Exploratory Data Analysis (EDA)

Telkom University

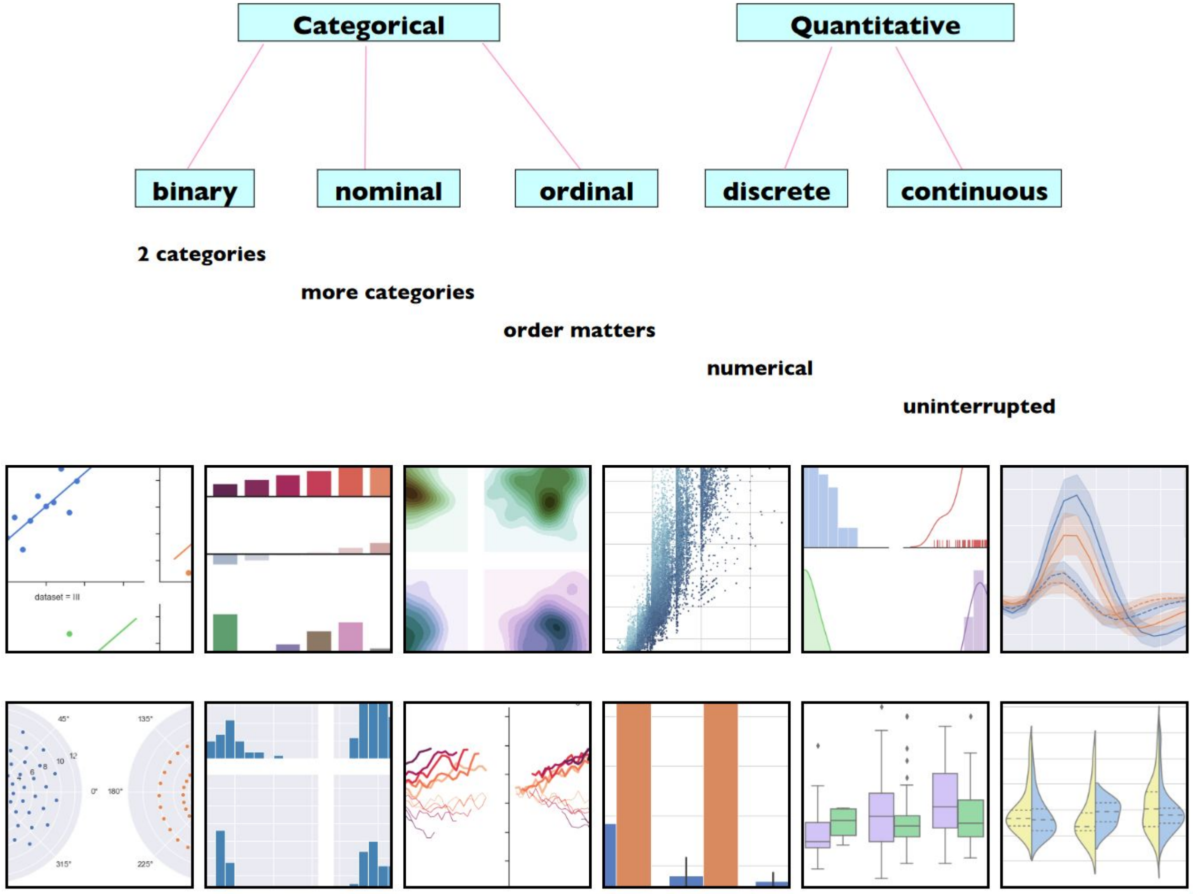
Dr. Risman Adnan Mattotorang

Outline – Exploratory Analysis

1. **Fundamentals of EA**
   - Objectives and Scopes of EA
2. **Statistical Summary of Data**
   - Numerical Data
3. **Statistical Hypothesis Testing**
   - Confirmatory Data Analysis
4. **Dimensionality Reduction**
   - LDA, PCA and SVD for EDA

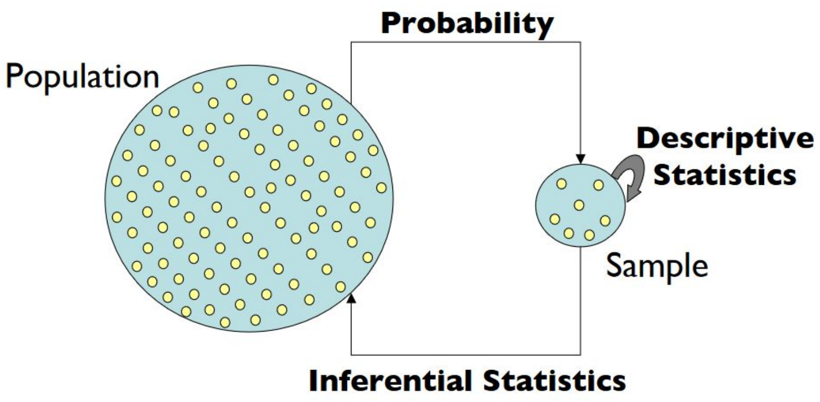
Recall: Key Principles of Visualization

1. **Review Types and Dimension of Data**
   - Univariate and Multivariate
2. **Show Comparisons**
3. **Show Causality/Systematic Structure**
4. **Visualize Multivariate Data**
5. **Integration of Evidences**
6. **Describe Evidences with Proper Scales**
7. **Data Understanding is the King!**



Scopes of EA

- **Not the same thing...**
- **Probability** teaches us how to predict future events.
- **Statistics** focuses on analysis of the frequency of past events.
- Statistics is how we learn from past experiences.
- **Exploratory Data Analysis (EA)** is concerned with “how to best learn from what data we have”.



Exploratory Analysis

- Before making ML/DL prediction and inference from data it is **essential to examine** all your variables.

- **Why? To listen to the data:**
  - To see **patterns & structures** in the data
  - To catch **mistakes and biases**
  - To find **violations** of statistical assumptions
  - To find **outliers/anomalies** in data
  - To assess **relationship** in data
    - **Explanatory** variables = **independent** variables
    - **Response** variables = **dependent** variables
  - To generate **hypotheses**...
  - and because … **if we don't, we will have troubles later**

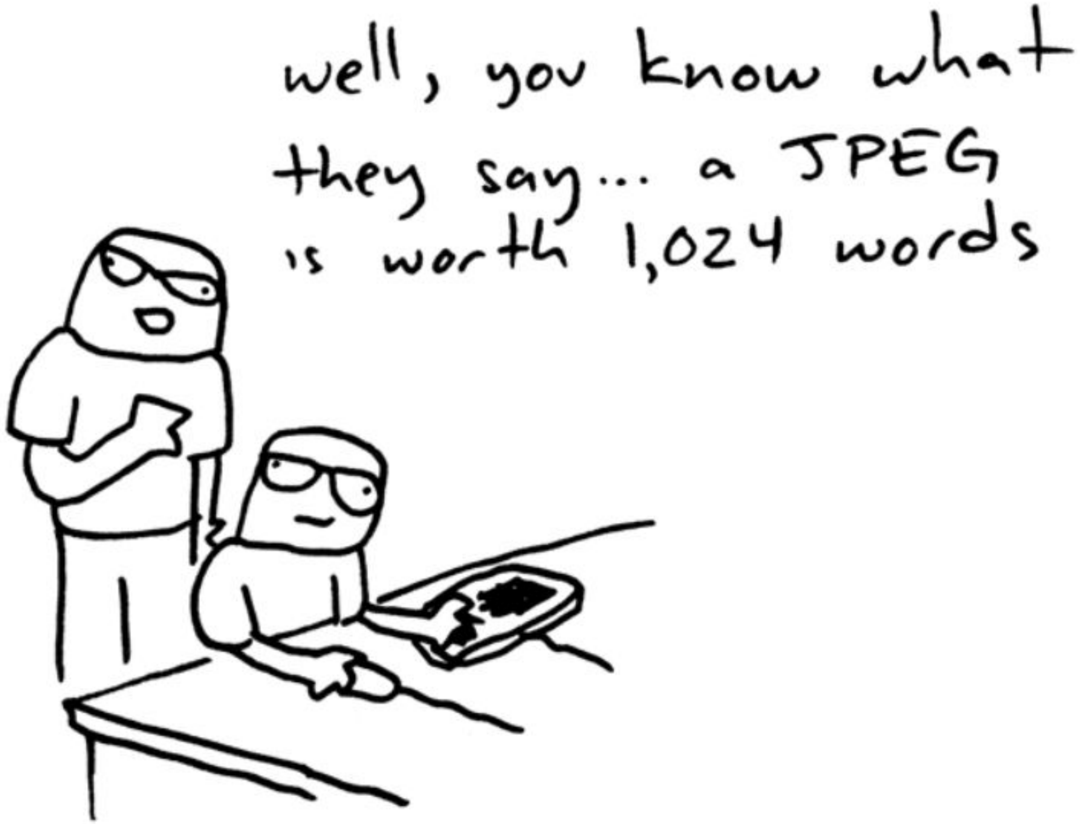
Exploratory Data Analysis

John Tukey (1961) EDA Focus on:

- Understand the data's **underlying structure**
- Develop **intuition** about the data set
- How the data were collected (to aid in **cleaning**)
- How to further **investigate** with statistical methods

EDA is Any Initial Investigations of Data:

1. **Basic**: Data Visualization and Numerical Summary
2. **Intermediate**: Statistical Hypothesis Tests with Confirmatory Analysis
3. **Advanced**: Unsupervised learning, PCA, SVD and Clustering



Recall: Data Wrangling

- **Data Wrangling**, sometimes referred to as data munging, is the process of transforming, mapping and manipulate data from one "**raw and messy**" data form into another format with the intent of making it more appropriate. If Pandas is not productive enough, you can try: DataWrangler, OpenRefine or AutoMunge.
- **Automunge** is a tool for automating the final steps of data wrangling of structured (tabular) data prior to the application of machine learning.

Outline – Exploratory Analysis

1. **Fundamentals of EA**
   - Objectives and Scopes of EA
2. **Statistical Summary of Data**
   - Numerical Data
3. **Statistical Hypothesis Testing**
   - Confirmatory Data Analysis
4. **Dimensionality Reduction**
   - LDA, PCA and SVD for EDA

Numerical Summaries of Data

Question: What summary we have concern?

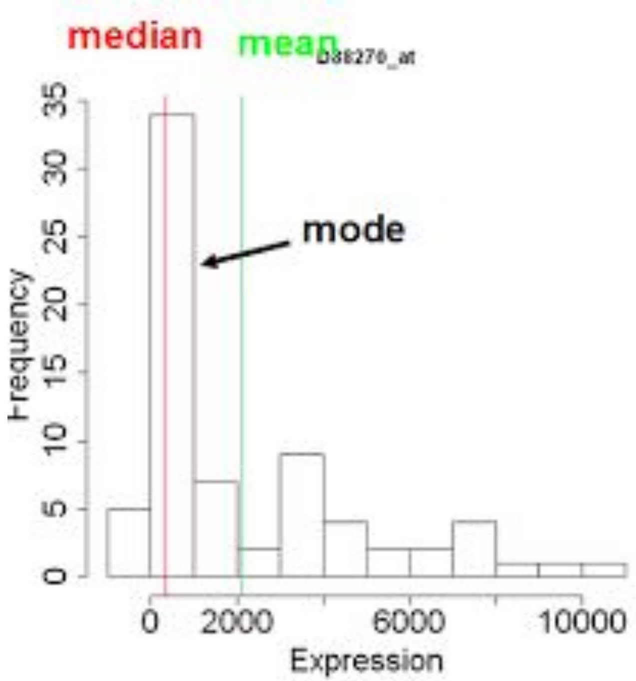
- **Central Tendency Measures.** Give a “center” around which the measurements in the data.
- **Variation or Variability Measures.** Describe “data spread” or how far away from center.
- **Relative Standing Measures.** Describe relative position of specific measurements in the data.

What is Numerical Summaries?

- Help to describe and compare samples
- Give information about corresponding parameters
- **Qualitative** data can be summarized by counts or percentages.
- **Quantitative** data can be summarized by measures of **location**, **scale** and **shape**.

Measures of Location

- For observations $x_1, x_2, \ldots\ldots x_n$ let x_j denote the *j*'th smallest observation:
  - **Sample Mean** $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
  - **Sample Median** $$x_M = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left[x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right] & \text{if } n \text{ is even} \end{cases}$$ $$= x_{\left(\frac{n}{2}+\frac{1}{2}\right)}$$
  - **Sample Mode** the value which occurs most frequently in the sample



Location: Mean

- **The Mean:** To calculate the average of a set of observations, add their value and divide by the number of observations.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Weighted means:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Trimmed:

$$\bar{x} = \alpha$$

Geometric:

$$\bar{x} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

Harmonic:

$$\bar{x} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Location: Median

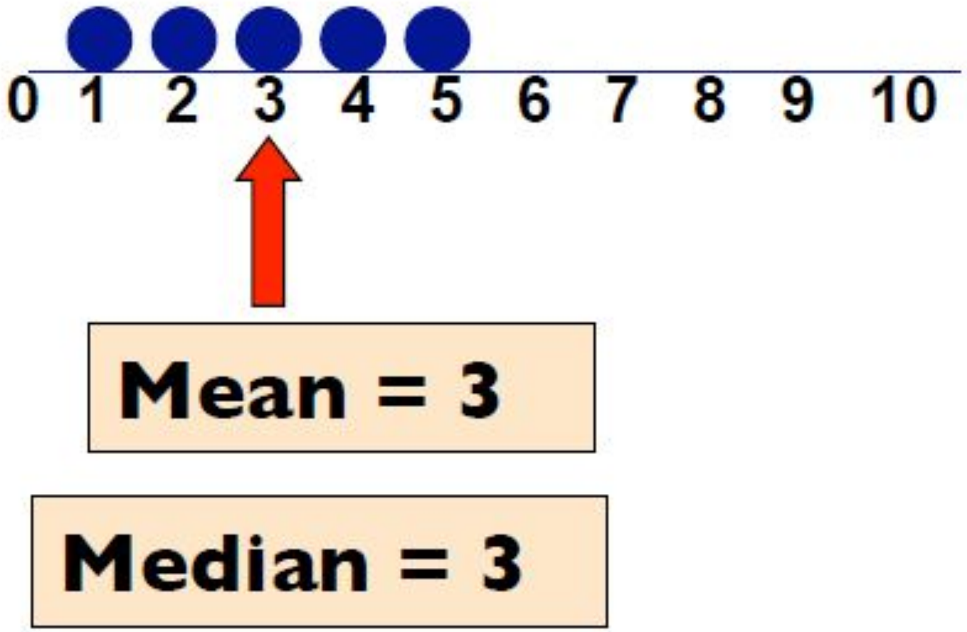
- **Median** – the exact middle value
- **Calculation**:
  - If there are an **odd** number of observations, find the middle value
  - If there are an **even** number of observations, find the middle two values and average them

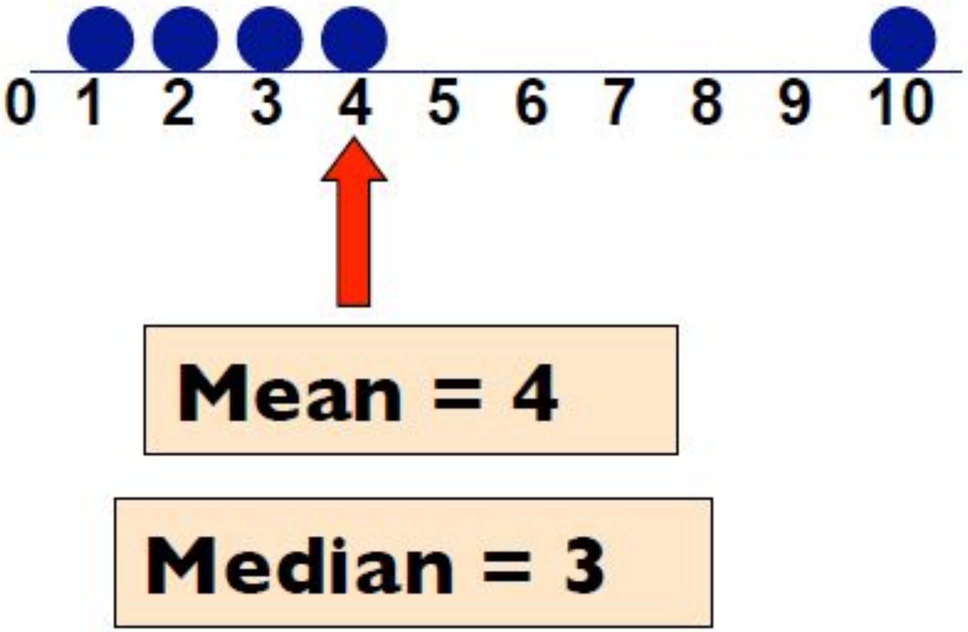
- **Example:** Age of participants: 17 19 21 22 23 23 23 38
- **Median =**

Median = (22+23)/2 = 22.5

Which Location Measure is Best?

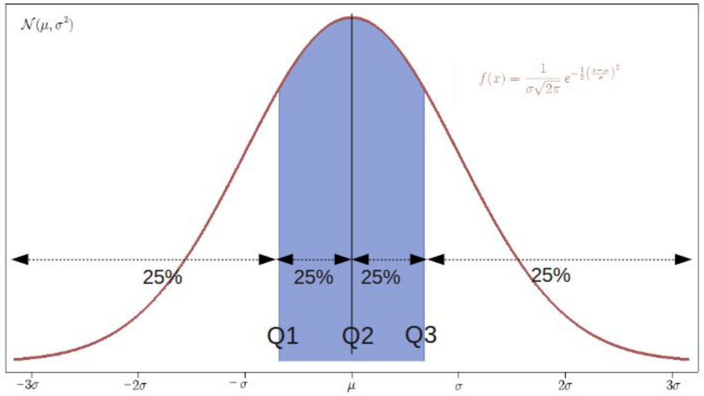
- **Mean** is best for symmetric distributions without outliers
- **Median** is useful for skewed distributions or data with outliers

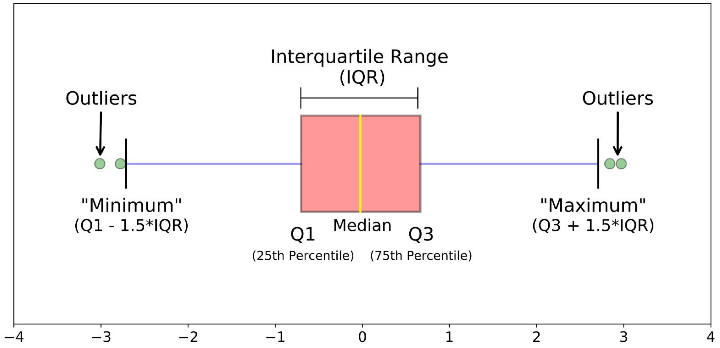




Location: Quantiles

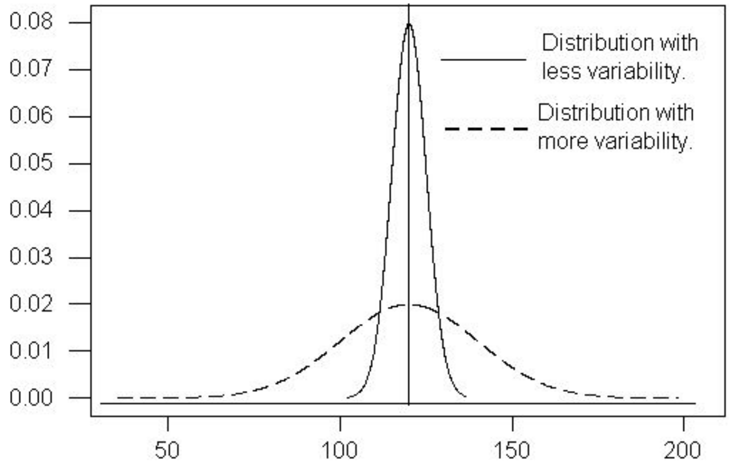
- **Sample Lower Quantiles:** $x_L = x_{\left(\frac{n}{4}+\frac{1}{2}\right)}$
- **Sample Upper Quantiles:** $x_U = x_{\left(\frac{3n}{4}+\frac{1}{2}\right)}$
- ***p*th Sample Percentile:** $x_{100p\%} = x_{\left(\frac{pn}{100}+\frac{1}{2}\right)}$
- **Five Number Summary:** $\left(x_{(1)}, x_L, x_M, x_U, x_{(n)}\right)$

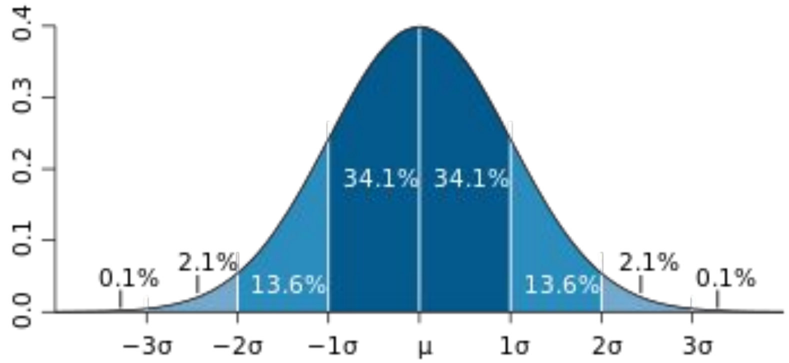




Measures of Scale: ANOVA

- **Sample Variance:** $\hat{\sigma}^2 = \dfrac{\sum_i^n (x_i - \bar{x})^2}{n-1}$
- **Sample Standard Deviation:** $\hat{\sigma} = \sqrt{\dfrac{\sum_i^n (x_i - \bar{x})^2}{n-1}}$
- **Inter-Quartile Range (IQR):** $x_U - x_L$
- **Sample Range:** $x_{(n)} - x_{(1)}$





Interesting Theoretical Results

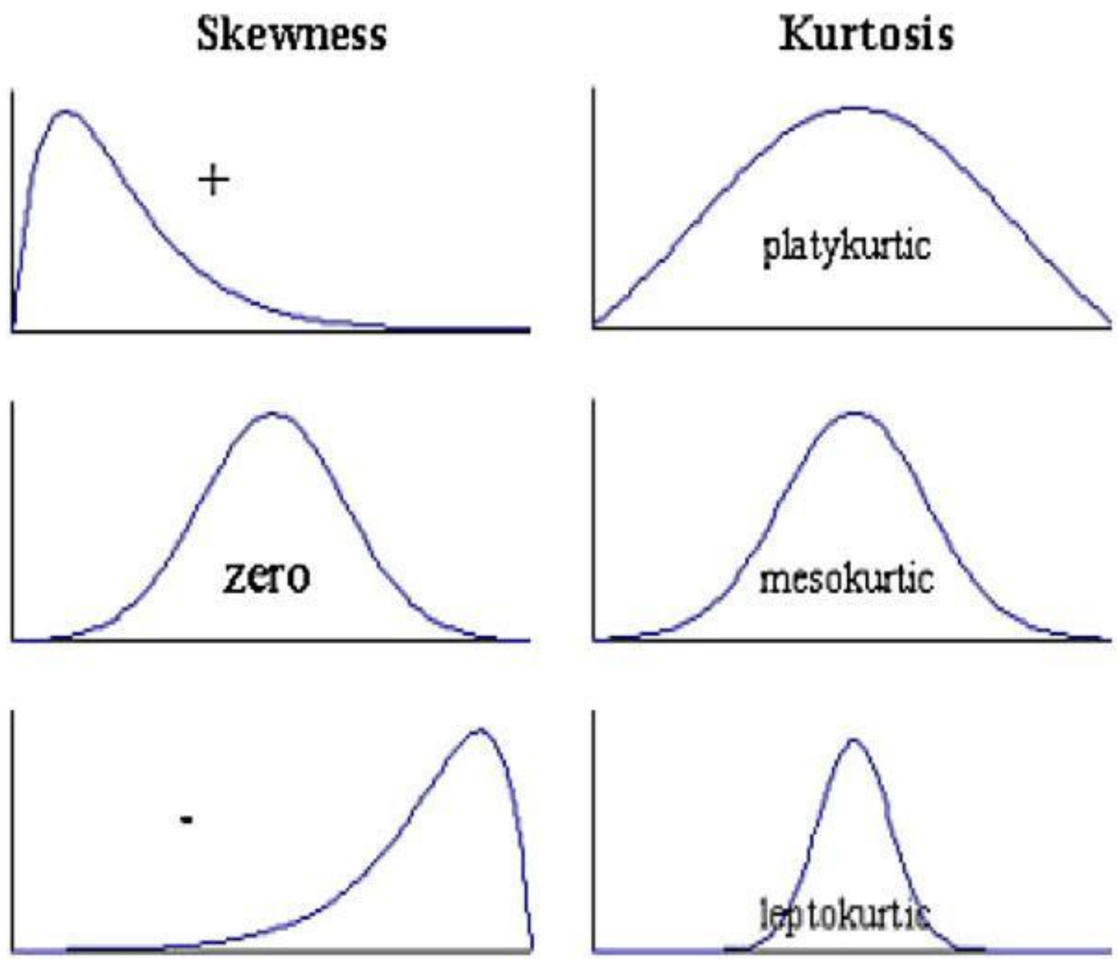
- Regardless of how the data are distributed, a certain percentage of values must fall within *k* standard deviations from the mean:

Note use of μ (mu) to represent "mean".

Note use of σ (sigma) to represent "standard deviation."

At least	within
$(1 - 1/1^2) = 0\%$	$k=1$ $(\mu \pm 1\sigma)$
$(1 - 1/2^2) = 75\%$	$k=2$ $(\mu \pm 2\sigma)$
$(1 - 1/3^2) = 89\%$	$k=3$ $(\mu \pm 3\sigma)$

Measures of Shape



- **Modality** number of peaks in the sample distribution
- **Skewness** is measure of symmetry in a distributions:
  - 0 : symmetric sample distribution
  - + : skewed to the right (long right-hand tail)
  - – : skewed to the left (long left-hand tail)
- **Kurtosis** is a measure of combined sizes of the two tails:
  - 3 : same peakedness as the normal distribution (**Mesokurtic**)
  - > 3 : more peaked - slim or long-tailed (**Leptokurtic**)
  - < 3: less peaked - flat, fat or short-tailed (**Platykurtic**)
  - Sometimes adjusted to give 0 for **Mesokurtic** distributions.

Skewness

- **Definition:**

$$Skewness = \frac{n}{(n-1)(n-2)}\sum\frac{(X_i - \bar{X})^3}{s^3} = \frac{n}{s^3(n-1)(n-2)}(S_{above} - S_{below})$$

Figure 1: Symmetrical Dataset with Skewness = 0

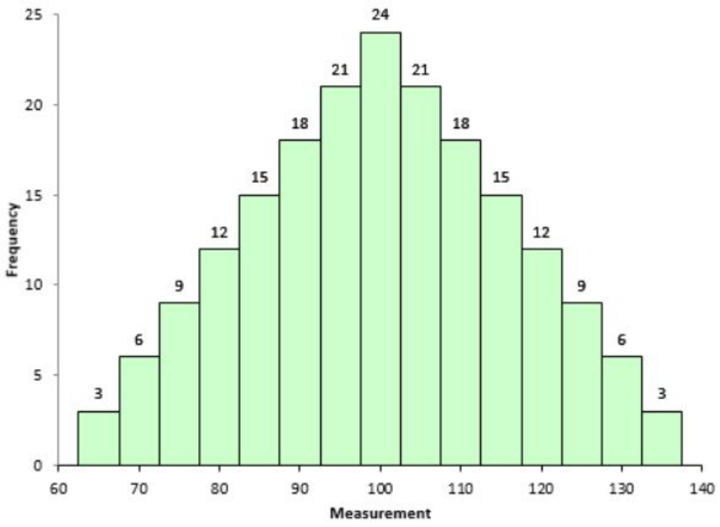


Figure 2: Dataset with Positive Skewness

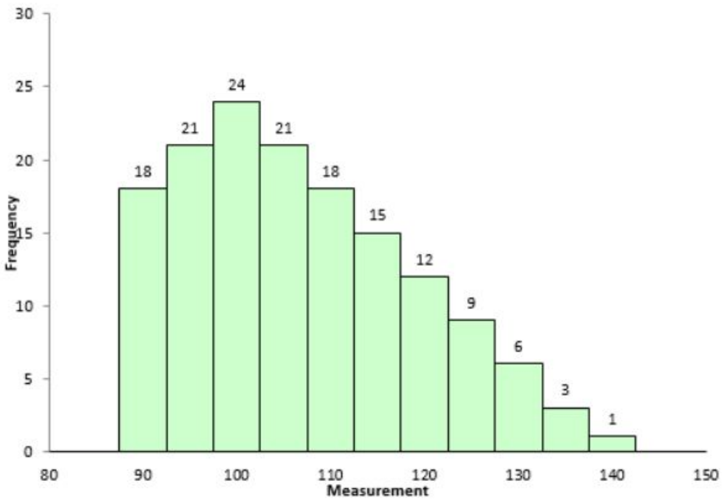
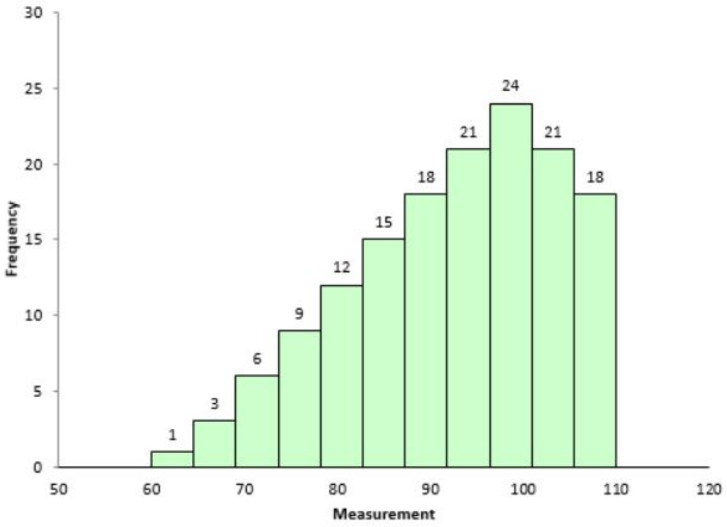


Figure 3: Dataset with Negative Skewness



Kurtosis

- Definition: $$Kurtosis = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \frac{(X_i - \bar{X})^4}{s^4} \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

- Kurtosis is the degree of peakedness of a distribution – Wolfram MathWorld

Figure 4: Negative Kurtosis Example

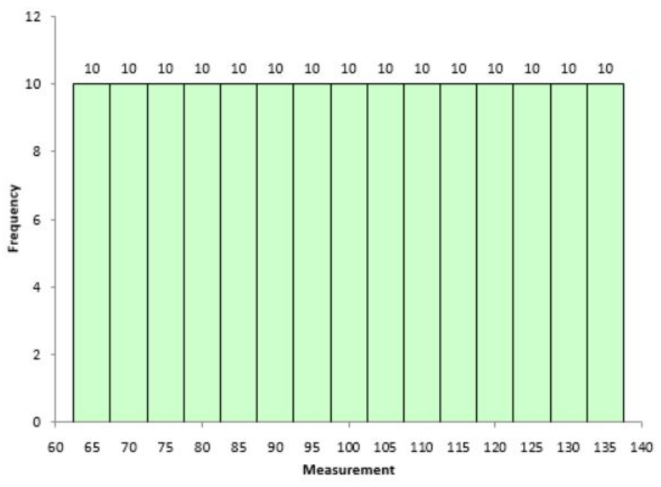
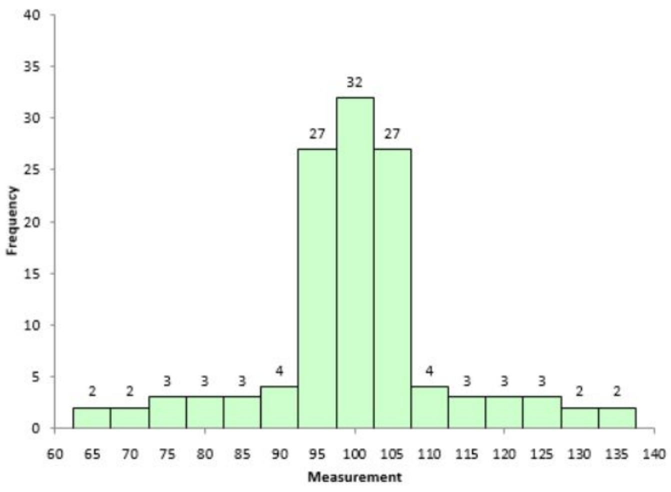


Figure 5: Positive Kurtosis Example



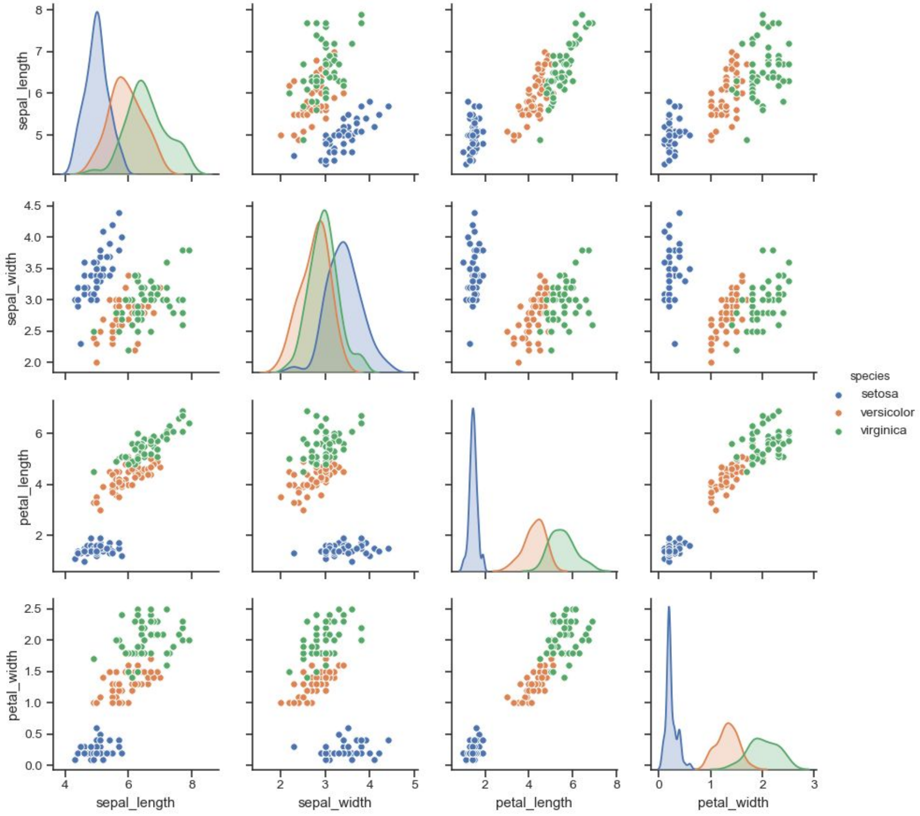
Covariance & Scatterplot

$$\mathrm{Cov}\,(X, Y) = \frac{\sum (X_i - \bar{X})(Y_j - \bar{Y})}{n}$$

- **Covariance** is a measure of the **joint variability**.
- **Positive Covariance:** variables show similar behavior.
- **Negative Covariance:** variables show opposite behavior.

- The sign of the covariance shows the tendency in the linear relationship between the variables. The magnitude of the covariance is not easy to interpret.

- **The normalized version of the covariance, the correlation coefficient, however, shows by its magnitude the strength of the linear relation.**



Correlation Coefficient (Bivariate)

- For observations $x_1, x_2, \dots\dots x_n$ and $y_1, y_2, \dots\dots y_n$ of two variables *X* and *Y*, **correlation coefficient** is a bivariate analysis that measures the strength of association between two variables and the direction of the relationship.

  - **Pearson** : measure the degree of the **linear** relationship

  - **Kendall** : measures the strength of **dependence** between two variables.

  - **Spearman:** measure the degree of **association** between two variables.

- **Note:** correlation does not imply cause!

Multivariate Analysis (MANOVA)

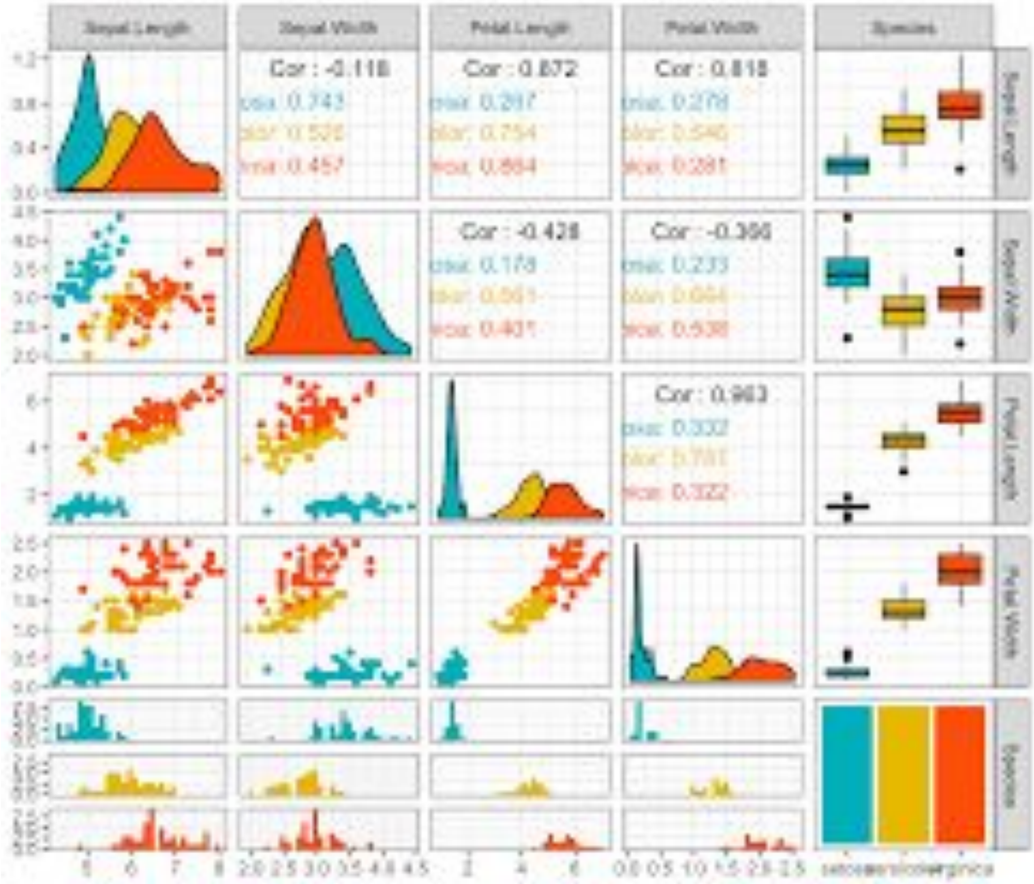
- **Multivariate Analysis of Variance (MANOVA)** is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time.

- **Uses for multivariate analysis include:**
  - Design for capability (also known as capability-based design)
  - Inverse design, where any variable can be treated as an independent variable
  - Analysis of Alternatives (AoA), the selection of concepts
  - Analysis of concepts with respect to changing scenarios
  - Identification of critical design-drivers and correlations

Multivariate Analysis (MANOVA)

MVA contains several process below:

1. **Plotting Multivariate Data:** Matrix Scatterplot and Profile Plot
2. **Calculating Summary of Statistics:** Means, Variances, Correlations
3. **Principal Component Analysis (PCA)**
4. **Linear Discriminant Analysis (LDA)**



Python Reference:
https://python-for-multivariate-analysis.readthedocs.io/

Outline – Exploratory Analysis

1. **Fundamentals of EA**
   - Objectives and Scopes of EA
2. **Statistical Summary of Data**
   - Numerical Data
3. **Statistical Hypothesis Testing**
   - Confirmatory Data Analysis
4. **Dimensionality Reduction**
   - LDA, PCA and SVD for EDA

Incorporating Models

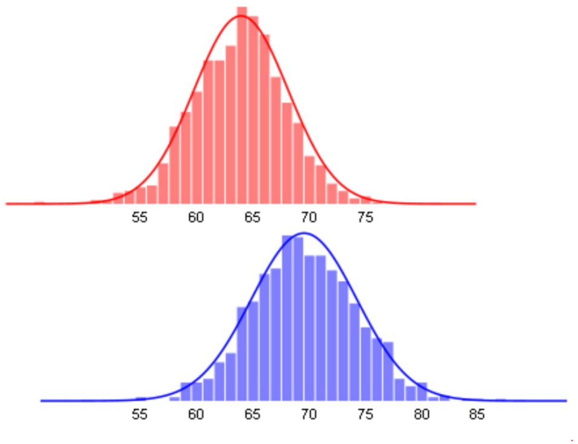
- **Exploratory data analysis** is sometimes compared to detective work: it is the process of gathering evidence. **Confirmatory data analysis** is comparable to a court trial: it is the process of evaluating evidence. **Exploratory analysis** and **confirmatory analysis** can—and should—proceed side by side" (Tukey; 1977).

Confirmatory	Exploratory
Test a priori hypothesis	Generates a posteriori hypothesis
Normally based on existing study	Discovers new knowledge
Stringent research restrictions	Less stringent research restrictions
Deals with knowns/unknowns	Deals with unknowns

Example: Heights by Gender

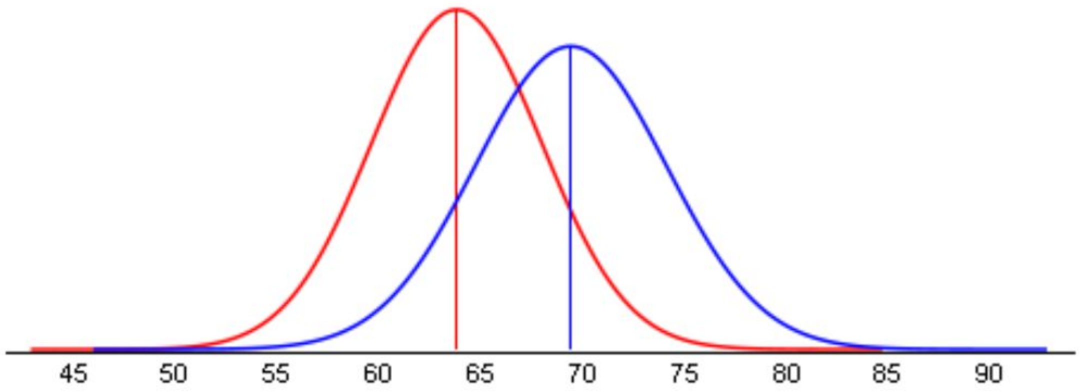
Gender	Male / Female
Height (in)	Number

- $\mu m = 69.4$ $\sigma m = 4.69$ $Nm = 1000$
- $\mu f = 63.8$ $\sigma f = 4.18$ $Nf = 1000$



Is the difference in heights significant?

- In other words: assuming no true difference, what is the probability that our data is due to chance?



Formulating Hypothesis

Null Hypothesis (**H0**) : $\mu m = \mu f$ (population)

Alternate Hypothesis (**Ha**) : $\mu m \neq \mu f$ (population)

A **statistical hypothesis test** assesses the likelihood of the null hypothesis.

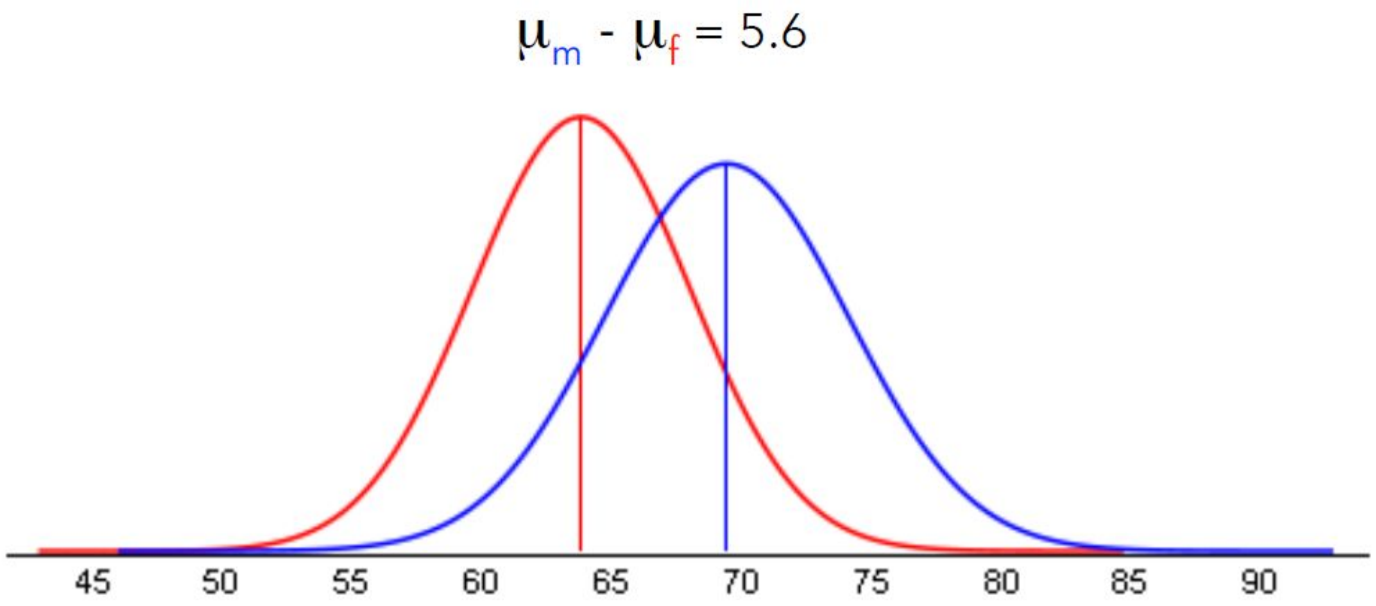
What is the probability of sampling the observed data assuming the population means are equal?

This is called the ***p*-value**. Testing procedure, compute a **test statistic.**

Testing Procedure

- Compute a **test statistic**. This is a number that in essence summarizes the difference.
- The possible values of this statistic come from a **known probability distribution**.
- According to this distribution, determine the probability of seeing a value meeting or exceeding the test statistic. This is the ***p*-value**.

$$Z = \frac{\mu_m - \mu_f}{\sqrt{\sigma^2_m / N_m + \sigma^2_f / N_f}}$$



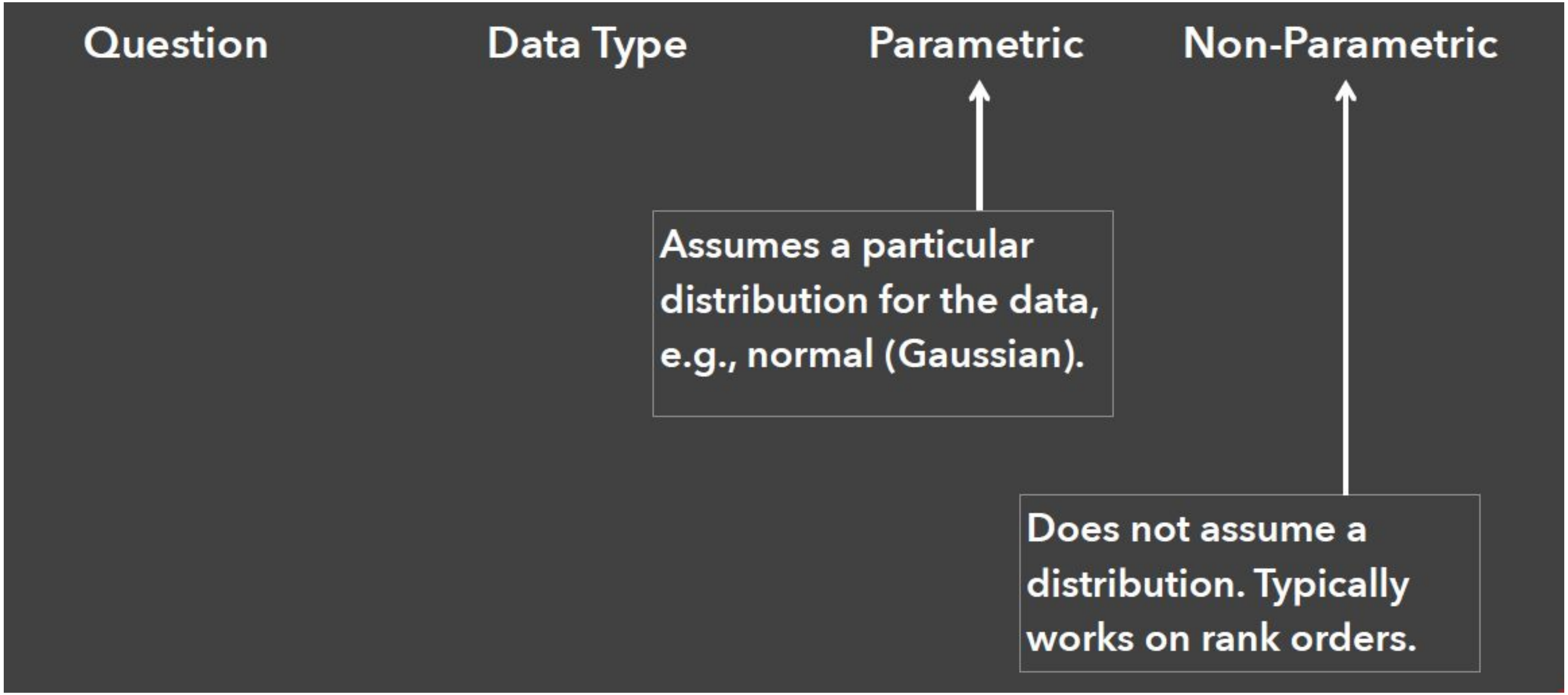
Statistical Significance

In general, a test statistic is selected or defined in such a way as to quantify, within observed data, behaviors that would distinguish the null from the alternative hypothesis, where such an alternative is prescribed, or that would characterize the null hypothesis if there is no explicitly stated alternative hypothesis. Two widely used test statistics are the ***t*-statistic** and the ***F*-test**.

Interpretation:

- The threshold at which we consider it safe (or reasonable?) to ***reject the null hypothesis***.
- If $p < 0.05$, we typically say that the observed effect or difference is **statistically significant**.
- This means that there is a less than 5% chance that the observed data is due to chance.
- Note that the choice of 0.05 is a somewhat arbitrary threshold (chosen by R. A. Fisher)

Common Statistical Methods



Common Statistical Methods

Question	Data Type	Parametric	Non-Parametric
Do data distributions have different "centers"? (aka "location" tests)	2 uni. dists > 2 uni. dists > 2 multi. dists	t-Test ANOVA MANOVA	Mann-Whitney U Kruskal-Wallis Median Test
Are observed counts significantly different?	Counts in categories		$\chi 2$ (chi-squared)
Are two vars related?	2 variables	Pearson coeff.	Rank correl.
Do 1 (or more) variables predict another?	Continuous Binary	Linear regression Logistic regression	

Outline – Exploratory Analysis

1. **Fundamentals of EA**
   - Objectives and Scopes of EA
2. **Statistical Summary of Data**
   - Numerical Data
3. **Statistical Hypothesis Testing**
   - Confirmatory Data Analysis
4. **Dimensionality Reduction**
   - LDA, PCA and SVD for EDA

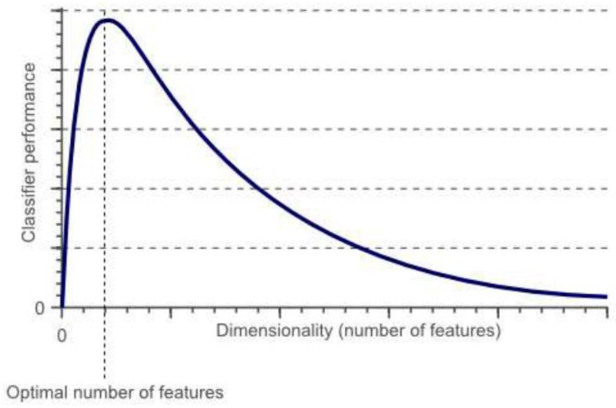
Dimensionality Reduction

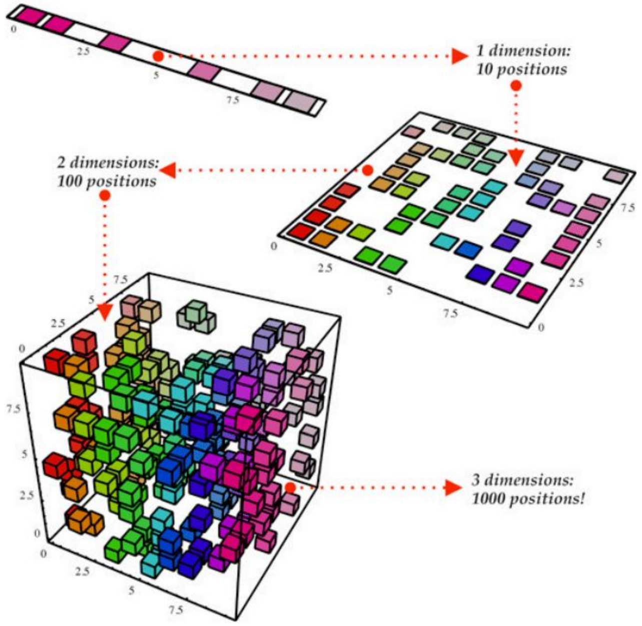
Problems: Curse of Dimensionality

- Theoretically, increasing features should improve performance. In practice, too many features leads to worse performance. Number of training examples required increases exponentially with dimensionality.

Solutions: Dimensionality Reduction

- Data can be represented by fewer dimensions (features). Reduce dimensionality by selecting subset (feature elimination). Combine with linear and non-linear transformations





Scopes of Dimensionality Reduction

- Dimensionality reduction techniques reduce number of features of the training dataset.
- This reduction is necessary to:

1. Eliminate the noise from the data
2. Visualize the data in 2 or 3 dimensions
3. Speed up the learning process
4. Enhance the learning results by eliminating correlated features.
5. Eliminate unnecessary features.
6. Compress the data size.

- **Two main approaches:**
  - **Projection** : project or map the data into a lower dimensional space (Eq. PCA, SVD).
  - **Manifold** **:** data in higher dimension is a manifold of lower dimension (Eq. MDS, LLE, etc).

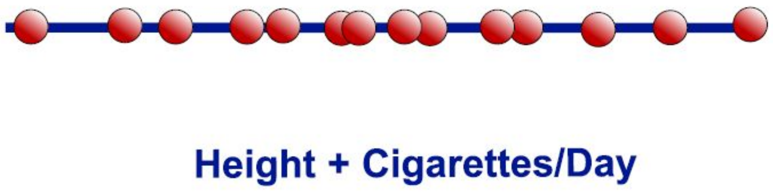
Solution: Dimensionality Reduction

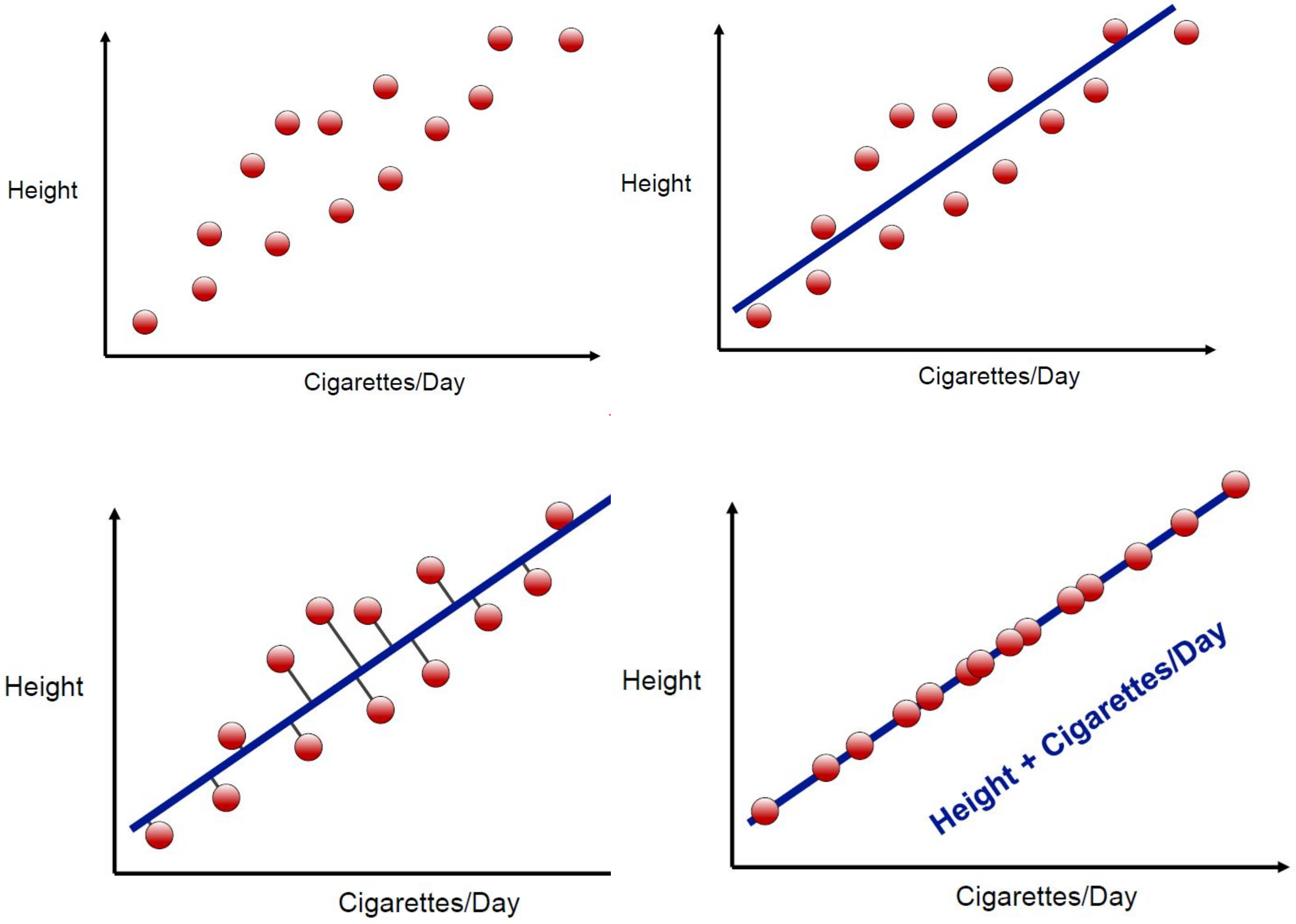
Sample Case:

- Two features: height and cigarettes per day
- Both features increase together (correlated)
- Can we reduce number of features to one?

Dimensionality Reduction:

Create single feature that is combination of height and cigarettes. This is Principal Component Analysis (PCA).



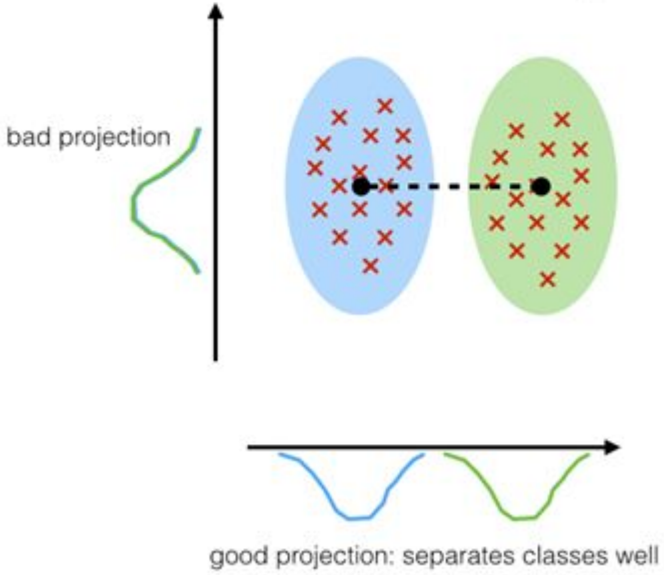


What we will learn

- **LDA: Linear Discriminant Analysis**
  - **Approach:** Find a linear combination of features that characterizes or separates two or more classes of objects or events.
- **PCA: Principal Component Analysis**
  - **Approach:** Uses an orthogonal transformation to convert bigger set of features intro smaller set of linearly uncorrelated variables called principal components.
- **SVD: Singular Value Decomposition**
  - **Approach**: Use matrix decomposition to find best projection axis with minimum reconstruction error.

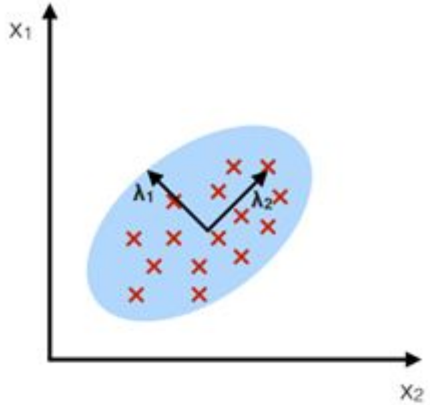
LDA vs PCA

LDA:
maximizing the component axes for class-separation



StatsQuest LDA

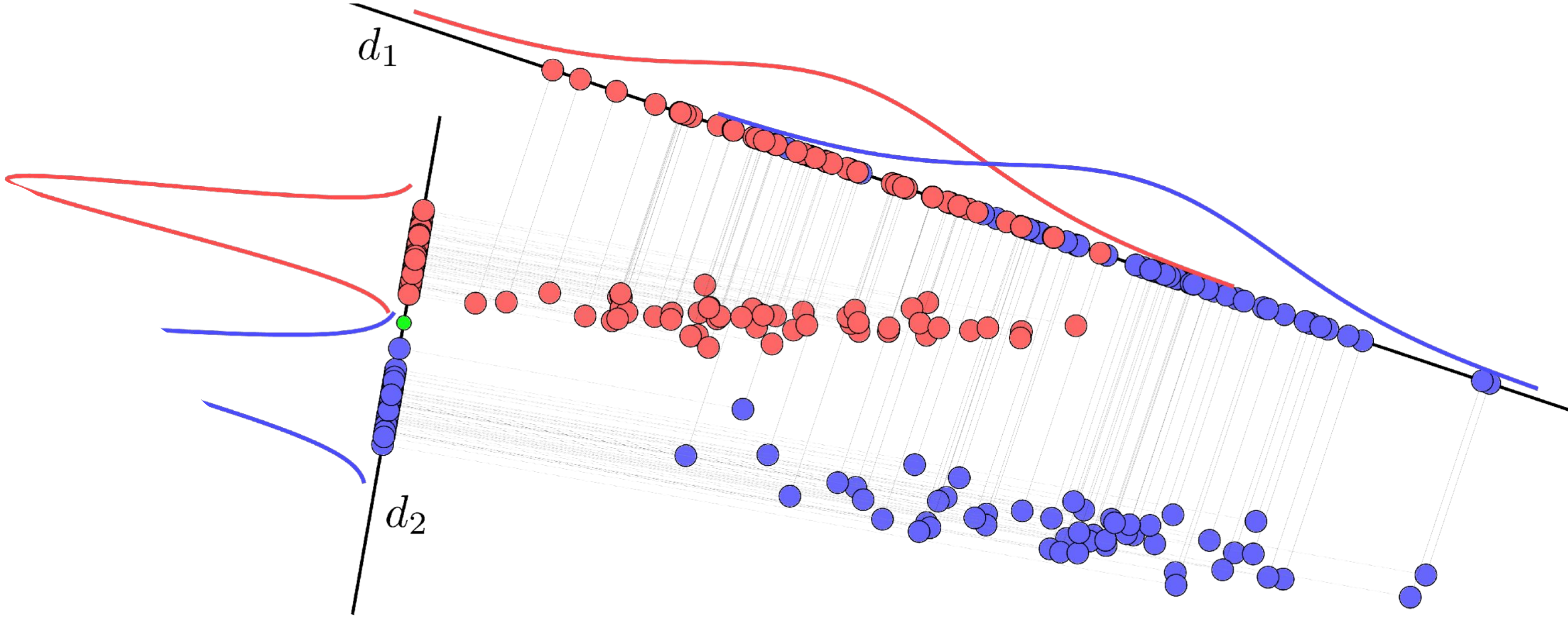
PCA:
component axes that maximize the variance



StatsQuest PCA

Linear Discriminant Analysis

- **LDA: Pick a new dimension that gives:**
  - Maximum separation of means of projected classes
  - Minimum variance between each projected class
- **Solution:** Eigenvectors based on between-class and within-class covariance matrices.



Linear Discriminant Analysis

- **LDA Definition**: A linear combination of features separates two or more classes. It is a supervised dimensionality reduction technique to be used with continuous independent variables and a categorical dependent variables. LDA does classification by assuming that the data within each class are **normally distributed**:

$$f_k(x) = P(X = x|G = k) = N(\mu_k, \Sigma)$$

- We allow each class to have its own mean μ_k , but we assume a common variance matrix Σ. Thus

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu_k)^T \Sigma^{-1} (x - \mu_k)\right\}$$

- **LDA Task:** Find *k* so that $P(G = k|X = x) \propto f_k(x)\pi_k$ is the largest.

Linear Discriminant Analysis

- **The linear discriminant functions are derived from the relation:**

$$\begin{aligned}\log(f_k(x)\pi_k) &= -\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)+\log(\pi_k)+C \\ &= x^T\Sigma^{-1}\mu_k-\frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k+\log(\pi_k)+C',\end{aligned}$$

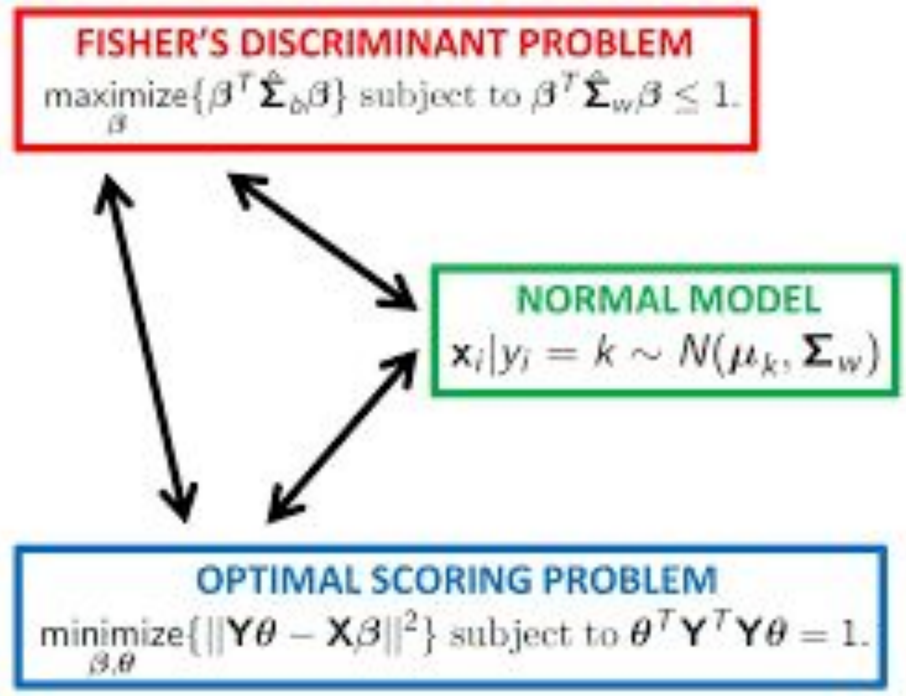
- And we denote: $\delta_k(x) = x^T\Sigma^{-1}\mu_k-\frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k+\log(\pi_k)$

- **The decision rule is:** $G(x) = \mathrm{argmax}_k\delta_k(x)$

LDA Algorithms

Two main algorithms for LDA:

- **Fisher Discriminant Analysis**
  - Gaussian classification rule with equal prior probabilities.
- **LDA by Optimal Scoring**
  - Use multivariate linear regression on derived responses.



Advanced Topic on LDA (Reading Material)

FDA Generalization:

- **FDA** – Flexible Discriminant Analysis
  - Allow non-linear decision boundary
- **PDA** – Penalized Discriminant Analysis
  - Expand the predictors into a large basis set, and then penalize its coefficients to be smooth
- **MDA** – Mixture Discriminant Analysis
  - Model each class by a mixture of two or more Gaussians with different centroids but same covariance, rather than a single Gaussian distribution as in LDA

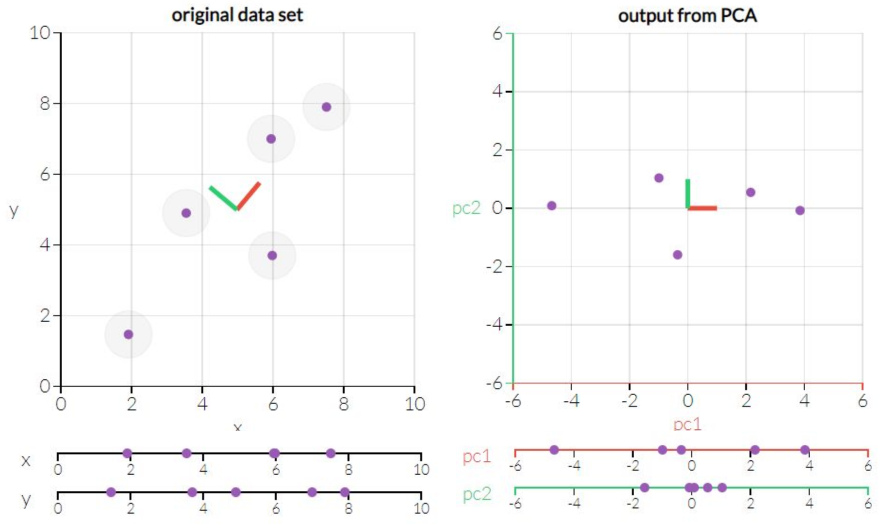
Principal Component Analysis

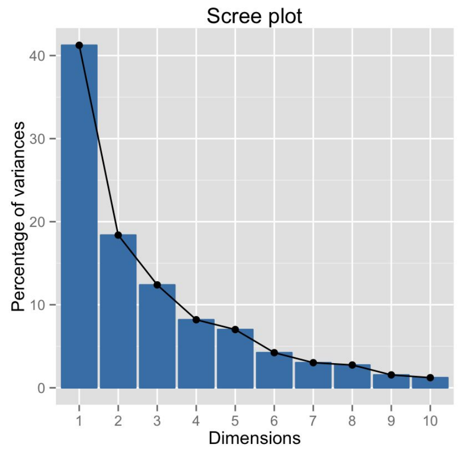
- **PCA is “an orthogonal linear transformation that transfers the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (first principal component), the second greatest variance lies on the second coordinate (second principal component), and so on.”**

- An exploratory technique to reduce the dimensionality of the data set, used to:
  1. Reduce number of dimensions in data
  2. Find patterns in high-dimensional data
  3. Visualize data of high dimensionality

Principal Component Analysis

- **PCA decomposes a dataset into a set of successive orthogonal components that explain a maximum amount of variance.**





Try: [PCA Web Demo]

Principal Component Analysis

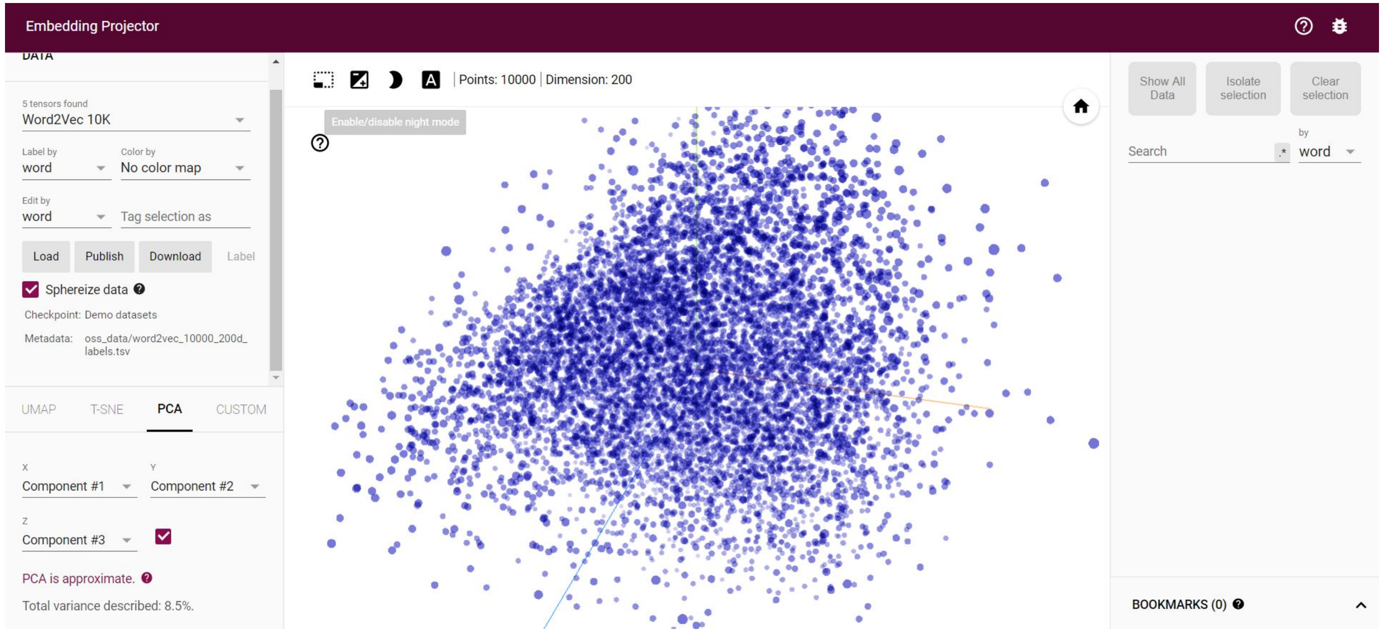
- **Keys to Interpret PCA:**
  - **Variance:** measure of the deviation from the mean for points in one dimension or measure of spread.
  - **Covariance:** measure of how much each of the dimensions varies from the mean with respect to each other.
    - + Covariance: both increase of decrease together
    - - Covariance: one increases other decreases
    - 0 Covariance: two dimension is independent.
  - **Covariance:** measure of how much each of the dimensions varies from the mean with respect to each other.
  - **Covariance Matrix:** Combination matrix of covariance
  - **Eigen Value and Eigen Vector:** $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$

$$\mathrm{var}(X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})}{(n-1)}$$

$$\mathrm{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)}$$

PCA for Embedding

- An **embedding** is a relatively low-dimensional space into which you can translate high-dimensional vectors. Embeddings make it easier to do ML on large inputs like sparse vectors representing words. Projector: https://projector.tensorflow.org/ Video lecture here.



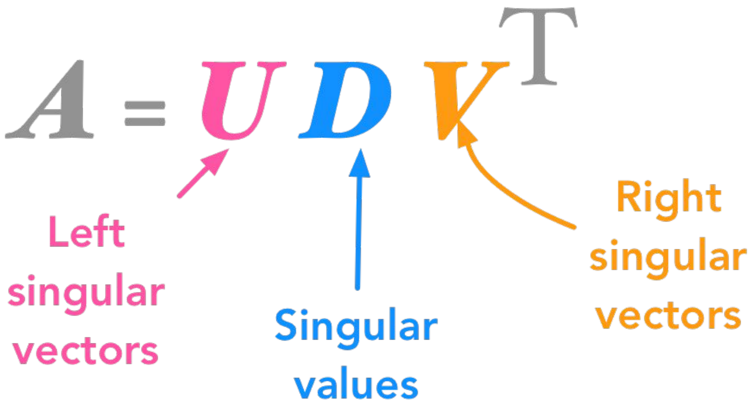
Singular Value Decomposition

- Let ***A*** be an $m \; x \; n$ matrix. The Singular Value Decomposition(SVD) of ***A***:

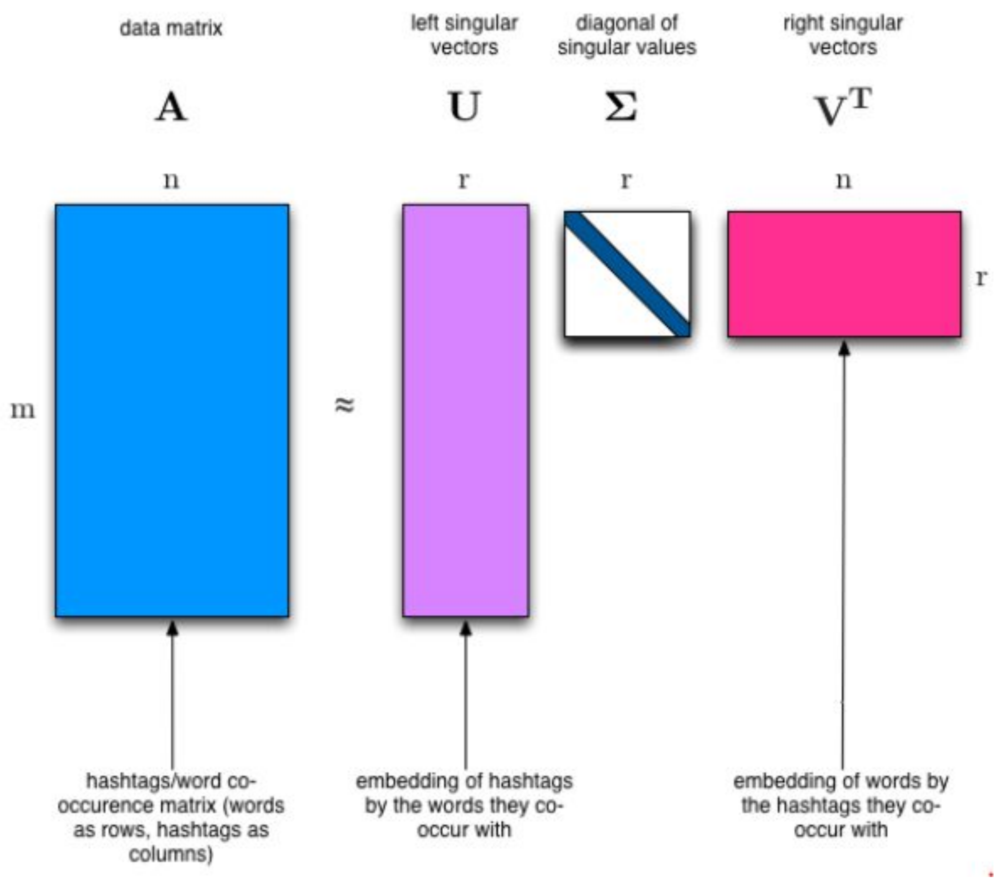
$$\boldsymbol{A} = \boldsymbol{U} \; \boldsymbol{D} \; \boldsymbol{V}^T$$

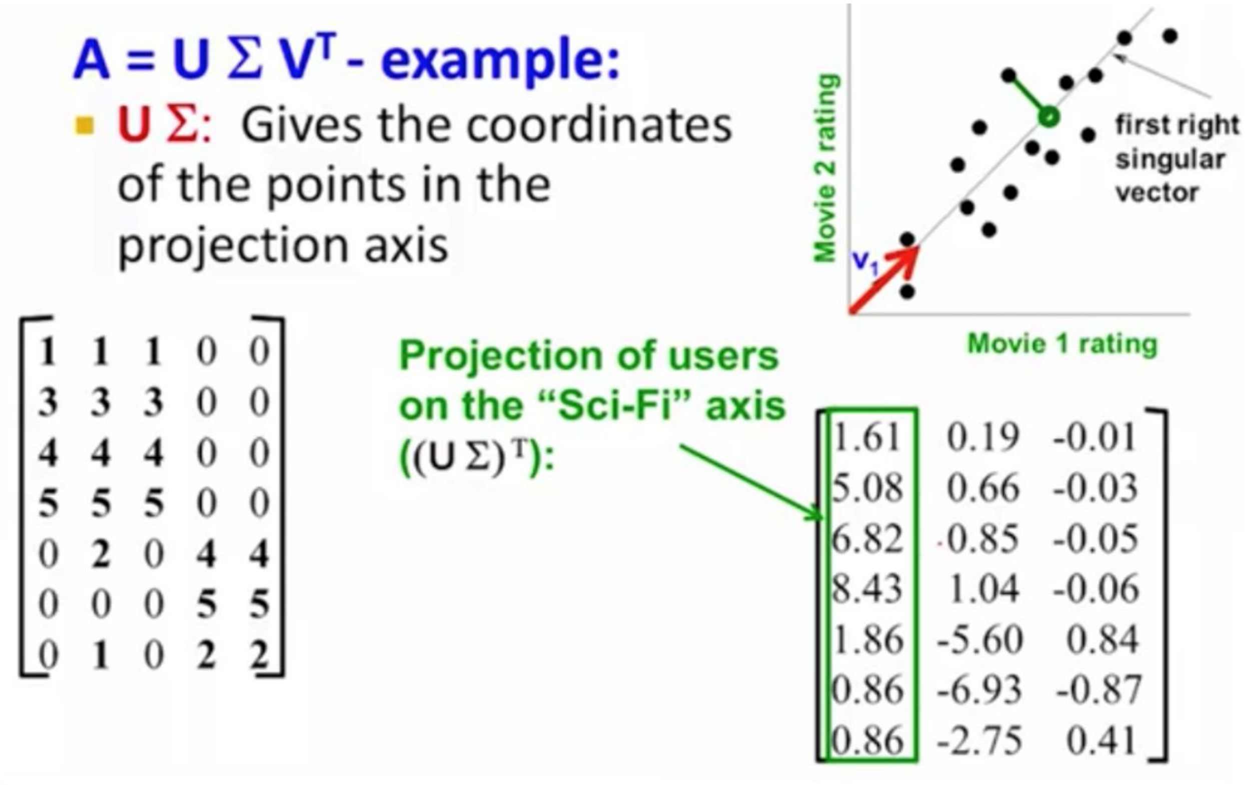
- *U* is orthogonal $m \; x \; r$, *V* is $n \; x \; r$ orthogonal and D is $r \; x \; r$ diagonal matrices.

- ***A***: Input data matrix (eg. *m* documents, *n* terms)
- ***U***: Left singular matrix (*m* documents, *r* **concepts**)
- ***D***: Singular value (strength of each **concept**)
- ***V***: Right singular matrix (*n* terms, *r* **concepts**)



Singular Value Decomposition



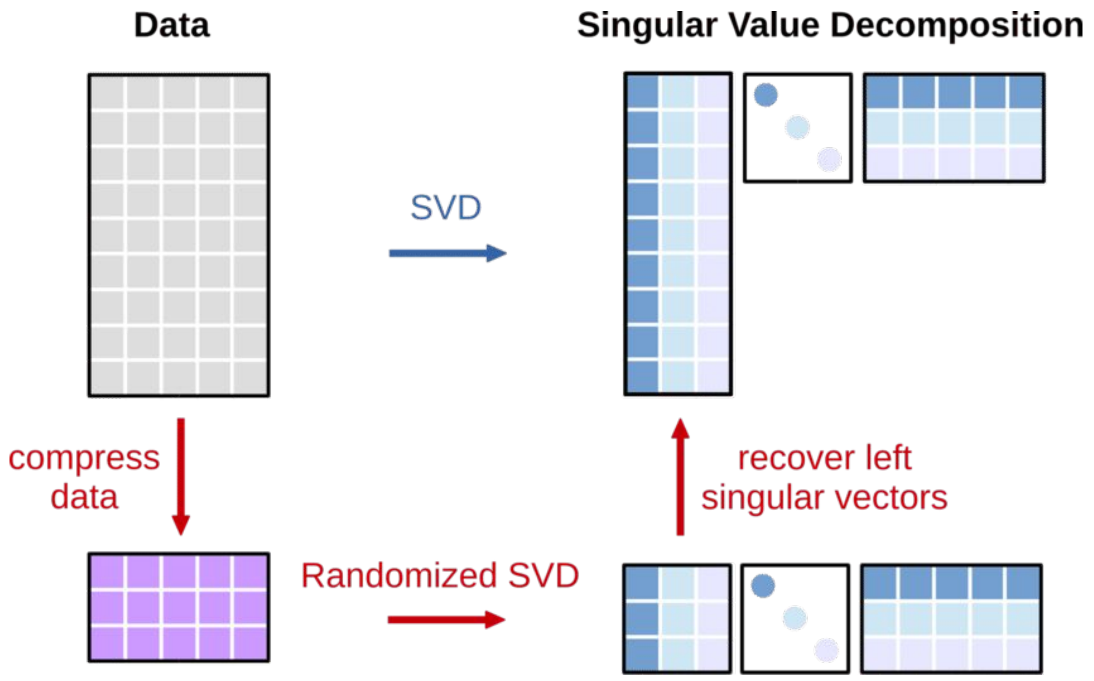


Singular Value Decomposition

- SVD allows an exact representation of **any matrix**, and also makes it easy to eliminate the less important parts of that representation to produce an approximate representation with any desired number of dimensions

- **SVD gives best axis to project on:**
    - Best = min sum of squares of projection error

- **In other words:**
    - Best = minimum reconstruction error.



Interpretation of SVD

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}}_{M} = \underbrace{\begin{bmatrix} .14 & 0 \\ .42 & 0 \\ .56 & 0 \\ .70 & 0 \\ 0 & .60 \\ 0 & .75 \\ 0 & .30 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} .58 & .58 & .58 & 0 & 0 \\ 0 & 0 & 0 & .71 & .71 \end{bmatrix}}_{V^{\mathrm{T}}}$$

- The key to understanding what SVD offers is in viewing the r columns of *U*, *Σ*, and *V* as representing **concepts** that are hidden in the original matrix *M*. These concepts are clear; one is "**science fiction**" and the other is "**romance**".

http://infolab.stanford.edu/~ullman/mmds/ch11.pdf

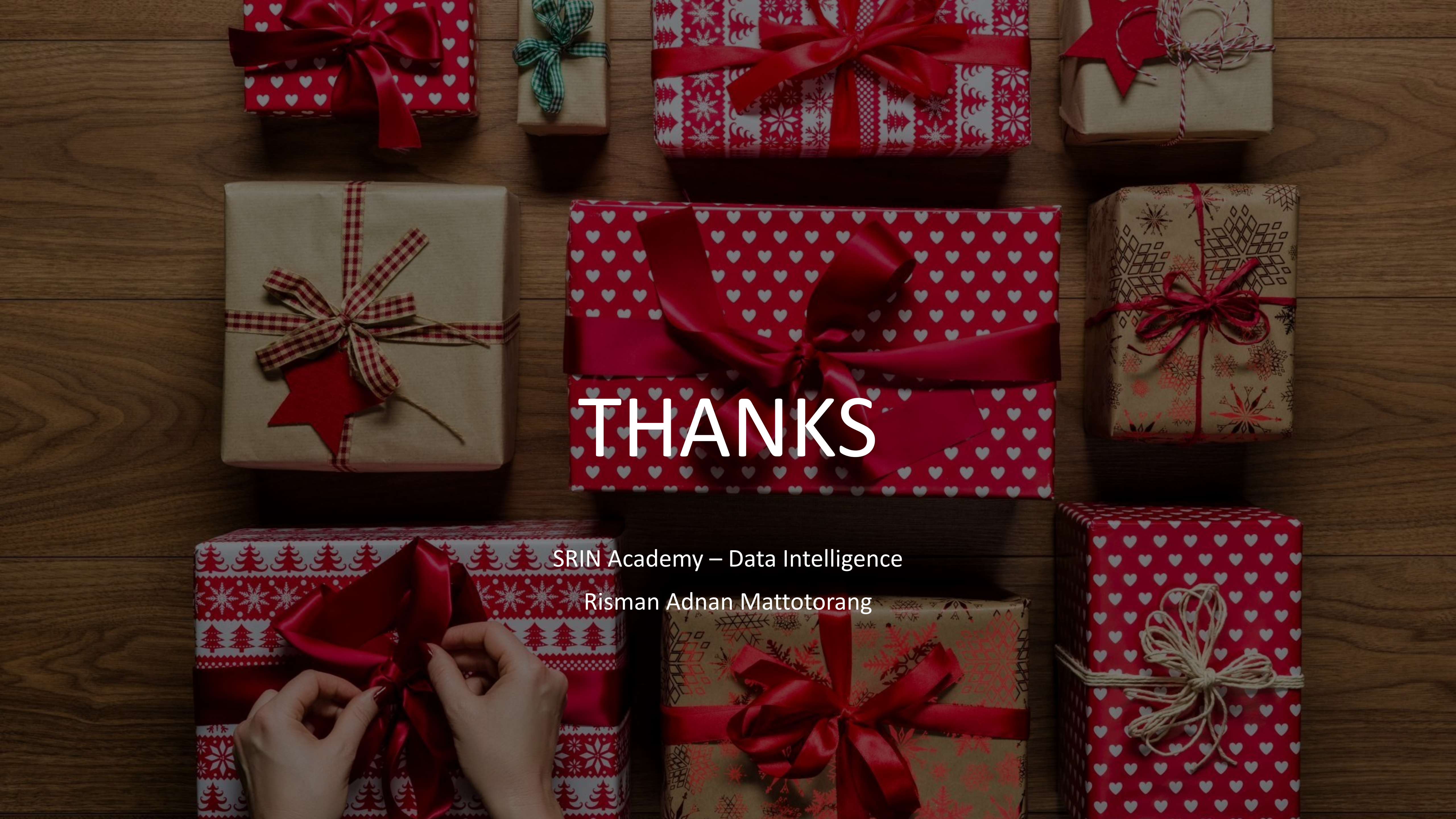
Dimensionality Reduction with SVD

- **Q**: How exactly dimensionality reduction in SVD?
- **A**: As simple as setting the smallest singular value to zero

$$
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 0.18 & \cancel{0} \\ 0.36 & \cancel{0} \\ 0.18 & \cancel{0} \\ 0.90 & \cancel{0} \\ 0 & \cancel{0.53} \\ 0 & \cancel{0.80} \\ 0 & \cancel{0.27} \end{bmatrix}
\text{x}
\begin{bmatrix} 9.64 & 0 \\ 0 & \cancel{5.29} \end{bmatrix}
\text{x}
\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0.71} & \cancel{0.71} \end{bmatrix}
$$

Homework

- StatQuests (PCA, LDA, SVD)
- Learn PCA, LDA and SVD in ScikitLearn
- Learn Pandas Framework



THANKS

SRIN Academy – Data Intelligence

Risman Adnan Mattotorang