

# NLP Project

Ndubuisi

2023-02-09

## R Markdown

```
library(readr)
library(sparklyr)
```

```
##
## Attaching package: 'sparklyr'
```

```
## The following object is masked from 'package:stats':
##
##   filter
```

```
library('tidyverse')
```

```
## — Attaching packages
## —————
## tidyverse 1.3.2 —
```

```
## ✓ ggplot2 3.3.6      ✓ dplyr 1.0.10
## ✓ tibble 3.1.8       ✓ stringr 1.4.1
## ✓ tidyr 1.2.1        ✓ forcats 0.5.2
## ✓ purrr 0.3.4
## — Conflicts ————— tidyverse_conflicts() —
## * dplyr::filter() masks sparklyr::filter(), stats::filter()
## * purrr::invoke() masks sparklyr::invoke()
## * dplyr::lag()     masks stats::lag()
```

```
library('rvest')
```

```
##
## Attaching package: 'rvest'
##
## The following object is masked from 'package:readr':
##
##   guess_encoding
```

```
library('lubridate')
```

```
##  
## Attaching package: 'lubridate'  
##  
## The following objects are masked from 'package:base':  
##  
##    date, intersect, setdiff, union
```

```
library('dplyr')  
library('ggplot2')  
library('tidytext')  
library('stringr')  
library(tm)
```

```
## Loading required package: NLP  
##  
## Attaching package: 'NLP'  
##  
## The following object is masked from 'package:ggplot2':  
##  
##    annotate
```

```
library(stringr)  
library(igraph)
```

```
##
## Attaching package: 'igraph'
##
## The following objects are masked from 'package:lubridate':
##
##    %--%, union
##
## The following objects are masked from 'package:dplyr':
##
##    as_data_frame, groups, union
##
## The following objects are masked from 'package:purrr':
##
##    compose, simplify
##
## The following object is masked from 'package:tidyr':
##
##    crossing
##
## The following object is masked from 'package:tibble':
##
##    as_data_frame
##
## The following object is masked from 'package:sparklyr':
##
##    %->%
##
## The following objects are masked from 'package:stats':
##
##    decompose, spectrum
##
## The following object is masked from 'package:base':
##
##    union
```

```
library(ggraph)
library(widyr)
library(janeaustenr)
```

## ##SCRAPPING DATA FROM REDDIT POLITICS PAGE

```
#Get post title
reddit_page<- read_html("https://www.reddit.com/r/politics/comments/10wgssm/discussion_t
hread_2023_state_of_the_union/")
reddit_page %>%
  html_node("._eYtD2XCvieq6emjKBH3m") %>%
  html_text()
```

```
## [1] "Discussion Thread: 2023 State of the Union"
```

```
#Get post comments  
reddit_page %>%  
  html_nodes("._1qeIAgB0cPwnLhDF9XSiJM") %>%  
  html_text()
```

## [1] "At 9 p.m. Eastern today, President Joe Biden is expected to address a joint session of Congress to deliver this year's State of the Union address."

## [2] "Selected Reporting in Advance of the SOTU:"

## [3] "AP News: State of the Union: Biden sees economic glow, GOP sees gloom (Analysis of the likely themes going into this year's State of the Union.)"

## [4] "NPR: As the nation has changed, so has the State of the Union speech (A look at the history of the State of the Union, from the Federalists to the present.)"

## [5] "AP News: Biden, Cabinet visiting 20 states after State of the Union (News.)"

## [6] "The Washington Post (metered paywall): Biden's State of the Union address to take aim at Silicon Valley | As critical legislation has stalled in Congress, the president plans to call on Republicans and Democrats to work together to regulate the industry (News.)"

## [7] "NBC: Biden's State of the Union address will make his case for re-election in 2024 | Biden will use the speech to reach more people and assure them that he's enacted plans to make their daily commutes shorter and their prescription drug bills lower (Analysis.)"

## [8] "The New York Times (metered paywall): For Biden, a Chance for a Fresh Start in a New Era of Divided Government | The president plans to use his first State of the Union address since Republicans took control of the House to call for bipartisan cooperation. Neither he nor many others expect that to happen. (Analysis.)"

## [9] "NPR: The State of the Union will showcase the state of the Biden-McCarthy relationship (Analysis.)"

## [10] "Where to Watch the Address:"

## [11] "C-SPAN: State of the Union Address & Republican Response (C-SPAN's stream begins at 8 p.m. Eastern)"

## [12] "PBS on YouTube: President Joe Biden's 2023 State of the Union address (PBS News hour special coverage begins at 6 p.m. Eastern)."

## [13] "The Washington Post on YouTube: Biden delivers 2023 State of the Union address (The Post's coverage begins at 8 p.m. Eastern)"

## [14] "The White House on YouTube: President Biden's State of the Union Address (Stream begins at 9 p.m. Eastern)"

## [15] "Responses to the SOTU:"

## [16] "This year there will be three major responses to Biden's speech. Two will be delivered by Republicans, one in English by Arkansas governor Sarah Huckabee Sanders and one in Spanish by first-term Representative Juan Ciscomani. The third will be delivered by progressive Democrat and first-term Representative Delia Ramirez on behalf of the Working Families Party. There will be a separate discussion thread for these responses, a link to which will be added here once available."

## [17] "Edit: Discussion Thread: Responses to the 2023 State of the Union"

## [18] "To sort this thread by 'newest comments first', click or tap here."

## [19] "To sort this thread by 'best comments first', click or tap here."

## [20] "Please vote on comments based on their contribution to the thread, and please remember that all subreddit rules remain in effect: stay civil, don't troll, no hateful speech, don't in any way support harm, don't spam, and so on."

## [21] "\"If anyone tries to cut Social Security, I will veto it... but I guess that's something no one's gonna try to do.\""

## [22] "Love seeing republican fury when they're called out on their political games"

## [23] "\"How dare he make us say it with our chest!\""

## [24] "Allowing the \"heckles\" to have a semi-back-and-forth only to back them into a \"good so then we all agree\" corner was beautifully done."

## [25] "Genuinely impressive, and they walked right into it."

## [26] "And he's funny enough to even get McCarthy to laugh when the dude is trying his

```

best to zone out. He's nailing it."
## [27] "'Speaker I don't want to ruin your reputation but I look forward to working with you.'"
## [28] "That was a good one lol"
## [29] "A Dem simply saying that they will work with a specific Republican is probably enough for that Republican to lose supporters. How far their party has fallen..."
## [30] "Using a SOTU to make a live power play is a new one."
## [31] "That was masterful"
## [32] "GOP shouting \"liar\" got a tad predictable it seems. They played themselves."
## [33] "Jesus Christ I just watched Joe Biden goad the republican party into defending social security and Medicaid"
## [34] "That was legit badass. LBJ style politicking live and in real time."
## [35] "Wow, did he just completely fuck their proposals to cut Medicare and Social Security in real time in front of the entire country?"
## [36] "He fully Regina Georged them."
## [37] "'So you agree? We keep Medicare and Social Security?'"

```

```

#Reddit Politics New posts webpage
reddit_pol_new <- read_html("https://www.reddit.com/r/politics/new/")

```

```

#Get time of comments
time <- reddit_pol_new %>%
  html_nodes("._2VF2J19pUIMSLJFky-7PEI") %>%
  html_text()
time

```

```

## [1] "18 minutes ago" "22 minutes ago" "34 minutes ago" "45 minutes ago"
## [5] "52 minutes ago" "55 minutes ago"

```

```

#Get URL of new posts
url<- reddit_pol_new %>%
  html_nodes("a.SQnoC3ObvgnGjWt90zD9Z._2INHsNB8V5eaWp4P0rY_mE") %>%
  html_attr("href")
url <- paste0("https://reddit.com",url)
url

```

```
## [1] "https://reddit.com/r/politics/comments/10xvcu1/bill_would_ban_diversity_equity_inclusion_at/"
## [2] "https://reddit.com/r/politics/comments/10xv9ph/supreme_court_might_have_easy_outs_on_elections/"
## [3] "https://reddit.com/r/politics/comments/10xv0i1/house_oversight_panel_requests_records_from/"
## [4] "https://reddit.com/r/politics/comments/10xurpr/rep_george_santos_work_for_firm_accused_of_ponzi/"
## [5] "https://reddit.com/r/politics/comments/10xunft/missouri_house_votes_against_limits_on_kids/"
## [6] "https://reddit.com/r/politics/comments/10xukw7/court_could_ban_abortion_pills_nationwide/"
```

```
#create a dataframe
reddit_pol_newdata <- data.frame(NewsPage=url, PublishedTime=time)
#Check the dimensions
dim(reddit_pol_newdata )
```

```
## [1] 6 2
```

```
# Filter dataframe by rows that contain a time published in minutes (i.e. within the hour)
reddit_pol_newdata2 <- reddit_pol_newdata[grepl("minutes|now", reddit_pol_newdata$PublishedTime),]
#Check the dimensions
dim(reddit_pol_newdata2)
```

```
## [1] 6 2
```

```
#create empty lists for titles and comments
titles <- c()
comments <- c()
```

```
# use a for loop to run through each page and gather post title and comments
for(i in reddit_pol_newdata2$NewsPage){

  reddit_pol_newdata2 <- read_html(i)
  body <- reddit_pol_newdata2 %>%
    html_nodes("._lqeIAgB0cPwnLhDF9XSiJM") %>%
    html_text()
  comments = append(comments, body)

  reddit_pol_newdata2 <- read_html(i)
  pname <- reddit_pol_newdata2 %>%
    html_node("._eYtD2XCVieq6emjKBH3m") %>%
    html_text()
  titles = append(titles, rep(pname,each=length(body)))

}
```

```
#create a dataframe
reddit_pol_dataframe <- data.frame(Headline=titles, Comments=comments)
dim(reddit_pol_dataframe)
```

```
## [1] 139 2
```

```
reddit_pol_dataframe
```



## Headline

[illegible]

## 104 Missouri House votes against limits on kids carrying guns  
## 105 Missouri House votes against limits on kids carrying guns  
## 106 Missouri House votes against limits on kids carrying guns  
## 107 Missouri House votes against limits on kids carrying guns  
## 108 Missouri House votes against limits on kids carrying guns  
## 109 Missouri House votes against limits on kids carrying guns  
## 110 Missouri House votes against limits on kids carrying guns  
## 111 Missouri House votes against limits on kids carrying guns  
## 112 Missouri House votes against limits on kids carrying guns  
## 113 Missouri House votes against limits on kids carrying guns  
## 114 Missouri House votes against limits on kids carrying guns  
## 115 Missouri House votes against limits on kids carrying guns  
## 116 Missouri House votes against limits on kids carrying guns  
## 117 Missouri House votes against limits on kids carrying guns  
## 118 Missouri House votes against limits on kids carrying guns  
## 119 Missouri House votes against limits on kids carrying guns  
## 120 Missouri House votes against limits on kids carrying guns  
## 121 Missouri House votes against limits on kids carrying guns  
## 122 Court could ban abortion pills nationwide  
## 123 Court could ban abortion pills nationwide  
## 124 Court could ban abortion pills nationwide  
## 125 Court could ban abortion pills nationwide  
## 126 Court could ban abortion pills nationwide  
## 127 Court could ban abortion pills nationwide  
## 128 Court could ban abortion pills nationwide  
## 129 Court could ban abortion pills nationwide  
## 130 Court could ban abortion pills nationwide  
## 131 Court could ban abortion pills nationwide  
## 132 Court could ban abortion pills nationwide  
## 133 Court could ban abortion pills nationwide  
## 134 Court could ban abortion pills nationwide  
## 135 Court could ban abortion pills nationwide  
## 136 Court could ban abortion pills nationwide  
## 137 Court could ban abortion pills nationwide  
## 138 Court could ban abortion pills nationwide  
## 139 Court could ban abortion pills nationwide  
##

#### Comments

## 1

As a reminder, this subreddit is for civil discussion.

## 2

In general, be courteous to others. Debate/discuss/argue the merits of ideas, don't attack people. Personal insults, shill or troll accusations, hate speech, any suggestion or support of harm, violence, or death, and other rule violations can result in a permanent ban.

## 3

If you see comments in violation of our rules, please report them.

## 4

For those who have questions regarding any media outlets being posted on this subreddit, please click here to review our details as to our approved domains list and outlet criteria.

## 5

Special announcement:

## 6

r/politics is currently accepting new moderator applications. If you want to help make this community a better place, consider applying here today!

## 7

I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.

## 8

Someone should write to businesses detailing that doing Business in Texas reflect negatively on their business image because they actively oppose these values.

## 9

Texas is continually baffling for me but this will be an interesting test case for DEI and structural racism.

## 10

This title is confusing. The article claims that they want to use merit as a means of admission.

## 11

Merit has always been the absolute deciding factor on admissions. But a hell of a lot more people always qualify than actually get admitted. So there are rules in place to make sure that it is not just rich white kids who get a chance at an education.

## 12

White supremacists keep throwing out these stories about totally unqualified black people getting things because of affirmative action, and those stories are nothing but bald face lies being used to justify discrimination.

## 13

As a reminder, this subreddit is for civil discussion.

## 14

In general, be courteous to others. Debate/discuss/argue the merits of ideas, don't attack people. Personal insults, shill or troll accusations, hate speech, any suggestion or support of harm, violence, or death, and other rule violations can result in a permanent ban.

## 15

If you see comments in violation of our rules, please report them.

## 16

For those who have questions regarding any media outlets being posted on this subreddit, please click here to review our details as to our approved domains list and outlet criteria.

## 17

Special announcement:

## 18

r/politics is currently accepting new moderator applications. If you want to help make this community a better place, consider applying here today!

## 19

I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.

## 20

What a bunch of clowns.

## 21

Looks like those clowns in Congress did it again!

## 22

Republicans have already established that you don't need to comply with congressional subpoenas, so I don't see why the Bidens would be compelled to turn anything over.

## 23

Exactly this. The speaker of the house ignored a congressional subpoena.

## 24

Zero congressional subpoenas should be honored until the GOP who ignored them are removed from positions of power.

## 25

It's the year 2026, and Republicans launch a new investigation into why Hunter Biden pops his collar and whether it's a communist code.

## 26

It's 2027 and the investigation into Hunter Biden's collar has revealed yet another instance of TFG selling classified American secrets.

## 27

During every congressional investigation/hearing into Trump's alleged crimes, every Republican complained about wasting time instead of working to fix the country's problems...

## 28

Now that they're in charge, their plans for the next 2 years include nothing but "boogey man issues" and culture war bullshit.

## 29

It's a feature, not a bug.

## 30

They have turned democracy into the WWE. The rubes in the Red States want big drama.

## 31

vague anti china language with no evidence or facts.

## 32

## 33

literal witch hunt.

## 34

This should definitely help with inflation.

## 35

They should turn over exactly the same number of documents released by the last administration about (Federal Employees) Ivanka & Jared's foreign engagements.

## 36

What office does Hunter and James hold again?

## 37

Remind me please.

## 38

I dunno, but dontcha wanna find out why the republicans are so interested in seeing his peen?

## 39

Says a lot that I couldn't even tell you who James is. Not even sure he is real. I hate this both ways BS, because Trump literally employed his children at the white house.

## 40

Do we get to see James' penis as well?

## 41

Here's a preview of our article by CBS News senior investigative correspondent Catherine Herridge:

## 42

Rep. James Comer of Kentucky, the Republican chairman of the House Oversight and Accountability Committee, has made the panel's first official requests for documents from President Biden's son, Hunter Biden, and his brother, James Biden, over foreign business pursuits.

## 43

In letters first obtained by CBS News, Comer informed Hunter and James Biden, as well as their then-business manager, that he is seeking documents and communications from the Bidens as part of his committee's probe into to any possible involvement by the president in their financial conduct, in particular in foreign business deals "with individuals who were connected to the Chinese Communist Party." Comer accused them in his letter of receiving "significant amounts of money from foreign companies without providing any known legitimate services."

## 44

In addition to that probe, Comer said the committee is considering writing a bill to strengthen federal ethics laws around public officials and their families.

## 45

Read more: <https://www.cbsnews.com/news/hunter-biden-james-biden-house-oversight-committee-james-comer/>

## 46

In addition to that probe, Comer said the committee is considering writing a bill to strengthen federal ethics laws around public officials and their families.

## 47

I imagine Congresscritters will strangely wind up exempt from this.

## 48

Rep. James Comer of Kentucky

## 49

he looks like Newt Khrushchev Gingrich

## 50

Someone should come up with a cost chart of these wasteful committees wasting our tax dollars.

## 51

After all these damning house investigations, Hunter Biden will never be able to hold office again. /s

## 52

As a reminder, this subreddit is for civil discussion.

## 53

In general, be courteous to others. Debate/discuss/argue the merits of ideas, don't attack people. Personal insults, shill or troll accusations, hate speech, any suggestion or support of harm, violence, or death, and other rule violations can result in a permanent ban.

## 54

If you see comments in violation of our rules, please report them.

## 55

For those who have questions regarding any media outlets being posted on this subreddit, please click here to review our details as to our approved domains list and outlet criteria.

## 56

Special announcement:

## 57

r/politics is currently accepting new moderator applications. If you want to help make this community a better place, consider applying here today!

## 58

I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.

## 59

Ship him and Bolsonaro back to Brazil.

## 60

Holy cow, just throw the guy out already.

## 61

Hell, if this was a private business, he'd be fired without recourse.

## 62

Good Lord, how can somebody be such a bad person...

## 63

Good job voting for this guy

## 64

Here's a preview of our story by reporters Caitlin Huey-Burns and Graham Kates:

## 65

Embattled GOP Rep. George Santos' work for a company that allegedly orchestrated a Ponzi scheme attracted scrutiny from federal regulators, according to a witness who has fielded questions from investigators.

## 66

The freshman New York lawmaker, whose litany of false and misleading claims about his past have left him with few allies, even among his fellow Republicans, previously worked for an investment firm known as the Harbor City Capital Corporation. The company was shut down by the Securities and Exchange Commission (SEC), which in 2021 called it a "classic Ponzi scheme."

## 67

Santos has said he wasn't aware of any wrongdoing at Harbor City, where he began working in 2020. He was not named in the SEC's complaint against the firm, but a childhood friend said federal investigators called her after she spoke publicly about witnessing Santos' investment pitch.

## 68

In an interview for the CBS Reports documentary "Campaign of Deceit," Tiffany Bogosian described how she bumped into Santos, her old junior high classmate, at a Starbucks in 2020. As they caught up, she mentioned that she'd become an attorney and had recently helped a client reach a \$2 million settlement.

## 69

Santos, then a regional director for Harbor City, promptly invited her and her client to dinner, Bogosian said in "Campaign of Deceit," which was released Thursday. She said when they arrived for the meal at Il Bacco, an Italian restaurant in Queens, she was brought into a "huge room" with "one table in the middle."

## 70

Read more: <https://www.cbsnews.com/news/george-santos-harbor-city-capital-alleged-ponzi-scheme-sec/>

## 71

This guy sounds like the perfect candidate for the Republican Presidential nominee. He just needs to beef up the culture war resume a bit.

## 72

This was known already.

## 73

As a reminder, this subreddit is for civil discussion.

## 74

In general, be courteous to others. Debate/discuss/argue the merits of ideas, don't attack people. Personal insults, shill or troll accusations, hate speech, any suggestion or support of harm, violence, or death, and other rule violations can result in a permanent ban.

## 75

If you see comments in violation of our rules, please report them.

## 76

For those who have questions regarding any media outlets being posted on this subreddit, please click here to review our details as to our approved domains list and outlet criteria.

## 77

Special announcement:

## 78

r/politics is currently accepting new moderator applications. If you want to help make this community a better place, consider applying here today!

## 79

I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.

## 80

Lets get rid of education, raise poverty levels by banning abortion, and increase the number of guns on the street...

## 81

Nothing could possibly go wrong.

## 82

Let's push for these policies in Red states explicitly.

## 83

"Don't want to go to school, you don't have to!"

## 84

Remove any pretense of needing an education in a red state

## 85

"you're naturally gifted, schooling is for meddlin' folk. Ditches always need dug."

## 86

Encourage them to smoke at age 10, drink at 15. Let 'em carry guns all the time and constantly encourage them to solve problems with their fists. Eventually, the problem will sort itself out.

## 87

/s -- I want none of these things to come to pass, but encouraging chaos in red states might be the right play. In twenty years, the blue states will own their land and make them work it, just like the feudal system.

## 88

They have prisons to fill and companies that want slave labor. They know their kids will go to private school.

## 89

Dumb poor people = cheap labor.

## 90

They know what they're doing.

## 91

How is it a restriction on gun rights when children don't have rights to guns in the first place?

## 92

Even without adult supervision is the thing that really fucking rubs salt in the wound.

## 93

This bill would not have stopped reasonable gun owners from taking older children own guns with them or anything like that.

## 94

Public schools are state institutions. Does that mean Missouri is fine with allowing kids to bring guns to school? I mean, there's no NEED for a kid to do it, but since Missouri is not in the habit of prosecuting people for what they "might" do, restricting guns in schools is technically no different than what they voted for.



## 95

Republicans are a death cult made up of ghouls and morons.

## 96

Theres a very good reason this states nickname is "Misery".

## 97

Theyve earned it on every concievable level.

## 98 Missouri's GOP-majority House on Wednesday voted against introducing a ban on children carrying guns without adult supervision in public. Just one Republican supported the proposal that was rejected in a 104-39 vote. Democratic Rep. Donna Baringer said authorities in her district wanted the ban so they could stop "14-year-olds walking down the middle of the street in the city of St. Louis carrying AR-15s." She added that police were powerless to do anything until the teens "actually brandish them, and brandish them with intent." Conservative lawmakers saw the proposal as an unwarranted restriction on gun rights. "While it may be intuitive that a 14-year-old has no legitimate purpose, it doesn't actually mean that they're going to harm someone. We don't know that yet," Republican Rep. Tony Lovasco said. "Generally speaking, we don't charge people with crimes because we think they're going to hurt someone."

## 99

So this is an example of the police making a request of lawmakers that is actually reasonable. And republicans respond by laughing in their fucking faces.

## 100

The police are not allowed to do anything until the child actually brandish is the weapon, and of course these are police we are talking about so at that point they just shoot the kid and anyone who happens to be anywhere near him.

## 101

Remove all speed limits, seatbelts, laws against driving on the sidewalk or the wrong side of the street, road signs, and drivers license requirements.

## 102

While these road safety measures might be intuitive, it doesn't mean a driver is going to harm someone. We don't know that yet.

## 103

Generally speaking, we don't issue citations because we think a pedestrian is going to get run over by someone doing 85 in a residential area.

## 104

Yeah they will definitely use that logic when they introduce bill to stop women from leaving the state to get abortion.

## 105

So, you end up with a scenario where a possible kid school shooter starts walking towards a school with an AR and the police have to ask what the kid's intent is?

## 106

No... the cop tell this Kid to drop the Gun and the kid is confused and frightened, makes a sudden move and the kid gets lit up.

## 107

This is what they want to continue?

## 108

From the article I've linked:

## 109

Of those taken into custody, a 15-year-old male and a 16-year-old male had guns in their possession. The 15-year-old was found in possession of a Glock 29 equipped with an auto sear, making the gun fully automatic, and the 16-year-old was in possession of a stolen Glock 19, police said.

## 110

Teens released to parents after shots fired at St. Louis officers

## 111

Florida GOP: "Hypothetically, educators may speak to children about homosexuality, and that simply mustn't be allowed under penalty of the law!"

## 112

Missouri GOP: "We cannot allow officers to intervene, should an unsupervised minor be in possession of a firearm in public. We mustn't act upon hypothetical situations in which the child may harm themselves or others, that would be unreasonable!"

## 113

That's your fucking republicans again. Turning our kids into child soldiers.

## 114

At what point is this not terrorism

## 115

It is absolutely terrorism, and once again it is the exact same type of terrorism the nazis used. Indoctrinating and arming children was one of their 1st fucking steps.

## 116

Recent grads, high school students, heck, 6 year olds, to them, this is not a problem.

## 117

Maybe they're the problem.....All of Them

## 118

Holy Republican's Batman

## 119

Tell those morons to come and hang out in east Kansas City MO some night.

## 120

I've given up on Missouri. If they don't care about their own families, then what can anybody else do?

## 121

Total shithole state.

## 122

As a reminder, this subreddit is for civil discussion.

## 123

In general, be courteous to others. Debate/discuss/argue the merits of ideas, don't attack people. Personal insults, shill or troll accusations, hate speech, any suggestion or support of harm, violence, or death, and other rule violations can result in a permanent ban.

## 124

If you see comments in violation of our rules, please report them.

## 125

For those who have questions regarding any media outlets being posted on this subreddit, please click here to review our details as to our approved domains list and outlet criteria.

## 126

Special announcement:

## 127

r/politics is currently accepting new moderator applications. If you want to help make this community a better place, consider applying here today!

## 128

I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.

## 129

Sweet let's have crooked judges make nonevidence based medical decisions for our country, while other modern countries scratch their heads at exactly how stupid we all are. 3r

d world country here we come!

## 130

District Court Judge Matthew Kacsmaryk in Amarillo, Texas, could either strike down the FDA's decades-old decision to approve mifepristone – the first of two pills used to end a pregnancy – or roll back more recent agency decisions making the pills available via telemedicine, mail delivery and pharmacy pickup.

## 131

He could give this order and it would immediately be stayed and overturned upon appeal.

A district court judge absolutely does not have the authority to counteract the FDA's regulatory decisions to approve drugs.

## 132

The 5th Circuit is extremely right wing, as is the SCOTUS. This could absolutely go through, regardless of what's right. If they decided the EPA doesn't have the right to protect our air by setting emissions standards, I could see them deciding the FDA shouldn't have the current authority it does.

## 133

Brett Kavanaugh- "hold my beer"

## 134

If you or anyone you care about is a woman of childbearing age, stock up on this lifesaving medicine right now.

## 135

It's a prescription, you can't just buy them over the counter.

## 136

The United States is on a steep steep decline. Not that long ago it was actually cool to live here

## 137

More Trump appointed judges deciding that the people are using their freedom wrong.

## 138

Drug dealers get a new product to push.

## 139

Cool. Guess it's time to gather a network of women and go underground again.

```
#remove bot messages from the comments
bot <- c(
  "As a reminder, this subreddit is for civil discussion.",
  "In general, be courteous to others. Debate/discuss/argue the merits of ideas, don't attack people. Personal insults, skill or troll accusations, hate speech, any suggestion or support of harm, violence, or death, and other rule violations can result in a permanent ban.",
  "If you see comments in violation of our rules, please report them.",
  "For those who have questions regarding any media outlets being posted on this subreddit, please click here to review our details as to our approved domains list and outlet criteria.",
  "I am a bot, and this action was performed automatically. Please contact the moderators of this subreddit if you have any questions or concerns.",
  "Special announcement:",
  "r/politics is currently accepting new moderator applications. If you want to help make this community a better place, consider applying here today!",
  "Free version on yahoo; https://www.yahoo.com/news/florida-parents-push-government-int-rusion-000143380.html"
)
```

```
reddit_pol_dataframe_no_bot <- subset(
  reddit_pol_dataframe, !(Comments %in% c(bot))
)

dim(reddit_pol_dataframe_no_bot)
```

```
## [1] 104 2
```

```
head(reddit_pol_dataframe_no_bot$Comments)
```

```
## [1] "Someone should write to businesses detailing that doing Business in Texas reflects negatively on their business image because they actively oppose these values."
## [2] "Texas is continually baffling for me but this will be an interesting test case for DEI and structural racism."
## [3] "This title is confusing. The article claims that they want to use merit as a means of admission."
## [4] "Merit has always been the absolute deciding factor on admissions. But a hell of a lot more people always qualify than actually get admitted. So there are rules in place to make sure that it is not just rich white kids who get a chance at an education."
## [5] "White supremacists keep throwing out these stories about totally unqualified black people getting things because of affirmative action, and those stories are nothing but bald face lies being used to justify discrimination."
## [6] "What a bunch of clowns."
```

```
#save dataframe in csv
write.table(reddit_pol_dataframe_no_bot,file = 'reddit_pol_data.csv', append = TRUE, sep
= ",", col.names = FALSE, row.names = FALSE)
```

# WORD COUNT

```
# read dataframe from csv file and convert to table
bp_data<- read.csv(file='reddit_pol_data.csv', stringsAsFactors=FALSE)

cleaned.bp <- as_tibble(bp_data)
```

```
#remove white spaces and unnest comments
cleaned.bp$Comments <- stripWhitespace(cleaned.bp$Comments)
cleaned.bp <- cleaned.bp %>%
  unnest_tokens(word, Comments)
```

```
#remove stop words
cleaned.bp %>%
  count(word, sort = TRUE)
```

```
## # A tibble: 3,551 × 2
##   word      n
##   <chr> <int>
## 1 the    723
## 2 to     427
## 3 a      354
## 4 of     341
## 5 and    327
## 6 in     243
## 7 that   215
## 8 is     201
## 9 i      152
## 10 this  129
## # ... with 3,541 more rows
```

```
data(stop_words)
cleaned.bp <- cleaned.bp %>%
  anti_join(stop_words)
```

```
## Joining, by = "word"
```

```
#remove numbers and unique characters
cleaned.bp %>%
  count(word, sort = TRUE)
```

```
## # A tibble: 3,053 × 2
##   word          n
##   <chr>      <int>
## 1 biden        67
## 2 republicans  64
## 3 trump        52
## 4 people       44
## 5 gop          42
## 6 2019         32
## 7 hunter       32
## 8 news         29
## 9 republican   29
## 10 laptop      26
## # ... with 3,043 more rows
```

```
nums <- cleaned.bp %>% filter(str_detect(word, "[0-9]")) %>% select(word) %>% unique()
cleaned.bp <- cleaned.bp %>%
  anti_join(nums, by = "word")
```

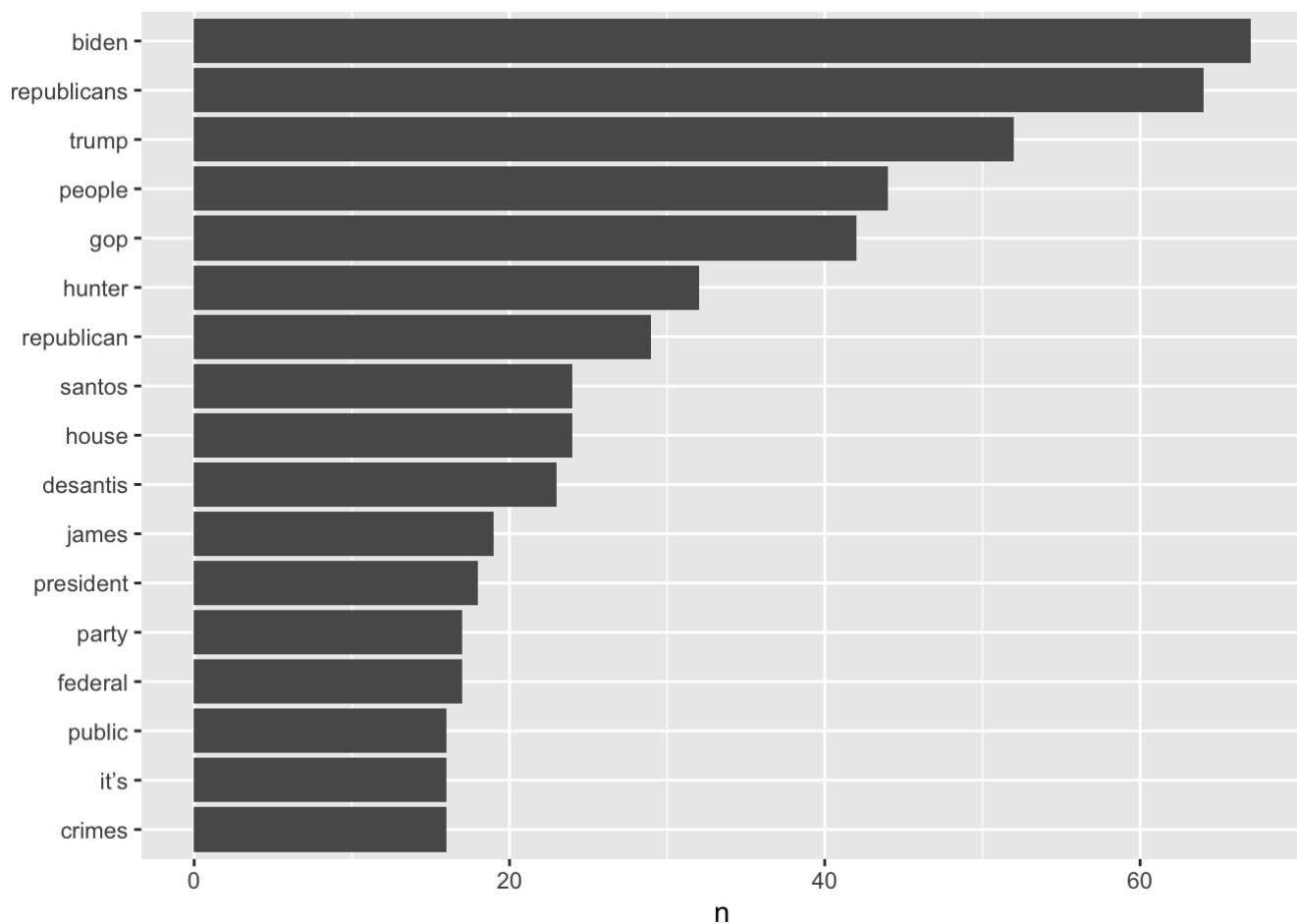
*#remove words with contextual or semantic meaning*

```
uni_sw <- data.frame(word = c("lol", "it's", "i'm", "subreddit", "moderators", "article", "news",
  "fuck", "laptop", "2019", "twitter", "time", "day", "it's", "shit", "201c", "201d"))

cleaned.bp <- cleaned.bp %>%
  anti_join(uni_sw, by = "word")
```

*#plot a graph with words that appear over 15 times*

```
cleaned.bp %>%
  count(word, sort = TRUE) %>%
  filter(n > 15) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()
```



## #SENTIMENT ANALYSIS

```
library(sentimentr)
```

```
#check for the sentiment score of words contained in the comments
bp_data$Comments <- as.character(bp_data$Comments)
sentiment_scores <- sentiment(bp_data$Comments)
head(sentiment_scores)
```

```
##      element_id sentence_id word_count  sentiment
## 1:           1           1          6 -0.4082483
## 2:           2           1          4  0.0000000
## 3:           2           2          4  0.0000000
## 4:           3           1          6  0.0000000
## 5:           3           2          5  0.0000000
## 6:           3           3          6  0.0000000
```

```
average_sentiment_score <- sum(sentiment_scores$sentiment)/length(sentiment_scores$sentiment)
average_sentiment_score
```

```
## [1] -0.02627684
```

*#The Average sentiment is -0.0278399 that means that the reaction of reddit users 24hours after president Biden's state of union address were negative.*

# CORELATION ANALYSIS

```
library(tm)
```

```
# remove carriage returns and new lines from text
cleaned.bp2<- as_tibble(bp_data)
cleaned.bp2$Comments <- gsub("\r?\n|\r", " ", cleaned.bp2$Comments)
```

```
# remove punctuation
cleaned.bp2$Comments <- gsub("[[:punct:]]", "", cleaned.bp2$Comments)
```

```
# force entire corpus to lowercase
cleaned.bp2$Comments <- tolower(cleaned.bp2$Comments)
```

```
#remove numbers from text
cleaned.bp2$Comments<- removeNumbers(cleaned.bp2$Comments)
```

```
# remove stop words
cleaned.bp2$Comments <- removeWords(cleaned.bp2$Comments, stopwords("SMART"))
```

```
# remove additional words
other.words <- c("lol","it's","i'm","subreddit", "moderators", "article","news", "fuck",
"laptop","2019","twitter","time","day", "it's", "shit", "201c", "201d")
cleaned.bp2$Comments <- removeWords(cleaned.bp2$Comments, other.words)
```

```
# removes additional remaining whitespace
cleaned.bp2$Comments <- stripWhitespace(cleaned.bp2$Comments)
```

```
# transform table into one-word-per-line tidytext format
clean.data <- cleaned.bp2 %>%
  unnest_tokens(word, Comments)
```

```
# most frequent words across corpus
clean.data <- clean.data %>%
  count(word, sort = TRUE)
clean.data
```



```
## # A tibble: 3,085 × 2
##   word          n
##   <chr>      <int>
## 1 republicans 66
## 2 biden       62
## 3 trump       50
## 4 people      44
## 5 gop         41
## 6 dont        35
## 7 hunter      29
## 8 im          28
## 9 republican  28
## 10 state      28
## # ... with 3,075 more rows
```

```
# preprocessing for tf-idf
posts.words <- cleaned.bp2 %>%
  unnest_tokens(word, Comments) %>%
  count(Headline, word, sort = TRUE)
total.words <- posts.words %>%
  group_by(Headline) %>%
  summarize(total = sum(n))

posts.words <- left_join(posts.words, total.words)
```

```
## Joining, by = "Headline"
```

```
# tf-idf
posts.words <- posts.words %>%
  bind_tf_idf(word, Headline, n)

posts.words <- posts.words %>%
  select(-total) %>%
  arrange(desc(tf_idf))
```

```
# compare tf-idf to gross frequency
unique(posts.words$word[1:30])
```

```
## [1] "canut"      "eat"        "cruel"      "travesty"
## [5] "wait"       "freedom"    "yinz"       "ha"
## [9] "nice"       "circles"    "extremist"  "unmasks"
## [13] "conservatives" "deny"      "bullying"   "idaho"
## [17] "whats"      "bad"        "strong"     "based"
## [21] "bully"      "disgusting" "endorse"    "work"
## [25] "monkeys"    "politicians" "submission" "recount"
## [29] "infighting" "thie"
```

```
clean.data$word[1:30]
```

```
## [1] "republicans" "biden"      "trump"      "people"     "gop"
## [6] "dont"         "hunter"     "im"         "republican" "state"
## [11] "desantis"     "house"      "santos"     "make"       "years"
## [16] "federal"      "good"       "president"  "trumps"     "crimes"
## [21] "hes"          "party"      "public"     "back"       "congress"
## [26] "james"        "states"     "doesnt"     "evidence"   "give"
```

```
posts.bigrams <- cleaned.bp2 %>%
  unnest_tokens(bigram, Comments, token = "ngrams", n = 2)
```

```
posts.bigrams %>%
  count(bigram, sort = TRUE)
```

```
## # A tibble: 5,232 × 2
##   bigram                n
##   <chr>                <int>
## 1 <NA>                  79
## 2 hunter biden         20
## 3 social security      12
## 4 evidence crimes      10
## 5 administration officials 9
## 6 dark brandon          9
## 7 demanded posts        9
## 8 employees trumps       9
## 9 officials employees     9
## 10 republicans routinely 9
## # ... with 5,222 more rows
```

```
bigrams.separated <- posts.bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ")
```

```
bigram.tf.idf <- posts.bigrams %>%
  count(Headline, bigram) %>%
  bind_tf_idf(bigram, Headline, n) %>%
  arrange(desc(tf_idf))
```

```
# explore
clean.data$word[1:30]
```

```
## [1] "republicans" "biden"      "trump"      "people"     "gop"
## [6] "dont"        "hunter"     "im"         "republican" "state"
## [11] "desantis"    "house"      "santos"     "make"       "years"
## [16] "federal"     "good"       "president"  "trumps"     "crimes"
## [21] "hes"         "party"      "public"     "back"       "congress"
## [26] "james"       "states"     "doesnt"     "evidence"   "give"
```

```
unique(posts.words$word[1:30])
```

```
## [1] "canut"      "eat"        "cruel"      "travesty"
## [5] "wait"       "freedom"    "yinz"       "ha"
## [9] "nice"       "circles"    "extremist"  "unmasks"
## [13] "conservatives" "deny"      "bullying"   "idaho"
## [17] "whats"      "bad"        "strong"     "based"
## [21] "bully"      "disgusting" "endorse"    "work"
## [25] "monkeys"    "politicians" "submission" "recount"
## [29] "infighting" "thie"
```

```
bigram.tf.idf$bigram[1:30]
```

```
## [1] "canut wait"      "cruel travesty"  "travesty freedom"
## [4] "nice work"       "work yinz"       "conservatives deny"
## [7] "deny whats"      "extremist circles" "whats extremist"
## [10] "bad gop"         "biden strong"    "strong unmasks"
## [13] "unmasks bad"    "bully politicians" "endorse bullying"
## [16] "gop disgusting" "politicians endorse" "country doesnt"
## [19] "doesnt fascist" "prevent recount"  "wow great"
## [22] "car mittens"     "gop infighting"  "lmao thie"
## [25] "roof car"        "strap roof"      "thie gop"
## [28] "eat eat"         "flavors choose"  "good eat"
```

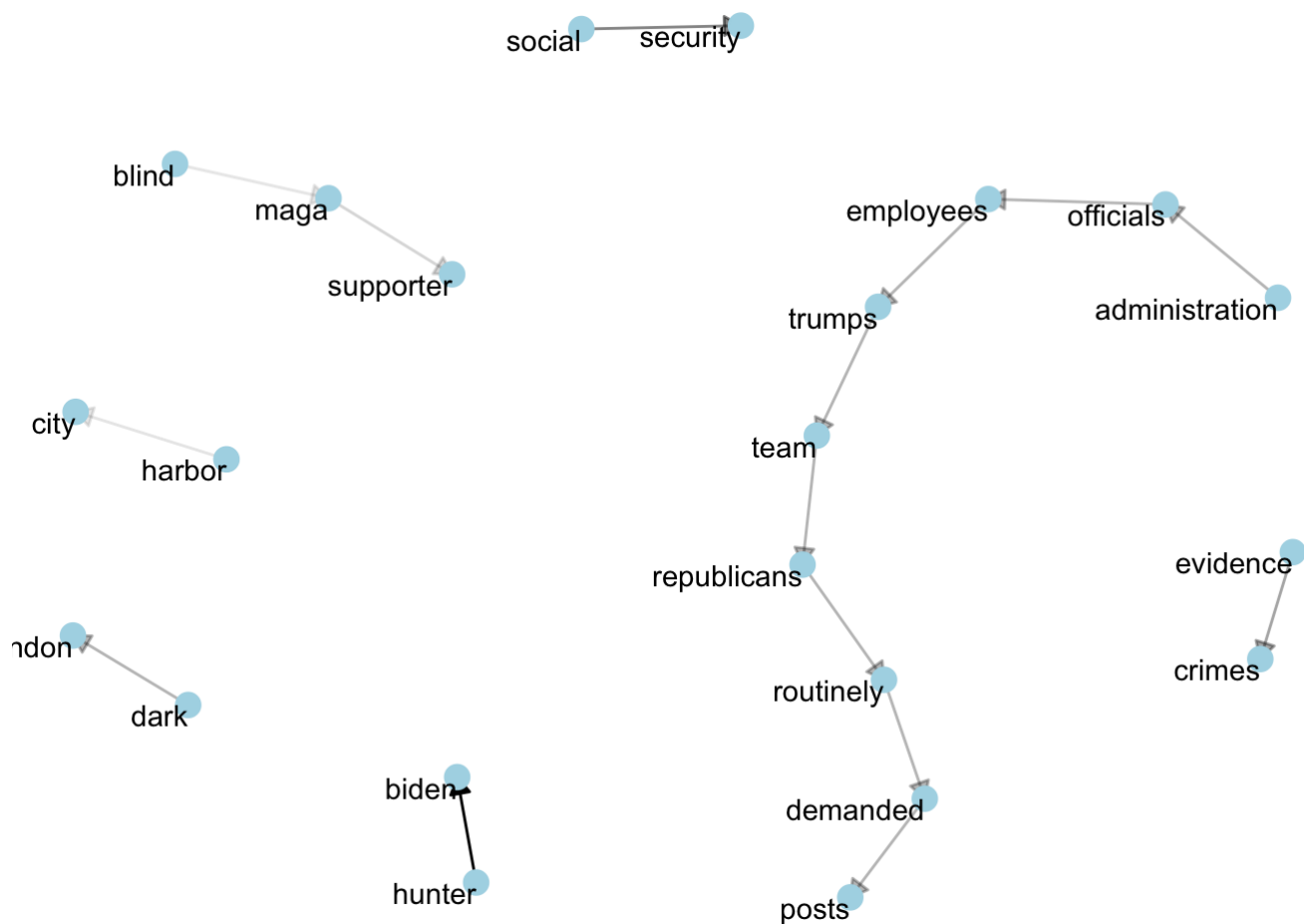
```
count_bigrams <- function(dataset) {
  dataset %>%
    unnest_tokens(bigram, Comments, token = "ngrams", n = 2) %>%
    separate(bigram, c("word1", "word2"), sep = " ") %>%
    filter(!word1 %in% stop_words$word,
           !word2 %in% stop_words$word) %>%
    count(word1, word2, sort = TRUE)
}
```

```
visualize_bigrams <- function(bigrams) {
  set.seed(2016)
  a <- grid::arrow(type = "closed", length = unit(.10, "inches"))

  bigrams %>%
    graph_from_data_frame() %>%
    ggraph(layout = "fr") +
    geom_edge_link(aes(edge_alpha = n), show.legend = FALSE, arrow = a) +
    geom_node_point(color = "lightblue", size = 4) +
    geom_node_text(aes(label = name), vjust = 1, hjust = 1) +
    theme_void()
}
```

```
viz.bigrams <- cleaned.bp2 %>%
  count_bigrams()
```

```
# filter out rare combinations, as well as digits and produce graph
viz.bigrams %>%
  filter(n > 5,
    !str_detect(word1, "\\d"),
    !str_detect(word2, "\\d")) %>%
  visualize_bigrams()
```



```
# count words co-occurring within Posts
clean.data2 <- cleaned.bp2 %>%
  unnest_tokens(word, Comments)

word_cors <- clean.data2 %>%
  group_by(word) %>%
  filter(n() >= 5) %>%
  pairwise_cor(word, Headline, sort = TRUE)
```

```
# explore top word correlations
word_cors
```

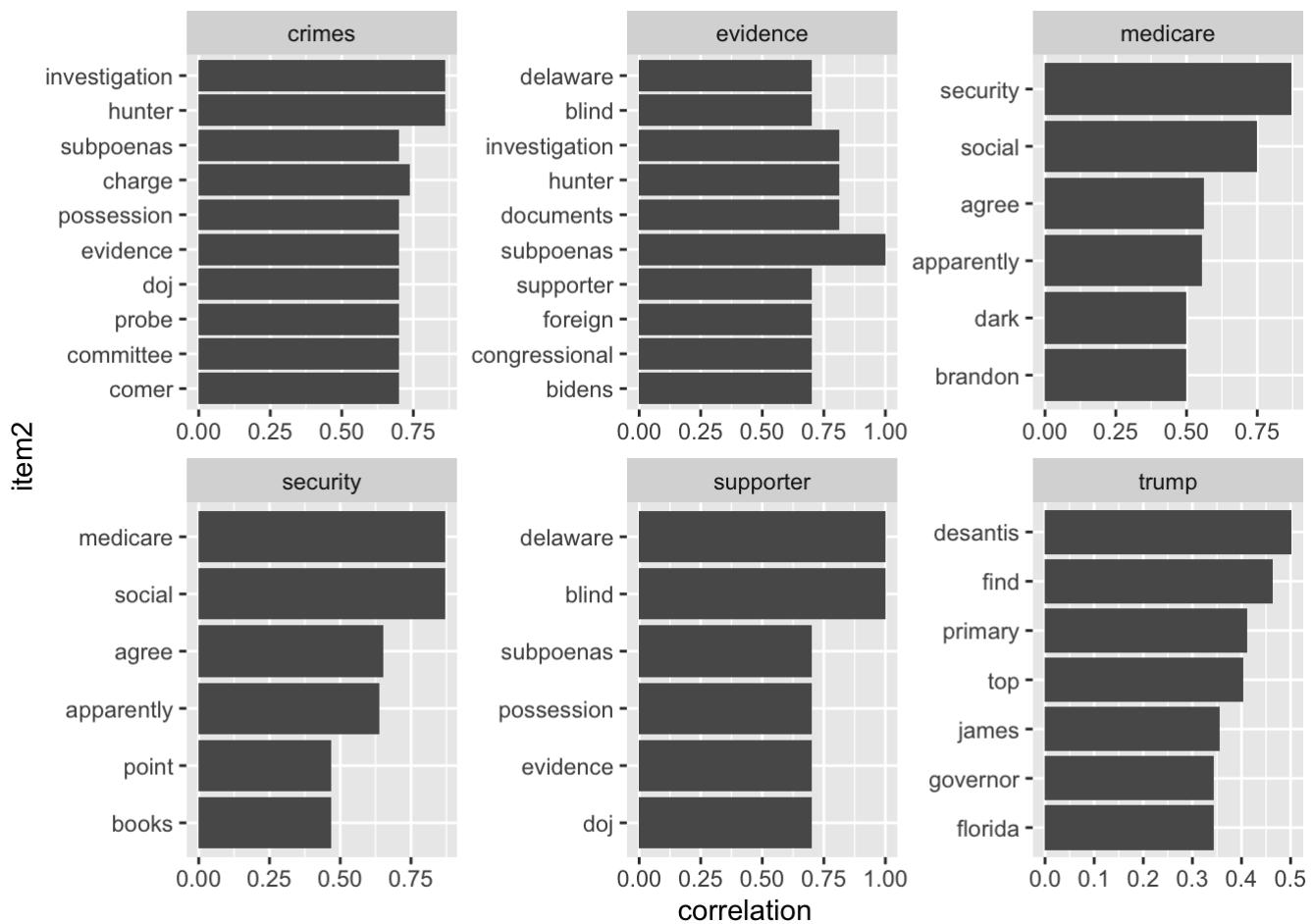
```
## # A tibble: 81,510 × 3
##   item1      item2      correlation
##   <chr>     <chr>          <dbl>
## 1 comer      committee      1
## 2 probe      committee      1
## 3 committee  comer          1
## 4 probe      comer          1
## 5 committee  probe          1
## 6 comer      probe          1
## 7 subpoenas  evidence       1
## 8 evidence   subpoenas      1
## 9 officials  administration 1
## 10 employees administration 1
## # ... with 81,500 more rows
```

```
# explore specific word correlations
# try a couple different words for fun
word_cors %>%
  filter(item1 == "hunter")
```

```
## # A tibble: 285 × 3
##   item1      item2      correlation
##   <chr>     <chr>          <dbl>
## 1 hunter  investigation      1
## 2 hunter  crimes             0.861
## 3 hunter  doj               0.811
## 4 hunter  committee          0.811
## 5 hunter  comer             0.811
## 6 hunter  probe             0.811
## 7 hunter  evidence          0.811
## 8 hunter  subpoenas         0.811
## 9 hunter  james             0.654
## 10 hunter  documents          0.654
## # ... with 275 more rows
```

```
# produce graph comparing 6 words of importance to their most correlated words
word_cors %>%
  filter(item1 %in% c("evidence", "crimes", "supporter", "trump", "medicare", "security"))
) %>%
  group_by(item1) %>%
  top_n(6) %>%
  ungroup() %>%
  mutate(item2 = reorder(item2, correlation)) %>%
  ggplot(aes(item2, correlation)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ item1, scales = "free") +
  coord_flip()
```

## Selecting by correlation



```
# create a graph showing relationship clusters using pairwise correlation scores
set.seed(2016)
```

