



Remaking the Social Sciences

Gary King

Institute for Quantitative Social Science
Harvard University

(talk at University of Virginia, 9/14/12)



The Spectacular Success of Quantitative Social Science

What university research has had the biggest impact on you?

- The genetics revolution?
- The Higgs-like particle?
- Exoplanets? The Mars rovers?
- Doubling the human life span in the last century?
- **Quantitative social science** (aka “big data,” “data analytics,” “data science”): transformed most Fortune 500 firms; established new industries; altered friendship networks (facebook); increased human expressive capacity (social media); changed political campaigns; transformed public health; changed legal analysis; impacted crime and policing; reinvented economics; transformed sports (seen MoneyBall?); set standards for evaluating public policy; etc.; etc., etc.



Cause of the Success: The Changing Evidence Base

The Last 50 Years:

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

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

- More of the above — improved, expanded, and applied
- Shrinking computers & the growing Internet
- The march of quantification: through academia, the professions, government, & commerce
- The replication movement: academic data sharing (e.g., Dataverse)
- Governments encouraging data collection & experimentation
- Biological sciences *becoming* social sciences
- Breathtaking advances in statistical methods, informatics, & software
- End of the Quantitative-Qualitative divide: Diverse perspectives but, with text, video, field notes etc. as data, we now all benefit from the same technologies

→ Change from studying problems to understanding and solving them



Examples of what's now possible with QSS

- **Opinions of activists:** 2000 interviews → billions of opinions in social media posts (1B every 2.5Days)
- **Exercise:** “How many times did you exercise last week?” → 500K people with cell phones accelerometers
- **Social contacts:** “Please tell me your 5 best friends” → all phone calls, emails, text messages, Bluetooth, social media connections, electronic address books
- **Economic development in developing countries:** Dubious or nonexistent governmental statistics → satellite images of human-generated light at night, or networks of roads and other infrastructure
- **Expert-vs-Statistician contests:** Whenever enough information is quantified (& a right answer exists), quantitative social science wins
- Many, **many**, more. . .

- **Need to move the social sciences:**
 - Humanities (on your own) → sciences (lab-style)
 - Large scale, interdisciplinary, collaborative work
- **The science model:**

Hires get: >\$2M in start up costs, 10 employees, 3500sf of lab space. (**Not an option!** but not needed either)
- **A new model for the social sciences:**

Greatly enhanced, shared, common infrastructure



Building Big Social Science: The Politics

- Common aspect of all great centers: **community**
- **Appeal to rational self-interest**; when paths cross, community forms
- Emphasize the whole range from **academic to plumbing**
- **Be university-wide** (many social scientists outside FAS)
- **Cooperate** with other university units, even if initially costly
- Preach and practice: **influence is more important than control**
- Give offense if you want, but **never take offense** (it's not about you)
- Focus on **infrastructure that scales**
- **Become an integral part of university administration** (turn administrative units into quasi-research social science projects, automate them, and harvest the extra revenue)
- **Acquire existing units**; don't expect the administration to pay for it all from scratch

IQSS Research Computing Environment (RCE):

- > 1500 shared computing cores
- Numerous active users
- A persistent environment, sign on from anywhere
- Growing via collaborative research project to extend platform to the Amazon cloud
- Making into open source software

- **Short courses:** R, Zelig, Stata, SAS, Numeric Data Resources, Atlas.ti, GIS, & many IQSS technologies
- **Research consulting:** one-on-one help with data management and analysis
- **Survey Consulting:** help with question wording, sampling, analysis, nonstandard data collection

OpenScholar: Faculty & Student Web sites



Personal Websites Exclusively for Harvard Scholars
Not a Harvard Scholar? Find out how to create a similar site at your institution.



Scholars at Harvard offers you the next generation scholarly web site. With it, you can easily create and manage a highly dynamic, powerful, and customizable site, with an easy-to-use interface and a look and feel that is uniquely yours.

Fact Sheet (PDF) Video Intro

See who's using
Scholars at Harvard!



Current Activity [\(more\)](#)

12 min 5 sec ago Venkatesh N. Murthy, Page updated: [Welcome to the Murthy Lab!](#)

© President & Fellows, Harvard University. Institute for Qu

Get Your Site Now



- > 1,259 **scholar** sites (growing fast)
- > 295 **project** sites
- Free at openscholar.harvard.edu
(\$5-20K saved per site)
- Installation at > 100 other Universities
- Growing via collaboration (with HUIT and HPAC) to extend to **department** sites

Dataverse Network



The Institute
for Quantitative Social Science
at Harvard University

Harvard's home for social science research.

Creating, sharing, and preserving scientific knowledge about human societies.

IQSS Dataverse Network

[Search](#) [User Guides](#) [Report Issue](#)



Welcome to the IQSS Dataverse Network!

Access the world's largest collection of social science research data across or browsing through one of the virtual data archives (called 'Murray Archive', which may easily be customized to appear as if it in fact is served by the IQSS Dataverse Network. This means that visibility, and access control for the data devolve to you, but all the guarantees, and software and hardware upgrades and maintenance IQSS. Learn more about the [Dataverse Network Project](#)

Show All Datasets

Filter by Organization

[Association, Society](#)
[Business, Industry](#)
[Government](#)
[School \(Harvard\)](#) ▾
[School \(Other\)](#)

Filter by Type

[Institutions, Depts, Centers](#)
[Journal Replication Archive](#)
[Large Collections](#)
[Research Projects](#)
[Scholars](#)

All Datasets

[ALL](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#)

Datasets: **376** | Studies: **41,876** | [File](#)

Name	
Ansolabehere, Stephen	View
Leeper, Thomas	View
Harper, Sam	View
Druckman, James	View
Reingold, Beth	View
California Welfare Laboratory	View
Gries, Peter	View
Sorting Guilty Minds	View
Mikhaylov, Slava	View
Pettingill, Lindsay	View

- Your own **easy-to-use virtual archive**
- Largest collection of social science data in the world
- Our installation: ~ 1000 dataverses, 51,000 studies, 715,000 files
- > 20 installations in other Universities
- Expand across humanities, social sciences, & sciences
- Free, open source

- Easy, standardized solution to R's power and diversity
- >100 statistical methods in common format
- Hundreds of thousands of users
- (Optional GUI through Dataverse, no R knowledge required)
- Huge *local* and *international* network effects: students learning from each other
- Growing via open source contributions

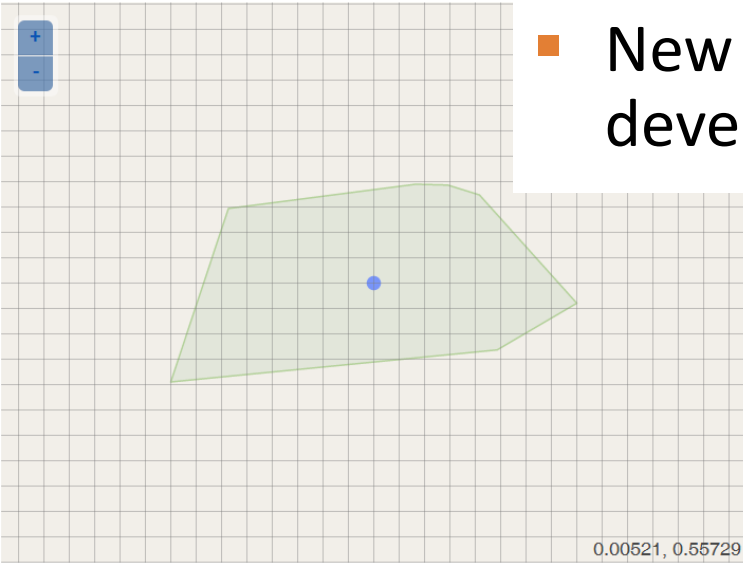
Program on Automated Text Analysis

- “Computer-assisted reading”, a new way to organize and make sense of huge amounts of unstructured text.
- Some services available now
- New (self-serve) approaches under development, available soon

Consilience con·sil·ience (kən·sil·yən
1. (original) *n.* The unity c
2. (new) *n.* An amazing n
derive insights from huge

< Document Sets **Document Set** ZengCheng2011

< > Display History Display Method Points



0.00521, 0.55729

★ Favorites

Label	Clusters	Compare to Current Clustering
-------	----------	-------------------------------

Label [+] "穿", "当下", "版", "话题", "具", "声", "宫", "线", "国际", "野蛮"

8.23% 13 [Detail >](#)

Label [+] "妥善", "X..8", "聚集", "观看", "新", "当事人", "直播", "连载", "妥善处理", "凤凰"



Grants Administration (Don't forget Plumbing!)

- Many active grants
- Increasing number of proposals submitted annually
- Pre- and post-award support
- Working closely with department staff
- Social scientists *come* for help & *stay* to interact on research & teaching

IQSS Activities Build a Community of Scholars

Administrative Services

- Support for PIs, Programs and Centers
- Sponsored Research Administration

Technology Services

- Research and Technology Consultants
- Technical Training
- Research Computing Environment (RCE)
- Repository for Social Science Data (Dataverse)
- Faculty Web Sites (OpenScholar)

Community & Collaboration

Technology Development

- OpenScholar
- Dataverse Network
- Zelig
- Text Clustering Tool
- Research Computing

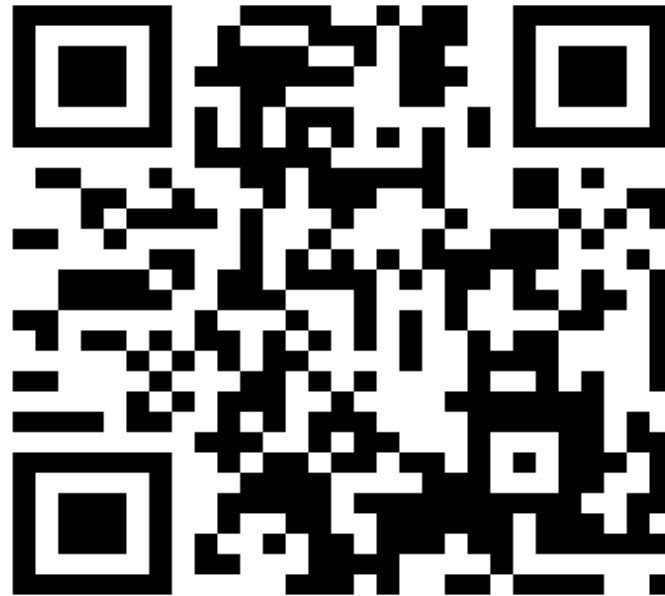


Research/Affiliated Programs

- Center for Geographic Analysis
- Program on Survey Research
- Program on Text Research
- RWJS in Health Policy Research
- Positive Political Economy
- Data Privacy Lab
- Workshops and Conferences
- Program on Methods



For more information.....



<http://GKing.Harvard.edu>