Computation and Visualization for Analytics

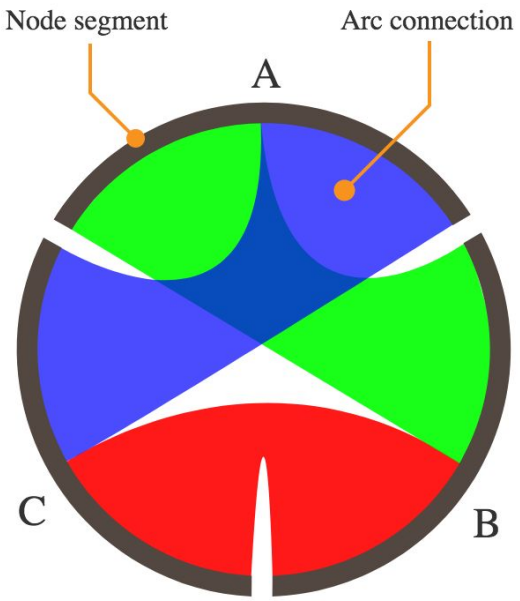
Spring 2021

Week 6.2

Visualizing

Relationships

Chord Diagram



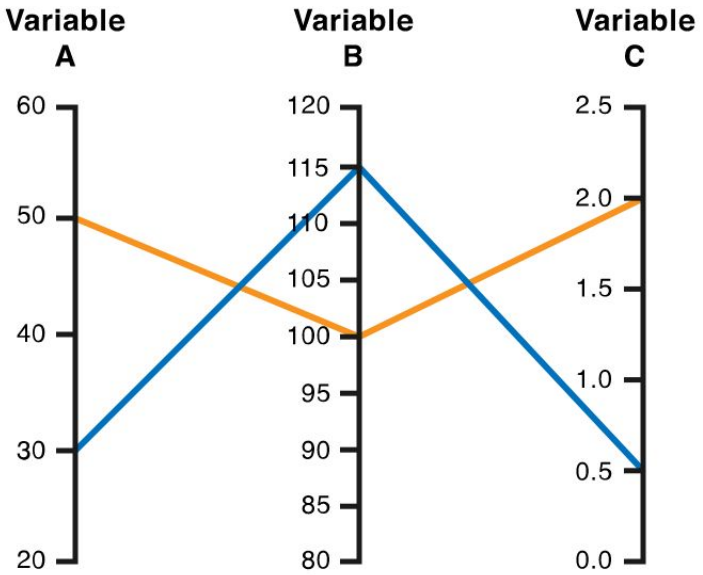
	A	B	C
A		10	10
B	10		10
C	10	10	

A $\xrightarrow{10}$ B

B $\xrightarrow{10}$ C

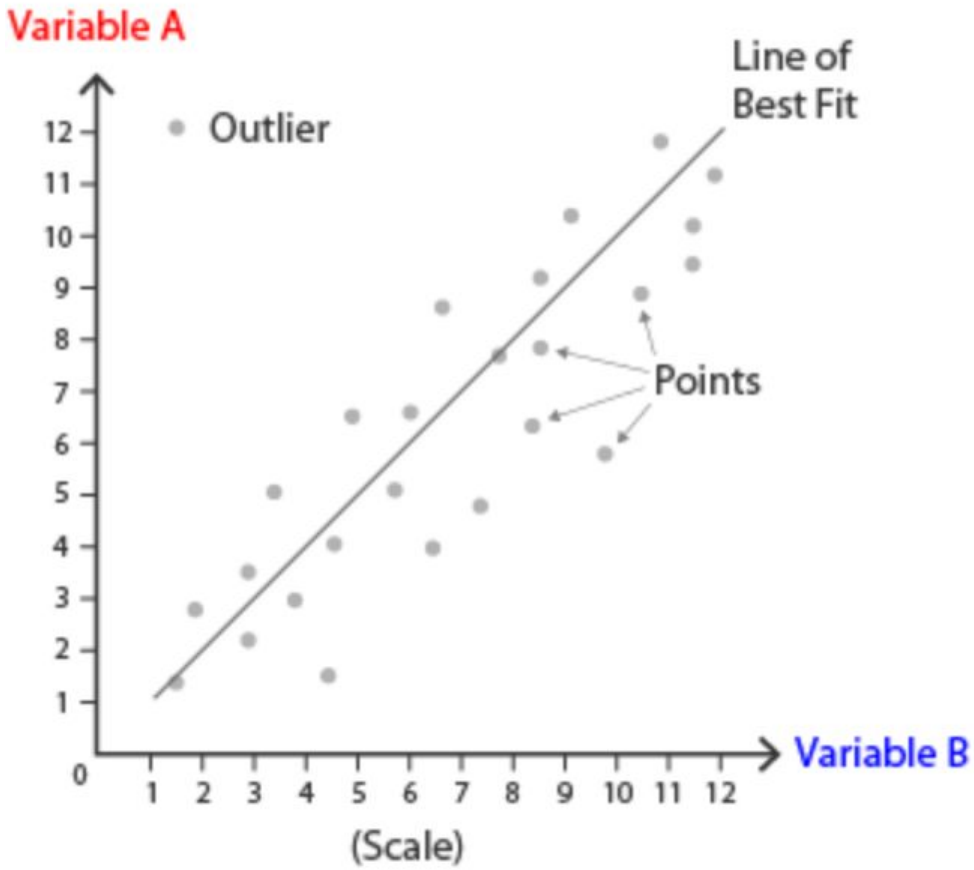
C $\xrightarrow{10}$ A

Parallel Coordinates Plot



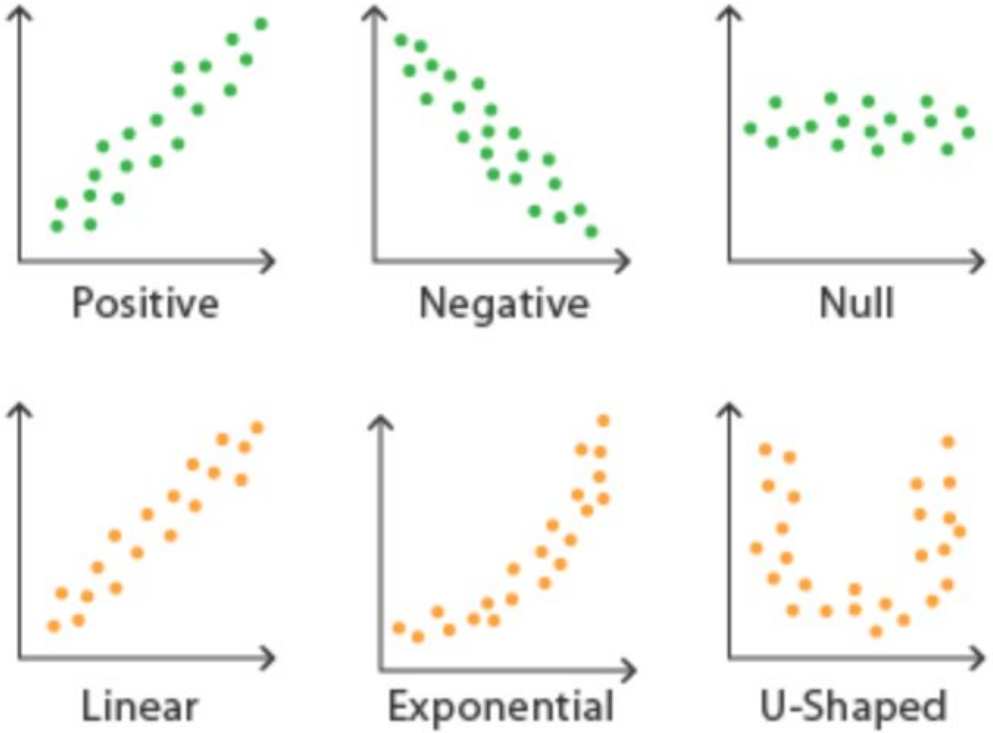
Data			
	Variable A	Variable B	Variable C
Item 1	50	100	2.0
Item 2	30	115	0.5

Scatter Plot



https://datavizcatalogue.com/

Types of Relationships



https://datavizcatalogue.com/

Association, dependence, correlation

- In everyday language, dependence, association and correlation are used interchangeably*
- Association is synonymous with dependence and is different from correlation*
- Association is a very general relationship: one variable provides information about another*
- Correlation is more specific: two variables are correlated when they display an increasing or decreasing trend*

Altman, N., & Krzywinski, M. (2015). Association, correlation and causation.

Association, correlation

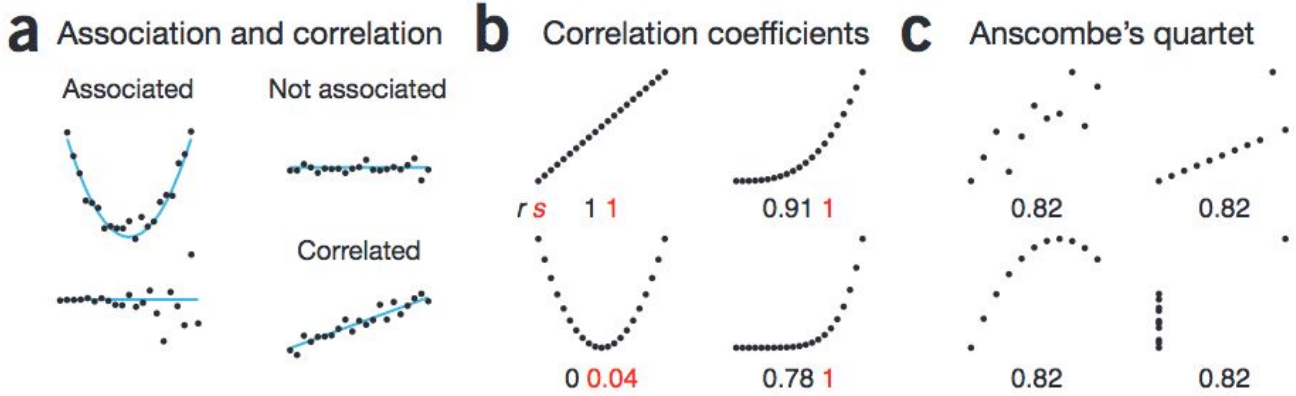


Figure 1 | Correlation is a type of association and measures increasing or decreasing trends quantified using correlation coefficients. (**a**) Scatter plots of associated (but not correlated), non-associated and correlated variables. In the lower association example, variance in *y* is increasing with *x*. (**b**) The Pearson correlation coefficient (*r*, black) measures linear trends, and the Spearman correlation coefficient (*s*, red) measures increasing or decreasing trends. (**c**) Very different data sets may have similar *r* values. Descriptors such as curvature or the presence of outliers can be more specific.

Altman, N., & Krzywinski, M. (2015). Association, correlation and causation.

Pearson Correlation Coefficient

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad \textbf{(Eq.3)}$$

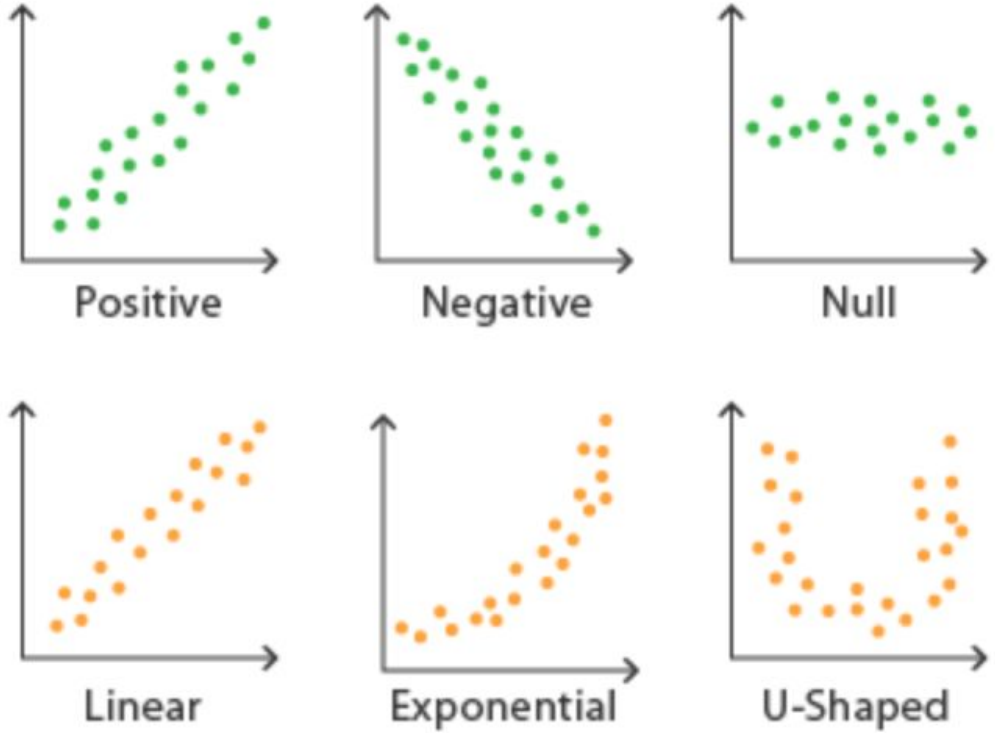
where:

n is sample size

x_i, y_i are the individual sample points indexed with *i*

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (the sample mean); and analogously for $\bar{y}$

Visual Correlation



https://datavizcatalogue.com/