Readme File for Replication Package of:

"Historical Family Types and Female Political Representation: Persistence and Change", by Aina Gallego, Didac Queralt, and Ana Tur-Prats

The Journal of Politics

1. Folder Contents

The replication folder includes data and scripts in two languages: Stata (.dta and .do) and R (.r). The replication files are organized in two folders:

Folder	Subfolder	Content
Codes		01_main_data_construction.do 02_master_results.do Figure_4_labor_market.do Figure_5_gender_attitudes.do Figure_A2_1860_2001_census.do Table_A9_female_literacy_1887.do Figure_A11_councilors_historic.do
Datasets	00_rawdata	regidors_1979.dta regidors_1983.dta regidors_1987.dta regidors_1991.dta regidors_1995.dta regidors_1999.dta regidors_2003.dta regidors_2007.dta regidors_2011.dta regidors_2015.dta epa_1979_q1_replication.dta epa_1979_q2_replication.dta epa_1983_q1_replication.dta epa_1983_q2_replication.dta epa_1987_q1_replication.dta epa_1987_q2_replication.dta epa_1991_q1_replication.dta epa_1991_q2_replication.dta epa_1995_q1_replication.dta epa_1995_q2_replication.dta epa_1999_q1_replication.dta epa_1999_q2_replication.dta epa_2003_q1_replication.dta epa_2003_q2_replication.dta epa_2007_q1_replication.dta

		epa_2007_q2_replication.dta epa_2011_q1_replication.dta epa_2011_q2_replication.dta epa_2015_q1_replication.dta epa_2015_q2_replication.dta cis_1092_replication.dta cis_1867_replication.dta cis_2194_replication.dta cis_2448_replication.dta cis_2597_replication.dta cis_2831_replication.dta cis_2911_replication.dta cis_3182_replication.dta cis_3273_replication.dta metroscopia_15077_replication.dta women_councilor_names.dta census_1860.dta census_1860_2001.dta census1887.dta municipal_controls.dta pop_1930_ine.dta pop.dta
	01_processeddata (the .dta files in this folder are generated by running the do files above)	main_analysis_dataset.dta epa_prov_year_gender.dta total_women_democ.dta women_councilors_all
RStudio	shapefiles	SECC_CE_20220101.CPG SECC_CE_20220101.dbf SECC_CE_20220101.prj SECC_CE_20220101.sbn SECC_CE_20220101.sbx SECC_CE_20220101.shp SECC_CE_20220101.shx
		Figure_2B_family_map.R
Figures		
Tables		

The folders *Figures* and *Tables* are empty and generated with the purpose of storing the results.

2. Software and Operation System

Most tables and figures have been generated using *StataNow 18.5 MP-Parallel Edition.* The analysis has been run on a computer that uses *Windows 11 Pro Education, Version 23H2*. Commands `preserve` and `restore` will only function in *MP* versions of *Stata* (as of early 2025).

To run the do files, the following commands need to be installed:

```
ssc install egenmore
ssc install blindschemes
ssc install coefplot
ssc install reghdfe
ssc install ivreg2
ssc install ranktest
ssc install acreg
ssc install ranktest
ssc install hdfe
ssc install norm
ssc install ciplot
ssc install outreg2
ssc install ftools
```

Figure 2(b) was generated using *R* (version 4.4.2 for Windows) in conjunction with *RStudio* (version 2024.12.1+563). An additional program, *RTools* (version 4.4), is required to plot the Figure 2b—a map. This analysis was run on a computer that uses *Windows 11*.

To run the script, the following commands need to be installed:

```
install.packages("sp")
install.packages("raster")
install.packages("tmap")
install.packages("dplyr")
install.packages("sf")
install.packages("ggmap")
install.packages("tibble")
install.packages("writexl")
install.packages("readxl")
install.packages("grid")
install.packages("RColorBrewer")
install.packages("viridisLite")
install.packages("readr")
install.packages("haven")
```

3. Instructions for Replication

After downloading the replication package, follow the following steps:

In Stata:

a) Open Stata, and the do file. Change the path to the user's location. This needs to be done at the beginning of the following do files: "01_main_data_construction.do" and "02_master_results.do". There is a note at the beginning of each do file that reads "`* New users change this path:`"
b) Run the do files in the following order:
   i. "01_main_data_construction.do" generates the main dataset
   ii. "02_master_results.do" is a master do file that generates all tables and figures in the paper and appendix. When using additional data to generate these results, the do file "02_master_results.do" invokes other do files. These additional do files are not numbered and the figure or table they generate is indicated in the file name (e.g., "Figure_5_gender_atittudes.do" creates Figure 5 in the paper)

In R (the code and shapefiles are located in the RStudio folder)

a) Open the "Figure_2B_family_map.R" . Change the patch to the user's location. There is a note at the beginning of the script that reads "`# New users change this path:` "
b) Run the script to obtain the map.

4. Files Description

- **regidors_`year'.dta:** there is one database for each election-year (1979, 1983, 1987, 1991, 1995, 1999, 2003, 2007, 2011, and 2015). Each dataset contains background information about all city councilors in Spanish municipalities for each election year. These databases were collated by the Spanish Government (*Ministerio de Hacienda y Función Pública*) and shared with the researchers. These databases are the source for our main dependent variable, namely the share of female councilors at the municipal level for the different elections.

- **epa_`year'_q`i'_replication.dta**: these files contain the raw microdata from the Active Population Survey (*Encuesta de Población Activa,* or EPA) conducted by the National Institute for Statistics (*Instituto Nacional de Estadística*, or INE). The raw data from 1999 onward is available at https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resultados&secc=1254736030639&idp=1254735976595#_tabs-1254736030639. Although earlier data are not available on the web, the INE shares it upon request. This replication package includes anonymized data from 1979 to 2015. Starting from these raw data, we have generated a year and quarter variables, kept relevant variables for our

analysis, and, in a few cases, harmonized some variable names for posterior analysis. This microdata is processed in the do file "Figure_4_labor_market.do", then saved as **epa_prov_year_gender.dta**. Our master file calls this dataset and merges it with the main one, generating Figure 4 on the main paper.

- **cis_`year'_replication.dta** and **metroscopia_15077_replication.dta**: These datasets are used for the replication of Figures 5(a)-(c), on gender attitudes, in the paper. CIS stands for *Centro de Investigaciones Sociólogicas* and it is an official polling agency of the Spanish government. Their data can be downloaded free of charge from the CIS website: https://www.cis.es/en/. *Metroscopia* is a private firm poll and their data is proprietary. We obtained permission to use it for research purposes. All datasets are anonymized at origin. In preparing these replication files we have translated question prompts and response categories from Spanish into English, included only the variables that are used in the analysis, and excluded Canary Islands (Atlantic Ocean) and Ceuta and Melilla (North of Africa), territories that are not part of the analysis. The do file "Figure_5_gender_attitudes.do" prepares these data and generates Figures 5(a)-(c) in the main paper.

- **women_councilor_names.dta**: this file contains the full name of women councilors between 1924-1930 and 1931-1939. We typed in these data from two sources: Gomez-Ferrer Morant and Del Moral Vargas (2015), and Nielfa Cristobal and Ruiz Franco (2015), respectively. The do file "Figure_A11_councilors_historic.do" starts with this dataset, generates relevant variables and saves them for two periods (**women_councilors_all.dta** contains the information for the autocratic and democratic periods, and **total_women_democ.dta** for 1931-1939, the democratic period); then, it merges in information for provinces without any women councilor (calling the data file **provincias.dta**) and population in 1930 at the province level (calling the file **pop_1930_ine.dta**, from INE historical data repository: https://www.ine.es/jaxi/Datos.htm?path=/t20/e245/p06/l0/&file=1930.px); finally, it generates Figure A11 in Appendix Section J.

- **census_1860.dta**: this dataset contains information from the historical 1860 Spanish census (data available at https://www.ine.es/inebaseweb/treeNavigation.do?tn=192209). Original information is at the historical district level, and we map the historical districts to modern municipalities and generate a dataset at the municipal level. This dataset contains our main explanatory variable, namely the historical family structure, and is merged into the councilor's dataset in the do file "01_main_ data_construction.do".

- **census_1887.dta**: this dataset contains information from the historical 1887 Spanish census (data available at: https://www.ine.es/inebaseweb/treeNavigation.do?tn=193862).

Original information is at the historical district level. This dataset is called in by the do file "Table_A9_female_literacy_1998.do" to replicate Appendix Table A9 on the analysis of women's literacy in the nineteenth century.

- **census_1860_2001.dta**: this dataset contains information on the family type for 1860 and 2001 and is aggregated at the province level. This dataset is called in the do file "Figure_A2_1860_2001_census.do" to generate Figure A2 in Appendix Section C to analyze the disappearance of stem families.

- **Municipal_controls.dta**: this dataset contains our time-invariant controls at the municipal level. We measure caloric suitability using the Caloric Suitability Index by Galor and Özak (2016), which captures the potential agricultural output (measured in calories) based on crops that were available for cultivation post-1500 CE. We compute the mean temperature and precipitation for 1970-2000, relying on data from the *WorldClim* website (available at https://worldclim.org/data/index.html). We include a ruggedness index, using data from Diego Puga's website (https://diegopuga.org/data/rugged/), and the latitude and longitude of the municipality's centroid using data from the Spain's National Institute for Geography (available at https://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=CAANE). Finally, we include information on coastal municipalities from Solé-Ollé and Viladecans-Marsal (2013), and municipal area from 2011 *Anuario Económico de La Caixa* (a yearly economic report that used to be published by a Spanish Bank).

- **Pop.dta:** this dataset contains population at the municipal level for each election year, from the Spanish Ministry of Interior Affairs (https://infoelectoral.interior.gob.es/es/inicio/).

- **main_analysis_dataset.dta**: this is the main dataset for our main analysis. The do file "01_main_data_construction.do" contains the code to construct it. This do file merges the several sources of raw data, generates relevant variables for the analysis, and saves the main file. The file "02_master_results.do" contains the code that replicates our main results in the paper. The figures and tables generated by this code will be stored in the *Figures* and *Tables* folders.

For additional details on variable definitions and sources, refer to Section E in the Appendix.

The following table summarizes, for all the tables and figures in the main text and appendix, their code or source to replicate them.

Table or Figure	Source
Figure 1	Theory. Authors' own elaboration
Figure 2(a)	"02_master_results.do"
Figure 2(b)	"Figure_2b_family_map.R"
Table 1	"02_master_results.do"
Table 2	"02_master_results.do"
Figure 3	"02_master_results.do"
Figure 4	"02_master_results.do"
Figures 5(a)-(c)	"02_master_results.do"
Figure 6	"02_master_results.do"
Table 3	"02_master_results.do"
Figure A-1(a)-(b)	Borrowed from Duranton et al. (2009)
Figure A-1(c)	Same as Figure 2(b)
Table A-1	Borrowed from Tur-Prats (2019)
Figure A-2	"02_master_results.do"
Figure A-3	"02_master_results.do"
Table A-2	"02_master_results.do"
Table A-3	"02_master_results.do"
Table A-4	"02_master_results.do"
Figure A-4	"02_master_results.do"
Figure A-5	"02_master_results.do"
Figure A-6	"02_master_results.do"
Figure A-7	"02_master_results.do"
Figure A-8	"02_master_results.do"
Table A-5	"02_master_results.do"
Table A-6	"02_master_results.do"
Table A-7	"02_master_results.do"
Figure A-9	Borrowed from Tur-Prats (2019)
Table A-8	"02_master_results.do"
Figure A-10	"02_master_results.do"
Table A-9	"02_master_results.do"
Figure A-11	"02_master_results.do"
Figure A-12	Borrowed from Verge (2012)
Table A-10	"02_master_results.do"
Figure A-13	"02_master_results.do"

References

Duranton, Gilles, Andres Rodriguez-Pose, and Richard Sandall (2009), “Family Types and the Persistence of Regional Disparities in Europe”, *Economic Geography* 85(1): 23-47.

Galor, Oded, and Ömer Özak (2016), “The Agricultural Origins of Time Preference”, *American Economic Review* 106(10): 3064-3103.

Gómez-Ferrer Morant, Guadalupe, and Marta Del Moral Vargas (2015). Las pioneras en la gestión local: Concejalas y alcadesas designadas durante la dictadura de Primo de Rivera y el Gobierno Berenguer (1924-1930). In G. Nielfa Cristóbal (Ed.), *Mujeres en los Gobiernos Locales: Alcadesas y Concejalas en la España Contemporánea*, pp. 43-48. Madrid: Biblioteca Nueva.

Solé-Ollé, Albert, and Elisabet Viladecans-Marsal (2013), "Do political parties matter for local land use policies?” *Journal of Urban Economics*, 78: 42-56.

Nielfa Cristóbal, Gloria, and Rosario Ruiz Franco (2015). La nueva ciudadanía de las mujeres en el ámbito municipal: Alcaldesas y concejalas en la Segunda Republica (1931-1939). In G. Nielfa Cristóbal (Ed.), *Mujeres en los Gobiernos Locales: Alcadesas y Concejalas en la España Contemporánea*, pp. 89-171. Madrid: Biblioteca Nueva.

Tur-Prats, Ana (2019), “Family Types and Intimate Partner Violence: A Historical Perspective”, *Review of Economics and Statistics* 101(5): 878-891.

Verge, Tania (2012), “Institutionalising gender equality in Spain: From party quotas to electoral gender quotas”, *West European Politics* 35(2): 395-414.