The Social Science Data Revolution

Gary King

Institute for Quantitative Social Science
Harvard University

(People, Power, & CyberPolitics Workshop, MIT, 12/8/11)

The Changing Evidence Base of Social Science Research

The Changing Evidence Base of Social Science Research

The Last 50 Years:

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

- Much more of the above

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

- Much more of the above
- Shrinking computers & the growing Internet: data everywhere

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

- Much more of the above
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: academic data sharing (e.g., Dataverse)

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

- Much more of the above
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: academic data sharing (e.g., Dataverse)
- Analogue-to-digital transformation of government records

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

- Much more of the above
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: academic data sharing (e.g., Dataverse)
- Analogue-to-digital transformation of government records
- Advances in statistical methods, informatics, & software

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

- Much more of the above
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: academic data sharing (e.g., Dataverse)
- Analogue-to-digital transformation of government records
- Advances in statistical methods, informatics, & software
- *The march of quantification*: through academia, professions, government, & commerce (*SuperCrunchers*, *The Numerati*, *MoneyBall*)

The Changing Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to...

- Much more of the above
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: academic data sharing (e.g., Dataverse)
- Analogue-to-digital transformation of government records
- Advances in statistical methods, informatics, & software
- *The march of quantification*: through academia, professions, government, & commerce (*SuperCrunchers*, *The Numerati*, *MoneyBall*)
- The end of the quantitative-qualitative divide

Examples of what's now possible

Examples of what's now possible

- Opinions of activists:

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise:

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week?

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers
- Social contacts:

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends"

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" $\rightsquigarrow$ continuous record of phone calls, emails, text messages, bluetooth, social media connections, electronic address books

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" $\rightsquigarrow$ continuous record of phone calls, emails, text messages, bluetooth, social media connections, electronic address books
- Economic development in developing countries:

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" $\rightsquigarrow$ continuous record of phone calls, emails, text messages, bluetooth, social media connections, electronic address books
- Economic development in developing countries: Dubious or nonexistent governmental statistics

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" $\rightsquigarrow$ continuous record of phone calls, emails, text messages, bluetooth, social media connections, electronic address books
- Economic development in developing countries: Dubious or nonexistent governmental statistics $\rightsquigarrow$ satellite images of human-generated light at night, or networks of roads and other infrastructure

Examples of what's now possible

- Opinions of activists: $\approx$1,000 interviews $\rightsquigarrow$ millions of political opinions in social media posts (1B every 4 days)
- Exercise: A survey: "How many times did you exercise last week? $\rightsquigarrow$ 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" $\rightsquigarrow$ continuous record of phone calls, emails, text messages, bluetooth, social media connections, electronic address books
- Economic development in developing countries: Dubious or nonexistent governmental statistics $\rightsquigarrow$ satellite images of human-generated light at night, or networks of roads and other infrastructure
- Many, many more...

One Example

One Example

of Automated Text Analysis

How to Read Billions of Social Media Posts

How to Read Billions of Social Media Posts

Daniel Hopkins and Gary King. "A Method of Automated Nonparametric Content Analysis for Social Science" *AJPS*. 54 (2010): 229-247

How to Read Billions of Social Media Posts

Daniel Hopkins and Gary King. "A Method of Automated Nonparametric Content Analysis for Social Science" *AJPS*. 54 (2010): 229-247

1. Downloaded & analyzed all English-language blog posts every day.

How to Read Billions of Social Media Posts

Daniel Hopkins and Gary King. "A Method of Automated Nonparametric Content Analysis for Social Science" *AJPS*. 54 (2010): 229-247

1. Downloaded & analyzed all English-language blog posts every day. (We learned: The university is not a research, not production, environment!)

How to Read Billions of Social Media Posts

Daniel Hopkins and Gary King. "A Method of Automated Nonparametric Content Analysis for Social Science" *AJPS*. 54 (2010): 229-247

1. Downloaded & analyzed all English-language blog posts every day. (We learned: The university is not a research, not production, environment!)
2. Commercialized in 2008:

crimson hexagon ⇝ 

⇝

Fast Company Names Crimson Hexagon Number Seven on "The 10 Most Innovative Companies in Web" List Leading Social Intelligence Firm Recognized For Revolutionary Measurement of Consumer Opinions in Social Media

Published: Wednesday, 16 Mar 2011 | 9:20 AM ET Text Size

CAMBRIDGE, Mass., Mar 16, 2011 (BUSINESS WIRE) -- Fast Company named

How to Read Billions of Social Media Posts

Daniel Hopkins and Gary King. "A Method of Automated Nonparametric Content Analysis for Social Science" *AJPS*. 54 (2010): 229-247

1. Downloaded & analyzed all English-language blog posts every day. (We learned: The university is not a research, not production, environment!)
2. Commercialized in 2008:

crimson hexagon ⇝ 

⇝

Fast Company Names Crimson Hexagon Number Seven on "The 10 Most Innovative Companies in Web" List Leading Social Intelligence Firm Recognized For Revolutionary Measurement of Consumer Opinions in Social Media

Published: Wednesday, 16 Mar 2011 | 9:20 AM ET Text Size

CAMBRIDGE, Mass., Mar 16, 2011 (BUSINESS WIRE) -- Fast Company named

3. CH collects *all* social media posts, runs huge servers with our methods

How to Read Billions of Social Media Posts

Daniel Hopkins and Gary King. "A Method of Automated Nonparametric Content Analysis for Social Science" *AJPS*. 54 (2010): 229-247

1. Downloaded & analyzed all English-language blog posts every day. (We learned: The university is not a research, not production, environment!)
2. Commercialized in 2008:

crimson hexagon ⇝ 

⇝

Fast Company Names Crimson Hexagon Number Seven on "The 10 Most Innovative Companies in Web" List Leading Social Intelligence Firm Recognized For Revolutionary Measurement of Consumer Opinions in Social Media

Published: Wednesday, 16 Mar 2011 | 9:20 AM ET Text Size

CAMBRIDGE, Mass., Mar 16, 2011 (BUSINESS WIRE) -- Fast Company named

3. CH collects *all* social media posts, runs huge servers with our methods
4. Crimson Hexagon Academic Grant Program to be announced soon

How to Read Billions of Social Media Posts

Daniel Hopkins and Gary King. "A Method of Automated Nonparametric Content Analysis for Social Science" *AJPS*. 54 (2010): 229-247

1. Downloaded & analyzed all English-language blog posts every day. (We learned: The university is not a research, not production, environment!)
2. Commercialized in 2008:

crimson hexagon ⇝ 

⇝

Fast Company Names Crimson Hexagon Number Seven on "The 10 Most Innovative Companies in Web" List Leading Social Intelligence Firm Recognized For Revolutionary Measurement of Consumer Opinions in Social Media

Published: Wednesday, 16 Mar 2011 | 9:20 AM ET Text Size

CAMBRIDGE, Mass., Mar 16, 2011 (BUSINESS WIRE) -- Fast Company named

3. CH collects *all* social media posts, runs huge servers with our methods
4. Crimson Hexagon Academic Grant Program to be announced soon (I.e., easy to do what I'll describe today)

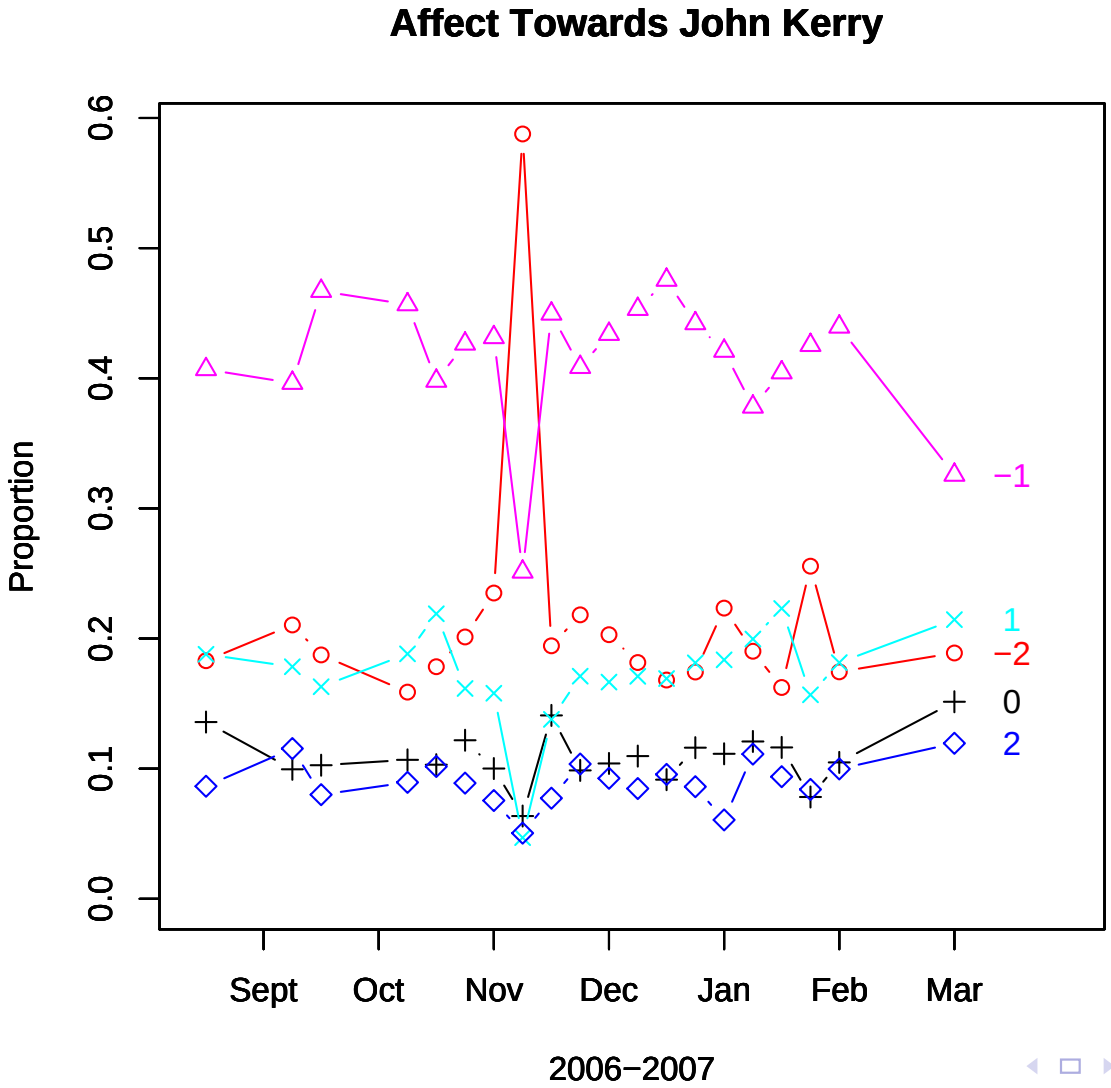
Example: Reactions to John Kerry's Botched Joke

Example: Reactions to John Kerry's Botched Joke

You know, education — if you make the most of it . . . you can do well. If you don't, you get stuck in Iraq.

Example: Reactions to John Kerry's Botched Joke

You know, education — if you make the most of it . . . you can do well. If you don't, you get stuck in Iraq.



Data and Quantities of Interest

Data and Quantities of Interest

- Input Data:

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam;

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama;

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime; support for different solutions to the Euro $ crisis)

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime; support for different solutions to the Euro $ crisis)
- Estimation

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime; support for different solutions to the Euro $ crisis)
- Estimation
  - Classifications add up to proportions only if accurate

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime; support for different solutions to the Euro $ crisis)
- Estimation
  - Classifications add up to proportions only if accurate
  - High classification accuracy $\not\Rightarrow$ unbiased category proportions

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime; support for different solutions to the Euro $ crisis)
- Estimation
  - Classifications add up to proportions only if accurate
  - High classification accuracy $\not\Rightarrow$ unbiased category proportions
  - 70% classification accuracy is high $\Rightarrow$ disaster for category proportions

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime; support for different solutions to the Euro $ crisis)
- Estimation
  - Classifications add up to proportions only if accurate
  - High classification accuracy $\not\Rightarrow$ unbiased category proportions
  - 70% classification accuracy is high $\Rightarrow$ disaster for category proportions
  - New methodology $\rightsquigarrow$ unbiased category proportions

Data and Quantities of Interest

- Input Data:
  - All social media posts (or other documents)
  - Categories (e.g., posts about US candidates: extremely negative, negative, neutral, positive, extremely positive, no opinion, not a blog)
  - Example documents from each category
- Quantities of interest
  - Computer science: individual document classification (spam filters, Google searches)
  - Social Science: category proportions (% of email which is spam; % negative comments about Obama; % of Egyptian posts supporting the regime; support for different solutions to the Euro $ crisis)
- Estimation
  - Classifications add up to proportions only if accurate
  - High classification accuracy $\not\Rightarrow$ unbiased category proportions
  - 70% classification accuracy is high $\Rightarrow$ disaster for category proportions
  - New methodology $\rightsquigarrow$ unbiased category proportions, (even when classification accuracy is low)

What Else Can We do With this?

- You choose:

What Else Can We do With this?

- You choose:
  - Data: country, documents, language

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.
  - (often pre-censorship)

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.
  - (often pre-censorship)
- You provide: example documents for each category

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.
  - (often pre-censorship)
- You provide: example documents for each category
- Results: Highly accurate category proportions over time

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.
  - (often pre-censorship)
- You provide: example documents for each category
- Results: Highly accurate category proportions over time
- Qualifications:

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.
  - (often pre-censorship)
- You provide: example documents for each category
- Results: Highly accurate category proportions over time
- Qualifications:
  - Opinion not sampled randomly; but no pop quizzes about unknown subjects

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.
  - (often pre-censorship)
- You provide: example documents for each category
- Results: Highly accurate category proportions over time
- Qualifications:
  - Opinion not sampled randomly; but no pop quizzes about unknown subjects
  - Measures the ongoing conversation: the classical notion of "activated public opinion"

What Else Can We do With this?

- You choose:
  - Data: country, documents, language
  - Categories: based on sentiment, topics, people, events, etc.
  - (often pre-censorship)
- You provide: example documents for each category
- Results: Highly accurate category proportions over time
- Qualifications:
  - Opinion not sampled randomly; but no pop quizzes about unknown subjects
  - Measures the ongoing conversation: the classical notion of "activated public opinion"
- Potential academic applications: very widespread

Some New Data Types

Some New Data Types

1. Unstructured text: emails (1 LOC every 10 minutes), speeches, government reports, blogs, social media updates, web pages, newspapers, scholarly literature

Some New Data Types

1. Unstructured text: emails (1 LOC every 10 minutes), speeches, government reports, blogs, social media updates, web pages, newspapers, scholarly literature
2. Commercial activity: credit cards, sales data, and real estate transactions, product RFIDs

Some New Data Types

1. Unstructured text: emails (1 LOC every 10 minutes), speeches, government reports, blogs, social media updates, web pages, newspapers, scholarly literature
2. Commercial activity: credit cards, sales data, and real estate transactions, product RFIDs
3. Geographic location: cell phones, Fastlane or EZPass transponders, garage cameras

Some New Data Types

1. Unstructured text: emails (1 LOC every 10 minutes), speeches, government reports, blogs, social media updates, web pages, newspapers, scholarly literature
2. Commercial activity: credit cards, sales data, and real estate transactions, product RFIDs
3. Geographic location: cell phones, Fastlane or EZPass transponders, garage cameras
4. Health information: digital medical records, hospital admittances, google/MS health, and accelerometers and other devices being included in cell phones

Some New Data Types

1. Unstructured text: emails (1 LOC every 10 minutes), speeches, government reports, blogs, social media updates, web pages, newspapers, scholarly literature
2. Commercial activity: credit cards, sales data, and real estate transactions, product RFIDs
3. Geographic location: cell phones, Fastlane or EZPass transponders, garage cameras
4. Health information: digital medical records, hospital admittances, google/MS health, and accelerometers and other devices being included in cell phones
5. Biological sciences: effectively becoming social sciences as genomics, proteomics, metabolomics, and brain imaging produce huge numbers of *person-level variables*.

Some New Data Types

1. Unstructured text: emails (1 LOC every 10 minutes), speeches, government reports, blogs, social media updates, web pages, newspapers, scholarly literature
2. Commercial activity: credit cards, sales data, and real estate transactions, product RFIDs
3. Geographic location: cell phones, Fastlane or EZPass transponders, garage cameras
4. Health information: digital medical records, hospital admittances, google/MS health, and accelerometers and other devices being included in cell phones
5. Biological sciences: effectively becoming social sciences as genomics, proteomics, metabolomics, and brain imaging produce huge numbers of *person-level variables*.
6. Satellite imagery: increasing in scope, resolution, and availability.

Some New Data Types

1. Unstructured text: emails (1 LOC every 10 minutes), speeches, government reports, blogs, social media updates, web pages, newspapers, scholarly literature
2. Commercial activity: credit cards, sales data, and real estate transactions, product RFIDs
3. Geographic location: cell phones, Fastlane or EZPass transponders, garage cameras
4. Health information: digital medical records, hospital admittances, google/MS health, and accelerometers and other devices being included in cell phones
5. Biological sciences: effectively becoming social sciences as genomics, proteomics, metabolomics, and brain imaging produce huge numbers of *person-level variables*.
6. Satellite imagery: increasing in scope, resolution, and availability.
7. Electoral activity: ballot images, precinct-level results, individual-level registration, primary participation, and campaign contributions

Some More New Data Examples

Some More New Data Examples

8. Social media: facebook, twitter, social bookmarking, blog comments, product reviews, virtual worlds, game behavior, crowd sourcing

Some More New Data Examples

8. Social media: facebook, twitter, social bookmarking, blog comments, product reviews, virtual worlds, game behavior, crowd sourcing
9. Web surfing artifacts: clicks, searches, and advertising clickthroughs. (Google collects 1 petabyte/72 minutes on human behavior!)

Some More New Data Examples

8. Social media: facebook, twitter, social bookmarking, blog comments, product reviews, virtual worlds, game behavior, crowd sourcing
9. Web surfing artifacts: clicks, searches, and advertising clickthroughs. (Google collects 1 petabyte/72 minutes on human behavior!)
10. Multiplayer web games and virtual worlds: Billions of highly controlled experiments on human behavior

Some More New Data Examples

8. Social media: facebook, twitter, social bookmarking, blog comments, product reviews, virtual worlds, game behavior, crowd sourcing
9. Web surfing artifacts: clicks, searches, and advertising clickthroughs. (Google collects 1 petabyte/72 minutes on human behavior!)
10. Multiplayer web games and virtual worlds: Billions of highly controlled experiments on human behavior
11. Government bureaucracies: moving from paper to electronic data bases, increasing availability

Some More New Data Examples

8. Social media: facebook, twitter, social bookmarking, blog comments, product reviews, virtual worlds, game behavior, crowd sourcing
9. Web surfing artifacts: clicks, searches, and advertising clickthroughs. (Google collects 1 petabyte/72 minutes on human behavior!)
10. Multiplayer web games and virtual worlds: Billions of highly controlled experiments on human behavior
11. Government bureaucracies: moving from paper to electronic data bases, increasing availability
12. Governmental policies: requiring more data collection, such e.g., "No Child Left Behind Act"; allowing randomized policy experiments; Obama pushing data distribution

Some More New Data Examples

8. Social media: facebook, twitter, social bookmarking, blog comments, product reviews, virtual worlds, game behavior, crowd sourcing
9. Web surfing artifacts: clicks, searches, and advertising clickthroughs. (Google collects 1 petabyte/72 minutes on human behavior!)
10. Multiplayer web games and virtual worlds: Billions of highly controlled experiments on human behavior
11. Government bureaucracies: moving from paper to electronic data bases, increasing availability
12. Governmental policies: requiring more data collection, such e.g., "No Child Left Behind Act"; allowing randomized policy experiments; Obama pushing data distribution
13. Scholarly data: the replication movement in academia, led in part by political science, is massively increasing data sharing

Enormous Emerging Opportunities for Social Scientists

Enormous Emerging Opportunities for Social Scientists

- For the first time: technologies, policies, data, and methods are making it feasible to attack some of the most vexing problems that afflict human society

Enormous Emerging Opportunities for Social Scientists

- For the first time: technologies, policies, data, and methods are making it feasible to attack some of the most vexing problems that afflict human society
- A massive change from studying problems to understanding and solving problems

Enormous Emerging Opportunities for Social Scientists

- For the first time: technologies, policies, data, and methods are making it feasible to attack some of the most vexing problems that afflict human society
- A massive change from studying problems to understanding and solving problems
- And then there's you & me:

Enormous Emerging Opportunities for Social Scientists

- For the first time: technologies, policies, data, and methods are making it feasible to attack some of the most vexing problems that afflict human society
- A massive change from studying problems to understanding and solving problems
- And then there's you & me:
  - In legislatures, courts, academic departments, . . . , change comes from replacement not conversion

Enormous Emerging Opportunities for Social Scientists

- For the first time: technologies, policies, data, and methods are making it feasible to attack some of the most vexing problems that afflict human society
- A massive change from studying problems to understanding and solving problems
- And then there's you & me:
  - In legislatures, courts, academic departments, . . . , change comes from replacement not conversion
  - Will we wait to be replaced? or put in the effort to convert and learn how to use the new information?

For more information

http://GKing.Harvard.edu