

Lesson 2: Linear Regression

This Lesson's Goals

Learn about linear regression

Make a figure for data from a linear regression

Do a linear regression in R

Summarise results in an R Markdown document

Math

$$y_i = a + bx_i + e_i$$

y_i = specific y value (dependent variable)

a = intercept

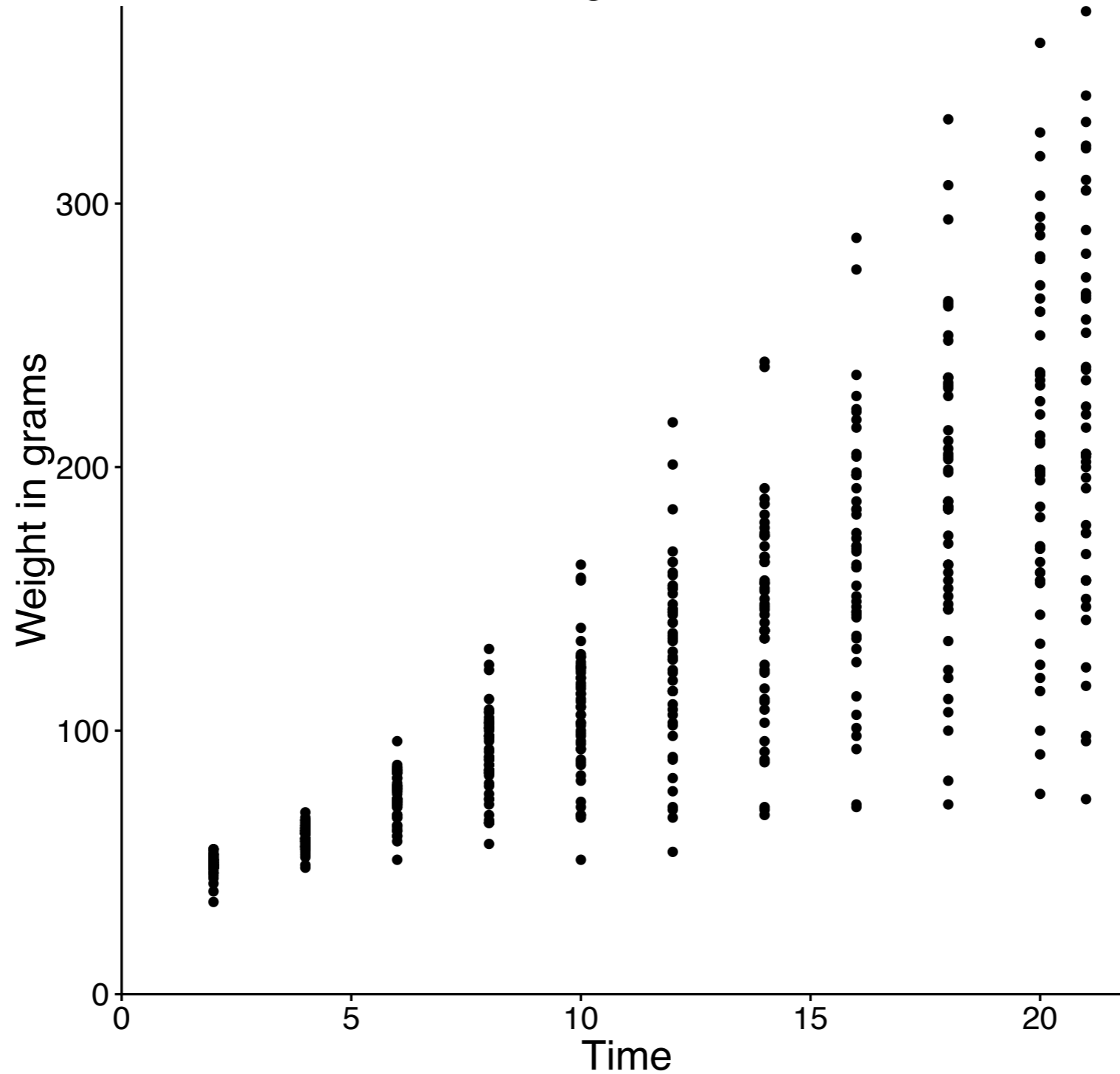
b = slope

x_i = specific x value (independent variable)

e_i = random variance or the residual

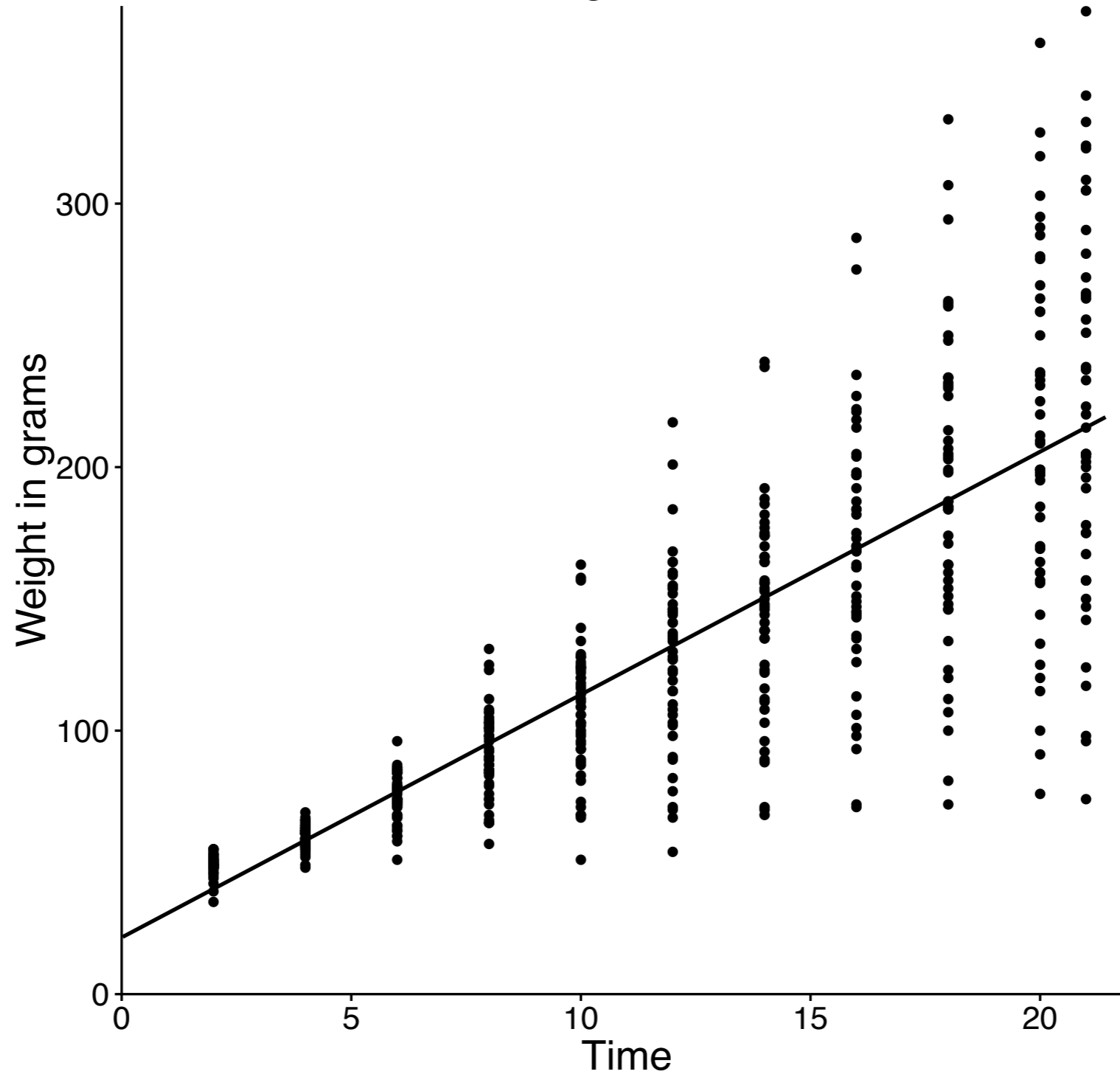
$$y_i = a + bx_i + e_i$$

Chick Weight Over Time



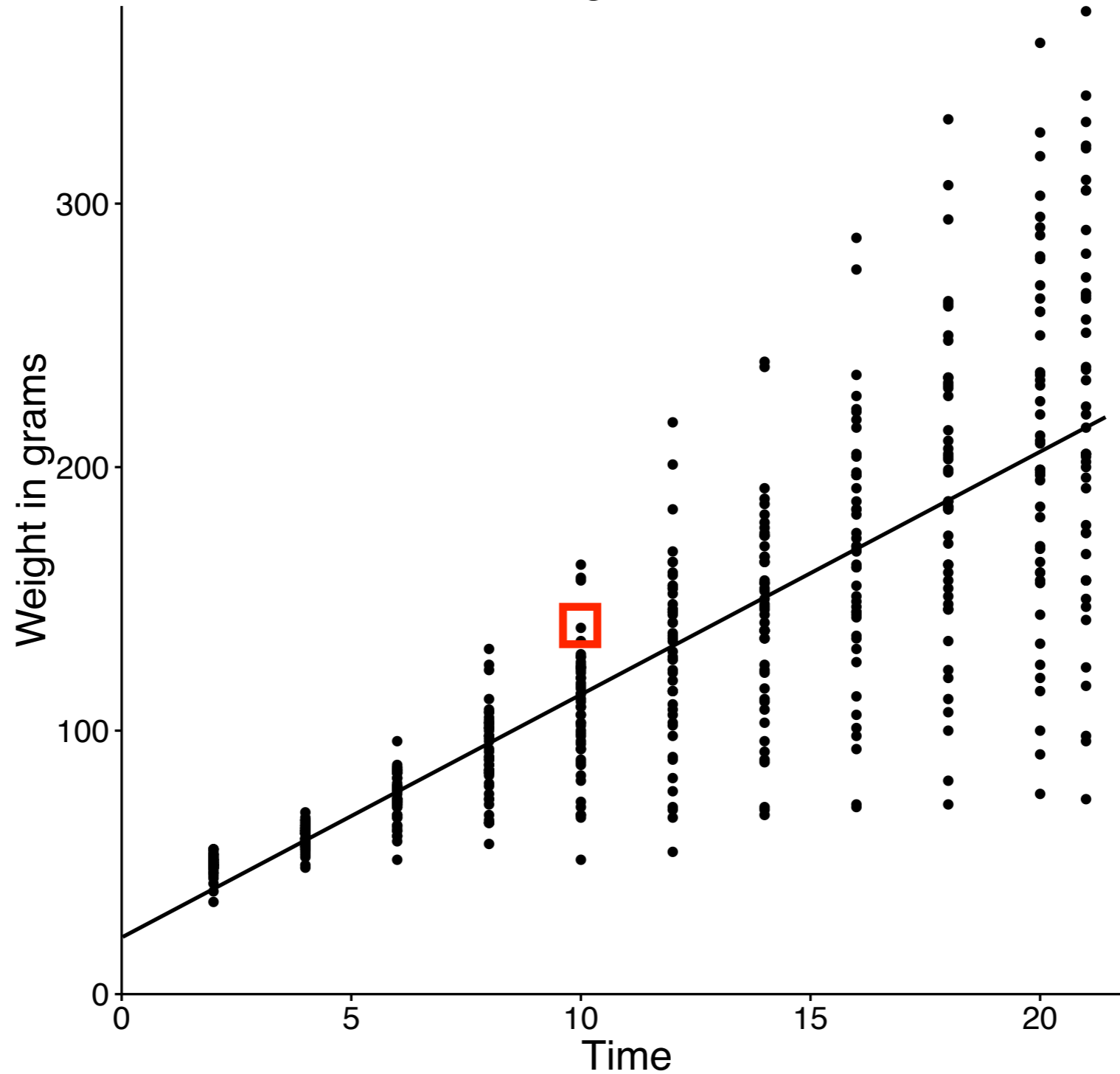
$$y_i = a + bx_i + e_i$$

Chick Weight Over Time



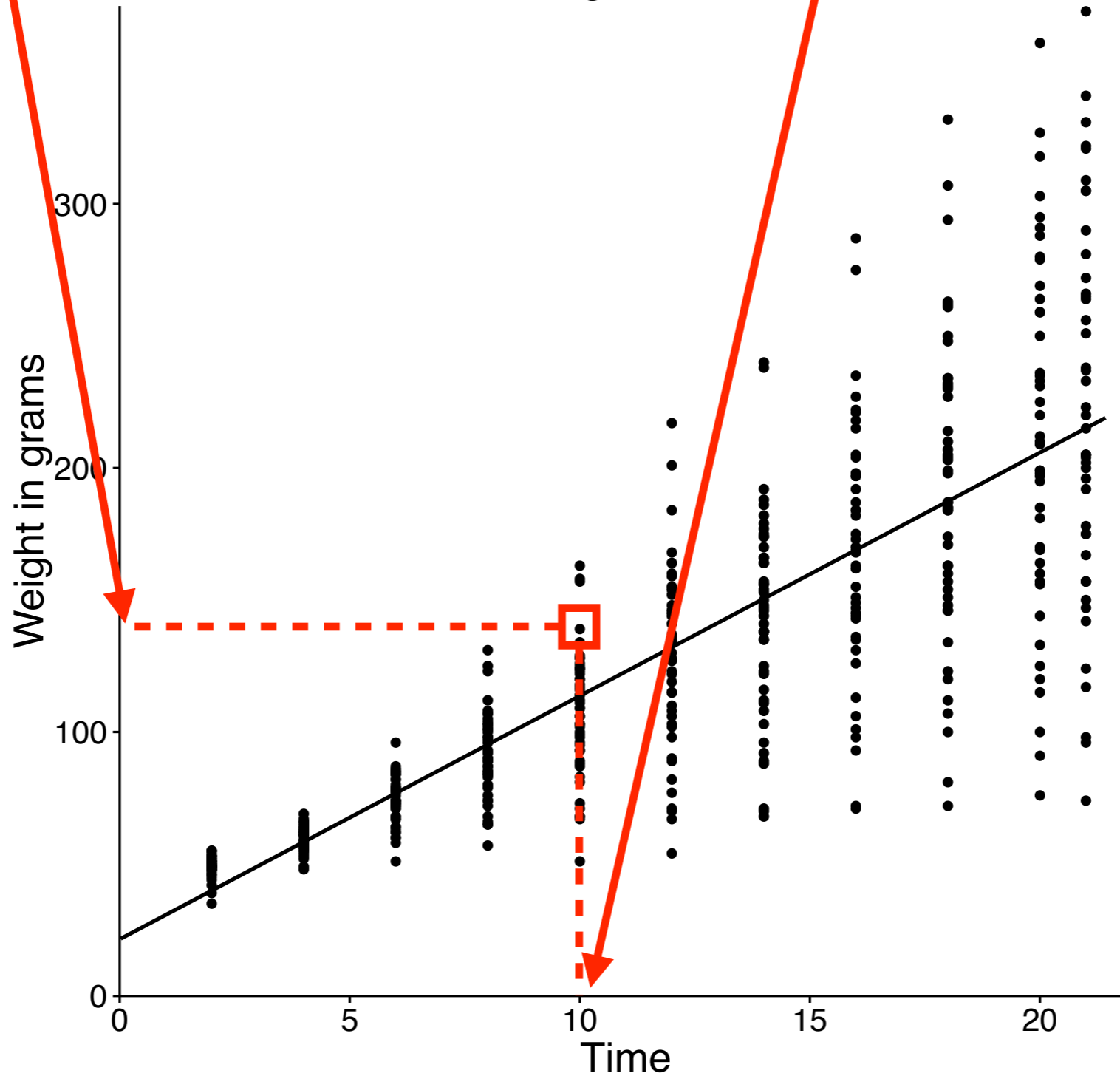
$$y_i = a + bx_i + e_i$$

Chick Weight Over Time



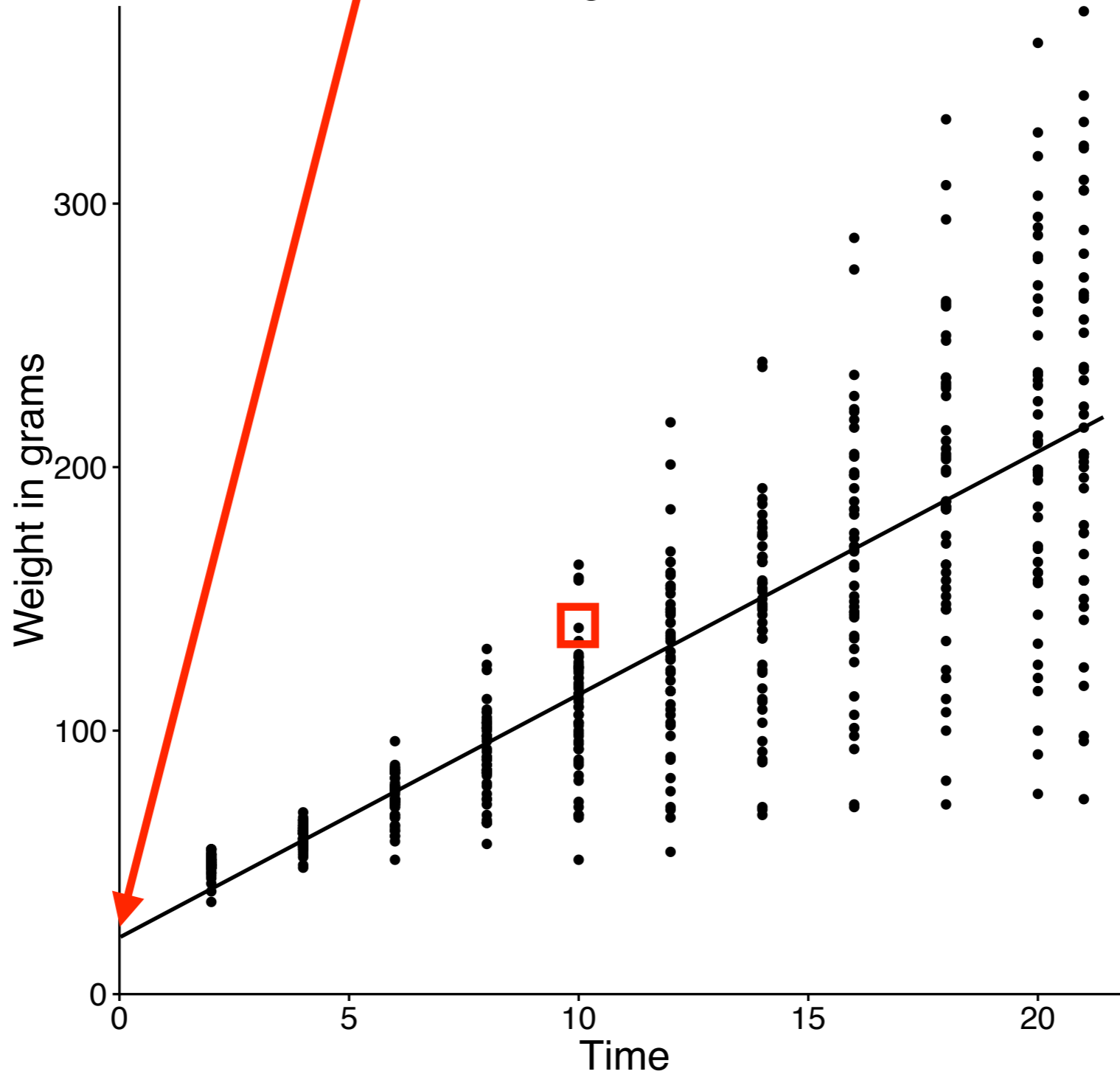
$$y_i = a + bx_i + e_i$$

Chick Weight Over Time



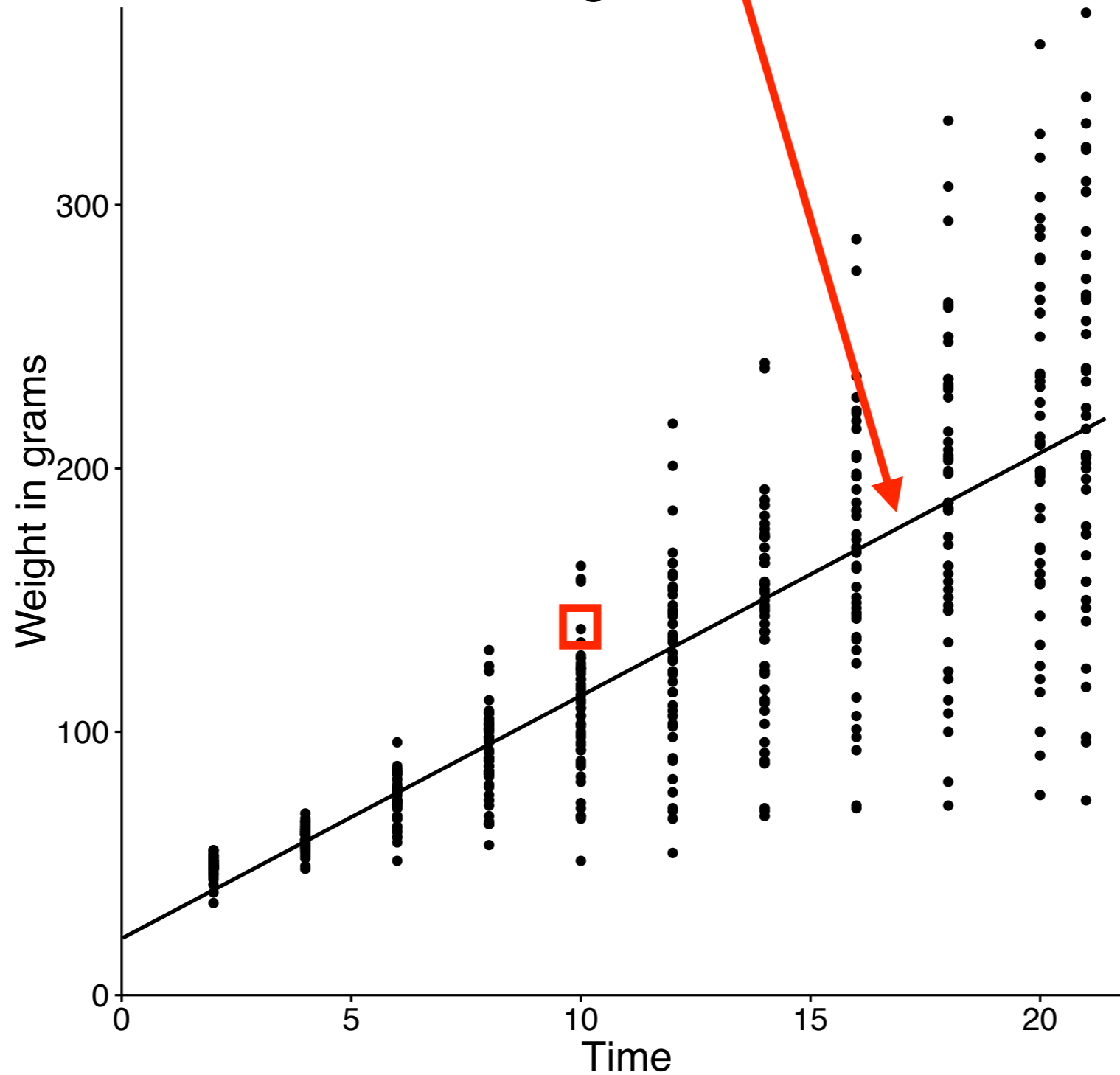
$$y_i = a + bx_i + e_i$$

Chick Weight Over Time



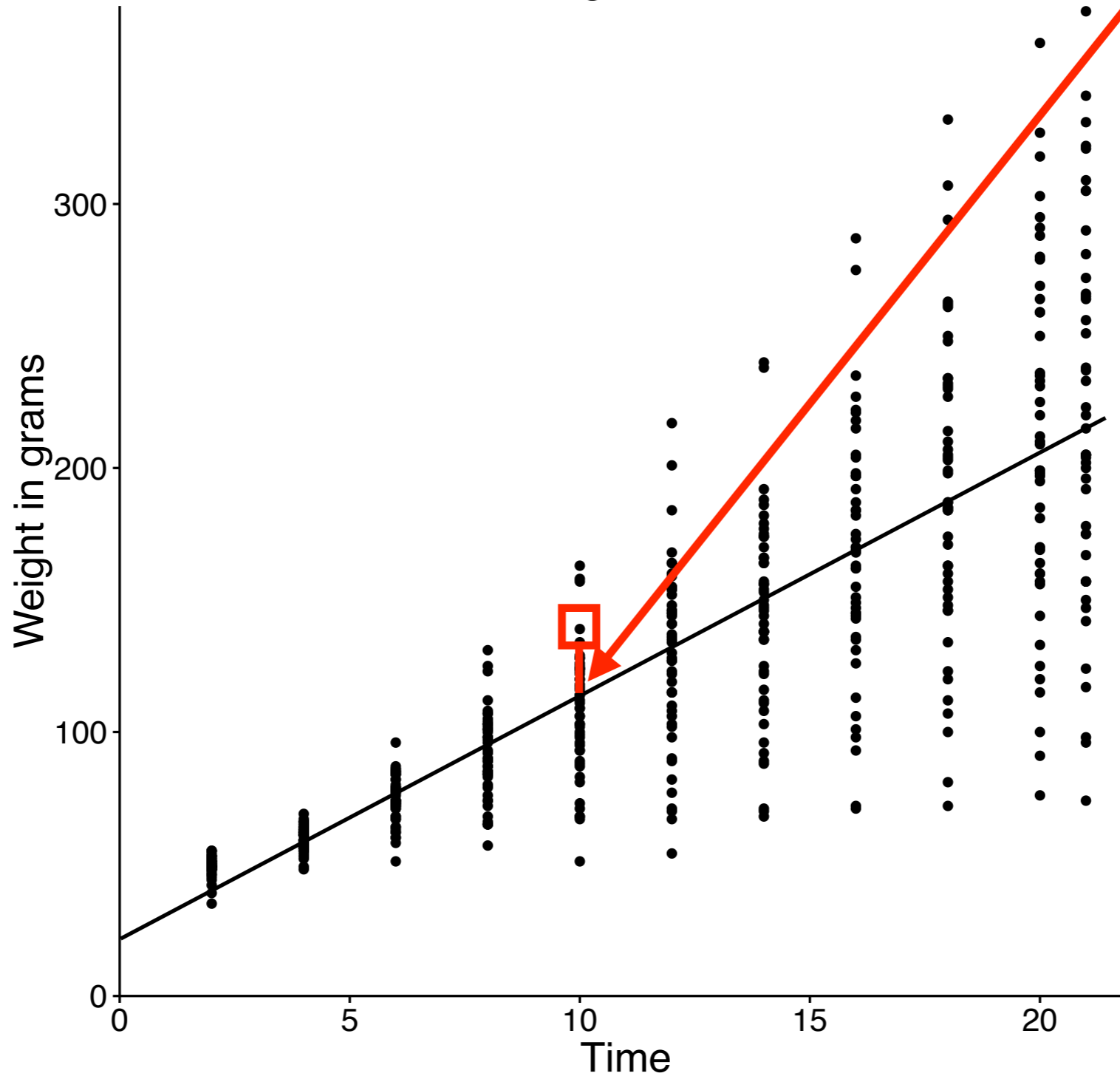
$$y_i = a + bx_i + e_i$$

Chick Weight Over Time



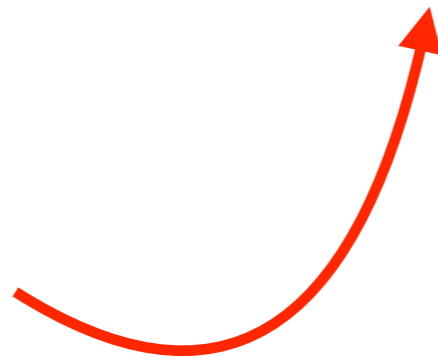
$$y_i = a + bx_i + e_i$$

Chick Weight Over Time



$$y_i = a + b x_i + e_i$$

categorical?



$$a + bx$$

continuous predictors

x = a set of continuous
data points

a = the value of y
when x is 0

b = the change in y for
one change in x

categorical predictors

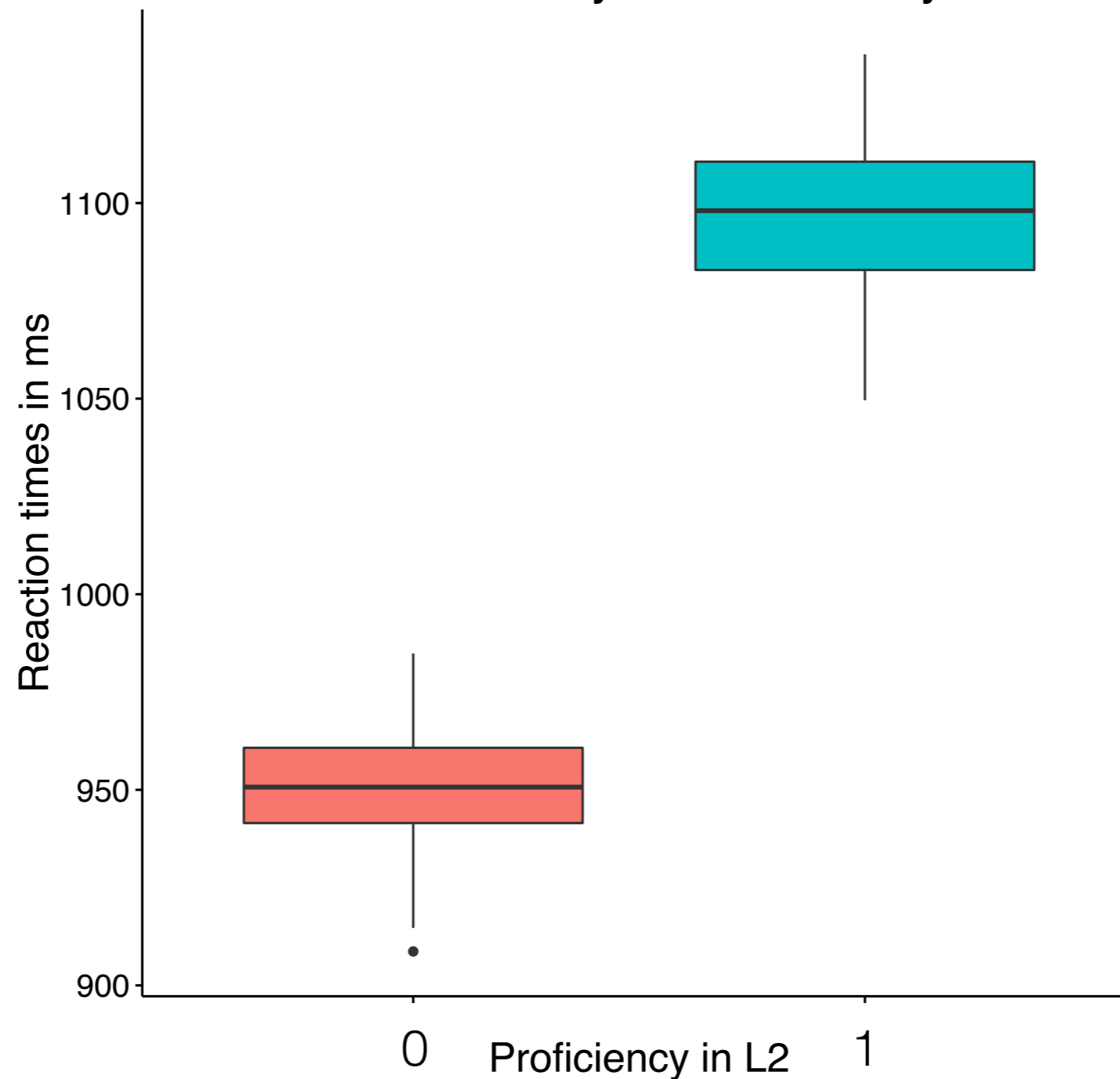
x = a set of
binary/categorical data points

a = the value of y when the x is
the **default level**

b = the change in y when x is the
non-default level

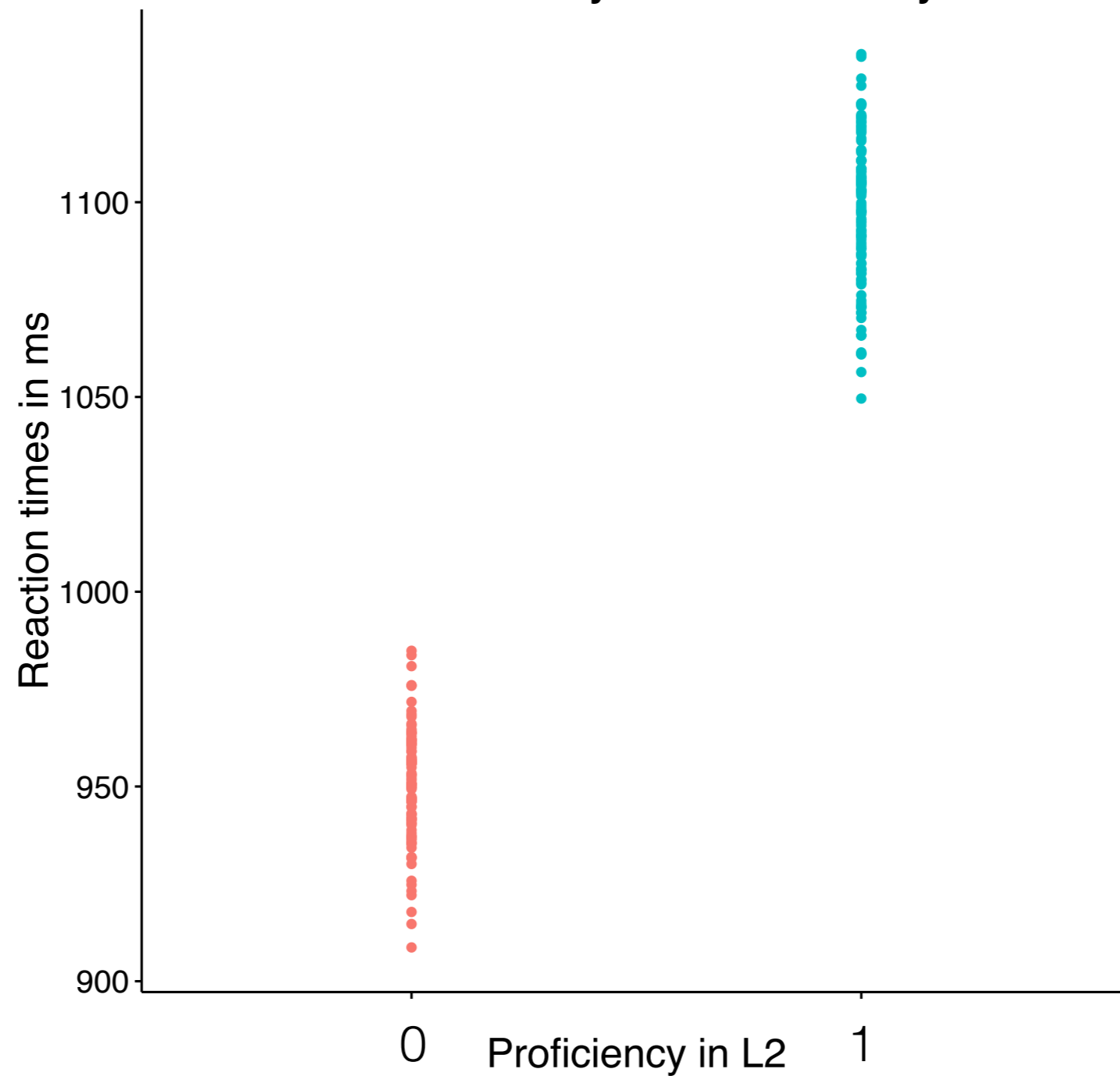
$$y_i = a + bx_i + e_i$$

Reaction Times by L2 Proficiency Level



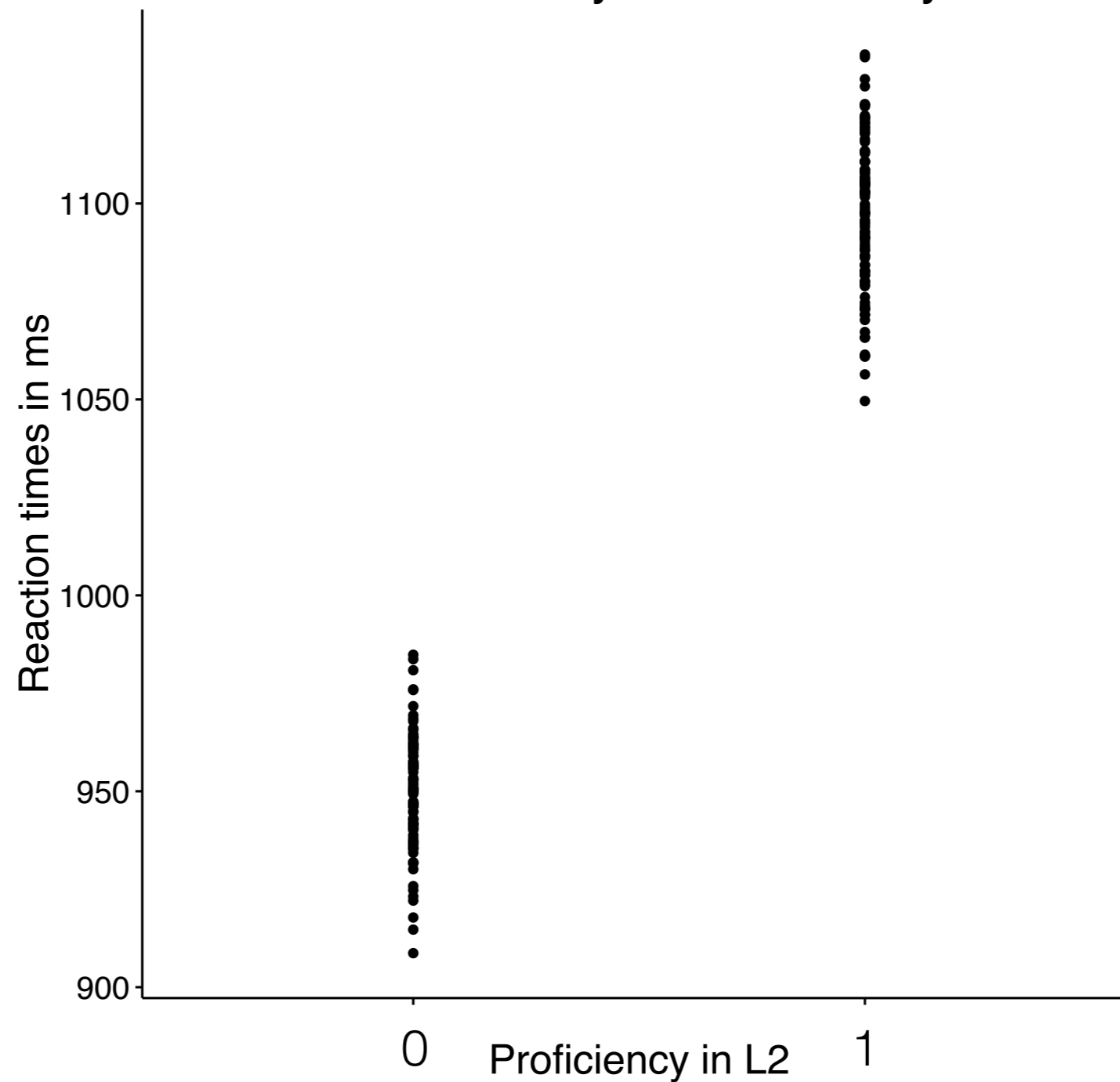
$$y_i = a + bx_i + e_i$$

Reaction Times by L2 Proficiency Level



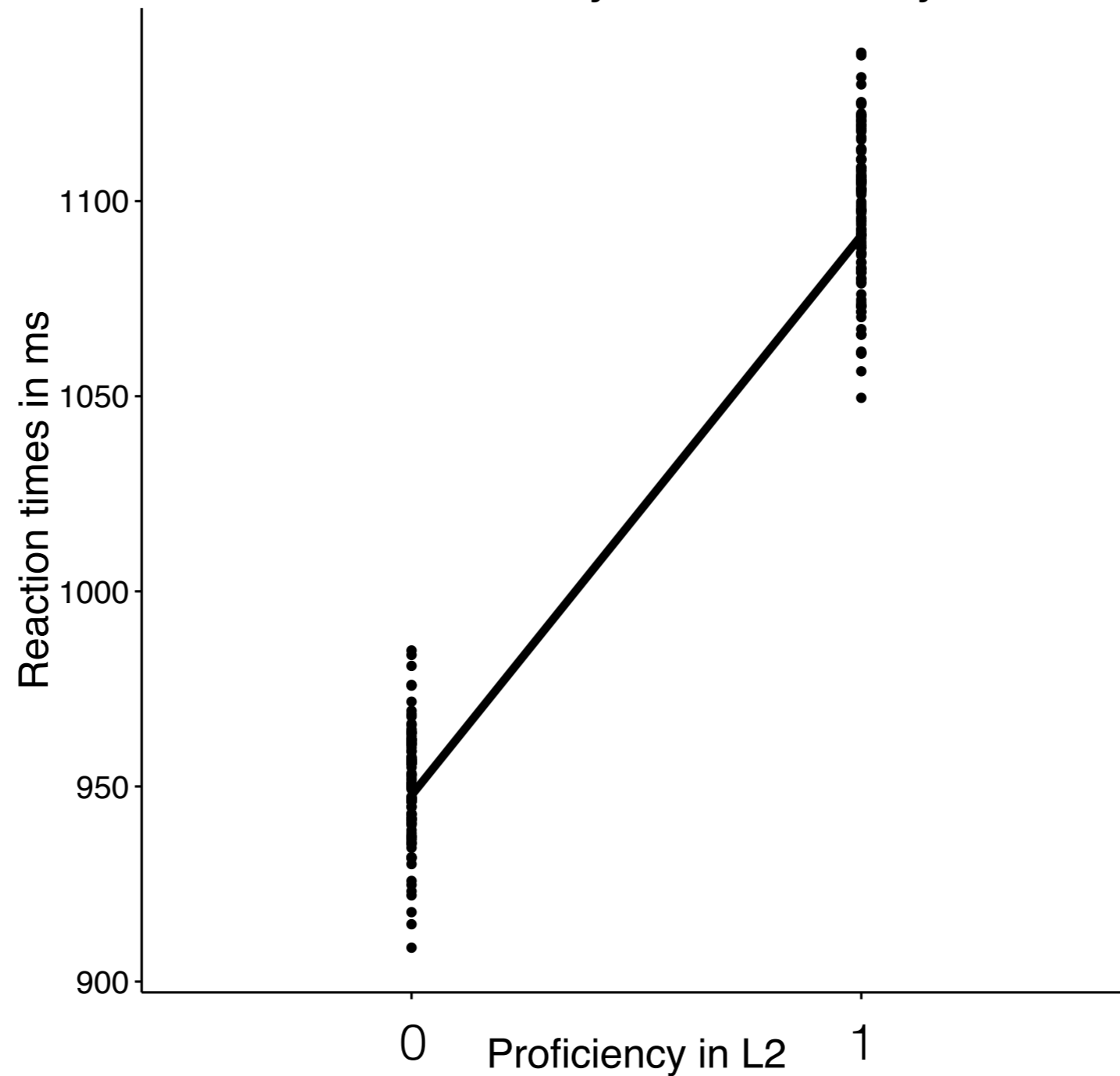
$$y_i = a + bx_i + e_i$$

Reaction Times by L2 Proficiency Level



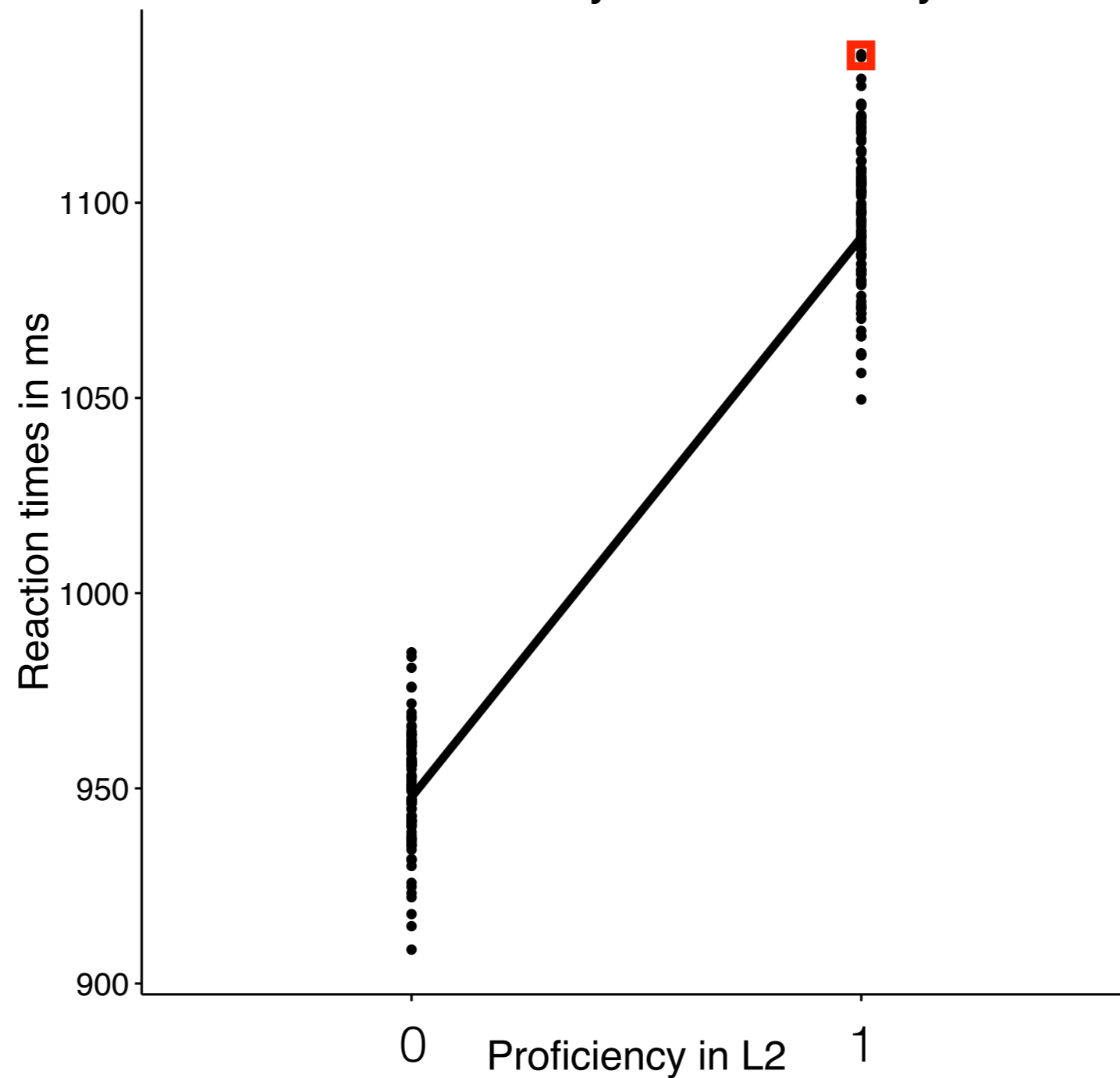
$$y_i = a + bx_i + e_i$$

Reaction Times by L2 Proficiency Level

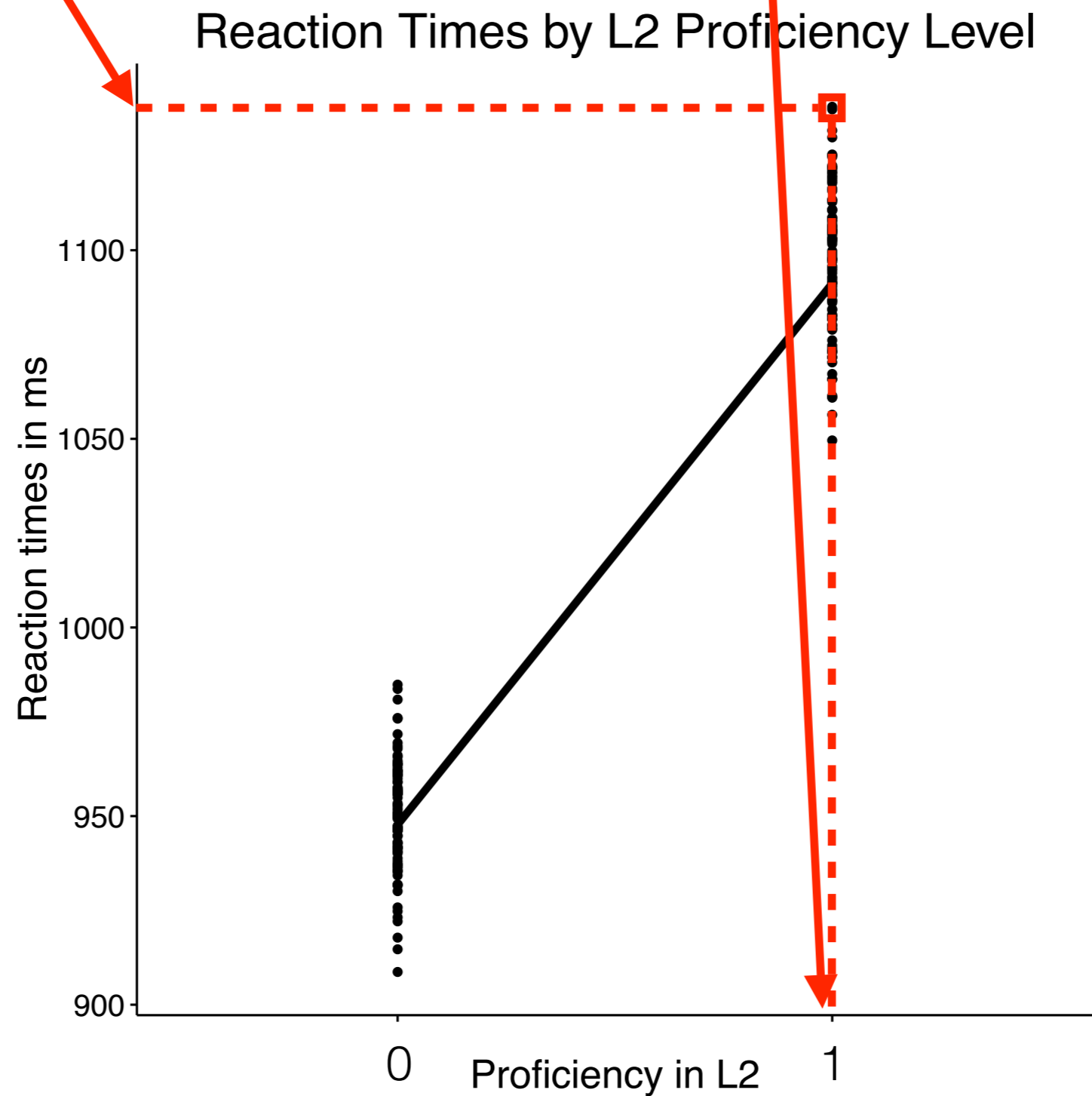


$$y_i = a + bx_i + e_i$$

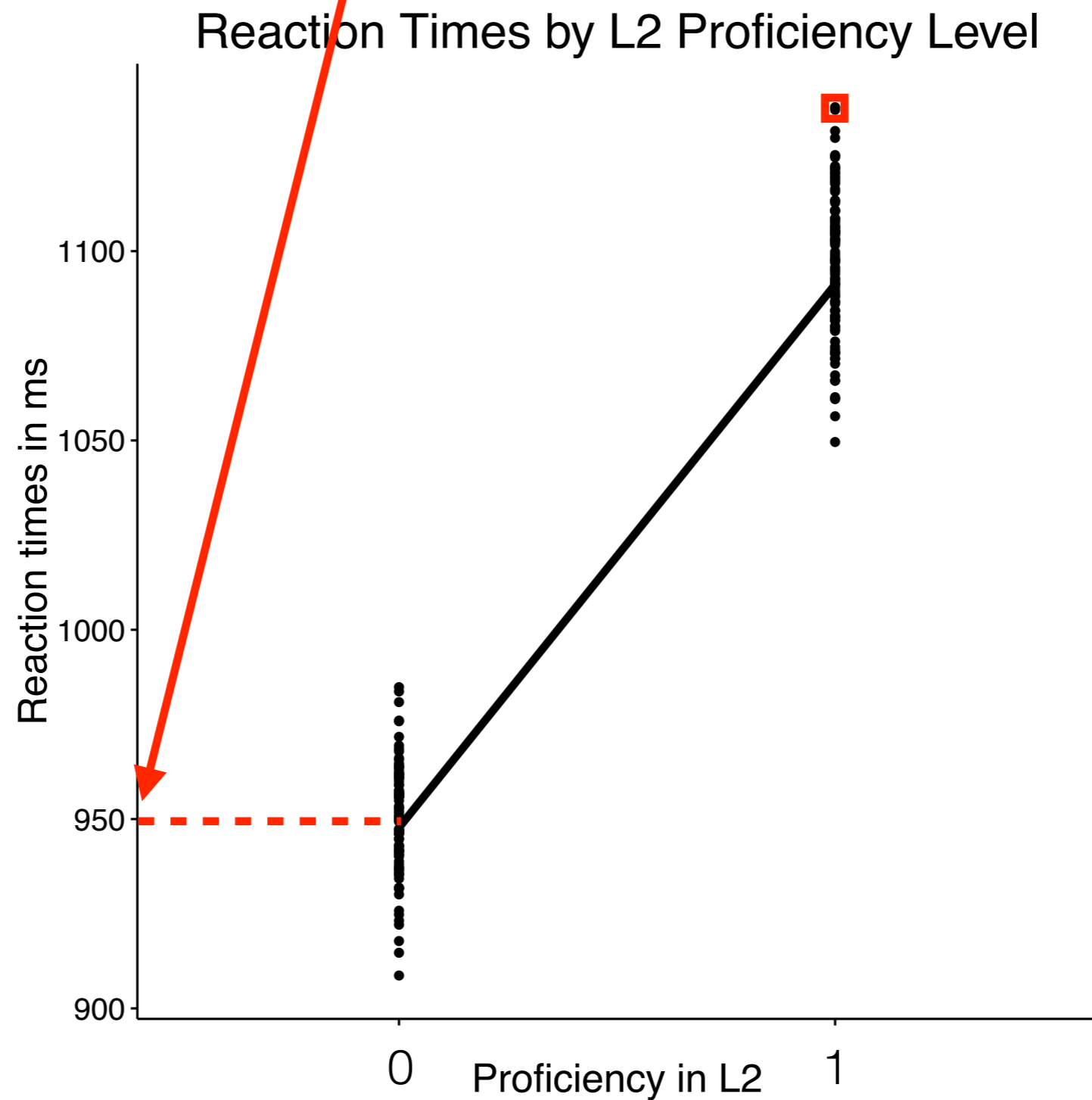
Reaction Times by L2 Proficiency Level



$$y_i = a + bx_i + e_i$$

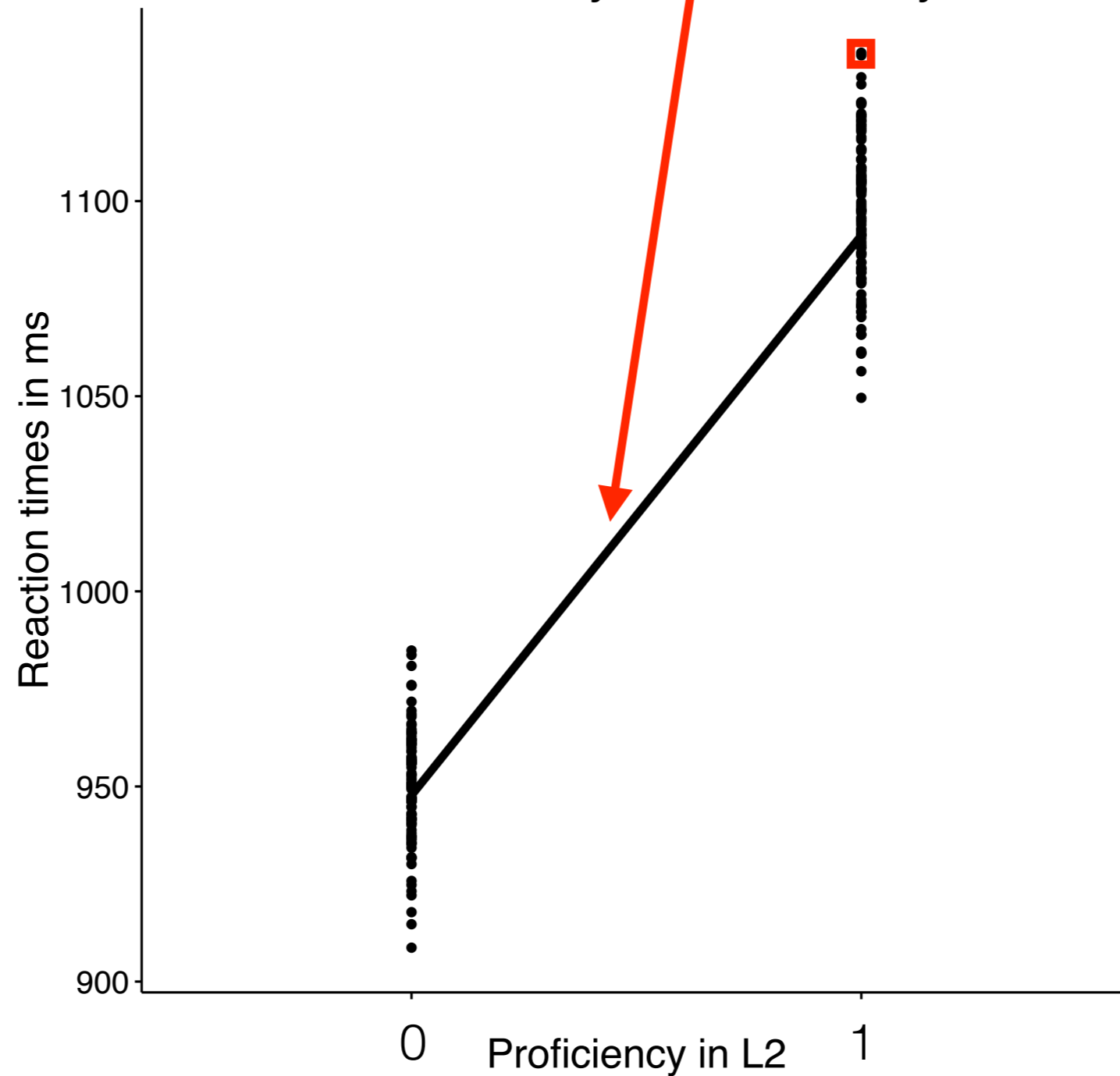


$$y_i = a + bx_i + e_i$$



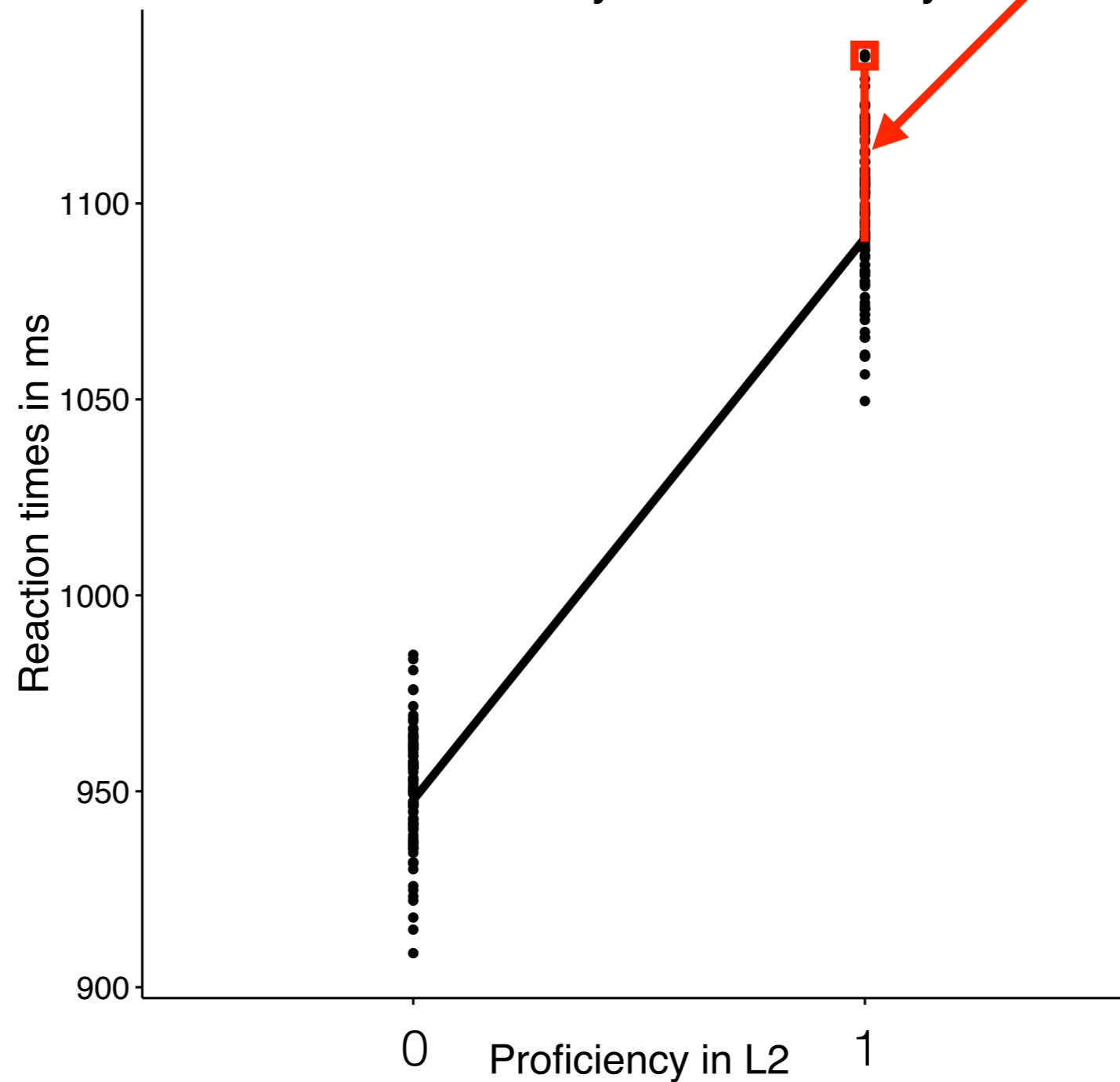
$$y_i = a + bx_i + e_i$$

Reaction Times by L2 Proficiency Level



$$y_i = a + bx_i + e_i$$

Reaction Times by L2 Proficiency Level



R Code

$$y_i = a + bx_i + e_i$$

`lm(weight ~ Time)`

Call:
`lm(formula = weight ~ Time, data = ChickWeight)`

Residuals:

Min	1Q	Median	3Q	Max
-138.331	-14.536	0.926	13.533	160.669

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	27.4674	3.0365	9.046	<2e-16 ***
Time	8.8030	0.2397	36.725	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38.91 on 576 degrees of freedom

Multiple R-squared: 0.7007, Adjusted R-squared: 0.7002

F-statistic: 1349 on 1 and 576 DF, p-value: < 2.2e-16

`> head(resid(continuous.lm))`

1	2	3
14.532575	5.926496	-3.679582
4	5	6
-16.285661	-21.891739	-22.497818

$$y_i = a + bx_i + e_i$$

`lm(rt ~ type)`

Call:
`lm(formula = rt ~ type, data = data_lesson)`

Residuals:

Min	1Q	Median	3Q	Max
-47.558	-10.712	0.535	11.073	40.874

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	950.31	1.704	557.57	<2e-16 ***
typelow	146.837	2.410	60.92	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.04 on 198 degrees of freedom
Multiple R-squared: 0.9493, Adjusted R-squared: 0.9491
F-statistic: 3711 on 1 and 198 DF, p-value: < 2.2e-16

`> head(resid(categorical.lm))`

1	2	3
-18.4180373	3.0870301	-16.0289276
4	5	6
11.5941267	10.6097411	-0.3063961

Lab

Dataset: Baby names per year
from USA Social Security Administration

Continuous Predictor: Does your name get more or less popular between the years of 1901 and 2000?

Categorical Predictor: Is your name more or less popular with females or males?

Continuous Predictor

y_i = frequency of name
 a = ? - will get from model
 b = ? - will get from model
 X_i = year
 e_i = ? - will get from model

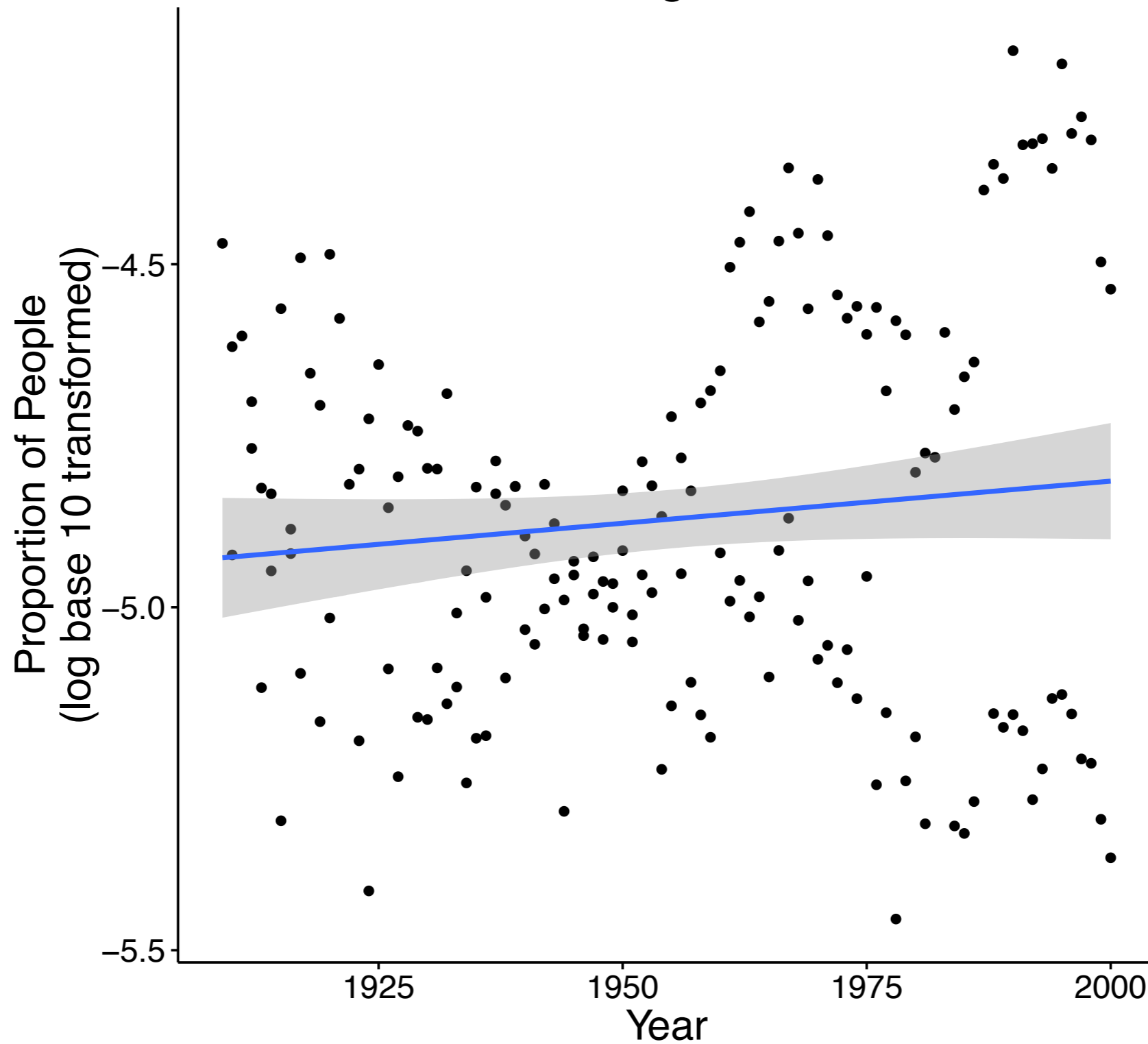
Categorical Predictor

y_i = frequency of name
 a = ? - will get from model
 b = ? - will get from model
 X_i = sex
 e_i = ? - will get from model

source: R package "babynames"

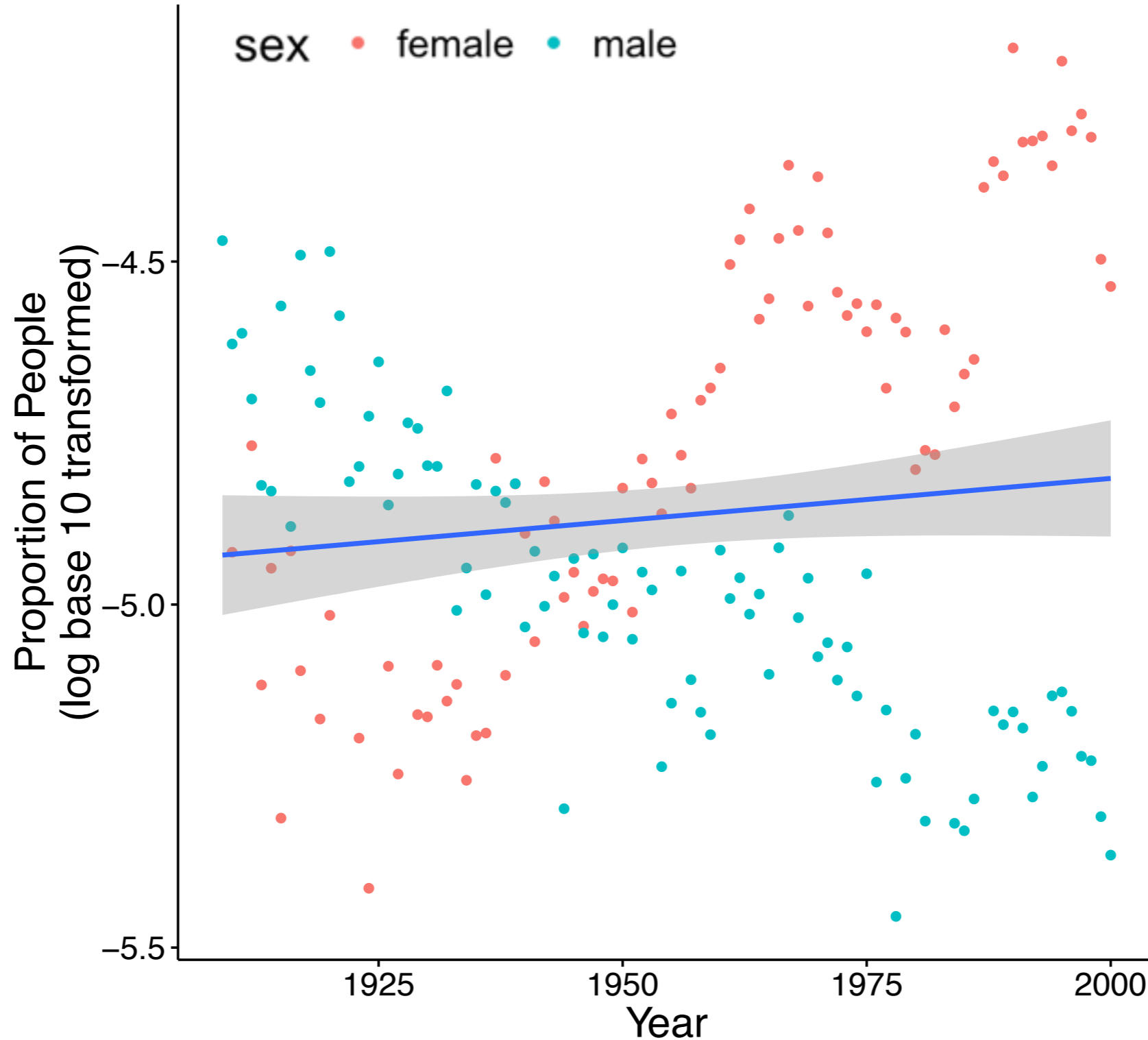
End of Lesson Food for Thought

Proportion of People with
the Name 'Page' Over Time



End of Lesson Food for Thought

Proportion of People with
the Name 'Page' Over Time



End of Lesson Food for Thought

Proportion of People with the Name 'Page' Over Time

