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.
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!
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!
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
ALL OUR IDEAS
CREATE. CONTRIBUTE. DISCOVER.

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

Add your own idea

10 votes on 269 ideas
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

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

You chose Get Bus Lanes on Broadway over Plant more trees
Now you have cast 2 votes (average is 10)

View all the results
<table>
<thead>
<tr>
<th>Idea</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep NYC's drinking water clean by banning fracking in NYC's watershed.</td>
<td>84</td>
</tr>
<tr>
<td>Invest in multiple modes of transportation and provide both improved infrastructure and improved safety</td>
<td>81</td>
</tr>
<tr>
<td>Plug ships into electricity grid so they don't idle in port - reducing emissions equivalent to 12000 cars per ship.</td>
<td>78</td>
</tr>
<tr>
<td>Implement congestion pricing in lower Manhattan</td>
<td>74</td>
</tr>
<tr>
<td>Continue enhancing bike lane network, to finally connect separated bike lane systems to each other across all five boroughs.</td>
<td>73</td>
</tr>
<tr>
<td>Composting! Provide municipal support for composting!!</td>
<td>73</td>
</tr>
<tr>
<td>Support and protect community gardens and create mechanisms to create new gardens and open space</td>
<td>72</td>
</tr>
<tr>
<td>Provide long-term leases for organic farms in unused public spaces, a garden at every public school and public housing development</td>
<td>72</td>
</tr>
<tr>
<td>Provide better transit service outside of Manhattan</td>
<td>72</td>
</tr>
<tr>
<td>Create a network of protected bike paths throughout the entire city</td>
<td>71</td>
</tr>
</tbody>
</table>
What are we trying to estimate?

<table>
<thead>
<tr>
<th>Vote</th>
<th>Session</th>
<th>Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>item 4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>item 3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>item 4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>item 3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>item 4</td>
</tr>
</tbody>
</table>

Public opinion

\[
\begin{bmatrix}
\theta_{1,1} & \theta_{1,2} & \ldots & \theta_{1,K} \\
\theta_{2,1} & \theta_{2,2} & \ldots & \theta_{2,K} \\
\vdots & \vdots & \ddots & \vdots \\
\theta_{J,1} & \theta_{J,2} & \ldots & \theta_{J,K}
\end{bmatrix}
\]

\(\theta_{j,k}\): how much respondent \(j\) likes item \(k\)
What are we trying to estimate?

Public opinion

\[
\begin{bmatrix}
\hat{\theta}_{1,1} & \hat{\theta}_{1,2} & \ldots & \hat{\theta}_{1,K} \\
\hat{\theta}_{2,1} & \hat{\theta}_{2,2} & \ldots & \hat{\theta}_{2,K} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\theta}_{J,1} & \hat{\theta}_{J,2} & \ldots & \hat{\theta}_{J,K}
\end{bmatrix}
\]

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

Score

\[
[\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_K]
\]

\(s_k\): probability that item \(k\) beats a randomly chosen item for a randomly chosen session
Recruited participants through Twitter, Facebook, blogs, etc.

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?

- Keep NYC's drinking water clean by banning fracking in NYC's watershed.
- Invest in multiple modes of transportation and provide both improved infrastructure and improved safety.
- Plug ships into electricity grid so they don't idle in port – reducing emissions equivalent to 12000 cars per ship.
- Continue enhancing bike lane network, to finally connect separated bike lane systems to each other across all five boroughs.
- Require all big buildings to make certain energy efficiency upgrades.
- Create more year-round Greenmarkets in under-served communities.
- Provide better transit service outside of Manhattan.
- Support and protect community gardens and create mechanisms to create new gardens and open space.
- Create a network of protected bike paths throughout the entire city.
- Implement congestion pricing in lower Manhattan.
- Invest in multiple modes of transportation and provide both improved infrastructure and improved safety.
- 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.”
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.”
Seed ideas

Score, $s_i$

Rank

1

25

20

30

40

50

60

70

80

 Uploaded ideas

Score, $\hat{s}_i$

Rank

1

50

100

150

200

239
Seed ideas

Score, $s_i$

Rank 1 25
20
30
40
50
60
70
80

Uploaded ideas

Score, $\hat{s}_i$

Rank 1 50 100 150 200 239

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, Dhwani Shah, Karen Levy

More information:

- http://www.allourideas.org
- http://blog.allourideas.org
- http://github.com/allourideas (open source)
- Follow us: @allourideas