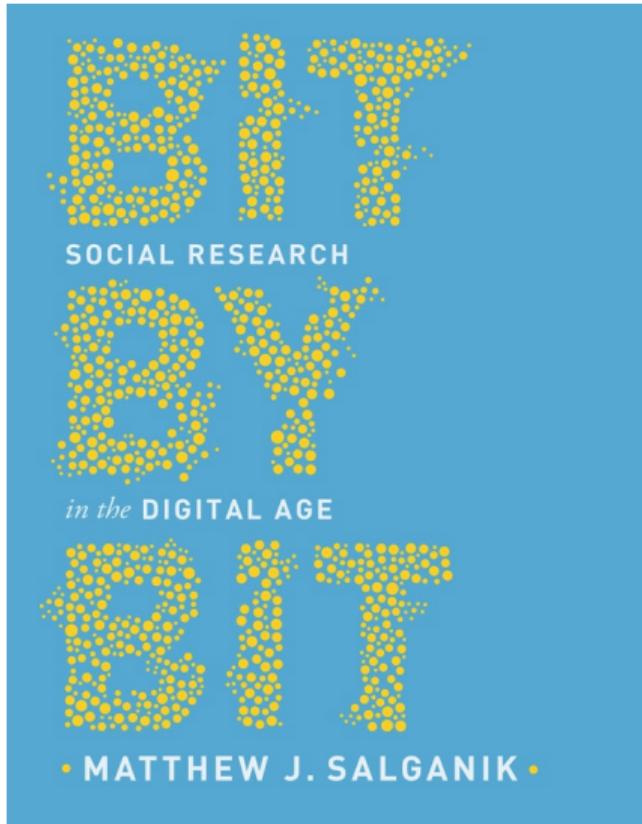


[Survey research in the digital age], [Probability and non-probability sampling], [Computer-administered interviews], [Combining surveys and big data], [Additions and extensions]

Matthew J. Salganik
Department of Sociology
Princeton University





- 1) Introduction
- 2) Observing behavior
- 3) Asking questions
- 4) Running experiments
- 5) Mass collaboration
- 6) Ethics
- 7) The future

Addition and extension 1 of 2: Wiki surveys

	Sampling	Interviews	Data environment
1st era	Area probability	Face-to-face	Stand-alone
2nd era	Random digital dial probability	Telephone	Stand-alone
3rd era	Non-probability	Computer-administered	Linked

Human-administered → Computer-administered

- ▶ enables change
- ▶ requires change

[home](#)
[winningest kittens](#)
[losingest kittens](#)
[newest kittens](#)
[add your kitten](#)
[facebook group](#)
[kittenwar myspace](#)
[faq](#)
[e-mail us](#)
kitten search:

[t-shirts and stuff](#)



kittenwar



Henry



Betty

VS.

[Click the cutest kitten picture!](#)

Can't decide? [Refresh the page](#) for a draw.

[Kittenwar has a brilliant new server, check it out! Thank you!](#)

[home](#)
[winningest kittens](#)
[losingest kittens](#)
[newest kittens](#)
[add your kitten](#)
[facebook group](#)
[kittenwar myspace](#)

[faq](#)
[e-mail us](#)

kitten search:

t-shirts and stuff

RESULTS



WIN LOSE



CLICK PICS FOR STATS

53% of people
agree that Henry is
cuter than Betty.



kittenwar



Young Japhy



Kizzibit

VS.

Click the cutest kitten picture!

Can't decide? [Refresh the page](#) for a draw.

[Kittenwar has a brilliant new server, check it out! Thank you!](#)

[home](#)
[winningest kittens](#)
[losingest kittens](#)
[newest kittens](#)
[add your kitten](#)
[facebook group](#)
[kittenwar myspace](#)

[faq](#)
[e-mail us](#)

[kitten search:](#)

[t-shirts and stuff](#)

RESULTS



WIN LOSE



[CLICK PICS FOR STATS](#)

51% of people
agree that [Kizzbit](#)
is cuter than [Young](#)
[Japhy](#).



kittenwar



Maria



Emelio Shikaka

VS.

[Click the cutest kitten picture!](#)

Can't decide? [Refresh the page](#) for a draw.

[Kittenwar has a brilliant new server, check it out! Thank you!](#)





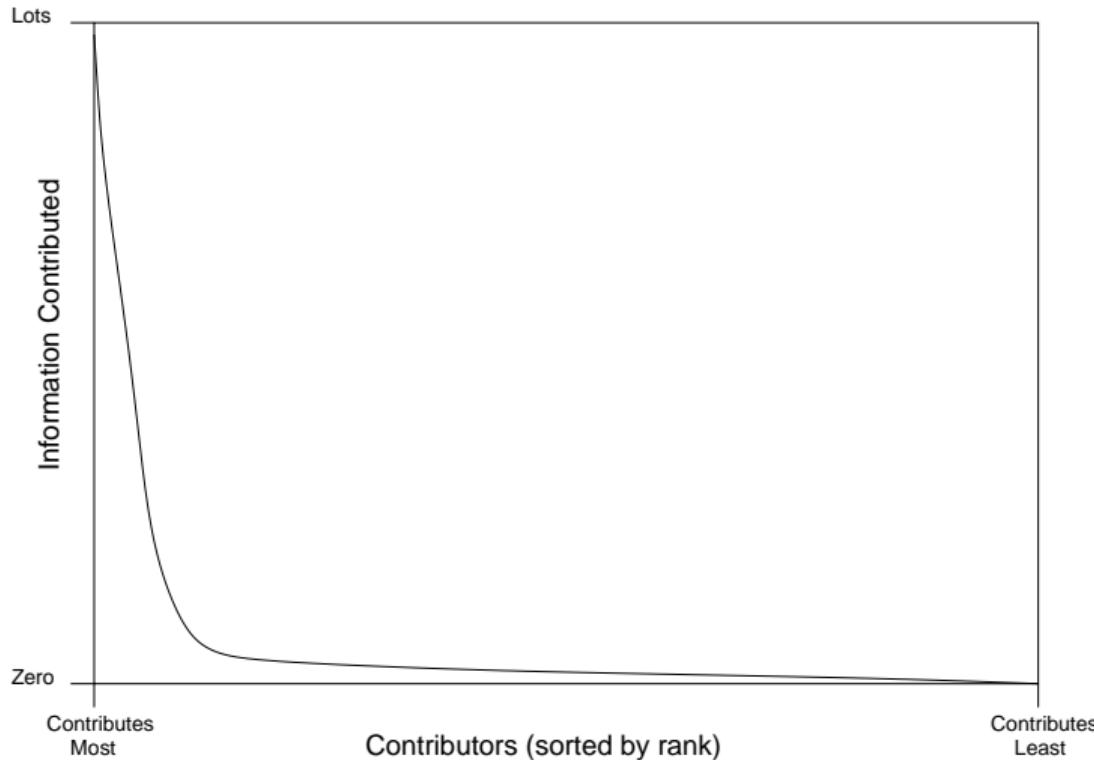
quantification or openness

quantification + openness =
wiki surveys

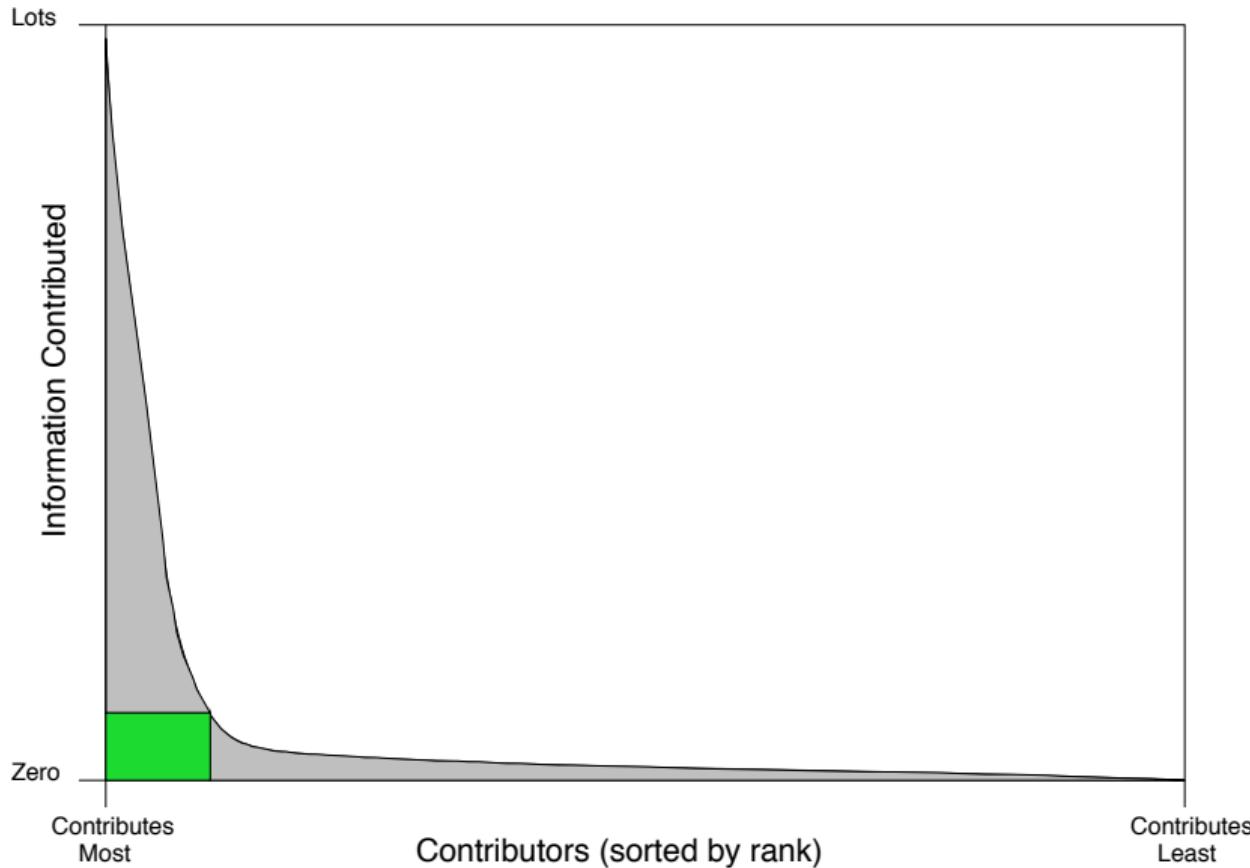
General principles of wiki surveys:

- ▶ greedy

Good web-based systems use
the **fat-head** and the **long-tail**



Surveys don't use the fat-head or the long-tail



General principles of wiki surveys:

- ▶ greedy

General principles of wiki surveys:

- ▶ greedy
- ▶ collaborative

General principles of wiki surveys:

- ▶ greedy
- ▶ collaborative
- ▶ adaptive



create
a question



collaborate
with others



discover
the best ideas

What is this?

All Our Ideas is a platform that enables groups to collect and prioritize ideas in a transparent, democratic, and bottom-up way. It's a suggestion box for the digital age.

[Learn more](#)

How does it work?

You can use All Our Ideas to create a website where visitors can vote on ideas and upload new ones. The intuitive and fun voting process yields powerful results.

[See an example](#)

Get started!

Click on the link below to get started with All Our Ideas. It is free, easy, and built on open source technology. Create your own interactive suggestion box and start discovering.

[Create a question](#)



This project is supported by a grant from the Center for Information Technology Policy at Princeton University.



SOURCE

All Our Ideas is open source software. Feel free to review, remix, or redesign. Also, you can use our API to create your own site.



[Log In](#)
[Create Your Own Question](#)
[About](#)



Which do you think is a better idea for creating a greener, greater New York City?

Seeded the wiki survey with 25 ideas:

- ▶ Require all big buildings to make certain energy efficiency upgrades
- ▶ Increase targeted tree plantings in neighborhoods with high asthma rates
- ▶ Establish a New York City Energy Planning Board

Which do you think is a better idea for creating a greener, greater New York City?

Focus on planting street trees before putting them in existing green space

Enforce low density zoning laws and do Not grant variances that are contrary to these protective laws.

I can't decide

10 votes on 269 ideas

Add your own idea

Which do you think is a better idea for creating a greener, greater New York City?

Plant more trees

Get Bus Lanes on Broadway

I can't decide

11 votes on 269 ideas

Add your own idea

You chose Enforce low density zoning laws and do Not grant variances that are contrary to these protective laws. over Focus on planting street trees before putting them in existing green space

Now you have cast 1 vote (average is 10)

[View all the results](#)

Which do you think is a better idea for creating a greener, greater New York City?

Provide funding to increase energy efficiency of buildings (PACE bonds/loans) creating green jobs, reducing emissions and utility bills.

Make sure that there are bike racks installed at or near all public schools and libraries.

I can't decide

12 votes on 269 ideas

Add your own idea

You chose [Get Bus Lanes on Broadway](#) over [Plant more trees](#)

Now you have cast 2 votes (average is 10)

[View all the results](#)

Which do you think is a better idea for creating a greener,
greater New York City?

Score

Keep NYC's drinking water clean by banning fracking in NYC's watershed.

84 [?]



Invest in multiple modes of transportation and provide both improved infrastructure and improved safety

81 [?]



Plug ships into electricity grid so they don't idle in port - reducing emissions equivalent to 12000 cars per ship.

78 [?]



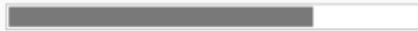
Implement congestion pricing in lower Manhattan

74 [?]



Continue enhancing bike lane network, to finally connect separated bike lane systems to each other across all five boroughs.

73 [?]



Composting! Provide municipal support for composting!!

73 [?]



Support and protect community gardens and create mechanisms to create new gardens and open space

72 [?]



Provide long-term leases for organic farms in unused public spaces, a garden at every public school and public housing development

72 [?]



Provide better transit service outside of Manhattan

72 [?]



Create a network of protected bike paths throughout the entire city

71 [?]



What are we trying to estimate?

Data

Vote	Session	Prompt	
1	1	item 4	item 1
2	1	item 3	item 1
3	1	item 4	item 3
4	2	item 3	item 4
5	2	item 4	item 2
:	:	:	:



Opinion matrix

$$\begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,K} \\ \theta_{2,1} & \theta_{2,2} & \dots & \theta_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{J,1} & \theta_{J,2} & \dots & \theta_{J,K} \end{bmatrix}$$

$\theta_{j,k}$: how much respondent j likes item k

What are we trying to estimate?

Opinion matrix

$$\begin{bmatrix} \hat{\theta}_{1,1} & \hat{\theta}_{1,2} & \dots & \hat{\theta}_{1,K} \\ \hat{\theta}_{2,1} & \hat{\theta}_{2,2} & \dots & \hat{\theta}_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}_{J,1} & \hat{\theta}_{J,2} & \dots & \hat{\theta}_{J,K} \end{bmatrix}$$

$\hat{\theta}_{j,k}$: how much respondent j likes
item k

What are we trying to estimate?

Opinion matrix

$$\begin{bmatrix} \hat{\theta}_{1,1} & \hat{\theta}_{1,2} & \dots & \hat{\theta}_{1,K} \\ \hat{\theta}_{2,1} & \hat{\theta}_{2,2} & \dots & \hat{\theta}_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}_{J,1} & \hat{\theta}_{J,2} & \dots & \hat{\theta}_{J,K} \end{bmatrix}$$

$\hat{\theta}_{j,k}$: how much respondent j likes item k



Score

$$[\hat{s}_1, \hat{s}_2, \dots, \hat{s}_K]$$

s_k : probability that item k beats a randomly chosen item for a randomly chosen session

Which do you think is a better idea for creating a greener, greater New York City?

Seeded the wiki survey with 25 ideas:

- ▶ Require all big buildings to make certain energy efficiency upgrades
- ▶ Increase targeted tree plantings in neighborhoods with high asthma rates
- ▶ Establish a New York City Energy Planning Board

Recruited participants through Twitter, Facebook, blogs, etc.



@NYCMayorsOffice

NYC Mayor's Office

Do you have ideas about how to make NYC
greener? Help update #PlaNYC.

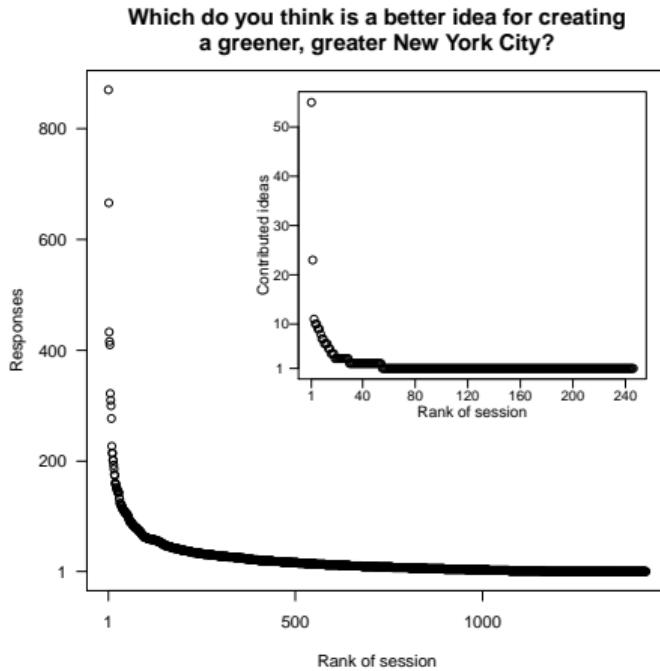
<http://bit.ly/9xeA88>

25 Oct via web Favorite Undo Retweet Reply

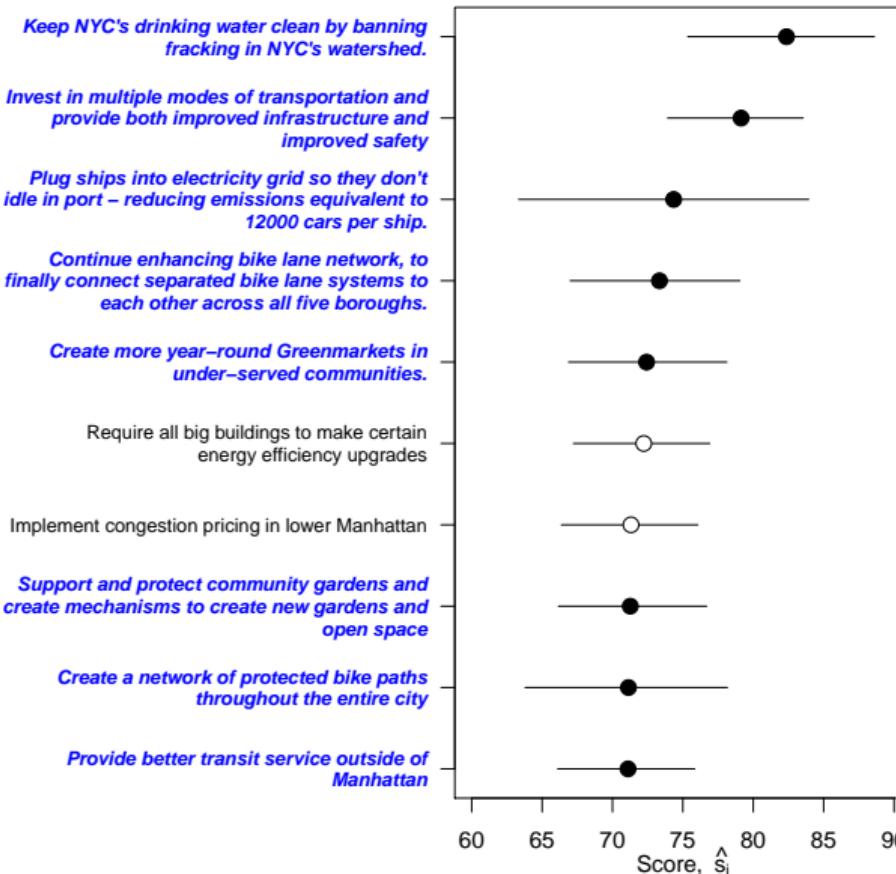
Retweeted by [allourideas](#) and 15 others



- ▶ 31,893 responses
- ▶ 464 ideas uploaded

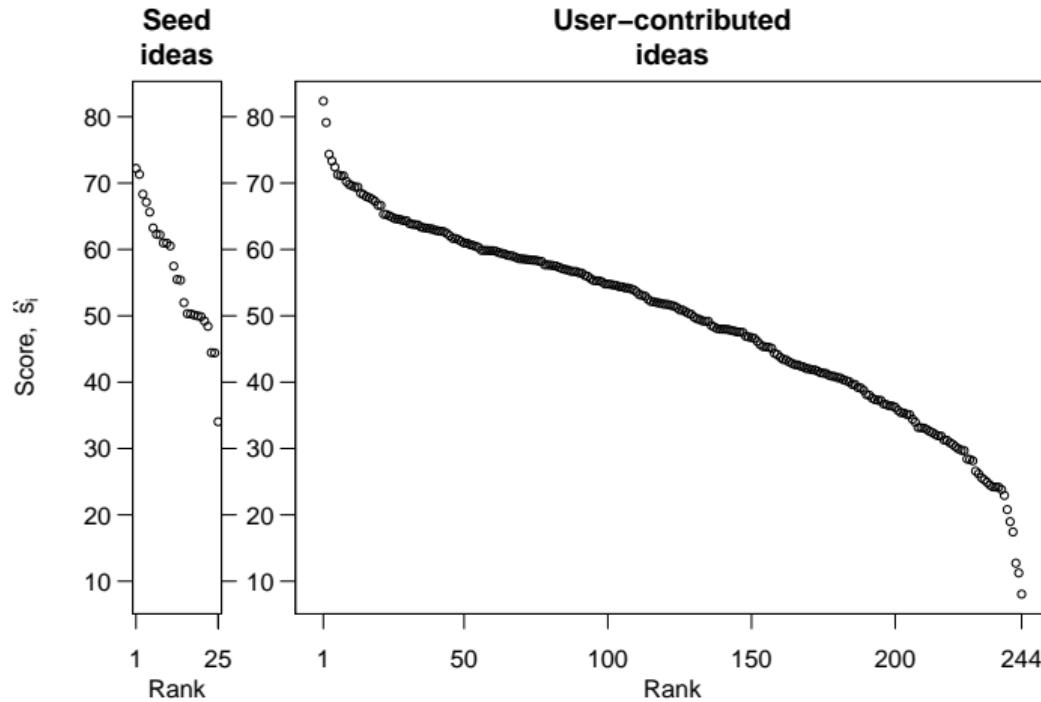


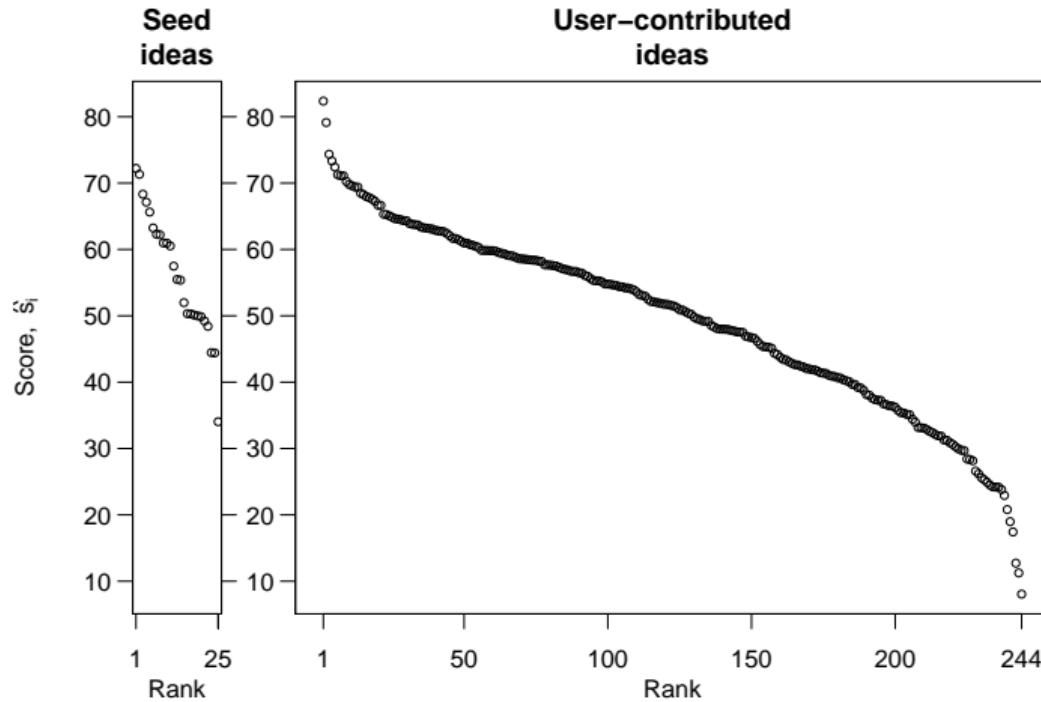
Which do you think is a better idea for creating a greener, greater New York City?



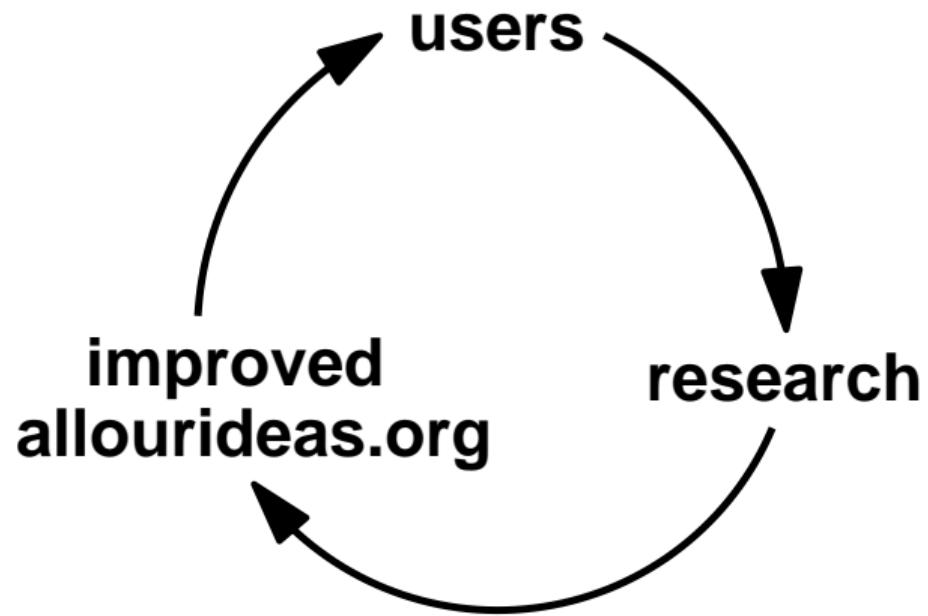
- ▶ Alternative framings: “Keep NYC’s drinking water clean by banning fracking in NYC’s watershed”

- ▶ Alternative framings: “Keep NYC’s drinking water clean by banning fracking in NYC’s watershed”
- ▶ Novel information: “Plug ships into electricity grid so they don’t idle in port - reducing emissions equivalent to 12000 cars per ship.”





variance + volume → extreme cases



Currently hosting:

18,000 wiki surveys with 975,000 ideas and 33 million votes



RESEARCH ARTICLE

Wiki Surveys: Open and Quantifiable Social Data Collection

Matthew J. Salganik¹, Karen E. C. Levy²

1 Department of Sociology, Center for Information Technology Policy, and Office of Population Research, Princeton University, Princeton, NJ, USA, **2** Information Law Institute and Department of Media, Culture, and Communication, New York University, New York, NY, USA and Data & Society Research Institute, New York, NY, USA

Salganik and Levy (2015)

Addition and extension 2 of 2: Amplified asking, a next step

Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,^{1*} Gabriel Cadamuro,² Robert On³

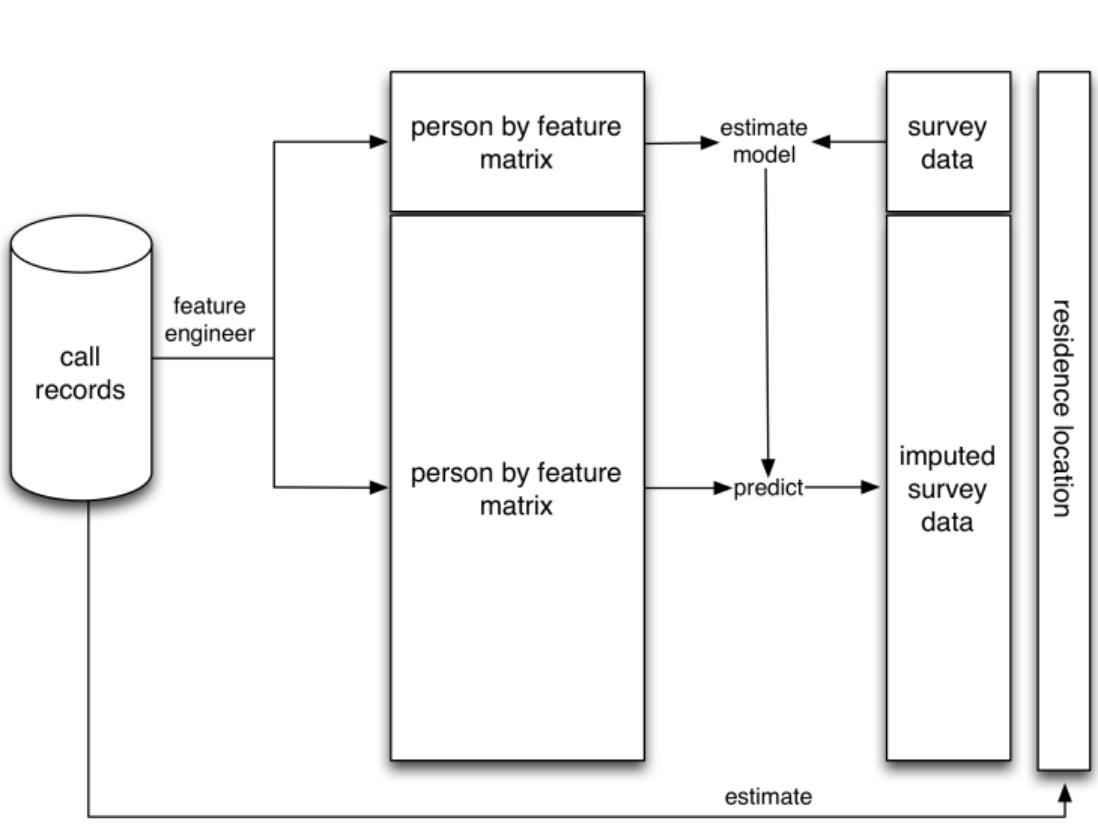
Combining satellite imagery and machine learning to predict poverty

Neal Jean,^{1,2*} Marshall Burke,^{3,4,5*} Michael Xie,¹ W. Matthew Davis,⁴
David B. Lobell,^{3,4} Stefano Ermon¹

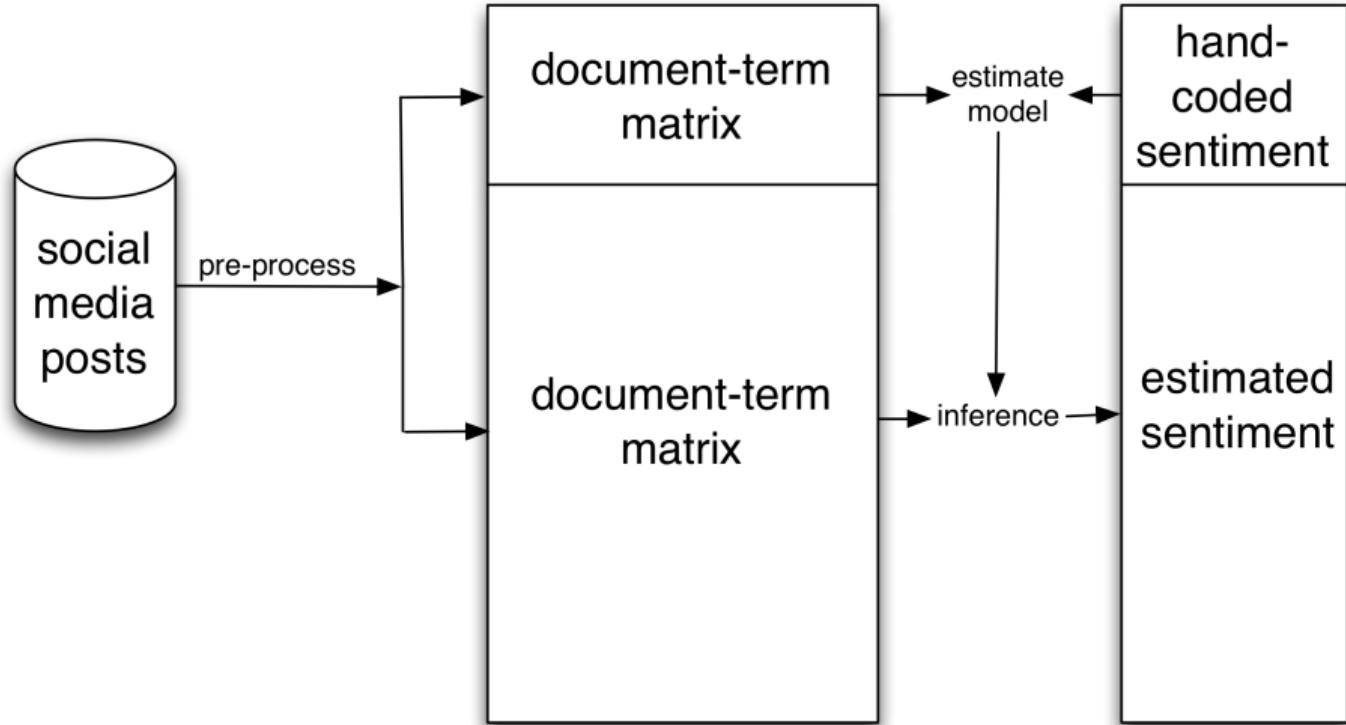
This paper is amazing and surprising (to me). First a digression

Supervised learning:

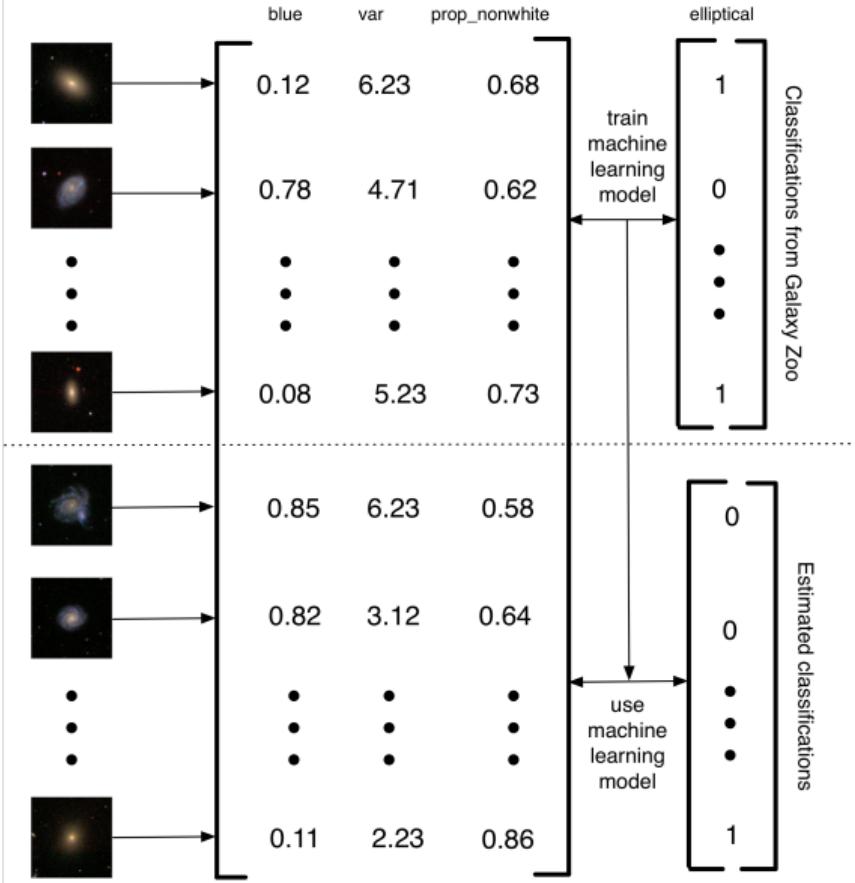
Lots of input-output pairs; goal is to develop a function that will predict the output from the input



See Chapter 3 of Salganik (2018)



See Chapter 2 of Salganik (2018)



See Chapter 5 of Salganik (2018)

Supervised learning:

Lots of input-output pairs; goal is to develop a function that will predict the output from the input

What if rather than engineering the features you could “learn” them automatically?

Deep learning

Yann LeCun^{1,2}, Yoshua Bengio³ & Geoffrey Hinton^{4,5}

LeCun, Bengio, Hinton (2015)

Combining satellite imagery and machine learning to predict poverty

Neal Jean,^{1,2*} Marshall Burke,^{3,4,5*}† Michael Xie,¹ W. Matthew Davis,⁴
David B. Lobell,^{3,4} Stefano Ermon¹

Combining satellite imagery and machine learning to predict poverty

Neal Jean,^{1,2*} Marshall Burke,^{3,4,5*}† Michael Xie,¹ W. Matthew Davis,⁴
David B. Lobell,^{3,4} Stefano Ermon¹

Artificial Intelligence Is Predicting Human Poverty From Space

August 18, 2016 // 02:00 PM EST

<http://dx.doi.org/10.1126/science.aaf7894>

https://motherboard.vice.com/en_us/article/artificial-intelligence-is-predicting-human-poverty-from-space

Live demo:

<https://www.google.com/maps/place/Kigali,+Rwanda/@-1.9546259,30.0345059,26517m/data=!3m2!1e3!4b1!4m5!3m4!1s0x19dca4258ed8e797:0xf32b36a5411d0bc8!8m2!3d-1.9705786!4d30.1044288>

But, most people had been using night lights



https://www.nasa.gov/multimedia/imagegallery/image_feature_2480.html

Prior research:

Nightlights + survey data to estimate wealth in places without surveys

Jean et al. (2016):
Day pictures + Nightlights + survey data to estimate wealth in places without surveys

Predicting poverty

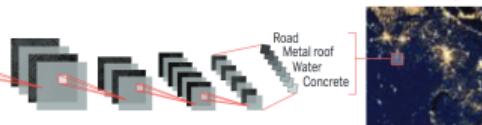
Satellite images can be used to estimate wealth in remote regions.

Neural network learns features in satellite images that correlate with economic activity

Daytime satellite photos capture details of the landscape



Convolutional Neural Network (CNN) associates features from daytime photos with nightlight intensity



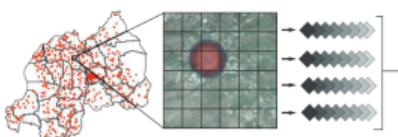
Satellite nightlights are a proxy for economic activity



Daytime satellite images can be used to predict regional wealth

Household survey locations

CNN processes satellite photos of each survey site



Features from multiple photos are averaged

Ridge regression model reconstructs ground truth estimates of poverty

- ▶ Start with CNN pretrained on ImageNet (e.g. hamsters and weasels)

Predicting poverty

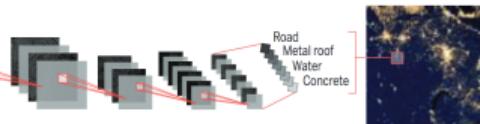
Satellite images can be used to estimate wealth in remote regions.

Neural network learns features in satellite images that correlate with economic activity

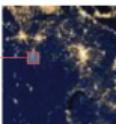
Daytime satellite photos capture details of the landscape



Convolutional Neural Network (CNN) associates features from daytime photos with nightlight intensity



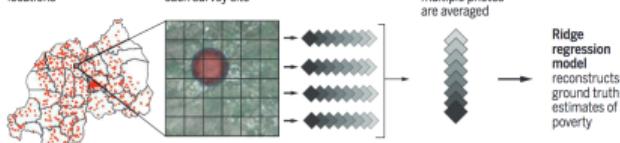
Satellite nightlights are a proxy for economic activity



Daytime satellite images can be used to predict regional wealth

Household survey locations

CNN processes satellite photos of each survey site

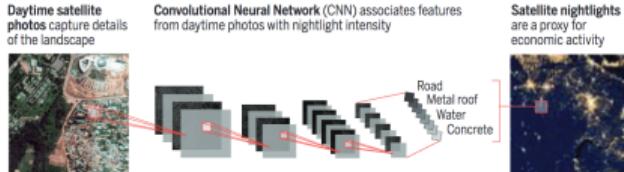


- ▶ Start with CNN pretrained on ImageNet (e.g. hamsters and weasels)
- ▶ Train CNN to predict nightlights from day pictures (lots of training data)

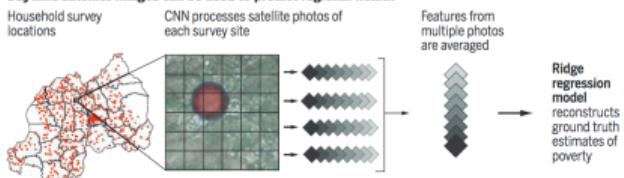
Predicting poverty

Satellite images can be used to estimate wealth in remote regions.

Neural network learns features in satellite images that correlate with economic activity

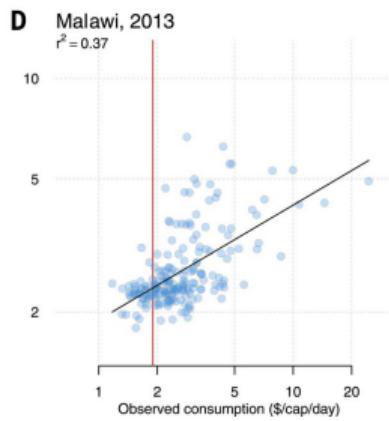
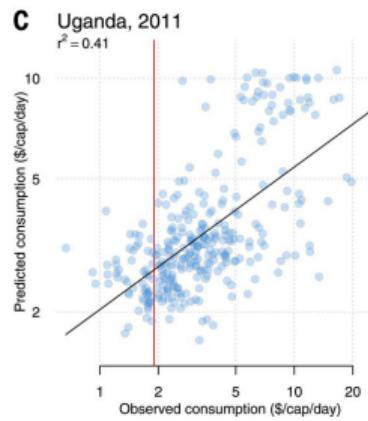
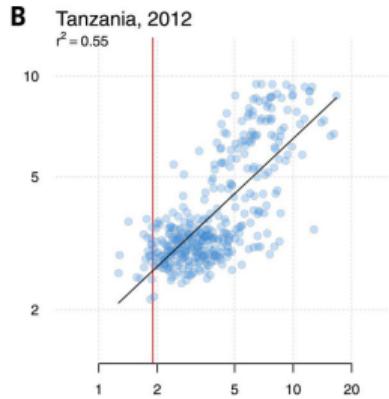
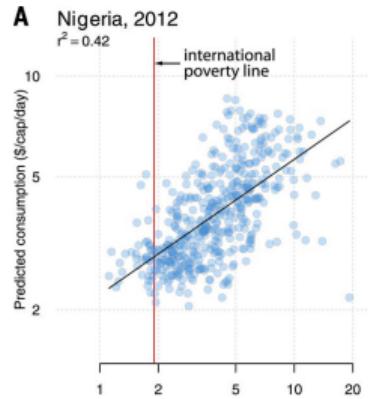


Daytime satellite images can be used to predict regional wealth

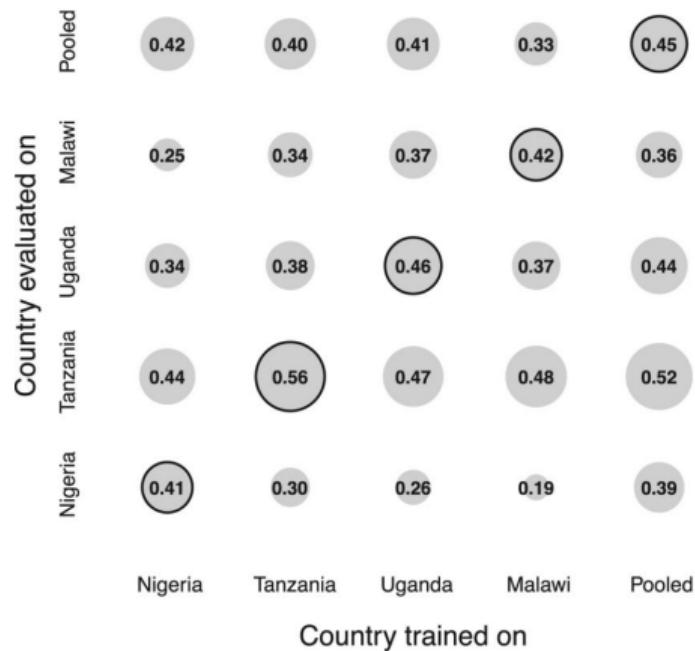


- ▶ Start with CNN pretrained on ImageNet (e.g. hamsters and weasels)
- ▶ Train CNN to predict nightlights from day pictures (lots of training data)
- ▶ Take features from CNN and train ridge regression to predict cluster mean survey response

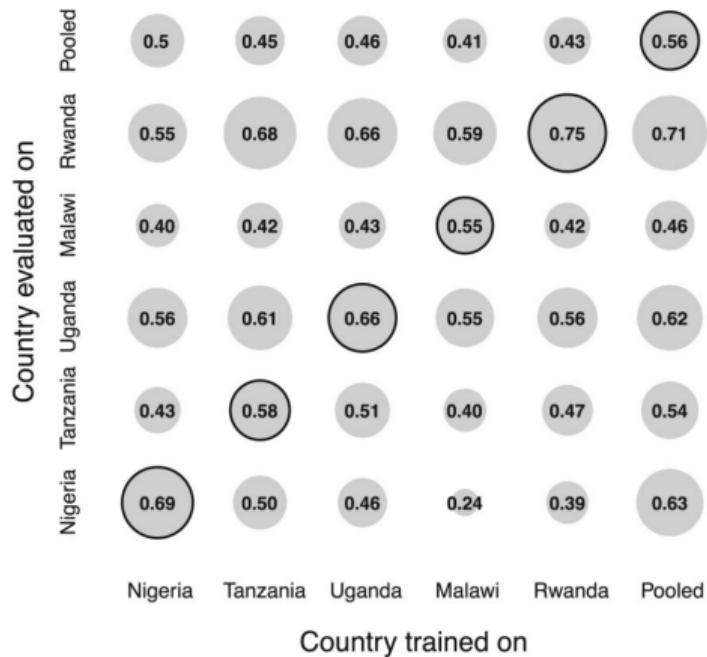
<http://dx.doi.org/10.1126/science.aah5217>

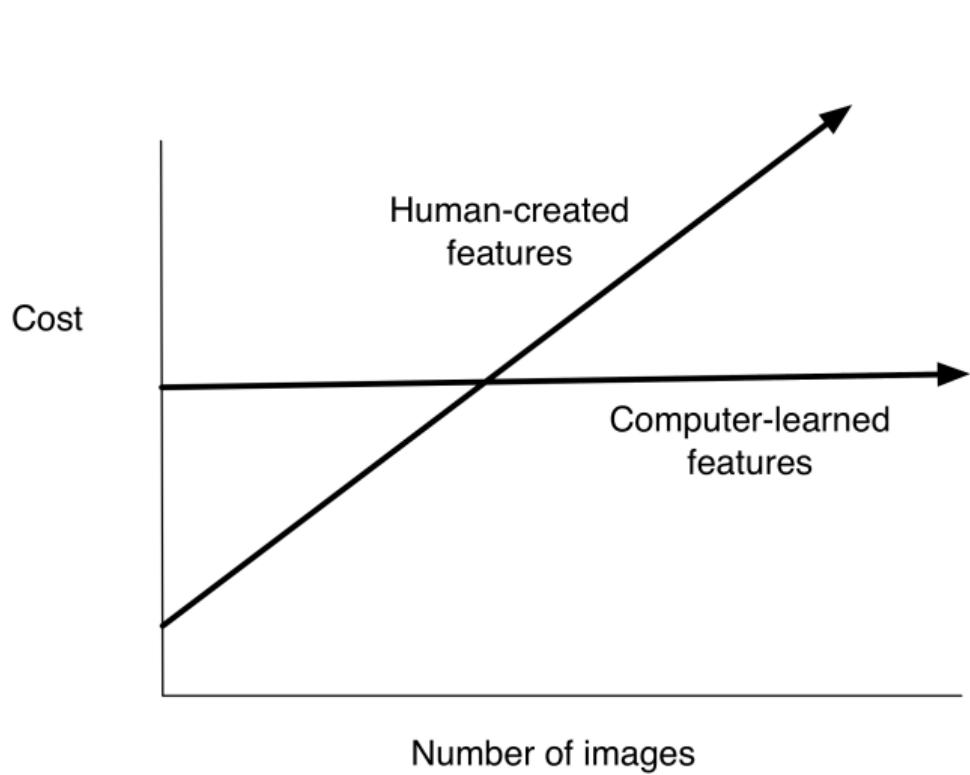


A Consumption expenditures



B Assets





 [nealjean / predicting-poverty](#)

[Watch](#) 22 [Star](#) 110 [Fork](#) 60

[Code](#) [Issues 1](#) [Pull requests 0](#) [Projects 0](#) [Wiki](#) [Insights](#)

Combining satellite imagery and machine learning to predict poverty

 18 commits  1 branch  0 releases  4 contributors  MIT

Branch: [master](#) [New pull request](#) [Create new file](#) [Upload files](#) [Find file](#) [Clone or download](#)

	imthexie select middle of pixel	Latest commit 975fddc on Mar 27	
	data/input	Clean replication code	10 months ago
	figures	Fixing cluster prefix in fig_utils.py	7 months ago
	model	Clean replication code	10 months ago
	scripts	select middle of pixel	3 months ago
	.gitignore	Clean replication code	10 months ago
	LICENSE	MIT License	6 months ago
	README.md	Update README.md	8 months ago
	requirements.txt	Clean replication code	10 months ago

<https://github.com/nealjean/predicting-poverty>

- ▶ Addition and extension 1 of 2: Wiki surveys
- ▶ Addition and extension 2 of 2: Amplified asking, a next step

[Survey research in the digital age], [Probability and non-probability sampling], [Computer-administered interviews], [Combining surveys and big data], [Additions and extensions]

Matthew J. Salganik
Department of Sociology
Princeton University

