Understanding Who Uses Twitter

State level estimates of those on Twitter

MOSAIC Methods Brief

AUTHORS
Mickey Jackson, SSRS
James McKinstry, SSRS
Cameron McPhee, SSRS
Trivellore Raghunathan, University of Michigan
Lisa Singh, Georgetown University
Michael Traugott, University of Michigan
Chintan Turakhia, SSRS
Nathan Wycoff, Georgetown University

measuring online social attitudes and information collaborative

SSRS | GEORGETOWN UNIVERSITY | UNIVERSITY OF MICHIGAN

MOSAIC Data Brief Series

MOSAIC (Measuring Online Social Attitudes and Information Collaborative) is a collaboration between SSRS, Georgetown University, and the University of Michigan. This collaboration will focus on understanding how to leverage survey data and social media data to better capture public opinion in reliable, valid, and scientifically rigorous ways. This methodology brief series is intended to share methodological results and data that researchers can use to better understand the population on social media. Methodology briefs will utilize traditional survey responses, open-ended survey responses, and/or social media posts in order to understand the strengths and limitations of each method for measuring public attitudes.

Survey research has long been a bedrock for understanding society and its needs. Social media data collected from, for example, Twitter users is providing new ways to perform social science research and, for some research questions, may be a viable supplement to or replacement for survey data. However, to understand what can and cannot be inferred from Twitter data, it is critical to understand who is using Twitter. Toward that end, two questions of interest include: "**How different are Twitter users from the general population**?" and "**Do Twitter use rates vary by states and region?**" Answering both of these questions is important to justify the use of Twitter data for generalizable social science research.

To date, estimates of the prevalence and characteristics of Twitter users have been provided only at the national level. To provide a richer picture of this population, the MOSAIC project has been investigating methods for producing estimates of Twitter use at the state level. This methodology brief provides preliminary state-level estimates and information about how they are produced.

Specifically, in this methodology brief, we apply a model-based small area estimation approach to a nationally representative survey to estimate the prevalence and characteristics of Twitter users at the state level.[1] Researchers may find these estimates useful as benchmarks when analyzing Twitter data for state-level public opinion research, or as explanatory variables in analyzing cross-state variation in other attitudes and behaviors.

This methodology brief highlights broad patterns in the estimates. The full set of estimates can be downloaded in tabular form at the MDI Data Portal.[2]

How Does Twitter Usage Vary by State?

We asked a nationally representative sample of 9,544 adults ages 18 or older whether they have a personal account on any of several social media sites. Data collection was conducted via Web from March 11 – June 13, 2021.

Based on this survey, we estimate that approximately 28% of U.S. adults have a Twitter account. However, there is substantial variation across states (See Figure 1 for a summary and the Appendix for more details). Our state-level estimates range from 20% (Wyoming) to 34% (Maryland), with the District of Columbia as an outlier at 47%. Geographically, most states with

---

[1] Here, we define a "Twitter user" as anyone who has an account on Twitter, regardless of whether or how often they use it.

[2] https://portals.mdi.georgetown.edu/public/mosaic/MOSAIC-Twitter-State-Estimates

relatively high Twitter usage are located in the Northeast and Mid-Atlantic regions, while states in the South and Mountain West have lower Twitter usage rates.
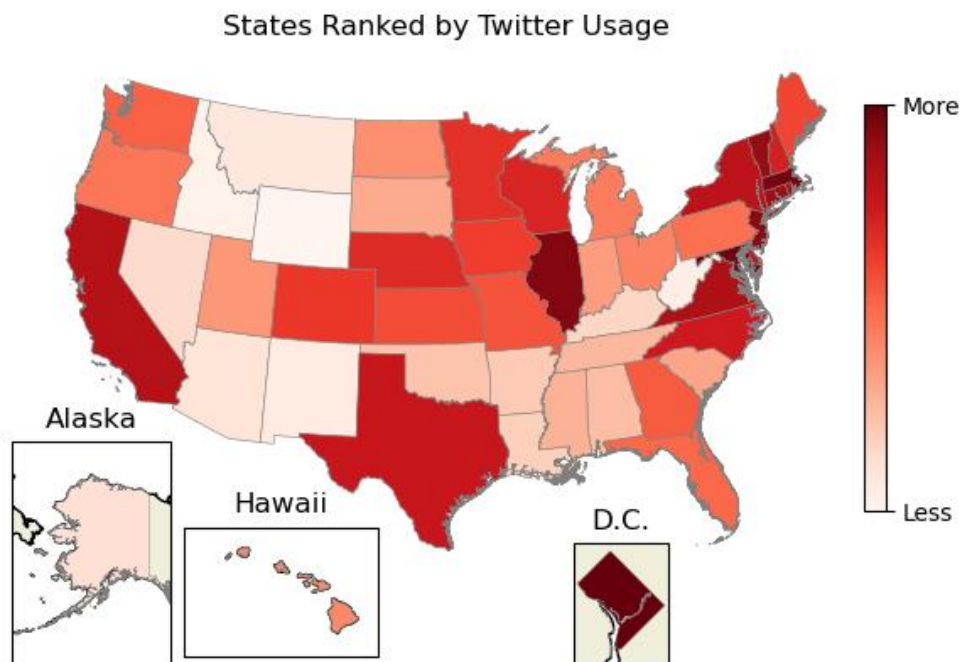
States Ranked by Twitter Usage

Figure 1. Estimated Rank of Percent of Adults that have a Twitter Account by State

How Are Twitter Users Different from Their States' General Populations?

Relative to their states' general populations, Twitter users are substantially less likely to have a low level of education (high school education or below) and substantially more likely to be under the age of 35 (see Figure 2). In comparison, differences by race/ethnicity and gender are relatively small. These state-level patterns mirror the Pew Research Center's national-level findings.[3] Figure 3 shows the six states having the largest differences in education between Twitter users and the general population. Figure 4 shows the six states having the largest differences in age between Twitter users and the general population. In both cases, the state with the largest difference is West Virginia.[4]

---

[3] Wojcik, S., & Hughes, A. (2019). Sizing up Twitter users. *Pew Research Center*, *24*. https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

[4] We used Kullback-Leibler Divergence in order to determine the difference in demographic categories between the population at large and Twitter users in each state. KL Divergence calculates a score that measures the difference between two distributions. It more effectively captures differences in minority groups than directly comparing percentage points.
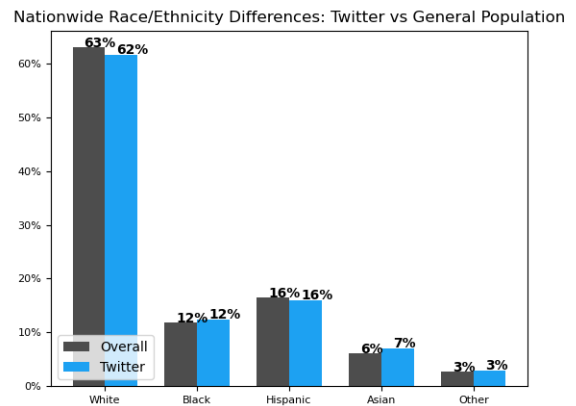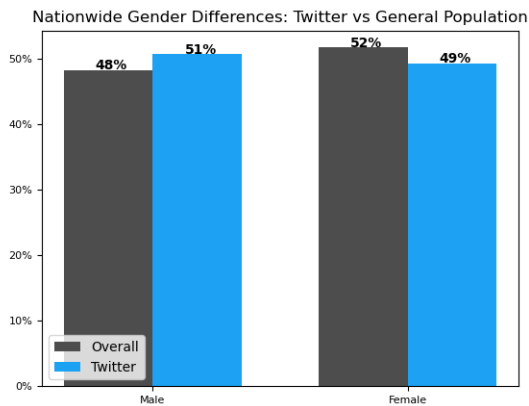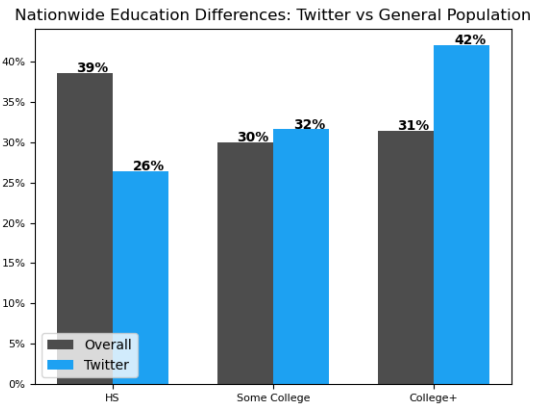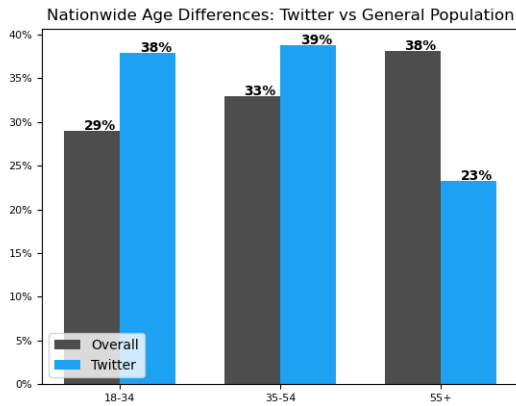
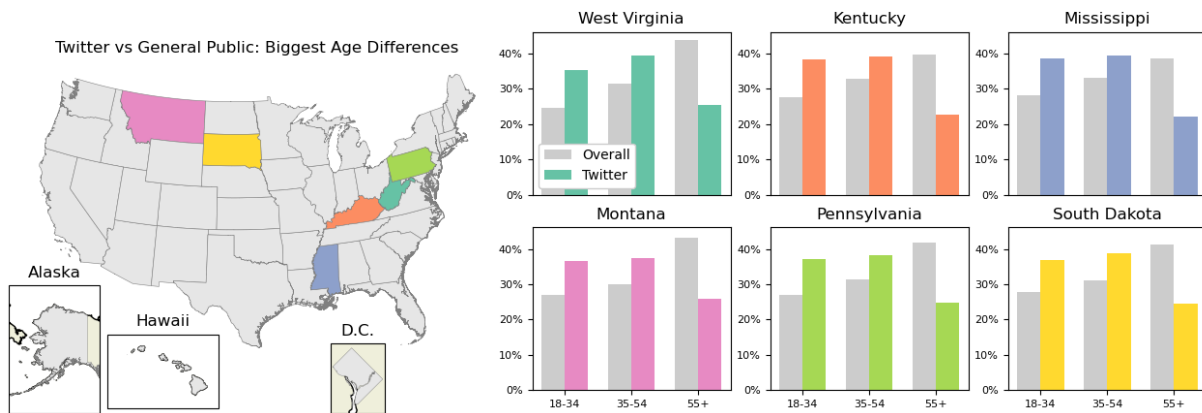Figure 2: Nationwide Age, Education, Gender and Race/Ethnicity Differences

Figure 3. Largest State-Level Age Differences between Twitter Users and the General Public
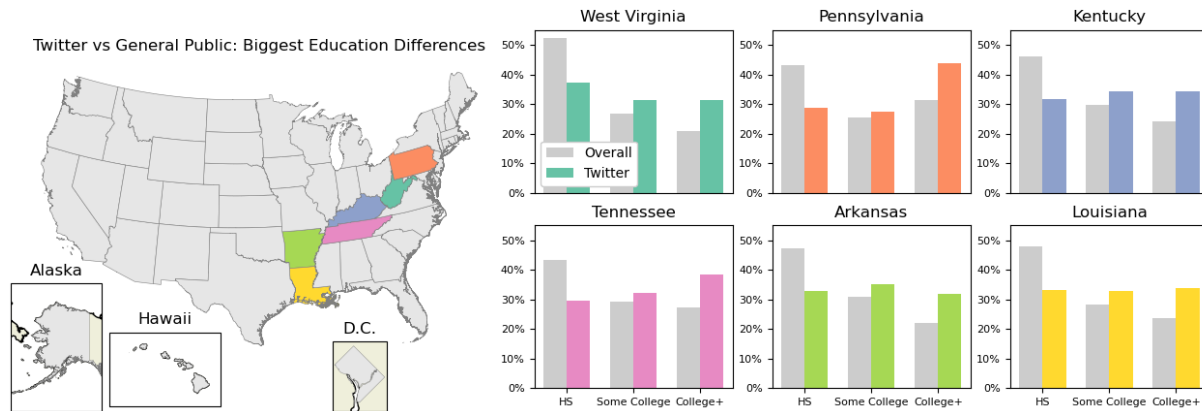
Figure 4. Largest State-Level Education Differences between Twitter Users and the General Public

How Did We Determine These Numbers?

Data are from the MOSAIC recruitment survey, which was conducted via the SSRS Opinion Panel of U.S. adults aged 18 and older. The SSRS Opinion Panel is a probability-based web panel of U.S. adults (including Hawaii and Alaska) and is recruited randomly based on a nationally representative ABS (Address Based Sample) probability design. Data collection was conducted via the web from March 11 – June 13, 2021, with 9,544 panelists completing the survey (9,468 in English and 76 in Spanish). Because data collection was conducted entirely via the web, the sample was limited to adults who use the Internet.

The state-level estimates are produced using a model-based small area estimation procedure called Multilevel Regression and Poststratification (MRP).[5] Due to its flexibility and relative simplicity, MRP is an increasingly popular technique for producing estimates within small geographic areas from moderately sized national samples.

Estimating Twitter use by MRP involved three main steps. First, we estimated a *multilevel logistic regression* on the survey data, where the dependent variable was an indicator of whether the respondent had a personal Twitter account. This model included two types of predictors:

- **Individual-level predictors** (state, race/ethnicity, educational attainment, gender, age, and interactions between gender and age, educational attainment and age, gender and

---

[5]Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*, *12*(4), 375-385. https://www.cambridge.org/core/journals/political-analysis/article/abs/bayesian-multilevel-estimation-with-poststratification-statelevel-estimates-from-national-polls/22A5EF78D027E76C782B3280D400FCC9

educational attainment, state and educational attainment, and state and age), modeled as random effects. Random-effects modeling implies that the state-level estimates become a "compromise" between a modeled estimate (based on relationships between demographics and Twitter use) and the raw state-level survey data, with the survey data carrying progressively more weight as the state-level sample size increases. This "partial pooling" allows for more precise estimates (at the risk of greater bias) for small states.

- **State-level predictors** (Census region, percent of adults using any social media[6], and percent of the 2020 presidential vote received by Donald Trump), modeled as fixed effects. These reduce unexplained cross-state variation and thus further improve the precision of state-level estimates. Region was included to account for the possibility of systematic regional differences in Twitter use. The percent of adults using any social media is intuitively expected to be related to state-level Twitter use. Finally, the 2020 Trump vote share was included as a proxy for state-level partisanship, as preliminary analysis of the survey data found that party identification was a strong predictor of Twitter use.

After estimating the model, the second step was to use it to *predict the percentage of adults with a Twitter account* in every poststratum. Poststrata are cells representing every possible combination of the individual-level predictors in the model; for example, Hispanic males in California ages 18 – 34 with some college education would be a single poststratum.

The third and final step was to estimate the state-level prevalence of Twitter users by taking *population-weighted averages* of the model-based poststratum-level estimates. We obtained an estimated population for each poststratum from the 2019 American Community Survey (ACS). We then multiplied each poststratum's ACS-estimated population by its model-estimated prevalence of Twitter users, summed the resulting estimates to the state level, and then divided the resulting sums by each state's population. This step can easily be adjusted to obtain estimates within even narrower groups (e.g., age groups within states).

One complication was the handling of Internet access, which is clearly related to Twitter use and therefore, in principle, should have been included in the MRP model and poststratification. However, Internet access could not be included in the model because the survey was conducted entirely by Web, which implies that all respondents had some form of Internet access.[7] Furthermore, the ACS does not ask directly about individual-level Internet access; it

---

[6] State-level social media usage estimates were obtained from: United States Census Bureau. (2021). Current Population Survey's 2019 Computer and Internet Use Supplement. Retrieved from https://www.census.gov/data/datasets/time-series/demo/cps/cps-supp_cps-repwgt/cps-computer.html.

[7] In analyses of MOSAIC survey data that rely on direct estimation rather than MRP, the exclusion of non-Internet-users is accounted for using a specialized weighting adjustment.

only asks whether *any household member* has access to the Internet *at home*. If not adjusted for, the exclusion of non-Internet-users would cause an overestimation of Twitter usage at both the state and national levels. To prevent this, we did include Internet access (using the household-level ACS item) in the poststratification table; but because Internet access could not be included in the model, we imputed a Twitter prevalence of 0% for all no-Internet-access cells. That is, we assumed that anyone who lacks Internet access at home is not a Twitter user. This assumption may cause us to slightly underestimate Twitter usage, since some persons without home Internet access may have Internet access elsewhere and thus could have a Twitter account. The need for this simplifying assumption is a limitation of the Web-only survey data on which this analysis was based.

Acknowledgments

Suggested Citation

Jackson, M. McKinstry, J., McPhee, C., Raghunathan, T., Singh, L., Traugott, M., Turakhia, C. & Wycoff, N. (2021*). Understanding who uses Twitter: State level estimates of those on Twitter*. MOSAIC Methods Brief: November 2021. *Measuring Online Social Attitudes and Information Collaborative.*

State Estimate Dataset

This dataset is available through the MDI portals:
https://portals.mdi.georgetown.edu/public/mosaic/MOSAIC-Twitter-State-Estimates
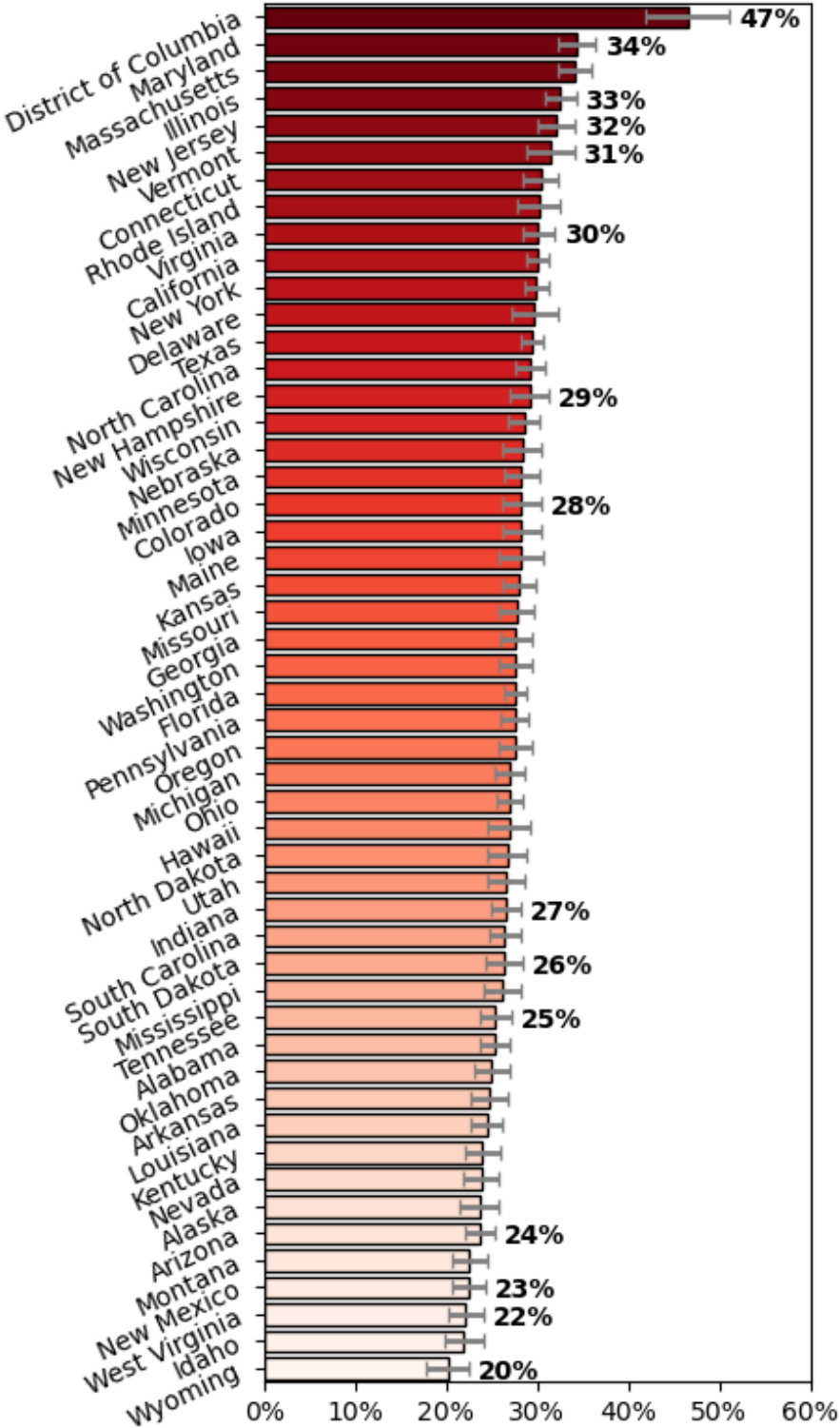
Figure A-1. Estimated Percent of Adults that have a Twitter Account by State