

Telling a Story with Data:

Using R for data journalism

Dr. Kokil Jaidka
jaidka@ntu.edu.sg
jaidka@sas.upenn.edu

Mining social media **for scientific discoveries about human behavior.**



Takeaways



HOW TO PLAN YOUR DATA
JOURNALISM PROJECT



HOW TO GET DATA FROM
THE NYTIMES API USING R



HOW TO VISUALIZE DATA
USING R

What is Data Journalism?

Traditional Journalism

- Relies on human sources, e.g., insiders, experts, scholars and scientists
- Quotes, pictures, text

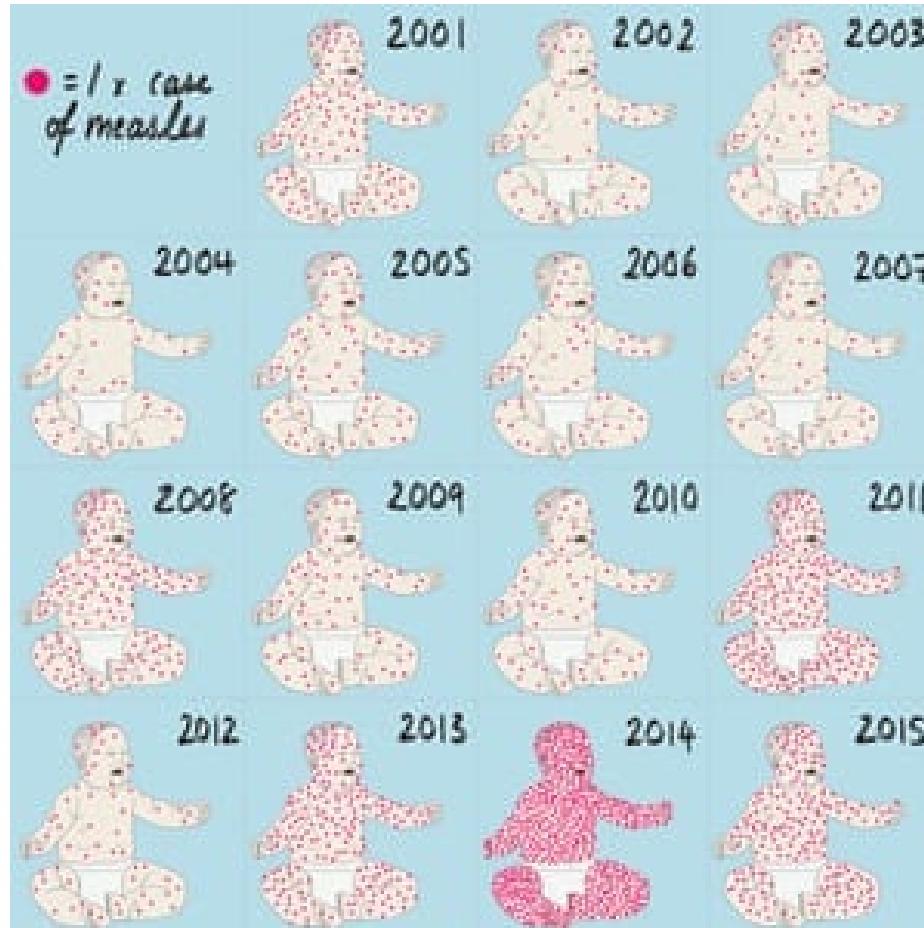
Data Journalism

- Relies on data sources, e.g., spreadsheets, websites, and databases
- Cool animations, Maps and Infographics
- Verifiability, Transparency
- Detail, Scale
- Discovery
- Investigative journalism

Let's look at some examples...

<https://www.americanpressinstitute.org/publications/reports/strategy-studies/how-data-journalism-is-different/>
<https://www.thebureauinvestigates.com/explainers/what-is-data-journalism>

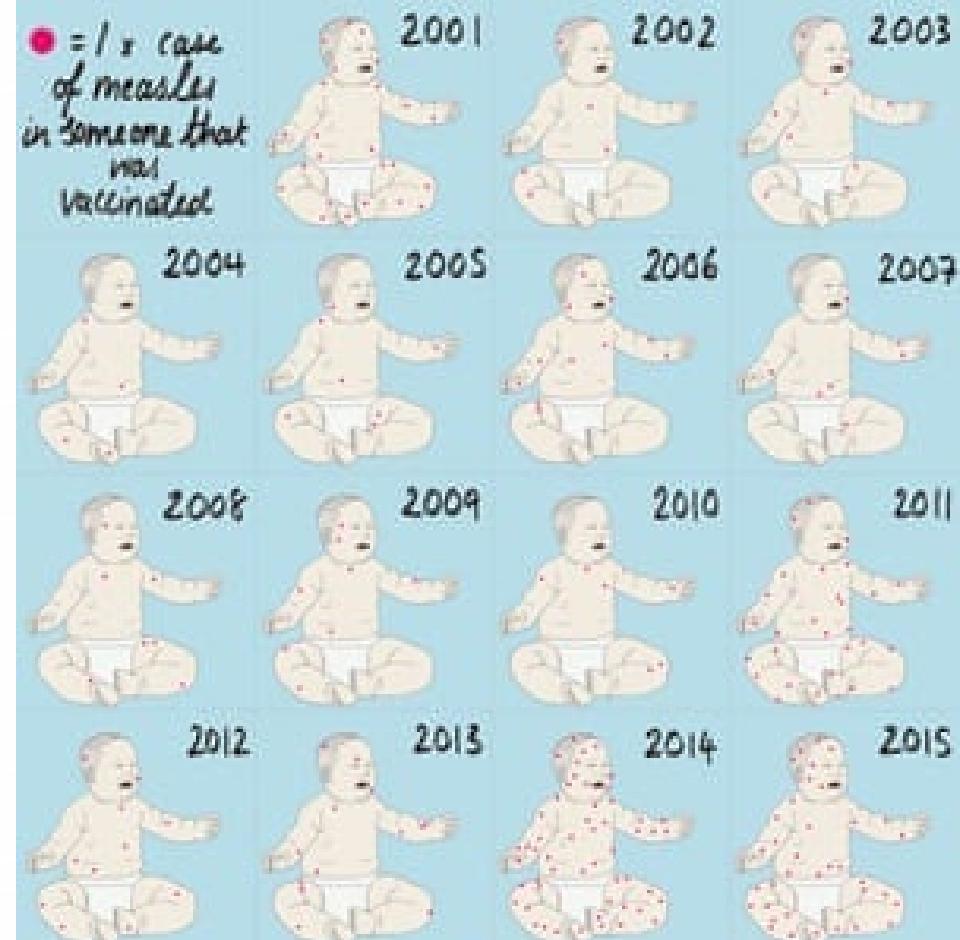
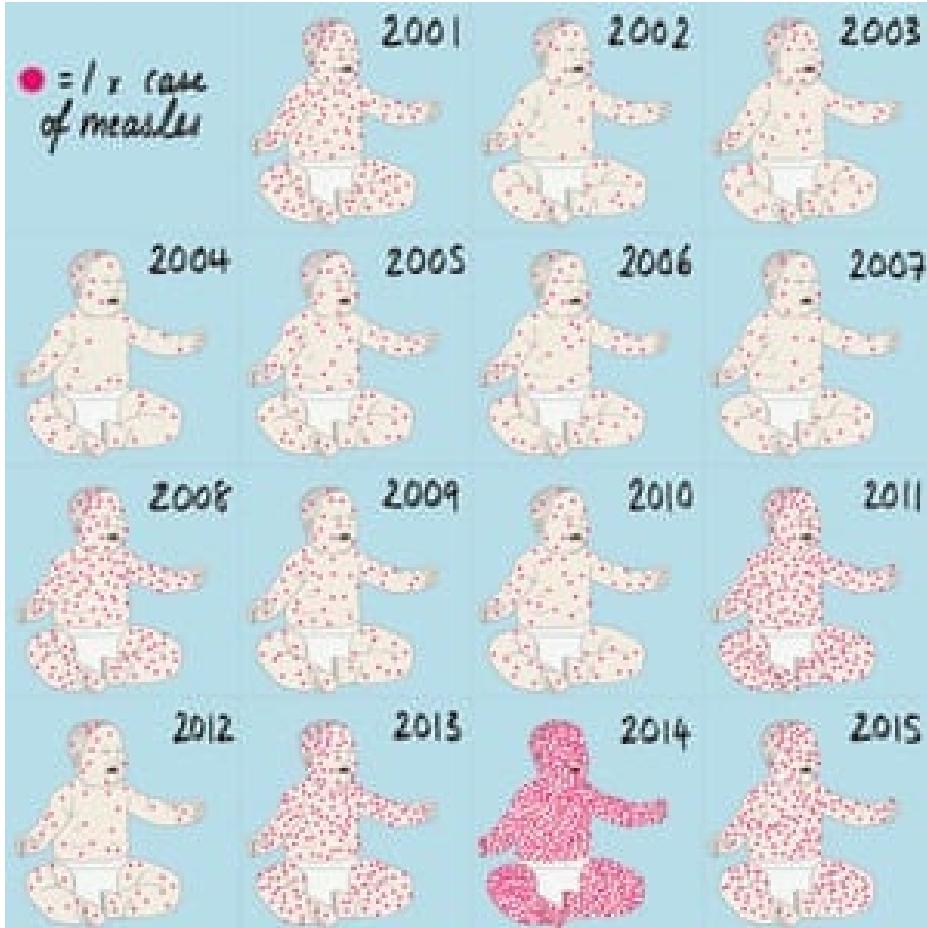
EXAMPLES



How measles outbreaks have spread amid the anti-vaccine movement. Guardian Datablog

<https://www.theguardian.com/news/datablog/2019/jan/31/measles-outbreaks-us-anti-vaccine-movement>

EXAMPLES

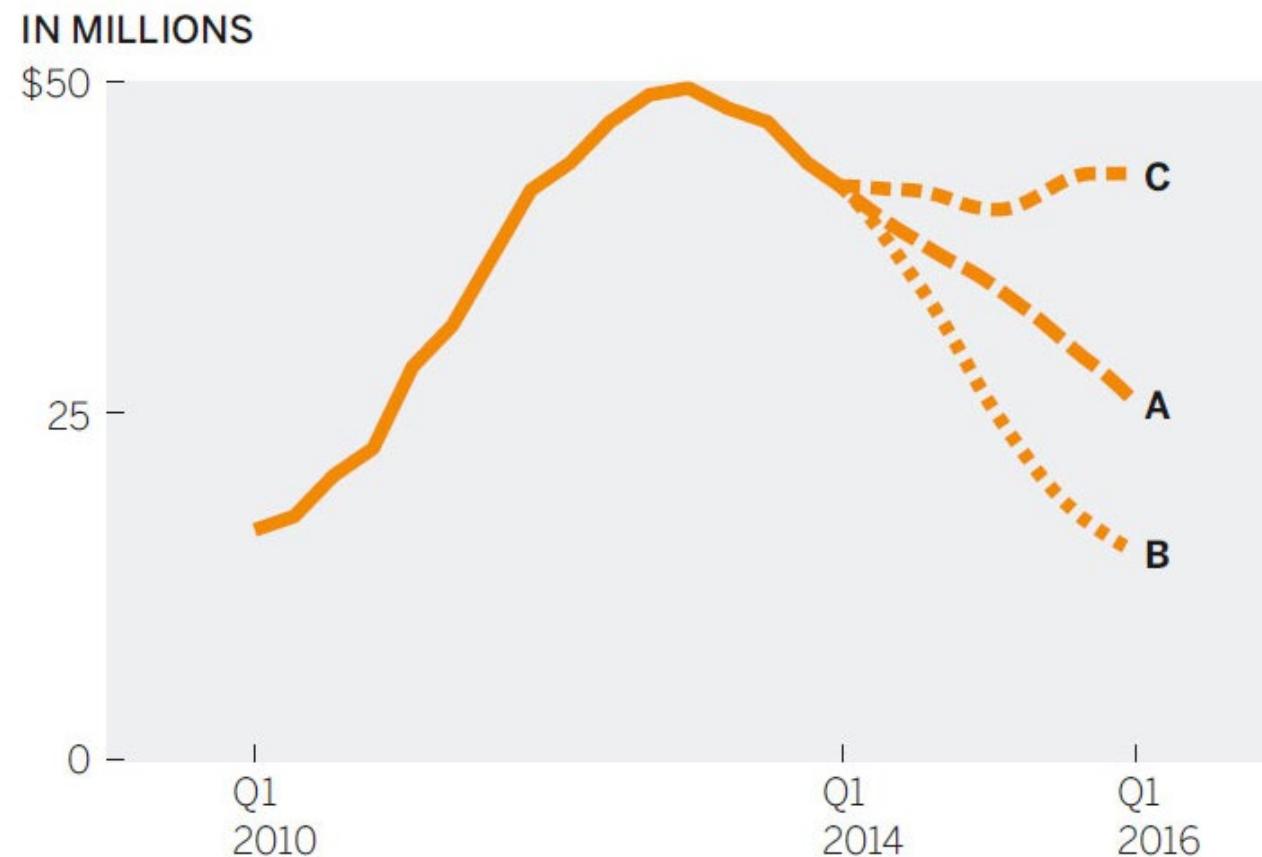


How measles outbreaks have spread amid the anti-vaccine movement. Guardian Datablog

<https://www.theguardian.com/news/datablog/2019/jan/31/measles-outbreaks-us-anti-vaccine-movement>

EXAMPLES

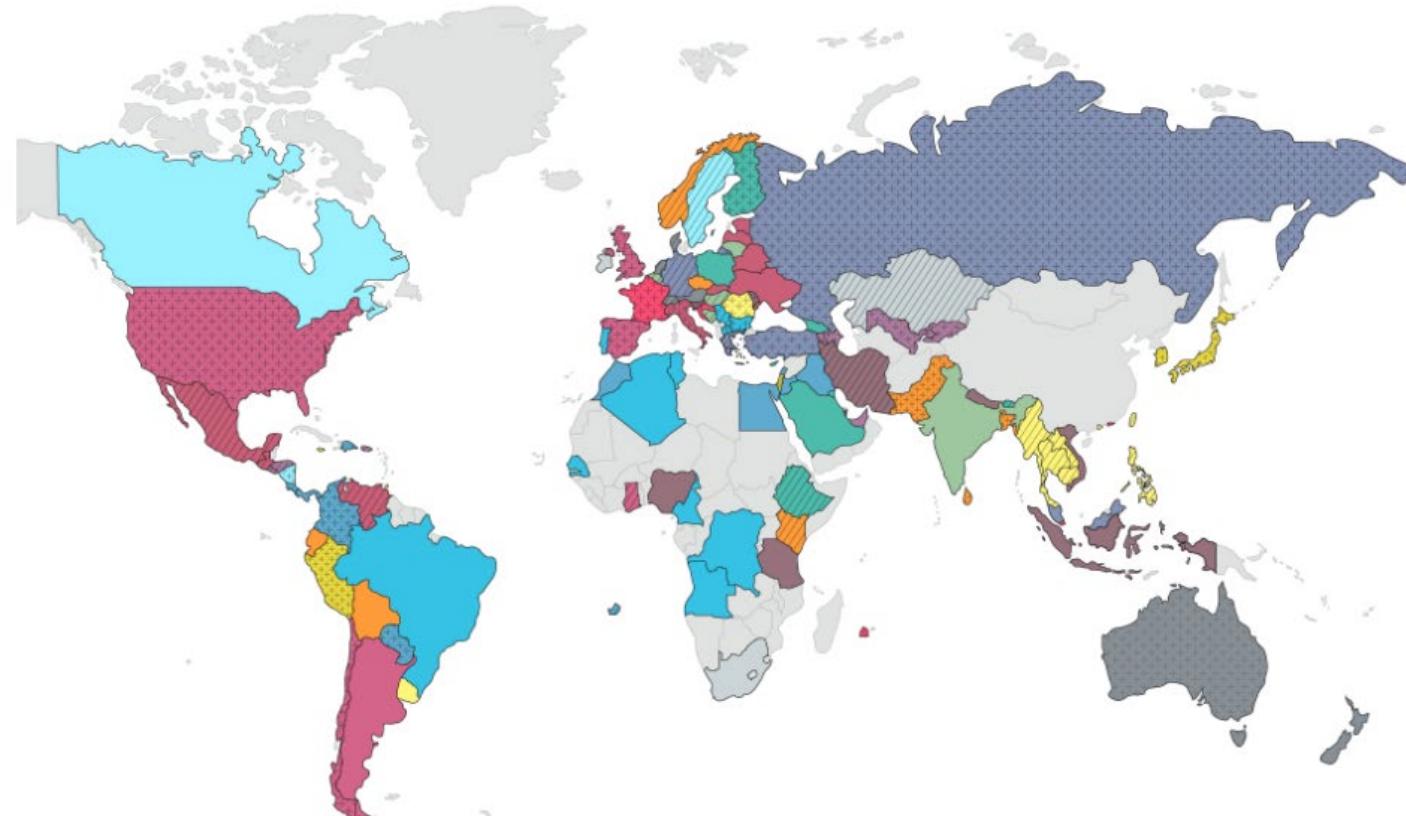
PROJECTIONS—THREE SCENARIOS



Berinato, S. (2016). *Good charts: The HBR guide to making smarter, more persuasive data visualizations.* Harvard Business Review Press.

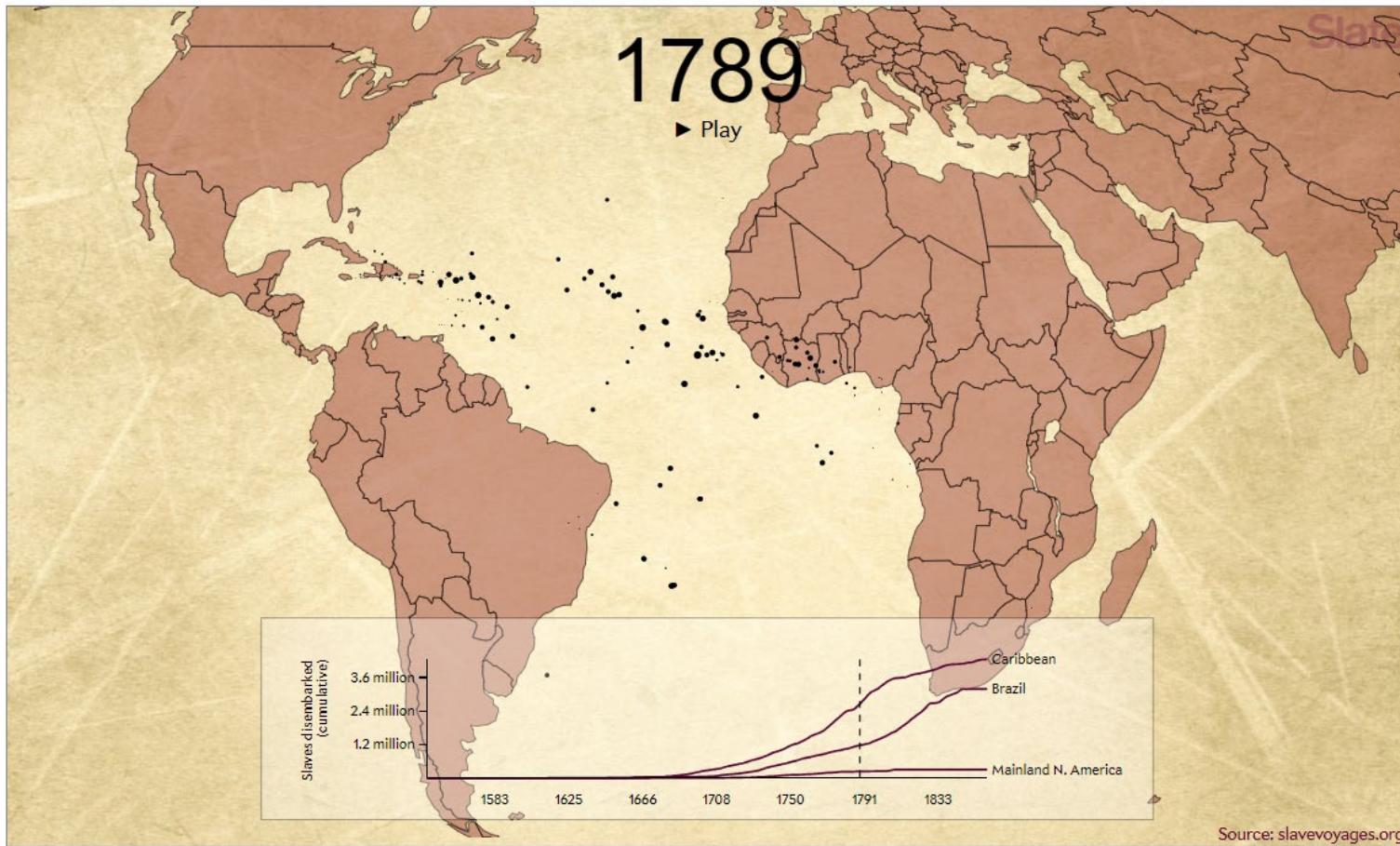
EXAMPLES

Most Popular*Netflix Shows by Country 2018



*Based on Google Trends data, not Netflix data. <https://www.highspeedinternet.com/resources/netflix-what-the-world-is-watching-in-2018>

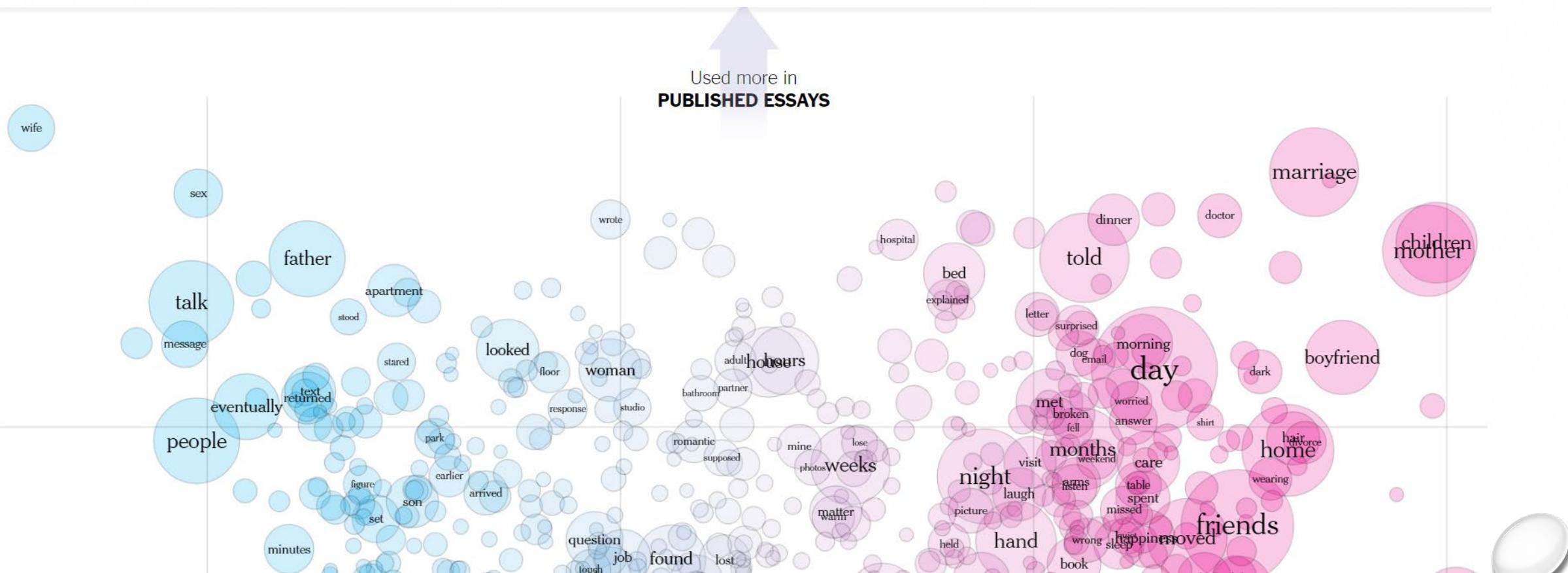
EXAMPLES



The Atlantic Slave Trade in 2 Minutes (2015). *The Slate*.

http://www.slate.com/articles/life/the_history_of_american_slavery/2015/06/animated_interactive_of_the_history_of_the_atlantic_slave_trade.html

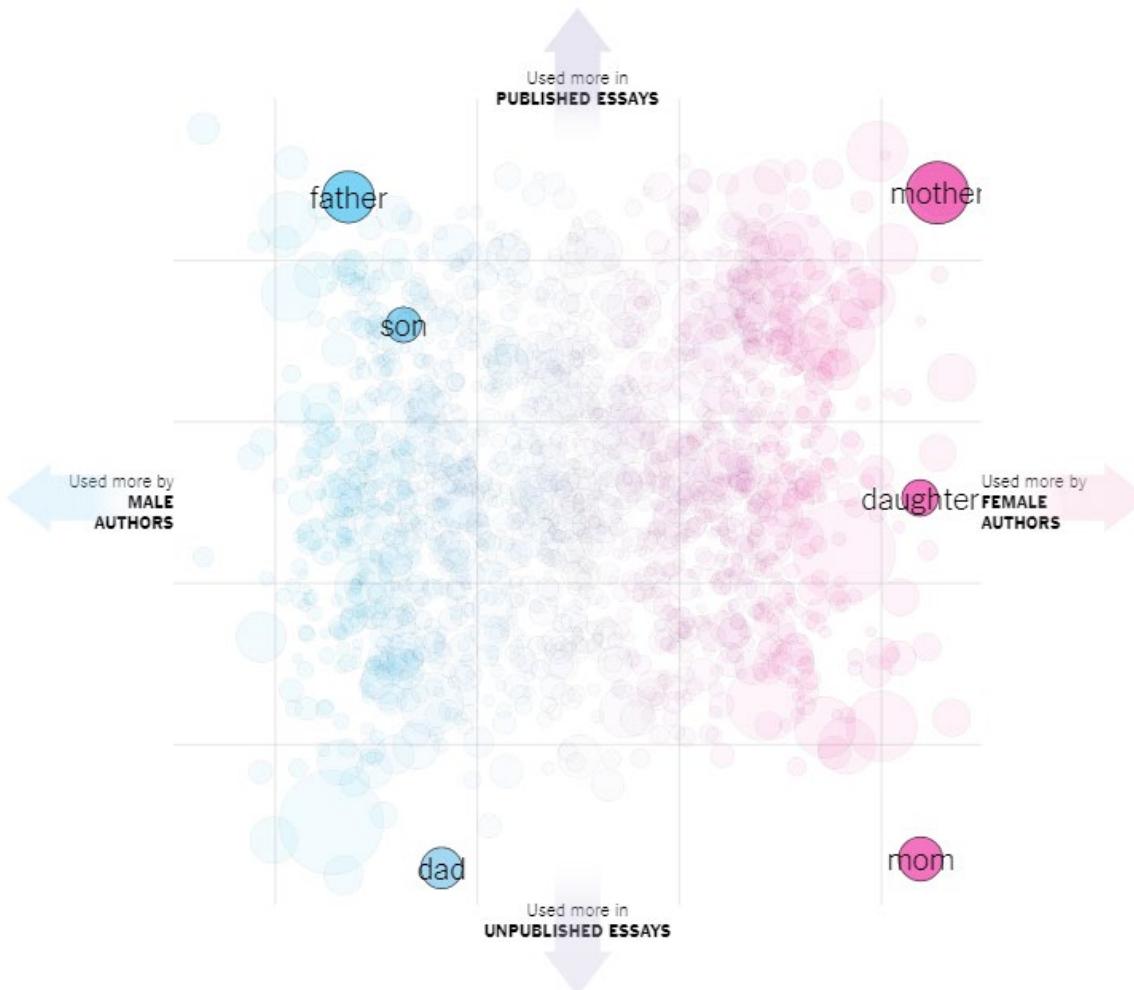
EXAMPLES



The Words Men and Women Use When They Write About Love (2017). The New York Times.

<https://www.nytimes.com/interactive/2017/11/07/upshot/modern-love-what-we-write-when-we-write-about-love.html>

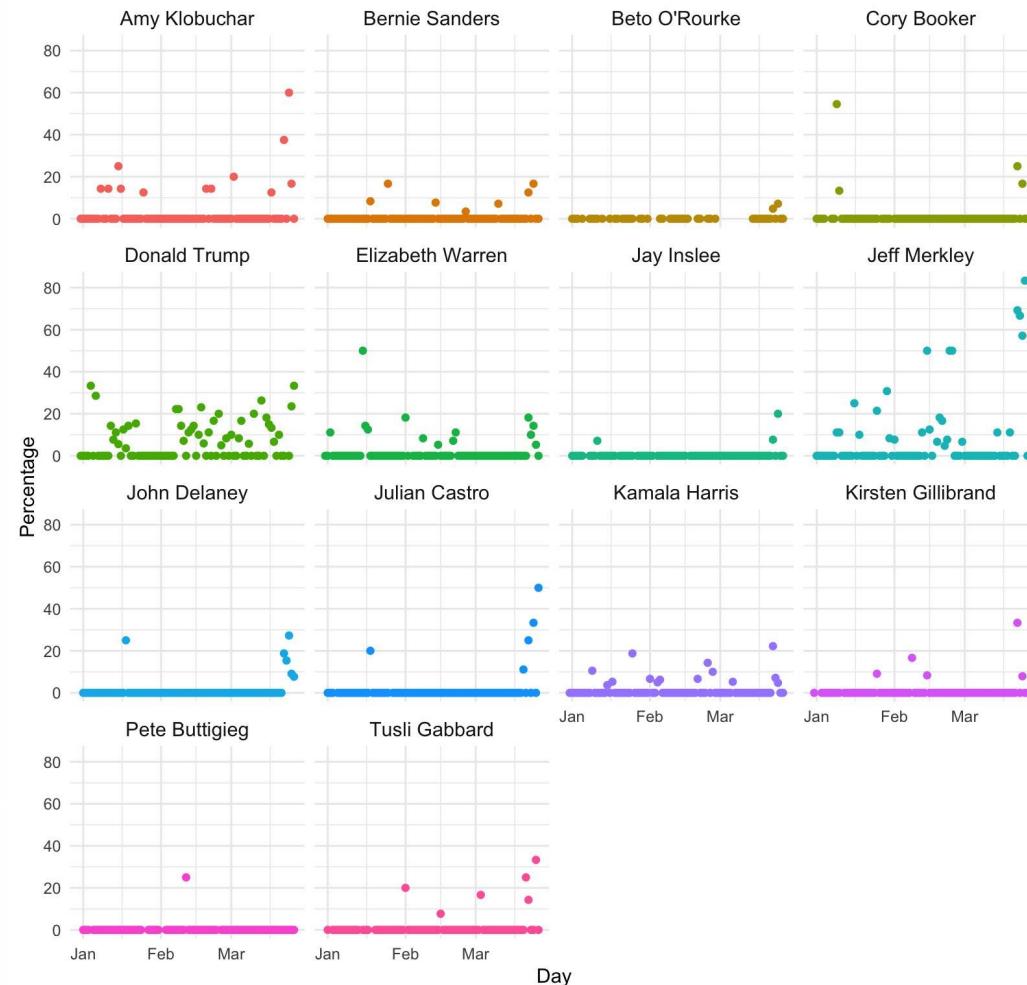
EXAMPLES



The Words Men and Women Use When They Write About Love (2017). *The New York Times*.
<https://www.nytimes.com/interactive/2017/11/07/upshot/modern-love-what-we-write-when-we-write-about-love.html>

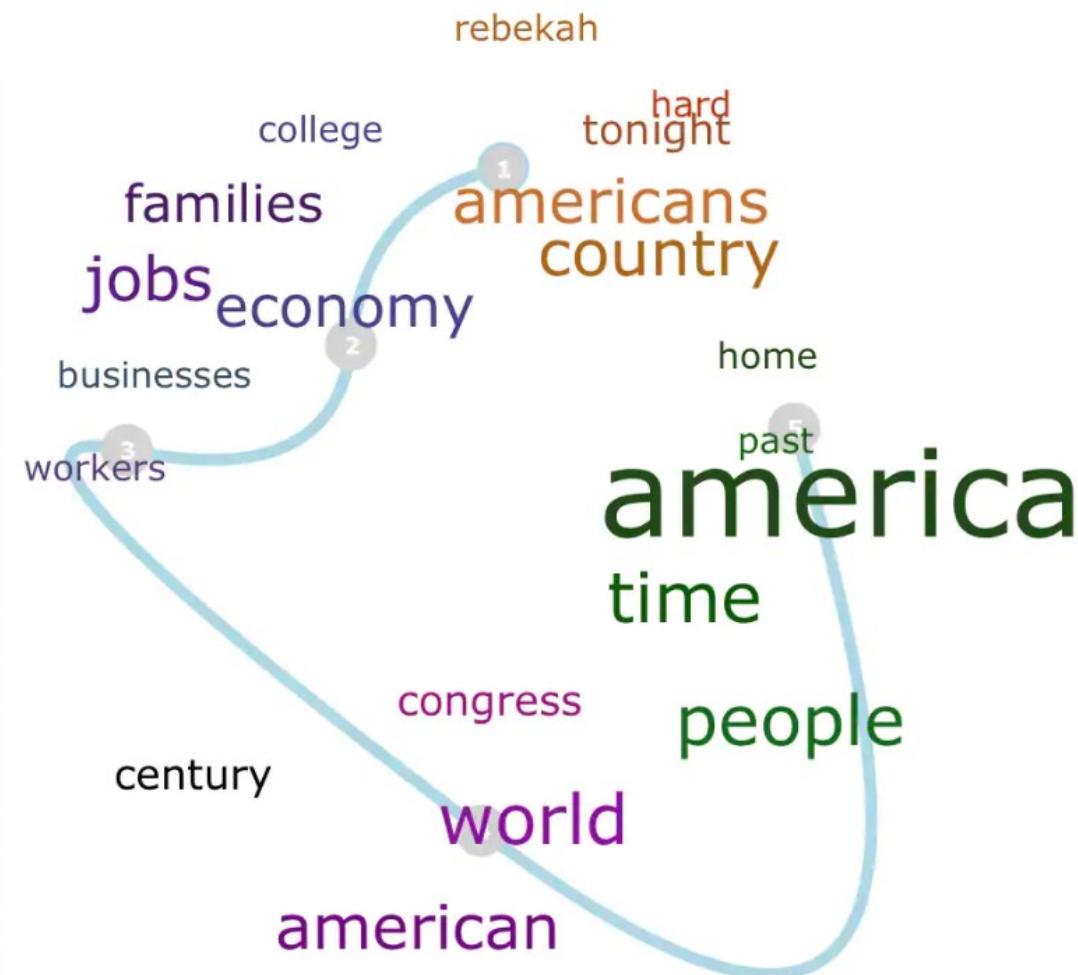
EXAMPLES

Percentage of Tweets Mentioning Russia, Mueller or Impeachment



<https://twitter.com/KevinReuning/status/1110681968331669504/photo/1>

EXAMPLES



The State of the Union address in a single figure (2015). *The Washington Post*.

<https://www.washingtonpost.com/news/monkey-cage/wp/2015/01/21/the-state-of-the-union-address-in-a-single-figure/>

Takeaways



**HOW TO PLAN YOUR DATA
JOURNALISM PROJECT**



**HOW TO GET DATA FROM
THE NYTIMES API USING R**



**HOW TO VISUALIZE DATA
USING R**

HOW TO PLAN A DATA JOURNALISM PROJECT

WHAT'S YOUR QUESTION?

WHAT'S YOUR DATA?

WHAT'S YOUR STORY?

How to plan a Data Journalism project

What is your question?

Hypothesis testing:

- **Traditional media sets the agenda for social media discussions (Drezner & Farrell, 2004)**
- **Intergroup contact on social media reduces prejudice (Pettigrew, 1998)**
- **People follow social media accounts that are congruent with their personal ideology (Barbera, 2015)**

SOCIAL MEDIA & COMPUTER-MEDIATED COMMUNICATION



With social media data, you can study:

- Choices in self-presentation (Gil-Lopez et. al, 2018)
- How user needs have changed and created new gratifications (Sundar & Limperos, 2013)
- Interpersonal relationships (Tong, Heide, Langwell, & Walther, 2008)
- Desirability and perceptions of authenticity (Sundar, 2008; Sydnor, 2018)
- Political ambivalence, participation, and knowledge (Cho, Shah, Eveland, & Kwak, 2005; Song & Eveland, 2015)

How to plan a Data Journalism project

What is your question?

- *What was the public reaction to the Trump – Kim Summit in Singapore?*
- *Did PAP's position on foreign talent change between 2008 and 2016?*
- *How is the public opinion among Singaporeans different from Malaysians re: the maritime dispute?*
- *What makes a K-pop song go viral?*
- *What makes a winning campaign speech?*

How to plan a Data Journalism project

What is the data that will answer your question?

Question	Ideal data/pre-requisites
• <i>What was the public reaction to the Trump – Kim Summit in Singapore?</i>	• Collection of public commentary (when/where/how)
• <i>Did PAP's position on foreign talent change between 2008 and 2016?</i>	• Definition of “position” • Laws on foreign talent from 2008-2016 • Statistics on foreign talent influx • Statistics on jobs
• <i>How is the public opinion in SG vs MY different re: the maritime dispute?</i>	• Collection of public commentary (when/where/how/what)
• <i>What makes a K-pop song go viral?</i>	• Statistics on viral and non-viral K-pop songs • Definition of “virality” • Statistics on campaign speeches
• <i>What makes a winning campaign speech?</i>	• Definition of “winning”

How to plan a Data Journalism project

What is your story?

Question

- *What was the public reaction to the Trump – Kim Summit in Singapore?*

Story



- Did PAP's position on foreign talent change between 2008 and 2016?



- *What makes a winning campaign speech?*



How to plan a Data Journalism project

What is your story?

- *What you found*
- *How you found it*
- *What it means*
- *What it doesn't mean*
- *What you assumed*

IN SUMMARY

WHAT'S YOUR QUESTION?

WHAT'S YOUR DATA?

WHAT'S YOUR STORY?

Takeaways



HOW TO PLAN YOUR DATA
JOURNALISM PROJECT

- WHAT'S YOUR QUESTION?
- WHAT'S YOUR DATA?
- WHAT'S YOUR STORY?



HOW TO GET DATA FROM
THE NYTIMES API USING R



HOW TO VISUALIZE DATA
USING R

HOW TO GET DATA FROM THE NYTIMES API USING R

- EXPLORE API
- CREATE OAUTH TOKENS
- GET JSON

Data Mining from New York Times

- Has an API: <https://developer.nytimes.com/>
- Needs one to create an app:
<https://developer.nytimes.com/my-apps>
- OAuth needs a key
- There's an R library for the API: nytimes; or you can use curl
- There is a way to try it out:
<https://developer.nytimes.com/apis>
- The response is in JSON format.



What kind of data can you pull?

- Real-time feed, older articles and article metadata
- Book and movie reviews
- User comments
- Geographic linked data
- Entities-related and tagged data



Example code

We are now ready to make an authenticated call to NYTIMES.

KEY
SECRET

library(nytimes)
library(jsonlite)

parameters

jsonlite::stream_in(url)

convert to dataframe

plot it

Let's jump into it...

<https://goo.gl/Vr5Vyw>

IN SUMMARY

- EXPLORE API
- CREATE OAUTH TOKENS
- GET JSON

Takeaways



HOW TO PLAN YOUR DATA JOURNALISM PROJECT

- WHAT'S YOUR QUESTION?
- WHAT'S YOUR DATA?
- WHAT'S YOUR STORY?



HOW TO RETRIEVE DATA USING R

- EXPLORE API
- CREATE OAUTH TOKENS
- GET JSON



HOW TO VISUALIZE DATA USING R

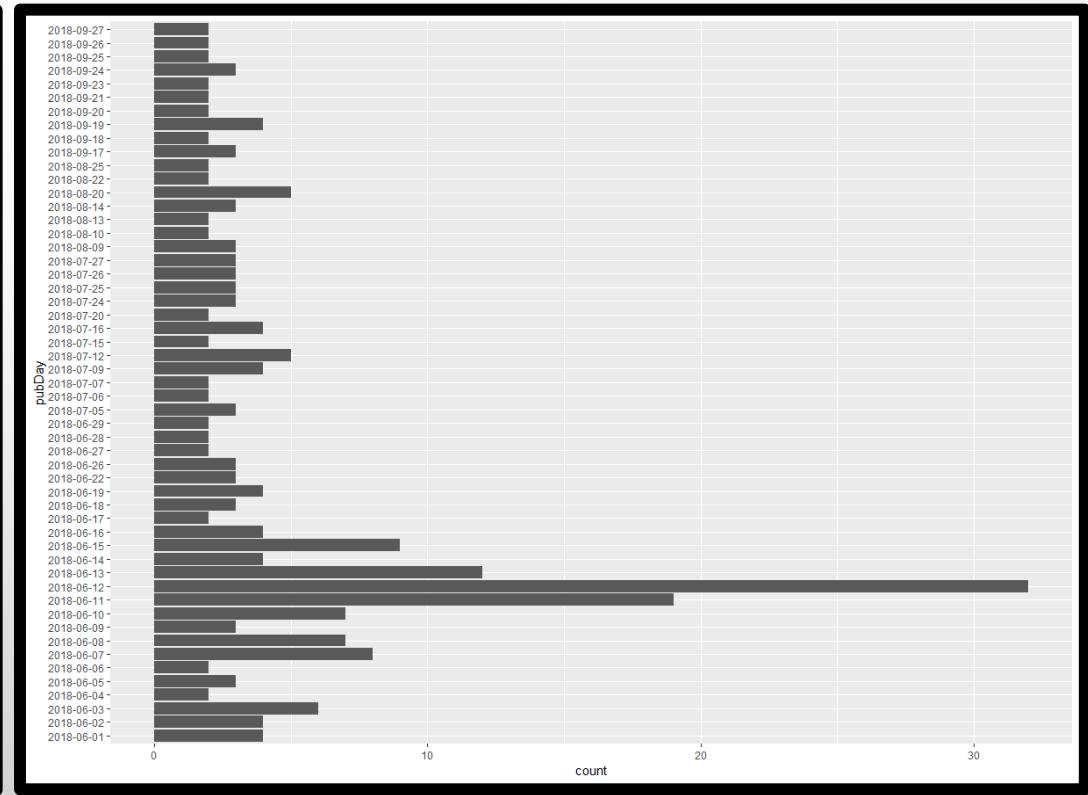
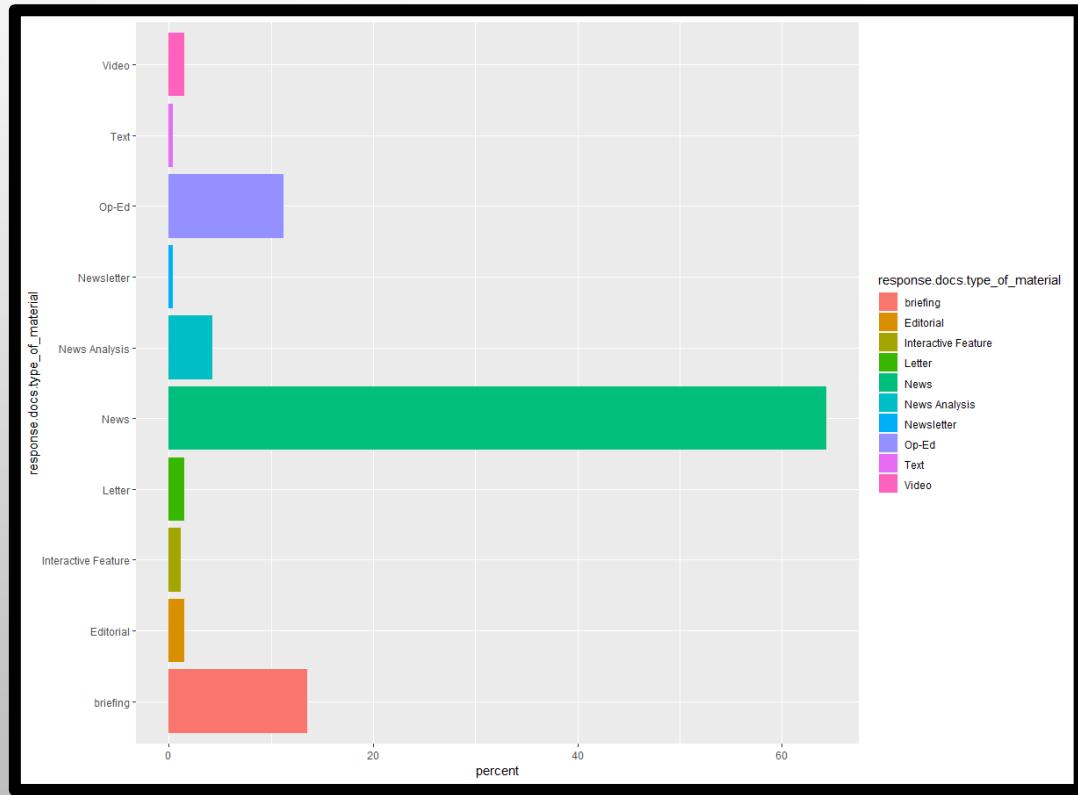
HOW TO VISUALIZE DATA USING R

- TABLEAU
- GGPLOT
- SHINY

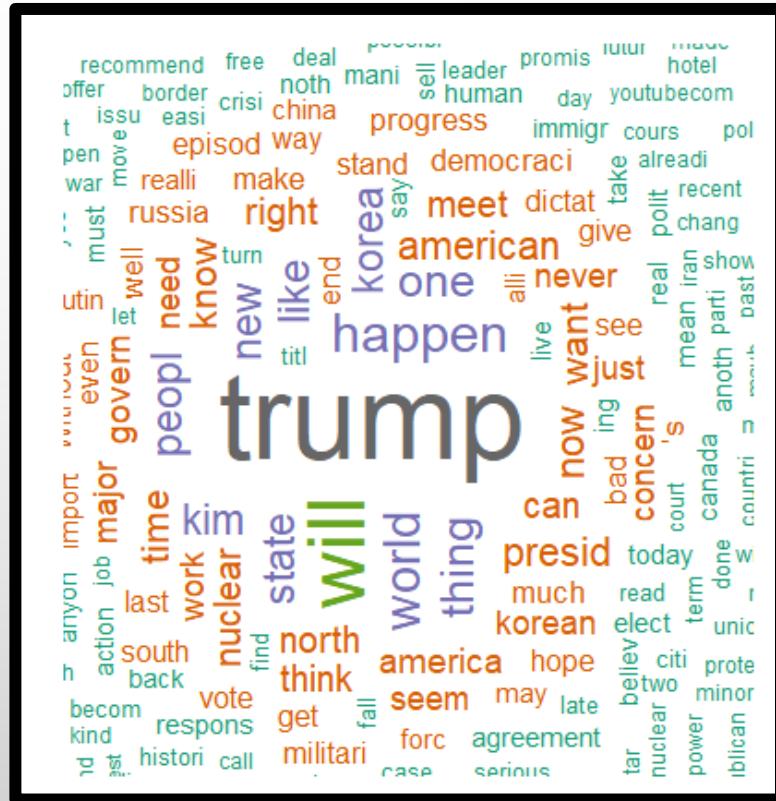
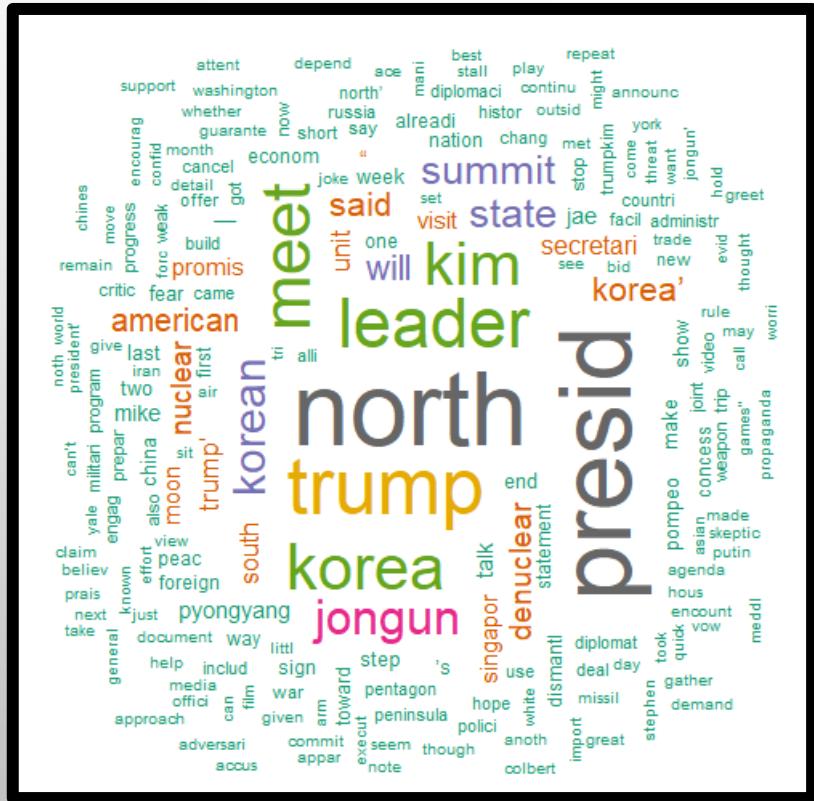
<http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html>

<https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/>

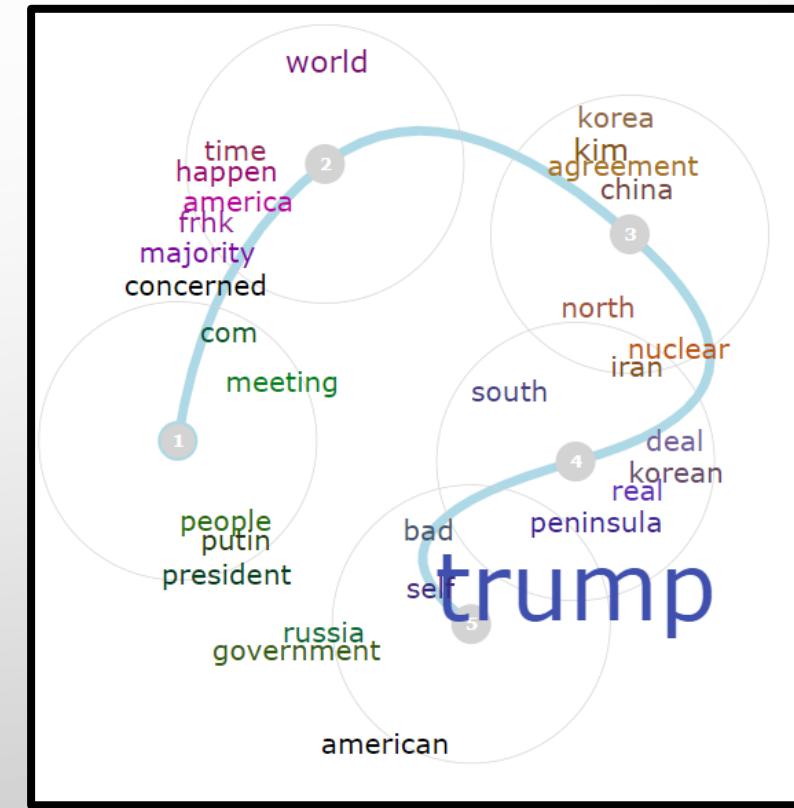
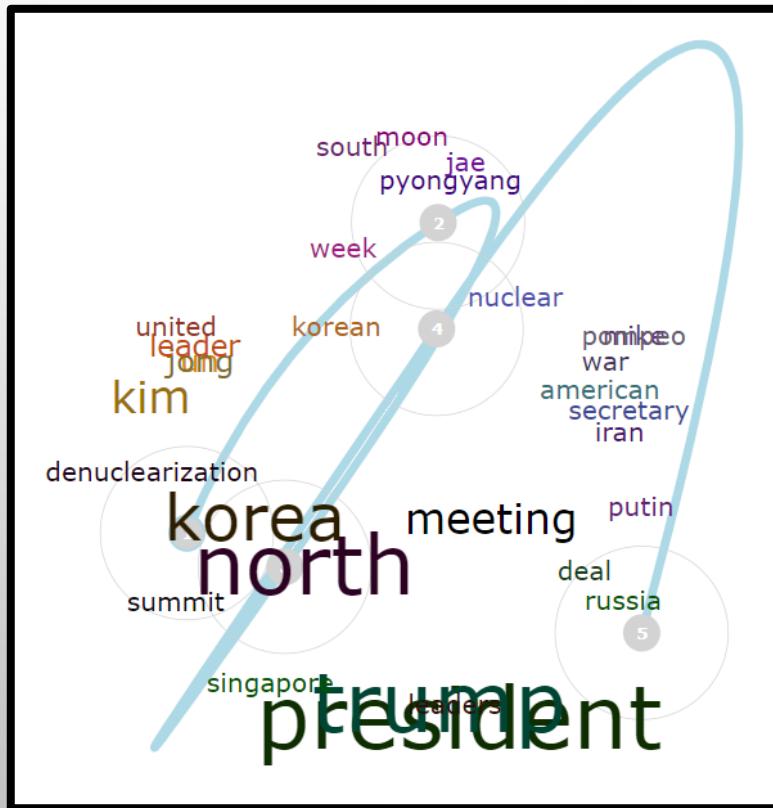
BAR PLOTS



WORD CLOUDS



TRAJECTORIES



HELLO SHINY

<https://www.nytimes.com/interactive/2016/12/26/upshot/duck-dynasty-vs-modern-family-television-maps.html>

```
install.packages("shiny")
library(shiny)
runExample("01_hello")
runExample("06_tabssets")
```

<https://shiny.rstudio.com/tutorial/>

Takeaways



HOW TO PLAN YOUR DATA JOURNALISM PROJECT

- WHAT'S YOUR QUESTION?
- WHAT'S YOUR DATA?
- WHAT'S YOUR STORY?

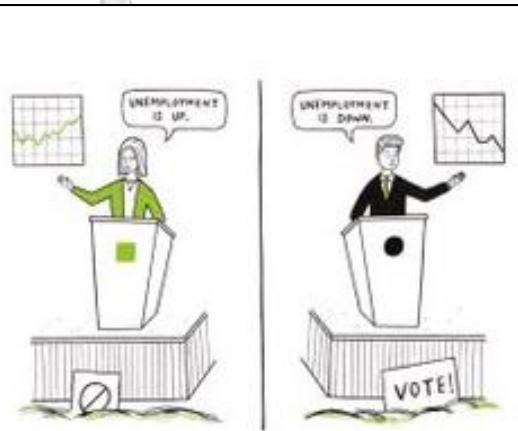
HOW TO RETRIEVE DATA USING R

- EXPLORE API
- CREATE OAUTH TOKENS
- GET JSON

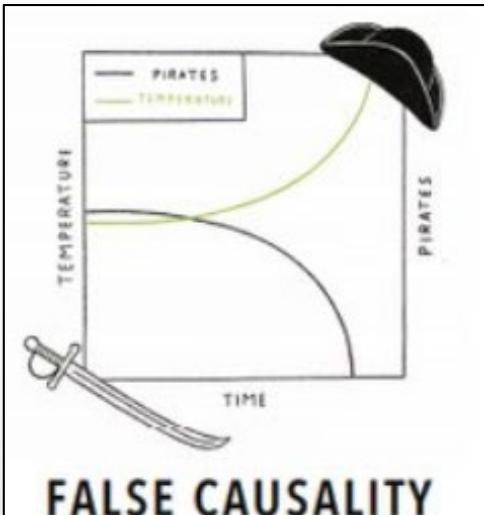
HOW TO VISUALIZE DATA USING R

- TABLEAU
- GGPLOT
- SHINY

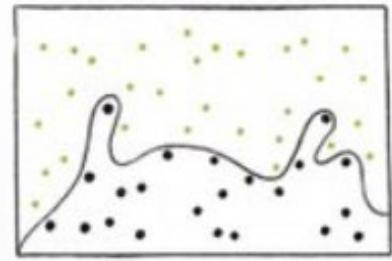
LIMITATIONS



CHERRY PICKING



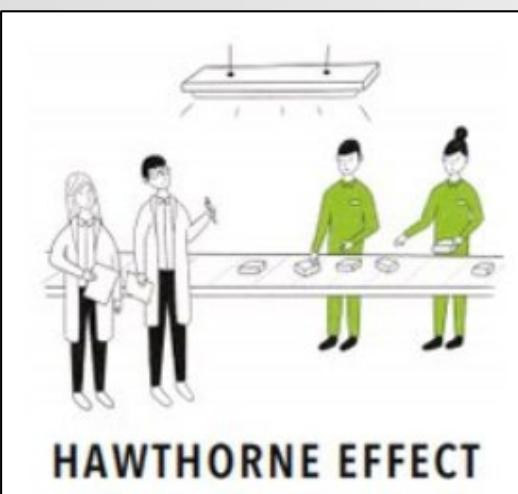
FALSE CAUSALITY



OVERFITTING



SAMPLING BIAS



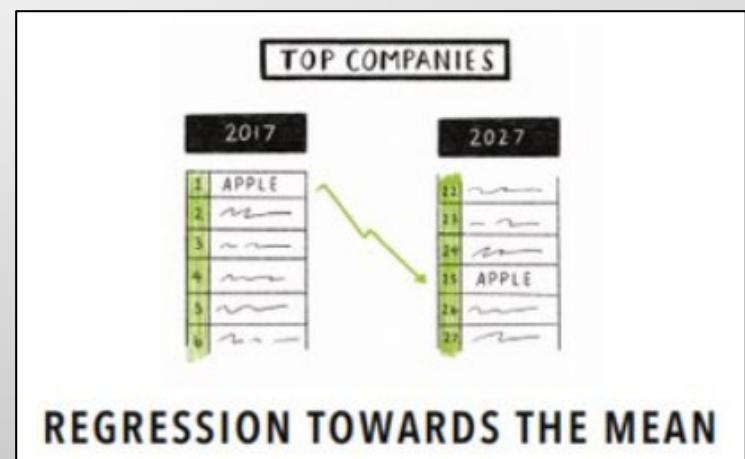
HAWTHORNE EFFECT



SURVIVORSHIP BIAS



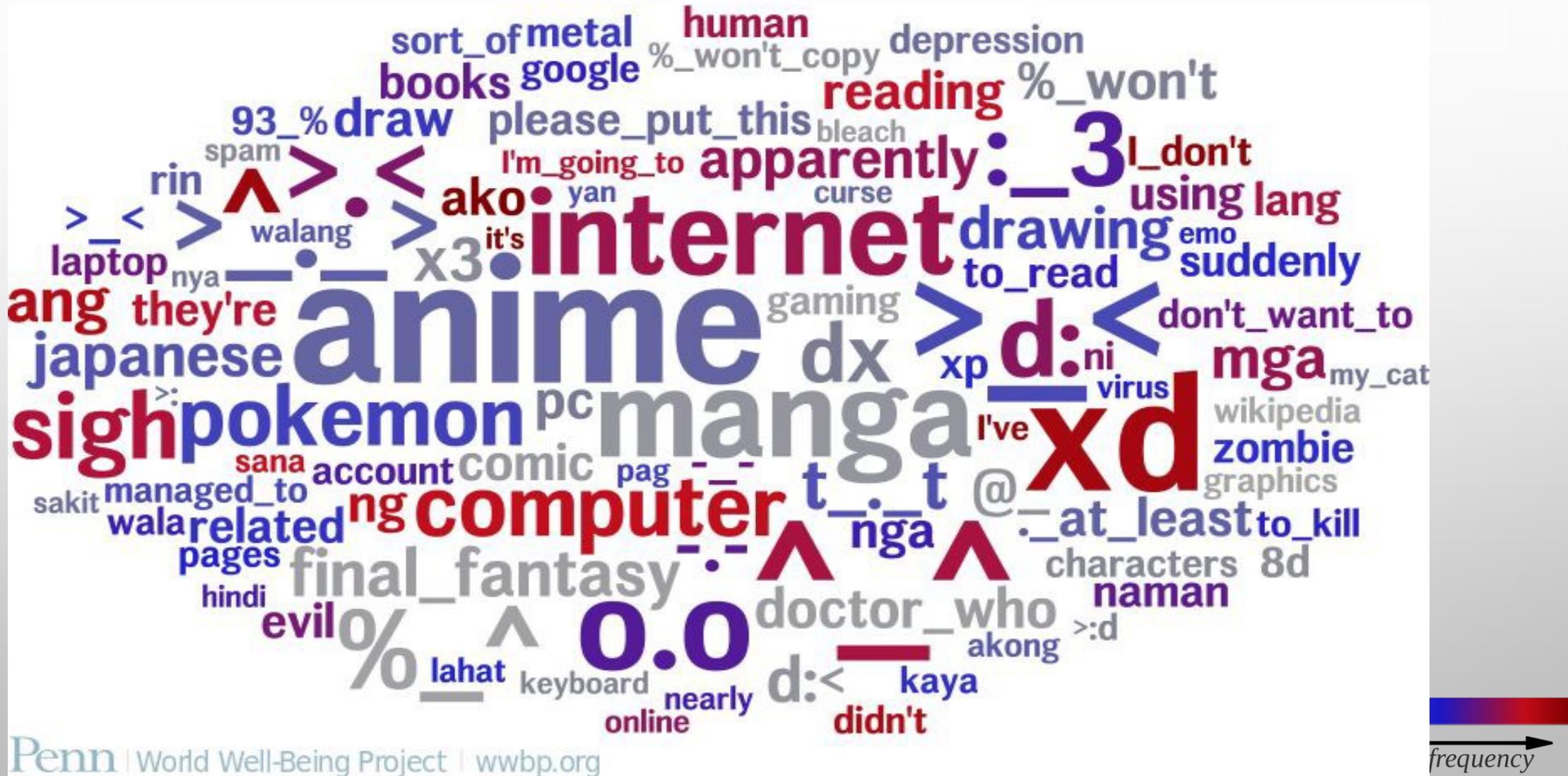
DATA DREDGING



Extraversion -- *sociable, assertive, active, energetic, talkative, outgoing*



Introversion



WORDS FROM FB POSTS ASSOCIATED WITH STRESS

High Stress

weird crap lonely
feels depressed kinda
strange feeling
ill felt bit urge
feel sad :(

stupid
people
annoying
rude
hate they're
hates
dumb smart
idiots idiot piss
act

mistakes
worst
deserve women
handful
handle hard
times bitch won't
strong control
dare admit hell

forgetting
forgotten
you've don't^{spread}
important regret
forget forgive
remember heard
remind happened sake
hearts

mornings haters
sundays hates suck
hate passion
tuesdays absolutely you
ugh mondays
to dislike
hating

worst
massive horrible
headache
attack
suffering
panic worse
hangover
my heart
strong
people
i can't do
talk
speeding
but
kill me
struggle
handle
legs
sore
neck
feet
tummy
killing
bodyhurts
ache
stomach
throat
head
hurt
pain

comfortable depressed
feel crap
pleasure felt starting
terrible does nt makes
guilty accomplished horrible
uncomfortable hits

karaoke dj join
pub drinks drink
cover lounge mic free
adies tonight
bar club tonite

trip shutter
cruise
florida trips adventure mini
roadcampingfield
Vacationplanning
packing **island** mission

feeling **lot** **bla bla** **feels**
motivated **today** **mission**
felt **accomplished**
alot **helped** **productive**
relaxed **lazy** **worked**

afternoon
breakfast accidents in_the airport
excited trip_to tonight's **first**
today six_months super nervous
taking_a_nap **the** this_morning
soooooo excited
bowl soooooo
duper pumped
excited soooo
super sooooooo
soostoked

sandwich nom
eating bacon
egg **breakfast** cheese
sausage pancakes toast
yum french eggs waffles
mmm

gift
present
conquered mystery
greatest army
yesterday opera
worry history
tomorrow today's
today called master

Low Stress

a a a
correlation strength

Gender



Gender



a a a
correlation strength

relative frequency

b b b
prevalence in topic

New Software and Tools

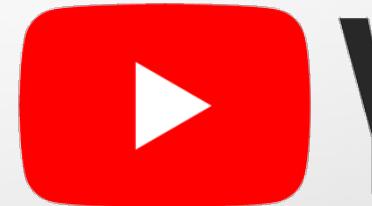
- [SocialMediaLab](#) – Timothy John Graham, The Australian National U; Robert Ackland, Australian National U:
http://vosonlab.net/papers/SocialMediaLab/SocialMediaLab_tutorial_ICA2017_Graham_and_Ackland.pdf
- **[Corpustools](#): An R Package for Text Analysis Beyond Bags of Words** – Kasper Welbers, U of Leuven; Wouter van Atteveldt, UvA
- **JamCode** – Christian Baden, Hebrew U of Jerusalem
- **Automatic Text Analysis Made Easy: Using [AmCAT](#), [NLPipe](#), and [R](#)** – Wouter van Atteveldt and others, U of Amsterdam
- **Facebook Page Data Extraction for Nonprogrammers: Introducing the [Netvizz](#) and [Facepager](#) Tools** – U of Amsterdam, Ernst-Moritz-Arndt-Universität Greifswald
- Lexhub - <http://lexhub.org/> - University of Pennsylvania
- NodeXL - <https://nodelx.codeplex.com/> - Social Media Research Foundation
- Tweet Archivist - <https://www.tweetarchivist.com/> -
https://www.youtube.com/watch?v=4D6_kwY7nZQ

Recap: Data Mining from Twitter

- Has an API
- Needs one to create an app: <https://apps.twitter.com>
- OAuth needs keys – and an access token
- There's an R library for the API: *twitteR*
- There is a way to try it out: <https://dev.twitter.com/rest/tools/console>
- The response is in JSON format.

Data Mining from YouTube

- Has an API:
<https://developers.google.com/youtube/v3/docs/>
- Needs one to create an app:
<https://console.developers.google.com/>
- OAuth needs keys – and an access token
- There's an R library for the API: tuber; or you can use curl
- There is a way to try it out:
<https://developers.google.com/apis-explorer/>
- The response is in JSON format.



YouTube

<https://developers.google.com/youtube/v3/getting-started>

<https://cran.r-project.org/web/packages/googlesheets/vignettes/managing-auth-tokens.html>

Data Mining from Facebook

- Has an API: <https://developers.facebook.com/docs/graph-api>
- Needs one to create an app: <https://developers.facebook.com/apps>
- OAuth needs keys – need an access token
- There is R library for the API: facebook.github.io/R/
- There is a way to try it out: <https://developers.facebook.com/tools/explorer/>
- The response is in JSON format

USE NETVIZZ

Netvizz

<https://apps.facebook.com/netvizz/>. Must have Facebook account.

The screenshot shows the Netvizz app interface with the following settings:

- Facebook Page ID:** A yellow box highlights the input field containing "153080620724".
- page id:** Input field with value "153080620724" (find page ids [here](#) or through Netvizz' search module).
- date scope:** Radio button selected for "posts between 2016-10-01 and 2016-11-07". A yellow box highlights the text "What posts to extract".
- data to get:** Radio button selected for "post statistics only (post metrics, stats per day and fans per country)".
- get post by page only or posts by page and users:** Input field with value "post by page only". A yellow box highlights the text "Data to output".

Video: <https://youtu.be/3vkKPcN7V7Q?t=8m39s>



Netvizz – Let's do this!

<https://apps.facebook.com/netvizz/>. Must have Facebook account.

https://apps.facebook.com/netvizz/?fb_source=search

148926458545119

FacePager

<https://github.com/strohne/Facepager> Must have Facebook account.

The screenshot shows the Facepager 3.7 application window. The top menu bar includes: Open Database, New Database, Export Data, Add Nodes, Delete Nodes, Presets, and Help. Below the menu is a toolbar with: Expand nodes, Collapse nodes, Copy Node(s) to Clipboard, Add Column, Add All Columns, Unpack List, and Copy JSON to Clipboard. The main interface consists of several panes:

- Object ID List:** A tree view showing a hierarchy of Facebook objects, primarily posts, under a seed node.
- Table View:** A grid showing extracted data with columns: Key and Value. Examples include: created_time (2015-11-09T17:20:13+0000), message (Hello, Facebook! I finally got my very own page. I hope you'll think of this as a place where ...), and id (424207551102424_425063411016838).
- Custom Table Columns:** A section titled "Custom Table Columns (one key per line)" containing the keys: id, created_time, and message.
- Data Request Commands:** A yellow box highlighting the "Data request commands" section in the bottom-left, which contains fields for Base path (https://graph.facebook.com/v2.6/), Resource (<user>/posts), Parameters (<user> <Object ID>), Maximum pages (100), Client Id, Scope (user_groups), and Access Token (redacted). It also includes a "Login to Facebook" button and a "Fetch Data" button.
- Status Log:** A log of API requests and responses, showing a series of fetches for the specified user and post IDs.

Annotations with yellow boxes highlight specific features:

- User Facebook credentials:** Points to the Access Token field in the "Data request commands" section.
- Customize field output to database file:** Points to the "Custom Table Columns" section.



Ernst Moritz Arndt
Universität Greifswald

Username testuserkw

Password Nz5INR7

<https://amcat.nl/accounts/login/>

<http://wiki.amcat.nl/3.4:Querying>

Found 758 articles in 0.084 seconds.

Experts Find No Way To Implement Trump's Wall
20-02-2016 | 1029 words | 1 Section A, Column C, National Deal, Pg. 1
... : Big, beautiful and impractical Mr. Trump has promised that the wall will be big, beautiful, tall ... bringing drugs and crime. And, yes, Mexico will pay for the Great Wall of Trump , as he has called it. But ... ambitious than I was imagining." The most common benchmark used for assessing Mr. Trump's wall is ...

Trump in New York: Deep Roots, but Little Sway
24-02-2016 | 1029 words | 1 Section A, Column C, National Deal, Pg. 1
... 's organization currently has no office in New York City, and the campaign has been forced to deal with is Wall Street. Here bank executives ... ending to him. "I don't Need Money" Mr. Trump says he has no need for banks. "I don't use Wall ...

Trump holds firm on immigration
01-03-2016 | 729 words | 1 NEWS, Pg. 1A

... from the beginning. "We will build a great wall along the southern border," said Trump, who was ... right away all they did not deserve payment for the wall , though Peña Nieto said later in a speech in Mexico City that he would help finance it. "We're not going to pay for the wall," Trump said.

2 Ex-Presidents of Mexico Say No Way They're Paying for Donald Trump's Wall
25-02-2016 | 381 words | 1 U.S. politics

... HIGHLIGHT: Vicente Fox and Felipe Calderón have a message for Mr. Trump : You want a wall on ... the border? You pay for it.. At almost all of his campaign rallies, Donald Trump leads a familiar ... refrain as he details his plan to build a giant wall along the border with Mexico. "Who's gonna pay ...

● Complex keyword queries

Trump Boasts of Rapport With Wall St., but It's Not Mutual

24-02-2016 | 1657 words | 1 Section B, Column D, Business/Financial Desk, Pg. 1

... By Donald Trump's reckoning, his relationship with Wall Street could not be better. "I am friends" If there were any prevailing doubts of his stature on Wall Street, Mr. Trump said the chief ... Goldman Sachs, drawing fire from her opponents. Part of the reason Mr. Trump has fewer recent ...

Donald Trump's for ...
16-02-2016 | 446 words | 1 NEWS

... Donald Trump is vowing to build a wall along the U.S. Southern border with Mexico, likening it to ... the Great Wall of China and even dubbing his proposed border barrier "The Great Wall of Trump ." It's ... debate. Yet, just as the Ming Dynasty's 13,000-mile wall failed to keep out the Manchurians, Trump's ...

Why Wall Street is yawning over Trump's big plan

10-02-2016 | 617 words | 1 MONEY

... driven sell-off might be far from over, according to ... That's because the plan has been ... are some theones: "Trump is a long shot" "I do think Wall Street is beginning to factor in a ... Political risk and election uncertainty have been cited as twin worries by Wall Street, but the ...

Border Security: 2 Former Presidents ...
25-02-2016 | 391 words | 1 Section A, Column C, National Deal, Pg. 1

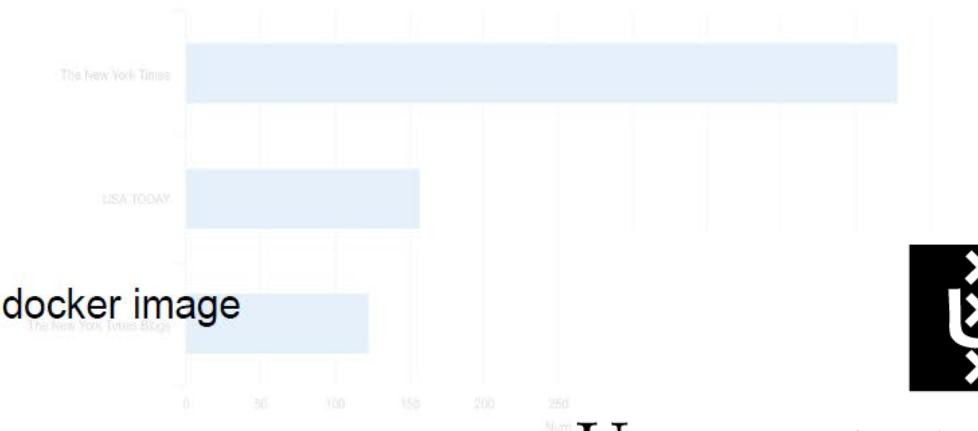
... At almost all of his campaign rallies, Donald Trump has insisted that Mexico will pay for a wall along the border with the United States. "Mexico ... " Who's gonna pay?" he asks. Without ... Mr. Trump's strong support, the U.S. political establishment has said little about the ...

Trump Gambles on immigration in Trip and Rally

01-03-2016 | 1603 words | 1 Section A, Column C, National Deal, Pg. 1

... made it clear that Mexico will not pay for the wall . " Mr. Peña Nieto did not dispute Mr. Trump at ... — 100 percent — they don't know it yet, but they will pay for the wall ." Mr. Trump said. "They ... 're great people, and great leaders, but they will pay for the wall ." Mr. Trump had billed the ...

Transcript of the Republican Presidential Debate in Houston
06-08-2015 | 100 words | 1 NEWS



AmCAT

Articles Query Coding Advanced

AmCAT Navigator
3.4.2rc1

Search



Projects

Annotator

Wouter van Atteveldt

Help

Found 758 articles in 0.084 seconds.

Experts Find Flaws in Trump Plan for Wall

20-05-2016 | 2658 words | - | Section A; Column 0; National Desk; Pg. 1

... : Big, beautiful and impractical Mr. Trump has promised that the wall will be big, beautiful, tall ... bringing drugs and crime. And, yes, Mexico will pay for the Great Wall of Trump , as he has called it. But ... ambitious than I was imagining." The most common benchmark used for assessing Mr. Trump's wall is ...

Trump in New York: Deep Roots, but Little Sway

24-02-2016 | 1926 words | - | Section A; Column 0; National Desk; Pg. 1

... 's organization currently owns 40 Wall Street; the commercial space at Trump Tower at 725 Fifth ... New York institution Mr. Trump has been forced to deal with is Wall Street. Here bank executives ... lending to him. "I Don't Need Money" Mr. Trump says he has no need for banks. "I don't use Wall ...

Trump holds firm on immigration; Republican candidate says there is only one core issue: 'Well-being'

01-09-2016 | 729 words | - | NEWS; Pg. 1A

... from the beginning. "We will build a great wall along the southern border," said Trump , who was ... right. Trump said they did not discuss payment for the wall , though Peña Nieto said later in a post to ... Donald Trump clarified his views on immigration in a speech Wednesday in Phoenix, saying the ...

2 Ex-Presidents of Mexico Say No Way They're Paying for Donald Trump's Wall

25-02-2016 | 368 words | - | US; politics

... HIGHLIGHT: Vicente Fox and Felipe Calderón have a message for Mr. Trump : You want a wall on ... the border? You pay for it. At almost all of his campaign rallies, Donald J. Trump leads a familiar ... refrain as he details his plan to build a giant wall along the border with Mexico. "Who's gonna pay ...

Trump Boasts of Rapport With Wall St., but It's Not Mutual

24-05-2016 | 1659 words | - | Section B; Column 0; Business/Financial Desk; Pg. 1

... By Donald Trump's reckoning, his relationship with Wall Street could not be better. "I am friends" If there were any prevailing doubts of his stature on Wall Street, Mr. Trump said the chief ... Goldman Sachs, drawing fire from her opponents. Part of the reason Mr. Trump has fewer recent ...

Donald Trump calls for 'Great Wall,' but even China's couldn't secure border

16-09-2015 | 486 words | - | NEWS; Pg. 5A

... Donald Trump is vowing to build a wall along the U.S. Southern border with Mexico, likening it to ... the Great Wall of China and even dubbing his proposed border barrier "The Great Wall of Trump ." It's ... debate. Yet, just as the Ming Dynasty's 13,000-mile wall failed to keep out the Manchurians, Trump's ...

Why Wall Street is yawning over Trump's big plan

10-08-2016 | 617 words | - | MONEY; Pg. 1B

... driven sell-off might be overblown. Wall Street's reaction to Donald Trump's economic plan has been ... are some theories: Trump is a long shot "I do think Wall Street is beginning to factor in a ... Political risk and election uncertainty have been cited as twin worries by Wall Street, but the ...

Border Security: 2 Former Presidents of Mexico Say Country Won't Pay for Trump's Wall

26-02-2016 | 390 words | - | Section A; Column 0; National Desk; FIRST DRAFT; Pg. 17

... At almost all of his campaign rallies, Donald J. Trump leads a familiar refrain as he details his ... plan to build a giant wall along the border with Mexico. "Who's gonna pay?" he asks. Without fail ... , Mr. Trump's adoring fans reply, "Mexico!" While the Mexican government has said little about Mr ...

Trump Gambles on Immigration in Trip and Rally

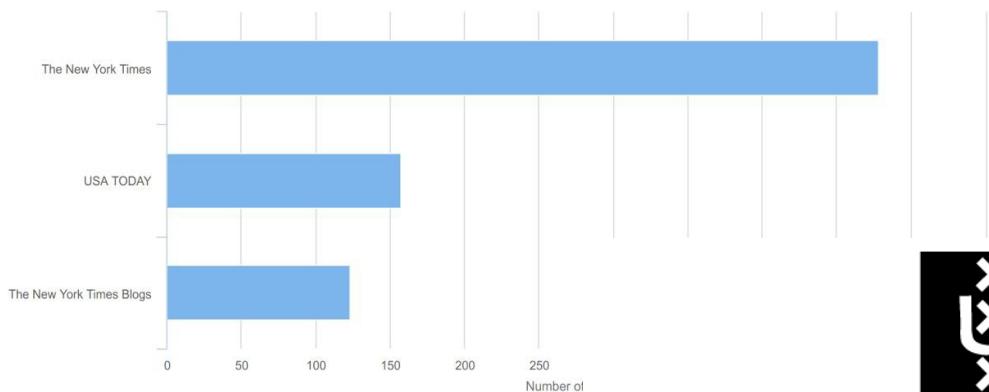
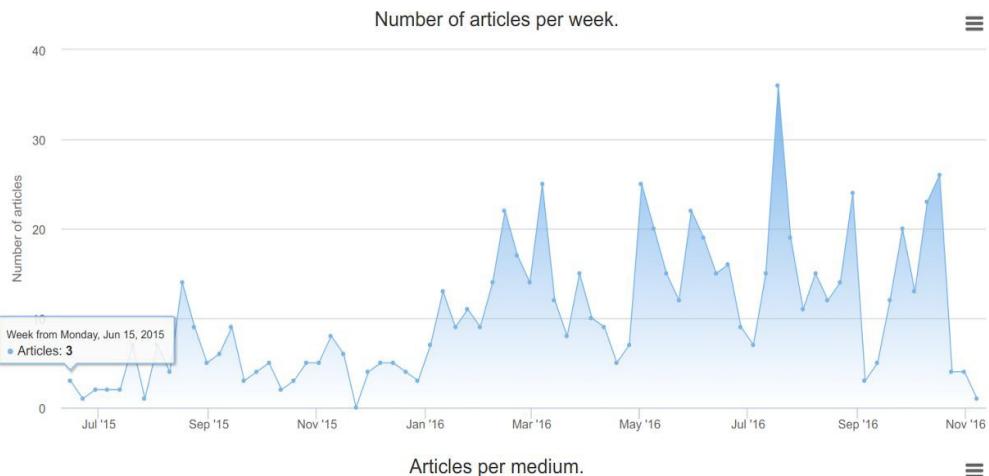
01-09-2016 | 1603 words | - | Section A; Column 0; National Desk; Pg. 1

... made it clear that Mexico will not pay for the wall ." Mr. Peña Nieto did not dispute Mr. Trump at ... -- 100 percent -- they don't know it yet, but they will pay for the wall , " Mr. Trump said. "They ... 're great people, and great leaders, but they will pay for the wall ." Mr. Trump had billed the ...

Transcript of the Presidential Debate in Houston

26-02-2016 | 23686 words | - | Section ; Column 0; Politics; Pg.

... building a wall along the southern border. TRUMP : (inaudible). BLITZER: It's about 315 miles ... the wall ? TRUMP: I will, and the wall just got 10



AmCAT

An API for power users

- All functionality available through API
- Use python/R to manage and analyse data
- Scrape and upload articles
 - Conduct automatic queries
 - Download text, metadata
 - Upload new article sets
 - Create projects, users, etc.
- Workflow:
 - Corpus/project management and explorative analysis using website
 - Reproducible queries using API
 - Download text or results and connect to other tools (topicmodels etc)

The screenshot shows an RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and a Project dropdown set to (None). The left sidebar has tabs for options.R, data.R, cndw.R, Untitled2.R, Untitled3.R, jiebo.R, Untitled4.R, Run, and Source. The main workspace shows R code for connecting to the AmCAT API, loading datasets, and creating a DocumentTermMatrix (dtm_a). The Environment pane lists variables like r, r_a, r_b, months, app_id, conn, and dtm_a. The Plots pane displays a word cloud with large words like 'hobben', 'dat', 'van', 'het', and 'nietvoor'. The bottom right corner features the University of Amsterdam logo.

```
library(amcat)
conn = amcat.connect("https://amcat.nl")
load("./Downloads/dataR.RData")
bigor2 = bigor
bigor = head(bigor)
bigor = subset(bigor2, date != "NA")
bigor$date2 = gsub(" d.", "T", bigor$date)
bigor$date2 = gsub(" ", " ", bigor$date2)
months = list("january", "february", "march", "april", "may", "june", "july", "august", "september", "october", "november", "december")
for (m in names(months)) bigor$date2 = gsub(paste0(m, " "), paste0(months[[m]], "T"), bigor$date2)
bigor$date2 = as.POSIXct(bigor$date2, format = "%Y-%m-%dT%T")
bigor$head = head(bigor, n = 100000)
amcat.upload(bigor, conn, title = "AmCAT test dataset", date = bigor$date2, provenance = NULL)
d = data.frames(domains(rep("x", 5), "y", "y"), headlines=c("a", "a", "b", "b", "a", "a"))
d$domhead = paste(d$domain, d$headline, sep = " ")
d$keep = 0
dtm_a = matrix(0, nrow = 100000, ncol = 100000)
dtm_a[1, 1] = 1
dtm_a[1, 2] = 1
dtm_a[1, 3] = 1
dtm_a[1, 4] = 1
dtm_a[1, 5] = 1
dtm_a[1, 6] = 1
dtm_a[1, 7] = 1
dtm_a[1, 8] = 1
dtm_a[1, 9] = 1
dtm_a[1, 10] = 1
dtm_a[1, 11] = 1
dtm_a[1, 12] = 1
dtm_a[1, 13] = 1
dtm_a[1, 14] = 1
dtm_a[1, 15] = 1
dtm_a[1, 16] = 1
dtm_a[1, 17] = 1
dtm_a[1, 18] = 1
dtm_a[1, 19] = 1
dtm_a[1, 20] = 1
dtm_a[1, 21] = 1
dtm_a[1, 22] = 1
dtm_a[1, 23] = 1
dtm_a[1, 24] = 1
dtm_a[1, 25] = 1
dtm_a[1, 26] = 1
dtm_a[1, 27] = 1
dtm_a[1, 28] = 1
dtm_a[1, 29] = 1
dtm_a[1, 30] = 1
dtm_a[1, 31] = 1
dtm_a[1, 32] = 1
dtm_a[1, 33] = 1
dtm_a[1, 34] = 1
dtm_a[1, 35] = 1
dtm_a[1, 36] = 1
dtm_a[1, 37] = 1
dtm_a[1, 38] = 1
dtm_a[1, 39] = 1
dtm_a[1, 40] = 1
dtm_a[1, 41] = 1
dtm_a[1, 42] = 1
dtm_a[1, 43] = 1
dtm_a[1, 44] = 1
dtm_a[1, 45] = 1
dtm_a[1, 46] = 1
dtm_a[1, 47] = 1
dtm_a[1, 48] = 1
dtm_a[1, 49] = 1
dtm_a[1, 50] = 1
dtm_a[1, 51] = 1
dtm_a[1, 52] = 1
dtm_a[1, 53] = 1
dtm_a[1, 54] = 1
dtm_a[1, 55] = 1
dtm_a[1, 56] = 1
dtm_a[1, 57] = 1
dtm_a[1, 58] = 1
dtm_a[1, 59] = 1
dtm_a[1, 60] = 1
dtm_a[1, 61] = 1
dtm_a[1, 62] = 1
dtm_a[1, 63] = 1
dtm_a[1, 64] = 1
dtm_a[1, 65] = 1
dtm_a[1, 66] = 1
dtm_a[1, 67] = 1
dtm_a[1, 68] = 1
dtm_a[1, 69] = 1
dtm_a[1, 70] = 1
dtm_a[1, 71] = 1
dtm_a[1, 72] = 1
dtm_a[1, 73] = 1
dtm_a[1, 74] = 1
dtm_a[1, 75] = 1
dtm_a[1, 76] = 1
dtm_a[1, 77] = 1
dtm_a[1, 78] = 1
dtm_a[1, 79] = 1
dtm_a[1, 80] = 1
dtm_a[1, 81] = 1
dtm_a[1, 82] = 1
dtm_a[1, 83] = 1
dtm_a[1, 84] = 1
dtm_a[1, 85] = 1
dtm_a[1, 86] = 1
dtm_a[1, 87] = 1
dtm_a[1, 88] = 1
dtm_a[1, 89] = 1
dtm_a[1, 90] = 1
dtm_a[1, 91] = 1
dtm_a[1, 92] = 1
dtm_a[1, 93] = 1
dtm_a[1, 94] = 1
dtm_a[1, 95] = 1
dtm_a[1, 96] = 1
dtm_a[1, 97] = 1
dtm_a[1, 98] = 1
dtm_a[1, 99] = 1
dtm_a[1, 100] = 1
dtm_a[1, 101] = 1
dtm_a[1, 102] = 1
dtm_a[1, 103] = 1
dtm_a[1, 104] = 1
dtm_a[1, 105] = 1
dtm_a[1, 106] = 1
dtm_a[1, 107] = 1
dtm_a[1, 108] = 1
dtm_a[1, 109] = 1
dtm_a[1, 110] = 1
dtm_a[1, 111] = 1
dtm_a[1, 112] = 1
dtm_a[1, 113] = 1
dtm_a[1, 114] = 1
dtm_a[1, 115] = 1
dtm_a[1, 116] = 1
dtm_a[1, 117] = 1
dtm_a[1, 118] = 1
dtm_a[1, 119] = 1
dtm_a[1, 120] = 1
dtm_a[1, 121] = 1
dtm_a[1, 122] = 1
dtm_a[1, 123] = 1
dtm_a[1, 124] = 1
dtm_a[1, 125] = 1
dtm_a[1, 126] = 1
dtm_a[1, 127] = 1
dtm_a[1, 128] = 1
dtm_a[1, 129] = 1
dtm_a[1, 130] = 1
dtm_a[1, 131] = 1
dtm_a[1, 132] = 1
dtm_a[1, 133] = 1
dtm_a[1, 134] = 1
dtm_a[1, 135] = 1
dtm_a[1, 136] = 1
dtm_a[1, 137] = 1
dtm_a[1, 138] = 1
dtm_a[1, 139] = 1
dtm_a[1, 140] = 1
dtm_a[1, 141] = 1
dtm_a[1, 142] = 1
dtm_a[1, 143] = 1
dtm_a[1, 144] = 1
dtm_a[1, 145] = 1
dtm_a[1, 146] = 1
dtm_a[1, 147] = 1
dtm_a[1, 148] = 1
dtm_a[1, 149] = 1
dtm_a[1, 150] = 1
dtm_a[1, 151] = 1
dtm_a[1, 152] = 1
dtm_a[1, 153] = 1
dtm_a[1, 154] = 1
dtm_a[1, 155] = 1
dtm_a[1, 156] = 1
dtm_a[1, 157] = 1
dtm_a[1, 158] = 1
dtm_a[1, 159] = 1
dtm_a[1, 160] = 1
dtm_a[1, 161] = 1
dtm_a[1, 162] = 1
dtm_a[1, 163] = 1
dtm_a[1, 164] = 1
dtm_a[1, 165] = 1
dtm_a[1, 166] = 1
dtm_a[1, 167] = 1
dtm_a[1, 168] = 1
dtm_a[1, 169] = 1
dtm_a[1, 170] = 1
dtm_a[1, 171] = 1
dtm_a[1, 172] = 1
dtm_a[1, 173] = 1
dtm_a[1, 174] = 1
dtm_a[1, 175] = 1
dtm_a[1, 176] = 1
dtm_a[1, 177] = 1
dtm_a[1, 178] = 1
dtm_a[1, 179] = 1
dtm_a[1, 180] = 1
dtm_a[1, 181] = 1
dtm_a[1, 182] = 1
dtm_a[1, 183] = 1
dtm_a[1, 184] = 1
dtm_a[1, 185] = 1
dtm_a[1, 186] = 1
dtm_a[1, 187] = 1
dtm_a[1, 188] = 1
dtm_a[1, 189] = 1
dtm_a[1, 190] = 1
dtm_a[1, 191] = 1
dtm_a[1, 192] = 1
dtm_a[1, 193] = 1
dtm_a[1, 194] = 1
dtm_a[1, 195] = 1
dtm_a[1, 196] = 1
dtm_a[1, 197] = 1
dtm_a[1, 198] = 1
dtm_a[1, 199] = 1
dtm_a[1, 200] = 1
dtm_a[1, 201] = 1
dtm_a[1, 202] = 1
dtm_a[1, 203] = 1
dtm_a[1, 204] = 1
dtm_a[1, 205] = 1
dtm_a[1, 206] = 1
dtm_a[1, 207] = 1
dtm_a[1, 208] = 1
dtm_a[1, 209] = 1
dtm_a[1, 210] = 1
dtm_a[1, 211] = 1
dtm_a[1, 212] = 1
dtm_a[1, 213] = 1
dtm_a[1, 214] = 1
dtm_a[1, 215] = 1
dtm_a[1, 216] = 1
dtm_a[1, 217] = 1
dtm_a[1, 218] = 1
dtm_a[1, 219] = 1
dtm_a[1, 220] = 1
dtm_a[1, 221] = 1
dtm_a[1, 222] = 1
dtm_a[1, 223] = 1
dtm_a[1, 224] = 1
dtm_a[1, 225] = 1
dtm_a[1, 226] = 1
dtm_a[1, 227] = 1
dtm_a[1, 228] = 1
dtm_a[1, 229] = 1
dtm_a[1, 230] = 1
dtm_a[1, 231] = 1
dtm_a[1, 232] = 1
dtm_a[1, 233] = 1
dtm_a[1, 234] = 1
dtm_a[1, 235] = 1
dtm_a[1, 236] = 1
dtm_a[1, 237] = 1
dtm_a[1, 238] = 1
dtm_a[1, 239] = 1
dtm_a[1, 240] = 1
dtm_a[1, 241] = 1
dtm_a[1, 242] = 1
dtm_a[1, 243] = 1
dtm_a[1, 244] = 1
dtm_a[1, 245] = 1
dtm_a[1, 246] = 1
dtm_a[1, 247] = 1
dtm_a[1, 248] = 1
dtm_a[1, 249] = 1
dtm_a[1, 250] = 1
dtm_a[1, 251] = 1
dtm_a[1, 252] = 1
dtm_a[1, 253] = 1
dtm_a[1, 254] = 1
dtm_a[1, 255] = 1
dtm_a[1, 256] = 1
dtm_a[1, 257] = 1
dtm_a[1, 258] = 1
dtm_a[1, 259] = 1
dtm_a[1, 260] = 1
dtm_a[1, 261] = 1
dtm_a[1, 262] = 1
dtm_a[1, 263] = 1
dtm_a[1, 264] = 1
dtm_a[1, 265] = 1
dtm_a[1, 266] = 1
dtm_a[1, 267] = 1
dtm_a[1, 268] = 1
dtm_a[1, 269] = 1
dtm_a[1, 270] = 1
dtm_a[1, 271] = 1
dtm_a[1, 272] = 1
dtm_a[1, 273] = 1
dtm_a[1, 274] = 1
dtm_a[1, 275] = 1
dtm_a[1, 276] = 1
dtm_a[1, 277] = 1
dtm_a[1, 278] = 1
dtm_a[1, 279] = 1
dtm_a[1, 280] = 1
dtm_a[1, 281] = 1
dtm_a[1, 282] = 1
dtm_a[1, 283] = 1
dtm_a[1, 284] = 1
dtm_a[1, 285] = 1
dtm_a[1, 286] = 1
dtm_a[1, 287] = 1
dtm_a[1, 288] = 1
dtm_a[1, 289] = 1
dtm_a[1, 290] = 1
dtm_a[1, 291] = 1
dtm_a[1, 292] = 1
dtm_a[1, 293] = 1
dtm_a[1, 294] = 1
dtm_a[1, 295] = 1
dtm_a[1, 296] = 1
dtm_a[1, 297] = 1
dtm_a[1, 298] = 1
dtm_a[1, 299] = 1
dtm_a[1, 300] = 1
dtm_a[1, 301] = 1
dtm_a[1, 302] = 1
dtm_a[1, 303] = 1
dtm_a[1, 304] = 1
dtm_a[1, 305] = 1
dtm_a[1, 306] = 1
dtm_a[1, 307] = 1
dtm_a[1, 308] = 1
dtm_a[1, 309] = 1
dtm_a[1, 310] = 1
dtm_a[1, 311] = 1
dtm_a[1, 312] = 1
dtm_a[1, 313] = 1
dtm_a[1, 314] = 1
dtm_a[1, 315] = 1
dtm_a[1, 316] = 1
dtm_a[1, 317] = 1
dtm_a[1, 318] = 1
dtm_a[1, 319] = 1
dtm_a[1, 320] = 1
dtm_a[1, 321] = 1
dtm_a[1, 322] = 1
dtm_a[1, 323] = 1
dtm_a[1, 324] = 1
dtm_a[1, 325] = 1
dtm_a[1, 326] = 1
dtm_a[1, 327] = 1
dtm_a[1, 328] = 1
dtm_a[1, 329] = 1
dtm_a[1, 330] = 1
dtm_a[1, 331] = 1
dtm_a[1, 332] = 1
dtm_a[1, 333] = 1
dtm_a[1, 334] = 1
dtm_a[1, 335] = 1
dtm_a[1, 336] = 1
dtm_a[1, 337] = 1
dtm_a[1, 338] = 1
dtm_a[1, 339] = 1
dtm_a[1, 340] = 1
dtm_a[1, 341] = 1
dtm_a[1, 342] = 1
dtm_a[1, 343] = 1
dtm_a[1, 344] = 1
dtm_a[1, 345] = 1
dtm_a[1, 346] = 1
dtm_a[1, 347] = 1
dtm_a[1, 348] = 1
dtm_a[1, 349] = 1
dtm_a[1, 350] = 1
dtm_a[1, 351] = 1
dtm_a[1, 352] = 1
dtm_a[1, 353] = 1
dtm_a[1, 354] = 1
dtm_a[1, 355] = 1
dtm_a[1, 356] = 1
dtm_a[1, 357] = 1
dtm_a[1, 358] = 1
dtm_a[1, 359] = 1
dtm_a[1, 360] = 1
dtm_a[1, 361] = 1
dtm_a[1, 362] = 1
dtm_a[1, 363] = 1
dtm_a[1, 364] = 1
dtm_a[1, 365] = 1
dtm_a[1, 366] = 1
dtm_a[1, 367] = 1
dtm_a[1, 368] = 1
dtm_a[1, 369] = 1
dtm_a[1, 370] = 1
dtm_a[1, 371] = 1
dtm_a[1, 372] = 1
dtm_a[1, 373] = 1
dtm_a[1, 374] = 1
dtm_a[1, 375] = 1
dtm_a[1, 376] = 1
dtm_a[1, 377] = 1
dtm_a[1, 378] = 1
dtm_a[1, 379] = 1
dtm_a[1, 380] = 1
dtm_a[1, 381] = 1
dtm_a[1, 382] = 1
dtm_a[1, 383] = 1
dtm_a[1, 384] = 1
dtm_a[1, 385] = 1
dtm_a[1, 386] = 1
dtm_a[1, 387] = 1
dtm_a[1, 388] = 1
dtm_a[1, 389] = 1
dtm_a[1, 390] = 1
dtm_a[1, 391] = 1
dtm_a[1, 392] = 1
dtm_a[1, 393] = 1
dtm_a[1, 394] = 1
dtm_a[1, 395] = 1
dtm_a[1, 396] = 1
dtm_a[1, 397] = 1
dtm_a[1, 398] = 1
dtm_a[1, 399] = 1
dtm_a[1, 400] = 1
dtm_a[1, 401] = 1
dtm_a[1, 402] = 1
dtm_a[1, 403] = 1
dtm_a[1, 404] = 1
dtm_a[1, 405] = 1
dtm_a[1, 406] = 1
dtm_a[1, 407] = 1
dtm_a[1, 408] = 1
dtm_a[1, 409] = 1
dtm_a[1, 410] = 1
dtm_a[1, 411] = 1
dtm_a[1, 412] = 1
dtm_a[1, 413] = 1
dtm_a[1, 414] = 1
dtm_a[1, 415] = 1
dtm_a[1, 416] = 1
dtm_a[1, 417] = 1
dtm_a[1, 418] = 1
dtm_a[1, 419] = 1
dtm_a[1, 420] = 1
dtm_a[1, 421] = 1
dtm_a[1, 422] = 1
dtm_a[1, 423] = 1
dtm_a[1, 424] = 1
dtm_a[1, 425] = 1
dtm_a[1, 426] = 1
dtm_a[1, 427] = 1
dtm_a[1, 428] = 1
dtm_a[1, 429] = 1
dtm_a[1, 430] = 1
dtm_a[1, 431] = 1
dtm_a[1, 432] = 1
dtm_a[1, 433] = 1
dtm_a[1, 434] = 1
dtm_a[1, 435] = 1
dtm_a[1, 436] = 1
dtm_a[1, 437] = 1
dtm_a[1, 438] = 1
dtm_a[1, 439] = 1
dtm_a[1, 440] = 1
dtm_a[1, 441] = 1
dtm_a[1, 442] = 1
dtm_a[1, 443] = 1
dtm_a[1, 444] = 1
dtm_a[1, 445] = 1
dtm_a[1, 446] = 1
dtm_a[1, 447] = 1
dtm_a[1, 448] = 1
dtm_a[1, 449] = 1
dtm_a[1, 450] = 1
dtm_a[1, 451] = 1
dtm_a[1, 452] = 1
dtm_a[1, 453] = 1
dtm_a[1, 454] = 1
dtm_a[1, 455] = 1
dtm_a[1, 456] = 1
dtm_a[1, 457] = 1
dtm_a[1, 458] = 1
dtm_a[1, 459] = 1
dtm_a[1, 460] = 1
dtm_a[1, 461] = 1
dtm_a[1, 462] = 1
dtm_a[1, 463] = 1
dtm_a[1, 464] = 1
dtm_a[1, 465] = 1
dtm_a[1, 466] = 1
dtm_a[1, 467] = 1
dtm_a[1, 468] = 1
dtm_a[1, 469] = 1
dtm_a[1, 470] = 1
dtm_a[1, 471] = 1
dtm_a[1, 472] = 1
dtm_a[1, 473] = 1
dtm_a[1, 474] = 1
dtm_a[1, 475] = 1
dtm_a[1, 476] = 1
dtm_a[1, 477] = 1
dtm_a[1, 478] = 1
dtm_a[1, 479] = 1
dtm_a[1, 480] = 1
dtm_a[1, 481] = 1
dtm_a[1, 482] = 1
dtm_a[1, 483] = 1
dtm_a[1, 484] = 1
dtm_a[1, 485] = 1
dtm_a[1, 486] = 1
dtm_a[1, 487] = 1
dtm_a[1, 488] = 1
dtm_a[1, 489] = 1
dtm_a[1, 490] = 1
dtm_a[1, 491] = 1
dtm_a[1, 492] = 1
dtm_a[1, 493] = 1
dtm_a[1, 494] = 1
dtm_a[1, 495] = 1
dtm_a[1, 496] = 1
dtm_a[1, 497] = 1
dtm_a[1, 498] = 1
dtm_a[1, 499] = 1
dtm_a[1, 500] = 1
dtm_a[1, 501] = 1
dtm_a[1, 502] = 1
dtm_a[1, 503] = 1
dtm_a[1, 504] = 1
dtm_a[1, 505] = 1
dtm_a[1, 506] = 1
dtm_a[1, 507] = 1
dtm_a[1, 508] = 1
dtm_a[1, 509] = 1
dtm_a[1, 510] = 1
dtm_a[1, 511] = 1
dtm_a[1, 512] = 1
dtm_a[1, 513] = 1
dtm_a[1, 514] = 1
dtm_a[1, 515] = 1
dtm_a[1, 516] = 1
dtm_a[1, 517] = 1
dtm_a[1, 518] = 1
dtm_a[1, 519] = 1
dtm_a[1, 520] = 1
dtm_a[1, 521] = 1
dtm_a[1, 522] = 1
dtm_a[1, 523] = 1
dtm_a[1, 524] = 1
dtm_a[1, 525] = 1
dtm_a[1, 526] = 1
dtm_a[1, 527] = 1
dtm_a[1, 528] = 1
dtm_a[1, 529] = 1
dtm_a[1, 530] = 1
dtm_a[1, 531] = 1
dtm_a[1, 532] = 1
dtm_a[1, 533] = 1
dtm_a[1, 534] = 1
dtm_a[1, 535] = 1
dtm_a[1, 536] = 1
dtm_a[1, 537] = 1
dtm_a[1, 538] = 1
dtm_a[1, 539] = 1
dtm_a[1, 540] = 1
dtm_a[1, 541] = 1
dtm_a[1, 542] = 1
dtm_a[1, 543] = 1
dtm_a[1, 544] = 1
dtm_a[1, 545] = 1
dtm_a[1, 546] = 1
dtm_a[1, 547] = 1
dtm_a[1, 548] = 1
dtm_a[1, 549] = 1
dtm_a[1, 550] = 1
dtm_a[1, 551] = 1
dtm_a[1, 552] = 1
dtm_a[1, 553] = 1
dtm_a[1, 554] = 1
dtm_a[1, 555] = 1
dtm_a[1, 556] = 1
dtm_a[1, 557] = 1
dtm_a[1, 558] = 1
dtm_a[1, 559] = 1
dtm_a[1, 560] = 1
dtm_a[1, 561] = 1
dtm_a[1, 562] = 1
dtm_a[1, 563] = 1
dtm_a[1, 564] = 1
dtm_a[1, 565] = 1
dtm_a[1, 566] = 1
dtm_a[1, 567] = 1
dtm_a[1, 568] = 1
dtm_a[1, 569] = 1
dtm_a[1, 570] = 1
dtm_a[1, 571] = 1
dtm_a[1, 572] = 1
dtm_a[1, 573] = 1
dtm_a[1, 574] = 1
dtm_a[1, 575] = 1
dtm_a[1, 576] = 1
dtm_a[1, 577] = 1
dtm_a[1, 578] = 1
dtm_a[1, 579] = 1
dtm_a[1, 580] = 1
dtm_a[1, 581] = 1
dtm_a[1, 582] = 1
dtm_a[1, 583] = 1
dtm_a[1, 584] = 1
dtm_a[1, 585] = 1
dtm_a[1, 586] = 1
dtm_a[1, 587] = 1
dtm_a[1, 588] = 1
dtm_a[1, 589] = 1
dtm_a[1, 590] = 1
dtm_a[1, 591] = 1
dtm_a[1, 592] = 1
dtm_a[1, 593] = 1
dtm_a[1, 594] = 1
dtm_a[1, 595] = 1
dtm_a[1, 596] = 1
dtm_a[1, 597] = 1
dtm_a[1, 598] = 1
dtm_a[1, 599] = 1
dtm_a[1, 600] = 1
dtm_a[1, 601] = 1
dtm_a[1, 602] = 1
dtm_a[1, 603] = 1
dtm_a[1, 604] = 1
dtm_a[1, 605] = 1
dtm_a[1, 606] = 1
dtm_a[1, 607] = 1
dtm_a[1, 608] = 1
dtm_a[1, 609] = 1
dtm_a[1, 610] = 1
dtm_a[1, 611] = 1
dtm_a[1, 612] = 1
dtm_a[1, 613] = 1
dtm_a[1, 614] = 1
dtm_a[1, 615] = 1
dtm_a[1, 616] = 1
dtm_a[1, 617] = 1
dtm_a[1, 618] = 1
dtm_a[1, 619] = 1
dtm_a[1, 620] = 1
dtm_a[1, 621] = 1
dtm_a[1, 622] = 1
dtm_a[1, 623] = 1
dtm_a[1, 624] = 1
dtm_a[1, 625] = 1
dtm_a[1, 626] = 1
dtm_a[1, 627] = 1
dtm_a[1, 628] = 1
dtm_a[1, 629] = 1
dtm_a[1, 630] = 1
dtm_a[1, 631] = 1
dtm_a[1, 632] = 1
dtm_a[1, 633] = 1
dtm_a[1, 634] = 1
dtm_a[1, 635] = 1
dtm_a[1, 636] = 1
dtm_a[1, 637] = 1
dtm_a[1, 638] = 1
dtm_a[1, 639] = 1
dtm_a[1, 640] = 1
dtm_a[1, 641] = 1
dtm_a[1, 642] = 1
dtm_a[1, 643] = 1
dtm_a[1, 644] = 1
dtm_a[1, 645] = 1
dtm_a[1, 646] = 1
dtm_a[1, 647] = 1
dtm_a[1, 648] = 1
dtm_a[1, 649] = 1
dtm_a[1, 650] = 1
dtm_a[1, 651] = 1
dtm_a[1, 652] = 1
dtm_a[1, 653] = 1
dtm_a[1, 654] = 1
dtm_a[1, 655] = 1
dtm_a[1, 656] = 1
dtm_a[1, 657] = 1
dtm_a[1, 658] = 1
dtm_a[1, 659] = 1
dtm_a[1, 660] = 1
dtm_a[1, 661] = 1
dtm_a[1, 662] = 1
dtm_a[1, 663] = 1
dtm_a[1, 664] = 1
dtm_a[1, 665] = 1
dtm_a[1, 666] = 1
dtm_a[1, 667] = 1
dtm_a[1, 668] = 1
dtm_a[1, 669] = 1
dtm_a[1, 670] = 1
dtm_a[1, 671] = 1
dtm_a[1, 672] = 1
dtm_a[1, 673] = 1
dtm_a[1, 674] = 1
dtm_a[1, 675] = 1
dtm_a[1, 676] = 1
dtm_a[1, 677] = 1
dtm_a[1, 678] = 1
dtm_a[1, 679] = 1
dtm_a[1, 680] = 1
dtm_a[1, 681] = 1
dtm_a[1, 682] = 1
dtm_a[1, 683] = 1
dtm_a[1, 684] = 1
dtm_a[1, 685] = 1
dtm_a[1, 686] = 1
dtm_a[1, 687] = 1
dtm_a[1, 688] = 1
dtm_a[1, 689] = 1
dtm_a[1, 690] = 1
dtm_a[1, 691] = 1
dtm_a[1, 692] = 1
dtm_a[1, 693] = 1
dtm_a[1, 694] = 1
dtm_a[1, 695] = 1
dtm_a[1, 696] = 1
dtm_a[1, 
```

AmCAT

File Edit Code View Plots Session Build Debug Profile Tools Help

optics.R x data x ondw x Untitled2* x Untitled3* x jiebo.R x Untitled4* x Run Source Addins

```
1 library(amcatr)
2 conn = amcat.connect("https://amcat.nl")
3 load("~/Downloads/dataR.RData")
4 bigor2 = bigor
5 bigor = head(bigor)
6 bigor = subset(bigor2, date != "NA")
7 bigor$date2 = gsub(" d. ", "T", bigor$date)
8 bigor$date2 = gsub(" m. ", "_", bigor$date2)
9 months = list("spalio"="10", "lapkričio"="11", "gruodžio"="12",
10           "sausio"="01", "vasario"="02", "kovo"="03", "balandžio"="04")
11 for (m in names(months)) bigor$date2 = gsub(paste0(m, " "), paste0(months[[m]], "-"), bigor$date2)
12
13 bigor$date2 = as.Date(bigor$date2)
14
15 bigor = head(bigor, 100)
16 amcat.upload.articles(conn, 1410, "LIFE2", headline = "-", text = bigor$text, date = bigor$date2,
17   provenance = NULL)
18
19
20
21 d = data.frame(domein=c(rep("x", 5), "y", "y"), headline=c("a", "a", "b", "c", "b", "a", "a"))
22 d
23 d$domhead = paste(d$domein, d$headline, sep = "_")
24 d$keep = 0
25 d[!d$keep %in% unique(d$domhead), ] = 1
23:52 (Top Level) R Script
```

Console ~/

```
Attaching package: 'scales'

The following object is masked from 'package:plotrix':

  rescale

> dtm.wordcloud(dtm_a)
Error in dtm.wordcloud(dtm_a) : object 'dtm_a' not found
> dtm_a = create_matrix(text_a)
> dtm_b = create_matrix(text_b)
> # dan kan je natuurlijk direct wordcloudjes maken (maar die zijn vast niet heel interessant)
> dtm.wordcloud(dtm_a)
> cmp = corpora.compare(dtm_a, dtm_b)
> cmp = plyr::arrange(cmp, over)
> cmp = arrange(cmp, -termfreq)
> with(head(cmp, 100), plotWords(x=log(over), words=term, wordfreq=termfreq, random.y = T, scale=5))
>
```

Environment History Presentation

Global Environment

- cmp 20157 obs. of 8 variables
- posts 10 obs. of 11 variables
- r 706381 obs. of 4 variables
- r_a 191770 obs. of 4 variables
- r_b 137079 obs. of 4 variables
- set 100 obs. of 2 variables

Values

- app_id "898983470244398"
- app_secret "aaaf59490d630a4e8a73550a4451a3775"
- conn List of 4
- dtm_a Large DocumentTermMatrix (6 elements, 1.4 Mb)

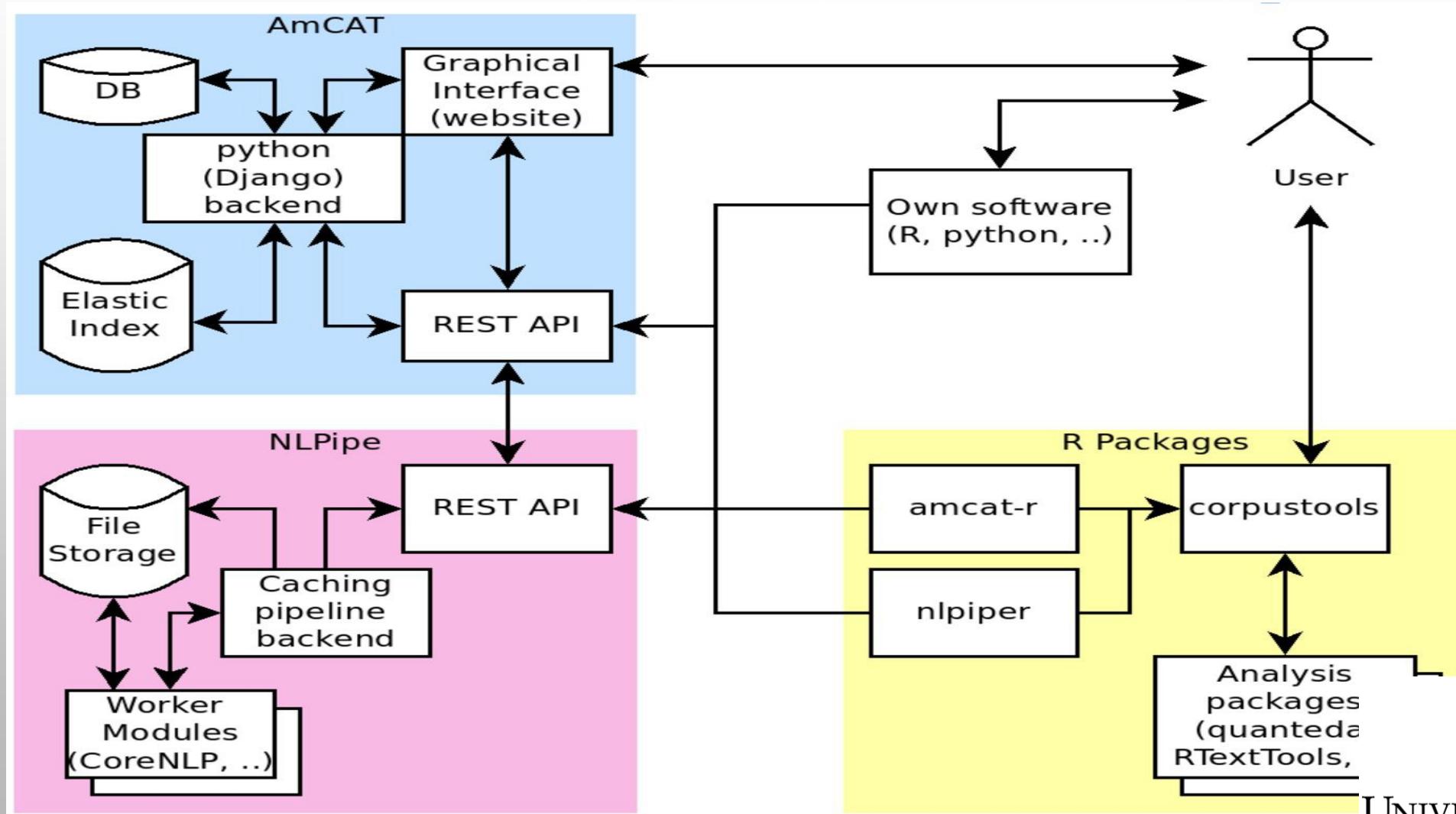
Files Plots Packages Help Viewer

Zoom Export Publish

Universiteit van Amsterdam



NLPipe



JamCode

Content Analysis Tool

LETTING THE CAT OUT OF THE BAG

Co-presence is related to relatedness, but the relationship is complicated.

- Macrosyntax: Headlines, Turns, Lists, etc.
- Syntax: Clauses, Parentheses, Sentences, etc.
- Conjunctions, Grammatical Roles & POS
- Anaphora, Coreference & Exophora
- Sequence & Proximity
- Stylistic devices: Rhymes, Puns, Alliterations
- Register & Contextual Knowledge

Intertextual Inference Context Model

JAMCODE

- ▶ Shallow parsing of Syntax and Macrosyntax
- ▶ Scoring based on Proximity & Sequence
- ▶ Imputation based on Contextual Discourse
- ▶ Multi-feature Probabilistic Relatedness

Micro & Macro Order clause, sentence, paragraph, title, ...

Trump's Valueless Foreign Policy
www.nytimes.com | Roger Cohen | 2 May 2017

So the threats were no more than bluster, and all is well. That is one view of President Trump's foreign policy at the 100-day-or-so mark.

Wrong.

Yes, there's no sign of the Wall, and NATO is no longer "obsolete," and the Iran nuclear deal is still in place, [...]

Defense Secretary Jim Mattis' national security adviser, recklessness and bellicosity. They have neutralized his ignorance even if nobody can help the president grasp its extent. Some of the loonier members of the

president's entourage have been fired or marginalized. Adults have taken charge. There is still a lot of noise, but "America First" has not upended the world.

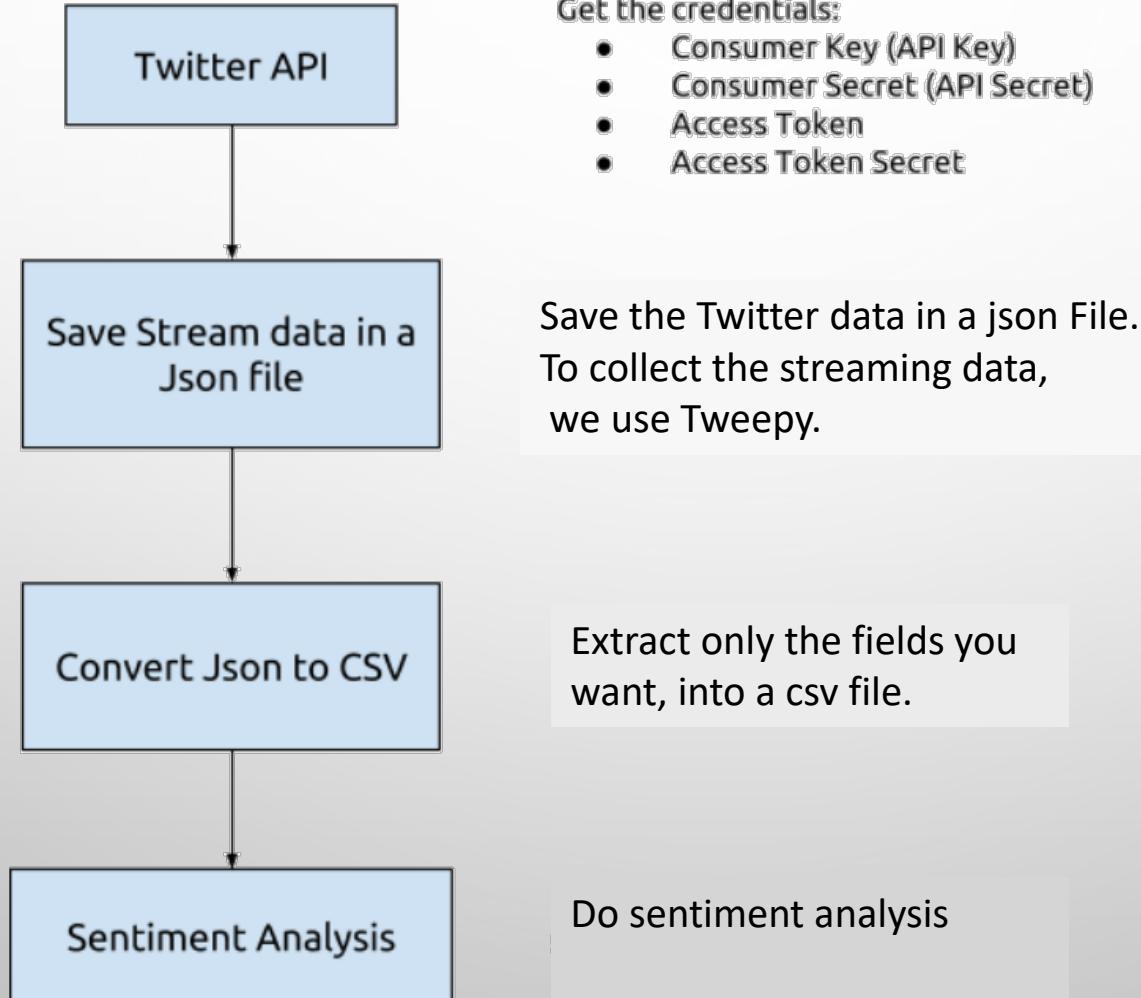
Local Coherence proximity ▶ probability



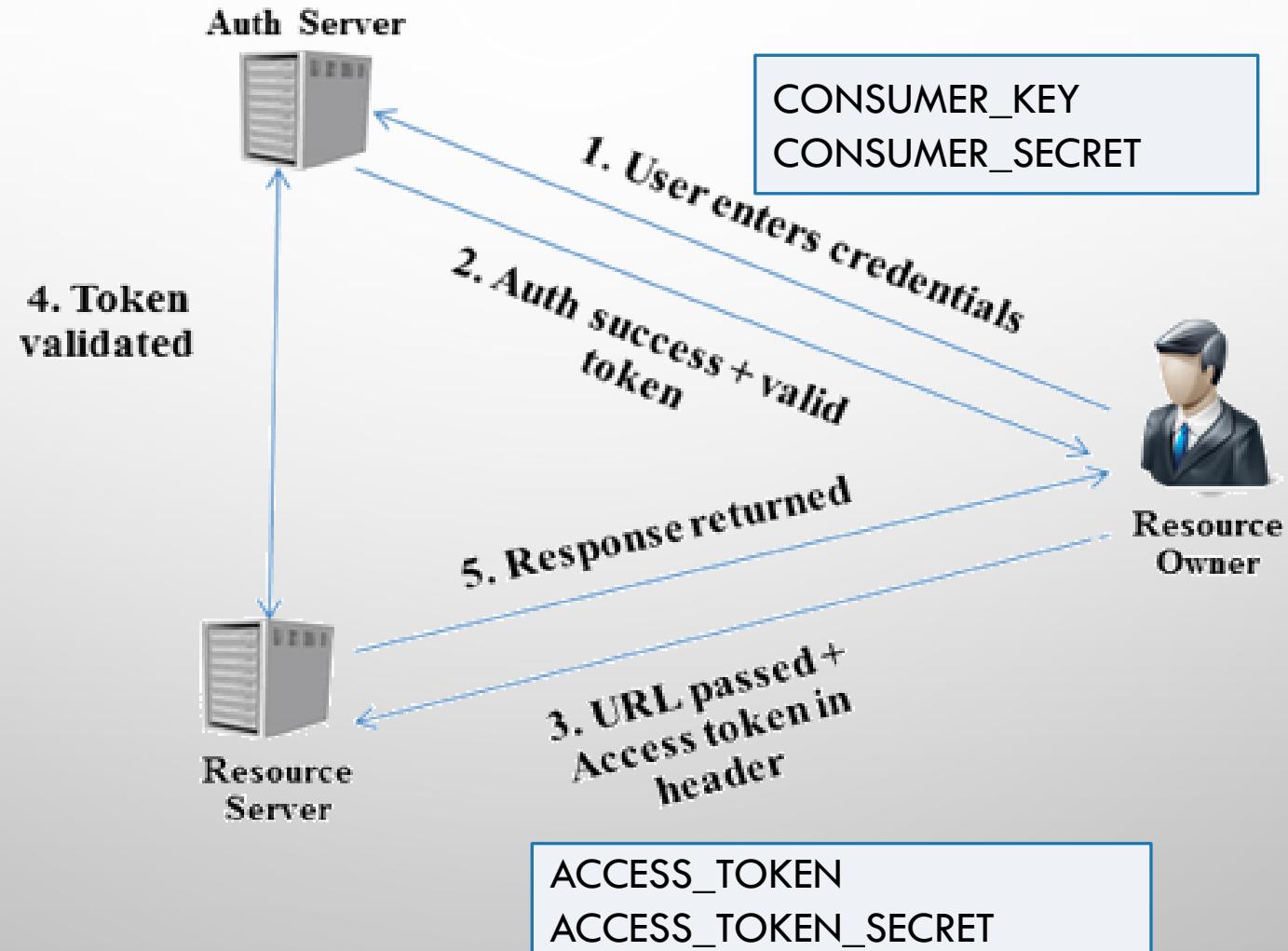
Thank you!

Slides will be up at
[github/kj2013](https://github.com/kj2013)

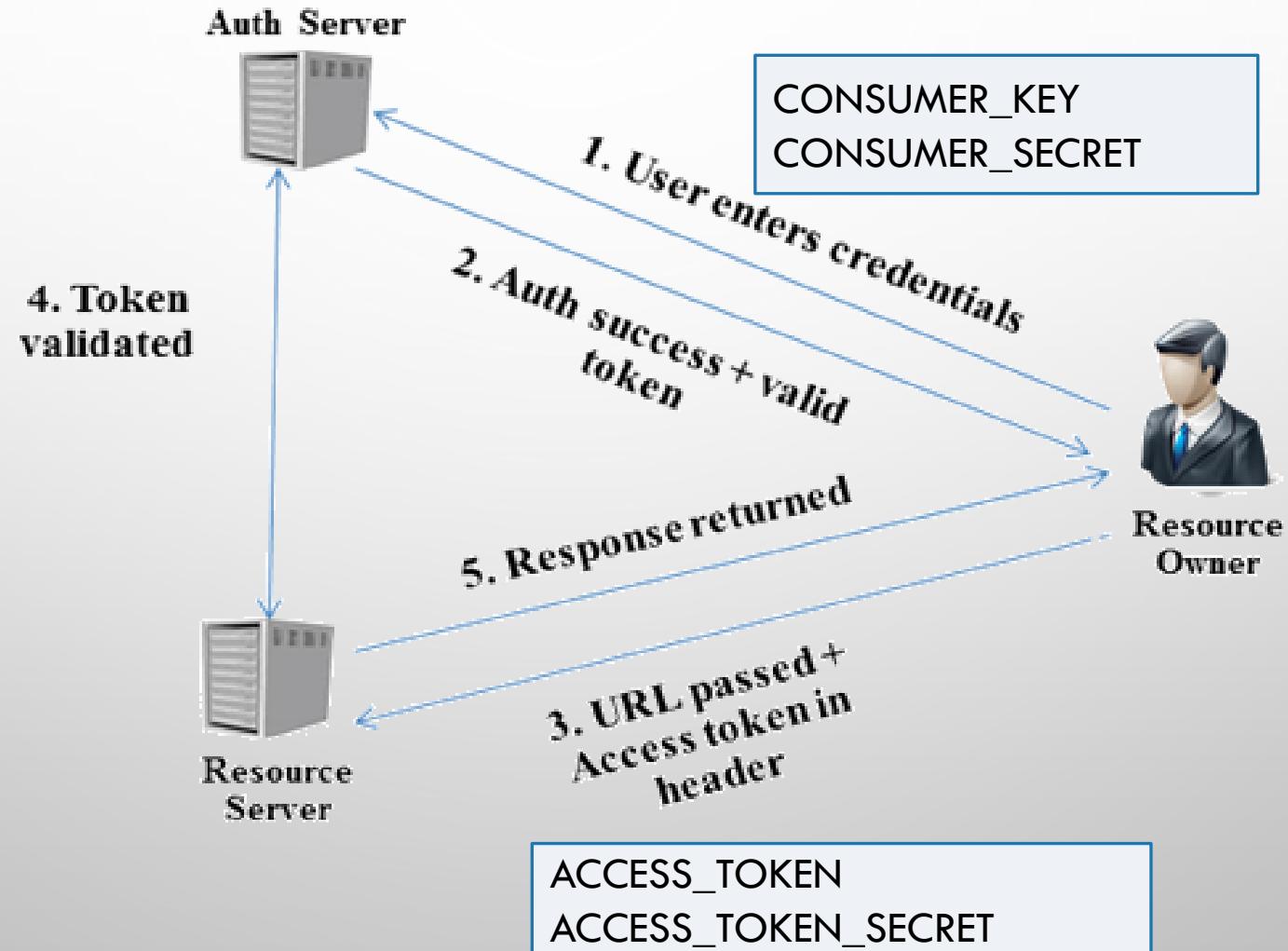
Recap: Data Mining from Twitter



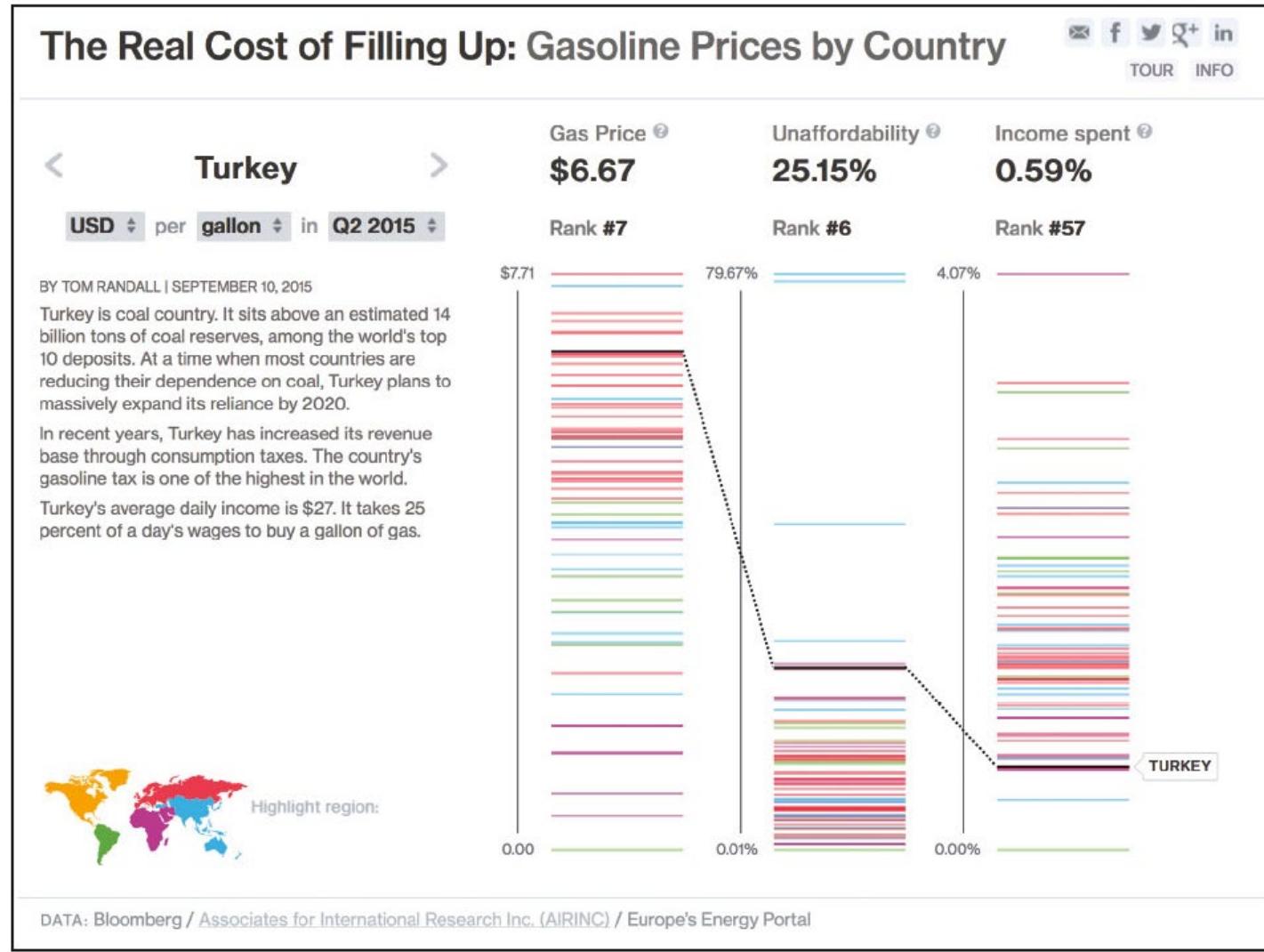
OAUTH dance



OAUTH dance



EXAMPLES



Berinato, S. (2016). *Good charts: The HBR guide to making smarter, more persuasive data visualizations.* Harvard Business Review Press.