R-language for statistical computing and visualization

Markus Kainu

September 2013

Contents

1 Open Research Methods in general

2 What is R? - Origin and characteristics

2.1 Origin

- R-language was iniated by two
- open source

2.2 R-language as a programming language

- object oriented, s-language bell laboratories (G)UI's, IDE Rstudio,
- contributed packages

2.3 R-language as a tools for statistical computing

- structure
- visual
- open source, community, licensing, teaching

2.4 Popularity of R-language

- enterprise level services

3 Who makes the R possible? - R-project

3.1 Organisation of the project

- development vs. user help

3.2 Development of the language

3.3 User support

- mailing lists - general vs. special interest groups blogs
- q & a sites

3.4 Contributed packages

- CRAN - task views R-forge
- Github
- bioconductor

4 What is R used for? - R in action

Why R-language is popular

4.1 Development of statistical methods

- new methods implemented first in R

4.2 Applied statistics

4.2.1 Bio/geo-sciences

- Geograhical Information Systems

4.2.2 Social sciences/economics

4.2.3 Humanities - analysis of natural languages

4.3 Business/enterprise analytics

- insurance, big data, banking, industry
- social media: facebook, google, twitter

4.4 Data journalism

- Guardian, New York Times, Chicago Herald Tribune

5 How does it work? - Visualising data in humanities using R-language

5.1 Word clouds

5.2 Networks maps

5.3 Spatial visualisation

5.4 Clustering

5.5