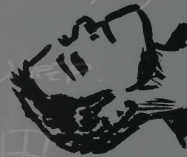# SGD Topics #1: Multivariate Visualisation
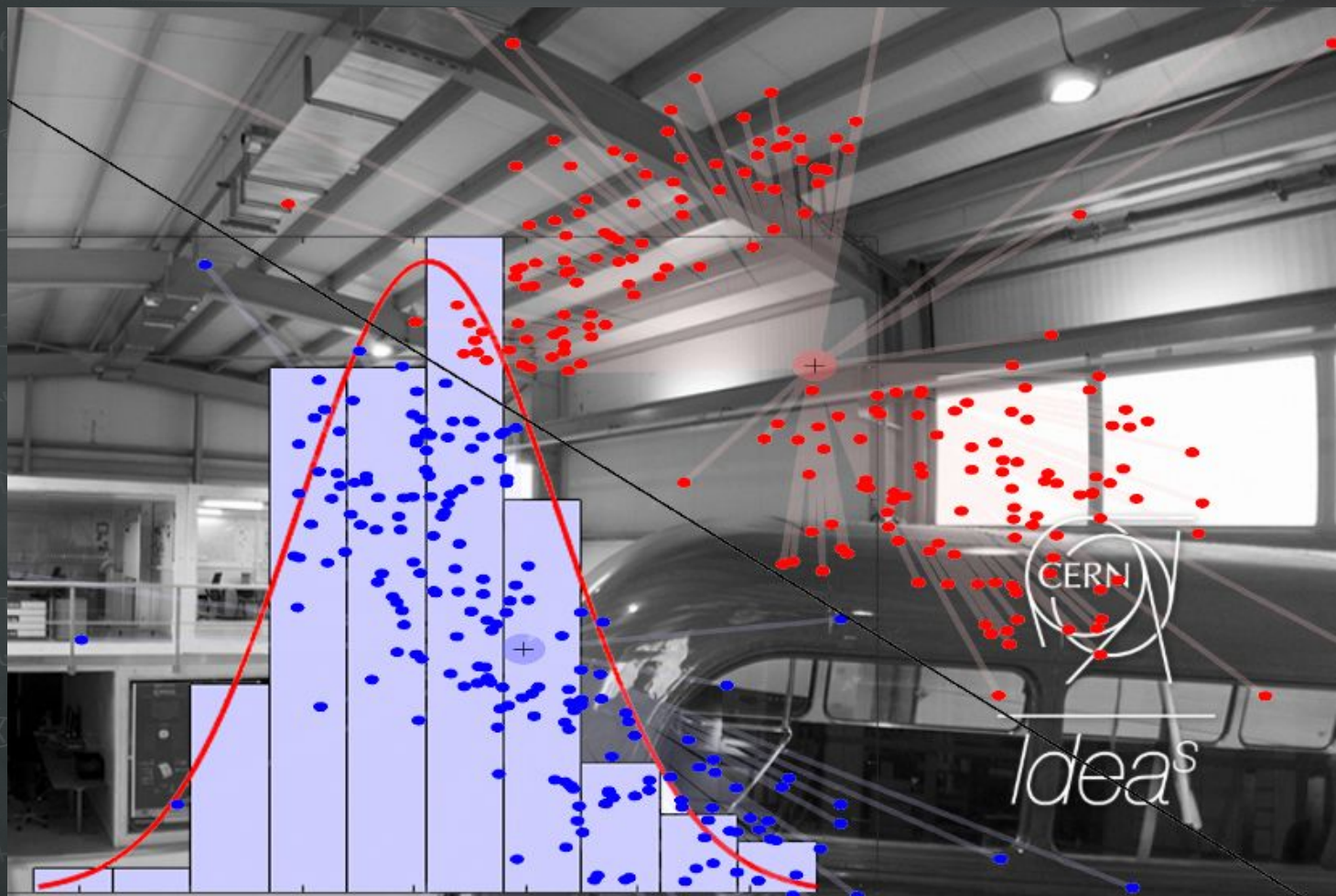
@vincecr0ft

@DrYacineHaddad

Vince Croft

Welcome!

- Hi! thanks to IdeaSquare!

Changes on the Way!

- SGD is on the way
- Two new meetups on the horizon
- Website under production

Visualising Data!

- The most important step in dealing with data.
- Continuous, categorical, binned/bucketed, ordered, etc
- What and why?
- Exploring data - EDA

THE CAVE

THE FIRE

THE ROADWAY

DIFFUSED DAYLIGHT

THE ROUGH ASCENT TO SUNLIGHT
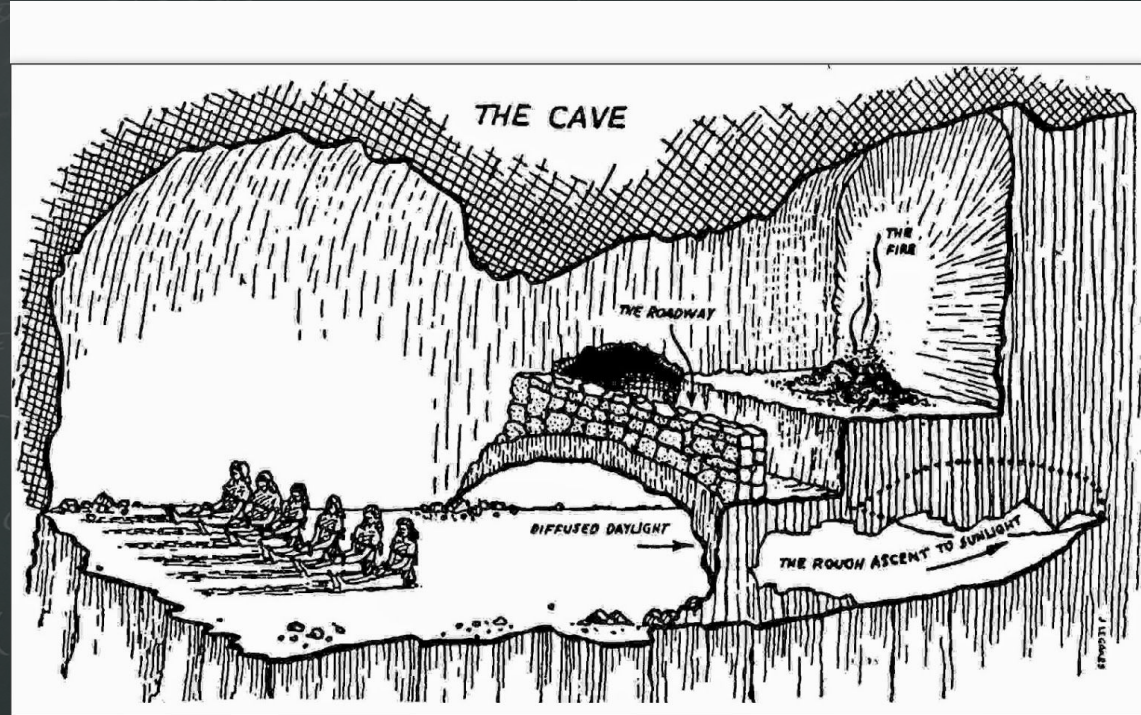
1D - Histograms and Summary data

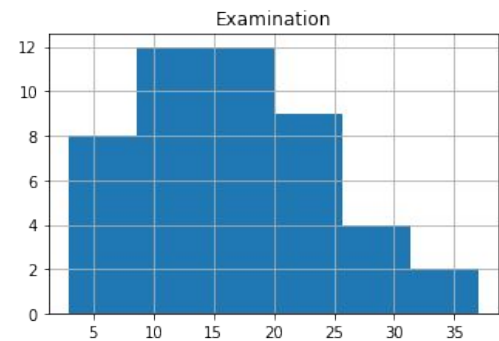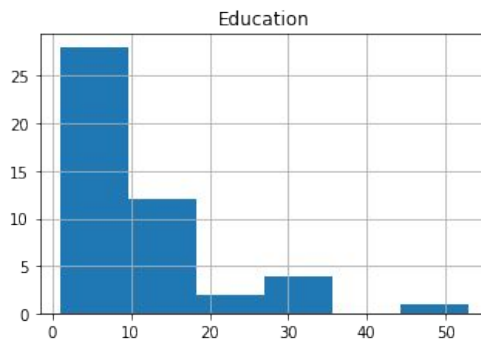- Data should be in reference to more data.
- Histograms let us visualise frequency and scale of observations
- Mean, standard deviation, etc let us understand a data set without seeing it all
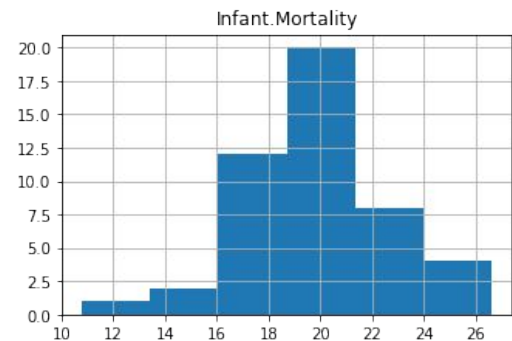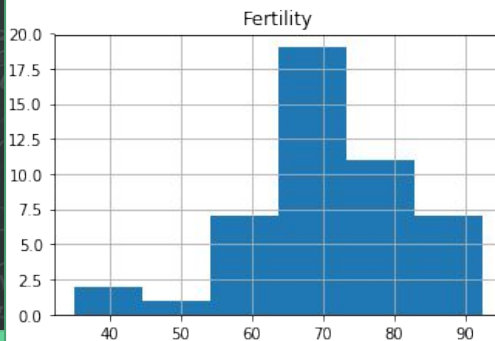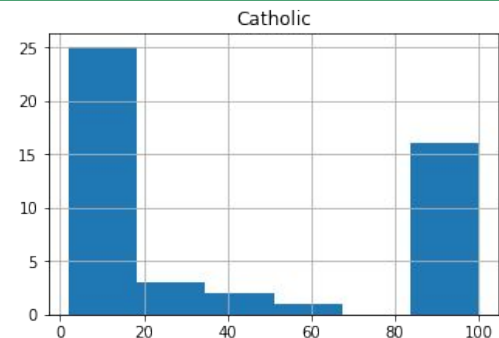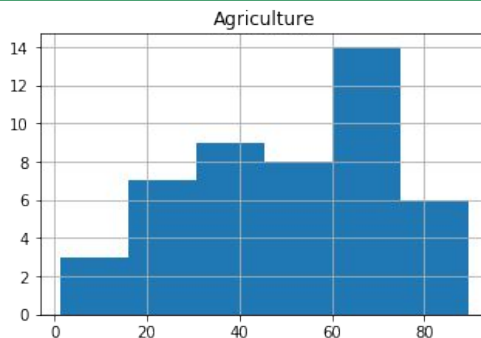
Correlation

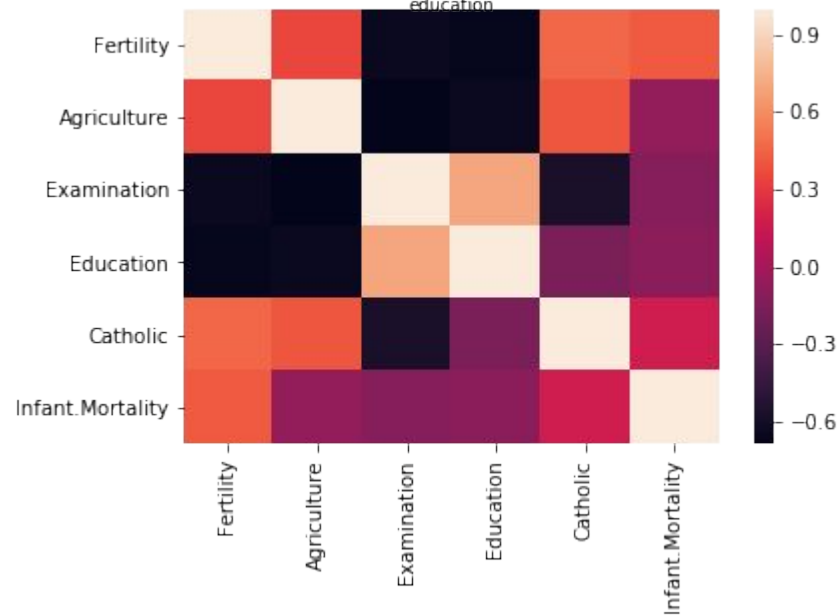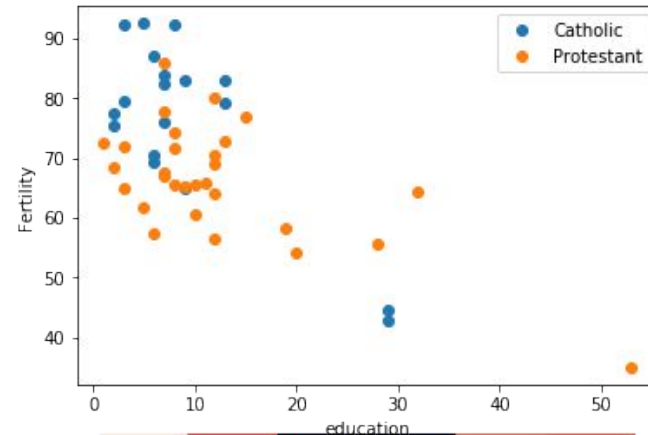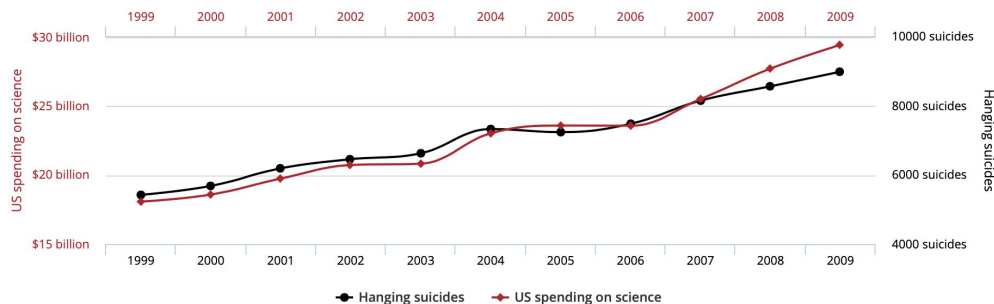- Picturing the relationship between two variables can give the most insight into how you data behaves
- Correlation does not imply causation

US spending on science, space, and technology
correlates with
Suicides by hanging, strangulation and suffocation
Correlation: 99.79% (r=0.99789126)

Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

A Picture Tells a Thousand Words

- Picturing the data is sometimes the only way to understand it
- The anscombe data sets have identical numbers of points lines of best fit standard deviations and correlations

Try it yourself!

https://github.com/SocieteGenevoiseDonnees/MultivariateVisualisation
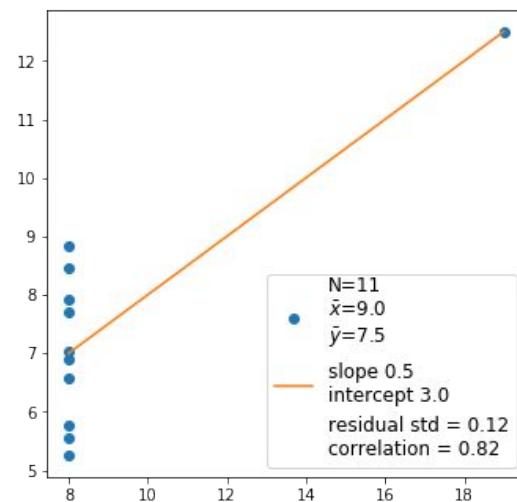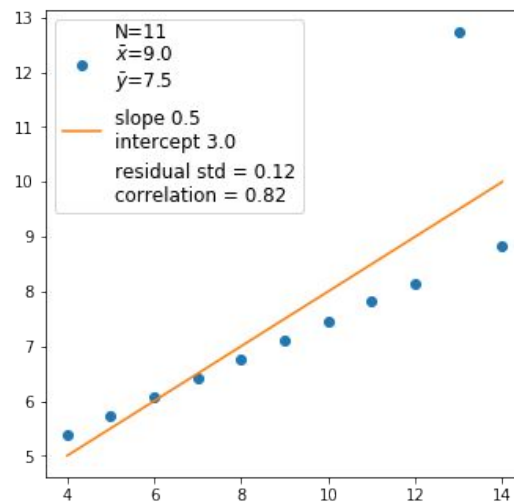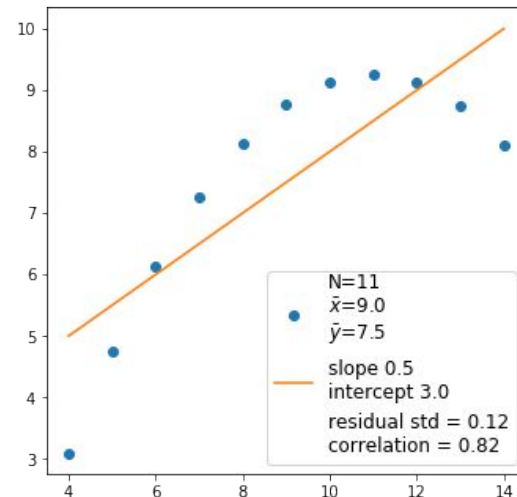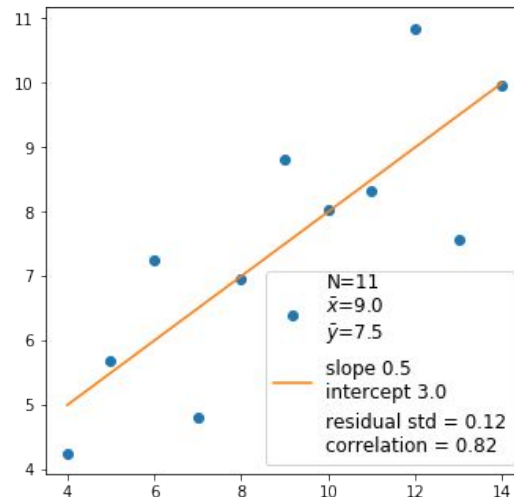
opendata.swiss ◯

Data   Organizations   Applications   About ▾   🔍

Find Swiss Open Government data

Learn more about opendata.swiss

6,385
Datasets

Search datasets...   🔍

Access the data catalogue using the API