# Lecture 4.1
# Many-level models
# (where many=3)

Warning: This lecture may go beyond this class. Last year it spilled over, and then we shortened Lecture 4.2.

Three Level Models more important than crossed effects, so this is ok.

Update:
Section for S043
for Wednesdays 2-3:30 p.m.
is moved to
Longfellow 207
starting 10/23.

Directions for non-HGSE folks: Longfellow 207 can be accessed by taking the stairs nearest Garden St up to the 2nd floor. Turn right and go through the door into the office suite, then turn left to go down the hallway. Longfellow 207 will be on your right.

# Today's Goals

Introduce the idea of three-level data using a canonical case study of student growth when students are nested in schools.

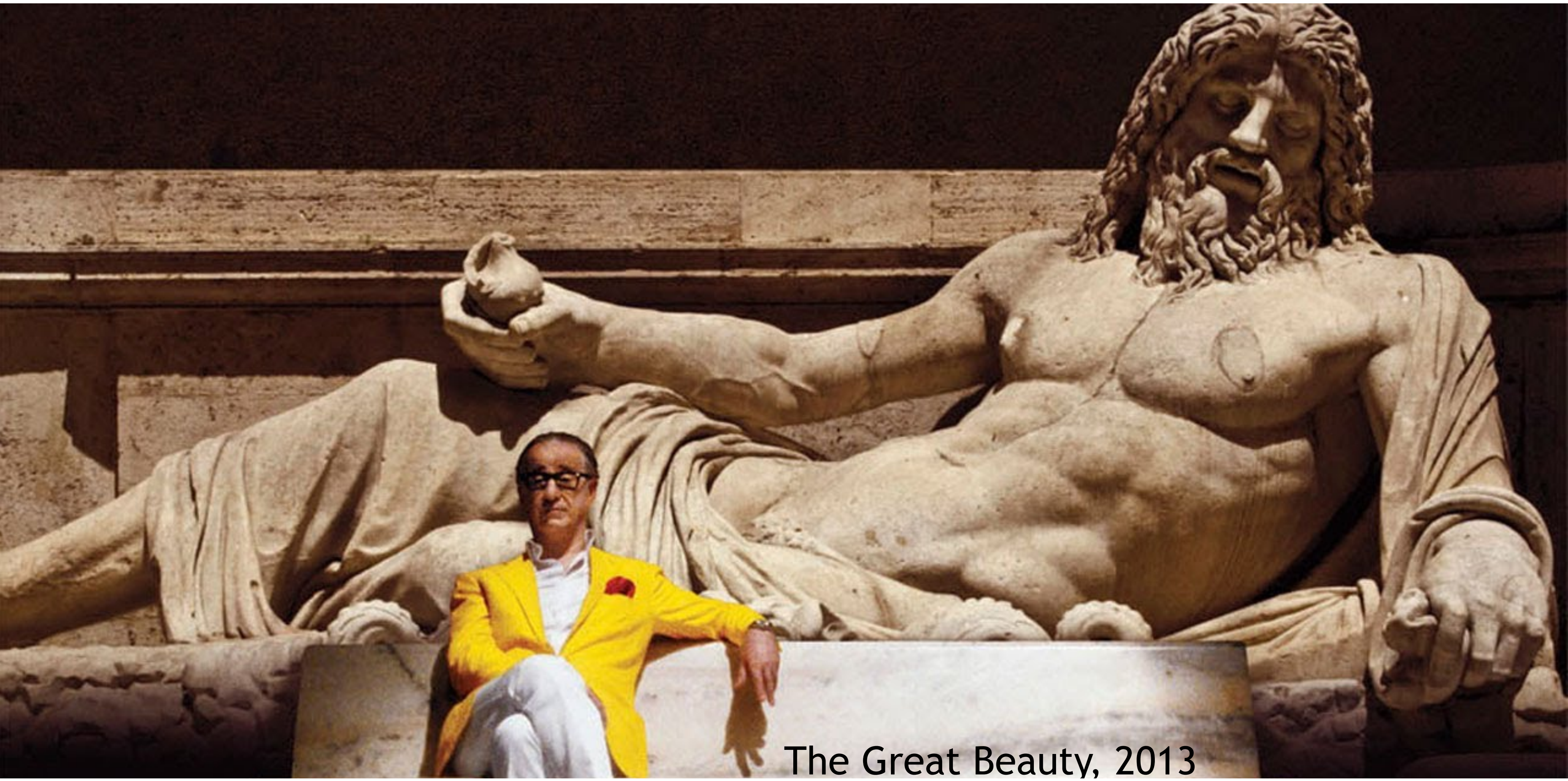
Learn how to fit these models

Learn how to add covariates to these models

Learn how to limit these models so they do not get too unwieldy.

Appendix: Some advice on fitting, etc., to look over at your leisure.

# An Important Special Case
# **Longitudinal data nested in a hierarchical structure**
(E.g. students growth for students in schools)



The Great Beauty, 2013

# READS Data

We are watching students across multiple time points. Outcome is reading ability. This is just like prior lectures.

We have student-level predictors (e.g., ELL status)

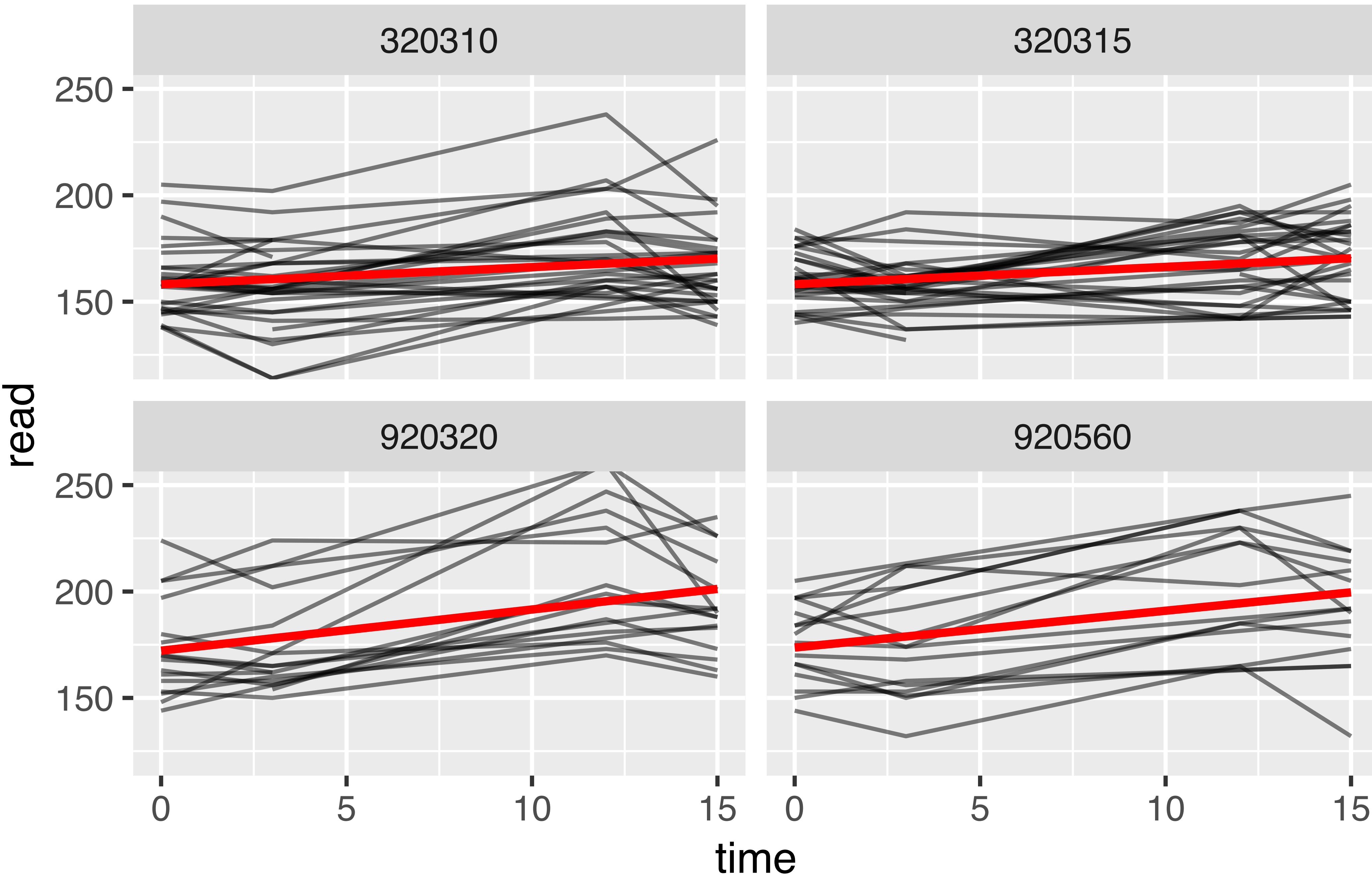
**BUT… t**hese students are inside schools!

We’ll also use school predictors, namely whether a school is identified as “high poverty” (i.e., a high proportion of students are eligible for free and reduced price lunches)

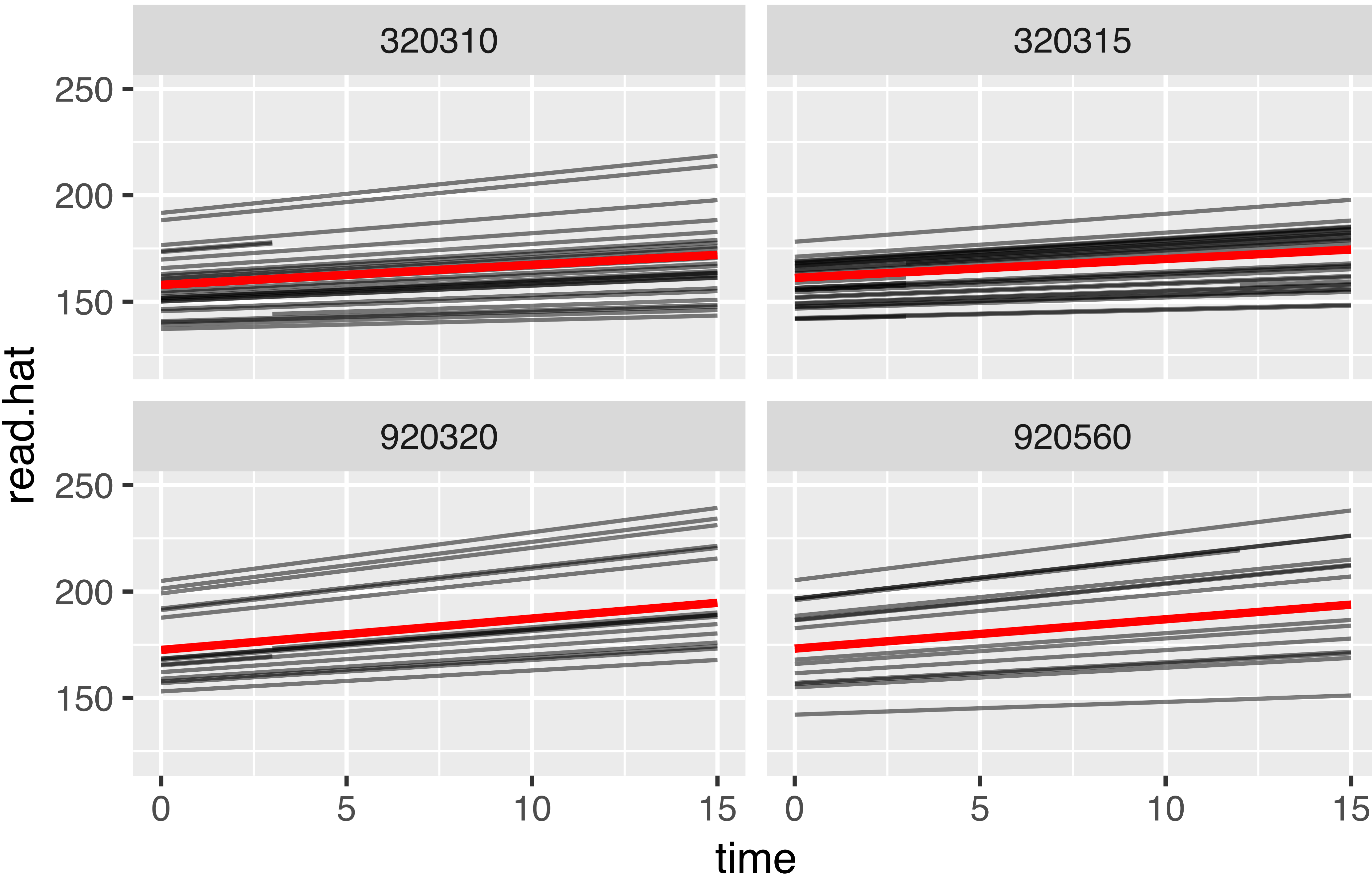
# A look at the data

```
> sample_n( dat, 15 )
             id    sch wave read gender poverty_school lep_new time
6624  CU-1012 260400   s4   NA      2        Poverty       0   15
765   DU-0552 320363   s3  192      1        Poverty       0    3
635   DU-0067 320304   s3  153      2        Poverty       0    3
10669 NH-0321 650309   f4  212      2    Not poverty       0   12
5060  RO-0005 780320   f3  190      2        Poverty       0    0
7542  NH-0246 650309   s4  198      1    Not poverty       0   15
6278  AS-0433 761321   s4  183      1        Poverty       0   15
8393  RO-0742 780384   s4  190      1    Not poverty       0   15
14    AS-0023 761312   s3  157      1        Poverty       1    3
11878 WA-0832 920336   f4   NA      2        Poverty       0   12
976   DU-1157 320339   s3  160      2        Poverty       1    3
8628  RO-1674 780412   s4  179      2        Poverty       0   15
6298  AS-0510 761332   s4  188      2    Not poverty       1   15
11667 RO-1467 780400   f4  168      2        Poverty       1   12
9523  CU-0462 260352   f4  168      2        Poverty       0   12
```

# Some students across 4 schools



# What our linear growth model says



# Some guiding research questions

RQ1: Do students at different schools grow at different rates, on average?

RQ2, RQ3: Do students identified as ELL grow at different rates than other students, and does that vary by school?

(These questions requires a student-level predictor.)

RQ4: Do students at high poverty schools have systematically different rates of growth than students at low poverty schools?

(This question requires a school-level predictor)

# Our three-level longitudinal linear growth model with school random effects

Time *t* for Student *i* in school *j* is modeled as:

$$read_{ijt} = \pi_{0ij} + \pi_{1ij} time_t + \epsilon_{ijt}$$

$$\left.\begin{aligned} \pi_{0ij} &= \beta_{00j} + u_{0ij} \\ \pi_{1ij} &= \beta_{10j} + u_{1ij} \end{aligned}\right\} \text{Level 2 has a bivariate normal error term}$$

$$\left.\begin{aligned} \beta_{00j} &= \gamma_{000} + r_{00j} \\ \beta_{10j} &= \gamma_{100} + r_{10j} \end{aligned}\right\} \text{Level 3 also has a separate bivariate normal error term}$$

Each student has its own slope and intercept.

These vary around each *school's* own average slope and average intercept.

# We now have lots and lots of random effects. We have stacked our models.

$$read_{ijt} = \pi_{0ij} + \pi_{1ij} time_t + \epsilon_{ijt}$$

Level 1

$$\pi_{0ij} = \beta_{00j} + u_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + u_{1ij}$$

Level 2

$$\beta_{00j} = \gamma_{000} + r_{00j}$$

$$\beta_{10j} = \gamma_{100} + r_{10j}$$

Level 3

Subscripts:
First is which level 1 coefficient
Second is which level 2 coefficient
Third is which level 3 coefficient

$$\epsilon_{ijt} \sim N(0, \sigma^2)$$

$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_0^2 & \sigma_0\sigma_1\rho \\ & \sigma_1^2 \end{pmatrix}\right]$$

$$\begin{pmatrix} r_{00j} \\ r_{10j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{10} \\ & \tau_{11} \end{pmatrix}\right]$$

Our two bivariate normal random effects, one set for student, one for school.

# What do the parameters mean?

$\beta_{00j}$ and $\beta_{10j}$ describe the average student *in school j*

$\gamma_{000}$ is the average student intercept *in the average school* (i.e., the average of $\beta_{00j}$ across schools).

$\gamma_{100}$ is the average student rate of growth *in the average school*

$u_{0ij}$ and $u_{1ij}$ are how student *ij* differs from the average student *in school j*

$r_{00j}$ and $r_{10j}$ are how school *j* differs from the average school.

# The variances and covariances are as before; we just have more of them

$$\epsilon_{ijt} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \overset{iid}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_0^2 & \sigma_0\sigma_1\rho \\ & \sigma_1^2 \end{pmatrix}\right]$$

$$\begin{pmatrix} r_{00j} \\ r_{10j} \end{pmatrix} \overset{iid}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{10} \\ & \tau_{11} \end{pmatrix}\right]$$

Notation is getting kind of hairy. We need all sorts of greek for our two covariance matrices. (It will get worse.)

# What are the parameters?

$\gamma_{000}, \gamma_{100}$: - fixed effects of average intercept and average slope

$\sigma^2$: residual variance (the degree of scatter of individual time points around individual student trends)

$\sigma_0^2, \sigma_1^2, \rho$: the variance of students' intercepts and slopes within their schools (and the correlation of these random effects)

$\tau_{00}, \tau_{11}, \tau_{01}$: the variance of schools' intercepts and slopes (and the covariance of these random effects)

Total of 9 parameters.

I used the tau notation in one level and the sigma notation in the other since otherwise we need, e.g., $\tau_{sch,00}, \tau_{sch,11}, \tau_{sch,01}$

# What is the reduced form?

Student's intercept is overall + school shift + student shift

Student's slope is, again, overall + school + student

$$
\begin{aligned}
read_{ijt} &= (\gamma_{000} + r_{00j} + u_{0ij}) + (\gamma_{100} + r_{10j} + u_{1jt})time_t + \epsilon_{ijt} \\
&= \gamma_{000} + \gamma_{100} time_t + [r_{00j} + u_{0ij} + (r_{10j} + u_{1jt})time_t + \epsilon_{ijt}]
\end{aligned}
$$

Overall average line

Lots of residual terms in brackets
(These average to zero across everything)

For a specific school j we have

$$
read_{ijt} = (\gamma_{000} + r_{00j}) + (\gamma_{100} + r_{10j})time_t + [u_{0ij} + u_{1jt} time_t + \epsilon_{ijt}]
$$

Overall regression for school j

Residual terms

# In R

```
lmer(read ~ time + (time|id) + (time|schid), data = dat)
```

Outcome is reading

Regressed on time

These are separate, so the random effects are independent

Time slope and intercept varies by student

Time slope and intercept also vary by school

# Note on R notation (two equivalent models)

```
# We can write this:
M2 <- lmer(read ~ 1 + time
              + (1 | id/schid), data = dat)


# Or we can write this
M2b <- lmer(read ~ 1 + time

             + (1 | id:schid)

             + (1|schid),

             data = lmm.data)
```

This is a random student intercept. The ":" makes sure you have a unique student ID for each student by school combo.

This is a random school intercept

Second way is better, and aligns with reduced form of model when written mathematically.

# Random R Tip / Concept Check

If you have students with unique IDs even across schools (e.g., no student ID is shared by two students in different schools) you don't need to say "student:school"

You can just say "student"

What R is doing is giving each student its own random effect, independent of school and time.

So you just need to count up the different students. The schools are irrelevant!

# Results

```
Model 1
----------------------------------------
(Intercept)                  166.72 ***
                              (0.60)
time                           1.32 ***
                              (0.03)
----------------------------------------
Num. obs.                  10730
Num. groups: id             3061
Num. groups: schid            60
Var: id (Intercept)          236.04
Var: id time                   0.45
Cov: id (Intercept) time       5.60
Var: schid (Intercept)        15.21
Var: schid time                0.04
Cov: schid (Intercept) time    0.46
Var: Residual                 98.87
```

$\gamma_{000}$

$\gamma_{100}$

Variances and covariances; profile confidence intervals show that all of these are statistically significant.

Take the square root of a variance to get a standard deviation

Take a covariance divided by the two standard deviations to get a correlation

RQ1: Do students at different schools grow at different rates, on average?

# Positive correlations

Notice that intercepts and slopes are positively correlated at both the student and school level.

Schools where students started higher had faster average growth; students who started higher also had faster average growth, even beyond being in schools that had higher average growth.

# Our model (repeated slide)

$$read_{ijt} = \pi_{0ij} + \pi_{1ij} time_t + \epsilon_{ijt}$$
$$\pi_{0ij} = \beta_{00j} + u_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + u_{1ij}$$
$$\beta_{00j} = \gamma_{000} + r_{00j}$$
$$\beta_{10j} = \gamma_{100} + r_{10j}$$
$$\epsilon_{ijt} \sim N(0, \sigma^2)$$
$$\begin{pmatrix} u_{0ij} \\ u_{1ij} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_0^2 & \sigma_0 \sigma_1 \rho \\ & \sigma_1^2 \end{pmatrix} \right]$$
$$\begin{pmatrix} r_{00j} \\ r_{10j} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{10} \\ & \tau_{11} \end{pmatrix} \right]$$

# Quick quiz: using LR tests

Our original model

```
lmer(read ~ 1 + time + (1 + time|id) + (1 + time|schid), data = dat)
```

1) How would we test a null-hypothesis that the average student in the average school does not grow over time?

```
lmer(read ~ 1 + (1 + time|id) + (1 + time|schid), data = dat)
```

2) How would we test a null-hypothesis that every student in a given school has the same rate of growth?

```
lmer(read ~ 1 + time + (1|id) + (1 + time|schid), data = dat)
```

3) How would we test a null-hypothesis that every school has the same mean rate of growth?

```
lmer(read ~ 1 + time + (1 + time|id) + (1|schid), data = dat)
```

# Finishing Lecture 4.1

Goals:

Level 2 covariates!

Level 3 covariates!

Recap!

# Adding Level 2 Covariates

http://graphicjoke.fiveforks.com/a-goat-walks-into-a-b

# Including student-level ELL status

ELL is a student-level variable (time-invariant). It indicates that a student is identified as an **E**nglish **L**anguage **L**earner

We can include it at the student level, and make it random at the school level, as in the following model:

Ok, the model is TOO BIG. See next slide.

# The model with ELL status

$$read_{ijt} = \pi_{0ij} + \pi_{1ij} time_t + \epsilon_{ijt}$$

Aside:
Why triple index on the gammas? In case we wanted level 3 (school-level) covariates

$$\pi_{0ij} = \beta_{00j} + \beta_{01j} ELL_i + u_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + \beta_{11j} ELL_i + u_{1ij}$$

$\beta_{00j} = \gamma_{000} + r_{00j}$ school average reading score

$\beta_{01j} = \gamma_{010} + r_{01j}$ school-specific ELL initial gap

$\beta_{10j} = \gamma_{100} + r_{10j}$ school average growth rate

$\beta_{11j} = \gamma_{110} + r_{11j}$ school-specific ELL/non-ELL growth rate difference

Variance-covariance matrix is 4X4 with 4 variances and 6 covariances. We've added two regression coefficients, two variances, and five covariances; our model has gotten a *lot* more complex, mostly from parameters we don't care about

# What's that model? It directly connects to math

ell predicts both initial status and rate of growth, which means the terms interact.

```
ell_mod <- lmer(read ~ 1 + ell + time + time:ell +
                (1 + time|id) +
                (1 + ell + time + time:ell|schid), data =
```

Rate of growth, *but not ell difference* varies across students (why?)

*Everything* varies across schools. This explodes into 4 different random effects at the school level.

$$\begin{aligned} read_{tij} = {} & \gamma_{000} + \gamma_{010}ELL_{ij} + \gamma_{100}time_{tij} + \gamma_{110}time_{tij}ELL_{ij} \\ & + u_{0ij} + u_{1ij}time_{tij} \\ & + r_{00j} + r_{01j}ELL_{ij} + r_{10j}time_{tip} + r_{11j}time_{tij}ELL_{ij} \\ & + \varepsilon_{tij} \end{aligned}$$

# R Syntax Reminder

**`time * ell_new`**

`is`

**`1 + time * ell_new`**

`is`

**`1 + time + ell_new + time:ell_new`**

# Fitted model

| | |
|---|---|
| (Intercept) $\gamma_{000}$ | 168.31 *** |
| | (0.61) |
| time $\gamma_{100}$ | 1.33 *** |
| | (0.04) |
| ell_new $\gamma_{010}$ | -9.59 *** |
| | (0.92) |
| time:ell_new $\gamma_{110}$ | -0.06 |
| | (0.07) |
| Var: id (Intercept) | 224.37 |
| Var: id time | 0.44 |
| Cov: id (Intercept) time | 5.47 |
| Var: schid (Intercept) | 14.46 |
| Var: schid time | 0.04 |
| Var: schid ell_new | 5.89 |
| Var: schid time:ell_new | 0.05 |
| Cov: schid (Intercept) time | 0.44 |
| Cov: schid (Intercept) ell_new | -5.59 |
| Cov: schid (Intercept) time:ell_new | 0.06 |
| Cov: schid time ell_new | -0.44 |
| Cov: schid time time:ell_new | -0.02 |
| Cov: schid ell_new time:ell_new | 0.40 |
| Var: Residual | 99.14 |

ell students begin lower than their peers (about 8 months behind; $\gamma_{010}$)

But appear to grow at similar rates ($\gamma_{110}$ )

Note SE is about same size as estimate. Not impressive. No evidence it is non-zero

RQ2: Do students identified as ELL grow at different rates than other students?

# Unpacking the variance components

```
Var: id (Intercept)                     224.37 u0ij
Var: id time                              0.44 u1ij
Cov: id (Intercept) time                  5.47 cov( u0ij, u1ij)
```

The variance of the $u_{0ij}$, the variance of $u_{1ij}$, the covariance of these two

```
Var: schid (Intercept)                   14.46 r00j
Var: schid time                           0.04 r10j
Var: schid ell_new                        5.89 r01j
Var: schid time:ell_new                   0.05 r11j
```

RQ3:
Does ELL growth rate gap differ by school

The variance of the $r_{00j}$, $r_{10j}$, $r_{01j}$, and $r_{11j}$

```
Cov: schid (Intercept) time               0.44 cov(r00j, r10j)
Cov: schid (Intercept) ell_new           -5.59
Cov: schid (Intercept) time:ell_new       0.06
Cov: schid time ell_new                  -0.44
Cov: schid time time:ell_new             -0.02
Cov: schid ell_new time:ell_new           0.40
```

The covariances of all the various pairs of $r_{00j}$, $r_{10j}$, $r_{01j}$, and $r_{11j}$

```
Var: Residual                            99.14
```

The variance of the $\varepsilon_{tij}$

# Testing whether $\gamma_{110} \neq 0$

I.e., testing whether ELL status is associated with growth rate. on average

**Wald:** First way to check is a Wald test; we did that already and failed to reject. Since this is a regression coefficient (fixed effect), we can probably trust this result.

**Confidence Interval:** Making a profile confidence interval (`confint(ell_mod)`) takes more than 30 minutes (that's when I got bored and quit). Might be quicker if you're only looking at that one CI, rather than CIs for all parameters

**LRT:** The reduced model would be

```
ell_red <- lmer(read ~ time + ell_new +
                  (time|id) +
                  (time * ell_new|schid), data = dat)
```

Use `anova(ell_mod, ell_red)` to compare them.

NOTE: What is weird about what we are testing, given the model?

# R(Studio) like a pro

Sometimes things take a long time in RStudio. Argh!

You can start another copy of RStudio so you can do other things while your models are converging.

On a PC, right-click on RStudio to launch another copy. Yay!

On Mac, “New Session” under “Session” menu

# Adding Level 3 Covariates

(Thanks to Siobhan for finding this picture.)

# Including a level-3 predictor

poverty_school is a school-level variable

We can include it at the school level as in the following model:

If our last model didn't fit, what makes you think this one will?

# Model with poverty_school

$$read_{ijt} = \pi_{0ij} + \pi_{1ij} time_t + \epsilon_{ijt}$$

$$\pi_{0ij} = \beta_{00j} + \beta_{01j} ELL_i + u_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + \beta_{11j} ELL_i + u_{1ij}$$

$$\beta_{00j} = \gamma_{000} + \gamma_{001} pov_j + r_{00j}$$
Does poverty of school predict… average reading?

$$\beta_{01j} = \gamma_{010} + \gamma_{011} pov_j + r_{01j}$$
the ELL gap?

$$\beta_{10j} = \gamma_{100} + \gamma_{101} pov_j + r_{10j}$$
average growth for non-ELL?

$$\beta_{11j} = \gamma_{110} + \gamma_{111} pov_j + r_{11j}$$
difference in ELL and non-ELL growth?

We've added four new regression coefficients.

# What’s that model?

poverty_school predicts both initial status, rate of growth, ell “effect” and the ell*time interaction.  That’s a three-way interaction!

```
pov_mod <- lmer(read ~ time * ell_new * poverty_school +
                (time|id) +
                (time * ell_new|schid), data = dat)
```

poverty_school is measured at the top level, so it can’t be random at any level; it doesn’t vary within any groups!

**<u>Note on interaction expansion:</u>**

```
time*ell_new*poverty_school turns in to
time + ell_new + poverty_school + time:ell_new +
time:poverty_school + ell_new:poverty_school +
time:ell_new:poverty_school
```

# Fitted model (regression coefficients only)

| | Model 1 |
|---|---|
| (Intercept) $\gamma_{000}$ | 167.09 *** |
| time $\gamma_{100}$ | 1.26 *** |
| | (0.04) |
| ell_new $\gamma_{010}$ | -8.44 *** |
| | (1.04) |
| poverty_schoollow poverty $\gamma_{001}$ | 3.91 ** |
| | (1.20) |
| time:ell_new $\gamma_{110}$ | -0.01 |
| | (0.08) |
| time:poverty_schoollow poverty $\gamma_{101}$ | 0.23 *** |
| | (0.07) |
| ell_new:poverty_schoollow poverty $\gamma_{011}$ | -3.90 |
| | (2.07) |
| time:ell_new:poverty_schoollow poverty $\gamma_{111}$ | -0.15 |
| | (0.15) |

Non-ELL students start higher and grow faster in wealthy schools

# Answering our research question

RQ4: Do students at high poverty schools have systematically different rates of growth than students at low poverty schools?

```
poverty_schoollow poverty                        3.91 **
                                                (1.20)
time:poverty_schoollow poverty                   0.23 ***
                                                (0.07)
ell_new:poverty_schoollow poverty               -3.90
                                                (2.07)
time:ell_new:poverty_schoollow poverty          -0.15
                                                (0.15)
```

Note: We could have simplified by removing ELL from our model to get average growth across all students in our schools.

# So complex! Reining it in.

# Wow!

These models get out of hand *really quickly!*

Fitting a piecewise model with a separate slope for school time and summer time would mean adding four new regression coefficients, three new variances, and 11 new correlations. That's a lot!

Adding student- or school-level predictors would also make the model far more complex

Adding a district-level and district-level predictors would make the model far, far, far more complex

# Being parsimonious

Sometimes there are empirical restrictions on the questions we can ask.

(To keep your sanity) try

1. Reducing the complexity of the question by simplifying the fixed part of the model
2. Completely pooling some associations, simplifying the random part (or parts) of the model

# Pooling coefficients

Our model from the last section:

```
pov_mod <- lmer(read ~ time * ell_new * poverty_school +
                  (1 + time|id) +
                  (1 + time * ell_new|schid),
                   data = dat)
```

If we assume that there are no **school-level differences** in growth rates, the ELL “effect,” or their interaction we can get the even simpler

```
pov_mod <- lmer(read ~ time * ell_new * poverty_school +
                  (1 + time|id) +
                  (1|schid), data = dat)
```

# Fitted model (regression coefficients)

```
                                          Model 1
-----------------------------------------------------
(Intercept)                               167.09 ***
time                                        1.26 ***
                                           (0.04)
ell_new                                    -8.44 ***
                                           (1.04)
poverty_schoollow poverty                   3.91 **
                                           (1.20)
time:ell_new                               -0.01
                                           (0.08)
time:poverty_schoollow poverty              0.23 ***
                                           (0.07)
ell_new:poverty_schoollow poverty          -3.90
                                           (2.07)
time:ell_new:poverty_schoollow poverty     -0.15
                                           (0.15)
```

How else can we simplify the model, and is it a good idea?

# Consider this more reduced model

```
pov_mod <- lmer(read ~ 1 + time + ell_new +
                  poverty_school +
                  time:poverty_school +
                  (1 + time|id) +
                  (1 + time + ell_new|schid),
                  data = dat)
```

5 fixed effects (the gammas)

2 random terms/student

3 random terms/school

What is this model, mathematically?

# What's that model?

$$read_{ijt} = \pi_{0ij} + \pi_{1ij} time_t + \epsilon_{ijt}$$

$$\epsilon_{int} \sim N(0, \sigma^2)$$

$$\pi_{0ij} = \beta_{00j} + \beta_{01j} ELL_i + u_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + u_{1ij}$$

ell no longer predicts rate of growth

$$(u_{0ij}, u_{1ij}) \sim N(0, \Sigma)$$

2x2 Matrix

$$\beta_{00j} = \gamma_{000} + \gamma_{001} pov_j + r_{00j}$$

$$\beta_{01j} = \gamma_{010} + r_{01j}$$

School poverty no longer predicts the ell "effect"

$$\beta_{10j} = \gamma_{100} + \gamma_{101} pov_j + r_{10j}$$

There's no more $\beta_{11j}$ to model

$$(r_{00j}, r_{01j}, r_{10j}) \sim N(0, \tau)$$

3x3 Matrix

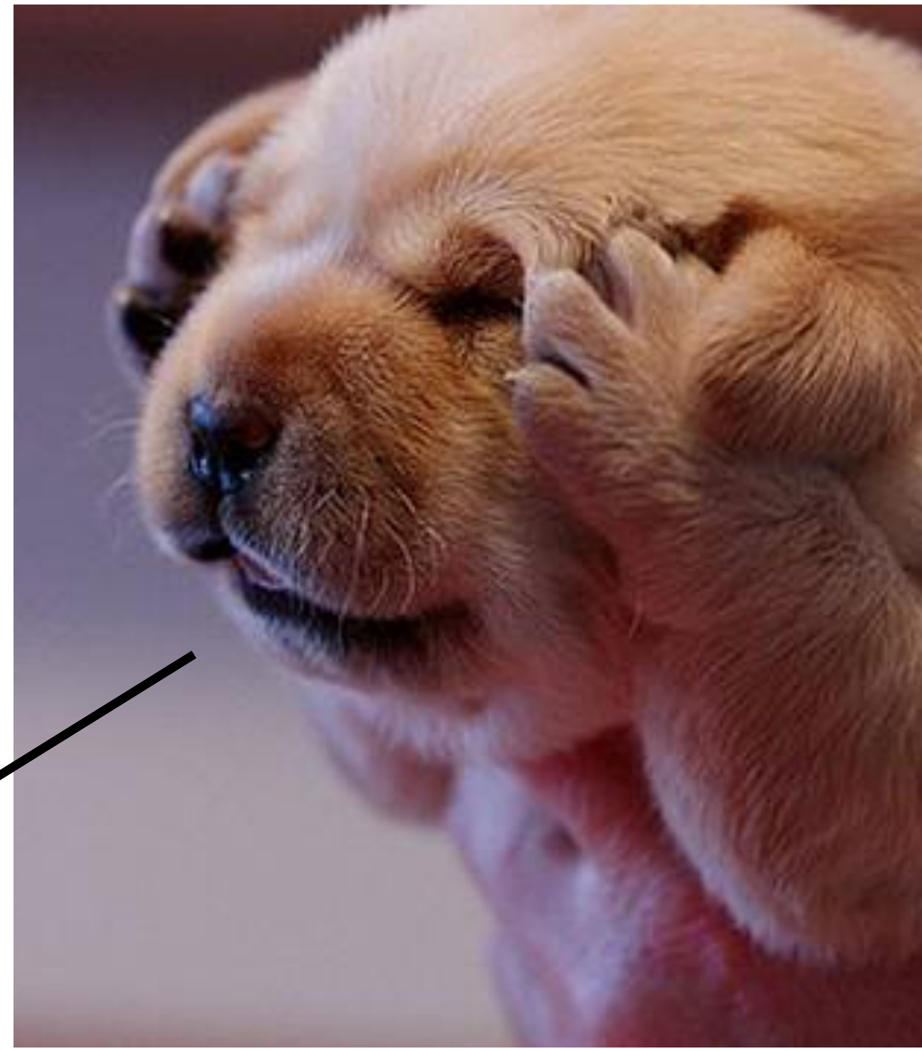
# Neat!

Look at all the cross-level questions we can ask and answer with these models!

So exciting!

Wonderful!

But a bit overwhelming.



# Recap on three level models

http://www.storytrender.com/39441/standing-elephant-stands-back-legs-reach-tree-top-leaves/

# Recap

http://cs179.org/lec41

A core data structure, growth of students in schools, is a three level model.

Three level models require you to specify equations one level up for each parameter in each level.

This can cause a lot of parameters to appear, and make fitting and interpretation hard.

The only way to survive is to impose structure, throwing away fixed and random effects so you don't get a huge blow-up of stuff.

# Appendix:
# Some practical advice

# Deciding what to pool

Pool things which are:

1. Higher level (these are harder to estimate)
2. Less likely to vary
3. Less interesting substantively

Model things which are:

1. Lower level
2. More likely to vary (intercepts!)
3. Interesting for how they vary

You can't model every part of the model; no models are true, but some models are interesting (and justifiable!)

# Failure to converge and you

Ah, for those halcyon days of OLS!

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Algebra (and matrix multiplication) was enough to estimate the model coefficients!

But MLE require iterative fitting! Argh!

# What can go wrong?

1. Sometimes the likelihood surface is too bumpy to find a maximum
2. Sometimes our path across our likelihood surface will get stuck at an edge (e.g., $\hat{\sigma}=0$)
3. Sometimes we'll enter something as both a random effect and a fixed effect without noticing
4. Sometimes we won't have enough observations to estimate models within each higher unit

# What to do

**Check your model and data for coding issues**

Make sure that the model is doing what you intend, and that they data are correct

# What to do

**Try centering variables within higher level units**

This can break artificially extreme correlations, which makes model estimation easier.

# What to do

**Play around with the lmer control parameters**

Try changing the optimizer (how lmer tries to find the top of the likelihood function), and maxit (how long lmer is willing to look)

SEE HANDOUT ON CONVERGENCE FOR THIS

# What to do

**Simplify, simplify, simplify**

You might not be able to estimate the model you want. Consider complete pooling for some of the parameters.

# Final Comments

1. Three-level models are very complex.
2. They are very unstable. Separating out which level gets what variance is not easy.
3. Consider restricting the slopes that are random.
4. Random intercept models will often be the best you can do.