

# Gallup's Most Important Problem Data Codebook

---

Policy Agendas Project  
2013

## ***Contact Information***

Visit [www.policyagendas.org](http://www.policyagendas.org) or contact us at [policyagendas@gmail.com](mailto:policyagendas@gmail.com).

## ***Citation Guidelines***

Please visit [www.policyagendas.org/page/how-cite](http://www.policyagendas.org/page/how-cite) for guidelines on how to cite the use of this or other Policy Agendas Project datasets.

## Overview of the Data Set

This data set codes public opinion polls from Gallup's Most Important Problem Survey using the Policy Agendas Project's content coding scheme. The data are available in both year and quarterly formats. Currently, only yearly data may be accessed through the Trends Analysis Tool.

## Content Coding

We coded all entries by the policy content of the poll responses, along with other variables of interest. We employed the standard Policy Agendas Project topic scheme in our policy content coding (with the addition of a category for those answering 'don't know' or an 'other' topic: major topic 25) and each entry was assigned one and only one content code. This allows researchers to trace activity on a particular topic across the period. The full description of our content categories may be found in the major topics codebook at <http://www.policyagendas.org/page/topic-codebook>.

## The Averaged Yearly Summary

This worksheet captures the aggregated proportions for each major category, on an annual basis, for all of the polls contained in the working data set. These annualized proportions were constructed by normalizing the percentage of responses in every major topic (e.g. Macroeconomics) by the total percentage of responses in any given year.

Since many of the individual poll takers permitted respondents to specify multiple answers to the MIP polling question, we needed a way to normalize across polls with varying total percentages of responses (many individual polls reported responses in excess of 100 percent). First, we summed all the reported response percentages, across all 23 major topics, for each poll. We then divided individual percentages by a divisor which was based on the summed total of the percentages for this poll, to create the normalized percentage for that major topic. This normalized percentage is the actual proportion of all responses that were in a single major topic. The normalized percentages for each poll now add to 100 percent (as opposed to 150 percent for poll A, 123 percent for poll B, and 100 percent for poll C, etc.). We then averaged these normalized percentages across the number of polls in a given year to get the annualized proportions reported in this worksheet.

For example, if the sum of the 23 major topic percentages were 150 percent in poll A (for simplicity, this example assumes that the results for the other 18 major topics were all 0 percent), then individual percentages for each major topic (e.g. 40 percent for Macroeconomics) were divided by this total to normalize the results to 100 percent of all reported responses.

Mathematically, these transformations looked like this:

Major Topic	Topic Name	Raw %	Divisor	Normalized %
1	Macroeconomics	40	100/150	26.67
2	Civil Rights	30	100/150	20.00
3	Health	20	100/150	13.33
4	Agriculture	25	100/150	16.67
5	Labor	35	100/150	23.33
Total	--	150	--	100.00

The individual columns in this worksheet are described in detail below:

## Variable Names and Descriptions:

### ***KeyID***

This column records the unique identifier for each observation. It has no substantive application.

### ***Year***

This column refers to the year of the corresponding MIP entry. Note that data are missing (no question was asked by Gallup) for 1953 and 1955.

### ***Percentage***

This column refers to the aggregated *proportions* for each major category, on an annual basis, for all of the polls contained in the dataset. Note this variable is a *proportion* and is not recorded as a percentage (i.e. multiplied by 100) in the database. We will fix this labeling error soon.

### ***MajorTopic***

This column records the Policy Agendas Project's major topic code that corresponds to the information found in the Description column. The codes and examples for this column are found in the Major Topics Codebook on the Policy Agendas Project website.

## The Quarterly Summary

Data aggregated at the quarterly level are also available, and the file is described below:

This worksheet captures the aggregated proportions for each major category, on a quarterly basis, for all of the polls contained in the working data set. These annualized proportions were constructed by normalizing the percentage of responses in every major topic (e.g. Macroeconomics) by the total percentage of responses to a single poll, and then averaging these proportions across the multiple polls in any given quarter. Because of the aforementioned low frequency of polling in the earlier years and the absence of polls in five of these early years, the quarterly data begins in the second half of 1956.

Since many of the individual poll takers permitted respondents to specify multiple answers to the MIP polling question, we needed a way to normalize across polls with varying total percentages of responses (many individual polls reported responses in excess of 100 percent). First, we summed all the reported response percentages, across all 23 major topics, for each poll. We then divided individual percentages by a divisor which was based on the summed total of the percentages for this poll, to create the normalized percentage for that major topic. This normalized percentage is the actual proportion of all responses that were in a single major topic. The normalized percentages for each poll now add to 100 percent (as opposed to 150 percent for poll A, 123 percent for poll B, and 100 percent for poll C, etc.). We then averaged these normalized percentages across the number of polls in a given quarter to get the quarterly proportions reported in this worksheet.

For example, if the sum of the 23 major topic percentages were 150 percent in poll A (for simplicity, this example assumes that the results for the other 18 major topics were all 0 percent), then individual percentages for each major topic (e.g. 40 percent for Macroeconomics) were divided by this total to normalize the results to 100 percent of all reported responses. Mathematically, these transformations looked like this:

Major Topic	Topic Name	Raw %	Divisor	Normalized %
1	Macroeconomics	40	100/150	26.67
2	Civil Rights	30	100/150	20.00
3	Health	20	100/150	13.33
4	Agriculture	25	100/150	16.67
5	Labor	35	100/150	23.33
Total	--	150	--	100.00

The number of polls in a given quarter, from the second quarter of 1956 to the first quarter of 2001, varies from no polls to six polls. By way of interpolation we constructed a complete dataset. As stated above, we averaged the normalized percentages across the number of polls to reach the quarterly proportions. However, in the case of no polls in a given quarter the normalized percentages from the last poll in the previous quarter and the first poll in the preceding quarter were averaged to achieve results.

The quarterly data follow the below format:

***Year***

The year column lists the year of the polls that were aggregated<sup>1</sup>.

***Quarter***

The quarter column lists the quarter (within the given year) of the polls that were aggregated.

***Proportion of Responses***

- P1: Quarterly proportion of responses that fell within the Macroeconomics topic.
- P2: Quarterly proportion of responses that fell within the Civil Rights topic.
- P3: Quarterly proportion of responses that fell within the Health topic.
- P4: Quarterly proportion of responses that fell within Agriculture topic.
- P5: Quarterly proportion of responses that fell within the Labor topic.
- P6: Quarterly proportion of responses that fell within the Education topic.
- P7: Quarterly proportion of responses that fell within the Environment topic.
- P8: Quarterly proportion of responses that fell within the Energy topic.
- P10: Quarterly proportion of responses that fell within the Transportation topic.
- P12: Quarterly proportion of responses that fell within the Crime topic.
- P13: Quarterly proportion of responses that fell within the Social Welfare topic.
- P14: Quarterly proportion of responses that fell within the Housing & Development topic.
- P15: Quarterly proportion of responses that fell within the Domestic Commerce topic.
- P16: Quarterly proportion of responses that fell within the Defense topic.
- P17: Quarterly proportion of responses that fell within the Science & Technology topic.
- P18: Quarterly proportion of responses that fell within the Foreign Trade topic.
- P19: Quarterly proportion of responses that fell within the International Affairs topic.
- P20: Quarterly proportion of responses that fell within the Government Operations topic.
- P21: Quarterly proportion of responses that fell within the Public Lands topic.
- P25: Quarterly proportion of responses that fell within the Don't Know/Other topic.

---

<sup>1</sup> Take note that for certain early years of this dataset there were no polls conducted, therefore, those years are denoted with a -9 in the spreadsheet. The years with no polls include: 1941, 1943, 1944, 1953, and 1955. It should also be noted that in certain early years only one poll was held in a given year. Therefore, the results of the one poll are reported as the average for that year. Such years are: 1942, 1950, 1951, 1952, 1954, and 1963. This should be considered when using the fiscal year sheet as well.