CAPTURING AND MAPPING QOL USING TWITTER DATA

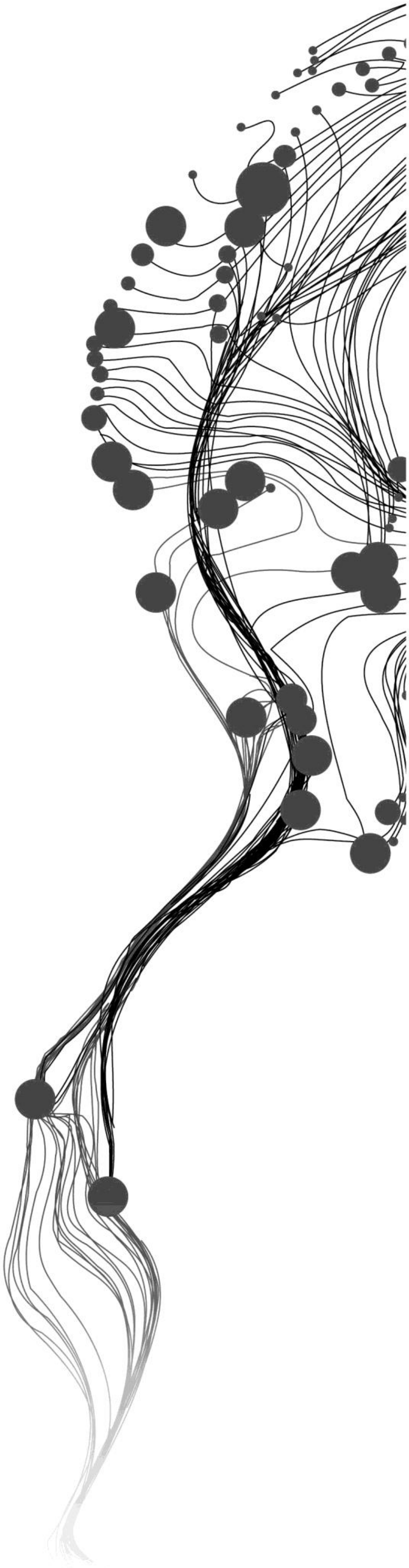
SLAVICA ZIVANOVIC

February, 2017

SUPERVISORS:

Dr. J.A. Martinez

Drs. J.J. Verplanke



CAPTURING AND MAPPING QOL USING TWITTER DATA

SLAVICA ZIVANOVIC

Enschede, The Netherlands, February, 2017

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Urban Planning and Management

SUPERVISORS:

Dr. J.A. Martinez

Drs. J.J. Verplanke

THESIS ASSESSMENT BOARD:

Prof.dr.ir. M.F.A.M. van Maarseveen (Chair)

Dr.ir. D. Hiemstra (External Examiner, University of Twente)

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 QoL. This study explores the use of social media in QoL research by capturing and analysing people's perceptions about their QoL using Twitter messages. The methodology is based on a mixed method approach, combining manual coding of the messages, automated classification, and spatial analysis. The city of Bristol is used as a case study, with a dataset containing 1,374,706 geotagged Tweets sent within the city boundaries in 2013. Based on the manual coding results, health, transport, and environment domains were selected to be further 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, Twitter data can be used to evaluate QoL. Second, based on people's opinions, there is a difference in QoL between Bristol neighbourhoods. And, third, Twitter messages can be used to complement QoL surveys but not as a proxy. 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.

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my supervisors, Dr. Javier Martinez and Drs. Jeroen Verplanke for their support, patience and constructive criticism. They allowed me to pursue my ideas and helped me find the right direction.

Special thanks to Dr. Ate Poorthuis, Dolly project and the Floating Sheep for providing me the necessary data, which helped a lot in shaping my research. #thankyou

I also want to acknowledge the European Commission and the Erasmus Mundus Sigma scholarship programme for giving me the opportunity to study at ITC, the Netherlands.

Thanks to my amazing friends here for nights and days of working and laughing together. And to the ones back home for encouraging me to follow my dreams.

I am forever grateful to my family for all the love and for always being there.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

IMD	Index of Multiple Deprivation
QOL	Quality of Life
VGI	Volunteer Geographic Information

1. INTRODUCTION

This section provides background information on quality of life research and connection between people and their living environment. Moreover, it covers main problems in today's research in the field of quality of life and gives justification for the study. Social media as a new source of data is also introduced, and possibilities are presented. Next, based on recognised knowledge gap, the research problem is identified. Furthermore, introduction part includes objectives of the research, general and specific, followed by research questions, and conceptual framework that serves as a guideline for this study.

1.1. Background on quality of life research and possibilities of social media as new data source

The condition of the living environment plays a major role in mental, physical, and social life of people. Most of the world's population live in rapidly growing cities where impacts of urbanisation have altered the conditions of the urban environment. Neighbourhoods are dynamic entities that are constantly evolving and are subjected to change and influences, both positive and negative. Objective quality of life conditions, including demographic, social and economic characteristics differ within urban areas, and disparities are becoming more visible. People living in different parts of the same city have different experiences and feelings of satisfaction with their neighbourhoods and objective conditions. Thus, it is necessary to regularly examine the relationship between people and the living surrounding to identify and measure differences in quality of life of the population (Pacione, 2003a).

Quality of life (QoL) is a multidisciplinary concept used by many researchers (Costanza et al., 2007; Haas, 1999; Pacione, 2003). Growing concern for differences within cities resulted in increased number of studies focused on community quality of life and well-being of the population. Quality of life is commonly defined as general satisfaction and well-being of individuals and communities in a specific surrounding across different domains. Policy makers, urban planners, and other researchers use results derived from the quality of life studies to address inequalities in a city, better understand issues, determine priority areas for intervention and allocate resources accordingly.

The non-existence of a unified methodology is one of the main issues in QoL research. For this reason, scientists are mainly focused on determining the effective way for defining, measuring and analysing quality of life. Quality of life can be measured in an objective and subjective approach and different sets of indicators are proposed and used by various researchers (Mohit, 2013). An objective approach measures the conditions within different domains of life, using official statistics and information about the living environment. On the other hand, a subjective approach shows levels of satisfaction people feel in a specific area and conditions. Although both objective and subjective measures are present in current research, in recent years, interest in using subjective measures has increased, as well as for combining both approaches (Ballas, 2013). This means that the importance of using people's perceptions in evaluating QoL is growing.

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 data are an important part of many urban policy initiatives (Shelton, Poorthuis, & Zook, 2015), and digital media is already used to analyse different aspects of cities and spatial distribution of various urban functions (Shelton et al., 2015). As a matter of fact, social media are often seen as a perceptive extension of human thought (Sui & Goodchild, 2011) and are used on a daily basis to express opinions. Moreover, an important characteristic of social media data is wide availability and constant multiplication in cyberspace, giving researchers the opportunity to go beyond official statistics (Shelton et al., 2015). At the same time, social media data can have both geospatial footprints and indicative words that can be used in the process of collecting and analysing information.

One of the ways to improve the existing methodologies in a subjective QoL approach is capturing people's perceptions using data derived from social media. Social media data represent one type of Volunteered Geographic Information (VGI), or according to Kitchin (2014) "data gifted by users" (p. 4). However, unlike, for example, OpenStreetMap, where people choose to make a contribution by updating the existing geographic datasets (Yang, Raskin, Goodchild, & Gahegan, 2010), social media offers spatial and temporal tagging of people's thoughts (Shelton, 2016), and the opportunity to use these in evaluating their quality of life.

In general, quality of life research offers interesting challenges, in both data collection process and methodology development. For the purpose of this study, quality of life is defined as a level of general satisfaction people feel with their living conditions. This includes the fact that people tend to express their personal opinions about their life, how they emotionally feel and how they see their living surrounding. Moreover, an important aspect of this research is the assumption that people tend to express their perceptions and opinions in a self-reported way using social media platforms. This requires us to develop suitable steps to understand the nature of social media messages and ways to use and analyse these in QoL research.

1.2. Research problem

The characteristics of neighbourhoods have a direct effect on inhabitant's life quality, which in turn shapes their perceptions about the living environment. Neighbourhoods are dynamic and change over time, and this change affects people's perceptions as well. One of the main challenges in quality of life research is finding the proper methods to measure these perceptions and efficiently capture the dynamics of the community.

Overall, the traditional collection of subjective perceptions can be time-consuming, expensive and slow (Bibo, Lin, Rui, Ang, & Tingshao, 2014; McCrea, Marans, Stimson, & Western, 2011). Due to this, data sources such as social media could play a significant role in capturing people's perceptions. However, a unified way for proper collection and analysis of these widely available data is not yet found, and it is not well understood how these data could be used in the quality of life research.

Currently, 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, this study will try to address this gap in knowledge and contribute to the current discussion by focusing on exploring the use of social media by developing indicators to capture people's perceptions about their life based on Twitter data.

1.3. Research objectives

In this section, general objective, specific objectives and following research questions are defined, followed by hypotheses.

1.3.1. General objective

The main objective of this study is to evaluate the applicability of social media data in capturing subjective QoL perceptions.

1.3.2. Specific objectives

1. To derive subjective QoL domains and evaluate different perceptions on QoL using content analysis of Twitter data
2. To apply and map the QoL perceptions in Bristol, United Kingdom
3. To compare subjective QoL perceptions with official survey results in Bristol, United Kingdom

1.4. Research questions

1. To derive subjective QoL domains and evaluate different perceptions on QoL using content analysis of Twitter data
    a) What are the steps and criteria for deriving subjective QoL domains using Twitter data?
    b) Which domains of subjective QoL are suitable to measure with Twitter data and why?

2. To apply and map subjective QoL perceptions in Bristol, United Kingdom
    a) What are the most significant subjective QoL perceptions about quality of life in Bristol?
    b) What are the geographic patterns of identified subjective QoL perceptions?
    c) Do the geographic patterns of identified subjective QoL perceptions show significant differences between subjective QoL in the neighbourhoods?

3. To compare subjective QoL perceptions with official QoL survey results in Bristol, United Kingdom
    a) Do results from this study reflect the results of an official survey?
    b) Which subjective perceptions derived from Twitter compare to which official survey domains?

1.5. Research Hypotheses

Several hypotheses are identified:

- Twitter data can be used to evaluate different perceptions on quality of life.
- The patterns of perceptions will reveal significant differences between QoL within neighbourhoods.
- The results of this study will reflect the results of the official QoL survey in Bristol.

2. SUBJECTIVE QUALITY OF LIFE AND ROLE OF SOCIAL MEDIA IN CAPTURING PEOPLE'S PERCEPTIONS

This section provides a brief literature review of the most important concepts of the research. It covers the key features of subjective QoL approach and the importance of people's perceptions when evaluating QoL in a particular area. It also includes the literature on importance and possibilities of social media as a new source of data in scientific research as well as the use of social media data in quality of life research and methods to analyse these data.

2.1. Subjective QoL research

Subjective approach in QoL research has 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 of well-being (Kapteyn, Lee, Tassot, Vonkova, & Zamarro, 2015), happiness (Diener, 2000), good life (Bonn & Tafarodi, 2013), life satisfaction (Carlquist, Ulleberg, Delle Fave, Nafstad, & Blakar, 2016) are commonly used to address the same phenomena (Carlquist et al., 2016). This lack of conceptual uniqueness is usually a major issue, both for researchers and policy makers and it is important for researchers to define clearly the concepts in the beginning phases of their studies.

Similarly, in the past few decades, defining subjective QoL has been a major challenge in social sciences and topic of many debates in different fields of study (Ballas, 2013). Nevertheless, the subjective approach in quality of life research is commonly defined as a measure of people's feeling of general satisfaction with their living conditions (Davern & Chen, 2010; Diener, 2000; Marans, 2003; Marans, 2015; Schuessler & Fisher, 1985).

Various studies emphasise the relevance of using subjective approach for capturing conditions of the living environment. For example, Moro, Brereton, Ferreira and Clinch (2008) used subjective indicators with data collected in the self-reported way done through the national quality of life survey to rank the level of satisfaction in Ireland. Davern and Chen (2010) used GIS technology to emphasise the spatial context of QoL, analysis, and map subjective well-being of people living in Victoria, South-East Australia. Similarly, Santos, Martins, and Brito (2007) used a survey to capture citizens 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, Islam, and Hamilton (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. Data were collected by telephone survey in Toronto and Bogota and by the face-to-face survey in Belo-Horizonte. Researchers in these examples used different approaches to address the issue. They used qualitative, quantitative and mixed method approach, as well as primary and secondary data. On positive side, methods used are versatile and adaptable to the needs of a researcher

When it comes to characteristics of the subjective QoL approach, there are several main points to cover. First, as can be seen from the examples above, different approaches and methods can be used to generate results. However, the most common measures of QoL are identified as indicators measured within different sets of domains. Although the focus of this study is the subjective part of the QoL evaluation, the important thing to mention is a significant difference between objective and subjective indicators.

Costanza et al. (2007) argue that objective indicators can be used to evaluate different opportunities to improve their life quality, but not directly measure the phenomena itself. That is why they suggest that subjective indicators should be used to provide meaningful insight into people's perceptions about their personal well-being. Pacione (2003b) wrote about subjective social indicators as a way to assess urban liveability, more precisely, the relation between people and their living environment. These indicators are focused on the self-reported perception of life satisfaction in a certain location and can be effectively used to assess differences in QoL between neighbourhoods (Moro et al., 2008). The opinions are often conflicting, favouring one approach over another. However, contemporary evaluations of QoL preferred the use of both approaches. It is more informative to find the connection between people's perceptions and objective conditions of their living environment.

Next, indicators are usually measured within different domains. The range of domains depends on of the needs of the measurements. As previously stated, in the subjective QoL, measurements mostly focus on self-reported, individual reports about the life satisfaction and life experiences to show the importance of the perceived need for a person's quality of life (Costanza et al., 2007). These needs are often classified into different domains (Costanza et al., 2007) and one of the goals of the assessing the subjective QoL conditions is to find the way to recognise them. The decision about domains is usually guided using a previously structured framework, based on QoL theory. Sirgy (2011) explains this as a top-down approach in QoL research, where domain selection is guided by theory and previous knowledge, and because of that, measures often have more credibility. On the other hand, Dluhy and Swartz (2006) introduce the expansion of community-based projects, where relevant domains and indicators are recognised by residents and community members. According to Sirgy (2011), this bottom-up approach is "essentially constrained in meaning or theoretical relevance" (p. 2). The conceptual framework, outlined in Figure 1 at the end of this chapter, introduces both approaches and serves as a guideline for present study.

Moreover, QoL domains also depend on the place, and the specific interaction people have with their surrounding (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. In fact, one of the challenges in this study is finding the way to connect people's perceptions generated from the data, and domains used in previous research in the same area. With attention to previously mentioned top-down and bottom-up approaches, this research can be defined as an attempt to combine these approaches, generating insights directly from the data and connecting them to well-known theory.

Examples of ranges of subjective QoL domains used by different authors depending on their needs and methods are introduced in Table 1. Even though various names are given for these domains, they can be summarised in several categories, as they all examine similar aspects within subjective QoL: quality of living, health, education, work, safety and security, community, emotional well-being, transport, and green spaces.

Table 1. Domains of subjective QoL used by different authors

Authors	Subjective QoL Domains
Bramston, Pretty, & Chipuer, 2002	Material well-being, Health, Learning, Intimacy, Safety, Community Involvement, Emotional well-being
Ibrahim & Chung, 2003	Family life, Social life, Working life, Education, Health, Wealth, Religion, Leisure, Self-development and Housing, Public Safety, Public Utilities, Politics, Transport, Media, Consumer goods and services, Healthcare, Environment

Das, 2008	Physical environment (housing, green areas, pollution), Economic environment (own economic conditions, cost of living), Social environment (security, traffic, health, welfare services)
Wills-Herrera et al., 2009	Standard of living, Health, Achieving in life, Personal relationship, Safety, Feeling part of the community, Future Security, Economic situation, State of Environment, Social conditions, Government, Business, Local Security
Eby, Kitchen, & Williams, 2012	Transportation, Recreation, Housing, Crime, Safety, Green space, Diversity, Integration
Rezvani, Mansourian, & Sattari, 2013	Physical Environment, Economic Environment, Social Environment
Haslauer, Delmelle, Keul, Blaschke, & Prinz, 2014	Living, Education, Work and Employment, Security, Health, Mobility, and Participation

In the United Kingdom, subjective QoL approach has been regularly used, especially in the past decade, to capture people's feeling about their life quality. Numbers of surveys are used to measure different aspects of subjective QoL on the local, national, European and international level. Table 2 shows different domains used to measure people's subjective perceptions in various surveys.

Table 2. Domains of subjective QoL measurements (Source: Eurofound (2016), ESS ERIC (2016), ONS (2016), Bristol City Council (2015))

Bristol QoL Survey	United Kingdom National Well-being	European QoL Survey (EQLS)	European Social Survey (ESS)
Health and healthy lifestyle	Health	Health	Health
Community cohesion	Our relationship	Perceived quality of society	Well-being
Keep Bristol working and learning	Personal well-being	Life satisfaction	Fear of crime
Personal finance	What we do	Employment	Media use
Crime and anti-social behaviour	The economy	Income	Politics
Vibrant Bristol	Education and skills	Education	Trust in institutions
Keep Bristol moving	Building successful places	Level of happiness	Immigration
Green capital	The natural environment	Family	Religion
	Where we live	Housing	Human values
	Governance	Work-life balance	Demographics

Domains are collected from official surveys on European, national and local level. Table 2 includes European QoL Survey (EQLS), European Social Survey (ESS), United Kingdom National Well-being Survey and Bristol QoL Survey.

In conclusion, many scientists agree on the importance of using subjective assessment in examining QoL and understanding the issues and needs of residents in a particular area. Also, there is an abundance of available methods to approach the evaluation. Moreover, there is a clear distinction between top-up and bottom-up approach in the domain definition. However, the common denominator that connects all of these approaches is a central role given to the people and their opinions about QoL.

The importance of local context is also emphasised. Not every area can be observed in the same manner, and all characteristics have to be taken into consideration. The methodological approach has to be designed in the way it covers relevant questions and addresses important issues. To choose appropriate domains for analysis, the type of information the study is looking for has to be known upfront.

2.2. Social media in studying people's perceptions

Conole, Galley, and Culver (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 amount of chosen individuals.

Many authors tried to engage in the complex issue of using social media data in scientific research as an inexhaustible source of people's thoughts, feelings, and observations. Although there are debates about the usability of these data, numerous authors agree that data derived from social media represents a possible new source for gathering knowledge about different social issues (Aladwani, 2015). Today, the problem is not how to get the data from social media, because there are various examples of organisations involved in extensively collecting data for several years (Zook & Poorthuis, 2015). The more important question is how to get meaningful insight.

In the last decade, social media gain popularity in studying people's perceptions (Lieske, Martin, Grant, & Baldwin, 2015) and among various options, Twitter is one of the most used platforms (Arribas-Bel, Kourtit, Nijkamp, & Steenbruggen, 2015; Bibo et al., 2014; Chen & Yang, 2014). Social media data were used in numerous studies, and, depending on the research topic, providing different types of information. For instance, companies often analyse messages from social media to get useful information about their brands. McKerlich, Ives and McGreal (2013) used content analysis of social media data to show positive and negative reactions on different brands. Similarly, Lo, Chiong and Cornforth (2016) demonstrated the usability of Twitter data in recognising the potential new customers for analysed products. In health science, various topics have been covered using social media. For example, Almazidy, Althani and Mohammed (2016) developed a framework for harvesting Twitter data in a disease outbreak to have an additional source of knowledge about disease spreading patterns. Furthermore, Twitter data are also used in disaster management. A good example is provided by Chatfield, Scholl and Brajawidagda (2013). They examined the usability of the Twitter tsunami early warning system in government and the role of people in a transfer of information. The purposes for analysing social media data in these examples were different. However, all studies focused on how people's opinions proved useful in assessing various phenomena and the role people had in producing knowledge and transfer information.

Similarly, using social media data gained popularity in urban planning. As mentioned before, one of the major advantages of social media is an opportunity to observe and analyse people's perceptions, needs, interests, etc. Hence, there is a possibility of gathering new knowledge from these data to inform decision makers and contribute to urban planning and design processes (Larsson, Söderlind, Kim, Klaesson, &

Palmberg, 2016). Even though it is not very obvious, there is a strong connection between online and physical space, especially when geo-located social media data are analysed. Messages produced in the online world have a strong relation to the physical location. Therefore, the spatial component of social media data is emphasised. For example, Tweet patterns may show the land use and diversity within the city, information about consumers and producers, proximity patterns, and so on. Moreover, there are possibilities for using social media information in geospatial science and urban planning (spatial segregation, social profile evaluation, measurement of satisfaction, traffic management, and so on.) (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, Lupi, & Simeone, 2014).

Some might argue against the use of social media due to the lack of scientific traditionality, but the richness and possibilities these data offer cannot be overlooked. Graham and Shelton (2013) hope that, based on the history of geography with diversity in theoretical and methodological paradigm and practices, the value of big data will be recognised in future research.

2.3. Social media in quality of life research

In the 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, Iacus, & Canova, 2015; Nguyen et al., 2016). Next, Bibo, Lin, Rui, Ang and Tingshao (2014) used 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, Harris, Kloumann, Bliss and Danforth (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, not much has been done when it comes to combining QoL research and social media data. Nevertheless, based on the studies that have embarked on this interesting and challenging issue, domains these researchers covered are listed in Table 2.

Table 3. Different authors and domains for studying subjective life quality using social media data

Authors	Domains
(Curini et al., 2015)	Overall perceived happiness and subjective well-being
(Bibo et al., 2014)	Subjective well-being
(Dodds et al., 2011)	Perceived happiness
(Nguyen et al., 2016)	Happiness, food and physical activities

The main challenges these authors encounter were about how representative the data are, issues with lack of technical knowledge, and limitation of the data itself. The samples used may not be representative of the whole population of the area analysed. Moreover, some population groups, like younger people, tend to be overrepresented. In addition, one of the major obstacles was to overcome the issue of lacking technical knowledge. The challenging part was working with new technologies to clean the data, reduce the noise level and prepare the data for further analysis, and perform the analysis. Furthermore, they recognized the limitations of data itself, because working with unstructured messages can be tricky.

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 dealing with new types of data. Especially study done by (Nguyen et al., 2016), as shown in Table 2, successfully used social media data to evaluate some of the domains that can be used in QoL studies as well.

Therefore, this research will focus on identifying which QoL domains can be derived directly from the data and capturing people's perceptions about their life quality within recognised domains.

2.4. Content analysis of social media data

The main part of capturing people's perception using Twitter data is going to be done using content analysis. Therefore, the aspects relevant for this study are reviewed.

Content analysis is widely used in a scientific research within different fields of study, both as a qualitative and quantitative technique (Hsieh & Shannon, 2005). They explain content analysis as flexible approach allowing many researchers to adapt the methods to their researched topic, but also emphasize downsides of the flexibility in lacking definition and exact procedural steps.

It is generally defined as an analysis of concepts and words stated in a certain text (Schwartz & Ungar, 2015). Bryman (2015) defines content analysis as a method of analysing documents to objectively and systematically quantify it based on previously defined categories. Objectivity is provided in generating the specific rules, which are going to be applied in an objective, transparent and systematic manner through the whole process of quantifying the analysed material (Bryman, 2015).

Content analysis is widely used in studying people's perceptions (Bryman, 2015). However, new opportunities with using and exploring social media data embrace new ways of text analysis and adaptation of traditional approaches to the new structures of text. Content analysis of social media messages is multi dimensional because it includes number of steps, beginning with initial analysis of words, spatial and temporal characteristics of messages, to the deep analysis of content, splitting data into pieces and capturing important connections between them (Croitoru, Crooks, Radzikowski, & Stefanidis, 2013).

The analysis of social media text messages requires a unique approach. Unlike the conventional analysis of surveys and interviews, this analysis is more data driven and exploratory, and the outcomes of the study are planned based on the information available (Schwartz & Ungar, 2015). It is completely dependent on the data and their characteristics. Moreover, the uniqueness of the analysis comes from the structure of the social media messages. The social media messages, in this case Twitter messages, are unstructured in nature (Chae, 2015). People use emoticons and acronyms, abbreviations, messages have spelling mistakes and often contain labelled words, etc. (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011) and this important characteristic have to be taken into account.

In the end, in this study, content analysis will include different sets of language processing methods. Text preparation is used to transform the unstructured forms of text into structured documents, using different techniques (Chae, 2015), and, when transformed and prepared for the analysis, such text can be used for analysis of key words, summarisation, analysis of word frequency, and so on.

2.5. Conceptual framework

This study explores subjective QoL derived from social media data. The conceptual framework (Figure 1) shows major concepts that are relevant to this research. The research focuses on subjective QoL, and the objective QoL part is added as an additional concept that is possible to use to understand the objective conditions beyond subjective perceptions better. The main goal is to check how data derived from social media could be usable in QoL research.

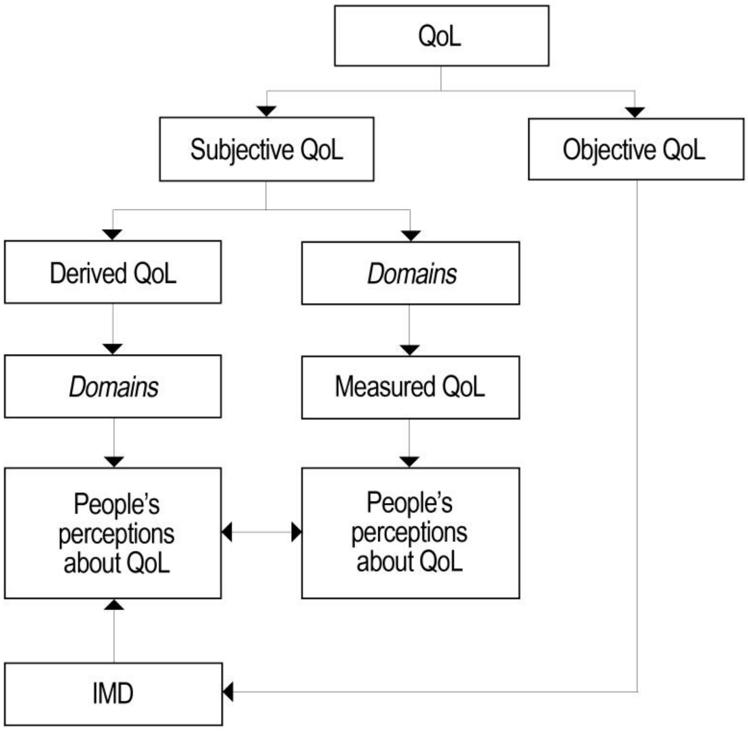


Figure 1. Conceptual Framework

In this research, two concepts of subjective QoL are observed, ***derived subjective QoL*** and ***measured subjective QoL.*** The difference between these concepts is in the approaches used for evaluating life quality. Derived subjective QoL is based on the inductive, bottom-up approach. Here, the evaluation is data driven, where data derived from social media are used to identify domains of subjective quality of life. Identified domains serve as guidelines for capturing people's perceptions about their subjective quality of life. The main idea behind this approach is to capture people's perceptions based on the things they are commenting about without previously asked questions. The result is the patterns of people's perceptions within derived domains. On the other hand, ***measured subjective QoL*** is based on the deductive, top-down approach. The evaluation is theoretically driven. People's perceptions are being measured based on previously defined QoL domains. This shows more traditional approach to subjective QoL where measuring is made using interviews, surveys, questionnaires, and such. The results are the patterns of people's perception within measured domains. This goes with a line with the previously mentioned distinction between bottom- up and top-down approaches captured by Sirgy (2011). The Central focus in this study is placed on derived subjective QoL while having in mind other concepts. The possibility of using social media data as a way to recognise subjective QoL domains and people's perceptions is investigated through the research. The domains are directly derived from the data and perceptions measured within these domains. Measured QoL is introduced in this study through the official QoL survey representing the people's perceptions obtained through way that is more traditional. To observe similarities and differences between derived and measured subjective QoL, the resulting perceptions from this study are compared to the perceptions from the official QoL survey.

The third concept from the framework is ***objective QoL***. The objective conditions of the living environment are often measured using the approach where index of multiple deprivation (IMD) is created. IMD is constructed in the way it includes the relevant indicators covering diverse aspects of people's life pointing to the differences in their life quality and levels of deprivation. In the studies that combine subjective and objective quality of life, IMD is often used to compare to present objective conditions of a living environment. Similarly, this study plans to use this measurement of objective conditions to evaluate what kind of association exists between the level of deprivation in the neighbourhood and subjective perception of the quality of life. In addition, to evaluate if there is a difference or similarity in the way level of deprivation effects measured and derived QoL. Moreover, the concentration of deprivation in a certain part of the city can show how that is connected with people using social media and in what way and to contextualise and quantify the spatial distribution of specific domains.

3. INTRODUCTION TO CASE STUDY AREA

This chapter provides an overview of a case study area, brief introduction to the specificity of the area and justification for case study area selection.

3.1. The city of Bristol

Bristol is located in the southwest of England. It is a sixth largest city in England, largest city and regional capital of this part of the country (Tallon, 2007). According to Census data from 2011, population size in Bristol was 428.100.

Bristol City region is an area of greater Bristol including the city of Bristol in the middle, South Gloucestershire in the north of the region, Bath and North East Somerset in the southeast and North Somerset in the southwest. The city of Bristol is the hub of the city region (Tallon, 2007).

The city of Bristol consists of 35 electoral wards, as illustrated in Figure 2. In may 2015, City Council made a change in boundaries and introduced new wards (Bristol City Council, 2015a). In this research, it was decided to do analysis and reporting in old ward boundaries, because of the possibility of connecting results from this study with indicators used in the official QoL survey in Bristol.

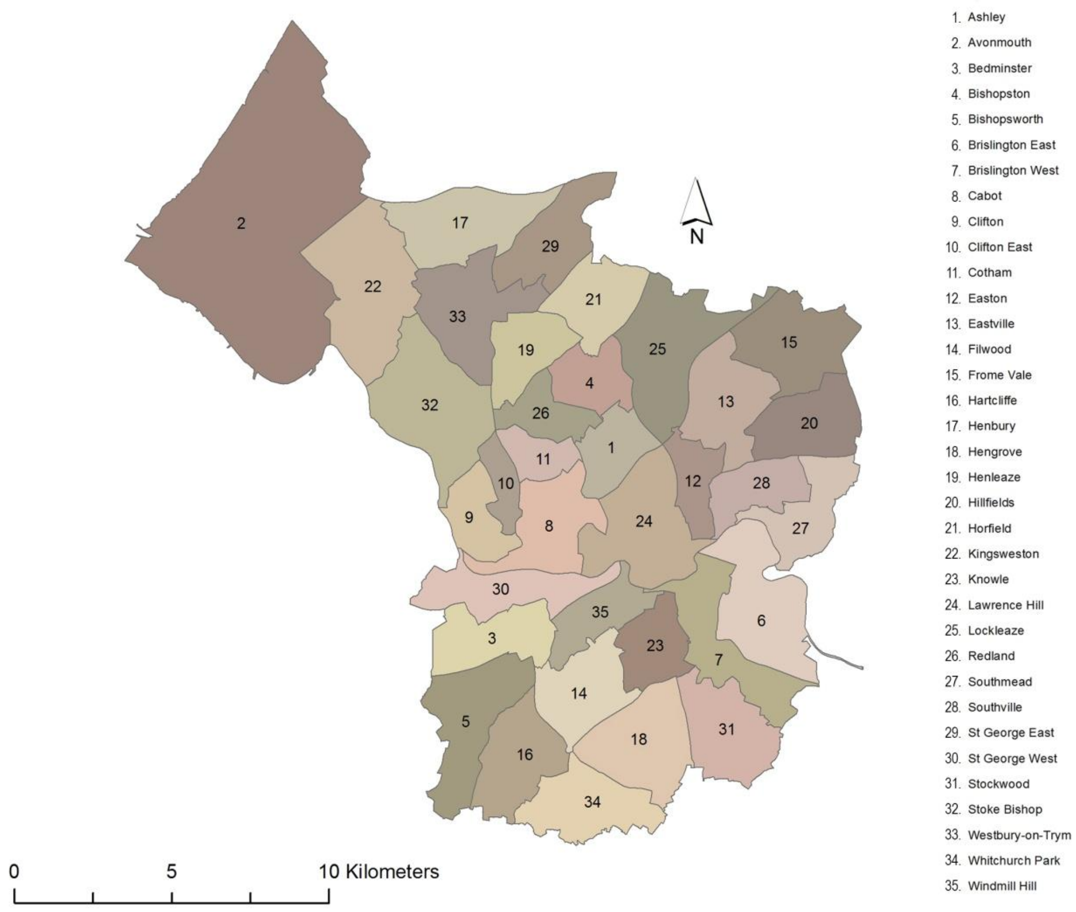


Figure 2. Bristol Wards in 2013 (source: own analysis based on data from Bristol City Council, 2015)

Bristol City Council established 14 neighbourhood partnerships in 2008. They are based on geographical closeness and made of two or three electoral wards. The idea of neighbourhood partnerships is to have every stakeholder involved in planning, problem-solving and decision making in Bristol.

Neighbourhood partnerships have regular meeting four times a year to discuss issues in the neighbourhoods and make decisions. They discuss topics like waste, recycling and clean neighbourhoods, parks and green spaces, dogs and dog ownership, neighbourhood safety, parking, planning and building control and within every topic, different parties have their responsibilities. Council, businesses, and citizens are in charge of specific tasks to keep the neighbourhood in the best conditions.

Bristol is a diverse city with many different cultures living together and sharing the living environment. Even though the city has a good 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.

Like many other cities in England, there is a significant difference between affluent and deprived areas in the city of Bristol (Tallon, 2007). Wealthy areas are located more in north-west part of the city, parts of the Henleaze and Redland Wards, while 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.

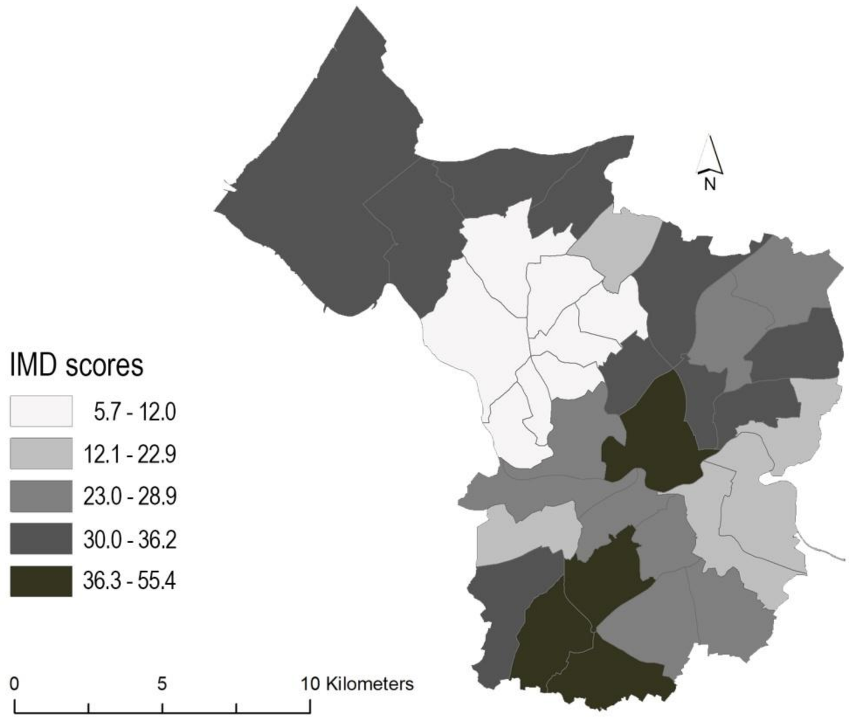


Figure 3. IMD for Bristol Wards where a higher value for the IMD indicates higher level of deprivation (source: own analysis based on data from Gov.UK, 2016)

Tallon (2007) connects inequalities with the existence of greater distance between jobs and housing, as new jobs are located in the northern parts of the city.

Bristol City Council (2015a) published a report on multiple deprivation in the city, and some of these issues are mentioned. According to the report, the city has several deprivation hotspots where problems are emphasised. 16% of residents live in the most deprived areas in England. Moreover, the highest levels of deprivation in the city of Bristol are in wards Whitchurch Park, Hartcliffe, Filwood and Lawrence Hill. Figure 3 shows that the wards with the highest level of deprivation are classified in the last category. Bishport Avenue in Bishopsworth ward and Hareclive in Hartcliffe ward are on the list of the most deprived hundred areas in England for index of multiple deprivation (IMD) in 2015.

3.2. Criteria for case study area selection

The specificity of using social media as the main source of data imposes specific requirements for selecting a case study. The criteria used took in consideration next characteristics of the city:

- Spoken language,
- Social media use, and
- Previous studies on quality of life (QoL)

English language

The main goal of the research is to check the applicability of Twitter data in the QoL research and capture people's perceptions using content analysis, which is an analysis of the text. Therefore, it was important that the city with predominantly English language be chosen for the analysis.

Twitter usage

Twitter emerged as a new social media platform in 2006, and since then a number of users is steadily rising. Today, based on the company fact, there are 313 million active monthly users, 82% of active users on mobile phones and more than 40 languages supported (Twitter.Inc, 2016). When it comes to Twitter usage, the United Kingdom is the second country in the world with over 15 million active users. Twitter is quite even with 49% males and 51% female users, and there are over 400 million Tweets sent daily.

History of QoL research

The suitability of this city also lies in the possibility to make a comparison between results derived from this research and previous studies on subjective QoL in the area. Bristol has a long history in QoL research (Mcmahon, 2002) and a good record of QoL data that can be used to compare and verify the results. The city has an official survey where they collect opinions of residents about various subjects (Bristol City Council, 2015). Data are analysed at ward level. The QoL domains used in the official reports, together with the literature on the topic, are going to be used to guide the domain selection for subjective QoL perceptions.

4. CAPTURING SUBJECTIVE QOL PERCEPTIONS – RESEARCH DESIGN AND METHODOLOGY

This section provides a description of the research design, data, methods, and tools used to answer the specific research questions. First, research design is briefly introduced. Ethical consideration as an important issue when analysing people's thought derived from social media are described. Next, necessary data are explained followed by a detailed overview of the steps in different parts of the analysis.

4.1. Research design

This research is designed to find the most appropriate way to capture subjective quality of life (QoL) using Twitter data. The main goal is to explore the potential social media has in producing meaningful results in QoL research. Analysis of social media is still something new in the field of QoL and doing so requires an exploratory approach. The starting point is to select the most appropriate traditional methods and techniques and adapt them for the purpose of the uniqueness of the data derived from social media.

This research is based on the mixed-methods approach, including both qualitative and quantitative methods to get a better understanding of the phenomena. Twitter data were analysed using a coding system and content analysis technique. The approach is inductive, which means that the results and observations are directly derived from the data. Moreover, the methodology includes semi-automatic approach, using manual coding and automated techniques. Using the data from social media, the domains of subjective QoL are derived and afterward compared with results from official QoL survey done in the city of Bristol.

Research design has several elements of cross-sectional design. It includes the content analysis and analysis of results from the official QoL survey. Although this research is not going to include conventional survey methods, data collected through social media are going to be analysed in a similar way. Moreover, there are elements of a case study, because the results are directly derived from the data collected for the specific area and analysed in local context.

Table 4 provides research design matrix summarizing data, tools, and methods necessary for capturing subjective QoL perceptions.

Table 4. Data needed and methods used

<table>
<tr><th>Research sub-objectives</th><th>Research questions</th><th>Analysis methods</th><th>Data and tools required</th><th>Anticipated results</th></tr>
<tr><td rowspan="2">To derive subjective QoL domains and evaluate different perceptions on QoL using content analysis of Twitter data</td><td>What are the steps and criteria for deriving subjective QoL domains using Twitter data?</td><td rowspan="2">Literature review
Content analysis</td><td rowspan="2">Literature
Twitter data
Atlas.ti
Excel</td><td rowspan="2">List of steps and criteria for deriving subjective QoL domains
List of dimensions suitable to measure with Twitter</td></tr>
<tr><td>Which domains of subjective QoL are suitable to measure with Twitter data and why?</td></tr>
</table>

To apply and map subjective QoL perceptions in Bristol, United Kingdom	What are the most significant subjective QoL perceptions about quality of life? What are the geographic patterns of identified subjective QoL perceptions? Do the geographic patterns of identified subjective QoL perceptions show significant differences between subjective QoL in the neighbourhoods?	Literature review Content analysis GIS spatial analysis	Literature Twitter data Excel ArcGIS	List of the people's perceptions about QoL Map of people's perceptions about QoL Discussion
To compare subjective QoL perceptions with official survey results in Bristol, United Kingdom	Do results from this study reflect the results of an official survey? Which subjective perceptions derived from Twitter compare to which official survey dimensions?	Literature review GIS spatial analysis Statistical analysis (paired sample t-test)	Official QoL survey data for Bristol Literature ArcGIS SPSS	Comparison between two studies Map showing the similarities and differences Discussion

4.2. Ethical consideration in analysing social media data

An increased number of studies exploring and using data derived from social media raised a series of questions regarding different ethical issues that could emerge.

An ethical approach to analysing big data is challenging because of its uniqueness and dynamic nature. The major concerns are about privacy, confidentiality, informed consent, and representativeness of a sample.

Bryman (2015) writes about specific questions emerging in the studies that involve internet and data collected from online sources. Bryman mentions that there is no clear difference between private and public space in the online world. Therefore it is hard to recognise the acceptable level of data use. Some authors (Bryman, 2015) suggest that consciously published data could be used without a form of consent as long as the authors of the data used are kept anonymous. Moreover, some of them argue that it is justifiable to use the data that can be publicly accessed without a password. In fact, one of the specific characteristics of Twitter is that users can choose if they want to have private or open accounts. For this reason, only open accounts with publically posted messages were considered for this research and authors are anonymized. Next, the way privacy issues should be tackled depends on how sensitive the topic is. If the topic is about children or violent behaviour, the privacy issue should be primarily solved.

Social media data are not representing the whole population of the area analysed. However, results of the analysis can give us a starting point for the further research to be based on. Moreover, we can argue that these spontaneously, freely written messages can give more sensible insight into people's subjective opinions than interviews and surveys, because there is no any influence of the researcher on researched people.

The best way to perform social media analysis is to ensure the ethical and sensitive approach to the data, and that anonymity is provided and data are used for scientific purposes only.

4.3. Data description

The data needed for this research are secondary data. The summary of the data is presented in Table 5, and detailed explanation is given in the next section.

Table 5. Overview of the data

Type	Source	Format	Year	Areal Unit
Twitter messages	DOLLY project	Excel	2012-2016	Points
QoL Bristol Survey	Bristol City Council	Excel	2005-2016	Wards
Index of multiple deprivation	Office for National Statistics UK	Excel	2015	LSOA

Twitter messages

The first type of data is messages posted by different users, collected from Twitter called Tweets. The tweet is a status message consisting of maximum 140 characters where people can express their opinions, thoughts, needs and so on. These messages were analysed and used for capturing subjective QoL. Tweets are short, unstructured text messages consisting of writing in different styles, slang, abbreviation, links, hashtags, and so forth. In Table 6 examples of the various types of Tweets are shown to illustrate their versatility and complexity.

Table 6. Examples of 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 yrs between July-September #r4today
What a lovely way to start an Autumn day :) http://t.co/gSnU9XFuFt
Hahaaaaa love it.. LOVE IT! People that cant see whats right in front of them.. Choosing to be ignorant! lol

Data used for this research are geo-tagged Tweets collected from January 2012 to September 2016 in the area of the city of Bristol. The Tweets are collected as part of the research in the University of Kentucky, in the Digital OnLine Life and You (DOLLY) project (Floating Sheep, 2016). DOLLY is an archive of billions of geo-tagged Tweets created for analysis and research in real time.

The dataset for this case study consists of 4,437,900 Tweets. Tweets and attributes are stored in .csv file format. Table 7 shows Tweet attributes and explanation.

Table 7. Attributes in the dataset and explanation

Attribute	Definition
id	Tweet Identifier
u_id	Tweet author identifier
lat	X coordinate of posted Tweet location
lon	Y coordinate of posted Tweet location
created_at	Time of Tweet creation
type	Type of Twitter user, private or corporation
u_location	Location of an user
u_lang	Language
URLs	Hashtags, labels used to tag the message
text	Text of the Tweet with maximum of 140 characters

QoL Bristol survey

The second type of data is indicator values derived from the official yearly survey on subjective QoL in Bristol. The QoL indicator values are calculated in the wards' level. The data from this survey will be used to compare results and see if there is a relation between them. Data are publicly available on the Bristol Council website (Bristol Council, 2016).

Since 2005, the city of Bristol uses an annual survey to collect people's perception about their quality of life. They used a set of 150 indicators within eight domains, and, in the last survey, closed in October 2015, approximately 30,000 households were invited to participate. The domains and indicators used in the city of Bristol were location specific and were not used in any other city in England.

The available dataset includes the results of the last survey, held in 2015, as well as data for previous five years. Even though survey questions were changed every year based on the specific problems in the city, key questions stayed the same, so it is possible to observe the trends over times. The data are available per electoral ward in excel database where all indicators and domains are included. The indicator values are in percentages.

Index of multiple deprivation in Bristol

The scores from multiple deprivation index are added as a data set representing objective conditions in the city of Bristol. Data are publicly available on the United Kingdom Government website (Gov.UK, 2016). The index of multiple deprivation (IMD) is the measure England uses to measure relative deprivation in small areas and can be observed as a measure of objective conditions of life quality of the people. The IMD is measured in England yearly since 2005. IMD combines various indicators to include a range of social, economic, environmental and housing characteristics and makes a single deprivation score.

Seven different domains and 37 indicators of deprivation are included in IMD. Domains of deprivation are:

- Income Deprivation
- Employment Deprivation
- Health Deprivation and Disability
- Education, Skills and Training Deprivation
- Barriers to Housing and Services
- Crime
- Living Environment Deprivation

IMD results are available in Excel dataset with scores for IMD for seven domains and six sub-domains at Lower-layer Super Output Area (LSOA) level. LSOA are small areas created to represent areas of approximately same population size, with an average of around 1,500 citizens. The ranks of the areas are based on scores and the larger the score, the more deprived the area is (and vice versa). For the purpose of this study, IMD scores are aggregated to ward level.

4.4. 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.

However, this research used a mixed approach, as it was intended to combine the theoretical knowledge about the subjective QoL, domains of the analysis and different approaches from the literature and insights from the data. QoL theoretical knowledge guided the steps for analysing data and extracting meaningful information, and therefore framed the research.

4.4.1. Preparation of Tweets for the analysis

Dataset used contained 4,437,900 Tweets. Different ArcGIS tools are used for preparing the data for further analysis. After clipping the data using the boundaries of the city of Bristol, the number of Tweets was reduced on 3,616,433 Tweets. Based on certain criteria, the year 2013 is chosen for the analysis.[1]

The justification for using the year of 2013 for research:

- Complete set of Tweets;[2]
- Publicly available shapefile for Bristol ward boundaries (Bristol City Council made a decision to change the boundaries of wards and introduce new boundaries in 2015. Spatial analysis played an important role in this study. Therefore year with available boundaries was chosen.)
- Previous studies on subjective QoL survey in Bristol in the same ward boundaries (One of the sub-objectives of this study is to compare ending results with results of official QoL survey in Bristol, and incorporate IMD as an objective measure of QoL. It was logical to use years that are more comparable)

Tweets for the year 2013 are aggregated into wards (administrative boundary) to see the spatial distribution of tweeting in the city of Bristol based on the total number of Tweets and prepare datasets for further analysis. The rest of the analysis is based on Tweets aggregated in wards. Twitter messages contain

[1] Preparation of Tweets flowchart is available in Appendix 1

[2] Looking only into dataset containing Tweets, year 2012 and 2016 were incomplete. Moreover, years 2014 and 2015 had strange numbers, not consistent with number of messages in other years and months available in dataset.

different information. Some of the key attributes of Tweets are the text of the message, date of creation, the number of retweets and favourites, id (Tweet identifier), coordinates, users, and so on.

In this study, descriptive statistic was used to show characteristics of Twitter usage and spatial distribution of these features. The analysis was done per ward in ArcGIS. ArcGIS was used to calculate the number of Tweets per ward. Normalisation of Twitter usage was done using population size to show the number of Tweets per capita per ward and to prepare the data for further analysis. The formula used for normalisation is:

$$TwPop = Pop/Tw$$

Where *Pop* is the size of population in ward and *Tw* is a number of Tweets in ward.

Next step was a visualisation of Twitter usage in Bristol in 2013. The most common difficulty in the visualisation of a large set of data is overplotting. Several studies addressed this issue and suggested possible options (Zook & Poorthuis, 2015). If the dataset is relatively small, the best solution is to plot slightly transparent points. Another possible answer is making heatmaps using kernel density or similar methods. However, it is hard to get meaningful insight from social media using these types of visualisations. For this study, an adequate method was to aggregate points into larger areas, as suggested by Zook and Poorthuis (2015). This allowed us to engage in the spatial domain and see the variations and spatial distribution of Tweets. 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.

4.5. Content analysis

Twitter data were processed using coding system and content analysis technique. Messages posted by the Twitter users were categorised based on the content.

The approach was semi-manual. It involved manual coding and automated analysis as most important components of the content analysis. Overview of steps is presented in Figure 4.

The content analysis of the Tweets was done using Atlas.ti, Excel and ArcGIS software. Atlas.ti is software for qualitative data analysis and it was used for manual coding of Tweets as a first step of the analysis. Microsoft Excel is spreadsheet-based software used as a part of Microsoft Office package. Even though it is a simple software, it offers options for doing an analysis of the text using a programming language called Visual Basic for Applications and different open source add-ins made specifically for text analysis.

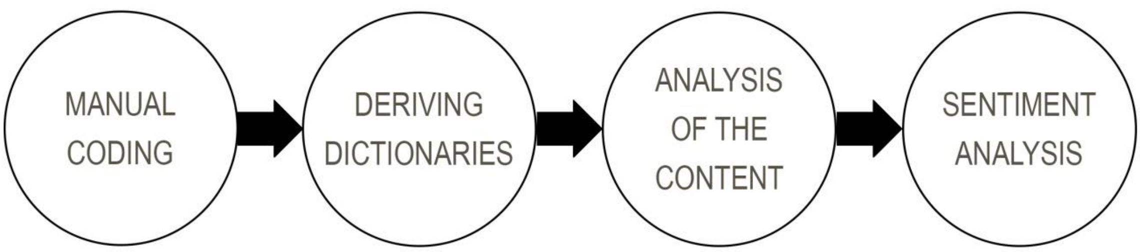


Figure 4. Content analysis steps

4.5.1. Qualitative analysis - Manual coding

The first phase in the content analysis was to derive subjective QoL domains for the analysis. Steps included are presented in Figure 5. In this research, domains were derived from manual coding using software Atlas.ti. For manual coding, a random sample was calculated for the area of Bristol, for the year 2013, based on the total number of Tweets (1,374,706).

Tweets were normalised based on the population size.

Random sample was calculated for the area of Bristol, based on the total number of Tweets (3616433). The sample size is calculated with confidence level 95%, and confidence interval 3, based on the formula:

$$ss = \frac{Z^2 * (p) * (1-p)}{C^2}$$

Where:

Z is Z value (for confidence level of 95% it is 1.96)

p is percentage picking a choice in decimals (.5 used for sample size needed)

C is confidence interval in decimals (0.03 for ± 3)

Formula for adding the population size:

$$SS\ new = \frac{SS}{1 + \frac{SS-1}{Pop}}$$

Where Pop is the population size.

The size of the sample used was 1067 Tweets for the area of Bristol. Based on the normalised values (Number of Tweets per capita per ward), the size of the sample for every ward was calculated and Tweets were selected using Random sample option in SPSS software. In this way, every Ward had a different number of Tweets in the sample based on the number of Tweets per capita.

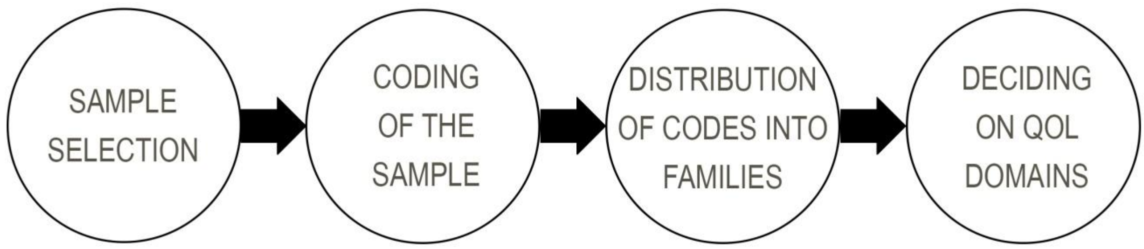


Figure 5. Manual coding steps

After the sample was prepared, coding was the next step. The first round of coding was done using a free coding technique to recognise QoL perceptions, derive domains and generate a codebook for further analysis. The random sample of Tweets was systematically reviewed and assigned with codes to label the possible QoL domains based on people's thoughts. Words and phrases indicating relevant domains for subjective QoL assessment were recognised. 66 free codes[3] were generated and a total number of 102 subjective QoL perceptions captured. Figure 6 illustrates examples of observations about QoL recognised in coding sample and their distribution in the city of Bristol.

Further, relations between codes were identified and more structured part of the coding analysis was done. Families of codes were defined and served as points for grouping similar codes. The families of codes were generated based on previously reviewed domains from different studies done on subjective QoL in Bristol and the United Kingdom and insights from the data. Table 8 shows code families with a number of observations and number of codes within families.

Table 8. Families of codes (subjective QoL domains)

Code Families	Number of Observations	Number of Codes
Health	31	19
Feeling of satisfaction	28	18
Community	14	6
Social life	10	7
Transport	8	6
City	3	3
Employment	3	2
Environment	2	2
Safety	2	2
Finances	1	1
Total	102	66

For the purpose of quality control, triangulation was done in manual coding by involving additional human coders. The resulting codes were similar to the initial coding and confirmed the results. In addition, they served as a guideline in the process of selecting which domains are going to be further analysed.

[3] List of all codes derived in qualitative analysis is available in Appendix 2

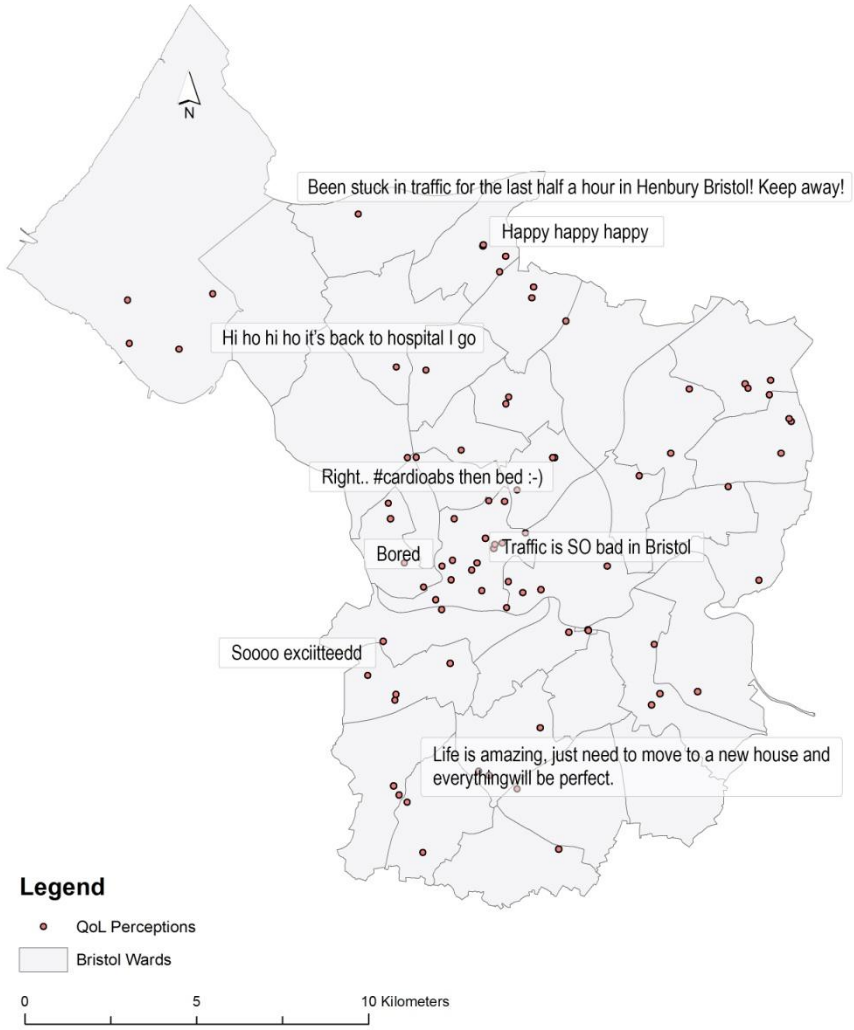


Figure 6. Examples of QoL perceptions captured in manual coding in Atlas.ti

It became apparent at this stage of the analysis that the two subtypes of perceptions can be distinguished in general, regardless of the domains:

- Emotional reaction where people express feelings, and
- Cognitive conclusions where people express opinions

Given these points, three domains (Figure 7) were selected to be analysed in automated part of the analysis. The selection was based on the exploratory nature of research and the characteristics captured in the process of analysing the random sample. Three domains were selected to examine how successfully would next step be able to manage to classify the perceptions in these domains.

Domains and justification for selection:

- Health - domain with a larger number of perceptions and more emotional reactions
- Transport - domain with a greater number of perceptions and more cognitive conclusions
- Environment - domain with fewer perceptions and mixed types of perceptions

Results from this part of the analysis served as a guideline for the next steps of the analysis. Furthermore, they provided an answer to the first two research questions which will be elaborated in the discussion section.

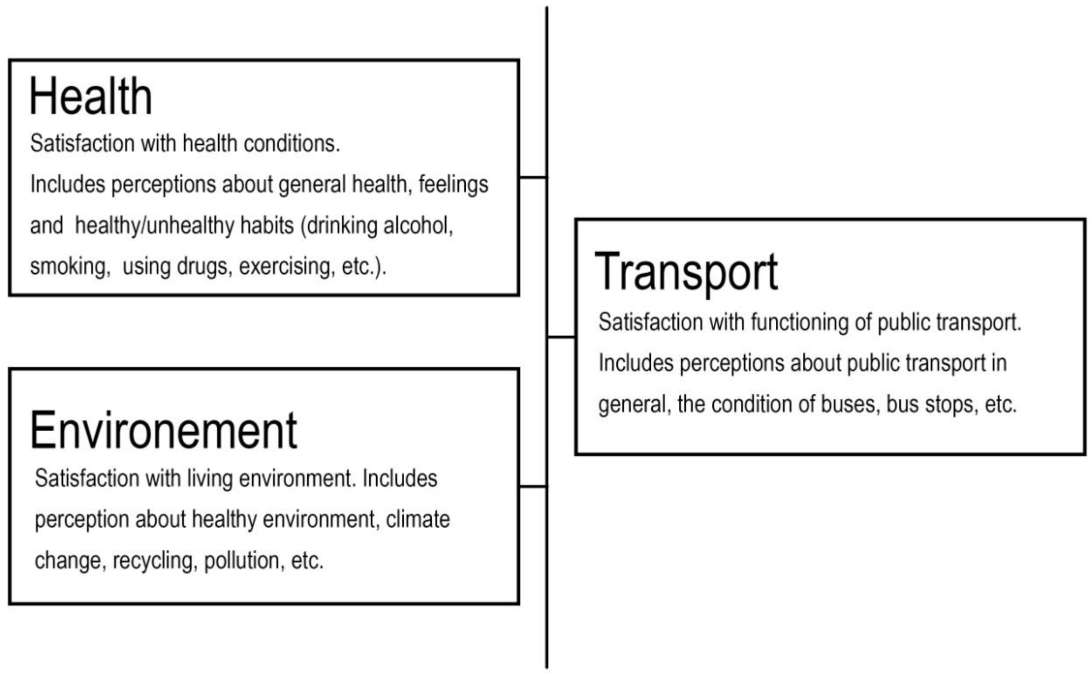


Figure 7. Subjective QoL domains and definitions

4.5.2. Text pre-processing

After manual coding, next part of the content analysis was text pre-processing, where collected Tweets were prepared for automated part of the analysis. The automated part requires text to become more structured and Tweets to be filtered and cleaned, in order to be properly executed.

In order to clean the Tweets and reduce the noise, next steps were done:

- Conversion of hashtags to words (Hashtags are converted to words with removing the symbol for hashtag (#) in front of the word. This step provides the possibility for the words that were hashtagged to be included in the filtering in the next stages of the content analysis, as the hashtags represent relevant indicative words);
- Removing punctuation;
- Converting all text to lower case;
- Removing usernames @username.

Tweets were filtered and cleaned using a programming language called Visual Basic for Applications used in Excel and using formula (LOWER) in Excel that converts all text to lower case.

The code used for removing punctuation, @, and # is:

```
Function RemPunct(Txt As String) As String
With CreateObject("VBScript.RegExp")
```

```
.Pattern = "[^A-Z0-9 ]"
.IgnoreCase = True
.Global = True
RemPunct = .Replace(Txt, "")
End With
End Function
```

4.5.3. Generating dictionaries

Text retrieval operations require a good strategy, having all the steps of the analysis defined and good structure to follow. However, the content analysis allows a certain amount of creativity in defining these steps due to the specific requirements of the researched topic. When creating a dictionary or a coding scheme for performing the analysis, it is important to know what kind of information we are looking for.

According to literature (Hsieh & Shannon, 2005; Schwartz & Ungar, 2015)generating dictionaries is one of the most challenging parts of the content analysis. It is essential for the analysis to produce a good set of indicative words and their synonyms to guide the retrieval of Tweets within different domains. Dictionaries are defined as a list of indicative words for a specific topic reflecting the relevant information generated based on previously defined domains. Search terms can be chosen based on the knowledge about subjective QoL domains, and closely related terms that can be grouped under the same domain.

According to Schwartz and Ungar (2015), 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 to a large collection of text.

In this research, dictionaries for selected domains 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. Moreover, the sample used for qualitative analysis in Atlas.ti was used for extracting words and phrases that are more frequent. Afterward, words and phrases relevant to the topic were manually extracted from the words and phrases frequency lists and assigned to corresponding domain dictionary.

Based on the number of domains recognised through the qualitative part of the analysis, dictionaries for three domains were constructed: health, transport, and environment dictionary. Dictionaries are simple and easy to use. They were data driven, which makes them specific for dataset used for this research. However, words and phrases were manually inspected and divided into categories based on the QoL literature and knowledge gained through the research. Every domain dictionary contained 25 indicative words.[4]

4.5.4. Quantitative analysis - Content classification

Classification of the content was the next step of the analysis. Before this stage, Tweets were cleaned and prepared for analysis, domains were derived and dictionaries generated.

The classification of the content was systematically done ward by ward by filtering Tweets for selected ward through the dictionary for every recognised domain.

[4] Dictionaries for every domain are available in Appendix 3

The filtering of Tweets was done in Excel using the formula:

=SUMPRODUCT(--ISNUMBER(SEARCH('words',A1)))>0

Where 'words' are the words from dictionaries used for filtering Tweets, and A1 is a field where text prepared for the analysis is. The results are true or false, where true is assigned to text containing words from dictionaries.

The result was a count of the Tweets containing one or more words or phrases for selected domain per ward. Tweets were selected, dictionaries with corresponding words and phrases were added, the operation of filtering the Tweets based on the words and phrases was selected and the result were Tweets expressing some form of the perception on subjective QoL within selected domains. 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 normalisation (Patterson & Hoalst-Pullen, 2014; Zook & Poorthuis, 2015). The advantages of using odds ratio are the opportunity to normalise our perceptions by any other variable and easy to understand results (Zook & 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:

$$OR = \frac{Pw/Ptot}{PopW/TwPop}$$

Where Pw is number of Tweets in ward related to the domain observed (for example, the number of Tweets about health in one wards), Ptot 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 amount of Tweets containing QoL perception based on the total tweeting population. The midpoint of a resulting ratio is 1. This indicates that the amount of perceptions is closest to the one expected based on the tweeting population size. An odds ratio greater than 1 indicates that there are more perceptions than expected based on the total tweeting population, and vice versa. Standard deviation classification method is used to visualise odds ratio of QoL perceptions in different domains, where mid values correspond to mid points of odds ratio.

4.6. Sentiment analysis

The final step of the content analysis was sentiment analysis of Tweets in different domains. This analysis is done using excel add-in[5] that offers different possibilities of analysing text. Classified content is categorised based on the semantic scores of the perceptions within domains.

The Tweets were classified into a five-point scale:

- highly positive (P+)
- positive (P)

[5] MeaningCloud™ (http://www.meaningcloud.com) has been used for Text Analytics purposes in the development of this research. Description of the tool is available in Appendix 4

- neutral (NEUT)
- negative (N)
- highly negative (N+)

Moreover, there was a sixth, uncategorised class where semantic analysis could not provide an answer about the sentiment of the message.

Next, positive and negative perceptions were counted and compared to check if they are significantly different. Paired sample t-test was used to detect if there is a significant difference between two groups, positive and negative perceptions.

For the case where the hypothesis is rejected, the assumption was made to distinguish wards with significantly higher percentages of positive and negative perceptions. The assumption was that if in one ward the percentage of tweets with positive perceptions is 5% bigger than those with negative ones, this ward is classified as positive, and vice versa. A dummy variable was created for both positive and negative classifications where if the assumption is true the value given is 1, else 0.

The resulting positive and negative perceptions were visualised using ArcGIS to show spatially similarities and differences in perceptions between wards in Bristol.

4.7. Comparison between derived and measured subjective QoL

The final part of the analysis included a comparison between perceptions derived in present study and opinions of residents captured in the official QoL survey in Bristol, referring to these results as derived and measured QoL as layed out in the conceptual framework (Chapter 1). Comparison was done statistically and spatially.

To show similarities between results, the null hypothesis was introduced following third hypothesis in this research. The null hypothesis was that two variables derived from two different studies are the same, i.e. the results of the present study will reflect the results of the official survey. The analysis was done to determine whether two means from two variables are statistically different. For the purpose of this, Paired Samples T-test was done in SPSS. Positive percentages of perceptions in health, transport and environment domain were used as variables derived in present study, and percentage of respondents satisfied with health services, percentage of respondents satisfied with bus services, and percentage of respondents satisfied with quality of parks and green spaces 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 overlapped with percentages of people satisfied in the corresponding topic.

- Health is overlayed with the percentage of respondents satisfied with health services.
- Transport is overlayed with the percentage of respondents satisfied with bus services.
- Environment is overlayed with percentage of respondents satisfied with quality of parks and green spaces

5. RESULTS

This chapter gives an overview of results obtained through the analysis. First, people's subjective perceptions about QoL in Bristol in 2013 are presented and described. Next, results for three subjective QoL domains observed in the study (health, transport, and environment) are presented. Here, perceptions are qualitatively described, and examples are given. Spatial distribution of perceptions is presented to demonstrate differences between wards. Moreover, sentiment analysis results are presented to show the amount of QoL perceptions that can be classified into one of the sentiment categories. In this part, the spatial distribution of positive perceptions is presented. Finally, perceptions about subjective QoL derived in this research are compared with finding from the official QoL survey done in the city of Bristol.

5.1. Subjective QoL perceptions in Bristol

People using Twitter in the city of Bristol in the year 2013 have opinions on different topics that can be categorised in various quality of life (QoL) domains. People's perceptions are captured and measured for three subjective QoL domains, selected and defined in previously described steps in the methodology section (Chapter 4) health, transport, and environment.

Table 9. Summary of QoL perceptions for Bristol, 2013

Bristol (2013)	
Number of geolocated Tweets	1,374,706
Number of QoL perceptions derived from tweets	61,970
Percentage of tweets containing QoL perceptions	4.51
Number of Tweets per tweeting population	22.22
Number of perceptions about health	25,187
Percentage of health perceptions	40.60
Number of perceptions about transport	31,247
Percentage of transport perceptions	50.60
Number of perceptions about environment	5,536
Percentage of environment perceptions	8.80

In this study, analysis used a total number of 1,374,706 Tweets posted in 2013 in Bristol, where 4.51% of Tweets can be categorised in one of the subjective QoL domains derived through the study. Table 9 shows summarised values for the area of Bristol, where numbers of QoL perceptions about health, transport, and environment in Bristol in 2013 are illustrated.

Figure 8 shows the relations between the percentages of perceptions in each domain in Bristol wards.

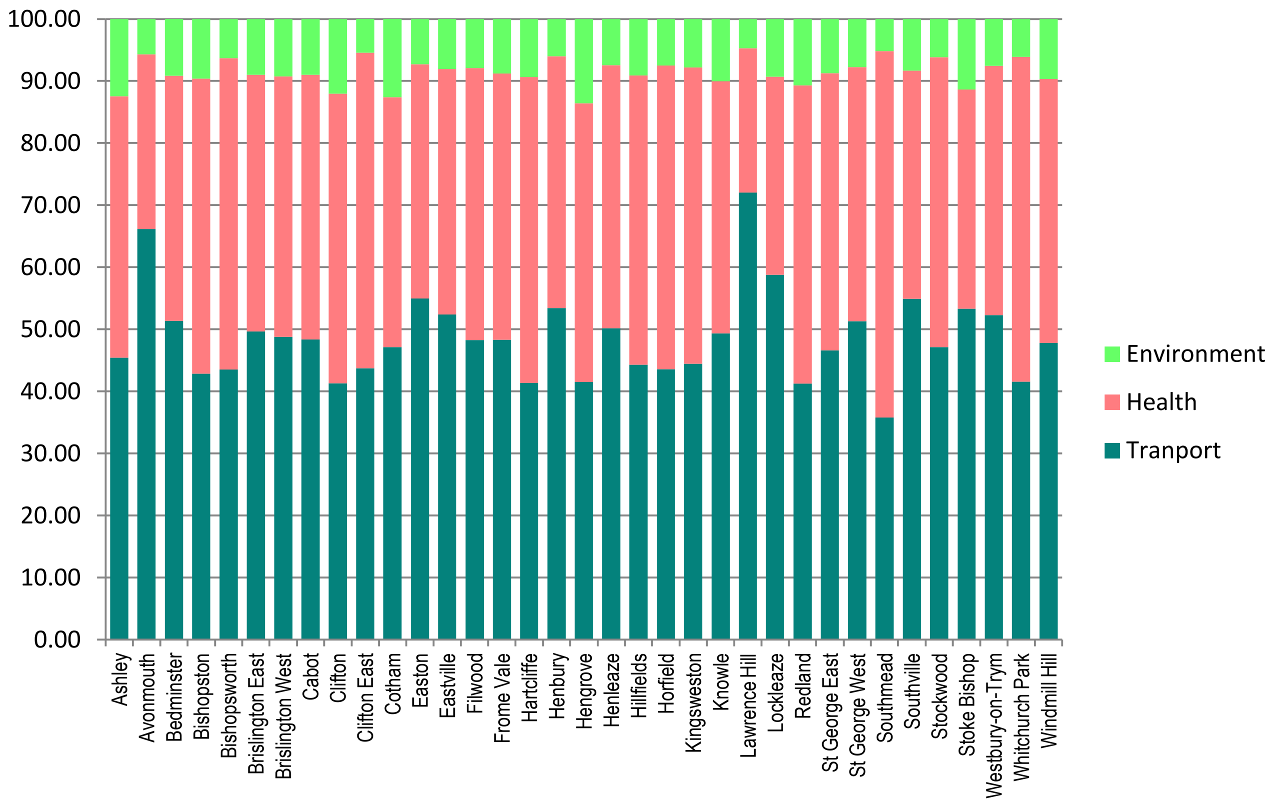


Figure 8. Percentages of perceptions per domain

There are differences between numbers of perceptions in each domain. While sample used for manual coding showed more perceptions about health, automated classification of Tweets derived more perceptions in transport domain. Moreover, there are significant differences in a number of perceptions between different wards in Bristol.

The difference in the number of perception is smaller between health and transport domains, while the number of perception in the environment domain is considerably lower. Wards can be distinguished by a number of perceptions in every domain, as shown in Figure 8. For example, wards Avonmouth, Lawrence Hill, Lockleaze have more perceptions in transport domain, while wards Bishopsworth, Southmead and Whitchurch Park in the health domain.

Even though there is fewer perception, the perceptions about the environment are more popular in some wards than another. For instance, Ashley, Cliffton, and Hengrove have more perceptions about environment compared to other wards.

For descriptive purposes, subjective QoL perceptions in Bristol per ward in 2013 are visualised in Figure 9. A number of perceptions directly reflects the amount of Twitter activity in the city. Therefore visualisation of a number of perceptions is not informative enough. To give meaningful insight, the number of Tweets per tweeting population (where a total number of Tweets for 2013 in Bristol stands as a proxy for tweeting population) is presented.

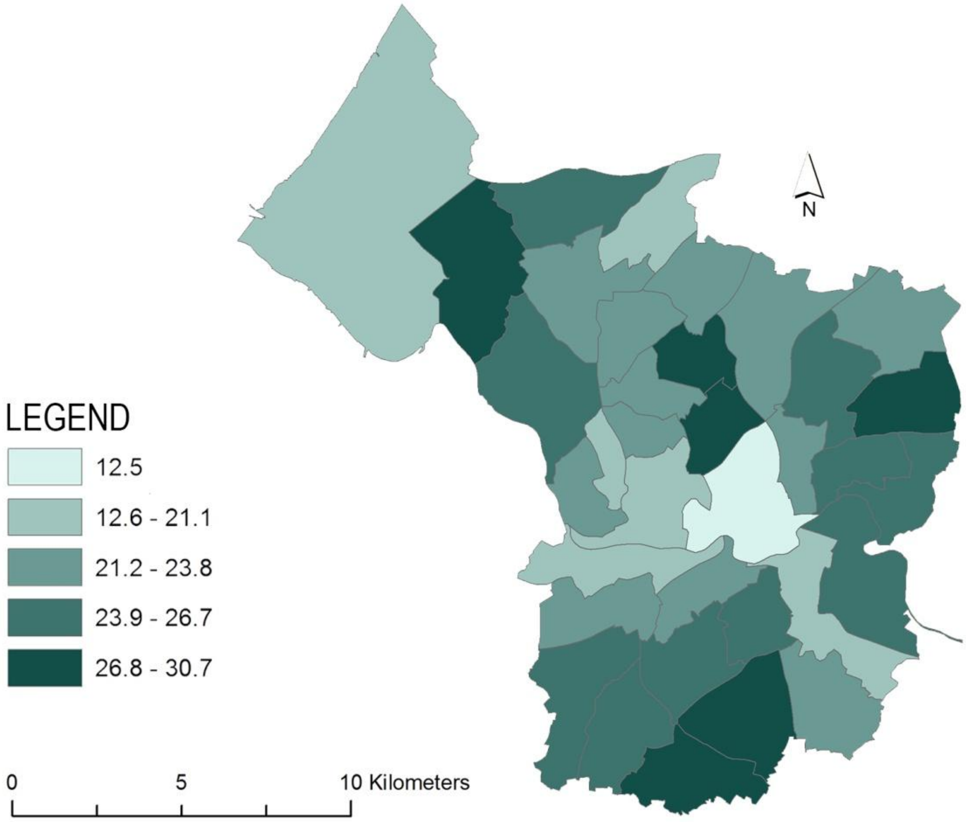


Figure 9. Subjective QoL perceptions per tweeting population

There is a variation in number of perceptions in Bristol wards. A lower number indicates more frequent tweeting about one of the subjective QoL domains. Lawrence Hill ward has the lowest value of 12.5, which means that every 12th Tweet from all the Tweets from 2013 in Bristol in this ward can be labelled as a subjective QoL perception. In the same time, six wards (Kingsweston, Bishopston, Ashley, Hillfields, Hengrove, and Whitchurch Park) have the highest values, where Tweets about subjective QoL are sparse; approximately every 25th perception can be categorised as QoL perception.

5.1.1. Perceptions about health

From all geolocated Tweets sent from within the administrative boundaries of Bristol in 2013, there are 25178 perceptions about health. In summary, 40.60% of all derived perceptions are about one of the topics derived in the health domain.

In the results from manual coding, two types of perception present in the data set are already distinguished: emotional reaction where people express feelings, and cognitive conclusions where people express opinions. Health domain is specific and different from transport and environment in the type of perceptions captured. Here, perceptions are more personal and indicate positive and negative feelings. Still, they are diverse and cover different topics.

Perceptions about health show that people tend to express their opinion about personal feelings of distress, pain, and discomfort, and different positive feelings as well.

> *"this headache couldnt get much worse"*

"soooo a sharp shooting pain in my left thigh has been pretty prominent for the last 36 hours"

"45 days without alcohol proud accomplishment"

Moreover, they comment about lifestyle and habits that can be classified as healthy and unhealthy. First, they comment about exercising, going to the gym, doing yoga, running, eating healthy food, and, on the opposite side, about the lack of physical activity, drinking alcohol, smoking, unhealthy eating and habits.

"exercise is so good"

"its been a week since i had a beer and i feel great"

"going start eating healthy and exercising tomorrow"

Although in a smaller scale, besides expressing feelings and commenting on lifestyle, perceptions where people express opinions about health issues in general are also present. For instance, people give opinions about the quality of healthcare, health services, hospitals, and relationship doctors and nurses have with their patients.

"local hospital the services are lets just say if theres someone critically sick theyd be dead at the waiting area"

"well aware information your guide to health wellbeing and community services in Bristol"

"good practice for the students at the dental hospital"

5.1.2. Perceptions about transport

From all geolocated Tweets sent from within the administrative boundaries of Bristol in 2013, there are 31307 perceptions about transport. In summary, 50,60% of all derived perceptions are about topic observed in the transport domain.

There are various types of perception within transport domain. The vast 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"

"the smells you find on public transport"

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"

"has some really funny tanlines from cycling in bristol sunshine"

"just been run over by a bike in town"

As it was already mentioned in the case study description, there is a strong encouragement for people to be engaged in the community development and the role online spaces have in connecting citizens with City Council and relevant organisations. As an example of this, there is a number of Tweets were people directly send Tweets to 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"

"bristolcouncil many traffic lights and bus lanes make driving to work a nightmare rarely pass a bus when I do it is empty"

"and now bristolcouncil think its fine to pull in on zig zag lines theyre on pedestrian crossings for a reason selfish idiots"

Moreover, transport domain also has a certain amount of perceptions expressing emotional reaction where people often feel some form of distress or excitement while using public transport, biking, walking on the streets, etc.

> *"on the train to clifton i feel so excited haha"*
>
> *"always get a headache when im on the bus"*
>
> *"omg this bus stinks and i feel sick as it is"*

5.1.3. Perceptions about environment

From all geolocated Tweets sent from within the administrative boundaries of Bristol in 2013, there are 5446 perceptions about environment. Compared to perceptions about health and transport, the number of perceptions is significantly lower. Only 8.8% of all derived perceptions are about environment. However, the lower number was expected due to the results of manual coding.

Here, perceptions are mostly about conditions of environment in general.

> *"better information planning amp management are needed to balance demands on coastal environment"*
>
> *"do we care too much about nature"*

Next, there are perceptions about sustainability and climate change.

> *"only domestic solutions are answer to climate change"*
>
> *"been brainstorming sustainability projects"*

Moreover, issues in the city with, for instance, garbage removal, recycling, unclean streets, and so on.

> *"desperate now to spend at least a day in a clean environment"*

5.2. Spatial distribution of QoL perceptions

QoL perceptions are visualised to identify and explore the differences in Bristol wards in perceptions on subjective QoL. This allows us to evaluate spatial dimension of QoL perceptions. Perceptions in health, transport and environment domains are visualised using odds ratio where the midpoint of a resulting ratio is 1. This means that the number of perceptions is closest to the one expected based on the tweeting population size. Values larger than 1 indicate that there are more perceptions than expected based on the total tweeting population, and vice versa.

Spatial distribution of perceptions about health and differences between Bristol wards are shown in Figure 10. Wards having values varying less from the mean represents wards with odds ratio closer to 1 and these words are located in the central area and southern outskirt of the city. Here, the amount of perceptions about health is the closest to one expected based on the tweeting population. More than half of the wards are classified into this class. Centre of the city and northern outskirt have slightly larger values.

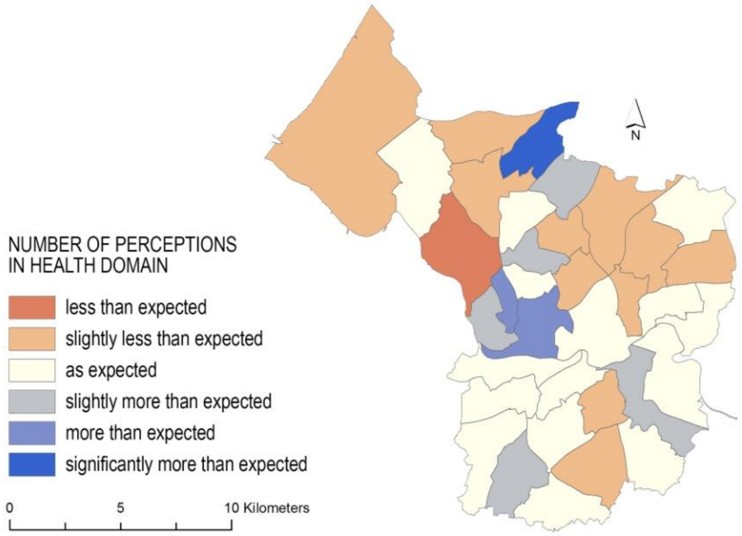


Figure 10. Spatial distribution of perceptions in health domain

However, number of perceptions smaller than expected form two larger clusters of wards in the northern part of the city. In Henbury ward the amount of perceptions is much higher than expected, compared to the other wards, while, on the other hand, in Stoke Bishop, there are much fewer perceptions than expected.

Spatial distribution of perceptions about transport and differences between Bristol wards are illustrated in Figure 11. Lawrence Hill ward stands out in the centre of the city with the highest value, as the ward having more perceptions than expected, compared to other wards

Wards, where people commented about transport as much as expected, are located in the central area and eastern part of the city. Wards with smaller numbers of perceptions than expected form one cluster of wards in the southern part of the city and several wards scattered around the city.

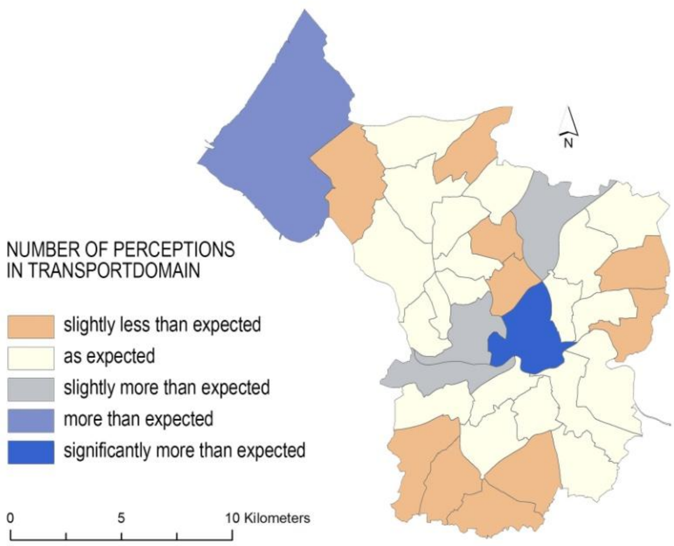


Figure 11. Spatial distribution of perceptions in transport domain

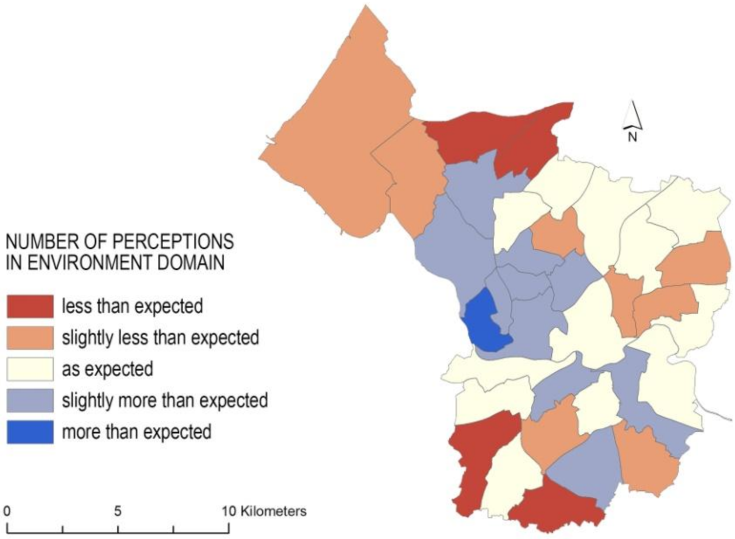


Figure 12. Spatial distribution of perceptions in environment domain

Spatial distribution of perceptions about environment and differences between wards is visualised in Figure 12. Wards, where people tweet about the environment as much as expected, are scattered around the city. Compared to health and transport domains, environment domain has more wards with less and more perceptions than expected. Wards with fewer perceptions than expected are located more in the northern and southern part of the city, while wards with more perceptions than expected are located more in the central part of the city. Clifton ward stands out with the highest value, showing the much higher number of perceptions than expected, compared to other wards. Although environment domain has significantly less perception captured in the analysis, this shows that the amount of perceptions in these ward is still higher.

5.3. Sentiments of perceptions

As a result of sentiment analysis, perceptions 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+). In general, there is a similar number of positive and neutral perceptions as shown in Table 10. Percentages of five sentiment groups for all perceptions for the city of Bristol are also presented in Figure 13, confirming the similar distribution of positive and negative Tweets. A detailed table with numbers of perceptions in different sentiment groups par ward is available in Appendix 8.

Table 10. Results from sentiment analysis summarised for the city of Bristol, 2013

	P+		P		Neutral		N		N+	
Domain	No	%	No	%	No	%	No	%	No	%
Health	2013	11.66	6208	35.97	1139	6.60	5998	34.75	1901	11.01
Transport	1967	10.37	7029	37.07	940	4.96	6386	33.68	2640	13.92
Environment	374	10.24	1448	39.65	199	5.45	1146	31.38	485	13.28
Bristol	4354	32.27	14685	112.69	2278	17.01	13530	99.81	5026	38.21

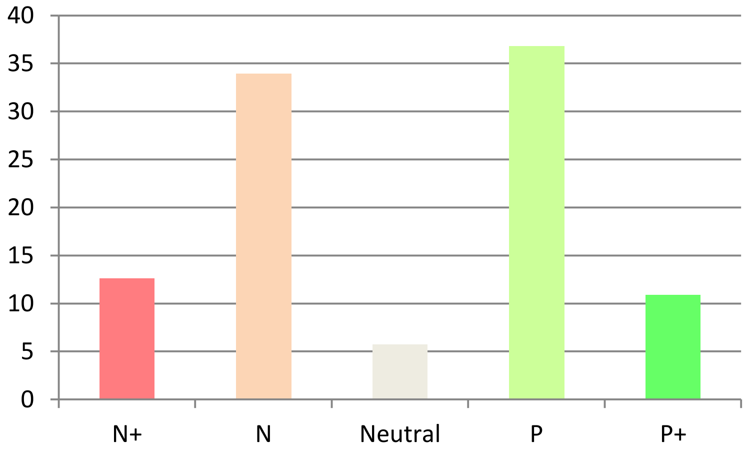


Figure 13. Percentages of perceptions in different sentiment groups for Bristol, 2013

Because of the similar number of positive and negative perceptions, instead of looking at each group separately, statistical analysis was done to detect the difference. One of the research hypotheses was set, saying that there is a significant difference between wards in average percentage of positive and negative Tweets for each domain.

A pared sample T-test was conducted to test the hypothesis. Results reject the null hypothesis, showing no significant difference between wards in positive and negative perceptions based on sentiment, with p-values for every domain p>0.05.[6]

5.3.1. Sentiment analysis of perceptions about health

Subjective QoL perceptions about health in Bristol are distributed in different sentiment categories. 68.7% of perceptions about health were given sentiment in the analysis, while 31.3% are characterized as perceptions without sentiment. Table 11 gives an example of Tweets distributed in five sentiment groups.

Table 11. Examples of Tweets in health domain distributed in sentiment groups

N+	"students health service are playing bye bye baby and it literally sounds like we are all at a funeral waiting to die"
N	"such a worst headache"
Neutral	"the pain that woman experiences while giving birth is similar to getting burned alive well then"
P	"loving working out got a real buzz for it again"
P+	"healthy eating is actually pretty fun that might sound odd but finding new and amazing recipes that are good for you is so inspiring"

[6] Paired sample T-test output is available in Appendix 5

5.3.2. Sentiment analysis of perceptions about transport

Subjective QoL perceptions about transport in Bristol are distributed in different sentiment categories. 60.57% of perceptions about health were given sentiment in the analysis, while 39.43% are characterized as perceptions without sentiment. Table 12 gives an example of Tweets distributed in five sentiment groups.

Table 12. Examples of Tweets in transport domain distributed in 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"

5.3.3. Sentiment analysis of perceptions about environment

Subjective QoL perceptions about environment in Bristol are distributed in different sentiment categories. 67.07% of perceptions about health were given sentiment in the analysis, while 32.93% are characterized as perceptions without sentiment. Table 13 gives an example of Tweets distributed in five sentiment groups.

Table 13. Examples of Tweets in environment domain distributed in sentiment groups

N+	"a criminal conspiracy at the environment agency"
N	"flood defence spending had increased when in fact its been cut"
Neutral	"i want to help the people in the south of England wales for being flooded and without power this christmas damn global warming"
P	"protect our natural resources for our health and future generations join me and take action"
P+	"your local councils environmental health team are quite handy for that"

5.4. Spatial distribution of positive and negative perceptions

There is no significant difference between wards with positive and negative perceptions (results of paired samples t-test), Nevertheless, we are interested to see how positive and negative perceptions are distributed in the city of Bristol. The assumption is made: if percentage of positive perceptions in wards is 5% bigger than percentage of negative perceptions, ward is classified as positive, and vice versa.

For the purpose of showing spatial distribution, we have wards with highest positive and highest negative perceptions visualised for each domain. Bristol wards with the highest percentages of positive and negative perceptions are shown.

The resulting visualisations and description are presented for every domain. We are observing these wards as the places where people are highly satisfied or highly dissatisfied, based on the perceptions we captured from Twitter and assumptions we made.

5.4.1. Spatial distribution of positive and negative perceptions in health domain

Spatial distribution of positive and negative perceptions about health is visualised in Figure 14 to show wards where people are more satisfied with a health condition and vice versa. Based on the assumption taken for calculation, there are 12 wards, where four wards have more positive than negative perceptions, and 8 have more positive than negative. Based on these perceptions, health conditions are better in the central and northern part of the city.

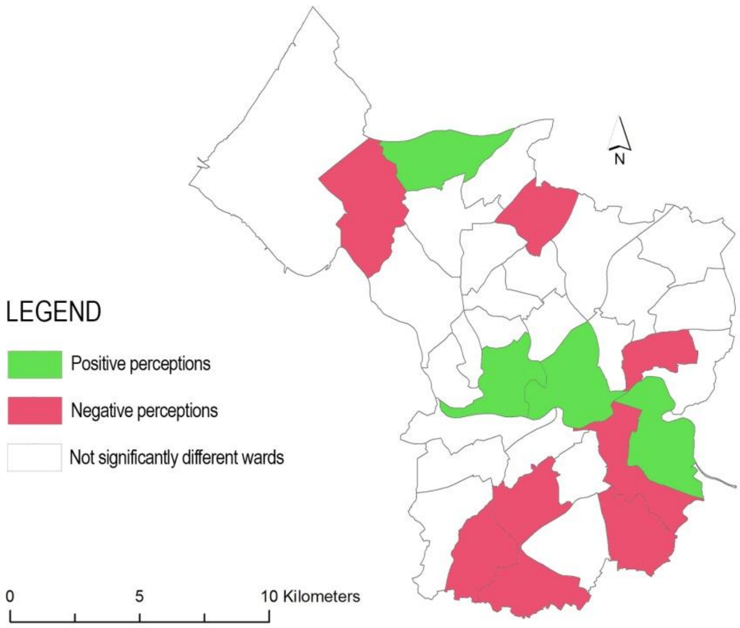


Figure 14. Spatial distribution of positive and negative perceptions in health domain

The highest percentages of satisfied people are in Henbury, Cabot, Lawrence Hill and Brislington East. On the other hand, there are twice as many wards with highly dissatisfied people. These are wards Kingsweston, Horfield, St. George West, Brislington West, Stockwood, Filwood, Hartcliffe and Whitchurch Park.

5.4.2. Spatial distribution of positive and negative perceptions in transport domain

Spatial distribution of positive and negative perceptions about transport is visualised in Figure 15 to show wards where people are more satisfied with transport condition and vice versa. 11 wards in transport domain have differences between positive and negative perceptions, 3 with more positive, and 8 with more negative perceptions. Looking at the highest percentages of satisfied people, 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. Similar to health domain, where a number of wards with more negative than positive perceptions is larger than the number of wards with more negative perceptions, transport domain has more highly dissatisfied wards. These are Kinsweston, Horfield, Henleaze, Easton, St. George West, Stockwood, Hengrove and Whitchurch Park.

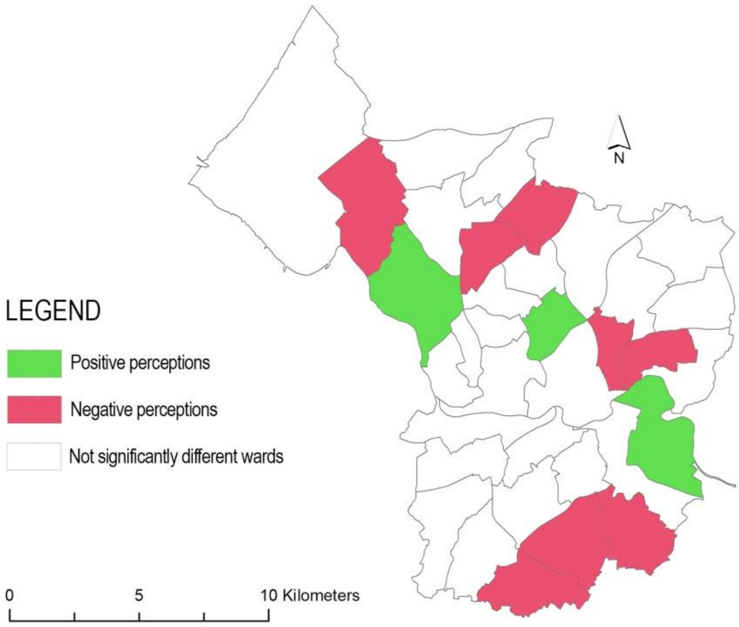


Figure 15. Spatial distribution of positive and negative perceptions in transport domain

5.4.3. Spatial distribution of positive and negative perceptions in environment domain

Spatial distribution of positive and negative perceptions about the environment is visualised in Figure 16 to show wards where perceptions are more positive than negative and vice versa. Comparing three observed domains, the environment is the only one where the number of wards with highest percentages of satisfied people is significantly larger than a number of wards with highest percentages of dissatisfied people.

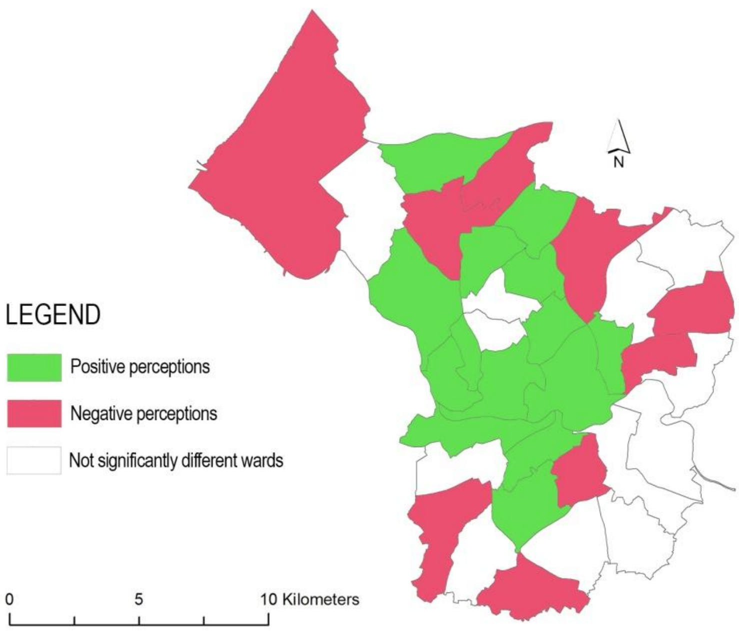


Figure 16. Spatial distribution of positive and negative perceptions in environment domain

Moreover, while transport and health have similar patterns of spatial distribution, the environment is slightly different. In total, there is a significant difference in 24 wards, 15 wards with more positive, and 9 with more negative perceptions. Wards with the highest percentage of positive perceptions are located more in the central parts of the city, while wards whit more negative perceptions are clustered in several parts of the city.

While there are no wards having highly positive perceptions for all three domains, wards St. George West and Whitchurch Park have highly negative perceptions in every domain.

5.5. Comparison between derived and measured subjective QoL

Subjective perceptions about QoL derived from all geolocated Tweets sent from within the administrative boundaries of Bristol in 2013 are compared to results from official QoL survey in Bristol in 2013. Three domains, health, transport, and environment are compared to subjective QoL indicators covering similar topics. Hence, people's perceptions about health are compared to the percentage of respondents satisfied with health services; people's perceptions about transport are compared to the percentage of respondents satisfied with bus services, and people's perceptions about the environment are compared to the percentage of respondents satisfied with quality of parks and green spaces.

For the purpose of analysing statistically significant similarity between results, null hypothesis is proposed saying that the two results are the same. Paired sample t-test[7] show that proposed null hypothesis can be rejected. This means that results from this study, when compared to results derived from the official QoL survey, have no significant similarities on the city level.

However, even though statistically significant similarity is not captured, results from three domains, together with proxies taken from the official survey are mapped to check spatial similarities in wards level. For the results of this study, for each domain, wards with the most positive perceptions are mapped. For the proxies, wards with most percentages of positive responses are mapped. These two maps are overlaid for every domain. In health and transport domain, there is no similarity between wards. This corresponds with the results of a statistical analysis showing that the two resulting variables are not correlated and are statistically different.

[7] Pared sample T-test output is available in Appendix 6

Differences between spatial distributions of positive opinions are shown in Figure 17. In the official QoL survey, people located in wards in the northern part of the city are more satisfied with conditions of health

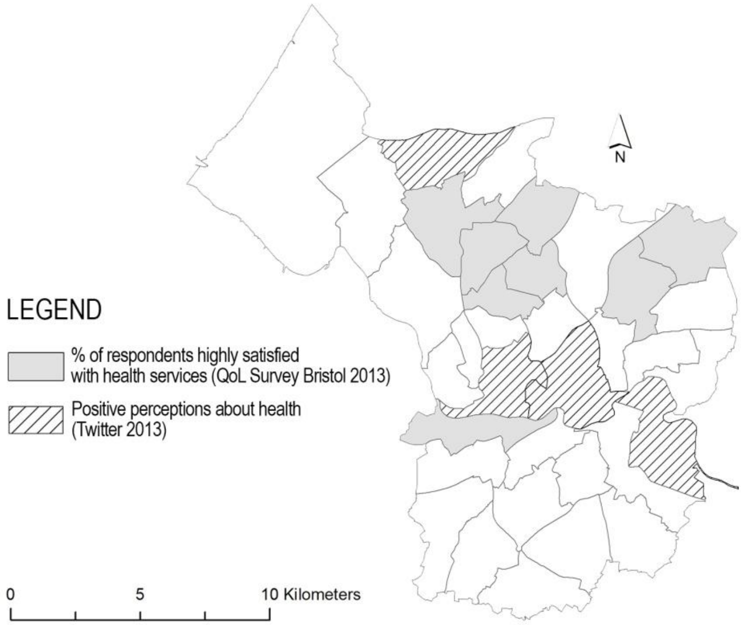


Figure 17. Comparison between derived and measured subjective QoL in health domain (source: own analysis based on data from Bristol City Council, 2015)

services, while positive perceptions about health derived from Twitter are located more in the central part of the city.

Similar to the health domain, when comparing opinions about transport, there are no similarities between wards. Figure 18 shows that positive perceptions about transport derived from Twitter are located in wards Stoke Bishop, Ashley and Brislington East, while respondents satisfied with bus services in official

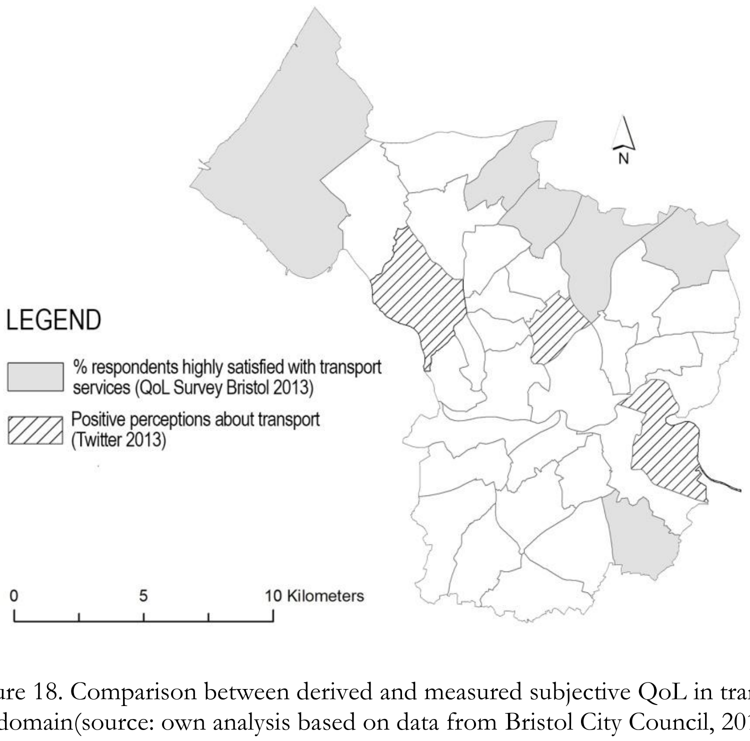


Figure 18. Comparison between derived and measured subjective QoL in transport domain(source: own analysis based on data from Bristol City Council, 2015)

QoL survey are mostly distributed in the northern part of the city.

Figure 19 show similarities and differences between two approaches in environment domain. Opposite from health and transport domain, where there is no similarity in spatial distribution, in environment domain there is a match in two results in three wards Henleaze, Clifton and Clifton East.

Results from both approaches are mostly distributed in central and northern part of the city.

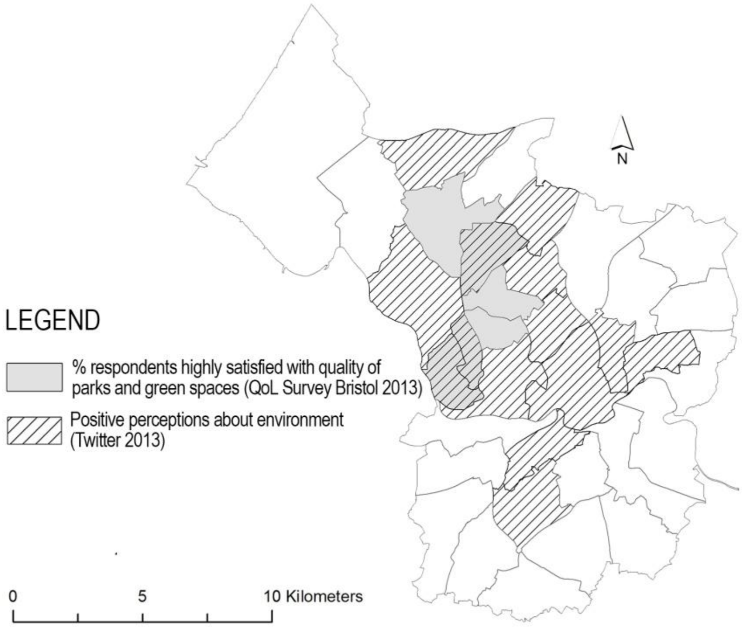


Figure 19. Comparison between derived and measured subjective QoL in environment domain
(source: own analysis based on data from Bristol City Council, 2015)

6. DISCUSSION

This section provides a discussion of the research findings. First, results from qualitative part of the analysis are discussed. This is followed by discussion on perceptions captured in health, transport, and environment domain and spatial distribution of subjective QoL perceptions. The sentiment of perceptions is discussed and how positive perceptions are spatially distributed. Moreover, comparison between subjective QoL perceptions derived from Twitter and perceptions from official QoL survey in Bristol, similarities, and differences, are elaborated. The relevance of understanding advantages and disadvantages of using social media in quality of life research is argued.

6.1. Deriving subjective QoL domains using Twitter data

Social media have showed to be a relevant source of data, applicable in capturing subjective quality of life (QoL) perceptions. Qualitative analysis of a random sample of Tweets successfully recognised people's perceptions about QoL and derived domains that are suitable to measure with Twitter data.

First, findings from qualitative analysis clearly suggested that QoL can be analysed using Twitter data and offered a general idea about the nature of messages indicating opinions about QoL. Possibilities to gain insights from the data, and still strengthen the process by effective use of theoretical knowledge are shown. While Twitter messages revealed QoL perceptions, QoL theory helped in classifying these perceptions into domains.

Captured perceptions and domains can easily be placed in theoretical frame of QoL studies. 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 the present study. Similarly to studies using traditional methods for collecting and analysing subjective QoL (for example Bramston, Pretty, & Chipuer, 2002; Eby, Kitchen, & Williams, 2012; Ibrahim & Chung, 2003) domains of general satisfaction, health, transport, city environment, relations with friends and family, natural environment, social interaction, finances and employment, and safety are derived.

Undoubtedly, every QoL perception derived from Twitter is subjective and personal. However, 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, for instance in psychology, health science, linguistic, happiness studies, and so on. The recognition of the second type of perceptions, where people evaluate conditions of the environment surrounding them, points 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.

Due to the difference in numbers of perceptions in derived domains, it is concluded that certain quality of life domains are more covered than another. Domains of general satisfaction, health, transport, community and social life have significantly more perceptions compared to other domains. Still, as opposed to other studies that used social media data for evaluating subjective QoL, where only perceived happiness (Dodds et al., 2011) or subjective well-being (Bibo et al., 2014) were observed, the possibility for deeper analysis of people's perceptions by introducing the variation of domains is captured.

The conclusion stated in the study of Nguyen et al. (2016), who successfully covered domains of overall happiness, healthy food, and physical activities as indicators for a difference in QoL, support the findings of this study. They concluded that social media could be used to provide a better understanding of well-being of communities. Certainly, differences in methodological approach contribute to differences between present study, and that of Nguyen et al. (2016). Still, findings are similar and support one another.

The benefit of including manual coding of a sample of Tweets is in having more transparent approach, instead of capturing perceptions only through black boxed automated classification. This part of the analysis gave an overall idea about the type of perceptions and domains that can be observed.

6.2. People's perceptions about QoL in Bristol

Content classification provided insight about perceptions in health, transport, and environment domains. Moreover, it gave an idea about variations in topics covered by perceptions. The analysis revealed differences in quality of life between wards in Bristol.

Results of content classification are spatially connected to Bristol wards and associated with wards characteristics regarding social issues and planning opportunities. Costanza et al. (2007), Moro et al. (2008), and Pacione (2003b) advocate the use of subjective measurements and self-reported perception of life satisfaction in a certain location to evaluate QoL and capture differences between the neighbourhoods. Moreover, they argue that objective measurements can be used to assess the objective conditions for residents to improve their QoL, but cannot directly measure the phenomena. Even though there is no correlation between positive and negative perceptions captured in different domains and IMD[8], we would like to build on this by having the result measuring the phenomena placed into local context and observed through the objective conditions. This will strengthen the meaning of the resulting subjective perceptions and give a better idea about reasons behind the existence of certain perceptions. For the purpose of this, we are using information about deprivation hotspots in Bristol and objective characteristics derived from index of multiple deprivation (IMD) and Census data introduced in previous chapters.

The first interesting 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 is labelled as one of the subjective QoL perceptions, 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 was and still is in the group of the most deprived 10% in England (Bristol City Council, 2015a).

Next, when looking at the total amount of perceptions, more perceptions are derived in health and transport domain, compared to the environment domain. While transport (50,6%) and health (40,6%) count for more than 90% of perceptions, only 8,8% of perceptions is classified in environment domain. This is consistent with findings from qualitative part of the analysis, where perceptions were mostly derived in health and transport domains, while the environment was one of the less covered domains.

Moreover, when looking at variations between perceptions in each domain, considerable difference in types of perceptions can be seen. Due to this, perceptions in every domain can be classified into subtypes, based on the main topics they cover. In health domain, when examining examples of Tweets, at least four types of perceptions can be distinguished: emotional reactions, feeling of physical pain, perceptions about healthy/unhealthy lifestyle, and opinions about the health system. While the first three subtypes are more about people's emotions, feelings of distress and lifestyle, opinions about health system have more direct implications for planning and decision making. These perceptions give insight into functioning of the health system and opportunity to address issues and work on improving conditions. Similarly, in transport

[8] Correlation analysis results is available in Appendix 7

domain at least three types of perceptions are captured: quality of public transport, quality of streets, and opinions about cycling. Additionally, environment domain includes perceptions that can be grouped into conditions of the environment, global warming and climate change, and perceptions with more local context, including garbage disposal and recycling. This recognition of various topics within QoL domains points to possibility for future researchers to dig deeper into this subject and cover discovered topics individually to measure subjective QoL of people in more details.

Additionally, Tweets show a significant level of correspondence between residents using social media, and Bristol City Council. This coincides with attempts from Bristol City Council's side to encourage people to use social media and online tools to contribute to the development of local community and participate in decision-making processes.

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.

The final analysis measured sentiment of captured QoL perceptions. The analysis revealed types of perceptions based on sentiment, and classified these perceptions into five sentiment categories. In general, across three domains, Tweets are similarly positive and negative in sentiment and it necessary to address both positive and negative perceptions 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 is discovered that prevalence of positive Tweets is the highest in environment domain when looking only into the highest positive and the highest negative perceptions. Based on this, it can be concluded that people Tweet less about the environment, but more positively. On the other hand, situation in health and transport domains is the opposite, with a greater presence of wards with highly negative perceptions. Map with IMD scores for Bristol wards is overlaid with pie charts illustrating the percentages of positive, neutral and negative perceptions in health, transport and environment domain.[9]

Spatial distribution of positive and negative perceptions about health indicates concurrence with some of the most deprived and affluent wards in Bristol. Map showing the IMD overlaid with pie chart showing the percentages of positive, neutral and negative perceptions in health domain is available in Appendix 9. While positive perceptions are mostly located in close vicinity to some of the most affluent parts of the city, negative ones are mostly matching wards with higher levels of deprivation. A contrasting finding can be seen in Lawrence Hill ward. While it represents one the most deprived areas in Bristol (Bristol City Council, 2015a) and wards with the highest percentage of people with bad health according to Bristol Census data (Bristol City Council, 2015), based on perceptions derived from Twitter, it is one of the wards with highly positive perceptions. If we go back a step, we might see that this was also the ward with the largest number of perceptions per tweeting population. In the scope of this study, we did not capture the reasons behind this inconsistency, and we could only speculate those reasons. However, this goes in line with findings from traditional QoL studies, where subjective perceptions not always matched the objective conditions of the area.

Positive and negative perceptions in transport domain have more similarities with characteristics of wards based on the level of deprivation. Map showing the IMD overlaid with pie charts showing the percentages of positive, neutral and negative perceptions in transport domain is available in Appendix 10. First, there are only three wards with highly positive perceptions, and they are located in central part of the city. Wards with highly negative perceptions match with wards with a higher level of deprivation. Two

[9] Produced maps are available in Appendices 9, 10 and 11

contrasting findings are captured in transport domain as well. Although ward Henleaze is characterized as one of the more affluent wards, highly negative perceptions are captured through the analysis of Twitter messages. On the other hand, ward Ashley, with part of the city called St Pauls being categorised in one of the most deprived 10% in England, have highly positive QoL perceptions derived from Twitter.

Environment domain has noticeably different patterns of wards with highly positive and highly negative perceptions compared to health and transport domains. Map showing the IMD overlaid with pie charts showing the percentages of positive, neutral and negative perceptions in environment domain is available in Appendix 11. We already mentioned significantly larger number of wards with highly positive perceptions. Still, in the environment domain, there are wards with highly negative perceptions, and they are located in the north-east part of the city, where levels of deprivation are higher. Wards with highly positive perceptions are distributed mostly in central and southern parts of the city. Here, we can also comment on captured mismatch. Wards in the south part of the city with highly positive perceptions are at the same time the ones with noticeable deprivation hotspots.

Bristol is England's greenest city (Bristol City Council, 2015b) and winner of The European Green Capital Award for 2015. Moreover Bristol has series of strategies covering topics such as environment, green energies, and climate change. Citizen's participation is reviewed in official QoL survey, where respondents show growing interest in issues emerging in their living environment. This can explain interest and diverse opinions about issues in environment domain captured from Twitter.

In general, based on the data used for this research, 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 correspond to low and high levels of deprivation, based on IMD. On the other hand, there are wards with the mismatch between the level of satisfaction based on positive and negative perceptions derived from Twitter and objective conditions in Bristol wards. These kinds of contrasting measurements are often in QoL research, especially when trying to place subjective perceptions into objective conditions.

In conclusion, sentiment classification is a special field in studying language, and there has been a big amount of research on this topic. Most of the studied on Twitter and sentiment in Twitter messages are done using machine learning and include the research on how algorithms can be developed and trained to recognise sentiment. In this study, the approach is simplified and adapted to the needs of social scientist. Tool used for the analysis purposes is called Cloud Meaning[10]. This tool is developed for sentiment analysis of text and can be used as an Excel add-in. It gave us information about a number of Tweets in each of five sentiment group and the possibility to truly capture differences between levels of satisfaction within observed domains and spatial distribution of positive and negative sentiment. Moreover, as notices 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 classifier useful for QoL research using Twitter data which justify our selection of the method used.

6.3. Reflection on comparison between derived and measured subjective QoL

It is relevant to recognise what are the possibilities in combining approaches in assessing subjective QoL to improve the planning and decision-making process. Following conceptual framework outlined in Chapter 2, results derived in this study are compared to the results derived from an official QoL survey done in Bristol in 2013.

[10] Appendix 4

Statistical analysis is performed to check the correlation between two results. Moreover, spatial similarities and differences are presented by visualising positive perceptions derived in both approaches.

First, statistically, there is no correlation between results derived in two studies. Moreover, when visualising positive perceptions from both studies by overlaying the maps, similarities are only captured in the environment domain. Yes, several wards in environment domain indeed show similarities with findings from the official QoL survey in Bristol, but not enough to say that the overall results from this study reflect results from the official QoL survey.

Moreover, there is one more setting where differences and similarities can be observed. It includes differences and similarities in the sense of the coverage of questions asked in the survey and types of perceptions captured from Twitter. For example, according to 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. On the other hand, perceptions derived from Twitter cover similar topics but are mostly oriented to quality and condition of buses, bus frequencies, congestion, how people feel inside the bus, and so on. Moreover, perceptions from Twitter cover a wider range of topics, compared to the proxy used for the comparison.

Furthermore, difference can be explained by the profile of respondents. Although demographics of Twitter users in the city of Bristol are unknown, characteristics of Twitter users in general can be observed. Moreover, demographics of Bristol QoL survey respondents can be explored. Here, only age of respondents is taken into account. According to Bristol QoL survey report (Bristol City Council, 2014), proportionally less young people responded in the 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 number of users is the highest in the 25-45 year-old age group (Ciuccarelli, Lupi, & Simeone, 2014). This difference in age of QoL survey respondents and Twitter users strengthen the suggestion of using data from social media as a 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 interesting 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 this is not directly connected with social media data used for the purpose of the present study, the thought is similar and we can build our conclusions on these ideas. Even though we captured only a few similarities when compared two studies, this lack of correlation between results can also be observed as a significant result and new knowledge. There is a possibility of using messages unintentionally produced by people to produce greater knowledge about issues in specific areas across various domains and fill the gaps in analysing cities.

In summary, several main similarities and differences in the approach are underlined. The main differences are in the size of the sample and methodology used for the analysis. Official QoL survey in Bristol is based on the smaller sample, while analysis of dataset we used covers larger population. Moreover, in this study insights are obtained from data itself, rather than theory, as it is done in more traditional approaches, counting QoL survey done in Bristol. Methods used in this research are adapted to the specific needs of social media data, compared to the traditional one. Moreover, official QoL survey in Bristol is done per ward, where households are interviewed, so we know for sure that location of perception is location where people live. With Twitter data, location problem is much more emphasised. According to Li et al. (2013)

geo tag on certain Tweets point to the presence of Twitter users in these sites. Moreover, they 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.

6.4. Reflection on usability of social media in QoL research

Social media data provide an opportunity for innovative research and new ways to discover patterns in QoL of people. Traditionally trained scientists are still skeptical about the usability of social media due to various issues associated with the analysis of social media data. However, other researchers recognise the potential and richness of these data. Moreover, comparing with traditional methods for analysing subjective QoL, harvesting and evaluating data from the social media offers a contemporary, fast and cost effective approach (Schnitzler, Davies, Ross, & Harris, 2016)

The main contribution of this study is recognition of possibilities Twitter data have in QoL research. Contemporary urban planning practice is embracing the positive characteristics of data derived from social media, and our study is only the small contribution towards better understanding of connections between location, people, and messages shared in online setting. In general, involvement of the community can be observed as a collaborative way for producing knowledge, facilitating participatory planning practice and joint decision making (Natarajan, 2015). Using city of Bristol exemplifies this claim. The cooperation between City Council and residents is at an exceptional level, and there are services created specifically for this kind of collaboration. They offer the opportunity to jointly make decisions about their neighbourhoods and take actions based on those decisions together. Likewise, social media data offer a novel and unobtrusive way of capturing people's perceptions for the purpose of 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 feeling about a system and possibility to capture general ideas about the functioning of this system. The availability and spatiality are key features of Twitter messages. The connection between physical and digital world is reflected through the spatiality of data and 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 the fact that, even though the Tweets are geo-tagged and connected with a specific point in space, it does not mean that opinion expressed is about that location. People can comment about public transport after they leave the bus or hospital service after they come back home. Nguyen et al. (2016) address this as "migration bias" (p. 86) and 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 big data are having 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 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 exceptional 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, Wadhwani, & More, 2016), and so on. The advantage of this research is the attempt to combine different techniques adapted for simple extraction of opinions from

Twitter data, and exploring how results of such study can be efficiently placed in a planning context and effectively used to improve the decision-making process and enhance quality of life of residents.

6.5. Limitations

Using social media data in scientific research can be challenging. This research has its limitation and, although some of them are briefly mentioned in previous sections, the ones that can be addressed and improved in future studies are presented.

Simple text classification is used, avoiding machine learning and advanced natural language processing algorithms, which is enough to provide insight for an urban planner or social scientist. When manually inspecting messages distributed in different domains using filtering through dictionaries, a certain level of error is captured. This happened because filtering with only one word and simple phrases are used. Therefore, there is a place for improvement. There are possibilities to classify text in more sophisticated ways using n-gram tokenization or carefully designed topic modeling (Bird, Klein, & Loper, 2009).

Messages posted on social media represent a biased sample. People using Twitter are not a representative sample of the population. Moreover, internet usage is very uneven among countries, within countries, and within cities (Warf, 2013). Warf also writes about underrepresented groups, such as children and elderly. In some countries gender is also relevant, and income plays an important role as well. There is also an issue of users representing themselves in a different way on social media than in everyday life. However, the same issue appears when dealing with surveys and interview.

Although Tweets used are geo tagged, a location issue when working with social media data is emphasised. It is known that person sending a message is present at a certain location. However, it still unknown what kind of function that location has (residence, work, leisure, travel, and so on). People can comment about certain thing, issue or location characteristic when already being in different location.

7. CONCLUSION

The main objective of this study was to examine the possibility of extracting people's opinions about subjective QoL from Twitter and determine whether Twitter data can be used as proxies for QoL survey data. A framework is developed to serve as a guideline for analysing social media data and help placing developed approach in the existing QoL theory.

A case study area is chosen based on the certain criteria, but particularly having in mind local characteristics in order to have result that can easily be placed in local context.

Methodological approach is designed and steps proposed for analysing data derived from Twitter for the purpose of assessing QoL, using the city of Bristol as a case study area. This study showed the relevance of using mixed method approach, with qualitative analysis producing input for quantitative part, and together generating meaningful results. Qualitative part revealed the variety of QoL domains that can be observed. As a result, health, transport and environment domain are chosen to be further observed. 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 opinions, there is a difference in QoL between Bristol neighbourhoods. And, the third one is that, while Twitter messages can be used to complement QoL surveys, they cannot be used as proxies.

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

Overall, we do not propose for social media to replace traditional surveys and interviews. We point to the fact that results derived from this and similar studies could be used to capture something that is maybe overseen in official surveys and provide additional knowledge. For example, Tweets captured in transport domain showed that people tweet about several additional issues compared to topics covered in survey questions, such as how people feel while using public transport or issues cyclist have. In addition, results could point to areas that need attention and provide starting point for further investigation.

7.1. Recommendations for future studies

Our study provided a strong argument to the usability of social media data in QoL research. However, it also left some place for the improvement and possibility to try different methodological variations to confirm or challenge the results.

A large group of people has limited access to the internet or no access at all. So the bias is towards people who use the internet. We have the tweeting population in the area, but we do not know the characteristics of that population. Some conclusion can be derived from the comparison with, for example, official statistics, but still, the assumption only can be made. However, there are studies proposing methods to derive population characteristics of Twitter users (Longley & Adnan, 2016). This offers the possibility to evaluate QoL of residents based on Twitter data with the addition of demographics also derived from Twitter.

We analysed only three domains, health, transport and environment due to the exploratory nature of the study and time constraints. However, we already captured some QoL domains that can be addressed in the future research. In the same time, we proposed possible sub-domains that can serve to develop indicators.

When analysing Twitter data for the purpose of QoL research, we recommend a multidisciplinary approach. Having a team where urban planner, linguist, computer scientist, sociologist, and so forth work together in collecting and analysing social media data offers endless opportunities in producing meaningful results and gathering new knowledge.

Academic research has interests in social media and big data. It is necessary to understand the opportunities in using social media as a tool to help improve the process of planning and decision making for better cities and future of urban areas. Here, two directions should be followed. First, to fit social media and data derived from social media into a theoretical framework, provide definitions and define roles. Second is to understand the practical part, how to efficiently prepare, analyse and use data derived from social media.

LIST OF REFERENCES

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. *Association for Computational Linguistics*, 30–38. Retrieved from http://dl.acm.org/citation.cfm?id=2021109.2021114%5Cnpapers3://publication/uuid/83CA53FE-43D1-4BD5-BCF2-D55B82CF0F99

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. http://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. http://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. http://doi.org/10.1007/s12061-015-9154-2

Ballas, D. (2013). What makes a "happy city"? *Cities*, *32*, S39–S50. http://doi.org/10.1016/j.cities.2013.04.009

Bibo, H., Lin, L., Rui, G., Ang, L., & Tingshao, Z. (2014). Sensing subjective well-being from social media. In D. Ślezak, G. Schaefer, So. T. Vuong, & K. Yoo-Sung (Eds.), *Active Media Technology* (Vol. 8610 LNCS, pp. 324–335). Warsaw: Springer International Publishing.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python.* (1st ed.). Cambridge: O'Reilly Media, Inc.

Bonn, G., & Tafarodi, R. W. (2013). Visualizing the Good Life: A Cross-Cultural Analysis. *Journal of Happiness Studies*, *14*(6), 1839–1856. http://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. http://doi.org/10.1023/A:1019617921082

Bristol City Council. (2014). Quality of life in Bristol: Survey results 2013, 82. Retrieved from http://www.bristol.gov.uk/sites/default/files/documents/council_and_democracy/consultations/qol2014final.pdf

Bristol City Council. (2015). *bristol.gov.uk*. Retrieved October 10, 2016, from https://www.bristol.gov.uk

Bristol City Council. (2015a). *Deprivation in Bristol 2015*. Bristol. Retrieved from https://www.bristol.gov.uk/statistics-census-information

Bristol City Council. (2015b). *State of Bristol*. Bristol. Retrieved from https://www.bristol.gov.uk/statistics-census-information

Bristol City Council. (2016). *Bristol City Council services*. Retrieved January 1, 2016, from https://www.bristol.gov.uk

Bryman, A. (2015). *Social research methods* (4th ed., Vol. 1). New York: Oxford University Press Inc.

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*, (1). http://doi.org/10.1007/s11482-016-9472-9

Chae, B. (Kevin). (2015). Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, *165*, 247–259. http://doi.org/10.1016/j.ijpe.2014.12.037

Chatfield, A. T., Scholl, H. J. Jochen, & 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. http://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. http://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 International Publishing.

Ciuccarelli, P., Lupi, G., & Simeone, L. (2014b). *Visualizing the Data City*. In B. Pernici, S. Della Torre, B. M. Colosimo, T. Faravelli, R. Paolucci, & S. Piardi (Eds.) (1st ed.). Milan: Springer International

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. http://doi.org/10.1111/j.1083-6101.2007.00393.x

Costanza, R., Fisher, B., Ali, S., Beer, C., Bond, L., Boumans, R., … Snapp, R. (2007). Quality of life: An approach integrating opportunities, human needs, and subjective well-being. *Ecological Economics*, *61*(2–3), 267–276. http://doi.org/10.1016/j.ecolecon.2006.02.023

Croitoru, A., Crooks, A., Radzikowski, J., & Stefanidis, A. (2013). Geosocial gauge: a system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science*, *27*(12), 2483–2508. http://doi.org/10.1080/13658816.2013.825724

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. http://doi.org/10.1007/s11205-014-0646-2

Das, D. (2008). Urban quality of life: A case study of Guwahati. *Social Indicators Research*, *88*(2), 297–310. http://doi.org/10.1007/s11205-007-9191-6

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. http://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. http://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. http://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. http://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. http://doi.org/10.1007/s11205-012-0067-z

Floating Sheep. (2016). *DOLLY*. Retrieved October 10, 2016, from http://www.floatingsheep.org/p/dolly.html

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221. http://doi.org/10.1007/s10708-007-9111-y

Gov.UK. (2016). *Office for national statistics.* Retrieved September 21, 2016, from http://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/measuringnationalwellbeing/2015-09-23

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. http://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. http://doi.org/10.1177/01939459922044153

Haslauer, E., Delmelle, E. C., Keul, A., Blaschke, T., & Prinz, T. (2014). Comparing Subjective and

Objective Quality of Life Criteria: A Case Study of Green Space and Public Transport in Vienna, Austria. *Social Indicators Research*, *124*(3), 911–927. http://doi.org/10.1007/s11205-014-0810-8

Hsieh, H. F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. http://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. http://doi.org/10.1023/A:1021305620042

Jansen, J. B., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter Power:Tweets as ElectronicWord of Mouth. *Journal of the American Society for Information Science and Technology*, *60*(11), 2169–2188. http://doi.org/10.1002/asi.21149

Kapteyn, A., Lee, J., Tassot, C., Vonkova, H., & Zamarro, G. (2015). Dimensions of Subjective Well-Being. *Social Indicators Research*, *123*(3), 625–660. http://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. http://doi.org/10.1007/s10708-013-9516-8

Larsson, J., Söderlind, A., Kim, H., Klaesson, J., & Palmberg, J. (2016). *European Handbook of Crowdsourced information*. In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, & R. Purves (Eds.). London: Ubiquity Press Ltd.

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. http://doi.org/10.1080/15230406.2013.777139

Lieske, S. N., Martin, K., Grant, B., & Baldwin, C. (2015). Planning Support Systems and Smart Cities. In S. Geertman, J. J. Ferreira, R. Goodspeed, & J. Stillwell (Eds.), *Planning Support Systems and Smart Cities* (pp. 41–60). Springer International Publishing.

Lo, S. L., Chiong, R., & Cornforth, D. (2016). Ranking of high-value social audiences on Twitter. *Decision Support Systems*, *85*, 34–48. http://doi.org/10.1016/j.dss.2016.02.010

Longley, P. A., & Adnan, M. (2016). Geo-temporal Twitter demographics. *International Journal of Geographical Information Science*, *30*(2), 369–389. http://doi.org/10.1080/13658816.2015.1089441

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. http://doi.org/10.1016/S0169-2046(02)00239-6

Marans, R. W. (2015). Quality of urban life & environmental sustainability studies: Future linkage opportunities. *Habitat International*, *45*(P1), 47–52. http://doi.org/10.1016/j.habitatint.2014.06.019

McCrea, R., Marans, R., Stimson, R., & Western, J. (2011). *Investigating Quality of Urban Life*. In R. W. Marans & R. J. Stimson (Eds.) *Investigating Quality of Urban Life* (Vol. 45). London: Springer Dordrecht Heidelberg.

Mcmahon, S. K. (2002). The development of quality of life indicators - a case study from the City of Bristol, UK. *Ecological Indicators*, *2*, 177–185. http://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. http://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. http://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. http://doi.org/http://dx.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., … Li, F. (2016). Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Applied Geography*, *73*, 77–88. http://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.

http://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. http://doi.org/10.1016/S0169-2046(02)00234-7

Rezvani, M. R., Mansourian, H., & Sattari, M. H. (2013). Evaluating Quality of Life in Urban Areas (Case Study: Noorabad City, Iran). *Social Indicators Research*, *112*(1), 203–220. http://doi.org/10.1007/s11205-012-0048-2

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. http://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. http://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. Retrieved from http://www.jstor.org/stable/2083289

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. http://doi.org/10.1177/0002716215569197

Shelton, T. (2016). Spatialities of data: mapping social media `beyond the geotag'. *GeoJournal*, 1–14. http://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. http://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. http://doi.org/10.1007/s11205-010-9692-6

Statista Inc. (2017). *Statista*. Retrieved January 12, 2017, from https://www.statista.com/statistics/257429/share-of-uk-internet-users-who-use-twitter-by-age-group/

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, *25*(11), 1737–1748. http://doi.org/10.1080/13658816.2011.604636

Tallon, A. R. (2007). Bristol. *Cities*, *24*(1), 74–88. http://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. http://doi.org/10.1007/s11205-012-0126-5

Twitter.Inc. (2016). *Twitter*. Retrieved July 26, 2016, from https://about.twitter.com/company

Warf, B. (2013). *Global Geographies of the Internet*. Springer Netherlands.

Waykar, P., Wadhwani, K., & More, P. (2016). Sentiment analysis in twitter using Natural Language Processing (NLP) and classification algorithm. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *5*(1), 79–81. Retreived from http://ijarcet.org/wp-content/uploads/IJARCET-VOL-5-ISSUE-1-79-81.pdf

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. http://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. http://doi.org/10.1016/j.compenvurbsys.2010.04.001

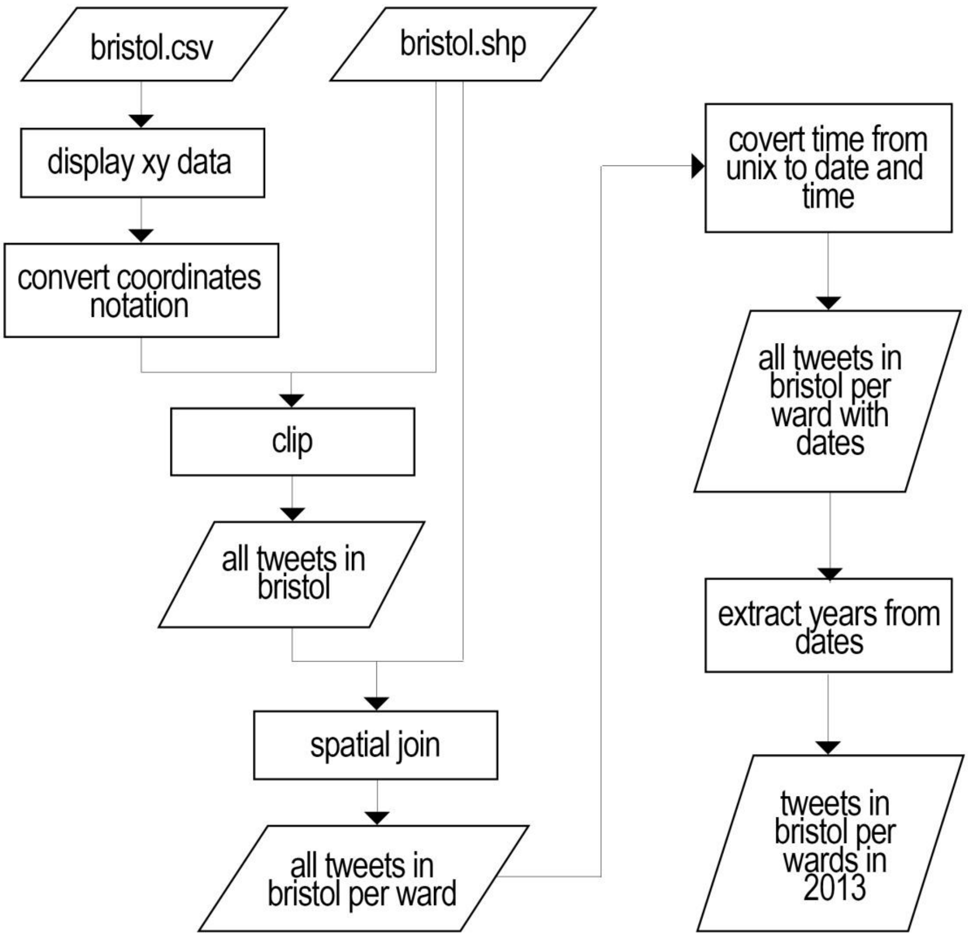
Zook, M., & P. (2014). Offline Brews and Online Views: Exploring the Geography of Beer Tweets. In M. Patterson & N. Hoalst-Pullen (Eds.), *The Geography of Beer Regions, Environment, and Societies* (pp. 201–

209). Dordrecht: Springer Science+Business Media.

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. Retrieved from https://www.questia.com/library/journal/1P3-3687980001/small-stories-in-big-data-gaining-insights-from-large

APPENDICES

APPENDIX 1 - PREPARATION OF TWEETS FLOWCHART



Formula for conversion of unix time in date and time:

- [dateAdd ("s", [time], "01/01/1970 00:00:00")]/1000

Deriving tweets for 2013 using new calculated date:

EXTRACT (YEAR FROM "DATE")=2013

APPENDIX 2 – LIST OF CODES FROM ATLAS.TI ANALYSIS

bad working hours	general dissatisfaction
being drunk	general satisfaction
biking	going out
boring	hate traffic
can't sleep	having headache
celebration	jobs
climate change	junk food
complaining about restaurant	loving the city
complaining about traffic	meditating
diet	neighbourhood partnership
doing yoga	nervous
drinking alcohol	pain
eating	parking
eating habits	personal relations
emotional	public transport
emotionally unstable	recycle
exercising	rush hour
family relations	security
feeling bad	shopping
feeling bored	smoking weed
feeling frustrated	traffic
feeling happy	traffic congestion
feeling sleepy	trains delayed
feeling tired	using drugs
flood alert	boring and lonely
gambling	weather
garbage	

APPENDIX 3 – DICTIONARIES

Transport
railway, transport, firstbus, bus, travel, train, rail, street, parking, traffic, cycle, cycling, bicycle, bike, biking, ride, car, driving, drive, jam, rushhour, cab, walking, hatetraffic, hatepublictransport
Health
flu, headache, pain, drink, drunk, alcohol, beer, smoking, hangover, alcoholic, alcoholism, gym, workout, exercise, exercising, yoga, hospital, doctor, disease, drug, diet, fitness, health, healthy, depression
Environment
environment, environmental, savetheplanet, forest, climate, climate change, recycle, werecycle, garbage, flood, flooding, pollute, pollution, renewable, energy, global warming, sustainable, sustainability, outside, earth, gogreen, neighborhood, nature, naturalresources

APPENDIX 4 – MEANING CLOUD

MeaningCloud™ (http://www.meaningcloud.com) has been used for Text Analytics purposes in the development of this research.

Meaning cloud is a tool providing different language. Options:

- Text Classification
- Sentiment Analysis
- Topics Extraction
- Language Identification
- Text Clustering

In this study it was used as an Excel Add In to perform Automated Sentiment Analysis of messages derived from Twitter.

It identifies the positive/negative/neutral polarity in any text, including comments in surveys and social media.

Differentiators:

- Extracts aspect-based sentiment.
- Discriminates opinions and facts.
- Detects polarity disagreement and irony.

APPENDIX 5 – Paired sample T-test 1

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HEALTH POS	32.0563	35	3.13359	.52967
	HEALTH NEG	32.7959	35	2.84348	.48064
Pair 2	TRANSPORT POS	29.0119	35	4.39788	.74338
	TRANSPORT NEG	30.6652	35	4.39852	.74349
Pair 3	ENV POS	32.1694	35	4.59560	.77680
	ENV NEG	31.1407	35	5.23885	.88553

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	HEALTH POS & HEALTH NEG	35	-.572	.000
Pair 2	TRANSPORT POS & TRANSPORT NEG	35	.352	.038
Pair 3	ENV POS & ENV NEG	35	-.415	.013

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	HEALTH POS - HEALTH NEG	-.73956	5.30134	.89609	-2.56064	1.08151	-.825	34	.415
Pair 2	TRANSPORT POS - TRANSPORT NEG	-1.65329	5.00730	.84639	-3.37336	.06678	-1.953	34	.059
Pair 3	ENV POS - ENV NEG	1.02869	8.28064	1.39968	-1.81581	3.87319	.735	34	.467

APPENDIX 6 – Paired sample T-test 2

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HEALTH POS	32.0563	35	3.13359	.52967
	% respondents satisfied with health services	79.171	35	5.2478	.8870
Pair 2	TRANSPORT POS	29.0119	35	4.39788	.74338
	% respondents satisfied with the bus service	53.060	35	8.9667	1.5157
Pair 3	ENV POS	32.1694	35	4.59560	.77680
	% respondents satisfied with quality of parks and green spaces	83.857	35	7.9139	1.3377

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	HEALTH POS & % respondents satisfied with health services	35	-.022	.900
Pair 2	TRANSPORT POS & % respondents satisfied with the bus service	35	-.228	.188
Pair 3	ENV POS & % respondents satisfied with quality of parks and green spaces	35	.164	.346

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	HEALTH POS - % respondents satisfied with health services	-47.11514	6.17132	1.04314	-49.23506	-44.99522	-45.167	34	.000
Pair 2	TRANSPORT POS - % respondents satisfied with the bus service	-24.04809	10.85037	1.83405	-27.77532	-20.32086	-13.112	34	.000
Pair 3	ENV POS - % respondents satisfied with quality of parks and green spaces	-51.68776	8.47418	1.43240	-54.59875	-48.77678	-36.085	34	.000

APPENDIX 7 – Correlation analysis

Correlations

		IMD score	HEALTH POS	HEALTH NEG	TRANSPORT POS	TRANSPORT NEG	ENV POS	ENV NEG
IMD score	Pearson Correlation	1	-.266	.253	-.191	.006	-.003	.266
	Sig. (2-tailed)		.123	.143	.272	.974	.988	.123
	N	35	35	35	35	35	35	35
HEALTH POS	Pearson Correlation		1	-.572	.246	-.279	.160	-.073
	Sig. (2-tailed)			.000	.154	.104	.359	.676
	N		35	35	35	35	35	35
HEALTH NEG	Pearson Correlation			1	.205	.570	-.067	.249
	Sig. (2-tailed)				.238	.000	.703	.149
	N			35	35	35	35	35
TRANSPORT POS	Pearson Correlation				1	.352	.072	-.187
	Sig. (2-tailed)					.038	.683	.283
	N				35	35	35	35
TRANSPORT NEG	Pearson Correlation					1	-.018	-.009
	Sig. (2-tailed)						.917	.958
	N					35	35	35
ENV POS	Pearson Correlation						1	-.415
	Sig. (2-tailed)							.013
	N						35	35
ENV NEG	Pearson Correlation							1
	N							35

APPENDIX 8

QoL perceptions per domains for 2013

	Transport	% of perceptions	Health	% of perceptions	Environment	% of perceptions
Ashley	2694	45.45	2497	42.12	737	12.43
Avonmouth	1014	66.14	432	28.18	87	5.68
Bedminster	827	51.37	645	40.06	138	8.57
Bishopston	486	42.86	539	47.53	109	9.61
Bishopsworth	601	43.52	693	50.18	87	6.30
Brislington East	677	49.63	564	41.35	123	9.02
Brislington West	713	48.77	614	42.00	135	9.23
Cabot	4820	48.38	4246	42.62	896	8.99
Clifton	504	41.31	569	46.64	147	12.05
Clifton East	544	43.73	583	46.86	117	9.41
Cotham	530	47.11	478	42.49	117	10.40
Easton	438	54.96	301	37.77	58	7.28
Eastville	609	52.36	449	38.61	105	9.03
Filwood	422	48.28	383	43.82	69	7.89
Frome Vale	518	48.32	460	42.91	94	8.77
Hartcliffe	460	41.37	548	49.28	104	9.35
Henbury	408	53.40	310	40.58	46	6.02
Hengrove	809	41.51	875	44.89	265	13.60
Henleaze	262	50.19	221	42.34	39	7.47
Hillfields	400	44.30	421	46.62	82	9.08
Horfield	618	43.58	694	48.94	106	7.48
Kingsweston	205	44.47	220	47.72	36	7.81
Knowle	409	49.34	337	40.65	83	10.01
Lawrence Hill	4758	72.05	1533	23.21	313	4.74
Lockleaze	795	58.76	432	31.93	126	9.31
Redland	425	41.26	495	48.06	110	10.68
St George East	432	46.60	414	44.66	81	8.74
St George West	522	51.33	416	40.90	79	7.77
Southmead	665	35.79	1097	59.04	96	5.17
Southville	1130	54.93	756	36.75	171	8.31
Stockwood	459	47.13	455	46.71	60	6.16
Stoke Bishop	1956	53.33	1296	35.33	416	11.34
Westbury-on-Trym	346	52.27	253	38.22	63	9.52
Whitchurch Park	380	41.58	478	52.30	56	6.13
Windmill Hill	471	47.82	419	42.54	95	9.64
Bristol	31307	50.60	25123	40.60	5446	8.80

% of perceptions about health in different sentiment groups

	P+	P	Neutral	N	N+	None
Ashley	9.33	27.51	4.97	24.59	7.33	26.27
Avonmouth	8.56	22.69	5.09	21.99	7.18	34.49
Bedminster	8.22	24.81	3.88	22.48	7.13	33.49
Bishopston	7.79	25.97	4.27	25.05	8.35	28.57
Bishopsworth	8.23	21.36	4.18	24.24	9.38	32.61
Brislington East	9.93	29.08	3.72	24.47	5.14	27.66
Brislington West	5.54	23.45	5.05	27.36	6.84	31.76
Cabot	7.70	25.11	4.03	19.22	6.62	37.33
Clifton	6.33	23.02	4.92	25.66	8.08	31.99
Clifton East	9.09	24.19	5.15	24.53	6.86	30.19
Cotham	6.49	23.64	3.97	24.69	6.69	34.52
Easton	6.64	26.25	2.99	23.26	9.30	31.56
Eastville	7.57	25.84	5.57	23.16	7.80	30.07
Filwood	5.48	20.10	4.96	31.33	9.66	28.46
Frome Vale	6.09	23.70	5.00	26.74	6.52	31.96
Hartcliffe	7.12	20.80	5.47	24.64	9.85	32.12
Henbury	11.61	22.58	3.23	21.61	6.77	34.19
Hengrove	7.66	23.54	4.23	24.46	8.80	31.31
Henleaze	12.22	23.98	2.71	26.70	6.79	27.60
Hillfields	8.79	27.32	4.99	26.84	7.84	24.23
Horfield	8.65	19.31	4.18	25.79	7.78	34.29
Kingsweston	7.73	23.18	5.00	26.82	10.00	27.27
Knowle	5.93	27.30	2.97	24.04	7.42	32.34
Lawrence Hill	9.00	26.61	4.31	21.27	7.50	31.31
Lockleaze	7.64	24.31	4.40	25.93	8.80	28.94
Redland	6.46	25.25	6.46	25.05	7.07	29.70
St George East	8.21	26.09	2.90	27.54	7.73	27.54
St George West	6.25	22.84	5.29	24.76	11.78	29.09
Southmead	7.84	23.15	4.38	24.89	8.11	31.63
Southville	7.80	27.91	4.23	21.83	9.39	28.84
Stockwood	5.93	21.98	3.96	28.79	6.37	32.97
Stoke Bishop	9.49	26.00	6.25	25.54	6.94	27.31
Westbury-on-Trym	8.30	23.72	4.35	24.90	7.11	31.62
Whitchurch Park	7.11	19.25	5.65	29.08	7.95	30.96
Windmill Hill	8.35	25.06	4.30	25.54	6.21	30.55
Bristol	8.01	24.71	4.53	23.87	7.57	31.38

% of perceptions about transport in different sentiment groups

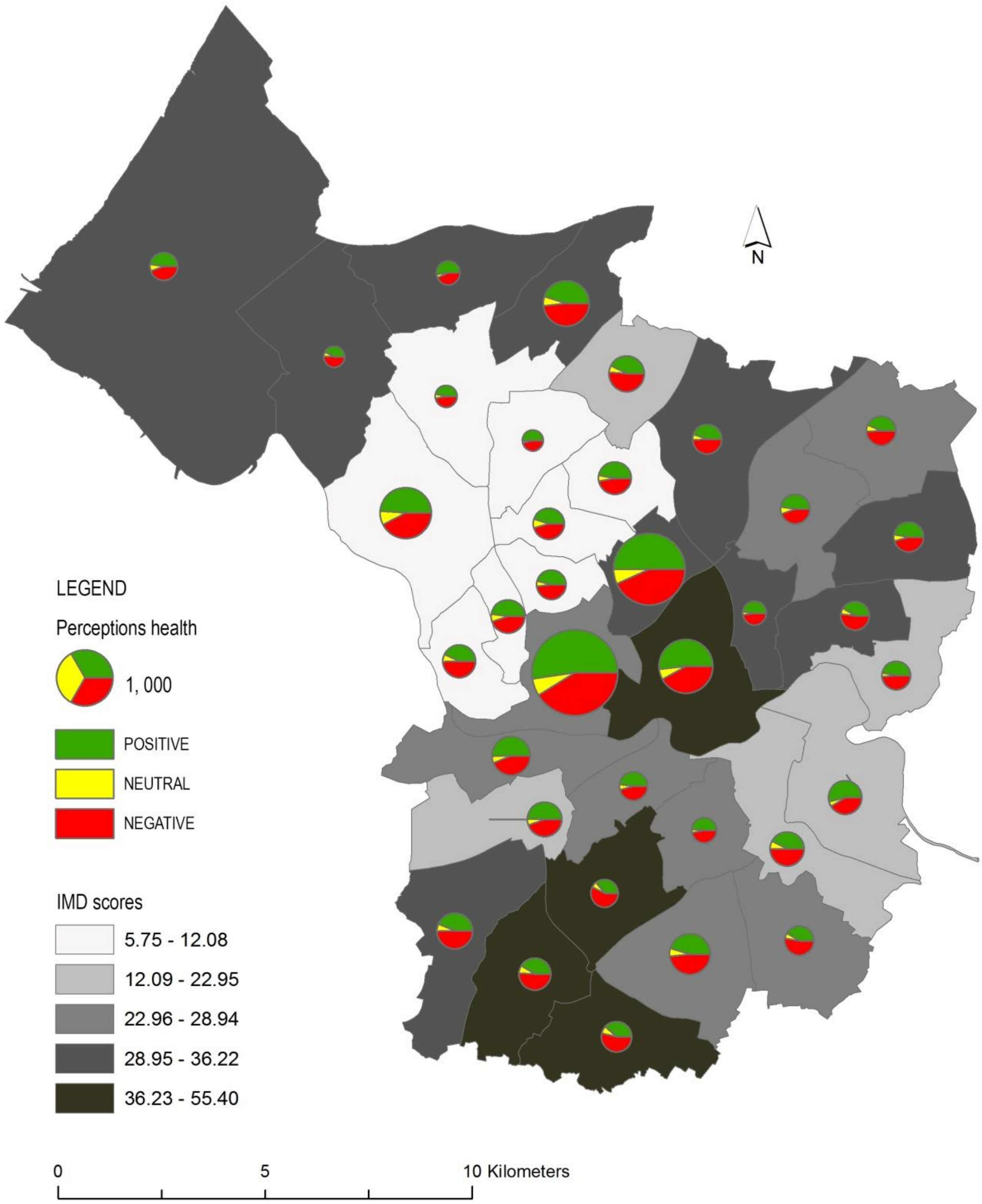
	P+	P	Neutral	N	N+	None
Ashley	13.30	43.54	5.19	29.42	8.55	31.37
Avonmouth	9.33	37.60	5.33	35.20	12.53	63.02
Bedminster	8.01	35.35	7.62	37.11	11.91	38.09
Bishopston	12.88	31.60	7.67	34.66	13.19	32.92
Bishopsworth	9.29	34.70	6.83	33.61	15.57	39.10
Brislington East	10.93	44.33	2.63	28.74	13.36	27.03
Brislington West	7.86	37.99	1.97	34.72	17.47	35.76
Cabot	9.72	38.22	4.93	33.81	13.32	40.66
Clifton	11.15	37.77	3.41	30.96	16.72	35.91
Clifton East	9.51	40.49	5.52	35.28	9.20	40.07
Cotham	10.36	35.80	3.85	36.09	13.91	36.23
Easton	7.78	33.46	4.67	38.52	15.56	41.32
Eastville	13.02	37.24	2.34	31.25	16.15	36.95
Filwood	10.47	35.47	4.05	31.08	18.92	29.86
Frome Vale	9.73	35.26	2.74	38.60	13.68	36.49
Hartcliffe	8.23	37.03	5.38	31.65	17.72	31.30
Henbury	11.63	40.12	2.91	31.98	13.37	57.84
Hengrove	8.90	34.66	4.36	35.23	16.86	34.73
Henleaze	9.73	33.51	3.24	44.86	8.65	29.39
Hillfields	13.94	34.26	4.78	33.47	13.55	37.25
Horfield	10.95	35.00	5.48	34.05	14.52	32.04
Kingsweston	9.29	33.57	5.71	38.57	12.86	31.71
Knowle	10.62	30.40	5.49	39.56	13.92	33.25
Lawrence Hill	9.41	33.62	4.88	40.31	11.78	51.30
Lockleaze	11.11	34.46	6.59	12.81	35.03	33.21
Redland	10.14	33.70	8.70	35.51	11.96	35.06
St George East	10.14	36.71	3.85	40.56	8.74	33.80
St George West	6.58	35.42	5.64	35.74	16.61	38.89
Southmead	9.23	38.74	5.41	29.95	16.67	33.23
Southville	9.59	34.05	4.43	36.48	15.45	38.14
Stockwood	7.81	37.92	5.95	34.20	14.13	41.39
Stoke Bishop	14.48	39.81	5.96	32.15	7.59	34.00
Westbury-on-Trym	10.58	36.51	6.35	32.80	13.76	45.38
Whitchurch Park	7.72	32.52	3.25	39.02	17.48	35.26
Windmill Hill	6.94	37.54	3.15	14.20	38.17	32.70
Bristol	10.01	36.24	4.86	33.78	15.11	39.43

% of perceptions about environment in different sentiment groups

	P+	P	Neutral	N	N+	None
Ashley	8.55	28.36	4.07	18.72	9.63	30.66
Avonmouth	2.30	27.59	0.00	27.59	9.20	33.33
Bedminster	4.35	24.64	4.35	17.39	10.87	38.41
Bishopston	11.93	23.85	4.59	22.94	6.42	30.28
Bishopsworth	8.05	21.84	1.15	22.99	12.64	33.33
Brislington East	7.32	22.76	4.07	23.58	8.13	34.15
Brislington West	4.44	25.93	3.70	20.00	6.67	39.26
Cabot	8.26	26.00	3.68	20.31	8.15	33.59
Clifton	6.12	22.45	4.76	14.29	6.80	45.58
Clifton East	5.13	35.04	3.42	23.08	6.84	26.50
Cotham	4.27	29.91	3.42	21.37	10.26	30.77
Easton	3.45	41.38	5.17	22.41	6.90	20.69
Eastville	7.62	20.00	1.90	18.10	7.62	44.76
Filwood	11.59	26.09	2.90	15.94	13.04	30.43
Frome Vale	7.45	24.47	3.19	22.34	6.38	36.17
Hartcliffe	4.81	22.12	3.85	15.38	15.38	38.46
Henbury	10.87	21.74	8.70	15.22	10.87	32.61
Hengrove	5.28	25.66	3.40	22.26	8.30	35.09
Henleaze	2.56	28.21	7.69	20.51	5.13	35.90
Hillfields	6.10	23.17	4.88	29.27	13.41	23.17
Horfield	7.55	30.19	5.66	16.98	4.72	34.91
Kingsweston	2.78	30.56	0.00	30.56	2.78	33.33
Knowle	2.41	24.10	6.02	26.51	9.64	31.33
Lawrence Hill	6.39	28.43	2.56	21.41	7.03	34.19
Lockleaze	7.94	17.46	2.38	22.22	19.05	30.95
Redland	3.64	30.00	4.55	20.00	10.91	30.91
St George East	2.47	28.40	2.47	18.52	11.11	37.04
St George West	5.06	24.05	2.53	32.91	10.13	25.32
Southmead	5.21	22.92	3.13	30.21	8.33	30.21
Southville	7.60	29.82	1.17	19.30	9.94	32.16
Stockwood	8.33	20.00	6.67	23.33	8.33	33.33
Stoke Bishop	7.21	30.77	3.85	22.12	6.73	29.33
Westbury-on-Trym	11.11	12.70	4.76	28.57	6.35	36.51
Whitchurch Park	1.79	28.57	1.79	23.21	14.29	30.36
Windmill Hill	7.37	29.47	5.26	18.95	9.47	29.47
Bristol	6.21	25.96	3.76	21.96	9.18	32.93

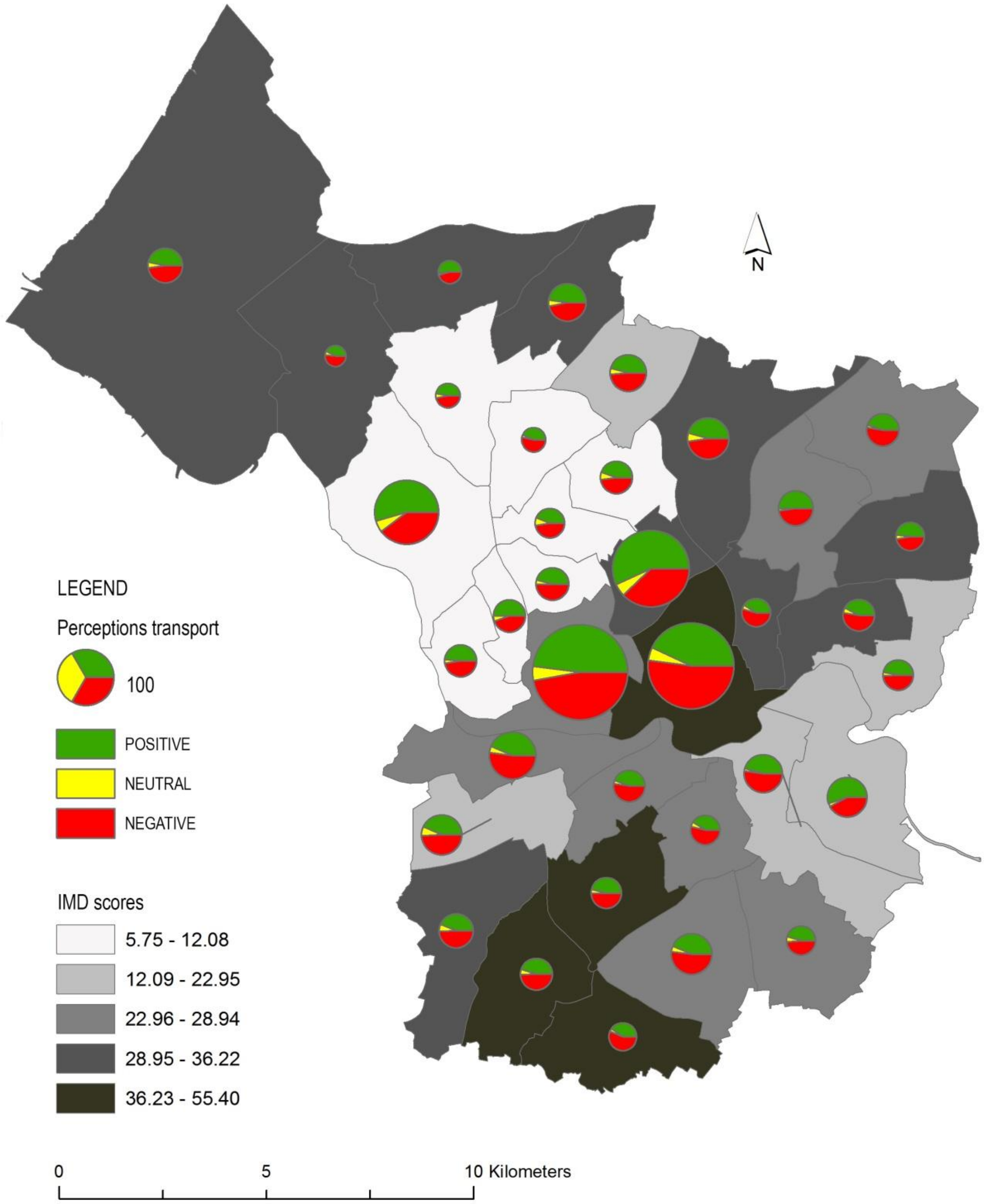
APPENDIX 9

IMD map overlaid with pie chart showing the percentage of positive, neutral and negative perceptions in health domain



APPENDIX 10

IMD map overlaid with pie chart showing the percentage of positive, neutral and negative perceptions in transport domain



APPENDIX 11

IMD map overlaid with pie chart showing the percentage of positive, neutral and negative perceptions in environment domain