Data Science Applications to Politics Research

Week 2 Seminar

Zach Dickson

London School of Economics

GV330

P-hacking

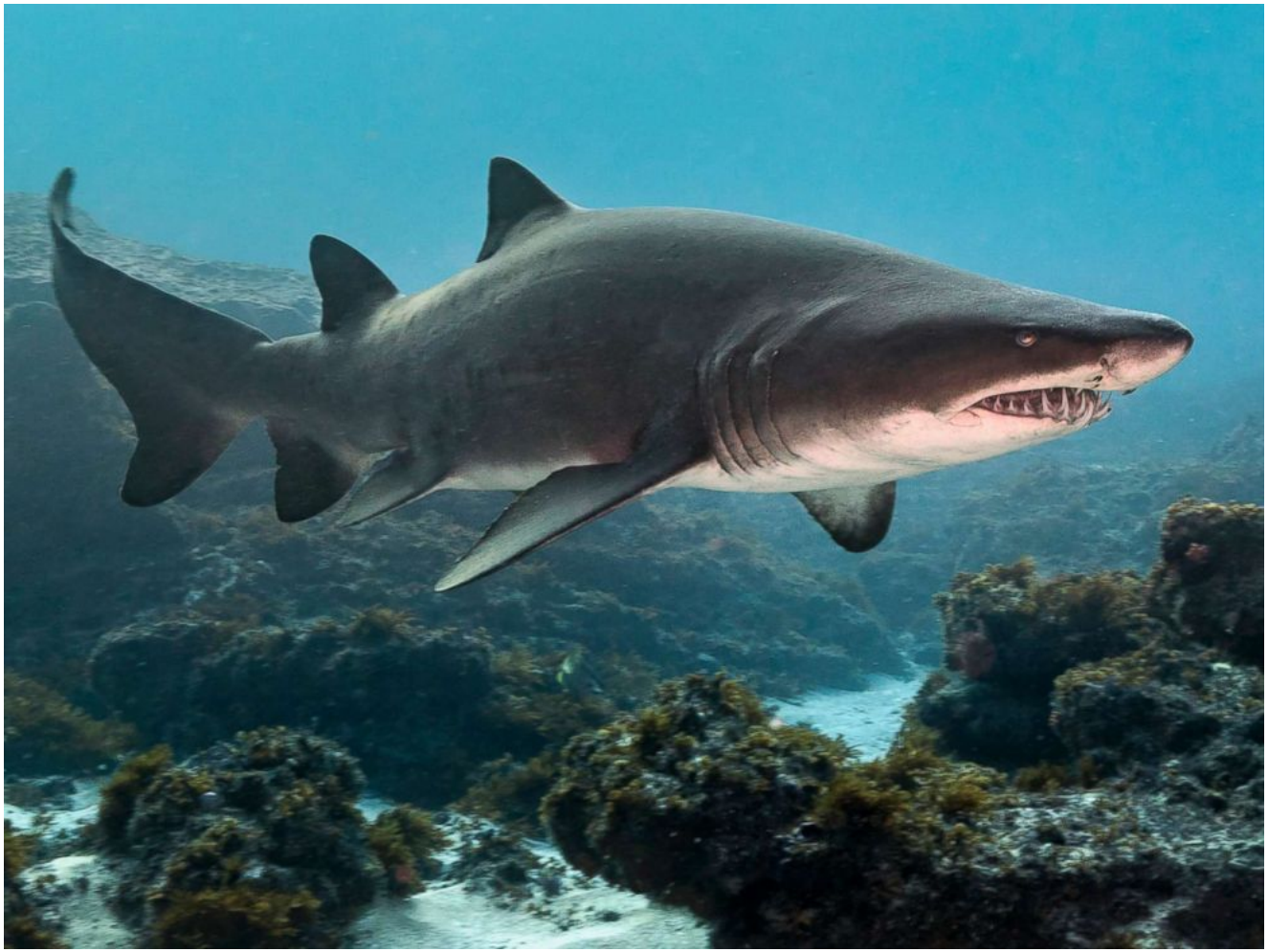
"If you torture the data long enough, it will confess." - Ronald Coase

P-hacking

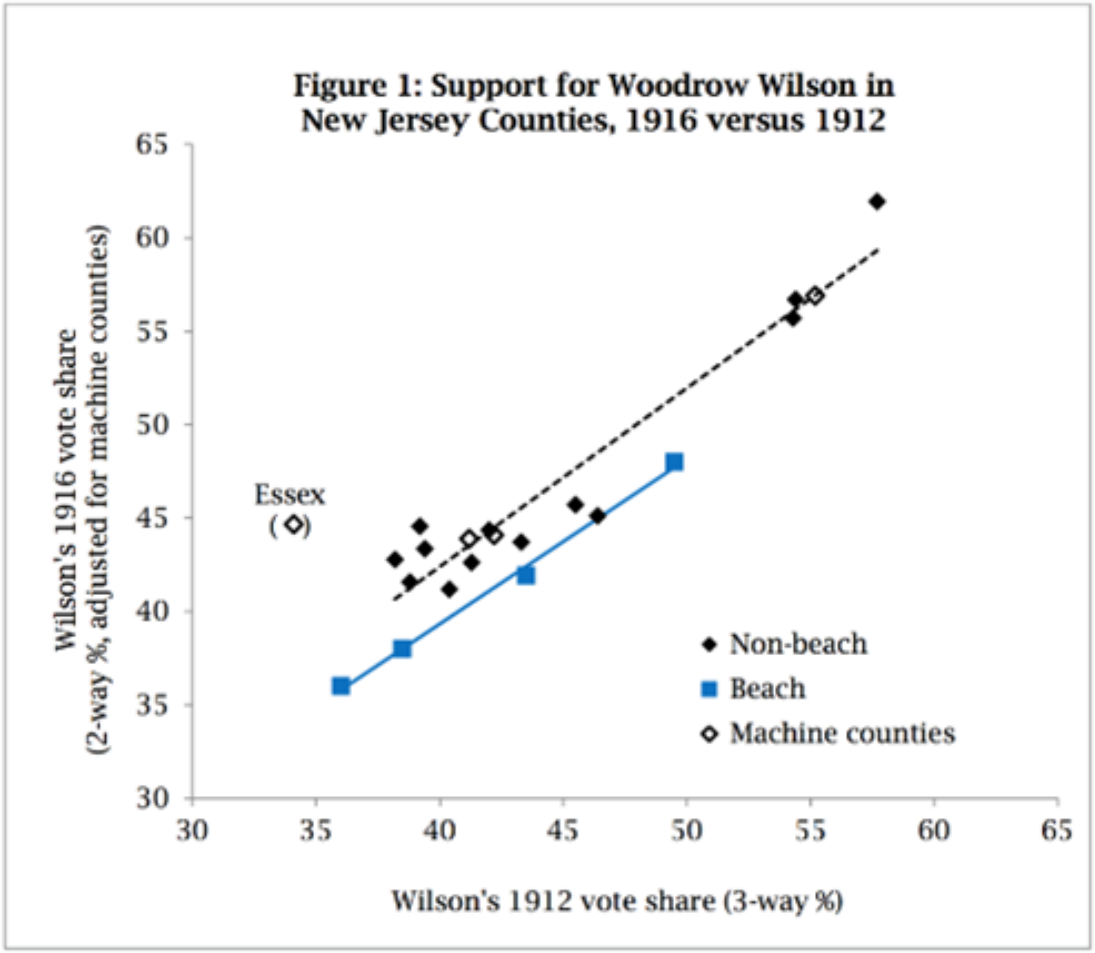
"If you torture the data long enough, it will confess." - Ronald Coase

- **Opportunity:** Researchers have many "degrees of freedom" in the design and analysis of a study $\rightarrow$ p-hacking (may not always be intentional, see Gelman & Loken 2013)
- **Motive:** Researchers have incentives (from journals, tenure requirements, etc.) to find statistical significance
- **Result:** Biased evidence base (also contributes to replication crisis)

Shark attacks and voting (Achen & Bartels 2017)



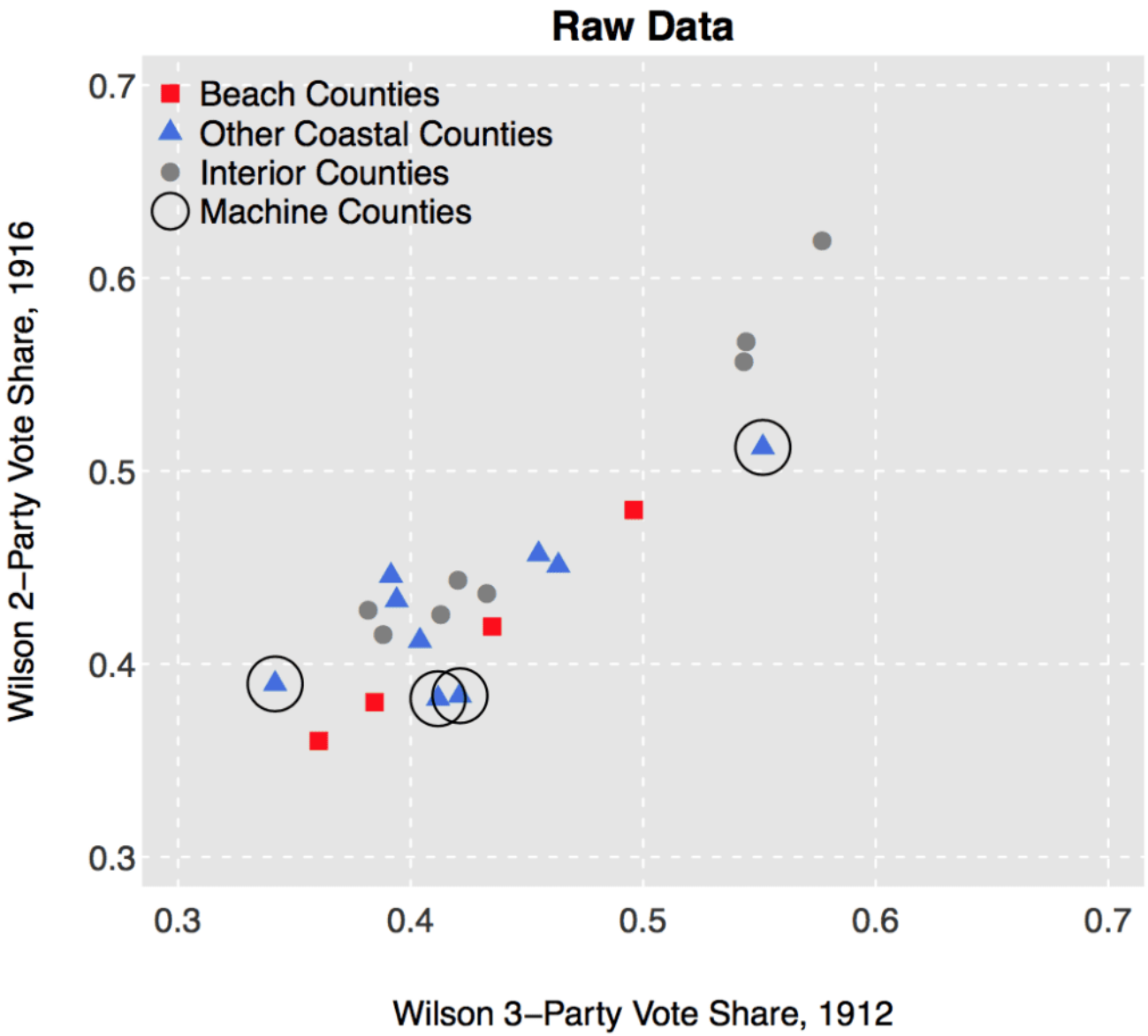
Shark attacks and voting (Achen & Bartels 2017)



"A dramatic series of shark attacks in NJ in 1916... voters significantly punished the incumbent president at the polls"

Achen, C., & Bartels, L. (2017). *Democracy for Realists*. Princeton University Press.

Fowler & Hall 2018: Re-analysis of Achen & Bartels



Critique of Achen & Bartels (Fowler & Hall 2018)

- Correlation: election results from different counties are not independent events (can inflate our sense of the "true" sample size)
- 'Forking paths': many different ways to analyze the data (e.g., choice of how to treat "outliers")

"We assemble data on every fatal shark attack in U.S. history and county-level returns from every presidential election between 1872 and 2012, and we find little systematic evidence that shark attacks hurt incumbent presidents or their party."

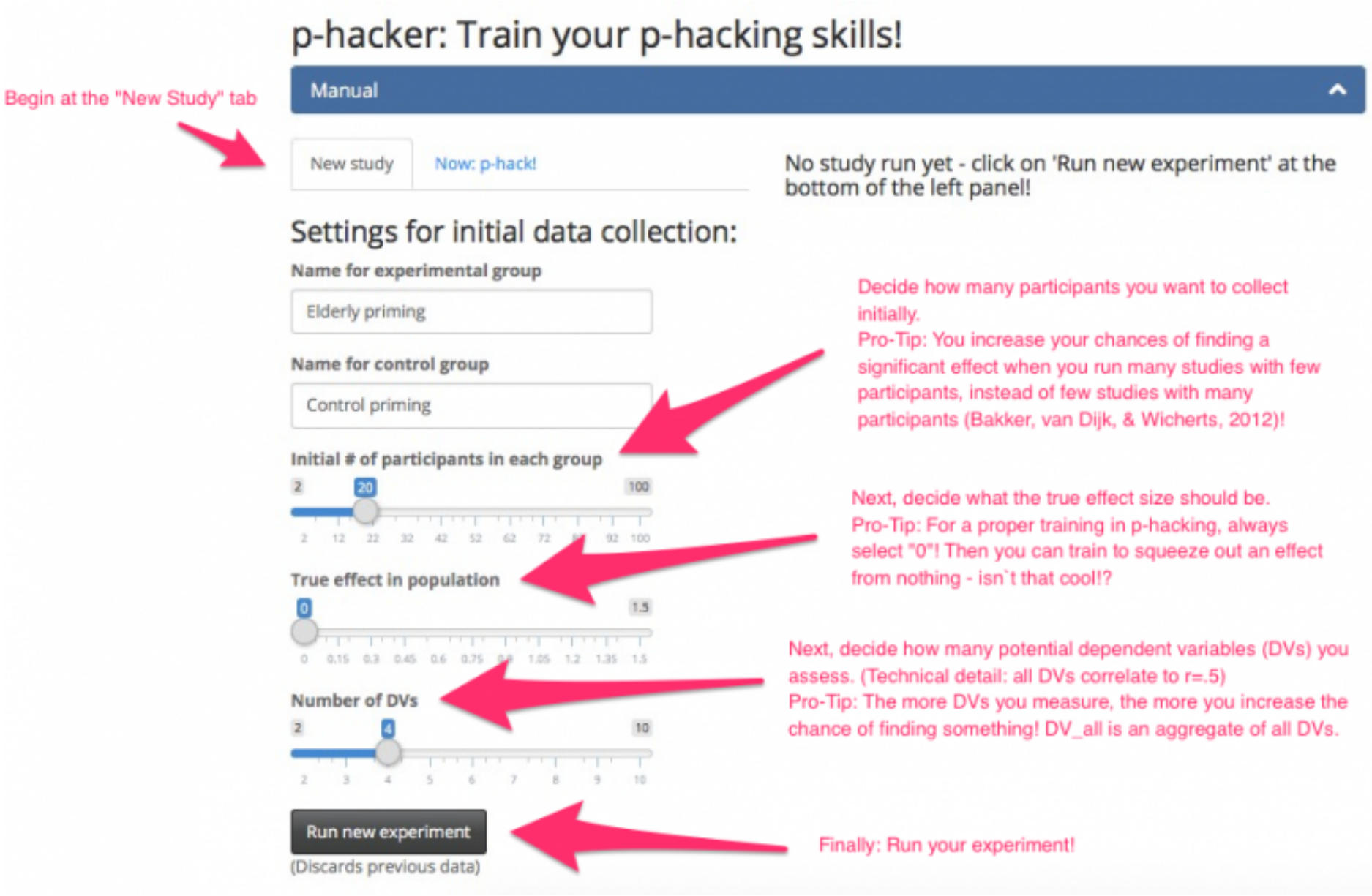
Fowler, A., & Hall, A. B. (2018). Do shark attacks influence presidential elections? Reassessing a prominent finding on voter competence. *The Journal of Politics*, 80(4), 1423-1437.

Activity

Go to
http://shinyapps.org/apps/p-hacker/

Schönbrodt, F. D. (2016). p-hacker: Train your p-hacking skills!

Set up your study



Inspect results of your study

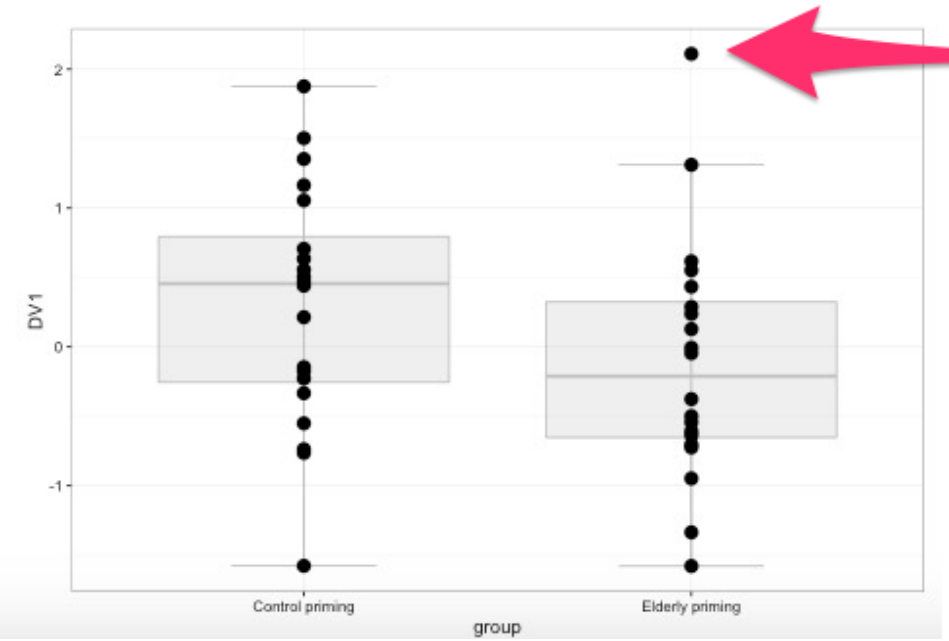
Tests for each DV

Name	N	Statistic	p-Value	Sign.	Actions
DV1	40	F(1, 38) = 2.26	p = .141		
DV2	40	F(1, 38) = 0.38	p = .543	ns	Save
DV3	40	F(1, 38) = 0.03	p = .859	ns	Save
DV4	40	F(1, 38) = 0.73	p = .400	ns	Save
DV_all	40	F(1, 38) = 0.1	p = .750	ns	Save

Not too bad! p = .14 in your first DV is a clear trend towards significance. (Discard the other DVs, focus on DV1)

Choose DV to plot

DV1



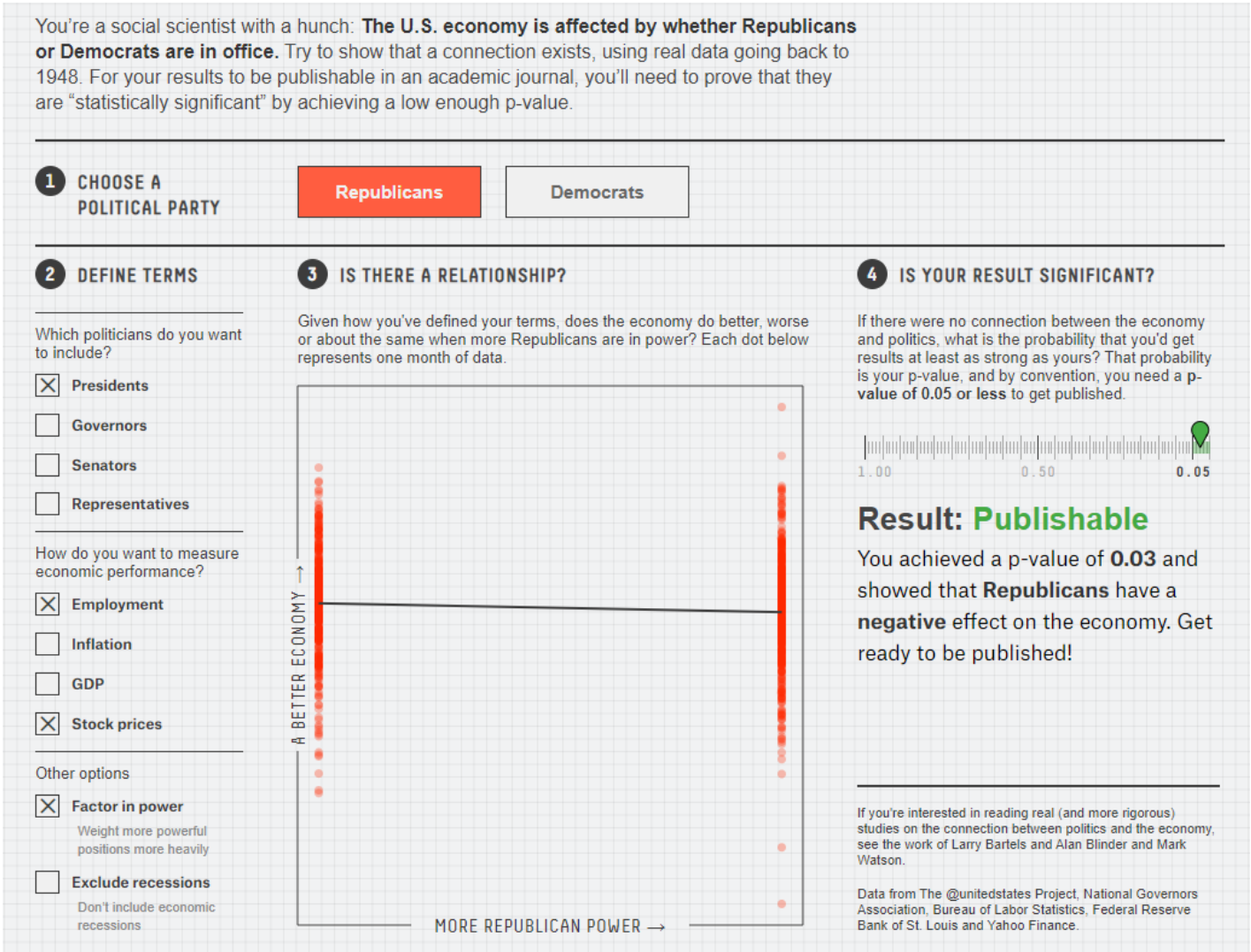
Here's a clear outlier (you can justify its exclusion with the box plot rule).

Try to click on a data point to exclude it!

Can’t get a significant result? “Now p-hack!”

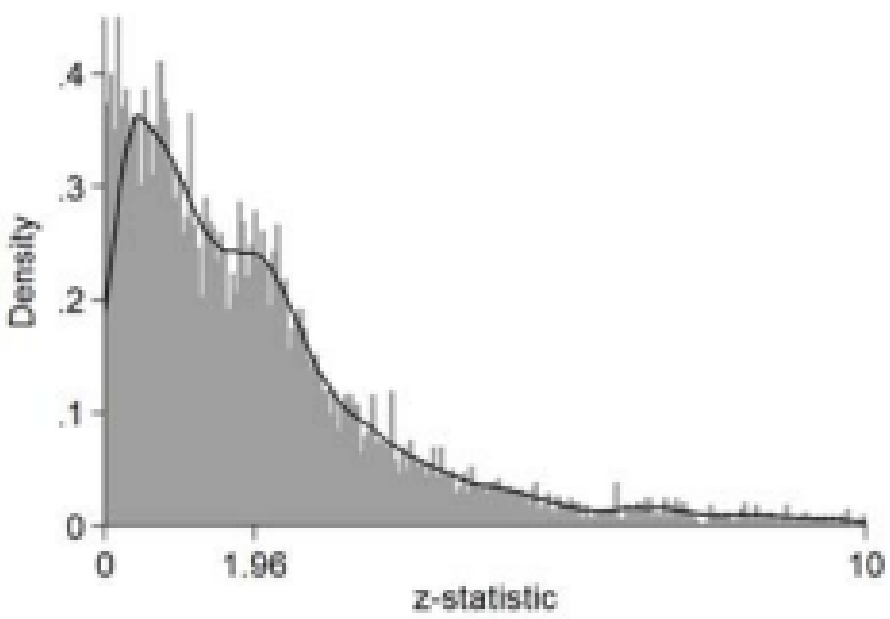
A similar tool

See also https://projects.fivethirtyeight.com/p-hacking/

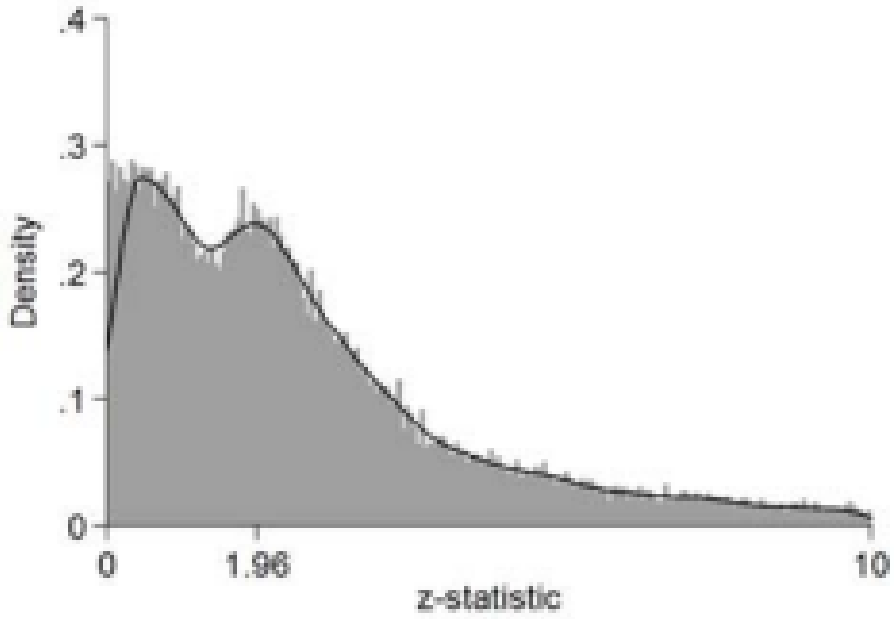


Brodeur et al. (2016)

Evidence of P-Hacking:



(a) Lab. experiments or RCT data.



(b) Other data.

Wicherts et al. (2016)

Identify 34 key researcher degrees of freedom (see article for full list):

Table 1
Checklist for different types of degrees of freedom in the planning, executing, analyzing, and reporting of psychological studies

Code	Related	Type of Degrees of Freedom
		Hypothesizing
T1	R6	Conducting explorative research without any hypothesis
T2		Studying a vague hypothesis that fails to specify the direction of the effect
		Design
D1	A8	Creating multiple manipulated independent variables and conditions
D2	A10	Measuring additional variables that can later be selected as covariates, independent variables, mediators, or moderators
D3	A5	Measuring the same dependent variable in several alternative ways
D4	A7	Measuring additional constructs that could potentially act as primary outcomes
D5	A12	Measuring additional variables that enable later exclusion of participants from the analyses (e.g., awareness or manipulation checks)
D6		Failing to conduct a well-founded power analysis
D7	C4	Failing to specify the sampling plan and allowing for running (multiple) small studies

...

Misconduct & Fraud: Rare(?) but serious

Misconduct & Fraud: Rare(?) but serious

- **Includes**: Falsifying some or all data and/or results, as well as plagiarism and other forms of misconduct

Misconduct & Fraud: Rare(?) but serious

- **Includes**: Falsifying some or all data and/or results, as well as plagiarism and other forms of misconduct
- **Result:** False or biased evidence base

Misconduct & Fraud: Rare(?) but serious

- **Includes**: Falsifying some or all data and/or results, as well as plagiarism and other forms of misconduct
- **Result:** False or biased evidence base
- **Note**: Fabrication of data (e.g., LaCour & Green 2014) less common than other “questionable research practices"

LaCour & Green 2014

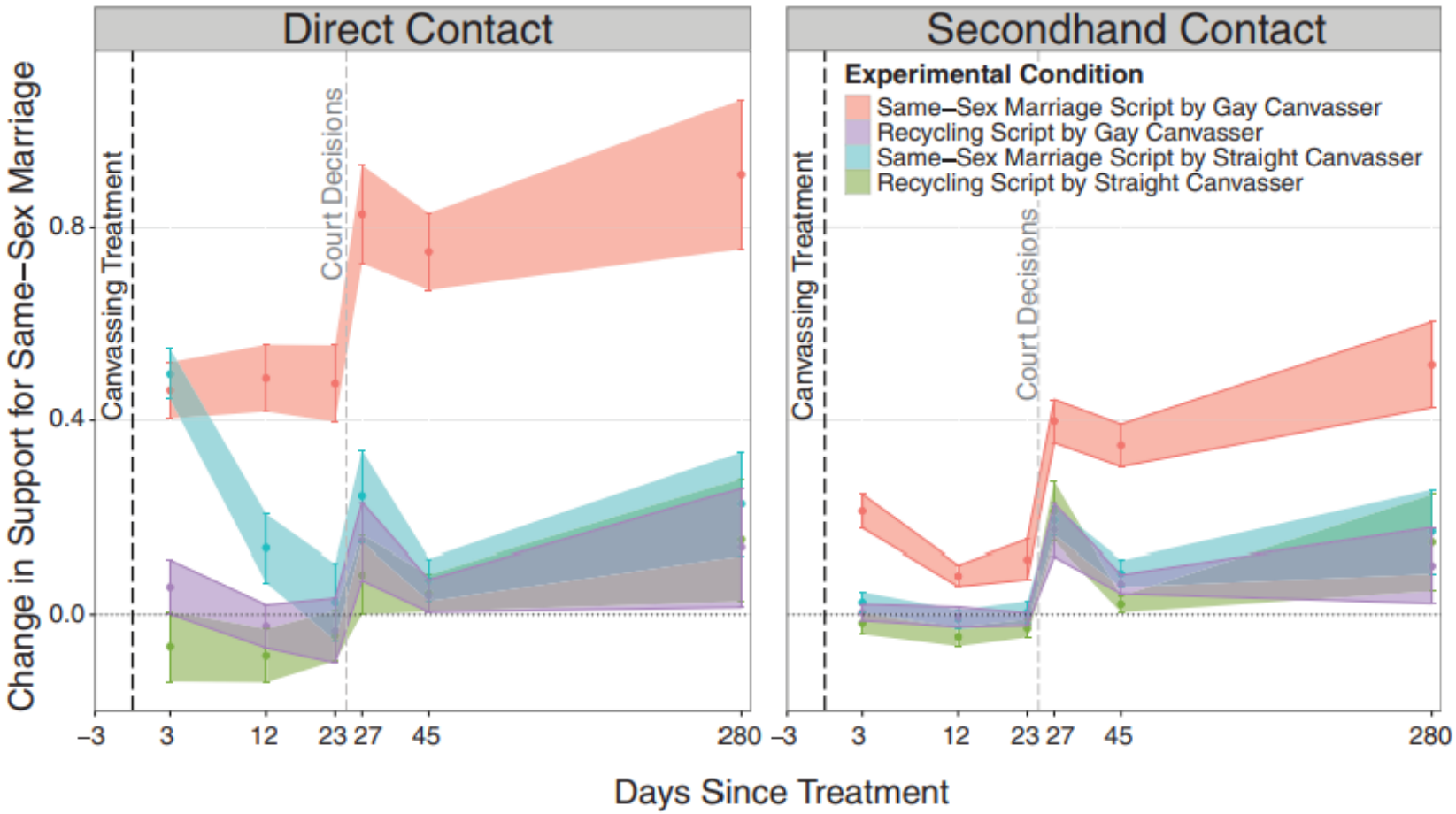


Fig. 1. Direct and secondhand effects on support for same-sex marriage, by assigned message and messenger, and time since treatment. The first vertical dashed line represents the canvassing intervention, which was administered between Internet survey waves 1 and 2. The second vertical dashed line represents the U.S. Supreme Court decisions striking down California's ban on same-sex marriage. The *y* axis is opinion change between the baseline survey and subsequent survey waves, with higher scores indicating more support for same-sex marriage. Points represent mean values, and bars display

Is extrasensory perception (ESP) real?

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

0022-3514/11/$12.00 DOI: 10.1037/a0021524

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size (*d*) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about psi, issues of replication, and theories of psi are also discussed.

Fraud?

"I would start one [experiment], and if it just wasn't going anywhere, I would abandon it and restart it with changes. . . I didn't keep very close track of which ones I had discarded and which ones I hadn't. . . .I was probably very sloppy at the beginning. I think probably some of the criticism could well be valid. I was never dishonest, but on the other hand, the critics were correct."

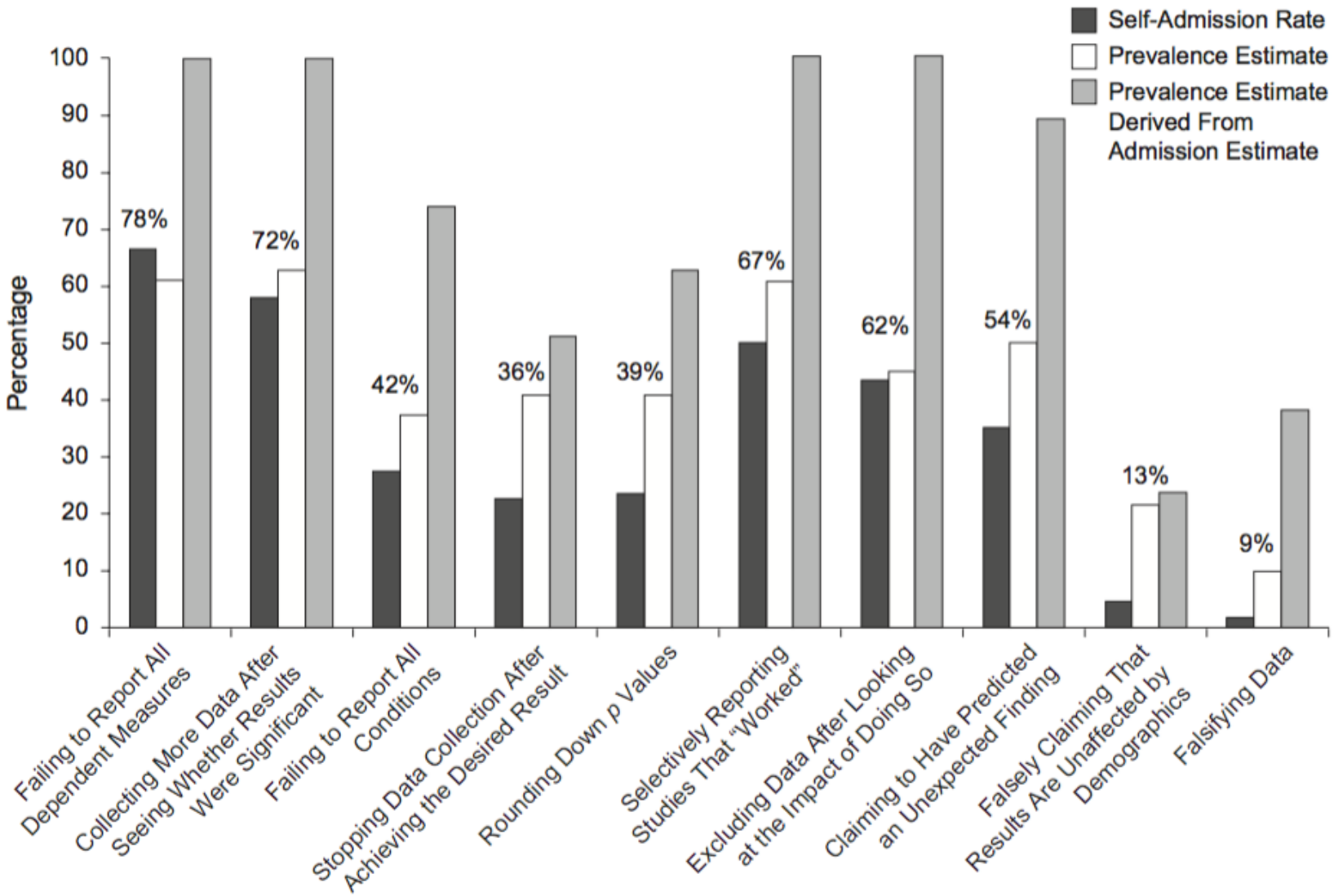
–Daryl Bem

"I'm all for rigor, but I prefer other people do it. I see its importance—it's fun for some people—but I don't have the patience for it."

–Daryl Bem, quoted in Slate Magazine

John et al. (2012)

Survey of 2000 psychologists on questionable practices:



What about solutions?

- **Transparency:** Pre-registration, open data, open code
- **Reproducibility:** Making data and code available
- **Replication:** Direct replications, pre-registered replications
- **Incentives:** Journals, funders, universities, etc.
- **Training:** Teaching good research practices
- **Tools:** Software to detect p-hacking, etc.
- **Punishment:** Retraction, job loss, etc.
- **Whistleblowing:** Encouraging reporting of misconduct