2022 Report by the Data Editor

This repository contains the code, data, and manuscript files for the 2022 report by the AEA Data Editor. If you are reading this on openICPSR, then only code and data are present.

Statement about rights

Raw data come from the JIRA system used by the AEA Data Editor and cannot be made available outside of the organization, as it contains names of replicators, manuscript numbers, and verbatim email correspondence. Anonymized data is publicly available at Vilhuber (2023).

- ☑ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☑ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permissions are documented in the `LICENSE.txt` file.

Locations

The repository at https://github.com/AEADataEditor/report-aea-data-editor-2022 contains text, code, data, and output from running the code.

The deposit at http://doi.org/10.3886/E189602V1 contains code and data, as well as output.

Citing the report

> Vilhuber, Lars. 2023. "Report by the AEA Data Editor." AEA Papers and Proceedings.

```
@article{ReportDE2023,
Author = {Vilhuber, Lars},
Title = {Report by the {AEA} Data Editor},
Journal = {AEA Papers and Proceedings},
Volume = {},
Year = {2023},
Month = {},
Pages = {},
DOI = {},
URL = {}}
```

Citing the code and data

> Vilhuber, Lars. and Linda Wang. 2023. “Code and Data for: Report for 2022 by the AEA Data Editor.” American Economic Association [publisher], http://doi.org/10.3886/E189602V1

Data

Summary of Availability

- ☐ All data **are** publicly available.
- ☑ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Data for pre-production verification

Anonymized files from the internal production system are provided in this repository, sourced from Vilhuber (2023). A copy is provided as part of this archive.

```
data/jira/anon/jira.anon.RDS
data/jira/anon/README.md
```

Data on lab members' names is directly downloaded from the Github repository associated with Vilhuber (2023), see `programs/config.R`.

openICPSR data on deposits (ICPSR, 2022)

The data are obtained on demand from the internal systems underlying the AEA Data and Code repository. The internal systems are accessible only to ICPSR staff, and were provided to the AEA Data Editor upon request. They are not accessible to others. The data were lightly hand-edited to account for formatting errors (double double-quotes and other issues related to the conversion from internal database representation to CSV).

Data were extracted on all published replication packages. The data files are in the folder `data/icpsr`. For additional use of the data, see the processing code.

Data on processing time

The data on processing times were extracted from the ScholarOne manuscript management system used by the AEA. Microdata are not available (even to the author), only summary statistics are provided as Excel sheets. These were simply reformatted for the report.

```
data/scholarone/dataEditorReport_20201128-20211127Revised.xlsx
```

Computational requirements

Software Requirements

- R 4.2.2
  - Package versions set to as-of **2021-12-01**, using the Rstudio Package Manager, except for Github installed versions
  - dplyr
  - here
  - tidyr
  - tibble
  - stringr

- readr
- splitstackshape
- digest
- remotes
- readxl
- writexl
- ggplot2
- ggthemes
- janitor
- dataverse
- "data.table" (github)
- "Rdatatable" (github)
- github("markwestcott34/stargazer-booktabs") (overrides standard stargazer!)

Packages are installed by `global-libraries.R`, and are sourced in the Dockerfile.

Programs

All programs are in the `programs` subdirectory:

```
programs/01_lab_members.R
programs/03_jira_dataprep.R
programs/04_table1_compliance.R
programs/05_table2_stats.R
programs/06_table3_response_options.R
programs/07_table4.R
programs/08_figure1_filesize.R
programs/09_table5_webstats.R
programs/99_write_nums.R
programs/config.R
programs/README.md
programs/README.pdf
programs/run_all.sh
programs/zenodo_pull.py
```

Running code

Each R file can be run independently (separate R sessions), in numerical order. See the `programs/README.md` file for further details.

The Python file `zenodo_pull.py` should be run manually.

The script `run_all.sh` is used within a (Linux) shell to implement the above run order, but is optional.

Mapping tables and figures to article

Table and figure numbers in the paper do not map to program names, due to editorial decisions. The table below maps files, figures/tables, and the programs used to generate them.

Name of file	Figure/ Table in article	Program to create
jira_response_options_mod.tex	Table 1	06_table3_response_options.R
n_journal_numbers_mod.tex	Table 2	05_table2_stats.R
plot_rounds_compare.png	Figure 1	07_table4.R
n_rounds.tex	Table 3	07_table4.R
n_webstats.tex	Table 4	09_table5_webstats.R
plot_filesize_dist.png	Figure 2	08_figure1_filesize.R
n_compliance_manuscript_mod.tex	Table 5	04_table1_compliance.R
n_updates_manuscript_mod.tex	Table 6	04_table1_compliance.R

In-text numbers are collected throughout all programs, and written out in `programs/99_write_nums.R` to `tables/latexnums.tex`.

License

See LICENSE.txt for data and code license.

References

- Vilhuber, Lars. 2023. "Process data for the AEA Pre-publication Verification Service." American Economic Association [publisher], https://doi.org/10.3886/E117876V4