

Capturing and mapping quality of life using Twitter data

Slavica Zivanovic · Javier Martinez  · Jeroen Verplanke 

© The Author(s) 2018

Abstract There is an ongoing discussion about the applicability of social media data in scientific research. Moreover, little is known about the feasibility to use these data to capture Quality-of-Life (QoL). This study explores the use of social media in QoL research by capturing and mapping people's perceptions about their life based on geo-located Twitter data. The methodology is based on a mixed-method approach, combining manual coding of the messages, automated classification, and spatial analysis. Bristol is used as a case study, with a dataset containing 1,374,706 geotagged Tweets. Based on the manual coding results, three QoL domains were analysed. Results show the difference between Bristol wards in number and type of QoL perceptions in every domain, spatial distribution of positive and negative perceptions, and differences between the domains. Furthermore, results from this study are compared to the official QoL survey results from Bristol, statistically and spatially. Overall, three main conclusions are

underlined. First, to an extent, Twitter data can be used to evaluate QoL. Second, based on people's perceptions, there is a difference in QoL between neighbourhoods in Bristol. And, third, Twitter messages can be used to complement QoL surveys, but not act as a proxy for traditional survey results. The main contribution of this study is in recognising the potential Twitter data have in QoL research. This potential lies in producing additional knowledge about QoL that can be placed in a planning context and effectively used to improve the decision-making process and enhance quality-of-life of residents.

Keywords Quality of life · Social media · Volunteered geographic information · Twitter data · Bristol

Introduction

Quality-of-life research and possibilities of social media as a new data source

Growing concern for differences within cities resulted in increased number of studies focused on community quality-of-life and well-being of the population (Costanza et al. 2007; Haas 1999; Pacione 2003a, b). Quality-of-life (QoL) is commonly defined as general satisfaction and well-being of individuals and

S. Zivanovic · J. Martinez (✉) · J. Verplanke
Department Urban and Regional Planning and Geo-Information Management, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, Netherlands
e-mail: zivanovic35831@alumni.itc.nl

J. Martinez
e-mail: j.martinez@utwente.nl

J. Verplanke
e-mail: j.verplanke@utwente.nl

communities in a specific surrounding across different domains (Davern and Chen 2010; Diener 2000; Marans 2003, 2015; Schuessler and Fisher 1985). QoL can be measured in an objective and subjective way with different sets of indicators proposed and used by various researchers (Mohit 2013). An objective approach measures QoL within different domains, using official statistics and information about the living environment, while a subjective approach evaluates levels of satisfaction people feel in or about a certain area. Although both approaches are present in current QoL research, in recent years, subjective measures are used more extensively. Interest in combining both approaches has increased as well (Ballas 2013).

Lately, new data sources, as well as new ways of collecting and analysing them, emerged in the scientific community. New technologies and new sources of information have been an important part of many urban policy initiatives (Shelton et al. 2015), and digital media has already been used to analyse different aspects of cities and spatial distribution of various urban functions (Shelton et al. 2015). Moreover, digital data are widely available and constantly multiplied in cyberspace, giving researchers the opportunity to go beyond official statistics (Shelton et al. 2015). Furthermore, social media data can have both geospatial footprints and indicative words that can be used in the process of collecting and analysing information.

Elwood et al. (2012) suggest that data produced on social media platforms can be observed as part of the Web 2.0 (participatory and social web), based on user generated content. According to these authors, people using social media are producing content and contributing to crowd-sourced sets of data by adding, knowingly or unknowingly (Harvey 2013), location to their posts. Social media data, when geo-located,¹ represent one type of Volunteered Geographic Information (VGI), or according to Kitchin (2014, 4) “data gifted by users”. However, unlike, for example, OpenStreetMap, where people choose to make a contribution by updating the existing geographic datasets (Yang et al. 2010), social media offers spatial

and temporal tagging of people’s raw thoughts (Shelton 2016).

An important aspect of present research is the fact that people tend to use social media platforms to express opinions about their life, how they emotionally feel and how they see their living surrounding in a self-reported way. This requires us to develop suitable steps to understand the nature of social media use and ways to analyse data derived from social media in QoL research.

Overall, the traditional collection of subjective perceptions can be time-consuming, expensive and slow (Bibo et al. 2014; McCrea et al. 2011). Due to this, data sources such as social media could play a significant role in capturing people’s perceptions. There is an ongoing discussion about the most appropriate measures of subjective QoL (Ballas 2013) and, moreover, about the applicability of social media in scientific research in general. Little is known about the feasibility to use social media data to capture people’s perceptions about their quality-of-life, and how traditional methods can be adapted for analysing data derived from social media. Therefore, the aim is to address this gap and contribute to the current discussion by exploring the use of social media data by capturing and mapping people’s perceptions about their life based on Twitter data within the context of subjective QoL research.

Subjective quality-of-life and the role of social media

Subjective QoL research

Subjective approaches in QoL research have a great potential in understanding the needs of individuals or communities. In various studies, depending on researched topics and areas of interest, subjective quality-of-life was introduced by different names and definitions. The terms well-being (Kapteyn et al. 2015), happiness (Diener 2000), good life (Bonn and Tafarodi 2013), and life satisfaction (Carlquist et al. 2016) are commonly used to address the same phenomena (Carlquist et al. 2016). Similarly, in the past few decades, defining subjective QoL has been a challenge and topic of many debates (Ballas 2013). Nevertheless, the subjective approach in quality-of-life research is commonly defined as a measure of

¹ Studies carried out by Leetaru et al. (2013) and Sloan and Morgan (2015) suggest that only a small percentage of Tweeter users (between 3 and 8%, depending on sampling and calculation) produced geotagged tweets.

people's feeling of general satisfaction with their living conditions (Berhe et al. 2014; Davern and Chen 2010; Diener 2000; Marans 2003, 2015; Schuessler and Fisher 1985; Tesfazghi et al. 2010).

The relevance of using a subjective QoL approach is emphasised by many researchers. For example, Moro et al. (2008) used subjective indicators with data collected in a self-reported way done through the national QoL survey to rank the level of satisfaction in Ireland. Similarly, Santos et al. (2007) used a survey to capture citizen's perceptions of life quality in Porto, Portugal, emphasising the importance of subjective measurements in defining urban policies and decision making. Some of the studies were more focused on evaluating the existing systems for measuring the subjective QoL. A good example is a study done by Wills-Herrera et al. (2009). They did a comparative, cross-cultural analysis of subjective well-being domains using Bogota, Belo-Horizonte, and Toronto as case studies to show how different global measurement systems can be applied at the city level.

Different methods have been used to capture and analyse QoL. However, the most common measures of QoL are identified as indicators, measured within different sets of domains, in objective or subjective way. Costanza et al. (2007) argue that objective indicators can be used to evaluate opportunities to improve people's life quality, but not directly measure the phenomena, and that subjective indicators should be used to provide meaningful insight into people's perceptions about their well-being. Pacione (2003b) indicated that *subjective social indicators* are a way to assess *urban liveability*, more precisely, the relation between people and their living environment. These subjective social indicators are focused on the self-reported perception of life satisfaction in a certain location and can be effectively used to assess differences in a neighbourhood QoL (Moro et al. 2008). The studies are often conflicting, favouring one approach over another. However, contemporary evaluations of QoL prefer the use of both approaches, since the combination is more informative to find the connection between people's perceptions and the objective conditions of their living environment.

Indicators are usually measured within different domains. The range of domains depends on the methodological approach and can be guided by theory or emerge from the residents themselves. As previously stated, in subjective QoL approaches

measurements mostly focus on self-reported statements about life satisfaction and experiences, to show the importance of the perceived need for a person's quality-of-life (Costanza et al. 2007). The decision about domains is usually guided by previously structured framework, based on QoL theory. Sirgy (2011) explains this as a top-down approach, where domain selection is guided by theory and previous knowledge, and, in his opinion, measures have more credibility. On the other hand, researchers like Dluhy and Swartz (2006) introduced the expansion of community-based projects, where domains and indicators are recognised by community members. According to Sirgy (2011, 2), this bottom-up approach is "essentially constrained in meaning or theoretical relevance".

In conclusion, many studies agree on the importance of using subjective assessment in examining QoL and understanding the issues and needs of residents in a particular area. In addition, there is an abundance of available methods to approach the evaluation and a clear distinction between top-down and bottom-up approaches in the domain definition. Their common denominator is a central role given to the people and their perception of QoL. The importance of local context is also emphasised. QoL domains depend on place, and the specific interaction people have with their surroundings (Tartaglia 2013). In the process of recognising domains for new research, study area and local context have to be included, and the domains covered in the official surveys and statistics have to be taken into account. The methodological approach has to be designed in a way it covers relevant questions and addresses important issues.

Social media in studying people's perceptions

Some authors prefer the term *social networks* while referring to *social media*. Conole et al. (2011) defined social networks as services that allow people to create public or private profiles, share their posts with chosen audience, and connect with a certain number of chosen individuals. Herein we will use the term *social media* as the data exchanged in a network to express perceptions, opinions, needs, interests, etc.

Although there are debates about the (re)usability of these data (Harvey 2013), numerous authors agree that data derived from social media represents a possible new source for gathering knowledge about

different societal issues (Aladwani 2015; Kusumo et al. 2017). Today, the problem is not how to get the data from social media, because there are many organisations involved in extensively collecting data for several years (Zook and Poorthuis 2015). The more important question is how to get meaningful insight.

Twitter² is one of the most used social media in studying people's perceptions (Arribas-Bel et al. 2015; Bibo et al. 2014; Chen and Yang 2014). For instance, in health science, various topics have been covered using social media data. Almazidy et al. (2016) developed a framework for harvesting Twitter data during a disease outbreak to have an additional source of knowledge about disease spreading patterns. Furthermore, Twitter data are also used in disaster management with an example provided by Chatfield et al. (2013). They examined the usability of the Twitter tsunami early warning system and the role of people in the transfer of information. Similarly, Kusumo et al. (2017) analysed the mapping of flood shelters and people's preferred shelter locations in Jakarta using Twitter data. Although the purposes for analysing social media data in these examples were different, all studies were focused on how people's opinions proved useful in assessing various phenomena, producing knowledge and transferring information.

One of the major advantages of social media is the opportunity to observe and analyse people's perceptions, opinions, needs, interests, etc. There is a possibility of gathering new knowledge from social media data to inform decision makers and contribute to urban planning and design processes (Larsson et al. 2016). Even though it is not very obvious, there is a strong connection between online and physical space, especially when geo-tagged social media data are analysed. Geo-tagged social media data include geographic coordinates of the location of the individual sharing the post. The advantage of Twitter, compared to other social media, is the possibility for the user to geo-tag Tweets which connects the message directly to the physical location where the message was sent from. Moreover, there are possibilities for using social media information in geospatial science and urban planning (e.g. spatial segregation, social profile

evaluation, measurement of satisfaction, traffic management) (Arribas-Bel et al. 2015).

One of the main benefits in using geo-tagged social media data is the possibility to integrate the results with more traditional research methods outcomes and different sources of knowledge (official statistics, urban plans, policies, etc.) and compare, complete and analyse the results and create better information about the dynamics of the urban area (Ciuccarelli et al. 2014a, b). Some might argue against the use of social media due to the lack of scientific tradition, but the richness and possibilities these data offer cannot be overlooked. Graham and Shelton (2013) expected that, based on the history of geography with diversity in theoretical and methodological paradigm and practices, the value of *big data* (large data sets produced in different manners with a potential to be mined for information, such as collection of Tweets) will be recognised in future research.

Social media in quality-of-life research

In quality-of-life research, Twitter was mainly used in health studies, evaluating quality-of-life based on health conditions. There are several studies where data collected from Twitter are used in creating indicators to assess the overall happiness and well-being of the population (Curini et al. 2015; Nguyen et al. 2016). Next, Bibo et al. (2014) used a Chinese social media platform similar to Twitter to assess the subjective well-being by collecting and analysing messages tagged with #SWB. They asked users to express their opinions and tag the messages with #SWB. Similarly, Dodds et al. (2011) tried to utilise data derived from Twitter to capture differences between several parts of the specific area in the matter of perceived happiness by using a previously developed tool named Hedonometer. Nguyen et al. (2016) used Twitter data to develop neighbourhood indicators for happiness, food, and physical activities. They used manual and automatic coding to capture indicative words to measure happiness, food consumption and leisure activities of the population. They concluded that social media provide formerly hard to obtain, costly data and can be used to give a better understanding of the community well-being.

Currently, there are few studies that have combined QoL research and social media data. These studies relate to overall perceived happiness and subjective

² Twitter is a free social networking service for interacting and networking with short messages "Tweets" in real time, restricted to 140 characters.

well-being (Curini et al. 2015), subjective well-being (Bibo et al. 2014), perceived happiness (Dodds et al. 2011) and Happiness, food and physical activities (Nguyen et al. 2016). The main challenges these authors encountered were about how representative the data were, issues with lack of technical knowledge, and limitation of the data itself. Using social media data involves a great deal of exploring in analysing the data and choosing proper methodology. Studies mentioned above used creative ways to adapt the traditional methods and develop new ones to address new types of data. Therefore, the present research will focus on identifying which QoL domains can be derived directly from the Twitter data and on capturing and mapping people's perceptions about their life quality within recognised domains.

Methodology, dataset and analysis

The methods described here explore the potential of using geo-located Twitter messages as a source of information about quality-of-life. The methodology herein suggested provides steps that are easily adaptable for utilising Tweets in (potentially) any geographic area and in any language. For the purpose of this research, the city of Bristol is selected as a case study area.

Case study area: the city of Bristol

Bristol is located in the southwest of England. It is the sixth largest city in England, and regional capital of this part of the country (Tallon 2007). According to mid-2016 population estimate, the population size in Bristol was 454,200. Bristol is a diverse city with many different cultures living together and sharing the living environment. Even though the city has a satisfying living condition, citizens are facing issues that affect their quality-of-life (McMahon 2002). In several parts of the city, wellbeing and health inequalities are emphasised. Moreover, Bristol has issues with traffic congestion, pollution and expensive housing compared to income. The Bristol City Council (2015) published a report on multiple deprivation in the city, where some of these issues (traffic accidents, congestion, air pollution) are mentioned. According to the report, the city has several deprivation hotspots

where problems are accentuated and 16% of its residents live in the most deprived areas of England.

Like many other cities in England, there is a significant difference between affluent and deprived areas in Bristol (Tallon 2007). As shown in Fig. 1, Bristol consists of 35 electoral Wards with wealthy areas located mostly in its north-west part of the city, in parts of the Henleaze and Redland wards. Deprived areas can be found in the eastern part of the city, in the wards of Easton and Lawrence Hill, and in the southern part, in the wards of Bishopsworth, Hartcliffe, Filwood, Knowle, and Whitchurch Park, and in the ward of Southmead in the northern part of the city.

Bristol was chosen as a case study because of an active use of social media platforms and rich history of official QoL surveys (Bristol City Council 2018) that offer possibility for comparison and further exploration.

Data description

The first type of data used are geo-located messages posted by Twitter users, collected from the Twitter social media platform called Tweets. Tweets are short, unstructured text messages consisting of maximum 140 characters written in different styles, slang, abbreviation, links, hashtags, and so forth. In Table 1 examples of the various types of Tweets are shown to illustrate their versatility and complexity.

Geo-tagged Tweets are messages containing location of the sender in the moment the message was posted online and these messages are the subject of this research. The Tweets used in this research were originally collected as part of the research at the University of Kentucky, in the Digital OnLine Life and You (DOLLY) project (Floating Sheep 2018), where DOLLY is an archive of billions of geo-tagged Tweets created for analysis and research in real time. The dataset used for this research consisted of geo-tagged Tweets collected from January 2012 to September 2016 in the area of the city of Bristol. Moreover, two additional datasets were used, scores from the QoL Bristol survey for 2013 and scores from the Index of Multiple Deprivation for 2015. Twitter data of 2013 have been chosen as they match the other two datasets and facilitate the comparison.

It is important to recognize some of the limitations of Twitter data. First, although the messages are geo-tagged, there is a risk of 'migration bias', since the

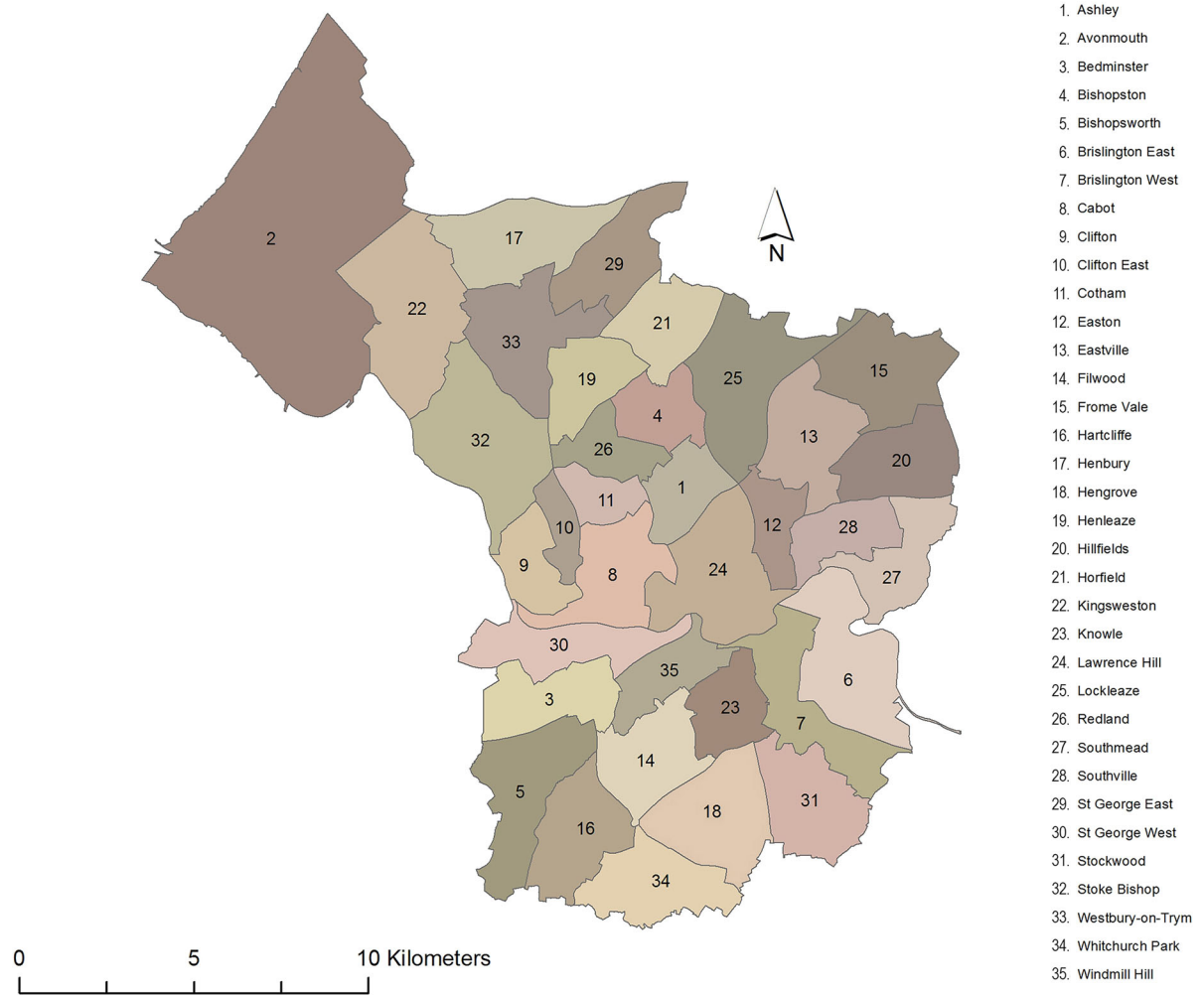


Fig. 1 Electoral wards in Bristol

Table 1 Examples of Tweets

Tweets

I think I've mistaken this whole situation and I feel like an idiot

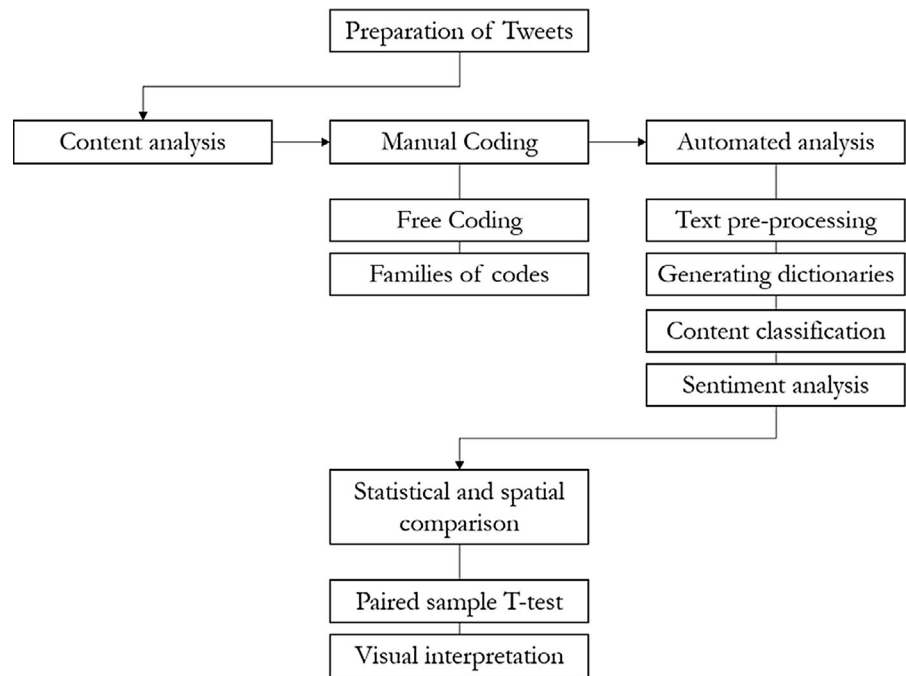
@username01 I bet the excitement was too much to handle haha

Why Labour won't talk about the economy: output across services sector rose at the strongest pace for 16 years between July-September #r4today

What a lovely way to start an Autumn day: <http://t.co/gSnU9XFuFt>

statement from the message about a specific location could be sent from a completely different location and different time. There is also a problem of representability, knowing that use of Twitter is very uneven (e.g. age of users, income of users, languages they use,

mobility of users, and access to mobile phones). Blank and Lutz (2017) investigated the representativeness of different social media platforms and found that Twitter users in Great Britain are significantly different from the total population in terms of age and

Fig. 2 Methodological framework

income (younger and wealthier) but not for education and gender.

Analysis of Twitter messages

Unlike conventional methods where capturing people's perceptions about observed phenomena is mostly theory driven, opinions derived from social media data require an approach that is more exploratory. It generates insights from the data, rather than theory. The steps of the analysis are shown in Fig. 2.

Preparation of Tweets

The dataset used contained a total of 4,437,900 Tweets. After clipping the data using the boundaries of the city of Bristol, the number of Tweets was reduced to 3,616,433. At this point of the analysis, the year 2013 was chosen to be further investigated because it coincided with the year in which the City of Bristol held its survey on QoL. Tweets for the year 2013 were aggregated into wards (administrative boundary) to see the spatial distribution of tweeting in the city of Bristol based on the total number of Tweets. The rest of the analysis is based on Tweets aggregated at ward level. Furthermore, the results were presented in boundaries that are meaningful for

policy makers and planners. In this case, the electoral wards are administrative boundaries used for policy makers to design interventions and target areas. Wards are also the boundary used by the Bristol City Council to report on QoL.

Content analysis

Twitter data were processed using a coding system and text analysis techniques where messages posted by the Twitter users were categorised based on the content. The approach was semi-manual and involved manual coding and automated analysis. The content analysis of the Tweets was done using Computer-Assisted Qualitative Data Analysis (CAQDAS) and Geographic Information System (GIS) software.³

For *manual coding*, the total number of Tweets (1,374,706) was used as a sampling frame to calculate a random sample for the area of Bristol, for the year 2013, where Tweets were normalised based on the population size. The size of the sample used was 1067 Tweets.

Free coding technique was used to recognise QoL perceptions, derive subjective QoL domains and generate a codebook for further analysis. Sixty-six

³ Atlas.ti and ArcGIS.

free codes were generated and a total number of 102 subjective QoL perceptions captured.

Families of codes were defined and served as points for grouping similar codes. They were structured based on previously reviewed domains from different studies done on subjective QoL in Bristol and in the United Kingdom, and from domains emerging from the data. Moreover, two additional human coders were involved for the purpose of quality control; triangulation and initial coding results were confirmed.

Transport and health domains emerged as the most predominant ones, while environment was added as environmental conditions play a relevant role when accessing the quality-of-life. Furthermore, selected domains are potentially informative for planners and policy makers.

Generating dictionaries

Automatic text retrieval operations require a thoughtful strategy, a coding scheme to follow. However, the content analysis allows a certain amount of creativity in defining these steps due to the specific requirements of the topic. Dictionaries are defined as a list of indicative words for a specific topic reflecting the relevant information generated based on previously defined domains. According to literature (Hsieh and Shannon 2005; Schwartz and Ungar 2015) it is essential to produce a good set of indicative words and their synonyms to guide the retrieval of messages. There are three ways to generate dictionaries: manual dictionaries, crowd-sourced dictionaries and dictionaries derived from the text. While manual dictionaries are widely used in the traditional content analysis, and crowd-sourced dictionaries are manual ones constructed on the opinions of the crowd, deriving dictionaries from text is an automated way to approach a large collection of text. Here, dictionaries were derived combining automated extraction and manual selection. First, the word frequencies were calculated for all Tweets from 2013 in an automated way using Excel. Afterward, words and phrases relevant to the topic were manually extracted from the frequency lists and assigned to the corresponding domain dictionary. As a result, dictionaries for three domains were constructed: health, transport, and environment. Every domain dictionary contained 25 indicative words.

Content classification

The classification of the content was systematically done ward by ward by classifying Tweets for each ward through the dictionary for every recognised domain. The result was a number of perceptions about subjective QoL in three analysed domains. Because the numbers itself do not say much and normalisation using population size assumes that all population tweet in the same rate, the normalisation was done using a slightly more refined calculation, calculating the odds ratio. Several authors addressed the issue of making a relevant spatial representation of patterns derived from Twitter as raw count and suggested the use of odds ratio (OR) normalisation (Zook and Poorthuis 2014; 2015). The advantages of using odds ratio are the opportunity to normalise our perceptions by any other variable and easy to understand results (Zook and Poorthuis 2015).

The normalisation was done by total tweeting population (the number of Tweets in 2013 for the city of Bristol is taken as a proxy for tweeting population). The formula used is:

$$OR = \frac{P_w/P_{tot}}{PopW/TwPop} \quad (1)$$

where P_w is the number of Tweets in a ward related to the domain observed (for example, the number of Tweets about health in one ward), P_{tot} is summary of all Tweets related to that domain in all wards (the city of Bristol), $PopW$ is the size of tweeting population in ward, and $TwPop$ is the total tweeting population for all wards (the city of Bristol).

In this case, odds ratio measures the number of Tweets containing QoL perception based on the total tweeting population.

Sentiment analysis

The final step of the content analysis was sentiment analysis of Tweets in different domains. Automated sentiment analysis was done using the Excel add-in MeaningCloud™ (<http://www.meaningcloud.com>) that offers different possibilities of analysing text. Automated sentiment analysis identified the positive/negative/neutral polarity in any text, including comments in surveys and social media. Automated sentiment analysis is based on differentiators: extracts

aspect-based sentiment, it discriminates opinions and facts, and detects polarity. Classified content is categorised based on the semantic scores of the perceptions within domains. The Tweets were classified into a five-point scale.

Next, positive and negative perceptions were counted and compared to check if they were statistically significantly different. Paired sample *t*-test was used to detect if there was a significant difference between two groups, positive and negative perceptions. The resulting positive and negative perceptions were visualised using ArcGIS to spatially show similarities and differences in perceptions between wards in Bristol.

Comparison between derived and measured subjective QoL

The final part of the analysis was a comparison between perceptions derived in present study and opinions of residents captured in the official QoL survey of Bristol, referring to these results as derived (from Tweet) and measured QoL (from survey). A comparison between the two was done statistically and spatially.

To test similarities between the Tweets results and the QoL survey, a null hypothesis was tested: the two variables derived from the two studies are the same, i.e. the results of the present study will reflect the results of the official QoL survey. For the purpose of this, a paired samples *t*-test was carried out in SPSS. Positive percentages of perceptions in analysed domains were used as variables derived in present study, and percentage of respondents satisfied with corresponding theme were used as variables from an official QoL survey in Bristol. Spatial comparison was done. Percentages of positive perceptions in health, transport and environment domain are overlaid with percentages of people satisfied in the corresponding topic using ArcGIS. Furthermore, the results were compared with Index of Multiple Deprivation (IMD), used as a measure of objective QoL.

Results

People using Twitter in the city of Bristol in the year 2013 have opinions on different topics that can be categorised in various QoL domains. Transport, health

and environment domains gave some relevant results and points to discuss (Table 2). Based on the highest percentage and versatility of the Tweets, transport is presented and discussed in detail.

From all of the geo-located Tweets sent from within the administrative boundaries of Bristol in 2013, the majority (50.42%) are perceptions about transport. There are various types of perceptions within the transport domain. The majority is about quality of public transport, buses, and bus stops (“as much as i love how cheap the mega bus to cardiff is why does it always have to be running late”; “lack of access to public transport is the single biggest barrier to youth accessing opportunities”). Additionally, people in Bristol give comments about parking places, conditions of streets, trains, and cycling (“park street looking gorgeous would love to be here in the winter to go sledging down it”).

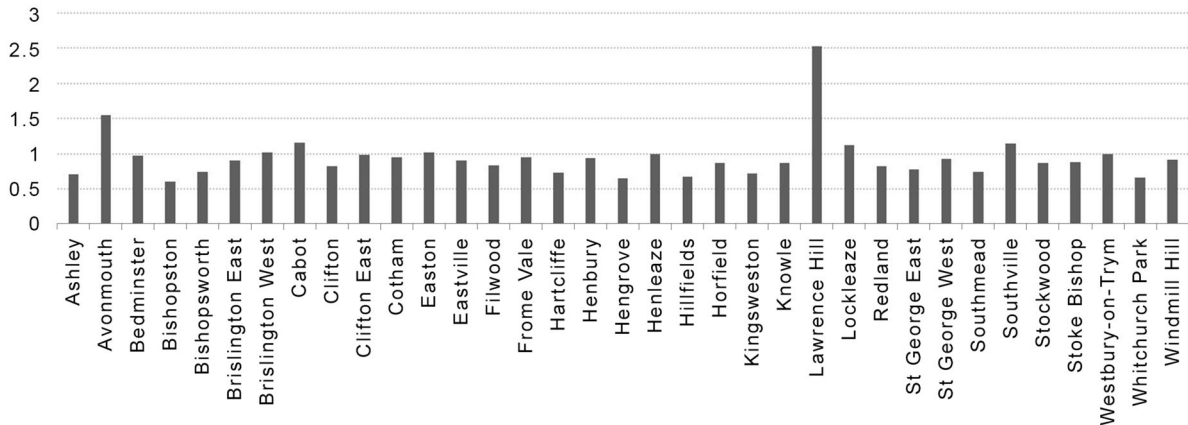
People are encouraged by the Bristol City Council to be engaged in the community development and voice their opinion through QoL surveys (Bristol City Council 2018). This could be reflected in a number of Tweets where people directly mention Bristol City Council Twitter account commenting on some of the burning issues regarding transport (“bristolcouncil no problem with riding on pavement at speed without consideration for other no”) Moreover, transport domain also has a certain amount of perceptions expressing emotional reaction, some form of distress or excitement while using public transport, biking, walking (“omg this bus stinks and i feel sick as it is”).

Content classification and odds ratio gave information about the spatial distribution of Tweets. Figure 3 shows odds ratio values for Bristol wards. In summary, people tweet as much as expected in more than half of the wards in Bristol, while there are several wards where tweeting activity is lower/higher than expected based on the total tweeting population.

The distribution of Tweets into sentiment categories gave us information about levels of satisfaction in Bristol wards. Subjective QoL perceptions about transport for the city of Bristol in 2013 are distributed in five sentiment groups: highly positive (P+), positive (P), neutral (NEUT), negative (N), and highly negative (N+). 60.57% of perceptions about transport were given sentiment in the analysis, while 39.43% are characterized as perceptions where the sentiment could not be categorized. Table 3 gives an example of Tweets distributed in five sentiment groups.

Table 2 Characteristics of tweets in Bristol (2013)

Tweets' characteristics	N	Percentage
Geolocated	1,374,706	
With QoL perceptions	61,970	4.51
With QoL perceptions about health	25,187	40.64
With QoL perceptions about transport	31,247	50.42
With QoL perceptions about environment	5536	8.93

**Fig. 3** Odds ratio values in transport domain in Bristol (2013)**Table 3** Examples of Tweets in transport domain distributed in sentiment groups

Sentiment group	Example of Tweets within sentiment groups
N+	“another big shout for stolenbikesbris because bike theft is such a major impediment to the development of mass cycling”
N	“i hate waiting for public transport”
Neutral	“not quite warm enough to cycle home in indoor clothes”
P	“im impressed the 40a bus is running on boxing day”
P+	“i love getting on to a warm bus”

Highly positive (P+), positive (P), neutral (NEUT), negative (N), and highly negative (N+)

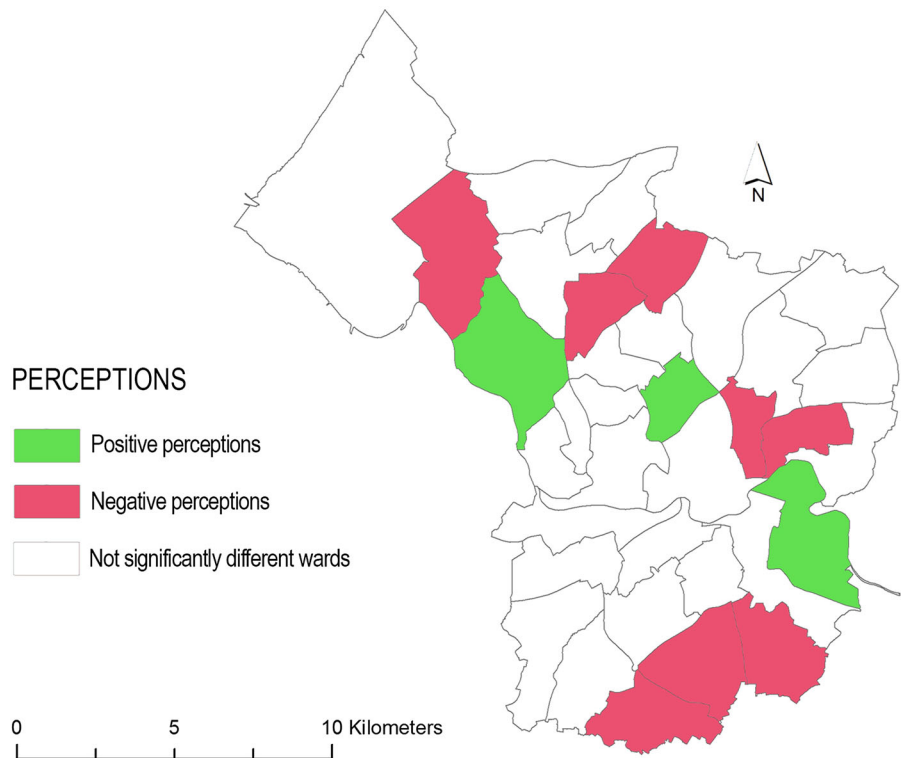
Statistically, there is no significant difference between positive and negative perceptions (at ward level), based on sentiment, with p values in transport domain $p > 0.05$. However, wards with highest positive and highest negative values are calculated and visualised for showing spatial distribution. These wards are observed as places where people have predominantly positive or negative perception, based on the perceptions captured from Twitter.

Spatial distribution of positive and negative perceptions about transport is visualised in Fig. 4. Eleven wards in transport domain have differences between

positive and negative perceptions, three with more positive, and eight with more negative perceptions. Considering the highest percentages of positive perceptions, transport conditions are the best in three wards, Stoke Bishop, Ashley, and Brislington East. Going north and south, the percentage of positive perceptions is decreasing.

The subjective perceptions about QoL derived from all geo-located Tweets sent from within the administrative boundaries of Bristol in 2013 are compared to results from the official QoL survey in Bristol in 2013. In the transport domain, based on the paired samples

Fig. 4 Spatial distribution of positive and negative perceptions in transport domain



t test (“Appendix”) the two results are significantly different ($p < 0.05$), and the variables are not significantly correlated.

Moreover, results from the present study compared to the Bristol Index of Multiple Deprivation (IMD) gave no significant statistical correlation. However, it is possible to observe positive and negative QoL perceptions in the local context and look for an explanation for the existence of certain perceptions. For this purpose, we used information about deprivation hotspots in Bristol and objective characteristics derived from the IMD (Fig. 5). The IMD map with scores for Bristol wards was overlaid with pie charts illustrating the percentages of positive, neutral and negative perceptions in transport domain. Positive and negative perceptions in transport domain have some similarities with the characteristics of wards based on the level of deprivation. First, there are three wards with positive perceptions, located in central, eastern and western part of the city and one in the ward with the lowest level of multiple deprivation. Wards with

highly negative perceptions match with wards with a higher level of deprivation.

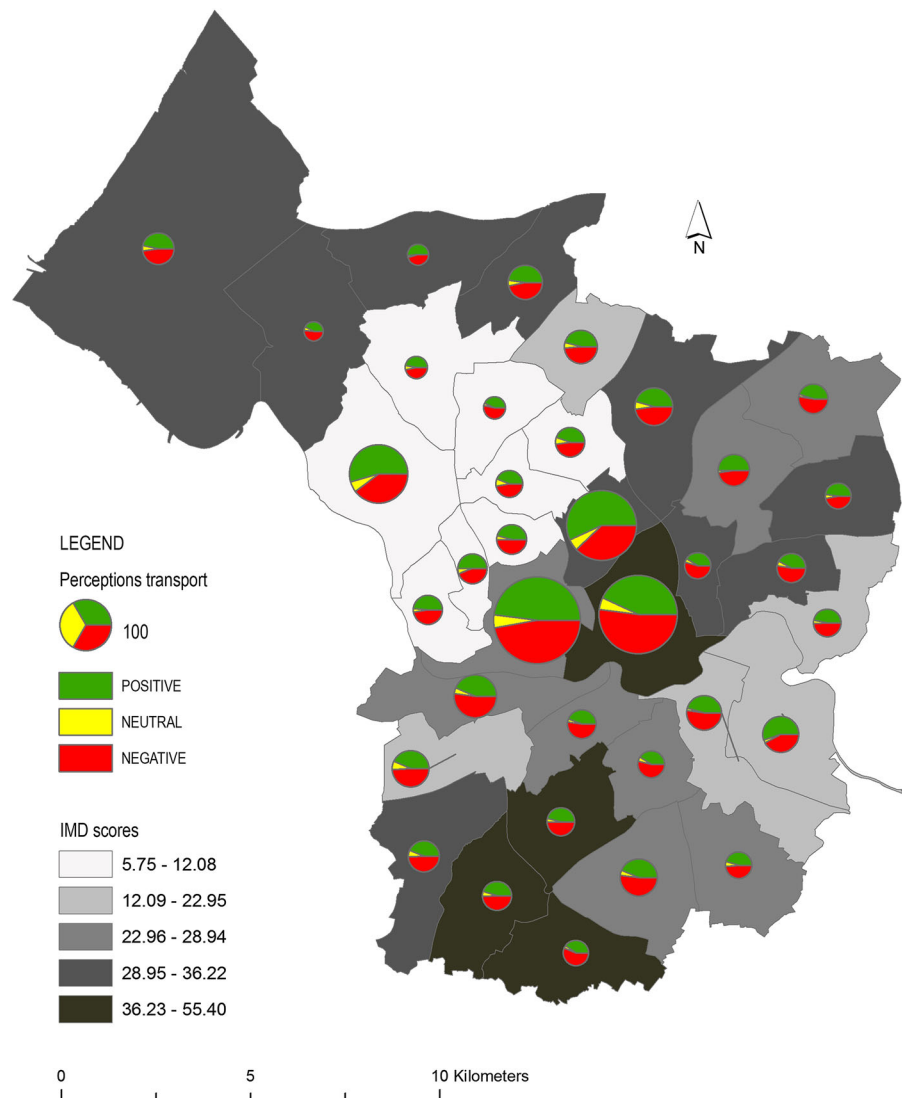
Discussion

Deriving subjective QoL domains using Twitter data

Social media have shown to be a relevant source of data, applicable in capturing subjective quality-of-life (QoL) perceptions. Qualitative analysis of a random sample of Tweets can successfully recognise people’s perceptions about QoL and derive domains that are suitable to measure with Twitter data. The benefit of including manual coding of a sample of Tweets is in having a more transparent approach, instead of capturing perceptions only through black-boxed automated classification. This part of the analysis gives an overall idea about the type of perceptions and domains that can be observed.

Findings from qualitative analysis offer a general idea about the nature of messages indicating perceptions about QoL. Possibilities to gain insights from the

Fig. 5 IMD overlaid with transport perceptions.
Source: own analysis based on English Index of Multiple Deprivation 2015 (IMD15)



data, and still strengthen the process by effective use of theoretical knowledge are shown. While Twitter messages reveal QoL perceptions, QoL theory helps in classifying these perceptions into domains. There is a line of similarity between summarised domains in subjective QoL research conducted in a more traditional way and domains derived from Twitter data in present study. Similarly to studies using traditional methods for collecting and analysing subjective QoL (for example Bramston et al. 2002; Eby et al. 2012; Ibrahim and Chung 2003), various domains of QoL are recognised.

Undoubtedly, most QoL perceptions derived from Twitter are subjective and personal. However, based

on obtained results, two types of perceptions can be distinguished:

- An *emotional* reaction where people express feelings. These perceptions are about how people feel within a certain domain and include Tweets where people express emotions like joy, happiness, excitement, and, on the opposite, feeling of dissatisfaction, sadness, and so forth.
- *Cognitive* conclusions where people express opinions. These perceptions are about how people feel about the observed topic and include Tweets where they express opinions about specific topic observed in their surroundings.

Emotions and feeling captured from social media are analysed vastly in various fields of study (psychology, health science, linguistic, happiness studies). However, the recognition of the second type of perceptions (cognitive) is valuable, pointing to a possibility for urban planners and decision makers to include the opinions of individuals derived from Twitter in recognising primary areas for specific policies and interventions. For example, people repeatedly pointing to a specific problem in the same part of the city.

People's perceptions about QoL in Bristol

The first significant finding is the fact that, when observing spatial distribution of Tweets per tweeting population, the ward in Bristol with the highest value, where every 12th Tweet indicates a clear QoL perception, is ward Lawrence Hill. This is also one of the most deprived wards in Bristol, and part of the ward called Old Market and The Dings is in the 10% of the most deprived wards in England (Bristol City Council 2015). Moreover, when looking at variations between perceptions, considerable difference in types of perceptions can be seen. Due to this, perceptions can be classified into subtypes, based on the main topics they cover. At least three subtypes are captured: quality of public transport, quality of streets, and opinions about cycling.

Spatial distribution of a number of perceptions gives a general idea about differences between Bristol wards in the sense of the quantity of perceptions and location with more frequent tweeting activity. Nevertheless, it is not informative enough to get a proper understanding of the level of satisfaction. Therefore, this study has taken a step in the direction of analysing the sentiment of captured subjective QoL perceptions to compare the wards according to the level of satisfaction. One of the most interesting findings is that the Tweets in this study are similarly positive and negative in sentiment and it is necessary to address both to get a better understanding of the level of satisfaction in Bristol wards. This is further explored by examining and interpreting their spatial distribution. It was found that there is a greater presence of wards with highly negative perceptions.

In general, the southern part of the city of Bristol is characterised as an area with higher level of deprivation. Additionally, there are wards in the city of Bristol

where positive and negative perceptions derived from Twitter converge with low and high levels of deprivation, based on the IMD. These kinds of contrasting measurements are often in QoL research, when trying to compare subjective perceptions with objective conditions. In cases where IMD is taken as an objective QoL measure the Tweets may converge or diverge with the relative measure of deprivation.

The tool used for sentiment classification gives us information about the number of Tweets in each of five sentiment groups and the possibility to capture differences between levels of satisfaction within observed domains and spatial distribution of positive and negative sentiment. Moreover, as noticed by Nguyen et al. (2016), only several studies addressed the issue of developing sentiment classification in domains of food and physical activity using social media. Similarly, not much has been done in developing sentiment classifiers useful for QoL research using Twitter data, which justifies our selection of the method used.

Reflection on comparison between derived and measured subjective QoL

It is relevant to recognise the possibilities of combining approaches in assessing subjective QoL to improve planning and decision-making process. Results derived in the present study are compared to the results derived from an official QoL survey done in Bristol in 2013. Statistically and spatially, we found no correlation between results derived in two studies.

Next to the spatial and statistical comparison, there is one more setting where the complementarity of Twitter data can be observed. It includes coverage of questions asked in the survey and types of perceptions captured from Twitter. For example, according to the QoL survey report, responses about transport mostly address satisfaction with information about public transport, the cost of public transport and satisfaction with bus lanes and bus stops. Perceptions derived from Twitter cover similar topics; however, they are mostly oriented to quality and condition of buses, bus frequencies, congestion, and how people feel inside the bus. This finding is consistent with previous studies on transport and well-being (e.g. Friman et al. 2017) where they demonstrate that satisfaction with travel is related to positive and negative emotional responses to critical incidents.

Moreover, perceptions from Twitter cover a wider range of topics, compared to the QoL survey used for

the comparison. While here the variety of topics is recognised, from personal feelings in the bus and at bus station, to opinions in different segments of transport in general, proxy used for comparison with official QoL survey is percentage of respondents satisfied with bus services.

Furthermore, differences between the derived QoL from Twitter and the QoL survey can be explained by the profile of respondents and age in particular. According to the Bristol QoL survey report (Bristol City Council 2014), proportionally less young people responded in the QoL survey. 59.3% of respondents was in the age group 50 years and older, where the highest response rate was in the age group 60–64. Conversely, 40.7% of respondents were from the age group 18–49, with the smallest response rate in the age group 18–24. Looking into Twitter demographics, younger population tend to use social media more. In the United Kingdom, in 2013, about two third of Twitter users were under the age of 34, with the highest percentage (47%) of users in the age group 18–24 (Statista Inc. 2017). However, studies show that, although the use of Twitter stays the highest in this age group, in the last decade, increase in the number of users is the highest in the 25–45 year-old age group (Ciuccarelli et al. 2014a, b). This difference in age of QoL survey respondents and Twitter users strengthen the suggestion of using data from social media as complementary data when evaluating QoL.

An idea we would like to address here is introduced by Goodchild (2007) and his analysis of Volunteer Geographic Information (VGI). He offers an interpretation of VGI serving as a way of producing information by employing people to act as sensors, capturing the change in the living environment and uploading it to the online world in appropriate form. Even though we captured only a few similarities between the derived QoL from Twitter and the official QoL survey, this lack of correlation between results can also be interpreted as the result or generation of new or complementary knowledge.

In summary, several main similarities and differences in compared approaches are underlined. The main differences are in the size of the sample and methodology used for the analysis. The official QoL survey in Bristol is based on a smaller sample, while the Twitter dataset we used covers a larger population. Moreover, in this study insights are obtained from the data itself, rather than theory or policy frameworks, as

it is done in more traditional approaches such as the QoL survey done in Bristol. Moreover, the official QoL survey in Bristol is done per ward, where households are interviewed, so we know for sure that the location of the QoL perception corresponds with the location where people live (no migration bias). With Twitter data, the location problem is much more emphasised. According to Li et al. (2013) geotags on certain Tweets point to the mere presence of Twitter users in these sites. Moreover, the authors distinguish three types of locations: residence, work, and tourist attractions. It is hard to check which location was used by the user at the moment of sending a message.

Reflection on usability of social media in QoL research

Compared with traditional methods for analysing subjective QoL, harvesting and evaluating data from social media offers a contemporary, fast and cost effective approach (Schnitzler et al. 2016).

Contemporary urban planning practice is embracing the positive characteristics of social media data, and this study is a contribution towards a better understanding of connections between location, people, and messages shared in online settings. In general, involvement of the community can be observed as a collaborative way of producing knowledge, facilitating participatory planning practice and joint decision making (Natarajan 2015). Using the city of Bristol exemplifies this claim. The City Council offers the opportunity to jointly make decisions and take actions based on those decisions together. Likewise, social media data offer a novel and unobtrusive way of capturing people's perceptions for evaluating characteristics of the neighbourhoods and communities.

Urban planning is traditionally placed in an offline setting. We experience the city as a system made of physical urban form and various functions. Social media offers insight into people's perceptions about a system and possibility to capture general ideas about the functioning of this system. Availability and spatiality are key features of Twitter messages. The connection between the physical and digital world is reflected through the spatiality of data and the existence of opinions. When the opportunity to give comments about something exists, people tend to use it, and that is linked to a particular location and stays kept in an online database. However, looking at this study, we have to

bear in mind that, even though the Tweets are geo-tagged and connected with a specific point in space, it does not mean that an opinion expressed is about that location. People can comment about public transport after they leave the bus, or hospital service when they are back home. Nguyen et al. (2016) address this as “migration bias” and therefore something that can reduce the strength of collected opinions.

Furthermore, Ballas (2013) recognised the value of subjective QoL studies in providing the insight for cities and regions and helped in creating policies and investments to improve life of their citizens. Correspondingly, Kitchin (2014) provided strong arguments supporting the role of big data in producing knowledge for shaping better cities. The emphasis is on an essential characteristic, the flexibility of data and diversity in use. This flexibility is reflected in the present study with producing meaningful output by adapting a set of different techniques for the desired purposes and producing new knowledge that can serve as an input for improvement of cities.

Many studies in different fields of science gave insight about social media data and methods for analysis, where some were focused on language characteristics (Agarwal et al. 2011), others on developing perfect algorithms (Waykar et al. 2016). The advantage of this research is the attempt to combine different techniques adapted for simple extraction of QoL opinions from Twitter data, and exploring how results of such study could be efficiently placed in a planning context and potentially used to improve the decision-making process and enhance quality-of-life of residents.

For this study ward level was a relevant unit of analysis as the Tweets were compared with the existing QoL survey. However, in future research Tweets could be aggregated at smaller areas such as LSOAs.⁴ Moreover, tweets could be analysed over time and capture to what extent persons change perceptions over time.

Limitations

Using social media data in scientific research can be challenging. In this research, simple text classification

is used, avoiding machine learning and advanced natural language processing algorithms, which could be useful as it provides insight for an urban planner or social scientist unfamiliar with those methods. There are possibilities to classify text in more sophisticated ways using n -gram tokenization or specifically designed topic modelling (Bird et al. 2009).

Messages posted on social media represent a biased sample. People using Twitter are not a representative sample of the population. Internet usage is very uneven among countries, within countries, and within cities, with underrepresented groups, such as children and elderly (Warf 2013). In some countries, gender is also relevant, and income plays an important role as well (Blank and Lutz 2017). Furthermore, some “power users” (Shelton et al. 2015, 202) may post a disproportionately large amount of tweets. In this study, considering that only a small percentage of users posted several Tweets (but not more than ten) we assume that their effect is negligible. Nevertheless, for further studies where Tweets are considered for QoL the percentage of power users and their amount of tweets should be considered outliers and removed from the dataset.

Although the Tweets used are geo-tagged, the migration bias is emphasised. It is known that a person sending a message is present at a certain location. However, it is still unknown what kind of function that location has (e.g. residence, work, leisure, travel). People can comment about a certain thing, issue or location characteristic while being in a different location.

Conclusion

The main objective of the present study was to examine the possibility of extracting people’s perceptions about subjective QoL from Twitter and determine whether Twitter data can be used as proxies for QoL survey data. We chose a case study in order to place the results in a local context where the use of QoL perceptions derived from Twitter data could be meaningful and compared to existing measures used by policy makers.

A methodological approach was designed and steps were proposed for analysing data derived from Twitter for the purpose of assessing QoL, using the city of Bristol as the case study area. This study shows the

⁴ Lower-layer Super Output Area (LSOA) level is small area unit created to represent areas of approximately same population size, with an average of around 1500 persons.

relevance of using a mixed method approach, with qualitative analysis (e.g. text analysis) generating input for quantitative analysis, and together generating meaningful results. The qualitative part revealed the variety of QoL domains that can be observed. As a result, health, transport and environment domains were chosen to be further analysed. The quantitative part classified Tweets into selected domains, capturing the amount of perceptions within observed domain and showing the differences between Bristol wards.

Three main conclusions are underlined. The first one is that Twitter data can be used to evaluate QoL of residents. The second one is that, based on people's perceptions, there is a spatial variation in QoL between Bristol wards. There is a difference between wards as their residents have diverse positive/negative QoL perceptions. The third one is that, while Twitter messages can be used to complement QoL surveys, they cannot be used as proxies or replace other QoL measurement tools. QoL derived from Twitter data could be used for triangulation or completeness of other QoL data. Twitter messages may be useful to indicate the emergence of concerns not identified by traditional QoL surveys but Twitter data limitations (e.g. migration and demographic bias) may render invisible certain segments of the population.

Urban planning observes the city as a complex combination of physical urban form and various functions traditionally placed in offline setting. Social media offers a possibility to capture people's ideas

about that system and its specific parts. In general, the findings of the present study reveal the importance of studying people's perceptions that can be easily elicited from social media. Also, the results, findings, and approaches used in the present study can be useful in designing future studies on subjective QoL using Twitter data, especially for urban planners and social scientists.

Acknowledgements This work was partly supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A3A2924563). Tweets dataset was provided by Dr. Ate Poorthuis, collected through the Dolly project (University of Kentucky) and the Floating Sheep.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest and comply with ethical standards.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

This appendix provides the paired samples *t* test results (Tables 4, 5, 6).

Table 4 Paired samples statistics for transport positive tweets and % respondents satisfied

	Mean	N	SD	SE Mean
Pair 1				
Transport positive tweets	29.0119	35	4.39788	.74338
% respondents satisfied	53.060	35	8.9667	1.5157

Table 5 Paired samples *t* test for transport positive tweets and % respondents satisfied

	Paired differences				t	df	Sig. (2-tailed)	
	Mean	Std. deviation	Std. error mean	95% confidence interval of the difference				
				Lower				Upper
Pair 1								
Transport positive tweets—% respondents satisfied	− 24.04809	10.85037	1.83405	− 27.77532	− 20.32086	− 13.112	34	.000

Table 6 Paired samples correlations for transport positive tweets and % respondents satisfied

	N	Correlation	Sig.
Pair 1			
Transport positive tweets & % respondents satisfied	35	– .228	.188

References

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. *Association for Computational Linguistics* (pp. 30–38). <http://dl.acm.org/citation.cfm?id=2021109.2021114%5Cnpapers3://publication/uuid/83CA53FE-43D1-4BD5-BCF2-D55B82CF0F99>. Accessed 1 October 2016.
- Aladwani, A. M. (2015). Facilitators, characteristics, and impacts of Twitter use: Theoretical analysis and empirical illustration. *International Journal of Information Management*, 35(1), 15–25. <https://doi.org/10.1016/j.ijinfomgt.2014.09.003>.
- Almazidy, A., Althani, H., & Mohammed, M. (2016). Towards a disease outbreak notification framework using Twitter minning for smart home dashboards. *Procedia Computer Science*, 82, 132–134. <https://doi.org/10.1016/j.procs.2016.04.019>.
- Arribas-Bel, D., Kourtit, K., Nijkamp, P., & Steenbruggen, J. (2015). Cyber cities: Social media as a tool for understanding cities. *Applied Spatial Analysis and Policy*, 8(3), 231–247. <https://doi.org/10.1007/s12061-015-9154-2>.
- Ballas, D. (2013). What makes a ‘happy city’??. *Cities*, 32(1), S39–S50. <https://doi.org/10.1016/j.cities.2013.04.009>.
- Berhe, R. T., Martinez, J., & Verplanke, J. (2014). Adaptation and dissonance in quality of life: A case study in Mekelle, Ethiopia. *Social Indicators Research*, 118(2), 535–554. <https://doi.org/10.1007/s11205-013-0448-y>.
- Bibo, H., Lin, L., Rui, G., Ang, L., & Tingshao, Z. (2014). Sensing subjective well-being from social media. In D. Ślęzak, G. Schaefer, So T Vuong, & K. Yoo-Sung (Eds.), *Active media technology* (Vol. 8610, pp. 324–335). Warsaw: Springer. https://doi.org/10.1007/978-3-319-09912-5_27.
- Bird, S., Klein, E., & Loper, E. (2009). In J. Steele (Ed.), *Natural language processing with python* (1st ed.). Sebastopol: O’Reilly Media, Inc. <https://doi.org/10.1097/00004770-200204000-00018>.
- Blank, G., & Lutz, C. (2017). Representativeness of social media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, 61(7), 741–756. <https://doi.org/10.1177/0002764217717559>.
- Bonn, G., & Tafarodi, R. W. (2013). Visualizing the good life: A cross-cultural analysis. *Journal of Happiness Studies*, 14(6), 1839–1856. <https://doi.org/10.1007/s10902-012-9412-9>.
- Bramston, P., Pretty, G., & Chipuer, H. (2002). Unravelling subjective quality of life: An investigation of individual and community determinants. *Social Indicators Research*, 59(3), 261–274. <https://doi.org/10.1023/A:1019617921082>.
- Bristol City Council. (2014). *Quality of life in Bristol: Survey results 2013*, 82. http://www.bristol.gov.uk/sites/default/files/documents/council_and_democracy/consultations/qol2014final.pdf. Accessed 10 October 2016.
- Bristol City Council. (2015). *Deprivation in Bristol 2015*. Bristol. <https://www.bristol.gov.uk/documents/20182/32951/Deprivation+in+Bristol+2015/429b2004-eeff-44c5-8044-9e7dcd002faf>. Accessed 10 October 2016.
- Bristol City Council. (2018). The quality of life in Bristol—bristol.gov.uk. <https://www.bristol.gov.uk/statistics-census-information/the-quality-of-life-in-bristol>. Accessed March 16, 2018.
- Carlquist, E., Ulleberg, P., Delle Fave, A., Nafstad, H. E., & Blakar, R. M. (2016). Everyday understandings of happiness, good life, and satisfaction: Three different facets of well-being. *Applied Research in Quality of Life*. <https://doi.org/10.1007/s11482-016-9472-9>.
- Chatfield, A. T., Scholl, H. J., & Brajawidagda, U. (2013). Tsunami early warnings via Twitter in government: Net-savvy citizens’ co-production of time-critical public information services. *Government Information Quarterly*, 30(4), 377–386. <https://doi.org/10.1016/j.giq.2013.05.021>.
- Chen, X., & Yang, X. (2014). Does food environment influence food choices? A geographical analysis through “tweets”. *Applied Geography*, 51, 82–89. <https://doi.org/10.1016/j.apgeog.2014.04.003>.
- Ciuccarelli, P., Lupi, G., & Simeone, L. (2014a). Reflections on potentialities and shortcomings of geo-located social media analysis. In B. Pernici, S. Della Torre, B. M. Colosimo, T. Faravelli, R. Paolucci, & S. Piardi (Eds.), *Visualizing the data city* (1st ed., pp. 55–61). Milano: Springer. <https://doi.org/10.1007/978-3-319-02195-9>.
- Ciuccarelli, P., Lupi, G., & Simeone, L. (2014b). In B. Pernici, S. Della Torre, B. M. Colosimo, T. Faravelli, R. Paolucci, & S. Piardi (Eds.), *Visualizing the data city* (1st ed.). Milan: Springer. <https://doi.org/10.1007/978-3-319-02195-9>.
- Conole, G., Galley, R., & Culver, J. (2011). Frameworks for understanding the nature of interactions, networking, and community in a social networking site for academic practice. *International Review of Research in Open and Distance Learning*, 12(3), 119–138. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>.
- Costanza, R., Fisher, B., Ali, S., Beer, C., Bond, L., Boumans, R., et al. (2007). Quality of life: An approach integrating opportunities, human needs, and subjective well-being. *Ecological Economics*, 61(2–3), 267–276. <https://doi.org/10.1016/j.ecolecon.2006.02.023>.
- Curini, L., Iacus, S., & Canova, L. (2015). Measuring idiosyncratic happiness through the analysis of Twitter: An application to the Italian case. *Social Indicators Research*,

- 121(2), 525–542. <https://doi.org/10.1007/s11205-014-0646-2>.
- Davern, M. T., & Chen, X. (2010). Piloting the geographic information system (GIS) methodology as an analytic tool for subjective wellbeing research. *Applied Research in Quality of Life*, 5(2), 105–119. <https://doi.org/10.1007/s11482-010-9095-5>.
- Diener, E. (2000). Subjective well-being. The science of happiness and a proposal for a national index. *The American Psychologist*, 55(1), 34–43. <https://doi.org/10.1037/0003-066x.55.1.34>.
- Dluhy, M., & Swartz, N. (2006). Connecting knowledge and policy: The promise of community indicators in the United States. *Social Indicators Research*, 79(1), 1–23. <https://doi.org/10.1007/s11205-005-3486-2>.
- Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, 6(12), 1–26. <https://doi.org/10.1371/journal.pone.0026752>.
- Eby, J., Kitchen, P., & Williams, A. (2012). Perceptions of quality life in Hamilton's neighbourhood hubs: A qualitative analysis. *Social Indicators Research*, 108(2), 299–315. <https://doi.org/10.1007/s11205-012-0067-z>.
- Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers*. <https://doi.org/10.1080/00045608.2011.595657>.
- Floating Sheep. (2018). *DOLLY*. <http://www.floatingssheep.org/p/dolly.html>. Accessed March 16, 2018.
- Friman, M., Olsson, L. E., Ståhl, M., Ettema, D., & Gärling, T. (2017). Travel and residual emotional well-being. *Transportation Research Part F: Traffic Psychology and Behaviour*, 49, 159–176. <https://doi.org/10.1016/j.trf.2017.06.015>.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>.
- Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261. <https://doi.org/10.1177/2043820613513121>.
- Haas, B. K. (1999). A multidisciplinary concept analysis of quality of life. *Western Journal of Nursing Research*, 21(6), 728–742. <https://doi.org/10.1177/01939459922044153>.
- Harvey, F. (2013). To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4587-2_3.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>.
- Ibrahim, M. F., & Chung, S. W. (2003). Quality of life of residents living near industrial estates in Singapore. *Social Indicators Research*, 61(2), 203–225. <https://doi.org/10.1023/A:1021305620042>.
- Kapteyn, A., Lee, J., Tassot, C., Vonkova, H., & Zamarro, G. (2015). Dimensions of subjective well-being. *Social Indicators Research*, 123(3), 625–660. <https://doi.org/10.1007/s11205-014-0753-0>.
- Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1–14. <https://doi.org/10.1007/s10708-013-9516-8>.
- Kusumo, A. N. L., Reckien, D., & Verplanke, J. (2017). Utilising volunteered geographic information to assess resident's flood evacuation shelters. Case study: Jakarta. *Applied Geography*, 88, 174–185. <https://doi.org/10.1016/J.APGEOG.2017.07.002>.
- Larsson, J., Söderlind, A., Kim, H., Klaesson, J., & Palmberg, J. (2016). In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, & R. Purves (Eds.), *European handbook of crowdsourced information*. London: Ubiquity Press Ltd. <https://doi.org/10.5334/bax>.
- Leetaru, K., Wang, S., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*. <https://doi.org/10.5210/fm.v18i5.4366>.
- Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, 40(2), 61–77. <https://doi.org/10.1080/15230406.2013.777139>.
- Marans, R. W. (2003). Understanding environmental quality through quality of life studies: The 2001 DAS and its use of subjective and objective indicators. *Landscape and Urban Planning*, 65(1–2), 73–83. [https://doi.org/10.1016/S0169-2046\(02\)00239-6](https://doi.org/10.1016/S0169-2046(02)00239-6).
- Marans, R. W. (2015). Quality of urban life and environmental sustainability studies: Future linkage opportunities. *Habitat International*, 45(P1), 47–52. <https://doi.org/10.1016/j.habitatint.2014.06.019>.
- McCrea, R., Marans, R., Stimson, R., & Western, J. (2011). In R. W. Marans & R. J. Stimson (Eds.), *Investigating quality of urban life* (Vol. 45). London: Springer. https://doi.org/10.1007/978-94-007-1742-8_3.
- McMahon, S. K. (2002). The development of quality of life indicators—A case study from the City of Bristol, UK. *Ecological Indicators*, 2(1), 177–185. [https://doi.org/10.1016/S1470-160X\(02\)00039-0](https://doi.org/10.1016/S1470-160X(02)00039-0).
- Mohit, M. A. (2013). Quality of life in natural and built environment—An introductory analysis. *Procedia—Social and Behavioral Sciences*, 101, 33–43. <https://doi.org/10.1016/j.sbspro.2013.07.176>.
- Moro, M., Brereton, F., Ferreira, S., & Clinch, J. P. (2008). Ranking quality of life using subjective well-being data. *Ecological Economics*, 65(3), 448–460. <https://doi.org/10.1016/j.ecolecon.2008.01.003>.
- Natarajan, L. (2015). Socio-spatial learning: A case study of community knowledge in participatory spatial planning. *Progress in Planning*, 111, 1–23. <https://doi.org/10.1016/j.progress.2015.06.002>.
- Nguyen, Q. C., Kath, S., Meng, H.-W., Li, D., Smith, K. R., VanDerslice, J. A., et al. (2016). Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Applied Geography*, 73, 77–88. <https://doi.org/10.1016/j.apgeog.2016.06.003>.

- Pacione, M. (2003a). Quality-of-life research in urban geography. *Urban Geography*, 24(4), 314–339. <https://doi.org/10.2747/0272-3638.24.4.314>.
- Pacione, M. (2003b). Urban environmental quality and human wellbeing—A social geographical perspective. *Landscape and Urban Planning*, 65(1–2), 19–30. [https://doi.org/10.1016/S0169-2046\(02\)00234-7](https://doi.org/10.1016/S0169-2046(02)00234-7).
- Santos, L. D., Martins, I., & Brito, P. (2007). Measuring subjective quality of life: A survey to Porto's residents. *Applied Research in Quality of Life*, 2(1), 51–64. <https://doi.org/10.1007/s11482-007-9029-z>.
- Schnitzler, K., Davies, N., Ross, F., & Harris, R. (2016). Using Twitter™ to drive research impact: A discussion of strategies, opportunities and challenges. *International Journal of Nursing Studies*, 59, 15–26. <https://doi.org/10.1016/j.ijnurstu.2016.02.004>.
- Schuessler, K. F., & Fisher, G. A. (1985). Quality of life research and sociology. *Annual Review of Sociology*, 11, 129–149. <http://www.jstor.org/stable/2083289>. Accessed 1 October 2016.
- Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social media: A systematic overview of automated methods. *The Annals of the American Academy of Political and Social Science*, 659(1), 78–94. <https://doi.org/10.1177/0002716215569197>.
- Shelton, T. (2016). Spatialities of data: mapping social media 'beyond the geotag'. *GeoJournal*. <https://doi.org/10.1007/s10708-016-9713-3>.
- Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142, 198–211. <https://doi.org/10.1016/j.landurbplan.2015.02.020>.
- Sirgy, J. M. (2011). Theoretical perspectives guiding QOL indicator projects. *Social Indicators Research*, 103(1), 1–22. <https://doi.org/10.1007/s11205-010-9692-6>.
- Sloan, L., & Morgan, J. (2015). Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PLoS One*. <https://doi.org/10.1371/journal.pone.0142209>.
- Statista Inc. (2017). *Statista*. <https://www.statista.com/statistics/257429/share-of-uk-internet-users-who-use-twitter-by-age-group/>. Accessed January 12, 2017
- Tallon, A. R. (2007). Bristol. *Cities*, 24(1), 74–88. <https://doi.org/10.1016/j.cities.2006.10.004>.
- Tartaglia, S. (2013). Different predictors of quality of life in urban environment. *Social Indicators Research*, 113(3), 1045–1053. <https://doi.org/10.1007/s11205-012-0126-5>.
- Tesfazghi, E. S., Martinez, J. A., & Verplanke, J. J. (2010). Variability of quality of life at small scales: Addis Ababa, Kirkos sub-city. *Social Indicators Research*, 98(1), 73–88. <https://doi.org/10.1007/s11205-009-9518-6>.
- Warf, B. (2013). *Global geographies of the internet*. Dordrecht: Springer. <https://doi.org/10.1007/978-94-007-1245-4>.
- Waykar, P., Wadhvani, K., & More, P. (2016). Sentiment analysis in Twitter using natural language processing (NLP) and classification algorithm. *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*, 5(1), 79–81.
- Wills-Herrera, E., Islam, G., & Hamilton, M. (2009). Subjective well-being in cities: A multidimensional concept of individual, social and cultural variable. *Applied Research in Quality of Life*, 4(2), 201–221. <https://doi.org/10.1007/s11482-009-9072-z>.
- Yang, C., Raskin, R., Goodchild, M., & Gahegan, M. (2010). Geospatial Cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems*, 34(4), 264–277. <https://doi.org/10.1016/j.compenvurbysys.2010.04.001>.
- Zook, M., & Poorthuis, A. (2014). Offline brews and online views: Exploring the geography of beer tweets. In M. Patterson & N. Hoalst-Pullen (Eds.), *The geography of beer regions, environment, and societies* (pp. 201–209). Dordrecht: Springer. <https://doi.org/10.1007/978-94-007-7787-3>.
- Zook, M., & Poorthuis, A. (2015). Small stories in big data: Gaining insights from large spatial point pattern datasets. *Cityscape: A Journal of Policy Development and Research*, 17(1), 151–160.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.