# Methodology

## Congressional social media data

This analysis examines a complete set of Facebook posts and tweets created on any account managed by any member of the U.S. Senate and House of Representatives between Jan. 1, 2015, and July 6, 2020. Researchers used the Facebook Graph API, CrowdTangle[1] API and Twitter API to download the posts. The resulting dataset contains more than 1.5 million Facebook posts from 712 different members of Congress who used a total of 1,391 Facebook accounts, and over 3.4 million tweets from 711 different members of Congress who used a total of 1,363 Twitter accounts. The broader data collection process is described here.

This analysis includes all text from these Facebook and Twitter posts, including image captions and emojis. Photo and video posts were not included in this analysis unless the post also contained meaningful text, such as a caption. Text that appeared only within images was not included in the analysis. Posts by nonvoting representatives were also excluded. In addition, for Rep. Jeff Van Drew of New Jersey (who switched from the Democratic to the Republican Party), researchers grouped social media posts according to the member's party affiliation at the time of each post or tweet.

### Identifying posts mentioning Latinx, Latino, Latina and Hispanic

To identify posts that included the terms Latinx, Latino, Latina and Hispanic, researchers used a case-insensitive regular expression – a pattern of keywords and text formatting – that consisted of the terms "Latinx," "Latino," "Latina" and "Hispanic." This pattern identified 24,793 posts on Twitter or Facebook that included any of these terms.

### Racial and ethnic group labels for members of Congress

To assign race and ethnicity for each member of the 116th Congress, researchers used lists of the Black, Asian and Pacific Islander and Hispanic members of Congress that are maintained as part of a collaborative project between the Office of the Historian and the Clerk of the House's Office of Art and Archives. In addition, according to the Congressional Research Service, there are four Native American members in the 116th Congress. Researchers used the Biographical Directory of the United States Congress to confirm the identity of these four members.

[1] CrowdTangle is a public insights tool owned by Facebook.

## Survey and Twitter data

The analysis of Twitter users in this report is based on two surveys of U.S. adults ages 18 years and older. The first was a survey of Twitter users conducted from Nov. 21 to Dec. 17, 2018, by Ipsos in English using KnowledgePanel (KP). The second was a nationally representative survey of U.S. adults (including both Twitter users and non-users) conducted from Oct. 29 to Nov. 11, 2019, on Pew Research Center's American Trends Panel (ATP). This report focuses on 2,791 respondents from the KnowledgePanel survey and 1,345 respondents from the ATP survey who said they use Twitter, agreed to allow researchers to examine their Twitter behaviors, and provided a valid Twitter handle. The margin of sampling error for the combined sample of 4,136 Twitter users is plus or minus 2.4 percentage points.

The two surveys had different designs and eligibility criteria. For the KP survey, a total of 7,850 panelists who had previously indicated that they have a Twitter account were invited to participate. Of the 4,829 who responded, 3,649 (76%) confirmed that they used Twitter. Among confirmed Twitter users, 3,293 (90%) agreed to provide their Twitter handle. Next, researchers reviewed each account and removed any that were nonexistent or belonged to institutions, products or international entities. In the end, there were 2,791 respondents who both completed the survey and provided a valid handle (76% of confirmed Twitter users).

KnowledgePanel members are recruited through probability sampling methods and include those with internet access and those who did not have internet access at the time of their recruitment (KnowledgePanel provides internet access for those who do not have it, and if needed, a device to access the internet when they join the panel). A combination of random-digit dialing (RDD) and address-based sampling (ABS) methodologies have been used to recruit panel members. (In 2009 KnowledgePanel switched its sampling methodology for recruiting members from RDD to ABS.) KnowledgePanel continually recruits new panel members throughout the year to offset panel attrition.

The ATP sample is comprised of adults who responded to ATP Wave 57. A total of 12,043 panelists responded out of 14,412 who were sampled, for a response rate of 84%.[2] On this survey, a total of 2,561 (21%) stated that they used Twitter. Of these Twitter users, 1,517 (59%) agreed to provide their Twitter handle. Researchers reviewed each account in a similar manner as for the KnowledgePanel respondents. This yielded 1,345 Twitter users who agreed to participate in the study and provided a valid handle (53% of confirmed Twitter users).

[2] See the survey methodology for additional details on ATP Wave 57.

The American Trends Panel (ATP), created by Pew Research Center, is a nationally representative panel of randomly selected U.S. adults. Panelists participate via self-administered web surveys. Panelists who do not have internet access at home are provided with a tablet and wireless internet connection. Interviews are conducted in both English and Spanish. The panel is managed by Ipsos.

The ATP was created in 2014, with the first cohort of panelists invited to join the panel at the end of a large, national, landline and cellphone random-digit-dial survey that was conducted in both English and Spanish. Two additional recruitments were conducted using the same method in 2015 and 2017, respectively. Across these three surveys, a total of 19,718 adults were invited to join the ATP, of which 9,942 agreed to participate.

**American Trends Panel recruitment surveys**

| Recruitment dates | Mode | Invited | Joined | Active panelists remaining |
|---|---|---|---|---|
| Jan. 23 to March 16, 2014 | Landline/ cell RDD | 9,809 | 5,338 | 2,303 |
| Aug. 27 to Oct. 4, 2015 | Landline/ cell RDD | 6,004 | 2,976 | 1,335 |
| April 25 to June 4, 2017 | Landline/ cell RDD | 3,905 | 1,628 | 684 |
| Aug. 8 to Oct. 31, 2018 | ABS/web | 9,396 | 8,778 | 6.403 |
| Aug. 19 to Nov. 30, 2019 | ABS/web | 5,900 | 4,720 | 4.681 |
| | **Total** | **35,014** | **23,440** | **15,406** |

Note: Approximately once per year, panelists who have not participated in multiple consecutive waves or who did not complete an annual profiling survey are removed from the panel. Panelists also become inactive if they ask to be removed from the panel.

PEW RESEARCH CENTER

In August 2018, the ATP switched from telephone to address-based recruitment. Invitations were sent to a random, address-based sample (ABS) of households selected from the U.S. Postal Service's Delivery Sequence File. In each household, the adult with the next birthday was asked to go online to complete a survey, at the end of which they were invited to join the panel. For a random half-sample of invitations, households without internet access were instructed to return a postcard. These households were contacted by telephone and sent a tablet if they agreed to participate. A total of 9,396 were invited to join the panel, and 8,778 agreed to join the panel and completed an initial profile survey. The same recruitment procedure was carried out on August 19, 2019, from which a total of 5,900 were invited to join the panel and 4,720 agreed to join the panel and completed an initial profile survey. Of the 23,440 individuals who have ever joined the ATP, 15,406 remained active panelists and continued to receive survey invitations as of August 2020.

The U.S. Postal Service's Delivery Sequence File has been estimated to cover as much as 98% of the population, although some studies suggest that the coverage could be in the low 90% range.[3]

[3] AAPOR Task Force on Address-based Sampling. 2016. "AAPOR Report: Address-based Sampling."

The American Trends Panel never uses breakout routers or chains that direct respondents to additional surveys.

## Weighting

The final sample of 4,136 Twitter users with valid handles was weighted in a multi-stage process. Both the KP and ATP samples begin with a base weight that reflects each respondent's probability of selection into the sample. In the first step, a propensity adjustment was made to the ATP sample's base weight so that the adjusted ATP sample more closely resembled the KP sample on gender, age, race, Hispanic ethnicity, education, region, party affiliation, volunteerism, voter registration, and metropolitan area. In addition to demographics, the propensity adjustment also balanced on whether panelists tweeted at least once between Oct. 28, 2019, and April 28, 2020, the number of tweets in that six month period, the number of followers, the number of accounts followed, and the number of tweets favorited per day. This was done to adjust for differences between the two samples in terms of the frequency of Twitter use that are likely attributable to differences in the design, timing and incentivization of the two surveys.

Next, the samples were combined and the weights were further adjusted using an iterative technique to match a set of benchmarks for gender, age, race, Hispanic ethnicity, education, region, party affiliation, volunteerism, voter registration, and metropolitan area. Because there are no official benchmarks for the population of Twitter users, population parameters were estimated using the set of respondents to ATP Wave 57 who indicated that they use Twitter (regardless of whether they agreed to provide their Twitter handle). For details on how the full Wave 57 sample was weighted see the survey methodology.

Sampling errors and statistical tests of significance take into account the effect of weighting at each of these stages. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls.

## Identifying posts mentioning Latinx, Latino, Latina and Hispanic

To identify posts that included the terms Latinx, Latino, Latina and Hispanic, researchers used a case-insensitive regular expression – a pattern of keywords and text formatting – that consisted of the terms "Latinx," "Latino," "Latina" and "Hispanic." This pattern identified 2,396 that included any of these terms.