Replication Codebook for “Money Talks? Financial Incentives and Party Support for Candidates from Underrepresented Groups”

Thiago N. Fonseca
Getulio Vargas Foundation, Cepesp, Brazil.

Débora Thomé
Getulio Vargas Foundation, Cepesp, Brazil.

João V. Guedes-Neto
Getulio Vargas Foundation, EBAPE, Brazil.

Datasets (what they are and how they’re used)

1) `test.data.gen` (party–level, gender focus)
Panel of party–state–office–year observations used to estimate treatment effects **for women**. Treated units are **federal deputy** races (votes for women count double for future funding); controls are **state deputy** races in the same state–year. Outcomes and “effort” measures are computed at the **party list** level within each state–year–office.

2) `test.data.race` (party–level, race focus)
Same structure and variables as `test.data.gen`, but outcomes/effort measure **Afro-Brazilian** candidates (Black/brown, i.e., “negros”) rather than women.

3) `gen.race.cand` (candidate–level, gender & race, competitiveness bins)
Candidate–party–state–office–year microdata used to study whether parties target **non-competitive** vs. competitive candidates. Includes each candidate’s funding and votes, plus `last.vote.share` to locate candidates relative to the **last elected** on their party list (the basis for the 0–10% “non-competitive” bin in the figures).

Shared conventions

- **Unit of observation:**
  - `test.data.gen` / `test.data.race`: **party × state × office × election.year**
  - `gen.race.cand`: **candidate × party × state × office × election.year**
- **Office:** open-list PR for **federal deputy** (treated) and **state deputy** (control).

- **States:** Brazil's 27 units (incl. DF), commonly two-letter codes (e.g., `SP`, `RJ`).
- **Parties:** official acronyms per TSE; values reflect the party at election time.
- **".share" variables:** stored/preferred as **proportions (0–1)**. In the paper we often report **percentages**; convert as needed (×100).

Variable dictionary

A) `test.data.gen` — party-level (women as the focal group)

Variable	Type	Definition / Coding	Notes & Formula
`election.year`	int	Election year (e.g., 2006, 2010, …, 2022).	Federal & state lower houses coincide.
`state`	str	State identifier (e.g., `SP`, `RJ`, `BA`) or full name.	Coterminous districts for federal/state deputies.
`office`	str	Legislative office: `federal_deputy` (treated) or `state_deputy` (control).	Maps the DiD treatment contrast.
`party`	str	Party acronym per TSE.	Use election-year labels (post-merger names where applicable).
`cand.share`	float (0–1)	**Share of women** on the party's list in that state–year–office.	# women candidates ÷ total party candidates. Party "effort" outcome.
`fund.share`	float (0–1)	**Share of party funds** allocated to **women** in that state–year–office.	(Σ funds to women) ÷ (Σ funds to all party candidates). Effort outcome.
`vote.share`	float (0–1)	**Party vote share earned by women** in that state–year–office.	(Σ votes for women on the party list) ÷ (Σ party list votes). Main outcome aligned to the vote-based reward.
`elect.share`	float (0–1)	**Share of seats won by women** within the party in that state–year–office.	# women elected from party ÷ # total party seats.
`incumb.share`	float (0–1)	**Share of incumbent women** on the party list.	Used as time-varying covariate in DR specs.
`coll.degree.share`	float (0–1)	Share of party candidates **with a college degree** (all candidates).	Time-varying covariate; proxy for candidate profiles.

B) `test.data.race` — party-level (Afro-Brazilians as the focal group)

Structure and variable names are identical to `test.data.gen`.
Interpret all ".share" variables **with respect to Afro-Brazilian candidates** (i.e., Black and brown, "negros" as per TSE self-ID). For example, here `cand.share` is the **share of Afro-Brazilian candidates**, `fund.share` the **share of funds to Afro-Brazilian candidates**, `vote.share` the **share of party votes cast for Afro-Brazilian candidates**, and `elect.share` the **share of seats won by Afro-Brazilian candidates**.

C) `gen.race.cand` — candidate-level (gender & race; competitiveness)

Variable	Type	Definition / Coding	Notes & Formula
`cand.id`	str	Unique candidate identifier (stable within election).	From TSE/CEPESP harmonization.
`gender`	str / int	Candidate's gender.	Suggested: `F`/`M` or 1 = female, 0 = male. Match paper coding.
`race`	str / int	Candidate's race category.	Suggested: 1 = **Afro-Brazilian (negro: preto+pardo)**; 0 = non-Afro. If multi-category is kept (white/Black/brown/Indigenous/etc.), document mapping used in analysis.
`election.year`	int	Election year.	—
`state`	str	State code or name.	—
`office`	str	`federal_deputy` or `state_deputy`.	Mirrors treatment/control context.
`party`	str	Party acronym.	—
`fund.share`	float (0–1)	Candidate's **share of party funds** within party × state × year × office.	Candidate funds ÷ total funds to that party list. Used in effort analyses.
`fund`	float (≥0)	Candidate's **total campaign funds** received.	Currency: BRL (nominal unless you store deflated; note in analysis if in real R$).
`vote.share`	float (0–1)	Candidate's **vote share in the district**.	Candidate votes ÷ total valid district votes (or party list votes, if that is your convention —use consistently with `last.vote.share`).
`elect.share`	int/ bool	Elected indicator.	1 = elected; 0 = not elected. (Kept as "share" for column compatibility.)

Variable	Type	Definition / Coding	Notes & Formula
`last.vote.share`	float (0–∞)	**Competitiveness ratio:** candidate's vote share **relative to the last elected** on the **same party list**.	`last.vote.share = vote.share_candidate / vote.share_last_elected_same_party_list`. The paper bins this (e.g., 0–10%) to define "non-competitive" candidates.

Notes on consistency & reporting

- Keep all "share" variables **as proportions (0–1)** in the data. Convert to **percent** only at presentation stage.
- If `fund` is deflated, record the **base year and deflator** in your replication README.
- For `race`, document the collapsing from TSE self-ID to **Afro-Brazilian vs. non-Afro** (negro = preto+pardo) used in the analysis.