

Democracy Fund + UCLA Nationscape User Guide

SEPTEMBER 2020

TABLE OF CONTENTS

Acknowledgements.....	3
Release	4
Methodology and Representativeness Assessment.....	5
Data Collection	6
Data Analysis, Weights, and Variance Estimation.....	7
Orientation to Data Files.....	8
Terms of Use	9
Appendices: Change Log, Weighting procedure.....	10

Acknowledgements

Nationscape is funded by The Democracy Fund, The Klarman Family Foundation, and UCLA with additional support from the Marvin Hoffenberg Chair in American Politics and Public Policy at UCLA through a partnership agreement between these parties.

The study was designed and data collection was managed by Nationscape Principal Investigators, Chris Tausanovitch (Associate Professor, Department of Political Science, UCLA) and Lynn Vavreck (Marvin Hoffenberg Professor of American Politics and Public Policy, UCLA) with design input from the Democracy Fund Voter Study Group, including former research director, John Sides (Professor, Department of Political Science, Vanderbilt University) and Robert Griffin, who currently holds that position.

UCLA graduate students staff Nationscape and execute the fielding of the survey each week. They are: Alex Rossell Hayes, Project Manager; Derek Holliday, Project Coordinator; and Aaron Rudkin, Project Developer. Tyler Reny worked as Project Coordinator from July 2019–August 2020.

Data collection is performed by LUCID, Inc., which collaborated with Nationscape personnel on aspects of the study's design, sampling plan, and feasibility. Lead personnel on the project at LUCID include: Lauren Astrachan, Analyst, Project Success; Ashley Burkardt, Lead, Project Success; Kyle Cousans, Senior Manager, Sales Engineering; Allison St. Martins, Associate Account Manager; Mary Kate Hutchinson, Senior Manager, Project Success; Wiley S. Davis, Analyst, Project Success.

Release

This document is a brief guide to Democracy Fund + UCLA Nationscape to accompany the second release of the Phase 1 data and the first release of the Phase 2 data in August 2020.

Nationscape is a 16-month election study conducted by researchers at UCLA. The Nationscape Project did not exist in election years prior to 2020. The project completes roughly 6,250 interviews each week. It started in July of 2019 and will conclude in January of 2021.

Phase 1 of the data, released in January of 2020, includes nearly 156,000 cases collected over 24 weeks. Data collection began with the week of July 18, 2019 and concluded with the week of December 26, 2019 (last interview on January 1, 2020).

Phase 2 of the data, released in September of 2020, includes a re-release of Phase 1 data and new data from January 2020 to July 2020 (Phase 2 data). This brings the total cases to 318,697.

Each weekly survey is released as its own dataset. There is no cumulative file release.

Current Release

This data release in Summer of 2020 includes a re-release of Phase 1 (Phase 1, version 20200814) and an initial release of Phase 2 data (Phase 2, version 20200814). The summer 2020 release is the second public release of Nationscape data. The re-release of Phase 1 (original version 20200131) is being done to add additional party identification variables so that users may generate their own composite party measures. It also includes a change to the Nationscape composite seven-category party measure (see notes attached to this document for details). Small changes to Phase 1 (version 20200131) weights may also appear due to this change.

If you find errors or have comments or questions about the data, please communicate with Nationscape Project staff by writing to Chris Tausanovitch at ctausanovitch@ucla.edu.

Citation

Citations to the datasets should be to the weekly studies as follows:

Tausanovitch, Chris and Lynn Vavreck. 2020. *Democracy Fund + UCLA Nationscape, October 10-17, 2019* (version 20200814). Retrieved from [URL].

Methodology and Representativeness Assessment

A detailed discussion of the sampling methodology, survey execution, and assessment of representativeness of the Nationscape Project can be found in:

Tausanovitch, Chris and Lynn Vavreck, Tyler Reny, Alex Rossell Hayes, and Aaron Rudkin. 2019. "Democracy Fund + UCLA Nationscape Methodology and Representativeness Assessment."

This document can be downloaded [here](#) by clicking on "Nationscape Methodology and Representativeness Assessment." These analyses were performed in July of 2019 and use an initial set of weights that have since been updated to allow for more specificity.

A description of the current weighting methodology is included at the end of this document.

Data Collection

Field dates: Nationscape conducts weekly surveys. The first wave went in to the field on July 18, 2019. For Phase 1 the last wave went in to the field on December 26, 2019. For Phase 2 the last wave went in to the field on June 25, 2020. The name of each dataset reflects its first field date. Each survey is in the field for one week.

Mode of interview: Interviews are conducted online anywhere the respondent has access to a networked computer or mobile device.

Interview length: The weekly questionnaires are designed for a 15-minute median administration time, not counting screening time by the sample provider. Observed median times to completion vary from 13 and a half to roughly 17 minutes depending on the week.

Languages: Nationscape pilot surveys were offered in Spanish or English with respondents able to choose their language question by question. Due to lack of use, the Spanish option was discontinued upon official launch of wave 1.

Participation rates: On average across all waves in this release, of those selected to be interviewed, roughly 12% decline immediately. Another 5% or so drop off elsewhere in the survey without completing. We remove about 8% for speeding or straight-lining through the survey. This results in an average yield of roughly 75% of the original invited sample, depending on wave. We define speeding as completing the survey in fewer than 6 minutes and straight-lining as selecting the same response for every question in the three policy question batteries.

Data Analysis, Weights, and Variance Estimation

Nationscape data can be analyzed using individual waves of the survey or any combination of waves. The number of completed interviews each week is roughly constant, so no week should dominate analysis in combined datasets, but researchers should be aware that some weeks have more completions than others.

Analyses should be weighted to accurately represent the population of interest. Sampling error calculations should account for the sample design and the effects of weighting on variance.

Weights

We provide weights in the “weight” variable that can be used for a national sample.

The sample is not a simple random sample nor is it a random sample of any kind. Please see the “[Democracy Fund + UCLA Nationscape Methodology and Representativeness Assessment](#)” for details on how weights are constructed to hit national targets and which national targets are used.

Instead, we use purposive sampling (selecting respondents based upon their characteristics) to obtain a sample that is constructed to be representative of the population in terms of a specified set of characteristics. The sample is not constructed from Lucid’s pre-existing targets. Nationscape staff designed the selection criteria and targets after several weeks of pilot waves in 2019 (see the Methodology Assessment linked above).

Researchers should take care in calculating margins of error or standard errors due to the non-random nature of the sample. Standard calculations based on random sampling will underestimate the true magnitude of random error in our sample. As a starting place, we recommend the adjustment proposed by Rivers and Bailey (2009).¹

1 Rivers, Douglas, and Delia Bailey. “Inference from matched samples in the 2008 US national elections.” *Proceedings of the joint statistical meetings*. Vol. 1. Palo Alto, CA: YouGov/Polimetrix, 2009.

Orientation to Data Files

Data files

The data files being released in the Summer 2020 release are constructed for delivery in Stata (.dta) format and we recommend this download. Each file contains a “readme” file to describe special features of that week’s data collection, if any.

Variables

There are roughly 200 variables in each weekly file.² They are named to reflect the topic they measure. For example, party identification in three categories is named “pid3.” Each file contains variable labels and value labels for each variable.

A spreadsheet of when each question was asked wave by wave is available for download along with this file.

Missing data

Missing data are mainly coded to indicate the reason they are missing using the following codes and categories:

888	Asked in this wave, but not asked of this respondent
999	Not sure, don’t know
“.”	Respondent skipped

Codebook

Each weekly survey has its own codebook and banner book of results. Nationscape staff have made every effort to maintain consistency across the survey waves on variable names, outcome categories, outcome codes and labels, and question wording. When variable names change it is most likely because a change was made to the question stem or outcome categories. Any deviations from consistency are noted in the “readme” file associated with each week’s data collection and at the digest at the end of this document.

² There is a rotating set of questions that vary over weeks of fielding, resulting in different numbers of variables in different weeks.

Terms of Use

Researchers using these data agree to:

1. Use them only for statistical purposes and not for investigation of specific survey respondents.
2. Make no use of the identity of any survey respondents discovered intentionally or inadvertently, and to advise Nationscape personnel of any such discovery immediately by emailing Chris Tausanovitch at ctausanovitch@ucla.edu.

3. Cite Nationscape data and documentation in work that uses Nationscape data and documentation. Citations to the data sets should be as follows:

Tausanovitch, Chris and Lynn Vavreck. 2020. *Democracy Fund + UCLA Nationscape, October 10-17, 2019* (version 20200814). Retrieved from [URL].

4. Acknowledge that the original collectors of the data, UCLA, LUCID, and Democracy Fund, and all funding agencies, bear no responsibility for the use of the data or for interpretations or inferences based upon such issues.

Appendix 1: Weekly Change Log

The following changes were made to the survey at the point noted below. All changes persisted through subsequent waves unless otherwise noted.

Wave 2: Disabled RealAnswer and ReCAPTCHA for last 700 completes in this week's data collection and in future waves. We also made a small edit to the question called "environment." We dropped "government" from "government investment" in the question stem and added "Not Sure" to the response options.

Wave 3: Launched 1 day late.

Wave 5: Began asking Gubernatorial and Senate general election ballot items only to respondents who have gubernatorial or senate elections in 2019 or 2020 based on their state as provided by Lucid.

Wave 10: New display logic was incorporated to show Democratic primary ranking questions to just those who indicated that they were voting in the Democratic primary or were not sure. These variables include rank_dems_1, rank_dems_2, and rank_dems_3.

Wave 12: We made a small edit to the gun ownership question. Categories are now: (1) Yes, I personally own a gun; (2) I don't, but a member of my household owns a gun; (3) No one in my house owns a gun; and (4) Not sure. The change is to the underlined text above.

Wave 16: Beto O'Rourke removed from democratic vote intent and ranking questions at 3:50PST on Nov 1 when he dropped out of the race.

Wave 17: Senate primary question wording changed from, "Someone more critical/supportive of the President" to "Someone more critical/supportive of President Trump."

Wave 37: Small change to the weighting procedure. All Nationscape waves weight to demographic variables, including household income. We offer users the choice to skip answering household income and model those who decline to answer as part of our weighting process. Due to a technical error, no respondents who declined to provide household income were admitted into wave 37 (2020-03-26). As a result, weighting for wave 37 excludes the "Unanswered" category for household income. Weights are otherwise unchanged.

In **Wave 43** (2020-05-07) we changed the stem of extra_return questions: "Some [old: Many] people live in states where many of the following things are restricted due to the spread of coronavirus. If restrictions were lifted on the advice of public health officials to do the following, how likely would you be to:"

In **Wave 46** (2020-05-28) we changed the stem of the extra_covid_wear_mask question from "Have you done any of the following things in response to the spread of coronavirus? - Worn a mask when going out in public" to "Have you done any of the following in the past week? - Worn a mask when going out in public." Because the time reference was changed we changed the variable name to *extra_covid_worn_mask*.

Partisanship: In waves 1–27 respondents who answered “Something Else” to *pid3* did not get a follow-up question asking if they lean closer to one party or the other; beginning in wave 28 (January 23, 2020) these respondents are shown the follow up question. This does not change *pid3* in any way. It does change *pid7* and *pid7_legacy*. See below.

Partisanship: Beginning with the 20200814 release of Phase 1 and new release of Phase 2 data, all party variables (including follow-ups) are included in the release so users may make their own composite seven-category party variable as they deem appropriate.

Partisanship: Beginning with the 20200814 release of Phase 1 and new release of Phase 2 data we have included two composite seven-category party variables that users may want to use depending on their needs. *pid7_legacy* is consistently coded across every wave of the survey from beginning to end. *pid7* exists only after wave 28 when those who answered “Something Else” to *pid3* were asked the follow-up question about leaning toward a party. The disposition for each of these composite variables is as follows:

PID3 Answer	Strong/Weak Rep	Strong/Weak Dem	Lean	pid7
Democrat	(Not Asked)	Strong	(Not Asked)	Strong Democrat
Democrat	(Not Asked)	Weak	(Not Asked)	Weak Democrat
Democrat	(Not Asked)	Skip (missing)	(Not Asked)	Weak Democrat
Republican	Strong	(Not Asked)	(Not Asked)	Strong Republican
Republican	Weak	(Not Asked)	(Not Asked)	Weak Republican
Republican	Skip (missing)	(Not Asked)	(Not Asked)	Weak Republican
Independent	(Not Asked)	(Not Asked)	Republican	Lean Republican
Independent	(Not Asked)	(Not Asked)	Democratic	Lean Democrat
Independent	(Not Asked)	(Not Asked)	Neither	Independent
Independent	(Not Asked)	(Not Asked)	Skip (missing)	Independent
Something Else	(Not Asked)	(Not Asked)	Republican	Lean Republican
Something Else	(Not Asked)	(Not Asked)	Democratic	Lean Democrat
Something Else	(Not Asked)	(Not Asked)	Neither	Independent
Something Else	(Not Asked)	(Not Asked)	Skip (missing)	Skip (missing)
Skip (missing)	(Not Asked)	(Not Asked)	(Not Asked)	Skip (missing)

Primary Party & Vote Choice: In Waves 1–33 all respondents were asked which party’s primary they would vote in (*primary_party*) and a follow-up asking who they would vote for (*dem_vote_intent* and *rep_vote_prim*). In Waves 34 and 35 (2020-03-05 and 2020-03-12) we began sorting respondents by state of residence. Those whose primaries/caucuses had occurred received past-tense questions (*primary_party_past*) and Democratic voters received *democratic_vote_alreadyvoted*. No past primary vote report was asked of Republican voters. Those who were yet to vote received the original questions. We continued to ask Democratic primary voters who they would vote for if the primary in their state were held today, even if they had already voted (*dem_vote_intent*).

Primary Party & Vote Choice: In Wave 36 (2020-03-19) an error in data coming from our vendor resulted in everyone being asked the prospective versions of the vote choice items: *dem_vote_intent* or *rep_vote_prim*, depending on their answer to the party-primary question.

Primary Party & Vote Choice: Beginning in Wave 37 (2020-03-26) the stem of *primary_party* changed to “Will you **or did you** vote in the Democratic or Republican primary or caucus in your state in 2020” and those who selected “The Democratic Primary/Caucus” or “not sure” received *dem_vote_intent* and *dem_vote_past*. Those who selected “The Republican Primary/Caucus” or “not sure” received *rep_vote_intent* and *rep_vote_past*. These questions remain on the survey until wave 47, then they are removed.

Primary Party & Vote Choice: In Wave 47 c we cut *dem_vote_intent*, *rep_vote_intent*, *rep_vote_prim*, *dem_vote_past*, and *rep_vote_past*. These items were replaced by two retrospective primary vote questions: *dem_prim_vote* and *rep_prim_vote*.

General Election Vote: In wave 47 (2020-06-04) we began pilot-testing general election vote questions. We added *vote_2020_v1*.

General Election Vote: In Wave 49 (2020-06-18) we changed the question stem of the 2020 vote question from “In the November 2020 general election for President, will you vote for...” to “If the election for president were going to be held now and the Democratic nominee was Joe Biden and the Republican nominee was Donald Trump, would you vote for...” This version is called *vote_2020* in Wave 49 and forward.

Appendix 2: Weighting Procedure

All targets are derived from the [2018 ACS](#) population over the age of 18, except for 2016 vote.

Weights for gender, census region, Hispanic ethnicity, race, education, age, household language, and country of birth are derived by directly comparing responses in the Nationscape survey with responses to the ACS. Targets appear in table below.

Nationscape respondents may choose not to provide household income. The target for non-response is constructed as the proportion of Nationscape respondents who did not provide an answer. Targets for response categories are based on ACS responses multiplied by the proportion of respondents who chose to answer the Nationscape income question. In effect, this means non-respondents are not weighted for income.

Metropolitan status is derived by crosswalking Nationscape respondents' ZIP codes to Census Bureau Public Use Microdata Areas (PUMAs) using the [2014 MABLE Geocorr](#) based on population percentage. PUMAs are assigned a metropolitan status based on the average status of ACS respondents within them. Groupings are below. Nationscape respondents' PUMA-level metropolitan statuses are weighted to ACS respondents' PUMA-level metropolitan statuses.

Interactions are calculated between several demographic variables. Because some interactions produce very small target populations, some variables are coarsened in the interactions. Interactions and coarsened categories are listed below.

The 2016 vote is derived from the [United States Elections Project](#) and [MIT Election Lab](#). We start with the final vote distribution and estimated turnout. Then, we adjust turnout downwards to account for the 2.7 percent of voters who were not eligible to vote in 2016 but were 18 at the time that our survey was fielded. We then adjust turnout upwards to account for the fact that in the target population we expect 6 percent of voters will be nonvoters in 2016 who lie about voting on our survey. This number comes from the best available estimate.³

Weighting Targets

Age

18–23	24–29	30–39	40–49	50–59	60–69	70+
9.4%	10.9%	17.4%	16.3%	17.0%	15.2%	13.8%

Birthplace

The United States	Another country
82.1%	17.9%

³ Jackman, Simon, and Bradley Spahn. “Why does the American national election study overestimate voter turnout?.” *Political Analysis* 27.2 (2019): 193–207.

Education

No high school diploma	High school diploma	Some college	Associate's degree	Bachelor's degree	Graduate degree
11.6%	27.3%	21.9%	8.4%	19.3%	11.5%

Gender

Female	Male
51.7%	48.3%

Hispanic

Mexican	Other Hispanic	Not Hispanic
9.8%	6.5%	83.7%

Household Income

\$19,999 or less	\$20k-\$34,999	\$35k-\$49,999	\$50k-\$64,999	\$65k-\$79,999	\$80k-\$99,999	\$100k-\$124,999	\$125k-\$199,999	\$200k or more	Unanswered
9.7%	10.6%	10.9%	10.5%	9.2%	10.4%	10.1%	14.6%	9.1%	4.9%

Language

English only	Spanish	Other language
78.1%	13.0%	8.8%

Race

AAPI	Black	White	Other race
6.9%	12.1%	74.1%	7.0%

Region

Midwest	Northeast	South	West
20.8%	17.4%	37.9%	23.8%

Urban

Rural	Suburban	Urban-suburban	Urban
24.1%	24.3%	42.2%	9.4%

Vote 2016

Clinton	Trump	Other vote	No vote
27.7%	26.5%	3.3%	42.5%

Metropolitan status grouping

	Rural	Suburban	Urban-suburban	Pure urban
Individual	0-1	2	3	4
PUMA average	0.0-1.6	1.6-2.4	2.4-3.2	3.2-4.0

Interactions

Interaction	Coarsening
Age × Hispanic	Combine 60–69 and 70+
Age × Race	Combine 60–69 and 70+
Education × Gender	None
Education × Hispanic	Combine Bachelor's degree and Graduate degree Combine Some college and Associate's degree
Education × Race	Combine Bachelor's degree and Graduate degree Combine Some college and Associate's degree
Gender × Race	None
Hispanic × Language	Combine English only and Other language
Hispanic × Race	Combine AAPI, Black, and Other race