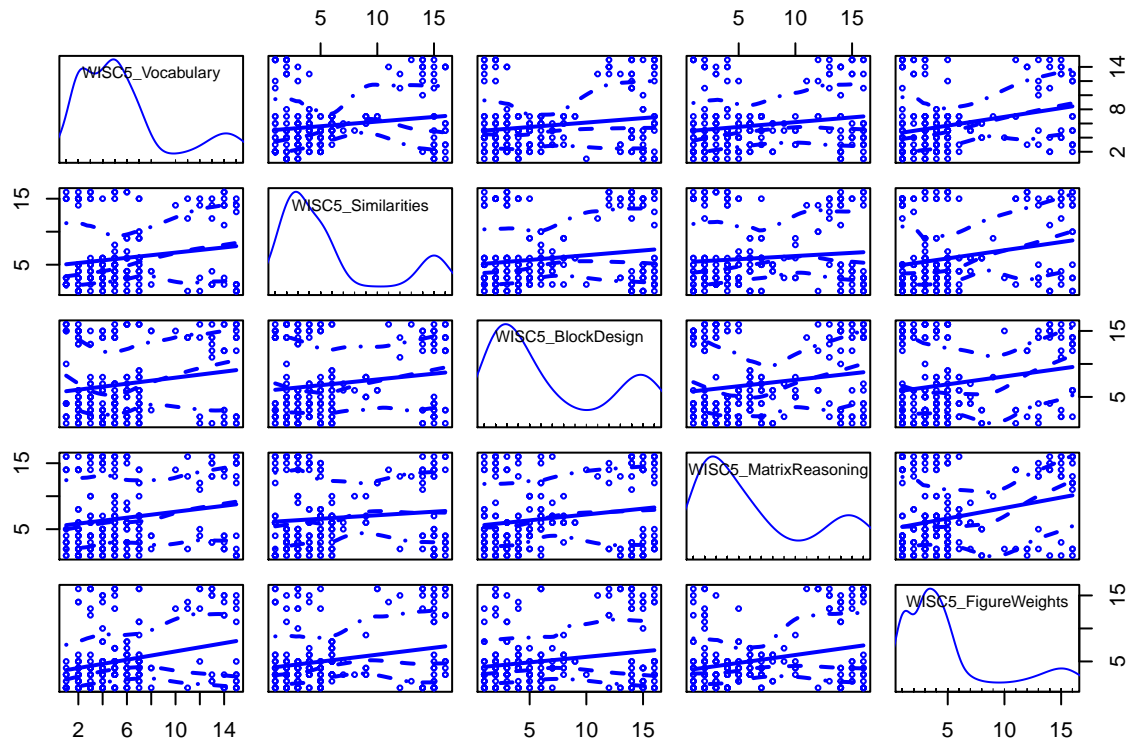
WISC4_PictureConcepts vs. mABC Total Standard Score

```
scatterplotMatrix(WISC4_full_dat[c('WISC4_Vocabulary', 'WISC4_Information', 'WISC4_Similarities', 'WISC
```

```
scatterplotMatrix(WISC5_full_dat[c('WISC5_Vocabulary', 'WISC5_Similarities', 'WISC5_BlockDesign', 'WISC
```
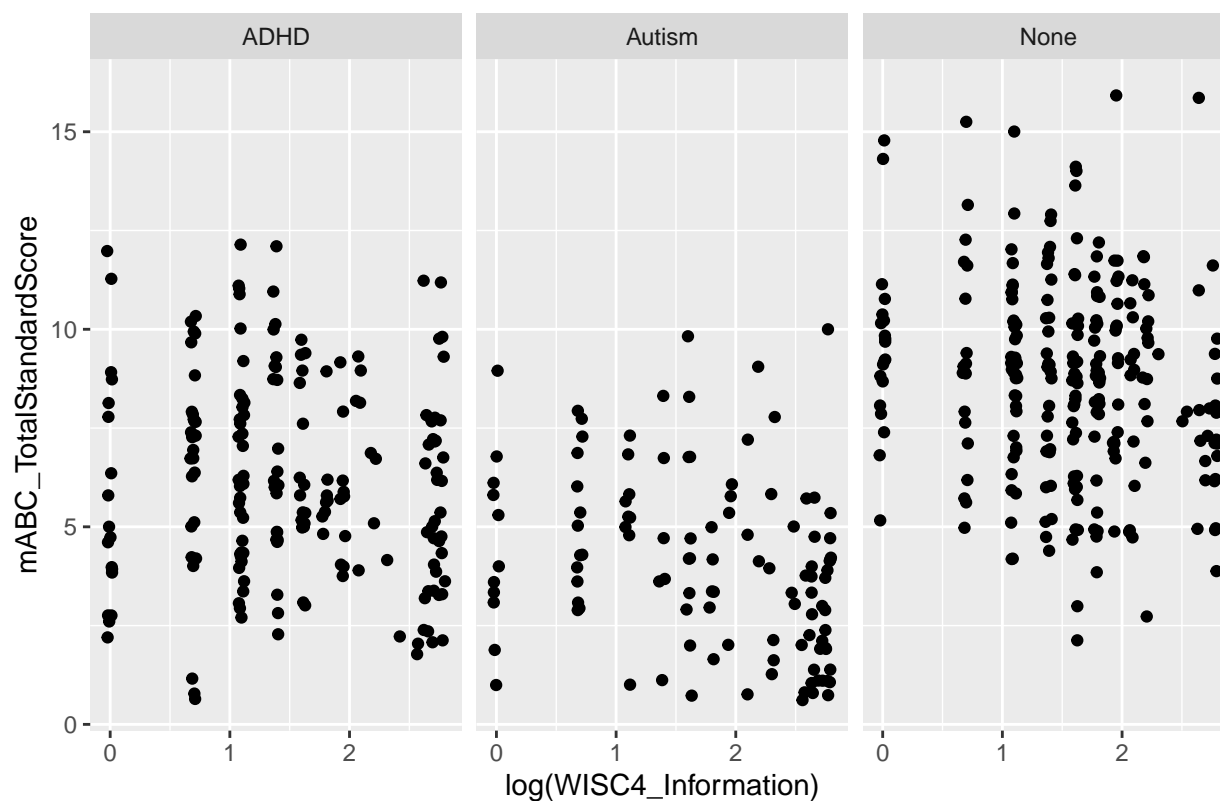


```
ggplot(full_dat, aes(x= log(WISC4_Vocabulary), y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('log(WISC4_Vocabulary) vs. mABC Total Standard Score')
```

log(WISC4_Vocabulary) vs. mABC Total Standard Score
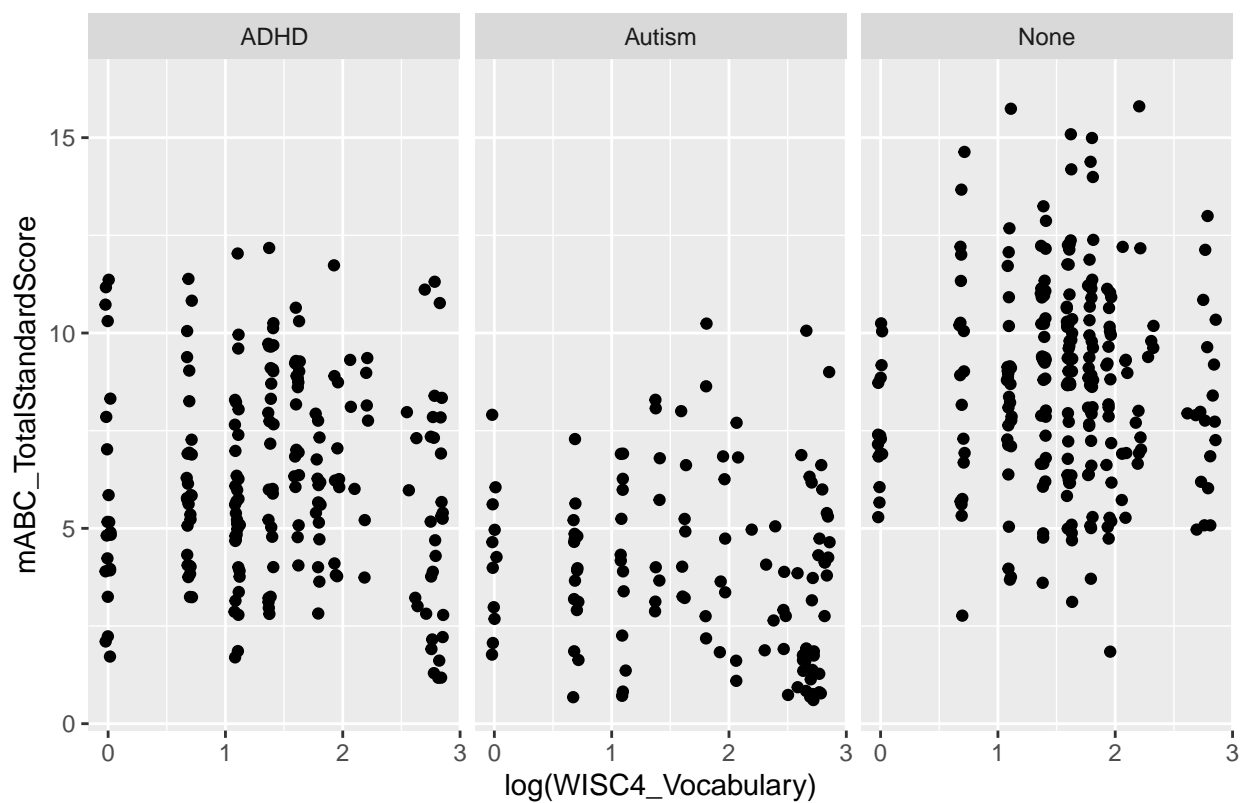

```
ggplot(full_dat, aes(x= log(WISC4_Information), y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('log(WISC4_Information) vs. mABC Total Standard Score')
```

log(WISC4_Information) vs. mABC Total Standard Score


```
ggplot(full_dat, aes(x= WISC4_Vocabulary + WISC4_Information + WISC4_Similarities + WISC4_BlockDesign +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_Design vs. mABC Total Standard Score')
```

```
## Warning: Removed 245 rows containing missing values (geom_point).
```

WISC4_Design vs. mABC Total Standard Score