

READ ME – Data and Code Supplement for:

“Media Slant and Public Policy Views”

Open ICPSR - 198941

Contributors:

Milena Djourelova (corresponding author), Cornell University. Email: mnd344@cornell.edu.

Ruben Durante, National University of Singapore, BSE, IZA, CESifo, and CEPR.

Elliot Motte, Universitat Pompeu Fabra.

Eleonora Patacchini, Cornell University, EIEF, IZA, and CEPR.

Overview

This folder contains the data, Stata code, and python code required to replicate the analyses in the “Media Slant and Public Policy Views” paper for the 2024 *AEA Papers and Proceedings*.

The folder contains several subfolders and files in addition to this readme document:

- 00_programs: Stata ado files
- 01_code/: Stata dofiles and Python scripts
- 02_data/: raw and final data
- 03_output/: Tables and Figure in paper
- requirements.txt: Python packages to install before replication
- data_files_summary.xlsx: list of all data files used in replication code

Statement about Rights

The authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

The authors of the manuscript have documented permission to redistribute and publish the data contained within this replication package.

Details on data sources and files

Data.Name	Data.Files	Location	Provided
TV segments	climate_allChannels_fulltext.csv ; immigration_allChannels_fulltext.csv; guns_allChannels_fulltext.csv; abortion_allChannels_fulltext.csv	02_data/01_TV_segments/01_raw/	TRUE

Cooperative Election Study	cumulative_2006-2021.dta; cumulative_cces_policy_preferences.dta	02_data/02_CCES/	TRUE
NIELSEN Focus Report 2012	Proprietary	NA	FALSE
County-Fips crosswalk	US_county_fips.txt	02_Data/03_controls/	TRUE
1996 Presidential Elections	presidential.dta	02_Data/03_controls/	TRUE
US Census 2010 county controls	county_controls_2010.dta	02_Data/03_controls/	TRUE
US Census 2010 zipcode-county crosswalk	zcta_county_rel_10.txt	02_Data/03_controls/	TRUE

The TV segments data are available in this present repository. The raw data was collected by the authors, and annotations of this data were generated with GPT-4, OpenAI's large-scale language-generation model. Upon generating the input prompts and output GPT completions the authors reviewed, edited, and revised the language to their own liking and take ultimate responsibility for the content of these annotations. Input prompts are located in 01_code/01_data_prep/01_GPT_questions.

The CES datasets were downloaded from the Harvard Dataverse repository. They are available at the URLs indicated in the data_files_summary Excel file. Users should properly cite Dagonel (2021) if using the cumulative_cces_policy_preferences.dta. Files should be placed under 02_data/02_CCES/.

The NIELSEN Focus Report 2012 data are proprietary and were acquired by the authors. They are not provided as part of this archive. Researchers interested in access to the data may contact Mr. Jonathan Wells, one of Nielsen's client managers, at jonathan.w.wells@nielsen.com. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Data used for county-level merging (fipscode-to-county crosswalk, zipcode-to-county crosswalk) and county-level controls from the US 2010 Census were downloaded from the source URLs listed in the data_files_summary Excel file. They are in the public domain and are provided as part of this archive. County-level vote share information for the 1996 presidential elections are from David Leip's US Election Atlas. We provide them here and they can also be acquired by visiting the US Election Atlas Web store at https://uselectionatlas.org/BOTTOM/store_data.php, looking under the "General Election Results by County" section, "US President General – County Level Vote Data" subsection, and the "1996 President" file.

Dataset list

See the data_files_summary Excel file for a full list of the datasets used by the code files in this archive.

Computational requirements

Software Requirements

- Stata (code was last run with version 15)
 - estout (as of 2019-05-31)
 - ftools (as of 2019-08-16)
 - reghdfe (as of 2019-11-13)
 - statastates (as of 2018-01-05)
 - zscore (as of 2010-07-22)
 - the program "0_setup.do" will install all dependencies locally, and should be run once when launching the 0_master.do
- Python (code was run on version 3.9.18)
 - the file "requirements.txt" lists all Python dependencies required. Please run "pip install -r requirements.txt" in a fresh conda environment as the first step. To set up a fresh conda environment, we recommend installing an Anaconda distribution (visit <https://www.anaconda.com/download>), opening an Anaconda prompt and typing `conda create -n <environment_name> python=3.9.18`. Then, activate the environment through `conda activate <environment_name>`, and navigate to this replication's package directory before installing the requirements.

Description of programs/code

- Programs in `programs/01_data_prep` will extract and reformat all datasets referenced above. Note that an openAI API key is needed for execution of the `02_GPT_labelling.py` script. Researchers interested in automatically using GPT models for data annotation should follow the steps indicated in OpenAI's quickstart guide to using the API at <https://platform.openai.com/docs/quickstart?context=python>
- Programs in `programs/02_analysis` generate all tables and figures in the main body of the article. The program `programs/02_analysis/0_master.do` will run them all.
- Programs in `programs/03_appendix` will generate all tables and figures in the online appendix. The program `programs/03_appendix/master_appendix.do` will run them for Tables A2 and A3. Table A1 is created manually based on output of `/03_appendix/Table_A1.py`
- Ado files have been stored in `00_programs/ado` and the `01_code/0_setup.do` files set the ADO directories appropriately.
- The program `programs/00_setup.do` will populate the `programs/ado` directory with updated ado packages, but for purposes of exact reproduction, this is not needed. The file `programs/00_setup.log` identifies the versions as they were last updated.

Controlled Randomness

Random seed for reproducibility of openAI GPT completions is set at line 70 of the `02_GPT_labeling.py` script. See the openAI API documentation for further details on pseudo-randomness within the GPT class of models.

The random seed for analysis in Stata is set at line 32 of the `0_setup.do`.

Memory and Runtime Requirements

The code was last run on a 6-core Intel-based laptop with 32GB of RAM and 500 GB of HDD, Windows 11.

Runtime of `0_master.do` file to produce all tables and figure: 7 minutes.

Runtime of `01_get_tv_transcripts.py` and `02_GPT_labelling.py`: 38 hours.

Instructions to Replicators

- Edit the 2nd, 6th and 7th lines of code in `01_code/0_setup.do` to adjust paths
- Run program `01_code/0_setup.do` once on a new system to set up the working environment for Stata.
- Download or acquire the data files referenced above. If obtained, the Nielsen data should be placed under `02_data/04_NIELSEN/`.

- Run program 01_code/02_analysis/0_master.do to replicate the tables and figure in the main body of the manuscript.
- Run program 01_code/03_appendix/master_appendix.do to replicate Tables A2 and A3 of the online appendix.

List of tables and programs

The code in the 01_code/ folder reproduces tables and figures in the paper, as explained below.

Figure/Table #	Program	Output file
Table 1	02_analysis/Table1.do	Table1.tex
Table 2	02_analysis/Table2_PanelA.do	Table_2A.tex
	;	;
	02_analysis/Table2_PanelB.do	Table_2B.tex
Table A1	Manually created based on output from 03_appendix/Table_A1.py	
Table A2	03_appendix/Table_A2.do	Table_A2.tex
Table A3	03_appendix/Table_A3.do	Table_A3.tex
Figure 1	02_analysis/Figure_1.do	Figure_1.eps ; Figure_1.pdf

References

Dagonel, A. (2021): “Cumulative CCES Policy Preferences,” Harvard Dataverse.
