

ES 2: Critical thinking with Python

aka, Introduction to Computing for Engineers

Steven Bell

8 April 2025



Logistics / due dates

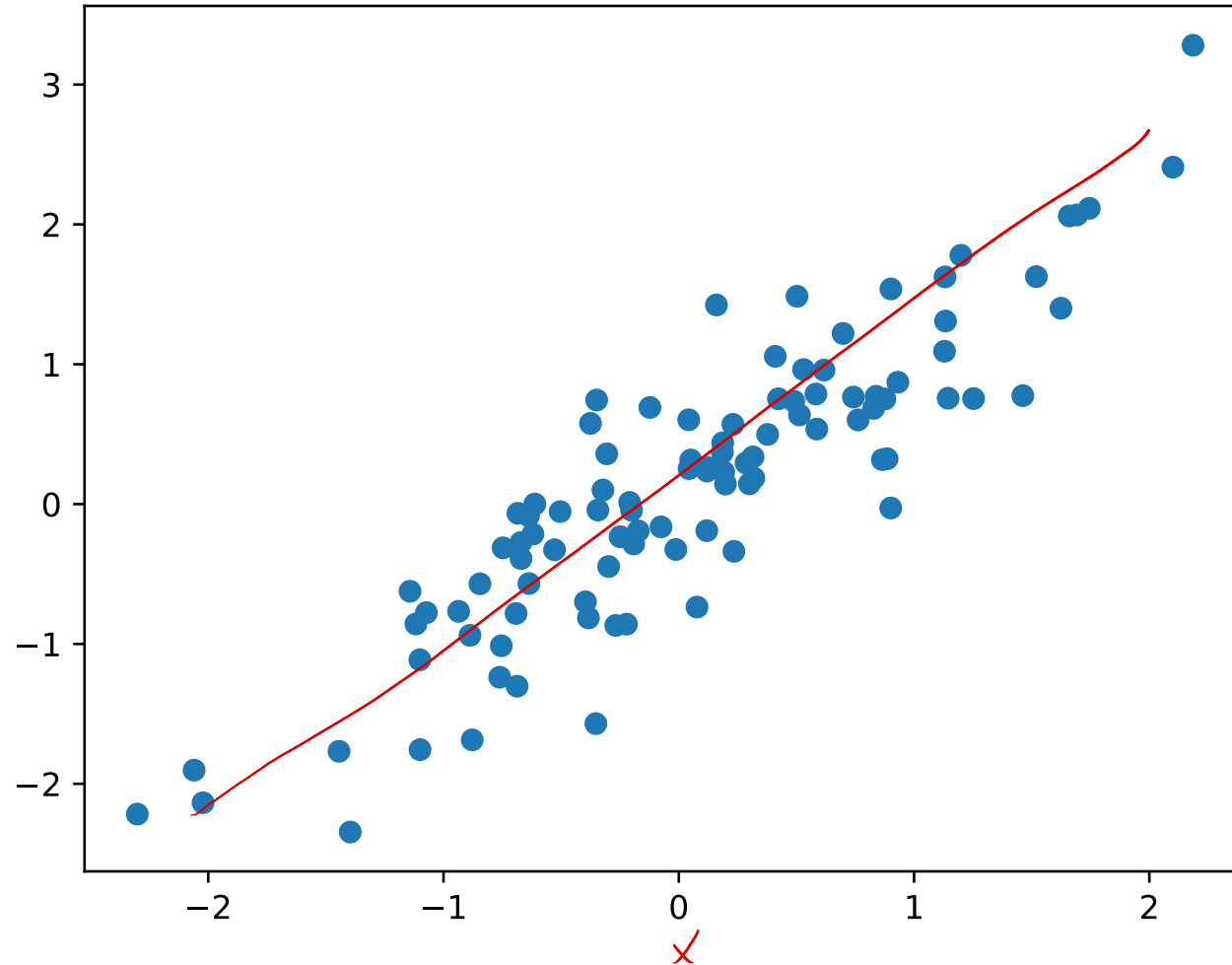
Roomba project due tonight (4/8)

Sine wave / ECG parts due next Monday (4/14)

Correlation

A summary statistic relating two data sets

Some nicely correlated data



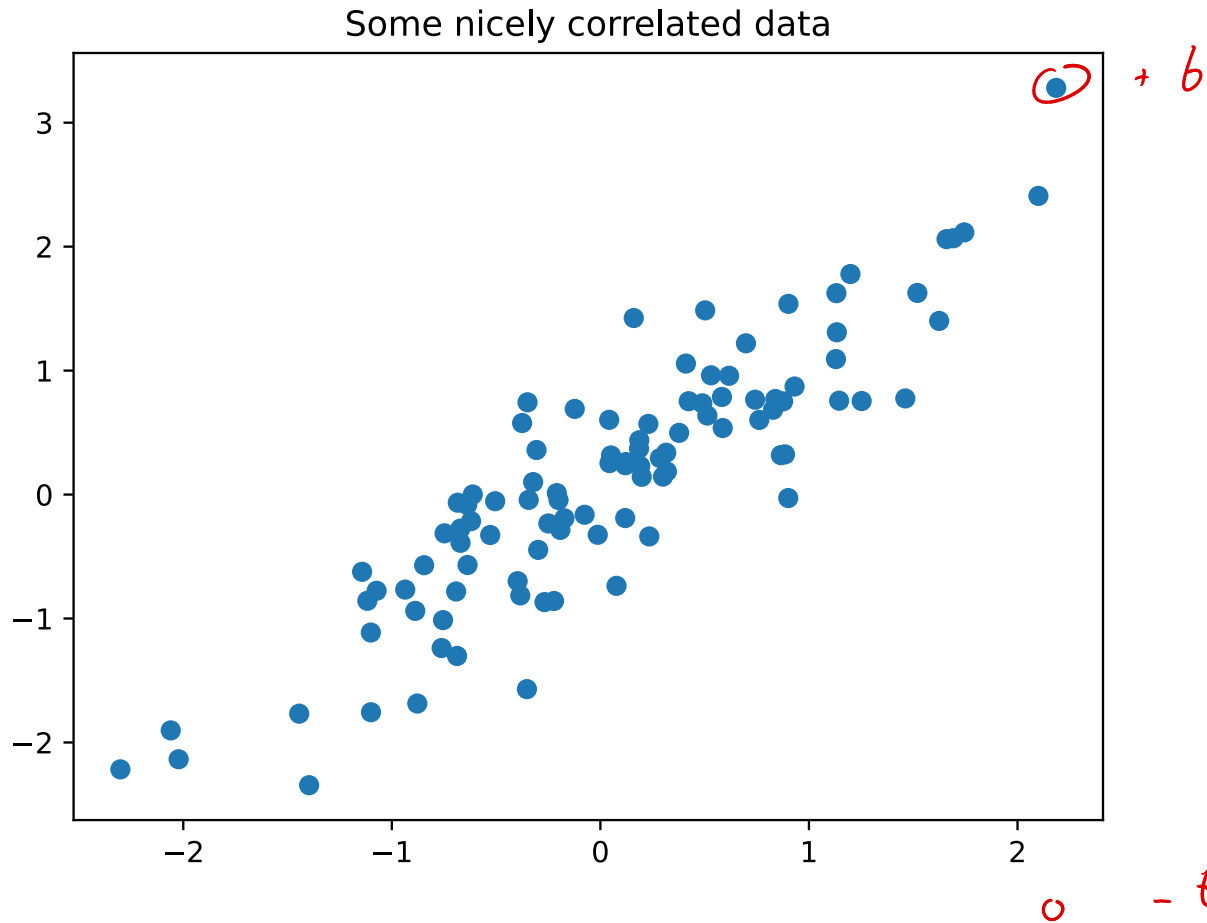
Y

$$Y = X$$

X

Correlation

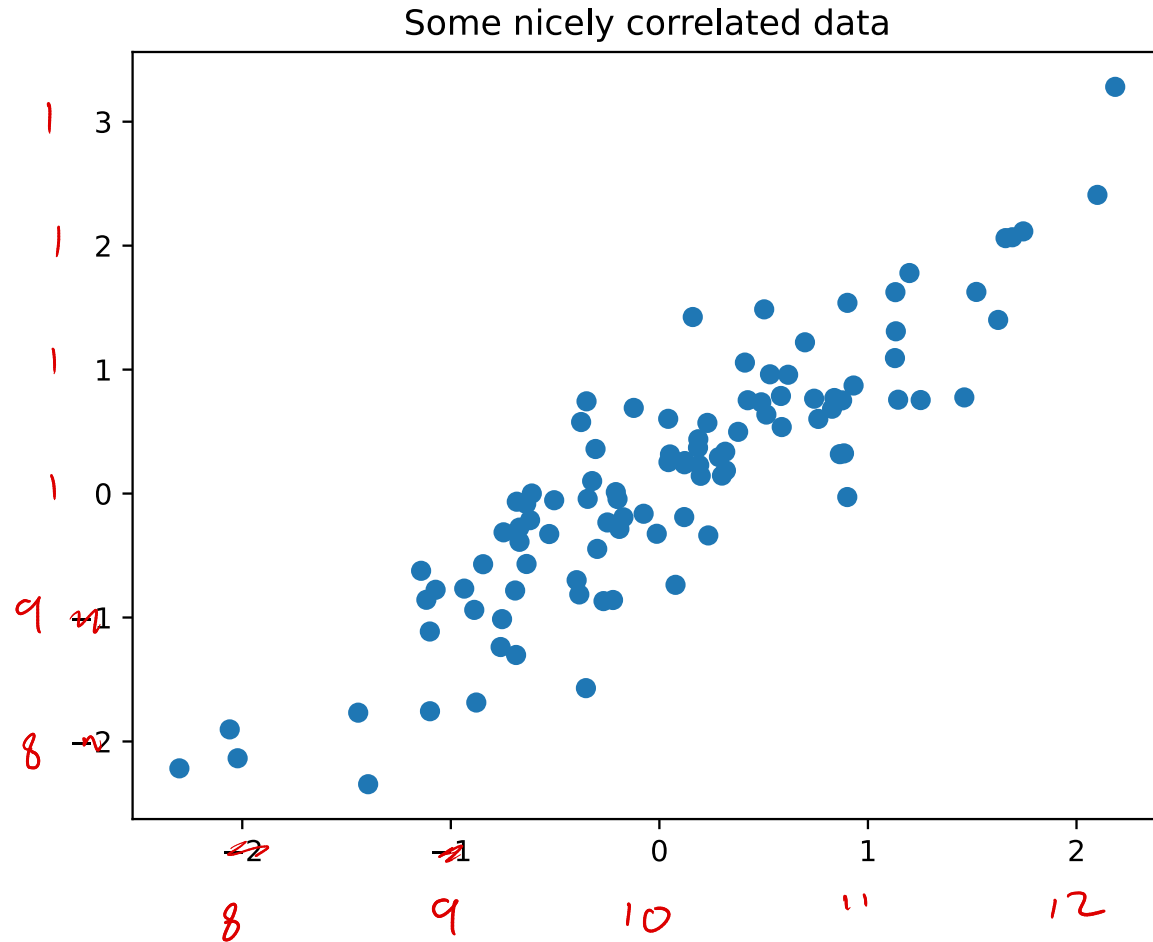
Key idea: multiply X and Y values, and take the mean



$$E[X Y]$$

Correlation

But we also need to account for shifting



$$E[(X - \mu_X)(Y - \mu_Y)]$$

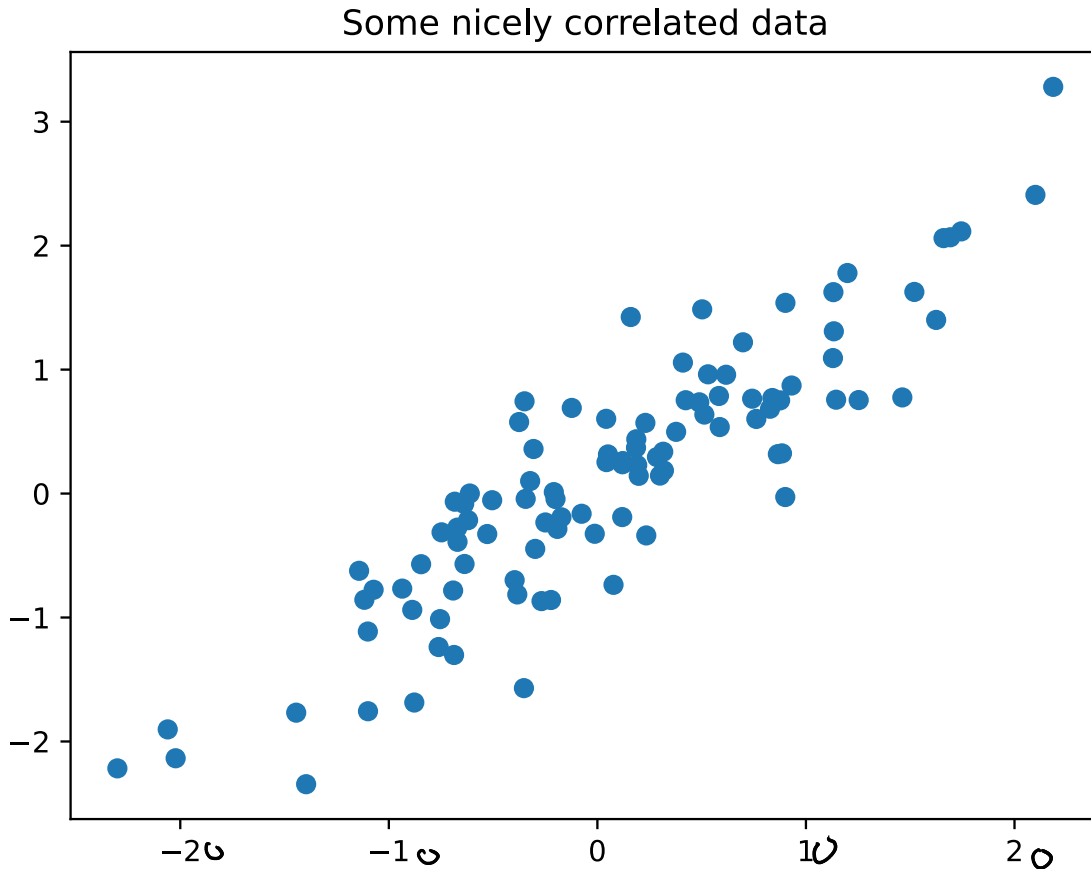
subtract mean

Correlation

And also for scaling

$$E[(X - \mu_X)(Y - \mu_Y)]$$

divide by standard deviation



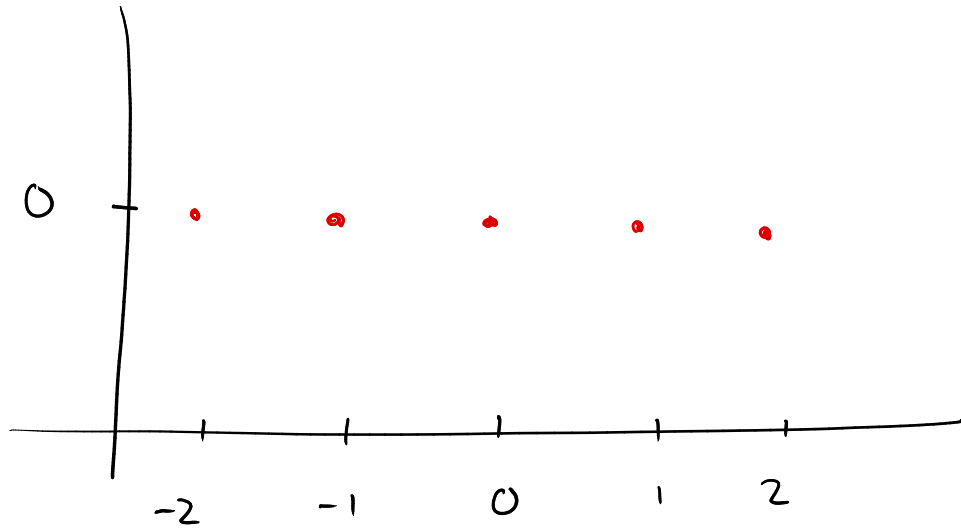
$$\frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Guess that correlation!

PollEV.com/stevenbell

$$\mathbb{E}[X Y] = 0$$

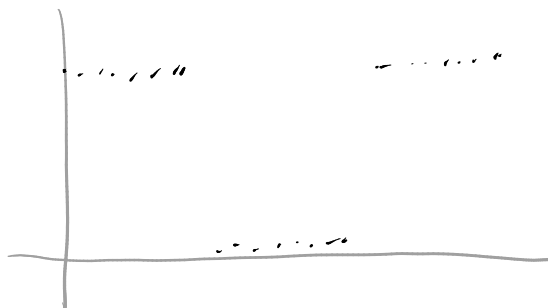
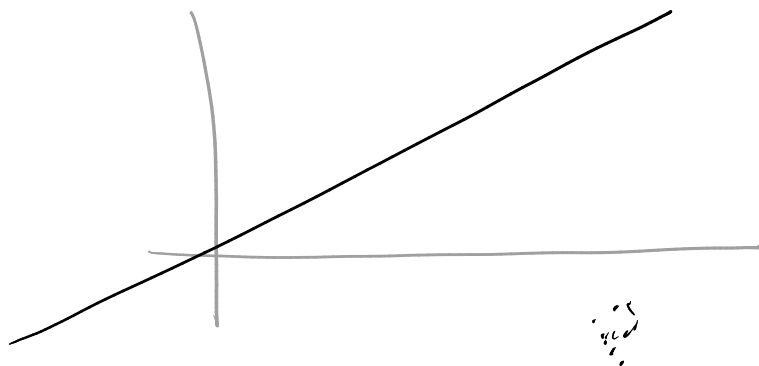
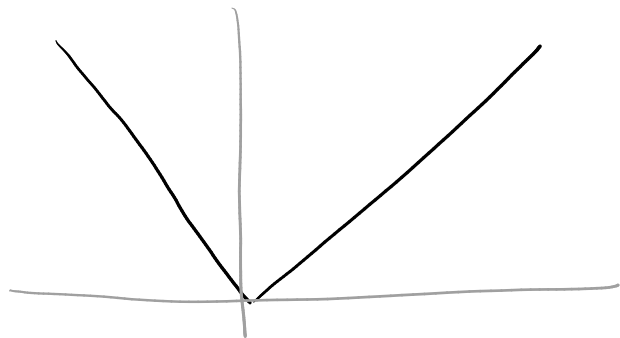
$$\rho_{x,y} = 0$$



Correlation

Is it possible for two datasets to be obviously related, but to have a correlation of zero?

Try making a scatter plot either with Python or on <http://robertgrantstats.co.uk/drawmydata.html>



Correlation

Is it possible for two data sets to be correlated, but have no meaningful connection?

Correlation

Is it possible for two data sets to be correlated, but have no meaningful connection?

Yes, of course! Correlation does not imply causation!

<https://tylervigen.com/spurious-correlations>

Correlations in the news

Google News search for "linked to"

Reading for Thursday:

[https://theconversation.com/
music-to-die-for-how-genre-affects-popular-musicians-life-expectancy-36660](https://theconversation.com/music-to-die-for-how-genre-affects-popular-musicians-life-expectancy-36660)