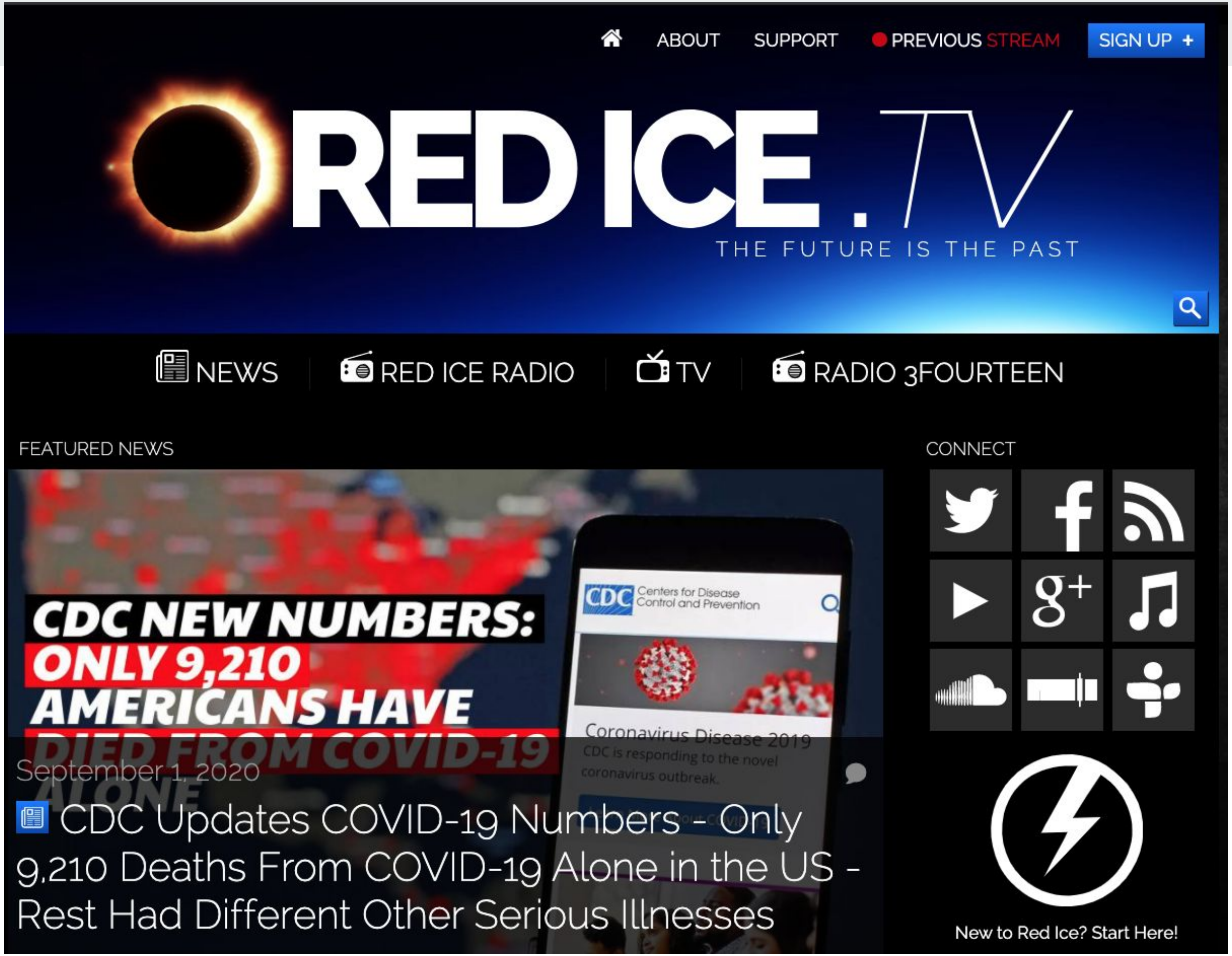
# Observations of 100 "Misleading" News Sites

Presentation 1
Grant Atkins
October 22, 2020

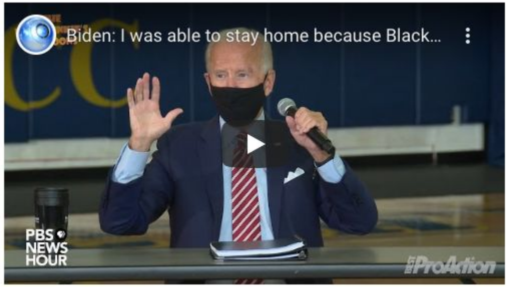
Old Dominion University
Web Archiving Forensics
CS 895



September 1, 2020

CDC Updates COVID-19 Numbers - Only 9,210 Deaths From COVID-19 Alone in the US - Rest Had Different Other Serious Illnesses

http://web.archive.org/web/20201006095632/https://redice.tv/



# BIDEN: I WAS ABLE TO STAY HOME BECAUSE BLACK WOMEN KEPT THE GROCERY SHELVES STOCKED.

**BRAVE REPORT**

http://web.archive.org/web/20201022022537/https://bravereport.com/
https://www.youtube.com/watch?v=o_OVWEb2kXI&feature=emb_title
https://www.snopes.com/fact-check/biden-pandemic-black-women-grocery/

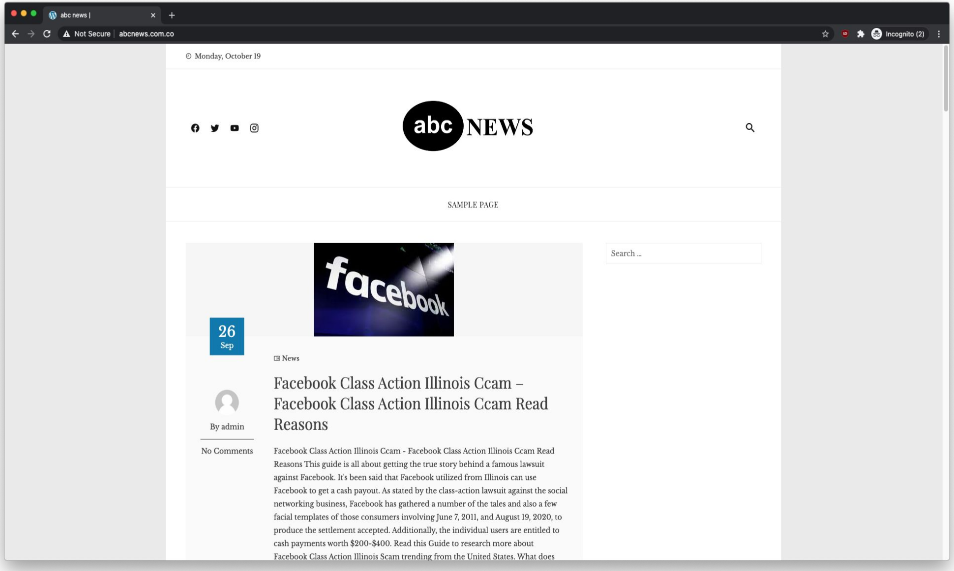
# Purpose

- Determine how well archived these sites are
- Observe the spread of past identified Misleading News sites across Twitter
  - Are they still being spread? Are they still getting archived?
- Identify top spreaders of these sites based on tweets

## Dataset & Sample

- Original dataset by Melissa Zimdars ~ 944 URIs started in 2016. Now up to 1001 (source)
- Sites tagged to indicate site types (e.g., bias, conspiracy, fake)
- Randomly sampled 100 top-level sites
- For each URI, retrieved the timemap and up to 10,000 tweets containing the URI domain

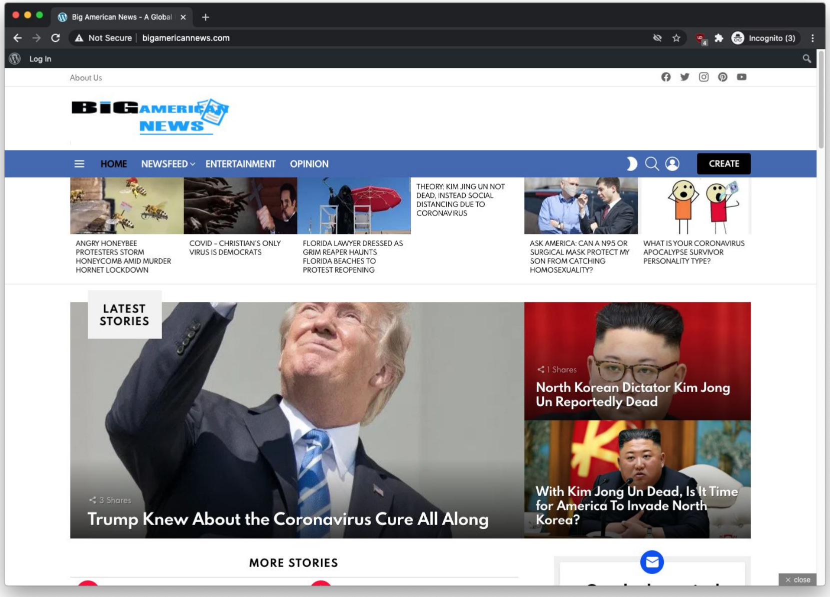
# Glimpse of the Dataset



Monday, October 19

Facebook Class Action Illinois Ccam – Facebook Class Action Illinois Ccam Read Reasons

Facebook Class Action Illinois Ccam - Facebook Class Action Illinois Ccam Read Reasons This guide is all about getting the true story behind a famous lawsuit against Facebook. It's been said that Facebook utilized from Illinois can use Facebook to get a cash payout. As stated by the class-action lawsuit against the social networking business, Facebook has gathered a number of the tales and also a few facial templates of those consumers involving June 7, 2011, and August 19, 2020, to produce the settlement accepted. Additionally, the individual users are entitled to cash payments worth $200-$400. Read this Guide to research more about Facebook Class Action Illinois Scam trending from the United States. What does



ANGRY HONEYBEE PROTESTERS STORM HONEYCOMB AMID MURDER HORNET LOCKDOWN

COVID – CHRISTIAN'S ONLY VIRUS IS DEMOCRATS

FLORIDA LAWYER DRESSED AS GRIM REAPER HAUNTS FLORIDA BEACHES TO PROTEST REOPENING

THEORY: KIM JING UN NOT DEAD, INSTEAD SOCIAL DISTANCING DUE TO CORONAVIRUS

ASK AMERICA: CAN A N95 OR SURGICAL MASK PROTECT MY SON FROM CATCHING HOMOSEXUALITY?

WHAT IS YOUR CORONAVIRUS APOCALYPSE SURVIVOR PERSONALITY TYPE?

Trump Knew About the Coronavirus Cure All Along

North Korean Dictator Kim Jong Un Reportedly Dead

With Kim Jong Un Dead, Is It Time for America To Invade North Korea?

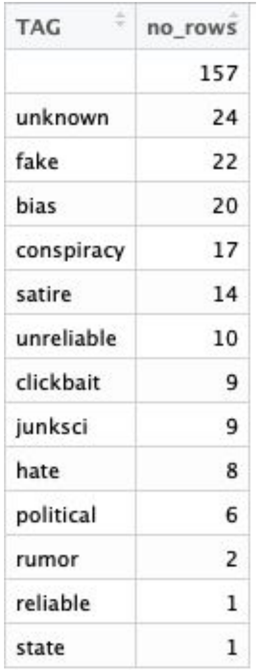
"http://abcnews.com.co/" - Fake, attempts to appear as https://abcnews.go.com/

"http://bigamericannews.com/" - Satirical, Fake with unproven stories (last stories were from May 2020)

# Sample Tags

Sites can be tagged multiple times (e.g., fake and satire)

Unknown was not classified at the time of dataset creation



| TAG | no_rows |
|---|---|
| | 157 |
| unknown | 24 |
| fake | 22 |
| bias | 20 |
| conspiracy | 17 |
| satire | 14 |
| unreliable | 10 |
| clickbait | 9 |
| junksci | 9 |
| hate | 8 |
| political | 6 |
| rumor | 2 |
| reliable | 1 |
| state | 1 |

http://web.archive.org/web/20171119174607/http://www.opensources.co/#

# How many are still alive?

| Status Code | Count |
|---|---|
| 200 | 92 |
| 402 | 1 |
| 403 | 1 |
| 522 | 1 |
| No HTTP Response | 5 |

TIL about HTTP/1.1 402 Payment Required

# 5 No longer provide an HTTP response

http://conservativebyte.com
http://ewao.com
http://theuspatriot.com
http://presstv.ir
http://usviewer.com

```
curl -ivL http://conservativebyte.com
* Could not resolve host: conservativebyte.com
* Closing connection 0
curl: (6) Could not resolve host: conservativebyte.com

curl -ivL http://ewao.com
* Could not resolve host: ewao.com
* Closing connection 0
curl: (6) Could not resolve host: ewao.com

curl -ivL http://theuspatriot.com
* Could not resolve host: theuspatriot.com
* Closing connection 0
curl: (6) Could not resolve host: theuspatriot.com

curl -ivL http://usviewer.com
* Could not resolve host: usviewer.com
* Closing connection 0
curl: (6) Could not resolve host: usviewer.com

curl -ivL http://presstv.ir
*   Trying 204.155.146.233...
* TCP_NODELAY set
TIMEOUT
```

# Status Codes don’t represent the content

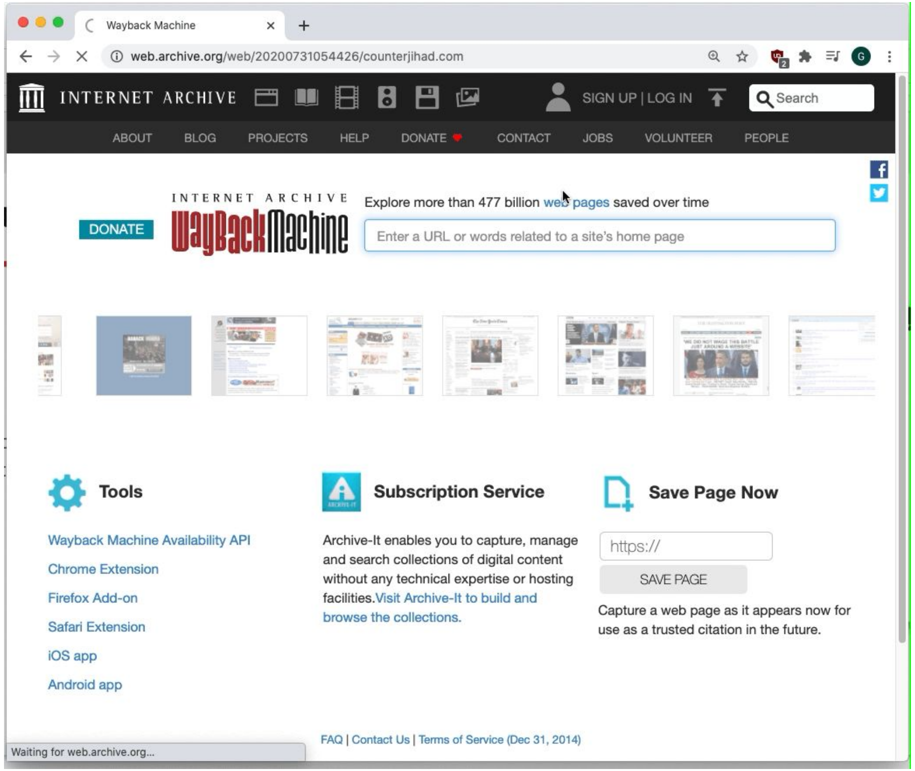
```
curl -iL http://qpolitical.com/
HTTP/1.1 200 OK
Server: nginx
Date: Tue, 20 Oct 2020 05:33:00 GMT
Content-Type: text/html; charset=utf-8
Transfer-Encoding: chunked
Connection: close
Expires: Tue, 20 Oct 2020 05:43:00 GMT
Cache-Control: max-age=600
X-Frame-Options: DENY

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"><html><head><meta http-equiv="refresh"
content="0;url=https://searchassist.verizon.com/main?ParticipantID=9asfsjdf9nwp7iqj2fp5wzabaada&FailedURI=http%3A%2F
%2Fqpolitical.com%2F&FailureMode=1&Implementation=&AddInType=4&Version=pywr1.0&ClientLocation=us"/><script
type="text/javascript">url="https://searchassist.verizon.com/main?ParticipantID=9asfsjdf9nwp7iqj2fp5wzabaada&FailedURI=
http%3A%2F%2Fqpolitical.com%2F&FailureMode=1&Implementation=&AddInType=4&Version=pywr1.0&ClientLocation=us";if(top.loca
tion!=location){var
w=window,d=document,e=d.documentElement,b=d.body,x=w.innerWidth||e.clientWidth||b.clientWidth,y=w.innerHeight||e.client
Height||b.clientHeight;url+="&w="+x+"&h="+y;}window.location.replace(url);</script></head><body></body></html>
```

# Adware persists in Web Archives

1) Memento: http://web.archive.org/web/20161118090952/buzzfeedusa.com
2) No HTTP header Location for redirect but as soon as you visit the memento….
    a) Only gives html redirects when giving a non-curl User-Agent.
    b) curl -H "User-Agent: googlebot" -iL http://ww1.buzzfeedusa.com
3) Redirects to http://external-115321085.us-west-2.elb.amazonaws.com/api/jefgold/search?p=buzzfeedusa&subid=89227033
4) Then to redirects to a live web Bing query

Indicating that there is Adware being stored in these mementos and JavaScript rewrites aren't good enough to currently catch these on the replay side.

# “Counterjihad.com” mementos redirect to live web



Explore more than 477 billion web pages saved over time

Enter a URL or words related to a site's home page

**Tools**

- Wayback Machine Availability API
- Chrome Extension
- Firefox Add-on
- Safari Extension
- iOS app
- Android app

**Subscription Service**

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit Archive-It to build and browse the collections.

**Save Page Now**

https://

SAVE PAGE

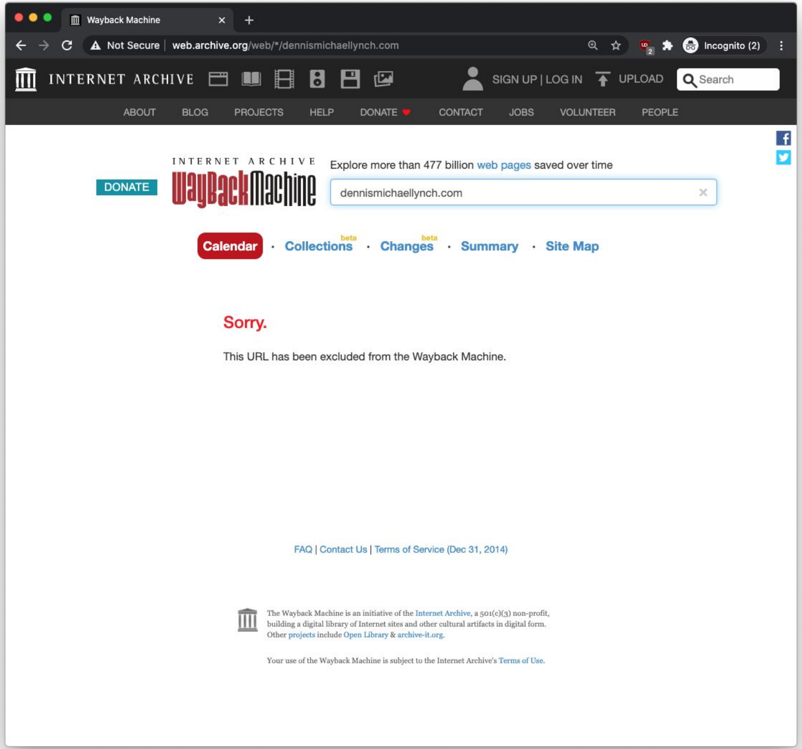
Capture a web page as it appears now for use as a trusted citation in the future.

FAQ | Contact Us | Terms of Service (Dec 31, 2014)

Waiting for web.archive.org...

http://web.archive.org/web/20200731054426/counterjihad.com

# Some sites don’t want to be archived



Explore more than 477 billion web pages saved over time

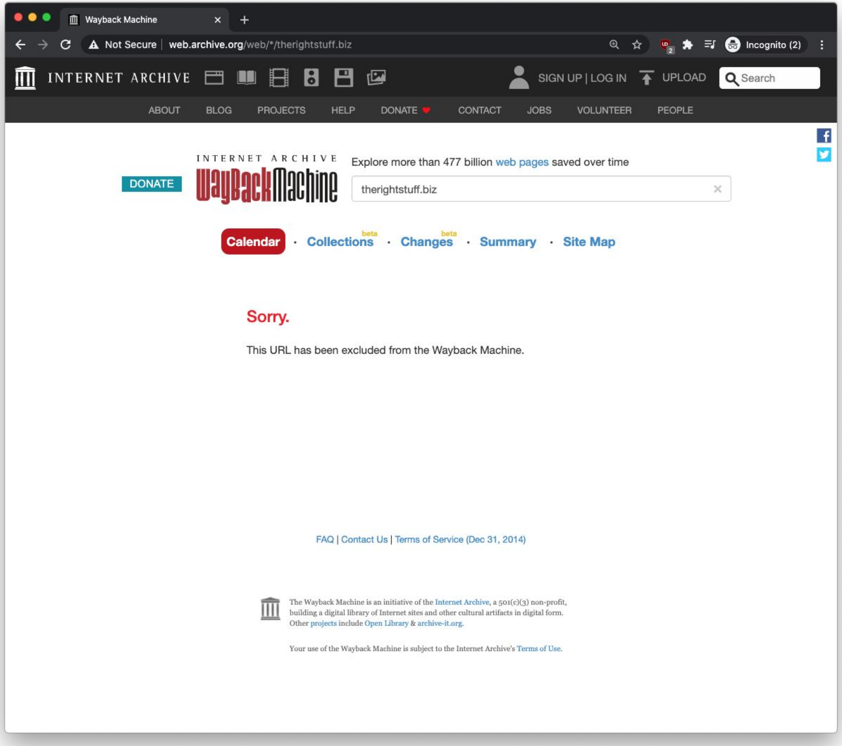
Sorry.

This URL has been excluded from the Wayback Machine.

The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit, building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org.

Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use.

http://web.archive.org/web/*/dennismichaellynch.com



Explore more than 477 billion web pages saved over time

Sorry.

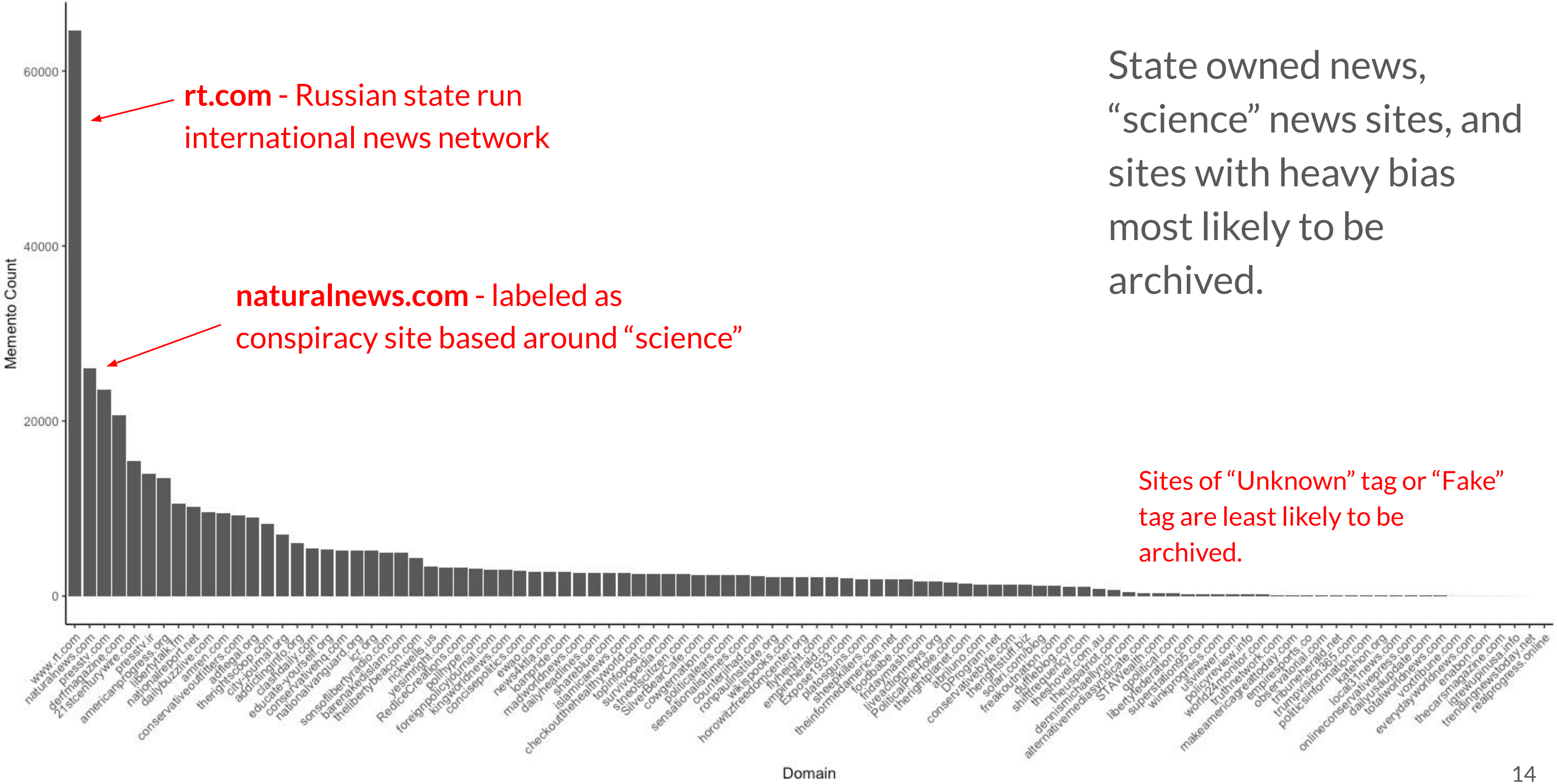
This URL has been excluded from the Wayback Machine.

The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit, building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org.

Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use.

http://web.archive.org/web/*/therightstuff.biz

# Memento Count by top-level site



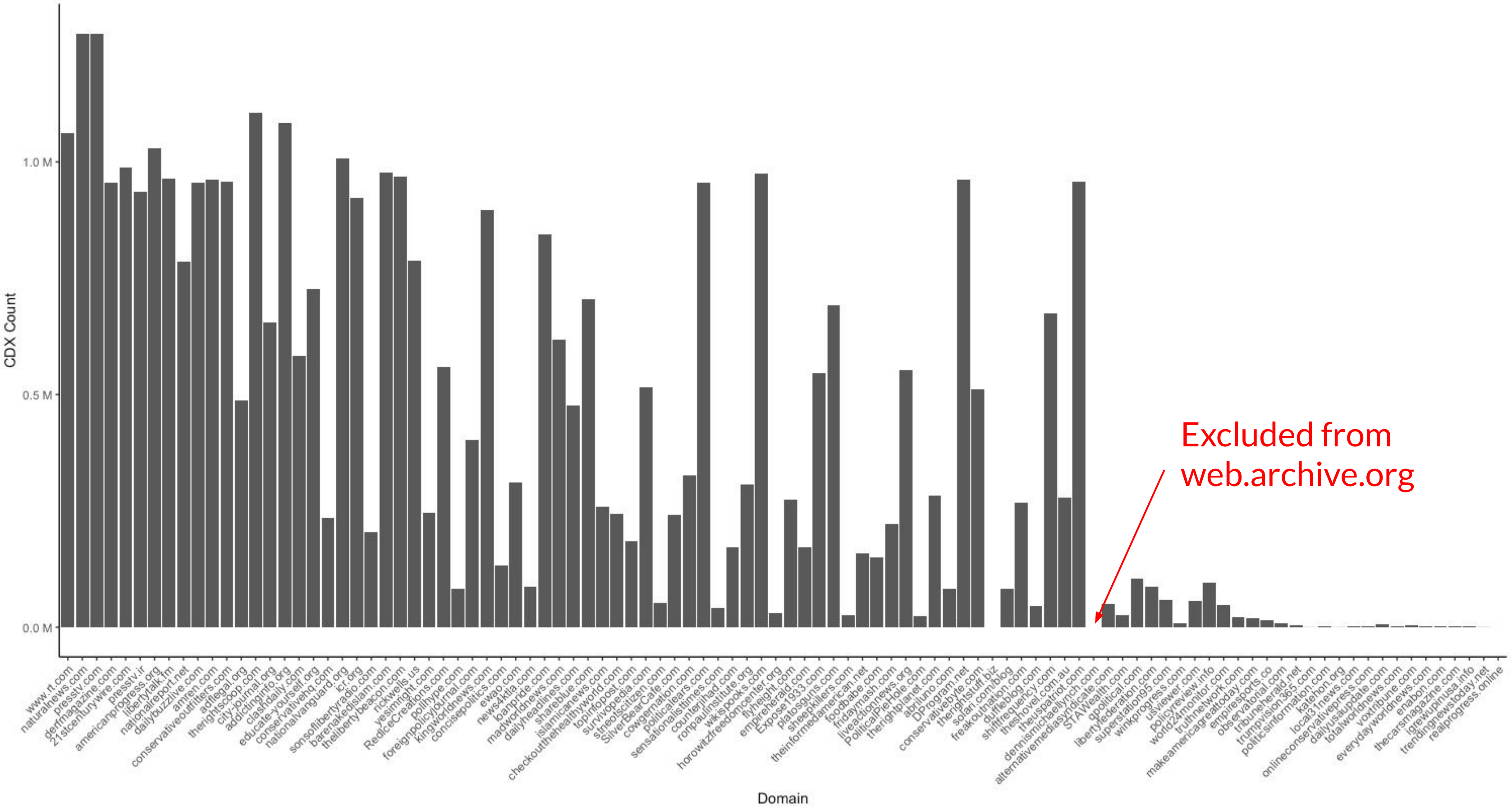
**rt.com** - Russian state run international news network

**naturalnews.com** - labeled as conspiracy site based around “science”

State owned news, “science” news sites, and sites with heavy bias most likely to be archived.

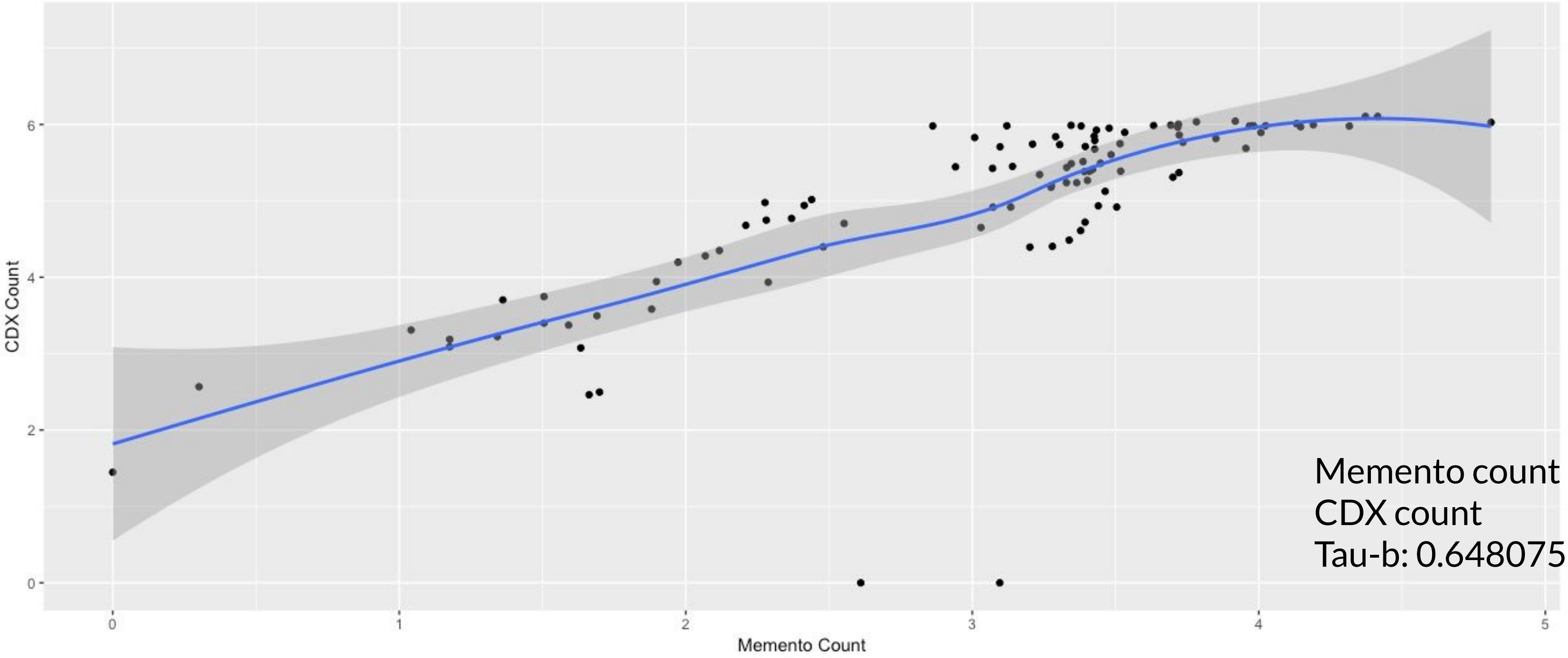
Sites of “Unknown” tag or “Fake” tag are least likely to be archived.

# CDX Count by top-level site



Excluded from web.archive.org

# Weak Positive Correlation between Memento Count and CDX Count



Memento count vs. CDX count
Tau-b: 0.6480752

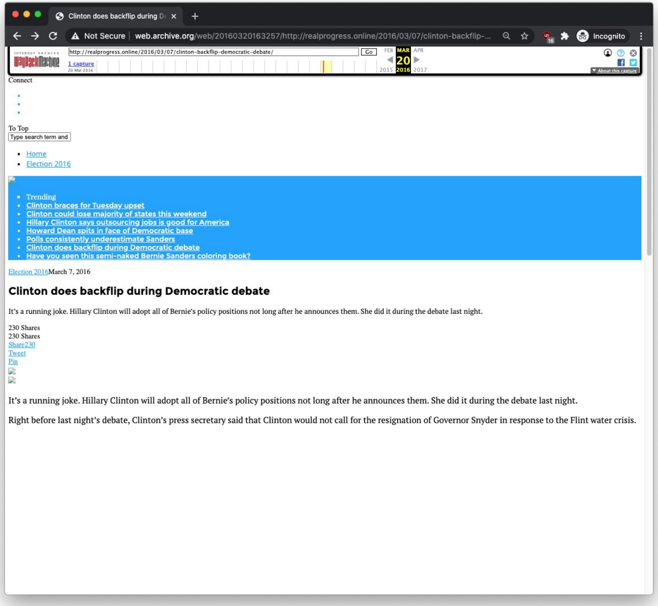
Log scale applied to both axis

# Archive Diversity

| Archive | Memento Count |
|---|---|
| web.archive.org | 264706 |
| wayback.archive-it.org | 127131 |
| webarchive.loc.gov | 3768 |
| web.archive.bibalex.org | 2955 |
| arquivo.pt | 1559 |
| archive.md | 1462 |
| wayback.vefsafn.is | 552 |
| swap.stanford.edu | 413 |
| web.archive.org.au | 147 |
| www.webarchive.org.uk | 19 |
| perma.cc | 4 |

# Top-level vs. deep link mementos

```
curl -s "http://web.archive.org/cdx/search/cdx?url=realprogress.online&matchType=prefix" | awk '{print
"https://web.archive.org/web/" $2 "/" $3};'
https://web.archive.org/web/20160318220246/http://realprogress.online/2016/03/05/hillary-clinton-told-crowd-outsourcing-good-america/
https://web.archive.org/web/20160319022817/http://realprogress.online/2016/03/05/seen-semi-naked-bernie-sanders-coloring-book/
https://web.archive.org/web/20160319153015/http://realprogress.online/2016/03/06/howard-dean-spits-face-democratic-base/
https://web.archive.org/web/20160320163257/http://realprogress.online/2016/03/07/clinton-backflip-democratic-debate/
https://web.archive.org/web/20160321155737/http://realprogress.online/2016/03/09/polls-consistently-underestimate-sanders/
https://web.archive.org/web/20160323173041/http://realprogress.online/2016/03/14/clinton-braces-tuesday-upset/
https://web.archive.org/web/20160325031702/http://realprogress.online/2016/03/24/clinton-using-dirty-tactics-trick-washington-voters/
https://web.archive.org/web/20160331034900/http://realprogress.online/2016/03/30/dc-democrats-try-keep-sanders-off-ballot/
https://web.archive.org/web/20160401034909/http://realprogress.online/2016/03/31/clinton-looks-set-lose-wisconsin/
https://web.archive.org/web/20160402112700/http://realprogress.online/2016/04/02/clinton-bracing-two-one-loss-wisconsin/
https://web.archive.org/web/20160622144108/http://realprogress.online/favicon.ico
https://web.archive.org/web/20160318220242/http://realprogress.online/robots.txt
https://web.archive.org/web/20160319022813/http://realprogress.online/robots.txt
https://web.archive.org/web/20160320161748/http://realprogress.online/robots.txt
https://web.archive.org/web/20160321155733/http://realprogress.online/robots.txt
https://web.archive.org/web/20160323173037/http://realprogress.online/robots.txt
https://web.archive.org/web/20160325031701/http://realprogress.online/robots.txt
https://web.archive.org/web/20160331034853/http://realprogress.online/robots.txt
https://web.archive.org/web/20160401034906/http://realprogress.online/robots.txt
https://web.archive.org/web/20160402112659/http://realprogress.online/robots.txt
https://web.archive.org/web/20160622144115/http://realprogress.online/wp-content/plugins/ad-inserter/css/devices.css?ver=1.6.1
https://web.archive.org/web/20160622144114/http://realprogress.online/wp-content/plugins/jetpack/css/jetpack.css?ver=3.9.2
https://web.archive.org/web/20160622144116/http://realprogress.online/wp-content/plugins/jetpack/modules/photon/photon.js?ver=20130122
https://web.archive.org/web/20160622144117/http://realprogress.online/wp-content/plugins/jetpack/modules/wpgroho.js?ver=4.4.2
https://web.archive.org/web/20160622144116/http://realprogress.online/wp-includes/js/comment-reply.min.js?ver=4.4.2
https://web.archive.org/web/20160622144116/http://realprogress.online/wp-includes/js/jquery/jquery-migrate.min.js?ver=1.2.1
https://web.archive.org/web/20160622144116/http://realprogress.online/wp-includes/js/jquery/jquery.js?ver=1.11.3
https://web.archive.org/web/20160622144118/http://realprogress.online/wp-includes/js/wp-embed.min.js?ver=4.4.2
```



web.archive.org/web/20160320163257/http://realprogress.online/2016/03/07/clinton-backflip-...

http://realprogress.online/2016/03/07/clinton-backflip-democratic-debate/

1 capture
20 Mar 2016

Connect

To Top
Type search term and

- Home
- Election 2016

- Trending
- Clinton braces for Tuesday upset
- Clinton could lose majority of states this weekend
- Hillary Clinton says outsourcing jobs is good for America
- Howard Dean spits in face of Democratic base
- Polls consistently underestimate Sanders
- Clinton does backflip during Democratic debate
- Have you seen this semi-naked Bernie Sanders coloring book?

Election 2016March 7, 2016

Clinton does backflip during Democratic debate

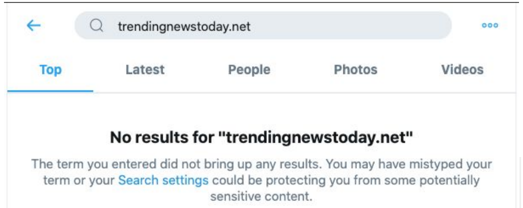
It’s a running joke. Hillary Clinton will adopt all of Bernie’s policy positions not long after he announces them. She did it during the debate last night.

230 Shares
230 Shares
Share230
Tweet
Pin

It’s a running joke. Hillary Clinton will adopt all of Bernie’s policy positions not long after he announces them. She did it during the debate last night.

Right before last night’s debate, Clinton’s press secretary said that Clinton would not call for the resignation of Governor Snyder in response to the Flint water crisis.

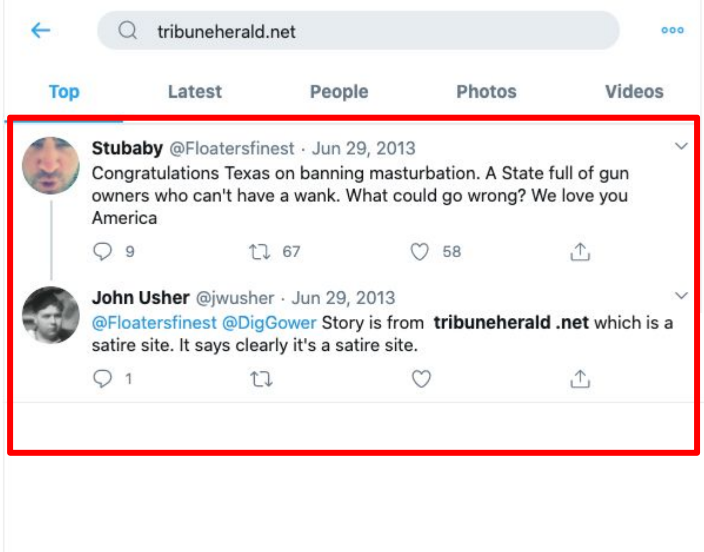
# 3 Sites were not found in tweets (...maybe 2)



**No results for "trendingnewstoday.net"**

The term you entered did not bring up any results. You may have mistyped your term or your Search settings could be protecting you from some potentially sensitive content.

Seems like someone fell for it...

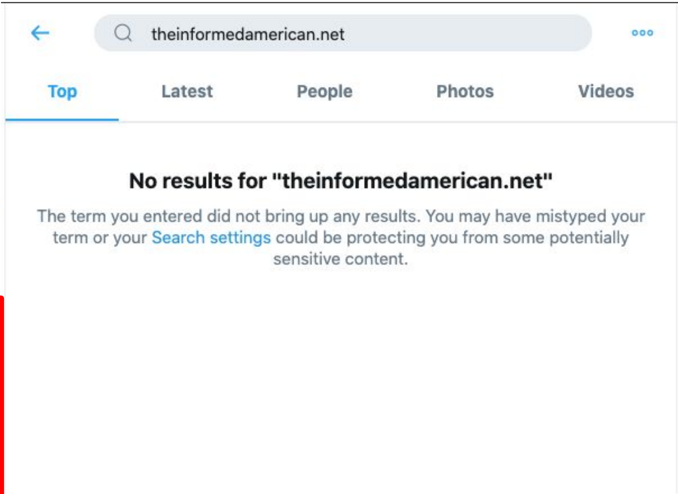


**Stubaby** @Floatersfinest · Jun 29, 2013
Congratulations Texas on banning masturbation. A State full of gun owners who can't have a wank. What could go wrong? We love you America

9 67 58

**John Usher** @jwusher · Jun 29, 2013
@Floatersfinest @DigGower Story is from **tribuneherald .net** which is a satire site. It says clearly it's a satire site.

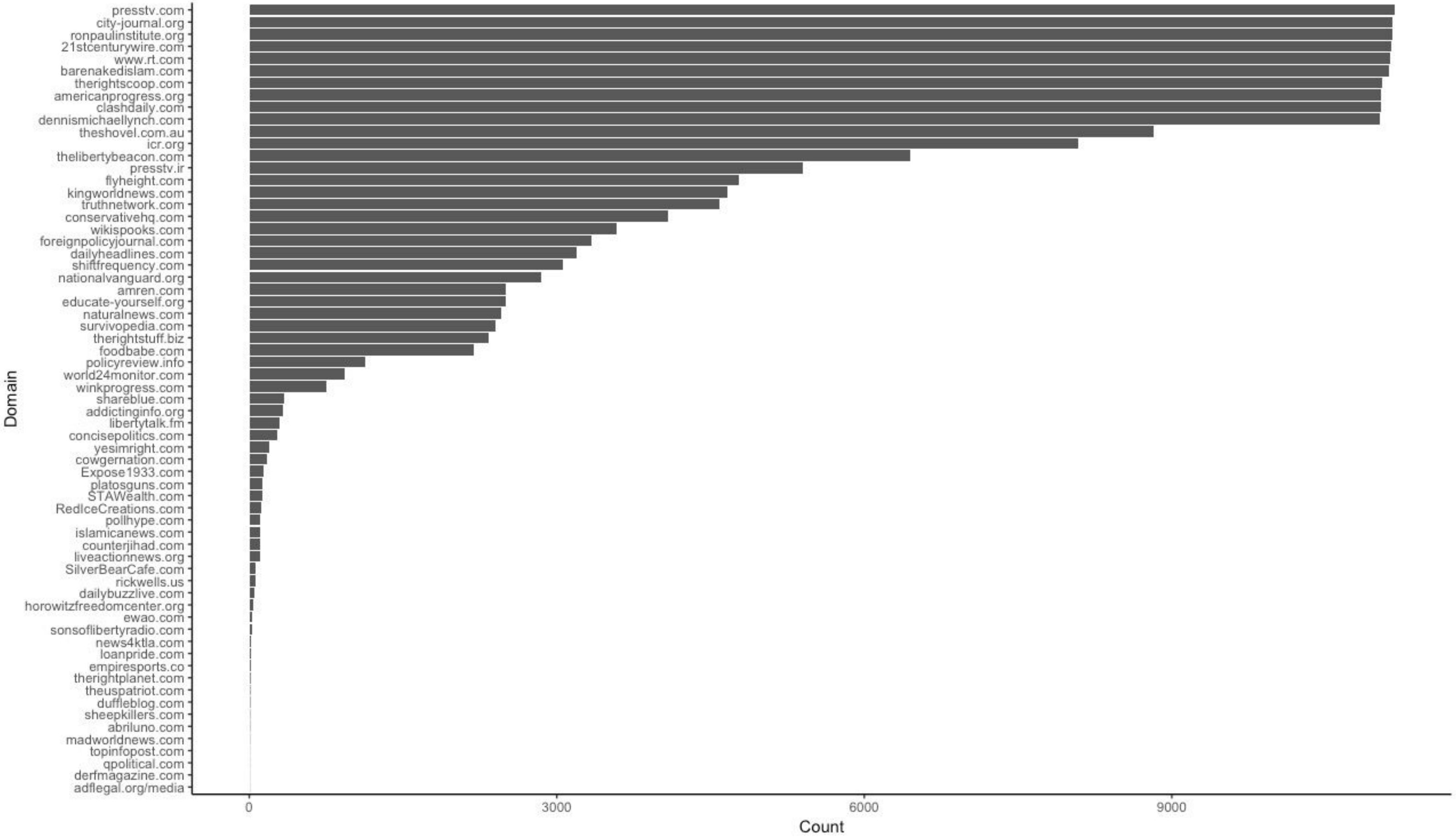
1



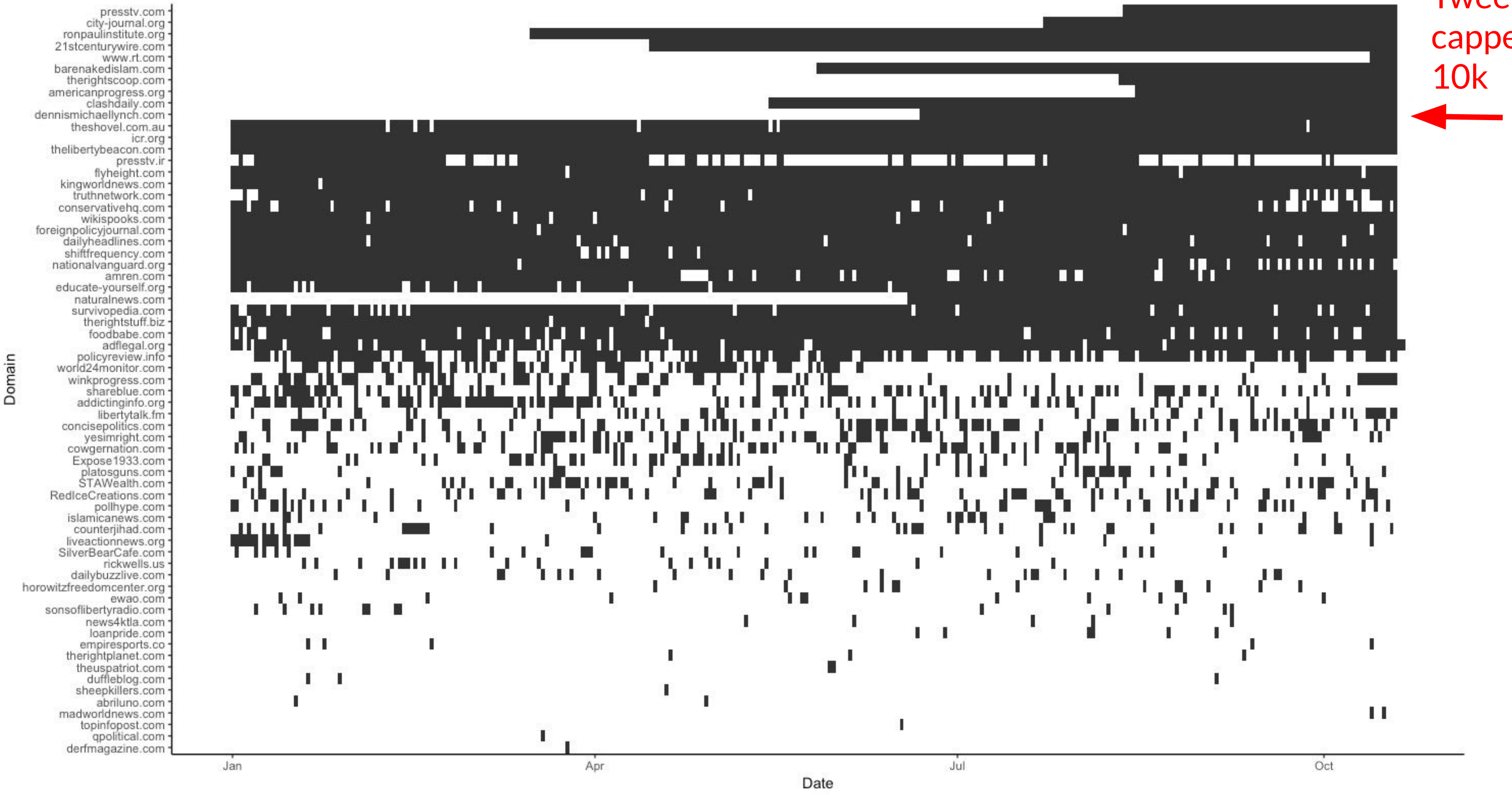
**No results for "theinformedamerican.net"**

The term you entered did not bring up any results. You may have mistyped your term or your Search settings could be protecting you from some potentially sensitive content.

# 65 sites still being tweeted in 2020

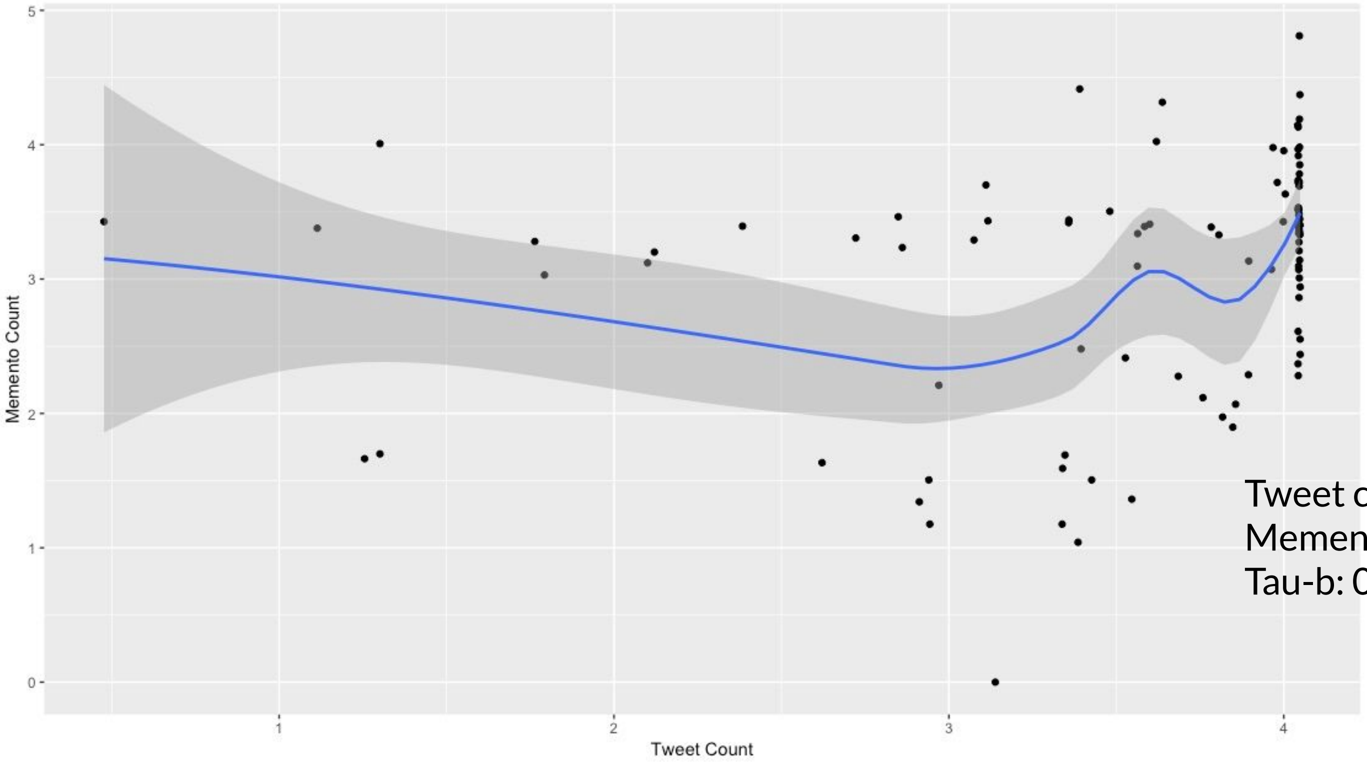


# Tweet distributions over time 2020



Tweets capped at 10k

# Weak Correlation between Memento Count and Tweet Count



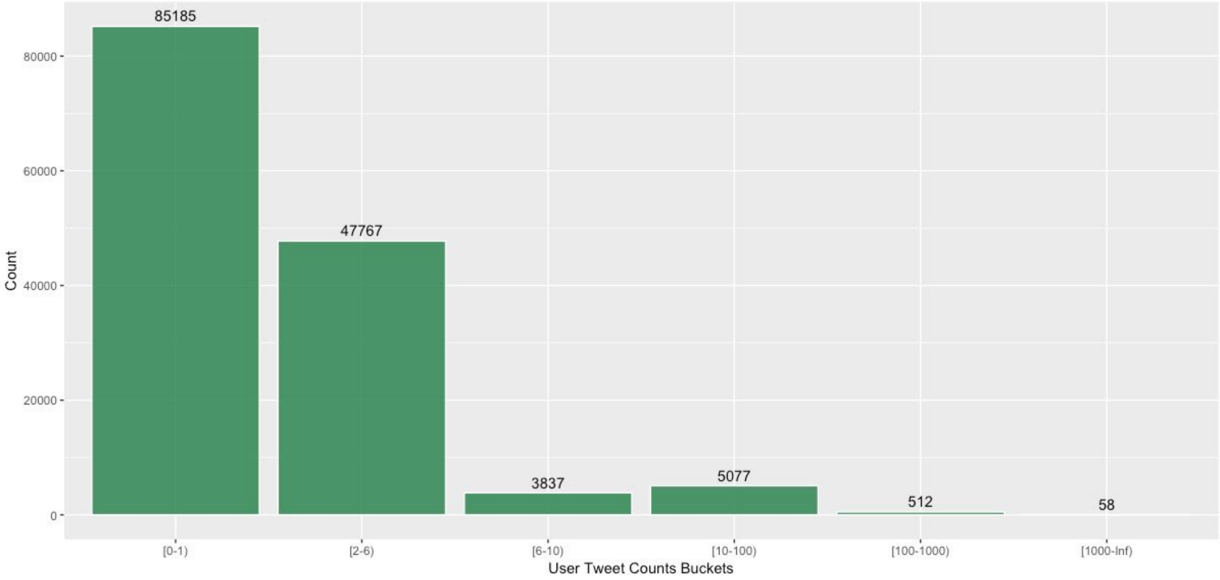
Tweet count vs. Memento count. Tau-b: 0.2224611

# 58 Users have over 1,000 tweets for these sites

640,365 tweets collected

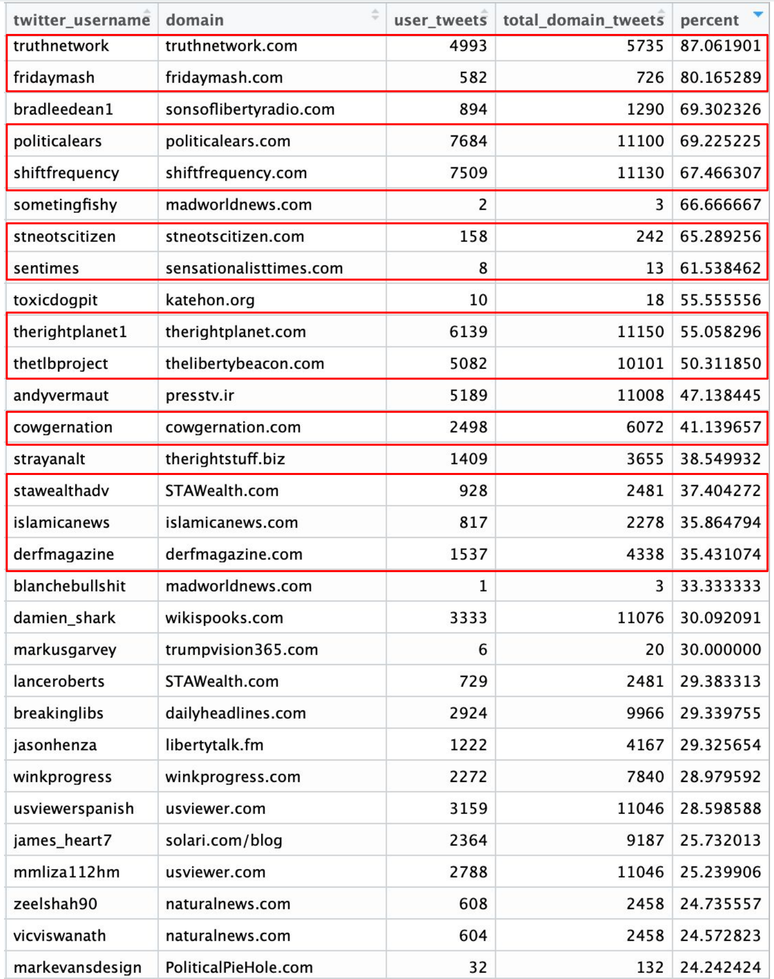
143,841 unique Twitter users

58 users with more than 1000 tweets connected to the sites collected.



Suspended/deleted accounts not shown from responses in twitter.com

# Twitter site accounts tweet their own material

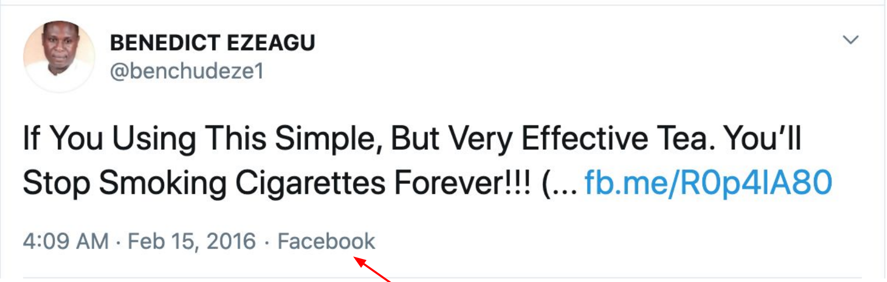


| twitter_username | domain | user_tweets | total_domain_tweets | percent |
|---|---|---|---|---|
| truthnetwork | truthnetwork.com | 4993 | 5735 | 87.061901 |
| fridaymash | fridaymash.com | 582 | 726 | 80.165289 |
| bradleedean1 | sonsoflibertyradio.com | 894 | 1290 | 69.302326 |
| politicalears | politicalears.com | 7684 | 11100 | 69.225225 |
| shiftfrequency | shiftfrequency.com | 7509 | 11130 | 67.466307 |
| sometingfishy | madworldnews.com | 2 | 3 | 66.666667 |
| stneotscitizen | stneotscitizen.com | 158 | 242 | 65.289256 |
| sentimes | sensationalisttimes.com | 8 | 13 | 61.538462 |
| toxicdogpit | katehon.org | 10 | 18 | 55.555556 |
| therightplanet1 | therightplanet.com | 6139 | 11150 | 55.058296 |
| thetlbproject | thelibertybeacon.com | 5082 | 10101 | 50.311850 |
| andyvermaut | presstv.ir | 5189 | 11008 | 47.138445 |
| cowgernation | cowgernation.com | 2498 | 6072 | 41.139657 |
| strayanalt | therightstuff.biz | 1409 | 3655 | 38.549932 |
| stawealthadv | STAWealth.com | 928 | 2481 | 37.404272 |
| islamicanews | islamicanews.com | 817 | 2278 | 35.864794 |
| derfmagazine | derfmagazine.com | 1537 | 4338 | 35.431074 |
| blanchebullshit | madworldnews.com | 1 | 3 | 33.333333 |
| damien_shark | wikispooks.com | 3333 | 11076 | 30.092091 |
| markusgarvey | trumpvision365.com | 6 | 20 | 30.000000 |
| lanceroberts | STAWealth.com | 729 | 2481 | 29.383313 |
| breakinglibs | dailyheadlines.com | 2924 | 9966 | 29.339755 |
| jasonhenza | libertytalk.fm | 1222 | 4167 | 29.325654 |
| winkprogress | winkprogress.com | 2272 | 7840 | 28.979592 |
| usviewerspanish | usviewer.com | 3159 | 11046 | 28.598588 |
| james_heart7 | solari.com/blog | 2364 | 9187 | 25.732013 |
| mmliza112hm | usviewer.com | 2788 | 11046 | 25.239906 |
| zeelshah90 | naturalnews.com | 608 | 2458 | 24.735557 |
| vicviswanath | naturalnews.com | 604 | 2458 | 24.572823 |
| markevansdesign | PoliticalPieHole.com | 32 | 132 | 24.242424 |

# Near duplicate tweets doesn’t necessarily mean they’re bots

Both of these tweets use a http://fb.me/ shortened URL indicating they were shared from Facebook (probably from a promotion) ~ hence the same text

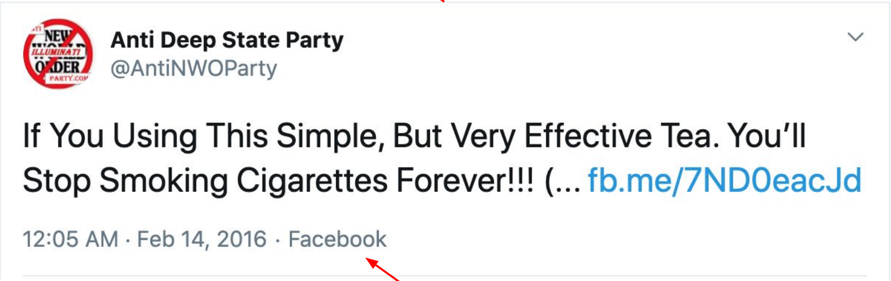
Indicates that Facebook was a large target for sharing misleading news.



BENEDICT EZEAGU
@benchudeze1

If You Using This Simple, But Very Effective Tea. You’ll Stop Smoking Cigarettes Forever!!! (... fb.me/R0p4lA80

4:09 AM · Feb 15, 2016 · Facebook



Anti Deep State Party
@AntiNWOParty

If You Using This Simple, But Very Effective Tea. You’ll Stop Smoking Cigarettes Forever!!! (... fb.me/7ND0eacJd

12:05 AM · Feb 14, 2016 · Facebook

Both redirect to
http://web.archive.org/web/20160218225616/http://checkoutthehealthyworld.com/if-you-using-this-simple-but-very-effective-tea-youll-stop-smoking-cigarettes-forever-incredibly/

# Takeaways

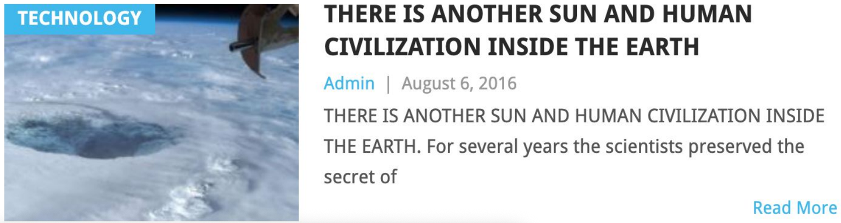
- A lot of misleading news sites live on (even if they're not posting new articles) and are well archived
- Adware/redirects are scattered throughout some of these Mementos
- For some sites a large majority of the tweets originate from the twitter site accounts
- Number of tweet linking to these sites are dropping but still remain high for frequently engaged news sites

# Backup Slides

- Data size: ~ 631 MB
- http://web.archive.org/web/20171119174607/http://www.opensources.co/#
- R & ggplot > Pandas & Seaborn (never again)
- Anti-semitic twitter users discovered:
    - https://twitter.com/__angrygoy__

# Personal favorite Screenshots

- http://web.archive.org/web/20161002194528/http://everydayworldnews.com/
- http://web.archive.org/web/20200730150738/http://sheepkillers.com/republicansdemocrats.html

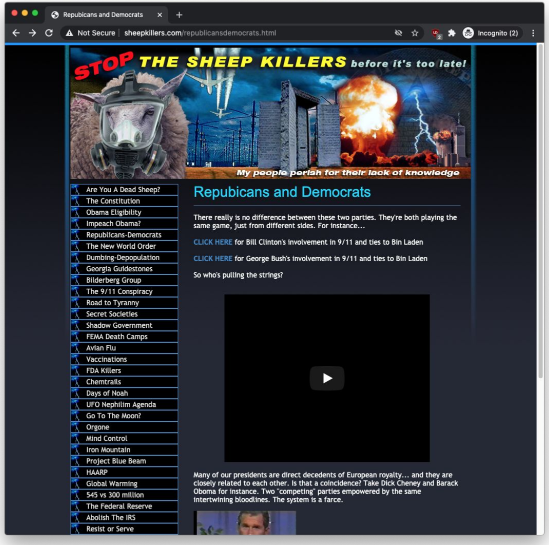


THERE IS ANOTHER SUN AND HUMAN CIVILIZATION INSIDE THE EARTH

Admin | August 6, 2016

THERE IS ANOTHER SUN AND HUMAN CIVILIZATION INSIDE THE EARTH. For several years the scientists preserved the secret of

Read More



Repubicans and Democrats

There really is no difference between these two parties. They're both playing the same game, just from different sides. For instance...

CLICK HERE for Bill Clinton's involvement in 9/11 and ties to Bin Laden

CLICK HERE for George Bush's involvement in 9/11 and ties to Bin Laden

So who's pulling the strings?

Many of our presidents are direct decedents of European royalty... and they are closely related to each other. Is that a coincidence? Take Dick Cheney and Barack Oboma for instance. Two "competing" parties empowered by the same intertwining bloodlines. The system is a farce.