

U.S. State-Level Data for Political Science

Updated August 26, 2020

This dataset contains information on more than 100 variables from a variety of sources for each U.S. state. DC is included but is missing data for many variables. The data are from 2010-2020, mostly from the last three years. All nominal variables are coded as numbers corresponding to categories. All binary (true/false) variables are coded as 1 (true) or 0 (false). Dataset compiled by Nathan Morse (nam@psu.edu).

Population	2
Overall Characteristics	2
Demographics	2
Race & Ethnicity	2
Economy	3
Economic Indicators	3
Environment	3
Politics	4
Voting	4
Polling	4
Party Control	5
Society	5
Culture	5
Crime & Criminal Justice	5
Guns	6
Police Brutality	6
Health	7
COVID-19	7
Public Health	8
Insurance	8
Health Policy	8
Drugs & Alcohol	9
Tips for using this dataset in R	10

Population

Overall Characteristics

Variable	Description	As of	Type	Source
pop_total	Population: Total	2020	Numeric	WPR
pop_growth	Population: Growth since 2010 (%)	2020	Numeric	WPR
pop_density	Population: Density (population/square mile)	2020	Numeric	WPR
pop_urban	Population: Urban (% of population)	2010	Numeric	Census

Demographics

Variable	Description	As of	Type	Source
pop_children	Age: Under 18 (% of population)	2018	Numeric	KFF
pop_elderly	Age: 65 or older (% of population)	2018	Numeric	KFF
pop_citizen	Citizenship: U.S. citizen (% of population)	2018	Numeric	KFF
pop_female	Gender: Female (% of population)	2018	Numeric	KFF
pop_hs	Education: High school degree or higher (% of population)	2020	Numeric	WPR
pop_college	Education: Bachelor's degree or higher (% of population)	2020	Numeric	WPR

Race & Ethnicity

Variable	Description	As of	Type	Source
pop_white	White (% of total population)	2018	Numeric	KFF
pop_black	Black (% of total population)	2018	Numeric	KFF
pop_hispanic	Hispanic (% of total population)	2018	Numeric	KFF
pop_native	American Indian/Alaska Native (% of total population)	2018	Numeric	KFF
pop_asian	Asian (% of total population)	2018	Numeric	KFF
pop_pacific	Hawaiian/Pacific Islander (% of total population)	2018	Numeric	KFF
pop_two	Two or more races (% of total population)	2018	Numeric	KFF

Economy

Economic Indicators

Variable	Description	As of	Type	Source
econ_poverty	Poverty rate (% of population under federal poverty level)	2018	Numeric	<u>KFF</u>
econ_income	Median annual household income (\$)	2017	Numeric	<u>KFF</u>
econ_ineq	Income inequality (Gini coefficient) <ul style="list-style-type: none"> 0 = Perfect equality 100 = Extreme inequality 	2020	Numeric	<u>WPR</u>
econ_homeless	Population of people experiencing homelessness	2019	Numeric	<u>KFF</u>
econ_unemp	Unemployment rate (% of labor force)	3/2020	Numeric	<u>KFF</u>
econ_cost	Cost of living index <ul style="list-style-type: none"> <100 = Lower than the US average 100 = Equal to the US average >100 = Greater than the US average 	2020	Numeric	<u>WPR</u>
econ_gsp	Gross state product (millions of \$)	2018	Numeric	<u>KFF</u>
econ_corptax	Corporate tax rate (% of income)	2020	Numeric	<u>WPR</u>
econ_taxes	Tax collections per capita (\$)	2018	Numeric	<u>KFF</u>
econ_spending	State expenditures per capita (\$)	2018	Numeric	<u>KFF</u>

Environment

Variable	Description	As of	Type	Source
enviro_air	Air quality (see <u>WPR</u> for details) <ul style="list-style-type: none"> ≤50 = good, little to no health risk >50 = moderate risk 	2020	Numeric	<u>WPR</u>
enviro_temp	Average temperature (F)	2020	Numeric	<u>WPR</u>
enviro_area	Land area (square miles)	2020	Numeric	<u>WPR</u>
enviro_human	Opinion: Agree that global warming is caused mostly by human activities (% of adults)	2019	Numeric	<u>Yale</u>
enviro_worried	Opinion: Somewhat or very worried about global warming (% of adults)	2019	Numeric	<u>Yale</u>
enviro_reg	Opinion: Somewhat or strongly support regulating CO2 as a pollutant (% of adults)	2019	Numeric	<u>Yale</u>

Politics

Voting

Variable	Description	As of	Type	Source
vote_clinton	2016 election: votes for Hillary Clinton (% of total votes)	2016	Numeric	WPR
vote_trump	2016 election: votes for Donald Trump (% of total votes)	2016	Numeric	WPR
vote_turnout16	Voter turnout rate for 2016 general election (% of voting-eligible population)	2016	Numeric	McDonald
vote_turnout18	Voter turnout rate for 2018 general election (% of voting-eligible population)	2018	Numeric	McDonald
vote_registered	Registered voters (% of voting-eligible population)	2018	Numeric	KFF
vote_id	Voter ID laws (see NCSL for details) <ul style="list-style-type: none"> • 0 = no ID required • 1 = non-strict, non-photo • 2 = non-strict, photo • 3 = strict, non-photo • 4 = strict, photo 	2020	Ordinal/ Nominal	NCSL
vote_mail	Mail voting laws <ul style="list-style-type: none"> • 0 = Excuse required for absentee voting • 1 = Absentee voting allowed for all • 2 = Ballots mailed directly to all voters 	2020	Ordinal/ Nominal	NYT

Polling

Variable	Description	As of	Type	Source
poll_trump	Trump approval rating (% of population)	2017	Numeric	Gallup
poll_dem	Political affiliation: Democrat or Democrat-leaning (% of population)	2017	Numeric	Gallup
poll_rep	Political affiliation: Republican or Republican-leaning (% of population)	2017	Numeric	Gallup
poll_class	Classification of public's political affiliations <ul style="list-style-type: none"> • -2 = Solid Republican • -1 = Leans Republican • 0 = Competitive • 1 = Leans Democrat • 2 = Solid Democrat 	2017	Ordinal	Gallup

Party Control

Variable	Description	As of	Type	Source
party_governor	Governor's party: Democrat (true/false)	2020	Binary	<u>KFF</u>
party_senate	State senate majority party: Democrat (true/false)	2020	Binary	<u>KFF</u>
party_house	State house majority party: Democrat (true/false)	2020	Binary	<u>KFF</u>
party_ag	State attorney general's party: Democrat (true/false)	2020	Binary	<u>KFF</u>

Society

Culture

Variable	Description	As of	Type	Source
culture_relig	Religiosity (% of adults who are highly religious)	2016	Numeric	<u>Pew</u>
culture_lgbt	LGBT population (per 100,000)	2020	Numeric	<u>WPR</u>
culture_ios	iPhone users (% of smartphone market)	2018	Numeric	<u>DeviceAtlas</u>
culture_football	College football revenue (\$)	2018	Numeric	<u>US DoEd</u>

Crime & Criminal Justice

Variable	Description	As of	Type	Source
crim_prison	Prison population	2017	Numeric	<u>SP</u>
crim_race	Racial/ethnic disparity in imprisonment (black to white ratio) <ul style="list-style-type: none">• 1 = Black imprisonment rates are directly proportional to population• >1 = Black imprisonment rates are disproportionately high	2017	Numeric	<u>SP</u>
crim_disen	Felony disenfranchisement rate (% of total population disenfranchised from voting due to felony convictions)	2017	Numeric	<u>SP</u>
crim_penalty	Death penalty status <ul style="list-style-type: none">• 0 = Inactive (death penalty is abolished)• 1 = Moratorium (legal, but governor doesn't use it)• 2 = Active (death penalty is legal and used)	2020	Nominal	<u>DPIC</u>
crim_murders	Homicide rate (deaths per 100,000)	2018	Numeric	<u>CDC</u>

Guns

Variable	Description	As of	Type	Source
gun_owners	Gun ownership rate (% of population)	2020	Numeric	WPR
gun_count	Number of registered guns	2020	Numeric	WPR
gun_deaths	Firearm death rate (deaths per 100,000)	2018	Numeric	KFF
gun_reg	Firearm registration required (true/false)	2020	Binary	WPR
gun_carry	Permit required to carry (true/false)	2020	Binary	WPR
gun_purchase	Permit required to purchase (true/false)	2020	Binary	WPR
gun_open	Open carry legal (true/false)	2020	Binary	WPR
gun_score	Gun law score <ul style="list-style-type: none"> 0.00 = F (weak gun laws) 4.00 = A (strict gun laws) 	2020	Ordinal	Giffords

Police Brutality

Variable	Description	As of	Type	Source
blm_killings	Number of people killed by police since 2012	8/20/20	Numeric	FE
blm_black	Black people killed by police since 2012 (% of police killings)*	8/20/20	Numeric	FE
blm_race	Racial disparity ratio for police killings (% of people killed by police who were Black divided by % of population that is Black) <ul style="list-style-type: none"> <1 = Black people are less likely to be killed 1 = Black people are killed directly proportional to population >1 = Black people are more likely to be killed 	8/20/20	Numeric	FE
blm_protests	Number of antiracism protests held in May and June 2020	6/30/20	Numeric	CCC
blm_arrests	Antiracism protests with arrests (% of protests that resulted in at least 1 arrest)*	6/30/20	Numeric	CCC
blm_injuries	Antiracism protests with injuries (% of protests that resulted in at least 1 participant or police injury)*	6/30/20	Numeric	CCC
blm_teargas	Antiracism protests with tear gas (% of protests at which police used tear gas)*	6/30/20	Numeric	CCC

* Race data were not available for all police killings. Cases with missing data were excluded from these calculations, although they may be disproportionately more likely to be racial minorities. Likewise, many protests had no information on arrests, injuries, and tear gas, so cases with missing data were excluded as well.

Health

COVID-19

Variable	Description	As of	Type	Source
covid_cases	Total confirmed cases	8/18/20	Numeric	CDC
covid_rate	Infection rate (cases per 100,000)	8/18/20	Numeric	CDC
covid_deaths	Total confirmed deaths	8/18/20	Numeric	CDC
covid_deathrate	Deaths rate (deaths per 100,000)	8/18/20	Numeric	CDC
covid_tests	Total tests performed	8/19/20	Numeric	CDC
covid_pos	Positivity rate (range) <ul style="list-style-type: none"> • 0 = 0-5% of tests • 1 = 6-10% of tests • 2 = 11-20% of tests 	8/19/20	Ordinal	CDC
covid_race	Racial disparity ratio for COVID-19 infections (% of infected people who are Black divided by % of population that is Black) <ul style="list-style-type: none"> • <1 = Black infection rates are disproportionately low • 1 = Black infection rates are directly proportional to population • >1 = Black infection rates are disproportionately high 	8/17/20	Numeric	KFF
covid_open	Openness Score <ul style="list-style-type: none"> • 0 = Full lockdown • 100 = Completely open 	8/19/20	Numeric	MultiState
covid_masks	Statewide mask mandates <ul style="list-style-type: none"> • 0 = No mask mandate • 1 = Required for certain industry employees only • 2 = Required inside business/public buildings • 3 = Broad public outside/inside mask mandate 	8/5/20	Ordinal/ Nominal	MultiState
covid_dist	Social distancing: change in mobility to retail and recreation sites (see Google for more details) <ul style="list-style-type: none"> • <0 = Residents visit retail and recreational sites less frequently now than in January 2020 • 0 = No change in visits to sites • >0 = Residents visit sites more frequently now than in January 2020 	8/16/20	Numeric	Google
covid_calls	Do Not Call reports: number of complaints of Do Not Call violations since the pandemic began	8/16/20	Numeric	FTC

Public Health

Variable	Description	As of	Type	Source
health_life	Life expectancy (years)	2015	Numeric	KFF
health_exercise	Frequent exercise rate (% of population that reports exercising frequently)	2017	Numeric	Gallup
health_vegan	Vegan index (number of all-vegan restaurants per million people)	2016	Numeric	HealthIQ
health_weight	Overweight rate (% of population reporting BMI over 25)	2018	Numeric	KFF
health_abort	Number of legal abortions	2015	Numeric	KFF
health_deaths	Death rate (deaths per 100,000)	2018	Numeric	KFF
health_heart	Heart disease death rate (deaths per 100,000)	2018	Numeric	KFF
health_mental	Mental illness rate (% of population)	2018	Numeric	KFF
health_suicide	Suicide rate (deaths per 100,000)	2018	Numeric	KFF

Insurance

Variable	Description	As of	Type	Source
ins_employer	Insured by employer (% of population)	2018	Numeric	KFF
ins_medicaid	Insured by Medicaid (% of population)	2018	Numeric	KFF
ins_medicare	Insured by Medicare (% of population)	2018	Numeric	KFF
ins_none	Uninsured (% of population)	2018	Numeric	KFF
ins_market	Type of marketplace (see KFF for details) <ul style="list-style-type: none"> • 0 = Federally-facilitated marketplace • 1 = State-based marketplace–Federal platform • 2 = State-based marketplace 	2020	Nominal	KFF

Health Policy

Variable	Description	As of	Type	Source
hp_medicaid	Annual Medicaid spending (millions of \$)	2018	Numeric	KFF
hp_medicare	Annual Medicare spending (millions of \$)	2014	Numeric	KFF
hp_family	Paid family leave laws in place (true/false)	2020	Binary	KFF
hp_sick	Paid sick leave laws in place (true/false)	2020	Binary	KFF

Drugs & Alcohol

Variable	Description	As of	Type	Source
drugs_alc	Alcohol consumption (gallons per capita)	2020	Numeric	<u>WPR</u>
drugs_alcrate	Alcoholism rate (% of adult population)	2018	Numeric	<u>KFF</u>
drugs_cigs	Average cigarette pack price (\$)	2020	Numeric	<u>WPR</u>
drugs_smokes	Smoking rate (% of population that regularly smokes)	2018	Numeric	<u>KFF</u>
drugs_overdose	Drug overdose rate (deaths per 100,000)	2018	Numeric	<u>KFF</u>
drugs_recpot	Recreational marijuana legalized (true/false)	2020	Binary	<u>DISA</u>
drugs_medpot	Medicinal marijuana legalized (true/false)	2020	Binary	<u>DISA</u>

Tips for using this dataset in R

Removing variables. To select only the variables you need, an easy tool is the `select` function in the package `dplyr`. The first argument is the name of the dataset (`data_full` in this example), followed by the names of variables you want to keep.

```
library(dplyr)
data <- select(data_full, state, pop_total, econ_taxes, poll_dem)
```

Removing an observation. Because DC is missing data for many variables, you may wish to omit it entirely.

```
data <- data_full[-(data_full$state == "District of Columbia"),]
```

Converting a numerically coded variable to a nominal variable. All nominal variables in this dataset are coded as numbers corresponding to categories. R may misinterpret these variables as numerical, but you can use `as.factor` to tell R that a variable is nominal.

```
data$poll_class <- as.factor(data$poll_class)
```