Big Data is Not About the Data!

$\mathsf{Gary}\ \mathsf{King}^1$

Institute for Quantitative Social Science Harvard University

Talk at the History of Evidence class, Harvard Law School, 11/17/2014

¹GaryKing.org

- Data:
 - easy to come by; often a free byproduct of IT improvements

- Data:
 - easy to come by; often a free byproduct of IT improvements
 - becoming commoditized

- Data:
 - easy to come by; often a free byproduct of IT improvements
 - becoming commoditized
 - Ignore it & every institution will have more every year

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases
- Where the Value is: the Analytics

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases
- Where the Value is: the Analytics
 - Output can be highly customized

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases
- Where the Value is: the Analytics
 - Output can be highly customized
 - Moore's Law (doubling speed/power every 18 months)

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases
- Where the Value is: the Analytics
 - Output can be highly customized
 - Moore's Law (doubling speed/power every 18 months)
 - v. Our Students (1000x speed increase in 1 day)

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases
- Where the Value is: the Analytics
 - Output can be highly customized
 - Moore's Law (doubling speed/power every 18 months)
 v. Our Students (1000x speed increase in 1 day)
 - \$2M computer v. 2 hours of algorithm design

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases
- Where the Value is: the Analytics
 - Output can be highly customized
 - Moore's Law (doubling speed/power every 18 months)
 v. Our Students (1000x speed increase in 1 day)
 - \$2M computer v. 2 hours of algorithm design
 - Low cost; little infrastructure; mostly human capital needed

- easy to come by; often a free byproduct of IT improvements
- becoming commoditized
- Ignore it & every institution will have more every year
- With a bit of effort: huge data production increases
- Where the Value is: the Analytics
 - Output can be highly customized
 - Moore's Law (doubling speed/power every 18 months)
 v. Our Students (1000x speed increase in 1 day)
 - \$2M computer v. 2 hours of algorithm design
 - Low cost; little infrastructure; mostly human capital needed
 - Innovative analytics: enormously better than off-the-shelf

• Opinions of activists:

• Opinions of activists: A few thousand interviews

 Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise:

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week?

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts:

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends"

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" → continuous record of phone calls, emails, text messages, bluetooth, social media connections, address books

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" → continuous record of phone calls, emails, text messages, bluetooth, social media connections, address books
- Economic development in developing countries:

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" → continuous record of phone calls, emails, text messages, bluetooth, social media connections, address books
- Economic development in developing countries: Dubious or nonexistent governmental statistics

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" → continuous record of phone calls, emails, text messages, bluetooth, social media connections, address books
- Economic development in developing countries: Dubious or nonexistent governmental statistics → satellite images of human-generated light at night, road networks, other infrastructure

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" → continuous record of phone calls, emails, text messages, bluetooth, social media connections, address books
- Economic development in developing countries: Dubious or nonexistent governmental statistics → satellite images of human-generated light at night, road networks, other infrastructure
- Many, many, more...

- Opinions of activists: A few thousand interviews → billions of political opinions in social media posts (1B every 2 Days)
- Exercise: A survey: "How many times did you exercise last week? → 500K people carrying cell phones with accelerometers
- Social contacts: A survey: "Please tell me your 5 best friends" → continuous record of phone calls, emails, text messages, bluetooth, social media connections, address books
- Economic development in developing countries: Dubious or nonexistent governmental statistics → satellite images of human-generated light at night, road networks, other infrastructure
- Many, many, more...
- In each: without new analytics, the data are useless

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

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

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information; need help
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, starting to analyze unstructured text, video, audio as data

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information; need help
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, starting to analyze unstructured text, video, audio as data
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information; need help
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, starting to analyze unstructured text, video, audio as data
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins
- Moral of the story:

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information; need help
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, starting to analyze unstructured text, video, audio as data
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins
- Moral of the story:
 - Fully human is inadequate

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information; need help
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, starting to analyze unstructured text, video, audio as data
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins
- Moral of the story:
 - Fully human is inadequate
 - Fully automated fails

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information; need help
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, starting to analyze unstructured text, video, audio as data
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins
- Moral of the story:
 - Fully human is inadequate
 - Fully automated fails
 - We need computer assisted, human controlled technology

The End of The Quantitative-Qualitative Divide

- The Quant-Qual divide exists in *every* field.
- Qualitative researchers: overwhelmed by information; need help
- Quantitative researchers: recognize the huge amounts of information in qualitative analyses, starting to analyze unstructured text, video, audio as data
- Expert-vs-analytics contests: Whenever enough information is quantified, a right answer exists, and good analytics are applied: analytics wins
- Moral of the story:
 - Fully human is inadequate
 - Fully automated fails
 - We need computer assisted, human controlled technology
 - (Technically correct, & politically much easier)

• Examples of Bad Analytics:

- Examples of Bad Analytics:
 - Physicians' "Verbal Autopsy" analysis

- Examples of Bad Analytics:
 - Physicians' "Verbal Autopsy" analysis
 - Sentiment analysis via word counts

- Examples of Bad Analytics:
 - Physicians' "Verbal Autopsy" analysis
 - Sentiment analysis via word counts
- Different problems, Same Analytics Solution:

- Examples of Bad Analytics:
 - Physicians' "Verbal Autopsy" analysis
 - Sentiment analysis via word counts
- Different problems, Same Analytics Solution:
 - Key to both methods: *classifying* (deaths, social media posts)

- Examples of Bad Analytics:
 - Physicians' "Verbal Autopsy" analysis
 - Sentiment analysis via word counts
- Different problems, Same Analytics Solution:
 - Key to both methods: *classifying* (deaths, social media posts)
 - Key to both goals: estimating %'s

- Examples of Bad Analytics:
 - Physicians' "Verbal Autopsy" analysis
 - Sentiment analysis via word counts
- Different 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:

- Examples of Bad Analytics:
 - Physicians' "Verbal Autopsy" analysis
 - Sentiment analysis via word counts
- Different 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	rT Text Size 🖃 +

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

• Examples of Bad Analytics:

1.

- Physicians' "Verbal Autopsy" analysis
- Sentiment analysis via word counts
- Different 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:



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



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

2. Worldwide cause-of-death estimates for



• Successful: single largest government program; lifted a whole generation out of poverty; extremely popular

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts:

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)
 - Mostly informal (subject to error & political influence)

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)
 - Mostly informal (subject to error & political influence)
 - Forecasts: inaccurate, inconsistent, overly optimistic

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)
 - Mostly informal (subject to error & political influence)
 - Forecasts: inaccurate, inconsistent, overly optimistic
- New customized analytics we developed:

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)
 - Mostly informal (subject to error & political influence)
 - Forecasts: inaccurate, inconsistent, overly optimistic
- New customized analytics we developed:
 - Logical consistency (e.g., older people have higher mortality)

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)
 - Mostly informal (subject to error & political influence)
 - Forecasts: inaccurate, inconsistent, overly optimistic
- New customized analytics we developed:
 - Logical consistency (e.g., older people have higher mortality)
 - More accurate forecasts

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)
 - Mostly informal (subject to error & political influence)
 - Forecasts: inaccurate, inconsistent, overly optimistic
- New customized analytics we developed:
 - Logical consistency (e.g., older people have higher mortality)
 - More accurate forecasts
 - \rightsquigarrow Trust fund needs \approx \$800 billion more than SSA thought

- Successful: single largest government program; lifted a whole generation out of poverty; extremely popular
- Solvency: depends on mortality forecasts: If retirees receive benefits longer than expected, the Trust Fund runs out
- SSA data: little change other than updates for 75 years
- SSA analytics:
 - Few statistical improvements for 75 years
 - Ignore risk factors (smoking, obesity)
 - Mostly informal (subject to error & political influence)
 - Forecasts: inaccurate, inconsistent, overly optimistic
- New customized analytics we developed:
 - Logical consistency (e.g., older people have higher mortality)
 - More accurate forecasts
 - \rightsquigarrow Trust fund needs \approx \$800 billion more than SSA thought
 - Other applications to insurance industry, public health, etc.

Example Substitution 1:

Example Substitution 1:



Example Substitution 1:

自由 "Freedom"

Example Substitution 1:







Example Substitution 1:

由



Example Substitution 1:





"Eye field"



Example Substitution 1:





"Eye field" (nonsensical)

Example Substitution 1: Homograph





"Eye field" (nonsensical)

Example Substitution 1: Homograph





"Eye field" (nonsensical)

Example Substitution 1: Homograph



"Freedom" CENSORED

"Eye field" (nonsensical)

Example Substitution 2:

Example Substitution 1: Homograph



"Freedom" CENSORED

"Eye field" (nonsensical)

Example Substitution 2:

和谐

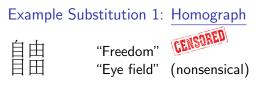




"Freedom" **GENSORED** "Eye field" (nonsensical)

Example Substitution 2:

和谐 "Harmonious [Society]" (official slogan)



Example Substitution 2:

和谐

"Harmonious [Society]" (official slogan)





"Freedom" (nonsensical)

Example Substitution 2:

"Harmonious [Society]" (official slogan)









Example Substitution 2:



"Harmonious [Society]" (official slogan) GENSOLED "River crab"





- "Freedom" (nonsensical)
- Example Substitution 2:

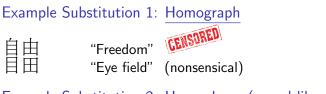


"Harmonious [Society]" (official slogan) CENSORED "River crab" (irrelevant)



和谐河解

"Harmonious [Society]" (official slogan) GENSIAED "River crab" (irrelevant)



和谐河蟹

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

"Harmonious [Society]" (official slogan) CENSTRED "River crab" (irrelevant)

They can't follow the conversation; Our methods can!



和谐河蟹

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

"Harmonious [Society]" (official slogan) **GENSIRED** "River crab" (irrelevant)

They can't follow the conversation; <u>Our methods can!</u> The same task:

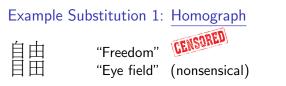


和谐河蟹

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

"Harmonious [Society]" (official slogan) GENSIRED "River crab" (irrelevant)

They can't follow the conversation; <u>Our methods can!</u> The same task: (1) Government and industry analyst's job,

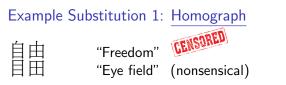


和谐河蟹

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

"Harmonious [Society]" (official slogan) **GENSTRED** "River crab" (irrelevant)

They can't follow the conversation; <u>Our methods can!</u> The same task: (1) Government and industry analyst's job, (2) language drift (#BostonBombings ~> #BostonStrong),



和谐河蟹

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

"Harmonious [Society]" (official slogan) **GENSIRED** "River crab" (irrelevant)

They can't follow the conversation; <u>Our methods can!</u> The same task: (1) Government and industry analyst's job, (2) language drift (#BostonBombings \rightsquigarrow #BostonStrong), (3) Child pornographers,



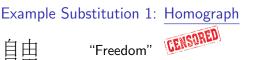
和谐河蟹

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

"Harmonious [Society]" (official slogan) GENSTRED "River crab" (irrelevant)

They can't follow the conversation; Our methods can!

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



和谐河蟹

"Eye field" (nonsensical)

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

"Harmonious [Society]" (official slogan) GENSIRED "River crab" (irrelevant)

They can't follow the conversation; Our methods can!

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

• To understand many documents, humans create categories to represent conceptualization, insight, etc.

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:
 - Unassisted Human Categorization: time consuming; huge efforts trying *not* to innovate!

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:
 - Unassisted Human Categorization: time consuming; huge efforts trying *not* to innovate!
 - Fully Automated "Cluster Analysis": Many widely available, but none work (computers don't know what you want!)

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:
 - Unassisted Human Categorization: time consuming; huge efforts trying *not* to innovate!
 - Fully Automated "Cluster Analysis": Many widely available, but none work (computers don't know what you want!)
- Our alternative: Computer-assisted Categorization

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:
 - Unassisted Human Categorization: time consuming; huge efforts trying *not* to innovate!
 - Fully Automated "Cluster Analysis": Many widely available, but none work (computers don't know what you want!)
- Our alternative: Computer-assisted Categorization
 - You decide what's important, but with help

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:
 - Unassisted Human Categorization: time consuming; huge efforts trying *not* to innovate!
 - Fully Automated "Cluster Analysis": Many widely available, but none work (computers don't know what you want!)
- Our alternative: Computer-assisted Categorization
 - You decide what's important, but with help
 - Invert effort: you innovate; the computer categorizes

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:
 - Unassisted Human Categorization: time consuming; huge efforts trying *not* to innovate!
 - Fully Automated "Cluster Analysis": Many widely available, but none work (computers don't know what you want!)
- Our alternative: Computer-assisted Categorization
 - You decide what's important, but with help
 - Invert effort: you innovate; the computer categorizes
 - Insights: easier, faster, better

- To understand many documents, humans create categories to represent conceptualization, insight, etc.
- Most firms: impose fixed categorizations to tally customer complaints, sort reports, retrieve information
- Bad Analytics:
 - Unassisted Human Categorization: time consuming; huge efforts trying *not* to innovate!
 - Fully Automated "Cluster Analysis": Many widely available, but none work (computers don't know what you want!)
- Our alternative: Computer-assisted Categorization
 - You decide what's important, but with help
 - Invert effort: you innovate; the computer categorizes
 - Insights: easier, faster, better
 - (Lots of technology, but it's behind the scenes)

1. What Members of Congress Do

• Data: 64,000 Senators' press releases

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it?

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship
 - Previous approach: manual effort to see what is taken down

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship
 - Previous approach: manual effort to see what is taken down
 - Data: We get posts before the Chinese censor them

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship
 - Previous approach: manual effort to see what is taken down
 - Data: We get posts before the Chinese censor them
 - We analyzed 11 million posts, about 13% censored

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship
 - Previous approach: manual effort to see what is taken down
 - Data: We get posts before the Chinese censor them
 - We analyzed 11 million posts, about 13% censored
 - Previous understanding: they censor criticisms of the government

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship
 - Previous approach: manual effort to see what is taken down
 - Data: We get posts before the Chinese censor them
 - We analyzed 11 million posts, about 13% censored
 - Previous understanding: they censor criticisms of the government
 - Results:

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship
 - Previous approach: manual effort to see what is taken down
 - Data: We get posts before the Chinese censor them
 - We analyzed 11 million posts, about 13% censored
 - Previous understanding: they censor criticisms of the government
 - Results:
 - Uncensored: criticism of the government

- Data: 64,000 Senators' press releases
- Categorization: (1) advertising, (2) position taking, (3) credit claiming
- New Insight: partisan taunting
 - Joe Wilson during Obama's State of the Union: "You lie!"
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' "
- How common is it? 27% of all Senatorial press releases!
- 2. Reverse Engineering Chinese Censorship
 - Previous approach: manual effort to see what is taken down
 - Data: We get posts before the Chinese censor them
 - We analyzed 11 million posts, about 13% censored
 - Previous understanding: they censor criticisms of the government
 - Results:
 - Uncensored: criticism of the government
 - Censored: attempts at collective action

For more information

GaryKing.org