R: A Hitchhikers Guide to Reproducible Research

- Wise up

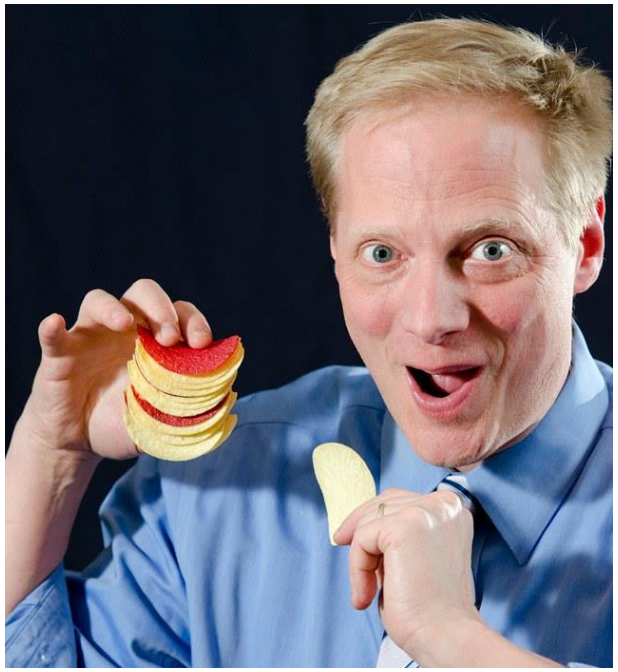
Brendan Palmer,
Clinical Research Facility - Cork &
School of Public Health
@B_A_Palmer

CRF-C
HRB Clinical Research Facility Cork

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

School of
Public Health

Don’t do what Donny Dont does!



“In short, peer review misses all the hard stuff, and a worrying amount of the easy stuff”
James Heathers, Northwestern University

#datathugs

Brian Wansink: The grad student who never said no

“Every day we would scratch our heads, ask "Why," and come up with another way to reanalyze the data with yet another set of plausible hypotheses. Eventually we started discovering solutions”

p-values should not define a study

nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 520 > Issue 7549 > Comment > Article

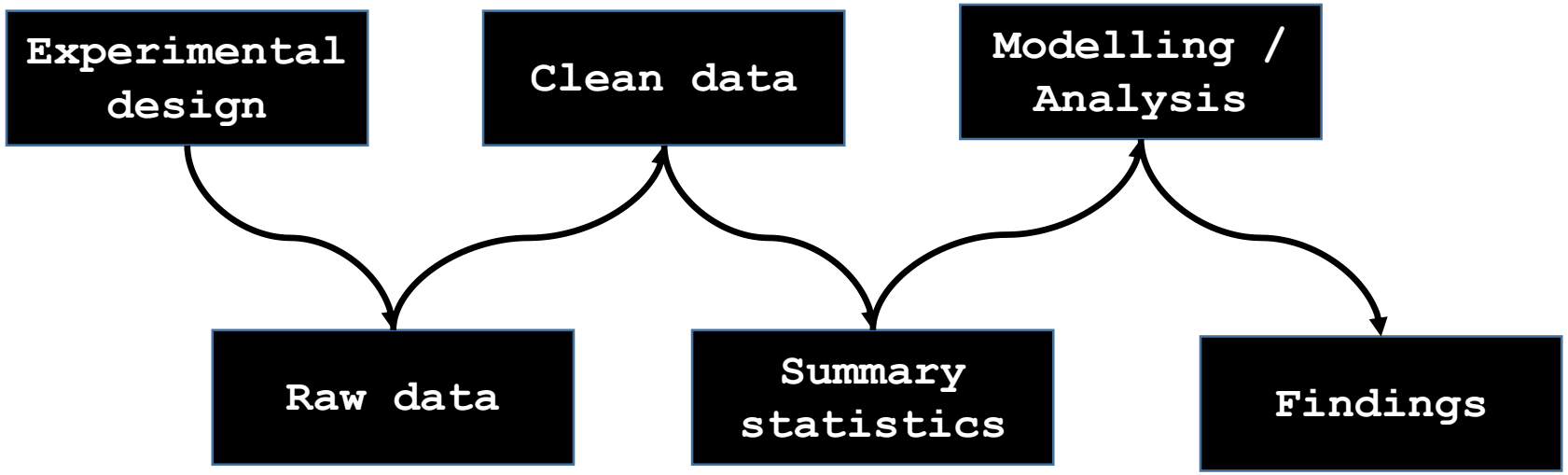
E-alert RSS Facebook Twitter

NATURE | COMMENT

Statistics: *P* values are just the tip of the iceberg

Jeffrey T. Leek & Roger D. Peng

28 April 2015



Extreme scrutiny

Required reading

Taylor & Francis
Taylor & Francis Group

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utas20

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

Ronald L. Wasserstein & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein & Nicole A. 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

To link to this article: https://doi.org/10.1080/00031305.2016.1154108

Eur J Epidemiol (2016) 31:337–350
DOI 10.1007/s10654-016-0149-3

CrossMark

ESSAY

Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations

Sander Greenland[1] · Stephen J. Senn[2] · Kenneth J. Rothman[3] · John B. Carlin[4] · Charles Poole[5] · Steven N. Goodman[6] · Douglas G. Altman[7]

Received: 9 April 2016 / Accepted: 9 April 2016 / Published online: 21 May 2016

Abstract Misinterpretation and abuse of statistical tests, confidence intervals, and statistical power have been decried for decades, yet remain rampant. A key problem is that there are no interpretations of these concepts that are at once simple, intuitive, correct, and foolproof. Instead, correct use and interpretation of these statistics requires an attention to detail which seems to tax the patience of working scientists. This high cognitive demand has led to an epidemic of shortcut definitions and interpretations that are simply wrong, sometimes disastrously so—and yet these misinterpretations dominate much of the scientific literature. In light of this problem, we provide definitions and a discussion of basic statistics that are more general and critical than typically found in traditional introductory expositions. Our goal is to provide a resource for instructors, researchers, and consumers of statistics whose knowledge of statistical theory and technique may be limited but who wish to avoid and spot misinterpretations. We emphasize how violation of often unstated analysis protocols (such as selecting analyses for presentation based on the *P* values they produce) can lead to small *P* values even if the declared test hypothesis is correct, and can lead to large *P* values even if that hypothesis is incorrect. We then provide an explanatory list of 25 misinterpretations of *P* values, confidence intervals, and power. We conclude with guidelines for improving statistical interpretation and reporting.

Editor's note This article has been published online as supplementary material with an article of Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process and purpose. The American Statistician 2016.

Pre-registration

- Documenting detail of method and analysis plan before analysing (or collecting) data

Why?

- Prevents rushing into data collection

- Better ensures that researchers collecting data understand what they are doing

- Prevents p-hacking

Does pre-registration slow projects down?

- Probably

- But also
  - prevents wasting time running sub-optimal designs

  - helps catch confounders

  - improves pedagogy

Does pre-registration slow projects down?

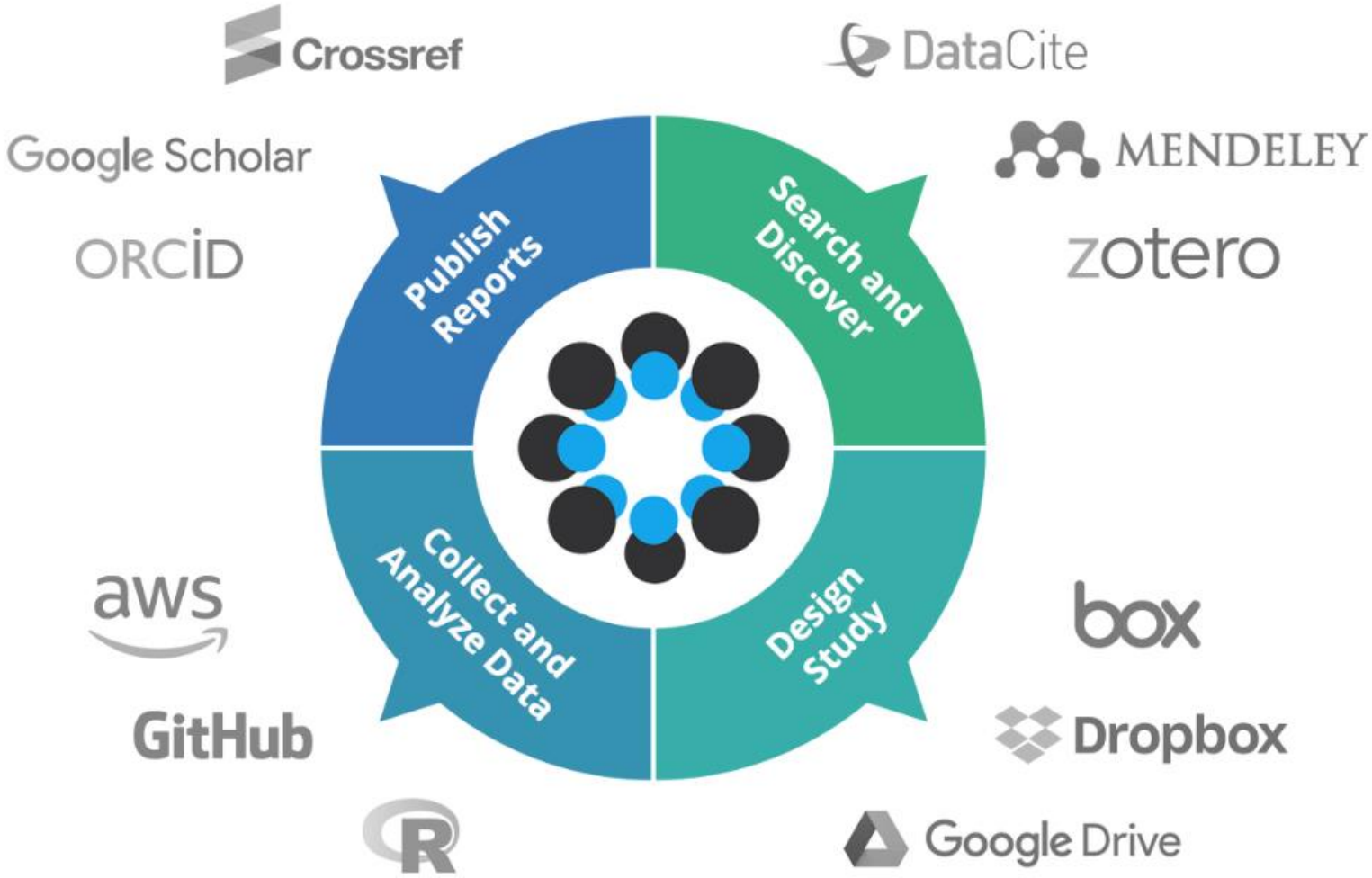
- Probably

- But also
  - prevents wasting time running sub-optimal designs

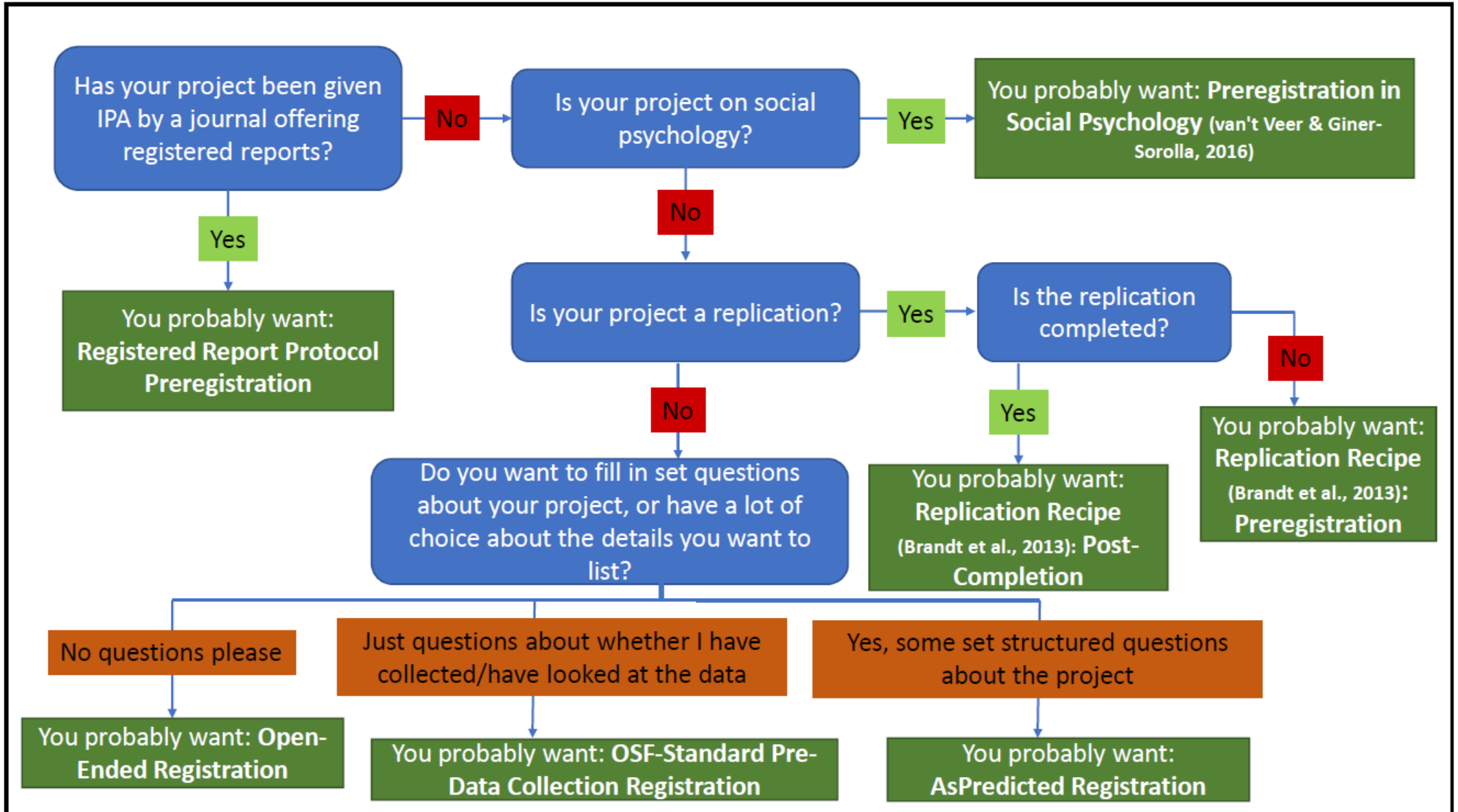
  - helps catch confounders

  - improves pedagogy

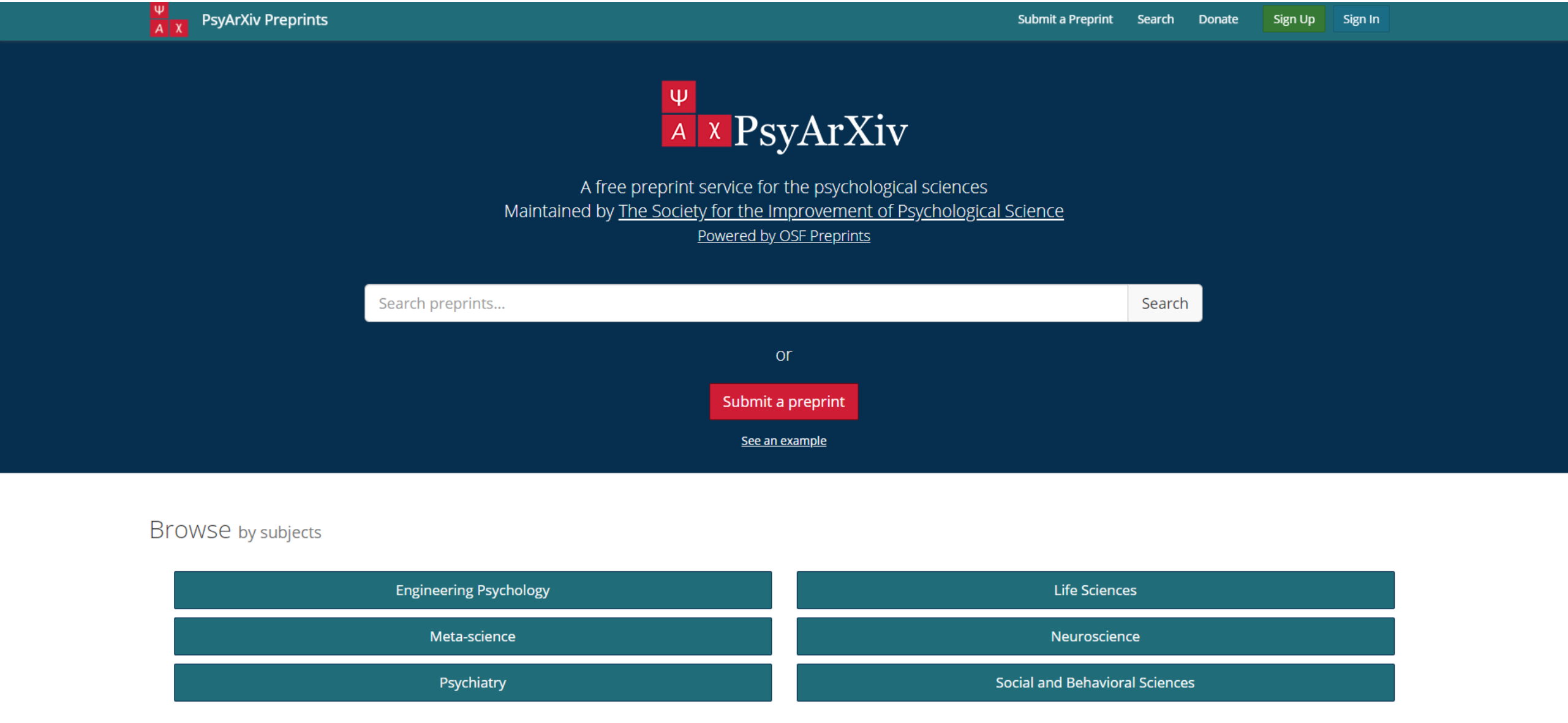
Open Science Framework



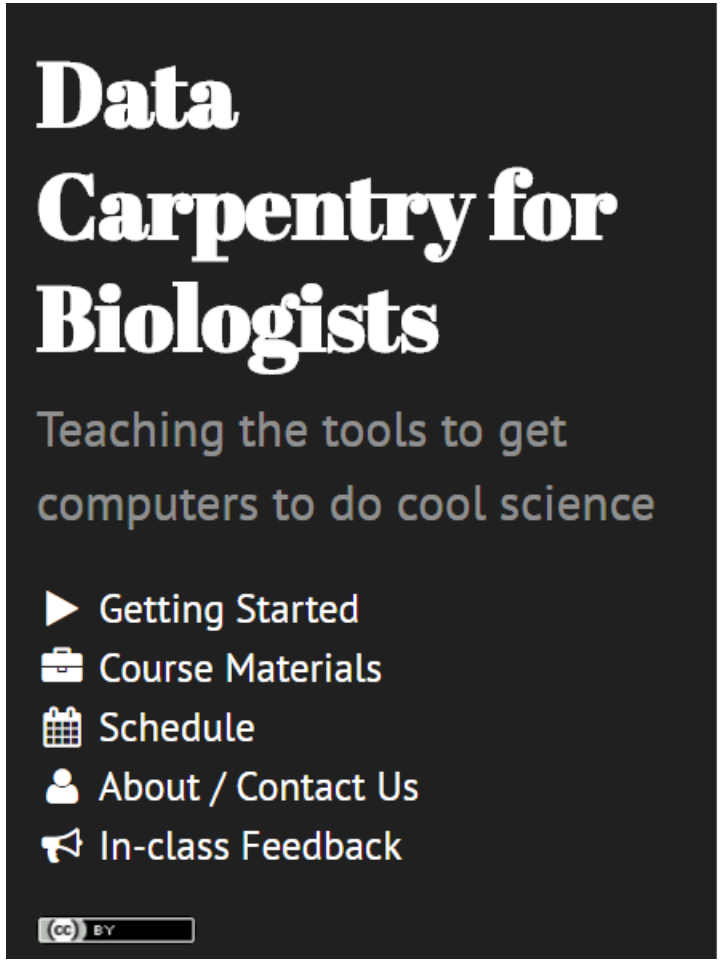
Pre-registration with OSF



PsyArXiv with OSF



And finally



Googling for Help

Notes

> Check that top Google results haven't change and adjust as needed
> Current top 3 hits:
>
> 1. https://blog.exploratory.io/selecting-columns-809bdd1ef615
> 2. https://dplyr.tidyverse.org/reference/select.html
> 3. https://stackoverflow.com/questions/21502465/replacement-for-rename-in-dplyr

Link to the data carpentry webpage

Digital Badge
- Responsible Conduct of Research

Digital Badge in the Responsible Conduct of Research

The UCC Digital Badge in the Responsible Conduct of Research is offered to researchers who complete a tailored, blended learning course on responsible, reproducible research. The digital badge aims enhance Research Integrity awareness among the research community in UCC by promoting frank discussions about responsible research. Transparent data practices are a key skill of RI and participants will also learn about the FAIR principles, good data management practices and reproducibility. The badge is available to researchers at all levels and career stages from PhD to PI.

- Coming to a School of Applied Psychology near you!!

meetup

Cork (Ireland) R-Users Group