Democracy Fund + UCLA Nationscape User Guide

DECEMBER 2021

TABLE OF CONTENTS

Acknowledgements 3

Releases 4

Methodology and Representativeness Assessment 6

Data Collection 7

Data Analysis, Weights, and Variance Estimation 8

Orientation to Data Files 9

Terms of Use 10

Digest Change Log 11

Sample Notes 15

Acknowledgements

Nationscape was funded by 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 staffed Nationscape and executed the fielding of the survey each week. They are: Alex Rossell Hayes, Project Manager; Derek Holliday, Project Coordinator; Aaron Rudkin, Project Developer. Tyler Reny worked as project coordinator from July 2019–August 2020.

Data collection was 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 included: 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.

Releases

This document is a brief guide to the Democracy Fund + UCLA Nationscape Project to accompany the third release of Phase 1 data, the second release of Phase 2 data, and the first release of Phase 3. This release completes the entirety of the Nationscape data release.

Nationscape is an 18-month election study conducted by researchers at UCLA. The project completed roughly 6,250 interviews each week. It started in July 2019 and concluded February 2021.

Phase 1 of the data, released in January 2020, included 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 August 2020, included a re-release of Phase 1 data and new data from January 2020 to July 2020 (Phase 2 data).

Phase 3 of the data, released December 2021, included a re-release of Phase 1 and Phase 2 data and new data from August 2020 to February 2021 (Phase 3 data). This release also contains data from three additional data collections that ran concurrently with weekly surveys on April 2–11, 2020, and July 15–25, 2020; and also one additional effort after the Capitol was attacked, January 21–February 3, 2021. The inclusion of these parallel data collections alongside weekly waves brings the total cases to 494,796.

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

Current Release

This data release, December 31, 2021, includes a re-release of Phase 1 (Phase 1, version 20211231) and a re-release of Phase 2 data (Phase 2, version 20211231). This final 2021 release is the third public release of Nationscape data. The first re-release of Phase 1 in August 2020 (from original version 20200131 to version 20200804) was done to add additional party identification variables such that users could generate their own composite party measures. It also included 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 also appeared due to this change.

This second and final re-release of Phase 1 and first re-release of Phase 2 (now version 20211231) removes 10,000 previously included cases due to suspect data quality. Specifically, we have removed college educated male respondents who completed the survey using the Firefox™ Internet browser. A substantial and unusual increase in the number of respondents with this set of characteristics occurred in Fall 2020. Given this data-quality concern, we have composed the final Nationscape sample excluding

all college educated men who completed the survey using Firefox.[1] In total, 10,358 respondents have been removed from Phases 1 and 2 due to these changes. Details on diagnostics relating to these removals can be found at the end of this document. Changes to the case weights may result from these adjustments to the sample.

ID Numbers

Because Nationscape respondent identification numbers are assigned sequentially and respondents from past releases have been removed from this release, respondent identification numbers from previous releases are not comparable to this release.

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

Citations

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

Tausanovitch, Chris and Lynn Vavreck. 2021. *Democracy Fund + UCLA Nationscape Project, October 10-17, 2019* (version 20211215). Retrieved from [URL].

Citations to this User Guide should be as follows:

Vavreck, Lynn and Chris Tausanovitch. 2021. "Democracy Fund + UCLA Nationscape Project User Guide," Release 3 (version 20211215). Retrieved from [URL].

1 The decision to remove these respondents was made by Nationscape staff. While Lucid offers several protections to ensure that respondents who attempt a survey are non-fraudulent and are unique, Nationscape staff chose a conservative approach to constructing the final sample and removed these respondents even though they were not flagged as fraudulent by staff at Lucid, Inc.

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, Lynn Vavreck, Alex Rossell Hayes, Derek Holliday, Tyler Reny, and Aaron Rudkin. 2021. "Democracy Fund + UCLA Nationscape Methodology and Representativeness Assessment," updated April 1, 2021.

Download the document.[2]

A description of the current weighting methodology is attached to the end of this document.

2 These analyses were performed in March 2021 and used the project's second set of weights from the second release (version 20200814). The final adjustment to the sample and to the weights implemented in the Phase 3 release did not change estimates by more than a few tenths of a point. An earlier version dated 2019 was performed at the start of the project in July 2019 and used weights that have since been updated to allow for more specificity.

Data Collection

Field dates

Nationscape conducts weekly surveys. The first wave went into the field on July 18, 2019. The last consecutive week of data collection went into the field on December 24, 2020. One additional wave was fielded on January 12, 2021, after the U.S. Capitol was attacked. Three concurrent waves were fielded in parallel to regular weekly waves in April and July 2020 and in January 2021. These parallel waves have roughly 10,000 completed interviews per wave. The last date of data collection for Nationscape is February 3, 2021, when the last parallel wave came out of the field.

Weekly surveys are named to reflect the first field date of the wave. Each regular weekly survey was in the field one week.

Parallel waves were in the field for two weeks.

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 13.5 to roughly 17 minutes depending on 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, of those selected to be interviewed, 5.1 percent decline immediately. Another 16.7 percent drop off elsewhere in the survey without completing. We remove 5.9 percent for speeding or straight-lining through the survey. This results in an average yield of 72.4 percent of the original invited sample, depending on wave. We define speeding as completing the survey in fewer than six minutes and straight-lining as selecting the same response for every question in the three policy question batteries.

Using the American Association for Public Opinion Research (AAPOR) standard definitions of case codes and outcome rates for surveys, the average Nationscape AAPOR cooperation rates are as follows:

Cooperation rate 1: 72.4 percent
Cooperation rate 2: 89.1 percent
Cooperation rate 3: 76.9 percent
Cooperation rate 4: 94.6 percent

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.

The three parallel Nationscape waves that were fielded concurrently with weekly surveys in April and July 2020, and separately after Joe Biden's inauguration in January 2021, exclude some questions that are a part of regular weekly Nationscape waves and include many that are not a part of regular waves. Notably, differences exist in the Nationscape issue battery (held constant throughout the regular waves) to account for changing national context. Methodology and weighting are unchanged. Variables that exist in these waves are directly comparable to those measured in the regular waves.

Weights

We provide weights in the "weight" variable that can be used for a national sample. Weights are constructed using self-reported 2016 vote. After the 2020 presidential election, we provide "weight_2020," which replaces self-reported 2016 vote with self-reported 2020 vote in the weight calculation. For waves after the 2020 election, we also include "weight_both," which uses both 2016 and 2020 vote.

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" (Representativeness Assessment linked earlier in this document; weighting description attached at the end) 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 nor is it constructed using Lucid's online tool for sampling. Nationscape staff designed the selection criteria and targets after several weeks of pilot waves in 2019 (see the Methodology Assessment linked above) and managed the sample on a daily basis in cooperation with staff at Lucid.

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).[3]

---

3 Rivers, Douglas, and Delia Bailey. "Inference from matched samples in the 2008 U.S. 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 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.[4] 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.

4 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.

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. 2021. *Democracy Fund + UCLA Nationscape Project, October 10–17, 2019* (version 20211215). Retrieved from [URL].

3. 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.

Digest 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.

Wave 3: Launched one day late.

Wave 5: Began asking Gubernatorial and Senate general election ballot items only to respondents who had 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 3:50 PST on November 1, 2020, 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 (March 26, 2020). As a result, weighting for wave 37 excludes the "Unanswered" category for household income. Weights are otherwise unchanged.

Wave 43: On May 7, 2020, 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:"

Wave 46: On May 28, 2020, 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 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)

PID3 Answer	Strong/Weak Rep	Strong/Weak Dem	Lean	pid7_legacy
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)	(Not Asked)	888 (missing)
Skip (missing)	(Not Asked)	(Not Asked)	(Not Asked)	Skip (missing)

Primary Party and 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 (March 5, 2020 and March 12, 2020), 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 and Vote Choice: In wave 36 (March 19, 2020) 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 and Vote Choice: Beginning in wave 37 (March 26, 2020) 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 remained on the survey until wave 47, then they are removed.

Primary Party and 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: IIn wave 47 (June 4, 2020) we began pilot-testing general election vote questions. We added *vote_2020_v1*.

General Election Vote: In wave 49 (June 18, 2020) 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…" The new version is called *vote_2020* in wave 49 and forward. The old version is *vote_2020_v1*.

Wave 60: All primary election questions from September 3, 2020, forward are exclusively in the past tense.

We added "Yes, I already voted" as an option for general election vote intention.

The "Warren vs. Pence" hypothetical presidential match-up was mistakenly asked of 55 respondents. These 55 respondents are coded as missing for Biden v. Pence or Harris v. Pence matchups.

Wave 62: On September 17, 20202, we changed *elect_conf_conduct* from a grid to a stand-alone, multiple-choice question. Response categories remained the same.

Wave 68: On October 29, 2020, we changed (and re-inserted) *elect_conf_vote* from a grid to a stand-alone, multiple-choice question. Response categories remained the same.

Wave 69: On November 5, 2020, we added "I did not vote in this race in 2020" as an option for House, Senate, and Governor vote intent questions.

We added "Received my ballot through the mail and returned it to an official ballot drop box" as an option for *extra_vote_mail_retr*.

Wave 70: On November 12, 2020, we split *switchers_text* into two variables that indicated whether respondents switched their votes to Biden (*_biden*) or Trump (*_trump*).

Wave 77: The original field period for Nationscape was through the end of 2020. This unanticipated wave was added on January 12, 2021, after the insurrection events at the U.S. Capitol on January 6, 2021.

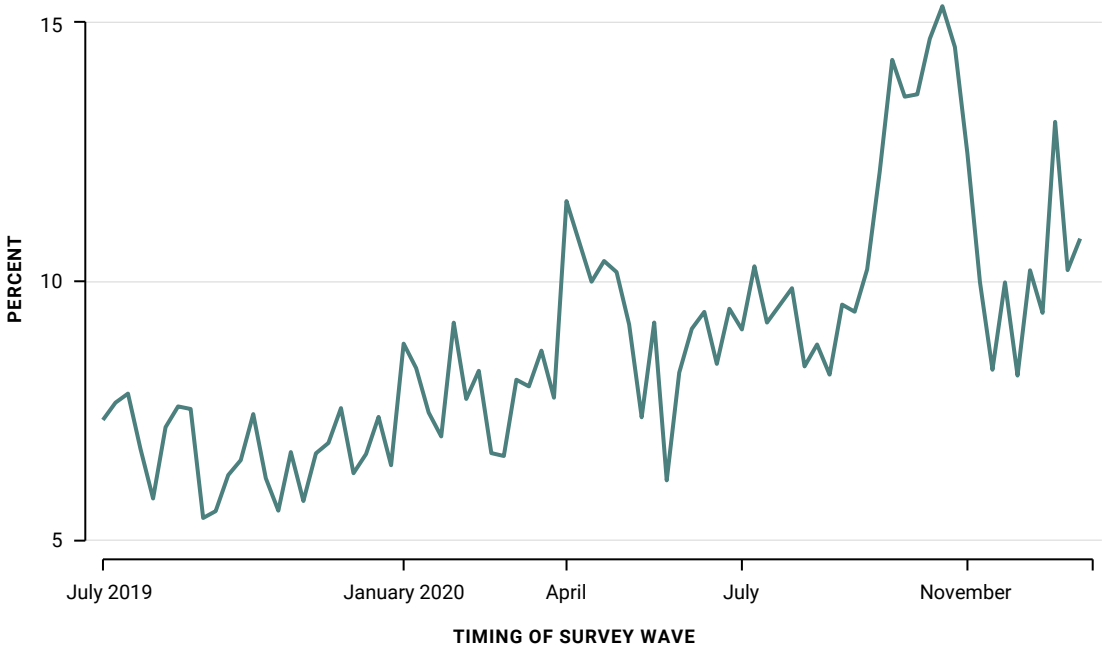
Sample Notes

Nationscape diagnostics revealed two anomalies as we prepared the Phase 3 release of Nationscape.[5] The first was related to how people were completing the survey (browser use). The second was related to demographics, specifically age and its intersection with education and gender. We describe each anomaly in turn.

Anomaly 1: Browser Use

Use of the Firefox Internet browser (on any device) trended upwards beginning in late 2019 and early 2020. However, this gradual trend was punctuated by substantial transient spikes in the share of respondents completing the survey using the Firefox browser occurring around April 2020, after Joe Biden secured the nomination, and during the general election campaign (September 2020 through early November 2020). The Fall 2020 spike in increased incidence of Firefox use subsided immediately after the general election (see *Firefox Browser Use by Week, 2019–2020* below).

Firefox Browser Use by Week, 2019-2020



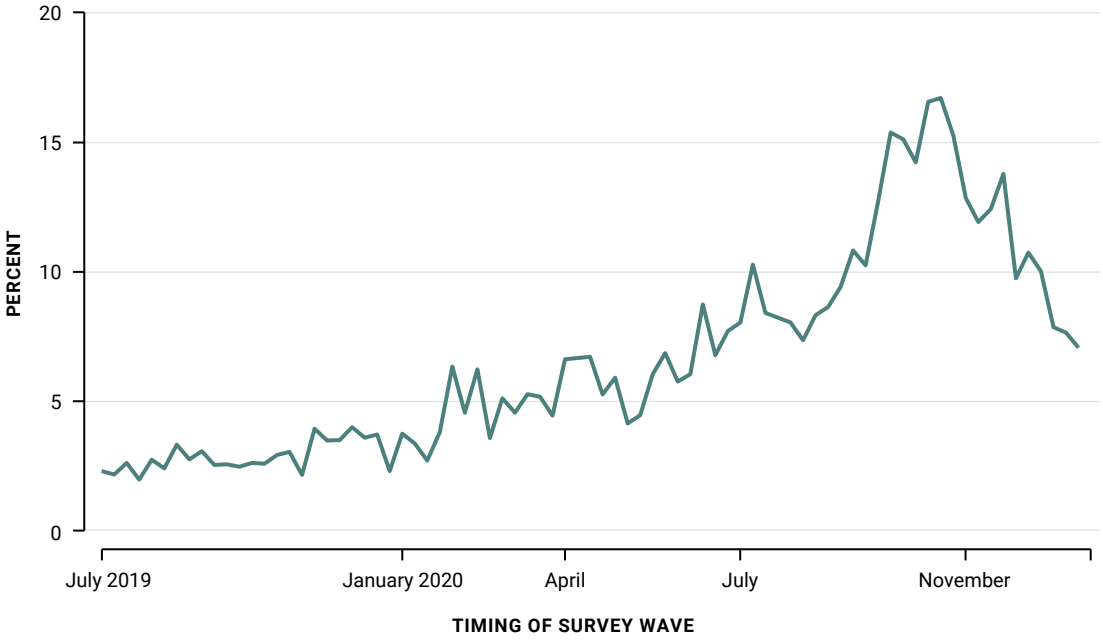
Data include all waves of Nationscape from July 2019 to January 2021.

5 Several people provided assistance as we investigated and thought about these irregularities. In addition to the Nationscape project staff, we thank Mark S. Handcock and Andrew Shapiro (Department of Statistics, UCLA), Jeffrey B. Lewis (Department of Political Science, UCLA), and Brian T. Hamel (Department of Political Science, Louisiana State University).

Anomaly 2: Age, Education, Gender

We also noted that the share of respondents who reported being exactly 40 years old were significantly over-represented relative to their target in the population. This difference was driven by men – and particularly by men who reported having a college education. Like the observed increase in Firefox use, the incidence and growth in college educated, male, 40-year-olds was largest during the general election period and declined quickly after Election Day (see *Share of College Educated Men who are 40, 2019–2020* below).

Share of College Educated Men who are 40, 2019-2020



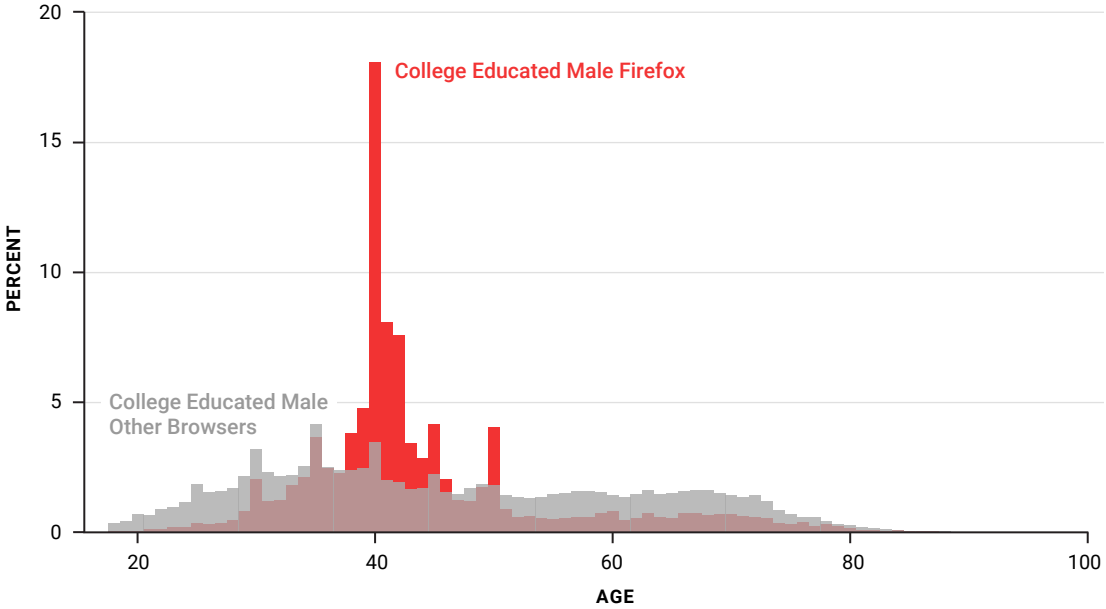
Data include all waves of Nationscape from July 2019 to January 2021.

Adjustment

Based on these trends, we investigated whether college educated men who completed the survey using Firefox were driving the irregularities with respect to the age distribution.

By examining the distribution of age among college educated men using Firefox compared to the distribution of age among college educated men using other browsers, we found that among these Firefox users, 40-year-olds were significantly over-represented (18 percent); the share among college educated, male non-Firefox users was roughly 3 percent (see *Distribution of Age* below).

Distribution of Age



Data include all waves of Nationscape from July 2019 to January 2021.

Due to the significant irregularities among this set of respondents (college educated men who completed the survey using Firefox), we composed the final Nationscape sample without respondents who completed the survey on Firefox and reported being college educated men.