Hierarchical Linear Models

John D. Lee and Linda Ng Boyle
10/1/2018

A Common Challenge

Repeated measures

– Multiple measures from the same person who might also be part of a group

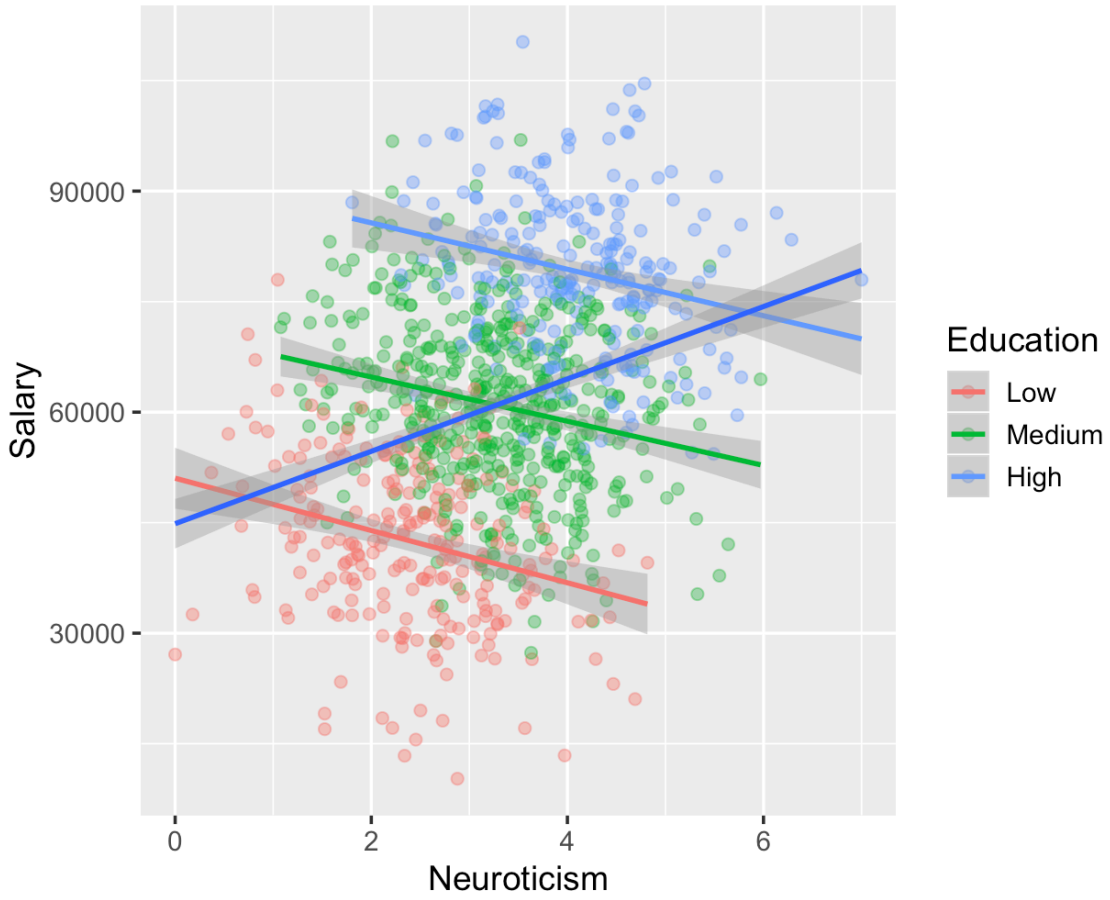
– Multiple exposures across people to the same stimuli

– Repeated measures ANOVA accommodates correlated observations

– Multi-level general linear models go further and explicitly model grouping effects

– Addresses Simpson's paradox, where individual trends don't reflect the aggregate trend

Simpson's Paradox



http://rpubs.com/lakenp/simpsonsparadox

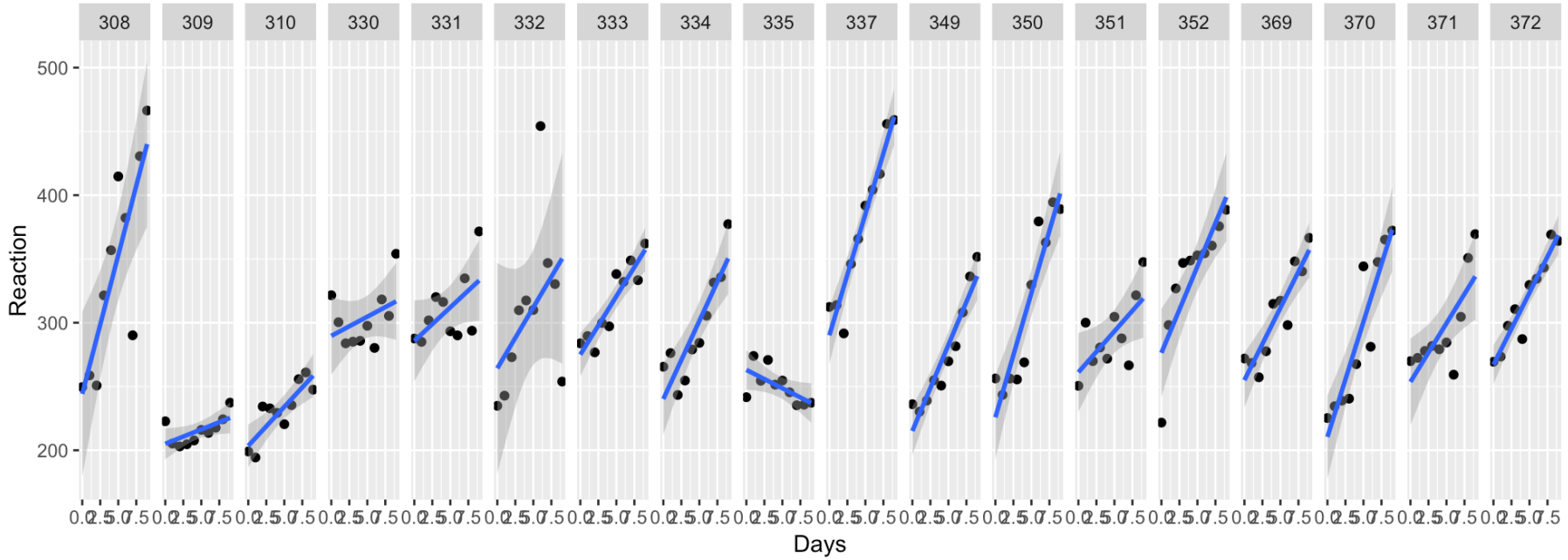
3/16

Useful packages

- `library(tidyverse)` for ggplot2, dplyr
- `library(skimr)` for skim
- `library(lme4)` for mixed effect, multi-level general linear models
- `library(lmerTest)` for extracting p-values
- `library(psycho)` for extracting APA reportable sentences

Effect of Sleep Disruption

```
sleep.df = sleepstudy

ggplot(sleep.df, aes(Days, Reaction)) +
   geom_point() +
   geom_smooth(method = "lm") +
   facet_grid(.~Subject)
```



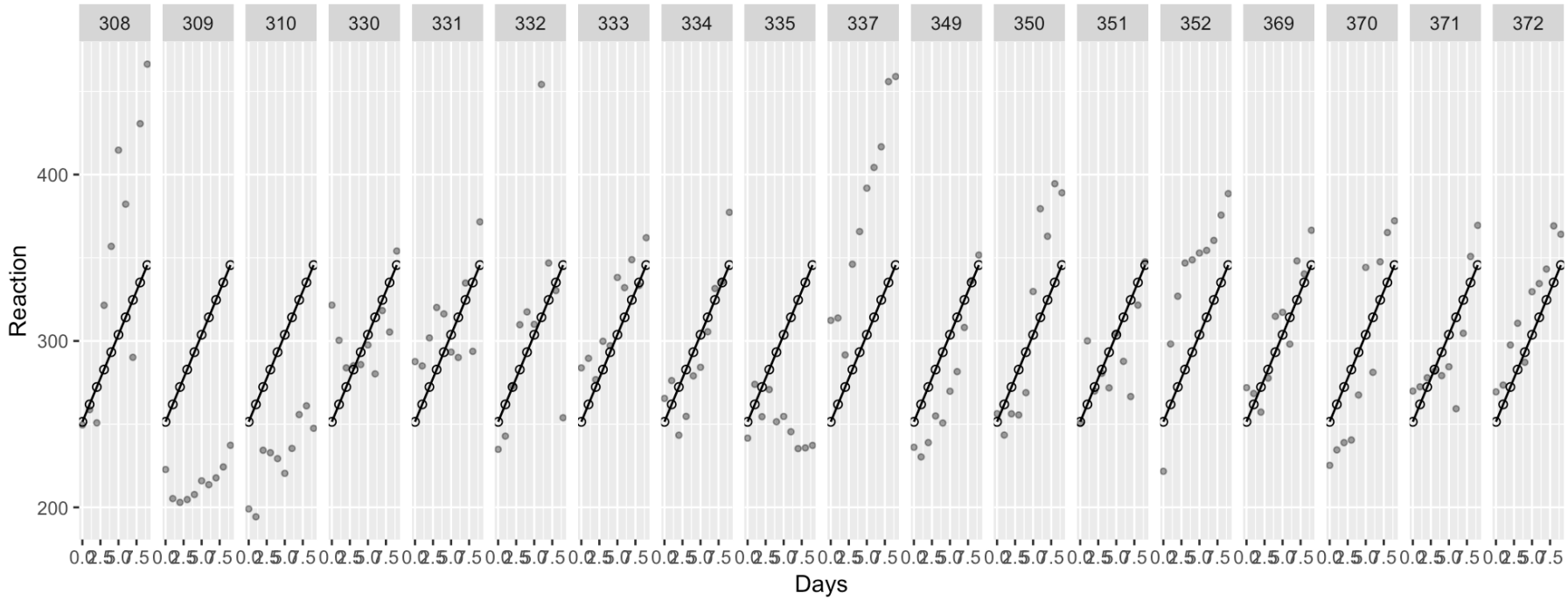
Uniform Effect of Sleep Disruption?

```
sleep.fit = lm(Reaction ~ Days, sleep.df)

sleep.df$prediction = predict(sleep.fit)

summary(sleep.fit)
```

```
##
## Call:
## lm(formula = Reaction ~ Days, data = sleep.df)
##
## Residuals:
##     Min       1Q   Median       3Q      Max
## -110.848  -27.483    1.546   26.142  139.953
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  251.405      6.610  38.033  < 2e-16 ***
## Days          10.467      1.238   8.454 9.89e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.71 on 178 degrees of freedom
## Multiple R-squared:  0.2865, Adjusted R-squared:  0.2825
## F-statistic: 71.46 on 1 and 178 DF,  p-value: 9.894e-15
```

Uniform Effect of Sleep Disruption?

```
ggplot(sleep.df, aes(Days, Reaction)) +
  geom_point(alpha = .4, size = 1) +
  geom_point(aes(y=prediction), shape = 21) +
  geom_line(aes(y=prediction)) +
  facet_grid(.~Subject)
```



7/16

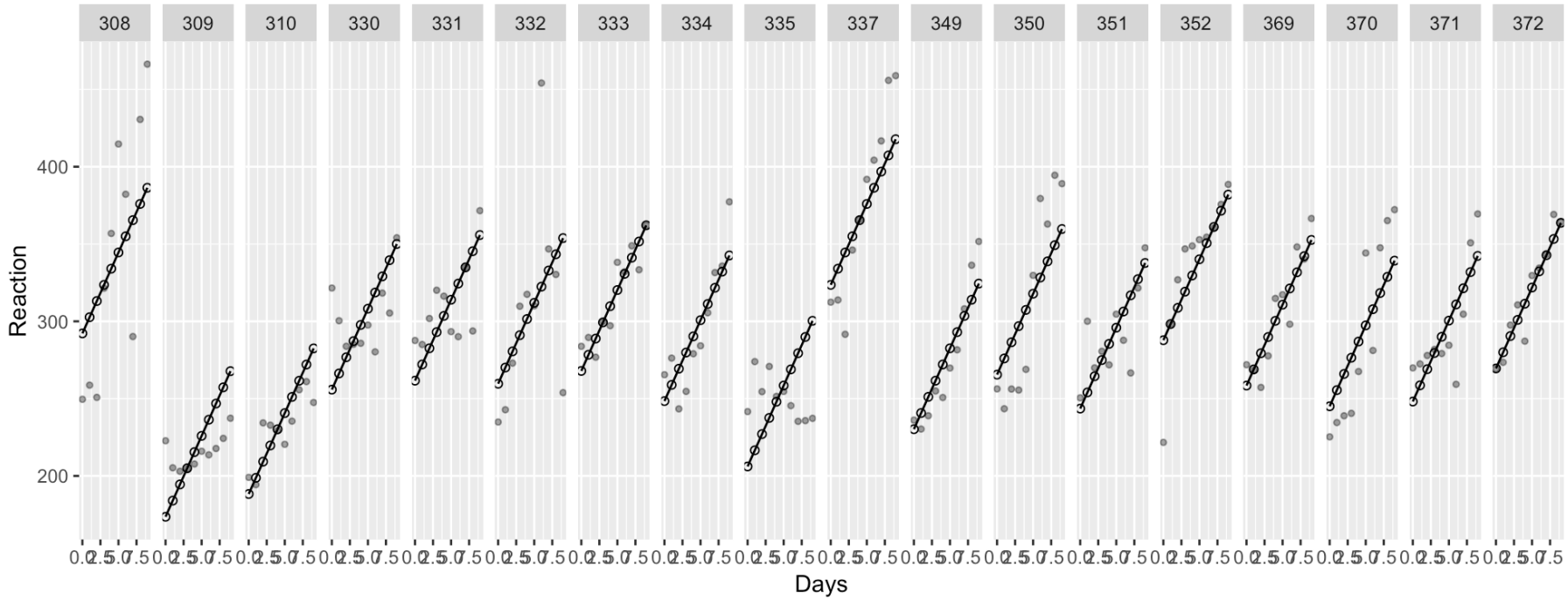
Random Effect for Intercept

```
sleep_int.fit = lmer(Reaction ~ Days + (1|Subject), sleep.df)
summary(sleep_int.fit)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Reaction ~ Days + (1 | Subject)
##    Data: sleep.df
##
## REML criterion at convergence: 1786.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2257 -0.5529  0.0109  0.5188  4.2506
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1378.2   37.12
##  Residual              960.5   30.99
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) 251.4051     9.7467  22.8102   25.79   <2e-16 ***
## Days         10.4673     0.8042 161.0000   13.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.371
```

Random Effect for Intercept

```
ggplot(sleep.df, aes(Days, Reaction)) +
  geom_point(alpha = .4, size = 1) +
  geom_point(aes(y=prediction_int), shape = 21) +
  geom_line(aes(y=prediction_int, group = Subject)) +
  facet_grid(.~Subject)
```



9/16

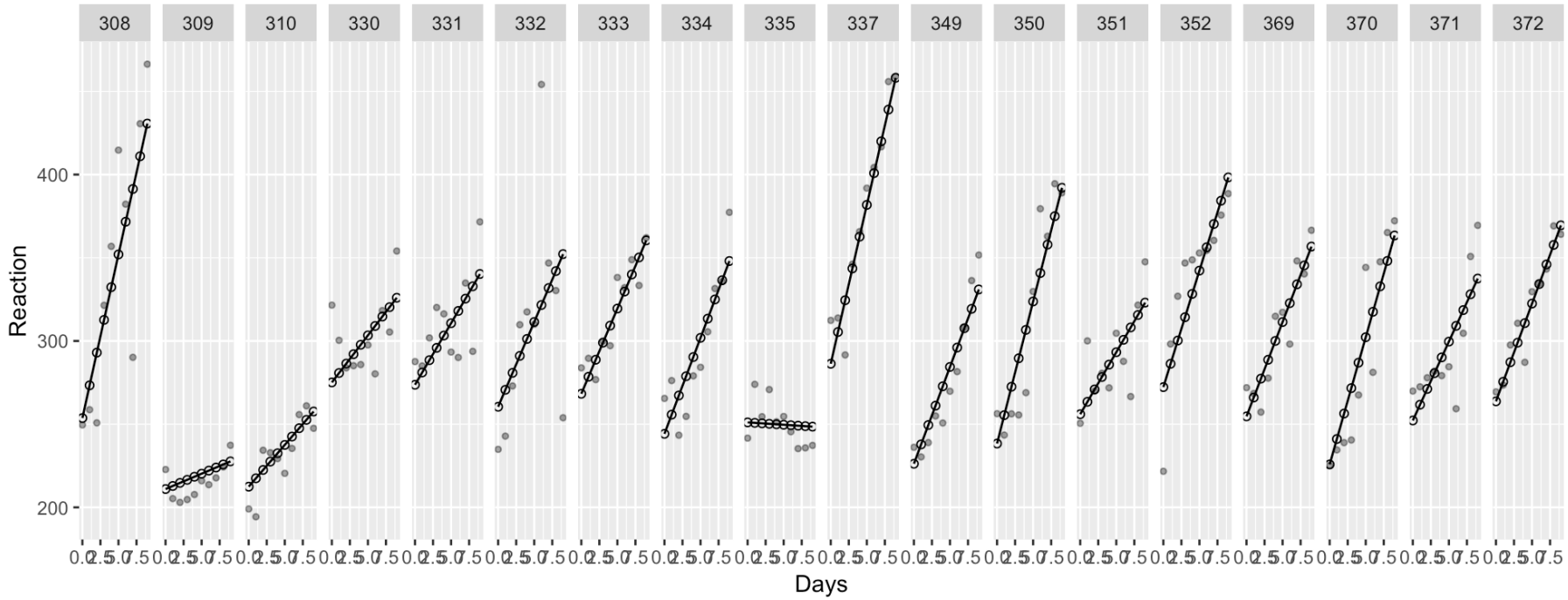
Random Effect for Intercept and Slope

```
sleep_int_slope.fit = lmer(Reaction ~ Days + (Days|Subject), sleep.df)
summary(sleep_int_slope.fit)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Reaction ~ Days + (Days | Subject)
##    Data: sleep.df
##
## REML criterion at convergence: 1743.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9536 -0.4634  0.0231  0.4634  5.1793
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 612.09   24.740
##           Days         35.07    5.922   0.07
##  Residual             654.94   25.592
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  251.405      6.825  17.000  36.838  < 2e-16 ***
## Days          10.467      1.546  17.000   6.771 3.26e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
```

Random Effects for Intercept and Slope

```
ggplot(sleep.df, aes(Days, Reaction)) +
  geom_point(alpha = .4, size = 1) +
  geom_line(aes(y=prediction_int_slope, group = Subject)) +
  geom_point(aes(y=prediction_int_slope), shape = 21) +
  facet_grid(.~Subject)
```



Summary for Publication

– The "psycho" package provides functions for summarizing model output

– It creates a publication-ready summary table

– It can even create sentences describing the output

– Parameter estimates can be extracted for plotting

```
library(psycho)
analyze(sleep_int_slope.fit, CI = 95) %>%
  summary(results, round = 2) %>%
  mutate(p = format_p(p)) %>%
  kable()
```

Variable	Effect_Size	Coef	SE	t	df	p	Coef_std	SE_std	CI_lower	CI_higher
(Intercept)	NA	251.41	6.82	36.84	17	< .001***	0.00	0.16	237.68	265.13
Days	medium	10.47	1.55	6.77	17	< .001***	0.54	0.08	7.36	13.58

Summary for Publication

psycho can create sentences describing the output

```
analyze(sleep_int_slope.fit, CI = 95)
```

```
## The overall model predicting Reaction (formula = Reaction ~ Days + (Days | Subject)) has an total expla
##    - The effect of Days is significant (beta = 10.47, SE = 1.55, 95% CI [7.36, 13.58], t(17) = 6.77, p
```

-The overall model predicting Reaction (formula = Reaction ~ Days + (Days | Subject)) has an total explanatory power (conditional R2) of 79.92%, in which the fixed effects explain 27.87% of the variance (marginal R2).

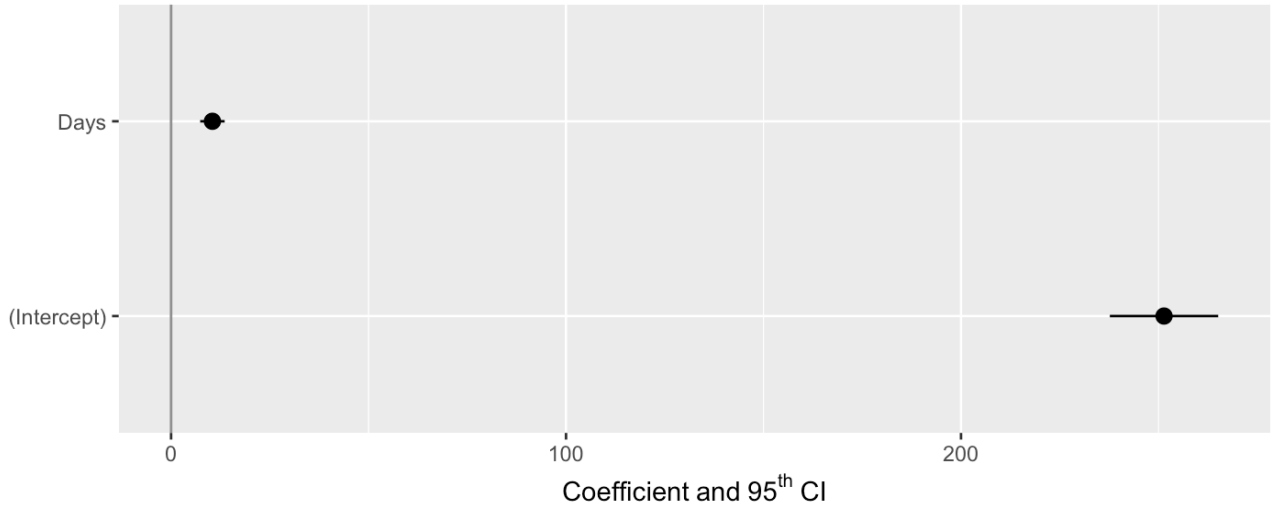
-The model's intercept is at 251.41 (SE = 6.82, 95% CI [237.68, 265.13]). Within this model:

-The effect of Days is significant (beta = 10.47, SE = 1.55, 95% CI [7.36, 13.58], t(17) = 6.77, p < .001) and can be considered as medium (std. beta = 0.54, std. SE = 0.079).

Summary for Publication

Parameter estimates can also be extracted for plotting

```
CI.lme.df = analyze(sleep_int_slope.fit)$summary

ggplot(CI.lme.df, aes(Variable)) +
    geom_hline(yintercept = 0, alpha = .5)+
    geom_pointrange(aes(y = Coef, ymin= CI_lower, ymax = CI_higher)) +
    coord_flip() +
    labs(x = "", y = expression(Coefficient~and~95^{th}~CI))
```



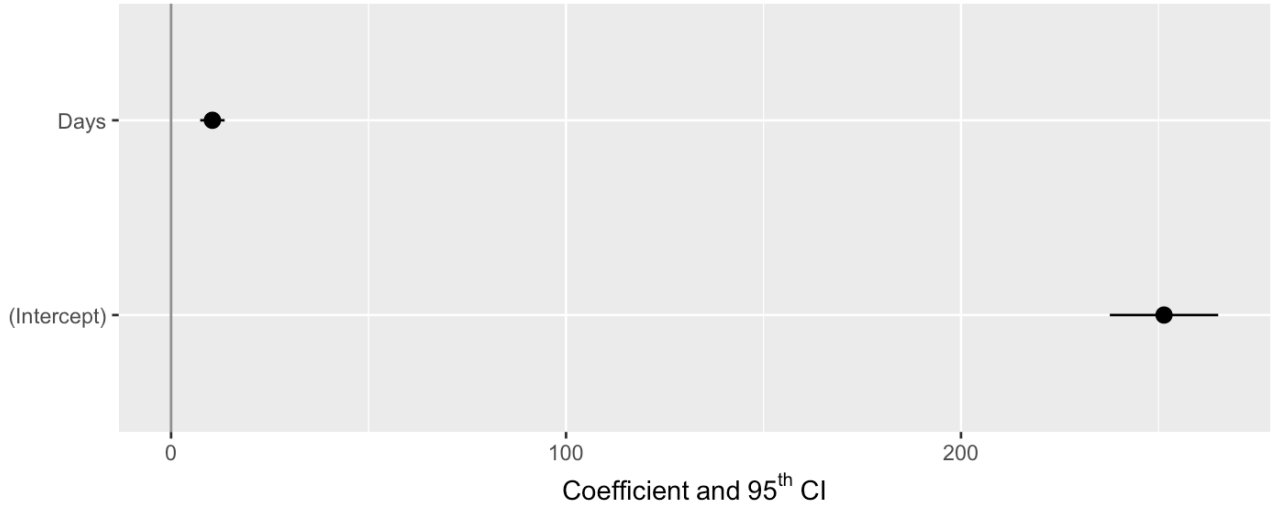
Exercise

Multi-level model fitting

– Fit a multi-level model with Subject as a random variable

– Use the `analyze` command from psych to extract an summmary of effects

– Plot the effect estimates



A Common Challenge

Repeated measures

– Multiple measures from the same person who might also be part of a group

– Multiple exposures across people to the same stimuli

– Repeated measures ANOVA accommodates correlated observations

– Multi-level general linear models go further and explicitly model grouping effects

– Addresses Simpson's paradox, where individual trends don't reflect the aggregate trend