

Why doing reproducible research?

Francisco Rodríguez-Sánchez

<https://frodriguezsanchez.net>

Reproducibility: good for you,
good for everyone

Automation (good code) saves time

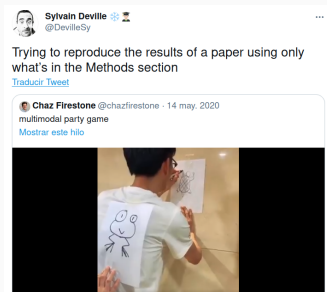


Trevor Branch
@TrevorABranch

...

My rule of thumb: every analysis you do on a dataset will have to be redone 10–15 times before publication. Plan accordingly. [#Rstats](#)

Code = fully traceable, reproducible analysis



Code advantages:

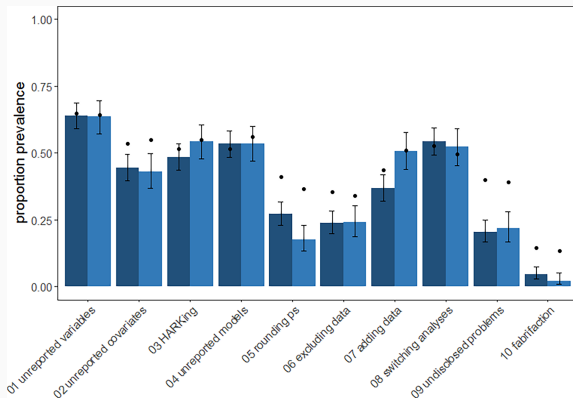
- Easier writing
- Easier, deeper review
- Reusable

Transparency prevents bad practices

RESEARCH ARTICLE

Questionable research practices in ecology and evolution

Hannah Fraser^{1*}, Tim Parker², Shinichi Nakagawa³, Ashley Barnett¹, Fiona Fidler^{1,4}



p-hacking, HARKing, data fabrication...

Transparency avoids unnecessary disputes

DOI:10.1063/PT.6.1.20180822a

22 Aug 2018 in [Research & Technology](#)

The war over supercooled water

How a hidden coding error fueled a seven-year dispute between two of condensed matter's top theorists.

Ashley G. Smart

Over the next seven years, the perplexing discrepancy would ignite a bitter conflict, with junior scientists caught in the crossfire. At stake were not only the reputations of the two groups but also a peculiar theory that sought to explain some of water's deepest and most enduring mysteries. Earlier this year, the dispute was finally settled. And as it turns out, the entire ordeal was the result of botched code.

Transparency brings better science



Alexey Shiklomanov

@ashiklom711

...

I'm co-author on a study currently published only as a publicly available discussion paper. My code was on GitHub.

A colleague read the paper, thought the results looked weird, checked my code, found a bug and emailed me about it.

This is how science should work. [#openscience](#)

Many journals (and funders) value/require reproducibility

As a condition for publication in ESA journals, all underlying data and statistical code pertinent to the results presented in the publication must be made available in a permanent, publicly accessible data archive or repository, with rare exceptions (see



Many journals value reproducibility

'Papers with exemplary **data and code archiving**
are **more valuable** for future research and [...]
will be given **higher priority** for publication'
(*Molecular Ecology*)

Many journals require reproducibility

EDITORIAL

ECOLOGY LETTERS  WILEY

From raw data to publication: Introducing data editing at Ecology Letters

‘We require the **data and code** for reproducing statistical results and generating figures and tables’

‘This material will need to be supplied at the **time of submission**’

Higher impact: cites, reuse, reputation

RESEARCH ARTICLE



The citation advantage of linking publications to research data

Giovanni Colavizza^{1,2}, Iain Hrynaskiewicz^{3,4}, Isla Staden^{1,5}, Kirstie Whitaker^{1,6},
Barbara McGillivray^{1,6*}

Colavizza et al 2020

ACADEMIC PRACTICE IN ECOLOGY AND EVOLUTION

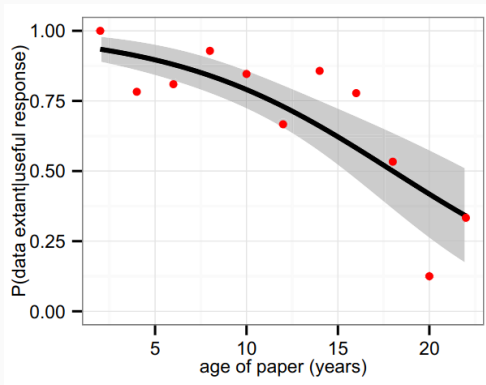
Ecology and Evolution
WILEY

Code sharing in ecology and evolution increases citation rates
but remains uncommon  

Maitner et al 2024

Let's stop losing data & code

The Availability of Research Data Declines Rapidly with Article Age



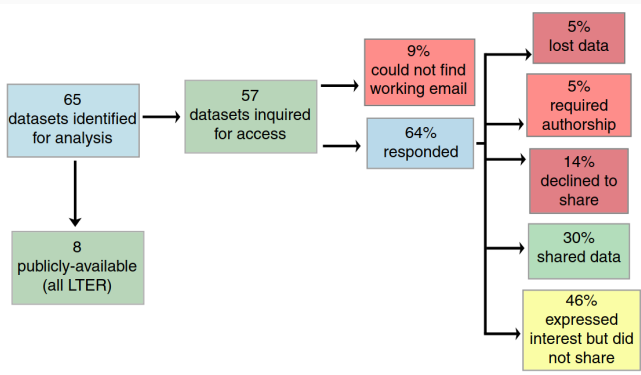
Vines et al 2014

Open data & code enable synthesis

REVIEW

Advances in global change research require open science by individual researchers

ELIZABETH M. WOLKOVICH^{*†}, JAMES REGETZ[‡] and MARY I. O'CONNOR[†]



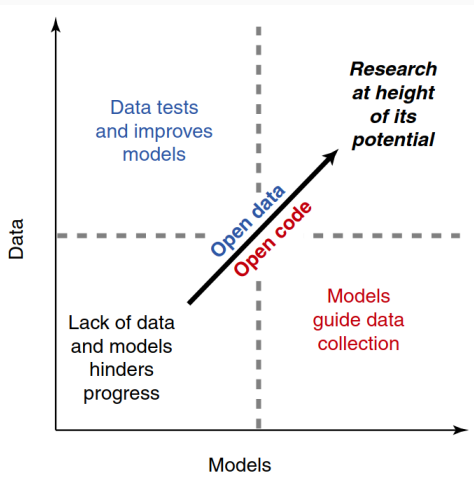
Wolkovich et al 2012

Open data & code enable synthesis

REVIEW

Advances in global change research require open science by individual researchers

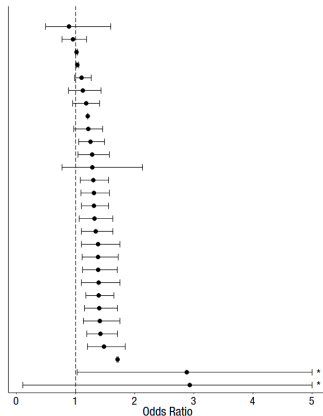
ELIZABETH M. WOLKOVICH^{*†}, JAMES REGETZ[‡] and MARY I. O'CONNOR[†]



Same data -> different results

Do soccer referees give more red cards to dark-skin players?

Team	Analytic Approach	Odds Ratio
12	Zero-Inflated Poisson Regression	0.89
17	Bayesian Logistic Regression	0.96
15	Hierarchical Log-Linear Modeling	1.02
10	Multilevel Regression and Logistic Regression	1.03
18	Hierarchical Bayes Model	1.10
31	Logistic Regression	1.12
1	OLS Regression With Robust Standard Errors, Logistic Regression	1.18
4	Spearman Correlation	1.21
14	WLS Regression With Clustered Standard Errors	1.21
11	Multiple Linear Regression	1.25
30	Clustered Robust Binomial Logistic Regression	1.28
6	Linear Probability Model	1.28
26	Hierarchical Generalized Linear Modeling With Poisson Sampling	1.30
3	Multilevel Logistic Regression Using Bayesian Inference	1.31
23	Mixed-Model Logistic Regression	1.31
16	Hierarchical Poisson Regression	1.32
2	Linear Probability Model, Logistic Regression	1.34
5	Generalized Linear Mixed Models	1.38
24	Multilevel Logistic Regression	1.38
28	Mixed-Effects Logistic Regression	1.38
32	Generalized Linear Models for Binary Data	1.39
8	Negative Binomial Regression With a Log Link	1.39
20	Cross-Classified Multilevel Negative Binomial Model	1.40
13	Poisson Multilevel Modeling	1.41
25	Multilevel Logistic Binomial Regression	1.42
9	Generalized Linear Mixed-Effects Models With a Logit Link	1.48
7	Dirichlet-Process Bayesian Clustering	1.71
21	Tobit Regression	2.88
27	Poisson Regression	2.93



29 teams: 2/3 found significant effect

73 teams testing the same hypothesis with the same data

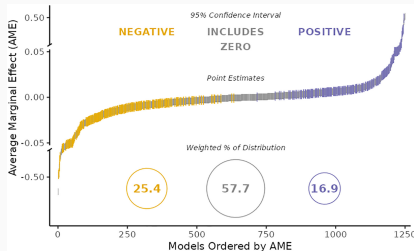
PNAS

RESEARCH ARTICLE

SOCIAL SCIENCES

OPEN ACCESS

Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty



‘This reveals a **universe of uncertainty** that remains hidden when considering a single study in isolation’

‘These results call for greater **epistemic humility** and **clarity in reporting** scientific findings’

132 teams asking same question with same data

How does sibling competition affect nestling growth in blue tits?

