In this post, we investigate how (& how often) members of the 116th House of Representatives refer to the 45th president of the United States on Twitter. TRUMP, POTUS, PRESIDENT TRUMP, @realDonaldTrump — options abound. Here, we consider how a House Rep’s stance towards (or opinion of) 45 influences the choice of referring expression, as well as how this stance aligns with the popularity of 45 in a House Rep’s congressional district.

A fully reproducible, R-based code-through.

**A very brief introduction**

Most linguistic variation is riddled with nuanced meaning, the source of which is often some type of socio-cultural value (Du Bois 2007). In the case of variation in reference, one dimension of this socio-cultural value is **status**. While “President Donald Trump” and “Donald Trump” point to the same referent, the former emphasizes the status of 45 as POTUS, while the latter downplays this status (Berg et al. 2019).

Similarly, “Mr. Trump” is a more deferential referring expression than “Trump”. We know this as speakers of English because of social convention: we refer to folks higher up in the food chain in different (ie, more formal) ways. A simple formality cline is presented below:

1. First name only < Last name only < Full name < Title and last name < Title and full name (Berg et al. 2019)

As a speaker, I can abide by this convention when referring to an elder/boss/POTUS/etc (by using forms towards the right of the cline), or I can flout it (by using forms to the left). In either case, I (theoretically) communicate my stance towards the referent to my **audience**.

In the case of a tweeting House Rep, this audience is their Twitter following (ie, ~their constituency). And if a House Rep is stancetaking when referring to 45 on Twitter, presumably how this audience feels about 45 mediates the polarity of the House Rep’s stance. Another, presumably safer, option would be to not refer to 45 at all. This is what we investigate here.

**Some open source data sets**

library(tidyverse)

**Legislators & vote margins**

We first grab some data/details about the 116th House of Representatives from a few online sources. For House Rep names, congressional districts, and twitter handles, we use a data set made available by [the @unitedstates project](https://theunitedstates.io/). The project is a fantastic resource; maintained by folks from GovTrack, ProPublica, MapLight & FiveThirtyEight.

leg\_dets <- '<https://theunitedstates.io/congress-legislators/legislators-current.csv>'

house\_meta <- read.csv((url(leg\_dets)),

stringsAsFactors = FALSE) %>%

filter(type == 'rep' & twitter!='') %>% #

select(type, bioguide\_id, icpsr\_id, last\_name, state, district, party, twitter) %>%

mutate(district = ifelse(district == 0, 'AL', district),

CD = paste0(state, '-', stringr::str\_pad(district, 2, pad = '0')),

twitter = toupper(twitter))

For **Trump vote margins by congressional district**, we utilize a data set made available by the [DailyKos](https://www.dailykos.com/).

url <- '<https://docs.google.com/spreadsheets/d/1zLNAuRqPauss00HDz4XbTH2HqsCzMe0pR8QmD1K8jk8/edit#gid=0>'

margins\_by\_cd <- read.csv(text=gsheet::gsheet2text(url, format='csv'),

skip = 1,

stringsAsFactors=FALSE) %>%

mutate(trump\_margin = Trump - Clinton) %>%

select(CD, trump\_margin)

**Tweets: 116th House of Representatives**

Next we gather tweets for members of the 116th House of Representatives using the rtweet package. Members took office on January 3, 2019, so we filter tweets to post-January 2. We also exclude retweets. (Last tweets collected on 8-19-19).

congress\_tweets <- rtweet::get\_timeline(

house\_meta$twitter,

n = 2000,

check=FALSE) %>%

mutate(created\_at = as.Date(gsub(' .\*$', '', created\_at))) %>%

filter(is\_quote == 'FALSE' &

is\_retweet == 'FALSE' &

created\_at > '2019-01-02' &

display\_text\_width > 0)

setwd("/home/jtimm/jt\_work/GitHub/x\_politico")

#saveRDS(congress\_tweets\_tif, 'congress\_tweets\_tif.rds')

saveRDS(congress\_tweets, 'congress\_tweets\_tif.rds')

Then we **join the Twitter and House lawmaker detail data sets**:

congress\_tweets <- congress\_tweets %>%

mutate(twitter = toupper(screen\_name)) %>%

select(status\_id, created\_at, twitter, text) %>%

inner\_join(house\_meta %>% filter(type == 'rep'))

For a high level summary of **how often members of 116th House have been tweeting since taking office**, we summarize total tweets by House Rep. The density plot below summarizes the distribution of House Reps’ tweeting habits by party affiliation. So, Democrats (in blue) a bit more active on Twitter.

total\_tweets <- congress\_tweets %>%

group\_by(party, twitter) %>%

summarize(all\_tweets = n())

total\_tweets %>%

ggplot( aes(all\_tweets,

fill = party)) +

ggthemes::scale\_fill\_stata() +

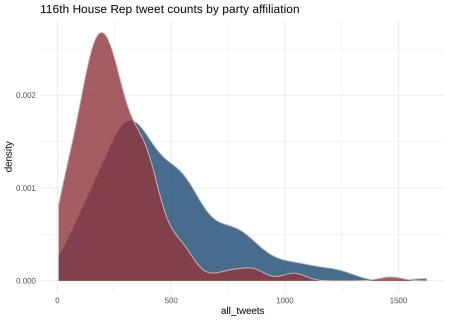
theme\_minimal()+

geom\_density(alpha = 0.8,

color = 'gray')+

labs(title="116th House Rep tweet counts by party affiliation")+

theme(legend.position = "none")



**Some additional summary statistics** about the tweeting habits of House Reps by party affiliation:

x <- list(

'REP' = summary(total\_tweets$all\_tweets[total\_tweets$party == 'Republican']),

'DEM' = summary(total\_tweets$all\_tweets[total\_tweets$party == 'Democrat']))

cbind(party =names(x$DEM), x%>% bind\_rows()) %>%

mutate(REP = round(REP), DEM = round(DEM))%>%

t(.) %>%

knitr::kable()

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| party | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| REP | 5 | 143 | 230 | 272 | 360 | 1466 |
| DEM | 23 | 278 | 404 | 463 | 591 | 1624 |

**Extracting referring expressions to 45**

With tweets and some legislator details in tow, we can now get a beat on how members of the 116th House refer to POTUS 45 on Twitter. Here we present a quick-simple approach to extracting Twitter-references to 45.

The code below summarizes the set of 45 referring expressions (in regex terms) that will be our focus here. It is not exhaustive. The list is ultimately a product of some trial/error, with less frequent forms being culled in the process (eg, #45). We have included “Trump Administration” in this set; while not exactly a direct reference to 45, it is super frequent and (as we will see) an interesting example.

s1 <- "Trump Admin(istration)?"

s2 <- '@realDonaldTrump'

s3 <- '(@)?POTUS'

s4 <- 'Mr(\\.)? President'

s5 <- "the president"

s6 <- '(Pres(\\.)? |President )?(Donald )?\\bTrump'

searches <- c(s1, s2, s3, s4, s5, s6)

potus <- paste(searches, collapse = '|')

**The procedure below** extracts instantiations of the regex terms/patterns above from each tweet in our corpus.

potus\_sum <- lapply(1:nrow(congress\_tweets), function(x) {

spots <- gregexpr(pattern = potus, congress\_tweets$text[x], ignore.case=TRUE)

prez\_gram <- regmatches(congress\_tweets$text[x], spots)[[1]]

if (-1 %in% spots){} else {

data.frame(doc\_id = congress\_tweets$status\_id[x],

twitter = congress\_tweets$twitter[x],

prez\_gram = toupper(prez\_gram),

stringsAsFactors = FALSE)} }) %>%

data.table:::rbindlist() %>%

mutate(prez\_gram = trimws(prez\_gram),

prez\_gram = gsub('\\.', '', prez\_gram),

prez\_gram = gsub('ADMIN$', 'ADMINISTRATION', prez\_gram),

prez\_gram = gsub('PRES ', 'PRESIDENT ', prez\_gram),

prez\_gram = gsub('@', '', prez\_gram)) %>%

left\_join(house\_meta)

**A sample of the output** is presented below.

set.seed(149)

potus\_sum %>% select(doc\_id:prez\_gram) %>% sample\_n(5) %>% knitr::kable()

| **doc\_id** | **twitter** | **prez\_gram** |
| --- | --- | --- |
| 1126141623803568128 | REPMATTGAETZ | REALDONALDTRUMP |
| 1110977855561838593 | REPJOEKENNEDY | TRUMP ADMINISTRATION |
| 1110615484335034373 | REPFRANKLUCAS | THE PRESIDENT |
| 1131334181815017472 | REPTEDLIEU | POTUS |
| 1136024493049163777 | REPWILSON | REALDONALDTRUMP |

Based on the above output, **the table below** summarizes the frequency of expressions used to reference 45 by party affiliation.

data.frame(table(potus\_sum$party, potus\_sum$prez\_gram)) %>%

spread(Var1, Freq) %>%

rename(prez\_gram = Var2) %>%

rowwise() %>%

mutate(Total = sum (Democrat, Republican)) %>%

arrange(desc(Total)) %>%

janitor::adorn\_totals(c('row')) %>% #Cool.

knitr::kable()

| **prez\_gram** | **Democrat** | **Republican** | **Total** |
| --- | --- | --- | --- |
| TRUMP | 6818 | 375 | 7193 |
| THE PRESIDENT | 3816 | 890 | 4706 |
| REALDONALDTRUMP | 2015 | 1989 | 4004 |
| POTUS | 1023 | 1802 | 2825 |
| TRUMP ADMINISTRATION | 2497 | 133 | 2630 |
| PRESIDENT TRUMP | 1649 | 776 | 2425 |
| DONALD TRUMP | 224 | 23 | 247 |
| MR PRESIDENT | 195 | 42 | 237 |
| PRESIDENT DONALD TRUMP | 15 | 12 | 27 |
| Total | 18252 | 6042 | 24294 |

**Party-level stance towards 45**

Based on the counts above, we next investigate potential evidence of stancetaking at the party level. Here, we assume that Reps are supportive of 45 and that Dems are less supportive. If House Reps are stancetaking on Twitter, we would expect that Democrats use less formal terms to downplay the status of 45 & that Republicans use more formal terms to highlight the status of 45.

To get a sense of which terms are more prevalent among each party, we consider the probability of each party using a particular expression to refer to 45. Then we calculate the degree of formality for a given expression as the simple ratio of the two usage rates – where the higher rate is treated as the numerator. Terms prevalent among Democrats are transformed to negative values.

The table below summarizes these ratios, which can be interpreted as follows: Reps are ~5.3 times more likely than Dem colleagues to refer to 45 on Twitter as POTUS; Dems are ~6 times more likely to refer to 45 as Trump.

ratios <- potus\_sum %>%

group\_by(party, prez\_gram) %>%

summarize(n = n()) %>%

group\_by(party) %>%

mutate(per = round(n/sum(n), 3))%>%

group\_by(prez\_gram) %>%

mutate(n = sum(n)) %>%

spread(party, per) %>%

mutate(ratio = ifelse(Republican > Democrat,

Republican/Democrat,

-Democrat/Republican),

ratio = round(ratio, 2)) %>%

filter(n > 60) %>%

select(-n) %>%

arrange(desc(ratio))

ratios %>% knitr::kable()

| **prez\_gram** | **Democrat** | **Republican** | **ratio** |
| --- | --- | --- | --- |
| POTUS | 0.056 | 0.298 | 5.32 |
| REALDONALDTRUMP | 0.110 | 0.329 | 2.99 |
| PRESIDENT TRUMP | 0.090 | 0.128 | 1.42 |
| THE PRESIDENT | 0.209 | 0.147 | -1.42 |
| MR PRESIDENT | 0.011 | 0.007 | -1.57 |
| DONALD TRUMP | 0.012 | 0.004 | -3.00 |
| TRUMP | 0.374 | 0.062 | -6.03 |
| TRUMP ADMINISTRATION | 0.137 | 0.022 | -6.23 |

**The visualization below** summarizes formality ratios for 45 referring expressions as a simple cline. Less formal terms (prevalent among Democrats) are in blue; More formal terms (prevalent among Republicans) are in red.

#cut <- 1.45

ratios %>%

mutate(col1 = ifelse(ratio>0, 'red', 'blue')) %>%

ggplot(aes(x=reorder(prez\_gram,

ratio),

y=ratio,

label=prez\_gram,

color = col1)) +

# geom\_hline(yintercept = cut,

# linetype = 2, color = 'gray') +

# geom\_hline(yintercept = -cut,

# linetype = 2, color = 'gray') +

geom\_point(size= 1.5,

color = 'darkgray') +

geom\_text(size=4,

hjust = 0,

nudge\_y = 0.15)+

annotate('text' , y = -5, x = 7, label = 'Democrat') +

annotate('text' , y = 5, x = 3, label = 'Republican') +

ggthemes::scale\_color\_stata() +

theme\_minimal() +

labs(title="Twitter-based formality cline") + ##?

theme(legend.position = "none",

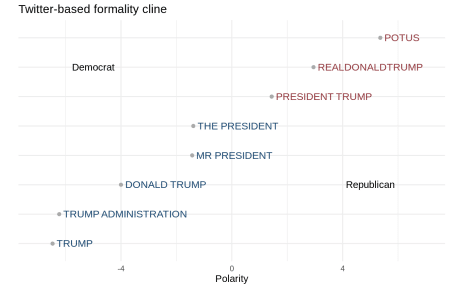
axis.text.y=element\_blank(),

axis.ticks.y=element\_blank())+

xlab('') + ylab('Polarity')+

ylim(-7, 7) +

coord\_flip()



**So, some real nice variation**. Recall our initial (& very generic) formality cline presented in the introduction:

1. First name only < Last name only < Full name < Title and last name < Title and full name

Compared to our House Rep, Twitter-based, 45-specific cline:

1. Trump Administration < Trump < Donald Trump < Mr. President < The President < President Trump < realDonaldTrump < POTUS

While alignment between (1) & (2) is not perfect, the two are certainly conceptually comparable, indeed suggesting that House Reps are choosing expressions to refer to 45 based on stance. Terms prevalent among House Dems downplay the status of 45 by excluding titles & explicit reference to the office (eg, TRUMP, DONALD TRUMP). In contrast, terms prevalent among Republicans highlight the status of 45 via direct reference to the office (eg, PRESIDENT TRUMP, POTUS). More neutral terms (eg, MR PRESIDENT, THE PRESIDENT) reference the office but not the individual.

While the Twitter handle @realDonaldTrump does not highlight the status of the presidency per se, it would seem to carry with it some Twitter-based deference. (I imagine the “real-” prefix is also at play here.) The prevalence of the acronym POTUS among Reps is interesting as well. On one hand, it is super economical; on the other hand, the acronym unpacked is arguably the most deferential. The prevalence of Trump Administration among Dems is also curious – it would seem to be a way to reference 45 without actually referencing (or conjuring images of) either the individual or the office.

**House Rep stance & 2016 presidential vote margins**

The next, and more interesting, piece is how stancetaking plays out at the House Rep level. While the formality cline presented above illustrates some clear divisions between how Dems and Reps refer to the president, its gradient nature speaks to individual variation.

In this section, we (1) present a simple method for quantifying House Rep-level variation in formality when referring to 45, and (2) investigate the extent to which district-level support for 45 in the 2016 presidential election can account for this variation.

ratios <- ratios %>%

mutate(polarity = case\_when(

ratio > 1.4 ~ 'Formal',

ratio < -2.5 ~ 'LessFormal',

ratio > -2.5 & ratio < 1.4 ~ 'Neutral'))

To get started, we first categorize each reference to 45 in our data set as either Formal (POTUS, REALDONADTRUMP, PRESIDENT TRUMP), Less Formal (TRUMP ADMINISTRATION, TRUMP, DONALD TRUMP), or Neutral (MR TRUMP, THE PRESIDENT). Reference to 45 (per legislator) is then represented as a (count-based) distribution across these three formality categories.

wide <- potus\_sum%>%

filter(prez\_gram %in% unique(ratios$prez\_gram)) %>%

left\_join(ratios %>% select(prez\_gram, polarity)) %>%

group\_by(CD) %>%

mutate(prez\_tweets = length(unique(doc\_id)))%>%

group\_by(CD, twitter, last\_name, party, polarity, prez\_tweets) %>%

summarize(n = n())

Formality distributions for a random set of House Reps are summarized in the plot below. So, lots of variation – and presumably 435 House Reps that refer to 45 with varying degrees of formality.

set.seed(171)

samp <- sample(margins\_by\_cd$CD, 10)

pal <- c('#395f81', 'gray', '#9e5055')

names(pal) <- c('LessFormal', 'Neutral', 'Formal')

wide %>%

filter(CD %in% samp) %>%

group\_by(CD) %>%

mutate(per = n/sum(n))%>%

select(-n) %>%

spread(polarity, per) %>%

ungroup() %>%

mutate(rank = rank(Formal),

lab = paste0(last\_name, ' (', CD, '-', substr(party, 1,1), ')')) %>%

gather(key = polarity, value = per, LessFormal, Formal, Neutral)%>%

mutate(polarity = factor(polarity,

levels = c('LessFormal', 'Neutral', 'Formal'))) %>%

ggplot (aes(x = reorder(lab, rank), y = per, fill = polarity)) +

geom\_bar(position = "fill",

stat = "identity") +

coord\_flip()+

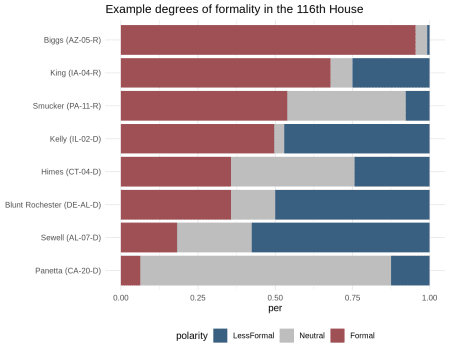
theme\_minimal()+

theme(legend.position = 'bottom',

axis.title.y=element\_blank()) +

scale\_fill\_manual(values = pal) +

ggtitle('Example degrees of formality in the 116th House')



**Based on these distributions**, we define a given House Rep’s degree of formality as the (log) ratio of the number of formal terms used to refer to 45 to the number of less formal terms used to refer to 45. Neutral terms are ignored.

Values greater than one indicate a prevalence for referring expressions that highlight the status of 45; values less than one indicate a prevalence for referring expressions that downplay the status of 45. The former reflecting a positive/supportive stance; the latter a negative/less supportive stance. A relative & rough approximation.

wide1 <- wide %>%

group\_by(CD) %>%

mutate(prez\_refs = sum(n)) %>%

spread(polarity, n) %>%

ungroup() %>%

replace(., [is.na](http://is.na)(.), 1) %>%

mutate(ratio = round(Formal/LessFormal, 3)) %>%

inner\_join(margins\_by\_cd)%>%

left\_join(total\_tweets)

So, to what extent does a congressional district’s collective support for 45 (per 2016 Trump margins) influence the degree of formality with which their House Rep refers to 45? Do House Reps representing districts that supported HRC in 2016, for example, use less formal terms to convey a negative stance towards 45, and mirror the sentiment of their constituents (ie, their ~Twitter followers & ~audience)?

The plot below illustrates the relationship between House Reps’ degrees of formality on Twitter & 2016 presidential vote margins for their respective congressional districts. As can be noted, a fairly strong, positive relationship between the two variables.

wide1 %>%

filter(prez\_tweets > 10) %>%

ggplot(aes(x = trump\_margin,

y = log(jitter(ratio)),

color = party) ) +

geom\_point()+

geom\_smooth(method="lm", se=T,

color = 'steelblue')+

geom\_text(aes(label=last\_name),

size=3,

check\_overlap = TRUE,

color = 'black')+

ggthemes::scale\_color\_stata()+

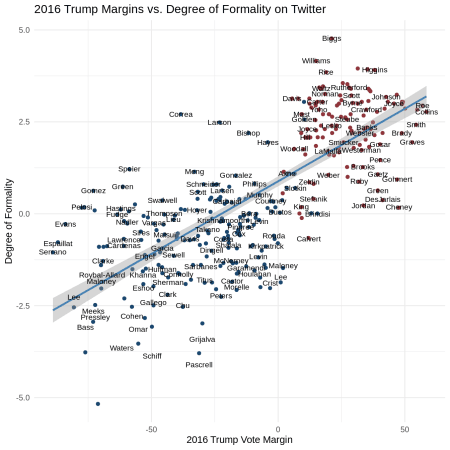
theme\_minimal()+

theme(legend.position = "none",

axis.title = element\_text())+

xlab('2016 Trump Vote Margin') + ylab('Degree of Formality')+

ggtitle('2016 Trump Margins vs. Degree of Formality on Twitter')



So, not only are there systematic differences in how Dems & Reps reference 45 on Twitter, these differences are gradient within/across party affiliation: **formality in reference to 45 increases as 2016 Trump margins increase**. House Reps are not only hip to how their constituents (the audience) feel about 45 (the referent), but they choose referring expressions (and mediate stance) accordingly.

**Prevalence of 45 reference**

Next we consider how often members of the 116th House reference 45 on Twitter, which we operationalize here as the percentage of a House Rep’s total tweets that include reference to 45.

wide2 <- wide1 %>%

mutate(party = gsub('[a-z]', '', party),

trump\_margin = round(trump\_margin,1),

per\_prez = round(prez\_tweets/all\_tweets, 2)) %>%

select(CD, last\_name, party, per\_prez, all\_tweets, trump\_margin) %>%

arrange(desc(per\_prez))

**The density plot below** summarizes the distribution of these percentages by party affiliation. A curious plot indeed. The **bimodal nature of the House Dem distribution** sheds light on two distinct approaches to Twitter & 45 among House Dems. One group that takes a bit of a “no comment” approach and another in which reference to 45 is quite prevalent.

wide2 %>%

ggplot( aes(per\_prez,

fill = party)) +

ggthemes::scale\_fill\_stata() +

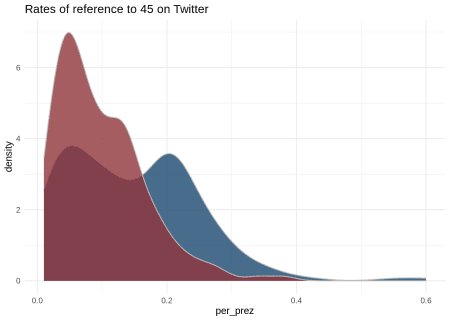
theme\_minimal()+

geom\_density(alpha = 0.8,

color = 'gray')+

labs(title="Rates of reference to 45 on Twitter")+

theme(legend.position = "none")



The table below summarizes 45 tweet **reference rates** for members of the 116th House, along with total tweets & 2016 Trump vote margins for some context. Lots going on for sure. Curious to note that Maxine Waters (CA-43) and Adam Schiff (CA-28) reference 45 on Twitter at the highest rates, despite being fairly infrequent tweeters in general. Almost as if they use Twitter for the express purpose of commenting on the president and/or defending themselves from the president’s Twitter-ire.

out <- wide2 %>%

DT::datatable(extensions = 'FixedColumns',

options = list(scrollX = TRUE,

fixedColumns = list(leftColumns = 1:3)),

rownames =FALSE,

width="450") %>%

DT::formatStyle('per\_prez',

background = DT::styleColorBar(wide2$per\_prez, "lightblue"),

backgroundSize = '80% 70%',

backgroundRepeat = 'no-repeat',

backgroundPosition = 'right')

**Rates of 45-reference, total tweets & 2016 Trump margins for members of the 116th House:**

Last question, then: to what extent does a congressional district’s collective support for 45 (per 2016 Trump margins) influence the rate at which House Reps reference 45 on Twitter?

The much talked about freshmen class of House Dems, for example, is largely comprised of folks from districts that supported Trump in 2016. As such, freshmen Dems are generally more centrist ideologically, representing districts with mixed feeling towards 45. Do they tend to play it safe on Twitter (and with their constituents), and keep the president’s name out of their Twitter mouths?

**Per the plot below**, this would seem to be the case (although freshmen Dems are not explicitly identified). Circe size reflects total tweet count. House members on both sides of the aisle representing districts with slimmer 2016 Trump margins reference 45 on Twitter at lower rates.

wide2 %>%

ggplot(aes(x = trump\_margin,

y = per\_prez,

color = as.factor(party),

size = all\_tweets) ) +

geom\_point()+

geom\_smooth(method = "lm", se = T)+

geom\_text(aes(label=last\_name),

size=3,

check\_overlap = TRUE,

color = 'black')+

ggthemes::scale\_color\_stata()+

theme\_minimal()+

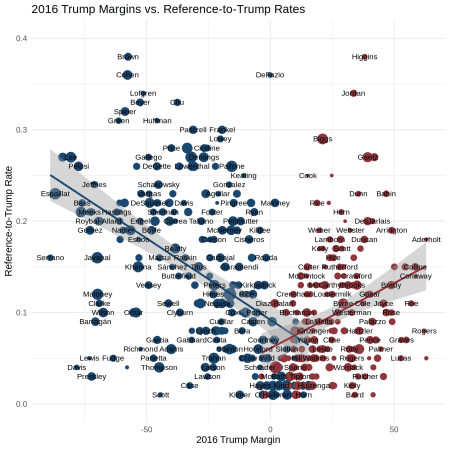
theme(legend.position = "none",

axis.title = element\_text())+

scale\_y\_continuous(limits = c(0,.4)) +

xlab('2016 Trump Margin') + ylab('Reference-to-Trump Rate') +

ggtitle('2016 Trump Margins vs. Reference-to-Trump Rates')



**Seemingly a no-brainer** if you don’t want to ruffle any feathers within an ideologically heterogeneous constituency, and if you want to fly under 45’s Twitter-radar. On the other hand, House Reps in safer (ie, ideologically more uniform) districts (especially Dems) are more likely to comment (or sound-off) on the doings of 45.

**Summary**

So, a couple of novel metrics for investigating variation with respect to the how & how often of 45-reference on Twitter in the 116th House. Simple methods (that could certainly be tightened up some) & intuitive results that align quite well with with linguistic/stance theory. Also some super interesting & robust relationships based in two very disparately-sourced data sets: 2016 Trump margins and Twitter text data (ca, present day).

The predictive utility of 2016 presidential voting margins seems (roughly) limitless. As does the cache of socio-political treasure hidden in the tweets of US lawmakers – for better or worse. A fully reproducible post. Cheers.

**References**

Berg, Esther van den, Katharina Korfhage, Josef Ruppenhofer, Michael Wiegand, and Katja Markert. 2019. “Not My President: How Names and Titles Frame Political Figures.” In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, 1–6.

Du Bois, John W. 2007. “The Stance Triangle.” *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction* 164 (3): 139–82.