

# *Telling a Story with Data:*

# Using R for data journalism

Dr. Kokil Jaidka    [jaidka@ntu.edu.sg](mailto:jaidka@ntu.edu.sg)  
[jaidka@sas.upenn.edu](mailto:jaidka@sas.upenn.edu)

# What is Data Journalism?

## Traditional Journalism

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

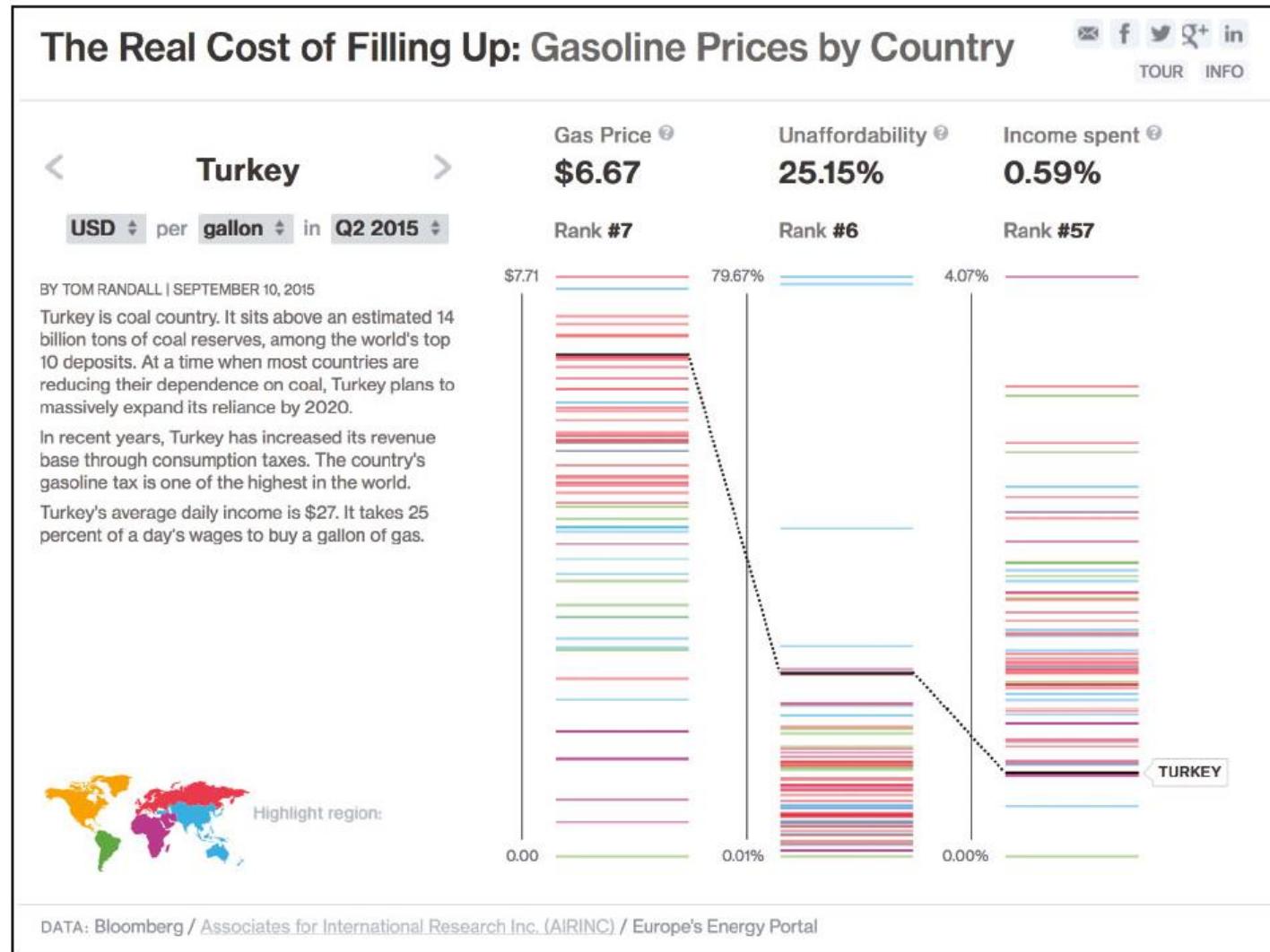
## Data Journalism

- 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>

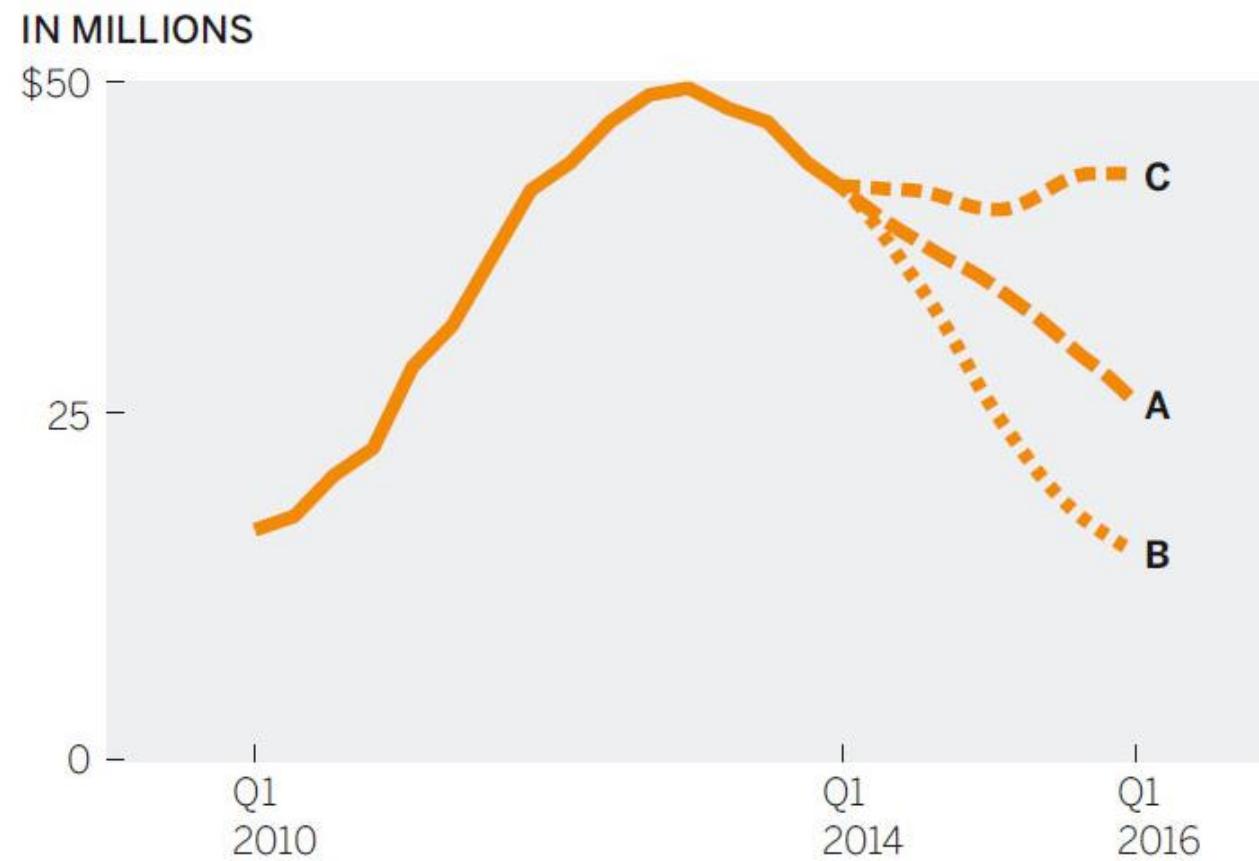
# EXAMPLES



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

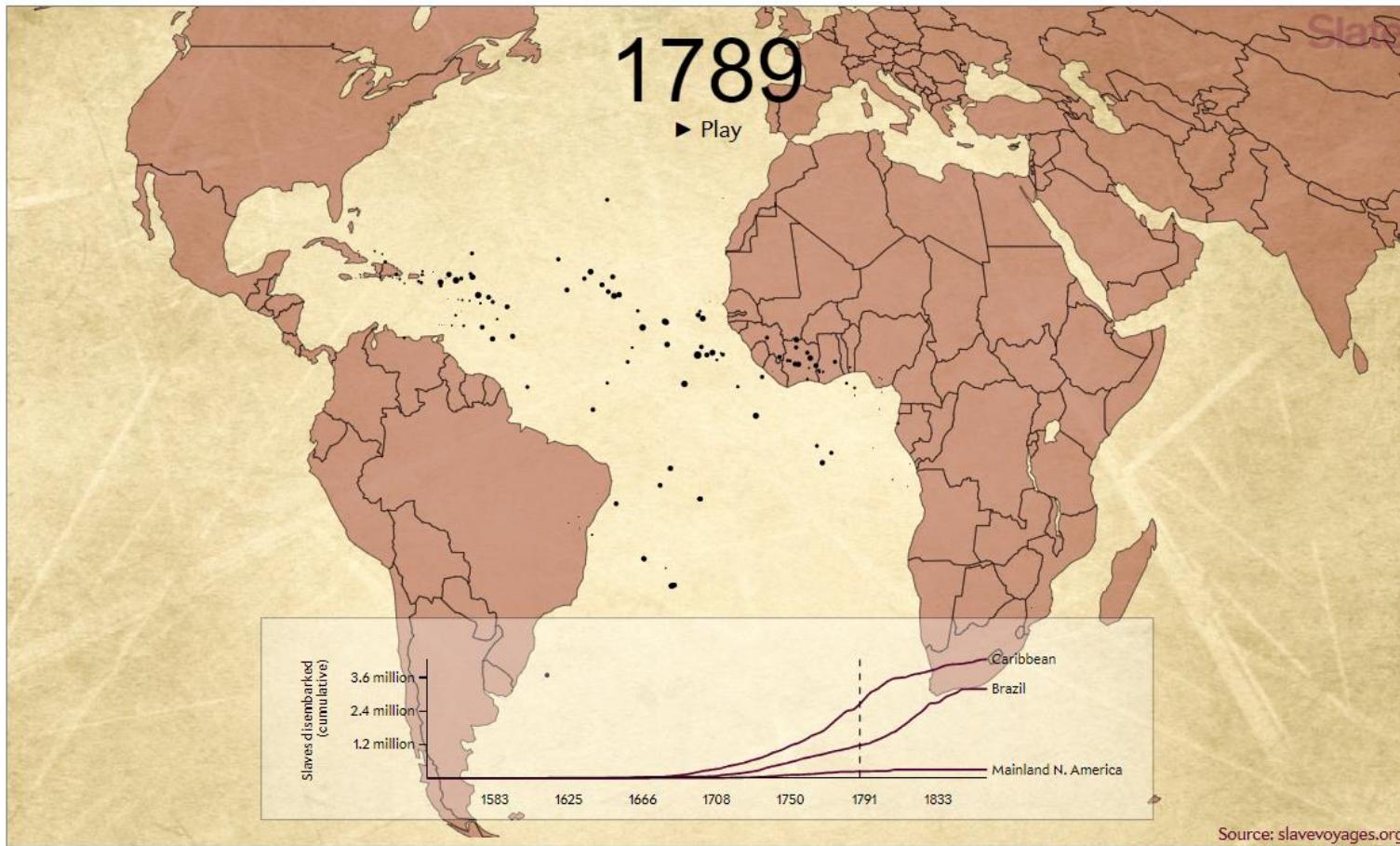
# EXAMPLES

## PROJECTIONS—THREE SCENARIOS



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

# EXAMPLES

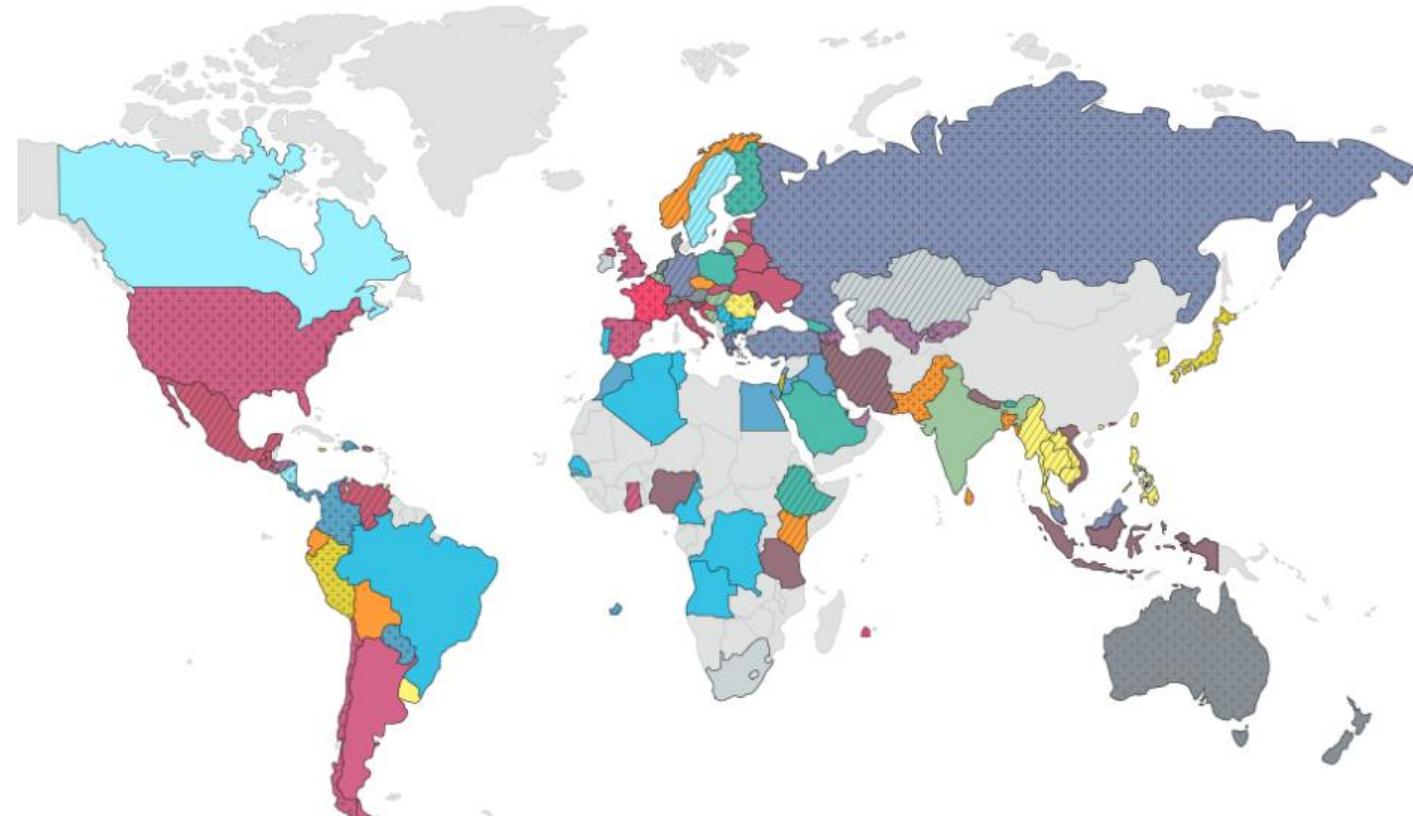


The Atlantic Slave Trade in 2 Minutes (2015). *The Slate*.

[http://www.slate.com/articles/life/the\\_history\\_of\\_american\\_slavery/2015/06/animated\\_interactive\\_of\\_the\\_history\\_of\\_the\\_atlantic\\_slave\\_trade.html](http://www.slate.com/articles/life/the_history_of_american_slavery/2015/06/animated_interactive_of_the_history_of_the_atlantic_slave_trade.html)

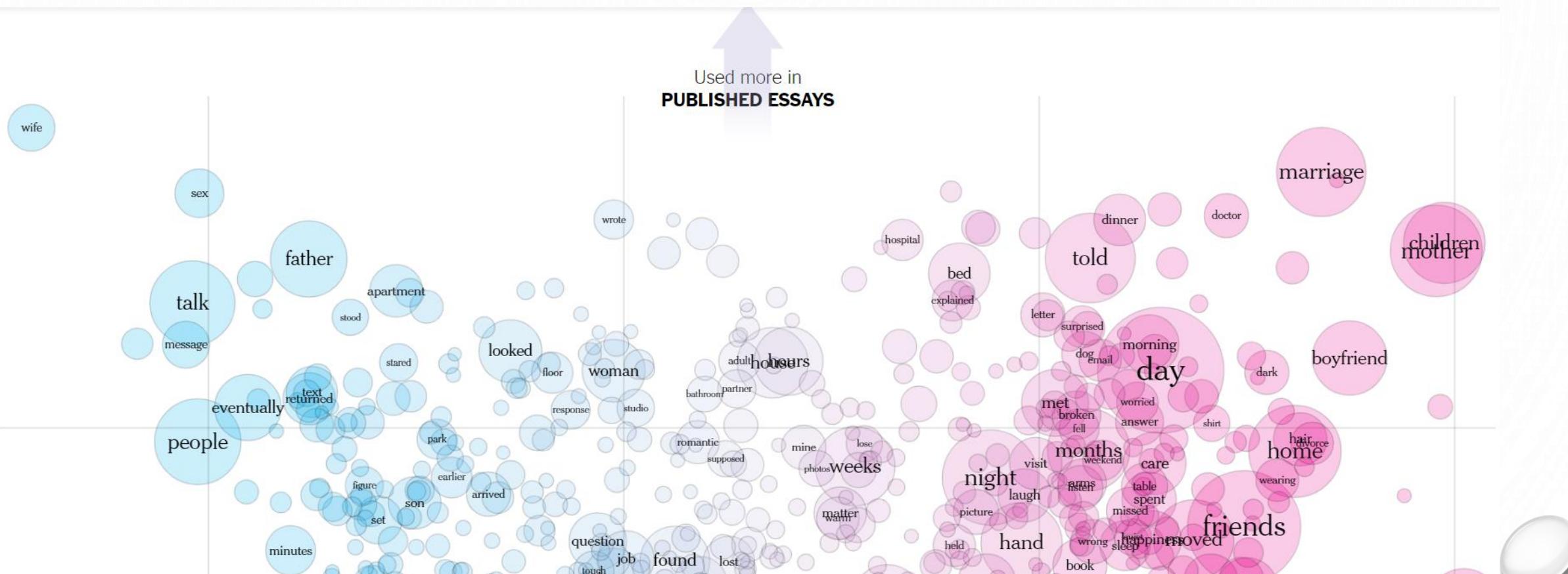
# EXAMPLES

## 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>

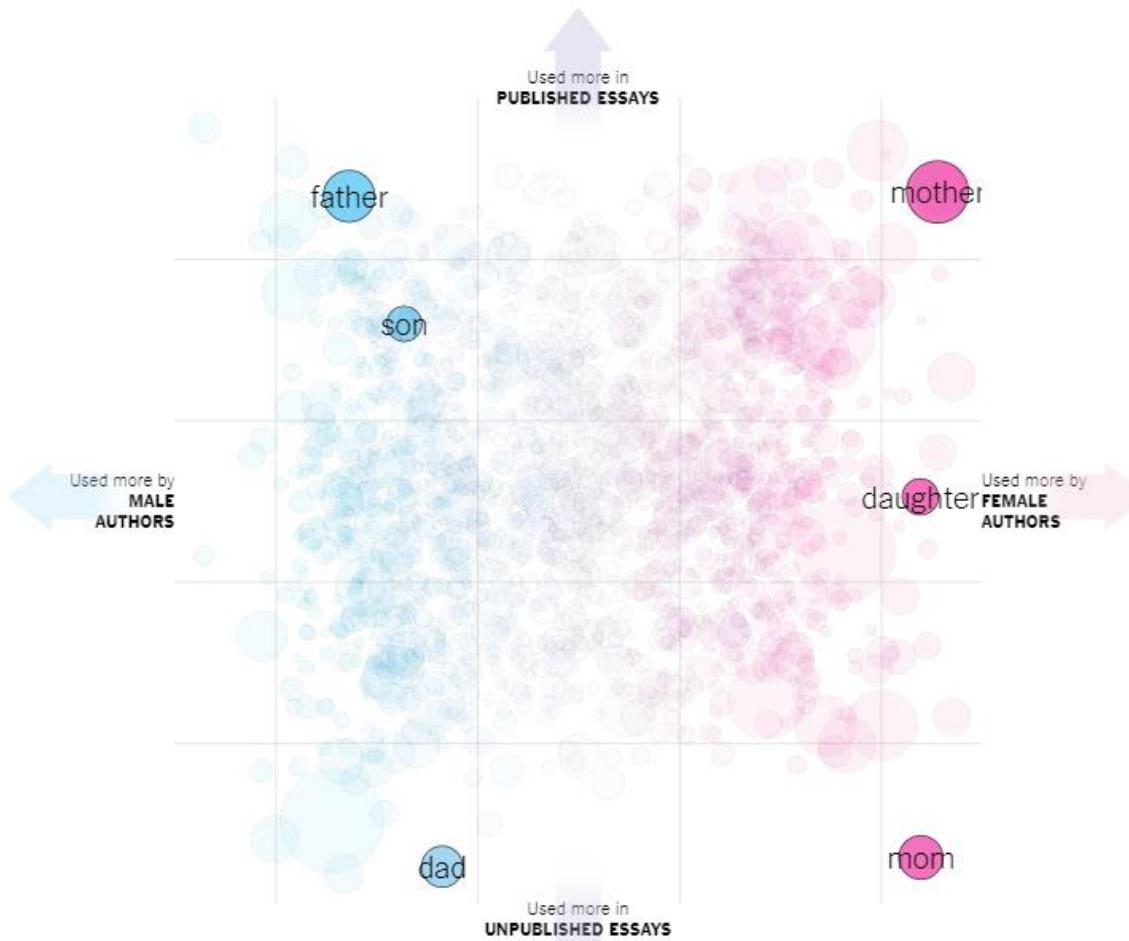
# EXAMPLES



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

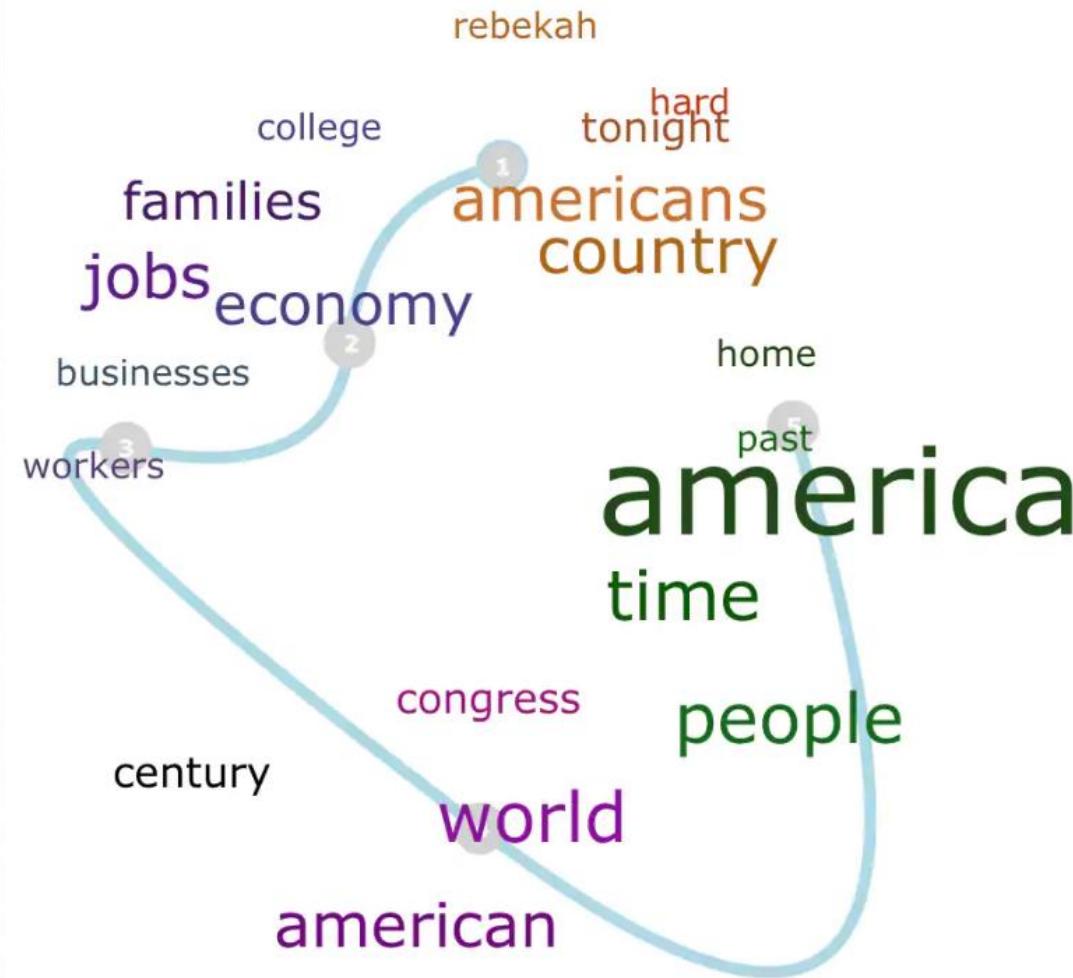
<https://www.nytimes.com/interactive/2017/11/07/upshot/modern-love-what-we-write-when-we-write-about-love.html>

# 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



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

# Takeaways



HOW TO PLAN YOUR DATA  
JOURNALISM PROJECT



HOW TO GET DATA FROM  
THE NYTIMES API USING R

# 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)**

# 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?*

<b>Question</b>	<b>Ideal data/pre-requisites</b>
<ul style="list-style-type: none"><li>• <i>What was the public reaction to the Trump – Kim Summit in Singapore?</i></li></ul>	<ul style="list-style-type: none"><li>• Collection of public commentary (when/where/how)</li></ul>
<ul style="list-style-type: none"><li>• <i>Did PAP's position on foreign talent change between 2008 and 2016?</i></li></ul>	<ul style="list-style-type: none"><li>• Definition of “position”</li><li>• Laws on foreign talent from 2008-2016</li><li>• Statistics on foreign talent influx</li><li>• Statistics on jobs</li></ul>
<ul style="list-style-type: none"><li>• <i>How is the public opinion in SG vs MY different re: the maritime dispute?</i></li></ul>	<ul style="list-style-type: none"><li>• Collection of public commentary (when/where/how/what)</li></ul>
<ul style="list-style-type: none"><li>• <i>What makes a K-pop song go viral?</i></li></ul>	<ul style="list-style-type: none"><li>• Statistics on viral and non-viral K-pop songs</li><li>• Definition of “virality”</li><li>• Statistics on campaign speeches</li></ul>
<ul style="list-style-type: none"><li>• <i>What makes a winning campaign speech?</i></li></ul>	<ul style="list-style-type: none"><li>• Definition of “winning”</li></ul>

# 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

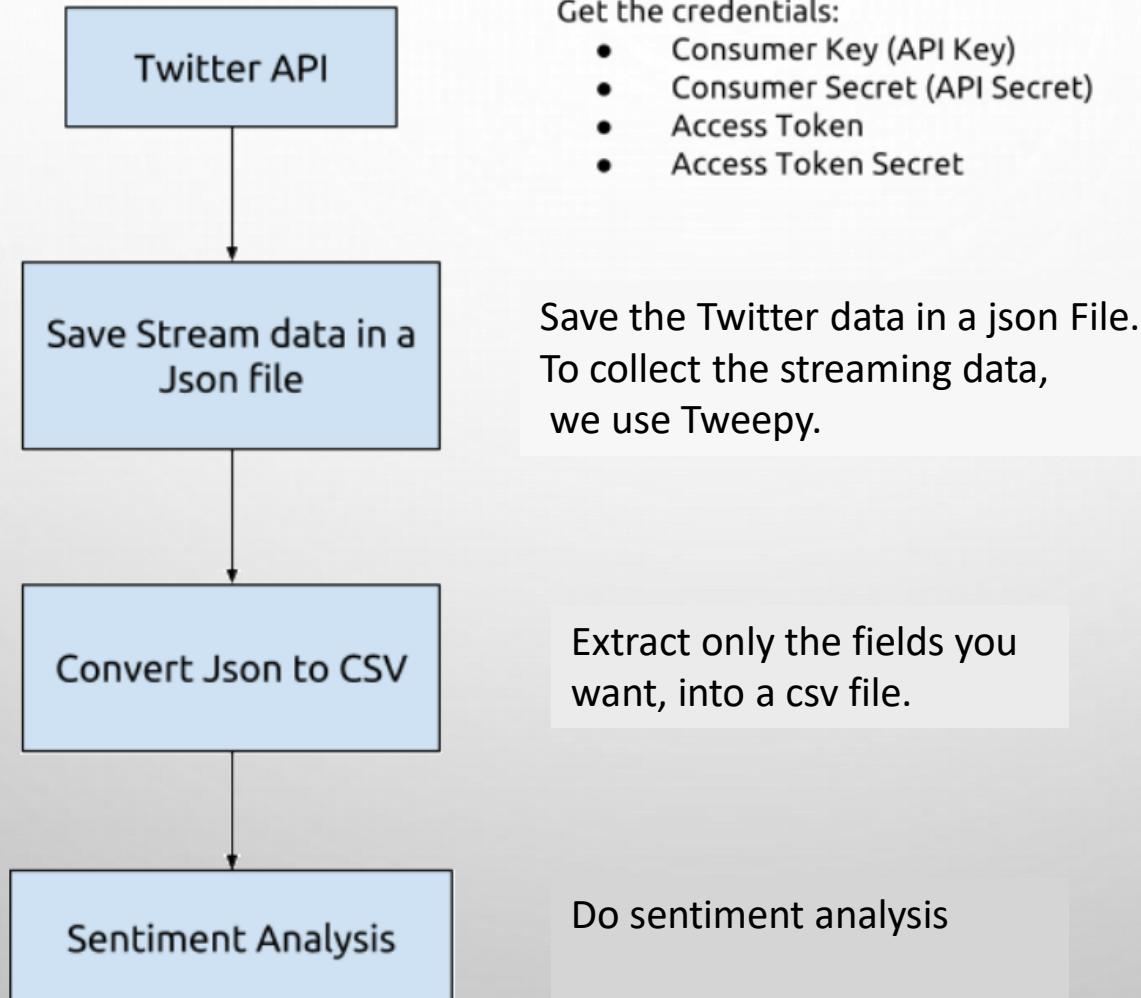


HOW TO VISUALIZE DATA  
USING R

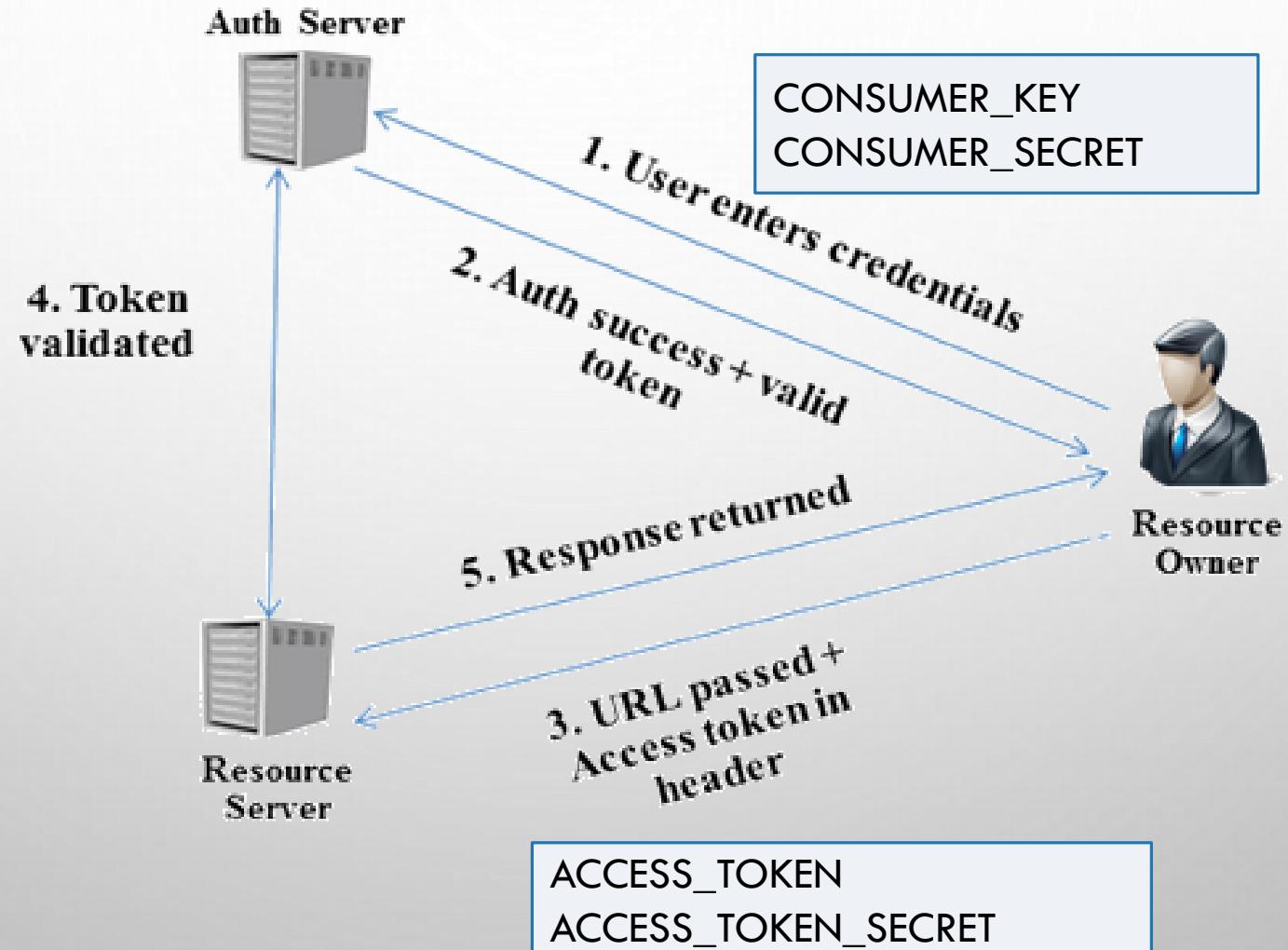
# HOW TO GET DATA FROM THE NYTIMES API USING R

- EXPLORE API
- CREATE OAUTH TOKENS
- GET JSON

# Recap: Data Mining from Twitter



# OAUTH dance



# 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 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/](https://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**

# Data Mining from YouTube

- Has an API:  
[https://developers.google.com/youtube/v3  
/docs/](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/](https://developers.google.com/apis-explorer/)
- The response is in JSON format.



YouTube

# 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...

# 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

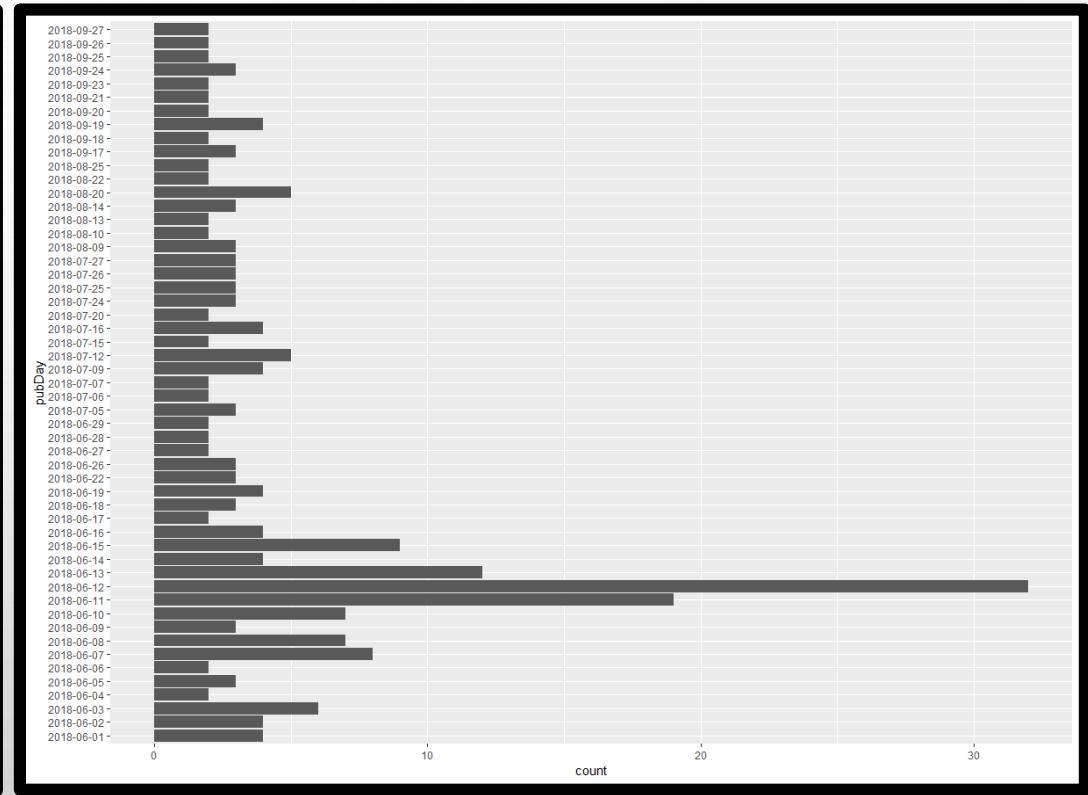
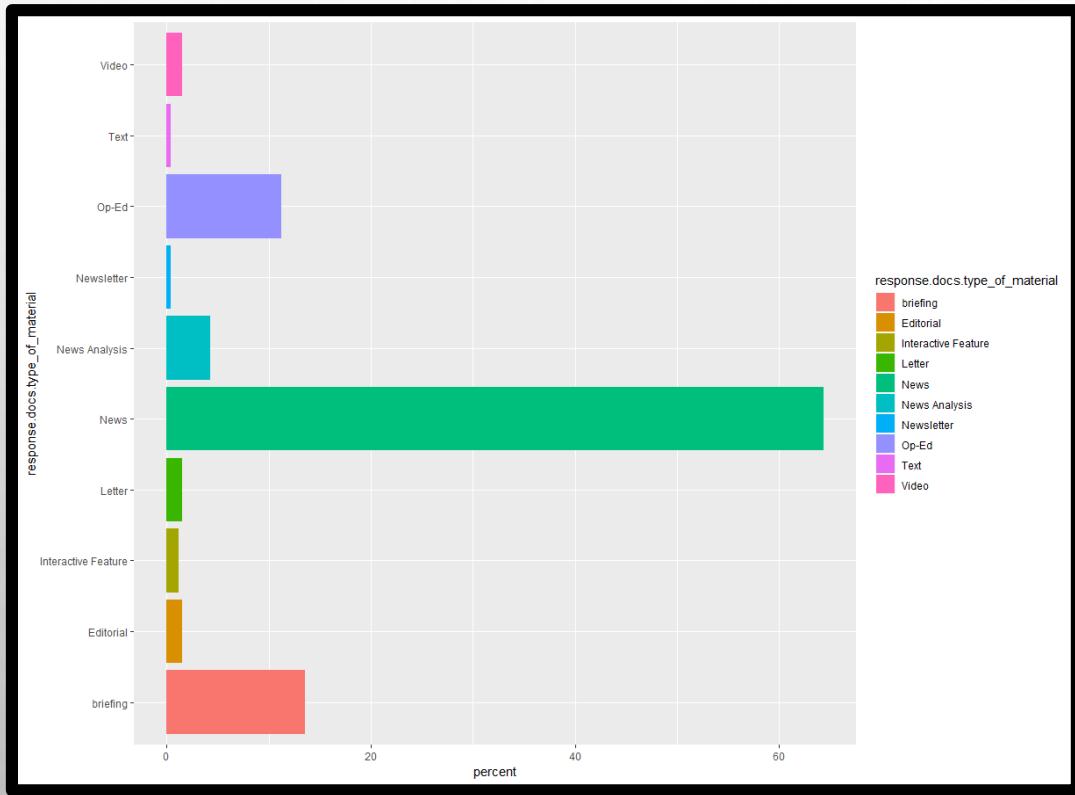


## HOW TO VISUALIZE DATA USING R

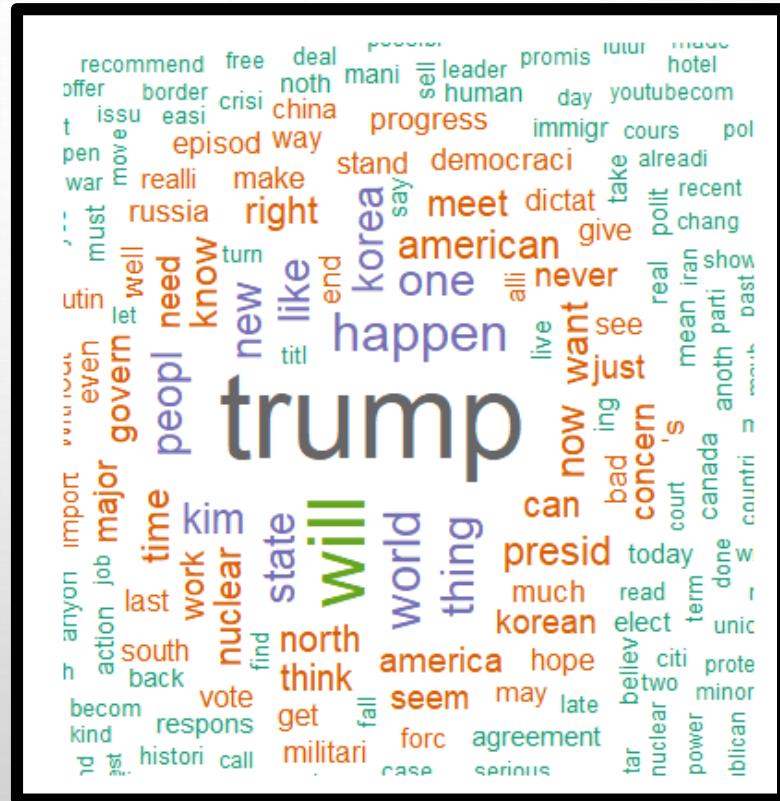
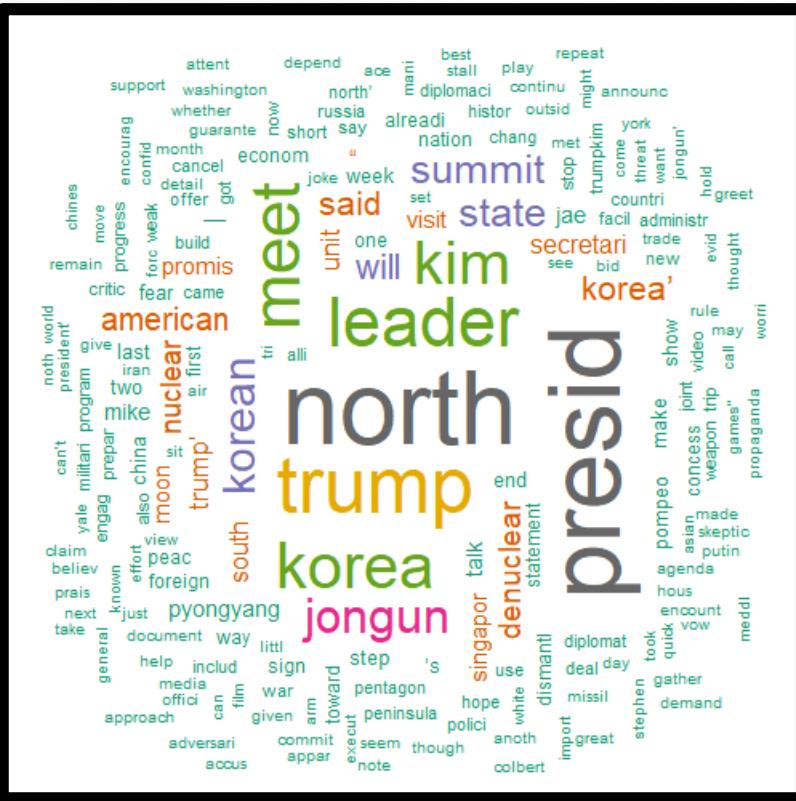
# HOW TO VISUALIZE DATA USING R

- TABLEAU
- GGPLOT
- SHINY

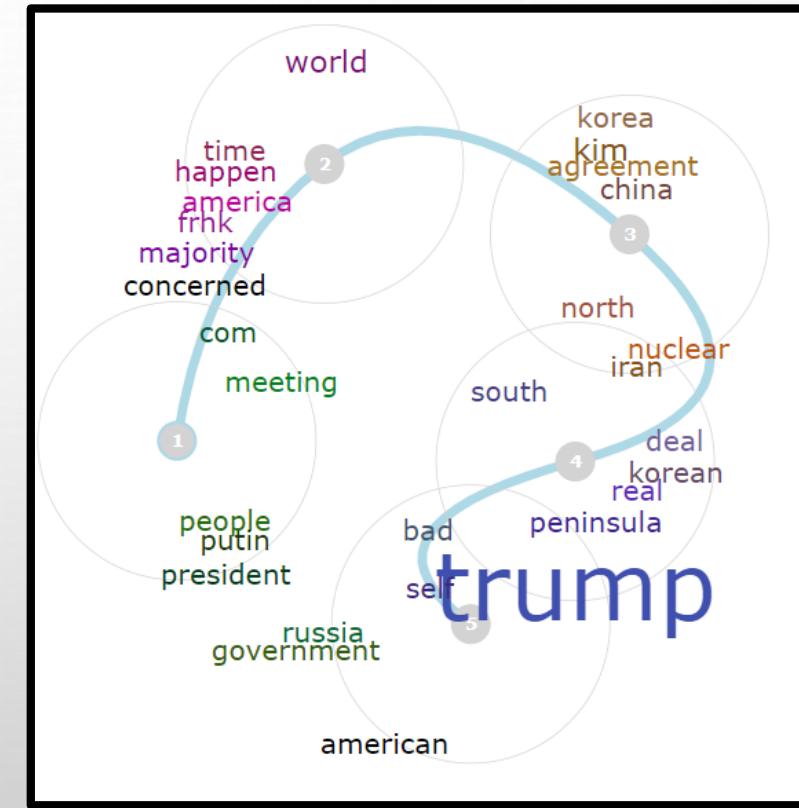
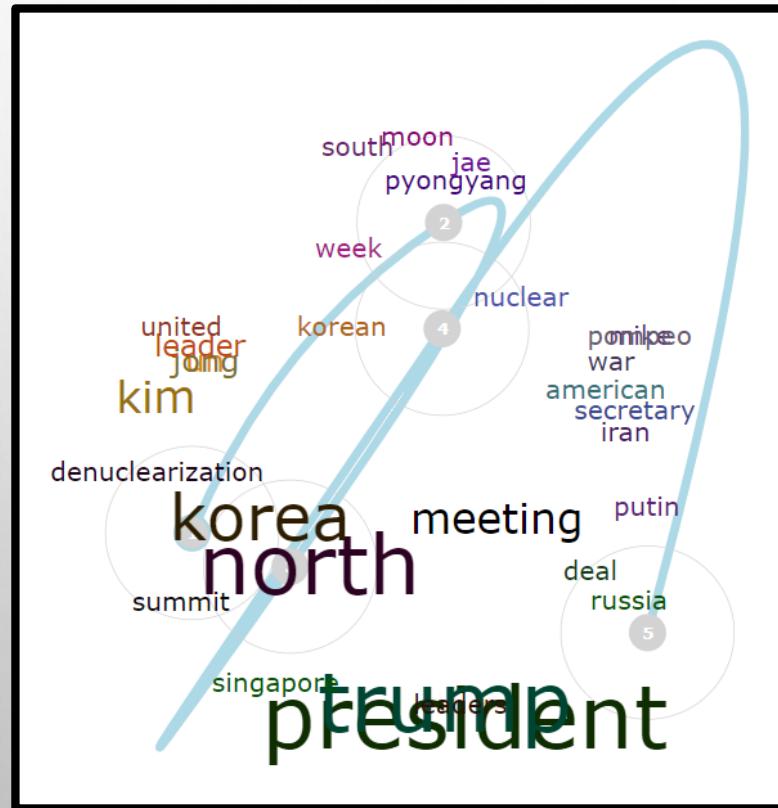
# BAR PLOTS



# WORD CLOUDS



# TRAJECTORIES



# Takeaways



## HOW TO PLAN YOUR DATA JOURNALISM PROJECT

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

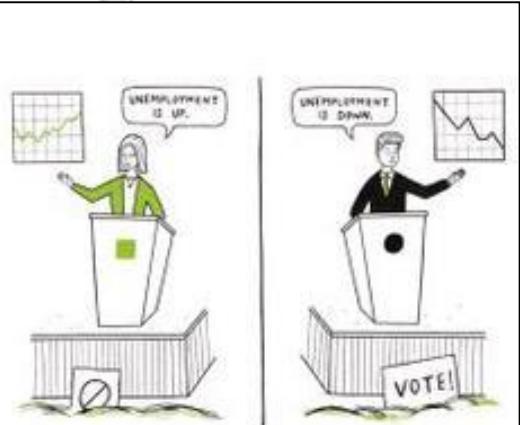
## HOW TO RETRIEVE DATA USING R

- EXPLORE API
- CREATE OAUTH TOKENS
- GET JSON

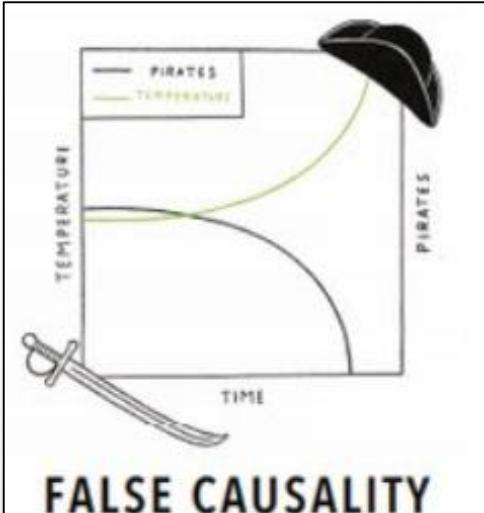
## HOW TO VISUALIZE DATA USING R

- TABLEAU
- GGPLOT
- SHINY

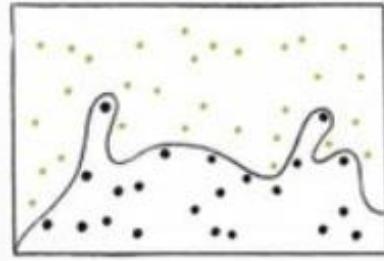
# LIMITATIONS



CHERRY PICKING



FALSE CAUSALITY



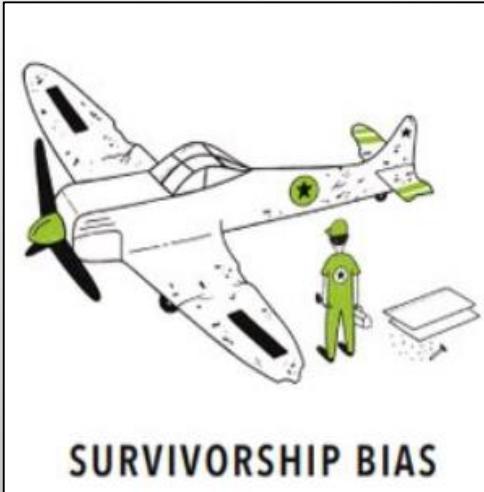
OVERFITTING



SAMPLING BIAS



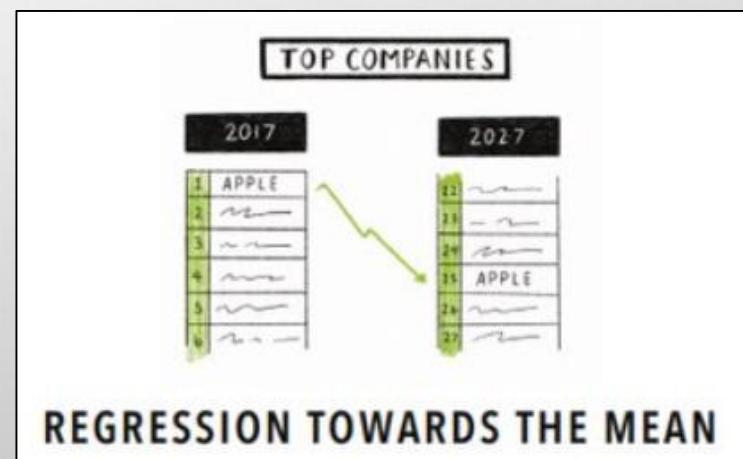
HAWTHORNE EFFECT



SURVIVORSHIP BIAS



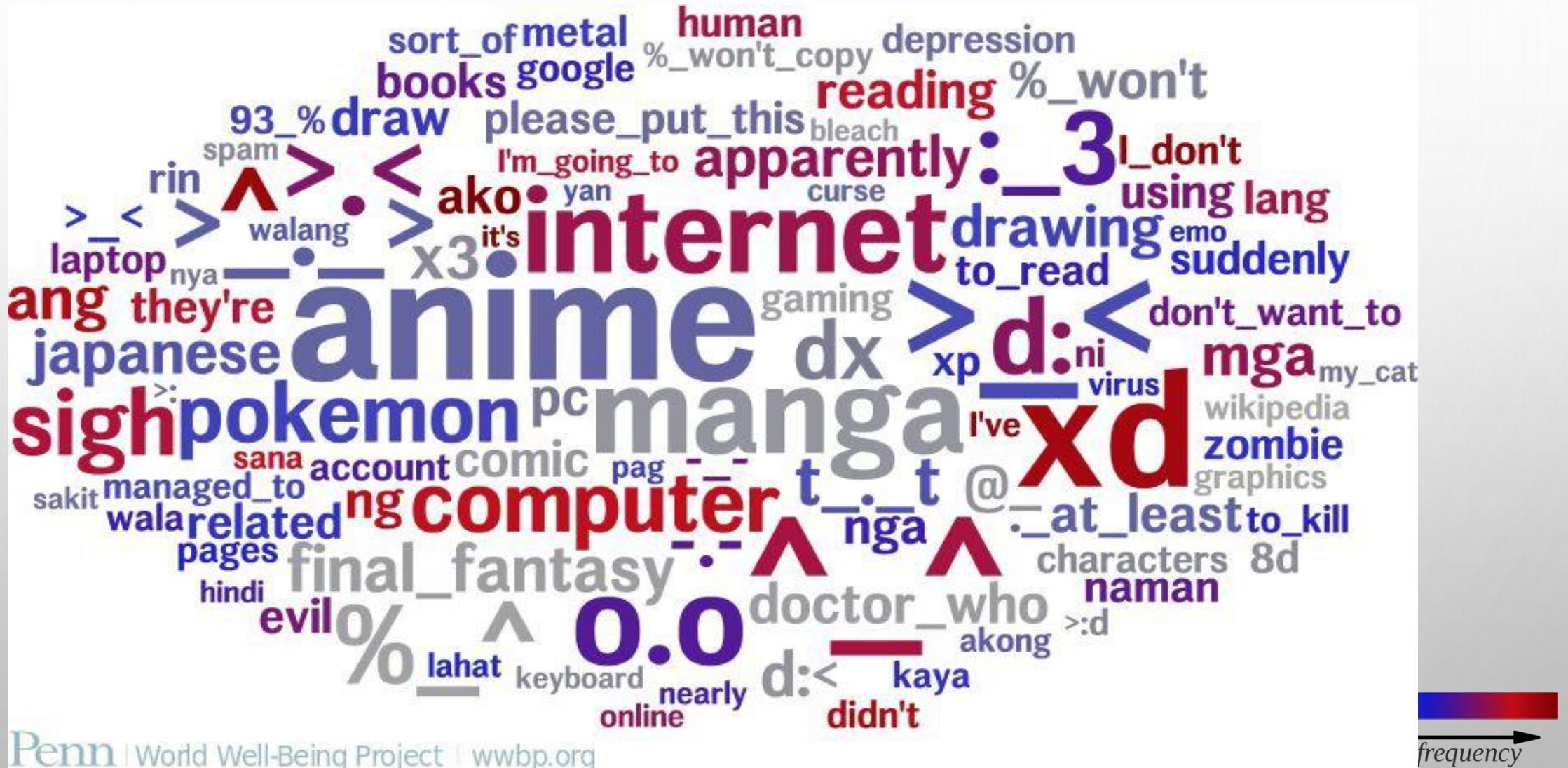
DATA DREDGING



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

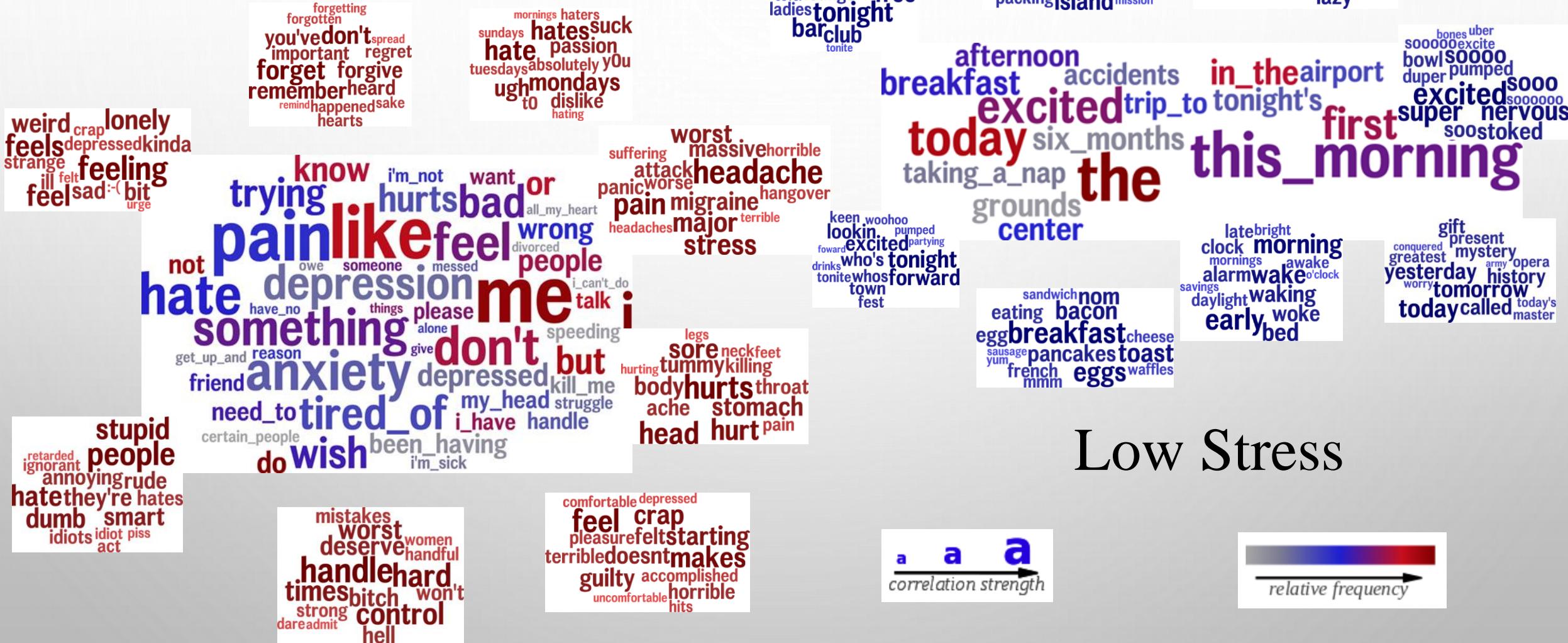


# Introversion



# WORDS FROM FB POSTS ASSOCIATED WITH STRESS

## High Stress



## 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](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](https://www.youtube.com/watch?v=4D6_kwY7nZQ)

# 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 "What posts to extract" section.
- 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 "Data to output" section.

**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](https://apps.facebook.com/netvizz/?fb_source=search)

148926458545119



UNIVERSITEIT VAN AMSTERDAM

# 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, mostly 'data' type, under a 'seed' node. Examples include 424207551102424\_5175265604, 424207551102424\_5129579555, etc.
- Selected Object Details:** A table showing key-value pairs for a selected object. Key columns include created\_time, message, and id. Value columns show their respective values: 2015-11-09T17:20:13+0000, Hello, Facebook! I finally got my very own page. I hope you'll think of this as a place where ..., and 424207551102424\_425063411016838.
- Custom Table Columns:** A section titled "Custom Table Columns (one key per line)" with a dropdown menu containing id, created\_time, and message.
- Data Request Commands:** A yellow box labeled "Data request commands" pointing to the "Parameters" section of the configuration pane.
- User Facebook Credentials:** A yellow box labeled "User Facebook credentials" pointing to the "Access Token" field in the configuration pane.
- Settings:** Includes "Base path" set to https://graph.facebook.com/v2.6/, "Resource" set to <user>/posts, "Parameters" set to <user> <ObjectID>, "Node level" set to 2, "Object types" set to seed,data,unpacked, and "Parallel Threads" set to 1.
- Status Log:** A log of API requests and responses, showing numerous entries like "Fetching data for 424207551102424 from https://graph.facebook.com/v2.6/42" and "Fetching data for 424207551102424 from https://graph.facebook.com/v2.6/42".



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 news from journalists and media outlets. Pg. 1  
20-02-2016 | 100% words | - | Section A; Column C; National Desk; Pg. 1  
... : Big, beautiful and impenetrable 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 | 100% words | - | Section A; Column C; National Desk; Pg. 1  
... 's organization, currently known as the Trump Organization, has been forced to deal with is Wall Street. Here bank executives are lending to him. You don't need money. Mr. Trump says he has no need for banks. "I don't use Wall ...

Trump holds firm on immigration. ○ Rep. Donald J. Trump, the presumptive Republican presidential nominee, has ...

01-09-2016 | 759 words | - | NEWS; Pg. 1A  
... from the beginning. "We will build a great wall along the southern border," said Trump, who was ... right away, and they did not discuss payment for the wall, though Peña Nieto said later in a speech in Mexico City that he would help finance it. ...

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

25-02-2016 | 368 words | - | 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

26-02-2016 | 1659 words | - | Section B; Column C; 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 ...

30-02-2016 | 405 words | - | NEWS; Pg. 1A  
... 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

10-03-2016 | 617 words | - | McGraw-Hill Construction Daily  
... driven sell-off might be far from over. The ... the 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 No Way They're Paying for Donald Trump's Wall

26-02-2016 | 360 words | - | Section A; Column C; National Desk; Pg. 1A  
... 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?" he asks. Without fail, Mr. Trump ... says he'll pay. ... We've got a nice government, he said, and about 20 ...

Trump Gambles on Immigration in Trip and Rally

01-09-2016 | 1003 words | - | Section A; Column C; 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 Republican Presidential Debate in Houston



# 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

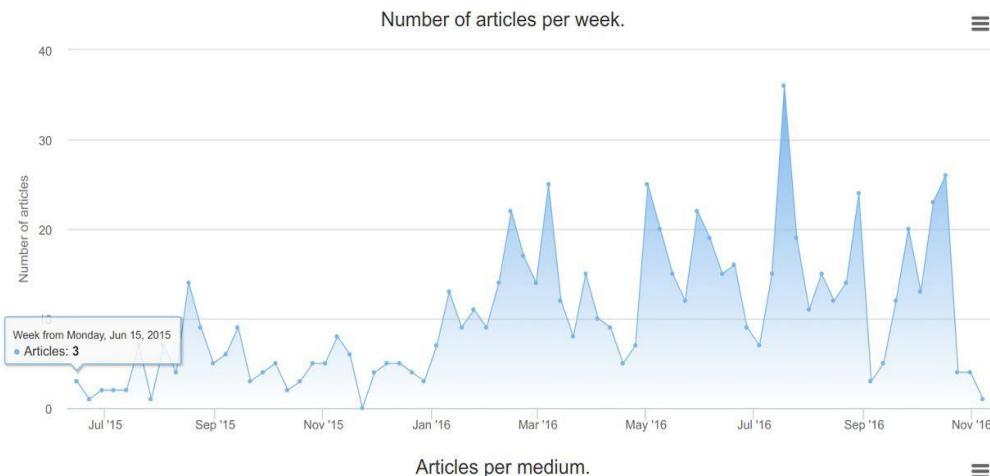
01-09-2016 | 1603 words | - | Section A; Column 0; National Desk; Pg. 1

... 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 Republican Presidential Debate in Houston

26-02-2016 | 23686 words | - | Section ; Column 0; Politics; Pg. 1

... 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



UNIVERSITEIT VAN AMSTERDAM

# AmCAT

# An API for power users



Attaching package: **scales**

## • WORKFLOW.

rescale

- Corpus/project management and explorative analysis using websites

## Reproducible queries using API

- Download text or results and connect to other tools (topicmodels etc)

```
> dtm.wordcloud(dtm_a)
```

Error in dtm.Wordcloud(a) : object 'dtm' not found

**Def**

> # dan kan je natuurlijk directe word

```
> dtm.wordcloud(dtm_a)
```

```
> cmp = corpora.compare(dtm_a, dtm_b)
> cmp = elmettareggendo(cmp, nvar)
```

```
> cmp = ptyx %>% arrange(cmp, over)
```

```
> with(head(cmp, 100), plotWords(x=loc
```

>



# 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

Go to file/function Run Source

```
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[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

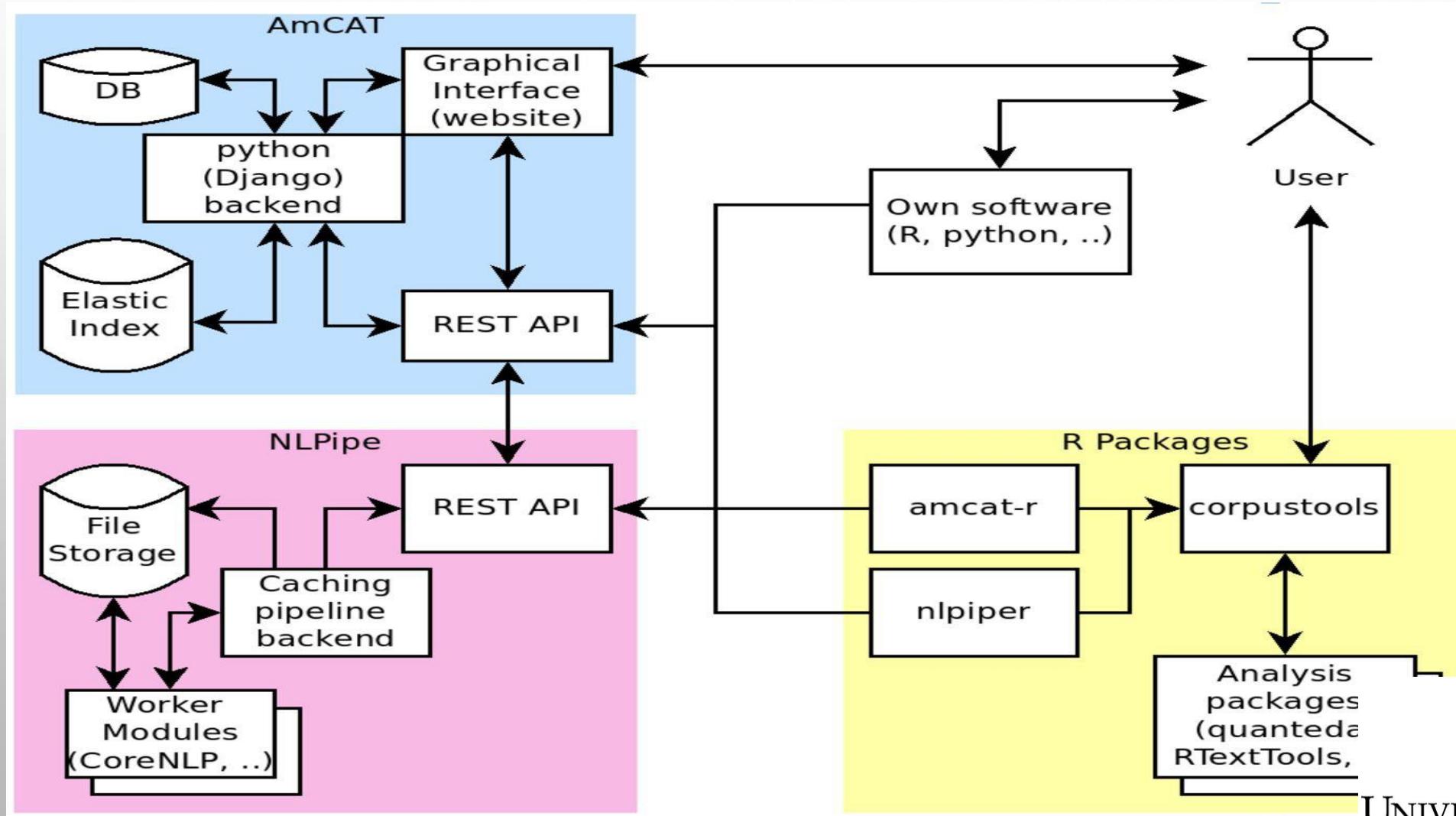
Word Cloud Visualization:

- Top 10 words (approximate frequency):

  - hebben
  - dat
  - van
  - wilders
  - trump
  - hi
  - nietvoor
  - aan
  - moeten
  - zeggen

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 M

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  
[kokiljaidka.wordpress.com](http://kokiljaidka.wordpress.com)

# Resources

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

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