THREE STUDIES USING BUSINESS NAMING PATTERNS TO DELIMIT REGIONAL GEOGRAPHIES

By

JESSE R. ANDREWS

Bachelor of Arts in History
University of North Carolina Greensboro
Greensboro, North Carolina
2013

Master of Arts in Geography
Appalachian State University
Boone, North Carolina
2017

THREE STUDIES USING BUSINESS NAMING PATTERNS TO DELIMIT REGIONAL
GEOGRAPHIES

Dissertation Approved:

Dr. Allen Finchum
_____

Dr Dale Lightfoot
_____

Dr. Jonathan Comer
_____

Dr. Laura Arata
_____

ACKNOWLEDGEMENTS

I want to thank my advisors, Dr. Allen Finchum, and Dr. Dale Lightfoot. Dr. Lightfoot provided me with an excellent stream of professional advice and support as an imitable model of an academic. Likewise, Dr. Finchum has proven a valuable collaborator, critic, and friend. Also, thanks to Dr. Jon Comer, whose quantitative method class reignited my interest in statistical analysis, and Dr. Stadler for providing much-needed bonhomie within the department.

Name: Jesse Andrews

Date of Degree: July 2022

Title of Study: THREE STUDIES USING BUSINESS NAMING PATTERNS TO DELIMIT REGIONAL GEOGRAPHIES

Major Field: Geography

Abstract: The following work contains three articles delimiting and interrogating the concept of vernacular regions using the spatial distribution and frequency of keywords in business names. The basics of this approach are not novel. Sociologist John Shelton Reed pioneered the method to define a vernacular South in his 1976 piece "The Heart of Dixie: An Essay in Folk Geography." However, these three articles extend this line of scholarship temporally, methodologically, and theoretically.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Introduction

The following work contains three articles delimiting and interrogating the concept of vernacular regions using the spatial distribution and frequency of keywords in business names. The basics of this approach are not novel. Sociologist John Shelton Reed pioneered the method to define a vernacular South in his 1976 piece "The Heart of Dixie: An Essay in Folk Geography." However, these three articles extend this line of scholarship temporally, methodologically, and theoretically.

The first article in this collection, titled "Paring Dixie Down," contained in chapter two, examines the effects of the old Dixie Highway on naming patterns in the areas it traversed. The highway has distorted prior scholars' attempts to map the South as a vernacular region via business names. This article serves as the foundational piece of this series and led to my initial interest in the methods used in the other two pieces. It is also essential to this series, as it was the first written and contains a broader critique of using place names to delimit regions. This article was published in the Fall 2020 issue of the *Southeastern Geographer* (Andrews & Finchum 2020), and since then, I have received several kind notes and incorporated several suggestions from scholars cited in this work.

The second article, chapter three, extends Reed's work back to the early 20th Century using 1910 and 1930 city guide data. I initially began this line of research to observe changes in the region's boundaries over time, with a particular interest in confirming the findings of Chapter

Two. The article largely succeeds at this aim. This approach generates mostly coherent historic

boundaries for the region, with one notable exception. However, the piece also reveals important

limits to the methods. This piece also historicizes Reed and his follower's scholarship and

questions some of the underlying assumptions as based on a particular place and time. The piece

argues that these assumptions do well in modeling the 1930 data, but they seem to break down

when modeling data from 1910, especially the word "Dixie." I intend to submit this article to the

Southeastern Geographer, as it is in some ways a follow-up to the first article.

The final piece uses colocational quotient (CLQ), a statistical method for examining

clustering in categorical variables, to offer a fine grain analysis of the spatial distribution of the

keywords "Great Plains," Midwest," "Prairie," and "High Plains" terms associated with the fuzzy

boundaries between the Midwest, Great Plains, and the Rocky Mountains. This piece grew out of

my earliest discussions of vernacular regions with Dr. Finchum. We discussed the findings of a

poster hanging in the department on the location of the perceptual Midwest. We began web

scraping regional terms in business names, and from this starting point, the project evolved into

chapter two.

After completing chapter two, I sought to return to the vernacular South. However, using

various ratio approaches on contemporary data did not provide interesting results. So, another

method was needed. My background in urban historical geography proved helpful here, as I had

become aware of the CLQ metric from readings on historical residential segregation (Cordoba &

Walter 2016; Cordoba et al. 2018). After checking the relevant scholarship, I recognized that

scholars had not used the metric to study regional geography. So, an article using the metric

could be methodologically innovative and address a topic that recent scholars have barely

explored. Initially, I used CLQ on a series of subregional terms within the South, producing

some legible but unsurprising maps. After further experimentation, the method worked better when delimiting fuzzy regional boundaries. So, I adapted the method to the Great Plains and Midwest and eventually extended the project westward to accommodate the term "high plains" and added the term "prairie" after noticing the interesting regional patterns that the keyword took.

It turns out that CLQ works particularly well in fuzzy borderlands regions, and the article establishes the method as a viable approach to delimiting regional differences using categorical point data. I intend to submit this article to the *Professional Geographer*. The *Professional Geographer* focuses on shorter empirical studies, such as this, and has published the last substantial article (Rossum & Lavin 2000) on delimiting the geography of the Great Plains.

Vernacular Regions

Perceptual or vernacular regions are those perceived to exist by their inhabitants and other members of the population at large. They exist as part of the popular or folk culture. Rather than being an intellectual creation of professional geographer, the vernacular region is the product of the spatial perception of average people. Rather than being based on carefully choose quantifiable criteria, such regions are the composites of the mental maps of the population.

-Terry G. Jordan (1978)

The study of vernacular or perceptual regions entered geography in the late 1960s. The first significant attempt to map these regions on a national level was Ruth Hale's (1971)

"A Map of Vernacular Regions in America." Hale used a survey-based approach to define vernacular regions. She mailed thousands of surveys across the country, asking which regions the participants felt they lived. Using this data, she constructed a regional map of the United States. Her work remains one of the most ambitious attempts at defining vernacular regions using surveys.

John Shelton Reed's "Heart of Dixie" (1976) sparked considerable scholarly interest in vernacular regions. Reed counted business names beginning with "Dixie" and "Southern" and divided those totals by "American" to create a ratio that he mapped via a crude form of interpolation. This approach was the first to move away from using interview data to map these regions and proved a popular, reproducible approach widely adopted to delimit vernacular regions. It was Reed's work, along with Terry Jordan's work on perceptual areas of Texas (1978) and Wilbur Zelinksy's (1980) mapping of the United States vernacular regions, that brought the technique prominence leading to a flood of publications attempting to define vernacular regions at a smaller state or regional scale (Lamme III & Oldakowski 1982; Shortridge 1980; Shortridge 1985; Good 1981; Raitz & Ulack 1981; Zdorowski & Carney 1985).

However, by the end of the 1980s, work on vernacular regions virtually disappeared with two important exceptions: Craig Colton's 1997 "The Land of Lincoln: Genesis of a Vernacular Region" and Douglas Heath's (1993) study of highly localized vernacular regions in the Allentown, Pennsylvania area. Colton's piece is one of the most imaginative pieces written on the concept. Writing between the time that studies on state vernacular regions had fallen out of fashion and before GIS transformed methods, Colten (1997) uses a mix of quantitative methods and fieldwork to identify a "Land of Lincoln" region stretching from his birthplace in central Kentucky to his adult home in central Illinois. This piece is an excellent example of using mixed

methods to examine a little-known vernacular region, and it is one of the few pieces that attempt to study a region that is not based on historical settlement patterns. Similarly, Heath's work is the only study of localized vernacular regions. Examining a group of counties in the Allentown-Bethlehem Pennsylvania area, Heath finds a group of sub-county level vernacular regions that largely correspond to the region's school districts.

In the past two decades, the meaning of "perceptual" and "vernacular" seem to have diverged in the scholarship. "Perceptual" has become a term used for survey-based studies (Lowry et al. 2008,) while vernacular is used by scholars in studies that quantify features of the landscape to study region (Liesch et al. 2015). Additionally, as GIS technology has become more accessible to non-specialist scholars and digital databases of businesses were added to academic libraries, scholars have begun revisiting vernacular regions. However, few of these works attempted to innovate in methods or theory despite using new statistical tools. Many were simply updates on past studies from the 1980s (Lamme III & Oldakowski 2007; Ambinakudige 2009).

After this brief revival of scholarship on vernacular regions, with one crucial exception outside of the field of geography (Bricker 2018), work on the subject has been virtually nonexistent. Except for my own, no works have been published on delimiting vernacular geographies in the past five years. This absence is unfortunate, as I believe there is still much fruitful work to be done on the subject, and I hope the following three proposed works can help fill those gaps and encourage the wider adoption of these methods in the discipline.

A final note. The following three articles are all inspired by this line of scholarship, particularly the use of business names to delimit region. However, none use the precise methods of past works. This update in methods is partly due to the particularities of the data and approach taken in each piece and partially due to advances in technology that make processing the data

much more efficient. Additionally, they contain critiques of past work. However, this should be by no means seen as a slight upon these scholars. They were working with much cruder tools and relied upon inspired use of limited data, clever methods, and time-consuming archival work compared to the relative ease of the brute force computation available to me. I hope the following three works illuminate a path forward in this area, updating what these earlier scholars have done with a combination of their methodical moxie and modern computational methods.

Bibliography

Ambinakudige, S. (2009). Revisiting "the south" and "dixie" delineating vernacular regions using GIS. *Southeastern Geographer*, *49*(3), 240–250.

Andrews, J. R., & Finchum, G. A. (2020). Paring Old Dixie Down: The Dixie Highway and the Mapping of a Vernacular South. *Southeastern Geographer*, *60*(4), 345–359.

Bricker, C. (2018). *Vernacular geography and perceptions of place: A new approach to measuring American regional and political subcultures* [Dissertation]. University of Iowa.

Colten, C. E. (1997). The land of Lincoln: Genesis of a vernacular region. *Journal of Cultural Geography*, *16*(2), 55–75.

Córdoba, H. A., Walter, R. J., & Foote, N. S. (2018). The residential segregation of San Antonio, Texas in 1910: An analysis of ethno-racial and occupational spatial patterns with the colocation quotient. *Urban Geography*, *39*(7), 988–1017.

Córdoba, H. A., & Walter, R. J. (2016). Measuring the Spatial Pattern of Ethnic Groups in San Antonio's Eastside: A Spatial Analysis Approach with the Colocation Quotient. *Papers in Applied Geography*, *2*(2), 160–177.

Good, J. K. (1981). The vernacular regions of Arkansas. *Journal of Geography*, *80*(5), 179–185.

Hale, R. (1971). *A Map of Vernacular Regions in America. Unpublished doctoral dissertation. Minneapolis: Univ. Of Minneapolis, 1971.*

Heath, D. E. (1993). Highly Localized Vernacular Regionalization in the Allentown-Bethlehem Area, PA-NJ. *The Professional Geographer*, *45*(3), 251–263. https://doi.org/10.1111/j.0033-0124.1993.00251.x

Jordan, T. G. (1978). Perceptual Regions in Texas. *Geographical Review*, *68*(3), 293–307. JSTOR. https://doi.org/10.2307/215048

Lamme III, A. J., & Oldakowski, R. K. (1982). Vernacular areas in Florida. *Southeastern Geographer*, *22*(2), 99–109.

Lamme III, A. J., & Oldakowski, R. K. (2007). Spinning a new geography of vernacular regional identity: Florida in the twenty-first Century. *Southeastern Geographer*, *47*(2), 320–340.

Liesch, M., Dunklee, L. M., Legg, R. J., Feig, A. D., & Krause, A. J. (2015). Use of business-naming practices to delineate vernacular regions: A Michigan example. *Journal of Geography*, *114*(5), 188–196.

Lowry, J., Patterson, M., & Forbes, W. (2008). The Perceptual Northwest. *Yearbook of the Association of Pacific Coast Geographers*, *70*(1), 112–126. https://doi.org/10.1353/pcg.0.0009

Shortridge, J. R. (1980). Vernacular regions in Kansas. *American Studies*, *21*(1), 73–94.

Shortridge, J. R. (1985). The vernacular middle west. *Annals of the Association of American Geographers*, *75*(1), 48–57.

Raitz, K., & Ulack, R. (1981). Appalachian vernacular regions. *Journal of Cultural Geography*, *2*(1), 106–119.

Reed, J. S. (1976). The Heart of Dixie: An Essay in Folk Geography. *Social Forces*, *54*(4), 925–939. https://doi.org/10.1093/sf/54.4.925

Rossum, S., & Lavin, S. (2000). Where Are the Great Plains? A Cartographic Analysis. *The Professional Geographer*, *52*(3), 543–552. https://doi.org/10.1111/0033-0124.00245

Zdorowski, R. T., & Carney, G. O. (1985). This land is my land: Oklahoma's changing vernacular regions. *Journal of Cultural Geography*, *5*(2), 97–106.

Zelinsky, W. (1980). North America's Vernacular Regions. *Annals of the Association of American Geographers*, *70*(1), 1–16. JSTOR.

CHAPTER II

Paring Old Dixie Down: The Dixie Highway and the Mapping

of a Vernacular South

Highlights

- Frequency of the term "Dixie" in phonebook listings grouped by ZIP Code recorded to create a regional map of the South.

- Concentrations of "Dixie" in the middle Ohio Valley have been noted as anomalous in prior work.

- Excluding listings within a 0.5 km buffer significantly changes the hotspot map—the entire Ohio Valley is removed from the hotspot, along with East Tennessee and South Florida.

- Dixie Highway appears to influence naming patterns in areas in the immediate vicinity significantly.

Abstract

This study attempts to approximate the boundaries of the American South using the frequency of "dixie" in the names of public establishments. We discover anomalous clusters of these establishments near the old Dixie highway system. This nascent highway system was a significant driver of tourism and commercial development during the early 20th Century. To verify the overrepresentation of "dixie" names along the route, we approximated the methods of prior studies, recording all instances of "dixie" from approximately 1,600 phone book directories in the contiguous 48 United States and mapping them by ZIP Code. Using this method, we find that 39 percent of instances of "dixie" in the ten states the system crossed lie within a 0.5-kilometer buffer of the route. After removing these instances, we demonstrate that their exclusion significantly shrinks the regional hotspot. We conclude that prior scholarship utilizing this technique has inflated the region's size by including instances of establishments named "dixie" to signal proximity to the Dixie Highway system. Likewise, we conclude that qualitative scholarship on naming conventions and the theoretical basis for using aggregated names as proxies for vernacular regions are underdeveloped, and further work is needed in both areas.

KEYWORDS: Vernacular Regions, Dixie Highway, American South

Introduction

This study approximates the boundaries of the American South using keywords from phone book listings. This approach is not novel. However, we argue that prior scholarship using this technique has inflated the region's size by including instances of the establishments whose name included "Dixie" due to proximity to one of the first interstate highway systems – the Dixie Highway. This early series of trunk roads was a significant source of regional mobility, tourism, and commercial development during the early 20th century (Ingram 2016). Portions of the old roadway still serve as a major thoroughfare in several large cities today, and per this study, the roadway appears to have had a significant effect on naming patterns in its immediate vicinity.

Recently, the continued use of the Dixie moniker for the portions of the roadway has caused controversy. There have been organized efforts to change the highway's name in many South Florida communities, with some success. Miami-Dade commissioners recently pledged to remove Dixie signage in the county (Burch 2020; Martinez 2020). This makes sense considering our study, which finds the Dixie-named businesses proximate to the Dixie Highway are most densely concentrated in this region.

This study utilizes well-developed methods to investigate vernacular regions. Since the popularization of the concept of vernacular regions in the mid-1970s, scholars have employed two mapping methods. One method relies on surveys of residents in a defined study area. Although it has recently fallen out of favor, this approach was frequently used throughout the late 1970s and early 1980s (Jordan 1978; Shortridge 1980; Lamme III & Oldakowski 1982). The alternative method employed in this study collects names and addresses of public establishments

14

listed in phone directories. This approach has been widely used in recent years to determine vernacular regions at various scales, although it also has its origins in the 1970s. (Weaver et al. 2018; Liesch et al. 2015; McEwen 2014).

The concept of vernacular regions and the methods of constructing them arose out of American geography and sociology in the 1970s have remained overwhelmingly provincial. With a few notable exceptions, studies have focused on the regions of the United States and particularly the United States South, along with the individual states that comprise the region. However, recently there has been work using the concept outside of North America, particularly in the Balkans (Musa et al. 2015; Vukosav & Fuerst-Bjeliš 2016). Nevertheless, the concept, with the need for widely accessible public data and a tradition of dispersed business ownership along with flexible naming conventions, remains primarily North American.

Literature Review

Sociologist John Shelton Reed was the first to attempt to map a vernacular region based on public business name listings. In his seminal piece 'The Heart of Dixie: An Essay in Folk Geography' (1976), Reed utilized phone books from 100 cities across the country to construct a ratio of keywords that indicate "southernness," including "south," "Dixie," and "southeast" compared to terms such as "national" and "American." Areas with high ratios of southern terms to national terms were used as a proxy for the South as a region. His work finds a core South centered on the Deep South and a large peripheral South in surrounding states. This peripheral South extends as far north as southern Indiana and encompasses most of Florida, east Texas, and

southeastern Oklahoma while excluding most of West Virginia and Northern Virginia. Significantly, Reed acknowledged the unusually high ratio of Dixie place names in Dayton and Cincinnati, Ohio, attributed it to the highway, and excluded the two cities from his analysis of that keyword. However, he did not attempt to quantify the effect of the highway on the study overall or acknowledge that its effects likely extended beyond southwestern Ohio.

Four years after Reed's initial study, geographer Wilbur Zelinsky (1980) published a more ambitious study that attempted to quantify numerous vernacular regions across the United States and Canada using similar methods but eschewing the ratio approach adopted by Reed, instead of using only term frequency. Both authors acknowledged that their studies had significant limitations, including using only terms that appeared at the beginning of the names of establishments. Additionally, both scholars limited their study area to larger cities, so data on smaller cities and rural areas were not examined. These limitations were primarily products of the technical constraints of the time. Phone books were not digitized, nor were advanced GIS and database technologies available to automate the search through data, so collecting data from the entire country was time and labor prohibitive.

In the four decades since, several follow-up studies have been conducted to measure the change in the extent of the vernacular South as measured by Reed and to do more extensive and more fine-grained analysis enabled by advances in computing technology. Reed and his coauthors (Reed et al. 1990) followed up his original study fifteen years later, using the same methods, and found a reduced Dixie area. Alderman and Beaver (1999) updated the original Reed study replicating his methods to determine if Reed's South had moved in the interceding quarter-century using 1988 and 1998 phonebooks. Ambinakudige (2009) and Cooper and Knotts (2010) use Reed's concept of word ratios with modern GIS technology. Notably, Cooper and

17

Knots continue to limit their search to areas corresponding to the 100 cities used in Shelton's original analysis. However, by using online phonebooks, they included establishments that contain Dixie anywhere in the name in their analysis. Ambinakudige went beyond the original methods of the Shelton piece by using phone directories that cover the entire lower 48 states and associating them with specific ZIP Codes. This approach creates a much larger study area and the potential for finer grain analysis. Furthermore, while Ambinakudige uses the word ratio concept, he utilizes a more sophisticated spatial statistic, local indicators of spatial association (LISA), to determine clustering in spatial data.

The most recent and thorough attempt using the ratio approach pioneered by Reed is an updated study by Cooper and Knotts (2017). The authors map an updated South using Reed's original ratio approach in this piece. However, instead of using the same 100 cities as Reed, as they had done in the 2010 study, Cooper and Knotts expanded the study area to include every city with over 50,000 in the eleven states of the Confederacy, along with Kentucky and Oklahoma. Most notably, the authors go beyond merely mapping a more expansive list of cities' ratio data. Cooper and Knotts, using a business analytics database, compare the size and type of business as well demographics of the owners that use "dixie" or "southern" in their name. They find that Dixie-named businesses, compared to "southern" businesses, tend to have lower revenues, were founded earlier, have lower credit ratings, were more likely to be owned by lone individuals, and were more likely to be retail or food service businesses.

Nonetheless, despite the different approaches, the studies produced roughly similar maps of the region, with the differences (beyond the differing time scales) likely arising from the varying methods used to delimit the boundaries. For instance, the Cooper and Knots study used kriging to draw the region's boundaries, likely to approximate the look of the older Reed and

Zelinsky pieces, both of which use an unspecified form of interpolation. In contrast, Ambinakudige found disconnected islands of Dixie, findings likely due to his use of a more expansive dataset and a more fine-grained statistic. Notably, all the previously cited literature finds that the number of Dixie-named businesses has declined over the past four decades.

Several other vital lines of inquiry in these studies of vernacular regions using keywords have been understudied and make the underlying data more challenging to interpret, especially across the extended temporal scales – a limitation we hope to address in future work. Most importantly, we need a better understanding of how particular names are chosen and the historical evolution of these naming practices in general, and how those patterns relate to "dixie" naming practices in particular.

Gunderman and Harty (2017) are the most recent scholars to examine business-naming patterns. Using data scraped from online phone listings, like this study, the authors map business names inspired by the long-touring rock band, the Grateful Dead. Additionally, the authors interviewed four businesses with Grateful Dead-inspired names. The owners interviewed acknowledge that Grateful Dead-inspired names are less recognized by the public than in the past; nonetheless, they express a desire to recreate a particular cultural atmosphere associated with the band. Another recent work dealing with the subject at the scale necessary to inform this study is Bletzer's (2003) article on Latinx business naming practices in rural South Florida, the results of which are specific to that particular community and challenging to universalize. Beyond these pieces, no recent qualitative work has been done on the practice of naming small businesses, scholarship that is sorely needed to develop a more nuanced theory of vernacular regions.

The Dixie Highway

The Dixie Highway system, today forgotten outside of some vestigial sections of the road that keep the original Dixie name, was one of the first attempts at an interstate highway system.

The highway, a hodgepodge of state, federal, and locally maintained roads constructed between 1915 and 1926, was the first attempt at a highway system running through the Southeast (Figure 1). The highway, first proposed by automobile booster, auto racing pioneer, and Florida real estate investor Carl G. Fisher, was envisioned to boost tourism in rapidly developing South Florida (Ingram 2016).

During this period, most roads through the South were disconnected hub and spoke systems connecting agricultural communities to local railroad hubs. Many were often poorly maintained dirt roads that frequently became impassable during heavy rains. Likewise, the system connected towns to a more extensive system of roads and offered a way to bypass the railroads' effective monopoly on moving agricultural goods, thereby appealing to rural farmers throughout the South. So, this proposal for a modern road system running from the Michigan-Canadian border to South Florida stitched together a coalition of midwestern and Florida business interests, town boosters along the potential routes, and farmers throughout the region looking for relief from excessive transportation fees charged by the railroads (Ingram 2016).

Figure 1. Outline of the Dixie Highway System (Shuttering 1919)

The completed Dixie Highway system was not a single route but a roughly 9,300-kilometer-long series of routing compromises originating in Sault St. Marie on the Canadian border and Chicago and running to South Florida. The system crisscrossed ten states: Michigan, Illinois, Indiana, Ohio, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, and Florida (Ingram 2016). New roadways have replaced much of the original route, including portions of I-75, I-95, I-26, and I-40, yet pieces of the original Dixie Highway remain throughout the areas it connected. Likewise, the highway lives on in popular memory through historical markers and events such as the Dixie Highway 90-mile yard sale, a yearly event running on the old highway route between Ringgold and Ackworth, Georgia (Jackson n.d.).

Data and Methods

This study employs similar methods as previous studies of vernacular regions. We collect all instances of the keyword "Dixie" from approximately 1,600 online phone listings in the contiguous 48 of the United States and map them according to ZIP Code (The Real Yellow Pages 2020). We obtained the keyword data for this study using commercially available scraping software. (ScrapeBox 2020) We recorded all instances of the term "dixie" in nationwide listings at yellowpages.com. These listings include business, educational, government, and non-profit organizations in over 16,000 communities. Before the initial analysis, we removed duplicate entries as well as firms whose names were derived from individuals named "Dixie" and Winn-Dixie grocery stores, a regional chain based in Jacksonville, Florida, with nearly 500 locations in Florida, Alabama, Louisiana, Georgia, and Mississippi (Winn-Dixie n.d.). Likewise, entries from Dixie County, Florida, were excluded if they referred directly to the county name. After this

21

process, we assigned all entries to polygons based on 2019 ZIP Code data. This data is
summarized in Table 1.

Table 1. Descriptive Statistics of ZIP Codes Containing Dixie

Populations	Observations	Percentage of Total ZIP Codes	Mean	SD	Range	Description
All Dixie	3718	12.2	1.2	2.42	1 to 78	Total Population of Dixie Establishment ZIP Codes
Dixie Highway Outside of Buffer	1880	6.2	1.5	2.31	1 to 78	Population of Dixie Establishments with Dixie Highway proximate locations removed

After initially plotting the data, we noted that many of the anomalous pockets of "Dixie"
were located near extant portions of the Dixie Highway. So, to establish proximity to the Dixie
Highway, we created a shapefile that charted the approximate route of the highway system as it
existed circa 1920. Unfortunately, much of the roadway no longer exists under the original
roadway markers. In these areas, where the road no longer exists, we used the pathway of the
road that replaced the original. These replacement roadways, at times, possibly do not follow the
precise route of the old road, but they are close enough to create a buffer at the scale needed for
this study. Finally, a 0.5-kilometer buffer was created around the highway to generate two study
populations: one population that included all instances of the "Dixie" keyword and one that
excluded the instances within the highway buffer.

Nonetheless, portions of the Dixie Highway still exist under the original name, most
prominently in southwestern Ohio, northern Kentucky, and South Florida, three of the areas, not

coincidentally, most affected by our analysis. In all 334 locations, approximately ten percent of the locations in the dataset and 56 percent located within the 0.5-km buffer have addresses on portions of the old Dixie Highway that retain the name. Finally, approximately 60 establishments, two percent of the total, are located along various other "Dixie" named roadways. We left these establishments in the dataset.

Finally, we adopt a similar methodology to previous studies of this type by creating hotspot maps to bound the region. We have created two maps, one that contains all instances of "dixie" and another that excludes instances of the term proximate to the highway. The exclusion of the establishments near the path of the original Dixie Highway system demonstrates that excluding these instances significantly shrinks the region as defined by previous studies while also shifting the core of the region to the south and west (Cooper and Knots 2010; Alderman & Beavers 1999; Reed 1976; Ambinakudige 2009; Reed et al. 1990).

However, unlike the original Reed study and most follow-up studies, this study does not employ a ratio approach. The ratio approach uses terms such as "American" or "national" divided by Dixie (or vice versa) to give a rough idea of the per capita usage of the term. Our approach uses ZIP Code polygons to calculate nearest neighbors and hot and cold spots through those associations. As ZIP Codes very roughly correlate with population, a weak per capita effect is present in the data mapping. Our study is concerned more with documenting the effects of a single feature on the naming landscape, an effect large enough that we do not believe creating more precise per capita data changes the study results.

To calculate the spatial hotspots, we use the Getis-Ord Gi* statistic in ArcGIS Pro 2.5 on "dixie" establishments data sorted by ZIP Code polygons. (ESRI 2020; Getis and Ord 1992). We employ a fixed distance band of 100 km after tests from spatial autocorrelation, indicating that

this distance had the highest levels of spatial autocorrelation at the scale of the study area. We calculated the Getis-Ord Gi* statistic twice – once on the entire dataset and once on the data with instances of Dixie removed within the 0.5-kilometer buffer around the route of the Dixie Highway.
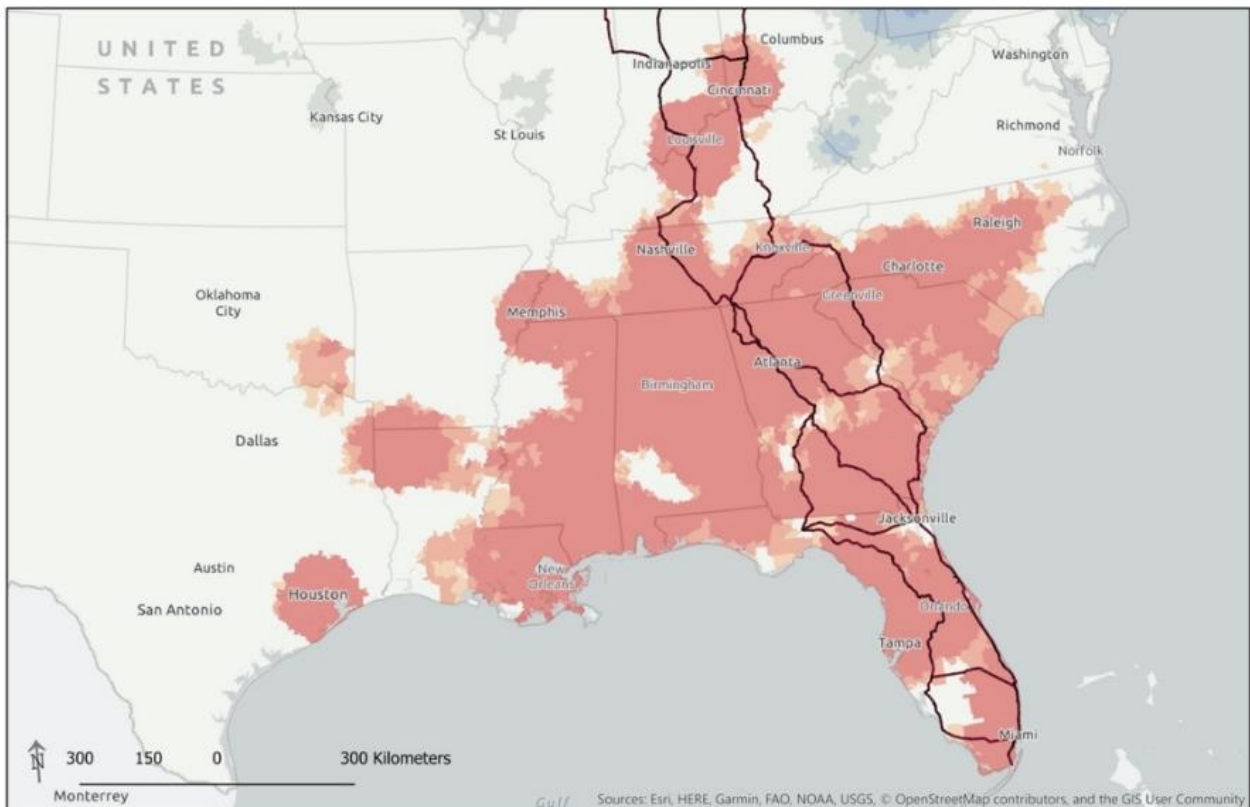
Results

Figure 2 Southeastern Dixie Hotspots

Our findings indicate that 39 percent of instances of "dixie" in the ten states that the system crossed lie within a 0.5 km buffer around the historic route of the highway. Thus, cursorily, it appears that the roadway substantially influences "Dixie" naming patterns outside the Deep South.

Figure 2 displays the initial results with all keyword locations included. Covering roughly 1.03 million km², the Dixie hotspot extends throughout much of the Southeast south of Virginia and Kentucky with a corridor extending north running along the middle Ohio Valley, around Louisville, Kentucky, and an additional corridor from approximately Lima to the Cincinnati metro area. These concentrations of "Dixie" in the Ohio Valley have been noted in prior work and grouped within the greater South. Apart from the continuous hotspot that covers much of the Southeast, smaller pockets surround the city of Houston and southeastern Oklahoma. Also, a disconnected hotspot surrounds the city of St. George in southwestern Utah, an anomaly we will consider further in the discussion section.
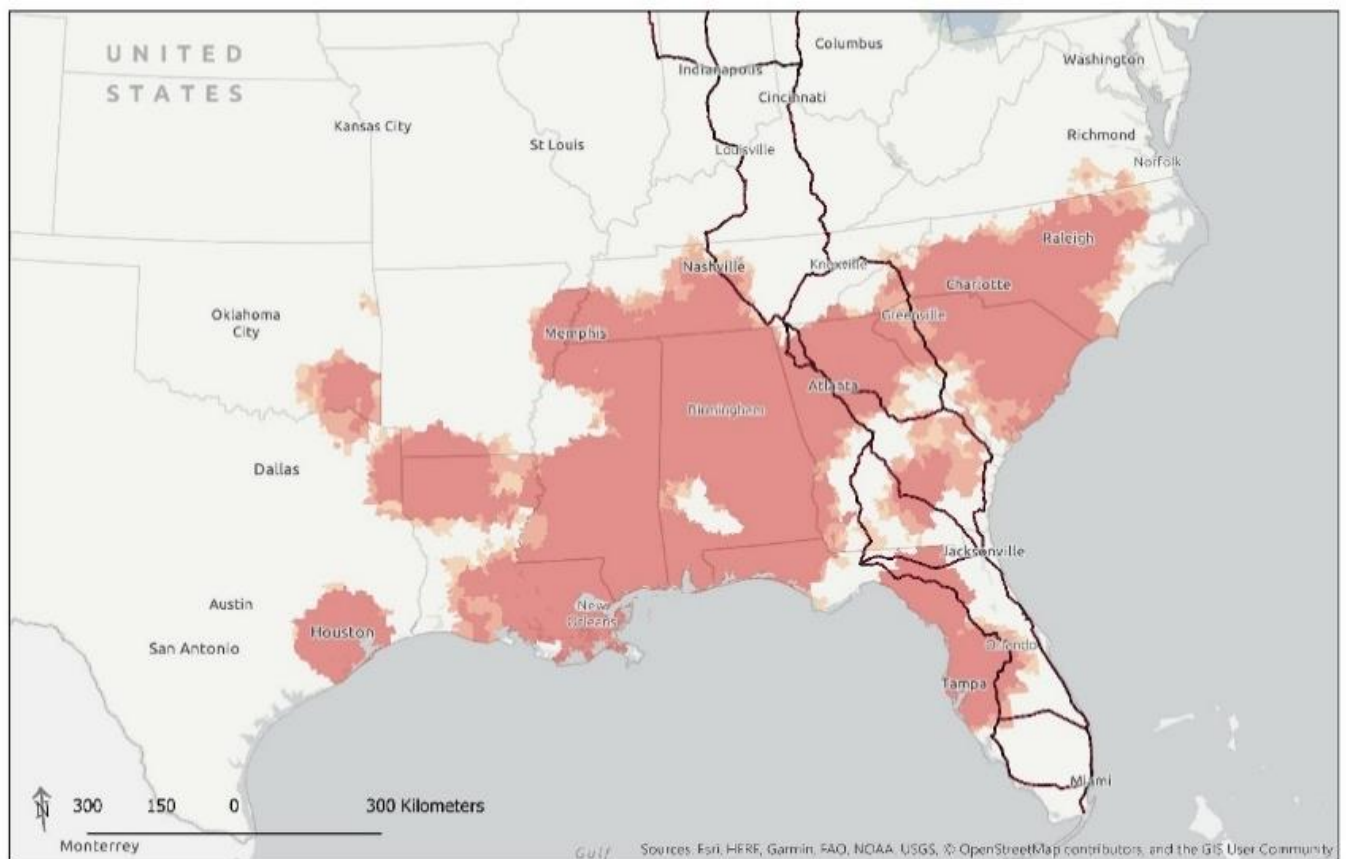

Figure 3 Southeast with Dixie Highway Proximate Establishments Removed

Removing Dixie locations within the 0.5-kilometer buffer around the historic Dixie

Highway route reduces the area of Dixie hotspots to approximately 857,000 km^2, reducing the

area of the hotspot approximately 18 percent, with the most acute reductions occurring in

peripheral areas of the region. As seen in Figure 3, southwestern Ohio/northern Kentucky and the

Louisville area disappear as well as Dixie hotspots in East Tennessee and the entire east coast of

Florida. In fact, in South Florida, Jacksonville, the Louisville area, southwestern Ohio, and East

Tennessee, virtually all instances of Dixie are within the 0.5-kilometer buffer along the highway

and therefore were removed from the second dataset. The results are more ambiguous in Georgia

but still surprisingly large. The buffer model removed approximately 40 percent of

establishments in north Georgia, while 55 percent of Dixie establishments in southern Georgia

were removed. The sheer number of establishments outside the buffer with the Dixie name

leaves much of the area within the Dixie hotspot, even after removing Dixie Highway-proximate

establishments.

Discussion

The hotspot maps contain several incongruities. The most conspicuous is the hotspot

located in the corner of southwestern Utah based around the city of St. George. Notably, St.

George contains the highest instances of Dixie usage in the dataset – the top two ZIP Codes with

the highest occurrences of the "Dixie" are in this region (see Table 1). These results, although

potentially surprising, are in line with Ambinakudige's study conducted a decade ago. This

region of Utah has been referred to as Dixie since it was settled by the Church of Latter-Day

Saints directive in the 1860s. The settlement was founded as a cotton-producing outpost in the

Great Basin (Arrington 1956; Cooper & Knotts 2017). Washington County, Utah, the region's center, currently contains a university and medical center with Dixie in the name. Nonetheless,

the concentration of Dixie usage in this small area and its connections to Southern identity warrant further study.

Winn-Dixie, a regional grocery store chain concentrated in Florida, central Alabama, and southeastern Louisiana, has nearly 700 locations, an order of magnitude larger than any other chain and the only one we removed entirely from the dataset (Winn-Dixie n.d.). However, two smaller-scale, regionally concentrated chains likely affected the hotspot analysis. In northeastern North Carolina, there is a chain of 25 convivence stores with Dixie in the name, but no other establishments using the term. Likewise, the Dixie Queen restaurant chain in and surrounding Memphis made up approximately half of the occurrences of "dixie" in this area. Also, a large portion of the intense hot spot centered around Dixie County in north-central Florida is likely due to establishments taking the county's name. Although we removed all listings that used the county name explicitly, many of the listings did not directly reference the county name, and therefore we left them in the dataset.

Conclusion

"Dixie" functions as more than a neutral regional signifier, as John Shelton Reed trenchantly summarizes:

A business or organization may use Southern in its name simply as a descriptive term, but Dixie is less likely to be used that way. The dual nature of Southern and the less ambiguous connotations of Dixie mean that while Southern can be substituted for Dixie, in general, the converse is not always true. For example, try substituting Dixie for Southern in the (genuine) organization names in the preceding paragraph. Dixie, in the name of an organization, links it to a symbol of the region's historic culture, a linkage that is often irrelevant and sometimes downright inappropriate. Dixie is, as one journalist observed, "a meaner word" than Southern (Reed 1976 p. 933)

Quantitative methods alone cannot discern the intent of individuals who name businesses. While using GIS buffers around the highway's historic routes demonstrates a robust connection between proximity to the historic Dixie Highway system and the name Dixie – one stronger than in the remainder of those states and regions. Further study of historical naming patterns in the region is needed as very little has been written about naming patterns before Reed's 1976 study, making it challenging to examine change during the highway's construction.

As we have reviewed, bounding vernacular regions through telephone directories is well established. Nevertheless, this method relies on several, often unstated, assumptions. The most crucial assumption is that public establishments' names function not merely as locational signifiers or marketing devices but also incorporate (at times inchoate) ideas about the region. Thus, scholars must employ an implicit model of the vernacular region. For instance, a business named Yankee Drugs located on Yankee Street and Yankee Drugs on Main Street are not of equivalent value in deciphering a vernacular region. This point may seem obvious, but it is an

important one, confirmed implicitly by the existence of exclusion criteria – such as chain stores – in these models.

The underlying results of mapping vernacular regions can be problematic when these assumptions are unmet. This fact informs some of the prior study models reviewed, particularly on which types of listings to exclude. For instance, the decision to include or exclude large chain stores from the analysis requires an (often unarticulated) model of the relationship between vernacular regions and public names, specifically, that the spatial origin of the name is what carries value in deciphering vernacular regions. Furthermore, finding meaning in the frequency of names with a term as culturally loaded as Dixie is difficult. Naming conventions are fluid and represent the collective results of countless individual decisions that must balance fundamental economic concerns, comport with personal values and ideologies while attempting to form linkages with the local community. The local naming landscape also influences them. Successful businesses with a particular name are likely to spawn similarly named businesses in the area even if, at first, a term only functions as a locational signifier. As a result, discerning the intent of any individual business name outside of direct interviews is virtually impossible.

Currently, scholarship in this area lacks in several crucial areas. First, we know little about how individuals choose public names. Moreover, we need underlying theoretical work on the dynamics of regions and public naming –a task that should be informed by quantitative work, such as this study. Finally, we lack an understanding of historical trends in business naming before the 1970s, making it difficult to understand the processes behind naming – specifically, the ability to identify changes in naming conventions and the relative rate of change in naming patterns on the landscape. Thus, presently, we need more qualitative work to solidify our understanding of how the building blocks of vernacular regions – names – are adopted and

proliferate on the landscape. With this further understanding, we can elevate the study of vernacular regions from a series of cartographic curiosities to an essential tool in the understanding of region.

References Cited

Alderman, D. H., and Beavers, R. M. 1999. Heart of Dixie Revisited: An Update on the

Geography of Naming in the American South. *Southeastern Geographer 39* (2): 190–205.

https://doi.org/10.1353/sgo.1999.0017.

Ambinakudige, S. 2009. Revisiting "The South" and "Dixie" Delineating Vernacular Regions

Using GIS. *Southeastern Geographer 49* (3): 240–50. https://doi.org/10.1353/sgo.0.0051.

ArcGIS Pro (Version 2.5). 2020. [Computer software]. ESRI.

Arrington, L. J. 1956. The Mormon Cotton Mission in Southern Utah. *Pacific Historical Review

25* (3): 221–38. https://doi.org/10.2307/3637013.

Bletzer, K. V. 2003. Latino Naming Practices of Small-Town Businesses in Rural Southern

Florida. *Ethnology 42* (3): 209–35. https://doi.org//10.2307/3773801.

Burch, A. D. S. 2020, January 20. 'We've Got to Change This': Has Dixie Highway Reached the

End of the Road? *The New York Times*. https://www.nytimes.com/2020/01/20/us/dixie-

highway-change-florida.html

Cooper, C. A., and Knotts, H. G. 2017. *The Resilience of Southern Identity: Why the South Still Matters in the Minds of Its People*. Chapel Hill: The University of North Carolina Press.

Cooper, C. A., and Knotts, H. G. 2010. Declining Dixie: Regional Identification in the Modern American South. *Social Forces 88* (3): 1083–1101.

Getis, A., and Ord, J. K. 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis 24* (3): 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Gunderman, H. C., and Harty, J. P. 2017. "The Music Never Stopped": Naming Businesses as a Method for Remembering the Grateful Dead. *Journal of Cultural Geography 34* (3): 373–95. https://doi.org/10.1080/08873631.2016.1264073.

Ingram, T. 2016. *Dixie Highway: Road Building and the Making of the Modern South, 1900-1930*. Chapel Hill: University of North Carolina Press.

Jackson, B. *Whatever Happened to Georgia's Dixie Highway.pdf*. (n.d.). Retrieved from http://georgiahistory.com/wp-content/uploads/2013/07/dixie_hwy.pdf

Jordan, T. G. (1978). Perceptual Regions in Texas. *Geographical Review 68* (3): 293–307. https://doi.org//10.2307/215048.

Lamme III, A. J., and Oldakowski, R. K. (1982). Vernacular Areas in Florida. *Southeastern Geographer 22* (2): 99–109. https://doi.org//10.1353/sgo.1982.0012.

Lewis, G. M. 1966. William Gilpin and the Concept of the Great Plains Region. *Annals of the Association of American Geographers 56* (1): 33–51. https://doi.org//10.1111/j.1467-8306.1966.tb00542.x.

Liesch, M., Dunklee, L. M., Legg, R. J., Feig, A. D., and Krause, A. J. 2015. Use of Business-Naming Practices to Delineate Vernacular Regions: A Michigan Example. *Journal of Geography 114* (5): 188–96. https://doi.org/10.1080/00221341.2014.965187.

Lowry, J., Patterson, M., and Forbes, W. 2008. The Perceptual Northwest. *Yearbook of the Association of Pacific Coast Geographers* 112–26. http://doi.org/10.1353/pcg.0.0009.

Martinez, S. 2020, February 5. Dixie Highway Name Condemned by Miami-Dade Commissioners. 'Get over it? ... Hell no.' *Miami Herald*. https://www.miamiherald.com/news/local/community/miami-dade/article239952068.html

McEwen, J. W. 2014. Louisiana: Apprehending a Complex Web of Vernacular Regional Geography. *Southeastern Geographer 54* (1): 55–71. https://doi.org//10.1353/sgo.2014.0001.

Musa, S., Šiljković, Ž., & Čelik, A. (2015). Vernacular regions as a determinant for development of creative tourism in Bosnia and Herzegovina. *ITHMC International Tourism and Hospitality Management Conference Sarajevo, Bosnia and Herzegovina 30 September-4 October 2015 BOOK OF PROCEEDINGS*, 601.

The Real Yellow Pages. 2020. Retrieved from https://www.yellowpages.com on June 10, 2020.

Reed, J. S. 1976. The Heart of Dixie: An Essay in Folk Geography. *Social Forces 54* (4): 925–39. https://doi.org/10.1093/sf/54.4.925.

Reed, J. S., Kohls, J., and Hanchette, C. 1990. The Dissolution of Dixie and the Changing Shape of the South. *Social Forces 69* (1): 221–34. https://doi.org/10.1093/sf/69.1.221.

ScrapeBox (Version 2.0.0.120). 2020. [Computer software]. ScrapeBox. scrapebox.com

Shortridge, J. R. 1980. Vernacular Regions in Kansas. *American Studies 21* (1): 73–94.

Shutting, R. J. 1919. *Outline of the Dixie Highway* [Map]. Dixie_Highway_Map.gif. https://georgiainfo.galileo.usg.edu/topics/maps/article/dixie-highway/dixie-highway-map-1919

Vukosav, B., & Fuerst-Bjeliš, B. (2016). Labels of Interest Groups as Indicators of a Vernacular

    Region: A Case Study in Croatia. *Tijdschrift Voor Economische En Sociale Geografie*,

    107(4), 454–467. https://doi.org/10.1111/tesg.12168

Winn-Dixie. (n.d.). Retrieved from https://www.winndixie.com/?a37a4 on June 10, 2020.

Weaver, R., & Holtkamp, C. (2016). Determinants of Appalachian Identity: Using Vernacular

    Traces to Study Cultural Geographies of an American Region. *Annals of the American*

    Association of Geographers, *106*(1), 203–221.

    https://doi.org/10.1080/00045608.2015.1090266

Zelinsky, W. 1980. North America's vernacular regions. *Annals of the Association of American*

    Geographers, *70* (1): 1–16. https://doi.org/10.1111/j.1467-8306.1980.tb01293.x.

Biography

Jesse R. Andrews (jesse.andrews@okstate.edu) is a Ph.D. candidate at Oklahoma State

University in Stillwater, Oklahoma. His research interests include historical, business, and urban

geography, along with the study of vernacular regions, particularly the American South.

CHAPTER III

RECONSTRUCTING A VERNACULAR SOUTH: 1910 TO 1930

Highlights

- Recorded frequency of the terms "Dixie" "Southern" "American" and "National" in historic city guides from 1910 and 1930.

- Created ratio-based maps on frequency of terms "Dixie" "Southern" to "American" and "National".

- Maps of "Southern" appear similar to previous maps of the region

- "Dixie" is highly divergent from modern maps

- 1910 concentrated in North Carolina Piedmont.

- By 1930 grows dramatically extends well into Midwest, likely a result of the Dixie Highway.

Reconstructing a Vernacular South: 1910 to 1930

Abstract: This study extends sociologist John Shelton Reed and his followers' work delimiting a vernacular South into the early 20th Century. Using historic city guides, I record and model the prevalence of "southern" and "Dixie" at the beginning of business names in over 300 cities between 1910 and 1930. Using these data along with the frequency of the terms "national and "American," I map the ratio of the "southern" and "Dixie" relative to "national" and "American" As in prior scholarship, mapping the term "southern" produces a cultural map of the Southern United States similar to how scholars using various methods have demarcated the region. However, interestingly, the prevalence of "Dixie" does not follow this pattern. The relative frequency of that term expanded dramatically between 1910 and 1930, while "southern" shrank proportionately during the same period. This observation is at odds with what has been detected in prior studies beginning in the mid-1970s. Notably, Dixie's expansion between 1910 and 1930 was the strongest along the route of the Dixie Highway, casting doubt on the idea that the term is purely a proxy for regional sentiment. Ultimately this study problematizes some of the assumptions behind vernacular regions. Likewise, it firmly establishes the technique as grounded in a particular set of structural norms that do not fully emerge until several decades into the 20th Century.

Reconstructing a Vernacular South: 1910 to 1930

Introduction

This study uses historical business name data to reconstruct the United States South as a vernacular region in the early 20th Century. The method used is similar to those pioneered by John Shelton Reed in "The Heart of Dixie: An Essay in Folk Geography" (1976) and subsequently adopted by several other scholars. Subsequently, Reed and others followed up that initial article with a series of updates to the original tracking changes to the vernacular region through the next three decades. Likewise, a series of articles inspired by Reed but using more advanced geospatial techniques have appeared in the past two decades. However, scholars have not applied these methods to reconstruct the boundaries of the vernacular South in the past. This study is an initial step in filling that gap.

Reconstructing vernacular regions using data from the past is not just an exercise in historical curiosity. Using these methods on historical data allows us to stretch models to their limits and, in the process, understand precisely what these techniques are measuring and occasionally gain insight into areas only tangentially related to the original purpose of the models. As we shall see, unstated assumptions underlying these models seem to break down for some of the data in 1910. Nonetheless, these failures are themselves opportunities to increase understanding of the underlying dynamics at play.

This study produces a series of maps that comport with existing scholarship on the region. However, these have scholarly value in that they also produce maps that are questionable reflections of the region. Nonetheless, as we will see, this data is equally valuable. The following

study explores these "bad" models as much as the "good" ones, as each model can reveal important information about the underlying data.

Literature Review

Perceptual or vernacular regions are those perceived to exist by their inhabitants and other members of the population at large. They exist as part of the popular or folk culture. Rather than being an intellectual creation of professional geographer, the vernacular region is the product of the spatial perception of average people. Rather than being based on carefully choose quantifiable criteria, such regions are the composites of the mental maps of the population.

  -Terry G. Jordan (1978)

The formal study of vernacular or perceptual regions entered academic geography and other related disciplines in the late 1960s. The first significant attempt to map these regions on a national level was Ruth Hale's (1971) underappreciated unpublished doctoral dissertation at the University of Minnesota, "A Map of Vernacular Regions in America." Hale used a survey-based approach to define vernacular regions. She mailed thousands of surveys across the country, asking which regions the participants felt they lived. With these data, she constructed a regional map of the United States. Her work remains one of the most ambitious attempts at defining vernacular regions using this method and directly inspired several scholars, including Wilbur Zelinsky, to explore the methods further.

Reed's "Heart of Dixie" sparked considerable scholarly interest in vernacular regions. Reed's piece and Terry Jordan's work on perceptual regions in Texas (1978), and Wilbur

Zelinsky's (1980) mapping of the United States vernacular regions using Reed's methods brought the technique prominence. Following these pieces, a flood of publications attempting to define vernacular regions at a smaller state or regional scale followed (Shortridge 1980; Good 1981; Raitz & Ulack 1981; Lamme III & Oldakowski 1982; Shortridge 1985; Zdorowski & Carney 1985).

However, by the end of the 1980s, work on vernacular regions virtually disappeared, with little being written on the subject into the late 1990s. One exception is Craig Colton's 1997 "The Land of Lincoln: Genesis of a Vernacular Region," one of the most imaginative pieces written using the concept. However, before GIS revolutionized methods, Colten (1997) used a mix of quantitative methods and fieldwork to identify a "Land of Lincoln" region stretching from his birthplace in central Kentucky to his adult home at the time in Illinois. This piece is an excellent example of using mixed methods to examine a little-known vernacular region. In addition, it is one of the few pieces that attempt to study a region not based on historical settlement patterns.

As GIS technology became more accessible to non-specialist scholars and digital databases of businesses populated academic libraries, scholars began revisiting vernacular regions. Likewise, in the past two decades, the meaning of "perceptual" and "vernacular" seem to have diverged in the scholarship. "Perceptual" has become a term more likely to be used for survey-based studies (Lowry et al. 2008), while "vernacular" is primarily used in studies that use quantified features of the landscape to study region (Liesch et al. 2015). However, few of these works attempted to innovate in methods or theory despite new statistical tools. Instead, many of them were updated past studies from the 1980s (Lamme III & Oldakowski 2007; Ambinakudige 2009).

After this revival of scholarship on vernacular regions, work on the subject, save a few examples, has been nonexistent (Bricker 2018; Vukosav & Fuerst-Bjeliš 2016). Except for my own, no articles that I am aware of have been published on delimiting North American vernacular geographies in the past five years. This absence of scholarly work is unfortunate, as there is still much fruitful work to be done on the subject.

Past Boundaries
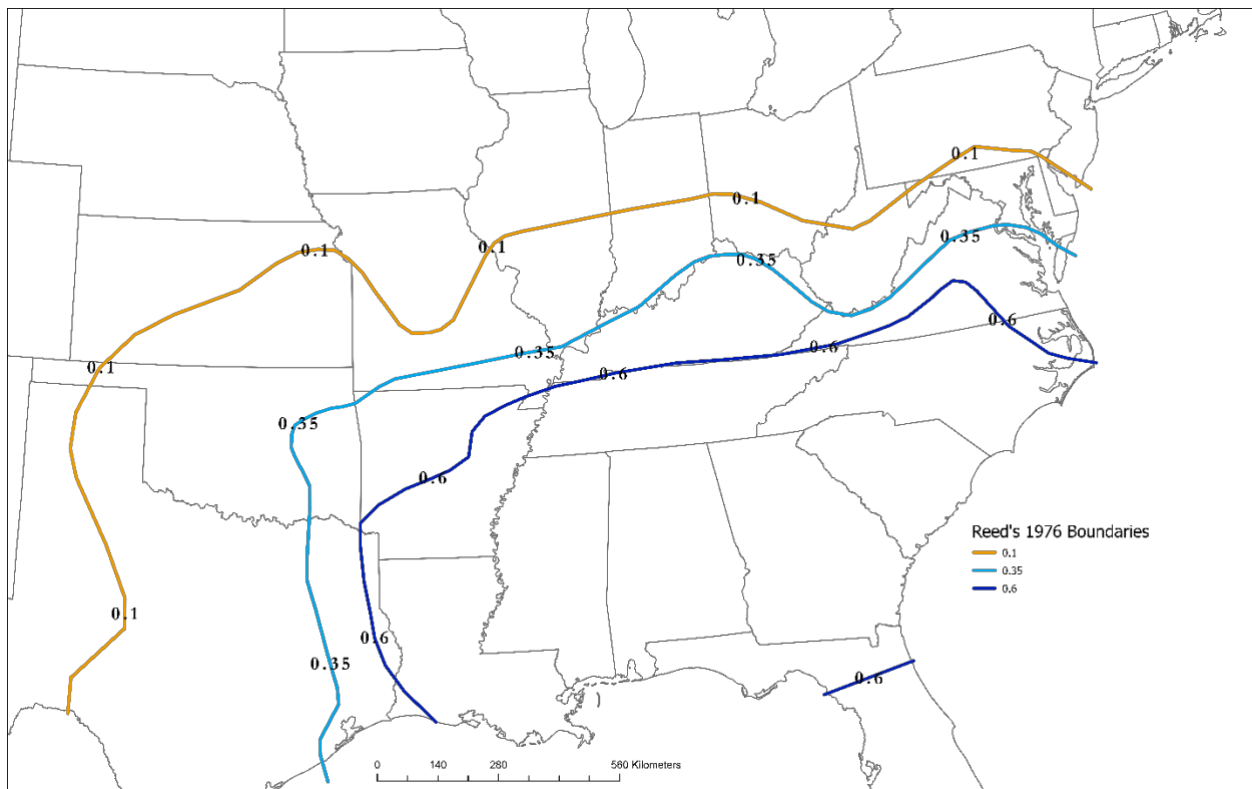
Figure 1. John Shelton Reed's Southern Ratio in 1976. Adapted from Reed (1976)

After Reed's initial study (Figure 1), two studies, one in 1990 (Reed et al.) and another in 1999 (Alderman & Beavers), attempted to recreate his methods precisely to examine change in regional identification over time. These attempts to detect change over time led to several conventions in visualizing this type of data, an issue we will discuss in-depth later in this study. Two other more recent studies use modern GIS techniques while keeping the keyword approach

to defining a vernacular region (Ambinakudige 2009; Cooper & Knotts 2010). Overall, these studies find a shrinking in the boundaries of the South using the "Southern" keyword and a precipitous drop in Dixie usage from the time of Reed's first study until 2010. The Southern boundaries these studies have drawn are similar, with some shrinking around the periphery. The Dixie boundaries are more interesting with some more variation, but all share easily identifiable patterns and a few potential anomalies.

I (Andrews & Finchum 2020) co-authored an earlier study that uses business naming patterns, which is important in understanding some anomalies within the data (Andrews & Finchum 2020). This study does not attempt to create boundaries for the region like the abovementioned studies. Instead, it examines the distribution of Dixie within .5 kilometers of the Dixie Highway and constructs hotspots with the proximate Dixie locations included and excluded. Excluding locations proximate to the highway removes the anomalous areas of Southwestern Ohio, the Louisville, Kentucky area, East Tennessee, and South Florida from the hotspot. Due to the removal of these keywords, these changes indicate that prior studies likely overestimated regional sentiment when modeling Dixie as a proxy for Southern regionalism.

Methods

Data

I collected all keyword data from the archive of city guides on ancestry.com (U.S. City Directories, 1822-1995 n.d.). Data was collected from 310 cities in the contiguous United States from city guides published circa 1910 and 1930 (Figure 2). I included all cities in the South and border states with data near the two target years in the database. In addition, nearly one hundred cities in states outside the South were included to give adequate coverage. Consequently, this is the largest and most geographically expansive sample of cities used for this type of research.

Figure 2. Cities included in the Dataset

Business names that began with "Dixie," "southern," "American," and "national" were recorded. Not all city guides were available for the target years of 1910 and 1930, so I used data

from proximate years for some locations. All data collected is within four years of the target years. I tallied all entries beginning with "Dixie," "Southern," and "American." The Dixie and Southern entries were divided by the number of American entries to create a ratio. This number provides a relative measure of "southernness." Finally, I plotted these numbers by city and created a series of maps using a relatively new and powerful form of interpolation – Empirical Bayesian Kriging.

Empirical Bayesian Kriging

Table 1. Formulas Used to Obtain Ratios

$$((National + American) + 1) \div 2 = AN\ Ratio$$

$$Dixie \div AN\ Ratio = DAN\ Ratio$$

$$Southern \div AN\ Ratio = SAN\ Ratio$$

After tallying all business names that began with southern, Dixie, national, and American in the 1910 and 1930 periods, I divided "southern" and "dixie" by the total American and National entries and then added one (Table 1). I added one to ensure all entries would have a non-zero ratio and reduce some of the variances in cities that had few listings.

These ratios were, in turn, used as input data for an Empirical Bayesian Kriging (EBK) model to generate ratio maps. EBK is a form of interpolation that automates setting parameters for a kriging model. EBK has been tested and is considered more accurate with large datasets than traditional kriging methods, especially with data such as this study, where the underlying spatial distribution is unknown (Krivoruchko & Gribov 2019).

The ratio data has a heavy leftward skew due to most locations outside of the South having very few Dixie or Southern-named establishments. Due to this skew, I use an empirical

transformation, a form of standard score transformation. This process converts the raw ratio numbers into a rank to form a normal distribution. This transformation is required to utilize several of the more powerful semivariogram models.

Semivariograms measure how semivariance changes with the distance between observed points. In most instances, the semivariance increases until the points' effects or influence become small enough to be indistinguishable from the dataset's mean semivariance. EBK obviates the need for the user to choose a particular semivariogram for the data. Instead, the technique creates numerous semivariograms through an iterative process. Semivariograms are created for neighboring points to develop subsets of the data, thereby producing many potential models. These models are, in turn, used to form a distribution. Next, localized versions of these distributions are used to calculate each raster's values. Unlike other forms of interpolation, the use of multiple potential models allows EBK to give relative confidence in predicting each raster value. These confidence levels indicate locations where the researcher should interpret the model's results more cautiously (Krivoruchko 2012)

I used a K-Bessel detrended model. K-Bessel is the most flexible and accurate widely used EBK model as well as the most computationally intensive. I used the detrended form of the model because the non-detrended models overestimate places far geographically south but with relatively low ratios, particularly South Florida and South Texas. I used three local model overlap factors with search neighborhoods of 75 and simulated 10,000 semivariograms for each dataset. The neighborhood search parameters were standard circular, with a maximum number of neighbors at 15 and a minimum of 10.

All model parameters were derived by trial and error using cross-validation diagnostics to judge the models' overall fitness using average Continuous Ranked Probability Score (CRPS), a

diagnostic that computes deviation from the predicted distribution function on each observation. The same semivariogram model and parameters were used for all datasets for consistency.

Visualizing the Models

The creation of acceptable EBK models is not the end of the process. The data must be visualized in a way to make it legible. The process of visualization is not a neutral one. Visualizing this type of data invariably involves making underlying interpretations. Likewise, decisions on displaying the data can lead to significant changes in how the data is interpreted. The data outputted by the model is continuous raster data. Thus, decisions made on displaying the data can potentially support very different narratives. The models presented in the results section of this study use the same binning as Reed did in his initial paper, as have subsequent works using the ratio approach. The adoption by subsequent scholars of Reed's visualization parameters makes sense in the context of those studies as they were attempting to make direct comparisons to Reed's work. For comparison purposes, unless otherwise noted, I followed this convention. However, Reed's binning configuration is not the only way to effectively display this data, an issue I will address further in the discussion section.

Results

This study's maps show significant intraregional differences, with a sharp division between the Upland and Deep South in line with this scholarship. The Upland South acts as a transition zone between the .6+ ratios of the former Confederate states and the .1 to .35 ratios in the Ohio Valley, the Ozarks, and a portion of Texas between Houston and Dallas (Figures 3 & 4). That this process forms legible boundaries is unsurprising as the model is specified to create graduated boundaries. However, the boundaries comport to well-documented cultural boundaries of the region, and the results are in line with prior scholarship using similar methods. These boundaries comport with past scholarship, which likely indicates that the business name ratio method detects salient aspects of Southern regional culture. Thus, the evidence suggests that the proportion of business names that use Dixie and Southern compared to national and American is a proxy for Southern regionalism.

The interpolation models produce maps similar to studies that use this approach on recent data. Thus, in this limited sense, the study was a success. If collected at a sufficient scope, historical city guide data can create historically plausible maps of a vernacular South in the early 20th Century. However, the models do more than verify the method's viability on historical city guide data. The study also hints at limitations of the method and large-scale changes in naming practices and the composition of business types in the early 20th Century.

Table 2 Correlation of Global Raster Values

MODEL	DIXIE 1910	SOUTHERN 1910	DIXIE 1930	SOUTHERN 1930
Dixie 1910	1	0.839	0.735	0.770
Southern 1910	0.839	1	0.886	0.955
Dixie 1930	0.735	0.886	1	0.935
Southern 1930	0.770	0.955	0.935	1

However, before we can understand the when, where, and if the region was shrinking, we first need to map the period between 1930 and 1970 to acquire a more definitive understanding of regional dynamics. Nevertheless, we can observe that the Southern ratio models have shown the most minor changes geographically among the four models. Between 1910 and 1930, the Southern ratio shrinks in areas north of the Mason-Dixon line and Ohio River but expands further west into Texas' heart (Figure 5). The 1910 model shows a .955 correlation with the 1930 model, the highest degree of correlation between any of the models (Table 2).

The Dixie models are more interesting to compare. Unlike the Southern ratio models, which change only slightly over the two decades, Dixie undergoes a substantial expansion in the twenty years between 1910 and 1930 (Figure 6). Thus, either mapping Dixie named businesses in 1910 does not do an adequate job of approximating the South, or the valence of the word or how business owners perceived the region changed dramatically in the two decades covered by the study. We can be confident that the former is not true because the 1910 and 1930 models of "Southern" are much more consistent, and the 1910 model roughly approximates most scholars' boundaries of the greater South. Thus, we are likely observing a genuine large-scale expansion of the term's usage as a business name during the two decades between 1910 and 1930.
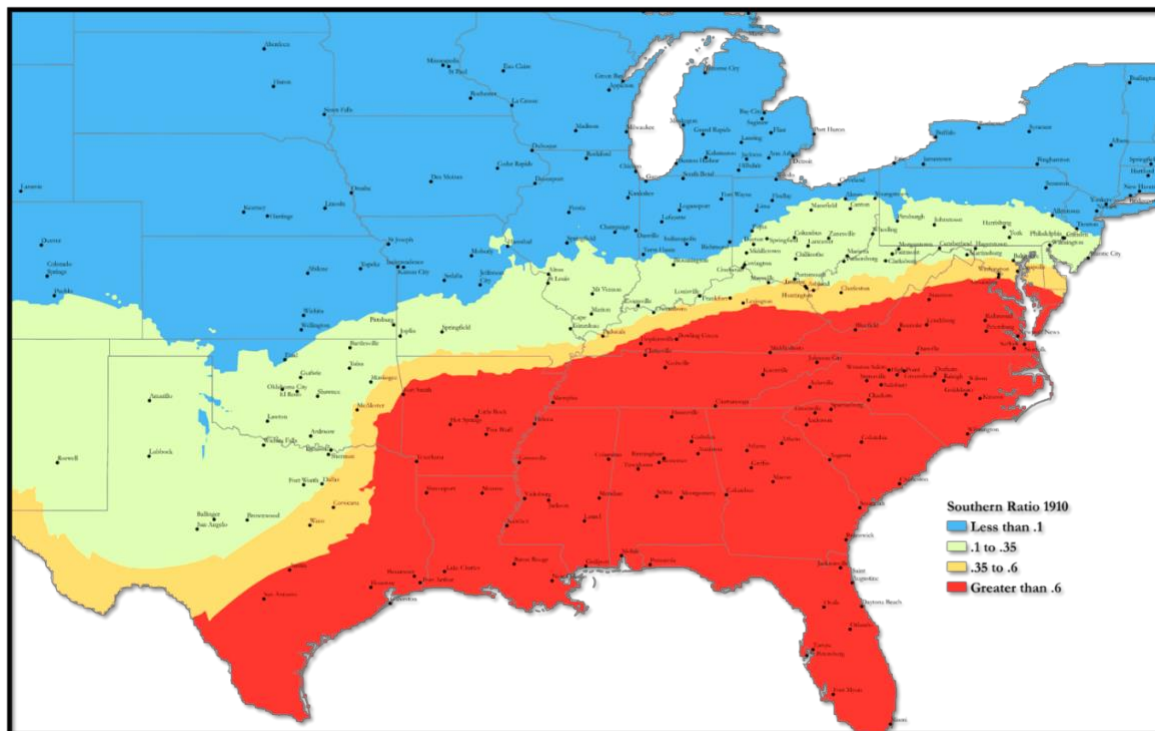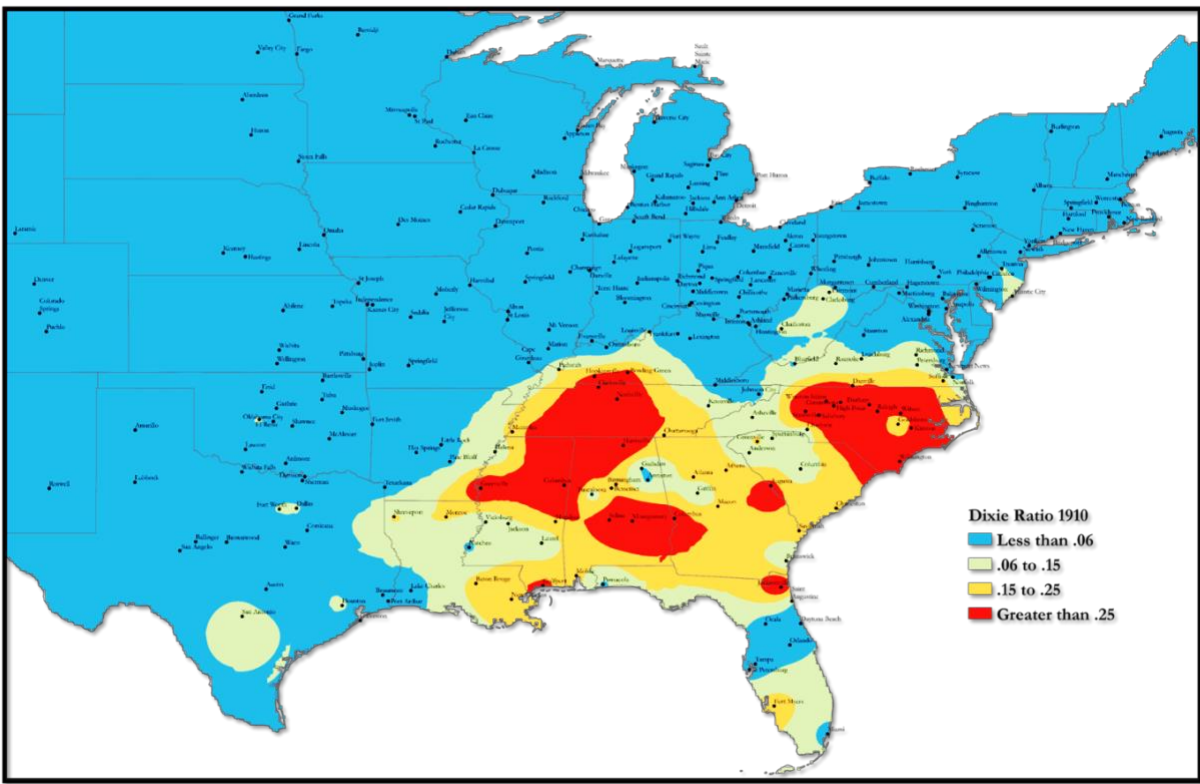

Figure 3. Southern Ratio in 1910

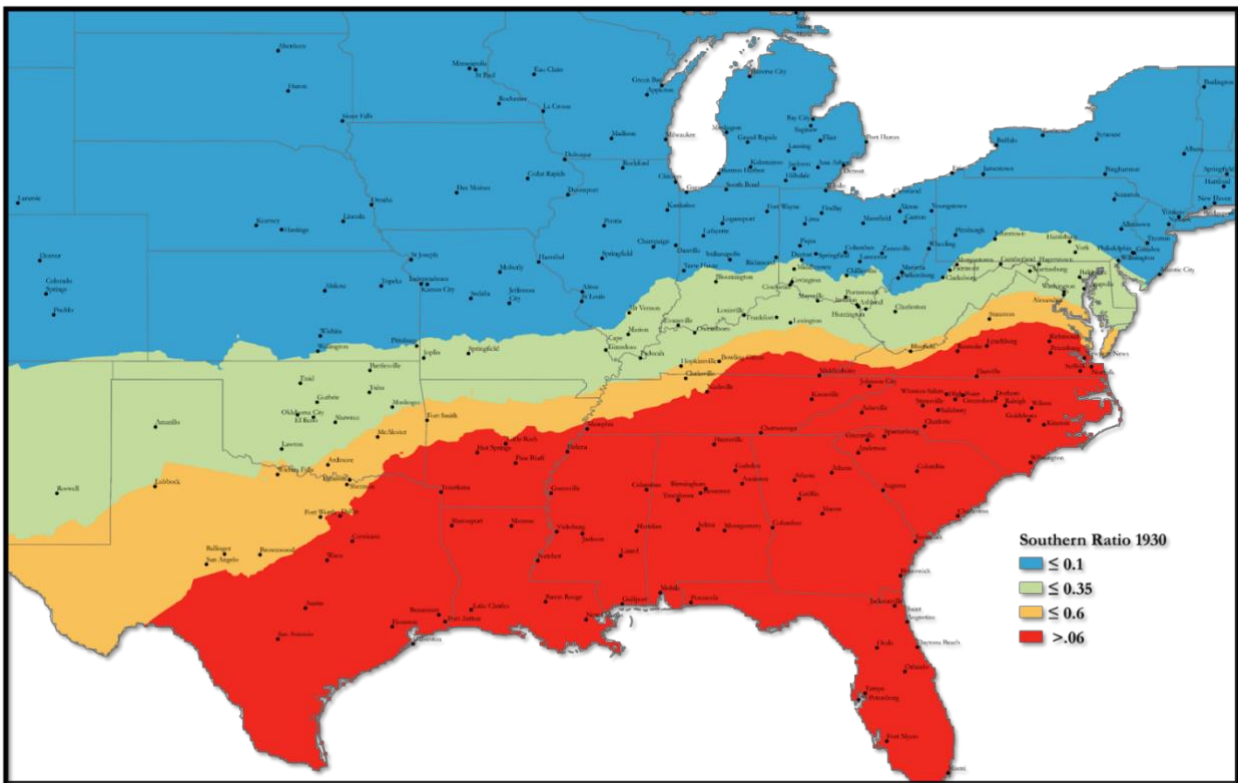**Figure 4. Dixie Ratio 1910**

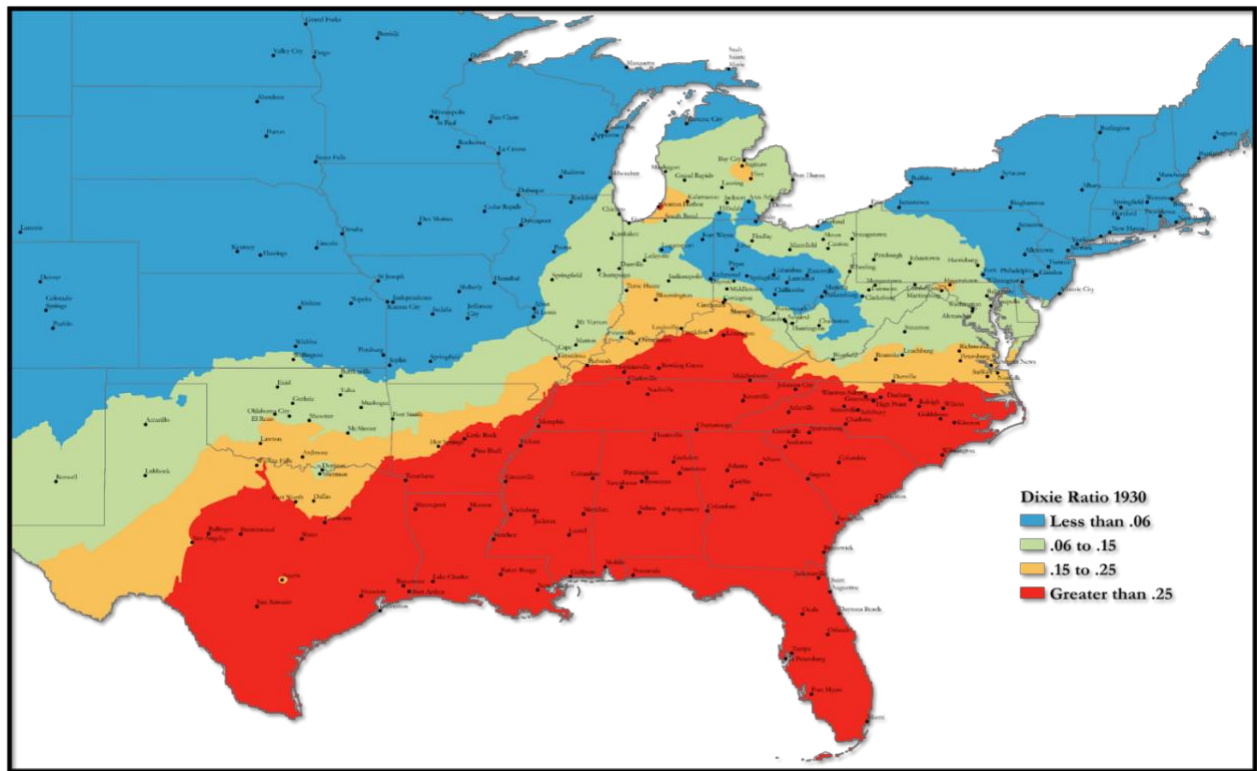
Figure 5. Southern Ratio in 1930

Figure 6. Dixie Ratio in 1930

The observed expansion also likely hints at structural changes that drove business name shifts during the two decades between 1910 and 1930, which we will explore in the next section.

Discussion

Vernacular regions, as measured, are ephemeral expressions of place. Therefore, any model attempting to use them must understand that naming patterns are embedded in more complex social dynamics contingent on place and time. With that understanding, this work is grounded in the following assumptions: naming practices are both representative and the product of relatively unregulated expressions of individual and community idiosyncrasies and the conventions embedded therein. They are also the products of localized power dynamics, social and political conventions, traditions, and market dynamics, all of which are fossilized as a name upon the christening of a new business or organization's name.

It is important to note that business names are typically resistant to change. Thus, names tend to appear and disappear via natural attrition by businesses opening and closing, as renaming a business is rare for various practical reasons (Cooper & Knott 2017). Cooper & Knotts' analysis of Dixie-named businesses confirms this pattern. In that study, the authors find businesses containing the word "Dixie" are significantly older than those that contain "southern." Not because one term was preferred as a regional designator but because the number of new businesses named "Dixie" has shrunk over time compared to "southern" named establishments.

Finally, I have noted some potential confusion concerning the function of the keywords in this study type of study. The misunderstanding can arise from the terms themselves, which could be interpreted as antithetically Southern; this is not the case. The term "national" was chosen by Reed because it was a term frequently used at the beginning of business names, likely guaranteeing a large enough sample in all the cities used in his study. Likewise, I selected "American" as a second control keyword because it was similar to "national" – a popular neutral term lacking regional connotations. Neither term is intended to function as anything other than a convenient control for the study terms. In fact, if these terms do have a more complex spatial relationship with the keywords that are the focus of the study, it could seriously undermine the results. Fortunately, including a second keyword allows us to measure potential spatial differences between the two control terms statistically.

Spatial Autocorrelation

Logically, this project's instrumental keywords "Dixie" and "Southern" display large degrees of global and local spatial autocorrelation as the entire project is predicated on discovering the levels of this phenomenon and mapping it. However, national and American, the control keywords are presumed to lack significant levels of spatial autocorrelation, globally or locally. Prior scholarship has not tested this assumption. Although when Reed introduced this method, he was aware of this limitation, briefly mentioning the potential for regional clusters of the control term to introduce error into the ratio measures. However, Reed, conducting the study in the mid-1970s, lacked the statistical tools to measure spatial autocorrelation among the control terms.

The potential for strong positive spatial autocorrelation among the control keyword is one reason why this study breaks with Reed's and subsequent scholarship use of the single ratio approach, instead using two roughly synonymous terms instead of one. It also permits the control terms' ratios to be tested for spatial clustering – checking if their relative frequencies exhibit statistically significant degrees of spatial variation. Calculating Local Moran's I on the two control terms' ratio to one another shows higher ratios of "national" relative to "American" in a large area surrounding the Great Lakes in the 1930 data (Figure 7). Intriguingly these higher levels of clustering seem to roughly follow the boundaries of Northern Cities Vowel Shift (Inland Northern American English 2022), an ongoing area of linguistic change. However, this could be purely coincidental.

In theory, this regional clustering could artificially lower the southern and Dixie ratios in these areas compared to the rest of the country if "southern" and "Dixie" indicate regional affiliation. Nevertheless, none of the other terms exhibit significant local clustering, and the use of two terms should ameliorate some regional clustering effects, as the study design intended.
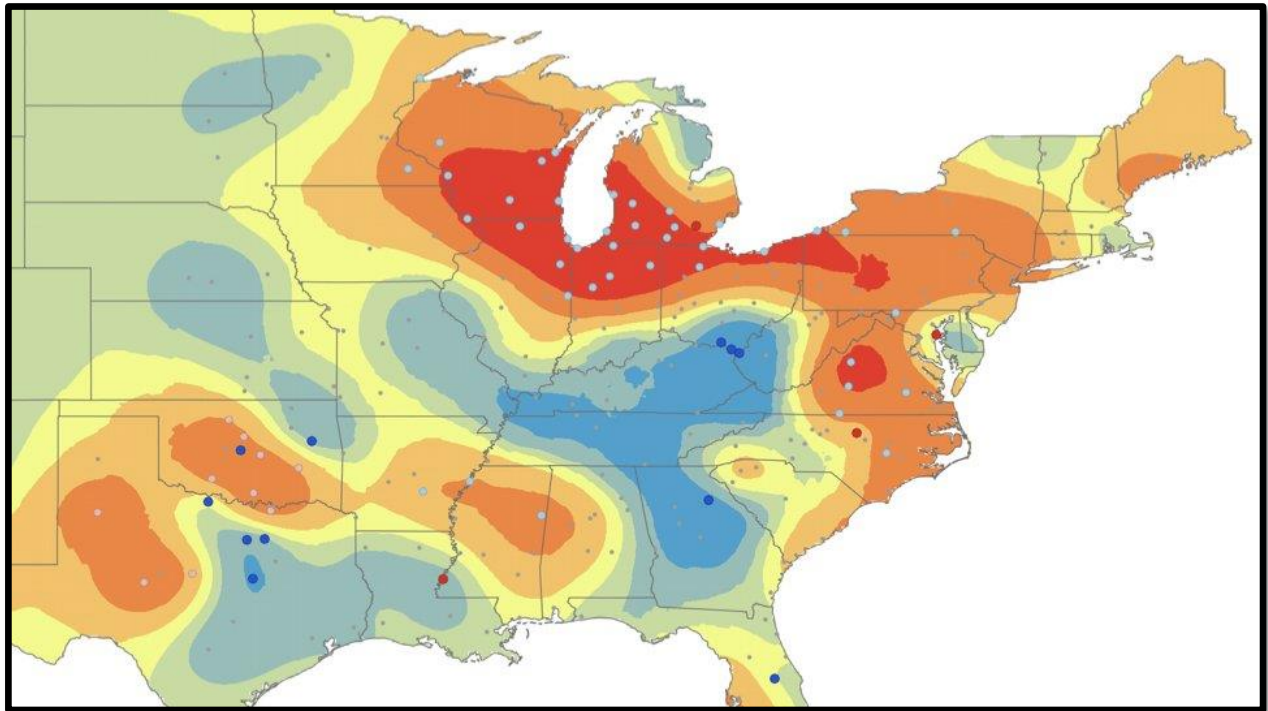
Figure 7. Relative ratio of the terms American (blue) and National (red)

Mapping or Visualizing? Intraregional Variation

As we have discussed, Reed's initial study's maps used a particular binning configuration. Reed chose this way to display the data because it comported with his notion of where the border should lie (Reed 1976). Additionally, the decisions about where to display map boundaries were made after he determined the results of his model. Reed's initial work used a rudimentary computerized form of interpolation that approximated hand-drawn methods, with Reed highlighting rounded numbers that approached his ideas of the region's boundaries. Reed is a preeminent scholar of the South and certainly better qualified than most to make those

designations. Nonetheless, regardless of expertise, making decisions to visualize the data in a way that comports to a predetermined spatial configuration can be problematic.

The primary effect, in this case, is that Reed's decisions on data visualization work to highlight a distinct border running along the edges of the Upland South. Reed's boundaries likely highlight a genuine cultural dividing line. Unfortunately, these borders do a poor job of showing intraregional differences. Highlighting variation within the region was not a salient issue for his study, as Reed's data lacked extreme outliers. However, among the models, the 1910 Dixie data is very different. This model has exceptionally high intraregional variation within the South with a series of extreme Dixie ratios in the Piedmont of North Carolina. We will discuss why the 1910 model is so different from the other models in the discussion section. Nonetheless, these differences emphasize the need to be cognizant of the temporal scale of the measured processes.

Subsequent follow-up work that has used the ratio approach also uses a similar classification system to the one that Reed pioneered. This decision makes sense, as those studies were conducted to measure change over time since Reed's initial study. Similarly, I have followed these rules in mapping out all four models. However, the classifications used by Reed are not the only way to effectively display this data. In fact, the 1910 Dixie data appears ill-served by Reed's configuration as this model contains extremely high ratio values in the North Carolina Piedmont and a few cities in the Deep South, with much lower values everywhere else.

The Dixie 1910 model mapped using Reed's binning configuration displays three areas of

high Dixie values with gaps in between. Thus, visualizing the data in this way show multiple

pockets of high Dixie values. However, as highlighted in Figure 8, this changes substantially

when the data is classified using Jenks Natural Breaks, a popular cluster analysis method.

Natural breaks attempt to minimize variance within classes and maximize the variance outside

classes through an iterative process (Jenks 1967). Applying natural breaks to this dataset in place

of Reed's highlights the high levels of variation within the 1910 Dixie data. Natural breaks

displays roughly the same Dixie boundaries between the first and second break, indicating where
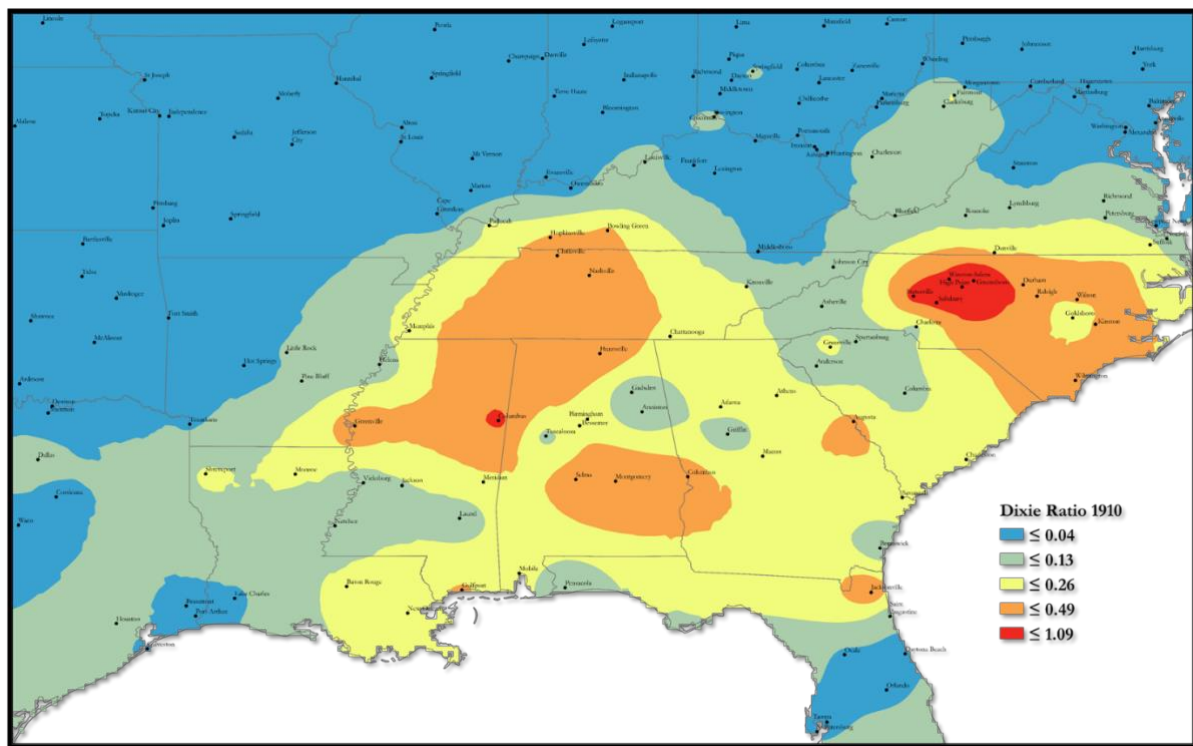

 Figure 8. Dixie Ratio 1910 by Natural Breaks

Dixie becomes noticeable on the landscape. However, using natural breaks on the data

emphasizes how much of an outlier the North Carolina Piedmont is respective to the remainder

of the South. So, in this case, natural breaks appears better at visualizing intraregional

differences in 1910, while the classifications created by Reed still work well enough to bound the region.

A Theory of Name Transition

So, what caused the rapid ascent of Dixie in the two decades between 1910 and 1930? The following is a speculative attempt to sketch out some of the conditions that led to the rapid ascent of Dixie-named businesses between 1910 and 1930. Unfortunately, historical changes in American small business naming practices are unstudied. Hence, dating the move from family names to more abstract branded names is currently impossible. However, anecdotally, I noticed a drop in family-named businesses by 1930 compared to the early part of the century.

In theory, more abstract names such as Dixie could have become more common in the face of increasing competition in the retail sector and the rise of advertising and marketing. At the same time, Southern, a directional and potentially ambiguous geographic signifier, seems to have remained proportionately the same. In an era of insipid corporate descriptives such as Standard Oil or United States Steel, names such as "southern" were not uncommon among larger corporate concerns. In comparison, "Dixie," with a few notable exceptions, was much rarer and only appears to have rapidly grown as a business name between 1910 and 1930.

In this model, increased urbanization led to changes to businesses' makeup, which changed naming patterns. So, by 1930 we see a composition ratio much closer to more contemporary studies using this type of data. To prove or disprove this speculation, we need more data, although tentatively, it appears compositional changes in businesses and naming practices affected the ratio data in these two decades. The critical question is precisely how much?

Dixie as a Metonym

Another issue left unaddressed in prior literature is when and how much Dixie was adopted as a metonym for the South within the region. This oversight is unsurprising considering former studies were conducted after this usage was well established. However, Dixie seems to have not been heavily used as a regional metonym within the South in 1910. This situation contrasts with Southern, a geographically descriptive term used to identify the region since the end of the 17th Century. In addition, Figure 9 illustrates that the period covered by this study had one of the most significant increases in the use of Dixie in print in the United States.
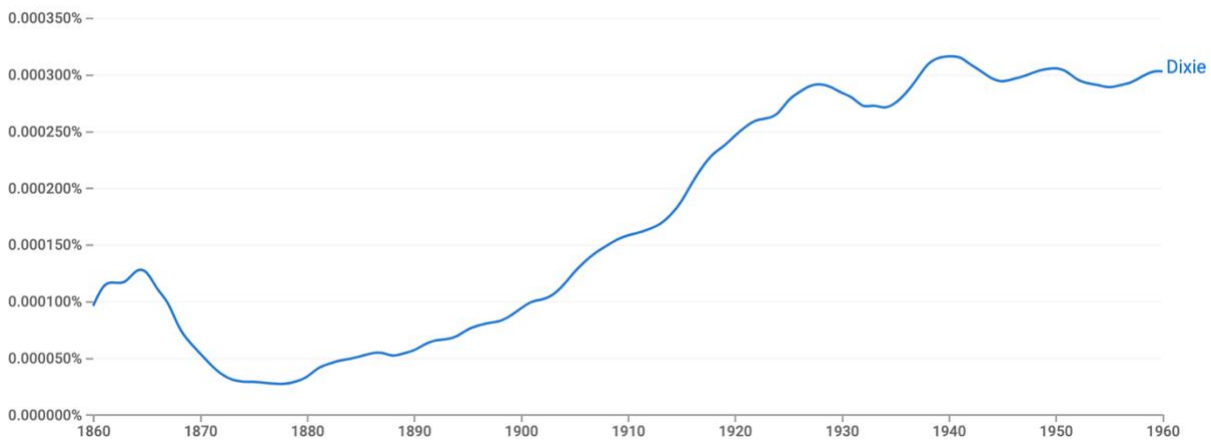

Figure 9. Use of "Dixie" in North America Publications 1860 to 1960

Table 3. OLS Regression of Percentage of Dixie Change between 1910 and 1930

Dixie Change	Coef.	Std. Err.	t	P	95% Conf.	Interval
1930 Population	-1.65^{-8}	9.45^{-8}	-0.17	0.862	-2.03^{-7}	1.70^{-7}
Southern State	.052119	.02168	2.40	0.017	.00940	.0948
Border State	.025756	.026712	0.96	0.336	-.02685	.07836
Population Growth	.009823	.002747	3.58	0.000	.00441	.01523
Percent Business Change	.075024	.025108	2.99	0.003	.02557	.12447
BusinessTotal1930	-.000019	.000190	-0.10	0.923	-.00039	.00035
American to National Ratio	-.00481	.008893	-0.54	0.589	-.02232	.01270
Constant	-.000195	.021098	-0.01	0.993	-.0417	.04136

The data bears out this speculation. OLS regression (Table 3) indicates that controlling for region, total population, and percent change in all businesses enumerated by the study, population growth is the most significant variable in the growth of Dixie names from 1910 to 1930. Urban growth itself seems to have been a strong predictor of the growth in Dixie-named businesses controlling for other likely factors offering some evidence that the change in business practices and makeup brought about by urbanization strongly influenced the growth of Dixie-named businesses.

Outliers and Anomalies: The North Carolina Piedmont and the Dixie Highway

Anecdotally, while collecting the data, I observed that the two areas of the highest Dixie concentrations differed in their business type composition. For example, the cities of the North Carolina Piedmont had by far the highest Dixie ratio in 1910. These firms tended to be retail-oriented and often had a single proprietor listed in the city guide. Conversely, Dixie businesses in the Deep South tended towards larger industrial concerns such as agricultural wholesale operations and resource extraction firms. In these cases, they are probably closer to asserting the "meaner" ideological connotations of Dixie noted by Reed and have remained Dixie hotspots.

During the first decade of the 20th Century, the North Carolina Piedmont was one of the first areas of the South to experience industrialization and rapid population growth in its numerous small cities and towns (Ayers 2020; Carleton 1990). Moreover, small businesses began to transition from personal names connected to the proprietor to more abstract names better suited to attract customers in the increasingly anonymous urban environs of the region. However, this explanation itself has some potential problems. Charlotte, the largest city in the region, had a relatively low Dixie ratio in 1910. Likewise, on the eastern edge of the region, the cities of Durham and Raleigh have elevated ratio numbers but nowhere near the ratios seen in the highest areas directly to the west. The highest area occupies the middle of Piedmont, running from roughly Burlington in the east to Statesville in the west and mainly comprising the modern-day North Carolina Triad, an area that was the one most industrialized in the South in 1910. Ultimately, while it may not be a satisfying answer, why Charlotte or any other individual place in the data set has an anomalous ratio likely comes down to chance. The addition of just a few

Dixie-named establishments can cause significant changes in the ratio data, especially in 1910 when there were fewer businesses overall. So, no individual data point should be overanalyzed. Instead, larger-scale patterns such as we see in the rest of the Piedmont cities are far more robust data in this study, especially with the relatively low numbers of establishments in the 1910 data.

By 1930 the most Dixie amenable sections of the country had likely undergone a similar transition as North Carolina had experienced at the turn of the century. Thus, except for the North Carolina Piedmont, which likely experienced the compositional changes earlier, we see an explosion of Dixie-named business throughout the South and, curiously, parts of the Midwest, a phenomenon I will contend with next.

Besides its rapid growth in the greater South, between 1910 and 1930, Dixie grew exponentially in parts of the Midwest, including Michigan, Indiana, Illinois, and Ohio. Portions of these states, which had virtually no Dixie-named establishments in 1910, have similar Dixie business density as the upper South in 1930. Looking beyond the results of the interpolation model at the raw data, this increase in Dixie businesses is not evenly dispersed in those states. Instead, the growth in Dixie is concentrated in relatively few cities and towns.

Nearly all Midwest cities with significant growth in Dixie establishments have at least one thing in common: they were all near the newly constructed Dixie Highway. Finchum and I (2020) studied Dixie-named businesses close to the old highway route using a current Dixie dataset and established the highway's effects on modern naming practices. This study appears to confirm these effects and extends them back to the beginning of the roadway.

The areas of high Dixie growth indicated by the model between 1910 and 1930 do not

wholly follow the route of the Dixie Highway (Figure 10). Some areas with high modern dixie

ratio scores, such as Lima, Ohio, are low in this model. As with Charlotte, this could be simply

random variation in the data. However, some city guides used in the dataset contain only

businesses within the city limits, while others contain information for suburbs or the entire

county. Unfortunately, these sources are often not clear on their geographic coverage. The more

geographically limited city guides that contained only the city limits or central business districts

missed locations on highways that bypassed downtown areas. Likewise, by design, interpolation

takes point data and disperses it across the landscape. This process is not ideal for modeling

points localized along narrow lines such as the Dixie Highway; nonetheless, the effects are still

evident to such a degree over the length of the highway to make any other cause exceedingly
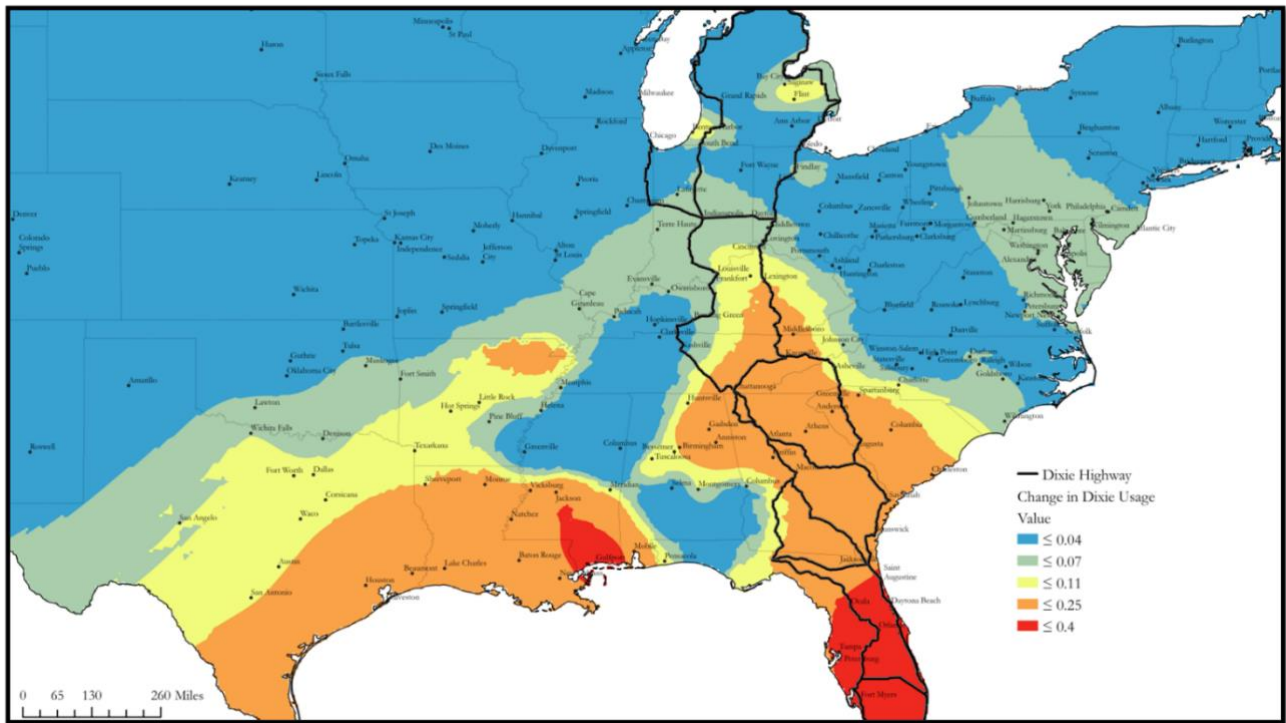
unlikely.

Figure 10. Percent Change in Dixie Between 1910 & 1930

Future Work

Much fruitful work remains to be done on this subject. Extending this work into the middle decades of the 20th Century is the natural next step. Unfortunately, the city guide selection on ancestry.com shrinks dramatically after approximately 1940, making acquiring an adequate amount of city data difficult. This dearth of sources after 1940 is not simply a quirk of the collection methods used for this study. Phonebooks were slowly replacing city guides during this period. So future work that wishes to extend the timeline to connect with Reed's initial work in the 1970s would need to acquire data from phonebooks, far fewer of which are currently digitized.

Also, as I have speculated, the composition of business types changed during the twenty years between 1910 and 1930, leading to business naming practices. Thus the 1910 data, especially the "Dixie" model, deviates widely from all other models created via these methods, detecting the earlier business naming patterns. Future work categorizing business into types could better understand this process, especially concerning the expansion of "dixie." The ability to quantify business types over time gives us additional data that could be used to modify the ratio models based on the period.

Finally, there is important work to be done on the shifting meanings and usage of Dixie from Reconstruction to the present. This study, along with previous work, hints at the term's growth as a metonym for the South in the early 20th Century, eventually acquiring its current association with (white) Southern regional identity. We currently lack scholarship that quantifies the evolving usage of Dixie in the half-century in between. Or, to rephrase the problem as Reed

might: When did Dixie get mean? How mean was Dixie in 1910? And how much meaner was it by the middle of the 20th Century?

Conclusion

Mapping the ratio of business names that begin with "national" and "American" to "southern" in 1910 and 1930 produces boundaries similar to those identified by scholarship beginning in the mid-1970s. Dixie produces very different results. The term was much less widespread in 1910 than expected and much more widely dispersed in 1930. The sea change in the frequency of "Dixie" within the South from 1910 to 1930 is likely explained by changes in the composition of businesses and evolving naming conventions. Dixie grew precipitously from a term with spotty distribution in the South to encompass the entire Southeast and spilling into the Midwest and Mid-Atlantic. Meanwhile, the concomitant expansion of the term into portions of the Midwest is likely due to the Dixie Highway.

Beyond the accuracy of mapping historical vernacular regions, this study also highlights many of the underlying assumptions inherent in past models. The most important being that the data from different years are comparable; for 1930, this seems to be the case. However, the 1910 data is more problematic and not directly comparable possibly due to changes in naming conventions and the underlying composition of business types.

Ultimately, if interpreted on their own terms, both models provide some insights into the functioning of the region in the first three decades of the 20th Century. However, demonstrates that scholars must be cautious when making comparisons across time while modeling historical data. Direct comparisons are only possible when the conditions that created the underlying configuration are sufficiently similar. Of course, subtle changes affecting the underlying dataset

will invariably occur in any study of data over time. However, sometimes large-scale discontinuities in data effectively make direct comparisons impossible among datasets that are temporally not far removed from one another. As this dataset demonstrates by the changing prevalence of "dixie," we cannot assume uniform change over time. As in this dataset, Dixie 1930 is more like 1970 than 1910 despite the shorter distance in time between the former. The challenges and promise of this type of data-driven historical scholarship are uncovering these historical discontinuities or. As evident in this study, either situation exposes a series of interesting questions beyond the initial scope of the study.

References Cited

Alderman, D. H., & Beavers, R. M. (1999). Heart of Dixie Revisited: An Update on the

Geography of Naming in the American South. *Southeastern Geographer*, *39*(2), 190–205.

https://doi.org/10.1353/sgo.1999.0017

Ambinakudige, S. (2009). Revisiting "the south" and "dixie" delineating vernacular regions

using GIS. *Southeastern Geographer*, *49*(3), 240–250.

Andrews, J. R., & Finchum, G. A. (2020). Paring Old Dixie Down: The Dixie Highway and the

Mapping of a Vernacular South. *Southeastern Geographer*, *60*(4), 345–359.

https://doi.org/10.1353/sgo.2020.0028

ArcGIS Pro (Version 2.8). 2021. [Computer software]. ESRI.

Ayers, E. L. (2020). *Southern Journey: The Migrations of the American South, 1790–2020*. LSU

Press.

Bricker, C. (2018). *Vernacular geography and perceptions of place: A new approach to

measuring American regional and political subcultures* [Dissertation, University of Iowa.]

Carlton, D. L. (1990). The revolution from above: The national market and the beginnings of industrialization in North Carolina. *The Journal of American History*, *77*(2), 445–475.

Colten, C. E. (1997). The land of Lincoln: Genesis of a vernacular region. *Journal of Cultural Geography*, *16*(2), 55–75.

Cooper, C. A., & Knotts, H. G. (2010). Declining Dixie: Regional Identification in the Modern American South. *Social Forces*, *88*(3), 1083–1101. https://doi.org/10.1353/sof.0.0284

Cooper, C A. & Knotts H.G. (2017) *The Resilience of Southern Identity: Why the South Still Matters in the Minds of Its People*. Chapel Hill: University of North Carolina Press

Good, J. K. (1981). The vernacular regions of Arkansas. *Journal of Geography*, *80*(5), 179–185.

Google Books Ngram Viewer. (n.d.). Retrieved March 25, 2021, from https://books.google.com/ngrams/graph?content=Dixie&year_start=1860&year_end=1960&corpus=28&smoothing=3&direct_url=t1%3B%2CDixie%3B%2Cc0#t1%3B%2CDixie%3B%2Cc0

Hale, R. F. (1971). *A Map of Vernacular Regions in America*. [Dissertation, University of Minnesota].

Inland Northern American English. (2022). In *Wikipedia*.

https://en.wikipedia.org/w/index.php?title=Inland_Northern_American_English&oldid=1065131 767

Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, *7*, 186–190.

Jordan, T. G. (1978). Perceptual Regions in Texas. *Geographical Review*, *68*(3), 293–307. JSTOR. https://doi.org/10.2307/215048

Krivoruchko, K. (2012). Empirical Bayesian Kriging. *ArcUser*, *6*(10).

Krivoruchko, K., & Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics*, *32*, 100368. https://doi.org/10.1016/j.spasta.2019.100368

Lamme III, A. J., & Oldakowski, R. K. (1982). Vernacular areas in Florida. *Southeastern Geographer*, *22*(2), 99–109.

Lamme III, A. J., & Oldakowski, R. K. (2007). Spinning a new geography of vernacular regional identity: Florida in the twenty-first Century. *Southeastern Geographer*, *47*(2), 320–340.

Liesch, M., Dunklee, L. M., Legg, R. J., Feig, A. D., & Krause, A. J. (2015). use of business-naming practices to delineate vernacular regions: A Michigan example. *Journal of Geography*, *114*(5), 188–196.

Lowry, J., Patterson, M., & Forbes, W. (2008). The Perceptual Northwest. *Yearbook of the Association of Pacific Coast Geographers*, *70*(1), 112–126. https://doi.org/10.1353/pcg.0.0009

Raitz, K., & Ulack, R. (1981). Appalachian vernacular regions. *Journal of Cultural Geography*, *2*(1), 106–119.

Reed, J. S. (1976). The Heart of Dixie: An Essay in Folk Geography. *Social Forces*, *54*(4), 925–939. https://doi.org/10.1093/sf/54.4.925

Reed, J. S., Kohls, J., & Hanchette, C. (1990). The Dissolution of Dixie and the Changing Shape of the South. *Social Forces*, *69*(1), 221–234.

Shortridge, J. R. (1980). Vernacular regions in Kansas. *American Studies*, *21*(1), 73–94.

Shortridge, J. R. (1985). The Vernacular Middle West. *Annals of the Association of American Geographers*, *75*(1), 48–57.

U.S. City Directories, 1822-1995. (n.d.). Ancestry.com. Retrieved March 25, 2021, from https://www.ancestry.com/search/collections/2469/

Vukosav, B., & Fuerst-Bjeliš, B. (2016). Labels of Interest Groups as Indicators of a Vernacular

Region: A Case Study in Croatia. *Tijdschrift Voor Economische En Sociale Geografie*, *107*(4),

454–467. https://doi.org/10.1111/tesg.12168

Zdorowski, R. T., & Carney, G. O. (1985). This land is my land: Oklahoma's changing

vernacular regions. *Journal of Cultural Geography*, *5*(2), 97–106.

Zelinsky, W. (1980). North America's vernacular regions. *Annals of the Association of American

Geographers*, *70*(1), 1–16.

CHAPTER IV

Using Collocational Quotient to Map a Vernacular Great Plains

I intend to submit this article to the *Professional Geographer*. The journal has published several

articles on the Great Plains as a region, and this work fits the journal's focus on shorter,

empirically focused pieces.

Abstract

This study uses collocational quotient (CLQ) to analyze the distribution of keywords used in business names to improve our understanding of the regional geography of the Great Plains. We employ the CLQ metric on four regional terms "Great Plains," "Midwest," "Prairie," and "High Plains," across a study area encompassing the central part of the United States. This novel use of CLQ allows the measurement of bidirectional spatial relationships between individual categorical variables, in this case, the words in the names of businesses. Additionally, this approach enables us to establish the degree keywords are spatially collocated. The study finds that using the technique on the various keyword pairings forms relatively well-defined areas of overlap that roughly correspond to the distribution of the physiographic region or, in the case of "Midwest," with commonly accepted definitions of the fuzzy border between the region and the Great Plains. The use of CLQ potentially offers a more sophisticated understanding of how vernacular regions are bounded and how the terms interact with one another on the margins of regions. By understanding how multiple regional terms interact spatially on the scale of individual businesses, we can better understand the regional boundaries, borderlands, and subtle differences in keyword placement relative to one another.

Keywords: Vernacular Regions, Midwest, Prairie, High Plains

Using Collocational Quotient to Map a Vernacular Great Plains

Introduction

This study uses colocational quotient (CLQ) to analyze the boundaries of the Great Plains. Herein we employ the CLQ metric on four regional terms "Great Plains," "Midwest," "Prairie," and "High Plains." This group of keywords cluster across a group of states running from Ohio in the east into the edge of the Northern Rockies in Colorado, Wyoming, and Montana and from the Canadian border in the north to North Texas in the south. The use of CLQ allows the establishment of bidirectional spatial relationships between individual categorical variables, in this case, the words in the names of businesses. Additionally, this approach enables us to establish how keywords are spatially colocated. Therefore, the use of CLQ potentially offers a more sophisticated understanding of how vernacular regions are bounded and how the terms interact with one another on the margins of regions. By understanding how multiple regional terms interact spatially, we can better understand the regional boundaries and subtle differences in keyword placement relative to one another.

In previous studies, attempts to delimit vernacular regions through establishment names have used hotspot analysis on single keywords or ratios of specific keywords to others in a defined study area. Both approaches have significant limitations. The individual keyword approach does not allow comparisons between differing terms, restricting our understanding of how regional naming patterns interact, especially in borderland areas. In contrast, the ratio

approach assumes no spatial autocorrelation between ostensibly neutral terms. So, although it enables comparisons of keywords, it requires finding terms that do not exhibit regional clustering for comparison. In contrast, CLQ enables us to quantify potentially asymmetric relationships between categorical variables. Thus, this metric allows us to understand the way multiple keyword variables interact with one another spatially in a much more sophisticated manner, including their deviation from randomness and their relative ratios within predefined areal units.

This study uses several of the keywords in two distinct senses. The first sense indicates the physiographic feature – for example, "Great Plains" as the expanse of grasslands that runs from west of the Mississippi to the Rocky Mountains. In contrast, the second sense is the density and distribution of the keyword such as "great plains" – potentially indicative of a vernacular region. Unless noted, we are referring to this second sense. Also, to clarify, the two uses are not coterminous, nor are they necessarily indications of the same underlying phenomena. The reason for differing boundaries between the vernacular region and physiographic features is manifold. Vernacular regions are defined by individual naming decisions, irrespective of the local landscape, leading to higher population areas providing more data. In the case of "Great Plains," major population centers exist along the physical feature's edge, indicating the dataset is skewed to the edges of the expansive grasslands. This skew towards more populated areas is not a limitation of the study as we do not expect keyword concentrations to define the physiographic feature. Instead, we attempt to detect what region residents of an area perceive themselves to be located.

Inspired by the expanding concept of borderlands, this study uses the CLQ metric to delimit these areas (Baud and Van Schendel 1997; Lauk 2019). Furthermore, as with other borderlands, the boundaries between regions are fuzzy, so no attempt will be made to draw lines

in this study, as any attempt at creating borders would be a misrepresentative reification of the concepts of vernacular region and borderlands. Instead, we visualize the keywords and describe the underlying patterns. Nevertheless, this novel use of the CLQ metric allows us to uncover spatial relationships between regions and keywords at scales, heretofore impossible.

Literature Review

Colocational Quotient

Locational quotient metrics have a long history in economic research, appearing as early as the 1940s (Miller et al. 1991). However, the CLQ metric was developed in the past decade by Leslie and Kronenfeld (2011). They developed the metric to understand spatial associations of categorical variables. Since its introduction, scholars have used this technique to examine a wide range of spatial phenomena. Leslie et al. (2012) uses the metric to examine clustering among different retail food establishments. Likewise, a series of papers have used CLQ to investigate the severity of vehicle crashes (Zou et al. 2017; Hu et al. 2018; Kuo & Lord 2020), crime (Yue et al. 2017; Wang et al. 2017; Pope & Song 2015), and residential segregation (Córdoba & Walter 2016; Vorotyntseva 2016; Córdoba et al. 2018). However, scholars have not used this technique to study regions to date.

Unlike older "locational" metrics, CLQ does not measure regional concentration. Instead, the metric measures the spatial relationship between different sets of categorical data. The technique has been described as a measure of "attraction" between multiple point data sets (Córdoba et al., 2018). In much of the scholarship, "attraction" is literal in that identifiable processes contribute to the comingling of the points at particular locations. Thus, push and pull

factors bring the variables into their relative proximities. However, we do not posit any such macro-scale human processes creating the boundaries delimited in this project. Instead, we look at the clustering of two variables indicative of perceptual transition zones.

Using ratio quotients is not a new approach to studying vernacular regions. Reed's "The Heart of Dixie: An Essay in Folk Geography" (1976), the first study to quantify a vernacular region using naming patterns, employed a cruder form of quotient analysis. So, this type of research is not new. However, this statistic allows us to look more precisely at boundaries between regions, obviating the need for arbitrary thresholds or interpolation -- two potential criticisms of the ratio approaches to defining vernacular regions using data aggregated to points or polygons.

Vernacular Regions

Terry Jordan (1978) describes vernacular regions as:

> …. perceived to exist by their inhabitants and other members of the population at large. They exist as part of popular or folk culture. Rather than being the intellectual creation of the professional geographer, the region is the product of the spatial perception of average people. Rather than being based on carefully chosen, quantifiable criteria, such regions are the composites of the mental maps of the population

The formal study of vernacular or perceptual regions[1] entered geography in the late 1960s. The first significant attempt to map these regions on a national level was Ruth Hale's (1971) underappreciated unpublished doctoral dissertation at the University of Minnesota, "A Map of Vernacular Regions in America." Hale pioneered a survey-based approach to defining vernacular regions. She mailed thousands of surveys across the country, asking which regions the participants felt they lived. With this data, she constructed a regional map of the United States. Her work remains one of the most ambitious attempts to define vernacular regions using this method and directly inspired several scholars, including Wilbur Zelinsky (1980), to explore these methods further.

John Shelton Reed, a prominent sociologist of the United States South, furthered the study of vernacular regions with his 1976 piece "The Heart of Dixie: An Essay in Folk Geography." This study was the first to use business naming patterns to delimit a vernacular region, in this case, the US South. Reed used local phonebooks from approximately 100 cities, tallying the total number of businesses in each city, beginning with the "Dixie," "southern," and "American." Next, Reed divided the terms Dixie and southern by the entries of American to create a ratio. With this ratio, Reed could demarcate a vernacular South using business names. The resulting map closely matched how various scholars have delimited the region, demonstrating the method's viability in detecting regional variation. Reed's piece, especially its use of business names, has led to numerous follow-up pieces over the years attempting to measure changes in the vernacular South Reed had initially charted. Most of these pieces dutifully replicated Reed's methods (Reed et al. 1990; Alderman & Beavers 1999; Cooper & Knotts 2010). However, a few others tried to use the general concept while utilizing more

advanced statistical techniques (Ambinakudige 2009; Andrews & Finchum 2020). Ultimately, Reed's and his followers' work showed the viability of using a ratio approach that tallied numbers of business names in cities and then used various forms of interpolation to map the vernacular South.

Great Plains and the Midwest

Scholars have done much less work charting vernacular regions outside the United States South. This lacuna, partially due to American academic geography's disregard for regional concepts closely linked to the physical environment, has left the idea of a distinct Great Plains cultural region understudied in the discipline. Instead, scholars have represented the region as a transitional zone between cultural regions. For example, Zelinsky's 1980 seminal study "North America's Vernacular Geography," inspired by Reed's use of business names, attempted to map all North America's major vernacular regions according to his somewhat opaque criteria. In this work, the Great Plains is not a distinct region; instead, portions of the area are in the Midwest, Southwest, West, and North. This regional configuration in Zelinsky's study comes from *a priori* commitments to a regional geography based on original settlement patterns and cultural homogeneity (Lubeke 1984). These theoretical commitments, most clearly articulated in Zelinsky's *The Cultural Geography of the United States* (1992), influenced the keywords that Zelinsky used to delimit the regions he draws in "North America's Vernacular Regions." In contrast, in the most comprehensive work on the Great Plains as a cultural region Shortridge (1988) combine several factors, including the initial settler population, to create a picture of the core Great Plains informed by historical settlement patterns and economic activity.

Lauk's excellent recent (2019) interdisciplinary volume on borderlands between the Midwest and the Great Plains provides several pieces arguing for a distinct Great Plains and a reexamination of this region's boundary and the Midwest. Lauk also argues for a Great Prairie region, to which this work lends some tentative support. In the same work, Laignen (2019) presents maps based on various criteria, including physiographic and business name data, to establish a boundary between the Midwest and the Great Plains. Notably, Laignen avoids drawing solid lines to demarcate the region instead of establishing transition zones. This work is similar to Rossum and Lavin's (2000) work aggregating large numbers of maps demarcating the region.

Only a few studies attempting to define vernacular regions in our study area have been published, most notably Shortridge's "The Vernacular Middle West" (1985). This article uses surveys of college students from around the country to draw cognitive maps of the Midwest region. These maps are combined with business name data used to create maps of the highest density of Midwest named businesses and organizations. The study finds notable differences in how students perceive the region inside and outside of it. Surprisingly, the study finds the average student places the core of the region in Kansas and Nebraska. In contrast, Shortridge finds the highest density per capita of "Middle West" businesses are in the Dakotas. The author speculates that the bucolic image of the yeoman farmer that dominates much of the country's impression of the region has shifted west to match these perceptions, in turn shifting away from the reality that much of the traditional Midwest was part of an emergent deindustrializing Rust Belt region.

Data and Methods

We acquired keyword data using ArcGIS Pro's Generate Points From Business Listings tool, searching for the keywords "Midwest," "Great pPains," "Prairie," and "High Plains" within the study area. These listings include businesses, educational, government, and non-profit organizations. Next, we excluded chains with over four locations, government, academic and duplicate listings, and businesses whose names refer to local municipalities or geographic features (e.g., Grand Prairie, Texas, Midwest City, Oklahoma). Midwest was the most numerous keyword after removing these entries from the data, with over 4,500 instances in the study area. Prairie is the second most numerous, with approximately 2,700 entries in the dataset. Meanwhile, the occurrences of High Plains and Great Plains are significantly lower, with 387 and 499 listings, respectively (Table 1).

Table 1. Keyword Frequency

Keywords	Listings
Midwest	4509
Prairie	2768
Great Plains	499
High Plains	387

We used the Colocational Analysis tool in ArcGIS Pro to compare the keywords. Nearest neighbors (NN) was used with eight neighbors. Nearest neighbors was used instead of distance due to the large difference in keyword density across the study area. A Gaussian local weighting scheme was used. The maximum number of permutations, 9,999 was used to calculate the p-values. The default .005 threshold was used to determine statistical significance.

Study Area

Prior research into vernacular regions at the subnational level has used the contiguous United States as the study area. This approach can potentially pick out anomalous pockets far from the perceived core region, such as the Dixie region in Utah (Andrews & Finchum 2020). However, this expansive approach can necessitate creating statistical neighborhoods that are too expansive to examine the data at the appropriate spatial scales. To avoid these issues, we created a study area focused on the most central keywords analyzed in the study – Great Plains (Figure 1).

Figure 1. Study Area

The study area runs north to south from approximately 49 ° N to 31 ° N and 89 ° W to 106°W, encompassing roughly 90 percent of the total instances of the keywords.

Spatial Autocorrelation

Table 2. Average Nearest Neighbor

Keyword	NN Ratio	Z-Score	P-Value
Midwest	0.347	-84.03	0.00
Prairie	0.453	-54.13	0.00
Great Plains	0.444	-23.74	0.00
High Plains	0.406	-22.40	0.00

As expected, all the terms exhibited high levels of spatial autocorrelation within the study area. All terms have nearest neighbor (NN) ratios well below one, indicating clustering. Likewise, all terms record Z scores between -84 and -22 (Table 2). These scores are likely exaggerated due to the composition of the study area, which includes an area of the southern United States that contains very few of any of the keywords and small portions of Canada that were not part of the study. The Z scores are so large that even with those caveats, all the keywords exhibit clustering well above any level of statistical significance.

Measures of Centrality

In the following, we create four deviational ellipses of the keywords, using the Directional Distribution tool in ArcGIS Pro 2.8. We used the mean center to calculate the centroids. The ellipses contain all keyword locations within a standard deviation of distance from

the mean center. The following measures are for businesses and organizations located within the study area. Thus, the following visualizations do not necessarily indicate the distribution of these terms nationwide. Nonetheless, using all locations in the contiguous United States produces only slight changes to the following data, as most keyword locations are within the designated area of study. "Midwest" is clustered in cities, especially Chicago (Figure 2). The centroid of the keyword is in northwestern Illinois, near the town of Galva. The directional distribution runs east-west from central Indiana to eastern Kansas and Nebraska. The term's western distribution is due to its relative popularity in the physiographic Great Plains. Meanwhile, the term sharply drops off east of Cleveland and south of the Ohio River. These factors create a standard deviational ellipses further west than most attempts to delimit the region.
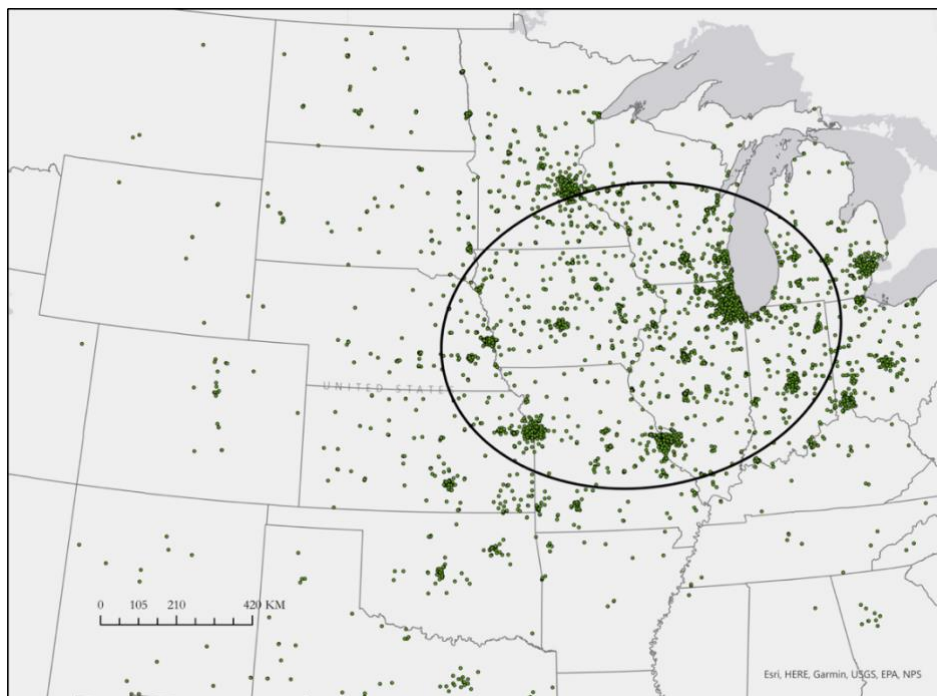

Figure 2. Locations of "Midwest" keyword and deviational ellipse of the term
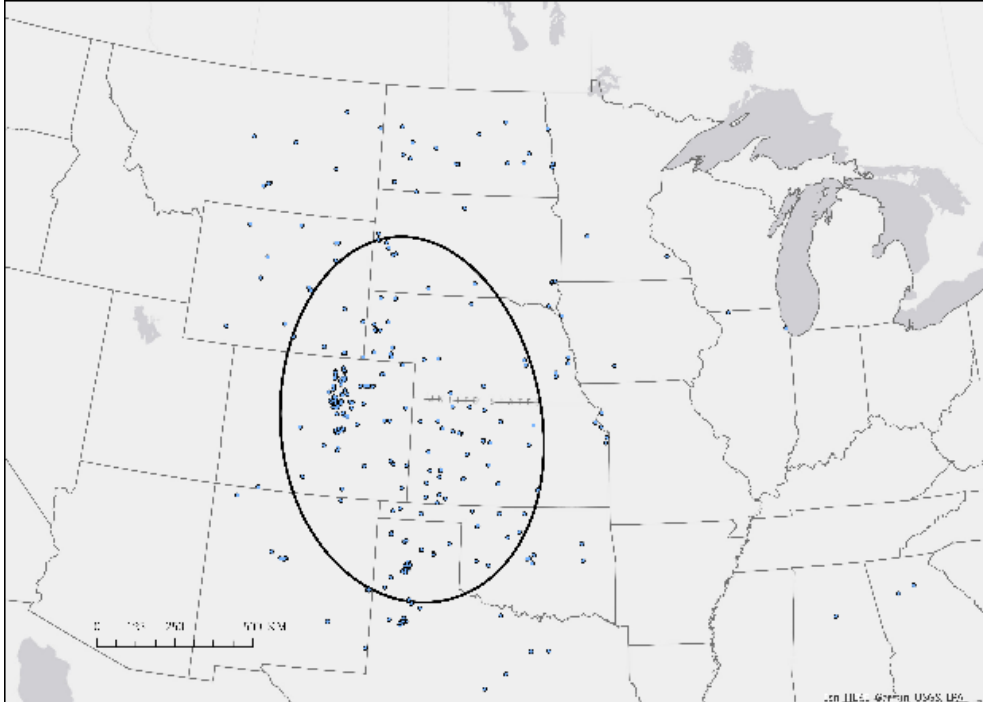
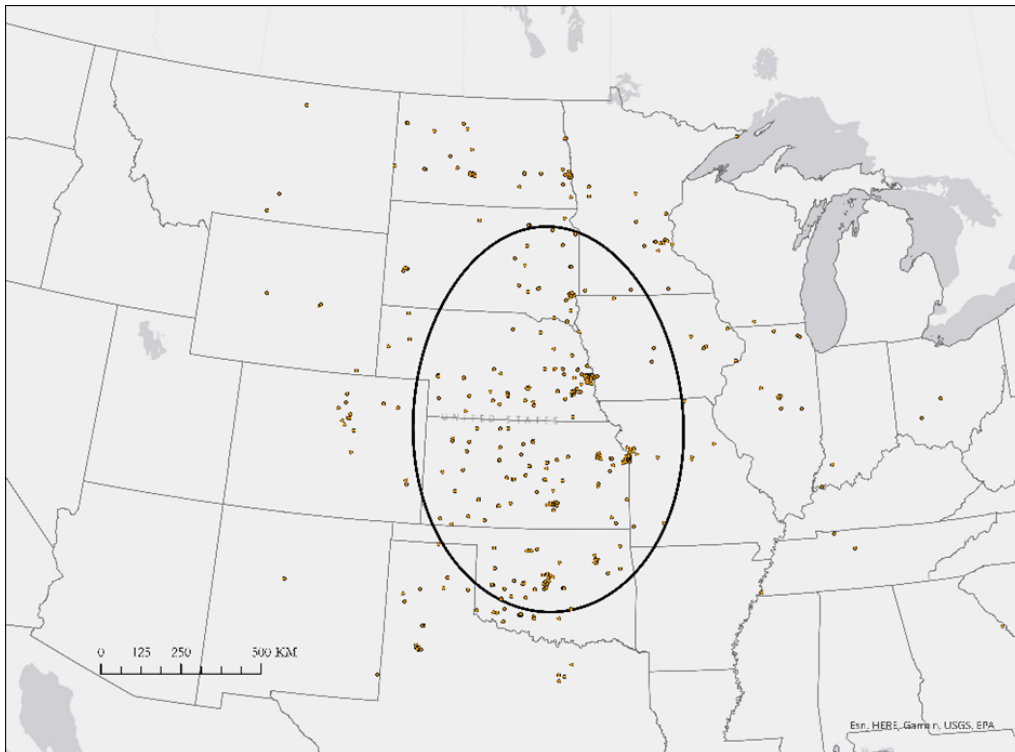Figure 3. Locations of "High Plains" keyword within the study area.

Figure 4. Locations of "Great Plains" keyword and deviational ellipse of the term
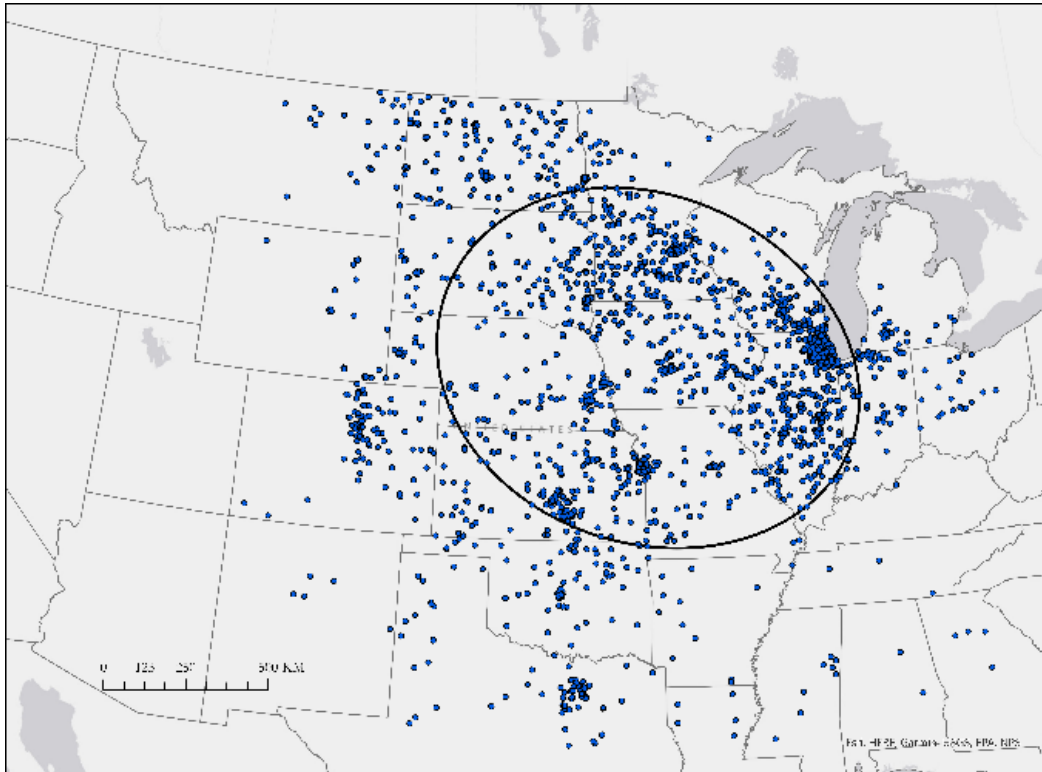
Figure 5. Locations of "Prairie" keyword within the study area

"Great Plains" is less clustered in cities and more dispersed across the landscape, with the ellipse

forming a slightly elongated circle (Figure 3). The centroid is in Washington County, Kansas,

approximately five kilometers southeast of the town of Haddam near the Nebraska border.

"High Plains" is centered in rural eastern Colorado (Figure 4). The ellipse has a slight southeast

to northwest orientation. The keyword is evenly distributed with less urban clustering than the

other keywords, instead clustering over the whole of the Front Range region

Prairie is centered in Iowa (Figure 5). However, this location represents the midpoints of a

slightly bimodal distribution more than an area of concentration like the other keywords. "Prairie"

has concentrations in central Illinois and southern Wisconsin, isolated from the major contiguous

area of Prairie named businesses in the upper Midwest and into Montana.

Two of the deviational ellipses, "Midwest" and "Prairie," are roughly east-west ovals. Indicating the points run in an easterly direction, as expected. "High Plains" exhibits a unique southeast-northwest orientation while "Great Plains" forms a circle indicating the data is evenly distributed north-south and east-west from the centroid. Overall, the dataset shows an overlapping east-west orientation, suitable for using CLQ to delimit boundaries.

Results

Conceptual Relationships

Before we can directly examine the results, we must sketch out what different spatial relationships using CLQ mean in the context of vernacular regions. The following is an attempt to sketch out a conceptual model of what distinctive spatial relationships indicated by CLQ might reveal about vernacular regions and their connection to keywords. Finally, to clarify, the relationships sketched out are speculative, scale-dependent, and only applicable in the context of keywords closely related to vernacular regions. Nonetheless, despite these reservations, we believe the following offers a start to understanding and interpreting vernacular regions using CLQ.

High degrees of colocation among two variable terms describe the same region; for example, "south and "southeastern" would likely fall into this category as descriptors of the American South. Isolation among terms indicates the opposite, with the degree of isolation likely showing the regional concepts' keywords' relative fuzziness. Unidirectional spatial relationships among variables likely indicate that the spatially correlated variable is potentially a subregion --

for instance, "new south" and "southern." New South could be clustered within "southern"; however, "southern" is likely much more dispersed and shows much less spatial correlation with "New South." Lastly, no spatial relationship between keywords indicates that they are likely not associated with regionally based naming patterns.

In this dataset, all terms exhibit perceptible east-west boundaries, representing areas of regional transition. This is intentional as we chose these terms to test the effectiveness of CLQ as these terms were likely to exhibit relatively well-defined boundary lines, as demonstrated in prior scholarship demarcating the regions. Also, while CLQ relationships are bidirectional, in practice, these keyword relationships map closely to one another; for instance, Midwest-Great Plains appears very similar to Great Plains – Midwest, so for concision, only one map is a map of each relationship is provided.

Great Plains – Midwest

Using the technique on the term Midwest in relation to instances of Great Plains, as seen

in Figure 6, provides one of the more defined boundaries. This map shows us that "Great Plains"

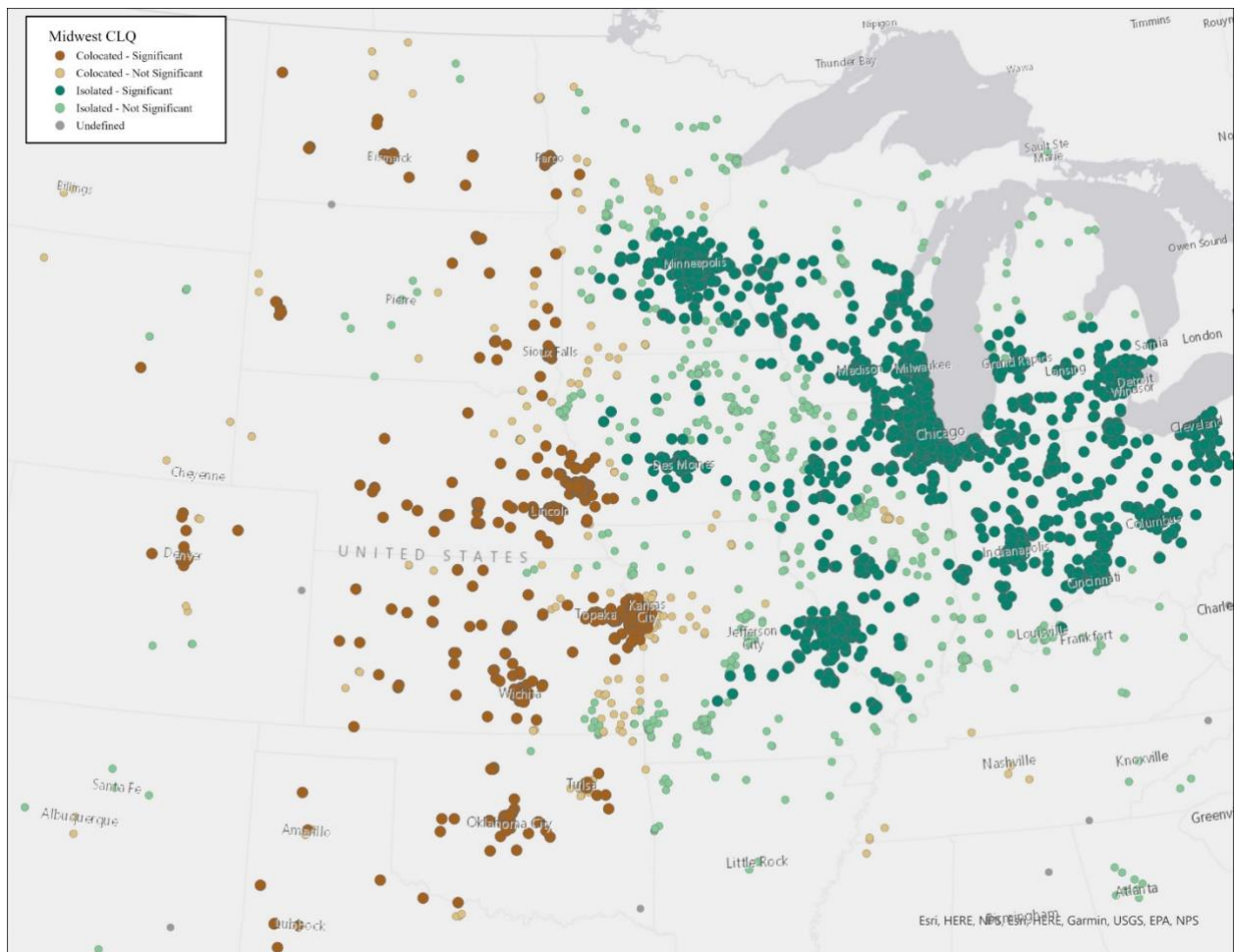businesses are situated further east than the physiographic Great Plains. The transition zone runs

Figure 6. Distribution of "Great Plains" compared to "Midwest" using CLQ

roughly along state borders beginning in the north at the boundary between Minnesota and the

Dakotas and moving south through Nebraska, Iowa, Kansas, and Missouri. The division becomes

less defined as it moves south of Kansas City, especially in southeastern Kansas, where both

terms are interspersed across the landscape in roughly equal numbers.

Compared to Great Plains, High Plains locations also provide a strong gradient running southeast to northwest from the North Texas Panhandle to eastern Montana (Figure 7). With concentrations in the Front Range region of Colorado and the Nebraska panhandle. This
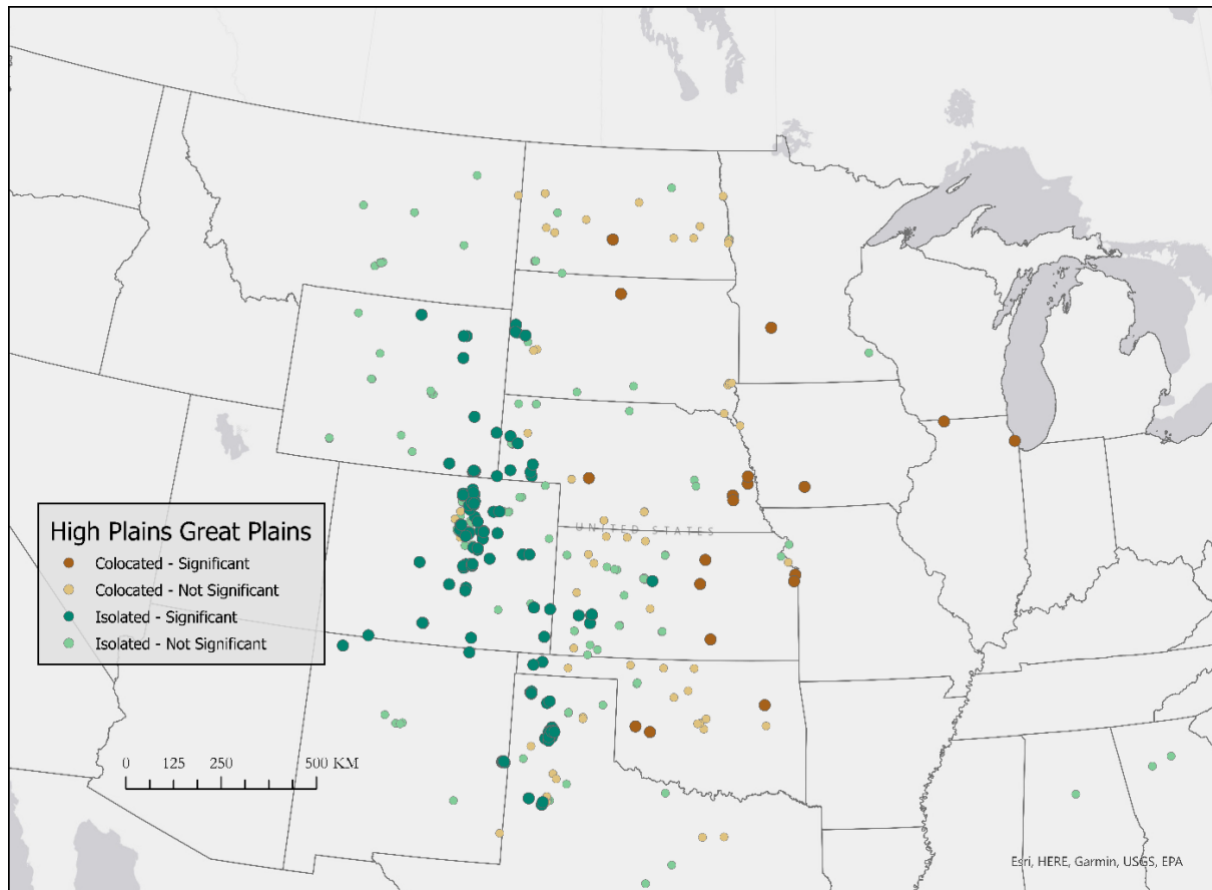

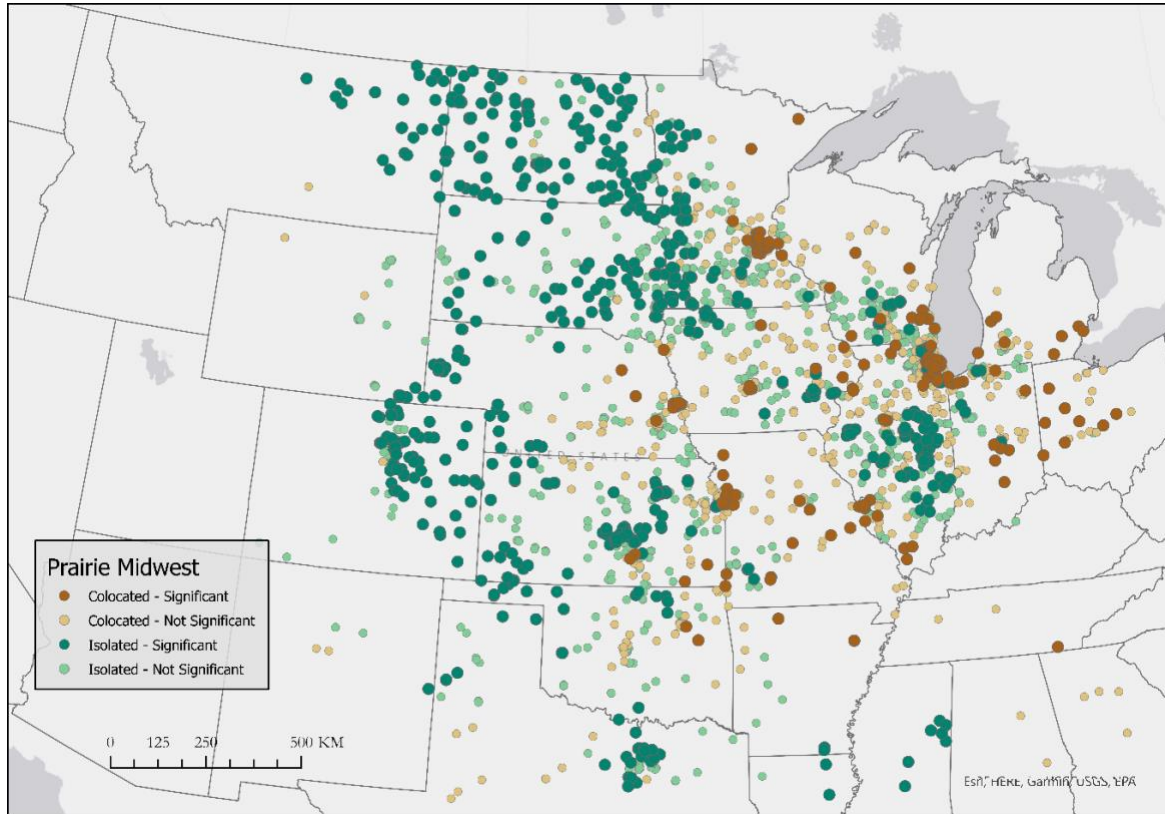Figure 7. Distribution of "Great Plains" compared to "High Plains"

region being overrepresented due to population. The only mildly surprising result is the extension of "High Plains" further east into Kansas than expected, a result at odds with the distribution in neighboring Nebraska, Oklahoma, and the Dakotas. However, this could just be a statistical anomaly with the small sample size in this area.

Prairie – Midwest

The comparison of "prairie" named businesses to the Midwest is perhaps the most interesting (Figure 8). These two keywords produce a less defined border than the prior keyword combinations. Nonetheless, several interesting patterns emerge. Notably, the terms are colocated in


Figure 8. "Prairie" Distribution Compared to "High Plains"

large urban areas, including Chicago, a "Midwest" stronghold, and Minneapolis – St Paul, Kansas City, Omaha, and St. Louis. These keywords exhibit the most robust urban-rural pattern. There is also an isolated "prairie" area in central Illinois, likely because Illinois is known as the "Prairie State" and another relatively remote area in south central Wisconsin centered on Madison, where prairie predominates. Beyond these areas, Prairie – Midwest follows the spatial patterns of Midwest – Great Plains, except Prairie extends notably further east in the Upper Midwest, encompassing most of the western third of Minnesota and parts of northwest Iowa.

Further south, the borders are much closer to the Midwest divide. Prairie exhibits some

unique configurations in the western end of the study area. The term is located further west in the

northern part of the study area, with numerous instances of the keyword located well into central

Montana. The term is virtually nonexistent in neighboring Wyoming and then becomes

extremely common again in the eastern half of Colorado. There is also an area of Mississippi and

northern Louisiana, which are exclusively Prairie and technically lie inside the study area but are

outside of the scope of this study.

High Plains – Prairie

High Plains and prairie are only colocated in the western part of the study area, as the

term "high plains" is the most geographically compact of the keywords used in this study (Figure

9). The terms High Plains and Prairie appear to be the most mixed in the areas of colocation,
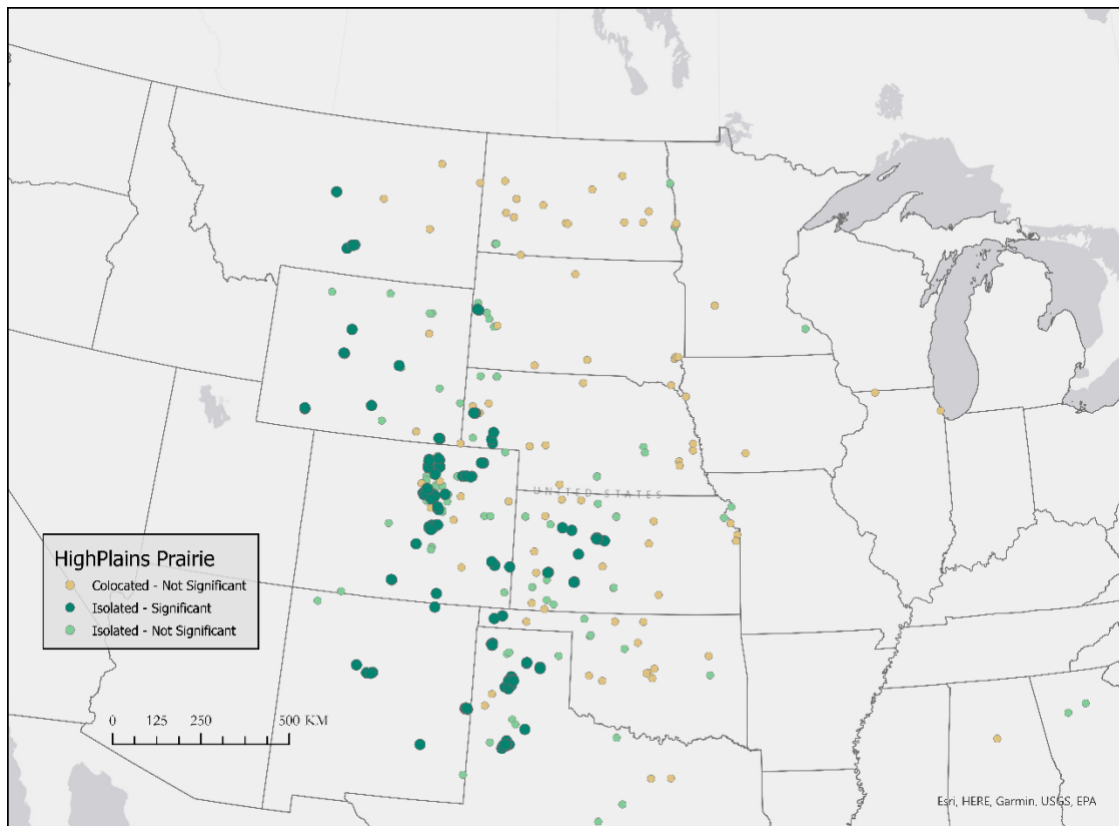
with extensive areas of colocation and isolation in the Front Range area of Colorado into western Kansas, the Nebraska Panhandle, eastern Wyoming, and central New Mexico.

Limitations

Determining the appropriate study area for use with spatial statistics involves tradeoffs. In this study, much of the southeastern area of the study area contains very few keywords. The inclusion of this area was necessitated by the need for a rectangular study area that includes Ohio, Indiana, and Illinois. This likely exaggerates the concentration of the nearest neighbor statistics for three of the keywords. However, due to nine "high plains" locations in the far southeast of the study area, the study area likely artificially increases dispersion numbers for

Figure 9. "High Plains" compared to "Prairie"   "high plains." This low density in the southeast of the study area also can skew results because this analysis uses K nearest neighbors to determine contiguity. However, since the density of points is low and these few points are not near the significant areas of colocation, the size of the study area has not had a substantial effect on the results. Another potential issue is the appropriateness of using K nearest neighbor for all the keyword combinations. For consistency, the same method was used for all comparisons. However, High Plains – Prairie is more spatially clustered than any other keyword pair, indicating that using a distance band instead of nearest neighbor analysis might be more appropriate for this pair of keywords. Nonetheless, this potential limitation does little to undermine the analysis as an overlap of those terms is not unexpected.

92

Discussion and Conclusion

Attempting to bound the Great Plains is a particular challenge because, unlike more well-defined regions such as "the South" or "New England," the borders are porous and encompass several more salient cultural areas, making ratio approaches potentially problematic. For instance, the eastern Dakotas patently have more in common culturally with the Upper Midwest than West Texas or the Oklahoma Panhandle. However, both are in the Great Plains. Thus, this region is partially based on physiographic, not cultural affinity. The keywords used in this study make that distinction clear, and this is why we have included the terms "prairie" and "high plains," as both are descriptives of landscape features without close connections to cultural regions.

CLQ works well in border areas where terms overlap. Due to the geographies and scale at play in this study, this closely mimics the distribution of these terms. We designed this study for keyword distributions with well-defined directional gradients and a Venn diagram-like pattern described above. Also, to clarify, we are not proposing CLQ as a replacement for interpolation-based approaches to mapping vernacular regions. Instead, we believe CLQ offers an additional tool for understanding relationships between names providing more context than simple ratio maps used in research to date, especially in transition areas or regions with overlapping or nested regional identities.

It is the study of these areas with nested regional identities that this technique perhaps has the most promise for future work. For instance, by using CLQ to compare subregional terms, such as 'New South," to terms that signify more extensive regional affiliations, such as "Southern," scholars could detect subregional variation that is not geographically contiguous.

The novel use of CLQ in this study provides an additional method to detect the boundaries of vernacular regions. This approach shows promise in demarcating regions with fuzzy boundaries, such as the Midwest and Great Plains. CLQ helps delimit a series of relatively well-defined border areas between the keywords in this study. The borders produced by this method are partially due to the nature of the ecotone separating the regions. Namely, the regions' physiography runs in a well-defined east to west direction. Likewise, population density and business density mimic this pattern, decreasing density across most of the study area as one moves east to west. Regions exhibiting these characteristics are good candidates for using this technique, providing another useful statistical tool for analyzing vernacular regions.

Bibliography

Alderman, D. H., & Beavers, R. M. (1999). Heart of Dixie Revisited: An Update on the Geography of Naming in the American South. *Southeastern Geographer*, *39*(2), 190–205.

Ambinakudige, S. (2009). Revisiting "the south" and "dixie" delineating vernacular regions using GIS. *Southeastern Geographer*, *49*(3), 240–250.

Andrews, J.R., & Finchum, G.A. (2020). Paring Old Dixie Down: The Dixie Highway and the Mapping of a Vernacular South. *Southeastern Geographer* 60(4), 345-359 doi:10.1353/sgo.2020.0028.

Baud, M., & Van Schendel, W. (1997). Toward a Comparative History of Borderlands. *Journal of World History*, *8*(2), 211–242.

Cooper, C. A., & Knotts, H. G. (2010). Declining Dixie: Regional Identification in the Modern American South. *Social Forces*, *88*(3), 1083–1101. https://doi.org/10.1353/sof.0.0284

Córdoba, H. A., Walter, R. J., & Foote, N. S. (2018). The residential segregation of San Antonio, Texas in 1910: An analysis of ethno-racial and occupational spatial patterns with the colocation quotient. *Urban Geography*, *39*(7), 988–1017.

Córdoba, H. A., & Walter, R. J. (2016). Measuring the Spatial Pattern of Ethnic Groups in San Antonio's Eastside: A Spatial Analysis Approach with the Colocation Quotient. *Papers in Applied Geography*, *2*(2), 160–177.

Hale, R. (1971). *A Map of Vernacular Regions in America. Unpublished doctoral dissertation. Minneapolis: Univ. Of Minneapolis, 1971.*

Hu, Y., Zhang, Y., & Shelton, K. S. (2018). Where are the dangerous intersections for pedestrians and cyclists: A colocation-based approach. *Transportation Research Part C: Emerging Technologies*, *95*, 431–441.

Jordan, T. G. (1978). Perceptual Regions in Texas. *Geographical Review*, *68*(3), 293–307. https://doi.org/10.2307/215048

Kuo, P.F., & Lord, D. (2020). Applying the colocation quotient index to crash severity analyses. *Accident Analysis & Prevention*, *135*, 105368.

Laingen, C. (2019). Geographies of the Borderlands. In J. Lauck (Ed.), *The Interior Borderlands: Regional Identity in the Midwest and Great Plains*. Center for Western Studies, Augustana University.

Lauck, J. K. (Ed.). (2019). *The Interior Borderlands: Regional Identity in the Midwest and Great Plains*. Center for Western Studies, Augustana University.

Leslie, T. F., Frankenfeld, C. L., & Makara, M. A. (2012). The spatial food environment of the DC metropolitan area: Clustering, colocation, and categorical differentiation. *Applied Geography*, *35*(1–2), 300–307.

Leslie, T. F., & Kronenfeld, B. J. (2011). The Colocation Quotient: A New Measure of Spatial Association Between Categorical Subsets of Points. *Geographical Analysis*, *43*(3), 306–326. https://doi.org/10.1111/j.1538-4632.2011.00821.x

Lewis, G. M. (1966). William Gilpin and the concept of the Great Plains region. *Annals of the Association of American Geographers*, *56*(1), 33–51.

Luebke, F. C. (1984). Regionalism and the Great Plains: Problems of concept and method. *The Western Historical Quarterly*, *15*(1), 19–38.

Miller, M. M., Gibson, L. J., & Wright, N. G. (1991). Location Quotient: A Basic Tool for Economic Development Analysis. *Economic Development Review*, *9*(2), 65.

Oxford English Dictionary. (2020.). *"prairie, n."*.

   https://www.oed.com/view/Entry/149304?redirectedFrom=praire&

Pope, M., & Song, W. (2015). Spatial relationship and colocation of crimes in Jefferson County,

   Kentucky. *Papers in Applied Geography*, 1(3), 243–250.

Reed, J. S. (1976). The Heart of Dixie: An Essay in Folk Geography. *Social Forces*, *54*(4), 925–939.

   https://doi.org/10.1093/sf/54.4.925

Reed, J. S., Kohls, J., & Hanchette, C. (1990). The Dissolution of Dixie and the Changing Shape of

   the South. *Social Forces*, *69*(1), 221–234.

Rossum, S., & Lavin, S. (2000). Where are the Great Plains? A Cartographic Analysis. *The*

   Professional Geographer, *52*(3), 543–552. https://doi.org/10.1111/0033-0124.00245

Shortridge, J. R. (1985). The Vernacular Middle West. *Annals of the Association of American*

   Geographers, *75*(1), 48–57.

Shortridge, J. R. (1988). *The Heart of the Prairie: Culture Areas in the Central and Northern Great*

   Plains. 17.

Vorotyntseva, N. (2016). *Measuring Segregation Patterns and Change: A Colocation Quotient Approach* [Dissertation, University of Connecticut]. https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=7330&context=dissertations

Wang, F., Hu, Y., Wang, S., & Li, X. (2017). Local indicator of colocation quotient with a statistical significance test: Examining spatial association of crime and facilities. *The Professional Geographer*, *69*(1), 22–31.

Zelinsky, W. (1980). North America's vernacular regions. *Annals of the Association of American Geographers*, *70*(1), 1–16.

Zelinsky, W. (1992). *The cultural geography of the United States*. Pearson College Division.

Zou, W., Wang, X., & Zhang, D. (2017). Truck crash severity in New York City: an investigation of the spatial and the time of day effects. *Accident Analysis & Prevention*, *99*, 249–261.

Yue, H., Zhu, X., Ye, X., & Guo, W. (2017). The local colocation patterns of crime and land-use features in Wuhan, China. *ISPRS International Journal of Geo-Information*, *6*(10), 307.

CHAPTER V

CONCLUSION

Conclusion and Future Work

These three pieces advance the study of vernacular regions in human geography. All three use a critical methodological insight gleaned from earlier scholarship – business naming patterns can be used to study regions and advance that concept in various ways. The first piece, "Paring Dixie Down: The Dixie Highway and the Mapping of a Vernacular South," examines prior scholarship using business naming patterns to map the US South and finds these methods are lacking without a nuanced understanding of local particulars and the ability to distinguish between names based on proximity versus affinity. While these two concepts are not mutually exclusive, we show in the piece that the distance to a major roadway heavily influenced the instances of Dixie named establishments in the immediate areas it ran through. This, in turn, led scholars to significantly overestimate Dixie as a term signaling regional affinity instead of operating as a locational signifier. Ultimately, this piece demonstrates the need for local knowledge when evaluating this type of data. Advances in technology have made this type of study much easier than in the past. Nonetheless, local knowledge is still needed to interpret the results correctly.

The second piece, "Reconstructing a Vernacular South," adapts prior scholars' methods to push the study of the vernacular South into the early 20th Century. This piece was initially conceived as an extension of the first piece. However, as the data collection area expanded and the results proved interesting outside of the narrow question addressed in the first article, the research questions were expanded. This piece finds that using similar methods to those pioneered by John Shelton Reed produces on historical sources produces interesting results. More importantly, it once again found the need to understand the context – in this case, the historical

context of the data. "Dixie" was a term that was not in widespread use as a business name in the first decade of the 20th Century, while during this time, the word "Southern" remains as relatively prevalent as it would appear in future studies. This highlights a second theme these three pieces implicitly address. Namely, methods in the social sciences are rarely universally applicable but instead depend on spatially and temporally limited assumptions. Awareness of these limitations can often provide insight into the underlying processes and advance understanding beyond the methods themselves. This piece manages to highlight this dynamic.

The third piece finds that colocational quotient (CLQ) works well in roughly delimiting vernacular regions in transitional areas with fuzzy borders. This approach worked particularly well with what is perhaps the fuzziest regional border in the United States, the boundary between the Midwest and Great Plains. Many attempts have been made to demarcate these two regions, and this piece is far from the final word. However, it points to a quantitative method that can accommodate the messiness of borderlands. The ability to demarcate these types of transitional regions is the capacity CLQ seems to offer the most promise as a research tool for exploring vernacular regions.

Future Work

These three works have explored the concept of vernacular region using business names in various ways. However, much potentially fruitful work remains to be done by expanding these studies temporally and spatially. The most obvious extension of this line of scholarship is extending the timeline for examining the terms "Dixie" and "Southern," although this is difficult with the city guide dataset I used. However, it is possible to collect the same data using phonebooks after 1930. As I mentioned in Chapter Four, extending the timeline backward using these methods is more complicated as they are much less linked to the region in 1910 than in 1930. This is likely due to fewer small retail businesses and changes in naming conventions. These effects presumably become more salient further back in time, at least in the case of keywords indicative of southern regionalism. However, it could be interesting to explore other regional terms to see if this pattern holds or is particular to the South.

Using additional sources of data could also expand this line of research. The use of phonebooks, which by the 1940s included rural directories and addresses, allows for more granularity in the data. This approach potentially overcomes one of the major weaknesses of the city guide data: its ties to cities, leaving rural areas largely uncovered. Unfortunately, currently, there is no digitized database of phonebooks nationwide with good temporal coverage, so collecting this data would be time-consuming. However, in theory, using phonebooks beginning in the 1940s, a researcher could put together statewide directories of keywords as point data, greatly expanding the possibilities of analysis.

Another promising avenue of historical research is the addition of city guides without widespread temporal coverage to the dataset. In Chapter Three, I used only cities with entries

around both target years. Only cities with multiple data points were used to make the data directly comparable and the project more manageable. A more ambitious project operating without my time and labor constraints could process much more data. Using all cities or even a portion of the ones with only entries for a particular decade could effectively quadruple the size of the dataset. This more extensive coverage could potentially offset issues of comparability using similar methods. More speculatively, the entirety of the city guide data from every year could be used in a project that created a system of spatial and temporal weights based on population, relative isolation, and years covered in the dataset to create an interpolated map of the region. However, absent a reliable process to digitize and read the data, this process would be highly time-consuming, and any potential weighting system would be challenging to test objectively.

Additionally, these point data would theoretically allow CLQ analysis of the same type used in Chapter Four. Although I am skeptical that historical analysis of the Great Plains and the Midwest would significantly change from the present boundaries, as that line is connected to physiographic form and the location of urban areas, neither of which have substantially changed in the past century. The more fruitful approach in extending CLQ further to study regionalism could involve using interregional terms. For instance, Midsouth and New South and how they interact within the South writ large. I conducted a preliminary analysis of several of these terms in various regional configurations, which worked well. However, none of the results was revelatory, but that could have been more due to my lack of originality in choosing terms. Nevertheless, the technique is viable using intraregional terms that do not exhibit strong regional contiguity but exhibit some other type of spatial clustering. For instance, "New South" exhibits more of an urban-rural divide than contiguous boundaries.

Even more ambitious work could be done in areas outside of North America. Some recent work has been done in Central Europe on vernacular regions using similar methods to those employed in these studies (Marek 2020). However, the work tends to be more theory-heavy, and scholars have yet to use business names in the same manner as in North American scholarship. Nonetheless, a change in geography, in addition to potentially informative attempts to delimit region, could also yield insights into a host of other structural differences between North America and new study areas. Including crucial differences in how public establishments are named and distributed spatially on the landscape, much like the data from the 1910s in Chapter Three, stretching the method gives us more insight into how the underlying structure of naming functions.

Although empirical work using naming patterns to delineate regions has fallen out of favor, as we have seen, this approach still offers many fruitful avenues of exploration, both via methodological innovation and using unutilized sources of data. These three works are potentially just the beginning of scholarship using this source material. Far more work can be done methodologically and by expanding the source materials. I hope that I, along with other scholars, can continue this line of inquiry in the coming years and the empirical study of naming patterns becomes a more permanent feature of scholarship in human geography.

Bibliography

Marek, P. (2020). Transformation of the identity of a region: Theory and the empirical case of the

    perceptual regions of Bohemia and Moravia, Czech Republic. *Moravian Geographical Reports*,

    28, 154–169. https://doi.org/10.2478/mgr-2020-0012

VITA

Jesse R. Andrews

Candidate for the Degree of Doctor of Philosophy

Doctor of Philosophy

Thesis: THREE STUDIES USING BUSINESS NAMING PATTERNS TO DELIMIT
REGIONAL GEOGRAPHIES

Major Field: Geography

Biographical: Jesse R. Andrews (is a Ph.D. candidate at Oklahoma State University in Stillwater,
Oklahoma. His research interests include historical, business, and urban geography,
along with the study of vernacular regions, particularly the American South.

Education:

Completed the requirements for the Doctor of Philosophy/Education in your major at
Oklahoma State University, Stillwater, Oklahoma in July 2022.

Completed the requirements for the Master of Arts Geography at Appalachian State
University Boone, North Carolina in December 2017

Completed the requirements for the Bachelor of Arts in History at the University of North
Carolina Greensboro Greensboro, North Carolina May 2013