

Cyberbullying and Hate Speech

What can social data tell us about cyberbullying
and hate speech online?

Contents

Abstract	3
Preface	4
Background and Methodology	5
1.0 Hate Speech Analysis	6
Progress and Challenges	8
Racial Intolerance.....	10
Homophobia	15
Transphobia	21
Correllation	28
2.0 Online Bullying	32
Progress and Challenges	33
Demographic Analysis	34
Recipient Responses	35
Timing.....	37
Appendix	45
About Brandwatch	46
About Ditch The Label	47

Abstract

This study of almost 19 million tweets analyses hate speech, masculinity and online bullying in the UK and the US.

Racial intolerance was the most common source of hate speech, with US tweets concentrated in southern states. In both markets, regions with racial hatred strongly correlated with misogyny. Socioeconomic factors may also play a role, with poorer areas hosting higher levels of hate speech.

Gender constructs remain prominent within family units and the data points to educators as playing a pivotal role in encouraging diversity. Significant progress has been made raising awareness around homophobia and transphobia, but hate speech persists within specific interest groups.

Online bullying was found to span a wide range of topics and responding to bullying often escalated the conflict.

For each of the topics, key areas of progress are outlined as well as challenges for those hoping to tackle these areas of prejudice might face.

Preface

This project represents the combined efforts of the Brandwatch Research Services team in conjunction with Ditch the Label. The contents, which can be divided loosely into two sections, provide supportive data and practical advice for campaigns hoping to challenge bullying and discrimination in online discourse.

Hate speech, the first of our three topics, is itself a contested term¹. For the purposes of this study, we define hate speech as expression or incitement of hatred towards recipients based on prejudice; for this study based on race, gender identity and expression, or sexual orientation. Further areas of discrimination, such as religion, go beyond the current scope but warrant further research. Analysis in this project assumes the aims of minimising hate speech and increasing awareness and debate.

Our research shows some hate speech to be concentrated in select geographic regions and demographics. The purpose of the project is not to vilify these groups (generalisations of this kind go against the core aims of the study), but rather to show the biggest opportunities for social progress. Reported rates of cyber bullying vary, with as high as 87% of youth being exposed according to a McAfee 2014 study² and as many as 34% being direct recipients³. In the second section of our study we analyse online bullying, questioning when bullying takes place, by which types of authors, and whether specific topics precede bullying remarks. The report recognizes bullying as a behaviour rather than identity and avoids 'bully' and 'victim' as static labels.

Bullying language is heavily context dependent. Here, we have separated insults used as friendly or teasing exchanges from those intended to offend. Much of this must be inferred at perlocutionary level, however features in the text (including those denoting humour in the recipient) are also used in segmentation.

This project sheds light on issues relating to hate speech, but should not be viewed as an argument for online censorship. Rather, the data points to the need for a nuanced approach, further open debate and awareness, and positive role models. While there are many signs of positive progress throughout, there are also key challenges to address for those hoping to facilitate social change.

1. Gagliardone, I., et al, UNESCO 2015. Countering Online Hate Speech, United Nations Educational, Scientific and Cultural Organization. Open access: <http://unesdoc.unesco.org/images/0023/002332/233231e.pdf>

2. Eichorn, K. et al, 2014. "2014 Teens and the Screen Study", <http://www.mcafee.com/us/about/news/2014/q2/20140603-01.aspx>

3. Patchin, J.W., Jinduja, S., 2015. "Lifetime Cyberbullying Victimization Rates", <http://cyberbullying.org/summary-of-our-cyberbullying-research>

Background and Methodology

Background

Award-winning anti-bullying charity Ditch the Label and leading social intelligence company Brandwatch teamed up to tackle questions surrounding hate speech and online bullying. The project captured a total of 19 million tweets from the US and the UK over the span of four years, to better understand progress and challenges across the following key areas:

1. Online Hate Speech

Five queries were written to capture hate speech across five topics: racial intolerance, misogyny, masculinity construct, homophobia and transphobia.

For each topic a separate query searched for neutral and supportive discussion about the topic.

The data was then analysed for key trends, author demographics and regional variation within the UK and the US.

2. Online Bullying

Bullying exchanges were collected from Twitter, not limited to specific hate speech areas of section one.

These were used to establish whether specific times and topics were more prone to online trolling, and whether responding to online bullying more often helps or hinders the recipient.

1.0

Hate Speech Analysis

—

Hate Speech Summary

Hate speech summary						
	Insult Prevalence	Gender Breakdown	Neutral/ Supportive Discussion	Most prominent US States	Most prominent UK Counties	Correlations
Racial Intolerance	7.7m tweets	59% male	Overtook racial insults in January 2015	Louisiana, Texas	Midlothian, East Ayrshire	Strong positive correlation with misogyny, moderate positive correlation with homophobia
Homophobia	390k tweets	64% male	More common than homophobic insults and growing	Rhode Island, Iowa	Denbighshire, Dundee	Moderate to strong positive correlation
Transphobia	456k tweets	60% male	More common than transphobic insults and growing	Alabama, Idaho	West Yorkshire, Coventry	Strong positive correlation with Republican support (US), moderate negative correlation with masculinity (UK)

Hate Speech Summary

Progress

Online support is growing: discussion about homophobia and transphobia consistently outnumbers hate speech in these areas. Discrimination as a talking point has grown by 14% (homophobia) and 38% (transphobia) each month on average over the past four years.

The Black Lives Matter movement has increased the profile of racial discrimination as a talking point online and gained traction on both sides of the Atlantic. While male authors are more likely to discuss the problems of racism, female authors now comprise a larger share in this conversation than in 2012.

Misogyny has grown significantly as a talking point since 2014 and engages both male and female authors online. Misogynistic comments are now also less likely to come from students (comprising 31% in 2012 compared with 20% in 2016).

Challenges

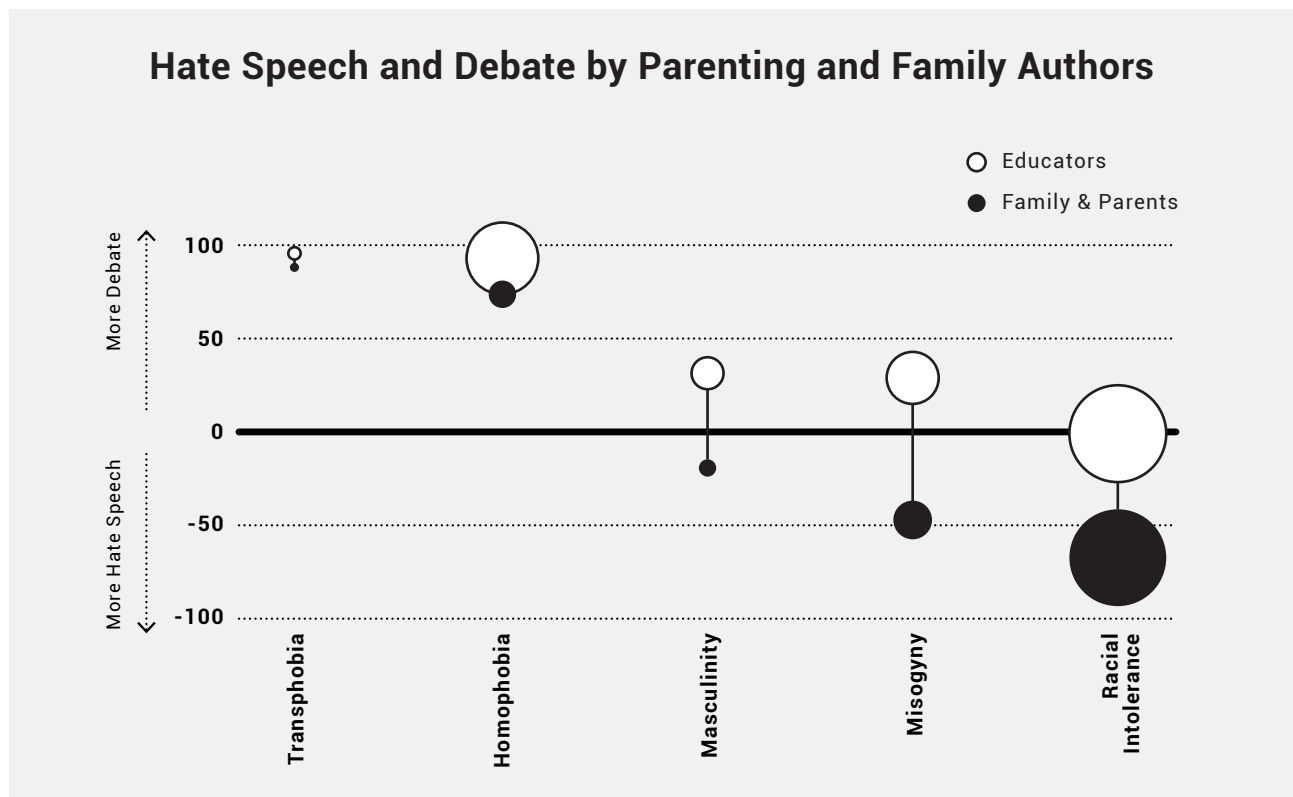
Masculinity constructs are a growing talking point, especially since the beginning of 2016. However, masculinity-related insults remain prevalent. This is especially the case among authors associated with family and parenting, suggesting these terms and attitudes may be transferred to future generations.

Intolerance in online conversation varies by region. In the US, Texas, Louisiana, Mississippi, Alabama, Florida and Southern Carolina host the most racially intolerant tweets (relative to population), while in the UK the most concentrated regions are Northumberland, Liverpool, Peterborough and Surrey.

Hate speech authors are overrepresented within specific demographics, including sports enthusiasts (more than twice as prominent) and executives (30% overrepresented). By contrast, teachers and scientists, as well as those interested in politics and environmental issues, were least present in online hate speech.

Female authors accounted for 52% of misogynistic language use. The data showed that misogynistic terms are used as insults between female authors, suggesting they sit within mainstream lexicon and may not consciously be considered misogynistic (though still typically used to insult females or perceived feminine behaviours).

Educators Play a Pivotal Role in Equality Debate



In the chart above, the white nodes represent teacher and lecturer conversation and blue nodes represent family and parents. The size of each node reflects how big a talking point it is within each group. Higher nodes indicate more neutral and supportive debate around the topic, while nodes below the zero line indicate that insults and hate speech are more common.

Authors associated with family and parenting were overrepresented in hate speech across all topics. The largest area of discussion was racial intolerance, suggesting a potential challenge for attitude transference between generations.

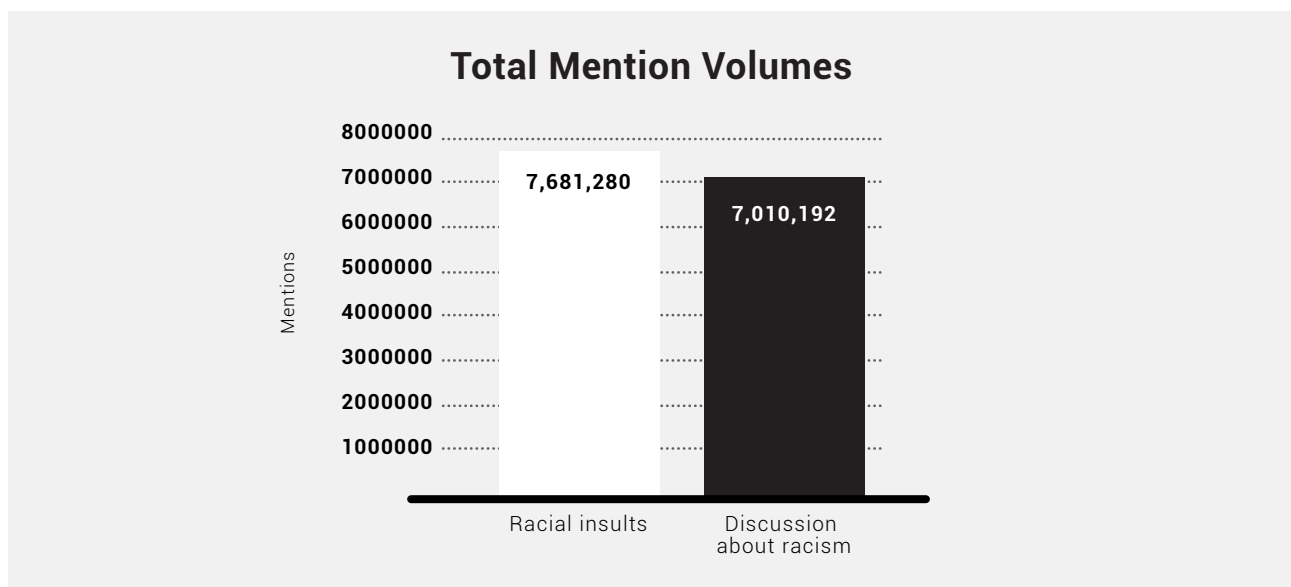
Pejorative discussion was also pronounced for masculinity constructs and misogyny. Home life is typically the most influential source of gender constructs for children and early adolescents again suggesting risk for negative attitudes to impact future generations.

By contrast, teachers and lecturers were more likely to encourage debate surrounding gender issues, highlighting one means by which gender constructs may be challenged.

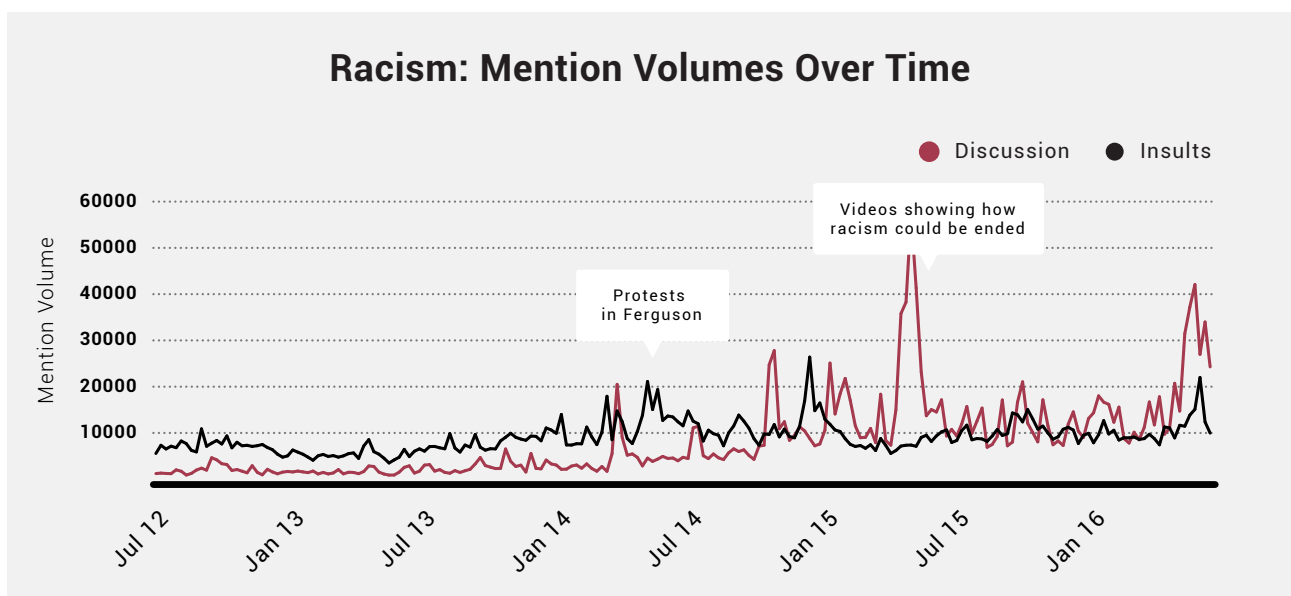
Conversation surrounding homophobia and transphobia was much more constructive on average among both parents/family members and educators. However, while educators were more vocal surrounding sexual orientation, transphobia remained a much smaller talking point. This could mark need for more visibility surrounding the rights of transgender individuals, both in the home and in educational contexts.

Racial Intolerance

Trend analysis



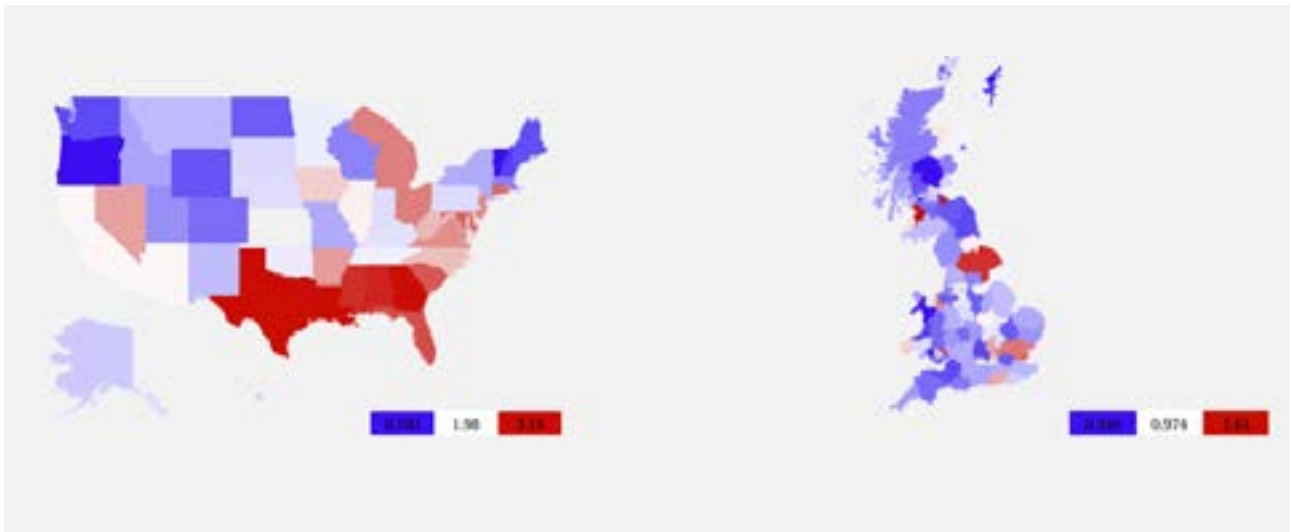
Across the time frame studied, neutral conversation around racial inequalities grew more strongly than racial hate speech. However, overall racial insults were 8% more visible than neutral discussion. The trend lines show that as of January 2015, weekly volumes of neutral conversation outgrew pejorative language.



June and July 2016 saw a peak of neutral conversation around racism. This was triggered by landmark political events in both the UK and US. Notably, the stories that triggered peaks in neutral conversation did not impact racist language. The limited overlap in topics driving neutral and negative peaks suggests that these conversations were held by two distinct author groups. We analysed lists of the most impactful authors, both within racial insults and among those discussing racism, and found an overlap between the two groups of less than five per cent. This divide may pose a challenge for those hoping to raise awareness of racial intolerance, as the authors they may hope to influence are likely not in their immediate networks.

Due to large volumes a 25% representative sample was taken for both queries. The trend chart shows 25% sample volumes while the bar chart shows actual volumes per query.

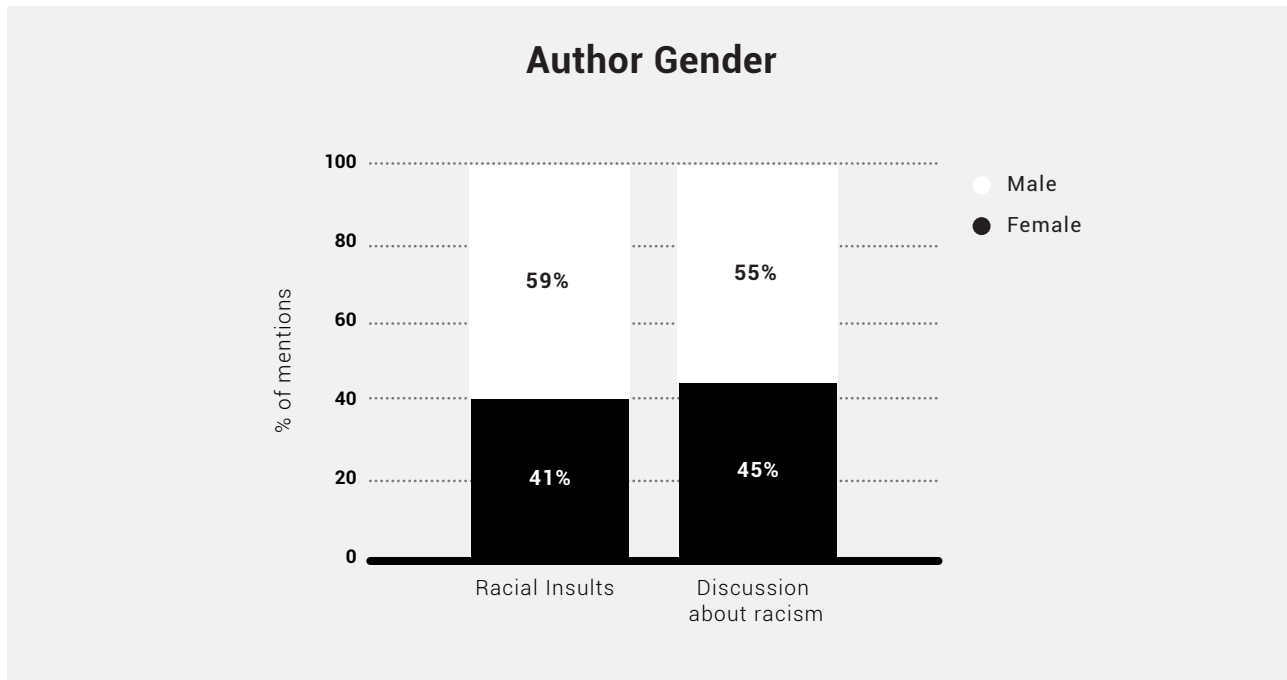
Heat Map Analysis



In the US, southern states showed the highest concentration of pejorative racial language with Texas, Louisiana, Mississippi, Alabama, Florida and Southern Carolina standing out. Vermont, Oregon and a majority of North-Western states emerged as more likely to host neutral or constructive debate.

In the UK, areas of heated negative discussion were scattered throughout the country with East Ayrshire, Midlothian, North Yorkshire, Blaenau Gwent and Torfaen emerging as most prominent. Despite both being neutral overall, Scotland appeared to lean slightly more towards the neutral end of the spectrum (blue) while England and Wales hosted larger shares of prevalently negative countries (shown in red shades in the maps above).

Demographic Analysis

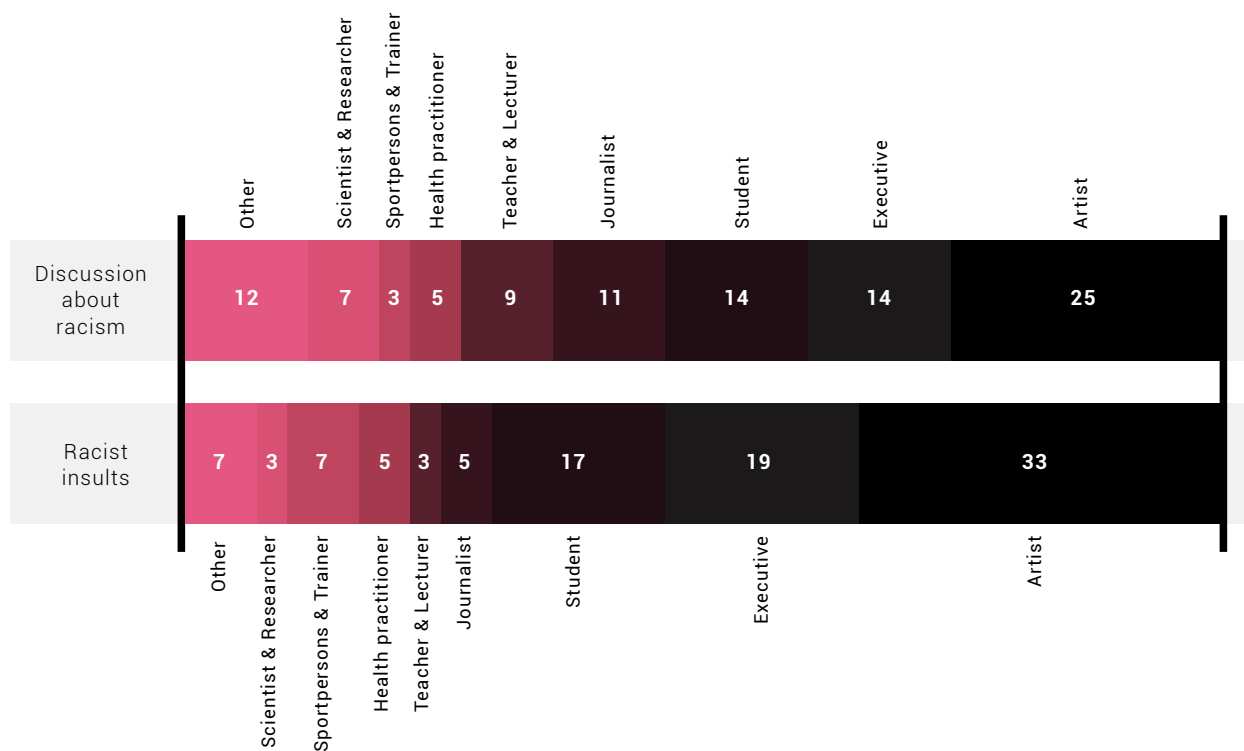


Overall, male authors were more active in discussing racial intolerance than females with the skew being even more pronounced for insulting language than for neutral.

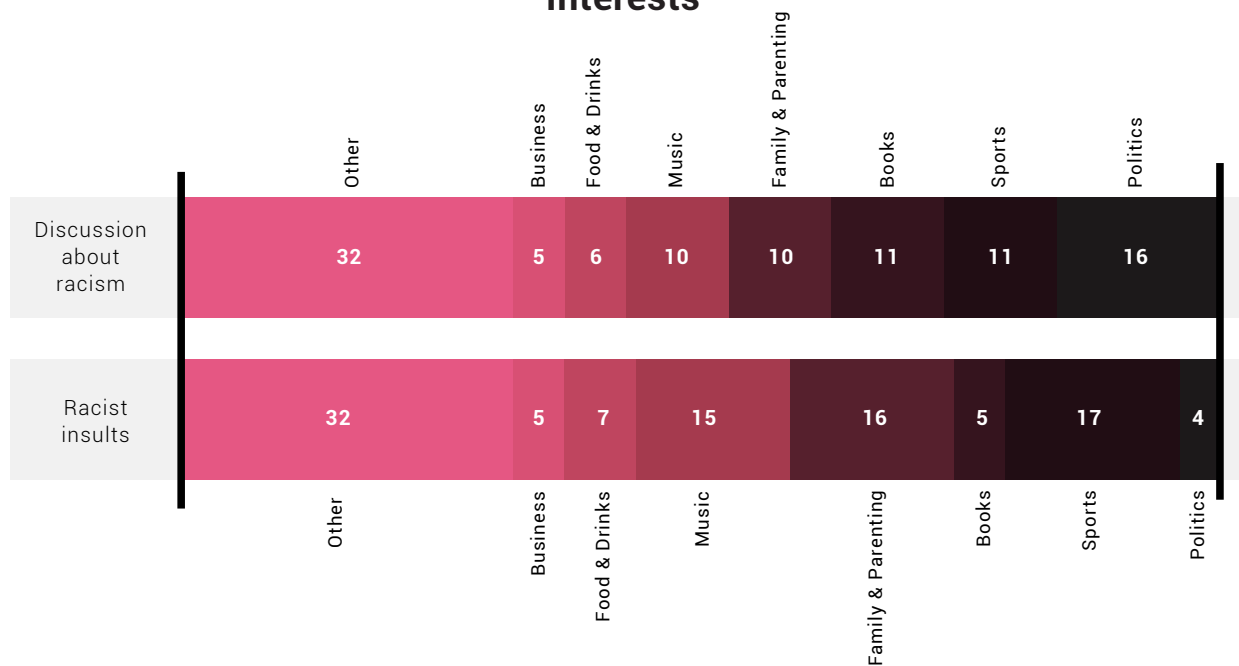
Authors working in educative fields (scientists & researchers, teachers & lecturers and journalists) were more active in neutral than in negative discussion. Students instead showed greater use of racial insults than neutral language, suggesting room for improvement around the reach and impact of conversations by educative professionals.

Politics was the strongest interest among neutral conversations of racial intolerance. This interest was four times weaker among authors using racial insults who were instead more likely to be interested in sports, music, family and parenting.

Professions



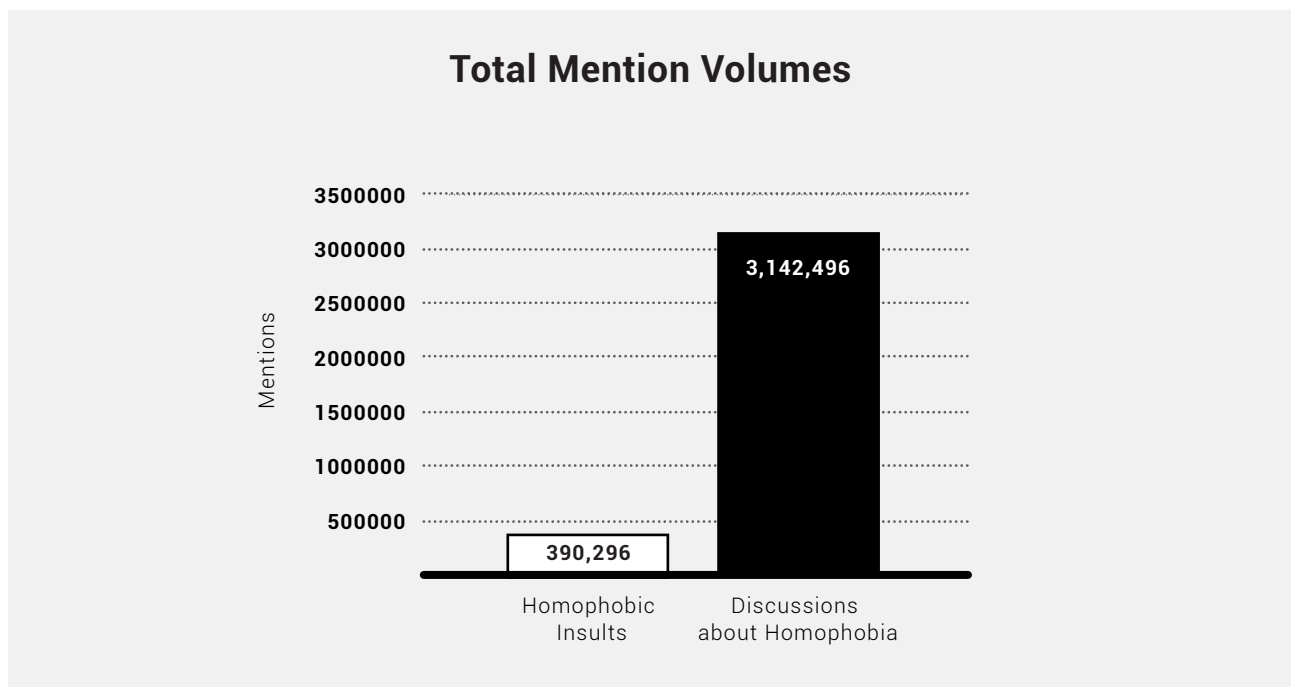
Interests



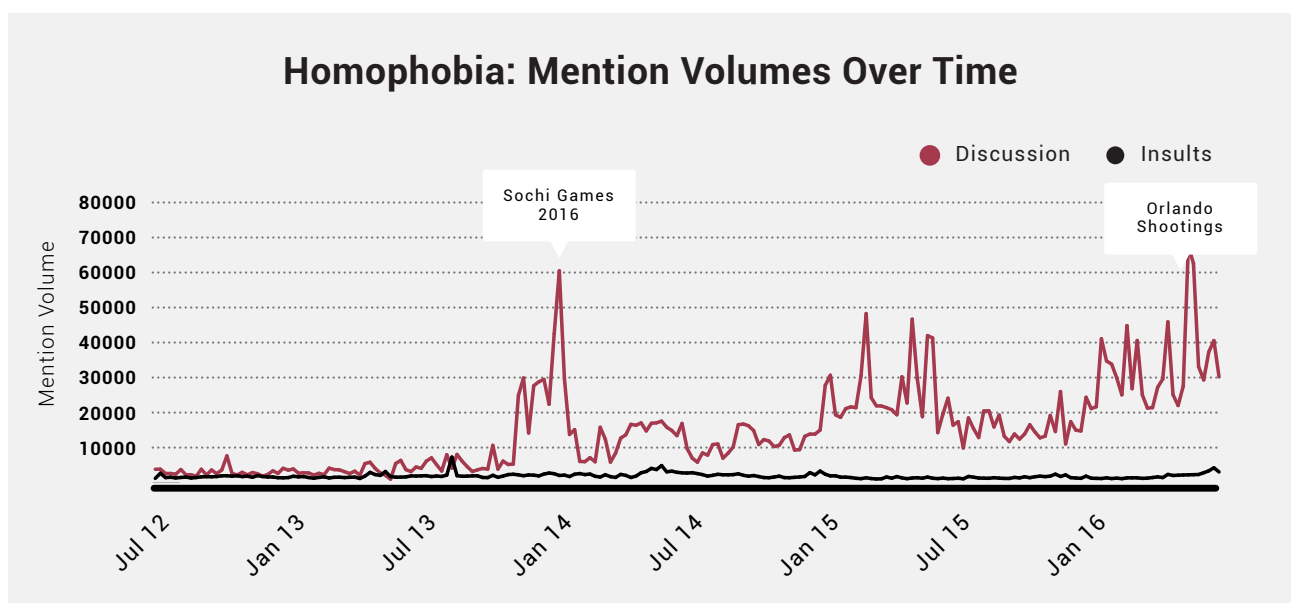
Professions and interests in category 'other' listed in appendix.

Homophobia

Trend analysis



Homophobic hate speech was eight times less common than neutral discussion about homophobia.

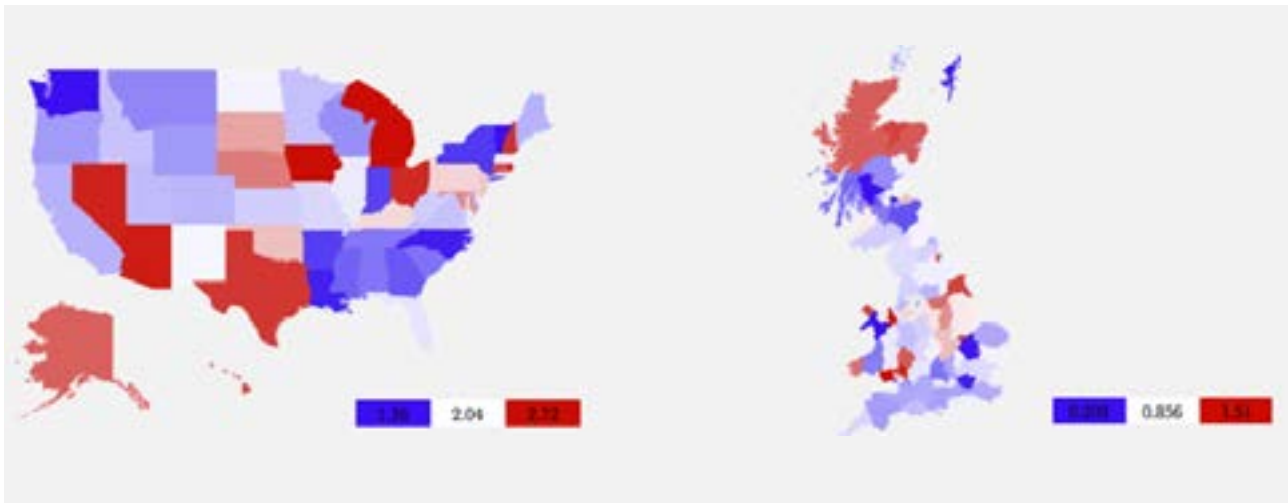


Neutral conversation peaked around two key events: the 2014 Winter Olympics in Sochi and the Orlando Shootings in June 2016. Both claimed around 2% of overall neutral homophobia discussion but showed minimal direct impact on the volume of homophobic hate speech. Engagement around these peaks was largely driven by actions/messages of support for the LGBT+ community by prominent authors such as **Channel 4** and **Germany's national sports team**, who rebranded their logo/opening ceremony outfits to reflect LGBT colours (Sochi), **gay rights activists** across the world who met for a day of protests (Sochi) and US presidential candidate **Hillary Clinton** who expressed support for the LGBT community on Twitter (Orlando).

Both these peaks were closely linked to politics and political figures, suggesting a sphere in which political actions and individuals can have a significant impact on awareness of homophobia. However, the relatively stable use of homophobic language throughout the time frame suggests more work is needed to enable change within these author communities.

The trend chart above shows % distribution for visual coherence.

Heat Map Analysis

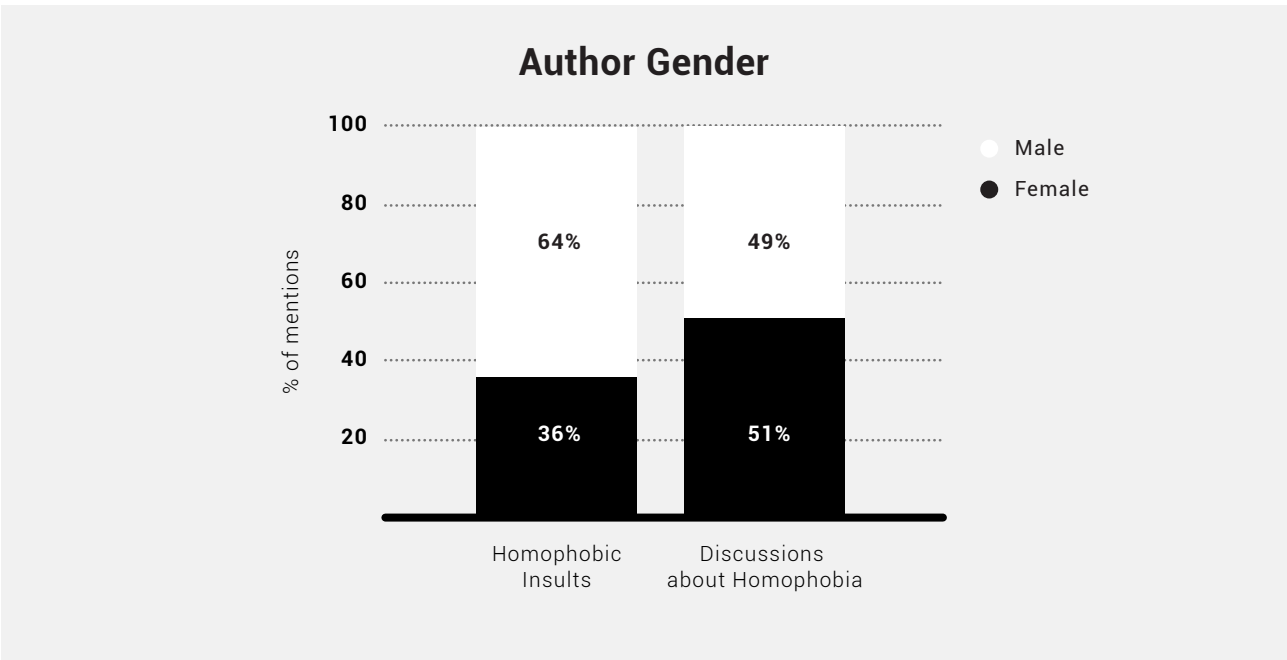


In the US, homophobic hate speech was strongly prevalent in Michigan, Ohio, Nevada, Arizona and Texas. Washington, Vermont, North Carolina and Louisiana were among the states to witness the lowest levels of homophobic engagement online.

The UK exhibited less homophobic discussion (relative to debate) in comparison to the US (23% and 56% respectively), with Scotland showing the most widespread level of homophobic language. Pockets of homophobic hate speech were seen in Wales (Glamorgan, Gwent and neighbouring Herefordshire) and England (East Yorkshire and Peterborough).

The UK exhibited a large number of counties in which homophobic hate speech was balanced by the neutral conversations around homophobia. By contrast, in the US individual states leaned more clearly towards either extreme suggesting that overall, suggesting that US states were more polarised around the topic of homophobia than counties in the UK.

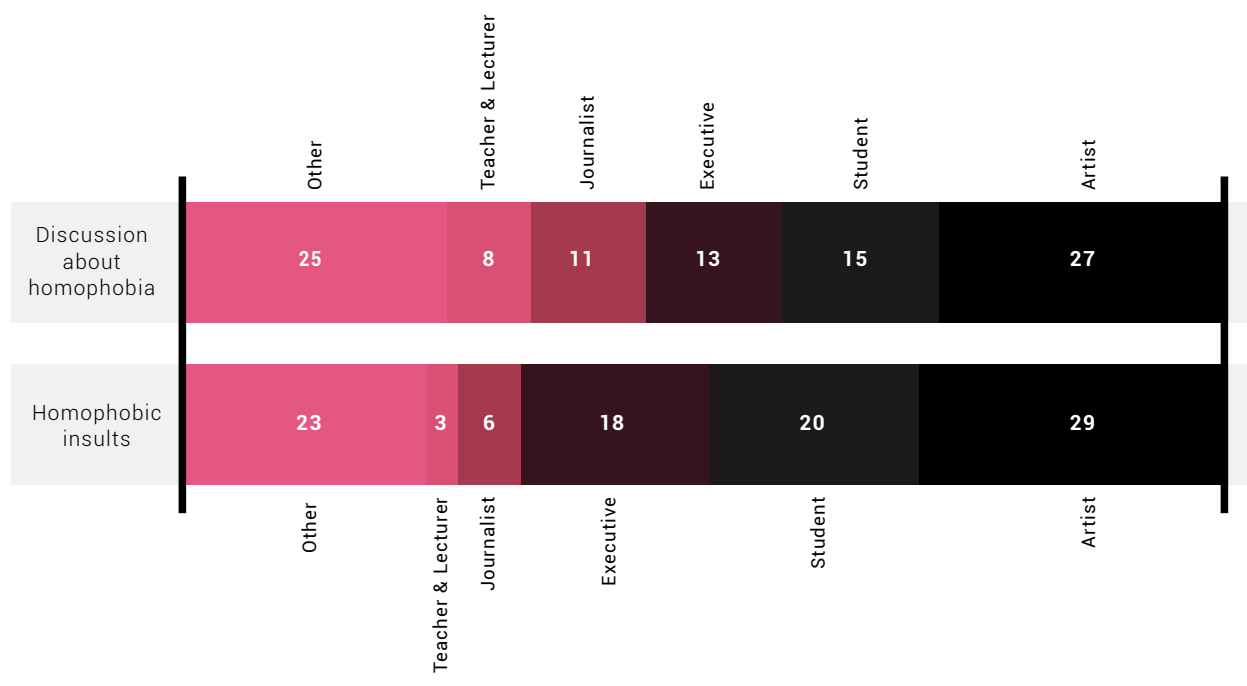
Demographic Analysis



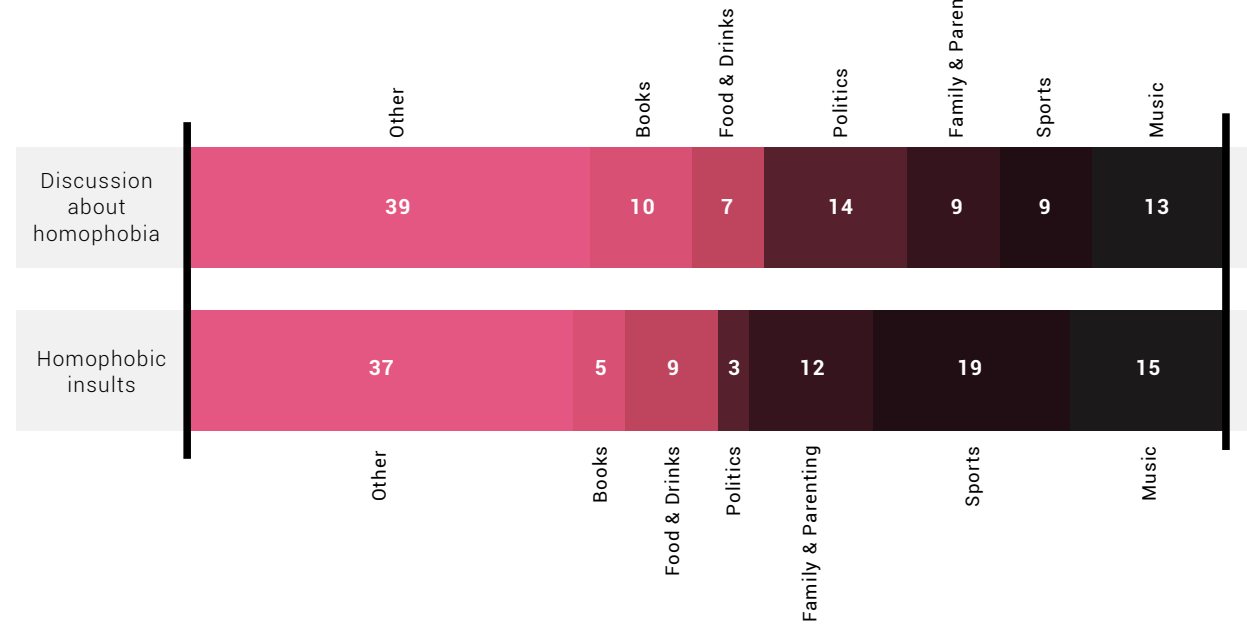
Gender analysis revealed a higher level of male engagement (64%) for homophobic insults, while an even split was observed for general discussion. Within homophobic hate speech, slight differences emerged between male and female conversation topics, with discrimination and insults tending to be gender equivalent (topics such as ‘fucking queer’ emerged in male conversation, while female content contained terminology such as ‘fucking lesbian’).

Authors engaging in homophobic insults were more likely to be students and executives, while also expressing a greater interest in sports. Yet topical interest remained similar across professions and interests, with no defining topical engagements unique to executives or students, or those interested in sports.

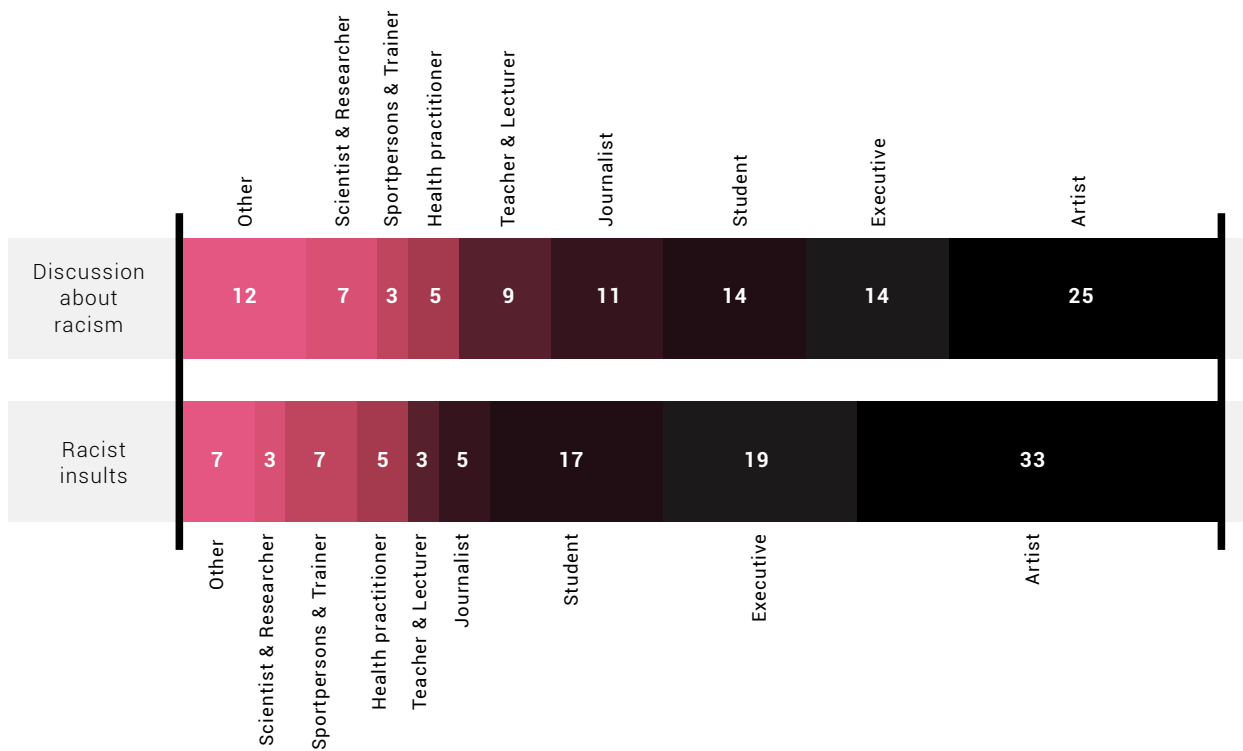
Professions



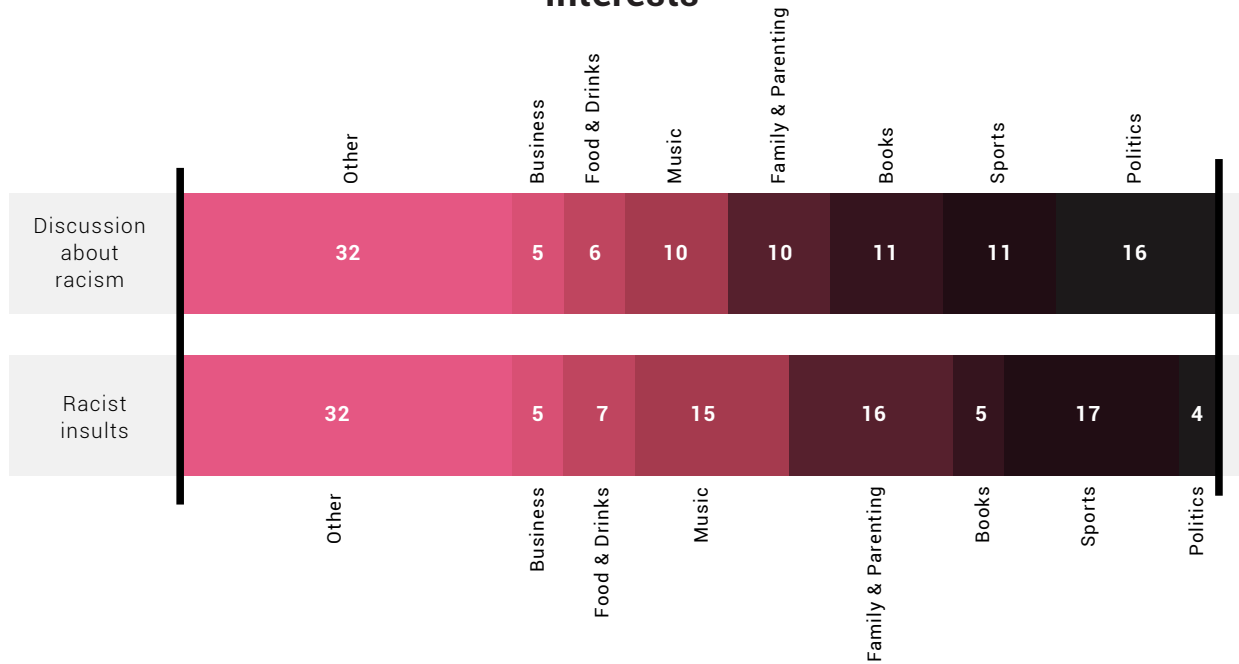
Interests



Professions



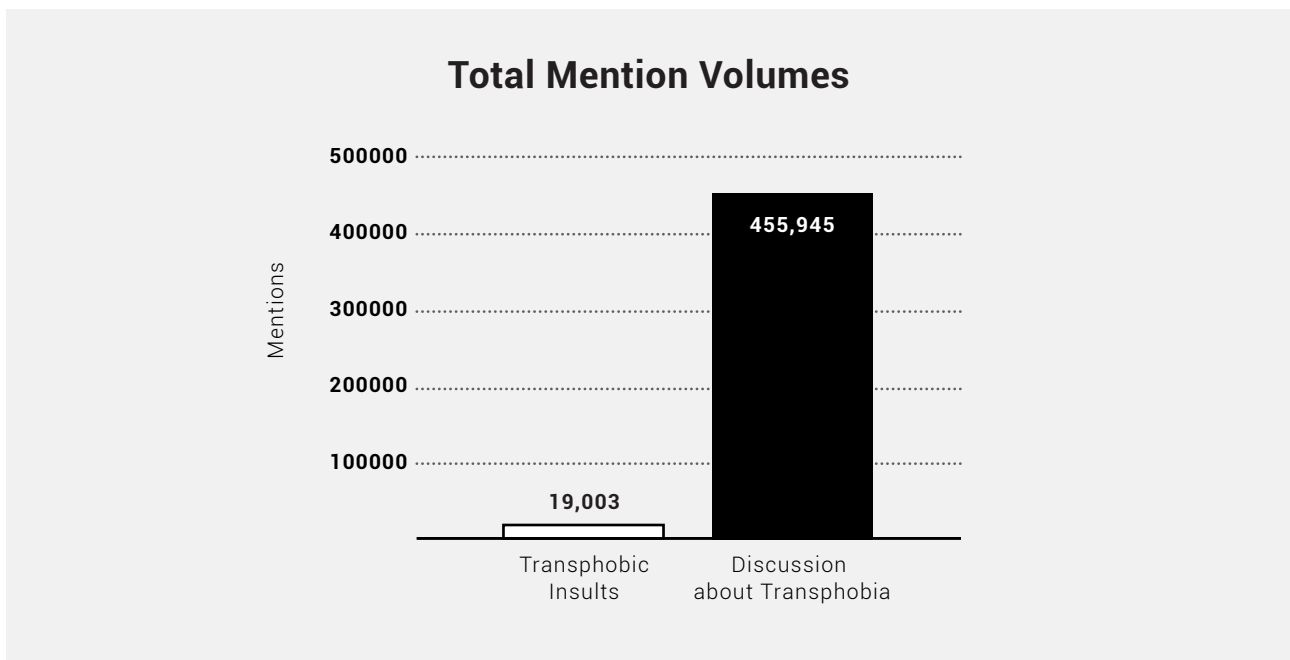
Interests



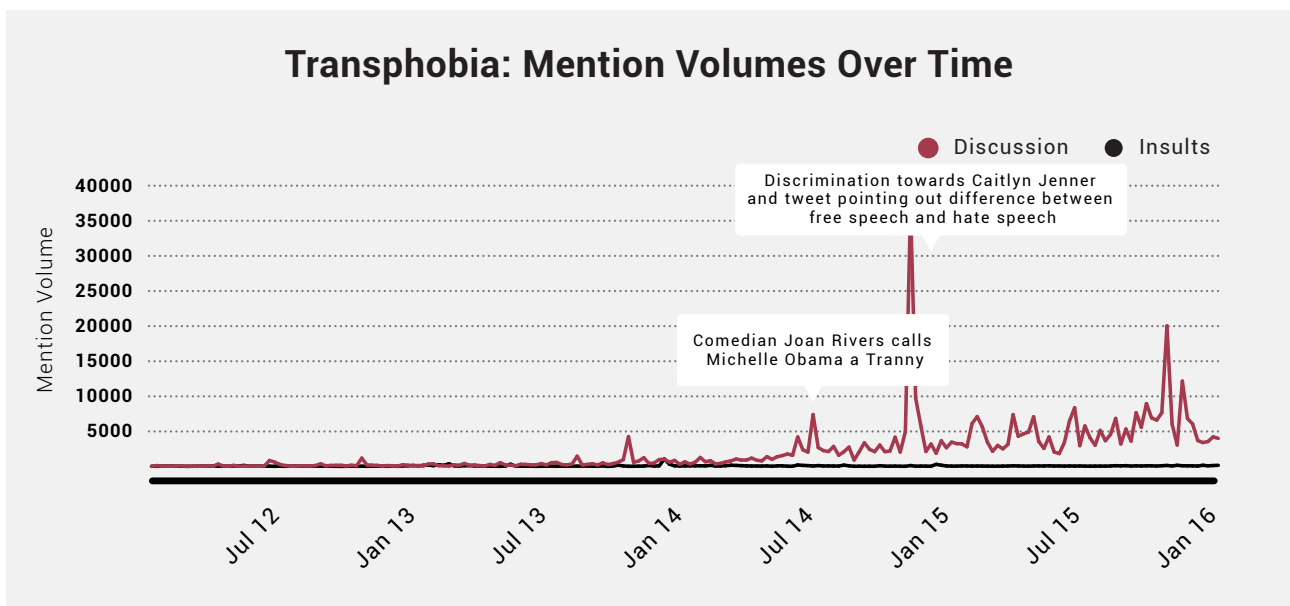
Professions and interests in category 'other' listed in appendix.

Transphobia

Trend analysis



Neutral discussion about transphobia was 24 times more prominent than transphobic hate speech, the strongest ratio across all discrimination areas.

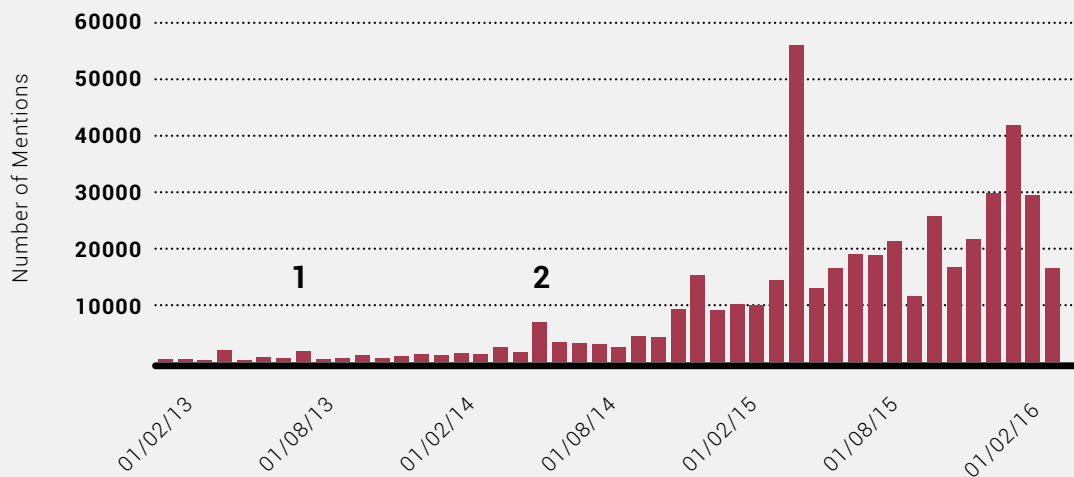


The largest peak in neutral transphobia conversation occurred in the week commencing June 1st 2015. A significant share was driven by engagement around **@beeteezy** content regarding racism, transphobia, homophobia and misogyny not being opinions, stating the difference between free speech and hate speech (12,262 retweets). Other engagement was driven by a transphobic tweet by **@DrakeBell** in which he insisted on calling Caitlyn Jenner 'Bruce'. **@elielcruz** encouraged other users to share her content to expose the author (2,700 retweets).

Transphobic language peaked on June 30th 2015 due to news of comedian Joan Rivers calling Michelle Obama a 'tranny'. Engagement came in the form of authors sharing news content involving footage of the comedian's comments, noting the statements as a 'bombshell'.

The trend chart above shows % distribution for visual coherence .

Transphobia Debate Has Gained Momentum



1 “There is an international day against homophobia and transphobia, and it is today”	“Disagreeing with a Trans person DOES NOT make you transphobic. It means you disagree with someone who happens to be Trans #gamergate”
2 “Fight transphobia every day! Today is Transgender Day of Remembrance 2013 #solidarity #support”	“Target customers need to stop being so transphobic I swear the next person who says they don’t want to use the bathroom if “men” can just... ”

As shown by the chart above, conversation about transphobia was scarcer in 2012 and 2013, however gained momentum from the second half of 2014. This may indicate an increased awareness and need to be vocal about issues affecting the transgender community. Over time, more authors have begun commemorating days such as International Day Against Homophobia and Transphobia and Transgender Remembrance Day. The share of total conversation around

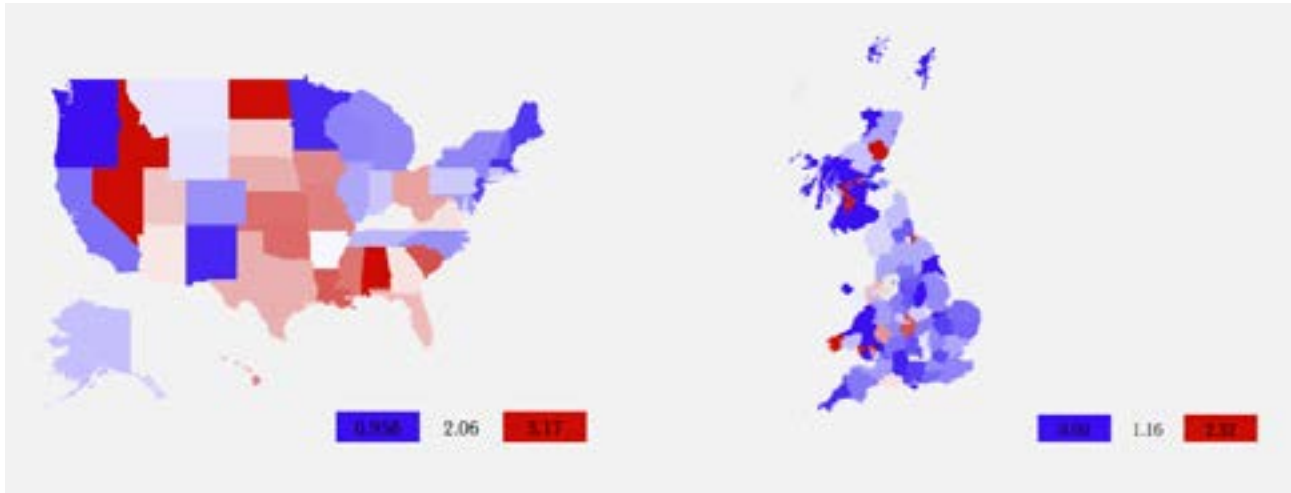
these two days was about 50% in 2012, it then decreased to around 14% in 2015 even though total conversation increased (due to a higher base level of discussion across the year).

Instances of transphobia or perceived transphobia shaming started emerging in the fall of 2014 with the divided opinions on whether Gamegate's introduction of a character identifying as transgender was transphobic. This discussion permeated the transphobia conversation through mid-2015. **Planet Fitness'** involvement in a lawsuit also led to claims that the attorney had made transphobic remarks on public radio. Public speeches were also scrutinized with authors calling out perceived transphobic slurs (including **Ellen** DeGeneres during the Oscars, Joan River, **National Review's** article on Laverne Cox).

TV series featuring transgender characters (Orange Is The New Black and SVU) prompted polarized discussion around the topic with some viewers praising the shows for celebrating diversity and shedding light on discrimination issues. Several authors suggested that SVU had been **transphobic** in the past. Popular culture sources such as these could be further leveraged to educate viewers on transphobic issues.

Legislation around bathroom rules for transgender individuals in early 2016 also played a role in igniting a debate which continued through summer 2016. Public issues could be an opportunity for DtL to participate in the social conversation by creating hashtags that foster constructive conversation.

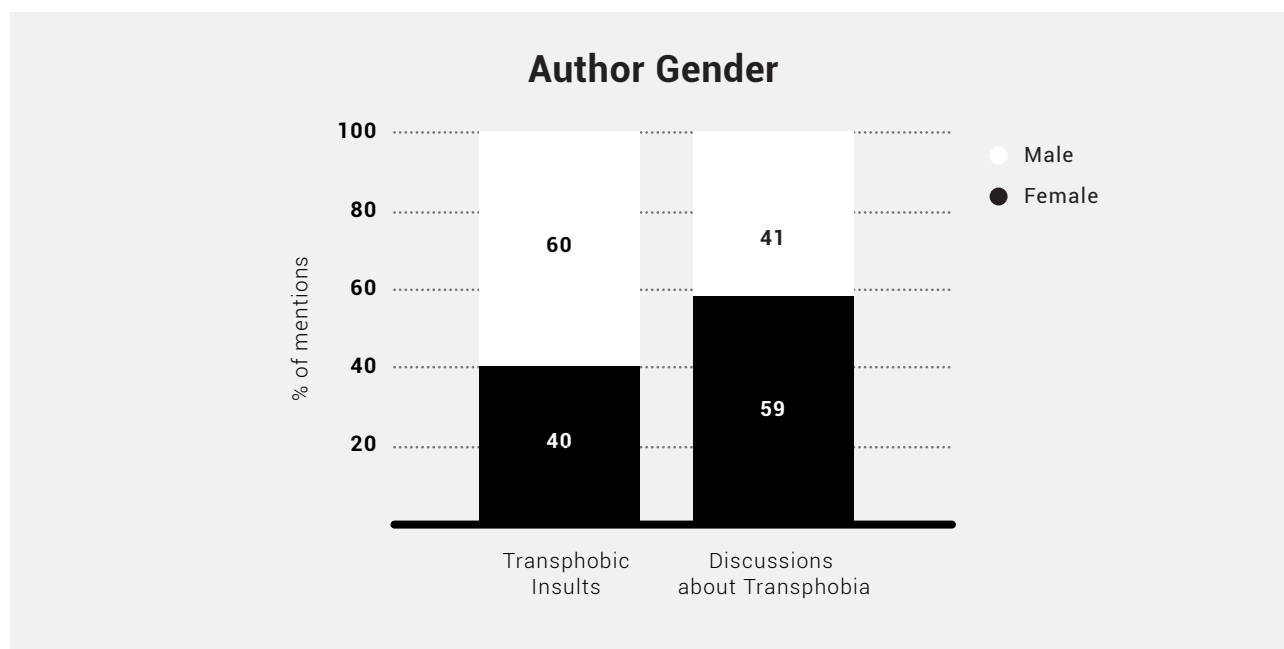
Heat Map Analysis



The maps above show the range of transphobic discussion across US states and UK counties. Nevada, Idaho, Alabama and North Dakota exhibited the highest ratios of transphobic discussion (relative to neutral discussion and debate). Kansas, Oklahoma and Louisiana also hosted above-average levels of transphobic language. By contrast, Washington, Oregon, Minnesota and New Mexico hosted the lowest transphobic scores.

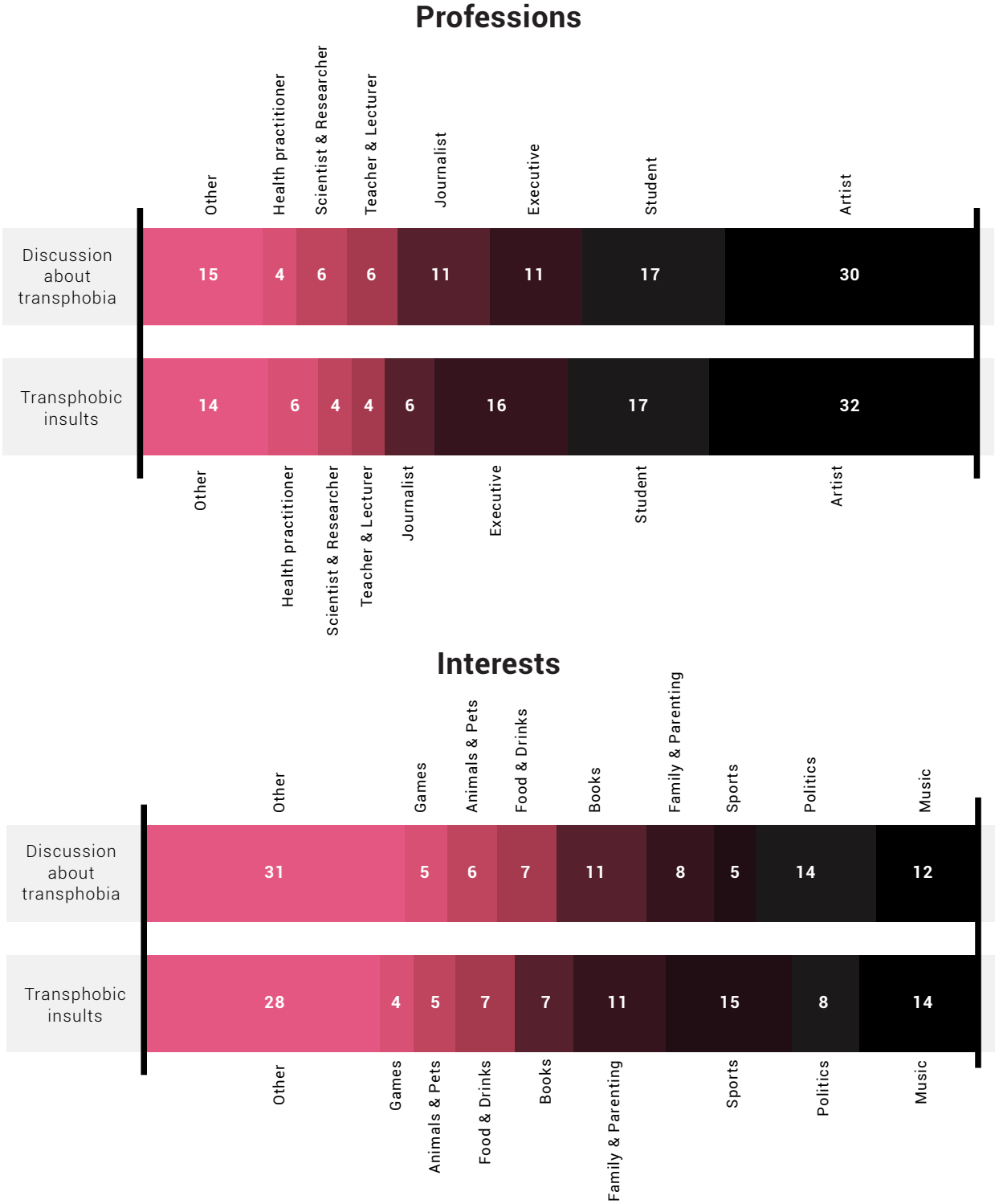
In the UK, Perth & Kinross, Lancashire, Derbyshire, Aberdeen City and Hampshire experienced the highest transphobic scores. Kent, Nottinghamshire and Sterling were the counties to host the lowest levels of transphobic conversation. The UK and the US differed to the extent that the UK hosted more neutral and low-level transphobic engagement, while in the US, transphobic conversation varied more widely between states.

Demographic Analysis



Gender distribution revealed a highly engaged female audience in neutral conversation around transphobia (59% female), with male authors more likely to use pejorative language (60%). In neutral conversation, Islamophobia emerged as a key talking point for both genders. This highlights the grouping of discriminatory mind-sets and engagement online, with popular content often encouraging authors to see discrimination as a wider issue pertaining to general human rights. Again, political reference was common here, with key **authors** such as Human Rights Campaign (@HRC) noting Clinton's appreciation of access and opportunity as key American values, put in contrast with transphobia, Islamophobia, xenophobia and racism.

Journalist and executive were the only professions to see a difference at query level, with ‘insults’ exhibiting 6% and 16% respectively, while general discussion hosted 11% and 11% respectively. As was seen for homophobia, authors were more likely to be interested in Sports discussion (15% compared with 5%)



Correlation Analysis

Correlation Overview (US)

Correlation Overview (US)										
	RACISM	TRANSPHOBIA	HOMOPHOBIA	MASCULINITY	MISOGYNY	AVERAGE ANNUAL INCOME	REPUBLICAN	NO POLITICAL AFFILIATION	DEMOCRAT	WEALTH DISTRIBUTION
RACISM		0.427	0.241	0.241	0.741	-0.219	0.150	-0.060	-0.120	0.116
TRANSPHOBIA	0.427		0.173	0.581	0.505	-0.379	0.622	0.276	-0.541	-0.038
HOMOPHOBIA	0.241	0.173		0.431	0.528	0.241	0.094	-0.154	-0.305	-0.066

The grid above displays correlations between the five discrimination topics, as well as outside topics: Average Annual Income*, those who self-identify as Republicans, no political affiliation, or Democrat**, and wealth distribution (Gini Coefficient)***. Correlations are based on a scale of -1 to 1, with zero representing no correlation, -1 representing a perfect negative and 1 representing a perfect positive correlation.

The strongest negative correlation was found for Misogyny and self-identified Democrats (-0.929), meaning that states with lower levels of misogyny tend to be stronger bases of Democratic support. The same was true to a lesser extent for transphobia (-0.541), suggesting that Democrat-strong regions are less likely to tolerate misogyny or transphobia in online discussion.

Racism and Misogyny saw the strongest uphill correlation (0.741) meaning states with high levels misogynistic language are also likely to exhibit less racial tolerance in the data. Homophobia and Transphobia had the weakest positive correlation (0.173), despite often sharing advocates within LGBT+ communities. This could potentially be a springboard to further investigate the relationship between these terms through the scope of social chatter.

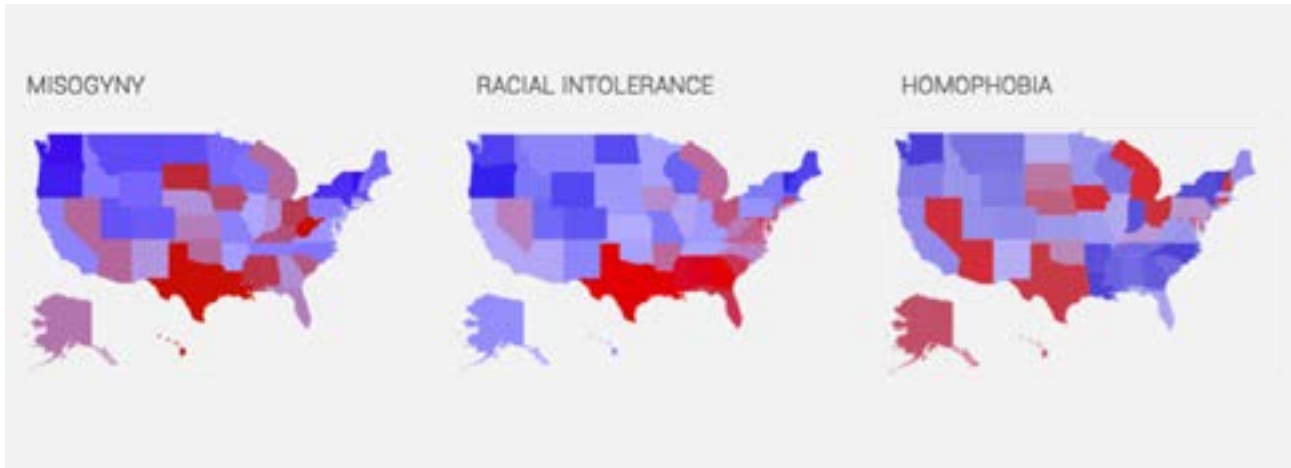
Methodology: Ratios for the five topics were found by dividing the insulters within the category by those who found the category to be intolerant. These ratios were then correlated against each other, as well as outside topics.

*Data is based off of the 2012 US Census results, released in 2013

**Data is based off of the 2014 Pew Research study

***Data is based off of the 2013 wealth coefficient research by Frank Gini

Correlation Analysis (US)



The heat maps above correspond to misogyny, racial intolerance and homophobia in the US, showing overlap across certain states. Texas hosted high levels of discriminatory content across all three constructs, and hosted the highest misogyny ratio (insults compared with neutral discussion) at 0.738 as well as the second highest racial intolerance ratio of 1.319. Nevada and Arizona both expressed high levels of homophobic engagement and slight levels of misogynistic content. Michigan similarly exhibited strong level of homophonic engagement, and slight levels of racial intolerance and misogynistic mentions.

While southern states including Texas, Louisiana, Mississippi, Alabama, Georgia, Florida and South Carolina hosted higher levels of racial intolerance online, the majority of southern states were neutral in the context of homophobia, with Texas and Arizona the only exceptions. Louisiana (1.621), Texas (1.319), and Georgia (1.298) saw the highest ratios of racial insults to race debate, meaning for almost every two racially insensitive tweets, there was one tweet about racial tolerance.

Washington and Oregon exhibited neutrality and low discriminatory scores across all constructs. Both held low ratios for topic insults to topic tolerance, with Washington ranking in the top five lowest ratios for all three categories, (0.426 racism, 0.068 homophobia, and 0.224 misogyny).

Correlation Overview (UK)

Correlation Overview (UK)										
	RACISM	TRANSPHOBIA	HOMOPHOBIA	MASCULINITY	MISOGYNY	VOTED YES TO LEAVE EU	CON 2015	UKIP	TURNOUT 2015	LONG TERM UNEMPLOYMENT
RACISM		-0.295	0.478	0.279	0.632	0.0139	-0.0658	-0.047	0.086	0.145
TRANSPHOBIA	-0.295		-0.315	-0.443	-0.334	0.334	-0.0813	0.095	-0.197	0.152
HOMOPHOBIA	0.478	-0.315		0.157	0.752	0.138	-0.0691	0.200	-0.142	0.163

The grid above displays correlations between the three measured topics (Racism, Transphobia and Homophobia), as well as outside topics: voted “yes” to leave the European Union, voted for the Conservative party in 2015, voted for UKIP in 2015, the share of voters who turned out for 2015 election, and long term unemployment in the UK. Correlations are based on a scale of -1 to 1, with zero representing no correlation, -1 representing a perfect downhill correlation and 1 representing a perfect uphill correlation.

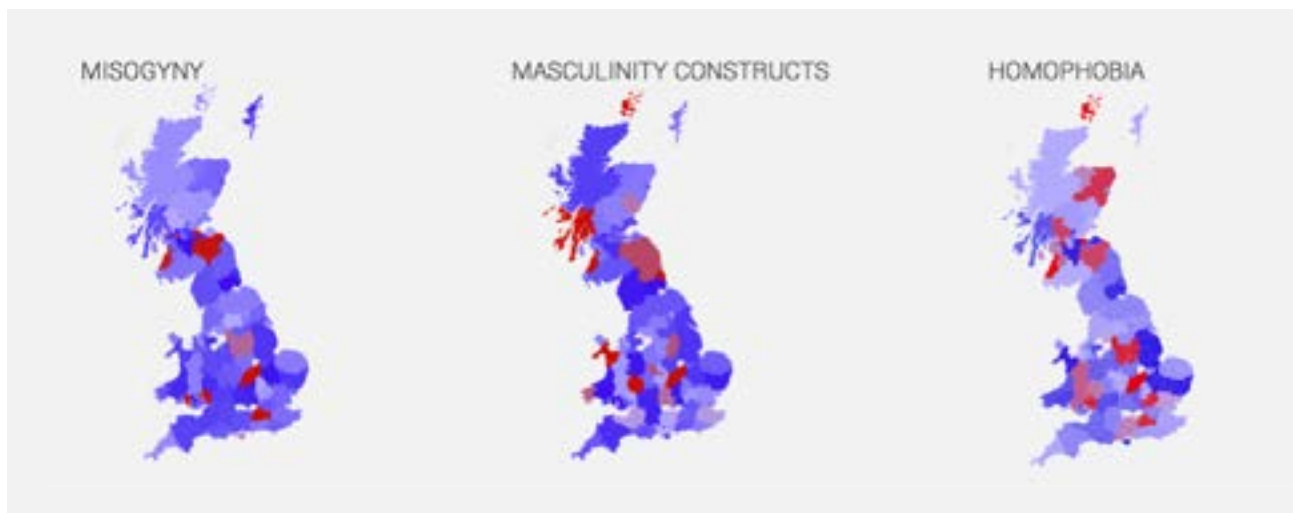
Homophobia to Misogyny saw the strongest positive correlation (0.752) among the topics, whereas Masculinity to Homophobia saw the weakest uphill correlation (0.157). Homophobia only had one downhill correlation, which did not feature as particularly strong, with Transphobia (-0.315), indicating the topic's tendency to be mentioned in conjunction with the other categories.

Interestingly, “Voting ‘Yes’ to leaving the EU” was not a reliable predictor of hate speech, including racial intolerance. This follows mainstream news reports of increases in hate crime following the referendum², and suggests that varying attitudes within each county make for a more nuanced picture of attitudes to race and nationality across the UK.

Variable census data collected from the 2015 General Election Results, British Election Study, 2015 (sourced 29/08/2016).

2. Lusher, A. 2016. “Racism Unleashed”, Independent, 07/28/2016.

Correlation Analysis (UK)



The heat maps above correspond to misogyny, masculinity constructs and homophobia in the UK, showing overlap across certain counties. Northamptonshire and South Ayrshire emerged as counties that exhibited high scores across all three constructs. Aside from Northamptonshire and South Ayrshire, regional distribution of discrimination appeared to be random and lacked consistency across constructs. This supports the notion that discrimination across the UK tended to be concentrated within a few regional pockets, while the majority of counties holding a more balanced range of views towards discrimination areas online.

2.0

Online Bullying

—

Online Bullying Summary

Progress

The data supports current Twitter advice to not necessarily interact with online bullying. Recipient replies led to escalated conflict in 44% of cases, compared with only 3% positive outcomes. This type of finding can be used to inform recipients who encounter online trolls.

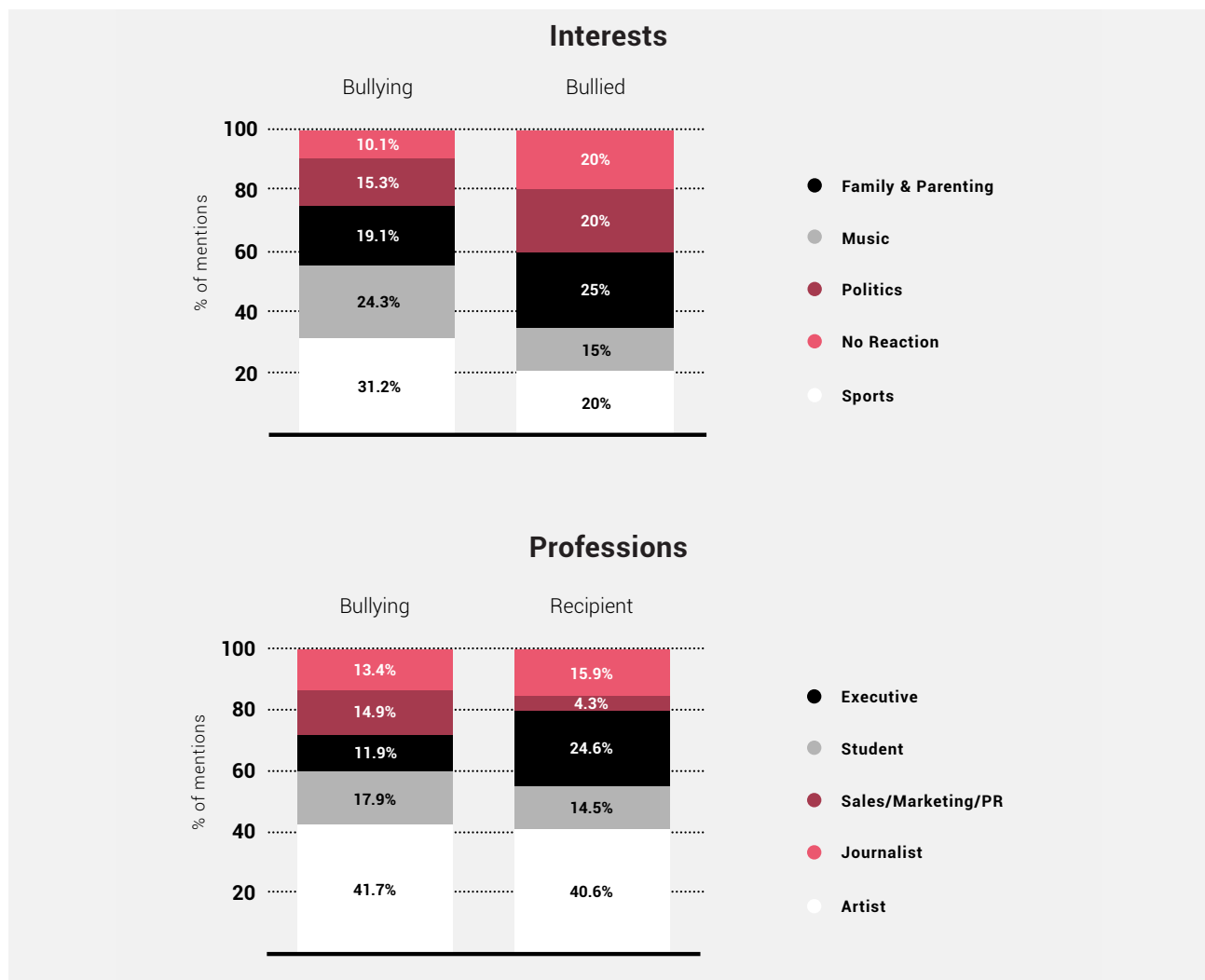
Less confrontational replies are less likely to escalate. Recipients of bullying who reply with offensive language were more likely to escalate the conflict, especially when relating to contentious issues such as politics. By contrast, more reasoned responses stood a better chance of diffusing the situation and enlisting support from others.

Challenges

Discrimination is visible within online bullying. The majority of insults related broadly to intelligence and appearance, however sexual orientation, religion and gender also featured within our sample.

Bullying scenarios are scattered. While politics was the most common topic to be met with bullying remarks, there were no consistent themes predictive of online trolling. The finding, that most topics can lead to online bullying, poses a challenge for tracking and countering bullying behaviour online.

Females More Likely to Engage in Troll Discussion



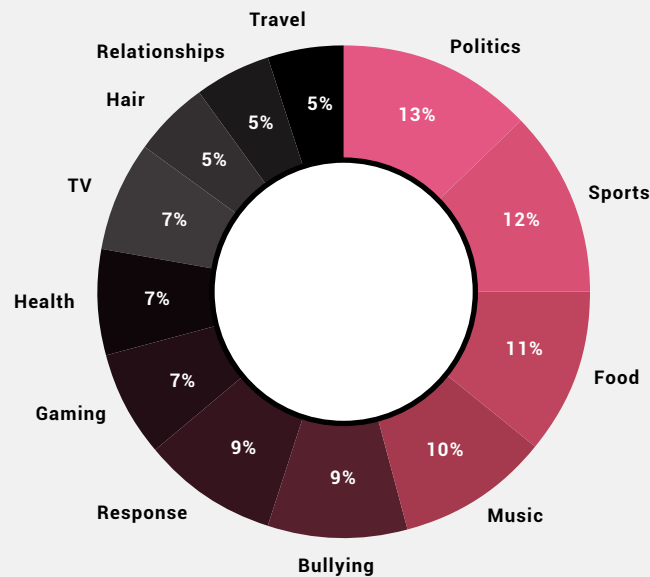
The charts above represent a demographic breakdown of Twitter trolls and their recipients. Females were more likely than male authors to be involved in this type of conversation, but the gender breakdown was roughly even between groups, with 66% of bullying authors identifying as female and 65% of recipients identifying as Male. Female authors tended to use insults relating to intelligence (dumb, stupid), appearance (fat, ugly), and derogatory animal terms (bitch, chicken). Male trolls also referenced intelligence in insults (stupid, dumb, moron) and appearance (ugly, fat), but were also more likely to use homophobic insults (faggot). Creating messaging around these prevalent themes for both genders could assist in educating and alleviating future online bullying behaviour.

Earlier in the study we identified executives as over-represented in online hate speech. However, as shown above, executives were also more likely to be the recipients of online trolling, perhaps due in part to their more prominent/visible online status.

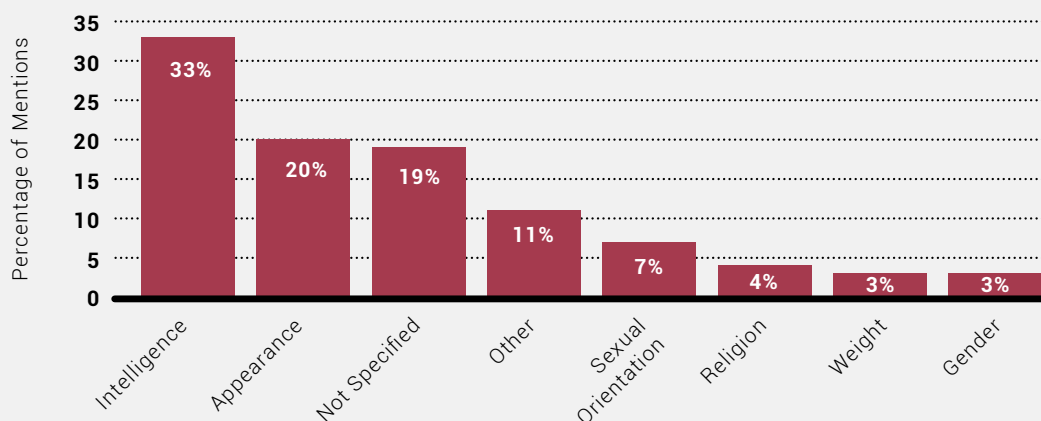
* Brandwatch Twitter demographics segment authors on an automated basis according to name and bio information.

Trolls Respond to Wide Range of Discussion Topics

Topics Shared by Recipients prior to Bullying



Topics of Troll Content

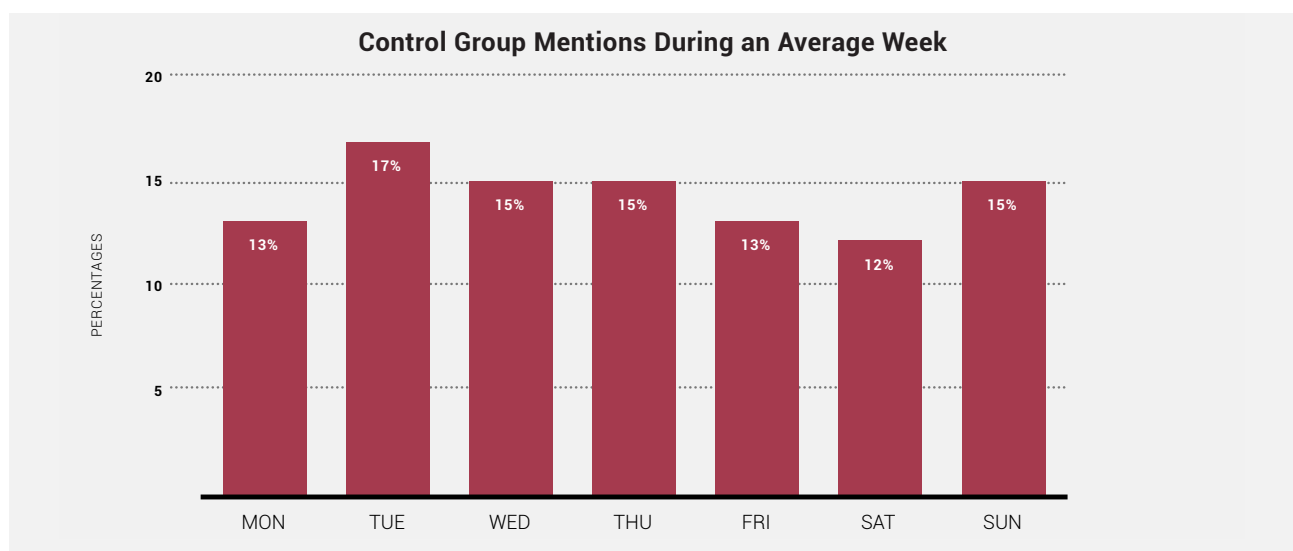
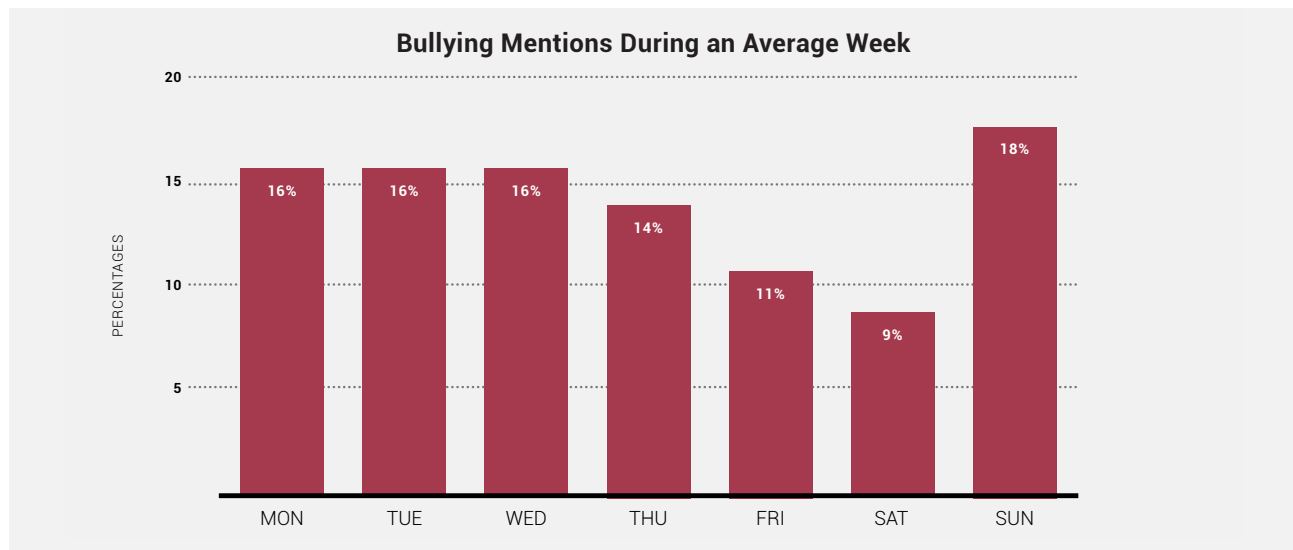


The charts above reflect content by those who were bullied in their three tweets prior to being targeted. While a few authors used hashtags, mostly around sports or television shows, no two authors used the same hashtag in our random sample. 80 topics emerged within our sample; the chart above reflects the top 12. The broad range of topics prior to bullying suggests that trolls are not restricted to a limited range of topics. Rather, any number of topics may be met by troll content.

Politics held the largest share of voice (13%) followed by Sports (12%), and Food (11%). Recipients who were trolled for their political views were more likely to be insulted based on their intelligence, more so than any other category of insult. Within pre-troll Sports conversation, the range of insults varied, with appearance and general insults holding the largest shares. The Response category indicates a bullied user responding to another peer, usually in the form of a yes or no response, but did trigger bullying which focused on insulting the user's intelligence as well as general insults.

*Sample of 100 bullied authors out of 400 bullied, confidence interval of 10

Bullying: Day of the Week

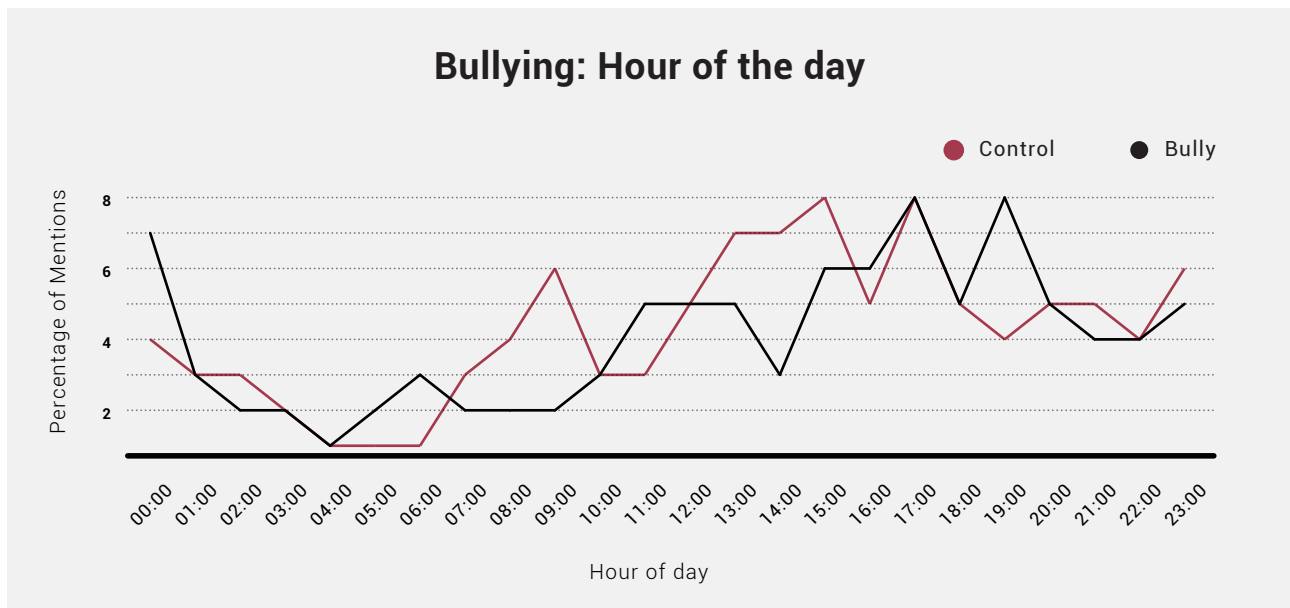


The charts above show shares of activity throughout an average week. Figure A represents when trolling takes place and Figure B shows a control group (representative of broader Twitter discussion).

Bullying authors tweeted most prolifically on Sunday (18%), followed by Monday, Tuesday and Wednesday, all at 16% of the week. In the control group, Tuesday featured as the most prolific day (17%), followed by Sunday, Wednesday, and Thursday all at 15%. This indicates Sunday as a bullying outlier and a day to monitor for such discussions.

On Sundays, appearance (ugly ,fat), and intelligence insults (stupid, dumb, idiot, moron) were most common, whereas on Wednesdays appearance outweighed intelligence insults. The trend of intelligence insults decreasing continued until volumes spiked on Sunday. Further research could reveal whether trolls are more critical of intelligence at the beginning of the week and more focused on appearance towards the end of the week, as well as which factors may influence these trends.

Bullying: Hour of the Day

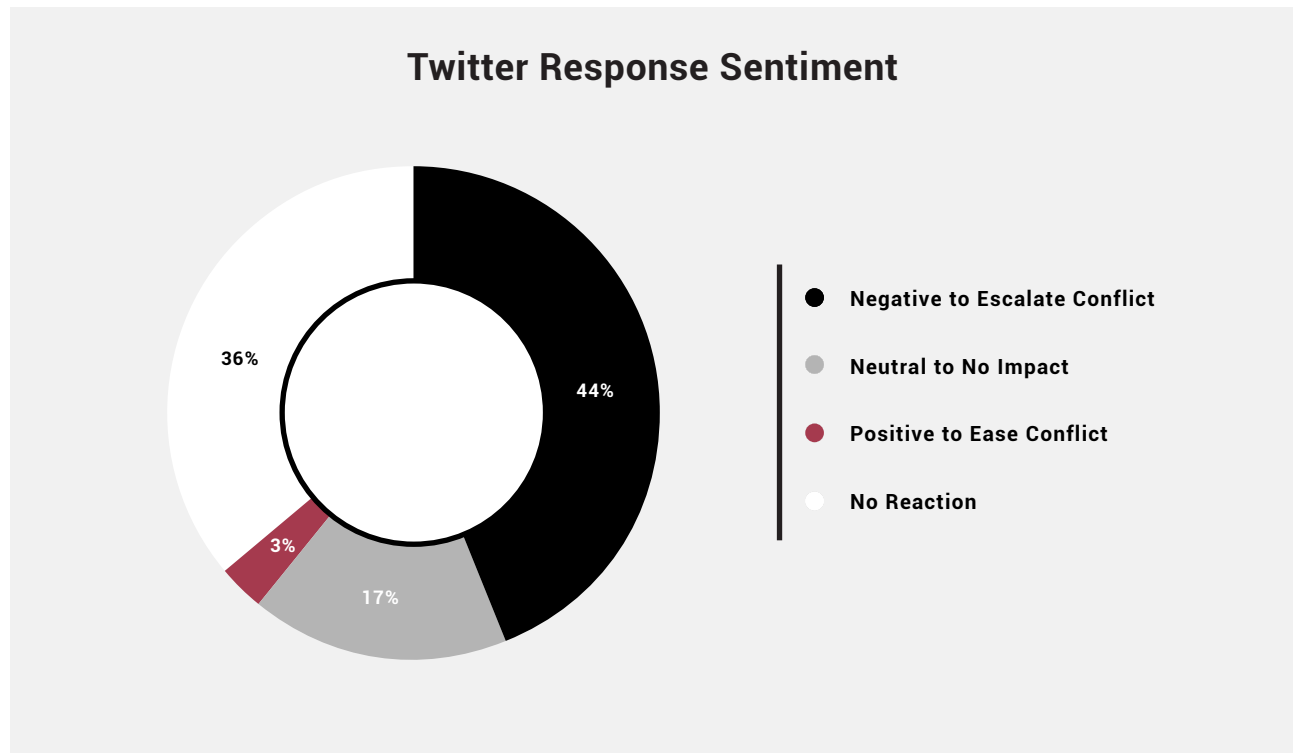


The chart above highlights the percentage of bullying activity throughout an average day, relative to a Twitter control group.

The groups shared one peak in volume, which occurred at 17:00/5:00 p.m., when bullying saw 8% of the day's volume and the control group held 7.5%. This featured as the largest peak in bullying conversation throughout the day, with content including homophobic, intelligence, and economic insults. Authors produced the fewest proportion of comments at 4 a.m., when insults focused predominantly on appearance (ugly) and intelligence (stupid).

Discrepancies between the two lines show that bullying was disproportionately common between six and eight in the evening and was less common around two in the afternoon, when there was an increase in general (non-bullying) activity.

Responding To Trolls Can Escalate Conflict



NEUTRAL OUTCOME 

"Deluded"

"Not at all. Just blinded by ignorant hatred and racist bigotry that brings shame on Hull."

"I'm proud of what I am"

"...a racist, homophobic sexist idiot?"

NEGATIVE OUTCOME 

"LOL you dumb b-----"

"Happiness is all that matters in life sweet pea..."

"so why are you talking to me????"

The chart above shows reactions to responses from bullying recipients. The examples feature represent archetypal bullying exchanges, with the inciting bullying shown first. As shown, negative reactions to responses from bullied individuals were the most likely, reflecting a tendency toward continuing or escalating the conflict. There is a scant likelihood of positive outcomes when engaging with trolls on Twitter.

Note that conversations categorized with 'positive reaction' do not necessarily entail positive reactions from the bullying author, but instead include positive outcomes for the recipient, including support received from other users.

Twitter's **Help Center** recommends several options for bullied individuals, including when to report abuse. These options include first unfollowing, then blocking, the user. According to the article, "Abusive users often lose interest once they realize you will not respond," recommending a "don't engage" policy. Further to this point, 36% of the 'no reaction' conversations were examples of non-engagement by bullied users.

Outcomes When Recipients Respond to Bullying

NEUTRAL REACTION



"Re, belief in gods. I have none. Atheism is a NON-belief. Are you really that stupid or are you trolling?"

"Why do you hate me so much. I'm asking you a few simple questions. There you go calling me this and that..Come on be responsible"

"I don't hate you, btw"

NO REACTION



"and this is based on WHAT, oh Trump and ur a chump 4 buying n 2 the BS!"

"Based on everything he says when I watch him. I'm a chump? Well YOU are blocked."

NEGATIVE REACTION



"Blacks r just as racist as any white racist they will have u try to feel guilty about #fuckem"

"I am very disappointed mate thought you were better then those comments you made. You can fuck off. #unfollowed #blocked"

"ur dumb, nothing I said was racist u complete spastic"

POSITIVE REACTION



"Proving your a fool. And a racist"

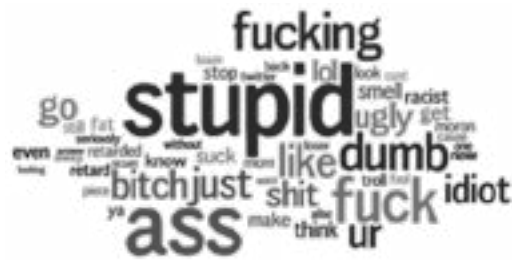
"I may be a fool – however definitely – not a racist – as a REPUBLICAN – in November – I'm voting for – HRC"

"For that I applaud you"

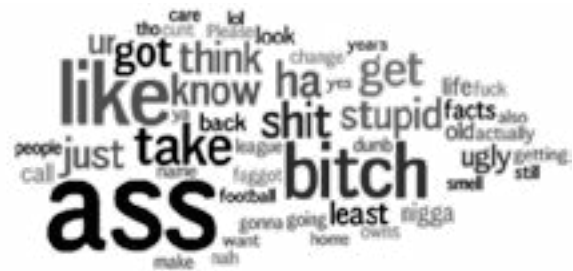
"we'll take you just this one time lol"

Lead-Up to Negative Reactions

Bullying Remark



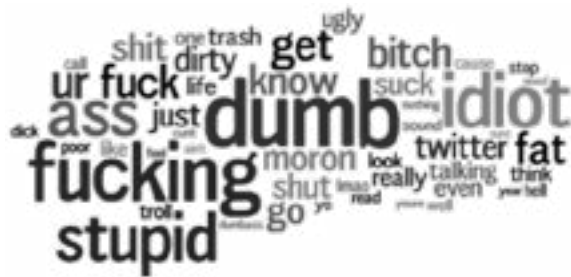
Response



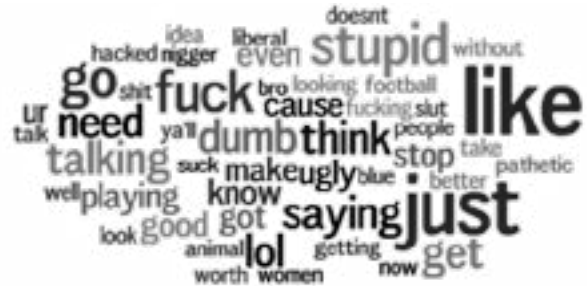
The above topic clouds represent the prominent phrases within conversations that ended with negative reactions from bullying authors. As shown, the two clouds reveal a striking similarity between the bullies and their victims, with several insults appearing prominently within both groups. Often, this bullying begets continued negativity and abuse. As these mentions were followed by further negative reaction and conversation, the finding suggests that engaging in abuse with similar language may simply perpetuate and escalate the conflict.

No Reactions

Bullying Remark



Response



The above topic clouds represent the prominent phrases within conversations that ended with no reaction from bullying authors. Although several negative phrases, including **dumb** and **ugly**, are prominent within the responses, they are far less likely to appear within responses by bullied individuals. This is understood when compared against the prominence of these outright negative terms in responses that prompted further negative reactions (see previous chart). The implication is that, if responding to bullying on Twitter, less offensive language is less likely to escalate the conflict.

Appendix

Hate speech summary

Insult prevalence		Homophobia		Transphobia		Masculinity Construct		Misogyny		
Professions	Interest	Professions	Interest	Professions	Interest	Professions	Interest	Professions	Interest	
Religious staff	Home & Gardening	Religious staff	Home & Gardening	Politician	Shopping	Religious staff	Home & Garden	Religious staff	Home & Garden	
Emergency worker	Shopping	Emergency worker	Shopping	Sales/ Maketing	Travel	Emergency worker	Shopping	Emergency worker	Shopping	
Politician	Automotive	Politician	Travel	Legal	Environment	Politician	Environment	Politician	Environment	
Software developer	Fashion	Sales/ Marketing	Environment	Sportsperson	Automotive	Legal	Travel	Legal	Travel	
Sales/ Marketing	Travel	Legal	Fashion	Software developer	Fashion	Sotware developer	Automative	Sales/ Marketing	Automotive	
Legal	Environment	Software developer	Automative		Photo & Video	Sales/ Marketing	Fashion	Software developer	Fashion	
	Photo & Video	Scientist & Researcher	Science		Science	Scientist & Researcher	Photo & Video	Scientist & Researcher	Photo & Video	
	Movies	Health practitioner	Photo & Video		Movies		Science	Health practitioner	Science	
	Science	Sportsperson	Movies		Business		Movies		Movies	
	Games		Technology		Fine arts		Technology		Fine arts	
	TV		Fine arts		TV		TV		Technology	
	Fine arts		Business		Technology		Fine arts		Business	
	Technology		Beauty/Health		Beauty/Health		Games		Beauty/Health	
	Beauty/Health & Fitness		TV				Business			
	Animals & Pets		Games							
				</						

About Brandwatch

Brandwatch is the world's leading social intelligence company. Brandwatch Analytics and Vizia products fuel smarter decision making around the world.

The Brandwatch Analytics platform gathers millions of online conversations every day and provides users with the tools to analyze them, empowering the world's most admired brands and agencies to make insightful, data-driven business decisions. Vizia distributes visually-engaging insights to the physical places where the action happens.

The Brandwatch platform is used by over 1,200 brands and agencies, including Unilever, Cisco, Whirlpool, British Airways, Heineken, Walmart and Dell. Brandwatch continues on its impressive business trajectory, recently named a global leader in enterprise social listening platforms by the latest reports from several independent research firms. Increasing its worldwide presence, the company has offices around the world including Brighton, New York, San Francisco, Berlin, Stuttgart, Paris and Singapore.

Brandwatch. Now You Know.

www.brandwatch.com | @Brandwatch

About Ditch the Label

We are one of the largest and most ambitious anti-bullying charities in the world. We are defiant, innovative and most importantly, proud to be different. Our mission is to reduce the effect and prominence of bullying internationally.

No more disempowerment. No more prejudice. No more bullying.

Each week, we provide award-winning support to thousands of young people aged 12-25, primarily through our website and digital partnerships. We also work with schools, colleges, parents/guardians, young people and other youth organisations. Innovation is at the core of all that we do and we believe that we can, and will beat bullying.

We commission and utilise research reports, like this one, to better understand the changing nature and climate of bullying and discrimination. This continuous learning process feeds directly into the improvement and evolution of our support programs which helps not only those who are being bullied, but those who are doing the bullying too.

Find out more at www.DitchtheLabel.org

Contact

Our address

15-17 Middle St, Brighton, BN1 1AL

Email

hello@DitchtheLabel.org

Web

www.DitchtheLabel.org

Tweet

[@DitchtheLabel](https://twitter.com/DitchtheLabel)

Brandwatch Research Services is a team of dedicated analysts based in Brighton, Berlin and New York, helping our customers better understand and use social data. For more research examples visit our report library www.brandwatch.com/reports