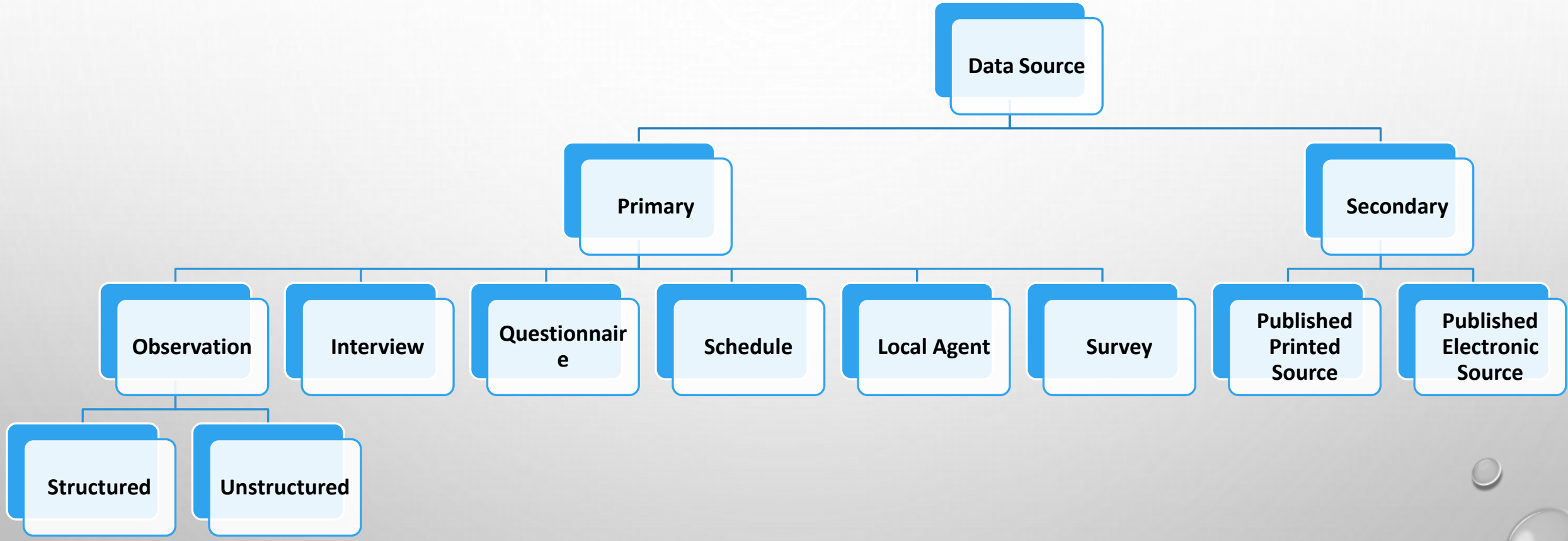
Data Sources

Data Source


Data Source
Primary
Secondary
Observation
Interview
Questionnaire
Schedule
Local Agent
Survey
Published Printed Source
Published Electronic Source
Structured
Unstructured

Example

- News Paper ?
- Diary
- Research Paper
- Review Articles
- Annual Report
- NGO published report
- Letter
- Lyrics

Terms And Symbol

- *Term is a word used to explain a particular identity*
- *Symbol is a mark or sign with a particular meaning*

Terms

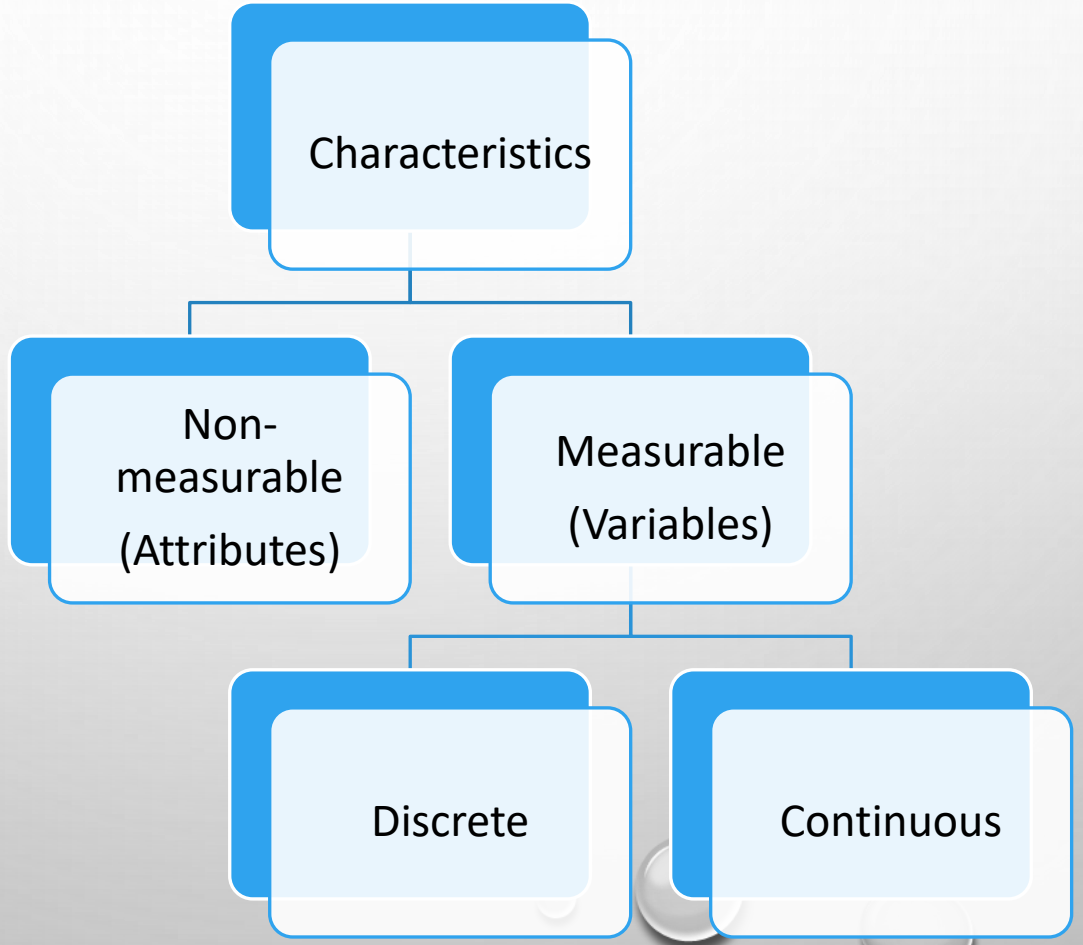
- Population
- Sample
- Data
- Observation
- Parameter
- Variables
- Characteristics
- Error
- Array
- Interval
- Attributes
- Class Size
- Class Mark
- Frequency
- Frequency Distribution
- Quantitative
- Qualitative
- Range
- Correlation
- Subscript
- Summation

Term

- ***Observation:*** *Measurement of an event is called Observation*
- ***Parameter :*** *A value calculated from a defined population is called parameter. e.g. mean height, birth rate, and mortality rate etc.*
- ***Statistics:*** *The quantity calculated to represent a character of population is known as* ***parameter*** *whereas quantity calculated to present the character of the sample is called* ***statistics.***

Characteristics

- *The term 'Characteristics' means Quality possessed by an Individual i.e. object. Height, weight, age , Hb% etc. are characteristics.*



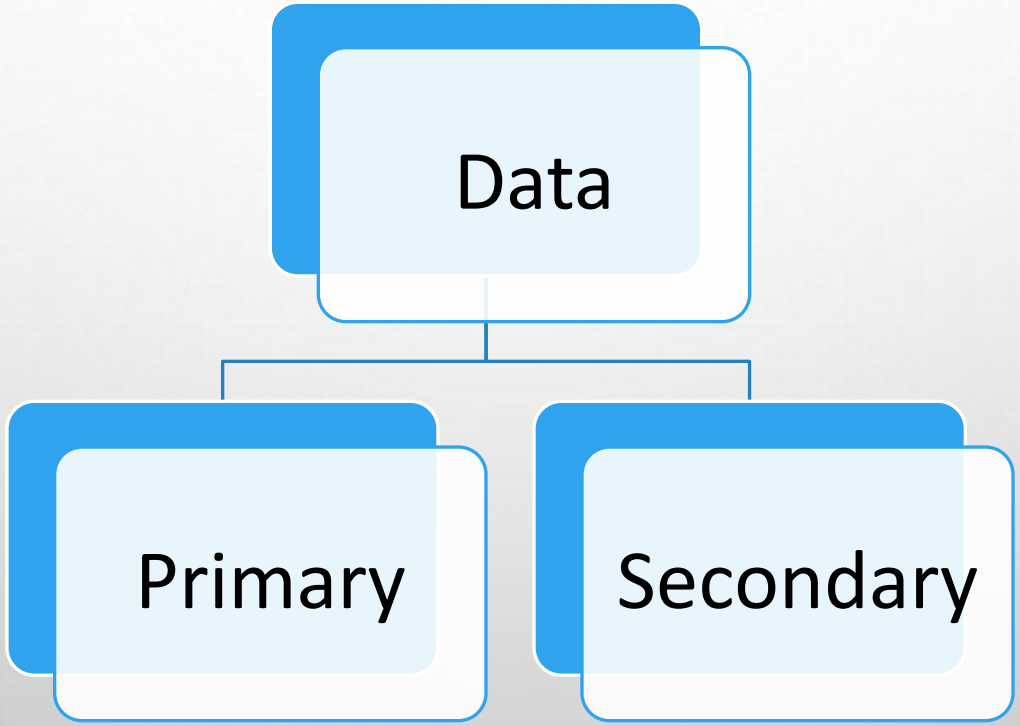
Symbols

Symbol	Name	Symbol	Name
∑	Sum	n	Size of a subsample
^	Hat, used above a parameter to denote an estimate	N	Total sample size
ANOVA	Analysis of variance	OR	Odds ratio
α	Alpha, probability of Type I error	*P*	Statistical probability
β	Beta, probability of Type II error; or population regression coefficient	X^2	Chi-square test or statistic
CI	Confidence interval	*r*	Bivariate correlation coefficient
CV	Coefficient of variation	*R*	Multivariate correlation coefficient
Δ	Delta, change	RR	Relative risk
δ	Delta, true sampling error	ρ	Rho, population coefficient
ε	Epsilon, true experimental error	SD	Standard deviation of a sample
H_0	Null hypothesis	SE	Standard error
H_1	Alternate hypothesis; specify whether 1 or 2 sided	SEM	Standard error of the mean
HR	Hazards ration	*t*	Student t; specify α level
κ	Kappa statistic	*U*	Mann-Whitney *U* (Wilcoxon) statistic
μ	Population mean	*z*	*z* score

$\sum$	Summation	x	An individual value, an observation
s	The standard deviation of sample data	x_1	A particular (1st) individual value
σ	The standard deviation of population data	x_i	For each, all, individual values
s^2	The variance of sample data	$\bar{x}$	The mean, average of sample data
σ^2	The variance of population data	$\bar{\bar{x}}$	The grand mean, grand average
R	The range of data	μ	The mean of population data
$\bar{R}$	The average range of data	p	A proportion of sample data
k	Multi-purpose notation, i.e. # of subgroups, # of classes	P	A proportion of population data
$\lvert y \rvert$	The absolute value of some term	n	Sample size
$>, <$	Greater than, less than	N	Population size
$\geq, \leq$	Greater than or equal to, less than or equal to		

Image Source: Google

Data Classification



Primary Data

- *Original and Unique*
- *Data collected directly by investigator / Agency for the first time for a specific purpose*
- *Raw data generated by experiment*
- *Data directly collected from population*

Advantage

•Specific to the problem under study.

•The quality of the data collected (for the investigator).

•If required, it may be possible to obtain additional data during the study period.

Disadvantages

1. *The investigator has to contend with all the hassles of data collection-*
   - *why, what, how, when to collect (Decision to be taken)*
   - *getting the data collected (personally or through others)*
   - *getting funding and dealing with funding agencies*
   - *ethical considerations (consent, permissions, etc.)*
2. *Ensuring the data collected is of a high standard-*
   - *all desired data is obtained accurately, and in the format, it is required in there is no fake/ cooked up data*
   - *unnecessary/ useless data has not been included*
3. *Cost of obtaining the data is often the major expense in studies*

Kumar, Ranjit. Research Methodology: A Step-by-Step Guide for Beginners. Los Angeles: SAGE, 2011. Print.
https://communitymedicine4all.com/2013/01/07/types-of-data-primary-and-secondary-data/

Secondary Data

- *Data collected by someone else for some other purpose (but being utilized by the investigator for another purpose).*
- *Gathering information with the **use of census data** to obtain information on the age-sex structure of a population, **the use of hospital records** to find out the morbidity and mortality patterns of a community, the use of an organization's records to ascertain its activities, and the collection of data from sources such as articles, journals, magazines, books and periodicals to obtain historical and other types of information, are examples of secondary data.*
-

Advantage and Disadvantage of Secondary Data

- **Advantages of using Secondary data**

- The data is already there- no hassles of data collection
- It is less expensive
- The investigator is not personally responsible for the quality of data

- **Disadvantages of using Secondary data**
- The investigator cannot decide what is collected (if specific data about something is required, for instance).
- One can only hope that the data is of good quality
- Obtaining additional data (or even clarification) about something is not possible (most often)

https://communitymedicine4all.com/2013/01/07/types-of-data-primary-and-secondary-data/

Example

- *Survey based study*
- *Opinion poll*
- *Exit poll*