# The Twitter Timeline

Peter Smyth

Diced project data manager

UK Data Service Webinar

20 April 2021

# What we will look at

1. Keys & Authentication
2. The new API and developers account
3. What is and what isn't in a users Timeline
4. Extracting a complete timeline ,
                    while avoiding the API rate limits
5. Stitching things together
6. Conversations
7. Analysis of the timeline contents.

# Keys & Authentication

- API v2 now uses a 'Bearer Token' for access
- Overall this makes things easier
- But you still need to get one before you start!
  - And to get one you need a Twitter account!

# The APIs

- Currently v1.1 and v2
- V1.1 well established
- V2 still being developed, but simpler to use

- We will assume that any one new to the Twitter API would choose v2
- This is a non-technical Webinar, so for today it doesn't matter

# The Old System (v1.1)

- You can however still use it if you have the right set of keys

- You can also use it with some v2 keys

- If you have old keys and legacy code, will still work

- No talk about when it will be shut down - that I've seen!

# The new system

- Dashboard!
- Changes to endpoints
- Changes to data delivered
  - Instead of getting everything
  - You now have to say what data you want returned

# Old System returned data

Twitter_Python > code > {} realDonaldTrump_20200909_from_13022267676551741441_inc.json > {} 0 > {} quoted_st

```
1  [{"created_at": "Wed Sep 09 16:32:24 +0000 2020", "id": 1303733001919049730,
2      "id_str": "1303733001919049730",
3      "full_text": "Congratulations @CorkyForSenate! https://t.co/NQk9w5TKde",
4      "truncated": false, "display_text_range": [0, 32],
5      "entities": {"hashtags": [], "symbols": [], "user_mentions": [{"screen_name":
6          "in_reply_to_status_id_str": null,
7          "in_reply_to_user_id": null, "in_reply_to_user_id_str": null,
8          "in_reply_to_screen_name": null, "user": {"id": 25073877,
9              "id_str": "25073877", "name": "Donald J. Trump",
10             "screen_name": "realDonaldTrump", "location": "Washington, DC",
11             "description": "45th President of the United States of America\ud83c\u
12                 "verified": true, "statuses_count": 55670, "lang": null,
13                 "contributors_enabled": false, "is_translator": false,
14                 "is_translation_enabled": true,
15                 "profile_background_color": "6D5C18",
16                 "profile_background_image_url": "http://abs.twimg.com/images/them
17                  "profile_text_color": "333333", "profile_use_background_image":
18                   "has_extended_profile": false, "default_profile": false,
19                   "default_profile_image": false, "following": false,
20                   "follow_request_sent": false, "notifications": false,
21                   "translator_type": "regular"}, "geo": null, "coordinates": null
22                    "place": null, "contributors": null, "is_quote_status": true,
23                    "quoted_status_id": 1303533048504483840,
24                    "quoted_status_id_str": "1303533048504483840",
25                    "quoted_status_permalink": {"url": "https://t.co/NQk9w5TKde",
26                        "expanded": "https://twitter.com/nhgop/status/130353304850
27                        "display": "twitter.com/nhgop/status/1\u2026"},
28                        "quoted_status": {"created_at": "Wed Sep 09 03:17:51 +0000
29                            "expanded_url": "https://twitter.com/NHGOP/status/1303
30                            "sizes": {"thumb": {"w": 150, "h": 150, "resize": "cro
31                            "medium": {"w": 1200, "h": 630, "resize": "fit"},
32                            "large": {"w": 1200, "h": 630, "resize": "fit"},
33                            "small": {"w": 680, "h": 357, "resize": "fit"}}}]},
34                            "source": "<a href=\"http://twitter.com/download/iphon
```

# Academic accounts

- 10,000,000 Tweets / month
- Access to Full Search
- Need to apply and be accepted

# Academic Accounts – Application process

- Make sure you have a Twitter Account before you start
- Complete relatively short online form
- Wait a day or so and hope

- Not sure what actual criteria are, but you are asked about You and Your project

# Academic Accounts – Application process

- Online form (Edge)

# Timelines  v  What do you see

Four things to consider:

1. What is included
2. What is not included
3. What is not included but you can get
4. What is not included and you cannot get

# Included in what you see

- Your own Tweets
- Tweets from those you follow (your friends)
- Twitter recommendations
- Adverts
- Tweet Warnings

# Not included (in what you see)

- All the tweets your friends see
- Their recommendations and Ads

# Timeline – What you can get

- All of your Tweets
- All of your Friends Tweets

# Timeline – What you **cannot** get

- Deleted Tweets
- Anything from banned accounts
- Adverts
- Recommendations (from Twitter)

# Example Twitter Feed

# Software choices (Python based)

1. Python Tweepy
2. Python twarc
3. TwitteR
4. Raw API (via Python Requests)

Plus many more possibilities

# Example : The UKDS Timeline

- Basic Stats:
  - 3248 Tweets
  - Earliest date is : 2018-06-05T14:32:34.000Z

  (Tweet rate of 3.1/day)

# Example : The UKDS Timeline

Made up of:

  691   Original Tweets

2296   Re-Tweets

 103   Quoted Tweets

 158   Replies

# Significance of the Tweet Rate

- The higher the Tweet rate, the shorter the Timespan of in the Timeline (limited to 3200 most recent Tweets)
- May be as short as a couple of weeks
- Need to collect them promptly (assuming you can)

# Significance of the Tweet Rate

- Example: DiCED
- Want 905 complete Timelines
- Total of 300030 unique Friends Timelines
- Took over 16 days to collect

# Significance of the Tweet Rate

- Situation easier with Academic account as you have access to full search and can specify start and end times

# Collecting the data

We need:

1. The UKDS Timeline
2. A list of the UKDS Friends
3. The Timelines of all of UKDS Friends

Although not part of the timeline, it is useful to have User details of the UKDS and their Friends (UKDS_friends.csv)

# Example: The UKDS User Info

```
"data": {
    "created_at": "2011-09-16T08:10:34.000Z",
    "description": "The UK\u2019s largest collection of UK and international social,
                economic and population data.
                Funded by @ESRC. Visit our blog: https://t.co/EBCD4NqGKW",
    "id": "374406950",
    "name": "UKDataService",
    "public_metrics": {
        "followers_count": 10137,
        "following_count": 3356,
        "listed_count": 378,
        "tweet_count": 10535
    },
    "url": "http://t.co/GUYNj9RsnE",
    "username": "UKDataService",
    "verified": false
  }
```

# A word about Trump

- Controversial Tweets
  - Free Speech
- Warning on Tweets
  - Opinion of Fact checker or Twitter
- Tweets Removed
  - No longer returned in searches
- Donald J Trump Banned!
  - He becomes a non-person to Twitter

# Non-Person on Twitter

- Not only can you not retrieve his Tweets
- His Id is removed from the Friends list of his 88M followers

# What data to keep

- In v2 of the API you make this decision in advance by what you ask for
- In v1.1 as you get everything, you may decide to get rid of the junk, but there will be key fields you want to keep

# What to keep  (Twitter_Fields.xlsx)

| Field | Essential | Likely |
|---|---|---|
| lang | | |
| text | Y | |
| author_id | Y | |
| id | Y | |
| created_at | Y | |
| conversation_id | Y | |
| possibly_sensitive | | |
| source | | |
| context_annotations | | |
| referenced_tweets | Y | |
| reply_settings | | |
| entities.mentions | | Y |
| entities.hashtags | | Y |
| public_metrics.retweet_count | | Y |
| public_metrics.reply_count | | Y |
| public_metrics.like_count | | Y |
| public_metrics.quote_count | | Y |
| entities.annotations | | |
| entities.urls | | Y |
| attachments.media_keys | | |
| in_reply_to_user_id | | Y |
| attachments.poll_ids | | |

# Timeline Reconstruction

1. Get the Target Timeline (UKDS)
2. Get List of Target Friends
3. Get Timelines from Friends
4. Combine 'Data' and 'Includes sections of Timeline Data
5. Combine (append) all of the Timelines
6. Sort by Date or ID

# 4. Combine 'Data' and 'Includes sections of Timeline Data

- If the Tweet was a Re-tweet, then you need to data in the 'Includes' section to get the full text of the Tweet.

- The same is true for a complete set of Entities ('hashtags, @manetions, URLs )

# What Next?

- The Timeline itself is just a Chronological sequence of events
- How you use it is upto you, but probably influenced by your research needs

# Conversations and Responses

- If you are going to do any kind of Text analysis, you will probably need to combine Tweets into conversations.

- Or even just re-construct multi-part Tweets

- API v2 helps here by providing a 'Conversation' end-point which will provide all of the Tweets with the same Conversation_id (You still have to assemble them yourself)

# Timeline Analysis

- What do you want to do?
- Have you kept the data to do it?
- Do you need additional software or Data?

# Timeline Analysis

Examples:

- Change of Tweet rate
- Trending (external) Hashtags
- Retweeting new/different sources

# Sentiment Analysis

- Tricky with limited text
- Some of which is hashtags, mentions and URLs
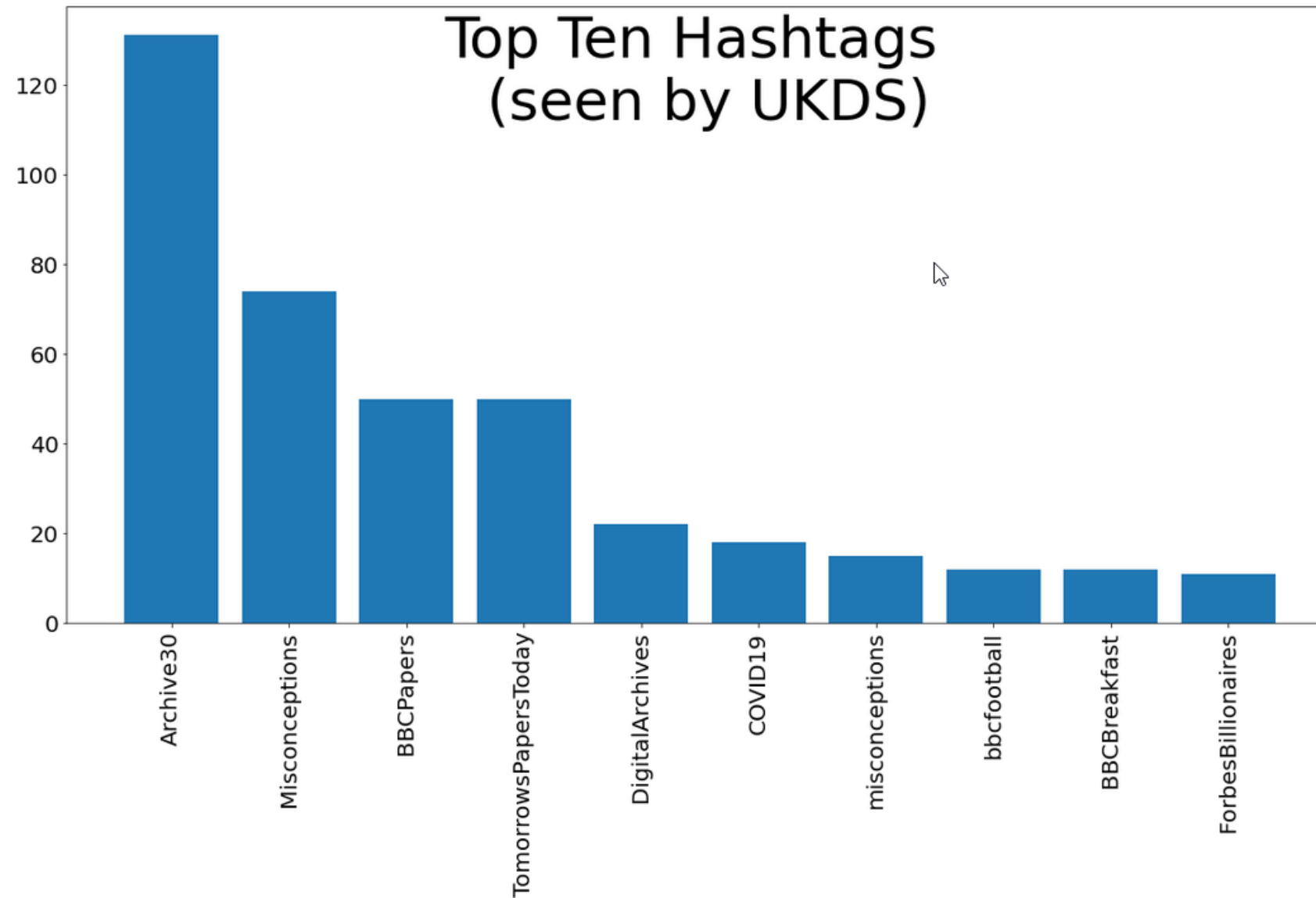- Emojis!
- Irony and sarcasm – No!

- (small demo)

# Basic Stats (and other things)

- How many Original Tweets v Retweets
- User info. – How old account, No. of Tweets, Friends and Followers
- User Description
- Tweet rate over time
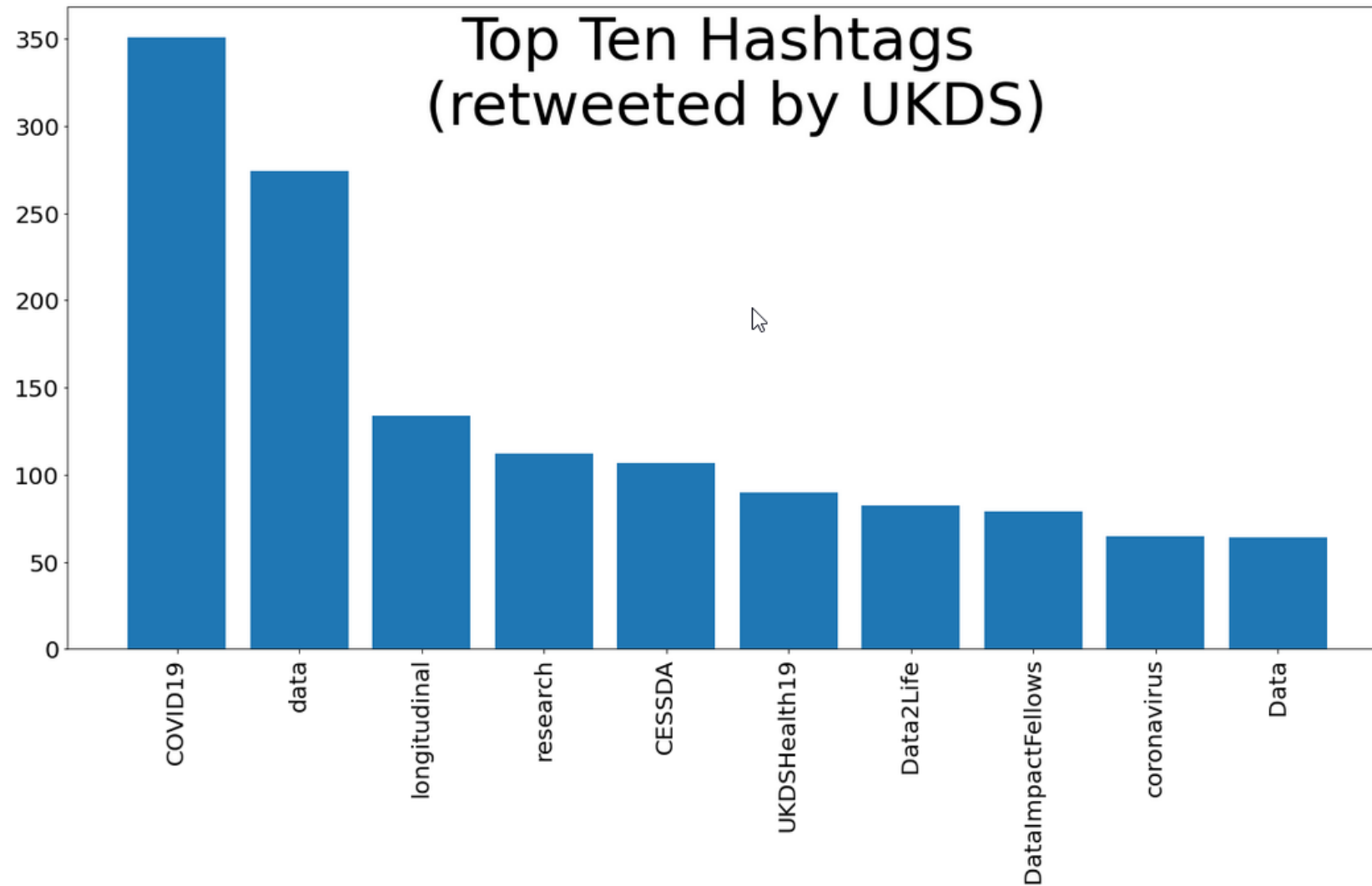- Relate @mentions and #hashtags to known trends or events
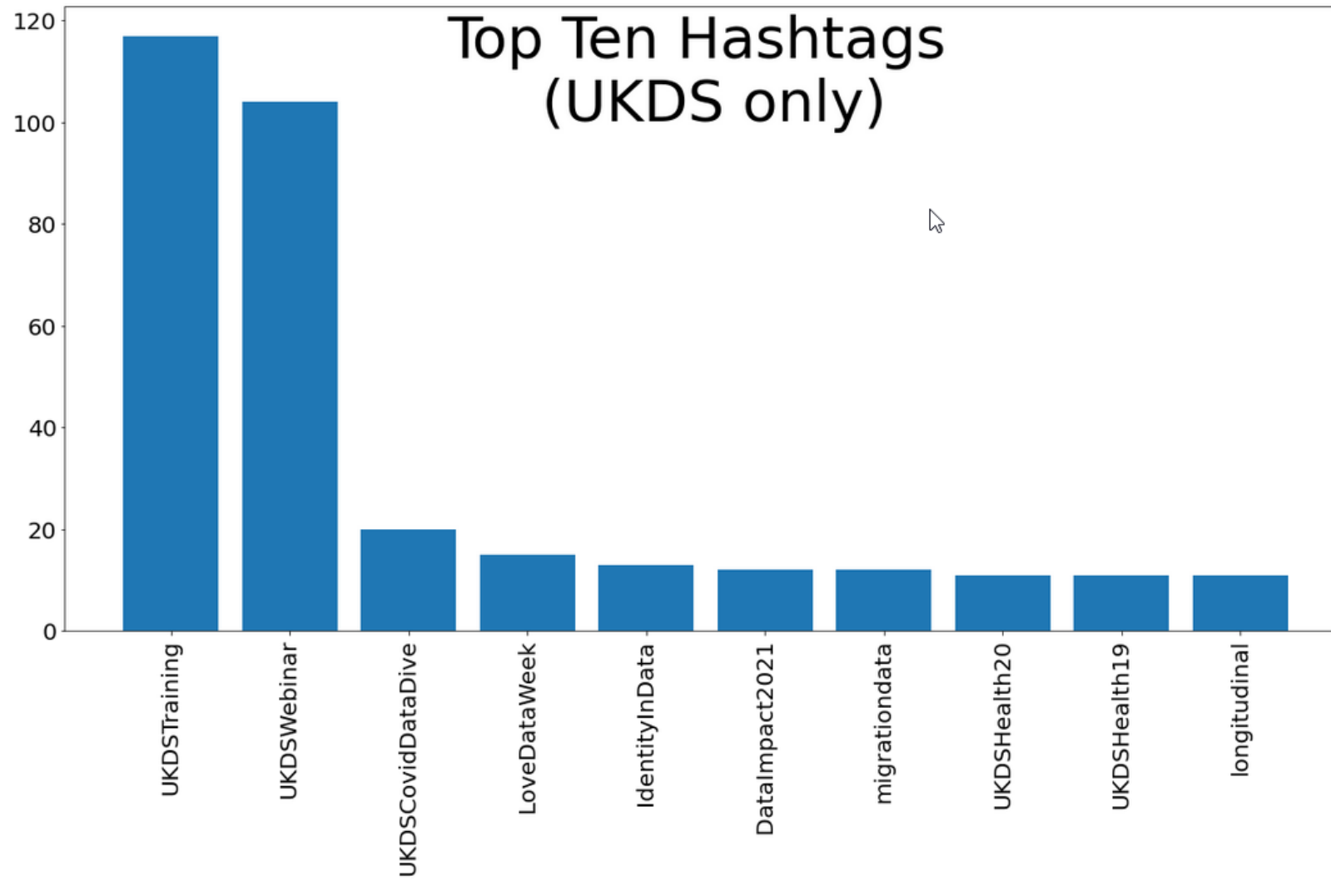
# Examples of Hashtag usage

# Examples of Hashtag usage



Top Ten Hashtags
(seen by UKDS)

# Examples of Hashtag usage

# Warning about shortened URLs

[{'start': 105,
'end': 128,
'url': 'https://t.co/KhiIOOPzVO',
'expanded_url':
'https://www.theyworkforyou.com/wrans/?id=2021-03-24.HL14586.h',
'display_url': 'theyworkforyou.com/wrans/?id=2021...'
}]

# Network Analysis

- Often Friends & Followers
- Could be accounts and #hashtags or @mentions
- Many additional s/w products support network graphs.
- Social Networks is a subject in its own right.

- (small demo)

# Twitter producers and consumers

- Friends and Family
- Hobbies and pastimes
- Commercial/Public services
- News Outlets
- Political Organisations

# Tweet Traffic

| Category | From You | | From Them | |
|---|---|---|---|---|
| | Original Tweets | Re-Tweets | Original Tweets | Re-Tweets |
| Friends and Family | Many | Few | Many | Few |
| Hobbies and pastimes | Many | None | Many | Few |
| Commercial/Public services | Few | None | Many | Many |
| News Outlets | Few | None | Many | Few |
| Political Organisations | Few | None | Many | Many |

# Summary

- Have a Plan
- Re-using a plan could be as simple as changing the account name
- The Timeline by itself is just data – you have to turn it into information
- There is much pre-built s/w to help with the techie bits.