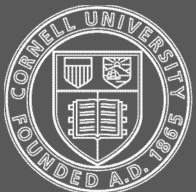


Using Roper Center Data to Satisfy Transparency Requirements

Peter K. Enns

Executive Director, Roper Center
Associate Professor, Dept. of Government
Cornell University
peterenns@cornell.edu

Data Under Constraint Workshop, Aug. 30, 2017

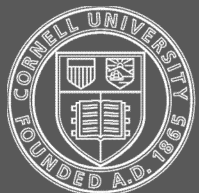


Cornell University



Overview

1. Why use the Roper Center to discuss data constraints and transparency?
2. How to satisfy transparency requirements with constrained data.
3. Heffington, Park and Williams' Most Important Problem Dataset as a case study.
4. Conclusions

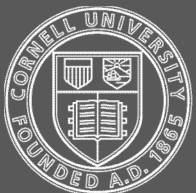


Cornell University



Why Discuss Transparency with Roper Center Data?

- Roper Center data cannot be made public:
<https://ropercenter.cornell.edu/terms-and-conditions/>.
- This constraint affects a lot of researchers and journals
 - Approximately 370 citations/year (google scholar)
 - Well over 100 *different* journals.
- The Roper Center wants to promote data transparency with Roper data.
- We believe the Roper Center's approach can also generalize to other types of data that are under constraint.



Cornell University

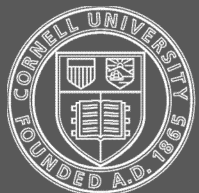


Number of *Different* Journals by Subject Area Citing Roper Data (2000-2016, *JSTOR*)

African American Studies (3) / African Studies (3) / American Studies (41) / Anthropology (5) / Archaeology (1) / Asian Studies (6) / Bibliography (4) / Biological Sciences (8) / Business (44) / Criminology & Criminal Justice (6) / Ecology & Evolutionary Biology (2) / Economics (38) / Education (11) / Feminist & Women's Studies (9) / Finance (1) / General Science (4) / Geography (1) / Health Policy (9) / Health Sciences (8) / History (19) / History of Science & Technology (1) / International Relations (16) / Irish Studies (1) / Jewish Studies (5) / Labor & Employment Relations (2) / Language & Literature (5) / Latin American Studies (7) / Law (30) / Library Science (3) / Management & Organizational Behavior (8) / Mathematics (2) / Music (1) / Peace & Conflict Studies (12) / Philosophy (3) / Political Science (298) / Population Studies (9) / Psychology (16) / Public Policy & Administration (18) / Religion (45) / Social Work (7) / Sociology (274) / Statistics (1) / Zoology (2)

Data Access & Research Transparency (DART):

- 1) Ensure that cited data are available at the time of publication through a trusted digital repository
- 2) Require authors to delineate clearly the analytic procedures upon which their published claims rely, and where possible to provide access to all relevant analytic materials
- 3) Maintain a consistent data citation
- 4) Ensure that journal style guides, codes of ethics, publication manuals, and other forms of guidance are updated and expanded to include improved data access and research transparency requirements



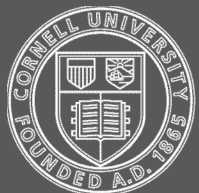
(3) Maintain a consistent data citation

The Roper Center provides a formal citation for every survey:

Fannie Mae. Fannie Mae's National Housing Survey Monthly Home Purchase Sentiment Index (HPSI) and Key Indicators - July, 2017 (a subset of the complete monthly Fannie Mae National Housing Survey), Jul, 2017 [dataset]. 31099569, Version 2. Penn, Schoen and Berland Associates [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], accessed Aug-26-2017.

And every survey question:

Gallup Organization. Gallup Poll, Mar, 2017 [survey question]. USGALLUP.041117.R12. Gallup Organization [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, iPOLL [distributor], accessed Aug-26-2017.

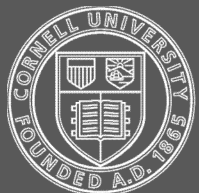


Cornell University



(2) Delineate analytic procedures and where possible provide access to all relevant analytic materials

- Syntax files to replicate the analysis can be published with the journal.
- If the journal requires replication, the Roper Center will provide the journal with the data (but data cannot be made public).

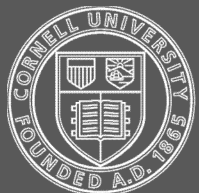


Cornell University



(1) Ensure that cited data are available at the time of publication through a trusted digital repository

- The Roper Center encourages researchers to archive replication data and replication materials with the Roper Center.

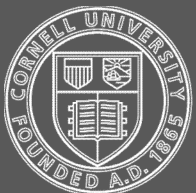


Cornell University



Benefits of Archiving Replication Data w/ the Roper Center

- Satisfies Roper Center terms and agreements.
- Satisfies DART requirements.
 - Ensure that cited data are available through a trusted digital repository.
 - Delineate analytic procedures and where possible provide access to all relevant analytic materials.
- We promote the data through our website, newsletter, social media, etc.
- Researchers and students at more than 300 member institutions can immediately access the data.
- Individuals at non-member institutions can request the data.

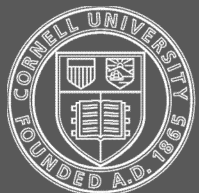


Cornell University



An Example: Most Important Problem Dataset

- Colton Heffington, Brandon Beomseob Park and Laron K. Williams, 2017. "The 'Most Important Problem' Dataset (MIPD): A New Dataset on American Issue Importance." *Conflict Management and Peace Science*
- The MIPD contains over 832,000 Americans' responses to the "most important problem" question from 1939-2015
- The MIPD also contains individual-level information on demographics, economic evaluations, partisan preferences, approval and party competencies.



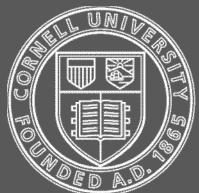
Cornell University



Most Important Problem Dataset (MIPD)

- First note in article:

“Replication materials for the analysis can be accessed on SAGE’s CMPS website. Replication data can be accessed for individuals affiliated with institutional members of the Roper Center for Public Opinion Research at goo.gl/GG9d87. Individuals not affiliated with a member institution can request the data on a case-by-case basis by emailing data-services@ropercenter.org.”



Cornell University



Home | Research | MIP Data | Government Data | Programs | Teaching | CV

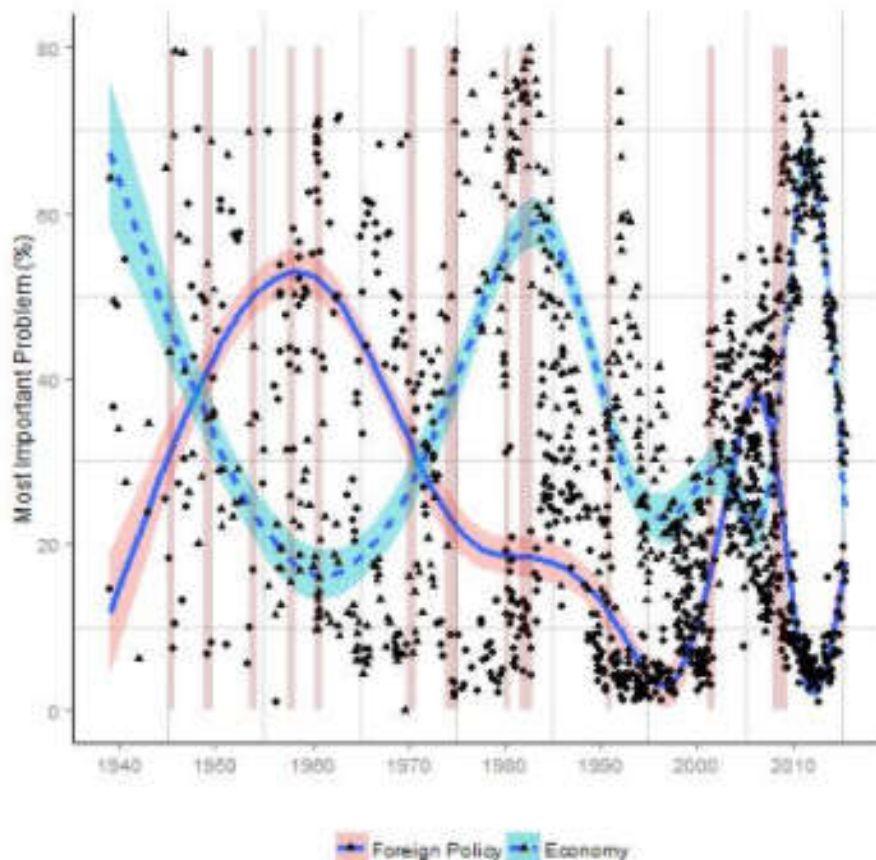
'Most Important Problem' Dataset updated through June 2015:

The 'Most Important Problem' Dataset (MIPD) contains over 832,000 Americans' responses to the "most important problem" question from 1939-2015. This page provides the various data sets, supporting materials and Stata files required for creating a variety of datasets to match specific research needs. Please cite the following: Colton Heffington, Brandon Beomseob Park and Laron K. Williams, forthcoming. "The 'Most Important Problem' Dataset (MIPD): A New Dataset on American Issue Importance", *Conflict Management and Peace Science*.

Stata data:

- All of the following files are only available from the Roper Center for Public Opinion Research
- [MIPD Individual Dataset \(Release 1.0\)](#): dataset containing the individual-level responses to the MIP question, as well as demographics, economic evaluations, presidential approval, and party competency questions.
- [MIPD Aggregate Dataset \(Release 1.0\)](#): Survey-level dataset containing the aggregate percentages of MIP responses based on the Comparative Agendas Project (CAP), Manifesto Research on Political Representation (MARPOR), and Singer (2011) coding schemes.
- [MIPD Annual Dataset \(Release 1.0\)](#): Annual dataset for the percentage of Americans identifying various categories as the "most important problem" facing the country today.

Most Important Problem Dataset



Colton Heffington, Brandon Beomseob Park and Laron K. Williams, 2017.

Public opinion researchers depend on certain questions as essential public opinion barometers, like presidential job approval or Bud Roper's right-direction/wrong-track measure. Perhaps no other question is as often used to determine what is foremost in the minds of the public than the open-ended "What do you think is the most important problem facing this country today?" Respondents offer their concerns in their own words, unaffected by potential bias introduced by limited lists of answers.

Data News

[Lemberg Center Poll: Six-City Study on Violence](#)

August 23, 2017

[The Summer of Love at the State Level](#)

July 26, 2017

[From Russia with Love: Young Adult Opinions in 1967](#)

July 19, 2017

Subscribe to Newsletter

Email *

Submit

[NEW SEARCH](#)

DOCUMENTATION

Survey questionnaires and documentation are available for free download.

[Survey References, PDF file](#)

[Download Adobe Acrobat Reader](#) 

DATASETS

The following datasets are available for download:

Study Documentation, Questionnaires and Codebooks:

[MIPD Individual Codebook PDF file](#)

[MIPD Aggregate Codebook PDF file](#)

Data Sets:

[MIPD Aggregate Dataset \(Release 1.0\)](#)

STATA file

[MIPD Annual Dataset \(Release 1.0\) STATA](#)

file

[Stata File to Generate Aggregate Datasets](#)

STATA do-file

[MIPD Individual Dataset \(Release 1.0\)](#)

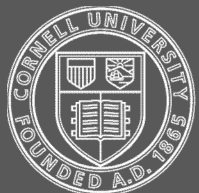
STATA file

Dataset Abstract

| | |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title: | The 'Most Important Problem' Dataset (MIPD): A New Dataset on American Issue Importance, Release 1.0 REPLICATION [USMISC2015-MIPD] |
| Survey Sponsor: | See survey notes for full citation |
| Survey Firm: | Not applicable |
| Study Date: | 1939-2015 |
| Sample: | |
| Sample Size: | |
| Survey Notes: | The Most Important Problem Dataset is an aggregated dataset based on over 600 Roper Center datasets. It contains more than 832,000 Americans' responses to the "most important problem" question from 1939-2015. This dataset was produced by researchers Colton Heffington, Brandon Beomseob Park and Laron K. Williams. *This is a replication dataset and has been made available in the form it was submitted by the researchers. This dataset has not been curated by the Roper Center in any way.* PLEASE CITE AS: Colton Heffington, Brandon Beomseob Park and Laron K. Williams (forthcoming). "The 'Most Important Problem' Dataset (MIPD): A New Dataset on American Issue Importance", Conflict Management and Peace Science. [dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor]. |
| Variables: | |
| Abstract: | Most Important Problem (MIPD) Individual: This dataset contains individual-level responses (from over 670 surveys totaling almost a million respondents) to the MIP question, in addition to demographics, partisan preferences (including vote intention, previous vote, ideology and party identification), approval (general approval and approval of specific policies), economic evaluations |

Most Important Problem Dataset (MIPD)

- Usage statistics since 1/23/17:
- 466 downloads
- 1151 page views

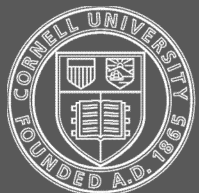


Cornell University



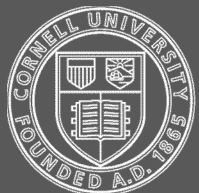
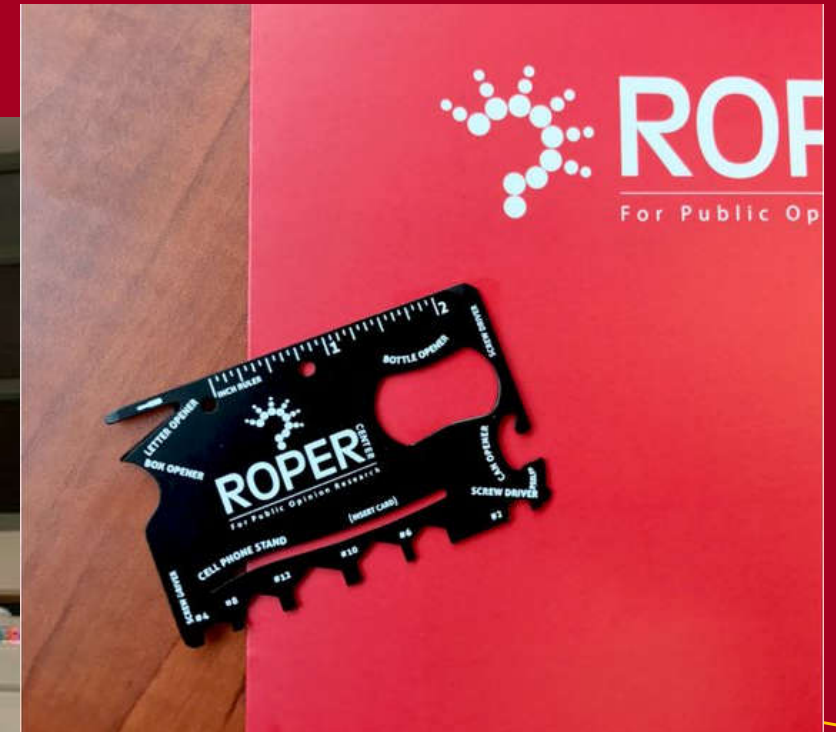
Conclusions

- Researchers and journals can easily satisfy transparency requirements and meet Roper Center terms and conditions.
- To do this, we strongly encourage researchers to archive their replication data with the Roper Center.
- Even when data from other sources seem constrained, lots of options may exist.
- Consider: replication syntax files, replication without making data public, archiving replication data with the original source.



Conclusions

- Publishing with Roper Center Data:
 - <https://ropercenter.cornell.edu/publishing-roper-center-data/>
- Contact the Roper Center with any questions:
 - roper-data@cornell.edu
 - 607.255.8129



Cornell University

