Article:
Hill, RL orcid.org/0000-0003-0099-4116, Kennedy, H and Gerrard, Y (2016) Visualizing Junk: Big Data Visualizations and the need for Feminist Data Studies. Journal of Communication Inquiry, 40 (4). pp. 331-350. ISSN 0196-8599

https://doi.org/10.1177/0196859916666041

Visualizing Junk: Big Data Visualizations and the need for Feminist Data Studies

Rosemary Lucy Hill, University of Leeds

Helen Kennedy, University of Leeds

Ysabel Gerrard, University of Leeds

Abstract

The datafication of culture has led to an increase in the circulation of data visualizations. In their production, visualizers draw on historical antecedents which define what constitutes a good visualization. In their reception, audiences similarly draw on experiences with visualizations and other visual forms to categorize them as good or bad. Whilst there are often sound reasons for such assessments, the gendered dimensions of judgements of cultural artefacts like data visualizations cannot be ignored. In this paper, we highlight how definitions of visualizations as bad are sometimes gendered. In turn, this gendered derision is often entangled with legitimate criticisms of poor visualization execution, making it hard to see and so normalised. This, we argue, is a form of what Gill (2011) calls flexible sexism, and it is why there is a need not just for feminist critiques of big data, but for feminist data studies – that is, feminists doing big data and data visualization.

Keywords: big data, data visualization, feminist data studies, gender, flexible sexism

Introduction

Data are increasingly valued as having the power to explain our social world. As data become more and more ubiquitous, so do data visualizations (Beer & Burrows, 2013) – that is, the visual representation of data in simple graphs and charts or more complex, often interactive, forms. An important way that people get access to data is through visualizations circulating in the media, online and elsewhere. As Gitelman and Jackson put it, 'd*ata are mobilized graphically'* (2013, p. 13, emphasis in original). Thus visualization as a field is professionalising, drawing on a number of historical antecedents which define conventions and what constitutes a good – or bad – visualization (Kennedy, Hill, Aiello, & Allen, 2016). Whilst there are usually sound design reasons for classifying visualizations as good or bad, the socio-cultural context in which visualizations are judged – from professional and non-professional perspectives – cannot be ignored (Kennedy, Hill, Allen, & Kirk, submitted). This means turning attention to the messy ways in which socio-cultural phenomena like sexism, classism and ageism can unconsciously shape judgements about data visualizations.

In this paper, we focus on the role of gender – and its intersections with class and age – in response to a data visualization called 'The Clicks Don't Lie' (see Figure 1), which appeared in the free, UK tabloid newspaper, The Metro, in July 2014. It shows data about the social media followers of pop musicians Rihanna and Shakira and was one of eight visualizations shown to focus groups on the X research project. The project explored how people engage with data visualizations and the factors that affect engagement. This particular visualization was widely derided: one participant described it as a visualization of 'junk', a phrase we adapt for the title of this paper to capture the derision that marked participants' responses to it. The criticisms levelled at the visualization revealed gendered, classed and age-related assumptions about the musicians, the relevance of social media, the colours used in the visualization, the imagined audience for the visualization and the trustworthiness of The Metro. But not all; alongside these derisive judgements were critiques of the visualization as poorly executed and failing to adhere to professional standards and conventions. We argue that gendered derision is entangled with legitimate criticisms of poor visualization execution and that, because of these entanglements, it is hard to see and so normalised. But we are not suggesting deliberate sexism on the part of our participants; rather, we argue that participants' responses to The

Clicks Don't Lie mobilized certain cultural discourses. Drawing on feminist research on the denigration of popular culture in order to consider the gendered dimensions of judgements about visualizations reveals how sexism takes various forms: how it is 'flexible', to use Gill's term (1993, 2011). This is why there is a need not just for feminists to critique big data, but for feminist data studies – that is, feminists doing big data and data visualization – to move both visualization and feminism forward in our increasingly data-driven world.

In what follows, we summarise debates about data's alleged neutrality, which we dispute, along with a number of feminist and other critics, we consider the relationship between gender, derision and popular culture, and we discuss the notion of 'flexible sexism'. We then turn to a close examination of the The Clicks Don't Lie, before analysing focus group participants' responses to it. We conclude by reflecting on what the responses mean for feminist data studies and data visualization.

<Figure 1: The Clicks Don't Lie>

Data and the myth of objectivity

As data become ubiquitous, they are seen to have more and more value and are increasingly relied upon, a phenomenon which has been called 'datafication' (Mayer-Schönberger & Cukier, 2013). Zealous belief in the explanatory potential of data is captured in the widely-cited words of Wired magazine editor-in-chief Anderson, who proposed that 'every theory of human behaviour, from linguistics to sociology, […] taxonomy, ontology and psychology' can be dismissed because, in what he calls the Petabyte Age, 'the numbers speak for themselves' (2008). This 'trust in numbers' is not new: as Porter noted in 1995 in a book of that name, numbers have long appealed because they can be understood from far away, by people distanced from and unfamiliar with the topic to which the numbers refer. More recently, Crawford (2013) made a similar point about big data: they allow us to view phenomena from afar (but, she goes on to argue, this means that we miss the detail that can be observed on closer scrutiny). Numbers are impersonal and seem objective, which minimises the need

for ‘personal trust’ (Porter, 1995, p. 223). As Porter notes, ‘a decision made by numbers (or by explicit rules of some other sort) has at least the appearance of being fair and impersonal’ (1995, p. 8). In other words, numbers look fair and objective.

A number of critics, Porter included, have highlighted the problems with assumptions about the objectivity of data (for example Day, 2014). These writers argue that data are not neutral, but rather are always made and shaped in particular ways: the titles of Gitelman’s (2013) edited collection Raw Data is an Oxymoron and Langlois, Redden and Elmer’s (2015) Compromised Data both communicate this point. Criticisms of the assumed objectivity of data have feminist histories, as feminist scholars have, for a long time, challenged such notions. Feminist critique (Haraway, 1988; Harding, 1986) has been important in questioning the very possibility of objectivity and, building on this, some feminist researchers have proposed that data themselves are masculine (for example Stanley & Wise, 1993). Scott (2010) argues that people working with quantitative data are well aware that they are not objective, but nevertheless, one outcome of objectivity criticisms is a heavy reliance on qualitative methods in feminist research. In this paper, we argue that it is precisely because of the types of problems that feminist scholarship has identified that there is a need not just for feminist critiques of quantitative methods, of data and associated assumptions of objectivity, but for feminism to do big data and data visualization. In other words, we need feminist data studies which is active in creating and communicating data. We say more about this in the conclusion.

The reception of a data visualization like The Clicks Don’t Lie takes place within multiple contexts: on the one hand, one in which data are seen as objective and therefore trustworthy and on the other, one in which sexism, classism and ageism play a part. We turn to these phenomena in the next section.

Gender, derision and popular culture

Popular culture has long been distinguished from high culture, in a hierarchy of aesthetic value in which the latter occupies first place: authentic, original and associated with the upper classes. Popular culture, on the other hand, is regarded as working class culture, and is thus seen to be of lesser value, manufactured and formulaic. Cultural researchers have worked to reclaim popular culture as both aesthetically worthwhile and worthy of study, with the work of the Centre for Contemporary Cultural Studies in Birmingham in the UK being particularly important. Building on this work, feminist researchers have argued that value judgements about media texts are gendered, with popular cultures that are aimed at girls and women subjected to harsh and often very visible derisions (see Baym, 1999). Social class, gender, and age (and race, although we do not discuss it here) intersect in the disdain levelled at popular culture, resulting in specific kinds of derision, depending on the media text in question. When popular cultures are targeted at girls and women, they become associated with a specific gendered identity: femininity. As Baym (1999) argues, popular cultural forms, such as soap operas, tend to be associated with emotionality, a stereotypically feminine trait. By extension, more 'highbrow' cultural forms, such as documentaries, are taken to denote reason and rationality, stereotypically masculine traits. The division between acceptable and derided media texts is therefore 'based on an assumed dichotomy between reason and emotion' (Jensen, 1992, p. 21), with reason trumping emotion in what is valued.

The Clicks Don't Lie depicts data about the social media followers of Shakira and Rihanna, two contemporary pop music artists. Pop music, like pop culture more broadly, is also feminised (Coates, 1997), seen as manufactured, falsely sentimental and simplistic. Pop's 'other' is rock, to which are attributed the characteristics of high culture (it is seen as authentic, emotionally honest, intellectually stimulating and deriving from the skilled performance of songwriter-musicians). The derision of pop music as a genre extends to its musicians, especially when they are women. No matter what the female pop musician actually does in the music making process, her autonomy and achievements are typically dismissed, and she is treated as little more than a pretty face used to sell music. This divide between rock and pop is clearly false: plenty of women pop musicians write and

produce their own emotionally honest and musically intricate songs, and Shakira and Rihanna are no exceptions.

Although Shakira and Rihanna are musicians, they have come to be associated with celebrity culture, which brings with it another set of gender, class, and age-related judgements. As Frith and McRobbie (1990 [1978]) note, when female musicians achieve fame, they are treated as celebrities, not musicians. Celebrities both male and female are perceived to have a largely young, female fan base (van der Graaf, 2014), meaning the derision of celebrity culture is also linked to age. The popular press frequently condemn the young 'fangirls' of celebrity culture for their inappropriate, overly emotional and obsessive fan practices, a condemnation challenged by feminist academics (Busse, 2013; Cann, 2015). When a media text is associated with popular, celebrity culture, gender, class and age-based judgements intersect. They inform and are informed by how we think about what are 'acceptable' forms of culture.

Thus responses to media texts cannot be divorced from the 'socially available ideologies and images' (Ang, 1985, p. 11) that shape the ways in which they attain their meanings. The Clicks Don't Lie is produced, circulates and is consumed in a context of value judgements about the particular gender, age and class intersection that it reflects, and these relate to pop music and celebrity culture. Despite widespread feminist critique, cultural texts that are seen to be for women or associated with femininity still occupy a contested space in our imaginings of what are acceptable forms of culture. This is the case with data visualizations as well as other cultural forms, in spite of assertions about the neutrality of data. The continued denigration of 'feminine' cultural texts takes place in a context of postfeminist discourse of equality that obscures sexist judgements. Gill's concept of 'flexible sexism', discussed in the next section, is valuable for understanding the processes at work in denigratory assessments of popular culture.

'Flexible sexism'

In recent years, the term 'sexism' has, according to Gill (2011), fallen out of fashion, in academic and everyday discourses. One reason for the hesitancy in talking about 'sexism' is the emergence of 'postfeminist' discourses in the 1990s (Gill, 2007), which maintains that sexism has now been eradicated, and that enduring imbalances are the result of individual failings, not structural inequalities (Rottenberg, 2013). This discourse renders sexism and related inequalities invisible and they therefore go unchallenged. However, as feminist scholarship makes clear, gender discrimination has not receded. Rather, it is taking on new forms. For Gill, this 'new sexism' manifests in 'agile, dynamic, changing, and diverse' ways, rather than as a 'single, unchanging "thing" (e.g., a set of relatively stable stereotypes)' (2011, p. 62). In short, sexism is *flexible*, and one of the key ways in which it operates is through the 'invalidation and annihilation of any language for talking about structural inequalities' (2011, p. 63). Gill makes this argument in relation to her research on new media work. In such environments, she argues, discourses about new media work help to 'produce sites in which discrimination flourishes', as they obscure inequalities such as sexism (2011, p. 62). She identifies an aversion amongst new media workers to discourses that highlight gender inequalities, which works to naturalize sexism within such settings (2002). The embeddedness and unspeakability of sexism therefore means that gender discrimination becomes difficult to identify.

To counter this, Gill (2011) argues that feminists should seek to uncover the manifestations of sexism in a range of different practices and cultural spaces. Data visualization is one such practice that needs to be closely examined in order to uncover hidden biases and sexist discourses, we argue. Our examination of participants' responses to The Clicks Don't Lie is therefore informed by the concept of flexible sexism. Below, we argue that participants' criticisms about the poor execution of the visualization are entangled with gender, class, and age-related derisions. This entanglement simultaneously illustrates the ways in which sexism is embedded within cultural discourses *and* illuminates the flexibility of sexism across cultural sites, including those which are considered objective, like big data. In the remaining sections, we illustrate this argument through a discussion of our focus group participants' responses to this particular visualization.

Researching responses to The Clicks Don't Lie

This paper draws on research undertaken on Project X, which aimed to understand people's engagements with data visualizations. The objective was not to study gender primarily, but rather to identify the range of factors that affect engagements. However, we were struck by the gendered character of discussion of The Clicks Don't Lie in distinction to the other visualizations, none of which were marked in this way, and this is why we single it out here. On the project, mixed methods were deployed including focus groups, diary-keeping, and interviews. Here we pay particular attention to the focus groups, as this was where participants saw and responded to The Clicks Don't Lie.

We recruited participants from pre-existing groups or communities, so they were therefore largely known to each other. We opted for homogeneity within focus groups because it results in understanding of others' lifestyles and situations and so facilitates discussion. We anticipated people would already feel comfortable in one another's company and therefore be more willing to open up (Kitzinger & Barbour, 1999; Krueger & Casey, 2000) and that the benefits of intergroup familiarity would outweigh the dangers of existing power imbalances within groups influencing the discussion, something that we worked to counter in the focus groups, ensuring that a range of views were heard and that no single participant dominated in any of the groups. What's more, the pilot study, data from which we reference here, was not drawn from an existing group but from representatives of all groups. Participants were therefore unknown to one another, yet discussion about The Clicks Don't Lie in this group was not dissimilar to that which took place in other groups.

We carried out nine focus groups with a total of 46 participants, in four geographical locations (which, given that migration data was a case study in the research, we characterise as: rural/high migration; rural/low migration; urban/high migration; urban/low migration). In addition to

the pilot group, we aimed to recruit participants who might be assumed to be interested in data, the visual, or migration, and so 'already engaged' in one of the issues at the heart of the project, and others about whom we could not make these assumptions. Selected groups included an art class, an open data group, a civil society group, a young farmers group, a rural community group, a religious migrant community group, two migrant groups and our pilot group (see Table 1).

Type of group	**No. participants**
art class (potentially interested in visual representation)	4
open data group (potentially interested in data);	8
2 East European groups (potentially interested in migration given personal and family histories);	2; 4
Asian/British Asian group (potentially interested in migration given personal and family histories)	6
Civil Society group (potentially interested migration, given their focus)	4
Young Farmers Association (potentially interested in migration, given their location)	6
Rural community (not assumed to be interested in data, migration or the visual)	6
Pilot group with representatives from most of the above categories	6
Total	46

Table 1: focus group participants.

27 participants were female and 19 male; ages ranged from 11 to 70, with the 30-39 age range best represented (18 participants). Employment sectors were extremely diverse, including fields like hairdressing and cleaning, local government, agricultural work, teaching, media, retail and

information services. All participants except four (two of whom were under the age of 16) had qualifications of some kind; 19 had completed tertiary education and 11 had higher degrees. As the study took place in the UK, most participants (n=31) self-reported as British, and other nationalities included German, Indian, Lithuanian, Pakistani, Polish and Thai. Focus groups lasted for two hours, during which we asked participants to evaluate eight visualizations circulating in the media (which can be viewed at http://seeingdata.org/developing-visualisation-literacy/rate-these-visualisations/). These represented a diversity of subject matters, chart types, original media sources, formats (print and online) and degrees of interactivity. Participants made notes about their first impressions, feelings and thoughts about each visualization, and this was followed by a full group discussion. This article focuses on discussion about The Clicks Don't Lie because of the strongly gendered character of this discussion.

The Clicks Don't Lie appeared in The Metro newspaper on 25th July 2014. The visualization and accompanying text (which can be found at http://seeingdata.org/developing-visualisation-literacy/clicks-dont-lie/) took up about a page of the newspaper. The story revealed that Shakira had just become the first person to receive 100m likes on Facebook and reflected on what this meant for her career. Thus the article, the accompanying visualization and the data within it are almost exclusively about Shakira, including a large cut-out photograph of her. Despite this, the visualization also features a large cut-out photograph of Rihanna, which is just as prominent as the image of Shakira. The visualization is framed by the two photographs, Shakira on the left and Rihanna on the right, which gives it some balance. The two photographs can be read as sexualising the musicians: Shakira with bare arms and midriff, armpits exposed, looking into the camera with long blond hair blowing about her head and partially covering her face. Her lips are parted and eyes slightly narrowed. These elements combine with others to suggest sexual availability, but the positioning of her hands is a yoga pose and suggests strength. Rihanna is squatting in front of a camera, thighs exposed between her bottom and the top of her knee-high boots, features which also serve to sexualise her.

Colour plays a vital role in the visualization. The dominant colours, bright orange and fuchsia pink, draw the different components of the visualization into an integrated text, creating a sense of a whole. The two musicians are bordered by them (Shakira in pink, Rihanna in orange), and this suggests a colour/musician relationship (which does not exist). This and the inclusion of the photograph of Rihanna were the source of some confusion for participants, as not all participants read the accompanying article, and so the visualization was assumed by some to illustrate data about both singers, as we discuss below. In fact, The Clicks Don't Lie almost exclusively represents data about Shakira's social media followers. Data are displayed in a range of chart types, some of which use the logos of Facebook, Twitter and Instagram, with one logo representing one million followers. Data are broken down by social media platform, by gender of social media followers, by marital status of fans and by the number of Facebook followers called Shakira. There is one piece of information about Rihanna, which refers to the number of retweets of one of her tweeted images (there is no comparable information for Shakira).

The Clicks Don't Lie is not a model visualization, and this is partly why we chose to include it in our study – we did not want all of our chosen visualizations to be defined as 'good'. Whilst the visualization uses the conventions of data visualization (graphically representing data through specific chart types and the inclusion of a data source, for example), there are a number of problems with it as a data visualization. First, it is difficult to compare the data. The size of the 'F's representing a million Facebook fans is much smaller than the Twitter birds and the Instagram cameras, both of which also represent one million followers (on Twitter and Instagram respectively). This means that the space given to each social media platform suggests similar quantities, whereas in fact, Twitter's space should be roughly one quarter of Facebook's to be accurate, and Instagram's one 25th. The number of Shakira's single fans, her Facebook followers called Shakira and their gender, are all represented in different ways: as a doughnut chart, as a figure in a circle, and as a pie/stacked bar chart. This inconsistency makes figures difficult to compare and comprehend.

Secondly, as noted, there is lack of clarity about who the various statistics reference, in part due to the two equal-sized photographs of Shakira and Rihanna, but also due to the lack of a formal key and the confusing use of colours in the visualization. Good visualizations have a clear key, and colour usually plays a vital role in indicating either what statistics refer to (for example, on pie charts where different colours are used to identify different categories) or to indicate the statistic itself (for example, on maps where varying shades of blue are used to indicate the varying depth of oceans). The Clicks Don't Lie does not utilise this convention and gives no indication that this is the case. With the lack of a formal key, the coloured borders around Shakira and Rihanna would seem to informally take on this function. However, the data in orange does not refer to Rihanna in most instances.

Thirdly, the positioning of the visualization within the newspaper is confusing. The title we use for The Clicks Don't Lie appears at the head of the article and is not repeated on the visualization; the visualization does not have a title of its own. Reading the article alongside the visualization, it is clear that all of the data, barring one item, are about Shakira. The visualization therefore does not stand alone, and this creates confusion. Finally, the visualization could be read as what influential data visualization specialist Tufte (1983) calls 'chartjunk duck'. Tufte argues that the majority of the ink on the page should be devoted to data, and not to other elements: such visualizations have a high data-ink ratio. The inclusion of the two photographs in The Clicks Don't Lie considerably lowers the data-ink ratio. Tufte also argued that 'chartjunk' (1983, p. 113) impedes the display of data and should be avoided. Graphics that are too decorative are a form of chartjunk and those that use the data as transport for decoration or other information are 'ducks'. In The Clicks Don't Lie, the photographs take up two thirds of the space and so could be viewed as the primary focus of the visualization: chartjunk duck. This is something about which our participants commented, as we show below.

The gendered derision of a data visualization

Despite our recognition of the limitations of The Clicks Don't Lie, we were surprised by the extent of the derision and derogatory language used to talk about it in our focus groups. Of our 46 participants, 27 people viewed The Clicks Don't Lie during the time set aside to engage with the eight visualizations. Some did not get around to looking at it because they spent time on other visualizations (i.e. they didn't get as far down the list of visualizations as The Clicks Don't Lie). Of these 27 people, 20 made derogatory comments about the visualization in their notes, three made comments that were neither positive nor negative, and four noted that they liked it. A further two people talked about The Clicks Don't Lie in the discussion, even though they had not taken the time to look at it during the allotted time. Of the four who made positive notes about it, three expressed dislike for it verbally. Thus of the 29 people who commented on the visualization in either their notes or the discussion, 27 expressed negative comments, which illustrates the predominance of negative feeling about it. These responses resulted only in part from its poor execution as a data visualization, we argue. We suggest that criticisms levelled at the visualization are also a product of the gender, age and class dimensions of discourses about pop culture and celebrity, which we discussed above. However, gender-based derision and comments about poor execution are entangled in complex ways. We discuss examples of both in the subsequent sections.

Poor execution

Many of our participants had expectations of what a good data visualization looks like and criticised The Clicks Don't Lie for not meeting them. For example, a number of participants noted that it lacked a title, with the result that they found themselves unsure of what they were looking at and how to make sense of it. In one focus group, the visualization was very unpopular with most participants for this reason, and Sarah was particularly annoyed by the lack of this simple visualization convention:

> There was no clear focus. Usually with those there's titles, like what the aim of the article is and the focus is; but it wasn't there. It was information and pictures. […] I think for me when I'm looking at them if it's a difficult layout or it's not right there… like the title thing got to me with the Rihanna thing because it's such a simple thing, but yet they didn't even say really in a title, a clear title, what the focus was. (Sarah, female, 34, white British retail worker)

The lack of a formal key combined with unsystematic use of colour, mentioned above, also proved problematic, making the visualization difficult to navigate and leading to miscomprehension, as Storm demonstrated:

> I didn't get that orange and pink thing because […] I assumed the pink was something to do with Shakira and the orange is obviously to do with Rihanna, because she's surrounded by the orange and she's surrounded by the pink. To which I said to myself: well why are they showing how many millions of Facebook fans Rihanna has got, and Twitter, vice versa, when if they're trying to show who's more popular they should show both Facebook statuses and both Twitter things. I didn't see the relevance of what they were saying. […] Shakira was everything represented in pink and Rihanna was everything in the orange. But then some of the things are a bit of both. It was a little bit higgledy-piggledy. (Storm, female 27, white British, employed)

The Metro's omission of a key for The Clicks Don't Lie confused our participants. A key is a convention of data visualizations that plays a vital role in enabling audiences to make sense of them. Nevertheless, Storm spent enough time looking at the visualization to enable her to make an important criticism of the data: the lack of comparability. Harry added to her criticism, saying: 'It doesn't give you comparatives' (Harry, female, 24, white British, worked in accountancy).

Storm and Harry were two of only three participants who critiqued the presentation of the data in the visualization; most of the criticisms were levelled at the subject matter and design. Even though a number of our participants worked with data (including Harry), few spent much time looking at the visualization to be able to make criticisms of it as a graphical representation of data. This speedy dismissal was apparent in some of the things they said, such as Ishmael's declaration, 'That's why it didn't take me very long. I looked at it and I was like, "Oh dear God, ignore!"' (Ishmael, female, 30, Indian, ICT worker). Participants' first impressions ruled their judgements, and these were frequently influenced by the subject matter (popular culture), the source of the visualization (The Metro), and the design and colour use (magazine-y, pink and orange).

Gender, class and age

Even participants who liked the subject matter of The Clicks Don't Lie (for example, they liked the musicians or they were interested in social media data) derided and disparaged it. For example, Haleema (female, 34, Pakistani Asian, unemployed) said that she liked it because she likes Rihanna and Shakira, but then made a link from the subject matter, calling it 'nothing', to her other engagements with popular culture. She talked about her husband's response to her family's use of social media and soap opera viewing. He did not approve and wanted them to watch serious programmes. In this assessment, Haleema likens the visualization - which is about social media, celebrities, and pop music - to soap operas, implying that they are all judged as not having value. The perception that the subject matter of the visualization has no value was evident in the language used to talk about it, which was consistently dismissive. Here are some examples of words and phrases used to describe it by a variety of participants:

- 'nonsense', 'why would you want to visualise junk?' (Noon, female, 32, Thai, student)

- ‘horrible’, ‘mess’, ‘garish’, ‘everything’s pink and orange’ (Jason, male, 34, white British, data scientist)
- ‘rubbish’ (Michelle, female, 26, white British, worked in beauty industry)
- ‘ugly’ ‘cluttered’, ‘chaff’, ‘pointless’, ‘it’s of no interest and no use’ (Horace, male, 27, white British, worked for an NGO)
- ‘not news’, ‘meaningless’ (Sara, female, 45, British Asian, careers advisor)
- ‘not really news; it’s just random information’ (Marek, male, 14, white British/Polish, school pupil)
- ‘null and void’ (Andrew, male, 32, white British, worked in agriculture)
- ‘why would anyone care’, ‘it doesn’t really matter’ (Kevin, male, 33, white British, technology worker)
- ‘a bit of gossip’, ‘no useful information on it’ (Manini, female, 39, Indian, adult education worker)
- ‘nothing really worth dwelling on’ (Mark, male, white British, local government data scientist)
- ‘there’s no information here’, ‘it’s just pointless data’ (Storm, female, 27, white British, employed)

Many of these criticisms, some of which are extremely dismissive (‘chaff’, for example), were levelled at the subject matter of the visualization. Social media and celebrity were seen to be trivial, uninteresting and therefore derisible by a number of our participants. Shakira and Rihanna were seen not as musicians but as celebrities, something that compounded what was perceived as the banality and meaningless of social media data. Kevin was one of a number of participants who spoke dismissively about the subject matter:

> What do we learn from knowing that one million and one people have liked her Facebook page? It doesn't really tell us anything other than one million and one people have liked her Facebook page. We already know she's a celebrity so it's not really adding any value intellectually at all. (Kevin, male, 33, white British, technology worker)

Kevin's dismissal of the social media data on which the visualization is based was tied up with his assumptions about Shakira's popularity. In these comments, different types of data are valued differently. Whilst some data might be seen as neutral and objective, not all data are equal: some data are more valued than others.

For Michelle, the subject matter implied a particular audience:

> If one of the young girls I work with looked at that they'd be like wow, look at how many followers she's got. It would probably interest them; whereas me, I'm not really interested. (Michelle, female, 26, white British, beauty industry worker)

Michelle assumed that social media interest young girls, not mature women. She distinguishes herself from those she believes the visualization to be aimed at – young girls. This assessment is based on the design of the visualization, another source of criticism, which draws on gendered and classed norms and values. For example, the design was associated with the style of women's celebrity and lifestyle magazines. As Sara put it, 'I thought the images reminded me of, you know, these women's gossip magazines' (Sara, female, British Asian, 45, careers advisor). The exchange below shows that Michelle and Sarah assumed that such magazines appeal to a certain class of women:

> [Michelle] I think it looks kind of cheap. It's the kind of thing you would see in a magazine, a cheaper magazine. I don't think you would see that in the middle of – I don't know, what's a more expensive magazine?
>
> [Sarah] The Times. The Lady. ((Laughter))
>
> [Michelle] You wouldn't see that in The Lady. You'd see that on Reveal: a total gossip based magazine that you could sit in a waiting room and read it and you know that every bit you're reading is rubbish.

This exchange assumes a consensus about what counts as quality (The Times, The Lady) and what is 'rubbish'. The Metro is a 'lowbrow' tabloid, whereas The Times is a conservative broadsheet newspaper. The Lady is a women's magazine associated with upper class lifestyles (for example, it includes advertisements for domestic staff). Quality, for Michelle and Sarah, is to be found in the cultural products of the upper class. The laughter that punctuated the focus group at the thought of The Clicks Don't Lie appearing in The Lady is an indicator of the scorn levelled at the visualization.

Similarly, the colours used in the visualization were subject to widespread critical commentary, and were assumed to indicate a young, female audience. Jason made an explicit judgement about the colours:

> It just looks a mess. It's very, very scrappy. I thought it was very garish, and the fact that they've used these colours, everything's pink and orange, and someone must have thought that was a really thing cool thing to do. But it's all just different things that are all mixed together. (Jason, male, 34, white British, IT worker)

Colours are 'culturally coded as masculine or feminine' (Gleeson & Frith, 2004, p. 104), with pink being used as a 'code' for recognising 'immature feminine identities' (Gleeson & Frith, 2004, p. 104). Colours, therefore, are not only used as tools to distinguish between masculinity and femininity, but also between different kinds of femininity. In this case, the colours suggest a femininity that is

distinctly ‘girly’ (Gleeson & Frith, 2004). When Jason writes off the colours in the visualization, he is summoning these gendered colour codings and their associations, even though he may not intend to do this.

The subject matter, design and colours of The Clicks Don’t Lie combined to connote to our focus groups that the visualization was intended for young girls interested in celebrity culture. As Ishmael put it, ‘pre-teen[s] would probably have loved it’. These associations position the visualization as trivial, and our participants worked to distance themselves from this demeaned imagined audience, even when they claimed some interest in Shakira, Rihanna and social media. For example, Sara said that, for her, the visualization was ‘meaningless’, and that she would ‘rather read something else which meant something’ to her, like ‘world matters’ or the news. Many participants worked to emphasise that they are not interested in this kind of content, and that they find other media, like the news more valuable. These statements are linked to longstanding assumptions about masculine, rational, ‘highbrow’ cultures as more valuable than the perceived frivolity of feminine, trivial, and ‘girly’ media.

Entangled derision

Some of the criticisms of The Clicks Don’t Lie were informed by sexist, classist and ageist cultural discourses, However, they were also, at the same time, a response to the poor execution of the visualization. This was particularly evident in comments in which participants identified it as chartjunk duck (that is, a visualization in which the data is secondary to the decoration), without, of course, using this term. For example, some participants were suspicious of the motives behind the production of the visualization, suggesting that the desire to communicate data was not the driving force behind it. Rather, because the data were seen to be of limited value, the visualization existed ‘more for the design. It’s more to look good than be informative’ (Andrew, 32, male, white British,

agricultural worker). For some participants, the visualization was made only to show images of women wearing little clothing:

> Part of me feels like it's just The Metro wanting to try and put pictures of attractive women in their newspaper on the basis that you'll sit on the bus, you'll see the bloke in front of you reading it and you'll want to go and get a copy yourself. It doesn't feel like it's given me any information. (Paul-Stewart, male, white British, 34, local government worker)

These comments draw simultaneously on feminist discourses which criticise sexualised media images of women (see Attwood, 2009) and on discourses about what kinds of visualizations are valuable. To see the visualization as a form of chartjunk duck is to suggest that the data are pointless in themselves, as many of our participants claimed they were. One participant (Harriet, late 20s/early 30s, female, white British, NGO worker) said that the objectified images of women got in the way of her engagement with the visualization, just as for other participants their judgement about the subject matter as pointless inhibited their in-depth consideration:

> [Harriet] There's so much in the media about the way that women should look. Like in The Metro, I'm sure there are a lot of pictures of […] Shakira, yes. Yeah, just that there's a lot of focus that's particularly in the media that I absorb. I do read magazines, but I still get annoyed, I guess is what I'm trying to say.
>
> [Interviewer] And that affected how you approached that particular thing?
>
> [Harriet] Yes, I would say definitely shaped how I approached that, absolutely.

Responses to visualizations are shaped by socio-cultural contexts and political beliefs, and gender plays an important role in whether responses are overtly feminist, like Harriet's, or reproductive of the status quo, such as when 'feminine' topics are derided and judged as worthless. Participants' criticisms of The Clicks Don't Lie for not meeting the standards of a 'good' visualization were entangled with gendered assumptions about pop culture, celebrity and young femininities.

Participants' scepticism seemed to be about the visualization failing to meet standards and this meant that the distinctly gendered derisions of the visualization, such as its focus on 'attractive women' (Paul-Stewart), its 'pointless' (Storm) data about celebrity and social media, and its use of 'fun' (Meg, female, white British, 80s, retired) and feminine colours, recede from view. The gendered dimensions of derisions of the visualization thus become invisible. In this way they are a manifestation of flexible sexism. This is because the entanglement of criticisms based on sexist, classist and ageist cultural discourses and others noting the poor execution of the visualization means that sexist critiques of The Click Don't Lie are obscured by criticisms about the professionalism of the data visualization.

To make critiques based on an apparent lack of professional standards gives the impression of unbiased judgement. However, our research uncovered the innate flexibility of sexism within the responses to The Clicks Don't Lie. We are not suggesting deliberate and intentional sexism on the part of our participants. Rather, we are arguing that, in their responses to The Clicks Don't Lie, they mobilised gendered cultural discourses and norms to forge their critiques. These discourses - which we argue are sexist - are so culturally embedded that they have become naturalised, and they play a role in assessments of cultural and media artefacts. Our research illustrates the ways in which sexism intersects with 'other axes of power'(Gill, 2011, p. 62), such as age and class. Indeed, perceptions of The Clicks Don't Lie's as feminine hindered our participants' ability to make informed judgements about the quality of the visualization *as* a visualization because their first impressions impeded thoughtful consideration. Postfeminist discourse suggests that sexist judgements of The Clicks Don't Lie should be read as neutral, apolitical statements, agnostic to derisive views about the value of women's culture. This in turn illustrates the dynamism of 'new sexism' (Gill, 1993), and points to the importance of analyzing manifestations of sexism across multiple media forms.

Conclusion: redeeming 'junk', or towards feminist data studies and data visualization

There was a significant gendered, class- and age-related component to focus group discussions of The Clicks Don't Lie. The lack of value attributed to the visualization derives from sexist, classist and ageist evaluations of the subject matter, the design (including colour use), the perceived lack of trustworthiness of The Metro newspaper, and, sometimes, the graphical representation of the data. This devaluing of the visualization was seen in the language of denigration consistently used by focus group participants, even by people who liked the visualization, the musicians or the fact that it was about social media data. These aspects of participants' responses are entangled with critiques of the poor execution of the visualization. This entanglement, we argue, is an example of flexible sexism.

When it comes to understanding how assessments of data visualizations are made, and how professional standards are discussed, taking sexism into account is vital. It was apparent from our focus groups that intertwined with ideas about chartjunk and what counts as valuable data are assumptions about what it is worthwhile visualizing and how best to visualize it. Elsewhere, two of us have argued that research into how people engage with data visualizations must consider users' social context (Kennedy et al., submitted). As we have seen in the discussion in this paper, this social context is shaped by gender, and by more than gender. Cultural discourses about young femininities that underscored the derision of The Clicks Don't Lie represent a particular intersection of ideas about gender, age and class. This again is an example of what Gill describes as 'thoroughly intersectional' (2011, p. 67), and thus flexible, sexism.

Because of judgements of the kind that we encountered in our research, there is clearly an urgent need for feminist critiques of big data, which seek to understand how gendered processes work and impact upon understandings of data. Such work is beginning to be undertaken, for example in feminist work on the processes of visualization (D'Ignazio, 2015; Kennedy et al., 2016) and on data creation (Bivens, 2015) which challenge techno-focussed discourses of data that leave important social questions aside. But there is also a need to go beyond critique, for feminists to do big data and data visualization, or what we call feminist data studies. In other words, feminism needs to engage

with data and visualization discourses and practices, to move both visualization and feminism forward in our increasingly data-driven world.

Arguably, this is also already under way, through initiatives like Feminist Accountancy (Waring, 1999), which considers areas of women's economic activity overlooked by mainstream accountancy, quantitative feminist sociology (Cohen, 2016) and through feminist uses of statistics to highlight inequalities (Scott, 2010). The possibility of feminist data visualization, which, according to proponent D'Ignazio, would be open to visualizing uncertainty and enable those reading the visualization to 'talk back to data' (D'Ignazio, 2015, n. p.) is also being explored. These projects seek to 'do' data and visualization in ways informed by feminist critique.

Cohen (2016) argues that there is a need for feminists both to critique and have an impact on how quantitative research is done, to ensure that issues relating to women's subordinated place in society are not overlooked. Feminist data studies needs to use data to: expose sexism in all its forms; identify and rectify gendered blindspots in current research; highlight where more research and visualization is needed; challenge gendered assumptions about data subject matters, data creation and visualization, and audiences; engage in debates about the potentials and problematics of big and quantitative data; ask feminist questions about the often-assumed neutrality and objectivity of data; and bring to light the embeddedness of sexism within all cultures, including data cultures. Feminist data studies should work to redefine what counts and what is counted. This might lead to better, fairer and more reflective data generation and visualization.

References

Anderson, C. (2008). The end of theory. Wired. Retrieved 12 January 2016, from http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Ang, I. (1985). Watching Dallas: soap opera and the melodramatic imagination. London: Methuen.

Attwood, F. (2009). Mainstreaming sex: the sexualization of Western culture. London: I.B. Tauris.

Baym, N. K. (1999). Tune in, log on: soaps, fandom, and on-line community. Thousand Oaks, Calif.: Sage Publications.

Beer, D., & Burrows, R. (2013). Popular culture, digital archives and the new social life of data. Theory, culture & society, 30(4), 47-71.

Bivens, R. (2015). The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. New Media & Society, 1-19.

Busse, K. (2013). Geek hierarchies, boundary policing, and the gendering of the good fan. Participations, 10(1), 73-91.

Cann, V. (2015). Girls and cultural consumption: 'typical girls', 'fangirls', and the value of femininity. In H. Savigny & H. Warner (Eds.), The politics of being a woman: feminism, media and 21st century popular culture (pp. 154-174). Basingstoke: Palgrave Macmillan.

Coates, N. (1997). (R)evolution now? Rock and the political potential of gender. In S. Whiteley (Ed.), Sexing the groove: popular music and gender (pp. 50-64). Abingdon: Routledge.

Cohen, R. L. (2016). Towards a quantitative feminist sociology. In L. McKie & L. Ryan (Eds.), An end to the crisis of empirical sociology? Trends and challenges in social research (pp. 117-135). Abingdon: Routledge.

Crawford, K. (2013, 1 April). The hidden biases in big data. Harvard Business Review. Retrieved 18/03/2016, from http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/

D'Ignazio, C. (2015). What would feminist data visualization look like? Retrieved from https://civic.mit.edu/feminist-data-visualization

Day, R. E. (2014). Indexing it all: the subject in the age of documentation, information, and data. London: MIT Press.

Frith, S., & McRobbie, A. (1990 [1978]). Rock and sexuality. In S. Frith & A. Goodwin (Eds.), On record: rock, pop, and the written word (pp. 371-389). London: Routledge.

Gill, R. (1993). Justifying injustice: broadcasters accounts of inequality in radio. In E. Burman & I. Parker (Eds.), Discourse analytic research: readings and repertoires of texts in action. London: Routledge.

Gill, R. (2002). Cool, creative and egalitarian? Exploring gender in project-based new media work in Europe. Information, Communication & Society, 5(1), 70-89.

Gill, R. (2007). Postfeminist media culture: elements of a sensibility. European Journal of Cultural Studies, 10(2), 147-166.

Gill, R. (2011). Sexism reloaded, or, it's time to get angry again! Feminist Media Studies, 11(01), 61-71.

Gitelman, L. (Ed.). (2013). Raw data is an oxymoron. Cambridge, MA: MIT Press.

Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), Raw Data is an Oxymoron (pp. 1-14). Cambridge, MA: MIT Press.

Gleeson, K., & Frith, H. (2004). Pretty in pink: Young women presenting mature sexual identities. All About the Girl. London: Routledge, 103-144.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. Feminist studies, 575-599.

Harding, S. G. (1986). The science question in feminism. Milton Keynes: Open University Press.

Jensen, J. (1992). Fandom as pathology: the consequences of characterization. In L. A. Lewis (Ed.), The adoring audience: fan culture and popular media (pp. 9-29). London: Routledge.

Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. Information, Communication & Society, 19(6), 715-735. doi: 10.1080/1369118X.2016.1153126

Kennedy, H., Hill, R. L., Allen, W., & Kirk, A. (submitted). Engaging with data visualizations: users, socio-cultural factors and definitions of effectiveness. First Monday.

Kitzinger, J., & Barbour, R. S. (1999). Introduction: the challenge and promise of focus groups. In R. S. Barbour & J. Kitzinger (Eds.), Developing Focus Group Research: Politics, Theory and Practice (pp. 1-21). London: Sage.

Krueger, R. A., & Casey, M. A. (2000). Focus Groups: a Practical Guide for Applied Research (3rd ed.). London: Sage Publications.

Langlois, G., Redden, J., & Elmer, G. (Eds.). (2015). Compromised data: from social media to big data. London: Bloomsbury Publishing.

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: a revolution that will transform how we live, work, and think. Boston, Mass: Houghton Mifflin Harcourt.

Porter, T. M. (1995). Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton: Princeton University Press.

Rottenberg, C. (2013). The rise of neoliberal feminism. Cultural Studies, 28(3), 418-437.

Scott, J. (2010). Quantitative methods and gender inequalities. International Journal of Social Research Methodology, 13(3), 223-236.

Stanley, L., & Wise, S. (1993). Breaking out again: feminist ontology and epistemology (2nd ed.). London: Routledge.

Tufte, E. R. (1983). The visual display of quantitative information. Cheshire, CT: Graphics press

van der Graaf, S. (2014). Much ado about Keanu Reeves: the drama of ageing in online fandom. In L. Duits, K. Zwaan & S. Reijinders (Eds.), The Ashgate research companion to fan cultures (pp. 35-48). Farnham: Ashgate.

Waring, M. (1999). Counting for nothing: What men value and what women are worth. Toronto: University of Toronto Press.