## Welcome to The Hacker Within

Matthew Brett

March 19, 2018

#### Plan of the talk

- the hacker culture
- ▶ the hacker in academia
- ▶ the hacker within
- topics

# In the hacker spirit

This presentation built with Markdown and Pandoc, and available at:

https://github.com/matthew-brett/thw-intro

Please see README.md for various copyrights on image files.

#### Hacker Culture

https://en.wikipedia.org/wiki/Hacker\_culture

What they had in common was mainly love of excellence and programming. They wanted to make their programs that they used be as good as they could. They also wanted to make them do neat things. They wanted to be able to do something in a more exciting way than anyone believed possible and show "Look how wonderful this is. I bet you didn't believe this could be done."

Richard Stallman, interview as shown in Hackers — Wizards of the Electronic Age

#### Hacker Heroes



#### Richard Stallman

- Give me back my printer
- ► Lisp Machines Inc vs Symbolics.
- ► Gnu Public License and copyleft

#### Hacker Heroes



Linus Torvalds

I decided just, how hard can it be? - Interview with LT

#### Hacker ethic

- Sharing
- Openness
- Decentralization
- ▶ Free access to computers
- World Improvement
- https://en.wikipedia.org/wiki/Hacker\_ethic

#### Hacker ethic

- All information should be free
- Mistrust authority promote decentralization
- ► Hackers should be judged by their hacking, not criteria such as degrees, age, race, sex, or position
- https://en.wikipedia.org/wiki/Hacker\_ethic

# The rise of open source

- Linux and Android
- Darwin and BSD.
- ► Firefox, Chrome, Python, R . . . (to be continued).

## How is this relevant to us?



Nullius in verba

#### Science demands mistrust

Science alone of all the subjects contains within itself the lesson of the danger of belief in the infallibility of the greatest teachers in the preceding generation... Learn from science that you must doubt the experts ... Science is the belief in the ignorance of experts

Richard Feynman, What is Science? (1969)

# Error is everywhere

The scientific method's central motivation is the ubiquity of error - the awareness that mistakes and self-delusion can creep in absolutely anywhere and that the scientist's effort is primarily expended in recognizing and rooting out error."

Donoho, David L, et al. 2009. Reproducible research in computational harmonic analysis. Computing in Science & Engineering 11, 8–18.

## Ubiquity of error

In my own experience, error is ubiquitous in scientific computing, and one needs to work very diligently and energetically to eliminate it. One needs a very clear idea of what has been done in order to know where to look for likely sources of error. I often cannot really be sure what a student or colleague has done from his/her own presentation, and in fact often his/her description does not agree with my own understanding of what has been done, once I look carefully at the scripts. Actually, I find that researchers quite generally forget what they have done and misrepresent their computations. . . .

# Ubiquity of error

Computing results are now being presented in a very loose, "breezy" way—in journal articles, in conferences, and in books. All too often one simply takes computations at face value. This is spectacularly against the evidence of my own experience. I would much rather that at talks and in referee reports, the possibility of such error were seriously examined.

- David L. Donoho (2010). An invitation to reproducible computational research. Biostatistics Volume 11, Issue 3 Pp. 385-388

#### Error and software

Reinhart & Rogoff (2010) "Growth in a Time of Debt". American Economic Review. 100: 573–78.

Herndon, Ash & Pollin (2014) "Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff". Cambridge Journal of Economics. 38: 257–279.

Picketty (2013) Capital in the Twenty-First Century.

## The solution is transparency

It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty — a kind of leaning over backwards. ... the idea is to try to give all of the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another.

Richard Feynman, Cargo Cult Science (1974)

## Science vs advertising

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures"

The wavelab front page

#### The hacker culture comes to the rescue

#### The story of R

- ▶ S, S+ and John Chambers
- ▶ R at the University of Auckland; Ross Ihaka, Robert Gentleman



Growth of R on StackOverflow - see The Impressive Growth of R.

# And many other examples

- Python vs Matlab.
- Unix on Windows.
- BioPython, Bioconductor.
- Your favorite project here.

#### The Hacker Within



http://www.thehackerwithin.org

#### The Hacker Within Ethos

#### We:

- chose open over closed, simple over easy.
- learn
- ► teach
- build
- give our stuff away.

# Stallman again

One of the important reasons for giving software away free is to enable the users to change it. This allows them to make better use of it, and also encourages and enables them to contribute to the effort. Furthermore, they develop self-reliance, confidence, and a sense of responsibility – Richard Stallman at https://www.gnu.org/gnu/yes-give-it-away.html

Is this the end of the talk?

Yes, it is the end of the talk.

# And now for something closely related

Some extra slides that might be relevant to questions.

## Reproducibility crisis

Open access, freely available online

Essay

# **Why Most Published Research Findings Are False**

John P. A. Joannidis

- http://dx.doi.org/10.1371/journal.pmed.0020124
- http://matthew-brett.github.io/teaching/ioannidis\_2005.html

## Begley and Ellis 2012

Scientists at Amgen (a drug company) tried to reproduce findings from 53 "landmark" studies.

... when findings could not be reproduced, an attempt was made to contact the original authors, discuss the discrepant findings, exchange reagents and repeat experiments under the authors' direction, occasionally even in the laboratory of the original investigator.

Of 53 studies, only 6 replicated (11%).

Glenn Begley and Lee Ellis (2012) "Raise standards for preclinical cancer research" Nature 483 pp. 531–533

#### Cargo cult science

In the South Seas there is a cargo cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to imitate things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas — he's the controller — and they wait for the airplanes to land. They're doing everything right. The form is perfect [...]. But it doesn't work.

Richard Feynman (1974) Cargo Cult Science.

## Do not trust yourself

The first principle is that you must not fool yourself—and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other scientists. You just have to be honest in a conventional way after that.

Richard Feynman, Cargo Cult Science (1974)

#### The difference between true and false

In studies for which findings could be reproduced, authors had paid close attention to controls, reagents, investigator bias and describing the complete data set. For results that could not be reproduced, however, data were not routinely analysed by investigators blinded to the experimental versus control groups. Investigators frequently presented the results of one experiment, such as a single Western-blot analysis. They sometimes said they presented specific experiments that supported their underlying hypothesis, but that were not reflective of the entire data set.

Begley and Ellis (2012).