

CSD 598 - Winter 2026

Descriptive statistics and data visualisation

Pertti Palo

3 Feb 2026

Last time: Ethics

Any questions? Any thoughts?

Why do data exploration?

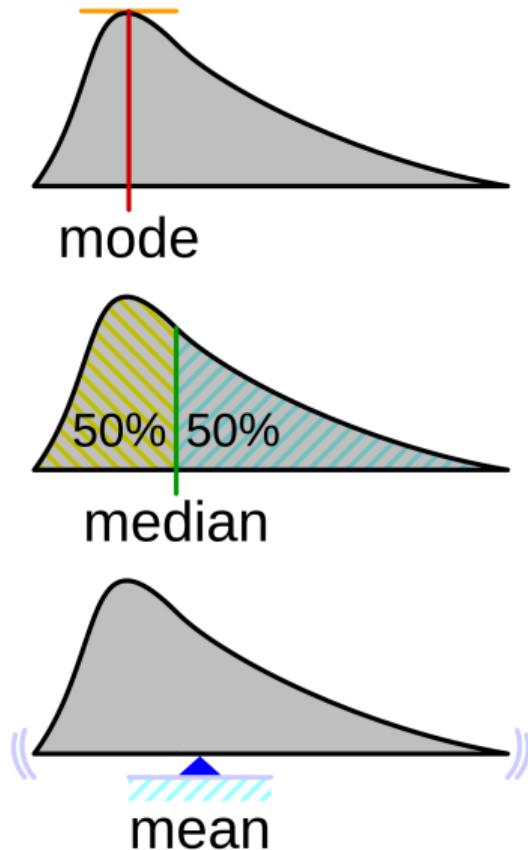


"Remember kids, the only difference between screwing around and science is writing it down."

Adam Savage, quoting ballistics expert Alex Jason

Descriptive stats

Central tendencies



Cmglee, CC BY-SA 3.0, via Wikimedia Commons

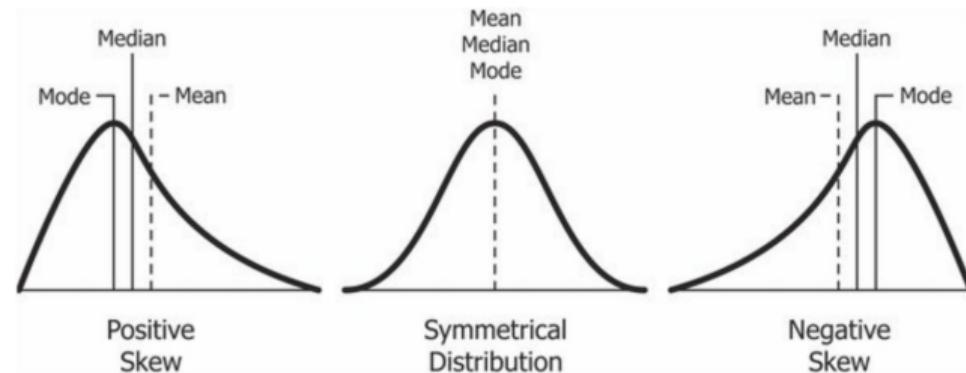
Dispersion and shape

Dispersion

- ▶ Range, quartiles: easy to calculate on samples
- ▶ Standard deviation (SD), variance: most used
- ▶ Average absolute deviation: more robust than SD

Shape measures

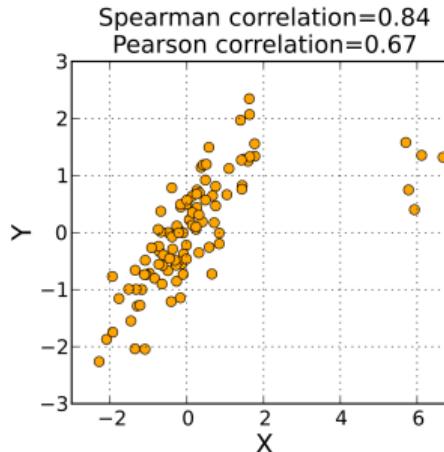
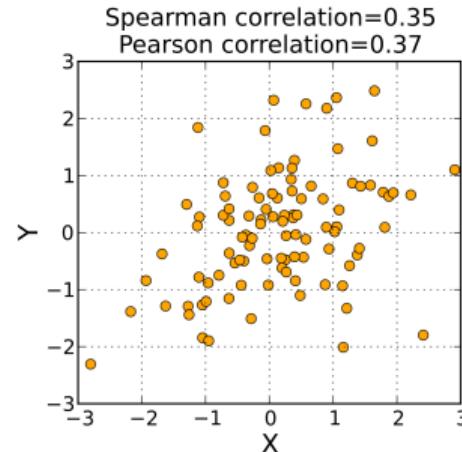
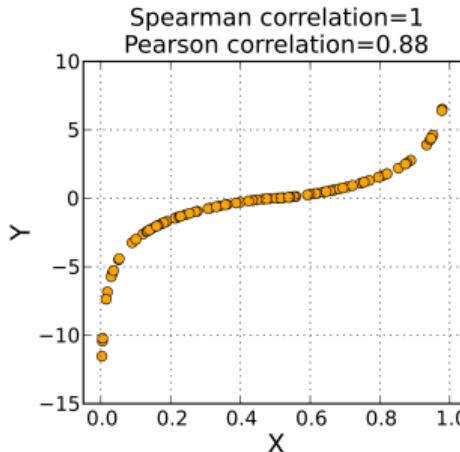
- ▶ Skewness: which way does the distribution lean?
- ▶ Kurtosis: tailedness, things get complex here



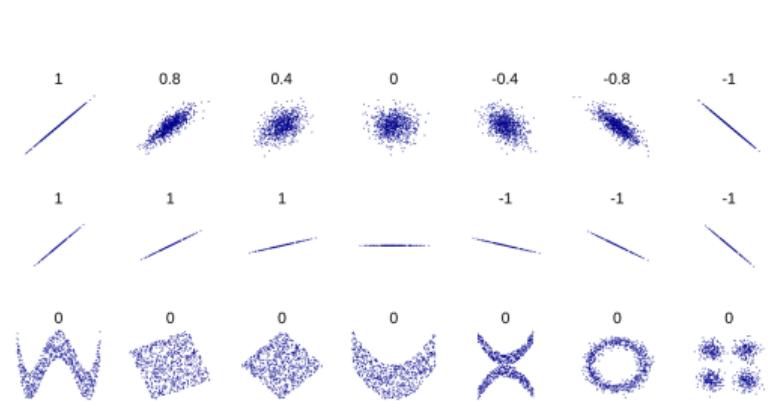
Diva Jain, CC BY-SA 4.0 international, via Wikimedia Commons

Dependence between variables: The usual

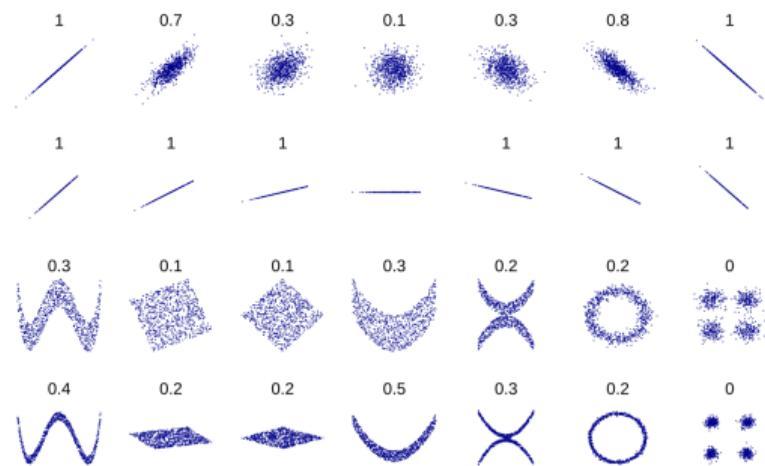
- ▶ Covariance: measure of the linear relationship between two variables.
- ▶ Pearson's ρ : correlation of continuous variables
- ▶ Spearman's ρ : correlation of ordering of variables



Dependence between variables: Pearson vs Distance correlation



Pearson correlation



Distance correlation

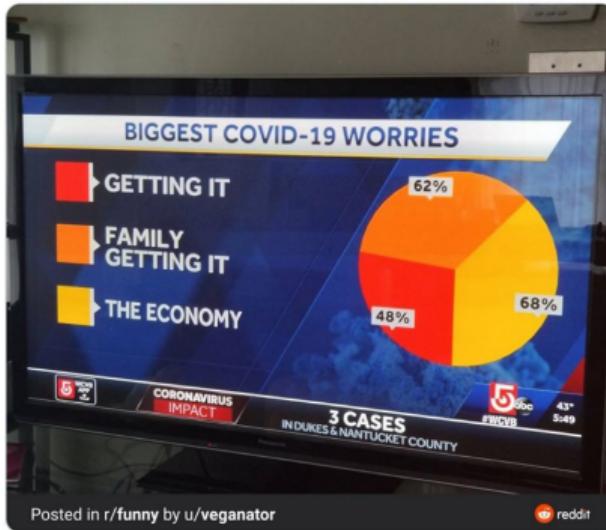
Denis Boigelot, CC BY-SA 3.0, via Wikimedia Commons

Visualisation



Dr. Lindsay Portnoy
@lpportnoy

This is not how pie charts work



9:48 PM · Mar 26, 2020 · Twitter for iPhone

8.1K Retweets 49.8K Likes

Choosing a good graph

Tell-show-do-apply

References