Understand Your Data

Types of Data Sources

- Primary **The data collector is the one using the data for analysis**
- Secondary **The person performing data analysis is not the data collector**

Data formats

- Structured
- Unstructured

Types of Variables

- Nominal
- Ordinal
- Interval
- Ratio

FIFA World Cup™   **Tournament Statistics**   🇧🇷 BRAZIL

EDITION	PLACEMENT	MP	W	D	L	GS	GA	GSA	GAA
Brazil 2014	Fourth	7	3	2	2	11	14	1.6	2
South Africa 2010	Quarter-finals	5	3	1	1	9	4	1.8	0.8
Germany 2006	Quarter-finals	5	4	0	1	10	2	2	0.4
Korea/Japan 2002	WON	7	7	0	0	18	4	2.6	0.6
France 1998	Runners-Up	7	4	1	2	14	10	2	1.4
USA 1994	WON	7	5	2	0	11	3	1.6	0.4
Italy 1990	Round of 16	4	3	0	1	4	2	1	0.5
Mexico 1986	Quarter-finals	5	4	1	0	10	1	2	0.2
Spain 1982	Second round	5	4	0	1	15	6	3	1.2
Argentina 1978	Third	7	4	3	0	10	3	1.4	0.4
Germany 1974	Fourth	7	3	2	2	6	4	0.9	0.6
Mexico 1970	WON	6	6	0	0	19	7	3.2	1.2
England 1966	Group matches	3	1	0	2	4	6	1.3	2
Chile 1962	WON	6	5	1	0	14	5	2.3	0.8
Sweden 1958	WON	6	5	1	0	16	4	2.7	0.7
Switzerland 1954	Quarter-finals	3	1	1	1	8	5	2.7	1.7
Brazil 1950	Runners-Up	6	4	1	1	22	6	3.7	1
France 1938	Third	5	3	1	1	14	11	2.8	2.2
Italy 1934	Preliminary round	1	0	0	1	1	3	1	3
Uruguay 1930	Group matches	2	1	0	1	5	2	2.5	1

MP : Matches played

W: Wins

D: Draws

L: Loss

GS : Goals scored

GA: Goals Against

GSA: Goals Scored Average

GAA: Goals Against Average

   http://www.fifa.com/fifa-tournaments/teams/association=bra/index.html

Types of data

Quantitative

Qualitative

Discrete
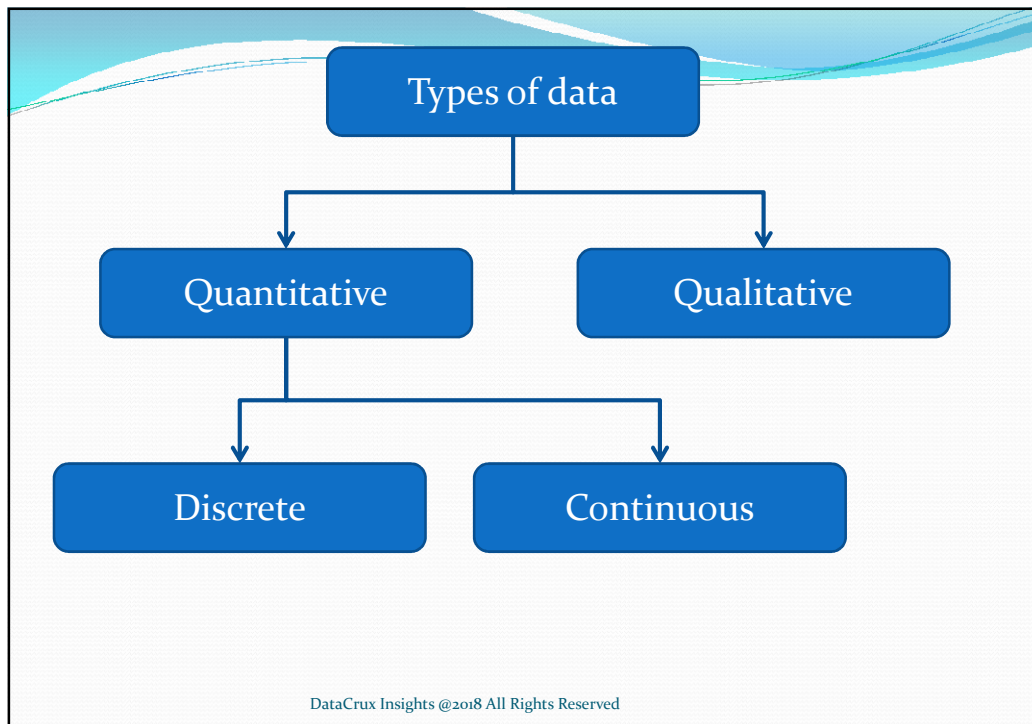
Continuous

Quantitative Data

- Described by a number
- Discrete or continuous
- Using operations such as average make sense

ATM Withdrawal

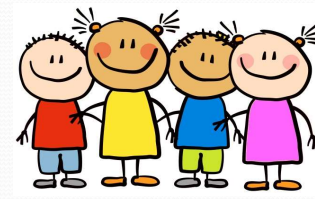**Number of employees in a company**

Time

Discrete Data

- Discrete data can assume only certain values.
- Typically, this type of data can be counted.

Number of bedrooms **Number of cars passing a toll booth** **Number of children in a family**

Continuous Data
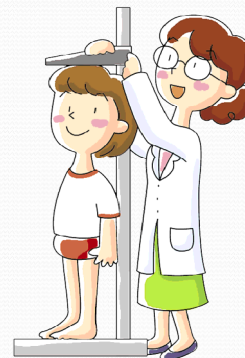
- Can take any value in a specific range.
- Outcome of measuring.

Time to reach school **Weight of tomatoes** **Height of a student**

Qualitative/Categorical Data

- Also known as categorical data
- Describes data that fits into categories
- Take on values that are names or labels

Gender

Color of ball

Nationality

Students				
Roll No	**Name**	**Age**	**Specialization**	**GPA**
1	Mark Rutherford	16	Maths	3.67
2	Peter D'silva	15	History	2.33
3	Veronica Little	16	Maths	3.00
4	Leela Apte	18	Maths	4.00
5	Tao Wan	18	History	2.06
6	Dinesh Gore	17	History	3.63
7	Sayuri Adams	15	History	2.56
8	James Lucas	17	History	3.90

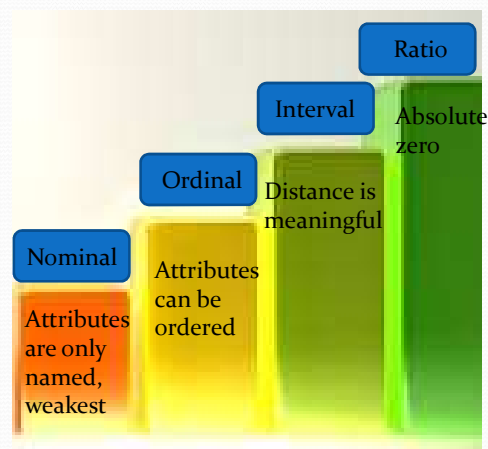| Variable	Type
Roll No	Categorical
Name	Categorical
Age	Continuous
Specialization	Categorical
GPA	Continuous

Levels of measurement

- Determines how data should be summarized and presented.
- Indicate type of statistical analysis to be used
- Also known as scales of measurement

Levels of measurement :

✓ Nominal : used to assign individuals/objects to categories

✓ Ordinal : used to rank order cases

✓ Interval : measurable, equally spaced, does not have a true zero

✓ Ratio : same as interval variable but they have a "true zero"

Ratio

Interval — Absolute zero

Ordinal — Distance is meaningful

Nominal — Attributes can be ordered

Attributes are only named, weakest

Variables

- Variable is a particular characteristic of data that can be observed and measured/counted.
- A variable can assume different values.
- e.g. Height, Weight, Gender
- 2 types of variables:
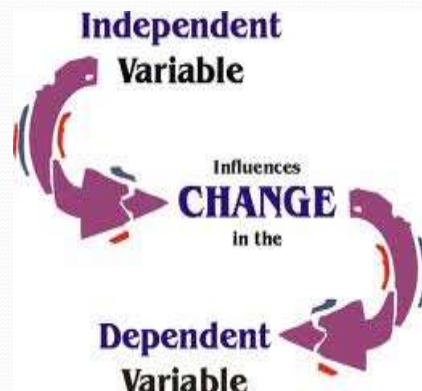 - Independent variable:
  - Dependent Variable

Variables / features

Independent variables
- Variable which stands alone
- Values are not affected by other variables

✓ Predictor
✓ Explanatory
✓ Control
✓ Input

Dependent Variables

✓ Response
✓ Outcome
✓ Target