INTERNATIONAL JOURNAL OF CARTOGRAPHY
https://doi.org/10.1080/23729333.2025.2581665

Visualising population distribution with choropleth maps: which classification methods are suitable for South African population data?

Lourens Snyman [a] and Serena Coetzee [a,b]

[a]Department, Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa; [b]United Nations University – Institute for Integrated Management of Material Fluxes and of Resources, Dresden, Germany

ABSTRACT
Choropleth maps are among the oldest and still one of the most frequently used techniques for visualising quantitative data, such as population density. Today, each geographic information system (GIS) has a plethora of options for categorising data into classes for choropleth maps. Each option has its pros and cons, depending on the data and the use case. This paper presents the results of a user study to assess the suitability of different data classification methods for effectively visualising population distribution with choropleth maps in South African metropolitan areas. The research focused on geographic accessibility as a use case: population density is visualised on choropleth maps, allowing decision makers to identify over- or underserved areas for the provisioning of facilities. Results show that respondents were more likely to provide correct answers when presented with maps visualising data classified according to quantiles and natural breaks (Jenks), suggesting that these are easier to interpret for assessing and understanding population distribution in South Africa.

ARTICLE HISTORY
Received 9 July 2025
Accepted 24 October 2025

KEYWORDS
Choropleth maps; population density; data classification

RÉSUMÉ
Les cartes choroplèthes sont parmi les cartes les plus anciennes et sont encore parmi les cartes les plus fréquemment utilisées pour visualiser des données quantitatives telles que des données de densité de population. Aujourd'hui tous les systèmes d'information géographique ont de nombreuses options pour catégoriser les données en classes pour faire des cartes choroplèthes. Chaque option a des avantages et inconvénients, selon les cas d'étude et les utilisations. Cet article présente les résultats d'une étude utilisateur visant à évaluer l'adéquation de différentes méthodes de classification de données pour la visualisation efficace des distributions de population dans les régions métropolitaines d'Afrique du Sud. Cette recherche est centrée sur l'accessibilité géographique comme cas d'utilisation : la densité de population est visualisée sous la forme de cartes

CONTACT Serena Coetzee serenacoetzee@unu.edu
Supplemental data for this article can be accessed online at https://doi.org/10.1080/23729333.2025.2581665.

choroplèthes, permettant aux décideurs d'identifier les zones sur- ou sous-desservies en services. Les résultats montrent que les participants ont plus de chance de donner les bonnes réponses avec des cartes classées en fonction des quantiles et des ruptures naturelles (Jenks), ce qui suggère que ces cartes sont plus faciles à interpréter pour évaluer et comprendre la distribution de la population en Afrique du Sud.

Introduction

Data visualisation, or rather geospatial data visualisation techniques, has evolved rapidly over the past decade. Today, any geographic information system (GIS) software application offers a plethora of built-in spatial analysis and visualisation techniques, such as choropleth maps, kernel density estimation heat maps, firefly maps, dot maps, graduated symbols, point density or isochrones, among others. These enable users to visualise spatial patterns in data quickly and effortlessly. With the increased processing capability of desktop computers, high-speed internet connections, cloud computing (Zhang et al., 2019), and parallel processing (Zhao et al., 2016), analysing and displaying large volumes of geographic data has become easier and faster, making it accessible to anyone.

Current proprietary and open-source GIS software development companies and communities are continually simplifying their applications, allowing non-GIS professionals to execute advanced spatial queries and visualisation techniques with just a few clicks. User interfaces are designed to guide end users, regardless of their GIS background, through logical steps to analyse geospatial data and prepare maps. Whether it is to create a topographic or choropleth map or to visualise travel distances with isochrones to optimise service locations, the software guides the user through the analysis and visualisation in a stepwise approach. Although these steps enable non-GIS users to create maps, the software cannot guarantee that the story is told effectively and that the visual message is communicated. 'Because anyone with the right software and an internet connection can now make and publish maps, mapmakers can also easily lie to themselves and others' (Monmonier, 2018).

Choropleth maps are one of the oldest and most frequently used techniques for visualising quantitative data in a GIS (Tyner, 2014). Slocum et al. (2014) noted that a choropleth map is 'the most commonly used (and abused) method of thematic mapping'. A choropleth map categorises observations into several classes, either manually or based on a data classification method. Today, GIS software offers a range of data classification methods for creating choropleth maps, allowing users to easily select and apply a method from a drop-down menu. Each data classification method has its advantages and disadvantages, which can vary based on factors such as the geographic scale and spatial distribution of the data. (Kraak et al., 2021) pointed out that classification could potentially increase uncertainty since patterns derived from a classification are influenced by the positioning of class breaks. Unfortunately, data classification methods also have the potential to distort spatial patterns, causing misleading representations and a potential oversimplification of information (Brewer, 2006; Evans, 1977; Monmonier, 2018). However, the target audience, such as policymakers, who will be interpreting the maps should also be considered (Tyner, 2014).

The challenge with data classification methods for choropleth maps in South Africa is selecting a classification method that effectively displays the country's unequal and dispersed population densities (demand). It should emphasise not only the city centres and their surroundings but also secondary or tertiary populated areas, such as townships and informal settlements, that are segregated from the city centre.

In this paper, we present the results of a user study that evaluated the effectiveness of four data classification methods for choropleth maps depicting population densities in South Africa. The following section describes choropleth maps with a specific focus on the underlying data classification methods used to visualise statistical data. This is followed by a brief historical overview of population distribution in South Africa, which describes the country's diverse population patterns and unequal access to service facilities. Lastly, the chapter explores the concept of geographic accessibility, which serves as the use case for this research, highlighting the key factors relevant to geographic accessibility analysis for the optimal positioning of service facilities. Section 3 describes the materials and methods used for choosing appropriate or suitable study areas, geographic units, and data classification methods that will be evaluated for this research. Results of the user study are presented and discussed in Section 4, followed by a conclusion which summarises key finding and presents recommendations for future research.

Literature review

Choropleth maps

The term 'choropleth' originates from the Greek words choros, meaning area, and plethos, meaning value (Kraak & Ormeling, 2011). A choropleth map is a simple, easy-to-use and easy-to-read technique for classifying and visualising data for statistical areas (Shaito & Elmasri, 2021; Tyner, 2014), such as enumerator areas, wards, municipalities or districts. Choropleth maps are also referred to as thematic maps. De smith et al. (2018) defined choropleth maps as maps that display information about an area based on a particular theme, using techniques such as colour shading, categorised into different classes. Juergens (2020) stated that 'an essential purpose of choropleth maps is the visual perception of spatial patterns'.

To create simple choropleth maps, Robinson (1995) identified three elements, including the size and shape of areas or polygons, the number of classes, and the class limits. The size and shape of polygons are frequently referred to as geographic units. A geographic unit refers to the spatial resolution or granularity of the data, such as enumerator areas, census tracts, wards, or municipalities. Kraak and Ormeling (2020) mentioned that 'it is a good cartographic practice to conveniently arrange the data before displaying them. This process is called classification', which is indeed a form of data generalisation. Five to seven classes are generally accepted throughout the literature (Brewer, 2015; Kraak & Ormeling, 2011; O'sullivan & Unwin, 2010). When choosing a specific number of classes, the choropleth map designer should also consider the upper and lower limits of each class. According to Robinson (1995), 'no aspect of choropleth mapping has received more space in the cartographic literature than methods to determine class limits'. Over the years, numerous data classification algorithms have been developed and integrated into GIS software applications. Since each data classification method has its own advantages and disadvantages, choosing an appropriate method should be considered carefully. One of the key aspects to consider is the overall distribution of the data. 'Different frequency distributions suggest different class interval systems' (Evans, 1977).

Some methods, such as standard deviation, are more effective when the data are evenly or normally distributed. In contrast, natural breaks are typically used when the data distribution is skewed (Całka, 2018). Numerous data classification methods, or methods for determining class limits, for choropleth maps are described in the literature, for example Slocum et al. (2014) highlighted six frequently used methods of data classification. These include equal intervals, quantiles, mean-standard deviation, maximum breaks, natural breaks, and the optimal method. Cromley (2019) observed that the most frequently used optimal classification is the Jenks optimal classification. Tyner (2014) mentioned arithmetic progression and geometric progression as additional classification methods.

De smith et al. (2018) compiled a comprehensive list of frequently used data classification methods for choropleth maps, which are typically used to analyse univariate data that include a single variable, such as population density or mortality rate. Besides those mentioned above, De smith et al. (2018) included unique values, exponential intervals, percentiles, and box plots. Other data classification methods that are worth mentioning include equal feature areas (Lloyd & Steinke, 1977), harmonic series (Kraak & Ormeling, 2011), and nested means (Dent et al., 2009; Kraak & Ormeling, 2011). For a more customised approach, class breaks could be manually adjusted to accentuate a particular phenomenon relevant to the analysis.

One disadvantage of choropleth maps when visualising statistical data is the tendency to 'overemphasize large, yet often sparsely populated, administrative areas because of their strong visual weight' (Besançon et al., 2020). This is also true for population data depicting urban and rural areas, as rural areas encompass a larger geographic space than urban areas (Harris et al., 2017). Bertin (1967/1983) noted that cartograms, also referred to as anamorphosis maps, are an effective solution for the problem of sparsely populated administrative areas. With cartograms, the size and shape of polygons are distorted based on a specific variable. For instance, when visualising population density, high density polygons would be enlarged and overemphasised and low-density polygons would be smaller, carrying less visual weight.

Furthermore, since statistical data are usually captured on various administrative boundaries, such as enumerator areas or wards, which are demarcated artificially, 'it is not possible to show variation within enumeration areas' (Tyner, 2014). Slocum et al. (2014) stated that choropleth maps are most effective and accurate when the size and shape of the polygons are fairly similar.

A challenge associated with choropleth maps is selecting a suitable data classification method that displays and communicates the data distribution in a clear and effective way (Slocum et al., 2014). Kraak et al. (2021) pointed out that classification could potentially increase uncertainty, as the patterns derived from a classification are influenced by the positioning of class breaks. Additionally, Jenks (1963) commented that 'a cartographer makes a series of judgments without really understanding what effect these judgments will have upon the reader's interpretation of the distribution'. Also, data classification methods for visualising data with choropleth maps could potentially distort spatial patterns, causing misleading representations and a possible oversimplification of information (Brewer, 2006; Evans, 1977 Monmonier, 2018). Schiewe (2023) noted that standard or traditional data classification methods are not ideal for visualising multi-temporal data, where temporal changes are either lost or incorrectly highlighted. Schiewe (2023) further introduced a change preservation metric that evaluates the effectiveness of a data classification method. The metric could be used as a data classification method that 'explicitly takes the preservation of changes into account'.

Population distribution in South Africa

South Africa has an uneven population distribution, which leads to unequal access to service facilities. Weir-Smith and Dlamini (2024) further mentioned that 'unequal spatial concentration is at the heart of economic imbalance in South Africa' and that post-apartheid policies did not resolve economic imbalances caused by the apartheid era. A report by the World Bank (2018) on poverty and inequality describes South Africa as one of the most unequal countries in the world, stating that 'inequality has increased since the end of apartheid in 1994'. Also, South Africa is ranked first among 164 countries in the World Bank's global database of Gini coefficients, which measures inequality of per capita consumption (or income, depending on the country). The 2018 Gini index for South Africa is notably high at 67 (World Bank, 2022). The Gini index is expressed as a value in the range from 0 to 1 or in percentages from 0 to 100. Zero represents perfect equality, whereby an index value of 1 or 100 means perfect *in*equality.

People in South Africa are highly segregated, and this segregation is inherently geographical (Brown & Chung, 2006). The Group Areas Act 41 of 1950 (apartheid) 'prohibited the multiracial use or occupation of urban land' (Strauss, 2019). Segregated zones were established in urban areas, allowing only certain race groups to live and work there. This resulted in the establishment of densely populated townships and informal settlements outside city centres and suburban living spaces. Mostly poor and pro-poor people reside in these areas, and access to basic services and public service centres is limited and insufficient.

Statistics South Africa[1] is the custodian of demographic data in South Africa. The last national census survey was conducted in 2022. Unfortunately, demographic data has not yet been released at a more granular level than local or metropolitan municipalities. Hence, for this research, demographic (or population) data from the previous national census conducted in 2011 were used. The total population is estimated to have reached 62 million in 2022, an increase from 51.7 million in 2011.

For Census 2011, the country was demarcated into approximately 103 000 enumerator areas 'based on specifications of administrative boundaries, size, and population density' (Statistics South Africa, 2012a), which were used to capture all the demographic data. These enumerator areas were subsequently classified into ten enumerator area types 'according to a set criteria profiling land use and human settlement within the area' (Statistics South Africa, 2012b). Of the 51.8 million people in South Africa, more than half (55.8%) live in formal residential areas, followed by traditional residential areas and informal residential areas, with 31.3% and 5.8%, respectively. Traditional residential areas are unique in the sense that it is 'community owned land under the jurisdiction of a traditional leader' (Statistics South Africa, 2012b). The population within municipalities is unevenly distributed. In addition to densely populated central business districts, most municipalities also feature scattered pockets of populated townships or informal settlements located in peri-urban areas or outside the city centre. Due to its history, the geographic distribution of the South African population presents unique challenges for visualising spatial patterns on a map for effective decision-making.

Geographic accessibility and population demand

'Accessibility is the most widely used metric in measuring the value of a location in public service delivery' (Church & Murray, 2009). Neutens (2015) noted that accessibility is multi-faceted and described five dimensions of health care accessibility: affordability, acceptability, availability, geographic accessibility and accommodation or appropriateness of a health service. This paper focusses specifically on geographic accessibility. Rodrigue et al. (2009) defined spatial or geographic access as 'the measure of the capacity of the locations to be reached by, or to reach, different locations' based on both time and distance (Ashiagbor et al., 2020) between people and service centres. Apart from travel-time and distance to a destination, accessing, waiting for transport and duration of travel could also be considered (Lucas, 2011). This links to the 'space–time of action' concept of Demoraes et al. (2020), meaning that the mode of transport, number of connections required, and availability of 'a rapid, mass public transport system near the place of residence' are also important factors which could influence accessibility. In connecting the population to public services, decision makers must understand where the population demand is in relation to public service facilities (Snyman & Coetzee, 2024). To achieve this, it is essential to create maps that visualise both supply (service centres or facilities) and demand (population) in order to identify gaps and shortfalls. Generally, choropleth maps are effective for visualising statistical data such as population densities (Tyner, 2014). Depending on the data distribution, data classification methods for choropleth maps could distort spatial patterns, resulting in a misleading representation of the data. For instance, if the data distribution is highly skewed, an equal interval data classification could over-generalise densely populated areas, giving an inaccurate visual impression of density patterns, which in turn could highlight inappropriate locations for the positioning of service centres, ultimately leading to poor service delivery and frustrated citizens. Another challenge with choropleth maps, specifically in a South African context, is the selection of a classification method that effectively displays unequal and dispersed population densities (demand). It should emphasise not only the city centres and their surroundings but also secondary or tertiary populated areas such as townships and informal settlements that are segregated from the city centre.

The CSIR guidelines for the provision of social facilities in South African settlements highlight six main service delivery categories. These include health and emergency services, public services, civic services, social services, educational services as well as parks and recreational services. Various facilities are listed within each category. The guidelines provide recommendations with regards to average population thresholds and acceptable travel distances for each kind of facility (Green, 2015).

'Improving service delivery to all people in South Africa is a key priority of government' (Green, 2012). Population demand refers to the 'number of people who may need the services' (Ma et al., 2018). Although visualising population demand is not the only metric used to analyse geographic accessibility, it is one of the prominent outputs of such a study, which is usually presented to a target audience. Other metrices include travel time or distance calculations to a service centre.

The purpose of choropleth maps for geographic accessibility analysis is to visually differentiate between densely and sparsely populated areas, while also identifying locations that are either over – or underserved for the optimal provision of service centres. The

Department of Public Service and Administration's geographic access guideline describes three facility location models. These include an expansion model, reduction model and relocation model (DPSA, 2012). The expansion model encompasses two approaches: greenfield and brownfield. For the greenfield approach, optimal locations for service centres are determined based on the population distribution (population demand), regardless of the current footprint of service centres. The brownfield approach considers the current footprint when determining optimal locations. If service centres are not optimally located (close to the people) due to possible settlement growth or movement patterns, they could either be relocated (relocation model) or closed down (reduction model).

Research design

Study areas

South Africa is categorised into ten enumerator area types: formal residential, informal residential, traditional residential, farms, parks and recreation, collective living quarters, industrial, small holdings, vacant, and commercial, highlighting the diverse landscape of the country (SuperCROSS, 2012). To effectively evaluate and assess the suitability of data classification methods for choropleth maps that depict population demand in a diverse geographic setting, four representative municipalities were selected as study areas where populations are distributed across all ten enumerator area types: City of Tshwane Metropolitan Municipality, Buffalo City Metropolitan Municipality, Mangaung Metropolitan Municipality, and Polokwane Local Municipality. See Figure 1.

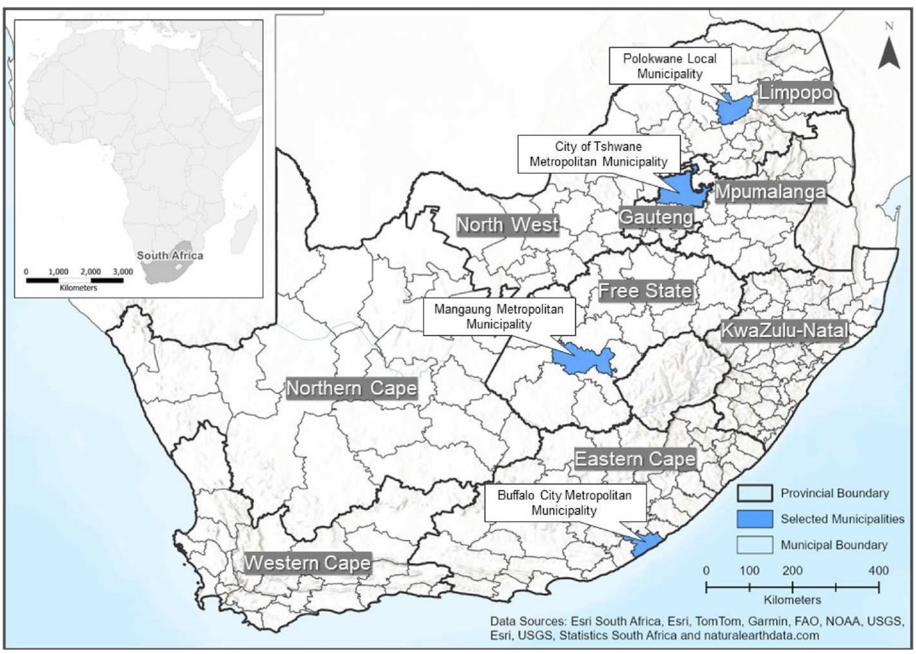

Figure 1. Municipalities in South Africa showing the four selected study areas.

Geographic units

Firstly, geographic units depicting population demographics were selected from the Census 2011 Community Profiles in SuperCROSS database, an official and freely available data source. These include a small area layer (SAL) and sub-places (SP). Small area layer polygons are the smallest geographical units with demographic data, whereas sub-places are aggregated polygons derived from small area layers that represent suburbs or villages. Secondly, since both small area layer and sub-place polygons vary in size, hexagons were created to represent equal-sized polygons, as we also wanted to determine whether equal – or varied-sized polygons could influence participants' interpretation of choropleth maps. Rather than computing spatial overlays, which present their own challenges, it was decided to superimpose the Census 2011 population data onto the hexagons. A point data set called Spot Building Count representing dwelling locations was aggregated per hexagon to indicate densities, or rather household densities, per hexagon.

The Spot Building Count data are maintained by Eskom,[2] the main electricity supplier in South Africa. Points are captured from Spot 5 imagery (European Space Agency, n.d.) and verified through various sources, including schools, 1:50 000 topographic data, and dwelling points captured by Statistics South Africa (n.d.). Hexagons were created with an area of two square kilometres, because the sub-place median polygon size of all four municipalities combined is approximately 1.5 square kilometres, which was rounded up to 2 square kilometres. Figure 2 illustrates the shape and size of the three geographic units in one of the selected study areas. Sub-places are shown in green, black polygons represent the small area boundaries and the red dashed lines show the hexagons. The illustration clearly highlights the varied-size and shape of the different geographic units. The aggregation of small areas into sub-places is also clearly visible.

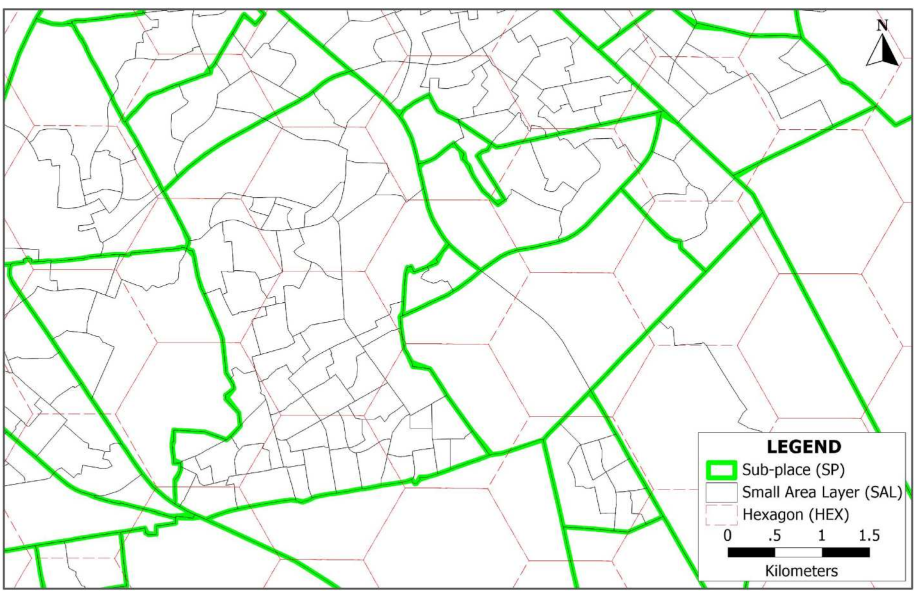

Figure 2. Sub-place, Small area layer and Hexagon polygons.

Data classification methods

While there are various GIS software applications available for spatial analysis, mapping, and visualisation (both open source and licenced), ArcGIS Pro[3] and QGIS[4] are considered the most popular and frequently used applications globally, including in South Africa (Coetzee et al., 2018). GIS Geography's (2022) ranking of the 30 best GIS software applications placed ArcGIS Pro in the top spot, followed by QGIS. Bolstad (2012) mentioned that ArcGIS is the most popular GIS software. Hence, for the purpose of this research, the effectiveness of data classification methods for choropleth maps available in either ArcGIS Pro or QGIS will be examined.

There are nine data classification methods available in both ArcGIS Pro and QGIS. Except for the manual and defined interval methods, which require users to set custom/manual limits for each class, this study initially identified seven data classification methods, namely equal interval, geometric interval, logarithmic scale, natural breaks (Jenks), pretty breaks, quantiles, and standard deviation. One of the key aspects recommended in the literature for selecting a data classification method for a specific data set is conducting a data distribution test, also known as a test for normality in the data. The household distribution by hexagon and population density by small area layer and sub-place are not normally distributed in any of the four study areas; on the contrary, they are highly skewed. The skewness and kurtosis measure the 'degree of normality of distributions, or the lack thereof' (Ho et al., 2015). A skewness value between – 0.5 and 0.5 suggests a normal distribution of frequencies (Hatem et al., 2022). The degree of skewness of the four study areas per geographic unit ranges from 2.49–8.02. This indicates that the majority of polygons exhibit a low population distribution (or low household count for hexagons and low population density for small area layers and sub-places), while only a small number of outliers show extremely high density.

Based on the skewness of the population distribution for each study area and geographic unit (hexagon, sub-place and small area layer), the standard deviation, equal interval and pretty breaks data classification methods were excluded from further analysis, as these methods are best suited for data that are normally distributed (Slocum et al., 2014; Tyner, 2014; Vasilca, 2019). Furthermore, equal intervals and pretty breaks are most effective when the data are uniformly distributed, meaning the data distribution has no peak and is consistent (Kraak et al., 2021). Hence, these methods would not effectively represent data distribution in South Africa, nor would the data variability or nuances stand out. The remaining data classification methods that were tested in the user study are: geometric interval, logarithmic scale, natural breaks (Jenks), and quantiles.

Geometric interval

The geometric interval data classification method is only available in ArcGIS Pro. The method attempts to group an equivalent number of features in each class while maintaining consistent class intervals. This is done by minimising the sum of squares for the data in each class. This method is usually preferred if the data are skewed (Evans, 1977).

Logarithmic scale

The logarithmic scale, which is only available in QGIS, creates an exponential increase between each class break, for example; 0-10, 10-100, 100–1000 and so on. This method is useful when data span a wide range of values. Since the legend is difficult to interpret, the map designer should consider changing the legend text manually. This method is also recommended in literature when the data distribution is highly skewed (Leydesdorff & Bensman, 2006).

Natural breaks (Jenks)

Jenks implemented a data classification algorithm in 1977, which was proposed by Fisher, to identify data belonging to the same class (as cited in Dent et al., 2009). The method, known as the optimal method, is now more commonly referred to as natural breaks (Jenks) in modern GIS software applications. The objective is to minimise the measure of data classification errors. Data with similar values are grouped into classes, with class breaks defined where there are significant differences between the data values. The number of features in each class is usually unevenly distributed, and the intervals between class breaks are not consistent. One advantage of this method is that it is effective for data that are not normally distributed (Całka, 2018).

Quantiles

Quantiles attempt to group an equal number of features in each class (De smith et al., 2018). As a result, class intervals are usually not consistent. This method is effective if the data are not normally distributed. One advantage of this method is that no class interval will be empty, meaning features are distributed across all classes (Vasilca, 2019), compared to methods such as equal intervals or pretty breaks, where some classes may be empty.

Map design

A total of 48 choropleth maps were created to depict the four study areas (local and metropolitan municipalities), three geographic units (hexagons, small area layers, and sub-places), and four data classifications. Following recommendations in literature, five class intervals were used on all the maps (Brewer, 2015; Kraak & Ormeling, 2011; O'sullivan & Unwin, 2010). To ensure consistency, a single colour scheme was used for all maps, ranging from light yellow to dark blue. We also kept the default legend settings. Figures 3–5 present dwelling count or population density by hexagon, small area and sub-place for the Buffalo City Metropolitan Municipality and the four selected data classification methods.

Geometric Interval	Logarithmic Scale
Natural Breaks (Jenks)	Quantiles
Geometric Interval	Logarithmic Scale
Natural Breaks (Jenks)	Quantiles
Geometric Interval	Logarithmic Scale
Natural Breaks (Jenks)	Quantiles

User study

'User studies can objectively establish which method is most appropriate for a given situation' (Kosara et al., 2003). User studies are conducted using a survey or questionnaire. Survey questionnaires have been used for many years (Clifford et al., 2016), including in the field of geography, and have proven to be an important tool for evaluating people's perceptions or interpretations of a subject.

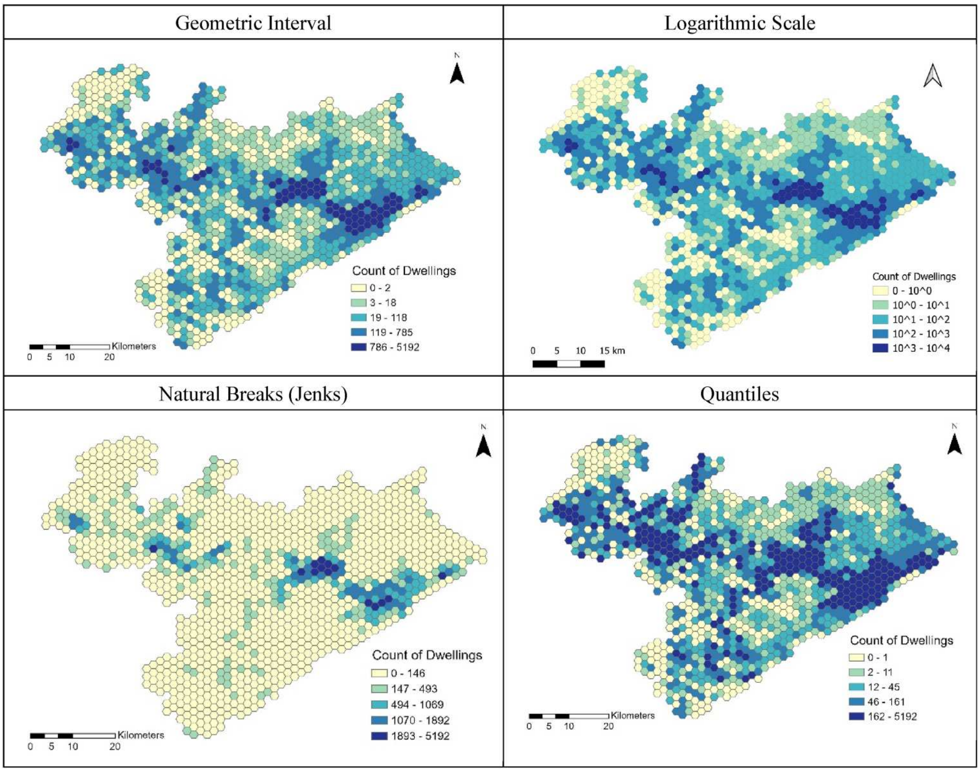

Figure 3. Buffalo City Metropolitan Municipality – Count of dwellings by hexagon.

For this paper, the user study includes a survey featuring an online questionnaire designed to assess respondents' interpretations of choropleth maps depicting population demand. The survey was created using Qualtrics, a secure online platform specifically designed for building and customising online questionnaires. Respondents were students from the University of Pretora with and without prior experience in geography and GIS. Students were chosen for the user study because they, in one way or another, represent the upcoming workforce of the country. In total 107 students participated in the user study. They were required to differentiate between densely and sparsely populated areas and to identify locations that are over – or underserved for the optimal provision of service centres. Four questions were designed, with increasing level of difficulty from 1 to 4 (See Table 1). Each question required respondents to click on a specific number of locations on the map. The underlying data classification method used for each choropleth map was not disclosed to respondents.

Respondents were assigned exactly the same questions in the same sequence (within-subject participant assignment). One disadvantage of this method is a potential learning effect, meaning that a response to a question could be influenced or affected by previous questions. Thus, to eliminate sequential dependencies of responses and the potential learning effect, both the geographic unit and locality (study areas) changed throughout the questionnaire. The questionnaire assessed respondents' ability to

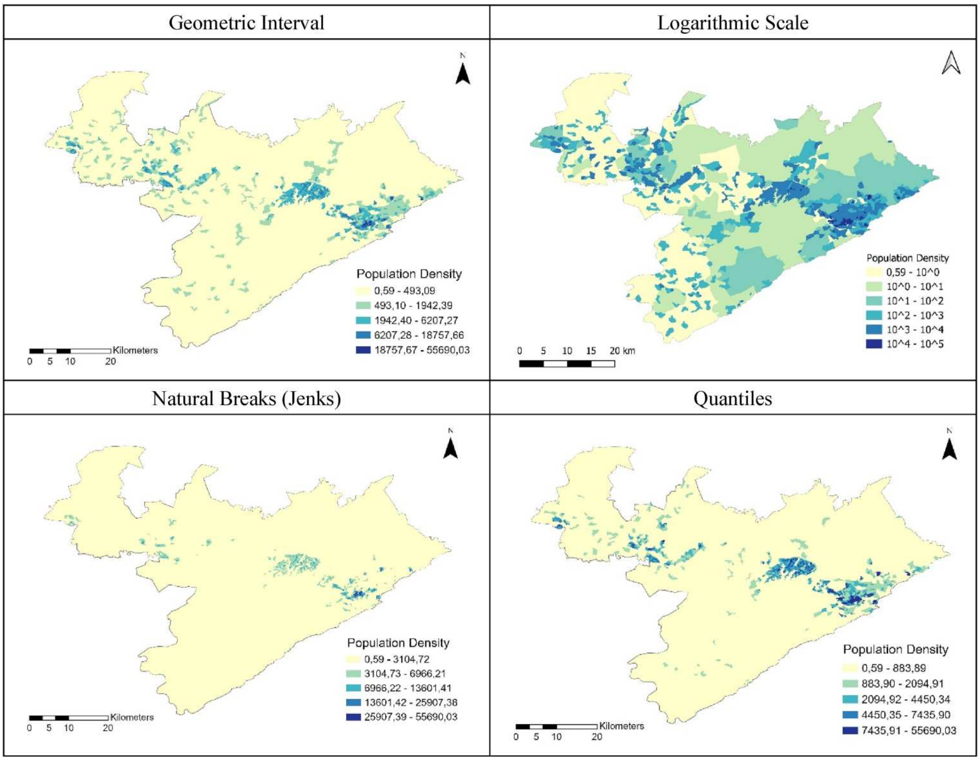

Figure 4. Buffalo City Metropolitan Municipality – Population density by small area.

solve one of the four geographic accessibility problems (see Table 1) and depict a specific data classification method, geographic unit, and study area.

The research design was approved by the Ethics Committee of <anonymized>. Participants who accepted the invitation to participate in the user study were informed about the purpose of the research and gave their consent before completing the questionnaire.

Results

The click locations of the respondents were assessed, validated, and flagged as either correct or incorrect, with 1 indicating a correct response and 0 indicating an incorrect one. Additionally, each question required respondents to click on a specific number of locations on the map, referred to here as click events.

For example, Figure 6 shows the actual click events of all respondents for Question 2: *Identify the best area for opening a new service centre. Ideally the new centre should be located in a densely populated cluster. The maximum threshold (catchment area of a service centre) is 10 km, i.e. a centre serves only people within 10 km from it. Anybody living further away is not served by that centre.*

Each dot represents a click event. Red dots were manually flagged as incorrect based on visual inspection. Yellow dots represent correct locations. Respondents should click on locations where there is a cluster of dark blue polygons (high density). Each dot represents

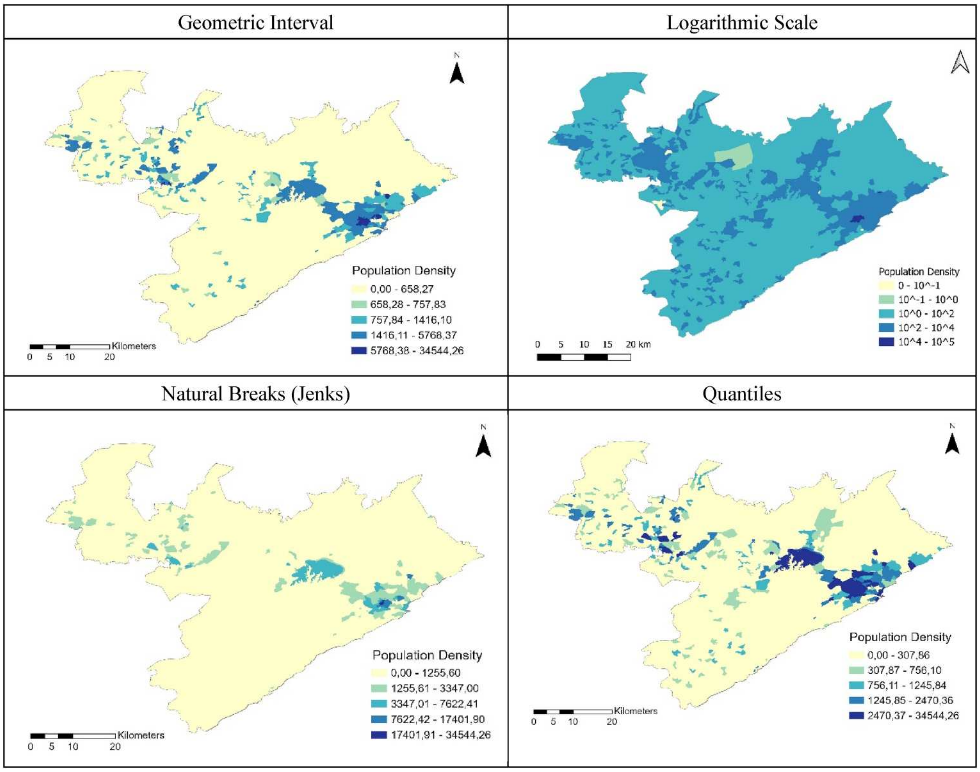

Figure 5. Buffalo City Metropolitan Municipality – Population density by sub-place.

Table 1. Geographic accessibility questions for choropleth maps with dwelling counts per hexagon. For the small area polygons and sub-places, the questions referred to population density.

Question 1	Identify three areas on the map where the dwelling count is very low. Click on the relevant areas.
Question 2	Identify the best area for opening a new service centre. Ideally the new centre should be located in a densely populated cluster (areas with a high dwelling count). The maximum threshold (catchment area of a service centre) is 10 km, i.e. a centre serves only people within 10 km from it. Anybody living further away is not served by that centre.
Question 3	Identify two locations on the map to add additional service centres. These centres should be located in high-density clusters (areas with a high dwelling count) with no other facility nearby (further than 10 km away from existing centres).
Question 4	Identify two service centres that are incorrectly placed and should be relocated to different locations. Remember the ideal location would-be high-density clusters (areas with a high dwelling count) with no other facility close by (within 10 km).

a click event. Yellow dots represent correct locations. Red dots were manually flagged as incorrect based on visual inspection.

Overall, the respondents performed well in the user study, achieving an average accuracy rate of 90%. Four respondents scored 100%, while three individuals recorded the lowest accuracy level of 65%. Females performed slightly better, achieving a rate of 90%, followed by males at 89%. Respondents with prior knowledge or experience in geography performed slightly better (92%) than those without previous experience (89%). Additionally, the frequency of map use does influence the percentage accuracy

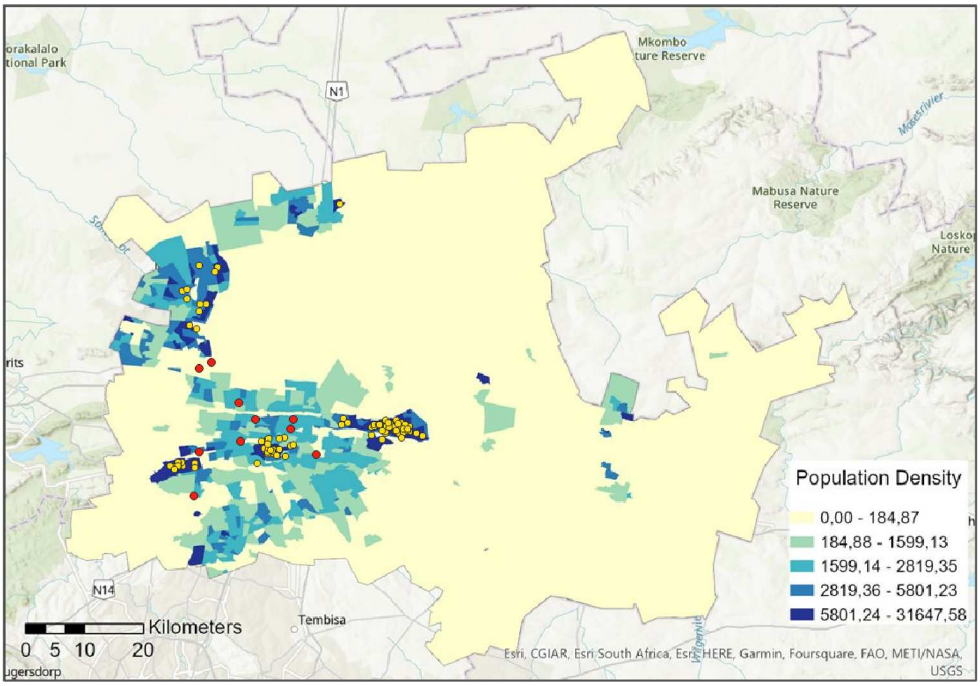

Figure 6. Choropleth map depicting correct and incorrect click events.

of responses. Students who indicated that they use maps once a week or every day performed well (92% and 90%) compared to those who use maps only once a month (88%) or rarely (86%).

Data classification methods

This section assesses the accuracy of responses, specifically the percentage of correct and incorrect answers among respondents, in relation to the four data classification methods. The first part includes a series of data tests to determine whether the data are fit for purpose and statistically relevant for analysis. This is followed by a results section in which response accuracies were evaluated based on the four data classification methods. Although the goal and primary objective of this research were to evaluate the suitability and effectiveness of different data classification methods for visualising population distribution with choropleth maps, we also sought to determine whether other variables are significantly associated with response accuracies. Apart from the four data classification methods, other predictor variables include the geographic accessibility question (Q1 – Q4), study areas, geographic units, respondents' perceived difficulty scores for each question, and their confidence ratings.

Data inspection

Firstly, a Shapiro – Wilk test (kwak & Park, 2019) was conducted to test for normality in the data. This test compared the percentage of correct answers given by each respondent to the 48 choropleth map questions in the questionnaire. Results from the Shapiro – Wilk test revealed a not normal distribution (sig. < 0.001). Additionally, a Friedman test was done to test for differences among the four data classification methods. This test is also recommended for nonparametric data distributions. A significance score, or *p*-value of less than 0.05 is generally considered to be statistically significant (Andrade, 2019). Results confirm a significance score of less than 0.001, indicating that there are statistically significant differences of responses between the data classification methods.

Lastly, a likelihood ratio test (Sur et al., 2019) was done to test the significance of data classification methods as a whole, in relation to response accuracies of respondents. Results suggest that data classification methods significantly affect or influence the outcome of response accuracies (sig. < 0.001).

The percentage accuracy of responses for each of the four data classification methods was notably high. Respondents were, however, more likely to provide correct answers for maps depicting the quantiles data classification method (92%), followed by natural breaks (Jenks) and geometric interval, with 91% and 89% accuracy, respectively. The logarithmic scale was ranked the lowest, with an accuracy of 87%. The box plot in Figure 7 shows the percentage accuracy per respondent for the four data classification methods. Quantiles and natural breaks (Jenks) exhibit less variation, albeit with fewer outliers (mild outliers are marked with black dots and extreme outliers are marked with stars) than the other two methods.

The data distribution of response accuracies per data classification method showed significant differences. The skewness for both natural breaks (Jenks) and quantiles was below −1, with values of −1.319 and −1.947, indicating a strongly negatively skewed distribution. The geometric interval and logarithmic scale had a skewness of −0.938 and −0.682, respectively, depicting a slightly more normal distribution. See Figure 8.

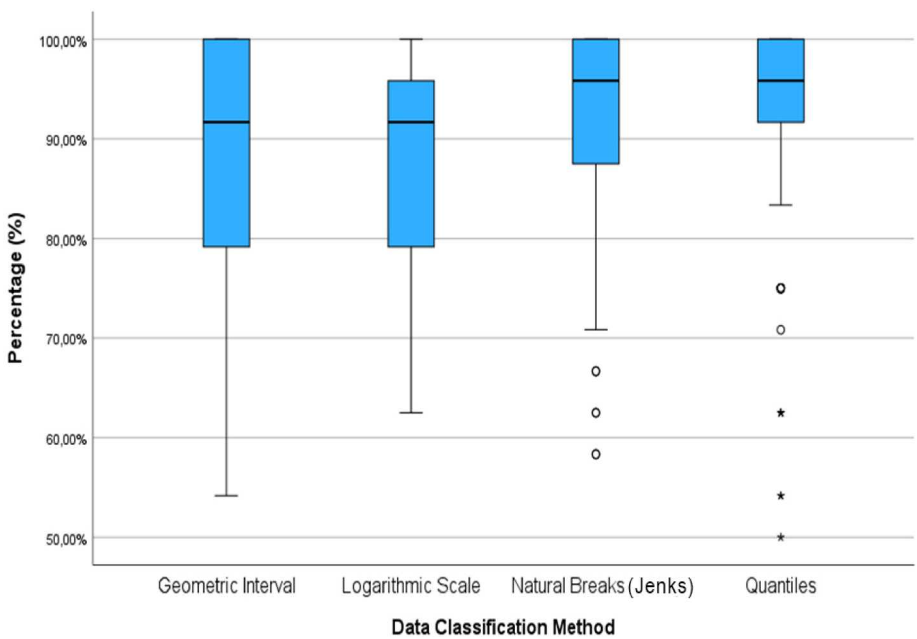

Figure 7. Accuracy score per data classification method.

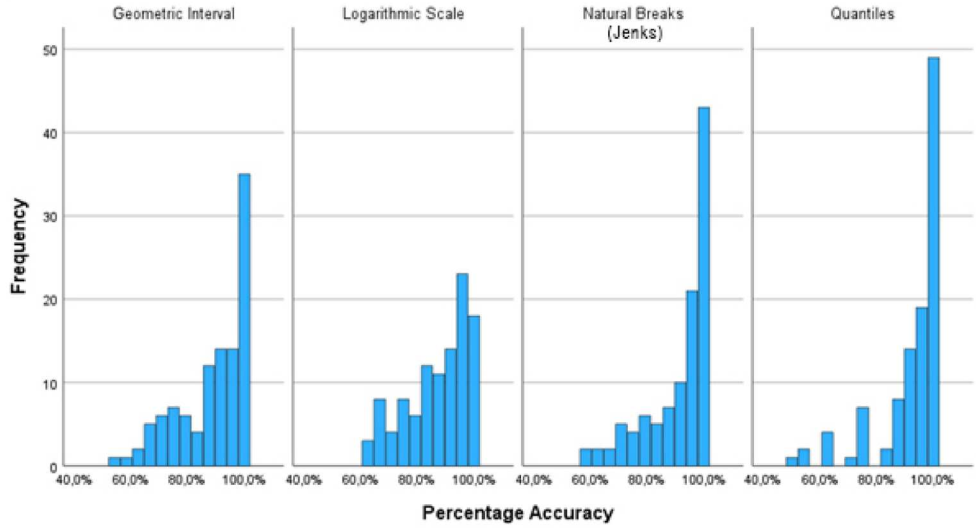

Figure 8. Histogram showing percentage accuracy of respondents per data classification method.

Significance of secondary predictor variables

Although the goal of this research was to evaluate the suitability and effectiveness of different data classification methods to visualise population demand with choropleth maps, we also wanted to test whether secondary predictor variables could be significantly associated with response accuracies in general, as well as for each data classification method individually. These variables include:

- Question – the four geographic accessibility questions,
- Location – the four study areas,
- GeoUnit – the defined geographic units (hexagons, small area layer, and sub-places),
- Confidence – the confidence level in answering a specific question, and
- Level – the rating of the perceived difficulty level of a question.

A logistic regression analysis was used to test the significance of predictor variables against a single dichotomous dependent variable, which represented respondents' answers as either correct or incorrect responses. These predictor variables were tested against the entire data set with all responses combined, as well as individually for each data classification method.

Overall, a statistically significant score of less than 0.001 was achieved for the following predictor variables; question, location and level (see Table 2). When evaluating responses for each data classification method individually, the geographic accessibility questions proved to be statistically significant for each method. Location was significant for both

Table 2. Significance score of predictor variables for each data classification method.

Predictor Variable	Geometric Interval	Logarithmic Scale	Natural Breaks (Jenks)	Quantiles	Overall
Question	0.000	0.000	0.000	0.000	0.000
Location	0.000	0.000	0.080	0.136	0.000
GeoUnit	0.599	0.141	0.001	0.092	0.884
Confidence	0.557	0.190	0.046	0.245	0.022
Level	0.000	0.000	0.000	0.318	0.000

the geometric interval and logarithmic scale methods. Geographic units and respondents' self-perceived confidence level were not significant overall, nor for any of the four data classification methods individually. The last variable; respondent's rating of their perceived difficulty level of a question proved to be significant for each data classification method, except for quantiles.

As discussed in the Map Design section, geographic accessibility questions were designed based on real-world scenarios. In theory, these questions also varied in increasing levels of difficulty. Question 1 was the easiest, while Question 4 was considered the most difficult. The results confirmed our theory. Most participants answered Question 1 correctly (95%), followed by Questions 2, 3, and 4 with 94%, 88%, and 82%, respectively. This was also true for each data classification method except quantiles, where respondents performed slightly better on Question 2 than Question 1 (97% vs 95%).

Table 3 shows the percentage accuracy per study area overall, as well as for maps depicting the four data classification methods. Respondents provided the most correct answers for choropleth maps showing the Mangaung Metropolitan Municipality (92%). Upon visual inspection, the municipality appears to have large, sparsely populated areas, with only a few high-density locations. The City of Tshwane Metropolitan Municipality ranked second, followed closely by the Buffalo City Metropolitan Municipality, which achieved scores of 89.9% and 89.5%, respectively. The most incorrect answers were for the Polokwane Local Municipality, with 88%.

Table 3 shows the percentage accuracy per study area for maps depicting the four data classification methods. The table also highlights variations in response accuracy between study areas and data classification methods. Hence, study areas influence respondents' interpretation of choropleth maps. For maps showing the Buffalo City Metropolitan Municipality, respondents provided the most accurate answers for the quantiles data classification method (93%), followed by the geometric interval method with 92%. Results for the City of Tshwane Metropolitan Municipality are notably different. Most correct answers were based on natural breaks (Jenks) and logarithmic scale (94% and 91%). Quantiles and natural breaks (Jenks) were best interpreted for maps showing the Mangaung Metropolitan Municipality (95% and 93% respectively), with natural breaks (Jenks) being the highest for the Polokwane Local Municipality (91%), followed by quantiles and geometric interval with 90%.

Questions rated by respondents as very easy would likely have the highest percentage of accuracy, followed by a gradual decrease, with questions rated as very difficult having the lowest percentage of accuracy. Results generally followed this pattern except for questions rated as very difficult. A percentage accuracy of 95% was calculated for questions considered

Table 3. Percentage accuracy per study area and data classification method

Study Area	Geometric Interval	Logarithmic Scale	Natural Breaks (Jenks)	Quantiles	Overall
Buffalo City Metropolitan Municipality	92%	85%	86%	93%	90%
City of Tshwane Metropolitan Municipality	85%	91%	94%	90%	90%
Mangaung Metropolitan Municipality	89%	91%	93%	95%	92%
Polokwane Local Municipality	90%	79%	91%	90%	88%

Table 4. Percentage accuracy per difficulty level and data classification method.

Difficulty Level	Geometric Interval	Logarithmic Scale	Natural Breaks (Jenks)	Quantiles	Overall
Very Easy	97%	91%	97%	95%	95%
Easy	93%	89%	93%	94%	92%
Neutral	83%	84%	86%	90%	86%
Difficult	69%	79%	84%	89%	80%
Very Difficult	67%	84%	91%	87%	84%

to be very easy (Table 4). This was followed by easy, neutral, and difficult, with 92%, 86%, and 80%, respectively. The percentage accuracy of questions that respondents considered very difficult was, however, higher at 84% compared to those considered difficult (80%).

By evaluating the perceived difficulty level of respondents per data classification method, a similar pattern was observed in which the percentage accuracy gradually decreased with each perceived difficulty level, ranging from very easy to very difficult. The overall percentage accuracy for quantiles was very high, regardless of the perceived difficulty level. The lowest percentage accuracy was recorded for the difficult and very difficult levels based on the geometric interval data classification method, with 69% and 67%, respectively. The highest percentage of accuracy (97%) was calculated for both geometric interval and natural breaks (Jenks) when the questions were considered to be very easy.

As indicated by the significance scores in Table 2, both geographic units and self-perceived confidence level per question were not identified as statistically significant predictors of response accuracies. In fact, response accuracy was 90% for maps showing population density by sub-place, hexagon and small area layers. For each question, respondents were asked to rate their confidence in their answers on a scale from 0 to 10, with 10 indicating complete confidence. The average and median confidence level per question was 8 overall, as well as for each data classification method individually, except for the logarithmic scale, which had an average confidence level of 7. Additionally, minimum and maximum confidence levels of 0 and 10 were recorded for all data classification methods.

Discussion

For this paper, it was essential to first assess whether choropleth maps are useful for visualising population distribution in South Africa, and secondly, whether choropleth maps are useful to identify over – and underserved areas for geographic accessibility analysis. The results from this study suggest that choropleth maps are indeed an effective and easy-to-use technique for visualising population demand since the average accuracy percentage was 90%. Four respondents scored 100%, and the lowest recorded accuracy was 65% among three individuals. These results correspond to findings by others (Sukraini et al., 2022; Shaito & Elmasri, 2021).

Although the percentage accuracy of responses for each of the four data classification methods was notably high, respondents were more likely to provide correct answers for maps depicting the quantiles data classification method (92%), followed by natural breaks (Jenks) and geometric interval, with 91% and 89% accuracy, respectively. The logarithmic scale was ranked the lowest, with an accuracy of 87%. In a previous study, Brewer and Pickle (2002) assessed seven data classification methods for classifying epidemiological data using choropleth maps. These methods included hybrid equal intervals, quantiles, box plots, standard deviation, natural breaks (Jenks), minimum boundary error, and shared area. Their map interpretation questions were designed in such a way that certain question types were more difficult than others. The findings from their study also indicate that quantiles, followed by minimum boundary error classification methods, were best interpreted by participants. This was followed by natural breaks (Jenks) and a hybrid version of equal intervals (Brewer & Pickle, 2002). These findings could encourage future research to better understand which data classification methods work best for different data types and visualisation tasks. Nevertheless, it would be useful to repeat the experiment with different study areas and users to validate the results about data classification methods.

One would have expected respondents to achieve the highest accuracy percentage for maps depicting the City of Tshwane Metropolitan Municipality since respondents were familiar with the area (Boscoe & Pickle, 2003). Although a high percentage of accuracy was recorded for all study areas, these findings indicate that different data classification methods are sensitive to the selection of study area. Another possibility to consider is that the maps (study areas) were not presented to respondents in a consistent order, which may have helped prevent a potential learning effect.

The size and shape of areas or polygons, also referred to as geographic units, were identified by Robinson (1995) as important elements for choropleth maps. Boscoe and Pickle (2003) further defined the ideal characteristics of geographic units, among others, as: 'a high degree of resolution, homogeneity of population size and land area, minimum population thresholds and land area thresholds, compactness of shape, audience familiarity and functional relevance'. Also, Schiewe (2019) evaluated the visual perception of spatial patterns with choropleth maps based on three effects: 'dark-is-more bias, area-size bias, and data-classification effect''. The results revealed that higher values are indeed associated with darker colours. The analysis also confirmed area-size bias, as 30–40% of participants overlooked smaller areas on the map. However, in this user study, geographic units were not found to be statistically significant predictors, either overall or in relation to any of the four data classification methods. Further research could investigate bias in South African map readers and compare this to international audiences.

Finally, results from this study confirm what Rautenbach et al. (2017) found when developing and evaluating a task taxonomy for spatial planning through a map literacy experiment using topographic maps; namely, that the accuracy score of participants correlated with their self-perceived difficulty level of a question.

Conclusion

The high percentage accuracy achieved during the user study confirms that choropleth maps with classes prepared according to the geometric interval, logarithmic scale, natural breaks (Jenks) and quantiles methods effectively visualise population distribution in South Africa. Results also indicate that one can use any of the four selected data classification methods to analyse supply and demand for the optimal positioning of service centres. However, respondents were more likely to provide correct answers for maps depicting the data according to the quantiles classification method, followed by natural breaks (Jenks) and geometric interval. Logarithmic scale was ranked lowest. A more focussed investigation on why some classification methods performed better or worse and whether this related to the characteristics and distribution of the data could shed light on the underlying reasons for these results. Also, future research could investigate the interaction between the different predictor variables and how they affect accuracy scores of respondents. Even though geographic units were not identified as a statistically significant predictor, the results suggest some variation in response accuracies between the four data classification methods and geographic units. Follow-up investigations could be useful to understand the impact of different geographic units on visualising choropleth maps. Tests could include different localities and geographic units such as voting districts, wards or municipalities.

The findings from our user study are useful for geographic accessibility of any kind of public or private facility, such as police stations, clinics, supermarkets and fast-food outlets. The user study involved the voluntary participation of students. The study could be repeated with actual decision makers who are, in one way or another, involved in decisions informed by population density or use choropleth maps for their decision making. Finally, results emanating from this study could be tested in other countries with similar population distribution characteristics to determine possible similarities and differences based on the use of these data classification methods.

Notes

1. https://www.statssa.gov.za/
2. https://www.eskom.co.za/
3. https://www.esri.com/en-us/home
4. https://qgis.org/

Acknowledgements

The authors are grateful to the participants in the user study, whose committed their time voluntarily and whose contributions made this research possible.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Lourens Snyman http://orcid.org/0000-0002-5715-8904
Serena Coetzee http://orcid.org/0000-0001-8683-8047

References

Andrade, C. (2019). The *P* value and statistical significance: Misunderstandings, explanations, challenges, and alternatives. *Indian Journal of Psychological Medicine*, *41*(3), 210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19

Ashiagbor, G., Ofori-Asenso, R., Forkuo, E. K., & Agyei-Frimpong, S. (2020). Measures of geographic accessibility to health care in the Ashanti Region of Ghana. *Scientific African*, *9*, e00453. https://doi.org/10.1016/j.sciaf.2020.e00453

Bertin, J. (1983). *Semiology of graphics*. The University of Wisconsin Press.

Besançon, L., Cooper, M., Ynnerman, A., & Vernier, F. (2020). An evaluation of visualization methods for population statistics based on choropleth maps. arXiv:2005.00324. https://doi.org/10.48550/arXiv.2005.00324

Bolstad, P. (2012). *GIS fundamental.* Eider Press.
Boscoe, F. P., & Pickle, L. W. (2003). Choosing geographic units for choropleth rate maps, with an emphasis on public health applications. *Cartography and Geographic Information Science*, *30*(3), 237–248. https://doi.org/10.1559/152304003100011171
Brewer, C. (2015). *Designing better maps: A guide for GIS users*. ESRI Press.
Brewer, C. A. (2006). Basic mapping principles for visualizing cancer data using geographic information systems (GIS). *American Journal of Preventive Medicine*, *30*(2), S25–S36. https://doi.org/10.1016/j.amepre.2005.09.007
Brewer, C. A., & Pickle, L. (2002). Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers*, *92*(4), 662–681. https://doi.org/10.1111/1467-8306.00310
Brown, L. A., & Chung, S. Y. (2006). Spatial segregation, segregation indices and the geographical perspective. *Population, Space and Place*, *12*(2), 125–143. https://doi.org/10.1002/psp.403
Całka, B. (2018). Comparing continuity and compactness of choropleth map classes. *Geodesy and Cartography*, *67*(1), 21–34.
Church, R. L., & Murray, A. T. (2009). *Business site selection, location analysis, and GIS*. NJ, John Wiley & Sons.
Clifford, N., Cope, M., Gillespie, T., & French, S. (2016). *Key methods in geography*. SAGE.
Coetzee, S., Smit, J., & Rautenbach, V. (2018). *Perceived supply and demand for GISc knowledge and skills in South Africa*. In Proceedings of the AfricaGEO 2018 Conference, 17-19 September 2018, Johannesburg, South Africa.
Cromley, G. A. (2019). Measuring differential access to facilities between population groups using spatial lorenz curves and related indices. *Transactions in GIS*, *23*(6), 1332–1351. https://doi.org/10.1111/tgis.12577
De smith, M. J., Goodchild, P. M. F., & Longley, P. P. A. (2018). *Geospatial analysis: A comprehensive guide to principles techniques and software tools*. The Winchelsea Press.
Demoraes, F., Gouëset, V., & Bouquet, M. (2020). Using "space-time of action" to assess changes in accessibility: The example of Bogotá between 1993 and 2009. *Cybergeo: European Journal of Geography*. https://doi.org/10.4000/cybergeo.34149
Dent, B. D., Torguson, J., & Hodler, T. W. (2009). *Cartography: Thematic map design*. McGraw-Hill Higher Education.
Department Of Public Service And Administration (DPSA). (2012). *Guideline: Improving geographic access to government service points*. Pretoria. https://www.dpsa.gov.za/dpsa2g/documents/sdot/2012/DPSA_Guideline%20on%20Access%20to%20Service%20Points.pdf.
Evans, I. S. (1977). The selection of class intervals. *Transactions of the Institute of British Geographers*, *2*(1), 98–124. https://doi.org/10.2307/622195
GIS GEOGRAPHY. (2022). *Mapping out the GIS software landscape*. https://gisgeography.com/best-gis-software/.
Green, C. A. (2012). *CSIR guidelines for the provision of social facilities in South African settlements*. CSIR Built Environment.
Green, C. A. (2015). *CSIR guidelines for the provision of social facilities in South African settlements*. Pretoria, CSIR Built Environment. https://www.csir.co.za/sites/default/files/Documents/CSIR%20Guidelines_revised_reprintNov2015.pdf.
Harris, R., Charlton, M., Brunsdon, C., & Manley, D. (2017). Balancing visibility and distortion: Remapping the results of the 2015 UK general election. *Environment and Planning A: Economy and Space*, *49*(9), 1945–1947. https://doi.org/10.1177/0308518X17708439
Hatem, G., Zeidan, J., Goossens, M., & Moreira, C. (2022). Normality testing methods and the importance of skewness and kurtosis in statistical analysis. *BAU Journal-Science and Technology*, *3*, 7.
Ho, A. D., U, Y., & C, C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, *75*(3), 365–388. https://doi.org/10.1177/0013164414548576
Jenks, G. F. (1963). Generalization in statistical mapping. *Annals of the Association of American Geographers*, *53*(1), 15–26. https://doi.org/10.1111/j.1467-8306.1963.tb00429.x

Juergens, C. (2020). Trustworthy COVID-19 mapping: Geo-spatial data literacy aspects of choropleth maps. *KN – Journal of Cartography and Geographic Information*, *70*(4), 155–161. https://doi.org/10.1007/s42489-020-00057-w

Kosara, R., Healey, C. G., Interrante, V., Laidlaw, D. H., & Ware, C. (2003). Thoughts on user studies: Why, how, and when. *IEEE Computer Graphics and Applications*, *23*(4), 20–25. https://doi.org/10.1109/MCG.2003.1210860

Kraak, M.-J., & Ormeling, F. (2011). *Cartography: Visualization of spatial data*. Guilford Publications.

Kraak, M.-J., & Ormeling, F. (2020). *Cartography: Visualization of geospatial data*. CRC Press.

Kraak, M.-J., Roth, R. E., Ricker, B., Kagawa, A., & Le Sourd, G. (2021). *Mapping for a sustainable world*. NY, United Nations.

kwak, S. G., & Park, S.-H. (2019). Normality test in clinical research. *Journal of Rheumatic Diseases*, *26*(1), 5–11. https://doi.org/10.4078/jrd.2019.26.1.5

Leydesdorff, L., & Bensman, S. (2006). Classification and powerlaws: The logarithmic transformation. *Journal of the American Society for Information Science and Technology*, *57*(11), 1470–1486. https://doi.org/10.1002/asi.20467

Lloyd, R., & Steinke, T. (1977). Visual and statistical comparison of choropleth maps. *Annals of the Association of American Geographers*, *67*(3), 429–436. https://doi.org/10.1111/j.1467-8306.1977.tb01152.x

Lucas, K. (2011). Making the connections between transport disadvantage and the social exclusion of low income populations in the Tshwane Region of South Africa. *Journal of Transport Geography*, *19*(6), 1320–1334. https://doi.org/10.1016/j.jtrangeo.2011.02.007

Ma, L., Luo, N., Wan, T., Hu, C., & Peng, M. (2018). An improved healthcare accessibility measure considering the temporal dimension and population demand of different ages. *International Journal of Environmental Research and Public Health*, *15*(11), 2421. https://doi.org/10.3390/ijerph15112421

Monmonier, M. (2018). *How to lie with maps*. University of Chicago Press.

Neutens, T. (2015). Accessibility, equity and health care: Review and research directions for transport geographers. *Journal of Transport Geography*, *43*, 14–27. https://doi.org/10.1016/j.jtrangeo.2014.12.006

O'sullivan, D., & Unwin, D. (2010). *Geographic information analysis*. NJ, John Wiley & Sons.

Rautenbach, V., Coetzee, S., & Çöltekin, A. (2017). Development and evaluation of a specialized task taxonomy for spatial planning – A map literacy experiment with topographic maps. *ISPRS Journal of Photogrammetry and Remote Sensing*, *127*, 16–26. https://doi.org/10.1016/j.isprsjprs.2016.06.013

Robinson, A. H. (1995). *Elements of cartography*. Wiley.

Rodrigue, J.-P., Comtois, C., & Slack, B. (2009). *The geography of transport systems*. Routledge.

Schiewe, J. (2019). Empirical studies on the visual perception of spatial patterns in choropleth maps. *KN – Journal of Cartography and Geographic Information*, *69*(3), 217–228. https://doi.org/10.1007/s42489-019-00026-y

Schiewe, J. (2023). Preserving change information in multi-temporal choropleth maps through an extended data classification method. *The Cartographic Journal*, *61*(2), 1–14.

Shaito, M., & Elmasri, R. (2021, 29 June–2 July). Map visualization using spatial and spatio-temporal data: Application to COVID-19 data. *In Proceedings of the 14th Pervasive Technologies Related to Assistive Environments Conference*, 284–291. https://doi.org/10.1145/3453892.3461336

Slocum, T. A., Mcmaster, R. B., Kessler, F. C., & Howard, H. H. (2014). *Thematic cartography and geovisualization*. Pearson.

Snyman, L., & Coetzee, S. (2024). Measuring geographic accessibility in data poor rural areas by augmenting the road network with a triangular irregular network: A case study in the O.R. Tambo district municipality of the Eastern Cape, South Africa. *Journal of Transport Geography*, *115*, 103808. https://doi.org/10.1016/j.jtrangeo.2024.103808

Statistics south africa. (2012a). Census 2011: How the count was done. Pretoria. http://statssa.gov.za/census/census_2011/census_products/Census_2011_How_the_count_was_done.pdf.

Statistics South Africa. (2012b). Census 2011: Metadata. Pretoria. https://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Metadata.pdf.

Strauss, M. (2019). A historical exposition of spatial injustice and segregated urban settlement in South Africa. *Fundamina*, *25*(2), 135–168. https://doi.org/10.17159/2411-7870/2019/v25n2a6

Sukraini, T. T., Yasa, I., & Wiguna, P. P. K. (2022). Comparing choropleth and graduated symbols: how different map types affect public understanding in COVID-19 map reading in Badung Regency, Bali, Indonesia. Geographia Technica, 17.

Sur, P., Chen, Y., & Candès, E. J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probability Theory and Related Fields*, *175*(1-2), 487–558. https://doi.org/10.1007/s00440-018-00896-9

Tyner, J. A. (2014). *Principles of map design*. Guilford Publications.

Vasilca, D. (2019). How to create an effective thematic map. Bulletin of university of agricultural sciences and Veterinary Medicine Cluj-Napoca. *Horticulture*, *76*(2), 258.

Weir-Smith, G., & Dlamini, S. (2024). Hit the road: Spatial characteristics of labor absorption in South Africa. *The Professional Geographer*, *76*(3), 331–342. https://doi.org/10.1080/00330124.2023.2300801

World BANK. (2018). *Overcoming poverty and inequality in South Africa: An assessment of drivers, constraints, and opportunities*. World Bank Publications.

WORLD BANK. (2022). Inequality in Southern Africa: An Assessment of the Southern African Customs Union. Washington, DC. https://documents1.worldbank.org/curated/en/099125003072240961/pdf/P1649270b73f1f0b5093fb0e644d33bc6f1.pdf.

Zhang, J., Xu, L., Zhang, Y., Liu, G., Zhao, L., & Wang, Y. (2019). An on-demand scalable model for geographic information system (GIS) data processing in a cloud GIS. *ISPRS International Journal of Geo-Information*, *8*(9), 392. https://doi.org/10.3390/ijgi8090392

Zhao, L., Chen, L., Ranjan, R., Choo, K.-K. R., & He, J. (2016). Geographical information system parallelization for spatial big data processing: A review. *Cluster Computing*, *19*(1), 139–152. https://doi.org/10.1007/s10586-015-0512-2