Multilevel analysis of longitudinal data in R

University of Leeds Q-Step

10 June 2016

Modelling growth using MLM

Preliminaries

In this session we will be relying on the following R packages:

```
library(lme4)
library(ggplot2)
library(faraway)
library(dplyr)
library(psych)
library(Hmisc)
library(Rmisc)
library(sjPlot)
```

Make sure to run these lines before proceeding. If you get an error message it is probably because one or more of those packages are not installed in your computer. Go ahead and install them by typing and running `install.packages("nameofpackage")` and then calling the library once more.

Data: Alcohol use amongst adolescents

We will analyse data on trajectories of alcohol use amongst teenagers. The dataset contains a measure of alcohol consumption recorded in 3 equally spaced occasions for a sample of 82 teenagers. It also includes other individual traits that may be relevant in explaining between-individual differences in patterns of change of alcohol consumption.

The dataset and the models that we will use today are discussed in much more detail (and still great clarity) in the book *Applied Longitudinal Data: Modeling Change and Event Occurrence* by Singer and Willet (2003) published in OUP.

To load the data from an internet url we use the command read.table and specify the online path where the dataset is stored. In this case the UCLA's Institute for Digital Research and Education which hosts a huge range of online resources for learning statistics and quantitative methods:

```
alco<- read.table("http://www.ats.ucla.edu/stat/r/examples/alda/data/alcohol1_pp.txt",
                  header=T, sep=",")
attach(alco) #this allows us not to have to specify the dataset in which the variables are contained
```

First, let us get a sense of the structure of the table and the variables it contains:

```
names(alco)
```

```
## [1] "id"      "age"     "coa"     "male"    "age_14"  "alcuse"  "peer"    "cpeer"
## [9] "ccoa"
```

In this session, we will use the following variables:

- id: personal identification number - our grouping variable
- age: Age of the individual
- coa: Whether individual had an alcoholic parent
- age_14: Age of individual centered at, i.e. minus, 14.
- alcuse: Number of alcohol units per week consumed by individual
- peer: Number of alcohol units per week consumed by individual's peers
- cpeer: Number of alcohol units per week consumed by individual's peers - centered at mean

Now, let us take a first peek at the data matrix:

```
head(alco)
```

```
##   id age coa male age_14   alcuse      peer     cpeer  ccoa
## 1  1  14   1    0      0 1.732051 1.2649111  0.2469111 0.549
## 2  1  15   1    0      1 2.000000 1.2649111  0.2469111 0.549
## 3  1  16   1    0      2 2.000000 1.2649111  0.2469111 0.549
## 4  2  14   1    1      0 0.000000 0.8944272 -0.1235728 0.549
## 5  2  15   1    1      1 0.000000 0.8944272 -0.1235728 0.549
## 6  2  16   1    1      2 1.000000 0.8944272 -0.1235728 0.549
```

Look carefully at the data and notice how every unique individual has 3 rows, each representing a period of observation of the three in which individuals were interviewed. This type of datasets are known a **long format** as they have many more rows (individual data points) than columns (variables). If you want to know more about longitudinal data formats and how to reshape data in r, check this out

Exploratory data analysis

Let us first let us get basic descriptive stats.

```
describe(alco)
```

```
## alco
##
##  9  Variables      246  Observations
## --------------------------------------------------------------------------------
## id
##         n missing  unique    Info    Mean     .05     .10     .25     .50
##       246       0      82       1    41.5     5.0     9.0    21.0    41.5
##       .75     .90     .95
##      62.0    74.0    78.0
##
## lowest :  1  2  3  4  5, highest: 78 79 80 81 82
## --------------------------------------------------------------------------------
## age
##         n missing  unique    Info    Mean
##       246       0       3    0.89      15
##
## 14 (82, 33%), 15 (82, 33%), 16 (82, 33%)
## --------------------------------------------------------------------------------
```

```
## coa
##        n missing  unique    Info    Sum    Mean
##      246       0       2    0.74    111  0.4512
## -------------------------------------------------------------------------------
## male
##        n missing  unique    Info    Sum    Mean
##      246       0       2    0.75    126  0.5122
## -------------------------------------------------------------------------------
## age_14
##        n missing  unique    Info    Mean
##      246       0       3    0.89       1
##
## 0 (82, 33%), 1 (82, 33%), 2 (82, 33%)
## -------------------------------------------------------------------------------
## alcuse
##        n missing  unique    Info    Mean     .05     .10     .25     .50
##      246       0      14    0.88   0.922   0.000   0.000   0.000   1.000
##     .75     .90     .95
##   1.732   2.646   3.000
##
## 0 (120, 49%), 1 (43, 17%), 1.41421353816986 (16, 7%)
## 1.73205077648163 (16, 7%), 2 (14, 6%)
## 2.2360680103302 (7, 3%), 2.44948983192444 (3, 1%)
## 2.64575123786926 (3, 1%)
## 2.82842707633972 (7, 3%), 3 (7, 3%)
## 3.16227769851685 (3, 1%)
## 3.31662487983704 (2, 1%)
## 3.46410155296326 (3, 1%)
## 3.60555124282837 (2, 1%)
## -------------------------------------------------------------------------------
## peer
##        n missing  unique    Info    Mean
##      246       0       8    0.95   1.018
##
## 0 (63, 26%), 0.894427180290222 (75, 30%)
## 1.26491105556488 (27, 11%)
## 1.54919338226318 (27, 11%)
## 1.78885436058044 (24, 10%), 2 (18, 7%)
## 2.19089031219482 (6, 2%)
## 2.52982211112976 (6, 2%)
## -------------------------------------------------------------------------------
## cpeer
##          n    missing     unique      Info      Mean
##        246          0          8      0.95 -0.0004405
##
## -1.018 (63, 26%), -0.123572819709778 (75, 30%)
## 0.24691105556488 (27, 11%)
## 0.531193382263184 (27, 11%)
## 0.770854360580444 (24, 10%), 0.982 (18, 7%)
## 1.17289031219482 (6, 2%)
## 1.51182211112976 (6, 2%)
## -------------------------------------------------------------------------------
## ccoa
##          n    missing     unique      Info      Mean
```

```
##         246          0        2       0.74 0.0002195
## 
## -0.451 (135, 55%), 0.549 (111, 45%)
## --------------------------------------------------------------------------
```
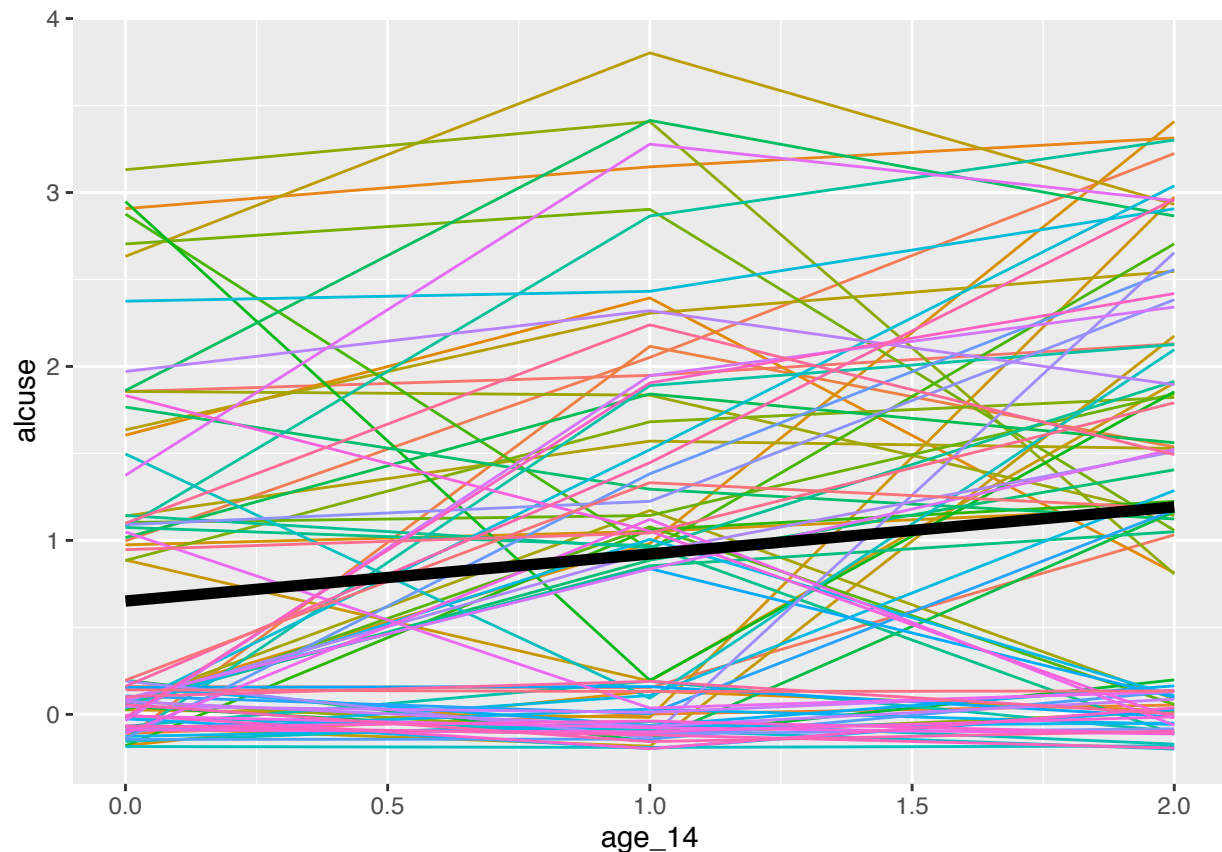
We may also want to know how many unique individuals there are:

```r
length(unique(alco$id)) #number of unique cases (employees)
```
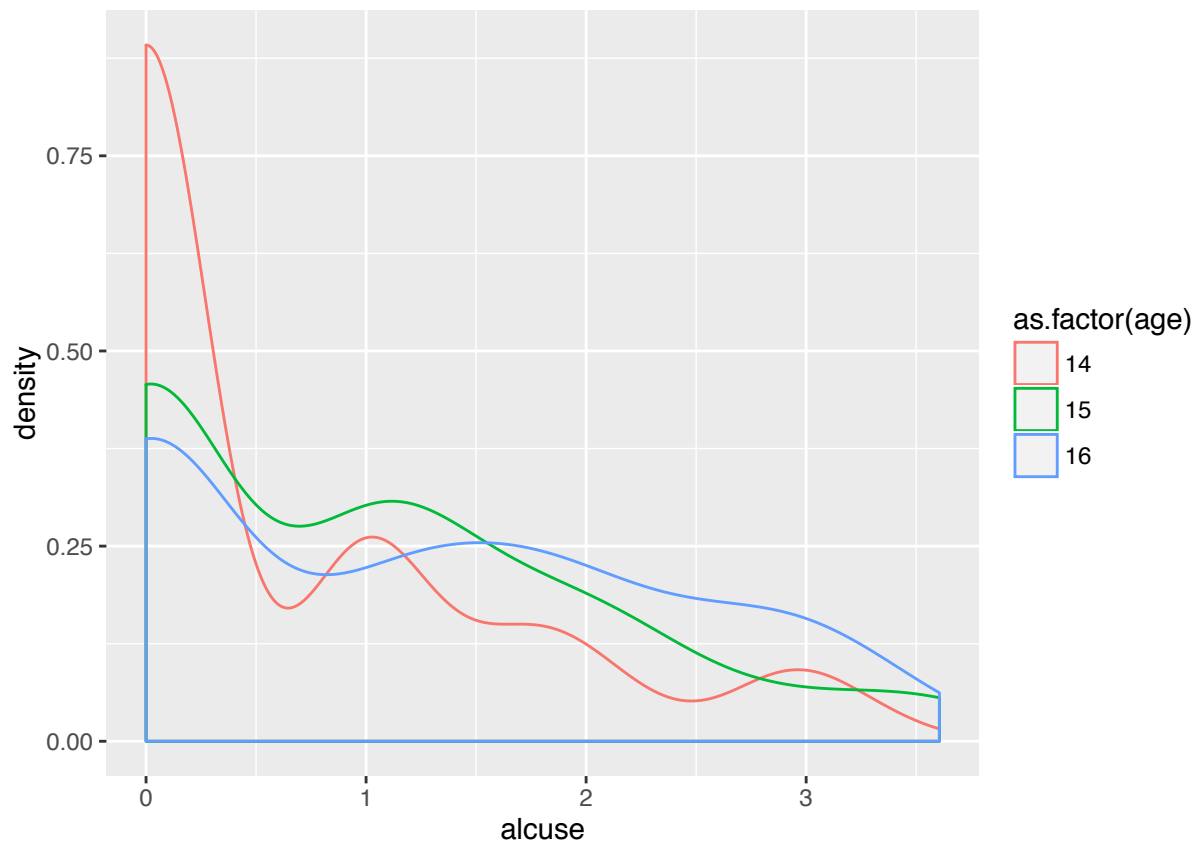
```
## [1] 82
```

It is easy to see the trajectories by means of a "spaghetti"" plot:

```r
ggplot(alco, aes(x=age_14, y=alcuse, colour = factor(id))) +
  geom_line(position=position_jitter(w=0.00, h=0.5)) + guides(colour=FALSE) +
   geom_smooth(method=lm, se=FALSE, colour="black", size=2)
```


Now, let us plot the dependent variable *alcuse* using density plots:

```r
ggplot(alco, aes(x=alcuse, colour=as.factor(age))) + geom_density()
```

We can also plot the evolution of average drinking over the three years. To do so, we will draw an error plot, which requires some previous calculations as below:

```
error <- summarySE(data=alco, measurevar="alcuse", groupvars="age", na.rm=FALSE, conf.interval=.95)

ggplot(error, aes(x=age, y=alcuse, group=1)) +
  geom_line() +
  geom_errorbar(width=.1, aes(ymin=alcuse-ci, ymax=alcuse+ci) , data=error) +
  geom_point(shape=21, size=3, fill="white")
```
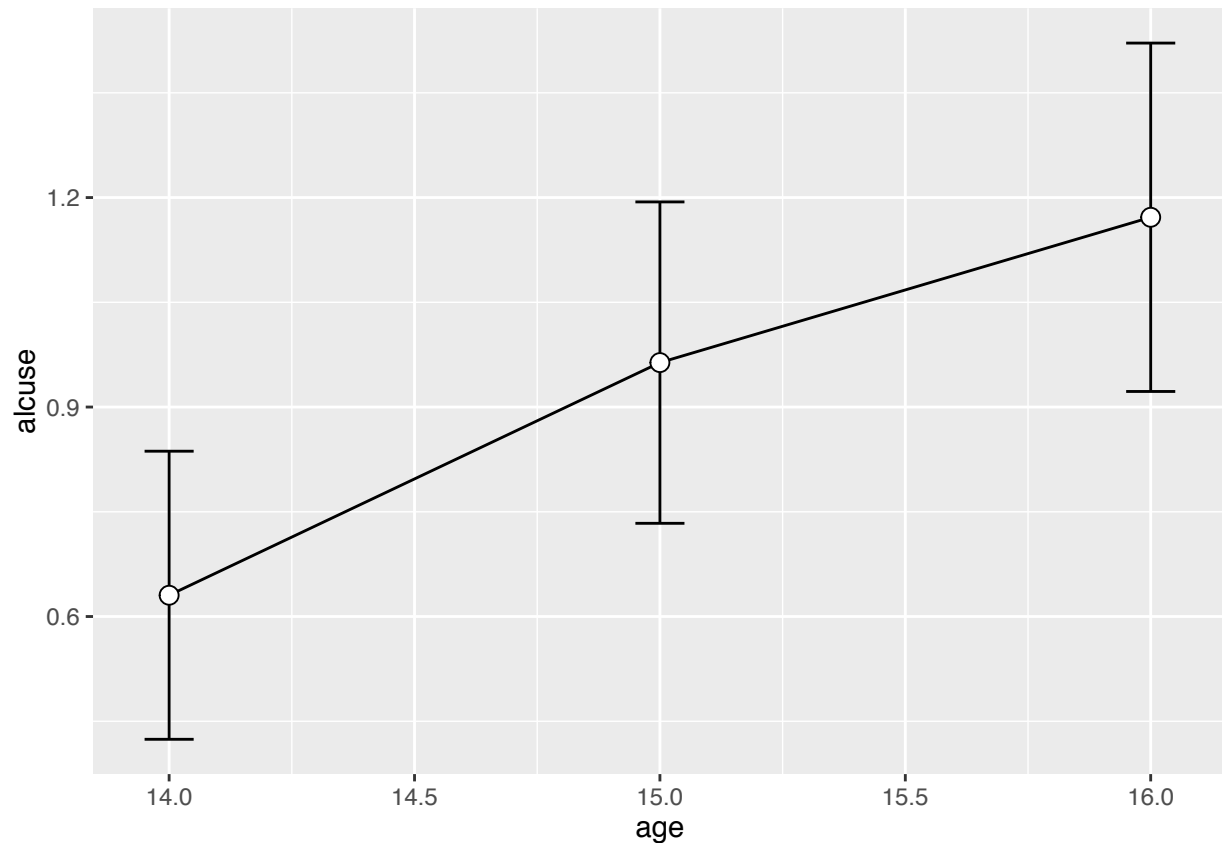
It seems that on average teenagers drink more alcohol as they grow older, which is not counter-intuitive. *Do they all follow the same pattern on growth though?*

To answer this question we need to plot drinking trajectories for each individual (we will plot only the 20 first cases for convenience)

```
alco20 <- filter(alco, id<= 20) # select subsample with command filter

ggplot(alco20, aes (x=age_14, y=alcuse)) +
  geom_line() + facet_wrap( ~ id) # facet_wrap produces a trellis of individual graphs by id
```
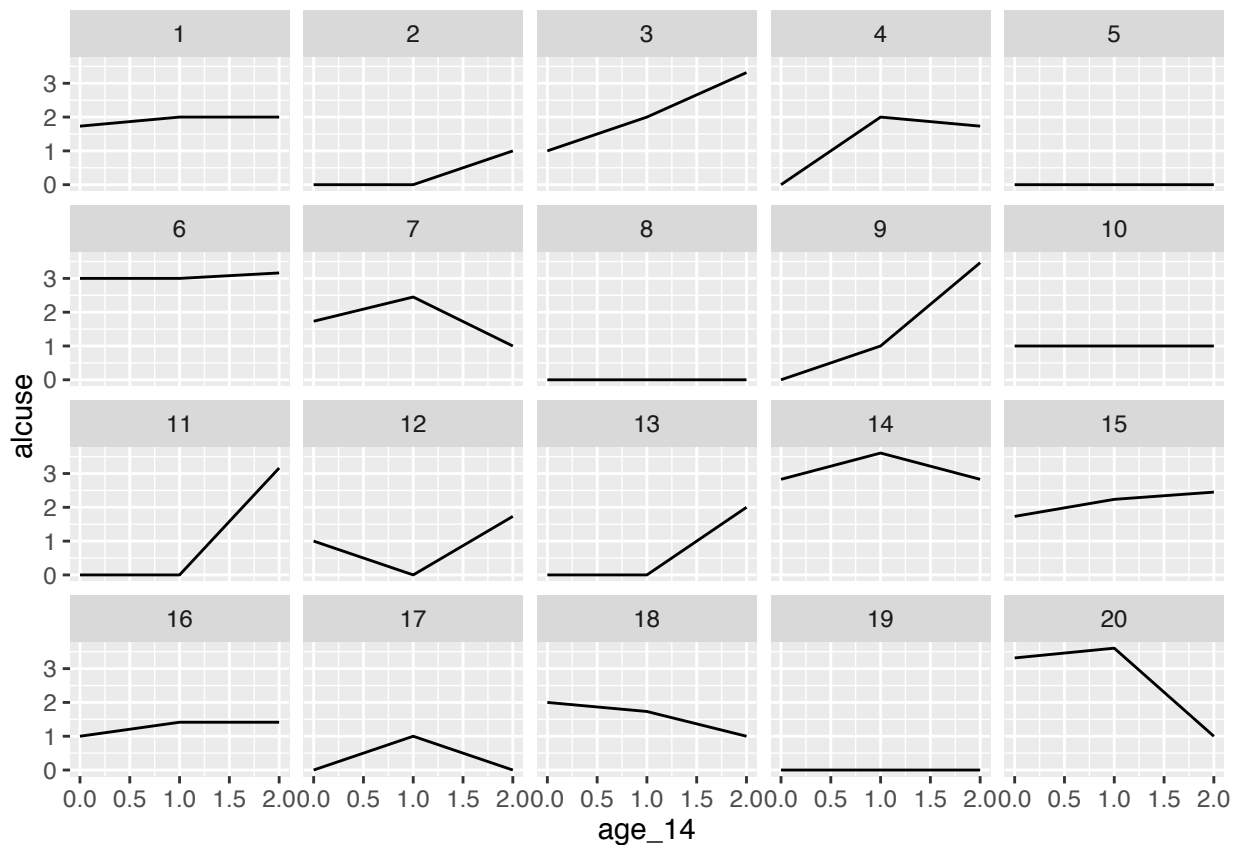
We can also, given that we are concerned with using linear models, fit a regression line on the 3 consecutive observations per each individual. This would be the first step in building a multilevel model (i.e. fitting the level 1 sub-model)

```
ggplot(alco20, aes (x=age_14, y=alcuse)) + geom_point () + geom_smooth(method=lm, se=FALSE) + facet_wrap
```
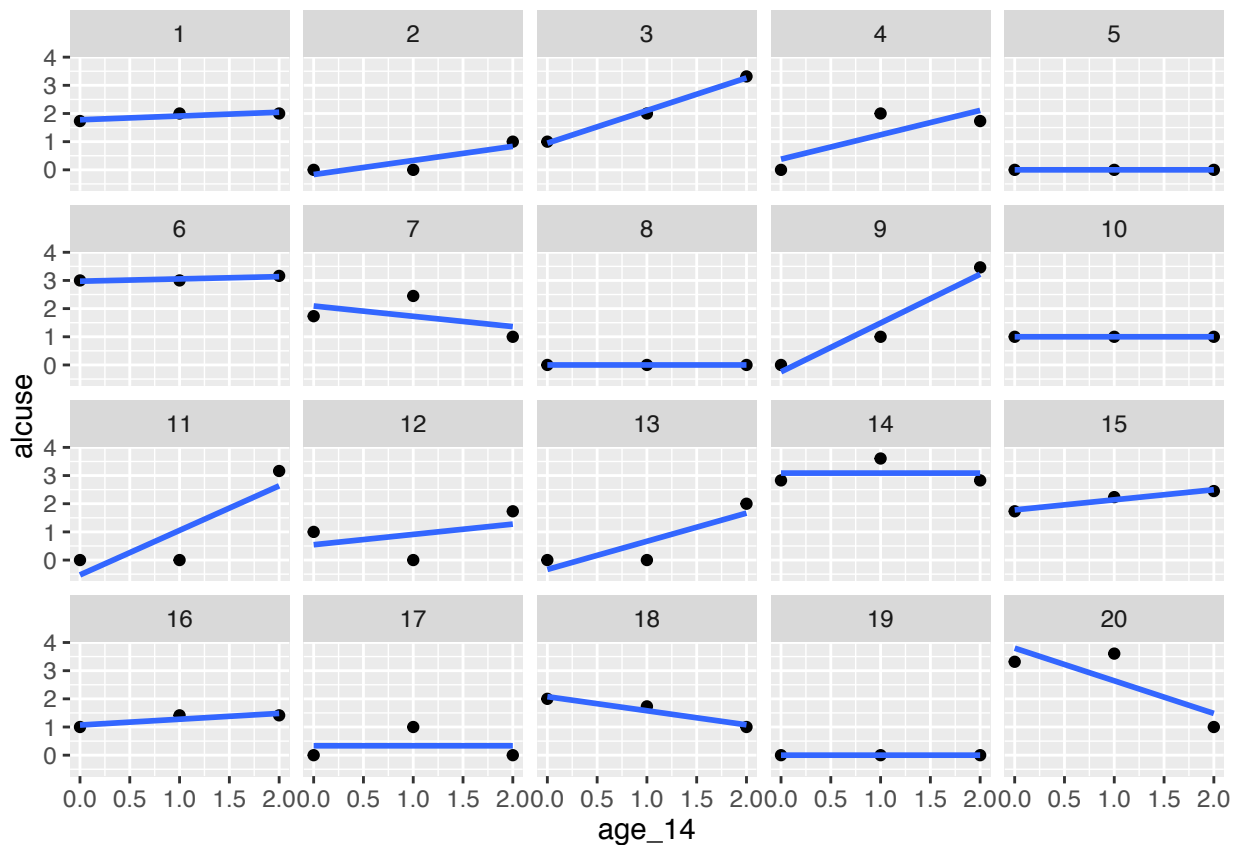
In this exercise we are aiming to understand why teenagers experience different alcohol consumption trajectories over time. Suppose that, after conducting a thorough literature review, we have identified two key factors that explain this variation: (a) whether the child has an alcoholic parent, and (b) the level of alcohol consumption in their network of peers. Again, before we start running models, let us check some exploratory plots.

First, we explore linear drinking trajectories broken by parent's alcoholism:

```
ggplot(alco, aes(x=age_14, y=alcuse, colour = factor(id))) +
  geom_smooth(method=lm, se=FALSE) + guides(colour=FALSE) +
  facet_wrap(~ coa) + geom_smooth(method=lm, se=FALSE, colour="black", size=2)
```
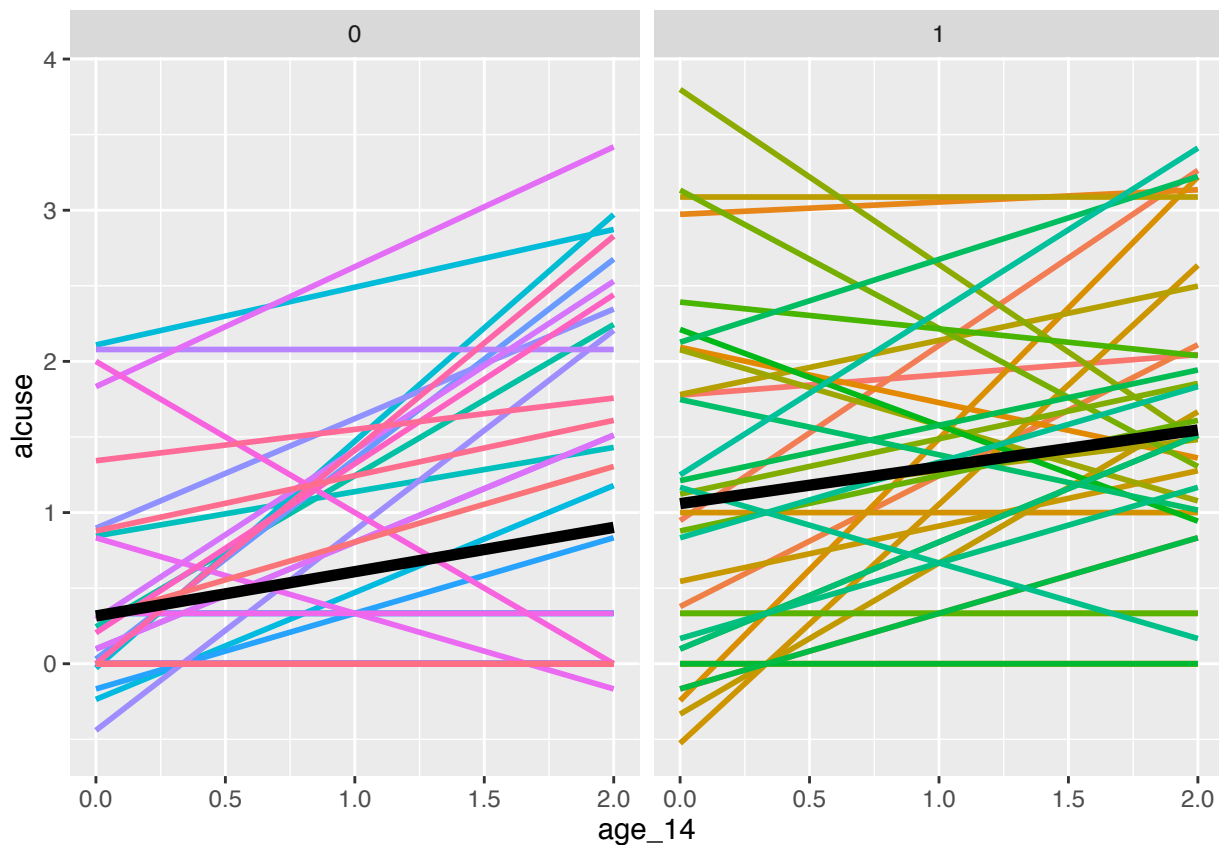
And we do the same by high or low alcohol consumption levels amongst peers:

```
alco$peerbin <- cut(alco$peer,
                    breaks=c(-Inf, 1.01756 , Inf),
                    labels=c("low","high"))


ggplot(alco, aes(x=age_14, y=alcuse, colour = factor(id))) +
  geom_smooth(method=lm, se=FALSE) + guides(colour=FALSE) +
  facet_wrap(~ peerbin) + stat_smooth(method=lm, se=FALSE, colour="black", size=2)
```
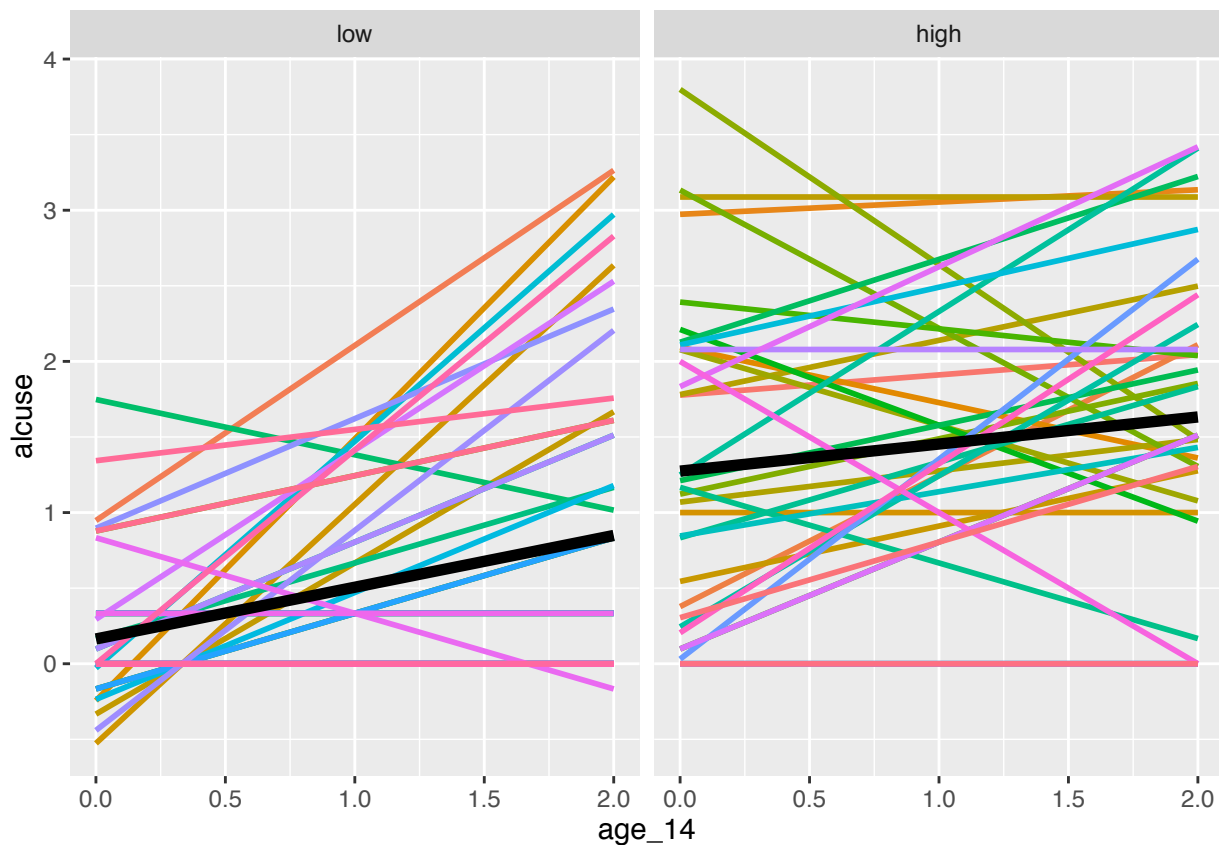
Growth models

Now that we have a preliminary picture of how teenager's alcohol consumption trajectories are patterned, we will fit a variety of models in order of increasing complexity. This will allow us to understand better the properies of the data we are trying to fit and how the model parameters capture them.

Model A: Unconditional Means model

The first model researchers will often fit first is a model without predictors and random effects for the intercept only. This is known as the unconditional means model (UMM) and allows us to explore the nature and source of variation in our our outcome variable:

```
model.a <- lmer(alcuse ~ 1 + 1|id , REML=FALSE)
summary(model.a)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: alcuse ~ 1 + 1 | id
##
##      AIC      BIC   logLik deviance df.resid
##    676.2    686.7   -335.1    670.2      243
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8865 -0.3076 -0.3067  0.6137  2.8567
```

```
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.5639   0.7509
##  Residual             0.5617   0.7495
## Number of obs: 246, groups:  id, 82
## 
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.92195    0.09571   9.633
```

This first model gives us two key bits of information:

1. The intercept $\gamma_{00} = 0.922$ which is often referred to as the population mean - i.e. the mean of alcohol use in the population.

2. The variance of the level 1 residual is a within-person deviation that assesses the distance between individual i's observed Y_{ij} and its true mean π_{0i}.

3. The variance of the level 2 residual is a between-person deviation that assesses the distance between the individual i's true mean π_{0i} and the population average true mean γ_{00}.

4. These two variances allows us to calculate intra-class correlation coefficient, which indicates the share of the total variance in the means that can be attributed to between-individual differences

```
model.a.vc<-as.data.frame(VarCorr(model.a))
model.a.icc<-model.a.vc[1,4]/ (model.a.vc[1,4] + model.a.vc[2,4])
model.a.icc
```

```
## [1] 0.5009395
```

The fact that 50% of the overall variation in patterns of alcohol consumption are due to level 2 or between person differences suggests we should try to predict that variation based on inter-individual differences. In other words, it makes sense to fit a multilevel model of some sort.

Model B: Unconditional Growth Model

Next step is to fit a model that introduces a **time variable** - in this case age_14 - as a **level 1 predictor**. The corresponding coefficient will indicate how alcohol consumption varies as teenagers grow old, or in other words the population average **rate of change**.

```
model.b <- lmer(alcuse ~ age_14 + (age_14|id) , REML=FALSE)
summary(model.b)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: alcuse ~ age_14 + (age_14 | id)
## 
##      AIC      BIC   logLik deviance df.resid
##    648.6    669.6   -318.3    636.6      240
## 
## Scaled residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.47999 -0.38401 -0.07553  0.39001  2.50685
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.6244   0.7902
##           age_14      0.1512   0.3888   -0.22
##  Residual             0.3373   0.5808
## Number of obs: 246, groups:  id, 82
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.65130    0.10508   6.198
## age_14       0.27065    0.06245   4.334
##
## Correlation of Fixed Effects:
##        (Intr)
## age_14 -0.441
```

The UGM model produces new information in contrast with the UMM:

1. In addition of an intercept (which we may want to call now, initial status) we get an age_14 or population growth rate coefficient which is positive and significant - i.e. as teenagers grow old they tend to drink more on average.

2. A second random effect variance that captures the deviation of subject-specific rates of change from the population rate of change.

3. Notice that by introducing the level 1 predictor, we have managed to reduce the residual variance from .56 to .34.

Model C: Introducing predictors I (COA)

Now, we are in a position to start introducing predictors at the between-subject level to account for the variation in longitudinal patterns of alcohol consumption. In practice, this means introducing effects on both the **intercept or initial status** and on **the rate of change**. In model C we introduce the variable COA (child of an alcoholic) both as main effect and as an interaction effect with time.

```
model.c <- lmer(alcuse ~ coa + age_14 + coa:age_14 + (age_14|id) , REML=FALSE)
summary(model.c)
```
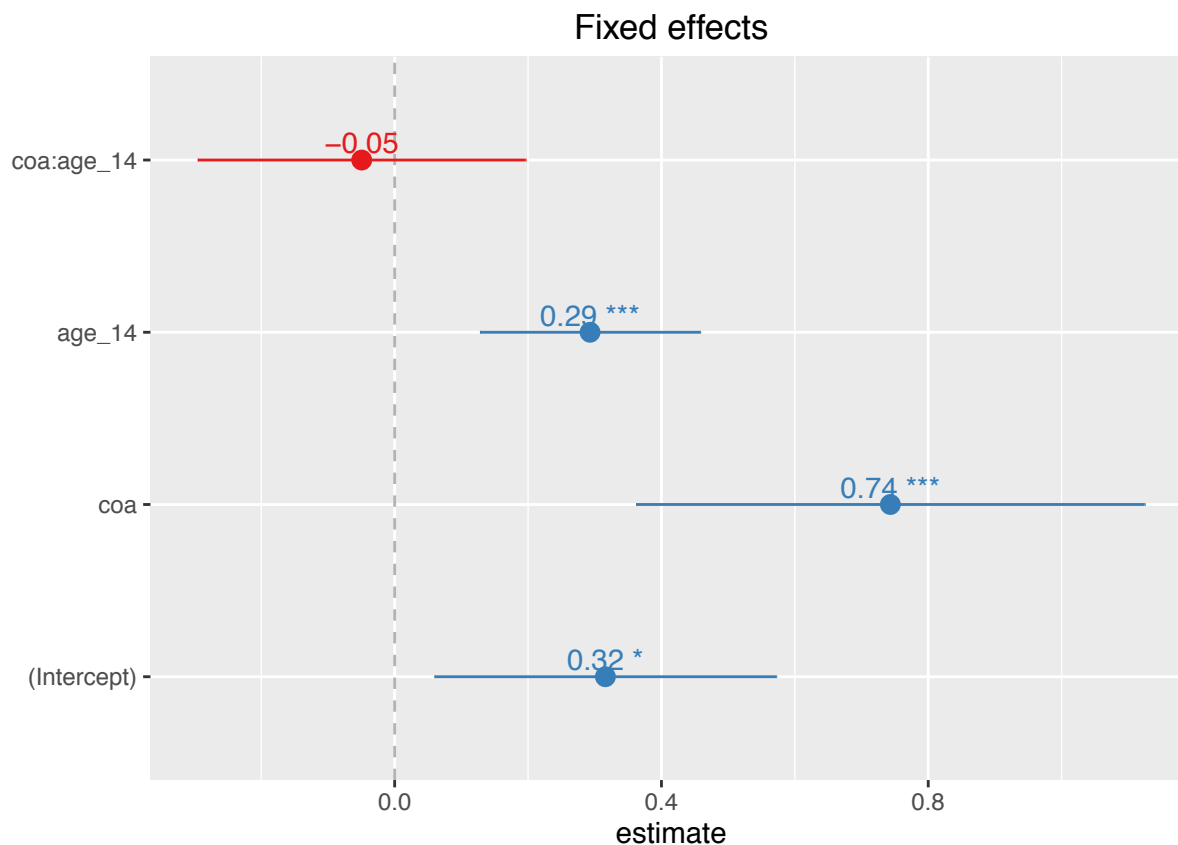
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: alcuse ~ coa + age_14 + coa:age_14 + (age_14 | id)
##
##      AIC      BIC   logLik deviance df.resid
##    637.2    665.2   -310.6    621.2      238
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5480 -0.3880 -0.1057  0.3602  2.3961
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.4876   0.6983
```

```
##          age_14       0.1506  0.3881   -0.22
##  Residual             0.3373  0.5808
## Number of obs: 246, groups:  id, 82
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  0.31595    0.13070   2.417
## coa          0.74321    0.19457   3.820
## age_14       0.29296    0.08423   3.478
## coa:age_14  -0.04943    0.12539  -0.394
##
## Correlation of Fixed Effects:
##           (Intr) coa    age_14
## coa       -0.672
## age_14    -0.460  0.309
## coa:age_14  0.309 -0.460 -0.672
```

The results show:

1. A positive and significant effect of COA on alcohol consumption - i.e. children of alcoholic parents have higher levels of consumption on average.

2. A negative, yet small and insignificant effect of COA on the rate of change (effect of time on alcohol use) - i.e. the alcohol consumption of children of alcoholic parents do not seem to grow faster or slower over time than that of children from who are not alcoholic.

3. Notice how the intercept has been halved and that the variance of the random effects for the intercept has fallen from .684 to .487.

```
sjp.lmer(model.c, type="fe")
```

Fixed effects

Model D: Introducing Predictors II (PEER)

Now we introduce PEER (alcohol consumption amongst the teenagers' network of peers) as main effect and interaction effect with age:
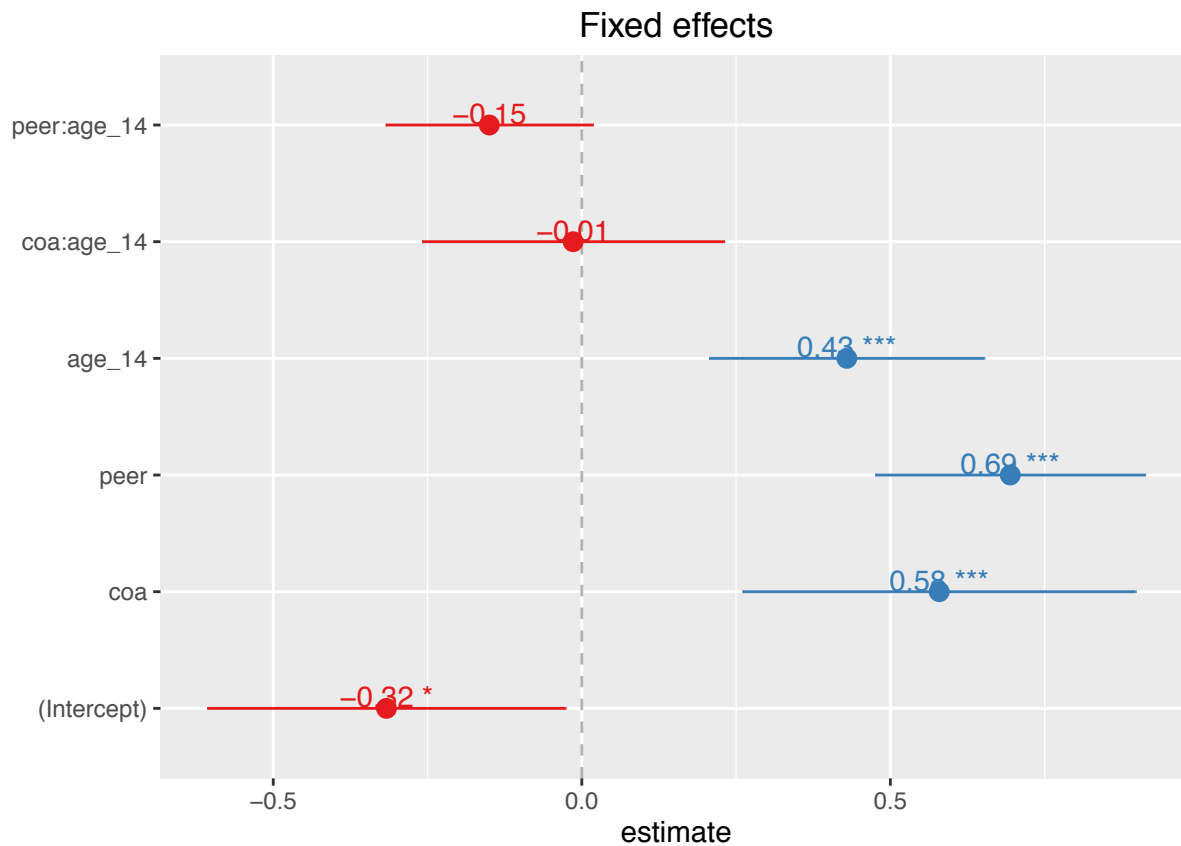
```r
model.d <- lmer(alcuse ~ coa + peer + age_14 + coa:age_14
                + peer:age_14 + (age_14|id) , REML=FALSE)
summary(model.d)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## alcuse ~ coa + peer + age_14 + coa:age_14 + peer:age_14 + (age_14 |
##     id)
##
##      AIC      BIC   logLik deviance df.resid
##    608.7    643.7   -294.3    588.7      236
##
## Scaled residuals:
##     Min       1Q   Median       3Q      Max
## -2.59554 -0.40005 -0.07769  0.46003  2.29374
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.2409   0.4908
##           age_14      0.1391   0.3730   -0.03
##  Residual             0.3373   0.5808
```

```
## Number of obs: 246, groups:  id, 82
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -0.31651    0.14806  -2.138
## coa          0.57917    0.16249   3.564
## peer         0.69430    0.11153   6.225
## age_14       0.42943    0.11369   3.777
## coa:age_14  -0.01403    0.12477  -0.112
## peer:age_14 -0.14982    0.08564  -1.749
##
## Correlation of Fixed Effects:
##             (Intr) coa    peer   age_14 c:g_14
## coa         -0.371
## peer        -0.686 -0.162
## age_14      -0.436  0.162  0.299
## coa:age_14   0.162 -0.436  0.071 -0.371
## peer:age_14  0.299  0.071 -0.436 -0.686 -0.162
```

We can visualise the fixed effects's coefficients:

```
sjp.lmer(model.d, type="fe")
```


Fixed effects

Model E: Dropping non-significant interactions

Given that the effect of COA on rate of change remains insignificant we decide to drop it from our model:

```
model.e <- lmer(alcuse ~ coa + peer + age_14 + peer:age_14
                + (age_14|id) , REML=FALSE)
summary(model.e)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: alcuse ~ coa + peer + age_14 + peer:age_14 + (age_14 | id)
##
##      AIC      BIC   logLik deviance df.resid
##    606.7    638.3   -294.4    588.7      237
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.59554 -0.40414 -0.08352  0.45549  2.29975
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.2409   0.4908
##           age_14      0.1392   0.3730   -0.03
##  Residual             0.3373   0.5808
## Number of obs: 246, groups:  id, 82
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -0.31382    0.14611  -2.148
## coa          0.57120    0.14623   3.906
## peer         0.69518    0.11126   6.249
## age_14       0.42469    0.10559   4.022
## peer:age_14 -0.15138    0.08451  -1.791
##
## Correlation of Fixed Effects:
##             (Intr) coa    peer   age_14
## coa         -0.338
## peer        -0.709 -0.146
## age_14      -0.410  0.000  0.351
## peer:age_14  0.334  0.000 -0.431 -0.814
```

We can visualise the fixed effects's coefficients:

```
sjp.lmer(model.e, type="fe")
```

Fixed effects
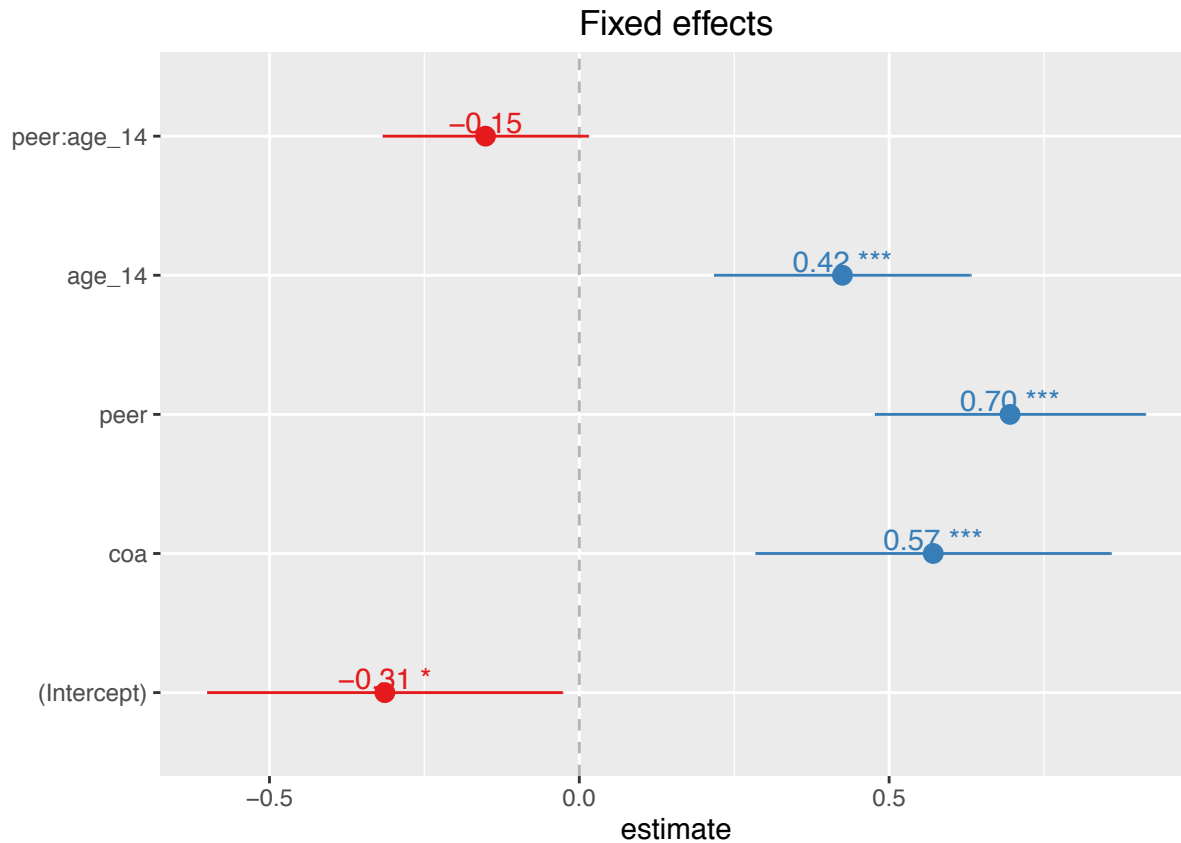
Model F: Introducing centered predictors

```r
model.f <- lmer(alcuse ~ coa + cpeer + age_14 + cpeer:age_14
                + (age_14|id) , REML=FALSE)
summary(model.f)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: alcuse ~ coa + cpeer + age_14 + cpeer:age_14 + (age_14 | id)
##
##      AIC      BIC   logLik deviance df.resid
##    606.7    638.3   -294.4    588.7      237
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.59554 -0.40414 -0.08352  0.45549  2.29975
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.2409   0.4908
##           age_14      0.1392   0.3730   -0.03
##  Residual             0.3373   0.5808
## Number of obs: 246, groups:  id, 82
##
## Fixed effects:
##              Estimate Std. Error t value
```

```
## (Intercept)    0.39387    0.10354    3.804
## coa            0.57120    0.14623    3.906
## cpeer          0.69518    0.11126    6.249
## age_14         0.27058    0.06127    4.416
## cpeer:age_14  -0.15138    0.08451   -1.791
##
## Correlation of Fixed Effects:
##             (Intr) coa     cpeer   age_14
## coa         -0.637
## cpeer        0.094 -0.146
## age_14      -0.336  0.000  0.000
## cpeer:ag_14  0.000  0.000 -0.431  0.001
```

We can visualise the fixed effects's coefficients:

```
sjp.lmer(model.e, type="fe")
```
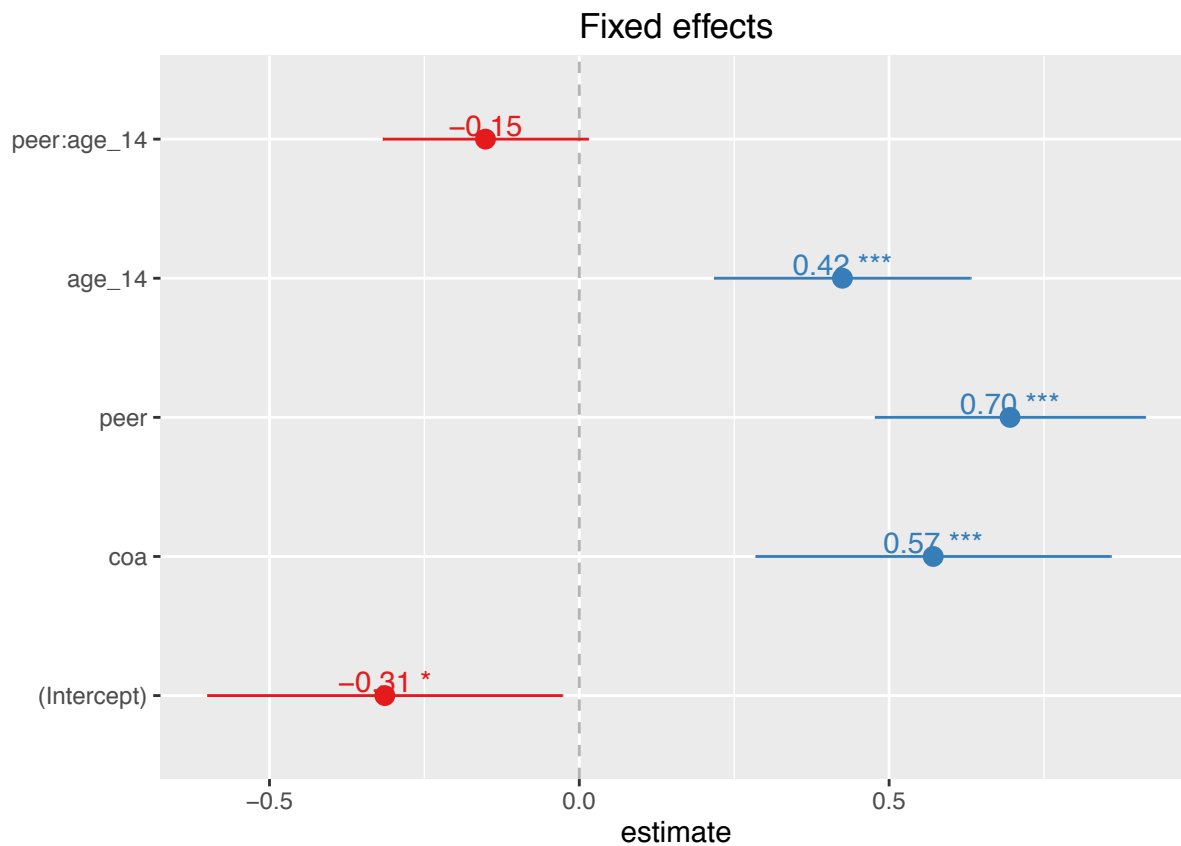

We can also retrieve

```
anova(model.a,model.b,model.c, model.d, model.e, model.f)
```

```
## Data: NULL
## Models:
## model.a: alcuse ~ 1 + 1 | id
## model.b: alcuse ~ age_14 + (age_14 | id)
## model.c: alcuse ~ coa + age_14 + coa:age_14 + (age_14 | id)
```

```
## model.e: alcuse ~ coa + peer + age_14 + peer:age_14 + (age_14 | id)
## model.f: alcuse ~ coa + cpeer + age_14 + cpeer:age_14 + (age_14 | id)
## model.d: alcuse ~ coa + peer + age_14 + coa:age_14 + peer:age_14 + (age_14 |
## model.d:      id)
##          Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## model.a   3 676.16 686.67 -335.08   670.16
## model.b   6 648.61 669.64 -318.31   636.61 33.5449      3  2.472e-07 ***
## model.c   8 637.20 665.25 -310.60   621.20 15.4085      2  0.0004509 ***
## model.e   9 606.70 638.25 -294.35   588.70 32.4993      1  1.192e-08 ***
## model.f   9 606.70 638.25 -294.35   588.70  0.0000      0  < 2.2e-16 ***
## model.d  10 608.69 643.74 -294.35   588.69  0.0126      1  0.9104569
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fitting Prototypical trajectories

Model B: Population average trajectories

```
model.b <- lmer(alcuse ~ age_14 + (age_14|id), REML=FALSE)

fixef.b <- fixef(model.b)

fit.b <- fixef.b[[1]] + alco$age_14[1:3]*fixef.b[[2]]

plot(alco$age[1:3], fit.b, ylim=c(0, 2), type="b",
     ylab="predicted alcuse", xlab="age")

title("Model B \n Unconditional growth model")
```
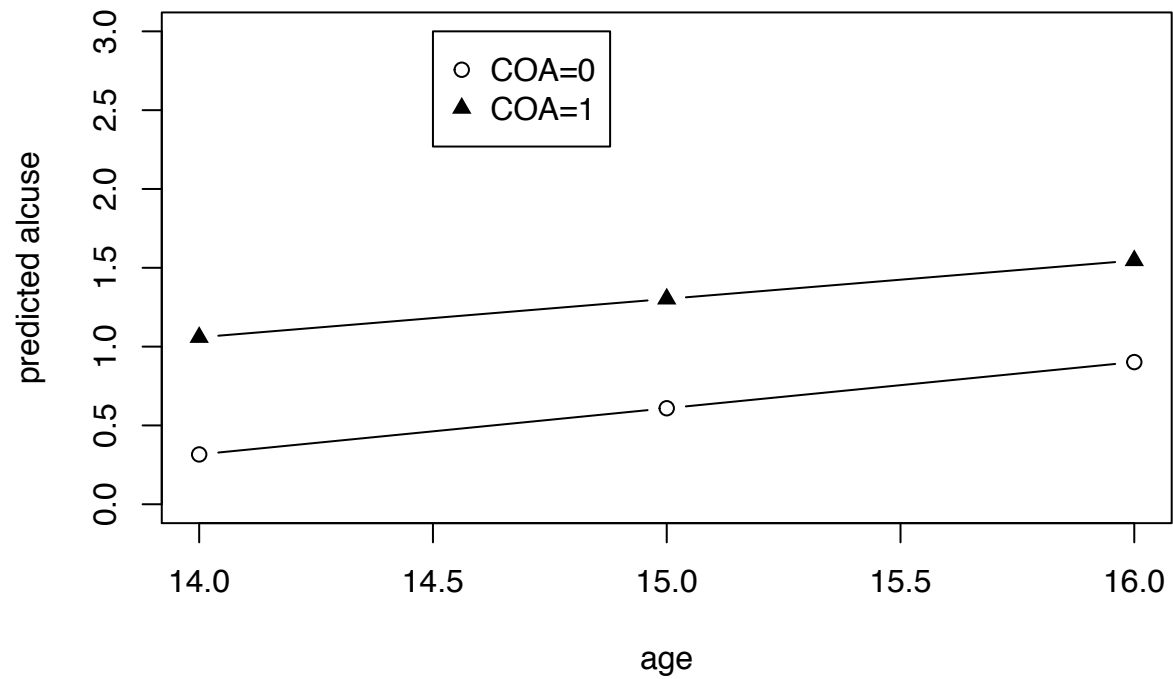
Model B
Unconditional growth model
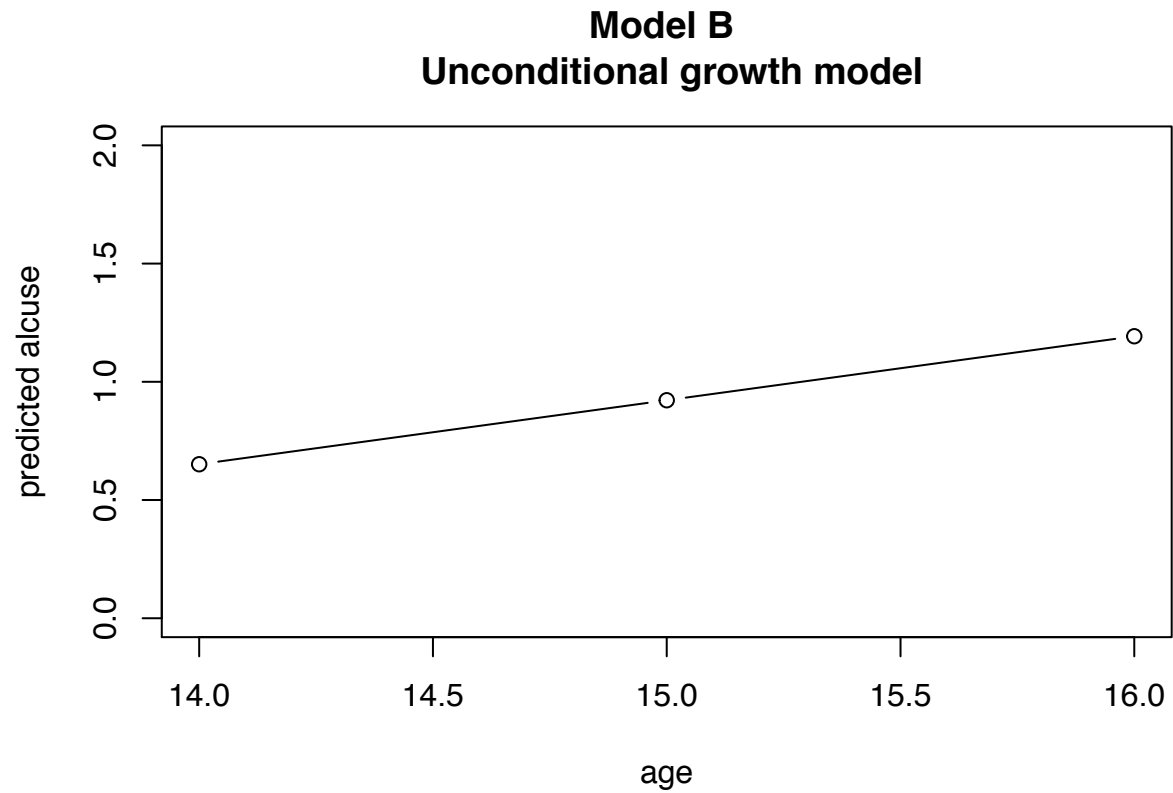
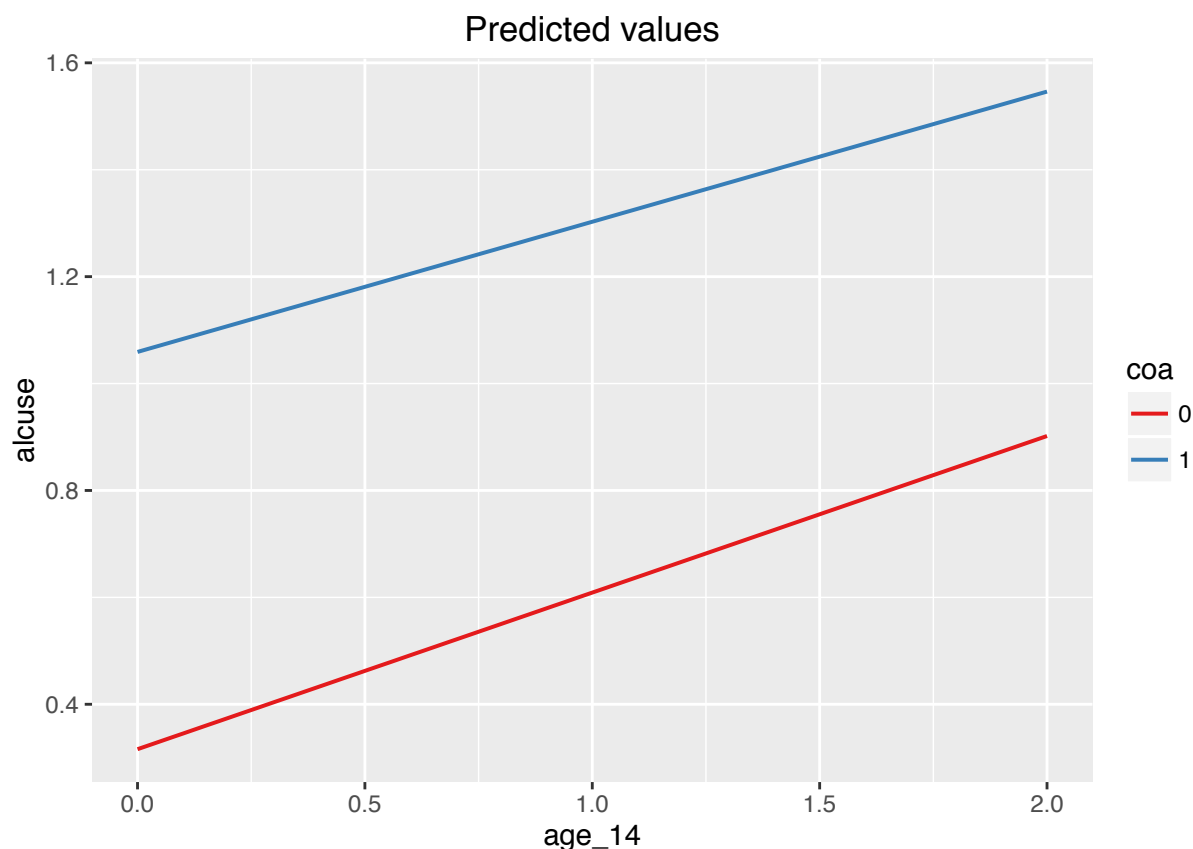**Model C: Trajectories by COA (Y/N)**

```
fixef.c <- fixef(model.c)
fit.c0 <- fixef.c[[1]] + alco$age_14[1:3]*fixef.c[[3]]
fit.c1 <- fixef.c[[1]] + fixef.c[[2]] +
          alco$age_14[1:3]*fixef.c[[3]] +
          alco$age_14[1:3]*fixef.c[[4]]
plot(alco$age[1:3], fit.c0, ylim=c(0, 3), type="b",
     ylab="predicted alcuse", xlab="age")
lines(alco$age[1:3], fit.c1, type="b", pch=17)
title("Model C \n Uncontrolled effects of COA")
legend(14.5, 3, c("COA=0", "COA=1"), pch=c(1, 17))
```

Model C
Uncontrolled effects of COA

It is also possible to use sjPlot's sjp.lmer function to do this when using categorical covariates:

```
sjp.lmer(model.c, type = "pred", facet.grid = FALSE,
         vars = c("age_14", "coa"))
```

Predicted values

Model E: Trajectories by COA (Y/N) & PEER (High 1.381 vs Low .655)

```
fixef.e <- fixef(model.e)
fit.ec0p0 <- fixef.e[[1]] + .655*fixef.e[[3]] +
            alco$age_14[1:3]*fixef.e[[4]] +
            .655*alco$age_14[1:3]*fixef.e[[5]]
fit.ec0p1 <- fixef.e[[1]] + 1.381*fixef.e[[3]] +
            alco$age_14[1:3]*fixef.e[[4]] +
            1.381*alco$age_14[1:3]*fixef.e[[5]]
fit.ec1p0 <- fixef.e[[1]] + fixef.e[[2]] + .655*fixef.e[[3]] +
            alco$age_14[1:3]*fixef.e[[4]] +
            .655*alco$age_14[1:3]*fixef.e[[5]]
fit.ec1p1 <- fixef.e[[1]] + fixef.e[[2]] + 1.381*fixef.e[[3]] +
            alco$age_14[1:3]*fixef.e[[4]] +
            1.381*alco$age_14[1:3]*fixef.e[[5]]
plot(alco$age[1:3], fit.ec0p0, ylim=c(0, 3), type="b",
     ylab="predicted alcuse", xlab="age", pch=2)
lines(alco$age[1:3], fit.ec0p1, type="b", pch=0)
lines(alco$age[1:3], fit.ec1p0, type="b", pch=17)
lines(alco$age[1:3], fit.ec1p1, type="b", pch=15)
title("Model E \n *Final* model for the controlled effects of COA")
legend(14.5, 3, c("COA=0, low peer", "COA=0, high peer",
       "COA=1, low peer", "COA=1, high peer"), pch=c(2, 0, 17,15))
```

Model E
Final model for the controlled effects of COA
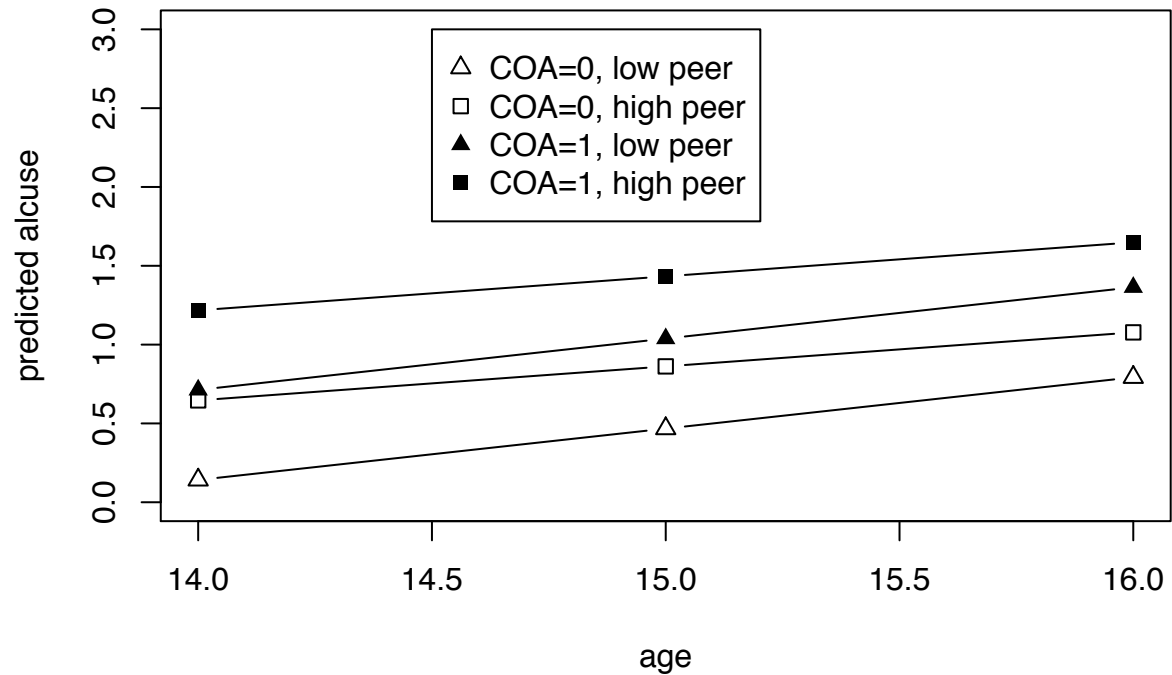
Model diagnostics

Finally, we will run a series of diagnostic plots to evaluate the quality of the model fit. First we need to run a few lines to prepare our data:

```r
modelf_diag <- fortify(model.f)
modelf_diag <- as.data.frame(modelf_diag)

re <- ranef(model.f)
str(re)
```
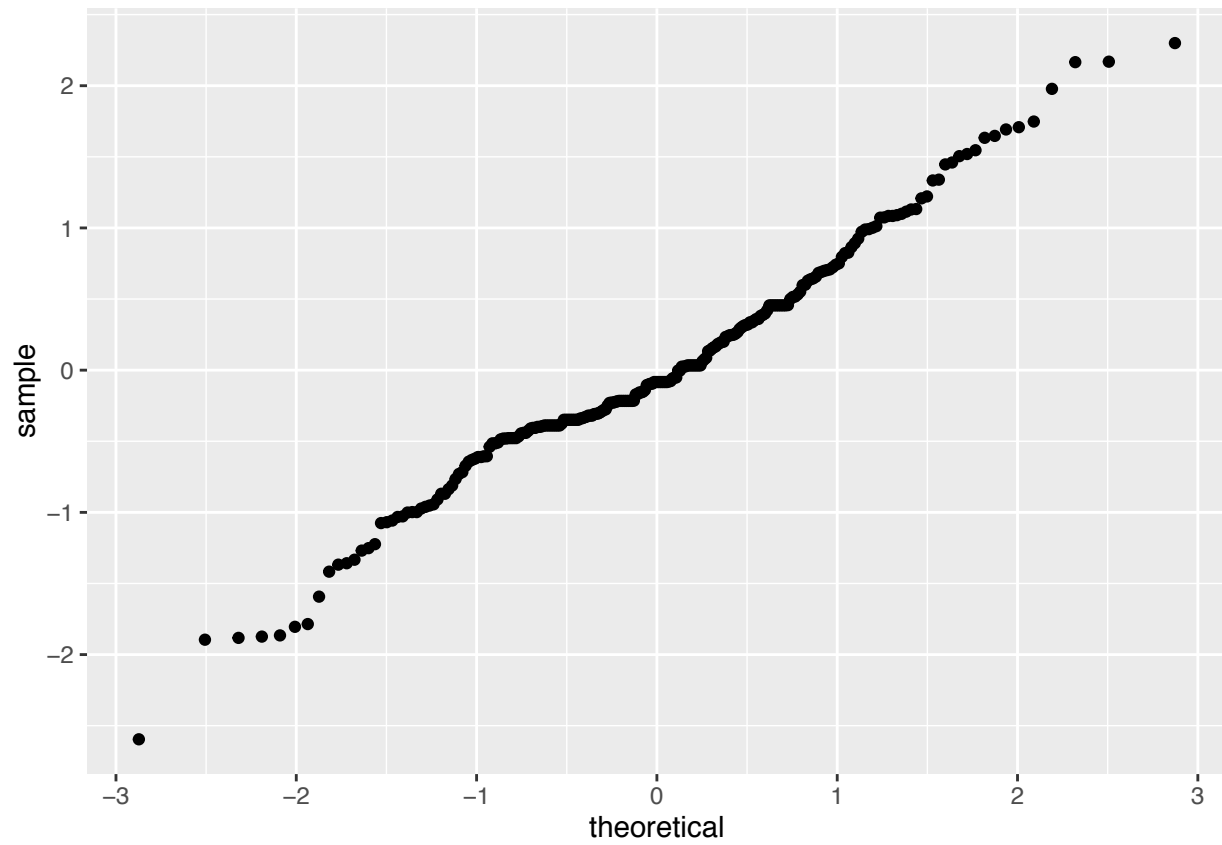
```
## List of 1
##  $ id:'data.frame':  82 obs. of  2 variables:
##   ..$ (Intercept): num [1:82] 0.33 -0.524 0.298 -0.418 -0.58 ...
##   ..$ age_14     : num [1:82] 0.0558 -0.0644 0.4897 0.1981 -0.3107 ...
##  - attr(*, "class")= chr "ranef.mer"
```
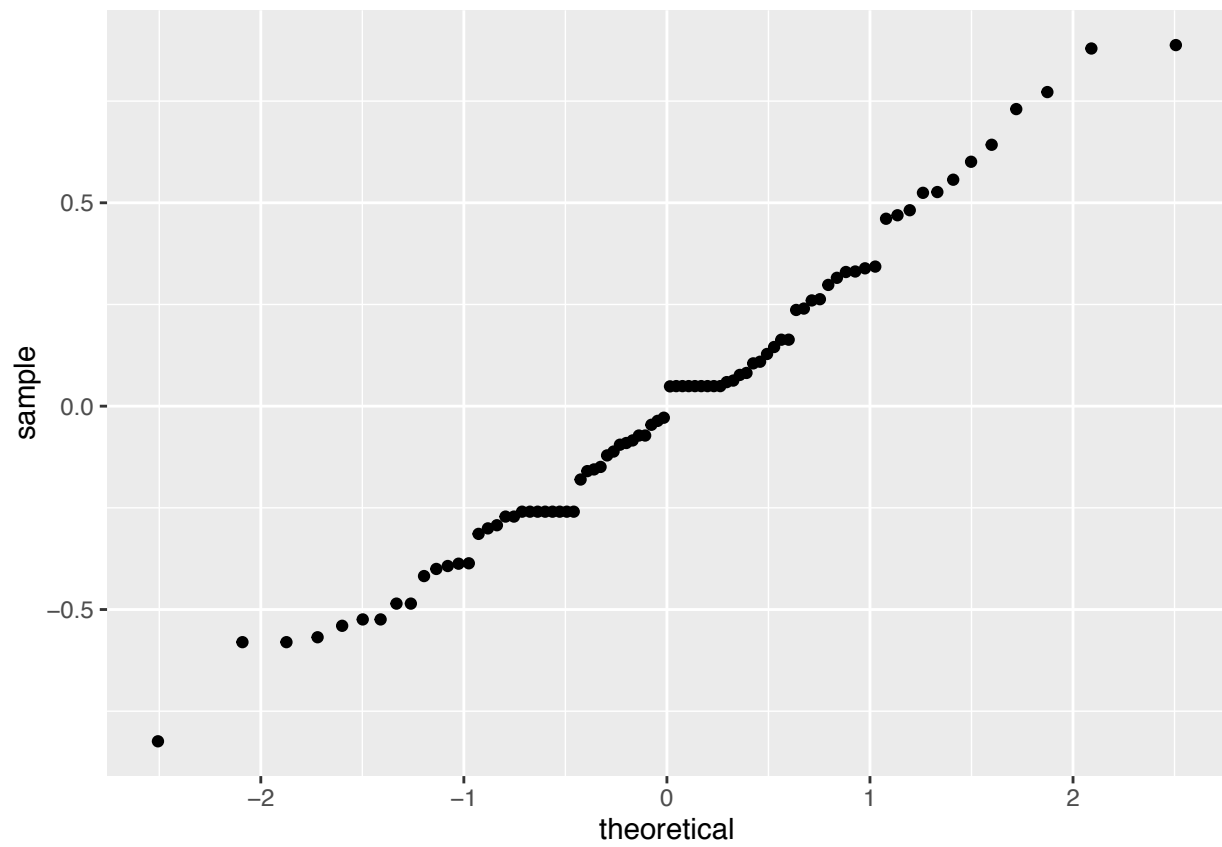
```r
names(re$id)[1] <- "Intercept"
```
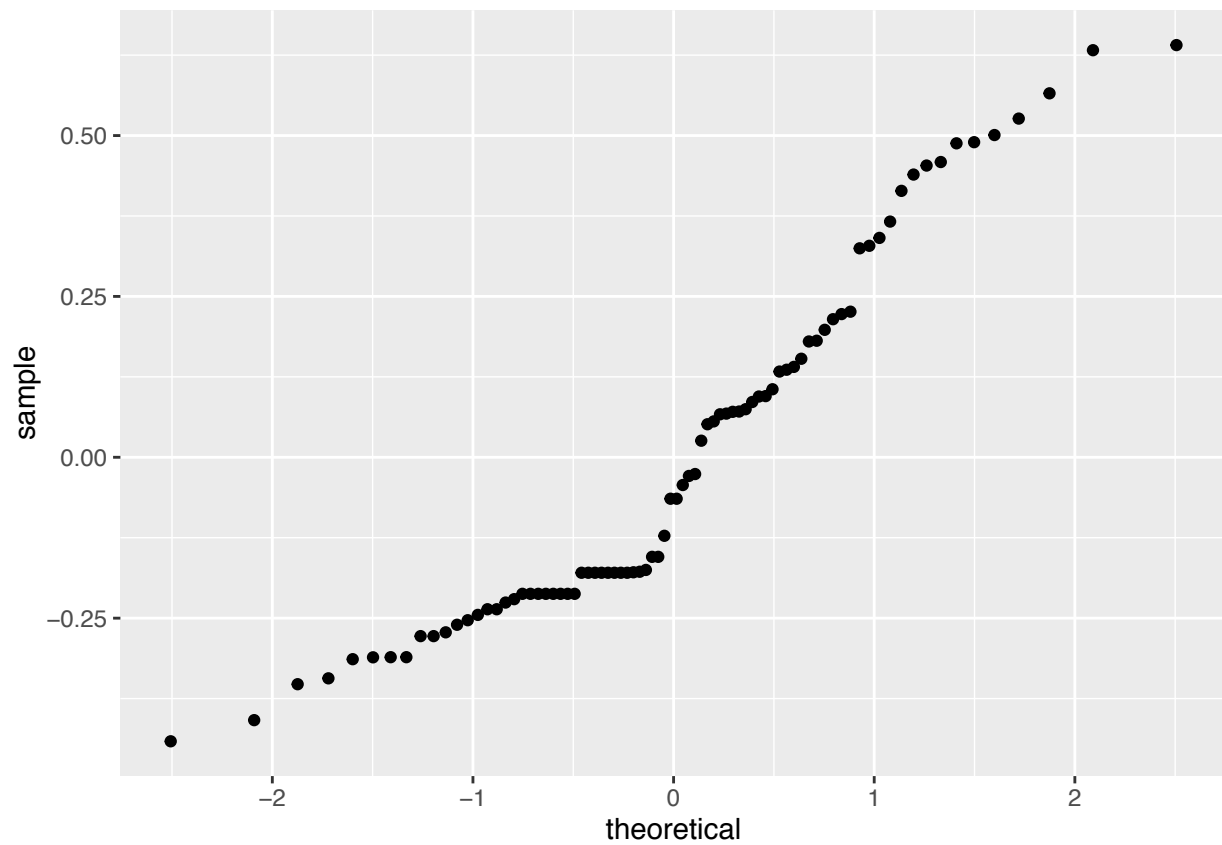
Normality

```r
ggplot(modelf_diag, aes(sample=.scresid)) +
  stat_qq()
```

```
ggplot(re$id, aes(sample=Intercept)) +
  stat_qq()
```
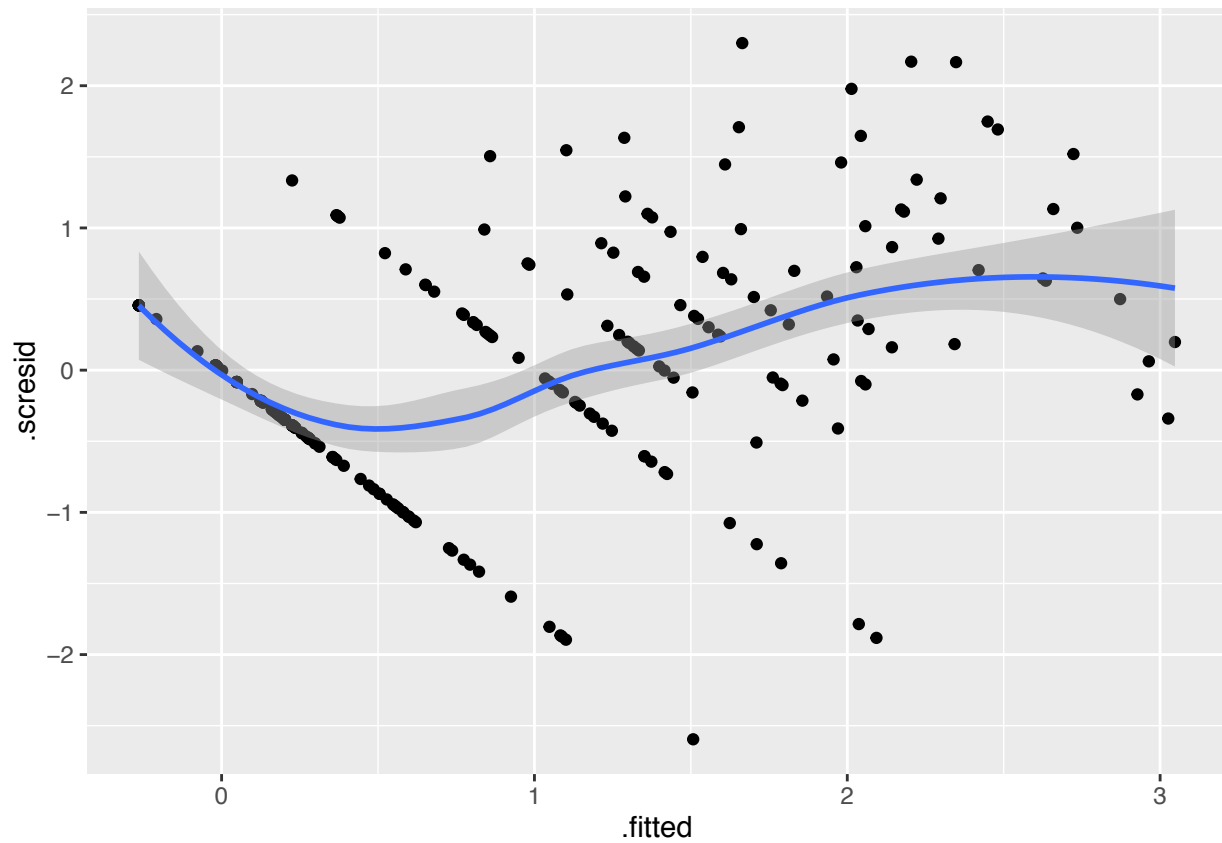
```r
ggplot(re$id, aes(sample=age_14)) +
  stat_qq()
```
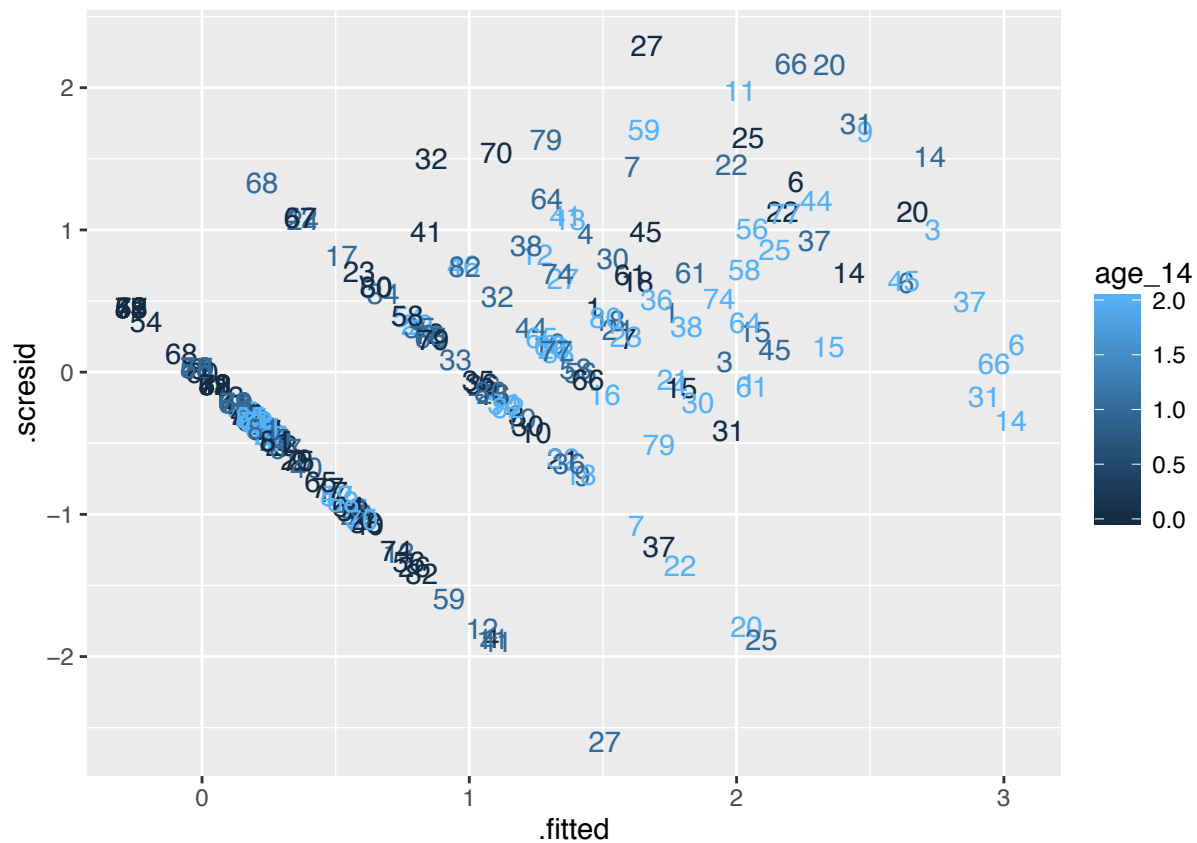
Homoskedasticity and linearity

```
#fitted vs residuals
#same as lme4 method -- plot(lmer2)
ggplot(modelf_diag, aes(x=.fitted, y=.scresid)) +
  geom_point() +
  geom_smooth()
```

Outliers

```
ggplot(modelf_diag, aes(x=.fitted,
                        y=.scresid, color=age_14)) +
  geom_text(aes(label=id))
```

Clustering of residuals

```r
#std residuals by Subject - most means near 0
ggplot(modelf_diag, aes(x=id, y=.scresid)) +
  geom_point() +
  stat_summary(color="red")
```

```
## No summary function supplied, defaulting to `mean_se()`
```