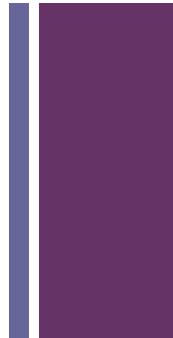


Applying Contextual Integrity to the Cambridge Analytica Case

Catherine Dwyer
Pace University
September 13, 2018



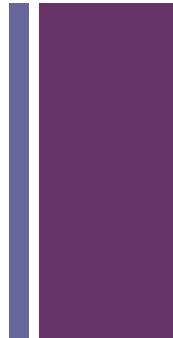
Outline



- Research question
- Application of Contextual Integrity (CI)
- Results to date
- Review of Findings
- Next steps



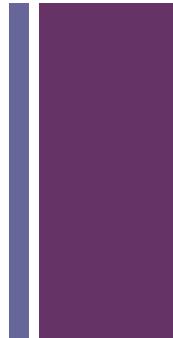
Research question



- Cambridge Analytica, a political data firm, used extensive set of Facebook data for predictive models that targeted political advertising during the 2016 US presidential election and the 'Brexit' vote in the UK
- The Cambridge Analytica case was widely perceived as violation of privacy norms for social media
- How can we use Contextual Integrity (CI) to identify in a granular, specific way, privacy norms for social media?



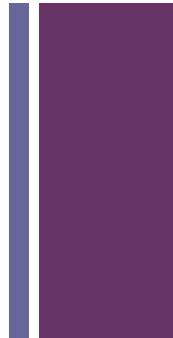
Application of CI



- CI examines whether information flow is appropriate as guided by legitimate, contextual information norms
- With CI we can examine news accounts of this case, in order to identify how privacy norms for social media
- CI describes four components of information norms – **actors, attributes, context, transmission principles**



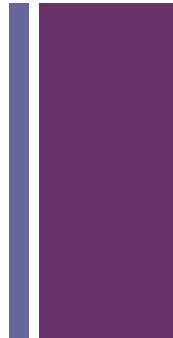
CI Elements



- Actors – senders and receivers of information
- Attributes – what information is being shared
- Contexts identified and how they are described
- Transmission principles – how did actions taken contradict what was expected for a particular context



Results to Date



- Methodology – collection of articles
- Quantitative analysis – descriptive statistics
- Qualitative analysis – identify potential norms



Methodology

- BYU Now Corpus (News on the Web), archives online news articles published in English <https://corpus.byu.edu/now/>
- Search for “Cambridge Analytica” 3/17/2018 through 5/02/2018
- URLs from hits were captured, then used to retrieve online articles
- Dataset: 2777 articles from 520 publications in 19 countries
- Descriptive statistics and NLTK Sentiment analysis calculated on collection



Mar 648 - cdwyer900@gmail.com × WQXR | New York's Classic 91.5 FM × NOW Corpus × Can you have your privacy and ×

Cathy

Secure | <https://corpus.byu.edu/now/>

NOW Corpus (News on the Web)

SEARCH FREQUENCY CONTEXT UNIVERSITY

FIND SAMPLE: [100](#) [200](#) [500](#) [1000](#)

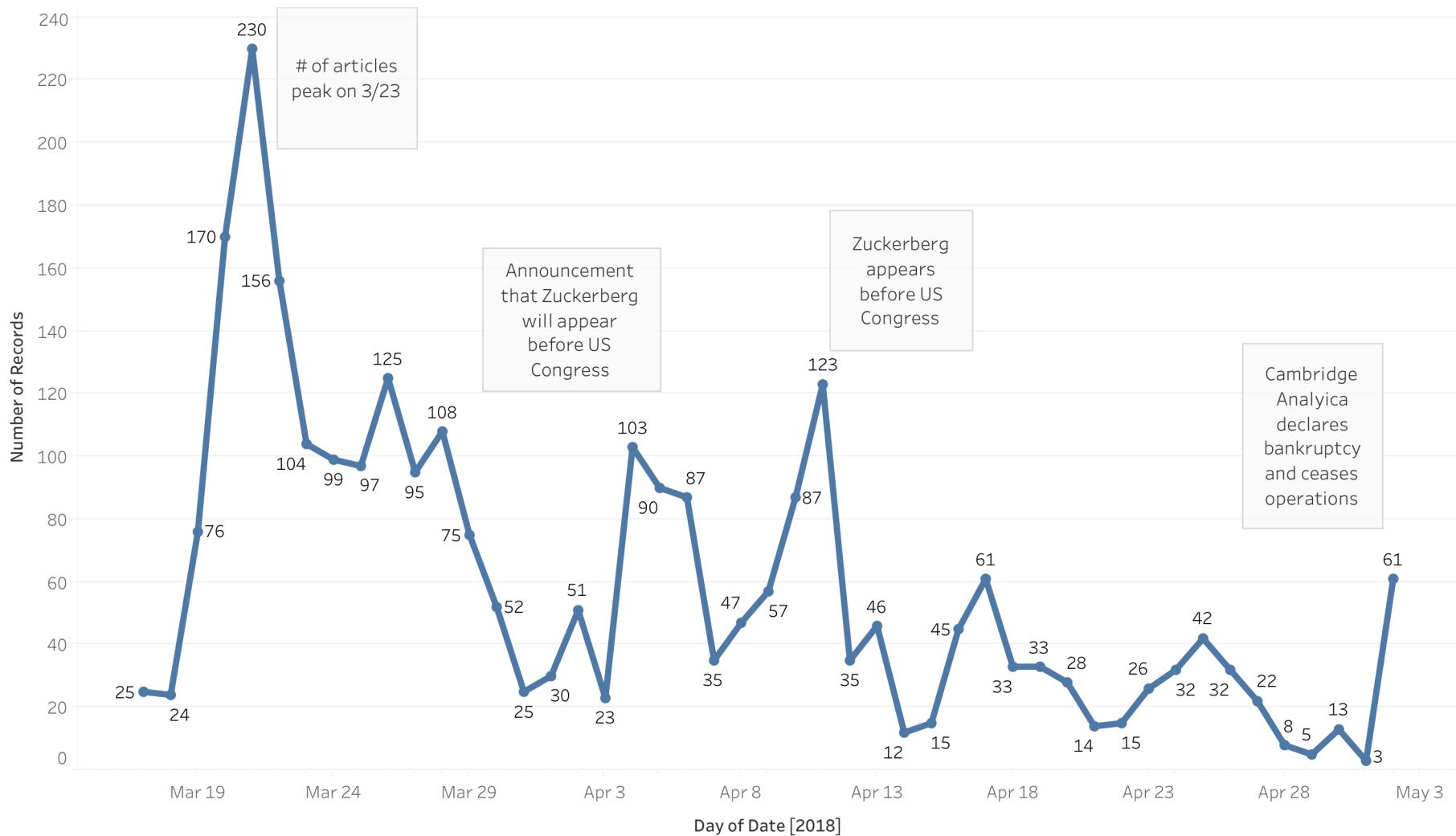
PAGE: << < 113 / 113 > >>

CLICK FOR MORE CONTEXT [?] SAVE LIST CHOOSE LIST CREATE NEW LIST [?] SHOW DUPLICATES

| ID | Date | Source | A | B | C | Text |
|-------|-------------|----------------|---|---|---|---|
| 11201 | 18-03-17 GB | Varsity Online | A | B | C | by revelations from a whistleblower who assisted Kogan at the data-analysis and influence firm Cambridge Analytica has alleged that the Cambridge ac |
| 11202 | 18-03-17 GB | Varsity Online | A | B | C | whistleblower, Christopher Wylie, has suggested that Kogan's tolls may have helped Cambridge Analytica influence the outcome of the 2016 US election |
| 11203 | 18-03-17 GB | Varsity Online | A | B | C | , grew uncomfortable with Kogan's connection to the firm SLC, which controls Cambridge Analytica, and distanced himself from Kogan. Das Magazin cla |
| 11204 | 18-03-17 GB | Varsity Online | A | B | C | data of tens of millions of Facebook users which was then passed on to Cambridge Analytica. According to Wylie, hundreds of thousands of Facebook u |
| 11205 | 18-03-17 GB | Varsity Online | A | B | C | racking up data on tens of millions of users, which was accessible to Cambridge Analytica. This contravened the social network's platform policy' which |
| 11206 | 18-03-17 GB | Daily Mail | A | B | C | had used the private details of millions of Facebook users without permission. # Cambridge Analytica is said to have exploited 50 million social media p |
| 11207 | 18-03-17 GB | Daily Mail | A | B | C | a' serious abuse' of the rules. # A whistleblower told how Cambridge Analytica -- owned by US hedge fund billionaire Robert Mercer, and headed at the |
| 11208 | 18-03-17 GB | Daily Mail | A | B | C | US voters, so they could be targeted with personalised political adverts. # Cambridge Analytica is accused of using 50 million Facebook profiles to build |
| 11209 | 18-03-17 GB | Daily Mail | A | B | C | their social network # Christopher Wylie, pictured, has blown the whistle on Cambridge Analytica and the extent to which they exploited Facebook to ha |
| 11210 | 18-03-17 GB | Daily Mail | A | B | C | people's profiles # Canadian data analytics expert Christopher Wylie, who worked with Cambridge Analytica, said: 'We exploited Facebook to harvest m |
| 11211 | 18-03-17 GB | Daily Mail | A | B | C | article # Share # The claims raise questions over the 2016 Brexit referendum as Cambridge Analytica, which claims it only pitched to work on the Leave |
| 11212 | 18-03-17 GB | Daily Mail | A | B | C | app called thisisyourdigitallife, built by academic Aleksandr Kogan. In collaboration with Cambridge Analytica, hundreds of thousands of users were pa |
| 11213 | 18-03-17 GB | Daily Mail | A | B | C | friends -- accumulating tens of millions of users' details. # Last night Cambridge Analytica said they' reject the claims made by Mr Wylie in what is paten |
| 11214 | 18-03-17 GB | Varsity Online | A | B | C | strategies to change voters' behaviour while working for the controversial data analysis firm Cambridge Analytica. # In a statement, released yesterday, |
| 11215 | 18-03-17 GB | Varsity Online | A | B | C | data to that the information had been destroyed, " he continued. # Cambridge Analytica, Kogan and Wylie all certified to us that they destroyed the data |
| 11216 | 18-03-17 GB | Varsity Online | A | B | C | said Kogan had produced social media algorithms which facilitated alleged efforts by data-analysis firm Cambridge Analytica to influence the outcome o |
| 11217 | 18-03-17 GB | Varsity Online | A | B | C | was elected. # In an interview with The Guardian/Observer, Wylie described how Cambridge Analytica courted the attention of Steve Bannon, the forme |
| 11218 | 18-03-17 GB | Varsity Online | A | B | C | which Bannon was taken while he visited the UK to give the impression that Cambridge Analytica, an independent company, had prestigious academic |
| 11219 | 18-03-17 GB | Varsity Online | A | B | C | suggested the company use the word " Cambridge " in its name. # Cambridge Analytica has rejected all the allegations that were put to it by The Guard |
| 11220 | 18-03-17 GB | Varsity Online | A | B | C | : " It is important to make clear that, despite its name, Cambridge Analytica has no connection or association with the University whatsoever. " # Facebo |



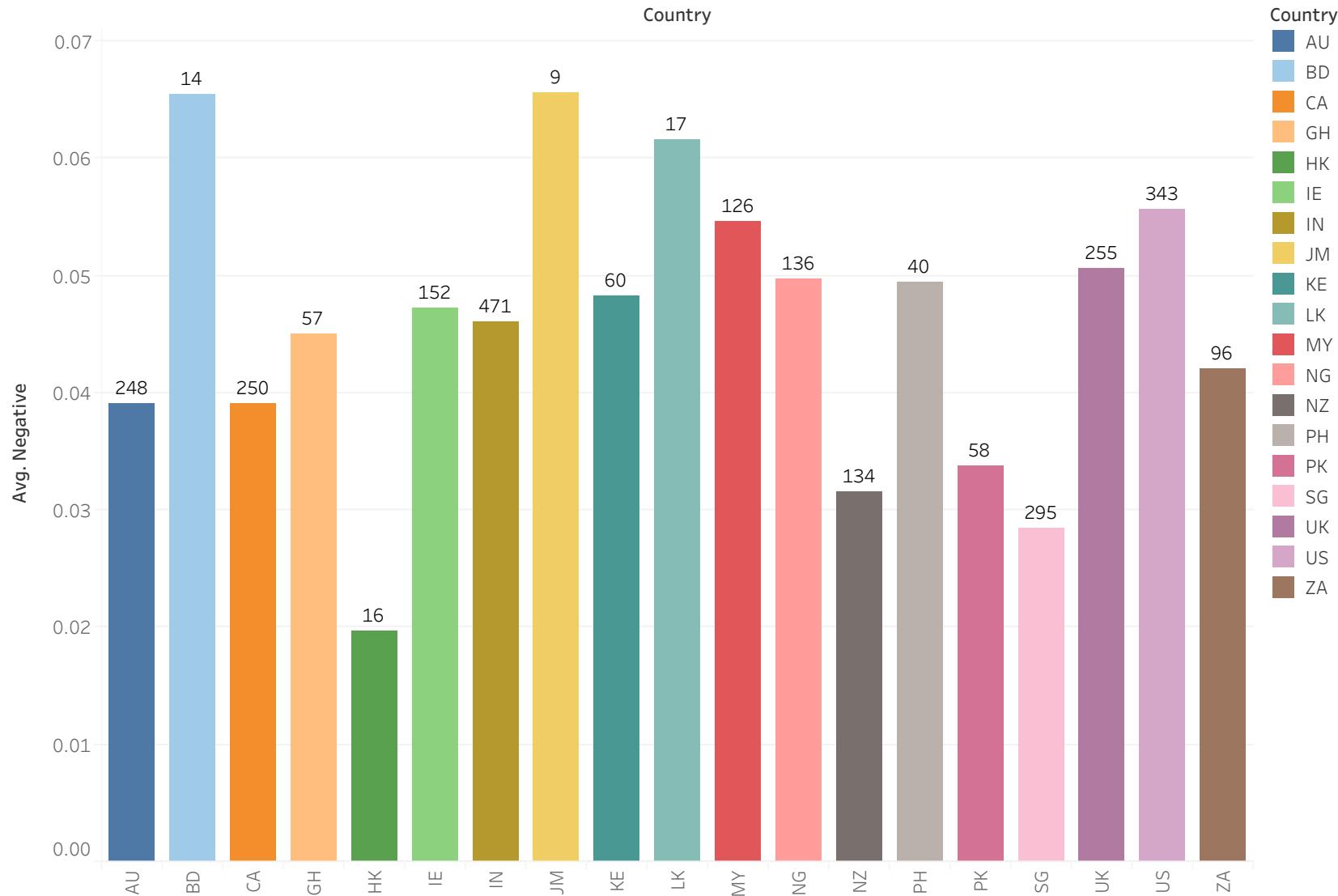
Publication timeline - 3/17 - 5/2/2018



The trend of sum of Number of Records for Date Day. The marks are labeled by sum of Number of Records. The data is filtered on Date, which ranges from 3/17/2018 to 5/2/2018.



Sentiment by country



Average of Negative for each Country. Color shows details about Country. The marks are labeled by sum of Number of Records.



NLTK – collocation tool (words commonly co-occur)

| Collocation results (1-10) | Collocation results (11-20) |
|----------------------------|-----------------------------|
| Cambridge Analytica | Robert Mercer |
| social media | chief executive |
| Donald Trump | Christopher Wylie |
| New York | Trump campaign |
| Alexander Nix | Aleksandr Kogan |
| Mark Zuckerberg | Steve Bannon |
| York Times | Facebook users |
| social network | Science Research |
| fake news | Global Science |
| Information Commissioner | user data |



Terms coded for CI

| Attributes | Contexts |
|----------------------|-------------------|
| user data | social media |
| personal information | social network |
| personal data | election campaign |
| Facebook data | consent decree |
| Facebook profiles | 2016 election |



Coding for CI (2)

| Actors | Transmission Principles |
|--------------------------|-------------------------|
| Cambridge Analytica | improperly obtained |
| Donald Trump | data breach |
| Alexander Nix | built models |
| Mark Zuckerberg | secretly recorded |
| Information Commissioner | gained access |



Close reading of language used to describe episode (NVivo)

- “50 million Facebook profiles harvested for Cambridge Analytica”
- “Permission from Facebook to harvest profiles in large quantities was specifically restricted to academic use”
- “built models to exploit what we knew about them and target their inner demons”



Application of CI Constructs to text

- “[Permission from Facebook] to harvest [profiles] in large quantities was specifically restricted to [academic use]”

Transmission principle

Context

Attribute



Review of Early Findings

- Lessons Learned: Scope (amount) of data collected was concern: “harvested”
- Power dynamic expressed in choice of language: “exploit”, “improper”
- Use of “model” and “modeling”
- Challenge – only so much text cleaning could be automated, did require editing ‘by hand’

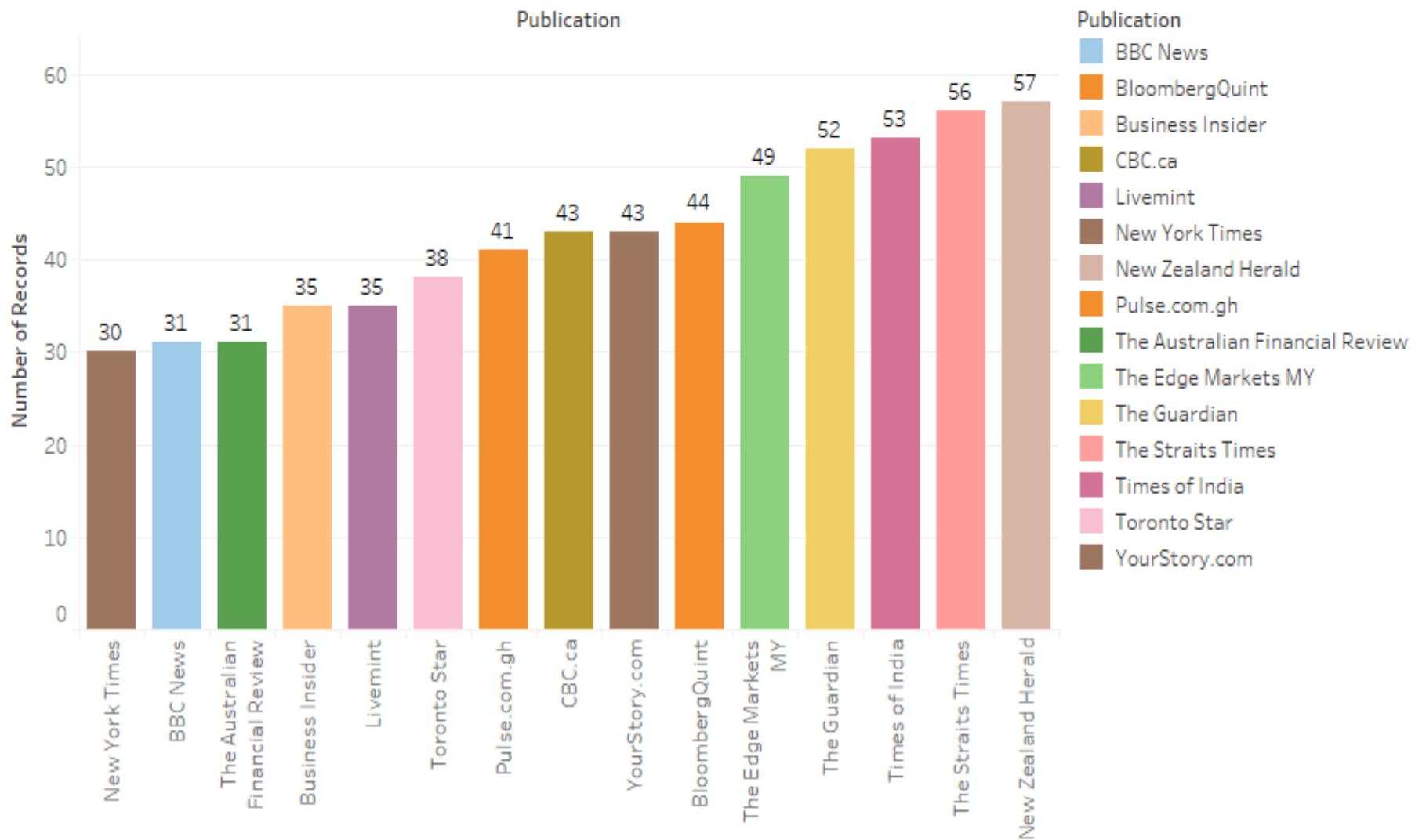


Next steps

- Continue mapping language in articles to elements of CI (actors, attributes, context, transmission principles)
- Compare norms across countries (e.g., India)
- Datasets and Python code will be shared via github
- Objective is to clearly define social media privacy norms, so they can inform efforts to strengthen online privacy with policy and legislation



Articles by publication





NLTK Concordance applied to collocation results

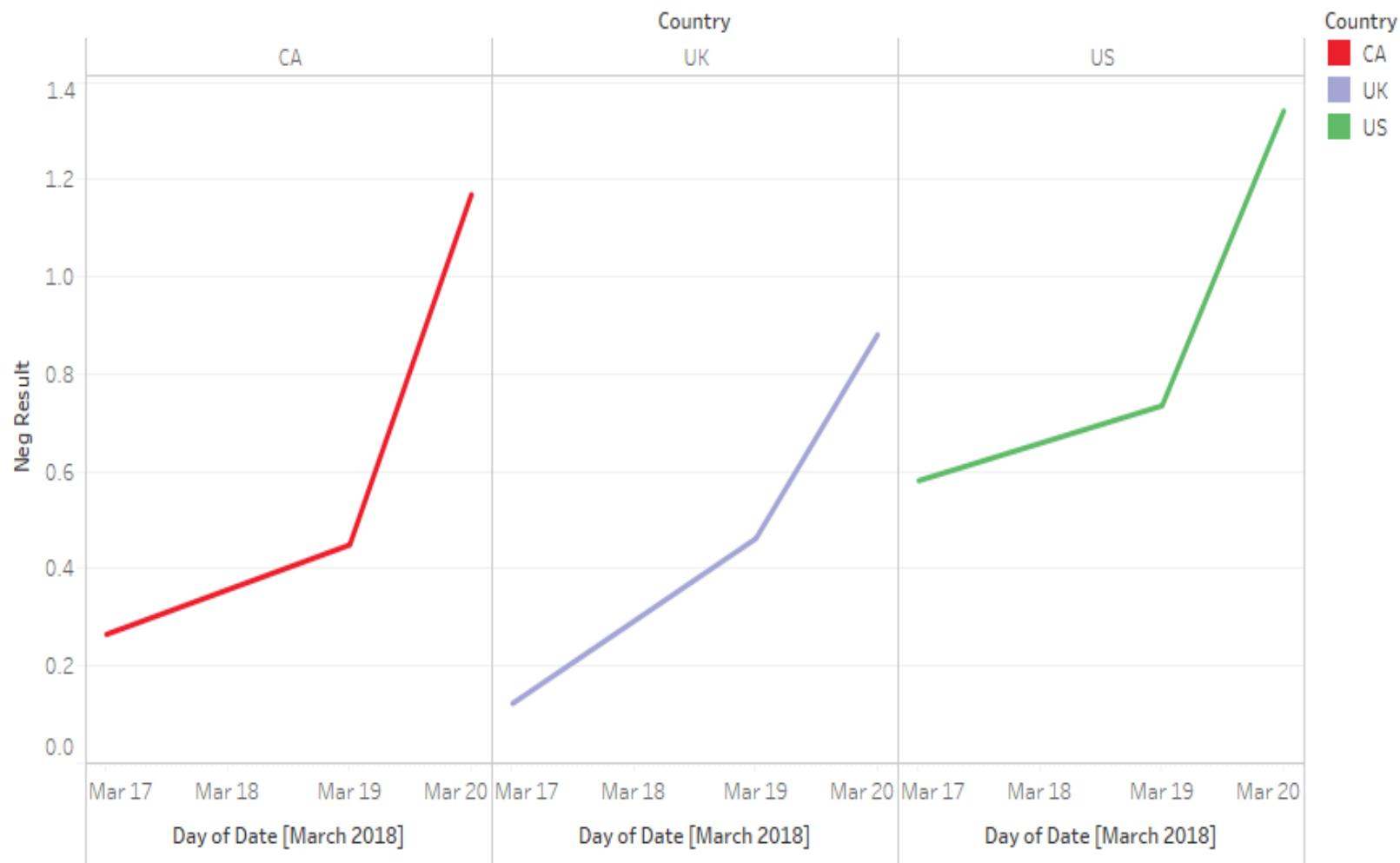
- NLTK concordance – shows every occurrence of specified word, along with context
- Cambridge Analytica had a database of millions of US voters that had its own algorithm to scan them, identifying likely political persuasions and **personality_traits**. They could then decide who to target ...
- People have been up in arms that our very democratic agency was being stolen from us by mysteriously dark forces aware of every aspect of our **psychological_profiles**.

Data Obtained From BYU Corpus

| 1 | | Date | Country | Publication | URL |
|----|-----|----------|---------|----------------------------|---|
| 2 | w1 | 18-05-02 | US | The Guardian | https://www.theguardian.com/uk-news/2018/may/02/cambridge-analytica-shutting-down |
| 3 | w2 | 18-05-02 | US | Slate Magazine | https://slate.com/technology/2018/05/fleeing-client-new-york-city.html |
| 4 | w3 | 18-05-02 | US | New York Post | https://nypost.com/2018/05/02/cambridge-analytic-shutting-down/ |
| 5 | w4 | 18-05-02 | US | The Guardian | https://www.theguardian.com/technology/2018/may/02/cambridge-analytica-shutting-down |
| 6 | w5 | 18-05-02 | US | Axios | https://wwwaxios.com/cambridge-analytica-shutting-down |
| 7 | w6 | 18-05-02 | US | KGO-TV | http://abc7news.com/technology/cambridge-analytic-shutting-down |
| 8 | w7 | 18-05-02 | US | PBS NewsHour | https://www.pbs.org/newshour/science/miles-to-go-with-cambridge-analytic-shutting-down |
| 9 | w8 | 18-05-02 | US | New York Times | https://www.nytimes.com/2018/05/02/us/politics/cambridge-analytic-shutting-down.html |
| 10 | w9 | 18-05-02 | CA | Toronto Star | https://www.thestar.com/opinion/star-columnists/2018/05/02/cambridge-analytic-shutting-down |
| 11 | w10 | 18-05-02 | CA | Toronto Star | https://www.thestar.com/news/canada/2018/05/02/cambridge-analytic-shutting-down |
| 12 | w11 | 18-05-02 | CA | iPolitics.ca | https://ipolitics.ca/2018/05/02/evening-brief-a-sad-day-for-canadian-politics |
| 13 | w12 | 18-05-02 | CA | IT World Canada | https://www.itworldcanada.com/article/hashtag-trump-cambridge-analytic-shutting-down |
| 14 | w13 | 18-05-02 | GB | Government Europa | https://www.governmenteuropa.eu/european-data-shielding |
| 15 | w14 | 18-05-02 | GB | BBC News | http://www.bbc.co.uk/news/business-43983958 |
| 16 | w15 | 18-05-02 | GB | Isle of Wight County Press | http://www.iwcp.co.uk/news/national/16200661.Digital-marketing-firms-shut-down |
| 17 | w16 | 18-05-02 | GB | Channel 4 News | https://www.channel4.com/news/cambridge-analytic-shutting-down |
| 18 | w17 | 18-05-02 | GB | TechRadar | https://www.techradar.com/news/cambridge-analytic-shutting-down |

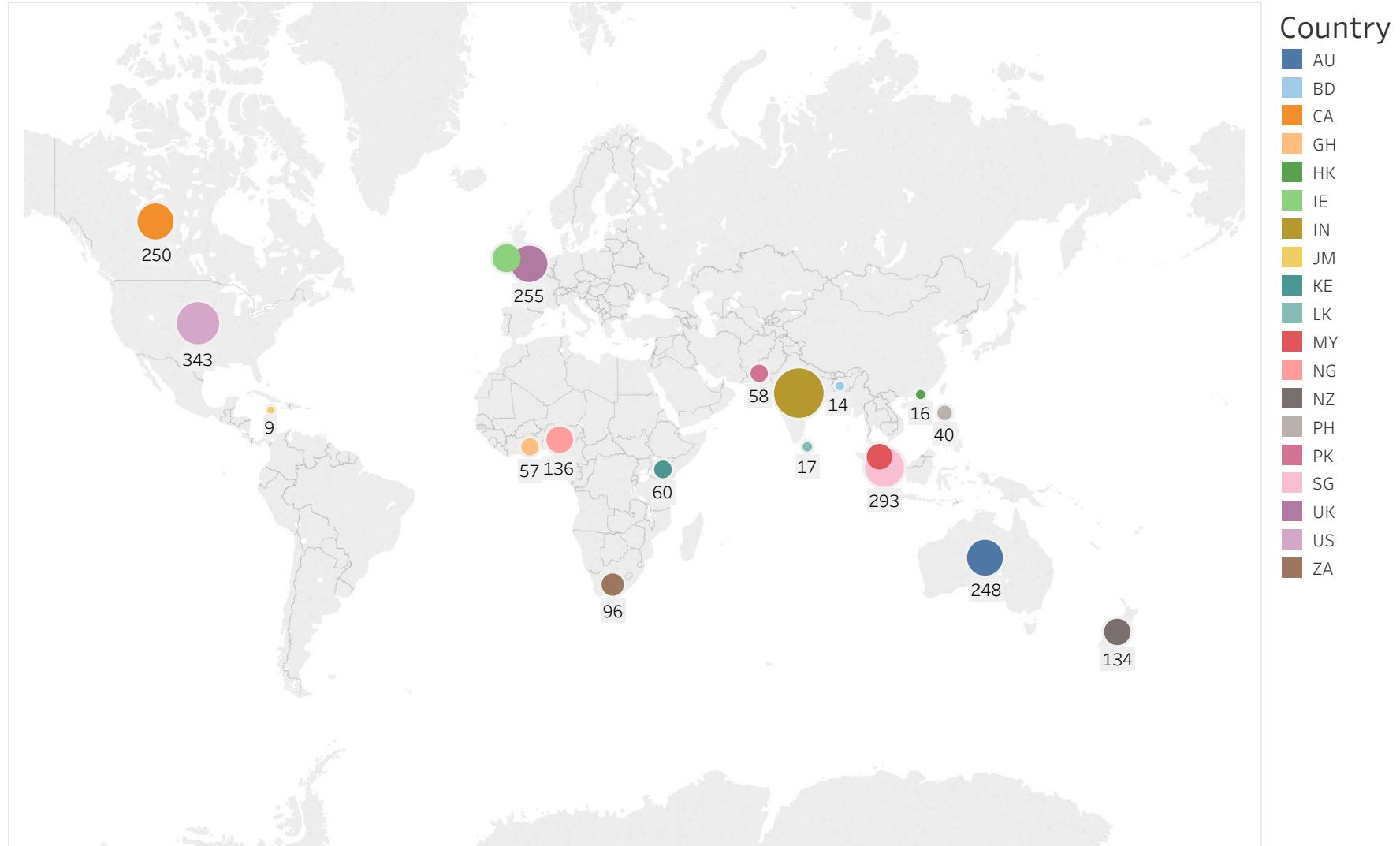


Sentiment analysis - articles 3/17, 3/19, 3/20





Articles by country - 3/17 - 5/2/18



Map based on Longitude (generated) and Latitude (generated). Color shows details about Country. Size shows sum of Number of Records. The data is filtered on Date, which ranges from 3/17/2018 to 5/2/2018.