

Wiki surveys: Open and quantifiable social data collection*

Matthew J. Salganik
Microsoft Research &
Dept. of Sociology, Princeton University

Joint work with Karen E.C. Levy

Rutgers, DIMACS
April 10, 2014

*This research is supported by Google, National Science Foundation, and the Center for Information Technology Policy at Princeton University.

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

VS.



Betty

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

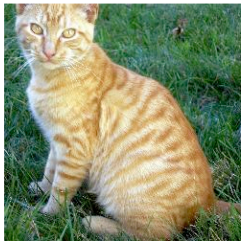


CLICK PICS FOR STATS

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

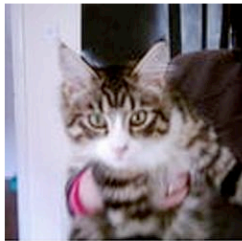


kittenwar



Young Japhy

VS.



Kizzibit

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

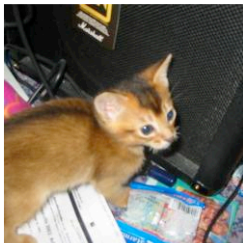


CLICK PICS FOR STATS

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

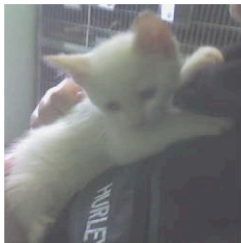


kittenwar



María

VS.

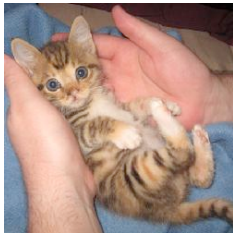


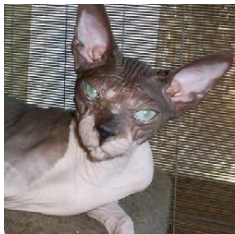
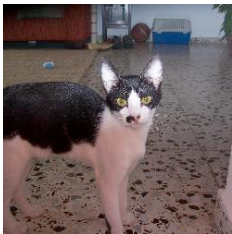
Emelio Shikaka

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!





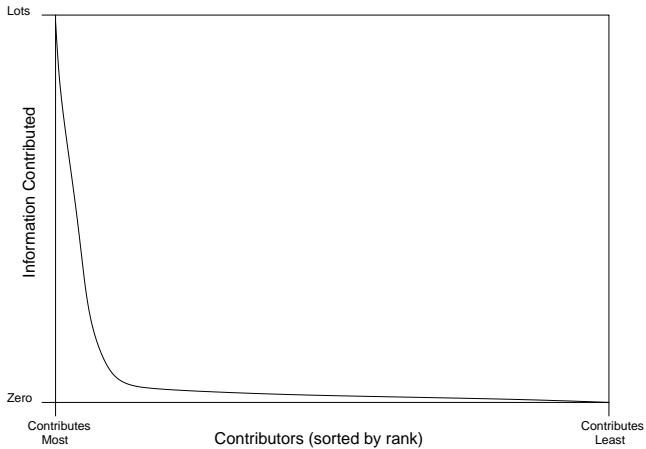
quantification or openness

quantification + openness =
wiki surveys

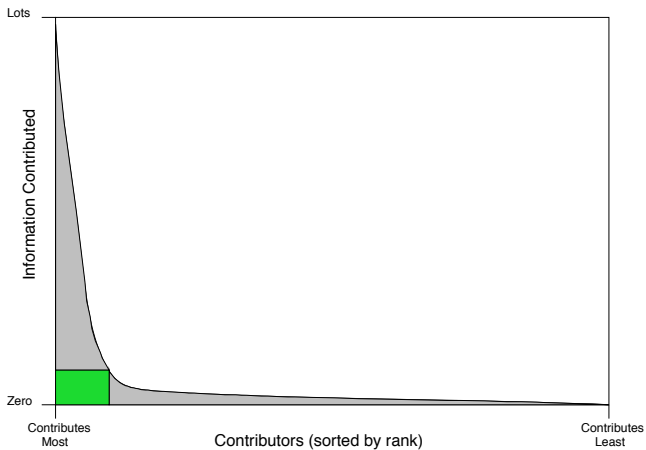
General principles of wiki surveys:

- ▶ 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.



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](#)



Learn how we are creating a greener, greater New York City.



Join the citywide conversation
to update PlaNYC

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



[Cast Votes](#) | [View Results](#) | [About this page](#)

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 [?]



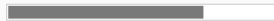
Implement congestion pricing in lower Manhattan

74 [?]



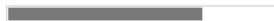
Continue enhancing bike lane network, to finally connect separated bike lane systems to each other across all five boroughs.

73 [?]



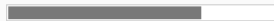
Composting! Provide municipal support for composting!!

73 [?]



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

72 [?]



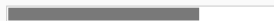
Provide long-term leases for organic farms in unused public spaces, a garden at every public school and public housing development

72 [?]



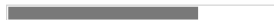
Provide better transit service outside of Manhattan

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
\vdots	\vdots	\vdots	\vdots



Public opinion

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

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

What are we trying to estimate?

Public opinion

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

$\theta_{j,k}$: 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

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

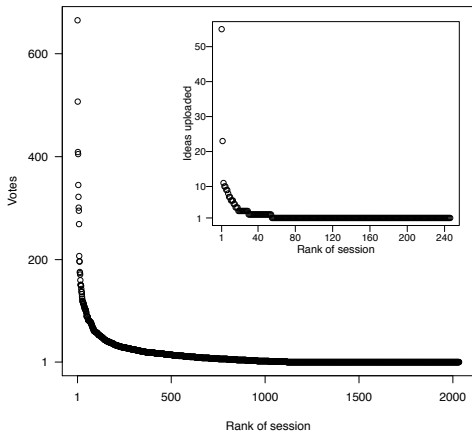
Retweeted by [allourideas](#) and 15 others



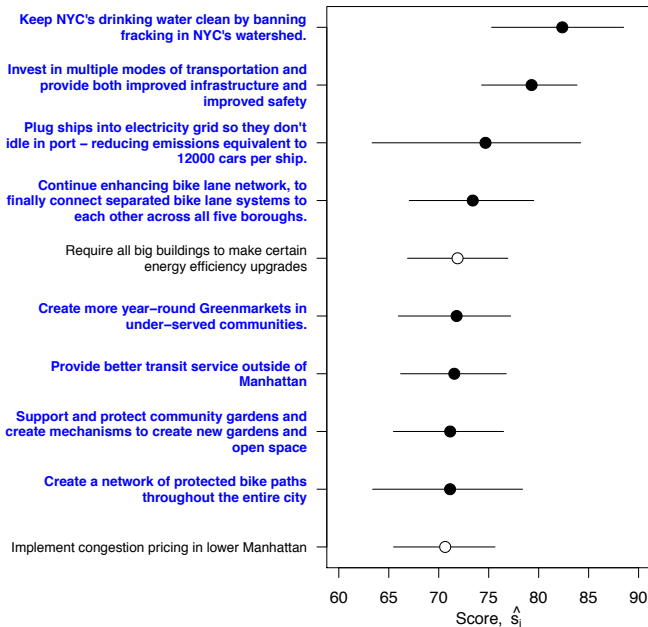
This is **not a random sample**, but random samples are possible

- ▶ 28,829 votes
- ▶ 464 ideas uploaded

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



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

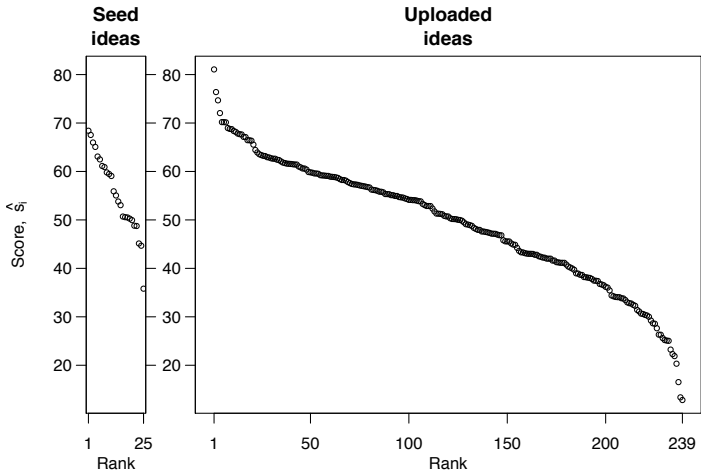


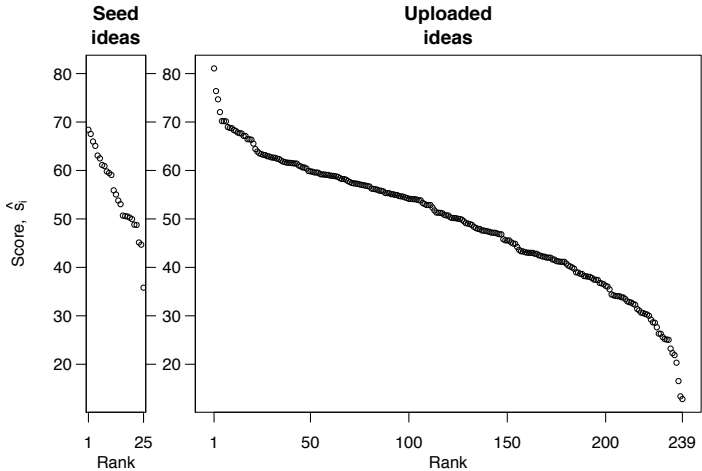
Two classes of uploaded ideas

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

Two classes of uploaded ideas

- ▶ 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 \rightarrow extreme cases

Currently hosting:
4,000 wiki surveys with 200,000 ideas and 5 million votes



Thank you

The All Our Ideas team, current and past:

- ▶ Peter Lubell-Doughtie, Adam Sanders, Pius Uzamere, Dhruv Kapadia, Chap Ambrose, Calvin Lee, Dmitri Garbuzov, Brian Tubergen, Peter Green, Luke Baker, Paul Yuan
- ▶ Josh Weinstein, Nadia Heninger, Bill Zeller, Bambi Tsui, Dhvani Shah, Karen Levy

More information:

- ▶ Salganik and Levy (2012) “Wiki surveys: Open and quantifiable social data collection”: arxiv.org/abs/1202.0500
- ▶ <http://www.allourideas.org>
- ▶ <http://blog.allourideas.org>
- ▶ <http://github.com/allourideas> (open source)
- ▶ Follow us: @allourideas

