Codebook for dataset used in *Guilty by Association*

Milena Ang

- `unique_id`: Unique identifier for observation
- `id_estado`/estado: State
- `id_distrito`/Distrito: Electoral distict
- `elec_year`: Electoral year
- `year_2012`/`year_2015`/year_2018: Dummies for each electoral year
- `voto_pro_gober`: Vote for candidates. For each district d in state s year y, Vote for candidates $= \frac{\text{votes}_{d,s,y}}{\text{effective_votes}_{d,s,y}}$; where votes is the number of votes cast in favor of the legislative candidates of the party of the governor that was in office in 2015 in each district* d of state s, and effective votes are the number of effective votes for legislator in that district. Source: INE (2018).
- `juicio`: Indicted: 1 if governor of state s in 2015 or after was issued an arrest warrant or arrested between July of 2015 and June of 2018. Source: Own.
- `log_pobtot`: Population (logged): Population by electoral district (logged). Estimates of the population for year 2012 and 2015 are based on the 2010 district-level information, and estimates for the population in year 2018 were posted in 2015 after the midterm elections. INEGI (2015)
- `log_pib`: State GDP (logged): State GDP base 2013 (logged) in the year before the election. 2017 values might be subject to updates. (Source: INEGI 2018).
- `econ_growth`: Economic growth: For each state s in year y, the economic growth is obtained with the next formula: $\frac{(Q_{2,y}-Q_{1,y})+(Q_{1,y}-Q_{4,y-1})+(Q_{4,y-1}-Q_{3,y-1})}{3}$; where $Q_{2,y}$ is the GDP in that state in the second quarter of the year, and so on. (Source: INEGI 2014).
- `hom_per100`: Homicides per 100,000: For each state s in year y, I obtained estimations of homicides by population: $\frac{Hom_{s,y01}+Hom_{s,y02}+Hom_{s,y03}+Hom_{s,y04}+Hom_{s,y05}+Hom_{s,y05}}{population_{s,y}} * 100000$; where $Hom_{s,y01}$ are the count of homicides in the month of January of year y for state s, and so on. Source of homicide counts: INEGI

- `mentions_prop`: Scandal: For each state s: $\text{scandal}_{s,y} = \frac{\text{mentions terms}_{s,y}}{\text{mentions}_{s,y}}$, where mentions $\text{terms}_{s,y}$ is the number of times the terms corrupción, peculado, and enriquecimiento ilícito were mentioned along with the words gobernador and the state name in the newspaper el Universal between Jan 1st of the election year and the day before the elections. $\text{mentions}_{s,y}$ is the number of times the words gobernador and the state name were mentioned in the newspaper el Universal between Jan 1st of the election year and the day before the elections. Source: Own using eluniversal.com.mx
- `prev_exp_fleg`: Candidate's experience. 1 if candidate had previous experience as a federal legislator,0 if not. Source: Own (thanks to Caro Plata for her invaluable help in gathering this data).
- `partido_gober`: Party of the governor. Source: Own.
- `st_elections_15_18`: Dummy variable equal to 1 if state had any local elections between 2015 and 2018 Source: Own.