Why R and RStudio Desktop?

Dr Andrew J. Stewart

Fellow of the Software Sustainability Institute

E: drandrewjstewart@gmail.com

T: @ajstewart_lang

G: ajstewartlang



Why R?

- R allows you to engage in reproducible research and in building reproducible workflows.
- Statistical packages in R reflect the latest advances in the fields of Statistics and Data Science.
- There are lots of amazing data visualisation packages available.
- In many institutions, the next generation of academics (M-level and PhD students, post-docs etc.) are learning R skills as part of their training.
- R coding skills are highly sought after in Industry/ONS/NHS/Civil Service...

Open Source Software

- Open Source Software is free and built/maintained/modified/improved by the community (which you are part of!) - check what license it is under for full details.
- You can use Open Source Software after you graduate so no more being tied to proprietary software (e.g., SPSS) that needs to be paid for.
- Open Source Software is supported by large corporations for example, Linux is supported by the Linux Foundation which includes Google, Microsoft, Intel, Samsung etc.
- Open Source is widely used across the public and private sectors and has had a huge impact on computing over the last couple of decades.

Open Source Software - UK Government

Guidance

Be open and use open source

Publish your code and use open source to improve transparency, flexibility and accountability.

Published 6 November 2017

From: Government Digital Service

Contents

- How open source differs to open standards
- How using open source will help your programme
- How being open will help your programme
- Publishing code when building your technology
- Buying technology that uses open source
- Related guides and sources

Related content

Open Source guidance

Make use of open standards

Hackney Council's approach to moving to a cloud-first model from the PSN

Define user needs

https://www.gov.uk/guidance/be-open-and-use-open-source



Blog

MOJ Digital & Technology

Organisations: Ministry of Justice

Why we code in the open

Dave Rogers and Steve Marshall, 21 February 2017 - Our services

At the Ministry of Justice, we code in the open, by default. This means whenever we write software, we make <u>our source code available to anyone and everyone</u>.

We are strong advocates of point 8 of the Digital by Default Service Standard:

Make all new source code open and reusable, and publish it under appropriate licences (or provide a convincing explanation as to why this can't be done for specific subsets of the source code).

So our teams have to make a case for not making their code freely available. This approach has resulted in us writing <u>tens of millions of lines of code</u>, in <u>hundreds of code repositories</u>, in the open since 2012.

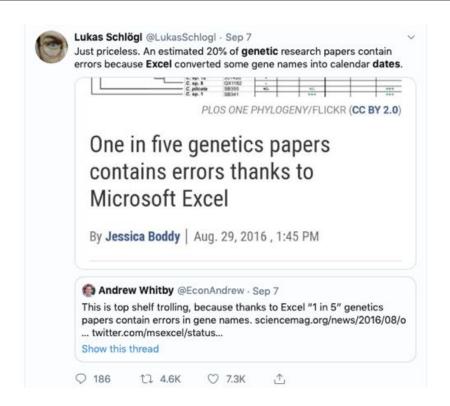
The Advantages of Open Source Software

Open software allows for transparency and reproducibility in your data processing, wrangling, visualisation modelling etc.

It allows others to spot and correct errors...



The Risks of Closed Source Software



Reinhart, Rogoff... and Herndon: The student who caught out the profs

By Ruth Alexander BBC News

① 20 April 2013









This week, economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors. Another surprise is that the mistakes, by two eminent Harvard professors, were spotted by a student doing his homework.

It's 4 January 2010, the Marriott Hotel in Atlanta. At the annual meeting of the American Economic Association, Professor Carmen Reinhart and the former chief economist of the International Monetary Fund, Ken Rogoff, are presenting a research paper called Growth in a Time of Debt.



R vs. SPSS

"SPSS is like a bus - easy to use for the standard things, but very frustrating if you want to do something that is not already pre-programmed.

R is a 4-wheel drive off-roader, with a bike on the back, a kayak on top, good walking and running shoes in the passenger seat, and mountain climbing and spelunking gear in the back.

R can take you anywhere you want to go if you take time to learn how to use the equipment, but that is going to take longer than learning where the bus stops are in SPSS."

Greg Snow, 2010, stackoverflow.com

In meme form...

If statistics programs/languages were cars...

Image credit Darren Dahly, @statsepi



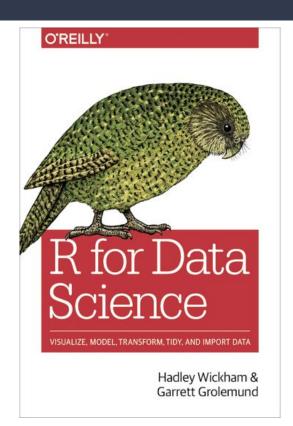








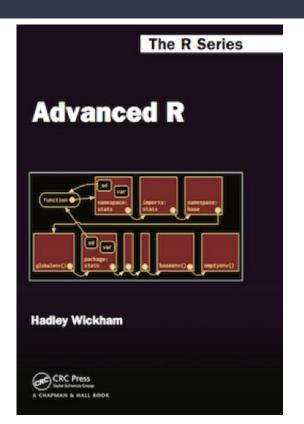
There are great R books to read...



Available electronically at:

http://r4ds.had.co.nz

There are great R books to read...



Available electronically at:

https://adv-r.hadley.nz

Hadley Wickham

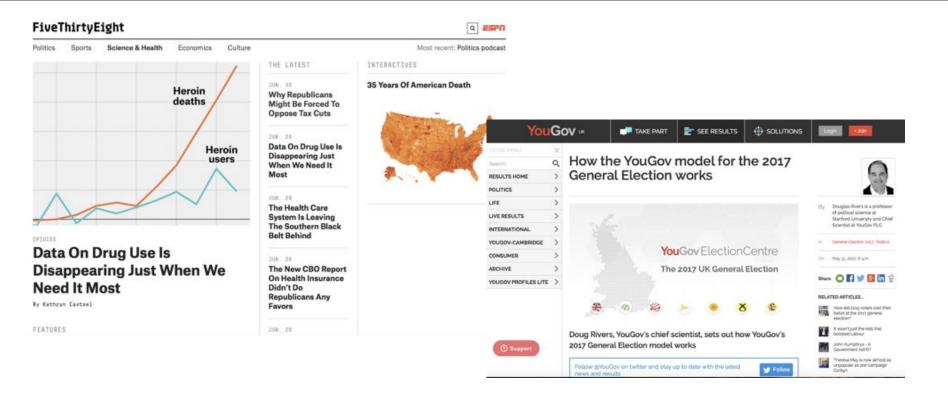


"The Man Who Revolutionized R"

Chief Scientist at RStudio, (lead) author of key R packages incl. ggplot2, tidyr, dplyr - all components of the tidyverse.

https://priceonomics.com/hadley-wickham-the-man-who-revolutionized-r/

R is widely used by a number of organisations



How to create BBC style graphics

Load all the libraries you need

Install the bbplot package

How does the bbplot package work?

Save out your finished chart

Make a line chart

Make a multiple line chart

Make a bar chart

Make a stacked bar chart

Make a grouped bar chart

Make a dumbbell chart

Make a histogram

Make changes to the legend

Make changes to the axes

Add annotations

Work with small multiples

Do something else entirely

https://medium.com/bbc-visual-and-da ta-journalism/how-the-bbc-visual-and-data-journalism-team-works-with-grap hics-in-r-ed0b35693535

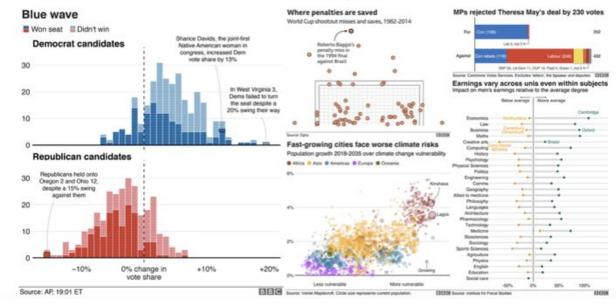
BBC Visual and Data Journalism cookbook for R graphics

Last updated: 2019-01-24

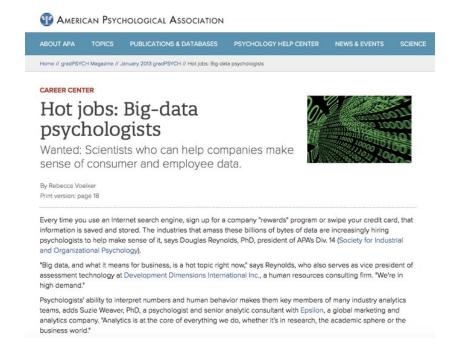
How to create BBC style graphics

At the BBC data team, we have developed an R package and an R cookbook to make the process of creating publication-ready graphics in our in-house style using R's ggplot2 library a more reproducible process, as well as making it easier for people new to R to create graphics.

The cookbook below should hopefully help anyone who wants to make graphics like these:



And Data Science is a growing employment destination for Psychologists



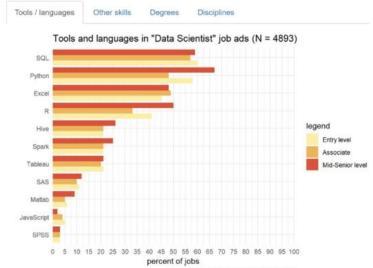


If your department teaches quantitative skills in SPSS, SAS, or Matlab, is it really preparing students for quantitative jobs outside of academia? #datascientist

#academia #rstats #python

Data scientist jobs

4893 data scientist job ads were included in this analysis.



Jobs were scraped from LinkedIn in Sept 2019 Locations included NYC, SF, Seattle, Boston, and Toronto Positions were Entry, Associate, and Senior levels #brainhackschool instructors be like: familiar with docker?

Me: 👑

Familiar with jupyter?

Me: 👑

Familiar with github?

Me: 😬

Familiar with binder?

Me: 😬

Familiar with python?

Me: 😬

It feels like I've spent these past 4 years of PhD on mars



7:05 PM · Aug 7, 2019 · Twitter for iPhone

There will be coding...



How I code:

- 1. Make it work
- Make it pretty

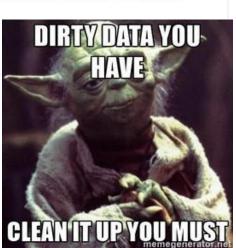
Make it work again because I probably broke it while trying to make it pretty

2:27 PM · Sep 13, 2019 · Twitter Web App

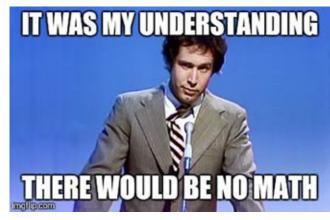
5.3K Retweets 32.4K Likes



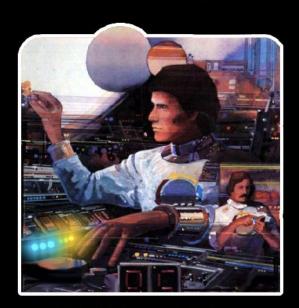








THE TWO STATES OF EVERY PROGRAMMER



I AM A GOD.



I HAVE NO IDEA WHAT I'M DOING.