What the Public Learned about Donald Trump and Hillary Clinton during the 2016 Campaign

Frank Newport,1 Lisa Singh,2 Stuart Soroka,3 Michael Traugott,3 and Andrew Dugan1

1 Gallup  2 Georgetown University  3 University of Michigan
Background

There is a good deal of research capturing trends in government/candidate approval and evaluations.

We are often able to connect these trends to campaign events, and/or shifts in media content.

We are rarely able to capture directly or over time the information that voters remember.

As a consequence, we have only a partial, or triangulated, view of how exactly campaigns shape attitudes.
Background

We explored the benefits of a different approach to measuring campaign dynamics: open-text responses to "read, seen or heard" questions about candidates.

We did so using data from the Gallup U.S. Daily poll, which interviewed a sample of 500 Americans each night from July 10 to November 7 (N = 58,941).

The base question is “Have you read, seen or heard anything about (Hillary Clinton / Donald Trump) in the last day or two?”

If “Yes,” respondents were asked what that was, and verbatim responses were recorded by the interviewers.
**Method**

We explored these data using automated content analysis, with very few changes to the raw text aside from some simple pre-whitening to exclude ‘stopwords’ (i.e., ‘and,’ ‘or,’ etc) and to remove plurals (i.e., ‘emails’ - > ‘email’)

We estimated the tone of comments using the Lexicoder Sentiment Dictionary, although we focus here both on tone and on simple word counts
Exhibits

The graphics we present here are intended to capture (in this order):

1. Simple word counts for each candidate, as a descriptive of the campaign generally
2. A comparison of the words that most uniquely identify the two candidates, as an illustration of the language that was most attached to / recalled about each campaign
3. Over-time word counts for each candidate, as a descriptive of campaign dynamics
4. Over-time sentiment analysis, across partisan groups, as an alternative illustration of trends in approval
5. A demonstration of links between sentiment and favorability
6. Correlations between words and responses to favorability questions, to identify the words most strongly associated with candidate approval
Levels of Information in the Electorate about Trump and Clinton

Did You Read, See or Hear Anything About Trump/Clinton in Last Day or Two?

Seven-day rolling averages: 2016

- % Yes, Clinton
- % Yes, Trump

GALLUP DAILY TRACKING
Basic Descriptives: Top Words for Clinton, Entire Sample, 7/10 – 11/7

e-mail
debate
health
liar
fbi
scandal
campaign
speech
people
going
foundation
talking
saw
president
benghazi
read
convention
issues
like
remember
investigation
last
think
information
said
trying
wikileaks
one
billclinton
state
money
everything
leaks

cant
good
new
lot
well
berniesanders
tax
wiki
campaigning
made
nothing
know
get
article

got
released
negative
regarding
night
today
political
server
anything
election
time
dnc
bad
coming

doesnt
done
right
see
dishonest

done
disclosed
good

cases
release
good
bad

country

done
disclosed
good
bad

country

health
liar
saw
get
like
president

email
Basic Descriptives: Top Words for Trump, Entire Sample, 7/10 – 11/7
Words here are ‘distinguishing’ based on a relatively simple formula, implemented using the wordcloud package in R:

Let $p_{i,j}$ be the rate at which word $i$ occurs in document $j$, and $p_j$ be the average across documents ($\sum_i p_{i,j} / \text{ndocs}$). The size of each word is mapped to its maximum deviation ($\max_i (p_{i,j} - p_j)$); and the category to which it is assigned (Democrat in blue or Republican in red) is determined by the document where that maximum occurs.
Basic Descriptives: Words Distinguishing “Read Seen or Hear” about Clinton and Trump, Entire Sample, Across Partisans, 7/10 – 11/7

Clinton, for Democrats (blue) and Republicans (red)

Trump, for Republicans (red) and Democrats (blue)
Campaign Dynamics: Weekly Trends in “Read, Seen or Heard” about Clinton, All Respondents, 7/10 - 11/7

Jul 10-16 email berniesanders fbi liar scandal
Jul 17-23 email convention liar benghazi speech
Jul 24-30 convention email speech dnc president
Jul 31-6 email liar convention speech campaign
Aug 7-13 email liar tax foundation scandal
Aug 14-20 email foundation liar health going
Aug 21-27 email foundation liar scandal health
Aug 28-3 email foundation liar scandal health
Sep 4-10 email liar scandal health
Sep 11-17 email liar scandal health
Sep 8-24 health debate email liar issues campaign
Sep 25-1 debate email tax liar campaign
Oct 2-8 debate email tax liar campaign
Oct 9-15 debate email tax liar campaign
Oct 16-22 debate wikileaks liar leaks
Oct 23-30 debate wikileaks liar campaign
Oct 31-5 email fbi investigation reopening scandal
Nov 6-7 email fbi foundation investigation scandal
Campaign Dynamics: Weekly Trends in “Read, Seen or Heard” about Trump, All Respondents, 7/10 - 11/7

Jul 10-16  mikepence dallas going shooting pick
Jul 17-23  convention speech wife national going
Jul 24-30  russia email convention speech going
Jul 31-6  muslim family 600 pollyan convention
Aug 7-13  isis family obama second people
Aug 14-20  campaign speech isis people louisiana
Aug 21-27  immigration campaign people louisiana trying
Aug 28-3  mexico immigration president speech going
Sep 4-10  mexico immigration president going speech
Sep 11-17  health speech talking abuse going
Sep 8-24  obama born talking new york bombing
Sep 25-1  debate missuniverse going women
Oct 2-8  tax debate women missuniverse paid
Oct 9-15  women debate video years ago
Oct 16-22  women debate election sexual rigged
Oct 23-30  women election going debate speech
Oct 31-5  email women fbi talking campaign
Nov 6-7  speech going people campaign
Campaign Dynamics: Topic Focus by Candidate Mentions over Time

Topics are initially determined using Latent Dirichlet Allocation, but the number of topics is then reduced manually, and uncoded keywords receiving many mentions in survey responses are added to existing topics, or used to create new topics. Each response is finally assigned a single topic based on the frequency of topic-coded words.
The Tone of respondents' 'Read, Seen or Heard' comments, where 'Trump Advantage' is the difference between the tone of responses about Trump and the tone of responses about Clinton. The 'tone' of candidate responses is based on % positive words minus % negative words, using the Lexicoder Sentiment Dictionary, applied to all open-ended responses to Gallup's 'read, seen or heard' question. Results here are shown using 3-day averages, from September 2\textsuperscript{nd} onward.
The Relevance of the Sentiment Dictionary to Candidate Favorability

Based on OLS regressions of candidate favorability (scaled from -1 to +1) on gender, age cohorts, race, region, and party ID, alongside counts of positive and negative words in R’s recollections of each candidate. N=55,795. Graphs show the estimated independent impact of positive and negative words (as defined in the LSD) on candidate favorability.
Words Most Strongly Correlated with (a binary measure of) Favorability for each Candidate

*Low Clinton Favorability:*
liar (.16), email (.13), scandal (.10), Benghazi (.08), dishonest (.05)

*High Clinton Favorability:*
debate (.07), campaign (.06), campaigning (.05), president (.05), speech (.05)

*Low Trump Favorability:*
women (.07), sexual (.05)

*High Trump Favorability:*
economy (.06), jobs (.06), makeamericagreat (.06), media (.06), speech (.06), America (.05), country (.05), plan (.05)
Conclusions

These findings offer a first glimpse into the messages / language that stuck with voters, and thus colored their candidate assessments, during the 2016 campaign.

Email-related scandals clearly dominated recalled words about Clinton. This is true for almost every week of the campaign.

There was no similarly common theme for Trump, whose multiple scandals produced a changing, and perhaps more easily overcome, narrative during the campaign.

This is evident not just in word counts, but in the words correlating most strongly with favorability: there is a language of support for Trump that does not exist for Clinton, and a language of not-support for Clinton that does not exist for Trump.
Conclusions

Negative recollections of Clinton appear to be the most powerful element of “read, seen or heard” comments where candidate favorability is concerned: these are the words most strongly predictive of both favorability for Trump, and not-favorability for Clinton.

These findings thus fit with popular accounts of the campaign, emphasizing Clinton’s lack of popularity over Trump’s popularity.

They also offer an opportunity to explore the source and timing of the public’s “read, seen or heard” comments. To what extent is an emphasis on email the result of media coverage? What news items are reported on but do not find their way in the public consciousness? We hope to address these questions in future work.