

What does Twitter have to say about Brexit?

A Reflection on the Collaboration Process

Allie Turner

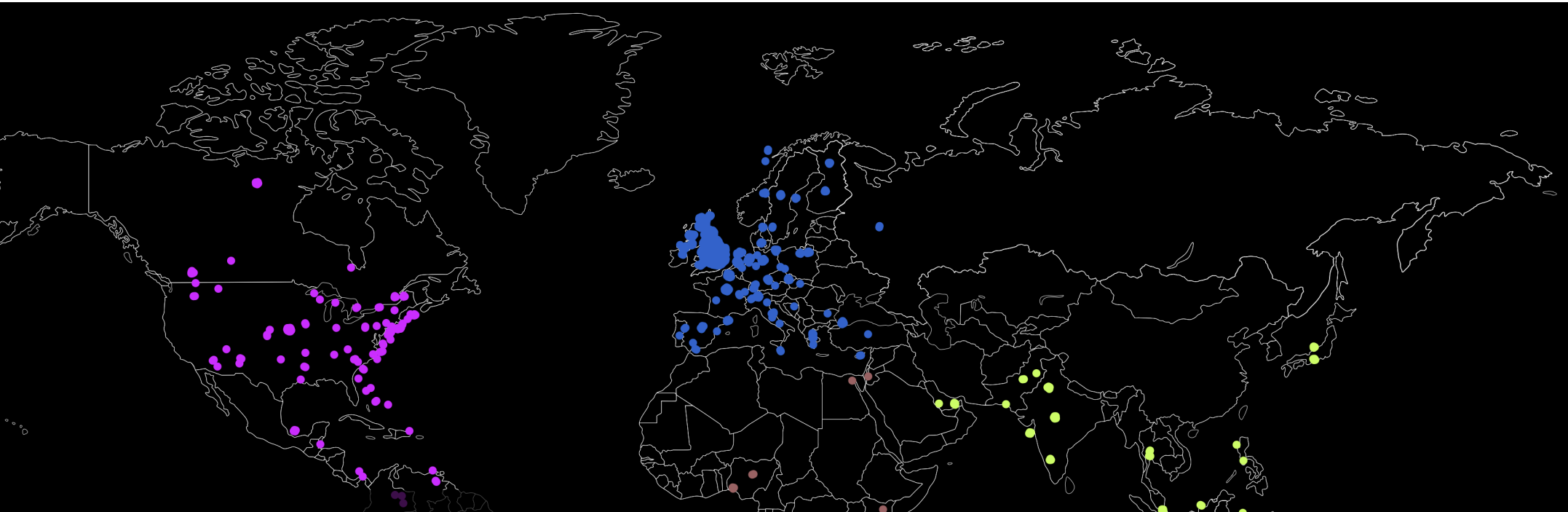
Finn Ickler

Luis de Sousa

Data Science for Design

December 2017

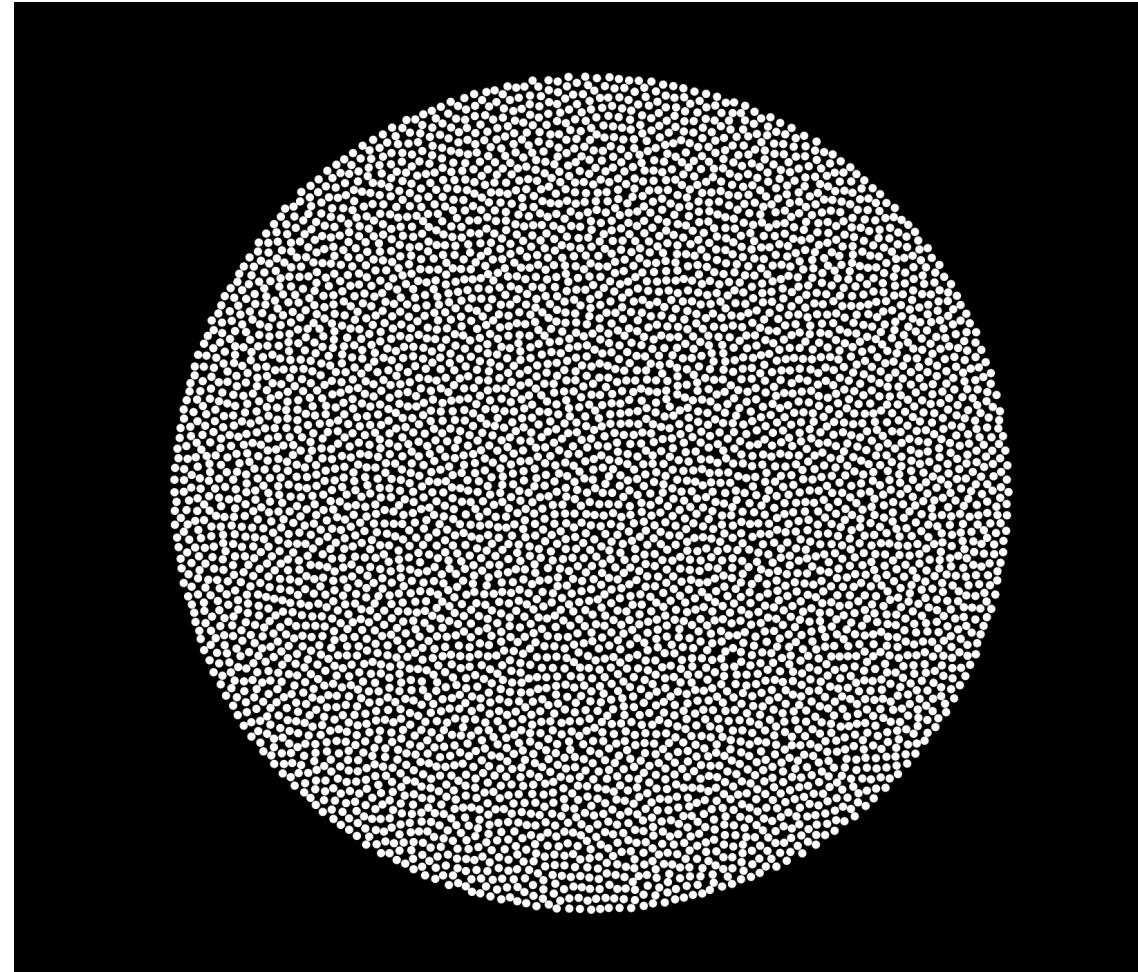
Our visualisation can be found at: <https://www.ffgti.org/Brexit>



Contents

Our data set came from the Neuropolitics Research Lab at the University of Edinburgh, a transdisciplinary research centre that utilises developments in cognitive neurosciences to shed light on new political attitudes, identities and behaviours. Their aim with providing this large data set was to provide new insights into how citizens' Brexit-related expectations are shaped in a digital world. How have their expectations about Brexit changed? Have their expectations changed from when they voted? What are the perceptions of digital authenticity and trust? We were asked to analyze what people we're saying, who was talking & what messages persisted.

Context	3
Development	4
Findings	6
Conclusion	11



Context

The tabular data provided to us was in json files, with categorical dictionaries within. With 32 different columns and large amounts of rows (different depending on the day and hour), there was a lot of data to analyze. Each json file was tweets that were tweeted for every hour of that day that included Brexit-related hashtags. Coming from five different days, each gathered Brexit-related Tweets from the following days: Referendum (23 June 2016) Pre-Referendum (21-22 June 2016) and Post- Referendum (22-23 September 2017, Florence Speech). We kept the same structure for our visualisation, believing it was important to show the relationships between the three timeframes and how the data changed between them.

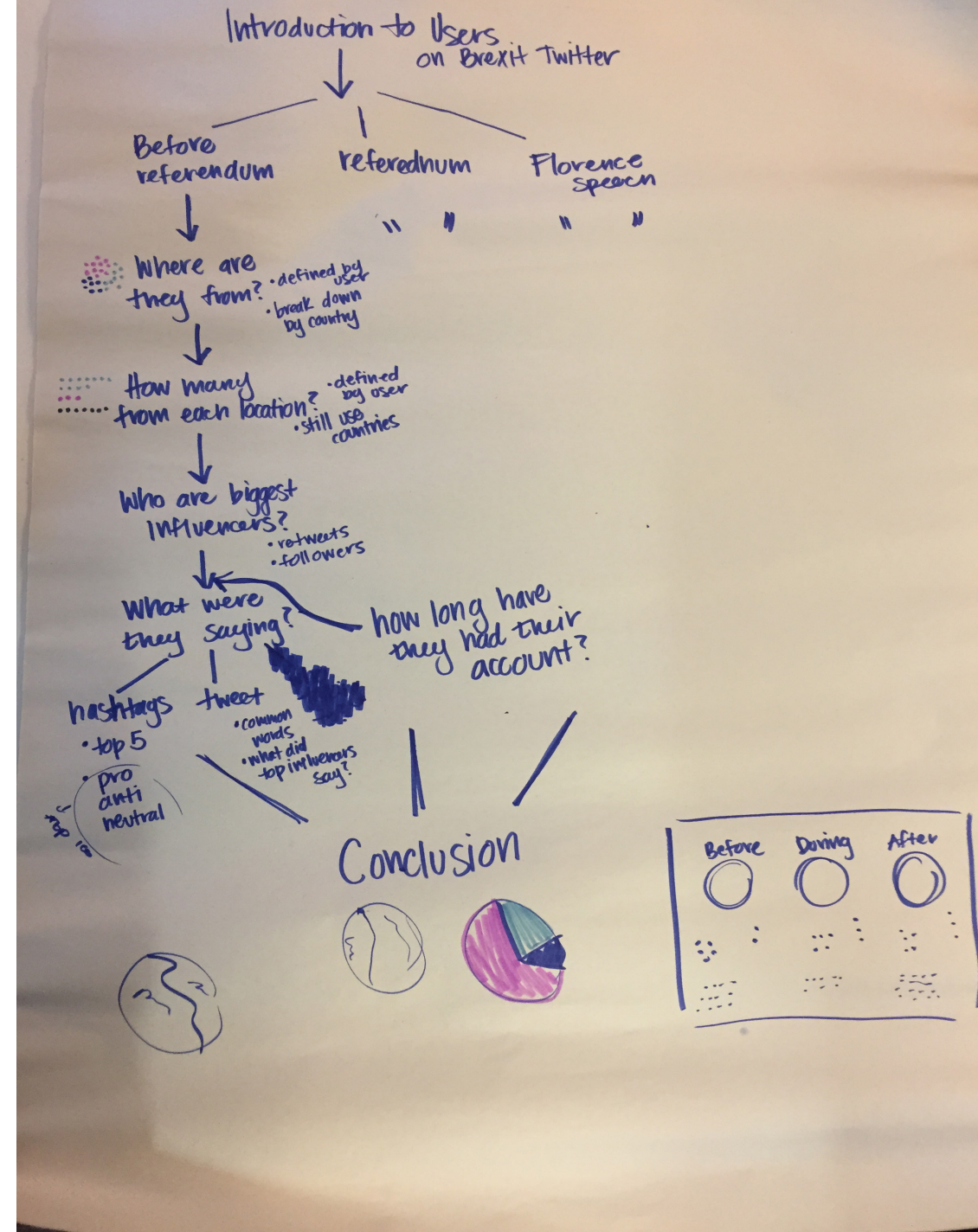
Before beginning the storyboard of our visualisation, we were each tasked with analyzing the data on our own, reflecting to see where we could bring our ideas together to form one cohesive story. Allie looked at the account ages of the users, comparing how long users had their accounts for when they tweeted, and hypothesized that a large subset of “Silicon Valley” users were bots that created their accounts within the hour that they tweeted. Finn analyzed the hashtags used and if they were pro-, anti- or neutral-brexit. He also looked at the location users defined in their profile and ran them through Google’s geolocation. Luis looked at the biggest influencers on Twitter, those who received the most retweets.



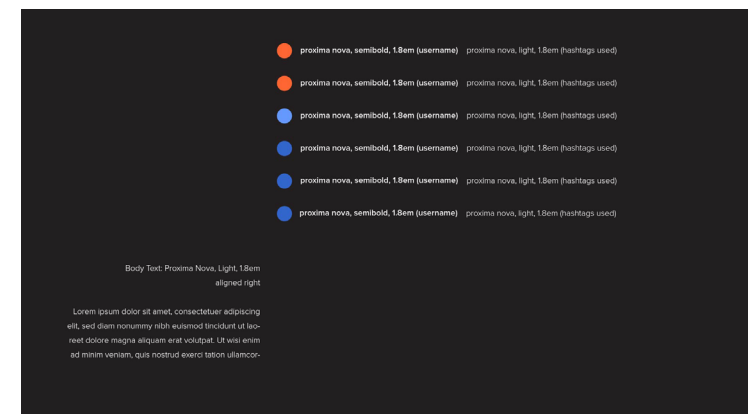
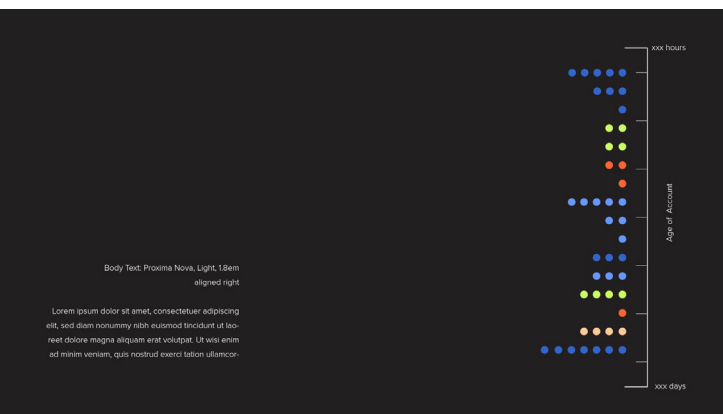
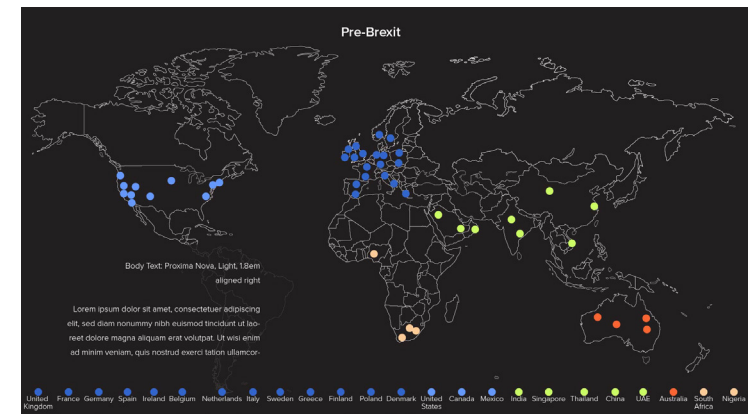
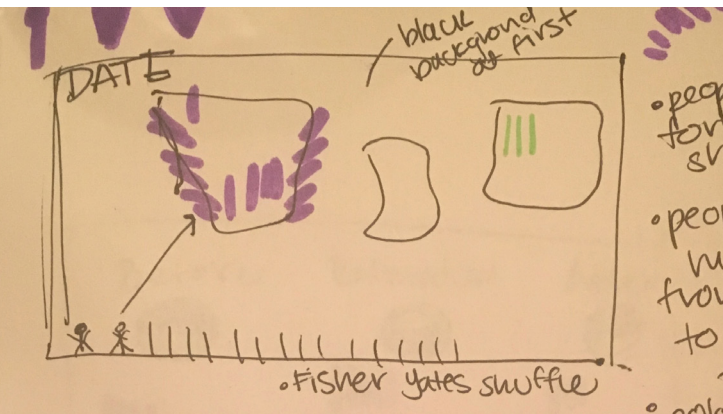
Development

With different hypotheses from all three of us, we mapped out one comprehensive storyboard that tells the story of what we found most important for the three separate timeframes. We were inspired by the Paradise Papers interactive site by SVT Nyheter, a Swedish public television broadcaster that released a visualisation about the 2,000 Swedes leaked within the documents (<https://www.svt.se/special/the-swedes-in-paradise-papers/>). Using a lot of data, similar to us, their visualisation was still clear and portrayed an understandable story, even to people with no background in the subject.

As Finn is proficient in JavaScript, Allie and Luis worked with him to define a style guide that he could convert into comparable d3.js code. Starting with an introduction to the project, the visualisation runs through countries represented by user profiles, account ages measured in months, the biggest influencers during that day, and what each influencer had to say with hashtags. This runs through for each timeframe (Referendum, Pre-Referendum and Post-Referendum) and ends with a summary of what the biggest influencers were saying on Twitter and whether it was pro-, anti-, or neutral-brexiteer.



Initial plan of a possibly story.



	screen_name	location	position	bogus	age
31	WishResonator	London, England		FALSE	223968535
63	JenDreYT	London, England	(51.5073219, -0.1276474)	FALSE	1326340
72	jacob_evering	London, England		FALSE	8034637
76	disclaimermag	London, England	(51.5073219, -0.1276474)	FALSE	71150945
79	james_ernest	London, England		FALSE	153396453
93	steventrent	London, England		FALSE	127326295
103	imscott	London, England	(51.5073219, -0.1276474)	FALSE	141394037
135	genevieve_joy	London, England	(51.5073219, -0.1276474)	FALSE	187784616
158	rclarke	London, England	(51.5073219, -0.1276474)	FALSE	288927193
159	MissJoHartley	London, England		FALSE	151588576
172	susioneill	London, England		FALSE	282203612
173	DanBloodyJohns	London, England	(51.5073219, -0.1276474)	FALSE	147327554
175	Jazzy_tizzle	London, England	(51.5073219, -0.1276474)	FALSE	167815741
218	popbitch	London, England	(51.5073219, -0.1276474)	FALSE	292694091
225	NARSTI	London, England		FALSE	185079172
235	_Eyniiiiii	London, England	(51.5073219, -0.1276474)	FALSE	160332006
236	bultifish	London, England	(51.5073219, -0.1276474)	FALSE	167662892
250	pad6	London, England		FALSE	109562565
269	abcdaisy_	London, England		FALSE	225008040
281	thecarolemalone	London, England		FALSE	80491519
301	FubarRadio	London, England	(51.5073219, -0.1276474)	FALSE	165351060
306	amzeratul	London, England		FALSE	227244242
312	jamesliamCook	London, England		FALSE	172770322

Several technical challenges arose while programming, both in the pre-visualisation data wrangling and in the processing of the data with JavaScript. During the pre-processing, we quickly noticed that the amount of raw data did not translate well to a straight forward force simulation like in the work done by the Swedish Newspaper. Even our high powered laptops where not able to fluently simulate over ten thousand nodes. Brexit day alone had 500,000 users we wanted to visualise. To fix this problem, we decided to let one circle in our simulation represent 100 users from our data.

Keeping the distribution was achieved by grouping the users into bins and then keeping the bins with more than 100 users. This was possible as we only needed the account age accurate to the month, and didn't have to handle the continuous time data. User data is compressed into only the relevant rows, 'Location', 'Position', 'Screenname' and 'Account Age' and saved as a csv file. The biggest problem during the development of the website was hot-swapping the different data sets to achieve the three different visualisations and getting the location data to appear correctly on a svg map. We decided on using a mercator-projection, but not showing the poles. This alteration caused many complications and we had to adapt our own conversion formula. In the end, we decided to set the resolution to a fixed 1219 x 700 pixels as this accommodates most laptops and does not have a negative impact on the viewing experience. We also abandoned the scrolling idea, as it is only suitable for visualisation of data where every possible state of the force simulation is reachable by every other state. As we are swapping data in and out, which takes a considerable time to load, skipping over triggers by scrolling too fast would break the visualisation.

Findings

The biggest Twitter influencers in this analysis were calculated by the amount of retweets their posts were generating. When this approach was initially used in relation to a single hour of the data set (Assignment 2), the biggest influencers were pro-Brexit. The top ones also recently created accounts, thus hinting that the Leave campaign on Twitter was stronger than the Remain counterpart.

After running the same analysis on all the hours before and during the referendum, we could not only validate the previous assumption, but most importantly (after associating respective hashtags and looking at the content of the retweets) we uncovered that many influential accounts were sharing messages of extreme positions, thus polarising people. One of these accounts even reached the top ten on the referendum date.

@Vote Leave #VoteLeave
@Stronger In #Remain
@TheOrdinaryMan #Dictatorship
@graham norton #remain
@Andrea Leadsom MP #IndependenceDay
@Mis Standen #EUisTheProblem
@Bloomberg #Brexit
@Bonnie Greer #Remain
@David Peterson #VoteLeave
@RT #Brexit (Leave)

@Vote Leave #VoteLeave
@LEAVE.EU #VoteLeave
@Emma Watson #Remain
@Russ #Remain
@Louise Mensch #VoteLeave
@BBC News (UK) #EURef
@J.K. Rowling #Remain
@Richard Branson #VoteRemain
@Vote Leave Media #VoteLeave
@TheOrdinaryMan #Dictatorship

@Nick Reeves - 48% #StopBrexit
@#Brexit Bin #StopBrexit
@Craig Whittington #Brexit (Leave)
@Richard Corbett #LeaveLies
@UKIP Nonsense ✳️ #StopBrexit
@James Melville #FlorenceSpeech (Remain)
@Ebrahim Hemmatnia #Iran
@Jeremy Cliffe #FlorenceSpeech
@Mireia #Catalunya+ #Referendum 7
@Graham Simpson #StopBrexit

Conclusion

As a group, our intention was to create a visualisation that answered the questions posed by the data holders, but also to make it perceptible to a wider audience. In that sense, we were happy with the result obtained.

After combining our individual explorations from Assignment 2, we could verify that the majority of our assumptions were effectively confirmed. Regardless of the origin of the accounts, we can now confirm that the majority of Twitter was indeed pro-Brexit. Such findings demonstrate the influence of social media on the general public, and how its importance cannot be disregarded as a valid representation of public perceptions. The biggest challenge of our investigation was that we could not gather an accurate location for all the users, and the ones that had it were self-determined. Should we have had such information, we would have been able to determine which accounts were bots created to undermine people's views, and their overall impact. Without such information, we could only hypothesize about it.

In regards to this project, our next steps would be to fix the aforementioned problems during the development stage, but also to compare the activity of the biggest influencers accounts', and how (in)active they are a year later.

Finally, we would like to liaise with the data holders, present them our findings, and hopefully generate new insights with their previous knowledge.

