Syllabus & Course Overview

Machine Learning and Big Data - DATA622

CUNY School of Professional Studies

Agenda

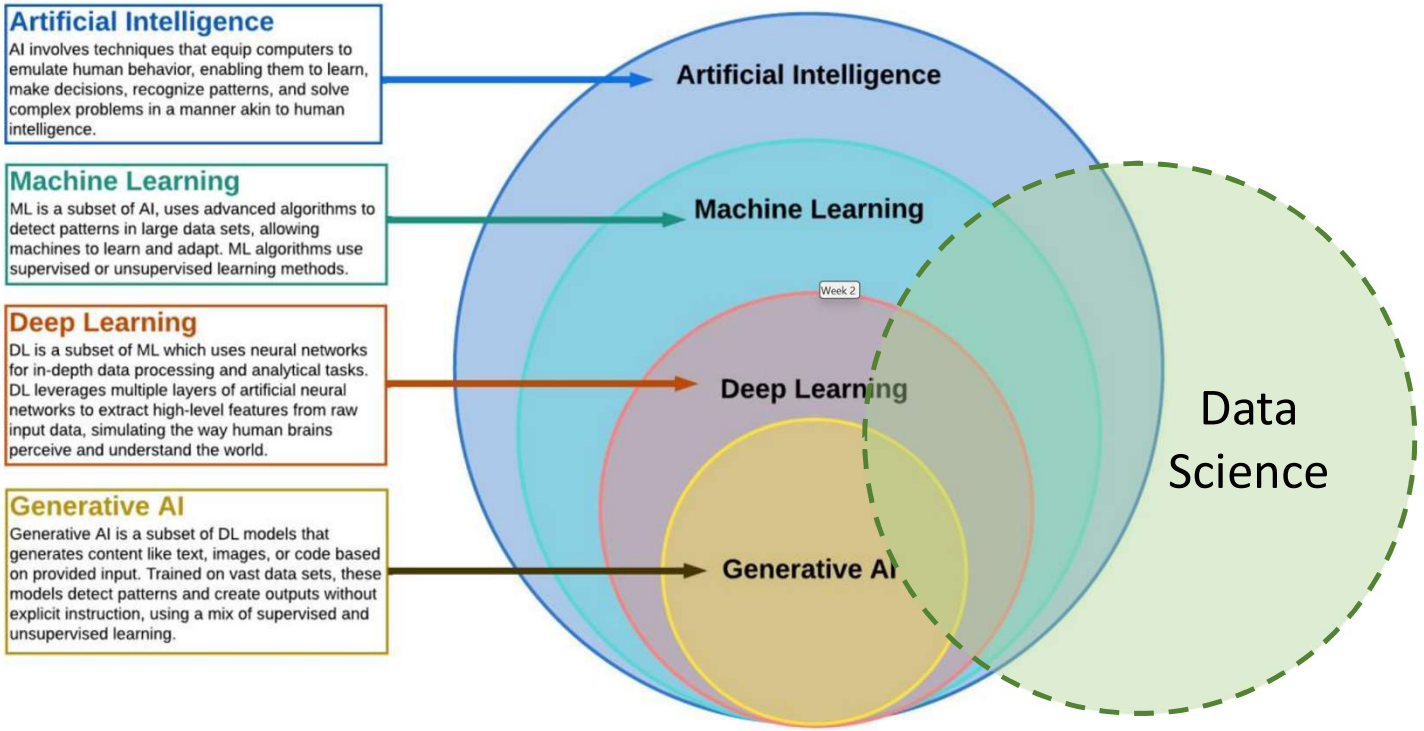
1. Introductions
2. Syllabus
3. Grading
4. Calendar
5. Textbooks
6. Week 1 – Tasks

Weekly calls

- Class calls will be recorded and posted online
- Thursday's 7PM ET
  Call to go over Weekly tasks and concepts.

Details on BrightSpace.

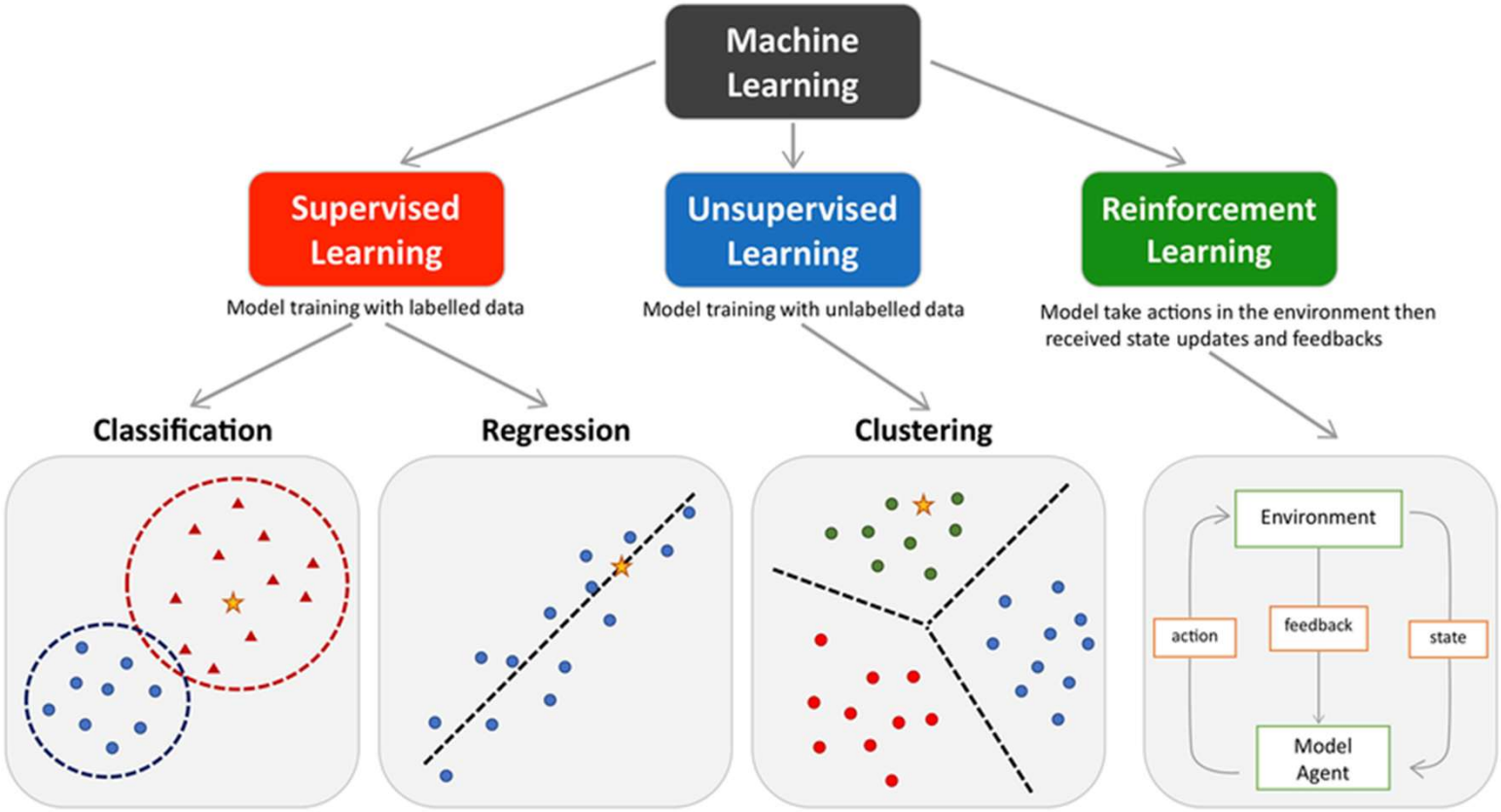
Machine Learning



Source: "The Application of AutoML Techniques in Diabetes Diagnosis: Current Approaches, Performance, and Future Directions", Zhuhadar, Lily Popova; Lytras, Miltiadis, 2023

CUNY School of Professional Studies

Types of Machine Learning

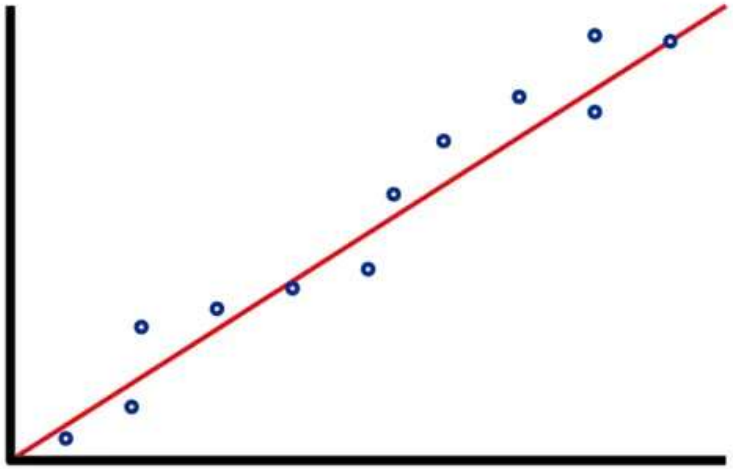


Source: "Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges", Pend et al, Sep 2021

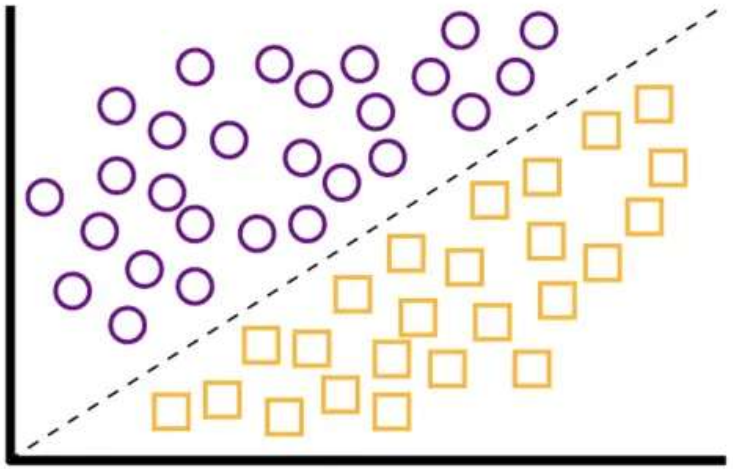
CUNY School of Professional Studies

Classification vs Regression

Regression

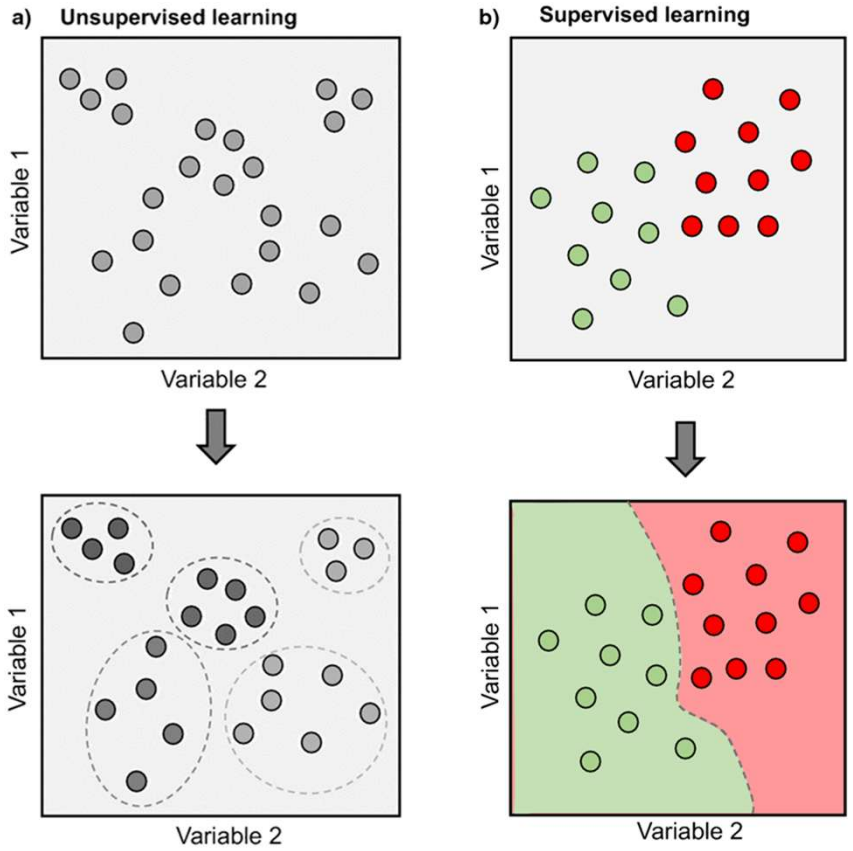


Classification



Unsupervised vs Supervised

Unsupervised learning
Input data is unlabeled
Has no feedback mechanism
Assigns properties of given data to classify it
Divided into Clustering & Association
Used for analysis
Algorithms include: k-means clustering, hierarchical clustering, apriori algorithm
A unknown number of classes



Supervised learning
Input data is labeled
Has a feedback mechanism
Data is classified based on the training dataset
Divided into Regression & Classification
Used for prediction
Algorithms include: decision trees, logistic regressions, support vector machine
A known number of classes

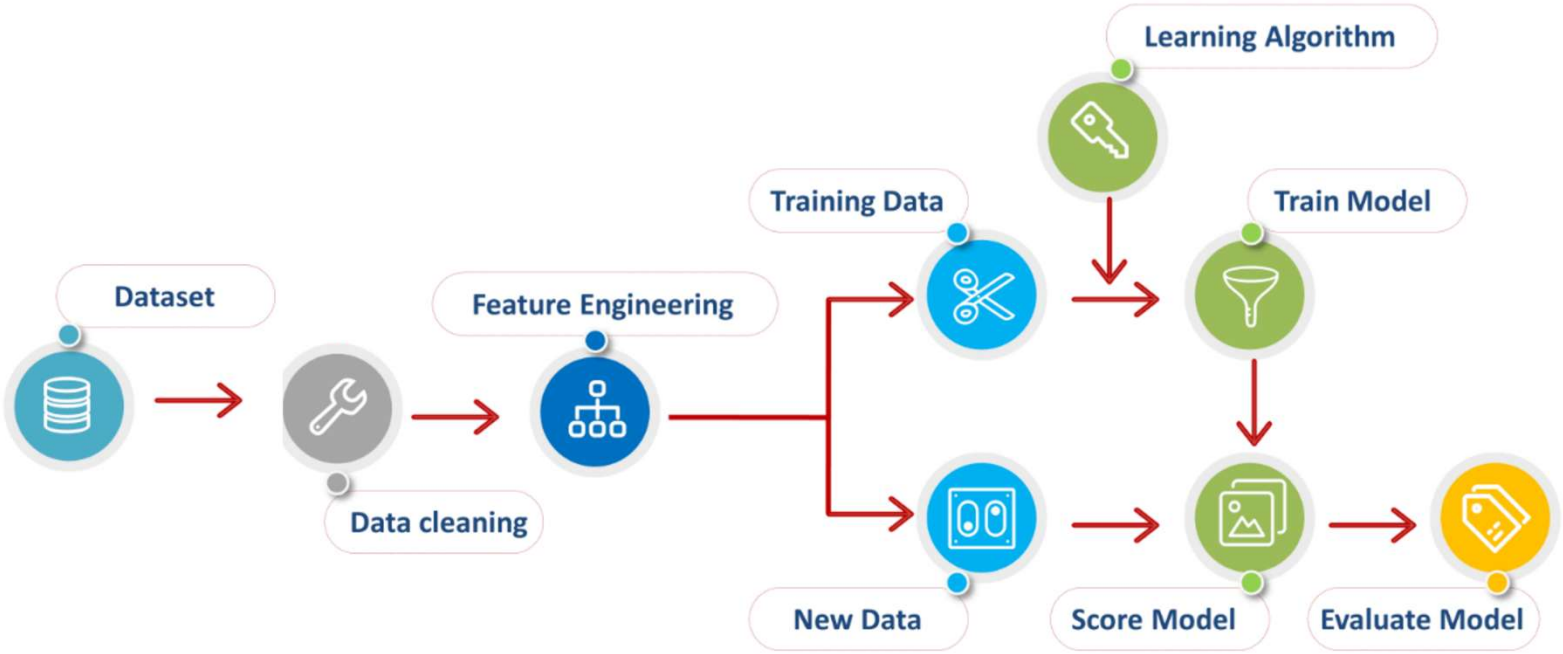
CUNY School of Professional Studies

Machine Learning Process

Source: "Neural Network for Porosity Prediction in Carbonate Formation", Huda F. Al-Saad, 2023

CUNY School of Professional Studies

Machine Learning Process



Source: "Neural Network for Porosity Prediction in Carbonate Formation", Huda F. Al-Saad, 2023

CUNY School of Professional Studies

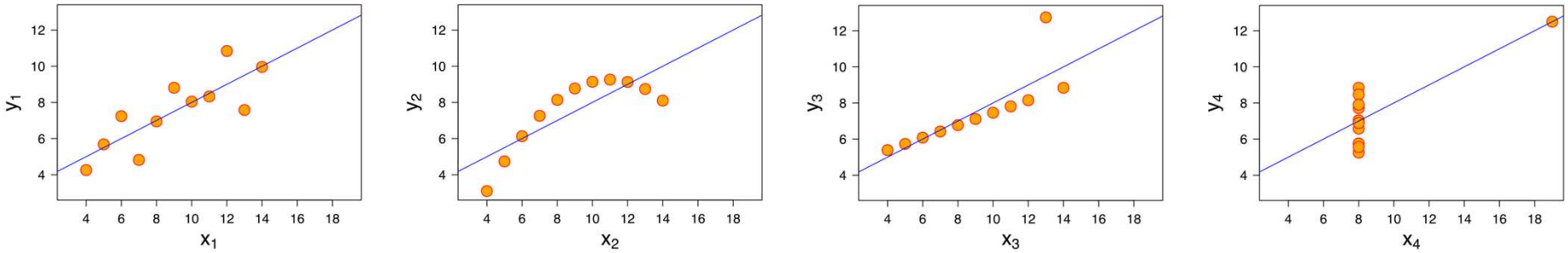
Machine Learning Algorithms

Source: "Neural Network for Porosity Prediction in Carbonate Formation", Huda F. Al-Saad, 2023

CUNY School of Professional Studies

Exploratory Data Analysis (EDA)

What do these data sets have in common?



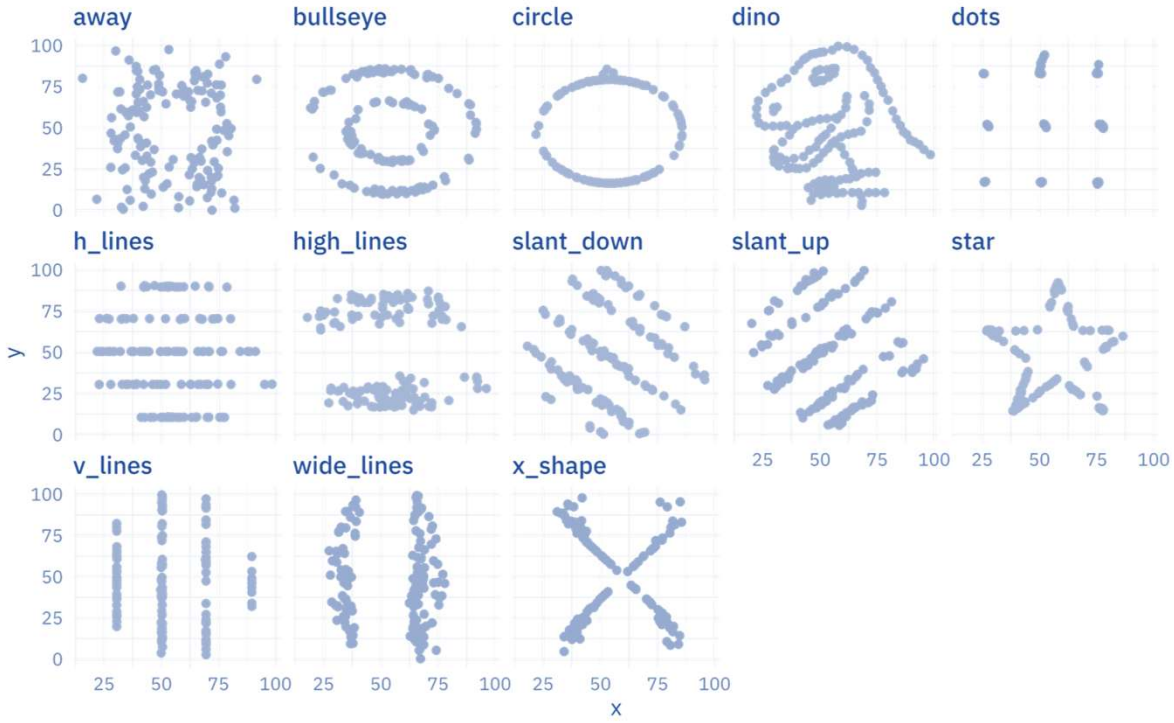
Not much. Except….

Identical means, variances, correlation, coefficients of determination, regressions

Property	Value
Mean of x	9
Sample variance of x: s_x^2	11
Mean of y	7.50
Sample variance of y: s_y^2	4.125
Correlation between x and y	0.816
Linear regression line	y = 3.00 + 0.500x
Coefficient of determination of the linear regression: R^2	0.67

Source: Anscombe's quartet from Wikipedia

Another example



Common statistical values for each group in the dataset

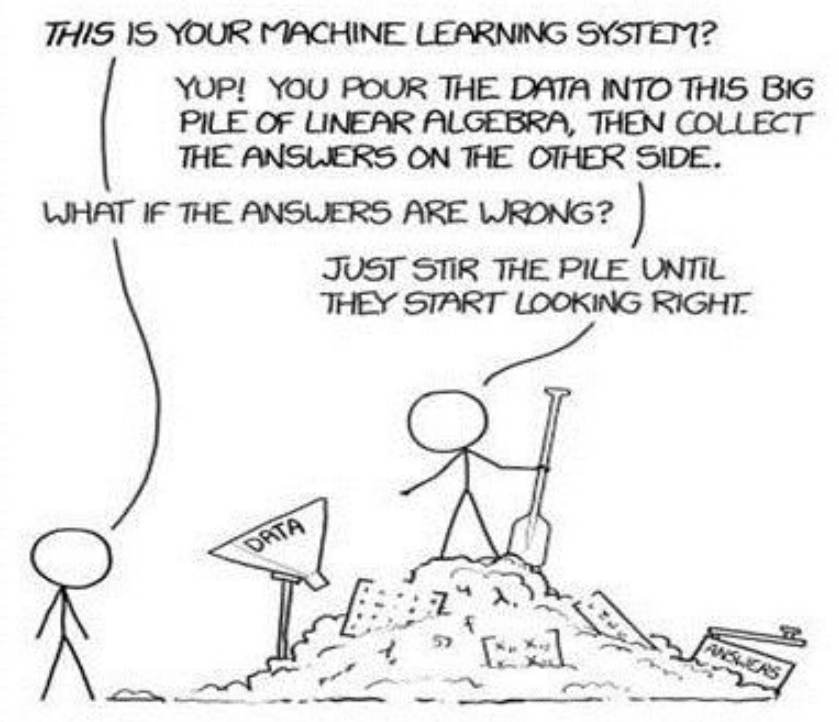
	Summary statistics					*Regression results*	
Dataset	Mean x	Mean y	Std Dev x	Std Dev y	Corr x y	Intercept	Coefficients
Away	54.27	47.83	16.77	26.94	-0.06	53.43	-0.10
Bullseye	54.27	47.83	16.77	26.94	-0.07	53.81	-0.11
Circle	54.27	47.84	16.76	26.93	-0.07	53.80	-0.11
Dino	54.26	47.83	16.77	26.94	-0.06	53.45	-0.10
Dots	54.26	47.84	16.77	26.93	-0.06	53.10	-0.10
H_lines	54.26	47.83	16.77	26.94	-0.06	53.21	-0.10
High_lines	54.27	47.84	16.77	26.94	-0.07	53.81	-0.11
Slant_down	54.27	47.84	16.77	26.94	-0.07	53.85	-0.11
Slant_up	54.27	47.83	16.77	26.94	-0.07	53.81	-0.11
Star	54.27	47.84	16.77	26.93	-0.06	53.33	-0.10
V_lines	54.27	47.84	16.77	26.94	-0.07	53.89	-0.11
Wide_lines	54.27	47.83	16.77	26.94	-0.07	53.63	-0.11
X_shape	54.26	47.84	16.77	26.93	-0.07	53.55	-0.11

Descriptive statistics can be misleading. Data visualization helps.

Source: Datasaurus Dozen from wikipedia

CUNY School of Professional Studies

What Exploratory Data Analysis isn't….



CUNY School of Professional Studies

Artificial General Intelligence (AGI)

What is Intelligence?

There are many definitions, here is one:

"Intelligence measures an agent's ability to achieve goals in a wide range of environments." S. Legg and M. Hutter

Narrow AI vs AGI

NARROW AI	ARTIFICIAL GENERAL INTELLIGENCE (AGI)
■ AI focused on a specific, singular or limited task	■ Not fully realized, with some developers questioning if it will be possible
■ Examples include image recognition, hyper-personalization, chatbots, predictive text	■ Seeks machines that can handle a range of cognitive tasks with little oversight
■ Trained on specific tasks by data scientists	■ The ability to learn, generalize, apply knowledge and plan for the future
■ Correlates questions or assignments to a specific data set to accomplish a task	■ Must consistently pass the Turing Test
■ No self-awareness, consciousness, ability to think	■ Single, general intelligence that possesses common sense and creativity and expresses emotions

Source: Definition from "A Collection of Definitions of Intelligence", Shane Legg Marcus Hutter, 2007, Image from Tech target under fair use