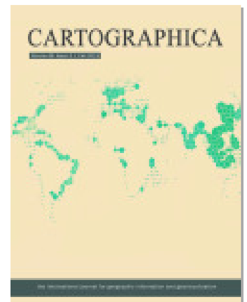
Towards Qualitative Geovisual Analytics: A Case Study Involving Places, People, and Mediated Experience

Ryan Burns, André Skupin

Cartographica: The International Journal for Geographic Information and Geovisualization, Volulme 48, Number 3, Fall 2013, pp. 157-176 (Article)

Published by University of Toronto Press



ARTICLES

Towards Qualitative Geovisual Analytics: A Case Study Involving Places, People, and Mediated Experience

Ryan Burns
Department of Geography / University of Washington / Seattle / WA / USA

André Skupin
Department of Geography / San Diego State University / San Diego / CA / USA

ABSTRACT

This article seeks to address the gap between the quantitative, summative data that are typically engaged in geovisual analytics and the more personal, experiential ways of knowledge construction accentuated by qualitative GIS. By incorporating diverse forms of data within a high-dimensional conceptual framework, we set out a type of qualitative geovisual analytics. This approach is attentive to the epistemological limitations of singular data sources and highlights the multiple ways of exploring neighbourhoods. The article reports on a project that used an online survey, including collection of personal impressions of San Diego neighbourhoods based on street-level video. Three attribute spaces are conceptualized: survey respondents' characteristics, attributes of San Diego neighbourhoods, and characteristics of the words used to describe these neighbourhoods. The self-organizing map (SOM) technique was used to reduce the dimensionality of these attribute spaces for visual exploration. This approach makes several important contributions, including a demonstration of "scaling up" the work that has been done in qualitative GIS. It productively integrates experiential data with a geovisual analytics approach and reaffirms the multiple meanings of visualization.

Keywords: geovisual analytics, visualization, qualitative data, attribute spaces, San Diego, neighbourhoods

RÉSUMÉ

Cet article vise à combler l'écart entre les données sommatives quantitatives qui servent habituellement en analyse géovisuelle et les façons plus personnelles et expérientielles de construire le savoir mis en évidence par les SIG qualitatifs. En intégrant diverses formes de données dans un cadre conceptuel à hautes dimensions, nous établissons un type d'analyse géovisuelle qualitative. Cette approche tient compte des limites épistémologiques des sources de données uniques et met en évidence les multiples façons d'explorer les quartiers. L'article présente un rapport dans le cadre duquel les chercheurs ont utilisé un sondage en ligne, y compris la collecte d'impressions personnelles de quartiers de San Diego basés sur des vidéos tournées au niveau de la rue. On conceptualise trois espaces représentatifs des caractéristiques : les caractéristiques des répondants au sondage, celles des quartiers de San Diego et celles des mots utilisés pour les décrire. La technique cartographique d'auto-organisation a servi à réduire la dimensionnalité de ces espaces représentatifs aux fins de l'exploration visuelle. Cette approche apporte plusieurs contributions importantes, qui démontrent notamment « l'expansion » des travaux effectués dans les SIG qualitatifs. Sa productivité inclut des données expérientielles et une approche analytique géovisuelle et réaffirme les multiples significations de la visualisation.

Mots clés : analyse géovisuelle, visualisation, données qualitatives, espaces représentatifs, San Diego, quartiers

I. Introduction

Geographers have long understood that different conceptualizations and representations of places have implications for knowledge produced about those places. The conceptualization and representational strategies adopted highlight particular aspects of places while obscuring others (Harley 1989; Rocheleau 1995). Geovisual analytics, a research program interested in the development and deployment of interactive suites of computational and visualization technologies, has facilitated the development of novel perspectives on spatio-temporal patterns and relationships, typically based on observable, measurable attributes of geographic features (Thomas and Cook 2005; Andrienko, Andrienko, Demsar, and others 2010). In contrast, as some who predominantly deal with qualitative data have noted, people's experiences and actions in geographic space are highly personal and characterized by differences in perception, cognition, and affect. These differences are informed by personal characteristics and background such as age, level of education, or socio-economic status. Recent theoretical engagements with place have thus argued that while places may sometimes be conceptualized and represented numerically in an aggregate manner, equal attention should be paid to the factors contributing to people's subjective experience of place and the associated complex interplay of individual attributes and momentary impression.

This article seeks to bridge the gap between the data that are typically engaged in geovisual analytics and more personal, experiential forms of knowledge by accentuating the distinct topologies of place formulated when visualizing personal impressions. Aligned with work on qualitative GIS, the work presented here expands the types of conceptual approaches taken within geovisual analytics. We accomplish this by distinguishing among a series of different *attribute spaces*, with particular consideration of people's mediated experience of place. The concept of attribute space adopted here aims to encompass the sets of attributes, qualities, and characteristics that describe a place, person, or mediated experience. The resulting high-dimensional models become the basis for expressing the location of individual entities and for expressing relationships among multiple entities through computational measures of similarity. Dimensionality reduction and spatial layout techniques, like the self-organizing map (SOM) method, allow the transformation of these high-dimensional spaces into low-dimensional geometries suitable for visual depiction (Skupin and Fabrikant 2003; Skupin and Agarwal 2008). While attribute spaces can be comprised of any sort of quality or characteristic of a group of objects, the SOM method has been used mostly within the geovisual analytics framework to explore aggregate attributes, such as census data or crime statistics, to the exclusion of the more personal ways in which people experience places. Our project juxtaposes traditional SOM-based depiction of census data with attribute spaces derived from people's mediated experience of geographic locales as well as people's background attributes.

The broader study reported here involves the generation and visual exploration of five different SOMs, using data gathered by an online map and video interface, and census data for neighbourhoods in the city of San Diego. By gathering both census data and research subjects' impressions of mediated neighbourhoods, different conceptualizations of *place* are contrasted to highlight the complexities inherent in place representations. This study builds on previous geovisual analytics work that conceptualized places and people as existing in high-dimensional attribute spaces. By thinking through the different attribute spaces in which places, people, and their utterances exist, this project questions the incompatibility of attribute spaces gathered from the census and those gathered from people's mediated experience. With the help of dimensionality reduction, the study elaborates three high-dimensional spaces constructed from the attributes of neighbourhoods, subjects' mediated experience of those neighbourhoods, and subjects' personal attributes.

The following section reviews literature related to the emerging field of qualitative geographic information systems (GIS). Researchers engaging qualitative GIS seek to open GIS to qualitative modes of representation and analysis, which have not traditionally had a strong presence in GIS more broadly. This discussion situates the current project within literatures that seek to more fully account for qualitative data in existing technologies. The second section illustrates that the geovisual analytics literature has highlighted important aspects of place through numeric and summative data but to the exclusion of more personal aspects of place experience, such as emotion and personal connections to places. This is not to denigrate the importance of quantitatively inclined work; in fact, the current article blurs the dubious boundary between quantitative and qualitative data. Rather, this article visualizes personal impressions of neighbourhoods to accentuate the different conceptualizations and representations of places that result in distinct topological relations between the places, and to accentuate potentially productive ways in which a high-dimensional framework can be used to explore different attribute spaces. As visualization is never purely quantitative, geovisual analytics is already well on its way toward this goal. The third section details the methodology used to gather, represent, and analyze the data set procured for the current project and is followed by a discussion of how we engage high-dimensional attribute space, largely based on the SOM method. The results illustrate both the potential of geovisual analytics to engage with qualitative epistemologies vis-à-vis qualitative GIS and the kinds of knowledge this approach may help produce. We argue that SOMs can be used to represent and explore the complex, high-dimensional attribute spaces related to personal impressions of places and that these form topologies quite distinct from attribute spaces of census and other common quantitative data sources.

II. Qualitative Knowledges and Geographic Technologies

Efforts to incorporate personal expressions, impressions, and emotions in a predominantly computational environment find immediate resonance in qualitative GIS. Although geographers have conducted mixed-methods research for decades, the representation and analysis of qualitative data within geographic information systems (GIS) has only recently become a major concern of researchers and practitioners. *Qualitative* here is defined not simply as nonnumeric but also richly contextual, personally experienced, and expressed in nuanced personal terms (Elwood and Cope 2009); in other words, *qualitative* refers to epistemology as well as data type. Partly responding to early critiques of GIS as a quantitative, positivist, and corporatist technology, researchers have developed ways of opening GIS to alternative epistemologies and situated knowledges (Sheppard 2005). These are constituted by the everyday experiences and knowledge through which people approach and utilize technologies. In this manner, the genealogy of qualitative GIS can be drawn to the critical GIS debates of the 1990s (Wilson 2009) and feminist GIS (e.g., Kwan 2002a). For all the diversity in particular approaches to this engagement with GIS, the overarching corpus of work seeks to position qualitative knowledge and experience of places alongside the numeric summative attributes of places.

Jung and Elwood (2010) have argued that qualitative GIS research typically falls into one of three categories: transforming qualitative data into a form easily represented cartographically, hyperlinking from GIS-based objects to external qualitative data artifacts, and modifying existing technologies on the software level to more readily handle qualitative data. Extending GIS in these ways is thought to benefit from a combination of qualitative and quantitative forms of reasoning. Whereas GIS is commonly – if mistakenly – understood to be best suited for quantitative representation and analysis, some working with qualitative data have productively contended that GIS can be conceptualized as a qualifying and qualitatively enabled technology (Kwan 2002b; Pain and others 2006). These lines of reasoning seek to show on the one hand how quantification can obscure the contexts and nuance through which people experience their environments but on the other how GIS can embed quantitative data with this richness.

The importance of explicitly integrating qualitative or qualified data with GIS stems from at least three concerns. First, reflecting longstanding methodological debates, qualitative data are seen as preserving much of the rich contextual information that may be lost in a purely quantitative study. In a community planning context in Humboldt Park, Chicago, Elwood (2009) effectively used qualitative data to visualize the tension between residents' perceptions of their neighbourhood and official administrative neighbourhoods. Elsewhere, Elwood (2006) has shown that even where quantitative data are mapped, they often must be understood in terms that lie outside the map. For instance, GIS-based mapping of vacant lots in Humboldt Park could be understood as *needs narratives*, a conclusion understood only in terms of the supplementary interview data collected.

A second impetus for the incorporation of qualitative data in GIS stems from the difficulty of much research-generated data to be represented quantitatively. Since different forms of data may represent distinct epistemological approaches to knowledge, greater diversity in knowledge representation, analysis, and synthesis may lead to more productive scholarship (Lawson 1995; Pavlovskaya 2006). The methodological challenge for GIS research, then, is to think about the dynamics and processes that may lie outside the purview of quantitative representational capacities. For instance, qualitative GIS has incorporated images, sketches, perceptions, and hypothetical scenarios – none of which have traditionally been considered "data" for GIS purposes (Al-Kodmany 2000; Matthews, Detwiler, and Burton 2005; Brown and Knopp 2008).

The third impetus lies in the political potential of integrating qualitative data in GIS. Feminist GIS in particular has emphasized the many different place-based experiences possible across different social groups and the particular representational strategies that may convey some of these experiences (Kwan 2002a; McLafferty 2005). Qualitative GIS, through its combination of quantitative and qualitative representations of these experiences, is most able to capture the ways in which experiences can be represented. As one example, Knigge and Cope describe how grounded visualization, through its juxtaposition of quantitative and qualitative data, can "build on the robust capacities of ethnographic and GIS techniques to . . . produce rigorous results and present them in ways that are not oppressive or overly exploitive" (Knigge and Cope 2006, 2035). Further, qualitative GIS has the potential to dislodge many of the ontological assumptions about social process; by representing relationships, emotion, gender, and class, qualitative GIS can be used to emphasize the more personal, dynamic ways people interact with space and with each other (Kwan 2002b; Pavlovskaya 2006).

If the discussion above seems to inadvertently construct an artificial binary relationship between quantitative and qualitative methods, it is worth noting that many within GIS have questioned the extent to which geographic technologies such as GIS are inherently quantitative or qualitative (Schuurman 2000). Much work in critical GIS has sought to break down this way of thinking, showing the personal narratives and knowledges that can be represented in GIS (Kwan 2002b). Indeed, as Pavlovskaya (2006) has argued, rather than distinguishing between quantitative and qualitative "toolboxes," it is often more realistic and analytically productive to think of a continuum along which most researchers operate – and along which it is possible to operate.

III. Geovisual Analytics

Geovisual analytics is characterized by the use of highly interactive displays to facilitate visual thinking (MacEachren and Kraak 2001; Andrienko and others 2007; Fabrikant and Lobben 2009). Building on visual analytics more broadly, the central problem geovisual analytics seeks to address is how to visualize large, complex geographic data sets in a manner that is cognitively intuitive (MacEachren and others 2004; Thomas and Cook 2005). Researchers within this field are concerned with developing such tools and approaches, as well as using these interfaces to generate (rather than merely confirm) hypotheses. Within cartography, this has long been echoed in the broadening use of maps, from static communication devices toward interactive exploratory platforms (MacEachren 1995). Rather than presenting *knowns*, the visualization process involves the exploration of *unknowns*; potential multiple meanings and interpretations become central to the process (Dykes, MacEachren, and Kraak 2005).

Another strong influence on the research program stems from visualization's intellectual lineage in scientific computing and exploratory data analysis (Buckley, Gahegan, and Clarke 2000; Hand, Mannila, and Smyth 2001; McCormick, DeFant, and Brown 1987; Thomas and Cook 2005). Both fields emphasize computation to generate the visualizations or extract meaningful patterns from data sets. For geovisual analytics this heritage influences the ways in which exploration of visual displays is undertaken. For instance, exploration might work toward extracting meaningful patterns or more "accurately" gleaning a coherent process from complex data sets (Slocum and others 2005, 45). At times human cognition and perception is augmented with computationally extracted patterns and the mobilization of several automated toolsets (Andrienko and others 2008; Andrienko, Andrienko, Demsar, and others 2010). In this manner, although no exclusion to qualitative *data* is implied, a qualitative *epistemology* may be elided.

Geovisual analytics emerged as the geographic counterpart to visual analytics, a field at its outset concerned with homeland security and terrorism risk management (Thomas and Cook 2005; Andrienko, Andrienko, Demsar, and others 2010; Anselin 2012). As suggested above, visual analytics and geovisual analytics have both moved beyond this topical application area to use various techniques in facilitating visual thinking more broadly (Andrienko and others 2007; Andrienko and others 2011; Keim and others 2010; Thomas and Cook 2006). Thus, although a relatively newly delineated field, geovisual analytics borrows many foundational concepts, motivations, methodologies, and technologies from geographic visualization and geographic information science (Anselin 2012). In fact, one might argue that geovisual analytics differs only from a *narrow* reading of geographic visualization, in which the latter purportedly lacks significant concern with sense-making and knowledge construction.

Within geovisual analytics, disparate analytical concepts have been developed to understand and theorize the relationships between visualized geographic objects (Yan and Thill 2008). The particularly useful concept of *attribute spaces* can be defined as the set of dimensions, or characteristics, of a given group of objects or phenomena. For instance, the attribute space of people may entail their socioeconomic status, sex, religion, and other related metrics. The attribute space of places may entail their demographic characteristics, the impressions visitors have of the place, or a place's multiple histories. Early work visualized places' attribute spaces as comprised by census data and places as temporally moving across this attribute space as the demographic makeup of places shifted (Skupin and Hagelman 2003). Following this, places' attribute spaces were extended to incorporate physical geographic data such as climate, geology, and topography (Skupin and Esperbé 2011). The present study mobilizes the concept of attribute spaces again to inflect a geovisual analytics project with concerns from critical human geography and an attention to the politics of knowledge representation, by looking at attribute spaces as comprised by personal impressions of places.

The tension between computational pattern extraction on the one hand, and the potential multiple meanings of visualizations on the other, speaks to the potential productive blending of epistemologies in geovisual analytics (Crampton 2001). This tension is important to recognize and broach because computation – and quantification more broadly – often bears "objective" clout in relation to more qualitative approaches. In spite of this potential, geovisual analytics has primarily focused on quantitative or quantified data and epistemologies. This focus is due in part to its roots in cognitive science, scientific visualization, and technology development. In contrast, data that is qualitative, emotional, and experiential can enrich geovisual analytics by illuminating the many potential meanings of geovisualizations.

The attribute space approach applied to qualitative, emotional, and experiential epistemologies tends to generate very high-dimensional spaces that are much more difficult to visualize than the low-dimensional geometries and topologies of traditional geographic visualization. A specific visualization method used in geovisual analytics, the SOM, is particularly adept at visualizing these high-dimensional spaces. As mentioned earlier, the SOM is a spatialization technique enabling the visualization of large and complex data sets (Skupin and Agarwal 2008). Combining elements of dimensionality reduction – similar to multidimensional scaling (MDS) and principal components analysis (PCA) – and clustering similar to *k*-means, the SOM allows the projection of high-dimensional data into low-dimensional display space.[1] Common geographic metaphors such as "nearness" and "region" factor strongly into the interpretation of SOMs. Compared to other approaches, such as MDS and PCA, the method makes efficient use of display space and more easily incorporates very large, high-dimensional data sets (Skupin and Fabrikant 2003).

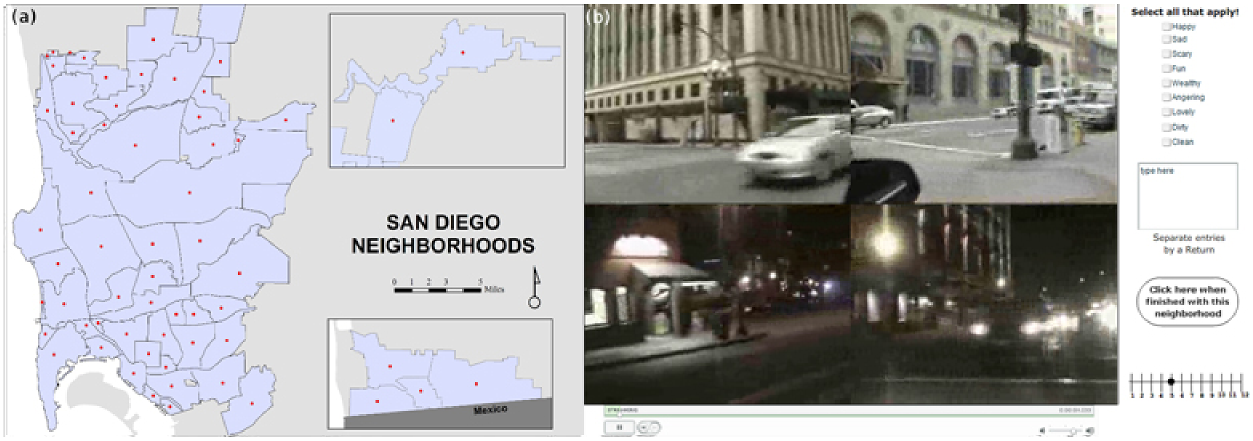


Figure 1. Panel (a) shows the map of San Diego presented to respondents; when a red dot was clicked, the interface in Panel (b) would show videos of that neighbourhood and allow respondents to describe the neighbourhood

SOMs have been used across many diverse geographic applications (Agarwal and Skupin 2008) but in most cases are attempts to extract patterns from, and produce knowledge about, very large, complex data sets. Some have used SOMs to analyze conference abstracts, showing how SOMs can highlight relations between individual objects and the entirety of the data set, while accentuating broad structures and patterns in a "knowledge domain" (Skupin 2004; Skupin and de Jongh 2005). In this case, the qualitative data of conference or article abstracts are often used, pointing to the ability of the SOM to handle this type of data. Research using census and crime data (Skupin 2007; Andrienko, Andrienko, Bremm, and others 2010) has demonstrated that SOMs can draw out strong relations between places, particularly when those relations span across several attributes. Skupin and Esperbé (2011) integrated physical-geographic and geologic attributes of places to increase the diversity of attributes considered by the SOM. Importantly, however, some have noted that the process of visualizing espouses particular representational and conceptual strategies that illuminate some processes but may obfuscate others (Skupin 2009). Thus, while important lessons have been learned through these visualizations, topics such as emotion, subjectivity, and people's descriptions of places have not yet been explored in geovisual analytics. Since this is a central way people experience places, it is important to recognize this as a valid topic of interest to the field.

IV. Methodology

The project reported here recognizes the many ways places can be represented in a technological environment. Whereas in geovisual analytics places and people are often represented by census-type metrics, people experience and come to know places in more personal and qualitative terms. To explore these place experiences, the concept of high-dimensional attribute spaces is employed, with dimensionality reduction operationalized through the SOM method. The different type of data collected ultimately represents places and their relationships differently, pointing to new productive means by which geovisual analytics may engage this type of data. The visualizations presented here are meant to illustrate the types of knowledge production that can occur in this context.

Visualized here are attribute spaces derived from three different data sets:

1. all 60 *neighbourhoods* within the city of San Diego,
2. subjects' *descriptions* of neighbourhoods they viewed in video form during an online survey, and
3. subjects' personal *characteristics*.

Neighbourhoods are represented via two different attribute spaces. One is based on population census data and the other is derived from subjects' utterances in response to videos depicting particular neighbourhoods. This two-pronged method highlights the different topologies between neighbourhoods that can emerge when considering different forms of data. In other words, borrowing from qualitative GIS, the relationships between neighbourhoods that we observe in visualizations may shift depending on the types of data represented. The descriptions and the personal characteristics were captured in a two-part online survey made available to students of San Diego State University over the age of 18.

Those who chose to participate in the study first accessed the survey website and then reported personal information such as sex, age, religion, and socioeconomic status. Other potentially useful information such as duration of time spent at each neighbourhood was not asked, as it would be significantly more difficult to acquire and less directly comparable among all respondents. After answering these questions, they watched 12 one-minute videos of neighbourhoods they chose from a geographic map of San Diego (see Figure 1). Neighbourhoods correspond to the 60 Community Planning Areas (CPA) within the city,

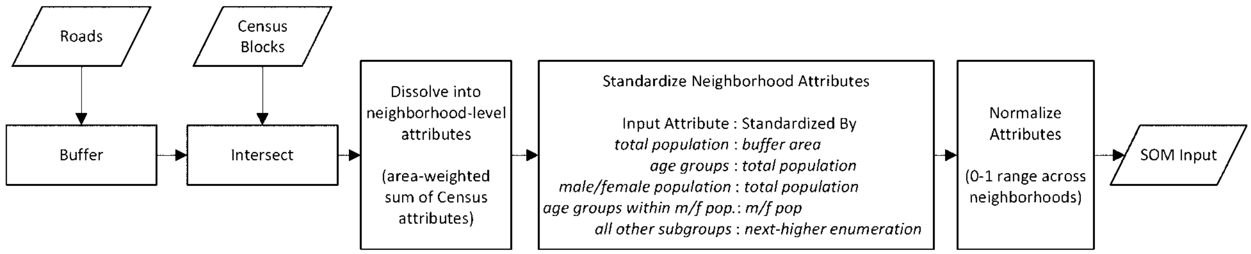


Figure 2. Census data within 100-foot (30.48-m) buffer of video locations, transformed into neighbourhood-level data ready for SOM training

which generally match the commonly understood neighbourhood structure of San Diego. One video for each neighbourhood had been recorded beforehand and showed the neighbourhood at certain times in the day: one near noon and one in the evening. As each video played, the respondents recorded their impressions of the neighbourhood, either by choosing from a list of nine predefined terms or by typing their own descriptors (there were no length restrictions for this option), or both. The nine predefined terms were *happy*, *sad*, *scary*, *fun*, *wealthy*, *angering*, *lovely*, *dirty*, and *clean*. The first and last videos for each respondent were discarded to account for a possible maturation effect (Kirk 2004). With 150 students having successfully completed the survey, there were a total of 1,500 video responses used for analysis.

The survey captured a static representation of a highly dynamic process. Respondents' descriptions would be expected to reflect the conditions seen in the videos, such as weather and precise filming location. Descriptions would also be expected to reflect respondents' familiarity and tacit knowledge of that neighbourhood, as well as their overall mood at the time of the survey and their sociocultural background. Thus, while this study treats the data as a static sample to visualize it, we acknowledge the highly dynamic nature of the data.

The text and personal characteristics data sets went through several processing steps to be usable as SOM input. Porter stemming (Porter 1980) is used to reduce each word to its stem (e.g., *angrily* and *angering* would both be reduced to *anger*). To reduce disproportionate influence of the more commonly chosen nine predefined terms, term frequencies are normalized via TF*IDF (Salton 1989), which increases the importance of infrequent terms, and then scaled to a 0–1 range (Skupin and Agarwal 2008). The result is an 80-dimensional attribute space for input to a SOM. Meanwhile, subjects' personal characteristics are coded numerically and then likewise scaled to a 0–1 range, resulting in a 40-dimensional data set ready for SOM training.

Next, census data are processed to reflect an approximate viewshed for each video (Figure 2), to acquire a comparable representation of places in the study (Burns and Skupin 2009). Each neighbourhood's summary census attributes should correspond to what one would expect to see when travelling along the roads from which videos were recorded. Under the assumption of a 100-foot field-of-view, San Diego census blocks are intersected with a 100-foot buffer constructed around each section of road along which videos had been collected. Each census block contributes to its intersecting buffer attributes that are weighted in proportion to the relative area size within the buffer. For example, attributes of a block occupying half of the buffer would be weighted at 50%. Weighted attributes are then averaged across all captured census blocks for each neighbourhood. Neighbourhood attributes then undergo standardization (i.e., division by suitable attributes) and normalization to a 0–1 range. The resulting data set contains 57 neighbourhoods and 136 attributes, in a form suitable for SOM training. Three neighbourhoods were excluded, since according to the census, the areas captured in the videos contained no population (Los Peñasquitos Canyon Preserve, Miramar Air Station, Scripps Reserve).

The three data sets above aimed to capture personal characteristics of the survey respondents and two different conceptualizations of neighbourhoods. A fourth data set is generated in which each utterance is assigned to the respondent who originally used it. This results in an 80-dimensional attribute space occupied by 150 subjects. A fifth data set is arrived at by transposing the neighbourhood-utterance data set, such that 60 neighbourhoods occupy a space defined by 60 different utterances.

The five data sets were used to train five distinct SOMs, each with a particular topology and organization. Onto each SOM, the input vectors used for neural network training are overlaid, allowing visualization of relationships between objects and the distribution of attributes across the SOM. Point symbols representing input vectors are placed at random locations within the two-dimensional extent of the respective most similar SOM neuron. This allows vectors sharing the *same* best-matching neuron to remain visible, alleviating the problem of coincident geometry frequently encountered in other SOM-based visualizations (Skupin 2002). Esri ArcGIS was used to produce GIS-compatible representation of all SOMs and SOM-based overlays and to generate a series of visualizations.

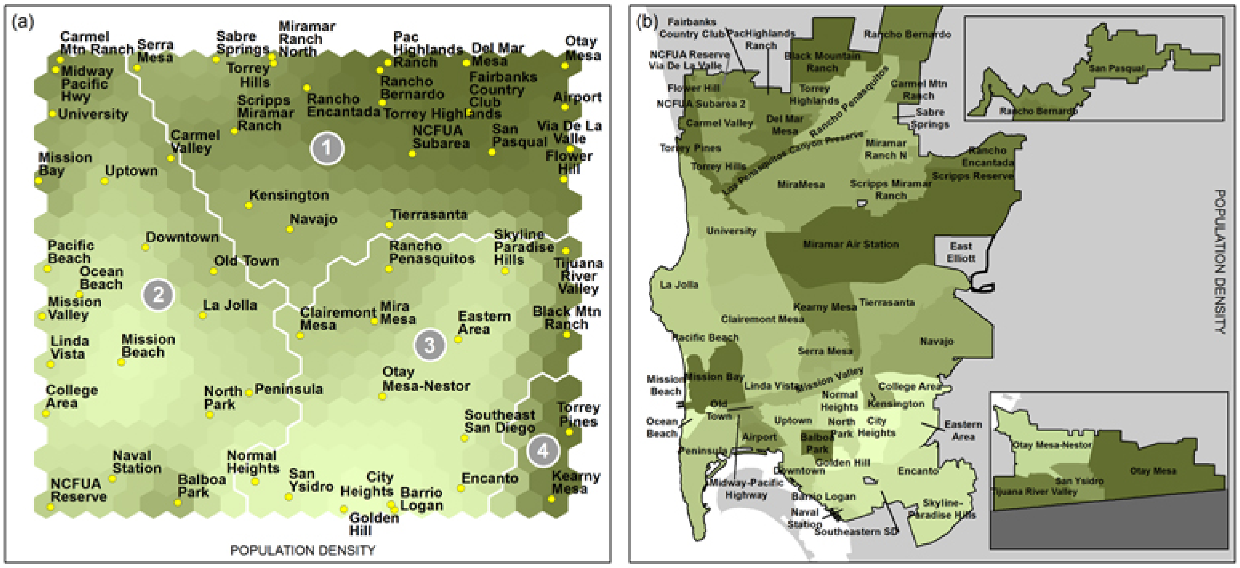


Figure 3. Panel (a) displays the population density phenomenon in attribute space as one component plane of the SOM, plus *k*-means clustering of neurons ($k = 4$) computed from all 136 attributes. Panel (b) displays the same phenomenon in geographic space.

V. Visualizations and Discussion

As suggested above, this study involved three forms of visualization. First, the SOM model itself is visualized, allowing visual examination of each dimension – or component plane – of the neural network. Second, the SOM is used as a base map onto which objects were mapped. Third, multiple attribute spaces are examined either by juxtaposition or by simultaneously letting different sets of attributes drive the geometry versus the symbology of a spatialization.

The following sections illustrate both the potential of geovisual analytics to engage with qualitative data and the kinds of knowledge this approach may help produce. In what follows, five SOMs are presented and various strategies for analyzing the visualizations are discussed. Two of these SOMs represent neighbourhoods in two distinct ways, two others represent people differently, and the last SOM represents descriptions of places. Finally, we argue that an approach based on the attribute space notion can lead to effective representations and explorations of the complex, high-dimensional attribute spaces related to personal impressions of places and that these attribute spaces form topologies quite distinct from those derived from census data.

1. VISUALIZATIONS

Neighbourhoods – Census

Displaying a particular attribute at the level of individual neurons is akin to typical single-attribute displays of choropleth maps. Accordingly, Figure 3 shows a single dimension – population density – visualized in attribute space and geographic space. On the left, the organization of the SOM has been constructed based on all 136 dimensions, but only one is showing. On the right is a standard choropleth map, showing the population density in geographic space. Again, both visualizations display one dimension of many, the choropleth map showing one dimension in geographic space, while the SOM component plane corresponds to that same dimension in attribute space. The SOM-based representation of the attribute tends to result in smoother patterns, since the SOM attempts to preserve attribute similarity. On the other hand, continuity in the geographic map would indicate possible spatial autocorrelation.

Note the similarity of places' attributes in the two spaces. Neighbourhoods falling in high population density areas of attribute space also exhibit high population density in the choropleth map, since both maps represent population density, though in different spaces. Some cartographic metaphors such as distance, clusters, and regions factor strongly into the interpretation of the SOM. For instance, the proximity of City Heights to Golden Hill and Normal Heights in attribute space indicates that they are similar across *many* attributes; in contrast, although College Area shares population density characteristics with the other three neighbourhoods, its distance from them suggests that other characteristics were markedly different.

With a SOM constructed from such a large number of attributes, continual engagement of the original 136-dimensional space is important as one explores its two-dimensional depiction. One approach is the U-matrix method

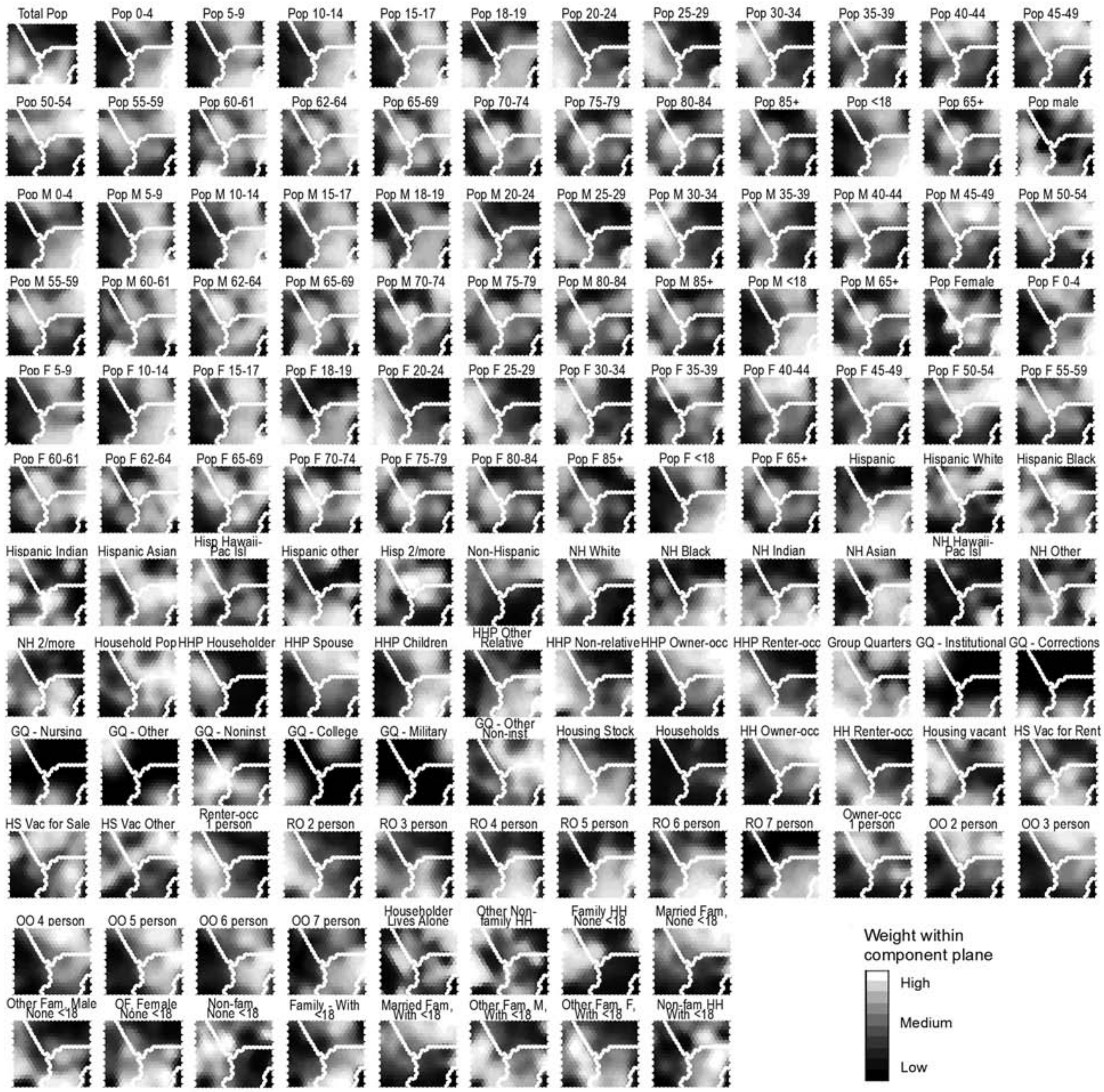


Figure 4. The 136 component planes that combine to comprise the SOM, plus overlay of *k*-means clustering of neurons

(Ultsch 1993), which visualizes differences in local distortion of high-dimensional relationships. But that method leaves the explicit delineation of cluster boundaries to the subjective choice of the analyst and is not capable of detecting non-local distortions. A more desirable approach would involve the strictly computational clustering of neurons, irrespective of their two-dimensional arrangement. To that end, neuron vectors were clustered using a *k*-means approach and the result projected onto the two-dimensional neuron lattice (Figure 3, left). Since this is computed in a strictly aspatial manner, without concern for neurons' two-dimensional arrangement, the degree of contiguity of the resulting two-dimensional cluster layout can serve to validate the success of SOM training, beyond what traditional validation approaches such as the quantization error are able to provide (Skupin and Esperbé 2011). For the purposes of our study, the overlay of cluster boundaries is meant to serve as a type of reference system that allows one to draw connections across multiple depictions of the SOM. Three clusters dominate the space (numbers 1, 2, 3), while the fourth cluster contains only two neighbourhoods.

In Figure 4, all 136 dimensions of the SOM are shown side by side, together with the *k*-means cluster boundaries. The full set of component planes is here provided to allow readers to explore the distribution of attribute weights

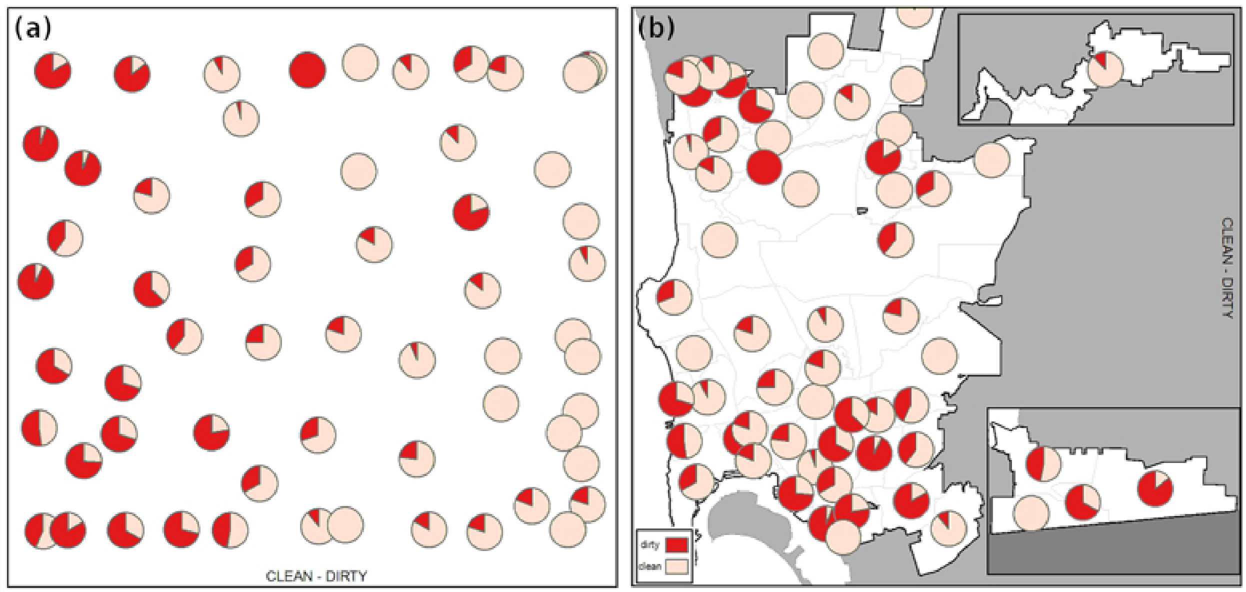


Figure 5. Neighbourhoods symbolized by the proportion of "clean" (light grey; online beige hue) descriptors to "dirty" (dark grey; online red hue) (Source: Burns and Skupin 2009). Panel (a) shows the phenomenon in attribute space with a heavier distribution of "dirty" on the left side of the map, and Panel (b) shows the phenomenon in geographic space, with "dirty" distributed more toward the south and far northwest.

across all neurons. One might argue that it is problematic to include such a large number of demographic variables, given the amount of duplication. For example, we include the overall proportion of the population aged 45–49 as well as a further breakdown into men aged 45–49 and women aged 45–49. The resulting component planes look indeed very similar. Likewise, neighbourhoods with a high proportion of people over age 65 also have large proportions of people over age 85. But the inclusion of *globally* correlated variables (e.g., overall age-group proportions and gender-specific age-group proportions) is done in the hopes of preserving interesting *local* patterns (e.g., those occurring in connection with military and correctional group quarters). Keep in mind also that there may be surprising heterogeneity in global correlation. In addition, the SOM method is less susceptible to the additive effects of duplicative variables, due to its inherent focus on preserving topological patterns instead of metric distance relationships.

Inclusion of *k*-means cluster boundaries serves two purposes – namely it helps to explain what causes these cluster structures, and the broad cluster tessellation helps to visually structure the continuous surface depiction of neurons. Thus we can see that cluster 1, for example, is characterized by lower population density, higher rate of housing ownership, and higher percentage of people in age cohorts 40–69 but with mostly two- to five-person households and small percentage of non-Hispanic blacks. Larger households, in both owner- and renter-occupied categories, tend to be found in cluster 3, as are Hispanic and black populations. Cluster 2 is meanwhile characterized by a relatively low number of children but elevated rates of people in their twenties and early thirties (with the exception of some neighbourhoods in the extreme bottom-left corner of SOM) and a high percentage of renter-occupied housing. Cluster 4 contains only the Torrey Pines and Kearny Mesa neighbourhoods, which, within the 100-foot buffer, are unique in that the video for the former captured a shopping district in which only seven people reside, and the latter video captured mostly a juvenile detention centre.

Neighbourhoods – Utterances

In contrast to the analysis of census data, a SOM was generated that expresses similarity of neighbourhoods in terms of how people characterize their mediated experience of them. The ways in which people experience and come to know places can result in distinct topologies between places. Whereas the previous visualization of census data lends useful information about places and their relations to each other, the data represent neighbourhoods' resident characteristics rather than the distinct experiences one may have in those neighbourhoods. The rich contextual and personal nature of emotions, experiences, and descriptions lend themselves more readily to a fluid reading of the map by encouraging the multiple meanings and interpretations advocated by geovisual analytics. Figure 5 illustrates this distinction. As before, Figure 5 shows another juxtaposition of attribute space and geographic

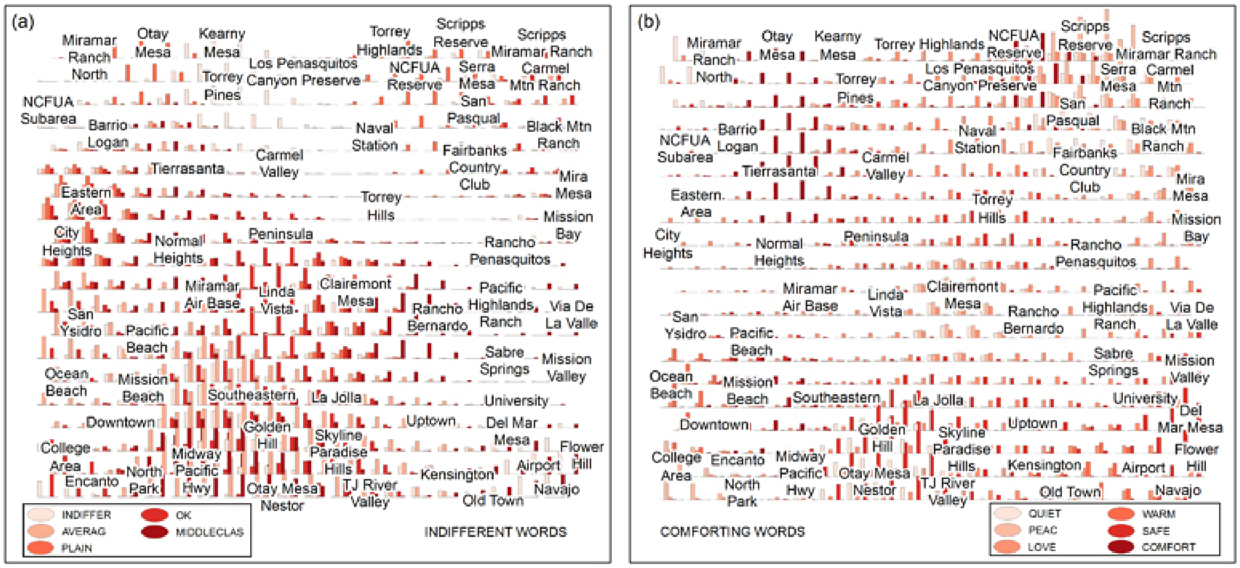


Figure 6. Indifferent words in Panel (a) are clustered toward the bottom-left, while comforting words in Panel (b) are more toward the right side of the map. Symbologies for both were derived from the same SOM.

space, but here the organization has been derived from the utterances people used to describe the neighbourhoods. Both maps display the same variable, and each object in attribute space corresponds to a neighbourhood in geographic space. In this visualization, the SOM model – the composite of the component planes – has been removed, leaving only the neighbourhoods that had been mapped onto it. The pie chart for each neighbourhood represents the ratio of times a neighbourhood was described as "dirty" relative to "clean." "Dirty" is represented with the red hue and "clean" with the beige hue. Note that these are 2 of the 80 total dimensions that contributed to the training of this SOM.

This figure shows how the SOM, without "understanding" semantics of the terms involved in training, tends to separate unlike terms in its organization. This organization here results because there was general agreement among participants about whether a neighbourhood would be described as "clean" or "dirty." Also, neighbourhoods with one of these descriptors tended to be described differently *overall* from neighbourhoods described with the opposite term. In other words, although a machine-based system cannot grasp the qualitative nuance of these descriptors, it can work through attribute spaces to render the differences visually; the researcher may then discern patterns. This figure additionally demonstrates people's impressions spatialized in both a geographic and a relational sense. One's impression is geographical because of its rootedness in a place; more complex, however, is how one's impressions of places are always shaped in relation to *other* places. Although not explicitly displayed in Figure 5, people's impressions of places are also always relational to one's own background, social standing, and perhaps even their mood when experiencing a place. This partially accounts for the fact that, to varying degrees, neighbourhoods are usually described as both dirty and clean, and usually not exclusively one or the other.

Comparing different word classes lends insights into the organization of the SOM and the survey respondents' predominant descriptions of particular neighbourhoods. Figure 6 juxtaposes two subjectively chosen classes of words to compare how the SOM organization was influenced by each class. In this figure, attention should be directed toward comparing the broad structural and pattern similarities and differences between the two panes, rather than to interpreting individual bar charts. Since the two visualizations in Figure 6 are based on the same SOM model with different dimensions represented, direct comparisons between the dimensions are possible. This geovisual analytic strategy might show, for instance, linkages between people's descriptions of neighbourhoods and the resulting SOM topology. The first class of words in Figure 6 comprises words that express indistinctive, indifferent feelings while the second comprises words that express feelings of comfort. Note that there is a strong discrepancy on the left side where neighbourhoods were described fairly commonly as "plain," "average," and "OK," but very rarely by any comforting words. Figure 7 is, again, the same SOM model as in Figure 6, with different dimensions represented. Figure 7 shows words with negative connotations. Where in Figure 6 indifferent words showed strong presence in the lower-left side of the map, neighbourhoods described with negative-connoting stems such as "dirti" (i.e., dirty, dirtiness), "ghetto," "poor," and "danger" (i.e.,

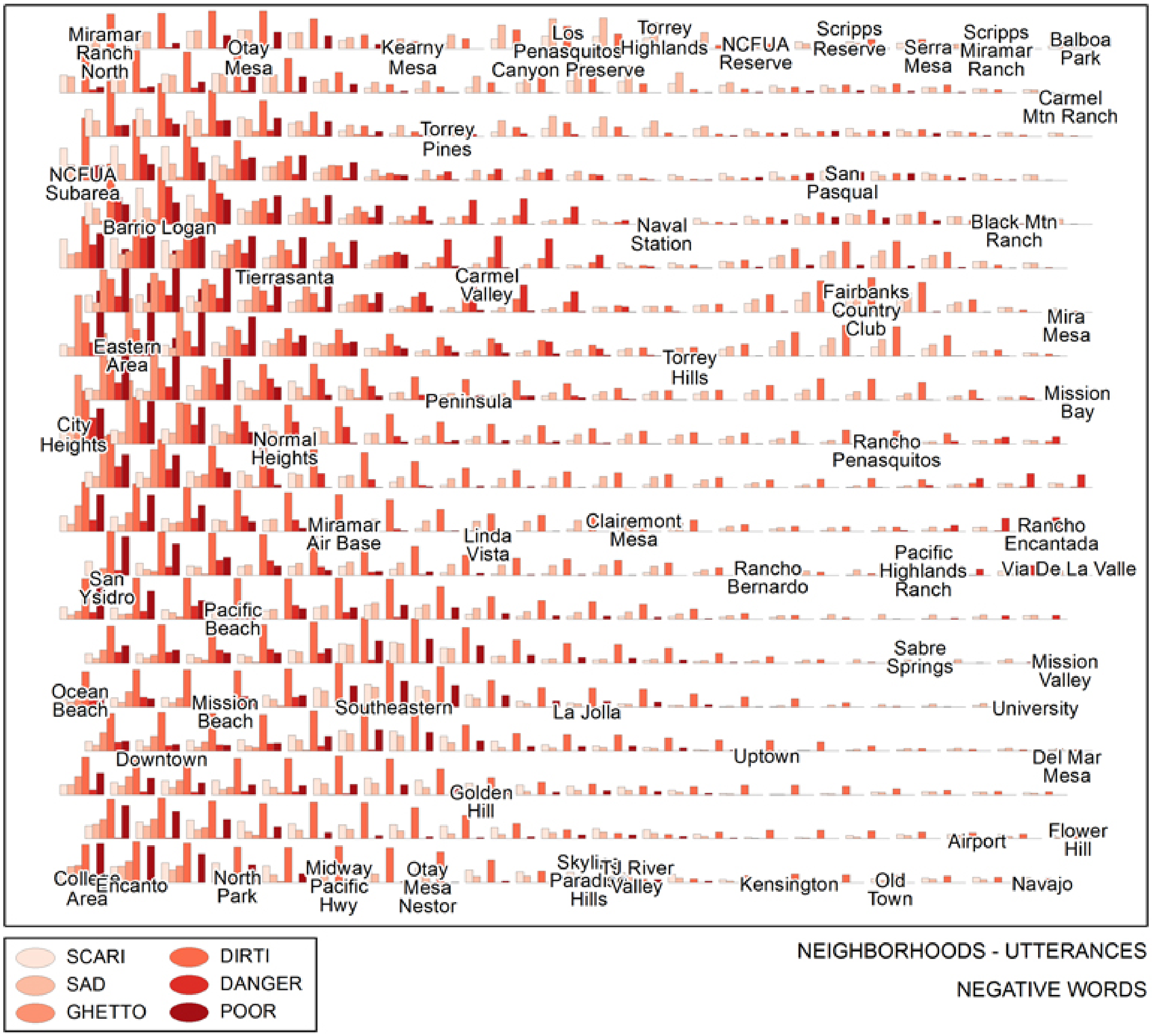


Figure 7. The distribution of negative words is heavier on the left side of the map

dangerous) were stronger on the left side of the map, more commonly in the upper-left corner. The discrepancies between positive and indifferent words are not seen in the upper-right-hand corner of the SOM, where neighbourhoods were described as "plain" but also with the stem "peac" (i.e., peaceful) and "comfort." Again, the peak of indifferent words in the middle-lower half of the SOM becomes particularly interesting when seen in relation to Figure 7, where negative-connoting words group largely to the left side of the map.

These visualizations illustrate strategies common to analysis of SOMs. In this particular case, there seems to be general agreement among research subjects about whether a neighbourhood can be described in positive, negative, or indifferent terms, since organization of these terms was mostly strongly polarized. Thus far we have also seen two ways of representing places, although these representations have been kept separated. Direct comparisons this way are limited, since the SOMs are based on different geometries. Comparing the attribute spaces may be undertaken differently.

COMPARING ATTRIBUTE SPACES

The discussion above showed how differences in the two representations of neighbourhoods result in different SOM topologies. As shown by Knigge and Cope (2006), qualitative data in a visualization context has the potential to lend new insights that complement quantitative measures of places. Here, the goal in comparing these representations of places is to complement purely quantitative representations of neighbourhoods. Two methods could be engaged for this comparison: comparison through

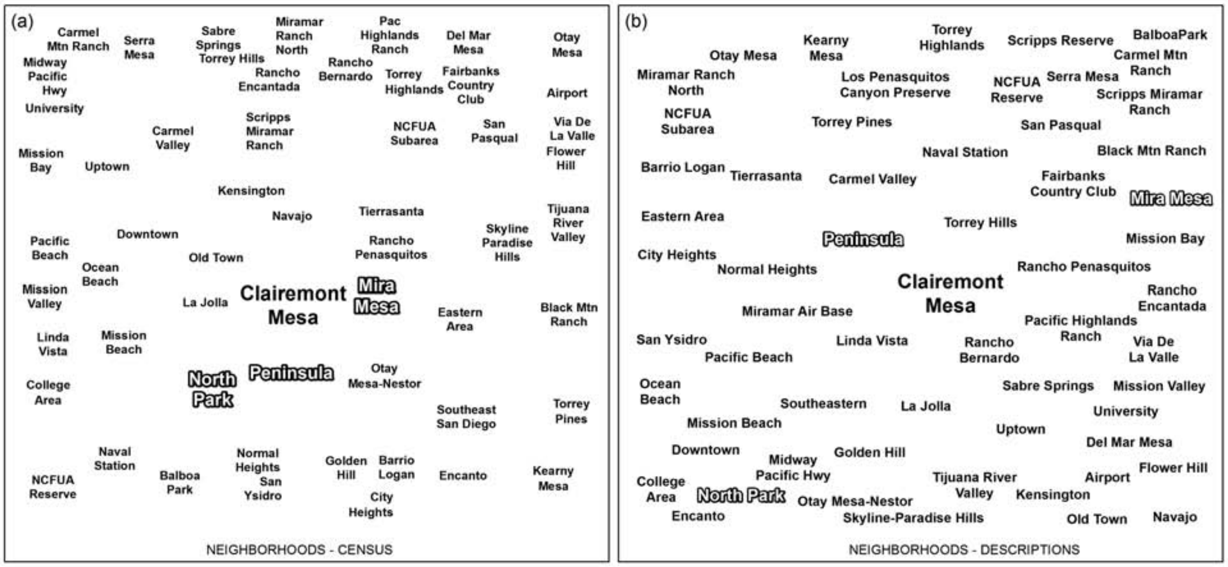


Figure 8. Juxtaposing two conceptualizations of neighbourhoods. Note the proximity of Mira Mesa, Peninsula, North Park, and Clairemont Mesa in Panel (a) and the dispersion of North Park and Mira Mesa in Panel (b)

juxtaposition, and use of one attribute space to train the SOM and another to symbolize objects mapped onto the SOM. Figure 8 juxtaposes the two conceptualizations of places. In the left pane is the SOM that has been trained using census data, and in the right pane is the SOM trained by people's descriptions of the neighbourhoods. The sets of markers in the figure highlight two common processes. First, the three neighbourhood labels with white text and black halos are three neighbourhoods with similar census attributes and thus were organized close together; in the right pane, they have spread to distant sides of the map. This suggests that while they share similar aggregate population attributes, people describe them very differently. Second, the neighbourhood label with grey text and a black halo highlights the centrality of Clairemont Mesa in both SOMs, suggesting its probable average character: most of its qualities probably represent the mean among all San Diego neighbourhoods.

The second method of comparing attribute spaces is to train a SOM with one attribute space and derive its symbology from another. Whereas the previous approach relies on comparing two separate visualizations, the second results in a single visualization that illustrates directly the relations between the two attribute spaces. The second approach also allows multiple component places of one attribute space to be represented on another attribute space's geometry. This approach here relies on the common unit of a neighbourhood: neighbourhoods' census data can be used to organize the SOM and the neighbourhoods' descriptions joined to the neighbourhoods for symbolization. If a correlation exists between the two attribute spaces, one would see distinct patterns emerge in the symbology. Figure 9 shows such a correlation. The SOM was trained by census values, but the symbology was derived from subjects' characterization of observed neighbourhood videos. Pastel hues have been used to denote positive words, while heavily saturated hues denote negative words. The size of the pie chart represents the number of times that neighbourhood's video was watched. Note that while significant variation exists in the distribution of hues, one can see distinctive regions emerge: many neighbourhoods with a strong minority presence have been organized to the bottom-right-hand corner of the map, an area that is dominated by negative descriptions. Another cluster of negative words appears further toward the top, near neighbourhoods that are sparsely populated and largely undeveloped. Positive descriptions stretch across the middle of the SOM with a few scattered throughout and are mostly beach neighbourhoods and neighbourhoods in the geographic north of San Diego – these northern neighbourhoods also being high-income suburban places. In this visualization geography, census attributes and personal impressions of neighbourhoods can be seen interacting on several levels.

Cross-symbolizing allows one to see interactions of multiple attribute spaces. Particular qualities of one attribute space may become more prominent when symbolized with another, and vice versa. These cross-symbolizations show that while some similarities exist between attribute spaces, they represent different aspects of places and result in distinct topologies between places. As we have explored personal descriptions of places, it may be useful to visualize the people themselves, to see personal characteristics and the utterances they used to describe the neighbourhoods we have visualized.

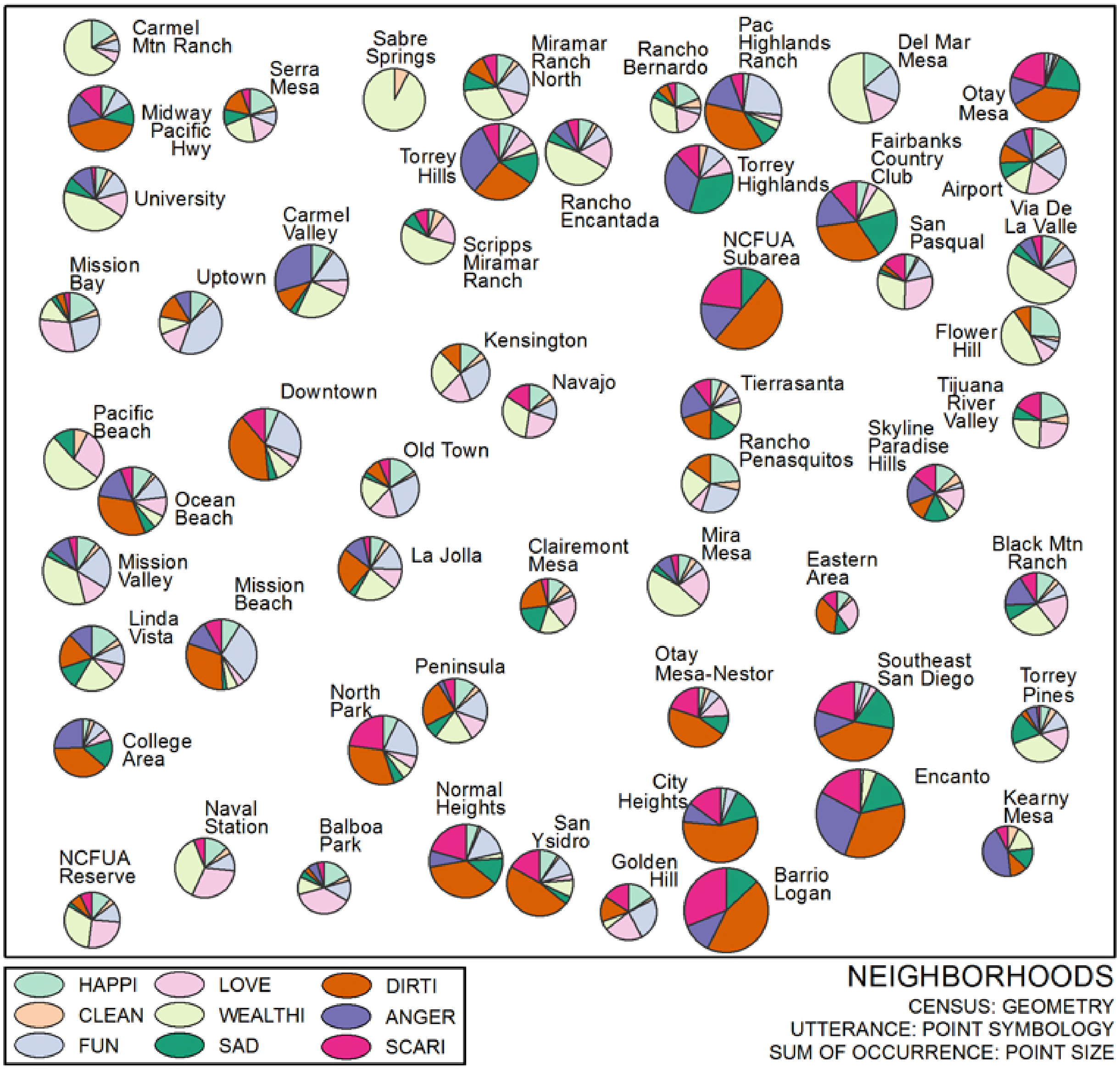


Figure 9. A SOM organized by census data and symbolized by utterances. Saturated hues represent negative utterances and pastel hues represent positive utterances. The size of the pie chart represents the number of times that neighbourhood's video was watched.

Subjects – Personal Attributes

Exploring SOMs related to survey respondents' personal characteristics reveals insights into the relations among respondents and the degree to which certain characteristics correlate with others. Visualizing subjects may also present a more complex understanding of neighbourhoods, since it is the subjects producing the descriptions; in other words, the descriptions have direct links to individual subjects. Essential to a qualitative approach, attention must be paid to the subjects contributing the descriptions. This is important to consider because who participates in the study and who describes the neighbourhoods has implications for the ways the neighbourhoods are described. Note that the topology for this SOM is derived from a different data set than the two described above. This makes direct comparison of the SOMs impossible. The SOM trained by subjects, where the dimensions come from their personal attributes, gives information about the subjects themselves rather than neighbourhoods. As mentioned earlier, this data set comes from the information respondents reported about themselves.

Figure 10 shows two representations of the same SOM, the left pane showing subjects symbolized by both sex and by religion and the right pane showing subjects symbolized by their socioeconomic status and year in school. In this figure, we observe survey respondents being sorted

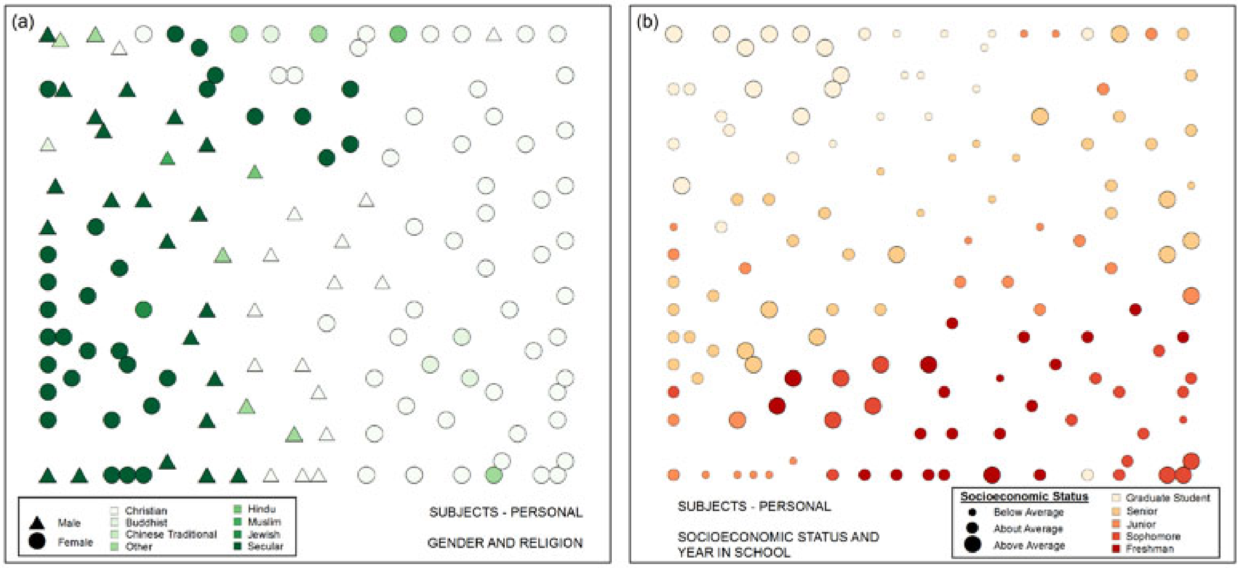


Figure 10. Survey respondents tended to divide along registers of religion and level in school. In Pane (a), men contained the boundary between Christians and those of other religions, while in Pane (b) graduate students with high socio-economic status diverged from others with this status.

along four dominant registers. First, the relative homogeneity of the right half of the left pane suggests that those who reported "Christian" as their religion tended to share other characteristics in common among themselves, characteristics not as often shared with those reporting a religion other than Christianity. In other words, if one's religion was Christian, we would likely see other characteristics in common among Christians. Second, sex acts as an axis of difference. In the left pane, men tend to fall in the middle of the SOM, with women falling on either side. This has the double effect of allowing men of all religions to be placed near each other and of forming the transition from Christianity to other religions. This can be explained as partly the result of men sharing more qualities in common with other men than with women. Third, the right pane shows a marked axis of freshmen and sophomores in the bottom-right and juniors, seniors, and graduate students in the upper-left half of the SOM. Nearly all graduate students coalesce to form an exclusive zone in the upper-left corner, and seniors form the adjacent exclusive zone. As before, this indicates that one's year in school serves as a useful predictor for other shared characteristics. Fourth, there are two distinct clusters of those who reported high socioeconomic status. The most prominent of these is in the upper-left corner of the right pane. This cluster consists solely of graduate students. The second cluster is toward the bottom-left of the SOM and consists of those from every other level in school. The separation of these clusters suggests that graduate students with a high socioeconomic status shared few other characteristics in common with others at different academic levels and high socioeconomic status.

Subjects – Utterances

Visualizing subjects based on their personal characteristics shows the kinds of people that participated in the survey; in contrast, we can say less about the subjects when we visualize them by the words they used. A SOM of respondents organized by the words they used reveals a strong diversity in terms, this diversity being typical across all subjects regardless of personal characteristics. In fact, one would expect this to be the case, since each respondent presumably could have watched videos of different neighbourhoods. Thus, two respondents with different personal characteristics may have used identical vocabulary, and vice versa. Since this SOM is organized by the words one used, and not the neighbourhoods described by those words, this SOM gives insight only into the vocabulary used by subjects; unsurprisingly, this vocabulary did not significantly vary along lines of subjects' personal characteristics.

Figure 11 illustrates these challenges involved in extracting meaningful information from a SOM where subjects were organized by the utterances they used. Symbology is derived by joining the subjects' personal characteristics to these objects, although these characteristics did not contribute to the SOM training. In other words, one attribute space was used for SOM training and a second attribute space used for symbolization. For more easy comparison, Figure 11 shares the same symbology schemata as Figure 10. Whereas in Figure 10 one can observe the organization of subjects by their personal characteristics, Figure 11 shows that these personal characteristics had little impact on the vocabulary used throughout the study.

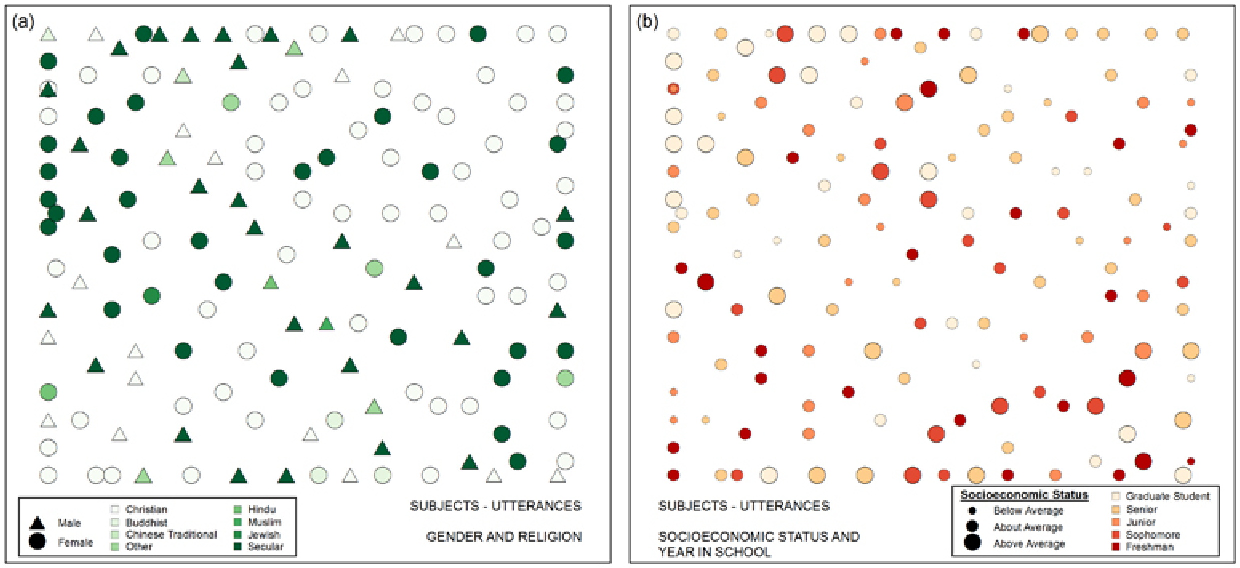


Figure 11. Personal characteristics had little impact on the vocabulary used by respondents. The distribution of respondents' personal characteristics appears to be more or less random in both Panels (a) and (b).

No conspicuous patterns exist in this SOM, and this randomness indicates a lack of correlation among the two attribute spaces (SOM organization and object symbolization).

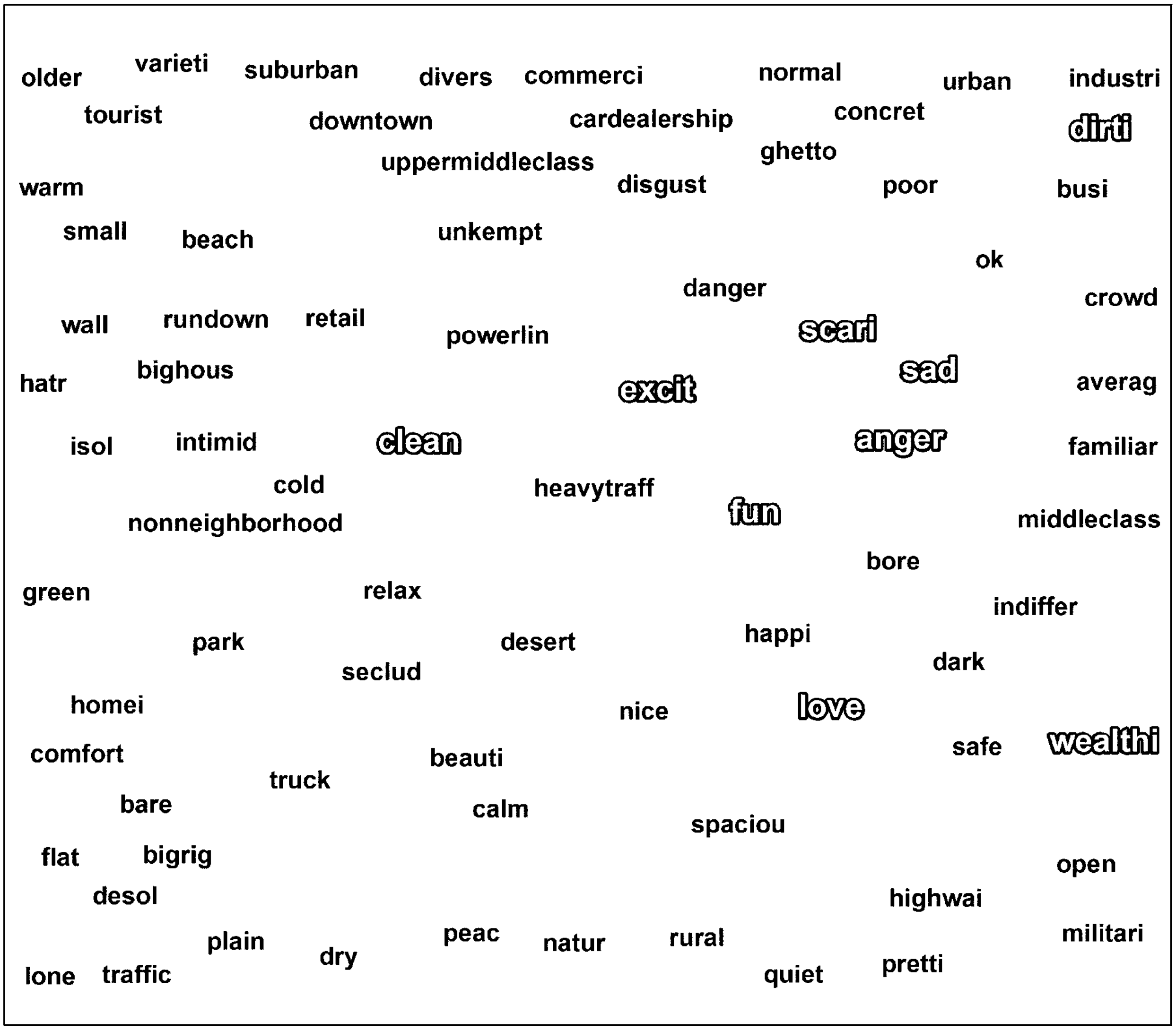
Visualizing subjects by their personal characteristics tells us more about the subjects than when we visualize them by their utterances. Since respondents were allowed to describe neighbourhoods openly and view any neighbourhood, we would not expect an individual's descriptions en masse to differ significantly from another's. If, however, each respondent had viewed the same neighbourhoods and chosen only from a list of predefined terms, more significant variation may have been visible.

Utterances – Neighbourhoods

The final SOM was of the individual words, organized by the neighbourhoods described *by* those words. For every word, its attribute space was the set of neighbourhoods it described. This is the transpose of the SOM where neighbourhoods were organized by their descriptors. We would expect this SOM to show a topological relationship between terms that may correspond with a general understanding of the terms (e.g., *nice* is near *lovely*) or present unexpected relations. The interest in this visualization stems from its mapping of the discourse (used broadly) used in describing the neighbourhoods. This will make visible all the terms used in the survey and show another way of thinking about the relationship between utterances and places. Since this visualization is the transpose of the SOM where neighbourhoods were organized by their descriptions, they could be similar; differences would be of particular interest.

Figure 12 shows the utterances mapped onto the SOM, with the SOM model removed. The nine predefined terms have been symbolized with a star to accentuate their location. This figure shows the resulting organization of the terms but not the topological network underlying them. Interestingly, the nine predefined terms fall relatively near each other, with seemingly contradictory terms organized close together. Aside from these terms, a predictable layout has emerged in this SOM. A look at the underlying SOM model can provide an explanation of why, among the 80 total utterances mapped, these 9 terms can be located near each other.

Figure 13 shows the model for this SOM, with 10 of the 60 total dimensions symbolized for each neuron of the SOM. Note the sharp change in pie charts between the seemingly opposite of the nine predefined terms. This marked shift indicates a break in the topological similarity between individual terms and, as mentioned above, nuances the general principle that nearness suggests similarity. While nearness can be suggestive of similarity, the topological structure of the SOM is a more accurate indicator. The neighbourhood dimensions have been symbolized by relative geography in the city, with the green hues being northern neighbourhoods, orange symbolizing central neighbourhoods, and blue-yellow-red symbolizing southern neighbourhoods. La Jolla, in pink, is also a northern neighbourhood but less forested than the other three, which are largely undeveloped in the videos. In Figure 13, most of the terms signifying undeveloped space (e.g., *spaciou*, *quiet*, *natur*) have amalgamated in the lower half, centre of the SOM. The central neighbourhoods form distinct regions in the upper-left-hand corner and around the stems "excit" and "danger." Most notable, however, is



UTTERANCE - NEIGHBORHOODS

Figure 12. Utterances here are organized by the neighbourhoods each utterance describes. The nine pre-chosen words have halos, and show a small amount of clustering.

the indication that southern neighbourhoods most strongly appear near negative stems such as "poor" and "scary."

VI. Conclusions

Places can be conceptualized and represented with a multitude of attributes. Metrics exist to capture different empirical representations of places, and different theoretical approaches have been developed to explain how particular conceptualizations mask processes made visible by others. Qualitative GIS has made explicit this tension and has sought to engage multiple forms of spatial reasoning. That research has also suggested that mixed methods research may productively lend insights into people's knowledge of their environments. Our project visualizes the intersections between these multiple attribute spaces and distinct conceptualizations of San Diego neighbourhoods. By questioning the relationship between personal characteristics and perceptions of places, we borrow insights from critical geographic theory and mixed methods, applying them in a geovisual analytics context.

This hybrid approach – working with qualitative descriptions in a visualization environment – opens new possibilities for both geovisual analytics and those working with qualitative data. First, this approach can "scale up" what is possible in qualitative research, as compared to the relatively small data sets that have typically been dealt with. While GIS typically involves visual and computational analysis of a limited number of dimensions, the

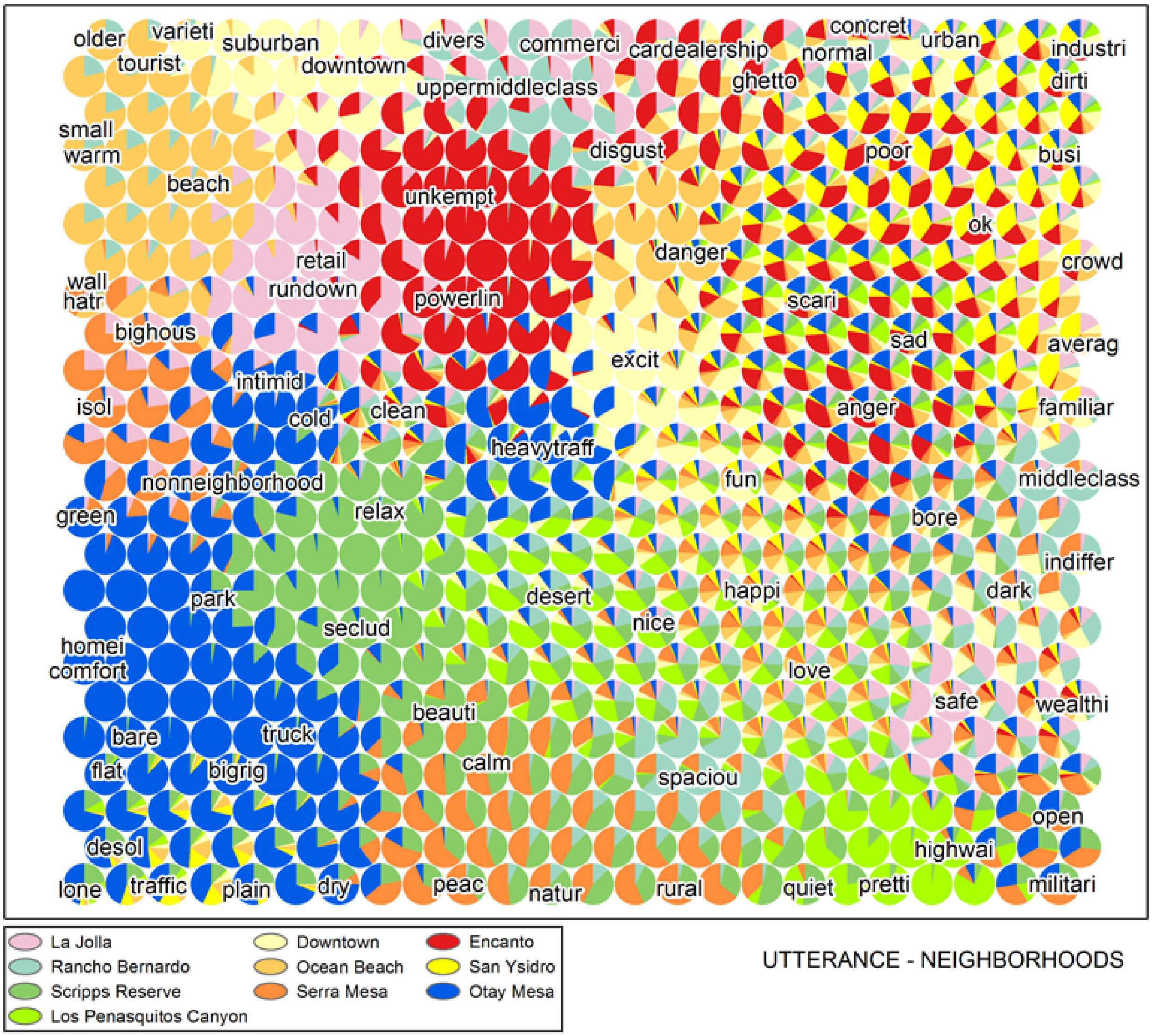


Figure 13. Ten dimensions of the SOM where utterances' geometries are determined by the neighbourhoods described with that utterance. This figure shows a sharp division between negative and positive pre-chosen stems and a slight clustering in the lower left of words denoting undeveloped space.

approach discussed here allows working with multi-faceted, high-dimensional data sets.

Second, this approach complements quantitatively driven geovisual analytics research by offering potential geovisual analytics engagements with qualitative data. Analyzing qualitative data allows the researcher to preserve the nuanced terms with which people describe their perceptions, experiences, and knowledge. According to the qualitative GIS literature, this richness is not easily retained through quantification. Our work demonstrates that approaches borrowed from information science, such as for representing and transforming text data, can be helpful. Integrating qualitative data also allows the researcher to move from abstract metrics such as census data to data grounded in individual backgrounds and lived experiences. More broadly, qualitative data can enrich the ways geovisual analytics explores geographic phenomena by illuminating a different empirical slice of reality.

Third, this hybrid approach accentuates the potential multiple meanings of visualizations. Since the current project's qualitative data reflects personal characteristics as well as complex understandings of San Diego neighbourhoods, the nature of the data lends itself to multiple analyses and engagements. Influences on SOM organization come from information gathered in the survey, such as personal backgrounds and characteristics, but also factors outside the data gathered in the survey, such as nuance in semantics and familiarity with the areas. Since this survey queried only 150 San Diego State University students, the diversity of backgrounds and perspectives is more limited than what might be expected in a larger-scale study; still, the amount of diversity in the present study has demonstrated potentially multiple meanings to each visualization. As shown here, multiple meanings emerge largely due to the combination of qualitative data, quantitative data, and attribute space notions. These meanings become more refined in the process of engaging the visualizations' environments, supporting a continual re-examination of detected patterns. In addition, neighbourhoods' attribute spaces *visualize* differently when looking only at census data or only at qualitative data, leading researchers to detect different relations between neighbourhoods. We can efficiently observe these distinctions through juxtaposed visualization and cross-symbolization, both techniques being easily executed with SOMs.

This study's limitations point to productive future research possibilities. First, all research subjects were San Diego State University students over the age of 18. Although the survey was held openly online and was marketed widely across the campus, the subjects represent a relatively homogeneous sample that may not be generalizable. Future studies might benefit from a larger and more diverse sample. Second, further studies might investigate additional interactions between attribute spaces. For instance, beyond the scope of this study was the influence a person's personal characteristics might have on the utterances they use. One potential question that could guide this research – left unanswered here – is how the presented approach could potentially skew, distort, or mislead our understanding of how people's personal characteristics and background affect their impression of mediated experience of geographic space. Third, exploring these additional attribute space interactions would be made much simpler with higher levels of interactivity. The current trend toward highly interactive and integrated visualization software environments within geovisual analytics might be useful to this potential work. Fourth, while our data sets captured a static glimpse of survey respondents' impressions of neighbourhoods, future research could add rich complexity to such data sets by treating impressions as dynamic and rapidly changing affective responses to the neighbourhoods represented. Fifth, generalizable lessons from our study are necessarily limited by its geographic scope, using data on only 60 neighbourhoods within a single city. In fact, use of the SOM method may be considered overkill and alternative dimensionality reduction techniques could be considered, such as multidimensional scaling. But that situation dramatically changes for much larger data sets involving multivariate attributes for perhaps hundreds or thousands of geographic regions. For example, SOM has already proven to be scalable to a data set of all 200,000-plus census block groups (Skupin 2009). That can lead to novel geometries, in particular when combined with multivariate clustering techniques, which have been a mainstay of geovisual analytics. The promise of that approach is to make it possible to consider and directly compare neighbourhoods in *different* cities. Finally, when much larger and more diverse groups of individuals are then given an opportunity to share their personal and mediated experiences of these diverse locations in a crowd-sourced manner – perhaps using a Mechanical Turk methodology – our conceptual approach intersecting the high-dimensional attribute spaces of places and people can indeed converge toward a qualitatively enriched geovisual analytics approach to the human experience.

Author Information

Ryan Burns is a doctoral candidate in geography at University of Washington and a research assistant with the Commons Lab at the Woodrow Wilson International Center for Scholars. His research interests are in the social and political implications of new mapping and mass collaboration technologies, and specifically how these technologies are used in humanitarian management. He is affiliated with UW's Simpson Center through the Certificate in Public Scholarship program, and serves on the leadership boards of two specialty groups of the Association of American Geographers.

Dr. **André Skupin** is Professor of Geography at San Diego State University. He holds a Dipl.-Ing. in cartography from the Technical University Dresden, Germany, and a PhD in geography from the State University of New York at Buffalo. Areas of interest and expertise include geographic visualization, visual data mining, information visualization, and spatio-temporal modelling; he has published work in *Proceedings of the National Academy of Sciences*, *PLoS ONE*, *Journal of Informetrics*, and *Journal of Visual Languages and Computing*. He has been involved in developing novel methods for analyzing human mobility, demographic change, and environmental sensor data in *n*-dimensional attribute space. Much of his research has addressed modelling knowledge artefacts through integrating natural-language processing with geographic principles and cartographic techniques.

Note

1. A more expansive description is provided by Kohonen (1990; 2001), the method's creator.

References

Agarwal, Pragya, and André Skupin, eds. 2008. *Self-Organising Maps: Applications in Geographic Information Science*. Chichester, UK: Wiley.

Al-Kodmany, Kheir. 2000. "Extending Geographic Information Systems to Meet Neighborhood Planning Needs: The Case of Three Chicago Communities." *URISA: Journal of the Urban and Regional Systems Association* 12 (3): 19–38.

Andrienko, Gennady, Natalia Andrienko, S. Bremm, T. Schreck, T. von Landesberger, P. Bak, and Daniel Keim. 2010. "Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatio-temporal Patterns." *Computer Graphics Forum* 29 (3): 913–22. http://dx.doi.org/10.1111/j.1467-8659.2009.01664.x

Andrienko, Gennady, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. 2010. "Space, Time and Visual Analytics." *International Journal of Geographical Information Science* 24 (10): 1577–600. http://dx.doi.org/10.1080/13658816.2010.508043

Andrienko, Gennady, Natalia Andrienko, Jason Dykes, Sara Fabrikant, and M. Wachowicz. 2008. "Geovisualization of Dynamics, Movement and Change: Key Issues and Developing Approaches in Visualization Research." *Information Visualization* 7 (3–4): 173–80. http://dx.doi.org/10.1057/ivs.2008.23

Andrienko, Gennady, Natalia Andrienko, Piotr Jankowski, Daniel Keim, Menno-Jan Kraak, Alan MacEachren, and Stefan Wrobel. 2007. "Geovisual Analytics for Spatial Decision Support: Setting the Research Agenda." *International Journal of Geographical Information Science* 21 (8): 839–57. http://dx.doi.org/10.1080/13658810701349011

Andrienko, Gennady, Natalia Andrienko, Daniel Keim, Alan MacEachren, and Stefan Wrobel. 2011. "Challenging Problems of Geospatial Visual Analytics." *Journal of Visual Languages and Computing* 22 (4): 251–6. http://dx.doi.org/10.1016/j.jvlc.2011.04.001

Anselin, Luc. 2012. "From SpaceStat to CyberGIS Twenty Years of Spatial Data Analysis Software." *International Regional Science Review* 35 (2): 131–57. http://dx.doi.org/10.1177/0160017612438615

Brown, Michael, and Larry Knopp. 2008. "Queering the Map: The Productive Tensions of Colliding Epistemologies." *Annals of the Association of American Geographers* 98 (1): 40–58. http://dx.doi.org/10.1080/00045600701734042

Buckley, Aileen R., Mark Gahegan, and Keith C. Clarke. 2000. "Geographic Visualization." In *A Research Agenda for Geographic Information Science*, ed. R.B. McMaster and E.L. Usery, 313–33. Boca Raton, FL: CRC Press.

Burns, Ryan, and André Skupin. 2009. "Visualization of Attribute Spaces Involving Places, People and Utterances." In *Proceedings of 24th International Cartographic Conference, Santiago, Chile*, 1–13. Santiago, Chile: International Cartographic Association. http://ns1.icaci.org/files/documents/ICC_proceedings/ICC2009/html/nonref/20_9.pdf

Crampton, Jeremy W. 2001. "Maps as Social Constructions: Power, Communication and Visualization." *Progress in Human Geography* 25 (2): 235–52. http://dx.doi.org/10.1191/030913201678580494

Dykes, Jason, Alan MacEachren, and Menno-Jan Kraak. 2005. *Exploring Geovisualization*. Amsterdam: Elsevier.

Elwood, Sarah. 2006. "Beyond Cooptation or Resistance: Urban Spatial Politics, Community Organizations, and GIS-Based Spatial Narratives." *Annals of the Association of American Geographers* 96 (2): 323–41. http://dx.doi.org/10.1111/j.1467-8306.2006.00480.x

Elwood, Sarah. 2009. "Multiple Representations, Significations, and Epistemologies in Community-based GIS." In *Qualitative GIS: A Mixed-Methods Approach*, ed. Meghan Cope and Sarah Elwood, 57–74. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9780857024541.d8

Elwood, Sarah, and Meghan Cope. 2009. "Introduction: Qualitative GIS: Forging Mixed Methods Through Representations, Analytical Innovations, and Conceptual Engagements." In *Qualitative GIS: A Mixed-Methods Approach*, ed. Meghan Cope and Sarah Elwood, 1–12. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9780857024541.d4

Fabrikant, Sara Irina, and Amy Lobben. 2009. "Introduction: Cognitive Issues in Geographic Information Visualization." *Cartographica* 44 (3): 139–143. http://dx.doi.org/10.3138/carto.44.3.139

Hand, D.J., H. Mannila, and P. Smyth. 2001. *Principles of Data Mining*. Cambridge, MA: MIT Press.

Harley, John Brian. 1989. "Deconstructing the Map." *Cartographica* 26 (2): 1–20. http://dx.doi.org/10.3138/E635-7827-1757-9T53

Jung, Jin-Kyu, and Sarah Elwood. 2010. "Extending the Qualitative Capabilities of GIS: Computer-Aided Qualitative GIS." *Transactions in GIS* 14 (1): 63–87. http://dx.doi.org/10.1111/j.1467-9671.2009.01182.x

Keim, Daniel, Jorn Kohlhammer, Geoffrey Ellis, and Florian Mansmann, eds. 2010. *Mastering the Information Age – Solving Problems with Visual Analytics*. Goslar, Germany: Eurographics Association.

Kirk, Roger E. 2004. "Maturation Effect." In *The Sage Encyclopedia of Social Science Research Methods*, ed. Michael S. Lewis-Beck, Alan Bryman, and Tim F. Liao, 618–19. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412950589.n534

Knigge, L., and Meghan Cope. 2006. "Grounded Visualization: Integrating the Analysis of Qualitative and Quantitative Data Through Grounded Theory and Visualization." *Environment & Planning A* 38 (11): 2021–37. http://dx.doi.org/10.1068/a37327

Kohonen, Teuvo. 1990. "The Self-Organizing Map." *Proceedings of the IEEE* 78 (9): 1464–80. http://dx.doi.org/10.1109/5.58325

Kohonen, Teuvo. 2001. *Self-Organizing Maps*. Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-56927-2

Kwan, Mei-Po. 2002a. "Is GIS for Women? Reflections on the Critical Discourse in the 1990s." *Gender, Place & Culture: A Journal of Feminist Geography* 9 (3): 271–79. http://dx.doi.org/10.1080/0966369022000003888

Kwan, Mei-Po. 2002b. "Feminist Visualization: Re-Envisioning GIS as a Method in Feminist Geographic Research." *Annals of the Association of American Geographers* 92 (4): 645–61. http://dx.doi.org/10.1111/1467-8306.00309

Lawson, Victoria. 1995. "The Politics of Difference: Examining the Quantitative/Qualitative Dualism in Post-structuralist Feminist Research." *Professional Geographer* 47 (4): 449–57. http://dx.doi.org/10.1111/j.0033-0124.1995.00449.x

MacEachren, Alan. 1995. *How Maps Work: Representation, Visualization, and Design*. New York: Guilford.

MacEachren, Alan, Mark Gahegan, William Pike, Isaac Brewer, Guoray Cai, E. Lengerich, and F. Hardisty. 2004. "Geovisualization for Knowledge Construction and Decision Support." *IEEE Computer Graphics and Applications* 24 (1): 13–17. http://dx.doi.org/10.1109/MCG.2004.1255801. Medline:15384662

MacEachren, Alan, and Menno-Jan Kraak. 2001. "Research Challenges in Geovisualization." *Cartography and Geographic Information Science* 28 (1): 3–12. http://dx.doi.org/10.1559/152304001782173970

Matthews, Stephen A., James E. Detwiler, and Linda M. Burton. 2005. "Geo-Ethnography: Coupling Geographic Information Analysis Techniques with Ethnographic Methods in Urban Research." *Cartographica* 40 (4): 75–90. http://dx.doi.org/10.3138/2288-1450-W061-R664

McCormick, B., T. DeFant, and M. Brown. 1987. "Visualization." *Scientific Computing in Computer Graphics* 21 (6): i–E-8.

McLafferty, Sara. 2005. "Women and GIS: Geospatial Technologies and Feminist Geographies." *Cartographica* 40 (4): 37–45. http://dx.doi.org/10.3138/1341-21JT-4P83-1651

Pain, Rachel, Robert MacFarlane, Keith Turner, and Sally Gill. 2006. "'When, Where, If, and But': Qualifying GIS and the Effect of Streetlighting on Crime and Fear." *Environment & Planning A* 38 (11): 2055–74. http://dx.doi.org/10.1068/a38391

Pavlovskaya, Marianna. 2006. "Theorizing with GIS: A Tool for Critical Geographies?" *Environment & Planning A* 38 (11): 2003–20. http://dx.doi.org/10.1068/a37326

Porter, M. 1980. "An Algorithm for Suffix Stripping." *Program* 14 (3): 130–7. http://dx.doi.org/10.1108/eb046814

Rocheleau, Dianne. 1995. "Maps, Numbers, Text, and Context: Mixing Methods in Feminist Political Ecology." *Professional Geographer* 47 (4): 458–66. http://dx.doi.org/10.1111/j.0033-0124.1995.00458.x

Salton, G. 1989. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Reading, MA: Addison-Wesley. http://www.iro.umontreal.ca/~nie/IFT6255/Introduction.pdf.

Schuurman, Nadine. 2000. "Trouble in the Heartland: GIS and Its Critics in the 1990s." *Progress in Human Geography* 24 (4): 569–90. http://dx.doi.org/10.1191/030913200100189111

Sheppard, Eric. 2005. "Knowledge Production through Critical GIS: Genealogy and Prospects." *Cartographica* 40 (4): 5–21. http://dx.doi.org/10.3138/GH27-1847-QP71-7TP7

Skupin, André. 2002. "A Cartographic Approach to Visualizing Conference Abstracts." *IEEE Computer Graphics and Applications* 22 (1): 50–58. http://dx.doi.org/10.1109/38.974518

Skupin, André. 2004. "The World of Geography: Visualizing a Knowledge Domain with cartographic Means." *Proceedings of the National Academy of Sciences of the United States of America* 101 (Suppl 1): 5274–8. http://dx.doi.org/10.1073/pnas.0307654100. Medline:14764896.

Skupin, André. 2007. "Where Do You Want to Go Today." In *Societies and Cities in the Age of Instant Access*, ed. Harvey Miller, 133–49. Dordrecht, The Netherlands: Springer. http://dx.doi.org/10.1007/1-4020-5427-0_9

Skupin, André. 2009. "Discrete and Continuous Conceptualizations of Science: Implications for Knowledge Domain Visualization." *Journal of Informetrics* 3 (3): 233–45. http://dx.doi.org/10.1016/j.joi.2009.03.002

Skupin, André, and Pragya Agarwal. 2008. "Introduction: What Is a Self-Organising Map?" In *Self-Organising Maps: Applications in Geographic Information Science*, ed. Pragya Agarwal and André Skupin, 1–20. Chichester, UK: Wiley.

Skupin, André, and Charles de Jongh. 2005. "Visualizing the ICA–A Content-Based Approach." In *Proceedings of 22nd International Cartographic Conference.* A Coruña, Spain: International Cartographic Association.CD-ROM.

Skupin, André, and Aude Esperbé. 2011. "An Alternative Map of the United States Based on an n-Dimensional Model of Geographic Space." *Journal of Visual Languages and Computing* 22 (4): 290–304. http://dx.doi.org/10.1016/j.jvlc.2011.03.004

Skupin, André, and Sara Fabrikant. 2003. "Spatialization Methods: A Cartographic Research Agenda for Non-geographic Information Visualization." *Cartography and Geographic Information Science* 30 (2): 99–115. http://dx.doi.org/10.1559/152304003100011081

Skupin, André, and R. Hagelman. 2003. "Attribute Space Visualization of Demographic Change." In *Proceedings of the Eleventh ACM International Symposium on Advances in Geographic Information Systems, New Orleans, LA, November 7–8*, 56–62. New York: ACM Press.

Slocum, T., R. McMaster, F. Kesler, and H. Howard. 2005. *Thematic Cartography and Visualization*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Thomas, J.J., and K.A. Cook. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA: IEEE Computer Society.

Thomas, J.J., and K.A. Cook. 2006. "A Visual Analytics Agenda." *IEEE Transactions on Computer Graphics and Applications* 26 (1): 10–13. http://dx.doi.org/10.1109/MCG.2006.5. Medline:16463473.

Ultsch, A. 1993. "Self-Organizing Neural Networks for Visualization and Classification." In *Information and Classification: Concepts, Methods, and Applications*, ed. O. Opitz, B. Lausen, and R. Klar, 307–13. Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-50974-2_31

Wilson, Matthew W. 2009. "Towards a Genealogy of Qualitative GIS." In *Qualitative GIS: A Mixed-Methods Approach*, ed. Meghan Cope and Sarah Elwood, 156–70. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9780857024541.d15

Yan, Jun, and Jean-Claude Thill. 2008. "Visual Exploration of Spatial Interaction Data with Self-Organizing Maps." In *Self-Organising Maps: Applications in Geographic Information Science*, ed. Pragya Agarwal and André Skupin, 67–85. Chichester, UK: Wiley. http://dx.doi.org/10.1002/9780470021699.ch4