Lectures 5-7

Simple Linear Regression

DSC 40A, Fall 2025

Agenda

- Simple linear regression. → Least squares solution
- Correlation.
- Interpreting the formulas.
- Connections to related models.
- HW1 due tonight
- HW2 will be released today
- Submit regrade requests → no need for emails
- Can ask private questions on Campuswire!

Least squares solutions

- Our goal was to find the parameters w_0^* and ${w_1}^*$ that minimized:

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (w_0 + w_1 x_i))^2$$

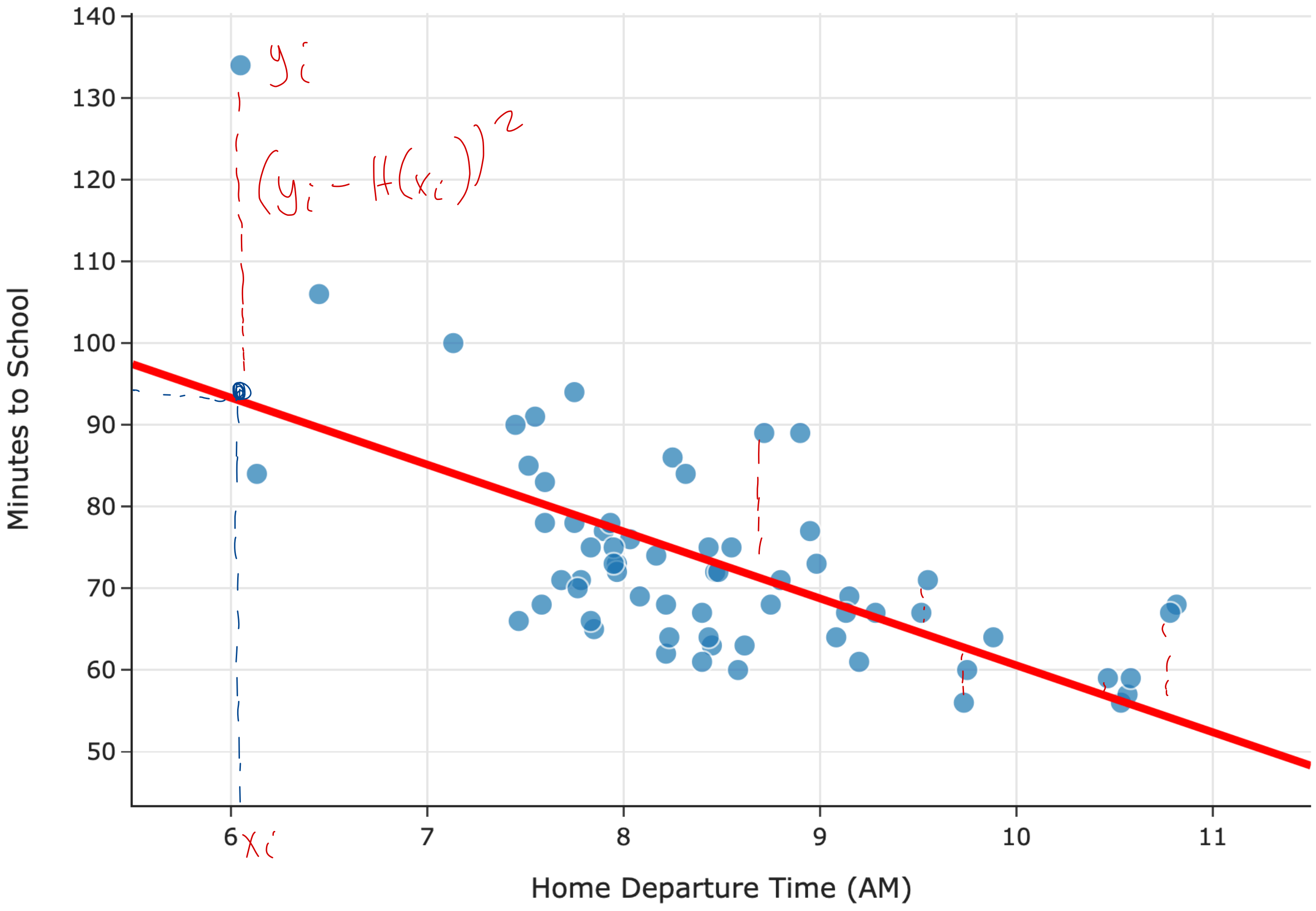
- To do so, we used calculus, and we found that the minimizing values are:

$$\boxed{w_1^* = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad\qquad w_0^* = \bar{y} - w_1^* \bar{x}}$$

- We say w_0^* and w_1^* are **optimal parameters**, and the resulting line is called the **regression line**.

optimal intercept optimal slope

Predicted Commute Time = 142.25 - 8.19 * Departure Hour



There is no other line that has a smaller MSE.

$$w_0^*, w_1^* = \arg\min_{w_0, w_1} MSE(w_0, w_1)$$

Now what?

We've found the optimal slope and intercept for linear hypothesis functions using squared loss (i.e. for the regression line). Now, we'll:

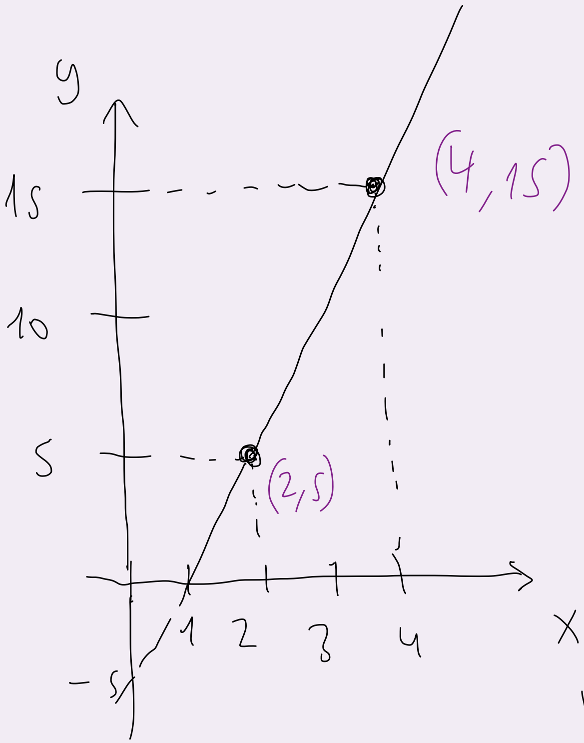
- See how the formulas we just derived connect to the formulas for the slope and intercept of the regression line we saw in DSC 10.
  - They're the same, but we need to do a bit of work to prove that.
- Learn how to interpret the slope of the regression line.
- Understand connections to other related models.
- Learn how to build regression models with **multiple inputs**.
  - To do this, we'll need linear algebra!

Question 🤔

Answer at q.dsc40a.com

Consider a dataset with just two points, $(2, 5)$ and $(4, 15)$. Suppose we want to fit a linear hypothesis function to this dataset using squared loss. What are the values of w_0^* and w_1^* that minimize empirical risk?

- A. $w_0^* = 2$, $w_1^* = 5$
- ~~B.~~ $w_0^* = 3$, $\underline{w_1^* = 10}$
- C. $w_0^* = -2$, $w_1^* = 5$
- D. $w_0^* = -5$, $w_1^* = 5$



w_1^*

$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{15-5}{4-2} = \frac{10}{2} = 5$

$\text{intercept} = \bar{y} - w_1^* \bar{x} =$

$\left(\bar{y} = \frac{20}{2} = 10, \quad \bar{x} = \frac{6}{2} = 3\right)$

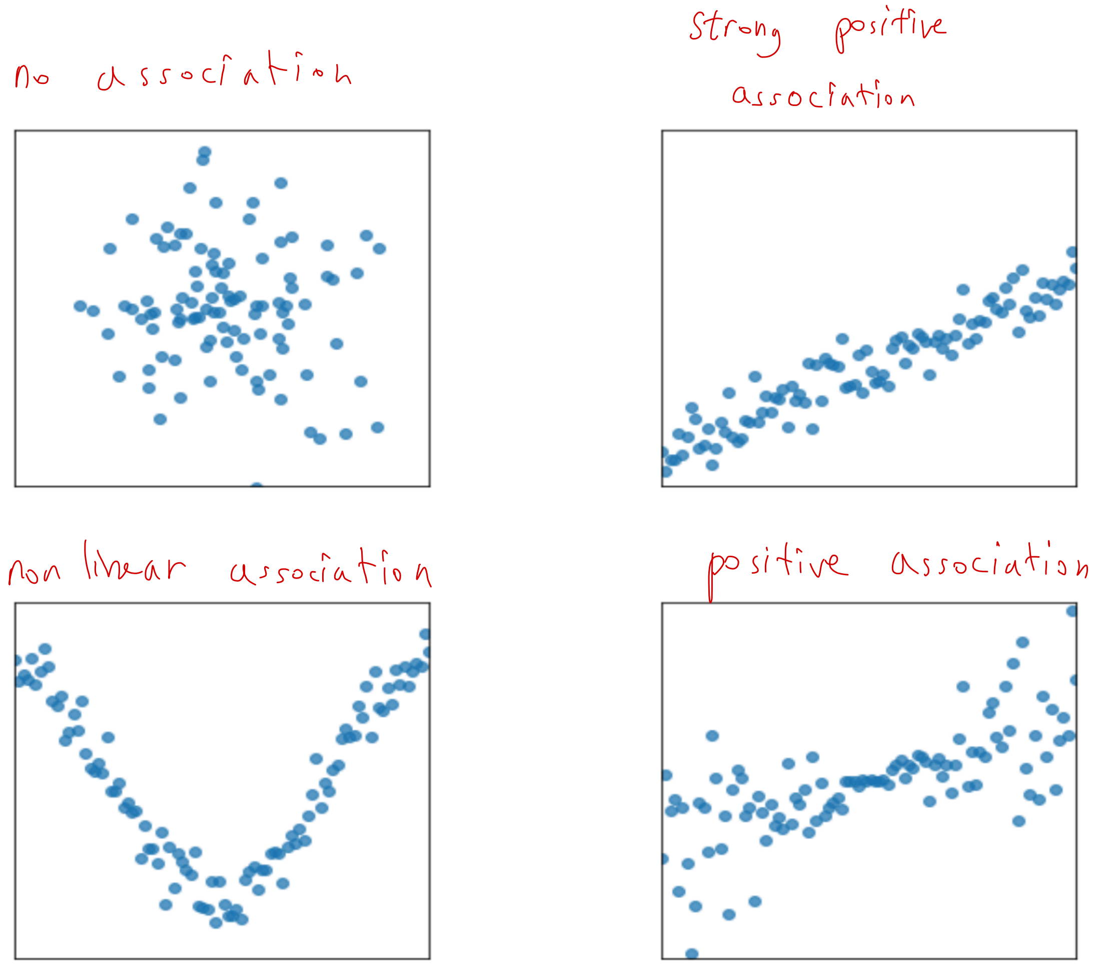
$w_0^* = 10 - 5 \cdot 3 = 10 - 15 = -5$

Correlation

Correlation = linear association

Quantifying patterns in scatter plots

- In DSC 10, you were introduced to the idea of the **correlation coefficient**, r.
- It is a measure of the strength of the **linear association** of two variables, x and y.
- Intuitively, it measures how tightly clustered a scatter plot is around a straight line.
- It ranges between -1 and 1.



r positive $\nearrow$ increase in $x \to$ increase in y

r negative $\nwarrow$ - " - $\to$ decrease in y

Pearson's

The correlation coefficient

- The correlation coefficient, r, is defined as the **average of the product of x and y, when both are in standard units**.
- Let σ_x be the standard deviation of the x_is, and $\bar{x}$ be the mean of the x_is.
- x_i in standard units is $\frac{x_i - \bar{x}}{\sigma_x}$.

$$\frac{x_i - \text{mean}(x)}{\text{std}(x)}$$

- The correlation coefficient, then, is:

$$r = \underbrace{\frac{1}{n} \sum_{i=1}^{n}}_{\text{averaging}} \underbrace{\left(\frac{x_i - \bar{x}}{\sigma_x} \right)}_{x \text{ in standard units}} \underbrace{\left(\frac{y_i - \bar{y}}{\sigma_y} \right)}_{y \text{ in standard units}}$$

alternative

covariance

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

plug in $y_i = x$ to get variance

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

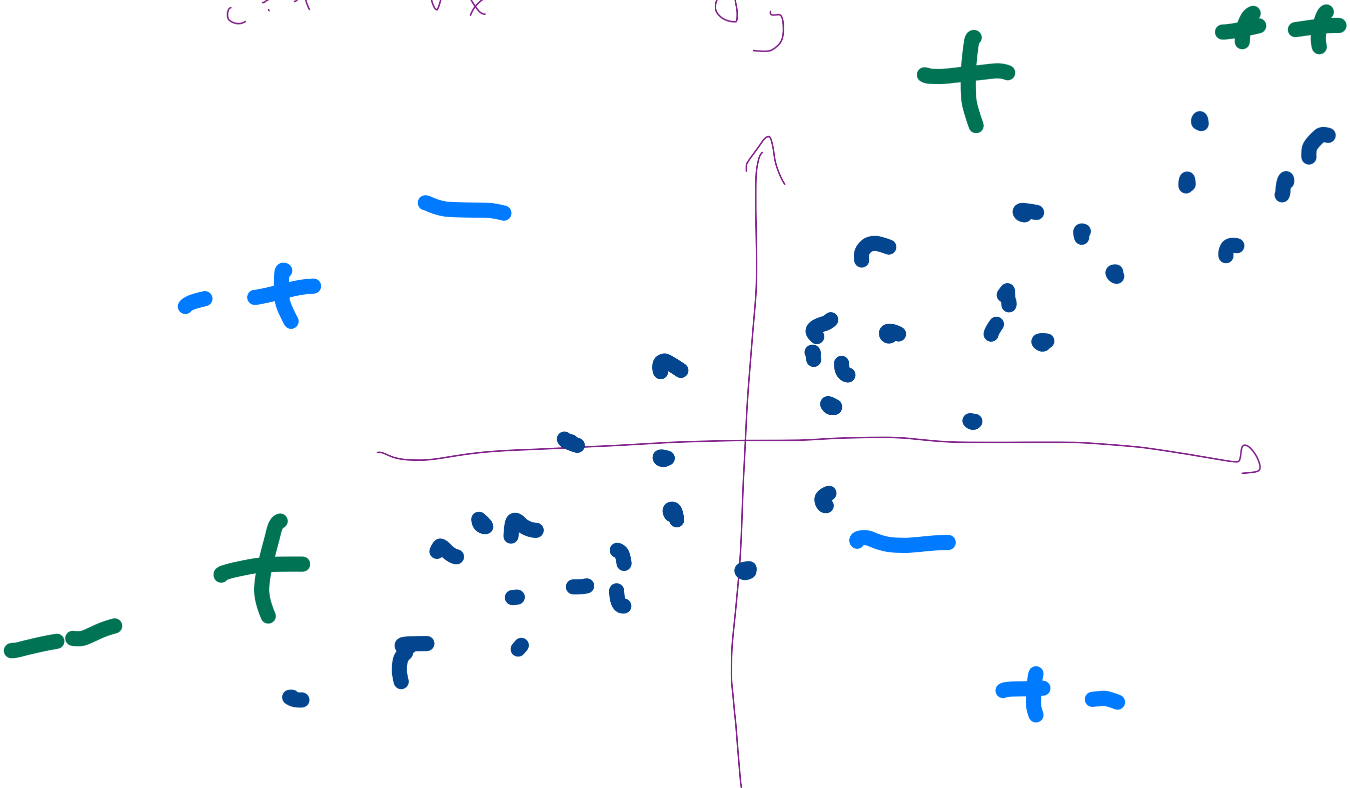
variance = covariance between x and itself

$$\sigma_x^2 = \sigma_{xx} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

std

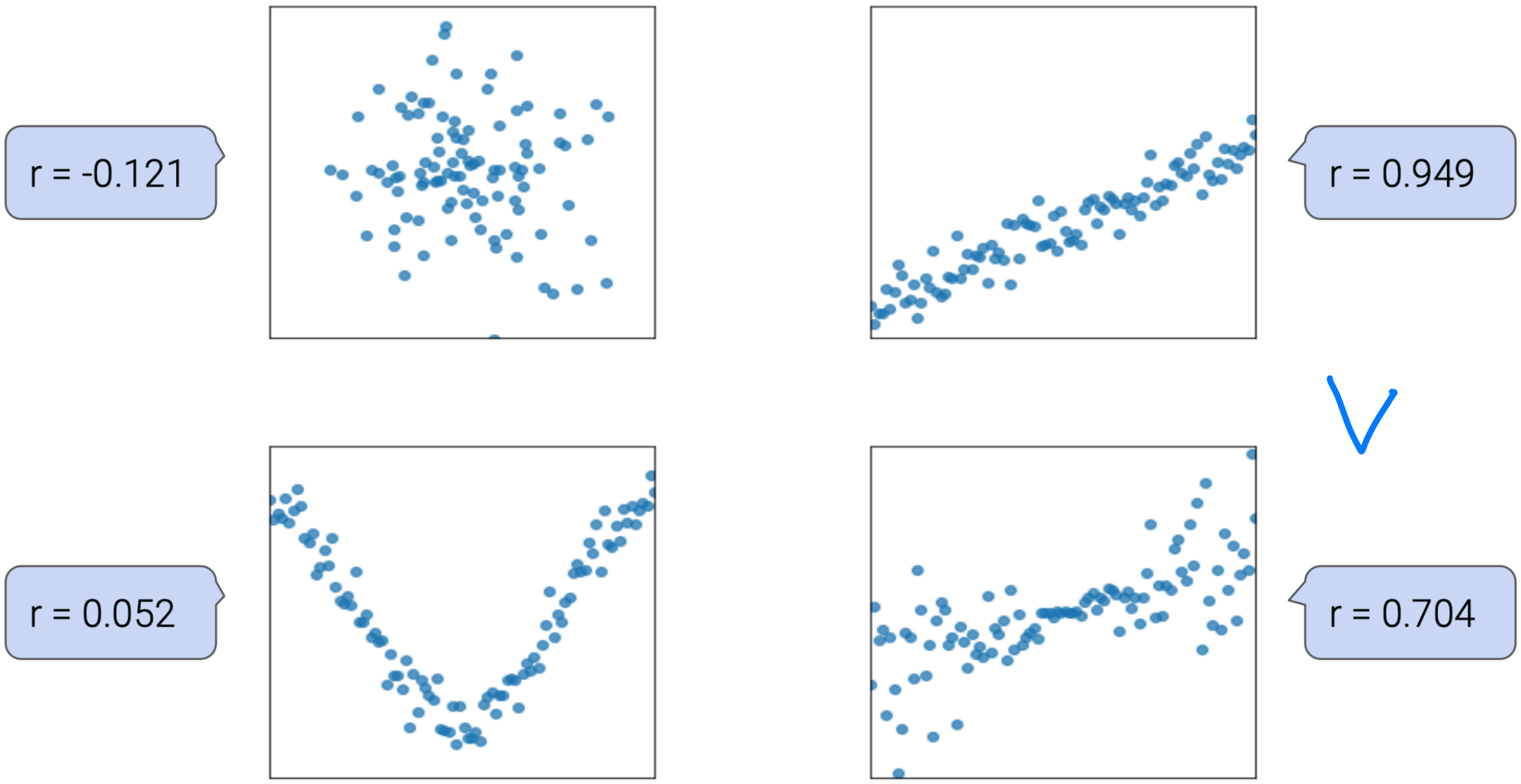
$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$r = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y}$$



$$r = \sum (+) + (+) + (-) + (-) > 0$$

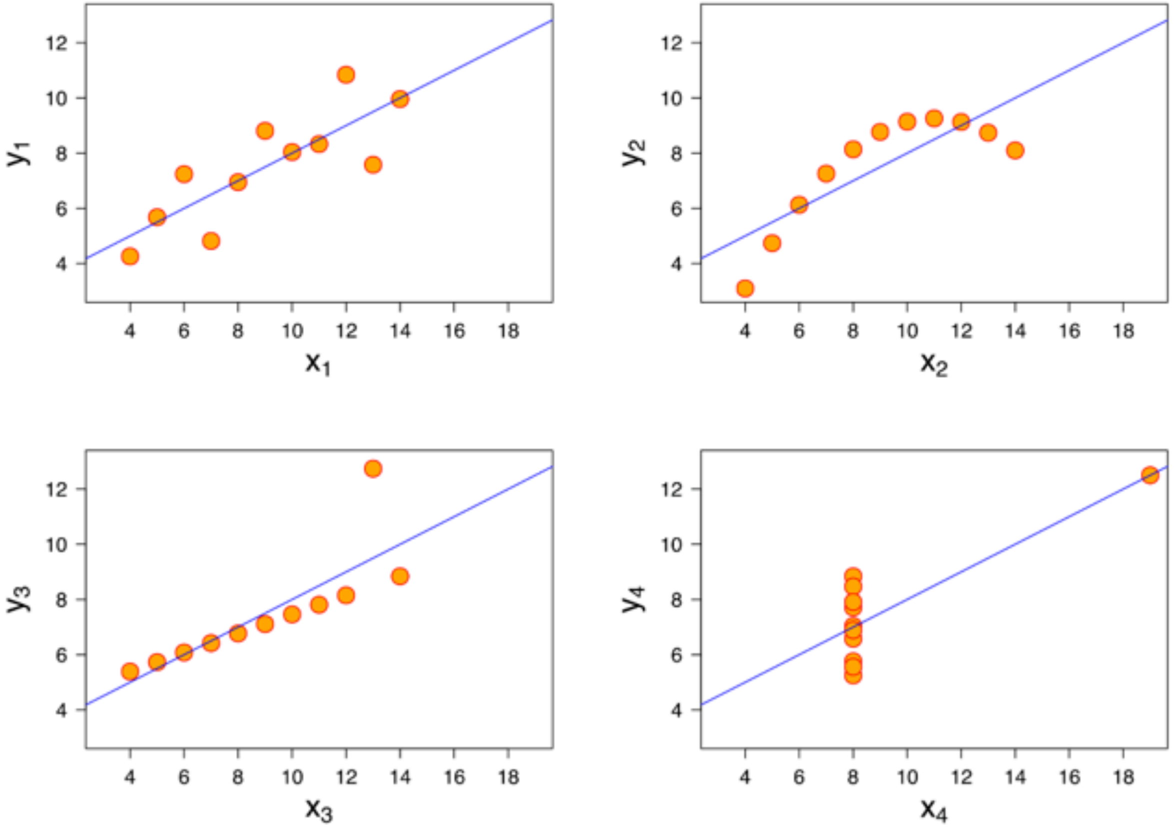
The correlation coefficient, visualized



Dangers of correlation

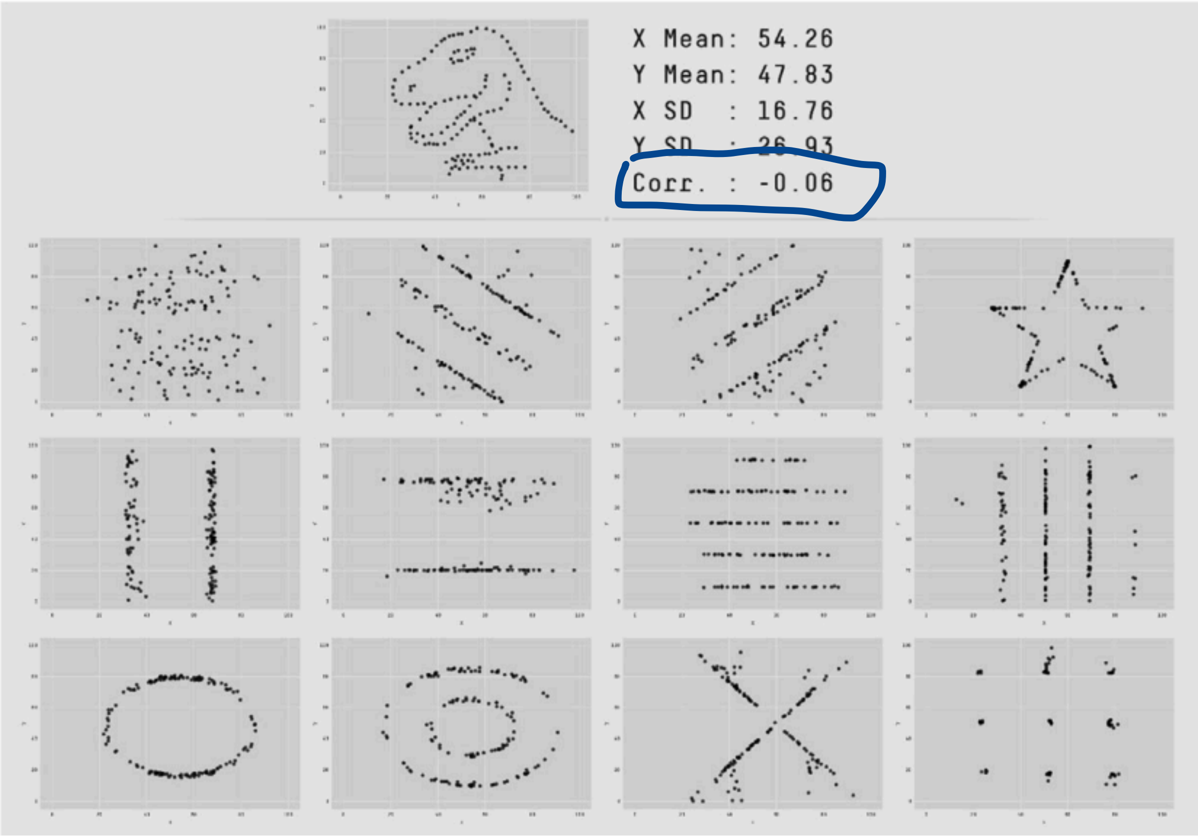
same mean, std, correlation

Anscombe's quartet (1973)



Dangers of correlation

Datasaurus dozen (2017)



Interpreting the formulas

Interpreting the slope

no units

units of y

$$w_1^* = r \frac{\sigma_y}{\sigma_x}$$

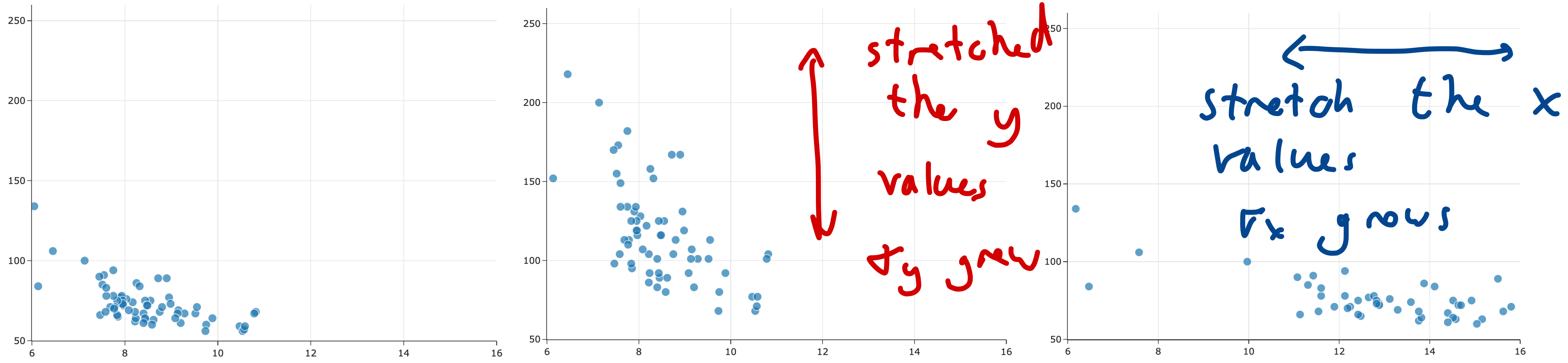
units of x

- The units of the slope are **units of y per units of x**.
- In our commute times example, in $H(x) = 142.25 - \underline{8.19}x$, our predicted commute time **decreases by 8.19 minutes per hour**.

Interpreting the slope

r is the same in all plots

$$w_1^* = r\frac{\sigma_y}{\sigma_x}$$



- Since $\sigma_x \geq 0$ and $\sigma_y \geq 0$, the slope's sign is r's sign.
- As the y values get more spread out, σ_y increases, so the slope gets steeper.
- As the x values get more spread out, σ_x increases, so the slope gets shallower.

Question 🤔

Answer at q.dsc40a.com

We fit a regression line to predict commute times given departure hour. Then, we add 75 minutes to all commute times in our dataset. What happens to the resulting regression line?

- A. Slope increases, intercept increases.
- B. Slope decreases, intercept increases.
- C. Slope stays the same, intercept increases.
- D. Slope stays the same, intercept stays the same.