Remaking the Social Sciences

Gary King

Institute for Quantitative Social Science
Harvard University

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

What university research has had the biggest impact on you?

o The genetics revolution?

o The Higgs-like particle?

o Exoplanets? The Mars rovers?

o Doubling the human life span in the last century?

⋮

o 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.

IQSS

The Last 50 Years:
o Survey research
o Aggregate government statistics
o In depth studies of individual places, people, or events

The Next 50 Years: Spectacular increases in new data sources, due to. . .
o More of the above — improved, expanded, and applied
o Shrinking computers & the growing Internet
o The march of quantification: through academia, the professions, government, & commerce
o The replication movement: academic data sharing (e.g., Dataverse)
o Governments encouraging data collection & experimentation
o Biological sciences *becoming* social sciences
o Breathtaking advances in statistical methods, informatics, & software
o 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

IQSS

- 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. . .

IQSS

- **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

IQSS

- 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

- > 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

- 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

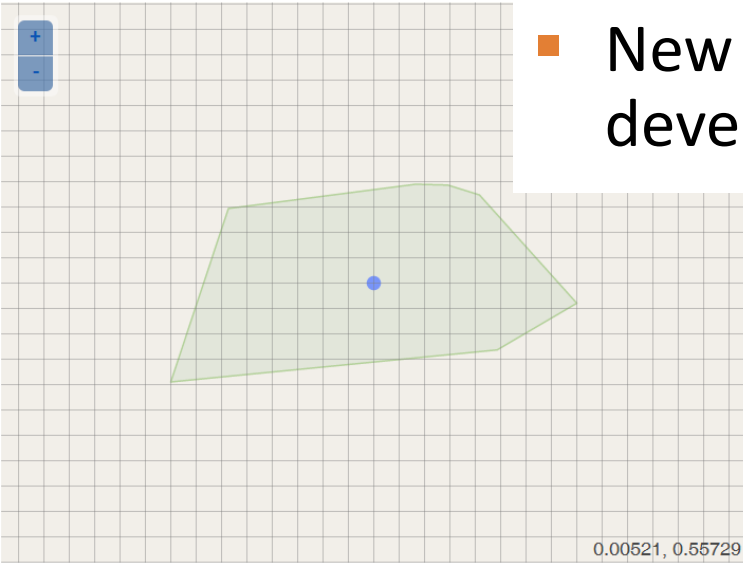- Growing via open source contributions

- "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

- 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 & teachng

IQSS

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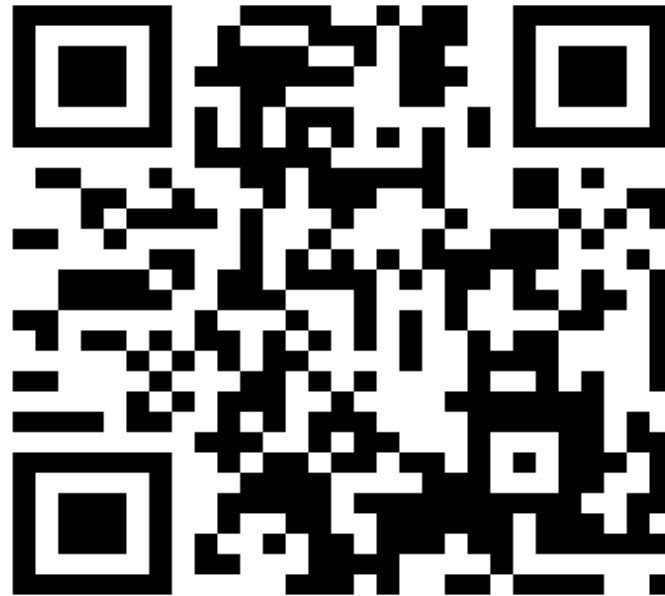- Sponsored Research Administration

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

Community & Collaboration

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

Technology *Development*
- OpenScholar
- Dataverse Network
- Zelig
- Text Clustering Tool
- Research Computing

http://GKing.Harvard.edu