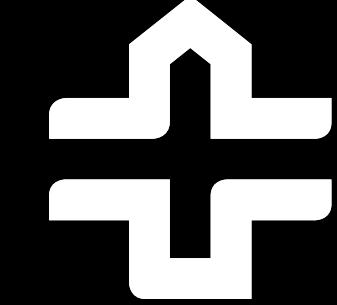
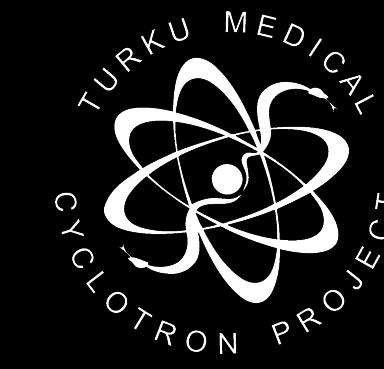


LINEAR MIXED EFFECTS MODELS

Turku PET Centre Brain Imaging Course 2024

Tuulia Malen, Turku PET Centre



Collect data

PET scans

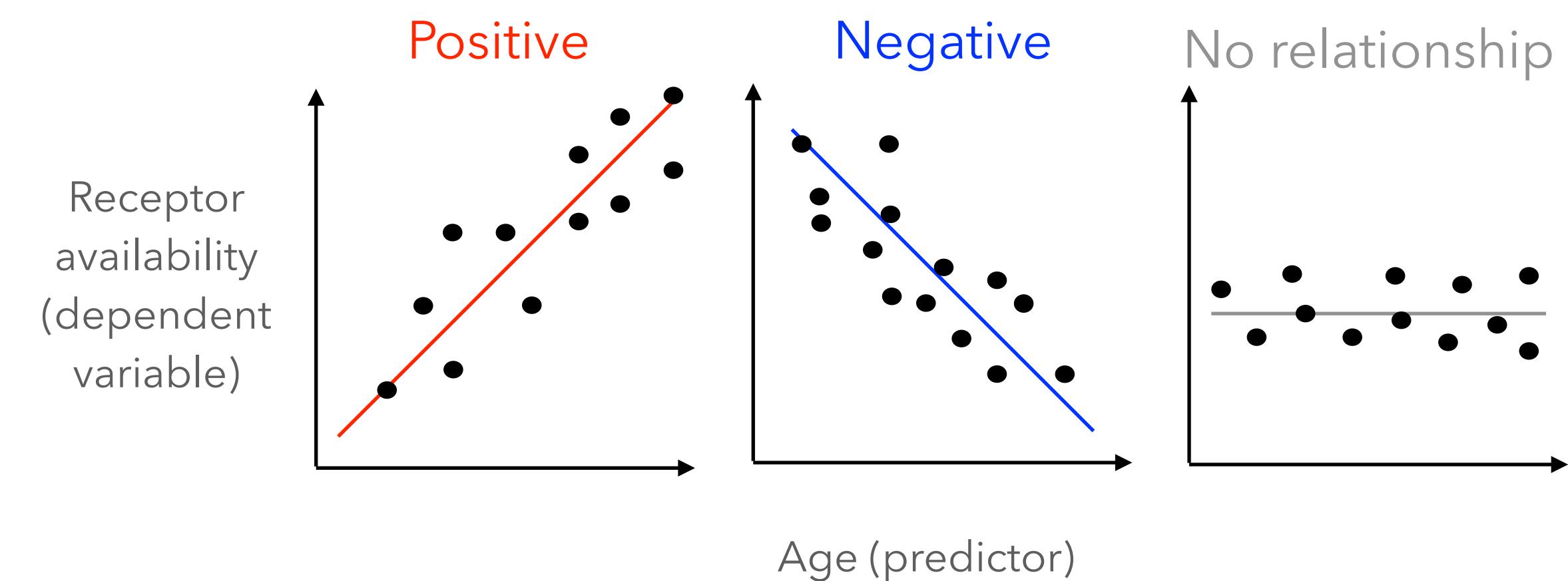
Image processing

From raw data to observations of interest
Tracer uptake in a PET scan
Receptor availability

Statistical analysis

Research question:
The effect of age on receptor availability?

Linear relationships:



Objectives

Linear mixed effects models

Theory:

- Linear model & linear mixed effects model
- Mixed effects in neuroimaging data
- Model diagnostics: Does the model fit the data sufficiently?

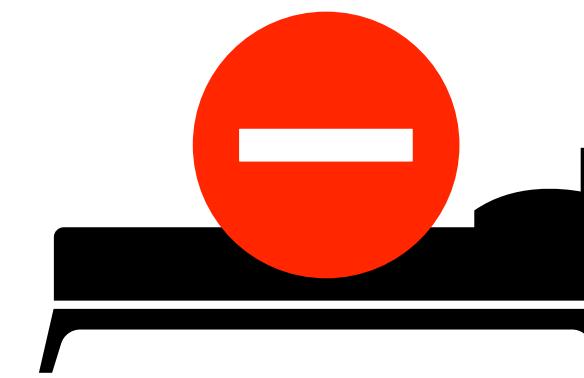
Practice:

- Running models and model diagnostics in R using RStudio
 - R is a language for statistical computing
 - RStudio is an environment for running R code

Linear model

- Linear model= linear regression
 - Regression: looking for relationship between variables
 - How predictor(s) explain/ predict dependent variable
 - Simple linear regression (1 predictor), multiple linear regression (>1 predictors)
 - Linear mixed effect model is one type of linear model

Example case: Sleep deprivation data



Sleep deprivation
0-9 days experiment
3 hours of sleep / night



Reaction time
in milliseconds (ms)?

Reaction	Days	Subject
249.5600	0	Sophie
258.7047	1	Sophie
250.8006	2	Sophie
321.4398	3	Sophie
356.8519	4	Sophie
414.6901	5	Sophie
382.2038	6	Sophie
290.1486	7	Sophie
430.5853	8	Sophie
466.3535	9	Sophie
222.7339	0	Andy
205.2658	1	Andy
202.9778	2	Andy

Sleepstudy dataset from lme4 package in R

Days= Number of days of sleep deprivation (3 hours/ night)

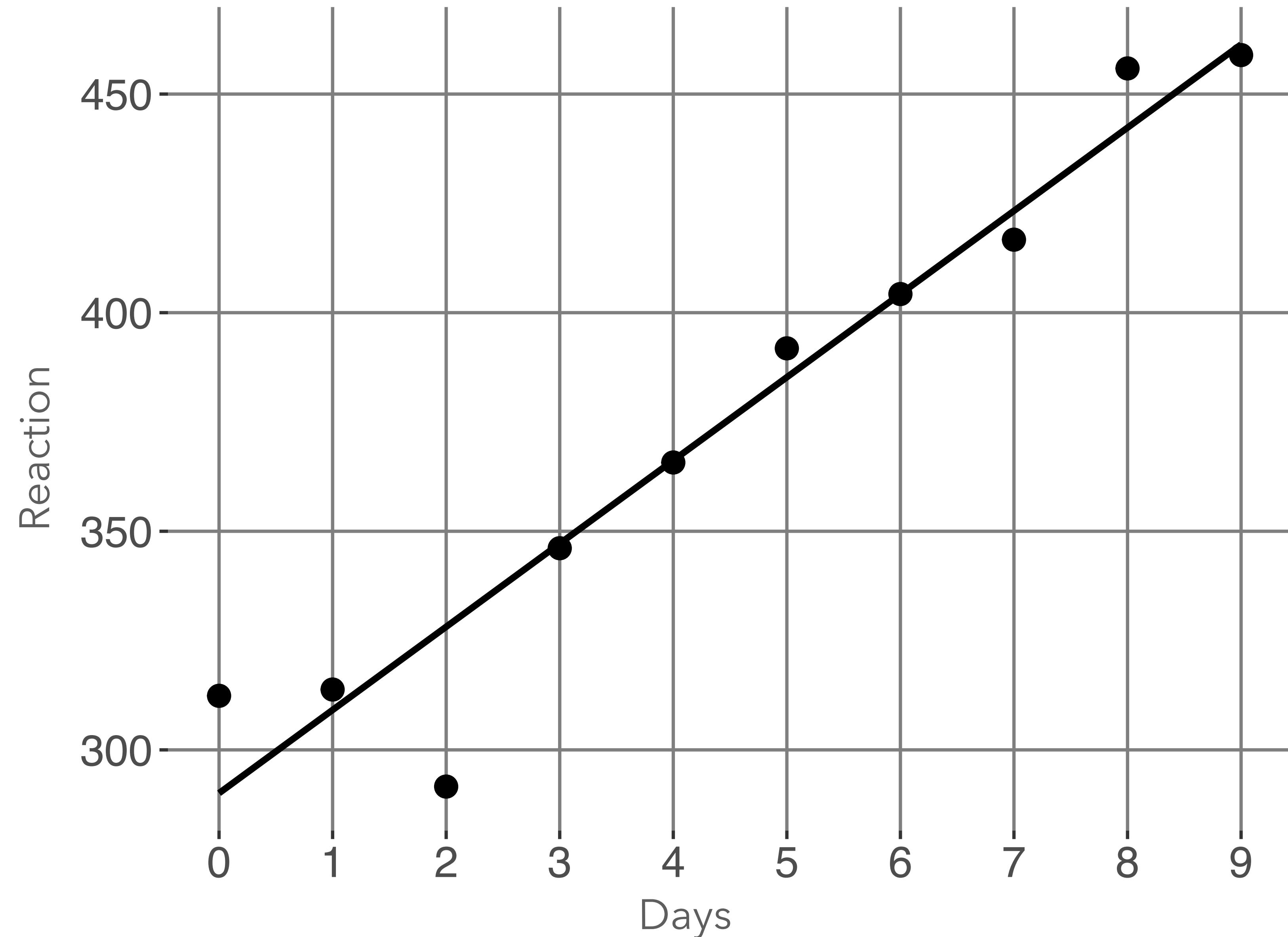
Reaction= Average reaction time (ms)

Subjects edited (mock-up names)

What is the effect of sleep deprivation on reaction time?

- Data from one subject
- Simple linear regression

$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

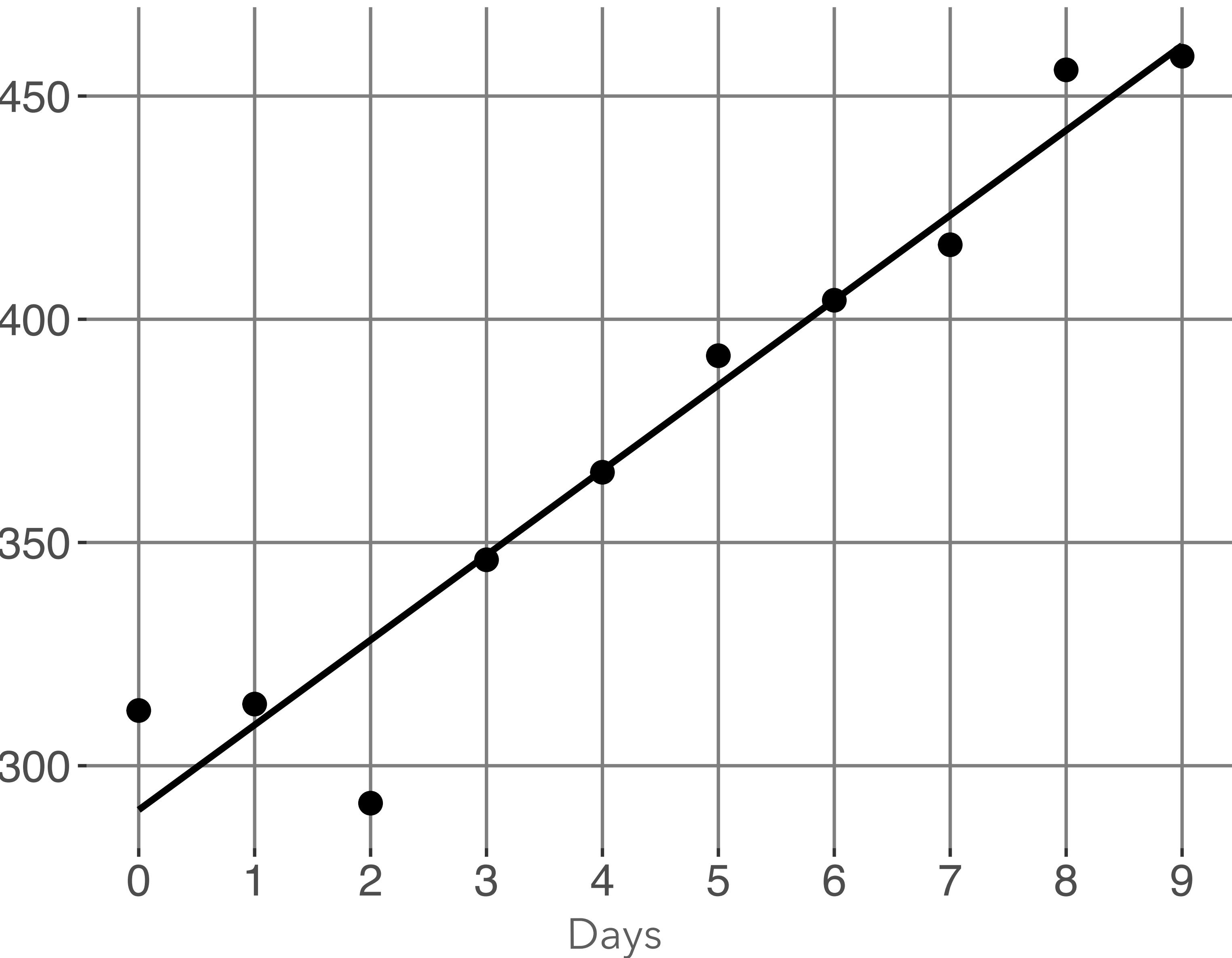


Dependent variable (y)

$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

Predictor variable (x)

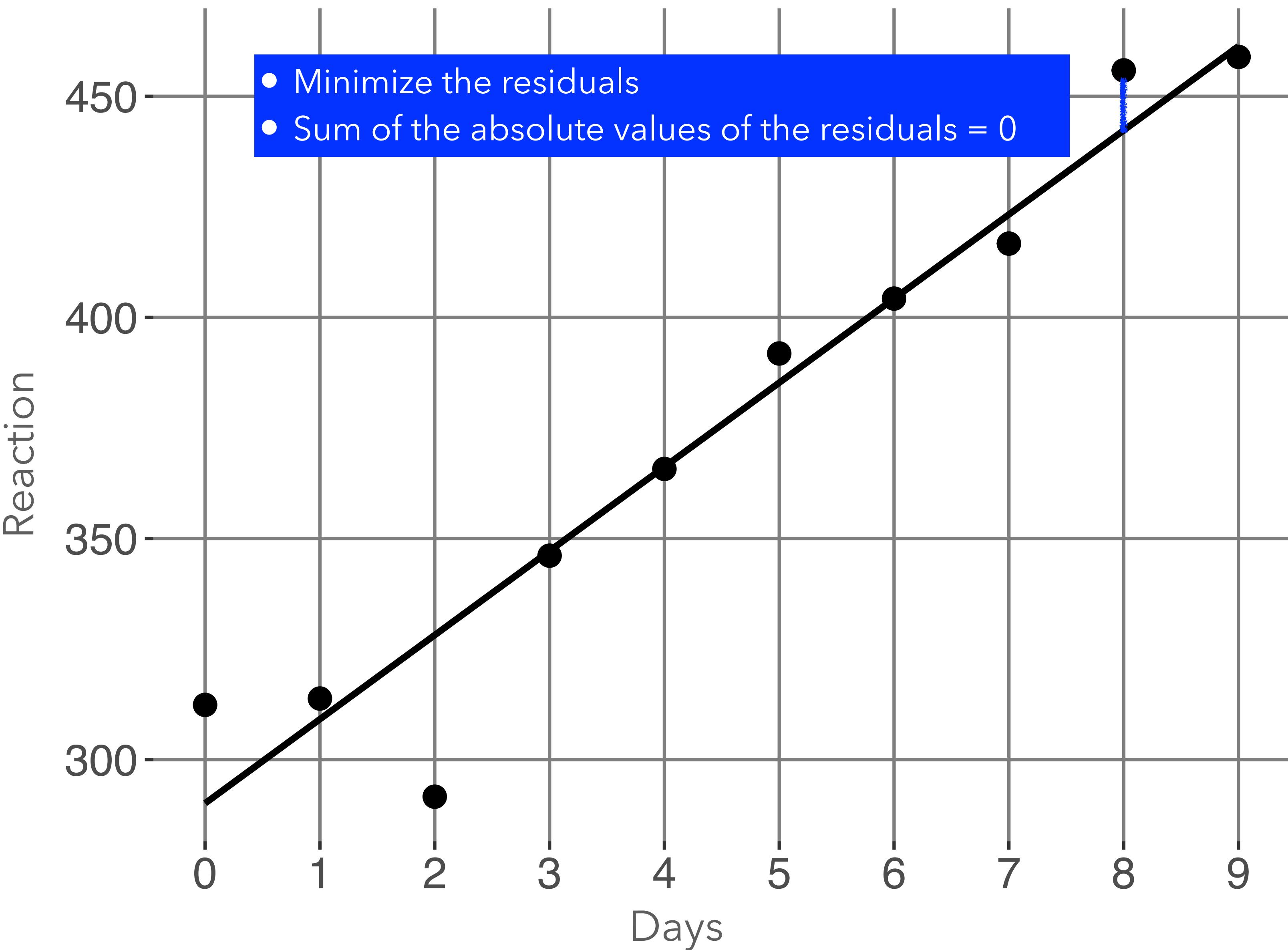
Reaction



Parameters

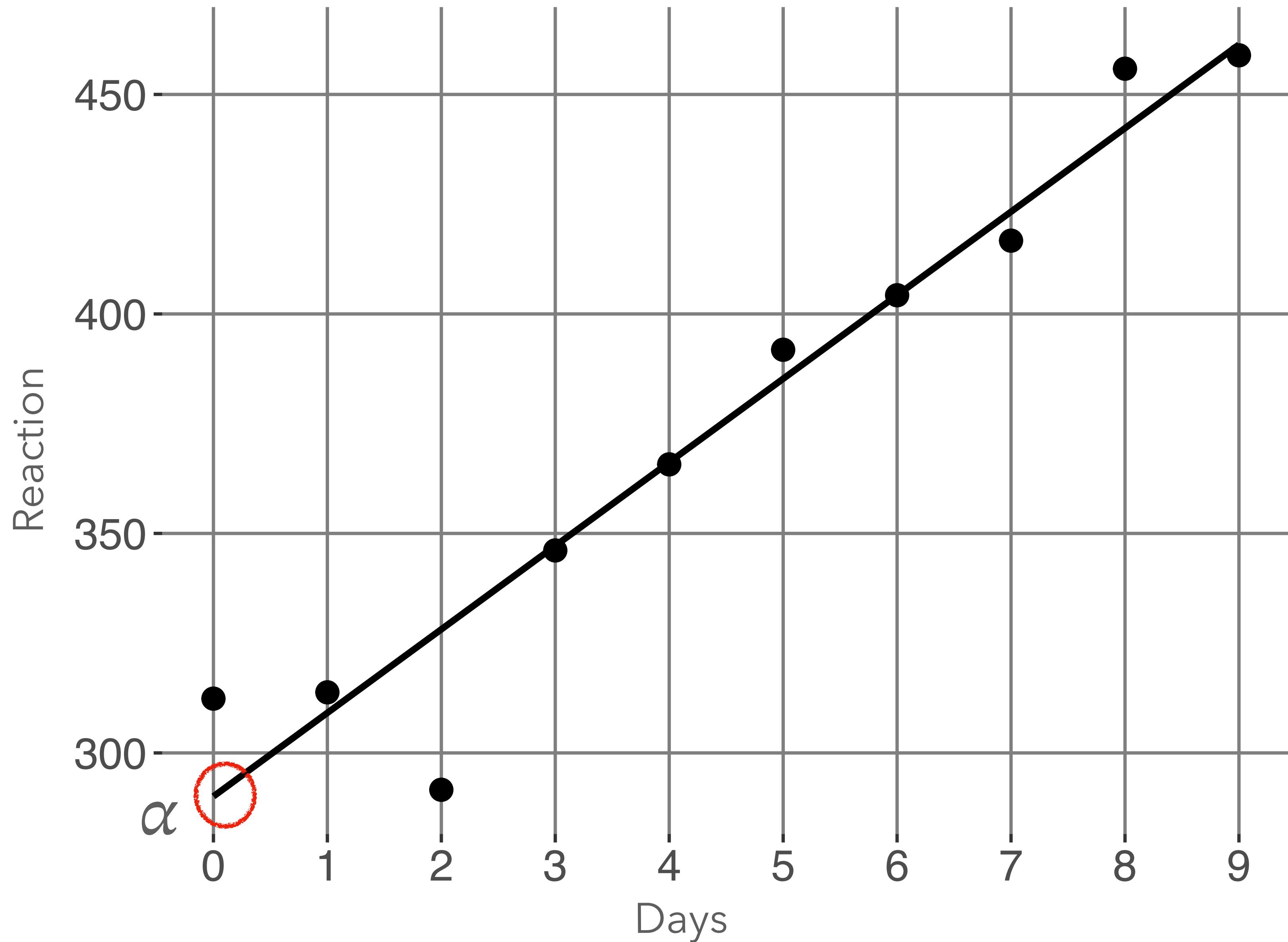
Reaction = $\alpha + \beta \cdot \text{Days} + \varepsilon$

Error
(Residuals)

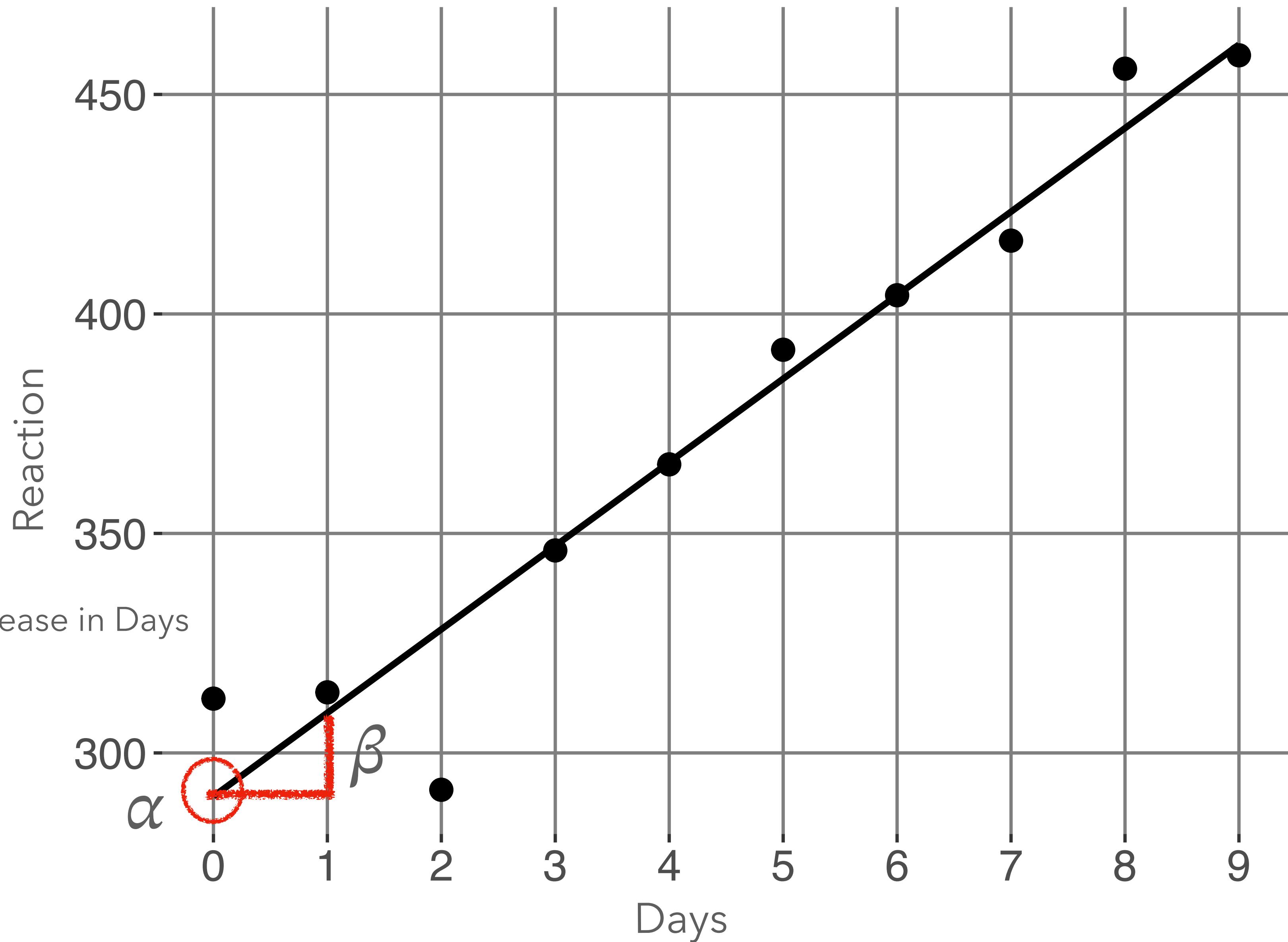


Intercept
• Reaction when Days= 0

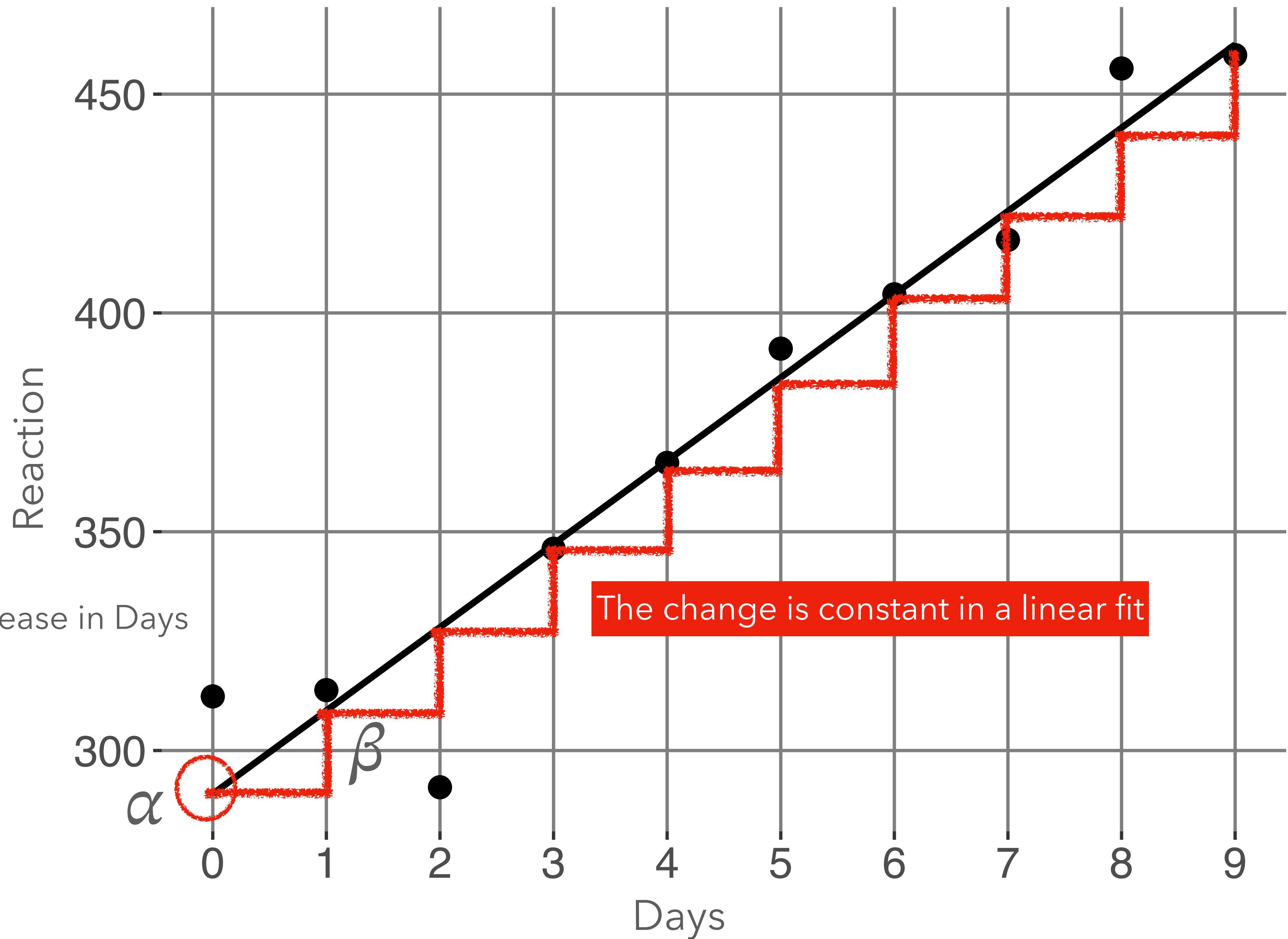
$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$



- $\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$
- Reaction when Days= 0
 - Main effect of Days on Reaction (slope)
 - The change in Reaction for one unit increase in Days
- Intercept
- Regression coefficient



- Intercept
- Reaction when Days= 0
- $\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$
- Regression coefficient
- Main effect of Days on Reaction (slope)
 - The change in Reaction for one unit increase in Days



$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

Prediction (fit)

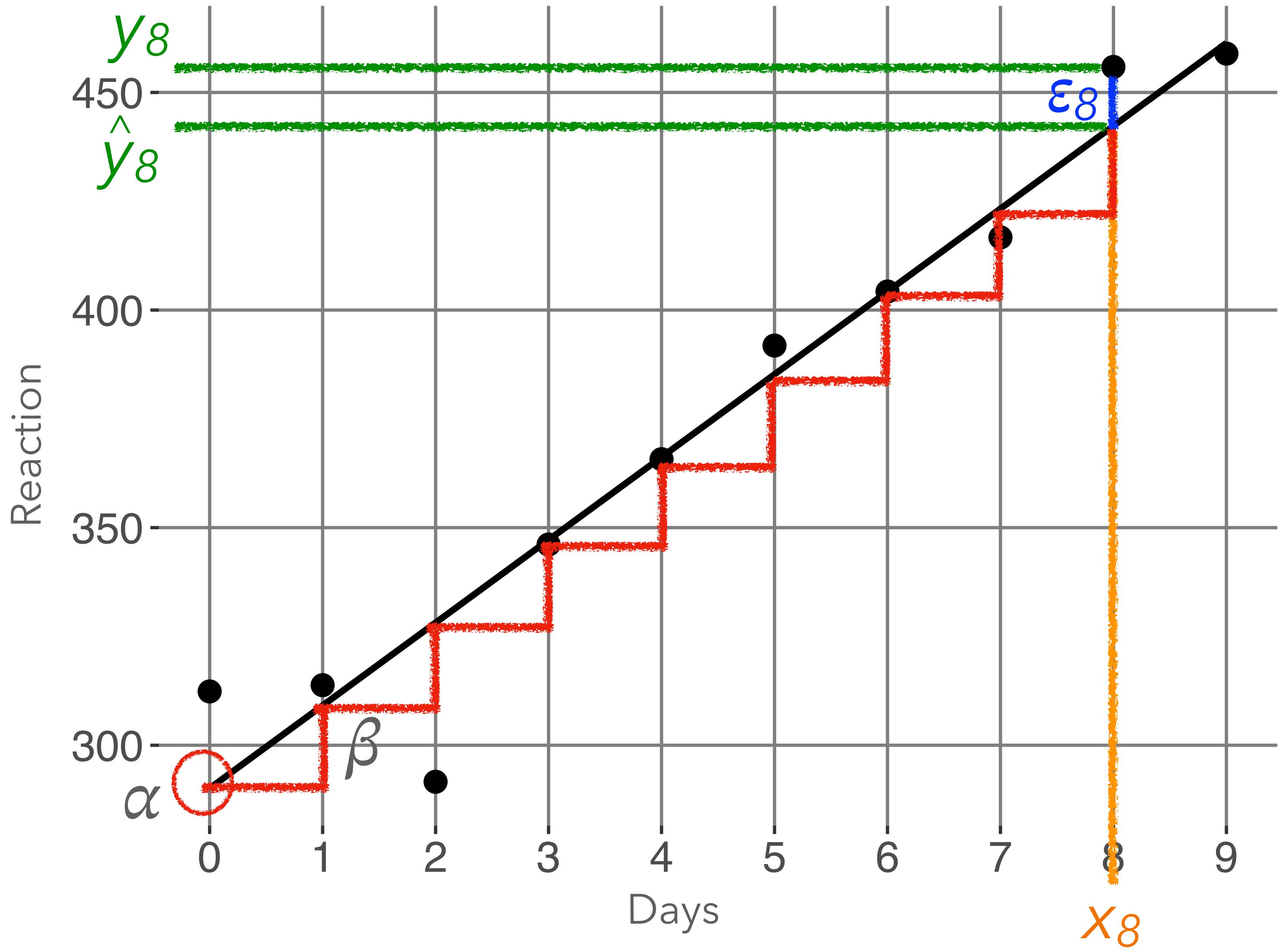
$$\hat{y}_8 = \alpha + \beta \cdot x_8$$

$$442 = 290 + 19 \cdot 8$$

Observation

$$y_8 = \alpha + \beta \cdot x_8 + \varepsilon_8$$

$$456 = 290 + 19 \cdot 8 + 14$$



$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

Prediction (fit)

$$\hat{y}_8 = \alpha + \beta \cdot x_8$$

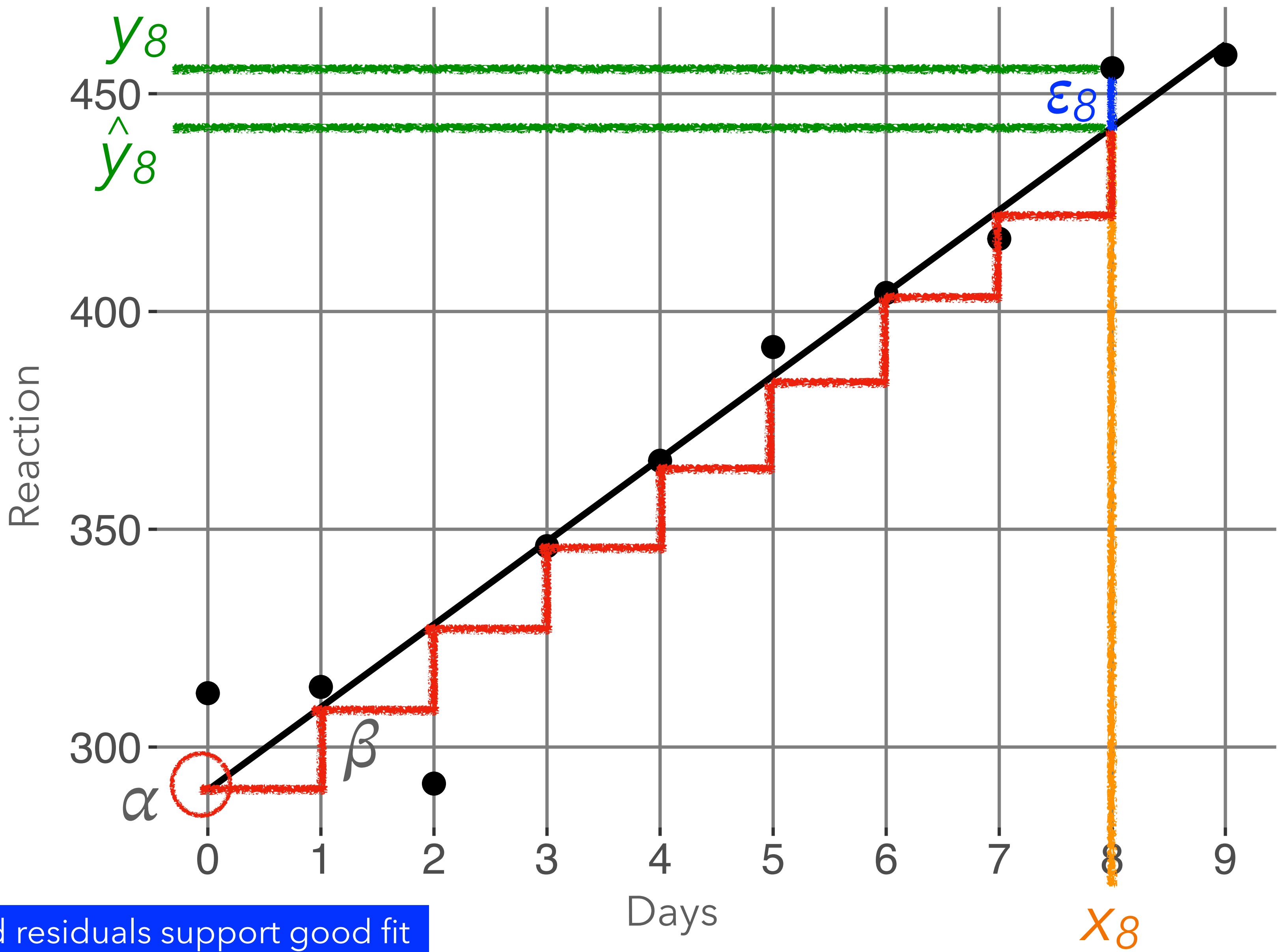
$$442 = 290 + 19 \cdot 8$$

Observation

$$y_8 = \alpha + \beta \cdot x_8 + \varepsilon_8$$

$$456 = 290 + 19 \cdot 8 + 14$$

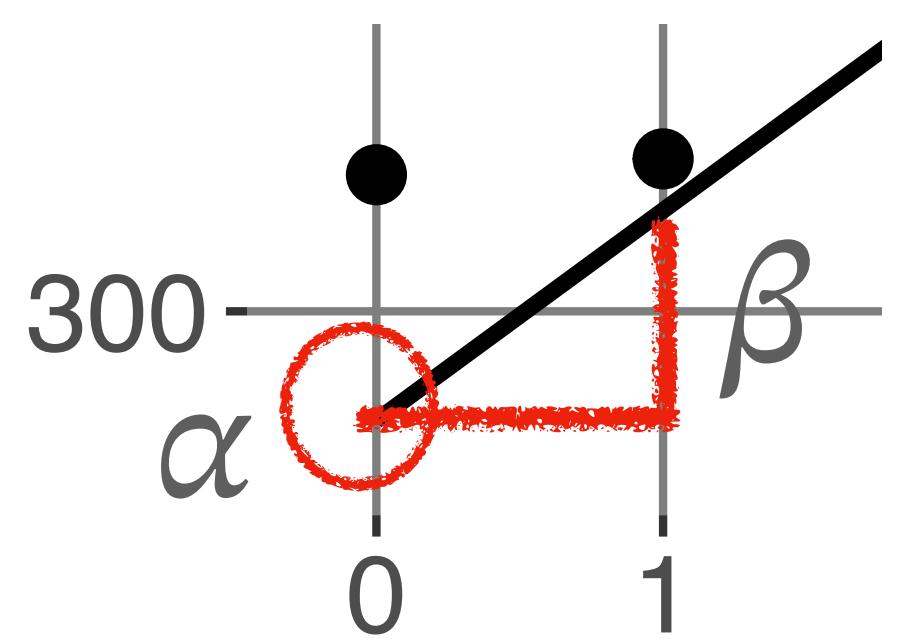
Randomly (not systematically) distributed residuals support good fit





Slope (regression coefficient)

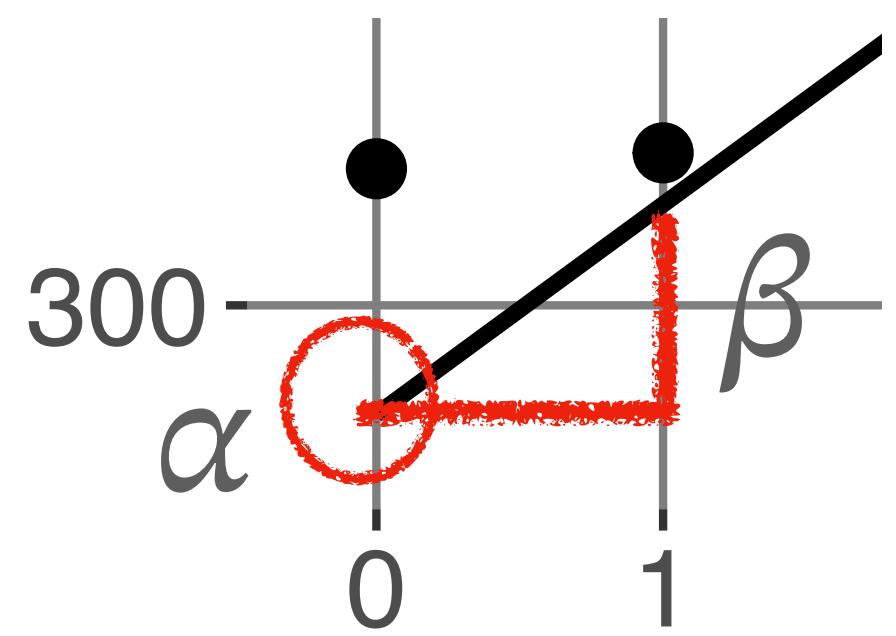
The change in the dependent variable for one unit increase in the predictor
The change in reaction time with one more night of sleep deprivation
'The effect of sleep deprivation on reaction time'





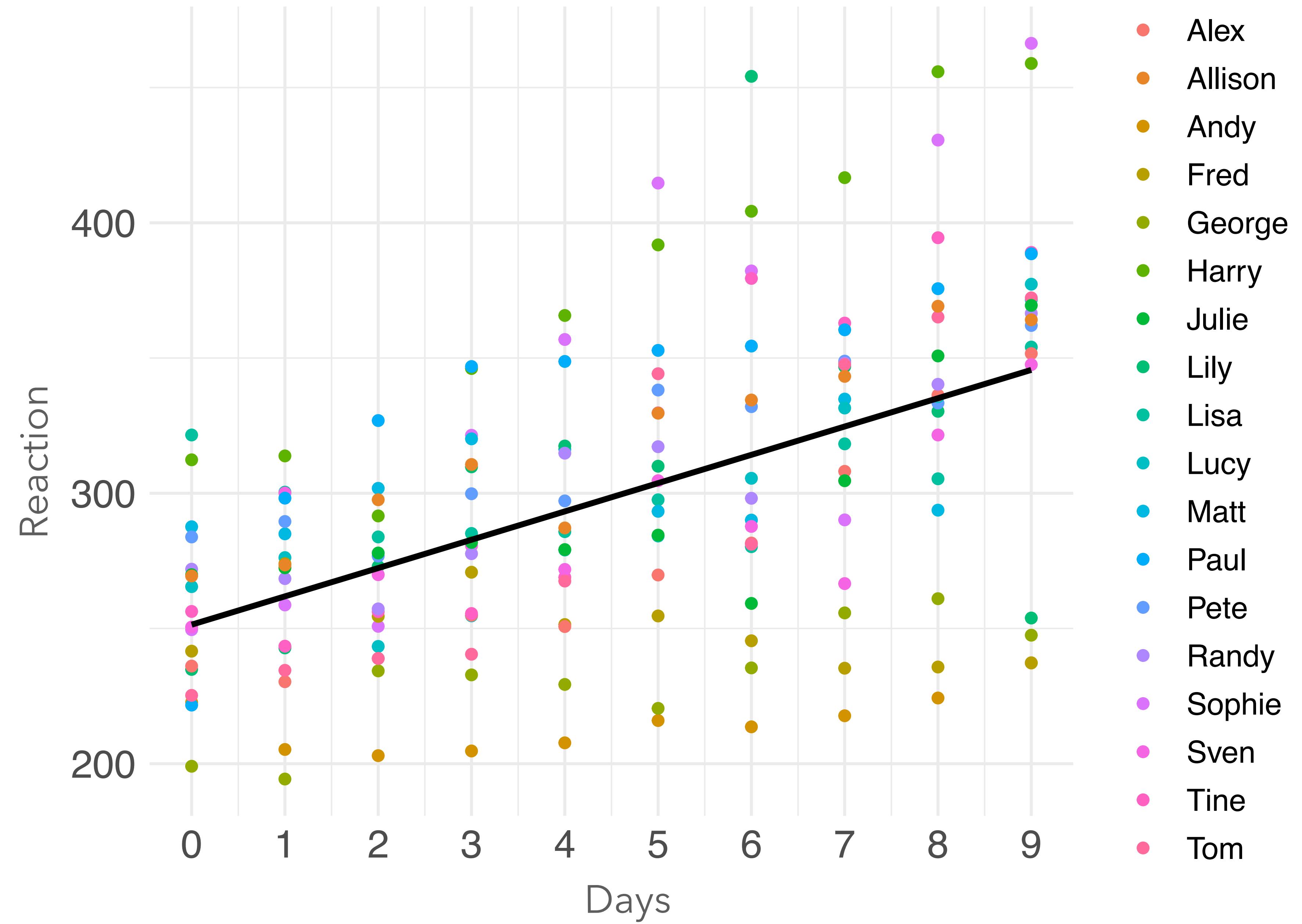
Slope (regression coefficient)

The change in the dependent variable for one unit increase in the predictor
The change in reaction time with one more night of sleep deprivation
'The effect of sleep deprivation on reaction time'



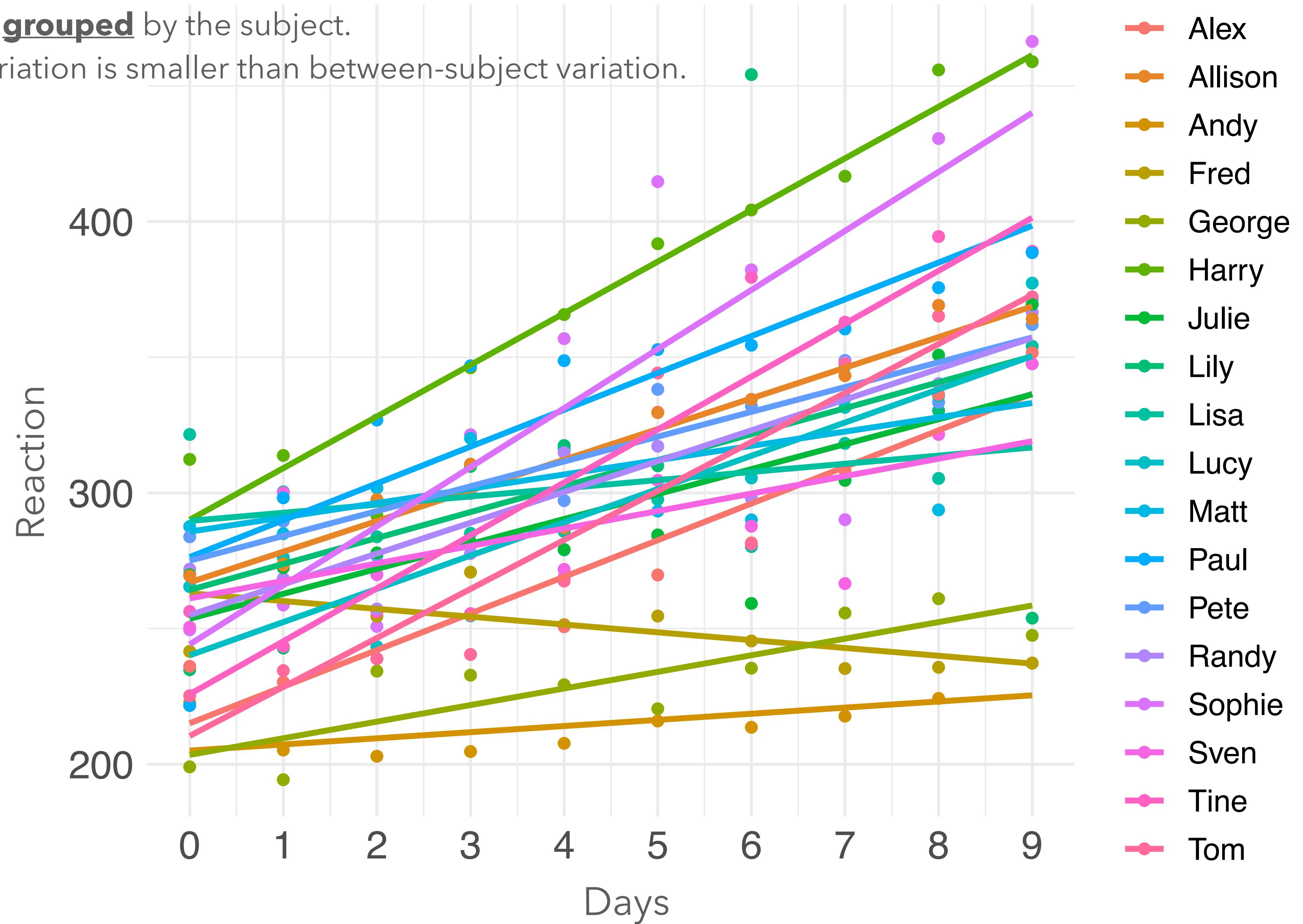
Intercept

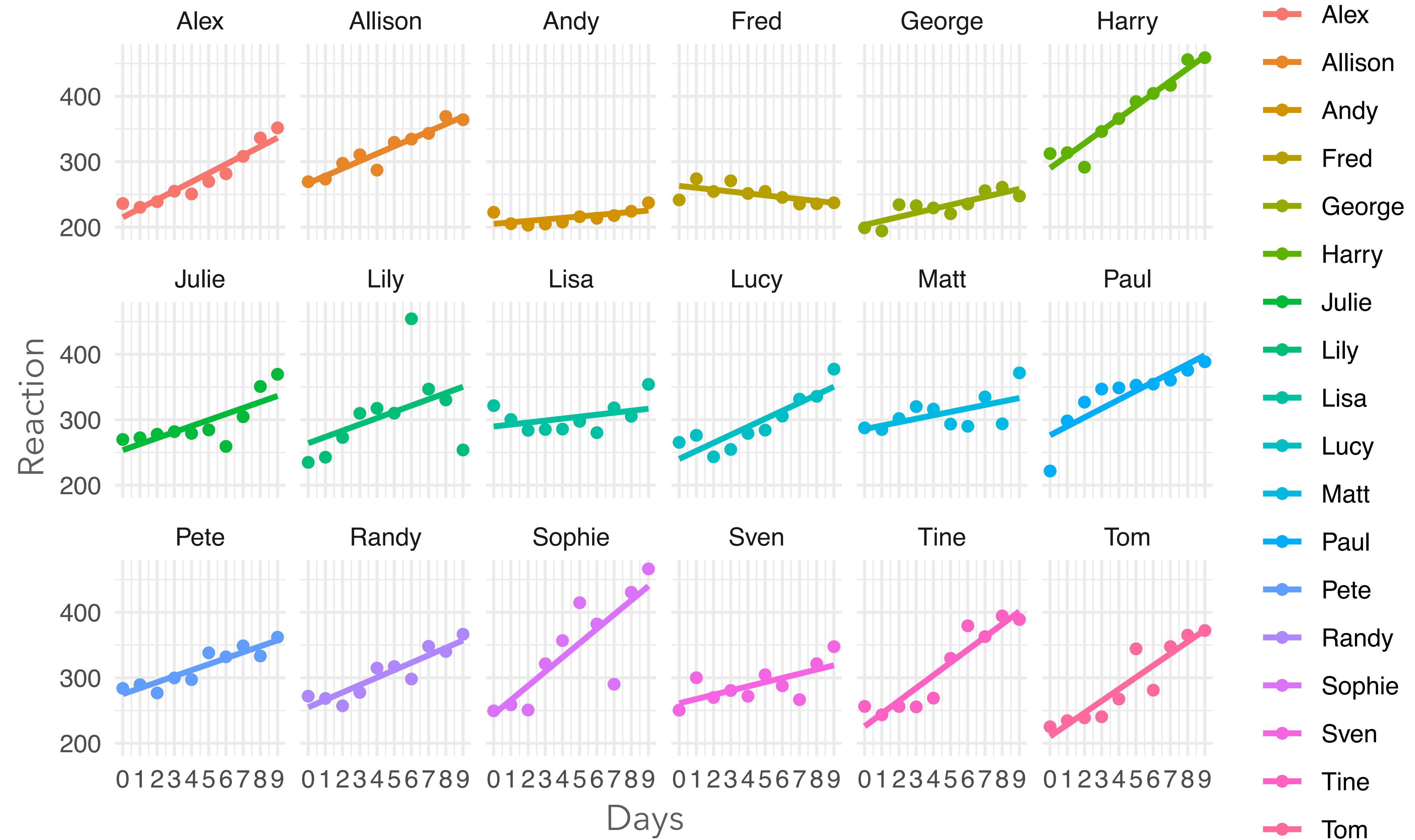
The overall level of the dependent variable
The reaction time at day zero before sleep deprivation
'Baseline reaction time'



Observations are **grouped** by the subject.

Within-subject variation is smaller than between-subject variation.





What is **generally** the effect of sleep deprivation on reaction time?
(No matter if it is Andy or Randy)

R + RStudio: language and environment for statistical computing and graphics, freely available

Preparation

The screenshot displays the RStudio interface with the following components:

- Script Editor (Top Left):** Shows the R script `lme.R` containing code to activate the `lme4` package and load the `sleepstudy` dataset.
- Environment (Top Right):** Shows the `Global Environment` pane with the `data` object, which has 180 observations and 3 variables.
- Console (Bottom Left):** Shows the R console output, identical to the script, indicating the activation of the `lme4` package and loading of the `sleepstudy` data.
- File Browser (Bottom Right):** Shows the file structure in the `skripti` folder, listing `example_sleepstudy_pixel_machine.R` and `lme.R`.

Preparation

The screenshot shows the RStudio interface with several panes:

- Code Editor (Top Left):** An R script named "lme.R" containing the following code:

```
1 # Activate the package
2 library(lme4)
3
4 # Load sleepstudy data from lme4 package
5 data <- sleepstudy
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
```
- Environment Pane (Top Right):** A red box highlights this pane. It shows the "Global Environment" with one object:

```
R > Global Environment >
data 180 obs. of 3 variables
```
- Console (Bottom Left):** Shows the same R code being run in the console.
- File Explorer (Bottom Right):** Shows the file structure in the "skripti" folder:

Name	Size	Modified
..		
example_sleepstudy_pixel_machine.R	12.5 KB	Aug 2, 2024, 11:15 AM
lme.R	648 B	Aug 2, 2024, 12:46 PM

A red annotation text "Environment to store objects, such as datasets" is placed inside the highlighted Environment pane.

Preparation

The screenshot shows the RStudio interface with several panels:

- Script Editor (Top Left):** A red box highlights the "lme.R" script editor. It contains the following R code:

```
1 # Activate the package
2 library(lme4)
3 
4 # Load sleepstudy data from lme4 package
5 data <- sleepstudy
6 
7 
8 
9 
10 
11 
12 
13 
14 
15 
16 
17 
18 
19 
20
```

Below the code, the text "Script= document to store commands" is displayed in red.
- Environment Browser (Top Right):** Shows the "Global Environment" tab. It lists a dataset named "data" with "180 obs. of 3 variables".
- File Browser (Bottom Right):** Shows the file structure at the path "Home > Desktop > lme > skripti". The contents are:| Name | Size | Modified |
| --- | --- | --- |
| .. | | |
| example_sleepstudy_pixel_machine.R | 12.5 KB | Aug 2, 2024, 11:15 AM |
| lme.R | 648 B | Aug 2, 2024, 12:46 PM |
- Console (Bottom Left):** Displays the same R code as the script editor, indicating it has been run.

Preparation

The screenshot displays the RStudio interface with the following components:

- Script Editor (Top Left):** Shows the R script `lme.R` containing code to activate the `lme4` package and load the `sleepstudy` dataset.
- Environment Browser (Top Right):** Shows the `Global Environment` tab with a data frame named `data` containing 180 observations and 3 variables.
- Console (Bottom Left):** Shows the R console output where the same commands have been run.
- File Browser (Bottom Right):** Shows the file structure in the `lme` directory, listing `example_sleepstudy_pixel_machine.R` and `lme.R`.

A red box highlights the word "package" in the script editor, and a red annotation text "There are different packages that include certain functions and/ or datasets" is overlaid on the script editor area.

```
lme.R
1 # Activate the package
2 library(lme4)
3
4 # Load sleepstudy data from lme4 package
5 data <- sleepstudy
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
```

```
R 4.3.2 · ~/ ...
> # Activate the package
> library(lme4)
> # Load sleepstudy data from lme4 package
> data <- sleepstudy
>
```

Name	Size	Modified
..		
example_sleepstudy_pixel_machine.R	12.5 KB	Aug 2, 2024, 11:15 AM
lme.R	648 B	Aug 2, 2024, 12:46 PM

Simple linear regression, main fixed effect of Days on Reaction

$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```

The screenshot shows the RStudio interface with the following components:

- Script Editor:** An R script named "lme.R" containing the following code:

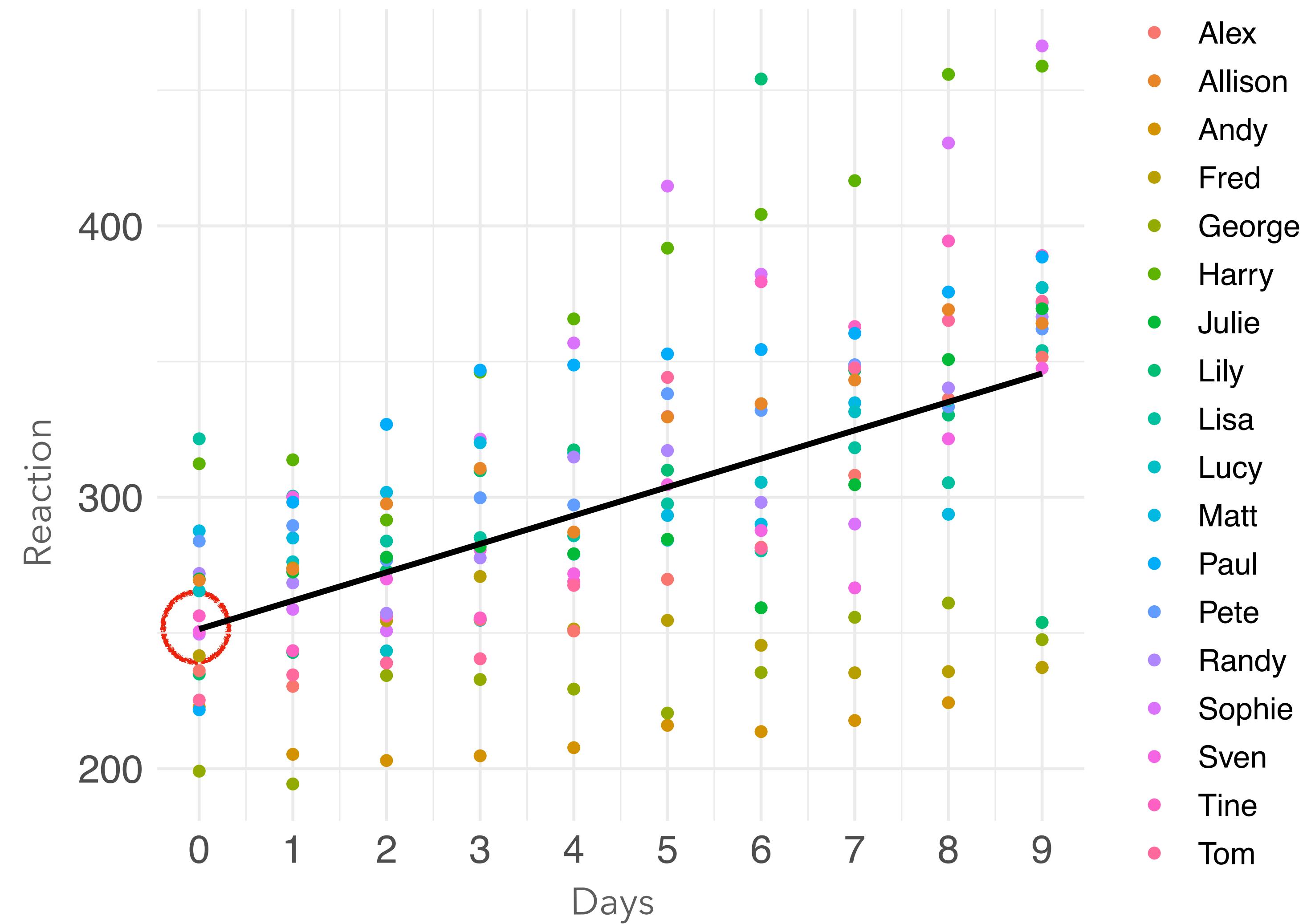
```
1 # Activate the package
2 library(lme4)
3
4 # Load sleepstudy data from lme4 package
5 data <- sleepstudy
6
7 # Model
8 fit_1 <- lm(Reaction ~ 1 + Days, data = data)
9
10 # Results
11 summary(fit_1)
12
13
14
15
16
17
18
19
20
```
- Environment:** Shows the global environment with objects "data" and "fit_1".
- Console:** Displays the R session history with the same commands as the script.
- File Browser:** Shows a directory structure under "skripti" with files "example_sleepstudy_pixel_machine.R" (12.5 KB, modified Aug 2, 2024, 11:15 AM) and "lme.R" (725 B, modified Aug 2, 2024, 12:49 PM).

$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```

Intercept

- Reaction when Days= 0

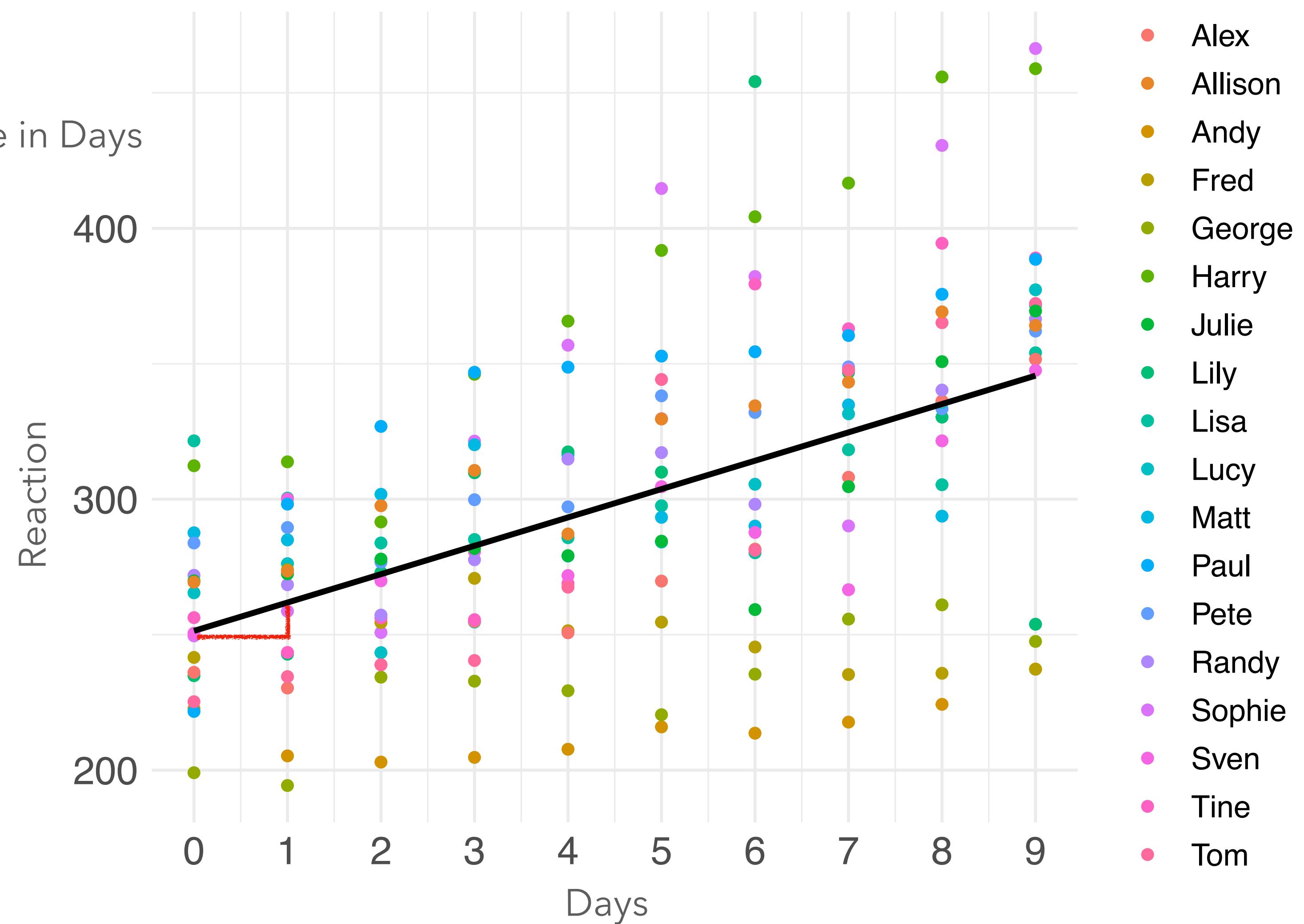


$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```

Regression coefficient

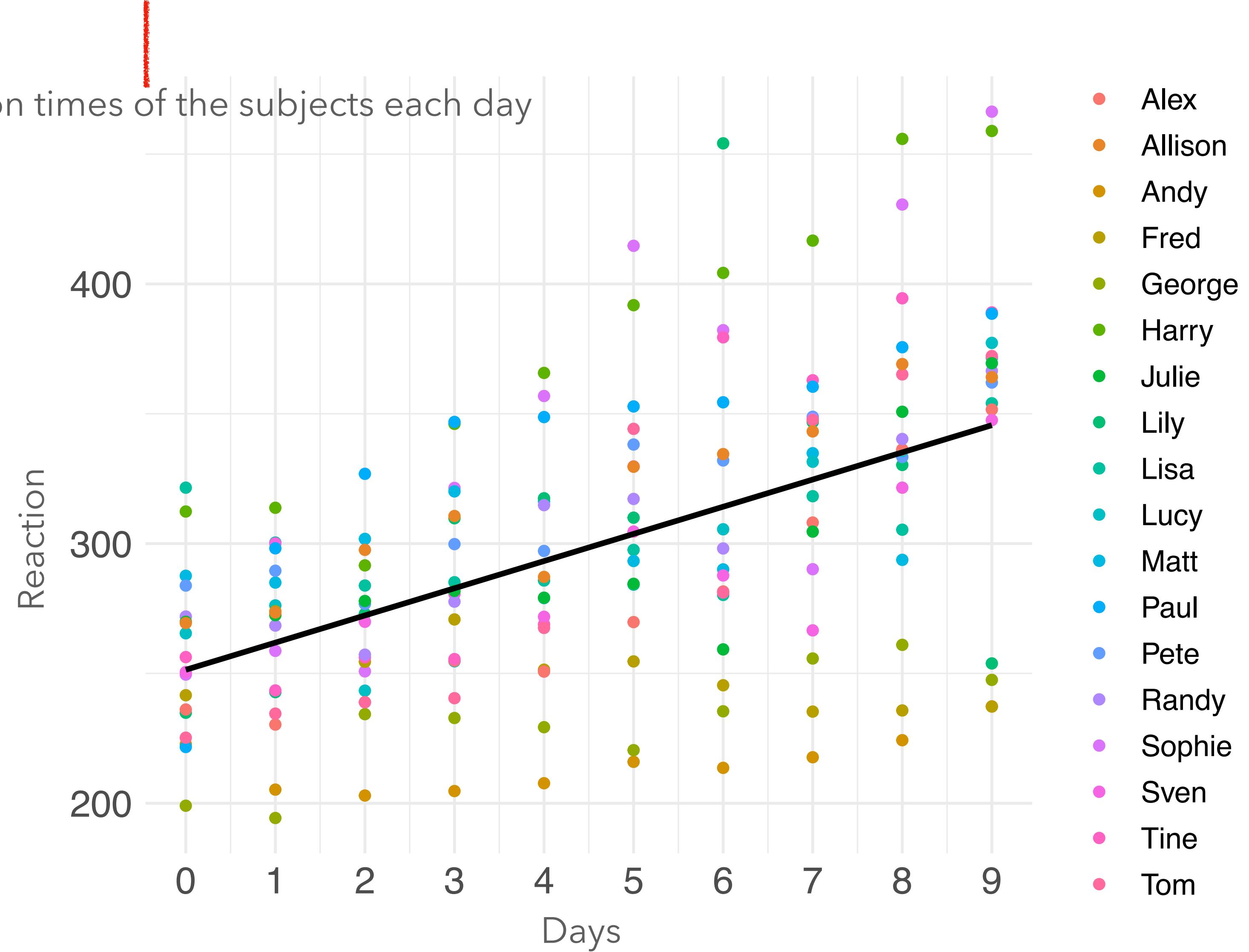
- Main effect of Days on Reaction (slope)
- The change in Reaction for one unit increase in Days



$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```

The reaction times of the subjects each day



RStudio

File Project Addins Go to file/function Run Source Environment History Connections Tutorial Import Dataset 239 MiB List R Global Environment Data data 180 obs. of 3 variables fit_1 List of 12

1 # Activate the package
2 library(lme4)
3 # Load sleepstudy data from lme4 package
4 data <- sleepstudy
5 # Model
6 fit_1 <- lm(Reaction ~ 1 + Days, data = data)
7 # Results
8 summary(fit_1)

13:1 (Top Level) R Script

Console Terminal Background Jobs R 4.3.2 · ~/
> summary(fit_1)

Call:
lm(formula = Reaction ~ 1 + Days, data = data)

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

Help pages:
[lme4::confint.merMod](#) Compute Confidence Intervals for Parameters of a [ng]lmer Fit
[lme4::glmer](#) Fitting Generalized Linear Mixed-Effects Models

Search Results R: Search Results Find in Topic

The search string was "lmer"

Vignettes:

[lme4::lmerperf](#) lme4 performance tips
[lme4::lmer](#) Fitting Linear Mixed-Effects Models using lme4

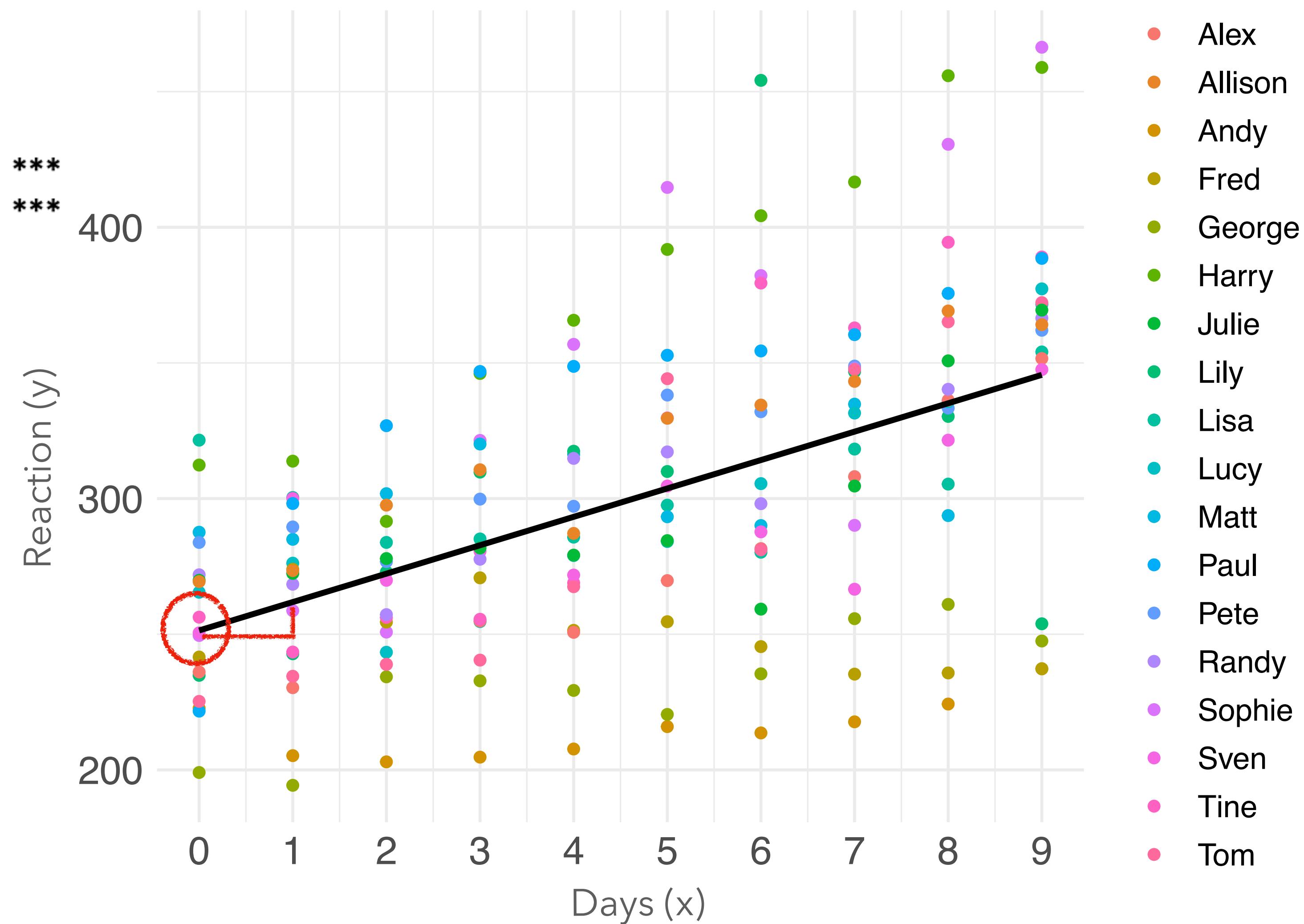
HTML source R code
PDF source R code

$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
summary(fit_1)
```

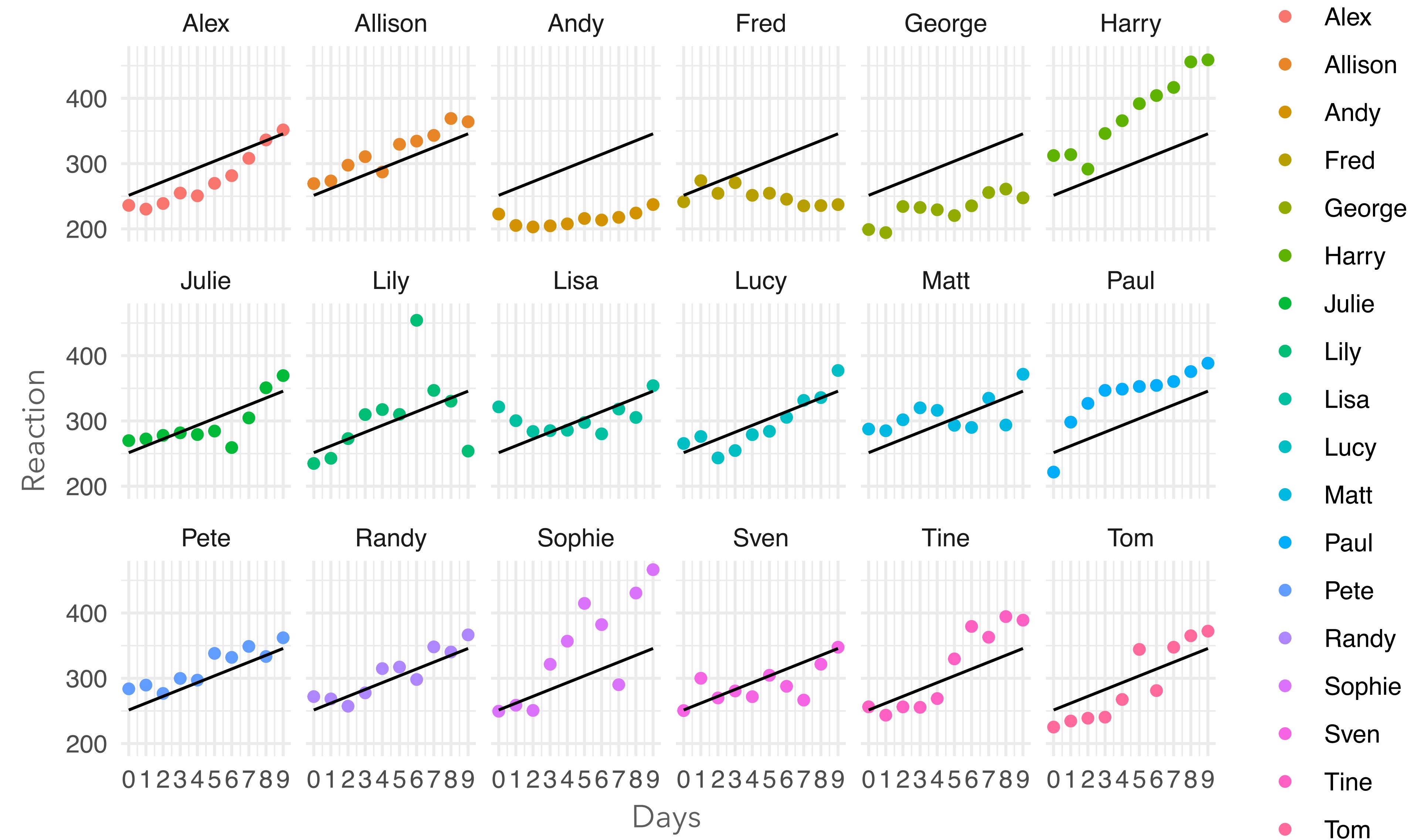
Coefficients:

	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	***



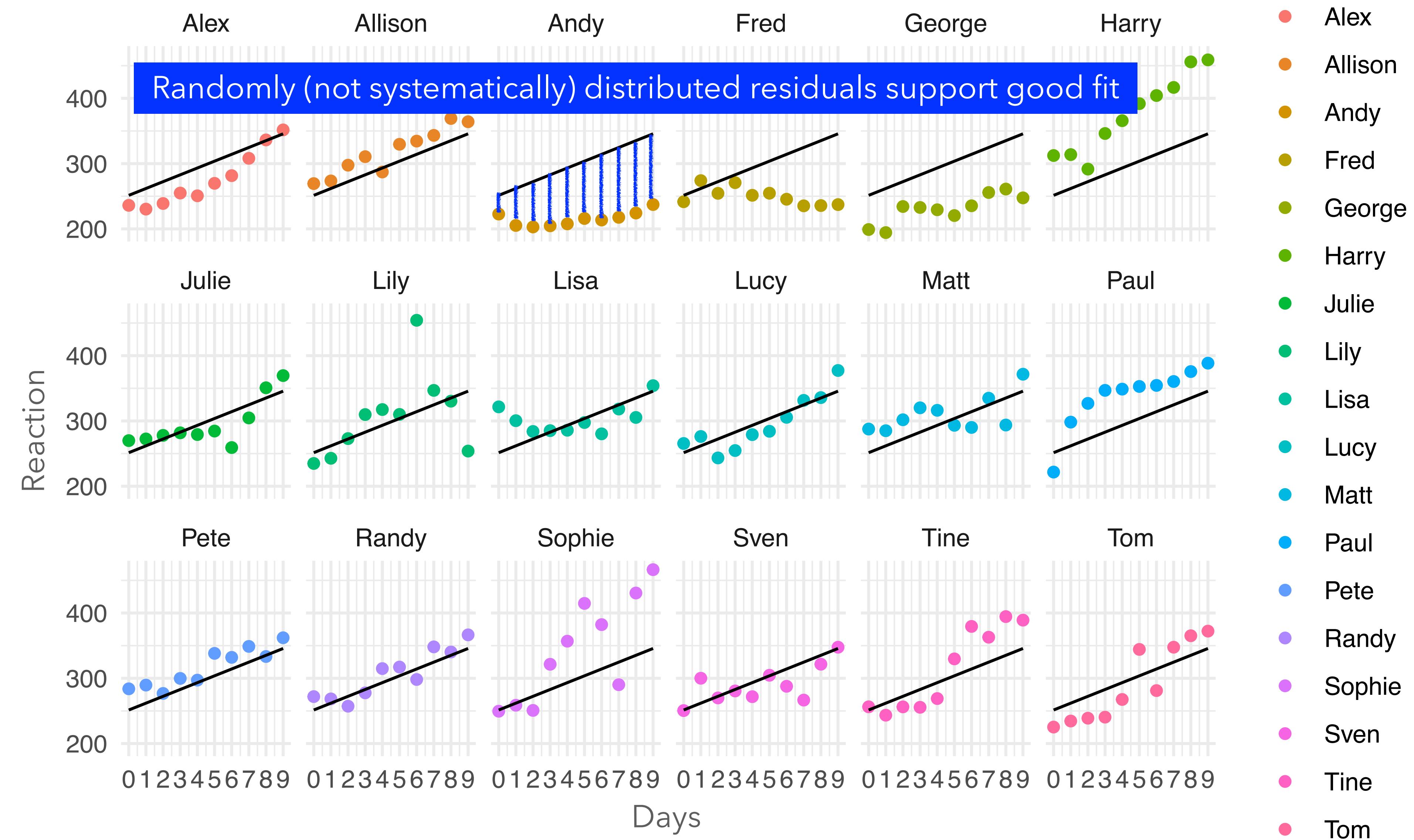
$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```



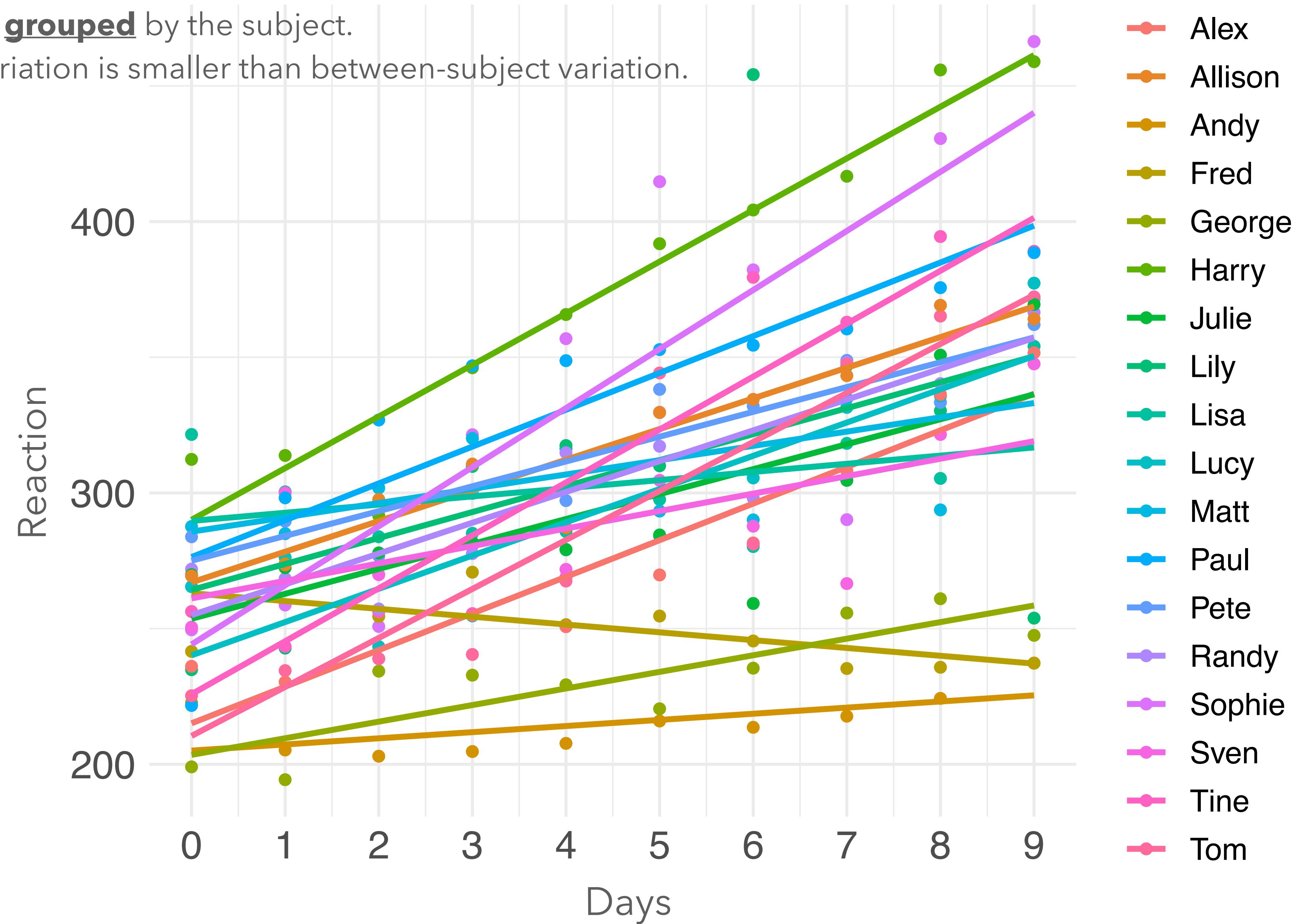
$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```



Observations are **grouped** by the subject.

Within-subject variation is smaller than between-subject variation.



$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
```

$$\text{Reaction} = (\alpha + b_{\alpha, \text{Subject}}) + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
```

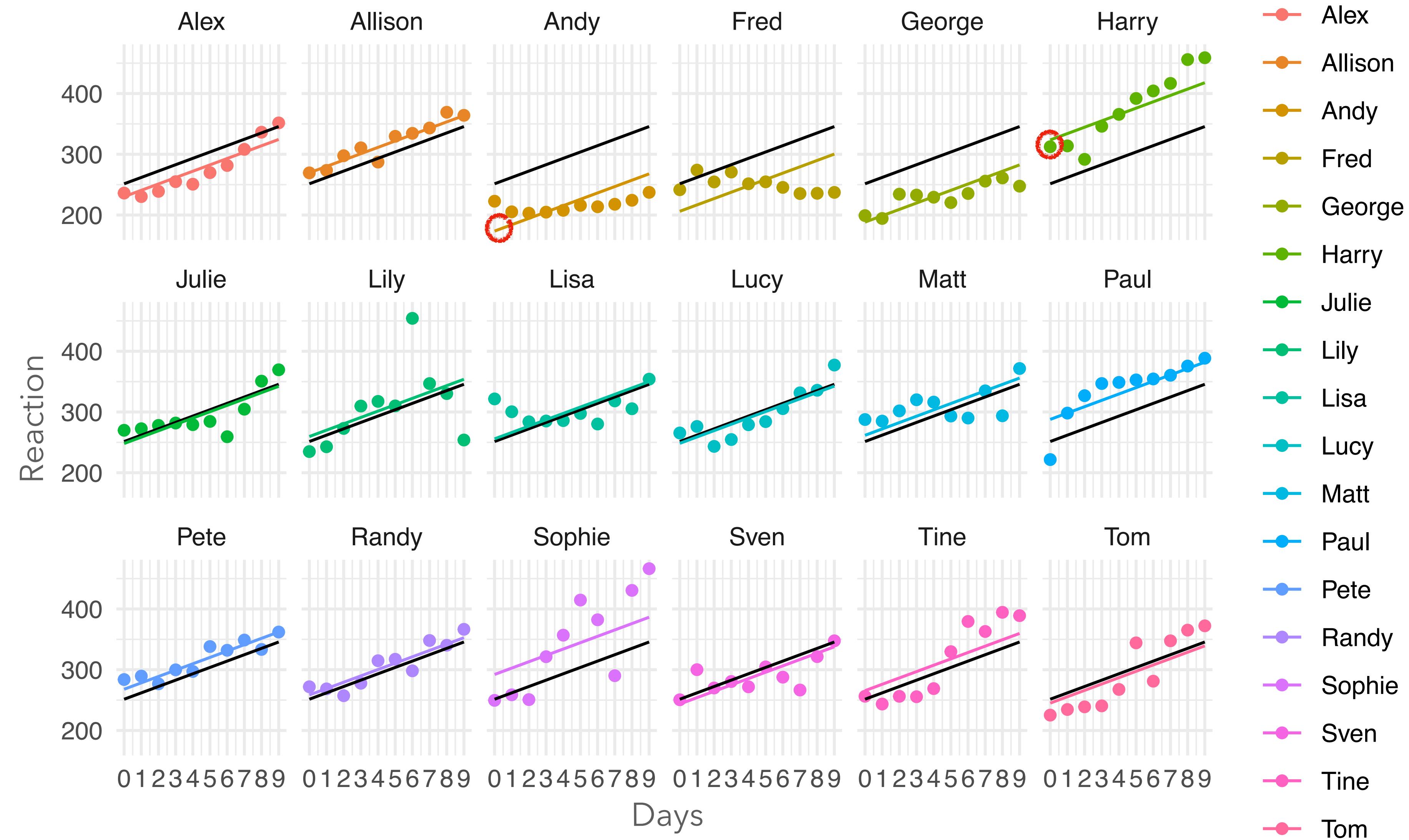
Random intercept (Subject-specific)

- Subjects are a random sample of all the possible subjects we could have had in our data
- Subject-specific intercept is a random variable

$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```

```
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
```



$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + \beta \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
```

```
summary(fit_2)
```

Random effects:

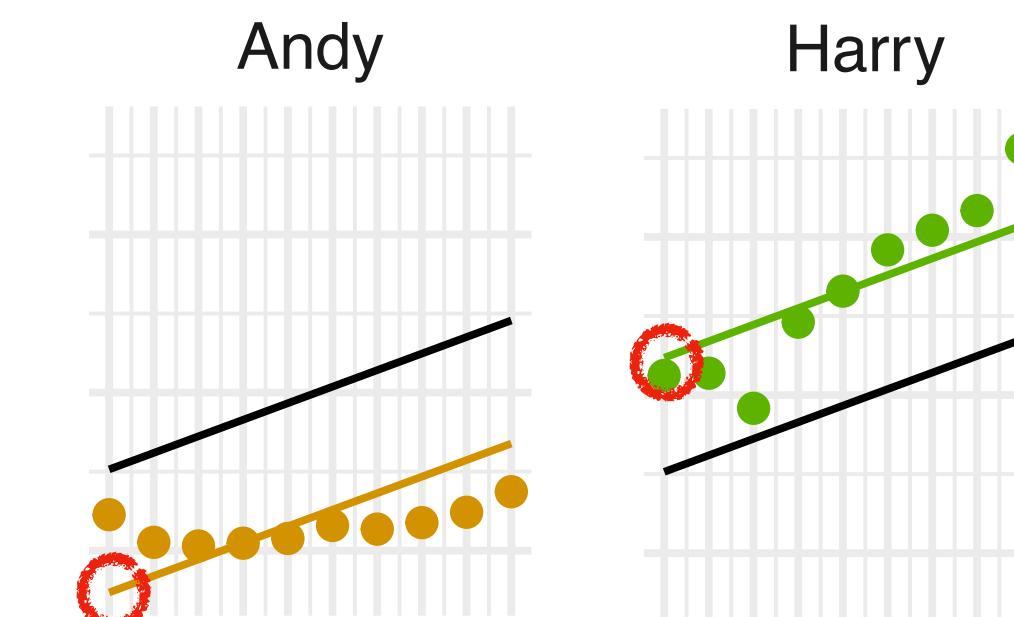
Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	1378.2	37.12
	Residual	960.5	30.99

Number of obs: 180, groups: Subject, 18

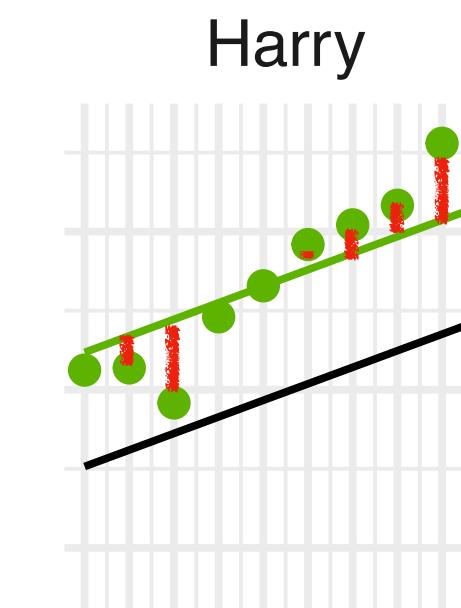
Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.4051	9.7467	25.79
Days	10.4673	0.8042	13.02

Between-subject variation in the intercept (on average)



Within-subject variation in the residuals (on average)



$$\text{Reaction} = \alpha + (\beta + b_{\beta, \text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject) , data = data)
```

$$\text{Reaction} = \alpha + (\beta + b_{\beta, \text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject) , data = data)
```



Random slope and not intercept (Subject-specific)

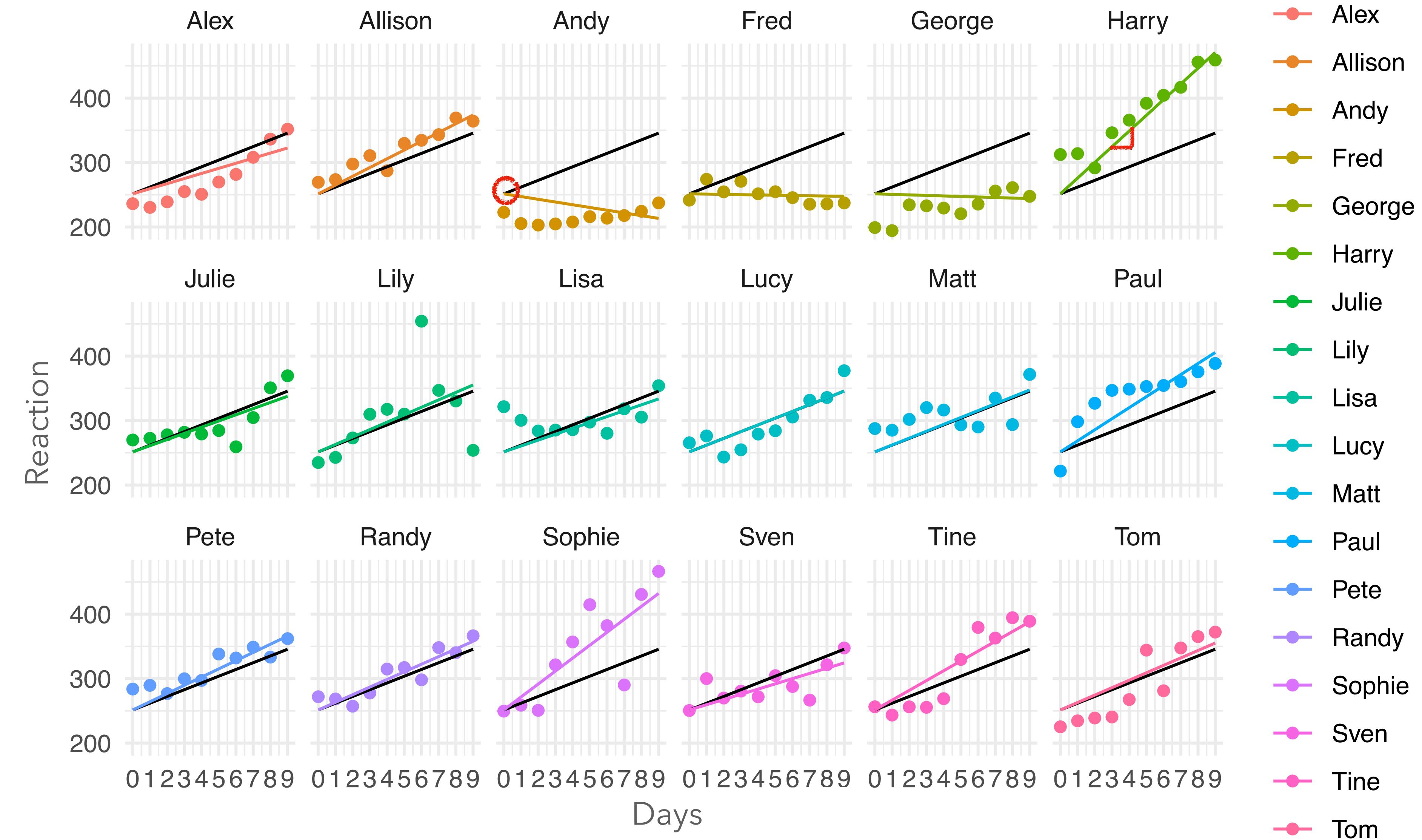
- Subject-specific effect of Days on Reaction
- Subject-specific slope is a random variable

$$\text{Reaction} = \alpha + (\beta + b_{\beta, \text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```

```
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
```

```
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject), data = data)
```



$$\text{Reaction} = \alpha + (\beta + b_{\beta, \text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject), data = data)
```

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	Days	52.71	7.26
Residual		842.03	29.02

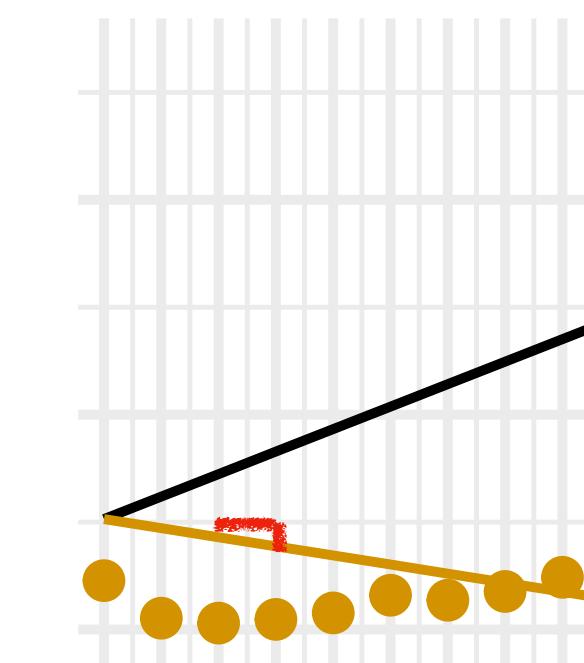
Number of obs: 180, groups: Subject, 18

Fixed effects:

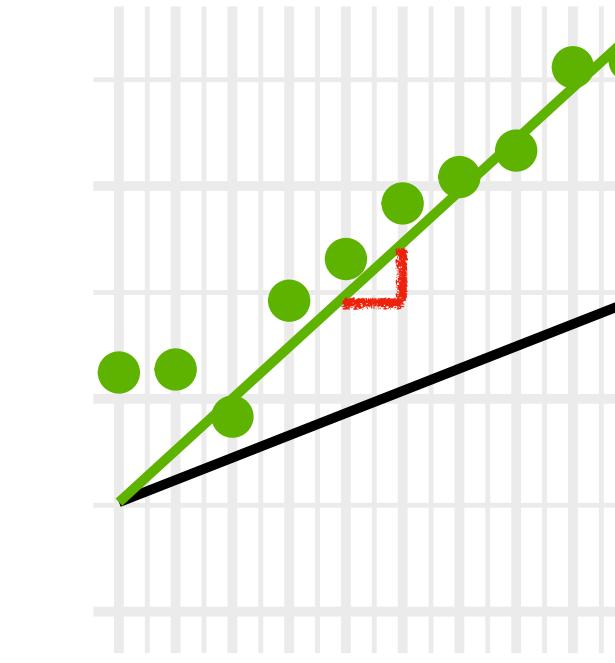
	Estimate	Std. Error	t value
(Intercept)	251.41	4.02	62.539
Days	10.47	1.87	5.599

Between-subject variation in the slope (on average)

Andy

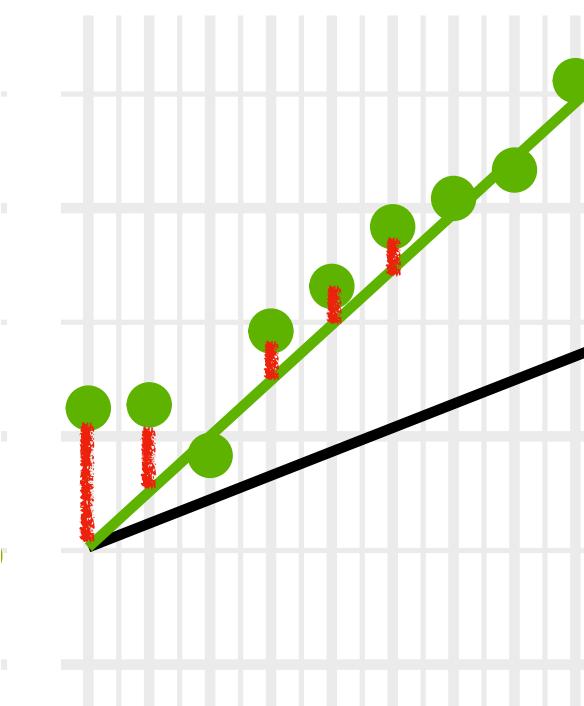


Harry



Within-subject variation in the residuals (on average)

Harry



$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + (\beta + b_{\beta,\text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject) , data = data)
fit_4 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject) , data = data)
```

$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + (\beta + b_{\beta,\text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject), data = data)
fit_4 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject), data = data)
```

Random intercept and slope (Subject-specific)

- Subject-specific intercept and effect of Days on Reaction

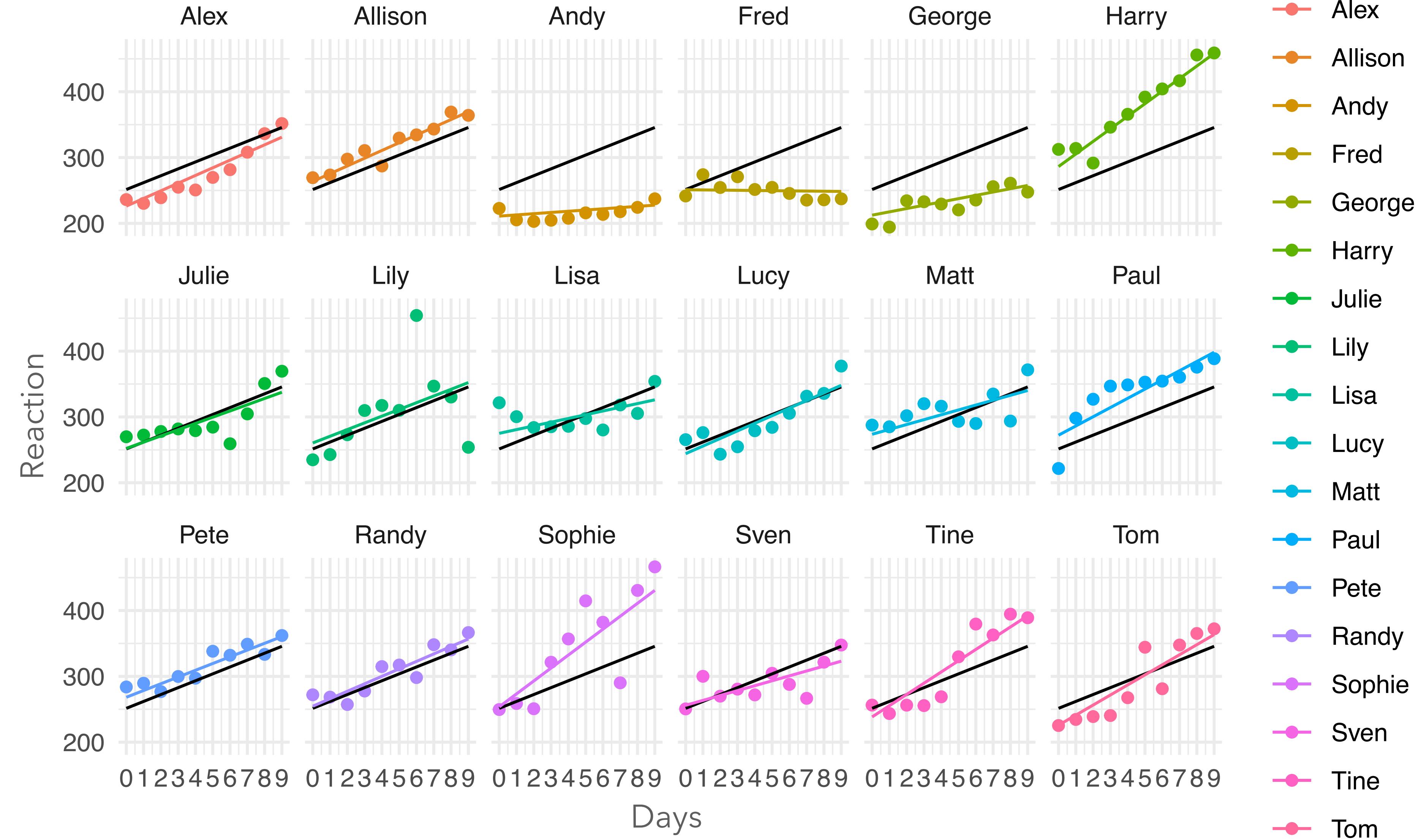
```
fit_1 <- lm(Reaction ~ 1 + Days, data = data)
```

```
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
```

```
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject) , data = data)
```

```
fit_4 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject) , data = data)
```

$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + (\beta + b_{\beta,\text{Subject}}) \cdot \text{Days} + \varepsilon$$



```

fit_1 <- lm(Reaction ~ 1 + Days, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 | Subject), data = data)
fit_3 <- lmer(Reaction ~ 1 + Days + (0 + Days | Subject) , data = data)
fit_4 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject) , data = data)

```

$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + (\beta + b_{\beta,\text{Subject}}) \cdot \text{Days} + \varepsilon$$

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	612.10	24.741	
	Days	35.07	5.922	0.07
	Residual	654.94	25.592	

Number of obs: 180, groups: Subject, 18

Between-subject variation in the intercept (on average)

Between-subject variation in the slope (on average)

Within-subject variation in the residuals (on average)

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.405	6.825	36.838
Days	10.467	1.546	6.771

Linear model:

Reaction ~ 1 + Days : linear regression with a fixed effect

Linear **mixed effects** models:

Reaction ~ 1 + Days + (1 | Subject) : linear regression with a fixed effect and a random effect (intercept)

Reaction ~ 1 + Days + (0 + Days | Subject) : linear regression with a fixed effect and a random effect (slope)

Reaction ~ 1 + Days + (1 + Days | Subject) : linear regression with a fixed effect and random effects (intercept and slope)

Mixed effects (aka hierarchical) model: both fixed and **random** effects (intercept and/ or slope)

Fixed effect

Aka. Population-level effect

Main effect:

- The effect of age on receptors
- Applied to all subjects
- Parameter

Interaction:

- Does the effect of age on receptor availability depend on sex
- The effect of age on receptors
- Separately for sexes
- Parameter

Random effect

Random intercept:

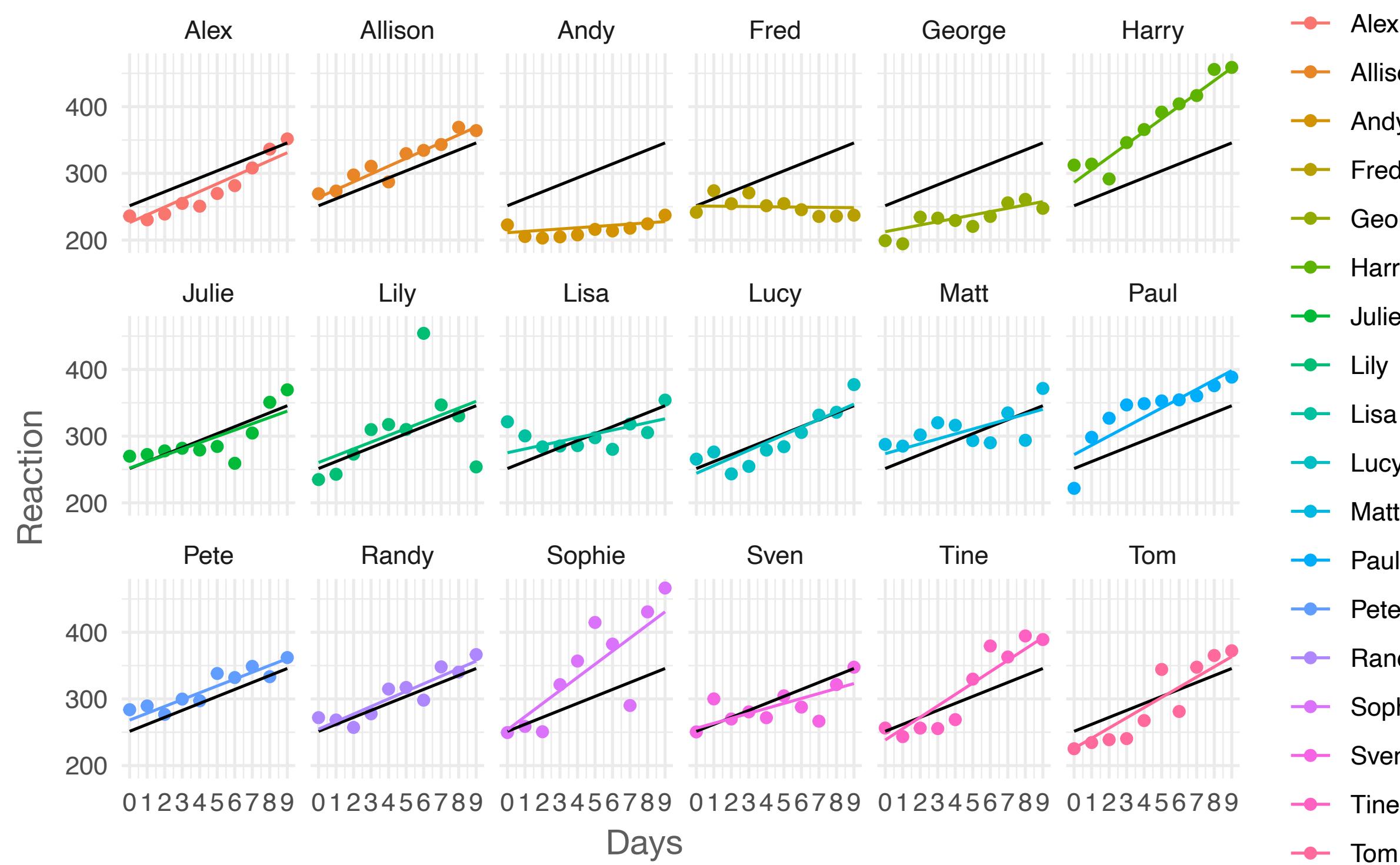
- The overall level of receptors may vary between subjects
- Subjects are a random sample of the population of subjects
- Random variable

Random slope:

- The age effect may vary between subjects
- Subjects are a random sample of the population of subjects
- Random variable

What is **generally** the effect of sleep deprivation on reaction time?
= Interest in the main effect (black slope)

What's the point of modeling the random intercept and slope? (colored slopes)

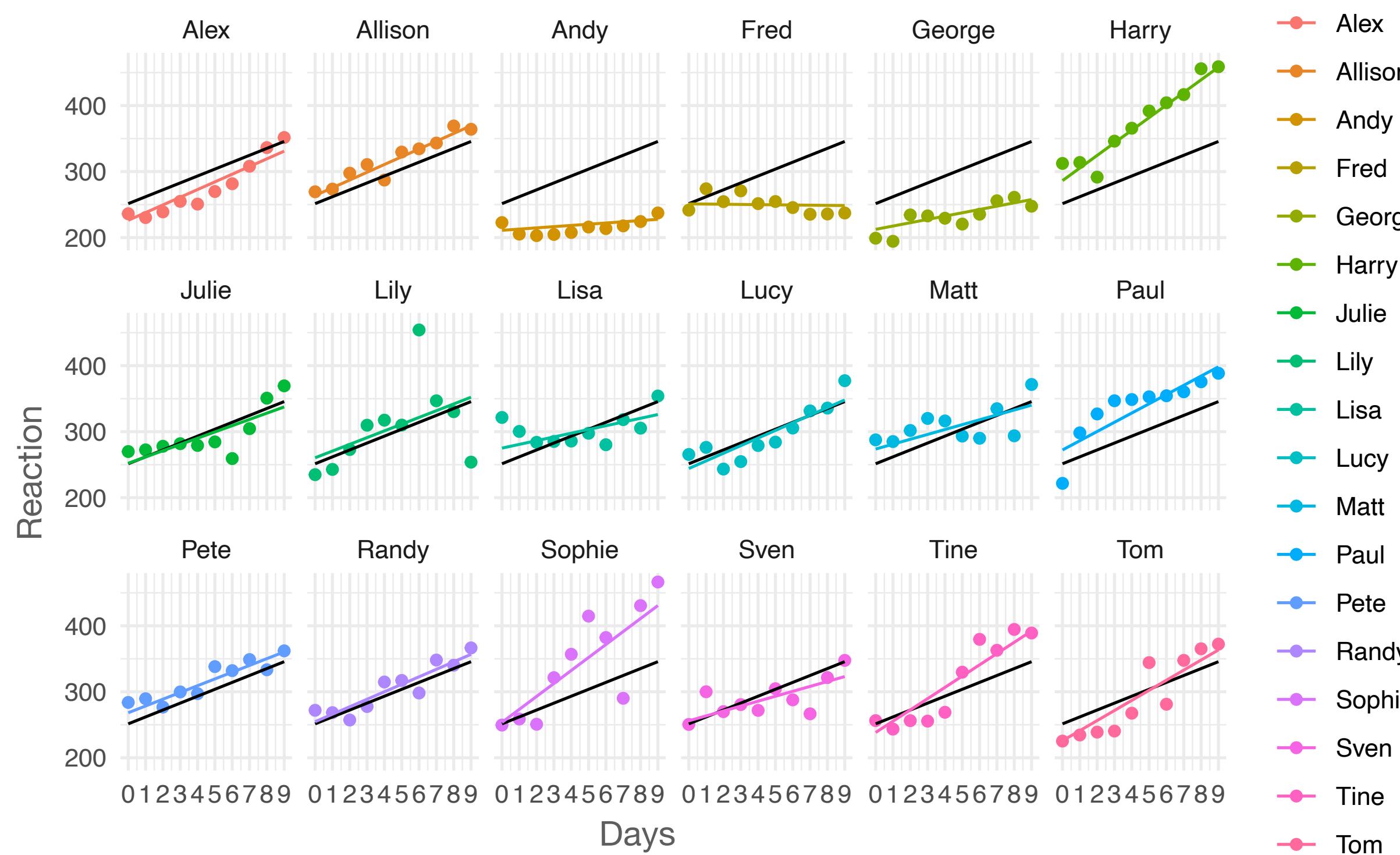


What is **generally** the effect of sleep deprivation on reaction time?

= Interest in the main effect (black slope)

What's the point of modeling the random intercept and slope? (colored slopes)

We get an idea of how much the subjects **vary on average**



summary(fit)

Random effects:

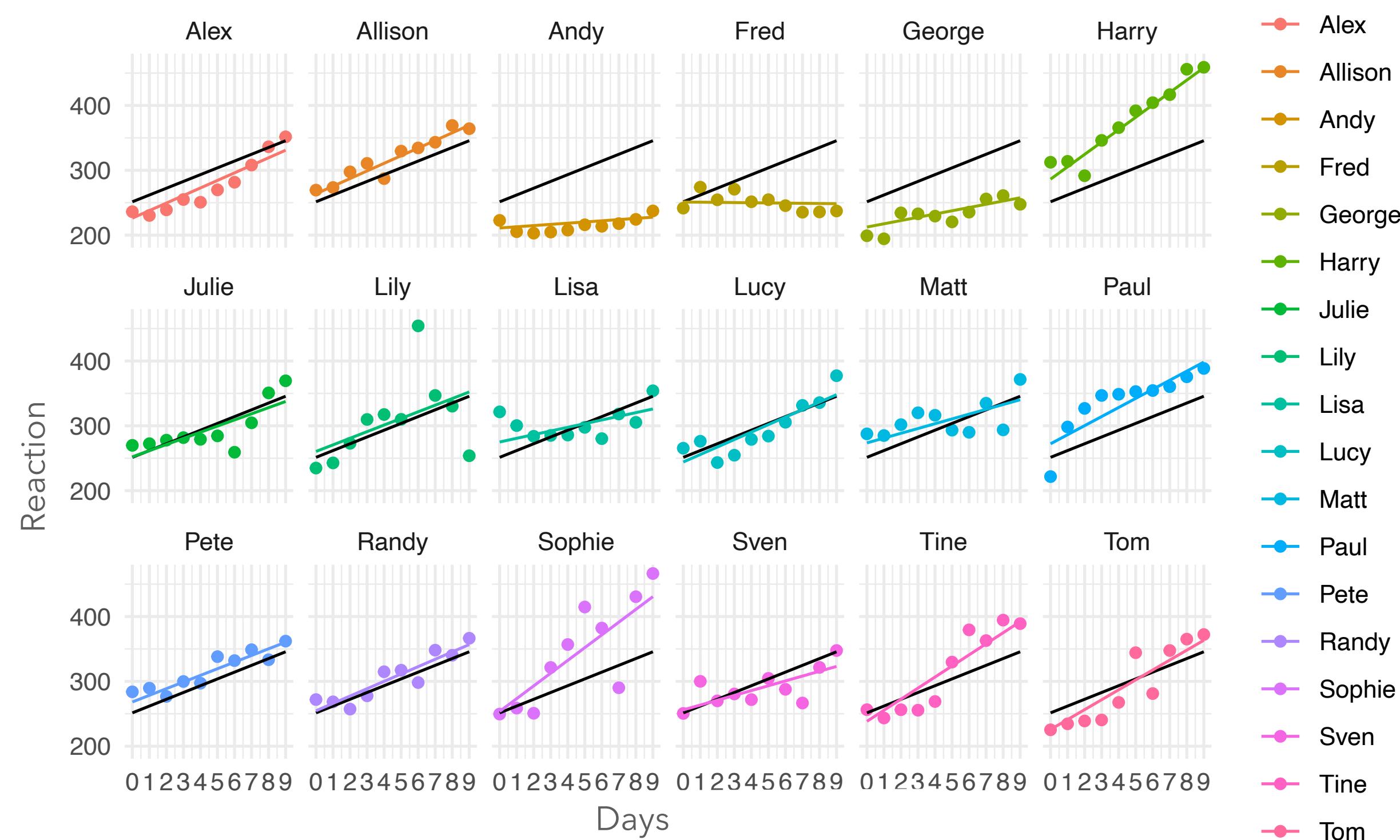
Groups	Name
Subject	(Intercept)
	Days

Variance	Std.Dev.
612.10	24.741
35.07	5.922
654.94	25.592

Residual

What are the **subject-specific** effects of sleep deprivation on reaction time?
= Interest in the subject-specific effects (colored slopes)

In this case, why not model **interaction** instead of the random intercept and slope?



Fixed interaction of days and subject:

Model the effect of Days on Reaction, separately for each subject

Random intercept and slope:

Model the effect of days on reaction time, allowing the reaction time and the effect of Days on Reaction to vary between subjects

What's the difference?

$$\text{Reaction} = \alpha + \beta_1 \cdot \text{Days} + \beta_2 \cdot \text{Subject} + \beta_3 \cdot \text{Days} \cdot \text{Subject} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days*Subject, data = data)
```

$$\text{Reaction} = \alpha + \beta_1 \cdot \text{Days} + \beta_2 \cdot \text{Subject} + \beta_3 \cdot \text{Days} \cdot \text{Subject} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days*Subject, data = data)
```

Interaction

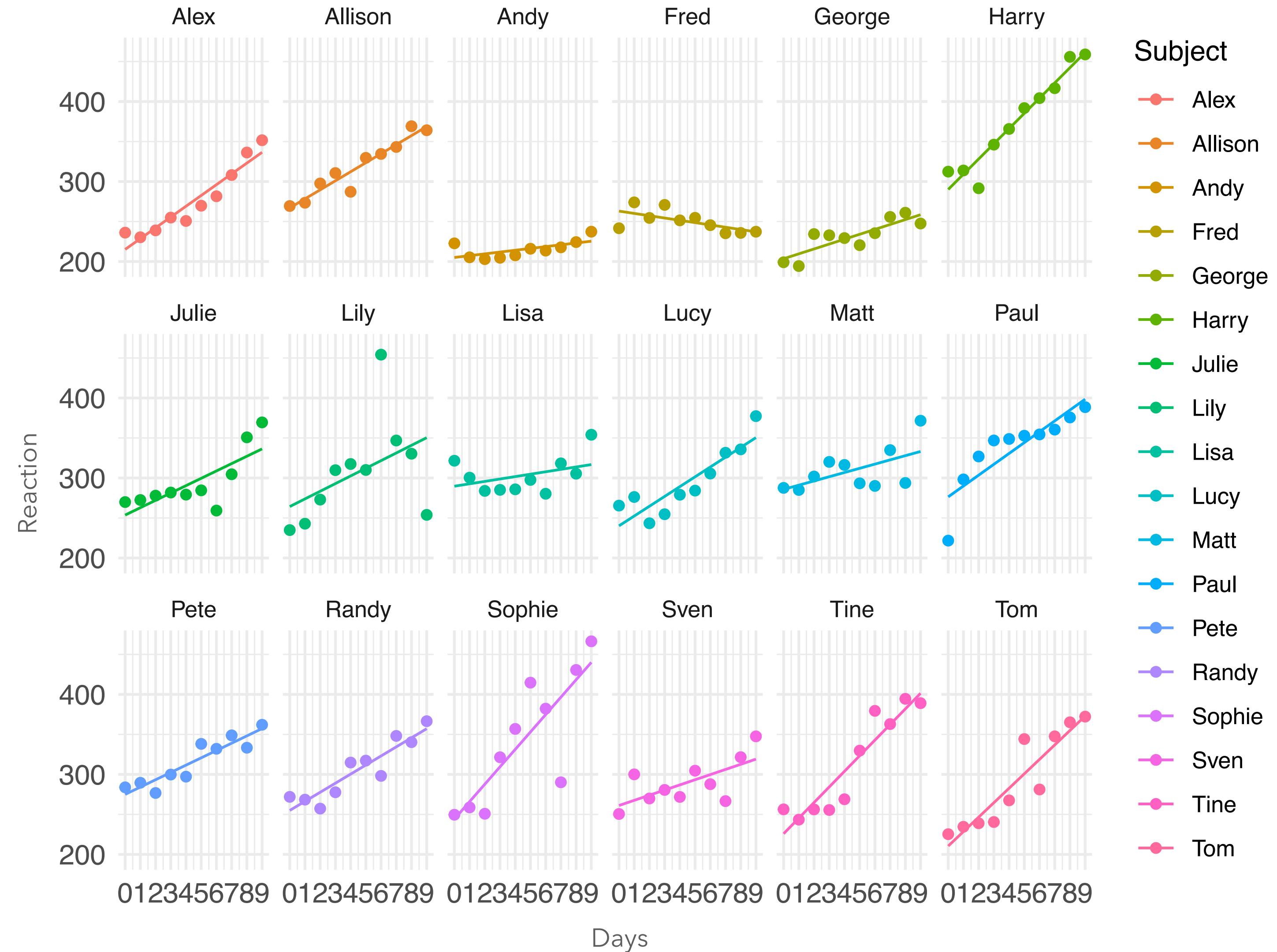
- Effect of Days on Reaction, separately for each Subject

$$\text{Reaction} = \alpha + \beta_1 \cdot \text{Days} + \beta_2 \cdot \text{Subject} + \beta_3 \cdot \text{Days} \cdot \text{Subject} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days*Subject, data = data)
```

- Minimize the residuals
- Sum of the absolute values of the residuals = 0

- Main effect missing



$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + (\beta + b_{\beta,\text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days*Subject, data = data)
fit_2 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject) , data = data)
```

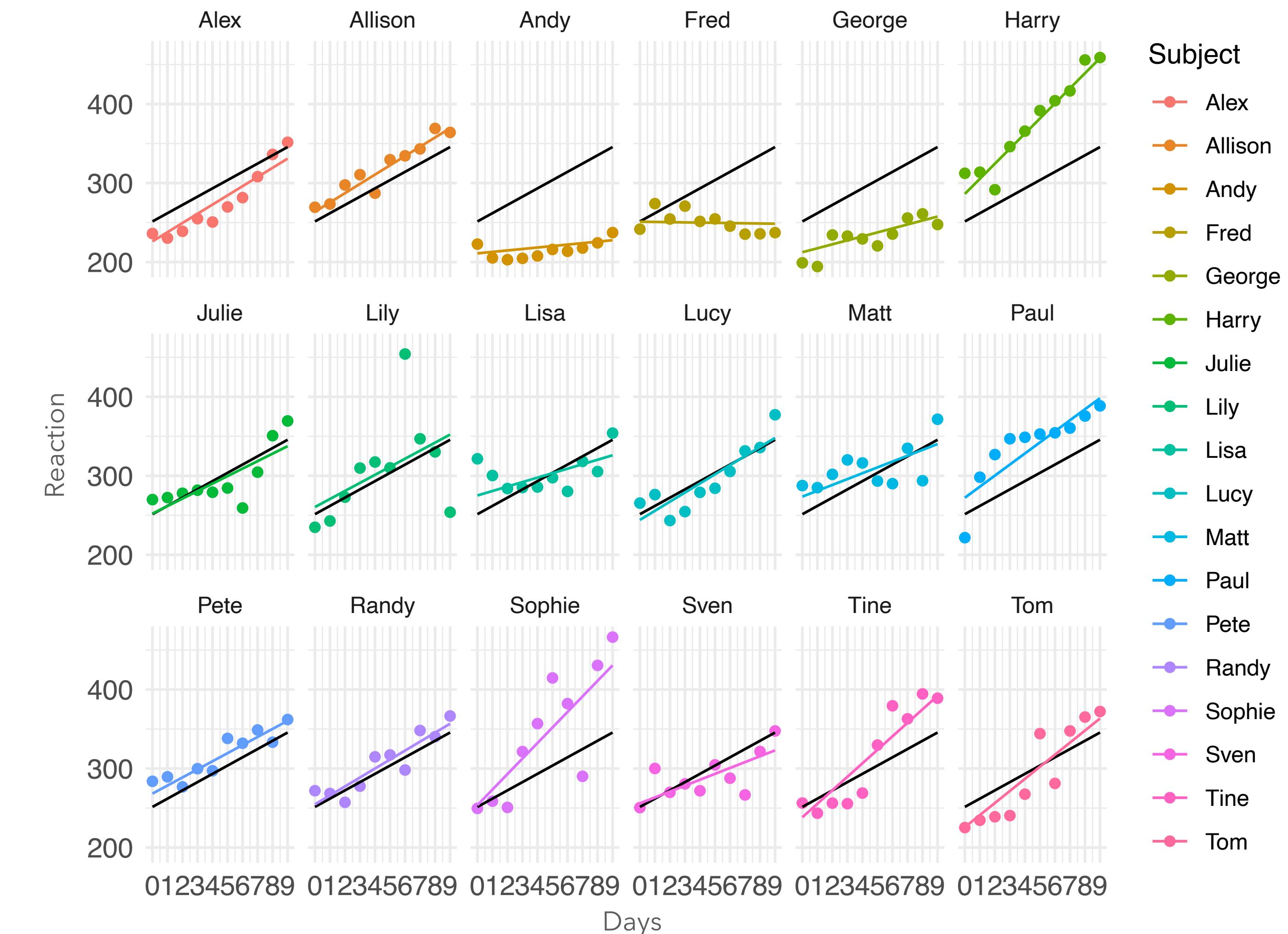
$$\text{Reaction} = (\alpha + b_{\alpha,\text{Subject}}) + (\beta + b_{\beta,\text{Subject}}) \cdot \text{Days} + \varepsilon$$

```
fit_1 <- lm(Reaction ~ 1 + Days*Subject, data = data)
```

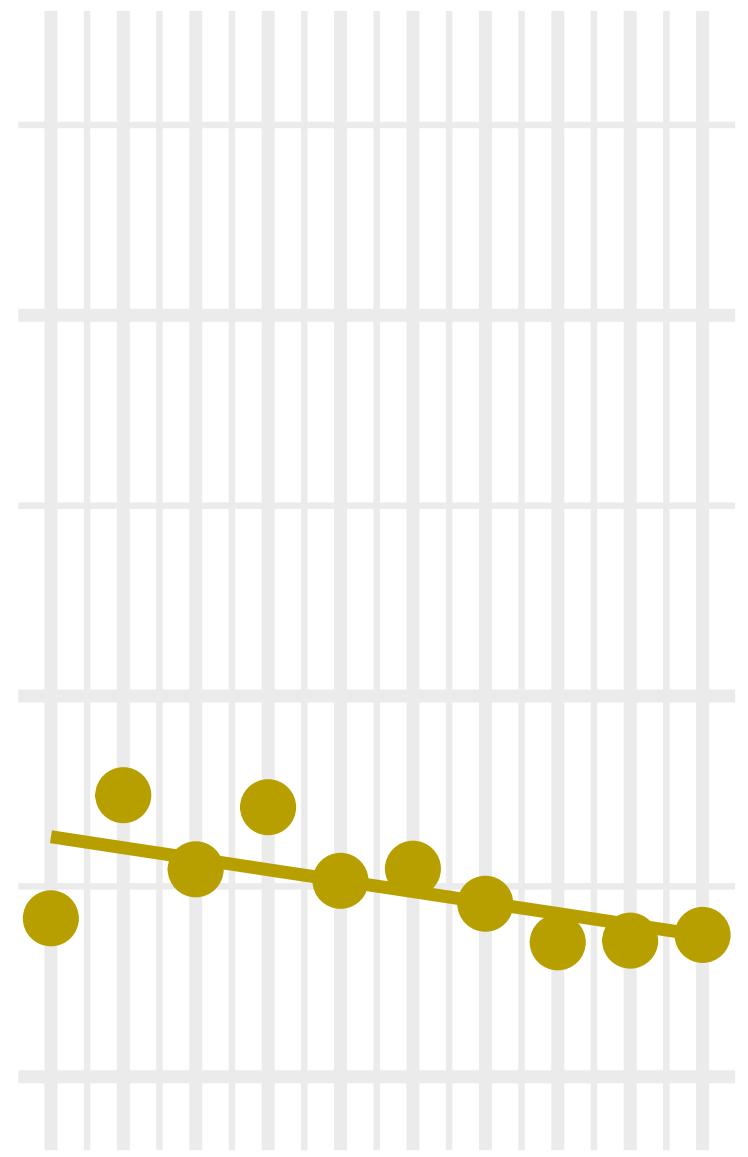
```
fit_2 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject) , data = data)
```

- Does not minimize the residuals
- Sum of the absolute values of the residuals $\neq 0$
- Shrinkage towards the mean (main effect)

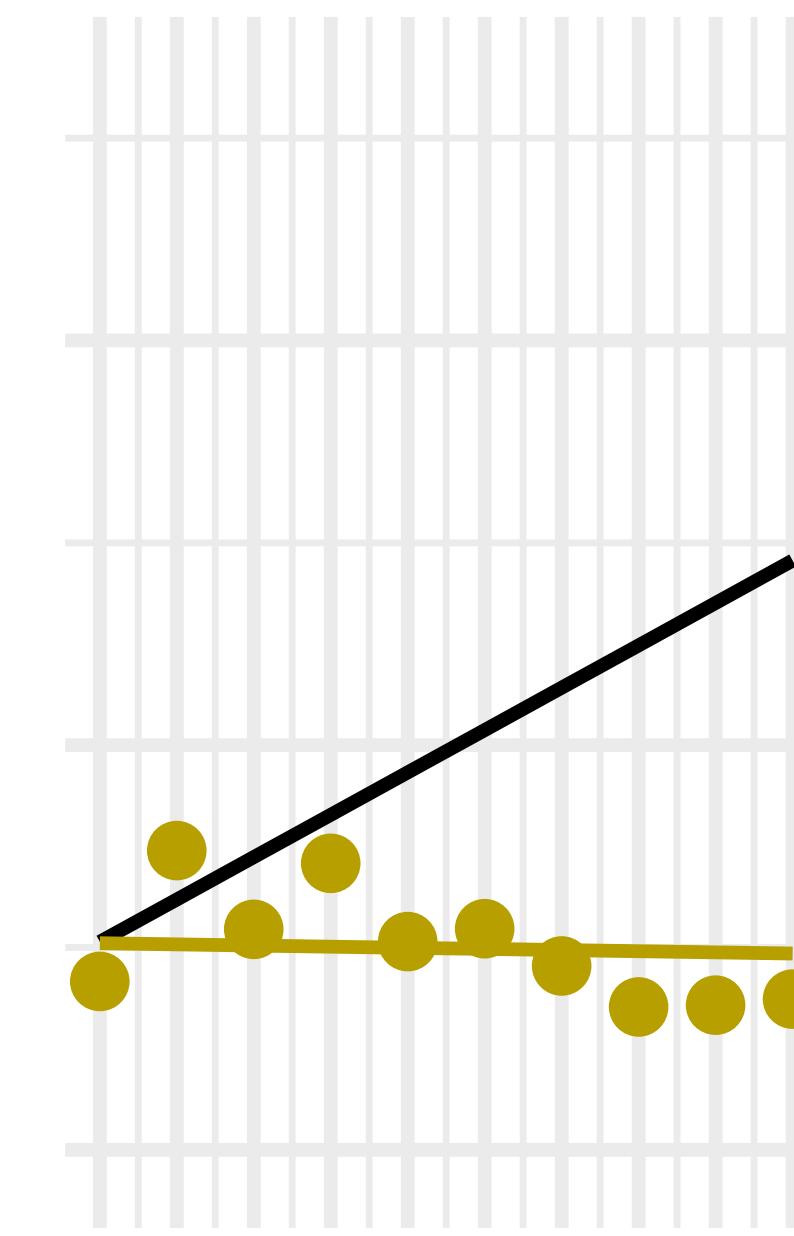
- Main effect



Fred



Fred



Fixed interaction

```
fit_1 <- lm(Reaction ~ 1 + Days*Subject, data = data)
```

Random intercept and slope

```
fit_2 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject) , data = data)
```

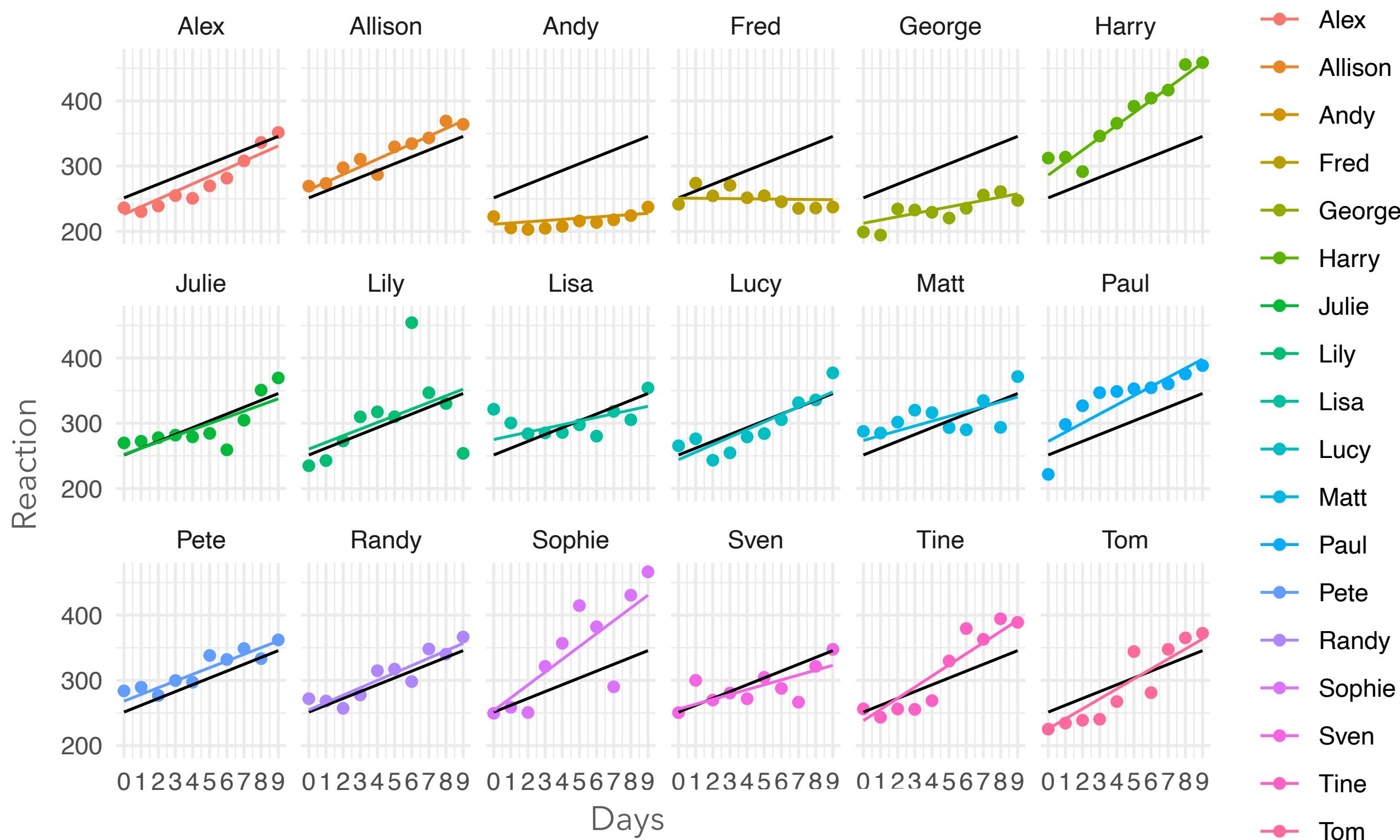
What are the **subject-specific** effects of sleep deprivation on reaction time?
= Interest in the subject-specific effects (colored slopes)

In this case, why not model **interaction** instead of the random intercept and slope?

Extreme, rare observations might be unreliable

-> Utilize the information from others

-> Random effects are pooled towards the main effect (shrinkage towards the mean)



However, both models could work
• Goals and preferences matter

Reporting results

```
fit_4 <- lmer(Reaction ~ 1 + Days + (1 + Days | Subject) , data = data)
summary(fit_4)
```

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	612.10	24.741	
	Days	35.07	5.922	0.07
Residual		654.94	25.592	

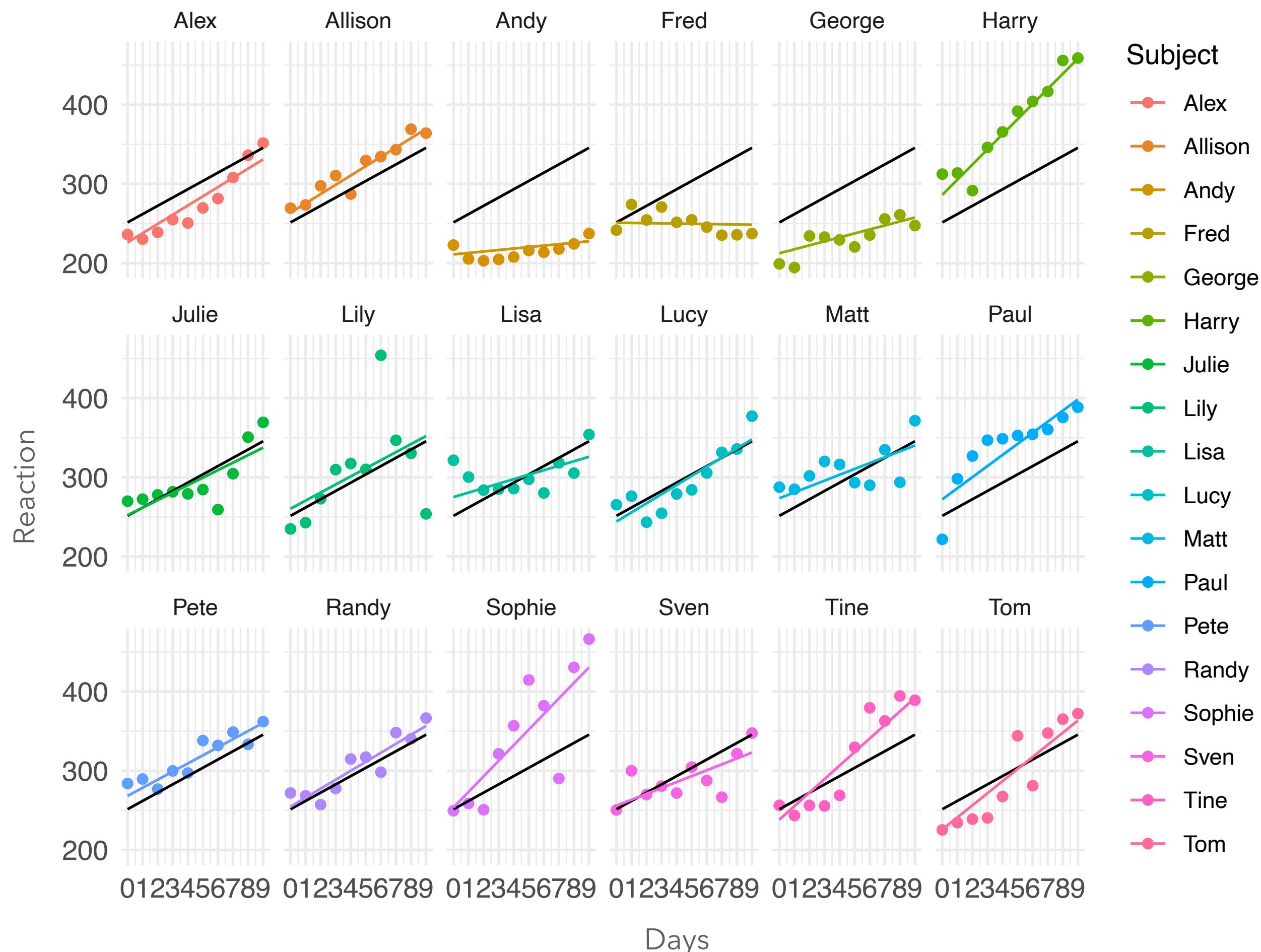
Number of obs: 180, groups: Subject, 18

Fixed effects:

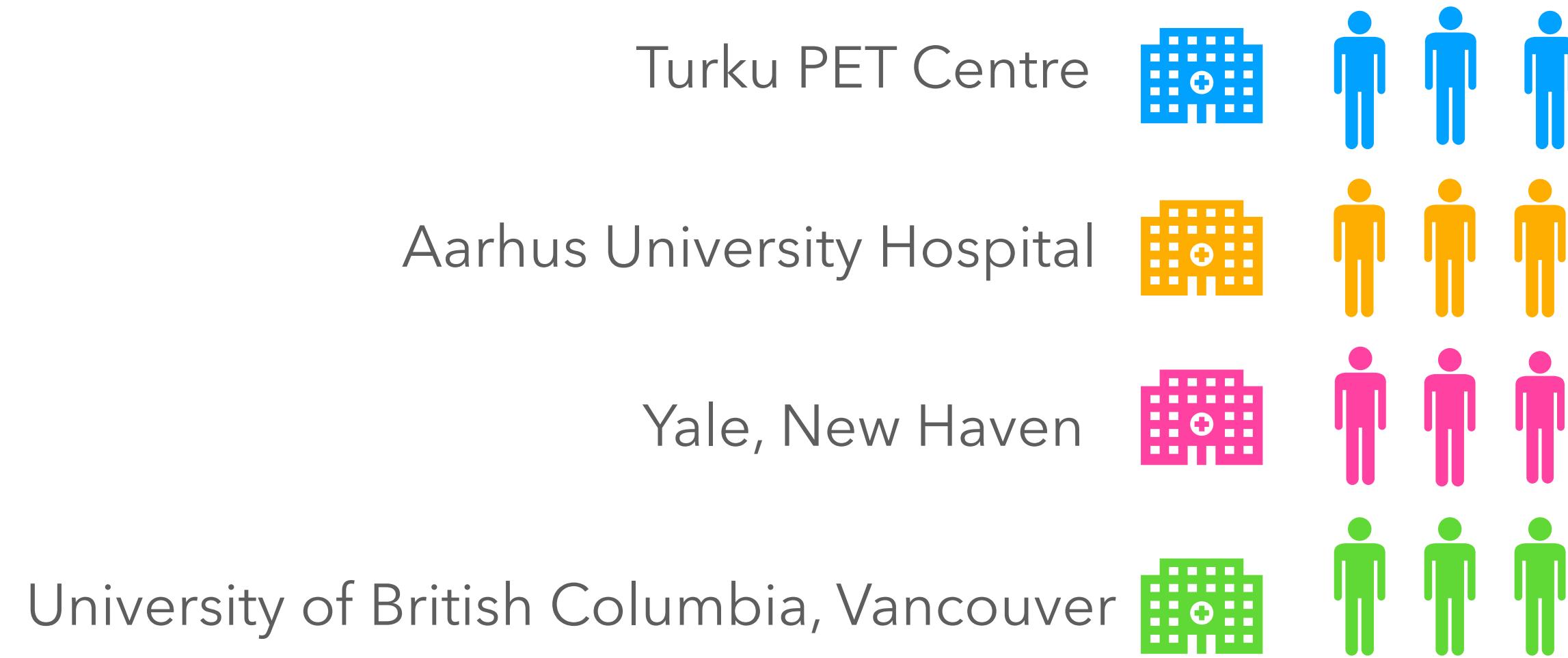
	Estimate	Std. Error	t value
(Intercept)	251.405	6.825	36.838
Days	10.467	1.546	6.771

Based on the data and our modeling:

- The baseline reaction time is 251 ms and it varies 25 ms between the subjects, on average.
- With each day of sleep deprivation, the reaction time becomes 10 ms longer, while the effect on average varies 6 ms between the subjects.



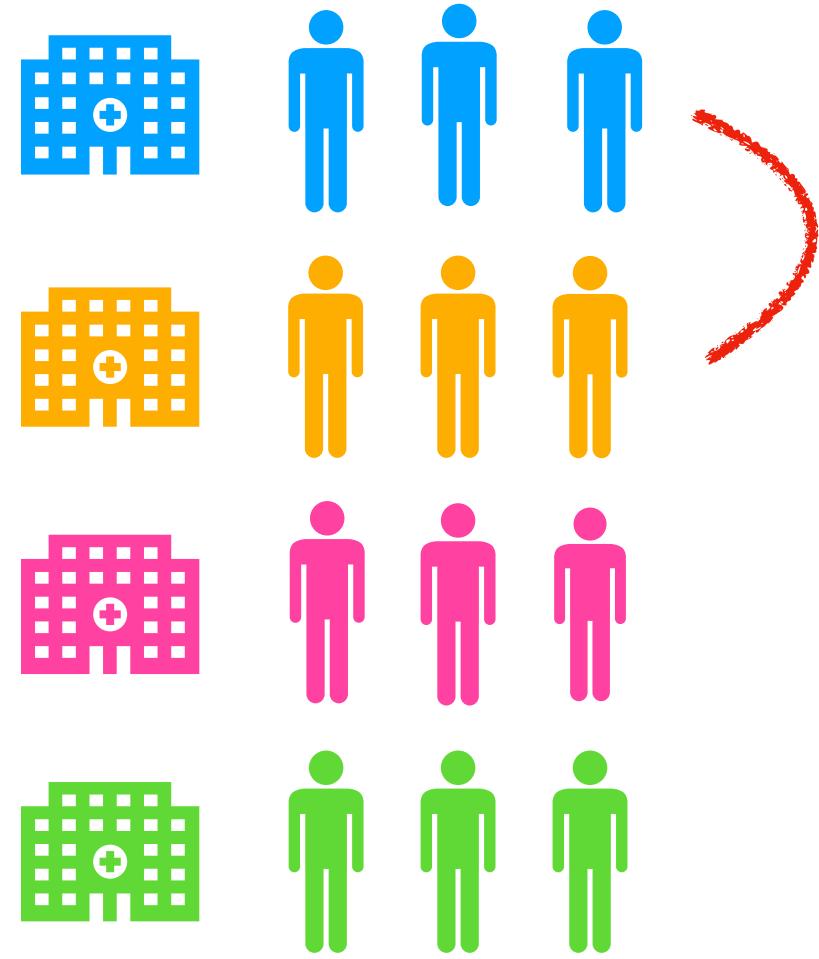
Fixed or random?



Fixed or random?

Fixed, if

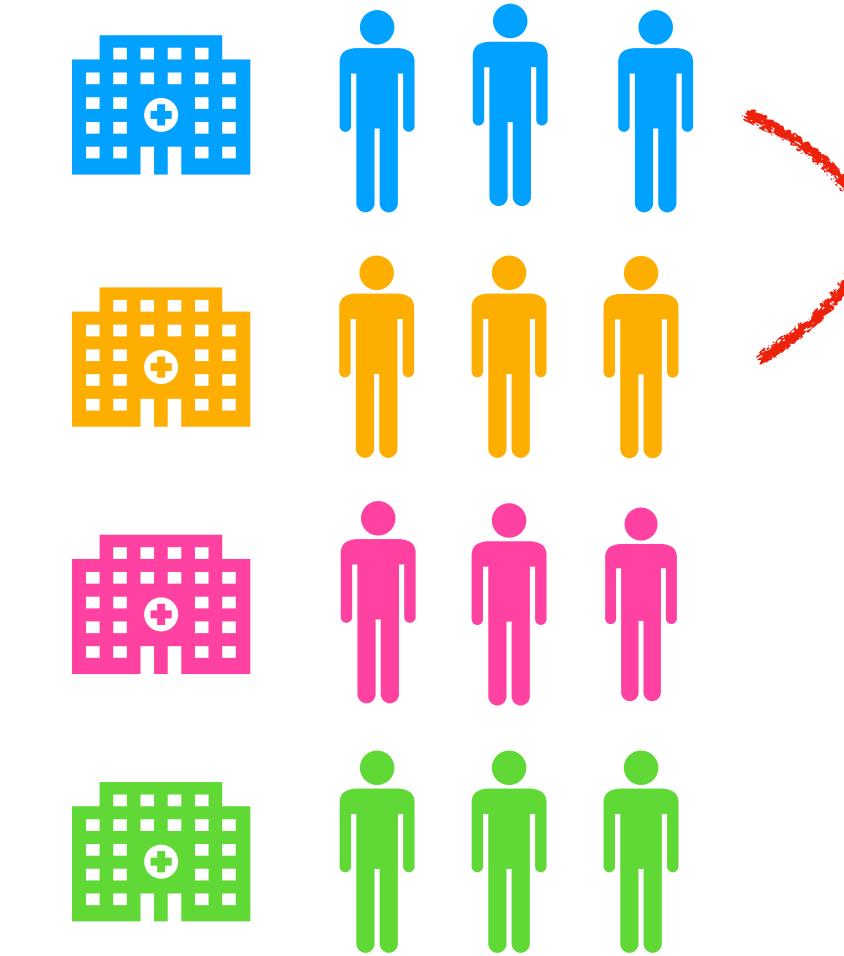
- Effect of site on PET measure
- The difference between the sites are the key interest
- The sites make the population of interest



Fixed or random?

Fixed, if

- Effect of site on PET measure
- The difference between the sites are the key interest
- The sites make the population of interest



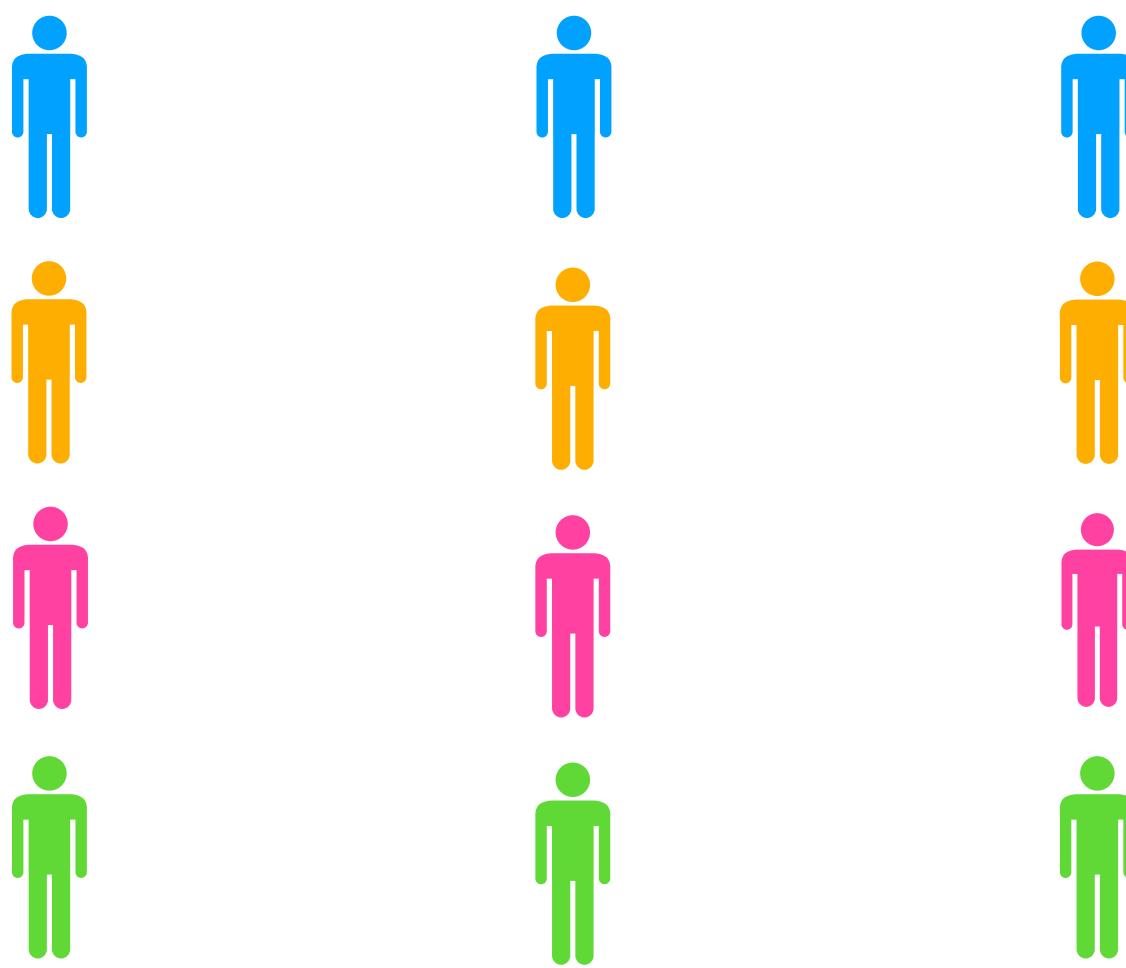
Random, if

- Effect of something else, e.g. age on PET measure
- Random sample of sites (from all possible sites): might affect but the contrasts not interesting
- Site produces random variation in the data: some site may have systematically higher PET measures (random intercept)
- 4+ levels (sites) recommended for random effects

Databases & registers!

Random effects in neuroimaging data

Measure 1 Measure 2 Measure 3

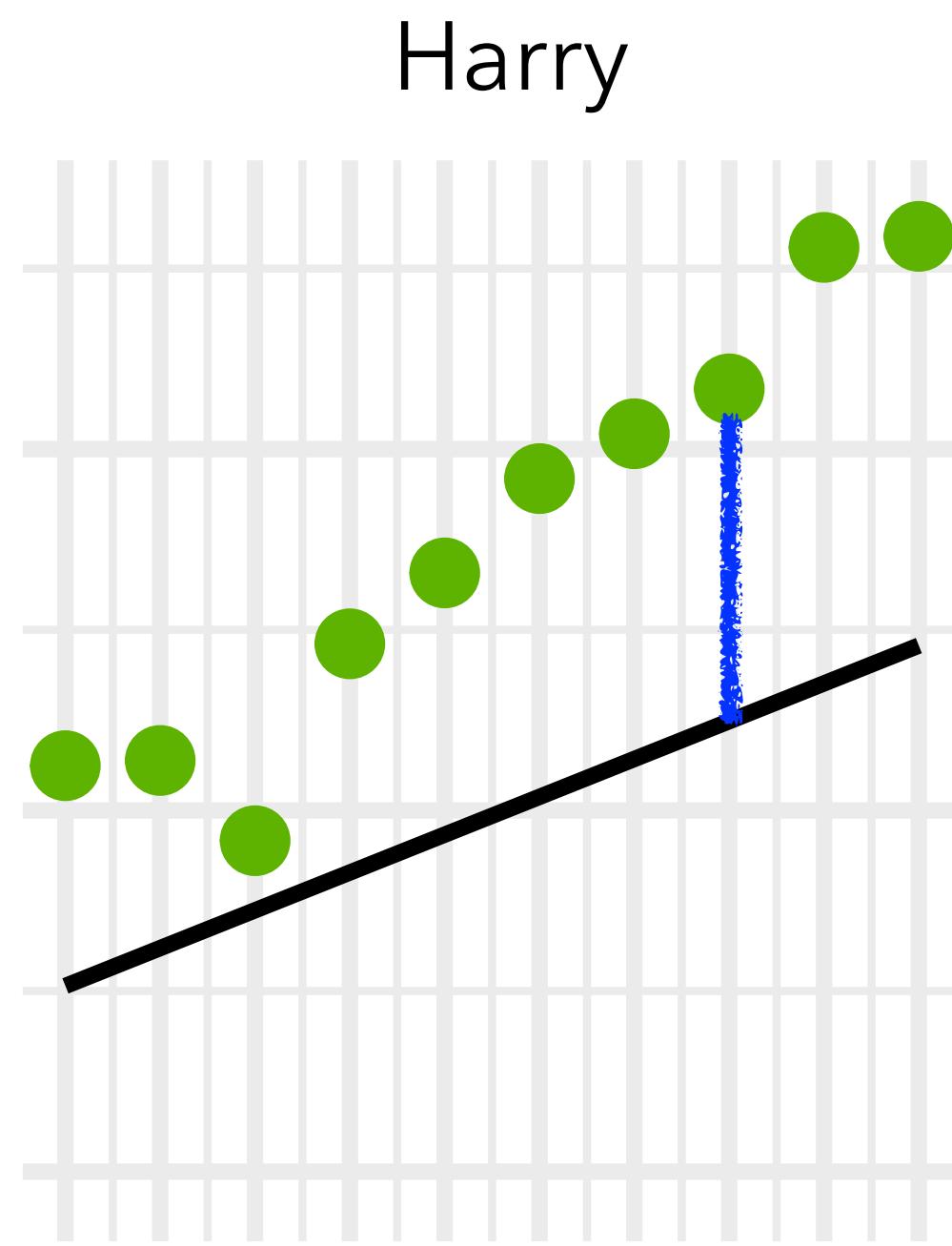


Model diagnostics: Does the model fit the data sufficiently?

'Post-marathon run'

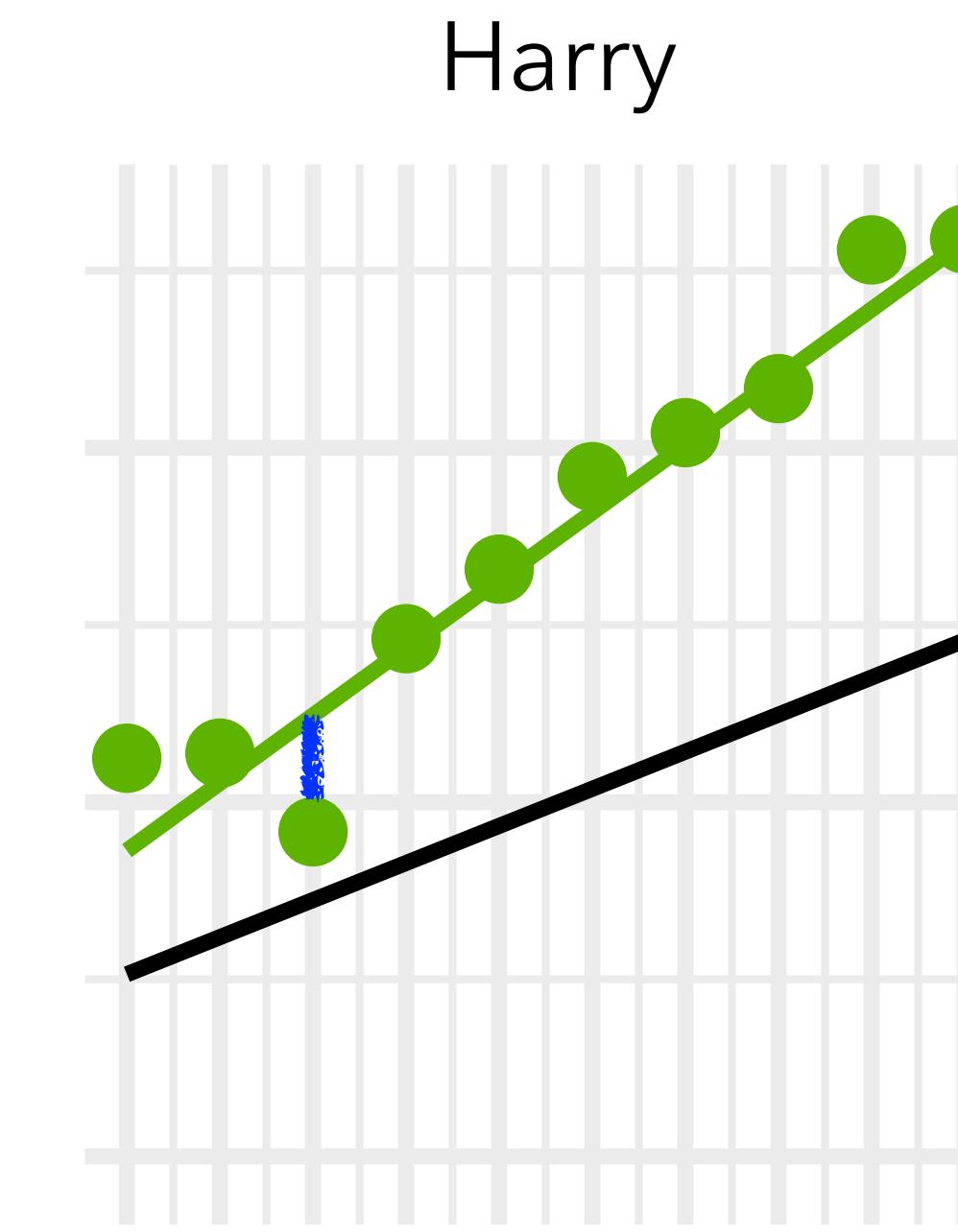
Residual size

Main effect



$\text{Reaction} \sim 1 + \text{Days}$

Main effect + random intercept and slope



$\text{Reaction} \sim 1 + \text{Days} + (1 + \text{Days} \mid \text{Subject})$

Residual variation: Model comparison

summary(fit)

Reaction ~ 1 + Days

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

Reaction ~ 1 + Days + (1 | Subject)

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	1378.2	37.12
	Residual	960.5	30.99

When the value is small, the observations are close to the fit (residuals are small)

Reaction ~ 1 + Days + (0 + Days | Subject)

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	Days	52.71	7.26
Residual		842.03	29.02

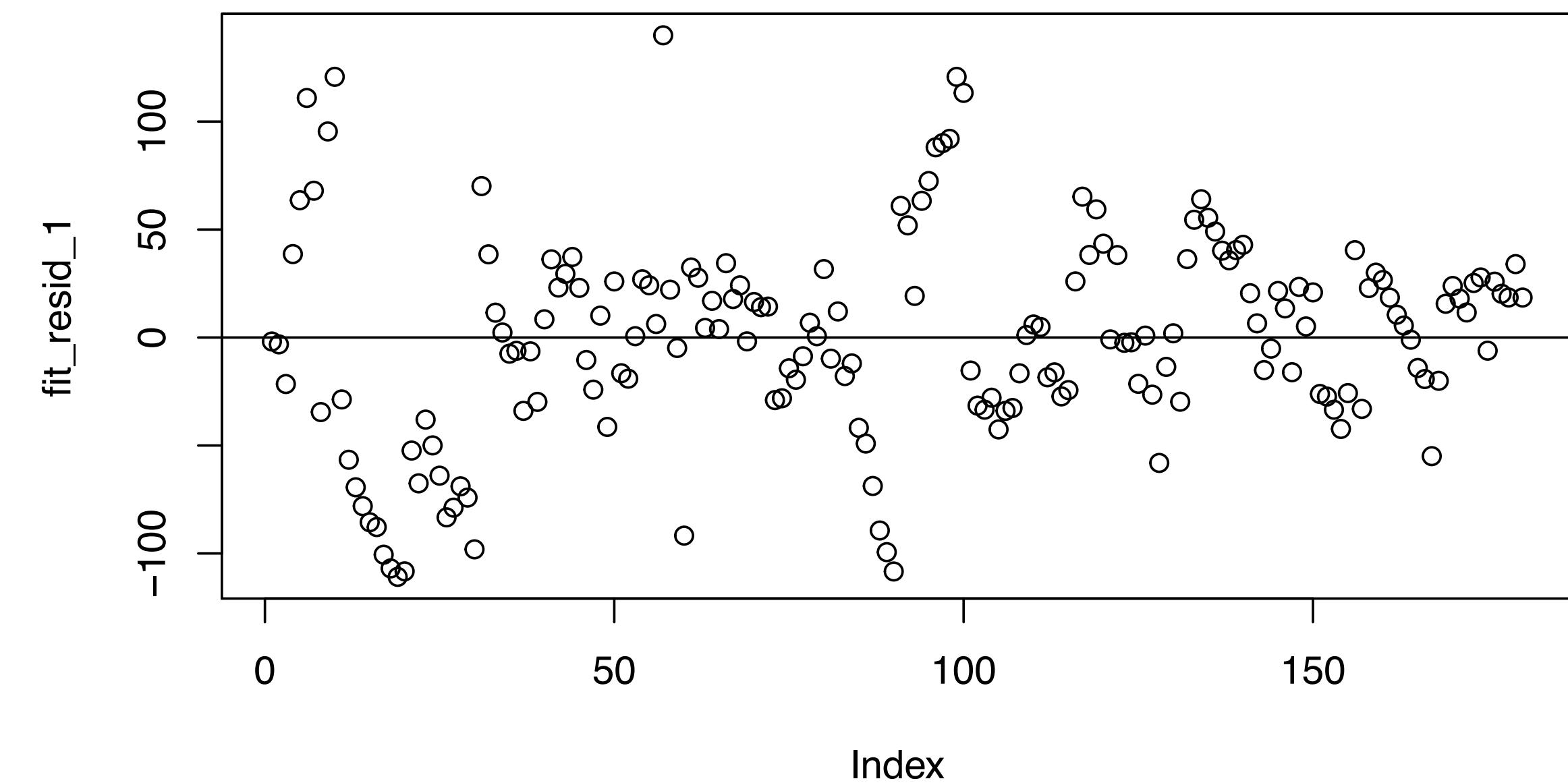
Reaction ~ 1 + Days + (1 + Days | Subject)

Random effects:

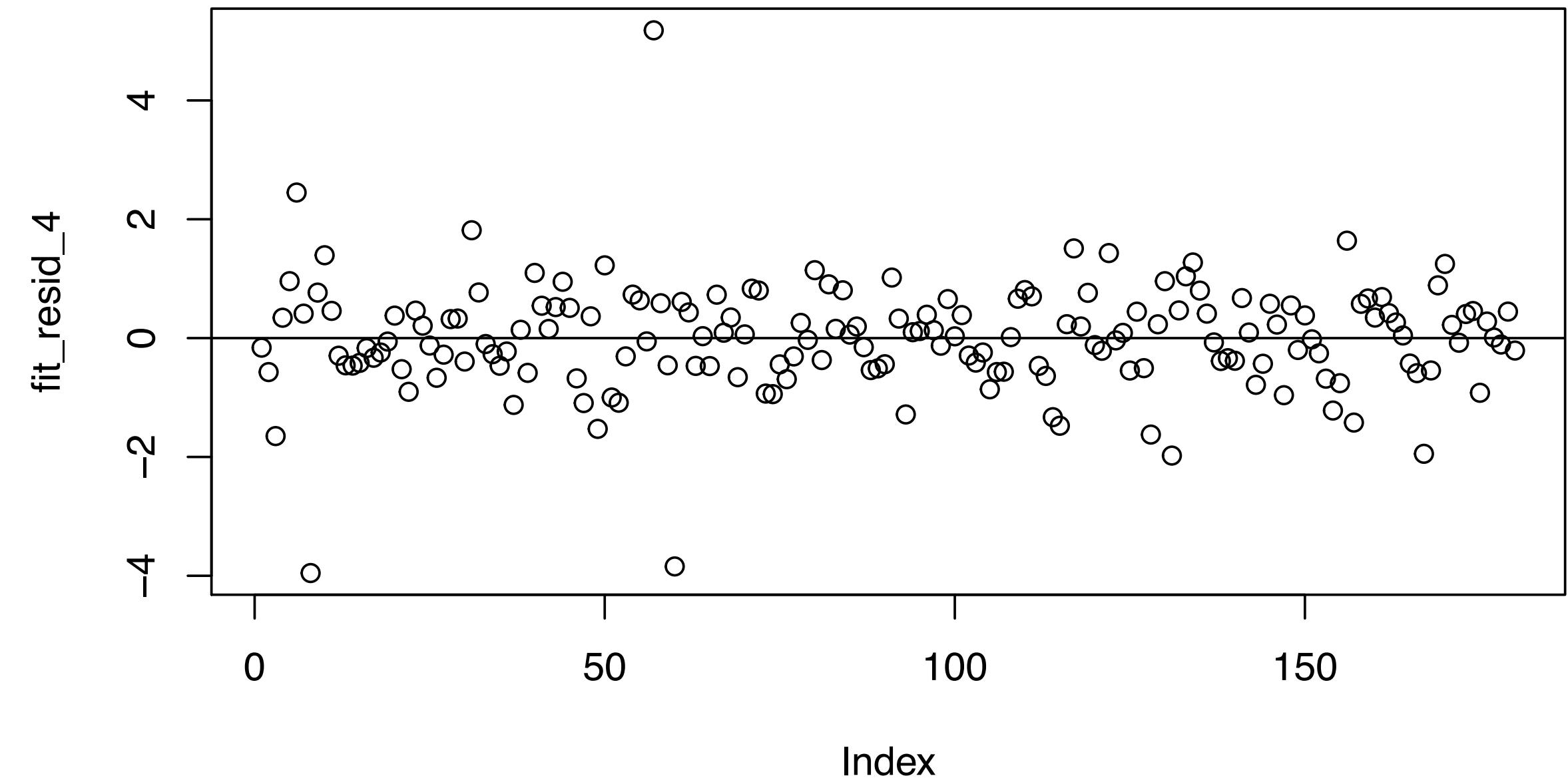
Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	612.10	24.741	
	Days	35.07	5.922	0.07
Residual		654.94	25.592	

Residual distribution

Main effect



Main effect + random intercept and slope



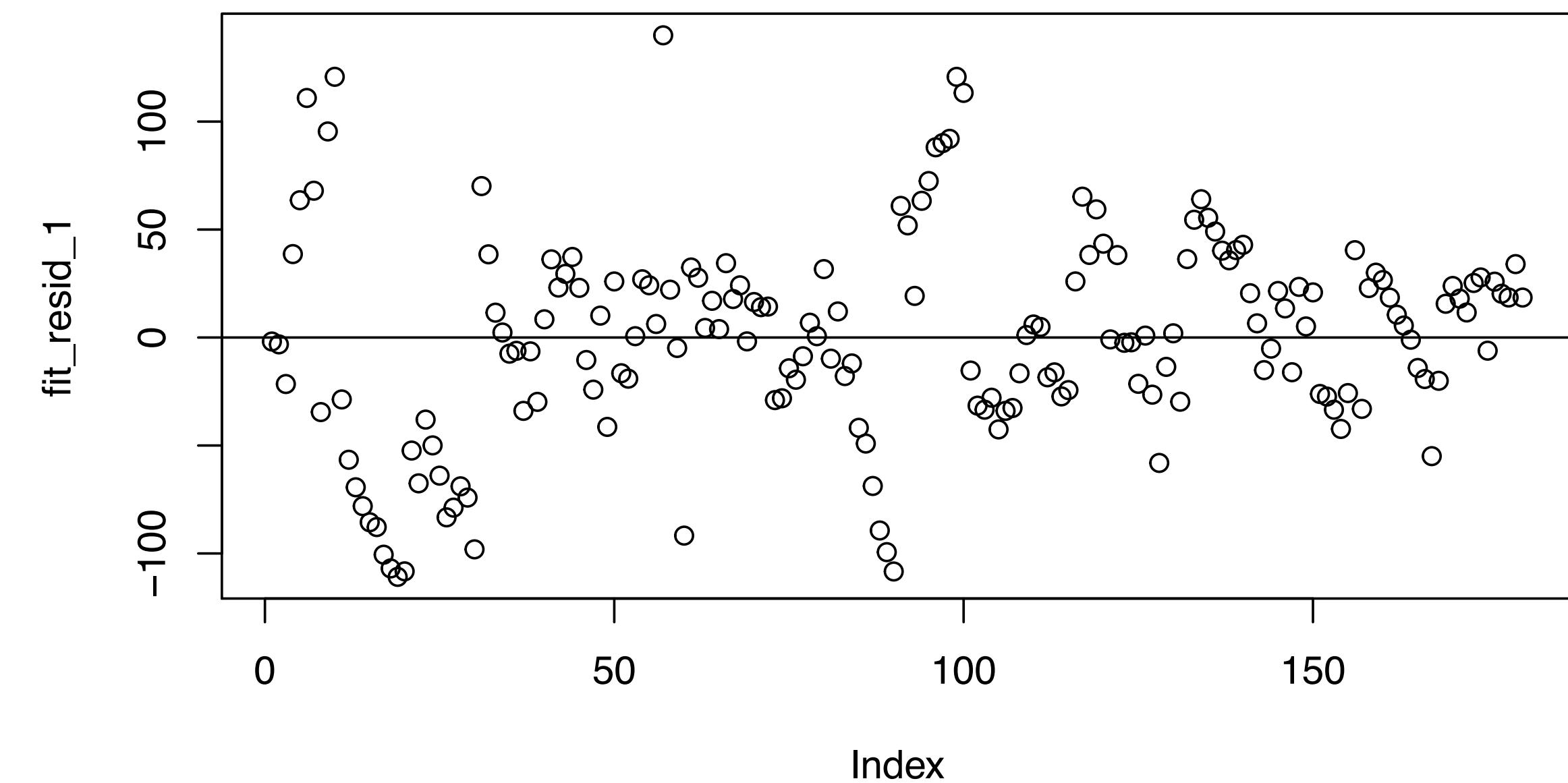
```
resid_1 <- resid(fit_1, scaled = TRUE)  
plot(resid_1)
```

R package 'stats'

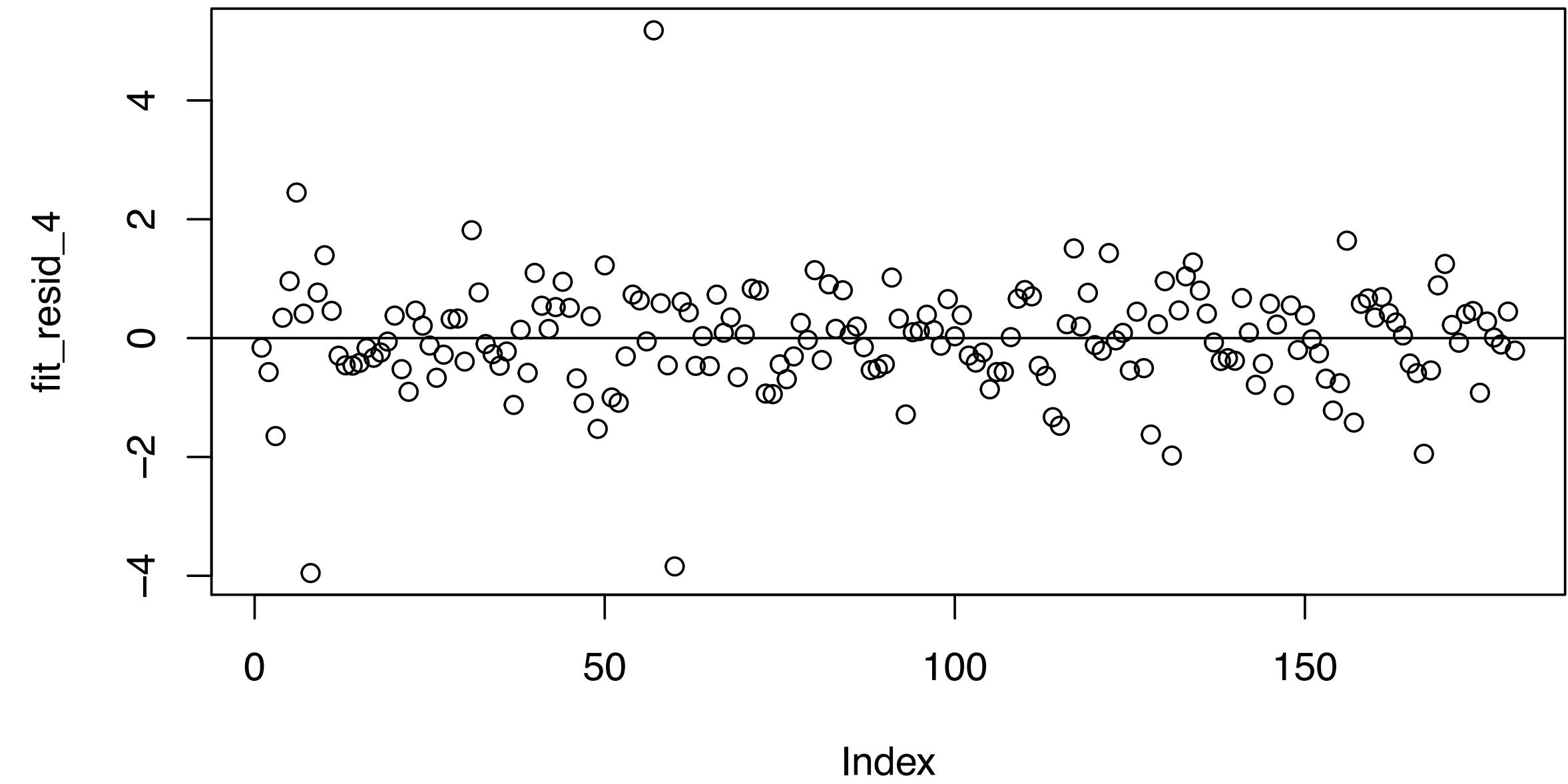
Residual distribution

Randomly (not systematically) distributed residuals support good fit

Main effect



Main effect + random intercept and slope



```
resid_1 <- resid(fit_1, scaled = TRUE)  
plot(resid_1)
```

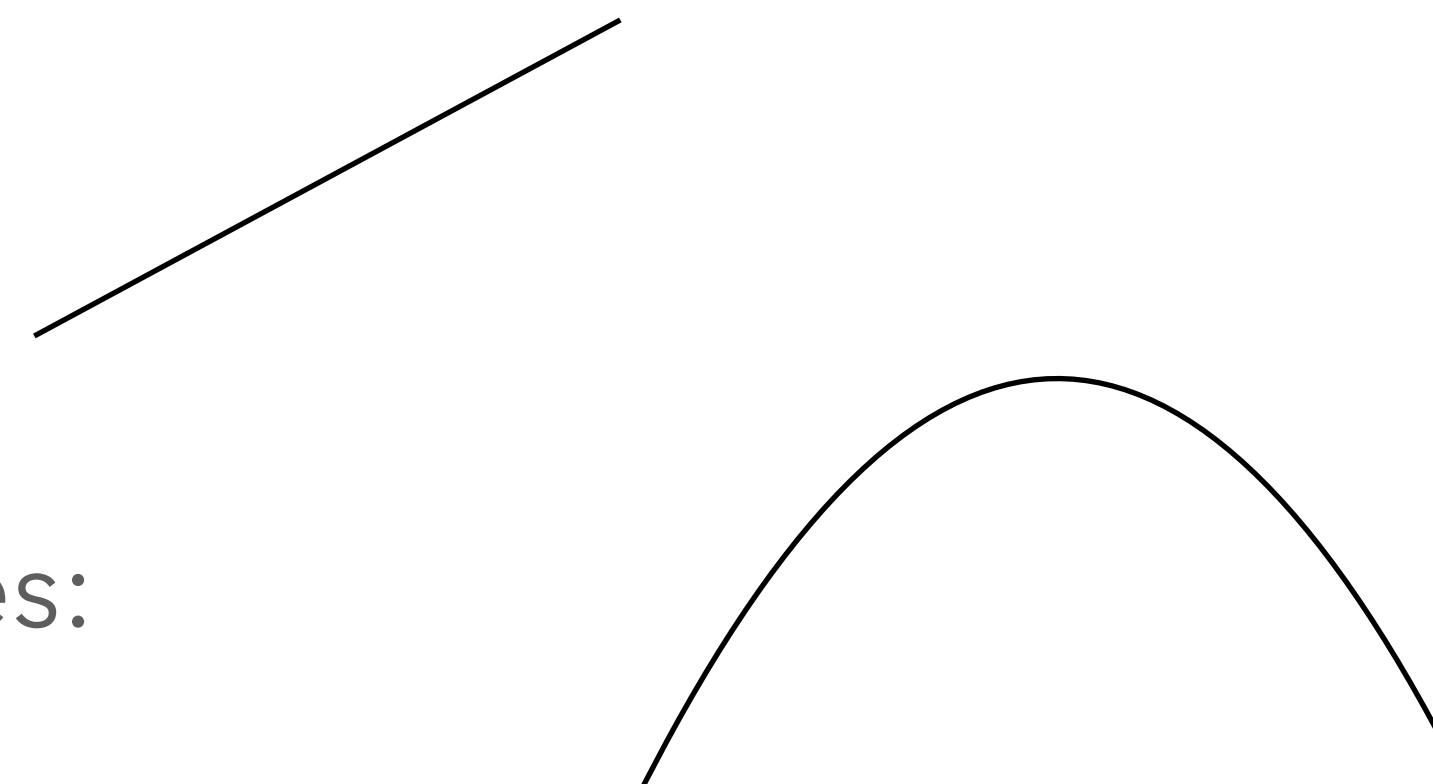
R package 'stats'

Quick word of nonlinearity

- Linear model is linear in terms of parameters (α, β), not variables (Reaction, Days)
- Linear model can be used to model also nonlinear relationship between the variables

Linear relationship between the variables:

$$\text{Reaction} = \alpha + \beta \cdot \text{Days} + \varepsilon$$

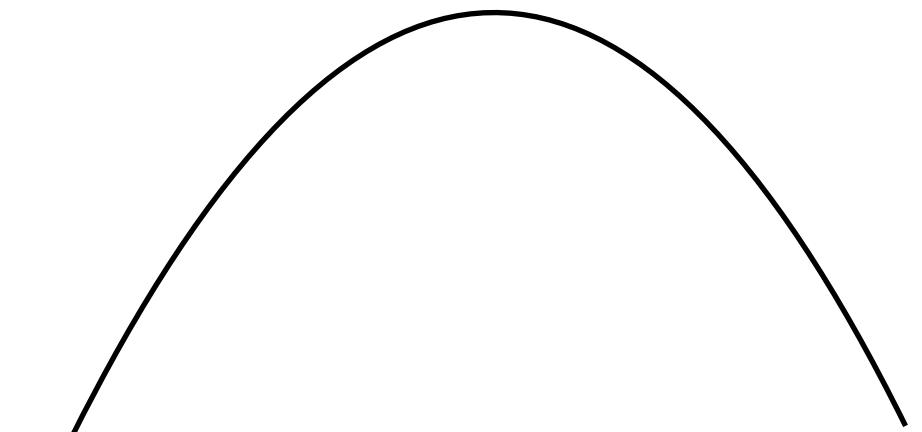


Quadratic relationship between the variables:

$$\text{Reaction} = \alpha + \beta \cdot \text{Days}^2 + \varepsilon$$

↓

Predictor to the 2nd power



Quick word of GLM

- General linear model vs. generalized linear model (GLM used for both)
 - Here we only considered general linear models

To take home

- Linear mixed effects model is one type of linear model
- What makes a linear model mixed?
 - Random effect(s) in addition to the fixed effect(s)
- Mixed effects modeling is useful when there is grouping structure from a variable that is a random sample from the population of interest
 - ‘Uninteresting’ variation in the data
 - We are not interested in comparing the levels (e.g. how Andy is different from Randy)

Thanks!

- **Thank you for your kind attention!**
- **Colleagues from Nummenmaa Lab, particularly Birgitta Paranko and Severi Santavirta**
- Academy Research Fellow / Assistant Professor **Joni Virta**
 - UTU Course: Sekamallit (Mixed effects models)

References

- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

Courses at the University of Turku, mainly:

- **Mixed effects (Sekamallit)** course in the University of Turku, Department of Mathematics and Statistics, by Academy Research Fellow / Assistant Professor **Joni Virta**
- Regression analysis and statistical learning (Regressioanalyysi ja tilastollinen oppiminen), by Professor **Henri Nyberg**

- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Posit team. (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. <http://www.posit.co/>.
- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2021). *_nlme: Linear and Nonlinear Mixed Effects Models_*. R package version 3.1-152, <URL: <https://CRAN.R-project.org/package=nlme>>.
- Figures made with ggplot2: H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- ChatGPT (<https://chatgpt.com>) for searching conceptual information