Data Science Applications to Politics and Policy

Week 2

Melissa Sands

London School of Economics

GV4L3

Overview

1. Modes of Statistical Inference
   - Descriptive Inference
   - Predictive Inference
   - Causal Inference
2. “Big data” and Causal Inference
3. Credibility Crisis in the Social Sciences

Overview

1. Modes of Statistical Inference
   - Descriptive Inference
   - Predictive Inference
   - Causal Inference
2. “Big data” and Causal Inference
3. Credibility Crisis in the Social Sciences

Three Modes of Statistical Inference

1. Descriptive Inference: summarizing and exploring data
   - Inferring "ideal points" from roll-call votes
   - Inferring "topics" from texts and speeches
   - Inferring "economic development" from satellite images
2. Predictive Inference: forecasting out-of-sample data points
   - Inferring future state failures from past failures
   - Inferring election outcomes from polls
   - Inferring individual level behavior from aggregate data
3. Causal Inference: predicting *counterfactuals*
   - Inferring the effect of a policy on voter turnout
   - Inferring the impact of social media use on political polarization
   - Inferring whether severe weather events make people more likely to believe in climate change

Descriptive Inference Example: Mapping the Ideological Marketplace (Bonica 2014)

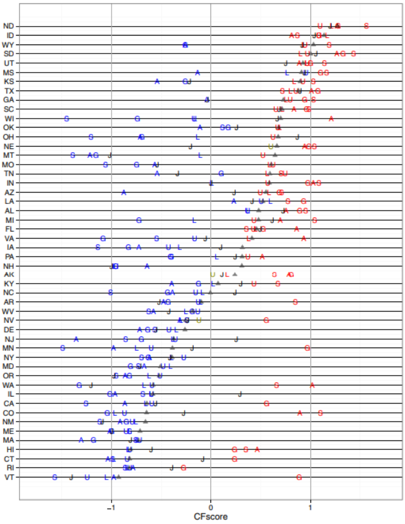
Develops method of measuring ideology of political candidates and campaign contributors using campaign finance data

- Data: Over 100 million campaign finance records from U.S. state and federal elections
- Generates ideal points using common pool of contributors who give to candidates across institutions and levels of politics
  - common-space campaign finance scores (CFscores)
- Assumption: contributors will on average prefer ideologically proximate candidates to those who are more distant.
  - Contribution amounts allocated based on distance between a candidate's ideal point and the contributor's ideal point
- Relies on donors who give to candidates for a variety of offices to bridge across institutions and levels of politics

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science*, 58(2), 367-386.

Bonica 2014

FIGURE 3 **Ideological Summary of State Politics (2010)**



Note: The symbols are interpreted as follows: G = Governor, A = Attorney General, S = Secretary of State, J = State Supreme Court (median), L = Lower Legislative Chamber (median), U = Upper Legislative Chamber (median), black triangle = mean ideal point of candidates elected in the state. The symbols are color coded by party (Dem = Blue; Rep = Red).

Predictive Inference Example: Voting behaviour (Nickerson & Rogers 2014)

- Campaigns use data to construct predictive models for more efficient targeting
- These models produce three types of “predictive scores” for each citizen in the voter database:
  - behavior scores: use past behavior and demographic information to calculate probabilities that citizens turn out, donate, etc.
  - support scores: survey a sample of citizens about their candidate/issue support. Use to gauge aggregate preferences.
  - responsiveness scores: predict how citizens will respond to campaign outreach using results of randomized field experiments. Model heterogeneous treatment effects and use to predict treatment responsiveness in target population.
    - Note that this can be considered *causal inference* – see next section

Example: Predicting voting behaviour (Nickerson & Rogers 2014)

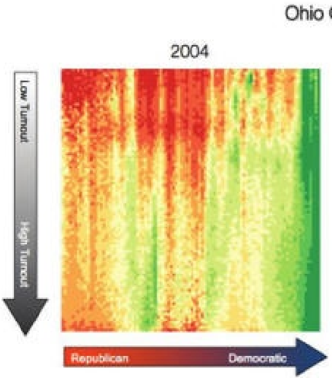
"Big data"

- voter files
- consumer and property data
- Census data (neighborhood characteristics)
- previous engagement (e.g., donating, volunteering)

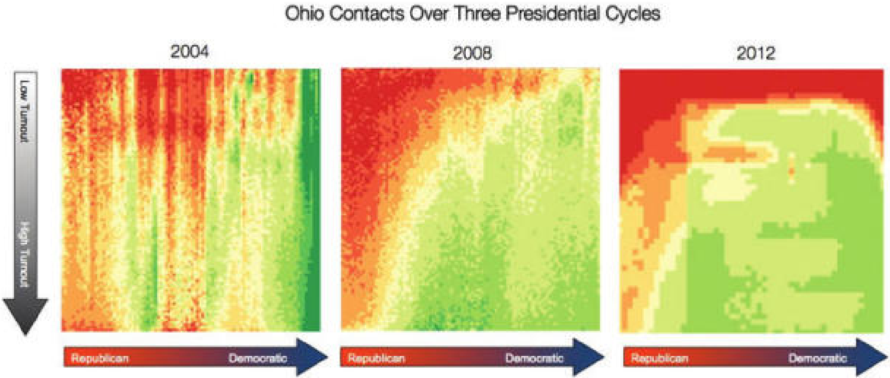
Analyses

- Simple regression techniques (e.g., OLS)
- Supervised learning algorithms

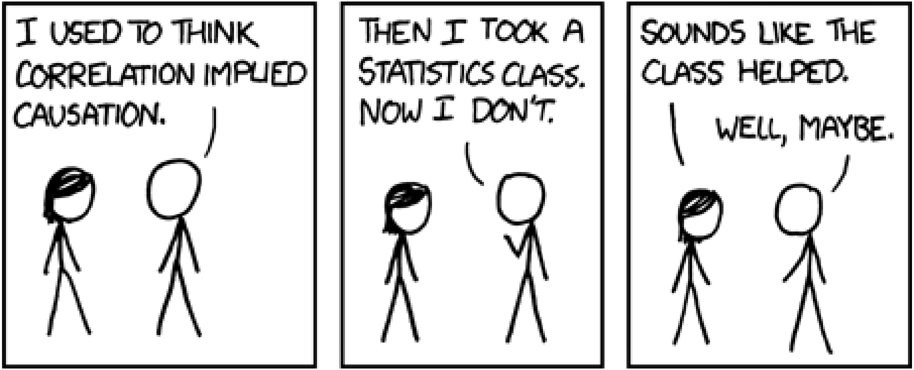
Nickerson & Rogers 2014



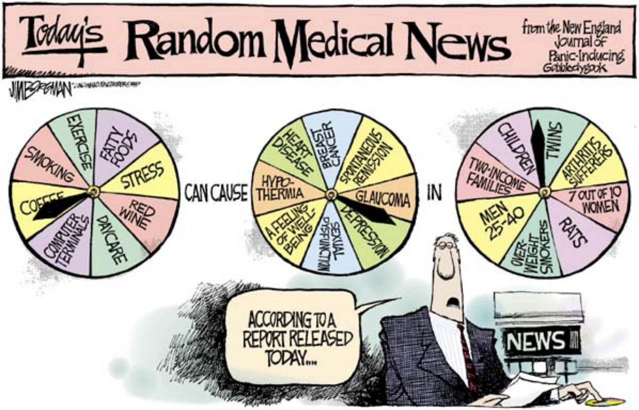
Nickerson & Rogers 2014



Causal Inference



Experiments vs. Observational Studies



Lancet 2003: comparing among individuals with the same age, gender, blood pressure, diabetes, and smoking, those with higher vitamin C levels have lower levels of obesity, lower levels of alcohol consumption, are less likely to grow up in working class, etc.

What is Causal Inference?

Causal inference $=$ inference about counterfactuals

A **causal effect** is a change in some feature of the world Y that *would result* from a change to some other feature of the world D.

Examples:

- Incumbency advantage:
  What *would* have been the election outcome if the candidate had not been an incumbent?
- Resource curse:
  What *would* have been the GDP growth rate without oil?
- Democratic peace:
  Would the two countries have fought each other if they had been both autocratic?
- Policy intervention:
  How many more disadvantaged youths *would* get employed under the new job training program?

Fundamental problem of causal inference

Causal Inference Example: GOTV Experiment (Gerber et al 2008)

- Voter turnout theories based on rational self-interested behavior generally fail to predict significant turnout unless they account for the utility that citizens receive from performing their civic duty.
- Two aspects of this type of utility, intrinsic satisfaction from behaving in accordance with a norm and extrinsic incentives to comply
- Gerber, Green, and Larimer (2008) test intrinsic motives in a large scale field experiment by applying varying degrees of extrinsic pressure on voters using to series of mailings to 180,002 households before the August 2006 primary election in Michigan.
  - Y_i: voted in primary (yes/no)
  - D_i: type of mailing

Example: GOTV Experiment (Gerber et al 2008)

- Civic Duty:
  - Encouraged to vote
- Hawthorne:
  - Encouraged to vote
  - Told that researchers would be checking on whether they voted
- Self:
  - Encouraged to vote
  - Told that whether one votes is a matter of public record
  - Shown whether members of their own household voted in the last two elections
- Neighbors:
  - Like Self but in addition recipients are shown whether the neighbors on the block voted in the last two elections elections

Social Pressure Treatment (Gerber et al 2008)

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	______
9995 JENNIFER KAY SMITH		Voted	______
9997 RICHARD B JACKSON		Voted	______
9999 KATHY MARIE JACKSON		Voted	______
9999 BRIAN JOSEPH JACKSON		Voted	______
9991 JENNIFER KAY THOMPSON		Voted	______
9991 BOB R THOMPSON		Voted	

Experimental results (Gerber et al 2008)

	Control (Not Mailed)	Civic Duty (Encouraged to vote)	Hawthorne (Encouraged & Monitored)	Self (Encouraged, Monitored, Shown Own Past Voting)	Neighbors (Encouraged, Monitored, Shown Own & Others' Past Voting)
Percent Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

Covariate balance (Gerber et al 2008)

With $n \simeq 180,000$, covariates are almost perfectly balanced:

TABLE 1. Relationship between Treatment Group Assignment and Covariates (Household-Level Data)

	Control	Civic Duty	Hawthorne	Self	Neighbors
	Mean	Mean	Mean	Mean	Mean
Household size	1.91	1.91	1.91	1.91	1.91
Nov 2002	.83	.84	.84	.84	.84
Nov 2000	.87	.87	.87	.86	.87
Aug 2004	.42	.42	.42	.42	.42
Aug 2002	.41	.41	.41	.41	.41
Aug 2000	.26	.27	.26	.26	.26
Female	.50	.50	.50	.50	.50
Age (in years)	51.98	51.85	51.87	51.91	52.01
N =	99,999	20,001	20,002	20,000	20,000

Note: Only registered voters who voted in November 2004 were selected for our sample. Although not included in the table, there were no significant differences between treatment group assignment and covariates measuring race and ethnicity.

Overview

“Big data” and Causal Inference

Having a lot of data can help us make causal inferences, under certain conditions. For example:

- If we have a randomized experiment with many participants
- In a regression discontinuity design (RDD) context, it helps to have many observations around the threshold

“Big data” and research design for causal inference are not substitutes.

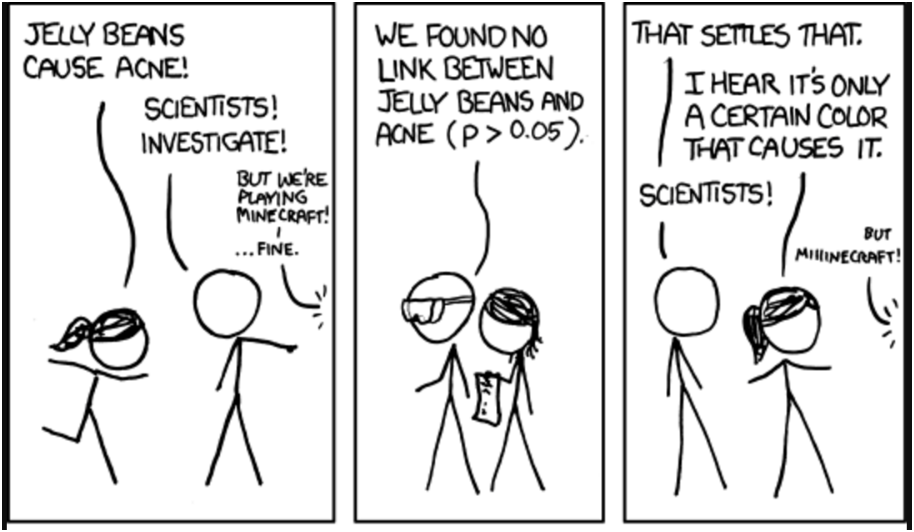
- Need exogenous variation and a solid research design in order to isolate causal effects (“identification precedes estimation”)
- Credibly identifying a source of exogenous variation requires creative insight about, and substantive knowledge of, the phenomenon under study.
- Computational techniques cannot replace “actual people doing actual thinking about social phenomena.” (Ashworth, Berry, & Bueno de Mesquita 2015)

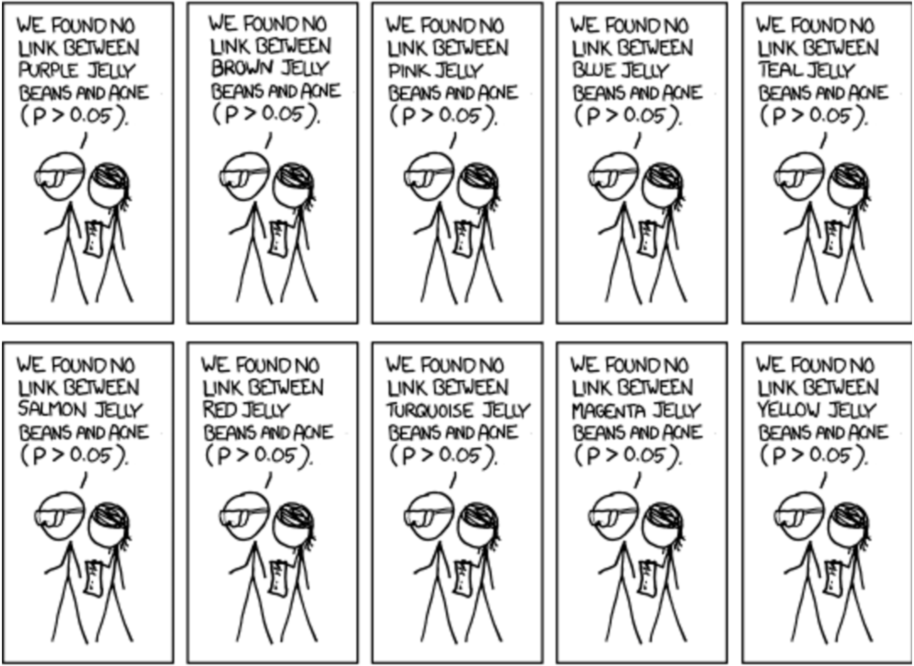
Overview

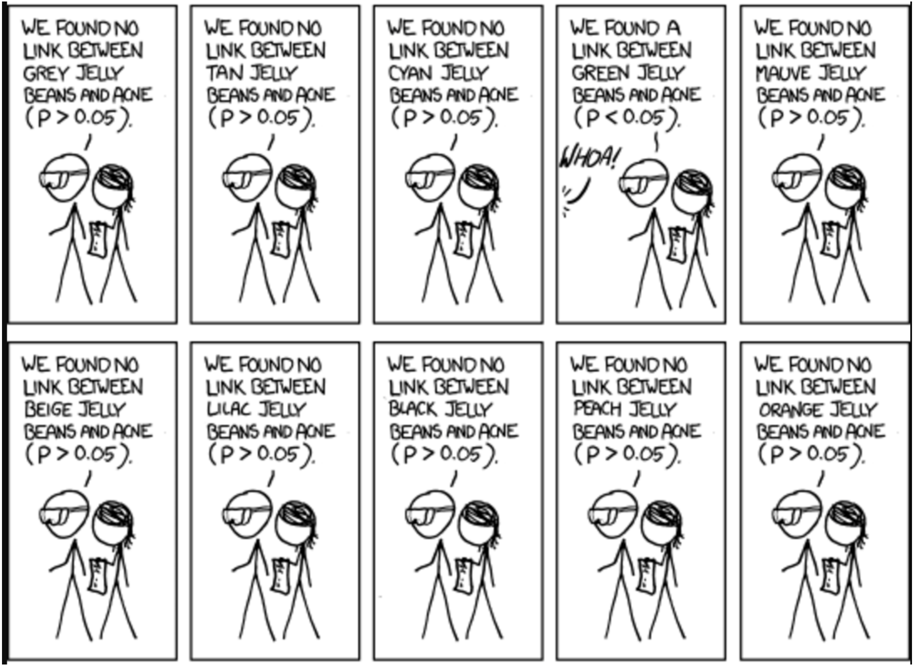
1. Modes of Statistical Inference
   - Descriptive Inference
   - Predictive Inference
   - Causal Inference
2. "Big data" and Causal Inference
3. Credibility Crisis in the Social Sciences

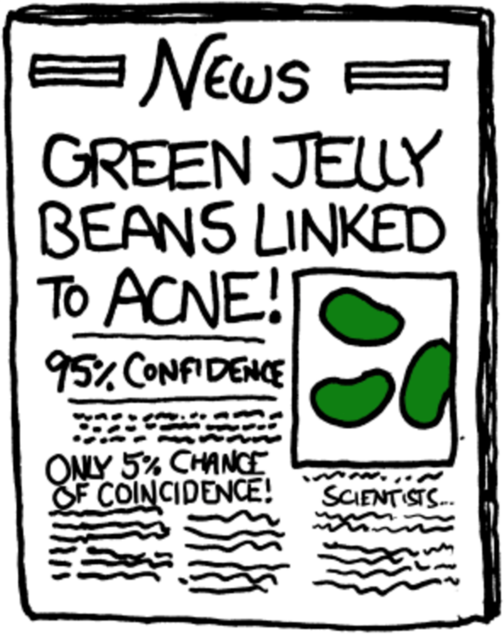
What is the problem?

Adapted from: Clark, J., Desposato, S., and McIntosh, C. 2017. "How to improve the credibility of (your) social science: A practical guide for researchers." Policy Design and Evaluation Lab (PDEL). University of California, San Diego.


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).


WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Source: XKCD

We often hear about ...

We often hear about ...

1. **"Replication crisis"**—studies fail to replicate (psych, econ, polisci, medicine, etc.)
2. **Publication bias**—published studies only represent fraction of results, biased toward statistically significant findings
3. **P-hacking/researcher degrees of freedom**—published studies use only a fraction of possible specifications, biased toward significance
4. **Misconduct/fraud**

$\rightarrow$ adds up to biased body of knowledge

Why do we have this credibility crisis?

Disciplines
- lack of **transparency** norm and tools
- **publishing** and **tenure** incentives
- **flexible research** questions and analysis

Researchers
- low incentive, capacity to **share data**
- **"file drawer problem"**
- **RDF** + **fraud** & **misconduct**

Social Science
- **"replication crisis"** for published work
- **biased** findings and collective knowledge

“Replication Crisis"

Social, behavioral, and medical studies often don’t replicate.

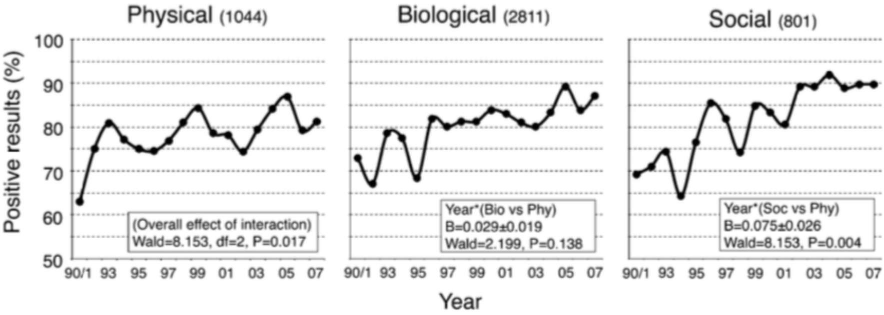
- Ideally, replications determine if original results are robust or due to *random chance*.
- In reality, failure to replicate often a result of ...
  - Lack of transparency in sharing data/code
  - Errors in data/code
  - Misconduct or fraud

Publication Bias (aka the “file drawer problem”)

- **Problem:** Studies more likely to be submitted/published when findings are significant $\rightarrow$ studies with null findings are hidden
- **Result:** Biased evidence base: we’re missing full universe of studies and results; what gets published could be due to random chance (e.g., if we expect 5% of results of all studies to be significant)

Fanelli (2010 & 2011)

Increase in % of papers with positive (i.e., they support the tested hypothesis) results over time, across scientific disciplines:



Franco, Malhotra, Simonovits (2014)

In a known population of conducted studies: strong results 60pp more likely to be written up than null results, 40pp more likely to be published:

Table 3. Cross-tabulation between statistical results of TESS studies and their publication status (column percentages reported). Pearson χ^2 test of independence: χ^2 (6) = 80.3, $P < 0.001$.

	Null (%)	Mixed (%)	Strong (%)
Not written	64.6	12.2	4.4
Written but not published	14.6	39.0	34.1
Published (non–top-tier)	10.4	37.8	38.4
Published (top-tier)	10.4	11.0	23.1
Total	100.0	100.0	100.0