Kenya Medical Training College -Port Reitz Campus
Department of Clinical Medicine
Year Two Semester One
Statistical Data
30th September 2020

Willis J. Opalla

Statistical Data

- Learning Objectives

    To apply the principles of statistical data in presentation, analysis and interpretation of data in healthcare.

Learning Outcomes

- By the end of this module, you should be able to

  1. Define data.

  2. Explain the criteria for classification of data.

  3. Explain the different types of data.

  4. Apply knowledge of vital statistics in calculation of demographic rates.

  5. Present, analyze and interpret data using frequency tables, charts and graphs.

Unit 3: Statistical Data

- **Unit 3 Contents:**
  - Primary, secondary, numerical and categorical data.
  - Data presentation: grouped and ungrouped data.
  - Vital statistics, calculation of demographic rates.
  - Introduction to data analysis, interpretation and presentation.

Statistical Data

- Data is the base of all operations in statistics.

- Definition:

  1. A systematic record of characteristics or different values observed in a particular item of interest.

  2. A collection of facts and figures to be used for a specific purpose such as a survey or analysis.

  3. A collection of facts, e.g. numbers, words, measurements, observations or descriptions of an item being studied.

Classification of Statistical Data

- Data can be classified, based on:
  1. Source: (Primary and Secondary data)
  2. Manipulability characteristics: (Quantitative, Qualitative or Numerical, Categorical data).
  3. Sophistication of measurement: (Nominal, Interval, Ordinal, Ratio).
  4. Precision of measurement: (Discrete or Continuous).
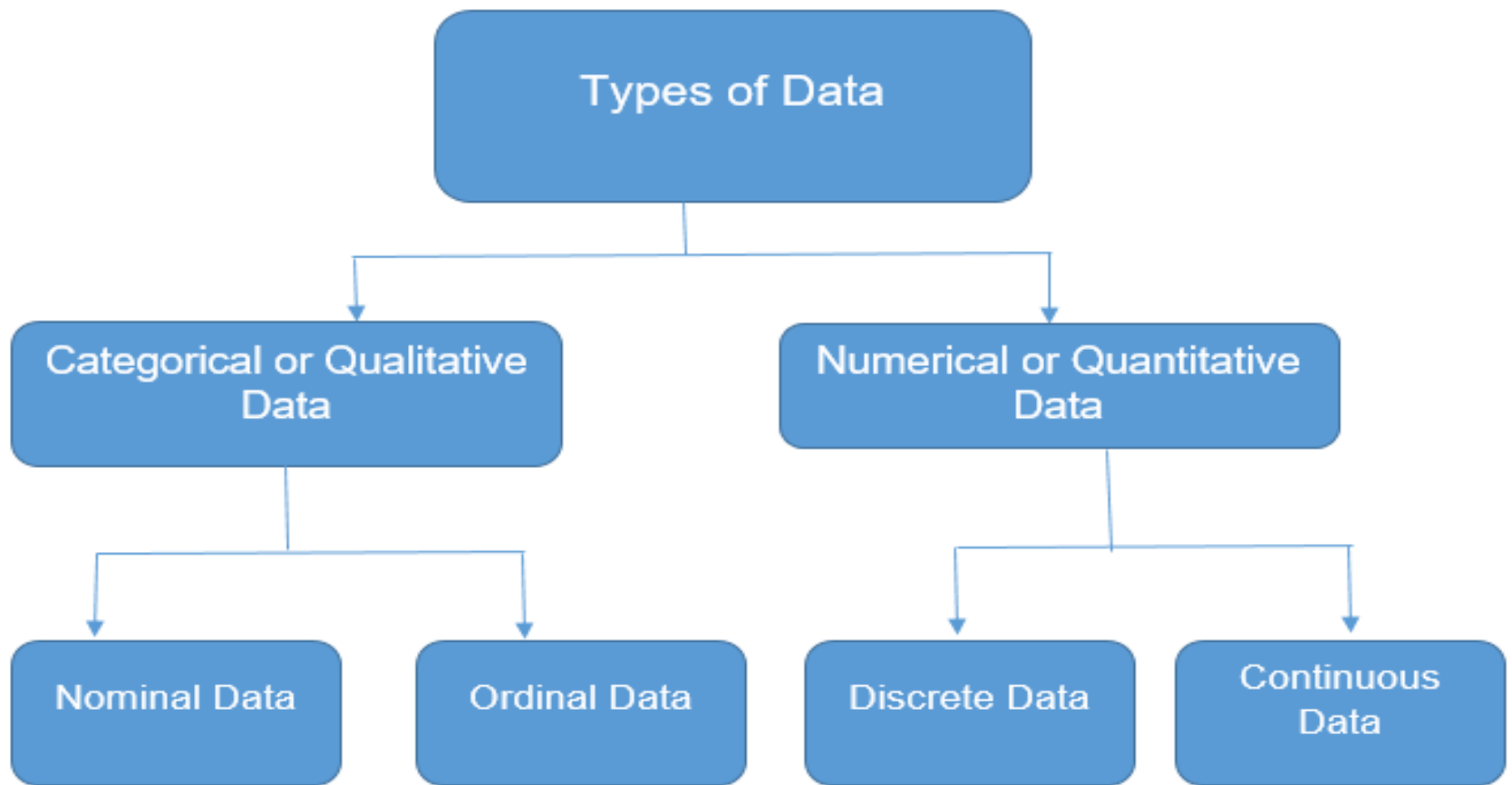 5. Manner of presentation: (Grouped or Ungrouped).
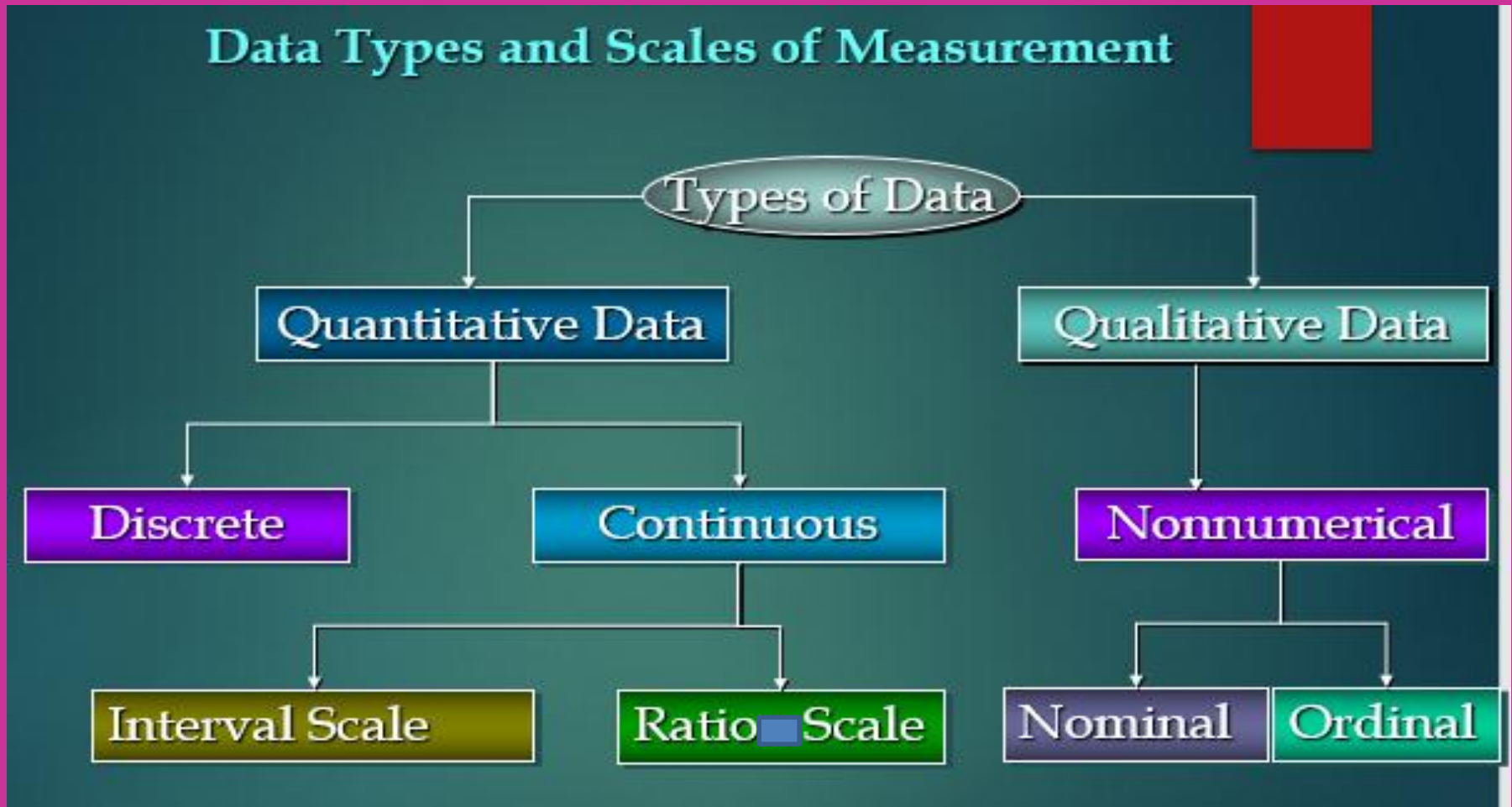
Types of Statistical Data

1. Primary data
2. Secondary data
3. Numerical (Quantitative) data
4. Categorical (Qualitative) data
5. Discrete data
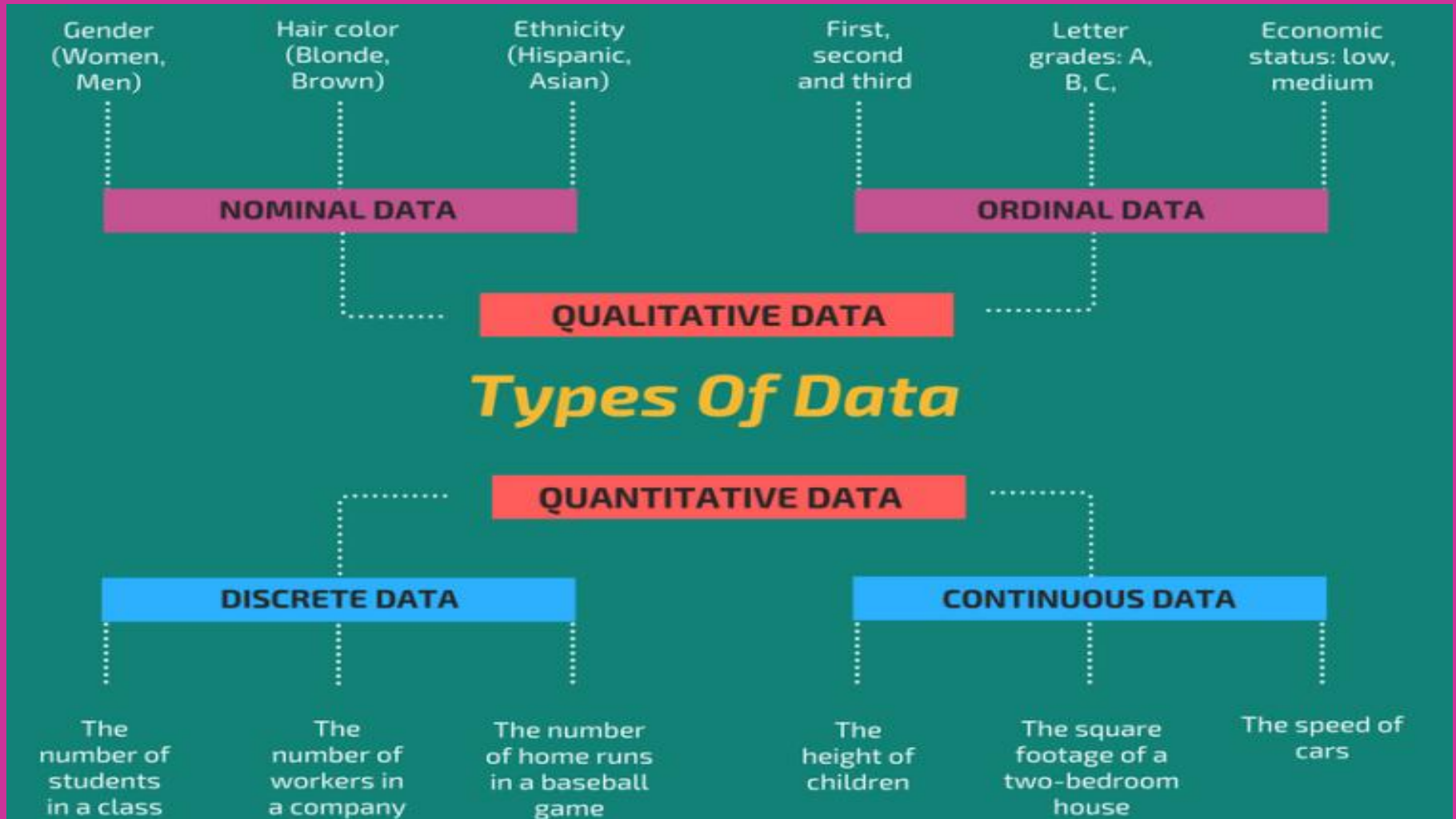6. Continuous data
7. Grouped data
8. Ungrouped data

Classification of Data Cont…

Data Types and Measurement Scales

Data Types and Scales of Measurement

Examples of Types of Data

Primary Data

- Are observations or facts which the investigator collects directly from respondents or research field and for the first time.

- Collected for the first time for a particular purpose.

- This data is 'pure' and original in the sense that no statistical manipulations have been performed on them.

Primary Data: Merits and Demerits

Merits	Demerits
Has a higher accuracy	Time consuming
Data interpretation is better	More resources are required
Addresses the key specific objectives	May be subject to inaccurate feedback
The investigator has greater control	Requires more skilled labour
Addresses relevant research concerns	Evaluated cost

Secondary Data

- Are observations or facts that an investigator obtains from what was collected earlier by other officials or researchers and stored in filed records or books.

- Data sourced from someplace that originally collected it.

- Collected by some researchers in the past, stored in filed records and is available either in published or unpublished form.

Secondary Data: Merits and Demerits

Merits	Demerits
1) Is readily available. 2) Much less expensive as compared with primary data. 3) Less time-consuming as compared with primary data.	1) Has a possibility that proper procedure might not have been followed during their collection. 2) May not be relevant in the present context. 3) May not be free from personal bias or prejudices. 4) May lack the needed accuracy or reliability. 5) May not be adequate. 6) Proper care and precaution may have to be taken before using them.

Quantitative and Qualitative Data

Types of Data in Statistics

Data can be classified into **two major groupings**:

Quantitative Data ("Numerical")

Data that can be measured with *numbers*, such as distance, duration, length, revenue, speed. Let's further classify these into two groupings:

Discrete

Whole numbers (integers) that cannot be divided, such as the # of eggs, # of wins, or # of dogs. You can't have 3.2 dogs. This data is binary

Continuous

Numbers that *can* be broken into finer and finer units (usually within a range). Weight, height, temperature are all examples (3.4981637081 lbs)

- Interval Scale Data
- Ratio Scale Data

Qualitative Data ("Categorical")

Non-numerical data that is usually textual and descriptive, like "mostly satisfied," "brown eyes," "female," "yes/no," etc.
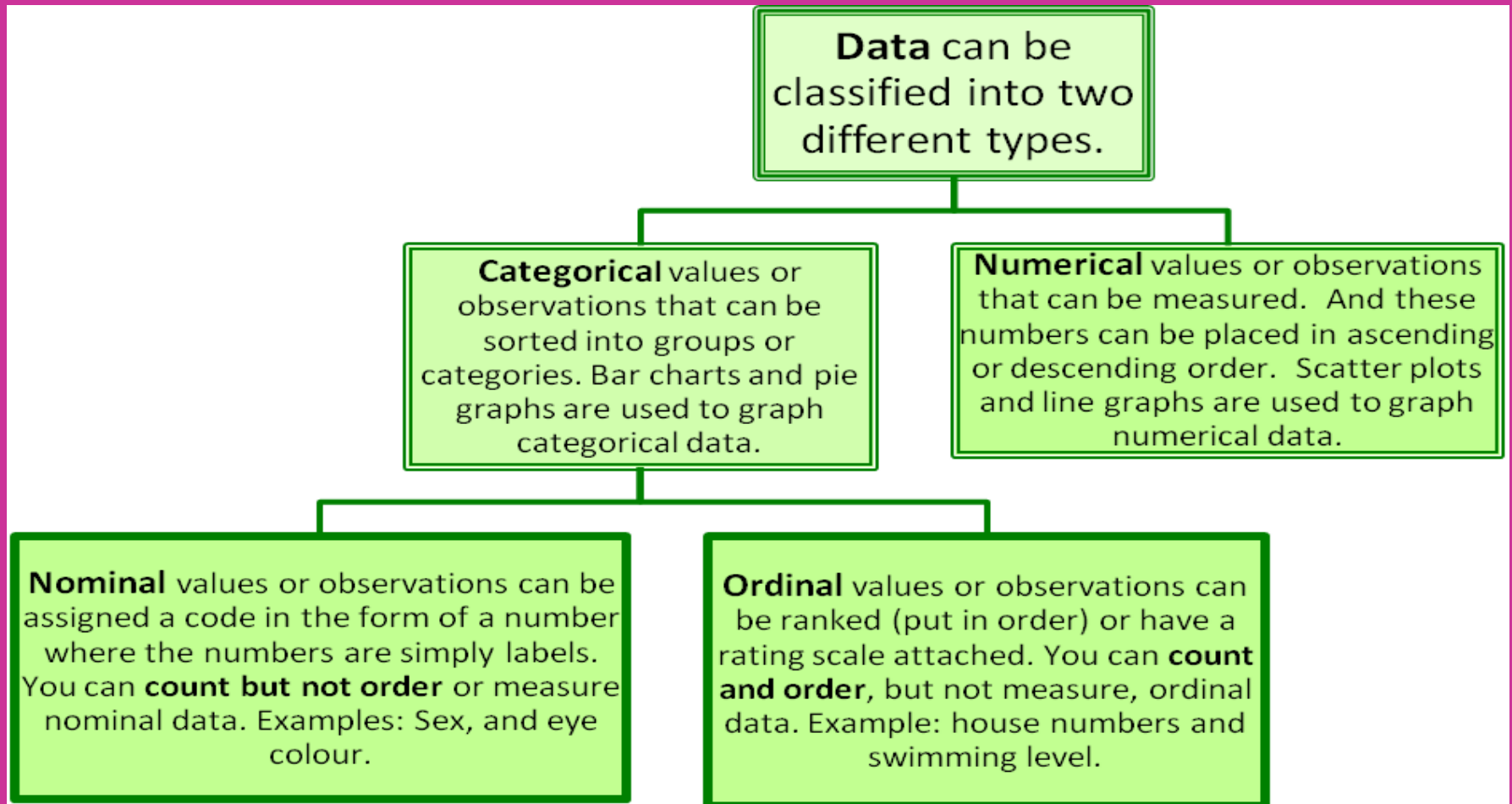
- Nominal Scale Data (Named)
  - Nominal with order
  - Nominal without order
  - Dichotomous
- Ordinal Scale Data (Ordered)

Quantitative and Qualitative Data

Data can be classified into two different types.

Categorical values or observations that can be sorted into groups or categories. Bar charts and pie graphs are used to graph categorical data.

Numerical values or observations that can be measured. And these numbers can be placed in ascending or descending order. Scatter plots and line graphs are used to graph numerical data.

Nominal values or observations can be assigned a code in the form of a number where the numbers are simply labels. You can **count but not order** or measure nominal data. Examples: Sex, and eye colour.

Ordinal values or observations can be ranked (put in order) or have a rating scale attached. You can **count and order**, but not measure, ordinal data. Example: house numbers and swimming level.

Quantitative Data

- Data which can be measured by assigning some numbers and not simply being observed.

- Can be numerically represented and calculations can be performed on them.

- e.g. data on the number of students playing different sports, number of discharged patients, number of beds in an ICU, etc.

- Also referred to as numerical data because measured in terms of numbers.

Qualitative Data

- Data in the form of descriptions that may be observed but cannot be computed or calculated.

- Data that reflects abstract or emotional reactions or attitudes of respondents to the phenomenon under investigation.

- e.g. attributes such as intelligence, honesty, wisdom, cleanliness, creativity, etc.

- This type of data are more exploratory than conclusive in nature.

- Also referred to as Categorical data since emphasis is on describing categories.

Qualitative (Categorical) Data
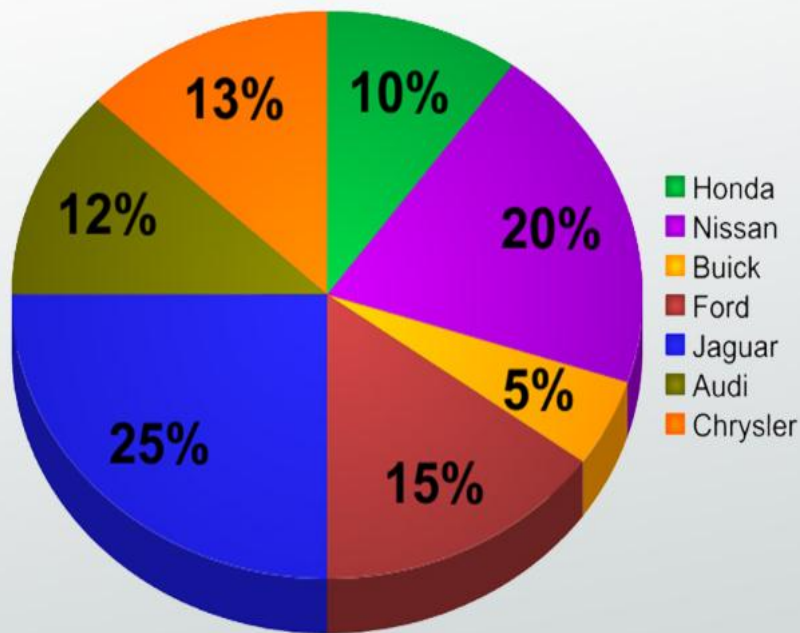
Discrete Data

- Data which only assume or take the form of certain specific values rather than a range of values, e.g. blood group or gender or number of hospital admissions in a given unit time or number of drugs in polypharmacy among the elderly, heart rate, number of children, etc.

- Commonly presented in bar charts.

Continuous Data

- **D**ata which can take values between a certain range with the highest and lowest values.

- The difference between the highest and lowest value is called the range of data, e.g. the gestational age, age of persons or height, weight or head circumference, etc, can take values even in decimals.

- Tabulated in a frequency distribution table.

- Graphically represented using histograms.

Ungrouped and Grouped Data

- Classification based on how data are presented:

Ungrouped vs. Grouped Data

Data can be classified as grouped or ungrouped.

Ungrouped data are data that are not organized, or if arranged, could only be from highest to lowest or lowest to highest.

Grouped data are data that are organized and arranged into different classes or categories.

Ungrouped and Grouped Data

Ungrouped Data	Grouped Data	
	Score	Frequency
96 67 28 32 66 65 69 33 98 96 76 49 52 64 76 83 92 93 68 52 79 81 30 68 69 83 86 43 45 39 83 75 66	20-29	3
	30-29	14
	40-49	12
96 92 75 83 76 83 85 62 37 65 63 44 62 31 36 38 42 39 83 87 56 58 88 89 93 42 53 69 90 55 66 49 52	50-59	8
	60-69	18
42 32 38 42 40 40 42 89 65 73 81 83 59 82 75 82 86 23 35 76 83 85 83 92 75 89 66 91 83 34 36 27 90 69	70-79	10
	80-89	23
	90-99	12
	Total	**100**

Ungrouped and Grouped Data

- Classification based on how data are presented:

⊙ **Ungrouped data** –
Data that has not been organized into groups. Also called as raw data.

Data	Frequency
2	8
3	4
5	6
7	7
8	2
9	5

⊙ **Grouped data** - Data that has been organized into groups (into a frequency distribution).

Data	Frequency
2 - 4	5
5 - 7	6
8 - 10	10
11 - 13	8
14 - 16	4
17 - 19	3

Sources of Data

1. Official records that are kept as a routine

e.g.

a) Patients' hospital medical records.
b) Hospital's books of accounts contain data on the hospital's financial management.

2. Published reports, commercially available data banks or literature in earlier research.

3. Surveys e.g. among patients.

4. Experiments.

Summary

- Data is systematic record of characteristics observed in a particular item of interest.

- Data can be classified, based on source, manipulability characteristics, sophistication of measurement, precision of measurement and manner of presentation.

- The sources of data include routinely kept official records, books and published reports, literature on earlier surveys and experiments.

References

- Heumann, C., Schomaker, M. and Shalabh (2016) *Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R*, Springer International Publishing Switzerland.