Visualizing Diversity: Spatial Data as a Resource Enabling Extension to Better Engage Communities

Justin Krohn  
*University of Missouri*, jlkvmn@missouri.edu

Jacqueline Davis-Manigaulte  
*Cornell University*, jad23@cornell.edu

Christopher Fulcher  
*Trust for Public Land*, chris.fulcher@tpl.org

Jennifer Sarah Tiffany  
*Cornell University*, jst5@cornell.edu

Follow this and additional works at: https://scholarsjunction.msstate.edu/jhse

Part of the *Education Commons, Life Sciences Commons, Medicine and Health Sciences Commons, and the Social and Behavioral Sciences Commons*

**Recommended Citation**


This Original Research is brought to you for free and open access by Scholars Junction. It has been accepted for inclusion in Journal of Human Sciences and Extension by an authorized editor of Scholars Junction. For more information, please contact scholcomm@msstate.libanswers.com.
Visualizing Diversity: Spatial Data as a Resource Enabling Extension to Better Engage Communities

Acknowledgments
The Center for Applied Research and Engagement Systems (University of Missouri and University of Missouri Extension) provided the technology infrastructure to conduct the spatial analysis required for this article. We would also like to express appreciation to Carol Parker and Julie Fox for their insights and suggestions.
Visualizing Diversity: Spatial Data as a Resource Enabling Extension to Better Engage Communities

Justin Krohn  
*University of Missouri*

Jacqueline Davis-Manigaulte  
*Cornell University*

Christopher Fulcher  
*Trust for Public Land*

Jennifer Sarah Tiffany  
*Cornell University*

Effective Extension programming relies on engaging people of all races, ethnicities, and cultures. Extension educators sometimes struggle with how best to engage communities that are not “traditional” program audiences. Centering data visualization on the strength of Black, Latino/Hispanic, Asian, Native American, and other potentially marginalized communities can assist Extension’s work to engage diverse staff, program participants, and advisory board members. For example, using maps to understand what languages people speak at home strengthens the connections between Extension programs and community participants and can inform staff recruitment and advisory board composition. However, maps of aggregated areas like counties can mask socioeconomic and demographic diversity. The level of analysis (nation, state, county, census tract, block group) and the analytic categories in the dataset shape the stories data can tell. Extension can draw on maps that use granular and comparable levels of analysis (census tracts and block groups, rather than counties) and reduce the risks of abstraction by engaging communities and acquiring primary data to reflect reality at finer levels of geography better. In this article, we will focus on visualizing data in two very different, highly diverse cities: New York City, NY (population 8,823,000), and Albuquerque, NM (population 929,000).

*Keywords:* DEI, mapping, languages, engagement, data, demographics, county, census

Engaging people of all races, ethnicities, and cultures stands as a high priority objective for Extension, an objective that is especially salient in urban Extension programming. Yet Extension educators sometimes struggle with how best to engage communities who are not “traditional” program audiences. Centering data visualization on the strength of Black, Latino/Hispanic, Asian, Native American, and other potentially marginalized communities can assist Extension’s work to engage diverse staff, program participants, and advisory board members.
Race/ethnicity and language are frequently used pathways to assess diversity and can be well visualized using maps. The level of granularity (national, state, county, census tract) shapes the stories the data can tell, as does the nuance and quality of inquiries used to develop the data. Using data to understand what languages people speak most comfortably can strengthen the connections between Extension programs and community participants and inform staff recruitment and advisory board composition. Knowledge of languages spoken at home by community members can aid Extension organizations to ensure that program offerings, volunteer opportunities, resource materials, and recruitment of new staff are designed to engage and welcome all community members. This includes, but is not limited to, translation of written resources, curricula, and presentations. Understanding how to use data to explore the diversity of languages in communities served by Extension opens doors to understanding other dimensions of diversity, equity, access, and inclusion.

People have been creating maps to visualize data and make “invisible” processes visible in understandable ways for centuries. The current practice of mapping natural and social processes has roots in Alexander von Humboldt’s publication of *Cosmos* and its accompanying atlas nearly 200 years ago (Cheshire & Uberti, 2021, p. 18). Extension has a unique role in using and acquiring local primary data and enhancing it with contextual data to tell a grounded story about people and places, the issues that impact communities/regions, and the programs and policies that may benefit human lives and natural processes. Extension can engage communities to minimize data abstraction and provide the needed contextual awareness that lifts local stories and strengthens diversity. We propose that understanding, critically exploring, and visualizing data are valuable components of Extension’s grounded storytelling and community engagement efforts.

In this article, we will focus on visualizing data in two very different, highly diverse cities: New York City, NY (population 8,823,000), and Albuquerque, NM (population 929,000). The power of data is in the details: Extension has traditionally operated at a county level and relied on county-level data. Analysis and visualization can move from county-level data that masks data diversity to census tract and block group unit level data that shines a light on the spatial and temporal variability of data, enabling Extension to meet people where they are. National, state, and county-level data can be highly abstract. Maps generated at that level of abstraction may often be only a blunt policy instrument, while what is needed is granular primary and secondary data to provide more precise identification of issues. The more granular we are in developing interventions (building on local knowledge of assets/issues), the better Extension can be at engaging people, building partnerships, and growing solutions to pressing problems. This tailored approach is also apt to be more sustainable and cost-effective from an economic point of view.
The Modifiable Areal Unit Problem and Its Significance for Urban Extension

A strength of Extension is its capacity to focus on people and places. Gathering and visualizing data can enhance this strength, especially if we ground-truth the data and discern when the data are overly abstract or grouped within arbitrary or overly large geographic units. Counties are important local government entities and are often relied upon to address local needs. However, counties vary widely in both size and population. County boundaries are often delineated arbitrarily and through political processes, with each state taking a different approach.

In geography, there is a phenomenon known as the Modifiable Areal Unit Problem (MAUP) (Openshaw & Taylor, 1979). The problem referred to by MAUP is that aggregation of information in an area with arbitrary boundaries can change with the redrawing of the unit boundaries, such as we see with counties and other common geographical units. These aggregated areas also mask spatial variability related to diversity and other socioeconomic and demographic data. Areas that have arbitrary and variable boundaries are not comparable. When we compare county-level data, we risk comparing apples and oranges because the size and populations of the counties vary so widely.

We see the MAUP problem play out across a variety of disciplines that deal with spatial statistics. Correlations found at an individual level become less pronounced as they are explored in larger aggregated areas. The reverse is also true where apparent relationships at an aggregated level disappear as the data are disaggregated (Openshaw, 1984). Useable knowledge informing effective interventions requires careful, critical choices as to the level of spatial data analyzed and generally draws on more granular geographic data.

Individual information often must be aggregated to protect privacy. It can be difficult to find the “Goldilocks” level of analysis needed to avoid MAUP, preserve confidentiality, and generate meaningful understanding. If an area is too small, there may not be a large enough population to explore questions of interest. If an area is too large, there is risk of overgeneralization and missing local assets and issues.

Not all information is affected by MAUP in the same way (Nelson & Brewer, 2017). The U.S. Census collects data at a variety of geographic levels. The smallest of these is the census block, which has no population requirements and is delineated based on natural or artificial boundaries. Census block groups are an aggregation of census blocks and generally have populations between 600 and 3,000 people. Census tracts are an aggregation of block groups and generally have populations between 2,500 and 8,000. Counties are then made of several census tracts, and each tract is nested neatly within its county, each block group within its tract, and each block within its block group. For privacy reasons, many statistics are suppressed at the census block level. However, levels such as census block group and census tract are found to preserve statistical relationships better than ZIP code tabulation areas (an approximation of ZIP codes.
which are based on routes, rather than areas) or counties (Krieger et al., 2002; Nelson & Brewer, 2017).

The Center for Applied Research and Engagement Systems (CARES; https://careshq.org) built a system for mapping America’s diversity using the CARES mapping engine for data at a variety of geographic levels, including county, census tract, and block group level data. Sub-county data lift the spatial variability of diversity, as shown in the following examples. For brevities sake, this article highlights diversity in two cities: Albuquerque and New York City. However, the data can be visualized for any community in the United States at https://cares.page.link/M2bp. The reader is encouraged to explore diversity in their community or region and overlay a wealth of associated socioeconomic and demographic data that may better inform conditions related to people and place.

Mapping Language Diversity

Knowledge of the languages spoken by people in their homes provides Extension with a vital resource for diversifying staff, program participants, advisory groups, and programming to engage communities effectively in Extension efforts. English is the principal and sometimes the sole language in which Extension currently operates at the same time as Extension often seeks to reach participants in communities where majorities speak a language other than English at home. There are strong efforts toward not only translating but also co-creating resources in multiple languages to fully engage the strengths of local communities and develop necessary relationships for authentic collaboration in programming. Like efforts to map data about human and natural processes, work to develop multilingual capacity is not entirely new, and there is history on which Extension can build its current efforts. For example, more than 100 years ago, Extension urban home demonstrations were sometimes “conducted in seven languages, as it was in Utica in 1918.” (Smith, 2013, p. 160). Visualizing granular data on language use can aid Extension in strengthening its capacity to create and offer programming in people’s preferred languages.

A simple language diversity index was calculated from American Community Survey (ACS 2015-2019) five-year estimates of the language spoken at home. Figure 1 offers an example from Albuquerque, NM, which measures the chance of any random two residents in an area speaking the same language, but it does not consider the relatedness of languages. The index is calculated as \( A = 1 - \sum (i^2) \), where \( i \) successively takes on the values of the total number of speakers of a language divided by the total population (Greenberg, 1956). With this measure, a score closer to 1 indicates a higher level of language diversity, while a score of zero indicates that everyone speaks the same language. Language diversity calculated this way at both the county and census tract levels around the city of Albuquerque, NM, is shown in Figure 1.

In Figure 1, the smoothing effect of county level aggregation with the index score is visible. The census tract level better illuminates the higher levels of language diversity within the city of
Albuquerque and areas within the surrounding counties with high Native American populations. On the maps, the areas with deeper colors have higher language diversity.

Figure 1. Maps Visualizing Languages Spoken at Home in Albuquerque and Nearby Communities, Contrasting County Level and Census Tract Data

New York has a much larger immigrant population. The importance of granular data is clear. The tract level map of New York highlights the spatial variation of diversity, while the county level map looks homogenous (see Figure 2). Much of Queens is a deep blue at the census tract level, which indicates a very high level of diversity, but there are pockets of low diversity that stand out, such as the area around the Corona neighborhood. The census tract level data also provides insight into Brooklyn, which appears evenly split, with a high diversity area in the southwestern portion and lower diversity in the northeastern portion. You can see an interactive version of this map online at https://cares.page.link/zQX6.
Extension educators are ideally positioned to understand and build upon the diversity illuminated by maps using census tract level data, drawing on their lived experiences and connection with the lived experiences of program participants, as well as their understanding of data. This exercise in mapping language diversity highlights the importance of analysis and visualization of data at the sub-county level. We can expand this example further by mapping the most common languages spoken in an area. By combining these two metrics, we hope to demonstrate the utility of sub-county level analysis and present its potential as a practical application that Extension personnel can use in their work.

We mapped the most common language spoken at home around Albuquerque as measured in the 2010-2014 ACS (see Figure 3). The tract level data show small areas of Native American languages and one census tract of other Indo-European languages that are lost at the county level map. At both the county and tract levels, “Spanish” and “Other” are the most common categories.
Changes in the census questions asked can illuminate or erase diversity. In this case, earlier ACS data included specific information about Native American languages that were dropped in the surveys starting in 2016. To enable direct comparison with newer ACS data, older, more detailed language categories were recoded into the newer categories, which caused a loss of precision in the data. One consequence is that areas with a high level of Native American speakers have been lumped into the “Other” category. Recognizing the consequences of this change in data gathering is important for Extension programs seeking to be equitable and inclusive, and the local knowledge and lived experiences of Extension educators and participants serve as a vital corrective to the abstract and apparently homogenous data.

In the more recent ACS data, the areas with high Native American populations again stand out. However, we also see additional census tracts with Vietnamese and Chinese speakers at home (see Figure 4). While the county level map did not change between survey years, at the tract level, we can see changes that would otherwise be missed at the county level. You can see an interactive version of this map online at https://cares.page.link/Ukva. Note that in this and other
maps identifying the most common language spoken at home, the language groups are consolidated to reflect 2015-2019 ACS categories so that Native American languages are placed in the “Other” category. This makes local knowledge of languages spoken and local community connections immensely significant. See our discussion in the section on “The Consequences of Consolidated Categories” and comparable visualizations of those data in Figures 7 and 8.

**Figure 4. The Most Common Language Spoken at Home Around Albuquerque as Measured in the 2015-2019 American Community Survey (ACS)**

The language maps of New York show striking contrasts between the county and census tract levels (see Figure 5). On the county level map, each county in the frame has more Spanish speakers than any other language. However, when we dig a little deeper to the census tract level, we see many more languages represented.
**Figure 5. The Most Common Language Spoken at Home Around New York City as Measured in the 2010-2014 American Community Survey (ACS)**

Drawing on data from the 2015-2019 ACS survey, the county level map shown in Figure 6 is the same as Figure 5, but the spatial distribution has changed between the years at the census tract level. For Extension personnel working in New York, maps such as these at the census tract level, along with input from local leaders, organizations, and individuals, can help inform programming that is responsive to changing community needs. You can access an interactive version of this map online at [https://cares.page.link/P5QW](https://cares.page.link/P5QW).
Figure 6. The Most Common Language Spoken at Home Around New York City as Measured in the 2015-2019 American Community Survey (ACS)

The Consequences of Consolidated Categories

How data is gathered (the questions asked, the reach of the survey) determines both accuracy and nuance. This is demonstrated in Figure 7 and Figure 8, which show changes in language categories used in the 2010-2014 ACS (32 options) and the consolidated categories used in the 2015-2019 ACS (14 options) (see Figure 7 and Figure 8, respectively). In both figures, the data used are from the 2010-2014 ACS. In the first panel on each figure, the data have been recoded to reflect the consolidated language categories used to generate the 2015-2019 ACS. The loss in nuance and granularity is clear. Much understanding about what languages people speak at home is lost when the consolidated data are mapped.

Figure 7, for example, demonstrates how the reduction in the number of language categories erases Indigenous languages spoken by large communities in and around Albuquerque by naming them “Other” rather than “Navajo” and “Other Native American,” as is done when the full data are mapped. Local knowledge and commitment to diversity, equity, and inclusion by
Extension can help to address this erasure. A similar loss of nuance when the 2015-2019 categories are used to map languages spoken in New York City is evident (see Figure 8).

Granularity and centering how data are visualized on the strengths and diversity of local populations matters from a public policy, economic, and equity perspective. Changes in the American Community Survey (ACS) reduced details gathered about Native North American languages and many other languages, making local knowledge an even more important complement to ACS and census datasets. As shown above, the impact of these changes can be readily visualized when maps using the earlier, more nuanced data are laid side-by-side with maps using more recent data.

*Figure 7. Language Diversity in Albuquerque and the Surrounding Area Shown in Two Versions of 2010-2014 ACS Language*

*Note.* The first panel consolidates the data into the 14 categories used in the 2015-2019 ACS while the second panel presents the full 32 language categories used in the original 2010-2014 survey (also see Figure 3).
**Figure 8. Language Diversity in New York City and the Surrounding Area Shown in Two Versions of 2010-2014 ACS Language**

The first panel consolidates the data into the 14 categories used in the 2015-2019 ACS while the second panel presents the full 32 language categories used in the original 2010-2014 survey (also see Figure 5).

**Representing Race and Ethnicity**

The unit of analysis (state, county, census tract, census block group) and the demographic categories used determine what maps illuminate. The maps shown in Figure 9 were made to emulate and extend the maps first produced by The Brookings Institute, where the authors conducted their analyses at the county level ([https://www.brookings.edu/research/mapping-americas-diversity-with-the-2020-census](https://www.brookings.edu/research/mapping-americas-diversity-with-the-2020-census)). Using the national average of representation by each racial or ethnic group, the Brookings Institute maps show counties that have a higher-than-average representation of specific groups. The contrast between the county level maps shown in Figure 9 and the more nuanced maps using census tract data shown in Figure 10 is evident. At the county level, the categories “Hispanic or Latino” and “two or more” races/ethnicities are highly represented around Albuquerque; around New York City, “Hispanic or Latino,” “Asian/Pacific Islander,” and “two or more” races/ethnicities are highly represented. At the
census tract level, we see many other groups represented in both Albuquerque and New York City, with all demographic groups but “Native American/Alaska Native” highly represented around New York City. There are multiple ways to represent the same type of data, illuminating different social relationships. The Brookings Institute determined that a relative value for mapping the racial/ethnic composition of an area’s population (“higher than the national average”) was of value. Framing questions and the geographic units mapped are key elements in generating knowledge that illuminates the strength of diversity and fosters inclusion rather than erasure of local assets.

Figure 9. Maps Showing Areas with Higher Than National Average Proportions of Non-White Racial and Ethnic Groups in the Albuquerque and New York City Areas Developed Using County-Level Data
Figure 10. Maps Showing Areas with Higher Than National Average Proportions of Non-White Racial and Ethnic Groups in the Albuquerque and New York City Areas Developed Using More Granular and Detailed Census Tract Level Data

The Path Forward

Mapping is inherently an abstraction of reality. It is a powerful tool that must be used mindfully. One role that Extension can play, given its presence in communities throughout the United States, is to ground maps in local understandings so they address locally important questions. Extension can draw on maps that use granular and comparable levels of analysis (census tracts and block groups, rather than counties) and reduce the risks of abstraction by engaging communities and acquiring primary data to reflect reality at finer levels of geography better. Extension has a unique role in interpreting local primary data and enhancing it with contextual data to tell a grounded story about people and places, the issues that are impacting communities/regions, and the programs and policies that will benefit human lives and natural processes.
The technologies are in place to make this happen. Leadership alignment is needed to train and position Extension to serve as a sentinel organization and partner with other anchor community organizations to utilize mapping fully as a tool for community engagement, program excellence, and promotion of diversity, equity, and inclusion.

This brief article focused on how data can be visualized, considered the differences between commonly used county level data and census tract data, and explored the impact of how data are categorized on the maps produced. The article focused specifically on selecting and visualizing data, which is a deliberate limitation of the article. We believe that well-understood and carefully visualized data are resources for lifting local stories and strengthening diversity within Extension. We hope this discussion of visualizing diversity data will provide a bridge to the practical use of maps to inform Extension practices.

References


*Justin Krohn* is a Research Project Analyst at the Center for Applied Research and Engagement Systems (CARES). He is also currently working on his Ph.D. in geoinformatics from the University of Missouri with a focus on remote sensing and conservation. Please direct correspondence to Justin Krohn at jlkvmn@missouri.edu.
Jacqueline Davis-Manigaulte serves as Director of Community Relations and Program Leader for Family and Youth Development at Cornell University Cooperative Extension’s New York City Programs. A Senior Extension Associate, her research and program focus areas include positive youth development, STEM programming, parenting, and adult education.

Chris Fulcher serves as Director of Business Development and Operations at the Trust for Public Land, a national non-profit that creates parks and protects land for people, ensuring healthy, livable communities for generations to come. He is an Assistant Extension Professor Emeritus at the University of Missouri and former Director for the Center for Applied Research and Engagement Systems (CARES).

Jennifer Sarah Tiffany, Cornell University Bronfenbrenner Center for Translational Research, serves as Executive Director for Cornell University Cooperative Extension’s New York City Programs as well as the Community Engagement in Research Core Component Director with Weill Cornell Medicine’s Clinical and Translational Science Center.

Decision on Author Order: We saw the conceptualization and development of this article as a joint effort, and each of us brought unique contributions. Justin Krohn is listed first because he designed and generated the maps and is shouldering the role of corresponding author. The other three authors are listed alphabetically by last name.

Acknowledgments

The Center for Applied Research and Engagement Systems (University of Missouri and University of Missouri Extension) provided the technology infrastructure to conduct the spatial analysis required for this article. We would also like to express appreciation to Carol Parker and Julie Fox for their insights and suggestions.

Resources

We hope this article encourages readers to use mapping to explore and visualize local data. There is no need to become a demographer or specialist in Geographic Information Systems (GIS). The following websites offer useful tools for exploring questions that are generated by and relevant to local communities and contexts.

Center for Applied Research and Engagement Systems. [https://careshq.org](https://careshq.org)
CARES hosts over 30,000 map layers and 6,000 data indicators from more than 700 datasets for topics ranging from agriculture, the environment, education, health, economics, demographics, and many more. Maps and data can be queried and exported.
The Community Data Tools developed by the University of Wisconsin-Madison’s Extension Division offers links to a range of data access and mapping sites and resources.

Map a Syst. https://www.mapasyst.org/
Map a Syst is a community dedicated to geospatial technology and education whose goal is “to inspire people to use mapping for outreach and engagement and provide resources to help people put mapping to work in their lives.”

Mapping 101. https://www.extension.iastate.edu/ffed/mapping-your-food-system/
Iowa State’s Mapping 101 offers easy-to-use tutorials that walk participants through QGIS—a free, open-source geographic information system (mapping software). Participants can develop their own maps for a project, program, or area of interest relevant to their particular place.

The website of the National Association of Realtors offers housing statistics at the national, regional, and metro-market level and can be used in tandem with environmental risk assessment sites like https://floodfactor.com/.

The National Equity Atlas offers actionable data and strategies to advance racial equity and shared prosperity with a focus on equity metrics. Data can be disaggregated by race/ethnicity, gender, nativity, ancestry, and income.

Online Mapping Tools. https://extension.unh.edu/nhnriguide/mapping-options/online-mapping-tools
The University of New Hampshire’s Online Mapping Tools site offers a variety of free and easy to use mapping resources.

Social Explorer. https://www.socialexplorer.com
Social Explorer provides resources and thousands of maps visualizing demography, economy, health, education, and more with a focus on the United States.

StatsAmerica. https://www.statsamerica.org/
StatsAmerica provides actionable data for economic developers to use in site requests, development metrics, grant writing, and strategic planning.

The U.S. Census provides a wide variety of data tools to view various census data including demographics, business dynamics, commuting, geocoding, and more.
Urban Institute. [https://www.urban.org/data-viz](https://www.urban.org/data-viz)

The Urban Institute’s data site provides statics, graphs, and maps on economic and social policy with strong resources on urban and equity issues.

University-Specific Resources.

For example, Cornell University’s Geographic Information Systems program ([https://guides.library.cornell.edu/gis](https://guides.library.cornell.edu/gis)) provide a variety of learning and mapping resources to users with library access.