

# Exploratory Data Analysis

*Gabriel Lahman*

*Rowan Lavelle*

*Jordan Majoros*

*Matthew Zelenin*

*11/11/2019*

```
library(dplyr)
library(tidyr)
library(alr4)
library(ggplot2)
```

## Research Question

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

## Filtering Data

```
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', 'CurrentlyTakingAtomoxetine'))
full_dat = subset(full_dat, visit == 1)
full_dat = subset(full_dat, !is.na(WISC_VERSION) & !is.na(SRS_VERSION))
full_dat = subset(full_dat, ADHD_Subtype %in% c('Combined', 'Hyperactive/Impulsive', 'Inattentive', 'No dx'))
WISC4_full_dat = subset(full_dat, WISC_VERSION == 4)
WISC5_full_dat = subset(full_dat, WISC_VERSION == 5)
```

## Summarizing Data

```
table(factor(full_dat$PrimaryDiagnosis), factor(full_dat$ADHD_Subtype))

##
##      Combined Hyperactive/Impulsive Inattentive No dx
## ADHD          150                3          32     0
## Autism         61                2          30    39
## None           0                 0           0   251
```

```
table(factor(full_dat$WISC_VERSION))
```

```
##
##    4    5
## 536  32
```

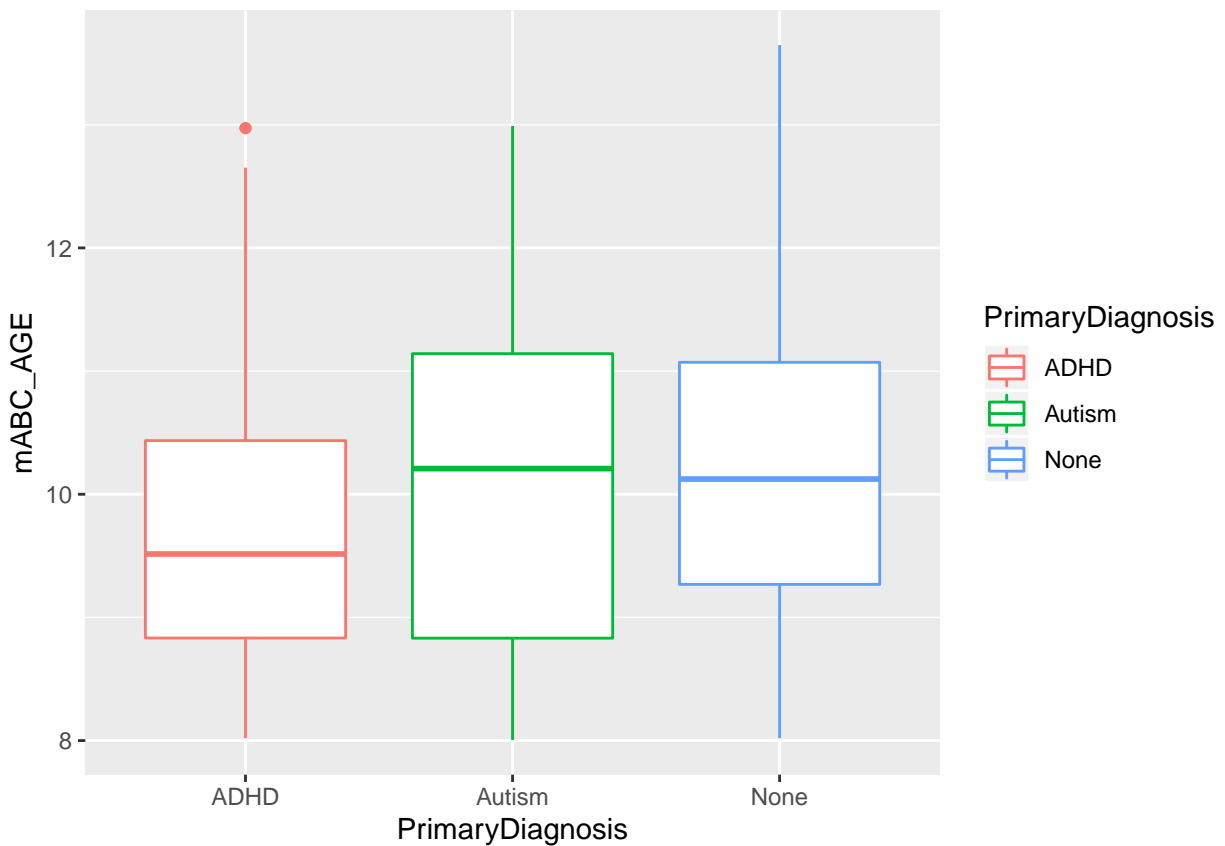
```
table(factor(WISC4_full_dat$PrimaryDiagnosis), factor(WISC4_full_dat$ADHD_Subtype))
```

```
##
##           Combined Hyperactive/Impulsive Inattentive No dx
##   ADHD           150                3           32      0
##   Autism          55                2           20     37
##   None              0                0            0    237
```

## Main Variables of Interest

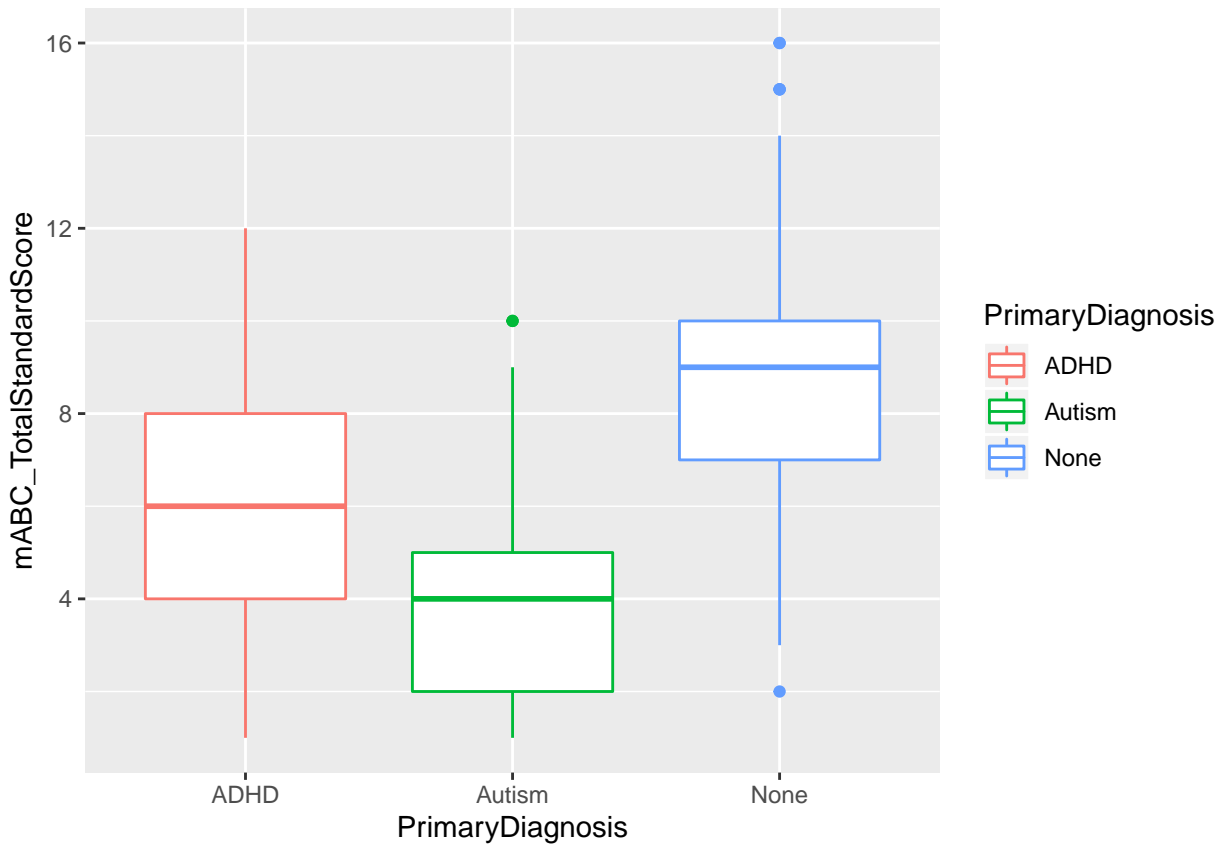
mABC\_AGE

```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=mABC_AGE, color=PrimaryDiagnosis)) +
  geom_boxplot()
```



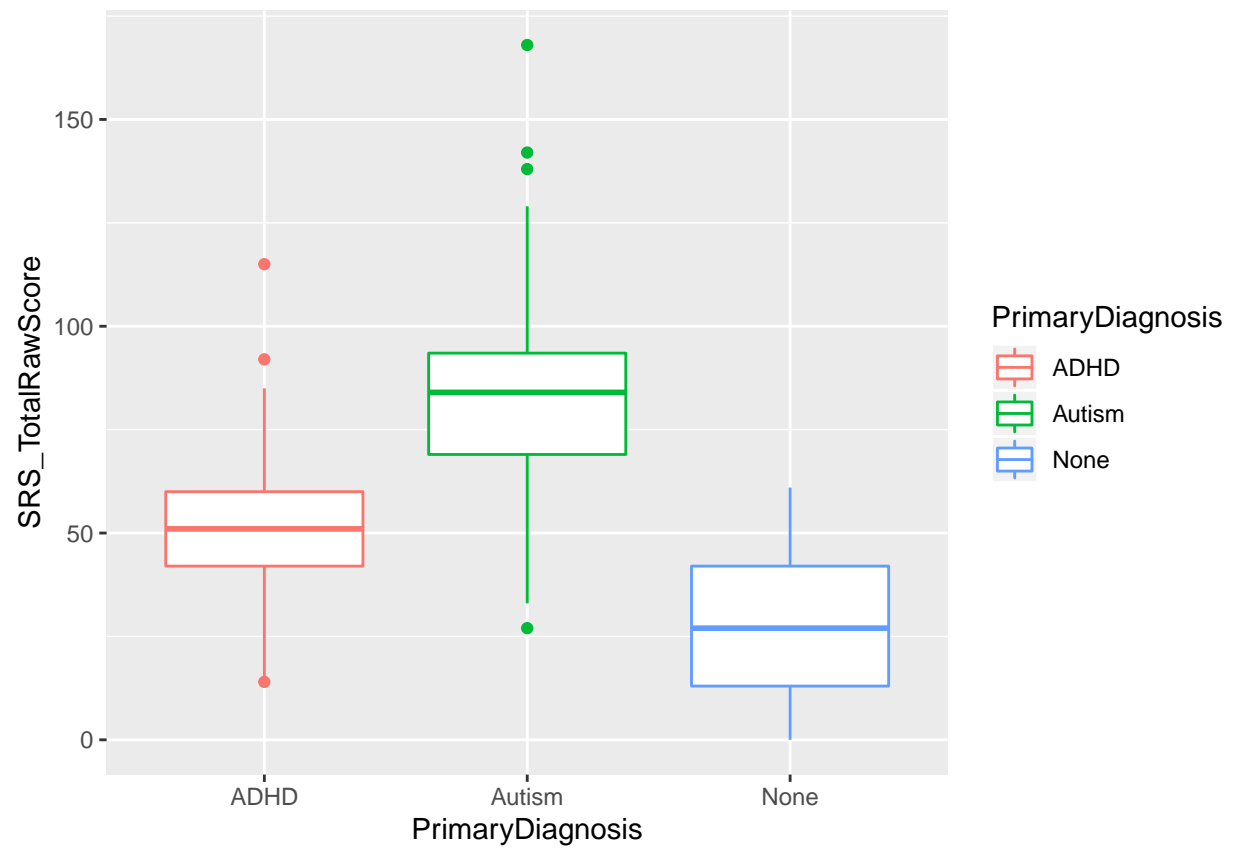
mABC\_TotalStandardScore

```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=mABC_TotalStandardScore, color=PrimaryDiagnosis)) +
  geom_boxplot()
```



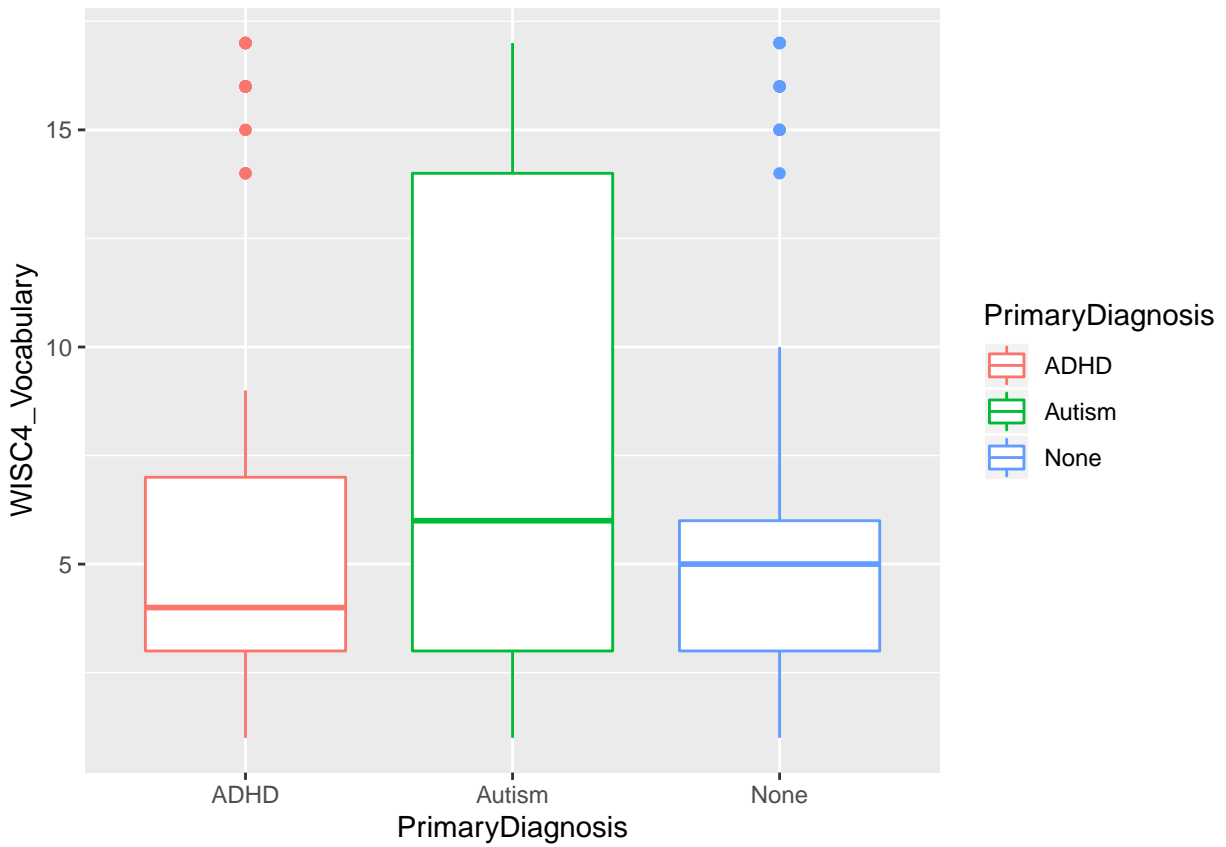
SRS\_TotalRawScore

```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=SRS_TotalRawScore, color=PrimaryDiagnosis)) +  
  geom_boxplot()
```



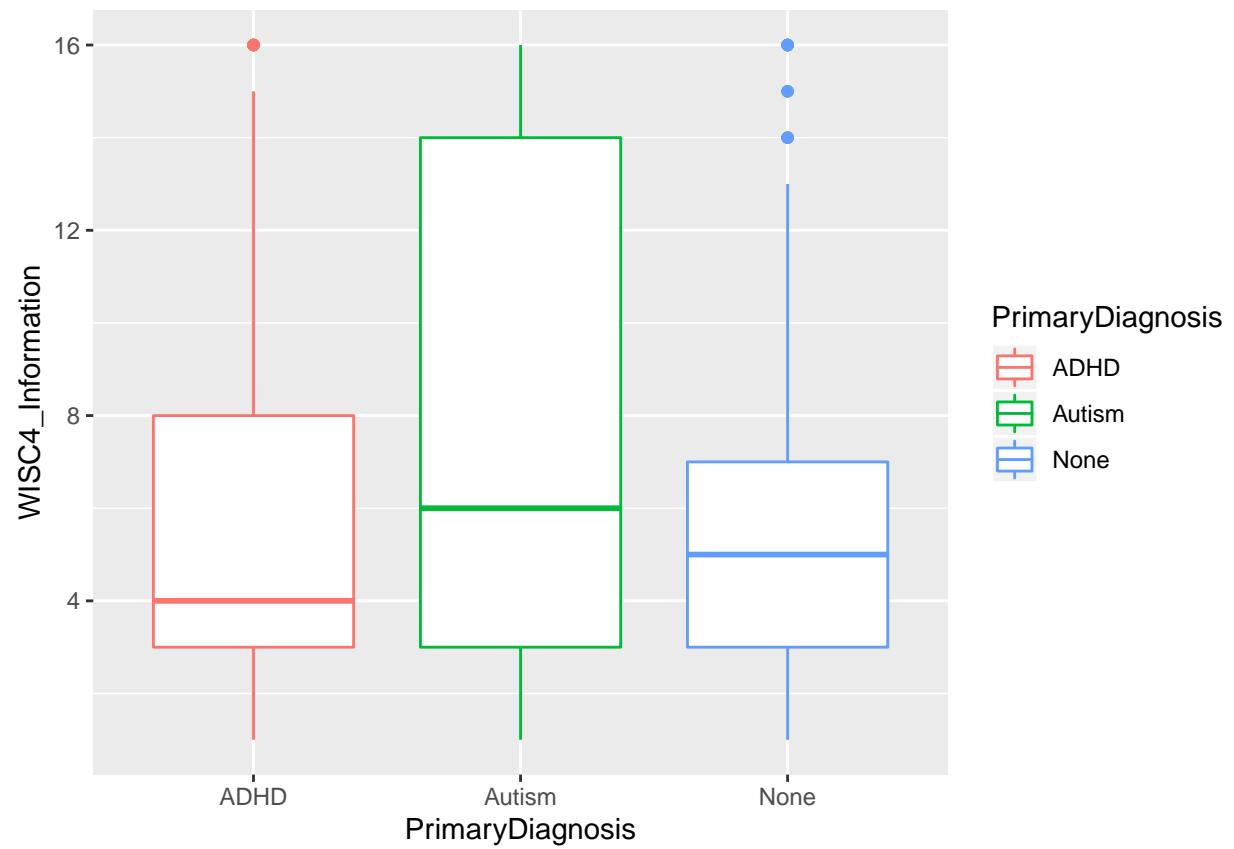
WISC4\_Vocabulary

```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=WISC4_Vocabulary, color=PrimaryDiagnosis)) +  
  geom_boxplot()
```



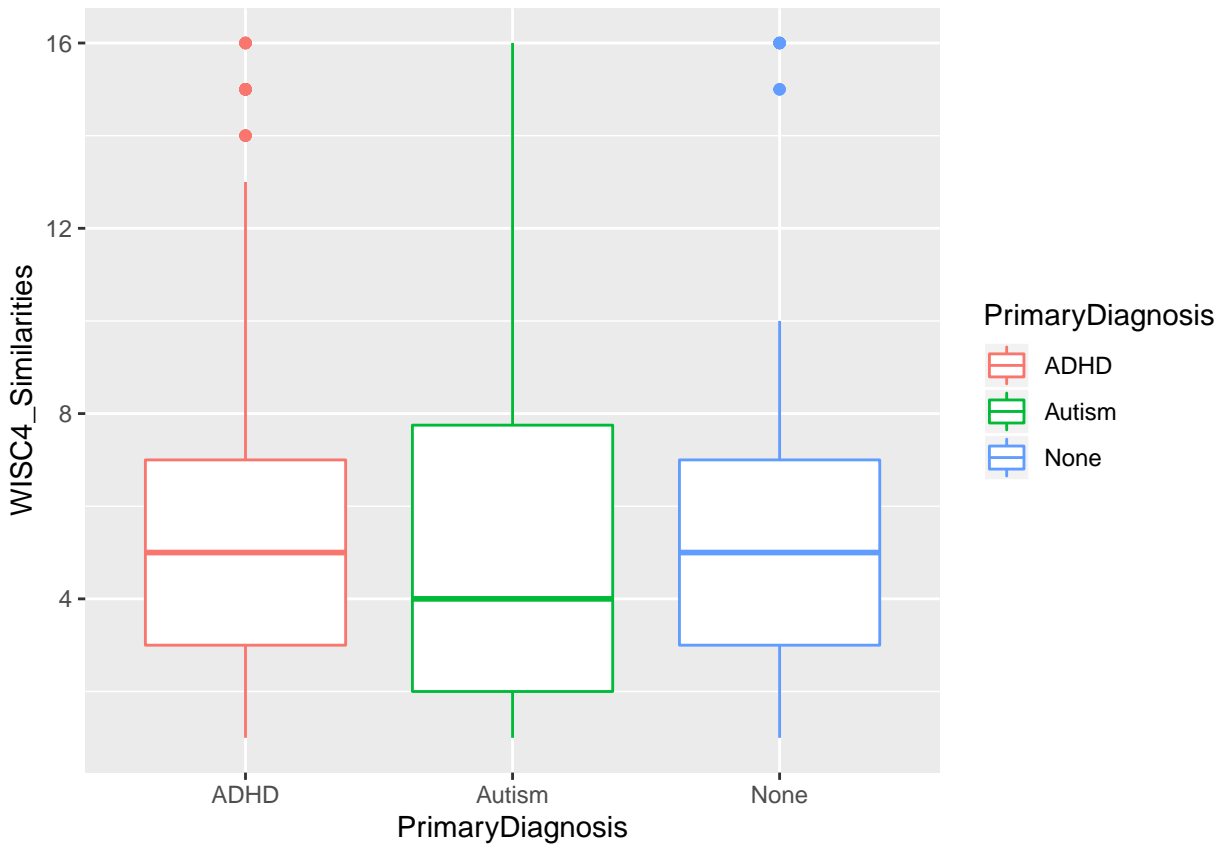
WISC4\_Information

```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=WISC4_Information, color=PrimaryDiagnosis)) +  
  geom_boxplot()
```



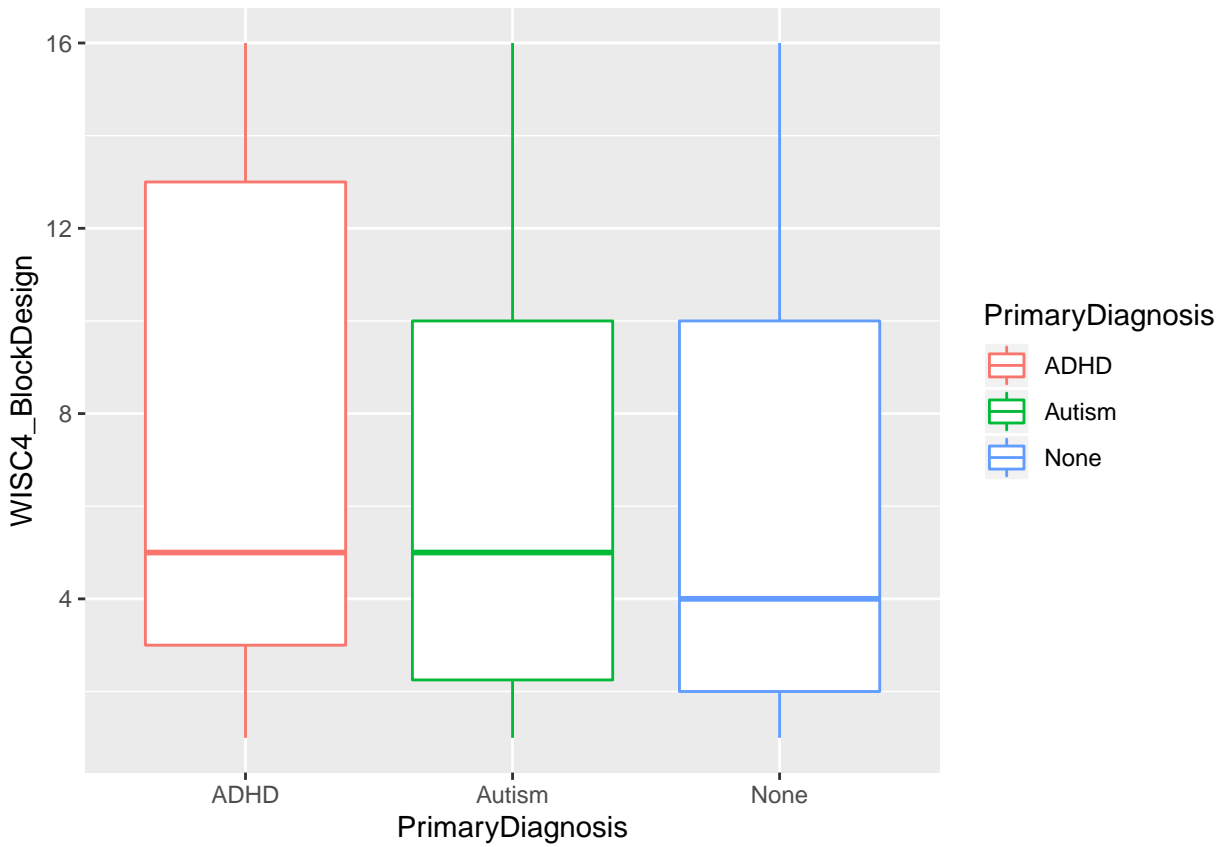
WISC4\_Similarities

```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=WISC4_Similarities, color=PrimaryDiagnosis)) +  
  geom_boxplot()
```



WISC4\_BlockDesign

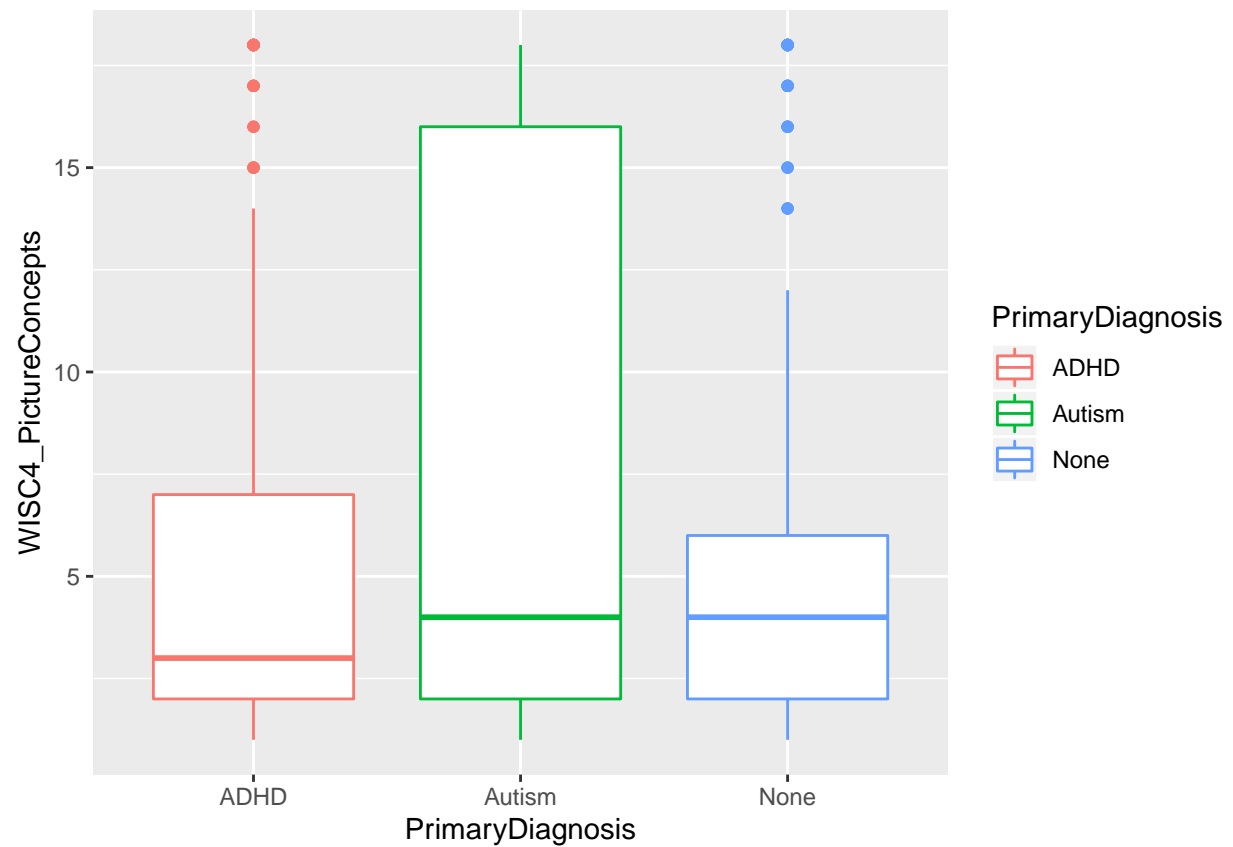
```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=WISC4_BlockDesign, color=PrimaryDiagnosis)) +  
  geom_boxplot()
```



WISC4\_PictureConcepts

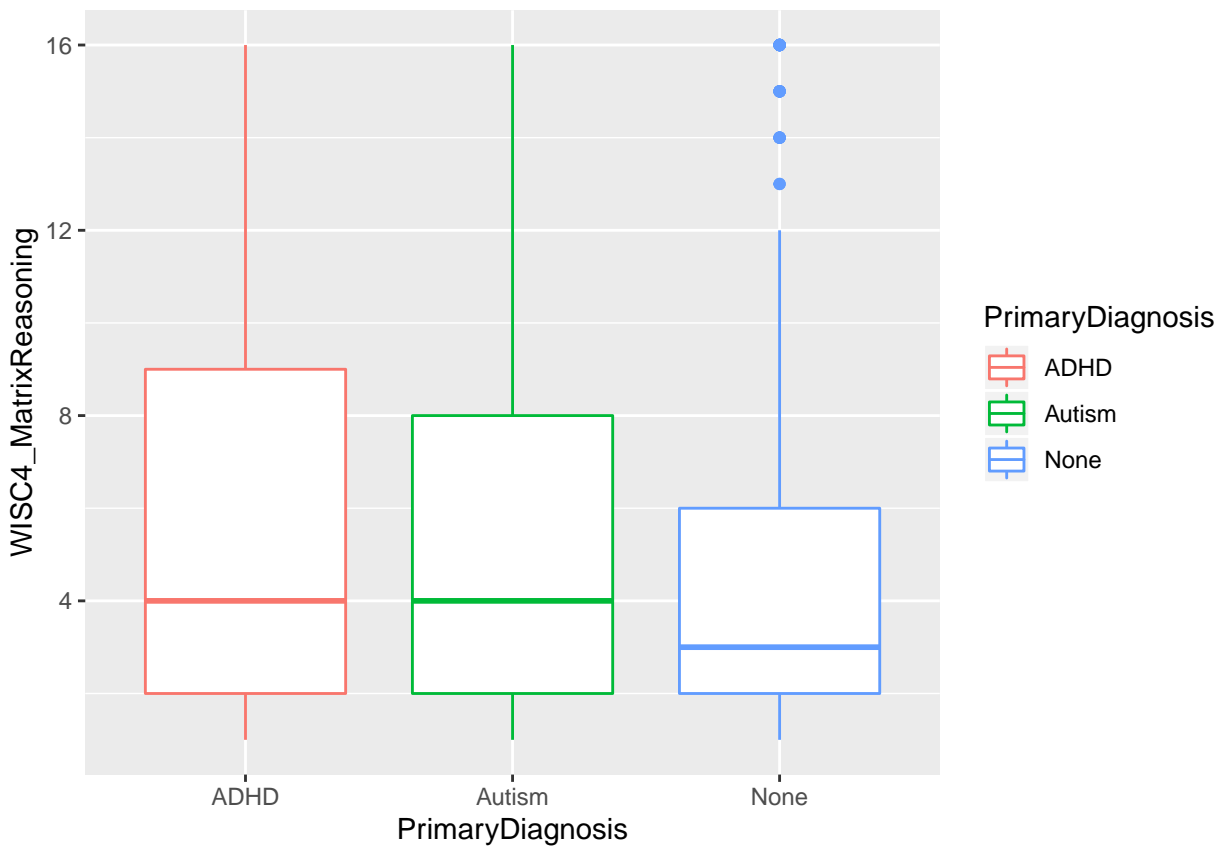
```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=WISC4_PictureConcepts, color=PrimaryDiagnosis)) +  
  geom_boxplot()
```





#### WISC4\_MatrixReasoning

```
ggplot(WISC4_full_dat, aes(x=PrimaryDiagnosis, y=WISC4_MatrixReasoning, color=PrimaryDiagnosis)) +  
  geom_boxplot()
```

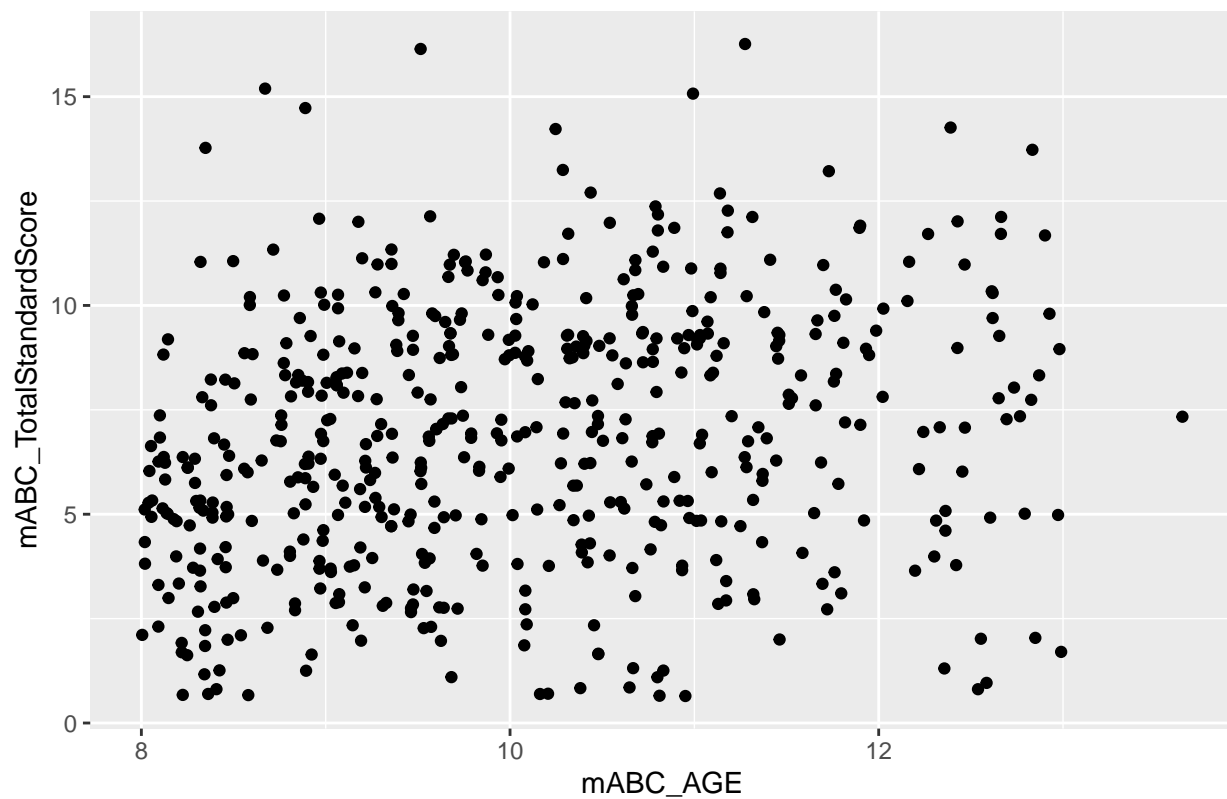


## Plots

mABC\_AGE vs mABC\_TotalStandardScore

```
ggplot(WISC4_full_dat, aes(x=mABC_AGE, y=mABC_TotalStandardScore)) +  
  geom_jitter() +  
  ggtitle('Age vs. mABC Total Standard Score')
```

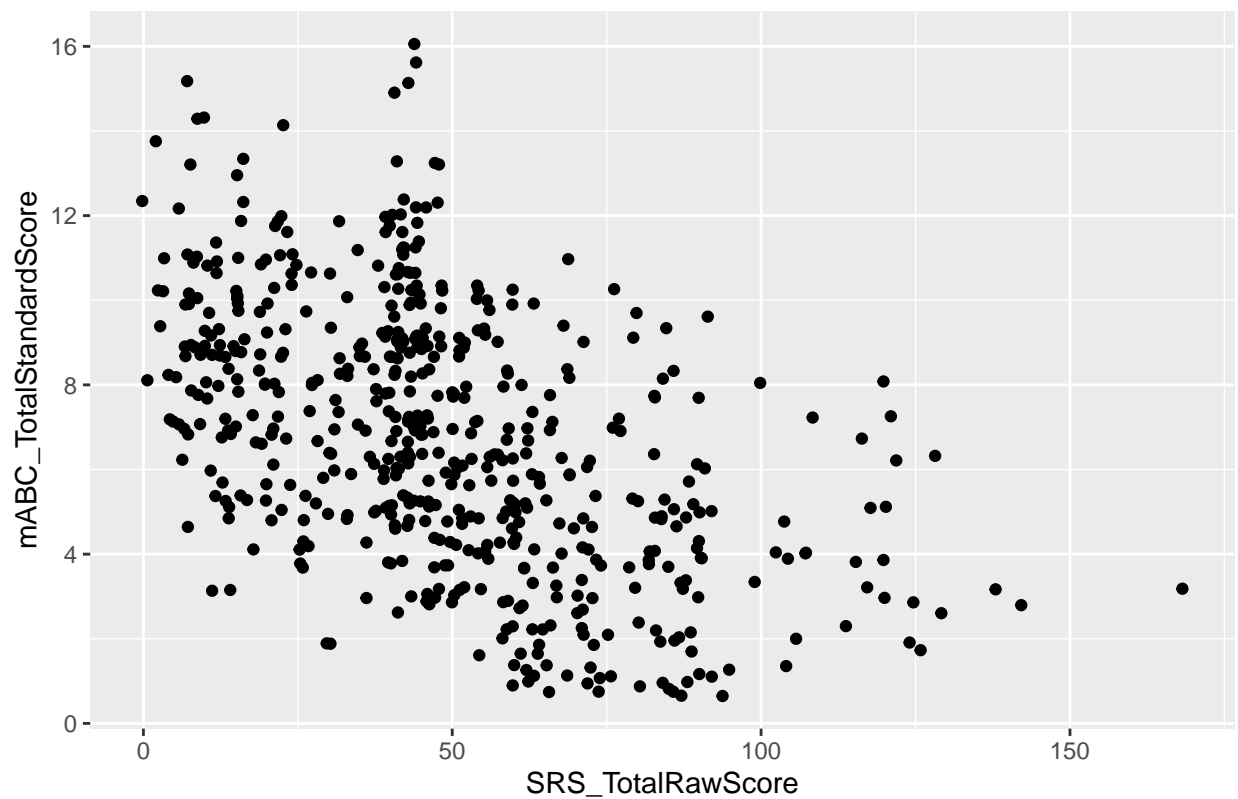
Age vs. mABC Total Standard Score



SRS\_TotalRawScore vs mABC\_TotalStandardScore

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

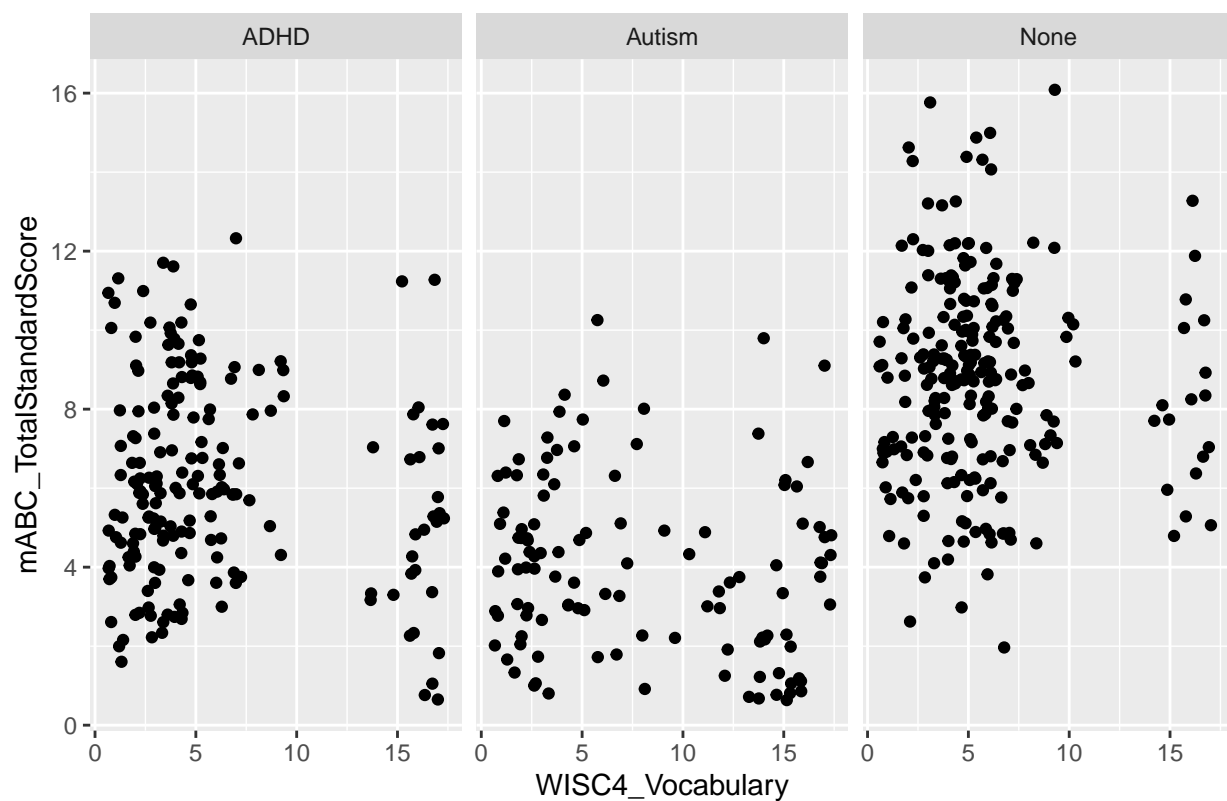
SRS\_TotalRawScore vs. mABC Total Standard Score



WISC4\_Vocabulary vs mABC\_TotalStandardScore

```
ggplot(WISC4_full_dat, aes(x= WISC4_Vocabulary, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_Vocabulary vs. mABC Total Standard Score')
```

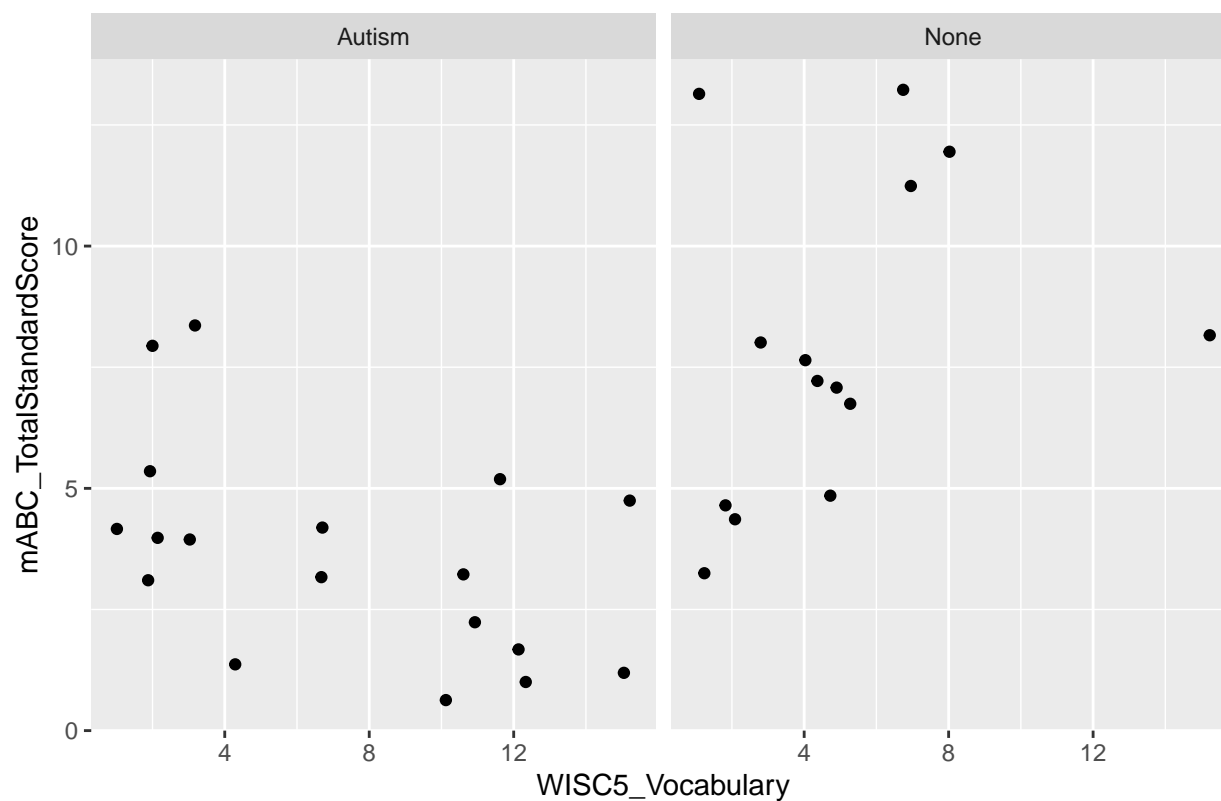
WISC4\_Vocabulary vs. mABC Total Standard Score



WISC5\_Vocabulary vs mABC\_TotalStandardScore

```
ggplot(WISC5_full_dat, aes(x= WISC5_Vocabulary, y=mABC_TotalStandardScore)) +  
  geom_jitter() +  
  facet_grid(. ~ PrimaryDiagnosis) +  
  ggtitle('WISC5_Vocabulary vs. mABC Total Standard Score')
```

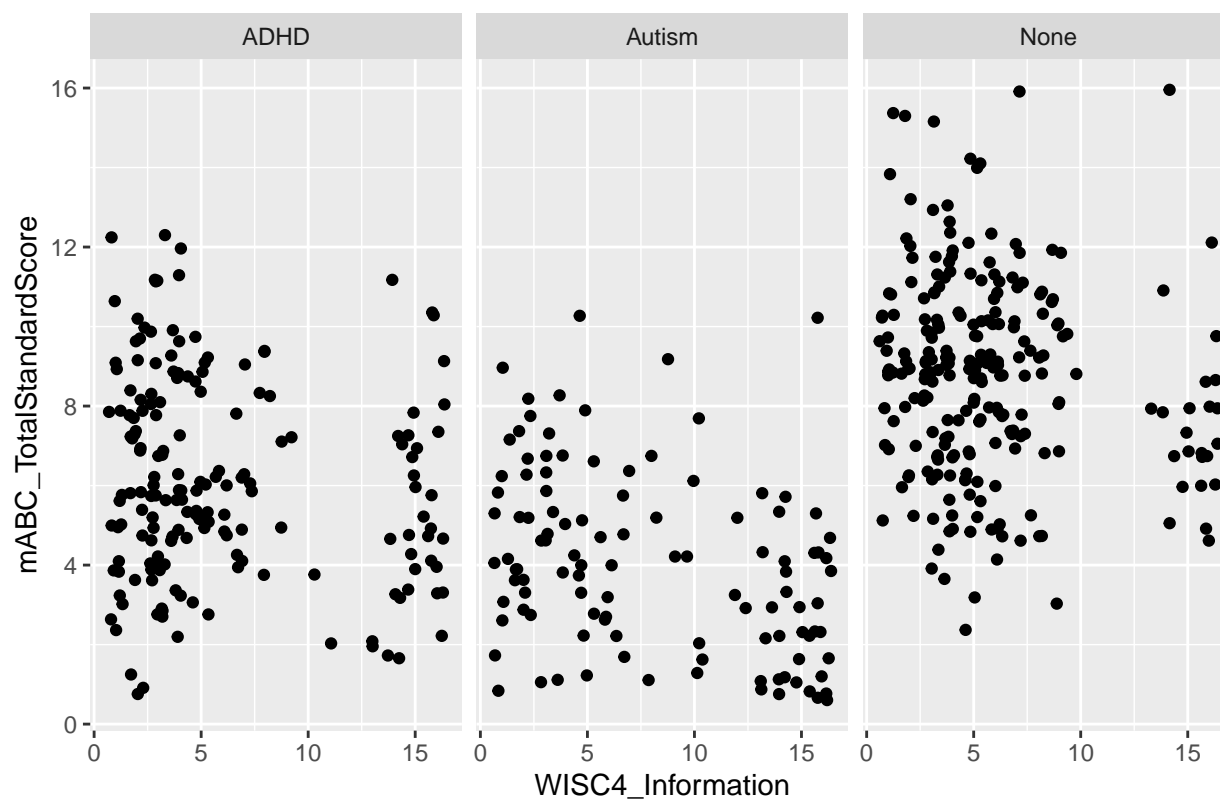
WISC5\_Vocabulary vs. mABC Total Standard Score



WISC4\_Information vs mABC\_TotalStandardScore

```
ggplot(WISC4_full_dat, aes(x= WISC4_Information, y=mABC_TotalStandardScore)) +
  geom_jitter() +
  facet_grid(. ~ PrimaryDiagnosis) +
  ggtitle('WISC4_Information vs. mABC Total Standard Score')
```

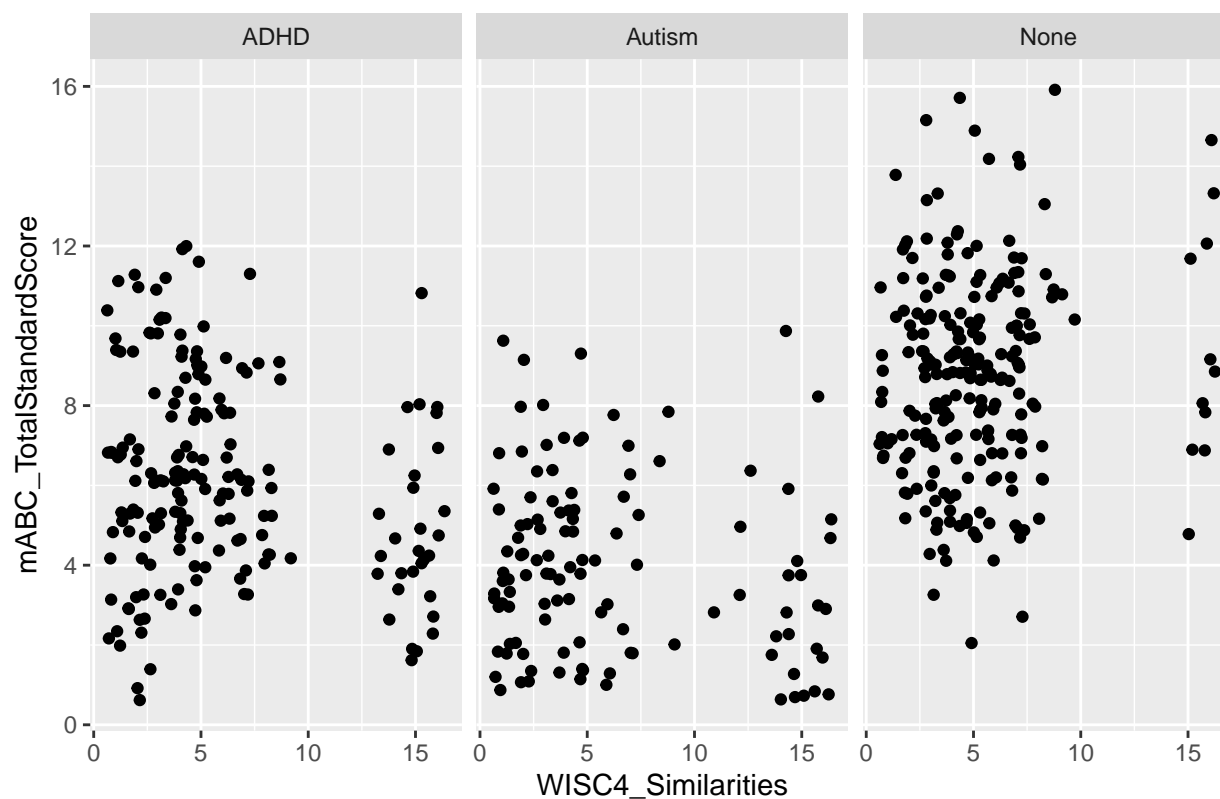
WISC4\_Information vs. mABC Total Standard Score



WISC4\_Similarities vs mABC\_TotalStandardScore

```
ggplot(WISC4_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

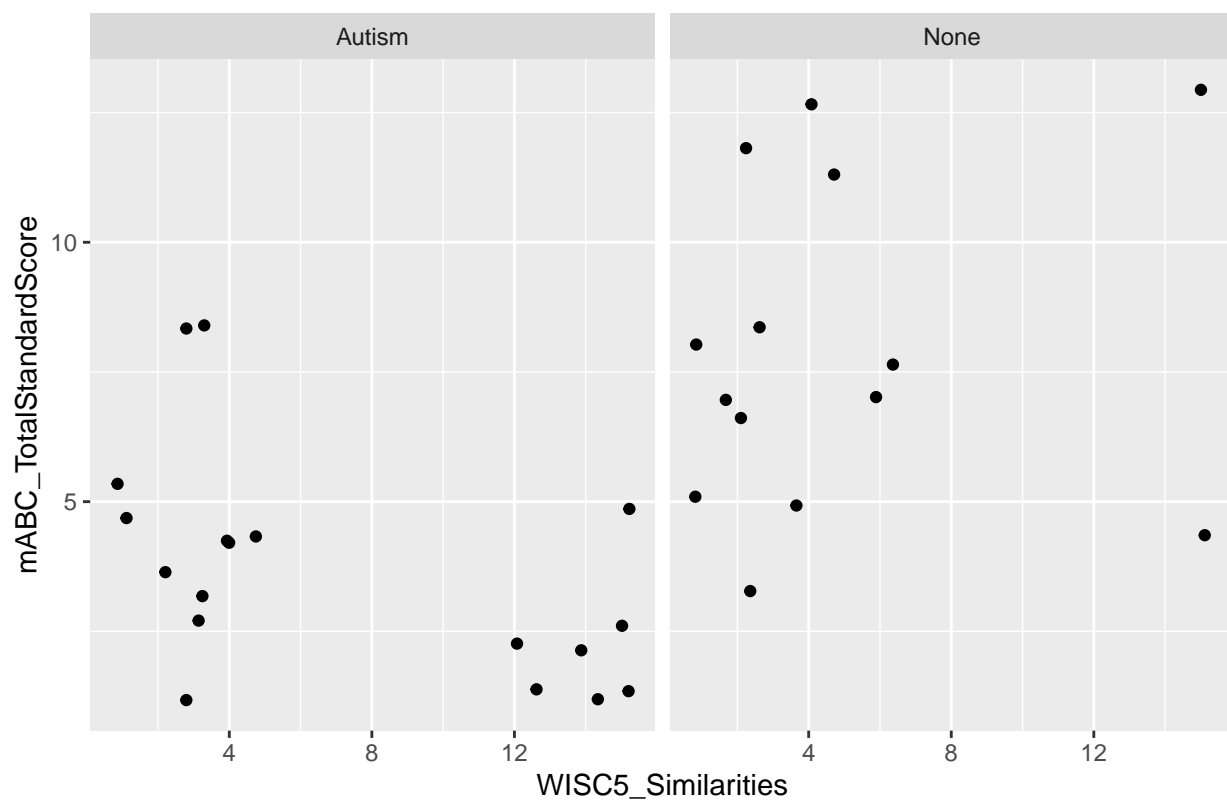


WISC5\_Similarities vs mABC\_TotalStandardScore

```
ggplot(WISC5_full_dat, aes(x= WISC5_Similarities, y=mABC_TotalStandardScore)) +  
  geom_jitter() +  
  facet_grid(. ~ PrimaryDiagnosis) +  
  ggtitle('WISC5_Similarities vs. mABC Total Standard Score')
```



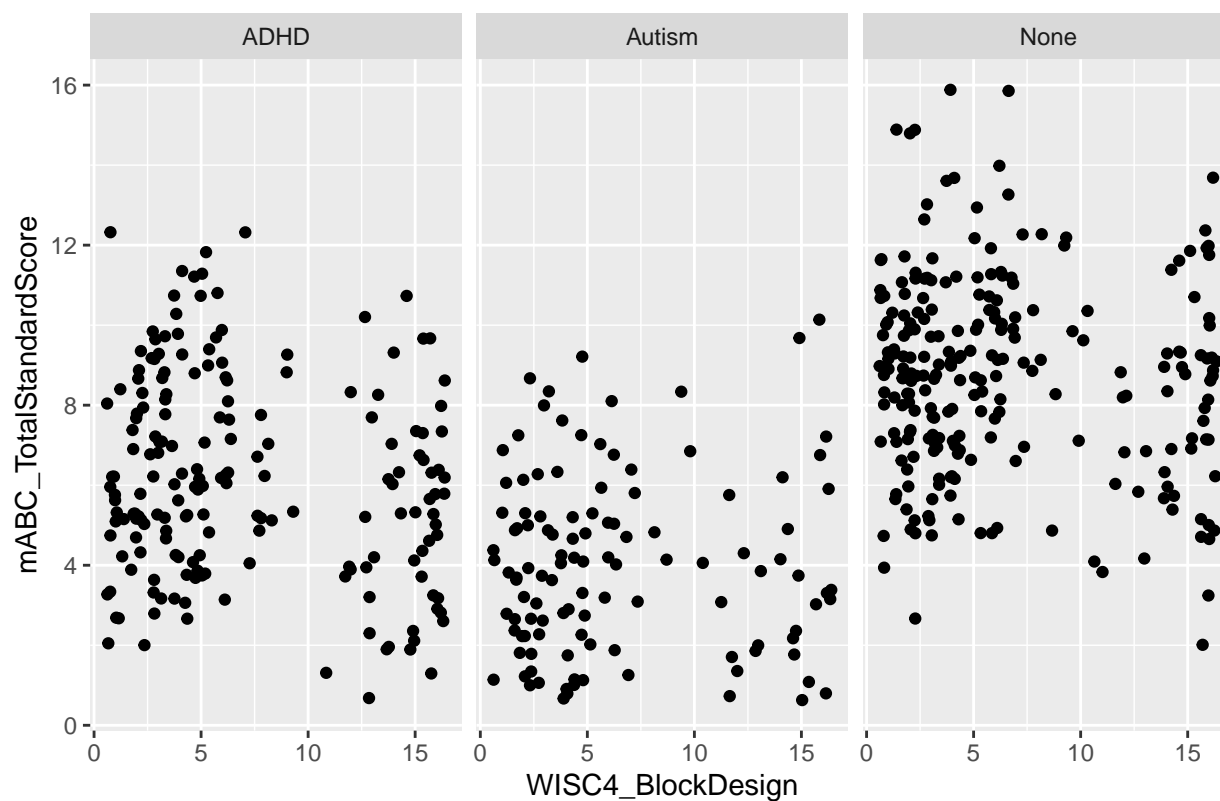
WISC5\_Similarities vs. mABC Total Standard Score



WISC4\_BlockDesign vs mABC\_TotalStandardScore

```
ggplot(WISC4_full_dat, aes(x= WISC4_BlockDesign, y=mABC_TotalStandardScore)) +  
  geom_jitter() +  
  facet_grid(. ~ PrimaryDiagnosis) +  
  ggtitle('WISC4_Design vs. mABC Total Standard Score')
```

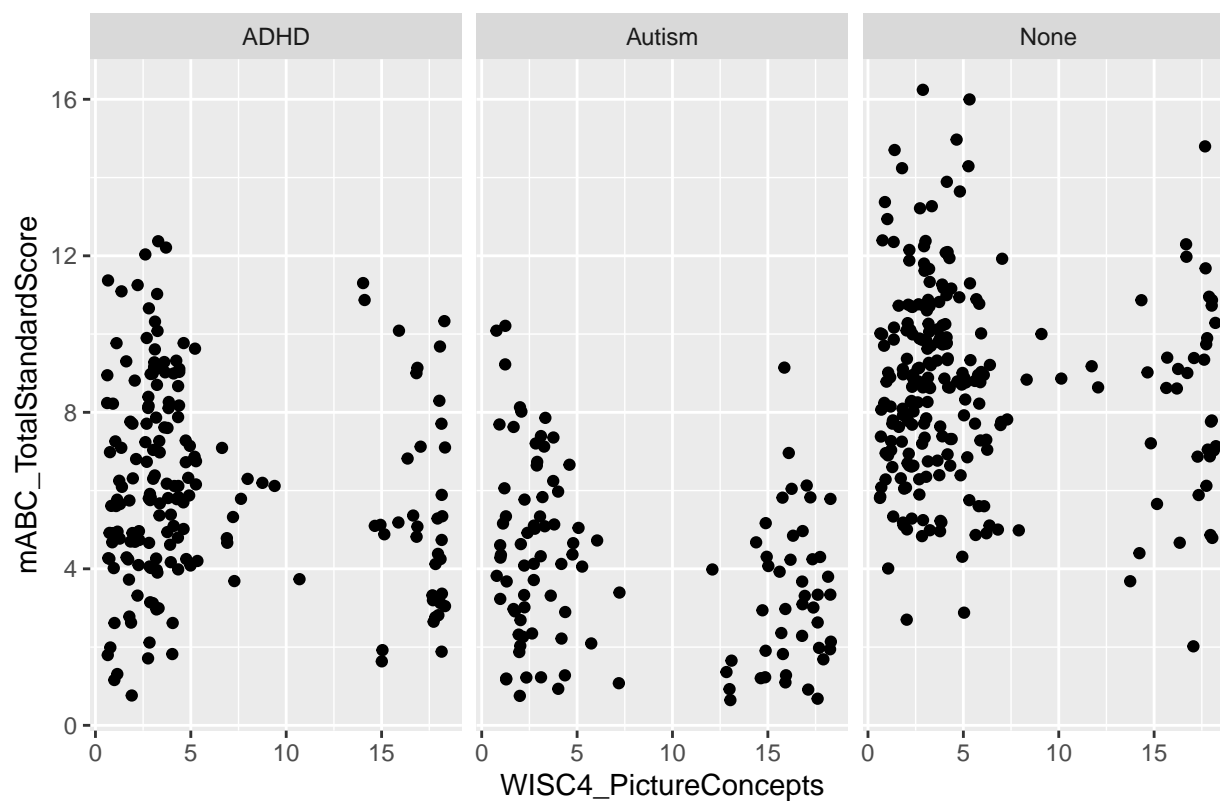
WISC4\_Design vs. mABC Total Standard Score



WISC4\_PictureConcepts vs mABC\_TotalStandardScore

```
ggplot(WISC4_full_dat, aes(x= WISC4_PictureConcepts, y=mABC_TotalStandardScore)) +  
  geom_jitter() +  
  facet_grid(. ~ PrimaryDiagnosis) +  
  ggtitle('WISC4_PictureConcepts vs. mABC Total Standard Score')
```

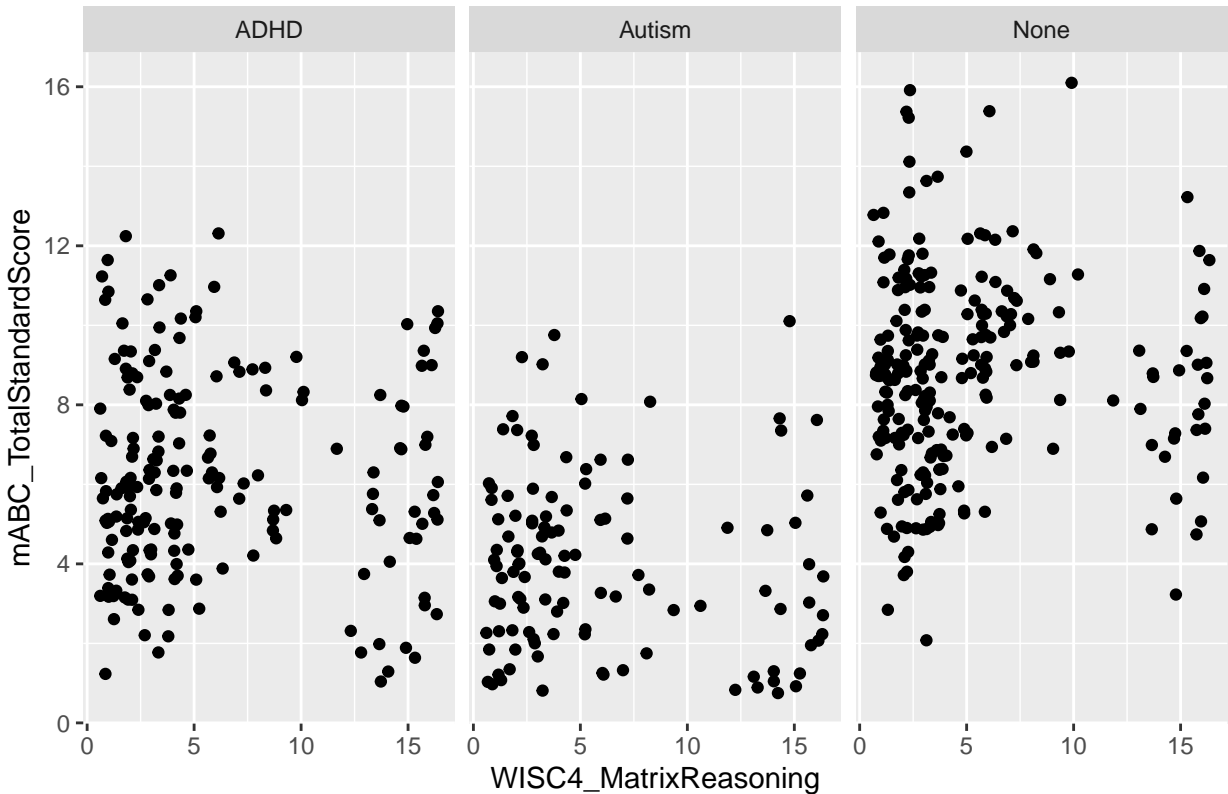
WISC4\_PictureConcepts vs. mABC Total Standard Score



WISC4\_MatrixReasoning vs mABC\_TotalStandardScore

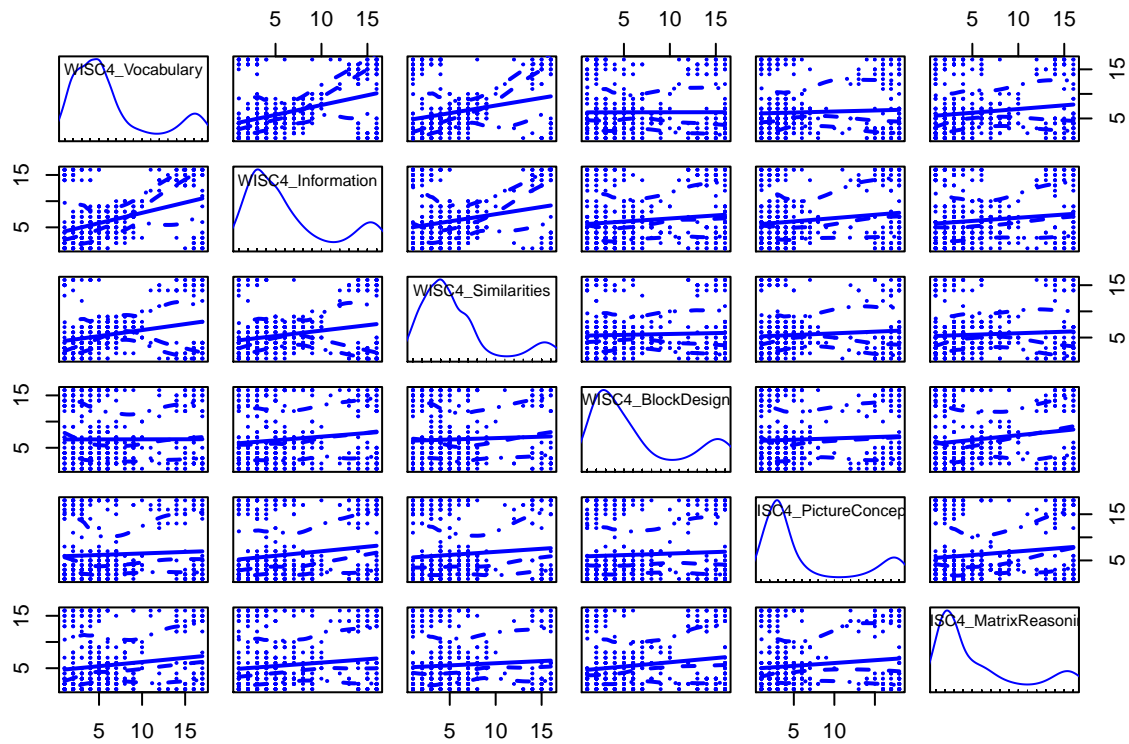
```
ggplot(WISC4_full_dat, aes(x= WISC4_MatrixReasoning, y=mABC_TotalStandardScore)) +  
  geom_jitter() +  
  facet_grid(. ~ PrimaryDiagnosis) +  
  ggtitle('WISC4_MatrixReasoning vs. mABC Total Standard Score')
```

## WISC4\_MatrixReasoning vs. mABC Total Standard Score



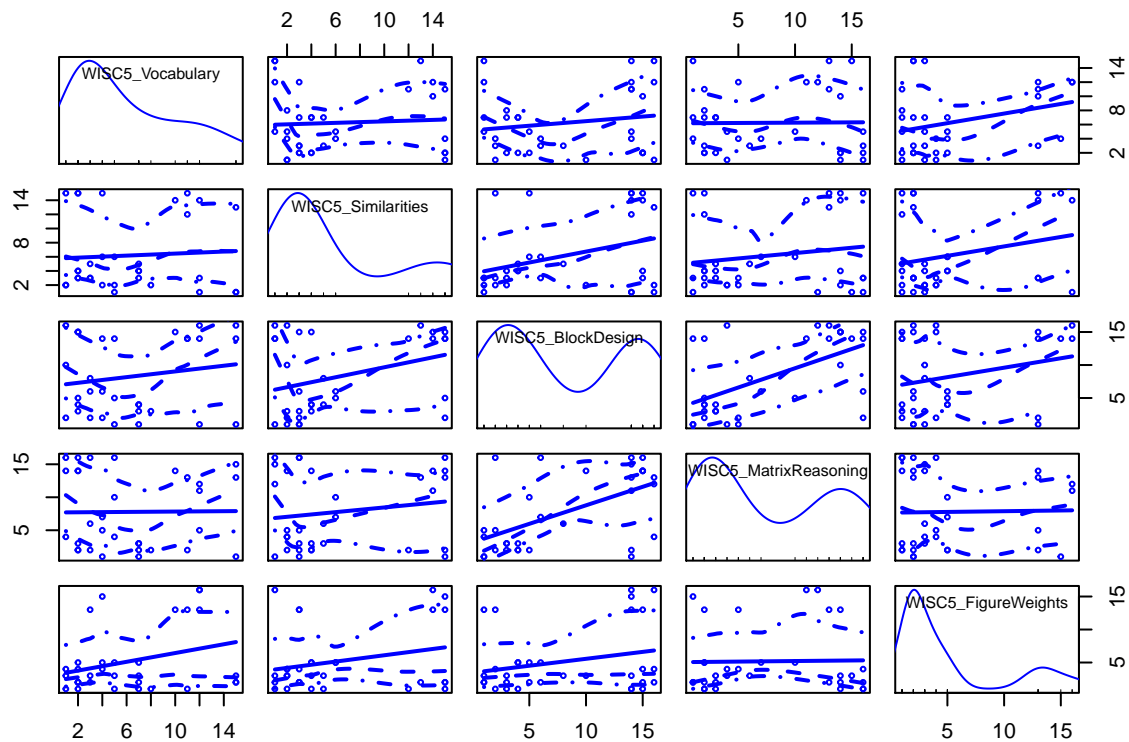
### Scatterplot Matrix for WISC\_4

```
scatterplotMatrix(WISC4_full_dat[c('WISC4_Vocabulary', 'WISC4_Information', 'WISC4_Similarities', 'WISC4_Scaled_Score')])
```



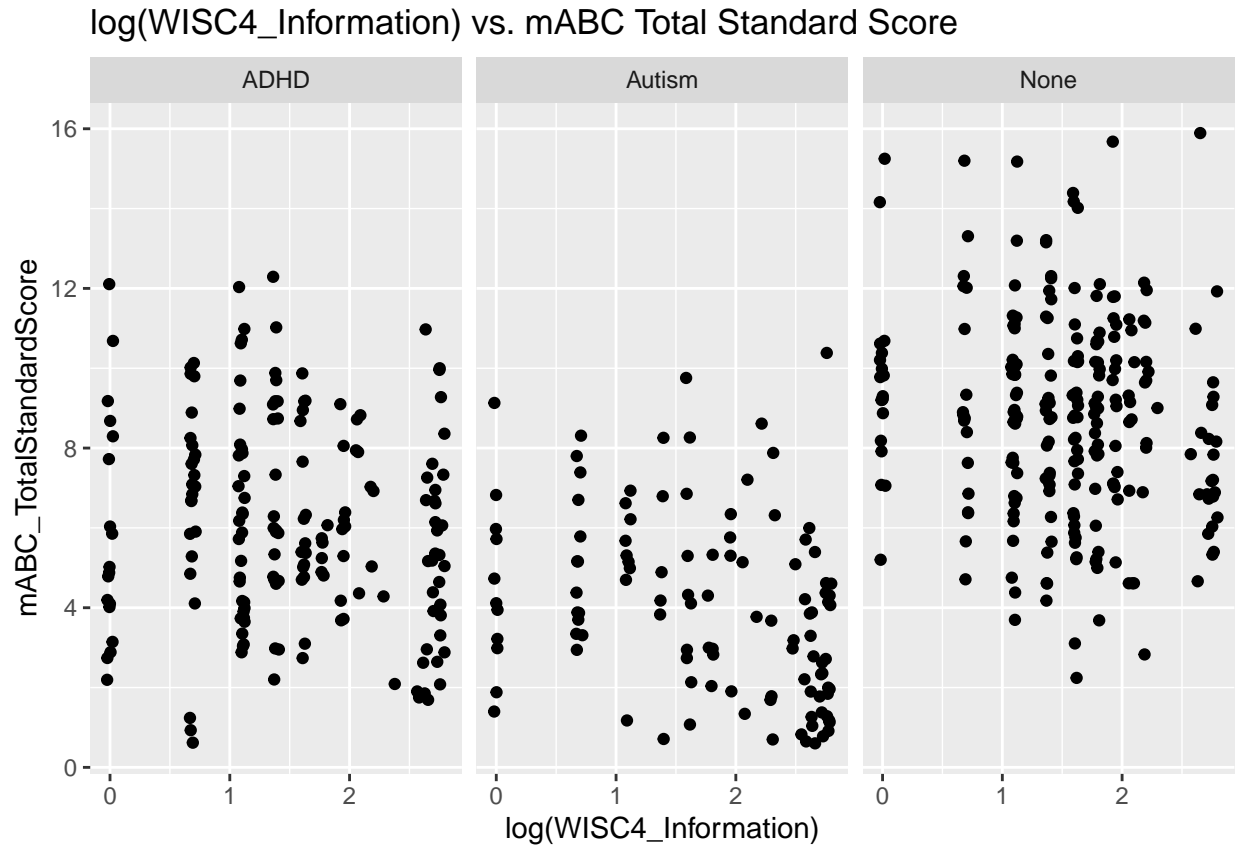
Scatterplot Matrix for WISC\_5

```
scatterplotMatrix(WISC5_full_dat[c('WISC5_Vocabulary', 'WISC5_Similarities', 'WISC5_BlockDesign', 'WISC5_MatrixReasoning', 'WISC5_FigureWeights')])
```



log(WISC4\_Information) vs. mABC Total Standard Score

```
ggplot(WISC4_full_dat, aes(x= log(WISC4_Information), y=mABC_TotalStandardScore)) + geom_jitter() +  
  facet_grid(. ~ PrimaryDiagnosis) +  
  ggtitle('log(WISC4_Information) vs. mABC Total Standard Score')
```



WISC4\_Total vs mABC\_TotalStandardScore

```
ggplot(WISC4_full_dat, aes(x= WISC4_Vocabulary + WISC4_Information + WISC4_Similarities + WISC4_BlockDe  
  geom_jitter() +  
  facet_grid(. ~ PrimaryDiagnosis) +  
  ggtitle('Sum of WISC4_Vars vs. mABC Total Standard Score'))
```

Sum of WISC4\_Vars vs. mABC Total Standard Score

