The Next Big [Social Science] Thing

Gary King[1]

Institute for Quantitative Social Science
Harvard University

National Academy of Sciences, 3/4/2016

[1] GaryKing.org

Big Data

Big Data is Not About the Data!

Big Data is Not About the Data!

- The most visible changes:

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes:

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
    - Improving scientific inference:

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
    - Improving scientific inference: using facts we know to learn about facts we don't know

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics,

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for:

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for: new methods to unlock secrets in new types of data:

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for: new methods to unlock secrets in new types of data:
  1. The changing evidence base

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for: new methods to unlock secrets in new types of data:
  1. The changing evidence base
  2. Data from outside organizations

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for: new methods to unlock secrets in new types of data:
  1. The changing evidence base
  2. Data from outside organizations
  3. Evidence-based public policy

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for: new methods to unlock secrets in new types of data:
  1. The changing evidence base
  2. Data from outside organizations
  3. Evidence-based public policy
  4. Meta-science

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for: new methods to unlock secrets in new types of data:
  1. The changing evidence base
  2. Data from outside organizations
  3. Evidence-based public policy
  4. Meta-science
  5. Merging science, technology, and social science

Big Data is Not About the Data!

- The most visible changes: new forms of big data (telescopes, microscopes, scanners, etc.)
- The most important changes: the methods that make the data actionable.
  - Improving scientific inference: using facts we know to learn about facts we don't know
  - The revolution is about the analytics, the methods of inference
  - The methods make *old data* and *new instruments* actionable
  - More data alone only makes research harder
- Look for: new methods to unlock secrets in new types of data:
  1. The changing evidence base
  2. Data from outside organizations
  3. Evidence-based public policy
  4. Meta-science
  5. Merging science, technology, and social science

  ...with examples from my research

1. The Evidence Base of Social Science Research

1. The Evidence Base of Social Science Research

The Last 50 Years:

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to...

- Advances in social science statistics

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to...

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science:

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms;

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries;

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks,

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns,

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health,

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing,

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics,

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to...

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy; & many others

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy; & many others
- Look for:

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy; & many others
- Look for:$\rightsquigarrow$ impact

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy; & many others
- Look for: $\rightsquigarrow$ impact $\rightsquigarrow$ data

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy; & many others
- Look for: $\rightsquigarrow$ impact $\rightsquigarrow$ data $\rightsquigarrow$ methods

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to. . .

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy; & many others
- Look for: $\rightsquigarrow$ impact $\rightsquigarrow$ data $\rightsquigarrow$ methods $\rightsquigarrow$ insights

1. The Evidence Base of Social Science Research

The Last 50 Years:

- Survey research
- Aggregate government statistics
- One-off studies of individual people, places, or events

The Next 50 Years: Additional data sources due to...

- Advances in social science statistics
- Shrinking computers & the growing Internet: data everywhere
- The replication movement: data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- *The march of quantification*: through academia, the professions, government, & commerce
- Real world impact of social science: transformed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis; impacted crime and policing, economics, even sports; set standards for evaluating public policy; & many others
- Look for: $\rightsquigarrow$ impact $\rightsquigarrow$ data $\rightsquigarrow$ methods $\rightsquigarrow$ insights $\rightsquigarrow$ data...

2. Corporate, Nonprofit, Government Data

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc.,

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck:

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook,

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook, Unix,

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook, Unix, gmail,

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook, Unix, gmail, snapchat,

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook, Unix, gmail, snapchat, Google,

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook, Unix, gmail, snapchat, Google, Twitter, etc.

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook, Unix, gmail, snapchat, Google, Twitter, etc.
- Look for:

2. Corporate, Nonprofit, Government Data

- At one time, almost all research data was inside the university
- Now, most data is outside, in corporations, governments, etc., controlled by our research subjects!
- We arrange access $\rightsquigarrow$ data becomes more informative $\rightsquigarrow$ the public becomes more concerned about privacy $\rightsquigarrow$ firms worry about research hurting business $\rightsquigarrow$ we obtain new access under new rules $\rightsquigarrow \cdots$
- Disruption will not stop. To upend the software-based tech industry requires only a few people, some creativity, and luck: Facebook, Unix, gmail, snapchat, Google, Twitter, etc.
- Look for: unique opportunities from fast changing research partnerships with groups outside the university

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)
  - Sentiment analysis via word counts (social media companies)

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)
  - Sentiment analysis via word counts (social media companies)
- Unrelated substantive problems, same analytics solution:

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)
  - Sentiment analysis via word counts (social media companies)
- Unrelated substantive problems, same analytics solution:
  - Key to both methods: *classifying* (deaths, social media posts)

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)
  - Sentiment analysis via word counts (social media companies)
- Unrelated substantive problems, same analytics solution:
  - Key to both methods: *classifying* (deaths, social media posts)
  - Key to both goals: *estimating %'s*

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)
  - Sentiment analysis via word counts (social media companies)
- Unrelated substantive problems, same analytics solution:
  - Key to both methods: *classifying* (deaths, social media posts)
  - Key to both goals: *estimating %'s*
- Modern Data Analytics: New method led to:

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)
  - Sentiment analysis via word counts (social media companies)
- Unrelated substantive problems, same analytics solution:
  - Key to both methods: *classifying* (deaths, social media posts)
  - Key to both goals: *estimating %'s*
- Modern Data Analytics: New method led to:
  1.

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

E.g.: How to Read a Trillion Social Media Posts & Classify Deaths without Physicians

- Examples of Bad Analytics:
  - Physicians' "Verbal Autopsy" analysis (from WHO)
  - Sentiment analysis via word counts (social media companies)
- Unrelated substantive problems, same analytics solution:
  - Key to both methods: *classifying* (deaths, social media posts)
  - Key to both goals: *estimating %'s*
- Modern Data Analytics: New method led to:

1. 

⇝


⇝

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

2. Worldwide cause-of-death estimates for

3. Evidence-Based Public Policy

3. Evidence-Based Public Policy

- Definition of "government policies":

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)
- Most large scale public policy evaluations fail:

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)
- Most large scale public policy evaluations fail:
  $\rightsquigarrow$ New "politically robust" methods

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)
- Most large scale public policy evaluations fail:
  $\rightsquigarrow$ New "politically robust" methods
- White House executive orders requiring openness

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)
- Most large scale public policy evaluations fail:
  $\rightsquigarrow$ New "politically robust" methods
- White House executive orders requiring openness
- Local governments competing to share data

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)
- Most large scale public policy evaluations fail:
  $\rightsquigarrow$ New "politically robust" methods
- White House executive orders requiring openness
- Local governments competing to share data
- Rise of systematic observational analyses (e.g., Social Security evaluation)

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)
- Most large scale public policy evaluations fail:
  $\rightsquigarrow$ New "politically robust" methods
- White House executive orders requiring openness
- Local governments competing to share data
- Rise of systematic observational analyses (e.g., Social Security evaluation)
- Look for:

3. Evidence-Based Public Policy

- Definition of "government policies":
  Giant experiments without control groups
- New movement to allow real experiments in government (e.g., Seguro Popular)
- Most large scale public policy evaluations fail:
  $\rightsquigarrow$ New "politically robust" methods
- White House executive orders requiring openness
- Local governments competing to share data
- Rise of systematic observational analyses (e.g., Social Security evaluation)
- Look for: new research (and new methods) from partnerships between researchers and governments

E.g.: Bias in Social Security Administration Forecasts

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures;

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods:

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed;

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative;

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results:

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000;

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts:

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
- New customized analytics we developed:

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
- New customized analytics we developed:
  - Logical consistency (e.g., older people have higher mortality)

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
- New customized analytics we developed:
  - Logical consistency (e.g., older people have higher mortality)
  - Far more accurate forecasts

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
- New customized analytics we developed:
  - Logical consistency (e.g., older people have higher mortality)
  - Far more accurate forecasts
  - $\rightsquigarrow$ Trust fund needs $>$ \$800 million more than SSA thought

E.g.: Bias in Social Security Administration Forecasts

- Social Security: single largest government program; lifted a whole generation out of poverty; extremely popular
- Forecasts: used for programs comprising $> 50\%$ of the US expenditures; e.g., if retirees draw benefits longer than expected, the Trust Fund runs out
- First evaluation of SSA forecasts in 85 years:
  - Methods: little changed; mostly qualitative; a time when we've learned more about forecasting than at any time in history
  - Results: unbiased until 2000; systematically biased after
  - Actuaries hunkered down, insulated themselves, refused to budge when Democrats & Republicans pushed hard for changes
  - In the process, they also insulated themselves from the facts: Especially since 2000, Americans started living unexpectedly longer lives (due to statins, early cancer detection, etc.)
- New customized analytics we developed:
  - Logical consistency (e.g., older people have higher mortality)
  - Far more accurate forecasts
  - $\rightsquigarrow$ Trust fund needs $>$ \$800 million more than SSA thought
  - Many other applications to different types of forecasts

The End of The Quantitative-Qualitative Divide

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, new methods make unstructured text, video, audio actionable

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, new methods make unstructured text, video, audio actionable
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins.

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, new methods make unstructured text, video, audio actionable
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins.
- But: There's always qualitative information that hasn't been quantified

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, new methods make unstructured text, video, audio actionable
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins.
- But: There's always qualitative information that hasn't been quantified
- Look for:

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, new methods make unstructured text, video, audio actionable
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins.
- But: There's always qualitative information that hasn't been quantified
- Look for: new methodological solutions to the quant-qual divide.

4. Meta-Science

4. Meta-Science

- Meta-science: must follow the rules of science

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:
  - *Dataverse*: largest collection of social science research data

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:
  - *Dataverse*: largest collection of social science research data
  - solve political problems technologically: Breaks the choice between individual credit and permanent archiving

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:
  - *Dataverse*: largest collection of social science research data
  - solve political problems technologically: Breaks the choice between individual credit and permanent archiving
  - dataverse automates the job of the archivist

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:
  - *Dataverse*: largest collection of social science research data
  - solve political problems technologically: Breaks the choice between individual credit and permanent archiving
  - dataverse automates the job of the archivist
  - a complete archive on your site, with no installations

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:
  - *Dataverse*: largest collection of social science research data
  - solve political problems technologically: Breaks the choice between individual credit and permanent archiving
  - dataverse automates the job of the archivist
  - a complete archive on your site, with no installations
  - Huge array of analytics and behavioral incentives feed back and improve science

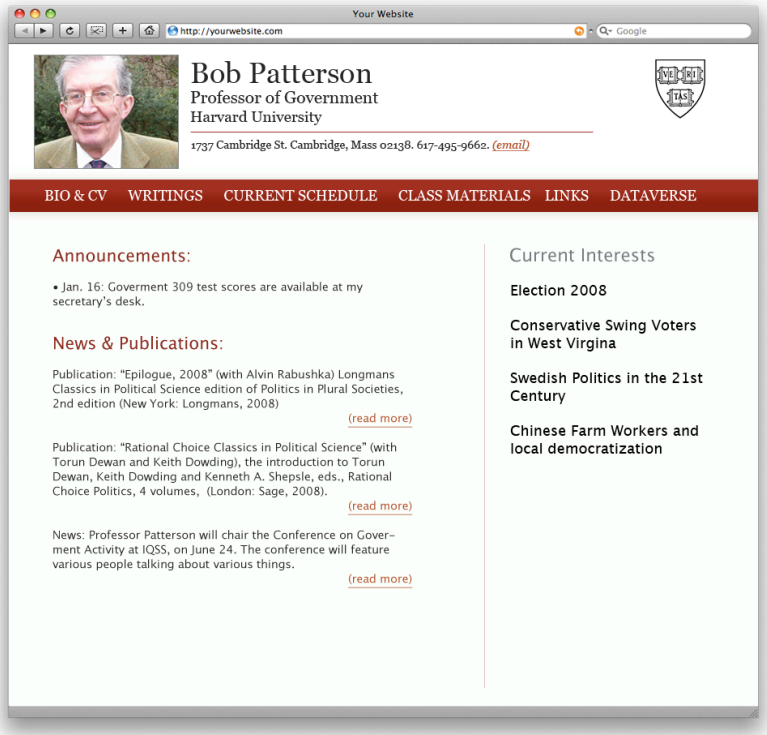
4. Meta-Science

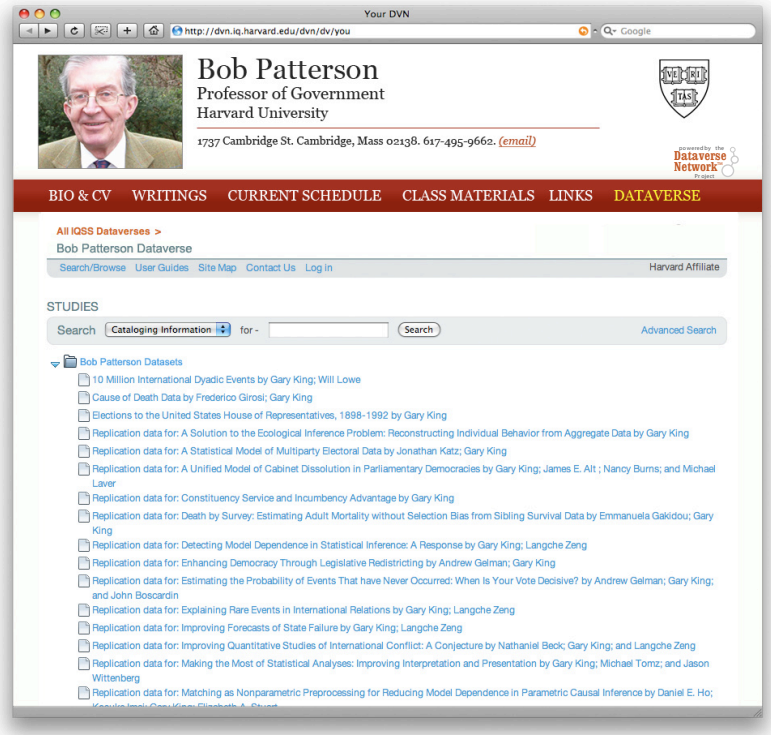
- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:
  - *Dataverse*: largest collection of social science research data
  - solve political problems technologically: Breaks the choice between individual credit and permanent archiving
  - dataverse automates the job of the archivist
  - a complete archive on your site, with no installations
  - Huge array of analytics and behavioral incentives feed back and improve science
- Look for:

4. Meta-Science

- Meta-science: must follow the rules of science
  - For example: [under embargo at *Science*]
- Meta-science: a subfield of the social sciences. E.g.:
  - *Dataverse*: largest collection of social science research data
  - solve political problems technologically: Breaks the choice between individual credit and permanent archiving
  - dataverse automates the job of the archivist
  - a complete archive on your site, with no installations
  - Huge array of analytics and behavioral incentives feed back and improve science
- Look for: huge potential in applying social science insights to improve science

Author Dataverse

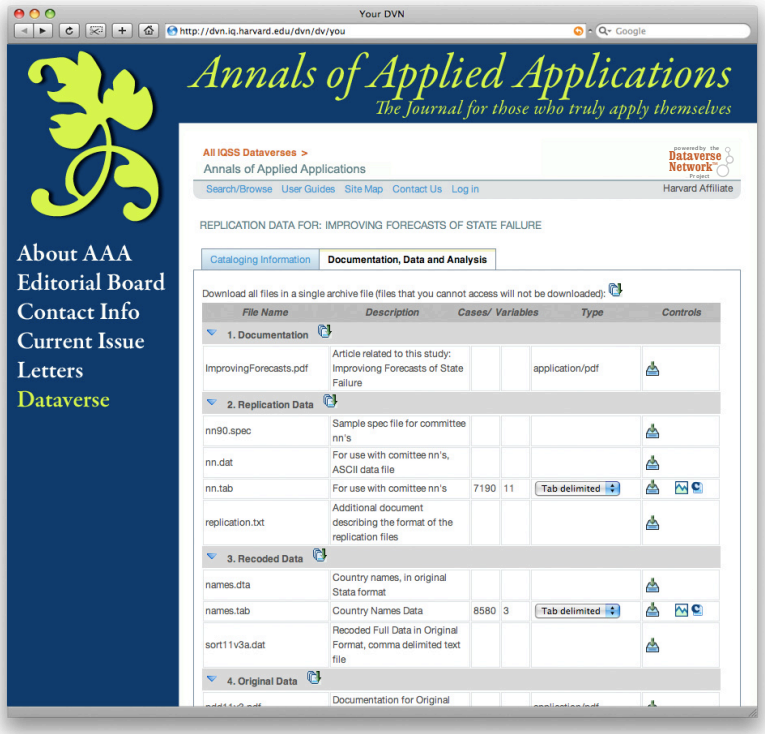




Your web site

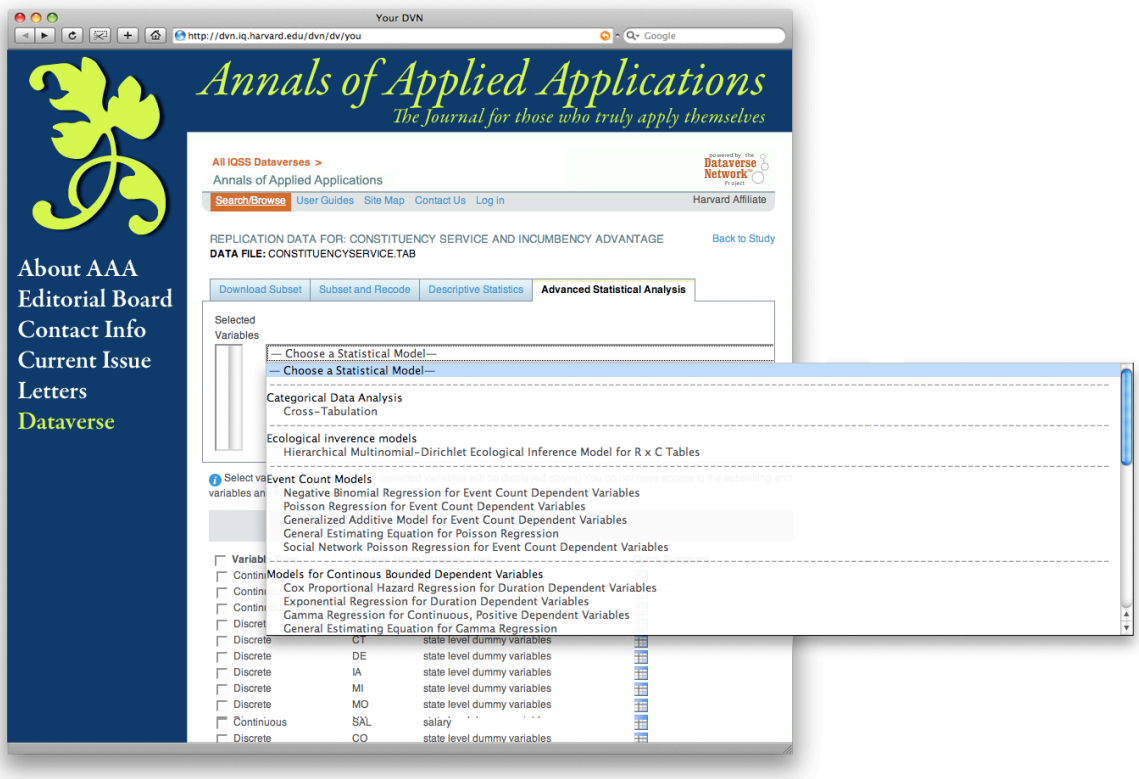
Your dataverse branded as your web site but served by the Dataverse Network, therefore requiring no local installation and providing an enormous array of services

Journal Dataverse



Journal Dataverse

Journal Dataverse



5. Merging Science, Technology, and Social Science

5. Merging Science, Technology, and Social Science

- Social Science:

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges,

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
    - Solved biological problems:

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
    - Solved biological problems: smoking,

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
    - Solved biological problems: smoking, obesity,

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
    - Solved biological problems: smoking, obesity, medical advice noncompliance

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems:

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming,

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity,

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity, teaching evolution

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity, teaching evolution
  - Solved problems in the science of science:

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity, teaching evolution
  - Solved problems in the science of science: replication and data sharing,

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity, teaching evolution
  - Solved problems in the science of science: replication and data sharing, science funding

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity, teaching evolution
  - Solved problems in the science of science: replication and data sharing, science funding
- Genomics, proteomics, brain scanning, etc., all producing huge numbers of person-level variables

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity, teaching evolution
  - Solved problems in the science of science: replication and data sharing, science funding
- Genomics, proteomics, brain scanning, etc., all producing huge numbers of person-level variables
- Look for:

5. Merging Science, Technology, and Social Science

- Social Science: understanding or solving society's challenges, through person or group-level analyses
- Discovering the underlying scientific cause sometimes doesn't solve the problem:
  - Solved biological problems: smoking, obesity, medical advice noncompliance
  - Solved physical science problems: human generated global warming, biodiversity, teaching evolution
  - Solved problems in the science of science: replication and data sharing, science funding
- Genomics, proteomics, brain scanning, etc., all producing huge numbers of person-level variables
- Look for: new science–social science partnerships, and merging of methods

E.g.: Humans are Horrible at Thinking of Keywords

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment:

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples:

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional,

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage,

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama,

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled:

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled: 8

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled: 8
- Unique keywords recalled by 43 undergrads:

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled: 8
- Unique keywords recalled by 43 undergrads: 149

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled: 8
- Unique keywords recalled by 43 undergrads: 149
- Keywords 42 of 43 failed to recall:

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled: 8
- Unique keywords recalled by 43 undergrads: 149
- Keywords 42 of 43 failed to recall: 98 (66%)

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled: 8
- Unique keywords recalled by 43 undergrads: 149
- Keywords 42 of 43 failed to recall: 98 (66%)
- $\rightsquigarrow$ Humans recognize keywords well, recall them poorly

E.g.: Humans are Horrible at Thinking of Keywords

- An experiment: "We have 10,000 twitter posts, each containing the word 'healthcare', from the time period surrounding the Supreme Court decision on Obamacare. Please list any keywords which come to mind that will select posts in this set related to Obamacare and will not select posts unrelated to Obama care."
- Examples: unconstitutional, coverage, obama, ACA...
- Median keywords recalled: 8
- Unique keywords recalled by 43 undergrads: 149
- Keywords 42 of 43 failed to recall: 98 (66%)
- $\rightsquigarrow$ Humans recognize keywords well, recall them poorly
- Thresher: New technology to discover the right keywords

E.g.: Following Conversations that Hide in Plain Sight

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1:

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1:

自由

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1:

自由 “Freedom”

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1:

自由 “Freedom”

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1:

自由 “Freedom”

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1:

自由 “Freedom”

目由 “Eye field”

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1:

自由 “Freedom” 

目田 “Eye field” (nonsensical)

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目由 "Eye field" (nonsensical)

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目由 "Eye field" (nonsensical)

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目由 "Eye field" (nonsensical)

Example Substitution 2:

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由	"Freedom"	
目田	"Eye field"	(nonsensical)

Example Substitution 2:

和谐

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2:

和谐 "Harmonious [Society]" (official slogan)

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2:

和谐 "Harmonious [Society]" (official slogan) 

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2:

和谐 "Harmonious [Society]" (official slogan) 

河蟹

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2:

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab"

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2:

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation;

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task:

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search,

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search, (2) Government and industry analyst's job,

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search, (2) Government and industry analyst's job, (3) language drift (#BostonBombings $\rightsquigarrow$ #BostonStrong),

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search, (2) Government and industry analyst's job, (3) language drift (#BostonBombings $\rightsquigarrow$ #BostonStrong), (4) Child pornographers,

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search, (2) Government and industry analyst's job, (3) language drift (#BostonBombings $\rightsquigarrow$ #BostonStrong), (4) Child pornographers, (5) Look-alike modeling,

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search, (2) Government and industry analyst's job, (3) language drift (#BostonBombings $\rightsquigarrow$ #BostonStrong), (4) Child pornographers, (5) Look-alike modeling, (6) Starting point for other automated text methods,

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search, (2) Government and industry analyst's job, (3) language drift (#BostonBombings $\rightsquigarrow$ #BostonStrong), (4) Child pornographers, (5) Look-alike modeling, (6) Starting point for other automated text methods, (7) Infinitely improvable classification,

E.g.: Following Conversations that Hide in Plain Sight

Example Substitution 1: Homograph

自由 "Freedom" 

目田 "Eye field" (nonsensical)

Example Substitution 2: Homophone (sound like "hexie")

和谐 "Harmonious [Society]" (official slogan) 

河蟹 "River crab" (irrelevant)

They can't follow the conversation; Thresher can.

The same task: (1) Long tail search, (2) Government and industry analyst's job, (3) language drift (#BostonBombings $\rightsquigarrow$ #BostonStrong), (4) Child pornographers, (5) Look-alike modeling, (6) Starting point for other automated text methods, (7) Infinitely improvable classification, eDiscovery

E.g. 2: Reverse Engineering Censorship in China

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- Everyone knows the Goal:

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- Everyone knows the Goal:
  Stop criticism and protest about the state,
  its leaders, and their policies

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. Stop criticism of the state

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. Stop criticism of the state
  2. Stop collective action

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx$ 13% censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests;

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx$ 13% censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests; new peace treaties;

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx$ 13% censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state,~~
  ~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests; new peace treaties; emerging scandals

E.g. 2: Reverse Engineering Censorship in China

- Previous approach: watch a few posts; see what's removed
- Data: We get posts before the Chinese censor them
- $\approx 13\%$ censored overall
- ~~Everyone knows the Goal:~~
  ~~Stop criticism and protest about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  1. ~~Stop criticism of the state~~ *Wrong*
  2. Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests; new peace treaties; emerging scandals
    - Disagreements between central and local leaders

For more information

GaryKing.org

Institute for Quantitative Social Science
Harvard University