Working with Data
A Workshop on Understanding, Visualising and Mapping Data
Chennai, 3-5 August 2012

First day
3rd August 2012

1. What is data?   [10:30-11:15]
2. What are data formats?   [11:30-12:00]
3. Presentation by Prajnya   [12:15-12:45]
4. What can be done with data?   [14:00-14:45]
5. Working with quantitative data   [15:00-15:45]
6. Working with spatial data   [16:00-16:45]

Working with Data
A Workshop on Understanding, Visualising and Mapping Data
Chennai, 3-5 August 2012

First day
3rd August 2012

1. What is data?	[10:30-11:15]
2. What are data formats?	[11:30-12:00]
3. Presentation by Prajnya	[12:15-12:45]
4. What can be done with data?	[14:00-14:45]
5. Working with quantitative data	**[15:00-15:45]**
6. Working with spatial data	[16:00-16:45]

Quantitative data

- variables with numeric values
- unit / standard

Quantitative data formats

- .xls
- .xlsx (.xml)
- .dta
- .ods
- .txt (comma-separated, tab-separated, fixed-width)

Quantitative data formats

- .xls
- .xlsx (.xml)
- .dta
- .ods
- .txt (comma-separated, tab-separated, fixed-width)

Quantitative data formats

- **.xls**
- **.xlsx** (.xml)
- **.dta**
- **.ods**
- **.txt** (comma-separated, tab-separated, fixed-width)

Quantitative data formats

- **.**xls
- **.**xlsx (.xml)
- **.**dta
- **.**ods
- **.**txt (comma-separated, tab-separated, fixed-width)

Softwares to work with quantitative data

- Microsoft Excel
- LibreOffice Calc
- Gnumeric
- SPSS
- Stata
- R (language, interpreter, plotter, software)
- Google Refine
- Data Wrangler
- Google Public Data Explorer
- ManyEyes
- Tableau Public

Softwares to work with quantitative data

- Microsoft Excel
- LibreOffice Calc
- Gnumeric
- SPSS
- Stata
- R (language, interpreter, plotter, software)
- Google Refine
- Data Wrangler
- Google Public Data Explorer
- ManyEyes
- Tableau Public

Softwares to work with quantitative data

- Microsoft Excel
- LibreOffice Calc
- Gnumeric
- SPSS
- Stata
- R (language, interpreter, plotter, software)
- Google Refine
- Data Wrangler
- Google Public Data Explorer
- ManyEyes
- Tableau Public

Softwares to work with quantitative data

- Microsoft Excel
- LibreOffice Calc
- Gnumeric
- SPSS
- Stata
- R (language, interpreter, plotter, software)
- Google Refine
- Data Wrangler
- Google Public Data Explorer
- ManyEyes
- Tableau Public

Softwares to work with quantitative data

- Microsoft Excel
- LibreOffice Calc
- Gnumeric
- SPSS
- Stata
- R (language, interpreter, plotter, software)
- Google Refine
- Data Wrangler
- Google Public Data Explorer
- ManyEyes
- Tableau Public

Working with quantitative data

- Collecting
- Storing
- Analysing
- Visualising

Working with quantitative data

- Collecting [scraping]
- Storing [databases]
- Analysing
- Visualising

Scraping

- Computer program to extract data.

- Understand the data, model the script.

- Inelegant. Last resort.

- Screen Scraping – from what is visually visible.

- Web Scraping – from HTML, XHTML.

Scraping Ethics

- Respect robots.txt.

- Timed access.

- Respect the license.

- No redundant downloads.

- Read ToS, ToU.

- Respect the privacy policy of the target.

- Respect response from the target.