Lab 5: Interactions: Graphing & Regressions

Wednesday Bushong

3/8/2017

Overview

Today we'll learn about testing all kinds of interactions and plotting them.

The Data

```
library(rgl)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
d <- read.csv("InteractionData.csv")

head(d)
```

```
##   Self_control Neurot Consci Gender     Grade
## 1     3116.576   19.2     14  males 4th grade
## 2     7890.500   16.5     25  males 4th grade
## 3     4363.364   15.1     17  males 4th grade
## 4     5992.400   21.0     20  males 4th grade
## 5     4676.384   18.8     16  males 4th grade
## 6     6242.744   19.4     18  males 4th grade
```

```
summary(d)
```

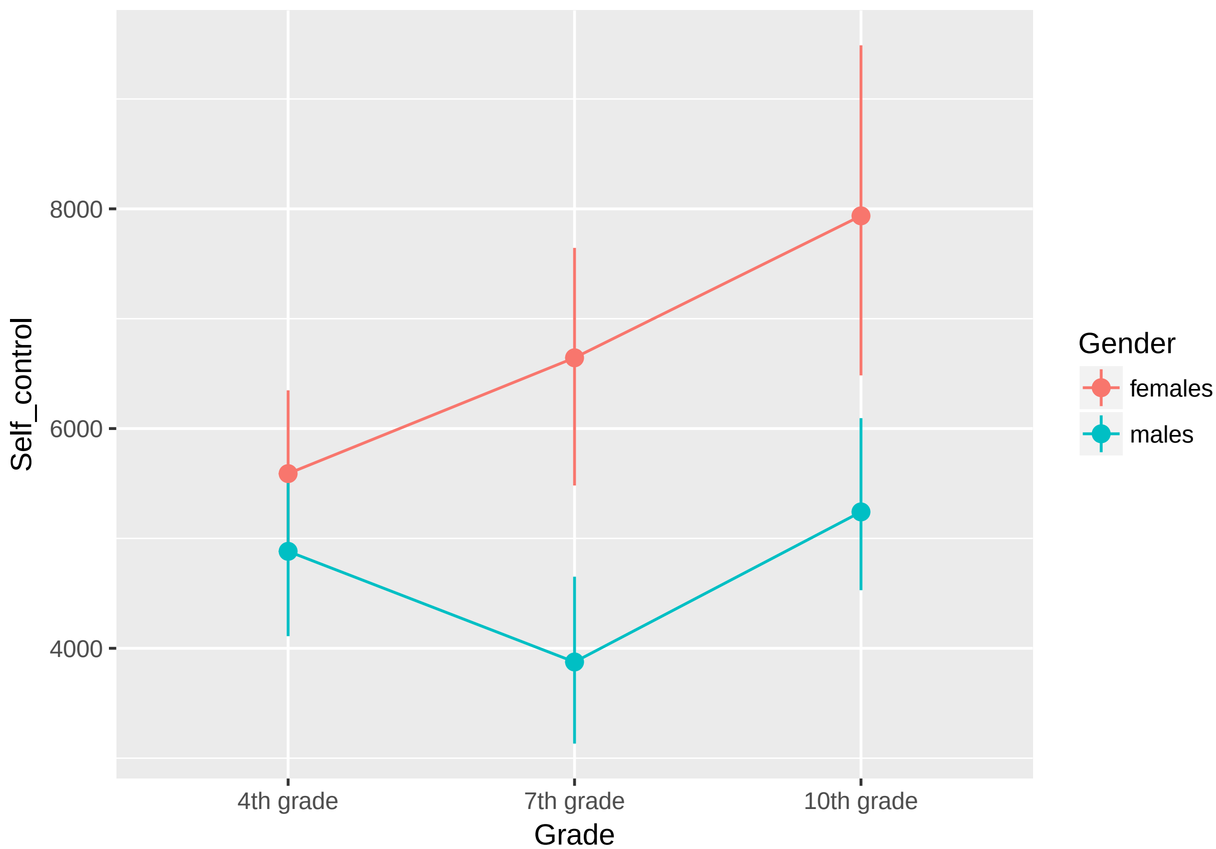
```
##   Self_control       Neurot          Consci          Gender
##  Min.   : 1306   Min.   : 9.40   Min.   : 8.00   females:50
##  1st Qu.: 3955   1st Qu.:15.35   1st Qu.:16.00   males  :50
##  Median : 5476   Median :17.75   Median :18.00
##  Mean   : 5690   Mean   :17.59   Mean   :18.12
##  3rd Qu.: 6941   3rd Qu.:20.25   3rd Qu.:21.00
##  Max.   :13450   Max.   :23.50   Max.   :25.00
##         Grade
##  10th grade:33
##  4th grade :34
##  7th grade :33
##
##
##
```

Categorical x Categorical Interactions

Plotting

```r
# relevel factors
d$Grade <- factor(d$Grade, levels = c("4th grade", "7th grade", "10th grade"))

p.gender.grade <- ggplot(d, aes(x = Grade, y = Self_control, color = Gender, group = Gender)) +
  stat_summary(fun.data = "mean_cl_boot", geom = "pointrange") +
  stat_summary(fun.y = "mean", geom = "line")
p.gender.grade
```



Regressions

Let's test the hypothesis that females' self-control incrases faster than males' across time in school. First, let's contrast code Gender. **Why do we want to do this?**

```r
contrasts(d$Gender) <- cbind(females = c(1, -1))
```

Now, how should we code Grade? Because this is a categorical variable with 3 levels, we'll have 2 contrasts for it. So choosing our coding scheme depends on the kind of question we want to ask. Since we want to know whether fmeles' self-control increases faster than males' across grade, one reasonable scheme is dummy coding with 7th grade as the reference level. That way, we get the Gender:Grade difference for 4th vs. 7th and 7th vs. 10th, which spans the space in a temporally sensible way.

```r
contrasts(d$Grade) <- cbind(grade_4 = c(1, 0, 0), grade_10 = c(0, 0, 1))
```

Testing for interactions

Now let's test for the interaction!

```r
m.gender.grade <- lm(Self_control ~ Gender * Grade, d)
summary(m.gender.grade)
```

```
##
## Call:
## lm(formula = Self_control ~ Gender * Grade, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4967.0 -1379.5   -83.5  1441.8  5514.2
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   5259.76     369.35  14.241  < 2e-16 ***
## Genderfemales                 1384.34     369.35   3.748 0.000308 ***
## Gradegrade_4                   -23.57     518.37  -0.045 0.963835
## Gradegrade_10                 1328.98     522.34   2.544 0.012580 *
## Genderfemales:Gradegrade_4   -1030.60     518.37  -1.988 0.049702 *
## Genderfemales:Gradegrade_10    -37.00     522.34  -0.071 0.943686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2121 on 94 degrees of freedom
## Multiple R-squared:  0.2877, Adjusted R-squared:  0.2498
## F-statistic: 7.594 on 5 and 94 DF,  p-value: 4.986e-06
```

Given our coding scheme,

- The positive main effect of gender means that females have greater self-control than males at the reference level of grade – that is, 7th grade.
- There is no main effect of our grade_4 contrast, meaning that there is not an overall difference in self-control between 4th and 7th grade.
- There is a positive main effect of the grade_10 contrast, meaning that there is an overall increase in self-control between 7th and 10th grade.
- There is a negative interaction between Gender and grade_4, meaning that the difference in self-control between females and males is smaller in 4th grade than in 7th grade – that is, that females increase in self-control more than males between 4th and 7th grade.
- There is no interaction between Gender and grade_10, meaning that the difference in self-control between females and males is not different between 7th and 10th grade.

Overall, we have some support for our hypothesis: females' self-control does increase faster than males', but only between 4th and 7th grade. Between 7th and 10th grade, females and males increase in self-control at about the same rate.

Testing simple slopes/simple effects

Given the interaction between Gender and grade_4, we can ask several simple effects questions. One that we may be interested in is whether males' self-control actually *decreases* between 4th and 7th grade (middle school is tough!).

We can do this by dummy-coding Gender so that males are the reference level. Then, the "main effect" of grade_4 will be the difference between 4th and 7th grade specifically for males.

```
d$Gender.dummy <- d$Gender
contrasts(d$Gender.dummy) <- cbind(females = c(1, 0))

m.gender.ss <- lm(Self_control ~ Gender.dummy * Grade, d)
summary(m.gender.ss)
```

```
##
## Call:
## lm(formula = Self_control ~ Gender.dummy * Grade, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4967.0 -1379.5   -83.5  1441.8  5514.2
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3875.42     514.36   7.534 2.99e-11 ***
## Gender.dummyfemales               2768.68     738.70   3.748 0.000308 ***
## Gradegrade_4                      1007.03     727.42   1.384 0.169517
## Gradegrade_10                     1365.97     738.70   1.849 0.067577 .
## Gender.dummyfemales:Gradegrade_4 -2061.20    1036.73  -1.988 0.049702 *
## Gender.dummyfemales:Gradegrade_10   -73.99    1044.68  -0.071 0.943686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2121 on 94 degrees of freedom
## Multiple R-squared:  0.2877, Adjusted R-squared:  0.2498
## F-statistic: 7.594 on 5 and 94 DF,  p-value: 4.986e-06
```

Our hypothesis is not supported – although males' self-control is numerically higher in 4th grade than 7th grade, this difference is not significant.

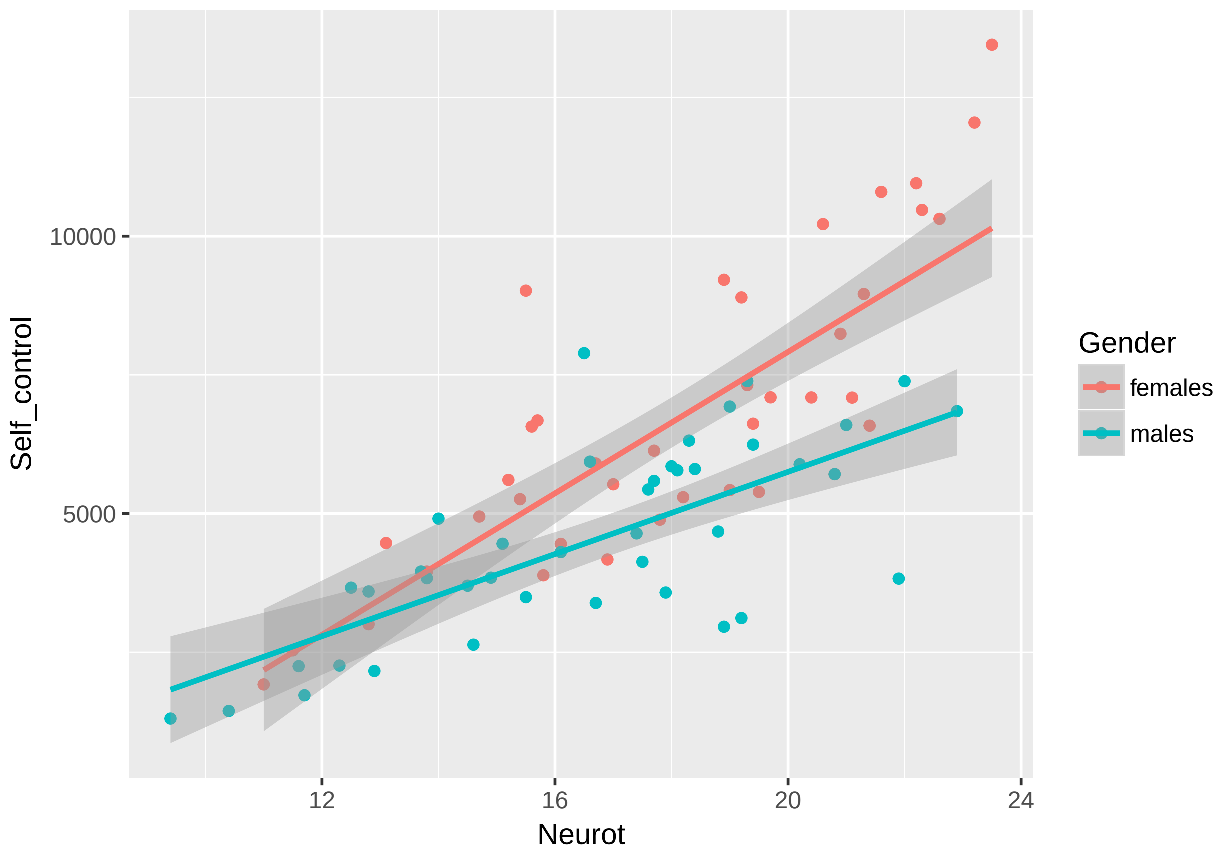
Categorical x Continuous Interactions

Now, let's consider the case of categorical x continuous interactions. Let's look at Gender and Neuroticism.

Plotting

First, let's plot! Here, I'll be making a scatterplot and dividing up the gender difference by color. We can also draw a best-fit regression line just for visualization purposes. (The function `geom_smooth` fits a separate linear model for each gender-split subset of the data, so we shouldn't take these lines as absolute truth).

```r
p.neuro.gender <- ggplot(d, aes(x = Neurot, y = Self_control, color = Gender)) +
  stat_summary(fun.y = "mean", geom = "point") +
  geom_smooth(method = "lm") # fits an lm to each subset of the data
p.neuro.gender
```



Regressions

Let's test the hypothesis that neuroticism affects self-control more for females than males.

Note that neuroticism is NOT centered, so we will need to do that here. **Why?**

```r
d$Neurot.centered <- d$Neurot - mean(d$Neurot)
```

Testing for interactions

```r
m.neurot.gender <- lm(Self_control ~ Neurot.centered * Gender, d)
summary(m.neurot.gender)
```

```
##
## Call:
## lm(formula = Self_control ~ Neurot.centered * Gender, data = d)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3290.1  -921.9  -131.4   949.3  3971.2
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   5616.82     148.31  37.872  < 2e-16 ***
## Neurot.centered                503.78      45.39  11.099  < 2e-16 ***
## Genderfemales                  758.95     148.31   5.117 1.59e-06 ***
## Neurot.centered:Genderfemales   133.43      45.39   2.940  0.00411 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1462 on 96 degrees of freedom
## Multiple R-squared:  0.6542, Adjusted R-squared:  0.6434
## F-statistic: 60.55 on 3 and 96 DF,  p-value: < 2.2e-16
```

Given our coding scheme (remember that Gender is contrast-coded),

- The positive main effect of neuroticism means that as neuroticism increases, self-control increases.
- The positive main effect of gender means that overall, females have higher self-control than males.
- The positive interaction means that the slope for neuroticism is steeper from females than males.

Testing simple slopes

Suppose we want to know the effect of gender on self-control for those who are highly neurotic – here, let's define that as those who have neuroticism scores 2 standard deviations above the mean. **How would we do this?**

```
d$Neurot.centered.2sds <- d$Neurot - (mean(d$Neurot) + 2*sd(d$Neurot))

m.highneurot.gender <- lm(Self_control ~ Neurot.centered.2sds * Gender, d)
summary(m.highneurot.gender)
```

```
##
## Call:
## lm(formula = Self_control ~ Neurot.centered.2sds * Gender, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3290.1  -921.9  -131.4   949.3  3971.2
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        8925.80     332.87  26.814  < 2e-16
## Neurot.centered.2sds                503.78      45.39  11.099  < 2e-16
## Genderfemales                      1635.36     332.87   4.913 3.69e-06
## Neurot.centered.2sds:Genderfemales   133.43      45.39   2.940  0.00411
##
## (Intercept)                        ***
## Neurot.centered.2sds               ***
```

```
## Genderfemales                       ***
## Neurot.centered.2sds:Genderfemales **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1462 on 96 degrees of freedom
## Multiple R-squared:  0.6542, Adjusted R-squared:  0.6434
## F-statistic: 60.55 on 3 and 96 DF,  p-value: < 2.2e-16
```

Continuous x Continuous Interactions

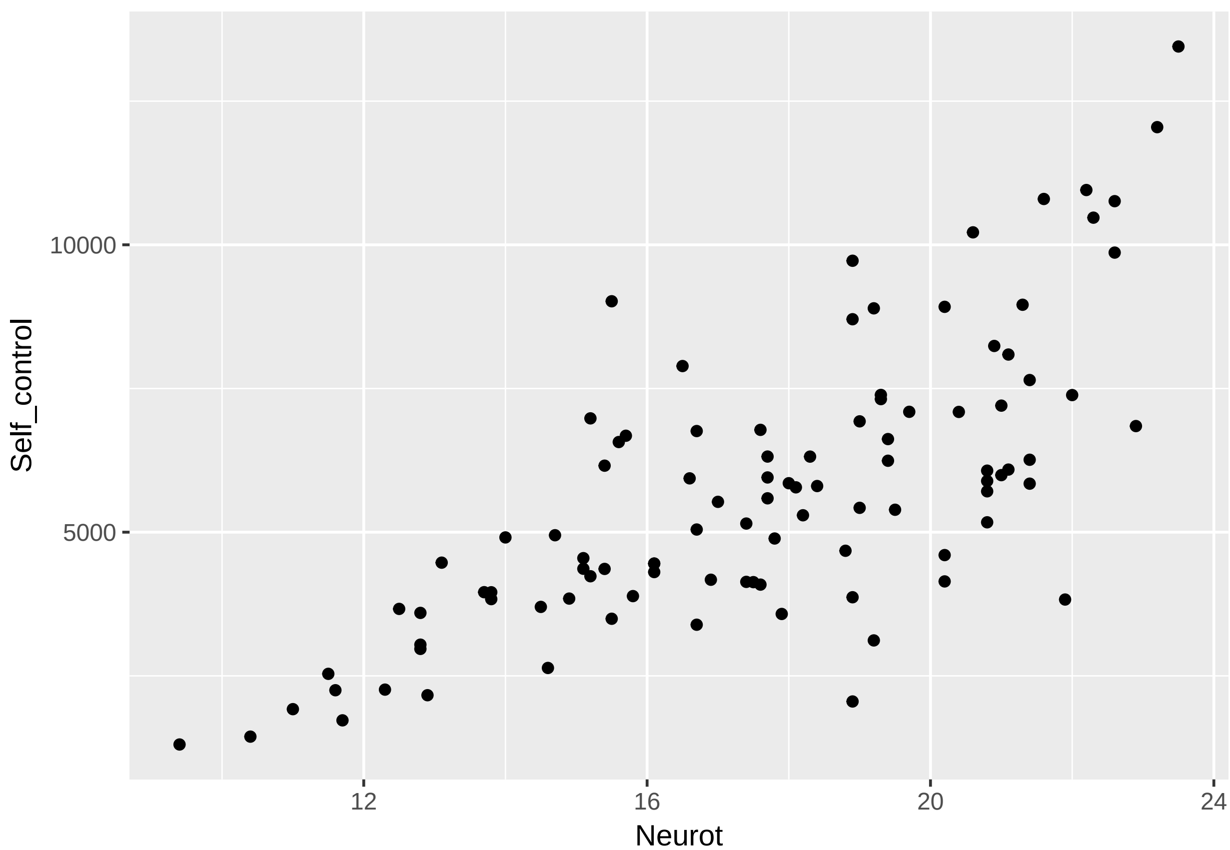
Now, let's look at the case of two continuous IVs. We'll use Neuroticism and Conscientiousness.

Plotting

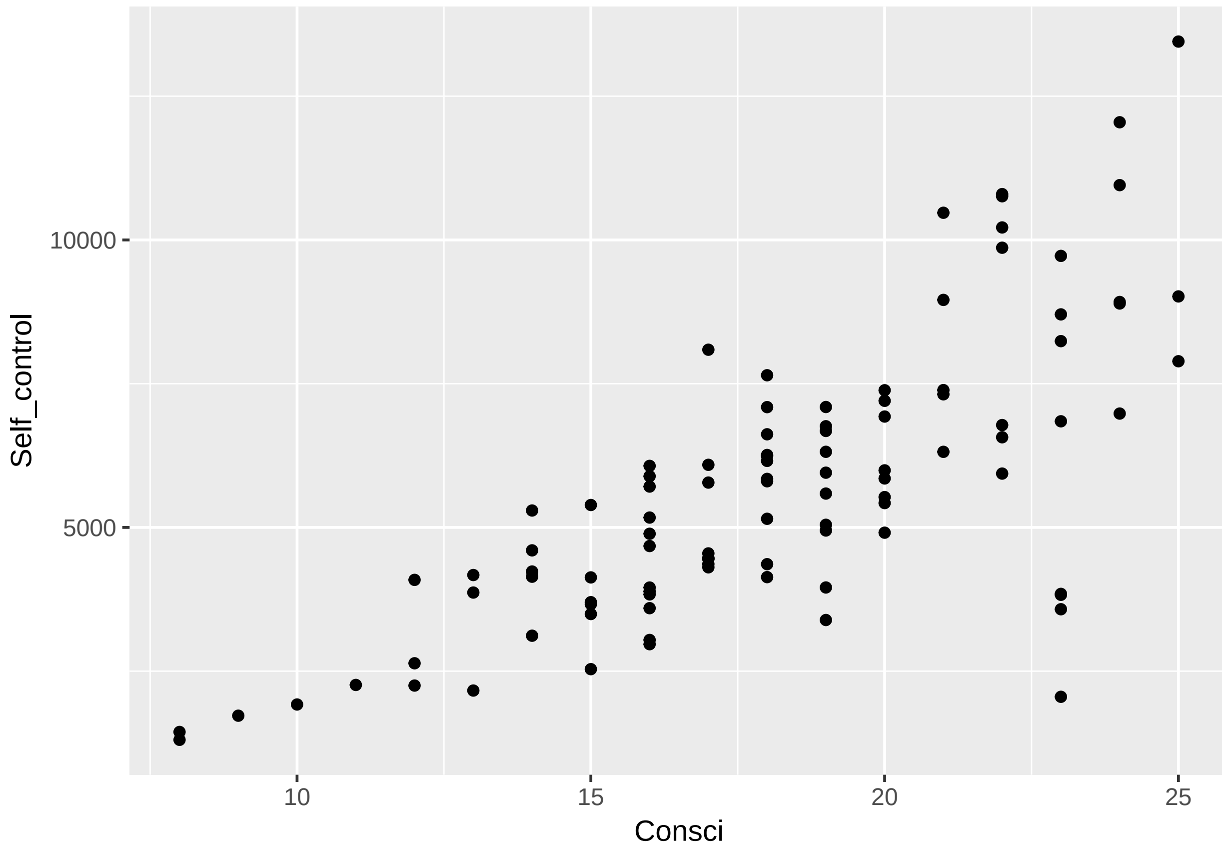
Visualizing continuous x continuous interactions is quite difficult! I'll show you a couple of strategies for visualizing the raw data. Then after we fit regression models we can make plots of the predictions as well.

In general, though, it's useful to plot the effect of both of the variables separately just to have a general idea of what the effects are like:

```
p.neurot <- ggplot(d, aes(x = Neurot, y = Self_control)) +
  geom_point()
p.neurot
```



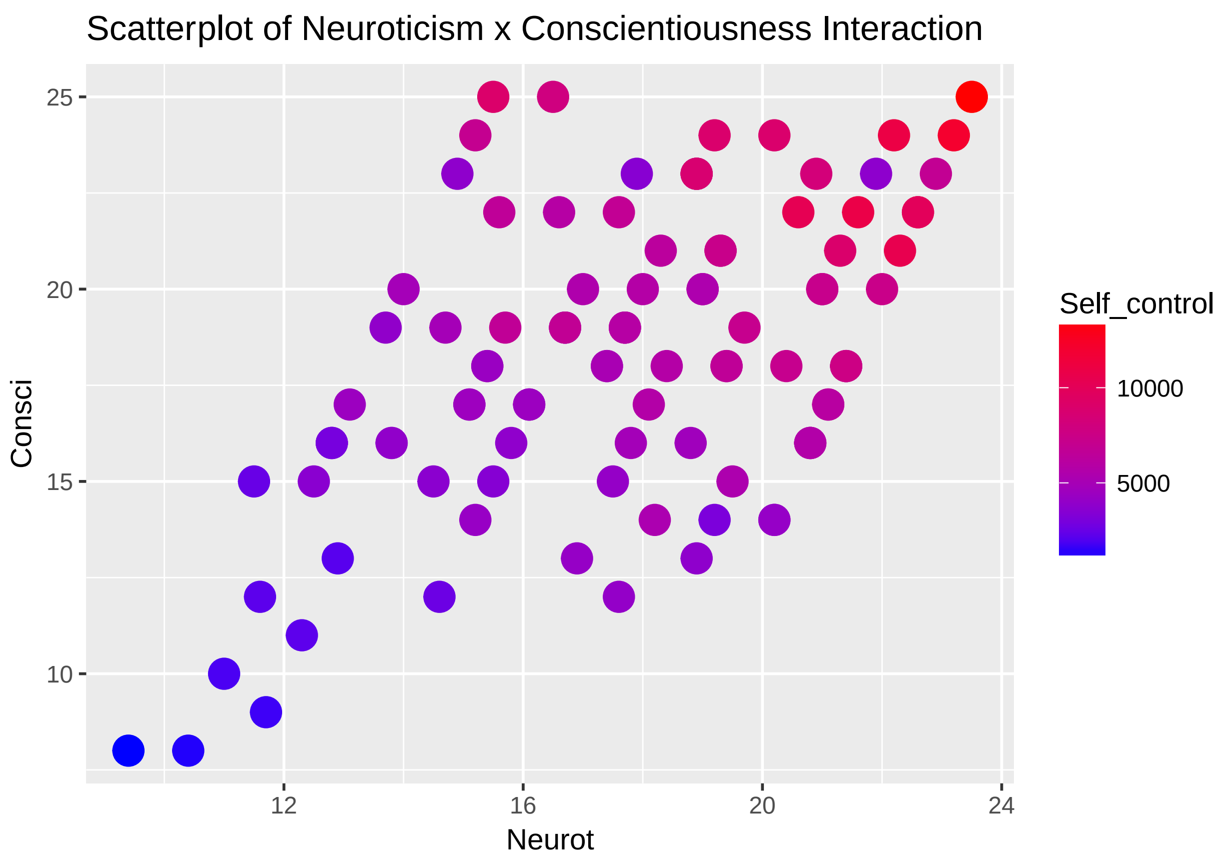
```r
p.consci <- ggplot(d, aes(x = Consci, y = Self_control)) +
  geom_point()
p.consci
```



It looks like each variable on its own has a positive effect. Now let's try to see whether there is an interaction!

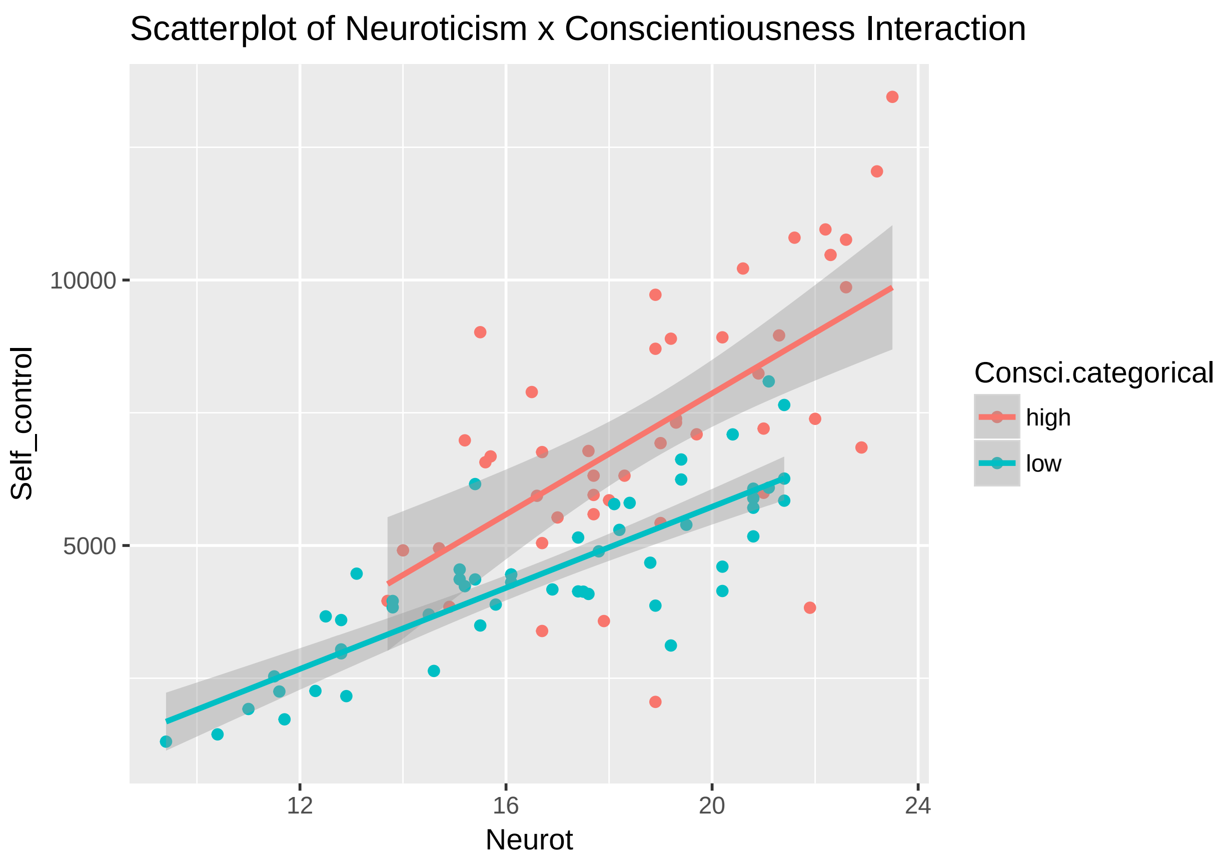
Option 1: Plot DV with color

```r
p.neurot.consci <- ggplot(d, aes(x = Neurot, y = Consci, color = Self_control)) +
  geom_point(size = 5) +
  scale_color_gradient(low = "blue", high = "red") +
  ggtitle("Scatterplot of Neuroticism x Conscientiousness Interaction")
p.neurot.consci
```



Option 2: Median Split

```
d$Consci.categorical <- ifelse(d$Consci > median(d$Consci), "high", "low")
p.neurot.consci2 <- ggplot(d, aes(x = Neurot, y = Self_control, color = Consci.categorical)) +
  geom_point() +
  ggtitle("Scatterplot of Neuroticism x Conscientiousness Interaction") +
  geom_smooth(method = "lm")
p.neurot.consci2
```



Option 3: 3D plot

```
p.3d <- plot3d(d[, c("Consci", "Neurot", "Self_control")])
```

Even a 3d plot is a bit hard to read. It's pretty unclear whether there's an interaction here. Let's test it with a regression!

Regressions

Again, we'll need to center the other continuous variable, `Consci`:

```
d$Consci.centered <- d$Consci - mean(d$Consci)
```

Testing for interactions

```
m.consci.neurot <- lm(Self_control ~ Consci.centered * Neurot.centered, d)
summary(m.consci.neurot)
```

```
##
## Call:
```

```
## lm(formula = Self_control ~ Consci.centered * Neurot.centered,
##     data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5756.2  -339.7    42.0   625.3  2399.7
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     5485.128    141.534  38.755  < 2e-16 ***
## Consci.centered                  338.321     40.707   8.311 6.21e-13 ***
## Neurot.centered                  368.216     47.622   7.732 1.04e-11 ***
## Consci.centered:Neurot.centered   29.727      8.485   3.504 0.000699 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1289 on 96 degrees of freedom
## Multiple R-squared:  0.7313, Adjusted R-squared:  0.7229
## F-statistic: 87.09 on 3 and 96 DF,  p-value: < 2.2e-16
```

There is an interaction after all – for higher values of conscientiousness, the effect of neuroticism on self-control is even higher.

Testing simple slopes

Let's say that we want to know what the effect of neuroticism is for people who aren't very conscientious.

```
d$Consci.1sd <- d$Consci - (mean(d$Consci) - sd(d$Consci))
m.neurot.lowconsci <- lm(Self_control ~ Neurot.centered * Consci.1sd, d)
summary(m.neurot.lowconsci)
```

```
##
## Call:
## lm(formula = Self_control ~ Neurot.centered * Consci.1sd, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5756.2  -339.7    42.0   625.3  2399.7
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                4171.483    221.040  18.872  < 2e-16 ***
## Neurot.centered             252.790     53.653   4.712 8.30e-06 ***
## Consci.1sd                  338.321     40.707   8.311 6.21e-13 ***
## Neurot.centered:Consci.1sd   29.727      8.485   3.504 0.000699 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1289 on 96 degrees of freedom
## Multiple R-squared:  0.7313, Adjusted R-squared:  0.7229
## F-statistic: 87.09 on 3 and 96 DF,  p-value: < 2.2e-16
```

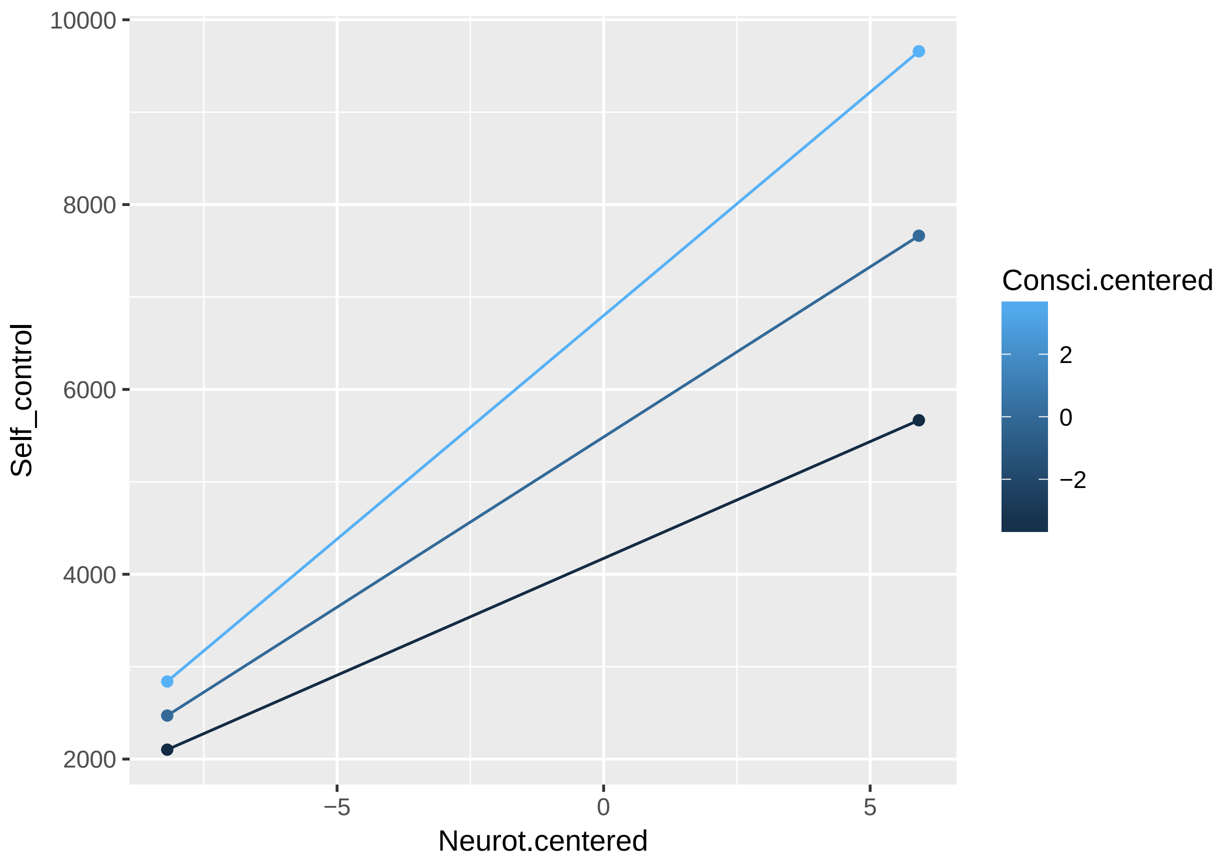
Neuroticism still has a main effect, so even those who are not very conscientious still have an effect of neuroticism on their self-control.

Plotting regression predictions

Now that we have a model of the continuous interaction, we can predict the slopes for a range of values! We can predict values of self-control for any given values of neuroticism and conscientiousness using the `predict()` function, which takes a model object and predicts values given new data. The general strategy we'll take is to plot the effect of neuroticism on self-control at various levels of conscientiousness. We'll do this by predicting self-control values for the range of neuroticism values, at 3 levels of conscientiousness.

```
# Get mean of consci, and +/-1 SD
consci.mean <- mean(d$Consci.centered)
consci.1.below <- consci.mean - sd(d$Consci.centered)
consci.1.above <- consci.mean + sd(d$Consci.centered)
# Create fake data based on this
fake.data <- data.frame(Neurot.centered = rep(range(d$Neurot.centered), 3),
                        Consci.centered = c(rep(consci.mean, 2), rep(consci.1.below, 2), rep(consci.1.al
# Predict values of self-control
fake.data$Self_control <- predict(m.consci.neurot, fake.data)

#  Plot!
p.predictions <- ggplot(fake.data, aes(x = Neurot.centered, y = Self_control, color = Consci.centered, g
  geom_point() +
  geom_line()
p.predictions
```



3D plotting regression planes

We can also draw a regression plane in 3-D space to visualize the interaction. The details are a bit too much to go into right now, but basically we will create a function to draw a surface from given slope values from our models. (Come see me if you want to understand how this works in more detail!)

```
# Plot regression plane with NO interaction
m.no.interaction <- lm(Self_control ~ Neurot.centered + Consci.centered, d)
f1 <- function(X, Z) {
  r = coef(m.no.interaction)[1] +
    coef(m.no.interaction)[2] * Z +
    coef(m.no.interaction)[3] * X
}
plot3d(d[c("Consci.centered", "Neurot.centered", "Self_control")])
plot3d(f1,
       xlim = range(d$Consci.centered),
       ylim = range(d$Neurot.centered),
       add = TRUE, col = "red", alpha = .5)


# Plot regression plane WITH interaction
m.interaction <- lm(Self_control ~ Neurot.centered * Consci.centered, d)
f2 <- function(X, Z) {
  r = coef(m.interaction)[1] +
    coef(m.interaction)[2] * Z +
    coef(m.interaction)[3] * X +
    coef(m.interaction)[4] * Z * X
}
plot3d(d[c("Consci.centered", "Neurot.centered", "Self_control")])
plot3d(f2,
       xlim = range(d$Consci.centered),
       ylim = range(d$Neurot.centered),
       add = TRUE, col = "blue", alpha = .5)

# Plot both!
plot3d(d[c("Consci.centered", "Neurot.centered", "Self_control")])
plot3d(f1,
       xlim = range(d$Consci.centered),
       ylim = range(d$Neurot.centered),
       add = TRUE, col = "red", alpha = .5)
plot3d(f2,
       xlim = range(d$Consci.centered),
       ylim = range(d$Neurot.centered),
       add = TRUE, col = "blue", alpha = .5)
```