

README

The empirical analysis can be fully replicated with the following datasets and R-scripts:

- “Data_Prep_Steinert.R”
- “Data_Analysis_Steinert.R”
- “Map_China_Steinert.R”
- “Validation_CHRD_Steinert.R”
- “complete_pprisdat.RData”
- “final_pprisdat.RData”
- “CECC_Version_23.07.2020.csv”
- “CHRD_fullinfo.csv”
- “complete_datevector.csv”

The key data source of the empirical analysis is the Chinese Political Prisoner Database compiled by the US Executive-Congressional Commission on China (see main article for more information). The dataset has been retrieved online on July 27, 2020 and this version of the dataset is stored in the file “CECC_Version_23.07.2020.csv”. Note that the Chinese Political Prisoner Database is continuously updated, thus, the information in the current version of the dataset might not necessarily align with the version from July 27, 2020.

Data preparation:

The raw data “CECC_Version_23.07.2020.csv” has been cleaned and prepared for the empirical analysis in the R-Script entitled “Data_Prep_Steinert.R”. After finishing the data preparation, I stored the cleaned data in the R data frames called “complete_pprisdat.RData” and “final_pprisdat.Rdata”. The former data frame contains all Chinese political prisoners that have been stored in the dataset at the date of retrieval. The latter object contains the subset of those Chinese political prisoners, where the information on the duration of political imprisonment is available. Drawing on “final_pprisdat.Rdata”, I hand-coded the criminalized action variables to the cleaned dataset.

Data analysis:

The data analysis can be fully reproduced with the R-Script “Data_Analysis_Steinert.R”. The R-Script makes use of the cleaned data frames “complete_pprisdat.RData” and “final_pprisdat.Rdata”. I also use the datafile “complete_datevector.csv” to create a monthly vector, which I use for the Figure 5 in the Online Appendix. “Data_Analysis_Steinert.R” reproduces the Figures 1 & 3-8 from the main article. It also reproduces the Figures 1-7 from the Online Appendix. Further, it reproduces the Tables 1-10 and 14 from the Online Appendix.

Validation with CHRD-data:

I conducted a data validation exercise with data provided by the NGO *Chinese Human Rights Defenders* as presented in the Online Appendix section 5. The data validation can be reproduced with the dataset

“CHRD_fullinfo.csv” and the R-Script “Validation_CHRD_Steinert.R”. This procedure reproduces the Table 13 presented in the Online Appendix.

Mapping the duration of political imprisonment across Chinese provinces:

I created a map of the duration of political imprisonment across Chinese provinces. This map presented in Figure 2 in the article can be reproduced with the R-Script “Map_China_Steinert.R”. Note that the R-Script may create problems due to the Chinese character encodings. In case the Chinese characters are not correctly displayed, I suggest to copy-paste the text in the PDF-document “Map_China_ChineseEncoding.pdf” to a new R-Script. Additionally, I suggest to change the encoding to “x_Chinese-Eten” in R-Studio.