

RESEARCH

Open Access



# A comparison of raster-based point density calculations to vector-based counterparts as applied to the study of food availability

Timothy Mulrooney<sup>1\*</sup> , Samuel Akinnusi<sup>2</sup> , Christopher McGinn<sup>3</sup> , Chima Okoli<sup>4</sup>  and Tony Esimaje<sup>4</sup> 

## Abstract

**Background** Proximity to food sources is one of the quantifiable factors measurable across space impacting diet-related health outcomes. Contemporary research has coined the terms ‘food desert’ and ‘food swamp’, sometimes combined with a poverty component, to highlight disproportionate access to healthy and unhealthy food sources. However, there are various ways to measure this proximity—i.e., food availability in this research. Dollar stores such as Dollar General, Family Dollar, and Dollar Tree are one emerging facet of the food environment that provides healthy and unhealthy food options yet have not fully been studied. With more ways to easily measure food availability within the confines of a GIS, this paper proposes a new raster-based Point Density metric to measure the availability of these Dollar stores. In this study, this raster-based metric was calculated for a 6-county region in central North Carolina and compared to six other availability metrics utilized in food security research. A novel Python-based tool to compute the Jaccard Index between these various availability metrics and a matrix to compare these pairwise Jaccard Index calculations was created for this raster-based metric, which is very easy to derive.

**Results** Using a pairwise Jaccard Index summarized and then averaged in a correlation table, the Point Density measure rated the highest (.65) when compared to 6 other popular vector-based techniques. Our results showed the density metric performed statistically better than Euclidean distance, drive-time, density, and point-in-polygon vector metrics when measuring availability for Dollar stores in Central North Carolina.

**Conclusions** Results reinforce the efficacy of this easy-to-compute metric comparable to vector-based counterparts that require more robust network and/or geoprocessing calculations. Results quantitatively evaluate food availability with an eventual goal of dictating local, regional, and even state-level policy that critically and holistically consider this metric as powerful and convenient metric that can be easily calculated by the lay GIS user and understood by anyone.

**Keywords** Food availability, GIS, Point Density, Vector-based, Raster-based, Food security

\*Correspondence:

Timothy Mulrooney  
tmulroon@ncsu.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

### Introduction

Food security is defined as the state where people have physical access to safe and affordable food, which allows for a healthy and active life. This can mean many different things at various spatial and temporal scales. However, according to the Food and Agriculture Organization (FAO), food security comprises the pillars of food availability, food access, and food utilization with the ultimate goal of food stability. Food availability represents the temporal, physical, and geographical proximity of healthy food to those who need it, while access accounts for those who are physically able to procure available food in various ways, such as individualized vehicular transportation, walking, public transportation, rideshares, etc. In some cases, food availability and access are used interchangeably; however, in this research, food availability is a subset of food access. As a result, food access depends upon availability—highly accessible foods are also highly available, but highly available foods may not be accessible to some people.

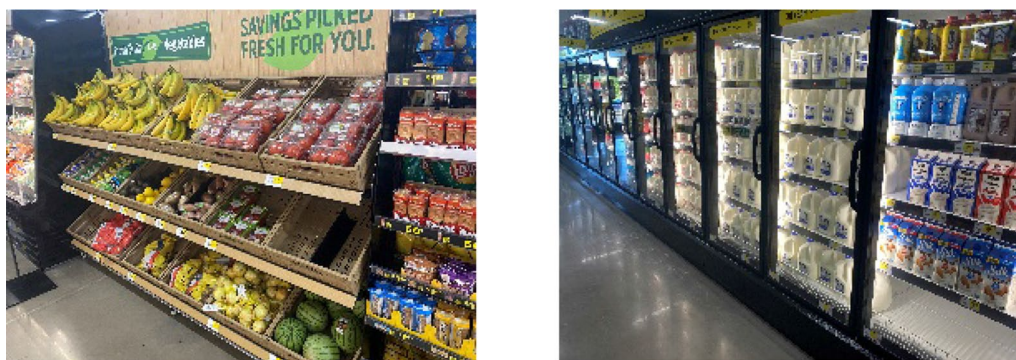
Quantitative methods elucidate relationships between food availability, food access, and food utilization. They help, though not fully explain it using metrics such as proximity, race/ethnicity, poverty status, and access to transportation, among other things. A Geographic Information System (GIS) is a powerful tool to examine quantitative spatial relationships between and among the various agents within the food environment at a local or global scale. These agents include costs of food, source locations (where people are traveling from), and destinations (where people are traveling to) to procure all forms of food, both healthy and unhealthy, as well as socio-economic-health-environmental variables stored within enumeration units. As a result, the application of GIS to food security studies is pervasive among research today [3, 6, 7, 9, 22].

There is a rich body of knowledge on the various ways to measure food availability within the confines of a

GIS. This includes count (number of grocery stores per ZIP code, for example), distance (distance to the nearest grocery store), drive-time (drive-time to the nearest grocery store), and density (number of grocery stores per square mile or population per ZIP code) metrics. Furthermore, unitless metrics such as ratios utilized by the RFEI (Retail Food Environment Index) measure the ratio of unhealthy food outlets versus healthy food outlets by enumeration unit. An accompanying mRFEI (Modified Retail Food Environment Index) represents the percentage of stores within an enumeration unit that are classified as healthy, further articulating the many ways in which availability can be adequately measured. There are many permutations of these metrics as well, which include buffers (number of grocery stores within a distance of a ZIP code, which can also be normalized by population and area), application of distance (Euclidean, Manhattan, or driving) which are increasingly more difficult (time and resource-wise) to compute, aggregation of distance (how many source and destination points are being used) and more complex ratio calculations above and beyond healthy/unhealthy food as per the RFEI and mRFEI.

Dollar stores, considered in this study to be Dollar General, Dollar Tree, and Family Dollar franchises in this study, have gained a foothold in the food environment. Not explicitly represented as supermarket stores in CAB (Commercially Available Business) databases, they appear in areas overlooked by major supermarkets and grocery stores. Many of these dollar stores provide staples such as vegetables, fruits, milk, and eggs, which indicate supermarkets, grocery stores, and a healthy food environment. Between 2009 and 2022, the number of dollar stores has doubled in just the study area alone (45 in 2009, 77 in 2016, and 93 in 2022) (Fig. 1).

While these food availability metrics have obvious utility, *there is little agreement on which metric best aligns with other food availability metrics.* Although there has



**Fig. 1** Healthy and fresh food offerings in Dollar General store within study area [21]

been ongoing discussion regarding the absence of a universally agreed-upon approach/metric for assessing food insecurity or food environment exposure, researchers have commonly employed methods such as ratio and proportion indicators [25, 28, 30] or, Geographic Information System (GIS) techniques [6, 14, 16, 22–24], qualitative methods [13], and more recently real-time information [26]. Furthermore, little work has explored using raster-based density techniques to measure food availability versus their vector-GIS counterparts. In this study, on the backdrop of measuring the availability of dollar stores and using the working hypothesis that a *raster-based point density metric to measure food availability is comparable to traditional vector-based counterparts*, we will explore:

- The development of a model to utilize point density calculations to measure the availability of dollar stores in Central North Carolina at the pixel scale and grouped within census block groups (using the *Point Density* spatial analyst tool).
- The use of this density metric to delineate, assess, and evaluate the most available and least available block groups on the backdrop of socio-economic variables within the study area.
- The comparison of this density-based metric against other traditionally used metrics using statistical techniques (using the *Jaccard Index* tool).

### Literature review

Food availability is largely geographical and can be measured using a GIS (Geographic Information System), which helps create, analyze, and render spatially related information in the digital environment. While studies have explored the physical factors such as climate that impact food security at a very low scale [29, 31], as applied to this study, GIS is used to measure food availability using various methods at much higher scales. These metrics are measured within various polygonal units of various sizes. Counties, subdivisions of states, are typically too coarse of a scale to express local-level food security which this research attempts to do. Finer units include census tracts, subdivisions of counties and census block groups, subdivisions of census tracts. Other units include ZIP (Zone Improvement Plan) codes that are not part of the census. They are smaller than counties, but larger than census tracts. However, they overlap counties which census units do not do.

One such method to measure availability is Euclidean distance, which measures the straight-line distance between a source or the center of an enumeration unit (such as a census tract or block group) and the nearest food source (such as a grocery store). This approach has

been used in several studies, including those by Misiaszek et al. [17], the Economic Research Service (2015), Zenk et al. [33], Lewis et al. [15], and Morris et al. [19], all of which utilized straight-line distance within a GIS to measure food availability. The enumeration units in these analyses vary, where Misiaszek et al. [17], Zenk et al. [33] (2005), and the Economic Research Service (2015) utilize census tracts, Chenarides et al. [9] utilize block groups, Lewis et al. [15] implement the ZIP code, and the USDA Food Access Atlas (2019) uses the centroid of 500-m cells/grids canvassed across a study area.

However, while Euclidean distance is easy to calculate, it does not accurately represent the practical food environment since people do not travel in straight lines to procure food. As a result, more resource-intensive network calculations can derive driving and walking distance/time given sources (places traveling from), destinations (stores to travel towards), and a network of roads or sidewalks with impedances (speed limits or travel time) provide a better representation of the practical food environment. For instance, Pearson et al. [27] and later Morland and Evenson [18] utilize this network distance between individual addresses and food locations, while works by Algert et al. [1], Mulrooney et al. [22] and Mulangu and Clark [20] utilize drive-time metrics. Cervigni et al. [7] measure walk-time and isochrones using these networking tools.

Despite the benefits of network calculations, they also have their challenges, particularly regarding the selection of sources and destinations. While it is easy to define healthy food outlets as destinations, the use of sources from which trips originate can significantly impact the results. Using multiple sources in a GIS can be expensive in terms of time and resources. For instance, in a study by Mulrooney et al. [22], more than 177,000 residential addresses were used as sources to travel to 193 potential destinations in North Carolina. Using Dijkstra's Shortest Path First (SPF) algorithm and a road network with over 98,000 vertices, calculating just one path requires between 177,000 and 9 billion calculations. Various sampling methods exist to approximate sources, including using the population-weighted block group centroid, as demonstrated by Berke and Shi [4] and the USDA Food Access Atlas [10], as well as random points [24] and random point distributions stratified by area and population [12]. A comparison of 291 population-weighted block groups with more than 177,000 individual addresses found both drive time and drive distance were within acceptable tolerances for population-weighted block group centroids using both tests of similarity and dissimilarity via t-tests and tests of equivalence.

In addition to distance-based measures, basic counts for a particular enumeration unit can be calculated

within the confines of a GIS using the *Spatial Join* functionality, which counts the number of food sources represented as points within an enumeration unit. These count values can be compared to other count values or further analyzed using buffers or normalized values. Count values normalized by population, area or combining both techniques with buffers can provide a more granular analysis. For instance, Brown-Amilian [6] found census tracts containing fewer Dollar stores have higher education attainment levels, less racial/ethnic diversity, and more income. Thornton et al. [30] built upon this by looking at the number of destinations within a distance of an enumeration unit, providing a more detailed analysis. Block et al. [5] explored the density of fast-food restaurants within a specific distance of a census tract when normalized by the area of the census tract.

Regardless of the measure, various limitations when utilizing GIS to assess the food environment exist. They include (1) the use of centroids as appropriate proxies for true source locations; (2) the size as well as odd and littoral shapes of enumeration units, especially in eastern North Carolina, which may influence results; (3) the relative location of food sources within an enumeration unit, where a food source near the border of a tract may be patronized by many in another tract but not counted for that tract depending upon the agglomeration method (count) utilized; and (4) the need and manner of normalization used. Nevertheless, their potential to provide spatially explicit information can help identify areas where interventions are needed to address food-related health disparities.

In this study, a density-based raster surface using the point density calculation will be created to assess and evaluate the availability of Dollar stores in Central North Carolina. Raster data are useful for representing continuous phenomena such as elevation or satellite imagery, and in the case of this research, store density. Prior work in this realm has explored distance and/or travel time to a given destination results in a travel-time surface. This aligns with the raster-based food desert analysis previously performed by the research team [24]. The application of cost-based surfaces is not new in studying food security. Yeager and Gatrell [32] developed a travel-time surface for rural Illinois by creating an interpolated travel distance surface. Hallett and McDermot [11] also developed a cost surface, representing the cost in dollars spent to travel to the nearest grocery store based on the IRS value of the cost to operate a motor vehicle (\$0.505/mile). Chen and Clark [8] expressed food access via both raster and 3-D surfaces as a product of spatial access and a store's hours of operation, thus creating food deserts that change diurnally. While other limitations of utilizing raster data in food security analysis may exist, it

will not be constrained by the discrete nature of vector data most often used in food environment analyses using GIS. Using statistical methods, results from this density metric will be compared to the previously utilized and aforementioned measures to determine how and to what degree it compares to vector-based counterparts.

## Materials and methods

### Study area

As part of a larger research project into food availability in North Carolina, we conducted a pilot study in six of North Carolina's central counties, including Alamance, Caswell, Chatham, Durham, Orange, and Person. This study area was selected due to its (1) proximity to the authors' host institution; (2) an area that has a manageable number of Dollar store that could be handled within the scope of this project; and (3) the combination of rural to suburban and urban regions in the study area. The region is known for its strong economy, high quality of life, and thriving culture and arts scene. The region is also racially and ethnically diverse, with a significant proportion of African Americans, Hispanics, and Asian Americans combined with a population of over 700,000 people. The study area has an area of 2675 miles<sup>2</sup> (6936.5 km<sup>2</sup>). The area is home to several major universities, including UNC Chapel Hill, North Carolina Central University, Elon University and Duke University, which provide a highly educated workforce and drive innovation and economic growth. In the study area, there are several malls and shopping centers and a large outdoor shopping complex. For groceries, there are many options, including Target, Food World, Food Lion, Harris Teeter, and Walmart. Family stores are also widely available in all six counties, including Dollar General, Dollar Tree, Family Dollar, and Big Lots. These stores offer a wide range of products at affordable prices, making them a popular choice for families and budget-conscious shoppers (Fig. 2).

### Data collection

The GIS vector data for county boundaries, census tracts, and Dollar stores were used for the spatial analysis in this study. The boundary data were retrieved from the NC OneMap (<http://www.nconemap.gov>), a public repository of spatial data for the state of North Carolina. The Dollar store dataset was extracted from the US business feature class provided by DataAxle. The Dollar stores were extracted by their name (Dollar General, Family Dollar, and Dollar Tree) from all businesses with the study area and a 10-mile buffer within the study area using the *Select by Attributes* and *Select by Location* tools in ArcGIS Pro (v. 3.0) and are current



**Fig. 2** Map of study area

through mid-2022. There are 94 Dollar stores within the study area (163 within the 10-mile boundary), up from 49 in 2009 (98 within 10-mile buffer). There are 420 census block groups in the study area, which range in population from 4 to 9460 and range in area from 0.075 mi<sup>2</sup> (0.194 km<sup>2</sup>) to 63.701 mi<sup>2</sup> (165.182 km<sup>2</sup>).

**Data processing and geostatistical analysis**

All geostatistical analyses were performed with the Esri ArcGIS Pro software with the help of geoprocessing toolsets. In addition to the point density metric, which serves as the focus of this research, six other measures of Dollar store availability analysis were performed and then compared to each other. The geoprocessing and statistical tools from the spatial analyst toolset, network analyst solvers and data management toolset were used

for availability measures in this research. Each metric, described in Table 1, are highlighted below:

- a. Point Density: In this measure, availability is measured to be the density of Dollar stores measured at the pixel level within 3 miles of a particular pixel and then grouped within block groups. The point density surface is generated by calculating the number of points within a specified distance of each pixel location in the study area and then representing the results as a continuous surface (raster layer). The *Spatial Analyst* extension of ArcGIS Pro software calculates the density of point features around each output raster cell to define a neighborhood around each raster cell. We utilized the point density spatial analyst tool to calculate the magnitude of dollar stores per unit area within this 3-mile neighborhood around each raster cell. In order to compare it to other avail-

**Table 1** A comparison figures for all the availability measures

Measure	Point Density	Drive Time	Join Count	Buffer (3 Miles)	Euclidean distance	Store density by area	Store density by population
Point Density	–	0.70	0.22	0.78	0.71	<b>0.81</b>	0.69
Drive Time	0.70	–	0.25	0.60	0.73	0.69	0.61
Join Count	0.22	0.25	–	0.22	0.25	0.19	0.22
Buffer (3 Miles)	0.78	0.60	0.22	–	0.59	0.70	<b>0.81</b>
Euclidean Distance	0.71	0.73	0.25	0.59	–	0.70	0.56
Store density by area	<b>0.81</b>	0.69	0.19	0.70	0.70	–	0.70
Store density by population	0.69	0.61	0.22	<b>0.81</b>	0.56	0.70	–
Average (AVG)	0.65	0.60	0.23	0.62	0.59	0.63	0.60
AVG without Join Count Jaccard Index	0.74	0.67	–	0.7	0.66	0.72	0.67

Bold represents highest Jaccard Index

ability metrics collected at the census block group level, the value of the resultant raster was extracted using the *Zonal Statistics* tool, also within the *Spatial Analyst* extension. This *Spatial Analyst* extension is a suite of tools focused explicitly on raster data calculation. Each census tract was assigned an average point density value for each of the pixels contained within it by joining the zonal statistics table to the census block group feature class. The resulting metric is a density value based on this average pixel density and is visualized in Fig. 6.

- b. Drive Time: In this metric, availability is measured at the block group scale to be the drive-time between the block group centroid and the nearest Dollar store. The *Closest Facility* calculation within the *Network Analyst* toolbar was used to calculate the drive-time between each source (420 block group centroids) and the nearest of possible destination representing the 164 Dollar stores within 10 miles of the study area. This result is a drive-time calculation in minutes for each block group.
- c. Join Count: In this metric, availability is measured to be the number of Dollar stores located within a census block group. This approach merely involves counting the number of Dollar stores within a census block group in the study area using the *Spatial Join* processing tool. Several researchers have adopted this method for food desert availability and accessibility measure [2, 6], and the resulting measure is simply a number, representing the number of stores in the block group.
- d. Buffer (3 miles): In this metric, the availability for a block group is calculated to be the number of Dol-

lar stores within 3 miles of a block group (as well as those within the block group). The *Spatial Join* tool was implemented; however, a search radius of 3 miles was specified in the *Spatial Join* parameters. The resulting measure is simply a number, representing the number of stores within the block group as well as the 3-mile buffer.

- e. Euclidean distance: In this measure, availability is calculated to be the Euclidean (straight-line) distance between a block group centroid and the nearest Dollar store. This was done using the *Near* geoprocessing function which calculates the distance between input features (block group centroids) and near features (Dollar stores). The resulting metric is a distance in miles.
- f. Store density by area: In this measure, availability at the block group level is measured to be the number of stores within a 3-mile area of the block group (Method d) normalized by the area of the block group. This method filters out larger block groups who may have high buffer values based solely on its size and the result is represented as the number of Dollar stores per square mile.
- g. Store density by population: In this measure, availability at the block group level is measured to be the number of stores within a 3-mile area of the block group (Method d) normalized by the population of the block group. This method filters out regions that may have more Dollar stores because they have higher populations, and the result is represented as the number of Dollar stores per 1000 population of the block group.

A summary of these calculations is below:

Metric	Calculation
Point Density	Tool: <i>Point Density</i> Calculates a magnitude-per-unit area from point features (dollar stores) that fall within a neighborhood around each cell $Point\ density = \frac{\#ofDS}{A}$ # of DS = number of Dollar stores within neighborhood of each output raster cell A = area of the neighborhood
Store density by area	Tool: <i>Calculate Field (on the attribute table)</i> $Store\ density\ by\ area = \frac{\#ofDS}{A}$ # of DS = number of Dollar stores within a block group A = area of each block group (in square miles)
Store density by population	Tool: <i>Calculate Field (on the attribute table)</i> $Store\ density\ by\ population = \frac{\#ofDS}{P}$ # of DS = number of Dollar stores within a block group P = population of each block group
Join Count	Tool: <i>Spatial Join</i> $Join\ count = \#of\ DS\ within\ a\ block\ group$
Buffer 3 miles	Tool: <i>Spatial Join</i> $Join\ count = \#of\ DS\ within\ 3\ miles\ of\ a\ block\ group$
Euclidean distance	Tool: <i>Near (Analysis)</i> Calculate distance between input feature in one layer (block group centroid) and closest feature in another layer (dollar store) Calculation rule: The distance between two points is the straight line connecting the points
Drive Time	Tool: <i>Closest Facility Solver (Network Analyst)</i> Calculation rule: Finds the one facility (dollar store) that is closest to a source (block group centroid) based on travel time using best driving routes

**Standardization of data**

A major goal of this project is to test the efficacy of a new metric (Method a) to measure food availability compared to proven and existing measures (Methods b–g). Given their varying units of measure, simple change detection techniques (subtracting the value of one from another and mapping or analyzing their differences, for example) between each of the metrics are not feasible. Furthermore, while units of measure each and unto themselves have powerful computational value, they have little value to the lay user. As a result, for each metric, every block group is assigned one of three values (Most Available, Least Available, Neither) based on the quintile classification of that particular metric. For example, Point Density (Method a) values for the 420 block groups range from

0 to 0.527214. The ‘Least Available’ block groups are denoted as the 84 (420 ÷ 5) block groups with the lowest values which range from 0 to 0.020521. The ‘Most Available’ block groups are the 84 block groups with the highest point density values whose values range from 0.310435 to 0.527214. The remaining 252 block groups are classified as ‘Neither’ for that metric. This was repeated for the six other metrics. Most of these classes were fairly easy to extract except for the Join Count (number of stores within a block group) method. The result of the count analysis had only five values ranging from 0 to 4. This lack of granularity saw exactly 84 block groups with one or more Dollar store, which were classes as ‘Most Available.’ The remaining 336 block groups with no Dollar stores are classified as ‘Least Available’ while no block groups are classified as ‘Neither’ using this method.

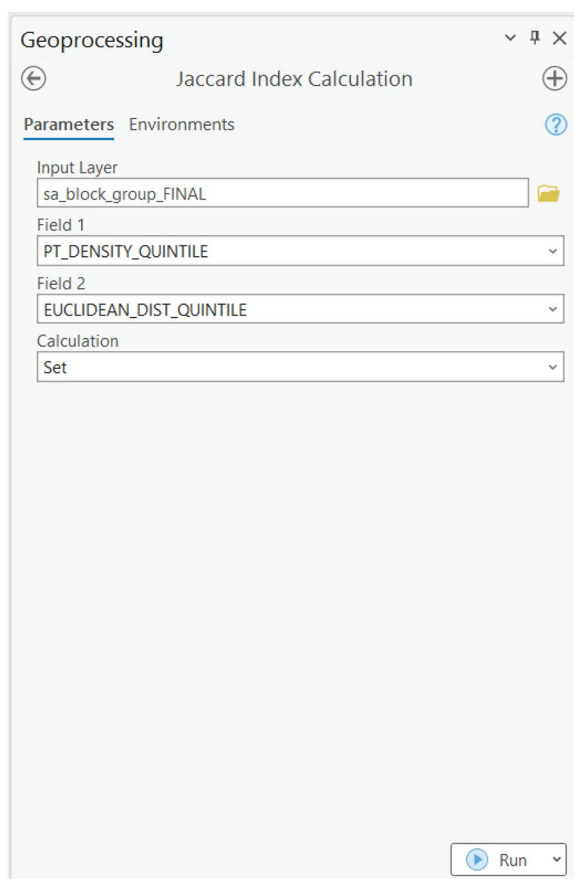
**Comparative analysis using the Jaccard Index**

The Jaccard Index (JI), also known as the Jaccard similarity coefficient or Jaccard similarity index, is a statistic used to measure the similarity between two data sets. It is calculated as the ratio of the intersection between two sets versus their union. The Jaccard Index ranges from 0 to 1, with higher values indicating greater similarity between the two sets. The formula for calculating the Jaccard Index is:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

where *A* and *B* are two sets of data, and in this case, class values (Least Available, Most Available, Neither) derived from the different availability metrics highlighted in Methods a through g.  $\cap$  represents the intersection of the two sets (values in common) while  $\cup$  represents the union of the two sets (420).

While the Jaccard Index can also account for binary vectors (Least Available and Null, for example) which will change size of the union, this analysis will utilize 420 as the union value since all block groups have been assigned a value and there are no Null values. While the Jaccard Index can be calculated by running some *Select by Attributes* queries and dividing by the total number of block groups (420) representing the Union of two metrics, the research team created a custom Jaccard Index Calculation tool using the in-built Python toolbox template to derive input parameters (input datasets, attributes, type of calculation) and custom Python code to run the calculations and output the results. Our Jaccard Index Calculation tool is an asset to any researcher or professional



**Fig. 3** Jaccard Index calculation tool

seeking to analyze and understand similarities between fields in their data (Fig. 3).

This Python-programmed ArcGIS-based Jaccard Index calculation tool has been used by the research team for the comparative analysis on the varying definitions of urban [23].

## Results

The Point Density metric (Fig. 5) utilizing little-used *Spatial Analyst* (in the food environment realm) tools was created and then grouped into 420 block groups in Central North Carolina (Fig. 6). By appearances, it appears much like its vector-based counterparts. Six other popular food availability metrics taken from prior research works and calculated in the vector GIS environmental were calculated as well, and all 420 block groups in study area were classified as ‘Least Available’, ‘Most Available’ or ‘Neither’ based on a simple quintile classification of

each of the metrics since simple change detection analysis techniques are not possible. Pairwise Jaccard Index calculations (Point Density vs. Euclidean Distance, for example) were performed between each metric and its six counterparts.

Table 1 represents the Jaccard Index values between all seven of the difference measures. Values closer to 1 represent higher agreement or similarity between the method of measures while values closer to 0 represent weak similarity between methods. For example, the *Point Density* and *Store Density by Area* metrics agreed with each other for 81% (tied of the highest between all 21 of the pairwise calculations) of the 420 block groups across the ‘Least Available’, ‘Most Available’ and ‘Neither’ classifications while *Euclidean Distance* and *Store Density by Population* agree with each other for 56% of the study area’s 420 block groups (Figs. 4, 5).

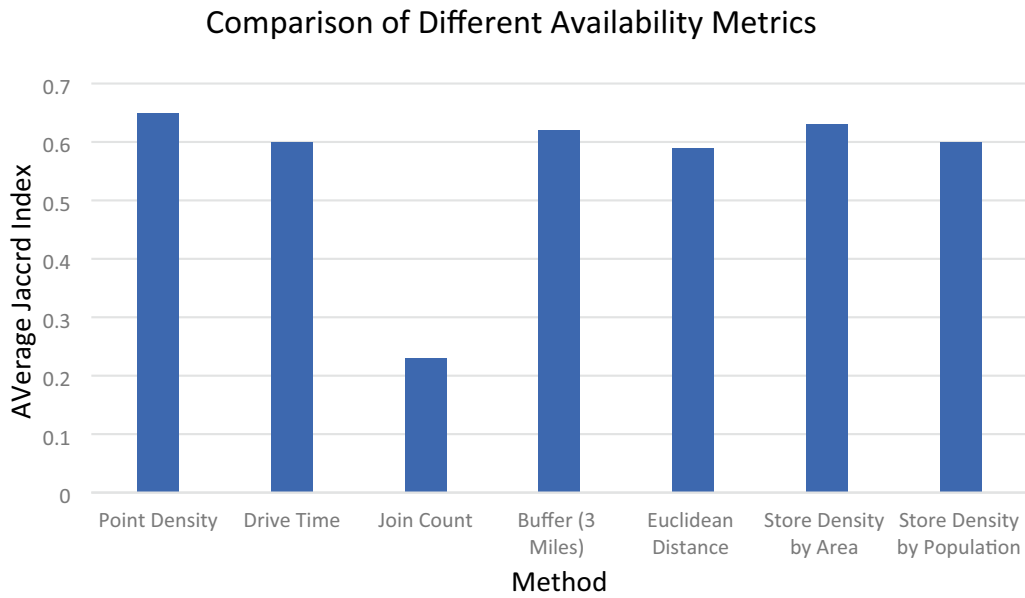
The Jaccard Indices for each of these pairwise calculations were averaged for each column/measure, resulting in the metric that best agreed with its six counterparts. Based on this, a general observation shows the Point Density outperformed other measures of availability adopted in this study. By far the Join Count method has the poorest performance with an average JI value of 0.23. However, this poor performance is due to the way the Join Count data were classified as either ‘Least Available’ or ‘Most Available’, with no ‘Neither’ classes assigned due to the lack of granularity with values. Even when the average of this Join Count outlier is removed from each of the metrics and the average is recomputed, the Point Density metric compares well to the other five food availability counterparts. These are highlighted in Figs. 2 and 3.

## Discussion

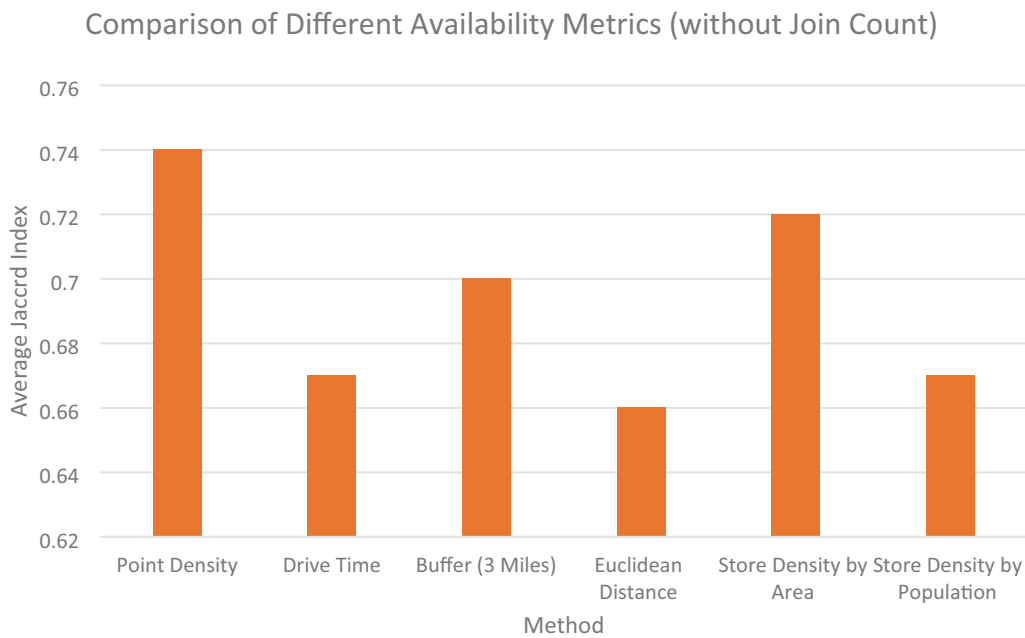
While often conflated with the concept of food access, the notion of food availability is largely geographical in nature and represents the proximity of food sources to a location. Food availability serves as one of the pillars of food security and is one that can easily be measured across place and space within the confines of a Geographical Information System (GIS).

While the research team is satisfied with these analyses and results, it is imperative to note the methods employed in this research to measure Dollar store availability were largely influenced by individual tool parameters, limitations, and choices by the research team. These influences include:

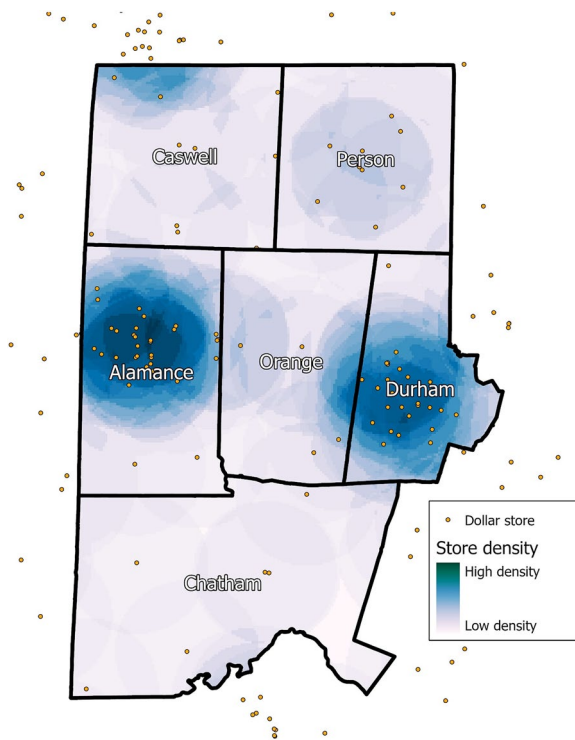




**Fig. 4** Performance of available measures based on average Jaccard Index values



**Fig. 5** Performance of available measures based on average Jaccard Index values excluding Join Count values



**Fig. 6** Dollar store point density surface

Influence	Explanation
Use of 3-class system	Measures of availability were created using various units of measure (drive-time in minutes, # of stores, # of stores per square mile, etc.) and converted to classes of 'Least Available', 'Most Available' or 'Neither' based on the quintiles of these units of measure. Since all block groups contained a value, a set comparison for the Jaccard Index was run as opposed to a <i>binary vector</i> where block groups take on only two values: '1' or '0' or in this case 'Least Available' or 'Null'. While there is a consistent number of block groups (420) in the union for set calculations, the number of block groups from the union of two binary vectors is variable depending upon the number of non-null block groups. Since measuring availability, either good or bad, served as the focus of this paper, sets were used to highlight the importance of retaining the most and least available block groups. As a result, the set calculation for the Jaccard Index was used instead of the binary vector which essentially measures only one category
Use of the Join Count Jaccard Index calculation	Values only ranged from 0 to 4 (Dollar stores located within the 420 polygonal block groups). As a result, the 84 block groups which contained a Dollar store were classified as 'Most Available' while all others were classified as 'Least Available'. As a result, there were no 'Neither' block groups in this method. Only 7 block groups contained more than one Dollar store and 77 contained exactly one Dollar store. While slightly different classes could've been created with this configuration (7—'Most Available', 77—'Neither' and 336 'Least Available'), both deviated significantly from the quintile configuration for other metrics, resulting in low Jaccard Indices. Because of this, a separate Jaccard Index average was calculated removing this outlier
Use of 3-Mile Buffer Length	This 3-mile buffer serves as a happy medium between the 1-mile and 10-mile buffers used to denote Low Access to urban and rural census tracts, respectively, by the USDA Food Access Atlas. However, research by Gallagher (2014), Schlundt et al. (2017) and Barnes et al. (2015) explicitly utilized 3-mile buffers as measures of food availability in their research

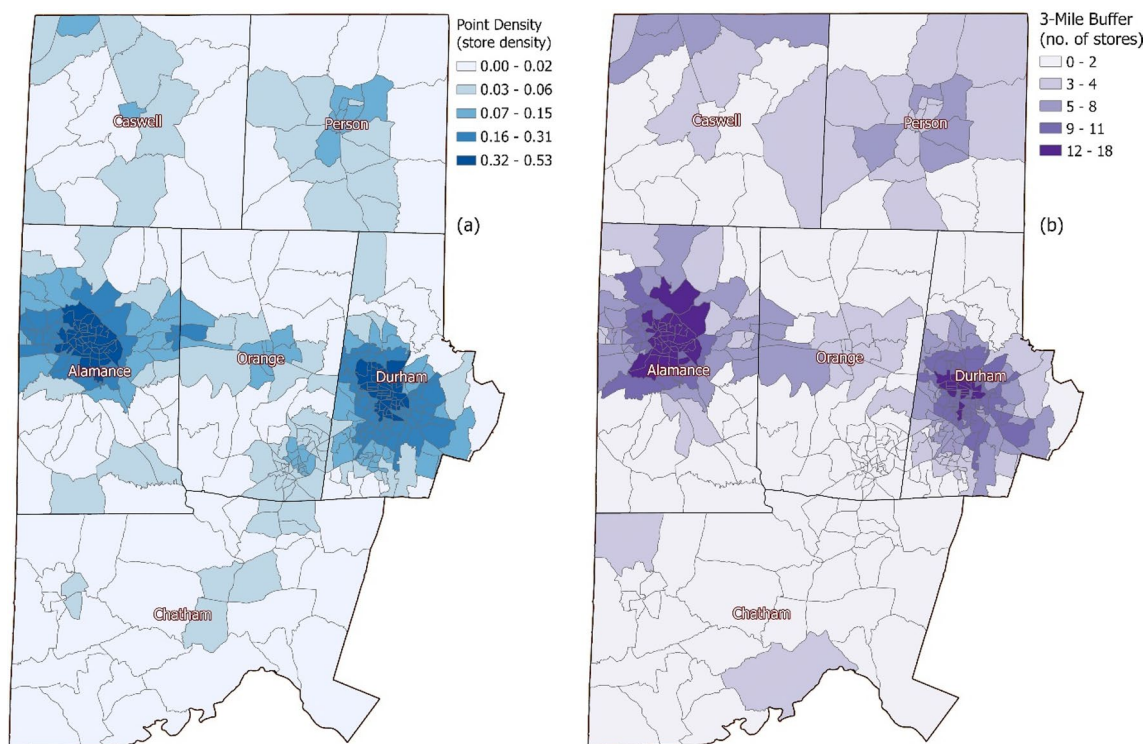
Influence	Explanation
The use of the centroid as a source	For drive-time and Euclidean distance calculations for availability, source locations were derived from block group centroids, where drive-time and Euclidean distance were calculated from these sources to the nearest destination (Dollar store), respectively. Other sources do exist. While individual address locations extracted from parcel data can be utilized as sources and these calculations run for each address and grouped at the block group level, they are computationally expensive to run and may not run on desktop computers. Research by Mulrooney et al. [22] showed population-weighted block groups serve as acceptable proxies for these individual addresses without compromising results while decreasing the number calculations by three orders of magnitude (291 source locations vs. 177,000 in their study) and may have an impact on research results over the geographic centroid

In support of this research, an ad hoc tool was developed by the research team to run a pairwise Jaccard Index between two attributes. It consisted of an interface using ArcGIS Pro tool builder requesting four parameters: input feature class, class attribute #1, class attribute #2 and type of Jaccard Index calculation (binary vector or set). Underlying custom Python code calculates the Jaccard Index and outputs the results. While this Jaccard Index could be calculated using the *Select by Attributes* functionality and hand-calculations, the research team foresees the utility of nominal and categorial attribute comparisons across fields such as biogeography, agriculture, remote sensing, environmental science, sociology and criminal justice, and plans to develop a custom tool to perform this within the vector and raster data environments.

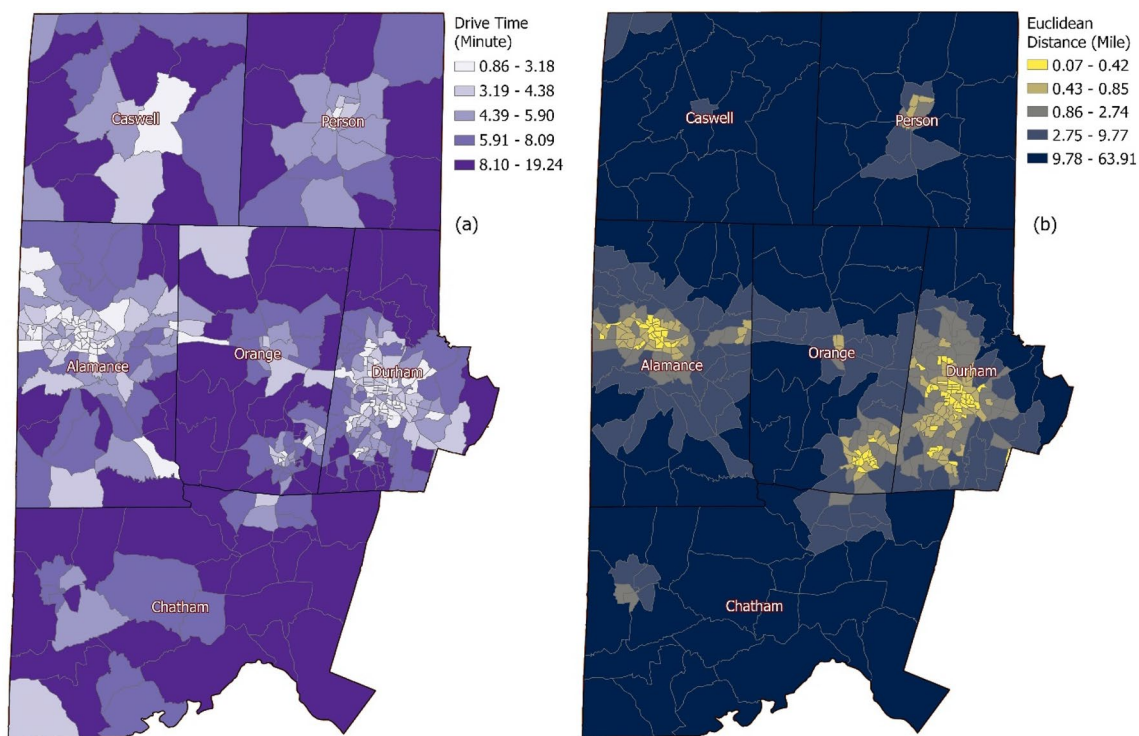
### Conclusions

In this study, the availability of Dollar stores such as Dollar General, Family Dollar and Dollar Tree was calculated using traditional vector techniques, as well as the introduction of a raster-based density calculation. Dollar stores, which serve as source of food, were chosen because of their adequate sample size and ubiquitous nature across urban/suburban/rural landscapes within a 6-county study area in central North Carolina, home

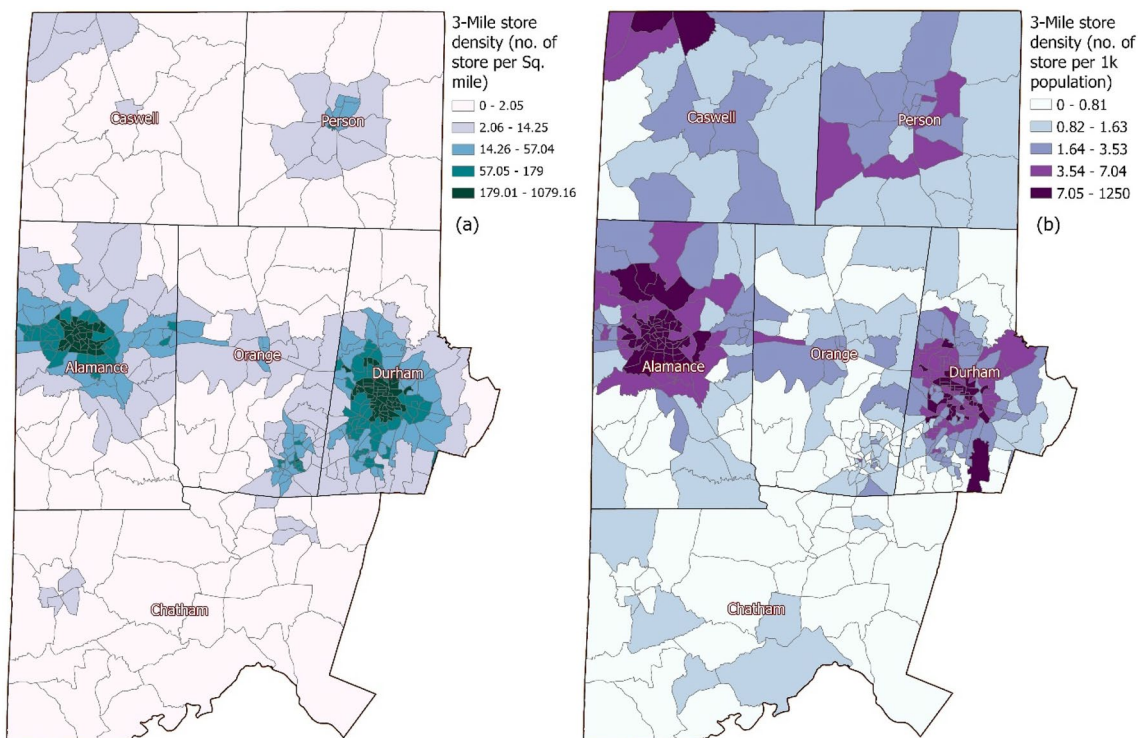
to more than 700,000 people. This raster-based density metric created for the study area essentially measures the density of Dollar stores within the study area as well as those within a 10-mile buffer of the study area. This was done because people living within the study area may be ‘closer,’ however that is defined for each of the metrics, to Dollar stores that are outside of the study area. The resulting density surface (Fig. 6) was grouped into census block groups (Fig. 7) and block groups were classified as ‘Least Available,’ ‘Most Available’ or ‘Neither’ (Fig. 10) based on a simple quintile classification of the resulting density metric. Other availability metrics such as drive-time to the closest Dollar store (Fig. 8a), a Join Count (Fig. 8b) statistic which basically counts the number of Dollar stores within each block group and areal density (Fig. 9a) which represents the density (# of stores within 3 miles of a block group per square mile) of Dollar stores were also calculated. Since each measure elicits its own distinct unit of measure that do not allow for simple comparison, each block group was classified as ‘Least Available,’ ‘Most Available’ or ‘Neither’ based on the aforementioned quintile classification (Fig. 10). A custom Python tool was created by the research team where a pairwise Jaccard Index which calculates the percent of agreement (via a value between 0 and 1) between the classes for each measure was computed and all 21



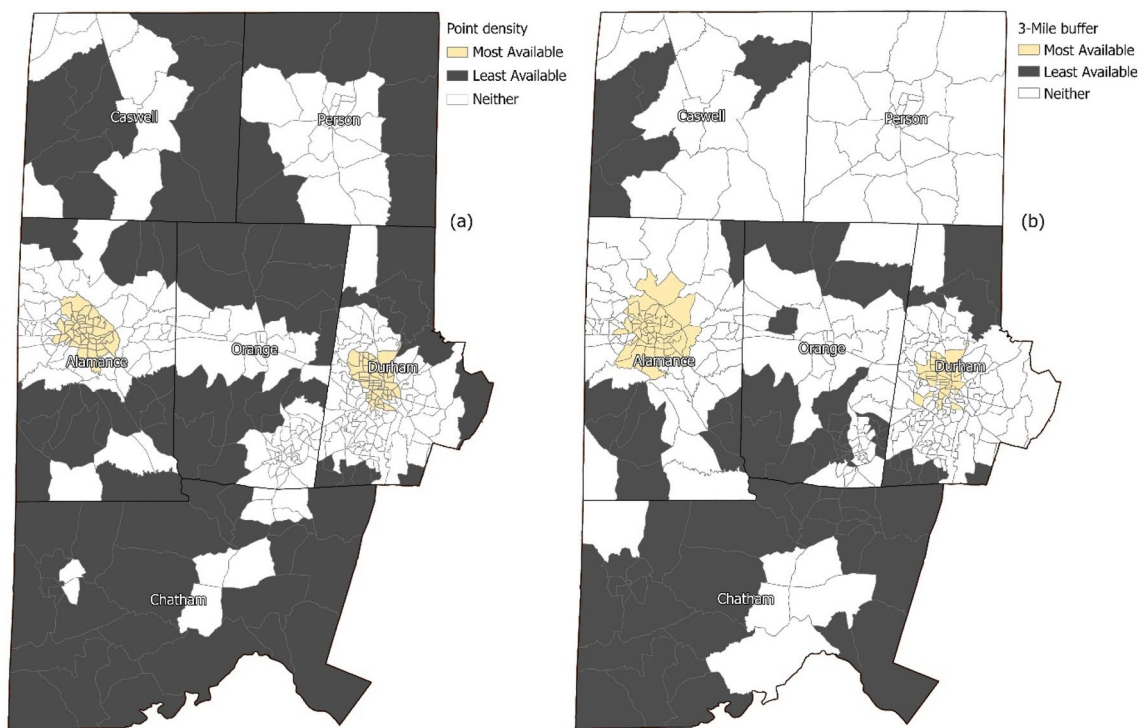
**Fig. 7** Spatial variation in food store availability based on point density (a) and 3-mile buffer (b)



**Fig. 8** Spatial variation in food store availability based on drive time (a) and Euclidean distance (b)



**Fig. 9** Spatial variation in food store availability based on store density by area (a) and store density by population (b)



**Fig. 10** Spatial variation of food store availability based on Jaccard availability metric (least available, most available, and neither) for point density and buffer methods

of these pairwise calculations which were subsequently placed into a resulting table and further summarized (Table 1). The Point Density metric performed slightly better than vector-based counterparts, even when outliers were removed. In summary, major results highlight:

- A density-based metric to measure food availability is easy to calculate and does not require more robust network calculations such as drive-time and drive-distance, geoprocessing calculations such as the Join Count and Buffer nor field operations such as density (by area or population) metrics.
- Using a pairwise Jaccard Index summarized and then averaged in a correlation table (Table 1), the Point Density measure rated the highest (0.65) when compared to 6 other popular vector-based techniques. Given the lack of granularity with the Join Count statistic which created coarse classifications, a new average Jaccard Index was calculated without the Join Count Jaccard Index. Even then, the average Jaccard Index for this metric (0.74) rated higher than its other 5 counterparts, including Drive-Time (0.67), Buffer (0.70), Euclidean Distance (0.66), Store Density by Area (0.72) and Store Density by Population (0.67).

Ancillary results from this research highlighted of the six counties in the study area, Alamance County has the best access to Dollar stores according to this Point Density metric. This is interesting because Alamance County has both a higher density (0.23 vs. 0.21) and even more Dollar stores (34 vs. 28) than Durham County, which has a population almost twice that of Alamance County. This county is situated between the larger cities of Greensboro and Durham, and is the subject for future research at a higher scale.

While further work may want to align these spatial relationships with socio-economic variables and long-term health outcomes at the block group level, this research highlights the efficacy and utility of easy-to-use density-based availability metrics not traditionally used in the spatial representation of the food environment. This metric does not require robust network calculations such as drive-time calculations and provides more granularity than simple point-in-polygon and even buffer calculations resulting from the *Spatial Join* operation. Insights ad.

Future work which quantitatively evaluates food availability with an eventual goal of dictating local, regional, and even state-level policy should critically and

holistically consider this metric as powerful and convenient metric that can be easily calculated by the lay GIS user and understood by anyone.

#### Acknowledgements

The authors wish to thank the reviewers for their meaningful comments which led to the overall improvement of this work.

#### Author contributions

TM: conceptualization, methodology, formal analysis using ArcGIS Pro (Version 3.0) software, investigation, writing—original draft preparation, writing—review and editing, funding acquisition, resources, supervision. SA: methodology, validation, formal analysis using ArcGIS Pro (Version 3.0) software, investigation, writing—original draft preparation, writing—review and editing, visualization. CM: writing—review and editing, supervision; TE: writing—Review and editing. CO: writing—review and editing.

#### Funding

This project received funding from the USDA National Institute of Food and Agriculture (NIFA) under the Agricultural and Food Research Initiative Competitive Program, grant number 2021-67021-34152. This work was also supported by the National Science Foundation under Grant No. 2226312 and NASA Award #22-MUREPDEAP-0002. The opinions, findings, conclusions, and recommendations expressed in this material are solely those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

#### Availability of data and materials

The datasets supporting the conclusions of this article are available in the Open Science Framework repository at <https://doi.org/10.17605/OSF.IO/BFDXZ>.

#### Declarations

##### Ethics approval and consent to participate

This research utilized no individualized data and did not require the permission or consent of an Internal Review Board (IRB). All store location data were provided by DataAxle while population data and census boundary outlines were provided via the United States Census.

##### Consent for publication

All authors consent to the publication of this manuscript and further affirm this manuscript is not currently under review with any other journals.

##### Competing interests

All authors certify that they have no affiliations with or involvement in any organization, industry or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript. The authors do not have a stake, financial or otherwise, in the industries related to the aforementioned research topic.

##### Author details

<sup>1</sup>Department of Environmental, Earth and Geospatial Sciences, North Carolina Central University, 1246 Mary Townes Science Complex, 1801 N. Fayetteville Street, Durham, NC 27707, USA. <sup>2</sup>Department of Environmental, Earth and Geospatial Sciences, North Carolina Central University, 1252 Mary Townes Science Complex, 1801 N. Fayetteville Street, Durham, NC 27707, USA. <sup>3</sup>Department of Environmental, Earth and Geospatial Sciences, North Carolina Central University, 1249 Mary Townes Science Complex, 1801 N. Fayetteville Street, Durham, NC 27707, USA. <sup>4</sup>Department of Environmental, Earth and Geospatial Sciences, North Carolina Central University, 1237 Mary Townes Science Complex, 1801 N. Fayetteville Street, Durham, NC 27707, USA.

Received: 21 June 2023 Accepted: 16 November 2023

Published online: 19 January 2024

#### References

1. Algert S, Agrawal A, Lewis D. Disparities in access to fresh produce in low-income neighborhoods in Los Angeles. *Am J Prev Med.* 2005;30(5):365–70.
2. Barnes TL, Colabianchi N, Hibbert JD, Porter DE, Lawson AB, Liese AD. Scale effects in food environment research: implications from assessing socioeconomic dimensions of supermarket accessibility in an eight-county region of South Carolina. *Appl Geogr.* 2016;68:20–7. <https://doi.org/10.1016/j.apgeog.2016.01.004>.
3. Benez-Secanho FJ, Miner J, Dwivedi P. Using advanced spatial statistical analyses to determine socio-economic constructs of fresh food availability in Georgia, United States. *J Agric Food Res.* 2021;6: 100204. <https://doi.org/10.1016/j.jafr.2021.100204>.
4. Berke EM, Shi X. Computing travel time when the exact address is unknown: a comparison of point and polygon ZIP code approximation methods. *Int J Health Geogr.* 2009;8:23. <https://doi.org/10.1186/1476-072X-8-23>.
5. Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: a geographic analysis. *Am J Prev Med.* 2004;27:211–7.
6. Brown-Amilian S. Dollar store access in the St. Louis metropolitan area, MO-IL, USA. *Papers Appl Geogr.* 2022;8(4):483–92. <https://doi.org/10.1080/23754931.2022.2071128>.
7. Cervigni E, Renton M, Haslam McKenzie F, Hickling S, Olaru D. Describing and mapping diversity and accessibility of the urban food environment with open data and tools. *Appl Geogr.* 2020;125(July): 102352. <https://doi.org/10.1016/j.apgeog.2020.102352>.
8. Chen X, Clark J. Interactive three-dimensional geovisualization of space-time access to food. *Appl Geogr.* 2013;43:81–6.
9. Chenarides L, Cho C, Nayga RM, Thomsen MR. Dollar stores and food deserts. *Appl Geogr.* 2021;134(2020): 102497. <https://doi.org/10.1016/j.apgeog.2021.102497>.
10. Economic Research Service (ERS), U.S. Department of Agriculture (USDA). (2019). Food Access Research Atlas. <https://www.ers.usda.gov/data-products/food-access-research-atlas/>
11. Hallett L, McDermott D. Quantifying the extent and cost of food deserts in Lawrence, Kansas, USA. *Appl Geogr.* 2011;31:1210–5.
12. Hillson R, et al. Stratified sampling of neighborhood sections for population estimation: a case study of Bo City, Sierra Leone. *PLoS ONE.* 2015;20(7): e0132850.
13. Hubley TA. Assessing the proximity of healthy food options and food deserts in a rural area in Maine. *Appl Geogr.* 2011;31(4):1224–31. <https://doi.org/10.1016/j.apgeog.2010.09.004>.
14. Kuai X, Zhao Q. Examining healthy food accessibility and disparity in Baton Rouge, Louisiana. *Ann GIS.* 2017;23(2):103–16. <https://doi.org/10.1080/19475683.2017.1304448>.
15. Lewis LB, Sloane DC, Nascimento LM, Diamant AL, Guinyard JJ, Yancey AK, et al. African Americans' access to healthy food options in south Los Angeles restaurants. *Am J Public Health.* 2005;95(4):668–73.
16. Mathenge M, Sonneveld BGJS, Broerse JEW. Mapping the spatial dimension of food insecurity using GIS-based indicators: a case of Western Kenya. *Food Security.* 2023;15(1):243–60. <https://doi.org/10.1007/s12571-022-01308-6>.
17. Misiaszek C, Buzogany S, Freishtat H. Baltimore City's Food Environment: 2018 Report; 2018.
18. Morland KB, Evenson KR. Obesity prevalence and the local food environment. *Health Place.* 2009;15(2):491–5. <https://doi.org/10.1016/j.healthplace.2008.09.004>.
19. Morris PM, Neuhauser L, Campbell C. Food security in rural America: a study of the availability and costs of food. *J Nutr Educ.* 1992;24(Supplement 1):525–585. [https://doi.org/10.1016/S0022-3182\(12\)80140-3](https://doi.org/10.1016/S0022-3182(12)80140-3).
20. Mulangu F, Clark J. Identifying and measuring food deserts in rural Ohio. *J Extension.* 2012;50(3):41.
21. Mulrooney T. Dollar General in Alamance County, NC. 2023.
22. Mulrooney T, Foster R, Jha M, Beni LH, Kurkalova L, Liang CL, Miao H, Monty G. Using geospatial networking tools to optimize source locations as applied to the study of food availability: a study in Guilford County,

- North Carolina. *Appl Geogr.* 2021;128: 102415. <https://doi.org/10.1016/j.apgeog.2021.102415>.
23. Mulrooney T, Liang CL, Kurkalova LA, McGinn C, Okoli C. Quantitatively defining and mapping rural: a case study of North Carolina. *J Rural Stud.* 2023;97(2022):47–56. <https://doi.org/10.1016/j.jrurstud.2022.11.011>.
  24. Mulrooney T, McGinn C, Branch B, Madumere C, Ifediora B. A new raster-based metric to measure relative food availability in rural areas: a case study in Southeastern North Carolina. *Southeast Geogr.* 2017;57(2):151–78. <https://doi.org/10.1353/sgo.2017.0015>.
  25. Murrell A, Jones R. Measuring food insecurity using the food abundance index: implications for economic, health and social well-being. *Int J Environ Res Public Health.* 2020;17(7):2434. <https://doi.org/10.3390/ijerph17072434>.
  26. Nwankwo W, Ukhurebor K. Big data analytics: A single window IoT-enabled climate variability system for all-year-round vegetable cultivation. *IOP Confer Seri Earth Environ Sci.* 2021;655: 012030. <https://doi.org/10.1088/1755-1315/655/1/012030>.
  27. Pearson T, Russell J, Campbell MJ, Barker ME. Do 'food deserts' influence fruit and vegetable consumption? – a cross-sectional study. *Appetite.* 2005;45:195–7.
  28. Salari M, Reyna M, Kramer M, Taylor H, Gari C. Food desert assessment: an analytical framework for comparing utility of metrics and indices; case study of key factors, concurrences, and divergences. *SSRN Electron J.* 2021. <https://doi.org/10.2139/ssrn.3823677>.
  29. Siloko IU, Ukhurebor KE, Siloko EA, Enoyozo E, Bobadoye AO, Ishiekwene CC, Uddin OO, Nwankwo W. Effects of some meteorological variables on Cassava Production in Edo State, Nigeria via density estimation. *Sci Afr.* 2012;13: e00852. <https://doi.org/10.1016/j.sciaf.2021.e00852>.
  30. Thornton LE, Lamb KE, White SR. The use and misuse of ratio and proportion exposure measures in food environment research. *Int J Behav Nutr Phys Act.* 2020;17(1):1–7. <https://doi.org/10.1186/s12966-020-01019-1>.
  31. Ukhurebor KE, Aidonjio PA. The influence of climate change on food innovation technology: review on topical developments and legal framework. *Agric Food Secur.* 2021. <https://doi.org/10.1186/s40066-021-00327-4>.
  32. Yeager CD, Gatrell JD. Rural food accessibility: An analysis of travel impedance and the risk of potential grocery store closures. *Appl Geogr.* 2014;53:1–10.
  33. Zenk SN, Schulz AJ, Hollis-Neely T, Campbell RT, Holmes N, Watkins G, Nwankwo R, Odoms-Young A. Fruit and vegetable intake in African Americans: income and store characteristics. *Am J Prev Med.* 2005;29(1):1–9. <https://doi.org/10.1016/j.amepre.2005.03.002>.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

