CLUSTER-BASED DELINEATION OF MEGAREGIONS IN THE UNITED STATES: IDENTIFYING ADMINISTRATIVE BOUNDARIES THAT REFLECT META-COMMUNITIES TO IMPROVE THE EFFECTIVENESS AND EFFICIENCY OF GOVERNMENT

A Thesis
Presented to
The Academic Faculty

By

Bryce Miller

In Partial Fulfillment
of the Requirements for the Degree
Master of City and Regional Planning

Georgia Institute of Technology

May 2021

© Bryce Miller 2021

CLUSTER-BASED DELINEATION OF MEGAREGIONS IN THE UNITED STATES: IDENTIFYING ADMINISTRATIVE BOUNDARIES THAT REFLECT META-COMMUNITIES TO IMPROVE THE EFFECTIVENESS AND EFFICIENCY OF GOVERNMENT

Thesis committee:

Dr. Catherine Ross
Harry West Professor of the School of City and Regional Planning
Georgia Institute of Technology

Dr. Subhrajit Guhathakurta
Professor and Chair of the School of City and Regional Planning
Georgia Institute of Technology

Dr. Arthi Rao
Research Scientist and Lecturer for the School of City and Regional Planning
Georgia Institute of Technology

Date approved: 28 April 2021

Faerie is a perilous land, and in it are pitfalls for the unwary and dungeons for the overbold ... In that realm a man may, perhaps, count himself fortunate to have wandered, but its very richness and strangeness tie the tongue of a traveler who would report them.

J.R.R Tolkien

For my wife Dana and my son Robert.

ACKNOWLEDGMENTS

I would like to thank the members of my thesis committee for their help in preparation of this work – Catherine Ross with her constant support, understanding, and enthusiasm, Subhrajit Guhathakurta for pushing me to consider theoretical bases and his recommendation of Manuel Castell's works, and Arthi Rao for her help working through data processes and guiding me to the important questions I needed to be asking.

Particular thanks go to the family and friends who support. With their help, I was able to focus on my work when I needed to and to take breaks when they knew I needed them. Old friends like Mark Parsons and Tyler Parker always encouraging me to continue on when things were hard and who acted as sounding boards for my ideas. New friends in my cohort who I commiserated and celebrated with as we do our best to solve the world's problems and keep each other sane.

The author gratefully acknowledges the support of their collegiate education by the Veteran Affairs Dependent Educational Assistance program.

Table of Contents

List of Tables

List of Figures

SUMMARY

Coordination and collaboration through governance at meta-urban scales have the potential to significantly improve quality of life while reducing the bureaucratic burden on society. Megaregional research and delineation has largely focused on scholarly inquiry into specific relationships using narrow datasets or on private efforts to identify market opportunities with opaque analysis methods. This work aims to provide a megaregion delineation that is transparent, data diverse, and comprehensible to a degree that the resulting boundaries are well suited to administrative implementation. The process developed leverages a combination of cluster analysis and metropolitan planning organization locations to identify sub-regions that share morphological characteristics and functional relationships. Recommendations are made for subsequent research into four areas: new data sources, process refinements, applications for megaregional planning, and implementation principles for megaregional government.

CHAPTER 1

Introduction

As society and the economy have globalized over the last century, connections between communities have grown in number, type, and complexity. The birth of regionalism coincided with this globalization as well as the beginning of the industrial revolution in Europe. Lower costs of travel and communication built new connections, and groups of communities began to be considered as units that were greater than the sum of their parts. Previously distinct municipalities began to meld, formally or morphologically, into a new class of cities whose political and economic influence directed the burgeoning global market.

Today, regional science and community planning have developed terms for these meta-urban forms consisting of highly interconnected communities separated by physical distance; one of these is "megaregion." Megaregions were initially proposed and gained recognition during the early 2000s, and several definitions from the last two decades are discussed in the literature review. In general, megaregions may be described as a set of adjacent geographies across a range of community densities that share morphological characteristics and functional relationships.

The concept of megaregions built on preceding terms and ideas by shifting a greater focus onto inter-community relationships and relaxing the constraints imposed by continuous urban environments. The resulting geographies encompass larger portions of the national population and economy while still maintaining meaningful identities. If these megaregions can be leveraged to increase coordination and collaboration between both public and private entities with similar interests, dead-weight losses could be reduced and the resulting gains used to help create greater equity and opportunity for residents. The next challenge for research then lies in delineating practical regions that maximize the commonalities of constituent parties' interests and motivators.

1.1 Current Delineation Shortcomings

Based on the literature reviewed by this work, megaregional delineation methodologies fall into one of two categories: low variable count and method opaque. Low variable count methods show clear results and definitive boundaries by focusing on a single data source (e.g. night-time light emission) or a narrow set of factors (e.g. population and employment growth rates). While the results of these efforts are important in understanding our societal structures, a reductionist tackling of individual facets fails to describe communities that are inherently multifaceted, complex, and highly interrelated. Further, the end goals of the research are often more scholarly than practical, seeking to understand phenoma rather than utilize them.

Conversely, research that utilizes a wide range of data types rarely describes the details of the methodology used and why various data sources were selected. Techniques this opaque cannot be reliably replicated without working directly with the authors, and, while such a system may suffice for internal analysis and individual projects, boundaries for governance must be transparent and changes easily understood by members of the community. These efforts are frequently undertaken by public entities or public-private partnerships with the purpose of analyzing a particular situation, both with regard to geography and to specific topics such as freight, and not to apply the methodology on a national scale.

An example of clear delineation is the Metropolitan Planning Organization's (MPOs) instantiating trigger: any urbanized area with a population over 50,000 must form an MPO with additional responsibilities added once the area is over 200,000 residents. Unfortunately, researchers in the last decade have found that a purely morphological connection does not exist above the regional or metropolitan scale in the built environment. Therefore, such a clear, concise, and physical trigger is not conducive (and may not exist) for describing community groups that are connected through non-permanent factors, and a more complex method is required to connect morphological and functional characteristics.

As urban populations grow, urbanized area footprints expand, and the disparities between municipal staff capacity and demand remain, megaregional government is poised to act as a intermediary between federal and state that control fiscal pipelines and the local community needs that those pipelines are meant to support. However, a governmental body requires definite boundaries, and megaregional research has yet to develop a common delineation methodology that is easily reproducible, highly transparent, and quantitatively rigorous.

1.2 Research Goals & Methods

This work adds value to the existing body of literature by focusing on a delineation process with the intent of administrative and governmental application rather than academic curiosity and investigation. From this requirement and the gaps in existing research, several are placed on the creation of the methodology in this text. A wide selection of data sources must be used to reflect the intricate nature of the societal phenomena in question. The process must easily documented such that future research can easily build upon its successes and address its failures, decision makers can parse its intentions and results into legislation, and members of the public can comprehend its logic, causal factors, and results without extensive knowledge in the field.

The delineation process utilizes clustering algorithms; and, although they do not inherently return concrete conclusions, cluster analysis excels at identifying patterns in settings too complex for human-driven, qualitative approaches. Their ability to identify communities that share morphological and functional characteristics across a broad range of input datasets addresses the shortfall of the low variable count methods, thereby developing megaregions that are well situated to tackle the concerns that already motivate inter-government coordination such as environmental stewardship, infrastructure investment, and business networks.

Cluster analysis generally seeks to group observations such that a distance metric between members of a cluster is minimized across the entire set. Ambiguity stems from the fact that a specific number of clusters is not identified as part of the process. Instead, the algorithm returns a set of partitions for every number of clusters from one to the number of observations. This work selects that optimal number of clusters by comparing in-cluster distances with the number of observations assigned to each cluster. Minimizing the mean of the average in-cluster distance between observations works to create regions that share morphological characteristics and functional relationships, and this value decreases as the number of clusters increases. Maximizing the number of members in each cluster reflects the desire to leverage resources among a large group of similar communities. Since both values fall as the number of clusters rises, the number of members is inverted, and a number of clusters is selected that minimizes the mean sum of the values across all datasets under consideration.

To ensure that the resulting regions are well-suited for government and similar to existing types of administrative boundaries, spatial distances are mixed with non-euclidean measures in the clustering process. Additionally, checks for contiguity and islands remove or add sub-regions after the clustering and region identification steps. These steps were inspired by conversations and research surrounding gerrymandering of political boundaries.

1.3 Subsequent Opportunities

Two main priorities emerge from the delineation process described in this work: process refinement and implementation strategy. The process laid out in the Methodology chapter includes two data source selection aspects. The first is a filtering process which requires a certain amount of spatial-temporal coverage and availability. The second describes how to add subsequent data sources in such a way that a minimum spanning tree is developed, preventing project size runaway. Still, there are opportunities for exploration of data in fields that are not currently a part of the megaregional delineation literature. Research in these areas may yield a more efficient set of data sources that accurately describe communities' interconnections.

As more efficient data source sets are identified, the clustering analysis and subsequent region identification process are also candidates for improvements. The time and resource constraints of this work prevent the complete exploration of all avenues and algorithms. New clustering algorithms might be developed that are more efficient, MPO prioritization can be performed over a wide set of parameters, and so on. Ultimately, the goal of the methodology laid out here is to provide a framework which prioritizes results that balance community homogeneity with economies of scale.

Regarding implementation, only high level recommendations are provided here. Adding an additional layer of government on top of already bureaucratic and slow systems could be highly unpopular with both law makers and the public, especially in the United States with its strong history of anti-centralized government figures and movements. Therefore, care should be taken to ensure that the implementation of megaregional government prioritizes incentives that far outweigh any costs—in funding or independence—to local governments and that megaregional government also reduces the burden of governance on society as a whole. Further commentary on how work in these area may proceed is covered in the concluding chapter of this work.

CHAPTER 2

Literature Review

2.1 History of Regionalism in Planning

The study of regions and megaregions is a natural extension of community planning that grows from the globalization of the modern world with roots firmly planted in the history of the field. From the mid 1700s through to the mid 1900s, technological developments enlarged the functional and morphological scale of communities. This trend resulted in larger cities, a greater percentage of the world population living in those urban areas, and dwindling lag in communication time. At the same time, socioeconomic advantage and disadvantage also concentrated; and these widening societal inequalities are likely primary sources of the modern issues that community planning struggles to address.

2.1.1 Patrick Geddes' Regional Survey

Although a man of many interests, Patrick Geddes is frequently considered to be one of the originators of the community planning field. Initially a student of biology, Geddes eventually began to seek the "coordination of man with his environment" (Tyrwhitt 1947), beginning with the improvement of his neighborhood in Edinburgh and eventually spending several years doing planning work in India where he solidified his synthesis of planning expertise, physical harmony, economics, and social complexities (Cherry 1974). This idea of harmony can be seen in the following excerpt from his *Report on the Towns in the Madras Presidency, 1915: Madura*.

> "Town-planning is not mere place-planning, nor even work-planning. If it is to be successful it must be folk-planning. This means that its task is not to coerce people into new places against their associations, wishes, and interest—as we find bad schemes trying to do. Instead its task is to find the right places for each sort of people; places where they will really flourish. To give people in fact the same care that we give when transplanting flowers, instead of harsh evictions and arbitrary instructions to 'move on,' delivered in the manner of officious amateur policemen." (Tyrwhitt 1947)

Considering the pandemic, riots, and other upheavals that have rocked the world during 2020 (and which may continue into the foreseeable future), this advice for planners from over a century ago remains timely.

Geddes referred to his school of planning as “seeking to undo as little as possible, while planning to increase the well-being of the people at all levels, from the humblest to the highest.” (Tyrwhitt 1947) He believed improvements with this goal in mind tended towards lower cost and greater returns. Geddes’ underlying rationale for this is illustrated in his comparison of community planning to chess.

> “The problem of city planning, as of chess, is to improve the situation by, as far as may be, turning its very difficulties into opportunities. Results thus obtained are both more economical and more interest, even aesthetically, than those that are achieved by clearing the board and re-setting all the pieces.” (Tyrwhitt 1947)

Some of Geddes’s idea on conversation can be seen decades later in preservationist movements by Jane Jacobs and her contemporaries as activists began to push back against urban renewal. However, it seems that this conservation-focused, low-impact methodology failed to integrate into general planning practice until the sustainability movement had gained widespread recognition and support. While the use of the “Geddesian tradition” (Chabard 2010) has often been used by planners attempting to redefine the field, the application of Geddes’ work to regionalism, especially in the 1940s, bears weight due to his championship of what he termed the “Regional Survey.”

> “The Greek City was at first merely the cultural centre of the rural life of the City State; and the Roman ‘Civitas,’ despite the excessive metropolitanism of Rome, was not just the municipal area but included the rural region together with the town, the ‘Pagus’ as well as the ‘Municipium.’ ”Our returning concept of the Region, and our pleas for a Regional Survey and a Regional Service, are thus but renewals of an ancient past... “The Regional Survey brings about the reunion of town and country and enables us to see that their activities are normally in harmony. Town and country can then again be considered together as City Regions, each occupying a definite geographical area.” (Tyrwhitt 1947)

The fear of the “German dream of a predominant World State,” which follows the preceding passage, never came to full fruition; and, although a world-wide government is a more feasible reality today, the “single central metropolis” sounds far-fetched in modern geopolitics. It does, however, seem that the economic forces of agglomeration will continue to draw populations out of the hinterlands and into urban or suburban settings. Today in regional planning, there is a significant focus on the interplay of global and local forces; and, while he may not be the true originator of the comparison or its use, the legacy of Patrick Geddes clearly lives on in the topic. Subsequently, other authors, whose legacy is similarly present in regional planning, revisited and expounded on ideas found in Geddes work.

In his introduction to *Patrick Geddes in India*, Lewis Mumford spoke highly of Geddes, saying that “The life and work of Patrick Geddes prefigure the age in which we now live. The tasks that he undertook as a solitary thinker and planner have become the collective task of our generation.” (Tyrwhitt 1947) Mumford went on to identify Geddes’ work and comprehensive regional surveys as highly influential on planning practice in his day, encouraging the consideration of the natural, built, and social environments that planners found themselves working in (Tyrwhitt 1947). Continuing, Mumford’s 1946 introduction identified Geddes and regionalism as a key to proper planning. More than 70 years later, the interaction of global and local scales are still a topic of discussion among modern planners.

> "Until Geddes applied his sociological insight and his biological knowledge to the region, regionalism was an archaic and backward looking movement, following the patter of nationalism, and paying more attention to a static and isolationist conception of the local community than to a dynamic one which placed the region in the midst of the currents of modern civilization. Just because Geddes respected the old roots of regional culture, he had no interest in limiting its expression to some historic moment of the past: if the roots were alive, they would keep on putting forth new shoots, and it was in the new shoots that he was interested.
>
> “If one part of Geddes’s thought and activity was attached to the region, indeed, to the village or the hamlet, another part was attached to the whole planet and to humanity. Geddes was a global thinker in practice, a whole generation or more before the Western democracies fought a global war. In short, one cannot appreciate Geddes’s regionalism unless one also appreciates his internationalism, his universal-ism (Tyrwhitt 1947).”

In today’s highly polarized political climate, this balancing of past and present, near and far, seems to be especially pertinent. Geddes’ balancing philosophy is also pertinent for local municipalities which find themselves competing in an increasingly globalized economy.

2.1.2 Jean Gottman’s Megalopolis

Although the term “megalopolis” was originated by Geddes in *Cities in Evolution* (1915) and used by Lewis Mumford in *The Culture of Cities* (1938), the term was popularized by the French Geographer Jean Gottman in his *Megalopolis: The Urbanized Northeastern Seaboard of the United States* (1961), which is still frequently referenced today. Robert Lang and Dawn Dhavale at the Metropolitan Institute at Virginia Tech based their megaregional work on many of the concepts explored by Gottman.

> “Gottmann’s original study … held that the region was unique in several ways, including its

> large size and commercial inventiveness. By the time Gottmann"revisited" the megalopolis in the late 1980s, he acknowledged that several other US regions could qualify as Megapolitan. He noted especially the cases of the Midwest and West Coast, but also saw a nascent megalopolis forming in the South around Atlanta (R. E. Lang and Dhavale 2005)."

In the 1987 review of his work referenced by Lang and Dhavale, Gottmann said that...

> "...the Megapolitan concept seems to have popularized the idea that the modern cities are better reviewed not in isolation, as centers of a restricted area only, but rather as parts of 'city-systems,' as participants in urban networks revolving in widening orbits."

Between 2010 and 2020, Google Scholar results for "megapolitan" were ~5,700 (~4,500 since 2015; 9.15M in a general search) while searching "megaregion" resulted in ~4,500 (~3,100 since 2015; 1.94K in a general search).

These high level results indicate that the term "megaregion" has yet to permeate society as generally as "megapolitan" has. An exhaustive study of whether this is reflective of a meaningful advantage to the concept of megapolitans over megaregions or simply an artifact of time and language is beyond the scope of this work; however, there are indications in the literature reviewed here that they represent morphologically distinct ideas and should be treated as related and not exclusive of each other.

2.1.3 Modern Views of Megaregions

Since early 2000s, there have been a number of definitions of megaregions propagating throughout planning-related literature, see Table 2.1. The megaregions resulting from the work by the Regional Plan Association (RPA) and by Dr. Catherine Ross have been particularly influential and seem to be considered the default by many academic and mainstream articles. For example, a 2012 presentation given at the Volpe National Transportation Systems Center titled "The Challenge of Transportation Planning for Megaregions"[1] gave the definition of megaregions delineated by the Georgia Tech Center for Quality Growth and Regional Development (CQGRD) *and* visually illustrated US megaregions using the 11 RPA America 2050 megaregions. Additionally, a 2017 USDOT-FHWA case study of the Arizona Sun Corridor referenced the image seen in Figure 2.13, which is based on work by Dr. Catherine Ross and CQGRD, as well as megaregional work by the RPA.

Ross et al. (2008) has characterized megaregions as linked networks of metropolitan centers and their surrounding areas that share or interact through environmental, economic, infrastructure and social factors

[1] https://www.volpe.dot.gov/sites/volpe.dot.gov/files/docs/The_Challenge_of_Transportation_Planning_for_Megaregions.pdf

Table 2.1: Sample definitions of 'Megaregion' from Literature

Year	Author	Definition
2006	Regional Plan Association	"...areas that comprise multiple, adjacent metropolitan areas that are connected by commuting patterns, business travel, environmental landscapes and watersheds, linked economies, and social networks [@RegionalPlanAssociation2006]."
2008	Florida et al.	"an integrated set of cities and their surrounding suburban hinterland across which labour and capital can be reallocated at a very low cost [@Florida2008]."
2009	Ross et al.	"...networks of metropolitan centres and their surrounding areas ... spatially and functionally linked through environmental, economic, and infrastructure interactions [@Ross2011]."
2012	Volpe National Transportation Systems Center	"Large networks of metropolitan centers and surrounding areas connected thru cultural, environmental, economic characteristics as well as infrastructure."
2014	UN-Habitat	"...several cities integrated with each other within the orbit of the overall region, surpassing mega- or meta-cities in terms of population, economic output and that further combine large markets, skilled labour and innovation [@UNHabitat2013]."
2017	USDOT-FHWA	"...areas that share infrastructure, economic linkages, environmental systems, topography, and culture [@MaricopaAssociationofGovernments2017]."
2018	Glocker (OECD)	"...an integrated system of cities and their surrounding region, which one can visit within a day using ground transport [@Glocker2018]."

(Ross 2011).

> "The fragmented political structure in which transportation planning occurs in the United States was at one time feasible because the urban problem was more localized in nature. However, due to population and economic growth, urban area expansion and increased relationships between urban areas and regions which are supported by progressive economic, communication and infrastructure connections, planners are faced with addressing problems that are system-related and thus cannot be spatially constrained to the political boundaries of a city, a county, or even to a single state (Ross 2011)."

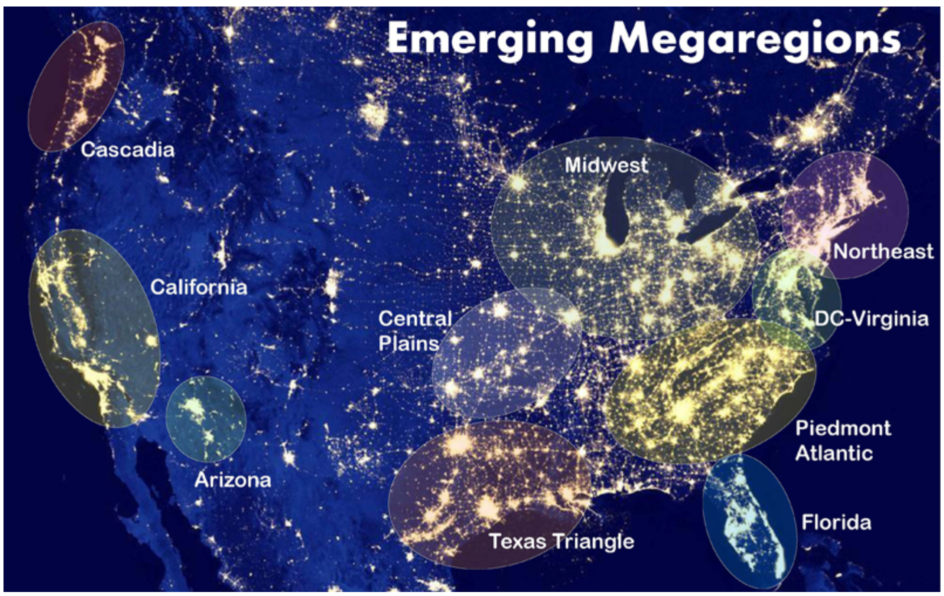

Figure 2.1: Federal Highway Administration (2016). Megaregions and Multi-Jurisdictional Planning. Washington, DC. Retrieved from http://www.fhwa.dot.gov/planning/megaregions/case_studies/, last accessed September 3, 2020.

In concert with the delineation of megaregions, a significant portion of US megaregions research seeks to illuminate the benefits, especially the economic, of megaregions, be that through freight and distribution, routing improvements, or policy coordination (Pain 2017). However, the voices against the development of governmental structures on this scale are frequently those who would be implementing these structures on a day-to-day, year-to-year basis (Pain 2017). Therefore, these concerns must be integrated into research and subsequent recommendations if the benefits of megaregions are to be realized without creating unnecessary or even detrimental bureaucratic bloat.

The FHWA 2017 report "Multimodal Planning at the Megaregional Scale" highlighted several key aspects of planning practice in relation to megaregions (Read et al. 2017). Despite a general awareness of megaregions, this scale of planning was rarely incorporated into local and regional plans. When megaregions *were* included, the primary focus was on transportation systems and infrastructure. This was compounded by several factors: the lack of federal and state legislation or structures to support megaregional planning, the difficulty in planning beyond jurisdictional boundaries, and the absence of practical guidance on how to incorporate megaregional context, trends, and issues (Read et al. 2017).

A number of case studies such as the one performed by Georg et al. on the Boston-Washington region have explored the nature of these large-scale geographies for a variety of purposes from morphological interest to freight policy application (Georg, Blaschke, and Taubenböck 2018). Their findings have reflected the "fuzziness" inherent in describing a multi-faceted urban fabric with boundaries and borders. While this may be a necessary acknowledgment for analysis and governance, it does lead to a discussion first and foremost of how megaregions are delineated. A common set of parameters and methodologies has yet to be identified. Often, the selection process depends on the funding source, the scope of the work, and datasets available to the researchers.

There is also little legislation to reference as significant steps to create functional megaregions have only entered the world stage in the last decade or so. In fact, this absence is core to the purpose of this work. Without clear administrative limits, powers to incentivize or enforce, or patent federal *and* state support, regional government may struggle to achieve its full potential and instead add weight to an already bureaucratic system. Delineating megaregions is how the theories and promises of regionalism are translated into legislation, organizations, and real-world impacts.

2.2 Identifying Regions and Megaregions

2.2.1 Differentiating Megaregions

In 2009, a staff report titled "Defining U.S. Megaregions" detailed the process used by the Regional Plan Association (RPA) to delineate megaregions as part of the America 2050 project, setting the stage for the following decade of discussion on megaregions in the United States. Five criteria quantitatively scored counties (whether they were within a core based statistical area (CBSA), their population density, projected population growth rate, population density growth rate, projected employment growth rate); then, in-house planners at the RPA used qualitative analysis to determine the score cut-off for inclusion in a megaregion and to refine boundaries based on local contexts. This process identified the eleven megaregions displayed in Figure 2.2.

Still, the report acknowledges the need for further research and refinement as the model performed well

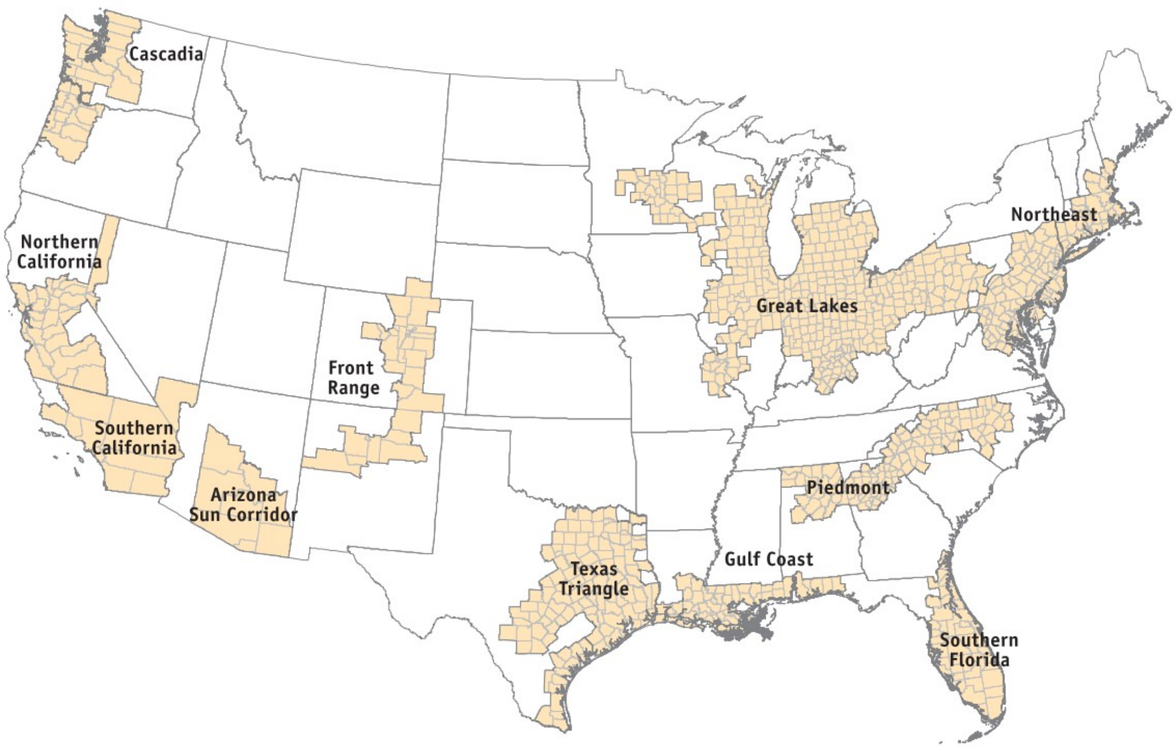

Figure 2.2: Regional Plan Association megaregions identified during the America 2050 project in 2009 (Hagler 2009).

identifying fast-growing, dense population centers but struggled with sparse, slow growing regions (Hagler 2009). This led to questions about connectivity between metropolitan areas like Pittsburgh and Philadelphia, Nashville and Memphis, and Salt Lake City and Denver. Other methods that were not integrated into the process but were discussed by the report include business flow analysis, intercity passenger and freight movement, natural systems, transportation systems, and cultural and economic connections.

As mentioned previously, Lang and Dhavale identified 10 “megapolitan” areas in a 2005 study that utilized the megapolitan geography popularized by Jean Gottman (R. E. Lang and Dhavale 2005). Their criteria were:

- Combines at least two, but may include dozens of existing metropolitan areas.
- Totals more than 10,000,000 projected residents by 2040.
- Derives from contiguous metropolitan and micropolitan areas.
- Constitutes an “organic” cultural region with a distinct history and identity.
- Occupies a roughly similar physical environment.
- Links large centers through major transportation infrastructure.
- Forms a functional urban network via goods and service flows.
- Creates a usable geography that is suitable for large-scale regional planning.
- Lies within the United States.
- Consists of counties as the most basic unit.

The resulting areas accounted for more than two thirds of US residents, see Figure 2.3. An interesting continuum defined by this study was the megapolitan spatial form from galactic to corridor. This spatial interpretation may be worth exploring with regard to natural environments and geography (R. E. Lang and Dhavale 2005). Later in the context of presentation to the Washington APA chapter on the Cascadia Megapolitan Area and high speed rail lines[2], Lang clarified some differences between megapolitans and megaregions.

> “Megapolitans are mostly continuous urban corridors and are best served by”Regional High Speed Rail“. Megaregions are proximate but discrete urban complexes that need”Express High-Speed Rail" to bridge the gaps between multiple megapolitan areas.
>
> In megapolitans, trains mostly compete with autos and should make multiple stops at key centers along the route. In megaregions, trains mostly compete with short-haul air service and should make very few stops in order to maintain maximum speed."

[2] https://www.washington-apa.org/assets/docs/the_cascadia_megapolitan_area.pdf

These differences imply an argument for spatially contiguous megapolitans and economically or socially (but not necessarily spatially) contiguous megaregions, supporting the idea of a super-metropolitan societal and administrative geographies. Although outside the scope of this work, it is worth noting that more linear and focused super-metropolitan forms could be a highly effective scale for specific situations.

Figure 2.3: Lang & Dhavale 2005, Map 2: The Megapolitans (R. E. Lang and Dhavale 2005)

Five years later in their book *Megapolitans and Megaregions: The Emergence of Large-Scale American Urban Systems*, Lang and Nelson proposed an updated version of US Megapolitan Areas, see Figure 2.4.

In their OECD paper for delineating megaregions, Glocker identified two major approaches: morphological and functional-network.

> "The morphological approach identifies megaregions based on continuous urban settlement areas that reach certain thresholds of density, dimension or degree of urbanisation. The underlying idea of this approach is that contiguous development results from a functioning as a megaregion. Thus, if multiple urban centres become integrated to the point where their labour markets and local supply chains overlap, the space between them tends to fill up with lower density development.
>
> "The functional or network approach defines a megaregion as an area of interactions between actors that can go in multiple directions and on several interconnected multiple layers. Identifying complex structures requires information on flows between the different parts of the megaregion. Such information can help capture material or immaterial flows. Material flows are directly observable and can be measured such as commuting flows or commodity flows. Immaterial flows

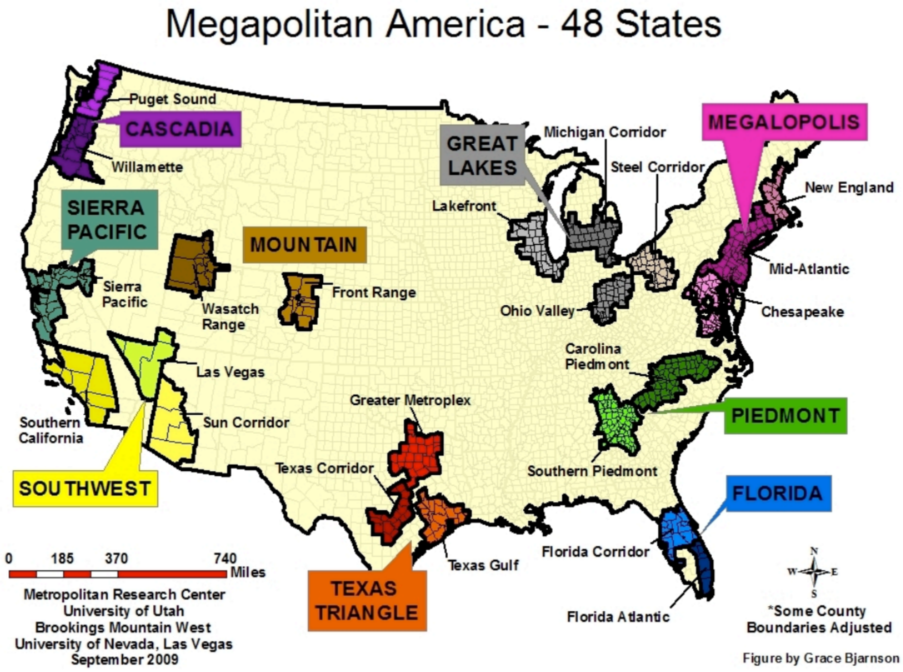

Figure 2.4: Lang & Nelson 2010, Megapolitan Areas (R. Lang and Nelson 2011)

include observable ones, such as email and telephone exchange, as well as non-observable ones such as knowledge flows (Glocker 2018)."

This distinction clarifies the back and forth between megaregions and megapolitans and will inform the remainder of this review of previous attempts at megaregional identification. A later review of datasets used confirms, Glocker's note that functional data is used less than morphological due to data availability at the local government scale. The remainder of this section addresses specific uses of various factors to identify megaregions and then posits possible datasets that could be used to supplement these more prominent methods.

2.2.2 Population

Despite disparities in the size of US counties (both in population and area), raw population is a prominent aspect of regional analysis. For example, the RPA's megaregion identification process relied heavily on population, population density, and population growth rate as three of their indicators (Hagler 2009). One of the greatest challenges in using population as a direct measure is that the centroid of an administrative boundary rarely aligns with the population-weighed centroid of the same geography. Other geographies are not without their flaws when measuring; for example, census tract boundaries have a much more consistent size in terms of population, but they often follow roads, bifurcating groups that likely share demographic characteristics. Thus population can form a strong foundation for community identification, but its limitations necessitate the use of addition data sources.

2.2.3 Proximity

In their OECD working paper on megaregions, Glocker developed a methodology based in network theory that characterized megaregions with spatial city locations and infrastructural link distances. This was accomplished through a population-weighted mean-shift clustering algorithm with a 300 kilometer search radius and resulted in 25 global megaregions of 10 million urban inhabitants or more (Glocker 2018). In the US, many of these megaregions had significant overlap with the Regional Plan Association megaregions, see Figure 2.5. This similarity to earlier methods indicates that algorithmic methods have the potential to compete with quantitative-qualitative analysis while requiring less data.

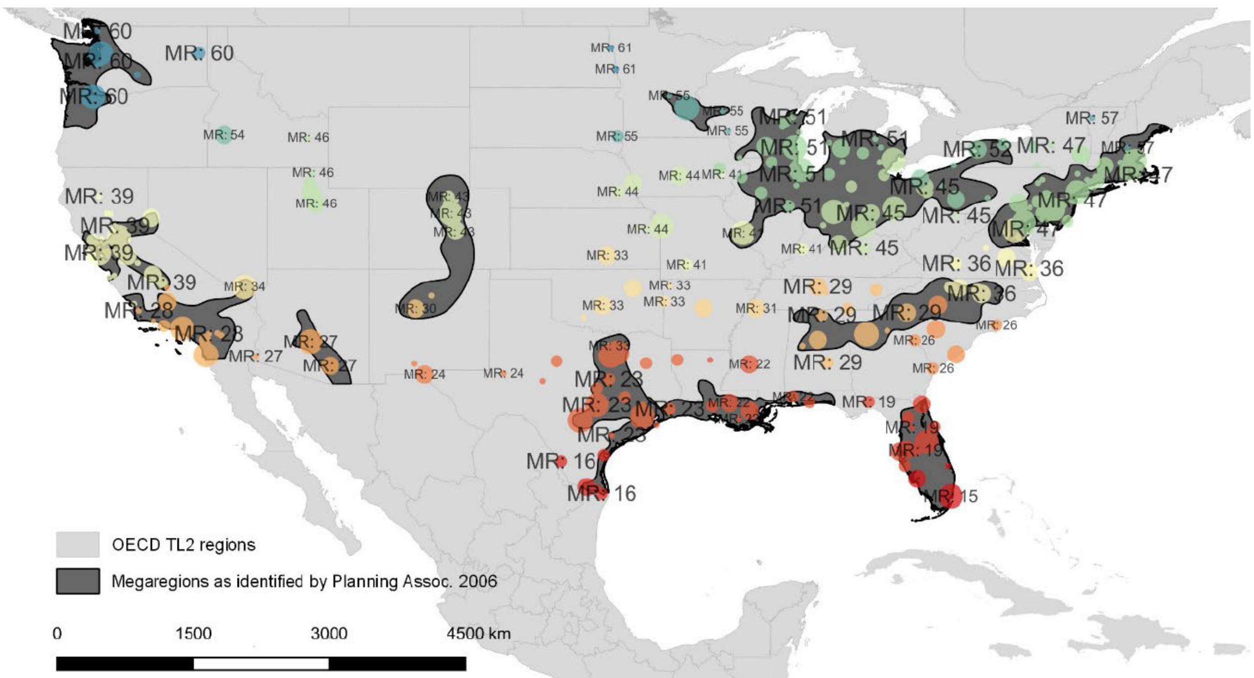

Figure 2.5: Comparison of OECD and RPA megaregions in the US (Glocker 2018)

Another method for measuring proximity and, by extension, contiguity has been light production or night-time light (NTL). In their paper on sustainability at megaregional scales, Marull et al. used data from the National Geophysical Data Center of NOAA across 4 sample years (1992, 2001, 2007, 2009) to measure the evolution of European megaregions (defined as “a contiguous lighted area with more than one major city of metropolitan region that also produces more than $100 billion in Light-based Regional Product” (Marull et al. 2013)). However, they qualify this delineation method as imperfect due to a non-exact relationship between luminosity and urbanization. Despite this, Marull et al. did find a significant connection.

> “Results show significant differences to a greater increase in GDPpc in the regions that were included in a megaregion in 1992, compared with those becoming part of a megaregion in 2001 or those not belonging to a megaregion in any of the periods analyzed. Results also show that the regions with low and medium levels of urban land show benefits in terms of GDPpc growth for belonging to a mega-region that are statistically significant. Given that a megaregion is a polycentric agglomeration of cities and its less dense hinterlands, these results could imply that the regions that benefits more from the formation of megaregions (in terms of GDPpc growth) would be the peripheral areas (Marull et al. 2013).”

In assessing the “Boswash” urban corridor, Georg et al. utilized several remote imagery based data sources including Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) “stable lights from 2010,” “man-made urban structures with a vertical extent,” artificial surface layers, and impervious surfaces (Georg, Blaschke, and Taubenböck 2018). These layers were then associated with population, income, and transportation infrastructure for validation purposes. Thresholds were set to produce spatially contiguous areas and overlaid on each other to highlight areas of agreement between datasets.

> "The spatial delimitation of regions is challenging—clear, crisp borders do not exist in complex real-world landscapes, and assumed or accepted boundaries are malleable due to conceptual logic, datasets or thresholds applied. Conceptual complexity leads to a struggle to construct regions in a consistent territorial layout.

> "The likelihood that a particular part of the Earth’s surface belongs to an urban corridor (or any other concept for a constructed territorial space) is determined by the amount of layers which yield positive results.

> “Our fuzzy delimitation of Boswash complements existing maps through the use of diverse input data and variables. We believe that our method to describe the area as a single connected area is mathematically reasonable and thus objective overall. Interactive GIS applications may even allow planners to overcome the binary view while performing queries like ‘show all raster cells which yield at least seven positive scores’ (Georg, Blaschke, and Taubenböck 2018).”

Nevertheless, earth observation (EO) data such as the NTL data used by Marull et. al or the DMSP-OLS are common in morphologically focused studies of megaregions and their development due to the dataset’s worldwide nature of the data and its availability over a number of years. Taubenböck and Wiesner analyzed spatial settlement patterns in five megaregions commonly agreed upon by literature using radar satellite data from the *Global Urban Footprint* from the German Aerospace Center (Taubenböck and Wiesner 2015). They

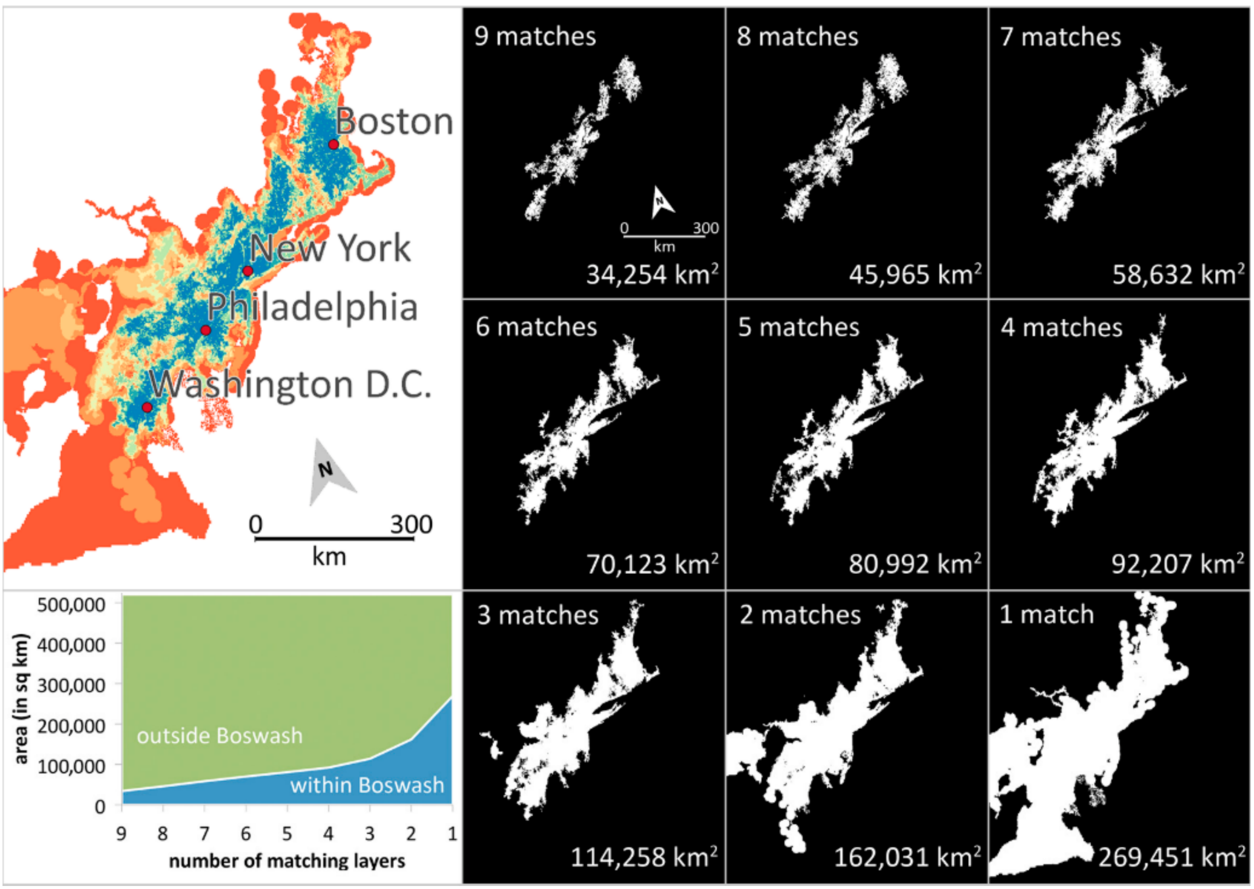

Figure 2.6: Probability-based Spatial Delimitation of the Boswash Urban Corridor. (Georg, Blaschke, and Taubenböck 2018)

found that settlement patterns did not follow similar across megaregions given the wide variation in spatial connectivity and monocentric-polycentric structure.

In a 2012 dissertation from Texas A&M University, Youngho Ko's morphological study of the RPA's 11 megaregions felt that the geographies did not share a connected form that could be identified through the non-functional relationships that were apparent in "2000 total population, 2000 population density, 2001 impervious land cover, and 2000 nighttime light emissions" (Ko 2012). Morphological megaregions either failed to emerge or were significantly different from the provided boundaries in Ko's methodology as human behaviors extended only to a certain degree (e.g. no one would commute six hours to work). Still, Ko does note that economic and functional interconnections do exist in polycentric urban structures.

> "The analysis results found that global scale spatial distributions of morphological characteristics had been inconsistently concentrated and clustered in high density subareas of each U.S. megaregion. The morphological clustering representation of the 11 U.S. megaregions, as the final result, concluded that the morphologically identified U.S. megaregions were not the same as the current U.S. megaregions. This study asks the urban and regional planning profession to balance the perspective between functional relationships and morphological characteristics in identifying U.S. megaregions (Ko 2012)."

2.2.4 Economics

Core Based Statistical Areas (CBSA) are built from an urban core with at least 10,000 people and any adjacent counties with a "high degree of social and economic[3] integration with the core as measured through commuting ties with the counties associated with the core";[4] their purpose being a consistent set of geographies for reporting data and statistics. Combined Statistical Areas are comprised of adjacent CBSAs with "substantial employment interchange"[5]. As will be discussed in the subsection on commuting, there are limits to associating social and economic connections between two geographies with their combined work-related travel.

2.2.5 Communication Data

Zhang et al. developed a complex weighted stochastic model using machine learning and an aggregated commuting network built from mobile phone signal data; while this particular study replicated an existing structure (the commute-shed), it demonstrates the power of mobile device data to describe complex patterns. This model was then used to measure spatial mesoscale structures in the Pearl River Delta and Guangdong-

[3]Bloomberg article on how megaregions are powering the world economy. https://www.bloomberg.com/news/articles/2019-02-28/mapping-the-mega-regions-powering-the-world-s-economy

[4]https://www.census.gov/topics/housing/housing-patterns/about/core-based-statistical-areas.html

[5]https://www.census.gov/topics/housing/housing-patterns/about/core-based-statistical-areas.html

Hong Kong-Macao Greater Bay megaregions in China (Zhang et al. 2020).

The leveraging of cellular and telecommunications data is already being used to develop better theories of community and social structure. However, this data is rarely publicly available, and utilizing private data to develop public governmental structures might be politically unpopular and costly to implement well.

2.2.6 Transportation

2.2.6.1 Infrastructure

A prime example of a functional relationship between regions is the transportation network. Within and between cities, roads, railroads, and waterways span administrative boundaries that—in the current paradigm—require multi-party coordination when large issues must be addressed. Population growth and increased freight traffic over the next several decades will continue to deteriorate the United States infrastructural assets. Megaregional planning may have the tools and scope needed to mitigate or alleviate these problems as capital investment can be comprehensive than piecemeal improvements performed by individual municipalities (Ross 2011).

Megaregions are also well positioned to address the user end of transportation problems through the coordination of public, private, and non-profit actors. Travel Demand Management (TDM) is already a function of many MPOs and other regional groups, but these efforts lack the ability to reach a majority of individuals who interact with their systems.

> "Although many actors including public, private, non-profit and civic organizations affect the prosperity of regions, there is neither a popular incentive nor mandate for these players to form an alliance. In such an alliance they could work together to achieve smart growth, manage climate change mitigation and adaptation strategies, and implement economic development. All of these activities are closely related to transportation infrastructure systems. Instead of cooperating, such actors often conflict and compete against each other within the same regions (Ross 2011)."

2.2.6.2 Commuting

Counter to the practice of utilizing commuter data in megaregion identification, Lang and Dhavale contend that. . .

> "A direct functional relationship as indicated by commuting does not exist at the Megapolitan scale (RPA 1967). The area is simply too big to make daily trips possible between distant sections. But commuting is just one—albeit key—way to show regional cohesion. Other integrating forces

exist such as goods movement, business linkages, cultural commonality, and physical environment (R. E. Lang and Dhavale 2005)."

This inability to identify megaregions through daily driver habits is reflected in Ko's dismissal of morphological megaregions. Commuter patterns *have* been successfully used to identify urban regions since the mid-twentieth century, but more recent research has identified complexities of non-core-centric movements between suburbs and hinterlands (He et al. 2019).

Nelson and Rae used 4,000,000 commuter flows at the census tract level to compare a visual heuristic and an algorithmic method for megaregional identification. While the visual exercise was limited to two case studies in California and Minnesota, the algorithmic method was used to analyze the entire continental US, see Figures @ref(fig.Commuter-regions) and @ref(fig.commuter-edges). The commuter regions were created with the "Combo" package developed at MIT's Senseable City Lab with the goal of modularity-based high-accuracy partitioning (Nelson and Rae 2016).

Figure 2.7: US regions based on commuting patterns (Nelson and Rae 2016)

He et al. developed a weighted-network of 2010 US Census Local Origin Destination Employment Statistics (LODES) data to identify county level commuter clusters. The analysis of the network with 3,091 nodes and

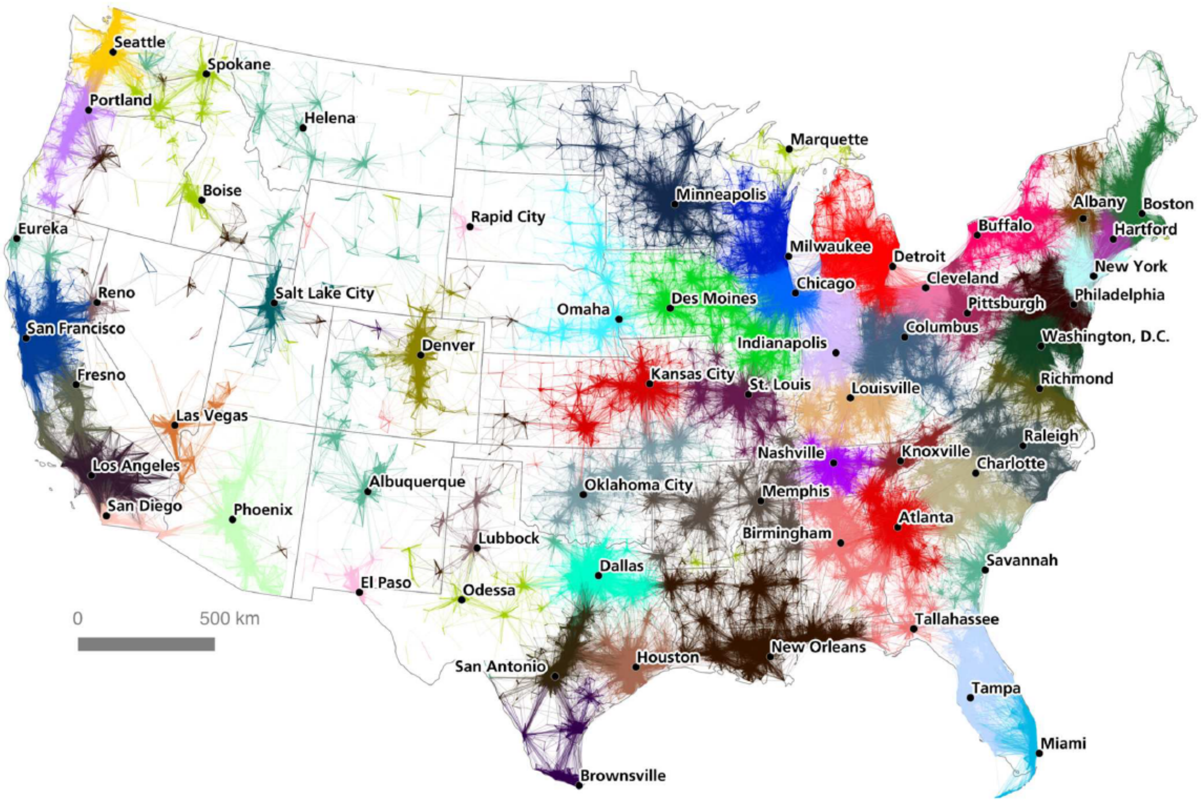

Figure 2.8: Partitioned commuter edges in the US (Nelson and Rae 2016)

17,222 edges produced 182 communities, see @ref(fig.he2019-comparison), that covered 90% of the commuters and of the general population (He et al. 2019). These clusters varied widely in size (average size of 49 counties with a standard deviation of 78) and had significant amounts of overlap. The researchers felt that allowing for communities to belong to multiple clusters prevents the creation of "…institutional structures and policies that are tailored not only to singular geographical entities, but to multitudinous, interacting identities that space and place assume (He et al. 2019)."

In comparing the resultant cluster sets with existing, large-scale delineations, He et al. found that

> "…the clusters found by our proposed method are much larger than existing delineations and account for much more of the inter-county commuting activity. We define the coverage rate as the number of commuters between all edges of the given sets of clusters divided by the sum of the total edges. The criterion captures how much of the total commuting activity is 'captured' by different modes of aggregation. The rate of coverage for all inter-county commuting was 86% for all clusters, compared to only 48% from OMB delineated MSAs and 77% from megaregions. The coverage rate for all within county, or same-county, commuting was 92% from clusters, compared to 86% from MSAs and 74% from megaregions." (He et al. 2019)

This comparison, while an interesting exercise, does appear to be comparing clusters derived not only from disparate datasets but also with different intentions. Although they can capture some aspects of inter-community relationships, commuting patterns do not inherently represent the socioeconomic networks between communities that are at the core of megaregions' usefulness.

> "Traditional delineations of geographic regions have relied on agglomerations of smaller geographies, historical and political boundaries, separating edges and central foci. The boundary characterization is important not only for scientific purposes of tracking and tracing historical evolution of urban systems but also for administrative purposes of allocating infrastructure investments and formulating economic development strategies. For example, boundaries of metropolitan areas in the United States are artifacts of delineation definitions, yet are central to tracking demographic and economic changes, funding allocations, determination of fair market rents, housing subsidies that depend on area median income and a host of other federal and state programs, even when the agencies caution their use for non-statistical purposes. These delineations are central but invisible to the lives of many." (He et al. 2019)

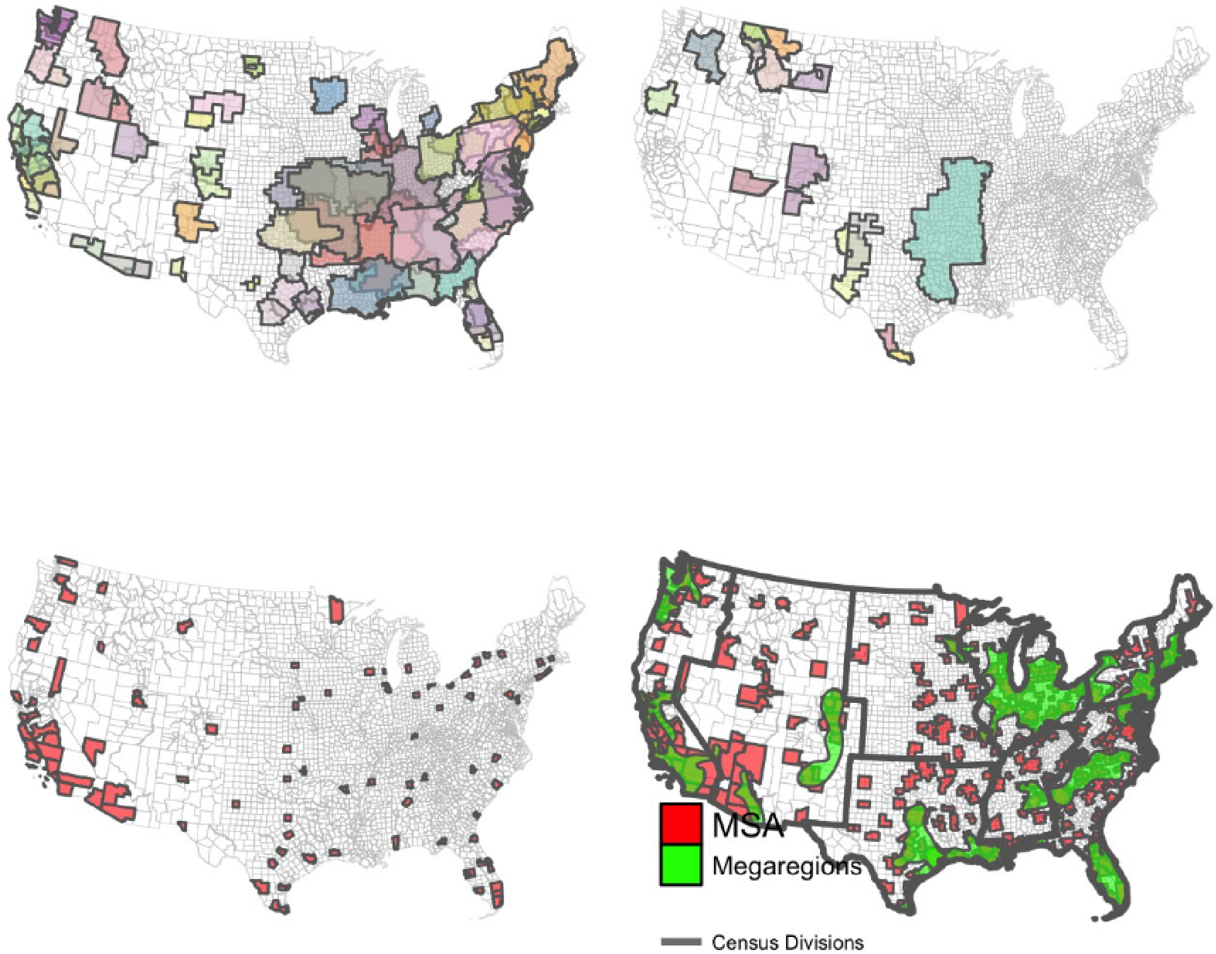

Figure 2.9: Non-nodal communities (top left), nodal communities (top right), and monads (bottom left), compared to MSAs (OMB) and megaregions (RPA) (bottom right) (He et al. 2019).

2.2.6.3 Freight

> "The economic foundation of cities is trade." - *Jane Jacobs, The Death and Life of Great American Cities*

Although the idea of the megaregion had not emerged during Jacob's writing years, her premise is still salient. Megaregional trade networks rely of a system of freight and telecommunication infrastructure, low barriers to trade, and social networks (Stich, Griffith, and Webb 2015). These networks in turn fuel the network's constituent metropolitans. Megaregions contain the overwhelming majority of US economic markets, and thus investment in these trade networks is key to continued economic health and growth. As success is spread throughout a megaregion's interconnected sub-markets, quality of life can improve even in communities that have little direct interaction with freight traffic.

> The 100 largest metropolitan areas in the US contain two-thirds of the country's population, generate 75 percent of its economic output and are its centers of advanced manufacturing, innovation, human capital, and technology (Tomer et al., 2013). Freight transportation investments are often considered to lead to higher levels of economic development and employment. In a 2012 national freight planning survey of Metropolitan Planning Organizations (MPOs), 87 percent of MPOs identified economic development as a primary motivation for conducting regional freight planning. However, freight volumes by themselves do not necessarily translate into related concentrated activity due to pass through freight volumes (Stich, Griffith, and Webb 2015).

Unfortunately, the increasing complexity of global trade is beyond the capacity of local municipal leaders to address, both in expertise and available time.

> Metropolitan leaders, particularly those in areas that make up megaregions, are unable to fully understand their role in domestic and global trade networks (US Department of Commerce, 2013). Therefore, the economic development strategies often in use by metro areas are disparate and inefficient, to the detriment of America's economic competitiveness and growth potential (Stich, Griffith, and Webb 2015).

In a report by Texas A&M supported by the USDOT, a study of freight patterns in the gulf coast highlights the incredible scale at which commodities move on a daily basis.

> "...more than 77 percent of commodities from megaregions were moved to domestic destinations by truck in 2002, and its portion in megaregions is projected to 80 percent in 2035, while non-megaregion areas rely less than 60 percent on truck in both 2002 and 2035 (Table 1). This means

> that megaregions will experience heavier freight traffic on highways compared to non-megaregion areas. Only 4-5 percent of commodities are carried by rail in megaregions, compared to 13 percent of rail usage in non- megaregion areas. Pipeline is frequently used in non-megaregion areas (approximately 26 percent in 2002) when compared to 4 percent in megaregions (Stich, Griffith, and Webb 2015)."

> "The average distance covered by truck freight is shorter (485 miles) than air (973 miles), rail (902 miles), and coastwise water (1,269 miles). Moreover, more than 65 percent of the tonnage of truck freight movements is estimated to move less than 100 miles (Puentes, 2008). The relatively short length of trucking implies that the freight movement policy between metropolitan areas at the megaregion level would be useful in relieving congestion caused by truck traffic on highways and ensuring just-in-time delivery of goods (Cortright, 2006) (Stich, Griffith, and Webb 2015)."

Stich et al. cite Detroit as an example of a disconnect between freight and economic development (Stich, Griffith, and Webb 2015). Two major freight gateways are located in close proximity to the region, yet little of the benefits are captured while negative externalities like pollution are dumped into the community. If regional economic developers were able to address this issue, the Detroit region would be able to tap into huge amounts of economic activity. By leveraging an entire megaregion's worth of economic actors, this could be chance for the community to regain some of its lost momentum.

The Arizona Sun Corridor study referenced previously identified three primary layers, along with a number of data sources, of industry and freight to model.

- "The financial layer, as represented by a firm synthesis model that looks at production, consumption, the evolution of firms (businesses), and the various factors that influence the birth, growth, location, and dissolution of firms.
- "The logistics layer, as represented by a supplier selection model that determines the transfer of goods between entities. It is this layer that generates the"buyers" and "suppliers" of all shipments.
- "The physical transportation layer, as represented by two models: a supply chain model that describes how goods are moved from origin to destination and a truck tour model that captures any touring behavior (i.e., making multiple deliveries in a single day) by trucks in the region." (Maricopa Association of Governments et al. 2017)

The team concluded with several areas where the study and resultant tools would benefit governance in the region.

> "This new model will assist… planners in coordinating policy development. It will also be an

effective tool for improving freight operations across the Sun Corridor megaregion. In addition, this model provides planners in the region with a new tool to better understand different transportation planning and policy scenarios. As the model is used and refinement continues, the project team hopes that future integration with the regional economic models will help PAG, MAG, and ADOT understand how freight movement interacts with other transportation factors and economic indicators in supporting the region's economic development (Maricopa Association of Governments et al. 2017)."

2.3 Data Sources for Megaregional Delineation

Although the the research works covered here is not an exhaustive list, they either include, reference, or build upon prominent papers on the topic. This section summarizes the common data sources and methodologies used in these works and then outlines trends and types of data that are utilized in the methodology.

2.3.1 Summary of Previous Data Criteria & Sources

Lang and Dhavale (R. E. Lang and Dhavale 2005):

- Combines at least two, but may include dozens of existing metropolitan areas
- Totals more than 10,000,000 projected residents by 2040
- Derives from contiguous metropolitan and micropolitan areas
- Constitutes an "organic" cultural region with a distinct history and identity
- Occupies a roughly similar physical environment
- Links large centers through major transportation infrastructure
- Forms a functional urban network via goods and service flows
- Creates a usable geography that is suitable for large-scale regional planning
- Lies within the United States
- Consists of counties as the most basic unit

Florida et al. (2008):

- 30 arc-second Light Intensity from the Earth Observation Program of NOAA's National Geophysical Data Center
- Estimated Light-based Regional Product (in same terms as GDP)
- Raw Population
- Patents from the US Patent and Trademark Office and the World Intellectual Property Office

- Highly cited scientific authors as a stand-in for scientific research activity

RPA (Hagler 2009):

- Member of a CBSA
- A population density greater than 200 residents per square mile
- A population density growth that met the following criteria:
  - The rate exceeded 0.15
  - The total growth by 2025 was more than 1,000 residents
  - The population density would increase by more than 50 people per square mile by 2025
- Employment growth that met the following criteria
  - The rate exceeded 0.15
  - The total growth by 2025 was more than 20,000 jobs

Youngho Ko (Ko 2012):

- 2000 Total Population
- 2000 Population Density
- 2001 Impervious Land Cover
- 2000 Nighttime Light Emissions"

Marull et al. (Marull et al. 2013):

- National Geophysical Data Center of NOAA across 4 sample years (1992, 2001, 2007, 2009) to measure the "contiguous lighted area with more than one major city of metropolitan region that also produces more than $100 billion in Light-based Regional Product."

UN-Habitat (2013):

- Population
- Economic Output
- Combining Large Markets
- Skilled Labor and Innovation
- Amalgamating Several cities

Taubenböck and Wiesner (Taubenböck and Wiesner 2015):

- Natural Earth data "urban hubs" greater than 100,000 residents
- Demographic information from LandScan data

- Radar satellite data from the *Global Urban Footprint* from the German Aerospace Center Landsat program

Nelson and Rae (Nelson and Rae 2016): - 4,000,000 commuter flows at the census tract level - "Combo" package developed at MIT's Senseable City Lab with the goal of modularity-based high-accuracy partitioning.

The Arizona Sun Corridor Study (Maricopa Association of Governments et al. 2017):

- Input/Output Tables, U.S. Bureau of Economic Analysis
- County Business Patterns, U.S. Census Bureau
- National establishment Time-Series, Walls & Associates
- Longitudinal Business Dynamics, U.S. Census Bureau
- Annual Survey of Manufacturers, U.S. Census Bureau
- Business Dynamics Statistics, Data Source U.S. Census Bureau
- Business Employment Dynamics, U.S. Bureau of Labor Statistics
- Statistics of U.S. Businesses, U.S. Census Bureau
- Non-employer Statistics, U.S. Census Bureau
- Commodity Flow Survey, U.S. Census Bureau
- Freight Analysis Framework, Federal Highway Administration (FHWA)
- TRANSEARCH, IHS, Inc., Global Insight Carload
- Carload Waybill Sample, Surface Transportation Board
- Air Carrier Statistics, U.S. Department of Transportation (DOT) Office of the Secretary (OST) – Research
- Trans-border Freight Database, US DOT OST – Research
- Port Import/Export Reporting Service, IHS, Inc. Port/Import Export Reporting Service
- National Highway Planning Network, FHWA
- Center for Transportation Analysis Railroad Network, Oak Ridge National Laboratory
- Vehicle Inventory and Use Survey, U.S. Census Bureau
- Vehicle Travel Information System, FHWA
- Truck GPS Data, American Transportation Research Institute, StreetLight Data
- National Performance Management Research Dataset, FHWA

Georg et al. (Georg, Blaschke, and Taubenböck 2018):

- Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) "stable lights from 2010"

- “Man-made urban structures with a vertical extent”
- Artificial Surface Layers
- Impervious Surfaces
- Layers were then associated with population, income, and transportation infrastructure for validation purposes
- Thresholds were set to produce spatially contiguous areas and overlaid on each other to highlight areas of agreement between datasets.

Glocker (Glocker 2018):

- Population-weighted mean-shift clustering algorithm with a 300 kilometer search radius

He et al. (He et al. 2019):

- 2010 US Census Local Origin Destination Employment Statistics (LODES) data

Chen et al. (2020):

- “GDP, population, employment, etc.” from the China City Statistical Yearbook 2015
- Daily intercity bus schedules to represent road passenger flows

2.3.2 Example Delineation Methods from Literature

As mentioned previously, the RPA’s America 2050 megaregions rely heavily on a small set of indicators and a qualitative process. While the indicators are important aspects of meta-communities and the qualitative process was carried out by a highly trained staff, this process is subject to many of the same abstract issues found in the gerrymandering of political boundaries, e.g. subjectivity in the wrong place can separate communities that should be considered jointly.

Another prominent process, Essentialist-Relational Approach, is described by Ross and Woo[6] as a three stage process of identifying core areas, identifying areas of influence, and considering regional characteristics. Core areas use agglomerations of population and employment, economic interactions of commodities, capital, and industry, societal travel patterns, infrastructure networks, proximity, and historical and cultural relationships. Hub and core areas were identified using Moran’s I measure of spatial autocorrelation and then interaction between these centers were measured with graph theory and Markov chains. Areas of influence relied on factors including natural, physical, and political environments as well as socioeconomic population characteristics. Candidate areas were identified using geographically weighted regression and socioeconomic characteristics were integrated using multivariate analysis and other mapping methods.

[6] Identifying Megaregions

More recently, the work done by Glocker (2018) is extremely informative and in line with other megaregional efforts by simply weighting population centers by proximity and size. This process is concerned with cities rather than counties or their international equivalents; therefore, the search radius of 300 kilometers takes on a different meaning and cannot be directly applied. However, the author does wish to express their sincerest appreciation for Glocker's willingness to share information and methodology during the initial research process for this work.

Other evaluations include Chen et al.'s 2020 discussion on China's megaregions and the relative lack of polycentrism, which was insightful and may be used in the future to help evaluate megaregions sets for the US, and the implementation of community detection through spatial networks by authors such as He et al. (He et al. 2019), where graph theory is used to group node-edge representations of communities.

Together, these and other megaregional delineation methods seek to identify communities that are both geographically proximate and characteristically similar. As community identification problems continue to be addressed in this field as well as others such as computer science and economics, there should be a regular review of applicable methods to ensure that megaregions encompass relevant communities as accurately and precisely as possible.

2.3.3 Proposed Data Sources

A set of data sources was selected from those used or referenced in previous works. The intent during the selection process was simply to reflect best practices on the topic while identifying a potential minimum set of sources that would sufficiently describe the multi-faceted nature of society previously discussed. The methodology chapter lays out the theories behind the source categorization and proposes evaluatory measures for the addition of new datasets. An explicit methodology was not implemented for the inclusion of initial datasets. Instead, an attempt was made to provide data coverage for the types of data sources that were repeatedly utilized in the existing body of research, laying a foundation on scholarly work in the hopes of making progress towards administrative application. The later refinement of the data source set is laid out in the methodology chapter and utilizes comparitive metrics from cluster analysis to identify data sources which result in more efficient computation.

List of Data Sources Utilized:

- Commodity Flow Survey
- County Business Patterns
- Healthcare Access

- Household Income (ACS S2503)
- IRS SOI Migration
- Koppen Climate Classifications
- Language Spoken (ACS S1601)
- LEHD LODES
- Medicare Reimbursement
- Metropolitan Planning Organizations
- National Land Cover Database (2016)
- Nativity & Citizenship (ACS B05001)
- Patent Creation
- Population (ACS B01003)
- Presidential Race Votes
- Proximity to Post-secondary Schools[7]
- Race & Ethnicity (ACS B02001)
- US Counties

2.3.3.1 Administrative Boundaries

Two administrative areas are utilized in this work: counties (with county equivalents) and Metropolitan Planning Organizations. Counties were frequently used in literature as the basic building blocks of megaregions, and a wide variety of data sources are available at the county level. MPOs are the primary regional government body in the US; thus, integrating their characteristics and spatial distribution into the delineation process may benefit initial coordination as megaregional organizations are established.

2.3.3.2 Culture and Demographics

Demographic data constitutes an important aspect of communities and is often used to identify sub-populations that are disadvantaged or vulnerable to shocks from economic changes or environmental shifts. Research with transparent methodology rarely explores this topic and its interactions with their work, but demographics are often referenced broadly by method-opaque works. These community data points are nonetheless vital to equitable and effective governance. For example, an understanding of a region's racial make-up and distribution sets the stage for addressing the challenges presented by institutional racism, acknowledging voter ideology may help highlight opportunities for bipartisan collaboration at key geographic junctions,

[7] See "Expanding College Access:The Impact of New Universities on Local Enrollment" by Patrick Lapid at UC Berkley (2017); "Student Choice of College: How Far Do Students Go For An Education" by Mattern and Wyatt in the Journal of College Admission (2009); and "Education Deserts: The continued Significance of 'Place' in the Twenty-First Century" by Hillman and Weichman at University of Wisconsin-Madison (2016).

and identifying common education hurdles may result in the more efficient removal of roadblocks preventing community growth.

Population

Population was one of the most common factors utilized in the megaregional methods reviewed. For this work, population and population density were included together as a single dataset during the delineation steps.

Cultural Heritage

Three datasets were used to capture various facets of communities cultural heritage: race and ethnicity, language spoken at home, and nativity and citizenship. Often a member of a minority population in one of these datasets is a minority in the others; nevertheless, there are important distinction possible when they are each considered individually. Race was pulled from ACS Table B02001 with categories of “Asian,” “Black,” “Multiple Races,” “Native American,” “Other,” “Pacific Islands,” and “White.” Areas where one race is predominant will likely have different interests to those communities that are more integrated, and, given cultural factors, these enclaves may be more likely in certain regions. Nativity and citizenship data came from ACS Table B05001 and had categories based on US citizenship and whether the individual was born in the US or abroad. This information helps differentiate ethnic communities that may appear homogeneous within the racial data. Language spoken at home was retrieved from ACS Table S1601 with categories of “Only English,” “Spanish,” “Indo-european,” “Asian-Pacific,” and “Other Languages”; while language use may often overlap with race, ethnicity, or foreign origin, communities’ ability to communicate or lack thereof can form natural boundaries between groups just as easily as race or immigrant origin.

Voting Trends

A nuanced approach to ideology would integrate detailed questions about topics ranging from economics to morality to the role of government. However, initial research into data availability on the subject did not identify any detailed datasets. Therefore, presidential election voting results serve as a generalized stand-in for ideology. Data was retrieved from MIT’s Election Lab for county vote tallies in the 2000-2016 presidential elections for the Democratic, Green, Republican, and other parties. The Election Lab also offered data on House and Senate races, but these were by Congressional district and would require additional analysis to confirm the reliability of a crosswalk from district to county.

Post-Secondary Education Proximity

Especially in households with heavy financial burdens, proximity to post-secondary education is an important factor in predicting whether students will attend a given location. Mattern and Wyatt found that higher SAT scores, parental education levels, and parental income were highly correlated with distance students traveled to attend a post-secondary school (Mattern and Wyatt 2009). Researchers at the University of Wisconsin-Madison note that there are significant portions of the United States that could be considered "education deserts"(Hillman and Weichman 2016) due to the lack of higher education in the area. Together, these two ideas point towards a spatial component of post-secondary education that may influence the needs of communities that megaregional government may be able to support. To create a spatial variable to describe this phenomena, average distance to all US post-secondary schools on file with the Education Department was calculated for each county and then divided by that county's average distance to all other counties in the contiguous United States.

2.3.3.3 Environmental

Several works in megaregional research reference how physical environments, built and natural, shape the societies within them. Therefore, two general environment datasets are included here to cover such variation. Climate and environment commonalities between regions may provide better focus and direction for tackling climate-change related issues facing communities today and in the years to come.

Koppen Climate Classifications

The Köppen Climate Classification separates environments by trends in seasonal temperature and precipitation variation. With 30 categories, the system is able to capture the general differences between locals while concisely describing similar areas. A moderate resolution raster was polygonized, intersected with county geometries, and the resulting area ratios summarized for each county GEOID.

National Land Cover Database

Where the Köppen system is concerned exclusively with natural climate, the NLCD raster shows the distribution of both natural and built environments. The classification categories are water, developed, barren, forest, shrubland, herbaceous, cultivated, and wetlands[8]. This allows for the differentiation of average community density and vegetation composition at the county level. The zonal histogram tool in QGIS was used to create summary statistics for each county from the NLCD 2016 data.

[8] https://www.mrlc.gov/data/legends/national-land-cover-database-2016-nlcd2016-legend

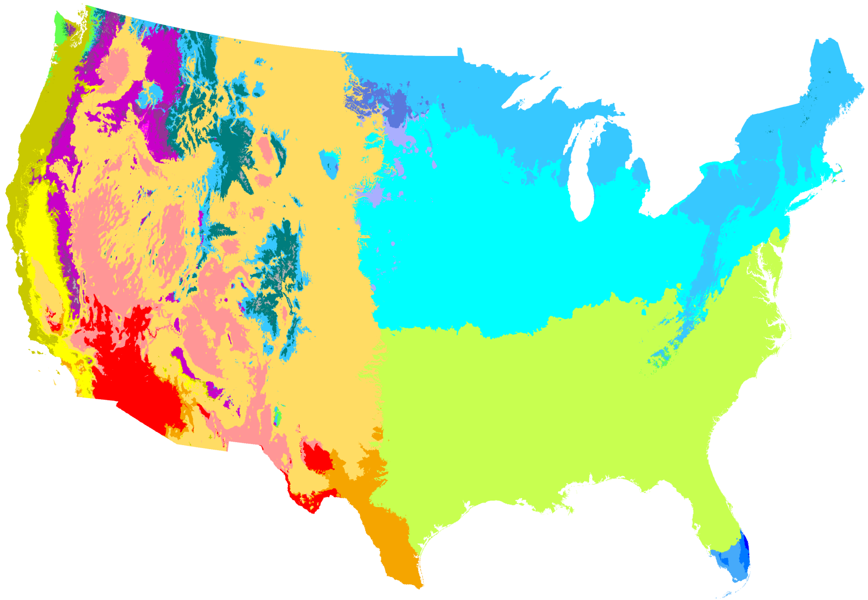

Figure 2.10: Map of Köppen Climate Classifications in the contiguous United States.

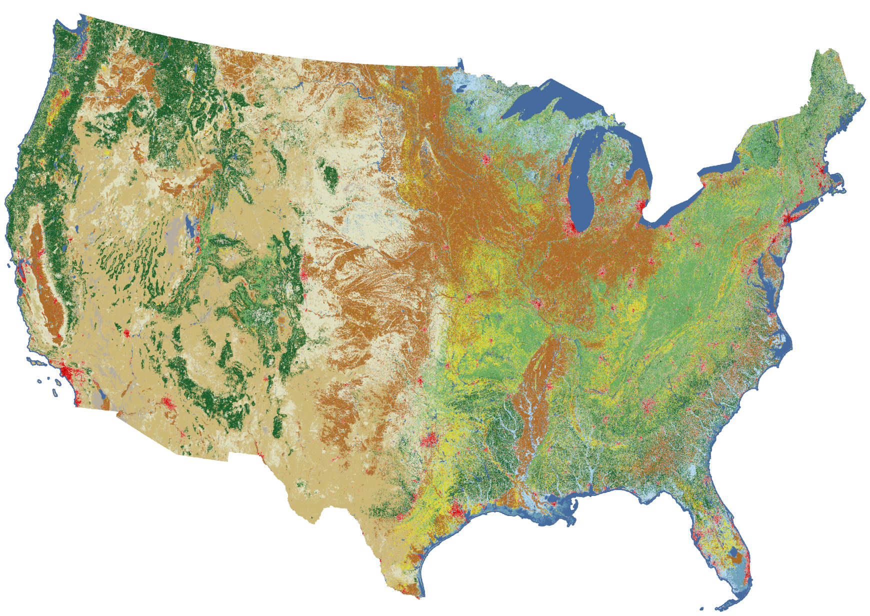

Figure 2.11: Map of land cover types in the National Land Cover Database.

Medicare Reimbursements

The social norms, environmental conditions, and personal behaviors of a population have been shown to have real and significant impacts of personal and public health. If these patterns can be shown to be spatially correlated, they may produce regions that can be used to streamline health outreach and education efforts. As the public health field has expanded its focus from treating individuals' illnesses to include preventing disease and promoting health on a societal scale, there has been a concurrent shift in opinion on how involved community planners should be in addressing the impact of the built environment on their communities (Kent and Thompson 2014). Unfortunately, the inter-related nature of public health and the built environment makes the identification of causal connections nearly impossible. On top of the challenge in designing experiments, health data is often obscured by privacy laws.

Medicare reimbursements at the county level morphologically reflect the functional data which forms the basis of Hospital Service Areas and Referral Regions. This data stand-in is used primarily to circumvent the low availability of public health related flow data, and it is provided by the same organization which provides the main body of research on these regions, the Dartmouth Institute for Health Policy & Clinical Practice.

2.3.3.4 Macro-Economics

Financial ties between communities were frequently cited in literature as a key functional relationship in megaregions. These connections act at greater physical distances than trends like daily commuting and redistribute wealth among non-base sectors of local economies. In analyzing the economy and freight patterns of their identified Gulf Coast Megaregion, Stich et al. utilized a capital overview, census-based labor overview, and location quotient to planning for the increase in freight volumes that are to be expected by 2035 (Stich, Griffith, and Webb 2015). This is an excellent example of the type of locally relevant research that can benefit entire megaregions, eliminating the need for individual municipalities and regions to conduct duplicate research while still providing plans and strategies that are meaningful and tailored to local needs. Two datasets were used to identify flows and static characteristics related to industry: the Commodity Flow Survey and County Business Patterns.

Commodity Flow Survey

The current challenge with utilizing data products such as the Commodity Flow Survey (CFS) and the Freight Analysis Framework (FAF) for local or regional analysis is their level of aggregation: state and metropolitan area. To quickly address this for use in this work's methodology, CFS data was allocated to counties proportionally according to CBP employment in the shipment's NAICS code across relevant counties.

Although the dataset used contains several variables related to each shipment's such as weight, distance traveled, and mode, this work simply considers the value of the shipments. This factor is the most directly indicative of economic flows and likely reflect other aspects mentioned above. Future research could apply all available values to the process to see if the resulting clusters vary to any significant degree.

CFS Freight-Value Flow Calculation:

1. Filter observations for the contiguous United States.
2. select origin, destination, NAICS code, and shipment value for each observation.
3. Sum shipment values by origin-destination-NAICS combinations.
4. Set up a square matrix for all county OD pairs in the contiguous US.
5. For each observation…
    1. Identify origin and destination counties associated with the CFS Area.
    2. Create vectors of employment in counties for the observation's NAICS code, divided by their sum, for origins and destination.
    3. Build a rectangular matrix from the two vectors by multiplying them together with all values summing to one.
    4. Multiply the matrix by the observation shipment value.
    5. Add entries to the appropriate OD pair in the distance matrix.
6. Entries in the OD matrix now represent the proportional commodity value shipped along the respective "edge."

Sources:

- 2017 Commodity Flow Survey Datasets
- CFS Geographies
- About the CFS

County Business Patterns

The County Business Patterns files provide information on the employment, payroll, and number of establishments in each county by NAICS code from 2-digit to 6-digit levels of aggregation. In addition to being an extensive dataset on employment by industry, the dataset also allows for the proportional dis-aggregation of the CFS down to the county level. For the process in this work, employment by NAICS code is the variable considered, and 3-digit NAICS codes were used for the calculation of the distance matrix. Payrolls and establishment counts could also be analyzed and may correlate with factors in other datasets such as household median income.

Sources:

- County File Layout
- CBP 2018 Record Files

2.3.3.5 Household Finances

The distribution of wealth in the United States and its related inequality has come to the forefront of the public mind over the last year with vulnerable populations suffering disproportionately in the COVID-19 pandemic and the associated economic downturn. As urban populations continue to grow amid stagnant wages, financial topics such as affordable housing and housing affordability will become even more pressing issues for local governments across the country. As megaregional government is established, datasets on physical housing stock and the finances associated with it can be analyzed for spatial connections that could then be leveraged to fund projects more effectively and legislate policies that better encourage a distribution of housing types that meet community needs.

Household Incomes

For this work, median household incomes and the percentage of households in various income brackets within each county were retrieved from ACS Table S2503, which also includes information on housing costs and differences between owner and renter policies.

IRS SOI Migration

The SOI division of the IRS reports year-to-year migrations on a state-to-state and a county-to-county level based on the number of returns that shifted from location to location as well as the exemptions and total AGI associated with those returns. These flows approximate the long-term movements of households, populations, and economic buying power.

LEHD LODES Commutes

The LODES dataset is frequently utilized in megaregional and regional delineation research to identify commuter sheds. In this work, only the "S000" field (total jobs/commutes) is implemented as a variable, but the other variables which split out observations by industry, age, or income could easily be plugged into the process as well.

2.4 Regional Government in Practice

2.4.1 United States

The primary form of legislated regional government in the US is the Metropolitan Planning Organization (MPO). Introduced by the Federal-Aid Highway Act of 1962, MPOs are created in urbanized areas with over 50,000 residents. An MPO's primary functions are to develop transportation improvement programs, develop regional transportation plans, and distribute federal funding. After 200,000 residents, the area is also designated as a Transportation Management Area which adds further responsibilities. Metropolitan transportation planning processes are governed by Title 23 of U.S.C. §§ 134–135. As of 2020, there are 408 MPOs in the United States.

2.4.1.1 Legislation

Federal Aid Highway Act of 1962

> §134. Transportation planning in certain urban areas[9]
>
> "It is declared to be in the national interest to encourage and promote the development of transportation systems, embracing various modes of transport in a manner that will serve the States and local communities efficiently and effectively. To accomplish this objective the Secretary shall cooperate with the States, as authorized in this title, in the development of long-range highway plans and programs which are properly coordinated with plans for improvements in other affected forms of transportation and which are formulated with due consideration to their probable effect on the future development of urban areas of more than fifty thousand population. After July 1, 1965, the Secretary shall not approve under section 105 of this title any program for projects in any urban area of more than fifty thousand population unless he finds that such projects are based on a continuing comprehensive transportation planning process carried on cooperatively by States and local communities in conformance with the objectives stated in this section."

FHWA—UMTA Joint Regulations on Urban Transportation Planning

- Required as condition for continuing federal assistance the designation by the governor of a metropolitan planning organization (MPO) in each urban area (*Transportation Planning: A Decision-Oriented Approach* 2001).

[9] https://uscode.house.gov/statutes/pl/87/866.pdf

- MPO must develop a unified planning program and a prospectus of the planning process (*Transportation Planning: A Decision-Oriented Approach* 2001).
- Transportation plan must consist of a long-range element and a transportation system management (TSM) element (*Transportation Planning: A Decision-Oriented Approach* 2001).
- MPO must develop a transportation improvement program (TIP) and an annual element detailing the following year's projects (*Transportation Planning: A Decision-Oriented Approach* 2001).

Intermodal Surface Transportation Efficiency Act (ISTEA)

After 40 years of intermittent updates, ISTEA[10] legislated a major increase in responsibilities for and mandates for other planning organizations to interact with Metropolitan Planning Organizations and metropolitan planning in general. Subsequent major transportation bills have provided updates to the language and connections to new programs and organizations, but the core of modern metropolitan planning policy is still primarily consists of language from ISTEA.

Section 1006(a)(b)(2)(B) indicates that local officials in urbanized areas should work with the local MPO to designate areas for arterials and highways.

Section 1007(d)(3)(E) rewrites Section 133 of Title 23 allows for a state and MPO to proceed through certain portions of the funding process when in regard to urbanized areas over 200,000 residents.

Section 1017(c) required MPOs to participate in developing a "national list of rights-of-way" with total mileage, estimates of total costs, and strategies to prevent losses.

Section 1024 refers to metropolitan planning in general, and rewrites Section 134 of Title 23. Many of these amending statements further designate and elaborate on the form and function of Metropolitan Planning Organizations. The overarching rationale was laid out as follows.

> It is in the national interest to encourage and promote the development of transportation systems embracing various modes of transportation in a manner which will efficiently maximize mobility of people and goods within and through urbanized areas and minimize transportation-related fuel consumption and air pollution. To accomplish this objective, metropolitan planning organizations, in cooperation with the State, shall develop transportation plans and programs for urbanized areas of the State. Such plans and programs shall provide for the development of transportation facilities (including pedestrian walkways and bicycle transportation facilities) which will function

[10] https://www.congress.gov/bill/102nd-congress/house-bill/2950

> as an intermodal transportation system for the State, the metropolitan areas, and the Nation. The process for developing such plans and programs shall provide for consideration of all modes of transportation and shall be continuing, cooperative, and comprehensive to the degree appropriate, based on the complexity of the transportation problems.

In particular, this "3C process" of "continuing, cooperative, and comprehensive" would continue on throughout most subsequent federal transportation planning legislation and practice.

From a practical standpoint, MPOs are designated for urbanized areas with over 50,000 residents through an agreement between the state governor and affected local units of government (comprising at least 75% of the area's population) where the MPO includes local elected officials, transportation agency officials, and relevant state officials. This same agreement identifies the MPO boundaries which "shall cover at least the existing urbanized area and the contiguous area expected to become urbanized within the 20-year forecast period and may encompass the entire metropolitan statistical area or consolidated metropolitan statistical area. . . ." If there are areas of air quality non-attainment, these are included unless otherwise specified by the governor-governments agreement.

Although re-designation typically requires similar participation to initial designation procedures, there is a somewhat edge case where a group representing at least 25% of the area's population when the urbanized area's population is between five and ten million and when the area is also under extreme non-attainment for ozone or carbon monoxide under the Clean Air Act.

In the case of multi-state coordination or MPOs which cross state lines, Congress requires the executive branch to appropriately encourage the formation agreements to task MPOs with transportation planning coordination. Otherwise, the legislation provides the following language on multi-state compacts.

> "The consent of Congress is hereby given to any 2 or more States to enter into agreements or compacts, not in conflict with any law of the United States, for cooperative efforts and mutual assistance in support of activities authorized under this section as such activities pertain to interstate areas and localities within such States and to establish such agencies, joint or otherwise, as such States may deem desirable for making such agreements and compacts effective."

When coordinating between multiple MPOs, consulting the relevant MPOs and the state must be included for all plans and programs.

However, the legislation does note that the contained mandates are not to interfere with other public agencies which have similar responsibilities to develop plans and coordinate transportation services and projects. More than one MPO may be designated for large urban areas at the governor's discretion.

ISTEA also lists fifteen "factors to be considered" when developing plans or programs required by US Code.

1. Preservation of existing assets
2. Consistency with applicable energy conversation goals
3. Traffic congestion
4. Land-use and development plans
5. Transportation enhancement expenditures
6. The effect of all transportation projects regardless of funding source
7. International connections and sites
8. Connectivity between the metro and other areas outside it
9. Management-system-identified needs
10. Preservation of rights-of-way
11. Efficient freight movement
12. Designed life-cycle costs
13. The overall effects of transportation decisions
14. Expanding and enhancing transit service and usage
15. Capital investment in transit security

Metropolitan Planning Organizations are federally mandated to develop two documents on a recurring basis: the long range plan and the transportation improvement program (TIP). The long range plans must contain an identification of transportation facilities with regard to the "factors to be considered," a financial plan for implementation, an assessment of measures needed to ensure the preservation of the existing transportation system and make the best use of that system while reducing congestion and maximizing mobility, and a prioritization of transportation system enhancement activities. This process is to involve clean air agencies and provide reasonable opportunity for public input from individuals and organizations impacted by the plan. Then, the plan must be published to the public for review and to the governor for informational purposes. The TIP is a program updated every two years with approval of the MPO and governor. It identifies a prioritized list of projects for the three years following the TIP's adoption and a financial plan for implementation with recommendations for funding sources public and private. Unless otherwise noted, the project selection process that involve the federal government must be in conformance with the current TIP. Like the long range plan, this process must provide the public with notice and a chance to comment.

If an urbanized area's population is over 200,000, it is designated as a Transportation Management Area

(TMA). MPOs within a TMA must follow the same process of developing and implementing transportation plans and programs. They must also operate a congestion management system that utilizes travel demand reduction and operational mangement strategies for facilities eligible for federal funding according to the relevant legislation. Each MPO within a TMA must also be certified every three years on the condition that they are 1.) in compliance with applicable federal law and 2.) their TIP has been approved by the MPO and the governor. If an MPO is unable to certified for more than two consecutive years, 20% of their funding is withheld. For such areas that are also not in attainment for air quality, no federal funding may be used for highway projects that will increase single-occupancy vehicle capacity except when associated with congestion management. For non-TMA MPOs, abbreviated plans and programs may be approved as needed by the federal government so long as they are in air quality attainment.

With regard to statewide planning, Section 1025(e) and 1025(f)(1) states that long-range plans and transportation improvement plans (TIPs) should be developed in cooperation with the local MPO. For both these items, the state is expected to provide reasonable opportunity to comment to organizations or individuals who will be impacted. Additionally, Section 1033(g) notes that pedestrian and bicycle facilities should be located according to plans developed by MPOs and states and "provide due consideration for safety and contiguous routes."

On Native American Reservations, MPOs are listed in Section 1032(b)(j) as one of the governmental entities that might be included in the development of TIPs for these areas as tribal governments choose to engage in transportation planning.

Section 1034(a) and 1034(d) indicate that the regulation, development, establishment, and implementation of certain systems and issues must be done in cooperation with MPOs. These systems are highway pavement of federal-aid highways, bridges on and off federal-aid highways, highway safety, traffic congestion, public transportation facilities and equipment, and intermodal transportation facilities and systems.

Section 3012 reiterates much of Section 1024 for the purpose of amending the language into the Federal Transit Act's Section 8 Metropolitan Planning. This ensures a well integrated system for MPOs to coordinate transportation as well as transit.

Section 3013(h)(1)(A) dictates that the MPO approves the use of certain block grant program funds for highway projects in transportation management areas.

Dealing with the Office of Intermodalism, Section 5002(c)(4) requires the director to involve states and MPOs in the collection of an intermodal data base hosted in the Bureau of Transportation Statistics. Shortly after, Section 5002(c)(6) requires technical assistance to MPOs serving urban areas with a population over one million in the process of gathering intermodal data.

Transportation Equity Act for the 21st Century (TEA-21)

TEA-21[11] made superficial change regional transportation planning and MPO responsibilities such as updating language, relevant organizations, and grammar and stylistic aspects of relevant US code.

Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU)

SAFETY-LU[12] added greater coordination requirements for items such as surface transportation program funds. TIP requirements were modified and the programs were to be updated every 4 years, the same as the minimum for air quality conformity demonstrations, rather than every 2.

Moving Ahead for Progress in the 21st Century (MAP-21)

Starting with MAP-21[13], MPOs in areas either in non-attainment or maintenance for air quality with a population over 1 million people had to begin developing plans that described how the region would reach emission and traffic congestion reduction goals. These goals were then coordinated with states, public transit providers, and other MPO plans to ensure consistency across planning efforts in the area. Non-attainment and maintenance zones also were required to update these plans more frequently than other areas: every four years rather than every five. This environmental focus is also seen in the authorization to develop mitigation plans for transportation project's environmental impacts.

Performance management followed a similar process where RTPs and TIPs had to include metrics and targets that were performance-driven and outcome-based. The legislation also began requiring multiple scenarios for to be used modeling processes.

Fixing America's Surface Transportation (FAST) Act

The FAST[14] act's primary focus at the regional level was to expand scope. Transit and active transportation options were included in more breadth for the purposes of transportation plans and TIPs. Resiliency and reliability, from both a transportation systems viewpoint and a wide community perspective, became higher priorities; and the list of interested parties that MPOs must give opportunity for commenting on plans was expanded (*Transportation Planning: A Decision-Oriented Approach* 2001).

[11] https://www.congress.gov/bill/105th-congress/house-bill/2400/text
[12] https://www.congress.gov/bill/109th-congress/house-bill/3
[13] https://www.congress.gov/bill/112th-congress/house-bill/4348
[14] https://www.congress.gov/bill/114th-congress/house-bill/22

2.4.1.2 The Current Functioning of MPOs

In 1997, the Advisory Commission on Intergovernmental Relations (ACIR) condensed the modern responsibilities of Metropolitan Planning Organizations into five points (*Transportation Planning: A Decision-Oriented Approach* 2001).

1. "Establish and manage a level playing field for effective multimodal and intergovernmental decision making in the metropolitan area."
2. "Develop, adopt and update a long-range multi-modal transportation performance plan for the metropolitan area that focuses on three types of performance: mobility and access for people and goods, system operation and preservation, and quality of life."
3. "Develop and continuously pursue an appropriate analysis program to evaluate transportation alternatives and support metropolitan decision making, scaled to the size and complexity of the region and to the nature of its transportation issues and the realistically available options."
4. "Develop and systematically pursue a multi-faceted implementation program designed to reach all the metropolitan transportation plan goals, using a mix of spending, regulating, operating, management, and revenue enhancement tools."
5. "Develop and pursue an inclusive and proactive public involvement program designed to give the general public and all the significantly affected subgroups access to and important roles in the four essential functions listed above."

Metropolitan Planning Organizations achieve these goals by primarily acting as conveners and funders. Regional plans help direct local and state goals, and TIPs incentivize various priorities for capital investment in the region. However, MPOs have little ability to enforce policy or mandate processes.

At the present time, there are no standardized or federal tools for evaluating MPOs or other regional organizations as a whole. Performance measuring instead occurs with regard to specific metrics (such as air quality compliance) or internally by the organization.

Although some organizations that house MPOs do deal in other areas, these policy agendas are typically set by state legislatures or governors and function primarily in an advisory or coordinating role. Transportation is the only field where this form of regional government has means to incentivize compliance with their priorities.

Without overarching structure, many other fields (such as housing and land use) are fractal at the regional scale, every jurisdiction following different procedures and a pantheon of advocate organizations vying for limited funding across a wide breadth of financial institutions and requirements.

2.4.1.3 Other Regional Structures

Census Bureau & OBM Statistical Areas

The Census Bureau and Office of Budget and Management define several regional boundaries based on population and economic factors with a based unit of counties: Micro- and Metropolitan Statistical Areas (MSAs), Core-Based Statistical Areas (CBSAs), and Combined Statistical Areas.[15] Metropolitan and Micropolitan Statistical Areas are each a subclass of CBSAs. Metropolitan statistical areas contain a minimum of one urbanized area with 50,000 or more and nearby communities have a high degree of socioeconomic connection, measured by commuting ties. Micropolitans have the same kinds of commuter connections, but their cores have 10,000 to 50,000 residents. Meanwhile, CSAs are comprised of adjacent MSAs with substantial economic relationships. Their purpose is not to replace MSA identifications but rather to connect them in a way that quickly communicates the degree of their connections.

Hospital & Medicare Based Regions

The Dartmouth Atlas Project developed "hospital service areas" (HSAs) and "hospital referral regions" (HRRs), see Figure 2.12, from a combination of Medicare user and provide information from the Centers for Medicare and Medicaid Services alongside data from the US Census, American Hospital Association, the American Medical Association, and the National Center for Health Statistics. HSAs are built from ZIP codes, and identify where the majority of Medicare users received treatment within a hospital. HRRs are collections of HSA based on "the greatest proportion of major cardiovascular procedures were performed, with minor modifications to achieve geographic contiguity, a minimum population size of 120,000, and a high localization index."[16]

The medicare reimbursement dataset from the Dartmouth Institute for Health Policy & Clinical Practice is utilized in the methodology of this work; medicare reimbursements are closely related to the basis of HRRs and HSAs and are adjusted for age, sex, and race[17].

Labor Market Areas

The Bureau of Labor Statistics maintains a list of Labor Market Areas (LMAs), see Figure 2.13, to identify economically integrate zones where an individual can find both a residence and employment. These are

[15]https://www.census.gov/topics/housing/housing-patterns/about/core-based-statistical-areas.html
[16]https://www.dartmouthatlas.org/faq/
[17]https://atlasdata.dartmouth.edu/downloads/general

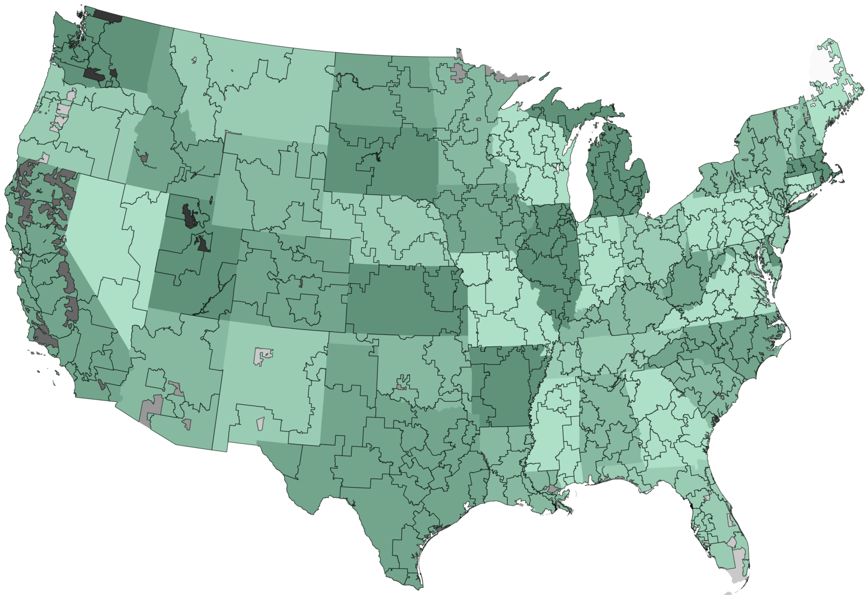

Figure 2.12: Dartmouth Institute for Health Policy & Clinical Practice (2020). Hospital Referral Regions. Retrieved from https://atlasdata.dartmouth.edu/downloads/supplemental, last accessed 6 April, 2021.

primarily composed of MSAs, but the bureau does define smaller labor markets under certain circumstances.[18] Individual counties were added to small LMAs if worker migration into or out of the county exceeded 25%, the commuting pattern is contiguous, and proximity to other small LMAs.[19]

Watersheds

Physical watersheds have been emulated in many flow-based regionalization techniques with commuting being a frequent example. The US Geologic Service is responsible for maintaining the maps and data associated with these natural regions. A map of the highest level of hydrologic unit regions is shown in Figure 2.14.

2.4.2 United Nations

Although the UN's sustainable Development Goals Report, 2019, gave "regional groups" that were primarily at the continental or sub-continental level[20] and based on their M49 standards rather than previous, and somewhat arbitrary, delineations of whether countries were "developed" or "developing," the World Cities Report is the primary location for references to "megaregions" in the body's literature. This regular report by the UN includes several references to "megacities," "megalopolis," and "megaregions." Included below are

[18] https://www.bls.gov/lau/laufaq.htm#Q06
[19] https://www.bls.gov/lau/laugeo.htm#geolma
[20] https://unstats.un.org/sdgs/report/2019/regional-groups/

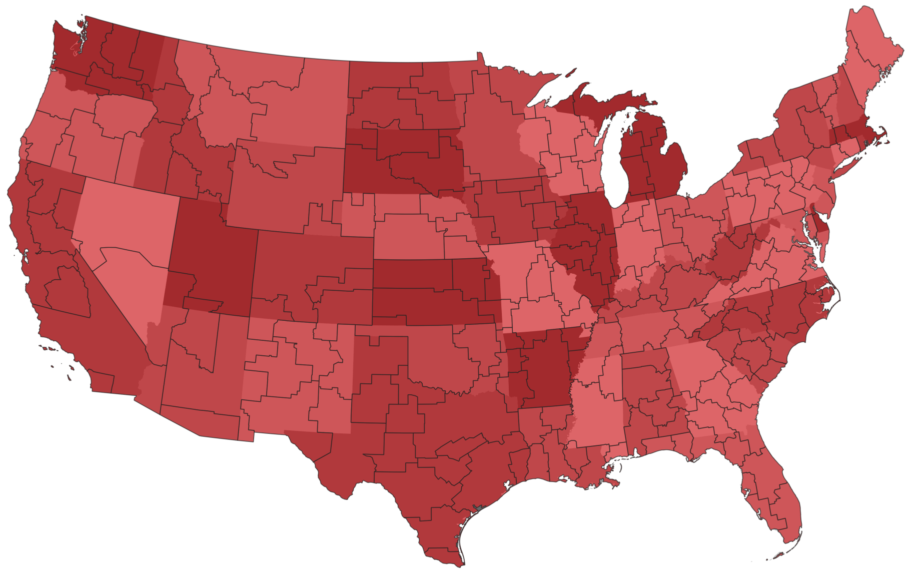

Figure 2.13: Pennsylvania State University (2010). Bureau of Economic Analysis Labor Markets. Retrieved from https://sites.psu.edu/psucz/data/, last accessed 6 April, 2021.

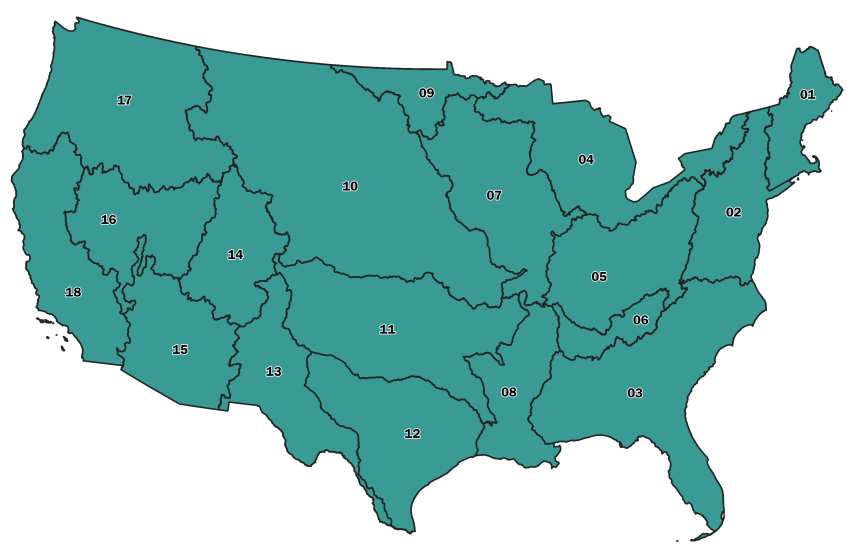

Figure 2.14: USGS HUC-2 Watersheds []

direct references to specific mega-urban-areas and then commentary on megaregions specifically. Megacities are defined here as metropolitan areas with 10 million people or more. Under this definition, 13% of the world's 2018 urban population lived in 33 megacities, increasing from 9% in 2000.

In the 2020 report, several specific examples were mentioned.

- US Northeast Megalopolis ("Boswash")
- Europe Rhine-Scheldt Delta Megaregion
- Pearl River Delta Megaregion
- Lond (megacity)
- New York City (megacity)
- Paris (megacity)
- Par-Am-Mun (megaregion from Paris to Munich via Amsterdam)
- Chi-Pitts (megaregion from Chicago to Pittsburgh)
- Greater Tokyo (megaregion)
- SoCal (megaregion from Los Angeles to San Diego)

Economic activity in concentrating in regions that encompass many local boundaries and cross international borders to the point that these regions' compete globally on the scale of nation-states. The interconnected geographies of megaregions account for nearly 40% of world GDP in 2015 with most existing in the already developed world, see Table 2.2, (United Nations 2020). The report goes on to warn that, as these regions compete with each other, the resulting productivity imbalance among regions combined with fragmented governance will exacerbate spatial and economic inequality between urban and rural communities. To help correct this trend, the report makes recommendations for developing nations; however, addressing government fragmentation and socioeconomic inequality would be beneficial regardless of a nation's status.

> In order to realize the economic value for sustainable urbanization and ensure inclusive prosperity, developing countries need strategies that ensure integrated spatial growth and development—to nurture nascent mega-regions within their territories, as well as those spanning neighbouring countries, so as to facilitate economic activities. For instance, it is vital to develop and implement national urban policies that maximize the benefits of urbanization and respond to these forms of interconnectivity and urban interdependence, as well as anticipating and managing the negative consequences of urban and regional growth [citation needed].

An endnote for this chapter notes that "Mega-regions are defined as 'areas of continuous light that contain at least two existing metro areas, have populations of five million or more, and generate economic output of

Table 2.2: From Table 3.3 of the UN World Cities Report 2020: 'Largest Mega-Regions in the World, 2015'

Megaregion	Cities	Output.Billions	Population.Millions
Bos-Wash	New York; Washington, D.C.; Boston	US$3,650	47.6
Par-Am-Mun	Paris, Amsterdam, Brussels, Munich	US$2,505	43.5
Chi-Pitts	Chicago, Detroit, Cleveland, Pittsburgh	US$2,130	32.9
Greater Tokyo	Tokyo	US$1,800	39.1
SoCal	Los Angeles, San Diego	US$1,424	22.0
Seoul-San	Seoul, Busan	US$1,325	35.5
Texas Triangle	Dallas, Houston, San Antonio, Austin	US$1,227	18.4
Beijing	Beijing, Tianjin	US$1,226	37.4
Lon-Leed-Chester	London, Leeds, Manchester	US$1,177	22.6
Hong-Shen	Hong Kong, Shenzhen	US$1,043	19.5
NorCal	San Francisco, San Jose	US$925	10.8
Shanghai	Shanghai, Hangzhou	US$892	24.2
Taipei	Taipei	US$827	16.7
São Paulo	São Paulo	US$780	33.5
Char-Lanta	Charlotte, Atlanta	US$656	10.5
Cascadia	Seattle, Portland	US$627	8.8
Ista-Burs	Istanbul, Bursa	US$626	14.8
Vienna-Budapest	Vienna, Budapest	US$555	12.8
Mexico City	Mexico City	US$524	24.5
Rome-Mil-Tur	Rome, Milan, Turin	US$513	13.8
Singa-Lumpur	Singapore, Kuala Lumpur	US$493	12.7
Cairo-Aviv	Cairo, Tel Aviv	US$472	19.8
So-Flo	Miami, Tampa	US$470	9.1
Abu-Dubai	Abu Dhabi, Dubai	US$431	5.0
Osaka-Nagoya	Osaka, Nagoya	US$424	9.1
Tor-Buff-Chester	Toronto, Buffalo, Rochester	US$424	8.5
Delhi-Lahore	New Delhi, Lahore	US$417	27.9
Barcelona-Lyon	Barcelona, Lyon	US$323	7.0
Shandong	Jinan, Zibo, Dongying	US$249	14.2

more than US$300 billion' (Florida, 2019)."

2.4.3 African Continent

Although Africa has yet to develop widely recognized megaregions, there are several sets of merging urban areas that will likely qualify in the near future: Johannesburg, Cairo, and Lagos to name a few.

2.4.3.1 The African Union

the Organisation of African Unity (OAU) was established in May 1963, consisting of 32 African nations, with the objective to unite the continent, improve quality of life, and eradicate the vestiges of colonialism. The African Union (AU) is the successor to the OAU and was launched in 2002 with 55 member states. This reorganization pivoted from the OAU's goals of de-colonialization towards intra-African cooperation and economic development. The AU's vision is "an integrated, prosperous, and peacful Africa, driven by its own citizens and representing a dynamic force in the global arena."[21]

Member states are grouped into five regions: north, west, south, east, and central, see Figure 2.15. The AU also subdivides its constituents into eight "Regional Economic Communities"[22] to encourage integration between members of the regions and in the African Economic Community, which aims to establish a common market in Africa.

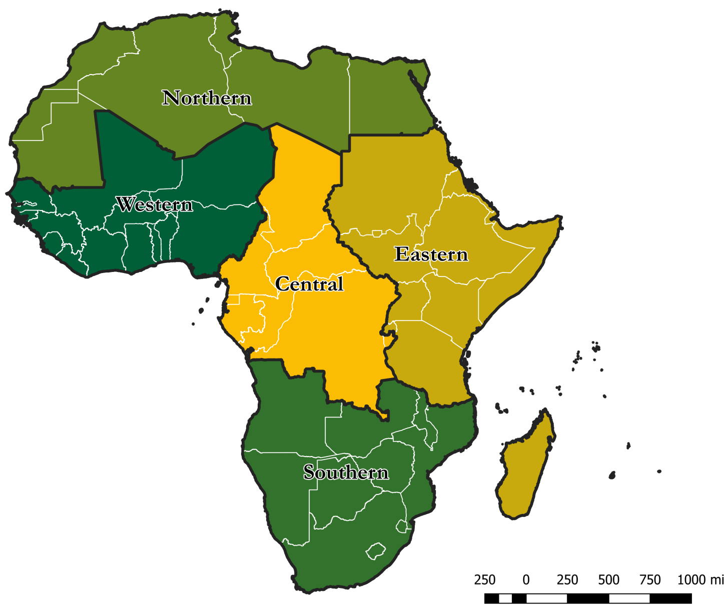

Figure 2.15: African Union Regions[23]

Although the Regional Economic Communities cover nearly all of continent, some nations are members of more than one organization. Kenya tops this list with membership in 4 RECs. Lists of member nations and a map of membership density (see Figure 2.16) are shown below. While these networks do not inherently indicate the presence of megaregions, their economic connectivity may provide the setting for their development in the future.

Regional Economic Communities:

- The Union of the Arab Maghreb[25]
  - *Morocco, Algeria, Tunisia, Libya, and Mauritania*

[21] https://au.int/en/overview
[22] https://au.int/en/recs
[23] https://au.int/en/member_states/countryprofiles2
[24] https://au.int/en/recs
[25] https://maghrebarabe.org/

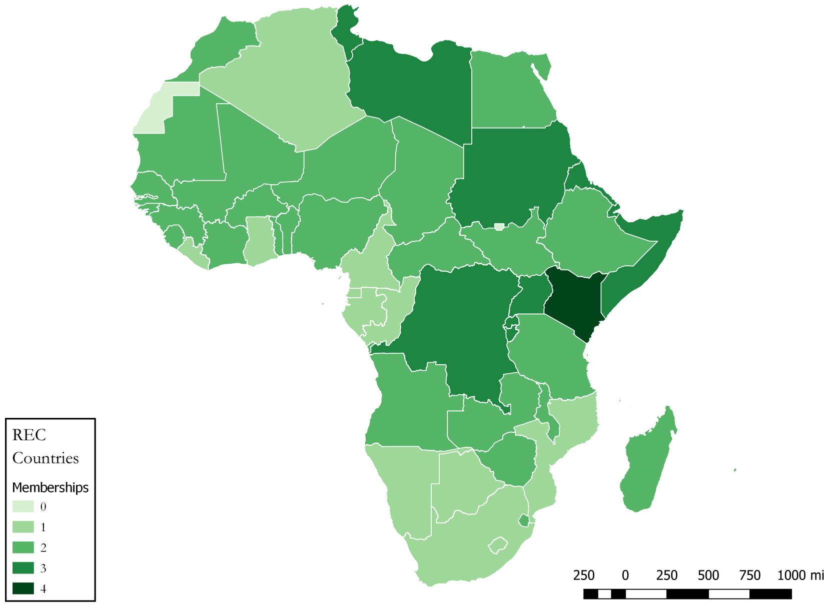

Figure 2.16: Regional Economic Community Membership[24]

- Common Market for Eastern and Southern Africa (COMESA) Secretariat[26]
  - *Burundi, Comoros, D.R. Congo, Djibouti, Egypt, Eritrea, Eswatini, Ethiopia, Kenya, Libya, Madagascar, Malawi, Mauritius, Rwanda, Seychelles, Somalia, Sudan, Tunisia, Uganda, Zambia, and Zimbabwe*
- Community of Sahel-Saharan States (CEN-SAD)[27]
  - *Benin, Burkina Faso, Cape Verde, Central African Republic, Chad, Comoros, Djibouti, Egypt, Eritrea, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Kenya, Liberia, Libya, Mali, Mauritania, Morocco, Niger, Nigeria, Sao Tome and Principe, Senegal, Sierra Leone, Somalia, Sudan, Togo, and Tunisia*
- East African Community (EAC)[28]
  - *Burundi, Kenya, Rwanda, South Sudan, Tanzania, and Uganda*
- Economic Community of Central African States (ECCAS)[29]
  - *Angola, Burundi, Cameroon, Central African Republic, Chad, D.R. Congo, Equatorial Guinea, Gabon, Republic of the Congo, Rwanda, and Sao Tome and Principe*
- Economic Community of West African States (ECOWAS)[30]

[26]https://www.comesa.int/
[27]https://au.int/en/recs/censad
[28]https://www.eac.int/
[29]https://au.int/en/recs/eccas
[30]http://www.ecowas.int/

  - *Benin, Burkina Faso, Cape Verde, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, and Togo*
- Intergovernmental Authority on Development (IGAD)[31]
  - *Djibouti, Eritrea, Ethiopia, Kenya, Somalia, South Sudan, Sudan and Uganda*
- Southern African Development Community (SADC)[32]
  - *Angola, Botswana, Comoros, D.R. Congo, Eswatini, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Nambia, Seychelles, South Africa, Tanzania, Zambia, and Zimbabwe*

2.4.3.2 Johannesburg and the Gauteng Region

Johannesburg lies in the Gauteng Provence of South Africa. The country's smallest and most densely populated, the province is estimated to be home to 15.5 million people within 18,182 sqkm in 2020. The rapid growth of the urban region led the University of Johannesburg, the University of Witwatersrand, Johannesburg, the Gauteng Provincial Government, and local governments from the province to establish the Gauteng city-Region Observatory (GRCO)[33] in 2008.

The city-region extends beyond the Gauteng provincial boundaries to include Pretoria, the country's administrative center, as well as several commercial and industrial cities. This combination of economic sectors accounted for approximately 45% of South Africa's total output. However, the region faces several challenges with high social and economic inequity being one of the most prominent as a remnant of the apartheid era.

2.4.4 China

With some of the largest cities in the, China was primed to develop meta-urban structures as its population, industrial capacity, and quality of life grew. In 2006, China's 11th Five-Year Plan added "city agglomeration" or "urban agglomeration" as a geographical planning unit, referencing "Beijing-Tianjin-Hebei," "Changjiang Delta," and "Zhujiang Delta" as existing agglomerations. It also indicates that regions in the initial stages of this development with focus on central cities to form megalopolises [citation needed].

> "The regions that have the conditions of urban agglomeration development shall strengthen unified planning and with megalopolis and megapolis as the leader, exert the functions of central cities and form several new city agglomerations with less land utilization, more employments, strong element concentration ability and rational population distribution."

[31] https://igad.int/
[32] https://www.sadc.int/
[33] https://gcro.ac.za/

In a morphological and function review of China's megaregions in 2020, Chen et al. analyzed the 20 megaregions approved by the State Council of China (with boundaries developed by other researchers due to the lack of official boundaries from the Chinese government).

Chinese Megaregions as of 2015:

- YRD: Yangtze River Delta
- PRD: Pearl River Delta
- BTH: Beijing–Tianjin–Hebei
- MYR: Middle Yangtze River
- CCQ: Chengdu–Chongqing
- LNP: Liaoning Peninsula
- SDP: Shandong Peninsula
- WTS: Western Taiwan Straits
- HCC: Harbin–Changchun
- CPL: Central Plain
- CAH: Central Anhui
- GZP: Guanzhong Plain
- SGX: Southern Guangxi
- TSM: Tianshan Mountains
- CSX: Central Shanxi
- CIM: Central Inner Mongolia
- CYN: Central Yunnan
- CGZ: Central Guizhou
- LXN: Lanzhou–Xining
- NNX: Northern Ningxia

These regions, shown in Figure 2.17, represent 26% of the country's land, 64% of its population, and 86% of its GDP [citation needed, Chen et al. 2020]; however, Chen et al. found that eight of the megaregions were morphologically and functionally monocentric, including Beijing, indicating that the benefits derived from megaregional economies were not inherently present or available.

An evaluation of China's megaregion policy by Su et al. also found that the policies implemented only succeeded on an economic level while failing in the domains of "rational urban growth, social equity, and environmental protection" (Su et al. 2017). Although the study was not exhaustive, the framework developed

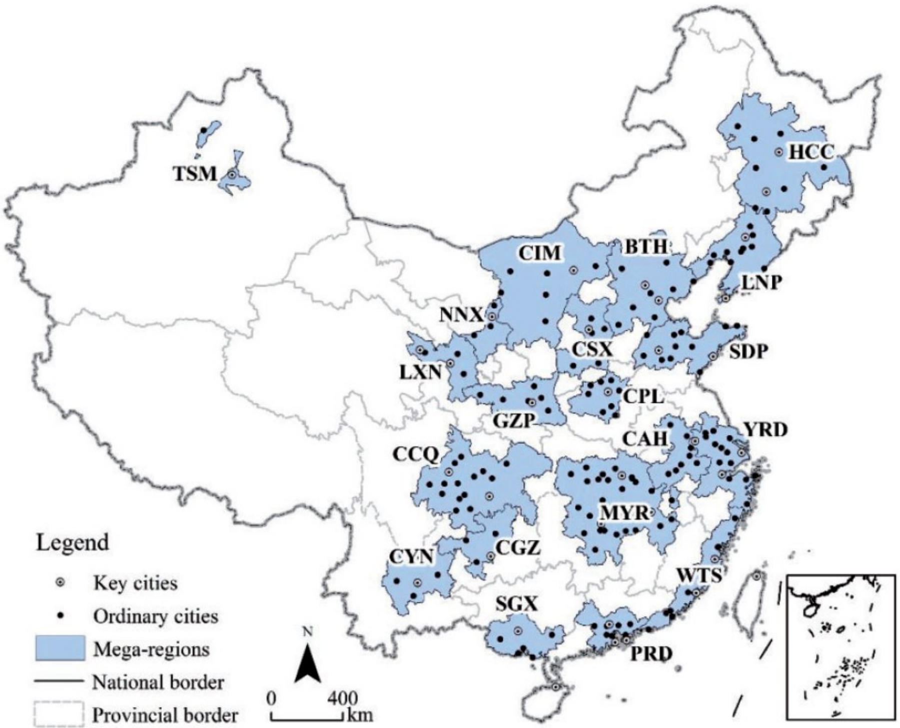

Figure 2.17: Chinese Megaregions, 2015 [citation needed, Fang 2015]

could be used as the basis for a more refined evaluation of other megaregions and of megaregional policy in other nations.

These disconnects between policy and ground-truth-ed reality highlight the pitfalls facing the application of megaregions. Without clear and relevant structures, definitions, and goals, regional governance will create bureaucratic bloat rather than more efficient and effective government.

2.4.5 European Union

The EU has used the Nomenclature of Territorial Units for Statistics (NUTS) since 1988 to gather data and statistics at a consistent regional level in order to be able to make international comparisons when the size of nation states varies greatly, see Figure 2.18. The statistics from this hierarchy of regions is then used to distribute funding from the European Regional Development Fund and the European Social Fund according to NUTS 2 regions' GDP when compared with the EU average[34], see Figure 2.19. Although there is significant attention from the EU on regional policy, discussion on megaregions is non-existent on their official

[34] "Less developed regions" had less than 75% GDP than the average, "transition regions" were between 75% and 90%, and "more developed regions" had more than 90% of the EU average GDP. See https://ec.europa.eu/eurostat/web/regions/background for further information.

website[35]. Still, by using a combination NUTS 3 EU regions and EO data, Marull et al. found the emergence of 12 megaregions in Europe, see Figure 2.20 (Marull et al. 2013): Am-Brus-Twerp, Brace-Lyon, Berlin, Frank-Gart, Glas-burgh, Lisbon, London, Madrid, Paris, Prague, Rom-Mil-Tur, and Vienna-Budapest. The same research saw significant economic benefit for NUTS regions within megaregions and recommended that modeling processes and policy priorities be altered to take encourage the development of these agglomerative economies.

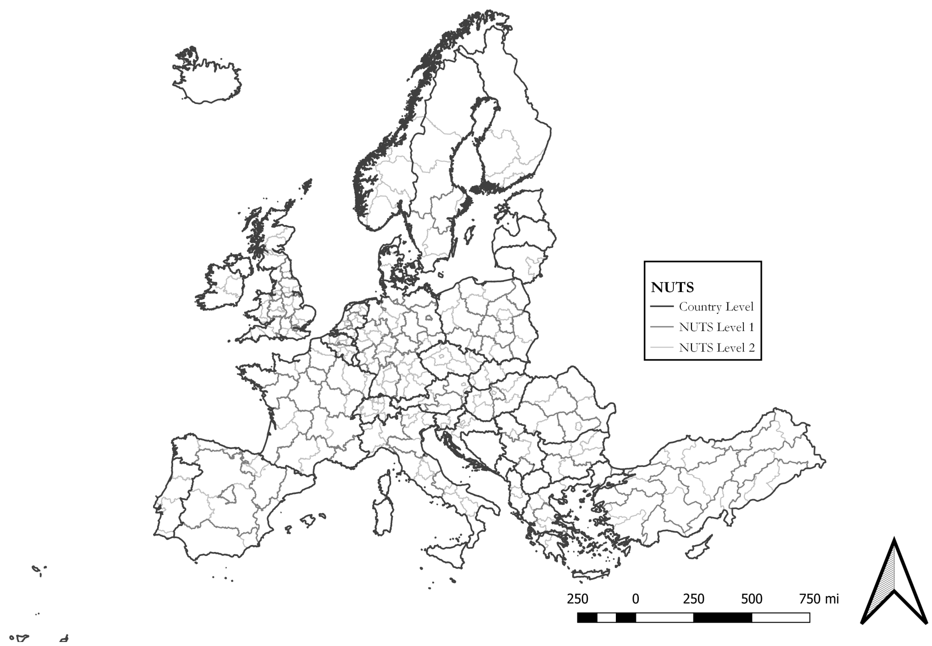

Figure 2.18: EU NUTS Regions for 2021.

2.4.6 United Kingdom

Especially given its recent separation from the European Union, a review of regional planning entities in the UK provides complementary context to their American counterparts. Planning occurred at the regional level in the UK during the latter part of the twentieth century, but the entities in charge of these processes were slowly dismantled and completely abolished by 2010 with spatial planning being returned entirely to local governments.

2.4.6.1 English Regions

These 9 regions of England had administrative functions between 1994 and 2011, but, since 2011, they are mainly used for statistical purposes, correlating to the NUTS Level 1 for the EU.

[35] As of 23 December 2020.

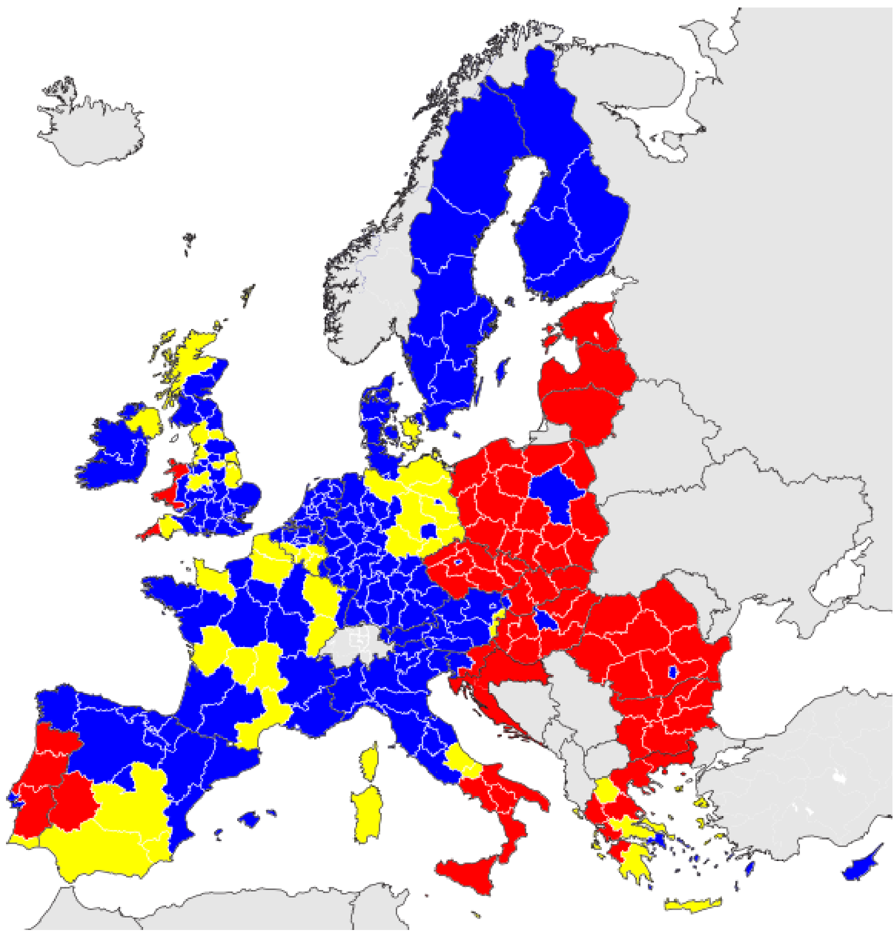

Figure 2.19: EU NUTS Regions by development status: red - less developed, yellow - transition, blue - more developed.

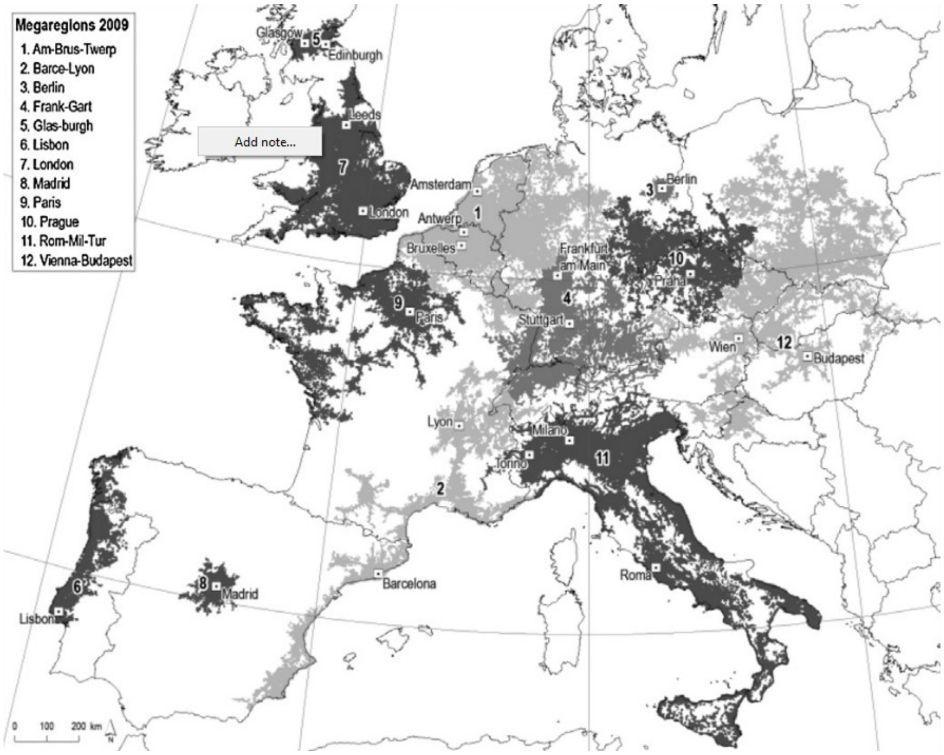

Figure 2.20: European megaregions developed from EU NUTS 3 and EO data.

2.4.6.2 Regional Development Agencies (RDA)

These agencies corresponded to the nine English Regions and operated between 1998 and 2010. Similar processes were also used in Ireland, Scotland, and Wales. These organizations had appointed boards comprised of local business leaders assisted by professional staff. Their primary activities were publishing regional Economic Strategy for the area, funding projects, coordinating with stakeholders, and working with the central government. They were replaced by Local Enterprise Partnerships after their official abolishment in 2012.

2.4.6.3 Regional Assemblies

These groups were created as part of the same legislation that created the Regional Development Agencies. Indirectly elected from counties councils, district councils, unitary authorities, and other interest groups, the assembly members oversaw the RDAs and developed a Regional Spatial Strategy. The London Assembly is the only remaining chamber after the others were dissolved between 2008 and 2010. These were replaced with Local Authority Leaders' Boards during the same period.

2.4.6.4 Greater London Authority (GLA)

Responsible for the City of London and the county of Greater London, the GLA consists of a executive mayor and a 25-member legislative body. The body was created by local referendum and derives its legal powers from the Greater London Authority Acts of 1999 and 2007. Its functions are concerned with transportation, economic development, and emergency services (such as police and fire). Given its powers and products, the body is something akin to a combination of a US municipal government and an MPO.

2.5 The Theoretical Nexus of Inter-municipal Politics and Regional Growth

2.5.1 Overview of Small State Political Theory

Small state political theory seeks to understand the behavior of low population nation-states in the context of the world stage. While the study of small states has lost prominence in political academia, a corollary can be drawn from these small states in world politics to local municipalities in within a country, coalitions of mayors, or even individual mayors of large cities in American politics. Patrick Geddes addressed the idea of small state politics well before it became a distinct theory in mid 20th century political science.

> "Generally speaking, the rulers of small states have accomplished more for their lands and cities, and even for general civilisation, than the rulers of great areas. David and Solomon, kings of Israel, are remembered more than the great Pharaohs or the Assyrian conquerors. The life of Pericles the Athenian stands higher in history than that of the spiritual tradition of the sacred cities, outweigh the glories of the great capitals which have risen and fallen throughout the ages. In fact, in west and east alike, do not the smaller states increasingly take initiative in advance of the greater? Indeed this must necessarily be so for great heterogeneous agglomerations of people cannot safely move faster than the pace of their less advanced members or than the level of agreement of their hostile parties (Tyrwhitt 1947)."

The influence of small nation-states is still highly relevant today despite the super-power-oriented nature of modern global politics. The primary characteristics of small powers can be summarized as dependence, status-quo orientation, supportive or international law and organizations, and risk aversion (Knutsen 2011). Several example of these patterns can be seen Middle Eastern countries, which exhibit varying degrees of influence on global politics. After exploring the bases of these small-state characteristics, this section will raise several questions and identify potential applications of theory to municipal governance in the United States.

2.5.1.1 Dependence on the Outside World

Small populations place inherent limits on states. Typically, small states are considered to have less than 10-15 million inhabitants, although some have placed this threshold as high as 30 million. Progress beyond an agrarian society requires greater specialization among citizens; and, as these small states gain more specialized citizens, their need to import non-specialized products grow. At the same time, limited academic settings and research industries lead to limitations in innovation, whether that is in physical goods to export or in foreign policy acumen (Steinsson and Thorhallsson 2017). Access to natural and human resources and capacities dictates much of a state's political power, but these can be supplemented by tools such as efficient economies, military power, and ideological influence.

Through coalition building, small states seek to reduce power asymmetry, make diplomacy more accessible, and limit the power of large nations. Groupings can be among small states, attachments to great powers, or a mixture of the two. These hierarchical relationships further enhance security, reduce uncertainty, and provide standards of behavior (Steinsson and Thorhallsson 2017). Unfortunately, some of these alliances have been literal ideological battle grounds, such as during the Cold War and subsequently the proxy wars in Middle East conflict between Iran and Saudi Arabia[36][37].

The means of connection between these parties can take several forms. In previous political eras, these were likely formal diplomatic relations and treaties where affiliated countries pledged military support should one of their number be invaded or attacked. In the current political schema, many inter-state compacts rely on economic agreements, the European Union being a prime example. This increasing economic integration has allowed for a greater number of countries to exist while simultaneously reducing the need for combining economies through national merges by expanding available markets and reducing barriers to trade (Alesina and Spolaore 1997). Easily accessible external markets are crucial for small states to overcome limits on their internal economies. Therefore, small states are motivated to support these systems and mechanisms which continue to move the global economy towards open markets.

2.5.1.2 Maintaining the Status Quo

In the last 70 years, small states have proliferated more than at any other time in history. Unprecedented periods of peace, economic stability, and democracy have set the stage for these nations to be able to overcome the structural barriers associated with their size. As this trend has progressed, the "function of military 'threats and opportunities' " (Alesina and Spolaore 2005) has decreased in the political calculations of minor

[36]https://www.brookings.edu/opinions/the-sunni-shiite-divide-in-the-middle-east-is-about-nationalism-not-a-conflict-within-islam/

[37]https://www.vox.com/2016/1/5/10718456/sunni-shia

nations while diplomatic and economic levers have gained more traction. Thus, small states have been known to successfully lobby larger powers to engender favorable foreign policy.

Although inter-nation violence has declined significantly during this modern period, small states are still vulnerable due to their economic and population constraints, which in turn lead to weaker military and diplomatic forces (Steinsson and Thorhallsson 2017). Thus, the safety of these nations relies heavily on domestic and international stability. Depending on the state's economic condition, this desire for stability and maintaining global order then manifests throughout national policy and politics.

2.5.1.3 Supporting International Law & Organizations

Since global peace and prosperity have given rise to these small nations, it is often in their best interests to maintain the current political system and work within it rather than to support wide scale change. These laws, practices, and organizations also serve to protect small states' rights and economies from colonizing powers; however, Russia in its conflicts with Ukraine is testing the practical application of this theory. Organizations such as the UN also provide the benefit of disseminating research and analysis that small states lack the resources to develop themselves while also conveying politically relevant information about major powers' interests and priorities in a setting where minimal diplomatic resources can still have an opportunity to strike deals and further agendas (Steinsson and Thorhallsson 2017).

Free trade is also key for small states. With smaller domestic markets, open economies free of tariffs and trade barriers are much more beneficial for small states than their bigger counterparts (Steinsson and Thorhallsson 2017). In order to provide their populations with their desired goods and services, these countries are more likely to focus on specific issues, sacrifice attention for other topics, and even surrender complete independence.

2.5.1.4 Aversion to Risk

External threats, perceived and actual, vary widely among modern small states. Luxembourg and Liberia must address very different issues, and, as such, small states are more likely to focus on particular issues whereas larger states have the bureaucratic weight to address many issues at once.
As with international violence, small state diplomacy often focuses on mitigating risks and leveraging available political influence to support the other three characteristics.

Hiding, either through political neutrality or natural geographic barriers, has also been a trend among small states to avoid unwanted attention from the wider world (Steinsson and Thorhallsson 2017). In turn, these intermediary actors also buffer tensions in global politics, if at some economic downsides, by encouraging large states to preserve their legitimacy and privilege in the eyes of other world powers.

While some issues faced by small states could be solved through mergers, the resulting population would be less homogeneous and the elected leaders supported by a smaller portion of voters. These leaders then have little reason to gamble their power base in favor of the greater margins for error afforded by agglomerated resources. Thus, low existential risk suppresses expansionist tendencies on the inter-national level and may even engender the shrinking of the average national size over time.

2.5.2 Application to Municipalities

Differences of situation and differences of scale must be addressed to apply these political theories to sub-state governmental organizations. Once these characteristics have been clarified, relationships and comparisons can be drawn more clearly.

Foremost is the need for self-defense. Cities and counties never have to provide militarization on the scale that a nation-state must. At most, their police force may deal with civil unrest. Outside invasion or more intensive disturbances quickly become the responsibility of the state or federal government. Comparison may still be made between policing and law enforcement, but these activities happen on the national and local level around the world. This highly stable situation allows municipalities to abandon the alliance making crucial to small nations in favor of competitive, non-cooperative behaviors.

Unless there are highly permeable borders between nations, as in the European Union, trade within a state encounters far fewer impediments. Meanwhile, any international or interstate trade is regulated under the purview of the federal government. According to Ross et al.,

> "... [at] present, there is no incentive for individual actors involved in local and regional planning to coordinate their efforts. Frequently, they instead compete against each other for resources, despite the presence of potential benefits of cooperation. Thus, in order for megaregion planning to be effective, a shift in how planning is conducted and perceived must occur. Several models for this shift have been proposed.
>
> Many of these models originate from Europe, where megaregion planning has progressed farther than in the United States. They include city-regionalism, functional polycentric development, reform-consolidation, market public choice, and new regionalism. None of these theories is without its drawbacks, however, and creating a framework for megaregion transportation planning in the U.S. will require further refinement. In addition to the trend of regional governance where public and private sectors and other interest groups form an alliance for regional interests rather than creating a new government, what is clear is that there remains an important role for the federal

government in providing leadership for megaregion planning efforts, while local and regional actors must develop the capacity and willingness to coordinate and undertake joint transportation (Ross 2011)."

In 2011, Xu and Yeh identified three regional governance concepts: federal intervention focused “reform-consolidation,” economism (Kwak 2017) and neoliberalism based “market public choice,” and public-private partnership institutionalized “new regionalism” (Xu and Yeh 2015). Unfortunately, all three of these ideas suffer from either unpopular or unfeasible bases that have likely prevented their broader societal acceptance or governmental adoption. Meanwhile, MPO-centric structures (e.g. regional commissions or councils of governments) have been relatively successful in the United States given their lack of enforcing power. Therefore, a lower-opposition route towards regional and megaregional governance may lie in this direction, co-leveraging offered incentives and enforced policies.

The basic framework for building an improved model of inter-municipal behavior that can effectively implement megaregional structures could be based on Steinsson and Thorhallsson’s three reasons for small states seeking hierarchical relationships: security and integrity, reduced uncertainty, and standards of behavior (Steinsson and Thorhallsson 2017). The implications of these ideas are briefly discussed here and the resultant recommendations are presented later in this work.

2.5.2.1 Security & Integrity

Cities and counties[38] within the US have a near complete absence of existential threats. Therefore, military protection and support is passed off to the federal level but is sometimes present at the local level in the form of personnel, bases, or supporting industries. Rather, this need is perceived and filled by police forces and militias, and local crime becomes a more present physical threat for most residents.

Structural threats to municipalities come from their parent state and less from the federal government. Many of these perceived threats are in terms of financial or infrastructural support. The incorporation of county land into cities may also fall under this umbrella for the county in question. Additionally, conflict between local and state goals may be used by local political leaders to energize public action or outcry.

2.5.2.2 Reduced Uncertainty

This ultimate reduction of external threats can maximize tendencies of municipal leaders to hold onto power and resist integration into larger, less homogeneous geographies that in times of danger proffer greater safety in numbers. In the context of regionalism and the usefulness of megaregions, this trend sets of the question

[38]In this work, “counties” is used as a general reference to counties and county equivalents.

of what size of administrative area is most effective and efficient at providing utility to residents. The calculus of maximizing boundary size to increase efficiency while maintaining resident homogeneity is what the process described in the methodology attempts to solve in order to describe the largest possible regions where residents face similar issues and challenges that planning staff can focus on in detail rather than finding a wider set of solutions for more disparate populations. It also seeks to develop practical administrative boundaries that encompass the needed locations for effective implementation of capital and programmatic investment. If the megaregion is workable scale for addressing these issues, government leaders at other scales should support its creation to in turn support high quality of life for their constituents.

2.5.2.3 Standards of Behavior

However, developing a set of cultural and legal expectations that have regional efficiency and effectiveness at its core may help reduce the impact of non-cooperative behaviors. Such a set would address the characteristics of small nation-states seen in municipalities by establishing meaningful opportunities for input in regional governance by other actors, protecting local interests while preventing consumptive competition, identifying systems and organizations that work to put municipalities on equal footing regardless or size or resources, creating insurance against real risks to municipalities and providing research and analysis to dispel incorrect perceptions of risk, and providing incentives to actively and constructively participate in and support the regional structures. Through plans and actions like these, megaregional governance can present itself as an enhancement to the abilities and capacity of local, state, and federal government and not another burden to be shouldered.

CHAPTER 3

Methodology

This methodology is primarily concerned with outlining a megaregional delineation process that produces boundaries robust enough to be easily legislated and practical enough for administrative functions of governance while still describing groups of communities with meaningful similarities and connections. The derivation of this process builds on the kind of data sources identified in literature and applies cluster analysis to delineate multi-faceted megaregion boundaries that reflect the societal phenomena created by the interaction of local communities and the globalized modern economy. This process also aims to be as strictly quantitative as possible so as to facilitate later reproduction, improvement, and adaptation to new information sets as data collection and distribution practices evolve and improve. Outcomes are described in the Results chapter of this work, and the potential application of the resultant boundaries to the implementation of megaregional governance through frameworks such small state political theory are covered in the Discussion chapter.

3.1 Developing Megaregions

3.1.1 The Role of Megaregions

Regional government has the potential to improve governance by agglomerating communities across administrative boundaries that might otherwise inhibit collaboration and coordination. As discussed earlier, municipalities often set the expectation that successes in neighboring communities is an inherent loss for them. This frequently occurs in the realm of economic development where businesses with "shop" among the locations they are considering to see who will offer the most advantageous incentives. The opposite occurs with affordable housing as perceived adverse impacts on image and property values are often cited a reasons to lobby for affordable housing to be located elsewhere. The proverbial game of hot potato has resulted in the continued concentration of poverty in already vulnerable and struggling communities.

The megaregional government would exist at a scale which provides the ability to bridge the gap between the global economy and local communities. The economic weight of constituent urban centers creates visibility on the national and world stage while a transparent connection to local motivators encourage responsiveness to those communities' feedback. As a preliminary step towards a policy and legislative framework for the implementation of megaregional governance, this chapter explores how the connective nature of megaregions and a multifaceted delineation process could facilitate the goals of improve efficiency and efficacy in government.

3.1.2 Global Cities

During the 20th century, several methods and theories were developed to explain the relationship between emerging urban mega-centers and the rapidly globalizing economy (Hall and Barrett 2012). One of these was Castell's "space of flows" that embedded key global cities as places where this flow "allow[s] for simultaneity of social practices without territorial contiguity" through telecommunication technology as high-level decision making centralized in contrast with the decentralization of back offices and service providers (Castells 1989). This structuralist approach has inspired further research into the firms and organizations that drive the global economy—for example, the Global and World Cities (GaWC) research network at Loughborough University. This research has also raised critiques of the approach's own power-based limitations that exclude many parts of the developing world (Hall and Barrett 2012).

Over the last two decades, the physical and intellectual expansion of the internet has led to a situation where individuals can engage with the global economy to the extent that some argue that there are no longer any "non-global" cities (Hall and Barrett 2012). Concurrently, social media and telecommunication have driven towards a realization of Castells' spatially distant society. Nevertheless, the global disruption caused by the COVID-19 pandemic has exposed the reality that these ethereal connections cannot be identically substituted for physically supported relationships[1]. It is on these face-to-face interactions that our global society's foundations are built, and that construction takes place on sidewalks, by parks, and at venues in local communities[2].

3.1.3 Local Communities

Both in perception and experience, the everyday interactions in local communities are often a world apart from those of the economic linchpin societies of global mega-corporations. Although telecommunication technology has vastly improve the world's capacity for long distance communication, local communities are still built up from municipal decisions focused on internal conditions, and impacts are less likely to be felt or directly measured by individuals outside the community. Conversely, the decisions made by global economic

[1]Sources include "Impact of the COVID-19 pandemic on quality of life and mental health in children and adolescents in Germany" (Ravens-Sieberer et al. 2021); "Age Differences in Stress, Life Changes, and Social Ties During the COVID-19 Pandemic: Implications for Psychological Well-Being" (Birditt et al. 2021); "Age Differences in COVID-19 Risk Perceptions and Mental Health: Evidence From a National U.S. Survey Conducted in March 2020" (Bruine de Bruin 2021); "Socioeconomic inequalities in the spread of coronavirus-19 in the United States: A examination of the emergence of social inequalities" (Clouston, Natale, and Link 2021); and "The mental and behavioral health impact of COVID-19 stay at home orders on social work students" (Lawrence et al. 2021).

[2]See Jane Jacob's *The Death and Life of Great American Cities.*

powers, private or public, can reshape the lives of thousands, if not millions, with little input from those individuals. Events like the housing crisis that sparked the 2008-2009 recession are exemplary of the everyday influence held by powerful market actors. Given this context, how can we describe the relationships between local areas in a way that global, economy-driving, societal meta-structures emerge from their amalgamation?

Christaller's Central Place Theory (CPT)(Hall and Barrett 2012) provides one framework for this question by forming the backbone of local areas from the geo-economic interplay of marketing and density. While the systems modeled by microeconomics are even more fickle those considered by macroeconomics, the positivist and economic theories used to develop methods such as CPT struggle to account for the irrational behavior of members of society(Hall and Barrett 2012). Meanwhile, Jane Jacobs and her subsequent adherents formed a movement around the social fabric of the neighborhoods and cities, perhaps the most difficult aspect of "community" to quantitatively measure and evaluate. Human interactions can range in intensity and frequency from being in another person's field of view across the park to intimate and lifelong friendships with a host of resultant societal outcomes.

Despite telecommunication technology's dramatic improvement over the last half century, the impact of the pandemic on in-person interaction has highlighted how far it has yet to go before remote connections compare to those that are face-to-face. Isolation, tribalism, social media echo chambers, have come to fill the voids left as trust building micro-interactions on sidewalks, visits with friends, gatherings at entertainment events, and a host of other societal behaviors evaporated into the ethernet in the face of the global pandemic. The subsequent protests, riots, and movements cited frustration, or even anger, with governmental action (or inaction) that has increased the distances between people, physically and economically. Therefore, it seems that the dreams of the futurists referred to by Castell(Castells 1989) have yet to be fully realized when it comes to the separation of our social fabric from the urban fabric and that government which is responsive to local discourse is still a crucial part of maintaining a cohesive society.

3.1.4 "Glocal" Fabric

What then connects these two systems, global civilization and local communities? CPT and other spatially centered thought suggests that it is the flow of goods and services resultant from contrasting forces such as agglomerative economies and bid-rent curves. To borrow some language from economics, the hierarchy of urbanization intensity separates populations by their differing utility functions while retaining, and sometimes altering, the connections that held together the precursor communities. Castell suggests that, while industry does connect to place through its need for labor, its organizational logic is spatially independent from the local scale and primarily concerned with the global social-power network. Although, he does assert that communities and local government should take action counteract the power imbalance resultant from the new, space-of-flows-driven economy, Castell does still separate these two facets of society to a large extent (Castells 1989).

Nevertheless, by observing flows of various metrics—freight, commutes, or migration, boundaries emerge as they do in the natural world. Watersheds are an intuitive example where barriers and routes of least resistance predict the *flow* of water and materials across the landscape. Similarly, there are economic, social, and demographic barriers that prevent or encourage various behaviors in populations. Commute times can only be so long before they encroach on working time. Financial barriers may discourage a family from moving to distant portions of the country or to places where they have no connections to leverage.

To help quantify relations between communities, Smith and Timberlake's 1995 "Conceptualizing and Mapping the Structure of the World System's City System" typology offers four functions of flows (economic, political, cultural, and social reproduction) with three types of flows (human, material, and information). At the time of their writing, the authors felt that gathering sufficient data to describe "a substantial number of the world's cities for even one point n time would be overwhelming." This problem, while still challenging, is now within the realm of possibility in today's world of "big data." However, a new issue arises in filtering the available data down into a subset that is, at once, both representative of communities *and* comprehensible by planners, decision makers, and the public after analysis is complete.

Examples provided by S&T of typology aspects:

- Flows: economic relations
  - Human: labor migration
  - Material: commodities
  - Information: business communications
- Political: invasions, foreign aid, treaties
- Cultural: visiting dance troupes, art exhibitions, feature films
- Social Reproduction: family migration, remittances, post cards

3.2 The Megaregion Delineation Process

Glocker (Glocker 2018) identifies two categories of megaregional definitions: morphological and functional. The Smith and Timberlake typology clearly identifies a broad dataset for the flows associated with functional relationships, and these terms can in turn be applied to static morphology. Combining these two methods provides a framework for delineation in which Structuralism retains its spatially unconstrained connections between major metros and the resultant megaregions can act as spatially grounded intermediaries between urban cores, hinterland local communities, and the broader economy per the Positivist worldview.

At the present time, the geographical boundary level with the most consistent documentation and broadest set of data in the United States is the county and its equivalents. States are too large to provide meaningful results for local communities, and, at the census tract level and below, struggles arise in clearly communicating flows of people and goods as even short trips can pass over multiple geographies. A fine-grained analysis would likely provide a different result from this county level analysis, but such a process is beyond the scope of this work's time and resources.

3.2.1 Important Principles

The use of megaregions and their governmental bodies bridges the global-local divide by grouping communities that are related via economic, political, and cultural morphologies and functions through a delineation process that leverages a spatial minimum spanning set of datasets and weighted clustering algorithms. The intention of this process is to define administrative boundaries of meta-urban communities that already exist so that government can better serve its residents while consuming fewer resources.

To achieve these goals, megaregional boundaries should be...

- contiguous, without holes or islands, and relatively cohesive with few spurs (e.g. low maximum betweenness centrality) to reflect the structure other existing administrative geographies in the US;
- built from counties and county-equivalents to minimize the splitting of local municipalities;
- centered around major MPO areas ensure the continuation of their developed regional plans, models, and relationships, with large, influential regions forming the basis for megaregion identification;
- developed from a minimum number (as a sort of "minimum spanning tree") of data sources that can meaningfully describe typology category; and
- based on data that is publicly accessible across space (the contiguous United States) and time (within the last ten years to reflect census frequency).

3.2.2 Proposed Process

1. Create 18 categories based on Glocker's two types of megaregion definitions and Smith and Timberlake's four functions of and three types of flows with the two societal flows (Cultural & Social) being combined.

    1. Hierarchy: (Morphological | Functional) -> (Economic | Political | Societal) -> (Human | Material | Information)

2. Add initial datasets to each category that meet given conditions

    1. Data Availability: dataset is publicly available in a stable (or multiple) source, covers the entire contiguous United States, and is available from at least within the last 10 years.
    2. Previously Reference: the type of data should either be referenced explicitly in megaregion research or practice or be discussed in regional science research as a variable for community identification.

3. Create contiguous clusters of counties for each dataset in each category.

    1. Run data under consideration through the R package ClustGeo[3] function "hclustgeo" with D0 being the non-euclidean distances between counties, D1 being the spatial distance between county centroids, the number of clusters (k) ranging from 2 to 48, and the mixing parameter (alpha) ranging from zero to one (exclusive) in increments of 0.1.
    2. Evaluate the mean of average in-cluster distances and the size of cluster membership for each k-alpha combination.
    3. For each dataset k value, select the alpha which produces the lowest value for the sum of normalized means of average in-cluster distances plus normalized inverse of number of cluster members (1 - normalize(cluster members)).
    4. Select a k value optimizes in-cluster distance and membership for all datasets under consideration.
    5. Return the county partition for the selected alpha and k values.
    6. Remove non-contiguous elements for each cluster.
    7. Fill any holes in the cluster.

4. Megaregion core, transition, and periphery areas are identified by MPO location within cluster groups.

    1. Intersect all cluster regions from Step 3.
    2. Iterate through MPOs in order of total population of member counties, and isolate cluster regions containing the primary MPO.
    3. "Core" areas are the intersected cluster regions that contain the primary MPOs. Other MPOs within this boundary are considered "core MPOs."
    4. "Transition" area are intersected cluster regions that share an underlying cluster region with the primary MPO at least 50% of the time.
    5. "Periphery" areas are intersected cluster regions that share at least one underlying cluster region with the primary MPO.
    6. Each area type is checked for contiguity and holes like the cluster regions in Step 3.6 and 3.7.

[3] "ClustGeo: an R package for hierarchical clustering with spatial constraints" by Chavent et al., 2017, https://arxiv.org/pdf/1707.03897.pdf

7. If the next MPO is in a megaregion core or periphery, continue to the next MPO.
8. Megaregion transition and periphery areas may overlap, but core areas must, by definition, not share any area with another megaregion.

5. This process is repeated for each category and then for each meta-category on the typological tree.

6. summary statistics are prepared for geographic and non-euclidean factors to compare core, transition, and periphery areas between typology groupings.

3.2.2.1 Visual Examples of Process Results

This section contains relevant visual references for the megaregion delineation process. Examples of specific megaregions are pulled from the Morphological-Societal data group of outputs for the Miami-Dade MPO.

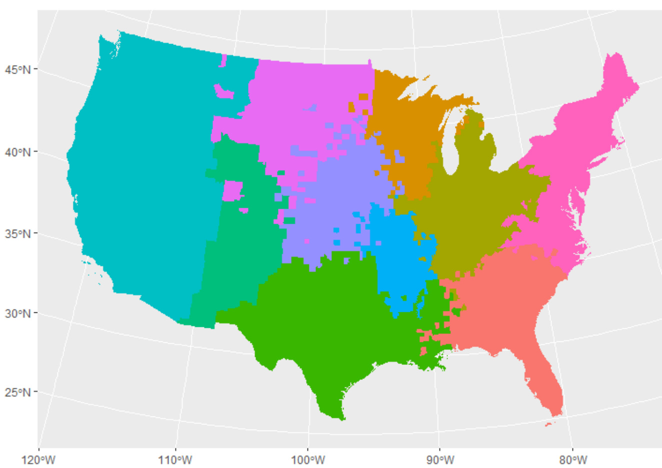

Figure 3.1: One set of raw cluster regions for the MS data group, Step 3.5.

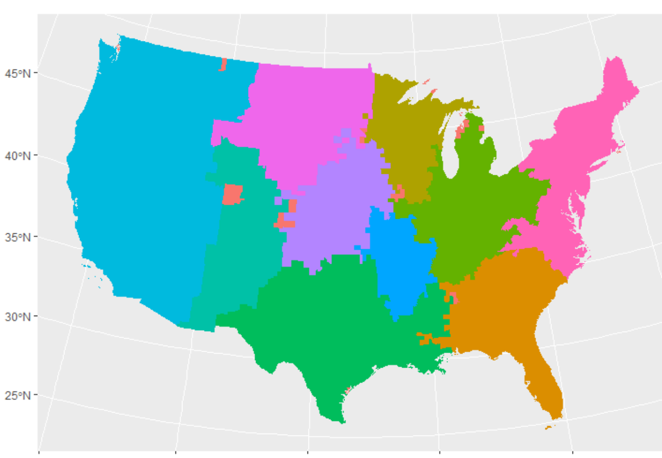

Figure 3.2: One set of contiguity and island adjusted cluster regions for the MS data group, Steps 3.6 and 3.7.

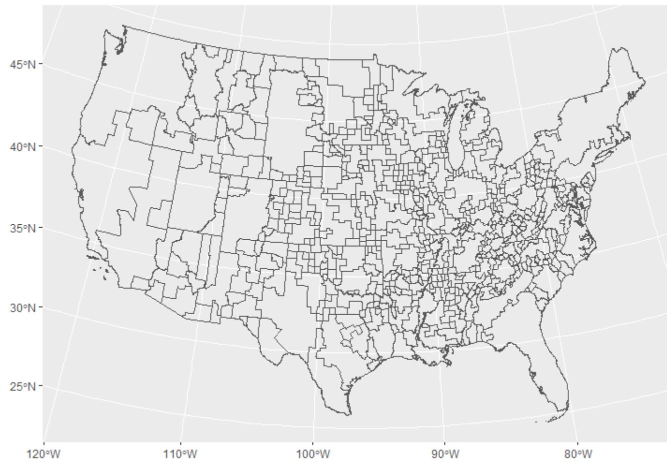

Figure 3.3: Intersection of all cluster regions resulting from the MS dataset group, Step 4.1.

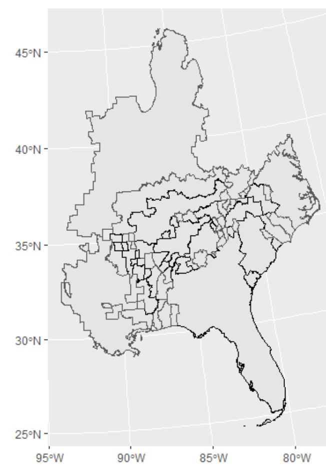

Figure 3.4: Intersection of cluster regions from the MS dataset group containing Miami-Dade, Step 4.2.

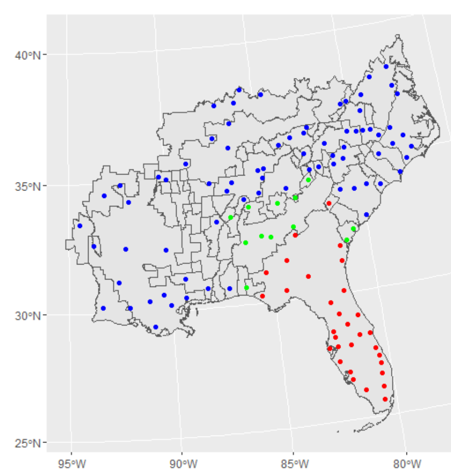

Figure 3.5: Example of intersected cluster regions containing core (red), transition (green), and periphery (blue) MPO centroids with the primary MPO being Miami-Dade, Steps 4.3 through 4.5.

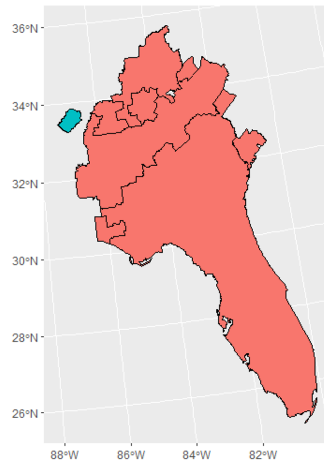

Figure 3.6: Contiguity check visualization for the transition zone of the Miami-Dade MS megaregion, Step 4.6.

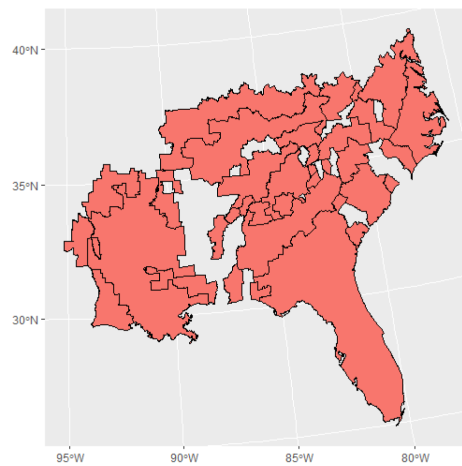

Figure 3.7: Contiguity check visualization for the periphery zone of the Miami-Dade MS megeregion, Step 4.6.

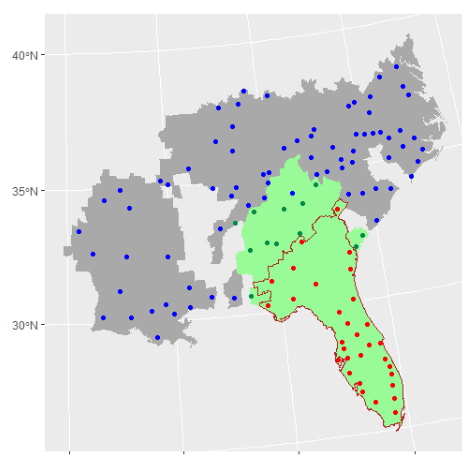

Figure 3.8: Core, transition, and periphery areas of the MS data group Miami-Dade megaregion with their associated MPO centroids.

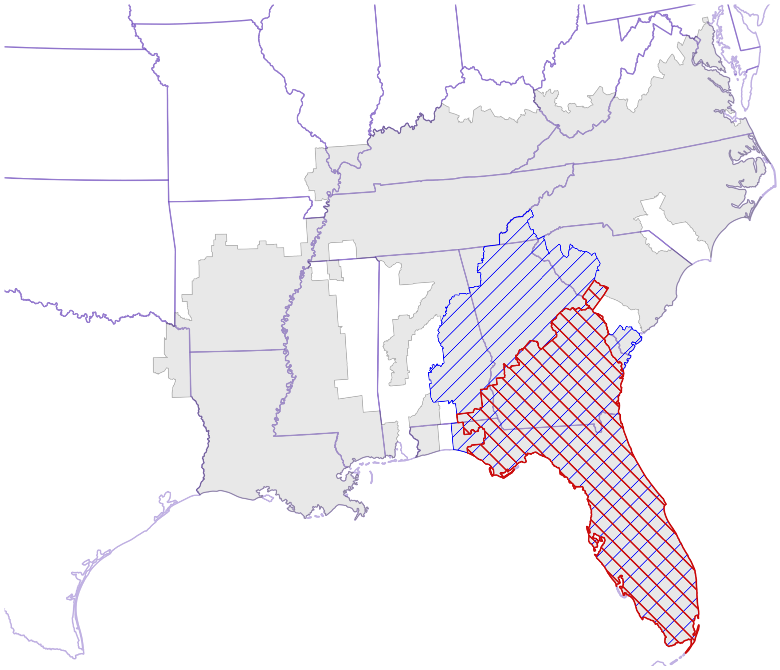

Figure 3.9: Miami-Dade Megaregion for the MS data group with state boundary overlay.

3.2.2.2 Typological Categories

Not all categories may have datasets identified due to the time constraints of this work and availability of resources; however, this process is intended to be an ongoing effort such that new iterations can be implemented with predictably and low additional effort. Relevant code and application processes are documented in the Appendices for this purpose.

The following bulleted list separates each typology category and provides an example of a dataset used or proposed for the delineation process. The "social" and "cultural" groups have been merged into a "societal" group for the purpose of this work as categorizing relevant datasets between them would not better serve the overall purpose. Datasets identification was not able to locate data for every category that met the process criteria.

- Functional
  - Economic
    - Human: LEHD LODES
    - Material: Commodity Flow Survey
  - Societal
    - Human: IRS SOI Migration
- Morphological

- Economic
  - Human: Household Income Distribution
  - Information: county Business Patterns
- Political
  - Human: Presidential Voting Records
- Societal
  - Human: Medicare Reimbursements, Population, and Racial Distributions
  - Material: Köppen Climate Classification, National Land Cover Database
  - Information: Language at home, Nativity, Patents, and Proximity to Post-secondary Education

3.2.3 Clustering Counties into Regions

Although simple methods of megaregional delineation are more easily communicated, reliably implemented, and aesthetically pleasing, if they do not accurately represent the reality of social and urban structures, they fail to describe the unique geographies under consideration here. A comprehensive theory of the nature of megaregions is beyond the scope of this work; nevertheless, repeating an analytic process with a variety of data sources contributes to the body of knowledge that would lead to such a theory.

The human mind is designed to recognize natural patterns to survive, but the patterns that emerge in urban forms are not entirely natural. Therefore, this process utilizes cluster analysis techniques to rigorously identify patterns within large dataset that cannot be easily comprehended. Still, the method does fall back on human interpretation when determining aspects such as the number of clusters partitioned. First datasets are processed through a range of number of clusters from 2, the smallest integer possible for the process, to 48, the number of partitions already existing within the contiguous United States, and then spatial contiguity and non-euclidean within-cluster distances act as two avenues to alleviate the issue of qualitative decision in the clustering pipeline.

The ClustGeo[4] package utilizes a combination of feature and constraint spaces to impose "soft" contiguity constraints on tree formation. As the mixing parameter approaches 1, clusters reflect the geographic distribution of counties. Although highly influential urban cores often have a great deal in common, the contiguity constraint requires the connection between them to be strong enough to also encompass nearby hinterlands, thereby linking urban, suburban, and rural communities.

To evaluate the performance of the partitions, the process balances low in-cluster distances (representing

[4]"ClustGeo: an R package for hierarchical clustering with spatial constraints" by Chavent et al., 2017, https://arxiv.org/pdf/1707.03897.pdf

clusters of increasingly similar communities) with high cluster membership (pushing towards large regions that can leverage economies of scale when tackling planning and governance issues). Meanwhile, the overlapping of megaregion transition and periphery areas was inspired by the work of Georg et al. in describing the "fuzzy" boundary of the Boston-Washington urban corridor (Georg, Blaschke, and Taubenböck 2018), and, according to Marull et al., these areas likely benefit most from the development of megaregional networks and economies (Marull et al. 2013).

3.2.3.1 Future Expansion

Continuing research into megaregions and other meta-regional societal structures will identify new data sources and produce improved data products for delineating boundaries. As this occurs, the follow steps may be used to determine if a new dataset should be added to the typological tree and if doing so should displace any of the existing data sources.

Add any additional datasets that meet the following criteria to their respective categories.

1. When comparing clusters from the new dataset, the Jaccard Coefficient should be greater than 0.50 for all previously included datasets in category.
2. If a new data source is added, the entire category should be reevaluated to determine whether each dataset in the group if making any contribution to the clustering performance; such evaluation to take to implement rigorous statistical identification of optimal culling of data sources.
3. Existing dataset may also be replaced if a new dataset performs better in terms of in-cluster distance and cluster membership *and* its Jaccard Coefficient is greater than 0.50 for all other datasets in the category.

CHAPTER 4

Results

This chapter focuses on the megaregion boundaries produced through the process outlined in the previous chapter. Some applications and opportunities for further research are mentioned here, and a review of major patterns, arenas for application, and strategies for implementation are found in the Discussion chapter. The delineation R-code, dataset sources, and resulting partitions of contiguous US counties are included as appendices to this work in the hopes that future research efforts aimed at implementing megaregional governance can quickly move beyond the scope of this work.

The subsections of this chapter deal with results from combinations of typological categories. Each sections contains a graphic comparing in-cluster distances with the number of cluster members, a map of the resulting megaregions' core, transition, and periphery areas, and a table of summary statistics comparing the megaregions to the nation as a whole.

Given the number of possible typology categories combinations, the resulting "typological tree" is diagrammed in Figure 4.1 along with the terms used here to distinguish between levels. "Data Categories" consider a focused set of data sources that describe patterns that share form (characteristic or flow), topic (economic indicators, political systems, or societal trends), and medium (resident populations, physical goods, or intangible information). "Subject Groups" gather datasets regardless of their medium, and "Type Divisions" gather data that represents flows between locations and data that describes locations into separate groups. Finally, the complete set of data is plugged into the process in an attempt to delineate megaregions for administrative use.

4.1 Individual Data Categories

Although some categories contain multiple data sources, each category seeks to identify similar trends. However, due to a lack of research on regional or megaregional delineation based on certain data sources, several categories lack any data source. These are prime opportunities for further research, and existing sets of data source may also be improved by swapping in sources that create clusters similar to those of an existing source with better performance or by adding sources if their contributions are significantly different from all sources currently in the set.

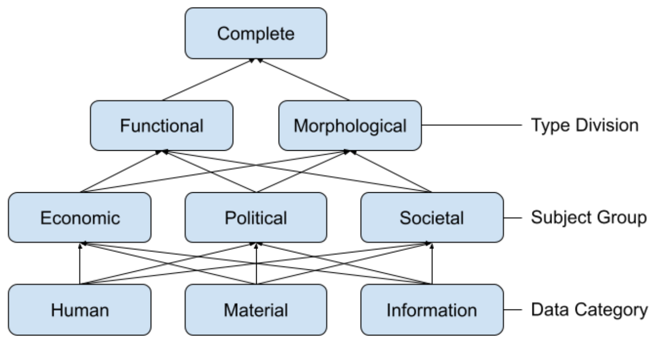

Figure 4.1: Typological tree for the combination of datasets when delineating megaregions.

Much of existing literature builds on the broader body of work in community planning and regional science. Commuter patterns, freight routes, population centers, and the like have all performed well in identifying core communities, but such qualitative prioritization of input ought to only serve as a starting point. It is entirely possible that a previously unused data source may outline the same communities as a more complex set with as much or greater accuracy. Should such a case occur, it would provide considerable insight into the socio-spatial nature of meta-communities.

4.1.1 Functional-Economic-Human

Datasets in Category:

- LEHD LODES

This category intends to track the movement of individuals as part of the macro-economic system. Here the LODES dataset provides commuting OD pairs for census tracts, but these have been aggregated to the county level for this analysis.

Of the three functional-economic groups, this one shows the strongest potential for an optimal number of clusters outside the range evaluated. After a stable period from 2 to ~15 clusters, the mean of in-cluster distance continues to fall steadily until 48 clusters, see Figure 4.2. Although the algorithm selected 46 clusters

Table 4.1: Summary Statistics for the Core Area Produced by the FEH Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	2964560	0.98	1201	64447	212929
# Member Counties	3106	3054	0.98	2	66	264
# Member MPOs	387	375	0.97	1	8	31
Intersecting States	48	41	0.85	1	3	7
Total Population	328016242	301623699	0.92	672008	6557037	17810084
Avg. Pop. Density	107	102	0.95	7	603	3181
Avg. Household Income	65036	62437	0.96	44977	66502	111360
% Difference D/R Votes	-0.024	-0.009	0.351	-0.5	0.038	0.518
% Pop. White	0.727	0.738	1.01	0.465	0.75	0.902
# of Patents	1046391	991606	0.95	491	21557	110582
Avg. Dist to Post-Secondary	886	945	1.1	755	1042	1719

to map in Figure 4.3, there are several areas where even visual inspection and intuition indicate that the distances are somewhat far to be driving on a regular basis. Nevertheless, daily economic patterns are important to developing an cohesive picture of community habits, and this observation simply indicates an opportunity for future research.

Table 4.1 reports summary statistics for these core areas. Use of a single summary statistic table will be continued without comment for other single data source categories, and transition area tables will be included for results with relevant. Because the clustering covers nearly the entire geographic extent of study, summary statistics are similar between the national level and the cluster regions for the single data source categories.

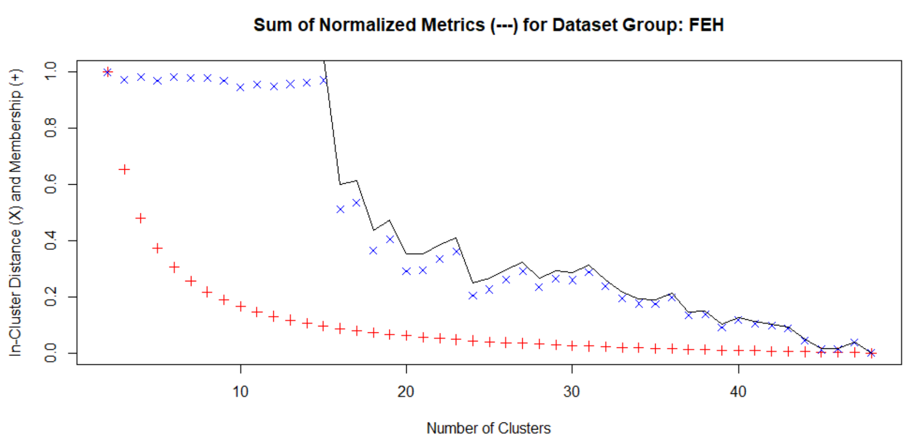

Figure 4.2: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the FEH data set group.

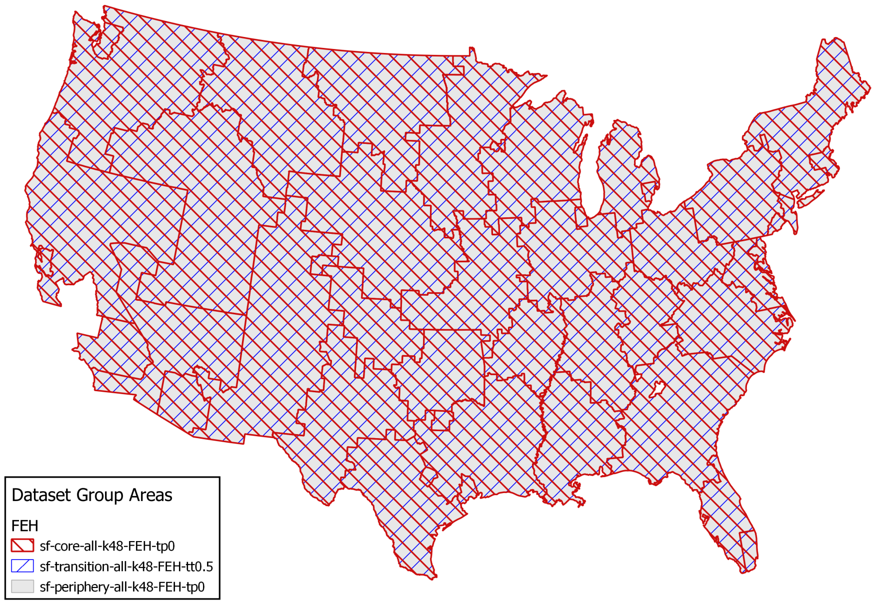

Figure 4.3: Map of megaregional core, transition, and periphery areas based on the clusters produced by the FEH dataset group.

4.1.2 Functional-Economic-Material

Datasets in Category:

- Commodity Flow Survey

County estimates for commodity flows produce some of the largest clusters in this entire work, see Figure 4.5, possibly indicating the long distances goods travel from port of entry to final use. It should be noted that each cluster region in this case is in contact with the US border, further supporting a connection to the import/export process. The in-cluster distance pattern shown in Figure 4.4 from 10 clusters on is also uniquely steady among its categorical companions. While this could be an artifact of the proportional distribution of flows, the trend is strong enough to be worth further research and comparison against other large-scale freight datasets.

4.1.3 Functional-Societal-Human

Datasets in Category:

- IRS Migration

The cluster sizing for year-to-year household migration falls between commutes and freight; this can be observed both by comparing the graph in Figure 4.6 and the map in Figure 4.7 to their companions. Freight's

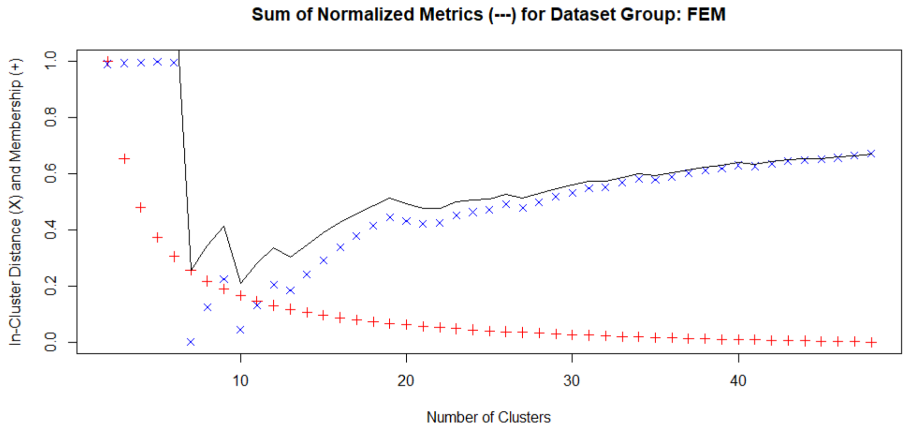

Figure 4.4: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the FEM dataset group.

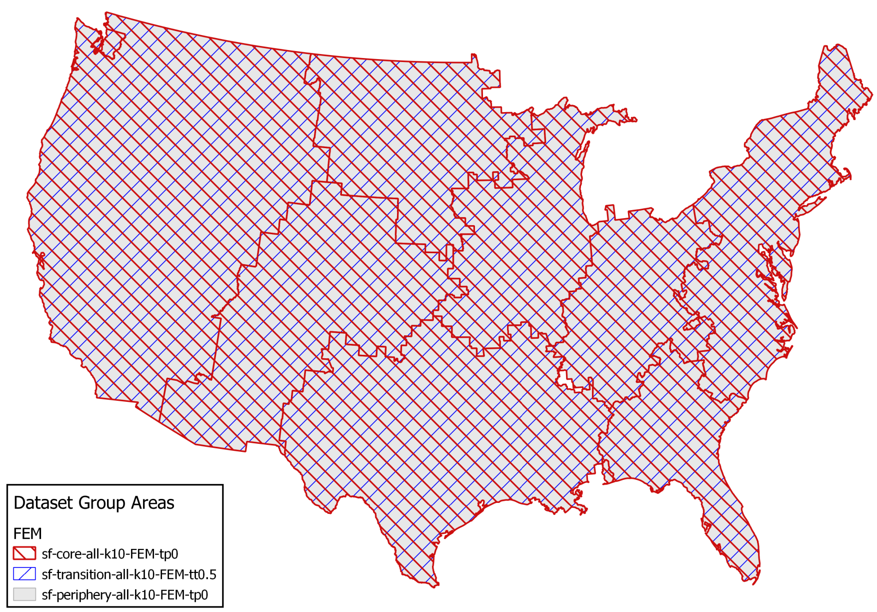

Figure 4.5: Map of megaregional core, transition, and periphery areas based on the clusters produced by the FEM dataset group.

Table 4.2: Summary Statistics for the Core Area Produced by the FEM Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	2956846	0.98	4309	328538	817017
# Member Counties	3106	3028	0.97	2	336	569
# Member MPOs	387	370	0.96	1	41	101
Intersecting States	48	9	0.19	1	9	16
Total Population	328016242	310316471	0.95	2994310	34479608	81402630
Avg. Pop. Density	107	105	0.98	28	189	695
Avg. Household Income	65554	63732	0.97	55449	67245	92904
% Difference D/R Votes	-0.024	-0.022	0.89	-0.137	0.014	0.179
% Pop. White	0.727	0.727	1	0.667	0.746	0.84
# of Patents	1046391	1002027	0.96	25440	111336	315367
Avg. Dist to Post-Secondary	886	937	1.12	764	1062	1720

optimal cluster number is clearly near 10 and commutes may be larger than inspected, but the migration clusters spend a significant range near an optimal level. It may be possible to further clarify an ideal number of clusters by adding the exemptions and aggregated AGI fields available from the same data source. The IRS also provides this data annually back into the 1990s, which would allow for several additional flow types: shifts in flows over time, aggregated flows, and time-lagged flow comparison.

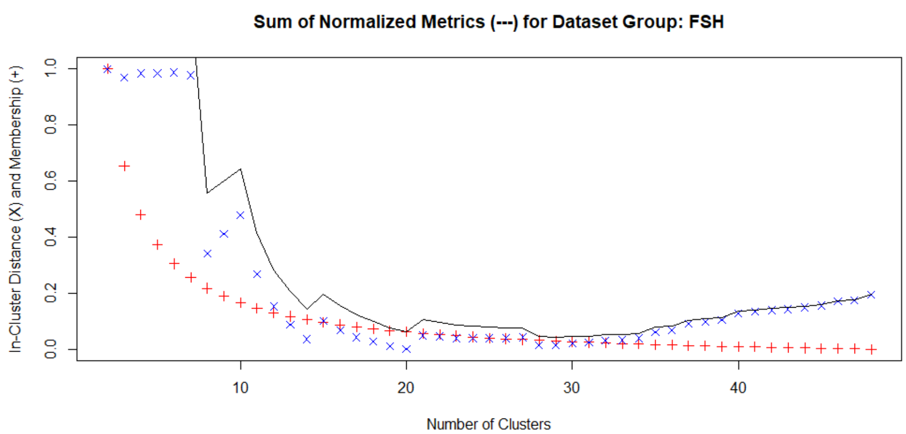

Figure 4.6: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the FSH data set group.

4.1.4 Morphological-Economic-Human

Datasets in Category:

- Household Incomes

The facet of ACS Table S2503 used for these analyses was the county median household income and the

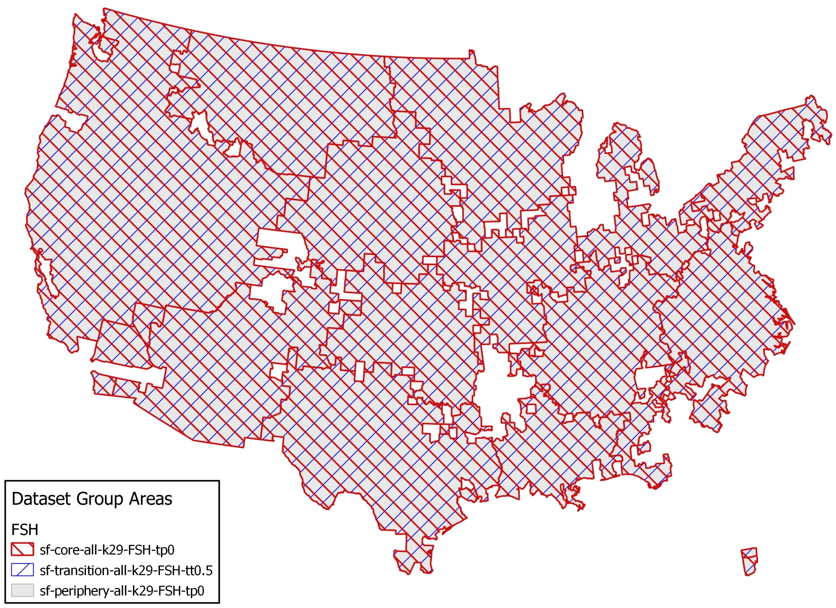

Figure 4.7: Map of megaregional core, transition, and periphery areas based on the clusters produced by the FSH dataset group.

Table 4.3: Summary Statistics for the Core Area Produced by the FSH Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	2593640	0.86	909	123507	560494
# Member Counties	3106	2532	0.82	1	121	415
# Member MPOs	387	235	0.61	1	11	45
Intersecting States	48	18	0.38	1	5	14
Total Population	328016242	177376982	0.54	1624151	8446523	29166378
Avg. Pop. Density	107	68	0.64	7	545	2869
Avg. Household Income	61553	59547	0.97	48047	62225	94974
% Difference D/R Votes	-0.024	0.062	-2.556	-0.346	0.062	0.455
% Pop. White	0.727	0.762	1.05	0.602	0.742	0.915
# of Patents	1046391	470104	0.45	1669	22386	96000
Avg. Dist to Post-Secondary	886	1015	1.18	779	1112	1715

Table 4.4: Summary Statistics for the Core Area Produced by the MEH Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	2922535	0.97	135306	487089	1000505
# Member Counties	3106	3008	0.97	302	501	820
# Member MPOs	387	371	0.96	40	62	95
Intersecting States	48	6	0.12	7	13	17
Total Population	328016242	313135767	0.95	29452447	52189294	74236617
Avg. Pop. Density	107	107	1	56	149	319
Avg. Household Income	65521	63554	0.97	53370	64353	74417
% Difference D/R Votes	-0.024	-0.028	1.171	-0.157	-0.006	0.188
% Pop. White	0.727	0.727	1	0.69	0.743	0.803
# of Patents	1046391	1032099	0.99	94476	172016	322878
Avg. Dist to Post-Secondary	886	908	1.02	769	965	1408

number of households in various income brackets. The optimal number of clusters is one of the lowest on the typological tree, but there are several values near 40 clusters that are nearly as low. This behavior could then be a reflection of labor markets varying on megaregional and local levels but not regionally or between small sets of regions. An investigation of correlation between this and other income and wealth datasets may reveal further clarification on this topic.

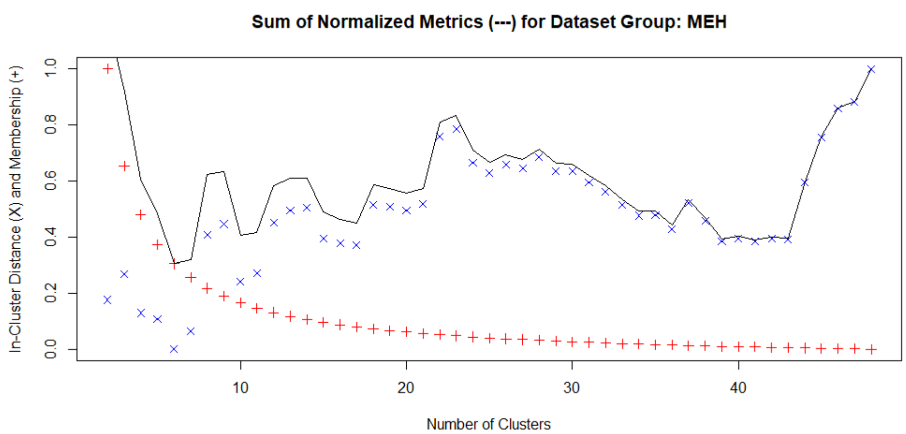

Figure 4.8: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the MEH data set group.

4.1.5 Morphological-Economic-Information

Datasets in Category:

- County Business Patterns

With more than three times as many cluster compared to the household incomes result, employment by

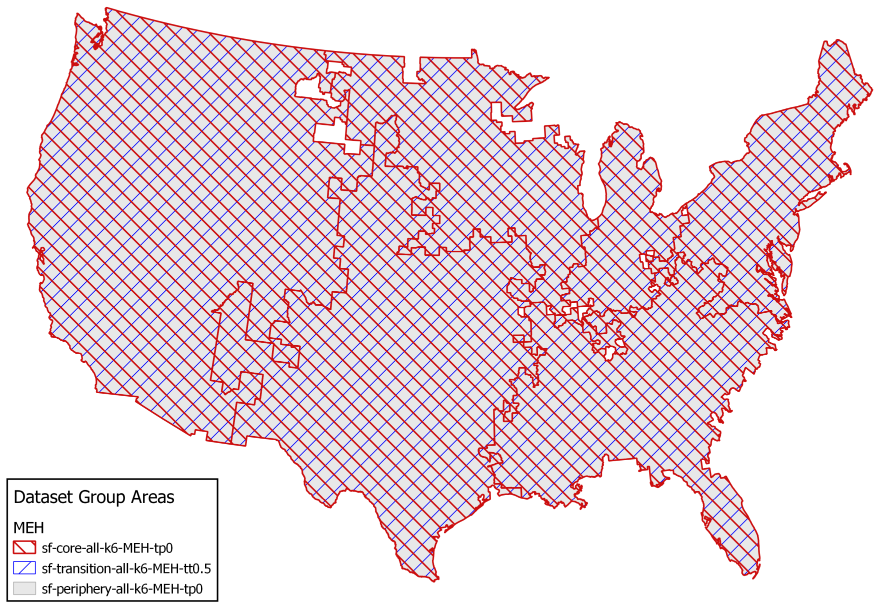

Figure 4.9: Map of megaregional core, transition, and periphery areas based on the clusters produced by the MEH dataset group.

industry appears to have trends and patterns close to the scale of median sized states. Further exploration of clustering trends could be explored by trying NAICS-code lengths besides three and by adding in payroll and establishment counts.

4.1.6 Morphological-Political-Human

Datasets in Category:

- Presidential Voting Patterns

Unlike many of the other single-category cluster regions, this result follows some state boundaries, especially in the Mountain West. Although there are clear state politics reasons that this is not a surprising pattern, no state-level elections were included in this category, but these trends may still be connected to state social identities or to state-level political messaging. The MIT Election Lab does provide House and Senate election data which may further clarify these patterns, and the addition of gubernatorial results may do the same. In developing a new level of governance, buy-in from the political class is crucial to ensure sufficient support for megaregional efforts to progress and subsequently improve citizens' and residents' quality of life, making this category a keystone for the success of more generalize sections on the typological tree.

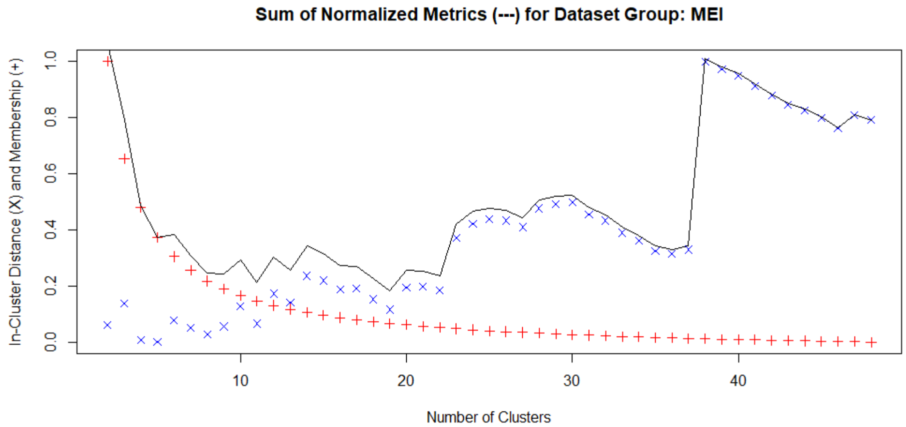

Figure 4.10: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the MEI data set group.

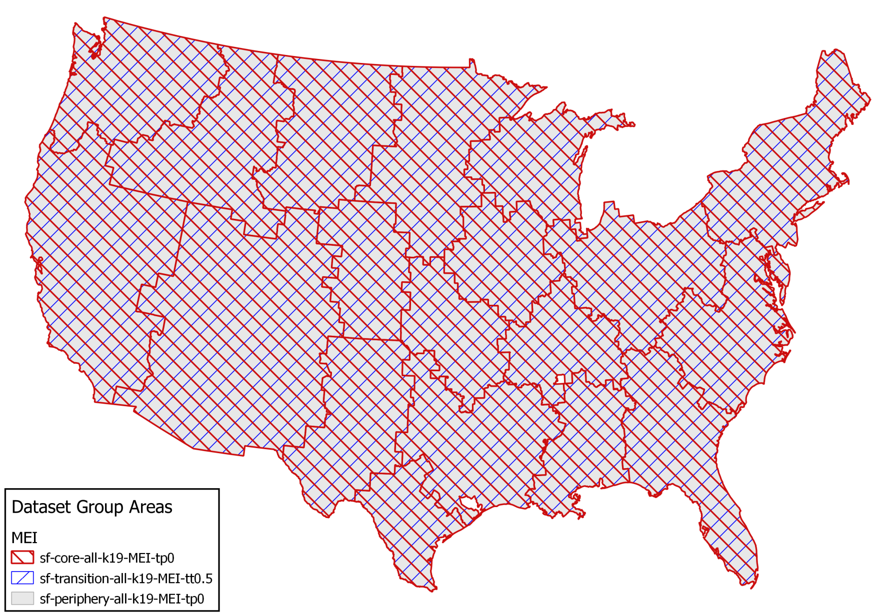

Figure 4.11: Map of megaregional core, transition, and periphery areas based on the clusters produced by the MEI dataset group.

Table 4.5: Summary Statistics for the Core Area Produced by the MEI Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	2971842	0.99	72509	156413	408588
# Member Counties	3106	3033	0.98	64	160	384
# Member MPOs	387	373	0.96	1	20	56
Intersecting States	48	19	0.4	1	6	10
Total Population	328016242	299934141	0.91	701172	15786007	50078385
Avg. Pop. Density	107	101	0.94	7	103	329
Avg. Household Income	65492	58513	0.89	51361	61625	76203
% Difference D/R Votes	-0.024	-0.008	0.335	-0.192	0.122	0.516
% Pop. White	0.727	0.733	1.01	0.606	0.789	0.906
# of Patents	1046391	961170	0.92	609	50588	244925
Avg. Dist to Post-Secondary	886	963	1.09	764	1030	1633

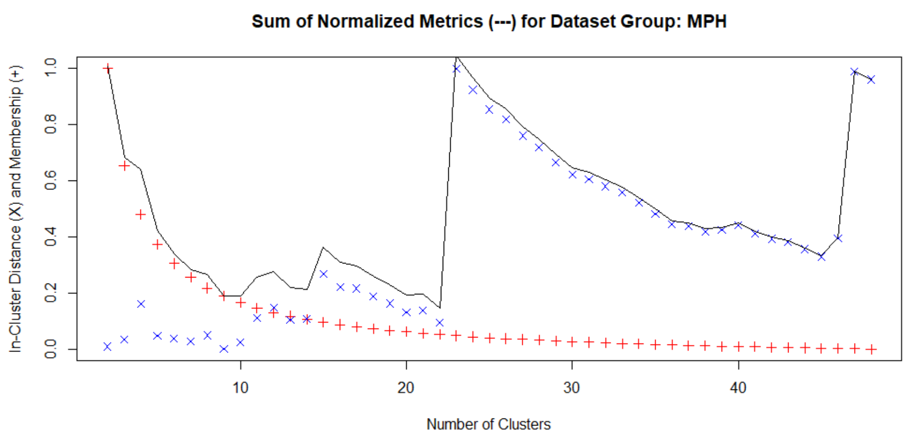

Figure 4.12: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the MPH data set group.

Table 4.6: Summary Statistics for the Core Area Produced by the MPH Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	2976784	0.99	67616	135308	263674
# Member Counties	3106	3034	0.98	43	138	227
# Member MPOs	387	377	0.97	2	17	56
Intersecting States	48	22	0.46	1	5	11
Total Population	328016242	310545448	0.95	1036040	14115702	53601202
Avg. Pop. Density	107	104	0.98	7	114	361
Avg. Household Income	65077	58274	0.9	48993	61162	76686
% Difference D/R Votes	-0.024	-0.015	0.61	-0.201	0.098	0.385
% Pop. White	0.727	0.73	1	0.601	0.777	0.894
# of Patents	1046391	1002289	0.96	1254	45559	257767
Avg. Dist to Post-Secondary	886	931	1.06	750	1005	1593

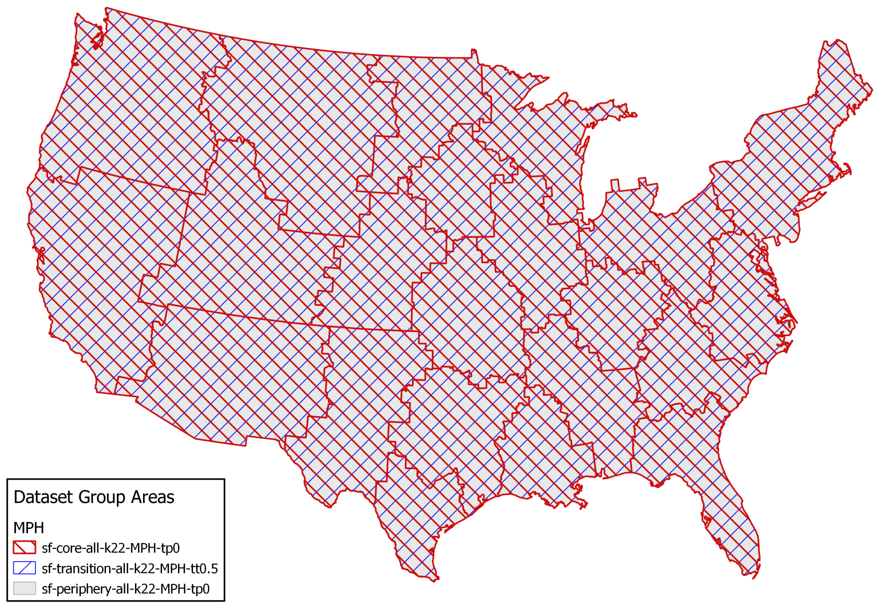

Figure 4.13: Map of megaregional core, transition, and periphery areas based on the clusters produced by the MPH dataset group.

4.1.7 Morphological-Societal-Human

Datasets in Category:

- Population
- Race & Ethnicity
- Medicare Reimbursement[1]

The first group with multiple data sources covered in this chapter, this category shows clear separation between several regions and a large transitional area in the Midwest from Michigan down into Tennessee, see 4.15. Later categories and groups will further refine these areas, but many of these emerge across several areas. This continued emergence lies at the basis of megaregional theory because it shows that communities can be connected in ways other than a continuous built environment. The cluster regions produced by each participating dataset are shown in Figures 4.16 through 4.18.

These regions contain 77% of the population in 63% of the area, resulting in a average population density ~25% higher than the national average, see Table 4.7. These areas also are responsible for 85% of patents and have a left-leaning voting pattern twice as strong as the national trend, but their residents tend to be farther from post-secondary education opportunities.

[1]It should be noted here that the data retreived was at the county level despite most of the datasets from Dartmouth being

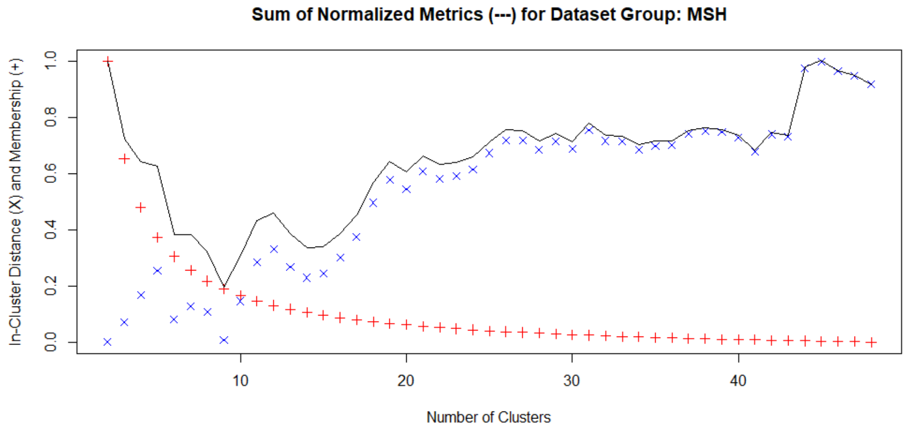

Figure 4.14: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the MSH data set group.

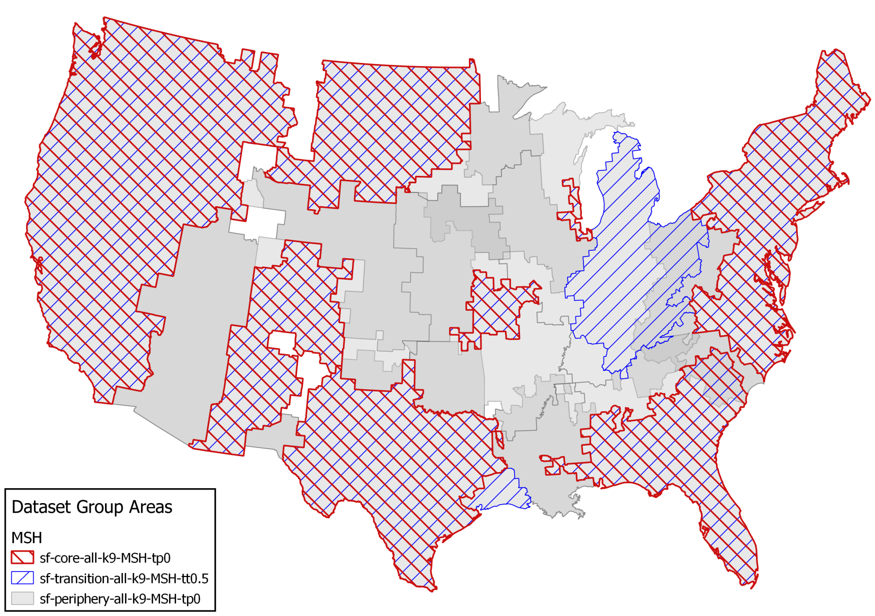

Figure 4.15: Map of megaregional core, transition, and periphery areas based on the clusters produced by the MSH dataset group.

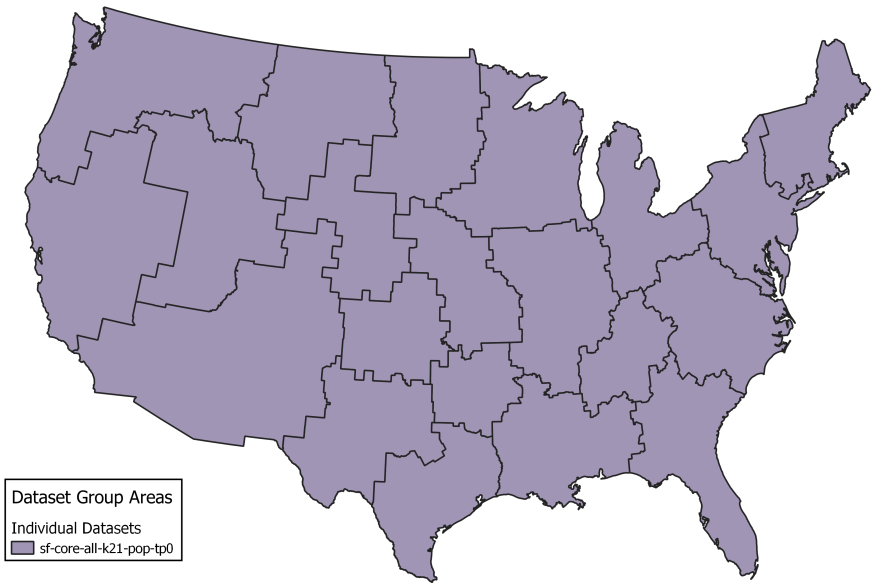

Figure 4.16: Map of clusters produced by the population dataset.

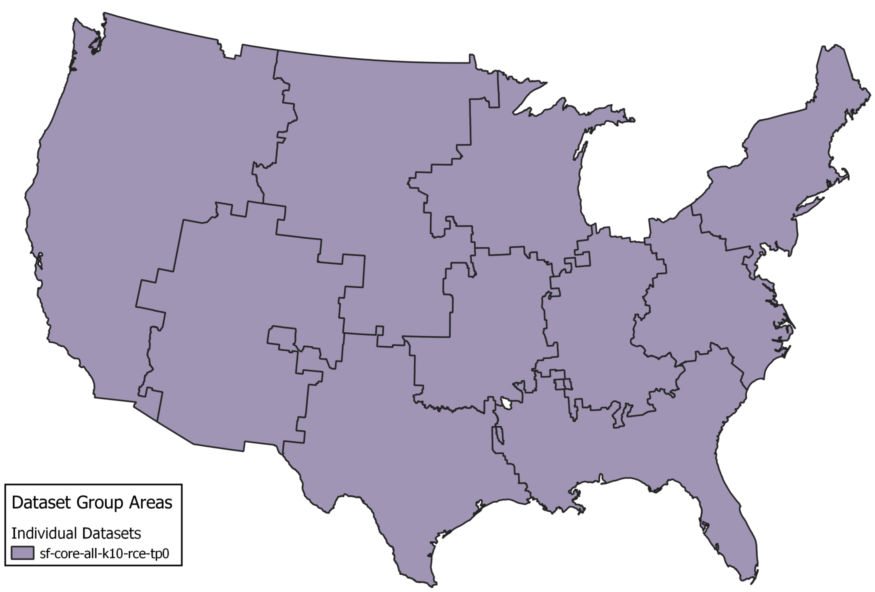

Figure 4.17: Map of clusters produced by the racial distribution dataset.

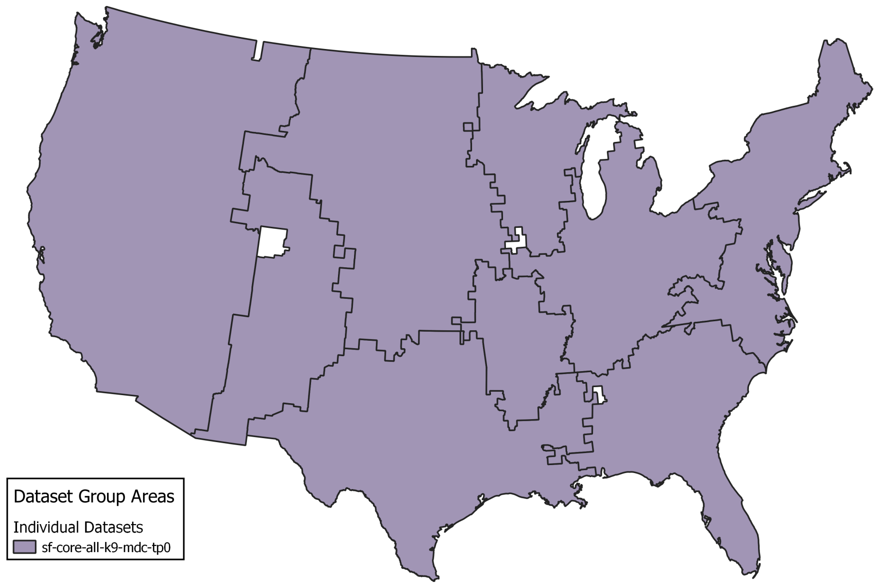

Figure 4.18: Map of clusters produced by the medicare reimbursement dataset.

Table 4.7: Summary Statistics for the Core Area Produced by the MSH Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	1692280	0.56	6974	211535	595622
# Member Counties	3106	1454	0.47	15	182	408
# Member MPOs	387	232	0.6	4	29	85
Intersecting States	48	8	0.17	3	6	15
Total Population	328016242	209299085	0.64	1690824	26162386	73741101
Avg. Pop. Density	107	124	1.16	8	215	864
Avg. Household Income	69607	66545	0.96	56326	67061	74785
% Difference D/R Votes	-0.024	-0.078	3.224	-0.164	0.032	0.251
% Pop. White	0.727	0.694	0.95	0.653	0.752	0.856
# of Patents	1046391	757261	0.72	2678	94658	306504
Avg. Dist to Post-Secondary	886	976	1.08	792	1024	1537

Table 4.8: Summary Statistics for the Transition Area Produced by the MSH Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	1891887	0.63	33073	236486	595622
# Member Counties	3106	1894	0.61	56	237	438
# Member MPOs	387	286	0.74	4	36	85
Intersecting States	48	8	0.17	3	7	15
Total Population	328016242	251466403	0.77	1690824	31433300	73741101
Avg. Pop. Density	107	133	1.24	8	137	313
Avg. Household Income	67508	62529	0.93	56326	65081	74785
% Difference D/R Votes	-0.024	-0.056	2.306	-0.164	0.034	0.251
% Pop. White	0.727	0.711	0.98	0.653	0.756	0.856
# of Patents	1046391	891641	0.85	2678	111455	306504
Avg. Dist to Post-Secondary	886	974	1.08	769	1021	1537

4.1.8 Morphological-Societal-Material

Datasets in Category:

- Köppen Climate Classification
- National Land Cover Database (NLCD)

While the NLCD does include values for developed land, these two datasets are the least concerned with individual members of society. Rather, they focus on the relationship between the built and natural environment. The separation between the Southeast coast the the Midwest / Great Lakes area is likely due in large part to the climate boundary there, similar to the transition from the great plains to the Rocky Mountains, see 4.20. The cluster regions produced by each participating dataset are shown in Figures 4.21 and 4.22.

Despite their indifferent stance on humanity, the variety of connections between populations and their living environment has been a growing field that continues to show strong indicators of impacts of one on the other. Therefore, these sources are yet another important facets of a holistic delineation process even though their appearance may not intuitively corroborate that narrative.

4.1.9 Morphological-Societal-Information

Datasets in Category:

- Harvard Patent Database
- Language Spoken at Home

focused on Hospital Service Areas and Hospital Referral Regions.

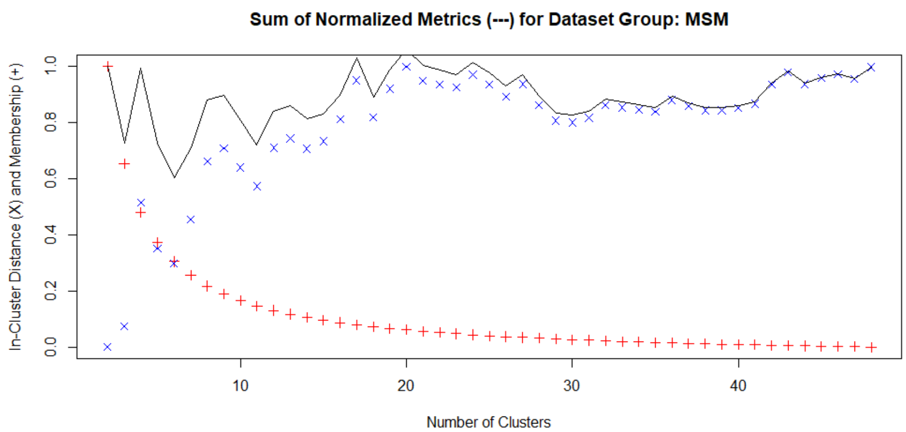

Figure 4.19: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the MSM data set group.

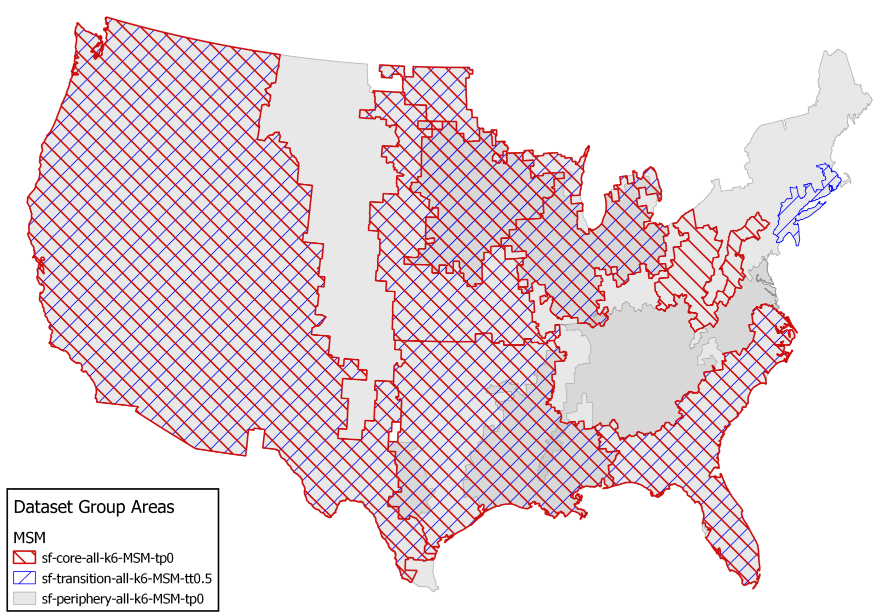

Figure 4.20: Map of megaregional core, transition, and periphery areas based on the clusters produced by the MSM dataset group.

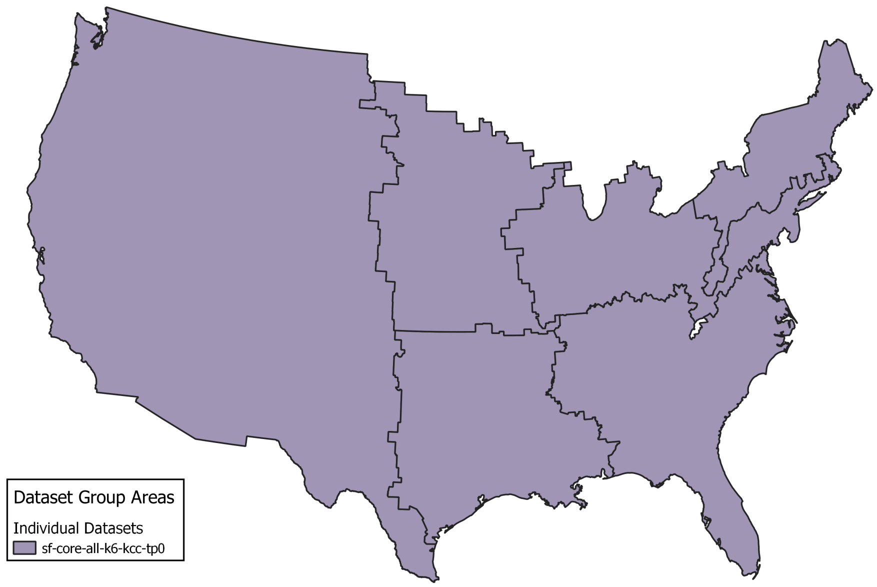

Figure 4.21: Map of clusters produced by the K"oppen Climate Classification dataset.

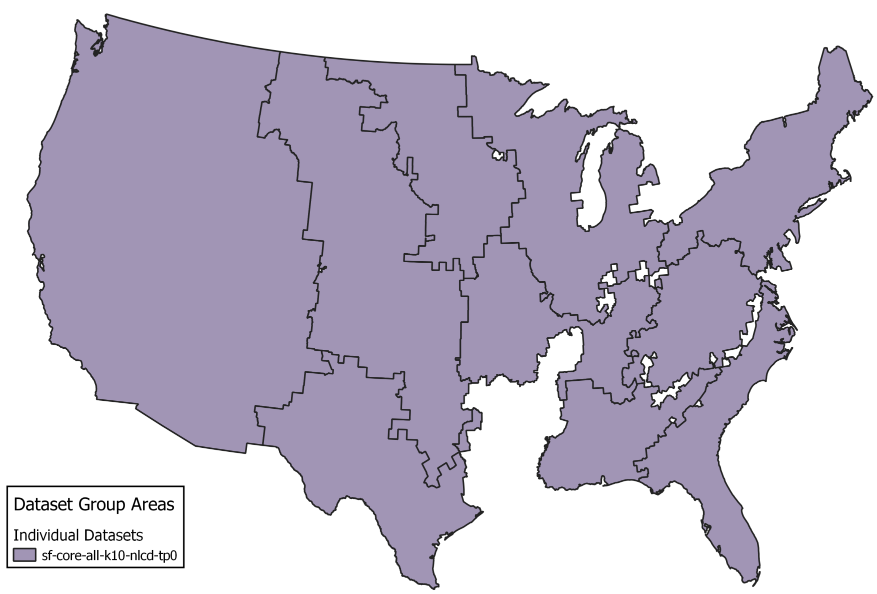

Figure 4.22: Map of clusters produced by the National Land Cover Database dataset.

Table 4.9: Summary Statistics for the Core Area Produced by the MSM Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	2344135	0.78	57820	390689	1135485
# Member Counties	3106	2240	0.72	135	373	540
# Member MPOs	387	255	0.66	18	42	68
Intersecting States	48	6	0.12	6	9	13
Total Population	328016242	196475832	0.6	9962104	32745972	73416426
Avg. Pop. Density	107	84	0.78	50	111	175
Avg. Household Income	63548	61163	0.96	53840	61096	71553
% Difference D/R Votes	-0.024	0.021	-0.862	-0.086	0.046	0.194
% Pop. White	0.727	0.744	1.02	0.69	0.78	0.872
# of Patents	1046391	590885	0.56	30052	98481	319025
Avg. Dist to Post-Secondary	886	893	1.01	805	954	1351

- Nativity & Citizenship
- Proximity to Post-secondary Education

Although the California and New York megaregions are present once again in these cluster regions, this group displays a much more distinct set of megaregions in the Midwest and Great Plains that balance large core areas with transitional zones, see Figure 4.24. These core areas are substantially more likely to vote for left-leaning candidates while the transition areas are more likely than the national average to vote towards the political right. The cluster regions produced by each participating dataset are shown in Figures 4.25 through 4.28.

The proximity to post-secondary education in Figure 4.28 shows a ripple-like pattern that may be a result of counties near the center of the country being closer to high densities of institutions along the east and west coasts. Distances were adjusted by dividing by the county's average distance to other counties, but the trend must be particularly strong. How exactly education deserts can be mapped to regions is a clear opportunity for research that could provide a bridge between two fields (education and community planning) that are not often well integrated in local governments due to the independence of school system boards from municipal governments.

Modern tech hubs like the Pacific Northwest and the Carolinas are conspicuously absent from the core areas. This could be related to their relatively recent prominence, but further study would be needed to clarify these relationships between time, population, and patent submission. Additionally, the eastern seaboard appears to be part of several periphery areas, and, with southern Florida not being a core or transition member, it may be that some of these port communities act more as distributors of immigrants rather than sinks.

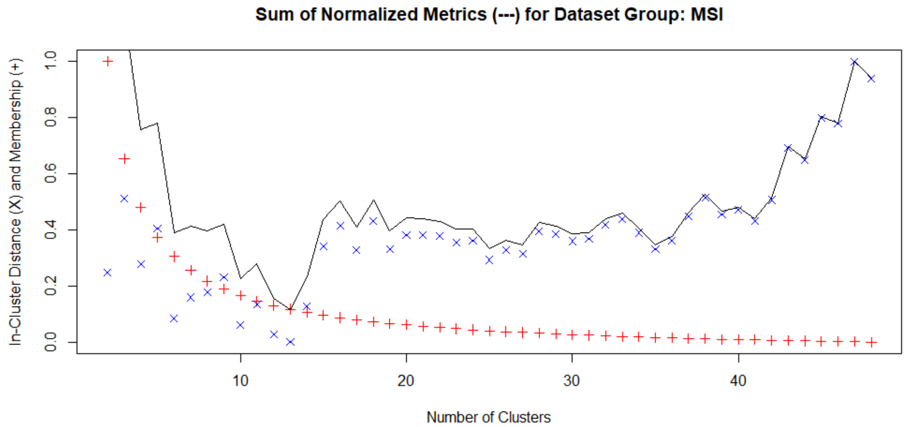

Figure 4.23: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the MSI data set group.

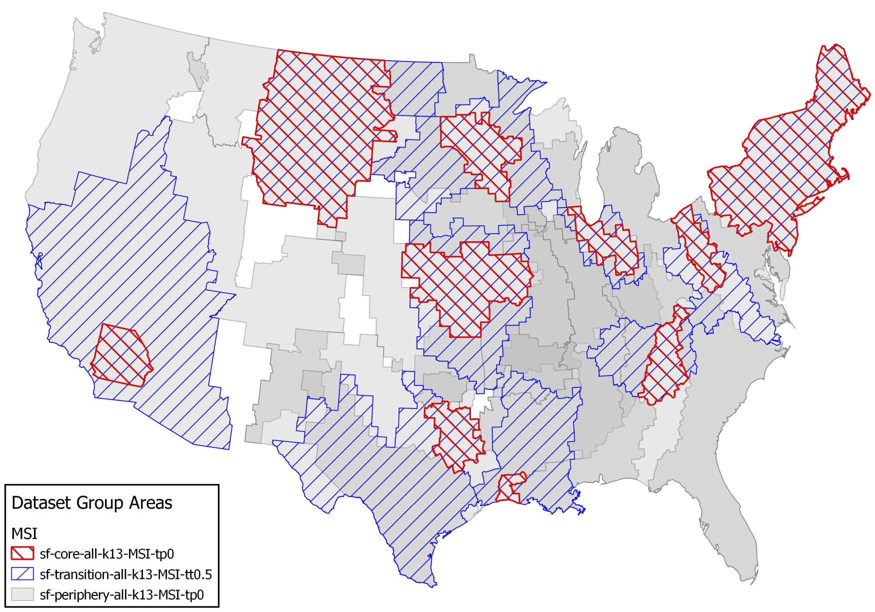

Figure 4.24: Map of megaregional core, transition, and periphery areas based on the clusters produced by the MSI dataset group.

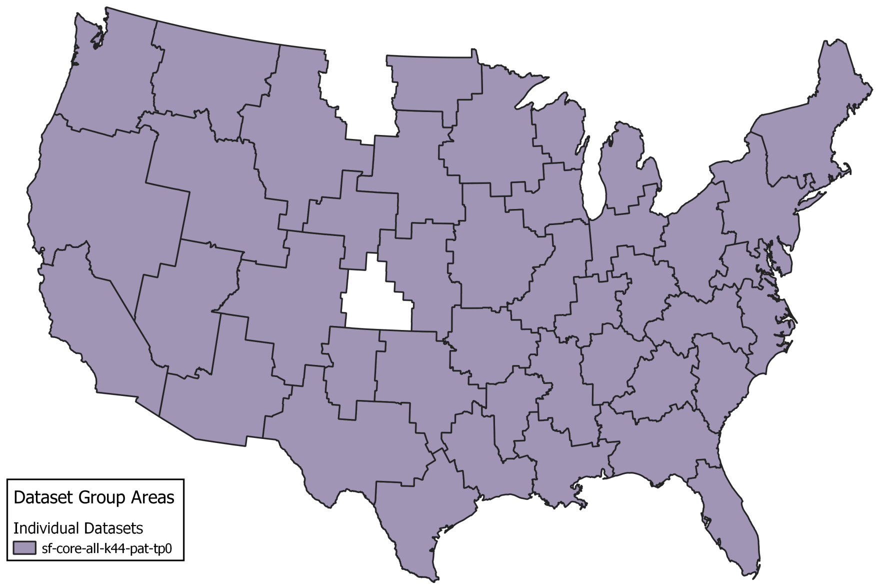

Figure 4.25: Map of clusters produced by the patent dataset.

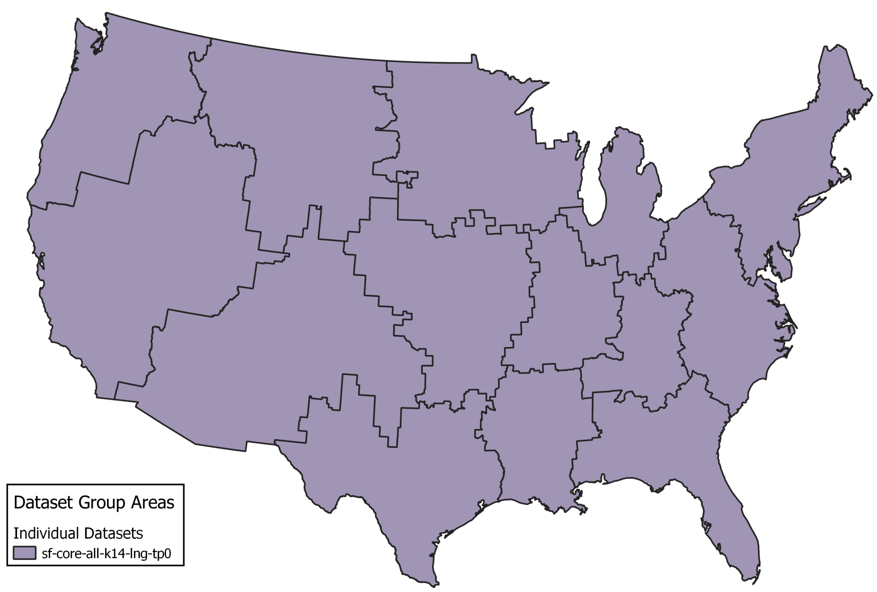

Figure 4.26: Map of clusters produced by the language spoken at home dataset.

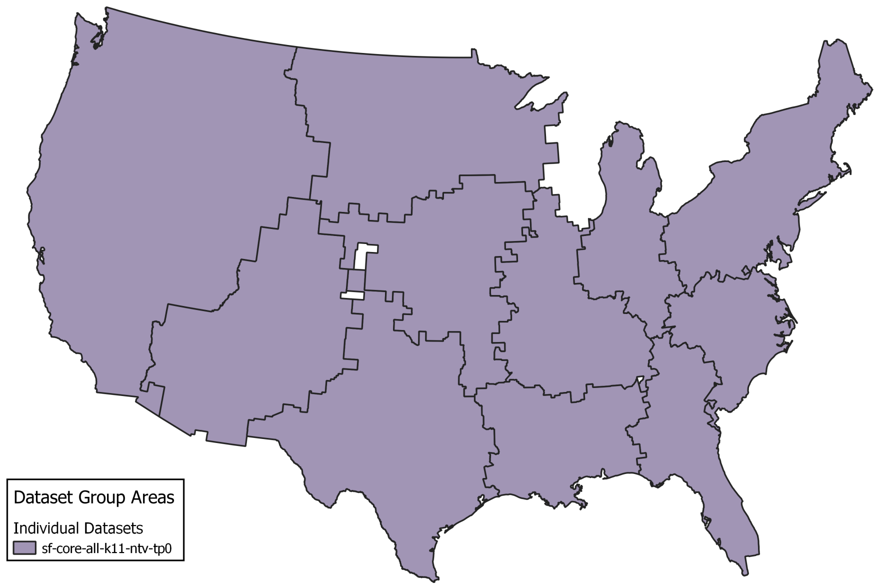

Figure 4.27: Map of clusters produced by the nativity and citizenship dataset.

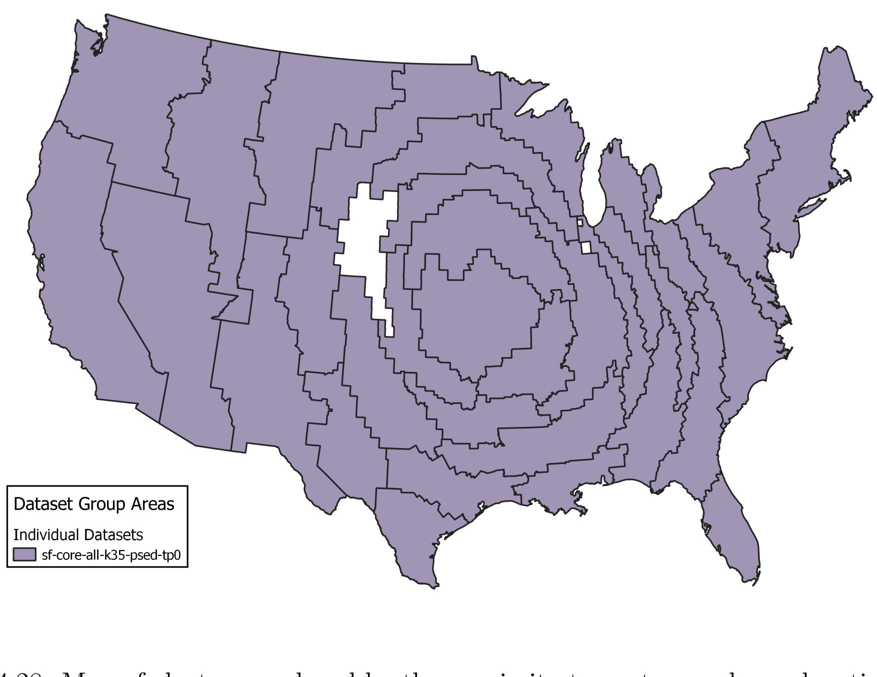

Figure 4.28: Map of clusters produced by the proximity to post-secondary education dataset.

Table 4.10: Summary Statistics for the Core Area Produced by the MSI Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	571311	0.19	5348	57131	184325
# Member Counties	3106	640	0.21	2	64	182
# Member MPOs	387	93	0.24	1	9	52
Intersecting States	48	10	0.21	1	4	10
Total Population	328016242	95638476	0.29	304799	9563848	50395244
Avg. Pop. Density	107	167	1.57	6	211	509
Avg. Household Income	70123	63970	0.91	49074	64083	74878
% Difference D/R Votes	-0.024	-0.127	5.219	-0.208	0.05	0.381
% Pop. White	0.727	0.718	0.99	0.605	0.745	0.874
# of Patents	1046391	390034	0.37	186	39003	248260
Avg. Dist to Post-Secondary	886	921	1.02	753	963	1461

Table 4.11: Summary Statistics for the Transition Area Produced by the MSI Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	1567597	0.52	37748	156760	402715
# Member Counties	3106	1540	0.5	84	154	250
# Member MPOs	387	199	0.51	5	20	52
Intersecting States	48	10	0.21	3	6	10
Total Population	328016242	195092845	0.59	1339772	19509284	50395244
Avg. Pop. Density	107	124	1.16	6	161	452
Avg. Household Income	67678	62854	0.93	49287	62969	74878
% Difference D/R Votes	-0.024	-0.064	2.646	-0.207	0.047	0.345
% Pop. White	0.727	0.712	0.98	0.615	0.753	0.871
# of Patents	1046391	746118	0.71	2463	74612	248260
Avg. Dist to Post-Secondary	886	909	1.03	754	971	1504

4.2 Subject Groups

At this stage, category datasets are combined by subject: economic, political, societal. How much various interactions alter interim steps and final outcomes may indicate the degree of similarity between categories and their underlying phenomena.

4.2.1 Functional-Economic

Datasets in Category:

- LEHD LODES
- Commodity Flow Survey

As shown in Figure 4.30, the combination of commuter and freight flows develops two large sets of relatively continuous core/transition areas along the northwest and southeast edges of the country. The underlying reasons for this pattern may be illuminated by studies that add data sources such as ports of entry and expand the already implemented datasets by industry. Exploration of a number of clusters greater than those used in this work may also provide further clarity.

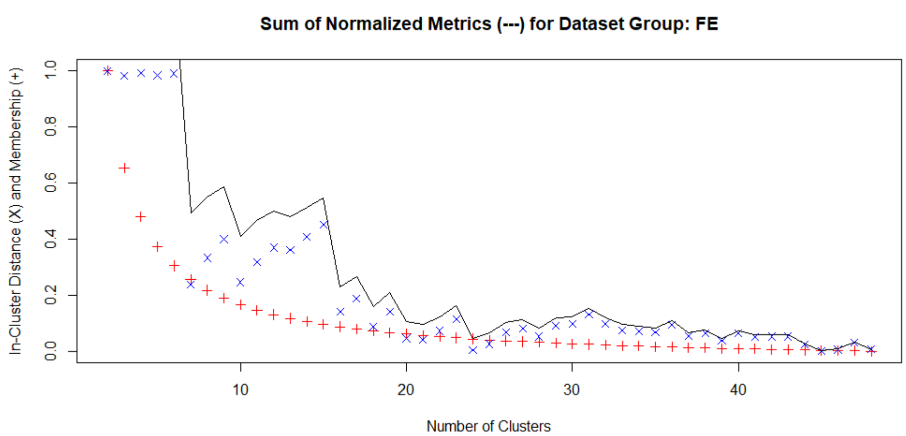

Figure 4.29: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the FE data set group.

4.2.2 Functional-Societal

The results for this section are currently identical to the Functional-Societal-Human category as no flow datasets have been identified for the other categories under this subject.

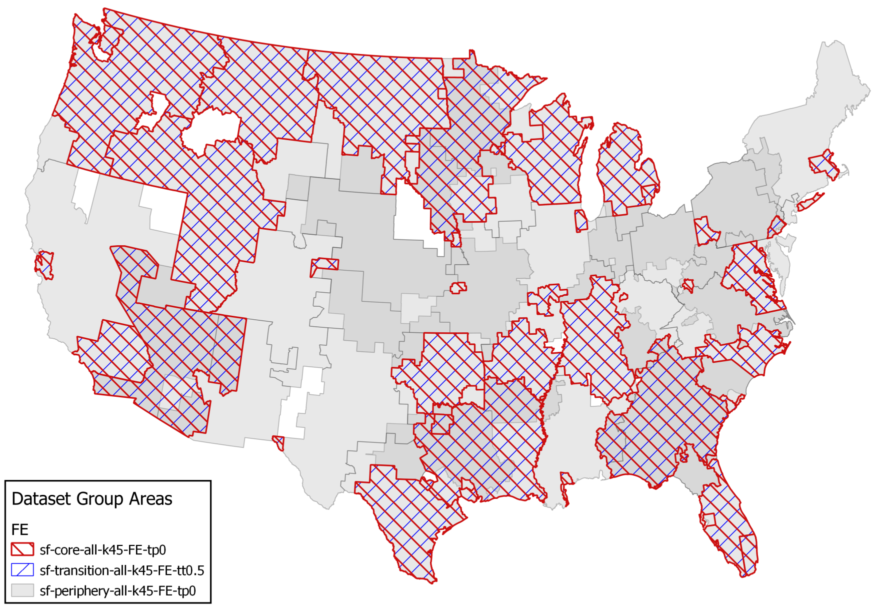

Figure 4.30: Map of megaregional core, transition, and periphery areas based on the clusters produced by the FE dataset group.

Table 4.12: Summary Statistics for the Core Area Produced by the FE Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	1481213	0.49	398	37030	166253
# Member Counties	3106	1458	0.47	1	36	236
# Member MPOs	387	198	0.51	1	5	28
Intersecting States	48	32	0.67	1	2	7
Total Population	328016242	193594444	0.59	183279	4839861	17810084
Avg. Pop. Density	107	131	1.22	7	752	3181
Avg. Household Income	67104	62658	0.93	43876	65887	111360
% Difference D/R Votes	-0.024	-0.049	2.033	-0.5	0.05	0.593
% Pop. White	0.727	0.705	0.97	0.465	0.745	0.902
# of Patents	1046391	662450	0.63	104	16561	110582
Avg. Dist to Post-Secondary	886	928	1.09	740	1028	1720

4.2.3 Morphological-Economic

Datasets in Category:

- Household Incomes
- County Business Patterns

Observing personal and corporate financial data together develops four tightly connected megaregions encircling the northern plains and the Midwest, see Figure 4.32. This could be a reflection of the collapse of manufacturing in the US since World War 2, and historical data may be able to see if this trend does indeed develop over time. These regions represent an interaction between employer locations and employee residences. Businesses located within these regions may be leveraging economies of scale and proximity to other companies within their industry. Knowledge of these megaregional scale could be useful for government in creating economic development plans by having a better understanding of the kinds of corporations that are more likely to be located in their communities. Private organizations would also benefit by having a reduced research load for placing new offices, relocating, or identifying possible labor force locations for recruitment efforts.

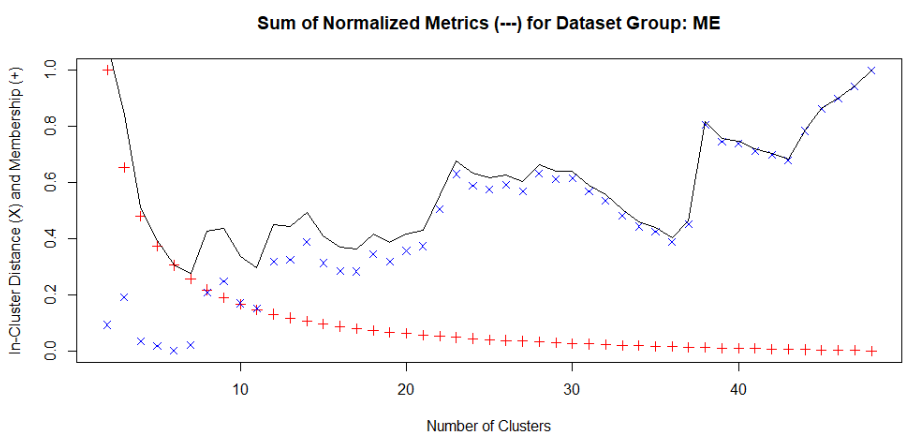

Figure 4.31: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the ME data set group.

4.2.4 Morphological-Political

This section experiences the same issue as Functional-Societal. However, it should be noted that the delineation process inherently considers some political aspects in the form of county and MPO boundaries.

4.2.5 Morphological-Societal

Datasets in Category:

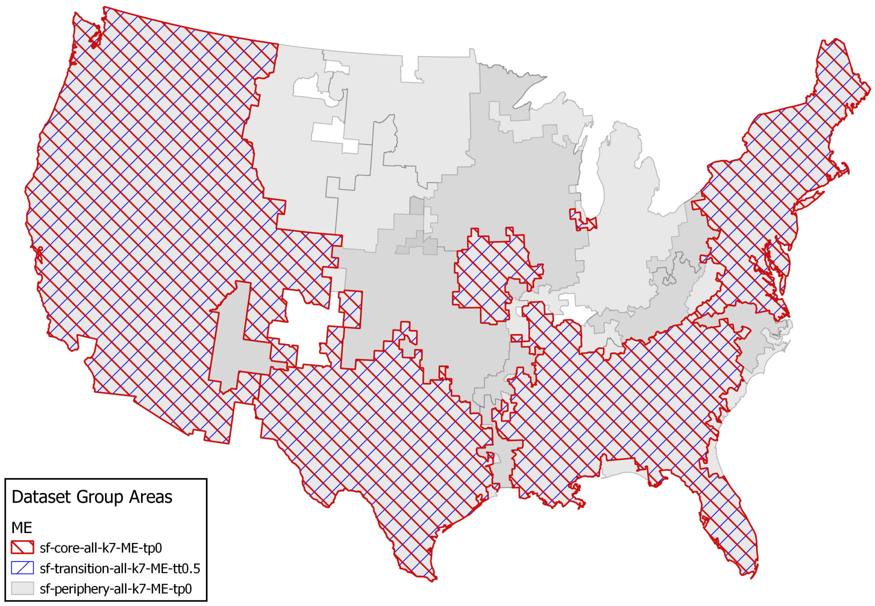

Figure 4.32: Map of megaregional core, transition, and periphery areas based on the clusters produced by the ME dataset group.

Table 4.13: Summary Statistics for the Core Area Produced by the ME Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	1820610	0.6	2635	303435	854150
# Member Counties	3106	1676	0.54	5	279	633
# Member MPOs	387	255	0.66	2	42	86
Intersecting States	48	6	0.12	2	8	15
Total Population	328016242	220378816	0.67	2804094	36729803	66526846
Avg. Pop. Density	107	121	1.13	69	294	1064
Avg. Household Income	67519	68262	1.01	54556	67983	80848
% Difference D/R Votes	-0.024	-0.032	1.299	-0.158	0.018	0.184
% Pop. White	0.727	0.711	0.98	0.689	0.736	0.832
# of Patents	1046391	753889	0.72	6467	125648	291431
Avg. Dist to Post-Secondary	886	903	1.03	767	972	1441

Table 4.14: Summary Statistics for the Transition Area Produced by the ME Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	1820610	0.6	2635	303435	854150
# Member Counties	3106	1676	0.54	5	279	633
# Member MPOs	387	255	0.66	2	42	86
Intersecting States	48	6	0.12	2	8	15
Total Population	328016242	220378816	0.67	2804094	36729803	66526846
Avg. Pop. Density	107	121	1.13	69	294	1064
Avg. Household Income	67519	68262	1.01	54556	67983	80848
% Difference D/R Votes	-0.024	-0.032	1.299	-0.158	0.018	0.184
% Pop. White	0.727	0.711	0.98	0.689	0.736	0.832
# of Patents	1046391	753889	0.72	6467	125648	291431
Avg. Dist to Post-Secondary	886	903	1.03	767	972	1441

- Population
- Race & Ethnicity
- Medicare Reimbursement
- Köppen Climate Classification
- National Land Cover Database
- Harvard Patent Database
- Language Spoken at Home
- Nativity & Citizenship
- Proximity to Post-secondary Education

This is by far the largest subject group, presenting a significant amount of variety in terms of data detail, source, and type; however, the four largest regions that appeared in the morphological-economic category are still represented. As mentioned previously, the ordering of MPOs by population during the region identification process may lead to a portion of the repeated emergence, but this is counteracted by the connection requirements of the core and transition areas. Opportunity attracts seekers, and this is supported by economic forces and the regional trends shown here. At this scale of background data, a comparison of the results in this section to other datasets such as linguistic variations may be of interest for future research, especially if these boundaries represent regions with strong cultural identities.

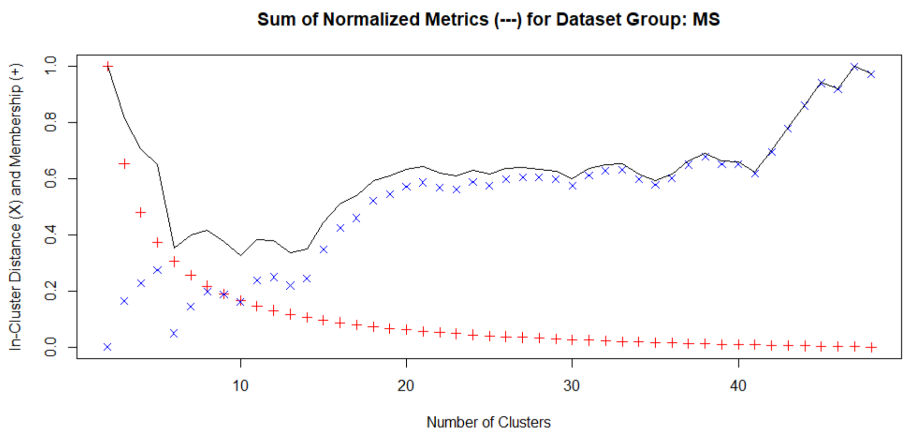

Figure 4.33: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the MS data set group.

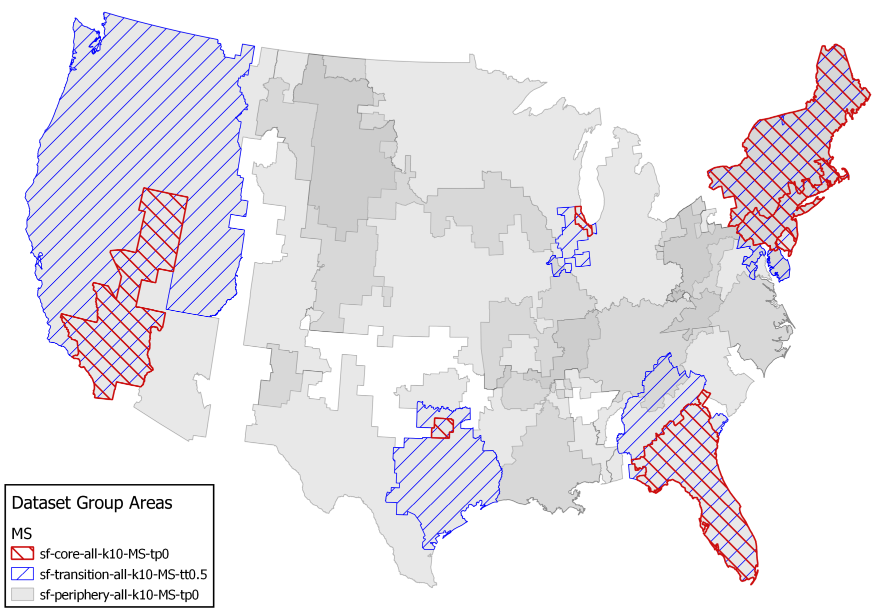

Figure 4.34: Map of megaregional core, transition, and periphery areas based on the clusters produced by the MS dataset group.

Table 4.15: Summary Statistics for the Core Area Produced by the MS Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	376588	0.12	2268	62765	143449
# Member Counties	3106	408	0.13	4	68	184
# Member MPOs	387	121	0.31	1	20	57
Intersecting States	48	6	0.12	1	5	11
Total Population	328016242	149573156	0.46	6561612	24928859	52396795
Avg. Pop. Density	107	397	3.71	218	1185	3225
Avg. Household Income	69461	71148	1.02	55227	70141	78927
% Difference D/R Votes	-0.024	-0.18	7.383	-0.371	-0.073	0.525
% Pop. White	0.727	0.675	0.93	0.587	0.662	0.738
# of Patents	1046391	604718	0.58	19975	100786	255329
Avg. Dist to Post-Secondary	886	976	1.08	772	1018	1460

Table 4.16: Summary Statistics for the Transition Area Produced by the MS Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	1177139	0.39	16893	196190	620427
# Member Counties	3106	1037	0.33	27	173	258
# Member MPOs	387	237	0.61	10	40	61
Intersecting States	48	5	0.1	1	7	12
Total Population	328016242	236743672	0.72	10760894	39457279	58776086
Avg. Pop. Density	107	201	1.88	95	322	637
Avg. Household Income	70932	72383	1.02	57091	70114	77007
% Difference D/R Votes	-0.024	-0.133	5.488	-0.244	-0.096	0.157
% Pop. White	0.727	0.693	0.95	0.662	0.693	0.712
# of Patents	1046391	979269	0.94	43692	163212	283533
Avg. Dist to Post-Secondary	886	992	1.09	763	1032	1524

4.3 Data Type Divisions

4.3.1 Functional

Datasets in Category:

- LEHD LODES
- Commodity Flow Survey
- IRS Migration

Despite this kind of data source being less common and available, the level of detail in connections between counties is much higher than the non-euclidean distances calculated between morphological observations. Variation among core area sizes is high in this group, but there are many areas where the distribution of major metropolitan areas can easily be inferred. Several large core areas do not display the contrasting symbology indicating transition areas, and it appears that this is connected to the root cause of the size variability already mentioned.

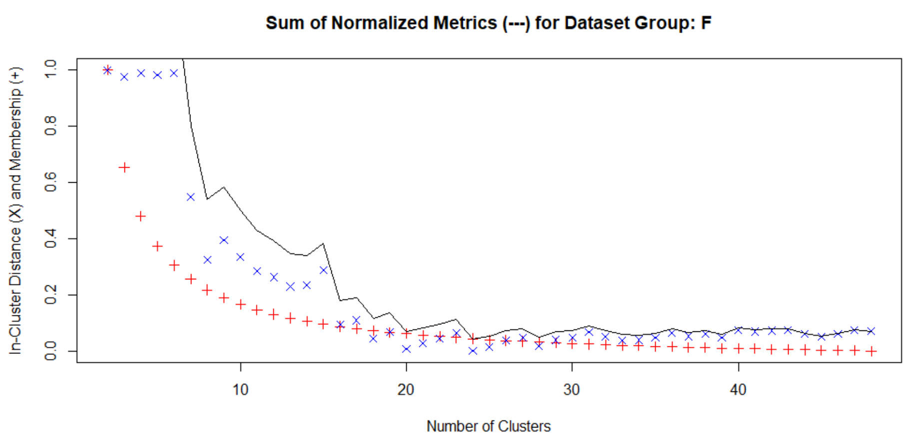

Figure 4.35: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the F data set group.

4.3.2 Morphological

Datasets in Category:

- Household Incomes
- County Business Patterns
- Presidential Voting Patterns
- Population

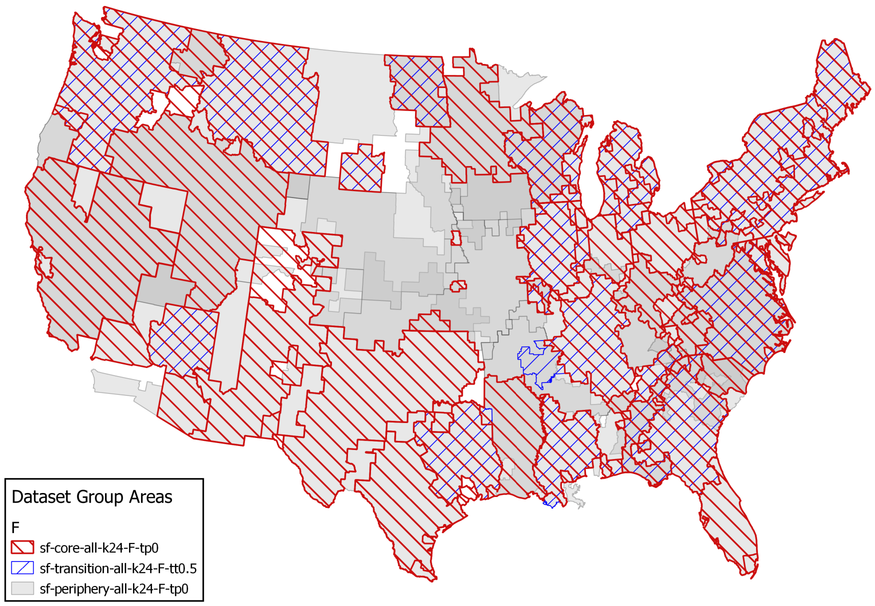

Figure 4.36: Map of megaregional core, transition, and periphery areas based on the clusters produced by the F dataset group.

Table 4.17: Summary Statistics for the Core Area Produced by the F Dataset Group

Factor	Nationally	Megaregions	Ratio	Min	Mean	Max
Total Area (sqmi)	3015119	5279226	1.75	912	72318	179881
# Member Counties	3106	5827	1.88	1	80	275
# Member MPOs	387	1006	2.6	1	14	59
Intersecting States	48	41	0.85	1	3	9
Total Population	328016242	758907220	2.31	183279	10395989	50881915
Avg. Pop. Density	107	144	1.34	8	306	3181
Avg. Household Income	65095	63223	0.97	46639	63235	94974
% Difference D/R Votes	-0.024	-0.064	2.647	-0.374	0.045	0.593
% Pop. White	0.727	0.734	1.01	0.567	0.758	0.916
# of Patents	1046391	2774850	2.65	104	38012	246043
Avg. Dist to Post-Secondary	886	953	1.09	752	1029	1715

- Race & Ethnicity
- Medicare Reimbursement
- Köoppen Climate Classification
- National Land Cover Database
- Harvard Patent Database
- Language Spoken at Home
- Nativity & Citizenship
- Proximity to Post-secondary Education

The primary areas already mentioned in the Northeast, Southeast, and West Coast maintain their form, but the small core and transition areas across the center of the country display how little structure can be gleaned from less homogeneous communities in these areas. Unfortunately, the corollary to this trend may be that the Rust Belt and old manufacturing centers will continue to struggle to stabilize many years into the future unless the constituent communities are able to come together to form a firm identity that planning efforts can use as a lodestar for their work. A Denver-base megaregion does appear uniquely distinct in this situation. Its connection to the plains to the east may be a relationship worth further investigation.

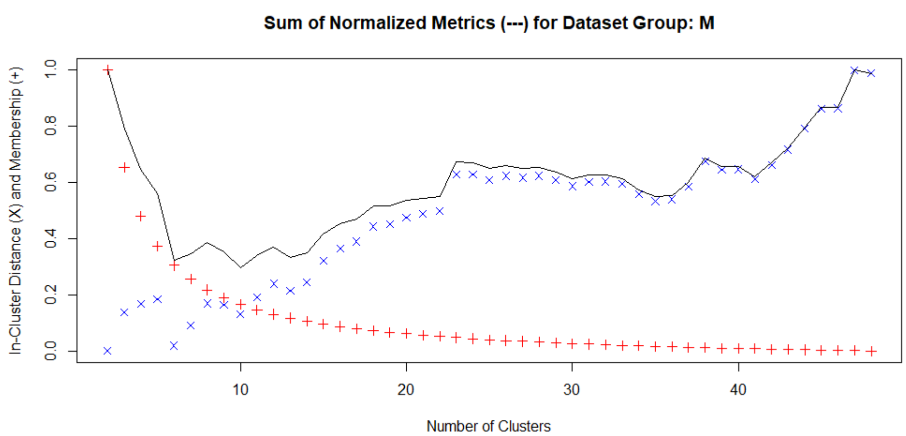

Figure 4.37: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the M data set group.

4.4 Complete-set Results

The merging of morphological and functional data poses several issues as most methods of analysis cannot handle both data types. Fortunately, performing cluster analysis outside the file shape constraints of a spatial file type such as the ESRI Shapefile allows for greater flexibility when calculating distances between

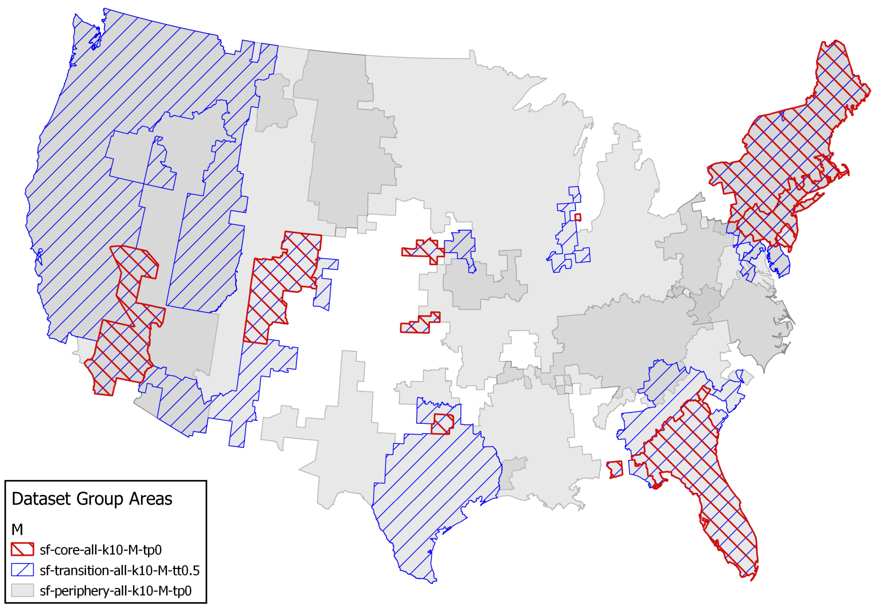

Figure 4.38: Map of megaregional core, transition, and periphery areas based on the clusters produced by the M dataset group.

Table 4.18: Summary Statistics for the Core Area Produced by the M Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	378683	0.13	336	37868	143415
# Member Counties	3106	441	0.14	1	44	183
# Member MPOs	387	125	0.32	1	12	57
Intersecting States	48	10	0.21	1	3	11
Total Population	328016242	131835983	0.4	162518	13183598	50764802
Avg. Pop. Density	107	348	3.25	29	704	2762
Avg. Household Income	69502	72416	1.04	55232	68912	92809
% Difference D/R Votes	-0.024	-0.135	5.553	-0.217	0.103	0.525
% Pop. White	0.727	0.698	0.96	0.632	0.749	0.903
# of Patents	1046391	545574	0.52	495	54557	248313
Avg. Dist to Post-Secondary	886	947	1.05	772	992	1465

Table 4.19: Summary Statistics for the Transition Area Produced by the M Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	1265887	0.42	1898	126589	536183
# Member Counties	3106	1116	0.36	2	112	250
# Member MPOs	387	259	0.67	1	26	63
Intersecting States	48	10	0.21	1	5	13
Total Population	328016242	232793460	0.71	488246	23279346	56063722
Avg. Pop. Density	107	184	1.72	89	214	360
Avg. Household Income	69817	65539	0.94	56445	66154	75738
% Difference D/R Votes	-0.024	-0.095	3.917	-0.172	0.047	0.336
% Pop. White	0.727	0.713	0.98	0.688	0.749	0.852
# of Patents	1046391	946659	0.9	556	94666	263689
Avg. Dist to Post-Secondary	886	959	1.06	768	1004	1534

observations and their associated clusters. Then a direct comparison can be more easily made as the exact same process is used in both cases.

In terms of the core and transition areas shown in Figure 4.40, there are several areas that contain only one MPO. These could be eliminated from the megaregional discussion since they are already managed by an MPO, but they were included in these results for the sake of completeness. Modifications to the border-hole-filling parameters and the relationships of transition areas may further refine these results.

Nevertheless, the patterns from below on the typological tree appear here with clear connections to various data sources. New York, Southern California, Atlanta, and the Texas Triangle are major anchors for megaregions. The Midwest is highly connected to major economic players, but struggles to produce a clear cluster identity. The Great Plains are largely absent from megaregional cores and transition areas.

The summary statistics, Tables 4.20 and 4.21, show how dominant these areas are. Megaregion core areas contain over 20% of the nation's population in 5% of its area, resulting in a density four-and-a-half times higher than national average. Expanding to the transition areas, these cover approximately half of the area, counties, and states of the contiguous United States, and contain ~85% of the population and MPOs and 95% of patents at nearly twice the population density of the country on average. This is achieved with an average racial distribution and distance to post-secondary education.

4.5 Impacts & Opportunities

While this work's methodology for megaregional delineation seeks to be as quantitatively focused as possible, qualitative decisions are nearly inescapable at the crossroads of government and data science. Both facets of

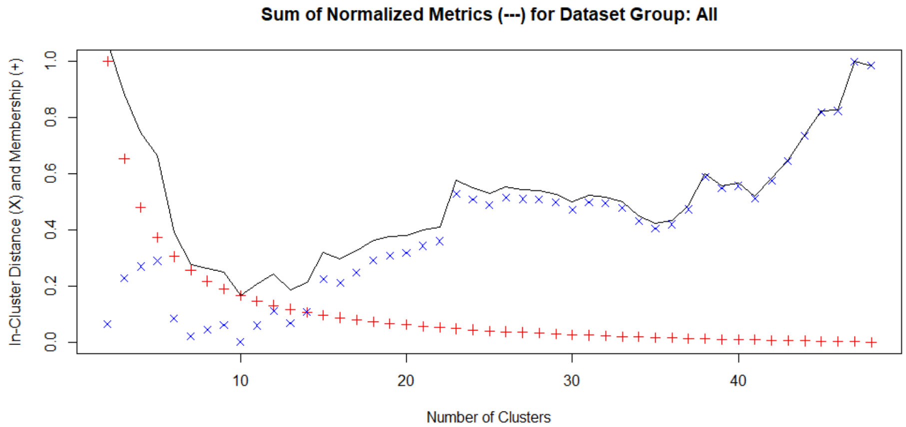

Figure 4.39: Graph of average cluster membership, the mean of average in-cluster distances, and their sum by the number of clusters for the all data set group.

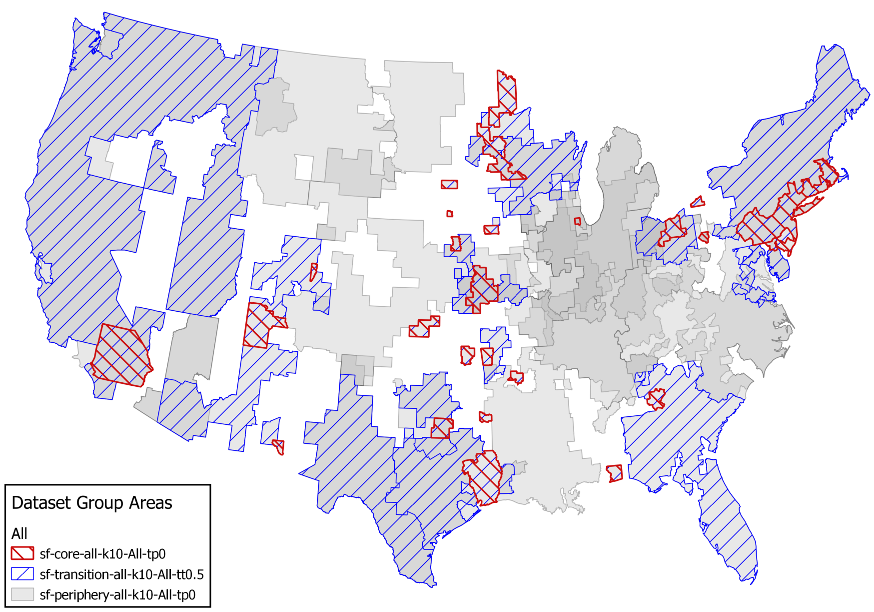

Figure 4.40: Map of megaregional core, transition, and periphery areas based on the clusters produced by the entire dataset group.

Table 4.20: Summary Statistics for the Core Area Produced by the Entire Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	139640	0.05	267	5371	27421
# Member Counties	3106	176	0.06	1	7	62
# Member MPOs	387	54	0.14	0	2	27
Intersecting States	48	19	0.4	1	1	6
Total Population	328016242	68504454	0.21	20173	2634787	34483224
Avg. Pop. Density	107	491	4.59	19	597	2762
Avg. Household Income	72835	61248	0.84	46871	63304	92809
% Difference D/R Votes	-0.024	-0.134	5.526	-0.253	0.079	0.593
% Pop. White	0.727	0.676	0.93	0.45	0.758	0.923
# of Patents	1046391	277660	0.27	7	10679	168414
Avg. Dist to Post-Secondary	886	878	0.98	772	923	1461

Table 4.21: Summary Statistics for the Transition Area Produced by the Entire Dataset Group

Factor	**Nationally**	**Megaregions**	**Ratio**	**Min**	**Mean**	**Max**
Total Area (sqmi)	3015119	1560121	0.52	267	62405	503269
# Member Counties	3106	1527	0.49	1	61	241
# Member MPOs	387	331	0.86	0	13	65
Intersecting States	48	23	0.48	1	3	13
Total Population	328016242	276116619	0.84	20173	11044665	57357994
Avg. Pop. Density	107	177	1.65	75	354	2762
Avg. Household Income	67586	61832	0.91	45545	62652	92809
% Difference D/R Votes	-0.024	-0.076	3.113	-0.173	0.061	0.396
% Pop. White	0.727	0.729	1	0.688	0.779	0.869
# of Patents	1046391	1005899	0.96	7	40236	268070
Avg. Dist to Post-Secondary	886	873	0.98	772	925	1531

data science have roles to play here. Quantitative tools have become increasingly competent with the advent of "big data" and proto-AI tools like machine learning. If government entities expand their collection and distribution of highly detailed observations of our built, social, and natural environment, cluster analysis will give undoubtedly give way to AI techniques to extract patterns too subtle and relationships too complex for human recognition. Until that point, qualitative methods can fill the gaps left by their quantitative counterparts. Nevertheless, there are many pitfalls in the use of relying on human intuition and intention in a process that could impact the governance and life quality of large populations. The gerrymandering of political boundaries is a prime example of this. Therefore, qualitative processes should be strictly confined to the places where empirical techniques do not yet reach.

Based on the results reviewed in this chapter, several of these gaps emerged, where answers to question raised further question with varying degrees of related-ness. The following is a brief list of places for subsequent work not already covered in the body of this chapter.

- When filling holes during the contiguity checks after the cluster analysis and region identification steps, should non-included areas adjacent to the border be considered "holes?"
- How do different MPO ordering methods alter outcomes? Do the same regions still appear?
- Does preventing the formation of core or transition areas substantially impact results?
- Resolving megaregions that share significant amounts of their transition areas, See the two Texas cores in Figure 4.40.
- How would altering the organization of the typological tree by swapping the position of "Functional|Morphological" with "Economic|Political|Societal" affect the middle layers of delineation results?

CHAPTER 5

Discussion

This chapter addresses two topics in addition to a review of patterns seen in the Results chapter: the applicability of megaregions to community planning sub-fields, with data source suggestions; and the societal need for megaregional government, implemented through the lens of small state political theory. Many small to medium sized municipalities face a disparity between available human resources and the day-to-day demands of local government, and often little remains for addressing upstream concerns. The result is staff making marginal progress on long-term solutions, only raising the levels of equity and quality of life in their communities when it suits the agenda presented by short-term issues and interests. Formal coordination at a megaregional level has the potential to improve government efficiency and efficacy by eliminating the duplication of these marginal efforts across multiple municipalities and leveraging agglomerated resources to address challenges that prevent progress on larger time and spatial scales.

Unfortunately, erosion of positive place-based identity in the United States has created an environment which makes such an implementation difficult as metropolitan areas, municipal leaders, and even some state officials may see megaregional government at best as an undue bureaucratic burden or at worst a direct threat to their sovereign independence and position of power. Therefore, the creation of megaregional government bodies must be careful to correctly incentivize support by focusing on the inherent interests of the parties in question. Small state political theory was initially developed in the twentieth century to analyze the behavior of small nation states in global politics, but its principles can be applied to local by simply the language referring to governments (e.g. “nation” to “city” or “state” to “county”). Megaregions can effectively bridge inter-regional gaps and integrate the interests of urban and rural populations, making them an ideal vehicle to deliver the proverbial antidotes to tribal isolation while produces net benefits for society as a whole.

5.1 Patterns from Results

Although the were several shortcomings with the process as discussed at the end of the Results chapter, there are several strong patterns that could be expected to persist through process refinement and dataset expansion. These refined megaregional boundaries will likely continue enclose the five to six primary areas from the initial results.

1. The West Coast
2. The Northeast
3. Florida-Georgia
4. The Texas Triangle
5. A western portion of the Midwest (tentative)
6. Kansas City (tentative)

5.1.1 The West Coast

When a higher number of clusters is selected by the delineation process, the largest portion of this region is typically subdivided into three sub-areas: California, with a centroid that shifts north-south depending on the data sources under consideration; the Pacific Northwest, with the eastern portions of Oregon and Washington participating to varying degrees; and the northern portion of Utah. Boundary sets with lower cluster numbers will add metropolitan areas from states such as Idaho, Arizona, New Mexico, and even Montana. A connection inland may be attributable to narratives about the coastal states centered around governance and costs of living, and a point inward can clearly be seen in Figure 4.7.

An important point for a megaregional government in this part of the county will be developing and incentivizing programs that appeal to a wide array political and ideological priorities. Even though this could be said of any body attempting to tie regions together, the longer distances between population centers in the western US when compared with the east makes knowledge spillover and other such physical-social phenomena less likely to occur in physical spaces.

5.1.2 The Northeast

The string of metropolitans along the Bos-Wash corridor formed a clear grouping in every category combination except the functional-economic; even then, several metros still emerged as small, but unconnected, core areas. With a body of literature on the region stretching back more than 50 years, this result was expected, but its presence does support the validity of the process by its ability to repeat the result. Depending on the data utilized, transition and periphery areas frequently spilled over in the Midwest and Southern states.

5.1.3 Florida-Georgia

Much like the Northeast, the Florida-Georgia megaregion, led by Atlanta Regional Commission MPO, was present in most delineations. Its variability was largely dependent on how strongly areas of states such as the Carolinas, Alabama, Mississippi, and Tennessee were connected to the Texas Triangle and Northeast megaregions. As mentioned before, the region is conspicuously absent in the morphological-societal-information delineation. A grasp of why will require subsequent research. Additionally, when splits occurred in the area, they frequently separated Miami-Dade from Atlanta. These two metropolitans represent substantial economic influence and population attraction, and they will form core aspects of any megaregional organization for this portion of the US.

5.1.4 The Texas Triangle

The last of the strong clusters, the influence of the "Texas Triangle"[1] cities was often present throughout the central US and even in the Midwest. Several clustering solutions separated the state into multiple areas, but, more often, they represented a unified block. In the complete set of clusters, two points of the triangle even form separate core regions but still share a significant portion of their transition areas, see Figure 4.40. Process refinement concerned overlapping transition areas may result in the re-merging of these two megaregions into a polycentric whole.

5.1.5 West-Midwest

The metropolitan areas in the Midwest formed a variety of clusters during the initial category stages, and it may be this that reduced their ability to form a distinct region in many of the settings with more data sources. Combined with the collapse of US-based traditional manufacturing centers over the last century, the weaker connection to the Bos-Wash corridor in the western portion of the region may mean that these metropolitans have the best capacity to form the core of a megaregion. However, an expansion of the data sources to fill out the typological tree may allow for this region to emerge more consistently.

5.1.6 Kansas City

This megaregion does not appear as consistently as the ones already discussed, but, when the Florida-Georgia and Texas Triangle megaregions are more restricted by data, it is usually present, see Figures 4.32, 4.15, and 4.24. This area may be an important connection between Texas and the Midwest with its strong connection to the rail and river networks for freight and industrial establishment that took place during the 1800s and 1900s.

5.2 Application to Situations

Although a great deal of literature exists exploring the subdivision of society and the built environment into megaregional elements, there is less on how these divisions can be leveraged to achieve the changes that

[1] Austin, Dallas, Fort Worth, Houston, and San Antonio form a triangle connected by I-10, I-35, and I-45.

society wishes to enact. This section discusses several existing arenas for application and then considers how the methods can be applied to new issues as they arise. The salience of megaregional geographies for inter-metropolitan transportation planning and policy making is intricately linked to the kinds of problems that community planning and its sister disciplines are attempting to address. Oden and Sciara at UT Austin analyzed 382 a survey of 382 MPO and found that these organizations perceived the costs of megaregional collaboration to exceed the potential benefits(Oden and Sciara 2020).

> "Despite significant current limitations, research and policy advocacy around megaregions may have significant value moving forward. Even if our survey results may have included some collaborative MPO work focused on large-scale corridor or other transportation activities that do not map cleanly to defined megaregions, these activities draw new attention to the importance of maintaining larger, highly connected functional systems in an era of federal and state devolution and under-investment. The content of the megaregional analysis and project work also implicitly shows that devolving responsibilities and fiscal burdens to increasingly local scales risks network ruptures and negative externalities that will generate costs at larger scales.
>
> “At a more aspirational level, expanding the spatial scale of planning to megaregional levels may help to prepare the ground for deeper discussions of prominent national challenges. . . . Orienting planning efforts to megaregional geographies conceivably would also provide new insights and approaches to challenges beyond funding, including the prioritization of infrastructure needs, the avoidance of environmental harms, and the mitigation of problematic socioeconomic patterns of human settlement (Oden and Sciara 2020).”

This section aims to identify opportunities for megaregional governance among the sub-fields of community planning. As specific applications are later defined during the process of legislating megaregions, data sources can be identified that are directly related to the desired policy envelope. This will help ensure that the administrative purpose of these governmental bodies align with the phenomena they are predicated upon.

5.2.1 Six Disciplines of Planning

Community planning is often separated into several sub-disciplines in order to more clearly organize and address challenges facing government and the public; and, although the categories used during the delineation process are a good candidate for separating domains into closely related groups, they do not match how issues are disperse among planning professionals. Instead, the following is a list of common specialization found in PCAB graduate planning programs in the United States.

- Economic Development
- Environment & Public Health
- Housing & Community
- Land Use
- Transportation
- Urban Design & Historic Preservation

These six specialties are each considered here for opportunities at megaregional scales and what data source forms might be implemented in this work's delineation process.

5.2.1.1 Economic Development

Economic development professionals are tasked with improving quality of life for residents and cultivating stable, productive tax bases to support government activity like emergency services, equity efforts, or legal arbitration. These effort may take a wide variety of forms from investments in human and infrastructural capital to post-secondary workforce education to locale marketing to potential residents.

However, the perception of winners and loser by both the public and their elected officials can turn efforts to optimize industry distribution and workforce location into bidding wars where public funds are the currency. The distribution of economic resources and opportunity are hotly contested topics at many geographic scales, but recent bids by various American metropolitan for corporate campuses for the likes of Microsoft, Amazon, and Toyota have drawn substantial media attention. In a race to the bottom, municipalities have courted major business investments with everything from tax write offs to land grants to attract them to their community. The result may be the expenditure of public funds for an outcome that the market would have brought about without government subsidy.

Another trend has been the shift in employment by industry. Amid general departures to foreign labor markets, any primary manufacturing or creation of physical goods remaining in the US has shifted towards smaller work forces and greater automation. Meanwhile, the export of services has grown to be major US export. The identification, analysis, and interpretation of trends like these require significant expertise; but, as experts are relegated to more centralized government entities, local governments lose direct access to those resources.

In an APA report, a task force on economic development noted that...

> "Planning for economic growth and development should occur at the regional level which corresponds most closely to the functioning economy. Unfortunately, economic development practice

> involves each municipality, county, or state competing with their peers. Place competition leads to imitation. But no silver-bullet exists; no one size fits all. Since cooperative, coordinated, and unique development strategies formulated at the regional level are extremely rare, planners need to find ways to contribute at the local level in spite of weak or nonexistent regional planning.
>
> "Economic developers rely heavily on financial incentives to attract investment in the near term which depletes resources needed for education and infrastructure that would benefit all companies in the long term. The shorter term emphasis on exploiting existing assets is different than the orientation of planners who want to develop assets for the long term.
>
> "Planners need to understand their region's role in the larger economic system (its economic base), find its unique features, and begin to identify innovative ways to promote economic growth and development. The most critical resource in the emerging knowledge economy is human capital; talent will largely determine regional competitiveness. With new technologies and global competition, planners need to attend to workforce development.
>
> "The region's economic base can be viewed as a mix of skills and occupations as well as a mix of traded and non-traded industries. One way to think of these two dimensions of the regional economy is that industries determine what places make (produced goods and services) whereas occupations determine what places do (activities of workers)." (Provo and Ph 2015)

Later, the task force reinforced inter-municipal cooperation with regard to the varying nature of a given municipality.

> "Finally, planners should understand the role their jurisdiction plays in the regional economy. Planners in central cities or suburban jurisdictions containing major employment nodes (traded sectors) should focus on economic growth facilitation and development enhancement described above. Planners in primarily residential jurisdictions should manage local economic activity (non-traded sectors), encourage efficient infrastructure use, and engage with education and training institutions for workforce development.
>
> "Above all, planners should foster cooperation among local jurisdictions with each making a unique contribution to the regional economy. Such cooperation is vital in the face of serious external competition (Provo and Ph 2015)."

Possible Responsibilities:

- Coordination of funding sources

- Allocation of job opportunities to where they will have the greatest impact
- Larger groups of local workforce developments that can be managed and supported by one agency
- Quality of life improvements
- Salaried employment (which implies health benefits and retirement)
- Life stability and financial security

Potential Data Sources:

- Employment related migration
- Business travel (mode and amount)
- Changes in wage over time by industry
- Workforce organization rates

5.2.1.2 Environment and Health

A governmental entity whose boundaries are more directly tied to environmental features such as climate, watersheds, and air quality, is better situated to address sources of upstream pollutants. The decades long legal battle between Florida and Georgia over water rights would have been an excellent opportunity for a megaregional body to step in, mediate, and provide expertise that was not as politically engaged while still being familiar with local circumstances.

Another factor impeding planning in the public health arena is data privacy. A megaregional government could be an ideal setting in which to analyze data if it is constructed from factors that identify communities with similar socio-demographics and general settings. This would result in a lower level of variability when compared with the national or state level, allowing for researchers to better tease apart subtle trends and make more precise policy recommendation therefrom.

The distribution of resources among regions, currently or historically, may also contribute to wealth and health trends. Such phenomena may reinforce concentrations of high and low socio-economic status populations, conflating other findings. It may be possible to address this challenge with historical exports-imports records, but this would likely need to be an independent undertaking with extensive collaboration with subject matter experts in anthropology and historical economics.

As the use of fossil fuels and the production of other pollutants continue, the adverse impacts of human health, especially in urban areas, becomes increasing apparent. However, the nature of air pollution often makes it difficult for states, much less individual municipalities, to address without significant cooperation from other governmental bodies.

Possible Responsibilities:

- Addressing sources of pollution upstream densely populated areas
- Providing climate-action solutions to local government tailored to the local environment
  - Many states share environments, so why duplicate efforts?
- Unifying licenses and regulations for environmentally integrated industries such as agriculture, fishing, and hunting
- How could the regions and megaregions defined in this work help address this issue?
- How can tools be identified for development?
- What scales and geographies are necessary to effectively tackle the issue?
- Where is the need greatest, and who is at greatest risk?
- The COVID-19 pandemic swept across the globe in a matter of weeks and has caused societal and economic upheaval for months on end.
- Did the patterns of the virus' spread follow regional trends?
- How could this be measured?
- How would responses differ if regional responses were taken?

Potential Data Sources:

- Causes of death
- Air, soil, and water quality
- Biodiversity
- Wetlands
- Transmissible disease patterns

5.2.1.3 Housing and Community Development

Housing supply has experienced several periods of unusual behavior in the last two decades. The oversupply in certain markets prior to the 2008 recession and the unusually strong demand during the COVID-19 pandemic in spite of general economic depression. To ensure equitable opportunities within a community, housing professionals have to balance financial tools and policies with the realities of politics and systemic suppression of opportunities for vulnerable populations. Unfortunately, short-term oriented financing and complex public subsidy programs create a setting that does not support housing that is affordable for these vulnerable or marginalized groups.

Possible Responsibilities:

- Coordinating funding sources
- Identifying locations for affordable housing and housing affordability measures based on regional dynamics
- Leverage larger funding sources for developers to bid on
  - Public sector often moves too slowly to directly develop real estate, so this could be more of a bid to develop situation.

Potential Data Sources:

- Household investments
- Average debt
- Housing affordability
- Renter-owner ratio

5.2.1.4 Land Use Planning

Land use is primarily concerned with the integration of other community planning fields within the built environment context. While much of this is best performed at the local level where day-to-day decisions are made, regional collaboration may be able to reduce the barrier to entry for private entities who wish to engage municipal government and, at the same time, alleviate burden on local staff in developing legal or ordinance based solution for the issues they are facing.

Formal structures for coordinating and incentivizing municipal behaviors at a regional and megaregional level could reduce barriers to entry for both businesses and residences while also reducing the burden of knowledge for practitioners and legislators. While unique circumstances often exist at the local level, legal language, policy rationale, and implementation methods could be standardized in a such a way that local government could still select tools that met their needs; and, if sufficient tools did not exist, megaregion-scale resources could be brought to bear to craft thorough options that could meet current and future situations.

Possible Responsibilities:

- Help reduce variability in zoning ordinances and building codes
  - Could develop a set of definitions and tools that local government can pick and choose from
  - Reduce legal load and associated man hours for development
- See notes on Environment and Housing

Potential Data Sources:

- Zoning code similarities

- Permitted uses
- Building density

5.2.1.5 Transportation

One of the clearest cases for a megaregional body is the coordination of infrastructure projects that cross state boundaries and have considerable impact on local communities but that are not accomplished at the national scale. The US infrastructure is severely under-invested and will require extensive efforts to address maintenance replacement shortfalls. Infrastructure links also form basis for functional relationships between components of megaregions, giving this scale of government a particular interest in how people, materials, and information flow along the network over time.

Meyer and Miller said in *Transportation Planning: A Decision Oriented Approach* that megaregions would be the drivers of national economic success.

> "It seems highly likely that the future economic success of the U.S. and other countries will be found in these megaregions, thus drawing attention to the transportation needs internal to each region, access between, and connections to the global market. The transportation connections within megaregions will need particular attention in that these connections not only support the economic activities of the megaregions themselves, but also serve as gateways to the rest of the nation. Longer distance, inter-megaregion transportation challenges will become even more apparent in future years as the nation's economy becomes more intertwined with the economic and financial activities occurring in all of the mega-regions in the country [citation needed]."

They then highlighted five economic trends that have begun to affect commercial transportation.

1. Importance of trade and globalization of the economy
2. Growth of service industries in the US of traditional manufacturing to increase competitiveness and emergence of high technology and knowledge-based industries
3. Industrial location and demographic trends, including increased flexibility for businesses in their location decisions and an aging population
4. Reduced government roles and increased privatization.

A 2012 presentation at the USDOT Volpe Center mentioned in the literature review also gave several reasons for using megaregions in transportation planning.

- Responds to reality of emerging large-scale regions.
- Better adapted to deal with global economic opportunities and environmental issues.

- Provides strategy to act globally, while providing a local focus on livability and sustainability.
- Improves health, mobility and employment opportunities across large-scale regions.
- Supports transportation innovation.

Possible Responsibilities:

- Coordinate inter-jurisdictional infrastructure projects
- Develop larger regional models of private and freight transportation flows
- Produce material for local government to improve walkability, public transit, and general accessibility
- Coordinate inter-urban rail connections between areas with local services such as buses and commuter rail

Potential Data Sources:

- Detailed, county-level freight flows
- Recreational travel flows
- Capital investment dollars
- Traffic counts
- Theoretical travel times between counties on network
- Intersection density
- Sidewalk presence
- Bike lane presence
- Transit connectivity

5.2.1.6 Urban Design & Historic Preservation

In the absence of morphological cohesion at the megaregional scale, urban design may have the least to contribute in a direct fashion. Instead, efforts in this are might focus on the development of a sense of regional identity and place. The impacts of such an identity are discussed in the next section in the context of Manuel Castell's space of flows, but the enhancement and preservation of culture and history have patent value in the integration of spatially disparate but demographically similar communities.

Possible Responsibilities:

- Coordinate urban and rural design, building, and land use policies to develop cohesive regional identity.
- Work with local communities to identify important connections between communities' cultures to inform design decisions made in other sub-fields.
- Developing messaging guidelines that create a common language between communities with varying

cultures, ideologies, and principles.

Potential Data Sources:

- Year built
- Value per square foot
- Presence of historic preservation groups
- Density of historic buildings and sites
- Evaluation of architectural style similarities

5.3 Implementing Megaregional Governance

The benefits of coordination and collaboration at the megaregional level are patent in the research reviewed earlier in this work, but the path forward towards its implementation is less clear. Two options for a megaregional body's role are convener, which is currently the primary avenue pursued by many MPOs, and administrator, more in line with the operation of entities like municipalities and states. Convening the voices and interests of a community has been a vital function of modern government, and the megaregional scale encompasses a wide variety of opinion and needs, making communication even more important. However, a stronger format may be needed to address the challenges facing the community planning profession and society at large.

The underlying reasons of this need for strong government can be found in the conclusion of *The Informational City*. Manuel Castell notes that the supersedence of the "space of places" by the "space of flows" (meaning the reorganization of exchange networks into a asymmetric, power-favoring, and geography-independent form) has resulted in the evaporation of social meaning from physical places (pg. 349). In this setting, even democracies can struggle to identify sources of oppression and inequity. Individual citizens are at an even greater disadvantage with fewer informational resources and less time to process and understand the constantly shifting milieu of market forces. First published in 1989, the following passage from *The Informational City* is eerily prescient of the social and political unrest in the US and around the world in 2020, (Castells 1989).

> "Faced with the variable geometry of the space of flows, grassroots mobilizations tend to be defensive, protective, territorially bounded, or so culturally specific that their codes of self-recognizing identity become non-communicable, with societites tending to fragment themselves into tribes, easily prone to a fundamentalist affirmation of their identity. While power constitues

> an articulated functional space of flows, societies deconstruct their historical culture into localized identities that recover the meaning of places only at the price of breaking down commmunication among different cultures and different places. Between ahistorical flows and irreducible identiteis of local communities, cities and regions disappear as socially meaningful places. The historical outcome of this process could be the ushering in of an era characterized by the uneasy coexistene of extraordinary human achievements and the disintegration of large segmetns of society, along wiht the widespread prevalence of senseless violence - for the impossibility of communication transforms other communities into 'aliens,' and thus into potential enemeis. The globalization of power flows and the tribalization fo local communities are part of the same fundamental process of historical restructing: the growing dissociation between techno-economic development and the corresponding mechanisms of social control of such development." (pg. 350)

Castell then outlines several "policy orientations" that could lead towards a revitalization of place-based systems, counteracting the divisive and isolating forces inherent to the current paradigm of space of flows and flows of power. Much like the Smith and Timberlake functions of flows, Castell divides these orientations into three broad categories: cultural, economic, and political. However, before considering these strategies, there must first be a clear identifications of the parties involved.

5.3.1 Entities in question:

In considering implementation, there are a number of entity types that will ultimately interact with megaregions. How their interactions differ across space, time, and scale are import questions to answer when there is a goal of reducing resistance to the process of integrating megaregional governance into the American administrative-political structure.

- Citizens & Residents (Individuals and Households)
- Private Organizations (Community Groups and Businesses)
- Governments (Local, State, and Federal)

5.3.1.1 Citizens & Residents

This category encompasses individuals and households. Although these two entities do not typically act substantially different from one another, they are both included here due to their usage in various datasets and methodologies. These are the least likely to interact directly with a regional or megaregional government body; however, they are the most important evaluatory point for efforts made by such bodies.

5.3.1.2 Private Organizations

This consists of groups ranging from informal to formal with a wide variety of focuses. In a vacuum, the maximum *potential* for business is found when there is no regulation restricting their action, but to maximize the average ability to succeed requires a careful balancing of market freedom and societal regulation. In Castell's view, the largest of these organizations are what drive the global economy from within a space of flows, separated in part from their physical location.

5.3.1.3 Government Bodies

How various levels of government, along with their staff and elected officials, react to regional government will depend greatly on their relationship with it. The federal government already works in a limited extent with regional government in the form of MPOs, and the policies recommended in this work will build off of that existing framework.

To encourage municipal buy-in as legislation is put in place, the empowering of regional government to enforce policies and offer incentives must result in greater benefit and freedom to local community than they might lose, thus appealing to the shift from managerial to entrepreneurial in local government objectives over the last several decades (Hall and Barrett 2012). State and local governments may perceive regional government as a threat to their autonomy if implemented poorly, but, done well, they may see it as a useful tool and way to reduce their administrative loads. This aspect of integrating regional governance will be analyzed through the lens of small state political theory.

5.3.2 Castell's Reconstruction of Place

As mentioned previously, the reflections on rebuilding the sense and importance of place fell into three categories: cultural, economic, and political.

5.3.2.1 Cultural

In order to succeed in this global context, Castell felt local societies needed to

- preserve local identities,
- leverage historic roots
- symbolically mark places associated with these identities and roots,
- actively preserve those symbols of recognition,
- express collective memories through actual communication,

- develop common-language terms with other community identities where communities are recognized as sub-cultures able to communicate through and with higher-order cultures (e.g. regions and megaregions), and
- connect the affirmation and symbolic practice of identity to economic policy and political practice.

All this works to overcome the dangers posed by tribalism and fundamentalism through improvements to inter-personal and inter-community communication. The best parties to support these processes vary by scale. Local community groups and non-profits are often well situated to work within municipal populations. Municipalities themselves already are those primarily responsible for historic preservation and the maintenance of local points of interest; their efforts could then be formally supported by higher-level bodies. Regional organizations can facilitate this between individual communities, and megaregions can do the same between regions and their interstitial hinterlands.

5.3.2.2 Economic

Cities and regions already form the physical basis of national economies, but they must also find their role in the new informational economy. As the ability to generate, process, and leverage information becomes more and more central to economic success, places need to understand how their resident labor forces respond to living conditions, education, and social settings and how those factors stimulate economic development. If a labor force is not supported by a well-defined cultural identity and the political power of its local government, it is much more vulnerable to the shifting tides of the global economy. However, the strictly physical nature of workers' home locations can tie economic activity, and by extension a community's economic power, to a geographic location. A skilled and cohesive group of laborers, whether they produce physical goods or services, is an important asset to any community and public investment should reflect that.

5.3.2.3 Political

Governments must play a central role in organizing the social control of places so as to defend against the disintegrating forces of the functional logic of the space of flows. Because their officials most directly reflect the special interests of a community, local government identify those interests and the needs of their residents. Specifically, Castell states that local government must reinforce power and capacity in two areas: the mobilization of their population's participation in collective strategy making, and connection with other communities in non-tribal ways that reflect their shared interests and assets. This second point would be accomplished through a coordinating megaregional body. This megaregional government can then aggregate those positions and use the economic and political power they represent to bargain with the powerful entities that direct globalized flows. In turn, megaregional government can return to the local level and distribute the bargained for aspects among communities according to their needs and desires ensuring that the flexibility of municipalities is used to its maximum potential.

This process can protect local government from inter-governmental conflict while still allowing them to make concrete efforts towards their citizens' goals. Coordination and collaboration through regional and megaregional government puts communities on a similar playing field to power-holder organizations by aggregating expertise and resources that would otherwise be spent making marginal progress after the management of day-to-day needs has been completed. Information technologies will facilitate coordination across large geographic regions, but it is ultimately the combination social mobilization and political will that produces meaningful results.

Castell's reflection seek to counter the runaway effects of a globalizing and centralizing of economic power. This is not to say that these forces are inherently wrong or bad; worldwide standards of living have risen steadily, and extreme poverty been reduced dramatically. Rather, it is that their impacts have rippled out into society with detrimental effects. To promote actions based on the Castell policy orientation, megaregional government must have a strong connection to small-scale actors and seek their support and buy-in for the formation of megaregional government and its subsequent actions. With inter-municipal conflict commonplace, these small actors struggle to leverage their collective economic and political weight. Megaregional government would provide a concentration point for this influence that would still be responsive to local needs and interests.

5.3.3 Small State Political Theory

Small state political theory has generally been used to describe the actions of countries like Qatar and Luxemburg on the world stage, how they seek to influence international policy, and what differences exist in how they utilize their resources. By changing the scale under consideration, this framework can be used to describe the rationale behind local governments' decisions and how policy could account for existing political trends among municipalities.

As stated previously, this theoretical framework describes the characteristics of small powers as dependence, status-quo orientation, supportive or international law and organizations, and risk aversion.

A lack of external threats removes local governments' security-based need to develop alliances. Alliances may still form in other domains such as the economy; however, these are relatively rare, and municipalities—and their elected officials—are more likely to act competitively than cooperatively in order to please voting blocks.

Security does exist at the local level in the form of law enforcement and the military industrial complex, but these are often perceived as intra-community issues rather than inter-community. Instead, threats to sovereignty are more likely to come from the parent state in the form of legislation, programs, and policies.

The absence of external threats implies the absence of an associated unifying force that can bring together disparate populations, making it more difficult to incentivize power sharing and collaboration between various levels and facets of government. To counteract this vacuum, standards of behavior can provide a setting in which the governmental and societal structures that already support local independence can be leveraged to distribute the benefits of megaregional government while preserving the benefits that creatures of the state enjoy.

The system that creates municipalities in the legal sense is also the system that allows them to operate with such freedom and independence. If the current situation is viable and profitable for a given municipality, it will be reluctant to advocate for systemic change. Therefore, a two pronged approach can be used in both gathering support for regional governance and in developing incentives within legislation: identifying the actual shortcomings of the current system both for the municipality itself and for its residents, and presenting new opportunities in the context of regional governance.

The focus of this work in highlighting shortcomings and opportunities comes from efficiency and efficacy.

5.3.4 Efficiency

The implementation should avoid increasing the size of government as much as is possible.

Who then is responsible for running and operating a regional government or governmental body? From a decision making standpoint, this can follow existing models used by regional commissions and councils of government where a body of elected officials from sub-regions nominate one of their members to represent them at the regional level. From a staff perspective, the benefits provided by a megaregional institution should be able to offset sufficient costs at the local level that municipalities can contribute funding in return for services and tools that are already tailored to achieve the goals outlined by Castell previously. This reallocating of fiscal and human resources into agglomerations will be able to serve many state and local governments more efficiently over time than any individual department could.

5.3.5 Efficacy

Once resources have been allocated more efficiently, megaregional institutions can begin to assess opportunities to produce recommendations and adjustments for local municipalities with the aim of greater impact per dollar spent. For example, there is no standard evaluation of MPOs. Many MPOs along with their parent organizations engage in internal assessments and reviews, but these can rarely be compared across localities.

By including an assessment process in the foundational legislation, megaregional government will be more transparent from the outside while being able to better evaluate the effectiveness of their own efforts. Other opportunities may include ordinance standards, program guidance, and fund distribution.

CHAPTER 6

Conclusion

Megaregions have the potential to greatly improve the efficacy and efficiency of government in the United States by more effectively leveraging resources and maintaining inter-community dialogues. Previous delineation efforts have been primarily scholarly and focused on limited datasets or highly private with opaque methodologies. This work proposed a method that focuses on public administration, transparency, and a multi-faceted group of datasets that reflect the complex nature of communities. To move forward in regional community planning, research must add the application of findings to the inquiry of phenomena.

Regionalism has been a movement in community planning for well over a century, and research in the last several decades has shown the potential benefits of coordination and collaboration at meta-community scales. Some governance does exist in the US at the regional level, but its use is limited when compared to local and state governments. Due to the inter-disciplinary nature of government, theories from other fields such as political science will facilitate an effective transition from academic consideration to policy implementation. Small state political theory is one example as it provides descriptive tools for identifying the basic interests and needs of municipalities and their officials and how they can be met while reassigning power and influence among sub-national government entities. Building bridges between this framework on local governance, megaregional theory, and other schools of thought mimics the relationships between physically disconnected communities and will help prepare legislation and policy that robustly describe megaregion communities.

Although some spatial theories about the built environment consider global economies and local communities separately, the connections that do exist between these two scales can be used as a basis for describing an administrative role for and delineation of megaregions. Through a series of cluster analyses, the process proposed in this work identifies a set of potential megaregional boundaries by combining datasets grouped by a typological tree. By nature, the analysis process allows for major economic cores to connect but when that connection's "strength" is sufficient to include the intermediary hinterlands as well. The resulting boundary set illuminates the underlying phenomena and how individual data sources contribute to the form of the megaregions delineated by the entire set of data sources.

With several portions of the contiguous United States showing repeated cohesion throughout a number of cluster sizes and data source types, the delineation process appears to have been successful in identifying megaregions while taking into account constraint based on the intended administrative purposes. There are still refinements that will produce results more comprehensible to decision makers and the general public, but the basis has shown an ability to integrate a wide array of functional and morphological information in a repeatable and transparent fashion. Once the shortfalls have been removed, the methodology will be ready to apply in real-world situations. Not every aspect of community planning is best achieved at the regional or megaregional scale, but there a number of opportunities where doing so will reduce the duplication of efforts by over-burdened local staff and better leverage agglomerated economic and political power to compete in a globalized market. Therefore, implementation strategies should ensure that induced marginal costs do not exceed the provided benefits.

In addition to the identification of six areas with emergent megaregions, the discussion of opportunities to apply megaregional government and what purposes such bodies should ultimately pursue is an outgrowth of the current national circumstances. The delineation of megaregion must therefore be an ongoing conversation as societal needs shift and evolve; and, with a highly data driven delineation process, these boundaries can be periodically redefined alongside those shifting cultural patterns. There is still a great deal of work to do in this area of study, and this work aims to serve as a first foray into practical delineation of megaregions. As society faces continuing challenges such as climate change, civil unrest, and growing inequity, megaregions can act as keystone conveners and responsive administrators to help communities make the most of their resources in a global economy made up of everyday people.

Appendix A

6.1 Cluster Analysis R-Code

```
# Libraries -------------------------------------

library(data.table)

library(tidyverse)

library(sf)

library(sfheaders)

library(ClustGeo)


# Functions -------------------------------------

d0d1_partition <- function(D0, D1, k, alpha){

  return(cutree(hclustgeo(D0, D1, alpha = alpha),k))

}

vector_normalize <- function(x){

  x <- as.numeric(x)

  if(length(unique(x))==1){

    return(rep(1,length(x)))

  }else{

    return((x - min(x, na.rm = T)) / (max((x - min(x, na.rm = T)), na.rm = T)))

  }

}


# Get Datasets ----------------------------------

files_data <- intersect(list.files('process-data/02-typology-categories', pattern = 'dat_
    ↪ '),

                        list.files('process-data/02-typology-categories', pattern = '.csv')
                            ↪ )

files_dist <- intersect(list.files('process-data/02-typology-categories', pattern = 'dist
    ↪ _'),
```

```
                         list.files('process-data/02-typology-categories', pattern = '.csv')
                            ↪ )
if(file.exists('dist-counties.zip')){
  unzip('dist-counties.zip')
  D1 <- as.dist(as.matrix(fread(file = 'dist-counties.csv')[-c(1193,2947),-c
      ↪ (1,1193+1,2947+1)]))
  file.remove('dist-counties.csv')
}

# Counties
crs.aeac <- paste0("+proj=aea␣+lat_1=20␣+lat_2=60␣+lat_0=40␣+lon_0=-96␣",
                   "+x_0=0␣+y_0=0␣+ellps=GRS80␣+datum=NAD83␣+units=m␣+no_defs␣")
sf_co <- st_transform(st_read('shapefiles/counties-contiguous-aeac-fixed.shp'), crs = crs
    ↪ .aeac)
sf_co <- sf_co[-which(sf_co$GEOID == '25019' | sf_co$GEOID == '53055'),]


# K-Alpha Values & Evaluation Metrics -----------
kvalues <- c(2:48)
alphastep <- 0.1
alphavalues <- seq(0+alphastep, 1-alphastep, alphastep)
vector.ka <- rep(0, length(kvalues)*length(alphavalues))
column.names <- paste0('a',gsub(pattern = '\\.', replacement = '', x = as.character(
    ↪ alphavalues)))
row.names <- paste0('k',kvalues)
matrix.names <- gsub(pattern = 'ds_|.csv', replacement = '', x = files_dist)

arr_kalpha_avgnb <- array(vector.ka,
                          dim = c(length(kvalues), length(alphavalues),length(matrix.names))
                             ↪ ,
                          dimnames = list(row.names, column.names, matrix.names))
arr_kalpha_micd <- arr_kalpha_avgnb
```

```r
for(i in 1:length(files_dist)){
  # average number of neighbors
  file_avgnb <- paste0('process-data/03-clustered-datasets/dat_kalpha_avgnb_',
                     gsub(pattern = 'dist_', replacement = '', x = files_dist[i]))
  mat_avgnb <- as.matrix(fread(file_avgnb,data.table = F))
  arr_kalpha_avgnb[,,i] <- mat_avgnb[,ifelse(dim(mat_avgnb)[2]==9,1,dim(mat_avgnb)[2]-8):
      ↪ dim(mat_avgnb)[2]]
  # mean of average in-cluster distances
  file_micd <- paste0('process-data/03-clustered-datasets/dat_kalpha_micd_',
                    gsub(pattern = 'dist_', replacement = '', x = files_dist[i]))
  mat_micd <- as.matrix(fread(file_micd,data.table = F))
  arr_kalpha_micd[,,i] <- mat_micd[,ifelse(dim(mat_micd)[2]==9,1,dim(mat_micd)[2]-8):dim(
      ↪ mat_micd)[2]]
}



# Alpha Selection ------------------------------
dat_alpha_select <- as.data.frame(matrix(c(0), nrow = length(kvalues), ncol = 1 + length(
    ↪ matrix.names)))
colnames(dat_alpha_select) <- c('k',matrix.names)
dat_alpha_select$k <- kvalues
for(j in 1:length(matrix.names)){
  for(i in 1:length(kvalues)){
    vec_alpha <- (1 - vector_normalize(arr_kalpha_avgnb[i,,j])) +
      vector_normalize(arr_kalpha_micd[i,,j])
    dat_alpha_select[i,j+1] <- alphavalues[min(which(vec_alpha == min(vec_alpha), arr.ind
        ↪ =TRUE))]
  }
```

```
}
write.csv(dat_alpha_select, file = 'process-data/03-clustered-datasets/dat_alpha_select.
    ↪ csv')


# K Selection ----------------------------------

vector.aam <- rep(0, length(kvalues)*3)
column.names <- c('alpha','avgnb','micd')
row.names <- paste0('k',kvalues)
matrix.names <- gsub(pattern = 'ds_|.csv', replacement = '', x = files_dist)

arr_k_select <- array(vector.aam,
                      dim = c(length(kvalues), length(column.names), length(matrix.names)),
                      dimnames = list(row.names, column.names, matrix.names))

arr_k_select[,,1]

for(i in 1:length(matrix.names)){
  # for each dataset
  for(j in 1:length(kvalues)){
    # for each k number of clusters
    # insert the optimal alpha value
    arr_k_select[j,1,i] <- dat_alpha_select[j,i+1]
    # find the corresponding avgnb for the kalpha value
    arr_k_select[j,2,i] <- arr_kalpha_avgnb[j,which(alphavalues==arr_k_select[j,1,i]),i]
    # find the corresponding micd for the kalpha value
    arr_k_select[j,3,i] <- arr_kalpha_micd[j,which(alphavalues==arr_k_select[j,1,i]),i]
  }
  # normalize the avgnb and the micd
  arr_k_select[,2,i] <- vector_normalize(arr_k_select[,2,i])
  arr_k_select[,3,i] <- vector_normalize(arr_k_select[,3,i])
```

```
}

arr_k_select[,3,]



if(length(files_data)==1){
  plot(x = kvalues, y = vector_normalize(as.numeric(arr_k_select[,2,])), pch = 3, col = '
      ↪ red',
       ylab = 'In-Cluster␣Distance␣(X)␣and␣Membership␣(+)',
       xlab = 'Number␣of␣Clusters', main = paste('Sum␣of␣Normalized␣Metrics␣(---)␣for␣
           ↪ Dataset␣Group:',id_opt))
  points(x = kvalues, y = vector_normalize(as.numeric(arr_k_select[,3,])), pch = 4, col =
      ↪  'blue')
  lines(x = kvalues, y = vector_normalize(as.numeric(arr_k_select[,2,])) +
          vector_normalize(as.numeric(arr_k_select[,3,])))
}else{
  plot(x = kvalues, y = vector_normalize(rowMeans(apply(arr_k_select[,2,],2,as.numeric)))
      ↪ , pch = 3, col = 'red',
       ylab = 'In-Cluster␣Distance␣(X)␣and␣Membership␣(+)',
       xlab = 'Number␣of␣Clusters', main = paste('Sum␣of␣Normalized␣Metrics␣(---)␣for␣
           ↪ Dataset␣Group:',id_opt))
  points(x = kvalues, y = vector_normalize(rowMeans(apply(arr_k_select[,3,],2,as.numeric)
      ↪ )), pch = 4, col = 'blue')
  lines(x = kvalues, y = vector_normalize(rowMeans(apply(arr_k_select[,2,],2,as.numeric))
      ↪ ) +
          vector_normalize(rowMeans(apply(arr_k_select[,3,],2,as.numeric))))

}
```

```
for(i in 1:length(matrix.names)){
  write.csv(arr_k_select[,,i],
            file = paste0('process-data/03-clustered-datasets/dat_k_select_',
                          gsub(pattern = 'dist_',replacement = '',x = matrix.names[i]), '.
                              ↪ csv'))
}



# BEST K
if(length(files_data)==1){
  dat_k_select_result <- data.frame('k'=kvalues,
                                    'metric' = vector_normalize(as.numeric(arr_k_select[,2,])
                                        ↪ ) +
                                      vector_normalize(as.numeric(arr_k_select[,3,])))
}else{
  dat_k_select_result <- data.frame('k'=kvalues,
                                    'metric' = vector_normalize(rowMeans(apply(arr_k_select
                                        ↪ [,2,],2,as.numeric))) +
                                      vector_normalize(rowMeans(apply(arr_k_select[,3,],2,as.
                                          ↪ numeric))))
}


write.csv(dat_k_select_result, file = 'process-data/03-clustered-datasets/dat_k_select_
    ↪ result.csv')

k_final <- dat_k_select_result$k[which(dat_k_select_result$metric == min(dat_k_select_
    ↪ result$metric))]
print(k_final)
```

```
# Partitions for Best K ------------------------

# Raw Partitions based on Best K and datasets' corresponding best alpha
dat_partitions <- as.data.frame(matrix(c(0), nrow = length(sf_co$GEOID), ncol = length(
    ↪ files_data)))
colnames(dat_partitions) <- paste0('p_',gsub(pattern = 'dist_',replacement = '',x =
    ↪ matrix.names))
rownames(dat_partitions) <- sf_co$GEOID

for(i in 1:length(files_dist)){
  print(files_dist[i])
  D0 <- as.dist(as.matrix(fread(paste0('process-data/02-typology-categories/',files_dist[
      ↪ i]), data.table = F)[-1,]))
  dat_partitions[,i] <- as.character(d0d1_partition(D0, D1, k_final, dat_alpha_select[
      ↪ which(dat_alpha_select$k==k_final), i+1]))
}

sf_partitions <- cbind(sf_co[,4:5], dat_partitions)
st_write(sf_partitions, dsn = 'process-data/03-clustered-datasets/sf_partitions.shp',
        driver = 'ESRI Shapefile', delete_dsn = T)







# Contiguous Partitions

dat_partitions_fixed <- as.data.frame(matrix(c(0), nrow = length(sf_co$GEOID), ncol =
    ↪ length(files_data)))
colnames(dat_partitions_fixed) <- paste0('p_',gsub(pattern = 'dist_',replacement = '',x =
    ↪  matrix.names))
```

```r
rownames(dat_partitions_fixed) <- sf_co$GEOID

for(i in 1:length(files_dist)){
  print(files_dist[i])

  sf_dissolved <- sf_partitions %>%
    group_by_at(i+2) %>%
    summarise()

  sf_multi <- sf_dissolved[which(st_geometry_type(sf_dissolved)=='MULTIPOLYGON'),] %>% st
      ↪ _cast(to = 'POLYGON')
  sf_single <- sf_dissolved[which(st_geometry_type(sf_dissolved)=='POLYGON'),]

  sf_poly <- rbind(sf_multi, sf_single)

  sf_noislands <- sf_poly %>%
    add_column(area = st_area(.)) %>%
    group_by_at(1) %>%
    filter(area == max(area)) %>%
    sf_remove_holes()

  dat_partitions_fixed[,i] <- as.numeric(as.character(st_within(x = st_centroid(sf_co),
      ↪ sf_noislands)))
  dat_partitions_fixed[is.na(dat_partitions_fixed[,i]),i] <- 0
}

head(dat_partitions_fixed,20)
sf_partitions_fixed <- cbind(sf_co[,4:5], dat_partitions_fixed)
st_write(sf_partitions_fixed, dsn = 'process-data/03-clustered-datasets/sf_partitions_
    ↪ fixed.shp',
         driver = 'ESRI␣Shapefile', delete_dsn = T)
```

```r
# Cluster Boundaries Shapefile ------------------

variables <- colnames(dat_partitions_fixed)
for(i in 1:length(files_dist)){
  for(j in 1:k_final){
    print(paste(variables[i],j))
    sf_cluster <- sf_partitions_fixed %>%
      group_by_at(i+2) %>%
      filter(.data[[variables[i]]] == j) %>%
      summarise() %>%
      add_column('variable'=c(variables[i])) %>%
      add_column('partition'=c(j)) %>%
      select(-1)
    if(i == 1 & j == 1){
      sf_cluster_regions <- sf_cluster
    }else{
      sf_cluster_regions <- rbind(sf_cluster_regions, sf_cluster)
    }
  }
}

st_write(sf_cluster_regions, dsn = paste0('process-data/03-clustered-datasets/sf_cluster_
    ↪ regions-',
                                          'k',k_final,'-',id_opt,'.shp'),
        driver = 'ESRI␣Shapefile', delete_dsn = T)
```

6.2 Megaregion Delineation R Code

```r
# 00 - Background Data -------------------------
library(igraph)
# 00-1 - Functions -----------------------------
```

```
st_rook = function(a, b = a) st_relate(a, b, pattern = "F***1****", sparse = F)

# 00-2 - Shapefiles ----------------------------

# Albers Equal Area Conic for North America
crs.aeac <- paste0("+proj=aea␣+lat_1=20␣+lat_2=60␣+lat_0=40␣+lon_0=-96␣",
                   "+x_0=0␣+y_0=0␣+ellps=GRS80␣+datum=NAD83␣+units=m␣+no_defs␣")

## US counties from the contiguous US, Albers Equal Area Conic, geometry fixed in QGIS
sf_co <- st_transform(st_read('shapefiles/counties-contiguous-aeac-fixed.shp'), crs = crs
    ↪ .aeac)
sf_co <- sf_co[-which(sf_co$GEOID == '25019' | sf_co$GEOID == '53055'),]



# 01 - MPO Geographies -------------------------

# 01-1 - MPO Counties --------------------------

sf_mpos <- st_transform(st_read('shapefiles/conus-mpos.shp'), crs = crs.aeac)

# MPO - County Crosswalk
centroids_mpos <- st_centroid(sf_mpos)
centroids_co <- st_centroid(sf_co)
within_co_mpos <- st_within(x = centroids_co, y = sf_mpos, sparse = F)
within_mpos_co <- t(st_within(x = centroids_mpos, y = sf_co, sparse = F))

# ggplot() + geom_sf(data = centroids_mpos)

xw_mpos_co <- data.frame('geoid'=as.character(sf_co$GEOID))
for(i in 1:length(xw_mpos_co$geoid)){
  if(sum(within_co_mpos[i,])==1){
```

```
    xw_mpos_co$mpo_id[i] <- as.character(sf_mpos$mpo_id[which(within_co_mpos[i,]==1)])
  }else if(sum(within_mpos_co[i,])==1){
    xw_mpos_co$mpo_id[i] <- as.character(sf_mpos$mpo_id[which(within_mpos_co[i,]==1)])
  }else{
    xw_mpos_co$mpo_id[i] <- 'none'
  }
}


# 01-2 - Counties Within MPOs -------------------
population <- fread('process-data/00-raw-data/ACSDT5Y2019.B01003/ACSDT5Y2019.B01003_data_
    ↪ with_overlays.csv')[-1, c(1,3)]
colnames(population) <- c('AFFGEOID','pop')
population$pop <- as.numeric(population$pop)
population$GEOID <- substr(population$AFFGEOID,10,14)

sf_co_mpos <- left_join(x=sf_co[which(xw_mpos_co$mpo_id!='none'),],
                      y = xw_mpos_co, by = c('GEOID'='geoid')) %>%
  left_join(y = population)


# 01-3 - MPO County Groups ----------------------
sf_mpocg <- sf_co_mpos %>% group_by(mpo_id) %>% dplyr::summarise(pop = sum(pop)) %>%
    ↪ arrange(desc(pop))
centroids_mpocg <- st_centroid(sf_mpocg)




# 02 - Iterating through MPOCGs -----------------

# 02-1 - MPO - Cluster Region Relationship ------
```

```
# Cluster Region
sf_cluster_regions <- st_read(paste0('process-data/03-clustered-datasets/sf_cluster_
    ↪ regions-',
                                     'k',k_final,'-',id_opt,'.shp'))


# MPO within Cluster Region
mat_mwcr <- st_within(x = centroids_mpocg, y = sf_cluster_regions, sparse = F)
# Cluster Region Neighbors
mat_crnb <- matrix(c(0), nrow = length(centroids_mpocg$mpo_id), ncol = length(centroids_
    ↪ mpocg$mpo_id))
# Identify which MPOs share cluster regions as a percentage
for(i in 1:length(centroids_mpocg$mpo_id)){
  for(j in 1:length(centroids_mpocg$mpo_id)){
    if(i*j %% 100 == 0){print(paste(i, j))}
    mat_crnb[i,j] <- sum(mat_mwcr[i,] == T & mat_mwcr[j,] == T) / sum(mat_mwcr[i,] == T |
        ↪  mat_mwcr[j,] ==T)
  }
}



# 02-2 - Cluster Region Intersections -----------

threshold_transition <- 0.5
threshold_periphery <- 0.0

sf_periphery_all <- st_sf(st_sfc(st_point(c(0,0))),crs = crs.aeac) # temporary sf_
    ↪ periphery_all to check against for i==1

# iterate through MPO county groups in order of population
for(i in 1:length(centroids_mpocg$mpo_id)){
```

```
for(i in 1:18){
  print(paste(i,'/',length(centroids_mpocg$mpo_id),
              sf_mpos$mpo_name[which(sf_mpos$mpo_id==centroids_mpocg$mpo_id[i])]))
  # intersection of all cluster regions
  if((i==1 |
      all(!st_within(x = centroids_mpocg[i,], y = sf_periphery_all, sparse = F))) &
     sum(mat_mwcr[i,]==T)!=0){
    print('New␣megaregion...')
    if(length(which(mat_mwcr[i,]==T))==1){
      sf_crintersection_current <- sf_cluster_regions[which(mat_mwcr[i,]==T),] %>% st_cast
          ↪ (to = 'POLYGON')
    }else{
      sf_crintersection_current <- st_collection_extract(x = st_intersection(sf_cluster_
          ↪ regions[which(mat_mwcr[i,]==T),]),
                                                          type = "POLYGON") %>% st_cast(to = '
                                                              ↪ POLYGON')
    }



    # MPOs that are always in the same set of cluster regions
    centroids_core <- centroids_mpocg[which(mat_crnb[i,] == 1),]
    cscw <- colSums(st_within(x = centroids_core, y = sf_crintersection_current, sparse =
        ↪  F)) # column sum core within
    sf_core_current <- sf_crintersection_current[which(cscw==max(cscw)),] %>%
      add_column(main_mpo = sf_mpos$mpo_name[which(sf_mpos$mpo_id==centroids_mpocg$mpo_id[
          ↪ i])]) %>%
      select(main_mpo, geometry)
    if(i==1){
      sf_core_all <- sf_core_current
    }else{
      sf_core_all <- rbind(sf_core_all, sf_core_current)
    }
```

```
# How many datasets are under consideration?
if(length(which(mat_mwcr[i,]==T))==1){
  # When there is also one dataset, the transition and periphery areas are the same as
      ↪  the core area
  # due to the absence of overlapping regions.
  sf_periphery_current <- sf_transition_current <- sf_core_current
  if(i==1){
    sf_transition_all <- sf_transition_current
    sf_periphery_all <- sf_periphery_current
  }else{
    sf_transition_all <- rbind(sf_transition_all,sf_transition_current)
    sf_periphery_all <- rbind(sf_periphery_all,sf_periphery_current)
  }
}else{
  # MPOs that share a cluster region with the primary MPO more than 50% of the time
  centroids_transition <- centroids_mpocg[which(mat_crnb[i,] > threshold_transition),]
  within_transition <- st_within(x = centroids_transition,
                                 y = sf_crintersection_current,
                                 sparse = F)
  sf_transition_current_parts <- st_cast(sf_crintersection_current[which(colSums(
      ↪ within_transition) > 0),],
                                         to = "POLYGON")
  # Check which parts are rook contiguous to each other
  rookt.nb <- st_rook(a = sf_transition_current_parts)
  rookt_contig <- components(graph_from_adjacency_matrix(adjmatrix = rookt.nb,
                                                         mode = "undirected"), mode = '
                                                            ↪ strong')$membership
  rookt_main <- rookt_contig[which(st_within(x = centroids_mpocg[i,],
                                             y = sf_transition_current_parts,
                                             sparse = F))]
```

```
if(length(rookt.nb)>0){
  sf_transition_current <- sf_transition_current_parts[which(rookt_contig==rookt_
      ↪ main),] %>%
    dplyr::summarize() %>%
    sf_remove_holes() %>%
    add_column(main_mpo = sf_mpos$mpo_name[which(sf_mpos$mpo_id==centroids_mpocg$mpo
        ↪ _id[i])]) %>%
    select(main_mpo, geometry)

  if(i==1){
    sf_transition_all <- sf_transition_current
  }else{
    sf_transition_all <- rbind(sf_transition_all, sf_transition_current)
  }

}

# MPOs that share at least one cluster region
centroids_periphery <- centroids_mpocg[which(mat_crnb[i,] > threshold_periphery),]
within_periphery <- st_within(x = centroids_periphery,
                              y = sf_crintersection_current,
                              sparse = F)
sf_periphery_current_parts <- st_cast(sf_crintersection_current[which(colSums(within
    ↪ _periphery) > 0),],
                                      to = "POLYGON")
# Check which parts are rook contiguous to each other
rookp.nb <- st_rook(a = sf_periphery_current_parts)
rookp_contig <- components(graph_from_adjacency_matrix(adjmatrix = rookp.nb,
                                                    mode = "undirected"), mode = '
                                                        ↪ strong')$membership
rookp_main <- rookp_contig[which(st_within(x = centroids_mpocg[i,],
                                           y = sf_periphery_current_parts,
```

```
                                                          sparse = F))]
      if(length(rookp.nb)>0){
        sf_periphery_current <- sf_periphery_current_parts[which(rookp_contig==rookp_main)
            ↪ ,] %>%
          dplyr::summarize() %>%
          sf_remove_holes() %>%
          add_column(main_mpo = sf_mpos$mpo_name[which(sf_mpos$mpo_id==centroids_mpocg$mpo
              ↪ _id[i])]) %>%
          select(main_mpo, geometry)

        if(i==1){
          sf_periphery_all <- sf_periphery_current
        }else{
          sf_periphery_all <- rbind(sf_periphery_all, sf_periphery_current)
        }

      }
    }
  }else{
    print('Outside␣a␣cluster␣region␣or␣already␣in␣an␣existing␣megaregion...')
  }
}




# 02-3 - Write Shapefiles -----------------------
folder <- 'process-data/04-megaregions/shapefiles/'
st_write(obj = sf_core_all, driver = 'ESRI␣Shapefile', delete_dsn = T,
         dsn = paste0(folder, 'sf-core-all-','k',k_final,'-',id_opt,'-tp',threshold_
             ↪ periphery,'.shp'))
st_write(obj = sf_transition_all, driver = 'ESRI␣Shapefile', delete_dsn = T,
```

```
        dsn = paste0(folder,'sf-transition-all-','k',k_final,'-',id_opt,'-tt',threshold_
            ↪ transition,'.shp'))
st_write(obj = sf_periphery_all, driver = 'ESRI␣Shapefile', delete_dsn = T,
        dsn = paste0(folder,'sf-periphery-all-','k',k_final,'-',id_opt,'-tp',threshold_
            ↪ periphery,'.shp'))
```

Works Cited

Alesina, Alberto, and Enrico Spolaore. 1997. "On the number and size of nations." *Quarterly Journal of Economics* 112 (4): 1026–56. https://doi.org/10.1162/003355300555411.

———. 2005. "War, peace, and the size of countries." *Journal of Public Economics* 89 (7): 1333–54. https://doi.org/10.1016/j.jpubeco.2003.07.013.

Birditt, Kira S., Angela Turkelson, Karen L. Fingerman, Courtney A. Polenick, and Akari Oya. 2021. "Age Differences in Stress, Life Changes, and Social Ties During the COVID-19 Pandemic: Implications for Psychological Well-Being." *The Gerontologist* 61 (2): 205–16. https://doi.org/10.1093/geront/gnaa204.

Bruine de Bruin, Wändi. 2021. "Age Differences in COVID-19 Risk Perceptions and Mental Health: Evidence From a National U.S. Survey Conducted in March 2020." *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 76 (2): e24–29. https://doi.org/10.1093/geronb/gbaa074.

Castells, Manuel. 1989. *The Informational City.* Massachusetts: Blackwell Publishers Ltd.

Chabard, Pierre. 2010. *Manifestoes and Transformations in the Early Modernist City.* Edited by Christian Hermansen Cordua.

Cherry, Gordon Emanuel. 1974. *The Evolution of British Town Planning; a History of Town Planning in the United Kingdom During the 20th Century and of the Royal Town Planning Institute, 1914-74.* New York: Wiley.

Clouston, Sean A. P., Ginny Natale, and Bruce G. Link. 2021. "Socioeconomic inequalities in the spread of coronavirus-19 in the United States: A examination of the emergence of social inequalities." *Social Science and Medicine* 268 (November 2020). https://doi.org/10.1016/j.socscimed.2020.113554.

Georg, Isabel, Thomas Blaschke, and Hannes Taubenböck. 2018. "Are we in boswash yet? A multi-source geodata approach to spatially delimit urban corridors." *ISPRS International Journal of Geo-Information* 7 (1). https://doi.org/10.3390/ijgi7010015.

Glocker, D. 2018. "The Rise of Megaregions: Delineating a new scale of economic geography," 51. https://www.oecd-ilibrary.org/development/the-rise-of-megaregions%7B/_%7Df4734bdd-en.

Hagler, Yoaf. 2009. "Defining U.S. Megaregions." *America2050*, no. November: 1–8. http://www.america2050.org/upload/2010/09/2050%7B/_%7DDefining%7B/_%7DUS%7B/_%7DMegaregions.pdf.

Hall, Tim, and Heather Barrett. 2012. *Urban Geography 4th Edition.* New York: Routledge.

He, Mark, Joseph Glasser, Nathaniel Pritchard, Shankar Bhamidi, and Nikhil Kaza. 2019. "Demarcating Geographic Regions using Community Detection in Commuting Networks," March. http://arxiv.org/abs/1903.06029.

Hillman, Nicholas, and Taylor Weichman. 2016. "Education Deserts: The Continued Significance of 'Place' in the Twenty-First Century." *Viewpoints: Voices from the Field.* https://www.acenet.edu/Documents/Education-Deserts-The-Continued-Significance-of-Place-in-the-Twenty-First-Century.pdf.

Kent, Jennifer L., and Susan Thompson. 2014. "The Three Domains of Urban Planning for Health and Well-being." *Journal of Planning Literature* 29 (3): 239–56. https://doi.org/10.1177/0885412214520712.

Knutsen, Bjørn Olav. 2011. "The European union as a small power. After the post-cold war." *Internasjonal Politikk* 69 (3): 547–51. https://doi.org/10.5860/choice.48-5952.

Ko, Y. 2012. "A study of unbalanced morphological understanding: Morphologicalland use patterns analysis of US megaregions." PhD thesis, Texas A&M University. https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/ETD-TAMU-2012-05-10847/KO-DISSERTATION.pdf?sequence=2%7B/&%7DisAllowed=y.

Kwak, James. 2017. *Economism: Bad Economics and the Rise of Inequality.* New York: Pantheon.

Lang, Robert E, and Dawn Dhavale. 2005. "Beyond Megalopolis: Exploring America's New "Megapolitan" Geography." Metropolitan Institute at Virginia Tech.

Lang, Robert, and Arthur Nelson. 2011. *Megapolitan America.* New York: Routledge.

Lawrence, Shawn A., Jeanette Garcia, Chris Stewart, and Celine Rodriguez. 2021. "The mental and behavioral health impact of COVID-19 stay at home orders on social work students." *Social Work Education* 00 (00): 1–15. https://doi.org/10.1080/02615479.2021.1883582.

Maricopa Association of Governments, Federal Highway Administration, Strategic Highway Research Program 2, Maricopa Association of Governments, Federal Highway Administration, and Strategic Highway Research Program 2. 2017. "Modeling a Multi-Modal Megaregion: Arizona's Sun Corridor Sheds Light on What Drives Industry and Freight: Freight Demand Modeling and Data Improvement Implementation Support." https://ops.fhwa.dot.gov/publications/fhwahop17033/fhwahop17033.pdf%7B/%%7D0Ahttps://rosap.ntl.bts.gov/view/dot/38459%7B/%%7D0Ahttps://trid.trb.org/view/1552568.

Marull, Joan, Vittorio Galletto, Elena Domene, and Joan Trullén. 2013. "Emerging megaregions: A new spatial scale to explore urban sustainability." *Land Use Policy* 34: 353–66. https://doi.org/10.1016/j.land usepol.2013.04.008.

Mattern, Krista, and Jeffrey Wyatt. 2009. "Student Choice of College :" *Journal of College Admission*, 18–30. https://files.eric.ed.gov/fulltext/EJ838811.pdf.

Nelson, Garrett Dash, and Alasdair Rae. 2016. "An economic geography of the United States: From commutes to megaregions." *PLoS ONE* 11 (11). https://doi.org/10.1371/journal.pone.0166083.

Oden, M., and G. Sciara. 2020. "The salience of megaregional geographies for inter-metropolitan transportation planning and policy making." *J* 80. https://doi.org/10.1016/j.trd.2020.102262.

Pain, Kathy. 2017. "Megaregions Imaginaries: Excursions Through a Dialectical Maze." *Geographical Review* 107 (3): 536–50. https://doi.org/10.1111/j.1931-0846.2016.12171.x.

Provo, John A, and D Ph. 2015. "APA Economic Development Task Force Report Planning for Economic Development."

Ravens-Sieberer, Ulrike, Anne Kaman, Michael Erhart, Janine Devine, Robert Schlack, and Christiane Otto. 2021. "Impact of the COVID-19 pandemic on quality of life and mental health in children and adolescents in Germany." *European Child and Adolescent Psychiatry.* https://doi.org/10.1007/s00787-021-01726-5.

Read, Anna, David Morley, Catherine Ross, and Sarah Smith. 2017. "Multimodal Planning at the Megaregional Scale." https://www.fhwa.dot.gov/planning/megaregions/reports/megaregions%7B/_%7Dreport%7B/_%7D2017/fhwahep17091.pdf.

Ross, Catherine L. 2011. "Megaregions: Literature Review of Organizational Structures and Finance of Multi-jurisdictional Initiatives and the Implications for Megaregion Transportation Planning in the U.S." Federal Highway Administration.

Steinsson, Sverrir, and Baldur Thorhallsson. 2017. "'Small State Foreign Policy' by Sverrir Steinsson and Baldur Thorhallsson. 2017." *The Oxford Research Encyclopedia of Politics*, 1–26.

Stich, Bethany, Kyle Griffith, and Peter Webb. 2015. "The Gulf Coast Megaregion: In Search of a New Scale to Understand Freight Transportation and Economic Development 7."

Su, Shiliang, Zebin Liu, Yifan Xu, Jing Li, Jianhua Pi, and Min Weng. 2017. "China's megaregion policy: Performance evaluation framework, empirical findings and implications for spatial polycentric governance." *Land Use Policy* 63. https://doi.org/10.1016/j.landusepol.2017.01.014.

Taubenböck, H., and M. Wiesner. 2015. "The spatial network of megaregions - Types of connectivity between cities based on settlement patterns derived from EO-data." *Computers, Environment and Urban Systems*

54: 165–80. https://doi.org/10.1016/j.compenvurbsys.2015.07.001.

Transportation Planning: A Decision-Oriented Approach. 2001.

Tyrwhitt, Jaqueline, ed. 1947. *Patrick Geddes in India*. London: Lund Humphries.

United Nations. 2020. *World Cities Report 2020: Unpacking the Value of Sustainable Urbanization*. https://doi.org/10.18356/c41ab67e-en.

Xu, Jiang, and Anthony Yeh. 2015. *Governance and Planning of Mega-City Regions: An International Comparative Perspective*. New York: Routledge.

Zhang, Wenjia, Chenyu Fang, Lin Zhou, and Jiancheng Zhu. 2020. “Measuring megaregional structure in the Pearl River Delta by mobile phone signaling data: A complex network approach.” *Cities* 104 (May): 102809. https://doi.org/10.1016/j.cities.2020.102809.