Candidate Characteristics Cooperative (C3) 2018 Data Codebook
(updated May 5, 2021)

Authors: Bernard L. Fraga, Eric Gonzalez Juenke, and Paru Shah

Overview

In 2018, we created the Candidate Characteristics Cooperative (C3), a hand-coded database of primary and general election candidates for state legislative elections held in 2018. We identified twenty methodologically diverse contributors across the country, and in return for coding a single state, contributors were offered access to the complete dataset during an embargo period of twelve months. We provided a list of the primary election candidates and relevant electoral data, and asked contributors to code the race, ethnicity, and gender of candidates with our guidance on categorizations. Contributors submitted their completed state files to the PIs, and we compiled these data into a single, uniform file.

The result of this project is a hand-coded database of nearly all state legislative primary and general election candidacies from 2018. Covering approximately 14,500 unique major-party and minor-party candidates contributors were able to identify the race/ethnicity of 94% of the candidates when using the techniques described above. The Candidate Characteristics Cooperative dataset also has similarly complete coding of candidate gender. The data are useful for those studying legislative representation and elections in the United States.

2018 Coding Contributors

Daniel Butler
Jason Casellas
Kennia Coronado
Kesicia Dickinson
Christian Dyogi-Phillips
Matthew Hayes
Robert Hogan
Danielle Lemi
Michelangelo Landgrave
Raymundo Lopez
Zoe Nemerever
Shayla F. Olsen
Spencer Piston
Robert Preuhs
Sara Sadhwani
Jamil Scott
Sono Shah
Christopher Stout
Danielle Thomsen
Erika Vallejo
Emily West

Citation

Fraga, Bernard L., Eric Gonzalez Juenke, Paru Shah. 2021. "Candidate Characteristics Cooperative Database, 2018 State Legislative Elections." Version 1; Published on 5/5/2021.

Candidate data purchased from Carl Klarner at: https://www.klarnerpolitics.org/

- Klarner, Carl, 2018, "State Legislative Election Returns, 1967-2016", https://doi.org/10.7910/DVN/3WZFK9, Harvard Dataverse, V3, UNF:6:pV4h1CP/B8pHthjjQThTTw==

Variable Description

id	C3 database unique id
year	2018
state	State postal abbreviation
fips	Federal numerical code for each state
chamber	Lower or Upper chamber
ddezo	District name (for named districts)
dname	Lower case collapsed dist name (for named districts)
dno	District number
geopost	Additional district number for MD, MN, SD, VT districts
mmdpost	multi-member district position number
fullname	Full name, including middle name and suffix
first	First name (lowercase)
last	Last name (lowercase)
middle	Middle name or initial (lowercase)
party	Party affiliation
incumbent	Incumbent =1
primaryoutcome	Winner or Loser in primary
generaloutcome	Won/Lost General Election
gender	(W)oman, (M)an, (U)nknown
race	Racial Code from Coding Scheme
ethnicity	Ethnic coding, though this is basically a + 1 on race
notes	Coder-provided notes on candidate coding

Race/Ethnicity Coding Guide – State Legislative Candidate Data Project – Updated May 30, 2018

The **Race** variable should be coded as follows for each candidate, one term per candidate:

- **White (Non-Hispanic White)**
  - European Ancestry (includes Spain), Armenian
- **Black**
  - Includes Haiti, Jamaica, Trinidad, African.
- **Latino (Hispanic)**
  - Includes any ancestry from Latin America not included in Black above. Does not include Spain unless accompanied by Latin American ancestry.
- **Asian American**
  - Includes Indian, Pakistani
- **MENA** (from "Middle East/North African" region)
  - Includes North African, Turkish, Egyptian, Iranian/Persian.
- **Native American**
  - Must indicate tribal membership. Includes Native Hawaiian and Alaska Native.
- **Multiracial**
  - Note racial identifiers in the Ethnicity field.

Ethnicity should be coded if you find evidence of country of origin (i.e. Mexican or Colombian for Latino; Indian or Chinese for Asian).

If you cannot find any information whatsoever regarding the racial/ethnic background of a candidate, leave it blank.

Notes variable is provided if there are additional identities (sexual orientation, religion, disability) that is provided by the candidate.

Detailed Codebook

---

id

---

type: numeric (int)

range: [1,14775] units: 1
unique values: 14,775 missing .: 0/14,775

---

year

---

type: numeric (int)

range: [2018,2018] units: 1
unique values: 1 missing .: 1/14,775

---

state State Abbreviation

---

type: string (str2)

unique values: 45 missing "": 0/14,775

examples: "IA"
"MI"
"NH"
"RI"

---

fips (unlabeled)

---

type: numeric (byte)

range: [1,56] units: 1
unique values: 45 missing .: 0/14,775

---
chamber (unlabeled)
---

type: string (str5)

unique values: 2 missing "": 0/14,775

tabulation: Freq. Value
11,845 "Lower"
2,930 "Upper"

---
ddezo (unlabeled)
---

type: string (str41)

unique values: 521 missing "": 13,259/14,775

warning: variable has embedded blanks

---
dname (unlabeled)
---

type: string (str34)

unique values: 85 missing "": 13,297/14,775

---
dno (unlabeled)
---

type: numeric (int)

range: [1,203] units: 1
unique values: 203 missing .: 143/14,775

---

geopost (unlabeled)

---

type: numeric (byte)

range: [1,7] units: 1
unique values: 7 missing .: 14,267/14,775

tabulation: Freq. Value
238 1
242 2
18 3
5 4
2 5
1 6
2 7
14,267 .

---

mmdpost (unlabeled)

---

type: numeric (byte)

range: [1,2] units: 1
unique values: 2 missing .: 14,339/14,775

tabulation: Freq. Value
187 1
249 2
14,339 .

---

fullname (unlabeled)

---

type: string (str35)

unique values: 14,176 missing "": 0/14,775

examples: "DAVID MILLARD"
"Jae Potter"
"Lisa Seger"
"Richard J. Wein"

---

first (unlabeled)

---

type: string (str14)

unique values: 2,503 missing "": 1/14,775

examples: "dan"
"jasmine"
"liz"
"rick"

---

last (unlabeled)

---

type: string (str19)

unique values: 8,503 missing "": 0/14,775

examples: "daigle"
"hill"
"mcvittyweber"
"schimmelpfenning"

---

middle (unlabeled)

---

type: string (str13)

unique values: 797 missing "": 9,433/14,775

examples: "j"

---

party (unlabeled)

---

type: string (str10)

unique values: 4 missing "": 0/14,775

```
tabulation:          Freq.  Value
                     7,148  "Democratic"
                     1,216  "MinorParty"
                     6,398  "Republican"
                     13     "WriteIn"

-------------------------------------------------------------------------------------------------------------
incumbent                                                                                  (unlabeled)
-------------------------------------------------------------------------------------------------------------

        type:  numeric (byte)

       range:  [0,1]                units:  1
  unique values:  2                 missing .:  0/14,775

    tabulation:  Freq.  Value
                 9,843  0
                 4,932  1
-------------------------------------------------------------------------------------------------------------
primaryoutcome                                                                             (unlabeled)
-------------------------------------------------------------------------------------------------------------

        type:  string (str13)

  unique values:  3                 missing "":  337/14,775

    tabulation:  Freq.  Value
                  337  ""
                 3,239  "l"
                 139  "notapplicable"
                 11,060  "w"
-------------------------------------------------------------------------------------------------------------
generaloutcome                                                                             (unlabeled)
-------------------------------------------------------------------------------------------------------------

        type:  string (str1)

  unique values:  2                 missing "":  3,225/14,775

    tabulation:  Freq.  Value
                 3,225  ""
                 5,533  "l"
                 6,017  "w"

-------------------------------------------------------------------------------------------------------------
gender                                                                                     (unlabeled)
-------------------------------------------------------------------------------------------------------------

        type:  string (str1)
```

unique values: 3 missing "": 0/14,775

tabulation: Freq. Value
10,104 "M"
113 "U"
4,558 "W"

---

race race

---

type: string (str15)

unique values: 8 missing "": 279/14,775

tabulation: Freq. Value
279 ""
289 "Asian American"
1,431 "Black"
613 "Latino"
67 "MENA"
22 "Multiracial"
54 "Native American"
36 "Unknown"
11,984 "White"

---

ethnicity (unlabeled)

---

type: string (str52)

unique values: 90 missing "": 14,447/14,775

examples: ""

---

notes (unlabeled)

---

type: string (str111)

unique values: 117 missing "": 14,029/14,775

examples: ""