Living the reproducible life

October 2018

Hadley Wickham
@hadleywickham
Chief Scientist, RStudio
Why?
Two primary computer interfaces: point & click GUI
Or: program with a CLI
What are the strengths of code?

1. Code is transportable
2. Code is repeatable
3. Code is read-able
4. Code is shareable
Code is transportable

Which leads to two powerful programming techniques
Top Questions

1. **Where to scroll to in a React app when route changes for screen reader**
   - Votes: 0
   - Answers: 0
   - Views: 2
   - Tags: react, react-router, accessibility, react-router-dom
   - Asked: 55 secs ago by dagata
   - Views: 9,694

2. **What if I schedule tasks for Celery to perform every minute and it is not able to complete it in time?**
   - Votes: 0
   - Answers: 0
   - Views: 2
   - Tags: Celery, scheduled-tasks, celery-task, celerybeat
   - Asked: 1 min ago by ravi
   - Views: 1

3. **Gerrit: Is there a way to push directly into master?**
   - Votes: 1
   - Answers: 0
   - Views: 5
   - Tags: gerrit
   - Modified: 2 mins ago by koyuwah
   - Views: 1,909

4. **Kubectl run-ing a tarballed image**
   - Votes: 0
   - Answers: 0
   - Views: 3
   - Tags: docker, kubernetes, kubectl
   - Asked: 2 mins ago by adiliberto
   - Views: 4,470

5. **AspNet Identity RequireUniqueEmail = false throws exception on CreateAsync**
   - Votes: 0
   - Answers: 0
   - Views: 7
   - Tags: asp-net-identity, unique, own
   - Modified: 2 mins ago by Dmitry Duka
   - Views: 1

6. **Square: Call to undefined function charge**
   - Votes: 0
   - Answers: 0
   - Views: 3
   - Tags: square-connect
   - Asked: 3 mins ago by John
   - Views: 4,104

7. **Python - Social Studio API - How to unpack JSON into pandas data frame**
   - Votes: 0
   - Answers: 0
   - Views: 4
   - Tags: python, json, pandas
   - Asked: 3 mins ago by Ulisses Sotomayor
   - Views: 36

8. **Issues running airflow scheduler as a daemon process**
   - Votes: 2
   - Answers: 1
   - Views: 15
   - Tags: python, amazon-ec2, ubuntu-18.04, airflow, apache-airflow
   - Answered: 3 mins ago by Tagar
   - Views: 2,993

9. **Flutter animation how to fade in/out gradually**
   - Votes: 1
   - Answers: 1
   - Views: 20
   - Tags: Flutter
   - Asked: 3 mins ago by Coffee
   - Views: 987
You can easily share code on GitHub
Code is repeatable
Analysis = Data + Decisions

Changes frequently!
Code is readable
Code lets you track every action you’ve performed
Good code also communicates intent

```r
library(tidyverse)
library(magick)

dir(pattern = ".png") %>%
  map(image_read) %>%
  image_join() %>%
  image_animate(fps = 1, loop = 25) %>%
  image_write("my_animation.gif")
```

https://twitter.com/ricardokriebel/status/849626401611411458
Much of my work is developing domain specific languages.
Code is shareable
All modern programming languages are open source

Free
Students can use same tools as practitioners.
Anyone can use best tools regardless of wealth.
Anyone can re-run your analysis

Bidirectional
You can fix problems
You can build your own tools
Why is sharing so important?

Learn from others

Open science

Easily critique
How?
Combine prose, code, & results
By JJ Allaire & Yihui Xie
https://rmarkdown.rstudio.com
Prose and code

Prose and results

.Rmd

.md

.html

.doc

.tex

.ppt

.pdf

...
Demo
# RStudio Formats

R Markdown formats from RStudio (see below for additional formats created by the R community)

## Documents
- **Notebook**
  - Interactive R Notebooks
- **HTML**
  - HTML document w/ Bootstrap CSS
- **PDF**
  - PDF document (via LaTeX template)
- **Word**
  - Microsoft Word document (docx)
- **ODT**
  - OpenDocument Text document
- **RTF**
  - Rich Text Format document
- **Markdown**
  - Markdown document (various flavors)

## Journals
- **jss_article**
- **acm_article**
- **acs_article**
- **ctex**
- **elsevier_article**

## Presentations
- **ioslides**
  - HTML presentation with ioslides
- **reveal.js**
  - HTML presentation with reveal.js
- **Slidy**
  - HTML presentation with W3C Slidy
- **Beamer**
  - PDF presentation with LaTeX Beamer
- **PowerPoint**
  - PowerPoint presentation

## More
- **flexdashboard**
- **bookdown**
- **Websites**
- **Tufte Handout**
- **Package Vignette**
- **Github Document**

- Interactive dashboards
- HTML, PDF, ePub, and Kindle books
- Multi-page websites
- Handouts in the style of Edward Tufte
- R package vignette (HTML)
- GitHub Flavored Markdown document.
Reproducible examples
By Jenny Bryan
https://reprex.tidyverse.org
Demo
Reproducible books
By Yihui Xie
https://bookdown.org
Advanced R
Hadley Wickham

Welcome

This is the website for work-in-progress 2nd edition of "Advanced R", a book in Chapman & Hall/CRC Series. The book is designed primarily for R users who want to improve their programming skills and understanding of the language. It should also be useful for programmers coming to R from other languages, as it explains some of R’s quirks and shows how some parts that seem horrible on a positive side.

This edition is a work in progress. If you’re interested in the electronic version of the 1st edition, you can find it online at http://adv-r.had.co.nz/.

Other books

You may also be interested in:

- "R for Data Science" which introduces...
Package websites
By Hadley Wickham
http://pkgdown.r-lib.org
Overview

ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

Installation

# The easiest way to get ggplot2 is to install the whole tidyverse:
install.packages("tidyverse")

# Alternatively, install just ggplot2:
install.packages("ggplot2")

# Or the the development version from GitHub:
# install.packages("devtools")
devtools::install_github("tidyverse/ggplot2")

Cheatsheet

Data Visualization with ggplot2 :: CHEAT SHEET

Links

Download from CRAN at https://cloud.r-project.org/package=ggplot2
Browse source code at https://github.com/tidyverse/ggplot2
Report a bug at https://github.com/tidyverse/ggplot2/ issues
Learn more at http://r4ds.had.co.nz/data-visualisation.html
Extensions at http://www.ggplot2-exts.org/gallery/

License

GPL-2 | file LICENSE

Citation

Citing ggplot2

Developers

Hadley Wickham
Author, maintainer
Winston Chang
Author
Lionel Henry
Author
Thomas Lin Pedersen
Fin
Why code?

1. Code is transportable
2. Code is repeatable
3. Code is read-able
4. Code is shareable