

Introducing Political Science Datasets



January 30, 2023

POLS 300: Research Methods in Politics and Public Policy

Dr. Isaac Hale

Spring Semester, 2023

OXY Occidental
College

Outline

1. Research question check-in
2. What is a data set?
3. Popular American Politics data sets
4. Popular Comparative Politics data sets
5. Popular international data sets: public opinion, parties, & elections
6. Advanced data sleuthing: ICPSR & Harvard Dataverse

Research Question Check-In

Research Question Check-in

- REMINDER: the first draft of your research question is due at midnight on Friday
- You are not wedded to it!
 - It's not a problem if you change in February
- An ideal research question:
 - Is short (1-3 sentences)
 - Has a dependent (outcome) variable
 - Has an independent variable that relates to the DV
 - Identifies the **unit of analysis**
 - AKA, what are you studying? Voters? Citizens? Countries? Legislators? Parties? Wars?
- Feel free to come to office hours! We can map it out

What *Is* a Data Set?

A Set of Data



A Data Set

- A collection of **data**, aka discrete values that convey:
 - Quantity
 - Quality
 - Statistics
- To be analyzed using statistical methods, your data set typically will have:
 - 1 observation per row
 - 1 variable per column
- A table with data from a paper, Wikipedia, or a news article is *not* necessarily a data set
 - These tables often summarize a data set, but hide the original data set used to generate them

	year	cleaned_vote_choice	province_recode	party_id	minority_thermometer	economic_retrospective_RECODED
1	2004	Liberal	British Columbia	none/DTS/DK	50	Got worse
2	2004	Liberal	Quebec	Liberal	90	Got worse
3	2004	Liberal	Ontario	Liberal	.	Got worse
4	2004	Liberal	Ontario	Liberal	90	Got worse
5	2004	Liberal	Ontario	Liberal	Really LIKE the group	Got Better
6	2004	Liberal	Manitoba	none/DTS/DK	70	Got worse
7	2004	Liberal	N&L	none/DTS/DK	.	Got Better
8	2004	Liberal	Ontario	Liberal	65	Got worse
9	2004	Liberal	British Columbia	Liberal	80	Got Better
10	2004	Liberal	New Brunswick	Liberal	Really LIKE the group	Stayed about the same
11	2004	Liberal	Quebec	Liberal	.	Got worse
12	2004	Liberal	N&L	none/DTS/DK	50	Got worse
13	2004	Liberal	N&L	none/DTS/DK	.	Got worse
14	2004	Liberal	New Brunswick	none/DTS/DK	75	Got worse
15	2004	Liberal	Ontario	Liberal	90	Got Better
16	2004	Liberal	Ontario	Liberal	.	Got worse
17	2004	Liberal	Ontario	Liberal	75	Got worse
18	2004	Liberal	Ontario	Liberal	95	Got Better
19	2004	Liberal	PEI	Liberal	50	Got worse
20	2004	Liberal	PEI	Liberal	.	Got worse
21	2004	Liberal	Ontario	Liberal	75	Got worse
22	2004	Liberal	Ontario	Liberal	80	Got worse
23	2004	Liberal	Ontario	Liberal	50	Got worse
24	2004	Liberal	Ontario	none/DTS/DK	80	Got Better
25	2004	Liberal	Quebec	Liberal	.	Got Better
26	2004	Liberal	N&L	none/DTS/DK	60	Got Better

NOT a Data Set

6.3 Summary Statistics

Table 2: Summary Statistics of All Variables (unweighted)

Variable	N	Mean	Std. Dev.	Min	Max
Vote choice	68,297	N/A	N/A	1	6
Minority Thermometer	65,262	66.62	25.43	0	100
Done for racial minorities	40,295	3.44	1.05	1	5
2019/2021 indicator	82,042	N/A	N/A	0	1
Party ID	75,214	N/A	N/A	1	7
Gender (female)	81,561	0.56	0.50	0	1
Quebec indicator	81,984	N/A	N/A	0	1
Income	75,292	2.93	1.40	1	5
Education	81,209	7.35	1.97	1	11
Age	81,241	49.69	16.96	17	115

Some Popular Types of Political Science Data

- Survey data
- Election results
- Country statistics
- International data
- Text analysis
- Scaled ideology data

Perusing the Data Buffet

- Today I will be *briefly* overviewing some of the most popular data sets
- I *highly* encourage you to use one of these in your poster
- This is **much** less work than collecting your own data
- It also makes your analysis stronger by giving you lots of *observations* (rows) that you can analyze

General Tips

- To analyze data, you will first need to download it
- Always download the data in Stata (.dta) format, if possible
- Stata can also read Excel and .csv files
- You will then need to access it on Stata on your virtual desktop
 - You can email the data to yourself
 - You can upload it to the cloud, then download it on Parallels
 - You can move it to the virtual desktop via USB drive
 - Or put it in a folder on your laptop that the virtual desktop can access

Survey Weights

- Later in the course we will talk about **weights** in survey data
- All survey data sets discussed today have one or more **weight** variables.
- These variables tell Stata to emphasize or deemphasize individual observations (rows) in analyses
- Q: Why do this?
 - A: If you do not use weights, your analysis will not be **representative**
- Example:
 - Survey sample is 5% age 18-24
 - Actual population: 12%
 - Solution: **weight** each 18-24 year-old respondent higher
 - Assumption: surveyed 18-24 year-olds have similar beliefs to unsurveyed 18-24 year-olds

American Politics Data Sets

The American National Election Study (ANES)



American National Election Studies

ANES Key Info

- Type: survey data
- Scope: all US presidential elections, 1952-present
- Observation unit: US citizens
- Good for analyzing:
 - Voter behavior
 - Voter attitudes
 - Changes in public opinion over time

ANES Important Links

- Website: <https://electionstudies.org/>
- Data download page:
<https://electionstudies.org/data-center/>
- Search survey questions:
<https://electionstudies.org/data-tools/anes-question-search/>
- Notes: you will need to register an account to download data (but not to do anything else)

The Cooperative Election Study (CES)



**COOPERATIVE
ELECTION
STUDY**

CES Key Info

- Type: survey data
- Scope: all US presidential + midterm elections, 2006-present
- Observation unit: US citizens
- Good for analyzing:
 - Voter behavior
 - Voter attitudes
 - Midterm elections (unlike the ANES!)
 - Congressional districts (unlike the ANES!)

CES Important Links

- Website: <https://cces.gov.harvard.edu/>
- Notes:
 - The CCES has a base survey AND “team” datasets
 - These team datasets have interesting questions, but fewer observations
- Search 2016/2018 team questions not in base survey:
<https://cooperativeelectionstudy.shinyapps.io/CESquestionsearch/>

The General Social Survey (GSS)

THE GENERAL SOCIAL SURVEY



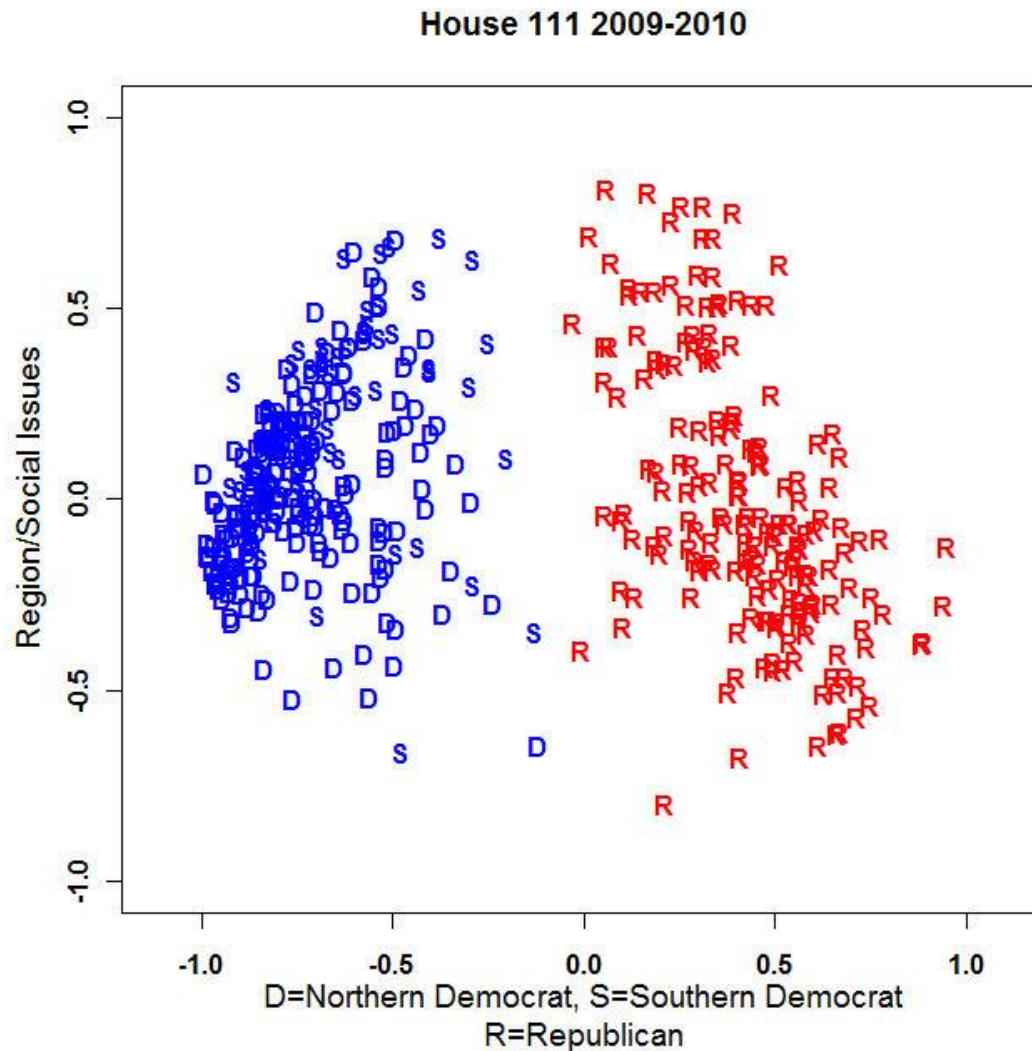
GSS Key Info

- Type: survey data
- Scope: 1972-present, annually (1972-94), even years (1994-present)
- Observation unit: US residents
- Good for analyzing:
 - Behavior not typically considered “political”
 - Attitudes not typically considered “political”

GSS Important Links

- Website: <https://gss.norc.org/>
- Data download page: <https://gss.norc.org/get-the-data/stata>
- Search survey questions and explore data: <https://gssdataexplorer.norc.org/>
- Notes:
 - The GSS is *less* likely to be a useful data set for you than the ANES, and CES – but it is very high quality and widely used

DW-NOMINATE



NOMINATE Key Info

- Type: scaled ideology data
- Scope: *every* 2-year US House and Senate session
- Observation unit: members of Congress
- Good for analyzing:
 - Ideology of members of Congress, congressional parties, and chambers
- “The modern study of the U.S. Congress would be simply unthinkable without NOMINATE legislative roll call voting scores.” – the Society for Political Methodology

NOMINATE Important Links

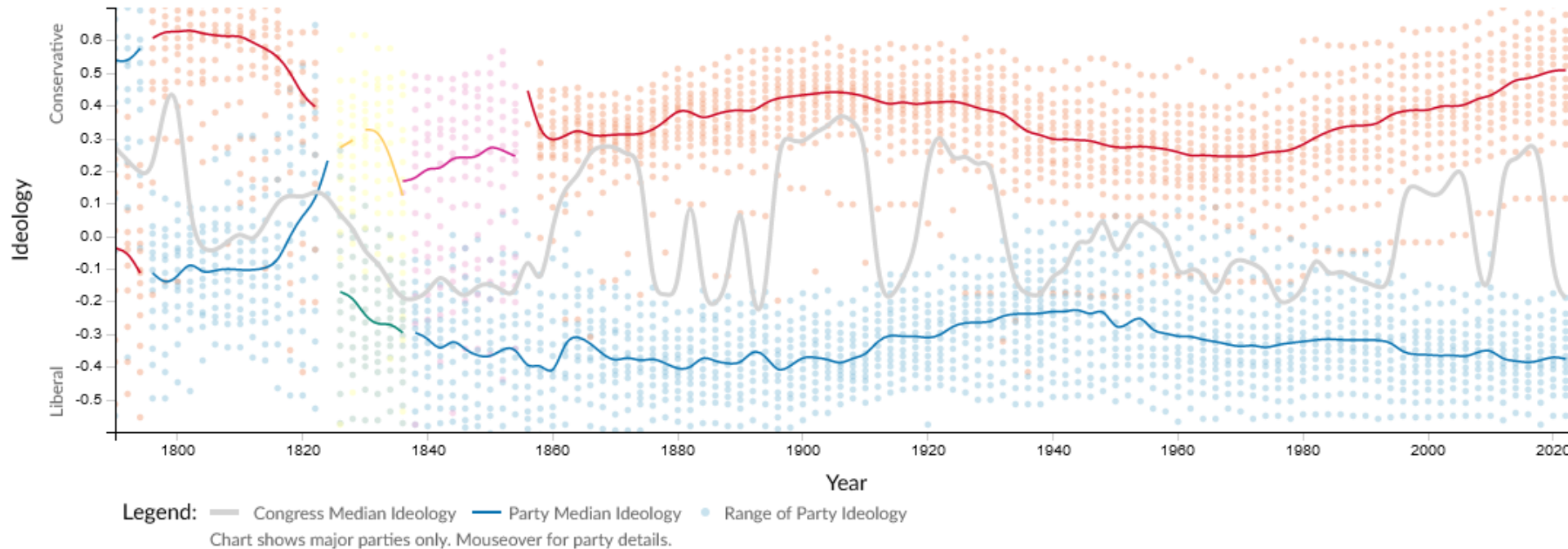
- Website: <https://voteview.com/>
- Data download page: <https://voteview.com/data>

Notes:

- NOMINATE works by mapping how often MCs *vote together*
- This turns out to be a very good (but imperfect!) measure of ideology
- There are two “axes” of NOMINATE. The 1st is ideology, the 2nd is harder to define

DW-NOMINATE Data Visual

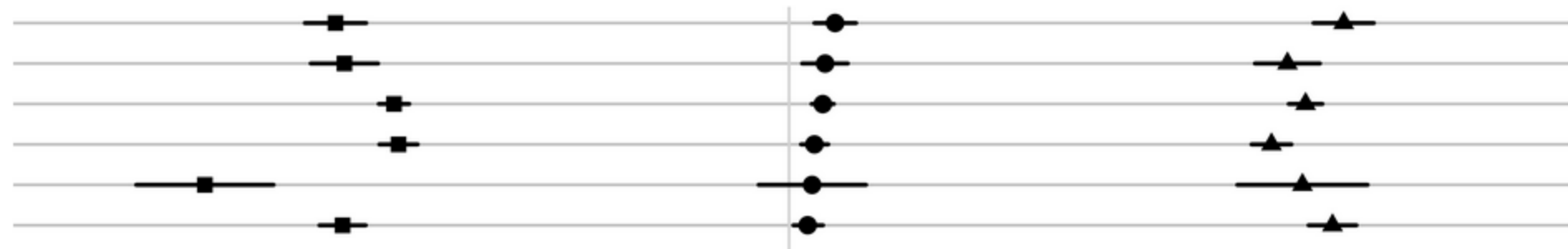
Congress at a Glance: Major Party Ideology



American Ideology Project (AIP)

The American Ideology Project

Research and Data on Ideology and Representation in the United States



AIP Key Info

- Type: scaled ideology data
- Scope: 2004-2021
- Observation unit: Geographic units (zip codes, states, counties, cities, congressional districts, state lege districts, school districts)
- Good for analyzing:
 - The left-right ideology of *public* ideology in the aggregate

Important Links

- Website:
<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BQKU4M>
- Notes:
 - The “American_ideology_project.pdf” file on page 2 is a guide to the data
 - You have multiple file formats you can download each data file in. Select “Stata Binary”
 - There are data files for both “presidential vote” (pvote) and “ideology” at each geographic level

Some Other Useful AP Data Sources

- [PollingReport.com](https://www.pollingreport.com/): aggregated polling data on major topics
- <https://polling.reuters.com/>: searchable polling from Reuters
- [CIRCLE](https://www.cIRCLE.nyu.edu/): data on civic + political engagement of young Americans
- [DIME](https://www.dimeproject.org/): an alternative to NOMINATE, based on donor networks
- [State Democracy Index](https://www.state-democracy-index.org/): scaled estimates of levels of democracy in US states, 2000-2018

Popular International Data Sets: Public Opinion, Parties, & Elections

Barometer Surveys

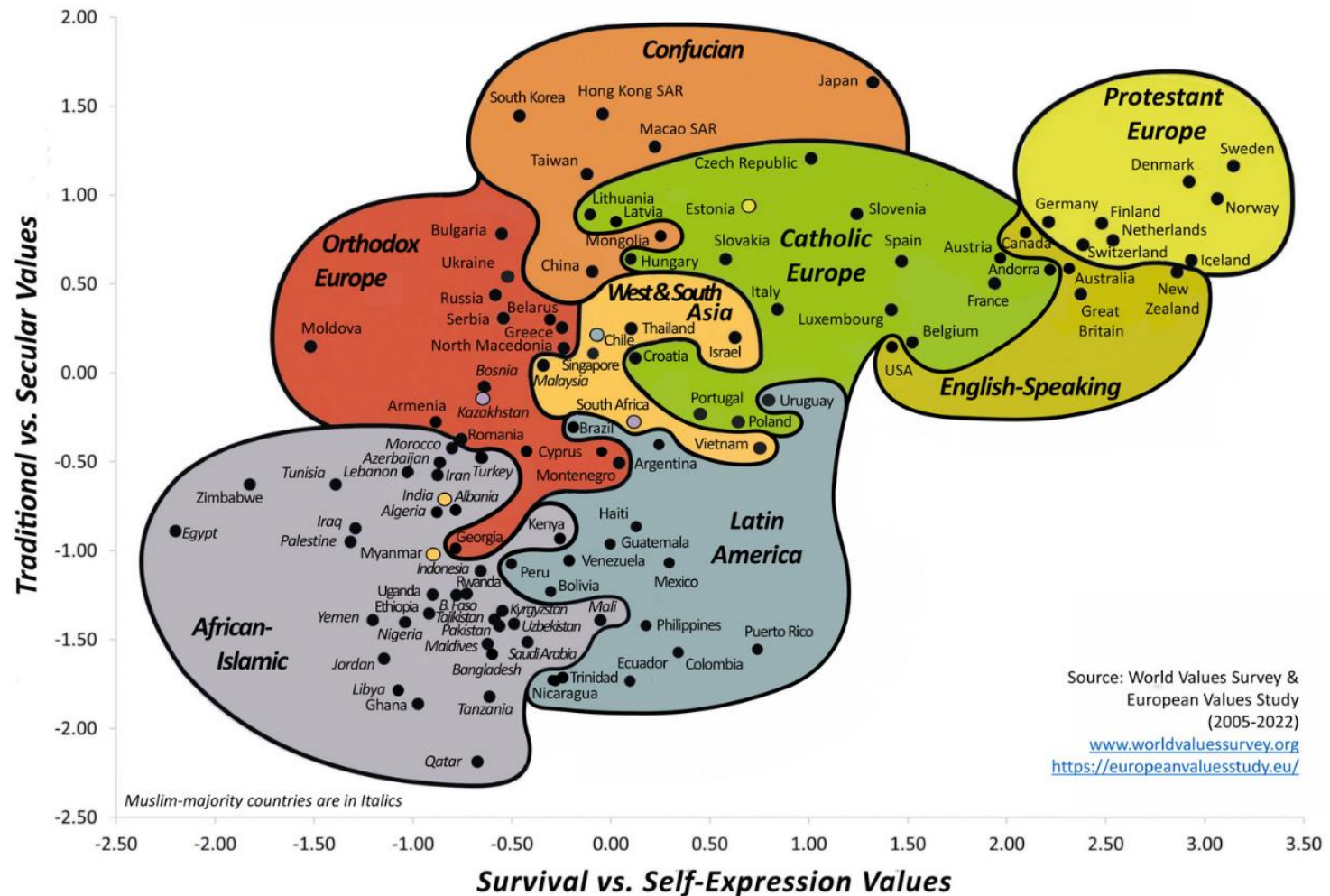
- These are massive public opinion surveys in different parts of the world. They provide data regionwide and by country.
- Africa: <https://www.afrobarometer.org/>
- Asia: <https://www.asianbarometer.org/>
- Europe: <https://europa.eu/eurobarometer/screen/home>
- Latin America: <https://www.vanderbilt.edu/lapop/>
- Middle East: <https://www.arabbarometer.org/>

Barometer Key Info

- Type: survey data
- Scope: different start dates, conducted in “waves”
- Observation unit: citizens in studied regions
- Good for analyzing:
 - Public opinion around the world
 - Comparing attitudes across countries in a region

World Values Survey (WVS)

The Inglehart-Welzel World Cultural Map 2022



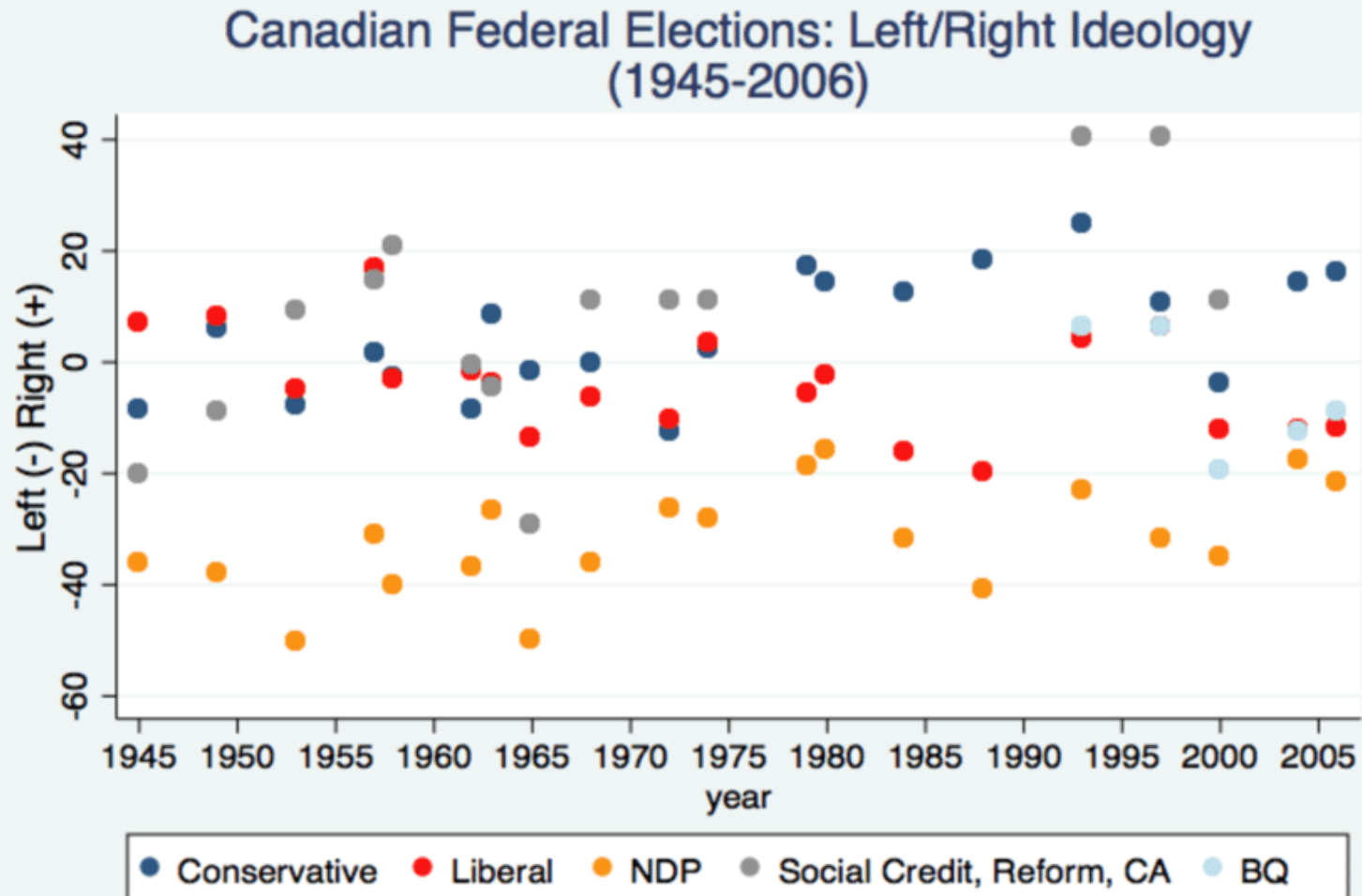
WVS Key Info

- Type: survey data
- Scope: 1981-present, 1x wave every 5 years
- Observation unit: residents of 120 countries
- Good for analyzing:
 - Individual-level public opinion across time, country, region, government type, etc.
 - Assessing political & cultural values

WVS Important Links

- Website: <https://www.worldvaluessurvey.org>
- Data download page:
<https://www.worldvaluessurvey.org/WVSContents.jsp>
- Search survey questions:
<https://www.worldvaluessurvey.org/WVSONline.jsp>
- Notes:
 - Reminder: the data are **not** representative unless you *weight* the. More info on WVS weights:
<https://www.worldvaluessurvey.org/WVSContents.jsp?CMSID=WEIGHT&CMSID=WEIGHT>

Comparative Manifestos Project (CMP)



Source: Comparative Manifesto Project

CMP Key Info

- Type: election results + scaled party data based on party *manifestos* (platforms)
- Scope: 1945-present, 50+ countries
- Observation unit: political parties with legislative representation
- Good for analyzing:
 - Party ideology/values over time
 - Comparing political parties

CMP Important Links

- Website: <https://manifestoproject.wzb.eu>
- Data download page:
<https://manifestoproject.wzb.eu/datasets>
- Search survey questions: https://visuals.manifesto-project.wzb.eu/mpdb-shiny/cmp_dashboard_dataset/
- Notes:
 - These data are amazing, but why might we be cautious?

Chapel Hill Expert Survey (CHES)



Chapel Hill Expert Survey

CHES Key Info

- Type: scaled survey data of political experts
- Scope: 1999-2019, selection of countries in Europe and Latin America
- Observation unit: political parties
- Good for analyzing:
 - Party ideology/values over time
 - Comparing political parties

CHES Important Links

- Website: <https://www.chesdata.eu>
- Europe data download page:
<https://www.chesdata.eu/ches-europe>
- Latin America data download page:
<https://www.chesdata.eu/chesla>
- Notes:
 - This is essentially an alternative to the CMP
 - Different pros, different cons

The Comparative Study of Electoral Systems (CSES)



A collaborative program of
research among election teams

Module 5 (2016-2021) focuses on citizens' perceptions of
political elites and out-groups, and the implications for
electoral democracy.



CSES Key Info

- Type: survey data + electoral system data
- Scope: elections in selected countries, 1996-present
- Observation unit: citizens
- Good for analyzing:
 - Voter behavior around the world
 - Impact of electoral systems on voter behavior + attitudes

CSES Important Links

- Website: <https://cses.org>
- Data download page: <https://cses.org/data-download/download-data-documentation/>
- Search survey questions: <https://cses.org/data-download/download-data-documentation/variable-table/>
- Notes:
 - Variable names differ across survey waves (see prior link)

Other National Election Surveys

- These are massive public opinion surveys, like the ANES, run domestically by political scientists in their own countries.
 - NOTE: English language data not always available!
- Many national election studies have been “folded into” the CSES (
- Some examples:
 - Canada: <http://www.ces-eeec.ca/>
 - UK: <https://www.britishelectionstudy.com/>
 - Germany: <https://www.gesis.org/en/elections-home/gles>

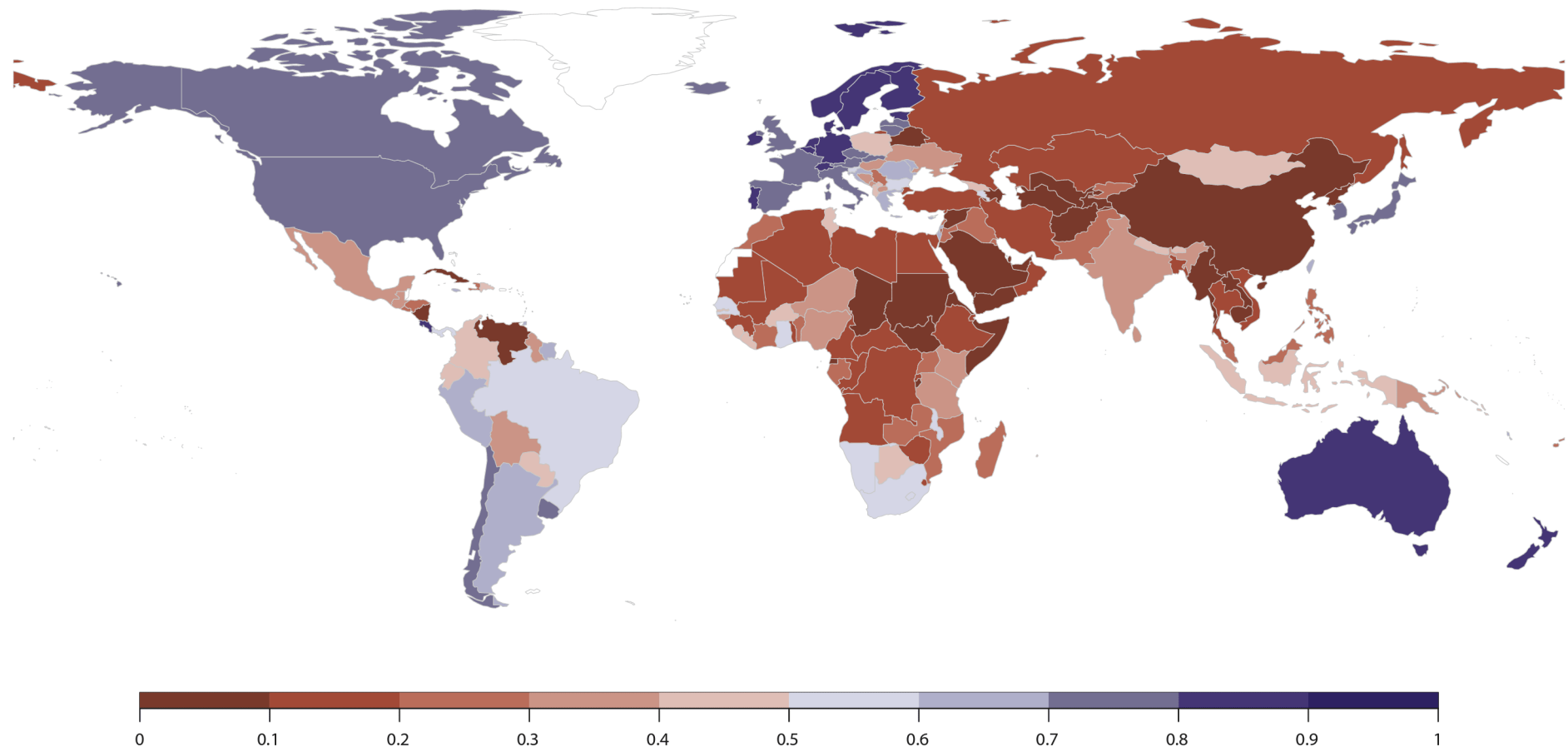
Some Other Useful Data Sources

- [CLEA](#): district-level legislative election results from 2,000+ elections in ~180 countries
- [Psephos](#): detailed election results from 1999-2022 from around the world
- [Party Personnel](#): data on candidate recruitment, committee assignment, and cabinet positions for 9 countries
- [Democratic Electoral Systems](#): country-level data on electoral rules and party systems, all democratic countries from 1946-2020

Popular International Data Sets: Democracy, Development, & International Relations

Varieties of Democracy (V-Dem)

FIGURE 2: STATE OF LIBERAL DEMOCRACY IN 2021



V-Dem Key Info

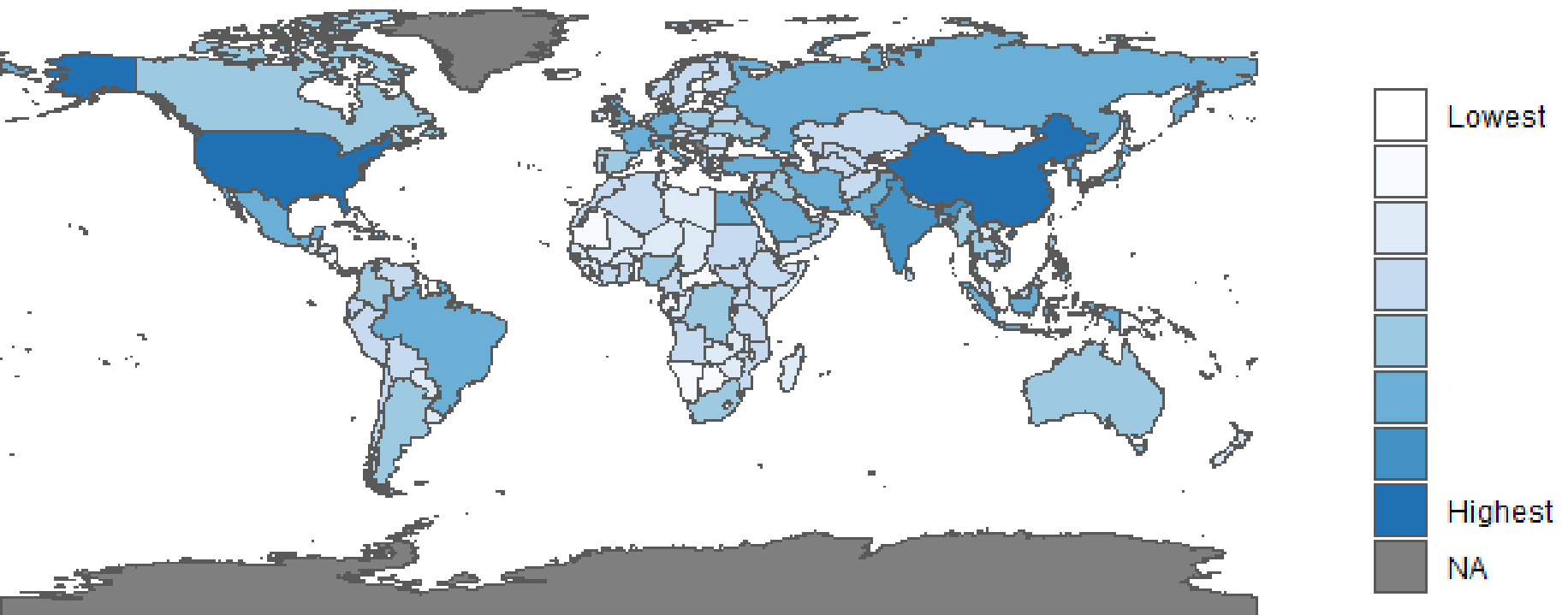
- Type: scaled measures of democracy
- Scope: all countries, 1789-2019
- Observation unit: country
- Good for analyzing:
 - Democracy (5 democracy indices available)
 - Country-level statistics
 - Easier to work with than World Bank data
- Caution: there is no “objective” measure of democracy!

V-Dem Important Links

- Website: <https://www.v-dem.net/>
- Data download page: <https://www.v-dem.net/data/the-v-dem-dataset/>
 - NOTE: you *probably* want the “Country-Year: V-Dem Core” dataset
- Search survey questions and visualize data: https://v-dem.net/data_analysis/VariableGraph/

Correlates of War (CoW)

2016 CINC Scores



CoW Key Info

- Scope: all countries, since 1816
- CoW hosts multiple data sets, including:
 - Conflict between states (threatened or actual)
 - Trade
 - World religion
 - Formal alliances between states
 - Countries with shared land/sea borders
 - Border changes
 - Intergovernmental organization membership
 - Colonial relationships
 - Data on military capacity
 - Historical data on wars
 - A historical list of states
 - Latitude/longitude of conflicts

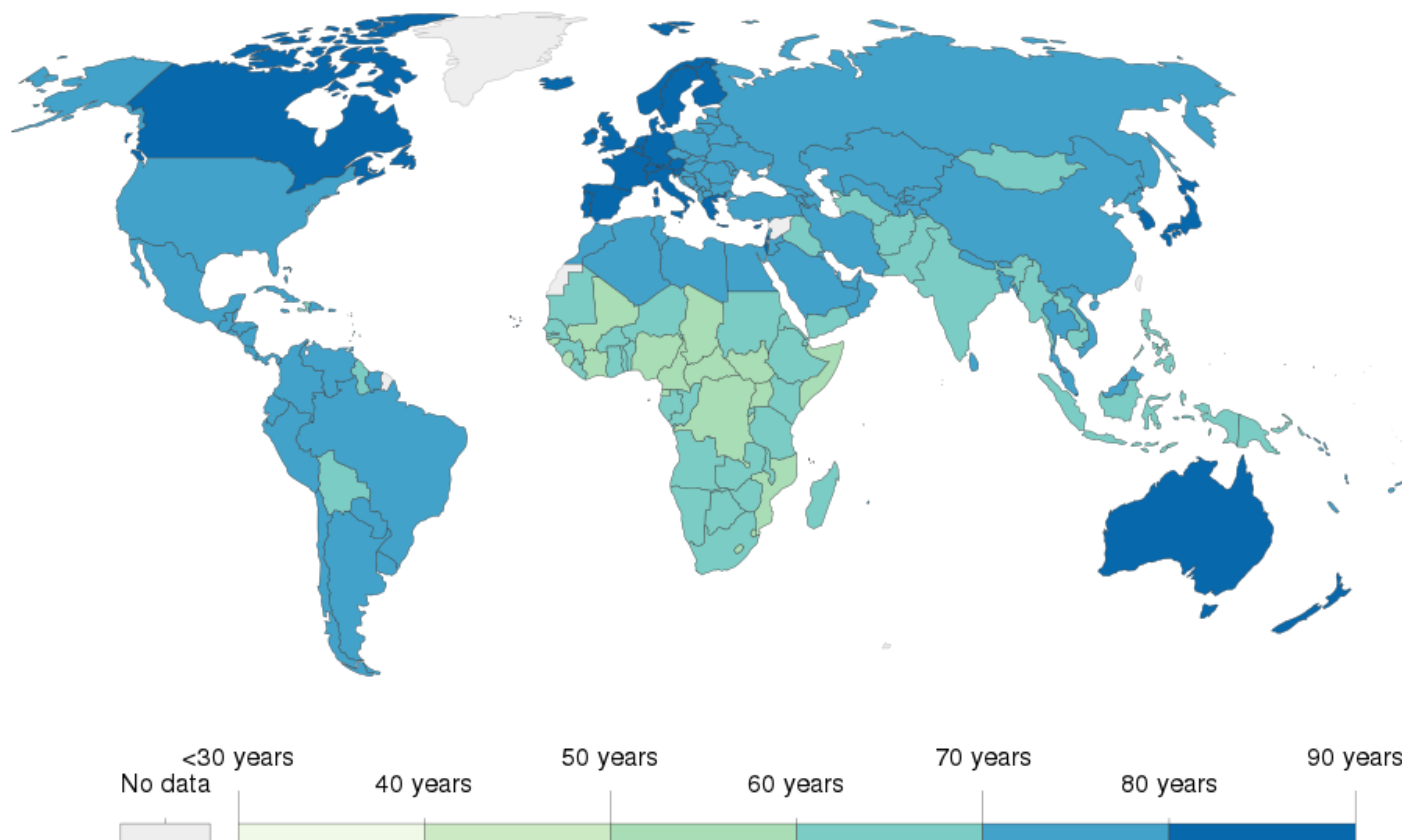
Important Links

- Website: <https://correlatesofwar.org/>
- Data download page:
<https://correlatesofwar.org/data-sets/>
- Notes:
 - The data vary in type:
 - Country-level (e.g. religion)
 - Country-pairs or country-groups (e.g., war or trade)
 - Organization-level (e.g. IGOs)

The World Bank

Life expectancy, 2016

The average number of years a newborn would live if age-specific mortality rates in the current year were to stay the same throughout its life.



World Bank Key Info

- Type: country summary data
- Scope: all countries, 1960-present
- Observation unit: countries
- Good for analyzing:
 - Country-level indicators, particularly economic
 - Country demographics

World Bank Important Links

- Website: <https://www.worldbank.org/en/home>
- Data download page:
<https://data.worldbank.org/about/get-started>
- The World Development Indicators database:
 - <https://databank.worldbank.org/source/world-development-indicators#>
- Notes:
 - The World Bank has a *ton* of data, so you will need to narrow by country/region, variable (“series”), and time period **before** downloading

Some Other Useful Data Sources

- [The Alliance Treaty Obligations and Provisions Project \(ATOP\)](#): military alliances from 1815-2018
- [POLITY5](#): an alternative to V-Dem
 - Site also hosts additional data on violence, state fragility, and conflict in India
- [All Minorities At Risk](#): data on the status of politically-active communal groups in all countries
- [Revolutionary Leaders and Revolutions](#): data on revolutionary leaders and governments

Advanced Data Sleuthing: ICPSR & Harvard Dataverse

ICPSR: The Google of Data Sets

- Trying to find data outside of one the data sets already mentioned? Search ICPSR!
- ICPSR archives over 500,000 data sets in the social sciences
- NOTE: it's quite possible the variable you want **is** in one of the aforementioned data sets, but you might not now until you look carefully!

For Everything Else, There's ICPSR!



An Example ICPSR Search

Search Results

Showing 1 - 50 of 488 results. [Show all](#)

GO

VIEW ALL

[search tips](#) ▼

Studies (488)

Variables (7)

Series (16)

Data-related Publications (12)

ICPSR Website (0)

 Summaries: ☒ Hidden

Sort by: Study Relevance ▼

⏮ ⏪ 1 2 3 4 5 ⏩ ⏭

Study Title/Investigator

Released/Updated

- [Data and Code for The Effects of the 1930s HOLC "Redlining" Maps](#)  2021-10-25
 Aaronson, Daniel; Hartley, Daniel; Mazumder, Bhashkar
- [Historic Redlining Scores for 2010 and 2020 US Census Tracts](#)  2021-10-15
 Meier, Helen C.S.; Mitchell, Bruce C.
- [The HOLC Maps: How Race and Poverty Influenced Real Estate Professionals' Evaluation of Lending Risk in the 1930s](#)  2022-12-29
 Fishback, Price; LaVoice, Jessica; Shertzer, Allison; Walsh, Randall
- [National Longitudinal Study of Adolescent to Adult Health \(Add Health\) Parent Study: Public Use, \[United States\], 2015-2017 \(ICPSR 37375\)](#) 2020-08-10
 Harris, Kathleen Mullan; Hotz, V. Joseph
- [ANES 1992 Time Series Study \(ICPSR 6067\)](#) 2016-06-23
 Miller, Warren E.; Kinder, Donald R.; Rosenstone, Steven J.; University of Michigan. Institute for Social Research. American National Election Studies

Harvard Dataverse: The Bing of Data Sets

- Harvard Dataverse hosts over 150,000 data sets
- It works very similarly to ICPSR

An Example Dataverse Search

redlining



Advanced Search

+ Add Data

☒ Dataverses (0)

☒ Datasets (4)

☒ Files (5)

Publication Year

2022 (4)

2016 (1)

2014 (1)

2013 (1)

2012 (2)

Subject

Social Sciences (1)

Author Name

Cohen-Cole, Ethan (1)

Connolly, N.D.B. (1)

Kenneth Brevoort (1)

Marciano, Richard (1)

Nelson, Robert (1)

More...

Author Affiliation

Alvarez & Marsal (1)

Federal Reserve Board (1)

1 to 9 of 9 Results

Sort ▾

Redlining in Boston



Apr 4, 2022 - Boston Area Research Initiative's Boston Data Portal

Nelson, Robert; Winling, LaDale; Marciano, Richard; Connolly, N.D.B., 2022, "Redlining in Boston", <https://doi.org/10.7910/DVN/WXZ1XK>, Harvard Dataverse, V1, UNF:6:X+Wfu6+NEIBG9UKk87G4tw== [fileUNF]

... Between 1935 and 1940 the federal government's Home Owners' Loan Corporation (HOLC) classified the neighborhoods of 239 cities according to their perceived investment risk. This practice has since been referred to as "**redlining**," as the neighborhoods classified as being the highest risk for investment were often ...

Producer URL: <https://dsl.richmond.edu/panorama/redlining>

Replication data for: Credit Card Redlining



Mar 13, 2013 - Review of Economics and Statistics Dataverse

Cohen-Cole, Ethan, 2013, "Replication data for: Credit Card Redlining", <https://doi.org/10.7910/DVN/PCDG7J>, Harvard Dataverse, V1

Related Publication Citation: Cohen-Cole, Ethan, (2011) "Credit Card **Redlining**." Review of Economics and Statistics 93:2, 700-713.

Replication data for: Credit Card Redlining Revisited



Apr 10, 2012 - Review of Economics and Statistics Dataverse

Kenneth Brevoort, 2012, "Replication data for: Credit Card Redlining Revisited", <https://doi.org/10.7910/DVN/MOTPKF>, Harvard Dataverse, V1

Related Publication Citation: Brevoort, Kenneth P., (2011) "Credit Card **Redlining** Revisited." Review of Economics and Statistics 93:2, 714-724.

Advanced Data Sleuthing

- What if the data you want is not on ICPSR or Harvard Dataverse?
- Data sets are often on authors' websites or on the Harvard Dataverse
- If you find a paper with data you want that's not online, you can email the authors!
 - But there is no guarantee they will respond in time
- For this class, I encourage (but do not require) you to use a data set that is already online and public

This is a Lot!

- Feel overwhelmed? That's ok!
- The point of today *is* to dump too much info on you. Why?
- When you are looking for data later in the course, you can open this slide deck and use it as a guide
- You can also use this as a **first stop** before coming to my office hours about finding data
- You do **NOT** need to pick a data set for your Research Question assignment....
- ... But you **DO** for the worksheet due next Friday (Feb. 10)

Q&A