# Media, Sports, Gambling, and Life through Statistics

Matt Smerglia, Stephen Wu

# Introduction

Goals:

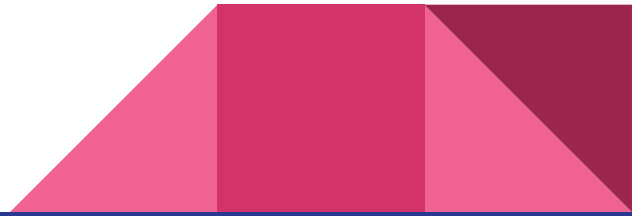
- Interest people in the field of statistics and its applications
- Make people more cognizant of poor visualizations and statistical confidence
- Discuss interesting statistics applications in real-life

Matt Smerglia

- STEM EE 2nd Year, Statistics for Industry and Sport Industry (PSP)

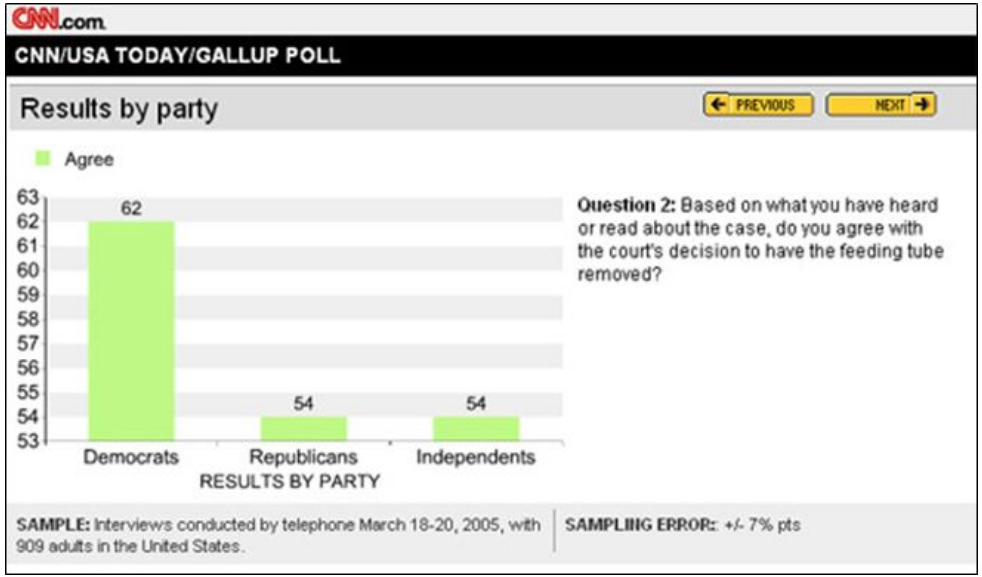
Stephen Wu

- STEM EE 1st Year, Computer Science Engineering



# Statistics in the Media

CNN posting Gallup poll about Terry Schiavo case

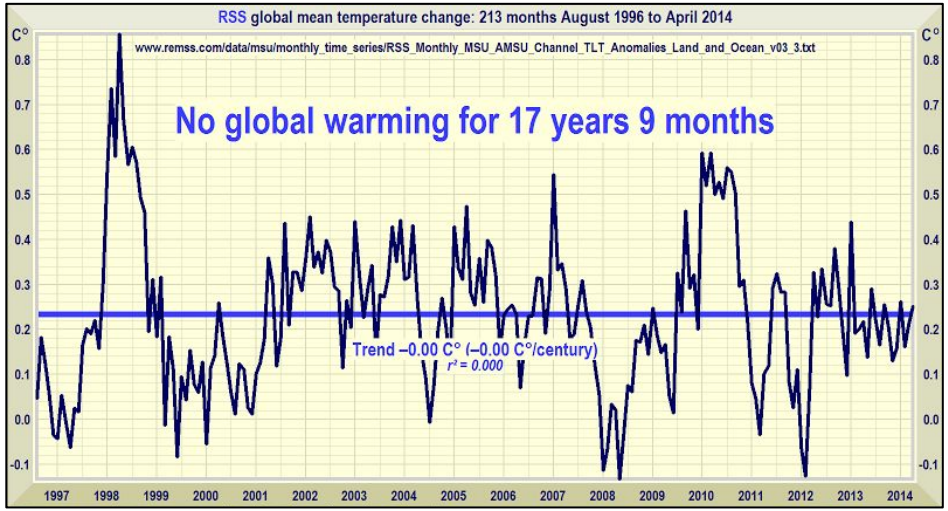




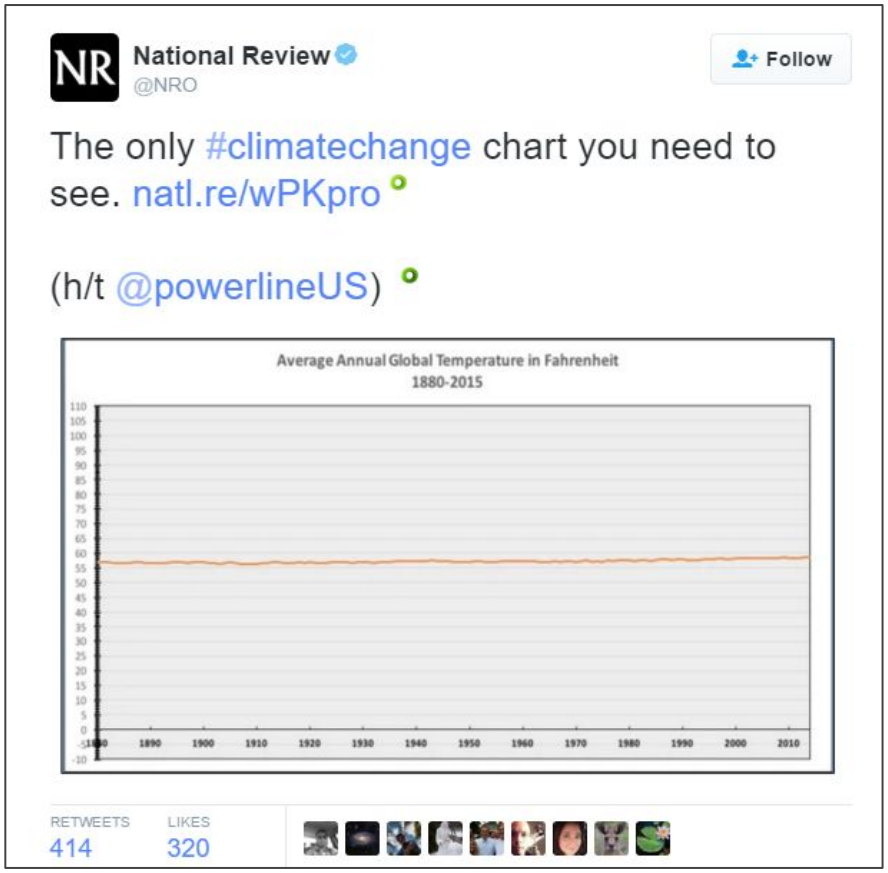
Fox News graphic comparing quantity of people on welfare to people with a full time job

What features of these graphs are potentially misleading?

# What global warming?



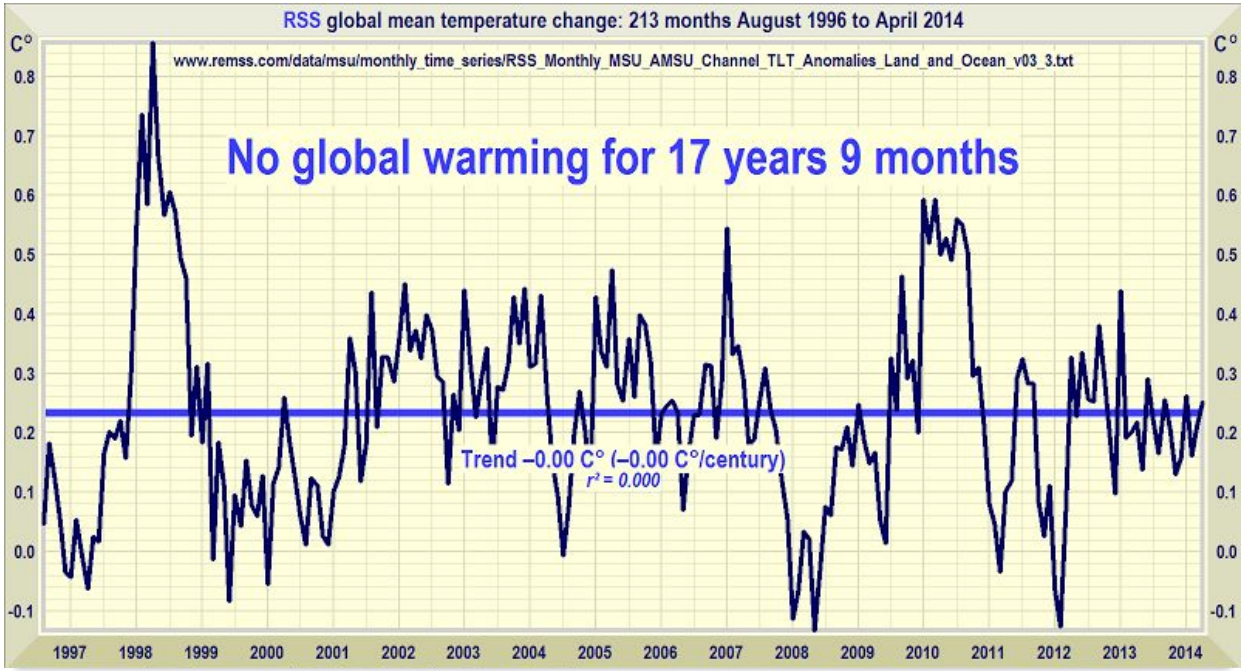

Satellite data from 17 years, RSS model, from Climate Depot




National Review tweet about climate change

What features of these graphs are potentially misleading?

# What global warming?




Satellite data, RSS, IPCC models, from *Climate Depot*

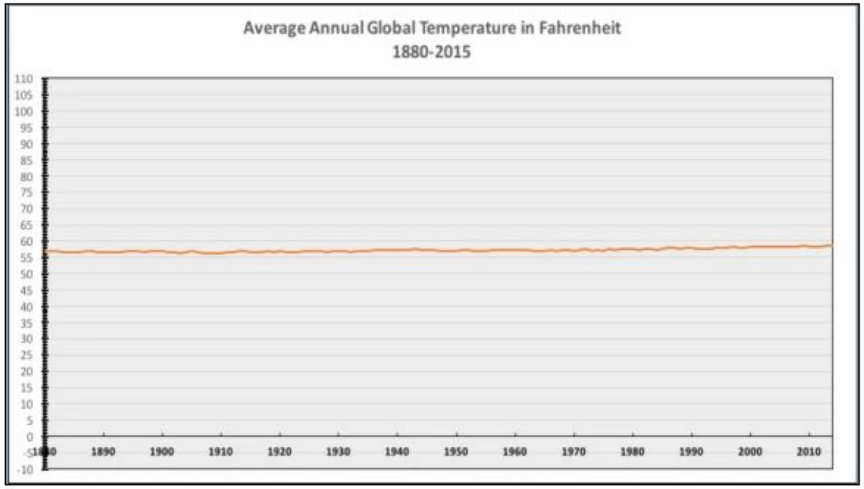
"Many of the alarmists on global warming, they've got a problem because the science doesn't back them up. In particular, satellite data demonstrate for the last 17 years, there's been zero warming."

— Ted Cruz, 2016 presidential candidate

Between 1997 and 1998, there was an unusually strong El Niño, and this caused 1998 to be one of the hottest years on record...When the 1998 measurement is included in the data, it looks as if there is no overall warming between 1998 and 2008 at all...Claims that global warming is not occurring that are derived from a cooling observed over short time periods ignore natural variability and are misleading."

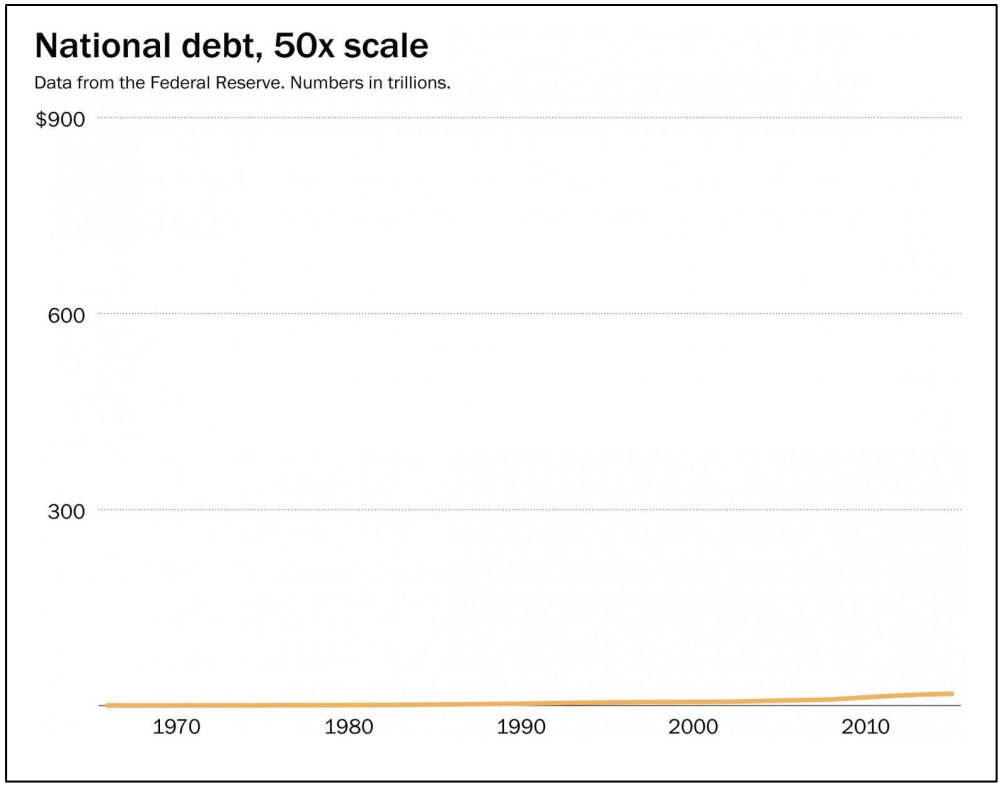
— Josh Willis, NASA Climate Scientist

# What national debt?




*National Review* graph about climate change

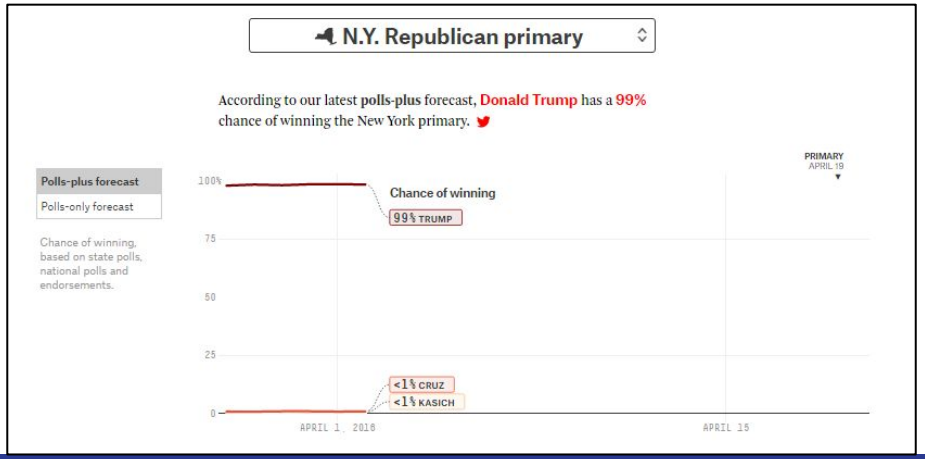
"The only #climatechange chart you need to see."
— *National Review*, very popular conservative news magazine

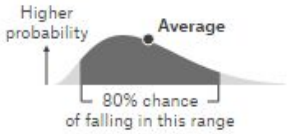



National debt over 40 years from Washington Post

# Political polling?

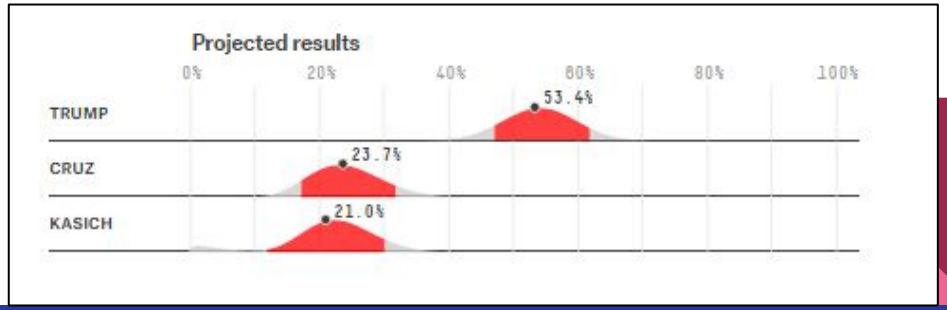
- Not all polls are created equal...Lots of room for bias.
- *FiveThirtyEight* creates a model based on congregating various polls and ranking them, but there's still room for error.




| POLLSTER | CALLS CELLPHONES | 538 GRADE | MEAN-REVERTED BIAS |
|---|---|---|---|
| SurveyUSA | ● | A | R+0.2 |
| Rasmussen Reports/ Pulse Opinion Research | | C | R+2.3 |
| Mason-Dixon Polling & Research, Inc. | ● | B+ | R+1.1 |
| Zogby Analytics, telephone | | C | R+0.5 |
| Public Policy Polling | | B- | R+0.7 |
| Research 2000 | | F | D+1.4 |
| YouGov | | C+ | R+1.2 |
| American Research Group | ● | C- | R+0.1 |
| Quinnipiac University | ● | B+ | R+0.9 |
| TCJ Research | | F | R+4.5 |
| Strategic Vision, LLC | | F | R+1.9 |
| Harris Interactive, online | | D+ | R+1.7 |

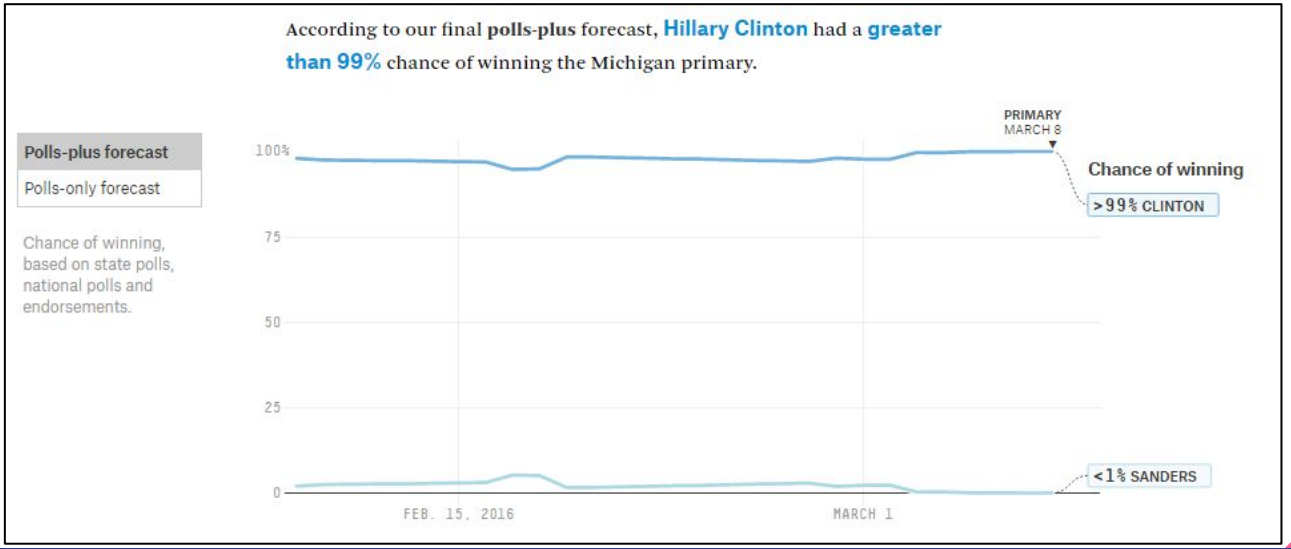
FiveThirtyEight primary polling



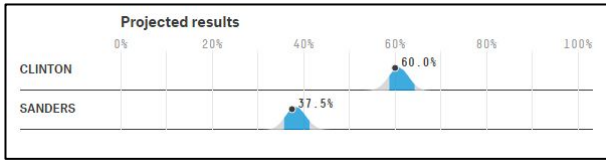
# Michigan Democratic Primary

- Pretty much everyone (CNN, Fox, MSNBC, 538) reported a *very, very* high chance of Hillary winning Michigan.

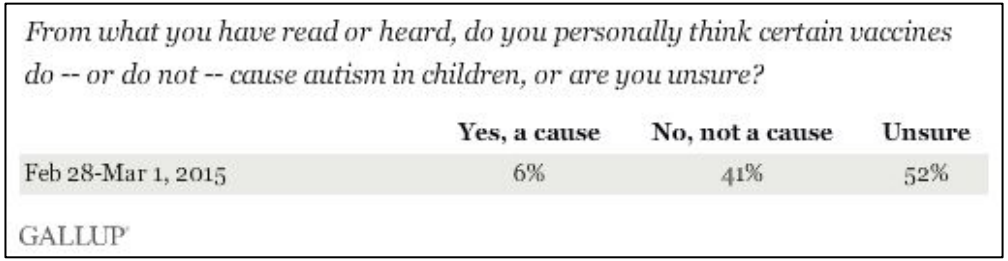
  She lost by 1.5 points to Sanders.




FiveThirtyEight primary polling



# Invalid science



*From what you have read or heard, do you personally think certain vaccines do -- or do not -- cause autism in children, or are you unsure?*

| | Yes, a cause | No, not a cause | Unsure |
|---|---|---|---|
| Feb 28-Mar 1, 2015 | 6% | 41% | 52% |

GALLUP

GALLUP polling in 2015

*1998*: Highly reputed science journal publishes study:
**Measles-mumps rubella (MMR) vaccine likely cause of autism**

Article has since been retracted, but damage has been done.

Retractions in scientific articles has gone up 10x in past decade.
"Most published research findings are false" — Stanford epidemiologist John Ioannidis

**Causes**? Competition, quality over quantity, publication bias, poor peer review, and **bad statistics**.

*Unreliable research: trouble at the lab*, The Economist
*Scientists Are Wrong All the Time, and That's Fantastic*, Marcos Wod, WIRED

# Statistical significance, p-values, and power

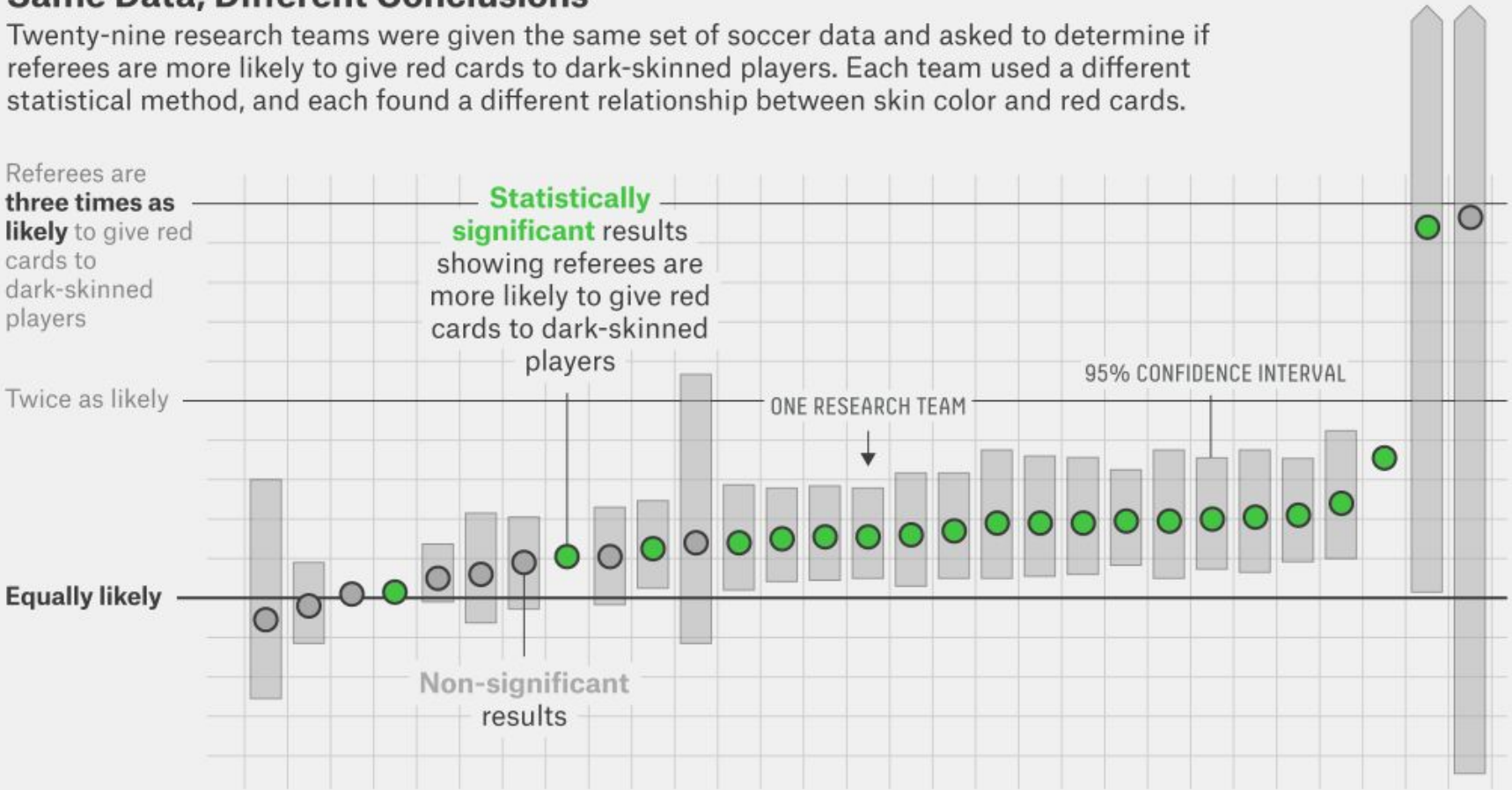
Science often accepts a result of “statistical significance” if its p-value is **5%**.

Media reports journal articles displaying **positive correlations**, but not **negative correlations.**

For the one MMR article that said vaccines caused autism, there were 20+ that didn’t...But which one was reported and widely discussed?

## Same Data, Different Conclusions

Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.

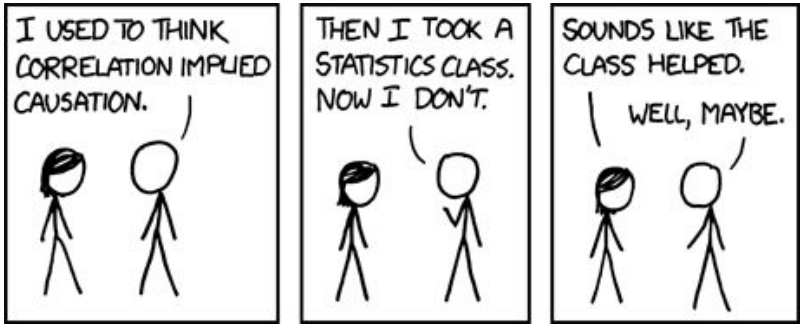



FIVETHIRTYEIGHT

SOURCE: BRIAN NOSEK ET AL.

"Failure is Moving Science Forward", *FiveThirtyEight*

# Correlation, Causation



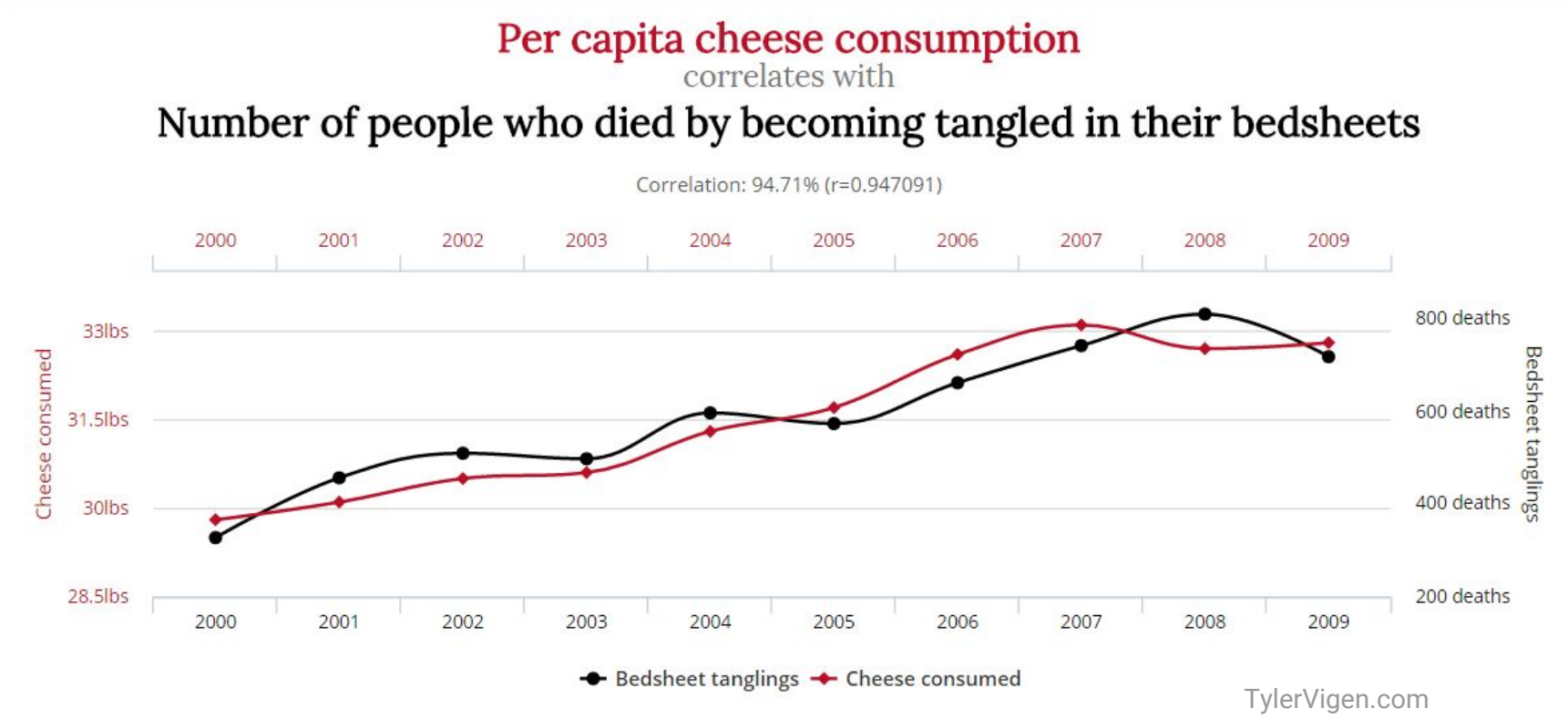

xkcd

- "Correlation does not imply causation."
  - Heard often, but practically everyone makes this mistake…
  - Correlation is *evidence* for causation, but not *proof*

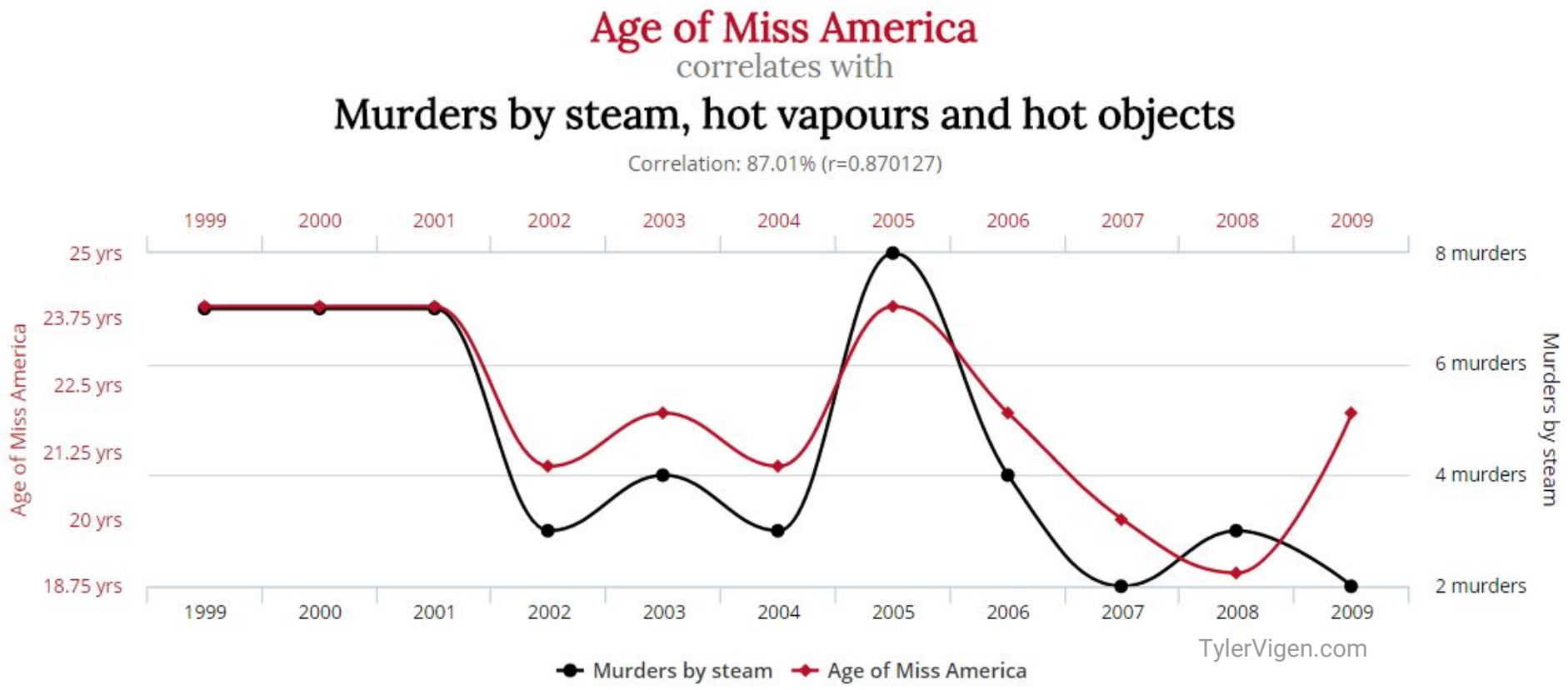
Based on statistics, we often cannot make assumptions as to whether event A causes event B. Social data is almost always correlative.

- "Bernie didn't win these states because of the lower African-American population." - npr
- "We encourage students to participate in activities because students who participate in at least one activity raise their GPA!" - OSU tour guide

# Spurious Correlations



# Spurious Correlations, continued



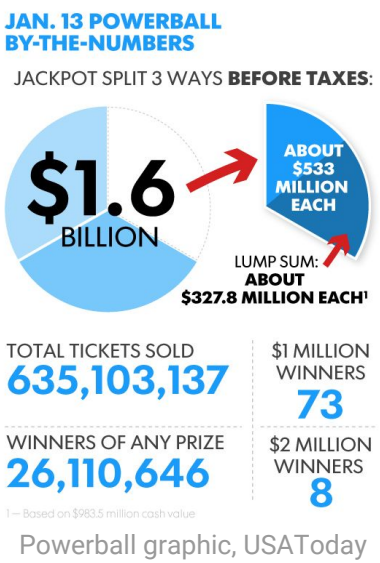
# January Powerball

- Recent powerball drawing: $1.59billion after 20 rollovers
- 1 in 292million odds — $2 / $3 tickets
- Would you gamble? What are the conditions that make the powerball worth it?

# Los Angeles Times Powerball Simulator

# Powerball by the Numbers

- Lump sum: $930 million —> **~$533 million** after taxes
  Annuity: **~$2.958m / month**



Powerball graphic, USAToday

- Bigger pot -> More players -> More winners -> Less expected value…
- Needed **$584 million** to buy every one of the **292 million combinations**

- Nearly 70% of lottery winners go broke in 7 years.
- Many have tragic stories — Jack Whittaker

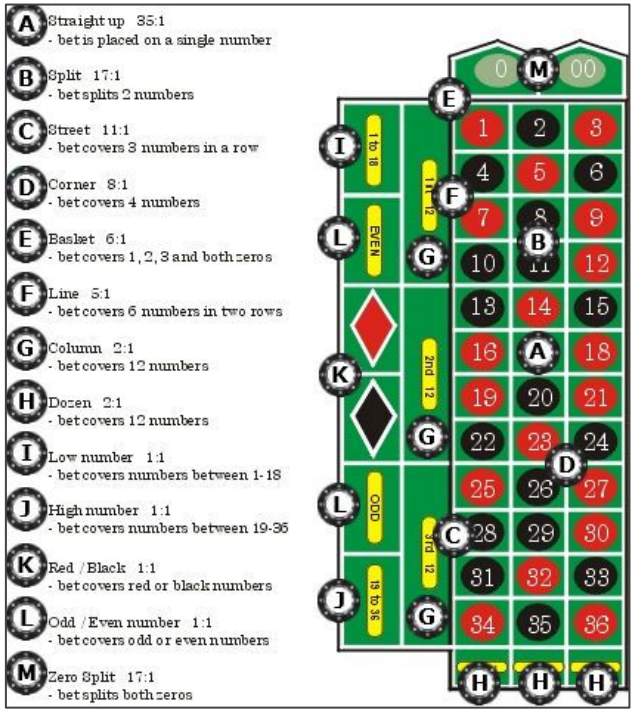
# Roulette & The Martingale

- Roulette, Craps, Slot Machines earn money by a built-in house edge.

- The Martingale system attempts beat the house.

Example:

1. Probability of hitting red or black: (18/38) ≈ 47%
2. Bet red/black. On loss, double down.
3. The first win recovers all previous losses.




Standard roulette odds from LiveDealer

# Martingale, continued

In action:
I have $1,000. I start with a $10 bet on red.

| Play | Bet | Result | Funds |
|---|---|---|---|
| 1 | $10 | LOSS, -$10 | $1000 - 10 = $990 |
| 2 | $20 | LOSS, -$20 | $990 - 20 = $970 |
| 3 | $40 | LOSS, -$40 | $970 - 40 = $930 |
| 4 | $80 | LOSS, -$80 | $930 - 80 = $850 |
| 5 | $160 | WIN, +$160 | $850 + 160 = $1010 |

On average, I should win almost every other play.
I keep betting red or black; it doesn't matter which.

In the case that I keep losing, I make it all back when I finally win, and I'm up my $10 bet.

# Martingale, continued

Here's what can happen in practice:

| Play | Bet | Result | Funds |
|---|---|---|---|
| 55 | $320 | LOSS #6, -$320 | $10500 - 320 = $10180 |
| 56 | $640 | LOSS #7, -$640 | $10180 - 640 = $9540 |
| 57 | $1280 | LOSS #8, -$1280 | $9540 - 1280 = $8260 |
| 58 | $2560 | LOSS #9, -$2560 | $8260 - 5700 = $5700 |
| 59 | $5120 | LOSS #10, -$5120 | $5700 - 5120 = $580 |

With $6,500, and a $25 starting bet:
you can handle ~8 losses in a row

$25(1 + 2^1 + 2^2 + 2^3 + \ldots + 2^7) = \$6{,}375$

Probability: $(1-.4737)^8 = 0.59\%$
(Of happening in 8 spins)

Let's test this model.

# WAR! What is it good for?

- W.A.R. is a popular sabermetric that has quickly become one of the most important statistics in baseball
- Stands for **Wins Above Replacement** - Represents the # of wins a player added to a team above what a replacement player would add
- How do we see WAR becoming more prevalent? Let's examine recent MVP and Cy Young Award Winners:

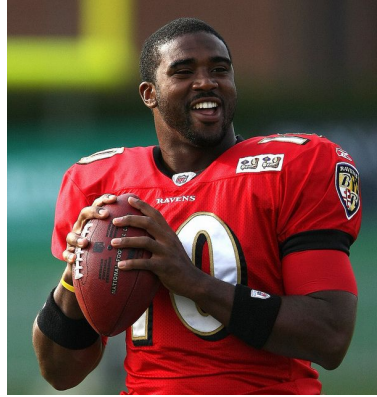
WAR Table from the Bleacher Report

| CALIBER OF PLAYER | WINS ABOVE REPLACEMENT |
|---|---|
| BENCH GUY | 0-1 WAR |
| ROLE PLAYER | 1-2 WAR |
| SOLID STARTER | 2-3 WAR |
| ABOVE-AVERAGE | 3-4 WAR |
| ALL-STAR | 4-5 WAR |
| SUPERSTAR | 5-6 WAR |
| MVP | 6+ WAR |

# QBR

- NFL uses newer sabermetric Total QBR - Quarterback Rating
- QBR is a number from 1-100 that gives a QB an overall rating of his performance by factoring in situational play  (1st Q v 4th Q, 3rd down, etc)
    - Much nicer than standard QB Rating! (Perfect QB rating is 158.3)

- Let’s see if QBR translates to MVPs:



Name the QB?

*photo credit: Keith Allison*

# PER

- The NBA also has a useful sabermetric to determine how efficient a player is, dubbed as PER
- Stands for **Player Efficiency Rating**
  - A per-minute rating developed by ESPN.com columnist John Hollinger.
  - "The PER sums up all a player's positive accomplishments, subtracts the nega accomplishments, and returns a per-minute rating of a player's performance.
- Once again, let's see if a high PER translates to MVPs:

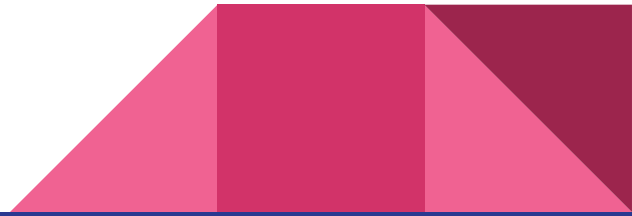
$$uPER = \frac{1}{min} \times \left(3P - \frac{PF \times lgFT}{lgPF} + \left[\frac{FT}{2} \times \left(2 - \frac{tmAST}{3 \times tmFG}\right)\right] + \left[FG \times \left(2 - \frac{factor \times tmAST}{tmFG}\right)\right] + \frac{2 \times AST}{3} + VOP \times \left[DRBP \times (2 \times ORB + BLK - 0.2464 \times [FTA - FT] - [FGA - FG] - TRB) + \frac{0.44 \times lgFTA \times PF}{lgPF} - (TO + ORB) + STL + TRB - 0.1936\,(FTA - FT)\right]\right)$$

$$PER = \left(uPER \times \frac{lgPace}{tmPace}\right) \times \frac{15}{lguPER}$$

Equations from *PER*, Wikipedia

# Resources

- ★ FiveThirtyEight.com
  - Nate Silver's FiveThirtyEight uses statistical analysis — hard numbers — to tell compelling stories about politics, sports, science, economics and lifestyle.
- ★ InformationIsBeautiful.net
  - Ideas, issues, knowledge, data — visualized!
- ★ Stats.org
  - Blog and newsletter about stats everyday!



# Questions?

Thanks for listening!

Fun STEM fact: The Scholars banquet is on 4/24 and is business casual!

And sign up for rock climbing!