Scientific Computing and Reproducible Research

Lunchtime Science, April 18th, 2018

Dr. Joachim Krois

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

Content

- The Crisis of Science
- Reproducible Research
- Scientific Computing and Software Tools

Is Science broken?

The Crisis of Science

The science 'reproducibility crisis' – and what can be done about it

March 15, 2017 by Ottoline Leyser, Danny Kingsley And Jim Grange, The Conversation

The 7 biggest problems facing science, according to 270 scientists

By Julia Belluz, Brad Plumer, and Brian Resnick | Updated Sep 7, 2016, 10:13am EDT

Why 'Statistical Significance' Is Often Insignificant

Researchers who want professorships are sometimes driven to publish suspect findings.

By Noah Smith

2. November 2017, 13:00 MEZ

Why Are Voters Ignoring Experts?

Jul 1, 2016 | **JEAN PISANI-FERRY**

Academics and policymakers may be tempted to respond to events like Brexit by dismissing what looks like a celebration of ignorance and retreating into be more humb decisions.

Problems with scientific research

How science goes wrong

Scientific research has changed the world. Now it needs to change itself

Europe

9 out of 10 experts agree: Britain doesn't trust the experts on Brexit

By Griff Witte June 21, 2016 Email the author

Science in crisis: from the sugar scam to Brexit, our faith in experts is fading

September 27, 2016 7.43am BST

SCIENCE THE STATE OF THE UNIVERSE. AUG. 21 2017 6:00 AM

Is Science Broken?

Or is it self-correcting?

By Daniel Engber

Science in Crisis

Published on Nov 15, 2017 in *Issue 175 - November 2017, Learning and Education, Politics*

http://scienceincrisis.info/

The 7 biggest problems facing science

1. **Academia has a huge money problem.**
2. **Too many studies are poorly designed. Blame bad incentives.**
3. **Replicating results is crucial. But scientists rarely do it.**
4. **Peer review is broken.**
5. **Too much science is locked behind paywalls.**
6. **Science is poorly communicated to the public.**
7. **Life as a young academic is incredibly stressful.**

(Belluz et al. 2016)

The Crisis of Science

The Crisis of Science

Mohr H. (1977) The Crisis of Science. In: Lectures on Structure and Significance of Science. Springer, Berlin, Heidelberg

Futures 91 (2017) 5–11

What is science's crisis really about?

Andrea Saltelli[a,b,*], Silvio Funtowicz[a]

[a] Centre for the Study of the Sciences and the Humanities (SVT), University of Bergen, Norway
[b] Institute of Environmental Science and Technology (ICTA), Universitat Autònoma de Barcelona, Spain

THE AMERICAN STATISTICIAN
2016, VOL. 70, NO. 2, 129–133
http://dx.doi.org/10.1080/00031305.2016.1154108

Taylor & Francis
Taylor & Francis Group

EDITORIAL

The ASA's Statement on *p*-Values: Context, Process, and Purpose

COMPUTER SCIENCE

Science. 2018 Feb 16;359(6377):725-72

Artificial intelligence faces reproducibility crisis

Unpublished code and sensitivity to training conditions make many claims hard to verify

NATURE | NEWS

Nature doi:10.1038/nature.2014.14763

Publishers withdraw more than 120 gibberish papers

Conference proceedings removed from subscription databases after scientist reveals that they were computer-generated.

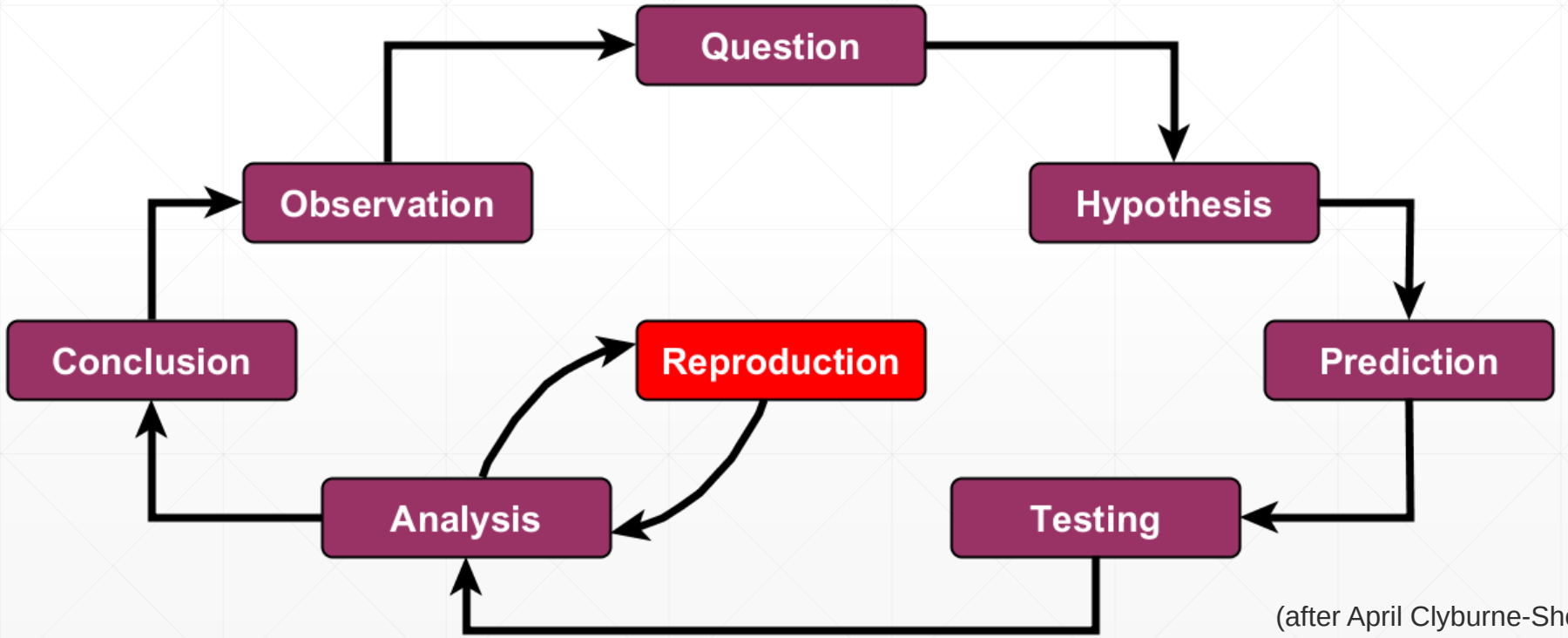
Richard Van Noorden

24 February 2014 | Updated: 25 February 2014

Psychology journal bans *P* values

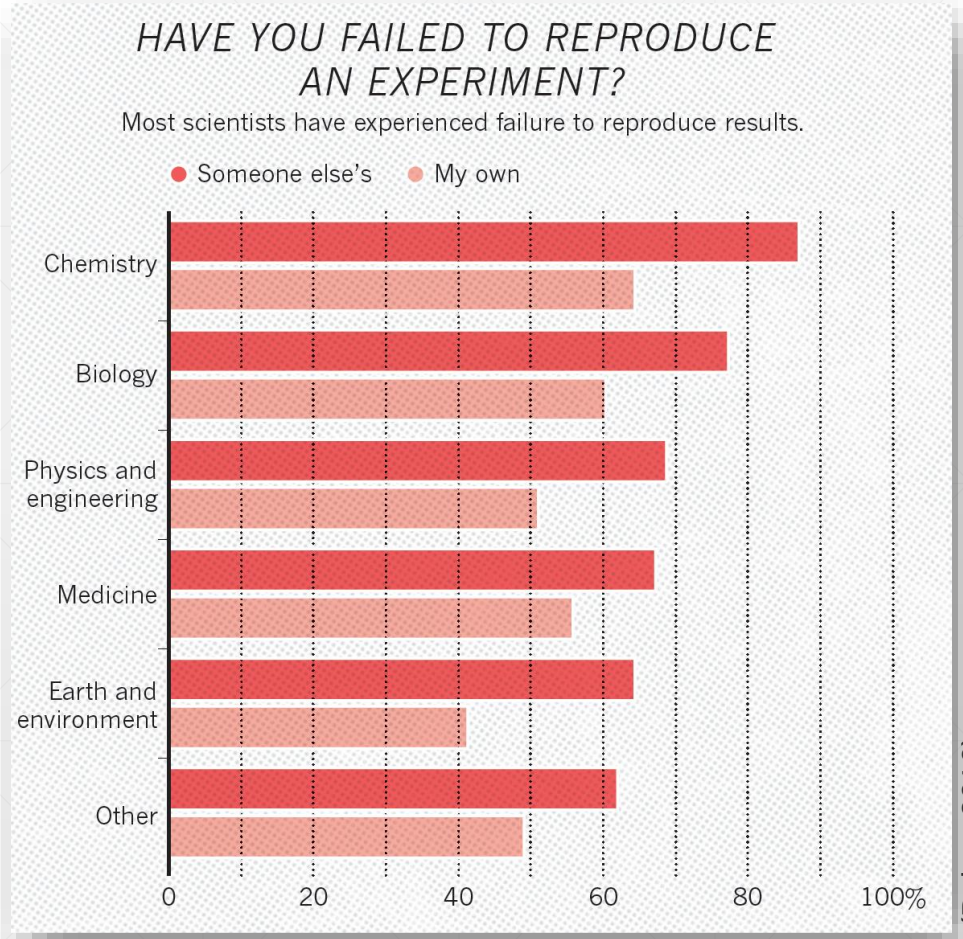
The Reproducibility Crisis

Replicating results is crucial. But scientists rarely do it.
(Belluz et al, 2016)



(after April Clyburne-Sherin, 2015)

The Reproducibility Crisis

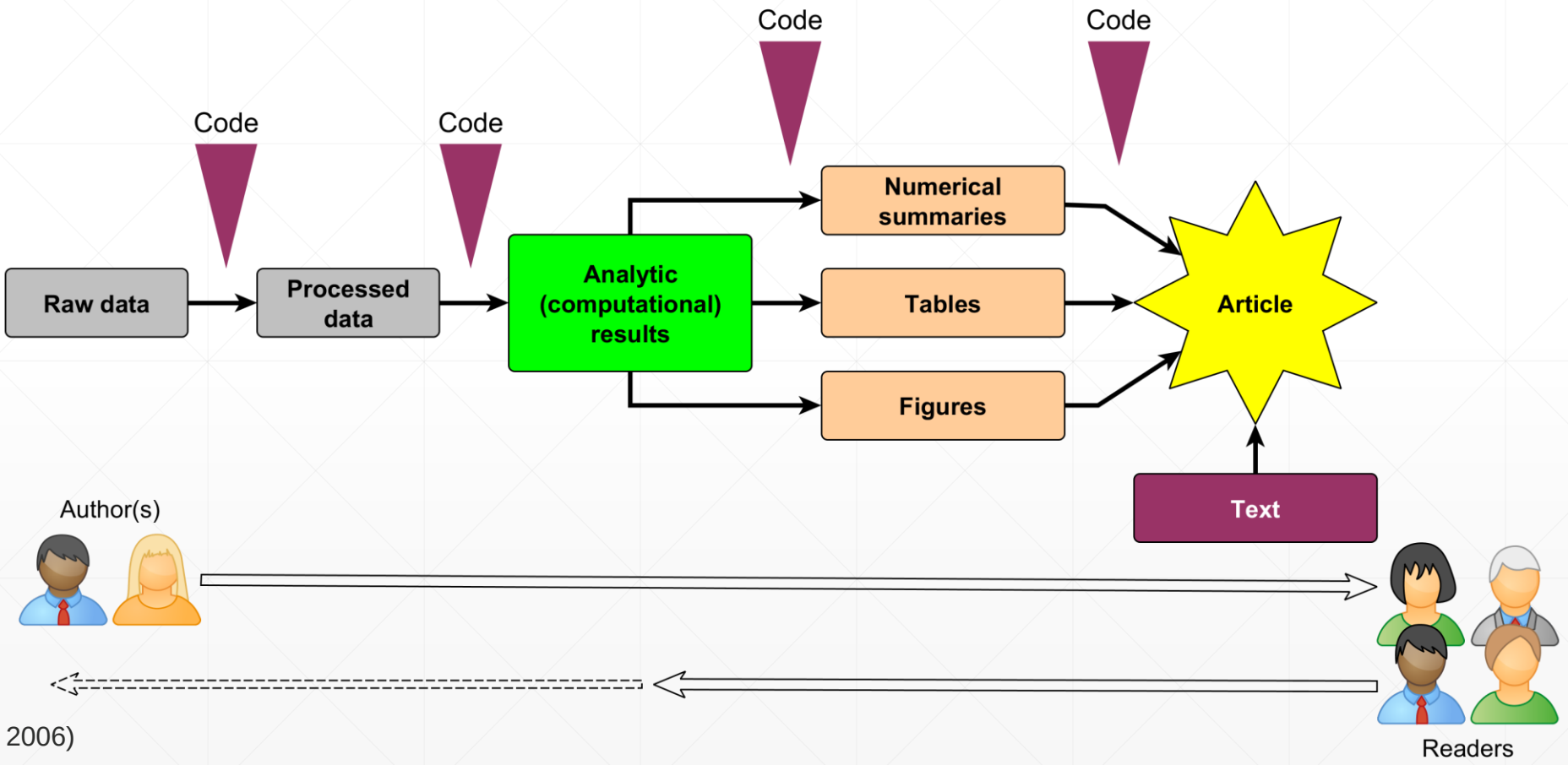


(Baker, 2016)

More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments.

How can we fix it?

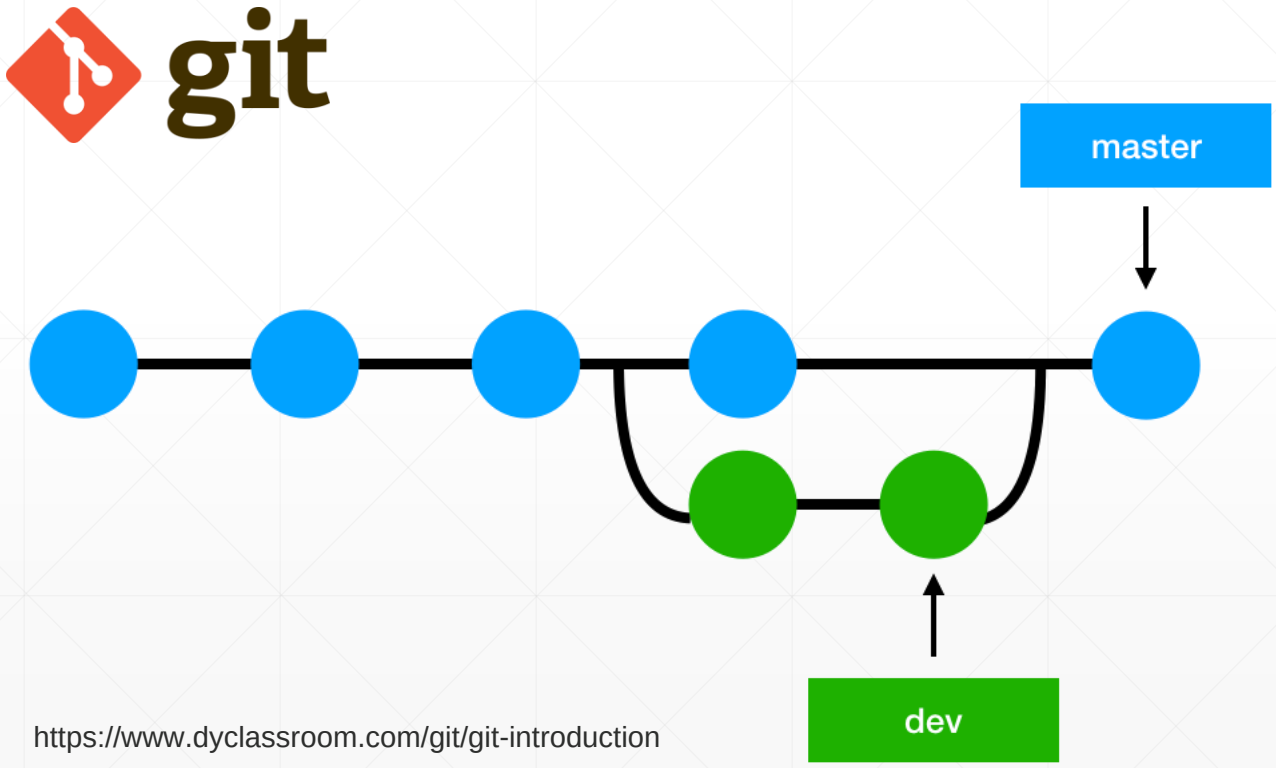
Reproducible Research



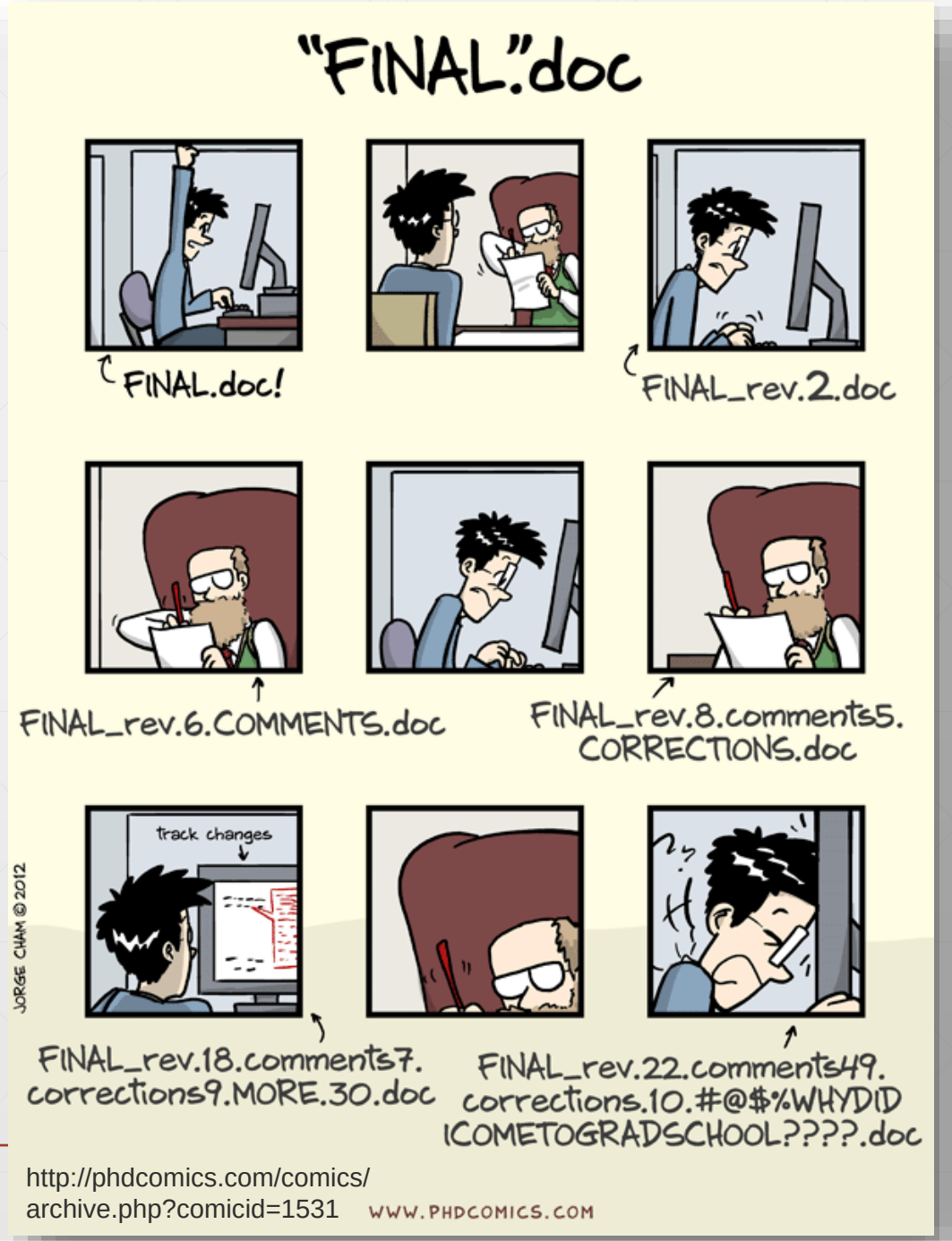
(after Peng et al. 2006)

The tools are out there, make use of them!

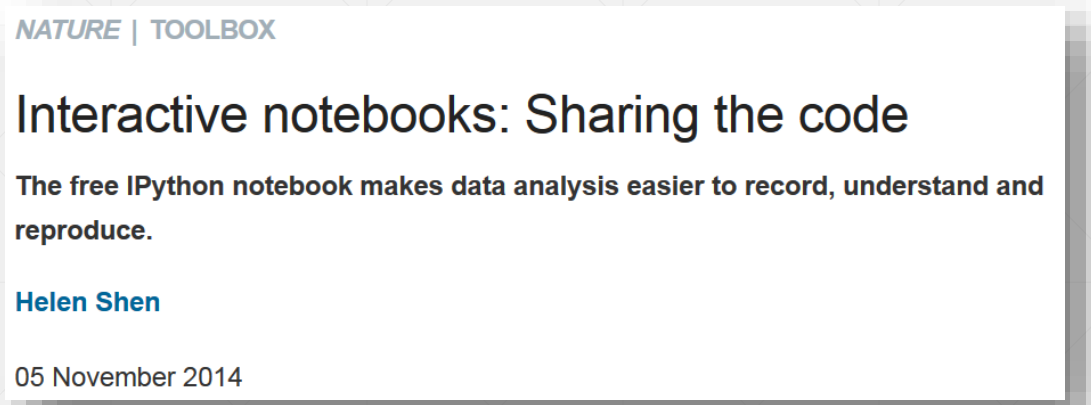
Version control

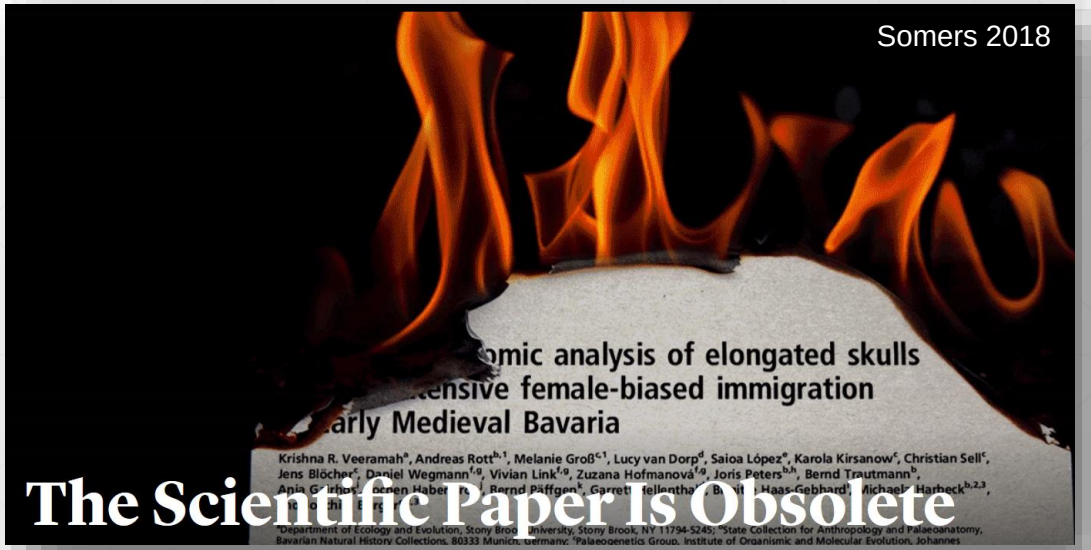


https://www.dyclassroom.com/git/git-introduction

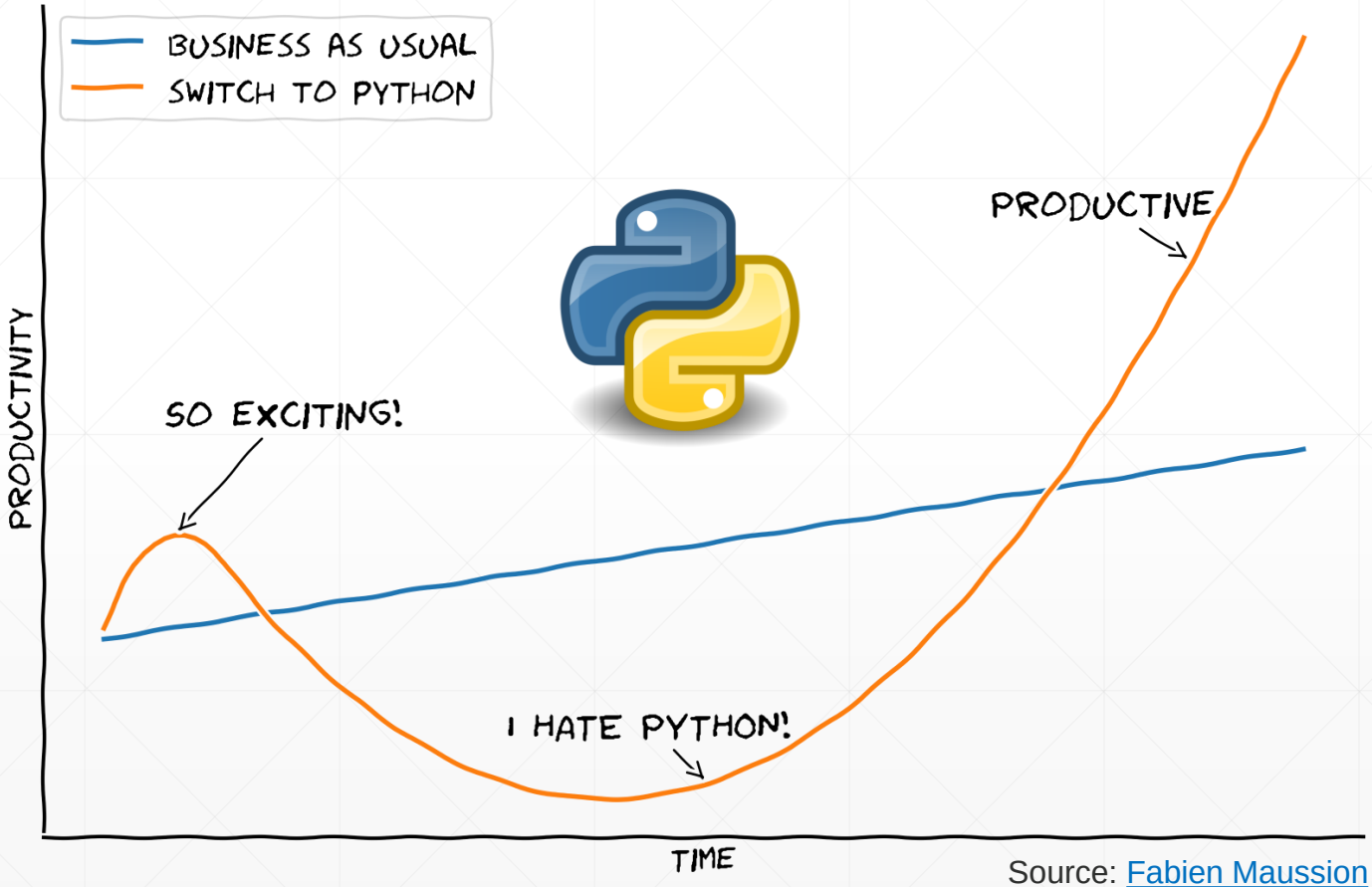


Literate Programming





Python Programming Language

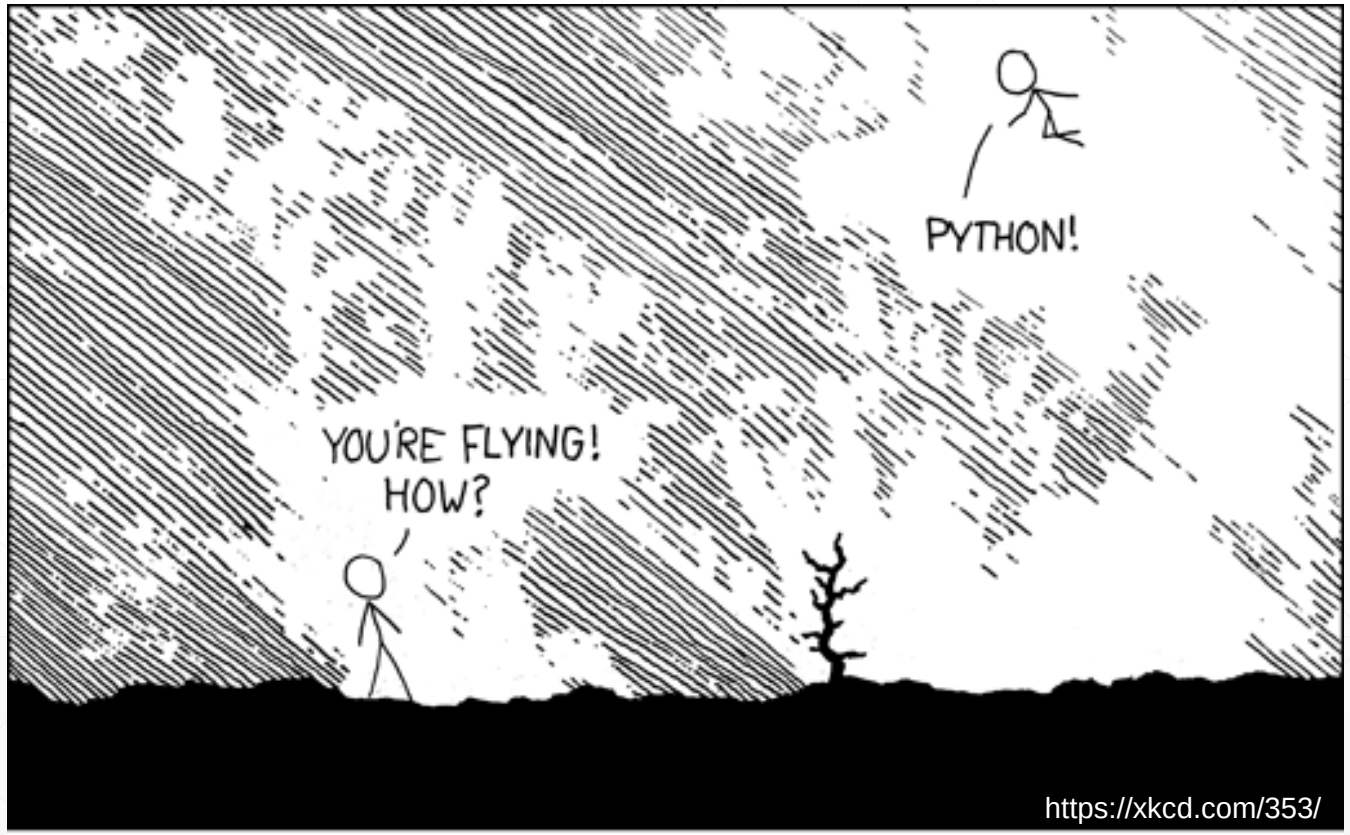


Source: Fabien Maussion

Guido van Rossum (BDFL)

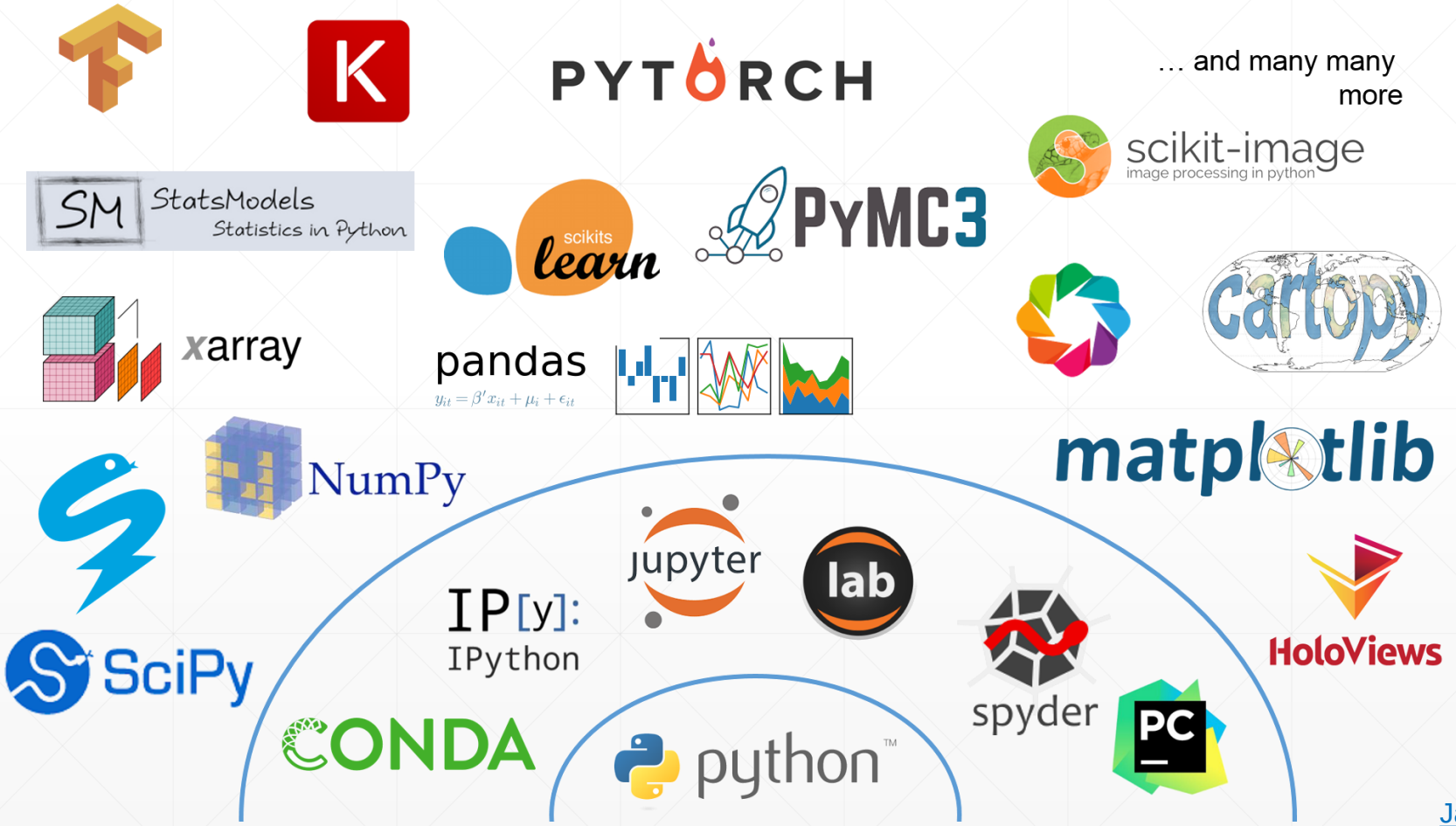
Source: SD Times, 2014

The Python Ecosystem




```
>>> import antigravity
```

The Python Ecosystem



(modified after Jake VanderPlas, 2016)

„…we live in a world where life-long learning is no longer optional…“

Andrew Ng @ The AI Podcast Ep. 32

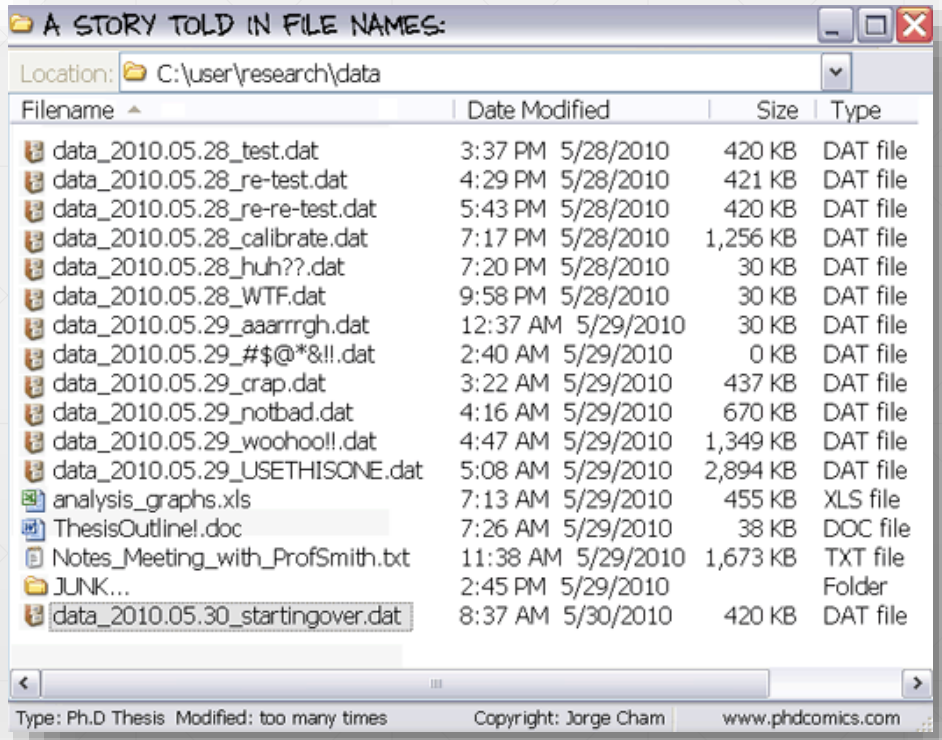
@Coursera

ANDREW NG
VP & Chief Scientist of Baidu;
Co-Chairman and Co-Founder of Coursera;
and an Adjunct Professor at Stanford University.

Demo time ☺

Summary: Reproducible Research

- Data management and informative naming (→ Cookiecutter)
- Version control (→ git)
- Documentation (→ sphinx)
- Use software that can be coded (→ Python)
- Literate programming (→ Jupyter Notebooks)
- Archive and share the materials (→ Open Science Framework and github)

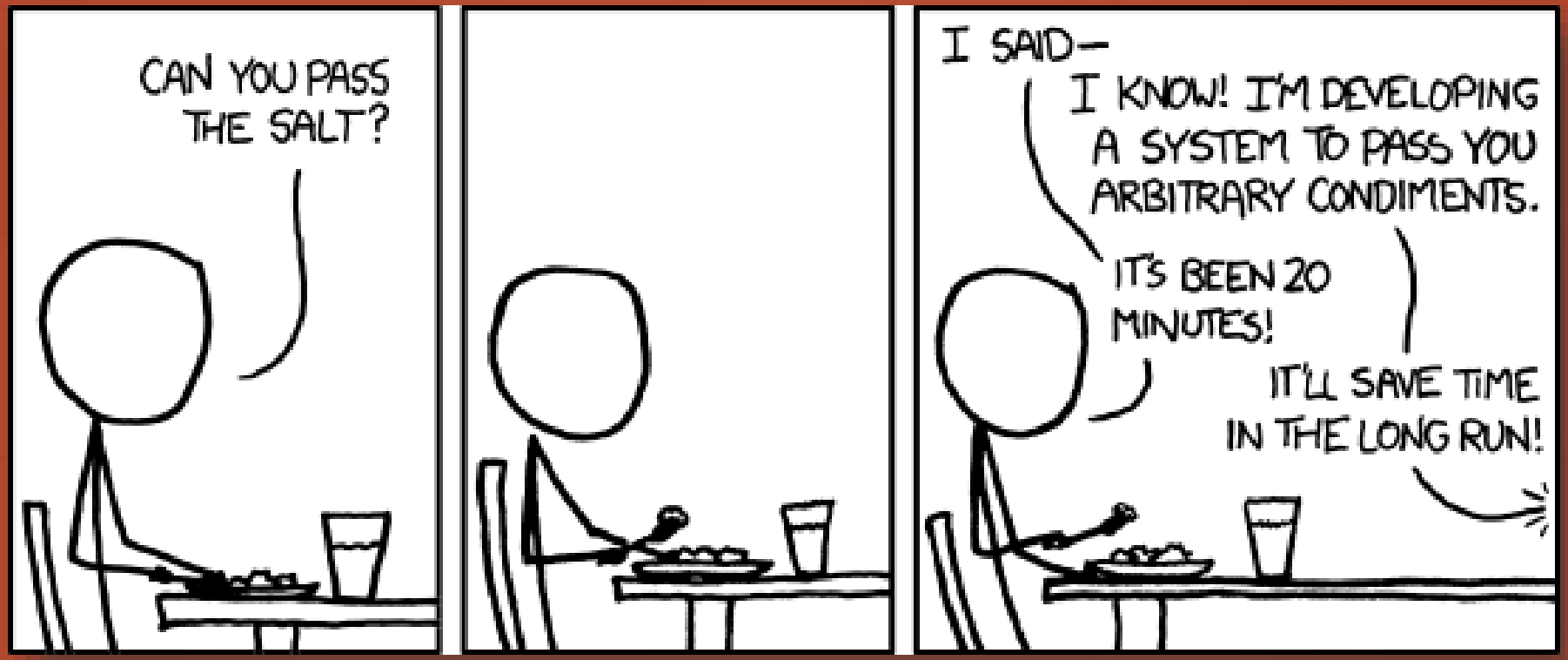


http://phdcomics.com/comics.php?f=1323

References

- Baker (2016) 1,500 scientists lift the lid on reproducibility. Nature, 533(7604):452-4.
- Belluz et al. (2016) The 7 biggest problems facing science, according to 270 scientists.
- Clyburne-Sherin (2015) The Reproducible Research Advantage.
- Engber (2017) Is Science Broken?
- Hutson (2018) Artificial intelligence faces reproducibility crisis, Science, 359, 6377.
- Leyser et al. (2017) The science 'reproducibility crisis' – and what can be done about it.
- Marvin (2014) Python creator proposes type annotations for programming language.
- Mohr H. (1977) The Crisis of Science. In: Lectures on Structure and Significance of Science. Springer, Berlin, Heidelberg
- Peng et al. (2006) Reproducible epidemiologic research. American Journal of Epidemiology, 163, 783–789.
- Pisani-Ferry (2016) Why Are Voters Ignoring Experts?
- Saltelli (2016) Science in crisis: from the sugar scam to Brexit, our faith in experts is fading.
- Saltelli and Funtowicz (2017) What is science's crisis really about?, Futures, Volume 91, 5-11.
- Shen (2014) Interactive notebooks: Sharing the code. Nature 515,1 51–152
- Smith (2017) Why 'Statistical Significance' Is Often Insignificant.
- Somers (2018) The Scientific Paper Is Obsolete - Here's what's next.
- VanderPlas (2016) Python's Data Science Stack.
- Van Noorden (2014) Publishers withdraw more than 120 gibberish papers, Nature, doi:10.1038/nature.2014.14763
- Wasserstein and Lazar (2016) The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108
- Witte (2016) 9 out of 10 experts agree: Britain doesn't trust the experts on Brexit.
- Woolston (2015) Psychology journal bans P values, Nature 519, 9, doi:10.1038/519009f

https://github.com/eotp/lunchtime-science-2018



https://xkcd.com/974/

Thanks for your attention