R-language for statistical computing and visualization for social and humanitarian sciences

Markus Kainu

January 23, 2014

Abstract

In sit amet sapien mollis ligula dignissim vehicula. Fusce turpis dui, semper consequat erat eget, laoreet rutrum diam. Maecenas neque urna, convallis quis est rhoncus, imperdiet interdum lorem. Curabitur tempor nisi sit amet est feugiat, sed pharetra neque fringilla. Etiam at enim mattis, pretium nulla eu, bibendum tellus. Morbi sed volutpat augue. Phasellus porta pretium nulla vel tincidunt. Pellentesque vel fermentum ante, non varius nunc.

Contents

1 Open Research Methods in general

The debate on open science in the context of Social and Humanitarian sciences has been predominantly focusing on open access to research publication and opening up the various types of digital research data (open research data). The use of open research methods has received a lot less attention due to obvious reason.

Firstly, research methods that researchers in humanitarian and social sciences have not being "software intensive" as is the case with digital humanities. Second, within the approaches that have used computational methods there has not been too many options to choose from and the use of proprietary analytical software like SPSS, Stata or SAS has been and still is common.

This was sensible choice in the era of closed data (King (2014))

Open research methods are referred as of a wider open science discussion.

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

King, Gary. 2014. "Restructuring the Social Sciences: Reflections from Harvards Institute for Quantitative Social Science." *PS: Political Science and Politics* 47: 165–172. http://journals.cambridge.org/repo_A9100Nlq.