Data Visualisation
CSE613

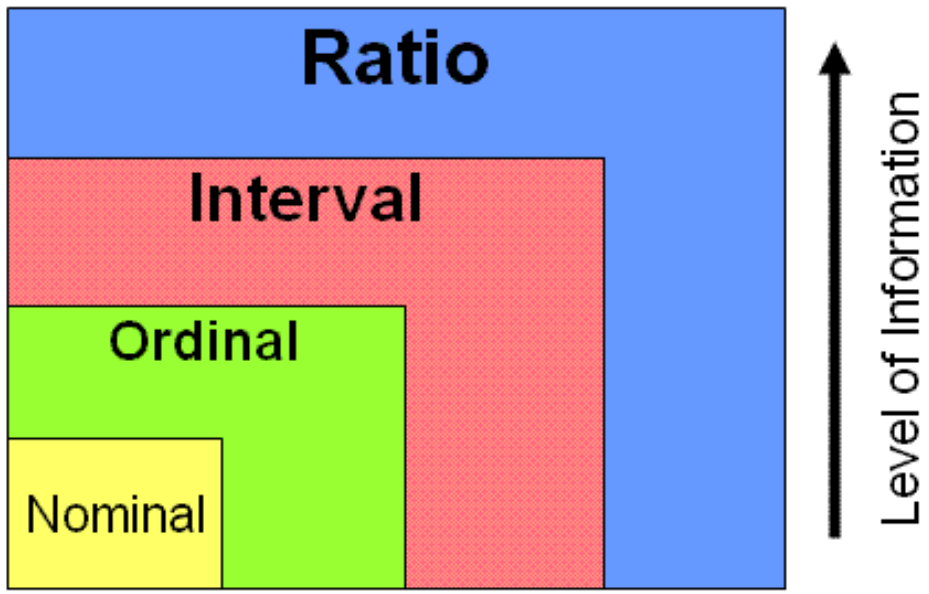
Prof. Ramesh Ragala

January 16, 2018

Different Types of Data in data visualization

- Quantitative Data
  - Continuous Data
  - Discrete Data
    - Interval Data
    - Ratio Data
- Qualitative Data
  - Ordinal Data
  - Nominal Data

Introduction: Semiotics

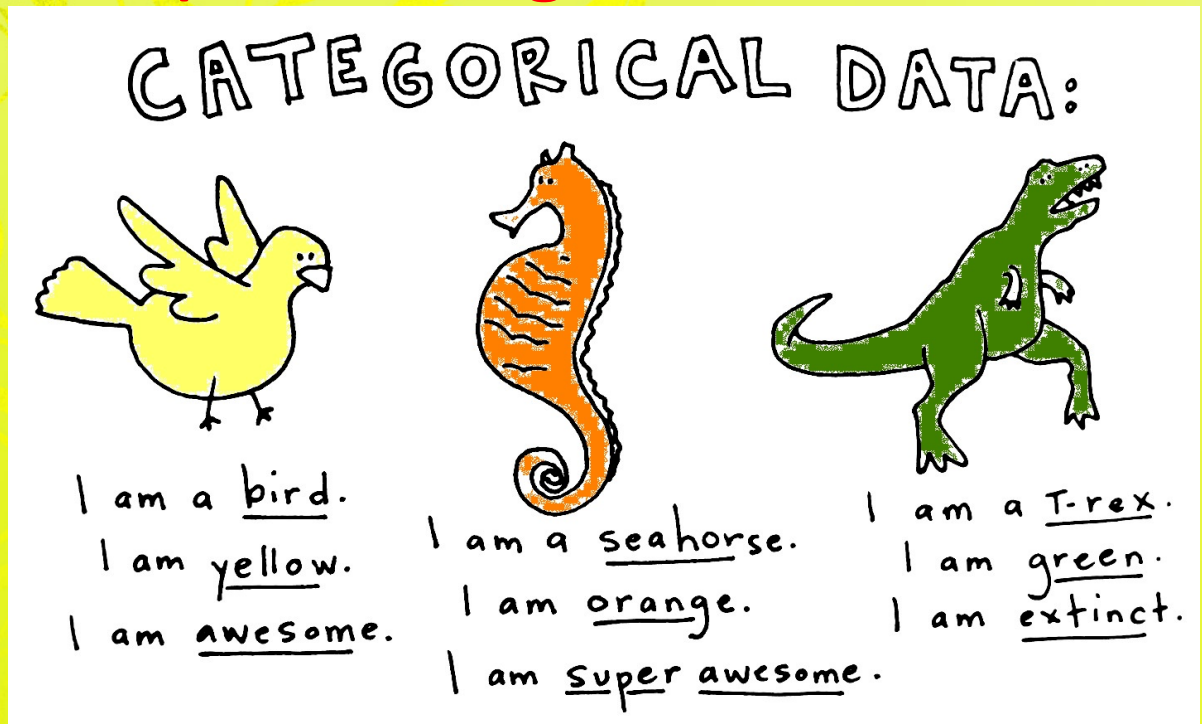


*Nonparametric statistics may be used to analyze interval and ratio data measurements.

Introduction: The essence of Semiotics

- **Level of Organisation:**
- A data variable is classified → 4 types → based on on the scale by which the values it contains are measured:
  - **Nominal/categorical data:**
    - The data values are categorical and not numeric.
    - A categorical variable is one that has two or more categories or labels or classes, but there is no intrinsic ordering to the categories.
    - simply Categorical variables represent types of data which may be divided into groups.
    - It is completely qualitative measurement.
    - Examples: age, gender, educational levels, countries, people names. operations: == and !=
    - Comparing two observations using the values for the variable, the observations will either be similar or different depending on whether the categorical value matches or not.

- **Example on Categorical Data:**



- if the categorical data has only two outcomes → binary or binomial data
- The Binomial data outcomes may pass/fail, live/dead or extinct/not extinct

Introduction: The essence of Semiotics

- **Examples on Categorical Variables**

	A	B	C	D	E	F	G	H	I
1	Name	Miles Per Gallon	Accceleration	Horsepower	weight	cylinders	year	price	Country
2	Volkswagen Rabbit Dl	43,1	21,5	48	1985	4	78	2400	Germany
3	Ford Fiesta	36,1	14,4	66	1800	4	78	1900	Germany
4	Mazda GLC Deluxe	32,8	19,4	52	1985	4	78	2200	Japan
5	Datsun B210 GX	39,4	18,6	70	2070	4	78	2725	Japan
6	Honda Civic CVCC	36,1	16,4	60	1800	4	78	2250	Japan
7	Oldsmobile Cutlass	19,9	15,5	110	3365	8	78	3300	USA
8	Dodge Diplomat	19,4	13,2	140	3735	8	78	3125	USA
9	Mercury Monarch	20,2	12,8	139	3570	8	78	2850	USA

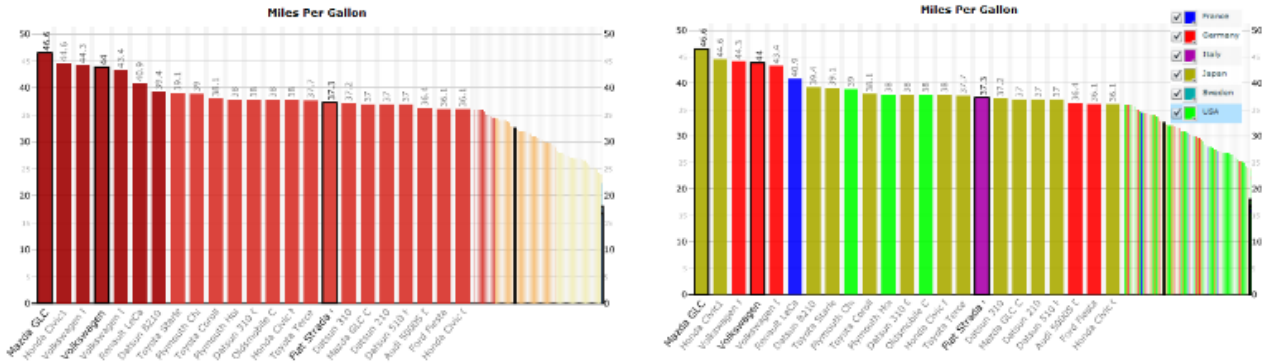


Figure: Classic car data set shown as bar chart for numerical variable "Miles per gallon" and coloured based on categorical variable Country.

Introduction: Semiotics

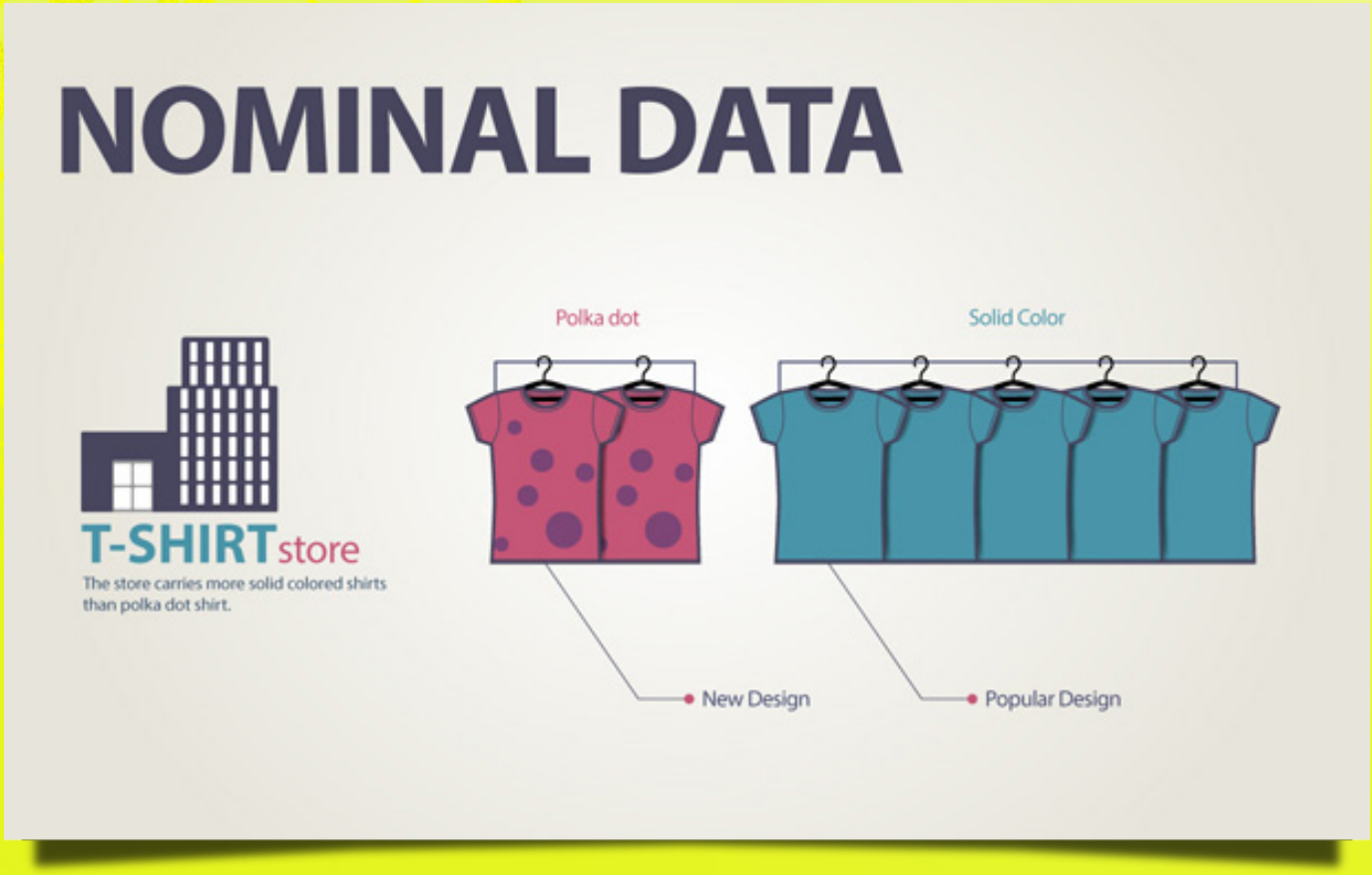


Figure: nominal level of measurement

Introduction: Semiotics

- A categorical variable (sometimes called a nominal variable) is one that has two or more categories, but there is no intrinsic ordering to the categories.
- A purely categorical variable is one that simply allows you to assign categories but you cannot clearly order the variables.
- If the variable has a clear ordering, then that variable would be an ordinal variable.
- The Nominal or categorical data has only meaning $\rightarrow$ how they are differing from one another.
- **Example:** Country names are Nominal data values $\rightarrow$ putting all country names in alphabetical order is not making any relationship to another.
- Assignment of numbers to categories has no mathematical meaning.
- Nominal categories should be mutually exclusive and exhaustive

Introduction: Semiotics

- **Where Can We Have Categorical Data:**
  - Social sciences : opinions on issues
  - Health sciences : response to treatments/drugs
  - Behavioral sciences : e.g. diagnose mental illness
  - Public health : AIDS awareness
  - Zoology : animals food preferences
  - Education : student's response to exams
  - Marketing : consumer preferences
  - Almost everywhere
- Distinction in categorical data are: Nominal Data and Ordinal Data

Introduction: Semiotics

- **Ordinal data values:**
  - The data values are categorical but ordered.
  - Comparing two observations using the values for that variable.
  - Operations: $==, !=, \leq$ and $\geq$
  - it is mainly used for obey ordering relations among data values
  - Ordinal data is that which has inherent order, but no inherent degree of difference between what is being ordered.
  - **Example:** The I^{st}, II^{nd} and III^{rd} place winners in a race are on ordinal scale
  - But we do not know **how much faster** first place was than second place
  - But we know only that one was faster than other.

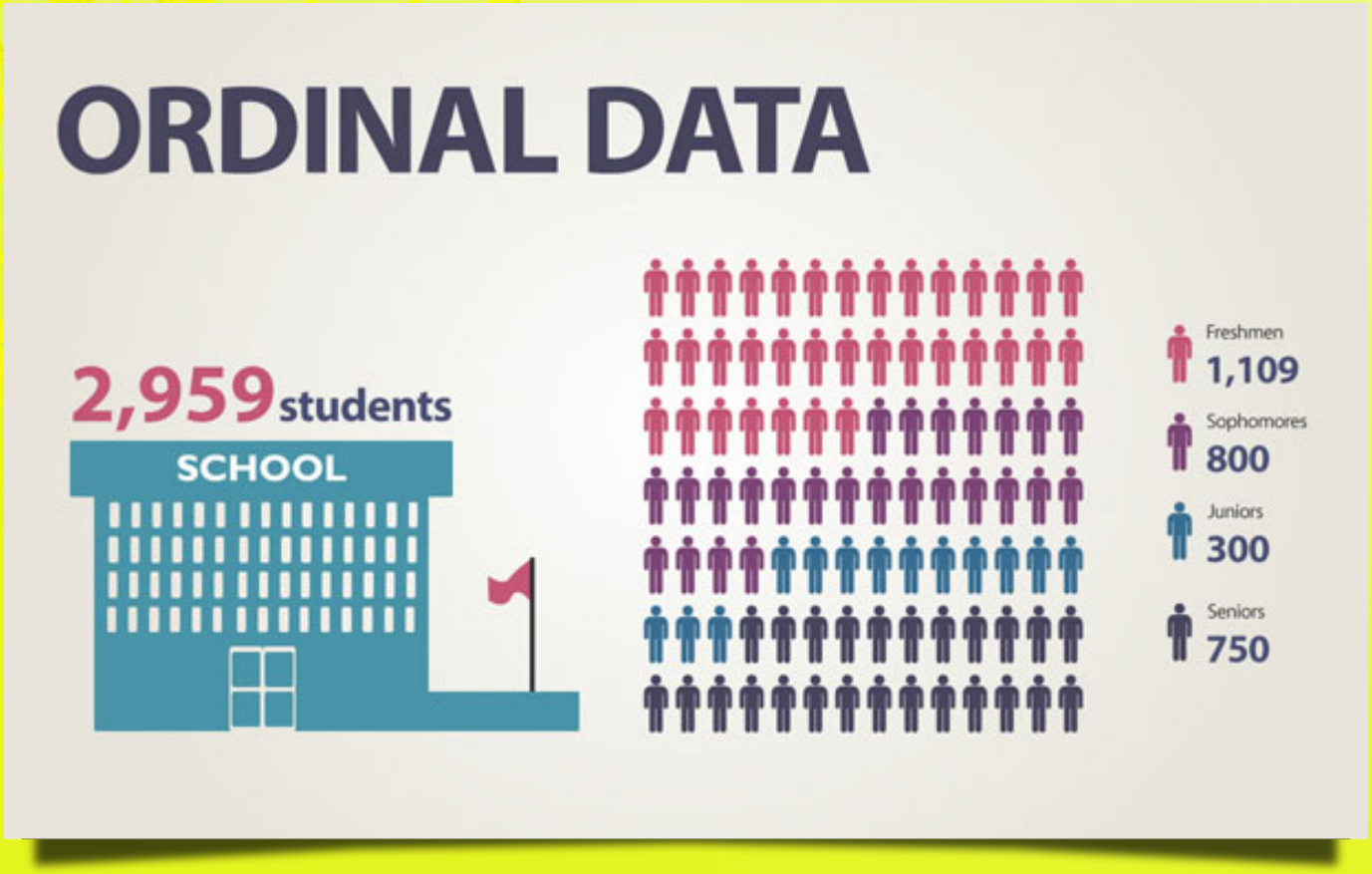
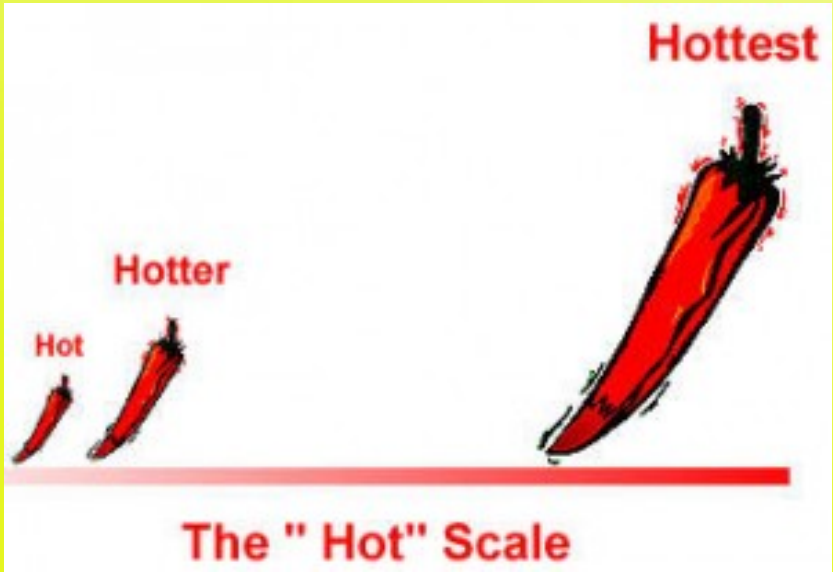


FIGURE: Ordinal level of measurement

Introduction: Semiotics



Introduction: Semiotics

- **Interval Data:**
  - The data values are numeric.
  - It represents the more sensitive type of data or sophisticated form of measurement.
  - simply, Interval data is data which exists on a scale with meaningful quantitative magnitudes between values.
  -
  - Data values can be compared quantitatively using basic arithmetic operations +,-,* and / not the values themselves.
  - The values are ordered. it includes negative numbers and zero. But zero is not absolute reference point.
  - Scale data is usually aggregated or converted to averages.

Introduction: Semiotics

- **Interval Data:**
  - **Example:1** The dataset does not contain an interval data variable, if there were a variable in a dataset that recorded the measurements of temperature. $\rightarrow$ it would be classified as a interval variable.
  - Temperature variable contains the values 40,60 and 80, we could say that compared with 40^oF, 80^oF is two times warmer than 60^oF (80-40)/(60-40), but not twice as hot because 0^oF is an arbitrarily chosen point on the scale.
  - **Example:2** if Sidda Reddy is rated as "6" on attractiveness and Durga Prasad a "3" $\rightarrow$ it does not mean Sidda Reddy is twice as attractive as Durga Prasad.

Introduction: Semiotics

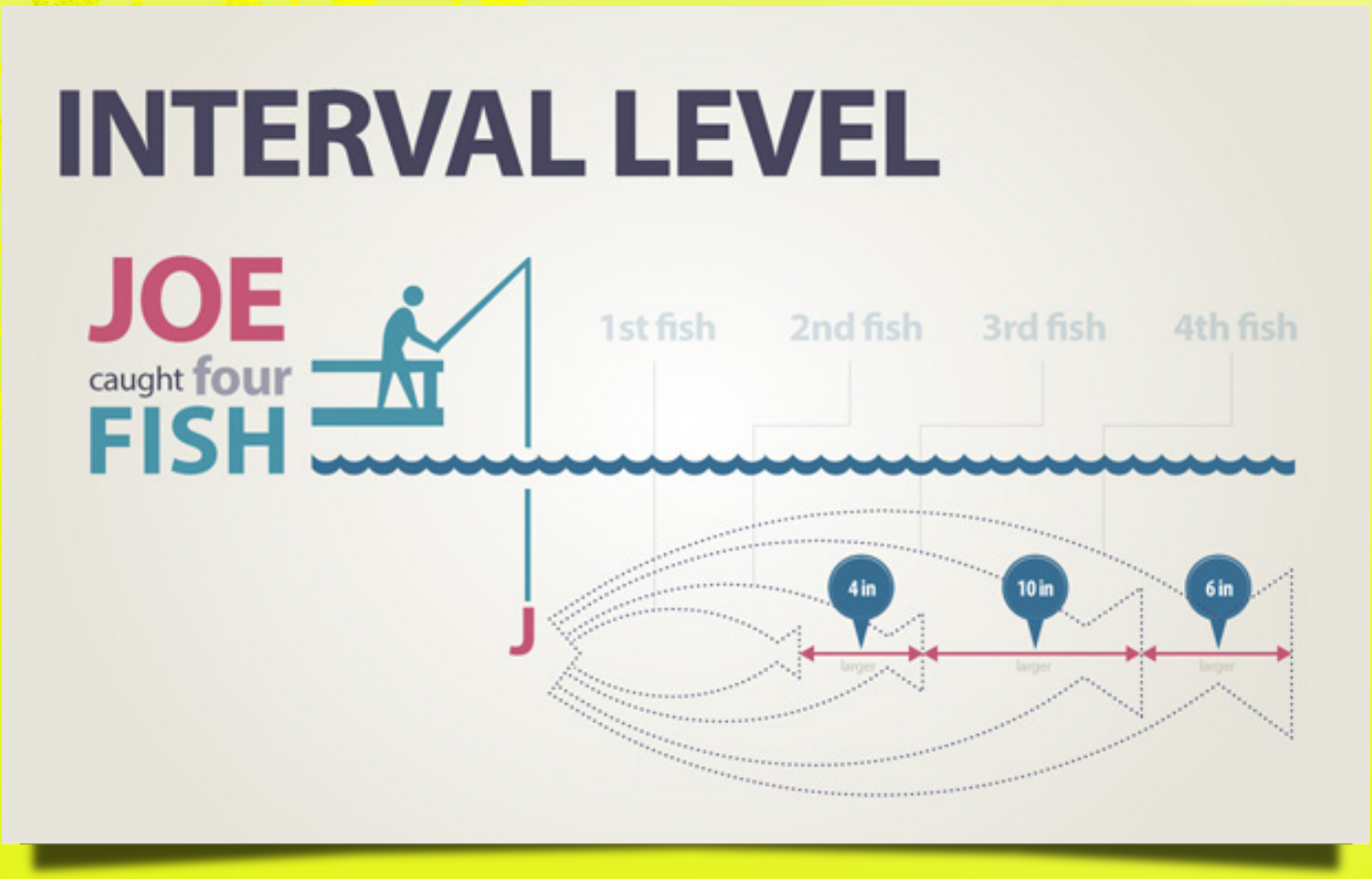


Figure: interval level of measurement

- The measurement between the sizes of the fish Joe caught in order of when he caught them.

- **Ratio Data:**
  - The Data Values are numeric and include an absolute zero.
  - This data values are allowed to compare quantitatively with other using basic arithmetic operations
  - Ratio data is data which, like interval data, has a meaningful order and a constant scale between ordered values, but additionally it has a meaningful zero value.
  - Supported Operations are ==, !=, $\leq$, $\geq$, -, / and *
  - The Ratio level of measurement applies to data that can be arranged in order.
  - In addition, both differences between data values and ratios of data values are meaningful. Data at the ratio level have a true zero.
  - **Example:** If one box weighs 50lbs and another 100lbs $\rightarrow$ the second box weighs twice as much as the first $\rightarrow$ this is not a case in interval data

Introduction: Semiotics

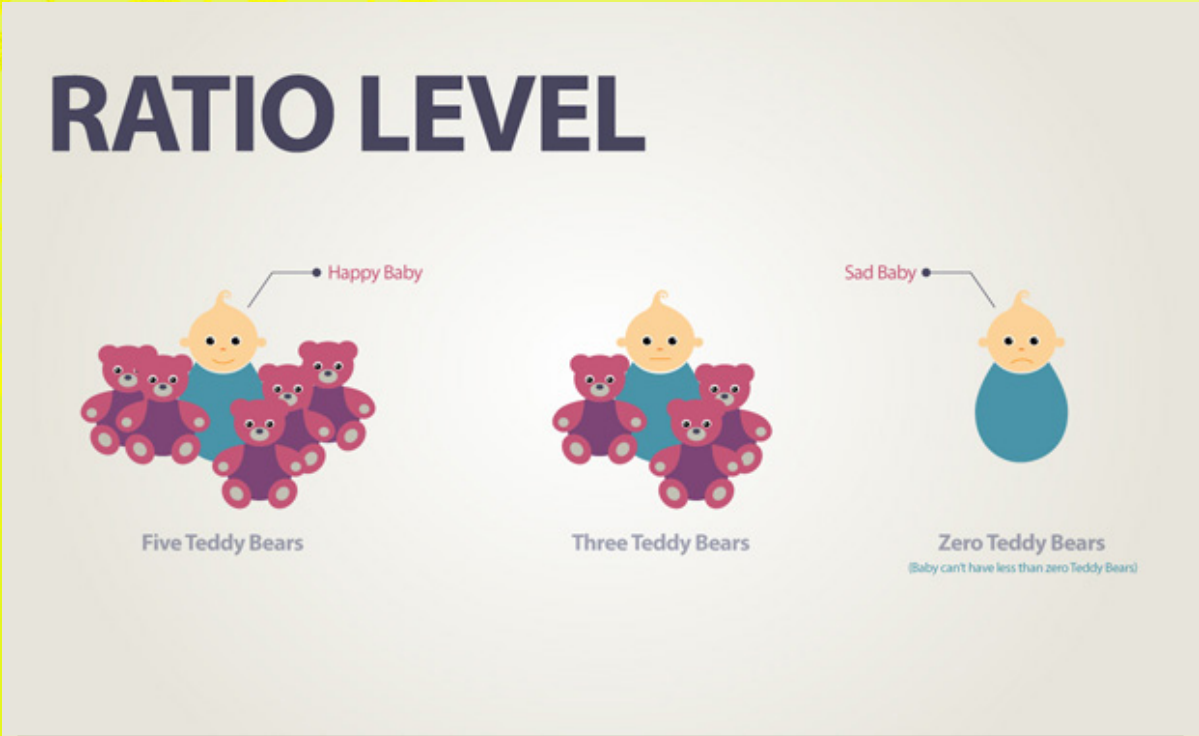


Figure: Ratio level of measurement

- The amounts of teddy bears a certain child has.
- Since we cant have less than zero teddy bears, then the ratio level has a true zero.

- **According to Bertin, The components are characterized by following way:**
  - Nominal Variables $\rightarrow$ Qualitative Components(N)
  - Ordinal Variables $\rightarrow$ Ordered Components(O)
  - Interval and Ration Variables $\rightarrow$ Quantitative Components(Q)