The 'Most Important Problem' Aggregate Dataset (MIPD) Codebook

Release 2.0

T. Murat Yildirim
Stavanger University

Laron K. Williams
University of Missouri

June 20, 2024

Contents

List of Tables

1 Overview

This document discusses the procedures used to create aggegate-level datasets containing the weighted percentages of MIP responses from the individual-level data. In this document, we will discuss weighting procedures, variable descriptions, and provide some basic descriptive statistics. Those interested in the original data collection or question wording can examine the codebook for the individual-level dataset, "MIPD Codebook–Release 2.0.pdf".

2 Reference

2.1 Location

The MIPD dataset can be found in the following locations:

- **Individual Data:** Roper Center for Public Opinion Research
- **Aggregate Data:** Laron Williams' personal website: `http://williamslaro.github.io`

2.2 Citation

This is the second release of the dataset. When using the dataset please cite both companion articles:

1. Heffington, Colton, Brandon Beomseob Park, and Laron K. Williams. 2019. "The 'Most Important Problem' Dataset (MIPD): A New Dataset on American Issue Importance," *Conflict Management and Peace Science.* 36.3: 312-335.
2. Yildirim, T. Murat and Laron K. Williams. Forthcoming. "Problem Importance across Time and Space: Updating the 'Most Important Problem Dataset"'. *Journal of Elections, Public Opinion and Parties*.

2.3 Acknowledgements

This project would not be possible without the helpful research assistance of Min Kyu Chang, Rachel Dicke, Ramsha Giliani, Dongjin Kwak, Ted Masthay, Andy Philips, and Stella-Leonie Wancke.

3 Overview of Aggregate Datasets

The MIPD aggregate dataset is structured so that it is easy to examine MIP measures over time for a variety of coding schemes, including the Comparative Agendas Project (CAP, Baumgartner and Jones 2002). We make a variety of coding decisions in creating these aggregate percentages.

1. **Temporal Domains:** Datasets are aggregated into four temporal domains: survey, annual, quarterly, and monthly. Keep in mind that surveys earlier than 1980 are rather scarce, so temporal domains beyond annual might have considerable missing data. Or, users can use the survey unit and use any aggregation procedure they like.

2. **Multiple Responses:** While the MIPD includes up to three MIP responses (if available), the aggregate MIP measures are only calculated based on the first MIP.

3. **Denominator:** We exclude all invalid responses from the calculations. This includes those that are "no opinion" (1502: including no answer, blank, not coded), "not applicable" (1503: including not asked, not applicable), and "refused" (1504: including unknown or refused). The denominator is calculated only by including those with valid responses. Valid responses also include "none", "other", and "don't know". The percentages can therefore be interpreted as the percentage of the electorate who answered the question and selected that category.

4. **Representative Samples:** The aggregate measures exclude those without a nationally-representative adult sample. This means excluding registered voters-only samples (`sample` = 2) and mixed samples of adults and youths (`sample` = 4).

5. **Question Wording**: we have coded about 20 different variations on the "most important problem" question. Some of these may be more appropriate for some empirical analyses than others. We select only those MIP questions that we think get at national assessments of the most important problem or issue currently facing the country. This means that we exclude a few surveys that ask variations of the MIP question wording that don't meet this standard (based on the `mipid` variable). We exclude the following question wordings:

    - 3: "What do you think is the most important problem facing this section of the country today?"
    - 17: "What do you think is the single most urgent problem facing the country today?"
    - 23: "What is the most important problem that you and your family face today?"
    - 30: "What do you think is the main problem facing the country today?"
    - 31: "What do you think will be the most important problem facing this country in the 21st Century?"

6. **Subgroups:** The percentages for the various categories are available across three subgroups of respondents. For example, selecting the `male` variable for the subgroup analysis will produce separate percentages of each coding scheme requested for males and females. These

variables will reflect, for example, the percentage of males selecting "economy". Variables are named based on the numeric values of the subgroups, but the variable labels provide more information about what the values represent (for example, a variable with the 1 suffix in the Gender dataset is coded as "male"). These are the subgroups available for download (though users can generate any splits they want based on the individual-level data):

   (a) Gender: female = 0, male = 1

   (b) Income quartiles: lowest 25th = 1, second 25th = 2, third 25th = 3, highest 25th = 4

   (c) Partisan identification: 1 = Democrat (including Strong and Lean), 2 = Independent, 3 = Republican (including Strong and Lean)

7. **Administrative Variables:** We include some additional administrative variables to help contextualize the individual data used to generate the aggregate percentages: number of surveys for that time period and number of valid (non-missing) respondents for that time period.

8. **Linear Interpolation:** For the time series datasets (based on annual, quarterly, and monthly data), missing data are linearly interpolated. Interpolated observations are those with values of 0 for the number of surveys (`numsurveys`) and missing values for the number of respondents (`numrespondents`). If you don't want the empirical analysis to include interpolated values, drop the observations based on those conditions.

3.1 Weighting Procedures

This aggregate dataset uses three weighting procedures to ensure that the individual-level responses can be effectively used to generate aggregate-level estimates of the percentage of respondents identifying different categories as the "most important problem" facing the country today.

3.1.1 Quasi-Response Weights

At the individual level we partition each response into quasi-responses containing separate descriptions of problems. The result is that we may code up to 22 quasi-responses for a single respondent. So that respondents who provide more quasi-responses are not weighted more than those who do not, we weight each respondent by the inverse of the number of quasi-responses. For example, a respondent who provides four responses will have a weight of 0.25 for each of their quasi-responses. This weighting procedure is combined with the national representative weights and aggregation weights (described below).

3.1.2 Population Weights

Beginning in the 1970s, polling agencies began supplying population weights in their surveys so that one can make inferences about the population overall. Table 1 shows the surveys and polling agencies that lacked population weights.

Table 1: Distribution of Missing Population Weights across Survey Houses

Survey House	# Surveys†	Missing Years
American National Election Study (ANES)	7	1966, 1972, 1978-1986
Decima Research	1	1989
National Opinion Research Center (NORC)	1	1944
Gallup	99	1939-1971

Note: †: number of surveys missing weights.

Even among those who provided population weights, there is some variation as to the "representative" value; most of the time the average respondent is weighted 1, but others are weighted 10, 100 (Gallup polls from 1993-1999) 1000 (*Los Angeles Times* and *New York Times* polls), or some other integer (Gallup had a representative value of 2 from 1962 to 1983; a *Washington Post* poll conducted by ICR Survey Research in April 1990 had a mean weight of 181,038). For example, early ANES studies either had population weights (e.g., 1960), no weights (e.g., 1966, 1972, 1978-1986), weights to remove duplicate respondents from panels (e.g., 1974 and 1976), or population weights to downweight oversamples of African Americans (e.g., 1964, 1968, 1970).

Population weights (if available) were used to calculate the survey-specific MIP percentages for each problem category, MIP$_p$ in the following manner:

$$\text{MIP}_p = \frac{\sum_{i=1}^{n}(w_i \times P_{pi})}{\sum_{i=1}^{n}(w_i \times M_i)} \tag{1}$$

where n is the number of non-missing responses, w_i is the population weight for observation i (for those surveys without population weights, $w_i = 1 \quad \forall \quad i$), P_i is a binary variable coded 1 if respondent i selected problem p as the MIP, and M_i is a binary variable coded 1 if respondent i provided a non-missing response to the MIP. The result is that each MIP$_p$ provides the percentage of respondents identifying that problem as the MIP (out of those who identified a problem).

3.1.3 Aggregation Weights

There is a significant amount of variability in the availability of suveys across time. Indeed, there might be 13 available surveys in 1960, but the preceding and following years may only have 4 each. Furthermore, even in the years where we have a large number of available surveys, the availability across quarters and months might vary. For example, consider Table 2, which shows the distribution of 29 surveys across months in 2007.

As we aggregate to a higher level, we should consider how many surveys come from each of the lower levels. Assume that we are calculating an annual measure of MIP responses by aggregating the data from the four quarters. A simple method would be to average the percentages from the 29 monthly surveys, but this would give greater weight to the months where we have a lot of surveys (such as May or September). A better method—and the one that we implement in the MIPD Aggregate Dataset—is to provide lesser weight to the surveys that come from more common periods and provide greater weight from the surveys that come from less common periods. This method provides equal weight to each of the lower time periods rather than the surveys themselves.

Consider the calculation of quarterly MIP responses for the first quarter of 2007 (Table 2). The simple method would be to calculate percentages for each individual survey and then merely take the mean of those six surveys. Our method weights each of the January surveys by 1/3, the February survey by 1, and the March surveys by 1/2. Essentially this gives equal weight to each month, rather than survey. The same process is used to generate the annual measures (giving weights to each quarter). Since the monthly measure already has a fine-grained temporal sequence, no aggregation weights are provided.

Table 2: Available Surveys in 2007 Across Months

	Available # of Surveys
January	3
February	1
March	2
April	3
May	4
June	1
July	3
August	3
September	4
October	1
November	2
December	2
Total	29

3.2 Variable Descriptions

This section describes some of the identifier variables that accompany the datasets. The variables describing the surveys (`studyid` through `respondents`) are only in the survey dataset (since the others combine multiple surveys).

- `studyid`: a string identifying each survey uniquely. These identification tags come from the polling agencies (e.g., Gallup, CBS) themselves and uniquely identify the files on the Roper Center's iPoll service.

- `prev_ts`: indicates the date of the previous presidential election in Stata's *dmy* format (e.g. 06nov2012).

- `next_ts`: indicates the date of the next presidential election in Stata's *dmy* format (e.g. 08nov2016).

- `fw_start`: the date that the field work for the survey began in Stata's *dmy* format (e.g. 04apr2013).

- `fw_end`: the date that the field work for the survey ended in Stata's *dmy* format (e.g. 11apr2013).

- `sample`: variable denoting the sample used in the survey. The vast majority of respondents were selected from a national sample of adults (98.5%), three surveys are of registered voters, and the remainder are from national adult samples with recalls or youth components. For aggregate analysis, scholars will want to exclude those surveys that are based on registered voters. Though some surveys are denoted *National Adult + Youth*, the youths ($<$18) are excluded from the dataset.

- `mipid`: this numeric variable groups the questions based on their wording into two dozen categories (see the "MIPD Codebook–Release 1.0.pdf" for more details).

- `weightavail`: binary variable coded 1 if the population weight was available. If a weight is not available, then all of the respondents are treated as equally likely to be chosen from the population.

- `unique_mip`: this variable provides the number of unique MIP responses by survey. This ranges from 10 to 319.

- `numrespondents`: this variable provides the number of respondents in that survey. This ranges from 450 to 5,791.

- The MIPD coding scheme uses Singer (2011) as its foundation. The dataset is in "long" format, which means that the MIP percentage variable (`perc`) varies across temporal domain, category, problem, and subgroup. The downloaded data from the Dashboard is based on the inputs from the user regarding the time period, temporal domain, subgroup, and category.

  - The `catname` variable identifies the coding scheme:

    - **MIPD (General)**: percentages in the 15 general categories of the MIPD coding scheme (see Table 3). For example, the Economy (1) general category includes all the 100-series specific categories, ranging from Economy (101) to Industrial Policy (110).
    - **MIPD (Specific)**: percentages of the specific categories of the MIPD coding scheme.
    - **Economic**: percentages of the economic specific categories (see the first column of Table 4); primarily made of 100-series but also includes some of the 500-series.
    - **Social Policy**: percentages in social welfare specific categories (200-series) (see the second column of Table 4).
    - **Foreign Policy**: percentages in the foreign policy and national security specific categories (600-series and some 500- and 700-series) (see the third column of Table 4).
    - We also convert our coding scheme to the Policy Agendas Project codes to make it easier for scholars to connect public priorities to policy outputs (**Policy Agendas Project**). For the most part, the PAP scheme grafts nicely onto the MIPD coding scheme (see Table 5). The exceptions are the following:
        - There are a lot of responses in the Morality category (for example, Morals/Values) that have no natural home in the PAP coding scheme, so we code them as Other (22).
        - A similar problem occurs with the Politics category (for example, Corruption), which we code as Government Operations (20).
        - There are only a few MIP responses that are in the Public Lands and Water Management (21), including Native American issues and Natural Resources: Water.
        - There are no natural categories in the PAP for responses that indicate no MIP or don't know, so we create None (23) and Don't Know (24) categories.
- The `problemname` variable provides the label for the problem within the `catname` coding scheme.
- The `subname` variable provides the name of the subgroup:
    - If no subgroup is selected this variable does not appear in the dataset.
    - Partisan identification: includes Democrats, Independents and Republicans (leaners are coded into partisan categories).
    - Gender: includes Males and Females.
    - Income: includes household income divided into quartiles (Lowest 25th, Second 25th, Third 25th, and Highest 25th).
    - Region: splits states into four regional categories:
        - Midwest: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota and Wisconsin.
        - West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington and Wyoming.

    - South: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia and West Virginia.
    - Northeast: Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island and Vermont.
- The `catdesc` and `problemdesc` variables provide short descriptions of the category and problem selected, while the `problemexamples` provides some common responses in that problem.

Table 3: Coding scheme for the Most Important Problem Dataset (MIPD)

Category	Sub-Category	Area?
Economy (100s)		
Economy (101)		
Employment (102)		
Inflation (103)	Gas/Energy (1)	
	Food (2)	
	Housing (3)	
	Health Care (4)	
	Education (5)	
	Transportation (6)	
Monetary Policy (104)		
Growth/Recession (105)		
Domestic Commerce (106)	Financial Services (1)	
	Pro-Business Intervention (2)	
	Business Development (3)	
	Business Ethics (4)	
	Small vs Big Business (5)	
	Stock Market (6)	
Government Regulation (107)		
Employment Compensation (108)	Minimum Wage (1)	
Taxes (109)	Level (1)	
	Fairness (2)	
	Reform (3)	
Industrial Policy (110)	Unions (1)	
	Strikes/Work Stoppages (2)	
	Automation (3)	
Social Policy (200s)		
Social Policy (201)		
Welfare (202)	Support (1)	
	Oppose (2)	
	Reform (3)	
Education (203)	Facilities (1)	
	Teachers (2)	
	Quality (3)	
	Funding for Students (4)	
	Funding for Schools (5)	
	Reform (6)	
	Higher Education (7)	
Health Care (204)	Affordable Care Act (1)	
	Medicare (2)	
	Medicare-for-All (3)	
	Medicaid (4)	
	Prescription Drugs (5)	
Retirement (205)		
Unemployment Insurance (206)		
Poverty (207)		
Inequality (208)		
Classes (209)		
Hunger (210)		
Housing (211)		
Homelessness (212)		
Transportation (213)		
Infrastructure (214)		
Child Care (215)		
Urban Issues (216)		
Specific Groups (217)	Refugees (2)	
	Elderly People (3)	
	Veterans (4)	
	Disabilities (5)	
Rights (300s)		
International Economic Relations (700s)		
Globalization (701)		
Trade (702)	Trade Deficit (1)	Yes
	Consequences (2)	
	Protections (3)	
	Free Trade (4)	
International Cooperation (703)		
Other Economies (704)		Yes
World Economy (705)		
Environment (800s)		
Environment (801)		
Climate Change (802)		
Pollution (803)	Air (1)	
	Water (2)	
	Acid Rain (3)	
Energy Sources (804)	Fossil Fuels (1)	
	Nuclear (2)	
	Renewable (4)	
Lack of Resources (805)	Water (1)	
	Energy (2)	
Natural Disasters (806)		
Public Health (807)	AIDS (1)	
	Cancer (2)	
	SARS (3)	
	COVID (4)	
Morality (900s)		
Morals/Values (901)		
Family Values (902)		
Religion (903)		
Culture (904)		
Abortion (905)		
Church/State Conflict (906)		
Teen Pregnancy (908)		
Feminism (909)		
Politics (1000s)		
Politics (1001)		
Corruption (1002)		
Trustworthiness (1003)		
Leadership (1004)		
Competence (1005)		
Representation/Accountability (1006)	Business Influence (1)	
	Special Interests (2)	
	Elite/wealthy (3)	
	Lower/Middle Class (4)	
	Self-Interested (5)	
	Compensation (6)	
	Accountability (7)	
Dishonesty (1007)		
Scandals (1008)	Watergate (1)	
	Savings and Loan (2)	
	Clinton-Lewinsky (4)	
Political Process (1009)	Campaign Finance (1)	
	Electoral System (2)	
	Electoral Administration (3)	
	Term Limits (4)	
	Two-Party System (5)	
	Apportionment (6)	
Ideology (1010)	Communism (1)	
	Fascism (2)	
	Socialism (3)	

Continued on next page

Table 3 – continued from previous page

Category	Sub-Category	Area?
Gender Rights (301)	Employment (1)	
	Political Participation (2)	
	Attitudes (3)	
LGBTQ+ Issues (302)	Gay Marriage (1)	
	Military (2)	
Reproductive Rights (304)		
Privacy Rights (306)		
Freedom (307)	Support for Constitution (1)	
	Constitutional Rights (2)	
	Symbolic Speech (3)	
Civil Rights (308)	Education (1)	
	Employment (2)	
	Housing (3)	
	Justice System (4)	
	Political Participation (5)	
	Affirmative Action (6)	
Racism/Discrimination (309)	Race Relations (1)	
Non-Violent Participation (310)	College (1)	
	Civil Rights (2)	
	War (3)	
Public Safety (400s)		
Crime (401)		
Law Enforcement (402)	Police Misconduct (1)	
	Support for Police (2)	
	Punishment (3)	
	Death Penalty (4)	
Gun Control (403)	Support (1)	
	Oppose (2)	
Drugs (404)		
Alcohol and Smoking (405)	Alcohol (1)	
	Smoking (2)	
Violence against Women (406)		
Political Violence (407)	College (1)	
	Race-Based (2)	
Fiscal Policy (500s)		
Budget (501)		
Debt (502)		
Deficit (503)		
Government Spending (504)	Too Little (-1)	
	Neutral (0)	
	Too Much (1)	
Social Spending (505)	Too Little (-1)	
	Neutral (0)	
	Too Much (1)	
Defense Spending (506)	Too Little (-1)	
	Neutral (0)	
	Too Much (1)	
Space Policy (507)		
Agricultural Policy (508)		
Immigration (509)	Policy (1)	
	Border Security (2)	
	Effects (3)	
	Government Assistance (4)	
	Rights (5)	
Foreign Policy (600s)		
Foreign Policy (601)	Isolationist (1)	Yes
	Multilateralist (2)	
	Position (3)	
National Security (602)		
Nuclear Weapons (603)	Arms Control (1)	Yes
	Proliferation (2)	
	Nuclear War (3)	
	Partisanship (4)	
	Conservativism (5)	
	Liberalism (6)	
	Republicans (7)	
	Democrats (8)	
	Federalism (9)	
	American Ideals (10)	
	Acceptance (11)	
	Division (12)	
Political Institutions (1011)	Government (1)	
	President (2)	
	Congress (3)	
	Supreme Court (6)	
	Courts/Judiciary (7)	
	Media (8)	
	CIA and FBI (9)	
	Bureaucracy (10)	
	Election (11)	
Political Actors (1012)	President: Number	
	Others (99)	
Experts/Science (1013)		
Youth Issues (1100s)		
Youth (1101)		
Youth Crime (1102)		
Gangs (1103)		
School Violence (1104)		
Employment (1105)		
Crimes against Youths (1106)		
Minority Groups (1200s)		
Minorities (1201)	Native Americans (1)	
	Hispanics (2)	
	Asian Americans (3)	
	Jews (4)	
	Muslims (5)	
Other (1300s)		
Other (1301)		
All (1302)		
None (1400s)		
None (1401)		
Don't Know (1500s)		
Don't Know (1501)		
No Opinion (1502)		
Not Applicable (1503)		
Refused (1504)		

Continued on next page

Table 3 – continued from previous page

Category	Sub-Category	Area?	Category	Sub-Category	Area?
Terrorism	Al-Qaeda (1)	Yes			
	Islamic State (2)				
	Islamic Terrorism (3)				
	Domestic (4)				
	General (5)				
War (605)	Correlates of War code	Yes			
Crises (606)	Cuban Missile Crisis (1)	Yes			
	Iran Hostages (2)				
	Peace Talks (3)				
	Cold War (4)				
Foreign Aid (608)	Anti-Foreign Aid (1)	Yes			
	Pro-Foreign Aid (2)				
World Hunger (609)					
Other Countries (610)		Yes			
Foreign Actors (611)		Yes			

Note: "Area" indicates categories that also code the countries/regions.

Table 4: Coding scheme for economic, social policy, and foreign policy categories (MIPD coding scheme values in parentheses)

Economic	**Social Policy**	**Foreign Policy**
Economy (101)	Social Policy (201)	Defense Spending (506)
Employment (102)	Welfare (202)	Foreign Policy (601)
Inflation (103)	Education (203)	National Security (602)
Monetary Policy (104)	Health Care (204)	Nuclear Weapons (603)
Growth/Recession (105)	Retirement (205)	Terrorism (604)
Domestic Commerce (106)	Unemployment Insurance (206)	Wars (605)
Government Regulation (107)	Poverty (207)	Crises (606)
Employment Compensation (108)	Inequality (208)	Foreign Aid (608)
Taxes (109)	Classes (209)	World Hunger (609)
Industrial Policy (110)	Hunger (210)	Other Countries (610)
Budget (501)	Housing (211)	Foreign Actors (611)
Debt (502)	Homelessness (212)	Globalization (701)
Deficit (503)	Transportation (213)	Trade (702)
Government Spending (504)	Infrastructure (214)	International Cooperation (703)
Social Spending (505)	Child Care (215)	Other Economies (704)
Defense Spending (506)	Urban Issues (216)	World Economy (705)
	Specific Groups (217)	

Table 5: Coding scheme for the Comparative Agendas Project (CAP) and the MIPD

CAP	MIPD	CAP	MIPD
Macroeconomics (1)	Economy, Employment, Inflation, Monetary Policy, Growth/Recession, Domestic Commerce (2,3), Taxes, Industrial Policy, Inequality, Classes, Budget, Debt, Deficit, Government Spending	Social Welfare (13)	Social Policy, Welfare, Retirement, Poverty, Hunger, Child Care, Elderly People, Disabilities, Social Spending
Civil Rights, Minority Issues and Civil Liberties (2)	Gender Rights, LGBTQ+ Issues, Reproductive Rights, Privacy Rights, Freedom, Civil Rights, Racism/Discrimination, Participation, Church/State Conflict, Minority Groups	Community Development and Housing Issues (14)	Housing, Homelessness, Urban Issues
Health (3)	Health Care, Alcohol and Smoking, Public Health	Banking, Finance and Domestic Commerce (15)	Domestic Commerce (1,4,5,6), Government Regulation, Natural Disasters
Agriculture (4)	Agricultural Policy	Defense (16)	Veterans, Defense Spending, National Security, Nuclear Weapons, Terrorism, War, Crises
Labor and Employment (5)	Employment Compensation, Unemployment Insurance, Youth Employment	Space, Science, Technology and Communications (17)	Space Policy, Experts/Science
Education (6)	Education	Foreign Trade (18)	Globalization, Trade
Environment (7)	Environment, Climate Change, Pollution	International Affairs and Foreign Aid (19)	Foreign Policy, Foreign Aid, World Hunger, Countries, Foreign Actors, International Cooperation, Other Economies, World Economy
Energy (8)	Energy Sources, Lack of Resources (2)	Government Operations (20)	Politics, Corruption, Trustworthiness, Leadership, Competence, Representation/Accountability, Dishonesty, Scandals, Political Process, Ideology, Institutions, Political Actors
Immigration (9)	Refugees, Immigration	Public Lands and Water Management (21)	Lack of Resources (1), Native Americans
Transportation (10)	Transportation, Infrastructure	Other (22)	Morals/Values, Family Values, Religion, Culture, Abortion, Feminism, Other, All
Law, Crime and Family Issues (12)	Crime, Law Enforcement, Gun Control, Drugs, Violence against Women, Political Violence, Teen Pregnancy, Youth, Youth Crime, Gangs, School Violence, Crimes against Youths	None (23)	None

Note: numbers in parentheses for MIPD categories are sub-category values.