Why R?

Unit Coordinator: Dr Danna R. Gifford
(danna.gifford@manchester.ac.uk)

Original author: Prof Andrew J. Stewart
(drandrewjstewart@gmail.com)

Why R?

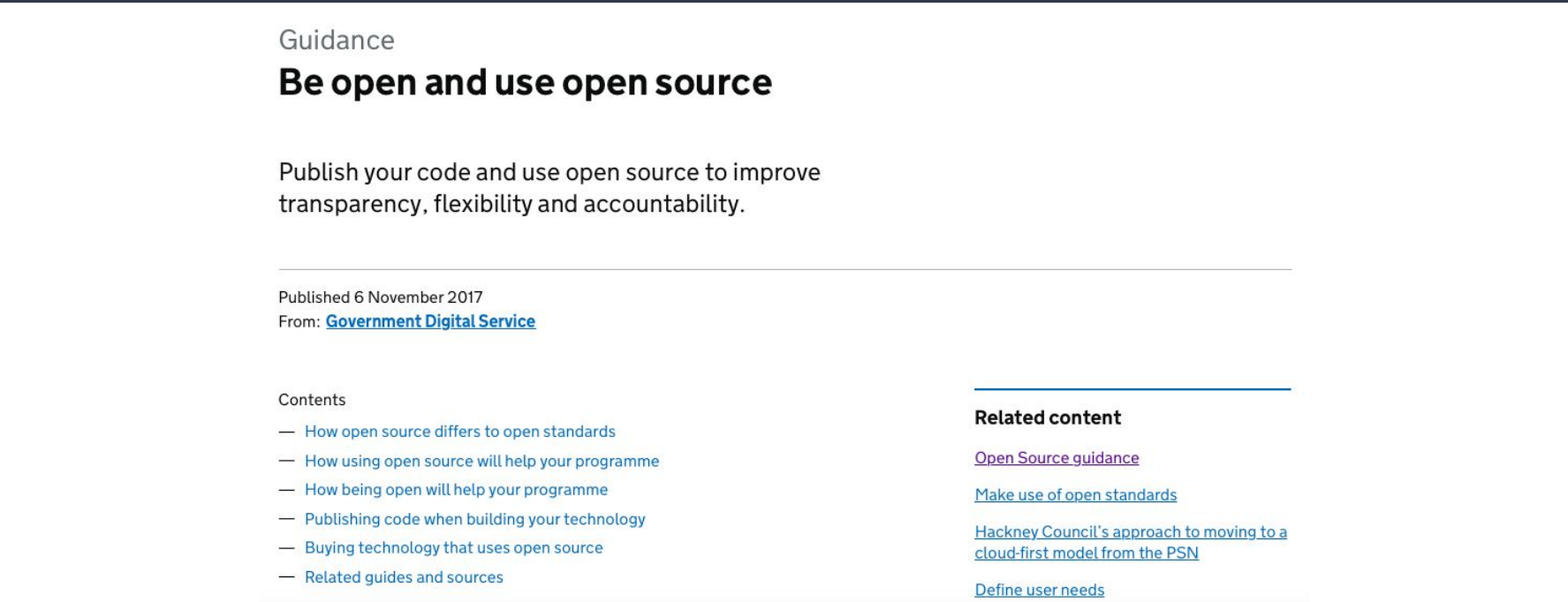
- R allows you to engage in reproducible research and in building reproducible workflows.
- Statistical packages in R reflect the latest advances in the fields of Statistics and Data Science.
- There are lots of amazing data visualisation packages available.
- In many institutions, the next generation 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.

https://opensource.org/history

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

GOV.UK

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 our source code available to anyone and everyone.

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 tens of millions of lines of code, in hundreds of code repositories, in the open since 2012.

https://mojdigital.blog.gov.uk/2017/02/21/why-we-code-in-the-open/

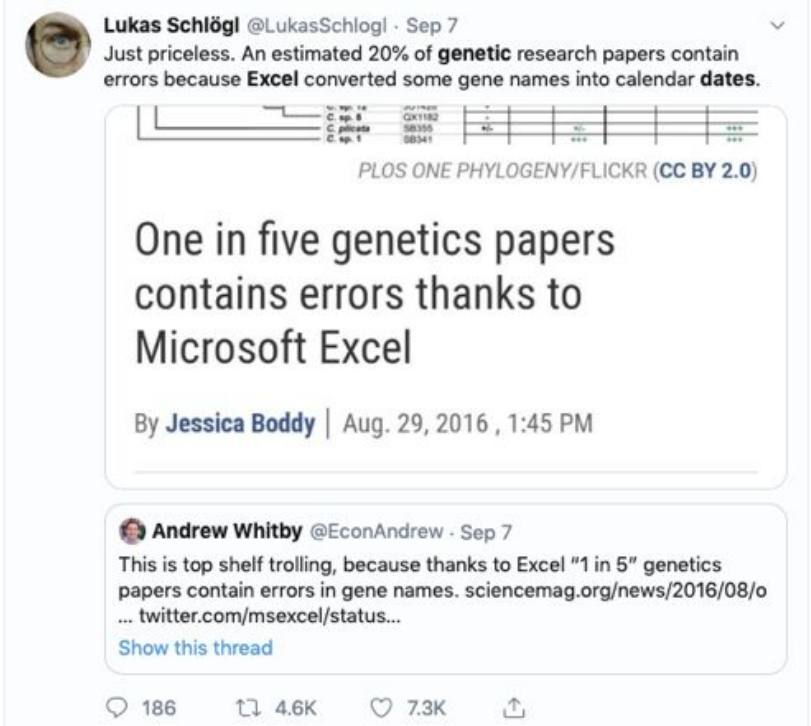
The Advantages of Open Source Software

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

It allows others to spot and correct errors...

https://arstechnica.com/information-technology/2019/10/chemists-discover-cross-platform-python-scripts-not-so-cross-platform/

The Risks of Closed Source Software



Lukas Schlögl @LukasSchlogl · Sep 7

Just priceless. An estimated 20% of **genetic** research papers contain errors because **Excel** converted some gene names into calendar **dates**.

PLOS ONE PHYLOGENY/FLICKR (CC BY 2.0)

One in five genetics papers contains errors thanks to Microsoft Excel

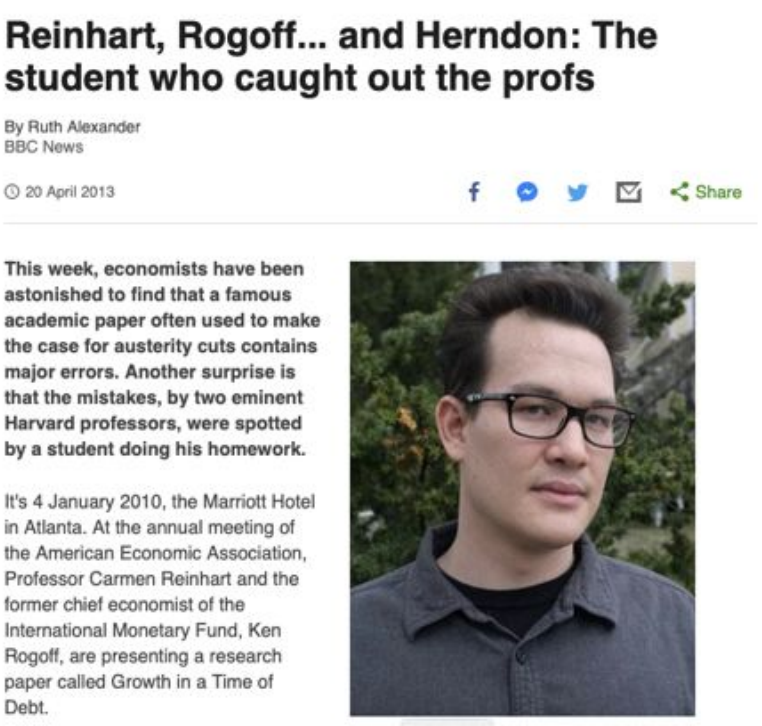
By **Jessica Boddy** | Aug. 29, 2016 , 1:45 PM

Andrew Whitby @EconAndrew · Sep 7

This is top shelf trolling, because thanks to Excel "1 in 5" genetics papers contain errors in gene names. sciencemag.org/news/2016/08/o ... twitter.com/msexcel/status...

Show this thread

186 4.6K 7.3K



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

By Ruth Alexander
BBC News

20 April 2013 Share

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.

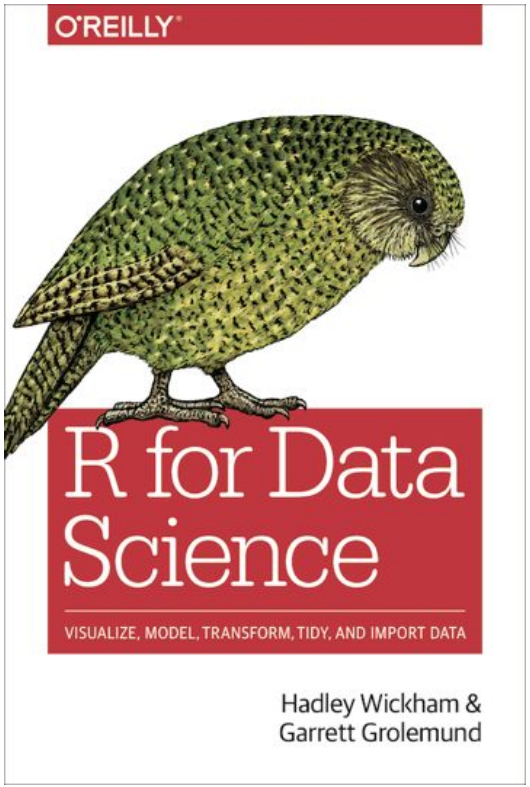
R can take you anywhere you want to go if you take time to learn how to use the equipment, but that is going to take longer than learning where the bus stops are in SPSS.”

Greg Snow, 2010, stackoverflow.com

In meme form...

If statistics programs/languages were cars...
Image credit Darren Dahly, @statsepi
Excel
SPSS
SAS
STATA
R

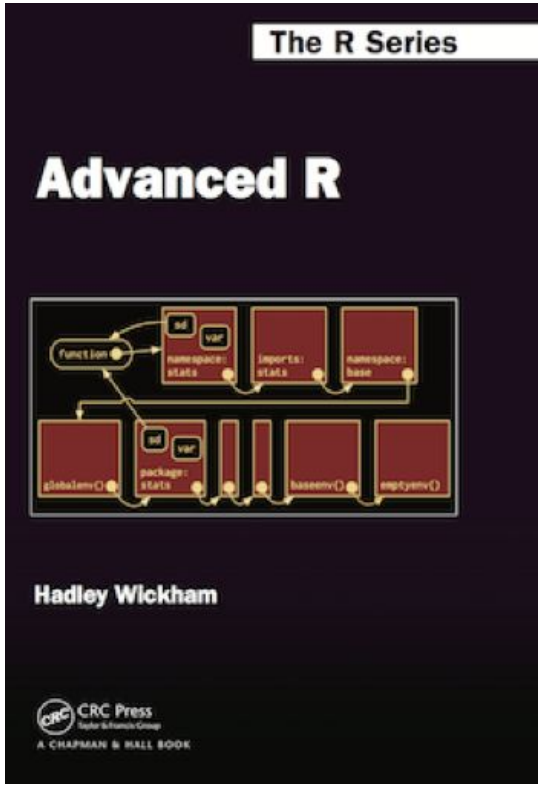
There are great R books to read...



Available electronically at:

http://r4ds.had.co.nz

There are great R books to read...



Available electronically at:

https://adv-r.hadley.nz

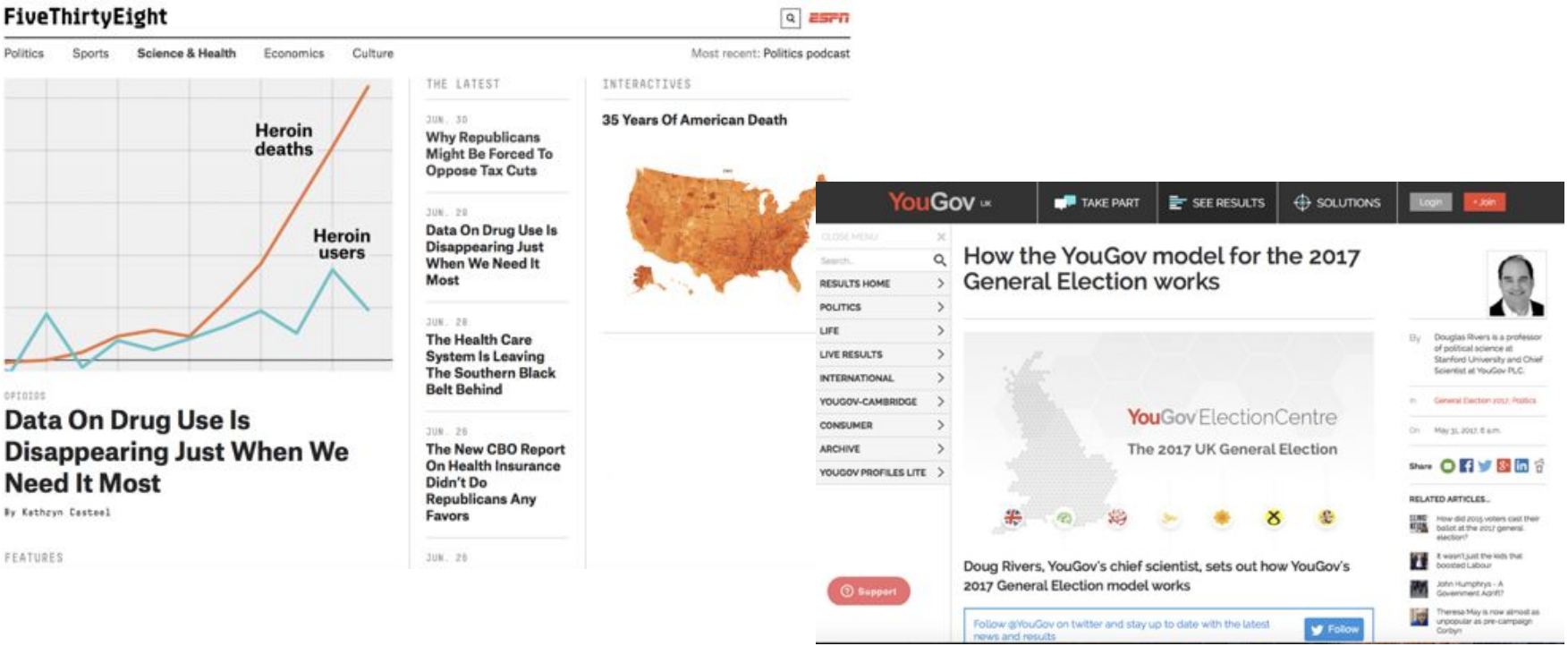
Hadley Wickham

“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

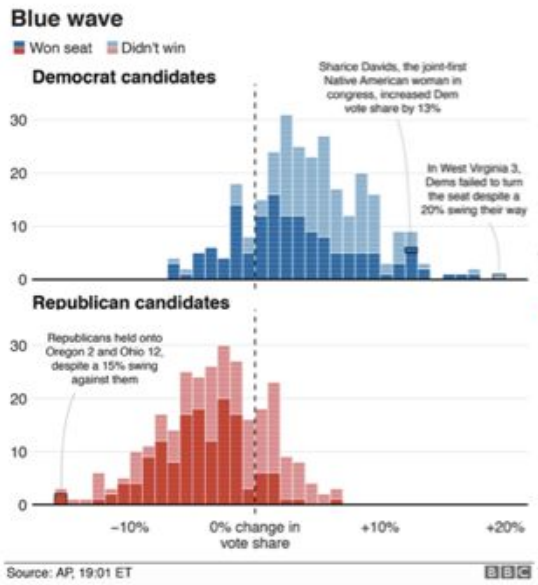
BBC Visual and Data Journalism cookbook for R graphics

Last updated: 2019-01-24

How to create BBC style graphics

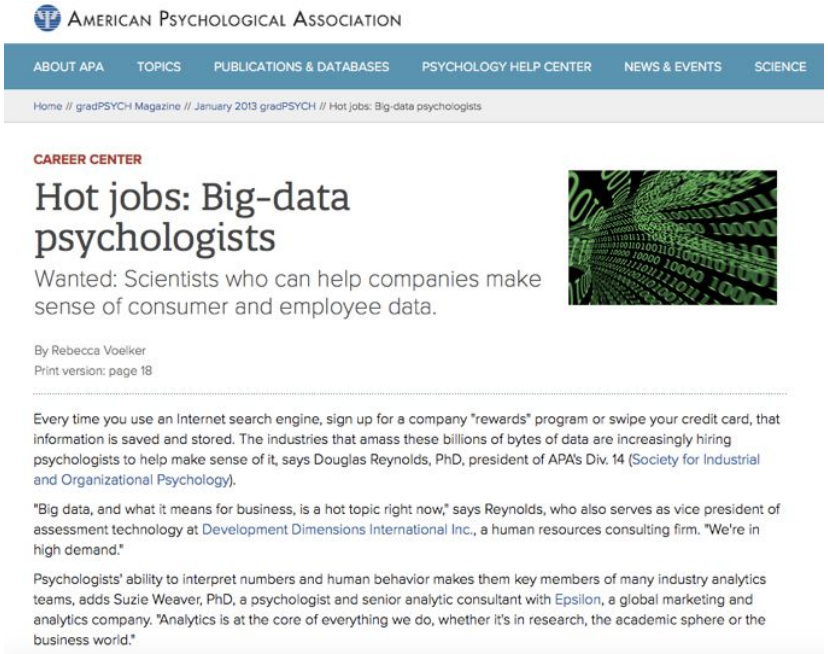
At the BBC data team, we have developed an R package and an R cookbook to make the process of creating publication-ready graphics in our in-house style using R's ggplot2 library a more reproducible process, as well as making it easier for people new to R to create graphics.

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



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

And Data Science is a growing employment destination for Psychologists



American Psychological Association

ABOUT APA | TOPICS | PUBLICATIONS & DATABASES | PSYCHOLOGY HELP CENTER | NEWS & EVENTS | SCIENCE

Home // gradPSYCH Magazine // January 2013 gradPSYCH // Hot jobs: Big-data psychologists

CAREER CENTER

Hot jobs: Big-data psychologists

Wanted: Scientists who can help companies make sense of consumer and employee data.

By Rebecca Voelker
Print version: page 18

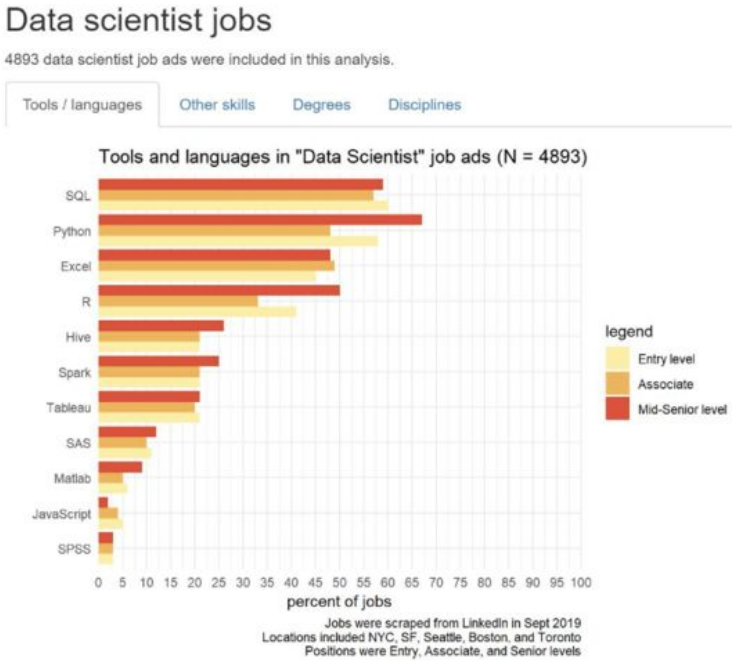
Every time you use an Internet search engine, sign up for a company "rewards" program or swipe your credit card, that information is saved and stored. The industries that amass these billions of bytes of data are increasingly hiring psychologists to help make sense of it, says Douglas Reynolds, PhD, president of APA's Div. 14 (Society for Industrial and Organizational Psychology).

"Big data, and what it means for business, is a hot topic right now," says Reynolds, who also serves as vice president of assessment technology at Development Dimensions International Inc., a human resources consulting firm. "We're in high demand."

Psychologists' ability to interpret numbers and human behavior makes them key members of many industry analytics teams, adds Suzie Weaver, PhD, a psychologist and senior analytic consultant with Epsilon, a global marketing and analytics company. "Analytics is at the core of everything we do, whether it's in research, the academic sphere or the business world."

Tyler Burleigh
@tylerburleigh

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



#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...

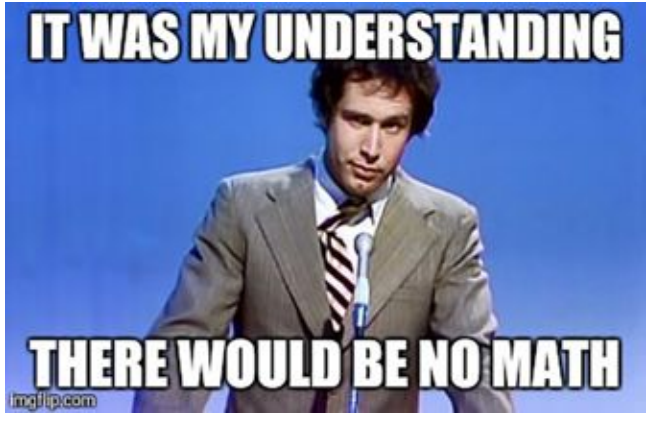
Kelly Vaughn
@kvlly

How I code:

1. Make it work
2. Make it pretty
3. 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.

Source unknown