R and related tools for research and teaching

Jouni Helske & Satu Helske

4 December 2015

Working more efficiently with R

Rcpp

git

github
SOCIAL CODING

Bitbucket

R Markdown

Shiny
by RStudio

R

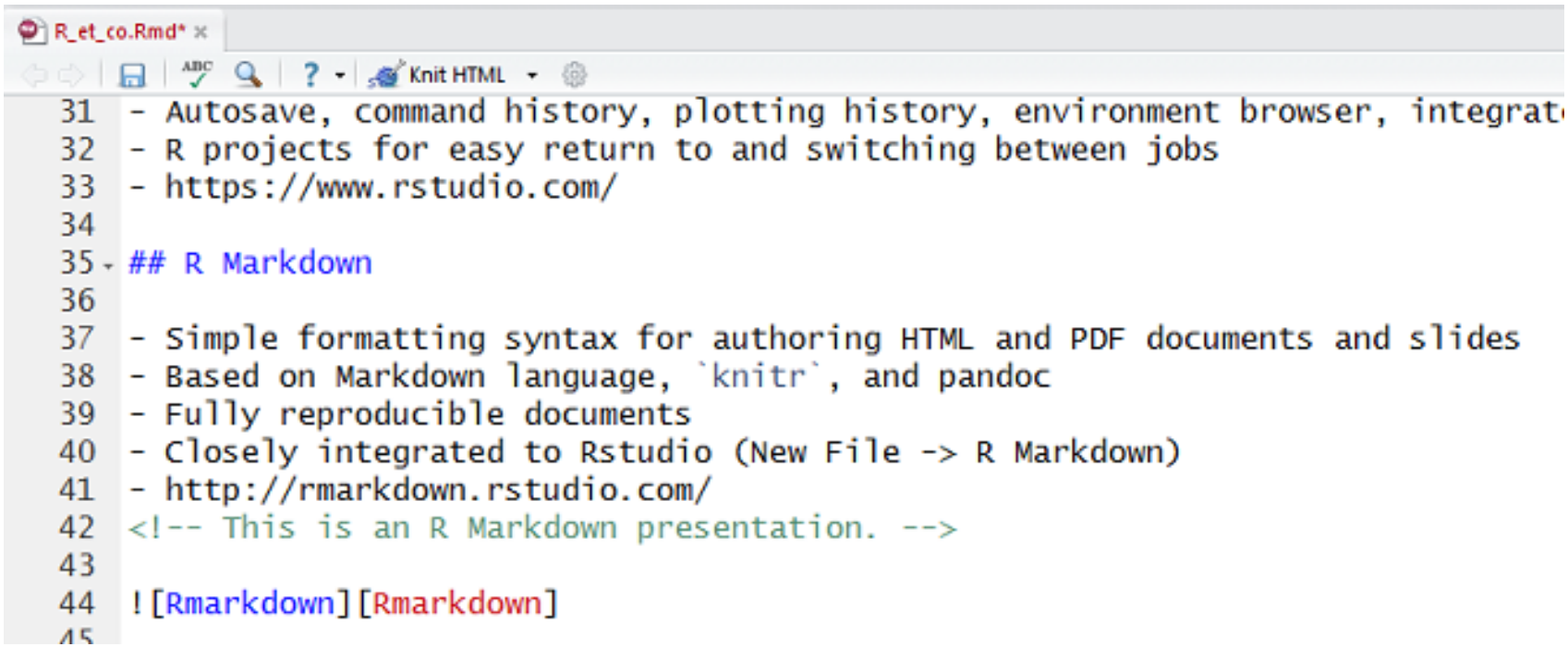
- More than 2 million users (Oracle estimate in 2012)
- 6th most popular programming language by IEEE Spectrum (in 2015)
- 3rd in number of scholarly articles in Google Scholar (in 2014; blog post by R.A.Muenchen)
- UseR! 2015 conference: 660 participants, 284 from industry
- R activity around the world: `http://rapporter.net/custom/R-activity`
- `https://www.r-project.org/`

Rstudio

- Open source graphical interface for R and related tools
- Code completion, syntax highlighting, code diagnostics
- Autosave, command history, plotting history, environment browser, integrated searchable help
- R projects for easy return to and switching between jobs
- `https://www.rstudio.com/`

R Markdown

- Simple formatting syntax for authoring HTML and PDF documents and slides
- Based on Markdown language, `knitr`, and pandoc
- Fully reproducible documents
- Closely integrated to Rstudio (New File -> R Markdown)
- `http://rmarkdown.rstudio.com/`



Rmd features

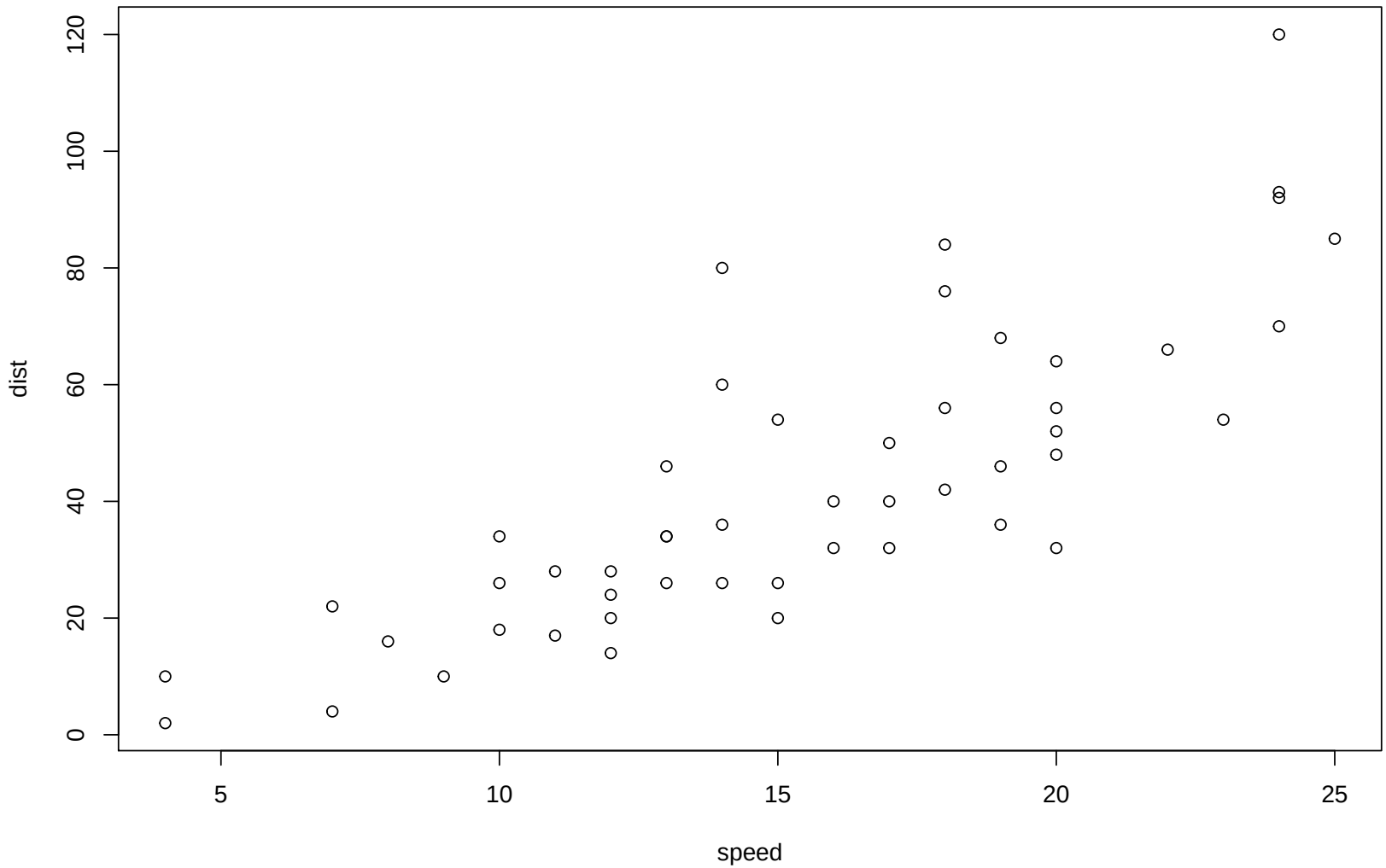
- Equations like in LaTex: \$y_i = \beta x_i\$ produces $y_i = \beta x_i$
- Embedding R code in the document:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Rmd figures

```
plot(cars)
```



Rmd tables

```
Type      | Freq   | %
----------| ------ |-----
Apples    | 7      | 44
Oranges   | 9      | 56
```

Type	Freq	%
Apples	7	44
Oranges	9	56

R package creation

- Sharing work with others
  - Co-workers, CRAN etc.
- Personal projects
  - Loading functions, data and other packages at once
- Easier integration with C/C++/Fortran codes with R
- Getting started with `package.skeleton`
- In Rstudio: New Project -> New Directory -> R package

Git

- Version control system originally for software development
- Useful for writing research articles, theses, R packages, ...
- Basics are easy and sufficient for small projects
- Embedded in Rstudio (Tools -> Version Control)
- `http://rogerdudler.github.io/git-guide/`

Github

- Web-based Git repository hosting service
- Interacting with other developers and users
- Bug tracking, feature requests, task management, ...
- Free for public repositories
  - Free private repositories e.g. in Bitbucket
- `https://github.com/helske`

Useful packages for graphics

- `ggplot2`
  - Plotting system based on the grammar of graphics
  - Easy to produce complex multi-layered graphics
- `ggvis`
  - Similar to `ggplot2`, faster but more restricted
  - Interactive graphics in RStudio or a browser
- `grid`, `gridBase`
  - Control and flexibility in appearance and arrangement
  - Basis for developing high-level functions

Useful packages for data manipulation

- `magrittr`
  - Piping via %>% operator
  - Improves readability and maintainability of code
- `dplyr` and `data.table`
  - Enhanced versions of `data.frame`
  - Fast and memory-efficient
  - More flexible data manipulation
  - Working with remote databases, automatic translation to SQL (`dplyr`)

Useful packages for reporting

- `knitr`
  - Dynamic report generation (PDF, html, Word, ...)
  - Easy to re-compile and update outputs
- `shiny`
  - Building interactive web applications from R
  - Web development skills not required

Useful packages for packaging

- `testthat`
  - Testing framework for R
  - Catching errors, warnings, messages, ...
- `devtools`
  - Simplifying tasks in package development
- `roxygen2`
  - Writing documentation for functions etc.
- `Rcpp`
  - Simple C++ integration in R
  - Writing C++ separately, Rcpp handles the ugly stuff

Learning

- Read
  - RStudio Online learning
  - Quick-R
  - OpenIntro
- Hands-on learning online
  - DataCamp
    - R tutorials and data science courses in browser
  - Codeschool
    - Learn R, Git, SQL, ...
- Hands-on in R
  - `swirl` package

Teaching I

- OpenIntro
  - Free material for teachers to use and modify
  - Textbooks, slides, excercises, ...
  - Statistics, R, SAS, ...
- DataCamp
  - Resources for building own courses
- `testwhat` package
  - Wrapper around `testthat` for checking exercises

Teaching II

- Shiny apps by Cal Poly State University
  - Correlation and regression game, sampling distribution demonstration, longest run of heads or tails, ...
- Using Github (experiences by Colin Rundel)
  - Assignments turned in via Github
  - Students learn version control
  - Simpler course administration
  - Keeping track of contributions

Getting help

- Stack overflow
- Google. . .