Rows - Datapoints
Columns - Headers

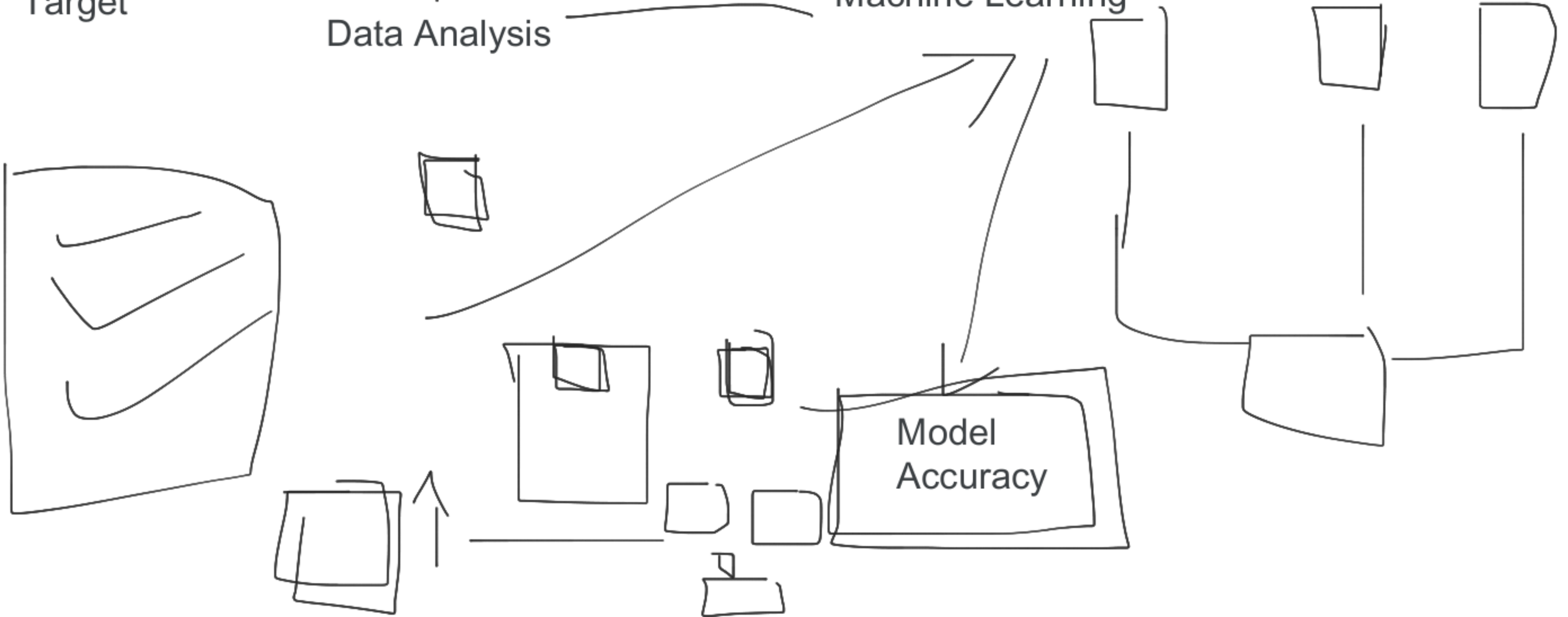
Feature
Target

Statistics

Data Analysis

Machine Learning

Data
Feature Engineering
Dimensionality Reduction
PCA
Hyper Parameter Tuning



1. Introduction of Statistics
a. Type of analysis
b. Types/Categories of Statistics - Descriptive & Infrential

2. Terminologies
a. Population (Mean & SD)
b. Sample (Mean & SD)
c. Variable
d. Parameter

3. Measure of Central Tendency
a. Mean
b. Median
c. Mode

4. Measure of Variance
a. Standard Deviation
b. Variance
c. Quartile Range

5. Random Variables & their Types

6. Pearson Correlation Coefficient

7. Co-variance

8. Spearman Rank Correlation

9. Probability
a. Addition Rule
b. Multiplication Rule
c. Combination
d. Permutation
e. Baye's Theorem

10. Hypothesis Testing
a. t-test
b. z-test
c. f-test (ANOVA)
d. Chi-Square
e. Type I & II errors
f. p-value, confidence

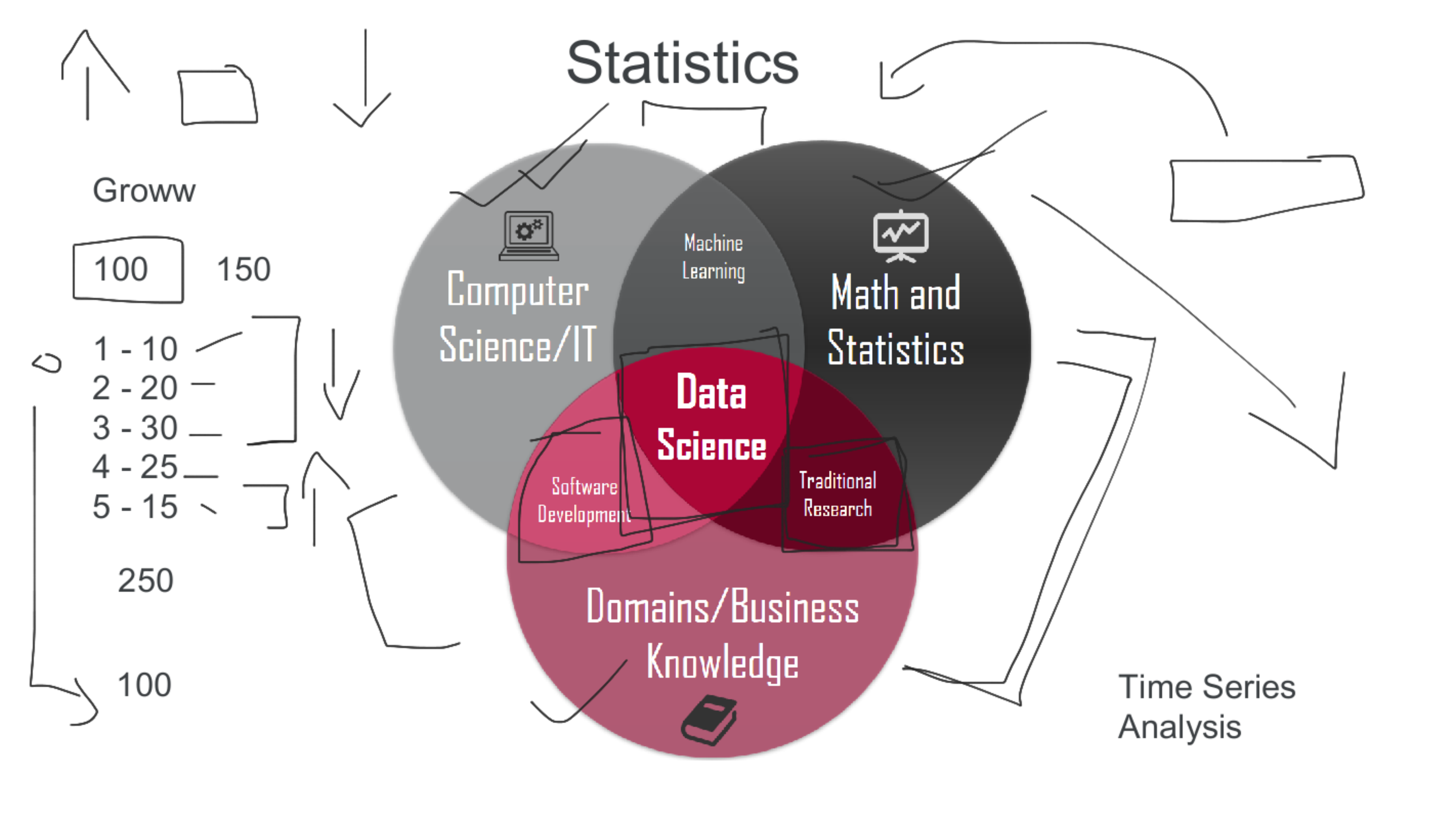
11. Central Limit Theorem

12. Skewness

13. Distributions
a. Normal/Gaussian
b. Binomial
c. Poissions'

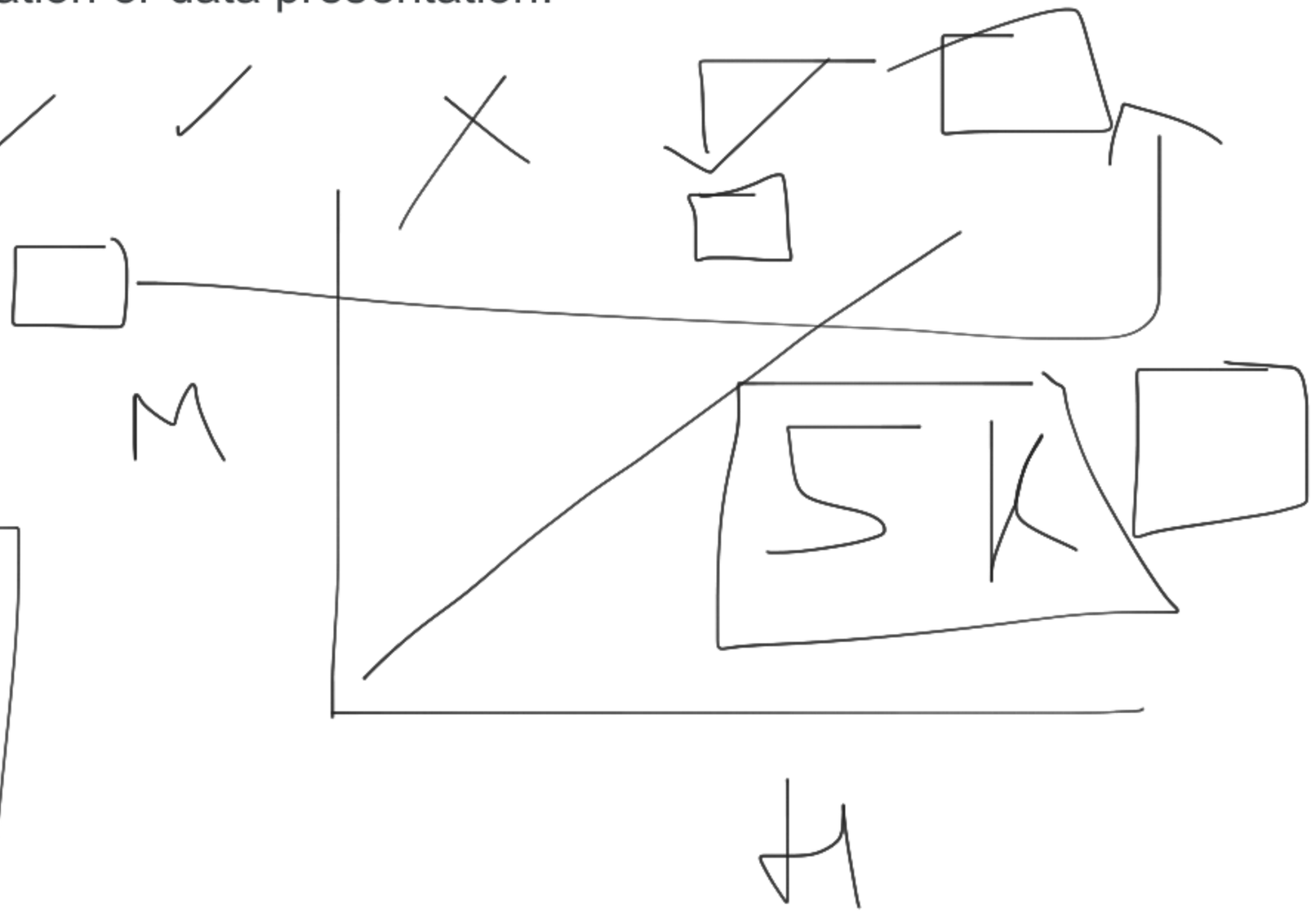
Extra Topics:

1. Variable Measurement Scales
a. Nominal
b. Ordinal
c. Interval
d. Ratio


Statistics
Groww
100
150
1 - 10
2 - 20
3 - 30
4 - 25
5 - 15
250
100
Computer Science/IT
Machine Learning
Math and Statistics
Data Science
Software Development
Traditional Research
Domains/Business Knowledge
Time Series Analysis

It is a mathematical science use to perform data collection, data analysis, data summarization, data organisation, data interpretation or data presentation.

Hr. of Study	Marks
1	10
2	20
3	30
4	40
5	50
6	60
7	70
8	80
9	90

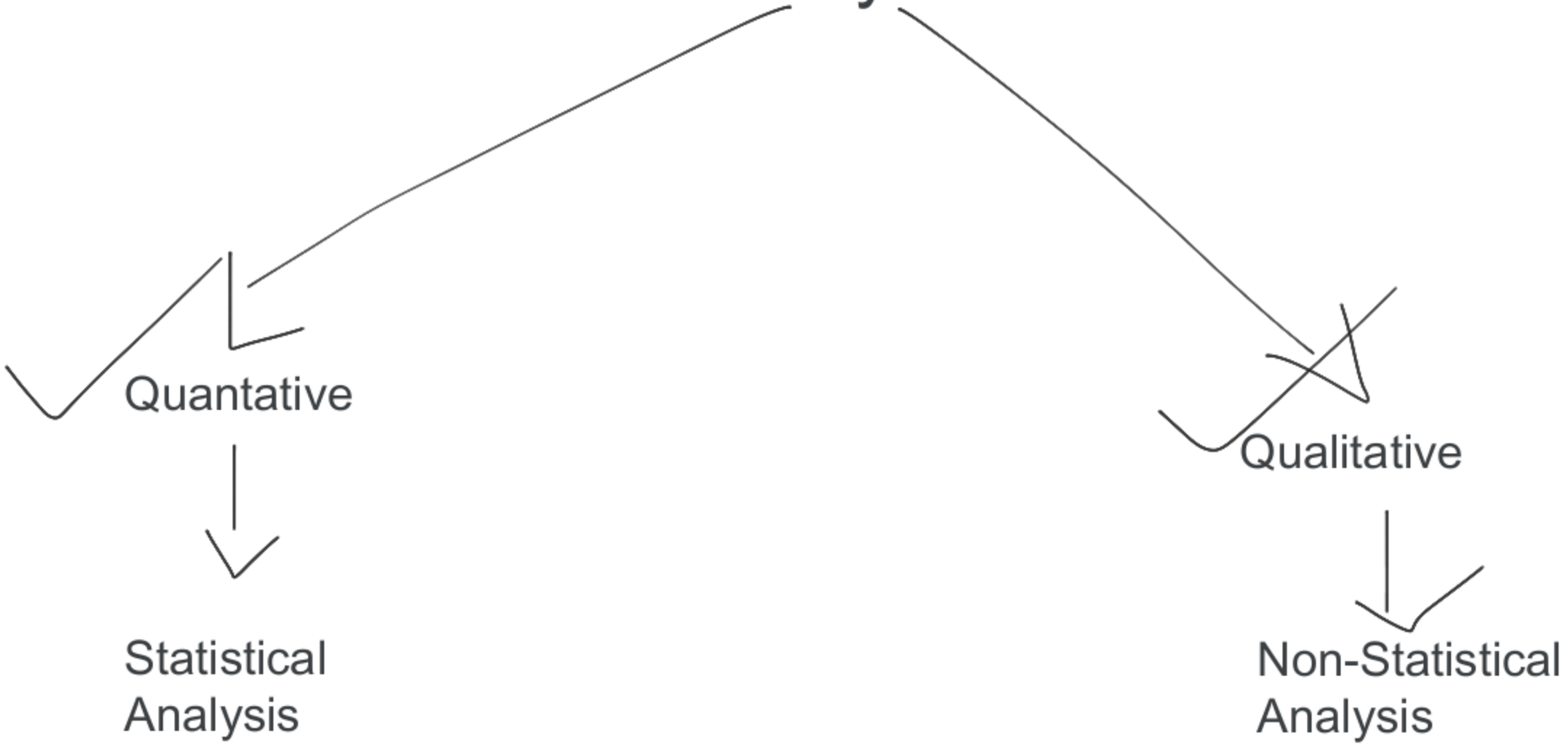


6 = ?

10 - ?

It is use to solve complex problems

Analysis



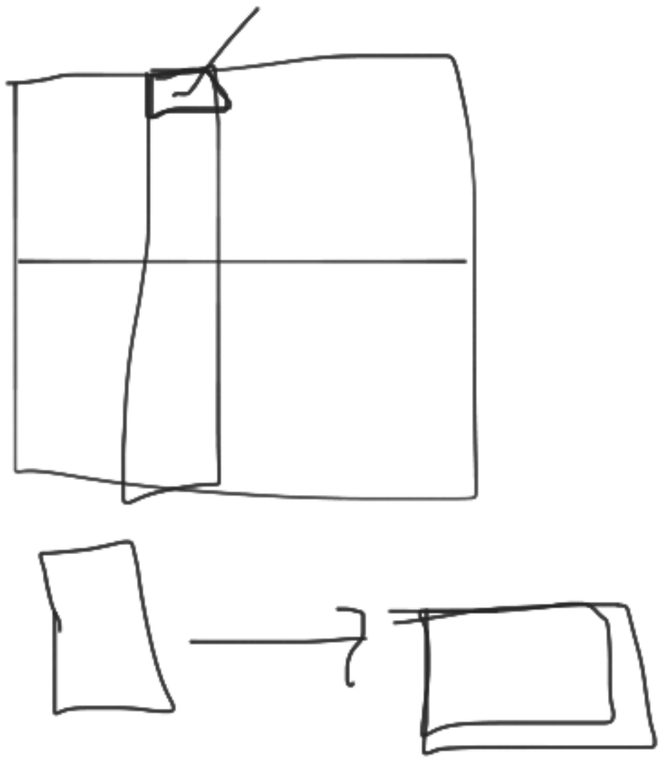
The shopkeeper sells 90 regular coffee per week.

Suppose, if you wanna purchase a cup of coffee, the cups that are available are in the following form - Small, Medium, Large

Categories of Stats

Mean
Median
Mode
SD
Variance

Probability
Hypothesis Testing
Distributions



Descriptive
Stats

It uses data to provide description

Population or Sample

Data
Analysis

Inferential
Stats

It uses data to provide description

Sample

Machine
Learning

Election for Class Monitor

→ Terminologies

Total Votes: 100

A, B, C

1. Population -> It is everything that being studied.
It is super set

2. Sample -> It is the portion or the subset of the population

A -> 25/100
B -> 35/100
C -> 30
NOTA -> 10

3. Variable -> It is a characterstics number, quantity that can be measure or counted

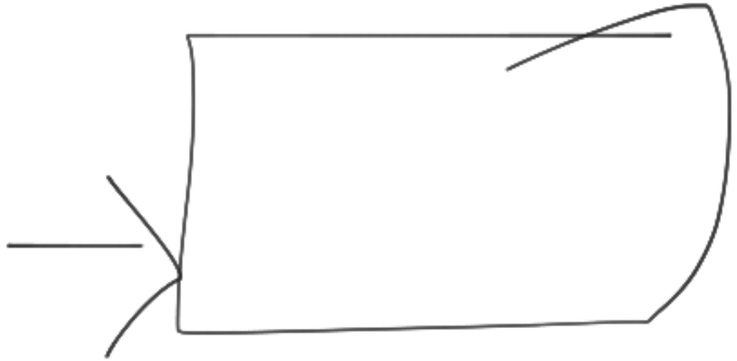
Votes of A -> 25

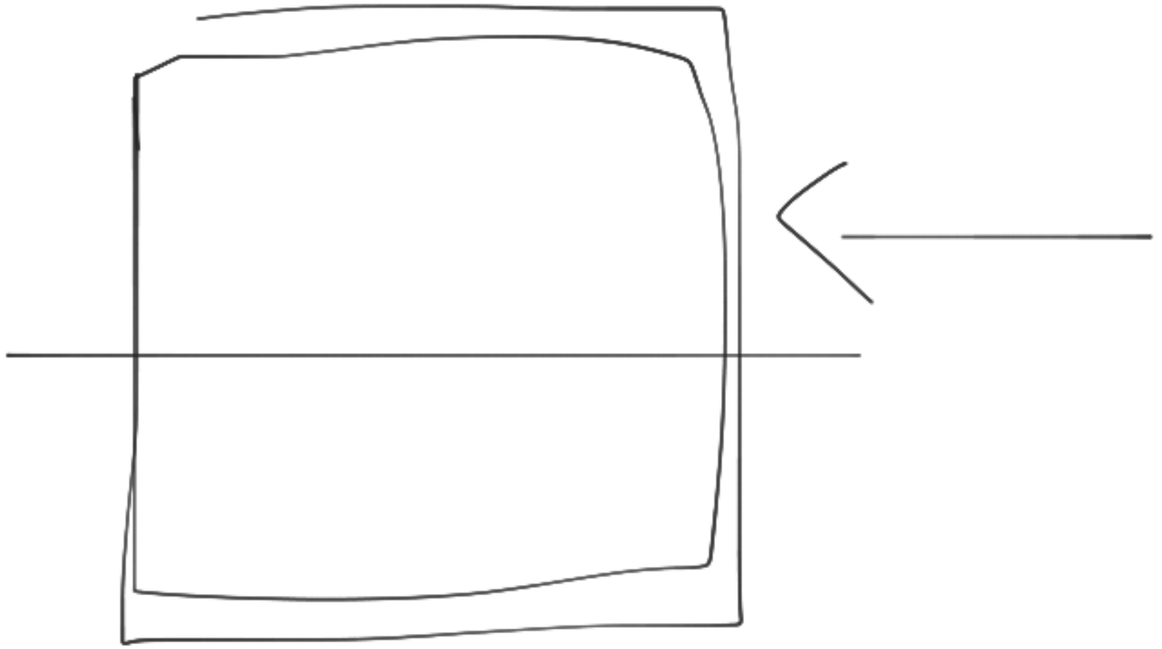
4. Parameter -> It is the characterstics of the population

Mean
Mode
Median

Q. You want to know the average cost of of statistics textbook. So, you have surveyed 25 textbooks.

Population -> All the stats textbooks
Sample -> 25
Parameter -> Average





Measure of Central Tendency

1,1,5,3,7,9,2,2,2,6,4

1. Mean -> It is a measure of average of all the values

$$\frac{\text{sum of all the values}}{\text{total number of values}} = \frac{42}{11} = 3.8$$

1,1,2,2,2,3,4,4,5,6,7,9

$$\frac{3+4}{2} = \frac{7}{2}$$

2. Median ->It is a measure of the central value

a. The values should be sorted in the ascending order
b. Choose the middle value
b(1) -> Odd -> 3
b(2) -> Even -> 3.5

3. Mode -> It is a measure of the values that occur more frequently -> 2