

# Codebook for datafiles

Recent Mexican election vote returns repository

<https://github.com/emagar/elecRetrns>

November 11, 2023

## 1 Variables grouped by type

### 1.1 Unit IDs

**edon** = state numeral 1:32.

**edo** = state abbreviation (may differ from commonly used abbreviations, so that sorting them alphabetically preserves the order set by edon, eg. Chiapas is cps, not chis).

**disn** = district identifier =  $\text{edon} * 100 + \text{district numeral}$ .

**cab** = cabecera, district's administrative center.

**inegi, ife** = municipal identifier codes used by the INEGI and the IFE/INE, respectively.

**mun** = municipality's name.

**emm** = unit's identifying code. It concatenates the state's /edo/ abbreviation (then a hyphen) two sequential digits for the election cycle (then a period) and, depending on the level of observation, the /inegi/ or district identifier. Using /emm/ as sort criterion returns a state-time-unit ordering.

**demar** = demarcación identifier =  $\text{inegi} + 1/100 \text{ demarcación numeral}$  (used for Nayarit municipal elections only).

**seccion** = voting precinct identifier =  $\text{/edon/} * 10000 + \text{sección electoral numeral}$  (as set by IFE/INE).

**casilla** = polling station identifier (as set by IFE/INE). There are four types of stations, coded B for /Básica/, C for /Contigua/, E for /Extraordinaria/, and S for /Especial/.

**circ** = secondary, proportional representation district (circunscripción plurinominal) the primary district belongs to.

**latitude, longitude** = coordinates indicating a polling booths's north-south and east-west position in a map. Available for federal deputy and presidential casilla-level returns in the 2006, 2009, 2015, and 2018 elections.

### 1.2 Temporal IDs

**yr, mo, dy** = year, month, day of the election.

**date.el, date.in** = date of the election and start of term, respectively.

**dextra** = dummy equal 1 for special elections (elección extraordinaria), 0 otherwise.

**danul** = dummy equal 1 for voided elections, 0 otherwise.

### 1.3 Voting

**v01, v02, ...** = raw vote for candidate 1, 2, etc.

**l01, l02, ...** = label of candidate 1's, 2's, ... party or coalition.

**c01, c02, ...** = candidate 1's, 2's, ... name.

**efec** = effective vote total, equal the sum of raw votes minus votes for write-in candidates minus invalid ballots. This is the denominator to compute vote shares.

**lisnom** = unit's total eligible voters (lista nominal).

**nr** = votes for write-in candidates (candidatos no registrados, void in Mexican election law).

**nul, nulos** = invalid ballots (votos nulos).

**tot** = total raw votes.

**win** = winner's party or coalition.

**ncand** = number of candidates running.

**dcoal** = dummy equal 1 if at least one candidate ran on a multi-party pre-electoral coalition, 0 otherwise.

**ncoal** = number of candidates who ran on multi-party pre-electoral coalitions.

**coalpan, coalpri, coalprd** = members of major-party coalitions ('no' indicates no coalition).

**dfake** = indicates an attempt to complete missing hegemonic era races (mostly in the 1960s and 70s) for the purpose of computing vote lags, made up of press reports and best guesses about what happened in the state's race.

### 1.4 Historical performance

**d.pan, d.pri, d.left** = first difference in the party's federal deputy vote share from last to present election.

**vhat.pan, vhat.pri, vhat.left** = predicted federal deputy vote share in the unit for the current election, a linear projection of the last five races.

**bhat.pan, bhat.left** = slope estimate of the party's autoregressive linear model for the unit. The PRI used as reference vote and has no slope estimate.

**alphahat.pan, alphahat.pri, alphahat.left** = party's alpha estimated for the unit.

**betahat.pan, betahat.left** = party's beta estimate for the unit. The PRI used as reference vote and has no beta estimate.

**dbackward** = dummy equal 1 if prediction with autoregressive model performed backwards, 0 otherwise.

### 1.5 Candidates and incumbents

**incumbent, runnerup** = winning and runner-up candidates' names.

**propietario, suplente** = primary and substitute candidate's name, respectively.

**part** = incumbent/candidate's party or coalition.

**part.2nd** = runner-up party or coalition.

**mg** = winner's margin = winner's vote share minus runner-up's vote share.

**dmujer** = dummy equal 1 if candidate/incumbent is a woman, 0 otherwise.

**race.after** = incumbent's status in the next consecutive race. The repository's README file describes categories and coding procedure.

**dreran** = dummy equal 1 if incumbent ran again in the next consecutive race for the same office.

**dreelected** = dummy equal 1 if incumbent won the next consecutive race for the same office.

**dcarta** = dummy equal 1 if member of Congress filed a letter of intent with the chamber's Junta to run for office again; 0 otherwise. Inapplicable before 2018.

**lista** = candidate's rank in senate two-member party lists. Top member of runner-up vote-getting list wins the state's third senate seat.

**drp** = dummy equal 1 if candidate ran for a PR seat, 0 otherwise.

**ddied** = dummy equal 1 if incumbent died in office, 0 otherwise.

## 2 Variables in alphabetical order

**alphahat.pan, alphahat.pri, alphahat.left** = party's alpha estimated for the unit.

**betahat.pan, betahat.left** = party's beta estimate for the unit. The PRI used as reference vote and has no beta estimate.

**bhat.pan, bhat.left** = slope estimate of the party's autoregressive linear model for the unit. The PRI used as reference vote and has no slope estimate.

**c01, c02, ...** = candidate 1's, 2's, ... name.

**cab** = cabecera, district's administrative center.

**casilla** = polling station identifier (as set by IFE/INE). There are four types of stations, coded B for /Básica/, C for /Contigua/, E for /Extraordinaria/, and S for /Especial/.

**d.pan, d.pri, d.left** = first difference in the party's federal deputy vote share from last to present election.

**danul** = dummy equal 1 for voided elections, 0 otherwise.

**date.el** = election date.

**date.in** = start of term date.

**dbackward** = dummy equal 1 if prediction with autoregressive model performed backwards, 0 otherwise.

**dcarta** = dummy equal 1 if member of Congress filed a letter of intent with the chamber's Junta to run for office again; 0 otherwise. Inapplicable before 2018.

**ddied** = dummy equal 1 if incumbent died in office, 0 otherwise.

**demar** = demarcación identifier = inegi + 1/100 demarcación numeral (used for Nayarit municipal elections only).

**dextra** = dummy equal 1 for special elections (elección extraordinaria), 0 otherwise.

**dfake** = indicates an attempt to complete missing hegemonic era races (mostly in the 1960s and 70s) for the purpose of computing vote lags, made up of press reports and best guesses about what happened in the state's race.

**disn** = district identifier = edon \* 100 + district numeral.

**dmujer** = dummy equal 1 if candidate/incumbent is a woman, 0 otherwise.

**dreelected** = dummy equal 1 if incumbent won the next consecutive race for the same office.

**dreran** = dummy equal 1 if incumbent ran again in the next consecutive race for the same office.

**drp** = dummy equal 1 if candidate ran for a PR seat, 0 otherwise.

**dy** = day of the election.

**edo** = state abbreviation (may differ from commonly used abbreviations, so that sorting them alphabetically preserves the order set by edon, eg. Chiapas is cps, not chis).

**edon** = state numeral 1:32.

**efec** = effective vote total, equal the sum of raw votes minus votes for write-in candidates minus invalid ballots. This is the denominator to compute vote shares.

**emm** = unit's identifying code. It concatenates the state's /edo/ abbreviation (then a hyphen) two sequential digits for the election cycle (then a period) and, depending on the level of observation, the /inegi/ or district identifier. Using /emm/ as sort criterion returns a state-time-unit ordering.

**ife** = municipal identifier codes used by the IFE/INE.

**incumbent, runnerup** = winning and runner-up candidates' names.

**inegi** = municipal identifier codes used by the INEGI.

**l01, l02, ...** = label of candidate 1's, 2's, ... party or coalition.

**latitude** = coordinates indicating a polling booths's north-south position in a map. Available for federal deputy and presidential casilla-level returns in the 2006, 2009, 2015, and 2018 elections.

**lisnom** = unit's total eligible voters (lista nominal).

**lista** = candidate's rank in senate two-member party lists. Top member of runner-up vote-getting list wins the state's third senate seat.

**longitude** = coordinates indicating a polling booths's east-west position in a map. Available for federal deputy and presidential casilla-level returns in the 2006, 2009, 2015, and 2018 elections.

**mg** = winner's margin = winner's vote share minus runner-up's vote share.

**mo** = month of the election.

**mun** = municipality's name.

**ncand** = number of candidates running.

**nccoal** = number of candidates who ran on multi-party pre-electoral coalitions.

**nr** = votes for write-in candidates (candidatos no registrados, void in Mexican election law).

**nul, nul0s** = invalid ballots (votos nul0s).

**part.2nd** = runner-up party or coalition.

**part** = incumbent/candidate's party or coalition.

**propietario, suplente** = primary and substitute candidate's name, respectively.

**race.after** = incumbent's status in the next consecutive race. The repository's README file describes categories and coding procedure.

**seccion** = voting precinct identifier = /edon/ \* 10000 + sección electoral numeral (as set by IFE/INE).

**tot** = total raw votes.

**v01, v02, ...** = raw vote for candidate 1, 2, etc.

**vhat.pan, vhat.pri, vhat.left** = predicted federal deputy vote share in the unit for the current election, a linear projection of the last five races.

**win** = winner's party or coalition.

**yr** = year of the election.