Metadata of the chapter that will be visualized in SpringerLink

Book Title	Geographical Information Systems Theory, Applications and Management
Series Title	
Chapter Title	Digital High-Scale Food Security Analysis: Challenges, Considerations and Opportunities
Copyright Year	2021
Copyright HolderName	Springer Nature Switzerland AG

Corresponding Author	Family Name	**Mulrooney**
	Particle	
	Given Name	**Timothy**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	North Carolina Central University
	Address	Durham, NC, USA
	Email	tmulroon@nccu.edu
	ORCID	http://orcid.org/0000-0001-9333-9641
Author	Family Name	**Wooten**
	Particle	
	Given Name	**Tysean**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	North Carolina Central University
	Address	Durham, NC, USA
	Email	
	ORCID	http://orcid.org/0000-0001-8355-3249

Abstract

Geospatial tools such as GIS (Geographic Information Systems) serve as a popular technology to assess and evaluate spatial dimensions of the food environment. While local-level policy decisions can be aided using GIS analysis and GIS data, little work has been invested in the holistic understanding of the data on which these decisions are made. In this paper, we address what entails high-quality geospatial data, challenges and opportunities that exist in the field of geospatial data development as applied to local-scale food environment research. We further explored factors of geospatial data quality assessment and quality control (QA/QC) for a commercially available business (CAB) database typically used in high-scale geospatial data analysis of the food environment. Factors related to the physical location of all food sources such as grocery stores and farmers markets and individualized vehicular transportation (roads) rated highest. They outweighed those related to land cover, utilities and zoning, which are more important in medium and low-scale (national level) analysis. When ranking various dimensions of data quality, subject matter experts found positional accuracy and attribute accuracy to be the most important in data development. However, errors related to temporal accuracy (age of data) exhibited the greatest number of errors within a CAB database. This schism serves as the impetus of this project and further addresses challenges between conceptual and practical geospatial data development policies and procedures.

Keywords (separated by '-'): Geographic information system - Geodatabase data development - Geospatial standards - GIS data quality - Food environment

Author Proof

Check for updates

Digital High-Scale Food Security Analysis: Challenges, Considerations and Opportunities

Timothy Mulrooney(✉) and Tysean Wooten

North Carolina Central University, Durham, NC, USA
tmulroon@nccu.edu

AQ1

Abstract. Geospatial tools such as GIS (Geographic Information Systems) serve as a popular technology to assess and evaluate spatial dimensions of the food environment. While local-level policy decisions can be aided using GIS analysis and GIS data, little work has been invested in the holistic understanding of the data on which these decisions are made. In this paper, we address what entails high-quality geospatial data, challenges and opportunities that exist in the field of geospatial data development as applied to local-scale food environment research. We further explored factors of geospatial data quality assessment and quality control (QA/QC) for a commercially available business (CAB) database typically used in high-scale geospatial data analysis of the food environment. Factors related to the physical location of all food sources such as grocery stores and farmers markets and individualized vehicular transportation (roads) rated highest. They outweighed those related to land cover, utilities and zoning, which are more important in medium and low-scale (national level) analysis. When ranking various dimensions of data quality, subject matter experts found positional accuracy and attribute accuracy to be the most important in data development. However, errors related to temporal accuracy (age of data) exhibited the greatest number of errors within a CAB database. This schism serves as the impetus of this project and further addresses challenges between conceptual and practical geospatial data development policies and procedures.

Keywords: Geographic information system · Geodatabase data development · Geospatial standards · GIS data quality · Food environment

1 Introduction

Patterns of negative health-related outcomes such as obesity, hypertension, and diabetes are spatial in nature and when mapped highlight patterns of geostatistical clustering and spatial autocorrelation. While lifestyle choices and genetics contribute to individual and household vulnerability that lead to these differential health outcomes, it is possible to identify social and environmental factors, sometimes associated with geographic location, that have an effect on larger groups, and might be considered as critical indicators to address in any mitigation plan. While "All Americans, rich and poor, have more access to healthy—and unhealthy—food choices than ever" [1], individual-level choice to purchase a particular item is dependent upon a variety of factors. There is, however, a

© Springer Nature Switzerland AG 2021
C. Grueau et al. (Eds.): GISTAM 2020, CCIS 1411, pp. 1–27, 2021.
https://doi.org/10.1007/978-3-030-76374-9_9

Author Proof

strong relationship between health and diet and it seems clear the accessibility of sources for fresh meats, fruits, and vegetables is an important factor in the overall health of a community. Even in low-income neighborhoods, food stamp recipients who live close to supermarkets ate more fresh food and vegetables [2]. Spatial proximity is the principal determinant to patronize a particular grocery for about half (48%) of US residents [3] and more than half (53.9%) of residents in Detroit often shopped within 2 miles of their residence [4]. Those who bypass the closest store cite reasons such as lower prices, lower prices on wanted items, better selection and better quality of fresh foods as reasons for doing bypassing these closer stores [3]. Lower income residents may not have the means to be as selective and are subjected to the grocery store and their options, or lack thereof, that geography dictates. While it is safe to say that geography is not a prime determinant in explaining or even justifying health outcomes, it does have more of a role than one would think.

The United States Department of Agriculture (USDA) has utilized the term *food desert* to underscore regions within low-income communities that have limited accessibility to fresh food via supermarkets. Although some research has focused on rural areas [5–8] most of the knowledge base on the subject has been associated with urban areas which bring about other variables such as pedestrian access and public transportation which are typically not options in rural regions. In addition, the number of large retailers is decreasing or consolidating, but increasing in size to accommodate all shoppers, both grocery and non-grocery [9]. Combined with the fact retailers are migrating to the suburbs from downtowns [10], retailers tend to locate near high-volume roads that are less accessible to non-vehicular individualized transportation (i.e. walking, public transit or riding a bike) [11]. Research [12] has highlighted this disparity of distribution when it found unhealthy food options greatly outweighed healthy counterparts in Los Angeles while other research found poor and minority neighborhoods had less healthy food options than their richer and whiter counterparts [13]. As a result, typical sources of fresh and ‘healthy’ foods such as supermarkets and farmers’ markets are being replaced by fast food restaurants and convenience stores which offer food options that are convenient (easily prepared and physically closer) and inexpensive, but typically less healthy. The long-term ramifications on community health far outweigh any of these tangible and intangible gains. In response to this increasing disproportion, research has explored the notion of *food swamps* which represent areas with tremendously high number or ratio of unhealthy food options compared to healthy options. Research at high scales [14, 15] has shown food swamps predict obesity and other negative health outcomes better than food deserts.

Geospatial tools such as Geographic Information Systems (GIS) serve as a popular technology to represent spatial dimensions of the food environment. A GIS serves as the means by which information about spatially-explicit phenomena can be created, stored, analyzed and rendered in the digital environment. GIS serves as the technological arm in the study of geography (i.e. the study of ‘where’). Experts in many dissimilar fields have seen the utility of GIS as a means of quantifying and expanding their research. GIS is used in disciplines such as business, sociology, justice studies, surveying and the environmental sciences. As applied to food security, GIS can be used to measure the distance between residents and large supermarkets or supercenters, or the density of food outlets

Author Proof

within an enumeration unit (census block group or zip code) as a commonly used proxy for availability and access [16, 17]. These regions of high and low access can be analyzed and mapped across both space and time [18] as shown in Fig. 1 [19], as well as the socio-economic factors that may help explain this access such as median family income (Fig. 2). These make powerful visual products disseminatable and understandable to the entire public that can have long-term practical and policy implications.

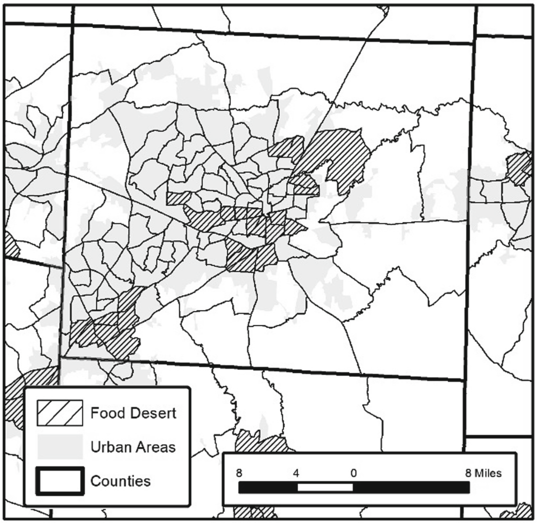


Fig. 1. Map of USDA food deserts in Guilford County, North Carolina [19].

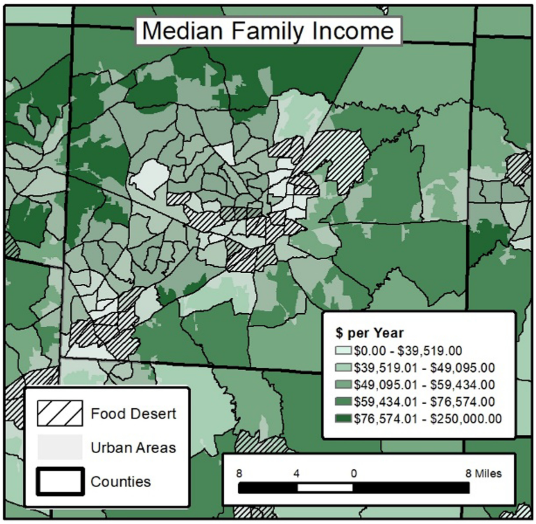


Fig. 2. Map of median household income combined with food deserts.

Author Proof

While many only see the output of GIS data and analysis in the form of maps, resources must be dedicated to creating high-quality data at a local scale. The manner in which these data are captured varies. Some methods include the use of a Global Positioning System (GPS) unit, extracting from or improving upon existing GIS data, downloading data from a web site, connecting to a web service, the use of an Unmanned Aerial Vehicle (UAV) or some other remote sensing platform, or creating data from an analog format via digitization or georectification. Regardless of the method, the resources (e.g., the computers, time and people dedicated to the process of collecting, creating, processing and cataloguing geospatial data) are the most time-consuming portion of a GIS-related project. This research holistically explores the types of geospatial data needed to perform high-quality analysis in support of analyzing and mapping spatial dimensions of the food environment at high scales. These database needs are quite different than data that may be required to remediate food insecurity at the individual/household level (such as Public Use Microdata) or coarser data at a national or sub-national scale. Little research has explored this field of database development, whether for the pure sake of science research and applied decision-making or policy that can be implemented in the field.

In the United States, food insecurity has been described as a "serious public health problem associated with poor cognitive and emotional development in children and with depression and poor health in adults" [20]. Given that women and children have much higher rates of food security than their male and more senior counterparts in the United States, some have called for a rights-based approach to addressing food security [20]. In support of understanding this multi-faceted problem, this research explores both technical and non-technical issues of the data required to represent the tangible and intangible food environment.

2 Literature Review

While the concepts 'food desert' and 'food swamp' have many theoretical definitions, they have applied applications. They exist in the real world and people have a practical understanding of them (Fig. 3).

Fig. 3. Sign outside of vacant grocery story building in Gibsonville, Guilford County, North Carolina.

Author Proof

Food security is considered to be the state "when all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life" [21]. Contemporary literature has used terms such as *availability*, *accessibility*, *proximity, disparity, inequality, density, variety, affordability, walkability, connectivity* and *quality* as well as the aforementioned *food desert* and *food swamp* to describe quantitative measures of the food environment and ultimately food security. These can all be captured using a GIS in some way, shape and form at various scales.

The mapping and demarcation of food-insecure areas within the digital environment has been made exponentially easier using GIS technologies. While first used as an aesthetic tool to map study areas [22] or highlight underlying explanatory variables such as income [23], GIS has since been used to measure real-world distances and calculate densities, quantitatively express proximity and render this proximity with statistical significance using a variety of analytical, geostatistical and cartographic tools. Among the first to do this in the field of food desert research were Donkin et al. [24], Lovett et al. [25] and Pearce [26] while more recent research [27, 28] has quantitatively calculated and mapped the spatial extent of the aforementioned food swamps at high (sub county) scales.

Within a GIS, ways to express spatial dimensions of the food environment vary. Some research has expressed access and availability (or lack of access and availability) as linear units such as kilometers or miles [29], travel time in minutes [30, 31] and densities such as the number of food options per square mile by census tract [32], as well as more complex metrics based on the cost to operate a car [33]. More recently, unitless metrics expressed as ratios [15, 34] have been used as alternatives to absolute measures because these absolute measures are meaningless if not placed within some context. A ten-minute drive time to the nearest fresh food source in an urban area means something much different than a ten-minute drive to the nearest fresh food source in a rural area. The proper and prudent use of absolute measures requires more analysis and interpretation. Food swamp research using GIS utilizes existing metrics such as the Retail Food Environmental Index (RFEI) and the Expanded RFEI [14] while others [27, 28] have derived their own metrics and to define spatial extents of food deserts and swamps using variations of the RFEI, Expanded RFEI, Modified RFEI [35] and Food Balance Metric [36].

However, coarser scales of analysis also exist. In studies that model the supply and demand forces from farm to plate at a national scale, geospatial data regarding farm locations, their arrangement, land cover, flood plains, rivers, climate and population change which support burgeoning sustainable planning, management and development efforts, especially in developing countries are required [37–39]. At this scale, food security at a small scale can be considered a function of the socio-economic and political environment regarding factors such as macro-economy, natural resources, market conditions, education, political climate, food safety/quality and health care practices. These are not considerations in local-scale analysis where proximity to known food sources are correlated with explanatory variables to define food-needy regions. In addition, the geospatial data needs required for local (community) level food analysis are scale dependent and much different in nature than data required at a coarser national or sub-national scale.

Author Proof

These geospatial data required for local research vary in scope, ranging from roads and business locations to sidewalks and municipal boundaries. For example, research in Vermont looked at the quality of food in conjunction with point-to-point distances along a vector road network in which was then grouped into polygonal enumeration units (towns/townships) [5]. Other pioneers [16, 36, 40] also used vector GIS data at some level (individual point, census block group, tract, etc.) to express food security using distance and density calculations derived from GIS data. In national-scale analysis of this type, analyzing thousands to hundreds of thousand sources traveling to thousands of destinations is resource-intensive and requires large, ancillary data layers such as roads in support of this analysis as well as the abovementioned interpretation to be useful. For a large county in North Carolina (~500,000 people) and using Dijkstra's Shortest Path First (SPF) algorithm with a road network of more than 98,000 vertices [41] a best-case scenario for calculating just one drive-time calculation between two locations requires a minimum of 98,000 calculations and a worst-case scenario of more than 9 billion ($98{,}000^2$) calculations [42]. There are literally trillions of possible calculations between sources and potential destinations at the national scale along a much more robust road network, which essentially make desktop computing solutions impossible. While applications using Python, Stata [43] and R programming solutions make this process more efficient than Esri's Network Analyst calculations using a GUI (Graphical User Interface), they are less intuitive for the average GIS user and impossible for the average computer.

In the United States, guidance on the mapping of the food environment begins with the United States Department of Agriculture Food Access Atlas (https://www.ers.usda.gov/data-products/food-access-research-atlas/go-to-the-atlas/). Food access takes into account both the availability or proximity of food sources to residents as well as having readily-available transportation. Information collected and mapped at the census tract level includes the aforementioned food desert metric (low income and limited access) as shown in Fig. 4 as well as individual components that make up this metric and ancillary measures such as income, poverty, race/ethnicity, vehicle access and high-density housing. They are provided as tabular data that can be brought into a GIS and mapped accordingly. As shown in Fig. 4, census tracts can take on varying sizes and shapes. These larger census tracts, one of which is 322 sq. miles (834 sq. km) in size, in the middle of the diagram located in Columbus, Pender and Sampson Counties in North Carolina are especially problematic because they may be too large to highlight high-scale food security patterns necessary for community-based research. As a result, higher-scale food environment analysis using block groups [30, 44, 45] or even pixels using raster-based calculations [27] better elucidate local-level patterns, drives local policy and decision-making, and ultimately serves as a focus of this research.

While there is boundless value in performing local-scale food environment analysis using GIS, little research has been performed on the actual themes or topics that necessitate high-quality research at a high scale. In particular, little work has been performed to determine how important roads are in food security research at the local level. What about elevation? In addition to the actual features, there are various questions about the individual attributes required for high-quality food desert research. Is income (at the census block group level) a necessary attribute for sub-county food desert research? What

Author Proof

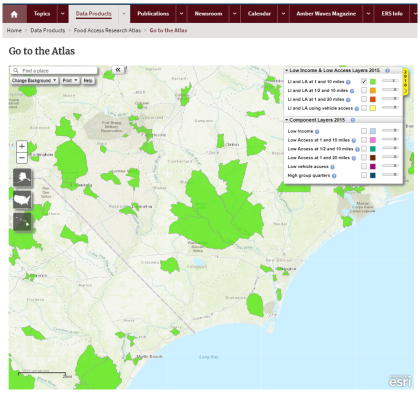


Fig. 4. USDA Food Access Atlas of Southeastern North Carolina showing low income and low access census tracts [19].

about road length? This research explores how can these themes and attributes can be prioritized when time and personnel constraints, which are a reality in the professional world, exist.

In a GIS sources and destinations used in the spatial assessment of the food environment are represented as points. Depending upon the focus and scale of analysis, the number of points utilized can range from the dozens [46, 47] to hundreds [40, 48] and even thousands. As a basis for this research on high-scale food security in North Carolina, GIS work highlighted metrics to measure food security at the block group level [27, 46, 49]. A variety of themes were used in these studies, ranging from roads, business locations and rivers to municipal boundaries, farmers' markets and fast food outlets. Each of these layers were developed or extracted from existing data at a scale appropriate for this type of detailed analysis. However, little insight is provided into what quality assessment was performed for the many consumers of these data. If a supermarket is not provided in the GIS database when one in reality exists (error of omission), one may be mapping food deserts and providing subsequent remediations where it is not needed. On the other hand, if a food source is attributed as a supermarket when it only serves a minimal sampling of fresh food (error of commission), researchers may not be properly identifying the food desert that exists in the area. The significance of data-driven considerations has caused researchers to think critically about the objective assessment, evaluation and reporting of data quality to data users.

In addition to determining what data layers best address the phenomena of food security, it is of paramount importance that these data are also correct. While most end-uses only want the end-products of GIS analysis typically in the form of maps, the largest cost of any GIS project is developing the data which go into high-quality research and the personnel resources attached to this data development. It goes without saying that in an era with limited resources such as personnel, space and time, database developers must be intentional in how, when and to what extent (temporal, spatial and topical) data must be developed. While attempts have been made to estimate the actual and tangible

Author Proof

costs [50, 51] and value [52, 53] of geospatial data, it is impossible to place a monetary value on data although various entities have tried to estimate it [54, 55].

Spatial data quality is the end-product of processes designed to ensure newly created data are correct (Quality Assurance) while also identifying existing data that are incorrect (Quality Control). Applications of QA/QC extend well beyond the GIS world, such as banking, manufacturing, software, medicine and even taxonomy [56]. While some research [57] has distinguished between QA and QC, the two concepts are usually termed as a pair and felt that one cannot exist without the other. Although the QA/QC of spatial data within a GIS is required as per Federal Geographic Data Committeee (FGDC) standards and various organizations have processes in place to ensure the various accuracies are adhered to that best fit their needs, resources and limitations, it is not has been at the forefront of GIS research when compared to other facets of Geographic Information Science.

Nonetheless, the resources dedicated to data creation, especially high-quality data, are extraordinarily high. One opinion is that data quality has no inherent value or worth, but is ultimately realized when an action is taken on information pertaining to data quality [58]. Along those same lines, the end goal of information quality was to satisfy customer needs, in this case being the many users who utilize these data, many of whom assume that the data have undergone some validation [59]. Various components contribute to spatial data quality to include: horizontal accuracy, attribute accuracy, temporal accuracy and attribute completeness.

Horizontal accuracy represents the distance a GIS data layer deviates from geographic reality. It essentially measures the distance a GIS data feature is from where it 'should' be. It is impossible to tell the exact location of where a feature should be placed, as geo-rectified imagery and even high precision Global Position Systems (GPS) data have inherent error attached to them. Some data used food security research (grocery stores) were created via the process of geocoding in which a relative location such as an address is converted to a point with absolute location (latitude and longitude). Researchers found the positional accuracy (the actual location versus what the geocoding algorithm represents as the address) of geocoded rural addresses to be poorer than urban counterparts [60–62]. This can be problematic a large study area.

Attribute accuracy describes how well the assigned attribute values match the actual characteristics of the objects. Attributes are the non-spatial characteristics of an entity used to describe each individual segment. Food source attributes are uniform across an entity, and serve to distinguish one object from another. Attribute values can be text descriptions (e.g., CONAME = 'Food Lion' or NAICS = '44511003') or numerical values (SALESVOL = 1655). In other cases, InfoUSA, a supplier of geospatial business data, uses domain fields to describe particular attributes. For example, the square footage of the store, represented by the field name SQFTCODE, can only have one of four values: A: 1–2, 499 Square Feet, B: 2,500–9,999 Square Feet, C: 10,000–39,999 Square Feet, D: 40,000 + Square Feet.

Attribute completeness measures the degree to which all required attributes have been populated. This does not necessarily mean that they are correct. For example, the SQFTCODE must be populated and can be one of only the four possible aforementioned values matched through a domain table. In some cases, it is left blank in the data. For the

Author Proof

SALESVOL attribute, which represents sales volume in thousands of dollars, it must be an integer. In some cases where it is not provided or unknown, a value of '0' is provided. These missing or unknown values may skew analysis when agglomerated with known values.

Temporal accuracy refers to the age of the data compared to the usage or publication date. Issues of temporal accuracy arise when the GIS data indicates that a feature is open but has since closed. The assessment of temporal accuracy can be problematic because time is rarely treated as a separate entity within spatial databases and even in metadata, except for historically explicit databases such as the decennial census [63].

Early pioneers of GIS recognized the importance of data quality largely due to the legal ramifications in publishing incorrect spatial information which may lead to accidents from the misuse of data [64]. Even then, they understood the reconciliation between accuracy, the cost of creating the most accurate of data and the inevitability that some error will still exist. This concession is what is referred to as *uncertainty absorption* [65]. Given multitude of individual GIS data features required for this type of analysis, it is impossible to field verify every single feature used in analysis.

As applied explicitly to GIS applications related to the quality of spatial food environment data, work has proliferated as research in the spatial analysis and representation of the food environment has increased and a need has arisen to answer questions about the validity of data on which decisions are made. Research [66] has understood these challenges, which include the reliability and validity of data (proper addresses and classifications of stores) as well detail and completeness (enough information is stored that can be useful in food environment analysis). Other research [67] further expounded on these dimensions to include the quality of geocoding processes, the definition of food outlet constructs (what is the definition of healthy, use of proprietary codes, etc.) and ways to measure access and via a reportable standard called Geo-FERN (Food Environment Reporting).

Comprehensive studies [68, 69] explored the quality of large spatial databases purchased from independent sources, referred to as Commercially Available Business (CAB) data, among and between disparate datasets and providers which serve as the basis for retail businesses. Larger-scale studies [70–73] were performed for Durham, Chicago, Albany and Pittsburgh respectively. All cited some degree of difference between different CAB databases such as InfoUSA, Dunn and Bradstreet, TDLinx, as well as field-based and automated methods, noting that caution must be taken when using CAB databases. Further research [74] reinforced the idea of uncertainty absorption within this narrow focus (validity of GIS data in measuring the food environment), highlighting the reconciliation that must be made between the sheer number of data sources provided by CAB databases, the time needed for field verification and the need for high-quality data.

As part of a study on food access and spatial disparity in rural Texas, the addresses of food sources provided via public lists such as Internet telephone directories, telephone directories and the Texas Department of Agriculture were ground-truthed [48]. 18.9% of food sources provided via these public lists could not be verified for a variety of reasons. These reasons included 1) businesses were no longer open 2) business where food source was formerly located was now occupied by non-food source 3) address did not exist or able to geocode and 4) located denoted as a food source was a residence with no apparent

Author Proof

food business. In addition, they found 35.7% of food sources within their study area were only identified through ground-truthing, as these food sources were not provided through public lists. In a similar study, field verification was performed on twenty-one different food source categories (Restaurant, Pub/Bar, Supermarket, Takeaway Food, etc.) across different combinations of socio-economic status (SES) and population densities (urban, rural, mixed) in England. For the rural low SES, more than 36% of food sources provided via a secondary source could not be found in the field [75].

Above and beyond these facets of data quality, the Federal Geographic Data Committee (FGDC) and spatial data transfer standards (SDTS) consider vertical accuracy (error in measured vs. represented elevation), data lineage (source materials of data) and logical consistency (compliance of qualitative relationships inherent in the data structure) components of data quality [76, 77]. Within the GIS community, temporal accuracy (age of the data compared to usage date) and semantic accuracy or "the quality with which geographical objects are described in accordance with the selected model" may also be considered elements of data quality [78] as well as metadata, the formal cataloguing of GIS data. Metadata has been used to describe data quality measures taken during the data development process and subsequent updates. Most generally thought of as "data about data", metadata serves as a formal framework to catalog the lifeline of a particular GIS data set. Feature-level metadata has been able to capture data quality information [79, 80], but is typically limited to quantitative measures of positional accuracy and qualitative information related to data lineage within eight of the more than 400 entries that comprise a complete FGDC-compliant metadata file. Even now, the population of these metadata elements is not fully automated and some entries must be done by the GIS data steward. Given the efficiency at which metadata population is done by each steward, data quality assessment done solely via the extraction of metadata entries is not advised.

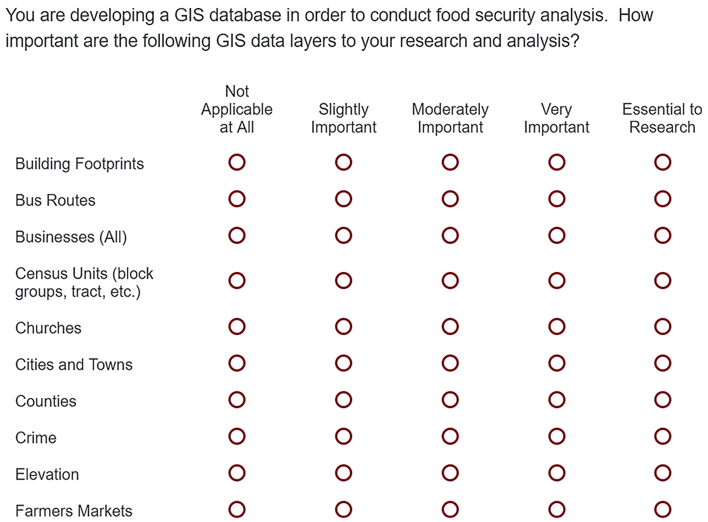
3 Procedures

As a means to prioritize data layers, attributes and dimensions of spatial data quality, a Likert-type survey was developed and distributed to the GIS community that focuses on local-scale food security research. It is composed of twelve questions that not only ask about users' GIS experience, but also asks users questions about their preferences for particular GIS data layers used in analysis (Fig. 5) and the attributes attached to those layers (Fig. 6).

As shown in these figures, respondents were asked to give responses to these questions on a 5-point Likert-type scale, representing "Not Applicable at All" through "Essential to Research". The Likert scale uses ordered responses on a bipolar scale to assess the level of favorability with a particular statement. Some scales do have an even number of responses (4, for example), which force respondents to choose one side of the mean or the other. However, this one does not.

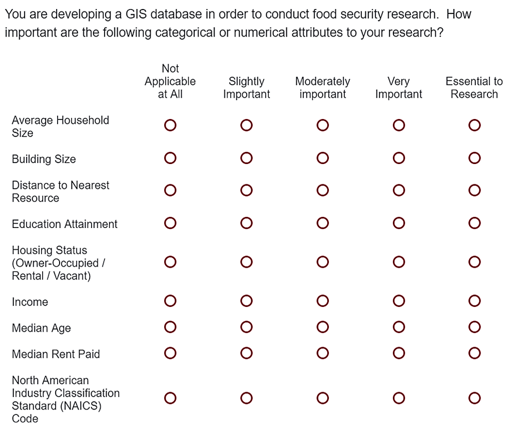
As applied to ranking dimensions of data quality, respondents were given a survey to rank six facets of data quality. An example of this survey and explanations of these facets are highlighted in Fig. 7.

Author Proof


You are developing a GIS database in order to conduct food security analysis. How important are the following GIS data layers to your research and analysis?

	Not Applicable at All	Slightly Important	Moderately Important	Very Important	Essential to Research
Building Footprints					
Bus Routes					
Businesses (All)					
Census Units (block groups, tract, etc.)					
Churches					
Cities and Towns					
Counties					
Crime					
Elevation					
Farmers Markets					

Fig. 5. Likert-type assessment used to rate importance of GIS data themes for use in food desert research. 23 layers were used in this assessment [19].


You are developing a GIS database in order to conduct food security research. How important are the following categorical or numerical attributes to your research?

	Not Applicable at All	Slightly Important	Moderately important	Very Important	Essential to Research
Average Household Size					
Building Size					
Distance to Nearest Resource					
Education Attainment					
Housing Status (Owner-Occupied / Rental / Vacant)					
Income					
Median Age					
Median Rent Paid					
North American Industry Classification Standard (NAICS) Code					

Fig. 6. Likert-type assessment used to rate importance of attributes for use in local-level food desert research. 18 attributes were used in this assessment [19].

This survey was created and distributed to the food desert community via message boards, e-mails and online forums in the Fall of 2017 and Spring 2018. 32 respondents answered the survey.

Author Proof

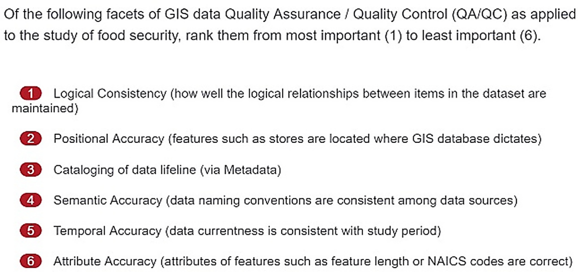


Fig. 7. Dimensions of spatial data quality that respondents were asked to rate using online assessment tool [19].

4 Results

4.1 Prioritization of Data Layers

Respondents were asked to rate data layers on 5-point Likert-type scale ranging from “Not Applicable at All” to “Essential to Research” where each response was assigned a point value as highlighted in Table 1.

Table 1. Respondents were asked the question *“You are developing a GIS database in order to conduct local-scale food security analysis. How important are the following GIS data layers to your research and analysis?”* regarding GIS data layers (street network, for example). The following scale assigned point values to their answers [19].

Response	Point value
Not applicable at all	1
Slightly important	2
Moderately important	3
Very important	4
Essential to research	5

For each layer, a weighted average based on responses was calculated from the values in Table 1 and ranked according to all 23 data layers in the survey. For example, for the Roads data layer, there were no responses for “Not Applicable at All”, one for

“Slightly Important”, five for “Moderately Important”, twelve for “Very Important” and the remaining fourteen responded with “Essential to Research”. This would compute to a value of 4.22 and this value would be ranked among the other 22 data layers selected for this survey. In this case, the Roads layer ranked 2nd amongst the 23 data layers in the questionnaire. The “Grocery Stores” data layer ranked with the highest with a score of 4.25, followed closely by “Roads”, “Farmers’ Markets” and “Urban Areas” as highlighted in Table 2.

Author Proof

Table 2. Rank of Layers/Themes as Voted by GIS User Community [19].

Rank	Layer
1	Grocery stores
2	Roads
3	Farmers markets
4	Urban areas
5	Census units (block groups, tract, etc.)
6	Cities and towns
7	Fast-food restaurants
8	Counties
9	Bus routes
10	Businesses (All)
11	Non-census sub-county units (boroughs, townships, etc.)
12	Schools
13	Zoning
14	Sidewalks
15	Land cover
16	States
17	Churches
18	Walking/Jogging trails
19	Building footprints
20	Crime
21	Utilities (Electrical/Gas/Cable/Phone)
22	Elevation
23	Golf courses

In addition, users were asked to name themes not mentioned in the above list that would be useful in this type of analysis. Themes mentioned include: Parks, Greenhouses, Arable Land, Irrigation Pathways, Rivers, Access to Water, Food Banks, Food Assistance Organizations, Community Gardens, Non-Profit Businesses, Health Agencies, Corner Stores, Partial Markets (Walgreens, for example), Liquor Stores, Bus Stops and County Agencies.

4.2 Prioritization of Attributes

The same conventions were applied to attributes used to describe the data layers from Table 1. After averaging values marked by uses, the “Distance to Resource” attribute was ranked highest, followed by “Income” and “Race/Ethnicity (by enumeration unit)”. These results are highlighted in Table 3.

Author Proof

Table 3. Rank of Attributes to Layers/Themes as Voted by GIS User Community [19].

Rank	Attribute
1	Distance to nearest resource
2	Income
3	Race/Ethnicity (by enumeration unit)
4	Population density
5	Average household size
6	Population
7	Education attainment
8	Housing status (Owner-Occupied/Rental/Vacant)
9	Transportation (# of vehicles by enumeration unit)
10	Median age
11	Median rent paid
12	Spending patterns (by enumeration unit)
13	Zoning type
14	North American industry classification standard (NAICS) Code
15	Road length
16	Building size
17	Number of employees by business
18	Speed limit

4.3 Dimensions of Data Quality

Using the facets of data quality addressed above, users were asked to rate six different dimensions of data quality from 1 (most important) to 6 (least important). These data dimensions speak to how the data are created, described and catalogued as part of the data development process. Scores for each facet were merely averaged and ranked. These rankings are highlighted in Table 4.

5 Opportunities for Development

5.1 Practical Applications of Data Quality Research

The importance and concepts of positional, temporal and attribute accuracies tie in with burgeoning opportunities in field of data quality assessment. These facets of data quality were rated highest of the six addressed in the survey as per Table 4. Research is beginning to realize the importance of testing data quality for store locations which entail a combination of field techniques and database analysis [69–73].

Author Proof

Table 4. Rank of Dimensions of Data Quality [19].

Rank	Facet of data quality
1	Positional accuracy (features such as stores are located where GIS database dictates)
2	Attribute accuracy (attributes of features such as feature length or NAICS codes are correct)
3	Temporal accuracy (data currentness is consistent with study period)
4	Logical consistency (how well the logical relationships between items in the dataset are maintained)
5	Semantic accuracy (data naming conventions are consistent among data sources)
6	Cataloging of data lifeline (via Metadata)

In support of this work, the research team developed a short field-based QA/QC project. 400 randomly selected food sources from an eleven-county region in southeastern North Carolina were divided between each of two major divisions of food ('healthy' vs. 'unhealthy') within urban and rural food sources. In order to maintain consistency in field verification for hypothesis testing, 100 urban healthy (UH) sources were randomly selected, as well as 100 rural healthy (RH), 100 urban unhealthy (UU) and then 100 rural unhealthy (RU). As a result, 200 urban features within the GIS database were field checked against 200 rural food sources in the same database. 200 healthy sources were to be checked against 200 unhealthy counterparts.

All 400 points were randomly selected and placed into a database for on-site field verification. The goal of field verification was to determine 1) if the business was actually located where the GIS database dictated 2) if the business was still in operation 3) if the business activity (fast food, for example) is attributed correctly. Also noted in the database were other issues that may contribute to questions of data integrity and subsequent food desert analysis, such as 1) geocoding errors where that point is located nearby, but not exactly where it should be and 2) points that could be attributed differently. This may occur where a small grocery store could have been attributed as a convenience store. Attributes were created specifically for field verification that contained placeholders for these notations that could be done in the field.

400 points were inspected to determine how well these GIS data and various permutations of these data aligned with geographic reality as well as cohorts against each other. Of the 400 total points inspected, 310 (77.5%) of them were accurate. Of the 90 AQ2 that were deemed as incorrect, the following is a summary of the errors (Table 5):

Author Proof

Table 5. Summary of errors in QA/QC process.

Description of error	Number of occurrences	Type of error
Food source permanently closed	32	Temporal accuracy
Point is actually a residential location	24	Attribute accuracy
Nothing exists at the point	18	Horizontal accuracy
New business occupying Location	9	Temporal accuracy
Does not sell food directly to public (Distributor)	3	Attribute accuracy
Business name is the same, but is not a food source	2	Attribute accuracy
Located far distance from actual feature	2	Horizontal accuracy

All 90 errors were generalized into one of seven general descriptions as shown in Table 3. The most popular error, representing 35.6% of all errors, was that the food source represented in the GIS databases, was permanently closed. One example of these temporal inaccuracies is shown in Fig. 7 (Fig. 8).

Fig. 8. Rural Supermarket Now Permanently Closed. This Location Was Represented in the GIS Database as Being Open.

These 90 errors were broken down between various cohorts of the food environment as shown in Table 4. Most notable is the difference between urban and rural accuracy. 82.5% of all 200 urban features checked were correct compared to 72.5% of rural counterparts using the same sample size. These differences were also expressed between healthy food (82% urban vs. 70% rural) and unhealthy food (83% urban vs. 75% rural). Of the three different cohorts of food sources field verified, all of them had urban accuracy to be greater than rural accuracy.

An independent t-test of two proportions was run between the two sets of results to determine if there was a difference between the percentages computed. Using the derived

Author Proof

accuracy percentages for each cohort ($\hat{p}_1$ and $\hat{p}_2$), the combined accuracy ($\hat{p}_0$) and the sample sizes for each cohort (n_1 and n_2), this test helps determine the criteria in order to reject the Null hypothesis (percentage from each cohort is equal to each other) and accept the alternate hypothesis (percent from each cohort are not equal to each other).

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0(1 - \hat{p}_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Permutations of the were run against each other using the test of two proportions as shown in Table 6. There are differences between urban and rural accuracy for the some of the six different cohorts of food stores inspected. Most significant was the distinct differences between the accuracy for all urban food sources and less accurate rural food sources at the $\alpha = .05$ level. These differences must be noted in working with unverified CAB data.

Table 6. Result for test of two proportions.

Null hypothesis	p-value
Urban Healthy (n = 100) = Rural Healthy (n = 100)	.0483**
Urban Unhealthy (n = 100) = Rural Unhealthy (n = 100)	.1664
All Urban (n = 200) = All Rural (n = 200)	.0170**
*$\rho < .1$ **$\rho < .05$ ***$\rho < .01$	

In addition to the actual quality of geospatial data being provided in a CAB database highlighting differences between urban and rural cohorts, other research has explored store-level metrics such as the linear shelf space of healthy food [81] and the amount of bruising of foods within stores [82]. These ideas further perpetuate the concepts of the relatively new idea of spatial justice/injustice which explores how access to both tangible and intangible assets [83, 84], such as food quality and even high-quality data in this case, vary across space. Further research opportunities into issues of data collection methods, field verification, data collection frequency and logical consistency can address the reasons for these distinct differences as applied to the narrow scope of spatial data accuracy within the confines of the food environment.

5.2 Standards-Based Approach to Database Development

Data standards such as the Spatial Data Standards for Facilities Infrastructure and Environment (SDSFIE) are used by the Department of Defense (DoD) to maximize interoperability and understandability across installations and branches by dictating naming conventions, attributes and domain values for spatial data layers. The name *spot_elevation_point* is denoted as "a point on the surface of the earth of known elevation" and is consistent across all DoD installations instead of using layer names such

Author Proof

as *point*, *landmark* or *landmarks*. The *spot_elevation_point* feature class contains 23 attributes, which is relatively little compared to the *road_centerline* feature class which contains 55 attributes. The FGDC has defined data standards for landmarks, addressing, thoroughfares and parcels (FGDC, 2011) in order to standardize attributes so features can geocoded, described and represented fully entirely by the GIS user community. While the development of a database dedicated solely to food security is still being realized, point and polygonal features representing municipal and census-based units such as zip codes, towns, census tracts and census block groups have attributes that can be seamlessly integrated with attributes that rank highly in this study such as supermarket density and access to transportation, as well as socio-economic indicators such as poverty, race/ethnicity, education attainment, population and population density. The development of these attributes may require further processing or the import of data using simple GIS operations from various spatial databases such as the 2010 Census, Esri Demographic Database, Esri Spending Patterns and American Community Survey.

In order to catalog both the data and the aforementioned processing, it is necessary to catalog administrative, structural and descriptive information about the geospatial data and the processes by which they were developed. Metadata serves as the formal means to describe a dataset, and provides the standardized framework for providing information about a dataset's lineage, attributes, age and creators using both qualitative and quantitative entries. In the GIS community, the FGDC-endorsed Content Standard for Digital Geospatial Metadata (CSDGM) is slowly being replaced by an International Standards Organization (ISO)-based metadata standard that accounts for evolving technologies such as remotely sensed imagery, online services and ontologies that did not exist when the original CSDGM (formally known as *FGDC-STD-001-1998*) was first published in 1998.

More than 400 individual elements comprise a complete metadata record and the state of North Carolina has developed a State and Local Government Profile, based on the ISO 19115, 19115-1 and 19119 standards. This standard streamlines these 400 elements into about 75 elements that best capture the information about a data layer which enable content consistency and improves the search and discoverability of data through online data repositories such as NCOneMap. This standard, as well as guidance for its use, is provided by the North Carolina Geographic Information Coordinating Council (NCGICC) through the NCOneMap online portal [85].

Using the State and Local Government Profile as a template, data layers developed in support of high-scale food security research should be cognizant of the following entries that already exist within this profile which speak explicitly to the aforementioned facets of data quality and help perpetuate data discoverability:

1) Topic Category: A theme keyword that adheres to at least one of the ISO Topic Categories.
2) Process Description: A repeatable element that provides a description of how the data were created and indicate the data source, where applicable. This process description should include any geoprocessing and/or field calculations used to derive spatial and attribute data derived for the sole purpose of food security research. This process description should also contain the source scale denominator and publication date

Author Proof

of source information, where available to clarify positional and temporal accuracy respectively.

3) Feature Catalogue: Entity and Attribute Descriptions and Citations referenced to ISO 19110, where possible.

In addition, the following Data Quality elements not explicitly addressed in this profile should be completed to catalog attempts to maintain the highest possible accuracies of data used in analysis. While not required, this cataloguing should strive to achieve popular positional (horizontal and vertical) accuracy standards such as the National Mapping Accuracy Standards (NMAS) for paper maps [86] and more recent National Standard for Spatial Data Accuracy (NSSDA) applied to purely digital data [87].

1) Attribute Accuracy Report: an explanation of the accuracy of the identification of the entities and assignments of values in the data set and a description of the tests used. This may be useful if food sources and/or destinations have been field checked for attribute errors.
2) Quantitative Attribute Accuracy Assessment: a value assigned to summarize the accuracy of the identification of the entities and assignments of values in the data set and the identification of the test that yielded the value.
3) Attribute Accuracy Value: an estimate of the accuracy of the identification of the entities and assignments of attribute values in the data set.
4) Logical Consistency Report: an explanation of the fidelity of relationships in the data set and tests used. This may be applicable if data used in the same analysis or derivation of attributes come from multiple data sources and/or at different scales.
5) Completeness Report: information about omissions, selection criteria, generalization, definitions used, and other rules used to derive the data set. Useful for both spatial data and attribute completion.
6) Horizontal Positional Accuracy Report: an explanation of the accuracy of the horizontal coordinate measurements and a description of the tests used. This may be useful when field checking the locations of food sources and/or destinations.
7) Horizontal Positional Accuracy Value: an estimate of accuracy of the horizontal positions of the spatial objects.
8) Horizontal Positional Accuracy Explanation: the identification of the test that yielded the Horizontal Positional Accuracy Value.
9) Vertical Positional Accuracy Report (where applicable): an explanation of the accuracy of the vertical coordinate measurements and a description of the tests used [76].

6 Conclusions

Spatial dimensions of the food environment can be measured using GIS. While GIS has increasingly become a powerful tool to map spatial dimensions of food security and the factors that help explain it, practitioners have little understanding of the challenges and opportunities in working with data at various scales. The comprehensive development of high-scale spatial data in support of the food environment elicits a number of both quantitative and qualitative considerations discussed in this paper. Among the considerations

Author Proof

included in this paper include the data themes necessary for research, the attributes for said themes, the importance of various dimensions of data quality, efforts to assess and evaluate data quality in the field, data quality and the role of metadata in the cataloguing of these data.

Given data and the people that develop it are the most expensive component of any GIS project, this is especially important when resources such as time, personnel, storage space, processing speeds and bandwidth must be compromised. This data development can take on many forms, ranging from the downloading of existing data, processing of existing data, extraction from currently existing databases such as the aforementioned CAB databases, geocoding or the use of remotely sensed imagery, either purchased, procured or captured using a UAS (Unmanned Aircraft System). Regardless of the methods, resources must be utilized in order to create the spatial information and derive the attributes that facilitate food security research while cataloguing the people, processes and resources via metadata that can be discoverable across various, especially online, platforms.

As highlighted in this paper, the GIS database requirements for food security analysis at a local scale are much different than those needs at the national/sub-national scale. National scale and sub-national (state) studies in food security explore the economics of food production and links between this food and those who need it using data such as land cover, supply chains, zoning, soil type, low-scale transportation networks (both road and railroad), state and county outlines using coarse data. High-scale analysis at the block group and even pixel scale requires more specialized data, analysis, attribution and cataloguing than data grouped at census tracts, the standard for much research, including the United States Department of Agriculture Food Access Atlas. Types of data required include high-scale road networks (which include speed limits and derived travel times), business locations and spending patterns. From a data development standpoint, the realization of a database in support of local-scale food security research requires a reconciliation between developing the correct data layers, developing them at an appropriate scale that allows for local-level (sub county) scale analysis, rendering within appropriate budgets (time, people, money, etc.) that can be practically applied through policy and/or decision-making.

Utilizing a survey of 32 GIS professionals who integrate GIS data in support of food environment research, they provided their opinions on the importance of various themes attached to food desert analysis the relative importance of dimensions of data quality. Themes directly contributing to the physical procurement of healthy food such as grocery stores, roads and farmers' markets were ranked highest by these professionals. Furthermore, analysis utilizing census tracts and block groups were ranked higher than counties, further articulating the opinion that county level analysis is too just coarse to guide meaningful decision making.

GIS-based exploratory data analysis is a useful tool for model development as it allows analysts to interrogate diverse geographically linked datasets to identify inherent patterns and develop testable hypotheses regarding factors contributing to those observed patterns. This data-driven approach minimizes bias from imposition of untested assumptions derived from studies for other purposes at other scales in other settings. Information related to proximity (physical distance from resources) and socio-demographics such as

Author Proof

income, race/ethnicity and household size were deemed as most important. These factors are essential to food desert research and specifically the USDA definition of a food desert, which contain both distance and poverty components. Lastly, dimensions of data quality were identified and users were asked to rank them in their order of importance. Positional accuracy and attribute accuracy ranked highest while the cataloguing of data in the form of metadata was ranked lowest. Research in the field of geospatial data quality assessment is evolving using field-based (virtual and otherwise) and programmatic techniques.

This focus on positional and attribute accuracy within this research was especially interesting because errors related to temporal accuracy (age of data) exhibited the most number of errors within a CAB database (food business locations) used for high-quality food environment research. In a field assessment and evaluation of 400 randomly selected data features in southeast North Carolina, 90 of these data points were found to be incorrect. 46% of the errors were related to temporal accuracy of the data, whether the business in question no longer existed at that location or a new type of business was occupying the food business location when checked in the field. 32% of errors were related to attribute errors where 1) the location was in fact a residential location) 2) the business name was correct, but it did not sell food and 3) the business did not sell directly to the public. The remaining 22% of errors were related to horizontal accuracy where the business location in the GIS was located far from the actual business, most likely due to geocoding error.

In exploring differences between various preselected cohorts of these data sources, distinct differences were found between accuracies for rural and urban cohorts. For n = 200, the geospatial data representing rural food sources (72.5%) was less accurate than urban cohorts (82.5%) at $\alpha = .05$. In addition, rural healthy food sources were statistically less accurate than urban healthy cohorts at that same significance level. While rural communities are disproportionately affected by unhealthy food environments [16] and some research has shown that disparities in food access are also greatest in rural communities [88, 89], this disproportionality also extends to the accuracy data sources within these regions. These schisms, which also include the difference between our concerns and perceptions with respect to geospatial data error and the true empirical error in geospatial data, serves as an impetus for future work and further addresses challenges between conceptual and practical geospatial data development policies and procedures.

High-quality data serves as the fundamental basis for decision-making. GIS data, whether provided through the United States Census or through other vendors can be easily converted to geospatial format if they are not already provided in that format. Another of the challenges in working with these data at various scales is its reliability, or lack thereof. Explanatory demographic data are typically collected within enumeration units such as the census block group, tract, county and state level through the American Community Survey (ACS), a program through the United States Census that samples data in non-decennial census years. Inherent in all ACS data is a sampling error, which represents “errors that occur from making inferences about the whole population from only a sample of the population” [90]. Within quantitative calculations of error is an enumeration unit’s determination of reliability which are a result of scale, sampling

Author Proof

methods and sampling size. Three classes of reliability exist for ACS data: High, Medium and Low. These classes can give users and decision makers insight into the data used for analysis at a particular scale. These factors must also be considered when developing data or overlaying them with other geospatial data given the propagation of error inherent in multiple inaccurate or unreliable data sources.

The specific focus of this work has been on the collection, integration, analysis, assessment and systematic description of geospatial data via formal metadata that is of a type and level of detail to be of practical value in the development, implementation and evaluation of interventions and policies addressing local-level food security. This holistic approach necessitates an understanding of the technical skills needed to develop high-quality geospatial data as well as the qualitative understanding to While the results of this work can be used as pure research in and of itself, it is anticipated that results can be used in helping to facilitate decision-making and dictate policy at directly addressing and remediating the phenomenon of food deserts as well a proliferating research in disparate fields such as meta-metadata (information about metadata), data mining, field assessment and data quality. Furthermore, this work addresses the technical components of geospatial database development such as attribution, naming conventions and metadata according to existing standards such as the ISO-based North Carolina State and Local Government Metadata Profile. While some minor questions still remain unanswered such as the potential for cross-validation or the integration of qualitative data given food desert research has been trending towards a mixed-methods approach (combining qualitative and quantitative data), it is our hope to further explore cost-effective methods for needs assessment that take into account both causal complexity, perhaps via longitudinal studies, and programmatic challenges imposed by the combination of the increase of chronic disease, the contribution of unhealthy eating to chronic disease, limited resources and increased demand. If done correctly, integrating GIS technologies with intervention planning has the potential to be a cost-effective means for organizations to conduct effective planning aimed at improving food and nutritional security at multiple spatial and temporal scales. Practical database development and the efficient use of resources serves as the cornerstone of this planning and implementation.

Nonetheless, the framework approach described in this research is flexible and broadly applicable, and can be useful for comparing and exploring spatial relationships among scales, accuracies and standards between different study areas if resources exist. We suggest that the approach, methods and results described in this paper be used to inform analysts and end-users of geospatial data research of any implicit or explicit error that may explain, elucidate, undermine and reinforce results using these data.

Acknowledgements. The project was supported by the Agricultural and Food Research Initiative Competitive Program of the USDA National Institute of Food and Agriculture (NIFA), grant number 2016-67023-24904.

This material is also based upon work supported by the National Science Foundation under Grant No. 1824949. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author Proof

References

1. Brat, I.: Do ‘food deserts’ cause unhealthy eating? Wall Street J. (2015). https://www.wsj.com/articles/do-food-deserts-cause-unhealthy-eating-1436757037. Accessed 21 Mar 2020
2. Rose, D., Richards, R.: Food store access and household fruit and vegetable use among participants in the US food stamp program. Public Health Nutri. **7**(8), 1081–1088 (2004)
3. Food Marketing Institute: U.S. Grocery Shopping Trends, 2016. Food Marketing Institute, Arlington, VA (2016)
4. Zenk, S., et al.: Food shopping behaviors and exposure to discrimination. Public Health Nutri. **17**(5), 1167–1176 (2013)
5. Van Hoesen, J., Bunkley, B., Currier, C.: A GIS-based methodology toward refining the concept of rural food deserts: a case study from Rutland County, Vermont. J. Agric. Food Syst. Comm. Develop. **3**(2), 61–76 (2013)
6. Gross, J., Rosenberger, N.: Food Insecurity in Rural Benton County: An Ethnographic Study. Working Paper No. 05-02, Rural Studies Program, Oregon State University, Corvallis, OR (2005)
7. Blanchard, T., Lyson, T.: Access to low cost groceries in nonmetropolitan counties: Large retailers and the creation of food deserts. Paper presented at the Measuring Rural Diversity Conference, Washington, DC, pp. 1– 24 (2006)
8. Morton, L., Bitto, E., Oakland, M.: Solving the problems of Iowa food deserts: food insecurity and civic structure. Rural Sociol. **70**(1), 94–112 (2005)
9. Clarke, G., Heather, E., Guy, C.: Deriving indicators of access to food retail provision in british cities: studies of Cardiff. Leeds Bradford. Urban Stud. **39**(11), 2041–2060 (2002)
10. Furey, S., Strugnell, C., McIlveen, H.: An investigation of the potential existence of “food deserts” in rural and urban areas of Northern Ireland. Agric. Hum. Values **18**, 447–457 (2001)
11. Mamen, K.: Facing Goliath: challenging the impacts of supermarket consolidation on our local economies, communities, and food security - policy brief. Oakland Inst. **1**(3), 1–8 (2007)
12. Lewis, L.B., et al.: African Americans’ access to healthy food options in South Los Angeles restaurants. Am. J. Public Health **95**(4), 668–673 (2005)
13. Powell, L.M., Slater, S., Mirtcheva, D., Bao, Y., Chaloupka, F.: Food store availability and neighborhood characteristics in the United States. Prev. Med. **44**, 189–195 (2007)
14. Cooksey-Stowers, K., Schwartz, M., Brownell, K.: Food swamps predict obesity rates better than food deserts in the United States. Int. J. Environ. Res. Public Health **14**, 1366 (2017)
15. Zenk, S.N., Powell, L.M., Rimkus, L., Isgor, Z., Barker, D.C.: Relative and absolute availability of healthier food and beverage alternatives across communities in the United States. Am. J. Public Health **104**, 2170–2179 (2014)
16. Morton, L., Blanchard, T.C.: Starved for access: life in rural America’s food deserts. Rural Realities **1**(4), 1–10 (2007)
17. Sharkey, J.R., Horel, S.: Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. J. Nutr. **30**(5), 620–627 (2008)
18. Chen, X., Clark, J.: Interactive three-dimensional geovisualization of space–time access to food. Appl. Geogr. **43**, 81–86 (2013)
19. Mulrooney T., Wooten, T.: A public participatory approach toward the development of a comprehensive geospatial database in support of high-scale food security analysis. In: Proceeding of the 6th International Conference in Geographical Information Systems Theory Application and Management (GISTAM 2020), pp. 21–32 (2020)
20. Chilton, M., Rose, D.: A rights-based approach to food insecurity in the United States. Am. J. Public Health **99**, 1203–1211 (2009)

Author Proof

21. Food and Agriculture Organization of the United Nations: Declaration of the World Summit on Food Security 2009. ftp://ftp.fao.org/docrep/fao/Meeting/018/k6050e.pdf. Accessed 29 Feb 2020
22. Wrigley, N.: 'Food deserts' in British cities: policy context and research priorities. Urban Stud. **39**(11), 2029–2040 (2002)
23. Guy, C.M., Clarke, G., Eyre, H.: Food retail change and the growth of food deserts: a case study of Cardiff. Int. J. Retail Distrib. Manage. **32**, 72–88 (2004)
24. Donkin, A.J., Dowler, E.A., Stevenson, S.J., Turner, S.A.M.: Mapping access to food at a local level. British Food J. **101**(7), 554 (1999)
25. Lovett, A., Haynes, R., Sunnenberg, G., Gale, S.: Car travel time and accessibility by bus to general practitioner services: a study using patient registers and GIS. Soc. Sci. Med. **55**(1), 97–111 (2002)
26. Pearce, J., Witten, K., Bartie, P.: Neighbourhoods and health: a GIS approach to measuring community resource accessibility. J. Epidemiol. Community Health **60**(5), 389–395 (2006)
27. 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. **57**(2), 151–178 (2017)
28. Rose D., Bodor J.N., Swalm C.M., Rice J.C., Farley T.A., Hutchinson P.L.: Deserts in New Orleans? Illustrations of urban food access and implications for policy. Paper presented at: University of Michigan National Poverty Center/USDA Research Conference on Understanding the Economic Concepts and Characteristics of Food Access. USDA, Washington, DC (2004)
29. Jago, R., Baranowski, T., Baranowski, J.C.: Distance to food stores and adolescent male fruit and vegetable consumption: mediation effects. Int. J. Behav. Nutri. Phys. Activity **4**, 35 (2007)
30. Jiao, J., Moudon, A., Ulmer, J., Hurvitz, P., Drewnowski, J.: How to identify food deserts: measuring physical and economic access to supermarkets in King CountyWashington. Am. J. Public Health **102**(10), 32–39 (2012)
31. Ver Ploeg, M., Breneman, V., Farrigan, T., Hamrick, K., Hopkins, D., et al.: Access to affordable and nutritious food–measuring and understanding food deserts and their consequences: United States Department of Agriculture (Economic Research), Service Washington, D.C. (2009)
32. Block, J.P., Scribner, R.A., DeSalvo, K.B.: Fast food, race/ethnicity, and income: a geographic analysis. Am. J. Prevent. Med. **27**, 211–217 (2004)
33. Hallett, L., McDermott, D.: Quantifying the extent and cost of food deserts in Lawrence, Kansas, USA. Appl. Geogr. **31**, 1210–1215 (2011)
34. Clary, C.M., Ramos, Y., Shareck, M., Kestens, Y.: Should we use absolute or relative measures when assessing foodscape exposure in relation to fruit and vegetable intake? Evidence from a wide-scale Canadian study. Prevent. Med. (Baltimore) **71**, 83–87 (2015)
35. Centers for Disease Control and Prevention: Children's Food Environment State Indicator Report (2011). http://www.cdc.gov/obesity/downloads/ChildrensFoodEnvironment.pdf. Accessed 3 Mar 2020
36. Gallagher, M: Examining the impact of food deserts on public health in Chicago. Mari Gallagher Research and Consulting Group. http://www.marigallagher.com/site_media/dynamic/project_files/1_ChicagoFoodDesertReport-Full_.pdf. Accessed 22 Apr 2029
37. Babatunde, R.O., Omotesho, O.A., Sholatan, O.S.: Factors influencing food security status of rural farmers in North Central Nigeria. Agric. J. **2**(3), 351–357 (2010)
38. Obioha, E.E.: Climate variability, environment change and food security nexus in Nigeria. J. Hum. Ecol. **26**(2), 107–121 (2009)
39. Soneye, A.: Farm holdings in Northern Nigeria and implication for food security: a remote sensing and GIS assessment. Afr. J. Food Agric. Nutri. Develop. **14**(2), 1–15 (2014)

Author Proof

40. McEntee, J., Agyeman, J.: Towards the development of a GIS method for identifying rural food deserts: geographic access in Vermont, USA. Appl. Geogr. **30**(1), 165–176 (2010)
41. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**, 269–271 (1959)
42. Melhorn, K.: Shortest Paths: Algorithm and Data Structures - The Basic Toolbox. Springer, New York (2008). https://doi.org/10.1007/978-3-540-77978-0
43. Huber, S., Rust, C.: Calculate travel time and distance with OpenStreetMap data using the open source routing machine (OSRM). Stata J. **16**(2), 416–423 (2016)
44. D'Acosta, J.: Finding food deserts: a study of food access measures in the phoenix-mesa urban area (Master's thesis) University of Southern California. Los Angeles, CA (2015)
45. Wang, Z.: Socio-demographic and economic determinants of food deserts (Master's thesis) University of Arkansas, Fayetteville, AR (2012)
46. Love, G., Mulrooney, T., Brown, L.: Using GIS to address food availability in Durham, North Carolina. North Carolina Geogr. **19**, 34–52 (2012)
47. Opfer, P.: Using GIS technology to identify and analyze 'food deserts' on the southern Oregon coast. Oregon State University Archives, pp. 1–52 (2010)
48. Sharkey, J.R., Horel, S., Han, D., Huber, D.: Association between neighborhood need and spatial access to food stores and fast food restaurants in neighborhoods of colonias. Int. J. Health Geogr. **8**(9), 1–17 (2009)
49. Major, E., Delmelle, E., Delmelle, E.: 2018 SNAPScapes: using geodemographic segmentation to classify the food access landscape. Urban Sci. **2**, 71 (2018)
50. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manage. **29**(4), 258–268 (2012)
51. Johnson, P.A., Sieber, R., Scassa, T., Stephens, M., Robinson, P.: The Cost(s) of geospatial open data. Trans. GIS **21**, 434–445 (2017)
52. Bernknopf, R., Shapiro, C.: Economic assessment of the use value of geospatial information. ISPRS Int. J. Geo-Inf. **4**, 1142–1165 (2015)
53. Both, A., Wauer, M., Garcia-Rojas, A, Hladky, D., Lehmann, J.: GeoKnow generator workbench – an integrated tool supporting the linked data lifecycle for enterprise usage. In: Proceedings of the 11th International Conference on Semantic Systems Posters and Demos, of SEM 2015, p. 92 (2015)
54. Koutnik, M.: The Benefits of GIS in Local Government: Examples from ESRI's User Community. Environmental Systems Research Institute, Inc. Redlands (1996)
55. Ledbetter, M., Hunt, C. Anderson, D.: Blueprints for a Citywide GIS: Scottsdale's Award-Winning System Provides a Profitable Example. GIS World, pp. 62–64, November 1997
56. Chapman, A.D.: Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen (2005)
57. Taulbee, S.M.: Implementing data quality systems in biomedical records. In: Gad, S.C., Taulbee, S.M. (eds.) Handbook of Data Recording, Maintenance, and Management for the Biomedical Sciences, pp. 47–75. CRC Press, Boca Raton (1996)
58. Dalcin, E.C.: Data quality concepts and techniques applied to taxonomic databases. Thesis for the degree of Doctor of Philosophy, School of Biological Sciences, Faculty of Medicine, Health and Life Sciences, University of Southampton (2004)
59. English, L.P.: Improving data warehouse and business information quality: methods for reducing costs and increasing profits. Wiley, New York (1999)
60. Bonner, M.R., Han, D., Nie, J., Rogerson, P., Vena, J.E., Freudenheim, J.L.: Positional accuracy of geocoded addresses in epidemiologic research. Epidemiology **14**(4), 408–412 (2003)
61. Cayo, M.R., Talbot, T.O.: Positional error in automated geocoding of residential addresses. Int. J. Health Geogr. **2**(1), 10 (2003)

Author Proof

62. Ward, M.H., et al.: Positional accuracy of two methods of geocoding. Epidemiology **16**(4), 542–547 (2005)
63. Longley, P., Goodchild, M., Maguire, D., Rhind, D.: Geographical Information Systems: Principles, Techniques, Management and Application, 2nd edn. Wiley, Chicester (2005)
64. Epstein, E.F: Litigation over information: The use and misuse of maps. In: Proceedings, IGIS: The Research Agenda 1, pp. 177–184. NASA, Washington, D.C. (1988)
65. Bedard, Y. Uncertainties in land information systems databases. In: Proceedings of the 8th International Symposium on Computer Assisted Cartography (Auto Carto 8), Baltimore, MD, pp. 175–184 (1987)
66. Forysth, A., Lytle, L., Van Riper, D.: Finding food: issues and challenges in using geographic information systems to measure food access. J. Transp. Land Use **3**(1), 43–65 (2010)
67. Wilkins, E., Morris, A., Radley, D., Griffith, C.: Using Geographic Information Systems to measure retail food environments: Discussion of methodological considerations and a proposed reporting checklist (Geo-FERN). Health Place **44**, 100–117 (2017)
68. Auchincloss, A., Moore, K., Moore, L., Diez-Roux, A.: Improving retrospective characterization of the food environment for a larger region in the United State during a historic time period. Health Place **18**(6), 1341–1347 (2012)
69. Liese, A., et al.: Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. Am. J. Epidemiol. **172**(11), 1324–1333 (2010)
70. Han, E., Powell, L., Zenk, S., Rimkus, L, Ohri-Vachaspati, P., Chaloupka, F.: Classification bias in commercial business lists for retail food stores in the U.S. Int. J. Behav. Nutri. Phys. Activity **9**, 46 (2012)
71. Hosler, A., Dharssi, A.: Identifying retail food stores to evaluate the food environment. Am. J. Prev. Med. **39**(1), 41–44 (2010)
72. Mendez, D., Kim, K., Hardaway, C., Fabio, A.: Neighborhood racial and socioeconomic disparities in the food and alcohol environment: are there difference by commercial data sources? J. Racial Ethnic Health Disparities **3**, 108–116 (2016)
73. Rummo, P., Albrecht, S., Gordon-Larsen, P.: Field validation of food outlet databases: the Latino food environment in North Carolina, USA. Public Health Nutri. **18**(6), 977–982 (2015)
74. Powell, L.M., Han, E., Zenk, S.N., et al.: Field validation of secondary commercial data sources on the retail food outlet environment in the United States. Health Place **5**, 1122–1131 (2011)
75. Lake, A.A., Burgoine, T., Stamp, E., Grieve, R.: The foodscape: classification and field validation of secondary data sources across urban/rural and socio-economic classifications in England. Int. J. Behav. Nutri. Phys. Activity **9**(1), 37 (2012)
76. Federal Geographic Data Committee (FGDC): Content Standard for Digital Geospatial Metadata Workbook. Federal Geographic Data Committee, Washington D.C. (2000)
77. United States Geological Survey (USGS): Spatial Data Transfer Standard (SDTS): Logical specifications. United States Geologic Survey, Reston, Virginia (1997)
78. Salgé, F.: Semantic accuracy. In: Guptill, S.C., Morrison, J.L. (eds.) Elements of Spatial Data Quality, pp. 139–152. Elsevier Science Ltd., New York (1995)
79. Devillers, R., Bédard, Y., Jeansoulin, R.: Multidimensional management of geospatial data quality information for its dynamic use within geographical information systems. Photogram. Eng. Rem. Sens. **71**(2), 205–215 (2005)
80. Qiu, L., Lingling, G., Feng, H., Yong, T.: A unified metadata information management framework for the digital city. In: Proceedings of IEEE's Geoscience and Remote Sensing Symposium, Anchorage, Alaska, pp. 4422–4424 (2004)
81. Farley, T., Rice, J., Bodor, J., Cohen, D., Blumenthal, R., Rose, D.: Measure the food environment: shelf space of fruits, vegetables and snack foods in store. J. Urban Health **86**(5), 672–682 (2009)

Author Proof

82. Cummins, S., et al.: Variations in fresh fruit and vegetable quality by store type, urban–rural setting and neighbourhood deprivation in Scotland. Public Health Nutri. **12**(11), 2044–2050 (2009)
83. Bailey, K., Grossardt, T.: Toward structured public involvement: Justice, geography and collaborative geospatial/geovisual decision support systems. Ann. Assoc. Am. Geogr. **100**(1), 57–86 (2010)
84. Soja, E.: Seeking Spatial Justice. University of Minnesota Press, Minneapolis (2010)
85. North Carolina Geographic Information Coordinating Council (NCGICC): North Carolina State and Local Government Metadata Profile for Geospatial Data and Services. North Carolina Geographic Information Coordinating Council, Raleigh (2018)
86. United States Bureau of the Budget: United States National Map Accuracy Standards. U.S. Bureau of the Budget, Washington, D.C. (1947)
87. Federal Geographic Data Committee (FGDC): Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy (FGDC-STD-007.3-1998) Federal Geographic Data Committee, Washington, D.C. (1998)
88. Morland, K., Wing, S., Diez Roux, A.: Neighborhood characteristics associated with the location of food stores and food service places. Am. J. Prevent. Med. **22**(1), 23–29 (2002)
89. Smith, C., Wright-Morton, L.: Rural food deserts: Low-income perspectives on food access in Minnesota and Iowa. J. Nutri. Educ. Behav. **41**(3), 176–187 (2009)
90. ESRI (Environmental Systems Research Institute): The American Community Survey (2014). https://www.esri.com/library/whitepapers/pdfs/the-american-community-survey.pdf. Accessed 29 Mar 2020

Author Proof

Author Queries

Chapter 9

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	
AQ2	Please check and confirm if the inserted citation of Fig. 8 and Table 5 is correct. If not, please suggest an alternate citation.	