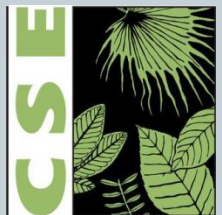# How to tell a story with numbers

- Identifying the story

- Identifying the numbers

- Identifying information sources
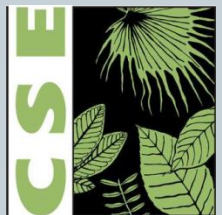
Interview the data

Data does not create meaning, people do

- Why was the data created and is it reliable?

- Who has created this data? (sources are sprouting like mushrooms)

- How was it created? sample size, duration etc

- Just like you would interview a source carefully to make sure that the information is good

Getting the raw data

Challenges

- Logical progression of ideas

- Getting the key ingredients (data, assisting data, case studies, images etc)
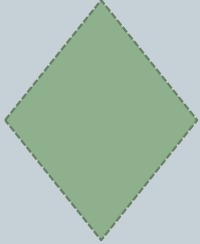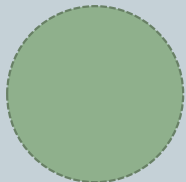
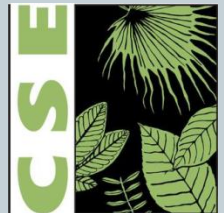Possible solution

- Use of flowcharts
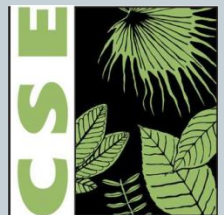
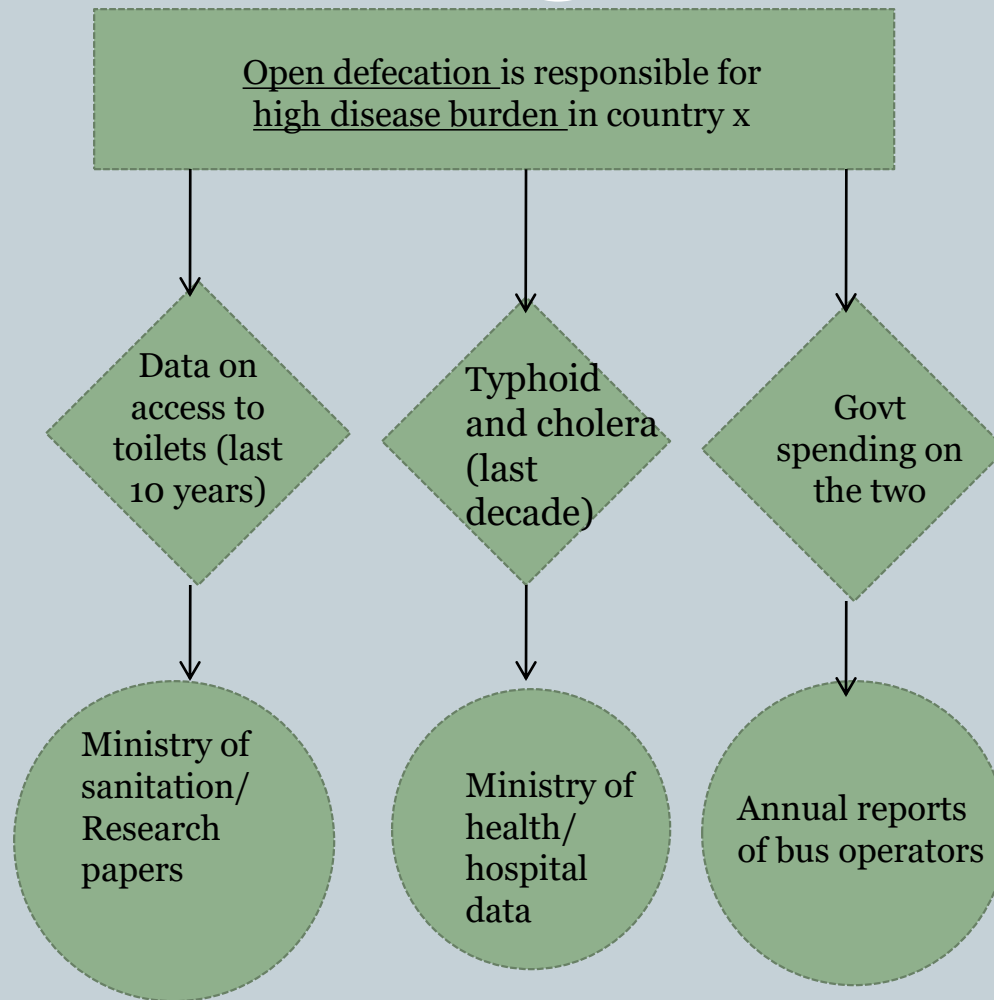# Flowcharts

Central idea/ premise

What are the numbers needed to tell the story

Information sources

Flowcharts

Class work

In each table:

Come up with 1 story idea from Churu visit

Create a flowchart around it

Analyse data

Analyse the story

Popular kinds of stories

Outliner stories

Trend stories

Correlation stories

Outliner stories

A value that is different from all the others

- Which city has the least crime?
- Why do students from this school have such good grades?

Eg: Rankings

A couple in Gujarat got married in just Rs 500

Trend stories

A trend is a pattern through time

- Crime has been decreasing over the last 10 years
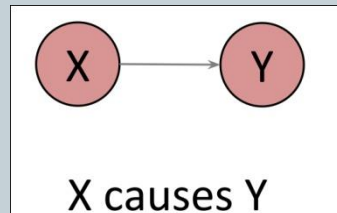- Has rabies deaths increased in India

Correlation stories
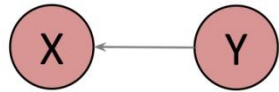
A correlation is when two variables change together

- More smoking causes more cancer.
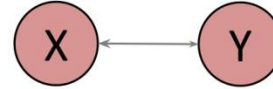- People buy more umbrellas when it's raining.

X causes Y

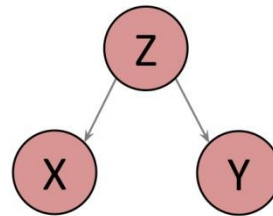Most popular type of correlation

Correlation stories (contd.)
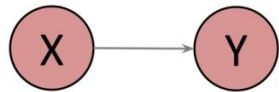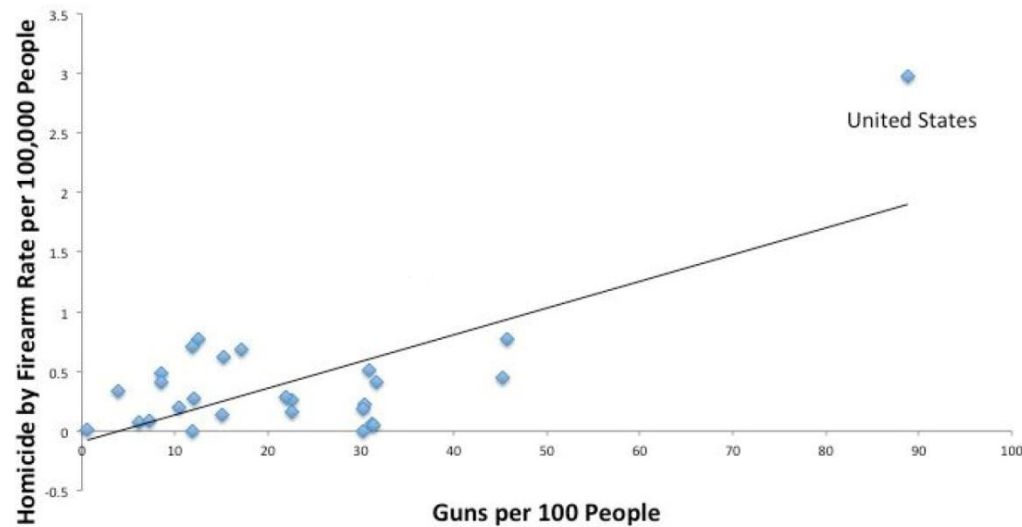

Y causes X

they cause each other

random chance

Z causes X and Y

Correlation stories (contd)

Guns and firearm homicides?

If you have a gun, you are likely to use it

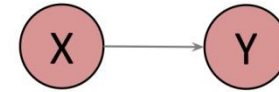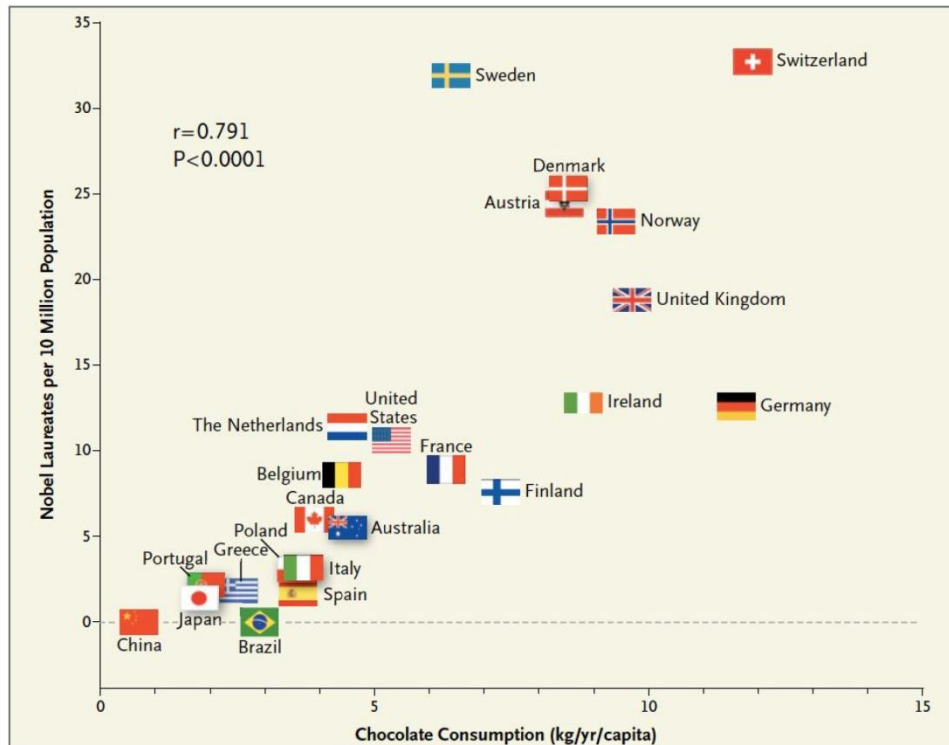If it is a dangerous locality, you will buy a gun

X causes Y

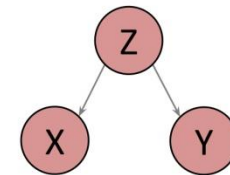Y causes X

Correlation stories (contd)

Chocolate and Nobel prizes

r=0.791
P<0.0001

Nobel Laureates per 10 Million Population

Switzerland
Sweden
Denmark
Austria
Norway
United Kingdom
United States
Ireland
Germany
The Netherlands
France
Belgium
Canada
Finland
Poland
Greece
Australia
Portugal
Italy
Spain
Japan
China
Brazil

Chocolate Consumption (kg/yr/capita)

X → Y

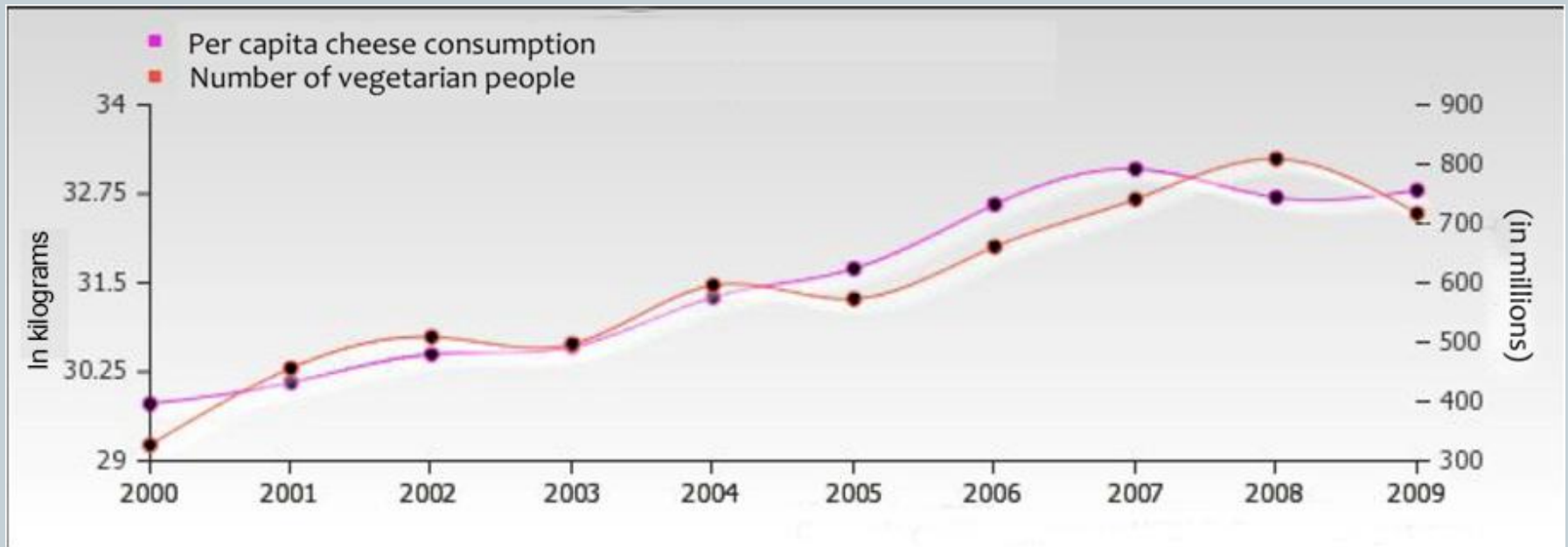X causes Y

Chocolates make
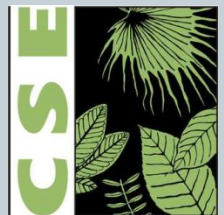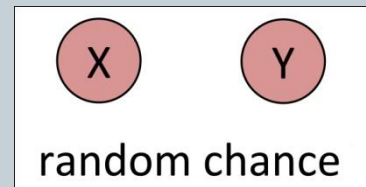you smarter

Z
X   Y

Z causes X and Y

1) Higher income makes
   you eat more chocolate
2) Higher incomes can
   fund better education

Correlation stories (contd)

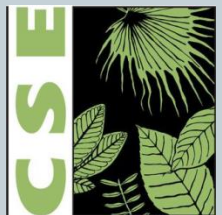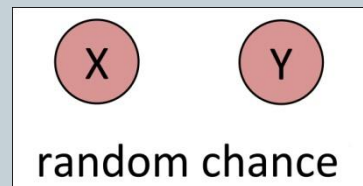There is no way to correlate the to parameters

X   Y

random chance

Correlation stories (contd)

The number of drowning cases increases in town X when the consumption of ice cream increases

There is no way to correlate the to parameters

X   Y
random chance

Correlation is not causation

Hard truths

- Only a small amount of raw data will be used finally for visualisation

- You will have to priorotise (choose between) raw data

- Always give a context... even if it 'seems' unnecessary

- Whenever you are using big numbers, try to give a reference for correlation (1,484 km²)

- Places to look for side stories:
Money, operation size, outreach, impact
Court judgements, laws
Case studies, geaographical locations
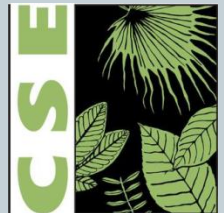
Analysis contd

- Look at percentages

- Look at comparable data to see the abberation

- Mean (average), median (middle value) and mode (most frequent value)

Analysis contd

- Always double check your final analysis… especially to ensure that no biases have crept in

- Also try to answer the WHY of the final analysis

Thank You

Rajit Sengupta

Down To Earth, Centre for Science and Environment, New Delhi, India

rajit@cseindia.org