Legislative and District Data
1972-2013

Author: Ella Foster-Molina
Affiliation: University of Rochester
Years: 1972-2013 (also 93rd-113th Congresses) House of Representatives and District Data
Published: October 2016

Citation and Use:
- Publications and research reports that use this data must cite it appropriately by including the following information:
  Ella Foster-Molina. *Legislative and District Data 1972-2013*. October 2016.
- This dataset incorporates data from ICPSR for the years between 1972 and 1982. Any use of the data for 1972-1992 is only available for ICPSR authorized users. Limited license information is included in the files sent along with this document to authorized users.
  - "When sharing data or other materials in these approved ways, you must include all accompanying files with the data, including terms of use. More information on permission to redistribute data can be found on the ICPSR Web site."[1]
  - "Authors of publications based on ICPSR data are required to send citations of their published works to ICPSR for inclusion in a database of related publications (bibliography@icpsr.umich.edu)."[2]
- This dataset incorporates data from the NHGIS for the years 1982-1992, which has a limited license for use of the data. If your analysis includes the years 1982-1992, the following provisions apply:
  - "Publications and research reports employing NHGIS data (either tabular or GIS) must cite it appropriately. The citation should include the following:
    Minnesota Population Center. National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota 2011."[3]
  - "In addition, we request that users send us a copy of any publications, research reports, or educational material making use of the data or documentation. Printed matter should be sent to:
    NHGIS
    Minnesota Population Center
    University of Minnesota
    50 Willey Hall
    225 19th Ave S
    Minneapolis, MN 55455"[4]

---

[1] See document labeled "Terms of Use" included with the other material sent.
[2] ibid
[3] https://www.nhgis.org/research/citation
[4] ibid

Data was compiled, decoded, and packaged from the following sources:

Census:
- census.gov (1992-2013)
- nhgis.org (1982-1992)
- ICPSR-00011 (1972-1982)
- ICPSR-08402 (1984-1992, only for CA, HA, LA, ME, MS, MT, NJ, NY, TX, WA)

Legislative and committee data, age/party/seniority:
- govtrack.us (1972-2013)

Cook scores:
- https://en.wikipedia.org/wiki/Cook_partisan_voting_index (2006-2013)

Black and Hispanic caucus membership:
- history.house.gov (1972-2013)

Nominate scores:
- http://voteview.com/dwnl.html (1972-2013)

Data Dictionary

MC and District Identification
- stateDist:
  - state and district information. States are abbreviated, so Kansas's first district is KS.1. For single district states, the district number is '0', ie WY.0. When two MCs served the same district in a congressional session because the original was replaced, both MCs are listed. Consider weighting regressions by the fraction of time served in Congress (fracServed). In general, 2-10 districts see an MC replaced in a given session.
- congNum:
  - Congressional session number.
- sponID:
  - govtrack id number for each member of Congress. All entries have a govtrack ID number.
- icpsr:
  - icpsr number for each member of Congress. Not all entries have an ICPSR number. Note that there are some discrepancies in how sources use ICPSR numbers, as some sources use 9's as the first digit to indicate party switches or other irregularities.
- lastName
  - Some nontraditional characters (hyphens, accents) are garbled in translation. All names are recognizable by the human eye, but id numbers should be used for matching individuals.
- firstName
- middleName

MC Personal Characteristics
- age
  - age at time of being sworn into Congress for that session.
- gender
  - gender of MC.
- party
  - party identification, Democrat (D) or Republican (R). If neither, classified as Independent (I).
- black
  - binary variable for being a member of the Congressional Black Caucus. 1 if a member of CBC, 0 if not. To my knowledge, all self identifying black members of Congress are members of the CBC.
- hispanic
  - binary variable for being a member of the Congressional Hispanic Caucus. 1 if a member of CHC, 0 if not. To my knowledge, all self identifying hispanic members of Congress are members of the CBC.

Data Dictionary

MC Time Served

- numberTerms
  - number of terms served in the House.
- daysServed
  - number of days of the full session served.
- fracServed
  - fraction of the congressional session that the MC served in. The vast majority of MCs served in the full session (2 years). Those that didn't filled vacant seats or left mid-term. This variable can be used to weight regressions to avoid double counting a district.

MC Legislative Information

- numSpon
  - number of bills an MC sponsored in a term.
- numCosp
  - number of bills an MC cosponsored in a term.
- numPassH
  - number of bills an MC sponsored and were approved by a full House vote in a term.
- numEnact
  - number of bills an MC sponsored and were signed into law by a full House vote in a term.
- passPrcnt
  - percent of bills that an MC sponsored and were enacted into law

Data Dictionary

MC Ideology and Committee Information

- cd
  - congressional district number. This is in many ways a repeat of stateDist, but also helps the eye keep track of which district is under discussion when manually scanning the table.
- statenm
  - state name. Again, a repeat of information from stateDist, but this is the state name used by nominate tables, all in caps, and often abbreviated (ie MINNESO).
- dwnom1
  - first dimension nominate score.
- dwnom2
  - second dimension nominate score.
- comPower
  - how powerful were the committees the MC served in. Example provided is of the 109th Congress. Committees vary by congress, so points assigned vary by Congressional session. Documentation available upon request. All committee information is currently available for the 109th-113th Congresses.

Committees in 109th Congress	Power
House Committee on Ways and Means	3
House Committee on Rules	2
House Committee on the Judiciary	2
House Committee on Appropriations	1
House Committee on Energy and Commerce	1
House Committee on Financial Services	1
House Committee on Resources	1
House Committee on Transportation and Infrastructure	1
House Committee on Agriculture	0
House Committee on Armed Services	0
House Committee on the Budget	0
House Committee on Education and the Workforce	0
House Committee on Government Reform	0
House Committee on Homeland Security	0
House Committee on House Administration	0
House Committee on International Relations	0
House Committee on Science	0
House Committee on Small Business	0
House Committee on Standards of Official Conduct	0
House Committee on Veterans' Affairs	0
House Permanent Select Committee on Intelligence	0

- chair
  - Binary variable. Takes the value of 1 if the MC was the chair of a committee in that congressional session, 0 if not. No MC chaired more than one committee. Committee information is currently available for the 109th-113th Congresses.

- rank
  - Binary variable. Takes the value of 1 if the MC was the ranking member of a committee in that congressional session, 0 if not. No MC was ranking member of more than one committee. Committee information is currently available for the 109th-113th Congresses.
- comName
  - names of the committees served on. Committee information is currently available for the 109th-113th Congresses.
- numCom
  - number of committees served on. Committee information is currently available for the 109th-113th Congresses.
- rankChair
  - rankChair = 0.5rank + chair. This creates a single variable for power in a committee, where being a ranking member is half as valuable as being the chair of a committee. Committee information is currently available for the 109th-113th Congresses.

Congressional Session Characteristics
- partyControl
  - 'D' if Democrats controlled Congress, 'R' if Republicans controlled Congress.
- demSeats
  - number of seats held by Democrats at the beginning of the Congressional session.
- repSeats
  - number of seats held by Republicans at the beginning of the Congressional session.
- otherSeats
  - number of seats held by third parties (independents, green party, etc.). Note that this dataset does not provide information about which third party MCs belonged to.

Data Dictionary

District Mobility and Nativity Distribution
- state
  - state name, two letter abbreviation (ie AK, CA). This is a repeat of information from stateDist, but also helps the eye keep track of which district is under discussion when manually scanning the table.
- district
  - congressional district number. This is in many ways a repeat of stateDist, but also helps the eye keep track of which district is under discussion when manually scanning the table.
- abroadPrcnt
  - percent of the district that lived abroad in the census used for this term. Censuses are updated yearly after the 109th Congress (2005-06). Prior to the 109th Congress, censuses were generally only updated once per decade. Use this variable cautiously if time matters.
- recentArrivalPrcnt
  - percent of the district that recently moved into the district from another county or state. This is a mild approximation, as a county can be in the same district as another, but the census does not track whether people recently moved into a district.
- totalPopBirthPlace
  - total population used for measuring mobility and/or nativity.
- prcntForeignBorn
  - percent of the district that was born in a foreign country.
- prcntExAliens (103-113 only)
  - percent of the foreign born population that became a citizen. This is intended to capture a measure of assimilation.

Data Dictionary

District Income Distribution

- totalHouseholds
    - Total number of households in the district at the time of the most recent census.
- under10k
    - percent of the district's households that made less than $10,000/year in 2009 dollars. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over10k
    - percent of the district's households that made over $10,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over15k
    - percent of the district's households that made over $15,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over25k
    - percent of the district's households that made over $25,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over35k
    - percent of the district's households that made over $35,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over50k
    - percent of the district's households that made over $50,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over75k
    - percent of the district's households that made over $75,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over100k
    - percent of the district's households that made over $100,000/year. The exact income changes before and after 2009 due to inflation and the census using

different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.

- over150k
  - percent of the district's households that made over $150,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- over200k
  - percent of the district's households that made over $200,000/year. The exact income changes before and after 2009 due to inflation and the census using different income buckets for reporting income prior to 1990. It is always as close as possible to the 2009 value listed. Contact author for exact values used.
- meanIncome
  - mean income of a district's households. No inflation adjustments used.
- medianIncome
  - median income of a district's households. No inflation adjustments used. This value was estimated for the 1980 census using the percent of the population in each income bracket. Thus, for the 1980s there is a heavy rounding error involved in the estimation.
- gini
  - inequality index estimated from the percent of the population in each income bracket.
- ses
  - socioeconomic status measure: estimated from the income and education of a district. This was estimated by finding the weight of the common factor between education and income.
- sesNorm
  - socioeconomic status measure, normalized: estimated from the income and education of a district. It is normalized to range between 0 and 100. This allows the magnitude of this coefficient to be directly compared to the coefficients for income and education, both of which are given in percentages.

District Employment Distribution
- totalEmploymentPop
  - total population used for the employment variables. Only includes those over the age of around 16, depending on the cutoff for that census.
- prcntUnemp
  - percent of the district's population that is unemployed but still in the labor force.
- prcntNotEmploy
  - percent of the district's population that is not in the labor force. This excludes those who are unemployed but in the labor force.

District Education Distribution
- prcntBA
  - percent of the district with a bachelor's degree or higher.
- prcntHS
  - percent of the district with a high school degree or higher.

Data Dictionary

District Racial Distribution
- totalPopRaceFile
  - total population for the race variables.
- prcntAsian
  - percent of the district that is Asian.
- prcntBlack
  - percent of the district that is black, including black hispanics.
- prcntBlackNotHisp
  - percent of the district that is black but not hispanic.
- prcntHisp
  - percent of the district that is Hispanic, both white and black.
- prcntMulti
  - percent of the district that is multi-racial.
- prcntWhite
  - percent of the district that is white but not Hispanic.
- prcntWhiteAll
  - percent of the district that is white, including white hispanics.
- prcntNotHisp
  - percent of the district that is not Hispanic.

District Age Distribution
- prcntOld
  - percent of the district that is over 62 or 60, depending on the age buckets used by the census for that decade.
- medianAge
  - median age in the district.