

# Analysis of Industrial Sector in Bangladesh

Rushad Faridi, PhD

University of Dhaka

May 29, 2017

# Outline

- 1 History of R
- 2 Advantages of R
- 3 Use of R in an NGO
- 4 Comparison between R and SPSS
- 5 Comparison between R and STATA
- 6 R is not owned by a company, its owned by its users
- 7 Prediction
- 8 Problems in real life
- 9 User experience

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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

# Data Analysis

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# Data Analysis

- R is created to deal with data

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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

# Data Analysis

- ▶ R is created to deal with data
- ▶ Unlike other programs, its not package

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# Data Analysis

- ▶ R is created to deal with data
- ▶ Unlike other programs, its not package
- ▶ It's a full fledged programming language

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# 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 statistical analysis

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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

# Data visualization

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R



# Data visualization

- Communication of data analysis is hugely dependent on data visualization

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# Data visualization

- Communication of data analysis is hugely dependent on data visualization
- Tables are good for summarizing results

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# Data visualization

- ▶ Communication of data analysis is hugely dependent on data visualization
- ▶ Tables are good for summarizing results
- ▶ Not good for communication

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# Data visualization

- ▶ Communication of data analysis is hugely dependent on data visualization
- ▶ Tables are good for summarizing results
- ▶ Not good for communication
- ▶ R's graphics capability is unmatched compared to any other statistical packages

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# Report Writing

- \* Now comes the most crucial part of report writing
- \* Non-profit, NGOs has to write sleuth of reports bas
- \* 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 p
- \* But still has to be formatted in excel
  - \* Unless used LaTeX (text processing engine)

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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

# 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
  - \* 50 Tables
  - \* 20 Charts
- \* This is report that has to be generated Every quart
- \* 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
  - \* When data changes, the scripts can be run and the a
  - \* But the catch is those are still raw tables
  - \* Raw tables and graphs code, you write for 50 tables
  - \* Then reproduce those raw tables and charts by runni
  - \* But the problem is that those are still raw tables
  - \* Therefore, exporting those raw tables still has to

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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

# R workflow: DRY

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

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

# R workflow: Reproducible research

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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

- \* Idea comes from literate programming
- \* Description of the code and the code comes together
- \* Early implementation in LaTeX and R, resulting into
- \* 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 into
- \* Docx
- \* PDF
- \* HTML
- \* Therefore every element of a document is in single file
- \* Ideally, we would just change the dataset and "Knit" the report
- \* Voila, the next quarter report is ready!
- \* Isn't that simply magical!



# Reading Data

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

- \* 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 addition to R family imagemagick package can
- \* Help to create high quality infographics from scrip

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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  
R

Analysis of  
Industrial  
Sector in  
Bangladesh

Rushad Faridi,  
PhD

History of R

Advantages of  
R

Use of R in an  
NGO

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

- \* STATA costs money, R is free
- \* STATA is easy to learn, GUI
- \* STATA presents results more user-friendly way
- \* The thing is that computing is changing drammmatical
- \* Statistics also changed dramatically
- \* The pace will only increase
- \* Notable changes :
- \* Much larger dataset
- \* Increased use of computer intensive analysis
- \* Data mining
- \* Dynamic computer graphics
- \* Automatic report writing
- \* Websites for accessing databases
- \* Websites for presenting results of statistical anal
- \* In all the above respects, R is ahead of any other

- \* The users drive R forward
- \* No company can compete with it, cause no company
- \* STATA will become tool for use in introductory st
- \* For use by people who only want to do relativey b
- \* You learnt hours/days/months of time spent in lear
- \* Then one fine morning you found that your could S
- \* R is only limited by the power of your computer.
- \* Students need more help with basic tasks in R
- \* On the other hand Stata is well designed makes it
- \* Stata becomes cumbersome when it is asked to do a
- \* R requires lots of basic skills to do simplest of
- \* But comes of its own when you do non-standard tas