# Week 2 Lecture 1:
Linear regression

EDS 222: *Statistics for Environmental Data Science*

Predicting flooding after storms
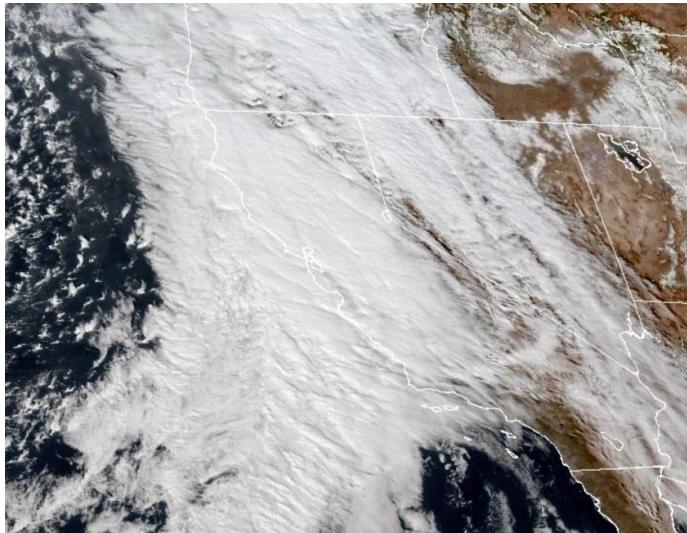

January 31, 2024
Image: GOES-West GeoColor via NOAA

February 4, 2024
Photo: AP Photo/Damian Dovarganes

Predicting flooding after storms
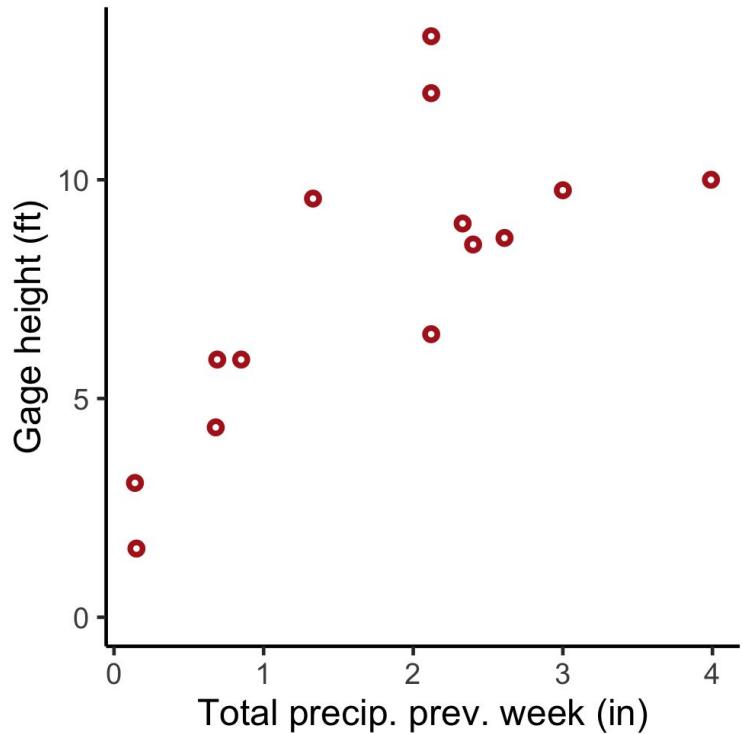
A storm hits LA, dropping 2" of rain in a week. How high would the LA river get?

What about a storm that drops 4"?

Today's agenda

➔ Fitting lines to points

➔ Quality of the fit
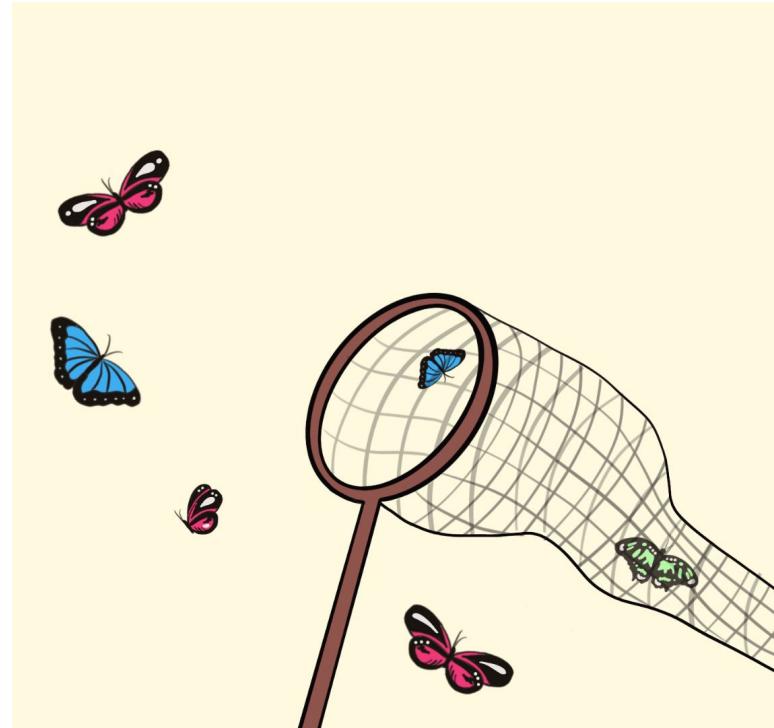
➔ Regression and RVs

Today's agenda

➜ **Fitting lines to points**

➜ Quality of the fit

➜ Regression and RVs
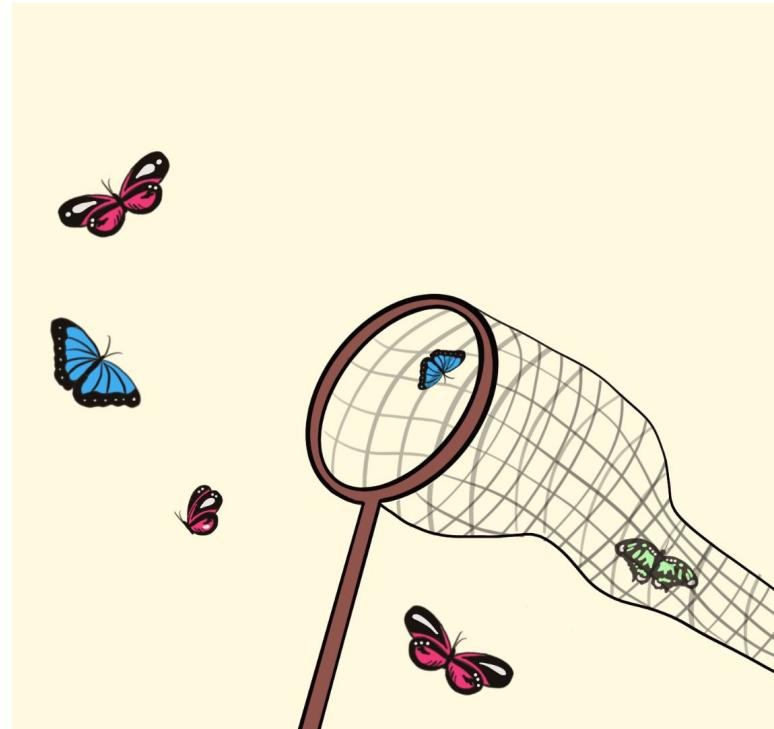
y=mx+b
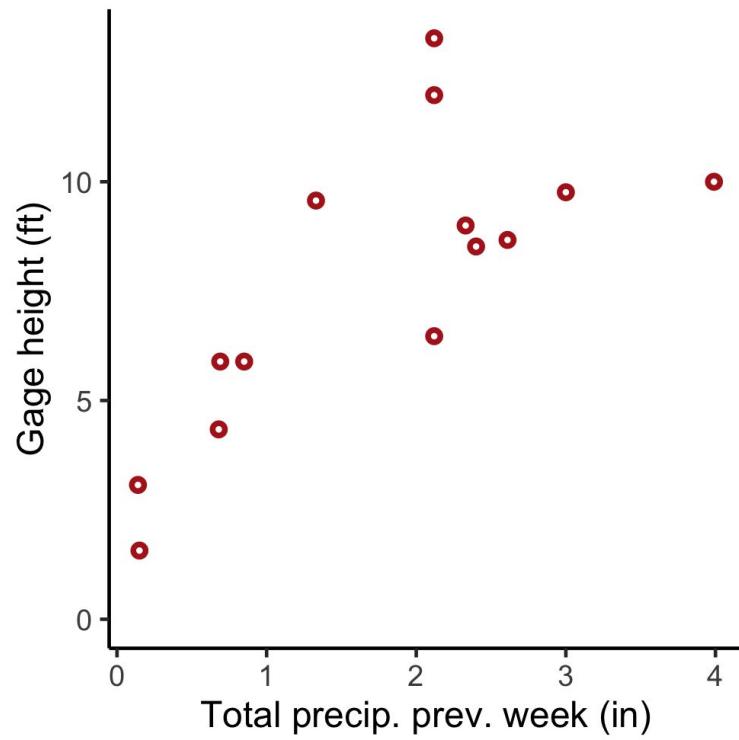
Residuals
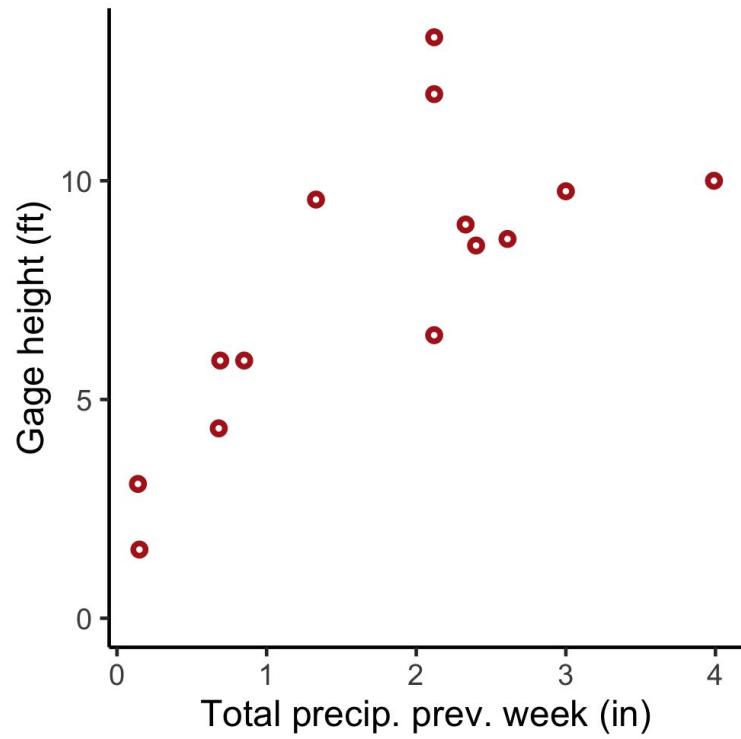
Exhaustive search
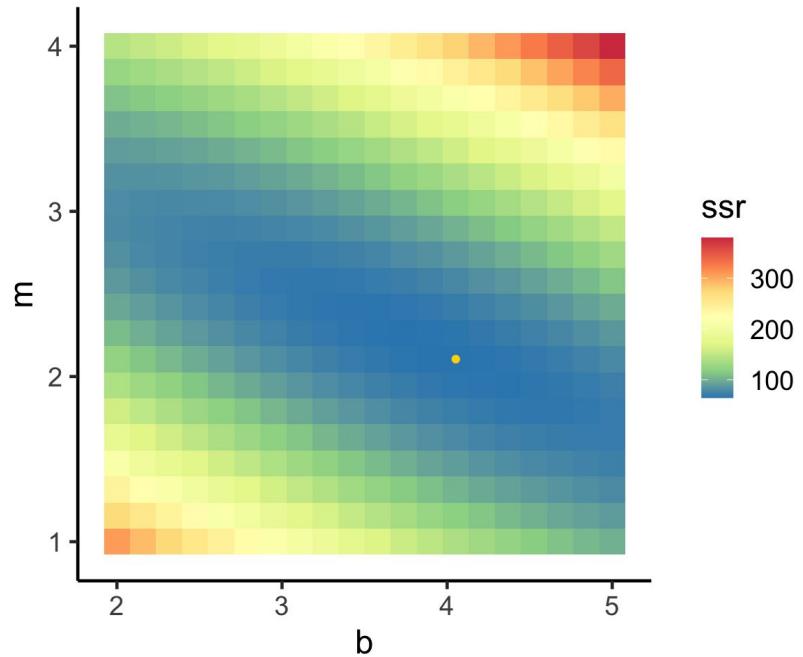
lm() solves it for you

```
> storm_mod <- lm(gage_ht ~ precip_wk_in, data = storm_data)
> summary(storm_mod)

Call:
lm(formula = gage_ht ~ precip_wk_in, data = storm_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6688 -1.1225 -0.6316  0.3928  4.7668

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.9133     1.1674   3.352  0.00576 **
precip_wk_in   2.1697     0.5645   3.844  0.00234 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.32 on 12 degrees of freedom
Multiple R-squared:  0.5518,    Adjusted R-squared:  0.5144
F-statistic: 14.77 on 1 and 12 DF,  p-value: 0.002337
```

# Fitting lines to points

Today's agenda

➔ Fitting lines to points

➔ **Quality of the fit**
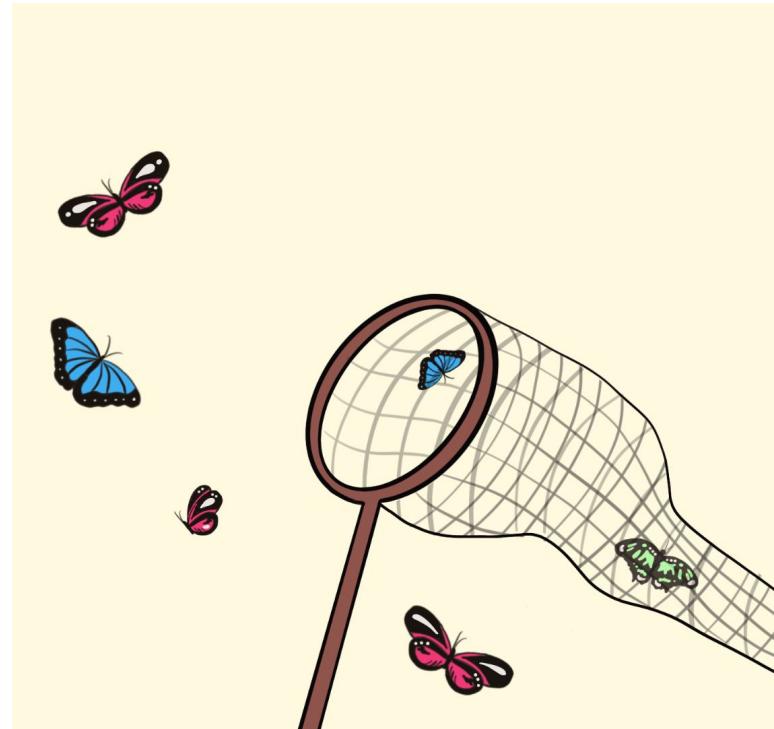
➔ Regression and RVs

The purpose of a model
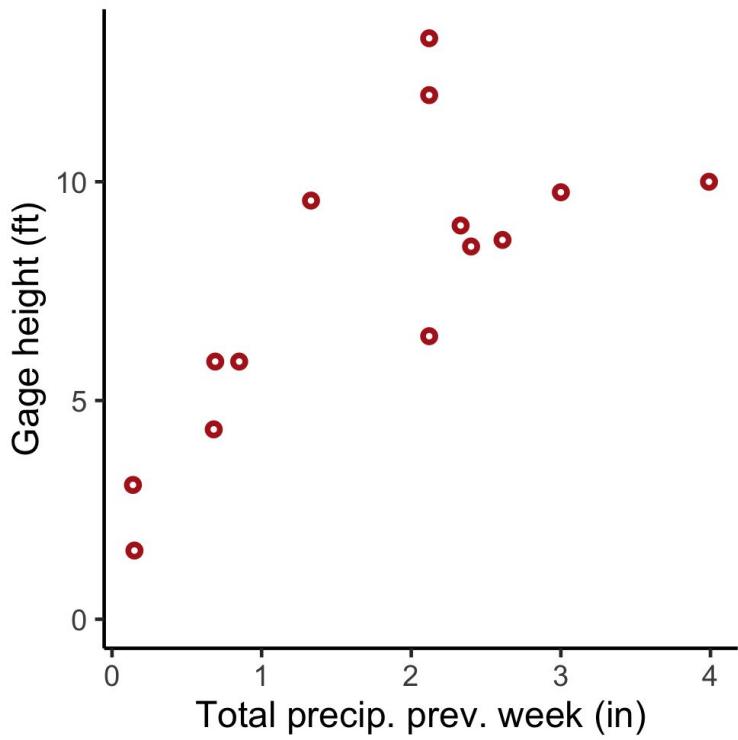
Variation of the response
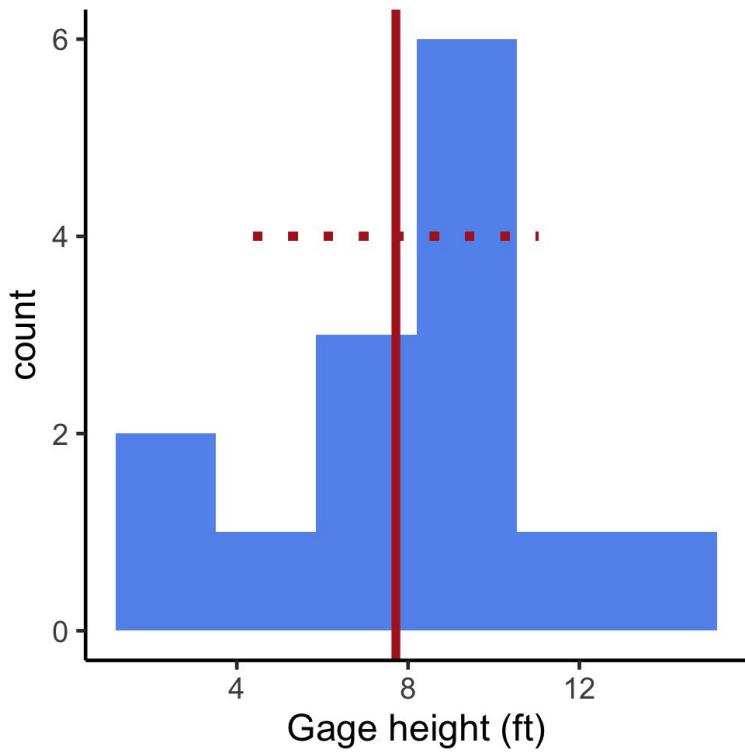
Variation of the residual
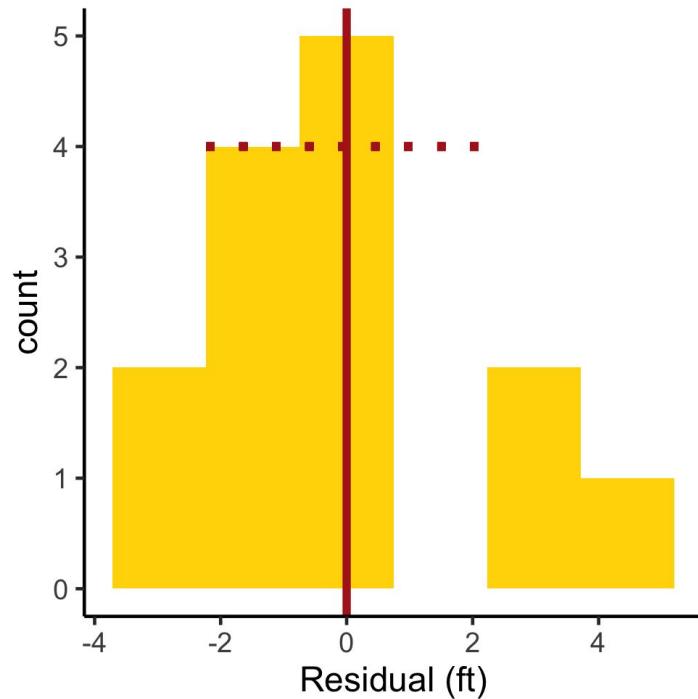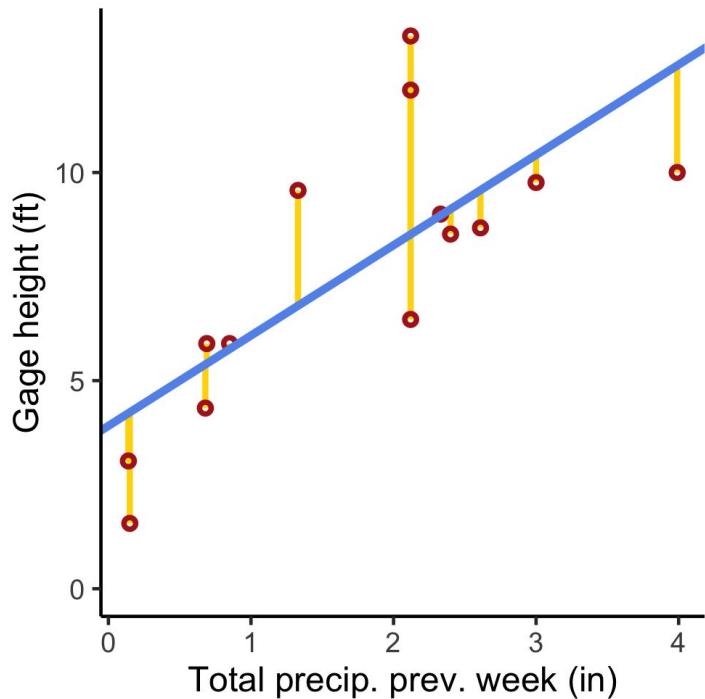
A measure of variance explained: R^2

Caution!
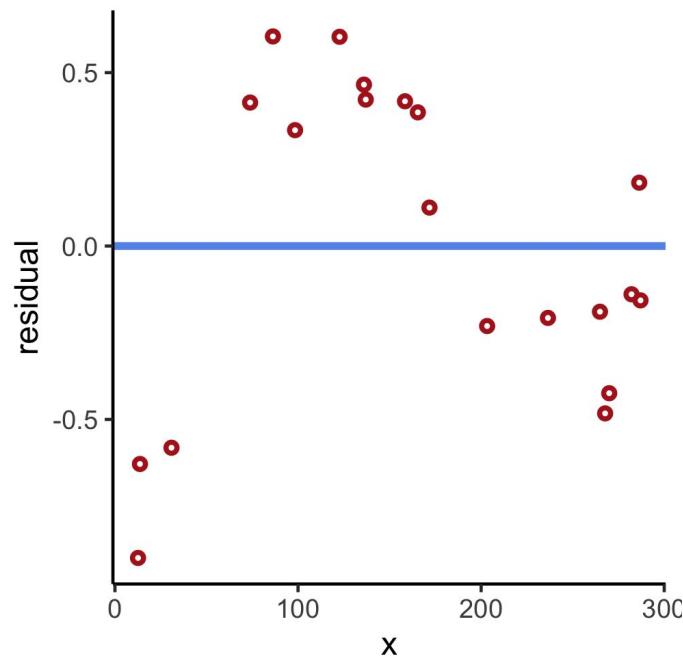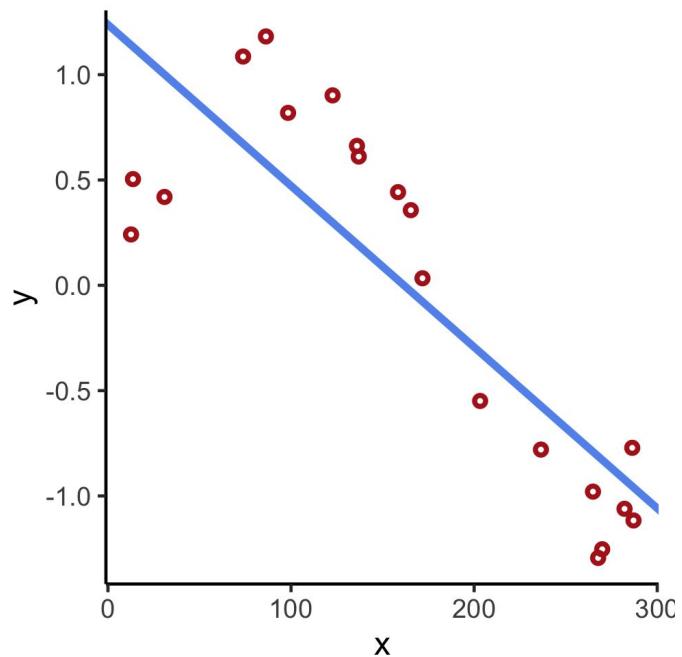
Quality of the fit

Today's agenda

➜ Fitting lines to points

➜ Quality of the fit
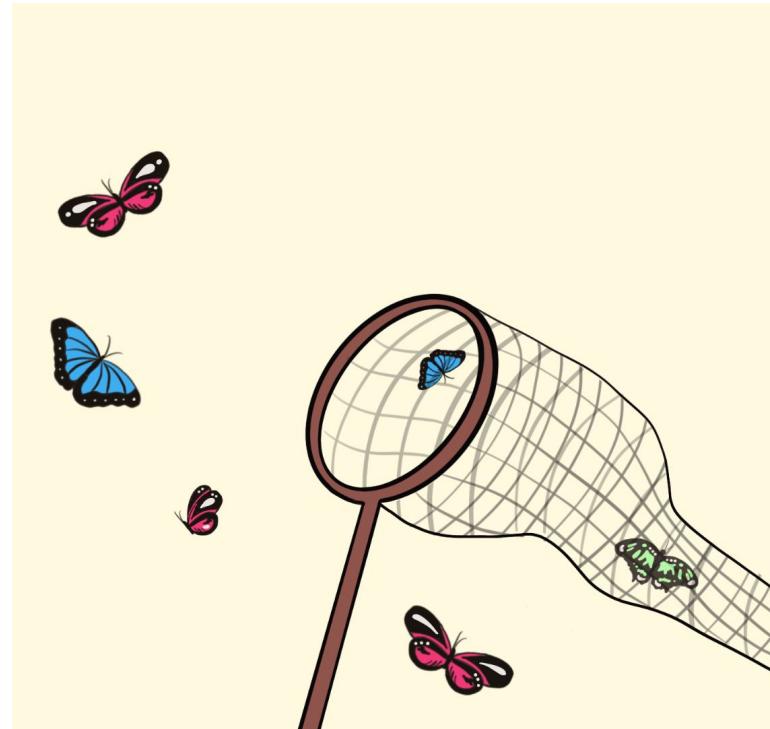
➜ **Regression and RVs**

Data generating process

Population, sample, and regression

Caution!
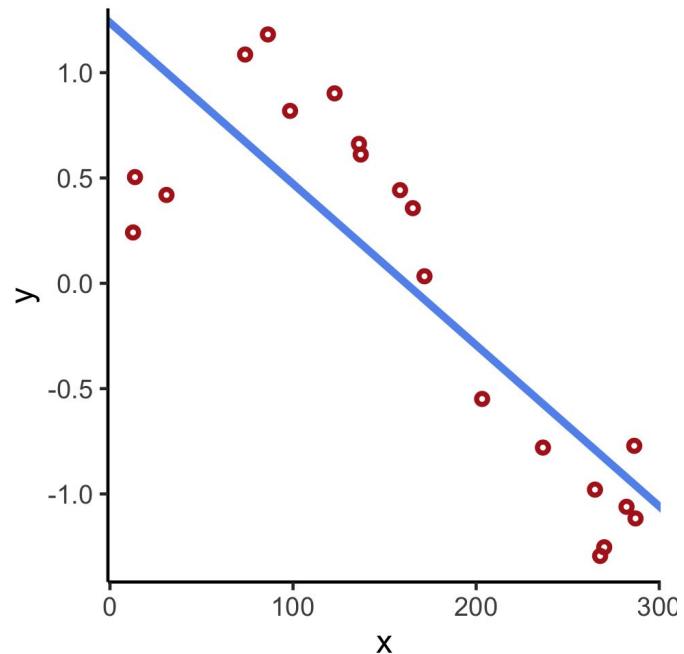
Recap

Fit lines to points

Families of random variables

Simulating data