

# Empowering Social Science to Understand and Ameliorate Major Challenges of Human Society

Gary King<sup>1</sup>

Institute for Quantitative Social Science  
Harvard University

Federal Interagency Conference on Social Science and Big Data, 9/30/2020

---

<sup>1</sup>[GaryKing.org](http://GaryKing.org)

# How to Empower the Social Sciences?

# How to Empower the Social Sciences?



# How to Empower the Social Sciences?

 Connections

 M

 ?

# How to Empower the Social Sciences?

 Connections

 Methods



# How to Empower the Social Sciences?

 Connections

 Methods

 Questions

# How to Empower the Social Sciences?

- ① **Connections:** Sharing data, people, expertise
- ② **Methods**
- ③ **Questions**

# How to Empower the Social Sciences?

- ② **Connections:** Sharing data, people, expertise
- ④ **Methods:** Big data is not about the data!
- ① **Questions**

# How to Empower the Social Sciences?

- ② **Connections:** Sharing data, people, expertise
- ④ **Methods:** Big data is not about the data!
- ① **Questions:** Stay flexible; let data suggest new questions

# How to Empower the Social Sciences?

 Connections

 Methods

 Questions

# How to Empower the Social Sciences?





# Mexican Universal Health Insurance Evaluation

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer:


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer: 


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer:  Commission an independent evaluation
- Our Study


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer:  Commission an independent evaluation
- Our Study
  - Randomized to communities:


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer:  Commission an independent evaluation
- Our Study
  - Randomized to communities: hospitals,


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer:  Commission an independent evaluation
- Our Study
  - Randomized to communities: hospitals, doctors,


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer:  Commission an independent evaluation
- Our Study
  - Randomized to communities: hospitals, doctors, medicines,


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - The government asked: how can one democratically elected government tie the hands of the next?
  - Answer:  Commission an independent evaluation
- Our Study
  - Randomized to communities: hospitals, doctors, medicines, insurance,


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc.


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:**


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!



# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:**


# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** 

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:**  “Politically Robust” Fail-Safe Design: randomization survives political interventions

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:** (CS) Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** (M) “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization



# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:** (CS) Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** (M) “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias,



# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:**  “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence,



# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- **How'd we get to do this?**
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- **Our Study**
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:**  “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs,



# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:**  “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)



# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:**  “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)
- Results

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:**  Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:**  “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)
- Results
  - **Health Care Impoverishment:** drastically reduced

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- **How'd we get to do this?**
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:** (CS) Commission an independent evaluation
- **Our Study**
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** (M) “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)
- **Results**
  - **Health Care Impoverishment:** drastically reduced
  - (?)

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:** (CS) Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** (M) “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)
- Results
  - **Health Care Impoverishment:** drastically reduced
  - (?) States qualify for federal funding by automatically signing up the poor,

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- **How'd we get to do this?**
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:** (CS) Commission an independent evaluation
- **Our Study**
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** (M) “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)
- **Results**
  - **Health Care Impoverishment:** drastically reduced
  - (?) States qualify for federal funding by automatically signing up the poor, but only 66% of the poor were aware they had health care!

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:** (CS) Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** (M) “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)
- Results
  - **Health Care Impoverishment:** drastically reduced
  - (?) States qualify for federal funding by automatically signing up the poor, but only 66% of the poor were aware they had health care! ~> immediately corrected

# Mexican Universal Health Insurance Evaluation

“Seguro Popular”: first health care for 50M Mexicans

- How'd we get to do this?
  - **The government asked:** how can one democratically elected government tie the hands of the next?
  - **Answer:** (CS) Commission an independent evaluation
- Our Study
  - **Randomized to communities:** hospitals, doctors, medicines, insurance, \$, etc. ~> the largest health policy experiment ever
  - **Problem:** Politicians intervene; most large evaluations fail!
  - **Solution:** (M) “Politically Robust” Fail-Safe Design: randomization survives political interventions
  - **Design:** Matched pair >> complete randomization
    - Lower bias, model dependence, research costs, & SEs (by 600%!)
- Results
  - **Health Care Impoverishment:** drastically reduced
  - (?) States qualify for federal funding by automatically signing up the poor, but only 66% of the poor were aware they had health care! ~> immediately corrected
  - **Externalities:** design implemented around the world



# Reverse Engineering Censorship in China

# Reverse Engineering Censorship in China

- How'd we get to do this?

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M),

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (S),

## Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (CS),  $\rightsquigarrow$  (?)

## Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed

## Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them

## Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- Everyone knows the Goal:

## Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- Everyone knows the Goal:  
Stop criticism and protest about the state,  
its leaders, and their policies

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~Everyone knows the Goal:~~  
Stop criticism and protest about the state,  
its leaders, and their policies *Wrong*

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~Everyone knows the Goal:~~  
Stop criticism and protest about the state,  
its leaders, and their policies *Wrong*
- What Could be the Goal?

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🔒),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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*

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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!

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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:

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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:
  
  
  - We can predict:

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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
  - We can predict:

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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
  - We can predict:

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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
  - We can predict:
    - Officials in trouble, likely to be replaced

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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
  - We can predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests;

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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
  - We can predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests; new peace treaties;

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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
  - We can predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests; new peace treaties; emerging scandals

# Reverse Engineering Censorship in China

- How'd we get to do this? trying (M), with (🚫),  $\rightsquigarrow$  (?)
- Previous approach: watch a few posts; see what's removed
- Data: Download all posts before the Chinese censor them
- ~~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
  - We can predict:
    - Officials in trouble, likely to be replaced
    - Dissident arrests; new peace treaties; emerging scandals
    - Disagreements between central and local leaders

# Reverse Engineering China's 50c Party

The Government Surreptitiously Fabricating Social Media Posts

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants:

# Reverse Engineering China's 50c Party

The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

# Reverse Engineering China's 50c Party

The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

Existing evidence?

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

Existing evidence? A few anecdotes;

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

Existing evidence? A few anecdotes; “no ground truth”;

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity;

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity; even several analyses with made up dependent variables!

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity; even several analyses with made up dependent variables!

**Our evidence:**


# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity; even several analyses with made up dependent variables!

**Our evidence:**

-  with journalists who found leaked archive in 1 county

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity; even several analyses with made up dependent variables!

### Our evidence:

- 📷 with journalists who found leaked archive in 1 county
- ❓ from fact of leak to reason for 50c




# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity; even several analyses with made up dependent variables!

### Our evidence:

-  with journalists who found leaked archive in 1 county
-  from fact of leak to reason for 50c
-  to systematize messy text, image, email archive

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity; even several analyses with made up dependent variables!

### Our evidence:

- (S) with journalists who found leaked archive in 1 county
- (?) from fact of leak to reason for 50c
- (M) to systematize messy text, image, email archive
- extrapolated results to all of China

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies

**Existing evidence?** A few anecdotes; “no ground truth”; “no successful attempts to quantify” 50c party activity; even several analyses with made up dependent variables!

### Our evidence:

- (S) with journalists who found leaked archive in 1 county
- (?) from fact of leak to reason for 50c
- (M) to systematize messy text, image, email archive
- extrapolated results to all of China
- In a poll(!), convinced 50c to acknowledge behavior

# Reverse Engineering China's 50c Party

The Government Surreptitiously Fabricating Social Media Posts

- ~~Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies~~ *Wrong*

# Reverse Engineering China's 50c Party

The Government Surreptitiously Fabricating Social Media Posts

- ~~Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies~~ *Wrong*
- Fabricates 450M social media posts a year!

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- ~~Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies~~ *Wrong*
- Fabricates 450M social media posts a year!
- Does not argue; does not engage on controversial issues

# Reverse Engineering China's 50c Party

## The Government Surreptitiously Fabricating Social Media Posts

- ~~Prevailing view of scholars, activists, journalists, social media participants: 50c party argues against those who criticize the government, its leaders, and their policies~~ *Wrong*
- Fabricates 450M social media posts a year!
- Does not argue; does not engage on controversial issues
- **Distracts**; redirects public attention from criticism and central issues to **cheerleading** and positive discussions of valence issues

# Externalities: Finding Those Hiding in Plain Sight

# Externalities: Finding Those Hiding in Plain Sight

Example Substitution 1:

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1:

自由

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1:

自由

“Freedom”

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1:

自由

“Freedom”

**CENSORED**

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1:

自由  
自由

“Freedom”

**CENSORED**

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1:

自由  
目田

“Freedom”

“Eye field”

**CENSORED**

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1:

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐

“Harmonious [Society]” (official slogan)

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐

“Harmonious [Society]” (official slogan)

**CENSORED**

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab”

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation;

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation; with **(M)** we can.

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation; with **(M)** we can.



# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation; with **(M)** we can.

**(?)**: (1) Analysts,

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation; with **(M)** we can.

**(?)**: (1) Analysts, (2) eDiscovery,

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation; with **(M)** we can.

**(?)**: (1) Analysts, (2) eDiscovery, (3) Long tail search,

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation; with **(M)** we can.

**(?)**: (1) Analysts, (2) eDiscovery, (3) Long tail search,  
(4) Language drift (#BostonBombings  $\rightsquigarrow$  #BostonStrong),

# Externalities: Finding Those Hiding in Plain Sight

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

Censors can't follow the conversation; with **(M)** we can.

- (?)**: (1) Analysts, (2) eDiscovery, (3) Long tail search,  
(4) Language drift (#BostonBombings  $\rightsquigarrow$  #BostonStrong),  
(5) Look-alike modeling, etc.



# The Effect of the News Media

# The Effect of the News Media

- What's the evidence?

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - Clever alternatives:

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - Clever alternatives: (1) New TV tower, behind a hill, in radio shadow;

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of “surprise” events; etc.

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)
- What's required to randomize

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)
- What's required to randomize
  - **Journalists:**

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)
- What's required to randomize
  - **Journalists:** total control over what's published & when



# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)
- What's required to randomize
  - **Journalists:** total control over what's published & when
  - **Scientists:**


# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)
- What's required to randomize
  - **Journalists:** total control over what's published & when
  - **Scientists:** total control over what's published & when


# The Effect of the News Media

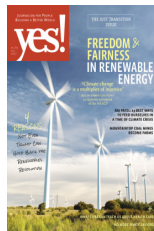
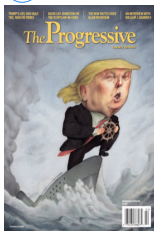
- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) 
- What's required to randomize
  - **Journalists:** total control over what's published & when
  - **Scientists:** total control over what's published & when
- 

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but  $\rightsquigarrow$  "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)
- What's required to randomize
  - **Journalists:** total control over what's published & when
  - **Scientists:** total control over what's published & when
-  48 media outlets let us randomize:

# The Effect of the News Media

- What's the evidence?
  - We can't randomize!
  - **Clever alternatives:** (1) New TV tower, behind a hill, in radio shadow; (2) Before/after studies of "surprise" events; etc.
  - Better but ~ "Profound biases," > 600% difference from truth
  - **Estimands:** different, of sometimes questionable relevance
- Switch to the obvious (hard) (?)
- What's required to randomize
  - **Journalists:** total control over what's published & when
  - **Scientists:** total control over what's published & when
-  48 media outlets let us randomize:



...

# The Effect of the News Media

- We choose a *policy area* (1 of 11)

## The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)

## The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles

## The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)

## The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - **Outlets Publish Simultaneously** (following their procedures)

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - **Outlets Publish Simultaneously** (following their procedures)
- Statistical Methods (M)

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news
  - One coin flip: to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - Outlets Publish Simultaneously (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news
  - One coin flip: to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - Outlets Publish Simultaneously (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news
  - One coin flip: to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - Outlets Publish Simultaneously (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- Results

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news
  - One coin flip: to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - Outlets Publish Simultaneously (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- Results
  - Causal effects:

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - **Outlets Publish Simultaneously** (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- Results
  - **Causal effects:** 1st day: 19.4% increase,

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news
  - One coin flip: to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - Outlets Publish Simultaneously (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- Results
  - **Causal effects:** 1st day: 19.4% increase, Total: 62.7% increase

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - **Outlets Publish Simultaneously** (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- Results
  - **Causal effects:** 1st day: 19.4% increase, Total: 62.7% increase
  - **Context:** 3 small media outlets

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- **Matched Pair Randomization** (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - **Outlets Publish Simultaneously** (following their procedures)
- **Statistical Methods** (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- **Results**
  - **Causal effects:** 1st day: 19.4% increase, Total: 62.7% increase
  - **Context:** 3 small media outlets  $\rightsquigarrow$  national conversation

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - **Match weeks on similarly of predicted news**
  - **One coin flip:** to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - **Outlets Publish Simultaneously** (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- Results
  - **Causal effects:** 1st day: 19.4% increase, Total: 62.7% increase
  - **Context:** 3 small media outlets  $\rightsquigarrow$  national conversation  $\rightsquigarrow$  policy agenda

# The Effect of the News Media

- We choose a *policy area* (1 of 11)
- Outlets volunteer for a *pack* of  $\approx 3$  (with our approval), following “project manager” protocol (e.g., Panama Papers)
- The pack chooses *subject* for articles
  - **We approve:** If rejected, outlets can still publish
  - **Requirement:** No breaking news (so we can randomize timing)
- Matched Pair Randomization (M)
  - Match weeks on similarly of predicted news
  - One coin flip: to choose weeks
    - **Treatment week:** publish & promote articles (usually Tuesday)
    - **Control week:** no compensation or special actions
  - Outlets Publish Simultaneously (following their procedures)
- Statistical Methods (M)
  - **Sequential Hypothesis Testing:** Saved \$ millions & years
  - **Automated Text Analysis:** To measure national conversation
- Results
  - **Causal effects:** 1st day: 19.4% increase, Total: 62.7% increase
  - **Context:** 3 small media outlets  $\rightsquigarrow$  national conversation  $\rightsquigarrow$  policy agenda  $\rightsquigarrow$  policy



# Designing COVID-19 Public Health Messaging in Italy

# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:


# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis: 


# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called



# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice



# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID:



## Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: 



## Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: . In 5 days,



# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: . In 5 days,
  - Signed up 20 academic collaborators



# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment;



# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities



# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey;



# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging



# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance



# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses,



# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper,



# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public



# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis:  Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- Results



# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working




# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups




# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying




# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying
- 




# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying
-  Impact of quarantine on mental health;

# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying
-  Impact of quarantine on mental health;  
Designed treatments

# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:**  Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** . In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying
-  Impact of quarantine on mental health;  
Designed treatments
- **Externalities**

# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:** 🇮🇹 Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** 🇮🇹. In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying
- **?** Impact of quarantine on mental health;  
Designed treatments
- **Externalities**
  - Day 7: Italy adopts our recommendations

# Designing COVID-19 Public Health Messaging in Italy

- Early in the COVID crisis: 🇮🇹 Italian Government called
- Urgent need: messaging to follow public health advice
- Side effect of COVID: 🇮🇹. In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- Results
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying
- ? Impact of quarantine on mental health;  
Designed treatments
- Externalities
  - Day 7: Italy adopts our recommendations
  - Day 9: Italy implements strategy nationwide

# Designing COVID-19 Public Health Messaging in Italy

- **Early in the COVID crisis:** 🇮🇹 Italian Government called
- **Urgent need:** messaging to follow public health advice
- **Side effect of COVID:** 🇮🇹. In 5 days,
  - Signed up 20 academic collaborators
  - Designed large scale experiment; pretest in multiple cities
  - Ran nationally representative survey; randomized messaging
  - Measured compliance
  - Completed analyses, wrote academic paper, made public
- **Results**
  - Public health messaging was working
  - High compliance for **all** subgroups
  - Even those who distrust government, or think they're lying
- **?** Impact of quarantine on mental health;  
Designed treatments
- **Externalities**
  - Day 7: Italy adopts our recommendations
  - Day 9: Italy implements strategy nationwide
  - Day 11: UN asks to repeat our experiments in other countries



# Dataverse: Automated Data Archiving

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution (?) Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,**


# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world,


# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** :

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** : 100s of contributors,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (OS): 100s of contributors, 1000s of dataverses,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (CC BY): 100s of contributors, 1000s of dataverses, 100,000s of datasets,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but ~~credit~~
  - **Distribute data yourself:** get credit, but ~~preservation~~
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (CC BY): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution (?) Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source (CS):** 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - **(M):**

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (OS): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - (M): data citations,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (CC): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - (M): data citations, UNFs,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but ~~credit~~
  - **Distribute data yourself:** get credit, but ~~preservation~~
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (CC): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - **(M):** data citations, UNFs, security,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (CC): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - **(M):** data citations, UNFs, security, privacy,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but ~~credit~~
  - **Distribute data yourself:** get credit, but ~~preservation~~
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (OS): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - **(M):** data citations, UNFs, security, privacy, metadata,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but ~~credit~~
  - **Distribute data yourself:** get credit, but ~~preservation~~
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (CC): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - **(M):** data citations, UNFs, security, privacy, metadata, automated analytics,

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but ~~credit~~
  - **Distribute data yourself:** get credit, but ~~preservation~~
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (CC): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - **(M):** data citations, UNFs, security, privacy, metadata, automated analytics, novel development, etc.

# Dataverse: Automated Data Archiving

- **Problem:** Forced choice (rock or hard place)
  - **Put data in archive:** get preservation, but credit
  - **Distribute data yourself:** get credit, but preservation
- **Solution** (?) **Dataverse:** get credit & preservation
  - **Keep data on my website,** branded as mine
  - **In fact really in the archive**
    - **Managed by professionals,** by 60 institutions around the world, in ways academics can't
    - **Open source** (OS): 100s of contributors, 1000s of dataverses, 100,000s of datasets, 10,000,000s of users
  - **(M):** data citations, UNFs, security, privacy, metadata, automated analytics, novel development, etc.
- **Every group here can participate!**

# Convincing Facebook to Make Data Available

# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data

# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?”


# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” [This](#) was [Cambridge Analytica](#).


# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” [This](#) was [Cambridge Analytica](#).
- 3 days later :



# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” [This](#) was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”



# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” [This](#) was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...



# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” This was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:



# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” This was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)





# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” This was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)
  - [No pre-publication approval](#) (like NO employees ever)





# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” This was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)
  - [No pre-publication approval](#) (like NO employees ever)
- We iterate, and I propose   a 2-part solution





# Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” This was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)
  - [No pre-publication approval](#) (like NO employees ever)
- We iterate, and I propose   a 2-part solution
  - [Outside academics](#): send proposals, no company veto





# Convincing Facebook to Make Data Available

## Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” This was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)
  - [No pre-publication approval](#) (like NO employees ever)
- We iterate, and I propose   a 2-part solution
  - [Outside academics](#): send proposals, no company veto
  - [Trusted 3rd party](#): Commission at [Social Science One](#) signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges





# Convincing Facebook to Make Data Available

## Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” [This](#) was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)
  - [No pre-publication approval](#) (like NO employees ever)
- We iterate, and I propose   a 2-part solution
  - [Outside academics](#): send proposals, no company veto
  - [Trusted 3rd party](#): Commission at [Social Science One](#) signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
- [Problem solved](#), without balancing  $\rightsquigarrow$  agreements, announcements, funding, 30+ people assigned at Facebook





# Convincing Facebook to Make Data Available

## Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this?](#)” [This](#) was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)
  - [No pre-publication approval](#) (like NO employees ever)
- We iterate, and I propose   a 2-part solution
  - [Outside academics](#): send proposals, no company veto
  - [Trusted 3rd party](#): Commission at [Social Science One](#) signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
- [Problem solved](#), without balancing  $\rightsquigarrow$  agreements, announcements, funding, 30+ people assigned at Facebook
- [Just one issue](#):

# Convincing Facebook to Make Data Available

## Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to share data
- In my hotel room packing, email arrives: “Hey what do we do about [this](#)?” [This](#) was [Cambridge Analytica](#).
- 3 days later : “Could you do a study of the 2016 election?”
- I’d love to , but...I need 2 things & you’ll only give me 1:
  - [Complete access](#) to data, people, etc. (like employees)
  - [No pre-publication approval](#) (like NO employees ever)
- We iterate, and I propose   a 2-part solution
  - [Outside academics](#): send proposals, no company veto
  - [Trusted 3rd party](#): Commission at [Social Science One](#) signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
- [Problem solved](#), without balancing  $\rightsquigarrow$  agreements, announcements, funding, 30+ people assigned at Facebook
- [Just one issue](#): Facebook’s implementation plan was **illegal!**

# How to Share Data Without Giving Anyone Access

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify!

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M) (?)

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”
  - Add special noise: privacy guarantees & can still find patterns





# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”
  - Add special noise: privacy guarantees & can still find patterns
  - Burgeoning literature; used in your cell & at US Census

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests.  $\rightsquigarrow$  (M)





# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
-   “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests.  $\rightsquigarrow$  
-  Facebook

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M) (?) “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests.  $\rightsquigarrow$  (M)
- (Facebook) Facebook
  - Data access provided to > 100 researchers around the globe

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
-   “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests.  $\rightsquigarrow$  
-  Facebook
  - Data access provided to > 100 researchers around the globe
  - Massive dataset (**17T** numbers) on effects of social media

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests.  $\rightsquigarrow$  (M)
- (FB) Facebook
  - Data access provided to > 100 researchers around the globe
  - Massive dataset (17T numbers) on effects of social media
  - All sides achieve goals; no balancing necessary

# How to Share Data Without Giving Anyone Access

- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests. ~→ (M)
- (FB) Facebook
  - Data access provided to > 100 researchers around the globe
  - Massive dataset (17T numbers) on effects of social media
  - All sides achieve goals; no balancing necessary
- Next step Need trusted open source software

# How to Share Data Without Giving Anyone Access

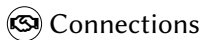
- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests. ~> (M)
- (FB) Facebook
  - Data access provided to > 100 researchers around the globe
  - Massive dataset (17T numbers) on effects of social media
  - All sides achieve goals; no balancing necessary
- Next step Need trusted open source software
  - (MS) add Microsoft ~> V1 software released

# How to Share Data Without Giving Anyone Access


- No problem; we'll de-identify! ~~de-identification~~
- (M?) “differential privacy”
  - **Add special noise:** privacy guarantees & can still find patterns
  - **Burgeoning literature;** used in your cell & at US Census
  - **Statistical properties:** biased, no uncertainty ests. ~> (M)
- (FB) Facebook
  - Data access provided to > 100 researchers around the globe
  - Massive dataset (17T numbers) on effects of social media
  - All sides achieve goals; no balancing necessary
- Next step Need trusted open source software
  - (MS) add Microsoft ~> V1 software released
  - **OpenDP initiative,** hundreds around the globe joining effort

# How to Empower the Social Sciences?

# How to Empower the Social Sciences?



# How to Empower the Social Sciences?




 Connections

 Methods

# How to Empower the Social Sciences?

- ② Connections
- ① Methods
- ③ Questions

# How to Empower the Social Sciences?

-  Connections
-  Methods
-  Questions

---

For more information  
[GaryKing.org](http://GaryKing.org)