



Fair?



---

---

---

---

---

---

---

---



Heads 6/6  
Heads  
Heads .5<sup>5</sup>  
Heads  
Heads .031  
Heads

---

---

---

---

---

---

---

---

Sample Size



Heads 6/6  
Heads  
Heads .5<sup>5</sup>  
Heads  
Heads .031  
Heads

you flipped 6 times;  
could have flipped more or less

---

---

---

---

---

---

---

---

## Null Hypothesis



Heads 6/6  
Heads .5<sup>5</sup>  
Heads .031  
Heads

what might we **refute**?  
the hypothesis that the coin is fair

---

---

---

---

---

---

---

---

## Statistical Significance

(p-value)



Heads 6/6  
Heads .5<sup>5</sup>  
Heads .031  
Heads

the probability of getting a result at least as  
extreme as this (is .031 enough to be convincing?)

---

---

---

---

---

---

---

---

## Null Hypothesis Significance Testing

how unlikely would it be for you to find the data  
you found if there were **no** effect of interest

---

---

---

---

---

---

---

---



## Sample Size



Heads 6/6  
Heads  
Heads .5<sup>5</sup>  
Heads  
Heads .031  
Heads

---

---

---

---

---

---

---

---

## Sample Size

how many is enough? 100? 10? 10000?

no right answer; but, all else equal, more data will allow you to be more confident in your findings

the "same" sample size can mean very different things

---

---

---

---

---

---

---

---

## Sample Size

how many is enough? 100? 10? 10000?

no right answer; but, all else equal, more data will allow you to be more confident in your findings

the "same" sample size can mean very different things

---

---

---

---

---

---

---

---

# Experimental Design

Is a good night's sleep better than an all-nighter for exam performance?

## Between Subjects

100 people

50 sleep 8 hrs before Intro Psych exams #1 and #2

50 pull all-nighters before Intro Psych exams #1 and #2

## Within Subjects

100 people

50 sleep 8 hrs before Intro Psych exam #1, and pull all-nighter before exam #2

50 pull all-nighter before Intro Psych exam #1, and sleep 8 hrs before exam #2

---

---

---

---

---

---

---

---

# Experimental Design

Is a good night's sleep better than an all-nighter for exam performance?

## Between Subjects

Sleep	No Sleep
S1 92	85 S6
S2 70	64 S7
S3 88	91 S8
S4 85	68 S9
S5 65	82 S10
<b>Avg 80</b>	<b>Avg 78</b>

## Within Subjects

---

---

---

---

---

---

---

---



By Daniel Engber

TWEET

SHARE

## Daryl Bem Proved ESP Is Real

Which means science is broken.

---

---

---

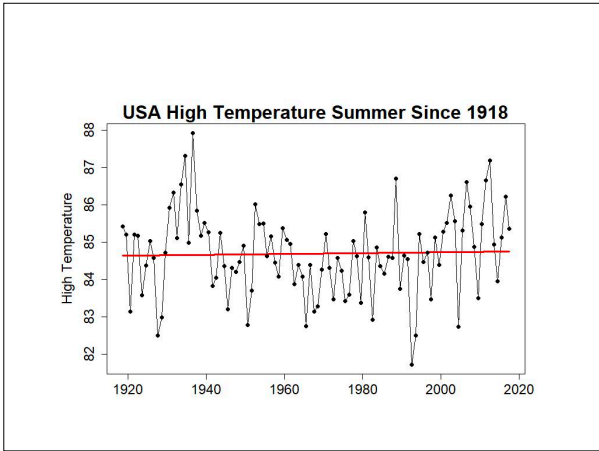
---

---

---

---

---




---

---

---

---

---

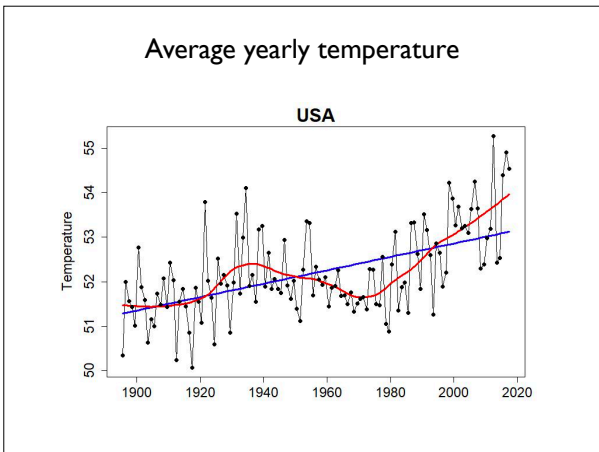
---

---

---

---

---




---

---

---

---

---

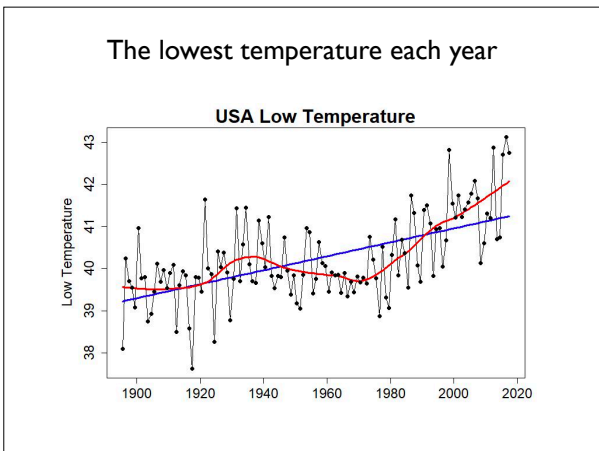
---

---

---

---

---




---

---

---

---

---

---

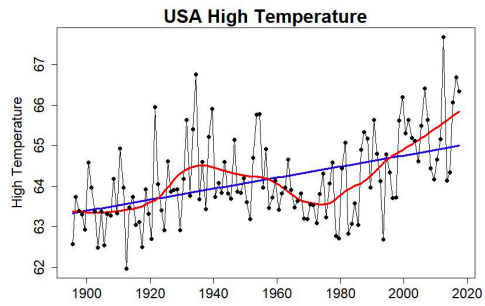
---

---

---

---

### The highest temperature each year



---

---

---

---

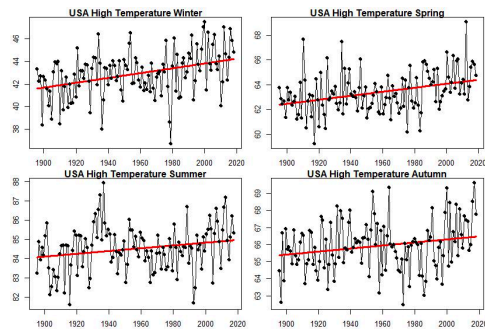
---

---

---

---

### The highest temperature each year, by season



---

---

---

---

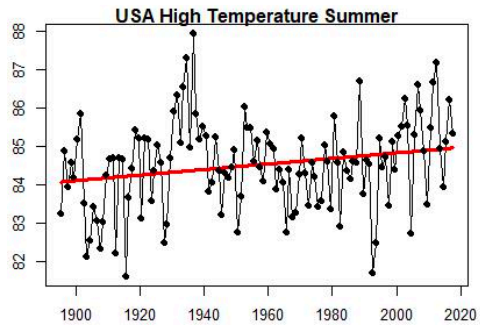
---

---

---

---

### The highest summer temperature each year



---

---

---

---

---

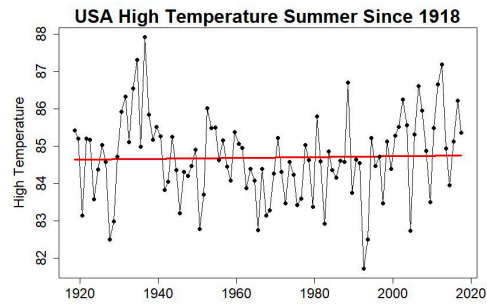
---

---

---



## The highest temperature each summer, since 1918



---

---

---

---

---

---

---

---

---

---

If you do enough tests, you can  
“discover” almost anything!

and if you only tell everyone about the tests  
that “worked”, you **mislead** your readers!

---

---

---

---

---

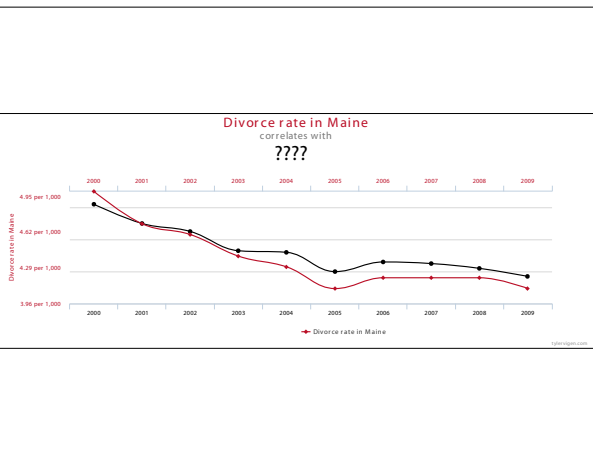
---

---

---

---

---



---

---

---

---

---

---

---

---

---

---

**What** correlates very highly with the divorce rate in Maine?

- A** # of jobs lost in Maine
- B** Church attendance
- C** lbs. of margarine consumed nationally

---

---

---

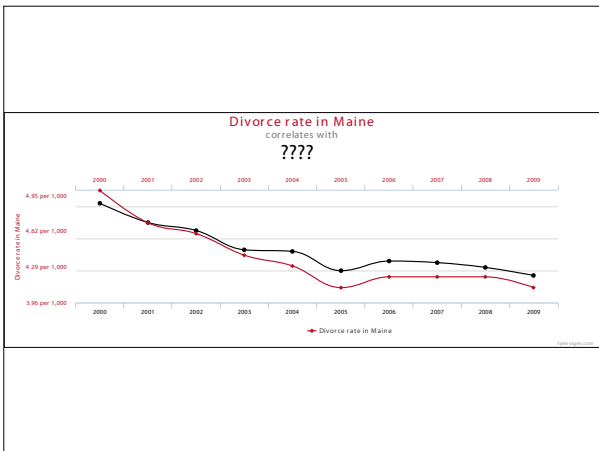
---

---

---

---

---



---

---

---

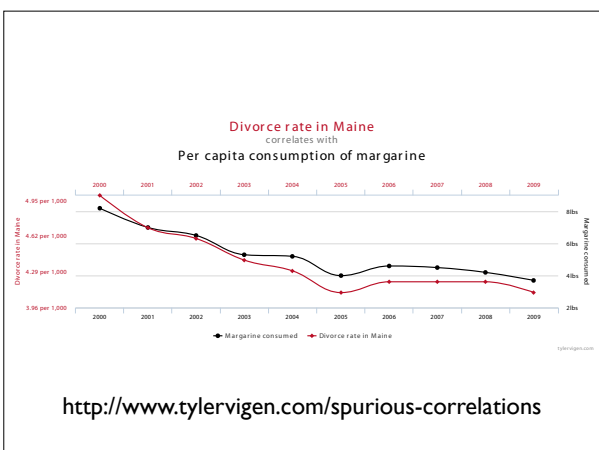
---

---

---

---

---



<http://www.tylervigen.com/spurious-correlations>

---

---

---

---

---

---

---

---



If you do enough tests, you can  
“discover” almost anything!

---

---

---

---

---

---

---

---

General Article



**False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant**

Psychological Science  
22(11) 1339–1346  
© The Author(s) 2011  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0956797611417832  
http://ps.sagepub.com  
SAGE

Joseph P. Simmons<sup>1</sup>, Leif D. Nelson<sup>2</sup>, and Uri Simonsohn<sup>1</sup>  
<sup>1</sup>The Wharton School, University of Pennsylvania, and <sup>2</sup>Ross School of Business, University of California, Berkeley

**Abstract**

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings ( $\leq .05$ ), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Этот текст является переводом оригинала, опубликованного в журнале Psychological Science. Оригинальный текст доступен по адресу: <http://ps.sagepub.com>.  
Этот текст является переводом оригинала, опубликованного в журнале Psychological Science. Оригинальный текст доступен по адресу: <http://ps.sagepub.com>.

---

---

---

---

---

---

---

---

Does listening to “When I’m 64”  
make you younger?

Yes! Subjects who listened to “When I’m 64” gave  
their age as 20 years old; subjects who listened to  
“Kalimba” gave their age as 22 years old

But wait...

---

---

---

---

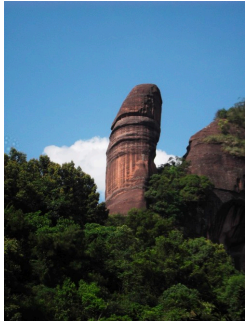
---

---

---

---





---

---

---

---

---

---

---

---

“women showed psi effects to highly  
arousing stimuli but men did not”

women, men

arousing, not arousing

positive, negative

4, 6, 8, or 10 exposures

---

---

---

---

---

---

---

---

“women showed psi effects to highly  
arousing stimuli but men did not”

3  
x  
3  
x  
3  
x  
5+



women, men

arousing, not arousing

positive, negative

4, 6, 8, or 10 exposures

at least 135  
possible tests!

---

---

---

---

---

---

---

---

“women showed psi effects to highly  
arousing stimuli but men did not”

3  
x  
3  
x  
3  
x  
5+



women, men  
arousing, not arousing  
positive, negative  
4, 6, 8, or 10 exposures

one was **bound** to  
yield a rare result

---

---

---

---

---

---

---

---

## “Researcher degrees of freedom”

undisclosed **flexibility** in experimental design  
and analysis that can lead to spurious “effects”

not always apparent; be on the lookout!

Take “research methods in experimental  
psychology” with Prof. Jeff Bowen!

---

---

---

---

---

---

---

---