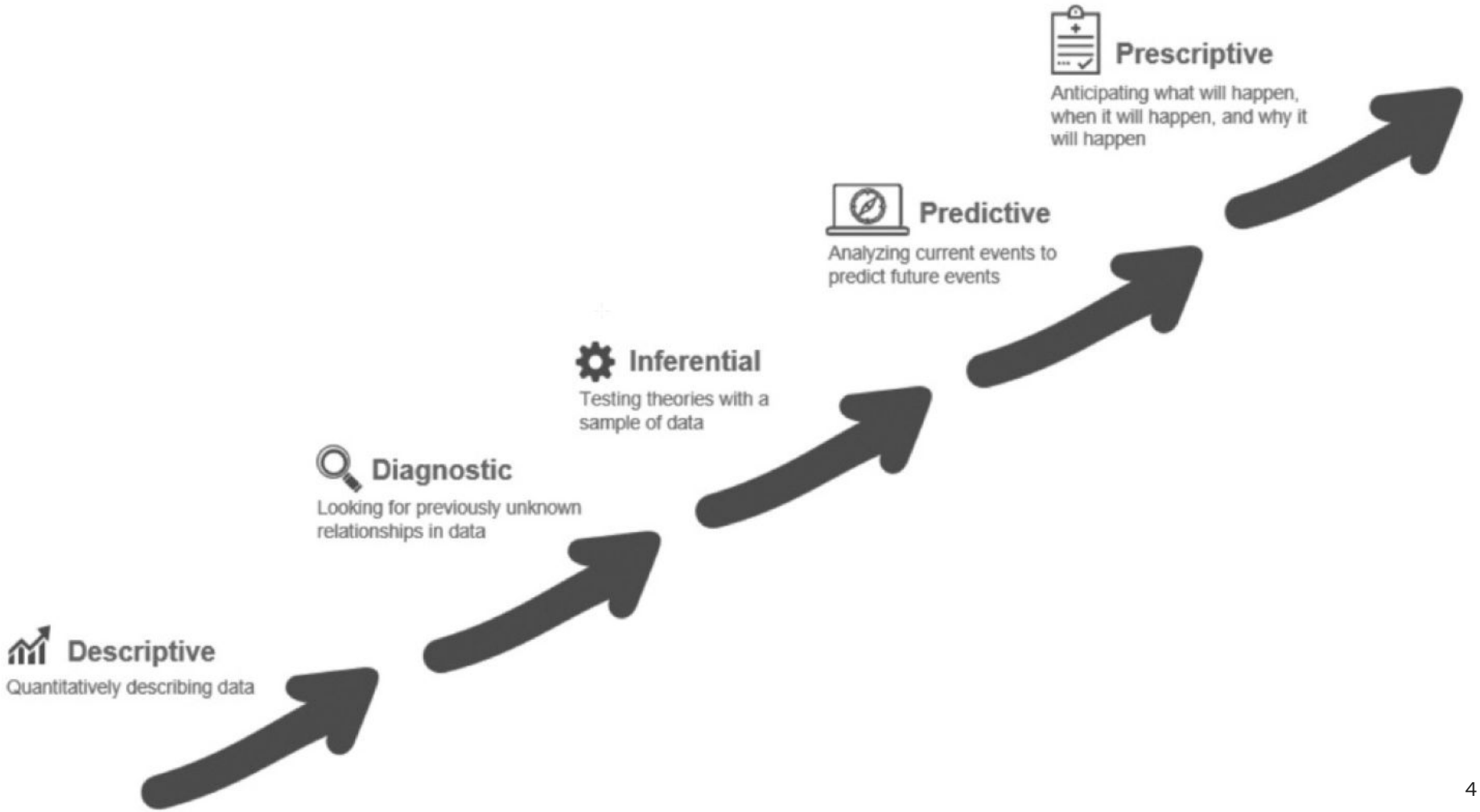
ISM3545C

Professor Ernesto Lee

Introduction to Analytics

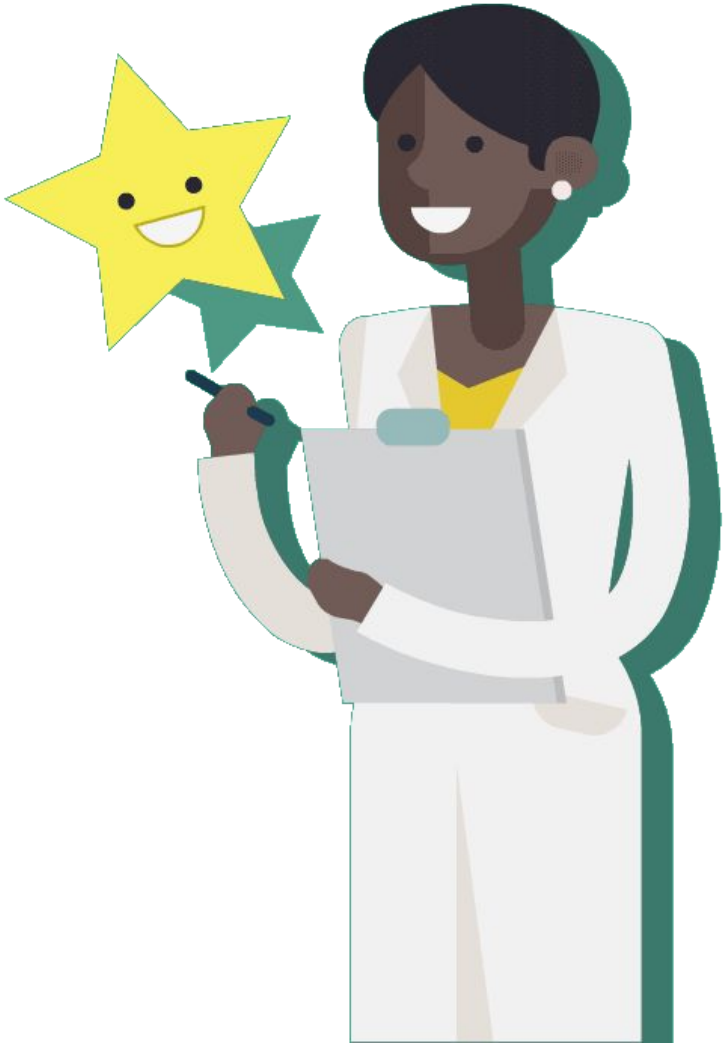
What is Analytics?

Analytics can be defined as the process of transforming raw data into insights.


Descriptive
Quantitatively describing data
Diagnostic
Looking for previously unknown relationships in data
Inferential
Testing theories with a sample of data
Predictive
Analyzing current events to predict future events
Prescriptive
Anticipating what will happen, when it will happen, and why it will happen

Analogy (for Analytics)

- Descriptive
- Diagnostic
- Inferential
- Predictive
- Prescriptive



We are going to study Descriptive and Diagnostic Analytics

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Correlation / Covariance

Terms

Outlier. An outlier is a number that lies outside of most of the other numbers in a set.

Mean. The mean, also referred to as the arithmetic mean, is the average of all the numbers in a given set of numbers.

Median. The median is defined as the middle of a sorted list of numbers in a set. The median is a better description of the data if there are lots of outliers.

Mode. The mode is defined as the most common number in a set of numbers.

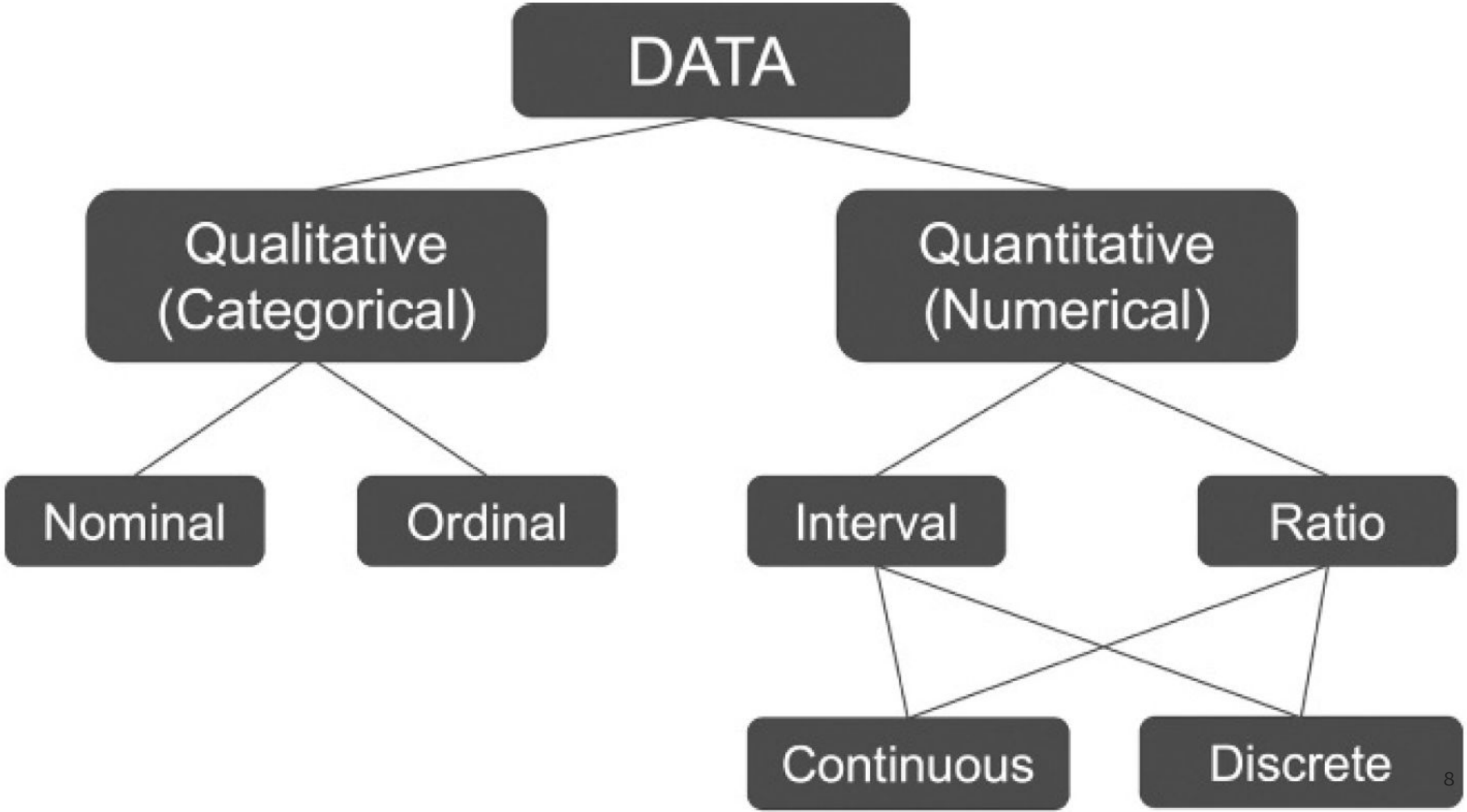
Variable. A variable is a characteristic, number, or quantity that can be measured. In a general sense, it can be seen as a column of data.

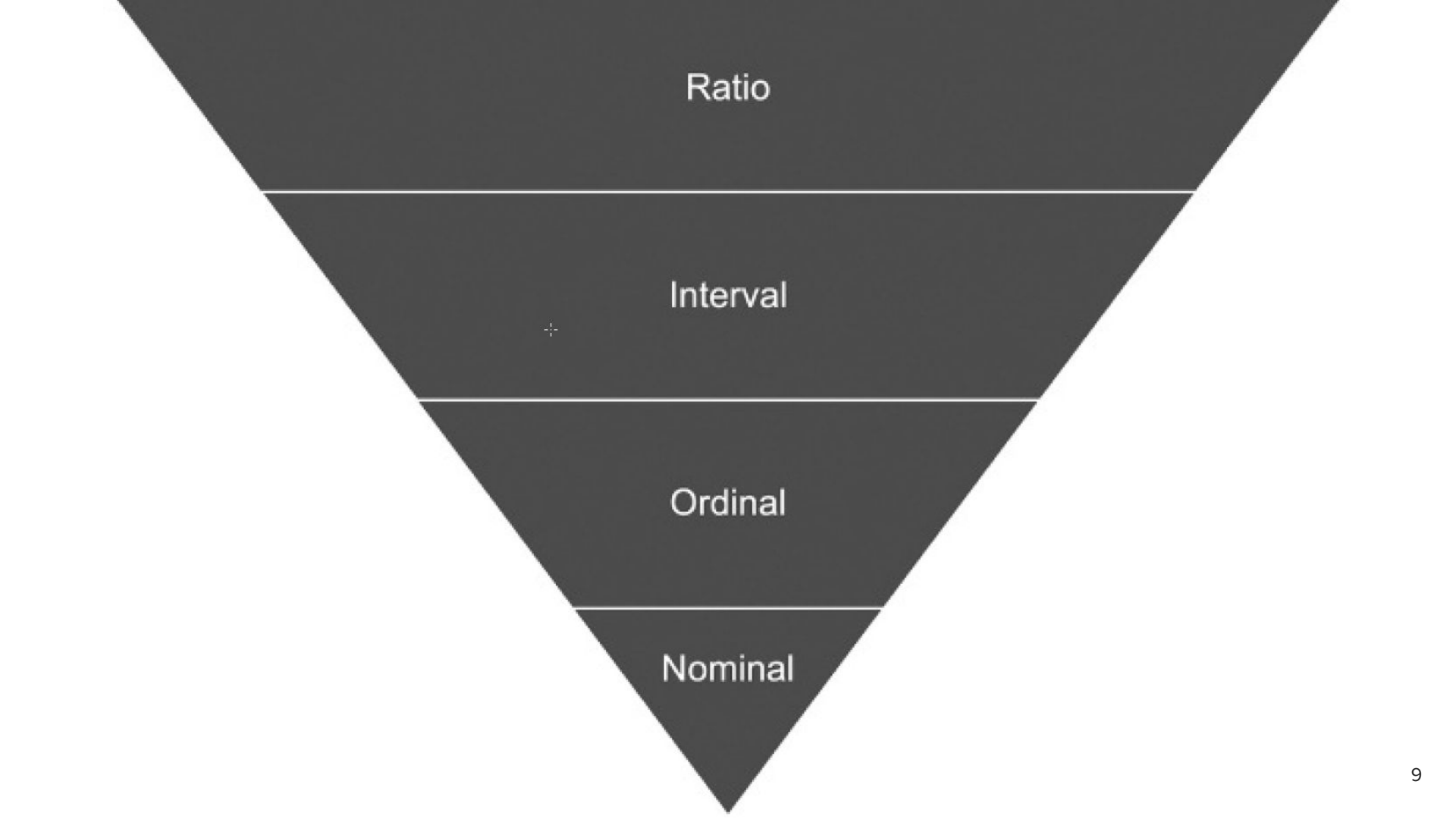
Quantitative data. Quantitative data, also called numerical data, is numerical information that can be measured or counted.

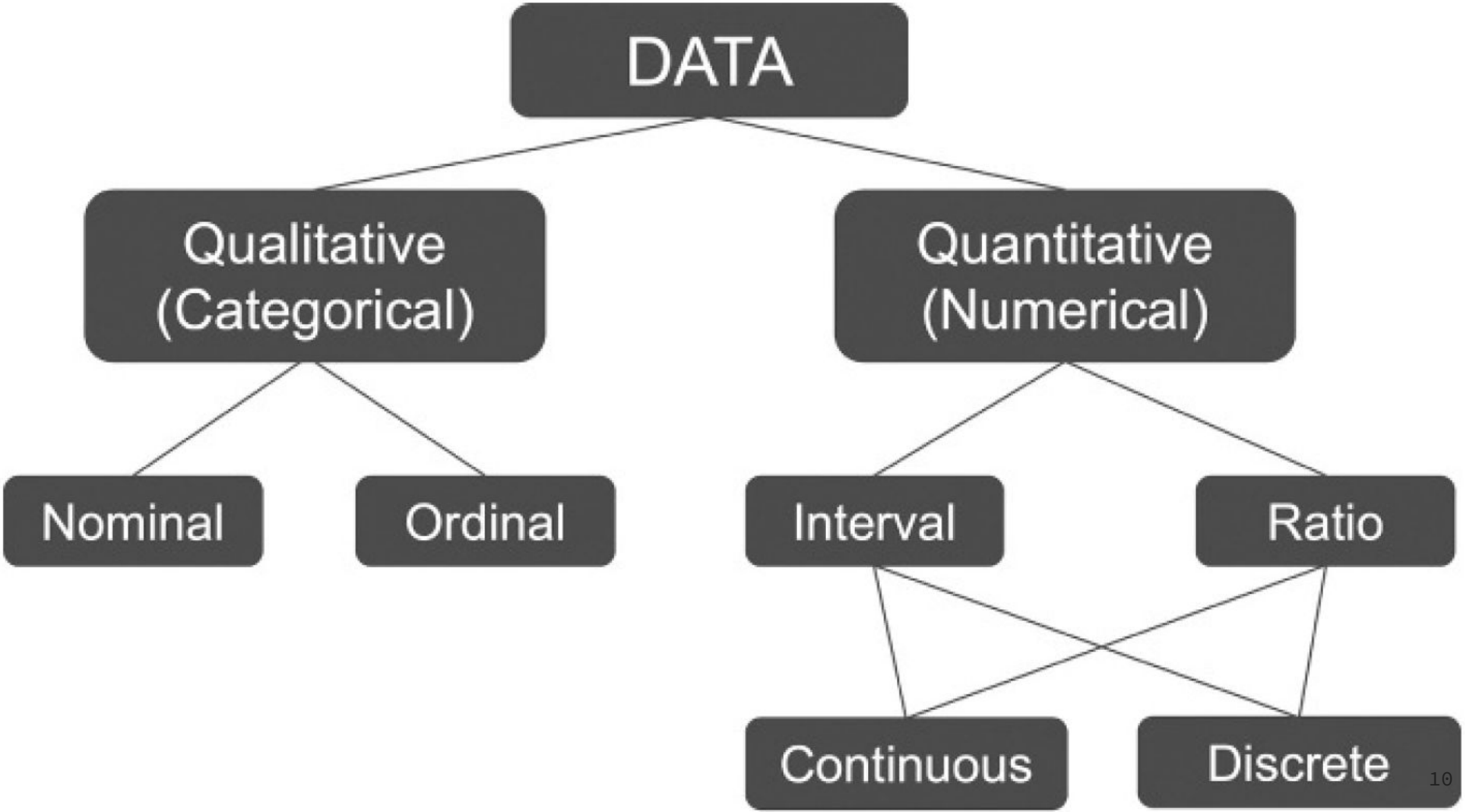
Qualitative data. Qualitative data, also called categorical data, is descriptive information about characteristics that are difficult to define or measure or that cannot be expressed numerically. Qualitative data can be put into categories like gender and sales regions.

Raw data. Raw data is data that is collected from a source and not cleaned, transformed, or edited in any way.

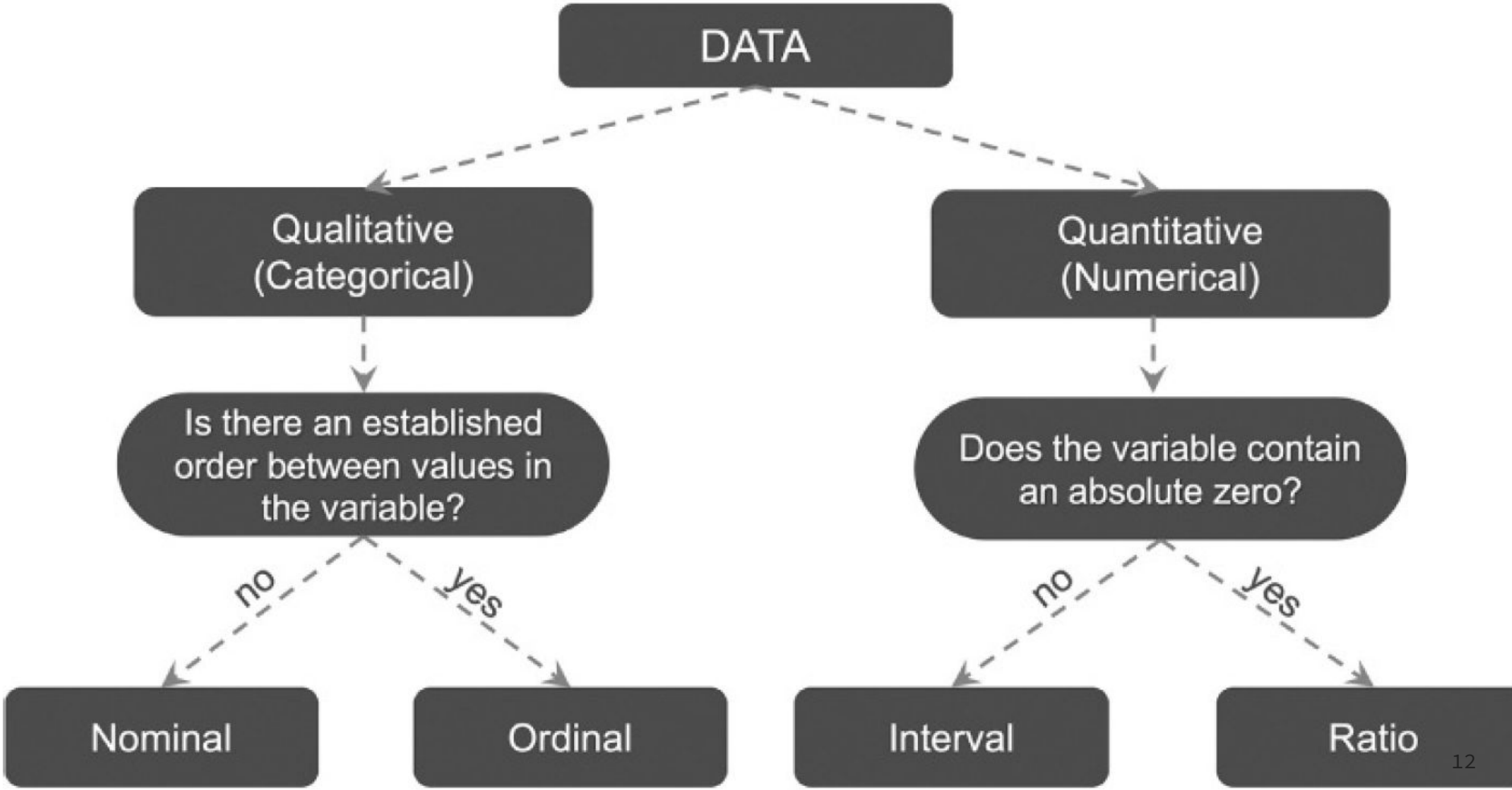
Aggregate data. Aggregate data is data that has been summarized using an aggregation (a cluster of things that have been brought together) such as sum, mean, or median, to


DATA
Qualitative (Categorical)
Quantitative (Numerical)
Nominal
Ordinal
Interval
Ratio
Continuous
Discrete

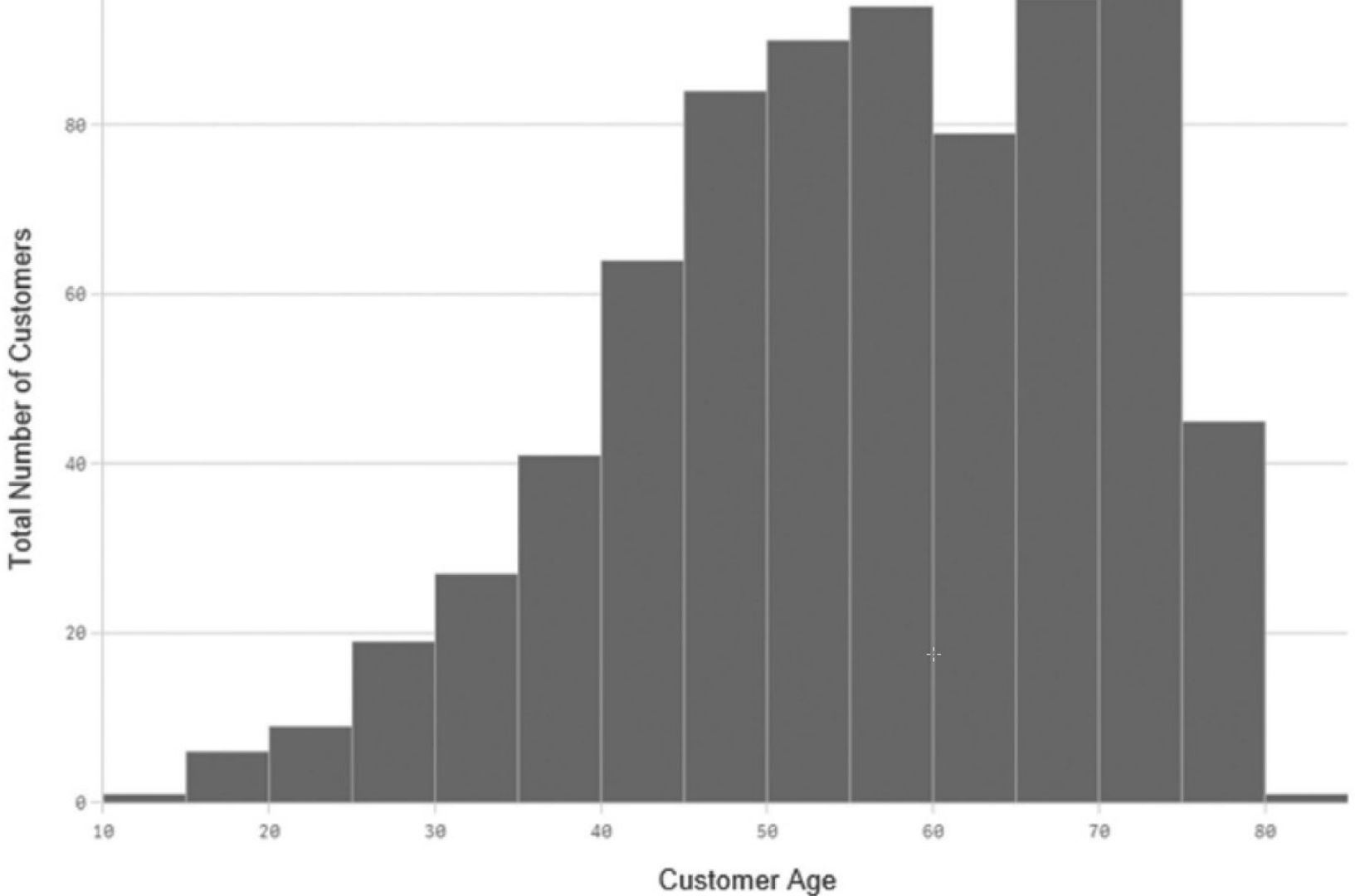

Ratio
Interval
Ordinal
Nominal

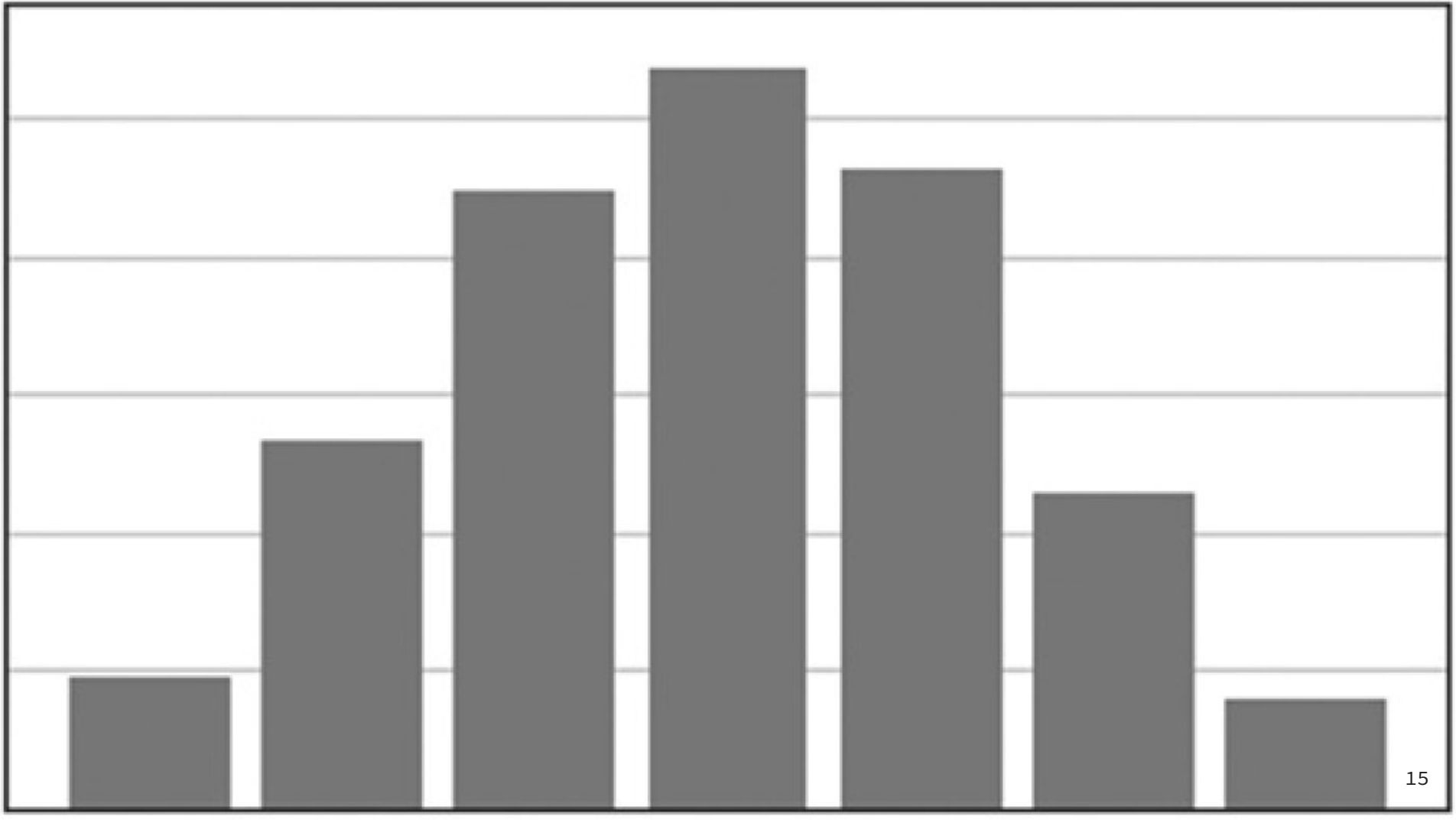

DATA
Qualitative (Categorical)
Quantitative (Numerical)
Nominal
Ordinal
Interval
Ratio
Continuous
Discrete

STATEMENT	NOMINAL	ORDINAL	INTERVAL	RATIO
Order of values in the variable is established	No	Yes	Yes	Yes
Can perform addition and subtraction (the difference between values in the variable can be evaluated)	No	No	Yes	Yes
Can perform multiplication and division (the variable contains an absolute zero)	No	No	No	Yes


DATA
Qualitative (Categorical)
Quantitative (Numerical)
Is there an established order between values in the variable?
Does the variable contain an absolute zero?
no
yes
no
yes
Nominal
Ordinal
Interval
Ratio

Organizing Data

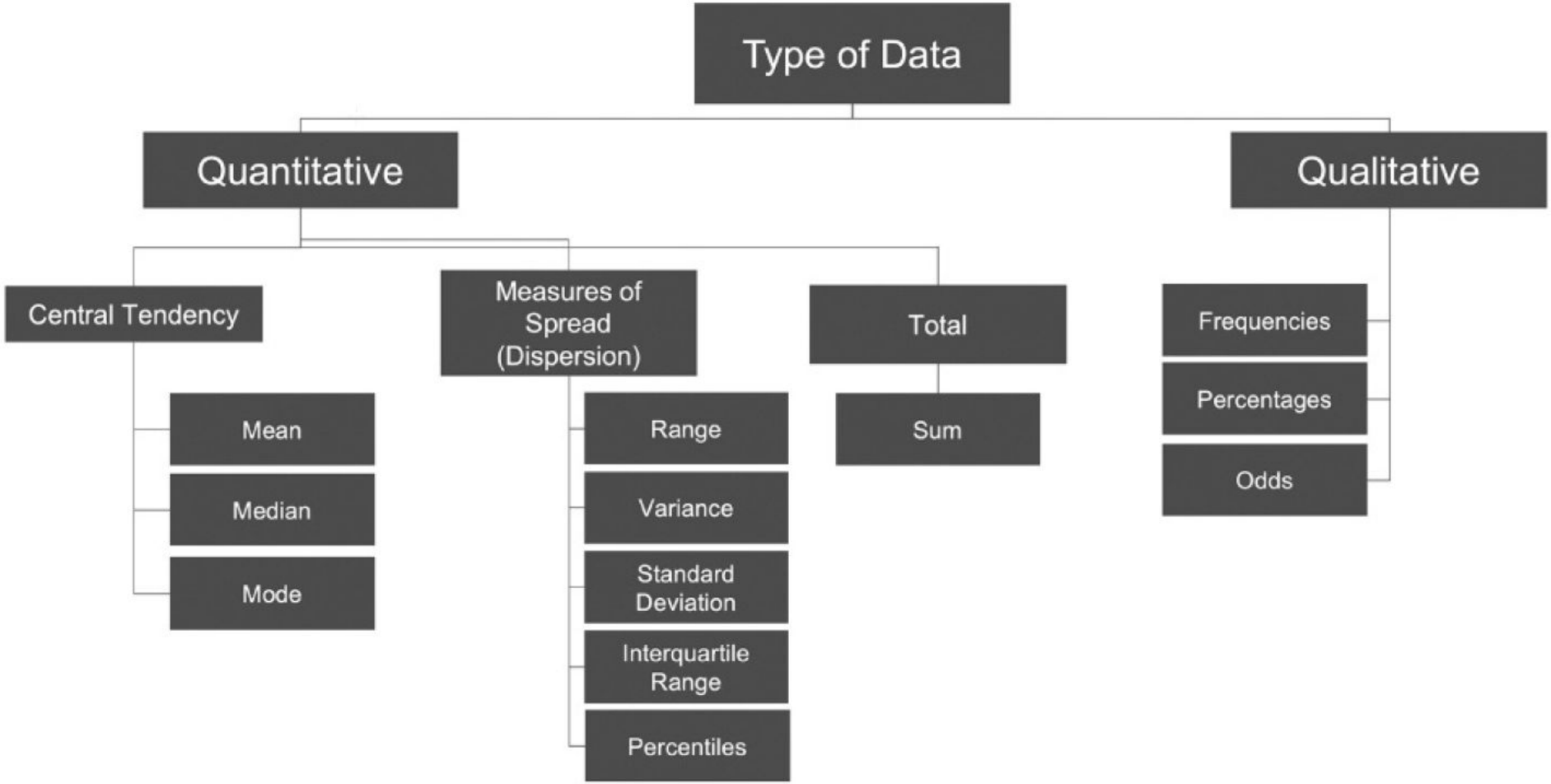

Total Number of Customers
0
20
40
60
80
10
20
30
40
50
60
70
80
Customer Age

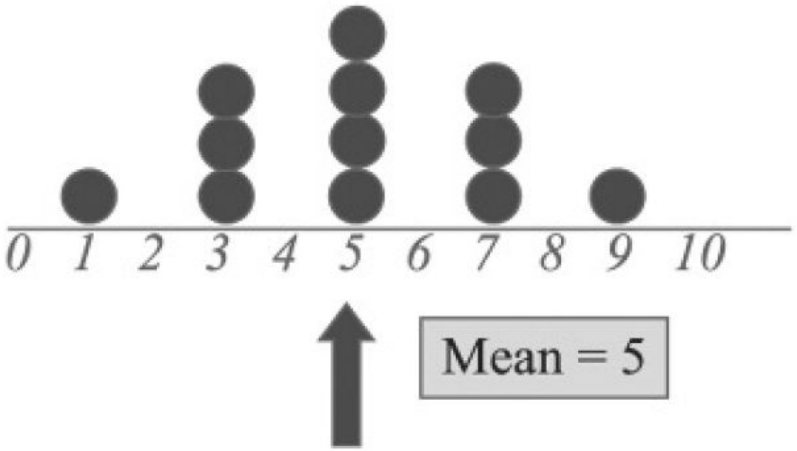


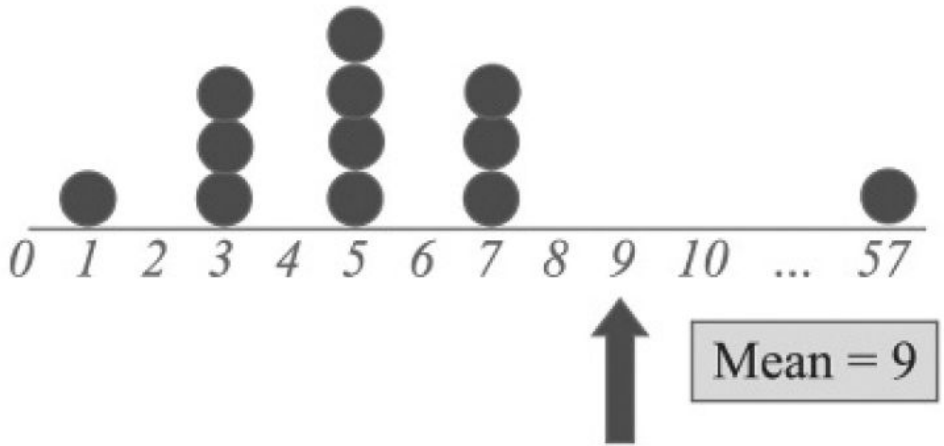
Why is this important?

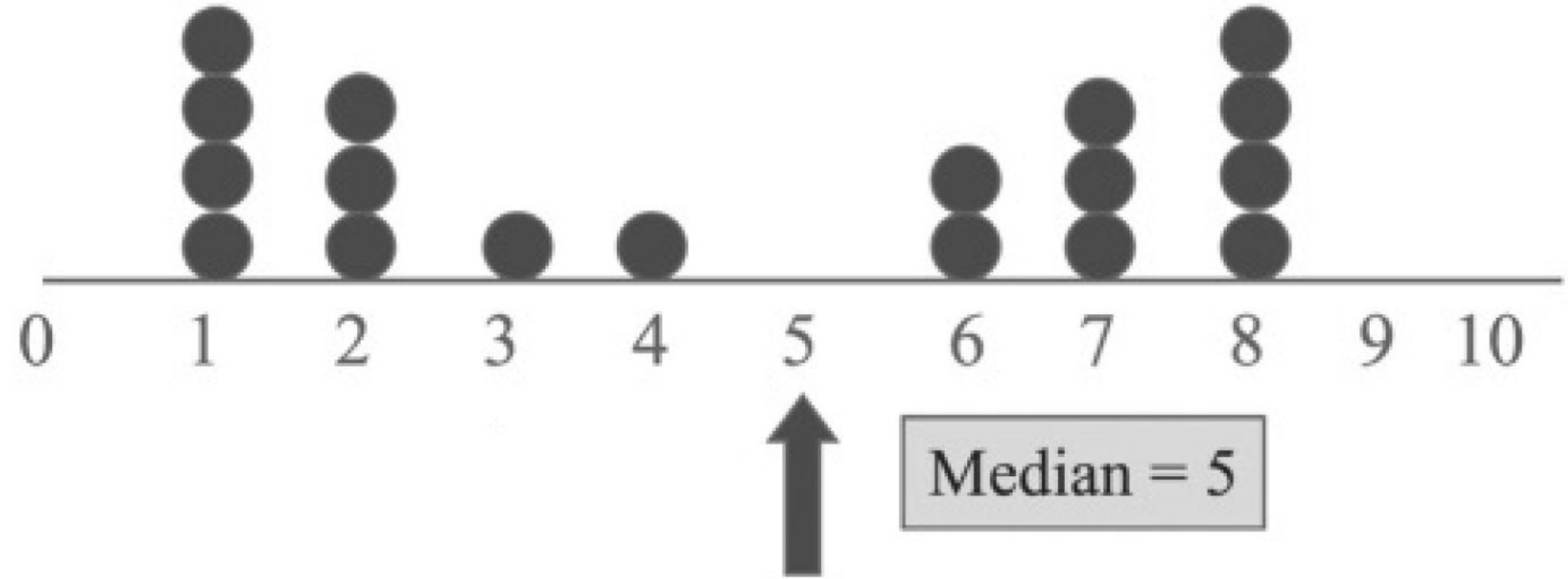
4 Types of Descriptive Analytics

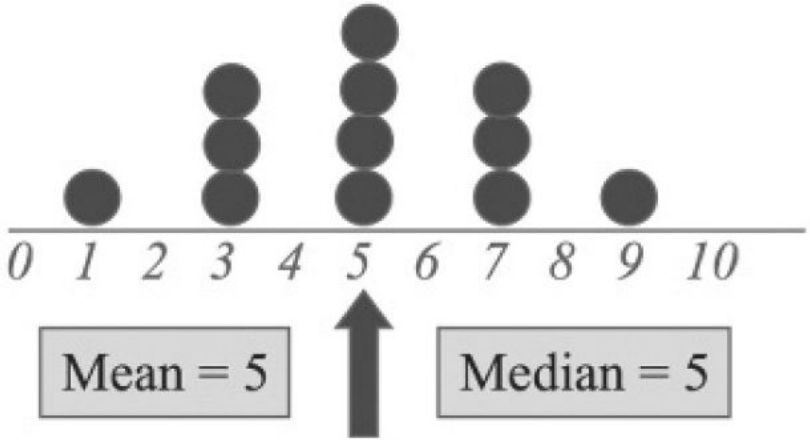
1. Central Tendency
2. Dispersion
3. Frequencies and Percentages
4. Totals

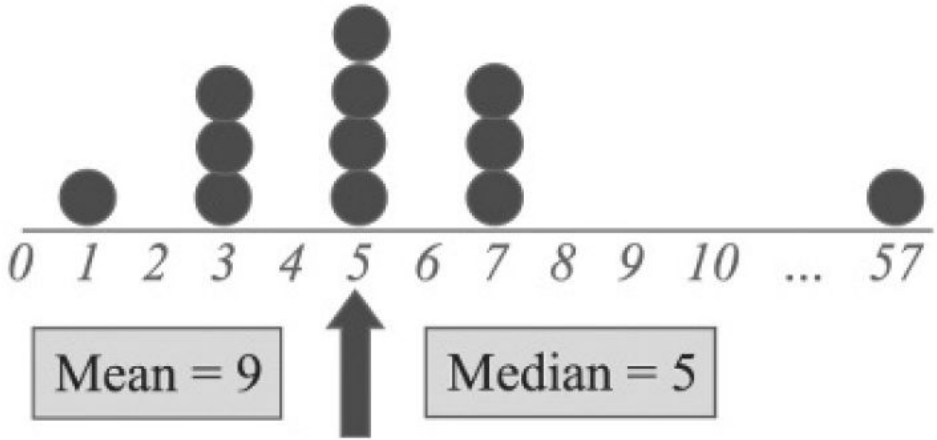

Type of Data
Quantitative
Qualitative
Central Tendency
Measures of Spread (Dispersion)
Total
Frequencies
Percentages
Odds
Mean
Median
Mode
Range
Variance
Standard Deviation
Interquartile Range
Percentiles
Sum


0 1 2 3 4 5 6 7 8 9 10
Mean = 5


0 1 2 3 4 5 6 7 8 9 10 ... 57
Mean = 9


0 1 2 3 4 5 6 7 8 9 10
Median = 5


0 1 2 3 4 5 6 7 8 9 10
Mean = 5
Median = 5


0 1 2 3 4 5 6 7 8 9 10 ... 57
Mean = 9
Median = 5

Mode

The most frequently occurring value in a set.

1,1,2,2,4,4,4,4,4,5,5

What is the mode?

MEASURE	NOMINAL	ORDINAL	INTERVAL	RATIO
Mode	Yes	Yes	Yes	Yes
Median	No	Yes	Yes	Yes
Mean	No	No	Yes	Yes