Applying Computational Analysis to Textual Data from the Wild: A Feminist Perspective

Shauna Julia Concannon
Open Lab, Newcastle
University
Newcastle, United Kingdom
shauna.concannon@ncl.ac.uk

Madeline Balaam
Media Technology and
Interaction Design
KTH Royal Institute of
Technology
Stockholm, Sweden
balaam@kth.se

Emma Simpson
Open Lab, Newcastle
University
Newcastle, United Kingdom
emma.simpson@ncl.ac.uk

Rob Comber
RISE SICS
Stockholm, Sweden
rob.comber@ri.se

ABSTRACT

With technologies that afford much larger-scale data collection than previously imagined, new ways of processing and interpreting qualitative textual data are required. HCI researchers use a range of methods for interpreting the ‘full range of human experience’ from qualitative data, however, such approaches are not always scalable. Feminist geography seeks to explore how diverse and varied accounts of place can be understood and represented, whilst avoiding reductive classification systems. In this paper, we assess the extent to which unsupervised topic models can support such a research agenda. Drawing on literature from Feminist and Critical GIS, we present a case study analysis of a Volunteered Geographic Information dataset of reviews about breastfeeding in public spaces. We demonstrate that topic modelling can offer novel insights and nuanced interpretations of complex concepts such as privacy and be integrated into a critically reflexive feminist data analysis approach that captures and represents diverse experiences of place.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation: Miscellaneous; H.3.1 Information Storage and Retrieval: Content Analysis and Indexing

Author Keywords

Topic modelling; feminism; human-data-interaction; data analysis; geodata; GIS; feminist GIS; critical GIS; text analysis.

This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI 2018 April 21–26, 2018, Montreal, QC, Canada

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5620-6/18/04.

DOI: <https://doi.org/10.1145/3173574.3173800>

INTRODUCTION

Researchers in Human-Computer-Interaction (HCI) have increasingly focused their efforts on in the wild deployments of technologies that aim to empower communities. Working in this way can support extended engagements and garner insights from in-situ user experiences. However, the data collected can be large and unwieldy and finding scalable ways of working with such rich data can be challenging [7, 3, 19].

Recent discourse within HCI has brought feminist perspectives to bear upon the theory, methodologies and practices of interactive systems research, calling for more direct engagement with gender and the development of a socio-technical theory of gender [47]. Bardzell highlights that “feminist HCI entails critical perspectives that could help reveal unspoken values within HCI’s dominant research and design paradigms and underpin the development of new approaches, methods and design variations” [4]. Bardzell and Bardzell outline key methodological positions characteristic of a feminist HCI methodology, such as a commitment to understanding participants’ experiences in their own terms, the development of mixed-methods approaches which support dialectal information gathering, and reflexivity [5].

Geographical information systems (GIS) have become prevalent in mobile devices, and are applied at various levels of urban place-making, from urban planning and crisis informatics to individual, persuasive and aesthetic interventions. Such systems have become instrumental in decision-making processes for cities, organisations and citizens, and researchers have grappled with the mechanisms through which to make sense of GIS and its relationship to place-making. For instance, although prominent accounts of place-making in HCI focus on experiential accounts (e.g. [18]), it is less evident that such approaches account for systemic roots of such experiences [51]. On the other hand, computational accounts such as crisis informatics and urban informatics for smart cities, draw our attention to the infrastructures of urban spaces, but may not

relate the personal, aesthetic and affective experiences we have in and of cities [48]. In this paper, we explore the opportunities of Feminist GIS to address this divide. In particular, we are interested in how we understand GIS and the data contained in such systems as a form of human-world relation [22], and the consequence of technological, or specifically, geographical information systems mediation of place-making. We are interested in how multiple accounts of a GIS dataset can introduce insights on the systemic and experiential making of places.

Specifically, in this paper we are interested in the exploration of Feminist GIS for uncovering the multiple perspectives which relate to the experiences of breastfeeding mothers. Previous studies have found that privacy concerns can be an important factor for breastfeeding parents [50, 30], with fear of breastfeeding in public causing ten percent of parents to abstain from breastfeeding altogether [13]. However, what privacy means in this context can vary dependent on a variety of personal preference factors. Furthermore, how concerns about privacy manifest in different demographic groups is unclear, although it has been established that breastfeeding in public is more common amongst older mothers, those from the least deprived and professional socio-economic groups [30].

With technologies that afford much larger scale data collection than previously imagined, new ways of handling, processing and interpreting the data are called upon. However, as acknowledged in critical [23] and feminist [29] data studies, data is not neutral, and from collection and analysis through to how we represent and report on data mindfulness of this non-neutrality should be employed. Feminist and critical studies of GIS seek to link individual experiences and agency with an understanding of how ‘social, economic, and political processes are constructed, maintained, legitimized, and resisted’ [25]. Our aim is to explore the suitability of scalable approaches for making the analysis of large datasets more manageable and to assist in the uncovering of place-based accounts from experiential data whilst maintaining critical reflexivity. Consequently, through this process, we assess how appropriate computational approaches are for a feminist GIS analysis, and ask how such an approach may benefit the work of the HCI community.

We present an application of Feminist GIS for the analysis of Volunteered Geographic Information (VGI) textual data. In particular, we explore some of the challenges of conducting analysis on large-scale VGI datasets which represent diverse experiences of place and situated knowledge. We demonstrate how official datasets can be brought together with VGI data to yield new insights on place-based experiential data. Using automatically produced topic models the researchers discovered new meanings within the data, that highlighted how socio-economic factors impact on different conceptions of privacy in public spaces for breastfeeding parents. Building on the work of Feminist HCI and drawing on discussions from Feminist and critical GIS, which echo many of the sensibilities that drive HCI researchers, we highlight how researchers can seek to address the inherent tensions that arise when dealing with qualitative and subjective data that scales.

RELATED WORK

In this section we will review the literature on geographical data and computational approaches to textual analysis. Theoretically, this paper draws Critical and Feminist perspectives on GIS and data studies. We draw attention to the complex ways in which data can mediate accounts of place-making.

Digital data and geographical knowledge

Advances in mobile technologies enable individuals to create huge volumes of rich digital data while embedded within spaces of everyday life. A smart phone, for example, is a camera, video and audio recorder, text editor and GPS tracker all in one. For researchers, this offers a unique opportunity to capture data generated in every day contexts, rather than lab settings, whilst using pre-existing technologies. An increasingly common output of in-situ deployments of pervasive and ubiquitous technologies is that the number of participants and amount of data generated can become hard to manage. This is particularly challenging when the data is richly qualitative in nature. With user-generated data in unstructured text form, content creators are free to respond from their subjective position and draw on their situated lived experience when selecting what and how to contribute. This can pose a number of issues for researchers trying to extract insights from and reveal pathways through this type of data. How can qualitative data that represents localised spatial knowledge be handled on scale? We will begin by consulting the literature on Geographic Information Systems and innovations which combine quantifiable geographic data with qualitative and subjective accounts of place.

Geographic Information Systems

GIS support the management, analysis and presentation of geographic data. Geographic databases were initially dominated by numerical information [40] and more strongly associated with quantitative geography¹. In 1987 Dueker described a GIS as an information system with a database consisting of ‘observations on spatially distributed features, activities, or events, which are definable as points, lines, or areas’ and ‘procedures to collect, store, retrieve, analyze, and display such geographic data’ [10, p.384]. Traditionally, GIS was firmly positioned in the domain of quantitative research [14, 36, 42, 49].

Traditional Geographical Information Systems (GIS) studies have focused on visualisation and mapping of geographical data, using quantitative and *positivist* research methodologies [33]. Critical Geography, meanwhile, is charged with the agenda of exposing the ‘socio-spatial processes that (re)produce inequalities between people and places’ [21, p.62]. Feminist critiques of quantitative approaches posit that they are incapable of reflecting the richness and complexity of women’s lives [26] and voice a scepticism towards the universally applicable generalisations put forth by such methods. Kwan [27] highlights that quantitative methods are not inherently problematic, but suggests that:

Feminist objectivity should be understood in terms of the situated knowledges based on particular “standpoints”

¹For a detailed review of the methodological positioning of GIS and how this has shifted over time c.f. [40]

or limited “positions” of women’s lived experiences in particular social and geographical context [27].

Feminist GIS calls for resistance to binary or reductive interpretations of geographic data [48]. O’Sullivan identifies the role of feminist geographers working with GIS as ‘establishing the viability of using GIS to see the world and the individual lives within it differently, whatever restrictions current technologies may place on such efforts’ [37, p.785].

Pavlovskaya [40] challenges the claim that GIS is inherently quantitative, suggesting that qualitative and quantitative approaches are not mutually exclusive. Although critical geographers have more often favoured theoretical and qualitative approaches, efforts have been made to bridge the methodological divide. Kwan [28] identifies these endeavours as hybrid geographies: approaching issues informed by critical geographies with quantitative of GIS methodologies, and those that seek to ‘cross the boundary between geo-spatial technologies (GIS and GPS) and a qualitative understanding of the lived experiences of individuals in various cultural contexts’ [28, p.758].

Using Hybrid approaches

Kwan highlights the potential strengths of quantitative approaches for feminist geography in describing measurable aspects of complex spatial relations and women’s lives. Quantitative data and methods, with their potential to provide a broad picture of spatial, temporal, social inequality, can be useful in political discourse and “a powerful instrument for initiating progressive social and political change” [27, p. 19]. More recently, feminist geographers, whilst maintaining a commitment to reflexivity and qualitative methods, have sought to reclaim quantitative analysis [40, 32, 27]. Efforts towards a hybrid approach have combined qualitative data such as audio, video, hand drawn graphics, photographs and videos, with GIS [49, 26, 9]. However, what a hybrid data analysis process should look like and the type of data required to do justice to the agendas laid out by feminist geographers needs further consideration.

In Feminist HCI this commitment to a flexible approach to methods is equally stressed; Bardzell and Bardzell reaffirm that feminist theory does not privilege qualitative over quantitative, but is primarily concerned with the mindful selection and acknowledgement of the methods best suited in a given context to “represent people on their own terms” [5]. Furthermore, they highlight that “methods are not simply tools found in the world that neutrally reveal data as it is”, making it the role of the researcher to be as transparent as possible about the processes employed.

Data sources

Kwan [27] prioritises the collection of primary quantitative data over a reliance on official statistics and stresses that great care is required when developing a coding scheme as ‘rigid categorization is a major weakness of quantitative methods’. However, research on critical, participatory and feminist approaches to GIS has highlighted VGI has the opportunity to be more accessible than traditional GIS, both in the breadth of peoples who can access and contribute, but also in the language

and data structures used. Within VGI data can be ‘communicated and stored in everyday terminologies, rather than in the codified language of many existing spatial databases’ [12, p.175]. Consequently, VGI has the potential to reconstruct power relations through shifts in who can participate, produce and access data, and the associated ‘knowledge’.

VGI: place-based data in the wild

VGI has the potential to provide more heterogeneous knowledge from a diverse group of participants [12]. Rather than viewing conflicting or contradictory experiences and accounts of place as problematic and indicating a lack of accuracy or validity, a critical GIS perspective seeks to understand whether these contradictions may be ‘indications of social and political difference’ [12].

How might these proposals from geographies be applied more widely to the interpretation of data and the field of HCI? And what are the practical challenges and obstacles associated with such an endeavour? We now focus our review on approaches to textual data, to examine some possible ways forward.

Interpreting Textual Data

In the previous section we addressed insights drawn from critical geography to understand how spatial knowledge is constructed using qualitative and quantitative approaches. Researchers in HCI attend to the ‘full range of human experience’ and have developed a rich range of methods for interpreting qualitative data that work ‘to understand social worlds in their complexity and contradiction’ [40]. In this section we will review computational approaches which can aid exploration and textual analysis, and explore their suitability for experiential place-based data.

The potential for placing automated topic analysis alongside more qualitative approaches has gained traction, particularly for research in sociology [56]. The approach described in this paper is most aligned with Gill et al. [15], which combines techniques such as keyword in context, frequency information and topic modelling as an empirical point of entry for human analysis. In particular, we are exploring the suitability of such methods for interpreting geo-referenced textual data.

Topic Models

Topic models, e.g. [6, 20, 16], are generative models for documents: a topic is modelled as a probability distribution over words, and a document as a distribution over topics. This captures the intuition that topics have distinctive but not mutually exclusive vocabularies, and that documents often discuss multiple topics. By estimating distributions from data, such models can discover the underlying topics in a collection of documents unsupervised, avoid a reliance on labeled training sets, and respond and adapt to the data. Topic models have proven useful for categorisation and filtering in many text domains, including news articles [52], academic abstracts [45, 39, 44] travel blogs [38] and digital humanities scholarship [24].

Previous work has shown that topic models, and in particular Latent Dirichlet Allocation (LDA) [6] can also identify topics in short documents such as tweets, which are extremely

short texts (140 characters or less) and feature irregular, non-standard English [55, 41, 43]. Applications of LDA have been used successfully on Twitter data to infer user attributes such as political orientation, gender and ethnicity [41], discover dialogue structure [46], categorise messages by style or function [40] and capture a good level of nuance between less clearly defined topics such as genre [8]. Unsupervised approaches have the potential to cast a new lens of enquiry on a dataset, and discover hidden or latent topics that can then be considered, contextualised and reflexively responded to in turn. Existing work in the HCI literature has typically used topic models for data from social media which changes over time and for the automation of content analysis in which the size of the dataset is too large for human annotation. This latter benefit of machine learning approaches is particularly pertinent as ubiquitous and pervasive technologies are increasingly supporting the collection of larger datasets. For example, Wang et al. [55] applied LDA to Facebook status updates to assign topics and assess which topics were more typically authored by individuals according to their gender. In this context this approach was identified as appropriate due to the large number of items in the dataset (half a million Facebook status updates). Xu et al. [57] applied LDA to crowdfunding updates, which enables them to classify the types of updates issued during campaigns. The resulting categories provide a useful indicator of the different discourse acts or functions of the various updates, such as ‘Social Promotion’, ‘Progress Report’ and ‘Answer Question’.

Interpreting topic models poses another challenge. Research in HCI focuses on how to visualize the output and make the navigation of automatically generated topics more intuitive [11]. How keyword queries and topic model exploration can be combined in interactive interfaces has also been explored, to support humanities and social sciences scholars to integrate this approach into their workflow [24].

In the majority of examples, LDA is used to describe the data. In this work, we use the topic models to ask questions of the data, and to ask questions of pre-existing, quantitative geographic data. Through this process we begin to unpack how such an approach can enable researchers to build up richer interpretations of place-based data, that combines qualitative and quantitative content to facilitate contextualised readings of situated knowledges. The potential of combining computational analysis with more traditionally qualitative approaches has been highlighted in examples such as [53, 54, 34], which combine Critical Discourse Analysis with topic modelling or computationally aided corpus studies.

In this paper we refocus the discussion on a tangential application of unsupervised topic modelling, and highlight how a data-led approach could speak to the values of feminist and critical approaches to big data. The unsupervised aspect of topic modelling with the potential to discover underlying topics with less researcher direction, provides a useful way to examine the data, outside of prior expectation. What needs to be considered, therefore, is not simply what the topics reveal, or how best to present them, but how to design for iterative and dialectical researcher-data dialogues. What should a hy-

brid approach look like and how can the skills of researchers and computational approaches be brought together for the analysis of textual data? The limitations of machine learning approaches are numerous; they cannot create datasets, nor account for contextual and material differences which disadvantage certain groups over others. Integrating this wide-lens and contextual perspective is crucial, but thinking about what computational approaches can bring to the table is necessary too.

CASE STUDY: FEEDFINDER

In this section we present a case study of an application of Feminist GIS for data analysis. The primary source materials used is a VGI dataset generated by breastfeeding parents using a mobile application called FeedFinder, created using a co-design process led by Balaam et al., as described in previous work [2]. FeedFinder was developed with the intention of creating a supportive health technology for women to make the decision to breastfeed in public. It is a mobile application that enables breastfeeding parents to find, review and share experiences of public breastfeeding sites. Reviews are comprised of criteria which is deemed important for meeting a woman’s breastfeeding needs; privacy, hygiene, comfort and baby facilities - each of which can be rated out of five, together with free form textual review comments.

The knowledge created by users of FeedFinder is unique, in that it is volunteered and created within a peer support context. Furthermore, as it elicits reflections on specific locations and experiences of place, the text review data produced is often focused and specific. As such this offers an opportunity to garner insights into immediately prominent concerns that mediate the lived experience of the breastfeeding parent. FeedFinder generates inherently spatial knowledge, in that the views expressed are directly responsive to particular places. While the star ratings provide a quantitative measure of how breastfeeding friendly and suitable a venue is, the free-form comments field enables individuals to share more nuanced responses. The majority of reviews on FeedFinder feature a comment. It is the textual review data that the analysis, presented here, particularly focuses on.

Furthermore the thematic concerns addressed in individuals’ comments are diverse [50]. In the comment data the various regions of concern that shape their experiences are attended to, and the specific priorities of the individual encoded. Though the benefits of breastmilk are well documented and breastfeeding remains top of the global public health agenda, rates in the UK fall well below the recommended guidelines. Public Health England (PHE) recently published a report detailing that a third of breastfeeding women choose not to breastfeed in public, and one in ten new mothers abstain from breastfeeding altogether due to the fear of breastfeeding in public [13]. Women who do breastfeed in public are much more likely to be older, white women with increased parity (more than one child) from the least deprived areas, as defined by the Index of Multiple Deprivation (IMD) [1]. A more detailed review of literature on breastfeeding in public is included in the authors’ prior work [50].

Prior to the development of FeedFinder and the associated research [46], research on breastfeeding practices has been dominated by survey data (e.g. [30]). Up until 2010 this data was collected through a national Infant Feeding survey, though now it is the responsibility of the hospitals (collecting birth data) and local authorities (in charge of public health practice since 2013). This move to local authorities for public health services has resulted in substantive financial cuts, meaning there is less resource to promote and sustain breastfeeding support - particularly in areas of most need. Other work has confirmed that users of FeedFinder map places to breastfeed in the most deprived areas of England (23% of venues reviewed were in the most deprived quintile, 16% were located in the least deprived quintile, with remaining venues distributed similarly across the three middle quintiles) [50], where we typically see less women breastfeeding overall. There could be a number of reasons for this including the IMD score of urbanised, city centre locations tending to be areas considered most deprived. However, we are unaware at this stage if the venues are mapped from women residing in these areas or if they are women who have travelled from outer city suburbs (generally considered less deprived).

Data generated by the FeedFinder App

FeedFinder launched in July 2013 and has seen an uptake of almost 10 000 users worldwide. For the context of this research, we present findings on the users within the UK only. Users are able to add a new venue and rate it based on four criteria; comfort, hygiene, privacy, baby facilities and also leave an additional text review. Formative design workshops with breastfeeding mothers identified these measures as key for contributing to a positive breastfeeding experience. 3444 total reviews were created by 606 users for 2535 venues over 43 months from May 2013 to January 2017.

Data Pre-processing

The initial step was to prepare the data for analysis, and integrate our VGI data, and in particular the review comments, with the associated data from the IMD. The IMD is the official measure of relative deprivation for small areas in England, based on the following domains: Income Deprivation, Employment Deprivation, Education, Skills and Training Deprivation, Health Deprivation and Disability, Crime, Barriers to Housing and Services and Living Environment Deprivation.

Using the geocoded data associated with each review venue, the data was cross-referenced with the IMD. The IMD decile for each venue location was established using the Lower-layer Super Output Areas (LSOA), neighbourhoods with populations <1500, and IMD rank data to provide contextual information on the relative deprivation of the location. IMD scores are ranked from 1 (most deprived) to 32,844 (least deprived) and in this paper will be described in relation to quintiles (20%). Our use of quintiles owes to the tradition of doing so in studies using the IMD (e.g. [35]). This facilitated an analysis of the textual data that could be then positioned in dialogue with an officially created dataset on contextual factors of deprivation. 2727 user reviews were included in the final dataset for analysis (717 entries were excluded on the basis that they

featured no comment data or because a postcode match was not generated, resulting in no associated IMD data).

Process

The process we followed was reflexive, and so the results are presented in line with the method. A preliminary frequency analysis was conducted on the corpus to assess whether there were empirically observed language-use patterns across IMD quintiles. Topic modelling was applied to explore the perceived distribution of topics within the dataset, as a data exploration tool in consultation with another researcher, and this in combination informed the resulting hand-coding of the dataset for two key privacy categories. The distribution of automatically extracted topics were measured across IMD quintiles, as were the distribution of the hand-coded themes.

Descriptive Statistics

Table 1 details the number of reviews classified as each of the IMD quintile together with the average number of characters and words. Quintile 1 refers to the most deprived and quintile 5 to the least deprived areas.

IMD Quintile	Number of Reviews	Mean no. of Characters	Mean no. of Words
1	634	135.39	22.75
2	673	143.63	24.26
3	682	139.37	23.91
4	435	154.06	26.08
5	303	149.81	25.63

Table 1. Number of and average character length of reviews

The majority of reviews were collected in most deprived quintiles (1 = most deprived, 5 = least deprived). The comments vary widely in length, ranging from 1 - 467 characters, with an overall mean character count of 144.45.

Differences in language choice by IMD score

In order to ascertain whether the language used to review venues differed across the most and least deprived venues a simple keyword analysis was conducted. Table 2 and table 3 detail the top 15 words most associated with the three least deprived and three most deprived deciles.

The keyword frequencies suggest that the language used to review venues differs in areas with different IMD ratings. Adjectives and descriptive language, such as ‘friendly’, ‘happy’ and ‘roomy’, surface as keywords in the three least deprived deciles. Together with ‘garden’ and ‘space’, it suggests that these reviews are commenting on largely positive qualities or provision requirements of the general environment.

Conversely the keywords for reviews in the more deprived areas seem generally related to basic provisions, as indicated through language closely tied to the practical experience of feeding, e.g. ‘chair’, ‘room’, ‘feeding’, and the logistics of finding (e.g. ‘located’, ‘top’) a suitable place to feed (e.g. ‘designated’, ‘separate’, ‘area’, ‘room’). A closer analysis of the frequency of the word ‘room’ in the review text of Quintiles 5 and 1 highlighted that the use of the word ‘room’

Word rank	Freq.	Keyness	Term
1	26	24.433	pub
2	40	20.867	disabled
3	22	20.507	ladies
4	42	15.602	mums
5	12	14.410	garden
6	7	14.038	weren
7	20	13.918	happy
8	13	13.370	make
9	170	11.741	friendly
10	49	11.692	space
11	8	10.678	health
12	15	10.439	buggies
13	52	9.693	lots
14	7	8.923	roomy

Table 2. Keywords associated with the three least deprived deciles

Word rank	Freq.	Keyness	Term
1	416	39.942	room
2	34	19.646	bolton
3	128	18.735	chair
4	41	18.509	nursing
5	49	16.637	rooms
6	17	15.542	scheme
7	55	14.525	separate
8	158	12.880	area
9	23	11.100	located
10	22	10.356	top
11	11	10.057	cooler
12	28	9.507	designated
13	40	9.358	bottle
14	359	9.336	feeding
15	10	9.142	mothercare

Table 3. Keywords associated with the three most deprived deciles

was significantly more present in Quintile 1 (raw frequency 278) than 5 (raw frequency 70), Log Likelihood of 34.69 (a Log Likelihood value of over 7 equates to 99% confidence of significance) ².

Topic modelling

Topic modelling was applied to the dataset using the Mallet Java implementation [31] of LDA, with stop words removed and sequence preserved. A hyperparameter optimization interval of 20, and 200 sampling iterations were specified for the production of 30 topics. The researcher experimented with 10, 20, 30, 40 and 50 topics and found that 30 provided a good level of granularity and coherence of topics without becoming unwieldy.

Table 4 displays 14 topics produced by LDA, order by how prevalent that topic is in the whole corpus (conveyed numerically by the ‘Distribution’ figure). Each topic is represented as a collection of words. It takes human effort to interpret and make sense of these and a human-readable label for each topic

²Calculations were carried out manually using Paul Rayson’s online log-likelihood calculator: <http://ucrel.lancs.ac.uk/llwizard.html>

added by the authors is supplemented in the ‘summary’ field. The results include some duplication and overlap of topics but it is possible to identify thematic grouping from the topics. 8 topics included ‘privacy’ or ‘private’ in the associated words list.

Making sense of the topics

The results of the LDA were discussed with two other researchers who had previously conducted a thematic analysis of the dataset. By talking through the topic keywords surfaced by LDA and looking at the reviews most highly associated with these topics we began a discussion around what constituted privacy. For each review a topic distribution was generated, so that a numerical value is provided to indicate the probability that a given ‘document’ is about each of the automatically generated topics. Table 5 provides an example review with the corresponding weighting for four most prominent topics.

In the original thematic analysis reviews ‘Privacy’ and ‘Designated areas’ were treated as unrelated concepts, with Designated areas representing a facilities concern. However, as we began to look through the topics which included words such as ‘privacy’ and ‘private’, it became apparent that the LDA topic models also featured words associated with designated areas. After consulting the corresponding reviews the decision to demarcate designated areas as separate to privacy concerns became questionable. However, the distinction between general comments of privacy and discretion, did seem differently motivated than the reviews which discussed the availability of a private space.

The concept of ‘privacy’ was not homogeneous and the concerns of privacy and the material affordances needed to support the individual requirements for privacy were not uniform. In terms of specific material elements taken purely from the keywords included in the topic models, we can extract the following provision considerations that inform the privacy requirements of mothers: ‘room’, ‘cubicle’, ‘door’, ‘lockable door’, ‘lock’, ‘curtain’, ‘wall’, ‘corners’, ‘chairs’, ‘seating’, ‘sofa’. The qualities of the space that are considered variously important, as identified from the LDA keywords, were: ‘hidden’, ‘quiet corner’, ‘separate’, ‘secluded’, ‘discreet’, ‘comfortable’, ‘comfy’, ‘spacious’, lighting (‘lit’, ‘dim’, ‘bright’), cleanliness (‘smelly’, ‘dirty’, ‘clean’, ‘cleanest’, ‘bin area’).

Topics and socio-economic factors

To assess whether the automatically extracted topics related to the socio-economic factors associated with a given venue, the variation of topic distribution for each review was analysed across IMD quintiles. A one way ANOVA analysis was conducted to assess which automatically generated topics were distributed significantly differently across IMD quintiles.

The ANOVA showed that of the total 30 topics, 14 had a significantly different distribution across groups. There was a statistically significant difference between IMD quintile groups for 13 of the total 30 automatically generated LDA topics as determined by one-way ANOVA. Details of f-scores and significance shown in table 6. A Tukey post hoc test revealed that the distribution of 9 topics were significantly different between IMD quintiles. Table 7 details the means and Standard

ID	Summary	Distribution	Keywords
22	Privacy & comfort	0.375	baby change comfy private facilities plenty privacy space chairs comfortable seating feeding room lots changing tables sofas seats good pram
14	Toilets & change facilities	0.195	toilet baby changing change disabled facilities table toilets loo clean ladies mat inside there's key station pram town spacious bathroom
27	Recreational provisions	0.116	children play area babies toys mums friendly kids breastfeeding small ots children's toddler older soft toddlers space great menu parents
3	Designated area	0.111	area feeding bottle chairs room water facilities floor toilet breast large separate private changing parent provided free breastfeeding warmers cooler
4	Welcoming atmosphere	0.108	breastfeeding friendly mums breastfeed babies staff welcoming happy centre issues time place supportive groups young families visit find mothers
23	Facilities & logistics	0.108	floor room baby change cafe toilets changing feeding area family ground located top lift upstairs space downstairs facility pushchair great
21	Food & positivity	0.078	coffee cafe free lovely cake cakes tea shop amazing welcoming run food drink delicious hot great home recommend nappies wonderful
17	Peer support	0.047	breastfeeding support group deal scheme group's closed peer mondays mums tuesday drop morning monday supporters sessions thursday long social clinic
9	Breastfeeding scheme	0.012	bolton friendly part breastfeeding cvs scheme delivered cafe training square community march barton premises tiffanys high covers
25	Separate room	0.294	room feeding changing baby chair comfy facilities feed separate clean private seats comfortable bit back facility sofa nice
1	Room with partition/lock	0.093	room door chair baby toilet lock changing cubicle lockable inside sign area pram downside main privacy sink people small
13	Comfort in private area	0.085	chairs rooms comfy nursing private rocking clean area side paper cubicles tub aren't dispenser spacious equipped curtains hidden wall
2	Cleanliness of BF area	0.035	nappy bins area smell easily doesn't basic difficult chair bags overflowing distracted due air low pleasant unpleasant changed

Table 4. LDA topics with significantly different distributions across IMD quintiles

Review	Topic	Composition
Only 1 room which was lockable but then a small cubicle inside that can be locked with a very cold and boring plastic chair with no arms sat in the corner Very uninviting! And was not helpful to all the other parents as there were people stood outside waiting by time I was ready to leave	T1	45%
	T19	16%
	T27	16%
	T3	4%

Table 5. Example review with LDA topic weightings

Deviation (SD) for those topics that were significantly differently distributed across the IMD quintiles. The results of all of the Post-hoc Tukey's HSD tests will be reported briefly here, followed by a summary of the key findings.

Post-hoc Tukey's HSD tests shows the mean weighting for Topic 1 (Room with partition/lock) was significantly higher in the most deprived, Quintile 1 compared to the least deprived Quintile 5 ($p=.02$), unless otherwise stated, the distributions across other quintiles was not significantly different from the others. Post-hoc Tukey's HSD tests shows the mean weighting for Topic 3 (Designated area) was significantly higher in the most deprived Quintile 1 compared to Quintiles 2 ($p=.01$), 4 ($p=.01$) and 5 ($p<.01$); Post-hoc Tukey's HSD tests shows the mean weighting for Topic 4 (Welcoming atmosphere) was

T	Selected key words	Anova	Sig.
1	Room with partition/lock	$F=(4,2722)=2.64$	$p=0.03$
2	Cleanliness of BF area	$F=(4,2722)=2.66$	$p=0.03$
3	Designated area	$F=(4,2722)=6.78$	$p<0.01$
4	Welcoming atmosphere	$F=(4,2722)=4.40$	$p<0.01$
9	Breastfeeding scheme	$F=(4,2722)=10.50$	$p<0.01$
13	Comfort in private area	$F=(4,2722)=3.37$	$p<0.01$
14	Toilets & change facilities	$F=(4,2722)=6.23$	$p<0.01$
17	Peer support	$F=(4,2722)=2.61$	$p=0.03$
21	Food & positivity	$F=(4,2722)=3.00$	$p=0.02$
22	Privacy & comfort	$F=(4,2722)=2.62$	$p=0.03$
23	Facilities & logistics	$F=(4,2722)=2.51$	$p=0.04$
25	Separate room	$F=(4,2722)=2.56$	$p=0.04$
27	Recreational Provisions	$F=(4,2722)=7.41$	$p<0.01$

Table 6. Details of the ANOVA results for topics T

significantly higher in the least deprived Quintile than Quintiles 1 ($p<.01$), 2 ($p<.01$), 3 ($p<.01$) and 4 ($p=.01$). Post-hoc Tukey's HSD tests shows the Topic 9 (Breastfeeding scheme) was significantly more characteristic of Quintile 1 than all other quintiles ($p<.01$). However, closer inspection of the reviews highlighted that the reviews classified in this category were authored by a small number of authors all relating to a particular scheme at a specific venue. Post-hoc Tukey's HSD tests shows the mean weighting for Topic 14 (Toilets & change facilities) is significantly lower in most deprived quintile than in Quintiles 3 ($p=.01$), 4 ($p<.01$) and 5 ($p=.01$). Post-hoc

Quintile	1		2		3		4		5		Total	
Topic	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	N
3	0.05	0.11	0.03	0.07	0.04	0.09	0.03	0.07	0.02	0.06	0.04	2727
4	0.03	0.08	0.04	0.08	0.03	0.08	0.04	0.07	0.05	0.11	0.04	2727
9	0.02	0.12	0.00	0.02	0.00	0.04	0.00	0.03	0.00	0.05	0.01	2727
14	0.05	0.07	0.06	0.09	0.06	0.09	0.07	0.10	0.07	0.11	0.06	2727
17	0.02	0.07	0.02	0.08	0.03	0.11	0.02	0.07	0.02	0.07	0.02	2727
21	0.02	0.07	0.02	0.05	0.02	0.06	0.02	0.06	0.03	0.08	0.02	2727
22	0.11	0.12	0.12	0.13	0.12	0.13	0.13	0.15	0.12	0.14	0.12	2727
23	0.04	0.08	0.04	0.08	0.03	0.08	0.03	0.07	0.02	0.04	0.03	2727
27	0.04	0.08	0.03	0.07	0.03	0.07	0.04	0.08	0.06	0.11	0.04	2727

Table 7. Mean distribution of selected topics across IMD quintiles

Tukey's HSD tests shows the mean weighting for Topic 17 (Peer support) differed significantly between Quintiles 4 and 3 ($p < .05$), with reviews in Quintile 3 more likely to mention support groups and schemes. Again, this topic was related to reviews written by a small and related set of authors in a particular area, promoting a particular scheme. Post-hoc Tukey's HSD tests shows the mean weighting for Topic 21 (Food & positivity) differed significantly between Quintile 2 and Quintile 5 ($p < .01$). Post-hoc Tukey's HSD tests shows the mean weighting for Topic 22 (Privacy & comfort) differed significantly between Quintile 1 and 4 ($p = .01$), but not between other quintiles. Post-hoc Tukey's HSD tests shows the mean weighting for Topic 23 (Facilities & logistics) was significantly lower in Quintile 5 than Quintile 2 ($p < .02$). Post-hoc Tukey's HSD tests show that the mean weighting for Topic 27 (Recreational provisions) was significantly higher in Quintile 5, the least deprived quintile, compared to every other quintile ($p < .01$).

In summary, reviews written about venues in the least deprived quintile were much more likely to comment on how supportive and friendly the environment was, how 'delicious' the food was, the availability of play areas for children and levels of comfort and seat quality. Reviews of venues in the most deprived areas were more likely to comment on the provision of a specific feeding area, discuss the availability of rooms, cubicles, curtains and locks. They more frequently referred to toilets and changing areas. Reviews about venues in the two most deprived quintiles also featured more spatial references (e.g. 'upstairs'), suggesting they included guidance on where feeding areas were located.

Hand coding comments for thematic content

In order to assess the validity of the automatically generated topics, the reviews were also hand coded for whether they discussed privacy. Based on the initial explorations with the data and the discussions conducted with the other researchers who conducted the original thematic analysis (as reported in [50]), reviews were coded for the following:

- Privacy: This includes any reviews that discuss discretion, finding nooks or seating positions which provide more privacy, and overt expression of concerns relating to privacy.
- Designated area: discusses the availability or quality of a designated feeding area (intentional/ad hoc).

Table 8 shows the results of the hand coding of reviews, detailing the raw figures for reviews that address privacy, a designated area (DA) for breastfeeding, and those that either address privacy or DA, per quintile. Figures shown in parenthesis represent the raw figures as a percentage of the total reviews per quintile.

IMD Quintile	Privacy	Designated Area (DA)	Privacy or DA
1	182 (28.71%)	183 (28.86%)	313 (49.37%)
2	195 (28.97%)	182 (27.04%)	319 (47.40%)
3	195 (28.59%)	137 (20.09%)	292 (42.82%)
4	131 (29.98%)	71 (16.25%)	188 (43.22%)
5	80 (26.36%)	34 (11.18%)	104 (34.32%)

Table 8. Number of comments coded as related to privacy or designated areas, or both, with percentage of total reviews given in parenthesis. Remaining reviews were coded as neither.

There are significantly more reviews which address either the provision of a designated areas for breastfeeding or privacy more generally in the most deprived Quintile (1), compared to the least deprived Quintile (5) ($\chi^2(2) = 17.53$, $p < 0.01$). However, the distribution of privacy related reviews is not significantly different in the least compared to the most deprived quintiles ($\chi^2(2) = 0.41$, $p = 0.52$). When considering reviews which discuss the provision of a designated breastfeeding area, we find that there are significantly more reviews which address the provision of a designated area for breastfeeding in the most deprived Quintile (1), compared to the least deprived Quintile (5) ($\chi^2(2) = 34.88$, $p < 0.01$). Our iterative analysis uncovered that, privacy, only when framed as relating to the availability of a separate private space, constitutes a greater priority for users of FeedFinder in venues which rank as more deprived, compared to those in less deprived areas.

These results reinforce the findings from the automatically generated topics. Furthermore, it highlights the significance of classification decisions. The relatively even distribution of privacy related reviews across all IMD quintiles suggests that there are a range of different notions of privacy, which although can be grouped under the same umbrella, attend to different priorities and concerns. For example, some reviews demonstrate an awareness of how different seating options can offer greater or lesser privacy or discretion, while others are focused less on discretion within a public setting and more

on the specific factors associated with the provision of a designated feeding area, whether there is a curtain or cubicle, if there is a door which is lockable and what the quality of the private area is.

Unpacking privacy in the context of FeedFinder

The criteria for classifying a review as privacy-related was not straightforward. While many reviews explicitly mentioned privacy, a great deal of others did not. Originally, reviews which discussed designated spaces for feeding and changing seemed to be a more descriptive account of facilities provision in the venues. In the thematic analysis the researchers also classified designated feeding areas separately from privacy discussion. Reviews which mentioned privacy aspects of designated areas were classified as ‘privacy-related’. For example:

Usually not a fan of the cubicle feeding venue but the 4 separate family rooms on the ground floor of Cabot Circus are great. Not too hot calming lighting and comfy seat more like a powder room at a posh bar or restaurant! Only niggle was being unable to work out the door lock so was interrupted twice.

[P]arents room at John Lewis. booths for breastfeeding are OK although there is no curtain and the chair there isn't very comfortable.

However, the decision made to discount those which referred to separate areas, as it was deemed more related to facilities provision, was revisited. Defining privacy as a researcher was complex - should the discussion of separate areas be recategorised as privacy? Does referencing the variability of available private spaces imply a desire for it or a mere observation thought to be of use to others? The ways in which this classification was conducted made a significant difference on how the result could be interpreted. If discussions of designated spaces that allow for feeding ‘in private’ was included, and privacy related reviews were recategorised on this boundary a different picture emerged. There was a correlation between designated areas and IMD, but not privacy alone nor privacy combined with designated places. So, deciding how to define privacy and the boundaries of classification became non-trivial.

From closer analysis of the review comments it becomes apparent that the rating scores are determined by the individual's concerns. So, for parents who are looking for a designated feeding area, there is a different hierarchy of concerns than for a parent who is more inclined to breastfeed publicly. Furthermore, the process of evaluation is shifted, while some were pleased that there was a private room which afforded them some privacy, others were much more motivated by the level of comfort within the room.

The focus on designated areas is arguably problematic in itself; that mothers feel the need for a separate, designated area implies (in line with prior studies) that many women are not comfortable breastfeeding in public. Despite the legal protection that supports mothers' rights to breastfeed in public, the lived reality of this experience reveals problematic power structures in the complex social and political reality of experiential accounts of place.

DISCUSSION

In this case study, we strived to bring qualitative and quantitative data together, and to bring empirically observed patterns in the data to bear upon an existing geographic information datasource, namely the IMD. Through this process new dimensions of the data were elucidated. The experiential data generated by breastfeeding parents, when introduced to the IMD data, surfaced distinct factors and practical concerns that were distributed differently in correlation with indexes of deprivation. This finding opens up more questions: are the values of individuals regarding privacy when breastfeeding differently motivated in such a way that is connected to socio-demographic factors? Are venues in more affluent areas better equipped with private breastfeeding areas so that parents do not need to mention it in reviews? Is it possible that these differences in review texts highlight the inequalities of parenting across the UK? That parents in areas of higher social deprivation strive to simply have access to basic amenities? While the causality remains elusive, this provides a set of practical recommendations for professionals looking to improve the levels of breastfeeding in public, and provides ‘evidence’ for health workers to take to local business owners, outlining the material needs of parents in their area.

Handling large-scale datasets typically requires, to some degree, a reductionist approach, which rallies against the feminist agenda to resist rigid categorisation. However, computers struggle with ambiguity and GIS must necessarily limit complexity as information needs to be plotted in a precise point on a map. However, by bringing together qualitative and quantitative data, multiplicitous accounts can simultaneously reside within or attach to a single given point. Topic models also attempt to reduce the dimensions of textual data, by identifying patterns in language use. However, topics are not mutually exclusive and are generated from the data itself rather than being imposed onto the data; in this way it aligns with the call to ‘rethink binaries’ [6] and gets closer to multiplicitous representation than other approaches to classification.

However, we acknowledge its limits. Ultimately, it offers another way of carving up the data into categories, but it also provides another lens on the data, and, we argue, offers a way to systematically evaluate and challenge existing structures. The topic models produced in this case study were not perfect. Some of this can be improved through adjusting the pre-processing of text before analysis; lemmatizing and stemming the text to reduce duplication and further tweaking the number of topics are two such adjustments that could be beneficial going forward. The output of the LDA is a list of words that requires interpretation in order to make sense; however, we do not see that as a limitation, but rather an important feature of a hybrid, human-computer approach. Furthermore, that the topic models produced by LDA feature some overlap and ambiguity is a natural result of working with human generated-data naturalistic textual data. After all, unlike an academic article or legal document the everyday language is less precise; and that is what makes it interesting but also challenging to work with.

Implications for Feminist GIS in HCI

How might the methods outlined above facilitate or be integrated into an application of a feminist, reflexive human-computer-collaborative approach to qualitative data? In the following section we highlight the implications for other researchers working with GIS datasets seeking to better understand how the systemic and experiential making of places.

Privileging Multiplicity

To refer back to the literature, analysing experiential data should ‘emphasize representational strategies premised on multiplicity rather than binaries’ [6]. We argue that topic models facilitate a non-binary approach because they describe instances within the data as a composition of multiple topics. Furthermore, the topic models facilitated a more nuanced interpretation of how privacy concerns manifest for different parents and the associated material requirements necessary to meet such needs.

Dataset Selection

Feminist GIS stresses that wherever possible, researchers should strive to create their own datasets and not only rely on existing, organisational/institutional top-down datasets. Using a VGI dataset created by the people whose experiences we were examining was imperative and afforded insights that could not have been captured via other methods.

Interrogate Existing Datasets

In this case study we demonstrate how VGI datasets can be brought together with existing institutional data to create new questions, expand understandings and reposition the centrality of experience in the interpretation of material spatial realities. Rather than dismissing the usefulness of the IMD, we were able to ask questions of it and generate new understandings by bringing together these different datasets.

Question Whose Voices are Heard

Critical GIS must be ‘reflexive in pro-actively questioning its own inclusivity’ [52], and this requires researchers to constantly reflect upon whose voices are being included and whose are not. Due to the way FeedFinder was co-designed with the communities it serves, as well as its deployment as a mobile application with a practical usage, the data collected includes more heterogeneous knowledge from a diverse group of participants [12]. The approach taken also enabled contradictions to inform new understandings [12] and highlight multiple subjectivities within the data. The dataset that we were working with was important in facilitating this.

Question Biases

By incorporating a data-driven empirical analysis using automatically produced topic models the researchers discovered new patterns within the data, that challenged their own assumptions about how ‘privacy’ was communicated through the reviews and the significance of designated areas for breastfeeding.

Transparency About Partial Pictures

In our description of the steps followed in the case study we have attempted to be transparent about the process, acknowledge ambiguities and tensions, and move back and forth

between readings. Kwan [27] argues that the ‘presentation of quantitative data should be accompanied by a description of the ambiguity or problems of the classification scheme’. Consequently, the findings are not clear cut, but we hope that fuller understanding of the experiences of breastfeeding parents in different socio-economic contexts has emerged through the process.

CONCLUSION

In this paper we presented a combination of computational and human approaches for textual analysis, suitable for large-scale experiential data. We have demonstrated the underlying methods and challenges behind exploring and analysing a real-life VGI dataset and positioning this experiential, qualitative dataset (FeedFinder reviews) in dialogue with a structural, quantitative dataset (IMD). We position this work alongside other efforts to develop a Feminist HCI methodology, as it begins to unpick the challenges and opportunities of dialectic information gathering, exploring how a heterogeneity of data sources and interpretations can be employed [5] and combined with computational approaches. An essential a priori step in achieving this is clearly to do with data collection. In order to empirically examine the ‘knowledges among very different and power-differentiated communities’ [17], we must first establish access to accounts of situated knowledge. As demonstrated by the work of Balaam et al. [2] and FeedFinder, this is possible.

We demonstrated that topic models can identify novel and insightful patterns in such data, providing an empirically grounded entry-point for researchers. This approach facilitated the experiential accounts of place to be brought to bear upon the purely quantitatively driven IMD data. This novel combination of existing structural data and experiential VGI data enables us to systematically interpret the impact of material-socio-economic factors on people’s experiences and concerns when breastfeeding in public. In particular, a key finding of this case study re-frames what breastfeeding parents may mean by privacy, and the material-factors, such as a separate room, cubicle, curtain or lockable-door, that populate the various individual regions of concern. Furthermore, we empirically show, that these concerns are systematically different depending on the contextual factors of deprivation in the area as captured by the IMD. As such, this paper contributes a reflexive process in which objective interpretation of data which preserves multiple subjectivities (as afforded through topic models) is positioned in dialogue with a more qualitative reading of the data, and through which user-generated data is used to ask questions of official data and vice versa.

ACKNOWLEDGEMENTS

This research was funded by EPSRC awards EP/M023001/1 (The Digital Economy Research Centre) and EP/L016176/1 (Centre for Doctoral Training in Digital Civics). Data supporting this publication is openly available under an ‘Open Data Commons Open Database License’. Additional metadata are available at: <http://dx.doi.org/10.17634/141304-11>. Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for access instructions.

REFERENCES

1. 2015. English indices of deprivation 2015. (2015). <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>
2. Madeline Balaam, Rob Comber, Ed Jenkins, Selina Sutton, and Andrew Garbett. 2015. FeedFinder: a location-mapping mobile application for breastfeeding women. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1709–1718.
3. Mara Balestrini, Yvonne Rogers, and Paul Marshall. 2015. Civically engaged HCI: tensions between novelty and social impact. In *Proceedings of the 2015 British HCI Conference*. ACM, 35–36.
4. Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1301–1310.
5. Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 675–684.
6. David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
7. Elizabeth Bonsignore, Alexander J Quinn, Allison Druin, and Benjamin B Bederson. 2013. Sharing stories “in the wild”: A mobile storytelling case study using StoryKit. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 3 (2013), 18.
8. Shauna Concannon and Matthew Purver. 2014. Interpreting Arts Audiences and Cultural Preference Through Twitter Data. In *Digital Intelligence*. Nantes, France.
9. Daniel Dorling. 1998. Human cartography: when it is good to map. *Environment and Planning A* 30, 2 (1998), 277–288.
10. Kenneth J Dueker. 1987. Geographic information systems and computer-aided mapping. *Journal of the American Planning Association* 53, 3 (1987), 383–390.
11. Jacob Eisenstein, Duen Horng Chau, Aniket Kittur, Eric P Xing, and others. 2011. Topicviz: Semantic navigation of document collections. *arXiv preprint arXiv:1110.6200* (2011).
12. Sarah Elwood. 2008. Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal* 72, 3–4 (2008), 173–183.
13. Public Health England. 2015. New mothers are anxious about breastfeeding in public. (2015). <https://www.gov.uk/government/news/new-mothers-are-anxious-about-breastfeeding-in-public>
14. Robin Flowerdew. 1998. Reacting to ground truth. *Environment and Planning A* 30, 2 (1998), 289–301.
15. Alastair J Gill, Saba Hinrichs-Krapels, Tobias Blanke, Jonathan Grant, Mark Hedges, and Simon Tanner. 2017. Insight workflow: Systematically combining human and computational methods to explore textual data. *Journal of the Association for Information Science and Technology* (2017).
16. Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* 101, suppl 1 (2004), 5228–5235.
17. Donna Haraway. 1988. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14, 3 (1988), 575–599.
18. Steve Harrison and Paul Dourish. 1996. Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*. ACM, 67–76.
19. Clint Heyer and Margot Brereton. 2010. Design from the everyday: continuously evolving, embedded exploratory prototypes. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. ACM, 282–291.
20. Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 50–57.
21. Phil Hubbard, Brendan Bartley, Duncan Fuller, and Rob Kitchin. 2002. *Thinking geographically: Space, theory and contemporary human geography*. A&C Black.
22. Don Ihde. 1995. *Postphenomenology: Essays in the postmodern context*. Northwestern University Press.
23. Rob Kitchin and Tracey P Lauriault. 2014. Towards critical data studies: Charting and unpacking data assemblages and their work. (2014).
24. Lauren Klein and Jacob Eisenstein. 2013. Reading Thomas Jefferson with TopicViz: towards a thematic method for exploring large cultural archives. *Scholarly and Research Communication* 4, 3 (2013).
25. LaDona Knigge and Meghan Cope. 2006. Grounded visualization: integrating the analysis of qualitative and quantitative data through grounded theory and visualization. *Environment and Planning A* 38, 11 (2006), 2021–2037.
26. Mei-Po Kwan. 2002a. Feminist visualization: Re-envisioning GIS as a method in feminist geographic research. *Annals of the association of American geographers* 92, 4 (2002), 645–661.
27. Mei-Po Kwan. 2002b. Quantitative methods and feminist geographic research. *Feminist Geography in Practice: Research and Methods* (2002), 160–73.
28. Mei-Po Kwan. 2004. Beyond difference: From canonical geography to hybrid geographies. *Annals of the Association of American Geographers* 94, 4 (2004), 756–763.

29. Koen Leurs. 2017. feminist data studies: using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review* 115, 1 (2017), 130–154.
30. Fiona McAndrew, Jane Thompson, Lydia Fellows, Alice Large, Mark Speed, and Mary J Renfrew. 2012. Infant feeding survey 2010. *Leeds: Health and Social Care Information Centre* (2012).
31. Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. (2002).
32. Sara L McLafferty. 1995. Counting for Women. *The Professional Geographer* 47, 4 (1995), 436–442.
33. Sara L McLafferty. 2002. Mapping women's worlds: Knowledge, power and the bounds of GIS. *Gender, Place and Culture: A Journal of Feminist Geography* 9, 3 (2002), 263–269.
34. Chrystie Mykietiuk, Shauna Concannon, and Paul Curzon. 2017. Narrative perspective, person references, and evidentiality in clinical incident reports. *Journal of Pragmatics* 117 (2017), 139–154.
35. Laura L Oakley, Mary J Renfrew, Jennifer J Kurinczuk, and Maria A Quigley. 2013. Factors associated with breastfeeding in England: an analysis by primary care trust. *BMJ open* 3, 6 (2013), e002765.
36. Stan Openshaw. 1998. Towards a more computationally minded scientific human geography. *Environment and Planning A* 30, 2 (1998), 317–332.
37. David O'Sullivan. 2006. Geographical information science: critical GIS. *Progress in Human Geography* 30, 6 (2006), 783–791.
38. Michael Paul and Roxana Girju. 2009a. Cross-cultural analysis of blogs and forums with mixed-collection topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*. Association for Computational Linguistics, 1408–1417.
39. Michael J Paul and Roxana Girju. 2009b. Topic Modeling of Research Fields: An Interdisciplinary Perspective.. In *RANLP*. 337–342.
40. Marianna Pavlovskaya. 2006. Theorizing with GIS: a tool for critical geographies? *Environment and Planning A* 38, 11 (2006), 2003–2020.
41. Marco Pennacchiotti and Ana-Maria Popescu. 2011. A Machine Learning Approach to Twitter User Classification. *ICWSM* 11, 1 (2011), 281–288.
42. John Pickles. 1995. *Ground truth: The social implications of geographic information systems*. Guilford Press.
43. Daniel Ramage, Susan T Dumais, and Daniel J Liebling. 2010. Characterizing microblogs with topic models. *ICWSM* 10 (2010), 1–1.
44. Daniel Ramage, Christopher D Manning, and Susan Dumais. 2011. Partially labeled topic models for interpretable text mining. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*. ACM, 457–465.
45. Daniel Ramage, Christopher D Manning, and Daniel A McFarland. 2010. Which universities lead and lag? Toward university rankings based on scholarly output. In *Proc. of NIPS Workshop on Computational Social Science and the Wisdom of the Crowds*.
46. Alan Ritter, Colin Cherry, and Bill Dolan. 2010. Unsupervised modeling of twitter conversations. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 172–180.
47. Jennifer A Rode. 2011. A theoretical agenda for feminist HCI. *Interacting with Computers* 23, 5 (2011), 393–400.
48. Nadine Schuurman and Geraldine Pratt. 2002. Care of the subject: Feminism and critiques of GIS. *Gender, Place and Culture: A Journal of Feminist Geography* 9, 3 (2002), 291–299.
49. Eric Sheppard. 2001. Quantitative geography: representations, practices, and possibilities. *Environment and Planning D: Society and space* 19, 5 (2001), 535–554.
50. Emma Simpson, Andrew Garbett, Rob Comber, and Madeline Balaam. 2016. Factors important for women who breastfeed in public: a content analysis of review data from FeedFinder. *BMJ open* 6, 10 (2016), e011762.
51. Alex Taylor. 2016. Data,(bio) sensing, and (other-) worldly stories from the cycle routes of London. *Quantified: Biosensing Technologies in Everyday Life* (2016), 189.
52. Jim Thatcher, Luke Bergmann, Britta Ricker, Reuben Rose-Redwood, David O'Sullivan, Trevor J Barnes, Luke R Barnesmoore, Laura Beltz Imaoka, Ryan Burns, Jonathan Cinnamon, and others. 2016. Revisiting critical GIS. *Environment and Planning A* 48, 5 (2016), 815–824.
53. Anton Törnberg and Petter Törnberg. 2016a. Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. *Discourse & Society* 27, 4 (2016), 401–422.
54. Anton Törnberg and Petter Törnberg. 2016b. Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media* 13 (2016), 132–142.
55. Yi-Chia Wang, Moira Burke, and Robert E Kraut. 2013. Gender, topic, and audience response: an analysis of user-generated content on facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 31–34.
56. Gregor Wiedemann. 2016. Computer-assisted text analysis in the social sciences. In *Text Mining for Qualitative Data Analysis in the Social Sciences*. Springer, 17–54.

57. Anbang Xu, Xiao Yang, Huaming Rao, Wai-Tat Fu, Shih-Wen Huang, and Brian P Bailey. 2014. Show me the money!: An analysis of project updates during

crowdfunding campaigns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 591–600.