

# A Note on Internet Use and the 2016 Election Outcome

Levi Boxell, *Stanford University*\*

Matthew Gentzkow, *Stanford University and NBER*

Jesse M. Shapiro, *Brown University and NBER*

September 2017

## Abstract

We use data from the American National Election Studies from 1996 to 2016 to study the role of the internet in the 2016 election outcome. Following the methodology in Boxell et al.'s (Forthcoming) study of political polarization, we compare trends in the Republican share of the vote between likely and unlikely internet users, and between actual internet users and non-users. Relative to prior years, the Republican share of the vote in 2016 was as high or higher among the groups least active online.

---

\*E-mail: lboxell@stanford.edu, gentzkow@stanford.edu, jesse\_shapiro\_1@brown.edu. We acknowledge funding by the Brown University Population Studies and Training Center, which receives support from NIH Grant P2C HD041020; Toulouse Network for Information Technology; the Stanford Institute for Economic Policy Research; the National Science Foundation (grant number: DGE-1656518); and the Institute for Humane Studies. The American National Election Studies and the relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.

# 1 Introduction

Many have hypothesized that the internet and social media impacted the outcome of the 2016 election. In a post-election interview, Hillary Clinton emphasized the role of social media in the election, citing fake news, Russian intervention, and Republicans' success in "marrying content with delivery and data" (Johnson 2017). Others have emphasized the Trump campaign's use of data to target messages online (Confessore and Hakim 2017).

There have been several attempts at examining these claims about the 2016 election empirically. Hampton and Hargittai (2016) argue that the internet is unlikely to have helped Trump because only a small percentage of Trump supporters use social media and because Trump did unusually well among the demographic groups least likely to use the internet. Allcott and Gentzkow (2017) show that while fake news was dominantly pro-Trump, it would have to be extraordinarily persuasive relative to other media technologies (e.g., TV ads) in order for it to have swayed the election.

We use data from the American National Election Studies (ANES) from 1996 to 2016 to study the role of the internet in the 2016 election outcome. Following the methodology in Boxell et al.'s (Forthcoming) study of political polarization, we compare trends in the Republican share of the vote between likely and unlikely internet users, and between actual internet users and non-users. Relative to prior years, the Republican share of the vote in 2016 was as high or higher among the groups least active online.

## 2 Data

We use data from the ANES (2015a, 2015b, 2016, 2017), which is a nationally representative survey that asks various demographic and political questions. We use the ANES 1948–2012 Time Series Cumulative, 2008 Time Series, 2012 Times Series, and 2016 Time Series datasets. We use data from survey waves in presidential election years from 1996–2016, inclusive, and we restrict attention to face-to-face surveys, excluding internet-based surveys that were conducted in more recent years. Our calculations weight responses from 1996–2012 by the type-0, face-to-face survey weights<sup>1</sup> and responses from 2016 by the post-election, face-to-face survey weights.<sup>2</sup>

---

<sup>1</sup>See *VCF0009x* in the 1948–2012 Times Series Cumulative file.

<sup>2</sup>See *V160102f* in the 2016 Time Series Study.

Our outcome variable is the party that the respondent voted for in the most recent presidential election. We construct this variable from responses to “How about the election for President? Did you vote for a candidate for President? (IF YES:) Who did you vote for?” which are then coded as either Republican, Democratic, Other, or refusals for respondents who said they voted for a presidential candidate.<sup>3</sup> Respondents who report not voting for a presidential candidate or who refuse to say who they voted for are excluded from our analysis.

We use three different measures of internet use. Our first measure, which we refer to as whether or not a respondent uses the internet, comes from responses to “Do you have access to the Internet or the World Wide Web [exc. 2008: (‘the Web’)]?” for 1996–2008<sup>4</sup> and “Do you or anyone in this household use the Internet at any location?” for 2012–2016.<sup>5</sup> Our second measure, which we refer to as whether or not a respondent observed campaign news online, comes from responses to “Have you seen any information about this election campaign on (the Internet/the Web)?” for 1996–2004,<sup>6</sup> “Did you read, watch, or listen to any information about the campaign for President on the Internet?” for 2008–2012,<sup>7</sup> and whether respondents “heard anything about the presidential campaign” on “Internet sites, chat rooms, or blogs” for 2016.<sup>8</sup> Our third measure, which we refer to as predicted internet access, is taken from Boxell et al. (Forthcoming)<sup>9</sup> and classifies respondents according to whether the respondent is in the top or bottom quartile in terms of the likelihood of having internet if they were a respondent in 1996, as predicted from the following covariates: age group, gender, race, education, and whether the respondent lives in the political south. Separately for each measure of internet use, we exclude respondents with missing or non-valid responses (as defined by the ANES) to the questions needed to construct the measure.<sup>10</sup>

### 3 Results

The figure shows, for each of our three measures of internet use, the proportion of voting respondents who voted for the Republican candidate in each presidential election. All three plots

---

<sup>3</sup>See *VCF0705* in the 1948–2012 Time Series Cumulative file and *V162058x* in the 2016 Time Series Study.

<sup>4</sup>See *VCF0744* in the 1948–2012 Time Series Cumulative file.

<sup>5</sup>See *prmedia\_useinet* in the 2012 Time Series Study and *V161007* in the 2016 Time Series Study.

<sup>6</sup>See *VCF0745* in the 1948–2012 Time Series Cumulative file.

<sup>7</sup>See *V085010* and *V085021* in the 2008 Time Series Study and *mediapo.net* in the 2012 Time Series Study.

<sup>8</sup>See *V161363d* in the 2016 Time Series Study.

<sup>9</sup>See SI Figure 12 and references therein for more details.

<sup>10</sup>For the predicted internet measure in 2016, we also drop respondents whose response for education is in the “95. Other SPECIFY” category.

show that, if anything, Trump outperformed relative to trend among those groups that are least active online. For two of the three measures, the 2016 election marked the first time since 1996 that the Republican candidate performed equal or better among the group that is less active online. It is important to note that the composition of internet users is changing over time. Therefore, trends in, say, the Republican share among actual internet users reflect both changes in which respondents use the internet as well as how respondents voted. Our measure of predicted internet use is constructed from a time-invariant function of covariates and is therefore less subject to this caveat.

The table shows, for each of our three measures of internet use, the change in the proportion of voting respondents who voted for the Republican candidate between 2012 and 2016, separately for more and less internet-active groups. The table also shows the difference in change in proportions between more and less internet-active groups. We report a 95 percent confidence interval on the change in proportions, and on the difference in change in proportions, based on a nonparametric bootstrap with 100 replicates.<sup>11</sup> We find that, compared to Romney, Trump performed relatively better among less internet-active groups, though we note that the confidence intervals are wide and always include 0.

Overall, our findings do not support the hypothesis that internet media and online campaign methods conferred an advantage to Trump compared to other Republican presidential candidates in the internet era.

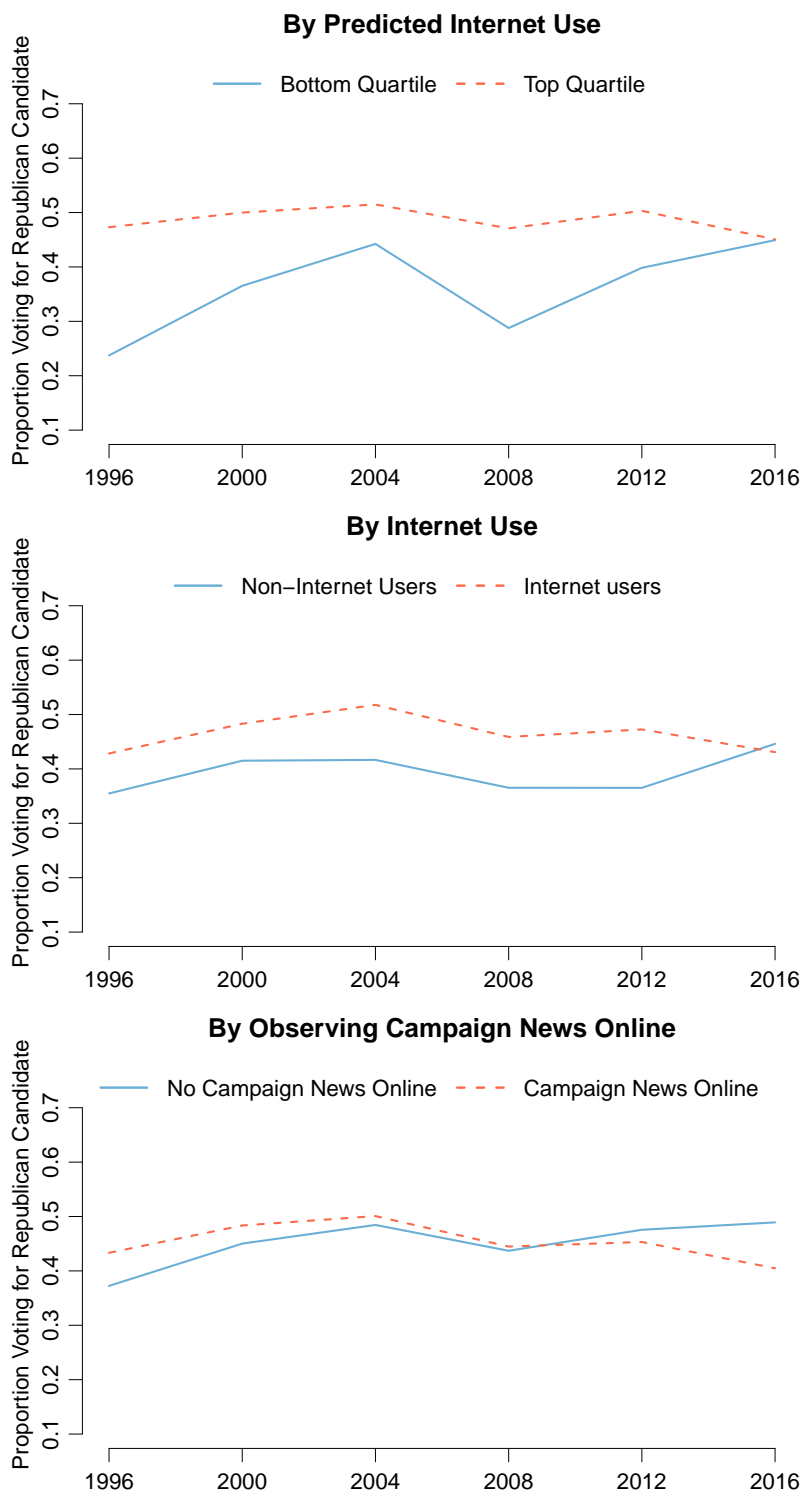
---

<sup>11</sup>See the SI appendix of Boxell et al. (Forthcoming) for details on the bootstrap procedure.

## References

- Allcott, Hunt and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*. 31(2):211–236.
- Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. Forthcoming. Greater internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences*.
- Confessore, Nicholas and Danny Hakim. 2017. Data firm says ‘secret sauce’ aided Trump; Many scoff. *New York Times*. Available at <https://www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html>. Accessed September 14, 2017.
- Hampton, Keith N. and Eszter Hargittai. 2016. Stop blaming Facebook for Trump’s election win. *The Hill*. Available at <http://thehill.com/blogs/pundits-blog/presidential-campaign/307438-stop-blaming-facebook-for-trumps-election-win>. Accessed June 14, 2017.
- Johnson, Eric. 2017. Full transcript: Hillary Clinton at Code 2017. *recode.net*. Available at <https://www.recode.net/2017/5/31/15722218/hillary-clinton-code-conference-transcript-donald-trump-2016-russia-walt-mossberg-kara-swisher>. Accessed September 21, 2017.
- The American National Election Studies (ANES; [www.electionstudies.org](http://www.electionstudies.org)). 2015a. The ANES 1948–2012 Time Series Study [dataset]. Stanford University and the University of Michigan [producers]. Accessed December 22, 2016.
- . 2015b. The ANES 2008 Time Series Study [dataset]. Stanford University and the University of Michigan [producers]. Accessed February 18, 2017.
- . 2016. The ANES 2012 Time Series Study [dataset]. Stanford University and the University of Michigan [producers]. Accessed December 22, 2016.
- . 2017. The ANES 2016 Time Series Study [dataset]. Stanford University and the University of Michigan [producers]. Accessed March 31, 2017.

Figure: Trends in Votes for Republican Presidential Candidate by Online Activity



Notes: Plot shows trends in the weighted proportion of voting respondents that voted for the Republican presidential candidate, separately for groups that are more and less active online. We measure online activity using predicted internet use, actual internet use, and whether or not the respondent observed campaign news online. See main text for details on variable construction.

Table: Votes for Republican Presidential Candidate by Online Activity, 2012-2016

Demographic Group	Change in Proportion	95% CI
By Predicted Internet Use:		
Bottom Quartile	0.051	(-0.0875, 0.1894)
Top Quartile	-0.053	(-0.1508, 0.0457)
<i>Difference</i>	<i>0.104</i>	<i>(-0.0715, 0.2786)</i>
By Internet Use:		
Non-Internet Users	0.081	(-0.1311, 0.2928)
Internet Users	-0.042	(-0.1005, 0.017)
<i>Difference</i>	<i>0.123</i>	<i>(-0.0993, 0.3445)</i>
By Observing Campaign News Online:		
No Campaign News Online	0.014	(-0.0773, 0.1045)
Campaign News Online	-0.048	(-0.1224, 0.0256)
<i>Difference</i>	<i>0.062</i>	<i>(-0.0505, 0.1745)</i>

Notes: Table shows the change between 2016 and 2012 (2016 minus 2012) in the weighted proportion of voting respondents that voted for the Republican presidential candidate, separately for groups that are more and less active online. We measure online activity using predicted internet use, actual internet use, and whether or not the respondent observed campaign news online. The difference row shows the difference in changes between the less active and more active group. The 95% confidence intervals are constructed via a nonparametric bootstrap at the respondent level with 100 replicates and taking the standard deviation of the statistic across replicates. See main text for details on variable construction and Boxell et al. (Forthcoming) for details on the nonparametric bootstrap procedure.