

Working with Twitter Data

Joseph Allen – Research Associate at the
UK Data Service

Joseph.Allen@Manchester.ac.uk



Please validate me!

- Twitter - @JosephAllen1234
- E-mail – Joseph.Allen@Manchester.ac.uk

Warning!

- Real Tweets can be racist, sexist, homophobic and beyond!



TODO - In this webinar

- Why Social Media data? Why Tweets?
- Cool project ideas - sentiment, word clouds and beyond.
- **break**
- Is it ethical to use Twitter data?
- Twitter Analytics & TwArxiv.
- Pipedream Automation demo.
- **break**
- Twitter API, Twarc and beyond!

Why Social Media data?



An Individual - Facticity

- Age - 26
- Sex - male
- Birth Year - 1994
- Number of siblings
- Salary
- Vegan



Assumptions

- Our data is recent enough.
- Our data is still relevant.
- Our individual didn't lie to us.
- Our individual isn't in denial.
- Events outside our dataset can be ignored.

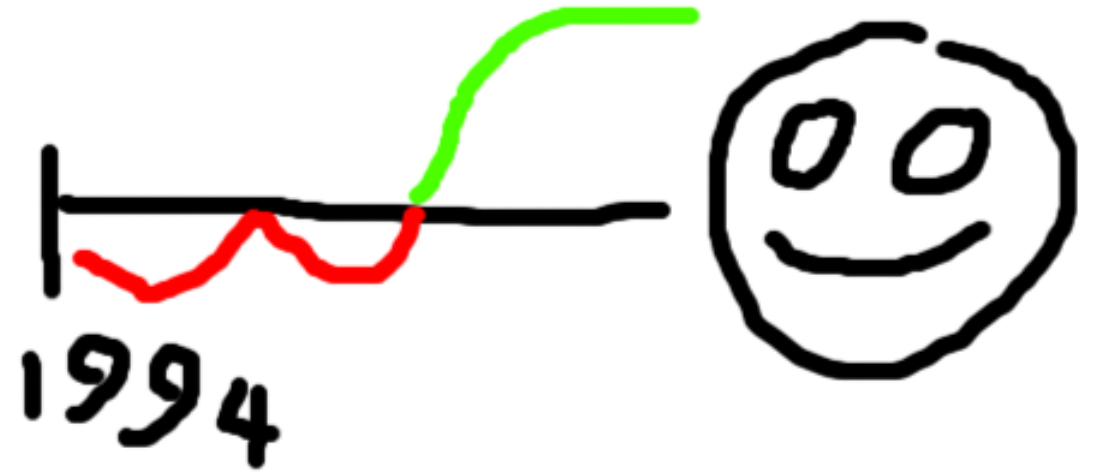


Expected Salary



An individual - Past

- Diet - Vegan
- How do they define vegan?
- Why are they vegan?
- How did they become vegan?
- Will they be vegan tomorrow?



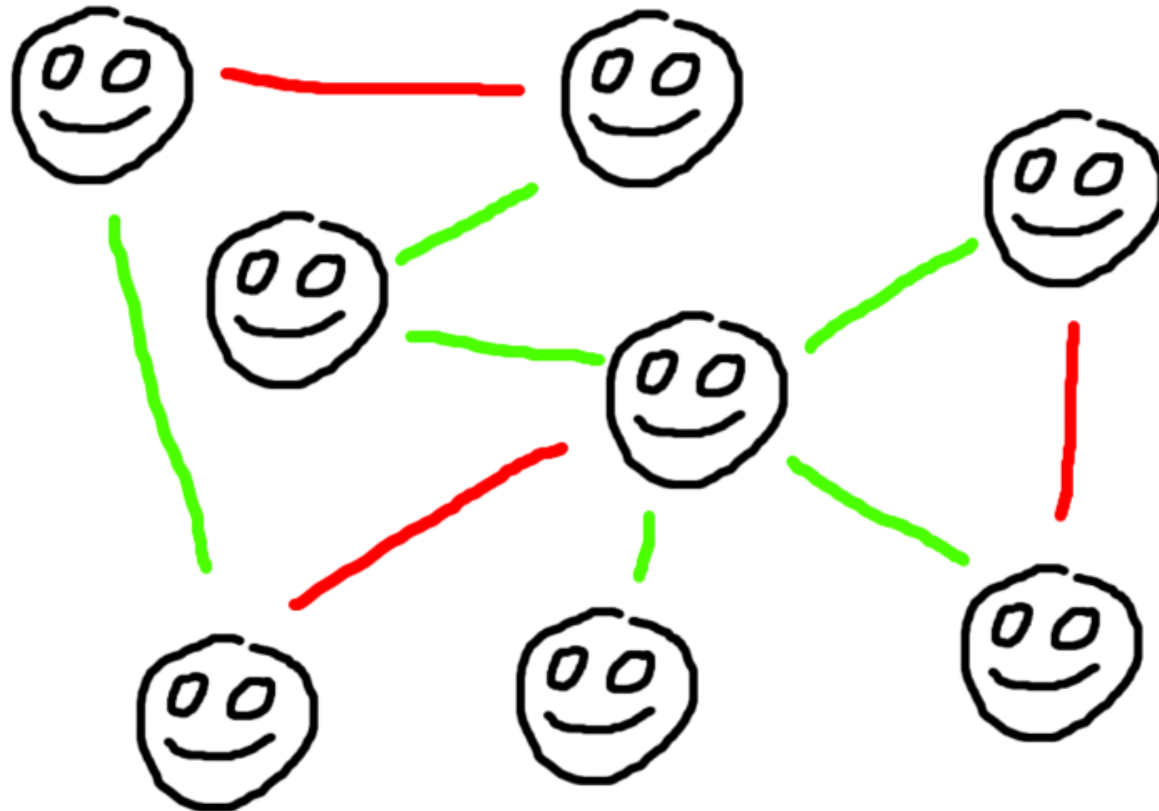
An individual - Transcendence

- Vegan, for now...
- Transcendence - The ability to transcend the immediate.
- Could we predict their future?
 - Vegans tend to continue being Vegan



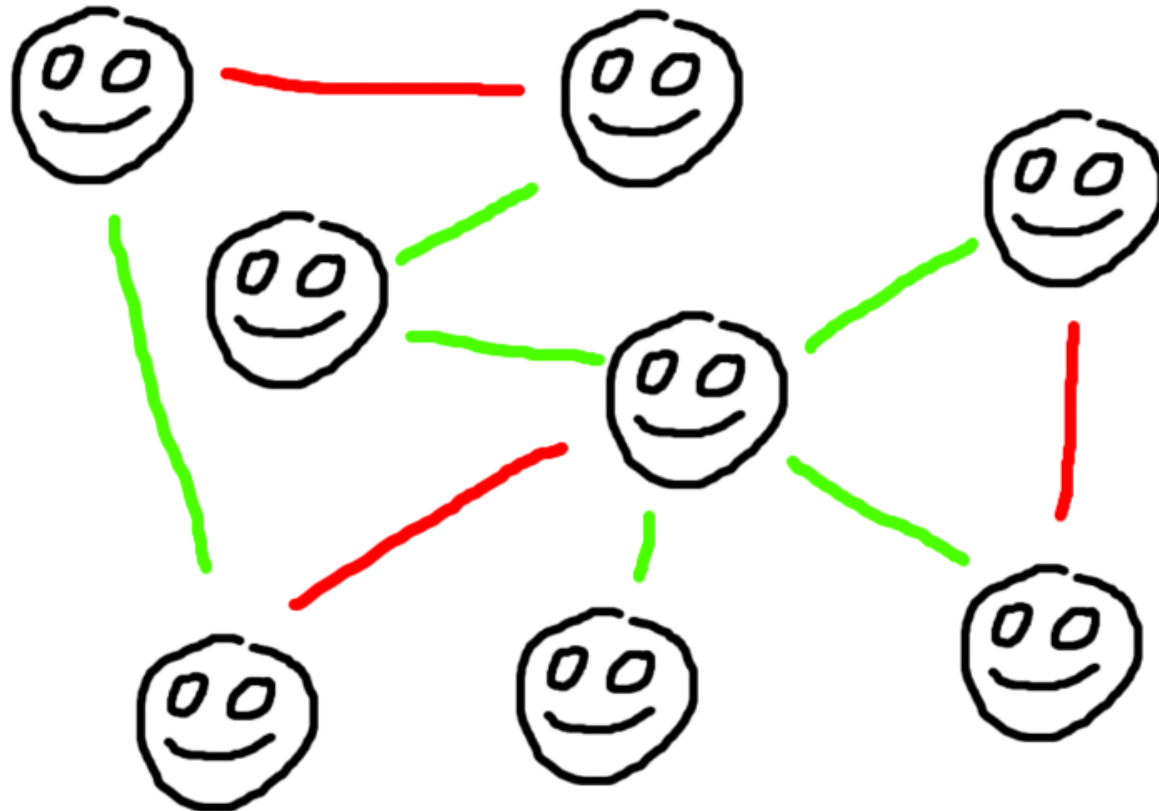
Enter Social Media

- Social Media - “Technologies that allow the sharing of information via virtual communities”.
- Free, well-documented and accessible data



The human approach

- How would a human predict “Will they be vegan tomorrow?”?
- We can ask.

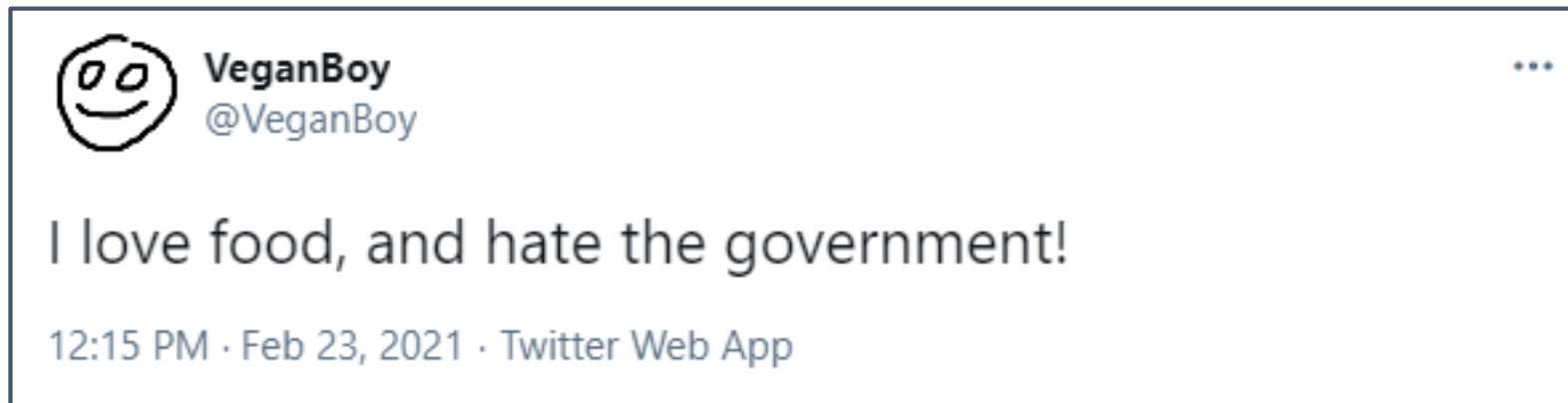


Why Twitter?



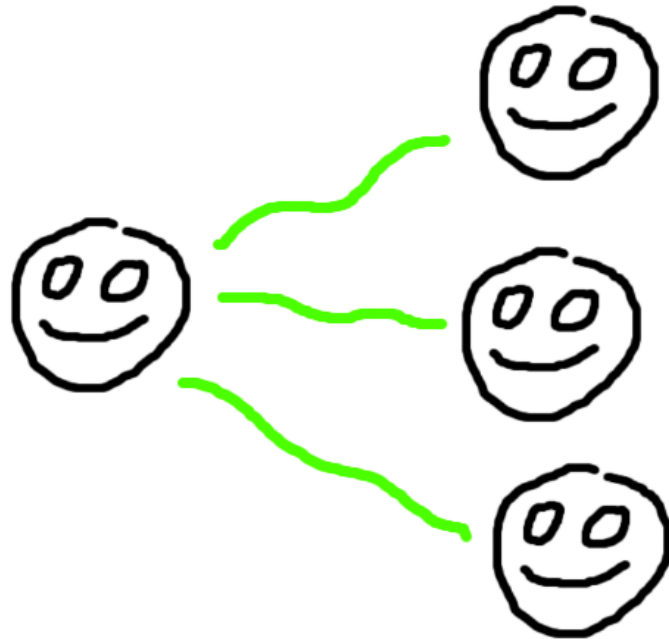
Enter the “Tweet”

- 280 character - succinct and high in sentiment
- Modern API for scraping and searching



Twitter makes this easy

- Twitter is a marketing platform
- They want you to find it valuable
- They want you to share your ad-friendly interests



Why not Twitter

- Open source packages quickly fall behind billion dollar companies.
- This is a marketing platform.
- Each platform has a bias:
 - Aged 24-35
 - 70% male
 - 80% of all tweets come from 10% of users.



Why Twitter

- Accessible
- Good Open source packages - twarc
- Academic tier
- Low-controversy
- Character limit



What cool things can we do with
Twitter data?



VeganBoy is negative



VeganBoy
@VeganBoy



I am getting pretty tired of this whole #Vegan thing

12:15 PM · Feb 23, 2021 · Twitter Web App



VeganBoy is positive



VeganBoy
@VeganBoy



It has never been easier to be Vegan, and it's finally getting cheap!



VeganBoy overshares



How sentiment analysis works? - Input

- I love cats - positive
- I am having an amazing day! - positive
- What a great view - positive
- I love you - positive
- I hate the weather here - negative
- I struggle to sleep - negative
- I hate it when I stub my toe - negative



Sarcasm pollutes models

- I love the government taxing my hard-earned money - negative
- I hate you, you are so silly- positive



How sentiment analysis works? - Output


- I love cats
- I am having an amazing day!
- What a great view
- I love you
- I hate the weather here
- I struggle to sleep
- I hate it when I stub my toe



So how does VeganBoy feel, mathematically?



A screenshot of a tweet from the user VeganBoy (@VeganBoy). The tweet text is "I love food, and hate the government!". The words "love", "hate", and "government" are underlined in green, brown, and red respectively. The tweet is timestamped "12:15 PM · Feb 23, 2021 · Twitter Web App".

 **VeganBoy**
@VeganBoy

I love food, and hate the government!

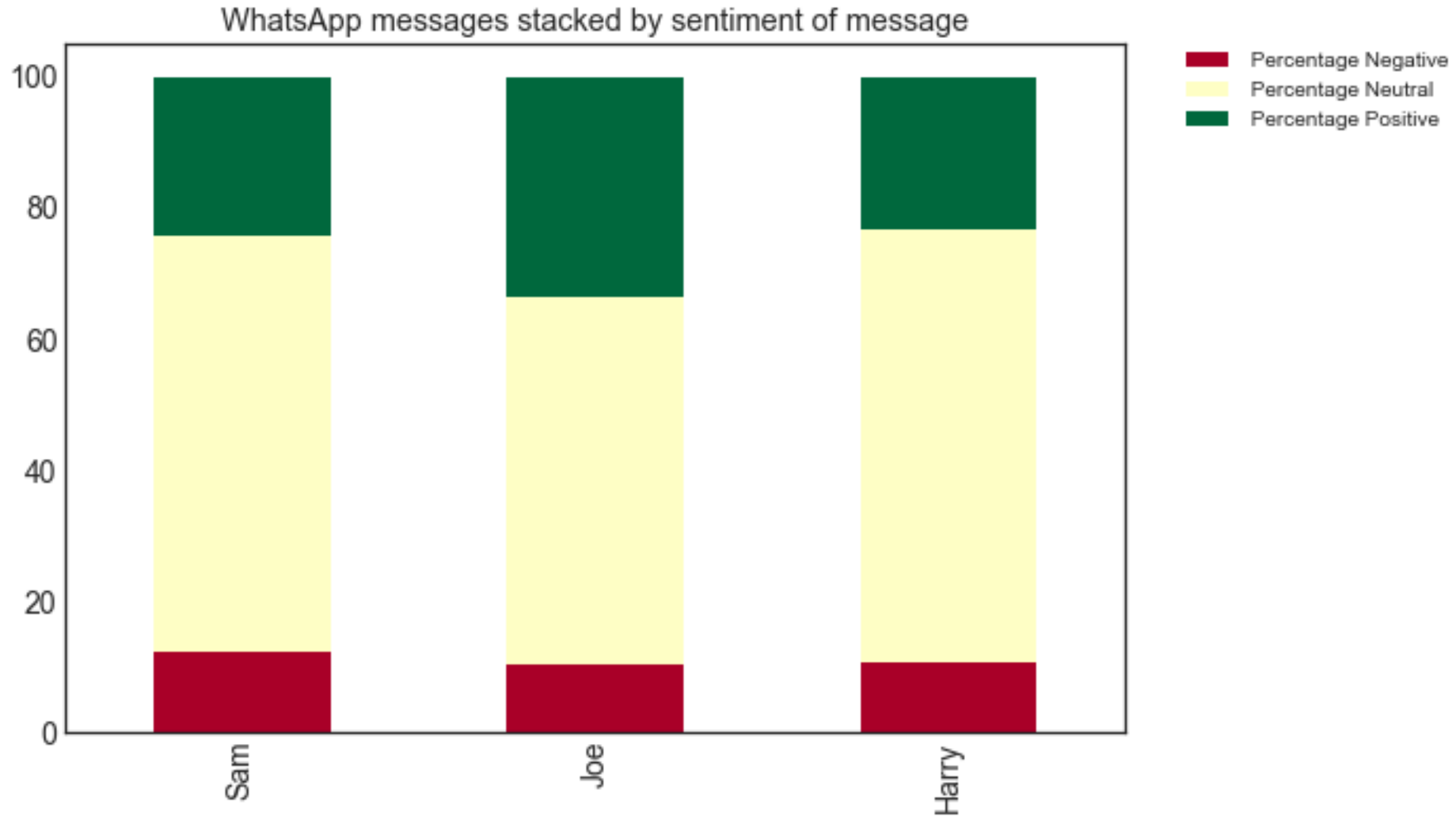
12:15 PM · Feb 23, 2021 · Twitter Web App



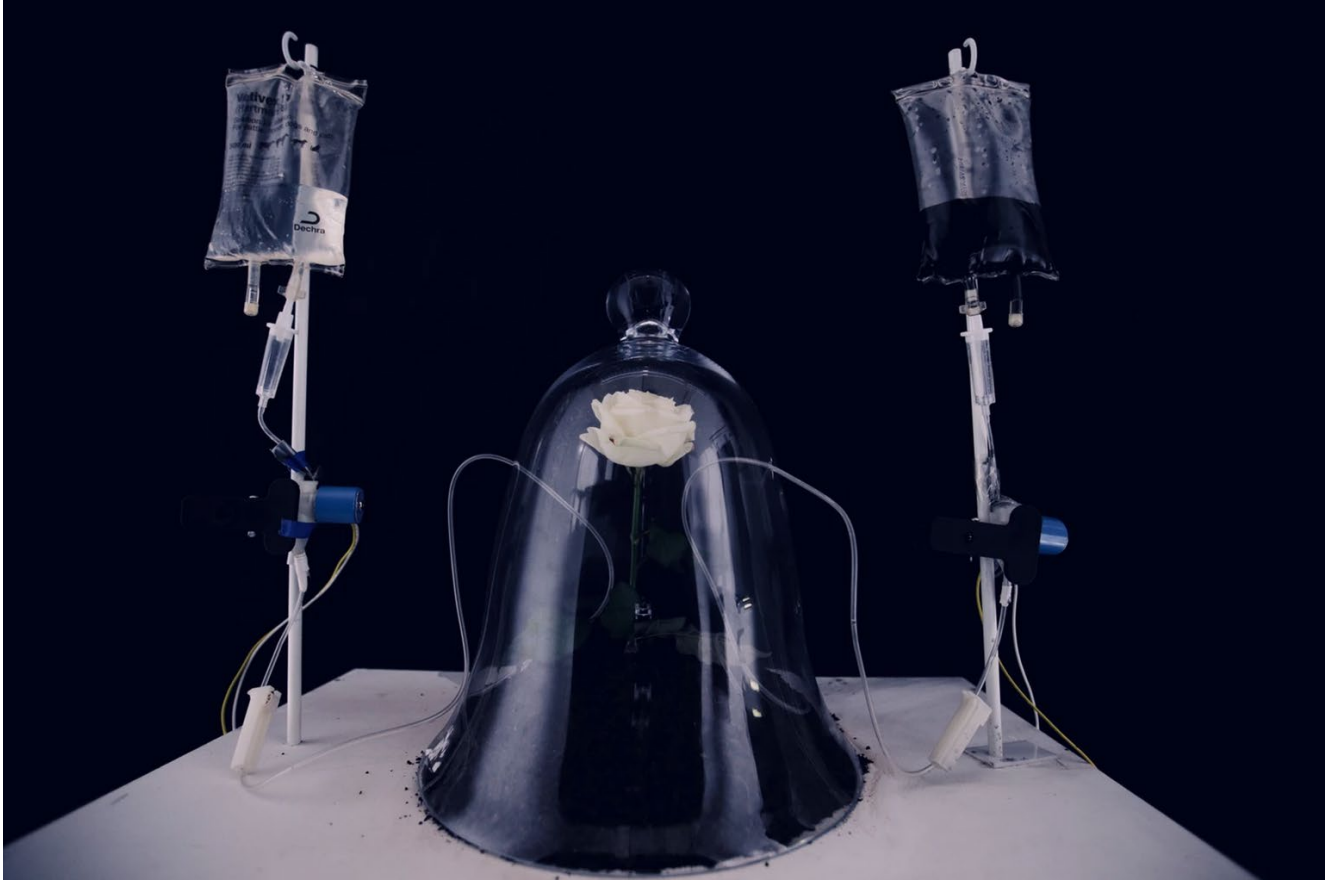
Cool Sentiment Analysis project 1 - Analyse your friends



Cool Sentiment Analysis project 2 - Analyse your friends 2



Cool Sentiment Analysis project 3 - The Rose



The Rose @rosetest18 · Jan 8, 2019
See @GOPSenate HATE-IN-PROGRESS @ChuckGrassley IS CHUCKIE insulting New Yorkers or lowans? Does CHUCKY want... [twitter.com/i/web/status/1...](#)
pos: 0.0
neg: 0.186
neu: 0.814
compound: -0.4939

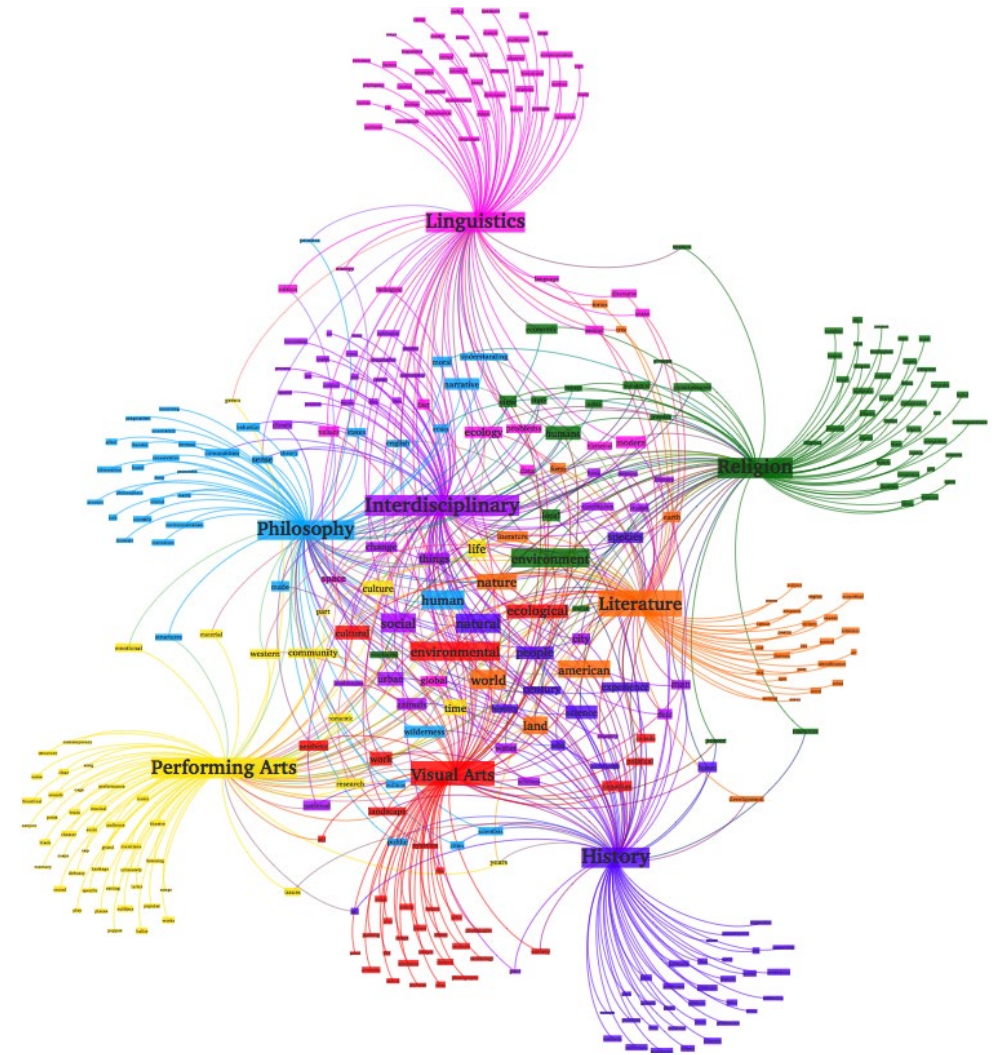
The Rose @rosetest18 · Jan 8, 2019
RT @BettyBowers: For your consideration, @TheAcademy:
Best Supporting Actor and/or Actress in Drama.
[twitter.com/LindseyGrahamS...](#)
pos: 0.372
neg: 0.0
neu: 0.628
compound: 0.7964

Lindsey Graham @LindseyGrahamSC · Jan 7, 2019
Good decision by President @realDonaldTrump to address the nation about the crisis at our southern border.
Great opportunity to explain to the American people why we need more border security funding.
The current situation at the border is indeed a national security crisis.



Topic Analysis

- Link similar words - love, want, care
- Similar topics - Vegan, veganism,veggie
- Opposite of veganism - bbq, steak



Is it ethical to use Twitter data?



Let's jump over to Mentimeter



We should be able to use this data

- This is public data, we all knew it was public.
- User's should know better than to post something that is racist, homophobic, sexist or beyond.
- It's a very important dataset for opinions, politics, share prices and beyond.
- It's public so there's no privacy concern.



We shouldn't use social media data

- Just because it's public, doesn't mean everybody look at it.
 - Do children know how Twitter data is used?
 - Did all users anticipate Twitter data would be used to target our adverts, and influence politics?
 - How will it be used in 5 years?
- Should users be informed their data is being used, and what for, how would they feel, how would you feel?
- Under GDPR, are the contents of a Tweet a breach of personal data?
- How do we give users the right to be forgotten and withdraw from studies?



University of Warwick on Social Media data

- Appears “publicly available”
- What is the solution for:
 - Informed consent
 - Anonymity
 - Right to withdraw
- More details here [Warwick Research & Impact Services](#)



Let's get technical – No
Programming required



Twitter Analytics



TwArxiv – twarxiv.org



TwArxiv

- Bastian - @gedankenstuecke
- <https://openhumans.org/>
- \$5,000 grants for grabs - www.openhumans.org/grants/

















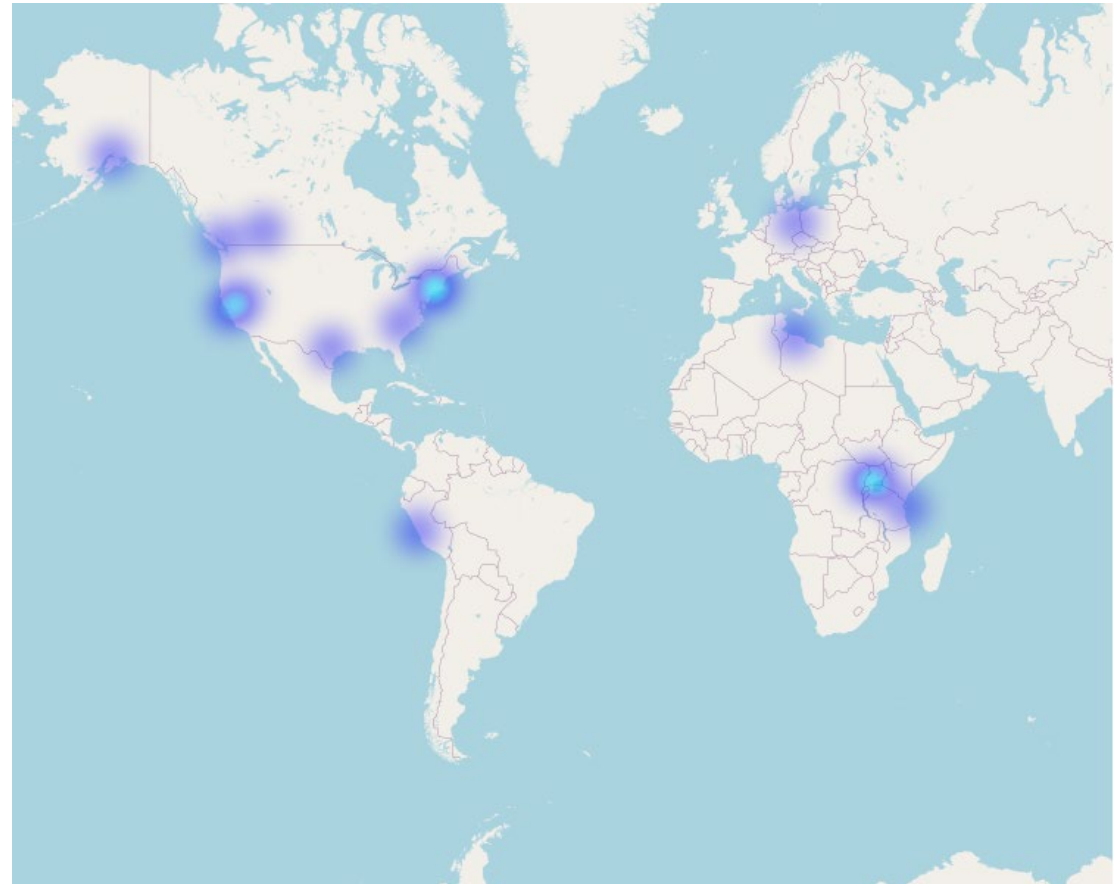
Where have I been tweeting?



Where has somebody else been tweeting?

Public visualizations

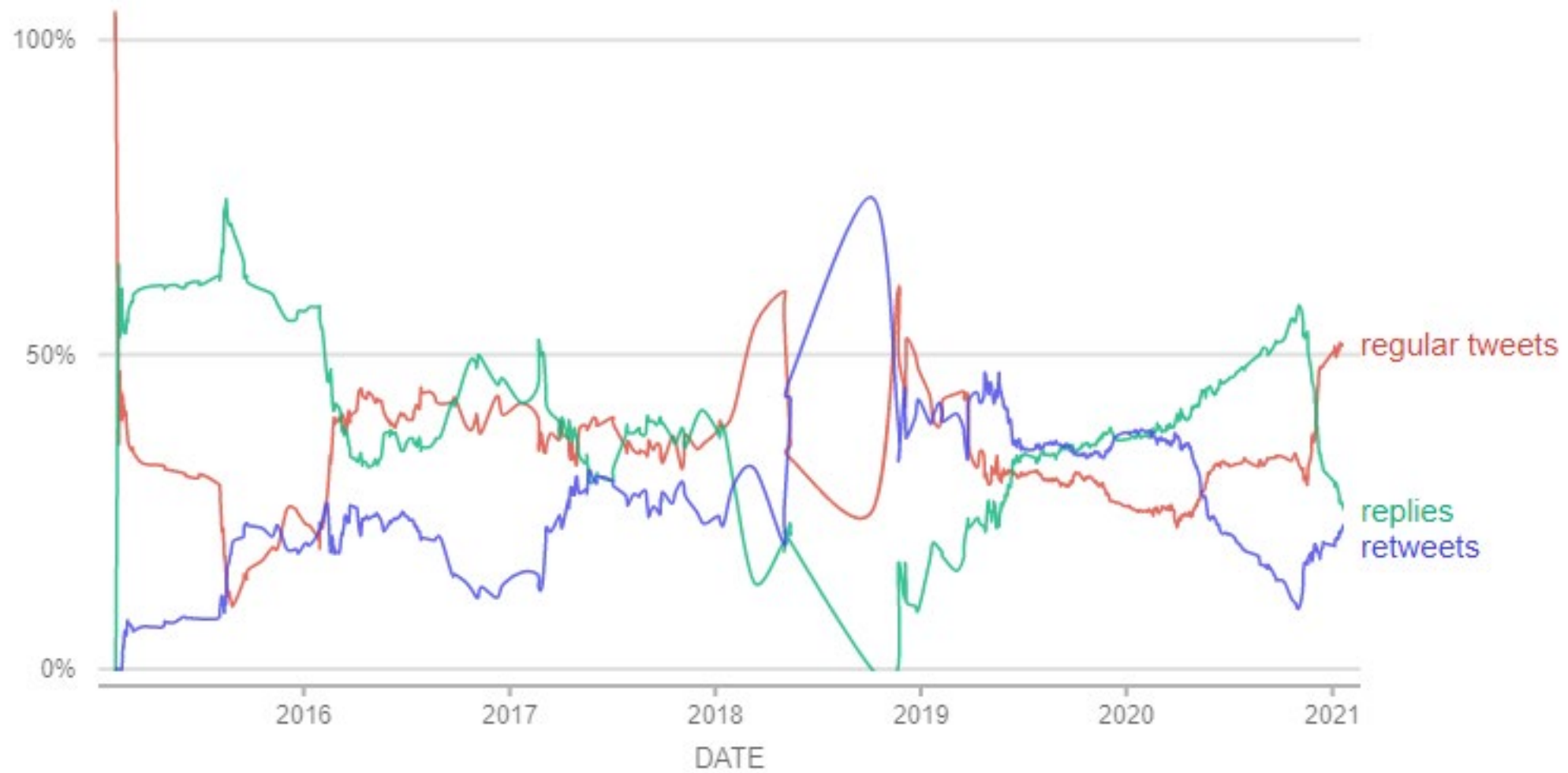
User	Graphs
02198761	General  Interactions 
04131072	General  Interactions 
05804139	General  Interactions 
08276171	General  Interactions 
09042053	General  Interactions 
09900540	General  Interactions 
10351608	General  Interactions 



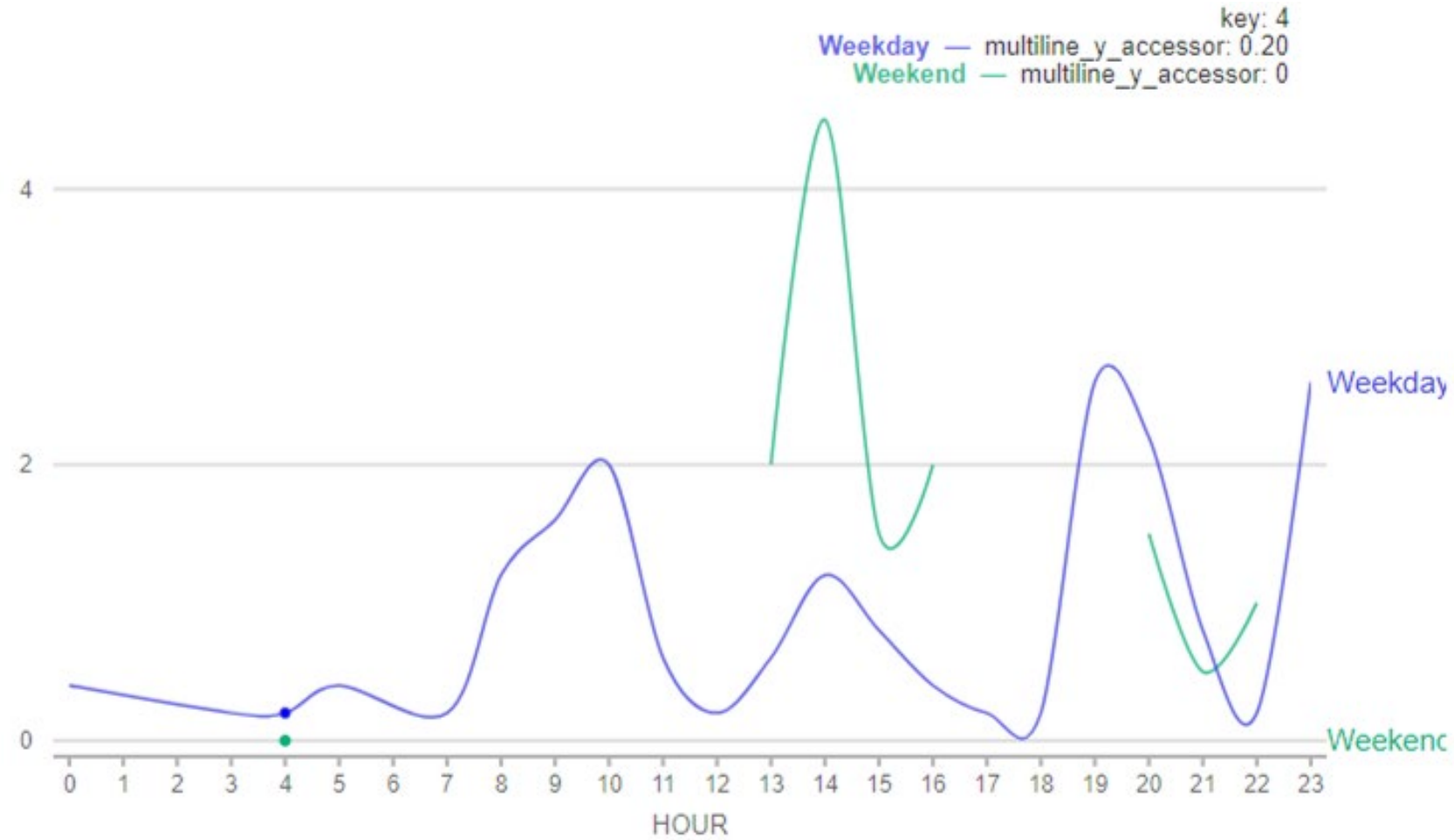
Tweets per day



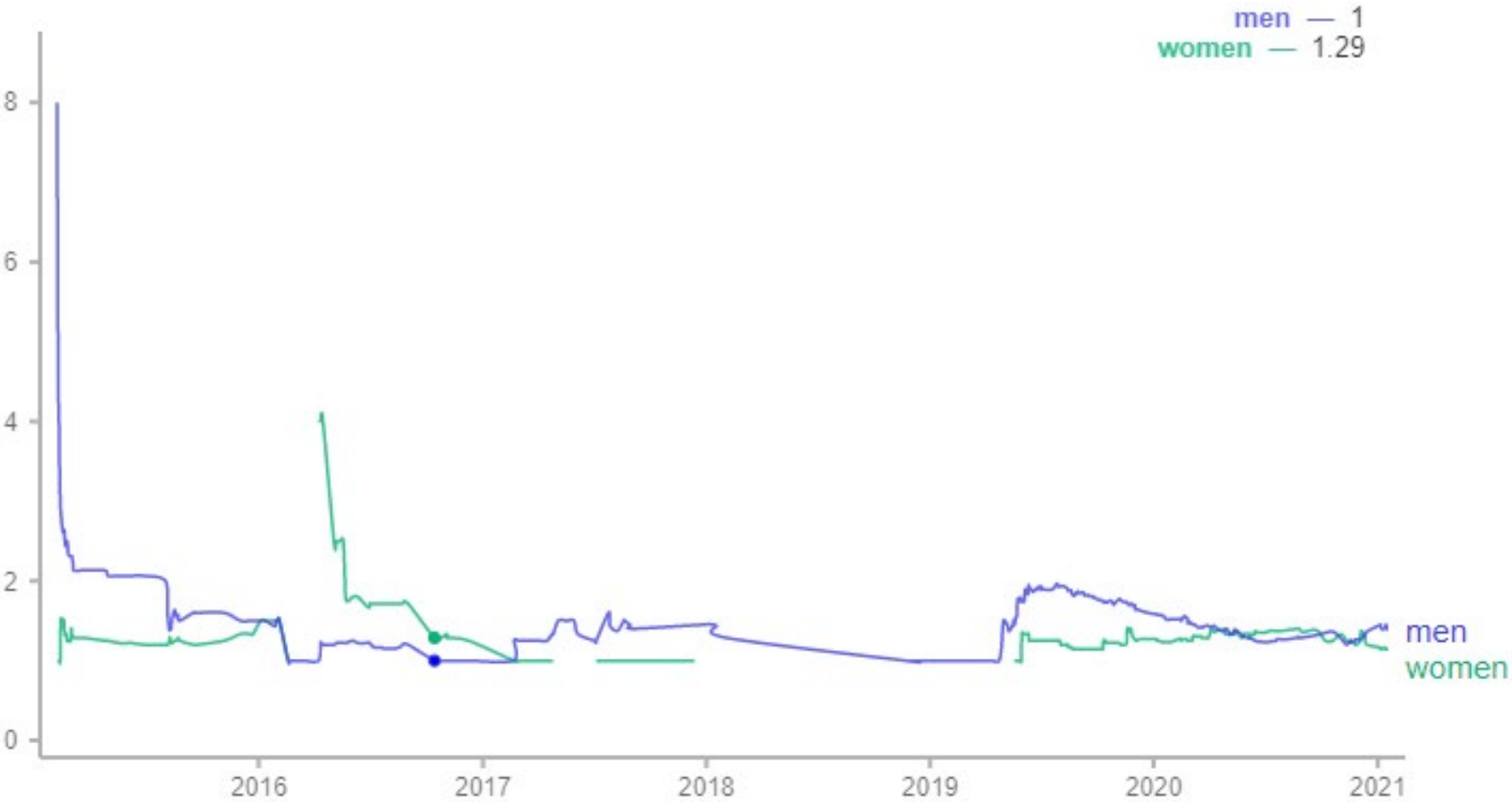
Tweet classifications



Tweets by hour



Replies by gender



Retweets by gender




Pipedream



What is Pipedream?

- Listen for Triggers
- Perform actions
- We can use our personal Twitter

Trigger workflows on any event or schedule



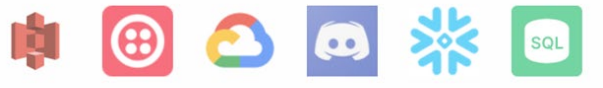
⋮

Run any Node.js code

```
// Require any npm package
const axios = require("axios")
// event contains your event data
console.log(event)
```

⋮

Use managed auth for 300+ apps




Pipedream Demo



Pipedream Review

- Requires no programming
- No messy API keys
- New Tweets only

Trigger workflows on any event or schedule



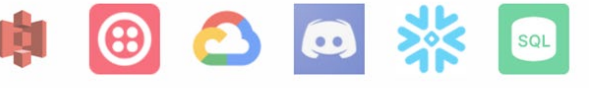
⋮

Run any Node.js code

```
// Require any npm package  
const axios = require("axios")  
// event contains your event data  
console.log(event)
```

⋮

Use managed auth for 300+ apps

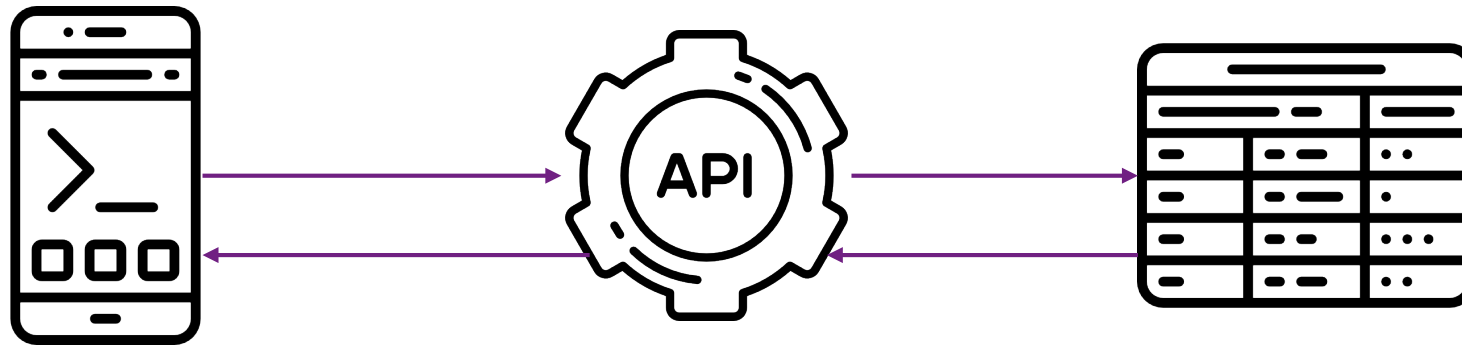


Twitter API



What's an API?

- API – Application Plugin Interface
- An API let's us access data



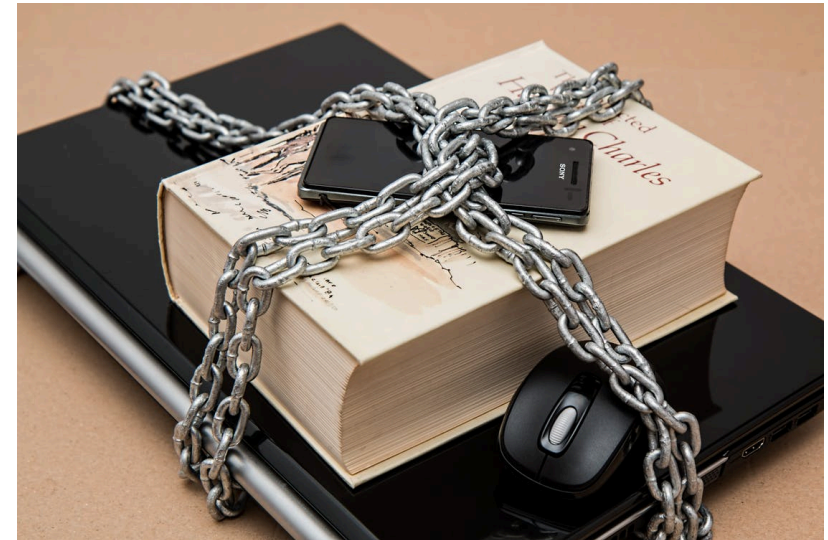
APIs as a restaurant

- A website/phone app as a food menu
- An API as a waiter
- A data source as a kitchen



Authentication

- APIs control the web
- We can't give everybody, all data, all the time
- Every social media chooses how they deliver this



All I wanted was Twitter data...

- Why are we talking about APIs and Authentication?
- Overwhelmed is the right feeling
- Without the ability to lock doors, we wouldn't build a home.



Twitter API

- Lookup Tweets and Users
- Search recent Tweets or historic data
- Filter or sample real-time tweets



Premium Tier

- Access 30 days of data using the premium tier
- More than 30 days of Tweets publicly available



Academic Tier

- Full archive search
- Requires your academic details



Twitter API review

- Requires strong programming
- Tiered Access
 - Free – 7 days
 - Premium – 30 days
 - Academic - Full access
- API keys, rate limits and beyond



Twarc - Demo



Exploring Twitter Data



What Tool is for me?

Tool	Programming	API	User Access	Tweet Volume	Tweet Age	Purpose
Twitter Analytics	X	X	Account	All	Full	Inspiration
Twarchive	X	X	Open data	All	Full	Inspiration
Pipedream	X	✓	All	500,000 / month	7 days	Slow collection
Twitter API	✓	✓	All	10,000,000 / month	Full archive	Historic collection
Twarc	✓	✓	All	10,000,000 / month	Full archive	Historic collection



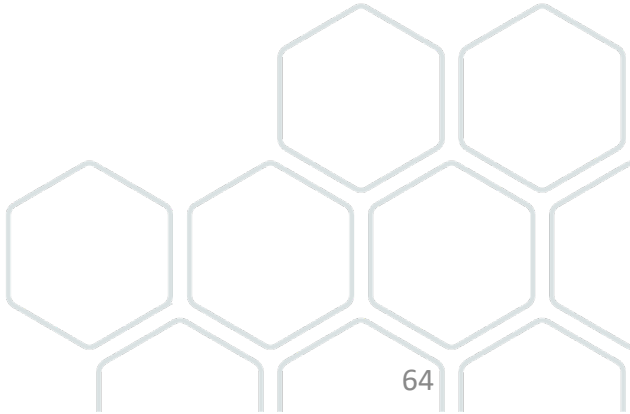
Thank you.

Joseph Allen

@JosephAllen1234

Joseph.Allen@Manchester.ac.uk

Any questions?



Sources

- [Blog post summary of this talk](#)
- [Warwick Ethical Approval](#)
- [Manchester Research IT](#)
- [Academic tier](#)
- [Slides](#)



Technical Resources

- [Using the Twitter API](#)
- [Using Twarc](#)
- [Tidying and analysing Twarc output](#)



I want to learn more!

- [Building Word Clouds](#)
- [Twitters tutorials](#)

