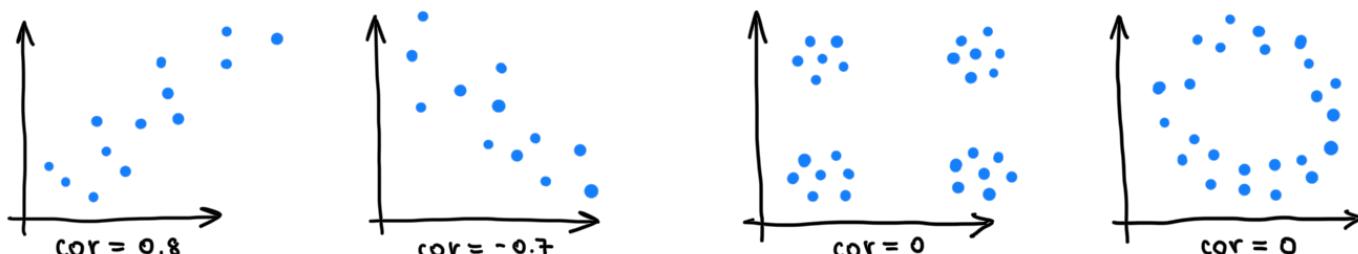


Multidimensional exploratory data analysis

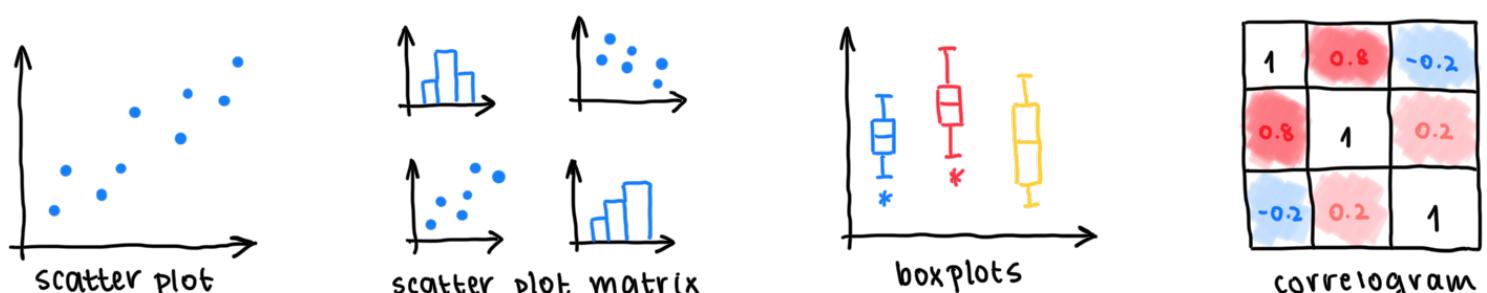
- exploration of several variables and their relationships simultaneously (whole table instead of individual columns)

Continuous variables

- Individuals measures of location (mean, median), variability (variance, standard deviation) or shape (kurtosis, skewness) are not often used since they say nothing about the relationship between variables
 - Measures of dependency
 - Sample covariance depends on scale, describes only linear relationships
 - Sample correlation has values between -1 and 1, too only describes linear relationships

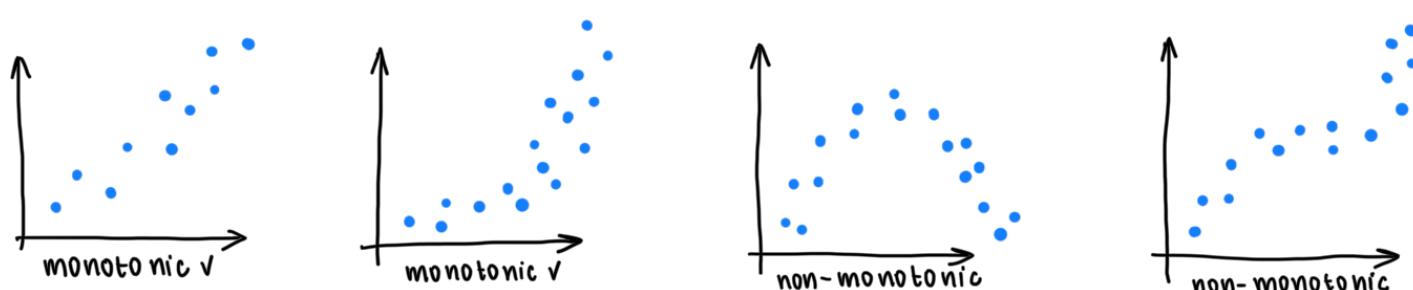


- Dimension reduction
 - PCA condenses existing variables into a smaller number of components while preserving maximal variability
 - Graphical methods
 - Histograms
 - Scatter plot (for 2 variables), scatter plot matrix
 - Boxplots
 - Correlogram



Categorical variables

- Measures of dependency (for ordinal)
 - Spearman's correlation coefficient - equal to the Pearson's correlation between the ranks of those two variables, describes monotonic relationships



- Kendal's tau - used as a test statistic to assess independence
 - Analysis of contingency tables
 - Pearson's chi squared test - tests for independence

- Fisher's exact test
 - Graphical methods
 - Grouped bar plot
 - Stacked bar plot

