Q-APS: Social Science Research Support for Scholars

The Productive Scholar: Technology Tools for Teaching and Research

April 17, 2014
Q-APS Overview

- Program for Quantitative and Analytical Political Science (Q-APS)
- http://q-aps.princeton.edu/
- Started in 2009 to support the intersection of political science and technical fields like statistics, game theory, and computer science
- Q-APS supports future and current social scientists through
  1. Teaching
  2. Training
  3. Infrastructure
  4. Exchange
  5. Consulting
- Goal: outline ways in which this support can be used by broader Princeton community
Teaching

Statistical Analysis
- the “quantitative” in Q-APS
- data–based description, prediction, or explanation

Theoretical Analysis
- the “analytical” in Q-APS
- characterizing strategic incentives
Teaching: Undergraduate

- Statistical Analysis
  
  **POL 245** Visualizing Data
  **POL 345** Quantitative Analysis and Politics
  **POL 346** Applied Quantitative Analysis

- Theoretical Analysis:
  
  **POL 250** Introduction to Game Theory
  **POL 347** Mathematical Models in the Study of Politics
Teaching: Undergraduate

- Statistical Analysis

Description of Observed Overall HS GPA
Teaching: Undergraduate

- Statistical Analysis

Inference for Mean Overall HS GPA

Years in Program

Overall HS GPA

80 82 84 86

1 4
Teaching: Undergraduate

- Theoretical Analysis

<table>
<thead>
<tr>
<th>Prisoner 1</th>
<th></th>
<th>Prisoner 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mum</td>
<td>-1, -1</td>
<td>-9, 0</td>
</tr>
<tr>
<td>Fink</td>
<td>0, -9</td>
<td>-6, -6</td>
</tr>
</tbody>
</table>

*The Prisoners’ Dilemma*
Teaching: Graduate

• Statistical Analysis
  - POL 571  Quantitative Analysis I
  - POL 572  Quantitative Analysis II
  - POL 573  Quantitative Analysis III
  - POL 574  Quantitative Analysis IV

• Theoretical Analysis:
  - POL 502  Mathematics for Political Scientists
  - POL 575  Formal Political Analysis I
  - POL 576  Formal Political Analysis II
Teaching: Graduate

- Statistical Analysis

![Graph showing statistical analysis results for different categories: distressed, problems, and fit, with predicted number of senders for each grade (Good, Tie, Bad, Very Bad) and aid status (Aid, No Aid).]
Teaching: Graduate

- Theoretical Analysis
Training

**\LaTeX** basics of math–friendly document-preparation software

**Politics Math Camp** math pre-fresher for incoming graduate students

**Statistical Programming Camp** basic statistical programming in \textbf{R}

**High-Performance Computing at Princeton** basic overview of using HPC at Princeton University

**Web-Scraping and Text-Processing with Python** extraction and manipulation of WWW and PDF content with \textbf{Python}

**Advanced Statistical Programming Camp** advanced statistical programming with \textbf{R}, \texttt{C++}, and HPC

*Follow links to access information and resources.*
Why learn how to scrape and process WWW content?
## Training: Web-Scraping and Text-Processing

**Candidate's Electoral Campaign Return**

**Financial Reports**
- Candidates
- Choose an Election
- Search option
- By candidates' summary
- By candidates' details
- By electoral district
- By contributor
- Description of the Return
- Introduction

**Contributions - Details**

**Selected Election**
- 41st general election (May 02, 2011)

**Return Type**
- Data as submitted
  - Original return as submitted to Elections Canada
  - May include updates to the original return by the agent and minor corrections made by Elections Canada

**Download (XML)**

<table>
<thead>
<tr>
<th>Name of contributor</th>
<th>Name of candidate Political party Electoral district</th>
<th>Date received</th>
<th>Class of contributor Part # of the return</th>
<th>Contribution given through (name of association)</th>
<th>Monetary $</th>
<th>Non-monetary $</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thirumugilan A.</strong></td>
<td>S/Isabaissan, Rathika / New Democratic Party / Scarborough—Rouge River</td>
<td>Apr 25, 2011</td>
<td>Individuals / 2a</td>
<td></td>
<td>400.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Jas N Aagard</strong></td>
<td>Barnes, Dave / Green Party of Canada / Brandon—Souris</td>
<td>Apr 08, 2011</td>
<td>Individuals / 2a</td>
<td></td>
<td>400.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
### Training: Web-Scraping and Text-Processing

#### Financial Reports

<table>
<thead>
<tr>
<th>Contributions - Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contributor Information</strong></td>
</tr>
<tr>
<td><strong>Full Name:</strong> Thimirugan A.</td>
</tr>
<tr>
<td><strong>City:</strong> Scarborough</td>
</tr>
<tr>
<td><strong>Province:</strong> ON</td>
</tr>
<tr>
<td><strong>Postal Code:</strong> M1H2Y7</td>
</tr>
</tbody>
</table>

#### Table: Contributions

<table>
<thead>
<tr>
<th>Name of contributor</th>
<th>Name of candidate/Party/Electoral district</th>
<th>Date received</th>
<th>Class of contributor/Part # of the return</th>
<th>Monetary $</th>
<th>Non-monetary $</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thimirugan A.</td>
<td>Sitasaliesan, Rathika / New Democratic Party / Scarborough–Rouge River</td>
<td>Apr 25, 2011</td>
<td>Individuals / 2a</td>
<td>400.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Jes N Aagaard</td>
<td>Barnes, Dave / Green Party of Canada / Brandon–Souris</td>
<td>Apr 08, 2011</td>
<td>Individuals / 2a</td>
<td>400.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Silar, Brian / New</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q-APS: Support for Scholars

Q-APS Teaching Training Infrastructure Exchange Consulting

April 17, 2014

12 / 25
Using **Python** and the tools covered in this workshop,

we were able to automate the data collection for over 265 thousand political contributions in Canada.

Unlike hiring many RA’s, the process based on these tools **scales** to updates to the source website or changes to the desired output.
Why learn tools to improve performance of large-scale analysis?
Training: Advanced Statistical Programming Camp
Training: Advanced Statistical Programming Camp
Training: Advanced Statistical Programming Camp
Training: Advanced Statistical Programming Camp

![Graph showing the relationship between dataset size and time taken for different approaches. The x-axis represents dataset size, the y-axis represents time in seconds. The graph includes lines for loop, parloop, cpp, and parc++ approaches.]
Training: Advanced Statistical Programming Camp

• Using R and the tools covered in this workshop,

• we are able to realize performance gains of over 100x relative to “naive” approaches and over 10x relative to “good” approaches.

• Waiting 1 second vs. 10 seconds for results isn’t very different. But, waiting 1 hour vs. 10 hours is very different.
Through collaboration with Research Computing @ Princeton . . .

- **Tukey**: Politics–specific computational cluster
- **Secure System for “Big Data”**: Distributed system for storage, retrieval, analysis, and visualization of big data
Exchange

• Support:
  • Post-Doctoral Fellows
  • Graduate Student Fellows

• Seminars:
  • Weekly Formal/Quantitative Seminar

• Conferences:
  • Hosted 2011 Summer Meeting of Society for Political Methodology
  • Co-Sponsored 2014 Asian Political Methodology Meeting
  • Co-Sponsored Annual Graduate Student Research Poster Session

Graduate Student Research Poster Session
Monday, January 13, 2014
300 Wallace Hall
12:00-2:00

Sponsored by the Department of Politics, Center for the Study of Democratic Politics (CSDP), and Program for Quantitative and Analytical Political Science (Q-APS)

Chantal Berman & Elisabeth Hugent
Was the Revolution Tweeted?
New Media and Old Networks in the Egyptian Revolution

Nhung Bui & Song Ha Joo
The Impact of Militarism and Personalism on Military Disputes: Unpacking Temporal and Spatial Variations

Brandon de la Cuesta & Romain Feraul
Who Gets to Rule?
A Structural Approach to Coalition Formation in sub-Saharan Africa

Benjamin Pittfield & Lauren Mattio
Reexamining the Firm-Regulator Relationship
Consulting

- The Q-APS consulting service began in Fall 2013

- Since then, we have provided over 500 consulting sessions and provided 400 hours of support

- Most of our clients come from the social science community: Politics, WWS, Sociology, and Economics

- But they come from departments like ORFE, EEB, and Electrical Engineering too!

- We support all levels of independent research: JPs, Senior Theses, Dissertations, and Faculty research projects
We frequently consult on the “usual suspects”:

• research design
• descriptive data analysis
• statistical inference (e.g., text analysis, multi-level models, causal inference)
• programming in R, Python, C++
• porting R, Python, C++ to Princeton’s HPC systems

But, the possibilities are not constrained to these.
Consulting: Who can benefit?

Descriptive Analysis of Textual
Consulting: Who can benefit?

Visualization of Geographical Data
Consulting: Who can benefit?

Predicted Authorship versus Actual Authorship

Statistical Inference from Textual Data
Consulting: Who can benefit?

Scraping Social Media to Measure Sentiments
Consulting: How?

• If you, your post-doc, your graduate student, or your undergraduate student can benefit from our services in support of independent research, send an email to:

q-apsconsulting@trac.princeton.edu

• Include:
  • Name
  • Department
  • Description of what you are working on
  • Description of the “problem”
Consulting: How?

If you misplace the email address, we are Google-able . . .

Statistical and Formal Theory Consulting | The Program for ...
Note: Consulting service office hours for the Fall Semester begin on Monday, September 16th. Our consultants are also available outside of office hours by ...