STATISTICS

COMPUTER SCIENCE

lec.1 notes

Statistics

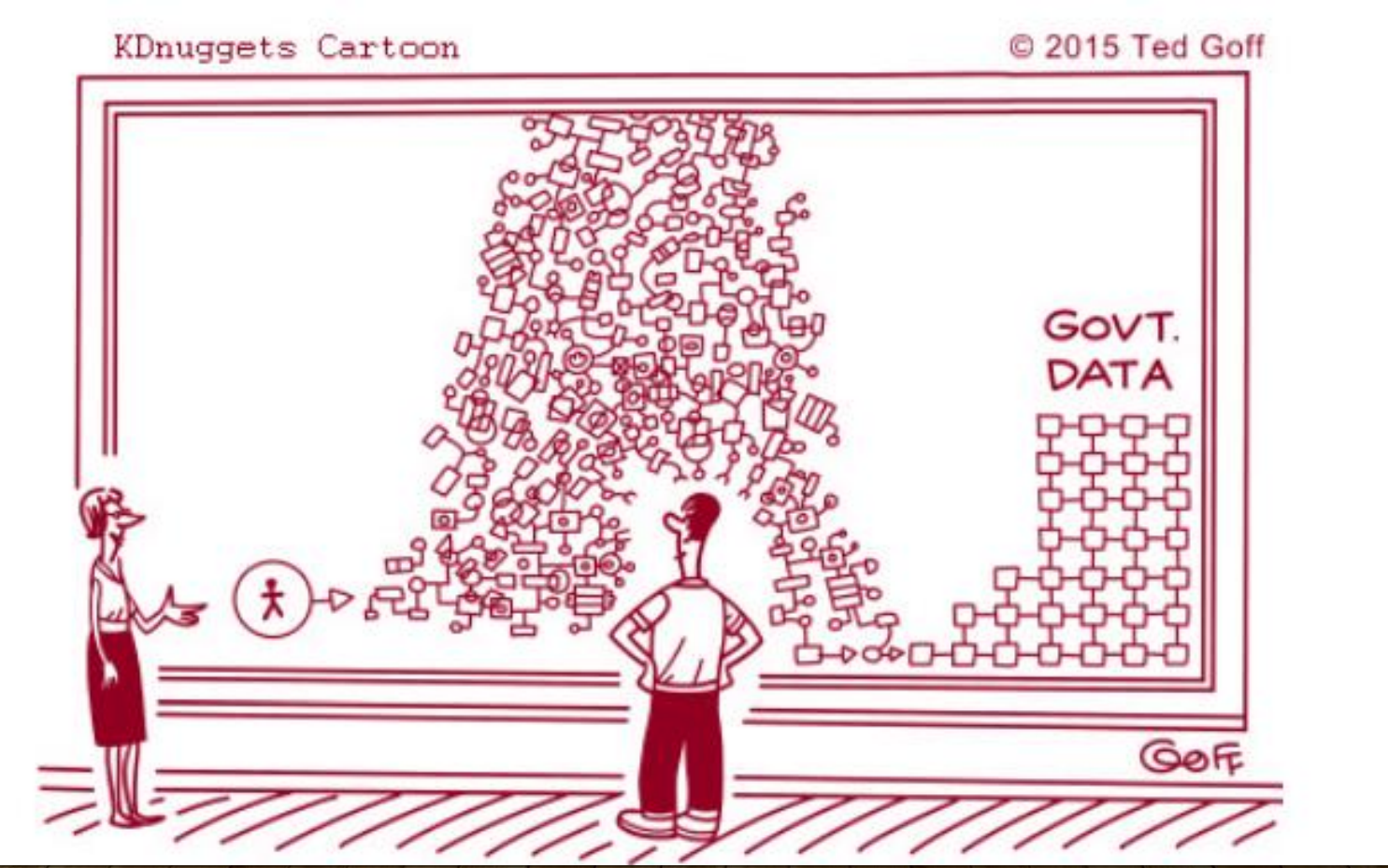
WHY STATISTICS

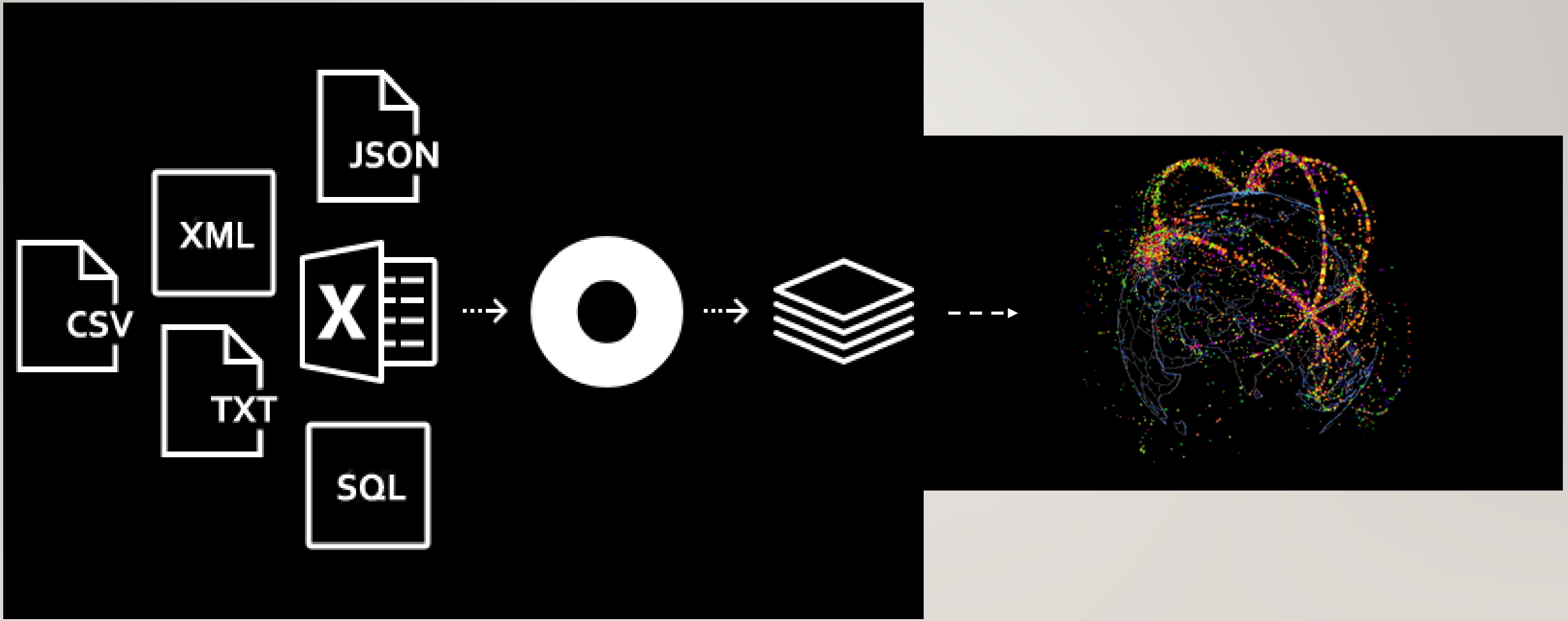
Statistics Limitations

"STATISTICS IS THE SCIENCE OF COLLECTING ,ORGANIZING PRESENTING, ANALYZING, AND INTERPRETING NUMERICAL DATA"

"STATISTICS IS A WAY TO GET INFORMATION FROM DATA."

BEFORE TRYING ANYTHING, YOU SHOULD KNOW YOUR DATA!

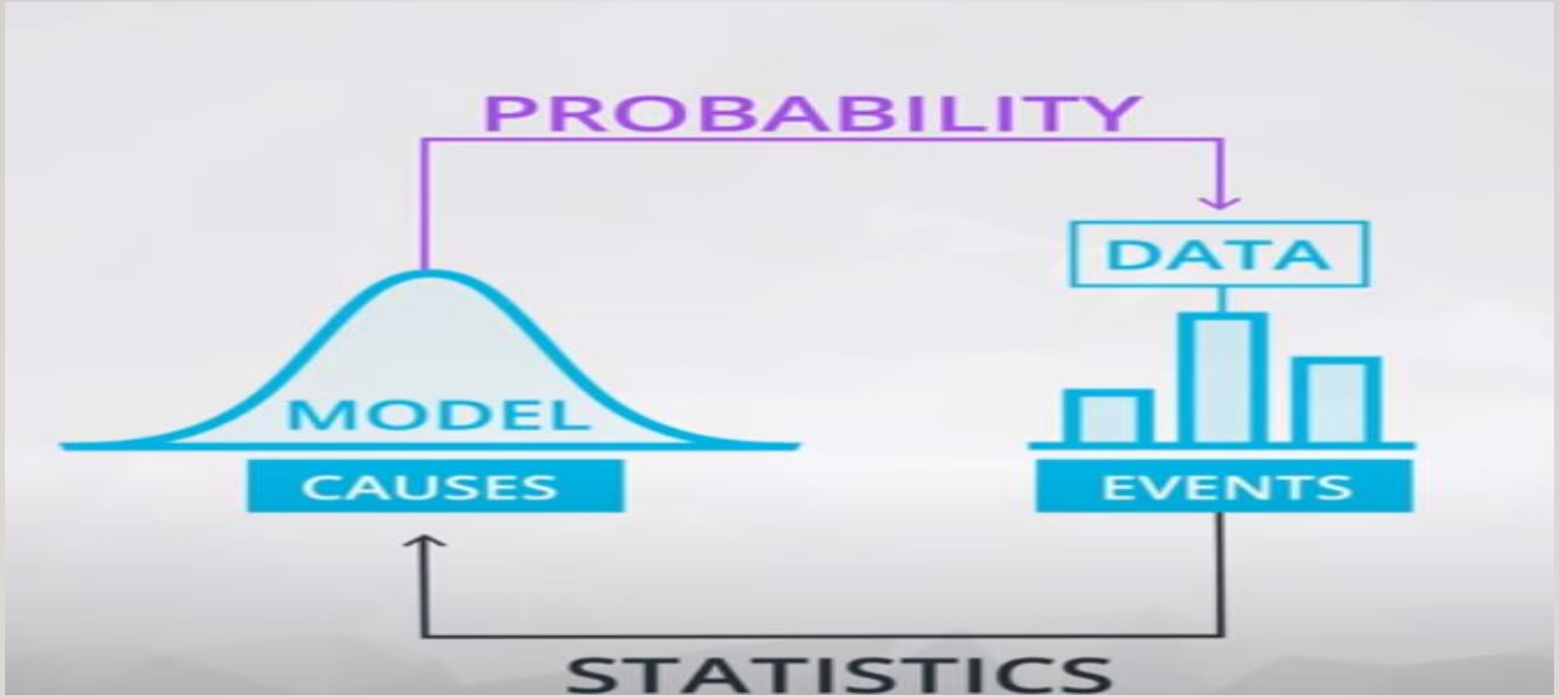



JSON
XML
CSV
TXT
SQL

COURSE CONTENT

“STATISTICS IS THE SCIENCE OF COLLECTING ,ORGANIZING PRESENTING, ANALYZING, AND INTERPRETING NUMERICAL DATA”

Probability Vs. Statistics



Probability: we are given description of causes & we are trying to predict data / future events

Statistics: we are given data & we are trying to infer possible causes

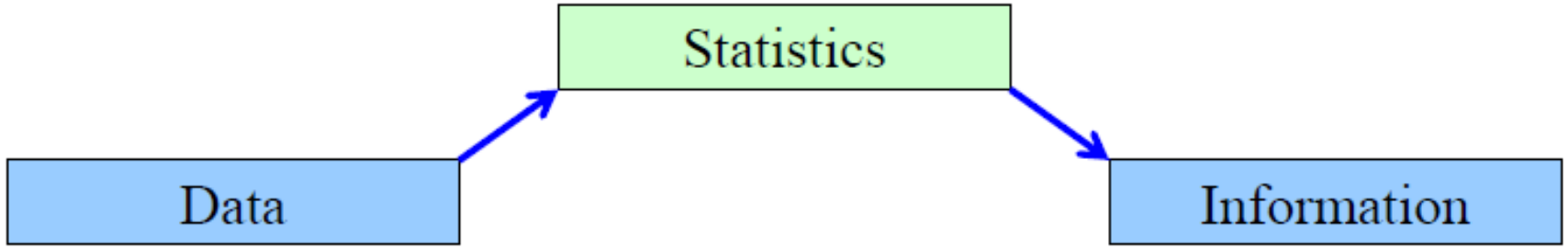
Probability is used to make inferences from samples to populations

Inferential Statistics uses Probability:

The chance of an event occurring

A measure of the likelihood that an event in the future will happen

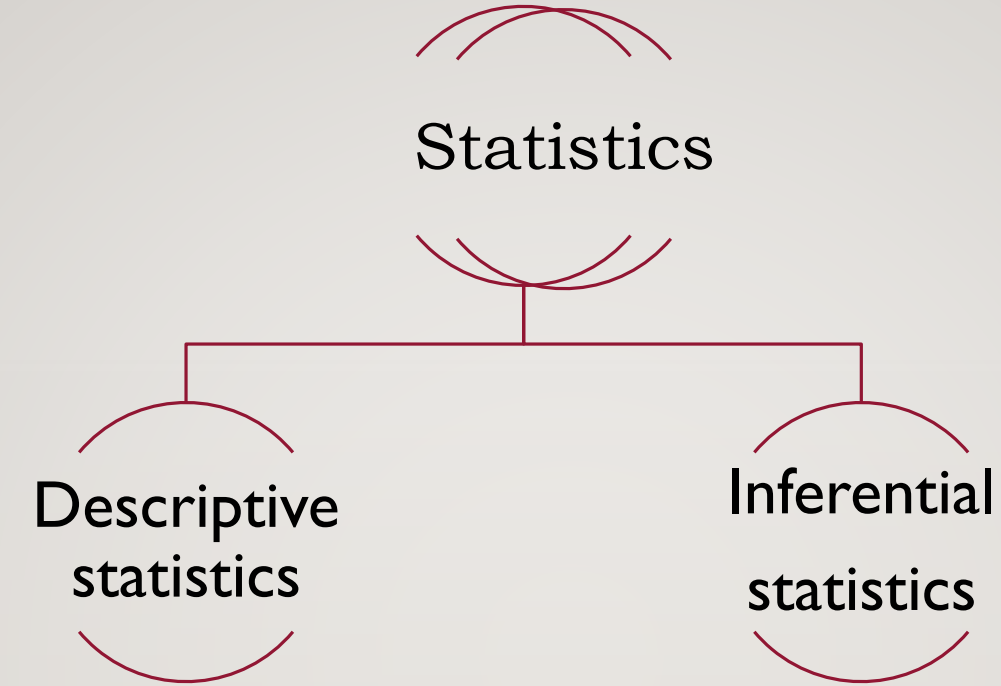
"Statistics is a way to get information from data"



Data: Facts, especially numerical facts, collected together for reference or information.

Information: Knowledge communicated concerning some particular fact.

Statistics is a ***tool*** for creating ***new understanding*** from a set of numbers.



Set of procedures for organizing and summarizing information

Generalizations and extract conclusions about population

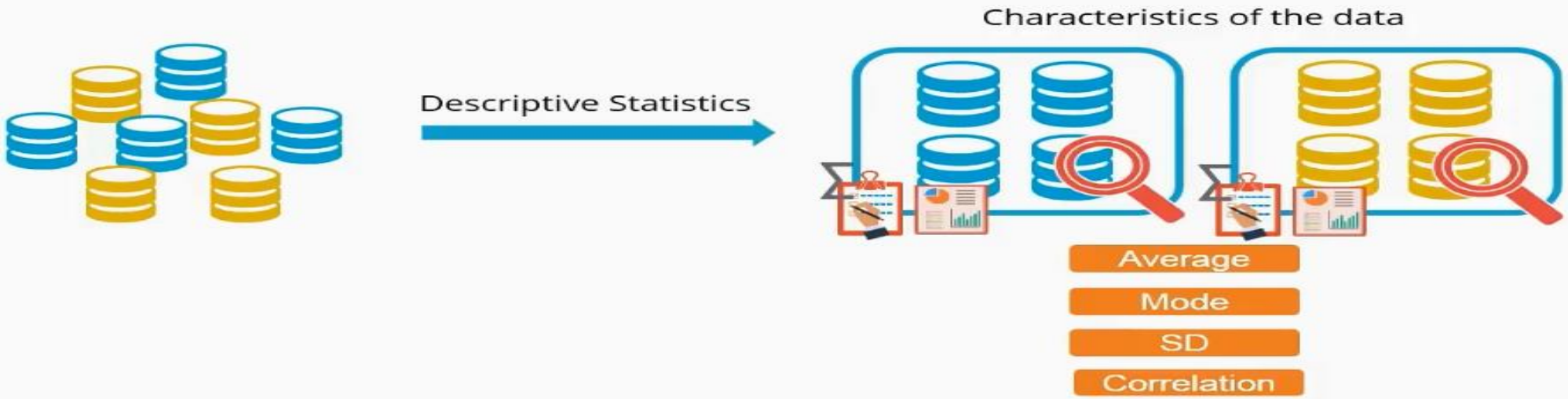
Statistical Terms

Variable

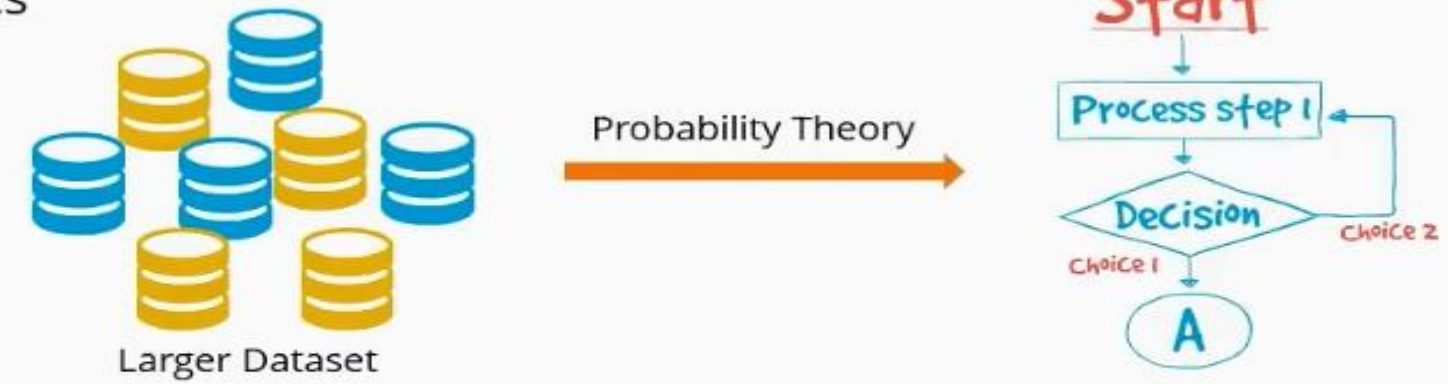
A variable is a feature characteristic of any member of a population differing in quality or quantity from another member.

Statistical Data Type
According to variable type

Descriptive Statistics helps organize data and focuses on the main characteristics of the data. It provides a summary of the data numerically or graphically.



Inferential Statistics



Modeling allows you to develop mathematical equations which describe the interrelationships between two or more variables.

Statistical Terms

Population

A population is the group from which data is to be collected.

ALL

Population

Sample

A sample is a subset of a population.

SELECTED

Population

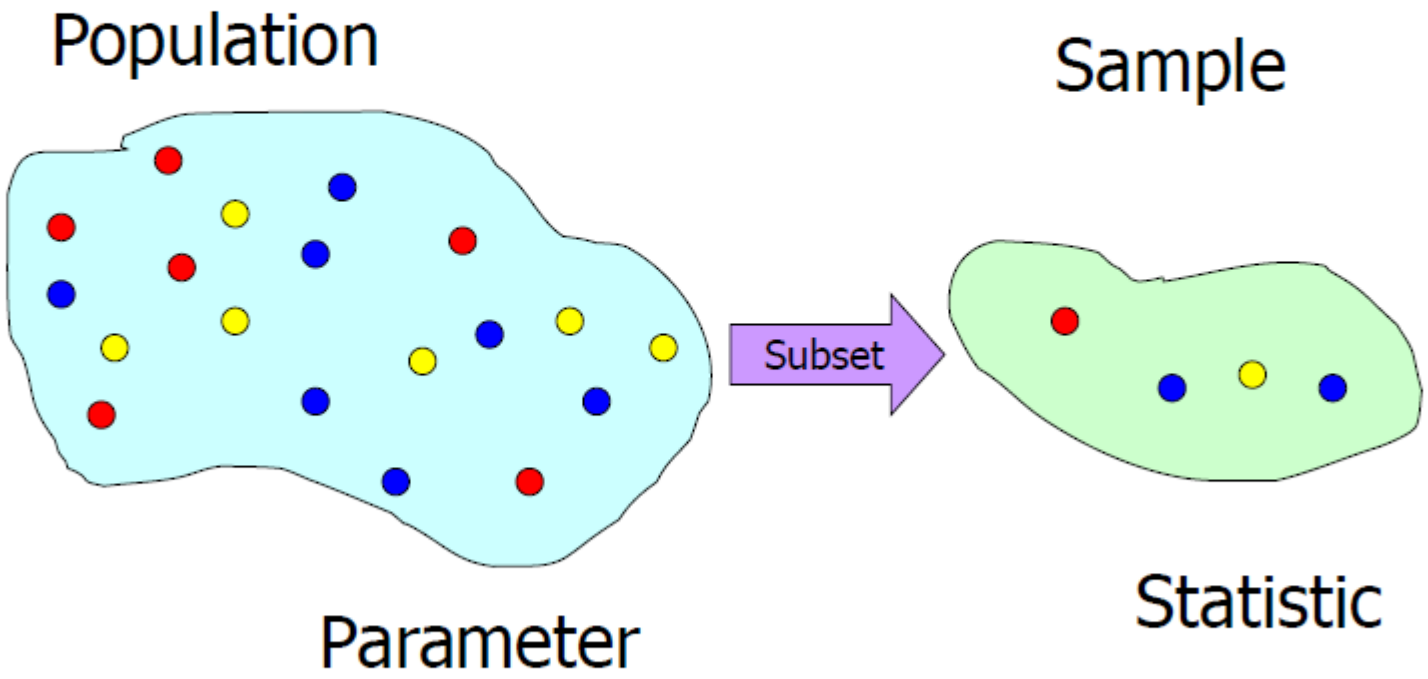
Sample

Parameter

A descriptive measure of a ***population.***

Statistic

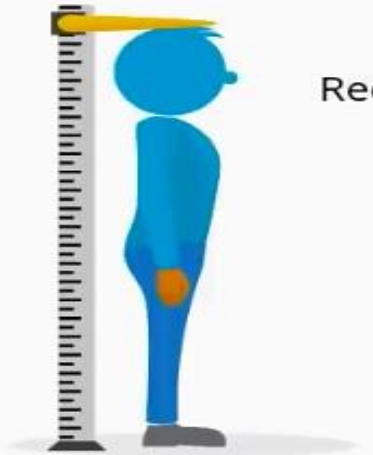
A descriptive measure of a ***sample.***



Populations have Parameters,

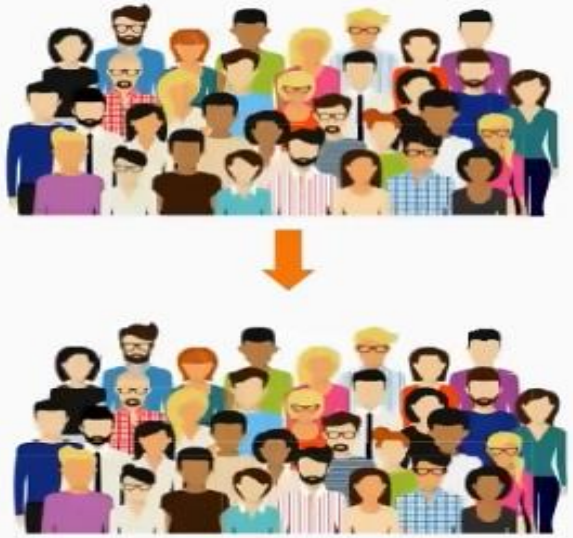
Samples have Statistics.

In the Descriptive Statistics method:



Record the Height

- Maximum Height
- Minimum Height
- Average Height



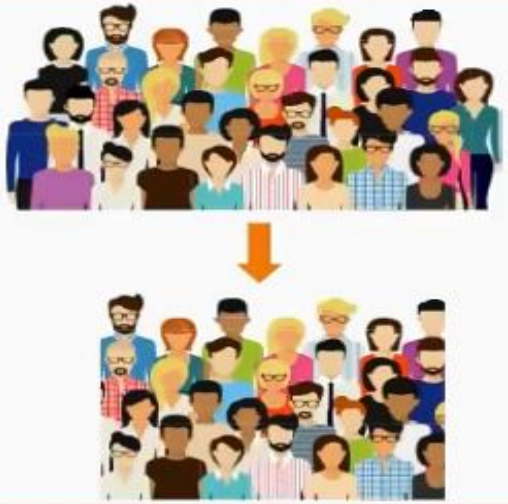
Inferential Statistics

It allows one to infer population parameters based on sample statistics and to model relationships within the data.

In the Inferential Statistics method:

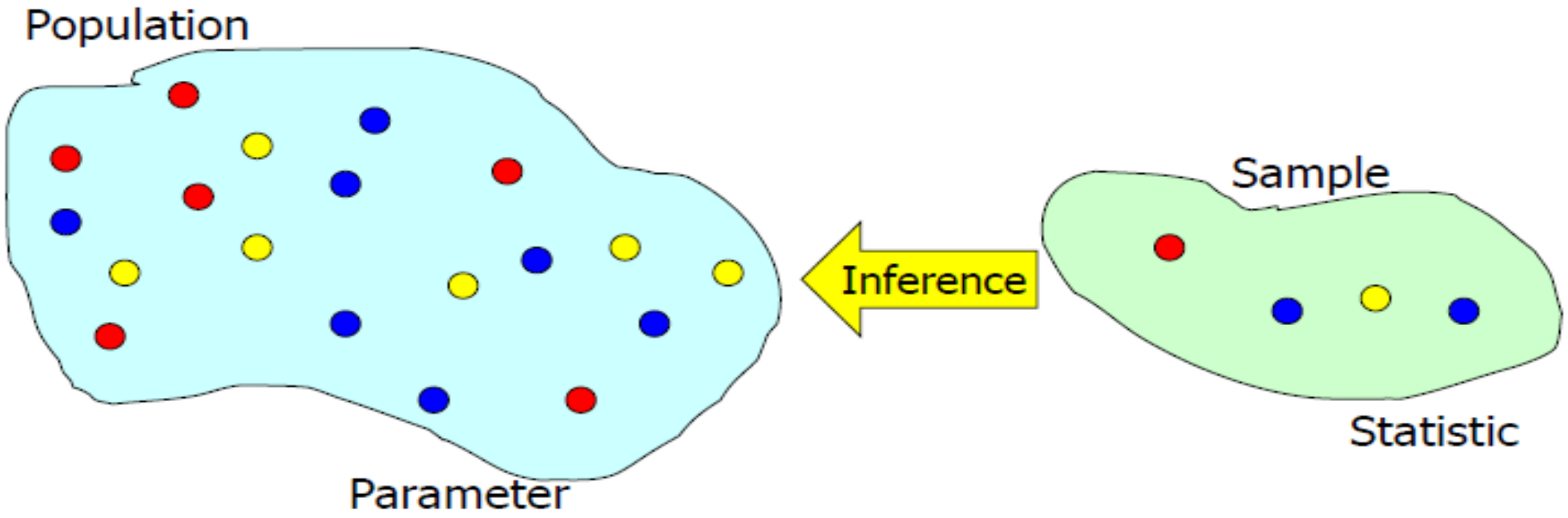
Height

- Tall
- Medium
- Small

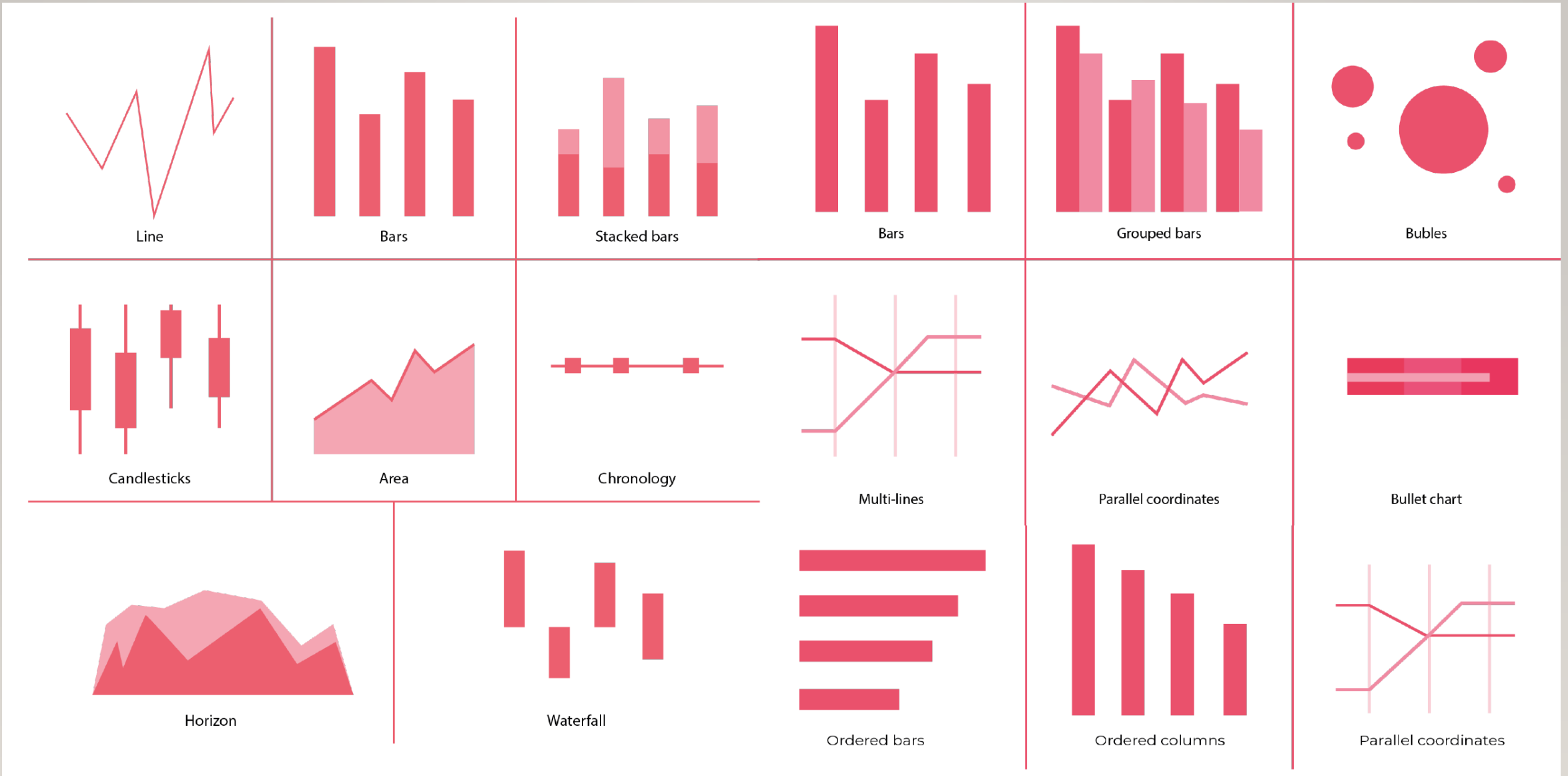


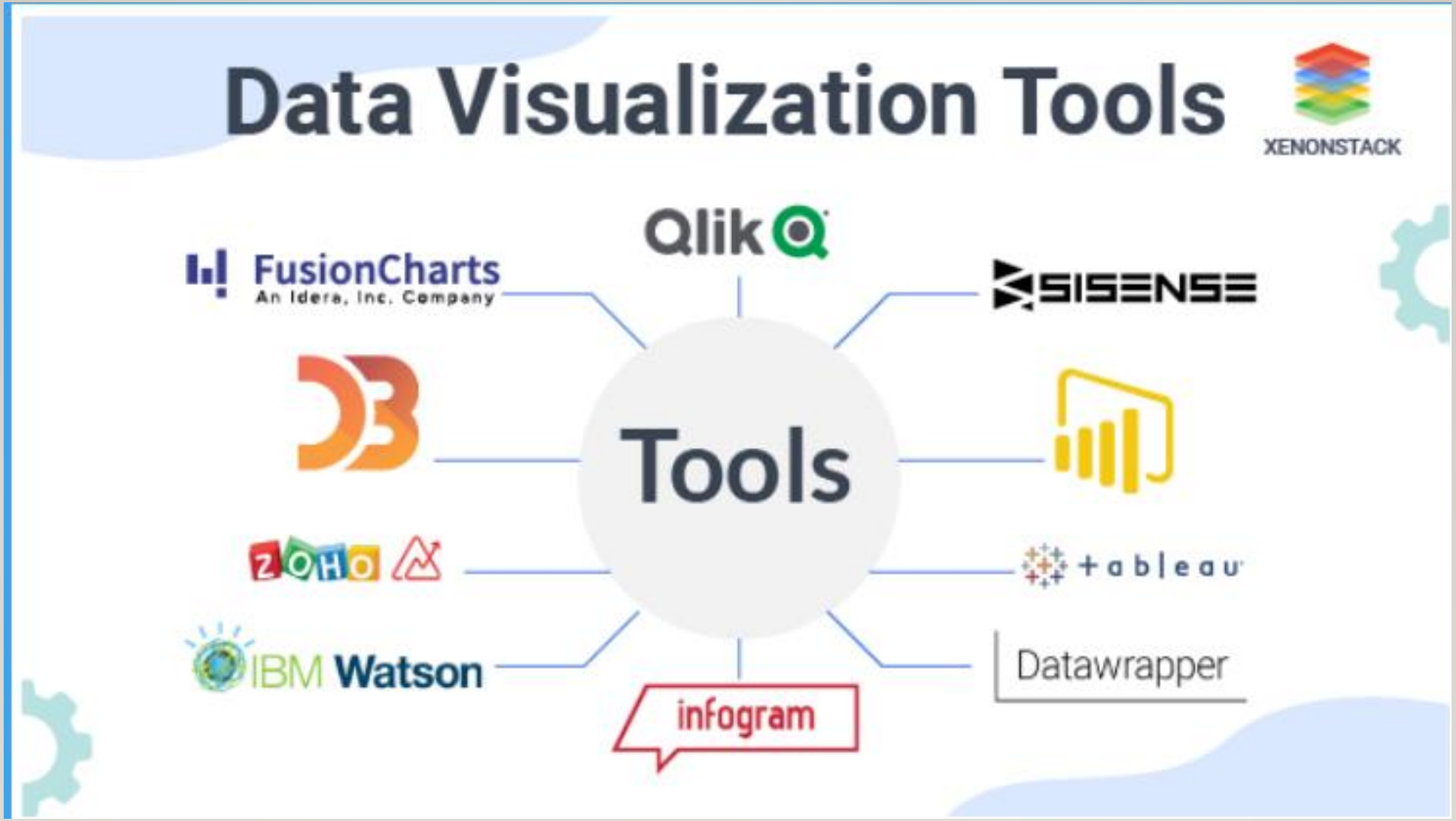
Statistical Inference…

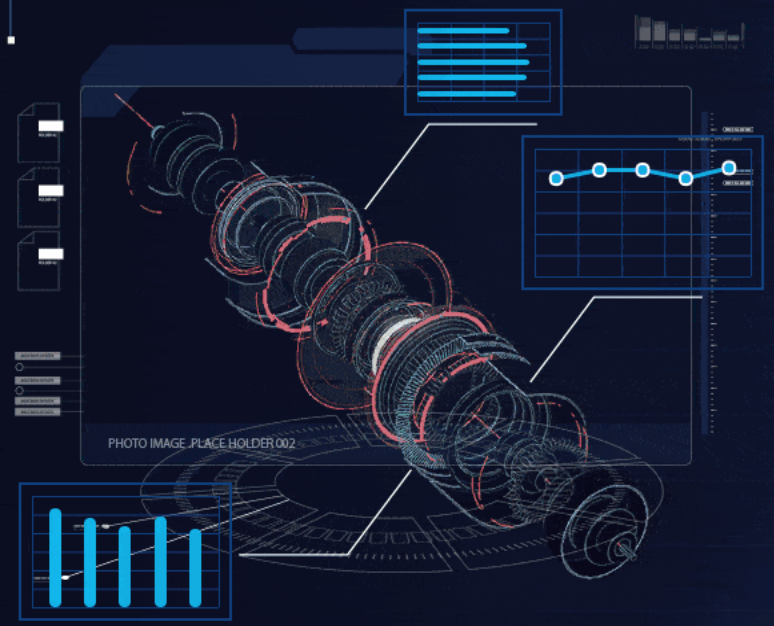
Statistical inference is the ***process*** of making an estimate, prediction, or decision about a population based on a sample.



What can we ***infer*** about a Population's Parameters based on a Sample's Statistics?


Line
Bars
Stacked bars
Bars
Grouped bars
Bubles
Candlesticks
Area
Chronology
Multi-lines
Parallel coordinates
Bullet chart
Horizon
Waterfall
Ordered bars
Ordered columns
Parallel coordinates


Data Visualization Tools
XENONSTACK
Qlik
FusionCharts
An Idera, Inc. Company
SISENSE
Tools
ZOHO
+ableau
IBM Watson
Datawrapper
infogram



Dasbate1

Dasbate2

按下 F11 即可結束全螢幕模式

Dasbate3

Dasbate4

1-A	1-B	2-A	2-B
1560	2660	1400	2700
3-A	3-B	4-A	4-B
3665	4235	890	1400

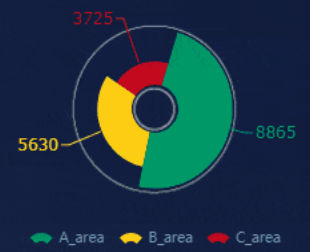
index	A_series	B_series	C_series	D_series
Current Value	12220	5855	12920	14200
Year-on-year	22%	14%	26%	32%
Early period	10%	5%	12%	22%

Distribution of sales amount

Sales Volume

User Statistics

Users	256	unit
Market share	26	%
Total sales	9163	set
Sales Amount	45195	10M



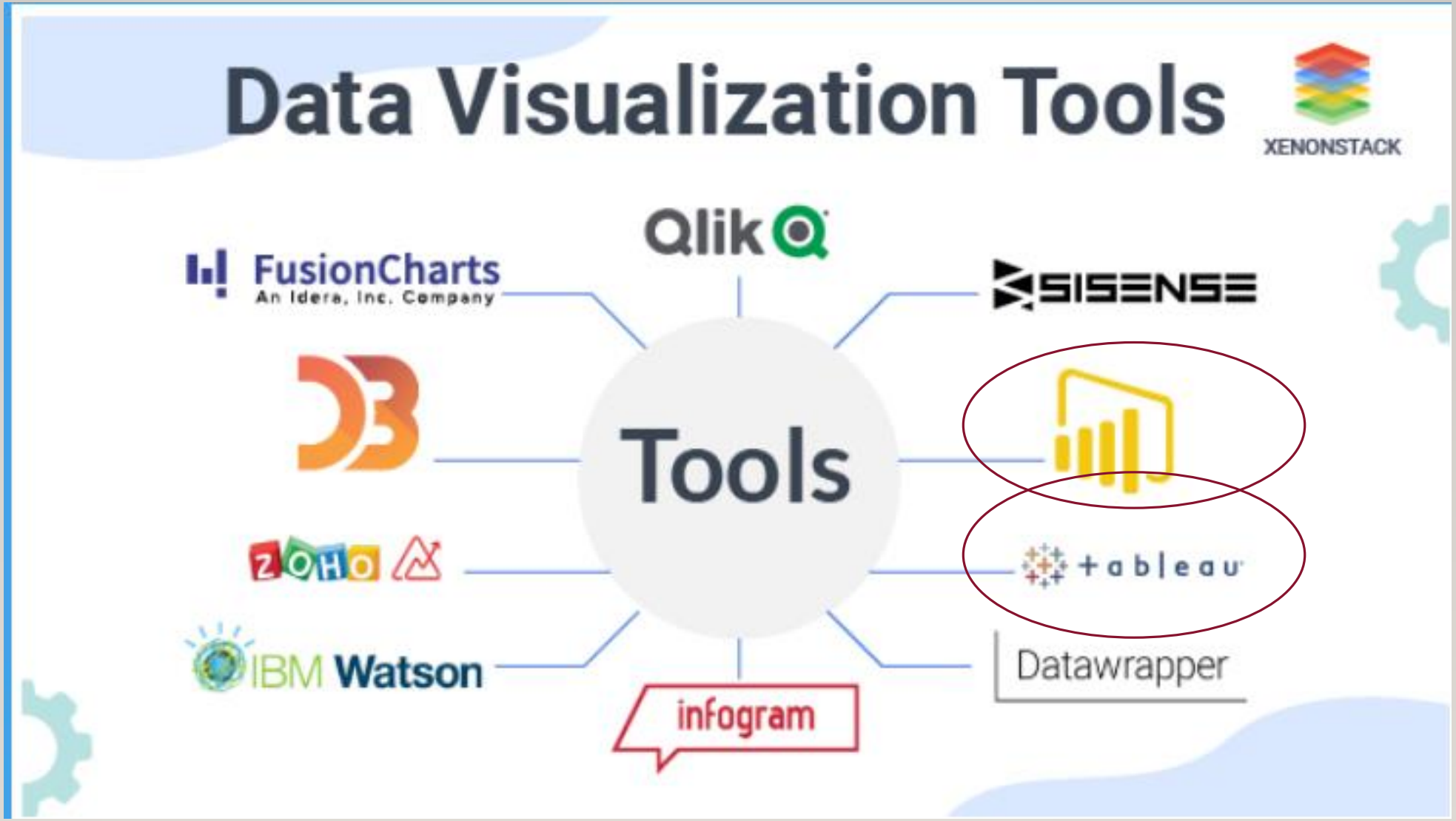
PV RMS AKV SD CRF IMF SHF

Market Analysis

Your task now

Anaconda 3
Jupyter Notebook

Visualizing Data Tools


Data Visualization Tools
XENONSTACK
Qlik
FusionCharts
An Idera, Inc. Company
SISENSE
Tools
ZOHO
+ableau
IBM Watson
Datawrapper
infogram