

FOR RELEASE FEBRUARY 25, 2016

# Seven-in-Ten Reddit Users Get News on the Site

*Interest in Democratic presidential candidate Bernie Sanders was clear among Reddit users well in advance of primary season*

**BY** Michael Barthel, Galen Stocking, Jesse Holcomb and Amy Mitchell

**FOR MEDIA OR OTHER INQUIRIES:**

Amy Mitchell, Director of Journalism Research  
Jesse Holcomb, Associate Director of Journalism Research  
Rachel Weisel, Communications Associate  
202.419.4372  
[www.pewresearch.org](http://www.pewresearch.org)

## About Pew Research Center

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping America and the world. It does not take policy positions. The Center conducts public opinion polling, demographic research, content analysis and other data-driven social science research. It studies U.S. politics and policy; journalism and media; internet, science and technology; religion and public life; Hispanic trends; global attitudes and trends; and U.S. social and demographic trends. All of the Center's reports are available at [www.pewresearch.org](http://www.pewresearch.org). Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

© Pew Research Center 2016

## Terminology

**“Subreddits”** are roughly equivalent to forums on message boards and are based around a subject (such as /r/politics) or the process used to discuss a variety of subjects (such as /r/AskAnAmerican or /r/explainlikeimfive). Users visiting a subreddit see a list of submissions. Any user can create a subreddit, and the creator assigns users to “moderate” the subreddit; these users can delete content, ban users and set subreddit rules.

**“Submissions” or “Posts”** are items submitted by users, roughly equivalent to a “post” on a blog. Submissions are made to a specific subreddit (e.g., a submission to /r/politics about voting results in South Carolina). They can be original content (e.g., a user’s thoughts on one candidate’s health care proposals), links to outside content (e.g., a link to a news story about one candidate’s fortunes or a video of a candidate’s speech) or a combination of the two. (In this report, the term “post” is generally used in reference to submissions.)

**“Comments”** are responses, reactions, additions or questions left by Reddit users to a post and/or to other comments.

**“Discussion threads”** are all the comments left in response to a single post and/or to other comments under that post.

**“Authors”** are Reddit users who submit a comment or a post.

**“Upvoting” and “Downvoting”** are ways for users to indicate their opinion of a post or comment they’ve read. Users can “vote” on a comment (or post) by clicking an up arrow (to “upvote”) or down arrow (to “downvote”). When there are more downvotes than upvotes, a comment is said to be “downvoted” and, depending on a user’s settings, may not be visible in the discussion.

For more information, please consult [Reddit’s FAQ](#).

# Seven-in-Ten Reddit Users Get News on the Site

*Interest in Democratic presidential candidate Bernie Sanders was clear among Reddit users well in advance of primary season*

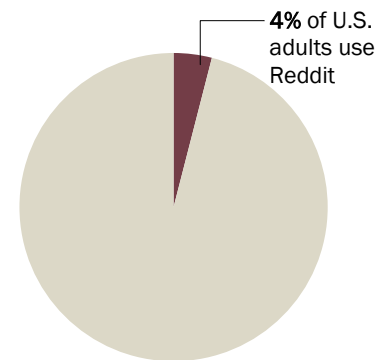
*Updated Edition, May 26, 2016: This new edition of our February 2016 report contains updated survey data. It takes into account newer weighting measures that the Pew Research Center has adopted for its American Trends Panel. Details on the weighting are provided in the appendix.*

The online discussion forum, Reddit, launched in 2005, is proving to be one of the more news-oriented social spaces on the Web. While it reaches a relatively small slice of the population, a large portion of its users report getting news there, and Reddit can drive substantial attention to news events.

A new survey from Pew Research Center finds that while just 4% of U.S. adults report using the site – a user base that is more likely to be young, male and liberal than the general public – 70% of Reddit users say they get news there. What’s more, 45% of Reddit users learn about the 2016 presidential campaign in a given week from the site. This is on par with the portion of Facebook (52%) and Twitter users (43%) who get news and information about the election on those platforms and outpaces most other social networking sites asked about.

Reddit attracts major newsmakers – President Obama conducted a Q&A session on the site (referred to, in the local parlance, as an “Ask Me Anything” or AMAs) in August of 2012, which attracted more than 20,000 comments, and more recently it played host to celebrities such as Amy Poehler in 2014 and Channing Tatum in 2015. With many posts linking to outside content, Reddit sends many of its [millions of daily visitors](#) to publishers, though [not as many as do Facebook or Twitter](#).

## Reddit attracts a small but news-oriented slice of the public



*Of the 4% of U.S. adults who use Reddit...*



Source: Surveys conducted Jan. 12-27 and Jan. 12-Feb. 8, 2016

PEW RESEARCH CENTER

The flow of news and information on the site is somewhat different than that of social networking sites such as Facebook, Twitter or Instagram. On Reddit, users actively choose to participate in specific discussion groups that interest them, called “subreddits,” rather than creating friend networks. These are roughly equivalent to forums or topics on other online message boards such as Digg or Fark. The names of these subreddits generally describe the topic being discussed (such as /r/politics) or the process they use to discuss a variety of subjects (such as /r/AskAnAmerican or /r/explainlikeimfive). Reddit users write and post an original “submission,” and then other users, or the original author, comment on that post or reply to other comments. Unlike on many more traditional news sites where comment sections are secondary to the articles, on Reddit the discussion among users is a main attraction. As such, Reddit is one site that exemplifies a new facet of people’s ability to connect with news by discussing it with other interested citizens outside the confines of their day-to-day, in-person networks. This report is the latest in a [series](#) that explores the role of news on individual social media sites.

To help understand the news dynamics on Reddit, Pew Research Center combined survey work with an inventory and analysis of three months of discussion about the 2016 presidential candidates on Reddit itself.

The opportunity to study the Reddit discussion in such depth was made possible through use of a large dataset of site comments collected by a [researcher](#) (a process that could take months to replicate on one’s own due to API rate limits) who made the May 2015 data publicly available initially in July 2015 and then updated the data set with new monthly data thereafter. Pew Research Center researchers initially downloaded the two most recent months available – May and June – and then later, when made available, added September.

The Center’s researchers then measured discussion surrounding each candidate. The size of the dataset necessitated that researchers adopt a hybrid approach: first human analysts determined the candidate described in a sample of comments, then researchers used these data to train a statistical model that utilized word frequencies to detect which, if any, candidate was referenced in the remaining comments.

The analysis focused on the 2016 presidential campaign, which was getting into full swing during this period with more than 20 announced contenders and the first of the debates. All comments<sup>1</sup>

---

<sup>1</sup> The original data captured by Jason Baumgartner was at the comment level, so for analysis purposes, researchers kept the level of analysis at comments. In addition, comments appear more frequently than do individual posts, and a greater share of users comment than produce their own posts. Because of this, they represent the closest proxy for participant conversation on the site. For more detail on the capture process and month selection, please see the [methodology](#).

naming one of the 21 presidential candidates identified for analysis<sup>2</sup> were captured and studied in depth. Among the highlights of the analysis:

- **Discussion around the 2016 presidential candidates was vigorous in the run-up to the debates and primaries, but did not dominate Reddit.** In the three months studied, Reddit users left more than 350,000 comments mentioning a presidential candidate. Even at that rate though, these comments amount to less than 1% of all comments left on Reddit (more than 165 million) during that period. Similarly, the vast majority of the comments naming a candidate (91%) appeared within 373 subreddits – a fraction of the roughly 84,000 active subreddits in the study, though many of the 373 rank among the most popular subreddits on the site.
- **Nonetheless, where the discussion did take off, it could dominate.** A third of all posts on /r/politics, one of the top 10 most active subreddits in period of study, had at least one comment naming a presidential candidate.
- **Most, but far from all, of the comments naming a candidate took place in forums that were oriented to politics.** During the three months studied, the majority of candidate-centered comments (60%) appeared in political subreddits, especially those focused on general politics (37%). But 40% of comments mentioning a presidential hopeful appeared in non-political subreddits such as information exchanges (11%), web culture (9%), or entertainment and sports (4%).
- **The data also reveal a strikingly high level of early interest in one particular candidate.** In the months before the early presidential debates last fall, Sen. Bernie Sanders was mentioned in more comments (about 165,000) than Hillary Clinton (85,000) and Donald Trump (73,000) combined. The Sanders phenomenon recalls previous presidential candidates who, while not leading in the polls, saw outsized levels of conversation or support in pockets of the internet, including [Ron Paul on Twitter](#), and [Howard Dean with Meetup.com](#) and the early blogosphere.
- **One of the most popular subreddits for Sanders was one dedicated to his presidential candidacy:** The subreddit /r/SandersForPresident received more than 200,000 comments overall in the three months studied, including 59,000 mentioning a candidate. Beyond generally discussing his candidacy, it also serves as a place for grassroots organizing as the banner across the top includes links for users to volunteer for Sanders, to

---

<sup>2</sup> The candidates studied were Ben Carson, Bernie Sanders, Bobby Jindal, Carly Fiorina, Chris Christie, Donald Trump, George Pataki, Hillary Clinton, Jeb Bush, Jim Webb, John Kasich, Lincoln Chafee, Lindsey Graham, Marco Rubio, Martin O'Malley, Mike Huckabee, Rand Paul, Rick Perry, Rick Santorum, Scott Walker and Ted Cruz.

attend Sanders events and to watch Sanders on television. Neither the Clinton nor Trump campaigns, on the other hand, seem to have robust “homes” on the site. Of all the 350,000 comments studied from these three months, just 61 appeared in a candidate-specific subreddit devoted to Clinton and 212 appeared in one dedicated to Donald Trump.

- **Heavy commenting activity is concentrated among a minority of users.** Most users who comment on Reddit only do so sparingly. In this case, fully 60% of those who named a candidate did so in just one comment in the time period studied. By contrast, 25% left three or more comments.
- **Mirroring public attention more broadly, conversation on Reddit grew as the campaign heated up.** The total number of comments made about presidential candidates rose 153% from June to September. This exceeded— but followed — the public’s expressed interest in the campaign: In [polls taken over similar periods](#), the percentage saying they were following news about the campaign “very closely” increased from 16% to 27% — a two-thirds jump.

## 1. Reddit news users more likely to be male, young and digital in their news preferences

While just 4% of U.S. adults report using Reddit, seven-in-ten of these users get news on the site. Overall, 2% of U.S. adults get news on Reddit.

Both Reddit users in general and users who get news on the site tend to be young, male, and to self-identify as liberal at higher rates than the overall public.

About seven-in-ten (71%) of Reddit news users are men, 59% are between the ages of 18 and 29, and 47% identify as liberal, while only 13% are conservative (39% say they are moderate). In comparison, among all U.S. adults, about half (49%) are men, just 22% are 18- to 29-year-olds and about a quarter (24%) say they are liberal.

As could be expected, Reddit news users are also heavy internet users: 47% report going online almost constantly (compared with [21% of U.S. adults overall](#)).

---

### Reddit users and news users more likely to be male and young

*% of U.S. adults, Reddit users and Reddit news users who are ...*

	U.S. adults %	Reddit users %	Reddit news users %
Men	49	67	71
Women	51	33	29
18-29	22	64	59
30-49	34	29	33
50-64	25	6	7
65+	19	1	<1
College degree	28	42	48
Some college	31	40	43
High school or less	41	18	9
White non-Hispanic	65	70	74
Black non-Hispanic	12	7	8
Hispanic	15	12	7
Other non-Hispanic	8	11	10
\$75k+	44	35	42
\$30,000-74,999	35	34	32
<\$30k	21	30	25
Liberal	24	43	47
Moderate	40	38	39
Conservative	35	19	13

Source: Survey conducted Jan. 12-Feb. 8, 2016. N (Reddit users) = 288, N (Reddit news users) = 230. U.S. population (except for political ideology) comes from 2015 March Current Population Survey.

PEW RESEARCH CENTER

---



To get a better understanding of the dynamics of news conversation on Reddit, the analysis in this report focuses on a current and often passionate conversation: the 2016 presidential candidates. Within Reddit users overall (not just Reddit news users), 45% report learning something about the presidential campaign or candidates on the site in a given week. This is especially pronounced among liberal Reddit users, of whom fully 59% said they learned something.

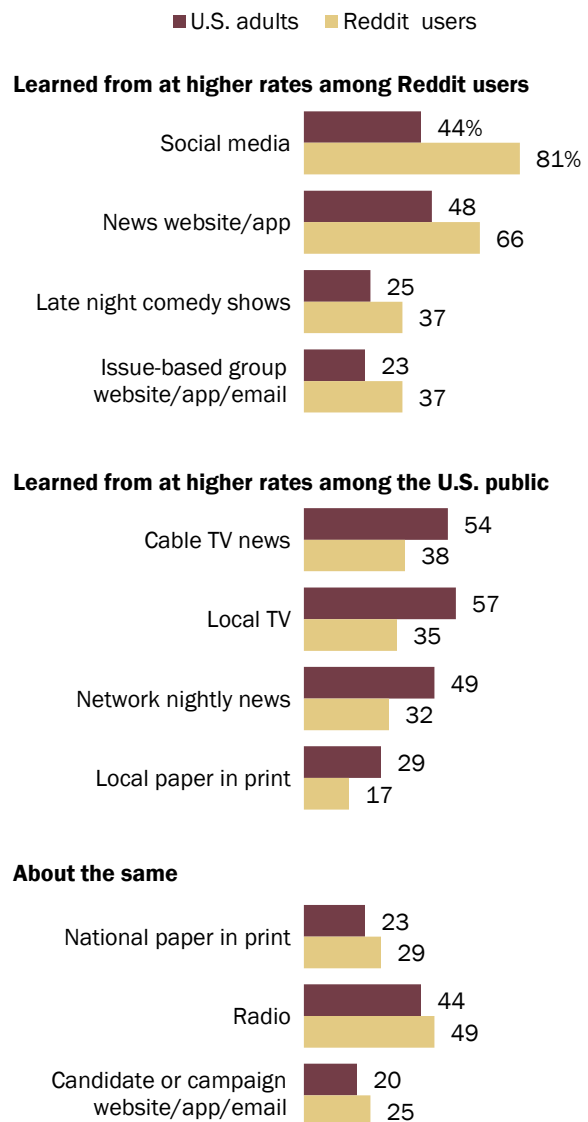
These users are not solely getting news on Reddit. They are also more likely than the general public to be learning about the election from other sources, including news websites or apps, late night comedy shows, and the apps, emails, or websites of issue-based groups. They are less likely, though, to be learning from nightly network news, cable or local TV news, or the print edition of a local daily newspaper.

And when asked for the one type of source they find most helpful for election information, social networking sites ranked first with 44% of Reddit users saying so. About one-in-five (18%) Reddit users specifically named Reddit as the most helpful source.

Many of these differences are in line with Reddit's younger, more liberal user base.

## Reddit users more likely to learn about presidential election from digital sources than the general public

*% who learned about the 2016 presidential election in the past week from ...*



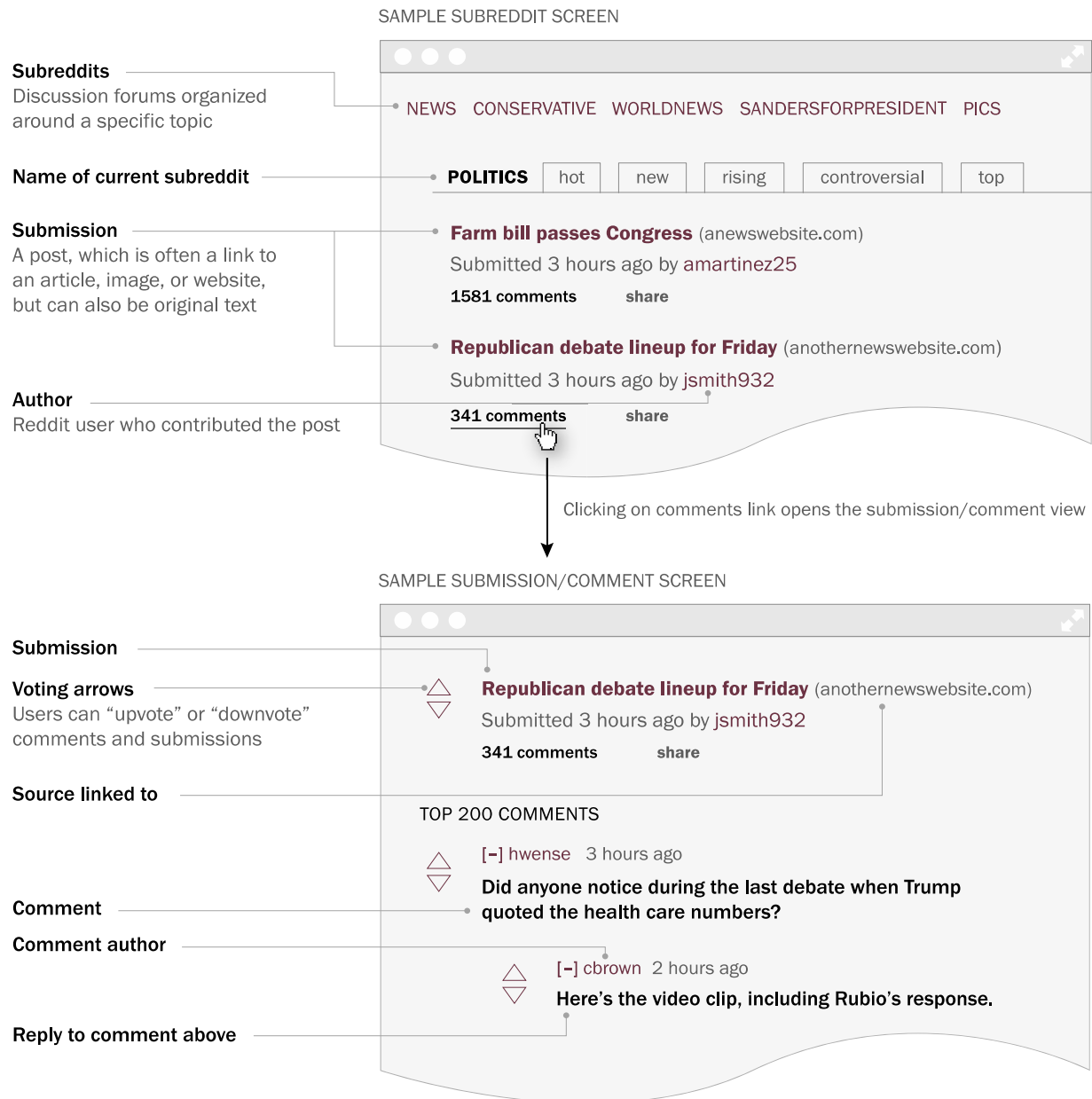
Source: Survey conducted Jan. 12-27, 2016

PEW RESEARCH CENTER

## Understanding Reddit

In its structure and function, Reddit contains elements of a discussion board, social networking site and messaging service. Users submit posts, known as submissions (referred to as “posts” in

### Parts of a Reddit screen



Source: Modeled after reddit.com

PEW RESEARCH CENTER

this report). These can be original content, links to outside content or a combination of the two. Indeed, fully 62% of posts studied here linked to another website (a quarter of these linked to common sites hosting images or videos).<sup>3</sup> Other users can then add comments to this post. In addition to commenting on posts, users can also rate both posts and comments by “upvoting” them (indicating that they are worth being seen by others) or “downvoting” them (indicating that they should not be seen). This voting drives the display of posts and comments on the site.

As a whole, Reddit is organized into subreddits, roughly equivalent to forums or topics on other online message boards. The names of these subreddits generally describe the topic being discussed (such as /r/politics) or the process used to discuss a variety of subjects (such as /r/AskAnAmerican or /r/explainlikeimfive).

Though many Reddit users get news on the site, news is only one of many types of things discussed there. Of the 10 subreddits that attracted the most comments overall in the three months studied here – not just those that named a presidential candidate – two involve video games, two involve sports, and four are general-interest forums such as /r/funny and /r/videos. The main general news subreddit, /r/news, was the ninth highest in total comments during these three months, while /r/worldnews was the tenth highest.

## Tracking discussion of the presidential candidates on Reddit

To shed light on how Reddit users discuss news and current events on the site, Center researchers took advantage of the publicly released work of Jason Baumgartner, a researcher who collected a massive dataset of site comments initially made publicly available in July 2015, with new monthly data added thereafter. Center researchers initially downloaded the two most recent months available, May and June, and then, when it became available, added September. The topic selected for analysis was discussion of the 2016 presidential campaign, operationalized by looking for mentions of the names of the leading presidential contenders in the discussions on Reddit.

---

### News is one of many types of discussion forums on Reddit

*Total # of comments (both those naming a candidate and not) in 10 subreddits with the highest number of comments in May, June and September 2015*

Subreddit	Description	Number of comments
AskReddit	Question and answer forum	12,374,057
leagueoflegends	Video game	3,198,934
funny	Humor	2,205,253
nfl	Sports	1,853,654
pics	Pictures	1,830,152
DestinyTheGame	Video game	1,792,667
videos	Videos	1,758,083
nba	Sports	1,692,089
news	News	1,676,465
worldnews	News	1,640,697

Source: Analysis of publicly available Reddit data from May, June and September 2015.

PEW RESEARCH CENTER

---

<sup>3</sup> Examples include Imgur, YouTube and Gfycat

It is important to note that this analysis is conducted on the comments that appear under posts, not the posts themselves. This decision was made in part because comments appear more frequently than do individual posts, and because they represent the closest thing to a proxy for participant conversation on the site. In addition, the initial data set made available to researchers was at the comment level. While in some ways this is a constraint, the clear focus on comments – which by their nature far exceed the number of original posts – does offer a window into how conversation happens on Reddit.

The analysis represents a first for the Center in that it employs a combination of in-house machine learning and human coding. Machine learning is a technique that allows “trained” statistical models based on word frequencies to stand in for humans in coding extremely large amounts of text or images. Working with a dataset of all comments posted in Reddit, researchers began by performing free-text searches for each candidate’s name. Due to polysemy, the fact that words can have multiple meanings (“trump” could refer to the Republican candidate or an action in a game of bridge), researchers took an extra step. In each of the three months, for all 21 candidates, if the name search returned less than 500 results, researchers coded all results. If it returned more than 500, researchers coded at least 500 but up to 3,000 as necessary for the classifier to achieve reliability. Overall, researchers hand-coded more than 50,000 comments. These results were used to train a machine learning model to classify all remaining comments. The model was structured to err on the side of not including a comment rather than including one in a classification category. That resulted in a final dataset of more than 350,000 comments from about 100,000 different authors on almost 90,000 posts in about 5,000 different subreddits over the three months. The findings presented here are intended to be a systematic analysis of a specific population: people who talk about presidential candidates on Reddit.

## 2. Discussion mentioning the 2016 presidential candidates was vigorous in run-up to primaries, but did not dominate Reddit

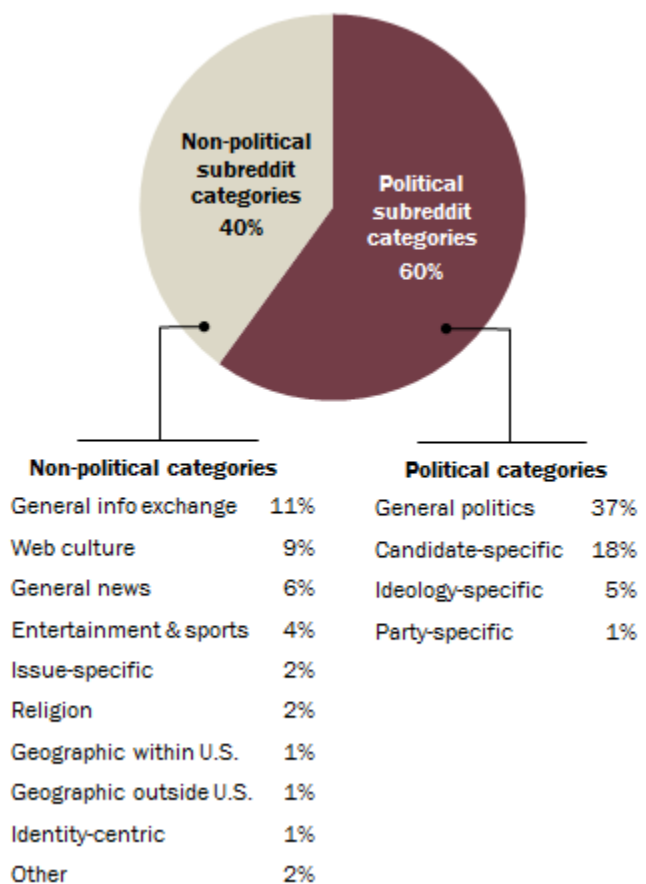
While conversation about the 2016 presidential candidates on Reddit was robust in the summer and early fall of 2015, in the prelude to the primaries, it was one of many topics of conversation on the site competing for attention among its users.

During the three months studied – May, June and September of 2015, more than a year before the general election and five to nine months before the first caucuses and primaries – researchers identified more than 350,000 comments mentioning a presidential candidate.<sup>4</sup>

These amounted to less than 1% of the 165 million-plus comments left on Reddit during that time period, a reflection of the diverse discussion around both news and non-news items that occur on the site. Similarly, of the nearly 84,000 active subreddits – those with any comments at all during the three months studied – comments mentioning candidates appeared in just 6%, or roughly 5,000. In addition, the conversation was further concentrated within those roughly 5,000 subreddits. Fully 91% of these comments appeared in 373 subreddits, many of which also happened to be among the most popular subreddits on the site at the time. Fully 85%

### 40% of comments naming presidential candidates appear in non-political subreddits

*% of comments mentioning at least one presidential candidate that appear in the following subreddit categories*



Source: Analysis of publicly available Reddit data from May, June and September 2015. Numbers may not add to 100% due to rounding.

PEW RESEARCH CENTER

<sup>4</sup> While there is not an exact comparison to discussion activity outside of Reddit, a few examples offer some perspective. [One study](#) of two weeks' worth of comments on the Facebook pages of nearly 200 local news outlets in one city – Denver, Colorado – yielded about 37,000 comments. Looking more closely at discussion about politics on a news website, the most-viewed article on The New York Times website on Feb. 7, 2016, [a story](#) about Bill Clinton's attacks on Bernie Sanders, had amassed about 2,600 comments by the following day. Other platforms for news discussion, however, give rise to numbers that are higher by orders of magnitude: 1.64 million tweets about the 2014 State of the Union speech [appeared on Twitter](#) within a four-hour time span.

of these comments appeared within the top 500 most-trafficked subreddits.

To gain a better sense of the kinds of subreddits included in these 373, researchers categorized them into 14 different topical categories: four that are clearly political in nature and 10 that are not. The total number of subreddits that fall into each of these categories varies considerably. Only comments falling into these 373 subreddits were included in the analysis that follows.

Here, as might be expected, the majority of comments (60%) appeared in political subreddits, especially those focused on general politics (37%).

A smaller, but still sizable, share (40%) of comments appeared in non-political subreddits such as information exchange (11%) and web culture (9%). More than 4,000 of these comments naming candidates, for example, appeared in a subreddit called /r/atheism over the course of the three months studied. About 800 appeared in one called /r/LateShow and nearly 700 in /r/forwardsfromgrandma.

In a broad social platform, one would not expect the candidate discussion to dominate the entire site, and the small slice of conversation devoted directly to the presidential candidates is not even that far out of sync with more mainstream media. For example, in 2011 – the last pre-election year – a [Pew Research Center content analysis](#) found that the 2012 presidential campaign accounted for 9% of the entire mainstream news agenda for the whole year.

Nonetheless, where the discussion did take off, it could dominate. The /r/politics subreddit is one of the top 20 forums on the site, and it hosted about 26,000 posts and a million comments in the three months studied.<sup>5</sup> This analysis identified about 9,000 posts on /r/politics where at least one comment named a presidential candidate – fully a third of all posts made in the period.

Conversely, in /r/news – the most visible setting for discussion of current events on Reddit – the 12,000-plus comments naming a candidate represented less than 1% of all /r/news comments posted during the three months.<sup>6</sup> While there may well have been other conversations about the campaign – in looking at the September /r/news posts with the most comments overall, not just those naming a candidate – election-related posts would not appear in the top 10. Instead, this candidate-driven discussion was overshadowed by topics such as Kim Davis, the county clerk who refused to issue marriage license to gay couples – a subject that accounted for three of the top 10 posts.

---

<sup>5</sup> For the month of September, in terms of total comments posted (not just those mentioning a candidate).

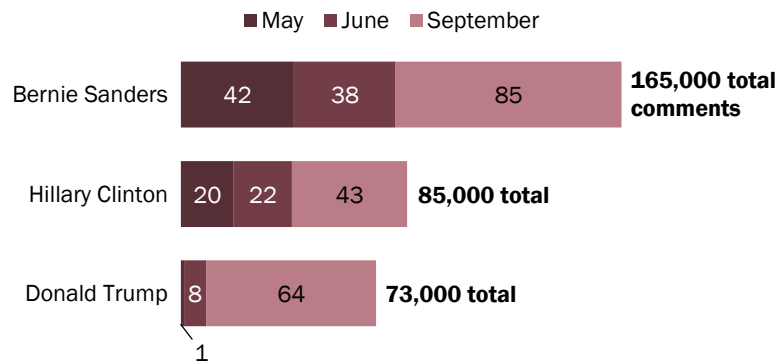
<sup>6</sup> Part of the reason for this may be due to the fact that political discussion is discouraged in this subreddit, which says in its sidebar that a post “will likely be removed if it primarily concerns politics.”

### 3. In discussions about presidential candidates, Sanders mentioned far more than others

The discussions of nearly all major presidential candidates grew over the three months studied, but within the liberal-leaning Reddit community, the discussions of some candidates far outweighed others. Three candidates received more than 70,000 comments across the three months (May, June and September 2015), while the vast majority were mentioned in fewer than 10,000 comments.<sup>7</sup>

#### The three most-named candidates all see increased attention from early summer into fall

Number of comments mentioning \_\_\_\_, in thousands



Source: Analysis of publicly available Reddit data from May, June and September 2015.

PEW RESEARCH CENTER

Only one candidate seemed to have an especially active subreddit dedicated to their campaign during the three months studied: Vermont Sen. Bernie Sanders. As discussed below, the /r/SandersForPresident subreddit attracted tens of thousands of comments and prominently offered users the chance to [engage with the campaign](#). Of all other subreddits dedicated to individual candidates, only those focused on Sen. Rand Paul (R-Ky.) attracted any significant number of comments, and even then the number of comments naming a candidate in /r/RandPaul only amounted to about 2% of those in /r/SandersForPresident. Some candidates even attracted subreddits dedicated entirely to making fun of the candidate, and two subreddits dedicated to Donald Trump, /r/The\_Donald and /r/donaldtrump, have argued over which is the real Trump 2016 subreddit, with [one accusing the other](#) of being run by Sanders supporters.

The discussions around the three candidates that received by far the most comments – Sen. Bernie Sanders, former Secretary of State Hillary Clinton and business mogul Donald Trump – offer a window into the kinds of conversation patterns that can take place on Reddit.

<sup>7</sup> See topline for full results.

## Bernie Sanders

Though Sanders was [less well known by the public than Clinton and Trump](#) during the time period studied, he was mentioned in about 165,000 comments – more than Clinton (about 85,000) and Trump (about 73,000) combined. The number of comments naming Sanders rose over time, doubling between June (about 38,000) and September (about 85,000).

One event that may have helped fuel the conversation around Sanders was an “Ask Me Anything” (AMA) session held by the candidate in May 2015 that attracted nearly 13,000 separate comments, far more than the average of 60 comments for AMAs that month. That was clearly not the only driver of comments, as the high level of interest in Sanders continued in June and roughly doubled by September, even without the AMA as a focus of attention.

One of the most popular subreddits for Sanders was one dedicated to his presidential candidacy:

[/r/SandersForPresident](#). While at least five of the candidates in the study have name-specific subreddits devoted to them, Sanders’ stands out. The candidate’s subreddit received more than 200,000 comments overall in the three months studied, with more than 59,000 that mention any candidate by name. In addition

to discussion, the channel also serves as a place for grassroots organizing, as the banner across the top includes links for users to volunteer for Sanders, to attend Sanders events and to watch Sanders on television. The information available does not indicate whether the subreddit was originally created by or as a part of his campaign, but it clearly serves as a pro-Sanders hub. Only

### Sanders mentioned in Reddit comments more than Clinton, Trump combined

Number of \_\_\_\_\_ that mention each candidate at least once

	Comments	Authors	Posts	Subreddits
Sanders	164,953	49,180	32,561	371
Clinton	84,749	30,442	20,804	365
Trump	73,220	38,720	20,023	361

Source: Analysis of publicly available Reddit data May, June, September 2015.

PEW RESEARCH CENTER



The homepage banner of the [/r/SandersForPresident](#) subreddit.

Retrieved Feb. 18, 2016.

PEW RESEARCH CENTER



4% of comments in /r/SandersForPresident were “downvoted,” less than half the rate we saw among comments from all subreddits (for more on downvoting, read [here](#)).

Not all of the subreddits devoted to Sanders are supportive; at least one criticizes the Reddit community’s focus on Sanders. Comments in this type of forum accounted for less than 1% of all comments about Sanders over the three months studied.

There is also evidence of a highly engaged base of commenters discussing Sanders. Almost one-in-four users (24%) who named Sanders in a comment posted three or more comments naming him. This is a higher proportion of users leaving three or more comments than for any other candidate studied.

The Sanders phenomenon on Reddit echoes former lesser-known candidates who garnered attention or support in digital platforms despite the fact that they did not have as high a profile in traditional media. In 2011, Texas Rep. Ron Paul, though behind other Republican candidates in the polls and in mainstream media coverage, enjoyed somewhat [more activity on Twitter](#) – with [more than 1.1 million tweets](#) referencing his campaign, which ranked him fifth among eight Republican candidates and gave him a more favorable tone in the Twitter conversation than any of the other candidates.

And back in the 2003-04 primary season, before social media era became a major force in politics, Vermont Gov. Howard Dean attracted a digital-savvy activist base that [marshalled web-based tools such as meetup.com to organize and discuss their candidate](#).

## Hillary Clinton

Hillary Clinton, the second-most discussed candidate on Reddit during the three months studied, was mentioned in about 85,000 comments. As with others, that number rose over time: In September, Clinton was named in 43,000 comments – about twice as many as May or June.

Unlike Sanders, whose campaign-specific subreddit attracted a great deal of discussion about the candidate, Clinton’s campaign and supporters do not appear to use a dedicated subreddit in the way that Sanders does. Of all the 350,000 comments that mentioned a candidate during the three months studied, just 61 appeared in a candidate-specific subreddit devoted to Clinton.

Nonetheless, users who mentioned Clinton seemed to be relatively active participants: 21% left three or more comments naming her within the three months, second only to 24% of comments mentioning Sanders.

---

### About one quarter of Reddit users who mention Sanders do so repeatedly

*% of Reddit users who mentioned the following candidates in at least one comment who left 3 or more comments naming the candidate*

	%
Bernie Sanders	24
Hillary Clinton	21
Rand Paul	16
Donald Trump	14
Marco Rubio	13
Jeb Bush	12
Carly Fiorina	12
Scott Walker	12
Ben Carson	12
Ted Cruz	11
Mike Huckabee	9
Chris Christie	8

Source: Analysis of publicly available Reddit data from May, June and September 2015.

Note: Only candidates with at least 5,000 total comments shown.

**PEW RESEARCH CENTER**

---

## Donald Trump

Donald Trump was named in about 73,000 comments on Reddit during the three months studied, making him the third-most mentioned candidate.

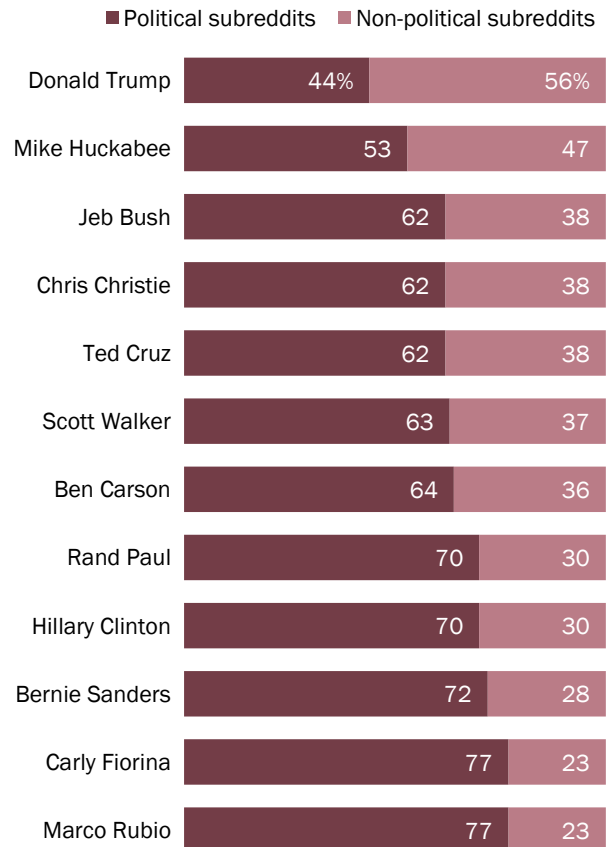
The Reddit discussion about Donald Trump stands out for its rapid growth over time, jumping from just 8,000 in June – the month Trump announced his candidacy – to nearly 64,000 in September. This eightfold increase outpaces the overall rise in the number of comments studied, which rose about 150% from June to September.

Not only was there more discussion about Trump as the campaign progressed – the scope of where he was being discussed broadened as well. Generally, comments about the Democratic candidates ranged across most subreddits, as both Clinton and Sanders were discussed each month in at least 300 of the 373 subreddits where candidates were mentioned most often. Comments naming Trump, however, rose from 146 in May to 268 in June and 355 in September.

Comments mentioning Trump also stand out for being more likely to occur in general information exchange subreddits such as /r/AskReddit or the “Ask Me Anything” forum. About one-in-five of these comments (19%) appeared in these subreddits, 8 percentage points higher than comments connected to any other candidate.

### More than half of comments naming Trump appear in non-political subreddits

% of comments mentioning \_\_\_\_ found in ...



Source: Analysis of publicly available Reddit data from May, June and September 2015.

Note: Only candidates with at least 5,000 total comments shown.

PEW RESEARCH CENTER

## 4. Discussion in Reddit circles more likely to focus on one candidate and one party

A question that arises with any social media discussion forum is the extent to which users and the messages they post venture beyond their own worldview or remain in their own silos of news and information. The data here do not offer a definitive answer on how this dynamic plays out on Reddit, but we can analyze the dynamic around comments mentioning candidates. The data here show that most comments – and even comments within a full discussion thread – generally narrow in on one candidate or party.

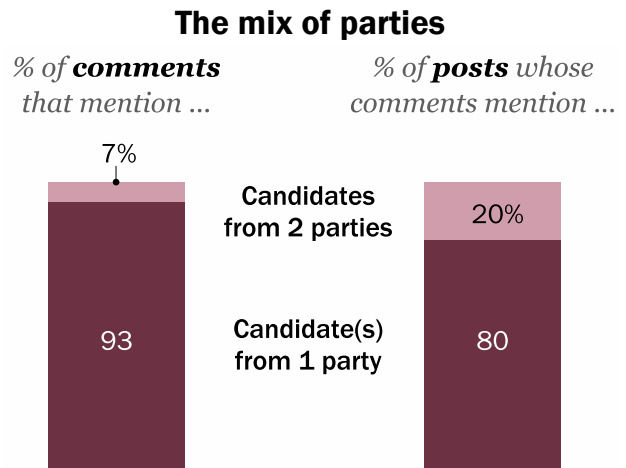
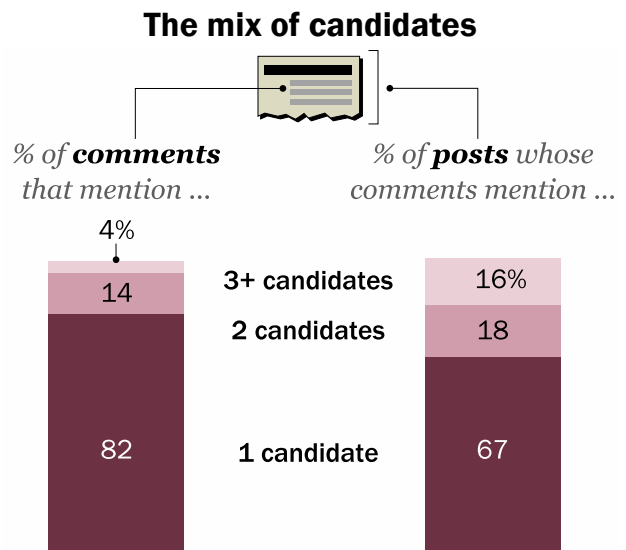
### Candidates and parties

Across the 350,000 comments studied, the vast majority mentioned just one candidate by name. Of all comments mentioning a candidate, about eight-in-ten (82%) name just one, while 14% name two and 4% mention three or more candidates.

Looking at the data another way suggests that at the broader level, an only slightly more inclusive discussion is occurring. In addition to examining multi-candidate mentions at the *comment* level, researchers also looked at the *post* level. Within the 373 subreddits where candidates were mentioned most often, more than 67,000 posts contained at least one comment naming a candidate. These posts, many containing hundreds or thousands of comments, can contain a lengthy conversation about politics and the campaign.

### Most Reddit comments mention one candidate

Within all comments that mention at least one presidential candidate ...



Source: Analysis of publicly available Reddit data from May, June and September 2015.

PEW RESEARCH CENTER

Within all posts with at least one follow-up comment naming a candidate, about two-thirds (67%) saw multiple candidates mentioned in their comment threads, and 16% had comment threads in which three or more candidates were mentioned. Most individual comments stick to just one candidate, in other words, and multiple candidates are just slightly more likely to be named within the broader conversation.

Looking at the party level to get a sense of when the candidate conversation cuts across party lines, the vast majority of both comments and posts studied here mention candidates from only one party. As with candidates, mentioning multiple parties is rare. Fully 93% of comments named candidates from only one party, and this pattern is no different at the post level. For eight-in-ten posts, their comments contain only mentions of one party's candidates.

## Upvotes and downvotes

Within the Reddit community, “upvotes” and “downvotes” are features users can employ to express approval or disapproval of a particular comment. Comments can be downvoted because people disliked them, or because they are off-topic, irrelevant or out of step with the rules of conversation.

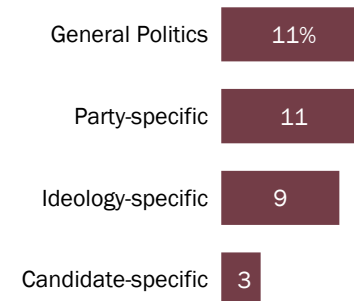
Of all the comments on Reddit during the time period studied – both those naming candidates and those not – fully 91% had a positive score. Just 5% of comments had negative scores, and 5% had a zero score, indicating either that the comment had received no votes or that there were roughly equal numbers of up and downvotes. Thus in general, downvoted comments are the exception rather than the norm.

This finding holds true for comments mentioning a presidential candidate. Across all of these comments in the three months studied, 11% were downvoted. Candidate-specific subreddits were especially unlikely to see downvoted comments – just 3% in all, compared with 11% of those in general political subreddits and of those in party-oriented threads. This could indicate that subreddits focused on a particular candidate attract a pretty supportive mix of users. What complicates this, though, is the control moderators of subreddits have over the rules about what is acceptable to post. Moderators could, for example, remove critical comments even before users have a chance to downvote them, or they could ban certain users altogether. Thus, it is unclear whether a low level of downvoting is tied to a naturally supportive user base or to a heavily controlled discussion arena.

---

### Candidate-specific subreddits least likely to see downvotes

*% of comments about presidential candidates with more downvotes than upvotes, among political subreddit categories*



Source: Analysis of publicly available Reddit data from May, June and September 2015.

PEW RESEARCH CENTER

---

## 5. Heavy commenting activity concentrated among a minority of users

The engagement level of Reddit users taking part in the candidate discussions over the three months studied echoes earlier research about other digital spaces. The majority of authors offered just a single comment naming a candidate, while fewer came back with multiple comments over the course of the study.

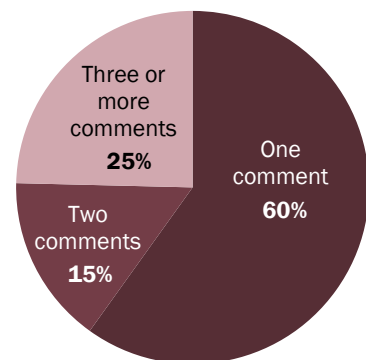
Fully six-in-ten of almost 100,000 authors posted just a single comment naming a presidential candidate. A quarter posted three or more.

Undoubtedly there may be instances where a user added additional comments without mentioning a candidate directly. But these findings reinforce much of what we have found in previous studies: A [2015 study](#) of local news habits found that just a small share (less than one-fifth) of commenters to Facebook pages of news outlets left more than two comments. Another study found that on [Twitter](#) a small core of active users (12%) tweeted 100 times or more during a four week period. Among all social media users, [we have found](#) that just a small fraction say they have posted videos or photos they themselves have taken of a news event.

---

### Most naming a candidate on Reddit do so in only one comment

*% of Reddit users who left \_\_\_\_\_ naming presidential candidates*



Source: Analysis of publicly available Reddit data from May, June and September 2015.

PEW RESEARCH CENTER

---

## Repeat commenters more likely to discuss multiple candidates and parties – and to be downvoted

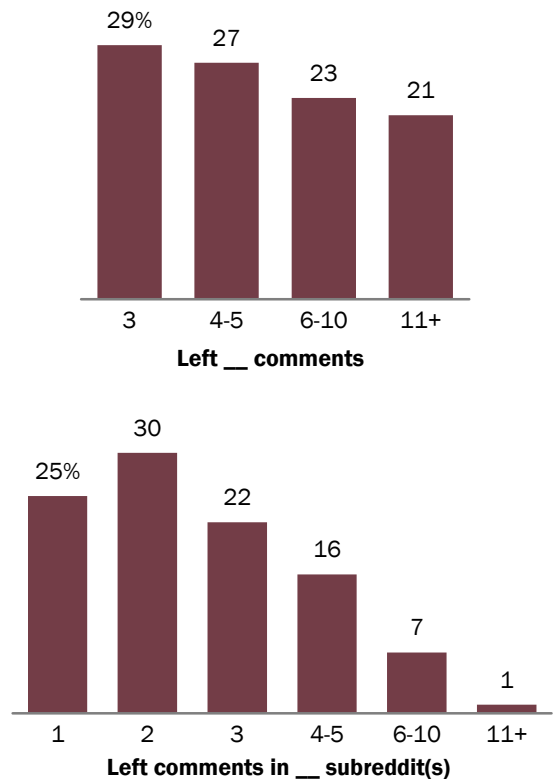
Highly engaged users – those who left three or more comments – tended to leave more than just three comments, as well as participate in multiple forums. About one-in-five of these commenters (21%) left more than 10 comments naming a candidate over the three months studied, and nearly half (45%) posted a comment naming at least one candidate in three or more subreddits.

Users who left three or more comments differed in terms of the number of candidates and parties they mentioned. Among these active commenters, 90% mentioned multiple candidates, with about two-thirds (64%) mentioning three or more. This far outpaces the 38% of all users who mentioned multiple candidates (19% of whom mentioned three or more).

This is also true at the party level. Nearly three-quarters (72%) of these highly engaged commenters named candidates from both the Republican and Democratic parties, compared with only about a quarter (26%) of all users in our analysis.

### Repeat commenters drive conversation on Reddit

*Of users who left three or more comments naming presidential candidates, % who ...*



Source: Analysis of publicly available Reddit data from May, June and September 2015.

PEW RESEARCH CENTER



Data on upvoting and downvoting behavior reinforce the notion that repeat commenters operate differently on Reddit than do one-time commenters.

Overall, 15% of authors had one or more comments naming a presidential candidate voted down. However, among repeat commenters, that jumps to nearly four-in-ten (36%). This could be because such commenters are more passionate on the subject and thus more likely to inspire a reaction, or it may more simply be that posting multiple times presents more opportunities to have a comment voted down.

### Repeat Reddit commenters stand out for whom they name and how their comments are received

Among \_\_\_\_, % who ...

	All users	Users who left 3 or more comments
	%	%
Named 1 candidate	62	10
Named 2 candidates	19	26
Named 3 or more candidates	19	64
Named 1 party	74	28
Named both parties	26	72
Had at least 1 comment downvoted	15	36

Source: Analysis of publicly available Reddit data from May, June and September 2015.

PEW RESEARCH CENTER

## Acknowledgments

This report was made possible by The Pew Charitable Trusts. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

This report is a collaborative effort based on the input and analysis of the following individuals:

### Primary researchers

Michael Barthel, *Research Associate*

Galen Stocking, *Research Associate*

Jesse Holcomb, *Associate Director, Journalism Research*

Amy Mitchell, *Director, Journalism Research*

### Research team

Elisa Shearer, *Research Analyst*

Jeffrey Gottfried, *Research Associate*

Kristine Lu, *Research Assistant*

### Graphic design

Margaret Porteus, *Information Graphics Designer*

### Machine learning advisor

Solomon Messing, *Director, Data Labs*

### Communications and web publishing

Shannon Greenwood, *Assistant Digital Producer*

Rachel Weisel, *Communications Associate*

In addition, the report benefited from feedback provided by the following individuals: Alex Leavitt (PhD candidate, Annenberg School for Communication & Journalism, University of Southern California) and J. Nathan Matias (PhD candidate, Center for Civic Media in the MIT Media Lab and fellow at Harvard's Berkman Center for Internet and Society). While the analysis was guided by our consultations with the advisers, the Pew Research Center is solely responsible for the interpretation and reporting of the data.

## Methodology

This report consists of two distinct research elements, with separate methodologies: a representative survey of U.S. adults conducted through Pew Research Center’s American Trends Panel and a content study of comments posted on the social forum reddit.com.

Reddit, whose slogan proclaims it to be the “front page of the Internet,” is part social network, part online forum (read [here](#) for more on how Reddit works). It features a broad mix of topics dispersed across a complex but structured environment. Moreover, its reach extends beyond the confines of the site itself, from driving traffic to external sites to attracting attention from political figures, including President Obama.

## Survey

### *American Trends Panel January 2016 wave*

The American Trends Panel (ATP), created by Pew Research Center, is a nationally representative panel of randomly selected U.S. adults living in households. Respondents who self-identify as internet users and provided an email address participate in the panel via monthly self-administered Web surveys, while those who do not use the internet or decline to provide an email address participate via the mail. The panel is being managed by Abt SRBI.

Data in this report pertaining to the Reddit usage and news usage numbers are drawn from the January wave of the panel, conducted Jan. 12-Feb. 8, 2016, among 4,654 respondents (4,339 by Web and 315 by mail). Panelists who have access to the internet but take surveys by mail were not sampled in this wave (i.e. mail respondents to this wave are all non-Internet users). The margin of sampling error for the full sample of 4,654 respondents is plus or minus 2 percentage points.

Members of the American Trends Panel were recruited from two large, national landline and cellphone random digit dial (RDD) surveys conducted in English and Spanish. At the end of each survey, respondents were invited to join the panel. The first group of panelists was recruited from the 2014 Political Polarization and Typology Survey, conducted Jan. to March 16, 2014. Of the 10,013 adults interviewed, 9,809 were invited to take part in the panel and a total of 5,338 agreed to participate.<sup>8</sup> The second group of panelists was recruited from the 2015 Survey on Government,

---

<sup>8</sup> When data collection for the 2014 Political Polarization and Typology Survey began, non-internet users were subsampled at a rate of 25%, but a decision was made shortly thereafter to invite all non-internet users to join. In total, 83% of non-internet users were invited to join the panel.

conducted Aug. 27 to Oct. 4, 2015. Of the 6,004 adults interviewed, all were invited to join the panel, and 2,976 agreed to participate.<sup>9</sup>

Participating panelists provided either a mailing address or an email address to which a welcome packet, a monetary incentive and future survey invitations could be sent. Panelists also receive a small monetary incentive after participating in each wave of the survey.

*This new edition of our February 2016 report contains updated survey data. It takes into account newer weighting measures that the Pew Research Center has adopted for its American Trends Panel.*

Under the revised protocol, the ATP data are weighted in a multi-step process that begins with a base weight incorporating the respondents' original survey selection probability and the fact that in 2014 some panelists were subsampled for invitation to the panel. An adjustment was made for the fact that the propensity to join the panel and remain an active panelist varied across different groups in the sample. The final step in the weighting uses an iterative technique that matches gender, age, education, race, Hispanic origin and region to parameters from the U.S. Census Bureau's 2014 American Community Survey. Population density is weighted to match the 2010 U.S. Decennial Census. Telephone service is weighted to estimates of telephone coverage for 2016 that were projected from the January-June 2015 National Health Interview Survey. Volunteerism is weighted to match the 2013 Current Population Survey Volunteer Supplement. It also adjusts for party affiliation using an average of the three most recent Pew Research Center general public telephone surveys. Internet access is adjusted using a measure from the 2015 Survey on Government. Frequency of internet use is weighted to an estimate of daily internet use projected to 2016 from the 2013 Current Population Survey Computer and Internet Use Supplement. The share of respondents who get news from 10 different social networks was weighted to match a Pew Research Center Journalism survey from March-April, 2016.

---

<sup>9</sup> Respondents to the 2014 Political Polarization and Typology Survey who indicated that they are internet users but refused to provide an email address were initially permitted to participate in the American Trends Panel by mail, but were no longer permitted to join the panel after Feb. 6, 2014. Internet users from the 2015 Survey on Government who refused to provide an email address were not permitted to join the panel.

The following table shows the unweighted sample sizes and the error attributable to sampling that would be expected at the 95% level of confidence for different groups in the survey:

<b>Group</b>	<b>Unweighted sample size</b>	<b>Plus or minus ...</b>
Total sample	4,654	2.4%
Reddit users	288	9.5%
Reddit news users	230	10.7%

Sample sizes and sampling errors for other subgroups are available upon request.

In addition to sampling error, one should bear in mind that question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls.

The Web component of the January wave had a response rate of 69% (4,339 responses among 6,301 Web-based individuals in the panel); the mail component had a response rate of 67% (315 responses among 474 non-Web individuals in the panel). Taking account of the combined, weighted response rate for the recruitment surveys (10.0%) and attrition from panel members who were removed at their request or for inactivity, the cumulative response rate for the January ATP wave is 2.9%.<sup>10</sup>

#### *American Trends Panel January 2016 early respondents*

Data in this report that pertains to learning about the 2016 presidential election from Reddit are drawn from the early respondents to the January 2016 wave of the panel. These 3,760 respondents consisted of 3,661 Web panelists who had completed the survey by January 27 and 99 mail panelists whose responses had been received by January 22. The margin of sampling error for these 3,760 respondents is plus or minus 2.3 percentage points. The weighting procedure for this early responder part of the January 2016 wave was the same as the full sample as discussed above.

The following table shows the unweighted sample sizes and the error attributable to sampling that would be expected at the 95% level of confidence for different groups in the survey:

<sup>10</sup> Approximately once per year, panelists who have not participated in multiple consecutive waves are removed from the panel. These cases are counted in the denominator of cumulative response rates.

<b>Group</b>	<b>Unweighted sample size</b>	<b>Plus or minus ...</b>
Total sample	3,760	2.3%
Learned about the election from Reddit	113	13.1%

Sample sizes and sampling errors for other subgroups are available upon request.

In addition to sampling error, one should bear in mind that question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls.

The Web component of the early responders of the January 2016 wave had a response rate of 58.1% (3,661 responses among 6,301 Web-based individuals in the panel); the mail component had a response rate of 20.9% (99 responses among 474 non-Internet users in the panel). Taking account of the combined, weighted response rate for the recruitment surveys (10.0%) and attrition from panel members who were removed at their request or for inactivity, the cumulative response rate for the early respondents to the January 2016 ATP wave is 2.5%.

## Content analysis

To gain a sense of how presidential candidates have been discussed within Reddit, researchers gathered publicly available comments and analyzed them using a mix of machine learning techniques and human content analysis.

Researchers examined comments that included the names of U.S. presidential candidates during the months of May, June and September 2015. Center researchers initially downloaded the two most recent months available – May and June – and then later, when made available, added September, when the campaign was getting into full swing and the candidates were engaging in the first debates.

## Data Acquisition

For this project, researchers pulled mentions of every major political candidate during May, June and September 2015 from a dataset of all Reddit comments published by Jason Baumgartner (<http://pushshift.io>) to reddit.com, archive.org and Google BigQuery in June 2015 and updated monthly since then. This dataset was constructed by requesting comments in large, concurrent batches from the Reddit application program interface (API). Researchers conducted tests on a

subset of the data (comments from the month of June) to ensure that relevant comments were not missing from the dataset overall. With the exception of those that had been deleted before Baumgartner’s software requested them, no comments were found to be missing from the data.

Researchers took several steps to identify and study a total of 21 presidential candidates. To explore the conversation about those candidates, researchers developed and refined SQL queries on the Google BigQuery dataset described above to identify relevant comments for each one. This iterative process was used to return as many comments that were about a candidate as possible without returning large swaths of comments that were clearly not related to the candidate.

Accordingly, researchers began with a candidate’s first and/or last name. All searches for candidates included both their first and last name as a string (e.g. “Ben Carson”) as well as their last name (“Carson”), with the exception of Rand Paul, for whom the candidate’s full name was used, as “Paul” would have captured too many off-topic comments. Depending on the prominence of the candidate’s first name in discussions about him or her, researchers also included it as an independent search term. For instance, Jeb Bush is often referred to just as Jeb, especially in informal discussions; whereas Scott Walker is infrequently referred to as just Scott – and if he was, it would be difficult to distinguish between discussions of the Wisconsin governor and others sharing his first name. Researchers then constructed SQL queries with regular expression search terms to account for possessive permutations of the name, variances in spacing and punctuation, and other characteristics of Web discussions.

Once the initial queries were constructed, researchers tested them for accuracy by combing the results for frequent off-topic discussions. Because researchers were using names as the search terms, famous individuals with similar names were often captured by the search results. For example, Rick Perry’s search results included discussions of actor Matthew Perry, singer Katy Perry, and filmmaker and actor Tyler Perry. We systematically excluded those names from the search results. The point of this exercise was not to narrow down the search results only to those that were about the candidate – we recognized early on that SQL was not well suited to this task – but to ensure that the result set was pruned of the most prominent false positives, which would lower the amount of off-topic comments in subsequent steps.

To be sure, there are natural limitations to this approach. For instance, comments may discuss candidates without naming them – opting instead to use pronouns such as “him” or “her.” It is also important to note that this analysis is conducted on the comments that appear under submissions, not the submissions themselves. This decision was made in part because comments appear more frequently than do individual posts, and because they represent the closest thing to a proxy for participant conversation on the site. In addition, the initial data set made available to researchers was at the comment level. (While researchers did some exploratory analysis of submission data

gathered directly from the Reddit API, analyzing the submissions would have made for a distinct and separate project altogether.) Despite these limitations, the data collected and studied here represent the essence of how conversation functions on this news-heavy platform.

## **Coding**

The resulting dataset consisted of over 2.2 million comments, a number large enough to render manual content analysis impractical. Instead, this project utilized manual analysis to train a machine learning model to determine which comments were discussing presidential candidates.

A team of coders first evaluated each comment that was identified from the database using a binary classification scheme – either the comment was about the specific presidential candidate or it was not. In each of the three months, for all 21 candidates, if the name search returned less than 500 results, researchers coded all results. If it returned more than 500, researchers coded at least 500 but up to 3,000+ in order to achieve reliability on the classifier. Overall, researchers hand-coded more than 50,000 comments. Four coders coded May and June data after reaching a Cohen’s Kappa of .86. A separate group of three coders coded September data. Coders reached a Cohen’s Kappa of .87, within the acceptable range of reliability.

Researchers then employed Support Vector Machines (SVMs), which are a frequently used class of machine learning models. When used in text analysis, they classify data by analyzing the language used across the entire document (in this case, each “document” consists of a single comment on Reddit). SVMs receive input in the form of training and test data, which is a set of documents that have been coded manually by trained researchers. Analyzing the training data, the algorithm identifies patterns in the language used for all documents associated with each code and develops a set of rules to classify subsequent documents. It is evaluated using a set of labeled documents not used in training.

Researchers tuned the SVMs such that they were able to predict new documents with a score of .80 or higher.

After the SVMs constrained the dataset to just those comments that were about presidential candidates, researchers further inspected the non-news or politically relevant subreddits with high numbers of comments in which a candidate was identified. Researchers randomly inspected 50 comments from these subreddits, and if the vast majority of them were not about the candidate in question, that subreddit was excluded from analysis. This occurred most frequently in the Entertainment category.

The final dataset included more than 350,000 comments.



## Subreddit Coding

Much of the analysis in this report was conducted on comments within broad subreddit categories, rather than individual comments or submissions. To determine the categories, manual content analysis was employed. A team of three coders with expertise in social media developed a codebook with 14 subreddit categories. After training on this codebook, these coders coded a sample of 48 subreddits by inspecting the subreddit description or wiki (where available) and frontpage posts and comments. This process achieved a Cohen's Kappa of .82.

All subreddits with at least 50 comments were coded by a member of this team, resulting in 409 subreddits. Using the method described above, 36 of these subreddits were removed because their comments were misidentified by the SVM. Subreddits were categorized accordingly:

- Web culture: 63
- Entertainment & sports: 55
- Issue-specific: 35
- Geographic (within U.S.): 32
- Candidate-specific: 28
- General information exchange: 24
- General politics: 24
- Geographic (outside U.S.): 19
- Ideology-specific: 18
- Identity-centric: 13
- General news: 12
- Religion: 11
- Party-specific: 3
- Other: 36

© Pew Research Center, 2016