History of R

Advantages of R

Use of R in an NGO

Data Analysis

* R is created to deal with data
* Unlike other programs, its not package
* It's a full fledged programming language
* Particularly developed to do deal with data and statistic

Data visualization

* Communication of data analysis is hugely dependent on dat
* Tables are good for summarizing results
* Not good for communication
* R's graphics capability is unmatched compared to any othe

Report Writing

* Now comes the most crucial part of report writing
* Non-profit, NGOs has to write sleuth of reports based on
* Typical work flow
* Excel, Microsoft word
* Basic calculation, preparation of tables in excel
* Formatting of tables in Excel
* Formatting of tables in word
* SPSS-EXCEL-MS-Word Workflow
* Generate tables and cross-tabs in SPSS.
  * Export those to excel
  * Formatting done in Excel
  * Copy and pasted in Excel
* STATA-EXCEL-MS-Word Workflow
* STATA has packages which transforms raw tables in publica
* But still has to be formatted in excel
  * Unless used LaTeX (text processing engine)
* Then copy and pasted in MS-Word

Problem with workflow

* Irrespective of the workflows mentioned above
* All suffer from a common problem
* The reports are not reproducible
* For example, let's you have written a report which inclu
* 50 Tables
* 20 Charts
* This is report that has to be generated Every quarter
* Therefore in a given workflow in a whole year
* 200 Tables and 80 charts has to be created in
* STATA and SPSS has scripting environment where all the co
* When data changes, the scripts can be run and the all tab
* But the catch is those are still raw tables
* Raw tables and graphs code, you write for 50 tables and 2
* Then reproduce those raw tables and charts by running the
* But the problem is that those are still raw tables
* Therefore, exporting those raw tables still has to be don
* Therefore export raw tables to Excel, formatting those ar
* Graphs has to be exported to images files("JPG","PNG" etc

R workflow: DRY

* DRY - Don't Repeat Yourself
* Coding/principle: don't write the same code more than onc
* R is a programming language - follows the same principle
* In the above workflow: don't repeat Table, don't repeat c
* You may say, "Hey its not the same chart, the data are di
* Well the numbers are different but it has exactly the sam
* So the output structure is same, input changes
* Ideal/ripe for DIY

R workflow: Reproducible research

* Idea comes from literate programming
* Description of the code and the code comes together
* Early implementation in LaTex and R, resulting into Rnw f
* LaTeX has a steep learning curve, easier is Rmarkdown.
* The whole report will be a single markdown script.
* Descriptions will be interlaced by the R codes depicting
* Whenever ready, these Rmarkdown files will be knitted to
* Docx
* PDF
* HTML
* Therefore every element of a document is in single place.
* Ideally, we would just change the dataset and "Knit".
* Voila, the next quarter report is ready!
* Isn't that simply magical!

Reading Data

* R can directly read texts and data from PDF
* From websites, called web scraping.
* All the major data sources has API for R
* WDI
* FAO
* UNDP
* And many more.

R has image processing capacity

* Recent additon to R family imagemagick package can do adv
* Help to create high quality infographics from scripting e

Comparison between R and SPSS

Comparison between R and STATA

R is not owned by a company, its owned by its users

Prediction

Problems in real life

User experience