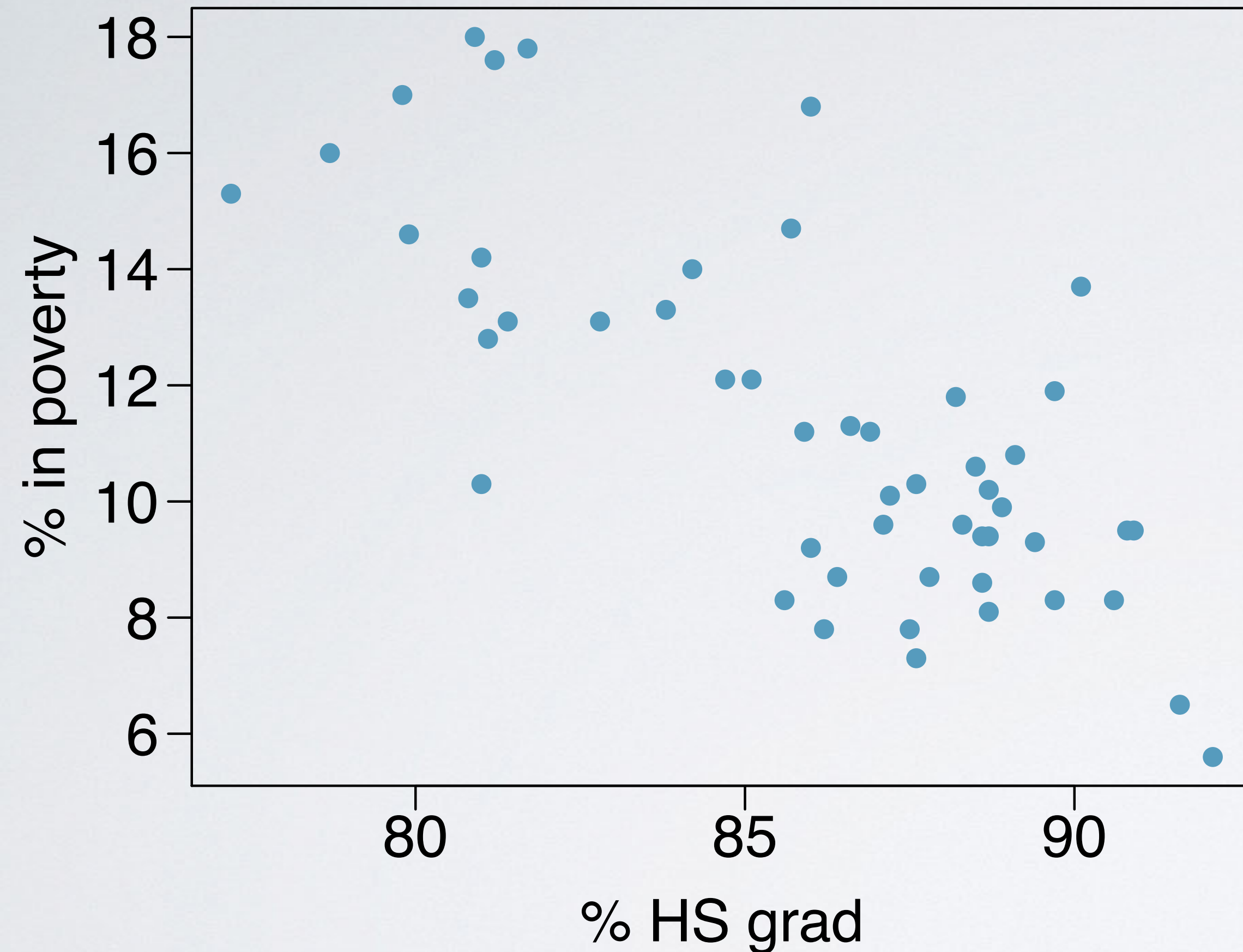


correlation

- ▶ definition
- ▶ properties

poverty vs. HS grad rate



- ▶ data: 50 states + DC
- ▶ poverty line: income below \$23,050 for a family of 4 in 2012

Response?

% in poverty

Explanatory?

% HS grad

Relationship?

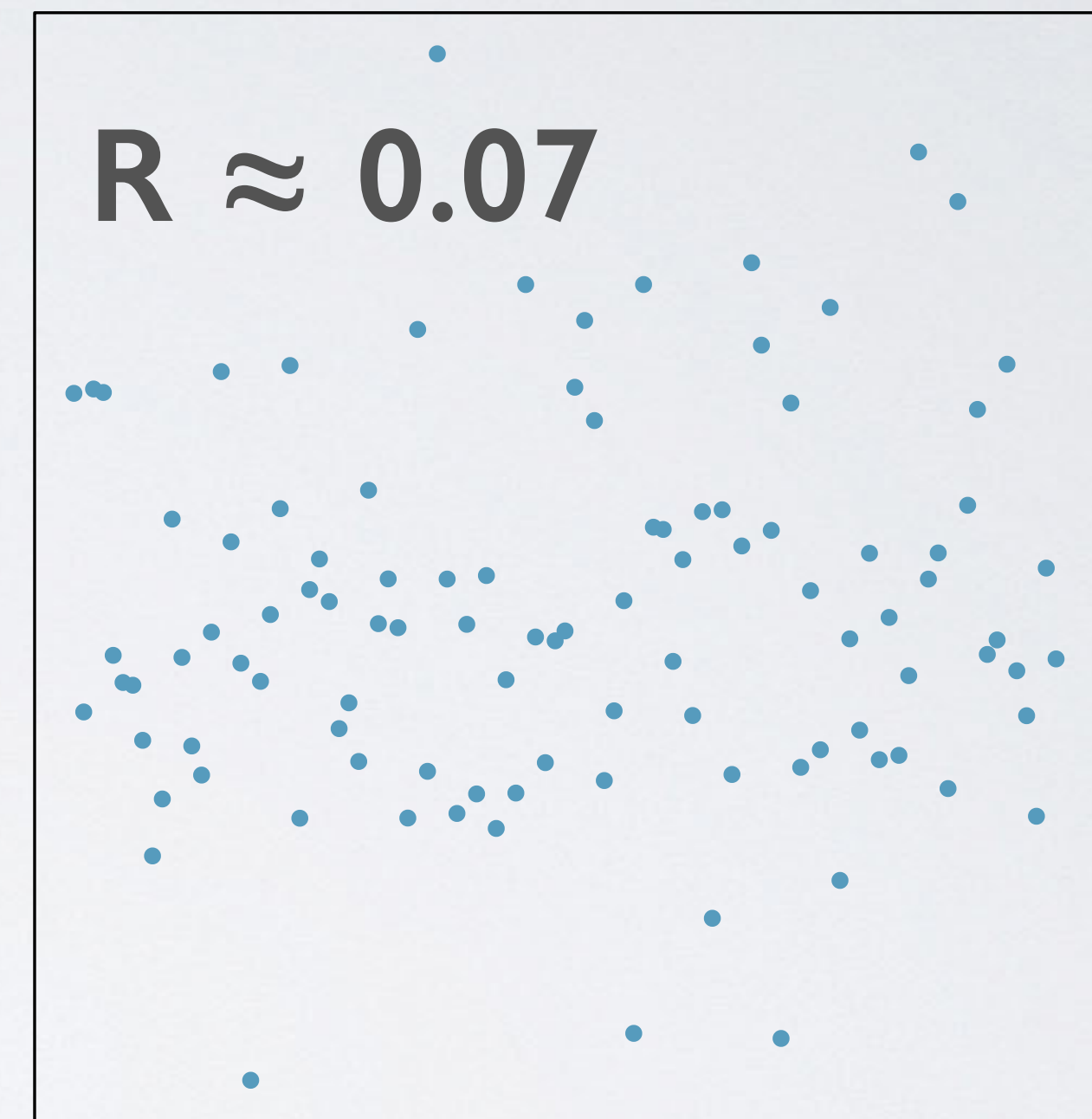
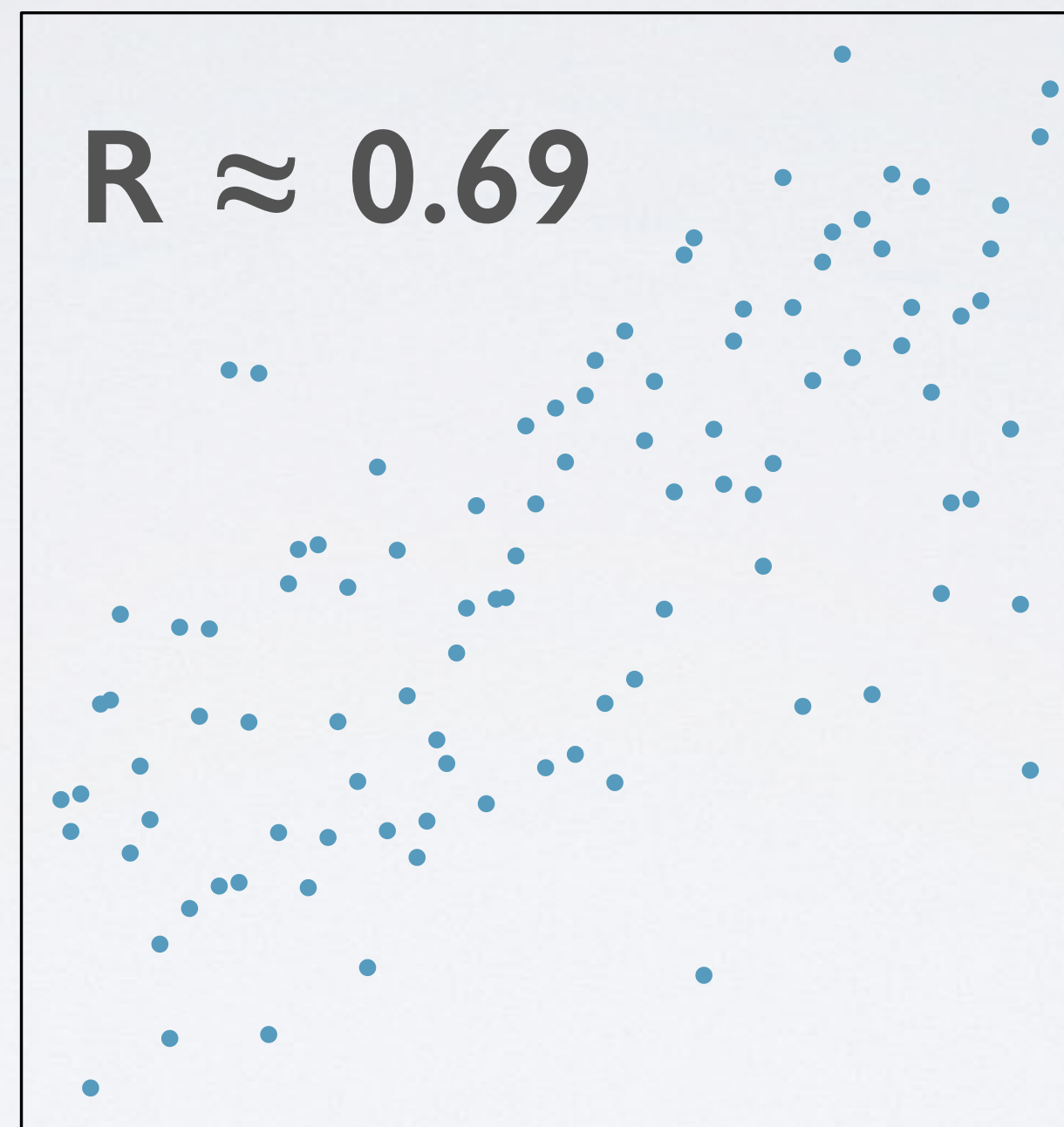
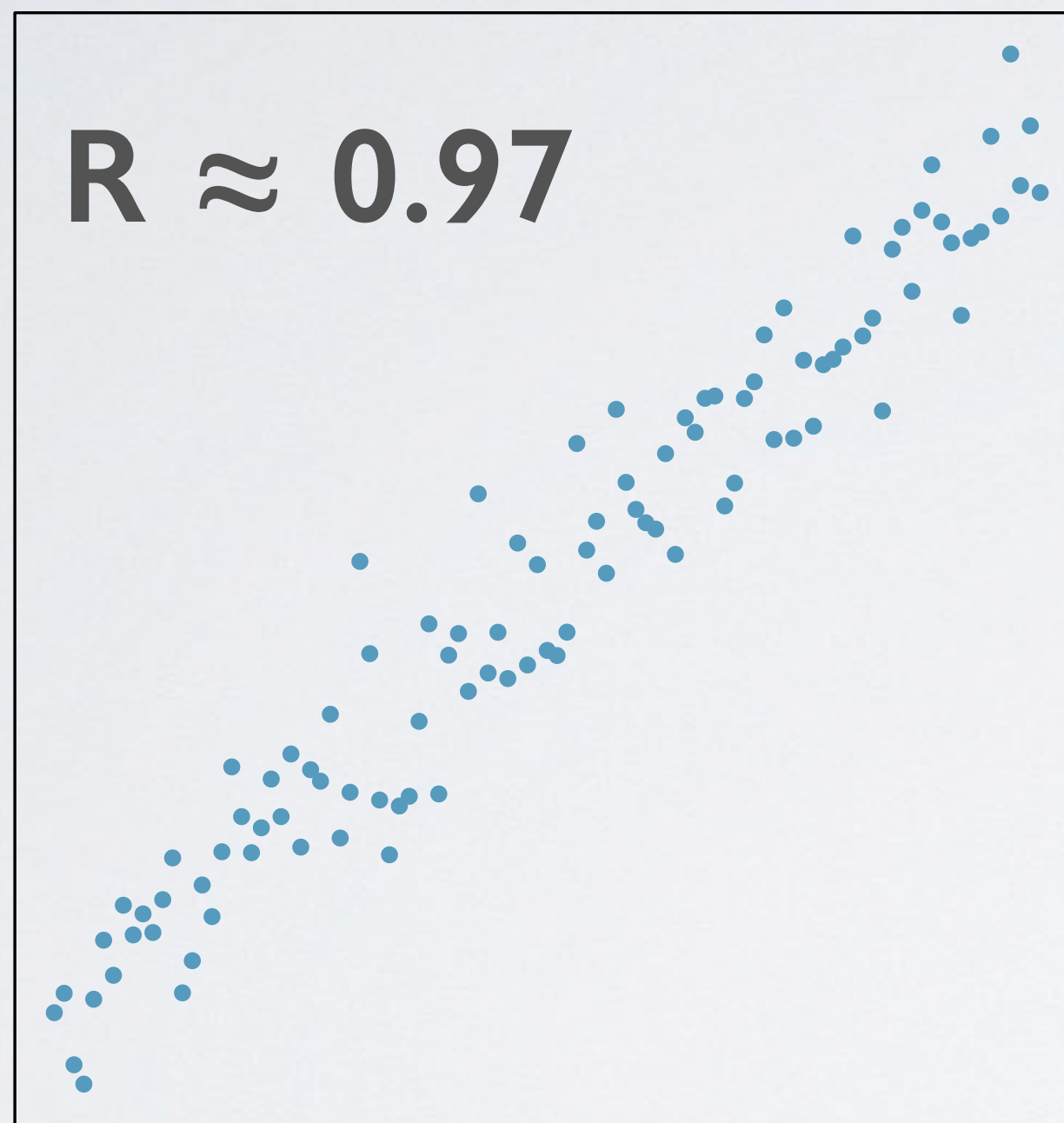
*linear, negative,
moderately strong*

correlation

- ▶ describes the strength of the linear association between two variables
- ▶ denoted as R

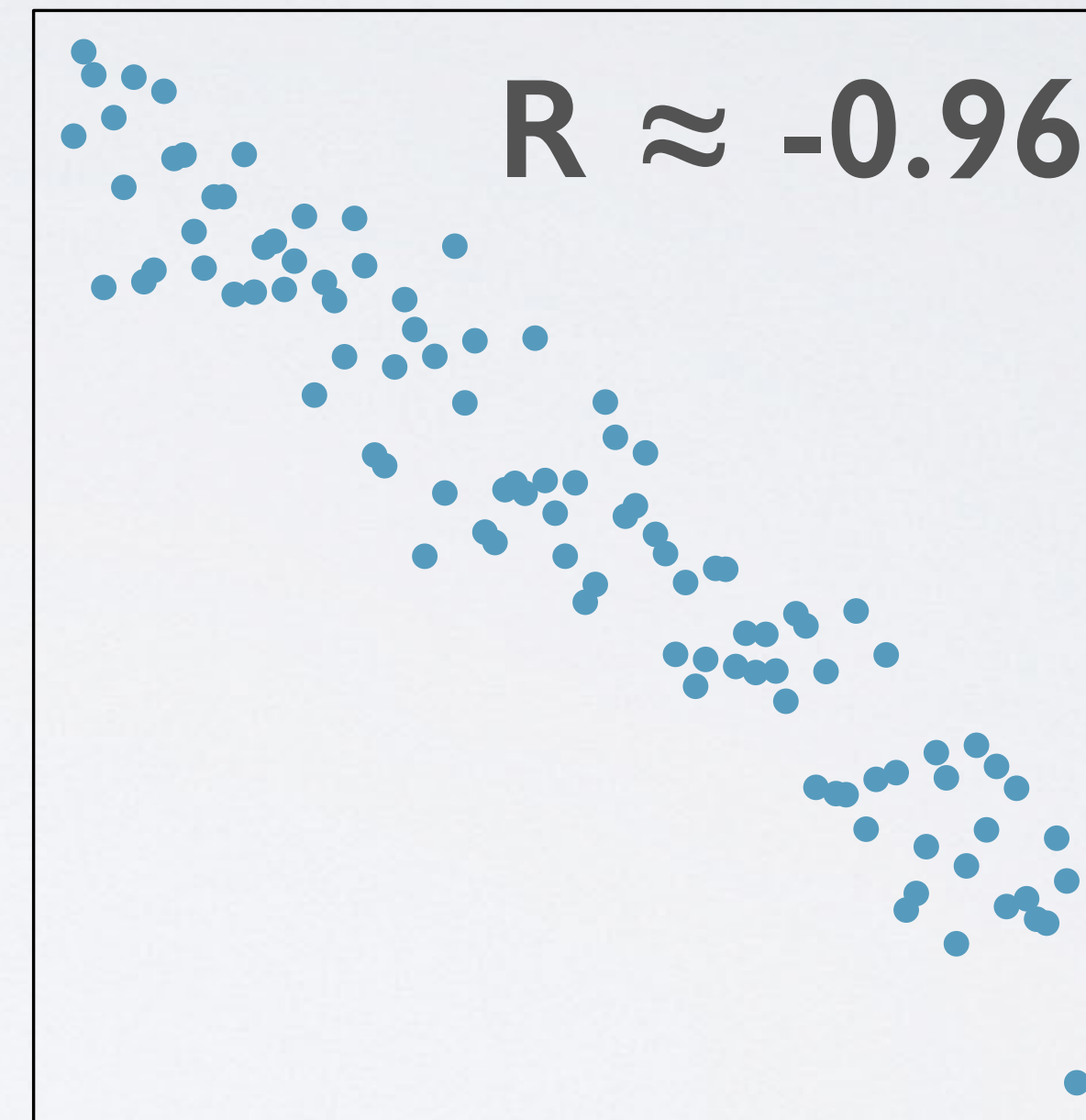
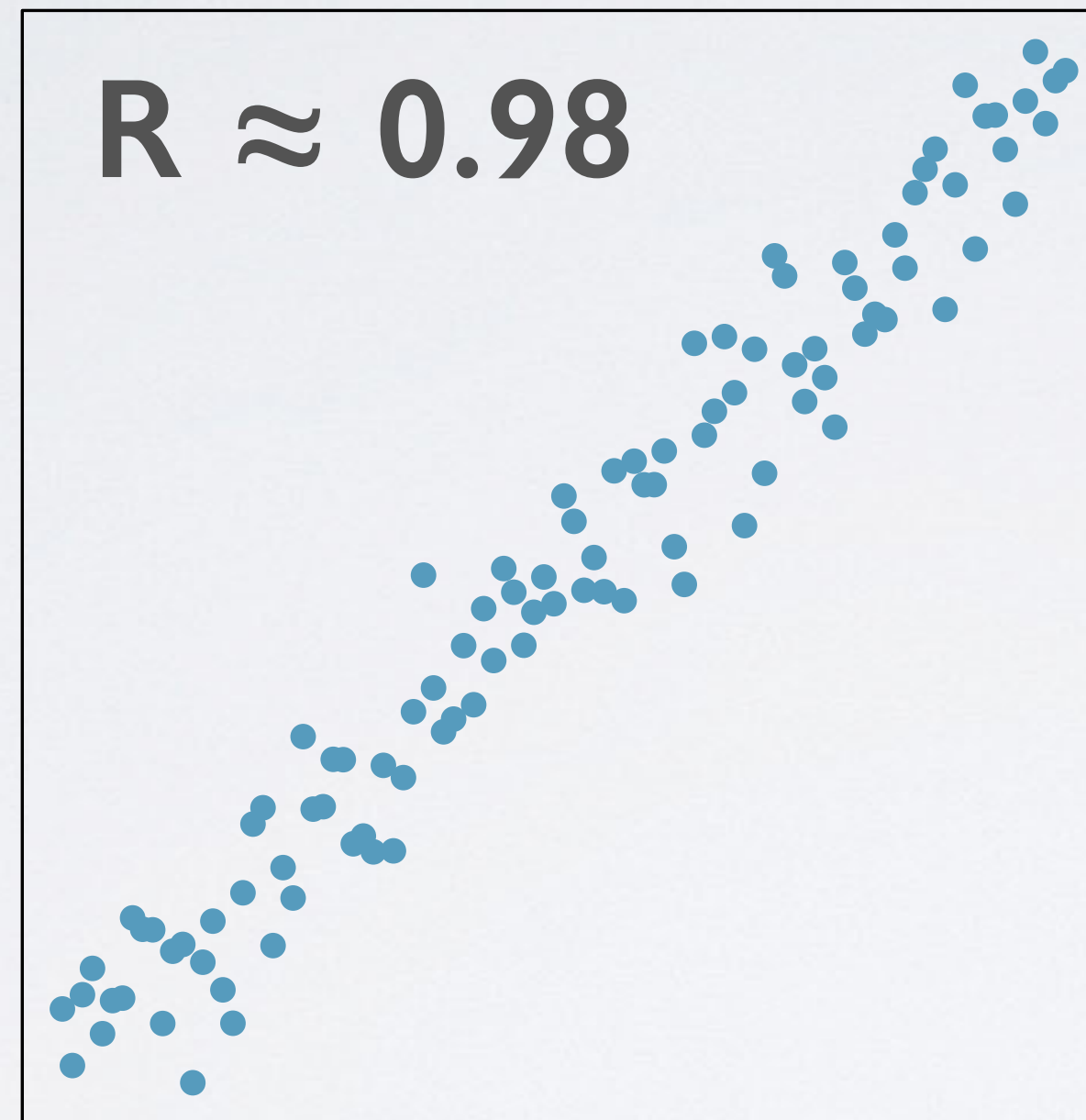
properties (I)

the magnitude (absolute value) of the correlation coefficient measures the strength of the linear association between two numerical variables



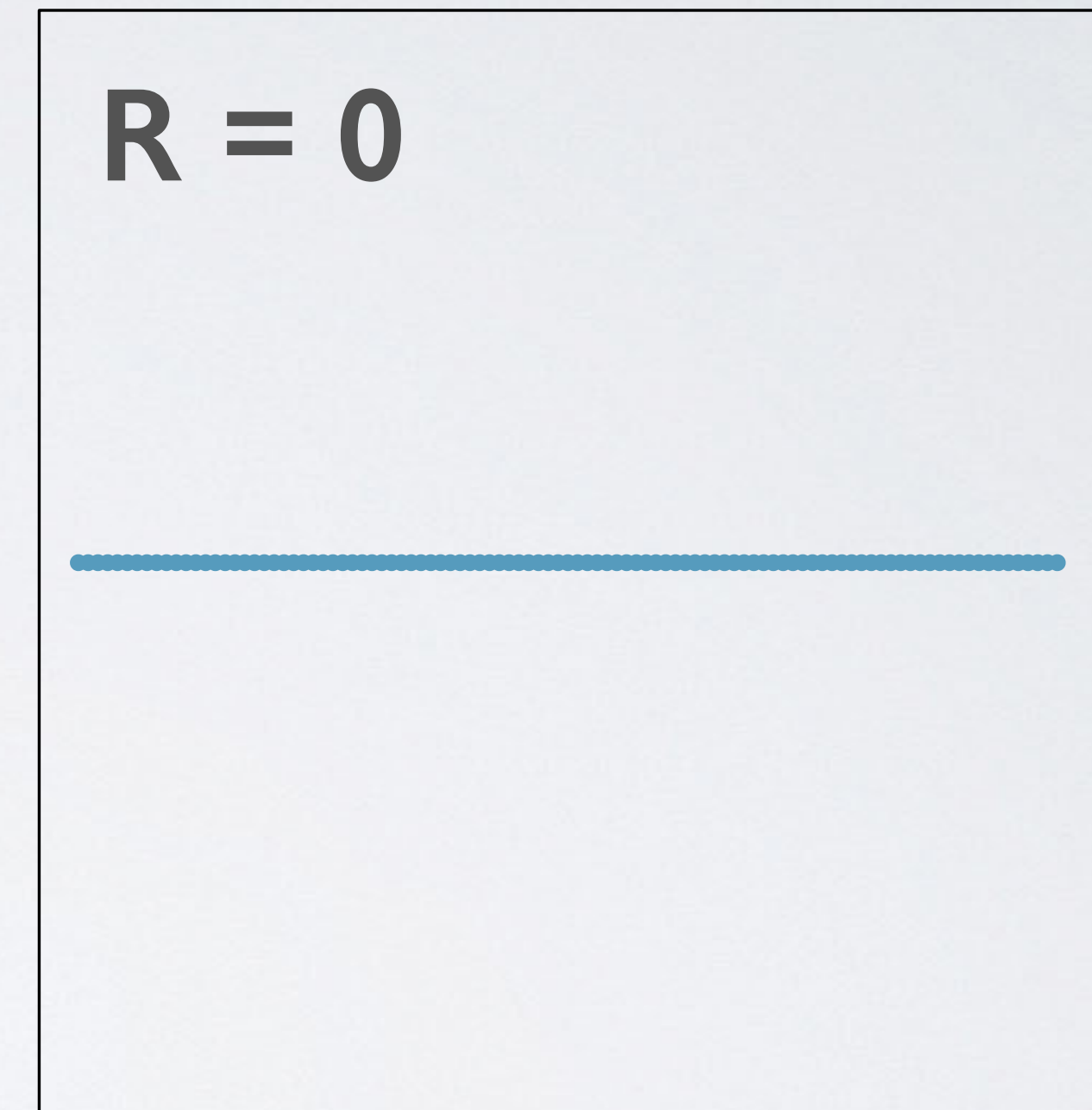
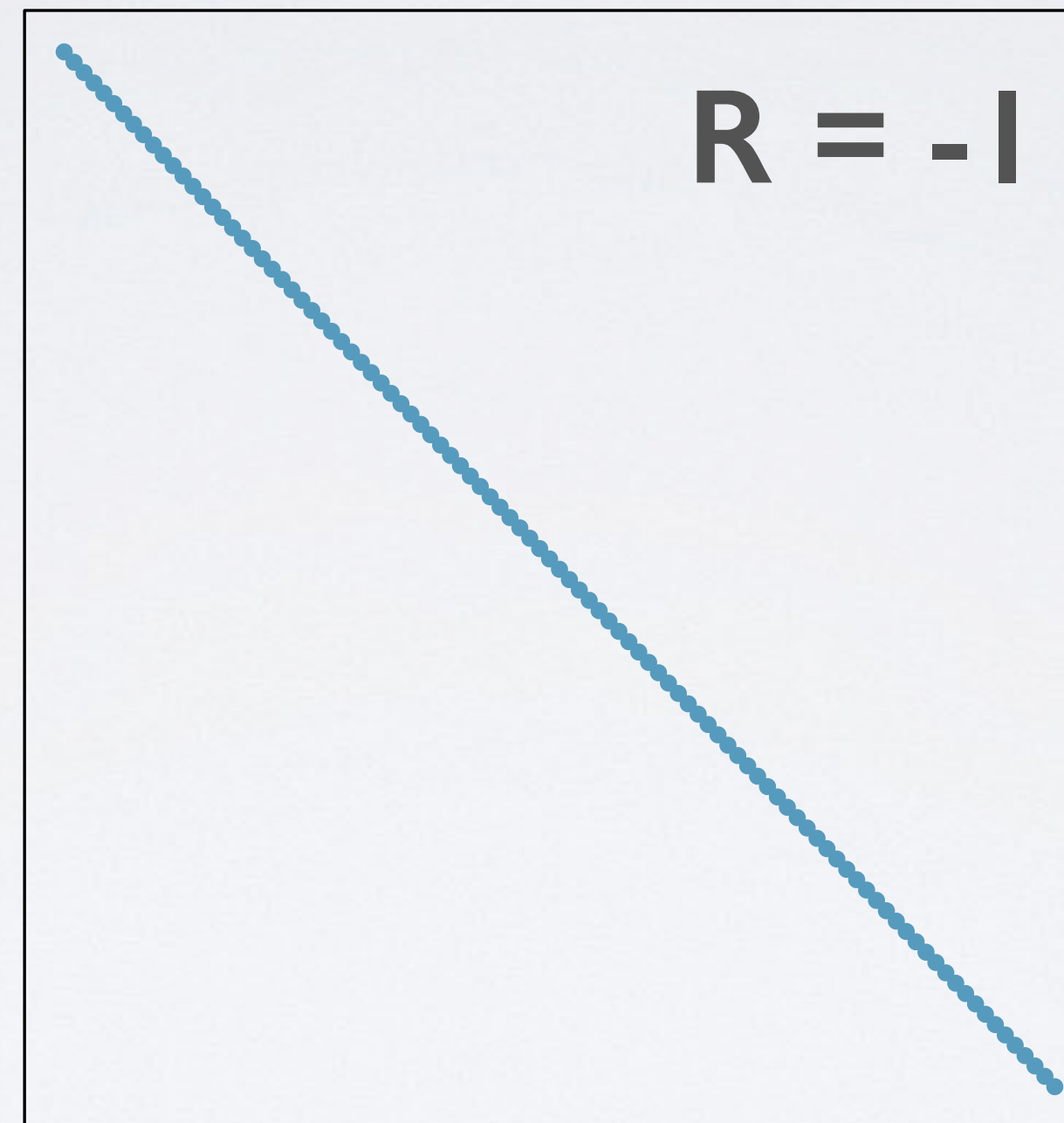
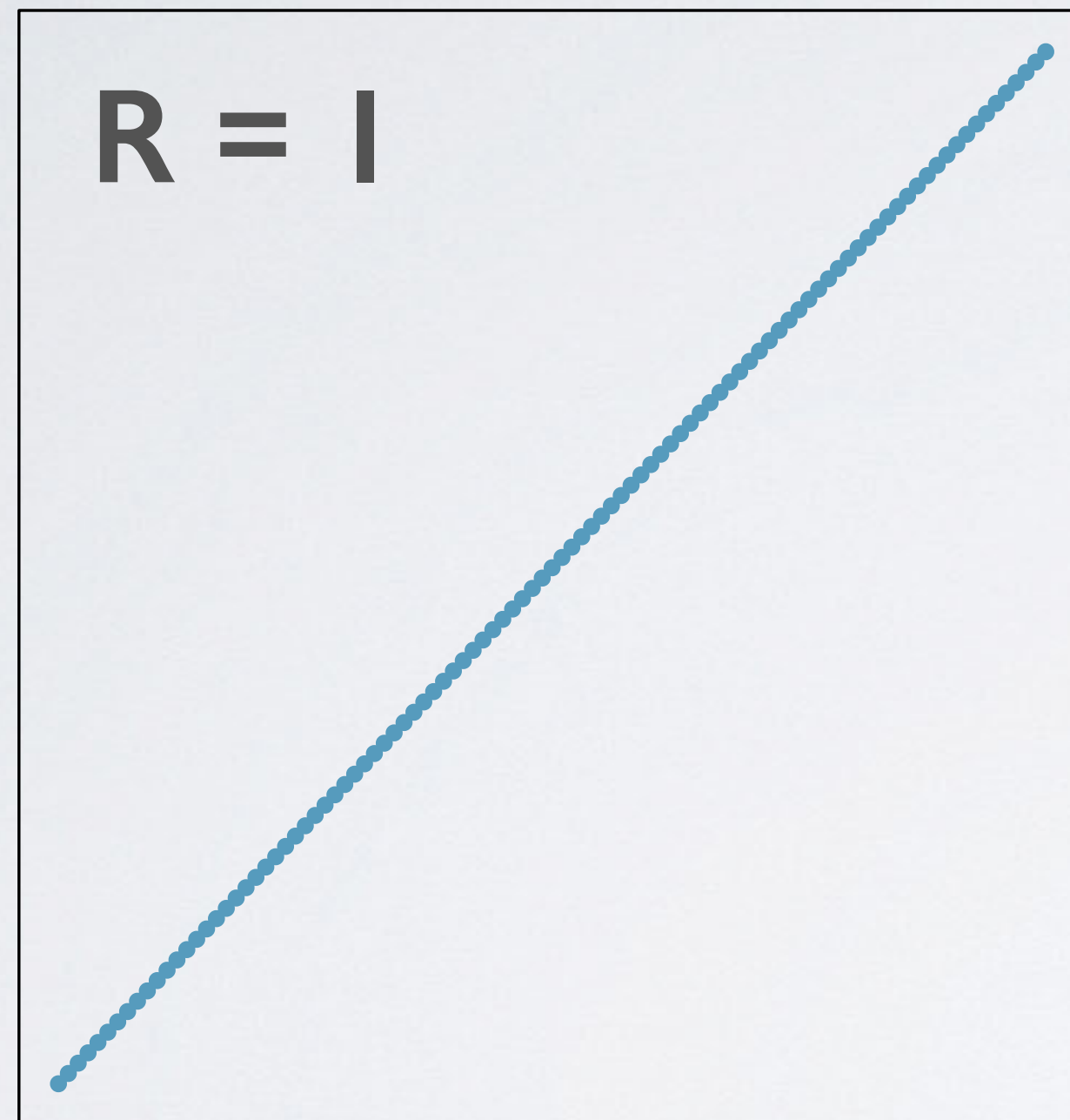
properties (2)

the sign of the correlation coefficient indicates the direction of association



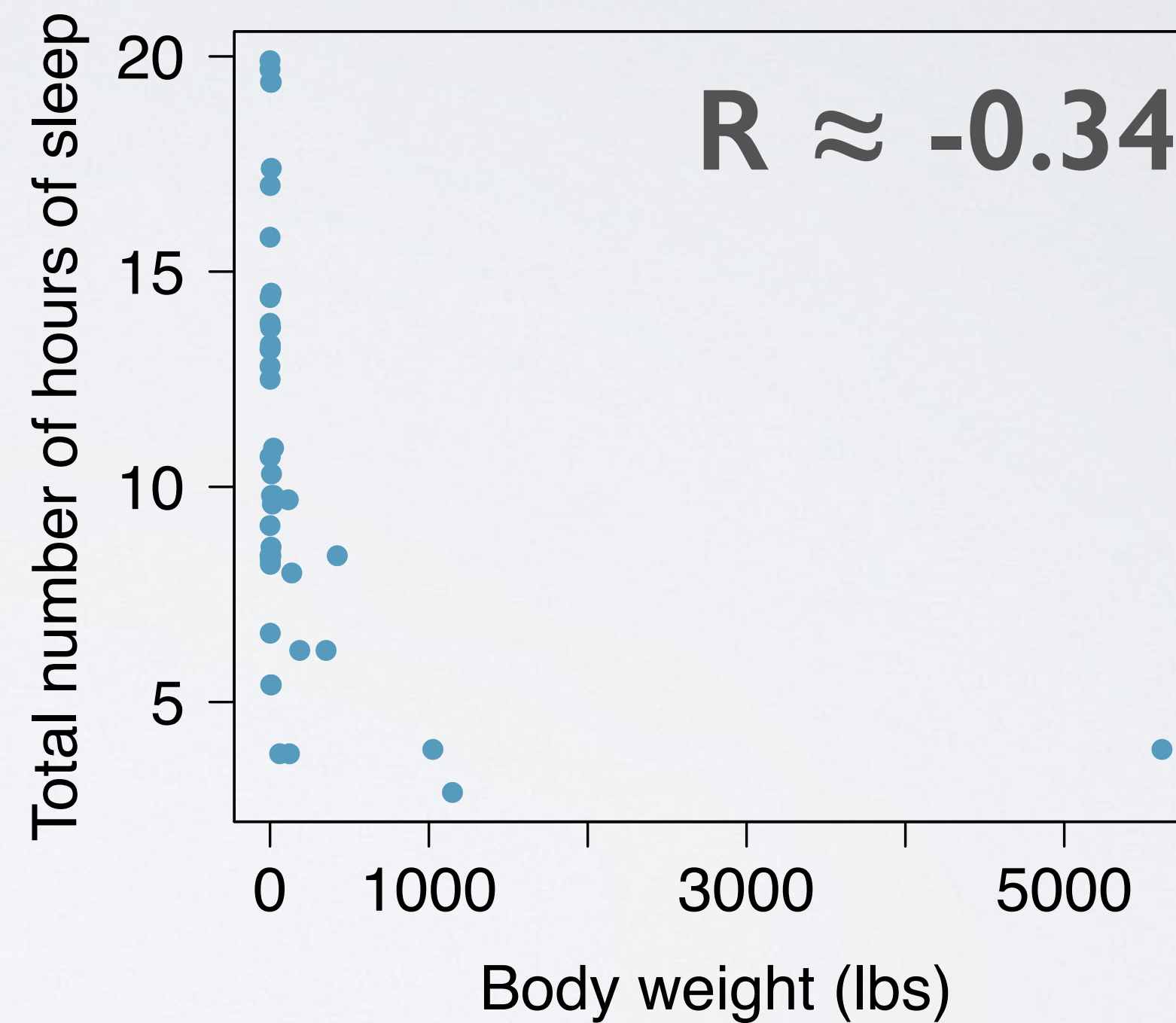
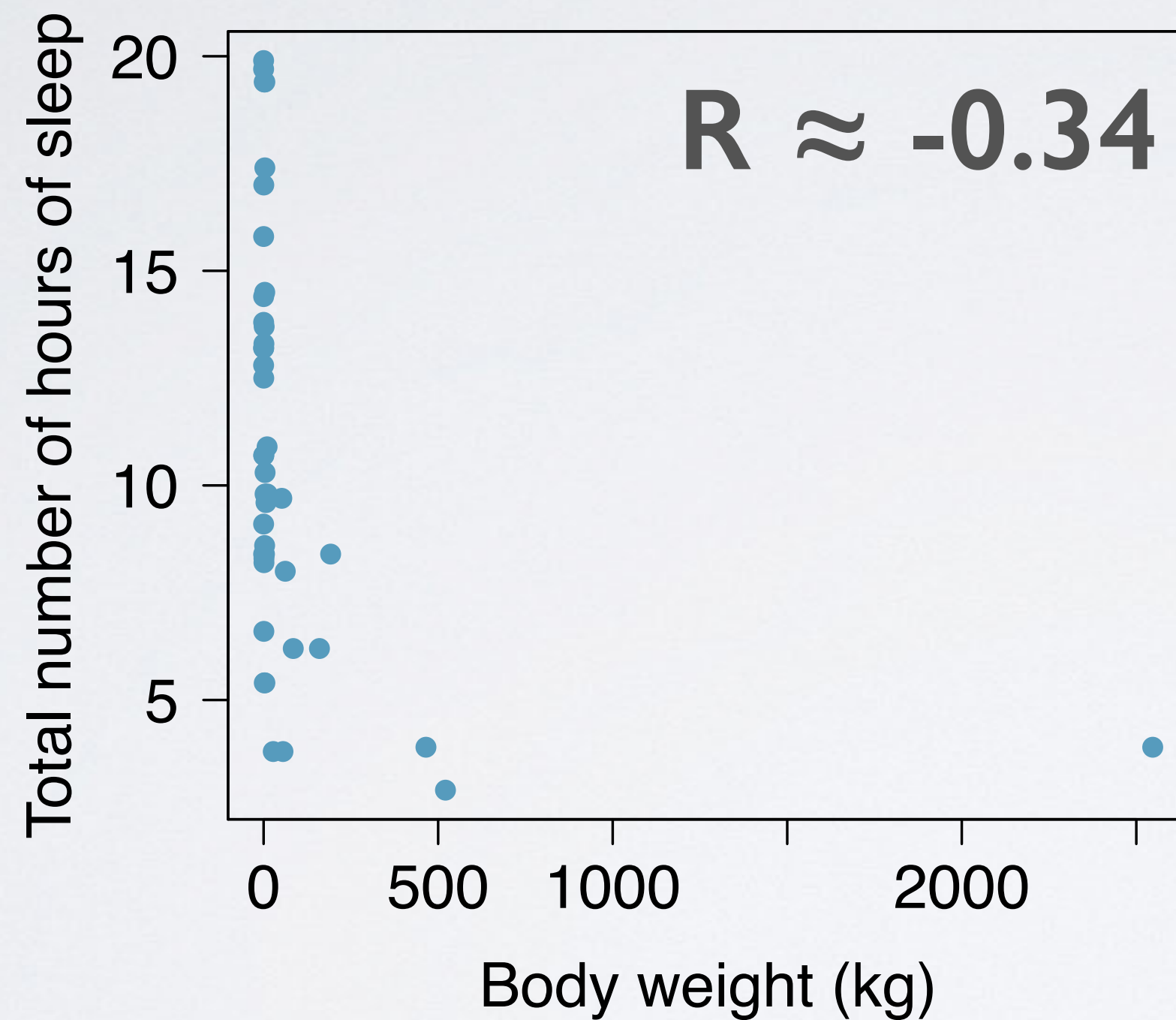
properties (3)

- ▶ the correlation coefficient is always between -1 (perfect negative linear association) and 1 (perfect positive linear association)
- ▶ $R = 0$ indicates no linear relationship



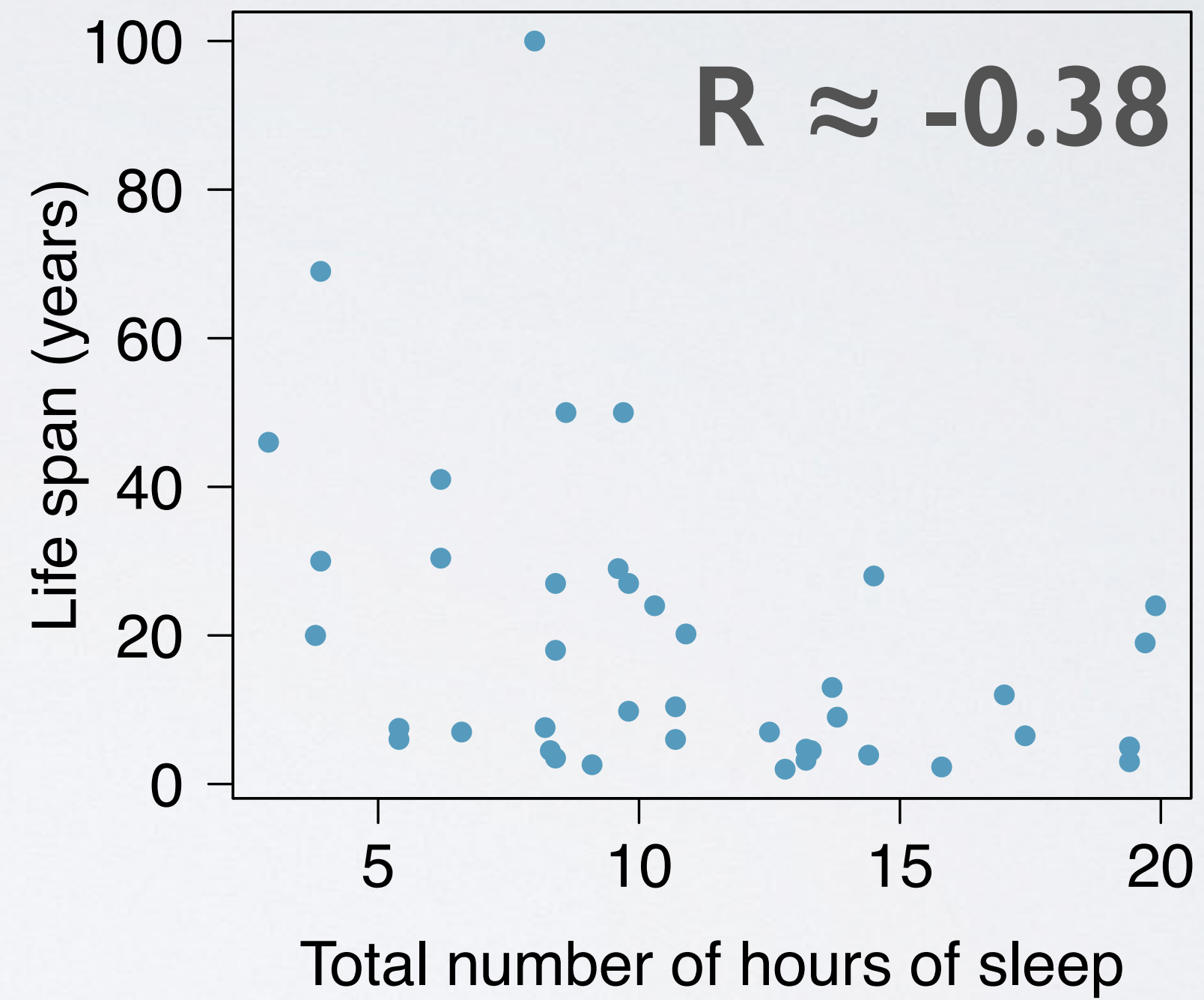
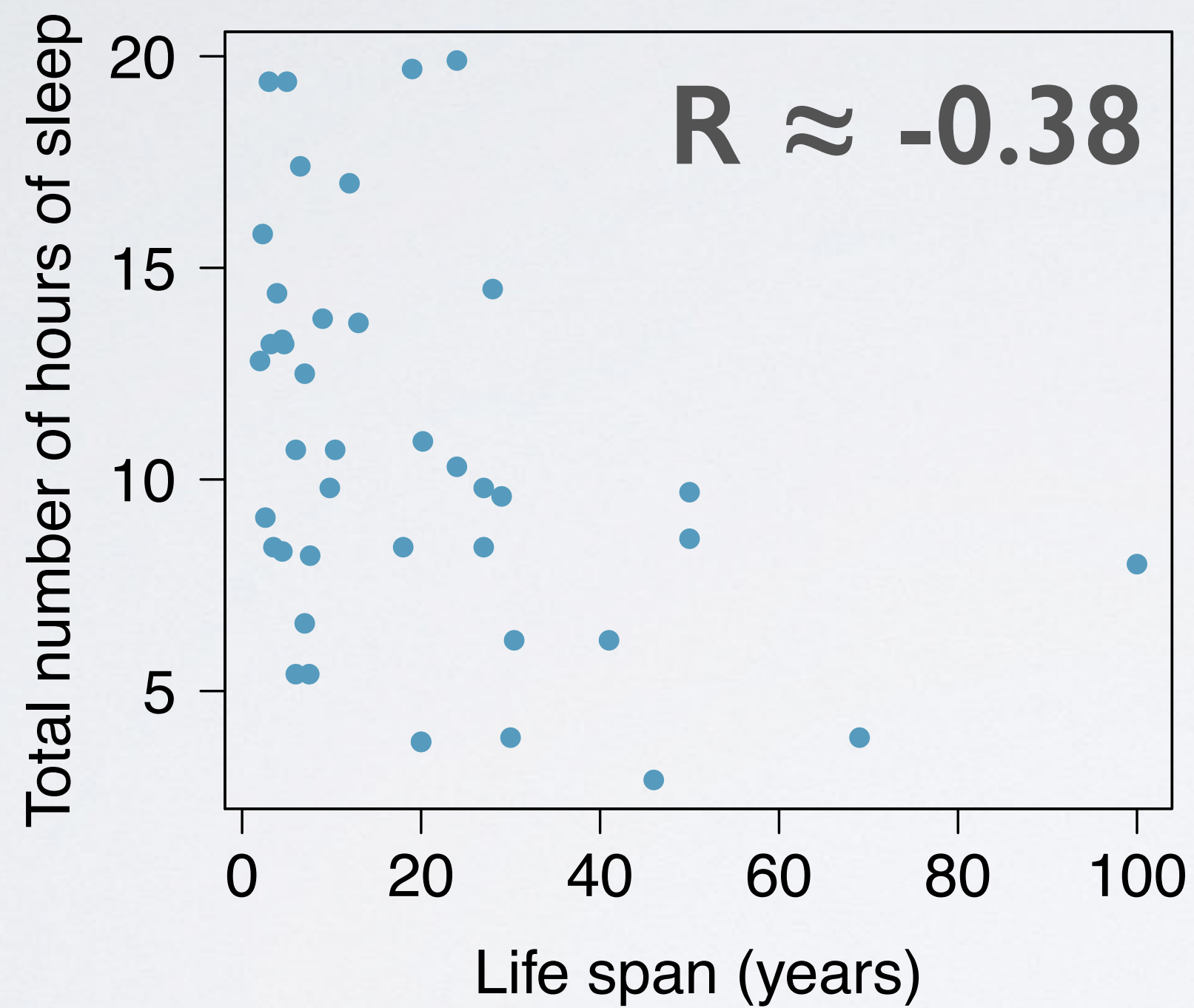
properties (4)

the correlation coefficient is unitless, and is not affected by changes in the center or scale of either variable (such as unit conversions)



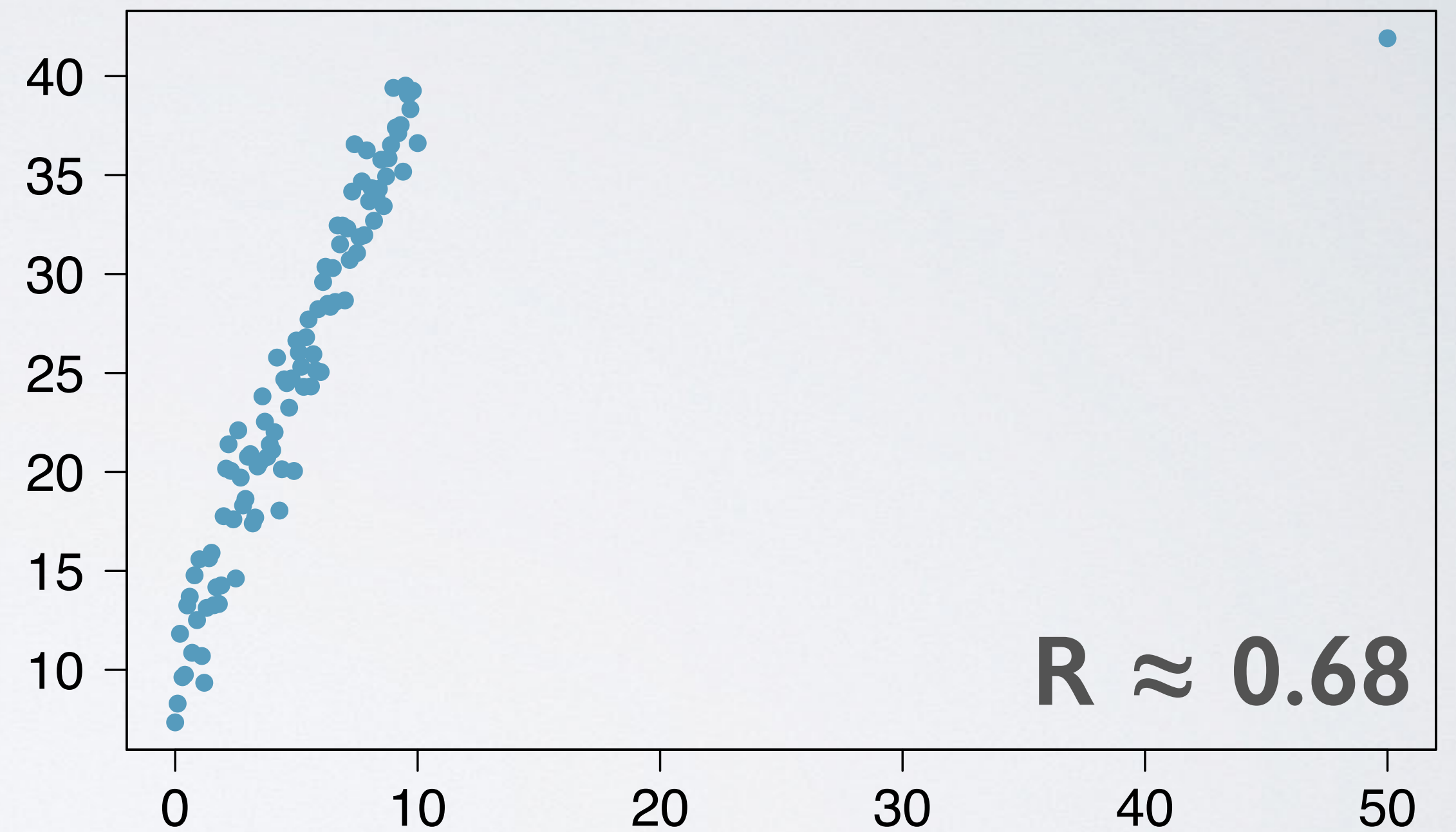
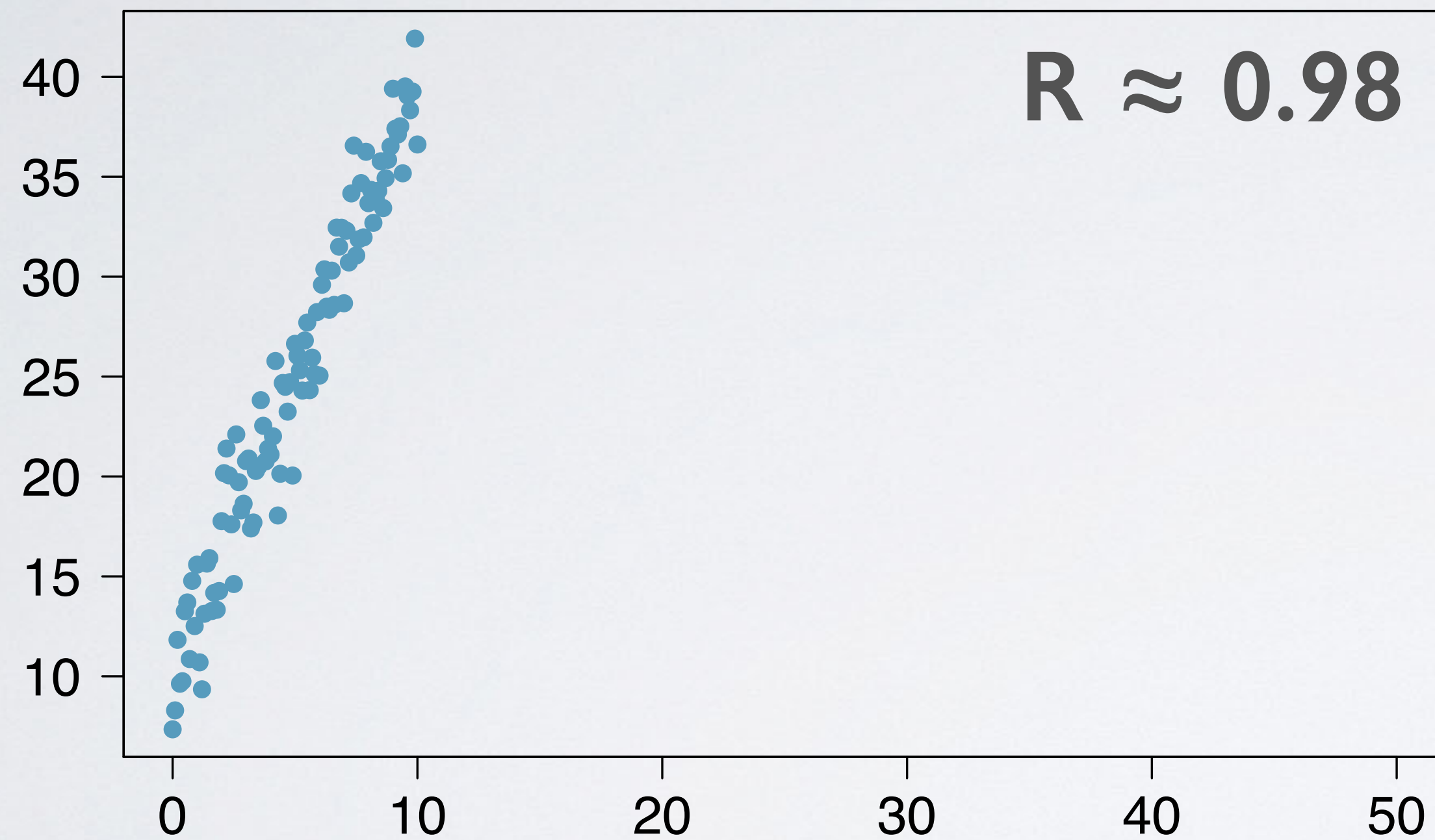
properties (5)

the correlation of X with Y is the same as of Y with X

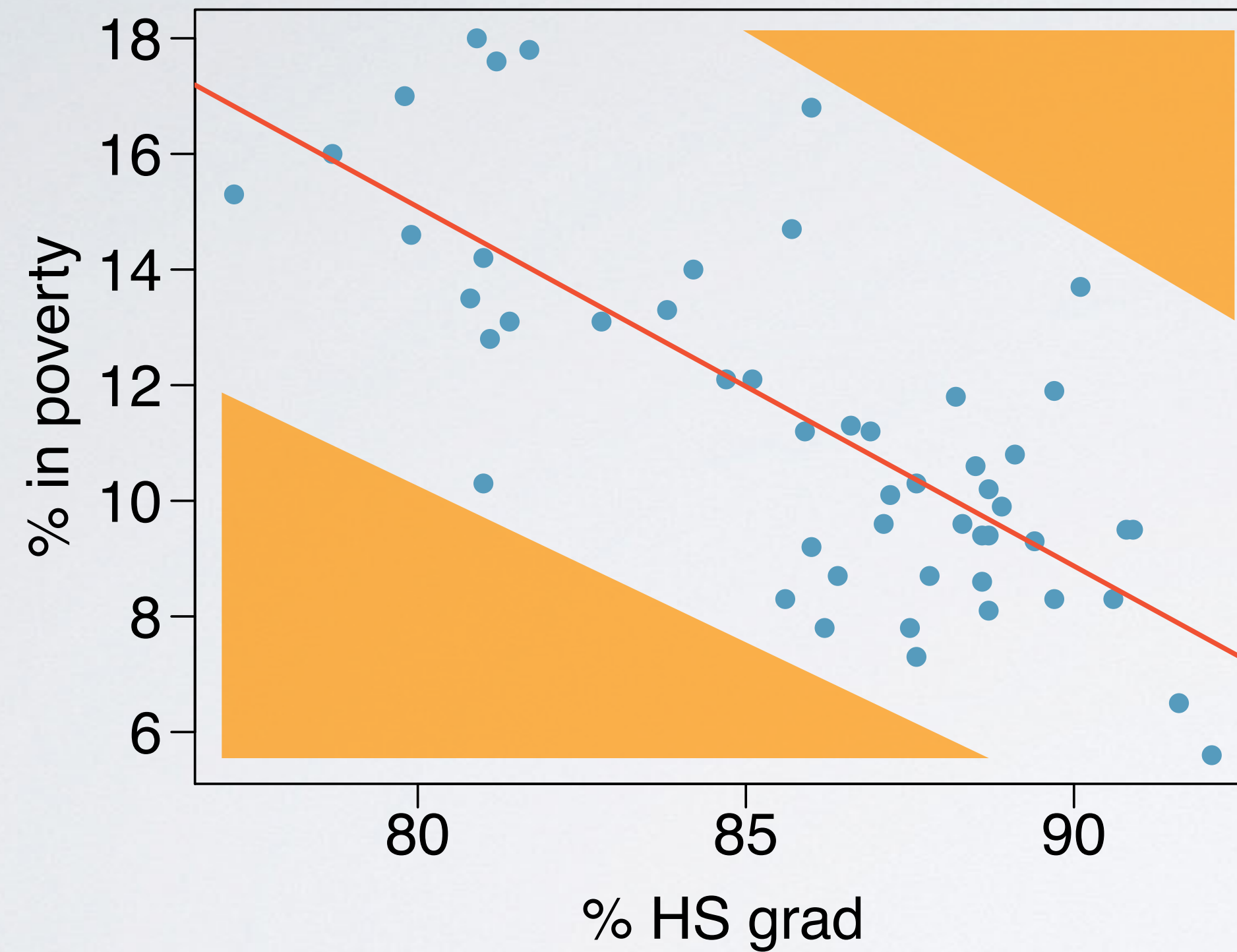


properties (6)

the correlation coefficient is sensitive to outliers



Which of the following is the best guess for the correlation between % in poverty and % HS grad?



~~(a) 0.6~~

☒ (b) -0.75

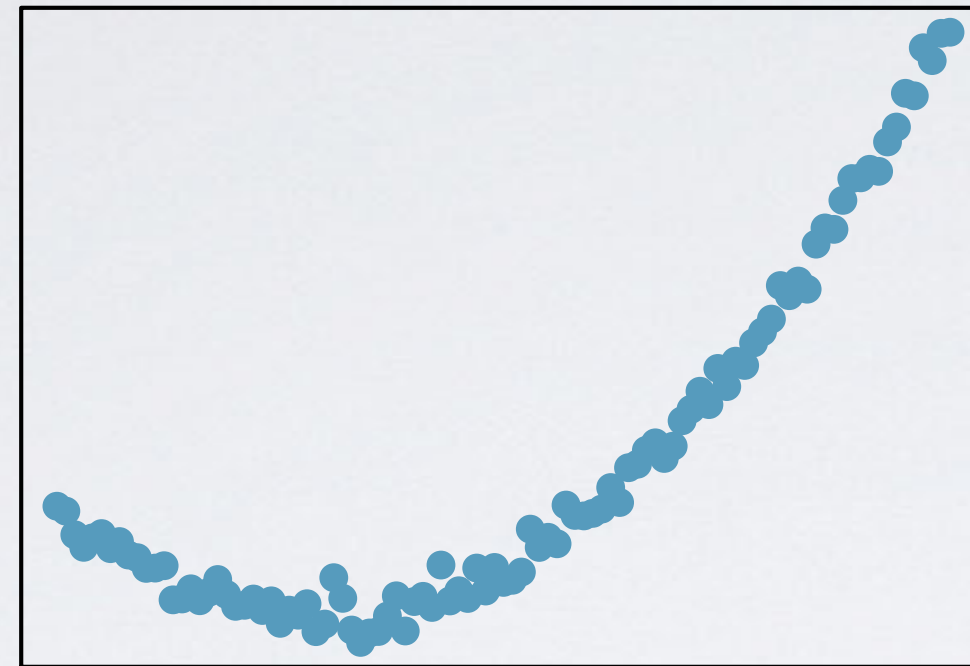
~~(c) -0.1~~

~~(d) 0.02~~

~~(e) -1.5~~

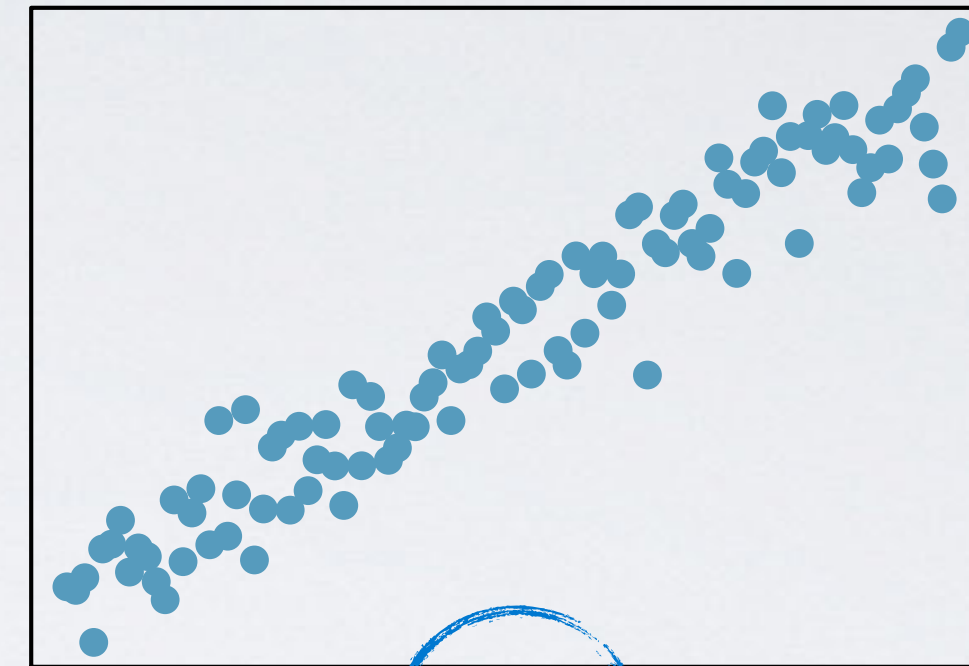
Which of the following has the strongest correlation, i.e. correlation coefficient closest to $+1$ or -1 ?

*very strong,
but not linear*



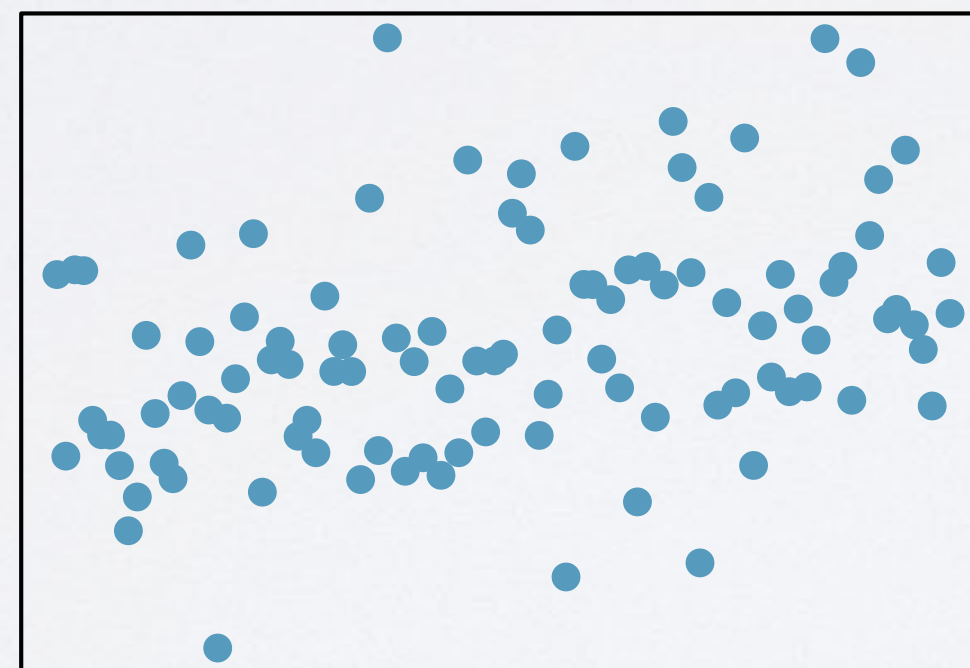
(a)

*strongest
linear*



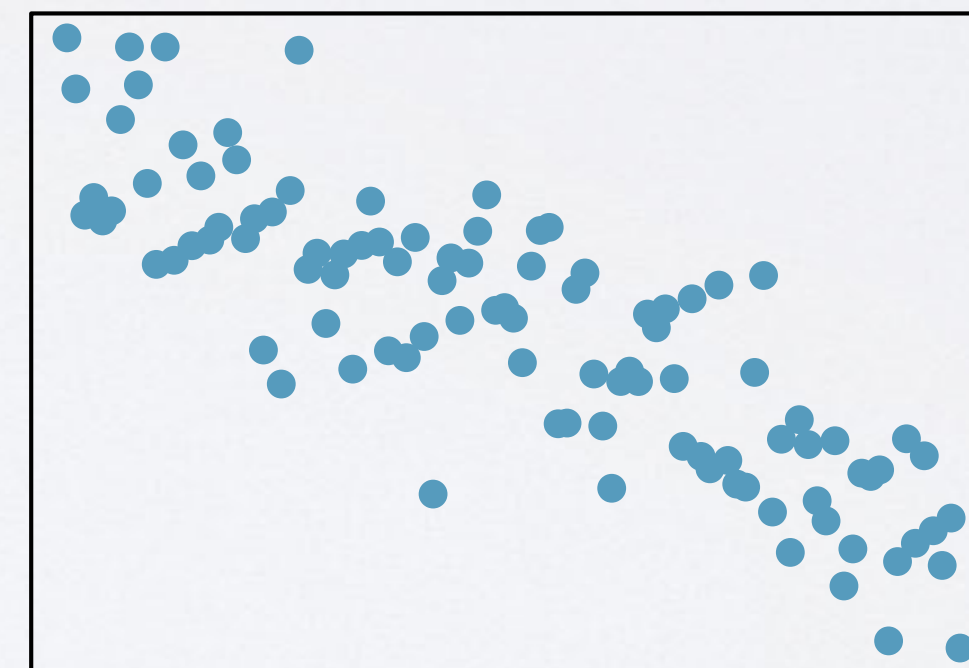
(b)

even weaker



(c)

weaker



(d)