# R Notebook

Neil Yetz & Gemma Wallace

## Load Libraries

```
library(psych)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  3.0.0     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## Warning: package 'ggplot2' was built under R version 3.6.3
## Warning: package 'tibble' was built under R version 3.6.3
## Warning: package 'tidyr' was built under R version 3.6.3
## Warning: package 'dplyr' was built under R version 3.6.3
## Warning: package 'forcats' was built under R version 3.6.3
## -- Conflicts ------------------------------------------------------------------------------------
## x ggplot2::%+%()   masks psych::%+%()
## x ggplot2::alpha() masks psych::alpha()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(lme4)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
library(lmerTest)
```

```
##
## Attaching package: 'lmerTest'
## The following object is masked from 'package:lme4':
##
##     lmer
## The following object is masked from 'package:stats':
##
```

```
##     step
```

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 3.6.3
```

```
library(modelr)
```

```
## Warning: package 'modelr' was built under R version 3.6.3
```

```
##
## Attaching package: 'modelr'
```

```
## The following object is masked from 'package:broom':
##
##     bootstrap
```

## class demo dataset

```
#Wide dataset will convert to long below
wide <- read_csv("mlm_grow.csv")
```

```
## Parsed with column specification:
## cols(
##   kid_id = col_double(),
##   team_id = col_double(),
##   txcond = col_double(),
##   perf_0 = col_double(),
##   perf_1 = col_double(),
##   perf_2 = col_double(),
##   perf_3 = col_double(),
##   perf_4 = col_double(),
##   perf_5 = col_double(),
##   perf_6 = col_double(),
##   inter_0 = col_double(),
##   inter_1 = col_double(),
##   inter_2 = col_double(),
##   inter_3 = col_double(),
##   inter_4 = col_double(),
##   inter_5 = col_double(),
##   inter_6 = col_double()
## )
```

## Format and output class demo

```
###### Stolen from Kim Start #######
long_perf <- wide %>%
  select(kid_id, txcond, perf_0:perf_6) %>%
  gather(key = measure, value = perform, perf_0:perf_6) %>%
  separate(measure, c("temp", "week")) %>%
  select(-temp) %>%
  mutate(week = as.numeric(week))


long_inter <- wide %>%
```

```
  select(kid_id, inter_0:inter_6) %>%
  gather(key = measure, value = interest, inter_0:inter_6) %>%
  separate(measure, c("temp", "week")) %>%
  select(-temp) %>%
  mutate(week = as.numeric(week))

grow <- full_join(long_perf, long_inter, by = c("kid_id", "week"))

grow <- grow %>%
  mutate(txcond.f = factor(txcond, levels = c(0,1), labels = c("control", "treatment")),
  kid_id.f = factor(kid_id))

###### Stolen From Kim End ######

# Simplify dataset
grow <- select(grow, kid_id, week, perform)
# Write to csv
#write_csv(grow, "grow.csv", na = "")
```

# Start demo

# Read in Data

```
grow <- read_csv("grow.csv")
```

```
## Parsed with column specification:
## cols(
##   kid_id = col_double(),
##   week = col_double(),
##   perform = col_double()
## )
```

# Describe data

```
describe(grow)
```

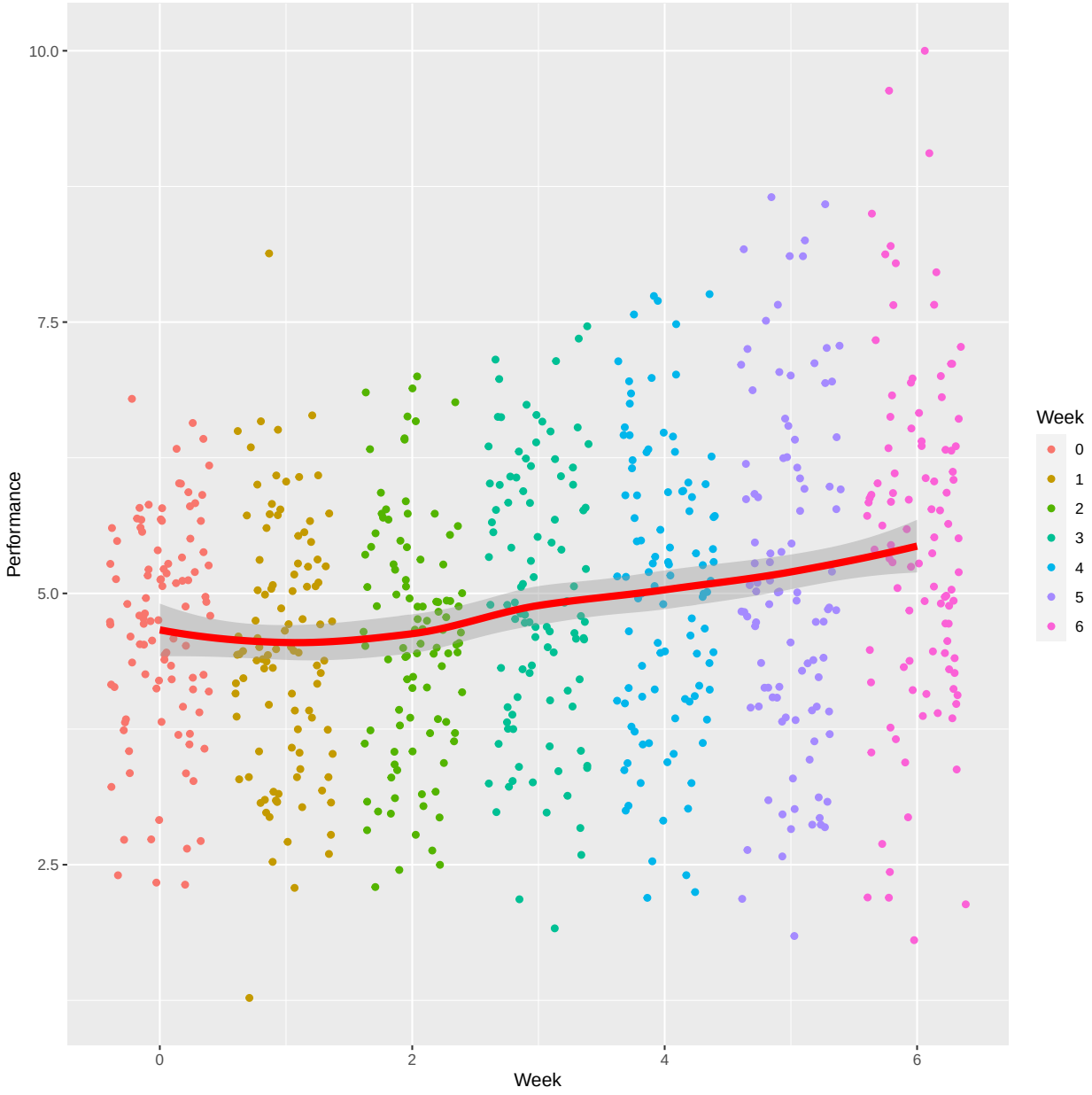
```
##         vars   n  mean     sd median trimmed    mad  min max  range skew
## kid_id     1 700 252.5 144.43 252.50  252.50 185.32 5.00 500 495.00 0.00
## week       2 700   3.0   2.00   3.00    3.00   2.97 0.00   6   6.00 0.00
## perform    3 700   4.9   1.33   4.88    4.87   1.32 1.27  10   8.73 0.29
##         kurtosis   se
## kid_id     -1.21 5.46
## week       -1.25 0.08
## perform     0.20 0.05
```

# Visualize the data

```
ggplot(grow, aes(x = week, y = perform)) +
  geom_jitter(aes(color = factor(week))) +
  geom_smooth(method = "loess", color = "red", size = 2) +
```

```
  xlab("Week") +
  ylab("Performance") +
  labs(color = "Week")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




## Baseline model

```
mod1 <-  lmer(perform ~ 1 + (1|kid_id), REML = TRUE, data = grow)
summary(mod1)
```

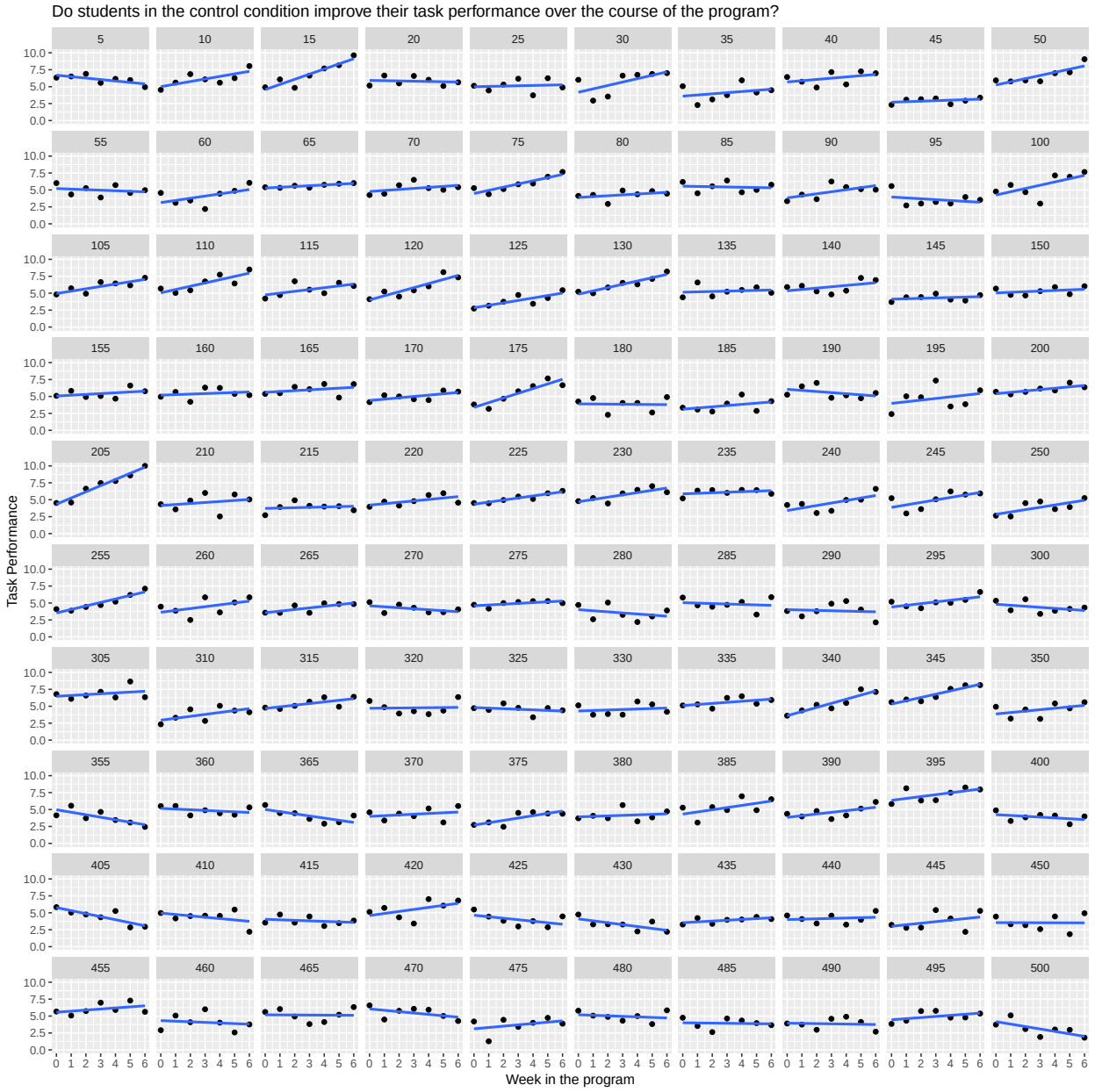
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
```

```
## Formula: perform ~ 1 + (1 | kid_id)
##    Data: grow
##
## REML criterion at convergence: 2166.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6469 -0.6424 -0.0085  0.5990  3.2746
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  kid_id   (Intercept) 0.8013   0.8951
##  Residual             0.9821   0.9910
## Number of obs: 700, groups:  kid_id, 100
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  4.90455    0.09703 98.99999   50.55   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Separated by subject

```
ggplot(data = grow, aes(x = week, y = perform)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  scale_y_continuous(limits = c(0,10)) +
  scale_x_continuous(limits = c(0,6), breaks = c(0,1,2,3,4,5,6)) +
  facet_wrap(~kid_id) +
  labs(title = "Do students in the control condition improve their task performance over the course of t
    x = "Week in the program", y = "Task Performance")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



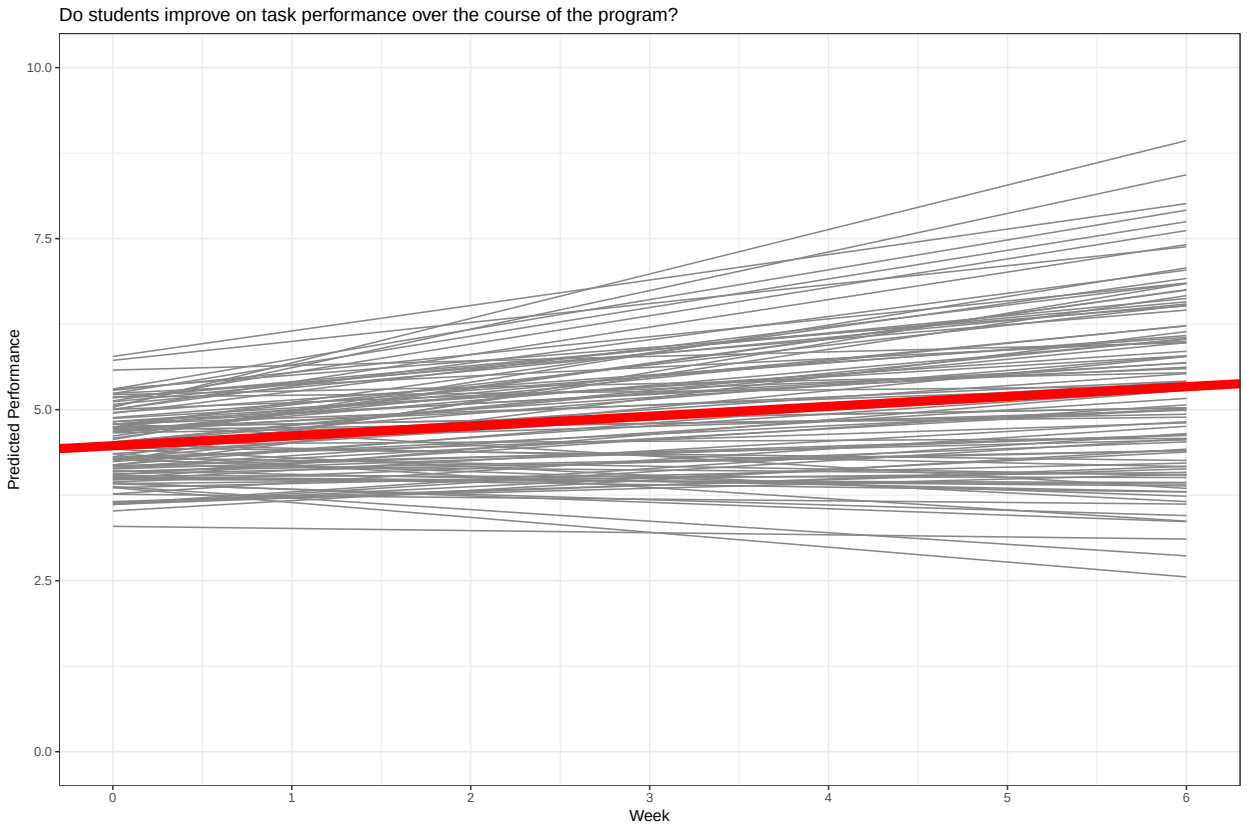

```
mod2 <-  lmer(perform ~ 1 + week + (1 + week|kid_id), REML = TRUE, data = grow)
summary(mod2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: perform ~ 1 + week + (1 + week | kid_id)
##    Data: grow
##
## REML criterion at convergence: 2038.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1997 -0.5740  0.0449  0.6342  3.0608
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  kid_id   (Intercept) 0.42532  0.6522
##           week        0.03774  0.1943   0.10
##  Residual             0.71247  0.8441
## Number of obs: 700, groups:  kid_id, 100
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  4.47310    0.08695 98.99841  51.442  < 2e-16 ***
## week         0.14382    0.02514 99.00129   5.722 1.13e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## week -0.292
```

```r
# Get predicted values
mod2.plot <- add_predictions(data = grow, model = mod2)

# Make plot
ggplot(data = mod2.plot, aes(x = week, y = pred, group = kid_id)) +
  geom_line(color = "grey53") +
  geom_abline(intercept = 4.4731, slope = .1438, color="red", size=3) +
  scale_y_continuous(limits = c(0,10)) +
  scale_x_continuous(limits = c(0,6), breaks = c(0,1,2,3,4,5,6)) +
  labs(title = "Do students improve on task performance over the course of the program?",
    x = "Week", y = "Predicted Performance") +
  theme_bw()
```




# Assignment

# read in data

```
long <- read_csv("Longitudinal.csv")
```

```
## Parsed with column specification:
## cols(
##   student = col_double(),
##   occas = col_double(),
##   gpa = col_double(),
##   job = col_double(),
##   sex = col_double()
## )
```

# Describe data

```
describe(long)
```

```
##         vars    n   mean    sd median trimmed   mad min max range  skew
## student    1 1200 100.50 57.76  100.5  100.50 74.13 1.0 200 199.0  0.00
## occas      2 1200   2.50  1.71    2.5    2.50  2.22 0.0   5   5.0  0.00
## gpa        3 1200   2.87  0.39    2.8    2.86  0.44 1.7   4   2.3  0.18
## job        4 1200   0.50  0.50    0.5    0.50  0.74 0.0   1   1.0  0.00
## sex        5 1200   0.52  0.50    1.0    0.53  0.00 0.0   1   1.0 -0.10
```

```
##         kurtosis   se
## student    -1.20 1.67
## occas      -1.27 0.05
## gpa        -0.02 0.01
## job        -2.00 0.01
## sex        -1.99 0.01
```

## Factor variables

```
long <- mutate(long,
               job = factor(job, levels = c(0,1), labels = c("half time", "full time")),
               sex = factor(sex, levels = c(0,1), labels = c("male", "female"        ))
               )
```

## Create occas squared

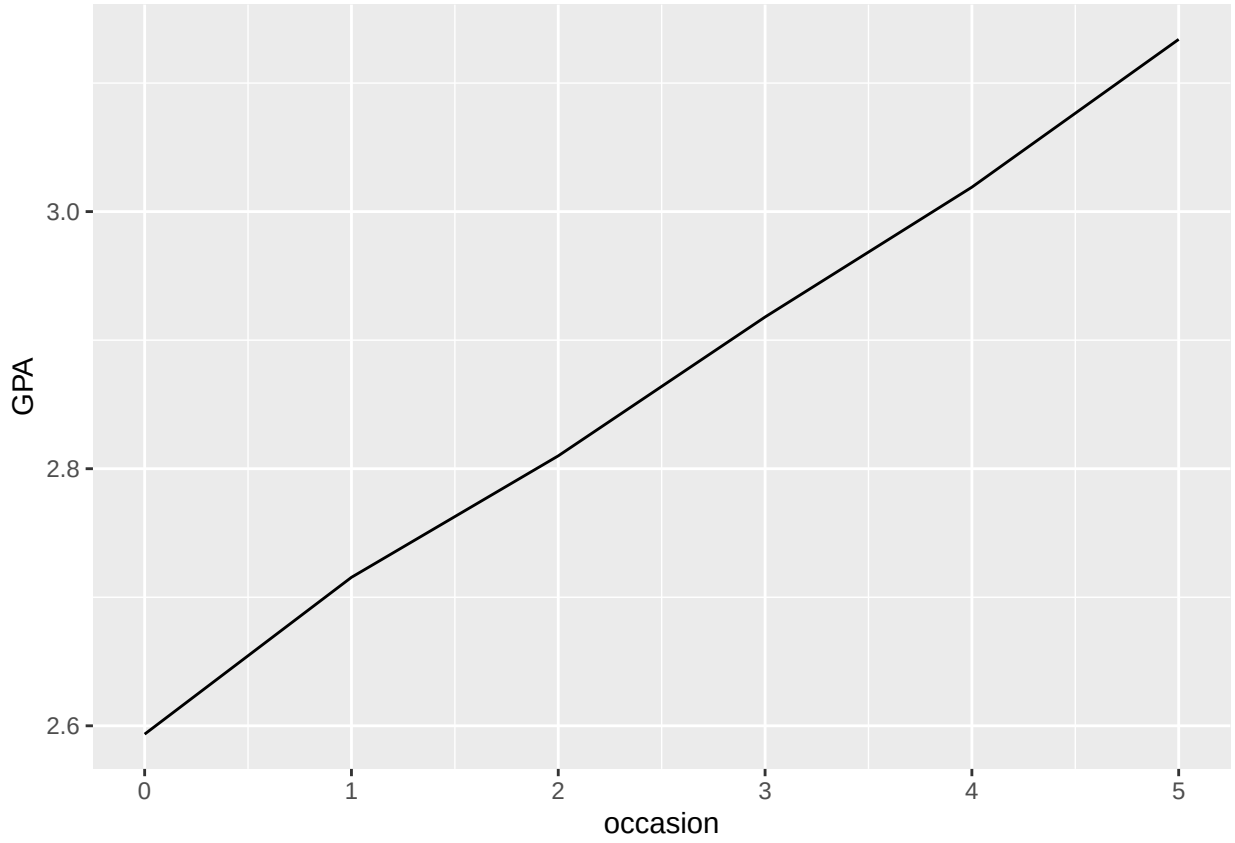
```
long <- mutate(long,
               occassq = occas**2)
```

## Aggregate data

```
agg_long <- aggregate(x=long$gpa,by=list(occasion = long$occas), FUN=mean)
agg_long
```

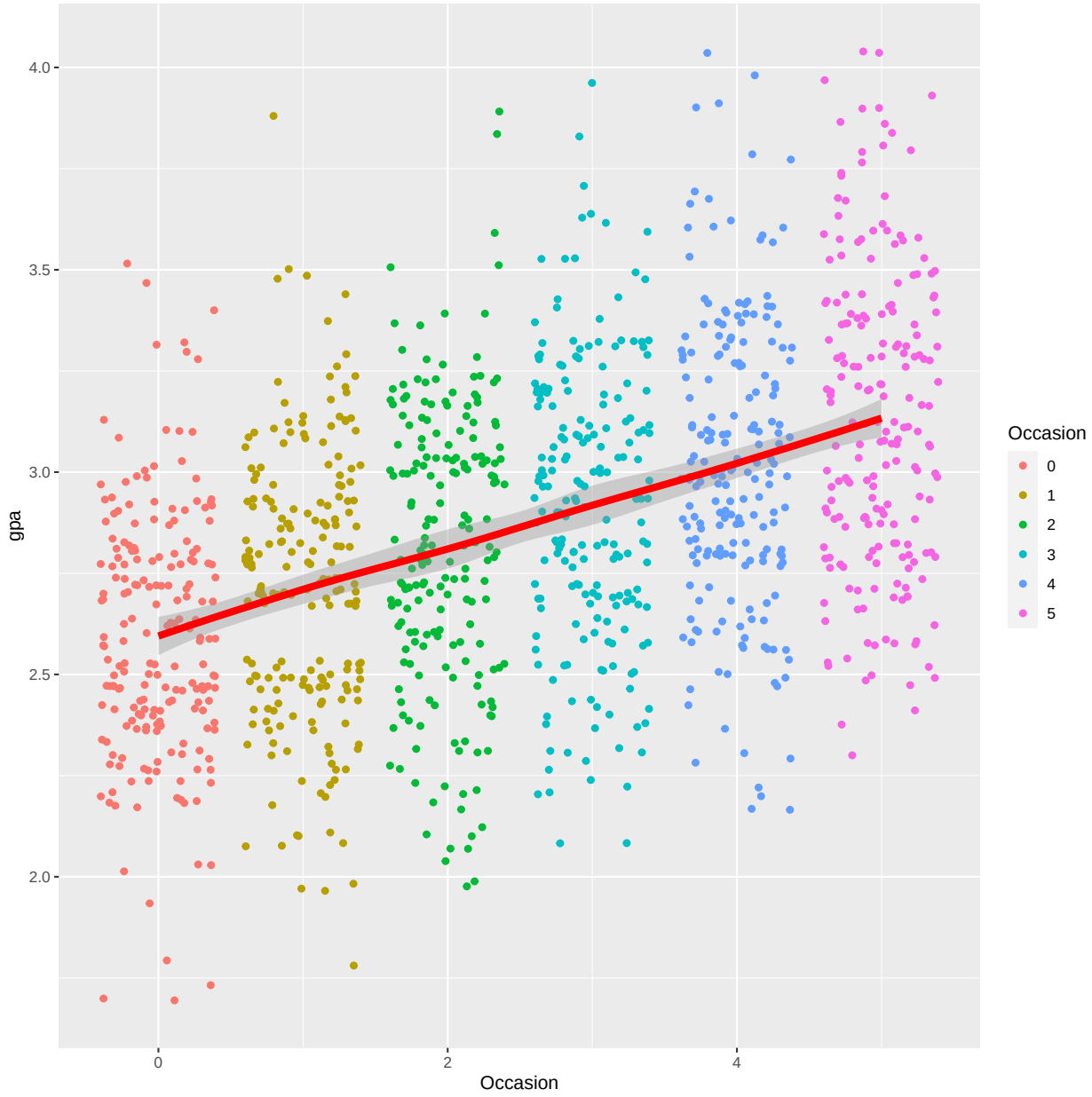
```
##   occasion      x
## 1        0 2.5935
## 2        1 2.7155
## 3        2 2.8100
## 4        3 2.9180
## 5        4 3.0190
## 6        5 3.1340
```

```
ggplot(agg_long, aes(x = occasion, y = x)) +
  geom_line() +
  ylab("GPA")
```



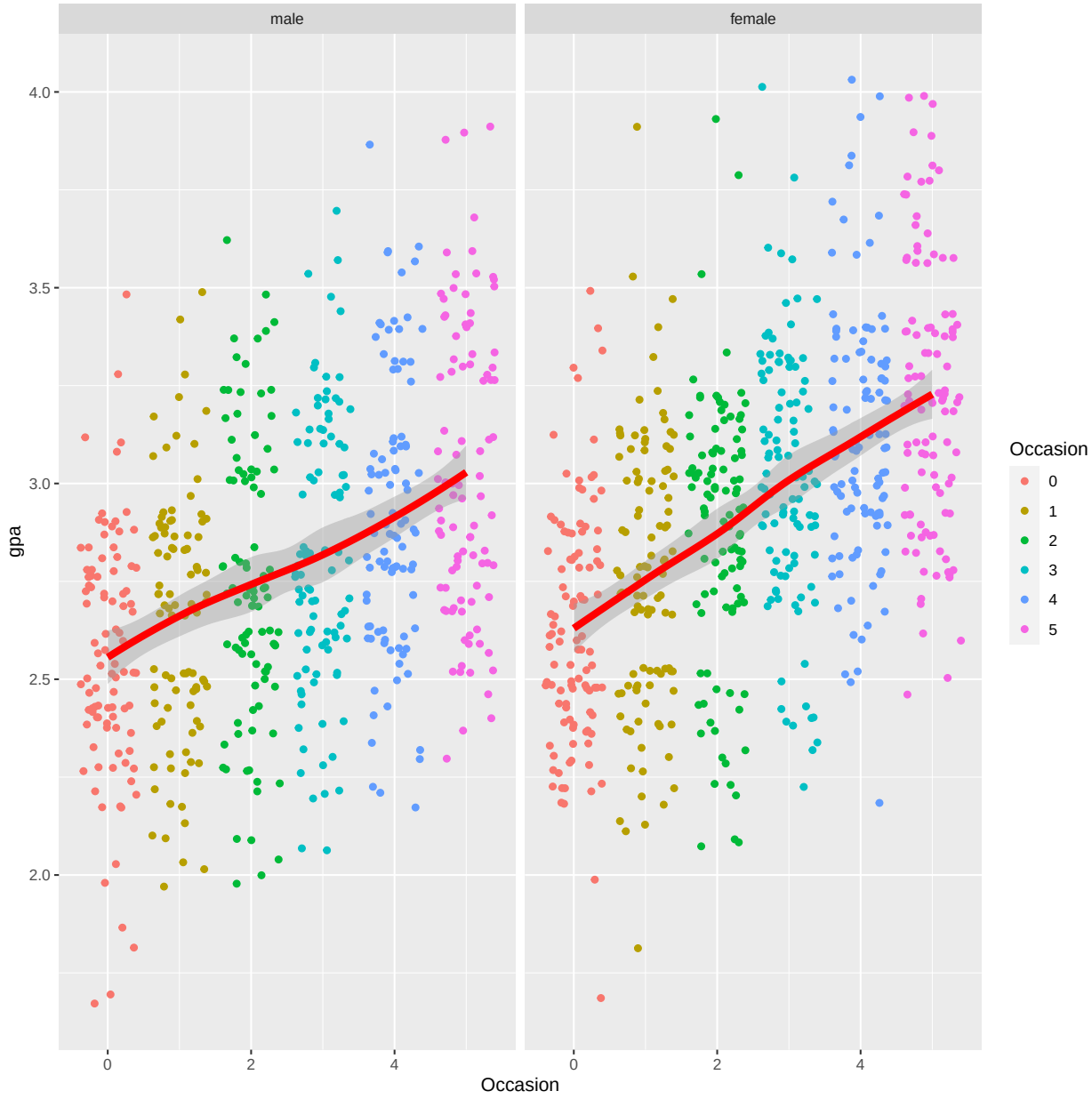

```
ggplot(long, aes(x = occas, y = gpa)) +
  geom_jitter(aes(color = factor(occas))) +
  geom_smooth(method = "loess", color = "red", size = 2) +
  xlab("Occasion") +
  labs(color = "Occasion")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



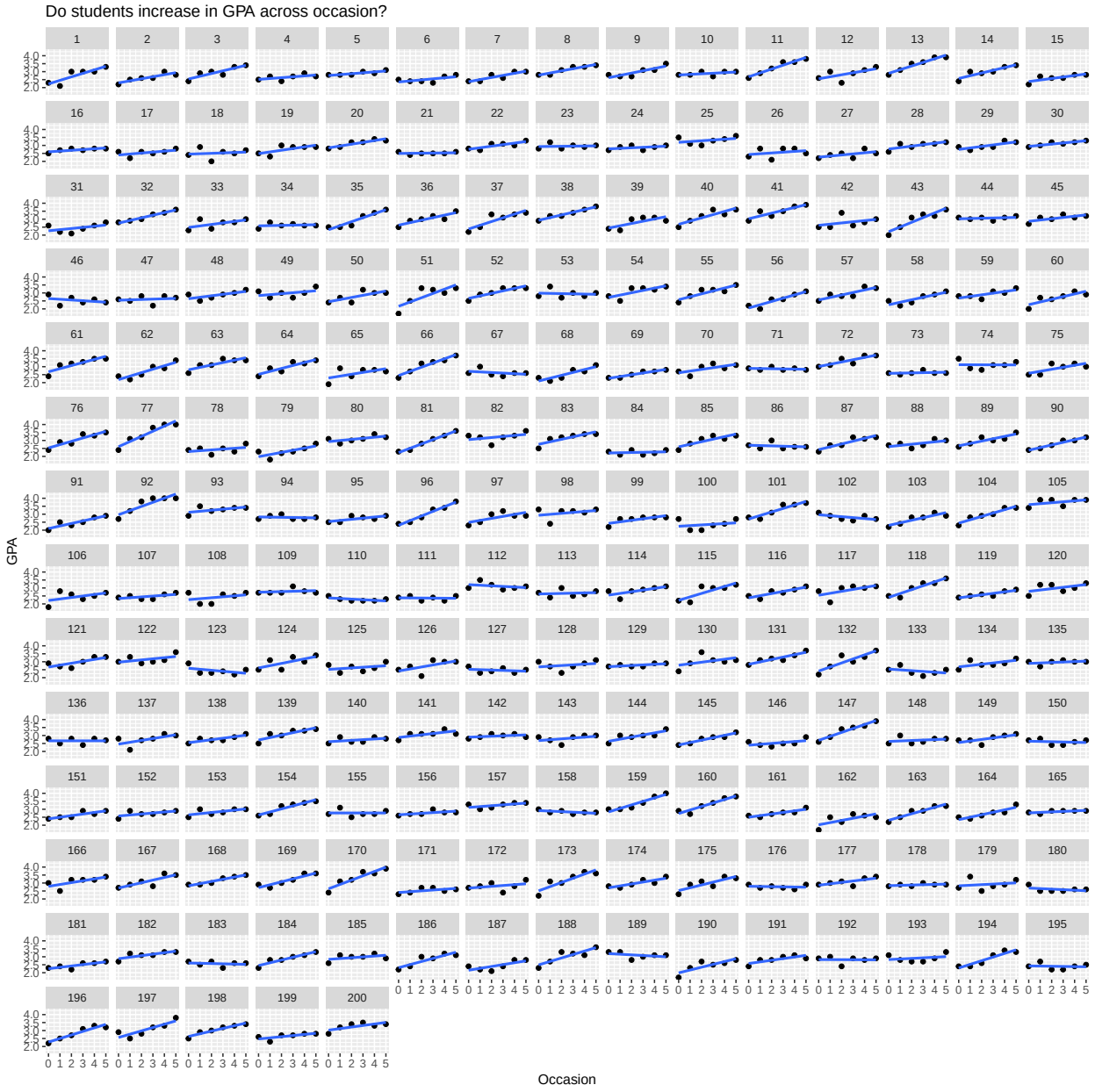

```
ggplot(long, aes(x = occas, y = gpa)) +
  geom_jitter(aes(color = factor(occas))) +
  geom_smooth(method = "loess", color = "red", size = 2) +
  xlab("Occasion") +
  labs(color = "Occasion") +
  facet_wrap(~sex)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(data = long, aes(x = occas, y = gpa)) +
geom_point() +
geom_smooth(method = "lm", se = FALSE) +
facet_wrap(~student) +
labs(title = "Do students increase in GPA across occasion?",
x = "Occasion", y = "GPA")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




# Baseline model

```
mod1 = lmer(gpa ~ 1 + (1|student), REML = TRUE, data = long)
summary(mod1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: gpa ~ 1 + (1 | student)
##    Data: long
##
## REML criterion at convergence: 919.5
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -3.6504 -0.5496  0.0603  0.6356  2.5736
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  student  (Intercept) 0.05714  0.2390
##  Residual             0.09759  0.3124
## Number of obs: 1200, groups:  student, 200
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   2.86500    0.01916 198.99999   149.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Added time as a random factor

```
mod2 = lmer(gpa ~ 1 + occas + (1 + occas|student), REML = TRUE, data = long)
summary(mod2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: gpa ~ 1 + occas + (1 + occas | student)
##    Data: long
##
## REML criterion at convergence: 261
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2696 -0.5377 -0.0128  0.5326  3.1938
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  student  (Intercept) 0.045192 0.2126
##           occas       0.004503 0.0671   -0.10
##  Residual             0.042389 0.2059
## Number of obs: 1200, groups:  student, 200
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 2.599e+00  1.836e-02 1.990e+02  141.59   <2e-16 ***
## occas       1.063e-01  5.884e-03 1.990e+02   18.07   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## occas -0.345
```

## ICC

ICC = variance for effect of interest / variance for effect + unexplained variance This is the underlying structure of the F statistic ICC for mod 2 = .005 / .005 + .042 = .106

Compare the ICCs across the different models

## Add covariate

```
mod4 = lmer(gpa ~ 1 + sex + (1|student), REML = TRUE, data = long)
summary(mod4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: gpa ~ 1 + sex + (1 | student)
##    Data: long
##
## REML criterion at convergence: 908.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5928 -0.5612  0.0553  0.6242  2.5290
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  student  (Intercept) 0.05200  0.2280
##  Residual             0.09759  0.3124
## Number of obs: 1200, groups:  student, 200
##
## Fixed effects:
##             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   2.78737    0.02681 197.99978 103.983  < 2e-16 ***
## sexfemale     0.14787    0.03700 197.99978   3.997 9.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sexfemale -0.725
```

```
mod5 = lmer(gpa ~ 1 + occas + sex + (1 + occas|student), REML = TRUE, data = long)
summary(mod5)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: gpa ~ 1 + occas + sex + (1 + occas | student)
##    Data: long
##
## REML criterion at convergence: 255.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1715 -0.5428 -0.0081  0.5317  3.2393
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  student  (Intercept) 0.044345 0.21058
##           occas       0.004504 0.06711  -0.16
##  Residual             0.042388 0.20588
```

```
## Number of obs: 1200, groups:  student, 200
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 2.541e+00  2.542e-02 2.134e+02  99.981   <2e-16 ***
## occas       1.063e-01  5.885e-03 1.990e+02  18.066   <2e-16 ***
## sexfemale   1.100e-01  3.372e-02 1.980e+02   3.263   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) occas
## occas     -0.276
## sexfemale -0.696  0.000
```

## Job + sex

```
mod6 = lmer(gpa ~ 1 + sex + job + (1|student), REML = TRUE, data = long)
summary(mod6)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: gpa ~ 1 + sex + job + (1 | student)
##    Data: long
##
## REML criterion at convergence: 912.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6049 -0.5523  0.0553  0.6306  2.5162
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  student  (Intercept) 0.05208  0.2282
##  Residual             0.09759  0.3124
## Number of obs: 1200, groups:  student, 200
##
## Fixed effects:
##                Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)    2.80309    0.03225 196.99978  86.914  < 2e-16 ***
## sexfemale      0.14885    0.03703 196.99978   4.019 8.31e-05 ***
## jobfull time  -0.03247    0.03699 196.99978  -0.878    0.381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) sexfml
## sexfemale   -0.586
## jobfulltime -0.555 -0.030
```

```
mod7 = lmer(gpa ~ 1 + occas + sex + job + (1+ occas|student), REML = TRUE, data = long)
summary(mod7)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
## lmerModLmerTest]
## Formula: gpa ~ 1 + occas + sex + job + (1 + occas | student)
##    Data: long
##
## REML criterion at convergence: 260.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1706 -0.5429 -0.0075  0.5322  3.2388
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  student  (Intercept) 0.044647 0.21130
##           occas       0.004504 0.06711  -0.16
##  Residual             0.042388 0.20588
## Number of obs: 1200, groups:  student, 200
##
## Fixed effects:
##                Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)   2.542e+00  3.028e-02  2.111e+02  83.951  < 2e-16 ***
## occas         1.063e-01  5.885e-03  1.990e+02  18.066  < 2e-16 ***
## sexfemale     1.101e-01  3.382e-02  1.970e+02   3.255  0.00134 **
## jobfull time -8.962e-04  3.378e-02  1.970e+02  -0.027  0.97886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) occas  sexfml
## occas       -0.232
## sexfemale   -0.570  0.000
## jobfulltime -0.540  0.000 -0.030
```