The Canadian Geographer
Le Géographe canadien
Canadian Association of Geographers
Association Canadienne des Géographes

Code clouds: Qualitative geovisualization of geotweets

Jin-Kyu Jung
School of Interdisciplinary Arts and Sciences, University of Washington-Bothell

The popularity of geotagged social media has provided many research opportunities for geographers and GIScientists in the digital age. This article reviews innovative approaches to studying spatially linked social media, and applies lessons taken from qualitative GIS and geographic visualization to improve these approaches. I introduce the idea of "code clouds" as a potential technique for the qualitative geovisualization of spatial information. Code clouds can depict and visualize analytic codes, or codes identifying key ideas and themes, that are generated through digital qualitative research. Rather than transforming qualitative forms of data into categories or numbers, code clouds attempt to preserve and represent the context of data as a visualized outcome of qualitative analysis. I use examples from an exploratory case study of geotweets in King County, WA, to demonstrate how code clouds can be applied to the production of meanings through qualitative geovisualization.

Keywords: code clouds, qualitative geovisualization, qualitative research, geotweets

Les nuages conceptuels : la géovisualisation qualitative des géogazouillis

À l'ère numérique, la popularité des médias sociaux géolocalisés ouvre la voie à de nombreuses possibilités de recherche pour les géographes et les spécialistes des SIG. Cet article passe en revue les nouvelles approches pour l'étude des médias sociaux fixés dans l'espace, puis tire des enseignements des SIG qualitatifs et de la visualisation géographique en vue d'améliorer ces approches. La notion de « nuage conceptuel » est dépeinte comme une technique pouvant servir à la géovisualisation qualitative de l'information spatiale. Les nuages conceptuels peuvent décrire et illustrer des codes analytiques ou des codes d'identification des idées et thèmes principaux qui se dégagent des études numériques privilégiant une approche qualitative. Plutôt que de transformer les bases de données qualitatives en des catégories ou des numéros, les nuages conceptuels tentent de préserver et de représenter le contexte de données en tant que résultat illustré à l'origine de l'analyse qualitative. Je m'appuie sur des exemples tirés d'une étude de cas exploratoire menée sur des géogazouillis dans le comté de King, État de Washington, afin de montrer de quelle manière les nuages conceptuels peuvent soutenir la production de significations à travers la géovisualisation qualitative.

Mots clés : nuages conceptuels, géovisualisation qualitative, recherche qualitative, géogazouillis

Introduction

People are living in a world inundated with data, to such an extent that researchers have begun exploring the importance of concepts including digital icebergs and Big Data (Gray 2007; Hey et al. 2007). Unprecedented quantities of data are instantly and continuously created and shared. Social media such as Twitter and Facebook contribute to these digital icebergs and Big Data phenomena. Popular social media seem to be the most effective way to get people interested in particular topics and events, or to get people to promote a product and stay in contact with distant friends and colleagues (Miller 2011). In 2011 Twitter users numbered more than 100 million, and they were posting 230 million tweets per day (Reisinger 2011). A year later, in 2012, the number of daily tweets increased dramatically to 500 million (Terdiman 2012). The US Library of

Correspondence to/Adresse de correspondance: Jin-Kyu Jung, School of Interdisciplinary Arts and Sciences, University of Washington-Bothell, 18115 Campus Way NE, Bothell, WA 98011-8246. Email/Courriel: jkjung5@uw.edu

DOI: 10.1111/cag.12133

Congress also committed to archiving all public tweets, thereby acknowledging the importance of social media for the study of social, cultural, economic, and political trends in contemporary society (Raymond 2010). However, the increasing amount and accessibility of Big Data also create a challenge for analyzing and representing them. Some researchers have approached this issue by developing systematic and algorithmic ways of analyzing data (Thatcher 2014). However, questions concerning how to adequately extract meaning from these large and messy data sources, and then how to represent this meaning, still remain (MacEachren et al. 2011).

Studying social media is a relatively new area for geographers, but is emerging as an important research focus (Miller 2011; Graham et al. 2013). It is a timely interest considering the fact that many social media are now spatially linked through georeferencing or geotagging. For example, Twitter is a popular microblog site for users to write a short 140-character status update, and to build a profile including a photo, ID, biographic data, and website (Fitton et al. 2009). Geotagged tweets can be mapped in real time, leading scholars to explore the best methods for tracking, monitoring, and visualizing this data (Erickson 2010; Kumar et al. 2011; Butts et al. 2012; Daraganova et al. 2012; Quercia et al. 2012; Takhteyev et al. 2012; Gundersen 2013). These studies demonstrate the potential implications of geographic data and spatial variability for the study of social media. Twitter is representative of web 2.0 desires to read, write, and share personal information (Schuurman 2009), and the association of locational information with tweets makes the application a form of maps 2.0 in which map readers also become map creators (Crampton 2009; Elwood 2010a). If we could develop ways to visualize spatially embedded social networks, we might be able to represent the everyday experiences of people in real time (Livehoods 2012).

Many of the examples in this paper explore this new convergence of GIS, geovisualization, and social media. Sui and Goodchild (2011) predicted that the rapid growth of the geographic webs and location-based social media would present opportunities as well as challenges for GIScientists and geographers. For instance, in the past GIScientists have tended to have access to either what Manovich (2011a) refers to as surface data (large volumes of quantitative data), or what he refers to as deep data (thick, ethnographic data often associated with qualitative data). Sui and Goodchild (2011) argue that social media combines both volume and depth. While this increased access to rich and plentiful data brings great opportunities, it also brings methodological challenges. We must still use the proper methods to extract meaning from the appropriate type of data. The problem is no longer the acquisition of Big Data, but the selection of appropriate methods to effectively understand and analyze that data. As Sui and Goodchild (2011, 1741) wrote, this is the moment when we begin to reconsider the use of “deep data for and about many.” Twitter data can be rich in detail about societal and spatial trends, but in order to realize this potential, we need to critically reflect upon the methods we use to analyze, interpret, and visualize them. I introduce the idea of “code clouds” as a method of analyzing and representing qualitative data (Big Data). Code clouds can depict and visualize analytic codes that represent key ideas and repeated themes generated from qualitative research. Rather than quantitatively transforming qualitative forms of data, code clouds represent the context of data as a visualized process and outcome of qualitative analysis, and they have the potential to be used with spatial information.

I begin in the following section with a review of current geographic approaches to mapping social and spatial media. I will argue that these innovative approaches may not sufficiently capture the context of social media, and literature on qualitative methods, geographic visualization, and qualitative GIS might help us reconfigure these geographic approaches to mapping social media. I then propose and explain the concept and application of code clouds as a method of integrating qualitative research and geographic visualization. Code clouds will be introduced to show the process and results of qualitative research through the visualization of analytic and interpretive codes. Examples of code clouds will then be presented from an exploratory case study using a 5 percent sample of all geotweets[1] in King County, Washington, during October 2012. These cases are not intended to show a full account of code clouds, but rather to demonstrate how this

[1] Geotweets are geographically referenced tweets that include locational information about where the tweet was authored. Leetaru et al. (2013) found that more than 3 percent of all tweets have location information available since Twitter first allowed tweets to include geographic data in 2009.

approach is performed to explore and visualize qualitative aspects of spatially integrated social media, and how it can foster discussions of qualitative geovisualization.

Geographic analysis of social media and qualitative geovisualization

New scholarship on social media has developed out of various theoretical and analytical sets of literature. Scholars have paid greater attention to the increasingly location-based nature of social media,[2] and efforts have been made to develop new geospatial applications designed to research these spatially linked social media (Kwak et al. 2010; Kumar et al. 2011; MacEachren et al. 2011; Sui and Goodchild 2011; Quercia et al. 2012; Tsou and Yang 2012; Graham et al. 2013).

Politics is one area in which we clearly see the influence of social media. Recently, social media have become essential for successful campaigns, and their use can greatly boost a candidate's popularity in politics (Sharif 2012). For instance, during the 2012 U.S. Presidential election candidates used social media to sell their campaigns to voters and to try to be more visible in online spaces. Seattle's local newspaper described this phenomenon as a manifestation of the "Twitterverse," where presidential debates came to resemble reality shows in which virtual audiences could cast ballots in the form of posting their opinions through social media (Parker 2012). Locally, the *Seattle Times* collected tweets and created a graph of "Favorable tweets about Obama and Romney," in which tweets favouring each presidential candidate were plotted and compared based on data from the Seattle Metropolitan area between 09/05/12 and 10/25/12. This graph was meant to represent trends of favourable tweets about presidential candidates, but it was not entirely clear how a tweet came to be classified as favourable during this process. 'Favor' is a highly subjective word that can be difficult to interpret and analyze.

A majority of social media research has focused on user-generated keywords or hashtags. A hashtag is the use of the '#' symbol to mark keywords or topics in a tweet, and it is often used to search, retrieve, and delineate information on social media sites (Small 2011; Tsou and Yang 2012; Twitter). Hashtags are central to finding key themes in tweets because it makes it possible to filter tweets. Hashtags are used on more than 70 percent of the Twitter accounts examined (Small 2011), and websites like #hashtags (www.hashtags.org) even automatically track and display hashtags in real-time. Small's (2011) research on Canadian politics on Twitter demonstrates the effective use of hashtags to study political conversations, political participation, and the nature of tagged tweets. Hashtags present a powerful, efficient, and systematic way to research social media; however, I will argue that we cannot fully see the contextual meanings of tagged/tweeted data with only hashtags search.

Another key development in spatially linked social media research entails the convergence of social network analysis, geographic analysis, and geographic visualization. Billions of real-time tweets are now captured, explored, and visualized through newly developed interfaces and applications. Tweets are grouped by users, topics, times, and locations, and are drawn from local to global scales (Kwak et al. 2010; Bhattacharjee 2013; Gundersen 2013). Kwak and co-researchers (2010) crawled the entire Twitter site in Korea, and mapped more than 1.47 billion social relations and thousands of trending topics for half a billion users. Geographic and social relationships are often computed together and visualized, and these cases demonstrate the geographic and geovisual turn in social network and social media research (Erickson 2010; Daraganova et al. 2012; Quercia et al. 2012; Takhteyev et al. 2012). Efforts to visualize social media on maps have also incorporated temporal and topical perspectives. For instance, prototypes of web-based geovisual analytic approaches have been developed to leverage Twitter in support of crisis management (Kumar et al. 2011; MacEachren et al. 2011). TweetTracker[3] by Humanitarian Aid and Disaster Relief (HADR), Ushahidi (http://www.ushahidi.com/), and SensePlace2 are all examples of visually enabled interfaces which support the understanding of spatial and temporal patterns identified through

[2] Many mainstream social media are now spatially linked (e.g., Foursquare, Flickr, Twitter, Yelp, Facebook, etc.).

[3] There are also similar mashup applications such as *TrendsMap* (http://trendsmap.com/), *Twitalyzer* (http://twitalyzer.com/), and *geotwitterous* (http://ouseful.open.ac.uk/geotwitterous/).

the analysis of geo-located tweets (MacEachren et al. 2011; Kumer et al. 2012). Most recently, Graham and co-workers (2013) extended the discussion of spatial analysis of social media to “geographic visualization.” Their spatially aware Treemap (http://www.treemap.com) visualizes both the number of tweets emanating from every country (by size), and the number of geocoded tweets as a proportion of that country’s Internet population (by shaded colour), to demonstrate the disparity among countries in terms of the use of tweets using the power of geographic visualization and its engagement with social media.

Many efforts have been made to expand the geographical analysis of social media beyond simply placing points of social media data on maps. We can detect the spatial distribution of tweets in relation to particular topics or places, and search for any meaningful spatial patterns/correlation with more advanced geographic analysis. However, the current discussions of geographic analysis of social media have not fully included qualitative and interpretive methods. Elwood (2009, 2010b) further argues that an emerging need is the handling of qualitative forms of spatial knowledge and human expression of spatial relationship, and she suggests that we consider adopting methods from works on critical visual methodologies.

Much of the research into visualization and geovisualization is associated with quantitative data, with fewer discussions of analyzing and visualizing qualitative data. One obvious exception is that of Self-Organizing Maps (SOM). SOM is based on an unsupervised learning algorithm using a collection of typically 2D nodes. Documents are located at these nodes, and the topological relationships between nodes are preserved. SOM allows visualizing 2D and 3D surfaces in order to reflect how data are distributed (Chen 2004; Leuthold et al. 2007; Skupin and Borner 2007; Hipp et al. 2012). SOM exemplifies the increasingly recognized power of visualization to highlight characteristics of data and to spatialize non-spatial data (Skupin and Fabrikant 2003). Leuthold et al. (2007) offer another example of the spatialization of non-spatial data with their generation of a 3D semantic space that represents the political landscape of Switzerland.

Manovich (2011b) introduces a new method for the visualization of media, called “direct visualization.” Direct visualization comes from his earlier work on “meta-media” (Manovich 2005; Lapenta 2012), in which a meta-media object contains both the original media structure (e.g., an image of the city) and the computer program that allows the user to generate descriptions of this structure (e.g., new 3D navigable reconstructions of the city with images). Manovich (2002) once defined visualization as a transformation of quantified data, which by itself is not visual, into a visual representation. However, he later redefined visualization to include data that are already visual such as text, photos, and videos. He argues that we now “create new visual representations of the ‘original’ (visual) data without translating them into graphic signs” (Manovich 2011b, 45). Direct visualization is visualization without the reduction and abstraction of data, and it gives us a new way of navigating, experiencing, and representing data, especially with an original form. Additionally, it offers a new possibility for qualitative geovisualization, and demonstrates the importance of preserving the context of data in the visualization process. It also shows us that visualized data need not be limited to numbers or geometric primitives. Rather, the visualized data could be multi-format qualitative data such as text, images, and video.

Code clouds: Integration of qualitative research and geographic visualization

Code clouds function to reflect and represent the context of data as a visualized outcome of qualitative analysis. While code clouds are related to content clouds, the concept is also influenced by discussions of qualitative GIS and geographic visualization. Qualitative GIS focuses on the integration of GIS and the qualitative. This scholarship reconsiders the importance of qualitative data and qualitative approaches with GIS by reconfiguring GIS itself (Cope and Elwood 2009; Knigge and Cope 2009; Wilson 2009). According to Cope and Elwood (2009), qualitative GIS does not simply incorporate non-numerical data, but includes rich descriptions of data, including data containing interpretations of situations and processes. Therefore, qualitative GIS is constituted by the integration of GIS and geographic visualization with situated, interpretive, and qualitative analysis. Qualitative GIS is associated with various qualitative methods including content analysis, grounded theory, discourse analysis, and

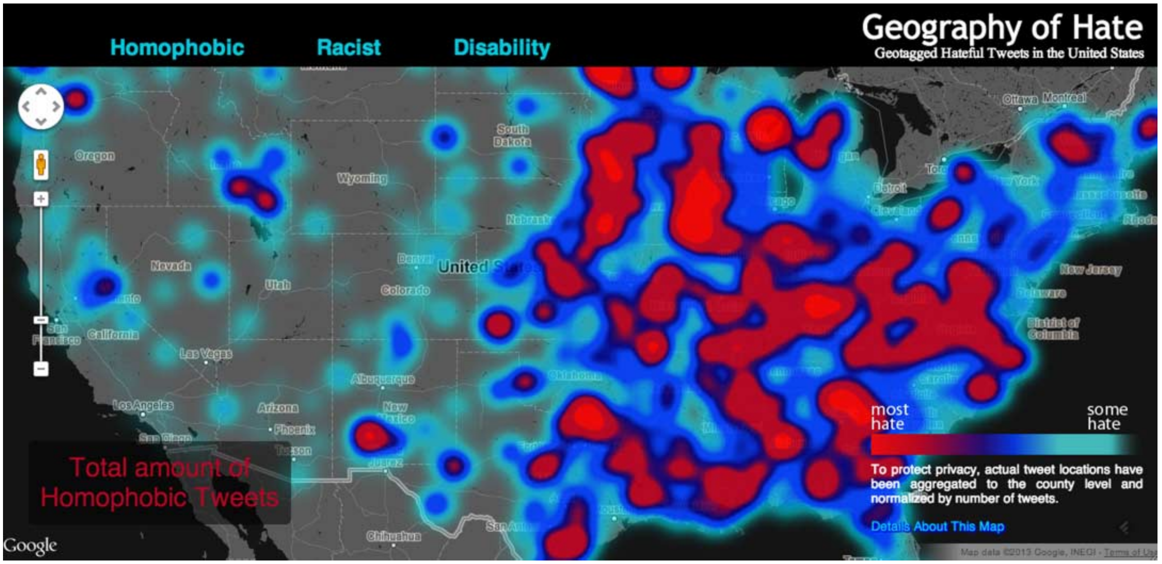

Figure 1
Hate heat map showing total number of homophobic tweets in the United States (http://users.humboldt.edu/mstephens/hate/hate_map.html)

visual analysis (Pavlovskaya 2002; Cieri 2003; Gahegan and Pike 2006; Kwan 2007). We can extend these discussions of qualitative GIS to qualitative geovisualization with code clouds.

Researchers and students at the Humbolt State University collected all of the geotagged tweets in the United States from June 2012 to April 2013 and created a heat map called "Geography of Hate" (Bhattacharjee 2013) (Figure 1). Hateful racist and homophobic tweets were categorized as positive or negative and aggregated at the county-level before being plotted on an interactive map.

This map allows readers to compare places with disproportionately high or low amounts of hateful tweets, and it is an example of geovisualizing the qualitative analysis of geotagged tweets. However, it fails to take into consideration the context of these tweets, and is mapped simply by the frequency of categories. For example, it is not clear how a term like "fag" or "faggot" was used, or whether people were using the "N word" in a racist manner or not. This example demonstrates the limitations of categorization and generalization, resulting in a poor visualization of the contexts of the tweets. Qualitative research is intended to capture the meaning and themes of data, and it requires continuous iterations of exploration and interpretation of how qualitative data are presented (Glaser and Strauss 1967; Davidson and Di Gregorio 2011). The context of tweets should not be sacrificed for the purpose of geographic visualization.

Cidell (2010) introduces a form of geographic visualization that does not necessarily require the transformation of qualitative data. Cidell (2010) argues that content clouds and mapping in GIS can create a spatial representation that offers a quick overlook of a dataset. For example, Figure 2 shows an analysis of the transcripts of public meetings concerning the purchase of a beltline railroad. A content cloud is a visual method of representing text data. The importance of each tag or word is symbolized with a font size or colour that is proportional to the frequency of the word (Ward et al. 2010).[4] Content clouds with an accompanying GIS map reveal a significant difference in terms of what participants discussed at meetings in different

[4] There are various visualization tools for visualizing texts such as tagClowd.com or Wordle.net. Jigsaw is a text visualization tool that was developed by John Stasko at the Georgina Institute of Technology (Ward et al. 2010). It uses several different views to present information such as graph, scatterplot, time line views, and texts. There are also other visualization options for text such as the WorldTree and TextArc visualization tool.

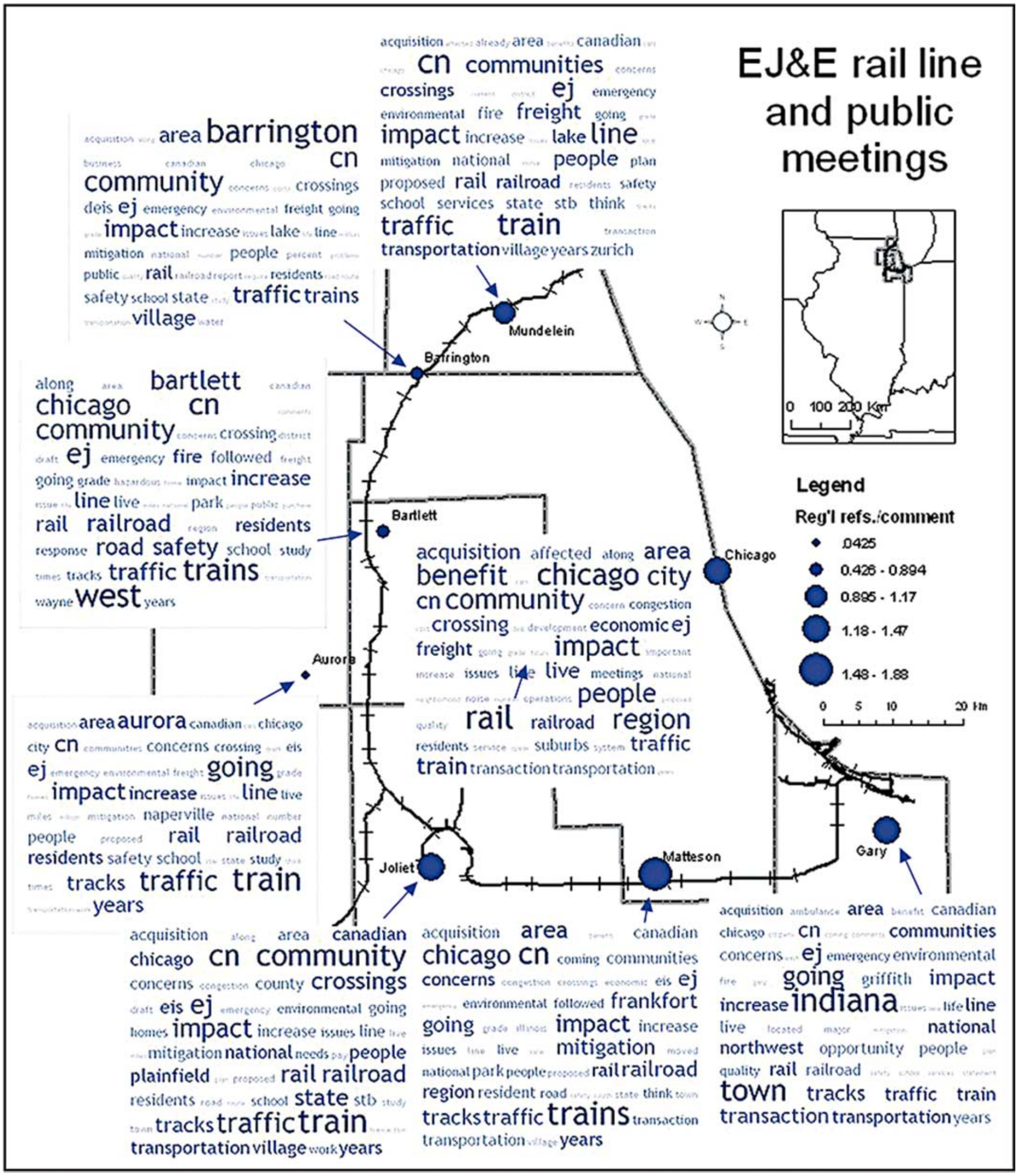

Figure 2
Content clouds and GIS: Cidell's (2010) "Content clouds from CN/EJ&E public meeting transcripts"

locations. This simple but inspiring integration of content clouds with GIS provides a potential prototype approach for integrating geovisualization, qualitative data, and qualitative analysis with spatially linked social media. Content clouds are also examples of direct visualization (Manovich 2011b) because data (words) in content clouds are visualized with the original form of data without any transformation.

Methodologically, content clouds are rooted in content analysis.[5] Content analysis is a technique for examining information in written or symbolic materials (Neuman 1997). It often refers to a quantitative measure of keywords that are typically deployed (Cidell 2010; Altheide and Schneider 2013), but it can also be a practical method for exploring massive amounts of social media and e-research documents in a digital era (Anderson and Kanuka 2003; Gray 2007; Hey et al. 2007). However, content clouds have several limitations. For example, content clouds are affective because they immediately make readers think that larger items are the most important pieces of information. This may not be the case at all, because they only illustrate counts and frequencies of words. Therefore, content clouds are susceptible to representing the most common words in a dataset, and not necessarily the most important ones. The importance of qualitative data lies in their potential to reveal context, and they should be "a source of well-grounded, rich descriptions, and explanations of processes in identifiable local contexts" (Miles and Huberman 1994, 1; Geertz 1973). Cope and Elwood's (2009) understanding of qualitative GIS emphasizes that qualitative data are important insofar as they allow us to pull out meaningful insights from them. I believe that codes, and especially analytic codes, are important for the representation of the contexts of qualitative data and the outcomes of qualitative analysis, and that they are an integral part of code clouds.

The coding process is the heart of qualitative analyses, especially in grounded theory (Strauss and Corbin 1997). Grounded theory is a method or theory used to inductively generate theories from empirical data that have been systematically gathered. The research begins with an area of study and allows "theory to emerge from the data" (Strauss and Corbin 1998, 12). Coding is a way of organizing and evaluating data to understand the contextual meanings of qualitative data, and the coding process allows researchers to find meanings of and relations between data by differentiating and combining them (Cope 2003, 2005). Coding is a rigorous qualitative method. Different researchers have suggested different coding procedures (Strauss 1987), but they generally follow two broad steps. The initial step is descriptive coding, and the next is analytic coding. Descriptive coding can be thought of as applying categorical labels to the data, while the analytic coding stage is more interpretive. Therefore, analytic codes are interpretive and reflective about the description of data (Cope 2003). Analytic codes show the process and the result of qualitative research better, and they can be a stepping-stone for further qualitative analyses (e.g., identifying emerging themes from analytic codes). Coding does not just entail continuously adding tags to data, but involves filtering data and developing general ideas or key themes. As a result, researchers can have a manageable dataset and deal with a huge amount of information without losing its contextual importance. Code clouds are designed to visualize analytic codes as a form of visualizing qualitative data and analyses, so it is important to remember that qualitative analysis should have preceded the creation of code clouds.

Technically, code clouds are generated in the same manner as content clouds, but methodologically they differ in important ways. Content clouds do not accurately represent qualitative data because they use input from unanalyzed and uncoded raw text. In contrast, code clouds enable readers to see what the data actually mean, and lead to further analysis and interpretation. Code clouds therefore make use of qualitative data transformation. We know that data transformation (e.g., data reduction, simplification, generalization, etc.) is a necessary part of nearly all quantitative analysis in order to understand and analyze the raw data. However, we often overlook the fact that data transformation is also an essential part of qualitative analysis. The purpose of qualitative analysis is not just showing the words themselves, but also the meanings they contain. Qualitative data transformation is the ability that coding lends

[5] Content analysis is typically understood by qualitative researchers as a quantitative technique because it quantitatively measures key words. However, it is often considered to be a qualitative technique by quantitative researchers because it works with qualitative data such as texts. Content analysis is the quantitative analysis of qualitative data.

us to analyze and interpret raw qualitative data, and that is what makes code cloud visualization a significant methodological advancement over content clouds.

Geotweet case study

This section presents two case studies that demonstrate the potential applications of code clouds to the production of meanings from geotweets. The study area is King County, the most populous county in Washington State and the home of the state's largest city, Seattle. This study uses a 5 percent sample of all geotweets generated in King County in October 2012. The original dataset was collected and provided by the DOLLY (Data on Local Life and you) project[6] at the University of Kentucky. A total of 14,858 geotweets were collected, and the dataset contains detailed information including user ID, user description, geographic coordinates, created date and time, place type, and tweet text. The geographic coordinates of each tweet are used to map them. I analyzed the main texts of the tweets through a coding process, using the computer-aided qualitative data analysis software (CAQDAS) called ATLAS.ti[7] to manage and facilitate the qualitative analysis process. Technically, Wordle (www.wordle.net) was used to generate code clouds of analytic codes. Other websites such as TagCrowd (http://tagcrowd.com), TwitScoop (www.twitscoop.com), and Twitter Trending Topics (twitter.com/trending-topics) all have the ability to generate a visual display of words. However, Wordle was chosen for its more flexible capabilities for visualizing text analysis (e.g., word counting) and layout (e.g., placement and shape) (Steele and Iliinsky 2010). Although Wordle is not made for advanced textual analysis and visualization, it generates code clouds well enough to reflect the contextualized meanings of source tweets.

I began by mapping basic spatial and temporal distribution of tweets, which is typical of quantitative analyses of geotweets. All tweets were mapped with embedded geographic coordinates (i.e., latitude and longitude) and grouped by tweet time. Tweets were more concentrated in the western part of King County, and more than half of the tweets were generated between late afternoon and midnight (Figure 3).

The next step was the coding process. I began an inductive and iterative coding analysis without any pre-defined categories or themes. Instead, I explored multiple possibilities and interpretations, so that new analytic codes were naturally developed from the original texts. I read and coded each tweet one by one. Coding procedures provided me with analytic tools for systematically and creatively handling masses of (raw) tweet data, and I started to draw out geographically focused areas and recurring themes. In particular I identified two themes: the University District (U-District) in Seattle, and the debate around two candidates in the 2012 U.S. Presidential election, Obama and Romney. The following discussions are based on these two themes, and they demonstrate the potential use of code clouds as a qualitative geovisualization approach in studying geotweets.

Theme 1. U-District

Many tweets were located within the U-District. Identifying the specific areas in which tweets are concentrated displays the potential for geographically analyzing spatially linked social media. It would not be possible to identify these locations if the tweets were not geotagged. Geographically focused tweets also prompt us to study specific topics and themes occurring within particular area, and their relationships to one another.

The busiest time for Twitter activities in the U-District was from 11:00 a.m. to 2:00 p.m. However, proportionally high numbers of tweets were generated in the late night as well. For instance, one tweet was generated at 3:00 a.m. at the University of Washington library, and it says, "Still at the library. I need sleep Ugh." A total of 218 analytic codes were created. Most common codes were generic, but closely related to the academic and social life at the University: 'UW,' 'building,' 'place,' 'party,' 'food,' 'campus,' 'frat,' 'drink,' and 'excitement.' There were also many tweets complaining about courses, classes, majors, instructors, and even TAs: 'complaint,' 'professor,' 'struggle,' 'disappointment,' 'exam,' 'hate,' 'homework, 'missing,' and 'changing major.' Seattle's gloomy fall weather was often mentioned as well: 'weather,' 'weird,' 'coffee,' and 'rain.' There were

[6] http://www.floatingsheep.org

[7] For further details of ATLAS.ti, see Friese (2012) and Lewins and Silver (2007). For examples of the parallel use of ATLAS.ti and GIS, see Jung and Elwood (2010).

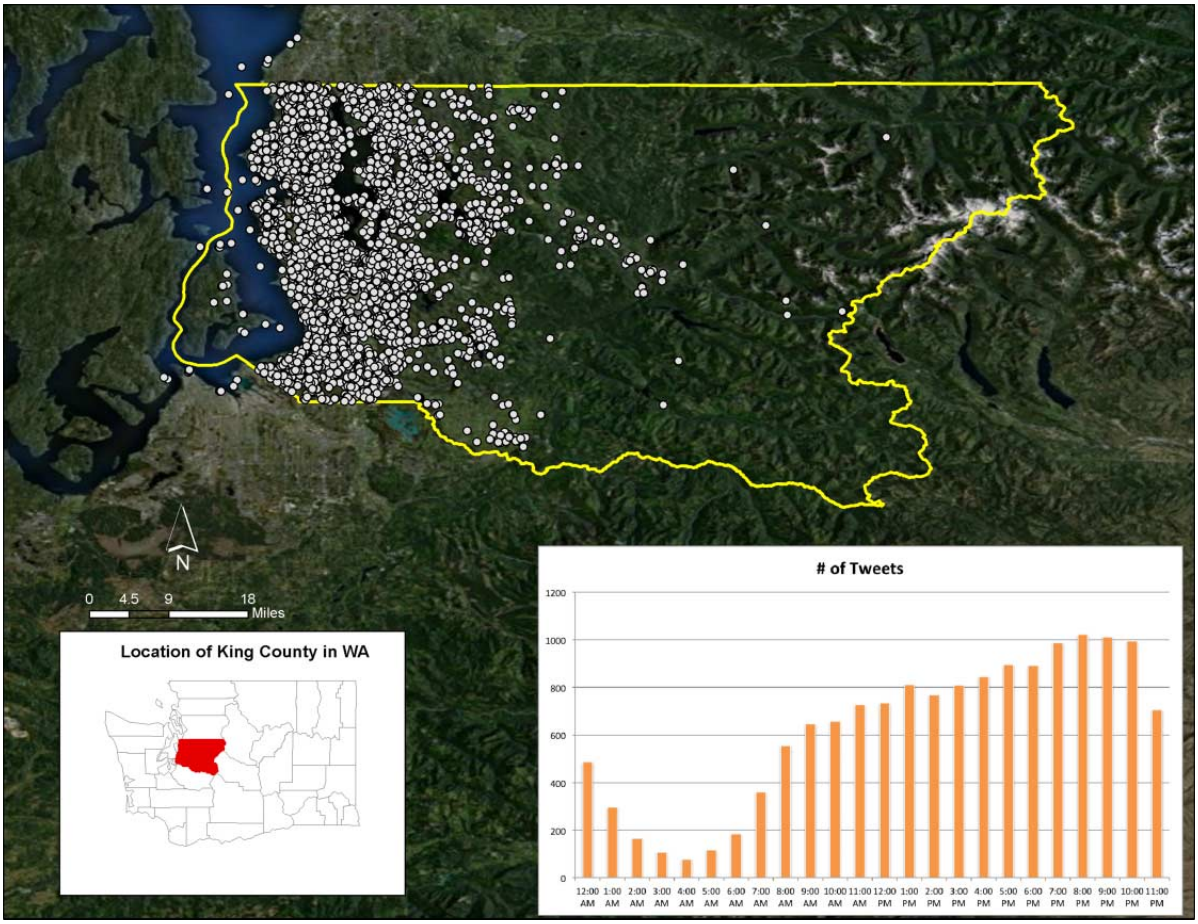

Figure 3
Spatial and temporal distribution of all tweets

also a great number of tweets that include symbols and acronyms specific to Twitter and the Net, such as 'lol,' '☺,' 'OMG,' and 'RT.'

Figure 4 compares the difference between content cloud and code cloud visualizations from the tweets in U-District. The largest words (most counted words) in content clouds, such as 'UW,' 'Seattle,' or 'Hall,' are not quite visible in code clouds. Code clouds present analytic codes that are more meaningfully linked to the context of tweets. The most common codes in the U-District show common activities (e.g., class), places (e.g., building, place), feelings (e.g., excitement), and concerns (e.g., changing major) associated with the university life.

There were also examples depicting why it is important to consider and interpret the meanings of a word. In certain instances the exact same word was used to express different meanings. I found three examples of the use of the word "sick" to be interesting:

> "I need to call in sick today so I can watch all my shows"
>
> (47.661980, −122.304321, 10/25/2012)
>
> "I'm sure his concert was sick as hell"
>
> (47.6618003, −122.300808, 10/24/2012)
>
> "Sick socks bro"
>
> (47.656224, −122.305537, 10/8/12)

The word "sick" can mean mental or physical illness (the first tweet) or it can probably mean

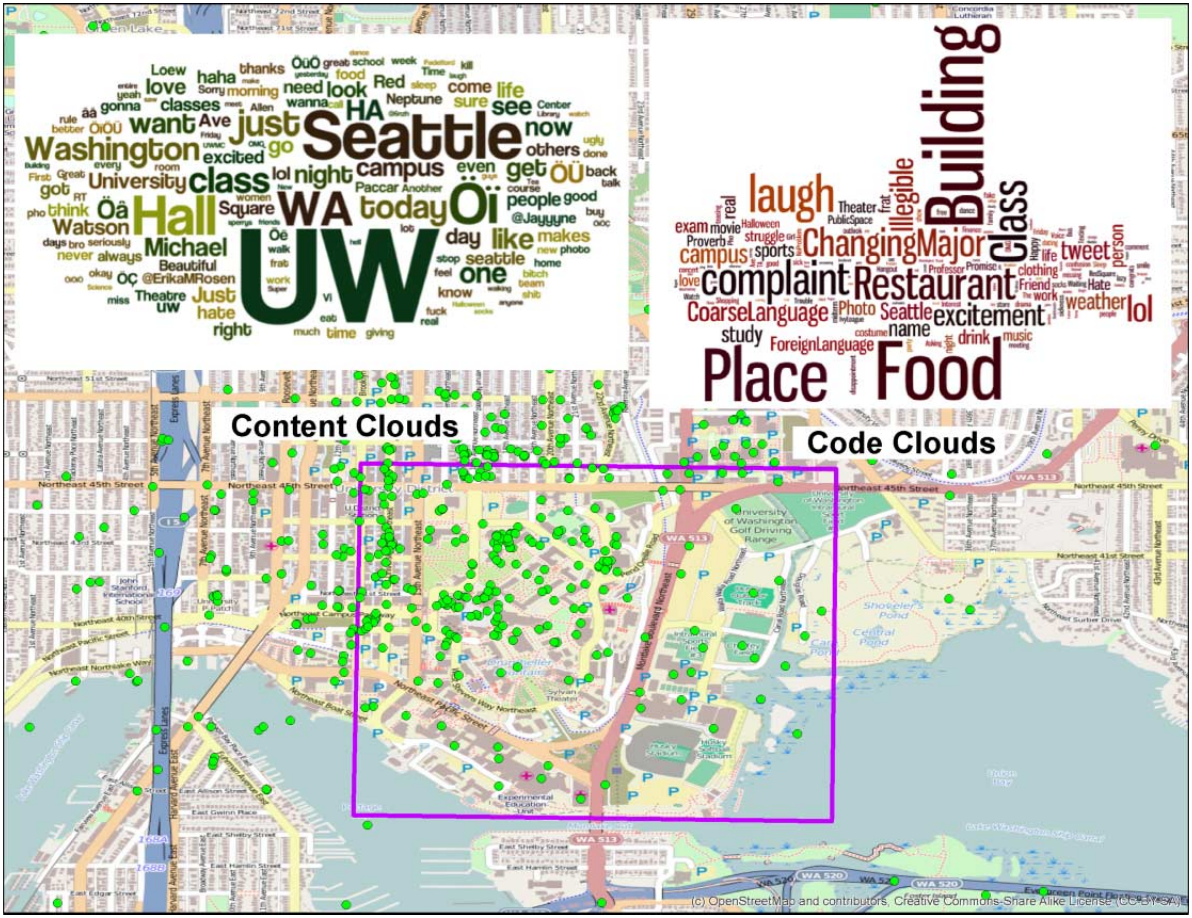

Figure 4
Content clouds and code clouds of geotweets in the U-District

extraordinarily good or attractive (the latter two). Content analysis might miss these nuanced differ-ences of the same word, making it difficult to reveal a tweet's true meanings in context. But, if we carefully read and qualitatively analyze tweets using a coding process, we can draw out the situated and interpretive meanings so that they are well-represented in code clouds. The production of contextual meanings through coding processes and their representation in code clouds are demonstrated again in the next example.

Theme 2. Tweeting for Obama or Romney?

My sample included many tweets showing a strong interest in the upcoming Presidential election and the ongoing debates between the Democratic and Republican candidates. This was a good example of the aforementioned "Twitterverse" phenomena. In my analysis I focused on reading the popularity of these two candidates in Twitter comments. I started with content analysis, and then moved to the coding process.

I first searched and retrieved all tweets that included either the word "Obama" or "Romney" through a key word search. Exactly duplicated tweets were removed by crosschecking them with "user ID" and "tweet time." I then carefully examined all retrieved tweets to determine if there were any tweets referring to people named "Obama" or "Romney" other than the two candidates. Fortuna-tely, I found no instances of this. I collected a total of

254 tweets that contained either of the two candidates' names: 154 tweets for Obama and 100 for Romney. I only looked at main tweet texts, and did not consider other information such as user IDs or user descriptions, even though this information might provide us with a new perspective on geotweets. After retrieving Obama and Romney tweets, they were plotted on a map (Figure 5). Overall, geotweets for both candidates were geographically spread out, and both candidates seemed popular across King County.

However, the coding process and visualization with code clouds illustrated the substantively different meanings being developed through twitter conversations across particular places in King County. A total of 131 and 75 analytic codes were created for Obama and Romney, respectively. Main campaign pledges were often discussed: 'health care,' 'tax,' 'job,' 'spending,' 'middle class struggle,' and 'school policy.' However, there were also tweets commenting on the candidates' performance, personalities, and even outlook: 'great opening,' 'losing temper,' 'repeating,' 'gesture,' 'calm,' 'mean,' 'salesmanship,' and 'hairstyle.' Politically incorrect remarks were found as well: 'al Qaeda,' 'Hellllllaaaaaaaa Ruddddeeeee,' and 'cocoa skin.' I purposefully created three 'selected' analytic codes/categories for each candidate indicating their favourability: 'favor,' 'not favor,' and 'neutral,' and then applied them to the tweets. Like any qualitative or quantitative categorization scheme, selected categories reduced some of the complexity of the data to make overall patterns apparent. It is especially important to consider the size of social media data, given the nuances and complexity of language. This can also accomplish things that simpler word counting techniques cannot because it teases out the nature of the comments about a particular candidate. The intention was not just to use these selected codes, but to add them to, and mix them with, other analytic codes generated during the coding process. The result from all tweets mentioning 'Obama' are as follows: 20 percent for 'Favor Obama,' 25 percent for 'Not Favor Obama', 23 percent for 'Neutral,' 5 percent for 'Not Favor Romney,' 14 percent for 'Favor Romney,' 4 percent for 'Michelle Obama' (not Obama), and the remaining 9 percent for ridicule/swear words that could not be taken into account. These results were quite different in the tweets related to Romney. Only 14 percent were 'Favor Romney,' but 21 percent were for 'Favor Obama,' and surprisingly, 43 percent were 'Not Favoring Romney.' 15 percent were 'Neutral' and about 7 percent of the tweets contained swear words that could not be included in any selected codes. Compared to the Obama tweets, more than half of these tweets were either 'Neutral' or 'Not Favoring Romney.' The outcomes from content analysis (e.g., searching tweets containing a word, 'Romney'), and the mapping of the geographic distribution of Romney tweets note equally strong support for Romney in King County. However, this was not the case once we qualitatively analyzed the meaning of each tweet. Most Romney tweets turned out to be non-favourable. Figure 6 shows the vivid difference between the content clouds and code clouds of Romney tweets.

The difference between content and context is clearly demonstrated in the following two tweets. Both contain the word, "Romney"; however, the natures and nuances of their comments are obviously different:

> "I personally think those were great opening lines by Romney" (47.60601285, −122.3178578, 10/3/2012)
>
> "My Romney, I have perfect cocoa skin. Your skin is as pasty and pale as your magic underwear."
>
> (47.76136017, −122.1601896, 10/3/2012)

The first tweet favours Romney, but the second one is quite sarcastic. The first tweet was coded for 'Favor Romney,' and the second was coded for 'Not Favor Romney.' If we had not considered the contextual meaning of the words, and had only looked at the content, both would have been coded for 'Favor Romney.' Different stories can be hidden in content, and we need to unearth these stories by applying qualitative interpretive approaches such as coding. In particular, the latter tweet reminds us of the importance of issues of subjectivity, including the researcher's positionality and the positionality of the respondents. The anonymous nature of digital online spaces makes it easy for tweets to be written quite sarcastically or falsely, and difficult for content analysis to capture these meanings. Considering the researchers' and respondents' positionalities and perspectives is an essential part of finding the situated meanings of tweet texts.

We can further study how tweets in specific locations are related to particular themes and topics

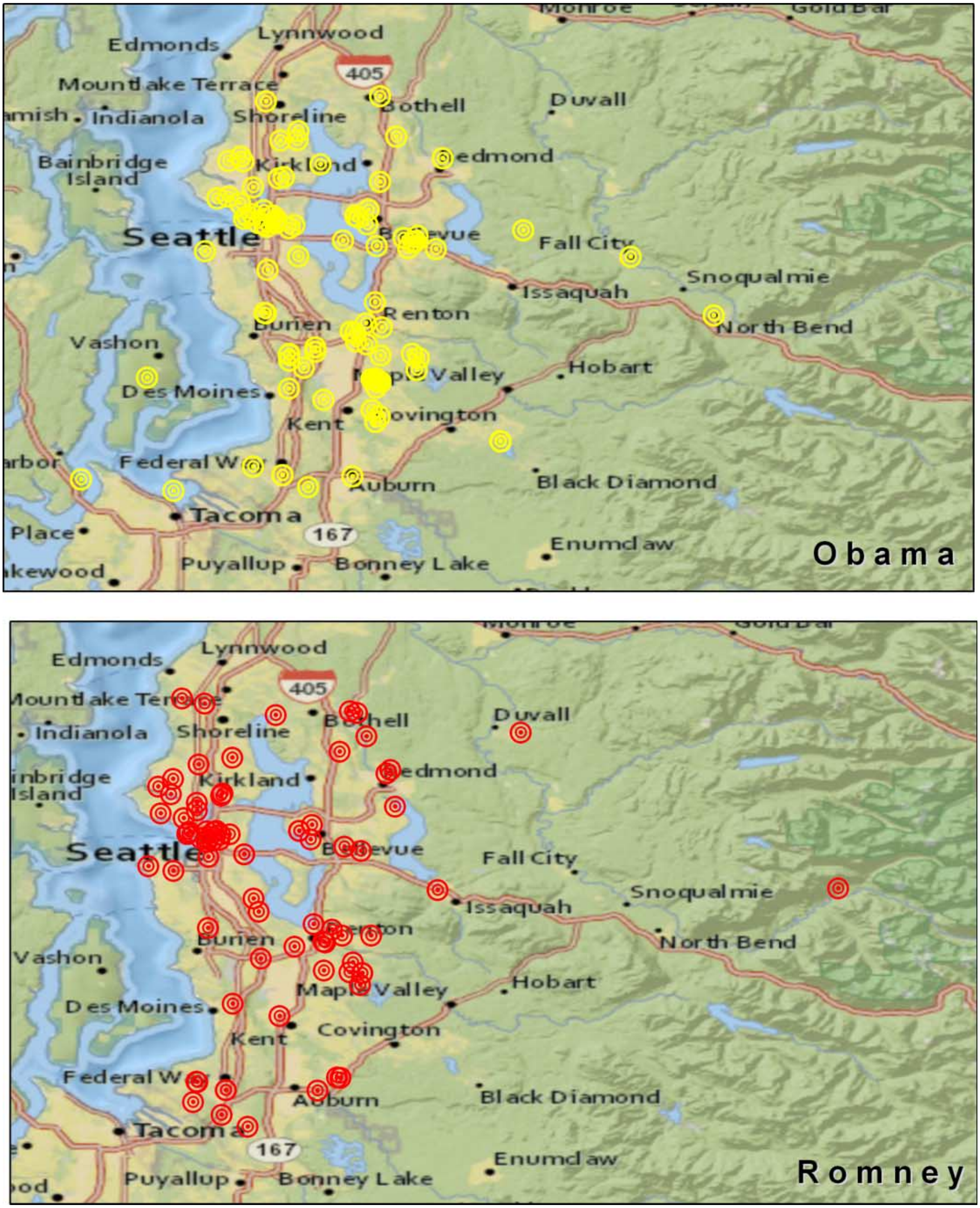

Figure 5
Geographic distribution of Obama and Romney tweets

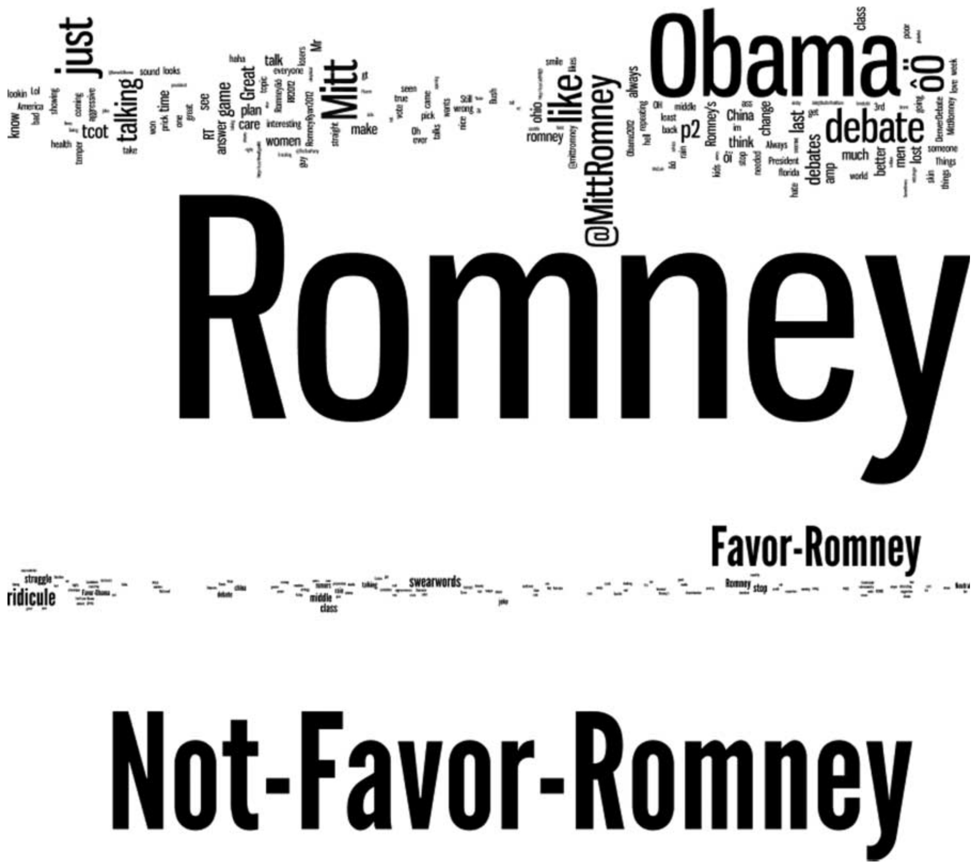

Figure 6
Content clouds and code clouds of Romney geotweets

when we display the geographic distribution of tweets with codes. For example, Figure 7 shows a map of the distribution of Romney tweets with coded themes. Not only can we show the tweet location, but we also can take into account the context of the tweets. In Figure 7, codes themselves have been used effectively to represent locations, and they function as texts themselves. However, codes are also used to locate mapped features. This is one example demonstrating the potential use of code clouds as a form of qualitative geovisualization.

Conclusion

This article discusses theoretical and practical methods of implementing qualitative analyses and geographic visualization together to engage with Big Data and geotagged social media. Code clouds provide robust way to explore, analyze, and represent the process and results of qualitative research. They display summaries of analytic codes by visualizing codes with different sizes and colours proportionate to their frequency. Code clouds themselves are forms of qualitative geovisualization because codes and their visualization are qualitative forms of data, and they can be displayed with spatial information. Also, code clouds can be part of a broader qualitative geovisualization that includes other quantitative and qualitative data such as numbers, texts, photos, videos, GIS maps, and hyperlinked multimedia. With this approach we can reflect the original intent of data throughout the research process, and represent the contextualized

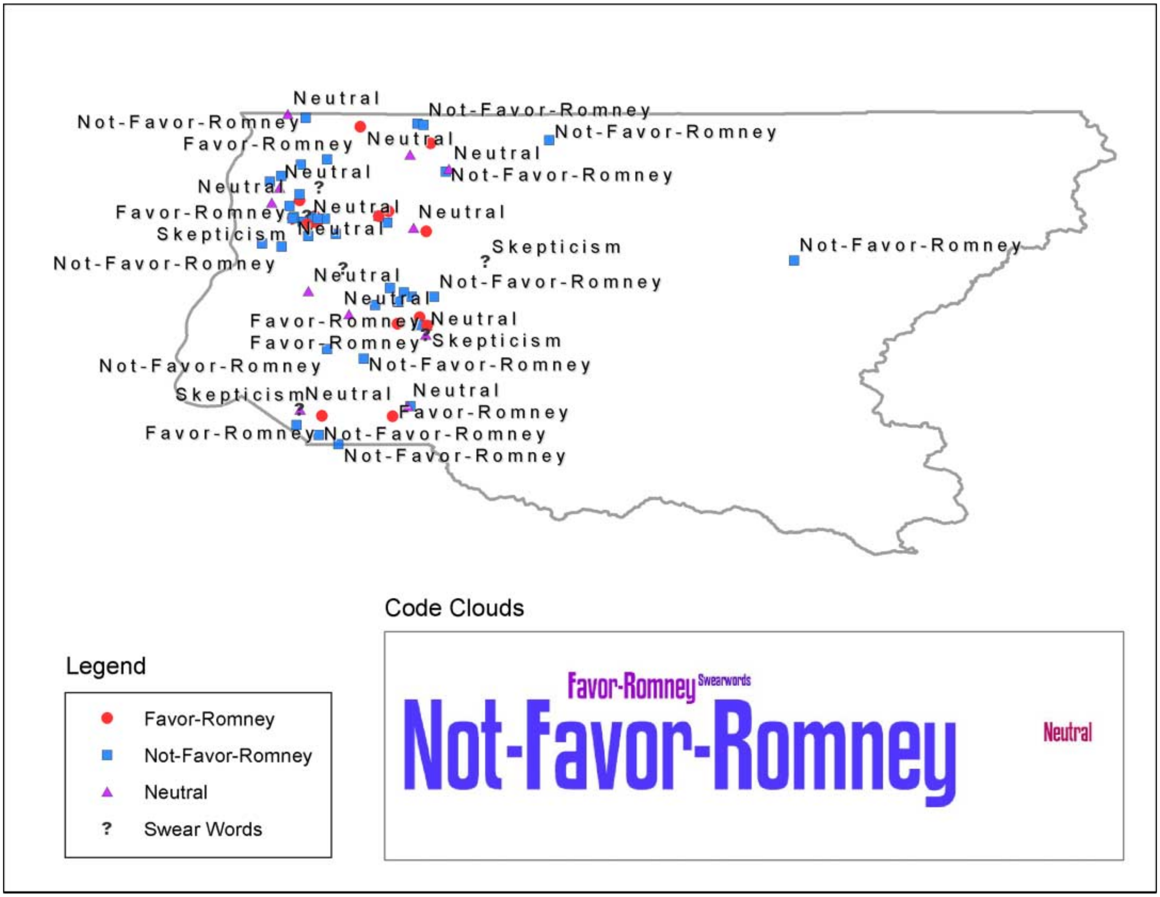

Figure 7
Distribution of Romney tweets with analytic codes and code clouds

meanings better. Code clouds also promote and extend a fuller engagement of earlier discussions in qualitative GIS and geographic visualization.

Spatially linked social media prove that we live in a world that Gray (2007) once called “data intensive,” or full of “data icebergs.” Gray argued that we are in the Fourth Paradigm in science, where the focus is on representing information in an algorithmic and systematic way. While many researchers have been devoted to algorithmic approaches to study Big Data, I argue that there is a lack of comprehensive and synthetic approaches that invite an integration of qualitative analysis and geographic visualization. Many innovative researchers have sought to combine different methodological approaches in creative ways in order to develop new hybrid approaches (Barewald 2010; Sui and DeLyser 2012; DeLyser and Sui 2013). Qualitative GIS researchers have demonstrated the power of mixed-methods research in human geography and GIScience (Kwan and Ding 2008; Cope and Elwood 2009; Knigge and Cope 2009; Elwood 2010b). Code clouds offer one way of carrying out mixed- or hybrid-methods by integrating qualitative analyses with geographic visualization for the representation of spatially integrated social media.

There are also several areas in which we may enhance the use of code clouds.

First, more collective analyses utilizing the locational and temporal dynamics of social media with qualitative analysis are needed. Such parallel approaches are useful because researchers generate key themes through qualitative analysis, and they concurrently apply more advanced geospatial analyses. This article pointed out the lack of holistic approaches that invite qualitative analyses and interpretive methods in studying spatially linked social media. However, this is not meant to disregard the importance of traditional, quantitative geographic analyses. Many creative efforts have been made to expand geographic analyses to study Big Data beyond simple mapping, and we should consider combining these spatial analytical methods with qualitative approaches.

Secondly, further engagements with various types of qualitative analysis are desirable. To fully understand the significance of qualitative geovisualization, we need to look closely at what kinds of data are mapped and visualized, and what associated qualitative data are collected, analyzed, and represented in the form of geovisualization. This article mainly focuses on qualitative analyses of textual data, especially the coding process. However, social media often contain various other forms besides texts, such as multimedia and weblinks. Our focus should be expanded from privileging textual analysis to include other forms of qualitative analyses such as visual methodologies (Rose 2001, 2003), meta-analysis (Gaber and Gaber 2007), discourse analyses (Dittmer 2010), and rhythmanalysis (Lefebvre 2004). These all represent various forms of qualitative data transformation, and provide new ways of creating and working with qualitative geovisualization. These methods therefore allow us to grapple with the questions of qualitative data and various ways of analyzing them.

Thirdly, social media allows us to examine smaller and subtler changes in shorter periods of time (Altheide and Schneider 2013). We can track continuous communications, enabling us to experiment and to develop new applications for tracking discourses. Therefore, we at least need to continue paying attention to the history and chronology of social media. Tracing and analyzing Retweets (RT) in Twitter might be a good starting point to research continuous conversation.

Fourthly, we should be aware of redundancy in the spatial data associated with social media. There are many automatically generated tweets with location-based service providers. For example, unwanted automatically generated Check-In data in Foursquare are posted on Twitter whenever users visit places that are defined as favourite places. There is a danger we may end up having boundless and useless redundant spatial and social data.

Lastly, in relation to Manovich's (2011b) direct visualization, we can extend the power of code clouds by generating clouds not only from the input of texts, but also from the input of other forms of data such as visuals. For example, photographic images themselves can be code and can be proportionally re-sized and visualized according to their importance.

Implementation of qualitative analysis and careful consideration of qualitative aspects of geovisualization complement current Big Data and social media research. Code clouds allow us to see and visualize the meanings of geotagged social media data, and help us to research the fruitful contexts of diverse socio-spatial, cultural, political and technical boundaries of knowledge.

Acknowledgements

I recognize the support of an organizer, Nadine Schuurman, and all the participants in the 2013 Alt.GIS conference on the development of earlier ideas of this article. I particularly thank Sarah Elwood and Jason Young for their constructive and thoughtful feedback, and extend my gratitude to the editor, the anonymous reviewers, and the editorial assistant for their continuous help that improved this article greatly. I also gratefully acknowledge the support for geotweet data from the DOLLY project at the University of Kentucky.

References

Altheide, D. L., and C. J. Schneider. 2013. *Qualitative media analysis*. Thousand Oaks, CA: SAGE.

Anderson, T., and H. Kanuka. 2003. *E-Research: Methods, strategies, and issues*. Boston: MA: Allyn & Bacon.

Baerwald, T. J. 2010. Prospects for Geography as an interdisciplinary discipline. *Annals of Association of American Geographers* 100(3): 493–501.

Bhattacharjee, R. 2013. Hate heat map shows where racist, homophobic tweets come from. http://news.msn.com/us/hate-heat-map-shows-where-racist-homophobic-tweets-come-from.

Butts, C. T., R. M. Action, J. R. Hipp, and N. N. Nagle. 2012. Geographical variability and network structure. *Social Networks* 34(1): 82–100.

Chen, C. 2004. *Information visualization: Beyond the horizon*. Singapore: Springer.

Cidell, J. 2010. Content clouds as exploratory qualitative data analysis. *Area* 42(4): 514–523.

Cieri, M. 2003. Between being and looking: Queer tourism promotion and lesbian social space in Greater Philadelphia. *ACME: An International E-Journal for Critical Geographies* 2(2): 147–166.

Cope, M. 2003. Coding transcripts and diaries. In *Key Methods in Geography*, ed. N. J. Clifford and G. Valentine. London: SAGE, 445–459.

———. 2005. Coding qualitative data. In *Qualitative research methods in human geography*, ed. I. Hay. New York: Oxford University Press, 310–324.

Cope, M., and S. Elwood. 2009. *Qualitative GIS: A mixed methods approach*. London: SAGE.

Crampton, J. 2009. Cartography: maps 2.0. *Progress in Human Geography* 33(1): 91–100.

Daraganova, G., P. Pattison, J. Koskinen, B. Mitchell, A. Bill, M. Watts, and S. Baum. 2012. Networks and geography: Modelling community network structures as the outcomes of both spatial and network processes. *Social Networks* 34(1): 6–17.

Davidson, J., and S. di Gregorio. 2011. Qualitative research and technology: In the midst of a revolution. In *The SAGE handbook of qualitative research* ed. N. K. Denzin and Y. S. Lincoln. Thousand Oaks, CA: SAGE, 627–644.

DeLyser, D., and D. Z. Sui. 2013. Crossing the qualitative–quantitative chasm II: Inventive approaches to big data, mobile methods, and rhythmanalysis. *Progress in Human Geography* 37(2): 293–305.

Dittmer, J. 2010. Textual and Discourse Analysis. In *The SAGE Handbook of Qualitative Geography*, ed. D. DeLyser, S. Herbert, S. Aitken, M. Crang, and L. McDowell. Thousand Oaks, CA: SAGE, 274–286.

Elwood, S. 2009. Geographic Information Science: New geo-visualization technologies—Emerging questions and linkages with GIScience research. *Progress in Human Geography* 33(2): 256–263.

———. 2010a. Geographic information science: Emerging research on the societal implications of the geospatial web. *Progress in Human Geography* 34(3): 349–357.

———. 2010b. Geographic information science: Visualization, visual methods, and the geoweb. *Progress in Human Geography* 34(3): 1–8.

Erickson, I. 2010. Geography and community: New forms of interaction among people and places. *American Behavioral Scientist* 53(8): 1194–1207.

Fitton, L., M. E. Gruen, and L. Poston. 2009. *Twitter for dummies.* Indianapolis, IN: Wiley Publishing, Inc.

Friese, S. 2012. *Qualitative data analysis with ATLAS.ti*. London: SAGE.

Gaber, J., and S. Gaber. 2007. *Qualitative analysis for planning & policy: Beyond the numbers.* Chicago, IL: The American Planning Association.

Gahegan, M., and W. Pike. 2006. A situated knowledge representation of geographical information. *Transactions in GIS* 10(5): 727–750.

Geertz, C. 1973. *The interpretation of cultures: Selected essays.* New York: Basic Books.

Glaser, B. G., and A. Strauss. 1967. *The discovery of grounded theory: Strategies for qualitative research.* New York: Aldine Transaction.

Graham, M., M. Stephens, and S. Hale. 2013. Mapping the geoweb: a geography of Twitter. *Environment and Planning A* 45(1): 100–102.

Gray, J. 2007. Jim Gray on eScience: a transformed scientific method. http://research.microsoft.com/en-us/collaboration/fourthparadigm/contents.aspx.

Gundersen, E. 2013. Visualizing 3 Billion Tweets. Map Box. June 19. http://www.mapbox.com/blog/visualizing-3-billion-tweets/.

Hey, T., S. Tansley, and K. Tolle. 2007. *The fourth paradigm: Data-intensive scientific discovery*. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf.

Hipp, J. R., R. W. Faris, and A. Boessen. 2012. Measuring 'neighborhood': Constructing network neighborhoods. *Social Networks* 34(1): 128–140.

Jung, J.-K., and S. Elwood. 2010. Extending the qualitative capabilities of GIS: Computer-aided qualitative GIS. *Transactions in GIS* 14(1): 63–84.

Knigge, L., and M. Cope. 2009. Grounded visualization and scale: A recursive analysis of community spaces. In *Qualitative GIS: A mixed methods approach* ed. M. Cope and S. Elwood. London: SAGE, 95–114.

Kumar, S., G. Barbier, M. Ali, and H. Liu. 2011. *TweetTracker: An analysis took for humanitarian and disaster relief*. Paper presented at The Fifth International AAAI Conference on Weblogs and Social Media. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2736.

Kwak, H., C. Lee, H. Park, and S. Moon. 2010. *What is Twitter, a social network or a news media*? Paper presented at 19th International conference on World Wide Web. http://an.kaist.ac.kr/traces/WWW2010.html.

Kwan, M.-P. 2007. Affecting geospatial technologies: Toward a feminist politics of emotion. *Professional Geographer* 59(1): 22–34.

Kwan, M.-P., and G. Ding. 2008. Geo-Narrative: Extending geographic information systems for narrative analysis in qualitative and mixed-method research. *The Professional Geographer* 60(4): 443–465.

Lapenta, F. 2012. Geomedia-based methods and visual research: Exploring the theoretical tenets of the localization and visualization of mediated social relations with direct visualization techniques. In *Advances in visual methodology*, ed. S. Pink. London: SAGE, 131–150.

Leetaru, K. H., S. Wang, G. Cao, A. Padanabhan, and E. Shook. 2013. Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday* 18(5): http://dx.doi.org/10.5210/fm.v18i5.4366.

Lefebvre, H. 2004. *Rhythmanalysis: Space, time and everyday life*, translated by S. Elden and G. Moore. London: Continuum.

Leuthold, H., M. Hermann, and S. I. Fabrikant. 2007. Making the political landscape visible: Mapping and analyzing voting patterns in an ideological space. *Environment and Planning B: Planning and Design* 34(5): 785–807.

Lewins, A., and C. Silver. 2007. *Using software in qualitative research: A step-by-step guide*. London: SAGE.

Livehoods. 2012. http://livehoods.org/.

MacEachren, A. M., A. C. Robinson, A. Jaiswal, S. Pezanowski, A. Savelyev, J. Blanford, and P. Mitra. 2011. *Geo-Twitter analytics: Applications in crisis management*. Paper presented at 25th International Cartographic Conference. http://www.geovista.psu.edu/publications/2011/MacEachren_ICC_2011.pdf.

Manovich, L. 2002. Data visualization as a new abstraction and anti-sublime. http://manovich.net/articles/.

———. 2005. Understanding Meta-Media, *CTheory*. http://ctheory.net/articles.aspx?id=493.

———. 2011a. Trending: The promise and the challenges of big social data. http://lab.softwarestudies.com/2011/04/new-article-by-lev-manovich-trending.html.

———. 2011b. What is visualisation? *Visual Studies* 26(1): 36–49.

Miles, M. B., and A. M. Huberman. 1994. *Qualitative data analysis: An expanded sourcebook.* Thousand Oaks, CA: SAGE.

Miller, G. 2011. Social scientists wade into the tweet stream. *Science* 333(6051): 1814–1815.

Neuman, W. L. 1997. *Social research methods, qualitative and quantitative approaches.* Boston: Allyn and Bacon.

Quercia, D., L. Capra, and J. Crowcroft. 2012. *The Social world of twitter: Topics, geography, and emotions.* Paper presented at The Sixth International AAAI Conference on Weblogs and Social Media. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4612.

Parker, K. 2012. *The social-media presidential campaign: Election by Twitterverse.* October 21. The Seattle Times. http://seattletimes.com/html/opinion/2019476817_parkercolumn-socialdebatexml.html.

Pavlovskaya, M. E. 2002. Mapping urban change and changing GIS: Other views of economic restructuring. Gender, *Place and Culture* 9(3): 281–289.

Raymond, M. 2010. *Twitter donates entire tweet archive to library of congress. Library of Congress.* http://www.loc.gov/today/pr/2010/10-081.html.

Reisinger, D. 2011. *Twitter tallies 200 million tweets per day.* CNET. July 1. http://news.cnet.com/8301-1350 6_3-20076022-17/twitter-tallies-200-million-tweets-per-day/.

Rose, G. 2001. *Visual methodologies: An introduction to the interpretation of visual materials.* 2nd ed. London: SAGE.

———. 2003. On the need to ask how, exactly, is geography visual? *Antipode*, 35: 212–221.

Schuurman, N. 2009. The new Brave New World: Geography, GIS, and the emergence of ubiquitous mapping and data. *Environment and Planning D: Society and Space* 27(4): 571–580.

Sharif, S. 2012. *Obama vs. Romney on Twitter: Data proves social media is critical for successful campaigns. Policymic.* http://www.policymic.com/articles/8991/obama-vs-romney-on-twitter-data-proves-social-media-is-critical-for-successful-campaigns.

Skupin, A., and K. Borner. 2007. Guest editorial. *Environment and Planning B: Planning and Design* 34(5): 765–766.

Skupin, A., and S. I. Fabrikant. 2003. Spatialization methods: A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science* 30(2): 99–119.

Small, T. A. 2011. What the hashtag? A content analysis of Canadian politics on Twitter. *Information, Communication & Society* 14(6): 872–895.

Steele, J., and N. Iliinsky. 2010. *Beautiful visualization.* Sebastopol, CA: O'Reilly.

Strauss, A. 1987. *Qualitative analysis for social scientists.* Cambridge: Cambridge University Press.

Strauss, A., and J. Corbin. 1997. *Grounded theory in practice.* Thousand Oaks, CA: SAGE.

———. 1998. *Basics of qualitative research: Techniques and procedures for developing grounded theory.* London: SAGE.

Sui, D. Z., and D. DeLyser. 2012. Crossing the qualitative-quantitative chasm I: Hybrid geographies, the spatial turn, and volunteered geographic information (VGI). *Progress in Human Geography* 36(1): 111–124.

Sui, D. Z., and M. F. Goodchild. 2011. The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science* 25(11): 1737–1748.

Takhteyev, Y., A. Gruzd, and B. Wellman. 2012. Geography of Twitter networks. *Social Networks* 34(1): 73–81.

Terdiman, D. 2012. Twitter hits half a billion tweets a day. CNET. October 26. http://news.cnet.com/8301-1023 _3-57541566-93/report-twitter-hits-half-a-billion-tweets-a-day/.

Thatcher, J. 2014. Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication* 8: 1765–1783.

Tsou, M.-H., and J.-A. Yang. 2012. *Spatial analysis of social media content (tweets) during the 2012 U.S. Republican presidential primaries.* Extended Abstract at Seventh International Conference on Geographic Information Science. http://www.giscience.org/proceedings/proceedings-ext-name.html.

Ward, M., G. Grinstein, and D. Keim. 2010. *Interactive data visualization: Foundations, techniques, and applications.* Natick, MA: A.K. Peters, Ltd.

Wilson, M. W. 2009. Towards a genealogy of qualitative GIS. In *Qualitative GIS: A mixed methods approach*, ed. M. Cope and S. Elwood. London: SAGE, 156–170.