DATA SCIENCE, STATISTICS, AND PUBLIC ADMINISTRATION

PMAP 8521: Program Evaluation for Public Service

August 26, 2018

No iCollege quiz today!

PLAN FOR TODAY

Who am I?

Data, data science, & public service

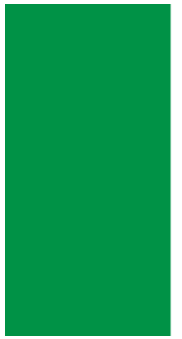
Evidence, causation, & evaluation

Class details

Playing with R

WHO AM I?

BRIGHAM YOUNG UNIVERSITY
FOUNDED
BYU
1875
PROVO, UTAH



§11 Now that we know the quiddity of truth through our knowledge of its definitions, let us learn the quality of truth, or how truth exists, and how it occurs and comes to be. We say that truth is a proposition with which the speaker's soul is tranquil,[26] knowing that what he said is indeed <as he said it> and that what he described is indeed as he described it. We also say that the truth is a proposition that the tongue articulates, whereby the soul informs another soul regarding something of which [the latter] is ignorant or of which it is content in knowing.[27] We also say that the truth is various modes that recount the form of beings and the existence of the intelligibles. We also say that truth is the ultimate verbal achievement of the soul // in its inquiry into whatever it did not know, as demonstrated in the following schema:

§12 ***The quality of truth is one of four:***

The first: Truth is a proposition with which the speaker's soul is content, knowing that what he said is indeed as he said it and what he described is indeed how he described it. The second: Truth is a proposition that the tongue articulates, whereby the soul informs another soul regarding something of which it is ignorant or which it is content in knowing. The third: Truth recounts the form of beings and the existence of intelligibles. The fourth: Truth is the ultimate verbal achievement of the soul in its inquiry into whatever it did not know.

26. *Sukūn al-nafs*, meaning the state of "rest" that is a result of unwavering certainty (as opposed to the frantic searching "movement" that accompanies uncertainty), and hence, "tranquility" of the mind. Below (*IM* 2.11) al-Muqammaṣ contrasts this content state with ignorance rather than with falsehood. For many [illegible] instance, Abū ʿAlī al-Jubbāʾī, *sukūn al-nafs* describes knowl- [illegible] h; see ʿAbd al-Jabbār, *Mughnī*, 12:13. Al-Muqammaṣ's use of [illegible] to the antiquity of the term, and there is, therefore, no [illegible] d al-Jabbār of forcing it on Abū ʿAlī, as suggested by Vajda, [illegible] " 151 and n. 10. See also Vajda, "Autour de la théorie," 139; [illegible] ī, 12:43, lines 9–10; and compare al-Qirqisānī in Hirschfeld, [illegible] lines 8–13. On this concept, which has Stoic origins, see [illegible] /peech, 43 and n. 21, as well as 47–49; and Vajda "Saadya,"

[illegible] ted here as "informing," is the silencing of the opponent [illegible] logical dead end (see Pines, "A Note on an Early Use of the [illegible] and n. 77; see also al-Qirqisānī, *Kitāb al-anwār*, 484–86; and

(١١) فإذ عرفنا ماهية الحقّ بمعرفتنا[43] حدوده،[44] فلنعرف الآن كيفية الحقّ، أو كيف يكون الحقّ ويقع ويجب. ونقول إنّ الحقّ مقالٌ تسكن إليه النّفس بعلمها[45] بأنّ ما قال <كما قال>[46] وما وصف كما وصف. ونقول أيضًا، إنّ الحقّ مقالٌ ينطق به اللسان مخبر،[47] أعني النّفس لانقطاع نفس أخرى في ما جهلت أو سكنت.[48] ونقول أيضًا إنّ الحقّ ضروب تحكي[49] صورة الموجودات ووجود المعقولات، ونقول أيضًا إنّ
٧أ الحقّ غاية ما إليه تصير النّفس من // [القول عن فحصها ع]ن ما كانت به جاهلة، وذلك على هذا [المثال]:

(١٢) *** كيفية الحقّ على أربعة أوجه ***

[أ]حدها، أنّ الحقّ مقال تسكن إليه النّفس بعلمها بأنّ ما[50] قال كما قال وما وصف كما وصف. والثاني، أنّ الحقّ مقال أن[51] ينطق به اللسان مخبر، أعني النّفس لانقطاع نفس أخرى في ما جهلت أو سكنت.[52] والثالث، أنّ الحقّ يحكي صورة الموجودات ووجود المعقولات. والرابع، أنّ الحقّ غاية ما إليه تصير النّفس من القول عن فحصها عن ما كانت به جاهلة.

٤٣. ف: بمعرفنا ٤٤. ف: حدورها ٤٥. ف: بعدمها ٤٦. للمقايسة، يُنظر لاحقًا الفقرة ١٢. ٤٧. ف: موجود (للمقايسة، يُنظر لاحقًا الفقرة ١٢). ٤٨. ربما يجدر التصويب: سكنت. ٤٩. ف: تحمي ٥٠. ف: بانما (للمقايسة، يُنظر، بلاو، النحو، المادة ٤٦) ٥١. يُنظر، بلاو، النحو، المادة ٣٥٩. ٥٢. ربما يجدر التصويب: سكنت.

al-Maqdisī, *Kitāb al-badʾ*, 1:51, line 11ff.). It seems, therefore, that this definition is intended to explain in what sense one can speak of truth in a disputation. The last part of the definition—which includes *juhūl* and *sukūn*—is somewhat awkward. With a minor correction (see n. 48 to the text) this sentence would read ". . . in order to help another soul out of a state of being either ignorant or silent (*sakatat*) about something." On "silence" in disputation, see, for instance, al-Qirqisānī, *Kitāb al-anwār*, 485, line 19.

NEAL A. MAXWELL INSTITUTE *for* RELIGIOUS SCHOLARSHIP

ALL HUMANITIES UNTIL...

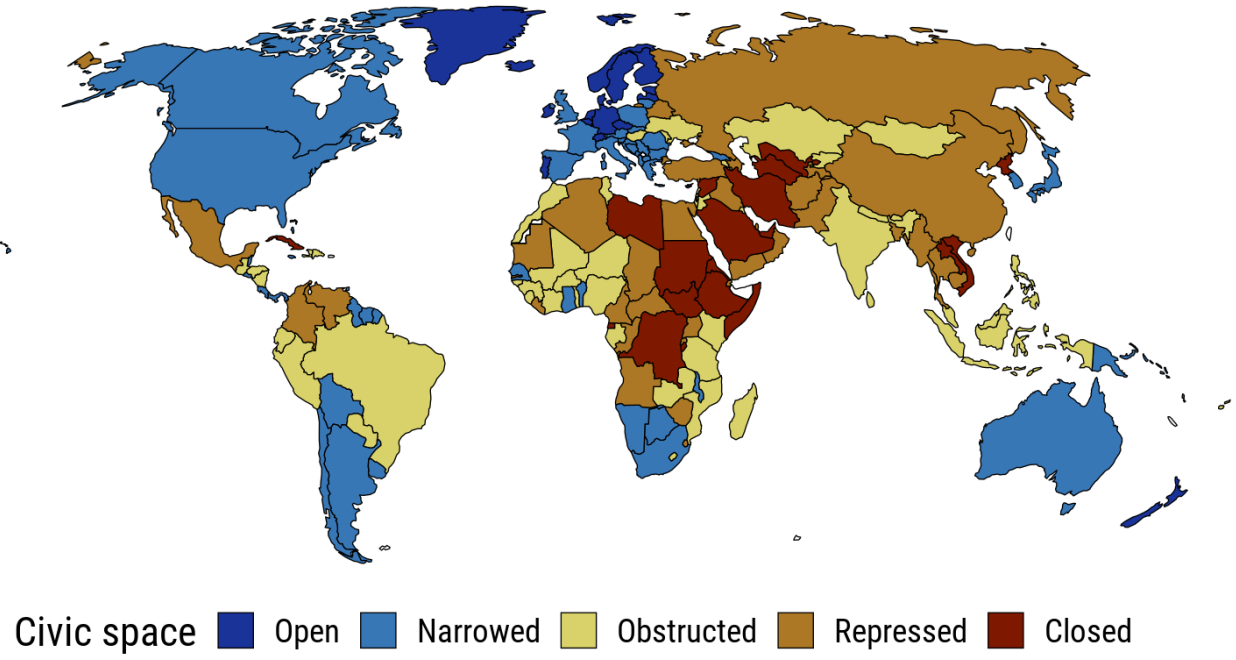
romney institute
BRIGHAM YOUNG UNIVERSITY

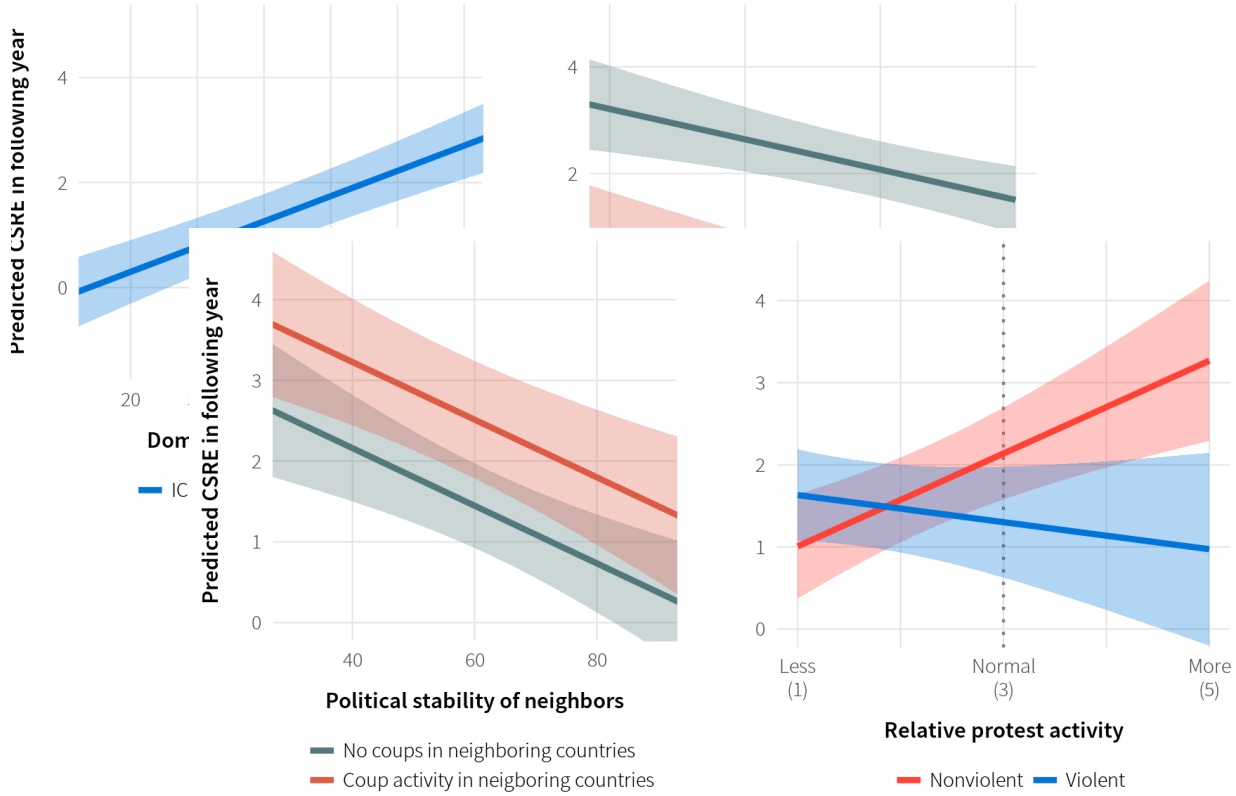
Economics

Statistics

180°

Duke | SANFORD SCHOOL OF PUBLIC POLICY

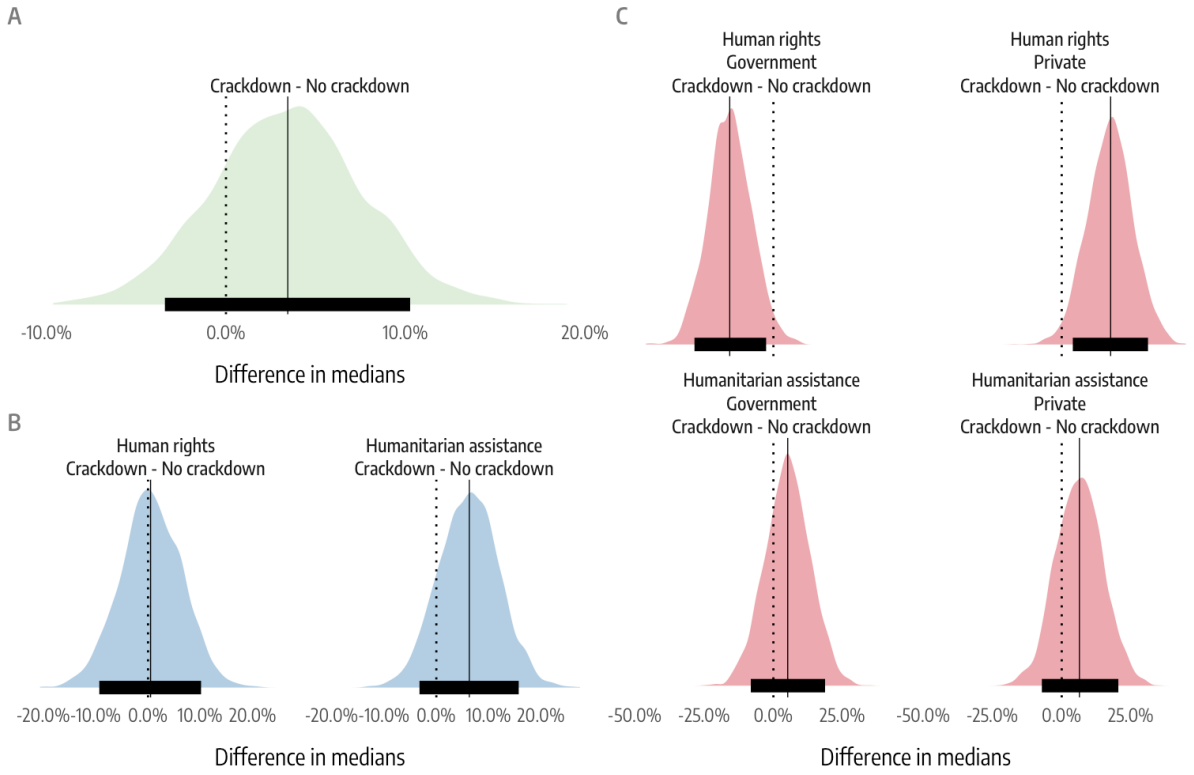


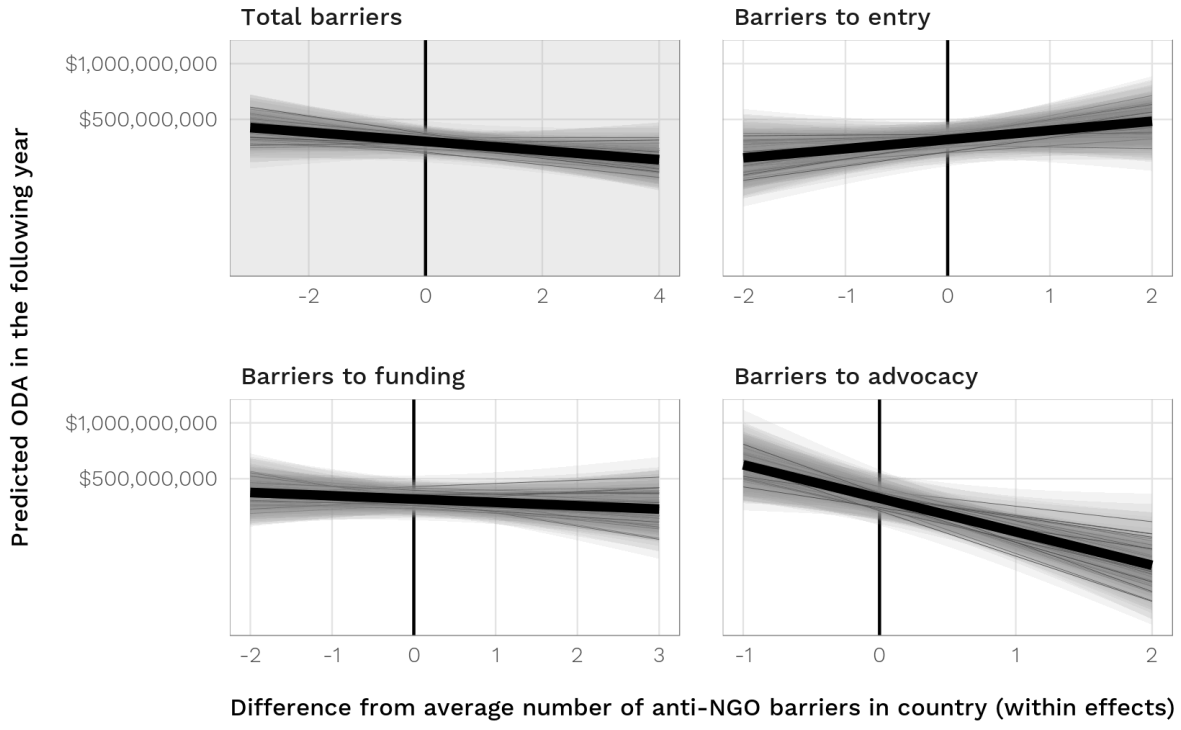


NGOS AND REPRESSION

Differences in donation likelihood in control and crackdown groups, conditioned by other experimental groups

90% credible intervals shown in black. Solid vertical line = median; dotted vertical line = 0





LOLZ

romney institute
BRIGHAM YOUNG UNIVERSITY

Georgia State University

ANDREW YOUNG SCHOOL
OF POLICY STUDIES

Economics

Statistics

DATA, DATA SCIENCE, & PUBLIC SERVICE

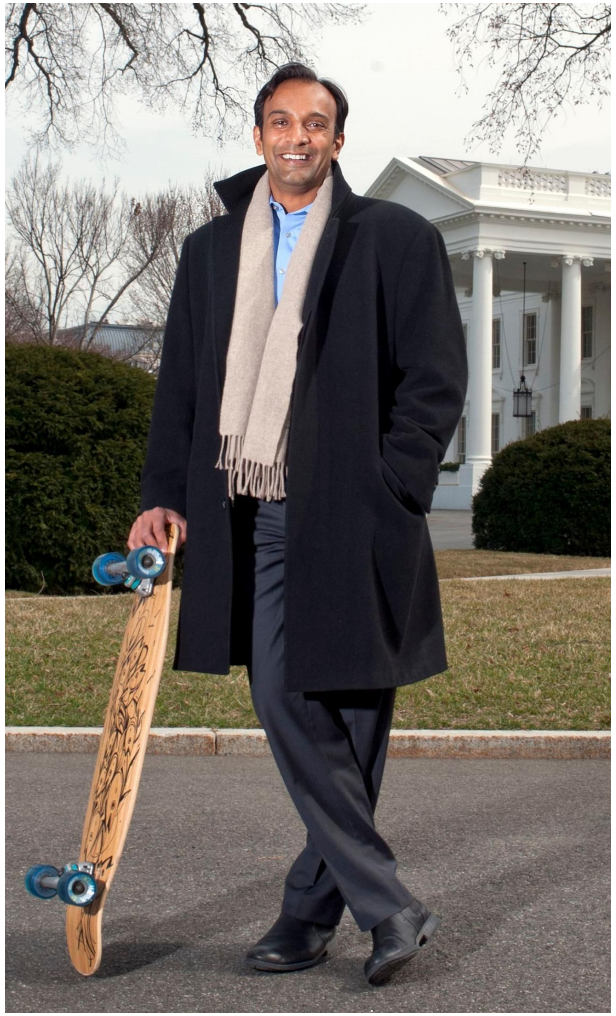
SUSAN CRAWFORD IDEAS 08.22.18 07:00 AM

WHY UNIVERSITIES NEED 'PUBLIC INTEREST TECHNOLOGY' COURSES

POLICYMAKERS AT ALL levels of government are struggling to thoughtfully harness data in the service of public values. Many public servants grew up in an era of firmly separate disciplines: You were either an engineer or an economist, either a programmer or a social worker, but never both. In an era in which data is everything, the risks to core democratic principles—equity, fairness, support for the most vulnerable, delivery of effective government services—caused by technological illiteracy in policymakers, and policy illiteracy in computer scientists, are staggering.

field aimed at addressing precisely this gap in interdisciplinary opportunities. This new area, "public interest technology," is still being defined; it encompasses designing public policy and laws with an awareness of how technology actually works, as well as ensuring that technology is being used to serve public values of fairness and equity. It means consciously thinking about the welfare of society in general, rather than the incentives of a single company.

DATA AND GOVERNMENT



"To responsibly unleash the power of data to benefit all Americans"

The White House

Office of the Press Secretary

For Immediate Release June 30, 2016

FACT SHEET: Launching the Data-Driven Justice Initiative: Disrupting the Cycle of Incarceration

"[O]ur criminal justice system isn't as smart as it should be. It's not keeping us as safe as it should be. It is not as fair as it should be.
Mass incarceration makes our country worse off, and we need to do something about it." –

President Barack Obama, July 14, 2015

M. HARRISON
NEWS WITH A TWIST

SUNLIGHT FOUNDATION

OPEN KNOWLEDGE INTERNATIONAL

U.S. CITY OPEN DATA CENSUS

POWERED BY OPEN DATA CENSUS

Atlanta, GA

0% Open

0% Score

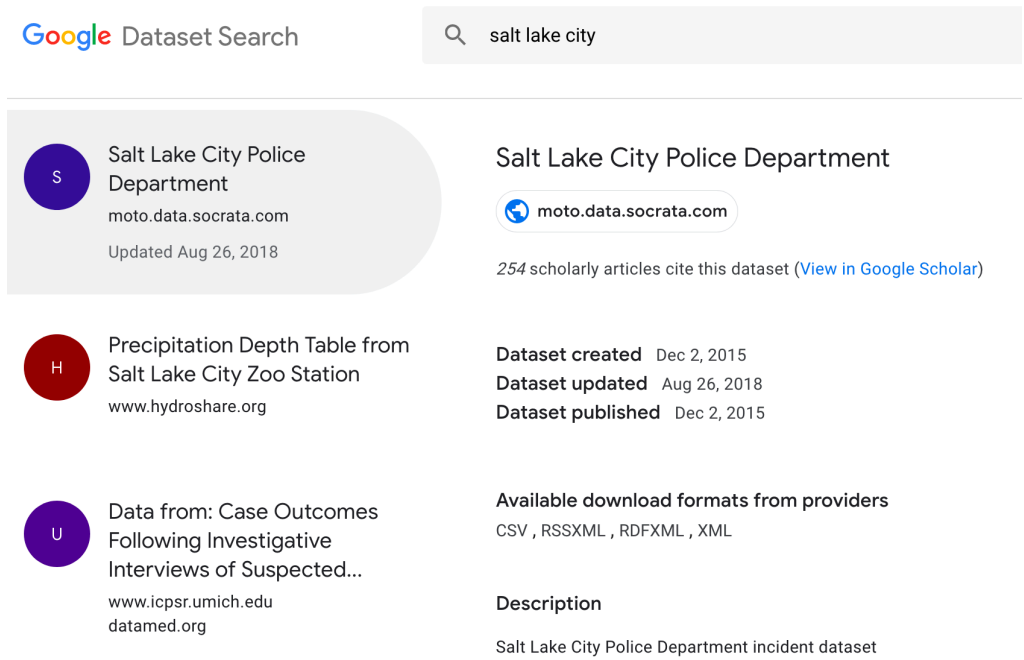
Breakdown

Dataset	Breakdown	Year	Score
Budget			
Business Listings			
Code Violations			
Construction Permits			
Crime Reports			
Emergency Calls			
Employee Salaries			
Lobbyist Activity			
Parcels			
Police Use-of-Force			
Procurement Contracts			
Property Assessment			
Property Transfers			
Public Facilities			
Restaurant Inspections			
Service Requests			
Spending			

Google Dataset Search Beta

Search for Datasets

Try boston education data or weather site:noaa.gov



How do you use all this data to make the world better?

WHAT IS "STATISTICS"?

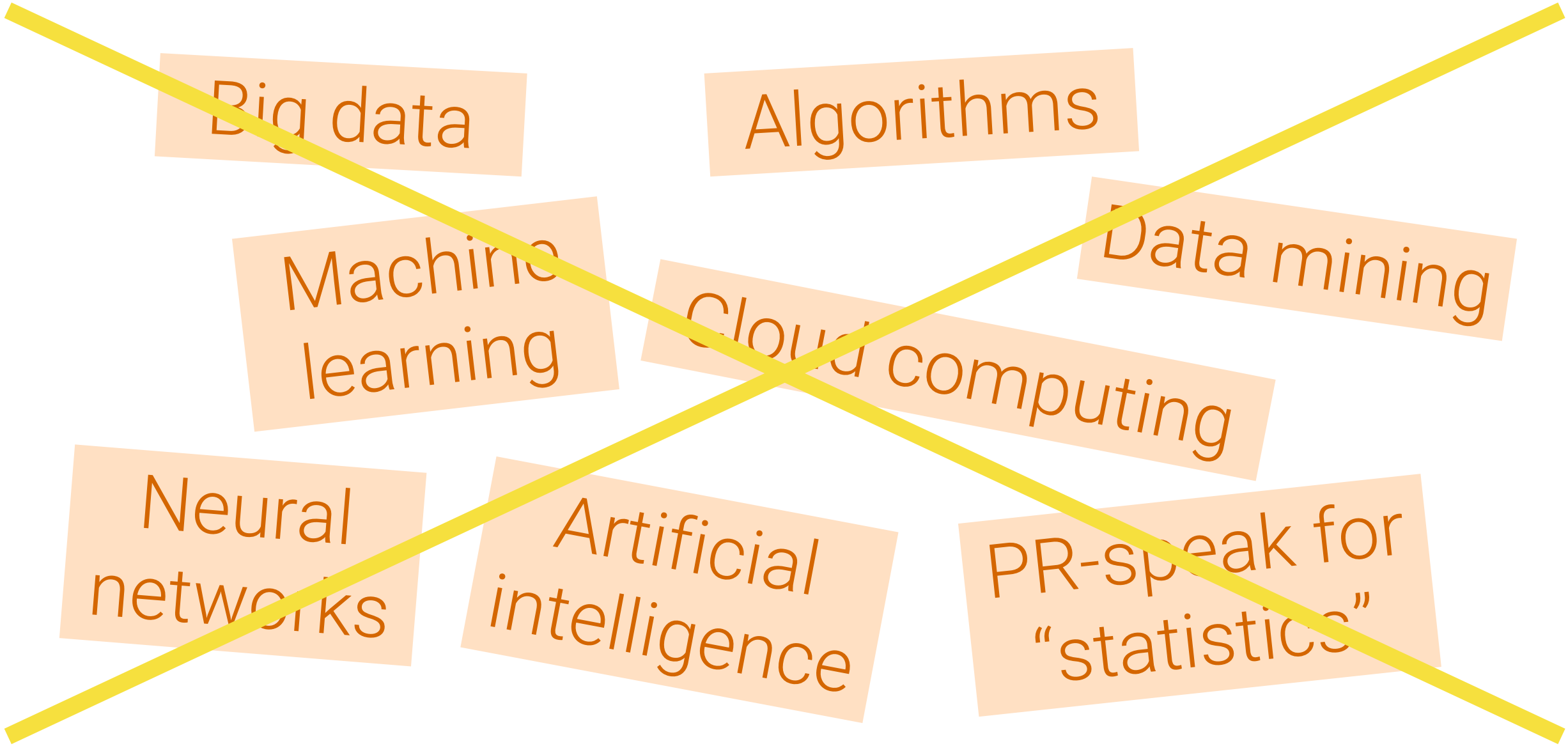
Collecting and analyzing data from a representative sample in order to make inferences about a whole population

WHAT IS "DATA SCIENCE"?

"Pile of obtuse theory about computer science"

"Scary spreadsheets and graphs"

WHAT IS "DATA SCIENCE"?

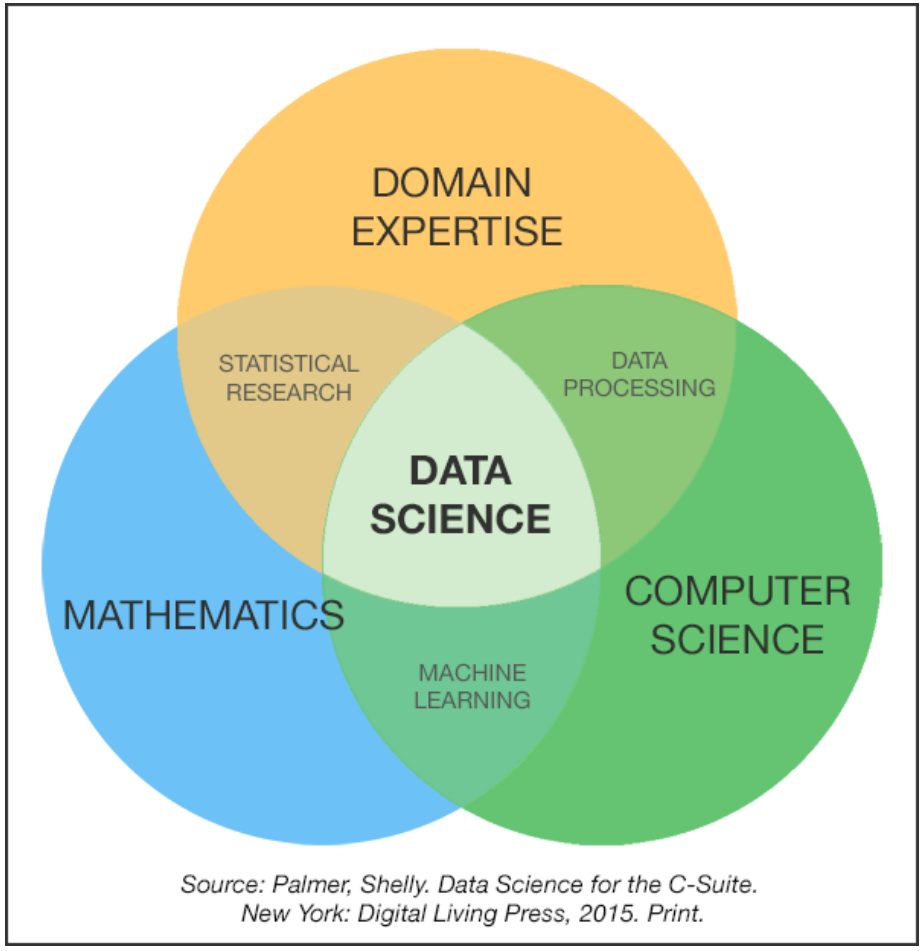


WHAT IS "DATA SCIENCE"?

Turning raw data into understanding, insight, and knowledge

Collect | **Analyze** | **Communicate**

WHAT'S THE DIFFERENCE?



Source: Palmer, Shelly. Data Science for the C-Suite. New York: Digital Living Press, 2015. Print.

Statistics

Collect

Analyze

Communicate

EVIDENCE, CAUSATION, & EVALUATION

What is the role of social scientists in the policy process?

What is the relationship between social science research and public administration?

EVIDENCE-BASED MEDICINE

MODERN EVIDENCE-BASED MEDICINE

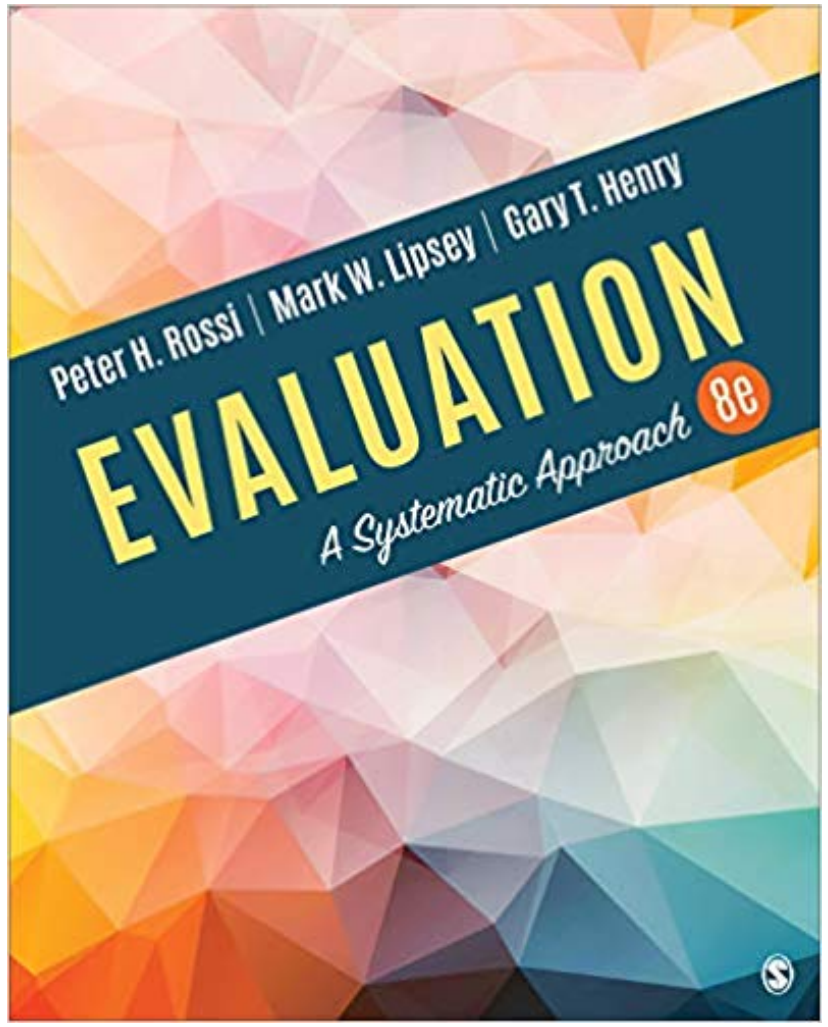
Apply evidence to clinical treatment decisions

Move away from clinical judgment and “craft knowledge”

Is this good?

How do we find and measure evidence for policies and programs?

TYPES OF EVALUATIONS



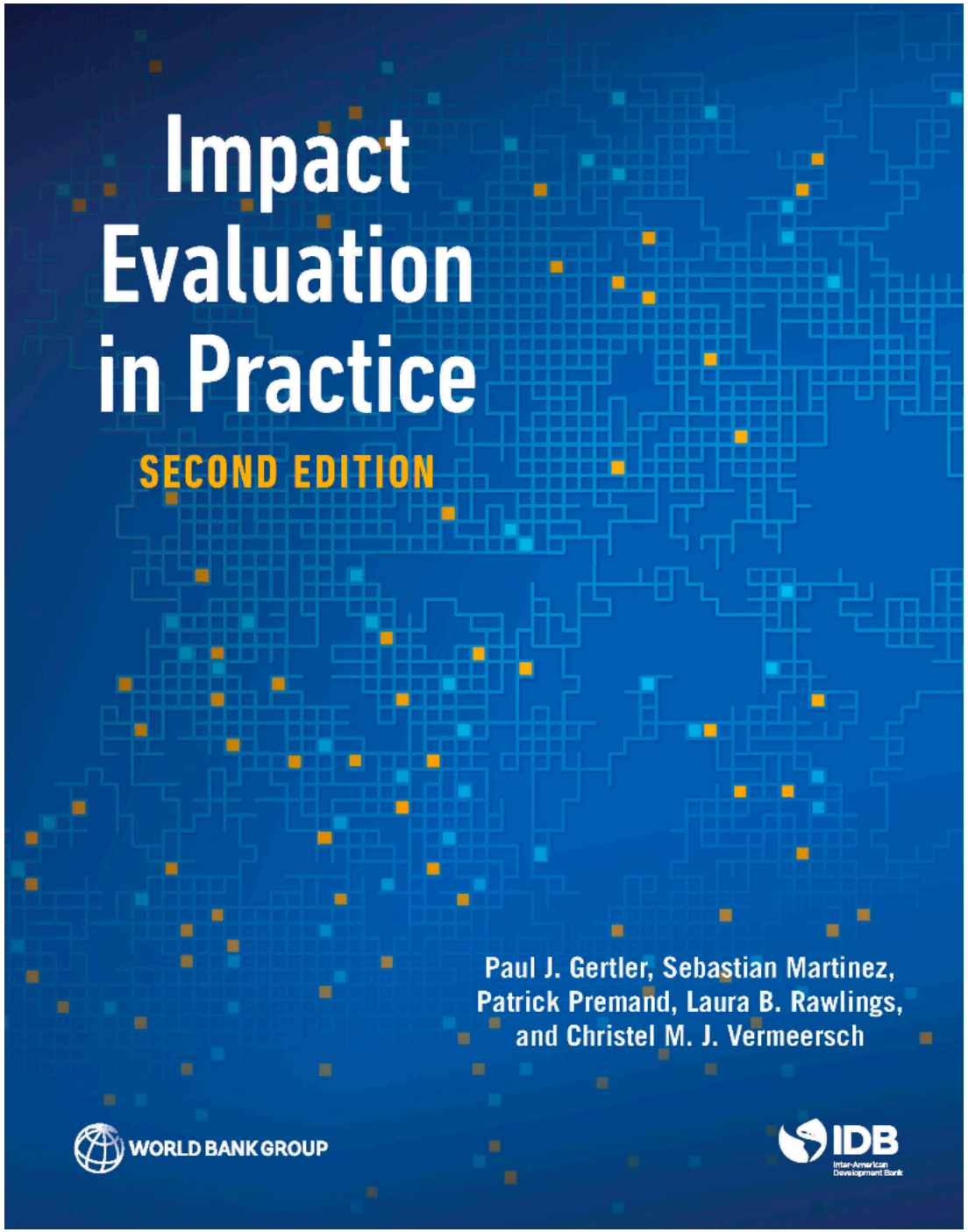
Needs assessment

Design and theory assessment

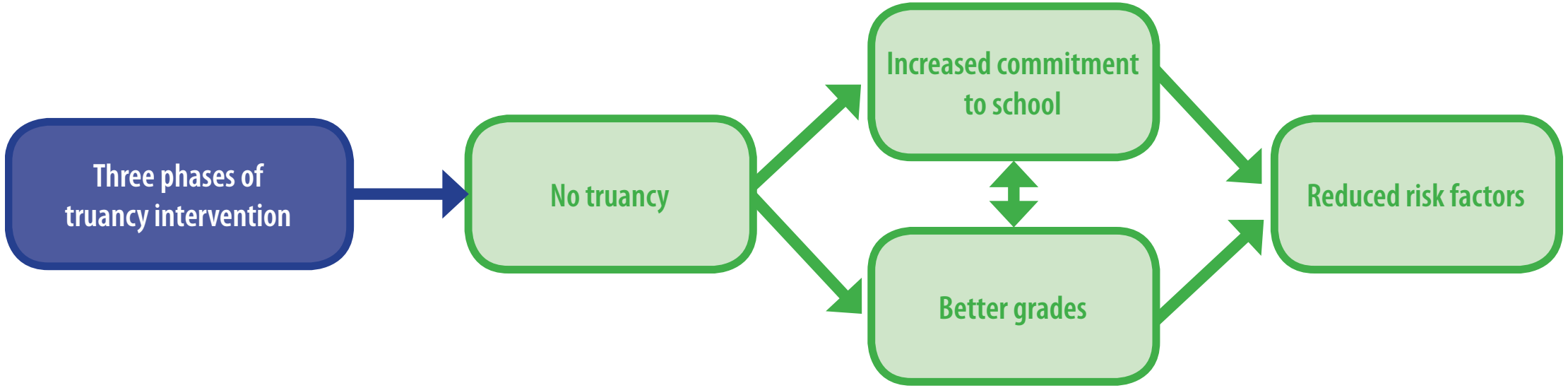
Process evaluation and monitoring

Impact evaluation

Efficiency evaluation (CBA)

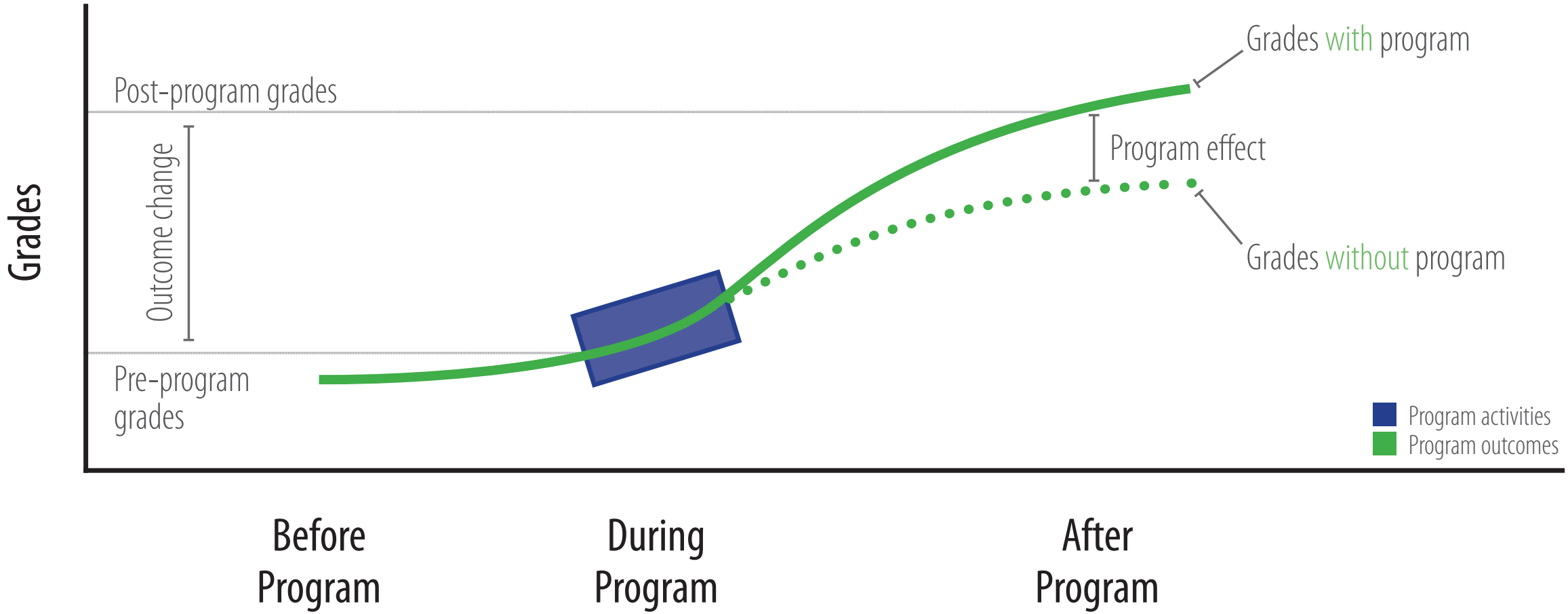

Impact Evaluation in Practice
SECOND EDITION
Paul J. Gertler, Sebastian Martinez,
Patrick Premand, Laura B. Rawlings,
and Christel M. J. Vermeersch
WORLD BANK GROUP
IDB
Inter-American Development Bank

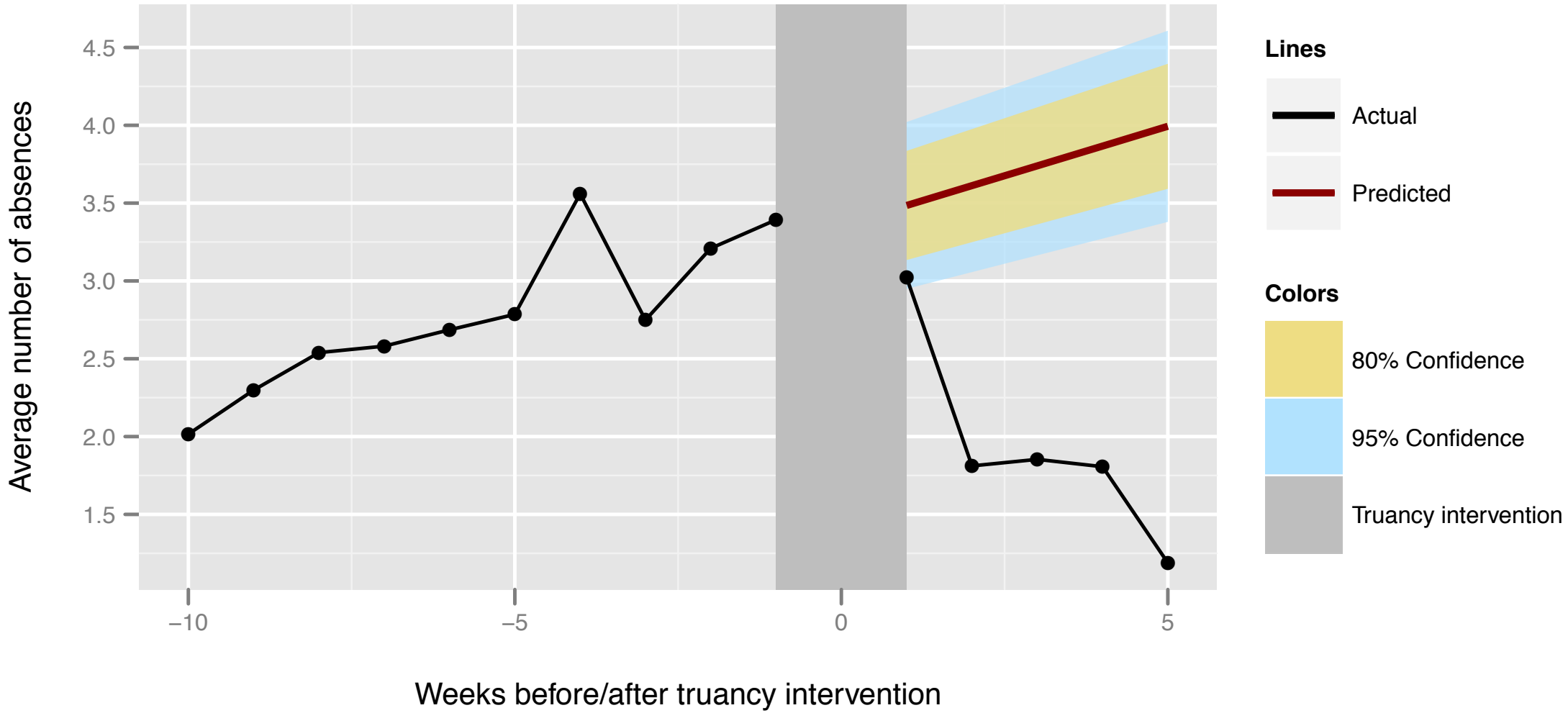
THEORIES OF CHANGE



* Because 11th and 12th graders who receive 3rd citations are generally unable to graduate from high school, district social workers no longer attempt to increase their commitment to school. As such, any outcomes that occur as a result of the alternative plans made for these students (work study programs, career development assistance, etc.) are only tangentially related to the outcomes of the truancy program itself. The system for creating alternative plans is an entirely separate program with its own logic model, goals, and outcomes.

THEORY → IMPACT




Average number of absences
4.5
4.0
3.5
3.0
2.5
2.0
1.5
−10
−5
0
5
Weeks before/after truancy intervention
Lines
Actual
Predicted
Colors
80% Confidence
95% Confidence
Truancy intervention

EVIDENCE-BASED POLICY

RAND health insurance study

Oregon Medicaid expansion

HUD's Moving to Opportunity

Tennessee STAR

POLICY EVIDENCE INDUSTRY

Jameel Poverty Action Lab (J-PAL)

Cochrane Collaboration

Campbell Collaboration

Should we have evidence for every policy or program?

No!

Science vs. art/craft/intuition

Ellie Murray
@EpiEllie

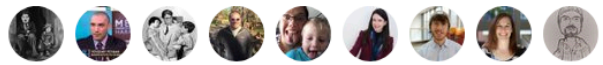
Follow

```
 ____________________________
|                            |
   IF U DONT SMOKE,
        U ALREADY
         BELIEVE IN
   CAUSAL INFERENCE
        WITHOUT
   RANDOMIZED TRIALS
|____________________________|
          (\__/)   ||
          (•ㅅ•)   ||
         /       づ
```

#HistorianSignBunny #Epidemiology

10:13 PM - 12 Jul 2018

200 Retweets **612** Likes



29 200 612

Should we have evidence for every policy or program?

No!

Science vs. art/craft/intuition

Smoking

Reducing the drinking age

Godwin's law

From Wikipedia, the free encyclopedia

Godwin's law (or **Godwin's rule of Hitler analogies**)[1][2] is an Internet adage asserting that "As an online discussion grows longer, the probability of a comparison involving Nazis or Hitler approaches 1";[2][3] that is, if an online discussion (regardless of topic or scope) goes on long enough, sooner or later someone will compare someone or something to Adolf Hitler or his deeds, the point at which effectively the discussion or thread often ends. Promulgated by the American attorney and author Mike Godwin in 1990,[2] Godwin's law

GODWIN'S LAW FOR STATISTICS

Correlation does not imply causation

Except when it does

Even if it doesn't,
this phrase is useless
and kills discussion

Not everyone found the news believable. “Facepalm. Correlation doesn’t imply causation,” wrote one unhappy Internet user. “That’s pretty much how I read this too… correlation is NOT causation,” agreed a Huffington Post superuser, seemingly distraught. “I was surprised not to find a discussion of correlation vs. causation,” cried someone at Hacker News. “Correlation does not mean causation,” a reader moaned at Slashdot. “There are so many variables here that it isn’t funny.”

David Robinson
@drob

Following

Correlation implies causation, don't @ me

1:12 PM - 22 Jun 2017 from Manhattan, NY

4 Retweets **56** Likes

2 4 56

Tweet your reply

David Robinson @drob · 22 Jun 2017

Replying to @drob

"Correlation implies casuation," the dean whispered as he handed me my PhD.

"But then why-"

"Because if they knew, they wouldn't need us."

5 46 169

John B. Holbein @JohnHolbein1 · Apr 7

Causality isn't achieved; it's approached.

3 1 8

Show this thread

John B. Holbein @JohnHolbein1 · Apr 7

Causality isn't binary; it's a continuum.

1 5 13

Show this thread

CORRELATION VS. CAUSATION

How do we figure out correlation?	How do we figure out causation?
Math and statistics	Philosophy. No math.

How do we know if X causes Y?

X causes Y if...

...we intervene and change X without changing anything else...

...and Y changes

WHAT IS CAUSALITY?

Y “listens to” X

X isn’t the only thing that causes Y

A light switch causes a light to go on, but not if bulb is burned out (no Y despite X) or if the light was already on (Y without X)

CAUSAL RELATIONSHIPS?

A light switch causes a light to be on

Lighting fireworks causes noise

Getting an MPA increases your earnings

Tariffs reduce trade

CAUSAL RELATIONSHIPS?

People wear shorts when ice cream trucks are out

Rooster crows are followed by sunrise

Colds go away a few days after you take vitamin C

CAUSATION

Causation =
Correlation + time order +
all other factors ruled out

How do you know if you have it right?

You need a philosophical model

That's what this class is for!

BREAK

CLASS DETAILS

Getting started with R

Evaluation, design, & causation

Theories of change

Logic models & results chains

Counterfacutal thinking & DAGs

Applied evaluation

Ethics of data science

Open science and collaboration

Politics and feasibility

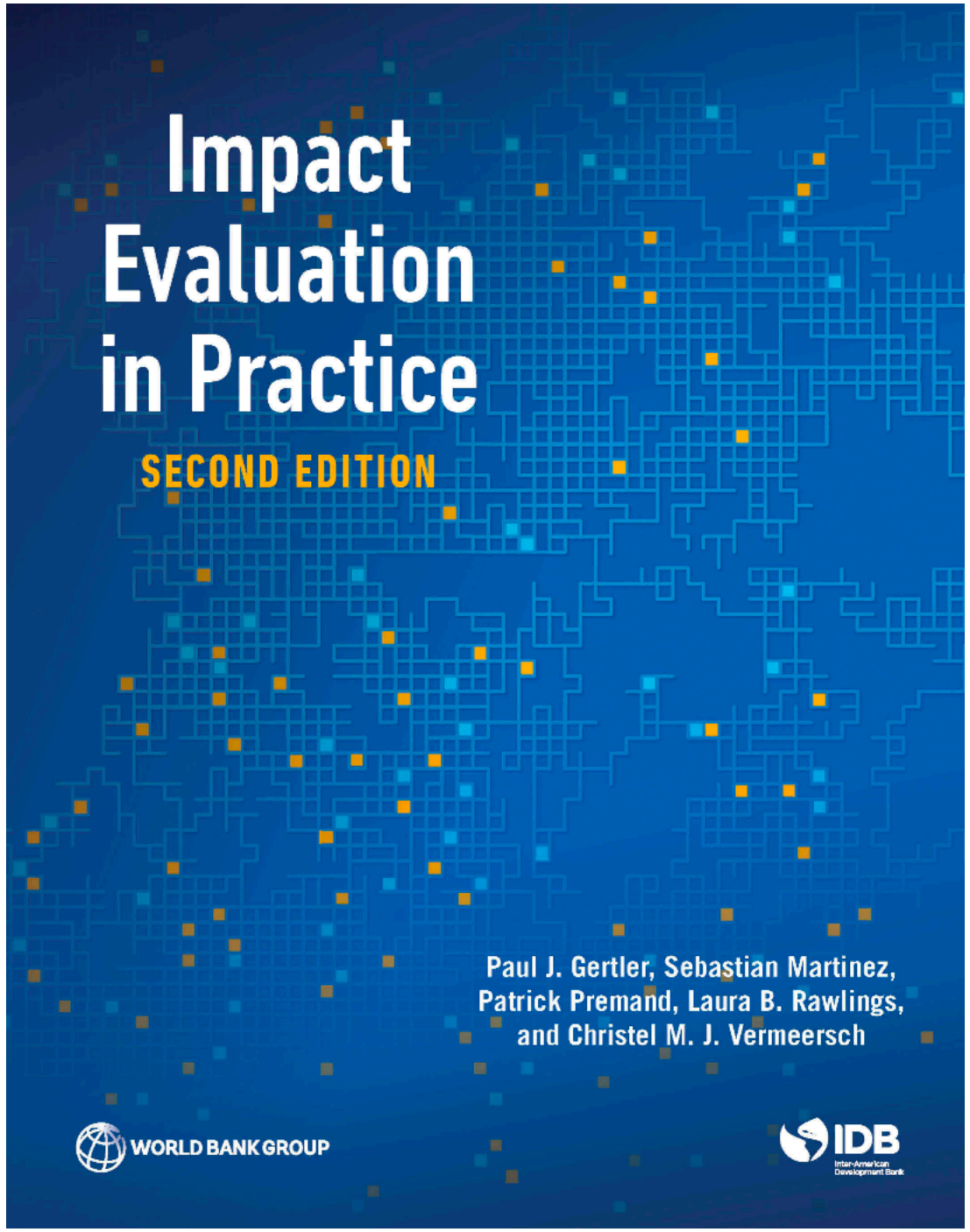
Communicating results

Statistical tools

Randomization & RCTs

DiDs, RDDs, and IVs

PROGRAM EVALUATION FOR PUBLIC SERVICE


Impact Evaluation in Practice
SECOND EDITION
Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch
WORLD BANK GROUP
IDB
Inter-American Development Bank

JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE
MASTERIN
METRIC
THE PATH FROM CAUSE TO EF

SCOTT CUNNINGHAM
CAUSAL INFERENCE:
THE MIXTAPE
(V. 1.7)

CLASSICAL STATISTICS CLASSES

Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Standard deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$$

Correlation

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

MODERN SUPER COOL STATISTICS CLASSES

```
mean(employee_salaries)
```

```
sd(employee_salaries)
```

```
cor(employee_salaries,
employee_experience)
```

MODERN SUPER COOL STATISTICS CLASSES

Modern and practical

Centered on data and tools

Quantitative, but not too mathy

Challenging, but not intimidating

CLASS TECHNOLOGY

THE TIDYVERSE

MODERN SUPER COOL STATISTICS CLASSES

```
strike_damages_month <- bird_strikes %>%
  group_by(Month) %>%
  summarize(total_damages = sum(Cost, na.rm = TRUE),
            average_damages = mean(Cost, na.rm = TRUE))

ggplot(data = strike_damages_month,
       mapping = aes(x = Month, y = total_damages)) +
  geom_col() +
  scale_y_continuous(labels = dollar) +
  labs(x = "Month",
       y = "Total damages",
       title = "Really expensive collisions happen in the fall?",
       subtitle = "Don't fly in August or October?",
       source = "Source: FAA Wildlife Strike Database")
```

SUCKING

There is no way to go from knowing nothing about a subject to knowing something about a subject without going through a period of much frustration and suckiness

Push through. You'll suck less.

Hadley Wickham, author of ggplot2 and the tidyverse

SUCKING

The New York Times

Opinion

SPORTING

(It's Great to) Suck at Something

By Karen Rinaldi

April 28, 2017

225

LEARNING R

Jesse Maegan
@kierisi

Following

My #rstats learning path:

1. Install R
2. Install RStudio
3. Google "How do I [THING I WANT TO DO] in R?"

Repeat step 3 ad infinitum.

7:19 AM - 18 Aug 2017

YOU CAN DO THIS.

GOAL FOR THE CLASS

Speak and do causation

Design rigorous evaluations

Change the world with data

PREREQUISITES

Math skills

None

Computer science skills

None

Statistical skills

Regression and differences in means

(ideally; you can survive without it, though)

MISCELLANEA

Class expectations

Late work

Technology

Participation

Other?

PLAYING WITH R

HERE WE GO!