Codebook for “Women and Party Building: Evidence from Municipal Governments in Brazil”

Tanushree Goyal and Cameron Sells

July 27, 2023

Overview

This codebook contains a description of the analysis datasets and raw datasets used in the replication files.

The names of the .rds datasets follow the format `data_TYPE_DESCRIPTION.rds`, where `TYPE` can be either "`raw`," "`cleaned`," or "`analysis`."

`raw` datasets are datasets that will be used to build the other datasets. All of the cleaned and analysis datasets can be reproduced using only the raw datasets and the replication code.

`cleaned` datasets are intermediate datasets that are built using the raw datasets, and are used to construct the analysis datasets

`analysis` datasets are the datasets that are used in the code_analysis.R and code_appendix.R scripts to reproduce the article's figures and tables. The main analysis datasets are **data_analysis_RDD_men.rds** and **data_analysis_RDD_women.rds**.

Party members are voters who have formally joined a political party. The party membership data used in this article comes from the database of party affiliations compiled by Brazil's Supreme Electoral Court (TSE). Brazilians normally join a political party by affiliating themselves with the local party office in the municipality where they reside. The municipal party offices are then required to report new affiliations periodically to their national-level party authorities and to the TSE. The TSE's records contain basic information on each member, including their name, their municipality, the date of their affiliation, the date on which they disaffiliated, and the date on which their affiliation was canceled for any other reason.

The TSE membership dataset does not list the gender of each party member. Instead, we code the gender of party members based on the members' first names. We perform most of the gender codings using the `genderBR` package in R created by Fernando Meireles, which uses data from Brazil's 2010 census to calculate the probability that a person with a given first name identifies as a man or a woman. When the `genderBR` package failed to classify a given member as male or female, we coded the member's gender by hand based on Portuguese naming conventions for given names. We were unable the code the gender of approximately 7% of the party members.

Although the membership dataset contains affiliations from as early as the late 1970s, the TSE did not standardize the reporting requirements of affiliations until much later, and the affiliations data for this earlier period is of poor quality. For this reason, our analysis focuses on affiliations that occurred between 2005 and 2020.

In order to ensure that we are comparing the performance of male and female mayors in the same set of municipalities, we restrict our regression discontinuity design (RDD) sample to municipality-elections in which the top two candidates in the mayoral election were one man and one woman.

This dataset despite having advantages of coverage, mandate, and scale, also has some limitations. First, because affiliations are reported by local party organizations, it is possible that some of the voters on the membership registry were registered with the party without their consent or knowledge. Second, although local party organizations have strong incentives to report affiliations, they have much weaker incentives to report disaffiliations. As a result, it is possible that many of the members on the registry were active within their party only for a brief period of time. Similarly, the nature of our data does not allow us to distinguish activists who regularly participate in their party's conventions and campaigns from less-active members who never participate in party activities. Importantly, however, there is no clear reason to expect that this measurement error in the dependent variable would be correlated with treatment status or the gender of a mayoral candidate. Consequently, our research design can still offer meaningful insight into how local election outcomes affect membership recruitment even if some party members are misclassified or misreported.

Main Analysis Datasets

data_analysis_RDD_men.rds and **data_analysis_RDD_women.rds**

These datasets are the main datasets used for all Regression Discontinuity Design (RDD) analyses in the main text and appendix. Their observations are male candidates and female candidates, respectively, in the subset of Brazilian mayoral elections where the top two vote-winners were one male candidate and one female candidate.

These datasets are built in `code_wrangling_RDD.R` using the aggregated party membership datasets, along with other data on candidate, municipality, and party characteristics.

Both datasets contain the same set of variables:

Variable	Description
`Year`	The year of the mayoral election. Most other variables are measured during the term following this year.
`MunCode`	The TSE municipality code
`Party`	The **2020-era** initials of the party, matched to the initials in the party membership dataset
`SIGLA_PARTIDO`	The initials of the party at the time of the election. This may be different from the `Party` variable due to renamings, splits, and mergers.
`NOME_CANDIDATO`	The name of the candidate that the observation represents
`Voteshare`	That candidate's voteshare in the mayoral election
`Winner`	Whether that candidate won the mayoral election
`Woman`	Whether the candidate is a woman
`Margin`	That candidate's margin of victory, which is positive if they won the election and negative if they lost
`Voters`	The number of voters in the municipality at the time of the election
`Female.Voters`	The number of female voters in the municipality at the time of the election
`Male.Voters`	The number of male voters in the municipality at the time of the election

`Recruit.Share`	The number of new members that the candidates' party recruited in that municipality, per 1,000 voters living in that municipality
`Gender.Gap`	The gender gap in membership recruitment for the candidate's party in that municipality, during the term following the election
`Attrition.Rate`	The share members who were recruited into the candidate's party during that term who subsequently disaffiliated from the party **or** switched to a different party
`Attrition.Rate.Female`	The attrition rate for female recruits
`Attrition.Rate.Male`	The attrition rate for male recruits
`Switch.Rate.Female`	The share of female recruits who subsequently switched to a different party
`Switch.Rate.Male`	The share of male recruits who subsequently switched to a different party
`Renominated`	Whether the candidate was renominated as mayor in the next election
`Advanced`	Whether the candidate was nominated for state of federal office in any subsequent election
`Cumulative.Female.Recruits.Share`	Cumulative recruitment of female party members in that municipality over past terms, across all parties, per 1,000 female voters at the time of the election
`Recruit.Share.Lagged`	The share of members recruited into that party per 1,000 voters, lagged by one 4-year mayoral term
`Gender.Gap.Lagged`	The gender gap in membership recruitment, lagged by one 4-year mayoral term
`Attrition.Rate.Lagged`	The membership attrition rate, lagged by one 4-year mayoral term
`University`	Whether the candidate had a university degree
`Incumbent`	Whether the candidate was the incumbent mayor
`Married`	Whether the candidate was married
`Age`	The age of the candidate
`Dynastic`	Whether the candidate shares the same surname as a previous mayor in their municipality, or a federal or state politician representing their state
`Insider`	Whether the candidate was affiliated with their party for at least one full mayoral term prior to the election

Time.In.Party	The number of years that the candidate has been affiliated with the party that they ran under in the mayoral election
Time.In.Politics	The number of years that the candidate has been affiliated with **any** Brazilian party
Prior.Office	The most recent local office that the candidate held prior to this election
First.Female.Mayor	Whether this election would be the first time that a woman was elected mayor in that municipality
Prior.Female.Mayors	The number of women who were previously elected mayor in that municipality prior to the current election
Challenger_Incumbent	Whether the candidate's opponent was an incumbent
Challenger_Time.In.Politics	The number of years since the candidate's opponent first affiliated with any party
Illiteracy_2000	The illiteracy rate in that municipality in 2000
UrbanizationRate_2000	The urbanization rate in that municipality in 2000
HDI_2000	The human development index in that municipality in 2000
HDI_2010	The human development index in that municipality in 2010
Population_2000	The population of the municipality in 2000
Female.Participation.Rate	The labor force participation rate for women
Voter.Gender.Ratio	The ratio of female to male voters
PT.Deputies.Voteshare	The chamber of deputies voteshare across all candidates affiliated with the Workers' Party (PT) in that municipality, in the most recent federal election (held 2 years prior to the mayoral election in question)
Gender.Gap.Year1	The gender gap in membership recruitment during the 1st year of the 4-year mayoral term
Gender.Gap.Year2	The gender gap in membership recruitment during the 2nd year of the 4-year mayoral term
Gender.Gap.Year3	The gender gap in membership recruitment during the 3rd year of the 4-year mayoral term. This is the year in the election cycle that typically sees the highest rates of affiliations.
Gender.Gap.Year4	The gender gap in membership recruitment during the 4th year of the 4-year mayoral term

`Woman.CouncilCandidatesShare.AllParties`	The proportion of city council candidate's in that municipality's **next** election who were female
`Woman.Councilors.Share.AllParties`	The proportion of elected city councilors (i.e., winners of the city council election) during the **next** term who were female
`Female.Donations.Ratio`	The ratio of average campaign donations to female candidates over average donations to male candidates, in the **next** city council election

Additional Analysis Datasets

data_analysis_CareerPathFederal.rds

A candidate-level dataset used in the career path analyses in the appendix, containing data on the first time that the candidate ran for the Federal Chamber of Deputies and their career trajectory up until that point.

This dataset is constructed in `code_wrangling_Supplementary.R` using the cleaned candidates dataset.

Variable	Description
`CPF_CANDIDATO`	The candidate's taxpayer identification number
`ANO_ELEICAO`	The year of the candidate's first run for the Federal Chamber of Deputies
`SIGLA_PARTIDO`	The candidate's party at the time of that election
`DESC_SIT_TOT_TURNO`	A description of the outcome of the election for that candidate
`DESCRICAO_CARGO`	The office that the candidate was running for
`Winner`	Whether the candidate was elected in that election
`Woman`	Whether the candidate was a woman
`Held.State`	Whether the candidate held state-level office in a prior term
`State.Win.Year`	The year that the candidate was elected to state-level office for a prior term, if applicable
`Held.Mayor`	Whether the candidate held mayoral office in a prior term
`Mayor.Win.Year`	The year that the candidate was elected to mayoral office for a prior term, if applicable
`Held.Council`	Whether the candidate held city council office in a prior term
`Council.Win.Year`	The year that the candidate was elected to city council for a prior term, if applicable
`Held.Vice`	Whether the candidate served as vice-mayor in a prior term
`Vice.Win.Year`	The year that the candidate was elected vice-mayor for a prior term, if applicable
`Held.Local`	Whether the candidate held any municipal-level office prior to the
`Held.Lower`	Whether the candidate held any municipal-level **or** state-level office prior to their first run for the Federal Chamber of Deputies

`All.Donations`	The aggregate campaign donations that the candidate received for that federal election, if available
`Party.Funds`	The aggregate campaign donations through party funds that the candidate received
`Candidate.Transfers`	The aggregate campaign donations through transfers from other candidates that the candidate received

data_analysis_CareerPathRenominationShifts.rds

A Federal politician-level dataset used in the career path analyses in the appendix, containing data on whether they ran for mayor or were renominated for the Chamber of Deputies in the next election.

This dataset is constructed in `code_wrangling_Supplementary.R` using the cleaned candidates dataset.

Variable	Description
`CPF_CANDIDATO`	The candidate's taxpayer identification number
`Year`	The year of the candidate's was elected to the Chamber of Deputies
`SIGLA_PARTIDO`	The deputy's party at the time of their election to the Chamber of Deputies
`DESCRICAO_SEXO`	The sex of the candidate
`Mayoral.Run`	Whether the deputy ran for mayoral office in the next election (which is held half-way into the congressional term that they are currently serving)
`Renominated`	Whether the deputy was renominated for the Chamber of Deputies by any party in the next Federal election

data_analysis_MemberBalance.rds

A party-level table of affiliation rates by gender used to make Figure 1.
This dataset is constructed in `code_wrangling_Supplementary.R` using the cleaned membership files for a select group of parties.

data_analysis_MunicipalitiesPanel.rds

A party-muncipality-term-level dataset used for the two-way fixed effects analyses in the appendix.
This dataset is constructed in `code_wrangling_Supplementary.R` using the cleaned party membership files and electoral data

Variable	Description
`MunCode`	The TSE municipality code
`Year`	The year of the election
`Treated.Women`	Equals TRUE if the party held the mayorship in that municipality during that term **and** the mayor was female, and FALSE otherwise
`Treated.Men`	Equals TRUE if the party held the mayorship in that municipality during that term **and** the mayor was male, and FALSE otherwise
`Opposition`	Equals 1 if the party did not hold the mayorship in that municipality during that term, and 0 otherwise
`Gender.Gap`	The gender gap in party membership recruitment during the following term
`Recruit.Share`	The number of new members recruited into that party during that term, per 1,000 voters in that municipality
`Female.Recruit.Share`	The number of new **female** members recruited into that party during that term, per 1,000 **female** voters in that municipality
`Male.Recruit.Share`	The number of new **male** members recruited into that party during that term, per 1,000 **male** voters in that municipality

data_analysis_PartySwitchingByGender.rds

A table that shows rates of party-switching for difference offices, by gender.
This dataset is constructed in `code_wrangling_Supplementary.R` using the cleaned party membership files and the cleaned candidates dataset.

Raw Datasets

data_raw_CampaignDonationsCouncil.rds

A dataset on campaign donations to city council candidates, used in `code_wrangling_RDD.R`

Variable	Description
`CPF_CANDIDATO`	The candidate's taxpayer identification number
`MunCode`	The TSE municipality ID
`ANO_ELEICAO`	The year of the election
`Office`	The office that the candidate was running for (Vereador, or city council in all cases)
`AllDonations`	The aggregate donations that the candidate received for that campaign
`Sex`	The sex of the candidate

data_raw_CampaignDonationsFederalState.rds

A dataset on campaign donations to federal and state congressional candidates, used in `code_wrangling_Supplementary.R`

Variable	Description
`CPF_CANDIDATO`	The candidate's taxpayer identification number
`ANO_ELEICAO`	The year of the election
`DESCRICAO_CARGO`	The office that the candidate was running for (Vereador, or city council in all cases)
`TIPO_RECEITA`	Donation type
`VALOR_RECEITA`	Value of the donation in *réis*

data_raw_Candidates.rds

A dataset on candidates, used in `code_wrangling_candidates.R` (a cleaned version of this dataset is used throughout the other scripts as well)

Variable	Description
ANO_ELEICAO	The year of the election
NOME_CANDIDATO	The name of the candidate
NUM_TITULO_ELEITORAL_CANDIDATO	The candidate's electoral ID
CPF_CANDIDATO	The candidate's taxpayer ID
SEQUENCIAL_CANDIDATO	The candidate's candidate number for that election
DESCRICAO_SEXO	The candidate's sex
DATA_NASCIMENTO	Date of birth
DESCRICAO_GRAU_INSTRUCAO	Educational attainment
DESCRICAO_ESTADO_CIVIL	Marital status
DESC_SIT_TOT_TURNO	Election outcome
SIGLA_PARTIDO	Party
DESCRICAO_CARGO	The office for which the candidate was running
MunCode	Municipality code (for municipal-level candidates only)

data_raw_electorate.rds

A dataset on the number of male and female voters in a municipality at the time of a given election, used in `code_wrangling_MemberGender.R`

Variable	Description
PERIODO	The year of the election
MUNICIPIO	The name of the municipality
MunCode	The TSE municipality ID
Voters	The number of voters of either gender

Female.Voters	The number of female voters
Male.Voters	The number of male voters

data_raw_FederalDeputiesElections.rds

A party-term-municipality-level dataset on the votes received by different parties in Federal Chamber of Deputies elections, used in `code_wrangling_RDD.R`

Variable	Description
CODIGO_MUNICIPIO	TSE municipality ID
ANO_ELEICAO	Year of the election
SIGLA_PARTIDO	Party initials
Votes	Total Chamber of Deputies votes won by that party in that municipality in that election, across all of its candidates
QTDE_VOTOS_LEGENDA	The number of votes cast on the party's *legenda* (as opposed to preference votes for a particular candidate)
All.Votes	The aggregate votes cast for Chamber of Deputies across all parties in that municipality in that election
Deputies.Voteshare	The party's voteshare in that municipality in that election
Legenda.Share	The share of the party's votes that were cast on the *legenda*

data_raw_filiados_PARTY.rds

A family of individual-level party membership datasets, where "PARTY" is the initials of the party in question. There are 35 of these datasets in total, and they are used in `code_wrangling_MemberGender.R` to construct the aggregate party membership datasets.

Note: These datasets originally contained a variable called `NOME_DO_FILIADO` which contained the member's full name. In the interest of protecting the members' privacy, we have replaced this variable with a variable called `NOME_DO_FILIADO_REDUCED`, which contains only the member's **first** name. We retain the first name because this variable is used to code the gender of party members..

Variable	Description
DATA_DA_EXTRACAO	The date that we downloaded this data
HORA_DA_EXTRACAO	The time that we downloaded this data
NUMERO_DA_INSCRICAO	Member's voter ID number
SIGLA_DO_PARTIDO	Party initials
NOME_DO_PARTIDO	Party name
UF	State initials
CODIGO_DO_MUNICIPIO	Municipality ID
NOME_DO_MUNICIPIO	Municipality name
ZONA_ELEITORAL	Electoral zone
SECAO_ELEITORAL	Electoral section
DATA_DA_FILIACAO	Date of affiliation
SITUACAO_DO_REGISTRO	Status of affiliation
TIPO_DO_REGISTRO	Type of affiliation
DATA_DO_PROCESSAMENTO	Date of processing
DATA_DA_DESFILIACAO	Date of disaffiliation
DATA_DO_CANCELAMENTO	Date of cancelization
DATA_DA_REGULARIZACAO	Date of regularization
MOTIVO_DO_CANCELAMENTO	Reason for cancelization

NOME_DO_FILIADO_REDUCED	The member's first name

data_raw_MayoralCandidatesData.rds

A dataset on personal characteristics of mayoral candidates in our RDD samples, used in `code_wrangling_RDD.R`

Variable	Description
ANO_ELEICAO	Election year
CODIGO_MUNICIPIO	Municipality ID
NOME_CANDIDATO	Candidate name
CPF_CANDIDATO	Candidate's taxpayer ID number
Dynastic	Whether the candidate shares the same surname as a previous mayor in their municipality, or a federal or state politician representing their state
Prior.Office	The municipal-level office that the candidate held prior to this election, if applicable
Incumbent	Whether the candidate is the incumbent mayor
Insider	Whether the candidate was affiliated with their party for at least one full mayoral term prior to the election
Time.In.Party	The number of years that the candidate has been affiliated with the party that they ran under in the mayoral election
Time.In.Politics	The number of years that the candidate has been affiliated with **any** Brazilian party
Prior.Office	The most recent local office that the candidate held prior to this election
First.Female.Mayor	Whether this election would be the first time that a woman was elected mayor in that municipality
Prior.Female.Mayors	The number of women who were previously elected mayor in that municipality prior to the current election
Renominated	Whether the candidate was renominated for mayor in the next election
Advanced	Whether the candidate was nominated for any state or federal-level office in any subsequent election
University	Whether the candidate has a university-level eduction
Married	Whether the candidate is married
Age	Candidate's age

data_raw_MunicipalElections.rds

A precinct-level dataset on electoral returns in municipal elections, used in `code_wrangling_RDD.R`

Variable	Description
DATA_GERACAO	The date that we downloaded this data
HORA_GERACAO	The time that we downloaded this data
ANO_ELEICAO	Election year
NUM_TURNO	Round of the election
DESCRICAO_ELEICAO	Description of the election
SIGLA_UF	State code
SIGLA_UE	District code (which in this case is the same as the municipality code)
CODIGO_MUNICIPIO	Municipality code
NOME_MUNICIPIO	Municipality name
NUMERO_ZONA	Zone number
CODIGO_CARGO	Office code
NUMERO_CAND	Candidate number
SQ_CANDIDATO	Candidate number
NOME_CANDIDATO	Candidate's full name
NOME_URNA_CANDIDATO	Candidate name used for the election
DESCRICAO_CARGO	Office that the candidate was running for
CODIGO_SIT_CANDIDATO	Code of the candidate's registration status
DESC_SIT_CANDIDATO	Description of the candidate's registration status
CODIGO_SIT_CAND_TOT	Code of the candidate's election status
CODIGO_SIT_CAND_TOT	Description of the candidate's election status
NUMERO_PARTIDO	Party ID number
SIGLA_PARTIDO	Party initials
NOME_PARTIDO	Party name

SEQUENCIAL_LEGENDA	Coalition's sequential number
NOME_COLIGACAO	Coalition name
COMPOSICAO_LEGENDA	Coalition composition
TOTAL_VOTOS	Total votes

data_raw_Municipality.rds

A dataset on municipality characteristics, used in `code_wrangling_RDD.R`

Variable	Description
MunCode	TSE's municipality ID
ibge7_code	IBGE's municipality ID
tse_code	TSE's municipality ID
Population_2000	Population in 2000
Population_2010	Population in 2010
HDI_2000	Human development index in 2000
HDI_2010	Human development index in 2010
UrbanizationRate_2000	Urbanization rate in 2000
Illiteracy_2000	Illiteracy rate in 2000
PerCapitaIncome_2000	Per capita income in 2000
Católica.Apostólica.Romana	Percent of the municipal population that identified as Roman Catholic
Evangélicas	Percent of the municipal population that identified as Evangelical
Sem.religião	Percent of the municipal population that identified as non-religious
Female.Participation.Rate	Labor force participation rate for women
Average.Campaign.Spending	Average campaign spending in that municipality

data_raw_PartyData.rds

A dataset on party characteristics, used in `code_appendix.R`

Variable	Description
`Party`	Party initials
`NOME_PARTIDO`	Party name
`NUMERO_PARTIDO`	Party number
`PartyIdeologyLeft`	Whether the party is left-wing, as classified by Bolognesi, Ribeiro, and Codato (2023)
`PartyPolicySeeking`	Whether the party is policy-seeking, as classified by Bolognesi, Ribeiro, and Codato (2023)
`PartyWomanFriendly`	Whether the party has the will to support female candidacies, as classified by Wylie (2018)
`PartyHighWomanLeadership`	Whether the party had a high share of women in the national party leadership, as classified by Wylie (2018)
`DALP_CandidateSelection`	Centralization score of the party's candidate-selection process, from the Democratic Accountability and Linkages Project (DALP)
`DALP_Strategy`	Centralization score of the party's strategy-selection process, from the Democratic Accountability and Linkages Project (DALP)
`DALP_Centralization`	The sum of `DALP_CandidateSelection` and `DALP_Strategy`