A TYPOLOGY OF MAIN STREET COMMERCIAL CORRIDORS USING CLUSTER ANALYSIS

Molly Balzano

A THESIS

in

Historic Preservation

Presented to the Faculties of the University of Pennsylvania in
Partial Fulfillment of the Requirements of the Degree of

MASTER OF SCIENCE IN HISTORIC PRESERVATION

2018

_____
Advisor
Donovan Rypkema
Lecturer in Historic Preservation

_____
Program Chair
Frank G. Matero
Professor

ACKNOWLEDGEMENTS

I would like to thank several people for their assistance with this project. Thank you to Donovan Rypkema, my advisor, for encouraging me to explore this topic and also keeping my goals and expectations within scope. Thank you to Patrice Frey, CEO of the National Main Street Center, for granting me access to Main Street data and shapefiles and patiently answering my many emails. Thank you also to the many folks at the National Main Street Center and the State Coordinating Main Street offices for providing me with the yearly statistics collected by each of their local Main Streets. While this data wasn't used for the cluster analysis, it was useful for the interpretation of findings and will be appended to a larger dataset made available to the National Main Street for future analyses.

Lastly, thank you to my partner, Sean Morse, for his careful edits and thoughtful suggestions throughout this project.

TABLE OF CONTENTS

Figures iv

Glossary of Terms v

1. Introduction 1

2. Main Street Background 9

3. Scholarship 14

4. Furthering the Main Street Approach 17

5. Data Acquisition 20

6. Methodology 25

7. Findings 39

8. Next Steps 43

Bibliography 45

Appendix 50

FIGURES

Fig. 1: Main Street Four-Point Approach Graphic 11

Fig. 2: Outlier Box Plot and Bar Graph 26

Fig. 3: Raw Data Variance Bar Graph 27

Fig. 4: Scaled Data Variance Bar Graph 28

Fig. 5: Correlation Matrix 29

Fig. 6: Single Linkage Method Graphic 31

Fig. 7: Complete Linkage Method Graphic 32

Fig. 8: Average Linkage Method Graphic 33

Fig. 9: Ward's Linkage Method Graphic 33

Fig. 10: Agglomerative Linkage Coefficient Table 33

Fig. 11: Within-Cluster Sum of Squares Graph 35

Fig. 12: Average Silhouette Graph 36

Fig 13. Gap Statistic Graph 37

Fig. 14: Dendrogram with Six Clusters 38

Fig. 16: 3-D Cluster Graph 39

Fig. 17: Heat Map Showing Cluster Breakdown 41

Fig. 18: Peer City Identification Tool 44

GLOSSARY OF TERMS

National Main Street Center – A nonprofit organization that oversees the State Main Street Coordinating Programs and the local Main Streets. They are responsible for publishing and teaching the revitalization methodology (including any new research or findings on how to improve the methodology) and keeping track of baseline statistics for each local Main Street to track their performance. They are based in Chicago, IL.

Main Street State Coordinating Program – A state-level department that works to assist local Main Streets in that state (there are a few local Main Streets that exist without a State Coordinating Program, but this is uncommon). There are currently 44 state Coordinating Programs. These departments are often funded by the state and can exist as their own department or housed within other state departments (some Main Street Coordinating Programs are housed within the state tourism department or the preservation department). There is usually one person, a State Coordinator, that works in the department, but there have been as many as eight people employed in one Main Street State Coordinating department. This depends on how many local Main Streets there are and how active the program is.

Local Main Street – A local nonprofit set up by residents to help stabilize or revitalize a town's commercial corridor. These can be set up a number of different ways. Some are distinct nonprofits, some operate as CDCs or BIDs, and others are housed within larger city departments. Local Main Streets are required to have a paid director (or manager) and can have other paid or volunteer staff. They are also required to have a volunteer Board of Directors.

The Main Street Organization – Wording used to describe the entire Main Street ecosystem: national, state, and local players.

1. INTRODUCTION

To contextualize how this project fits in with Main Street's existing model, this paper begins with a history of the Main Street organization, the conditions that engendered its creation, and Main Street's current operational structure.

In the mid-20th century, the U.S. government responded to the changing urban landscape (migrations from city to suburb, the Great Migration, white flight, etc.) and the ensuing disinvestment and blight with urban renewal practices, viewed by today's scholars as exceedingly destructive to both America's built fabric and communities' longstanding social webs. Urban renewal practices involved the systematic demolition and clearance of existing, historic infrastructure in the hope that a new physical landscape would cure all of the social and economic woes of struggling communities.

> What troubled the reformers was not so much the belief that these "sordid quarters" took a heavy toll on their tenants as the fear that they would degrade the working class and destroy the whole society. This fear grew out of the widespread belief in environmental determinism, the notion, as one architecture critic put it, that man "is molded by his environments." ("Be the man what he may," he said, "be his aspirations of the highest, the good that is in him will be stifled if his house be bad and his surroundings worse").[1]

[1] Fogelson, Robert M., *Downtown Its Rise and Fall, 1880-1950*, New Haven: Yale University Press, 2008.

Of course, government clearance policies left countless problems in their wake.[2] They eventually fell out of favor after having little impact on reviving downtowns, but in their stead, rose a new approach to revitalization. In the 1970s, city officials and planners believed that massive infrastructure improvements would help to improve downtowns. Cities invested in highways, mass transit, "convention centers, malls, cultural centers, football stadiums, and baseball diamonds." Academics have mixed feelings about whether or not these improvements had beneficial effect. Author and historian Robert M. Fogelson wrote, "Freeways have probably done more to spur decentralization than to curb it; and the [modern] urban redevelopment projects have probably done as much to weaken the central business district as to strengthen it."[3] He goes on to write,

> In my view, the decline of downtown was a result not so much of the deterioration of mass transit and the proliferation of private automobiles, of too much traffic and too little parking, as of the American vision of the "bourgeois utopia"—and of the local, state, and federal policies that helped Americans to realize it.[4]

Clearly, residents in these communities, beset by demolition and rampant development in their neighborhoods, felt similarly. Anxious to rehabilitate their towns, but averse to the demolition of their historic neighborhoods (and the inevitable erosion of the social and cultural life fostered there), residents responded by creating a different

---

[2] Klemek, Christopher, *The Transatlantic Collapse of Urban Renewal: Postwar Urbanism from New York to Berlin*, Chicago, IL: University of Chicago Press, 2011.
[3] Fogelson, Robert M., *Downtown Its Rise and Fall, 1880-1950.*
[4] Ibid.

model for revitalization: a community-centered, community-led approach to addressing local problems. In the 60s and 70s, all across the country, neighbors got together and formed the first Community Development Corporations (CDC), Community Action Agencies (CAA), Business Improvement Districts (BID), and other community nonprofits. These organizations, headed by private individuals rather than government officials, were created to address issues that the centralized system of American governmental planning was ill-equipped to tackle.[5]

In their early inception, neighborhood nonprofit organizations served predominately two purposes: if a community received little attention and very few public resources, the nonprofit worked to leverage non-traditional funds and partnerships to keep the neighborhood and its residents afloat and, if possible, attract new private investment to the area. Conversely, if a community received unwanted attention and aid—say, large-scale plans constructed and enacted entirely by outside entities—neighborhood nonprofits gave residents a voice and helped them advocate for their own needs and aspirations during the redevelopment process.[6]

One of the first CDCs was the Bedford-Stuyvesant project, conceived in 1966 by Robert Kennedy. Franklin Thomas, head of the Bedford-Stuyvesant Restoration Corporation at that time, worked with then vice president of the Ford Foundation,

---

[5] Ibid.

[6] Hoffman, Alexander Von, "History Lessons for Today's Housing Policy," *History Lessons for Today's Housing Policy*, August 2012, Accessed February 23, 2018, http://webcache.googleusercontent.com/search?q=cache:http://www.jchs.harvard.edu/sites/jchs.harvard.edu/files/w12-5_von_hoffman.pdf.

Mitchell Sviridoff, to expand this model. In 1980, they conceived of a large independent organization to assist CDCs, known as the Local Initiatives Support Corporation (LISC), that would distribute grants, give loans, and offer technical assistance to CDCs. After only four years, "LISC had obtained more than $70 million from 250 corporations and foundations and three federal agencies and set up 31 branch offices, which raised funds from local sources."[7]

Although philanthropic and nonprofit support helped the movement to grow, it was government funding, particularly federal funding, that helped community development to thrive on a large scale. The Housing and Community Development Act of 1974 replaced destructive urban renewal programs with community development block grants (CDBGs) designed to aid the work of local community organizations. Governments had finally seen the benefits of community-led efforts and moved to support them.[8] Three years later, additional federal programs, such as the Urban Development Action Grant, were created to fund additional efforts in inner-city areas suffering extreme economic distress.

With new funding streams available, local governments turned to neighborhood nonprofit organizations and contracted them to pursue their own redevelopment work. These new funding streams, in tandem with rising support from state and federal officials, spurred the creation of additional CDCs, Community Action Agencies, Business

[7] Hoffman, Alexander Von, "The Past, Present, and Future of Community Development," Shelterforce.org, July 17, 2017, Accessed February 23, 2018, https://shelterforce.org/2013/07/17/the_past_present_and_future_of_community_development/.
[8] Ibid.

Improvement Districts, and other organizations to enact redevelopment work.[9] Following in the wake of this grassroots momentum, the National Trust for Historic Preservation launched Main Street in 1980. The Main Street program was developed as a commercial stabilization program, designed around the belief that commercial corridors were a valuable part of a community, not just economically but also as social and cultural hubs for the surrounding neighborhood. The hope was that revitalization of a commercial corridor would have a catalytic impact, attracting additional investment to the community at large.[10]

While there are several commercial stabilization programs—including BIDs and commercial-focused CDCs—Main Street is predominately focused on maintaining and reviving older commercial corridors. Older, historic communities have, on average, been hit the hardest by changing industries in the U.S., and their commercial corridors have suffered the worst from the shifting retail environment (first, consumers moving to the suburbs; next, the emergence of big box stores; and now, e-commerce). Fogelson, remarking on the fall of downtown through the early and mid-twentieth century, wrote,

> People who had moved to the periphery were no longer going downtown—or were going downtown less often. Instead, they were patronizing the outlying business districts, shopping at chain stores, doing business at branch banks, and relaxing at neighborhood restaurants and movie theaters.[11]

[9] Ibid.
[10] National Main Street Center, "Main Street Impact," Accessed March 28, 2018, https://www.mainstreet.org/mainstreetimpact.
[11] Fogelson, Robert M., *Downtown Its Rise and Fall, 1880-1950.*

Main Street sought to address these issues, and their success has led to steadily increasing membership. Today, Main Street is pervasive in the United States, similar in reach and impact to other commercial stabilization programs. [12]

While Main Streets can be considered comparable to other commercial stabilization programs in some ways, they differ in their organizational structure, and, by consequence, their suitability and impact in different types of communities (of note, some local Main Streets are structured as BIDS or CDCs). For example, BIDS require substantial momentum and neighborhood cohesion to organize and are most appropriate in areas with business vacancy rates below 20% (they can still exist in low income areas, but not areas with high commercial vacancy).[13] To establish a BID, local officials must confirm the majority of businesses support the creation of the program, and then the BID is authorized by state legislation.[14] BIDs are generally funded through taxes levied on business owners, but many draw on public funds (some BIDs are quasi-governmental). CDCs, by contrast, can be started by just few motivated community members; they don't require the broad cohesion of a BID and can be impactful in areas with high rates of vacancy not serviceable by a BID. CDCs are organized as 501(c)3s and are eligible for an array of government funding, including federal grants authorized

[12] Abello, Oscar Perry, "Business Improvement Districts Are More Than Just a Name on a Trash Can," NextCity, August 7, 2015, Accessed March 14, 2018, https://nextcity.org/daily/entry/business-improvement-districts-support-small-business

[13] "Starting a BID," NYC Small Business Services, https://www1.nyc.gov/site/sbs/neighborhoods/starting-a-bid.page.

[14] Armstrong, Amy, Ingrid Gould, Amy Ellen Schwartz, and Ioan Voicu, "The Benefits of Business Improvement Districts: Evidence from New York City," Furman Center for Real Estate and Urban Policy – NYU, Furmancenter.org, July 2007, Accessed March 2, 2018. http://furmancenter.org/files/publications/FurmanCenterBIDsBrief.pdf.

under Section 4 of the HUD Demonstration Act of 1993. In addition, as non-profit institutions, CDCs are tax-exempt and may receive unlimited donations and grants from private and public sources.[15]

> A significant portion of funding comes from local government and through state and federal grants, such as the U.S. Department of Housing and Urban Development's Community Development Block Grant. CDCs can also receive funding from philanthropic foundations like the Ford Foundation and the Surdna Foundation. CDCs may also apply for funding through intermediary organizations (like the Local Initiative Support Corporation and NeighborWorks America nationally and local organizations like Pittsburgh's Neighborhood Allies) that receive government resources and then allocate funding to community groups.[16]

In contrast to CDCs and BIDs, local Main Streets can be structured in a number of different ways. The organizational structure of Main Streets is adaptable, depending on the needs of a particular community, and this makes them well-suited to respond to the needs of almost any commercial corridor. Main Streets are flexible in other areas as well. Like CDCs, they don't require broad cohesion and can be started by just a few motivated community members; in addition, they can be effective in communities that have experienced severe disinvestment and have a high vacancy rate. Unlike CDCs and BIDS, Main Streets have state *and* national oversight. Neither CDCs nor BIDS have centralized, administrative oversight (there was a national organization that oversaw

---

[15] Rachid Erekaini. "What is a Community Development Corporation?" NACEDA, Sept. 17, 2014, Accessed March 2, 2018, https://www.naceda.org/index.php?option=com_dailyplanetblog&view=entry&category=bright-ideas&id=25%3Awhat-is-a-community-development-corporation-&Itemid=171

[16] Ibid.

CDCs, called NCCED, but it dissolved in 2006).[17] This lack of oversight means success can vary widely from community to community and there is often little-to-no communication or shared insights exchanged between entities. [18] The State and National Main Street organizations provide much-needed support for local Main Street members—monetary and educational—and broaden the network of any individual community. Notably, Main Street has a forum for local Main Street members to exchange ideas and ask questions, and they host a yearly conference to update communities on recent research, trends, and opportunities for funding.

While Main Street may have a slightly narrower focus—older, historic commercial corridors—they are now comparable in size, scope, and impact to CDCs and BIDs. Today there are more than 1,600 communities with Main Street programs (close to the number of CDCs, and almost double the number of BIDs). There are Main Streets in 46 states, ranging in size from tiny rural, single-road Main Streets, to dense commercial strips in major metropolitan cities.[19]

[17] Simon, Harold, "Season of Change," Shelterforce.org, September 23, 2006, Accessed March 10, 2018. https://shelterforce.org/2006/09/23/season_of_change/.
[18] Rachid Erekaini, "What is a Community Development Corporation?"
[19] National Main Street Center, "The Programs," Accessed March 28, 2018, https://www.mainstreet.org/theprograms.

2. MAIN STREET BACKGROUND

By the middle of the 20th century, older, historic Main Streets had suffered: businesses closed, local jobs disappeared, and storefronts left vacant looked worse for wear as each year went by. The National Trust for Historic Preservation launched their pilot program in 1980 in attempt to combat disinvestment in these historic commercial corridors. The pilot program was an attempt to ascertain whether or not a targeted program of revitalization could breathe new life into local main streets: save the historic buildings, bolster local businesses, and reinstate Main Street as a social and cultural hub for the community. The National Trust developed a series of steps that residents could take—regardless of their access to public or private resources—to turn around their downtowns. The hope was that incremental efforts, small at first, would eventually create noticeable improvement and attract new public and private investment to the area.[20]

This program for Main Street revitalization was piloted in three small American towns: Galesburg, IL; Madison, IN; and Hot Springs, SD. A Main Street Manager—akin to a CDC director—was assigned to each city to guide the efforts and tweak the program as needed. After three years and considerable progress, the program was deemed a

[20] National Main Street Center, "The Main Street Movement," Accessed March 28, 2018, https://www.mainstreet.org/themovement.

success. It was formalized, incorporating learnings from the pilot cities, and steadily expanded to more areas.

Main Street is unique in its approach in that it is specifically designed to maximize small amounts of capital and grassroots effort for maximal return. The National Main Street Center calls its tailored approach for downtown revitalization the "Four-Point Approach." The Four-Point Approach leverages towns' existing assets—historic infrastructure and any other defining features—and significant sweat equity to achieve change over time. This approach is best described as asset-based community development or place-based community development (activating a space using design and events to spur interest, public, and private investment). Interventions are conceived at the local level and implemented by residents and community organizations. Main Street believes, and has proven in many communities, that incremental changes, especially with local buy-in, are more stable and long-lasting than the big, quick fixes often employed by planners—demolition and reconstruction chief among them. The revitalization approach is designed to be grassroots—a way for residents to bring back their Main Street, even if local government isn't participating or there's little money to be found.

The Four-Point Approach is meant to be all-encompassing, addressing the myriad reasons that a commercial corridor has declined or struggles to survive. The Four-Point

Approach divides the oft-daunting, comprehensive task of revitalization into four focus areas for improvement: Organization, Economic Vitality, Promotion, and Design.[21]

Fig. 1: Main Street Four-Point Approach Graphic, mainstreet.org

To improve organization, Main Street aids residents in organizing a team—a local manager and a board of directors—to lead the grassroots revitalization effort. Main Street also works to connect disparate stakeholders and organizations already doing work in the town (say, a CDC, local library, community center, or gardening club). Main Street works with these groups to form a cohesive goal for commercial corridor

[21] National Main Street Center, Kennedy Smith, and Josh Bloom, *The Main Street Approach: A Comprehensive Guide to Community Transformation,* Report, Accessed September 10, 1017, http://www.mainstreet.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=01cf95e3-5e71-ae73-902f-1b0e9494ceaa&forceDialog=0.

revitalization. By unifying existing organizations and volunteers, Main Street helps to maximize existing resources and efforts and point these toward targeted goals.[22]

To improve economic vitality, Main Street helps to stabilize and grow existing businesses and works on attracting new businesses to fill vacancies. The National Main Street Center and State Main Street Coordinating Programs offer a variety of resources to business owners. There are classes on a wide-range of topics, from tax assistance, to social media marketing, to succession planning. Many states also offer free services, such as market analyses, and offer financial aid, such a storefront improvement grants and revolving loans to aid business owners. In addition to these efforts, Main Street works to attract new businesses and grow local entrepreneurs to fill store vacancies and build a robust commercial environment.[23]

To help promote the commercial corridor, Main Street schedules activities and events to bring people back downtown and shopping local once again. Main Street also helps communities to develop their unique image and offerings (say, their proximity to a natural resource, their food or music or art scene, or their historic significance). Main Street works to market this unique sense of place to generate awareness and attract people back to the area.[24]

Lastly, to improve design and the physical appearance of the commercial corridor, Main Street helps residents complete small, incremental improvements that

[22] Ibid.
[23] Ibid.
[24] Ibid.

aren't financially prohibitive—possible for even the most resource-strapped communities (e.g. grooming public spaces, painting, plantings). State Coordinating Programs and The National Center often provide free design review, architectural assistance, and preservation education to help with these improvements.[25]

The Four-Point Approach is uniquely crafted to revitalize and maintain commercial corridors—it can be employed in areas where other approaches have failed. Barring BIDs, a small number of nonprofits focus on commercial corridor revitalization exclusively (and none of them come near the scope and influence of Main Street). Of the nonprofits that *do* focus on commercial revitalization, scarce few are poised to deal with communities that have faced severe disinvestment, loss of a primary industry (say, manufacturing or resource extraction), high vacancy, or other challenges.

Main Street does exclude most modern commercial areas. It focuses primarily on those commercial corridors constructed before the 1950s (though a few were constructed as late as the 1970s). While this excludes suburban retail centers and other newer commercial construction, it allows Main Street to focus on the unique needs of older commercial corridors that were built as central nodes in their cities—housing small, local businesses, employing local residents, and often serving as a social hub for the surrounding neighborhood. The Main Street approach is designed for these types of multifaceted historic commercial corridors, attempting to address all or most aspects of

[25] Ibid.

the economic, social, and cultural features of the corridor and the surrounding community.

3. SCHOLARSHIP

> While detailed statistics are kept to track the activity of Main Street related investments, the reporting is based on descriptive statistics and economic impact results only; there is a dearth of serious study of this widely renowned and successful program. [26]

Both CDCs and BIDS have been researched extensively by governmental organizations, NGOs, and academics. Much has been written about their efficacy and opportunities for improvement. Academics have likely been drawn to studying CDCs and BIDs because they have been in existence for over 50 years, appear to have organizational staying power, draw from public funds, and their collective interventions affect a significant number of people in the U.S. and abroad.[27]

It's possible that academics haven't yet felt the draw towards Main Street because the Main Street Organization was established later (the first pilot program in 1980 was launched to little fanfare), it grew slowly, and it only reached a capacity of note in the last decade. In addition, there may have been some uncertainty about Main

[26] Mason, Randall, "Economics and Historic Preservation, A Guide and Review of the Literature," Brookings Institute, Brookings.edu, September 2005, Accessed February 23, 2018, https://www.brookings.edu/wp-content/uploads/2016/06/20050926_preservation.pdf

[27] "An Overview of the Literature on Community Development Corporations," RA Berger and G. Kasper, *Nonprofit Management and Leadership*, Winter 1993.
See also: *Journal of the Community Development Societ*y and the *Journal of Urban Affairs.*

Street's staying power as an organization—they transitioned from a financially dependent arm of the National Trust to financially independent subsidiary in 2013 and fought hard to stay afloat during that time. However, despite a few minor setbacks, the Main Street Organization has grown increasingly since its inception and markedly since becoming an independent subsidiary in 2013, ultimately proving the staying power of the organization and the usefulness of their programming.

Perhaps researchers in planning and preservation overlook Main Street because the organization straddles two fields. As a subsidiary of the National Trust for Historic Preservation, Main Street may appear to many planners and community developers like a preservation organization. Alternatively, to preservationists, Main Street may look as though it extends too far beyond the confines of preservation, more akin to a community development organization.

In any case, Main Street has been now been in existence for 38 contiguous years, has proven their long-term sustainability as an independent subsidiary (financially and programmatically independent from the National Trust for Historic Preservation), and has grown steadily since its founding. Main Street consultant Donovan Rypkema writes,

> In the last 25 years, some 1,700 communities in all 50 states have had Main Street programs. Over that time, the total amount of public and private reinvestment in those Main Street communities has been $23 billion. There have been over 67,000 net new businesses created, generating nearly 310,000 net new jobs. There have been 107,000 building renovations. Every dollar invested in a local Main Street program leveraged nearly $27 of other investment. The

> average cost per job generated—$2,500—is less than a tenth of what many state economic development programs brag about.[28]

Main Street has long reached the capacity and national influence that makes it deserving of robust study and scholarship. Research is long overdue in exploring the following areas: 1) how Main Street has impacted communities; 2) where there are opportunities for greater efficacy; and 3) the cost-benefit of choosing Main Street over other commercial stabilization or community development interventions.

The National Main Street Center does publish case studies to show how the program has been effective and collects simple, descriptive statistics for each of its local Main Streets (this includes: net new jobs, net new business, public investment dollars, private investment dollars, and number of buildings rehabilitated).[29] But, unfortunately, they have not been able to launch more rigorous studies. Rypkema writes,

> Main Street data as currently gathered, while useful, does not meet the standards of robust, defensible research. There is no ongoing measurement of preservation-based commercial revitalization not affiliated with Main Street, except in limited ways through CDBG. There is no comparison of what is happening in Main Street communities and similar non-Main Street communities.[30]

[28] Rypkema, Donovan, "Heritage Conservation and the Local Economy," Report, August 2008, Accessed March 1, 2018, http://www.globalurban.org/GUDMag08Vol4Iss1/Rypkema%20PDF.pdf.
[29] Mason, Randall, "Economics and Historic Preservation, A Guide and Review of the Literature."
[30] Rypkema, Donovan and Caroline Cheong, "Measuring Economic Impacts of Historic Preservation: A Report to the Advisory Council on Historic Preservation," Report, August 2008, Accessed March 1, 2018, http://www.globalurban.org/GUDMag08Vol4Iss1/Rypkema%20PDF.pdf.

The National Main Street Center, like most nonprofits, has limited resources to undertake this research, but academics can explore many of these topics. I have attempted to fill some of the research gaps with this project. I believe a Main Street typology would help to striate Main Street's many local communities into statistically similar groups for both: 1) more targeted programming; and 2) more focused research on the efficacy of Main Street in distinct community types.

4. FURTHERING THE MAIN STREET APPROACH

The Four-Point Approach, as it is currently conceived, is applied too broadly: while somewhat adaptable, the same method is applied to vastly different communities with markedly different needs. Communities can try to tailor their revitalization approach to their specific needs, but each community must endeavor to customize programming from the core methodology. And while the methodology is designed to be nimble (If one route—a certain economic strategy or brand/identity—isn't moving the needle, the town can change course), it can be challenging to implement when community members with little experience can't determine which direction to go and can't afford tailored guidance or consulting. Local Main Streets can work with their state-level Main Street Coordinating Program, but again, the universe of possible

options (where to start, how to position the town, which communities are comparable) is vast, for both the local program and the state coordinators.

The mass-application of the Four-Point Approach could be refined to target specific programming to different types of communities. In fact, Main Street has recognized the need for more specialization in their approach—in August 2017, they launched Urban Main, an offshoot of Main Street that's designed specifically to address the needs of Main Streets in major metropolitan cities. This program more adequately addresses the needs of urban communities and urban businesses—aiding with unique political environments, metropolitan transportation issues, security, and gentrification—but this specialized approach only exists for urban Main Streets.[31]

All Main Streets, not just those in urban areas, could benefit from more targeted programming and recommendations tailored to their unique needs. While it is infeasible for the National Main Street Center to develop comprehensive programming for each of its 1,600 members, the Four Point Approach could be significantly improved if it were tailored to address the needs of similar *types* of communities. A cluster analysis can be used to group Main Street communities into statistically similar clusters, based on comparable attributes. This segmentation would be useful to several types of practitioners within the Main Street organization. A segmentation would enable the National Main Street Center to develop targeted strategies for each community-cluster,

[31] National Main Street Center, "Urban Main," Accessed March 28, 2018, https://www.mainstreet.org/themovement.https://higherlogicdownload.s3.amazonaws.com/NMSC/390e0055-2395-4d3b-af60-81b53974430d/UploadedImages/UrbanMain/NMSC30_FAQ_GENERAL_2.pdf.

increasing the efficacy of revitalization efforts. In addition, at the local level, a segmentation would enable communities to identify comparable Main Streets and ascertain the strategies most likely to succeed in their own community. Lastly, communities could leverage this information to make a case for funding or policy that has succeeded in communities similar to their own (for better or worse, in the existing climate of data-driven decision making, statistical findings can hold more weight than empirical evidence).

While this type of statistical analysis hasn't been done for Main Streets before, several scholars have used cluster analysis to segment neighborhoods (researchers have used cluster analysis to examine everything from health outcomes to demographic makeup of different neighborhoods).[32] Perhaps the most comparable project to this research is the Peer City Identification Tool developed by the Community Development and Policy Studies (CDPS) division of the Federal Reserve Bank of Chicago. The CDPS created a front-end application to explore the similarities and differences between 960 American cities.

> [The Peer City Identification Tool] is a data comparison and visualization instrument that can help policymakers and practitioners understand a municipality in the context of peer cities. Drawing on city-level indicators from the American Community Survey and historical Decennial Census records, the

[32] Reibel, Michael, and Moira Regelson, "Quantifying Neighborhood Racial and Ethnic Transition Clusters in Multiethnic Cities," *Urban Geography* 28, no. 4 (2007): 361-76, doi:10.2747/0272-3638.28.4.361. ;and, Pedigo, Ashley, William Seaver, and Agricola Odoi, "Identifying Unique Neighborhood Characteristics to Guide Health Planning for Stroke and Heart Attack: Fuzzy Cluster and Discriminant Analyses Approaches," *PLoS ONE* 6, no. 7 (2011), doi:10.1371/journal.pone.0022693.

> PCIT performs a cluster analysis to identify groups of similar cities along economic, demographic, social, and housing dimensions. [33]

The Peer City Identification Tool used a hierarchical cluster analysis and Ward's Linkage method, the same method used for this project (explained in section 6 of this paper). Advanced data analysis such as this were valuable comparisons for their selection of variables, analytic processes, and expected outcomes and issues. Several studies were used to inform the variable selection and cluster method used to create the Main Street typology.

5. DATA ACQUISITION

Before collecting any data, it was necessary to define the parameters for which Main Street communities would be included in this analysis. For the sake of time, this analysis could only be performed on those local Main Streets that had already been mapped in GIS by the National Main Street Center. To date, the National Center has mapped 1028 of their Main Street commercial corridors. Additional mapping for any Main Streets not yet included in the National Center's shapefile was outside the scope of this project. Second, only accredited and affiliate Main Streets were included for this study, excluding general members. Accredited Main Streets have to meet certain

[33] Federal Reserve Bank of Chicago, "Peer City Identification Tool," Accessed April 2, 2018, https://www.chicagofed.org/region/community-development/data/pcit.

criteria—determined using a 10-point accreditation check-list[34]—to prove they are actively implementing the Main Street Four-Point Approach. Affiliate Main Streets are those Main Streets that are working towards accreditation and implementing the Main Street methodology, but have not yet met all criteria for accreditation. Using these parameters, 905 local Main Streets were selected for the cluster analysis.

For this analysis, data was pulled by census block group. Block groups were used because census block-level data is only available for decennial censuses. The American Community Survey (ACS) uses surveys and statistical analyses to forecast population and demographic changes. The 2015 ACS was used for this study because the 5-year ACSs use 60 months of data to generate their estimates; whereas the 1-year ACS's use just 12 months of data and have a larger margin of error.[35]

Most commercial corridors have limited downtown housing, so in order to gather demographic data, the boundaries of the study area had to be expanded beyond the confines of the commercial area to the Primary Trade Area (sometimes referred to

[34] To pass the 10-point checklist, a community: "1. Has broad-based community support for the commercial district revitalization process, with strong support from both the public and private sectors 2. Has developed vision and mission statements relevant to community conditions and to the local Main Street program's organizational stage 3. Has a comprehensive Main Street work plan 4. Possesses an historic preservation ethic 5. Has an active board of directors and committees 6. Has an adequate operating budget 7. Has a paid professional program manager 8. Conducts a program of ongoing training for staff and volunteers 9. Reports key statistics 10. Is a current member of the Main Street America™ Network." National Main Street Center, "Main Street Tier System Overview," Accessed March 28, 2018, https://higherlogicdownload.s3.amazonaws.com/NMSC/390e0055-2395-4d3b-af60-81b53974430d/UploadedImages/Join/Main_Street_America_Tier_System_Overview.pdf

[35] American Community Survey Office, "ACS Summary File Technical Documentation," September 2016, Accessed January 18, 2018, https://www2.census.gov/programs-surveys/acs/summary_file/2015/documentation/tech_docs/2015_SummaryFile_Tech_Doc.pdf.

as the Retail Trade Area or Consumer Trade Area).[36] A Primary Trade Area (PTA) delineates the area in which consumers/visitors to the downtown area likely live. While there are many ways to determine a PTA (real drive time, retail gravity zones, etc.), the most common method is to draw a smooth buffer that extends two miles beyond the periphery of the commercial zone. According to the Urban Land Institute, a 2-mile buffer represents an approximate 5-10-minute drive and should capture roughly 80% of consumers.[37] While drive time can differ in rural, mid-size, and urban communities depending on transportation infrastructure and geographic obstacles, this is the method most widely used by both planners and private businesses to capture a commercial area's surrounding market.[38]

For each Main Street community, block groups that intersected with the PTA were selected, then data was pulled and totaled for those block groups. With so many communities, only publicly available data that could be obtained at scale was within scope. The following variables were selected for this analysis: population (density was calculated), income, education, race, household makeup, length of tenure, own vs. rent, employment, median home value, median rent, industries, and distance to the next nearest Main Street. Of the demographic and housing data that was considered, only those variables that would reasonably differentiate a community were chosen. For

[36] Ooi, Joseph T.l., Gaylon E. Greer and Phillip T. Kolbe, "Investment Analysis for Real Estate Decisions," Dearborn Real Estate Education, *Journal of Property Investment & Finance* 24, no. 3 (2006), doi:10.1108/jpif.2006.24.3.268.1.
[37] Beyard, Michael D., and W. Paul O'Mara, *Shopping Center Development Handbook*, Washington, DC: ULI-Urban Land Institute, 1999.
[38] Ibid.

example, a small difference in the gender makeup of a community is likely irrelevant, so this variable was excluded.

While many Main Streets have strong programing and events that certainly contribute to their success, it was not within the scope of this project to include that information. In addition, while it was possible to include some housing information (median year of housing, percent of vacant buildings, and tenure), nonuse features of the built environment were also excluded. Nonuse variables are variables that represent "values for which economic methods are ill-suited (values of beauty, memorial power, attachment, and other 'priceless' qualities)."[39] A more robust cluster analysis may include qualitative data about programming (specific design, promotion, or economic strategies); however, it was not feasible to incorporate that data into this project.

The variables that were chosen were selected for one of two reasons: they were used in comparable research employing cluster analysis; or, they were suggested by practitioners in the field—chiefly, Donovan Rypkema, Principal of PlaceEconomics, and Josh Bloom, Principal of Community Land Use and Economics Group—who have worked with countless Main Street communities across the U.S. In addition, the selected variables were affirmed by my own primary research during my time as an intern at the National Main Street Center in Chicago. While working at the National Main Street Center, I spent three months composing case studies on 11 local Main Street. My research included over 45 interviews with Main Street State Coordinators, local Main

[39] Mason, Randall, "Economics and Historic Preservation, A Guide and Review of the Literature."

Street Managers/Directors, private Main Street consultants, National Main Street Field Officers, and local politicians. During this research, interviewees conveyed what they believed to be the most impactful aspects of their towns—features that retain longtime residents, attract new ones, and draw in visitors. Interviewees cited some of the following aspects as a key to their success: their Main Street's proximity to a major urban area (serving as a bedroom community for local residents that worked in the city or as weekend destination for urban residents that wanted to get away to the Main Street); their Main Street's proximity to another Main Street (enabling regional strategies to draw tourists, sometimes shared financial resources and volunteer labor); their Main Street's resident makeup (some said families were the most important, some said younger residents, some said older residents, and others said diverse residents); and, the design of their Main Street (cost, quality, and history of buildings).

Local Main Streets' self-reported data was left out of this analysis for three reasons: these variables are not available both pre- and post-intervention, they are inconsistently recorded, and an analysis based on this type of data would be difficult to replicate—both by other researchers studying Main Street and by researchers using this methodology to study other community development organizations. Using only publicly-available data makes this process easier to emulate and makes it easier for Main Street to re-pull data in the future. In addition, the census data gathered for this project is available pre- and post-Main Street intervention, enabling additional analyses, including

regression models to test the impact of Main Street or to determine variables with outsize affect.

6. METHODOLOGY

Using R, the census variables were organized into tabular data, cleaned, and summarized via preliminary inferential analyses. Any anomalous data was either corrected or excluded where corrections were infeasible. A cluster analysis cannot be done using missing values, so NA values were excluded. In addition, a few outliers had to be removed as they had an overwhelming effect on the data. Nine Main Streets have a significantly larger area than average (most of these are in Montana where towns are very spread out). These towns were removed from the dataset as they negatively affected the model. Ideally, these nine towns can be fit back into appropriate clusters based on other attributes.

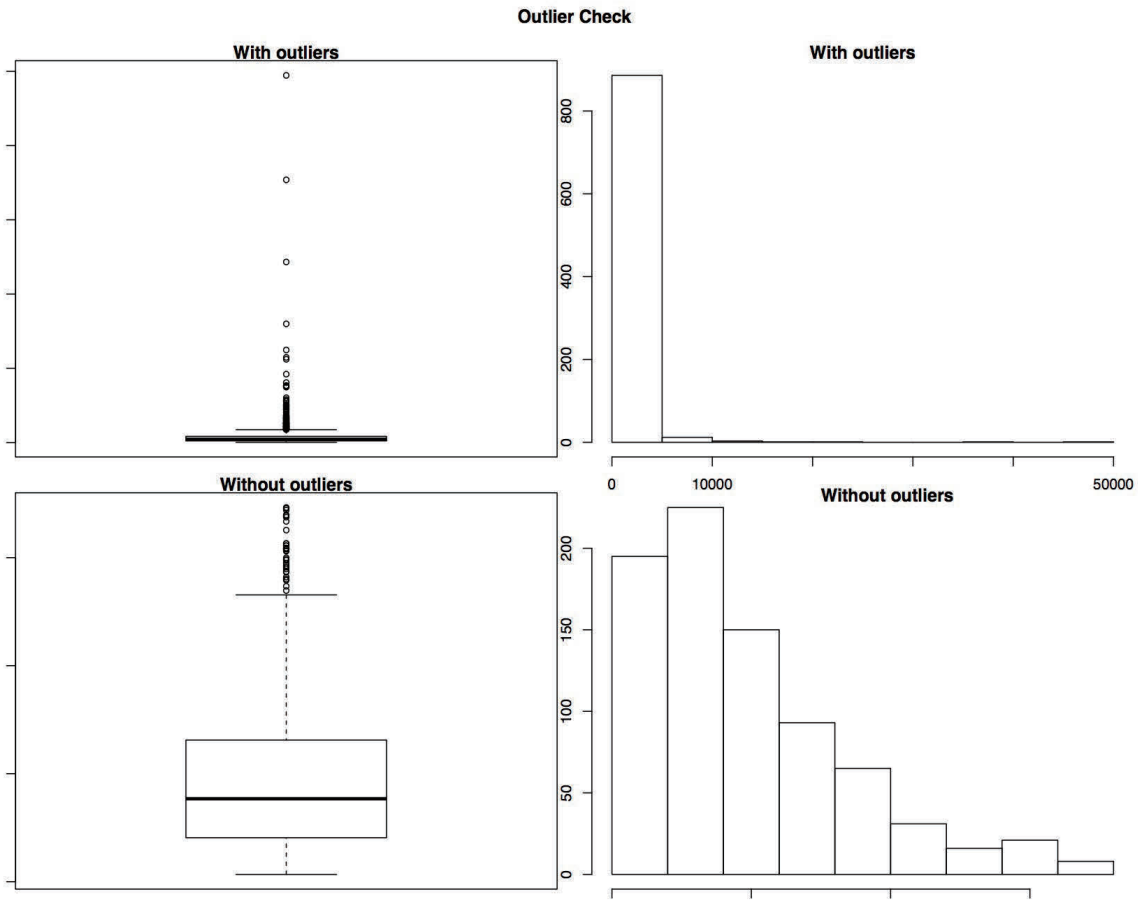


Fig. 2: Boxplot and bar graph showing outliers and skew for size of the Main Street (square miles). The dataset was negatively affected by outliers with large Main Street corridors.

After the data was cleaned and organized, the variables were transformed in preparation for the cluster analysis. With variables of vastly different scales (e.g. income, area of the main street, percentage of vacant homes), it is best practice to standardize the variables to make them comparable. This way, one variable doesn't disproportionately dictate the outcome of the model.[40] To achieve this, a z-score is

[40] Kabacoff, Robert I., *R in Action*: *Data Analysis and Graphics with R*, Shelter Island, NY: Manning, 2015.

calculated for each value, denoting how many standard deviations a raw number is from the mean of all numbers for a particular variable.

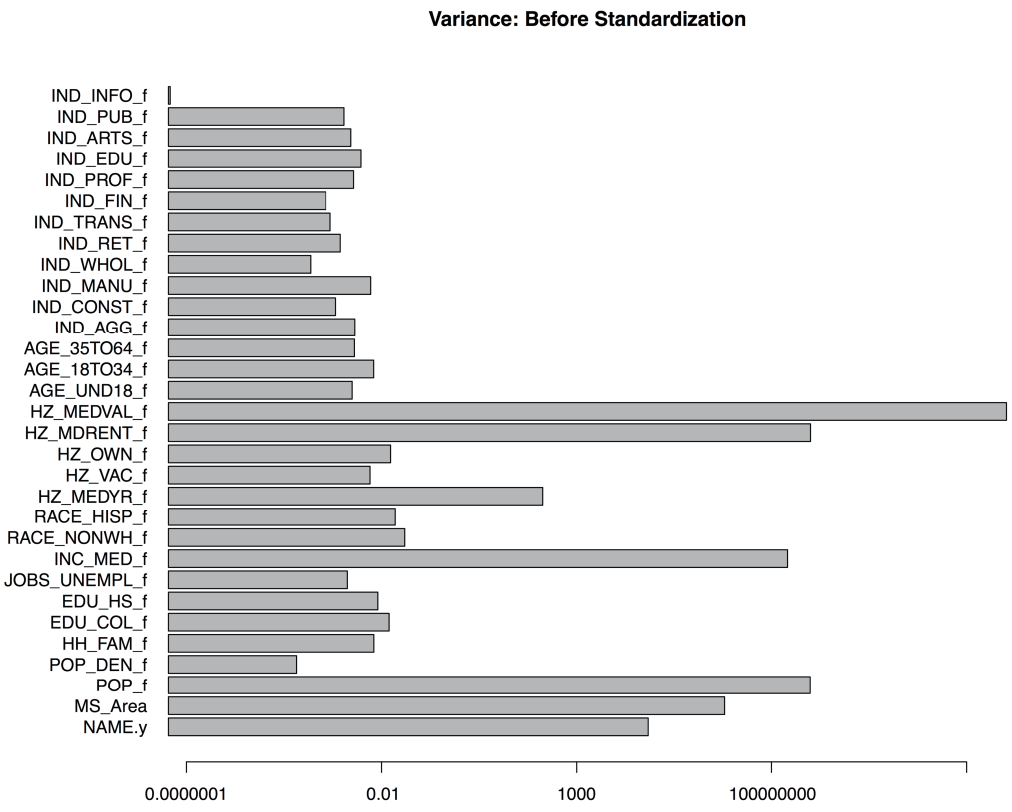


Fig. 3: Raw variables have significantly different variance

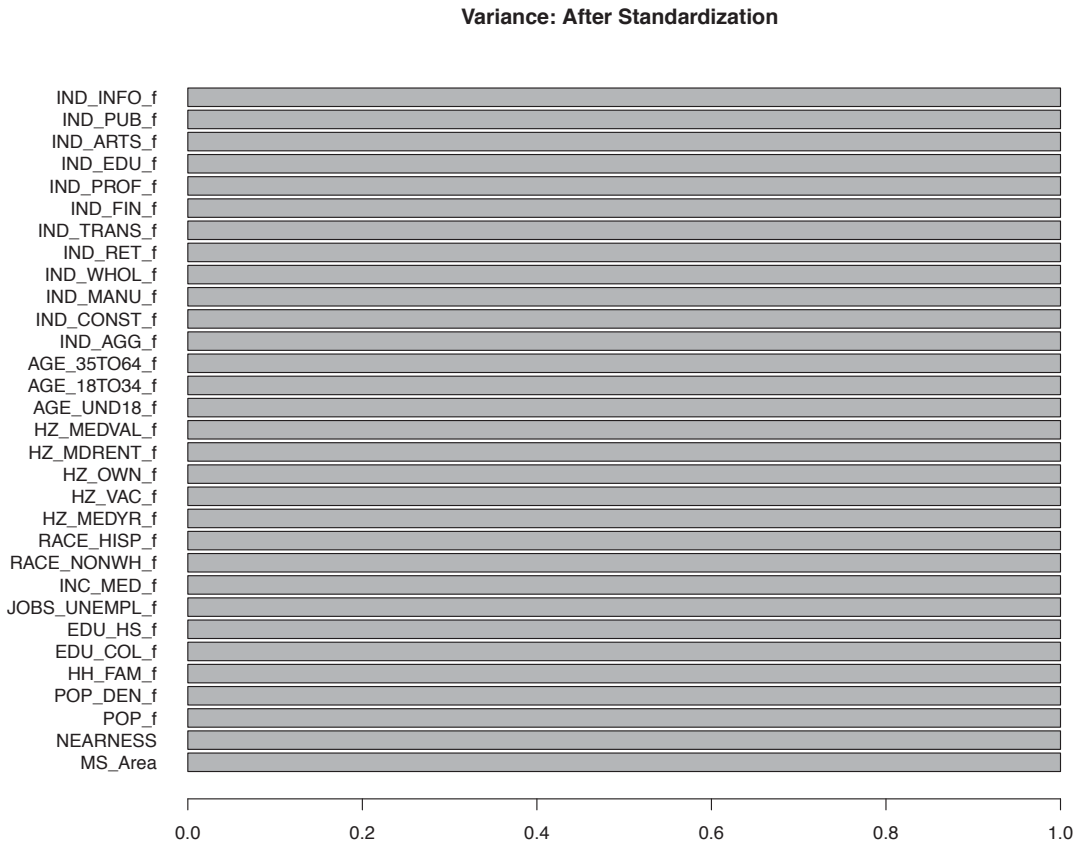


Fig. 4: Scaled variables have the same variance

After the data is scaled, the variables were analyzed to determine their correlation. Highly correlated variables can also have an outsize impact on the cluster—if the variables are highly correlated, they essentially count as the same variable with double the weight. In other words, the cluster will be hyper dependent on highly correlated variables to the exclusion of other variables. A correlation matrix shows that several variables were highly correlated: median income and college education; people under 18 and people living in a family household; and median home value and median

rent. Of these, median home value and median rent are the most highly correlated (almost 1).

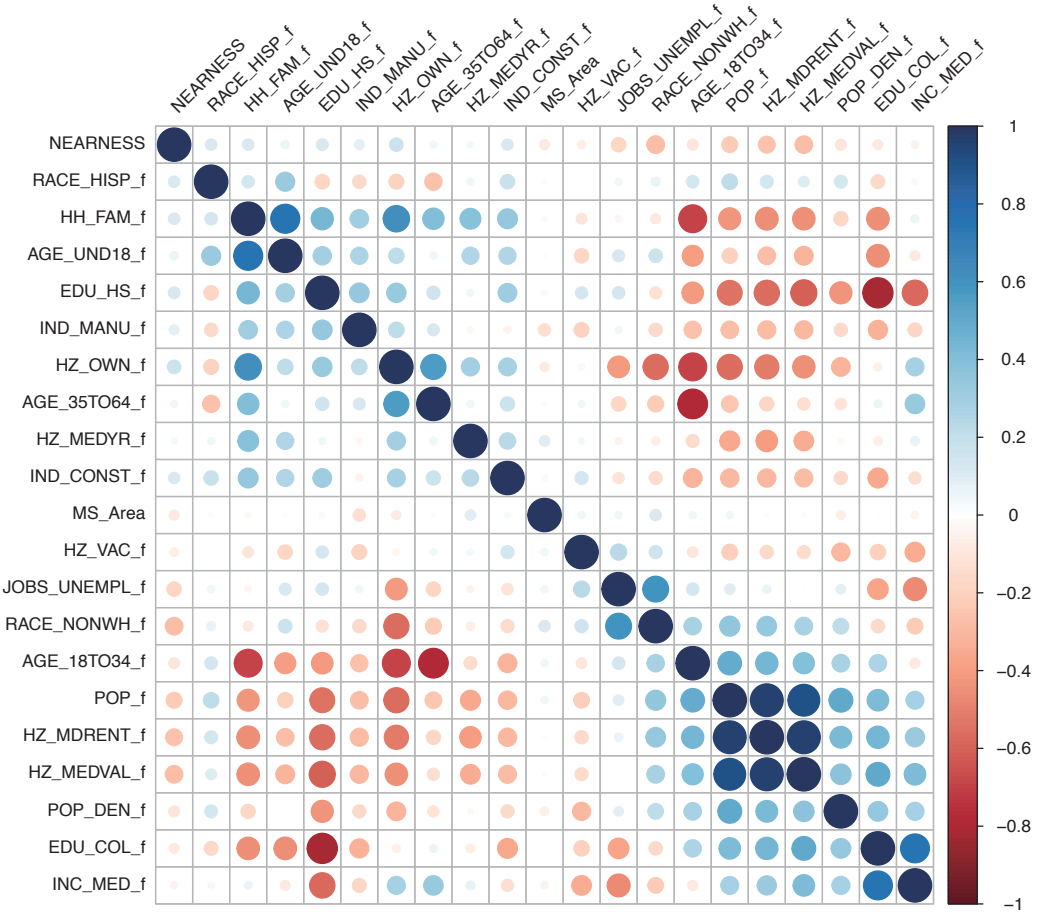


Fig. 5: Correlation Matrix showing correlated variables

There are several ways to deal with highly correlative data—the most popular is Principal Component Analysis (PCA). PCA reduces the set of variables to linearly uncorrelated components. This method is affective, but it masks the initial variables put

into the model, making it difficult to interpret the final results.[41] For the sake of interpretability, I chose to simply remove median rent, one of the highest correlated variables.

After this preparation, the data was ready to be clustered. There are multiple ways to cluster data, but with each method, the primary goal is the same: to separate objects into distinct groups so that objects within a group are as homogenous as possible and objects across groups are as heterogeneous as possible. The two most popular clustering methods are k-means clustering and hierarchical clustering. K-means clustering uses a pre-specified number of clusters and groups the observations into that specific number of groups. K-means can be useful when a researcher already has an idea of how the data will naturally split (or if the researcher runs exploratory models to determine the ideal number of groups). Hierarchical clustering is an alternative approach that does not rely on a pre-specified number of clusters and is much easier to visually interpret.[42] Hierarchical clustering was chosen for this project because of its interpretability, important for an organization with wide membership and varying skills and proficiencies. A hierarchical cluster clearly shows where the communities were divided at each step of clustering, even beyond the chosen number of clusters.[43]

There are two types of hierarchical clustering: agglomerative hierarchical clustering, or bottom-up clustering (also called AGNES), and divisive hierarchical

[41] Ibid.
[42] Ibid.
[43] Ibid.

clustering, or top-down clustering (also called DIANA). Agglomerative clustering begins with all of the observations and iteratively determines which observations are most similar; whereas, divisive clustering begins by determining iteratively which splits would make groups most different. Agglomerative hierarchical clustering was used for this project.

When clustering objects, there are several ways to calculate the similarity—or linkage—between two observations. The linkage method can have a significant impact on the way the data is clustered. The most common linkage methods are single, complete, average, and ward, shown below.

Single Linkage Method

Fig. 6: Single Linkage (Minimum Distance): "Groups are formed from the individual entities by merging nearest neighbors… an object will be added to a cluster so long as it is close to any one of the other objects in the cluster, even if it is relatively far from all the others."[44]

[44] Moral, Roger Del, "On Selecting Indirect Ordination Methods," *Classification and Ordination,* 1980, 75-84, http://www.math-stat.unibe.ch/unibe/portal/fak_naturwis/a_dept_math/a_dept_ms/content/e237483/e237655/e243381/e281679/files281690/Chap11_ger.pdf.

Complete Linkage Method

Fig. 7: Complete Linkage (Maximum Distance): "Complete linkage clustering proceeds in much the same manner as single linkage clustering, with one important exception: At each stage, the distance between clusters is determined by the distance between the two elements, one from each cluster, that are most distant. Thus, complete linkage ensures, that all items in a cluster are within some maximum distance of each other…it can be highly sensitive to outliers."[45]

Average Linkage Method

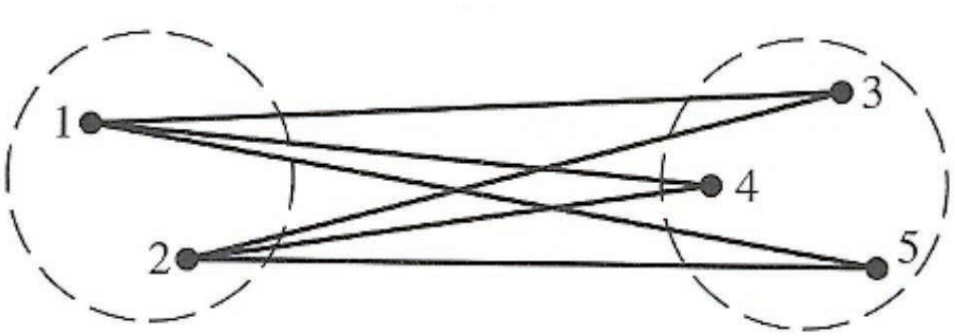


Fig. 8: Average Linkage (Average Distance Between All Pairs of Items): "Average linkage treats the distance between two clusters as the average distance between all pairs of items where one member of a pair belongs to each cluster."[46]

[45] Ibid.
[46] Ibid.

Ward's Linkage Method

Fig. 9: Ward's Linkage (Minimum Within Group Variance): "Seeks to join the two clusters whose merger leads to the smallest within-cluster sum of squares."[47]

Calculating the agglomerative coefficient identifies the amount of clustering structure for each linkage method. The closer to 1 the coefficient, the stronger the clustering structure. The table below shows that Ward's linkage method produces the strongest clustering structure.

Linkage Method			
Single	Complete	Average	Ward's
0.83826	0.67287	0.90533	0.97041

Fig. 10: Agglomerative Linkage Coefficient Table

The Main Street data was clustered using an agglomerative hierarchical cluster, Euclidean distance, and Ward's linkage method. The hierarchical method of clustering

[47] Ibid.

produces a dendrogram, or a tree-like representation of how the different objects in the group were split. In this case, the "leaves" of the dendrogram represent each individual Main Street. As Main Streets are grouped together, they form branches. The resulting tree can be "cut" at any point along the tree to derive the number of branches/groups. There are several ways to determine the ideal number of clusters: graphing the total within-cluster sum of squares for each number of clusters, graphing how well each object lies within its cluster, and graphing the total inter-cluster variation.[48]

The first method used to determine where to cut the dendrogram—or the optimal number of clusters—is the Elbow Method, which graphs the total within-cluster sum of squares. The ideal number of clusters will be the one in which the total intra-group variation—or total within-cluster sum of squares—is minimized.[49] The Elbow Graph for the Main Street data shows that as the number of groups increases, the within-group sum of squares—or, the total intra-group variation—declines. A notable bend in the graph would suggest an ideal number of clusters, as it suggests that that number of clusters results in a significant improvement in the model (if the line flattens out, it suggests that further clustering adds less improvements to the model). The following graph has few notable elbows suggesting it may be acceptable to choose a variety of cluster numbers.

[48] "K-means Cluster Analysis," *K-means Cluster Analysis - UC Business Analytics R Programming Guide*, Accessed February 20, 2018, https://uc-r.github.io/kmeans_clustering#gap.
[49] Ibid.

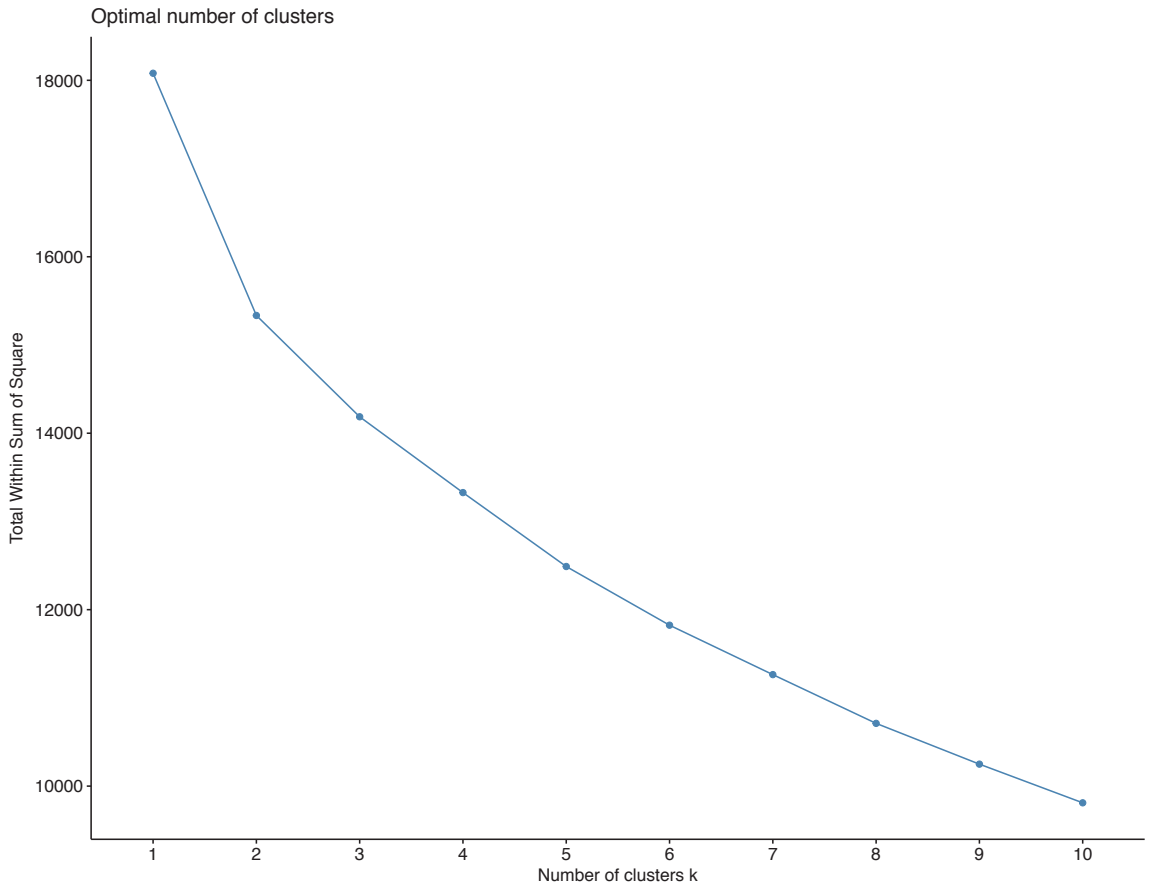


Fig. 11: Elbow Method: Total Within-Cluster Sum of Squares

There is a modest bend between three and six clusters. An additional method helps to determine which of these cluster numbers is preferable.

The second method used to ascertain the optimal number of clusters is the Silhouette Method. "The silhouette method determines how well each object lies within its cluster. A high average silhouette width indicates a good clustering."[50] The following

[50] Ibid.

graph shows a significant decline at four clusters, a plateau between four and six clusters, then a gradual rise, suggesting poorer clustering silhouettes after six clusters.

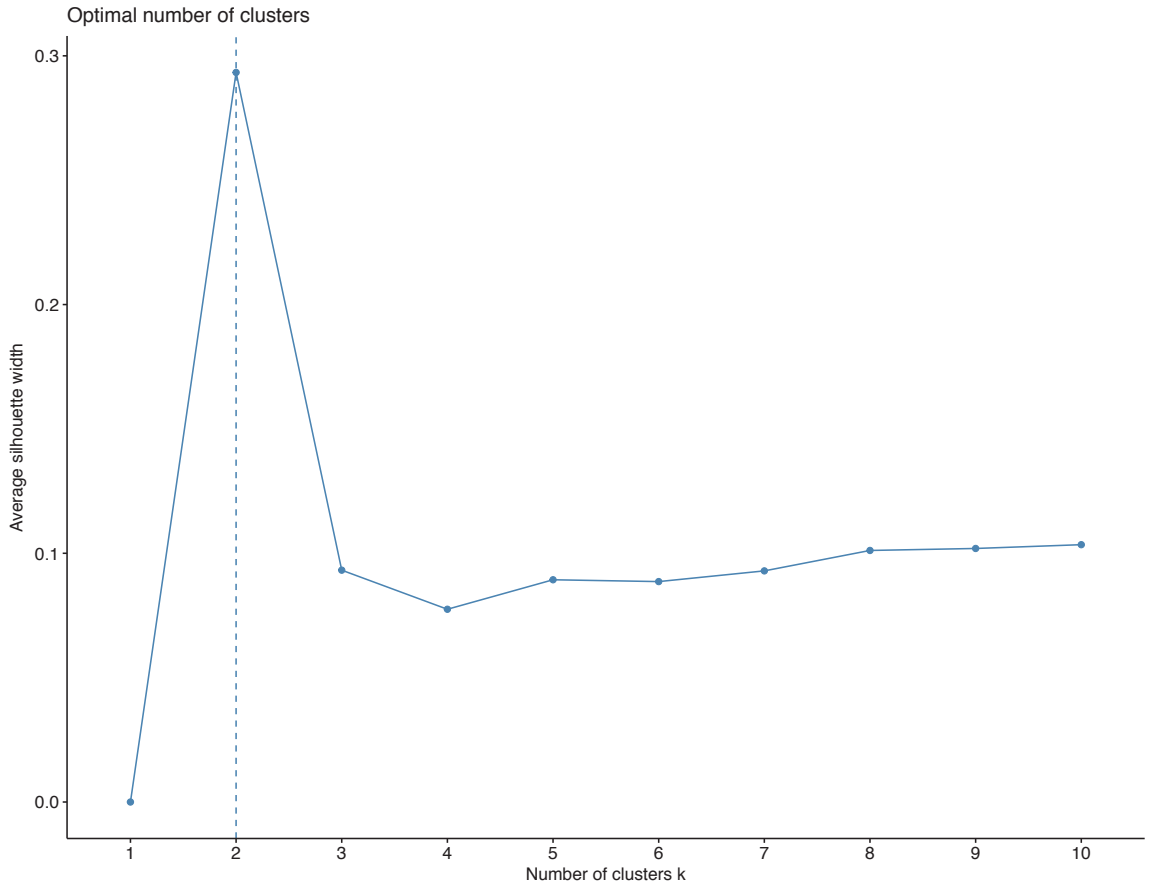


Fig. 12: Silhouette Method: Average Cluster Silhouette Width

Because the silhouette method suggests three appropriate cluster numbers—four, five, and six—a third method was used to check the optimal number of clusters. The Gap Statistic Method compares the total within-cluster variation for different cluster numbers to a uniform distribution (i.e. data with no discernable clustering). The

larger the gap statistic, the more the clustering structure differs from a uniform distribution. [51]

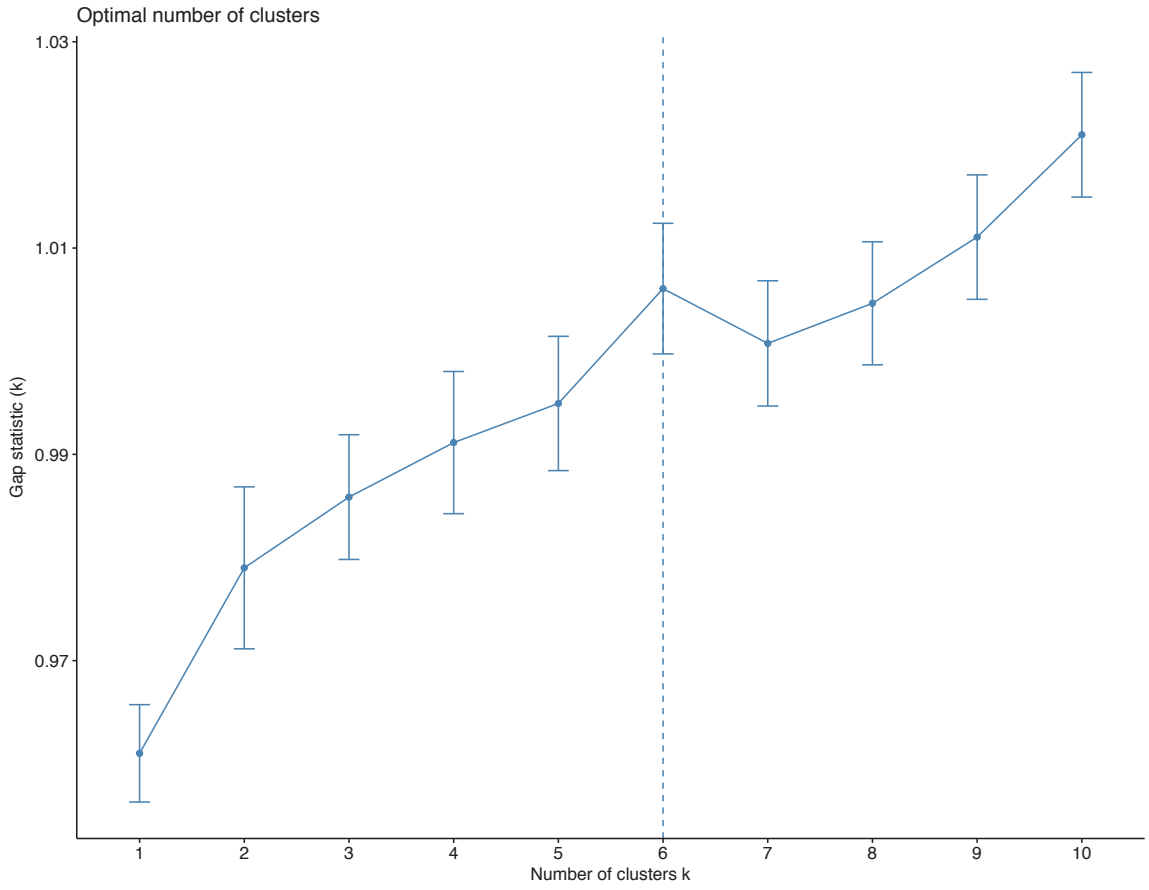


Fig. 12: Gap Statistic Method: Total Inter-Cluster Variation Compared to Normal Distribution

The graph shows an uptick in the gap statistic—or cluster structure—at six clusters.

Taken together, the three methods point to six as the optimal number of clusters; however, with several feasible options, the decision of how many clusters to

[51] Ibid.

should consider the number of clusteres most suitable for the Main Street organization (a realistic number of groups for which to develop programming).

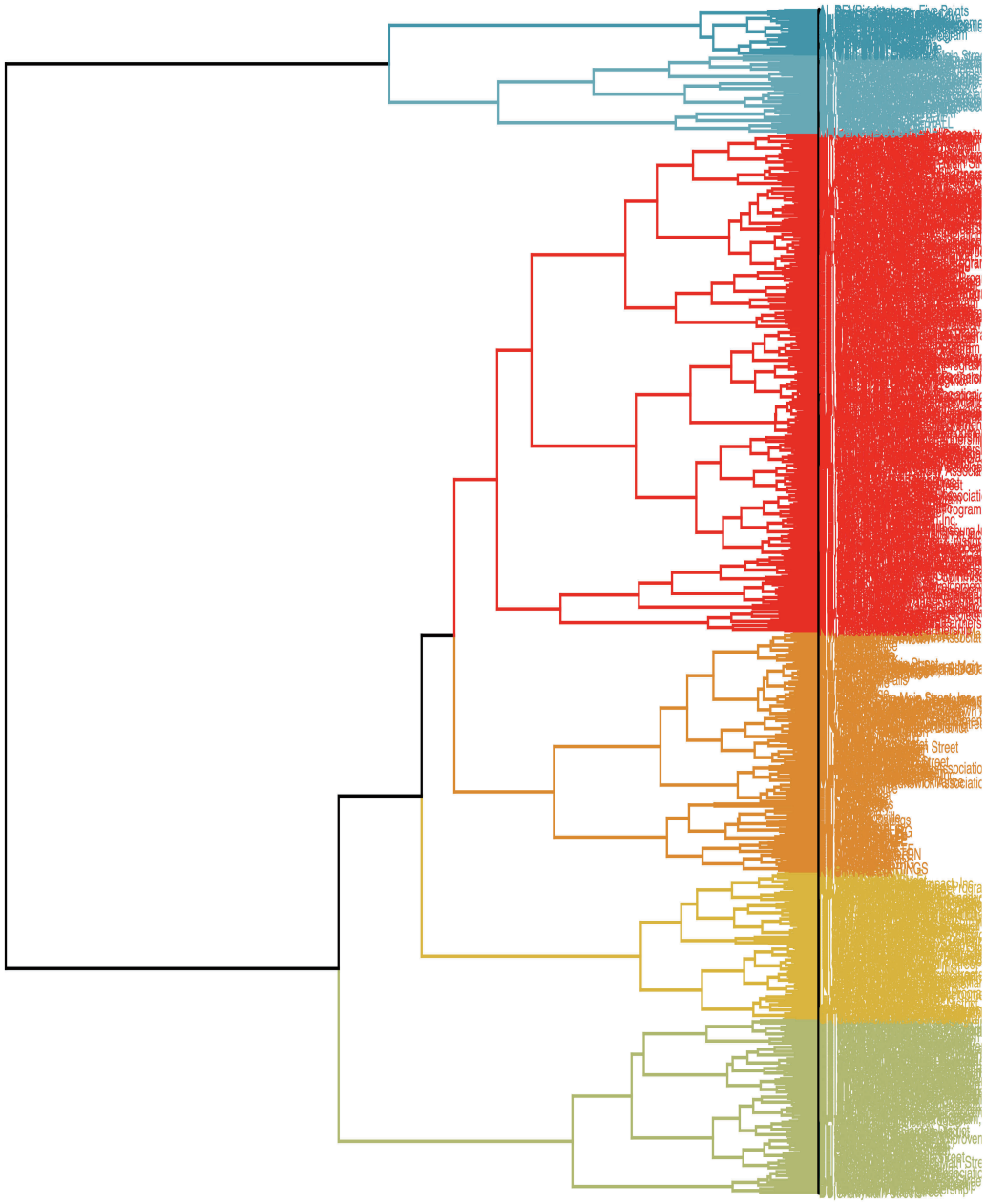


Fig. 14: Dendrogram showing six clusters

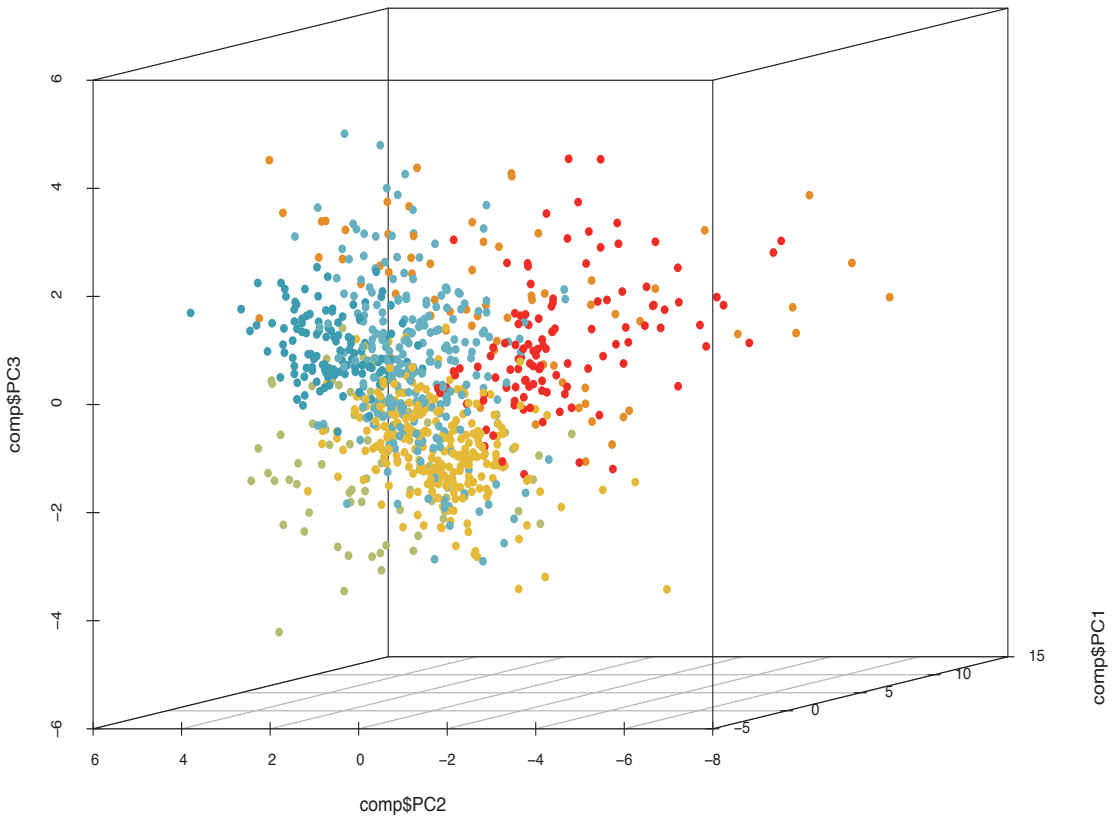


Fig. 15: Main Street clusters graphed along three axes (top three dimensions)

7. FINDINGS

While different methods can help to suggest an ideal number of clusters, this decision is just as much informed by the organization. If there is only organizational capacity for a simple typology with a few subgroups, that can be accommodated by

trimming the dendrogram at a higher point. Of course, this means that the neighborhoods within these larger groups will be more diverse, but there is still significant value-add with a segmentation. Alternatively, if the organization has capacity to develop tens of tracts for distinct groups, a larger number of clusters may be appropriate. A smaller number of clusters will have fewer members, but may parse similar neighborhoods into different groups when they would make more sense together. There are tradeoffs in each scenario.

In discussions with various consultants, state and national coordinators, and from my own time interning at the National Main Street Center, I believe the National Organization's time and resources are too limited to manage tens of clusters. The previous methods suggested six clusters were ideal, but 10 showed similar promise. I chose six clusters as I believe this better reflects the National Main Street Center's capacity. If the Main Street Center decides they would like to drill down further, they can easily ascertain where each split happens by looking at the dendrogram.

Looking at the six clusters, it is clear that they diverge in several key areas. Cluster 1 Main Streets have, on average, an older population, a more diverse population, higher rates of housing vacancy and unemployment, fewer college educated, and a higher rate of manufacturing workers. Cluster 2 has more children under 18 and more elderly over 65, more family households, lower rates of college educated, a greater percentage of Hispanic, and more agricultural workers. Cluster 3 has a significantly larger than average population of young adults 18 to 34 (very few kids or

elderly), very few family households, more renters than owners, and a higher number of college grads. Cluster 4 has fewer college educated, slightly higher rates of homeownership, and a greater number of manufacturing workers. Cluster 5 Main Streets are very dense, have a large population of young adults age 18 to 34 (few young or elderly), higher rates of college educated, more renters than homeowners, very high income and high home value. Lastly, Cluster 6 has more college educated, higher income, more older adults age 35 to 64, a higher rate of homeownership, and, on average, older housing stock.

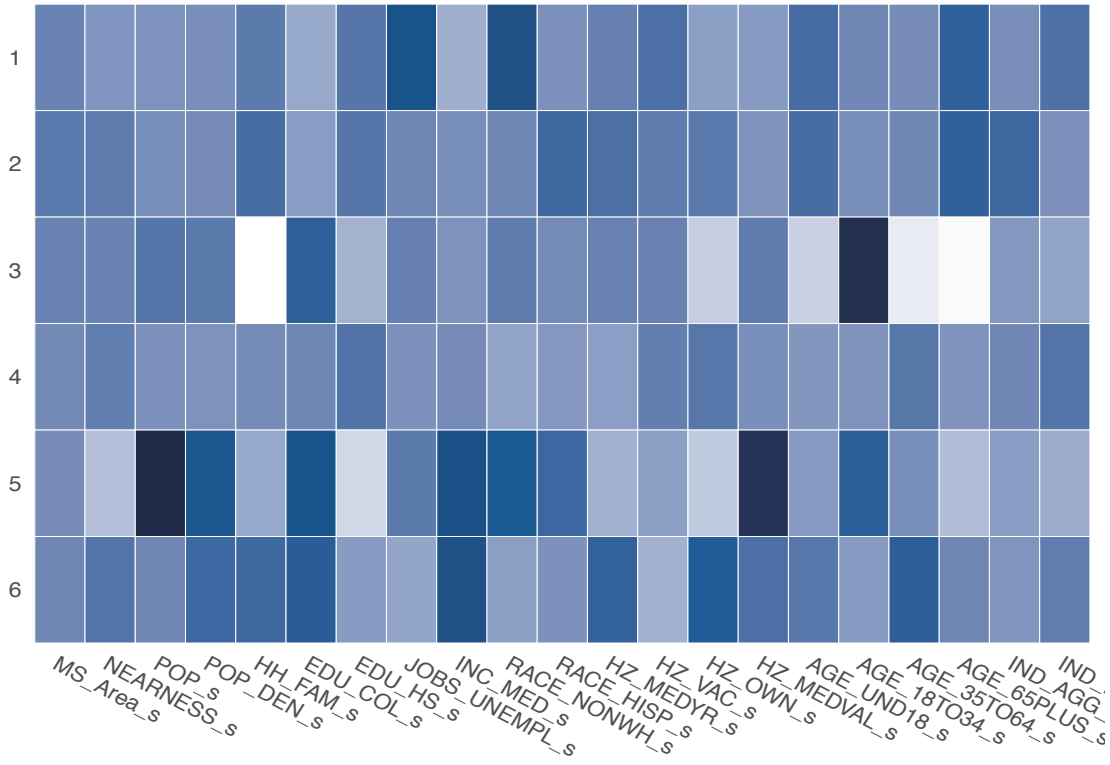


Fig. 16: Heat map showing Main Street clusters by different variables

With these six groups, the National Main Street Center is poised to develop more targeted programming and education designed for the specific attributes and demographics of each groups. This solution is not as time consuming or costly as individualized programming but more targeted than the current approach, and is poised to add significant value to The National Main Street Center's Four Point Approach. This typology will also be useful in helping local communities to identify comparable and/or aspirational communities—these clusters are a starting point for communities looking for guidance on addressing issues within a comparable context.

This analysis could also have implications for other national organizations with a broad scope and diverse membership. This study serves as an example of how collecting and analyzing publicly available data can lead to significant opportunity for programming and implementation specialization. This method can be replicated to optimize the allocation of certain government subsidies (such as real estate tax credits and grants) and strengthen the application of community development interventions (say, by LISC, HUD, NHS, etc.).

In addition, this thesis should make clear why robust and regular data-keeping practices are worthwhile. Ideally, with the initial lift of this project, upkeep will not be overly taxing to the organization and be worthy of any additional effort required. This thesis could serve as the starting point for additional statistical analyses to obtain a fuller understanding of what's working and what needs work within the Main Street Organization.

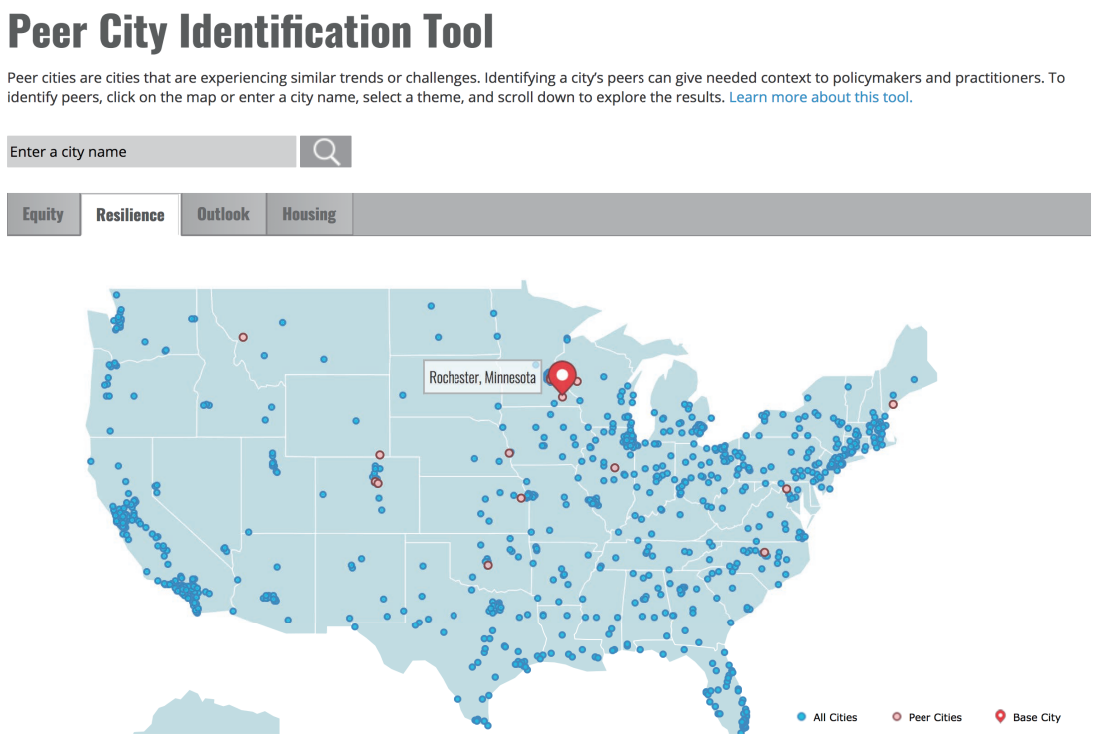
Large-scale data-driven projects, such as this one, can provide a big-picture view that helps to optimize resource allocation and streamline top-level decision making. These types of analyses may not be appropriate in guiding every-day, nuanced decision making as they fall short in capturing an individual community's on-the-ground sense of place, but they can add impactful insights to round out case studies, surveys, and other qualitative understandings.

8. NEXT STEPS

The Main Street Organization can use the code written for this project to gather additional variables as they see fit. With additional variables, Main Street could cluster communities based on specific topics (say, housing, resiliency, demographics), similar to what was done by The Federal Reserve Bank of Chicago for the Peer City Identification Tool. A front-end application or other user-friendly tool will make these findings more accessible for the National Organization, the State Coordinating Offices, local Main Streets, consultants, and researchers, who will hopefully be moved by open accessible data to undertake more advanced projects regarding Main Street.

As mentioned in this thesis, certain limitations restricted the scope of this project. Moving forward, this analysis could be expanded to include additional variables (perhaps even qualitative survey results) as practitioners and researchers see fit.

Hopefully, this project serves as a useful tool for Main Street and a robust example of the efficacy of large-scale data analysis for preservation and asset-based placemaking.



Resilience Peer Group

Download Data

Peer Cities	Unemployment rate	Labor force participation rate	Change in labor force participation rate	Labor share of manufacturing	Change, labor share of manufacturing, 1970-2016	Median family income	Change in median family income, 2000-2016
PCIT-960 Median	**7.7%**	**64.5%**	**-0.1%**	**9.5%**	**-58.2%**	**$60,862**	**-10.5%**
Peer Group Median	**5.4%**	**70.8%**	**0.8%**	**6.8%**	**-58.6%**	**$71,913**	**-5.3%**
Arvada, Colorado	5.3%	70.5%	-1.7%	9.2%	-52.5%	$87,422	-5.3%
Cheyenne, Wyoming	6.4%	68.1%	1.4%	3.3%	-49.7%	$73,475	7.6%
Denver, Colorado	5.4%	70.9%	3.3%	5.3%	-65.2%	$71,913	2.2%
Durham, North Carolina	6.5%	69.7%	1.9%	6.8%	-61.6%	$68,255	-8.6%
Eau Claire, Wisconsin	5.2%	71.2%	0.8%	10%	-56.3%	$66,070	-8.2%
Frederick, Maryland	5.6%	72.8%	-0.2%	6.2%	-62.1%	$76,298	-8%
Lawrence, Kansas	4.9%	69.1%	-1.6%	6.7%	-56.7%	$74,701	-0.7%
Minneapolis, Minnesota	6.9%	74.1%	2%	8.5%	-58.6%	$72,970	2.9%
Missoula, Montana	7.4%	70.8%	0.8%	3.4%	-64.5%	$67,229	9.4%
Moore, Oklahoma	4.4%	72.1%	1%	7.7%	-59.5%	$69,783	0.1%
Normal, Illinois	4.3%	68.4%	-1.8%	5.2%	-49.6%	$82,896	-6.4%
Omaha, Nebraska	5.4%	69.8%	0.4%	9.2%	-46.4%	$66,867	-9.9%
Portland, Maine	5.7%	70.8%	1.6%	6.6%	-60.3%	$69,337	-2.6%
Richfield, Minnesota	5.5%	71.5%	0.3%	8%	-59.7%	$71,650	-13%
Rochester, Minnesota	4.7%	71.2%	-1%	8.2%	-50%	$83,609	-5.7%

Fig. 17: Federal Reserve Bank of Chicago Peer City Identification Tool

Bibliography

Abello, Oscar Perry. "Business Improvement Districts Are More Than Just a Name on a Trash Can." NextCity. August 7, 2015. Accessed March 14, 2018. https://nextcity.org/daily/entry/business-improvement-districts-support-small-business.

American Community Survey Office. "ACS Summary File Technical Documentation." September 2016. Accessed January 18, 2018. https://www2.census.gov/programs-surveys/acs/summary_file/2015/documentation/tech_docs/2015_SummaryFile_Tech_Doc.pdf.

Anderson, M. R. "Cluster analysis for applications." *New York: Academic* (1973).

Armstrong, Amy, Ingrid Gould, Amy Ellen Schwartz, and Ioan Voicu. "The Benefits of Business Improvement Districts: Evidence from New York City." Furman Center for Real Estate and Urban Policy – NYU. Furmancenter.org. July 2007. Accessed March 2, 2018. http://furmancenter.org/files/publications/FurmanCenterBIDsBrief.pdf.

Beyard, Michael D., and W. Paul OMara. *Shopping Center Development Handbook*. Washington, DC: ULI-Urban Land Institute, 1999.

Campos, A., and R. C. Oliveira. "Cluster Analysis Applied to the Evaluation of Urban Landscape Quality." *WIT Transactions on Ecology and the Environment* 204 (2016): 93-103.

Caruso, Gina, and Rachel Weber. "Getting the Max for the Tax: An Examination of BID Performance Measures." *International Journal of Public Administration* 29, no. 1-3 (2006): 187-219. doi:10.1080/01900690500409088.

Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. "NbClust: AnRPackage for Determining the Relevant Number of Clusters in a Data Set." *Journal of Statistical Software* 61, no. 6 (2014). doi:10.18637/jss.v061.i06.

Desai, Manisha, and Melissa D. Begg. "A Comparison of Regression Approaches for Analyzing Clustered Data." *American Journal of Public Health* 98, no. 8 (2008): 1425-1429.

Dono, Andrew L., and Linda S. Glisson. *Revitalizing Main Street: A Practitioners Guide to Comprehensive Commercial District Revitalization*. Washington, DC: Main Street, National Trust for Historic Preservation/National Trust Main Street Center, 2009. Accessed September 10, 2017. http://www.mainstreet.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=05c4ce2a-39c7-0d15-a938-733989956055&forceDialog=0.

Federal Reserve Bank of Chicago. "Peer City Identification Tool." Accessed April 29, 2018. https://www.chicagofed.org/region/community-development/data/pcit.

Fogelson, Robert M. *Downtown Its Rise and Fall, 1880-1950*. New Haven: Yale University Press, 2008.

Francaviglia, Richard V. *Main Street Revisited: Time, Space, and Image Building in Small-Town America*. University of Iowa Press, 1996.

Gorham, Matthew. "Guide to the Bedford Stuyvesant Restoration Corporation Publication and Photograph Collection." NYU Digital Library Technology Services. January 20, 2017. Accessed April 28, 2018. http://dlib.nyu.edu/findingaids/html/bhs/arc_124_bed_stuy_restoration_corp/bioghist.html.

Hoffman, Alexander Von. "History Lessons for Today's Housing Policy." *History Lessons for Today's Housing Policy*. August 2012. Accessed April 28, 2018. http://webcache.googleusercontent.com/search?q=cache:http://www.jchs.harvard.edu/sites/jchs.harvard.edu/files/w12-5_von_hoffman.pdf.

Hoffman, Alexander Von. "The Past, Present, and Future of Community Development." Shelterforce.org. July 17, 2017. Accessed April 28, 2018. https://shelterforce.org/2013/07/17/the_past_present_and_future_of_community_development/.

Jacquez, Geoffrey M. "Spatial Cluster Analysis." *The Handbook of Geographic Information Science 395* (2008): 416. Accessed October 9, 2017. https://www.biomedware.com/files/jacquez_ch22_preprint.pdf.

"K-means Cluster Analysis." *K-means Cluster Analysis - UC Business Analytics R Programming Guide*. Accessed February 20, 2018. https://uc-r.github.io/kmeans_clustering#gap.

Kabacoff, Robert I. *R in Action: Data Analysis and Graphics with R.* Shelter Island, NY: Manning, 2015.

Kendig, Hal. "Cluster Analysis to Classify Residential Areas: A Los Angeles Application." *Journal of the American Institute of Planners* 42, *No. 3* (1976): 286-294.

Klemek, Christopher. “The Transatlantic Collapse of Urban Renewal: Postwar Urbanism from New York to Berlin.” Chicago, IL: University of Chicago Press, 2011.

LeSage, James P. "Regression Analysis of Spatial Data." *Journal of Regional Analysis and Policy* 27 (1997): 83-94. Accessed September 10, 2017. https://ageconsearch.umn.edu/bitstream/130445/2/27-2-7.pdf.

Mason, Randall. “Economics and Historic Preservation: A Guide and Review of the Literature.” Brookings Institute. Brookings.edu. September 2005. Accessed February 23, 2018. https://www.brookings.edu/wp-content/uploads/2016/06/20050926_preservation.pdf.

Mason, Steven, and Jonathan Schroeder, David Van Riper, and Steven Ruggles. *IPUMS National Historical Geographic Information System: Version 12.0* [Database]. Minneapolis: University of Minnesota. 2017. http://doi.org/10.18128/D050.V12.0.

Moral, Roger Del. "On Selecting Indirect Ordination Methods." *Classification and Ordination*. 1980. 75-84. http://www.math-stat.unibe.ch/unibe/portal/fak_naturwis/a_dept_math/a_dept_ms/content/e237483/e237655/e243381/e281679/files281690/Chap11_ger.pdf.

National Main Street Center, Kennedy Smith, and Josh Bloom. *The Main Street Approach: A Comprehensive Guide to Community Transformation.* Report. Accessed September 10, 1017. http://www.mainstreet.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=01cf95e3-5e71-ae73-902f-1b0e9494ceaa&forceDialog=0.

National Main Street Center. “Main Street Impact.” Accessed March 28, 2018. https://www.mainstreet.org/mainstreetimpact.

National Main Street Center. “Main Street Tier System Overview.” Accessed March 28, 2018. https://higherlogicdownload.s3.amazonaws.com/NMSC/390e0055-2395-4d3b-af60-81b53974430d/UploadedImages/Join/Main_Street_America_Tier_System_Overview.pdf

National Main Street Center. “The Main Street Movement.” Accessed March 28, 2018. https://www.mainstreet.org/themovement.

National Main Street Center. “The Programs.” Accessed March 28, 2018. https://www.mainstreet.org/theprograms.

National Main Street Center. “Urban Main.” Accessed March 28, 2018. https://www.mainstreet.org/themovement.https://higherlogicdownload.s3.amazonaws.com/NMSC/390e0055-2395-4d3b-af60-81b53974430d/UploadedImages/UrbanMain/NMSC30_FAQ_GENERAL_2.pdf

Ooi, Joseph T.l., Gaylon E. Greer and Phillip T. Kolbe. "Investment Analysis for Real Estate Decisions.” *Dearborn Real Estate Educatio*n. Journal of Property Investment & Finance 24, no. 3 (2006). doi:10.1108/jpif.2006.24.3.268.1.

Pedigo, Ashley, William Seaver, and Agricola Odoi, "Identifying Unique Neighborhood Characteristics to Guide Health Planning for Stroke and Heart Attack: Fuzzy Cluster and Discriminant Analyses Approaches," *PLoS ONE* 6, no. 7 (2011), doi:10.1371/journal.pone.0022693.

Pendola, Rocco, and Sheldon Gen. "Does ‘Main Street’ Promote Sense of Community? A Comparison of San Francisco Neighborhoods." *Environment and Behavior* 40, no. 4 (2008): 545-574. Accessed September 10, 2017. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.515.2731&rep=rep1&type=pdf.

Pryor, Susie, and Sanford Grossbart. "Ethnography of an American Main Street." *International Journal of Retail & Distribution* Management 33, no. 11 (2005): 806-823.

Rachid Erekaini. “What is a Community Development Corporation?” NACEDA, Sept. 17, 2014, Accessed March 2, 2018, https://www.naceda.org/index.php?option=com_dailyplanetblog&view=entry&category=bright-ideas&id=25%3Awhat-is-a-community-development-corporation-&Itemid=171

Reibel, Michael, and Moira Regelson, "Quantifying Neighborhood Racial and Ethnic Transition Clusters in Multiethnic Cities," *Urban Geography* 28, no. 4 (2007): 361-76, doi:10.2747/0272-3638.28.4.361.

Robertson, Kent A. "Can Small-City Downtowns Remain Viable?" *Journal of the American Planning Association* 65, no. 3 (1999): 270-83. doi:10.1080/01944369908976057.

Robertson, Kent A. "The Main Street Approach to Downtown Development: An Examination of the Four-Point Program." *Journal of Architectural and Planning Research* (2004): 55-73.

Ryan, Brent D. *Design after Decline: How America Rebuilds Shrinking Cities*. Philadelphia: University Of Pennsylvania, 2014.

Rypkema, Donovan. “Heritage Conservation and the Local Economy.” Report. August 2008. Accessed March 1, 2018, http://www.globalurban.org/GUDMag08Vol4Iss1/Rypkema%20PDF.pdf.

Rypkema, Donovan and Caroline Cheong, “Measuring Economic Impacts of Historic Preservation: A Report to the Advisory Council on Historic Preservation.” Report. August 2008. Accessed March 1, 2018. http://www.globalurban.org/GUDMag08Vol4Iss1/Rypkema%20PDF.pdf.

Rypkema, Donovan and Randall Mason. *Measuring Economic Impacts of Historic Preservation.* Report. November 2011. Accessed September 10, 2017. http://www.placeeconomics.com/wp-content/uploads/2016/08/Economic-Impacts-v5-FINAL.pdf.

Simon, Harold. “Season of Change,” Shelterforce.org, September 23, 2006, Accessed March 10, 2018. https://shelterforce.org/2006/09/23/season_of_change/.

Smith, Joshua and Mitsuru Saito. "Creating Land-Use Scenarios by Cluster Analysis for Regional Land-Use and Transportation Sketch Planning." *Journal of Transportation and Statistics 4*, No. 1 (2001): 39-49.

"Starting a BID." NYC Small Business Services. https://www1.nyc.gov/site/sbs/neighborhoods/starting-a-bid.page

Utah League of Cities and Towns. *Cluster Analysis of Utah’s Cities and Towns.* Report. April 1, 2007. Accessed October 8, 2017. http://www.ulct.org/ulct/wp-content/uploads/sites/4/2013/02/2007-ClusterAnalysis.pdf.

Utah League of Cities and Towns. *ULCT City & Town Cluster Analysis 2016.* Report. April 4, 2016. Accessed October 8, 2017. http://www.ulct.org/wp-content/uploads/sites/4/2016/09/Cluster-Analysis-2016.pdf.

Wears, Robert L. "Advanced Statistics: Statistical Methods for Analyzing Cluster and Cluster-Randomized Data." *Academic Emergency Medicine* 9, no. 4 (2002): 330-341.

APPENDIX: CODE

```
#****************************************#
#      UNIVERSITY OF PENNSYLVANIA
#     DEPT OF HISTORIC PRESERVATION
#       MASTERS THESIS - FALL 2018
#      MAIN STREET CLUSTER ANALYSIS
#
#        CODE BY: MOLLY BALZANO
#         mdbalzano@gmail.com
#
#         Completed: 04-29-2018
#        Last Update: 00-00-0000
#****************************************#

#******************************************************#
#     WORKFLOW
#
#  1.  Set up work space, install & call packages
#  2.  Upload Main Street shapefiles
#  3.  Inspect & clean Main Street shapefile attribute tables
#  3.  Calculate the nearest Main Street for each Main Street
#  4.  Upload census block group shapefiles
#  5.  Select all census block groups within 2-mile radius of a Main Street
#  6.  Get ACS data for all census blocks within Main Street buffers
#  7.  Calculate additional variables
#  8.  Calculate z-scores for all variables
#  9.  Cluster analysis
#  10. Create visuals
#  11. Save data
#
#******************************************************#

#******************************************************#
#     TIPS
#
#  1. Change from readOGR to st_read to open shapefiles faster
#  2. Change projection depending on location (currently using national UTM)
#
#******************************************************#
```

```
# # # # # # # # # # # # # # # # # # # # # # # # # # # #
#  (1) Set up work space, install & call packages   #
# # # # # # # # # # # # # # # # # # # # # # # # # # # #


# Clear workspace
rm(list=ls())

# Set working directory (this is the Mac file directory format)
setwd("")
wd <- getwd()

# Creae a list of packages to install
packages_list <- c("foreign", "rgdal", "sf",  "raster", "broom", "tigris",
                   "sp", "rgeos", "censusapi", "tidycensus", "tidyverse",
                   "viridis", "cluster", "cleangeo", "geosphere", "dplyr",
                   "spatstat","factoextra", "dendextend", "devtools",
                   "leaflet", "Hmisc", "plyr")

# Loop through list of packages. If packages aren't installed, install packages
a <- lapply(packages_list, function(x){if(! x %in% installed.packages())
install.packages(x)})

# Loop through list of packages, call libraries
a <- lapply(packages_list, function(x){library(x, character.only = TRUE)})

# If census API is not installed, install
if (! "censusapi" %in% installed.packages()) devtools::install_github("hrecht/censusapi")

# Call census API library
library("censusapi")  # connect to census API


# # # # # # # # # # # # # # # # # # # # # # # # # # # # #
#  (2) Upload Main Street shapefiles, inspect & clean   #
# # # # # # # # # # # # # # # # # # # # # # # # # # # # #
```

```
# Create a list of folders & file names
dirs <- list.files(path = "./Main_Streets", full.names = FALSE, recursive = FALSE)

# Create vectors to store each shapefile
shapes <- vector( "list", length(dirs))


#*****************************************************
# Read strings as characters, not factors
# stringsAsFactors = FALSE
# Check if there's an issue below forcing factors to strings
#*******


# Loop through Main Street folder, open each shapefile
for (b in 1:length(shapes)){
  shapes[[b]] <- try(readOGR(paste0("./Main_Streets/", dirs[b])), TRUE)
}

#set names
names(shapes) <- paste0("ms_", '', dirs)


# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
#  (3)  Inspect & clean Main Street shapefile attribute tables  #
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #


# Clean up shapefile attribute column names and add missing data
shapes[["ms_boston"]]$ST <- "MA"
shapes[["ms_boston"]]$MUNC <- "Boston"
shapes[["ms_boston"]]$SHAPE_Area <- shapes[["ms_boston"]]$Shape_Area
shapes[["ms_dc"]]@data[["MUNC"]][10] <- "Washington DC"
shapes[["ms_dc"]]@data[["MUNC"]][11] <- "Washington DC"
shapes[["ms_delaware"]]$ST <- "DE"
shapes[["ms_kentucky"]]$ST <- "KY"
shapes[["ms_kentucky"]]$MUNC <- shapes[["ms_kentucky"]]$NAME
shapes[["ms_kentucky"]]$SHAPE_Area <- shapes[["ms_kentucky"]]$Shape_Area
shapes[["ms_louisiana"]]$ST <- "LA"
shapes[["ms_louisiana"]]$MUNC <- shapes[["ms_louisiana"]]$COMMUNITIE
shapes[["ms_louisiana"]]$NAME <- shapes[["ms_louisiana"]]$COMMUNITIE
```

```
shapes[["ms_louisiana"]]$SHAPE_Area <- shapes[["ms_louisiana"]]$Shape_Area
shapes[["ms_michigan_oaklandcounty"]]$ST <- "MI"
shapes[["ms_michigan_oaklandcounty"]]$MUNC <-
shapes[["ms_michigan_oaklandcounty"]]$Community
shapes[["ms_michigan_oaklandcounty"]]$NAME <-
shapes[["ms_michigan_oaklandcounty"]]$Community
shapes[["ms_michigan_oaklandcounty"]]$SHAPE_Area <-
shapes[["ms_michigan_oaklandcounty"]]$Shape_Area
shapes[["ms_montana"]]$ST <- "MT"
shapes[["ms_montana"]]$MUNC <- shapes[["ms_montana"]]$NAME
shapes[["ms_montana"]]$SHAPE_Area <- shapes[["ms_montana"]]$Shape_Area
shapes[["ms_new_jersey"]]$ST <- "NJ"
shapes[["ms_new_jersey"]]$MUNC <- shapes[["ms_new_jersey"]]$MUNI
shapes[["ms_new_jersey"]]$SHAPE_Area <- shapes[["ms_new_jersey"]]$Shape_Area
shapes[["ms_north_carolina"]]$NAME <- shapes[["ms_north_carolina"]]$PROGRAM
shapes[["ms_oklahoma"]]$ST <- "OK"
shapes[["ms_oklahoma"]]$MUNC <- shapes[["ms_oklahoma"]]$Community
shapes[["ms_oklahoma"]]$NAME <- shapes[["ms_oklahoma"]]$MSName
shapes[["ms_oklahoma"]]$SHAPE_Area <- shapes[["ms_oklahoma"]]$Shape_Area
shapes[["ms_oregon"]]$NAME <- shapes[["ms_oregon"]]$Community
shapes[["ms_oregon"]]$SHAPE_Area <- shapes[["ms_oregon"]]$Shape_Area
shapes[["ms_orlando"]]$NAME <- shapes[["ms_orlando"]]$BUISNESSNE
shapes[["ms_orlando"]]$SHAPE_Area <- shapes[["ms_orlando"]]$Shape_area

# Select only the shared columns in each attribute table
for (c in 1:length(shapes)){
  print(c)
  shapes[[c]]@data <- data.frame("ST" = shapes[[c]]$ST, "MUNC" = shapes[[c]]$MUNC,
"NAME" = shapes[[c]]$NAME, "SHAPE_Area" = shapes[[c]]$SHAPE_Area)
}

# Loop through the list of shapefiles, change projection, and bind together into one
shapefile
for (c in 1:length(shapes)){
  if (!exists("ms_states")){
    ms_states <- spTransform(shapes[[c]], CRS("+proj=utm +north +zone=4
+ellps=WGS84"))  # Read shapefile and transform to correct projection
  }
  else{
    ms_states <- rbind(ms_states,spTransform(shapes[[c]], CRS("+proj=utm +north
+zone=4 +ellps=WGS84")))  # if the merged dataset does exist, merge to it
```

```
  }

}

# Check projection
proj4string(ms_states)

# Check for NAs
check_ms_states <- ms_states@data[rowSums(is.na(ms_states@data)) >0 ,]
View(check_ms_states)

# Fix any outstanding errors (if this doesn't work, exlcude)
ms_states@data[["MUNC"]][514] <- "Metuchen"


# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
#  (3) Calculate the nearest Main Street for each Main Street   #
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #


# Create a new column to store distance to the nearest Main Street
ms_states@data$Nearness <- ''

# For each Main Street, calculate distance to nearest Main Street
for(d in 1:nrow(ms_states)){
  current_state <- ms_states[d,]
  e <- gDistance(current_state, ms_states, byid=TRUE)
  print(min(e[e>0]))
  poly_index <- which.min(e[e>0])
  print(poly_index)
  print(ms_states[poly_index,]$NAME)
  name <- ms_states[poly_index,]@data$NAME
  current_state@data$Nearness<- name
  attribute_data = data.frame(current_state)
  ms_states[d,]<- attribute_data
}

# Check data
head(ms_states@data)
```

```
############################
#  (4) Upload Census block group shapefiles   #
############################


# Make sure there is single shapefile in every folder with valid attributes
blockshpfolder_list <- list.files(path = "./Block_Groups", full.names = FALSE, recursive =
FALSE)

# Open block group shapefiles and combine into one shapefile
# Takes ~2 minutes per state
for (file in blockshpfolder_list){
  # if the merged dataset doesn't exist, create it
  shp_folder <- paste(getwd(),'Block_Groups',file,sep="/")
  shp_name <- gsub('.shp','',list.files(path= shp_folder,pattern = "\\.shp$"))

  if (!exists("blocks_group")){
    blocks_group <- spTransform(readOGR(dsn=shp_folder, layer=shp_name),
CRS("+proj=utm +north +zone=4 +ellps=WGS84"))
  }

  # if the merged dataset does exist, merge to it
  else{
    blocks_group <- bind(blocks_group,spTransform(readOGR(dsn=shp_folder,
layer=shp_name), CRS("+proj=utm +north +zone=4 +ellps=WGS84")))
  }

}

# Check projection
proj4string(blocks_group)


##########################################################
#  (5) Select all census block groups within a 2 mile radius of a Main Street   #
##########################################################


# Select all block groups with a 2-mile buffer of each Main Street
# Takes ~4 hours to run
for(g in 1:nrow(ms_states)){
```

```
  current_ms <- ms_states[g,]
  print(current_ms$NAME)

  # Draw a 2-mile buffer around each polygon
  state_buffer <- gBuffer(current_ms, width = 3218)

  # Select block groups whose centroid falls within the buffer
  blocks_within <- gIntersects(state_buffer, blocks_group, byid =TRUE)

  # Create a vector with the blockgroups selected
  blocks_selected <- blocks_group[as.vector(blocks_within),]
  print(length(blocks_selected))

  # For each blockgroup in the vector, add the Main Street's data
  if(length(blocks_selected) > 0){
    blocks_selected$NAME <- current_ms$NAME
    blocks_selected$ST <- current_ms$ST
    blocks_selected$MUNC <- current_ms$MUNC
    blocks_selected$NEARNESS <- current_ms$Nearness
    blocks_selected$SHAPE_Area <- current_ms$SHAPE_Area
    blocks_selected$MS_Area <- gArea(current_ms)/1609.344
    blocks_selected$Block_Area <- gArea(blocks_selected)/1609.344
  }

  # Create a new vector with the combined data
  if(!exists("blocks_selected_all"))
  {blocks_selected_all = blocks_selected}
  else
  {blocks_selected_all = bind(blocks_selected_all,blocks_selected)}
}

# Create a data frame from the blocks vector
blocks_sel_frame <- data.frame (blocks_selected_all)

# Save file
write.csv(blocks_sel_frame, "blocks_sel_frame.csv", row.names = FALSE)


# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
#  (6) Get ACS data for all census blocks within Main Street buffers  #
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
```

```
# Connect to census API
# Get API here: http://api.census.gov/data/key_signup.html
# Census_api_key
census_api_key("828af1537ea1388037a9e60cf6bf7688f4f8b361")

# Cache data for use in future sessions
options(tigris_use_cache = TRUE)

# Download all available census variables to find the names of the desired variables
availablevars <- listCensusMetadata(name="acs5", vintage=2015)

# Test the variables come through for one county
test_census <- get_acs(geography = "block group",
                       variables =c(POP = "B01003_001E",
                                    HH_TOT = "B09019_002E",
                                    HH_FAM = "B09019_003E",
                                    HH_NONFAM = "B09019_024E",
                                    EDU_TOT = "B15003_001E",
                                    EDU_HSDIP = "B15003_017E",
                                    EDU_HSGED = "B15003_018E",
                                    EDU_COL1YR = "B15003_019E",
                                    EDU_SOMECOL = "B15003_020E",
                                    EDU_ASSC = "B15003_021E",
                                    EDU_BACH = "B15003_022E",
                                    EDU_MAST = "B15003_023E",
                                    EDU_PROF = "B15003_024E",
                                    EDU_DOC = "B15003_025E",
                                    JOBS_TOT = "B23025_002E",
                                    JOBS_UNEMPL = "B23025_005E",
                                    INC_MED = "B19013_001E",
                                    INC_AGG = "B19025_001E",
                                    HZ_TOT = "B25001_001E",
                                    HZ_OCU = "B25002_002E",
                                    HZ_VAC = "B25002_003E",
                                    HZ_OWN = "B25003_002E",
                                    HZ_RENT = "B25003_003E",
                                    HZ_MEDYR = "B25035_001E",
                                    RACE_WH = "B02001_002E",
                                    RACE_HISP = "B03002_012E",
```

```
HZ_MDRENT = "B25064_001E",
HZ_MEDVAL = "B25077_001E",
AGEM_TOT = "B01001_002E",
AGEM_LESS5 = "B01001_003E",
AGEM_5TO9 = "B01001_004E",
AGEM_10TO14 = "B01001_005E",
AGEM_15TO17 = "B01001_006E",
AGEM_18TO19 = "B01001_007E",
AGEM_20 = "B01001_008E",
AGEM_21 = "B01001_009E",
AGEM_22TO24 = "B01001_010E",
AGEM_25TO29 = "B01001_011E",
AGEM_30TO34 = "B01001_012E",
AGEM_35TO39 = "B01001_013E",
AGEM_40TO44 = "B01001_014E",
AGEM_45TO49 = "B01001_015E",
AGEM_50TO54 = "B01001_016E",
AGEM_55TO59 = "B01001_017E",
AGEM_60TO61 = "B01001_018E",
AGEM_62TO64 = "B01001_019E",
AGEF_TOT = "B01001_026E",
AGEF_LESS5 = "B01001_027E",
AGEF_5TO9 = "B01001_028E",
AGEF_10TO14 = "B01001_029E",
AGEF_15TO17 = "B01001_030E",
AGEF_18TO19 = "B01001_031E",
AGEF_20 = "B01001_032E",
AGEF_21 = "B01001_033E",
AGEF_22TO24 = "B01001_034E",
AGEF_25TO29 = "B01001_035E",
AGEF_30TO34 = "B01001_036E",
AGEF_35TO39 = "B01001_037E",
AGEF_40TO44 = "B01001_038E",
AGEF_45TO49 = "B01001_039E",
AGEF_50TO54 = "B01001_040E",
AGEF_55TO59 = "B01001_041E",
AGEF_60TO61 = "B01001_042E",
AGEF_62TO64 = "B01001_043E",
INDM_TOT = "C24030_001E",
INDM_AGG = "C24030_003E",
INDM_CONST = "C24030_006E",
```

```
                    INDM_MANU = "C24030_007E",
                    INDM_WHOL = "C24030_008E",
                    INDM_RET = "C24030_009E",
                    INDM_TRANS = "C24030_010E",
                    INDM_INFO = "C24030_010E",
                    INDM_FIN = "C24030_011E",
                    INDM_PROF = "C24030_017E",
                    INDM_EDU = "C24030_021E",
                    INDM_ARTS = "C24030_024E",
                    INDM_OTH = "C24030_027E",
                    INDM_PUB = "C24030_028E",
                    INDF_AGG = "C24030_030E",
                    INDF_CONST = "C24030_033E",
                    INDF_MANU = "C24030_034E",
                    INDF_WHOL = "C24030_035E",
                    INDF_RET = "C24030_036E",
                    INDF_TRANS = "C24030_037E",
                    INDF_INFO = "C24030_040E",
                    INDF_FIN = "C24030_041E",
                    INDF_PROF = "C24030_044E",
                    INDF_EDU = "C24030_048E",
                    INDF_ARTS = "C24030_051E",
                    INDF_OTH = "C24030_054E",
                    INDF_PUB = "C24030_055E"),
            key=census_api_key, year = 2015,
            county = "Maricopa",
            state = "AZ",
            geometry = TRUE, output = "wide")
View(test_census)

# If there are incorrect vars (return NA); find correct ones
test_census2 <- get_acs(geography = "block group",
            variables =c(VAR1 = "",
                    VAR2 = "",
                    VAR3 = "",
                    VAR4 = "",
                    VAR5 = ""),
            key=census_api_key, year = 2015,
            county = "Maricopa",
            state = "AZ",
            geometry = TRUE, output = "wide")
```

```
View(test_census2)

# Create data frame to store the census data
county_state_list <- data.frame(unique(blocks_sel_frame[,c('COUNTYFP','STATEFP')]))
dim(county_state_list)

# Select desired variables for all blockgroups in the US
# For each county in the count_state_list, select the following variables by block group
for (h in 1:nrow(county_state_list)){
  print(h)
  census_data <- get_acs(geography = "block group",
                         variables =c(POP = "B01003_001E",
                                      HH_TOT = "B09019_002E",
                                      HH_FAM = "B09019_003E",
                                      HH_NONFAM = "B09019_024E",
                                      EDU_TOT = "B15003_001E",
                                      EDU_HSDIP = "B15003_017E",
                                      EDU_HSGED = "B15003_018E",
                                      EDU_COL1YR = "B15003_019E",
                                      EDU_SOMECOL = "B15003_020E",
                                      EDU_ASSC = "B15003_021E",
                                      EDU_BACH = "B15003_022E",
                                      EDU_MAST = "B15003_023E",
                                      EDU_PROF = "B15003_024E",
                                      EDU_DOC = "B15003_025E",
                                      JOBS_TOT = "B23025_002E",
                                      JOBS_UNEMPL = "B23025_005E",
                                      INC_MED = "B19013_001E",
                                      INC_AGG = "B19025_001E",
                                      RACE_WH = "B02001_002E",
                                      RACE_HISP = "B03002_012E",
                                      HZ_TOT = "B25001_001E",
                                      HZ_OCU = "B25002_002E",
                                      HZ_VAC = "B25002_003E",
                                      HZ_OWN = "B25003_002E",
                                      HZ_RENT = "B25003_003E",
                                      HZ_MEDYR = "B25035_001E",
                                      HZ_MDRENT = "B25064_001E",
                                      HZ_MEDVAL = "B25077_001E",
                                      AGEM_TOT = "B01001_002E",
                                      AGEM_LESS5 = "B01001_003E",
```

```
AGEM_5TO9 = "B01001_004E",
AGEM_10TO14 = "B01001_005E",
AGEM_15TO17 = "B01001_006E",
AGEM_18TO19 = "B01001_007E",
AGEM_20 = "B01001_008E",
AGEM_21 = "B01001_009E",
AGEM_22TO24 = "B01001_010E",
AGEM_25TO29 = "B01001_011E",
AGEM_30TO34 = "B01001_012E",
AGEM_35TO39 = "B01001_013E",
AGEM_40TO44 = "B01001_014E",
AGEM_45TO49 = "B01001_015E",
AGEM_50TO54 = "B01001_016E",
AGEM_55TO59 = "B01001_017E",
AGEM_60TO61 = "B01001_018E",
AGEM_62TO64 = "B01001_019E",
AGEF_TOT = "B01001_026E",
AGEF_LESS5 = "B01001_027E",
AGEF_5TO9 = "B01001_028E",
AGEF_10TO14 = "B01001_029E",
AGEF_15TO17 = "B01001_030E",
AGEF_18TO19 = "B01001_031E",
AGEF_20 = "B01001_032E",
AGEF_21 = "B01001_033E",
AGEF_22TO24 = "B01001_034E",
AGEF_25TO29 = "B01001_035E",
AGEF_30TO34 = "B01001_036E",
AGEF_35TO39 = "B01001_037E",
AGEF_40TO44 = "B01001_038E",
AGEF_45TO49 = "B01001_039E",
AGEF_50TO54 = "B01001_040E",
AGEF_55TO59 = "B01001_041E",
AGEF_60TO61 = "B01001_042E",
AGEF_62TO64 = "B01001_043E",
INDM_TOT = "C24030_001E",
INDM_AGG = "C24030_003E",
INDM_CONST = "C24030_006E",
INDM_MANU = "C24030_007E",
INDM_WHOL = "C24030_008E",
INDM_RET = "C24030_009E",
INDM_TRANS = "C24030_010E",
```

```
                    INDM_INFO = "C24030_010E",
                    INDM_FIN = "C24030_011E",
                    INDM_PROF = "C24030_017E",
                    INDM_EDU = "C24030_021E",
                    INDM_ARTS = "C24030_024E",
                    INDM_OTH = "C24030_027E",
                    INDM_PUB = "C24030_028E",
                    INDF_AGG = "C24030_030E",
                    INDF_CONST = "C24030_033E",
                    INDF_MANU = "C24030_034E",
                    INDF_WHOL = "C24030_035E",
                    INDF_RET = "C24030_036E",
                    INDF_TRANS = "C24030_037E",
                    INDF_INFO = "C24030_040E",
                    INDF_FIN = "C24030_041E",
                    INDF_PROF = "C24030_044E",
                    INDF_EDU = "C24030_048E",
                    INDF_ARTS = "C24030_051E",
                    INDF_OTH = "C24030_054E",
                    INDF_PUB = "C24030_055E"),
                key=census_api_key,year = 2015,
                county = as.character(county_state_list[h,]$COUNTYFP),
                state = as.character(county_state_list[h,]$STATEFP),
                geometry = TRUE,output = "wide")

  # Combine all data into a single data frame
  if(!exists("census_data_all"))
  {census_data_all <- data.frame(census_data)}
  else
  {census_data_all <- bind(census_data_all,data.frame(census_data))}
}

# Drop the margin of error columns
census_data_clean <- census_data_all[, -grep("_[[:digit:]]{3}M",
colnames(census_data_all))]

# Drop duplicate columns
census_data_clean2 <- census_data_clean[,-c(51,52,76,77,97,98)]

# Replace NAs with column mean
for(i in 1:ncol(census_data_clean2)){
```

```
  census_data_clean2[is.na(census_data_clean2[,i]), i] <- mean(census_data_clean2[,i],
na.rm = TRUE)
}

# Create a data frame with the census block groups that match those in the Main Street
block group data frame
census_data_req <- data.frame(census_data_clean2[(census_data_clean2$GEOID %in%
blocks_sel_frame$GEOID),])

# Merge the Main Street data frame and the census block group data frame using GEOID
combined_data <- merge(census_data_req,blocks_sel_frame,by="GEOID")

# Check for NAs
apply(combined_data, 2, function(x) any(is.na(x)))

# Save file
write.csv(combined_data, "combined_data.csv", row.names = FALSE)

# # # # # # # # # # # # # # # # # # # # #
#  (7) Calculate additional variables   #
# # # # # # # # # # # # # # # # # # # # #


# For each Main Street, sum the block group data to get totals
# Where appropriate, calculate percentages
# (Missing vars: age, population of city; county seat; nearness to geographic feature)
aggregate_ms <- combined_data %>%
group_by(ST,MUNC,NAME.y,SHAPE_Area,MS_Area,NEARNESS) %>% dplyr::summarize(
  BLOCK_AREA_f = sum(Block_Area),
  POP_f = sum(POP),
  POP_DEN_f = sum(POP)/sum(Block_Area),
  HH_FAM_f = sum(HH_FAM)/sum(HH_TOT),
  HH_NONFAM_f = sum(HH_NONFAM)/sum(HH_TOT),
  EDU_COL_f =
(sum(EDU_ASSC)+sum(EDU_BACH)+sum(EDU_MAST)+sum(EDU_PROF)+sum(EDU_DOC)
)/sum(EDU_TOT),
  EDU_HS_f =
(sum(EDU_HSGED)+sum(EDU_HSDIP)+sum(EDU_COL1YR)+sum(EDU_SOMECOL))/sum(E
DU_TOT),
  EDU_LESSHS_f = (sum(EDU_TOT)-EDU_COL_f-EDU_HS_f)/sum(EDU_TOT),
  JOBS_UNEMPL_f = sum(JOBS_UNEMPL)/sum(JOBS_TOT),
```

```
  JOBS_EMPL_f = (sum(JOBS_TOT)-sum(JOBS_UNEMPL))/sum(JOBS_TOT),
  INC_MED_f = mean(INC_MED),
  INC_MEAN_f = sum(INC_AGG)/sum(POP),
  RACE_WH_f = sum(RACE_WH)/sum(POP),
  RACE_NONWH_f = (sum(POP)-sum(RACE_WH))/sum(POP),
  RACE_HISP_f = sum(RACE_HISP)/sum(POP),
  HZ_MEDYR_f = mean(HZ_MEDYR),
  HZ_OCU_f = sum(HZ_OCU)/sum(HZ_TOT),
  HZ_VAC_f = sum(HZ_VAC)/sum(HZ_TOT),
  HZ_OWN_f = sum(HZ_OWN)/sum(HZ_OCU),
  HZ_RENT_f = sum(HZ_RENT)/sum(HZ_OCU),
  HZ_MDRENT_f = mean(HZ_MDRENT),
  HZ_MEDVAL_f = mean(HZ_MEDVAL),
  AGE_UND18_f =
(sum(AGEM_LESS5)+sum(AGEM_5TO9)+sum(AGEM_10TO14)+sum(AGEM_15TO17)+su
m(AGEF_LESS5)+sum(AGEF_5TO9)+sum(AGEF_10TO14)+sum(AGEF_15TO17))/(sum(AG
EM_TOT)+sum(AGEF_TOT)),
  AGE_18TO34_f =
(sum(AGEM_18TO19)+sum(AGEM_20)+sum(AGEM_21)+sum(AGEM_22TO24)+sum(AGE
M_25TO29)+sum(AGEM_30TO34)+sum(AGEF_18TO19)+sum(AGEF_20)+sum(AGEF_21)
+sum(AGEF_22TO24)+sum(AGEF_25TO29)+sum(AGEF_30TO34))/(sum(AGEM_TOT)+su
m(AGEF_TOT)),
  AGE_35TO64_f =
(sum(AGEM_35TO39)+sum(AGEM_40TO44)+sum(AGEM_45TO49)+sum(AGEM_50TO54)
+sum(AGEM_55TO59)+sum(AGEM_60TO61)+sum(AGEM_62TO64)+sum(AGEF_35TO39)
+sum(AGEF_40TO44)+sum(AGEF_45TO49)+sum(AGEF_50TO54)+sum(AGEF_55TO59)+s
um(AGEF_60TO61)+sum(AGEF_62TO64))/(sum(AGEM_TOT)+sum(AGEF_TOT)),
  AGE_65PLUS_f = (sum(AGEM_TOT)+sum(AGEF_TOT)-AGE_UND18_f-AGE_18TO34_f-
AGE_35TO64_f)/(sum(AGEM_TOT)+sum(AGEF_TOT)),
  IND_TOT_f = sum(INDM_TOT),
  IND_AGG_f = (sum(INDM_AGG)+sum(INDF_AGG))/IND_TOT_f,
  IND_CONST_f = (sum(INDM_CONST)+sum(INDF_CONST))/IND_TOT_f,
  IND_MANU_f = (sum(INDM_MANU)+sum(INDF_MANU))/IND_TOT_f,
  IND_WHOL_f = (sum(INDM_WHOL)+sum(INDF_WHOL))/IND_TOT_f,
  IND_RET_f = (sum(INDM_RET)+sum(INDF_RET))/IND_TOT_f,
  IND_TRANS_f = (sum(INDM_TRANS)+sum(INDF_TRANS))/IND_TOT_f,
  IND_FIN_f = (sum(INDM_FIN)+sum(INDF_FIN))/IND_TOT_f,
  IND_PROF_f = (sum(INDM_PROF)+sum(INDF_PROF))/IND_TOT_f,
  IND_EDU_f = (sum(INDM_EDU)+sum(INDF_EDU))/IND_TOT_f,
  IND_ARTS_f = (sum(INDM_ARTS)+sum(INDF_ARTS))/IND_TOT_f,
  IND_OTH_f = (sum(INDM_OTH)+sum(INDF_OTH))/IND_TOT_f,
```

```
  IND_PUB_f = (sum(INDM_PUB)+sum(INDF_PUB))/IND_TOT_f,
  IND_INFO_f = (IND_TOT_f-IND_AGG_f-IND_CONST_f-IND_MANU_f-IND_WHOL_f-
IND_RET_f-IND_TRANS_f-IND_FIN_f-IND_PROF_f-IND_EDU_f-IND_ARTS_f-IND_OTH_f-
IND_PUB_f)/IND_TOT_f
)

# Inspect data frame
dim(aggregate_ms)
head(aggregate_ms)
tail(aggregate_ms)
summary(aggregate_ms)
str(aggregate_ms)

# Fix errors: change data types
aggregate_ms$NEARNESS <- as.numeric(as.character(aggregate_ms$NEARNESS))

# Check for NAs
check_aggregate_ms <- aggregate_ms[rowSums(is.na(aggregate_ms)) >0 ,]
dim(check_aggregate_ms)


# # # # # # # # # # # # # # # # # # # # # # # # #
#  (8) Calculate z-scores for all variables   #
# # # # # # # # # # # # # # # # # # # # # # # # #


# Reset RStudio plot panel (if errors)
graphics.off()
par("mar")
par(mar=c(1,1,1,1))
options(scipen=5)

# Check for outliers
source("http://goo.gl/UUyEzD")
outlierKD(aggregate_ms, MS_Area)
outlierKD(aggregate_ms, HZ_MEDVAL_f)

# Remove outliers
agg_clust <- agg_clust[-c(473),]

# Subset variables for cluster
```

```
agg_clust <- agg_clust[, -c(4,7,11,14,16,18,19,23,26,32,33,44)]

# Name rows
agg_clust$ID <- paste(agg_clust$ST, agg_clust$MUNC, agg_clust$NAME.y, sep=",")
clust_names <- agg_clust[['ID']]
row.names(agg_clust) <- clust_names

# Check variance
mar <- par()$mar
par(mar=mar+c(0,5,0,0))
barplot(sapply(agg_clust, var), horiz=T, las=1, cex.names=1)
barplot(sapply(agg_clust[,-c(1,2,3,35)], var), horiz=T, las=1, cex.names=1, log='x', main =
"Variance: Before Standardization")
par(mar=mar)
plot(sapply(agg_clust[,-c(1,2,3,35)], var))

# Calculate z-scores for all columns to scale
zscores <- agg_clust[,-c(35)]
zscores[,4:34] <- data.frame(scale(zscores[,4:34]))
barplot(sapply(zscores[,4:34], var), horiz=T, las=1, cex.names=1, main = "Variance: After
Standardization")
plot(sapply(zscores[,4:34], var))


# # # # # # # # # # # # # # #
#  (9) Cluster Analysis   #
# # # # # # # # # # # # # # #


# Remove extra columns
zscores <- zscores[,-c(24,26:34)]

# Graph correlation matrix
#install.packages("corrplot")
library(corrplot)
ms_corr <- cor(zscores[,4:24])
corrplot(ms_corr, method="circle", order="hclust", tl.col="black", tl.srt=45, tl.cex=0.8)

# Get principal component vectors
pc <- princomp(zscores[,4:24])
plot(pc)
```

```
plot(pc, type='l')
summary(pc)

# First for principal components
pc <- prcomp(zscores[,4:24])
comp <- data.frame(pc$x[,1:9])

# Choose PCA or remove highly correlated variables
zscores <- zscores[, -c(18)]

# First, determine best linkage method
# Linkage methods to assess
m <- c( "average", "single", "complete", "ward")
names(m) <- c( "average", "single", "complete", "ward")

# Function to compute coefficient
ac <- function(x) {
  agnes(zscores[,4:23], metric = "euclidian", method = x)$ac
}

# Compute coefficient (closer to 1 suggests strong clustering structure)
map_dbl(m, ac)

# Check divisive clustering
ms_hc2 <- diana(zscores[,4:23], metric = "euclidian")
# Divise coefficient; amount of clustering structure found
ms_hc2$dc

# Determine optimal number of clusters (Elbow Method)
fviz_nbclust(zscores[,4:23], FUN = hcut, method = "wss")

# Determine optimal number of clusters (Silhouette Method)
fviz_nbclust(zscores[,4:23], FUN = hcut, method = "silhouette")

# Determine optimal number of clusters (Gap Statistic Method)
set.seed(123)
gap_stat <- clusGap(zscores[,4:23], FUN = hcut, nstart = 25,
                    K.max = 10, B = 50)
fviz_gap_stat(gap_stat)

# Determine optimal number of clusters (30 Indices)
```

```
library("NbClust")
nb <- NbClust(zscores[,4:23], min.nc = 3,
          max.nc = 10, method = "ward.D2")

# Graph optimal number of clusters
fviz_nbclust(nb)


# # # # # # # # # # # # #
#  (10) Create Visuals  #
# # # # # # # # # # # # #


# Name rows
zscores$ID <- paste(zscores$ST, zscores$MUNC, zscores$NAME.y, sep=",")
zscores_names <- zscores[['ID']]
row.names(zscores) <- zscores_names

# Install color palette
#install.packages("wesanderson")
library(wesanderson)
colors <- wes_palette(n=6, name="Zissou1", type = "continuous")

# Require packages
require(magrittr)
require(ggplot2)
require(dendextend)

dend <- zscores[,4:23] %>% dist(method="euclidian") %>%
  hclust(method="ward.D2") %>% as.dendrogram %>%
  set("branches_k_color", k = 6, value = colors) %>% set("branches_lwd", 0.7) %>%
  set("leaves_pch", 19) %>% set("leaves_cex", 0.5) %>%
  set("labels", row.names(zscores)) %>% set("labels_cex", 0.5) %>%
  set("labels_colors", k = 6, value = colors)
ggd1 <- as.ggdend(dend)
ggplot(ggd1, horiz = TRUE) +
  geom_point()

# Cut dendrogram at 6 clusters
clusMember6 <- cutree(dend, k = 6, order_clusters_as_data = TRUE)
```

```
# 3D Plot (looks better)
library(scatterplot3d)
comp$clusMember6 <- clusMember6
colors <- colors[as.numeric(comp$clusMember6)]
scatter1 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 15)
scatter2 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 30)
scatter3 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 45)
scatter4 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 60)
scatter5 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 75)
scatter6 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 92)
scatter7 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 105)
scatter8 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 120)
scatter9 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 135)
scatter10 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 150)
scatter11 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 165)
scatter12 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 185)
scatter13 <- scatterplot3d(comp$PC1, comp$PC2, comp$PC3, pch = 16, color=colors,
angle = 195)

# Add cluster number to Main Street data frame
final_ms <- agg_clust
final_ms$cluster6 <- clusMember6

# Add cluster number to zscore data frame
zscoresf <- zscores
zscoresf$cluster6 <- clusMember6

# Group by cluster
ms_heatmap <- zscoresf %>% group_by(cluster6) %>% dplyr::summarize(
```

```
  MS_Area_s = mean(MS_Area),
  NEARNESS_s = mean(NEARNESS),
  POP_s = mean(POP_f),
  POP_DEN_s = mean(POP_DEN_f),
  HH_FAM_s = mean(HH_FAM_f),
  EDU_COL_s = mean(EDU_COL_f),
  EDU_HS_s = mean(EDU_HS_f),
  JOBS_UNEMPL_s = mean(JOBS_UNEMPL_f),
  INC_MED_s = mean(INC_MED_f),
  RACE_NONWH_s = mean(RACE_NONWH_f),
  RACE_HISP_s = mean(RACE_HISP_f),
  HZ_MEDYR_s = mean(HZ_MEDYR_f),
  HZ_VAC_s = mean(HZ_VAC_f),
  HZ_OWN_s = mean(HZ_OWN_f),
  HZ_MEDVAL_s = mean(HZ_MEDVAL_f),
  AGE_UND18_s = mean(AGE_UND18_f),
  AGE_18TO34_s = mean(AGE_18TO34_f),
  AGE_35TO64_s = mean(AGE_35TO64_f),
  AGE_65PLUS_s = AGE_UND18_s - AGE_18TO34_s - AGE_35TO64_s,
  IND_AGG_s = mean(IND_AGG_f),
  IND_MANU_s = mean(IND_MANU_f)
)

# Install libraries
library("reshape2")
library("scales")

# Make a heatmap for each cluster
msorder <- list(6,5,4,3,2,1)
ms_heatmap$Order <- as.numeric(msorder)
ms_heatmap$Name <- with(ms_heatmap, reorder(cluster6, Order))
ms_heatmap <- ms_heatmap[,-c(1,23)]

ms_heatmap.m <- melt(ms_heatmap)
ms_heatmap.s <- ddply(ms_heatmap.m, .(variable), transform)

ggplot(ms_heatmap.s, aes(variable, Name)) +
  geom_tile(aes(fill = value), colour = "white") +
  scale_fill_gradient2(low = "white", mid = muted("steelblue"),
                       high = "black", midpoint = 1, space = "Lab",
                       na.value = "grey50", guide = "colourbar") +
```

```
  scale_x_discrete("", expand = c(0, 0)) +
  scale_y_discrete("", expand = c(0, 0)) +
  theme_grey(base_size = 20) +
  theme(legend.position = "none",
        axis.ticks = element_blank(),
        axis.text.x = element_text(angle = 330, hjust = 0))


# # # # # # # # # # #
#  (10) Save data   #
# # # # # # # # # # #


# Save data
write.csv(final_ms, "ms_clusters.csv", row.names = FALSE)

# Save data
write.csv(zscoresf, "zscore_clusters.csv", row.names = FALSE)
```