Final Project

Leah Schultz

12/20/2017

```
library(dplyr)
library(lme4)
library(ggplot2)
library(merTools)
library(sjPlot)
library(broom)
library(tidyr)
library(lcmm)
library(plyr)
oysup <- read.csv("~/1-descriptives-and-graphs-leahschultz/oysup_teacher_self.csv")
purpose <- read.csv("~/Dropbox/Lab & Research/OYSUP Project/oysup_self.csv")

oysup <- oysup %>%
  dplyr::select(FAMID, extra_7s:extra_10s, agree_7s:agree_10s,
                consc_7s:consc_10s, neuro_7s:neuro_10s, open_7s:open_10s,extra_7t:extra_10t,
                agree_7t:agree_10t, consc_7t:consc_10t, neuro_7t:neuro_10t, open_7t:open_10t)
dems <- purpose %>%
  dplyr::select(SEX2, MPEDUC2)
oysup <- cbind(oysup, dems)
```

First, restructuring data:

```
oysup_long <- tbl_df(oysup) %>%
  gather(c(-FAMID, -SEX2, -MPEDUC2), key = "grade", value = "value") %>%
  separate(grade, into = c("variable", "grade"), sep = "_", convert = T) %>%
  separate(grade, into = c("grade", "source"), sep = -2) %>%
  mutate(grade = as.numeric(grade)) %>%
  spread(variable, value)
oysup_long
```

```
## # A tibble: 8,592 x 10
##    FAMID  SEX2 MPEDUC2 grade source agree consc extra neuro  open
##  * <int> <int>   <int> <dbl>  <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1  1001     2       3     7      s    NA    NA    NA    NA    NA
##  2  1001     2       3     7      t    NA    NA    NA    NA    NA
##  3  1001     2       3     8      s    NA    NA    NA    NA    NA
##  4  1001     2       3     8      t    NA    NA    NA    NA    NA
##  5  1001     2       3     9      s   5.0   5.0     5   3.5   5.0
##  6  1001     2       3     9      t   2.5   2.5     3   3.5   3.5
##  7  1001     2       3    10      s   5.0   5.0     5   5.0   4.5
##  8  1001     2       3    10      t   4.5   4.0     5   1.0   3.5
##  9  1002     2       3     7      s   4.5   4.0     5   3.5   4.5
## 10  1002     2       3     7      t   4.0   4.5     5   1.5   5.0
## # ... with 8,582 more rows
```

```
oysup_long$grade2 <- oysup_long$grade^2
```

As a complement to the SEM growth models and TICs models that I'm already doing for my analyses of personality change during adolescence, I want to use latent class mixed models to understand any consistent trends that might be occurring in individuals' development, both from student- and teacher-reports. It is clear that not everyone is changing in the same ways during this time, but are there distinct clusters of students changing in specific ways?

```
oysup_long_e <- subset(oysup_long, subset = !is.na(extra))
oysup_long_e_s <- subset(oysup_long_e, subset = source=="s")

model_1 <- lcmm(fixed = extra ~ grade + grade2,
                   random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=1, idiag=FALSE, link="linear",data=oysup_long_e_s)
```

```
## Be patient, lcmm is running ...
## The program took 0.31 seconds
```

```
model_2 <- lcmm(fixed = extra ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=2, idiag=FALSE, link="linear",data=oysup_long_e_s)
```

```
## Be patient, lcmm is running ...
## The program took 1.17 seconds
```

```
model_3 <- lcmm(fixed = extra ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=3, idiag=FALSE, link="linear",data=oysup_long_e_s)
```

```
## Be patient, lcmm is running ...
## The program took 2.1 seconds
```

```
model_4 <- lcmm(fixed = extra ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=4, idiag=FALSE, link="linear",data=oysup_long_e_s)
```

```
## Be patient, lcmm is running ...
## The program took 6.18 seconds
```

```
model_5 <- lcmm(fixed = extra ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=5, idiag=FALSE, link="linear",data=oysup_long_e_s)
```

```
## Be patient, lcmm is running ...
## The program took 12.39 seconds
```

```
summarytable(model_1, model_2, model_3, model_4, model_5)
```

```
##         G    loglik npm      BIC    %class1  %class2 %class3  %class4
## model_1 1 -3560.439   7 7168.939 100.000000
## model_2 2 -3478.861  10 7026.381  15.119917 84.88008
## model_3 3 -3478.861  13 7046.978  16.996872 83.00313       0
## model_4 4 -3478.861  16 7067.576  18.978102 81.02190       0  0.00000
## model_5 5 -3442.851  19 7016.154   6.673618 32.95099       0 60.37539
##         %class5
## model_1
## model_2
## model_3
## model_4
## model_5       0
```

Looks like after 2 classes, a third class cannot be estimated. Two classes seem to fit the data best. Let's look at how many subjects are in each class (since it gives percentages but not #s).

```r
# How many subjects per class in 2-class model?
pp_model2 <- postprob(model_2)
```

```
##
## Posterior classification:
##   class1 class2
## N 145.00 814.00
## %  15.12  84.88
##
## Posterior classification table:
##      --> mean of posterior probabilities in each class
##         prob1  prob2
## class1 0.8691 0.1309
## class2 0.0372 0.9628
##
## Posterior probabilities above a threshold (%):
##          class1 class2
## prob>0.7  83.45  96.81
## prob>0.8  73.10  93.37
## prob>0.9  60.00  88.82
##
```

```r
pp_model2[1]
```

```
## [[1]]
##   class1 class2
## N 145.00 814.00
## %  15.12  84.88
```

```r
#First, we need a data frame that specifies which class subjects are in
membership <- as.data.frame(matrix(nrow=959,ncol=2))
colnames(membership) <- c("FAMID","Class_2")
membership[,1:2] <- model_2$pprob[,1:2]
oysup_long_e_s <- merge(x = oysup_long_e_s, y = membership, by="FAMID")

# How are the classes different?

# Extraversion in 7th grade
oysup_long_e_s7 <- subset(oysup_long_e_s, subset = grade=="7")
t.test(data = oysup_long_e_s7, extra ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  extra by Class_2
## t = -6.7374, df = 88.907, p-value = 1.547e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.0985833 -0.5981712
## sample estimates:
## mean in group 1 mean in group 2
##        3.506494        4.354871
```

```r
oysup_long_e_s8 <- subset(oysup_long_e_s, subset = grade=="8")
t.test(data = oysup_long_e_s8, extra ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  extra by Class_2
## t = -13.204, df = 126.22, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.559991 -1.153335
## sample estimates:
## mean in group 1 mean in group 2
##        3.040909        4.397572
```

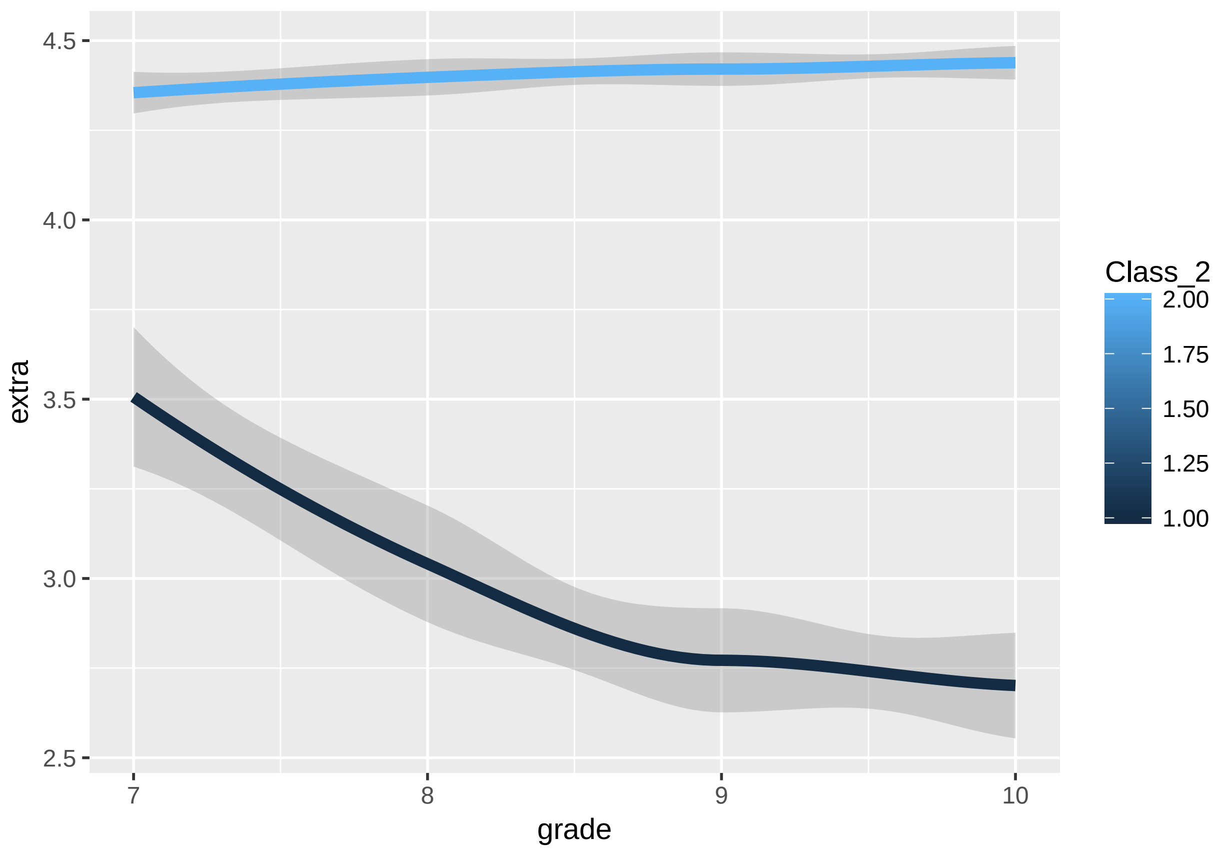
```r
oysup_long_e_s9 <- subset(oysup_long_e_s, subset = grade=="9")
t.test(data = oysup_long_e_s9, extra ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  extra by Class_2
## t = -24.043, df = 172.81, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.784162 -1.513445
## sample estimates:
## mean in group 1 mean in group 2
##        2.771739        4.420543
```

```r
oysup_long_e_s10 <- subset(oysup_long_e_s, subset = grade=="10")
t.test(data = oysup_long_e_s10, extra ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  extra by Class_2
## t = -28.217, df = 167.69, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.858423 -1.615375
## sample estimates:
## mean in group 1 mean in group 2
##        2.701493        4.438392
```

```r
oysup_long_e_s$FAMID <- as.character(oysup_long_e_s$FAMID)
plot_e_s <- ggplot(oysup_long_e_s, aes(x = grade, y = extra, group = FAMID, color = Class_2)) +
  geom_smooth(aes(group = FAMID, color = Class_2), size = 0.5) +
  geom_smooth(aes(group = Class_2), method = "loess", size = 2, se=T)
plot_e_s
```



Group 2 is stable, while Group 1 is decreasing over time.

Let's repeat with teacher-reported data.

```
oysup_long_e <- subset(oysup_long, subset = !is.na(extra))
oysup_long_e_t <- subset(oysup_long_e, subset = source=="t")

model_1 <- lcmm(fixed = extra ~ grade + grade2,
                random = ~grade, nwg = FALSE, subject = "FAMID",
                ng=1, idiag=FALSE, link="linear",data=oysup_long_e_t)
```

```
## Be patient, lcmm is running ...
## The program took 0.24 seconds
```

```
model_2 <- lcmm(fixed = extra ~ grade + grade2,
                mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                ng=2, idiag=FALSE, link="linear",data=oysup_long_e_t)
```

```
## Be patient, lcmm is running ...
## The program took 1.03 seconds
```

```
model_3 <- lcmm(fixed = extra ~ grade + grade2,
                mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                ng=3, idiag=FALSE, link="linear",data=oysup_long_e_t)
```

```
## Be patient, lcmm is running ...
## The program took 2.23 seconds
```

```r
model_4 <- lcmm(fixed = extra ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=4, idiag=FALSE, link="linear",data=oysup_long_e_t)
```

```
## Be patient, lcmm is running ...
## The program took 6.25 seconds
```

```r
model_5 <- lcmm(fixed = extra ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=5, idiag=FALSE, link="linear",data=oysup_long_e_t)
```

```
## Be patient, lcmm is running ...
## The program took 32.76 seconds
```

```r
summarytable(model_1, model_2, model_3, model_4, model_5)
```

```
##         G    loglik npm      BIC     %class1  %class2   %class3  %class4
## model_1 1 -3287.927   7 6623.768 100.0000000
## model_2 2 -3260.668  10 6589.784  66.9861555 33.01384
## model_3 3 -3260.668  13 6610.319  61.7678381  0.00000 38.232162
## model_4 4 -3241.093  16 6591.704   7.3482428 30.24494 62.406816  0.00000
## model_5 5 -3227.377  19 6584.806   0.4259851 50.69223  8.732694 27.15655
##          %class5
## model_1
## model_2
## model_3
## model_4
## model_5 12.99255
```

```r
pp_model2 <- postprob(model_2)
```

```
##
## Posterior classification:
##   class1 class2
## N 629.00 310.00
## %  66.99  33.01
##
## Posterior classification table:
##      --> mean of posterior probabilities in each class
##         prob1  prob2
## class1 0.8950 0.1050
## class2 0.1751 0.8249
##
## Posterior probabilities above a threshold (%):
##          class1 class2
## prob>0.7  89.83  75.16
## prob>0.8  78.38  63.87
## prob>0.9  60.89  45.48
##
```

```r
pp_model2[1]
```

```
## [[1]]
##   class1 class2
## N 629.00 310.00
## %  66.99  33.01
```

A two-class solution seems to work best here, too.

```
membership <- as.data.frame(matrix(nrow=939,ncol=2))
colnames(membership) <- c("FAMID","Class_2")
membership[,1:2] <- model_2$pprob[,1:2]
oysup_long_e_t <- merge(x = oysup_long_e_t, y = membership, by="FAMID")

oysup_long_e_t7 <- subset(oysup_long_e_t, subset = grade=="7")
t.test(data = oysup_long_e_t7, extra ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  extra by Class_2
## t = 13.365, df = 253.94, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.057597 1.423123
## sample estimates:
## mean in group 1 mean in group 2
##        4.027397        2.787037
```

```
oysup_long_e_t8 <- subset(oysup_long_e_t, subset = grade=="8")
t.test(data = oysup_long_e_t8, extra ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  extra by Class_2
## t = 21.846, df = 359.79, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.359893 1.628954
## sample estimates:
## mean in group 1 mean in group 2
##        4.124468        2.630045
```

```
oysup_long_e_t9 <- subset(oysup_long_e_t, subset = grade=="9")
t.test(data = oysup_long_e_t9, extra ~ Class_2)
```

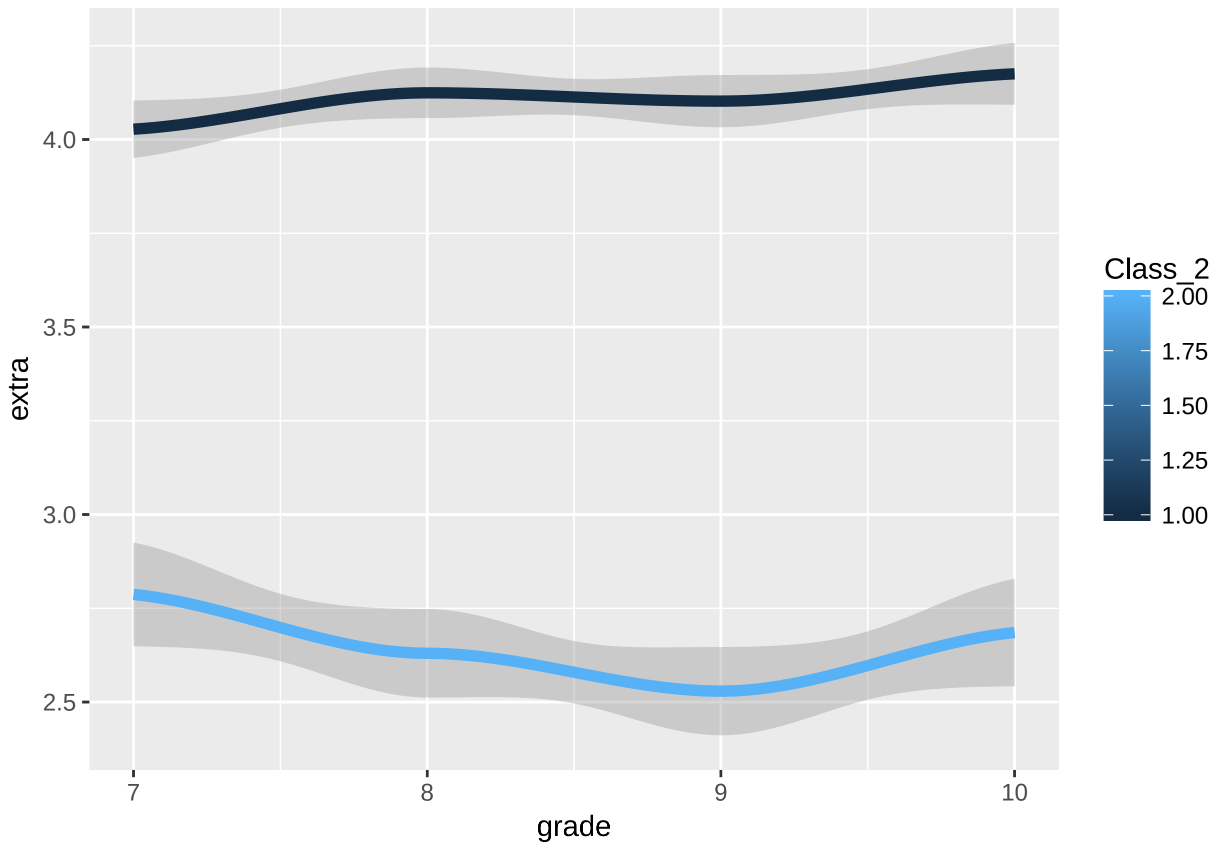
```
##
##  Welch Two Sample t-test
##
## data:  extra by Class_2
## t = 23.395, df = 414.64, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.441064 1.705445
## sample estimates:
## mean in group 1 mean in group 2
##        4.102273        2.529018
```

```
oysup_long_e_t10 <- subset(oysup_long_e_t, subset = grade=="10")
t.test(data = oysup_long_e_t10, extra ~ Class_2)
```

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

```
##
## data:  extra by Class_2
## t = 19.882, df = 249.24, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.342158 1.637300
## sample estimates:
## mean in group 1 mean in group 2
##        4.175159        2.685430
```

```
oysup_long_e_t$FAMID <- as.character(oysup_long_e_t$FAMID)
plot_e_t <- ggplot(oysup_long_e_t, aes(x = grade, y = extra, group = FAMID, color = Class_2)) +
  geom_smooth(aes(group = FAMID, color = Class_2), size = 0.5) +
  geom_smooth(aes(group = Class_2), method = "loess", size = 2, se=T)
plot_e_t
```



Agreeableness

```
oysup_long_a <- subset(oysup_long, subset = !is.na(agree))
oysup_long_a_s <- subset(oysup_long_a, subset = source=="s")

model_1 <- lcmm(fixed = agree ~ grade + grade2,
                random = ~grade, nwg = FALSE, subject = "FAMID",
                ng=1, idiag=FALSE, link="linear",data=oysup_long_a_s)
```

```
## Be patient, lcmm is running ...
## The program took 0.33 seconds
```

```
model_2 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=2, idiag=FALSE, link="linear",data=oysup_long_a_s)
```

```
## Be patient, lcmm is running ...
## The program took 2.61 seconds
```

```
model_3 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=3, idiag=FALSE, link="linear",data=oysup_long_a_s)
```

```
## Be patient, lcmm is running ...
## The program took 1.78 seconds
```

```
model_4 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=4, idiag=FALSE, link="linear",data=oysup_long_a_s)
```

```
## Be patient, lcmm is running ...
## The program took 5.81 seconds
```

```
model_5 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=5, idiag=FALSE, link="linear",data=oysup_long_a_s)
```

```
## Be patient, lcmm is running ...
## The program took 34.87 seconds
```

```
summarytable(model_1, model_2, model_3, model_4, model_5)
```

```
##          G    loglik npm      BIC   %class1  %class2  %class3  %class4
## model_1 1 -3094.514   7 6237.097 100.00000
## model_2 2 -3028.166  10 6125.001  89.89583 10.10417
## model_3 3 -3028.166  13 6145.602  12.29167 87.70833  0.00000
## model_4 4 -3028.166  16 6166.203  12.70833  0.00000 87.29167  0.00000
## model_5 5 -3001.037  19 6132.545   0.00000  3.43750  0.00000 66.97917
##          %class5
## model_1
## model_2
## model_3
## model_4
## model_5 29.58333
```

Two classes seem to fit the data best. Let's look at how many subjects are in each class.

```
# How many subjects per class in 2-class model?
pp_model2 <- postprob(model_2)
```

```
##
## Posterior classification:
##   class1 class2
## N  863.0   97.0
## %   89.9   10.1
##
## Posterior classification table:
##      --> mean of posterior probabilities in each class
```

```
##         prob1  prob2
## class1 0.9618 0.0382
## class2 0.1365 0.8635
##
## Posterior probabilities above a threshold (%):
##          class1 class2
## prob>0.7  95.71  79.38
## prob>0.8  93.05  72.16
## prob>0.9  88.41  58.76
##
```

```
pp_model2[1]
```

```
## [[1]]
##   class1 class2
## N  863.0   97.0
## %   89.9   10.1
```

```
# How are the classes different?

membership <- as.data.frame(matrix(nrow=960,ncol=2))
colnames(membership) <- c("FAMID","Class_2")
membership[,1:2] <- model_2$pprob[,1:2]
oysup_long_a_s <- merge(x = oysup_long_a_s, y = membership, by="FAMID")

# agreeableness in 7th grade
oysup_long_a_s7 <- subset(oysup_long_a_s, subset = grade=="7")
t.test(data = oysup_long_a_s7, agree ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  agree by Class_2
## t = 5.0405, df = 62.723, p-value = 4.212e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.3684726 0.8526336
## sample estimates:
## mean in group 1 mean in group 2
##        4.435115        3.824561
```

```
# agreeableness in 8th grade
oysup_long_a_s8 <- subset(oysup_long_a_s, subset = grade=="8")
t.test(data = oysup_long_a_s8, agree ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  agree by Class_2
## t = 10.295, df = 80.237, p-value = 2.492e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.005707 1.487646
## sample estimates:
## mean in group 1 mean in group 2
##        4.473343        3.226667
```

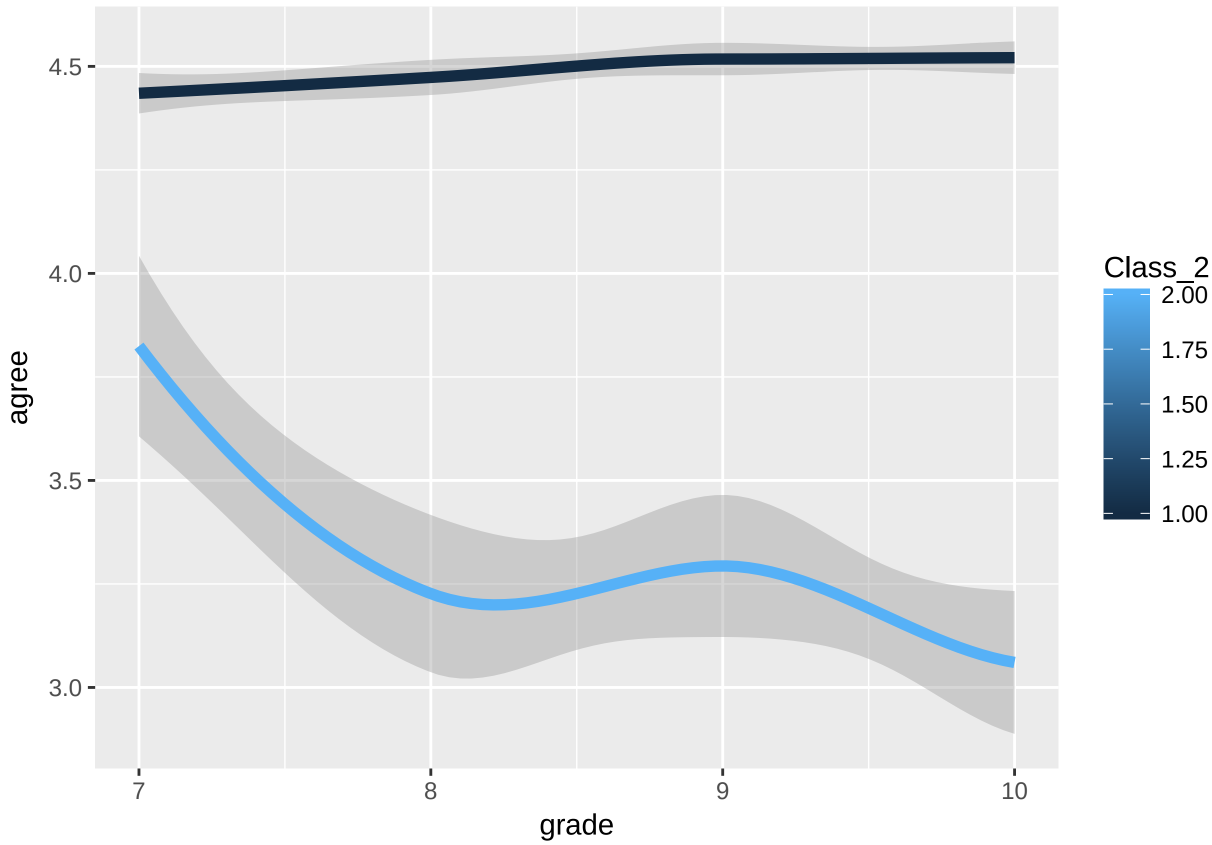
```r
# agreeableness in 9th grade
oysup_long_a_s9 <- subset(oysup_long_a_s, subset = grade=="9")
t.test(data = oysup_long_a_s9, agree ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  agree by Class_2
## t = 14.655, df = 100.54, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.058541 1.389998
## sample estimates:
## mean in group 1 mean in group 2
##        4.517748        3.293478
```

```r
# agreeableness in 10th grade
oysup_long_a_s10 <- subset(oysup_long_a_s, subset = grade=="10")
t.test(data = oysup_long_a_s10, agree ~ Class_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  agree by Class_2
## t = 20.423, df = 101.59, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.318690 1.602406
## sample estimates:
## mean in group 1 mean in group 2
##        4.520988        3.060440
```

```r
oysup_long_a_s$FAMID <- as.character(oysup_long_a_s$FAMID)
plot_a_s <- ggplot(oysup_long_a_s, aes(x = grade, y = agree, group = FAMID, color = Class_2)) +
  geom_smooth(aes(group = FAMID, color = Class_2), size = 0.5) +
  geom_smooth(aes(group = Class_2), method = "loess", size = 2, se=T)
plot_a_s
```



Again, one group seems to be relatively stable and high, while the other seems to decrease in agreeableness over time.

Let's repeat with teacher-reported data.

```
oysup_long_a <- subset(oysup_long, subset = !is.na(agree))
oysup_long_a_t <- subset(oysup_long_a, subset = source=="t")

model_1 <- lcmm(fixed = agree ~ grade + grade2,
                   random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=1, idiag=FALSE, link="linear",data=oysup_long_a_t)
```

```
## Be patient, lcmm is running ...
## The program took 0.34 seconds
```

```
model_2 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=2, idiag=FALSE, link="linear",data=oysup_long_a_t)
```

```
## Be patient, lcmm is running ...
## The program took 1.1 seconds
```

```
model_3 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=3, idiag=FALSE, link="linear",data=oysup_long_a_t)
```

```
## Be patient, lcmm is running ...
```

```
## The program took 4.27 seconds
```

```
model_4 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=4, idiag=FALSE, link="linear",data=oysup_long_a_t)
```

```
## Be patient, lcmm is running ...
## The program took 21 seconds
```

```
model_5 <- lcmm(fixed = agree ~ grade + grade2,
                   mixture = ~grade, random = ~grade, nwg = FALSE, subject = "FAMID",
                   ng=5, idiag=FALSE, link="linear",data=oysup_long_a_t)
```

```
## Be patient, lcmm is running ...
## The program took 32.83 seconds
```

```
summarytable(model_1, model_2, model_3, model_4, model_5)
```

```
##          G    loglik npm      BIC    %class1  %class2   %class3   %class4
## model_1 1 -2835.665   7 5719.243 100.000000
## model_2 2 -2809.308  10 5687.064   5.324814 94.67519
## model_3 3 -2789.652  13 5668.287   5.431310 87.85942  6.709265
## model_4 4 -2789.652  16 5688.821   5.537806 86.79446  0.000000  7.667732
## model_5 5 -2786.800  19 5703.651   3.620873  0.00000 44.408946 47.710330
##           %class5
## model_1
## model_2
## model_3
## model_4
## model_5 4.259851
```

```
pp_model3 <- postprob(model_3)
```

```
##
## Posterior classification:
##   class1 class2 class3
## N  51.00 825.00  63.00
## %   5.43  87.86   6.71
##
## Posterior classification table:
##      --> mean of posterior probabilities in each class
##          prob1  prob2  prob3
## class1 0.7573 0.1780 0.0647
## class2 0.0442 0.8941 0.0617
## class3 0.0869 0.1873 0.7258
##
## Posterior probabilities above a threshold (%):
##           class1 class2 class3
## prob>0.7   66.67  90.18  49.21
## prob>0.8   39.22  82.18  44.44
## prob>0.9   31.37  63.52  33.33
##
```

```
pp_model3[1]
```

```
## [[1]]
##   class1 class2 class3
## N  51.00 825.00  63.00
```

```
## %   5.43  87.86   6.71
```

A three-class solution seems to work best here, actually.

```
membership <- as.data.frame(matrix(nrow=939,ncol=2))
colnames(membership) <- c("FAMID","Class_3")
membership[,1:2] <- model_3$pprob[,1:2]
oysup_long_a_t <- merge(x = oysup_long_a_t, y = membership, by="FAMID")

means_7 <- oysup_long_a_t %>%
  dplyr::filter(grade == "7") %>%
  dplyr::group_by(Class_3) %>%
  dplyr::summarize(mean = mean(agree))
means_7
```

```
## # A tibble: 3 x 2
##   Class_3     mean
##     <int>    <dbl>
## 1       1 3.750000
## 2       2 4.171806
## 3       3 2.166667
```

```
means_8 <- oysup_long_a_t %>%
  dplyr::filter(grade == "8") %>%
  dplyr::group_by(Class_3) %>%
  dplyr::summarize(mean = mean(agree))
means_8
```

```
## # A tibble: 3 x 2
##   Class_3     mean
##     <int>    <dbl>
## 1       1 3.166667
## 2       2 4.085833
## 3       3 2.780702
```

```
means_9 <- oysup_long_a_t %>%
  dplyr::filter(grade == "9") %>%
  dplyr::group_by(Class_3) %>%
  dplyr::summarize(mean = mean(agree))
means_9
```

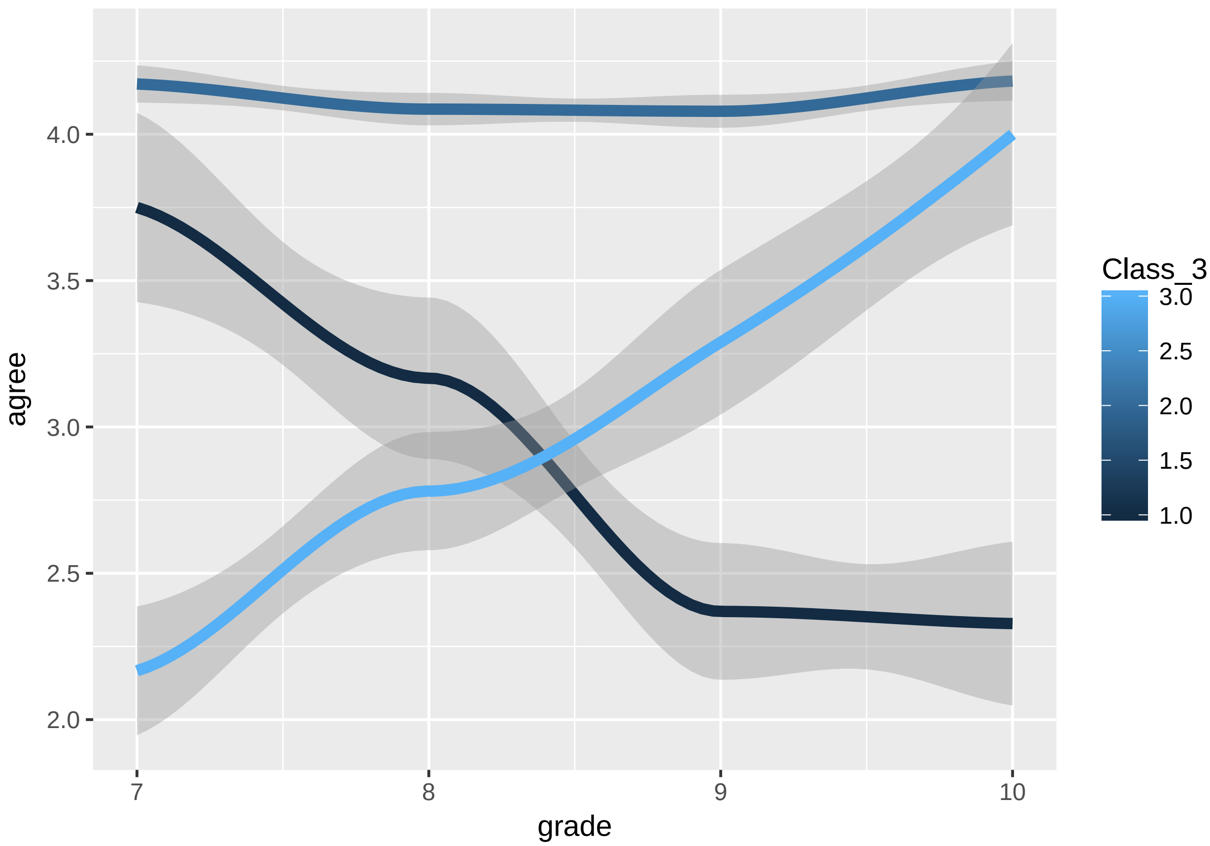
```
## # A tibble: 3 x 2
##   Class_3     mean
##     <int>    <dbl>
## 1       1 2.369565
## 2       2 4.078584
## 3       3 3.289474
```

```
means_10 <- oysup_long_a_t %>%
  dplyr::filter(grade == "10") %>%
  dplyr::group_by(Class_3) %>%
  dplyr::summarize(mean = mean(agree))
means_10
```

```
## # A tibble: 3 x 2
##   Class_3     mean
##     <int>    <dbl>
```

```
## 1       1 2.328125
## 2       2 4.181818
## 3       3 4.000000
```

```
plot_a_t <- ggplot(oysup_long_a_t, aes(x = grade, y = agree, group = FAMID, color = Class_3)) +
  geom_smooth(aes(group = FAMID, color = Class_3), size = 0.5, se=F) +
  geom_smooth(aes(group = Class_3), method = "loess", size = 2, se=T)
plot_a_t
```



There appear to be three distinct trends for students' development in agreeableness (according to teachers' perceptions): one where students are stable, one where students begin high and decrease, and one where students begin low and increase.