The Social Organization of the R Project

John Fox

McMaster University

useR 2008 Dortmund

Introduction

What is Problematic About Open-Source Software Development?

- Typical questions (particularly posed by economists): Why do people (or organizations) participate in open-source software development, and is their participation rational?
- A different point of view: Participation in voluntary associations is a normal social activity.
- What is problematic is why and how a voluntary association can produce a complex, integrated product such as software.

Introduction

Stated Motivations of R-Core Developers

- To leverage one's own efforts by building a mutually useful product:

  > *"[M]y feeling is that I gain great benefit from open source software. This is tremendously valuable to me, being able to use all of these other tools, and I feel both a moral and a practical obligation to contribute back into this sea of tools that are, I think, very important for the development of our profession."*

- An economist might find the "practical obligation" an expression of rationality.

Introduction

Stated Motivations of R-Core Developers

- To work on the cutting edge of statistical computing:

  "[It's] very satisfying ... to work on a day-to-day basis with people with whom one has common interests and can get a lot of pleasure from working with."

Introduction

Stated Motivations of R-Core Developers

- To provide statistical computing facilities to those who could not otherwise afford them:

> *"One of the nicest sort of things [is that] other people in the Philippines or Bolivia or Mexico ... can have a world class statistical software system [when] they could never afford any of the commercial systems."*

Introduction

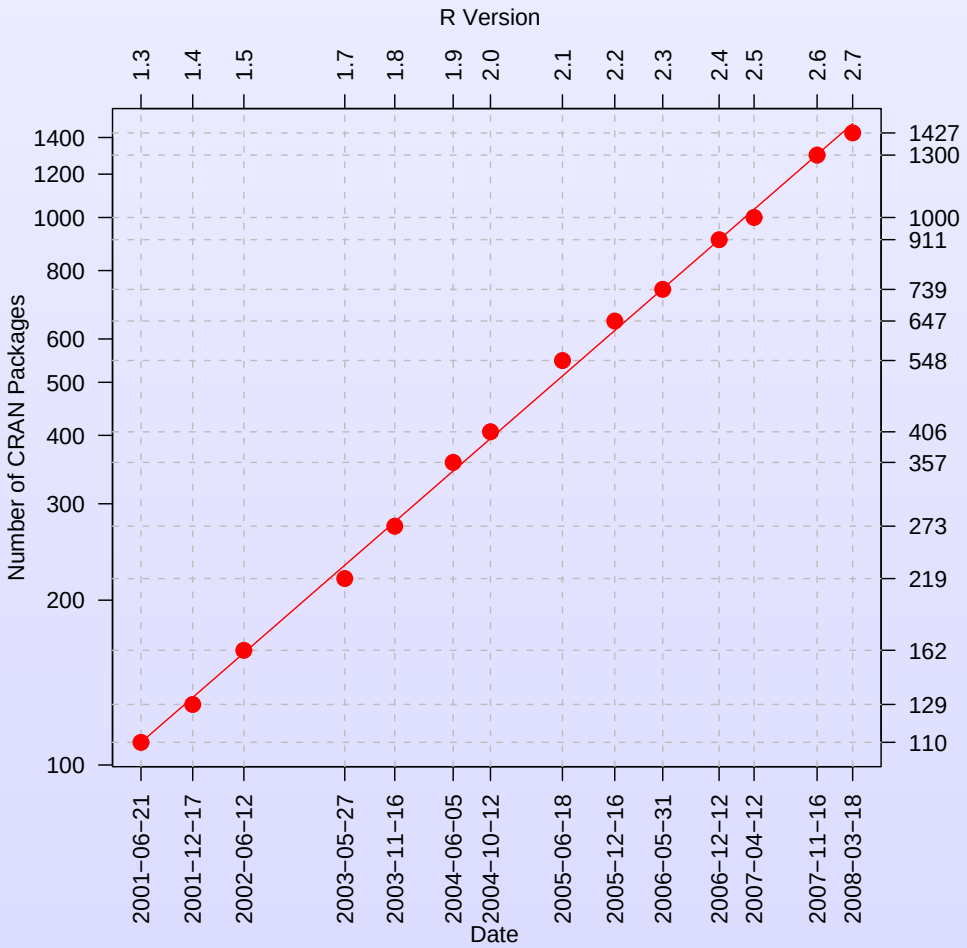
Stated Motivations of R-Core Developers

- Statisticians are habituated to cooperation:

> *"But statistics itself is a collaborative field. You can't actually do anything in statistics, or at least nothing of interest, unless you cooperate with a subject matter expert. So basically . . . cooperation is built into the subject, and that might have had some influence on it, but maybe you have to be predisposed to collaboration if you're going to be in statistics."*

The Trajectory of the R Project

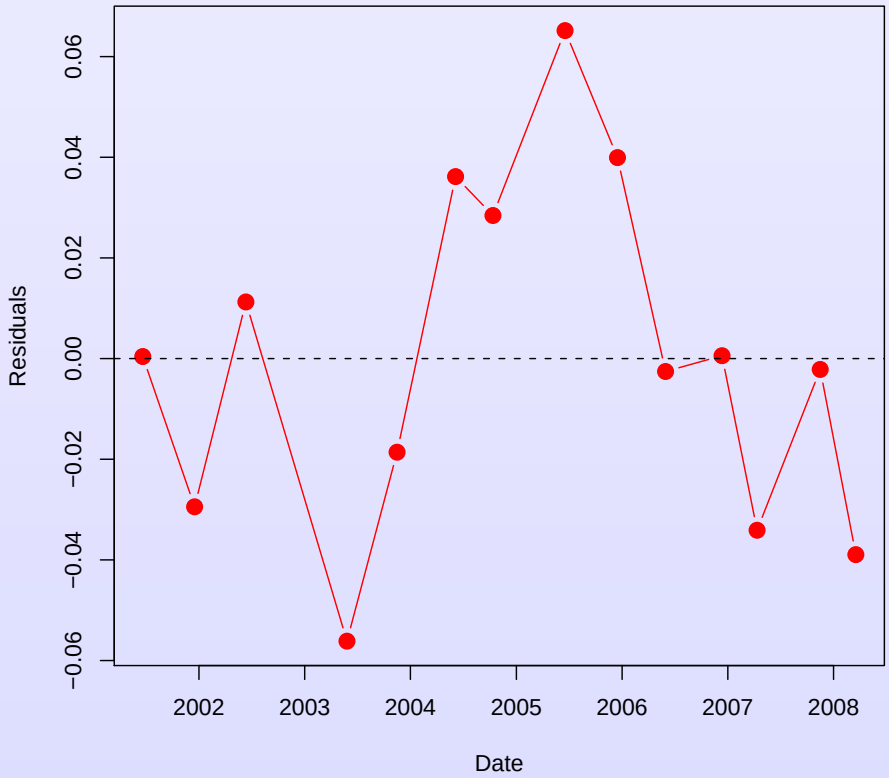
The growth in CRAN packages is approximately exponential



Source of Data: https://svn.r-project.org/R/branches/.

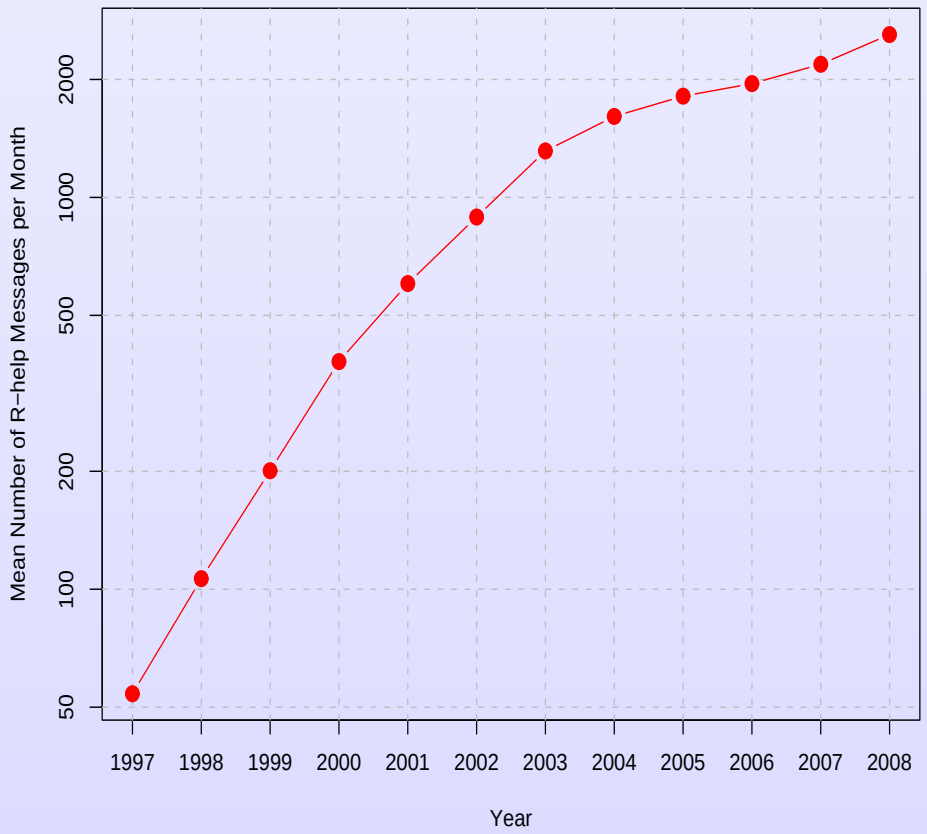
The Trajectory of the R Project

But Tukey would want us to plot the residuals



The Trajectory of the R Project

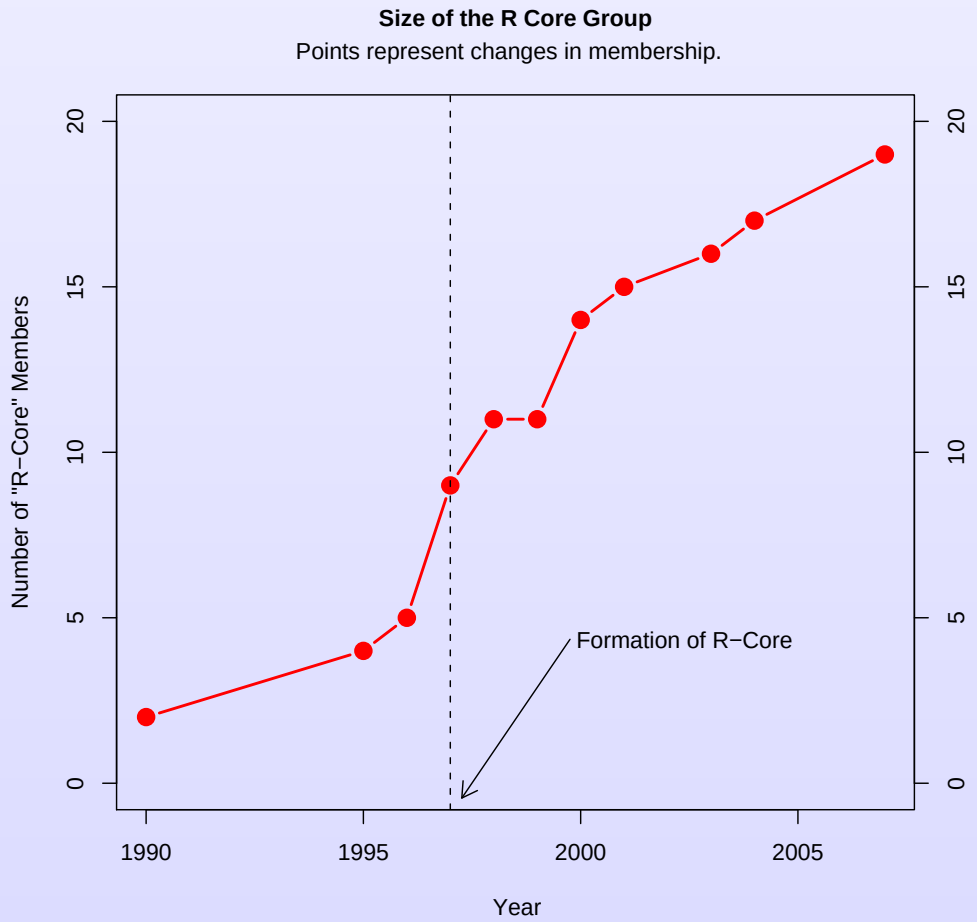
The growth rate in the number of messages on R-help has declined



Source of Data:
https://stat.ethz.ch/pipermail/r-help/.

The Trajectory of the R Project

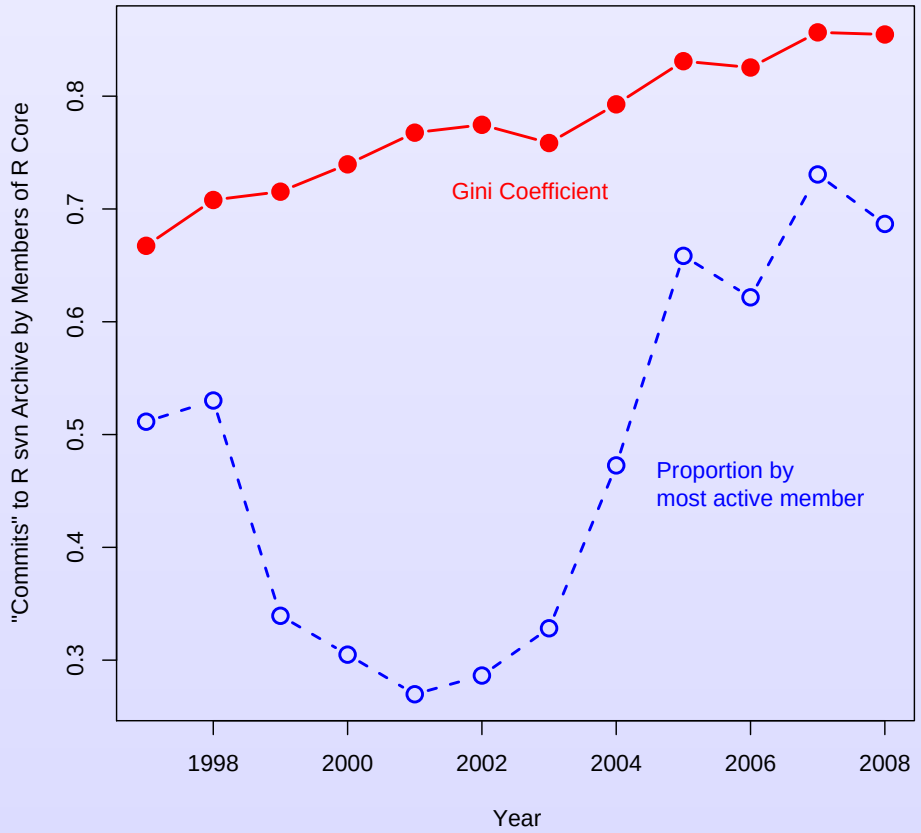
The size of the R Core group has doubled



Sources of Data: Interviews with members of R Core, contributors(), https://svn.r-project.org/R.

The Trajectory of the R Project

Activity in the R svn archive by members of R Core has become more unequal



Source of Data:
https://svn.r-project.org/R

The Development and Organization of the R Project

- A member of the R Core group on how decisions get made:

> *"[W]e [have] a system that [is] democratic but the person who [is] going to do the work [gets] more votes than anybody else."*

Development of the R Project

	Stage		
	Initial	*Transitional*	*R-Core*
Approximate Dates	1990-94	1994-97	1997-
Recruitment	Some student participation	Demonstrated interest	Semi-purposive, by invitation
Division of Labour	None	Developing	Semi-formal
Hierarchy	None	Original developers, contributors	Informal

Development of the R Project

	Stage		
	Initial	*Transitional*	*R-Core*
Principal Mode of Cooperation	Direct collaboration	Anarchic voluntarism	Role enactment + voluntarism
Planning	None	Implicit	Partial
Decision-Making	Joint	Individual	Semi-consensus
Resolution of Disagreements	Discussion	Largely unnecessary	Discussion, preemption, avoidance
Principal Goal	Personal Development	Reproduce functionality of S	Various, partly conflicting

Why Did R Succeed?

- The initial developers opened up the project, eventually forming the R Core group (cf., Octave, LispStat) and releasing R under the GPL.
- The Core group is immensely talented, with complementary skills.
- The project had an initial target: reproducing the functionality of S.
- Much of the necessary software beyond the basic R system was already available in S "libraries" (e.g., MASS, survival, nlme).
- The S language had already penetrated the statistics community.
- S is relatively easy to use (cf., LispStat).

Why Did R Succeed?

- The package system, introduced early on, permitted participation with minimal direct intervention by R Core.
- The package system serves partially to circumvent disputes.
- The R Core group successfully leveraged information technology (e.g., version control, e-mail lists, package automation, distribution via the Internet).
- R clearly improved on S: e.g., lexical scoping, name spaces, package system.
- R runs of all widely used computing platforms (Windows, Mac OS, Linux/Unix).
- R is free (in both senses).

Can This Success Continue?

Positive Factors

- R has a great deal of momentum.
- The basic R system is essentially sound, and much of the dynamism of R is in package development.
- Many of the factors leading to the initial success of R continue to apply (e.g., talent of R Core).
- R has attracted a very large user and developer base.
- R is highly visible (e.g., in books and journal articles).
- R has powerful advocates.

Can This Success Continue?

Negative Factors

- The decision-making procedures of R Core were perhaps better suited to an earlier stage in the development of the software and a smaller Core group
  - E.g., failure to resolve long-standing issues (no multi-threading, weakness in handling very large data sets).
- There is no general plan for the development of R.
- Possible over-dependence on a few key individuals without a clear plan for succession.
- The current organization of CRAN may not be sustainable (e.g., users already suffer from information overload).