

## Essential Information:

1. Title and Author(s) of article
2. If the codebook covers multiple datasets, please divide the codebook into sections where each section corresponds to one of the datasets.
3. All variables used in the analysis dataset(s) must be fully defined within the codebook. If there are variables in the analysis data that are not used to produce results within the manuscript or appendices, they should be removed from the dataset. There must be a 1:1 match between variables in codebook and variables in analysis data.
4. All values (including missing values) must be included for each variable with complete value labels. This includes any categorical, dichotomous, binary, and/or ordinal variables.
5. Must be saved as a PDF.
6. If full data citations for original data sources have not been included in the README, they must be included in the codebook. We recommend placing data citations in a References section at the bottom of the codebook.

## Codebook Template

[Article Title]

[Authors] – [Affiliation] – [E-mail] – [Twitter (optional)]

### A. Codebook for datafile.dta

Variable	Variable Definition	Value & Value Label (s)	Source
Gender	Respondent is male or female	1 = Male 2 = Female	
Age	Age of respondent	1 = 18 – 25 2 = 26 – 30 3 = 31 – 35 4 = 36 – 40 5 = 40 & older . = Missing	
Female_labor	Percentage of women on the labor market		The World Bank (2017)
ChRef (binary)	Whether the sponsor is Chair of the committees to which the bill is referred	1 = Sponsor is Chair of one of the committees to which the bill is referred 0 = Otherwise	

### B. References

The World Bank. 2017. “Gender Data Portal.”  
<http://datatopics.worldbank.org/gender/about> (August, 2017).



## R PRO-TIP

Both the ‘code book’ and ‘dataMaid’ library packages can be used to generate codebooks. Again, please review the output to ensure that all variables, variable definitions, values and value labels are present



## STATA PRO-TIP

The 'codebook' or 'label book' commands can be used to generate some of the codebook. These commands depend on the robustness of your data within Stata so you will need to review the output to see if any variables, variable definitions, values or value labels are missing. You will also need to add original data source citations to the final product (if they have not been included in the README)

## What should I include in a data citation?

A data citation should lead a secondary user directly back to the original source data. We recommend using ICPSR's Citing Data as guidance: <https://www.icpsr.umich.edu/web/pages/datamanagement/citations.html>

Here is an example of a complete data citation:

**US Census Bureau. 1993-2011. Public Elementary-Secondary Education Finance Data.**  
<https://www.census.gov/programs-surveys/school-finances/data/tables.html>



## PRO-TIP

The citation above tells you the data producer or author (US Census Bureau), the date of the data collection (1993 - 2011), the title of the dataset (Public Elementary-Secondary Education Finance Data) and provides a direct link to the tables used.

The user should be able to follow the citation and immediately gain access to the data. In some cases, this may not be possible. For example, with online databases, you may not be able to directly link a user back to the original data; however, you can provide the data citation alongside instructions with the key terms or query used to extract the original data.

If the data cannot be directly accessed due to access restrictions, licensing agreements, or copyright, please provide a data citation to the source of the original data as well as detailed instructions for requesting access to the exact data file(s) you used.

If you have additional questions or concerns about data access restrictions, please contact the SPPQ Editors.

## Examples of complete and successful codebooks:

1. Kitchens, Karin, 2020, "Replication Data for: Partisan Politics and Public Education: Finding the Formula for (Electoral) Success", <https://doi.org/10.15139/S3/GoWQQR>, UNC Dataverse, V1, UNF:6:Dgld/5Qb7sVJh7onKNDJmA== [fileUNF]
2. Sung Eun Kim; Johannes Urpelainen; Joonseok Yang, 2020, "Replication Data for: State Policy and Lobbying in a Federal System: Evidence from the Production Tax Credit for Renewable Energy, 1998-2012", <https://doi.org/10.15139/S3/KUT29A>, UNC Dataverse, V1, UNF:6:8vOafbQnnFLzjmCERFoNWg== [fileUNF]
3. Cohen, Jeffrey, 2020, "Replication Data for: Relative Unemployment, Political Information and the Job Approval Ratings of State Governors and Legislatures", <https://doi.org/10.15139/S3/WK5EK6>, UNC Dataverse, V1, UNF:6:2ZnrhmdHcf3dyZYAC+pVQA== [fileUNF]
4. Karch, Andrew, 2020, "Replication Data for: Bill Content, Legislative Outcomes, and State-Level Resistance to National Policies", <https://doi.org/10.15139/S3/PLKGGGA>, UNC Dataverse, V1, UNF:6:oKsexBbLwflBNKYgiBtQ6Q== [fileUNF]
5. Butcher, Jordan; Squire, Peverill, 2020, "Replication Data for: An Update to the Squire State Court of Last Resort Professionalization Index", <https://doi.org/10.15139/S3/OIAAIZ>, UNC Dataverse, V1, UNF:6:RpIPGm8WU5nf5BUfVT/zw== [fileUNF]