Software Engineering Standards in the R Community

by Oliver Keyes, Jennifer Bryan, David Robinson

Abstract Oliver's abstract for useR conference goes here.

Introduction

Introductory section which may include references in parentheses (R Core Team, 2012), or cite a reference such as R Core Team (2012) in the text.

Best Practices

We considered a number of development practices...

- **Tests** Tests allow... [cite testthat].
- **Vignettes** Vignettes provide long-form documentation, such as a tutorial. Developers have the choice of several vignette builders, of which the most popular are the built-in **Sweave** and the **knitr** package. The **knitr** package also offers a choice of format, between LaTeX and R Markdown.
- **roxygen documentation**: **roxygen**, and more recently **roxygen2**, make it easy to...

Practices in CRAN

Some figures and analyses, like Figure **??**.

Vignettes

This section will discuss vignettes.

- Only 1241 out of 6287 (19.7) packages currently provide any vignettes.
- Figure 2 shows the use of vignettes over time. From 2009 onward, vignettes were infrequently found in CRAN, making up less than 5% of packages. In 2014, their use grew sharply to about 20%. Since their inception, knitr and R markdown have each grown in popularity as tools for vignette development, though Sweave and LaTeX are still more common overall.

Testing

- Most packages don't have tests
- Tests over time (graph)
- Some things that will require new data:
- Of those that have "Other", what do those tests look like?
- How many tests do testthat and RUnits have?

Other practices

- Go over advantages and history of
- Semantic versioning
- Upstream repository
- Change log
- Figure 3, practices are becoming more common. GitHub in particular.

Conclusion

- Most packages in CRAN don't take advantage of the newest methods and practices in package development. However, that proportion is rapidly increasing in packages published or updated in the last two years.

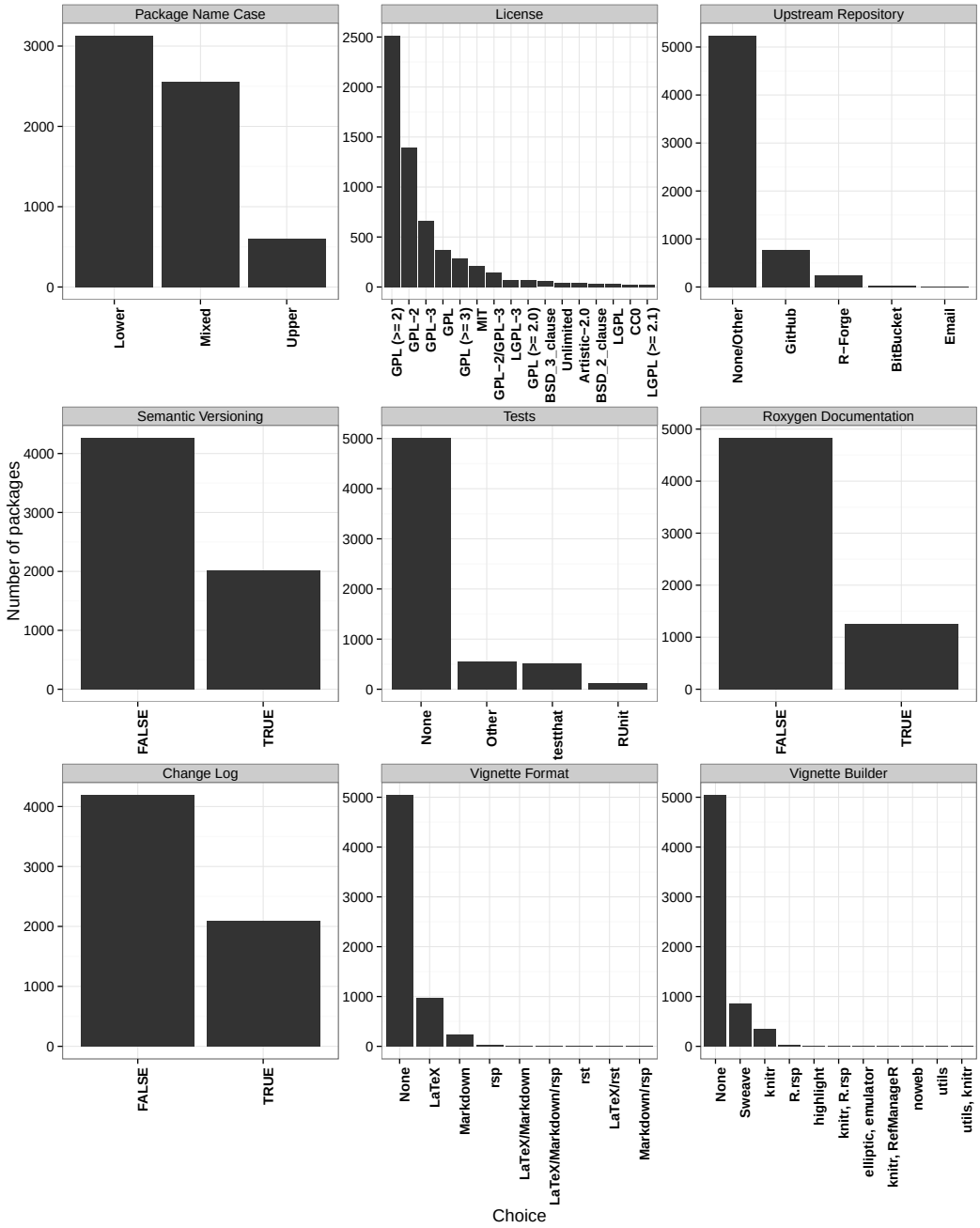

Figure 1: Overview of software development practices on CRAN. Note that only the first 15 choices of license (out of 100 used in CRAN) are shown.

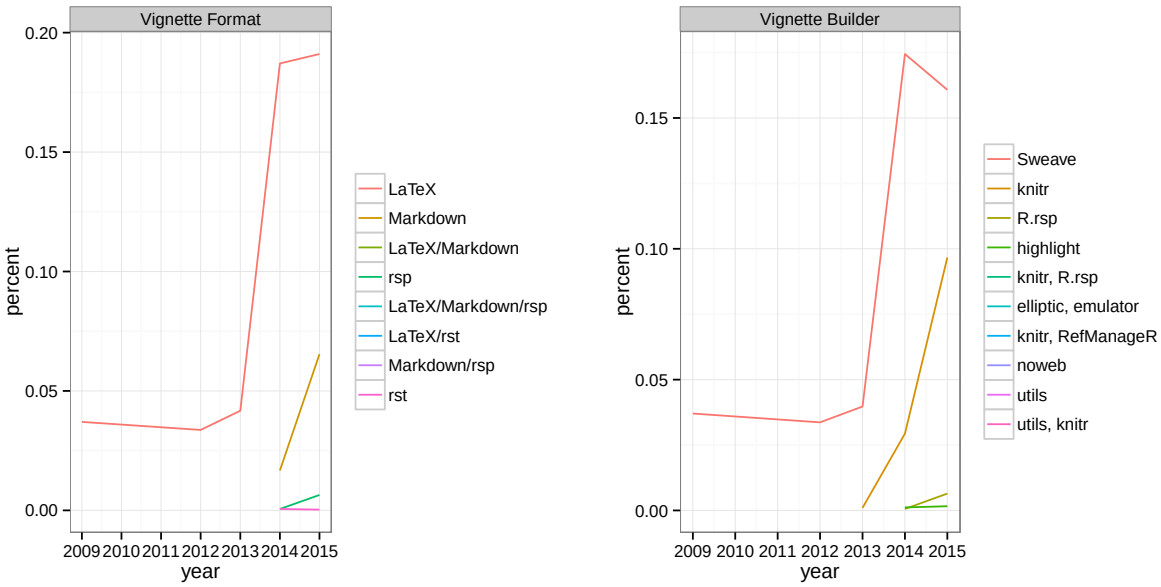

Figure 2: Use of particular vignette formats over time.

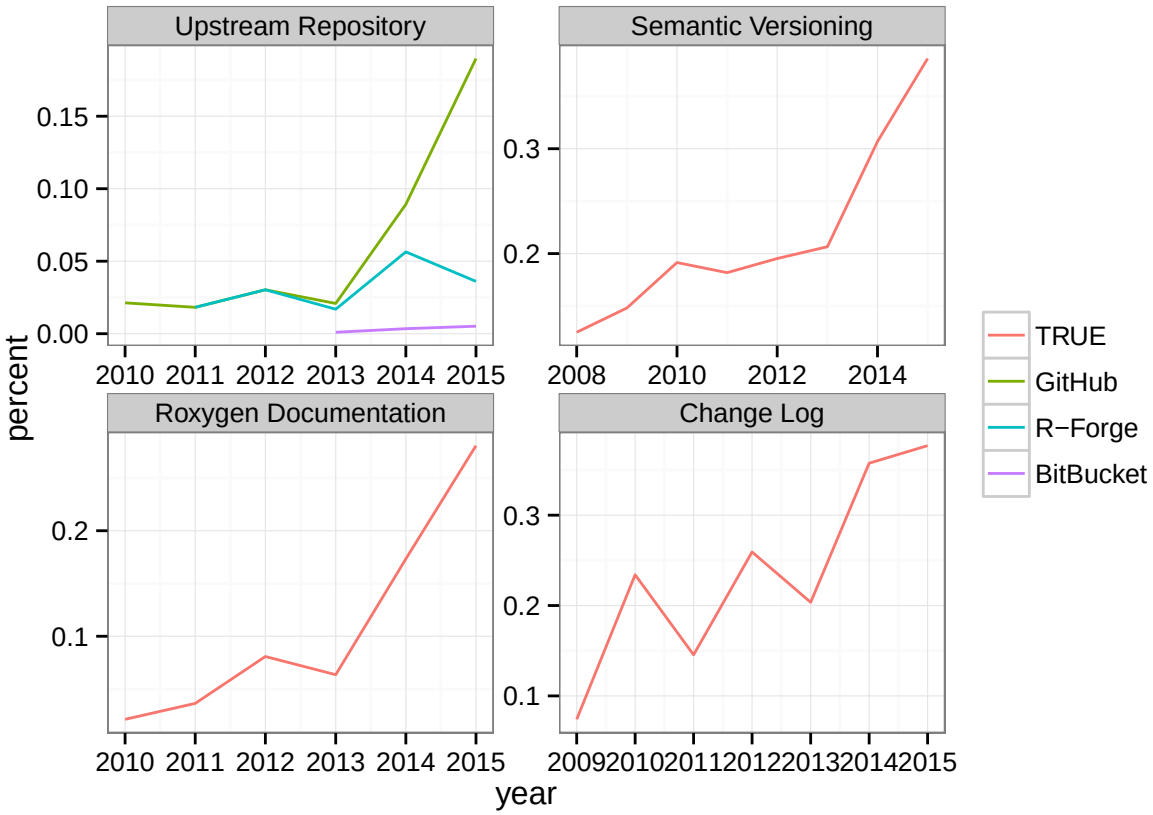

Figure 3: Use of semantic versioning etc over time.

Bibliography

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org/. ISBN 3-900051-07-0. [p1]

Oliver Keyes
Wikimedia Foundation

ironholds@gmail.com

Jennifer Bryan
University of British Columbia

jenny@stat.ubc.ca

David Robinson
Princeton University

admiral.david@gmail.com