

Program for Quantitative and Analytical Political Science at Princeton

Kosuke Imai

Department of Politics
Center for Statistics and Machine Learning

Presentation at the Advisory Council Meeting
Department of Politics, Princeton University
February 26, 2016

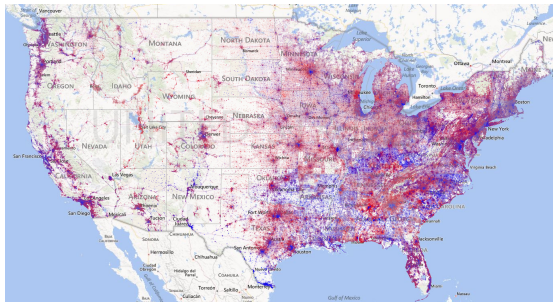
- Internet and computing revolution over the last two decades
- Transformed social science research and education
- Data, Data, and **Data!**
- Past: government data, national survey data
- Today: more of old types of data and lots of new data
 - Randomized experiments and surveys conducted by researchers
 - Administration records: voter files, contributions, lobbying, ...
 - Economic data: trade, company information, finance, ...
 - Military data: casualty, insurgent attacks, ...
 - Social media data: websites, blogs, tweets, cell phones, ...
 - GIS data: satellite, climate, natural resource discoveries, ...
 - Text, images, sounds: news, speeches, bills, commercials, ...

Examples of Recent Dissertation Research

- ① Number and size of European states and their evolution over time
 - Traditional way: rich historical analysis of a few important cases
 - New way: Geocode state boundaries every 10 years over 400 years
- ② Lobbying and trade policies
 - Traditional: hire research assistants hand-code lobbying reports
 - New: write a computer code to scrape and analyze 800,000 reports
- ③ Impact of supreme court opinion on lower courts' decisions
 - Traditional: read lots of opinions!
 - New: use plagiarism detection software to compare 700,000 opinions

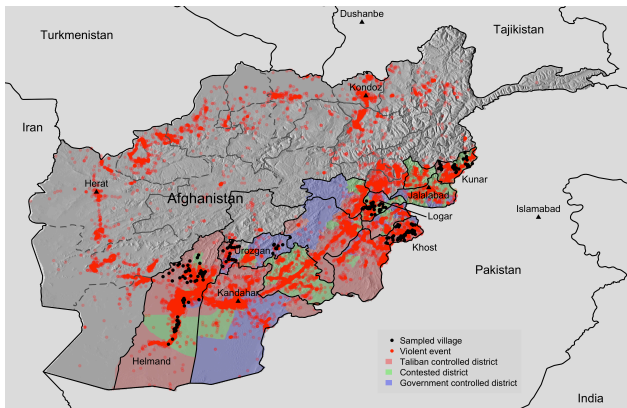
Examples of Recent Faculty Research

- 1 Analysis of 160 million geocoded US voters from L2 to study neighborhood effects on turnout



- 2 Analysis of 20 million roll calls in the state legislature to study polarization
- 3 Analysis of the careers of all federal employees to understand bureaucratic policy making and personnel politics

4 Analysis of geocoded insurgent attacks and US airstrikes in Afghanistan and Iraq to study civil war dynamics

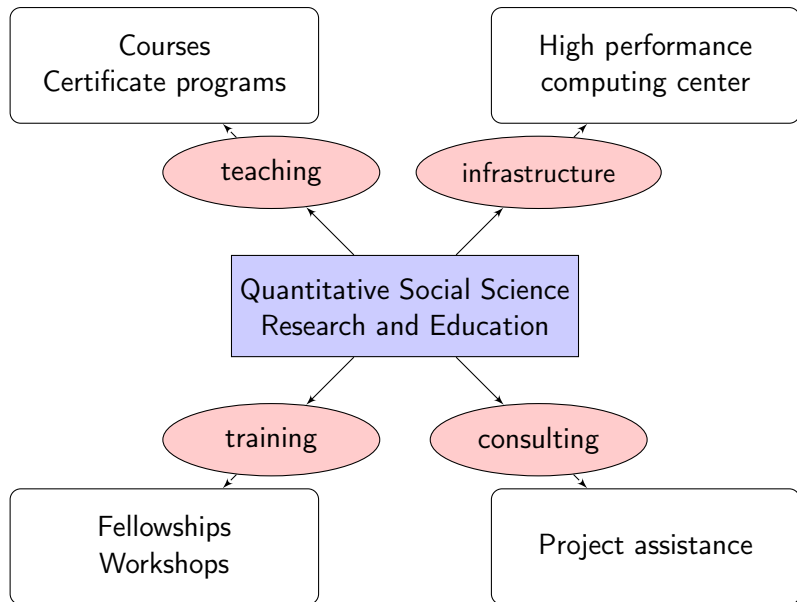


5 Analysis of 3 trillion product level imports and exports between 150 countries over 25 years

Challenges

- ① Data availability is NOT the problem
 - Need for new algorithms, models, and methods
 - Need for new computing technologies
 - parallel computing
 - database management and visualization
 - natural language, image, and video processing
- ② Everyone, not just statisticians and methodologists, are analyzing data
 - Need for methodological training at all levels
 - Undergraduates, not just graduate students
 - Curriculum development
 - Independent research
 - Job market in virtually all areas
- ③ Importance of ideas and theory
 - Big data \rightsquigarrow what and where to look?
 - Game theory
 - How to connect theory to data

Overview of Quantitative Social Science at Princeton



① Undergraduate teaching

- 8 years ago \rightsquigarrow No politics undergraduate course in statistics
- Now, we offer 3 courses
 - POL245 Visualizing Data (Freshman Scholars Institute): 30 students
 - POL345 Quantitative Analysis and Politics: 30 \rightsquigarrow 100 students
 - POL346 Applied Quantitative Analysis: 3 \rightsquigarrow 35 students
- Courses in game theory and political economy
- Certificate program: Statistics & machine learning, Political economy

② Graduate teaching

- 12 years ago \rightsquigarrow No sequence
- Creation of sequences in statistics and game theory
- Improved placements: Caltech, Chicago, Harvard, MIT, Yale, etc.
- Options outside of academia: Facebook, political consulting, etc.

Undergraduate Teaching: Let's Look at the Data First

- Introductory statistics courses are **unpopular**
 - Non-politics introductory statistics courses in social sciences
 - 5 year average: 2008/09–2013/14
 - Courses in ECO, PSY, SOC, and WWS

	Lectures	Assignments	Readings	Precepts	Overall
Statistics	3.2	3.3	3.1	3.6	3.1
All PU courses	3.8	3.7	3.7	4.0	3.9

- Politics introductory statistics courses:

	Lectures	Assignments	Readings	Precepts	Overall
POL 245	4.4	3.9	3.5	3.9	4.3
POL 345	4.0	3.8	3.7	4.2	4.1

- What's the secret of success?

Why Teaching Introductory Statistics Courses is Hard

- 1 Students are **not interested in statistics**:

	Professor	Distribution Requirement	Departmental	Certificate Program	General Interest
Statistics	0%	20%	71%	3%	6%
All PU courses	6%	12%	32%	7%	42%

“Professor Imai tried hard to make statistics interesting. But, statistics is boring.”

- 2 Students have **weak mathematical and programming background**

“as a person not naturally inclined towards statistics and probability, I don't feel at all qualified to pass judgement on how the course might have been improved.”

The Problems of Standard Introductory Statistics Courses

- The Problem for social science students: **Boring!!**
 - only teaches statistical concepts but not actual data analysis
 - often does not teach basic statistical programming
- How can we make these materials more **fun** and **relevant**?
- Emphasize **applications** rather than methods themselves
 - forecasting election outcomes
 - gender and racial discrimination in labor market
 - disputed authorship of the Federalist Papers
- Use of statistics in junior papers and senior thesis
- Statistics as a useful skill for post-graduate career
- Guest speakers from industry

Recent Emails from Alumni

- **Senior thesis research:**

“I took your POL 345 course last semester, and I think I’m going to need to use statistical analysis skills in my research this semester. I’m currently studying abroad in Beijing, and working on a research project to assess which methods can best improve the quality of education and access to education for migrant children in Beijing...”

- **Graduate school:**

“When I first decided to take Professor Imai’s statistics course as a senior majoring in Politics at Princeton, I was excited at the idea of gaining new skill sets but wasn’t entirely sure why statistics and learning R would be useful for me. In retrospect, I am very grateful for having taken the course, which helped me become fully ready for my present graduate studies in political science...”

- **Finance:**

“It was a pleasure, and quite a coincidence, running into you on Wall St. the other week. I have actually been meaning to send you an e-mail to thank you for what Pol 345 has done for me. I recently returned from London for two months of training with DB, and spent a large majority of the time discussing options pricing models, which is highly contingent on statistical assumptions...”

- **Small-town newspaper:**

“I graduated from Princeton last June, and took POL 345 a year ago. I was a history major, and my job now has little to do with statistics – I’m a sports reporter for a small-town newspaper. But I did find a way to employ **R** quite usefully...”

- ① High performance computing center at Princeton
 - Collaboration across different departments
 - Originally built for natural science departments
 - Social sciences and humanities are now utilizing too
 - Growing demands for computational resources

- ② Need training to use these and other resources
 - Computing workshops
 - Consulting services
 - 700 hours
 - 100 unique clients from 9 departments
 - led by research specialists and postdocs

Concluding Remarks

- Everyone – undergrads, grad students, faculty – analyzes data
- Need for better curriculum and infrastructure at all levels
- Importance of human resources: teaching, training, consulting
 - methodologist faculty
 - research specialists
 - postdocs, graduate and undergraduate fellows
- Coordination with other departments at Princeton: Center for Statistics and Machine Learning
- Collaboration with private sectors, non-profits, and government agencies: data collection/sharing, joint projects, job placements
- Internationalization:
 - Political Economy conferences in Venice
 - Political Methodology conferences in Asia