ANDY RUDDOCK
EXPLORING MEDIA RESEARCH

THEORIES, PRACTICE, AND PURPOSE

13

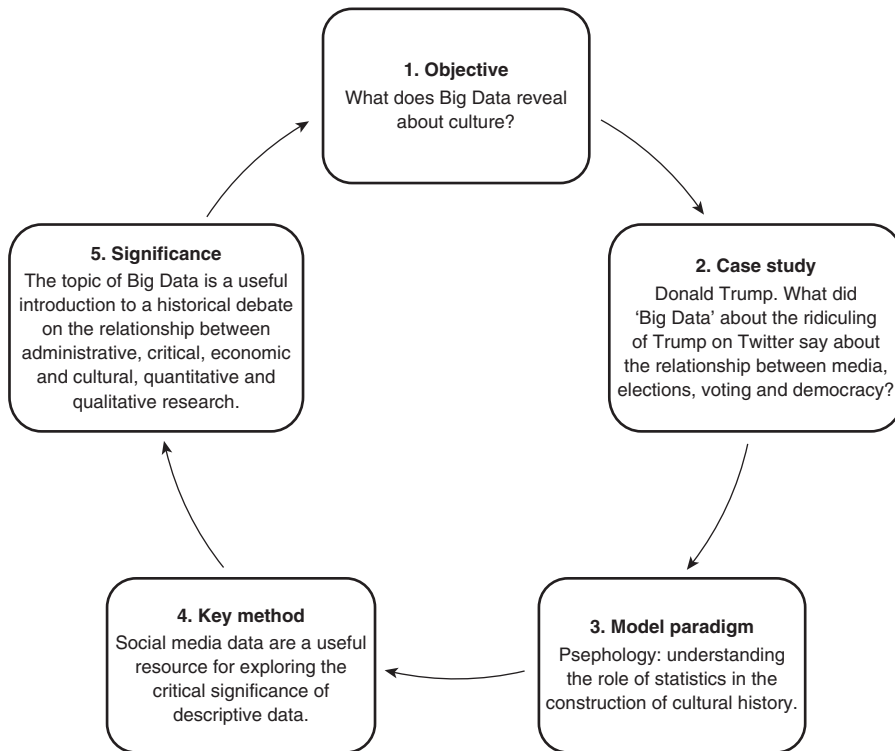# BIG DATA: HOW CAN WE USE IT?

Figure 9   Using Big Data

OBJECTIVE: UNDERSTANDING HOW TO USE SOCIAL MEDIA DATA IN A CRITICAL FASHION

Big Data refers to 'continuous gathering and analysis of dynamically collected, individual-level data about what people are, do and say' (Couldry and Powell, 2014, 1). This chapter outlines the complexities of using Big Data in research by asking what descriptive, quantitative data really tell us about media culture. It also shows how these conceptual issues affect practical challenges in using social media data sets in qualitative data assessment (QDA) software. In particular, the chapter explains how to work through conceptual and methodological problems to find ways to develop valuable critical insights into the impact of social media on political participation from descriptive statistics.

The discussion is grounded in research on Twitter's role in Donald Trump's 2015–2016 campaign to win the US Republican Party's nomination as its candidate for the US Presidency. The case study has been chosen for the following reasons. The debate about what one can and cannot do with Big Data reflects the bigger question about how much information, especially qualitative information, tells us about how society works. Psephology – the study of voting behaviour – has played a leading role in this conversation. Current discussions about the difficulties of understanding voting preferences with Big Data echoes many conceptual and methodological issues that psephology raised. In 2015–2016 such discussions centred on Donald Trump, a political candidate whose appeal perplexed not only American but global media audiences. Social media played an enormous role in this international media drama. Trump was a prolific Twitter user, which exposed him to a great deal of voter feedback through tweets. As such, his candidacy provided an opportunity to assess how much social media data could contribute to critical understanding of modern politics. This is all the more so since, as a figure whose public profile was built in part through a reality television career, Trump's campaign was a landmark in the convergence between political communication and entertainment. Eminent psephologist Sir David Butler thought the transformation of elections into media dramas was the defining political 'fact' of the post-Second World War world. Trump's eventual election as President of the United States demonstrated two things: the prescience of these observations and the necessity of considering why most political experts failed to predict his success, despite access to sophisticated polling mechanisms and banks of data. In hindsight, the role that Twitter played in *obscuring* Trump's popular appeal, during the primaries, was a portent of things to come. It was an illustration of how Big Data can be seductively misleading because of its narrative flexibility.

So, Big Data provokes big questions about methods and mediatized politics. The chapter identifies and answers these puzzles as follows. First, it examines the public furore around Donald Trump's primary electoral campaign and Twitter's role in building his controversial persona. Second, it links confusion about the impact of

Twitter on Trump with academic observations over the ambiguous status of Big Data as a resource. Third, debates about Big Data and elections are connected to psephology, the science of voting behaviours. Here the work of Big Data analyst Nate Silver is compared with the career of Sir David Butler, the Oxford don who is famed as the founder of this field. The comparison shows that many questions about Big Data recycle earlier discussions on the difficulties of predicting voting action. Finally, the chapter translates scholarly observations about evidence, Big Data and mediated elections into practical steps for using Twitter data to pinpoint trends in media politics.

The key points of this chapter are that:

- Big Data concerns the value of numbers in understanding culture.
- This issue has taken on particular importance given the increasingly influential role of media in electoral politics.
- Big Data reveal more about the biases of media and information industries than the feelings and thoughts of media users.

CASE STUDY: BIG DATA, TWITTER AND THE 2015–2016 US PRESIDENTIAL ELECTION

Public angle: What is Big Data and why does it matter?

In 2015, a group of political scientists wrote a chilling *Washington Post* article on how Big Data corrupted American politics (McDonald, Licari and Merivaki, 2015). New methods for analysing voter registration records afforded granular understanding of the audiences for political messages, such that those messages could be tailored to suit a panoply of interests. Trouble was, some states charged for access to the data. Ergo, Big Data had turned elections into a game for rich people who could find out how to say the right things to the right people.

At one point during the US primaries of the same year, it seemed Donald Trump hadn't quite taken the point. That September, the celebrity mogul exposed himself to the popular scrutiny of Twitter, inviting the public to ask him anything they liked via the #Asktrump hashtag. Within hours media outlets like *Time* (White, 2015) and *The Huffington Post* (Satran, 2015) ridiculed the stunt as a débâcle. An avalanche of abuse crashed down – payback, it seemed, for a candidate for whom vitriol was a signature campaign move. Trump had called Mexican immigrants rapists and murderers, promised to build a wall along the US–Mexican border, and advocated a blanket ban on all Muslims entering the US. Twitter delivered a public backlash. Some of the more amusing/trenchant tweets were reported ('So a Muslim, a woman, and a Mexican walk in a bar,

who's rights do you alienate first? #AskTrump'; Satran, 2015, no page). Trump was likened to an over-ripe vegetable. Twitter users had, it seemed, done far better than professional politicians or journalists to befuddle Trump's populist appeal. Here was the popular power of social media. Twitter's ability to capture historical snapshots in downloadable form also seemed to promise that if ordinary people can lay the mighty low, so too a wider spectrum of researchers could analyse how such feats worked.

Except Trump wasn't brought down by the incident; if anything, his social media outbursts became more impudent. Certainly, the incident showed what some people thought about the man. It also made the point that the last half-century has seen a media-driven shift in the nature of political communication, where today much campaigning is 'reactive' and responds to unexpected challenges thrown at candidates (McNair, 2006). Ed Miliband's disastrous tilt at the UK premiership in 2014 came at the end of a campaign where much was made of a photo of him struggling to eat a sandwich. But what's the evidence that these moments really matter, and if so, why do they matter? Especially given that Trump won. At the very least, one can say that the data generated around #AskTrump in fact said very little about his appeal. So what did it signify? The political biases of Twitter users? The entertainment value of electoral politics? The role of social media in cultivating smugness? The only thing we can be confident about is that the Trump phenomenon was an opportunity to reflect on the kind of knowledge that Big Data generates.

Scholarly context: Big Data

Across the humanities, resistance to quantitative methods reacts to the notion that the only knowledge worth having is that which comes from the value-free measurement of tangibles – like a person's decision to vote for a certain candidate (Kolakowski, 1992; Fay, 1993). This theme is picked up in the Big Data literature. Squabbles about what Big Data can and can't tell us reflect cornerstone enigmas in social research: 'puzzles about the locus and nature of human life, the nature of interpretation, the categorical constructions of individual entities and agents, the nature and relevance of contexts and temporalities, and the determinations of causality' (Wagner-Pacific, Mohr and Breiger, 2015, 1). So, Big Data is the latest instance of a familiar gambit, where quantitative methods are presented as value-free tools when their effective use depends on the ability to explain why they are nothing of the sort. The apparently automatic algorithms that capture data emerge from thoroughly human processes of scientific development utilizing the training and experience of people who apply their expertise to technical problems in specific ways (Wagner-Pacific et al., 2015).

What's new about Big Data?

Big Data has reignited discussions on the validity of statistics as a mode of cultural research. What is new is the ferocity of these debates, and the level of

public interest that they elicit. David Beer (2015, 1) observed that 'general reliance on statistics and probability calculations ... have suddenly become a central part of culture'. Critics think Big Data has produced a 'change' where numbers become more effective because their use value doesn't have much to do with representing what people are really like. Big Data is attractive to media and businesses because it dispenses with the obligation to engage with viewers, readers and listeners on their own terms. For all the rhetoric of wanting to please the public, audiences are inconvenient realities for media companies (Bratich, 2005). Whether it's to attract advertising revenue or defend public subsidy on the basis of public service, media organizations have historically pursued more sophisticated methods to enumerate audience quantity and quality (Ang, 1991; Balnaves and O'Regan, 2010; Balnaves, O'Regan and Goldsmith, 2013). One way or another, commercial newspapers, radio and television interests used numbers to tell advertisers plausible stories about who was paying attention to their content, and this shackled numbers to representative obligations (Couldry and Powell, 2014). Social media have changed the game. Instagram, Twitter, Facebook, Snapchat, etc. still have to sell evidence about the size and spending power of their customers, but the data comes from those customers in the course of their usage. Social media don't have to 'ask the audience' what they are doing because their users *must* provide such data to be users at all.

Consequently, Big Data represents a profoundly changed mode of communication, expression and social existence. According to Couldry and Powell (2014), Big Data threatens the idea of common culture by altering media ecology. In the most dystopian of views, the technology and business models of mobile media oblige users to provide a constant stream of data about their habits and preferences, attenuated from notions of identity or expression, thus removing meaning and culture from the communication equation.

Worse still, Big Data's potential value is inaccessible to most media researchers. Cracking it can require collaboration with computer scientists, and it is often hard to connect their technical expertise to the purpose of cultural analysis (Shaw, 2015). Unsurprisingly, some conclude that Big Data has reinforced a new 'knowledge divide' (boyd and Crawford, 2012). Taking all of this into account, social media can be extremely poor research resources. Twitter is a case in point. Sure, it's great at archiving the quotidian and not-so-ordinary thoughts that pass through public conversation. Regrettably, the platform only makes a small sample of its tweets available. This data is, after all, the revenue stream of the platform, meaning that 'tracking the ongoing public activities of more than 5000 Twitter users at a time is now only possible by working with Twitter's licenced third-party API provider Gnip, at a cost well beyond the funding available to most research projects' (Bruns and Burgess, 2012, 4).

Lewis, Zamith and Hermida's (2013) research on Twitter's role in the 2011 Arab Spring is an interesting case study of how these tendencies turn into real strangleholds on knowledge. The researchers wanted to examine 'how social

media, and more specifically Twitter, can offer a platform for the co-construction of news by journalists and the public', using the tweets of Andy Carvin. Carvin, who worked for America's National Public Radio, was a prolific tweeter credited with playing an important role in connecting the protestors with the outside world – a symbol, if you like, of Twitter's democratic potential. There was nothing democratic, however, about accessing data. The authors used several different kinds of software to harvest Carvin's tweets, but only gained access to the full data set when it was given to them by the man himself. More to the point, Carvin only had them because Twitter gave him the file, which belonged to them.

Even if you can access the 'firehose' and tools to make sense of it, it isn't clear the data are worth the effort. Media researchers seek rich explanations for how people live with media. Even researchers who do appreciate the value of studying Twitter as a new style of open political discourse admit that claims have their limits. Bruns and Burgess (2012), for example, qualified their study of the #ausvotes hashtag as a distinct political narrative in the Australian Federal Election of 2011 with the recognition that this conversation circulated among a minority of tech savvy, affluent 'political news junkies', where caustic comments directed at campaign messages 'can be understood as a mildly subversive, if largely inconsequential, form of speaking truth to power' (2012, 395).

For critics, then, Twitter is at the cusp of the argument that there are good reasons to ask whether Big Data does more harm than good when it comes to understanding the social world. In fact, there is almost an architectural desire to make that world disappear into bundles of data. Consequently, a good deal of effort has gone into detailing how Big Data builds barriers to knowledge: Twitter frequently features in exemplification. Twitter's capacity to obscure political conversations, and Donald Trump's prolific use of the platform, thus made the entrepreneur's political ambitions a significant historical moment in research on media, elections and democracy.

Big Data, Donald Trump and Nate Silver

This is all the more so given how one of Big Data's leading luminaries, Nate Silver, used Trump to explain the limitations of statistics in understanding voters. Nate Silver's '538' blog offered interesting reasons for why scientific polls using sophisticated surveying methods produced at best mystifying, and at worst downright misleading, accounts of Trump's appeal. Silver won renown as a precociously gifted statistician with almost supernatural predictive ability. Counterintuitively, he spent a lot of time bursting the bubble around statistical modelling methods. In *The Signal and the Noise* (2012), Silver warned that we are not enjoying an era of better predictions of everything from the weather to voting, and Trump's early 2016 standings in news polls were a case in point. More or less across the board, polls showed Trump leading the Republican race for months. However, primary campaign polls were historically bad at calling

the winner. In these kinds of contests, popularity was one thing, voting was another. This was likely to be especially true of Trump, given the different reasons being propounded for his success. Was he a lightning rod for resentment against the political establishment? A symptom of a 'power failure' within the Republican leadership? The benefactor of wildly disproportionate media coverage? Whatever the answer, the lesson was that polls did a poor job of explaining America's political mood. Did Trump's poll success represent genuine popularity? And what had brought that about? Both questions remained mysteries (Silver, 2016a). Indeed, Silver fell victim to his own prediction. After months of dismissing Trump's nomination chances, the latter-day Nostradamus confessed The Donald's victory proved the limitations of his methods (Silver, 2016b).

Silver's criticisms represented profoundly important perspectives on the use of numbers to understand society through election case studies. Throughout the 20th century, scholars had used statistics to produce forensic explanation of voting behaviours, but were aware of two important qualifications in this pursuit. The first was that media were driving significant changes in what elections represented as events in cultural history. The second was that the use of numbers in political science was itself a form of cultural history, where the role of human interpretation and the biases in data collection, storage and use were factors to be considered at the analytical stage. No one claims that numbers provide an objective picture of what happens in elections. Nate Silver became important as a figure who applied the same argument to a new era of numbers with apparently impeccable objectivity credentials – the era of Big Data.

Silver's observations about Trump connect current debates on Big Data to a longer tradition where elections have been used to make sense of the power of numbers as tools for cultural historians. It is, in many respects, a variant on quarrels over the relative merits of quantitative and qualitative methods. It also relates to the question of how media affect political practice, where elections can be understood as events that either drive or exhibit the forces of mediatization in action. Finally, although Big Data does represent different modes of generating and analysing quantitative data about how culture works and how people behave, this knowledge is best understood through critical research principles that have been at play in media studies for several decades. In particular, the idea that media industries affect knowledge through the structure of their message systems is now an insight that reaches more visibly into the research process itself, given the deliberate space between the amount of data generated about media users and the amount of that data that becomes available to public scrutiny. In this respect, Big Data research is characterized by a remarkably clear view that any method of any value must be informed by well-developed theory. To put this more simply, the new Big Data era, and people like Silver, underline the counterintuitive position that knowledge begins with the detailed analysis of what isn't and can't be known, given the tools available to most analysts most of the time. To begin to understand, let's go back to post-war Britain.

KEY PARADIGM: PSEPHOLOGY

The birth of psephology

At the end of the Second World War, a young British Army lieutenant called David Butler returned to his studies in political science at Nuffield College, Oxford. Back at his desk, he commenced using statistics to provide sophisticated, multidimensional explanations for British voting patterns. He believed that new data methods could revolutionize understanding of how Britain voted (Butler, 1998). Butler's work came to the attention of none other than Winston Churchill. In 1950, Britain's most experienced parliamentarian asked the 25-year-old political scientist to explain the nation's post-war mood. Soon, Butler was recognized as the founder of psephology – the study of electoral facts (Ranney, 1976).

Butler was the Nate Silver of his day. His knack for analysis established him as the driving force behind the Nuffield election studies, 'one of post-war political science's most massive enterprises' (Ranney, 1976, 217). These studies became the authoritative accounts of British political facts. Part of their success was attributed to Butler's scepticism about statistical truths. Early on, Butler advised that a detailed understanding of why people voted as they did in the past did not translate into an ability to predict how they would do so in the future (see Bale, 2012). He regarded election research as an inherently 'messy' business, given that it involved the complicated intersection of cultural and institutional life. The combination of statistical skill combined with a critical awareness of the limitations of his art placed Butler at the centre of British television election coverage for 60 years. Retiring from media work in 2015, Butler did his best to avoid predicting the outcome of the British election of that year (Wallop, 2015). The question missed the point of his life's work.

Psephology and Big Data

The Butler/Silver analogy ties the new interests in Big Data to key themes in studies of media and power. Butler and Silver shared the view that statistics offer valuable tools for building intricate cultural histories. Silver was unashamedly 'pro science and pro-technology', but added 'numbers have no way of speaking for themselves. ... We imbue them with meaning ... we may construe them in self-serving ways that are detached from their reality' (Silver, 2012, 8). Butler was similarly smitten with the new explanatory power of post-war social statistics. The post-war provision of social housing in the UK, for example, provided a wealth of unprecedented data on lifestyle gradations within the working classes that could break voting patterns down into finer classifications (Butler, 1955). Yet the key to locking this potential lay in subordinating numbers to the broader dictates of cultural history:

> I have ... wondered at the complex of family history and neighbourhood pressure, of house ownership and class status, that has determined their

> vote. ... I have always been conscious of the gaps in my knowledge of British economic geography and local history. (Butler, 1983, 15)

Butler wanted to dismiss electoral 'myths', by which he meant contestable simplifications of voting behaviours, but warned there was more to the project than making sophisticated statistical associations between voting and demographic factors. To take such a view was to lose sight of elections as historical snapshots. Butler feared that the very science he was pioneering ran the risk of writing self-defeating cultural histories with no people in them.

A second sentiment shared by Silver and Butler was that media had changed elections, complicating data analysis in the process. Silver complained that television encouraged data analysis that served dramatic effect more than edification. This had more to do with the transition of politics into a form of media entertainment than it did with the improved powers of polls (Silver, 2012). Asked to define landmarks in post-war electoral politics, Butler named the arrival of television and the rise of public relations. Butler's interest in PR presaged many of the themes that were to characterize Silver's intervention in the Big Data arena. The former had been struck by the 1997 Labour Party's innovative understanding of how detailed surveillance of the media scene, accompanied with data from qualitative and quantitative voter research, produced a nimble campaign where opposition messages were rapidly countered with crowd-pleasing counter-arguments. Labour's spectacular rebranding in the public mind was produced by Big Data and small data. It used qualitative data to listen to audiences, in focus groups, but it also kept a forensic eye on opposition press coverage and learned to rebut criticisms as soon as they were made. The main effect of this, however, has been to make performative politics the supreme measure of campaigning.

Butler believed media had changed elections as historical events. He considered elections as events that made the mechanics of politics more visible by placing them 'under strain'. By the late 20th century, media had become central to campaigning practice, and elections displayed the media's political role in full flow. During elections, societies consciously put their full sense-making capacities to political use, and media had become key ingredients in this process (Butler, 1998). Consequently, the study of elections was more about making sense of this change, rather than simply predicting winners and losers.

So Butler's psephology defined elections as political, historical and cultural rituals, wherein the scholar's role was to explore the complexity of voting behaviours. In a similar vein, Silver argued that media harnessed sophisticated data-gathering methods to manufacture the very myths that Butler abhorred. Silver feared that the spectacular nature of media elections transformed data into something that was used to tell stories, rather than the truth. So the new topic of Big Data reflects old ideas about the role of media in the production of knowledge, and the role of media studies as a topic that plays close attention to how political truth is created and shared in society. Silver used Trump as an exemplar. There's a world of difference between saying that Donald Trump was ahead in the polls and concluding that he was a person that a sizeable portion of

American voters actually liked. In the early 21st century, this was a media matter since poll stories are a staple of political news.

When we add into the mix how media users contributed to the data explosion through social media reactions to candidates, and how these reactions became part of the media story about how audiences read elections, as it did in #Asktrump, then we are left with an intriguing convergence between psephology and Big Data. The staging of elections as media events has affected the use of statistics in understanding voters. We might say, for example, that Silver's comments about the mystery of Trump's popularity were an example of the myth making that Butler deplored. At any rate, there are dangers at play in how media and statistics collude to conjure impressions of media reality, and the provision of Big Data by social media users adds to the confusion, not least because it conflates the identities of citizen, audience and participant. But the point is not to dismiss the value of Big Data; it is to appreciate how knowing what we can say with it starts by thinking about what it can't do.

KEY METHOD: CAPTURING AND ANALYSING TWEETS

Fortunately, the alliance between Big Data and QDA software makes it pretty easy to see how these conceptual observations transfer into research practice. NVivo, for example, easily seduces with its power to automatically capture, sort and analyse social media data with the tick of a box. Evidently, the lesson so far is that there's no point in producing word clouds, charts and the like unless we think carefully about what this information says. To see how this works in practice, let's consider what following NVivo guidelines for analysing social media content tells us about the media relationship that Trump established with the rest of the world.

Choosing a case study

The first step in using Big Data is to think about what to look at. Given the plethora of evidence, it is important to get the story straight on how a particular data set relating to a particular event has been selected. When it comes to elections, if we follow Butler's advice then the task is to choose evocative historical snapshots of the cultural mood at a significant moment. With that in mind, the data set we are going to work on is the reaction to a tweet that Trump sent shortly after it became clear that he had won the race as Republican nominee.

On March 5, 2016 Trump tweeted a picture of himself eating a taco alongside the legend 'Happy #CincoDeMayo! The best taco bowls are made in Trump Tower Grill. I love Hispanics!'. The tweet met a wave of criticism as a crass effort to win over Hispanic voters, given his racist attack on Mexico. The message seemed to glory in ignorance. Far from seeking forgiveness, Trump revelled in his victory by promoting his commercial interests, appropriating the very culture that he had gone out of his way to malign (Parker, 2016).

So, here was an incident where it was possible to see what social media users thought about Trump at a key moment when criticisms about his campaign came to a head in public. Trump's victory in the face of unapologetic racism evoked the possibility that his campaign had worked beyond expectations because of his bigotry. The decision to celebrate a political milestone by basking in racism seemed to support this view. But what would Twitter users make of it, and what could we say about this political moment by capturing tweets?

Capturing and making sense of tweets

Having selected a plausible case study, the next step is to figure out how to load it into NVivo. Here, we encounter a series of good news/bad news scenarios. Positively, there are a range of tutorials on YouTube that will teach you how to catch tweets, then load them into the software. Negatively, two problems immediately emerge. The first is that we discover that the PC version of the software has more functionality than the one available to Mac users. The second is that whatever the platform, Twitter controls the amount of tweets that it is possible to capture, so at base one can only view about 1% of the chatter about the event, and there is no way of telling how these tweets are selected. So, the very first practical issue one faces isn't figuring out the technology, but figuring out the point of engaging with research so confined by the proprietorial interests of media, software and hardware companies.

Here, the strategy is to go for the PC option, based on techniques outlined in NVivo's tutorials (QSR International, 2015). Following instructions, this study did a Twitter search for the hashtag #trumptaco, captured those tweets via a plug-in called NCapture, then loaded the captured tweets as a data set into NVivo for PC. This process captured 11,099 tweets. The next task was to infuse this data with meaning. The key lies in using descriptive statistics to ask questions as well as find answers. Once we understand that, the simplest features of NVivo's descriptive analysis in fact represent useful tools for addressing the limitations of mediated democracy and our ability to fathom how it works.

Critically visualizing descriptive data

Remember, the idea is to critically evaluate evidence by placing it in a theoretical context. So the analysis here confines itself to considering the complexity of descriptive data. The process described above gathers what looks like an impressive data set, ordered to produce instant information, and provides an architecture for further study. Captured tweets identify the tweeter and his or her location. Locations are mapped, meaning that we can either look at global pictures or drill down into specific localities, or indeed take a comparative look at reactions in different places. We can start all of this by simply hitting a 'map' tab in NVivo. The program also 'clusters' bundles of linguistically similar tweets.

NVivo produces tree diagrams that map together Twitter user names according to similar content. Other information includes the biographies attached to each Twitter handle, and of course the tweets themselves.

There's no doubt that this gathering and sorting function is useful, but it's just as clear that the harvested and sorted data mean nothing without critical reflection. There's no way of knowing why these tweets in particular were selected (QSR International, 2015). Also, if you try repeating this exercise over a number of days, you'll get different results. Try a Twitter search ncapture for 'Trump' and 'taco' right now, and you'll see what I mean. The 'wow' factor of the map also dissolves once you realize that Twitter users can edit their location if they want to (QSR International, 2015).

So we're faced with a filtering process that we can't control, which makes it incumbent on us to consider how the choices that we do make, as researchers, might aggravate this problem. Let's go back to the beginning of the chapter, and the fact that political journalism about Trump-related Big Data probably told us very little about what was going on in the Republican primaries. Here, again, it would be possible to research how people attacked Trump's chutzpah. We could do this by using a word search. In this case, I ran a search for the 50 most used words in the sample, setting the minimum word length to four letters (to suit taco; anything less would have missed out a key term). 'Racism' occurs 828 times, with 'racist' making a further 506 appearances. So we could discuss how Trump's tweet made the issue of racism visible. Except we'd quickly notice, looking at these tweets, that it's very difficult to assess how racism entered the public vernacular. Almost 600 of those references are retweets of a CNN tweet about US Senator Elizabeth Warren accusing Trump of basing his campaign on pure racism. We might also note that 'racism' is 29th in the list of 50 most used words, and along with 'racist' is one of only two key words that gives any hint at any real political content. We are reminded that we are dealing with algorithms, not content, here by the fact that numbers one, two and three in the list are 'https', 'trump' and 'taco'.

It's tempting to say that descriptive data gained from the combination of Twitter and NVivo doesn't tell us much about anything other than what happens in Twitter and NVivo. Unless, of course, we want to think about how Big Data contributes to the transformation of mediatized politics into an exercise in enter-taining myth making – the critical line of inquiry introduced by Butler and continued by Silver. If that is the case, then simple descriptive Big Data from Twitter demonstrates why Big Data makes it harder to figure out what people make of politics.

Something interesting happens with this data set if we use the chart func-tion, tabulating the Twitter users and hashtags that dominated the conversation. Producing a three-dimensional chart produces a picture that can be related to critical views on the alienating aspects of electoral politics (see Figure 10).

The immediate impression is telling. @the realdonaldtrump stands like a skyscraper surveying the rest of the Twitterverse. For all the media contro-versy about the tweet, the main thing you could capture from Twitter in the
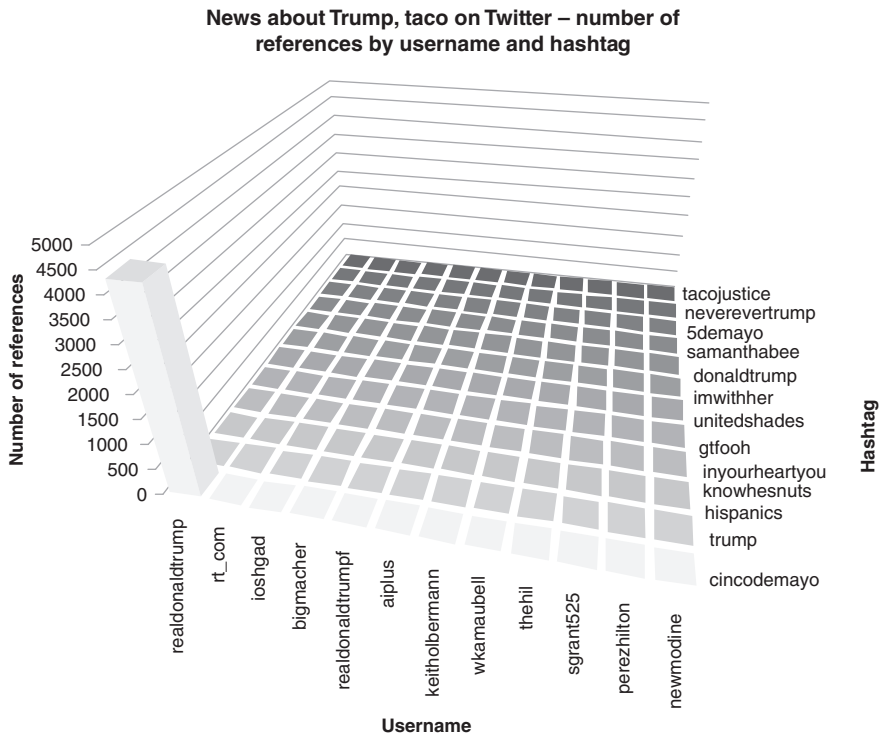
**News about Trump, taco on Twitter – number of
references by username and hashtag**

Figure 10   Tracking Trump tweets

days afterwards was the voice of the man himself – his unapologetic micro-blogged advertorial accounted for almost 50% of the traffic captured. To be sure, the hashtags signal dissent (#inyourheartyouknowhesnuts, #never-trump). But looking at the top sources, it's hard to see a connection with popular resentment. Other listed tweeters included comedians Josh Gad and Jeff Dwoskin, broadcasters W. Kamau Bell and Keith Olbermann, news source *The Hill*, and celebrity blogger Perez Hilton. So, the evidence tells us this: if you used Twitter to figure out the political mood in America at the time when the world realized that Donald Trump would be the Republican's official can-didate for President, then what you would have seen is a public conversation dominated by the man himself, where the most visible opposition came from other parts of the media sector, including significant contributions from the entertainment world.

In spite all of the reservations about what Big Data really says, and the admitted limitations of methods for importing and analysing this evidence, this is a pretty interesting finding that can ground some useful questions. The first is that, for all the advent of social media, televised people's forums and the like, political communication remains an elite game dominated by the logic of media businesses that want to enrapture audiences. If you looked at Twitter

223

four days after the tweet, what you saw was a conversation dominated by Trump and, to a lesser degree, other broadcasters and entertainers. It is interesting that Keith Olbermann was one of them. Four years earlier, the famous American news anchor had caused a stir by criticizing the prevalence of aggressive hate speech in American political discourse. Olbermann feared that the effect of entertaining political journalism had turned politics into a gladiatorial sport, where the only way to respond to criticism was to destroy rather than debate the opposition. Trump's open taunting of his critics seemed to underline the significance of Olbermann's earlier observations. What the data did say, as mapped in Figure 10, was that whatever the event meant to the public, what it looked like when filtered through NVivo was an elite conversation/row between media elites. In that respect, this Big Data set about this event raises a question that complements the interest first developed in David Butler's psephology and continued in Nate Silver's work: that media have significant effects on politics. There are two possibilities here. The first is evidence that social media enhance the transformation of political communication into a form of entertainment.

The second is perhaps more important. This small project suggests a 'big' problem in the sense that it is the relationship between Twitter and NVivo that 'describes' this vision of entertaining politics. The experience here is one where the descriptive first impression of this controversy was that it was, above all else, a drama born of well-established media industries where news, theatrics and comedy coexist. In the context of critical debates about Big Data, it is important to consider the factors that indicate that this 'impression' is the product of the creation, harvesting and analysis of Big Data. In that sense, the fact that Big Data told us virtually nothing about how ordinary people felt about Trump's taco is really quite useful. If nothing else, it contributes to the conversation on why his victory was so surprising, and dismaying to so many: it was just too easy to write evidence-based stories about why it would never happen.

SIGNIFICANCE: BIG DATA ALERTS US TO THE TENSIONS BETWEEN DIFFERENT KINDS OF KNOWLEDGE

Big Data prompts considerable social unease. Our reliance on mobile and social media means that we generate rafts of personal information without meaning to or understanding where it goes and what it does. These technologies make an 'always on' world. We announce where we are, what we like and what we do to those interested in managing such things for political and/or economic reasons. Hence the public debates on cyber security and privacy. Clearly, Big Data has become a focus for research on the intersections of media technologies, media users and political power, and it is incumbent upon us, as media researchers, to consider what we can – and can't – do with it.

When making sense of what Big Data means for society and media research, one useful place to start is by thinking critically about a sense of optimism on social media and participatory culture. Twitter, for example, has been celebrated as a method that ordinary people can use to humiliate powerful people. When they do, they generate masses of information that can be easily downloaded into QDA software. This allows us to gather and analyse evidence about key media events, like elections, in completely new ways. The data, and the tools, are remarkably accessible: the data is public and the software is relatively easy to use to produce results. One of NVivo's key features, for example, is the capacity to download and analyse social media content from Twitter and Facebook. This capacity promises a unique insight into media events, as they happen. Not only can citizens do more, but researchers can say more about them.

The generation and capturing of Big Data has certainly changed what we can say about media users. However, figuring out what Big Data from social media says is incredibly difficult in practice – impossible, in fact, without critical insights. What looks like a flood of data is really a trickle that is controlled by the technologies of social media and analytical tools. In this sense, research using Big Data is subjected to a kind of media dependency that is itself unique. If researchers have never been so data rich, they have also, arguably, never been so tied to media message systems, broadly defined. According to some critics, Big Data is the latest version of a myth where systematic methods of gathering and ordering data reflect an objective system of data gathering and analysis that is unaffected by any kind of human bias (boyd and Crawford, 2012). The political consequences of this myth become more profound when it is tied to the commodification and gatekeeping power of the private media platforms that gather and selectively share Big Data.

That said, researchers who have revolutionized the applications of statistics to understanding elections have directed a considerable amount of time working through these issues, and their insights into what you can't claim for numbers and Big Data are a valuable position from which to generate critical insights into media politics that *leverage* the limitations of empirical analysis. When it comes to Twitter, and NVivo, bearing these points in mind makes simple descriptive analysis a useful means of demonstrating the transformation of electioneering into a form of media entertainment, and resisting the temptation to create an image of widespread popular participation in those politics by cherry-picking the odd satirical quote that the odd member of the public might direct at elite politicians. In this sense, this 'simple' use of Big Data poses a very big question for media and democracy. Does the ability to critique the political establishment in very public ways through social media disguise ongoing imbalances, where electoral politics remains an elite sport? And if so, what does this say about the degree of participation that happens in media cultures? These questions are further explored in the following chapters on policy and audiences.

CHAPTER SUMMARY

- Big Data reminds us that all forms of evidence about media influence have to be interpreted through well-established critical insights into the relative merits of quantitative and qualitative methods.

- Big Data is not simply evidence, but is the product of message systems. It is a form of media representation that has to be interpreted.

- Using simple descriptive methods is a useful way to demonstrate how Big Data works as a narrative about new modes of media power.

- Used this way, Big Data is also a useful tool to assess levels of popular participation in media culture.