DC 2A and MC 2A

CLASSIFICATION OF DATA

IN GEOGRAPHY

By
SYFUJJAMAN TARAFDER

What is Data?

Data is a set of observation pertaining to an individual (object, area or anything).

Data may be-

- **Purely Quantitative** (Textual), Nominal Data
- **Semi Quantitative** (Rank), Ordinal Data
- **Purely Quantitative** 1. Interval Data (Interval is important, not the ratio); 2. Ratio Data (Interval and Ratio both are important)

Geographical Data

Geographical data is the data pertaining to geographical phenomena or object or area.

- **Non-Spatial** (Data without spatial attributes. Sectoral data.). Statistical analyses are performed.
- **Spatial** (Data of objects, points or area with spatial/locational attributes. District wise Population.

N. B. All data pertaining to anything on earth's surface have spatial properties but when we use it without its spatial properties, it becomes non-spatial.

Data is the information about geographical features. Data is important for **quantitatively**-

- Describing geographical area
- Describing Geographical object
- Describing a group.

Classification/ Grouping of Data

Huge volume of data are grouped and tabulated for

- summarise the data
- organise the data in meaningful ways and so that similarities and dissimilarities can be readily apprehend.
- categorise the data based on its characteristics.
- Comparing data
- Statistical treatment of the data

Data can be classified in four different ways:

- **Time basis**: Time points (Hour, day, month, year etc.)
- **Geographic basis**: Administrative area (Village, Block, District, State, Country)
- **Qualitative basis**: By attributes (Age, Sex, Caste…)
- **Quantitative basis**: Magnitude of variation

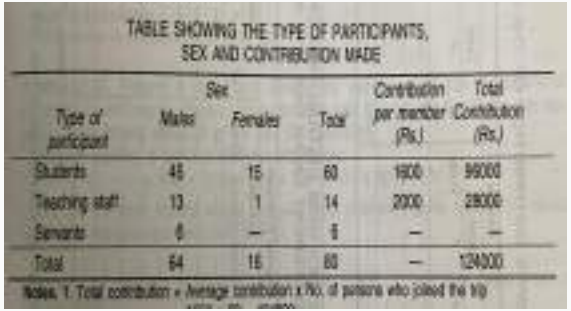
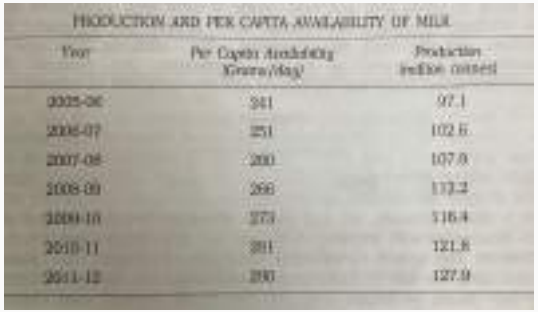


TABLE SHOWING THE TYPE OF PARTICIPANTS, SEX AND CONTRIBUTION MADE

Type of participant	Sex		Total	Contribution per member (Rs.)	Total Contribution (Rs.)
	Males	Females			
Students	45	15	60	1600	96000
Teaching staff	13	1	14	2000	28000
Servants	6	–	6	–	–
Total	64	16	80	–	124000

Notes. 1. Total contribution = Average contribution x No. of persons who joined the trip



PRODUCTION AND PER CAPITA AVAILABILITY OF MILK

Year	Per Capita Availability (Grams/day)	Production (million tonnes)
2005-06	241	97.1
2006-07	251	102.6
2007-08	260	107.9
2008-09	266	112.2
2009-10	273	116.4
2010-11	281	121.8
2011-12	290	127.9

N. B. The grouping of data into different classes or groups is important to summarise the data and organise in a meaningful way.

QUANTITATIVE CLASSIFICATION

There are many ways of grouping or organising data into different classes, and the success of analysis depends on effective and meaningful groupings of data.

Classification is important for-

- Statistical analysis, and
- Mapping

Quantitative Classification

Exclusive Class boundary: Upper limit of each class is exclusive of that class.

10- 15

15- 20

20- 25

25- 30

Inclusive Class boundary: Both upper and lower limit for inclusive in that class.

10- 14

15- 19

20- 24

25- 29

Class Interval/ Width

Both the above classification schemes, there may have either equal or unequal class intervals.

Equal Class Interval:

Simple- (Chose common class interval based on range in the dataset and the chosen number of classes.)

- 10- 15
- 15- 20
- 20- 25
- 25- 30

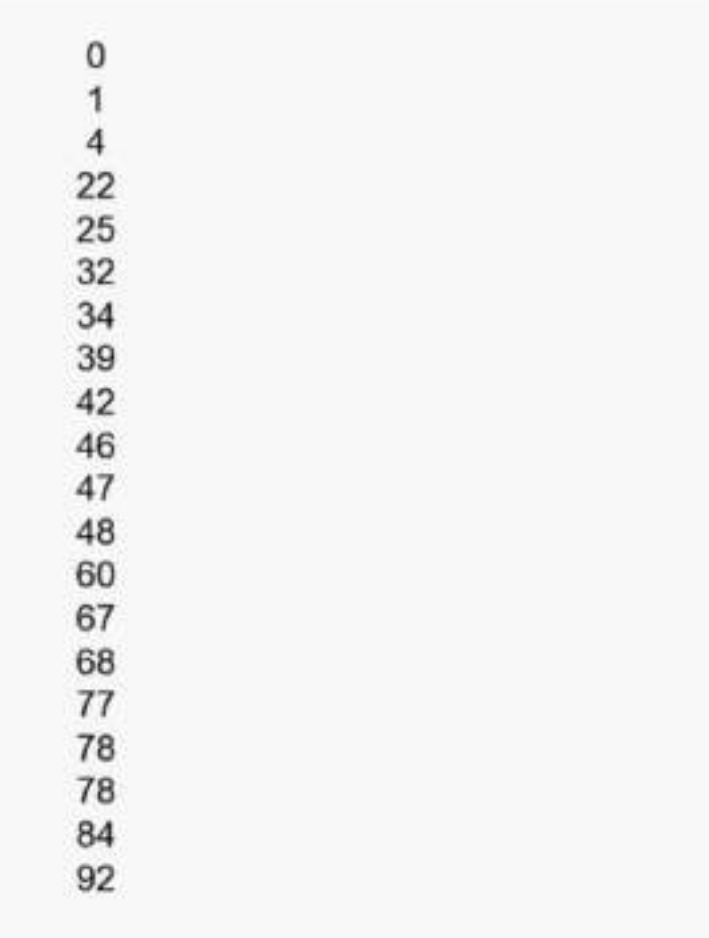
- 10- 14
- 15- 19
- 20- 24
- 25- 29

Unequal class widths:

Arbitrary- (Chose class widths in each class based on the distribution of values.)

- Large widths in initial classes,
- Small widths in initial classes,
- Gradual increase in width

0
1
4
22
25
32
34
39
42
46
47
48
60
67
68
77
78
78
84
92



Exercise

Class Interval/ Width

Both the above classification schemes, there may have either equal or unequal class intervals.

Equal Class Interval:

Standard Deviation- (Identify class boundary using mean and standard deviation)

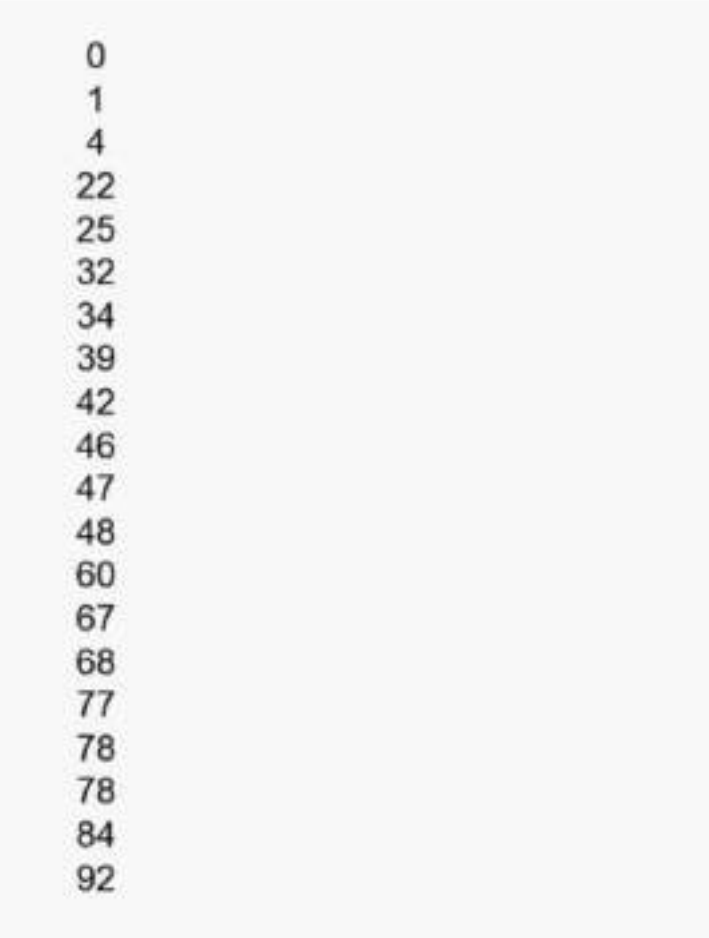
e. g. If, mean= 10.5; SD= 2.34;

Unequal class widths:

Geometric Progression- (Increasing class width in successive classes following geometric progression of initial width)

- 10-20
- 20-40
- 40-80
- 80-160

0
1
4
22
25
32
34
39
42
46
47
48
60
67
68
77
78
78
84
92



Exercise

Class Interval/ Width

Both the above classification schemes, there may have either equal or unequal class intervals.

Unequal Class Interval:

Quantile- (Equal number of observations rather than equal width. Quartiles- Q1, Q2 (Median), Q3 are used as boundary of the classes.

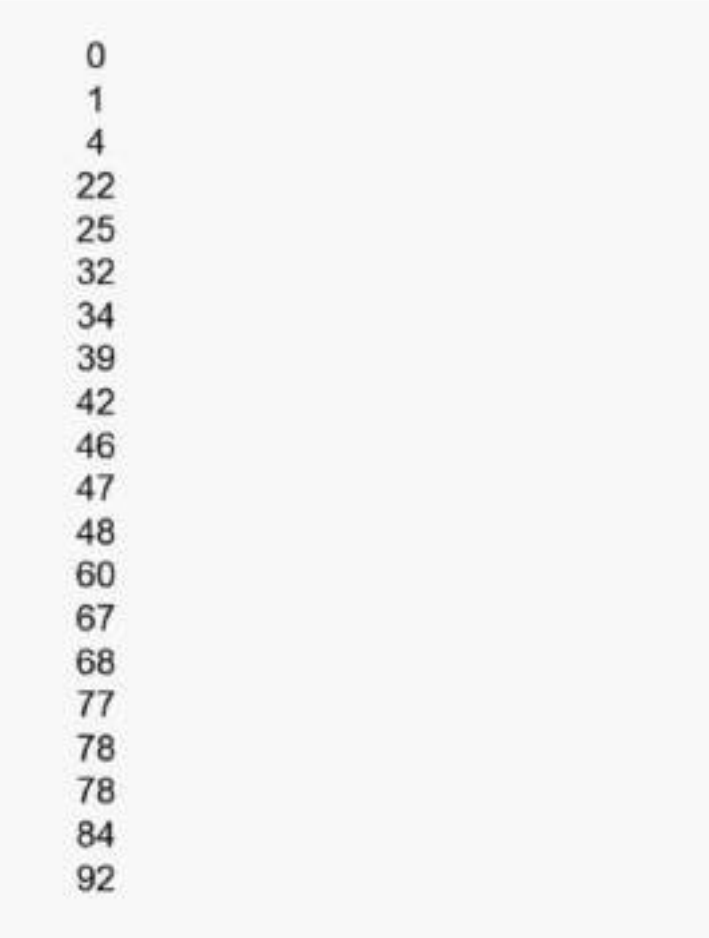
- Lowest- Q1
- Q1- Q2
- Q2- Q3
- Q3- Highest

Unequal Class Width:

Natural breaks- (Consider largest gaps between successive values and create class limits).

- Largest gap
- Next larger gap
- Next larger gap

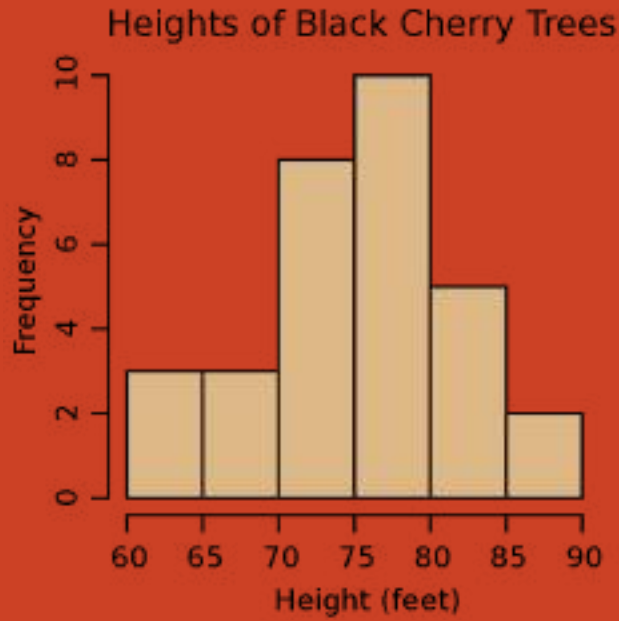
0
1
4
22
25
32
34
39
42
46
47
48
60
67
68
77
78
78
84
92



Exercise

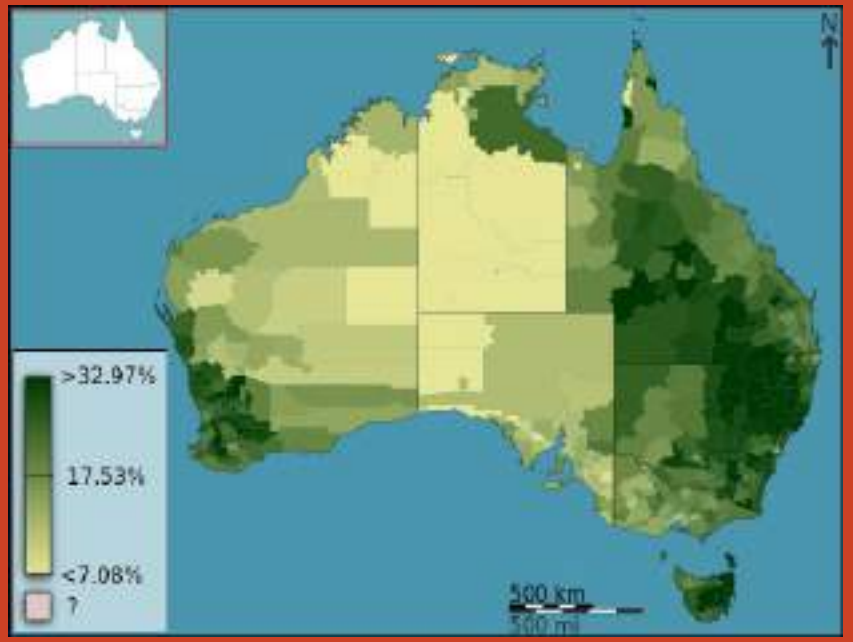
Final point

Classification helps to understand the distribution of observations within the entire range)

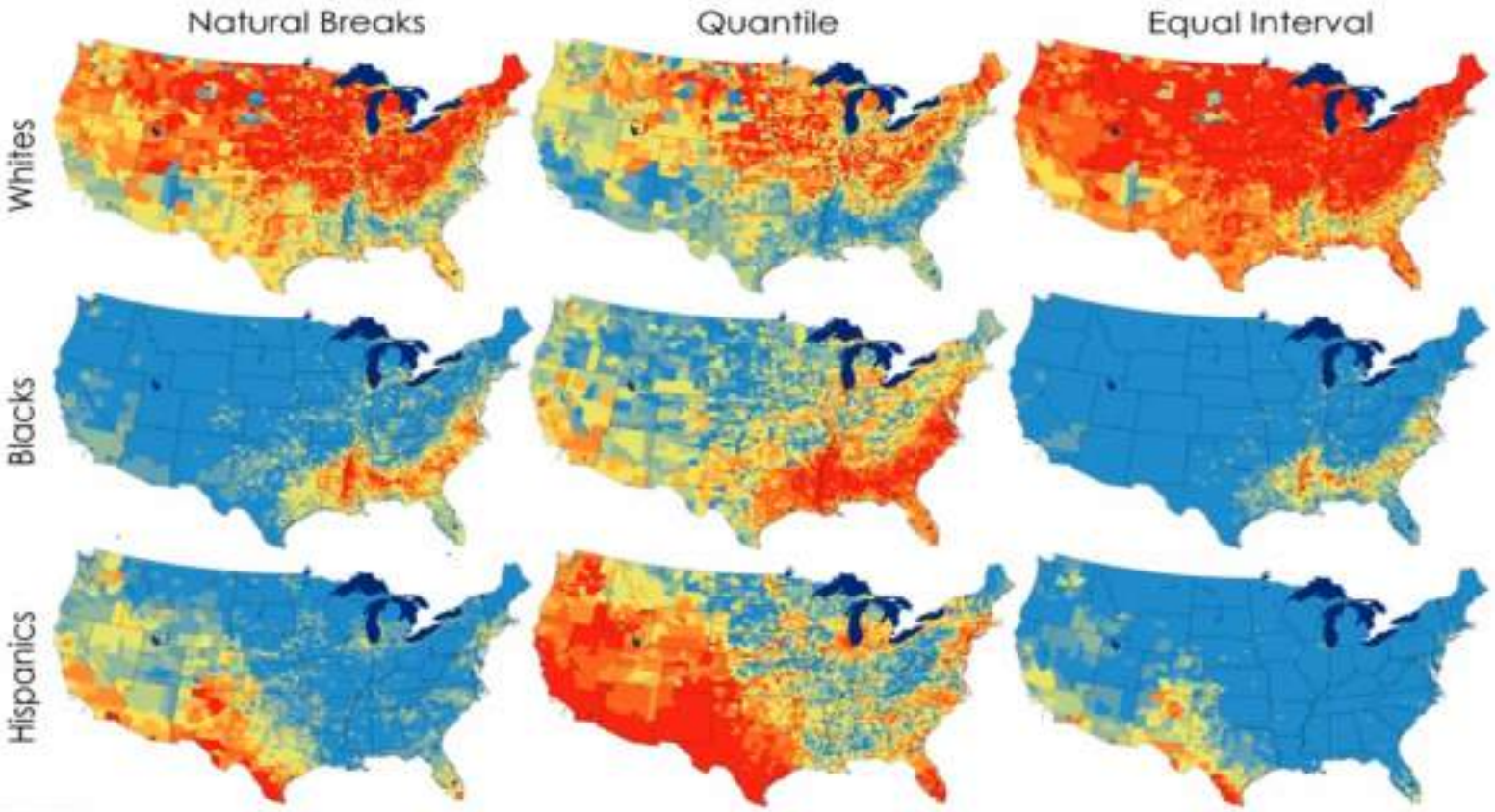


Final point

Classification helps grouping spatial units, thus helps preparing thematic maps.



How does the classification scheme alter the perception of race/ethnicity distributions across the country?



http://ucgis.edu/~mlivolsi/lab1.html

Thanks!

Presenter:

Syfujjaman Tarafder,
Assistant Professor
Gour Mahavidyalaya