

Computational ecosystems for social science*

Markus Kainu ⁺¹, Joona Lehtomäki ^{‡2}, Juuso Parkkinen ^{§3}, Juha Yrjölä ^{¶4}, Måns Magnusson ^{||5},
and Leo Lahti ^{**6}

¹Aleksanteri Institute, University of Helsinki, Finland

²Department of Biosciences, University of Helsinki, Finland

³Reaktor Innovations Oy, Finland

⁴Kansan Muisti ry, Finland

⁵Linköping University, Sweden

⁶Department of Veterinary Bioscience, University of Helsinki, Finland

May 29, 2015

Keywords: social science; elections; open government data; statistical programming; machine learning

1 Title & authors

[TODO lisätäänkö Mikko Tolonen tekijälistalle? Lisäksi pitää kysyä tahtooko Juha olla nyt mukana kun ei olekaan vaalijuttua + kansanmuistia posterissa]

2 Background

The recent explosion in open data availability has created novel opportunities for research, journalism and citizen science. High-quality machine readable data streams are increasingly available on political decision making, historical processes, welfare, traffic, and other aspects of society. There is a great need for analytical tools to take advantage of these new data streams in computational social science, digital humanities, and related fields.

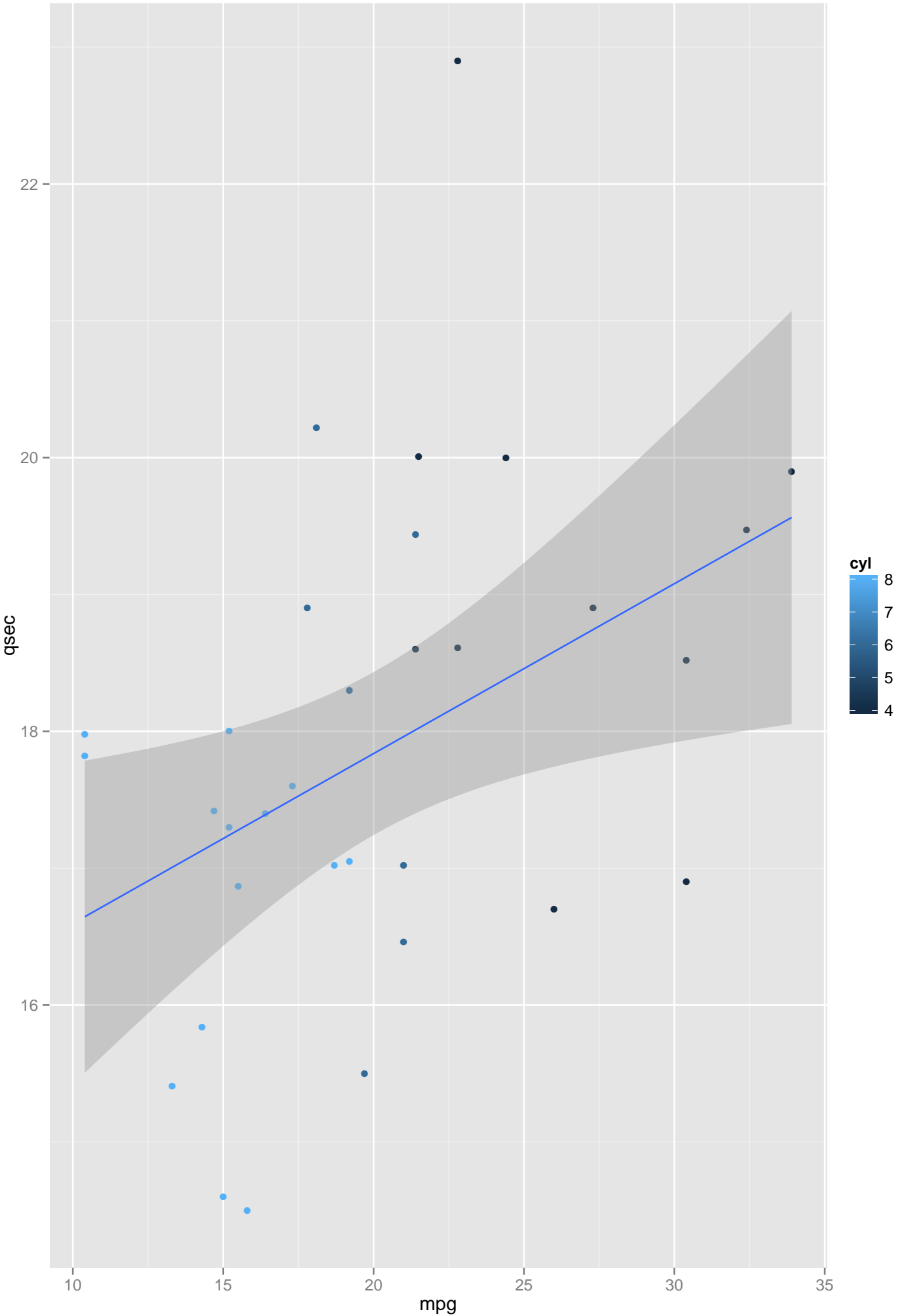
3 Open source data analytics

Efficient data analysis relies on customized, reproducible analysis workflows that are best developed jointly by the user community. Availability of ready-made algorithms for standard data analysis tasks allows an individual researcher to avoid reinventing the wheel, leaving more time to solve the specific research problems. Solutions have emerged in data intensive research fields, such as bioinformatics and particle physics, based on open source statistical programming languages. In computational social sciences and digital humanities, analogous statistical software libraries are now emerging and have a huge potential to contribute to transforming the field. However, these resources are currently highly scattered and come in various formats, hindering wider adoption. Specific web-based tools are available, but more flexible computational tools are urgently needed for fully powered data processing and analysis.

4 rOpenGov developer community and ecosystem

rOpenGov is a statistical ecosystem focused on open source data analysis algorithms relevant to computational social sciences and digital humanities. We provide a discussion forum and flexible algorithms for reproducible data analysis

*Extended abstract submitted for International Conference on Computational Social Science in June 8-11, 2015 in Helsinki, Finland
⁺markuskainu@gmail.com
[‡]joona.lehtomaki@helsinki.fi
[§]juuso.parkkinen@iki.fi
[¶]juha.yrjola@iki.fi
^{||}mans.magnusson@gmail.com
^{**}leo.lahti@iki.fi



1

