2012

Commuting patterns and labor markets: a new regional classification for Louisiana

Deepa Acharya
Louisiana State University and Agricultural and Mechanical College, deepa.acharya18@gmail.com

Commuting Patterns and Labor Markets: A New Regional Classification for Louisiana

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in

The Department of Agricultural Economics and Agribusiness

by
Deepa Acharya
B.S., Tribhuvan University, 2008
August 2012

Acknowledgements

I want to thank my dad and mom, Divakar and Kalawati Sharma for being my support system and the driving force behind all success in my life. I also thank my brothers Dhruba and Prakash for their love and support throughout my study period, and every step in my life. I thank my sister Daya for always believing in me and being there for me. I am grateful to my brother-in-law Prakash Subedi for his help and encouragement. I could not have done this without my husband Manish Poudel. His love and faith in me always kept me going.

I want to express gratitude to my major advisor Dr. Ashok K. Mishra for providing me an opportunity to study at LSU, and the help and guidance during my study period. I want to thank my advisory committee members Dr. Matt Fannin and Dr. Krishna P. Paudel for their suggestions and guidance. I am also thankful to the Dr. Gail Cramer for his help and encouragment.

I really appreciate the technical support from Som Bohra and Dr. Michael Kellar. I am thankful to my friends Sandamali and Michelle for their support. Finally, I am grateful to LSU, my friends in and outside the department, and Nepali community without whom my stay in Baton Rouge would not have been so pleasant and full of wonderful memories.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BEA	Bureau of Economic Analysis
BOEMRE	Bureau of Ocean Energy Management, Regulation and Enforcement
CCC	Cubic Clustering Criterion
CBSA	Core Based Statistical Area
CEA	Component Economic Areas
COMPAS	Community Policy Analysis System
CSA	Combined Statistical Area
CZ	Commuting Zone
EA	Economic Area
ERS	Economic Research Service
ESS	Error Sum of Squares
FEA	Functional Economic Area
FIPS	Federal Information Processing Code
GDP	Gross Domestic Product
GIS	Geographic Information Systems
GOM	Gulf of Mexico
IMPLAN	IMpact analysis for PLANning
IRS	Internal Revenue Service
LED	Local Employment Dynamics
LEHD	Longitudinal-Employer Household Dynamics
LMA	Labor Market Area
LSU	Louisiana State University

MSA	Metropolitan Statistical Area
NIST	National Institute of Standards and Technology
OCAMA	Oklahoma City
OCS	Outer Continental Shelf
OMB	Office of Management and Budget
QCEW	Quarterly Census of Employment and Workforce
RAC	Residential Area Characteristics
RMSSTD	Root Mean Square Standard Deviation
SAS	Statistical Analysis System
SMA	Standard Metropolitan Statistical Area
UK	United Kingdom
UTRR	Undiscovered Technically recoverable Resources
WAC	Working Area Characteristics

Abstract

Regional classification and labor market study form an important part of any regional development efforts. Successful formation and implementation of developmental policies for a region requires a sound knowledge of the labor market situation and socioeconomic background of the region, which in turn leads ultimately to regional welfare. We find literature in the area of regional classification to be very inadequate. This study classifies Louisiana using a clustering approach in two different ways. First of all, Ward's method has been used to classify Louisiana into labor markets based on two-way commuting flow between the parishes. Eight geographical clusters are formed and compared with eight Metropolitan Statistical Areas in Louisiana. Secondly, a regional classification for Louisiana is delineated based on four socioeconomic variables using K-means clustering method. Based on goodness-of-fit criteria, nine regional clusters have been formed.

1. Introduction

1.1 Overview

Louisiana is located in the southeastern portion of the U.S. and borders the Gulf of Mexico (GOM)[1]. According to the 2010 U.S. Census it has a population of 45,333,72 and land area of 43,204 square miles. The presence of various oil and gas industries, as well as one of the busiest ports in the U.S. (New Orleans), make Louisiana an important part of U.S. economy. According to the 2002 U.S. Census, the mining sector (which includes oil and gas extraction) in Louisiana was comprised of 1,503 establishments with 46,871 employees working in them. Furthermore, Louisiana is also famous for various marine resources, including navigation, recreation, and commercial fishing. The value of aquaculture products from Louisiana was $264,063,740 in 2011. Similarly, the gross farm value of agricultural produce in Louisiana was $3,824,167,187 in 2011 (LSU AgCenter, 2011). Louisiana parishes are home to several large-scale plantations, including cotton plantations.

These industries require a large workforce to fulfill their labor demands. To help determine where that workforce is coming from, a labor market study of this state is important. The advances in infrastructure and technology in the U.S. have considerably reduced the impact of geographical location and, therefore, labor is highly mobile. In such a context, then, it is

[1] Five states namely Texas, Louisiana, Alabama, Mississippi and Florida border the Gulf of Mexico in the United States (Fig. I). GOM is a major tropical sea of North America. It pays a vital role in the national economy of the U.S., as offshore operations in the GOM are the major source of U.S. domestic natural gas and oil. According to the 2011 Report of the Bureau of Ocean Energy Management Regulation and Enforcement (BOEMRE), there are 3,302 active offshore production platforms for the production of natural gas and oil in the GOM. Reports from the Outer Continental Shelf (OCS) oil and gas assessment done by BOEMRE in 2006 estimated the quantity of undiscovered technically recoverable resources (UTRR) in the Outer Continental Shelf of the GOM to range from 66.6 to 115.3 billion barrels of oil and 326.4 to 565.9 trillion cubic feet of natural gas (BOEMRE 2010).

necessary to identify labor markets beyond the traditional definition of counties or parishes. Hence, study of labor market is important.

The aim of this study is to form a regional classification system based on commuting[2] patterns. Further, we classify Louisiana into regions based on socioeconomic variables to gain a deeper understanding of the state of Louisiana. This is expected to promote regional growth and welfare by assisting researchers and policy-makers as they seek to develop better roadways, encourage development, and provide additional support for underserved areas.

The U.S. Census Bureau has divided the country broadly into four regions, namely, Northeast, Midwest, South and West for representing decennial census data. Regions are classified into nine divisions, two in each region except the South, which has three divisions. Divisions are sub-divided into states. States are further classified into counties and counties into county equivalents[3]. There are 3,143 counties and county equivalents. County divisions are composed of smaller geographical units called places (or parts). Places are composed of further smaller divisions called census tracts (or parts), which usually have 1,500 to 8,000 people residents. Census tracts further branch out into sub-units called block groups (or parts), which, according to the U.S. Census Bureau, ideally contains 600-3000 people. Finally, the smallest sub-division is the census block (U.S. Census Bureau, 2011). According to U.S. Census Bureau (2010), there are 1,148 census tracts, 3,471 block groups, and 204,447 census blocks in Louisiana. All states, counties, tracts, block groups and census blocks are represented by their

---

[2] Johnson (2006) has defined commuters as the workers who are identified as residents of a different location than that of their jobs when data are recorded.

[3] Louisiana is sub-divided into parishes and Alaska into boroughs, which are county-equivalent geographical units of these States.

standard Federal Information Processing System (FIPS)[4] codes. For example, a state is represented by a two-digit code followed by a three-digit code for a county. Further, the Census Bureau has also classified the county in several other classifications (urban/rural areas, micropolitan/metropolitan statistical areas, non-metro areas and so on).

Federal statistical agencies use various geographical entities for collection, tabulation, and publication of federal statistics. The Office of Management and Budget (OMB) has defined metropolitan statistical areas (metro) as core urban areas with populations of 50,000 or more. Similarly, micropolitan areas contain an urban core of at least 10,000, but a population of less than 50,000. Each metro or micro area is composed of one or more counties and includes the counties containing the core urban area, as well as any surrounding counties that have a high degree of socio-economic integration with the urban core, as measured by commuting to work (U.S. Census Bureau, 2010).

Combined statistical areas (CSAs) are bigger areas containing both metropolitan statistical areas and micropolitan statistical areas. CSAs consist of two or more adjacent core-based statistical areas (CBSAs) in which there is at least a 15 percent employment interchange between cores, as measured by commuting. When this exchange is 25 percent or higher between a pair of CBSAs, they are combined into a CSA. On the other hand, if the measure is between 15 and 25 percent, the decision for a combination is reached by a local opinion in both areas. CBSAs are smaller geographical entities than micro areas, which have a minimum population of 10,000. At least 25 percent of people living in the outlying areas of the CBSAs commute to the core.

[4]"FIPS codes are a standardized set of numeric or alphabetic codes issued by the National Institute of Standards and Technology (NIST) to ensure uniform identification of geographic entities through all federal government agencies. The entities covered include: states and statistically equivalent entities, counties and statistically equivalent entities, named populated and related location entities (such as, places and county subdivisions), and American Indian and Alaska Native areas." (U.S. Census Bureau, 2010)

Currently there are eight metropolitan statistical areas in Louisiana. They are:

1. New Orleans-Metairie-Kenner Metropolitan Statistical Area

   It has a population of 4,544,228 and contains seven parishes namely, Jefferson, Orleans, Plaquemines, St. Bernard, St. Charles, St. John the Baptist and St. Tammany.

2. Baton Rouge Metropolitan Statistical Area

   It has a population of 802,484 and contains nine parishes namely, Ascension, East Baton Rouge, East Feliciana, Iberville, Livingston, Pointe Coupee, St. Helena, West Baton Rouge and West Feliciana.

3. Shreveport-Bossier City Metropolitan Statistical Area

   It has a population of 398,604 and contains three parishes namely, Bossier, Caddo and De Soto.

4. Lafayette Metropolitan Statistical Area

   It has a population of 273,738 and contains two parishes namely, Lafayette and St. Martin.

5. Houma-Bayou Cane-Thibodaux Metropolitan Statistical Area

   It has a population of 208,178 and contains two parishes namely, Lafourche and Terrebonne.

6. Lake Charles Metropolitan Statistical Area

   It has a population of 199,607 and contains two parishes namely, Calcasieu and Cameron.

7. Monroe Metropolitan Statistical Area

It has a population of 176,441 and contains two parishes namely, Ouachita and Union.

8. Alexandria Metropolitan Statistical Area

It has a population of 153,922 and contains two parishes namely, Grant and Rapides.

1.2 Rationale of the Study

Federal and state governments use counties as the basic geographic units for data collection, tabulation, and dissemination. According to the U.S. Census Bureau, the major legally defined political and administrative units of the U.S. are the states and counties[5]. Therefore, they form the primary geographic units for reporting data (U.S. Census Bureau, 2005)[6]. The vast amount of socioeconomic data available at the county level makes counties ideal as the unit of analysis for this study. However, using the county divisions for defining labor markets has several limitations. One major drawback is that county boundaries are politically defined, and were not created to define a labor market. Hence, there is a need for a broader regional classification, which is not limited to county contiguity and extends beyond county boundaries.

The results of this study can be used in policy design regarding local labor and employment for Louisiana. We see this precedent in other, similar research. The findings of a Greek study in 2005-2007, for instance, calculated labor market areas (LMAs) in Greece on the basis of commuting flows and were later used in empirical analysis with the goal of formulating economic development and social cohesion policy proposals (Prodromidis 2008). In addition,

[5] The concept of counties as administrative unit is traced back to England and was brought to the colonies by early settlers.

[6] The cities of Philadelphia and San Francisco are spread over the entire counties, so they have a single government for the city and the county.

here in the United States, the 2000 U.S. Census asked for journey-to-work and place-of-work information primarily for planning highway improvements and developing public transportation services. Police and fire departments continue to use this data to plan smooth emergency operations in areas of high employment concentration (US Census Bureau 2004).

In addition, the regional classification we propose may influence key decisions in the economic development of Louisiana. The results of our work may aid policymakers in developing plans to improve the public transportation system. Furthermore, the labor market developed from commuting should aid real estate business decisions as developers analyze optimal housing locations. Hence, labor market classification is one of the initial and most important steps in reaching the goal of optimal human resources in the Louisiana market economy.

In the long run, regional classification leads to regional welfare through the effective policy recommendation and subsequent implementation with the help of information provided about the local labor market. The establishment of any new industry in an area starts with a detailed study of the labor market conditions for that area. The labor market uses information from localized studies and bases its subsequent decisions on their analysis. The feasibility study for an industry establishment takes into account the availability of local labor and its appropriateness to the specific requirements of the employer. If local labor cannot fulfill the demand, questions such as who will migrate/commute to the area to fill the gap need to be answered before development continues. Information about the labor market can impact employment promotion and subsequent socioeconomic development of the region. The commuting aspect of the labor market is a vital piece of the puzzle and has implications for resource-use, employment, and migration.

1.3 Objectives

The major objective of this study is to identify and describe an alternative regional classification of Louisiana based on commuting patterns and socioeconomic data. The classification based on labor-commuting data will be used as a measure of local labor markets in Louisiana. In addition, we use socioeconomic variables to cluster Louisiana parishes into relatively homogenous groups with similar socioeconomic characteristics. These classifications will help with recognition of larger regions beyond the traditional geographical classification formed by counties, cities, and metropolitan statistical areas. This study will inform labor market issues such as unemployment so that policymakers can better planning and manage the state's human resources. Similarly, the results will help policymakers implement effective practices that could reduce regional socioeconomic disparity.

Specifically, we aim to accomplish several objectives:

- Classify Louisiana into two regional classifications: functional and homogenous
- Cluster several parishes into common regions based on commuting and socioeconomic behavior
- Map clusters for ease of description and analysis
- Compare and contrast clusters obtained from commuting data with metro areas of Louisiana in terms of geography
- Describe the socioeconomic characteristics of the clusters.

2. Literature Review

The literature review is divided into four sections. The first part explores the different types of classifications and their importance as described by various regional economists. The second part describes the use of labor commuting as a basis for regional labor market classification. Next, we give an overview of the use of socioeconomic variables in regional classification. Finally, we discuss in greater detail the use of socio-economic variables for regional classification.

2.1 Types and the Importance of Regional Classification

Regional economists have defined regions and classified them in various ways. Hoover and Giarratani (1999) define a region as a geographical area that is considered an entity for purposes of description, analysis, administration, planning or public policy. Description allows information to be handled and presented more conveniently. When there is more interdependence of units or activities, analysis of information is very useful. Hoover and Giarratani (1999) describe two major types of regions: homogenous, and functional. In addition, they describe nodal regions and administrative regions as derivatives of the first two types of regions.

Homogeneous regions are differentiated on the basis of the internal uniformity of a place and show similarity of the place. Any change will affect the whole region in a similar way. An example of a homogenous region is the winter wheat belt in the central part of the U.S. It is a homogenous agricultural region because all parts use the same method to grow the same crop, wheat. A region where the line of demarcation is defined by its economic interdependence is called a functional region. Areas of a functional region are characterized by greater interaction with one another as compared to outside areas. There is usually a vast amount of transference of goods and services within a functional region.

A nodal region is a type of functional region, which is distinguished by a single main nucleus (the principal city of the region), subordinate centers, and the remaining rural territory. The nodal region is differentiated from other functional regions because it considers the role of each entity in the interaction pattern. Administrative regions are constructed for management purposes; in their formation, both homogeneity and functional interaction are considered.

Brown and Holmes (1971) recognized locational entities, which are homogenous in some aspects as regions. They classified regions into two broad groups: formal, or uniform, regions and functional regions. Formal/uniform regions are constrained by contiguity and the descriptive variables are attributes of the areas being grouped. In contrast, the areas in a functional region are functionally complementary to each other. They are comprised of locational entities with greater interaction or connection to each other than with outside areas. In the case of functional regions, variables describing the region are interactions between the areas being grouped.

A nodal region is a special case of a functional region, having a single focal point in which the notion of dominance or order is introduced. In other words, when the focal point of the interdependence is a single characteristic such as labor markets, it is considered a nodal region. A nodal region is different from a functional region, as it does not have symmetrical relationship, especially if a single interaction is considered. When within-group interaction is stronger than between-group interaction, without considering the role of each entity in the interaction pattern, it is called a functional region. On the other hand, nodal regions have groupings based on both interactions between and rank within locational entities, with a single locational entity dominating all others.

Ilbery (1981) gives both divisive and agglomerative methods of classification. The divisive method is deductive; the universal set is sub-divided in a series of steps. Agglomerative classification is inductive; it attempts to allocate individuals to groups, facilitating the examination of regularities and significant interrelationships. Ilbery also points out that much of the previous literature on delimitation of uniform regions in geographical classifications has frequently adopted a divisive approach. Agglomerative methods of classification can be distinguished from divisive methods in five main ways:

i) Enumeration, not definition, specifies the universal set.

ii) Derivation of theoretical classes is not possible.

iii) Assumption about the order of interrelationships among the variables used to differentiate the classes is not made.

iv) Agglomerative methods are usually used as a sampling framework for further scientific enquiry.

v) They are more realistic, although there is an inherent difficulty of assigning elements to the correct classes.

A sound understanding of the geography and concept of a region is one of the first steps in formulating and implementing any economic policy. Anderson (1975) provides three explanations for the importance of classifying regions:

i) Classification simultaneously facilitates the presentation and understanding of the specific features of a multivariate distribution.

ii) It allows the compilation of statistics in such a way that we can easily see the significant patterns by making associations and differences more readily discernible.

iii) It can stimulate further research and ultimately lead to the development of theory.

In the same way, Blien et al. (2006) give three major areas where the typology of regions could be informative with respect to labor markets:

i) To study the effects of special policy measures;

ii) To get a sound knowledge of the spatial structure of the economy; and,

iii) To provide insight for other research studies.

2.2 Use of Commuting Data in Classifications

The importance of commuting data in labor markets can be found in various pieces of literature. The Economic Research Service (ERS, 2010) highlights the importance of using commuting zones (CZs) and labor market areas (LMAs) for regional classifications, reasoning that local county boundaries do not contain the whole economy and labor market of the area. It should be defined as measured by the interrelationships between buyers and sellers of labor in that region. ERS used county and county-equivalent level commuting data to define 741 commuting zones in 1990. After taking minimum population requirements into account, they further grouped these into 394 labor markets. In 2000, they updated the zones using the same methodology, delineating 709 commuting zones.

We can find several examples in literature regarding policy implications of labor studies on commuting, as it affects employment and migration. Holmes (1972) concluded that strong linkages exist between out-commuting and out-migration based on his study in mid-eastern Pennsylvania. In his earlier work, he examined commuting as an alternative to out-migration in certain Australian situations. Renkow et al. (1997) presented different scenarios where commuting and migration are substitutes and complements. They concluded that commuting and migration would be substitutes for those households that found local wages lower than distant wages, and who had to either commute or migrate to maximize their income. Similarly, commuting and migration were complements in two situations: when a particular household changed its place of residence without changing its place of work due to preferred residential amenities in the new place and when people choose a new workplace and new residence simultaneously.

Poole (1964) has concluded that study of data of commuting patterns is one of the most important things required by a regional investigator for regional analysis regarding an urban economic base study, measuring economic growth or structural change. In his case study of characteristics and commuting patterns of the work force of Oklahoma's largest single employer, the Oklahoma City Air Material Area (OCAMA), he recognized that failure to account for commuting patterns can result in exaggerated and erroneous base employment figures with subsequent inaccuracies in the base employment-to-service employment ratio, base employment-to-total employment ratio, and base employment-to-total population ratio. Poole points out that data generated on commuting indicates the geographic reach of the local labor market. The author further emphasizes the importance to the concerned economic development organization of commuting pattern studies in providing more accurate information to prospective industries as well as established industries contemplating expansion regarding labor market characteristics, capabilities, and limitations. Additionally, spatial delimitation of the local labor market is required for the estimation of local labor supply for prospective industries.

Models to estimate the employment changes using commuting data have been attempted by several economists. One such example is a model developed by Davis et al. (2004) to estimate employment growth using commuting data for Minnesota. Employment growth was calculated in terms of number of jobs in the county, as the sum of changes in labor force size plus number of in-commuters, minus the number of unemployed people and the number of out-commuters. The study concluded that in-commuting and increased labor force were the key determinants of the labor market adjustments.

Johnson et al. (2006) use the Community Policy Analysis System (COMPAS) model to study the impact of different industries in a region and labor markets. COMPAS model is based on inter-sectorial linkages, because change in any industry or sector can cause consequent changes in other sectors, as all are linked in the economy. Johnson (2006) gives equations to quantify labor supply in a labor market using commuting as one of the variables: labor demand equals labor supply in an economy in equilibrium. According to the author, labor demand is a function of wage, while labor supply is a sum of four components: resident labor force, in-commuters and the unemployed, minus out-commuters. Johnson (2006) asserts that the major factor that determines the impact of employment changes in the local economy is commuting.

Other examples of labor market classification studies based on commuting can be found in the context of European counties. Kristiansen (1998) provides functional economic classification of Danish municipalities using journey-to-work files. Similarly, Prodromidis (2008) classifies Greek municipalities on the basis of commuting data and provides the labor markets by focusing on the largest commuting nodes[7]. The European studies use commuting data for classification in a similar manner to the American studies, although there are slight variations in the methodology used for classification.

2.3 Use of Socioeconomic Variables in Regional Classification

Literature shows that socio-economic variables have been used to form groupings to study regions. Celik et al. (2011) did a study using socioeconomic variables to find the similarities and dissimilarities of 14 cities in the East Anatolia region in Turkey. The study was

[7] These studies have been further discussed in the review about methodologies section of the literature review.

an effort to determine the comprehensive activities that would help to accelerate the development of the region. The authors highlight the need to balance the geographical, functional and social inequalities to achieve economic development. They shed light on the fact that East and South Anatolia differ primarily due to an imbalance of economic resources, income distribution and equality of opportunity. They argue that using only Gross Domestic Product (GDP) to estimate the true socio-economic development of a region is insufficient. Instead, using various socio-economic variables in addition to GDP gives a better indication of where a region stands in ranking with its counterparts.

Celik et al. (2011) classify East Anatolia by using variables under nine broad topics: an overall welfare indicator, along with demographic, educational, financial, industrial, health, agricultural, infrastructure, and constructional indicators. They find some variables influence the difference between the provinces more than other variables. In particular, employment indicators and industrial indicators cause a significant statistical difference between the clusters of provinces. The authors conclude that the major difference in development between the clusters of cities is due to employment, industrial, financial and other welfare indicators. They suggest that regional disparity between the provinces should be taken into account when planning any future projects for development in the East Anatolia region. Furthermore, an effort should be initiated to give higher priority to investments in and promotions of below-average regions (in terms of development indicators compared to their counterparts).

Rovan and Sambt (2003) use socio-economic variables to cluster Slovenian municipalities. They emphasize that, for most of these, national welfare is best served by keeping them at a sustainable level. In addition, they point out the relevancy of classification by highlighting that geographical proximity does not necessarily mean socio-economic similarity.

The variables used for the study fall into four broad categories of variables: demographic, economic, social and standard of living. Demographic variables incorporate indices of aging, population growth and daily migration. Similarly, income tax base per capita and share of agricultural population fall under economic variables. Social variables include unemployment and number of students per thousand inhabitants and standard of living is determined by number of cars per thousand inhabitants. The study uses Ward's clustering procedure and K-means clustering procedure to obtain two large clusters and four smaller clusters, then compares the means of the variables for each.

In a 1994 study, the Economic Research Service (ERS) formed several classifications of non-metro counties to depict socio-economic diversity in rural America. They classify them in this way to identify groups of counties that shared common social and economic characteristics, enabling policymakers to further group them according to pertinent topics. ERS classifies the U.S. non-metro counties into seven broad overlapping types commonly known as "ERS county typologies." Four are classified based on the particular economic activity most depended on in that county: farming, manufacturing, mining and government. The other three broad topics are labeled according to the most relevant policy for that county: persistent poverty, Federal lands and retirement destination. Those counties that do not confirm to any particular groups are grouped separately as "unclassified counties."

ERS (1994) provided an update of the above classification, commonly known as the "1989 update." This update helped access the changes that had occurred in the non-metro counties between 1979 and 1986 and also encompassed some changes in the definition and concept of the classifications. This time the counties were classified into six non-overlapping types based on the primary economic activity of the non-metro counties, namely, farming-dependent, mining-dependent, manufacturing-dependent, government-dependent, services-dependent and non-specialized. Similarly, the counties were classified into five broad overlapping groups with more importance from rural policy point of view, namely, retirement-destination, Federal lands, commuting counties, persistent poverty, and transfers-dependent.

These classifications were delineated based on certain criteria. Certain thresholds had to be fulfilled for a county to be named as a certain type. For example, a county could be classified as manufacturing-dependent when manufacturing contributed to at least 30% of weighted annual total labor and proprietor income over a period of three years between 1987 and 1989. There were 556 farming-dependent counties, 190 retirement-destination counties, 146 mining-dependent counties, 270 federal lands counties, 506 manufacturing-dependent counties, 381 commuting counties, 244 government-dependent counties, 535 persistent poverty counties, 323 services-dependent counties, 381 transfers-dependent counties, and 484 non-specialized counties from the 1989 classification.

A large number of unclassified counties in the 1989 update prompted a need for revision and was done by Cook and Mizer (1994) for ERS as a “1990” update of ERS county typology. The types of groupings of counties were the same as in the 1989 update. It was felt that metro counties needed to be also included in the classifications. Hence, a new county typology was given by ERS (2005), which included all 3141 counties, county-equivalents and independent cities in the U.S. While the classification categories based on economic variables were the same six groupings as in the 1989 typology, the classification categories based on policy variables were different from the previous classifications. The new classification contained seven policy-based classifications, which were not mutually exclusive, namely, housing stress, low-education, low-employment, persistent poverty, population loss, non-metro recreation, and retirement destination.

These groupings were determined by similar criteria as the previous typologies. For example, for a county to be mining-dependent, at least 15% of average annual labor and proprietors' earnings had to come from mining sector. Among the economic types, there were 440 farming-dependent, 128 mining-dependent, 905 manufacturing-dependent, 381 Federal/state government-dependent, services-dependent 340 and 948 non-specialized counties in the 2004 typology. Similarly, there were 537 housing stress, 622 low-education, 460 low-employment, 386 persistent poverty, 601 population loss, 664 non-metro recreation, 440 retirement destination counties among the policy types.

Harris (1943) classifies U.S. cities by function based on the activity of greatest interest. The author puts forward that he improves on previous functional classifications, as the previous ones did not have sufficient criteria for distinguishing types and were poor in classifications that had more than one well-known type. Statistics on occupations and employment are the key determinants of the principal activities in each city. Nine major types of cities are identified: manufacturing (M), retailing (R), diversified (D), wholesaling (W), transportation (T), mining (S), educational (E), resort or retirement (X), and others (including political, P). Manufacturing (comprising 44% of the total metropolitan districts and 43% of smaller centers), retailing and diversified types of cities were the most numerous.

In order to rule out the related local service employment apart from their primary activities, Harris assigns different percentage values (high or low) to different functions. For example, the principle criterion used in grouping the manufacturing cities is that employment in manufacturing must equal at least 74% of total employment in manufacturing, retailing and wholesaling, while the principle criterion in grouping the wholesale centers states that employment in wholesaling must equal a minimum of 20% of the total employment in manufacturing, wholesaling and retailing.

Shields and Deller (1996) provide homogenous classification of Wisconsin counties for analytical comparisons, based on eight economic sectors: forest-related tourism, manufacturing and forestry, agriculture, tourism and government, manufacturing, urban, diversified and trade, and other. The authors recognize that geographical proximity is not a sufficient criterion for defining viable regions for the purpose of economic analysis. Principal component analysis is used to transform data into broad indices. The indices are then used to generate clusters of counties having similar economic structure.

At first, principal components or the linear combinations of the original variables, whose coefficients are the eigenvalues of the correlation coefficient matrix were developed. Interpreting with the help of loading scores, the absolute values of the principal component values closer to one are regarded to be important from the viewpoint of cause of variation in data. On the other hand, the values closer to zero are regarded as contributing little to the principal component. Finally, scores or the coefficients of the principal components are used to cluster the counties. Similar counties formed by the statistical techniques of principal components and cluster analysis could be grouped in the spatial sense or scattered across the state.

The clustering procedure uses 11 iterative algorithms, which minimize squared Euclidian distance for clustering. The cluster analysis minimizes the variations in the variables within a group, while at the same time maximizing the differences between different groups. In other words, economically similar counties are grouped together, while economically different counties are excluded from the group. The variables studied are the economic indicators of the county: the labor market, product market, and industry structures in agriculture, forestry, manufacturing, services, trade, government, and tourism. Using the clustering procedure, eight clusters, which could aid in different regional economic policy analysis are identified. Out of eight, seven clusters belong to one in each of the categories described above, while the eighth cluster, Madison, is a single county, Dane (which includes the state capitol and University of Wisconsin).

2.4 Methodologies Used in Classification

Various grouping procedures have been applied to combine locational entities into regions. Ilbery (1981) highlights the objective of the grouping procedure, which is to decrease the actual distance between observations in a group. Several techniques are available for estimating the distance between two observations, including the nearest, furthest, and total distance methods, as well as the group average and centroid replacement methods. These techniques, based on inter and intra-group distances are usually led by Ward's (1963) error sum of squares (ESS). A matrix of distance values can be obtained by this method by calculating ESS for each pair of observations, while at the same time distances of individual members to group centroids can be kept at a minimum**.** The author goes on to say that besides distance, which can be used to measure the degree of similarity or dissimilarity between observations, another popular criterion is the correlation coefficient. This value measures 'shape' distance between individuals. Depending on whether or not the data are normally distributed, Pearson's product moment and Spearman's rank correlations are the most commonly used coefficients.

According to Anderson (1975), hierarchical[8] methods of agglomerative classification use a calculation of similarity or dissimilarity between every pair of observations in the original dataset. After the two observations with smallest 'distance' between them are combined, the similarities between this new observation and others have to be calculated again. The procedure is repeated until all the observations form a single group. A linkage tree called a dendrogram is used to show the stages in the grouping process. Finally, a subjective decision has to be made to decide which step of the procedure provides the optimal number of groups.

Literature shows that some classifications of LMAs based on commuting focus on a "core" municipal area where there is a high inflow of commuters from surrounding areas. Prodromidis (2008) delineates Greece's labor market areas on the basis of a 15% commuting threshold by examining the 2001 inter-municipal travel-to-work flows of Greece. Two-way commuting from the fringes to the core and vice versa (similar to the UK self-containment algorithm and North American labor market definitions) is used for the study. Commuting origins and destinations are codified in a non-symmetrical iteration matrix and clustered without placing any restriction on contiguity. The iterative process attaches surrounding municipalities with significant commuting flows (to the city cores) with the city-cores in order to identify the boundaries of the major travel-to-work areas. This method uses the same building blocks and commuting data as done by Eurostat in a previous classification of labor market. However, Eurostat does not use two-way commuting, so Prodromodis' study is regarded as an improvement. Prodromidis' classification shows that, within Greece, Athens, Thessaloniki, and the urban centers of Patras, Iraklion, Larisa, Volos, and Ioannina have the largest LMAs surrounding them.

[8] Hierarchical groups are formed by adding members to the group in such a way that each addition reduces the number of subsets by causing least impairment to the objective function (Ward, 1963).

Kristensen (1998) uses input-output technique to analyze commuting data for Danish municipalities to develop functional economic areas (FEAs). Journey-to-work data is used to form clusters among 275 Danish municipalities. The study attempts to remove "urban bias"[9] prevalent in most studies by justifying that urban bias is only appropriate in areas where hinterlands have strong links to the urban areas, and only certain parts in Denmark have strong urban links. The study aims to minimize the level of subjectivity by choosing the cut-off level[10] and degree of closedness[11] carefully. A set of municipalities is considered closed when a group of municipalities in the set have commuting flows only to others within the set and open when the commuting flows go beyond the set. Kristensen uses open sets for the study, which is a modification of an earlier algorithm developed by Hewings (1996), who uses rigorous restriction and closed sets in the algorithm. Kristensen develops a matrix with pairs of municipalities with commuting flow. The dimension of the matrix is 275 by 275, as all the municipalities are included. Then, the following steps form an algorithm:

(i) A value is assigned to all the positive entries (t_{ij}) in the journey-to-work matrix.

(ii) All diagonal elements are assigned the value of one.

(iii) MI, MJ and MJI matrices[12] are created.

(iv) An ordering of closed matrices based on their appearance is generated.

---

[9] Urban bias refers to the focus on central urban nodes surrounded by the hinterlands, during the development of any FEAs. This assumption comes from the fact that hinterlands depend primarily on the urban nodes for the supply of goods services, jobs, income and growth (Kristensen, 1998).

[10] Cut-off level is the preferred point in time or space, at which the computer stops during the formation of clusters. Cut-off level affects the level of aggregation, as larger cut-off levels with larger commuting flows means that there will be fewer FEA's (Kristensen 1998).

[11] Closed sets have the same pattern of forward and backward linkages (Kristensen, 1998).

[12] MI contains municipalities with forward linkages, MJ contains the ones with backward linkages and MJI contains the potential closed sets that we are interested in.

(v) Open sets are sorted based on similarity.

(vi) The journey-to-work matrix is rearranged based on the new generated ordering.

The coefficients of the journey-to-work matrix are calculated from actual flows. Finally, the location quotient (LQ_{ij}) is calculated as the value of the entry (t_{ij}) in the matrix divided by the total of the respective column (T_j), which as a whole is divided by the total of respective row (T_i) divided by the grand total (T_{ij}) [$LQ_{ij} = (t_{ij}/T_j)/(T_i/T_{ij})$][13]. Decision rule of LQ > 0 is implemented in order to ensure that all entries are positive. A cut-off level of 0.75 is used for the location quotient during the formation of clusters. Using this methodology, 43 clusters, including one closed cluster, are obtained.

Andersen (2000) also attempts to delineate functional economic areas for Denmark. He focuses on core-commuting areas called commuting nodes. Besides average commuting data, shopping data are also collected as an observation of travel behavior. The criterion used for classification is that the group of municipalities in an economic area has to have a higher level of interaction within the group compared to interaction with municipalities from outside areas. Another condition is that at least one municipality is to be regarded as the center.

The algorithm decides whether a municipality/municipal couple is a center by calculating the value of coefficients k_1. Mathematically, k_1 is less than the number of employees living and working in the municipality/municipal couple divided by the number of employees living in the municipality/municipal couple. Similarly, coefficient k_2 is equal to the number of people living and working in the group of municipalities divided by the total number of in-commuters and out-commuters in an area. The higher value of k_2 shows that the area is closed and has no interaction with other areas. The algorithm first shows the possible centers where people both live and work. Each municipality is assigned to the center area, with which it has the highest level of

[13] Symbols T_i, T_j and T_{ij} denote the totals of the row, column and the grand totals respectively.

interaction. A drawback of the study is that ad-hoc algorithm is used for grouping the municipalities, as a result of which the results can not be guaranteed to be unique[14] and urban-bias may be present.

Several attempts to delineate labor markets for the U.S. have been done in the past. Tolbert et al. (1987) uses commuting data from the 1980 census to delineate labor market areas to be used for statistical and planning purposes in rural America. With their work, they aim to develop a geographical standard, which captures the variations in local economic and labor force activities. Previous studies represent labor markets as Standard Metropolitan Statistical Areas (SMAs) because of the easily accessible data on urban areas. These studies omit the rural areas by definition, and as a result, they provide little help in the research on non-metro employment patterns. The 1980 county group designations of labor markets based on state planning district and Metropolitan Statistical Area (MSA) boundaries is not satisfactory either, as they are confined within the state boundaries. Census county groups bounded within the States do not provide a good measure of labor markets, as the area formulations between different states are inconsistent and limited by the arbitrary interstate lines. Bureau of Economic Analysis (BEA) economic areas formed from 1960 and later 1970 census commuting-to-work data focuses on an urban center and surrounding counties. So, they are not found to be very useful for non-metro labor market research.

Hence, the study by Tolbert et al. (1987) forms the most up-to-date delineation of LMAs, which focuses on both metropolitan areas and non-metro areas, using journey-to-work files of county and county equivalents in the 50 states and District of Colombia. First, frequency matrices with rows representing the county of residence or county of origin, and columns

[14] When it is not unique, it cannot be said with certainty that no more areas could be separated from the resulting areas.

representing county of work or county of destination are created. The numbers in the matrices represent how many people commute to the county of work; the extreme right hand side of the matrix shows total employed people in a county of origin. The diagonal represents the people working in the same county as the county of residence. Tolbert et al. convert absolute commuting flows in the frequency matrices to flow matrices, in which the commuting flows are expressed as proportional measures to account for the wide variations in the county populations.

For counties i and j, proportional flow measure is defined as the sum of commuters from counties i and j, divided by the resident labor force of the smaller county. Having the smaller county's resident labor force in the denominator helps to establish even highly asymmetrical commuting patterns as an evidence of a strong labor market tie. Moreover, use of volume of shared commuters in a relative rather than an absolute basis ensures that larger counties do not dominate the analysis. Using resident labor force instead of daytime labor force or non-resident labor force helps ensure that it is constant across all versions of frequency matrices and not sensitive to commuting direction. First of all, frequency matrices with rows representing he counties of residence and columns representing the counties of work are prepared. The diagonals represent workers who do not commute to another county. Substantial geographic overlap ensures that interstate LMAs are identified wherever appropriate.

Then, the measure of association ($P_{ij} = P_{ji}$) were computed for each pair of counties in a frequency matrix using as the sum of the number of persons commuting from county i to j and the number of persons commuting from county j to i, divided by the minimum among the resident labor force of the two counties, i or j. The main diagonal of the flow matrices is set to zero ($P_{ij} = P_{ji} = 0$ when $i = j$). Finally, proportional flow measures are expressed as distance measures ($D_{ij} = D_{ji}$) as one minus the measure of association ($P_{ij} = P_{ji}$). Flow matrices resulting from the frequency matrices represent the similarity matrices used in the delineation analysis. Six different matrices for six different regions of the country are prepared. The regions are: West, Southwest, Midwest, Central, Southeast and Northeast. Hierarchical cluster analysis based on the extent of commuting is used to cluster the counties. They use dendrogram[15] to interpret the results of the hierarchical cluster analysis. County clusters with normalized distances of less than 0.98 are grouped into the same LMAs conditional upon their fulfillment of 10,000 minimum population criteria. Three hundred eighty-two LMAs fulfilling the minimum population criterion of at least 10,000 are formed.

Tolbert and Sizer (1996) update the 1980 delineation of U.S. commuting zones and labor markets using the journey-to-work file from the 1990 census and the same methodology as used previously. They view local labor markets as a set of relationships between employers and workers that exist in space bounded by places of work and residence. This spatial conception of labor markets dictates the methods, data sources, and procedures of classification. Tolbert and Sizer justify the use of counties as the units of analysis by pointing out that a vast amount of socioeconomic and political data can be obtained at the county level. They identify 741 commuting zones when forming clusters, not using the Census Bureau's population minimum of 100,000. When the minimum criterion is used, only 341 labor market areas are identified. Beginning with a distance of 0.7 and continuing to the maximum, dendrograms are used to depict between-cluster average distance in a vertical manner.

The Bureau of Census has several definitions of economic areas. Johnson and Kort (2004) write that Bureau of Economic Analysis (BEA) economic areas are formed of relevant regional markets surrounding metropolitan or micropolitan statistical areas. They are based on

[15] A dendrogram is a tree-diagram, often used to illustrate the arrangement of clusters obtained from hierarchical clustering.

homogeneity of regions and are not limited by state boundaries. These economic areas consist of one or more economic nodes (metropolitan or micropolitan statistical areas) surrounded by counties that are economically dependent on these nodes or regional centers of economic activity. BEA economic areas serve as regional markets for labor, commodities and information. They are based on the U.S. Office of Management and Budget's (OMB's) definitions of urbanization-based statistical areas. OMB uses the term Core-based Statistical Areas (CBSAs) for the areas based on urban cores, having a minimum population of 10,000. Metropolitan statistical areas (MSAs) are CBSAs whose population exceeds 50,000. Combined statistical areas (CSAs) are those contiguous CBSAs that fulfill OMB's criteria for interdependence.

Using commuting data from the 2000 decennial census, redefined statistical areas from OMB of February 2004, and newspaper circulation data from the Audit Bureau of Circulations for 2001, BEA economic areas were redefined in 2004. The redefined BEA economic areas are largely based on CSAs, MSAs and micropolitan areas. Newspaper readership data are used to measure regional markets in less populated parts of the country. The number of economic areas increased from 172 in the previous delimitation of 1995 to 179 in 2004. Similarly, the number of Component Economic Areas (CEAs) decreased from 348 to 344.

The grouping procedure is accomplished in three phases: economic nodes are identified, counties are assigned to CEAs and finally, CEAs are aggregated to form BEA economic areas. It starts with 3,141 counties in the U.S., and first produced 344 nodes composed of 37 micropolitan areas performing as nodes and 305 MSA-based CEA nodes. From them, 165 combinations of CEAs that did not meet the Economic Area (EA) criteria were left out to form the final 179 BEA economic areas. The fixation of maximum rate of total out-commuting at 8% and the maximum rate of commuting from one economic area to another at 4% ensures that there is limited market interdependence.

Kongari et al. (2011) use trade data from IMPLAN (IMpact Analysis for PLANning)[16], and commuter data from Local Employment Dynamics (LED) to classify labor market in the Gulf of Mexico (GOM) region. IMPLAN uses multiregional input-output models and estimates county-to-county trade flows. U.S. freight survey data is combined with algorithms from Oak Ridge National Laboratories that give dollar worth of physical quantities. This is part of a cooperative agreement between the authors and the Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) to reevaluate the previously formed thirteen regions known as "on shore areas," along the GOM coast using a more recent dataset and theory. They propose a linkage coefficient that quantifies the strength of the bond between the two counties. The strength of the linkage coefficient is inversely proportional to the linkage coefficient's numerical value.

Kongari et al. (2011) developed a matrix of linkage coefficients and cluster using PROC CLUSTER in SAS. They place no constraint on the geographical contiguity of the parishes. Using this methodology, the authors form regions based on total industry trade, specific oil and gas industry trade, and commuting patterns. At first, they prepare a matrix of trade or commuting flows. From there, they calculate the linkage coefficient by subtracting imports and exports between two counties divided by the sum of the total trade of the two counties minus the imports and exports between two counties from one as a whole. The counties are aggregated using SAS

[16] IMPLAN is a database created by MIG Inc. for creating complex social accounting matrices and multiplier models of local economics.

in an iterative manner repeatedly until the desired number of clusters are obtained. Finally, the clusters were mapped using GIS[17] and also represented in dendrograms.

They observed that county trade is less affected by state boundaries, as compared to labor markets along the GOM coast. Using a 40-industry subset of the trade data, 100 regions are formed. Similarly, one hundred regions are formed from the commuting data of the original 534 counties in the GOM. In the same way, from the original 13 formed in the BOEMRE classification, 39 new regions are formed from the oil and gas trade regional clustering.

Kongari et al. also test the homogeneity of the individual counties on the basis of socio-economic variables. R-square values from regressions are calculated, a higher R-square meaning that the counties clustered together tended to share similar socio-economic characteristics. The results from the calculated R-square values show that the all-trade data dominated rest of the clustering approaches for classification. Moreover, they find that re-classifying the same GOM counties into more numerous regions does not necessarily create more homogenous regions of county groupings. We referred to Kongari et al. (2010) for a definition of the linkage coefficient for the labor market classification based on commuting for this study. This study uses Ward's method for grouping the regional clusters based on commuting. On the other hand, regional classification based on socioeconomic variables is done by using K-means clustering.

It is obvious from the literature that classification of regions and labor markets forms an important part of regional studies and development scholarship. Classifications based on clustering can be identified in a number of ways; there is no specific rule to guide which classification methodology is better than the rest. It depends mostly on the subjective decision of

[17] GIS is a technology, which allows us to view, understand, question, and, interpret, by visualizing data in different ways that reveal relationships, patterns, and trends in the form of maps, globes, reports, and charts. It helps to capture, analyze, and manage geographically referenced information systematically by integrating hardware and software with data (gis.com, 2011).

the researcher. Most classifications of region focus on a central urban node and the more rural regions surrounding those areas. Few classifications of U.S. counties as a whole, based on commuting and socioeconomic variables, are available. Some U.S. states are classified based on certain economic criteria, but the literature lacks any reference for regional market and labor market classification for Louisiana. Hence, this study will try to fill that gap.

3. Data and Methodology

3.1 Data

Secondary data from journey-to-work data files of the 2009 Local Employment Dynamics (LED)[18] from the U.S. Census Bureau are used for the delineation of labor markets based on commuting. The data files are obtained from the Longitudinal-Employer Household Dynamics (LEHD)[19] homepage of the U.S. Census Bureau website[20] from an application called "OnTheMap." The Quarterly Census of Employment and Wages (QCEW) provides the workplace data for this application. The data from QCEW does not include self-employed people, railroad workers and Federal government employees such as military personnel. Statistical Administrative Research System (StARS) provides the data of residence (Murakami, 2007). It combines the data from a variety of federal sources, such as Social Security, Internal Revenue Service (IRS), Medicare, Medicaid, and Veterans' Affairs. The addresses are provided as Census Block codes.

LEHD provides data organized in the form of origin-destination files. The variables provided in the files are number of workers with their respective home and work Census Block Federal Information Processing System (FIPS) geographical codes (geocodes). Geocodes of the census blocks are represented by 15-digit FIPS codes. The data from LEHD is synthesized data, meaning that the distribution of origin-destination flow is synthetic. It is slightly distorted in

[18] LED is a voluntary partnership between state labor market information agencies and the U.S. Census Bureau to develop new information about local labor market conditions (U.S. Census Bureau, 2010). A variety of data regarding labor market can be obtained from the LED.

[19] LEHD is an innovative program of the U.S. Census Bureau, which combines federal and state administrative data on employers and employees using modern statistical and computing techniques with core Census Bureau censuses and surveys (US Census Bureau, April 22, 2011).

[20] <http://lehd.did.census.gov/led/onthemap/>

order to protect confidentiality of workers. However, the distortion is an insignificant amount, so as not to affect the real data; the counts of workers living and working in a specific block are real.

Parishes are used as the basic unit of analysis for this study. Hence, only the first five FIPS codes shown below in Table 1 are required to distinguish different parishes; the other ten codes have no relevance to this study. The origin-destination files provide the total number of commuters in different age, earning and industry categories. But this study is only concerned with the total number of commuters in all categories as a whole. A representation of the LEHD data of origin-destination is given below:

Table 1 Origin-destination file from LEHD

Work Geocode	**Home Geocode**	**Number of Commuters**
220010000000000	**22001**0000000000	0
220017000000000	**22003**5000000000	7
220035000000000	**22001**0000000000	5
220037000000000	**22003**7000000000	0
220037000000000	**22005**9000000000	9

Source: Local Employment Household Dynamics, U. S. Census Bureau (2011)

For the description of clusters using socio-economic variables, we use four variables: in-commuters by civilian labor force ratio, unemployment rate, median household income and number of establishments. Initially, additional socioeconomic variables widely used in literature were included, but as they had high correlation, we omitted them from the study. The effect of correlation is discussed later in the results section of the study.

In-commuters data is from LEHD[21], civilian labor force data is from the Bureau of Labor Statistics (BLS), U.S. Census Bureau[22], median household income and unemployment data is from the Economic Research Service (ERS)[23] and data on the total number of establishments is from County Business Patterns, U.S. Census Bureau[24]. All variables use data from 2009.

In-commuters to civilian labor force ratio is an important variable as it helps give an idea of the composition of the labor force. Civilian labor force consists of people 16 years and older, both employed and unemployed. It excludes those who are in the army or who are institutionalized. Median household income is used in this study, rather than average household income, because it is regarded to be more stable since it is not affected by very high or low values. Unemployment is a serious social problem and can be a challenge to policymakers trying to solve the issue. Establishment means a distinct physical place of business, rather than an entire business.

3.2 Methodology

Two different types of methodologies are used for the classifications. The first clustering is done for commuter-based classification and uses Ward's method. The second clustering is done for socio-economic variable-based classification uses k-means clustering.

3.2.1 Clustering for Commuter Based Classification

The conceptual model of a local labor market based on commuting is based on the observation that whenever there is abundance of employment in a county or parish, people from

---

[21] http://lehd.did.census.gov/led/onthemap/la/od/

[22] http://censtats.census.gov/cgi-bin/usac/usacomp.pl

[23] (http://www.ers.usda.gov/data/unemployment/RDList2.asp?ST=LA

[24] http://censtats.census.gov/cgi-bin/cbpnaic/cbpcomp.p

surrounding counties will in-commute to a job in that county. It may also be due to higher wage in that county, compared to the current one they are living in. Similarly, if people see prospects of better jobs or more opportunity of employment in another county, they will out-commute and find jobs elsewhere**.**

In this study, we attempt to form a labor market based on commuting via clustering procedure. Fisher and Ness (1971) point out that it is impossible to find the optimal clustering procedure. Hence, we use admissible clustering procedures. SAS (Statistical Analysis System) clustering algorithm PROC CLUSTER is applied for grouping the regions. It is a hierarchical clustering algorithm used to form the clusters by using one of the eleven agglomerative methods of clustering. Agglomerative methods of clustering are a type of hierarchical clustering, which use the “bottom-up” approach. This approach is based on grouping smaller clusters into large ones. In contrast, divisive clusters use “top-down” approaches and are based on splitting big clusters into smaller ones.

Ward’s method (Ward, 1963) is used for clustering the regions. Ward’s method has better distance between-cluster as compared to some other methods like centroid and average in the initial results. Celik et al. (2011) use Ward’s method to cluster provinces in the East Anatolia region in Turkey using socio-economic variables. In their study, results from Ward’s method are more anticipated than other clustering procedures. Blien et al. (2006) analyze model-based classification of regional labor markets designed to access labor market policy in Germany and use Ward’s hierarchical clustering procedure. They point out that Ward’s method gives more uniform clusters, as compared to other similar clustering methods and also has less singular clusters. Massart and Kaufman (1983) use cluster analysis for the interpretation of analytical chemical data. They suggest transforming the raw data before clustering, so that the gross size or the range of variation does not influence the classification. Here in our study, the use of the linkage coefficient (Equation 4), which comes later in the chapter, does the work of transformation of the data.

Ward's method (1963) is based on minimizing within-cluster variance. This systematically provides the objective function value associated with n to 1 number of groups. Ward clarifies that the number of groups to be formed is in fact given by the changes in objective function values as the number of clusters decreases. One advantage of this method is that number of clusters does not need to be specified in advance, as it systematically provides the objective function value associated with n to 1 number of groups. However, for this study, we specify the number of clusters ourselves. In the initial results, the statistics determining the optimum number of clusters do not seem to show significant improvement at various levels. This may be caused by low variation in the data due to clustering on the basis of only one variable or low sample size. Hence, we specifically set the number of clusters at eight to compare and contrast with the eight Metropolitan Statistical Areas in Louisiana.

Ward's method gives the measure of distance between two clusters in terms of the increase in the sum of squares when we merge them. The objective function in Ward's clustering is based on minimizing the error sum of squares (ESS). A mean is set to represent all scores in a group; individual scores in a group are not considered. The ESS acts as an indicator, which represents the "loss" of information, caused by representing the mean score of the group, instead of individual scores. Equation 1 below gives ESS:

$$ESS = \sum_{i=1}^{n} {x_i}^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right)^2 \text{................(Equation 1)}$$

In equation 1, x_i is the value of the score, a measure of rating of the i^{th} individual. The aim of the procedure is to repeat the grouping process until the groups can be combined from n groups to 1 group in the best possible way so as to minimize ESS.

PROC CLUSTER takes data in either coordinate or distance form. Both rows and columns in a distance matrix correspond to the objects to be clustered (SAS documentation, 2011). On the other hand, coordinate matrix has observations in the rows and variables on the columns. The data is fed as distance measure in the study. PROC TREE statement in SAS is used to create dendrograms. The dendrogram output allows us to visualize cluster membership at different levels of the cluster tree (SAS documentation). Finally PROC FREQ statement provides the frequency distribution of parishes in different clusters.

The linkage coefficient (C_{ij}) is the only variable used for clustering. Number of in-commuters and out-commuters for each pair of parishes and total number of commuters in each of the parishes are the variables for calculating C_{ij}. Cubic Clustering Criterion (CCC) and overall R-squared measure goodness-of-fit. No constraint is placed on the regions; in terms of contiguity, a region could be composed of parishes not contiguous to one another. Starting with 64 parishes/counties in Louisiana, parishes of eight clusters form a labor market.

First, commuters' matrix (O) is developed by entering parish of work or the destination parish in the first column, and home parish or origin parish in the first row in a Microsoft Excel worksheet. Then, the observation for each home-to-work parish pair is entered as the total number of people coming from the county-of-residence to work into the work-county from the home county. The total number of commuters in each county is calculated by adding in-commuters and out-commuters. First of all, to get total in-commuters, the values across each row are summed and to get total out-commuters, the values across the corresponding columns are summed. As the data of commuters at the county level is synthesized from original data of commuters at the census block level, the number of people commuting within the same county of residence must be forced to zero. All the data is mined and the coefficient values calculated in the Microsoft Access 2010 database. An example of a commuter matrix for five counties represented by A, B, C, D and E is shown in Table 2.

By transposing the in-commuters matrix, we obtain the out-commuters matrix (I) by placing the home parish in a row and work parish in a column. Finally, we derive the journey-to-work or commuters' matrices by adding these two matrices (Kongari et al., 2011).

$I = O^T$ ……………………………………………………………………………(Equation 2)

$T = O + I$……..……………………………………………………………........(Equation 3)

Table 2 Commuters matrix

O_{ij}	A_j	B_j	C_j	D_j	E_j	$\sum i$	T_i
A_i	0	4	12	5	1	22	22 + 17 = 39
B_i	7	0	2	11	3	23	23 + 20 = 43
C_i	6	8	0	9	15	38	38 + 18 = 56
D_i	1	1	1	0	5	8	8 + 30 = 38
E_i	3	7	3	5	0	18	24 + 18 = 42
$\sum j$	17	20	18	30	24		

The classification methodology used defines a linkage coefficient that quantifies the strength of the bond between two counties (Kongari et al., 2011). The linkage coefficient between two counties is defined by the total number of commuters between a pair of parishes, divided by the total number of commuters in the two parishes, and then subtracted from one as a whole. The formula for the strength of the bond between two counties is given in equation 4 below.

$$C_{ij} = 1 - \frac{T_{ij}}{T_i + T_j - T_{ij}} \text{ ....(Equation 4)}$$

Where,

C_{ij} = coefficient of linkage between two counties A_i and A_j

T_{ij} = number of commuters between two counties A_i and A_j

T_i = total number of commuters in county A_i

T_j = total number of commuters in county A_j

The denominator represents the subtraction of the intersection set from the union of two sets. It can be represented in set form as:

$A \cup B = A + B - A \cap B$ ……………………………………………………(equation 5)

An example of the final linkage matrix is given below:

Table 3 An example of a coefficient matrix

C_{ij}	A_j	B_j	C_j	D_j	E_j
A_i	0	A_i B_j	A_i C_j	A_i D_j	A_i E_j
B_i	B_i A_j	0	B_i C_j	B_i D_j	B_i E_j
C_i	C_i A_j	C_i B_j	0	C_i D_j	C_i E_j
D_i	D_i A_j	D_i B_j	D_i C_j	0	D_i E_j
E_i	E_i A_j	E_i B_j	E_i C_j	E_i D_j	0

Using the above definition, a linkage coefficient matrix as shown in the table above is developed with diagonal elements being zero. In other words, these are the values of C_{ij} for each combination of county pairs. The table is symmetric, with the coefficients being identical on both sides of the diagonal. For example, the coefficient of linkages between parishes A and B (A_iB_j) will have an equal value to the coefficient of linkage between parishes B and A (B_iA_j). Such developed matrix is passed to PROC CLUSTER for the purpose of classification.

The objective of the clustering procedure is to group similar observations together. The link is stronger with smaller coefficient (Cij) values. Hence, when a coefficient approaches zero, both parishes combine to form a new functional region. SAS interprets the coefficient as a distance. So, the linkage increases as the distance decreases. If we want to form a group of ten clusters from a group of 50 parishes, SAS finds the parishes with the shortest distance and groups them as one. It is an iterative algorithm. The same step is then repeated until the desired number of clusters is reached. In other words, this iterative procedure optimizes the number of groups specified *a priori* by the researcher. Finally, this clustering process can also be represented in the form of dendrograms in SAS to make the process of clustering clear.

In the final matrix, missing values are filled with the coefficient values of 1. This is done by assuming that there are no commuters between two parishes, where the number of commuters is not given in the data from LEHD. When we assume there are no commuters between two parishes, the linkage coefficient value becomes one (1), which is the weakest linkage. Initial results of clusters which have matrices missing many observations, yield more scattered clusters, while the clusters obtained after using one (1) for the missing values yield clusters with groups of neighboring parishes. Similarly, there are some other missing values in the case of parishes with only one-way commuting. In such cases, the missing values are filled with the corresponding coefficient values from the opposite side of the diagonal, as the matrix is symmetric with same values on either sides of the diagonal. After the clusters form, we use Geographic Information System (GIS) to map them.

3.2.2 Clustering for Socioeconomic Variables Based Regional Classification

To cluster the second group based on socioeconomic data, we apply PROC FASTCLUS in SAS, which is a non-hierarchical clustering algorithm (K-means). Celik et al. (2011) use the non-hierarchical k-average technique to group provinces in the East Anatolia region of Turkey, using 49 socio-economic variables. The K-means clustering algorithm assumes K-clusters determined *a priori* and defines one centroid per cluster, which is the mean of the observations. Then it uses Euclidian distance as a measure for assigning each observation to a centroid. The process is repeated to reach a minimum objective function. Goodness-of-fit is measured with Cubic Clustering Criterion (CCC), pseudo-F statistic, and overall R-squared.

4. Results and Discussion

4.1 Results of the Commuter Based Regional Classification

Formation of eight clusters using Ward's method on commuting data yields the result shown in table 4. It shows the parishes belonging to each cluster along with the total number of parishes in each cluster, the percentage of parishes in each cluster and the location of the clusters in Louisiana.

Table 4 Cluster result using commuters

Cluster Number	Parish Name	Number of Parishes	Percentage of Parishes	Location
1	Bossier Caddo	2	3.23	Northwest
2	Jefferson Orleans	2	3.23	Southeast-central
3	Lafourche Terrebonne	2	3.23	Southeast
4	East Baton Rouge Livingston St. Tammy Tangipahoa St. Charles St. John Baptist Ascension Iberville West Baton Rouge East Feliciana Assumption Washington St. James Pointe Coupe Plaquemines St. Bernard St. Helena	18	28.13	Southern-central
5	Beauregard Vernon Calcasieu Jefferson Davis Allen Evangeline	9	14.06	Southwest

(Table 4 continued)

5	Cameron Acadia St. Landry			
6	Iberia Lafayette St. Martin St. Mary Vermilion	5	7.81	Southern-central
7	Grant Rapides East Carroll West Carroll Catahoula Concordia Franklin Richland Natchitoches Sabine Caldwell La Salle Avoyelles Winn Madison Tensas De Soto Red River	18	28.13	Northern
8	Morehouse Ouachita Jackson Lincoln Claiborne Webster Union Bienville	8	12.5	Northern-central

Cluster 1 (Bossier and Caddo) consists of the same parishes as the Shreveport-Bossier City MSA. Similarly, Cluster 2 corresponds very closely to New-Orleans-Metairie-Kenner MSA, but the MSA has seven parishes and the cluster only contains two parishes (Jefferson and Orleans); the other parishes in the MSA form clusters in other groups. Cluster 3 consists of

Lafourche and Terrebonne, which lie in the Houma-Bayou Cane-Thibodaux MSA in southeast Louisiana.

Cluster 4 and Cluster 7 are the biggest clusters with 18 parishes each (more than 29% of the observations). Cluster 4 lying in Southern-central Louisiana is mostly formed of parishes surrounding East Baton Rouge Parish. It corresponds very closely to Baton Rouge MSA. East Baton Rouge Parish contains the City of Baton Rouge, which is important from a number of viewpoints. Baton Rouge is the capital of Louisiana; hence, most of Louisiana's government employees commute to this area on a daily basis. Moreover, Louisiana State University, the state's largest university, is also there. This brings commuters with university-related jobs to this parish. Exxon Oil also has a facility there, bringing in a large number of oil and gas industry employees.

Cluster 5 consisting of seven parishes in Southwestern Louisiana includes parishes in Lake Charles MSA (Calcasieu and Cameron) and seven surrounding parishes (Vernon, Jefferson Davis, Beauregard, Allen, Acadia, St. Landry and Evangeline). Cluster 6 consists of parishes surrounding the Lafayette MSA (Lafayette and St. Martin) in Southern-central Louisiana. Surrounding parishes include St. Mary, Iberia, and Vermillion. Cluster 7 is another larger cluster consisting of 18 parishes in Northeastern and Central Louisiana. This cluster is formed of La Salle, Winn, Natchitoches, and Avoyelles parishes around the Alexandria MSA (Grant and Rapides) and parishes in Northern Louisiana including some touching the adjoining parishes in the state of Mississippi. They include East Carroll, West Carroll, Madison, Franklin, Richland, Tensas, Sabine, Catahoula, Concordia, Caldwell, Red River and De Soto.

Cluster 8 lying in Northern-central Louisiana consists of Union and Ouachita parishes from Monroe MSA and six other surrounding parishes (Morehouse, Jackson, Lincoln, Claiborne, Bienville and Webster). Overall, we observe that the clusters are not very uniform. Table 5 (below) shows the socioeconomic properties of the parishes in each cluster by averaging the unemployment rate, median household income, number of establishments and the ratio of commuters by civilian labor force of each group of parishes per cluster.

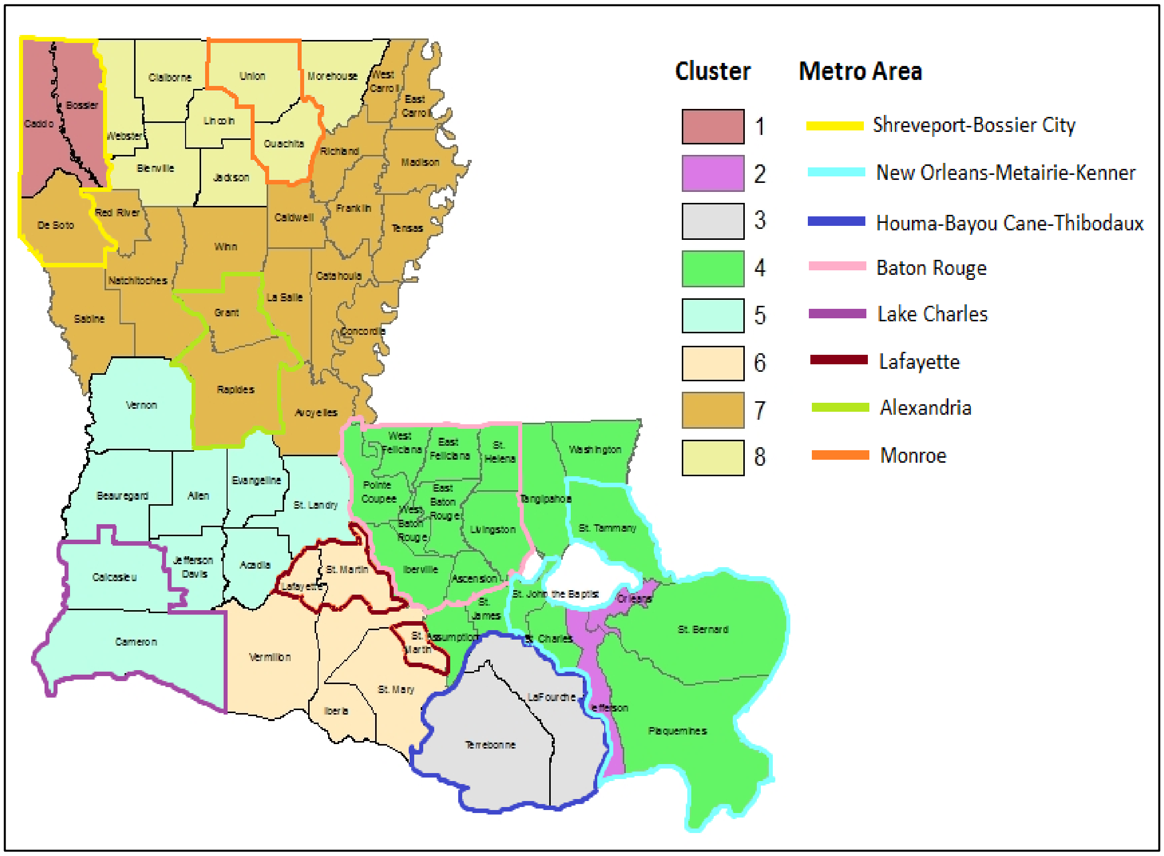

Figure 1 Map of Louisiana showing labor market based on commuting

Figure 1 represents a map showing the labor market of Louisiana based on commuting. The highlighted borders show the places where the MSAs are situated. From these results, it seems like labor markets formed of commuting correspond very closely to the MSAs defined by the census. However, in most cases, they are much larger than the MSAs, as shown by the commuting patterns. This labor market defines the area as shown by interrelationships between all possible parish pairs. The MSAs take only the urban parishes and their surrounding parishes into account, while this labor market treats clustering all parishes equally.

Table 5 Socioeconomic characteristics of the clusters

Cluster Number	Unemployment Rate (%)	Median Household Income ($)	Number of Establishments	In-commuters /Civilian Labor Force
1	6.600	43,574.000	4,353.000	0.430
2	6.900	40,835.500	10,126.000	3.476
3	4.550	47,737.000	2,444.000	0.638
4	7.211	41,442.111	1,674.889	0.442
5	6.856	39,212.444	1,140.889	0.293
6	6.360	41,930.200	2,520.400	2.415
7	9.322	30,824.722	483.277	0.221
8	9.075	33,633.250	946.750	0.229

Table 5 shows the mean of socioeconomic characteristics of the eight clusters. Four variables describe the clusters: unemployment rate (%), median household income ($), number of establishments and in-commuters by civilian labor force ratio. Cluster 1 (Bossier and Caddo parishes) and Cluster 2 (Jefferson and Orleans parishes) have similar unemployment rates (6.6% and 6.9% respectively) and median household incomes ($43,574 and $40,835.5 respectively). But because the number of establishments in Cluster 2 is significantly higher (10,126) compared to Cluster 1 (4,353), it has a much higher ratio of in-commuters to civilian labor force with a value of 3.476, compared to Cluster 1 with a value of just 0.43.

Cluster 3, containing Lafourche and Terrebonne parishes has the lowest unemployment rate with a value of 4.55, and the highest average median household income with a value of

47,737. It has a fairly large number of establishments (2,444) and, as a result, the ratio of commuters to civilian labor force is quite high (0.638). Cluster 4 has fairly high median household income ($41,442.111), though the unemployment rate is rather high (7.211%). Clusters 5 and 6 and Clusters 1 and 2 seem to be similar to each other in terms of unemployment rates and median household income categories.

While the number of establishments is very similar between Cluster 2 and Cluster 5, there are surprisingly different values of the ratio of in-commuter to civilian labor force. Statistics show that Cluster 7 and Cluster 8 are formed with poorer parishes. They have the highest unemployment rates with values of 9.322% and 9.075% respectively. Similarly, they have the lowest median household incomes with the values of $30,824.722 and $33,633.250 respectively. The number of establishments in these clusters is very few; Cluster 7 has an average of around 483 and Cluster 8 has an average of around 947. As a result, they also have low values of in-commutes to civilian labor force ratios.

The map of Louisiana showing the various clusters obtained from commuting data, with the spatial distribution of the respective means of the clusters' socioeconomic characteristics is shown in Figure 2. Short denotations have been assigned to each variable name for clarity on the map. UR denotes unemployment rate, MHI denotes median household income, NOE denotes the number of establishments and IC/CL denotes the ratio of in-commuters by civilian labor force.

A dendrogram tree of the linkage distance of clusters at various semi-partial R-squared values is shown in Figure 3. Semi partial R-squared values give an idea of the loss of homogeneity due to merging two clusters to form a single new cluster. Hence, lower values are better. The parishes are labeled by their corresponding names. Different parish names with their corresponding FIPS codes are given in the Appendix. We can see which parishes were joined in the different stages of clustering by using a dendrogram. For example, we see that uniting Bossier and Caddo parishes form the first stage cluster. Similarly, Jefferson and Orleans parishes form the second cluster. Parishes unite with one another forming bigger clusters in each stage or iteration until there is only one cluster. In this way, dendrogram gives a visual assessment of the clustering process and helps determine the optimum number of clusters.

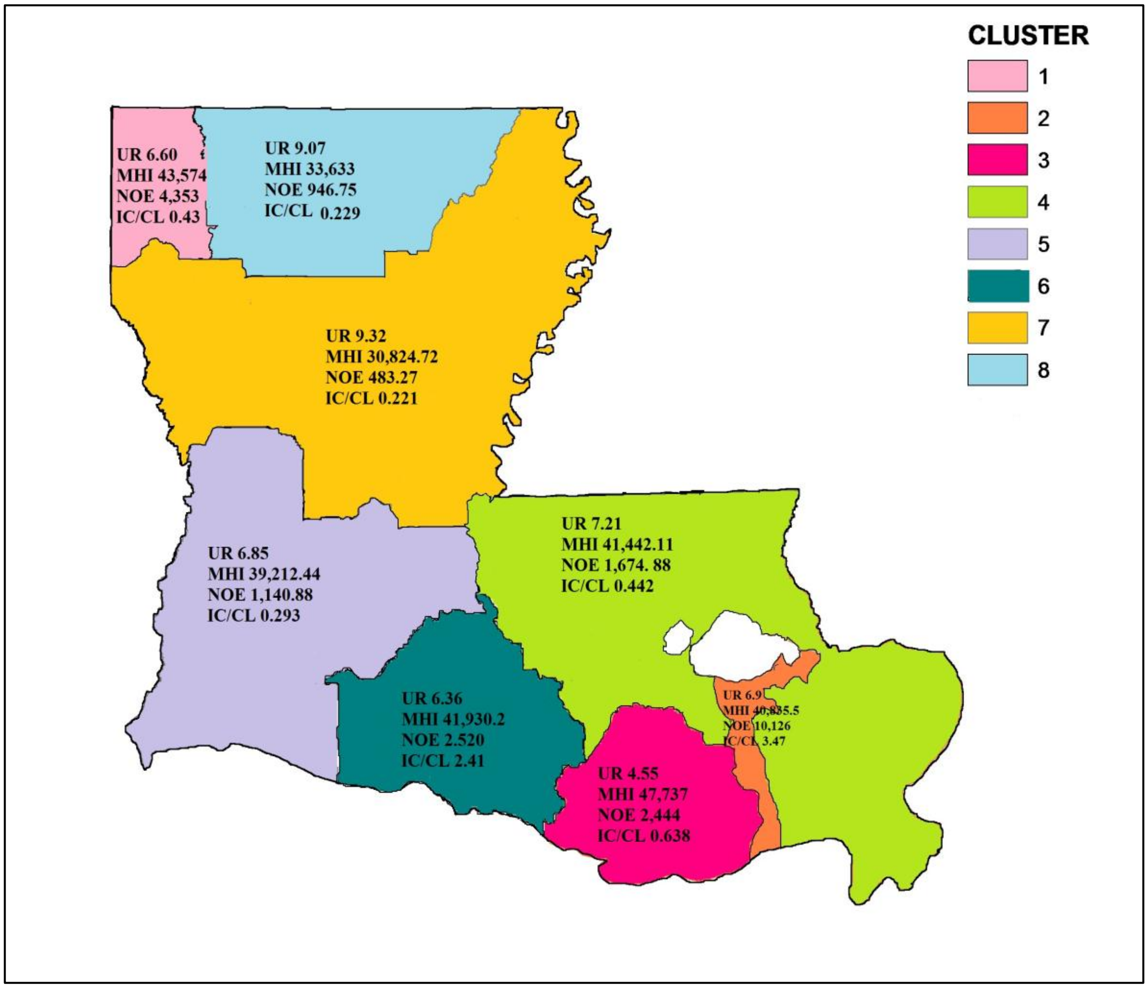

Figure 2 Map of Louisiana showing the spatial distribution of the socioeconomic characteristics of the regional clusters

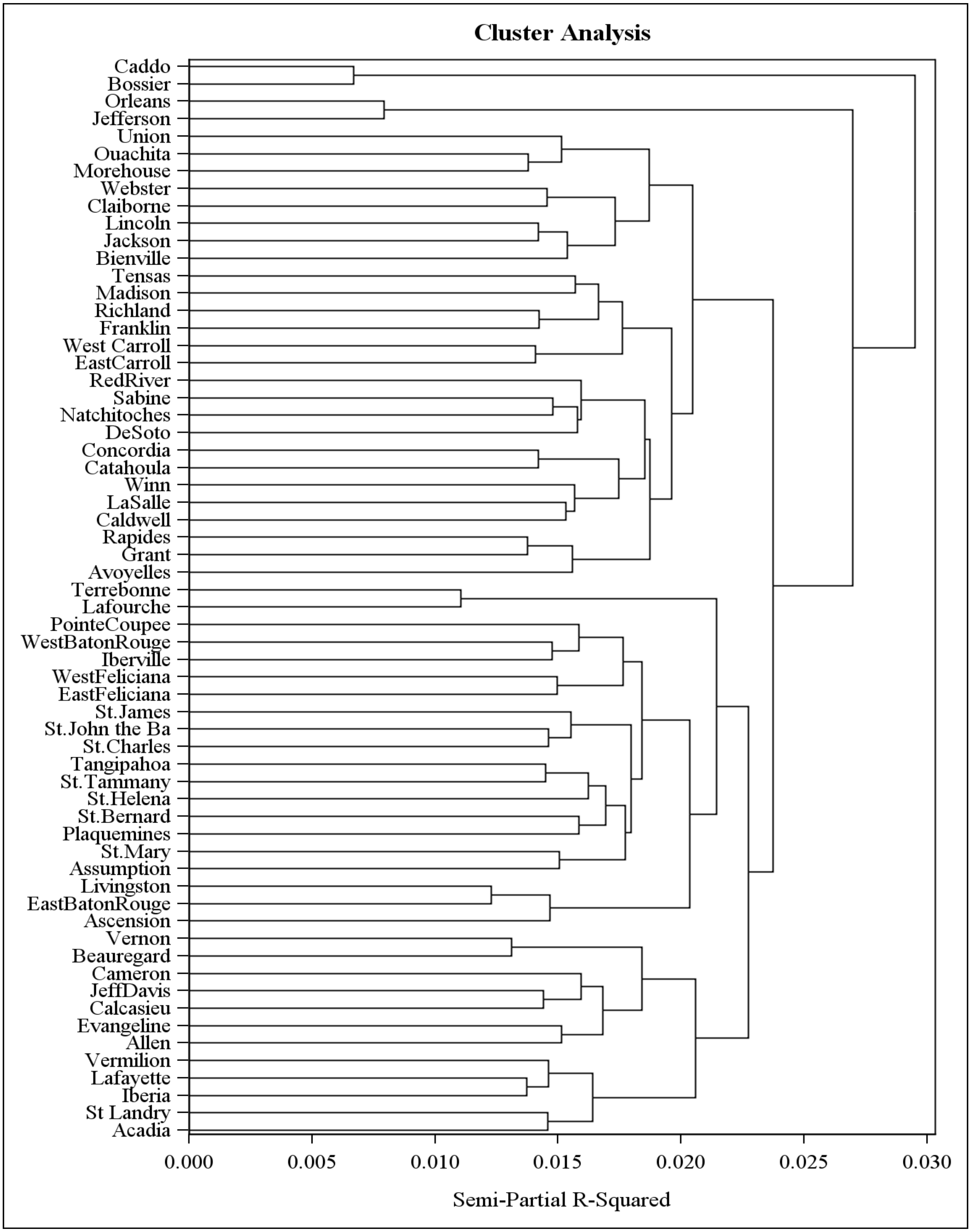

Figure 3 Dendrogram tree of the clustering process

As seen by the relationship between the R-squared statistic (Figure 4) and the number of clusters, clusters do not seem to have a very close association within themselves. The R-squared shows the proportion of between-cluster variation explained by the variables. Here, the R-squared is increasing with the number of the clusters and reaches a maximum around 64 clusters.

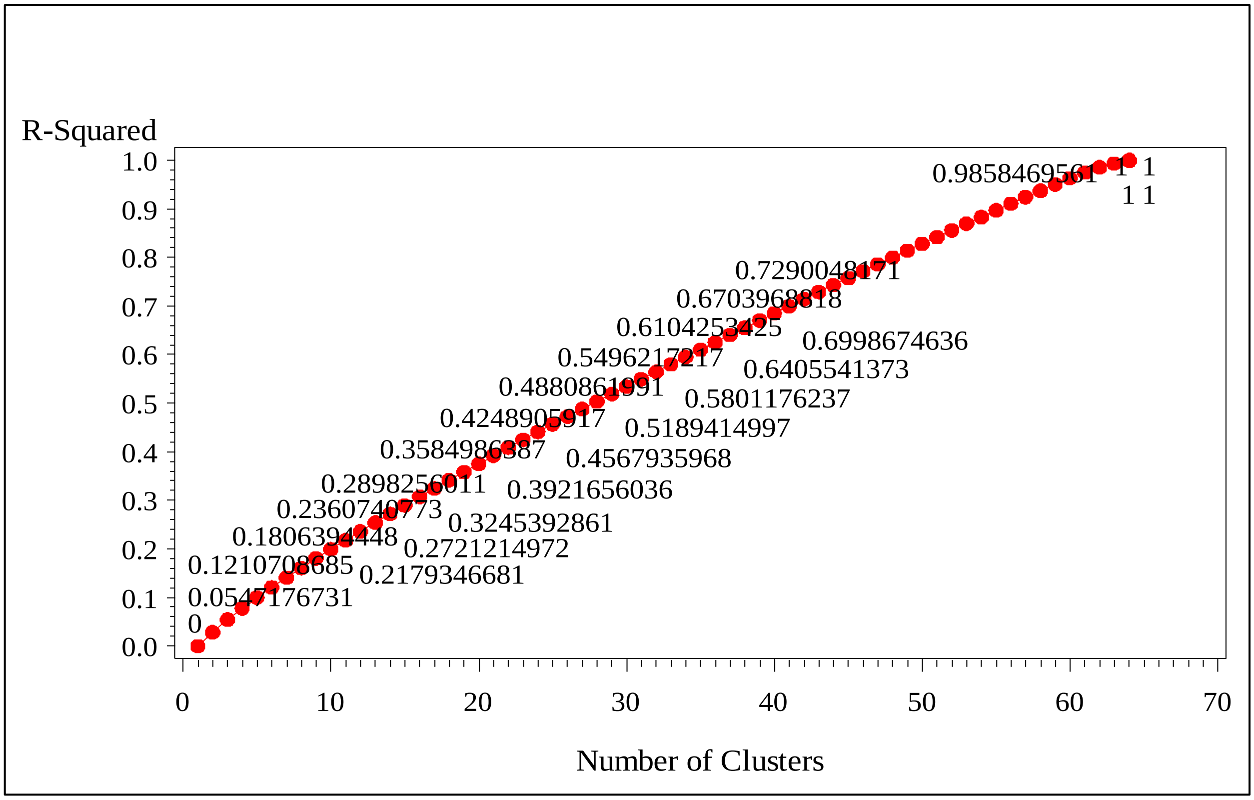

Figure 4 R-squared values with the number of clusters

Similarly, the graph of Root-Mean-Square Standard Deviation (RMSSTD) versus number of clusters (Figure 5) shows that there is not much difference in standard deviation up to the 60th cluster, after which it drops significantly. RMSSTD is a measure of homogeneity within clusters. Ideally, we want the deviation among the clusters to be higher, so that the clusters have higher degrees of separation and significant difference between them.

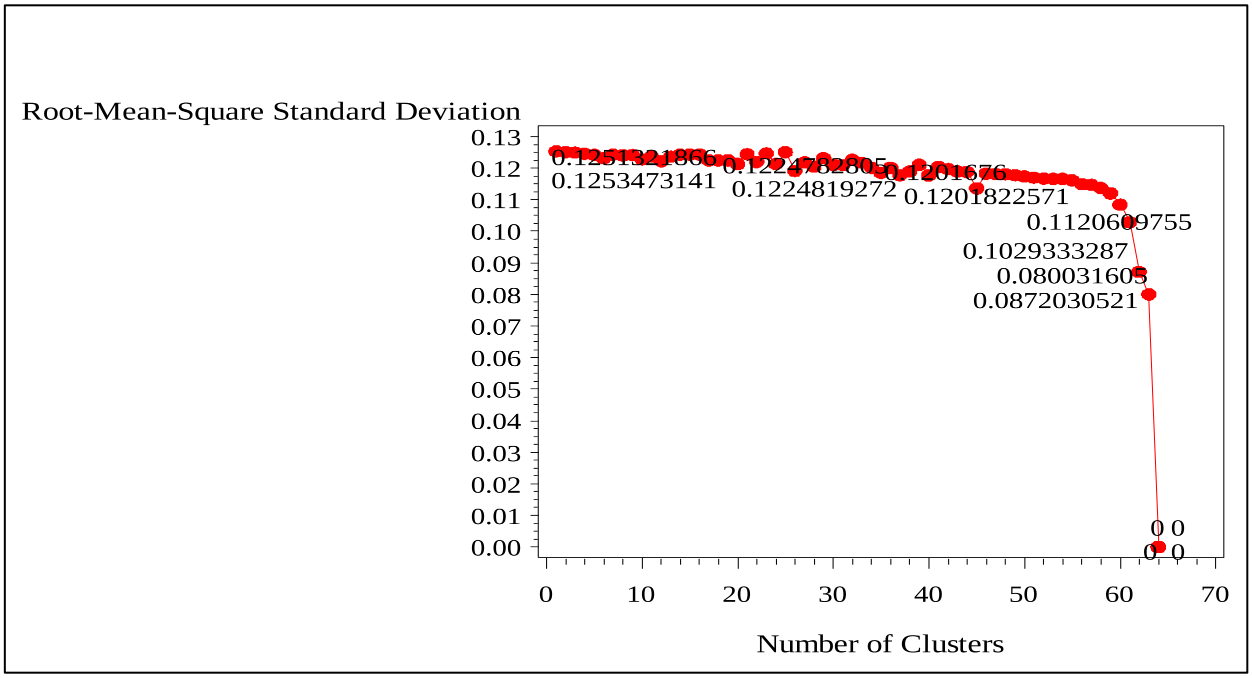

Figure 5 Root Mean Square Standard Deviation with the number of clusters

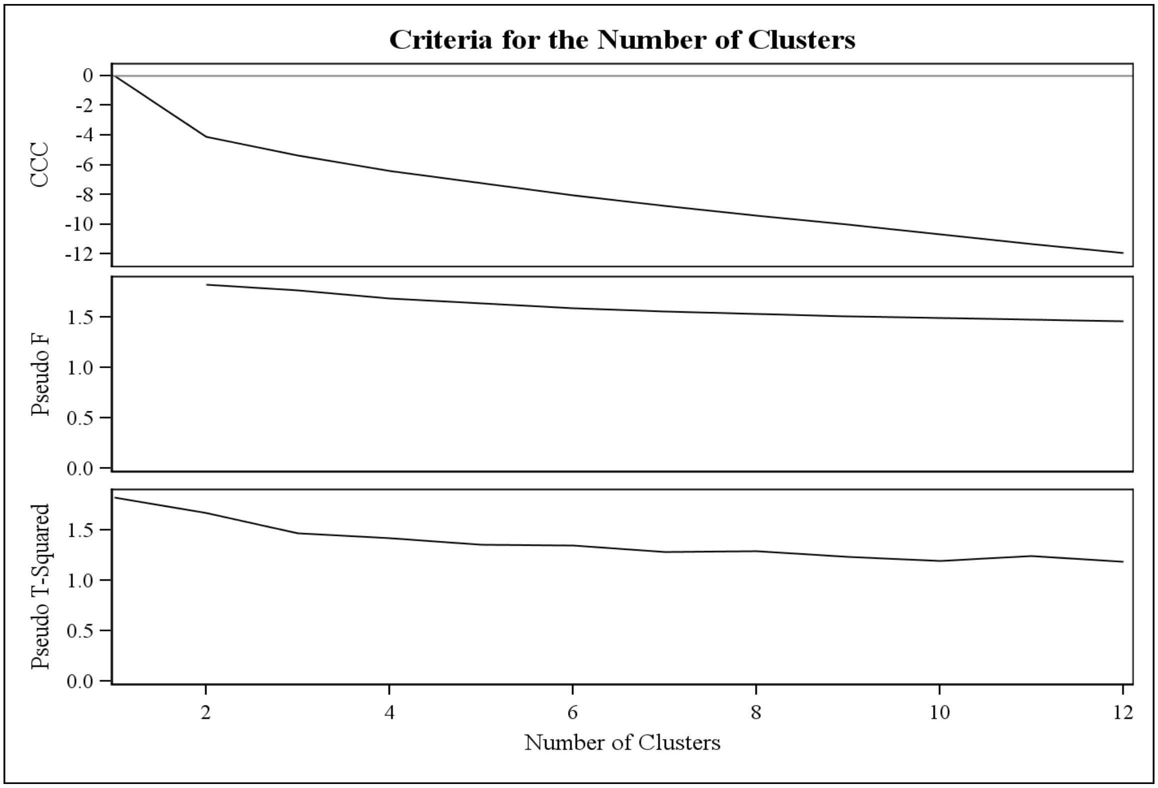

Figure 6 Criteria for the number of clusters

An output of the Cubic Clustering Criterion (CCC) and pseudo-F criterion for 12 clusters is shown in Figure 6. CCC is a statistic developed by SAS for determining the optimum number of clusters. A higher value of CCC is favored. Similarly, pseudo-F[25] is a statistic which, which captures the tightness of the clusters. Here, the criterions do not peak, making it difficult to determine the total number of clusters for best results. This could be due to the fact that only one variable has been used in the clustering, and the variable C_{ij} might not have been able to account for much of the variation in the clusters. Another reason could be a low sample size.

4.2 Results of the Socioeconomic Variables Based Classification

In this part of the study, goodness-of-fit criteria were used to determine the optimum number of clusters. Goodness-of-fit criteria contain three statistics: R-squared, Cubic Clustering Criterion (CCC), and pseudo-F (Table 6). The researcher may want to classify regions according to individual requirements, but the goodness-of-fit criteria provide the statistics to help make the decision easier. Plotting the above values (Figure 7) clarifies where the peaks of these values occur, in order to decide the best number of clusters.

First of all, the values of Pearson’s correlation coefficient are calculated to make sure no significant correlation existed between the variables (Table 7) and results show no correlation value of any major concern. Some other variables are initially included for clustering, but removed after calculating the values of the Pearson’s correlation coefficient. The goodness-of-fit criteria used for determining the optimum number of clusters do not work when the variables are correlated which is why it is important to rule out significant correlation between the variables before going ahead with the clustering.

[25] Mathematically, pseudo-F equals the ratio of mean of the sum of squares between cluster groups to the mean of the sum of squares within groups.

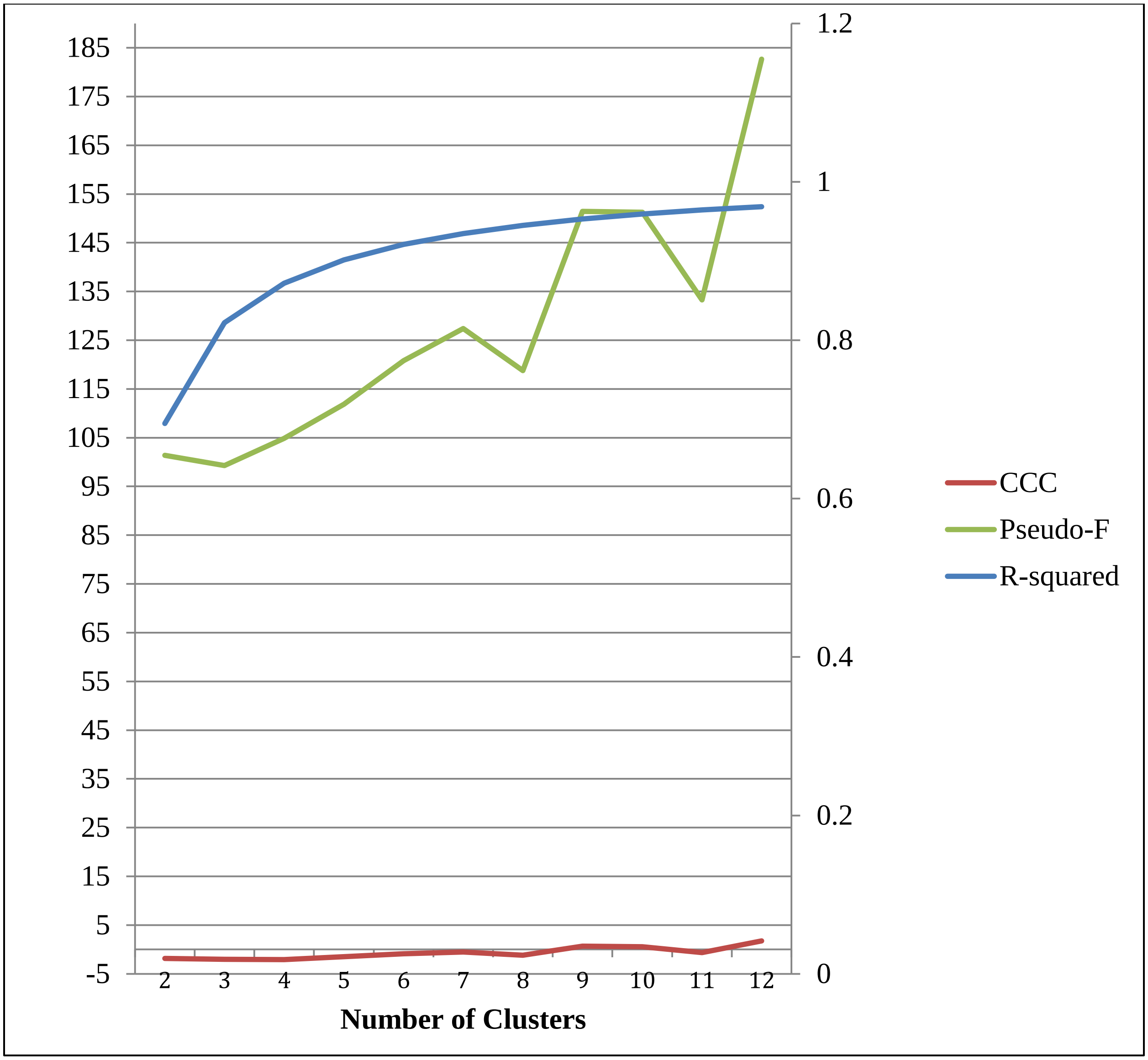

Figure 7 Goodness-of-fit criteria with the number of socioeconomic clusters

A plot of the graph of CCC, pseudo-F statistic, and overall R-squared shows that each statistic peaks at around nine clusters (Figure 7) then falls. Hence, Louisiana is segmented into nine clusters based upon those four socioeconomic variables chosen for the study. The result in Table 8 is obtained by using k-means clustering to the socioeconomic data after determining that the optimum number of clusters is nine.

Table 6 Goodness-of-fit criteria values

Number of Clusters	R-squared	CCC	Pseudo-F
2	0.6947	-1.836	101.41
3	0.82231	-1.998	99.26
4	0.87195	-2.108	104.87
5	0.90135	-1.506	111.87
6	0.92084	-0.897	120.78
7	0.9346	-0.525	127.41
8	0.945	-1.177	118.78
9	0.95295	0.669	151.41
10	0.95933	0.537	151.27
11	0.96447	-0.614	133.3

Table 7 Correlation matrix of the socioeconomic variables

Pearson Correlation Coefficients, N = 64 Prob > |r| under H0: Rho=0				
	Unemployment Rate	**Median Household Income**	**In-commuting /Civilian Labor**	**Number of Establishments**
Unemployment Rate	1.000	-0.635 <.0001	-0.229 0.068	-0.377 0.002
Median Household income	-0.635 <.0001	1.000	0.228 0.069	0.319 0.0103
In-commuting /Civilian Labor	-0.229 0.068	0.228 0.069	1.000	0.544 <.0001
Number of Establishments	-0.377 0.002	0.319 0.010	0.544 <.0001	1.000

Table 8 Cluster result using socioeconomic variables

Cluster Number	Parish Name	Number of Parishes	Percentage of Parishes
1	Acadia Allen Caldwell De Soto East Feliciana Grant Iberville Jackson Jefferson Davis Lincoln Pointe Coupe Rapides Sabine St. Bernard St. Mary Tangipahoa Union Vermilion Webster West Carroll	20	31.25
2	East Baton Rouge Jefferson Lafayette	3	4.69
3	Ascension St. Tammany	2	3.13
4	Assumption Beauregard Calcasieu Iberia La Salle St. Martin Vernon West Baton Rouge	8	12.5
5	Cameron St. Charles	2	3.13
6	East Carroll Madison Tensas	3	4.69
7	Avoyelles Bienville Catahoula Claiborne	15	23.44

(Table 8 continued)

7	Concordia Evangeline Franklin Morehouse Natchitoches Red River Richland St. Helena St. Landry Washington Winn		
8	Bossier Lafourche Livingston Plaquemines St. James St. Johns Terrebonne West Feliciana	8	12.5
9	Caddo Orleans Ouachita	3	4.69

Table 9 represents the mean of each variable per cluster. Cluster 2 has the highest mean of the total establishments, and hence the highest ratio of in-commuters per labor force (Table 9). Since the parishes in Cluster 3 have the highest median household income, their rate of in-commuter per civilian labor force is less than most of other clusters mean. Cluster 2 and 3 seem to be formed of wealthier parishes. Cluster 2 includes East Baton Rouge, Jefferson and Lafayette, while Ascension and St. Tammany are in Cluster 3. All these parishes are parts of metropolitan areas. According to Ascension Economic Development Corporation[26], Ascension Parish is one of the fastest-growing parishes of the U.S. and is the fastest-growing parish in Louisiana. It boasts one of the best school systems in Louisiana.

[26] http://www.ascensionedc.com/

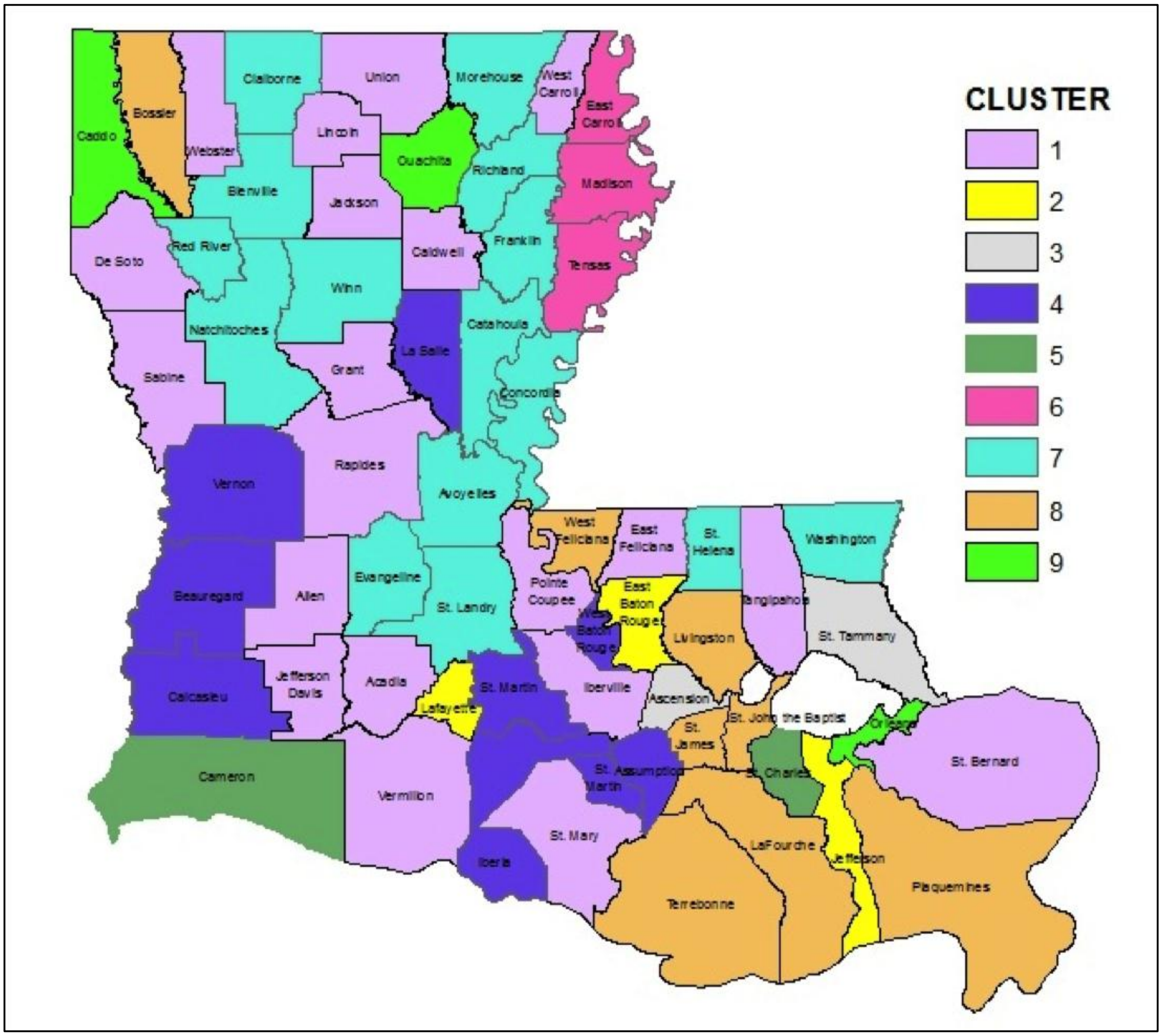

Figure 8 Map showing regional classification of Louisiana based on socioeconomic variables

Similarly, East Baton Rouge contains the capitol city, Baton Rouge. Lafayette parish contains the city of Lafayette and the remaining two parishes Jefferson and Ascension contain the city of New Orleans. Grand Isle, a popular tourist destination, is situated in Jefferson Parish. The presence of these important cities cause Clusters 2 and 3 to have more establishments, higher median household income, and lower unemployment rates; overall, they seem to be better off in socioeconomic development compared to the other clusters. Figure 8 shows the map of Louisiana obtained by clustering the parishes based on socio-economic characteristics.

Table 9 Mean of the four variables per cluster

Cluster Means				
Cluster	**Unemployment Rate**	**Median Household Income**	**In-commuting /Civilian Labor**	**Total Establishments**
1	7.970	36,873.500	0.300	782.900
2	5.666	46,349.666	5.805	10,658.666
3	5.500	59,931.500	0.303	3,909.000
4	6.775	42,310.375	0.315	1,153.875
5	6.000	55,993.000	0.640	552.500
6	11.133	24,042.333	0.214	142.000
7	9.280	30,793.866	0.269	482.866
8	6.475	48,752.125	0.484	1,352.750
9	7.266	37,088.666	0.435	6,286.000

Cluster 6 seems to be formed of poorer parishes, as it has the highest unemployment rate, lowest median household income, and lowest number of establishments and lowest ratio of in-commuters to civilian labor force. An interesting observation is that these parishes are the same as the ones delineated as non-white majority parishes by the Rural Policy Research Institute (2006). Neighboring effect of poorer counties from Mississippi adjoining these parishes could also be a contributing factor for this. The parish clusters are scattered and there does not seem to be any obvious patterns in the socioeconomic characteristics. However, the northern parishes seem to be less developed compared to the southern parishes.

5. Summary and Conclusions

5.1 Summary

Regional classification and labor market study form important steps towards balancing the development of any region and implementing regional economic policy. The literature shows that various parts of the U.S. and Europe have been classified using commuting data and socio-economic variables. The variables and methodology used for the classification in those studies varies greatly according to the purpose of the study and subjective decision of the researcher. However, Ward's method of clustering has been used fairly widely. Regional economists have defined the two broad categories for the classifications of regions to be functional and homogenous. Two more types of classifications named administrative and nodal classifications are derived from them. The government does administrative classifications for the purpose of local governance. In the case of Louisiana, the state has been classified into 64 geographical units called Parishes.

We have delineated two types of regional classifications for Louisiana. The first type is a dynamic approach and the second type is a static approach. First of all, we classified Louisiana parishes on a functional basis into labor market areas (LMAs). These functional relationships between parishes were based on commuting behavior, as shown by journey-to-work files. We defined a coefficient, referring to Kongari et al. (2011), and used it as a variable for classification. We mapped the clusters for the ease of description. We also provided the socio-economic profiles of each clusters. We compared the LMAs with the eight metropolitan areas in Louisiana, and found that these overlap the metropolitan areas. Present maps of metropolitan statistical areas (MSAs) show only the parishes included in the metro and the parishes surrounding them as micropolitan statistical areas. They completely rule out the rural parishes, and those parishes who are not a part of MSAs or micropolitan statistical areas. The LMAs in the form of commuting clusters take all parishes into account and show their close association, regardless of whether they belong to any MSAs or not.

This study can first help policy makers make decisions regarding labor policy. As all the parishes in the commuting clusters are interrelated by commuting linkage, implementation of labor market policy into the whole LMA instead of only the parish in question should be considered. In addition, entrepreneurs might get insight into the best places to establish certain industries by taking into account the concentration of labor in various parishes. The study shows that northern labor markets formed of Clusters 7 and 8 are poorer, as shown by the values of the socio-economic variables. Hence, labor policies directed at removing poverty should be focused on Northern Louisiana. This classification can also provide useful guidance to the Department of Transportation Planning. Roads among the parishes in a cluster should have a good network in the shortest way possible to ensure efficiency of labor flow.

In the second part of the study, parish classification was further augmented to form clusters based on socio-economic variables using k-means clustering. These clusters show parishes with similar socio-economic behavior. This study can help policy makers and government officials by using a shotgun approach when implementing policy. For example, if there is a budget for development works, the fund can be used in the regions that need them most based on the socio-economic characteristics. Parishes need to have balanced socio-economic development for any state to prosper. If similar economic policy is implemented in different regions with different outcomes, a deeper study should be conducted in order to analyze causation. Then appropriate policy recommendations can be made to change the strategy or bring in new policy.

We can clearly see that there is huge disparity in the clusters' economic characteristics. Regions lagging in development should take initiative to implement policies that specifically target the area that is behind. For instance, Cluster 6 has very few establishments, low median household income, and a high unemployment rate as compared to other clusters. This observation can incite policy makers to take a more holistic approach when implementing policy in areas with similar socio-economic characteristics.

5.2 Limitations of the Study

The dataset from the Census Bureau has administrative limitations, particularly when place of work and place of payroll diverge. For example, construction workers may have a different place of work than the place where they receive their payroll. This causes ambiguity in answering where the place of work is. Problems may also arise in stating the place of residence when people with temporary jobs (such as summer jobs) may indicate their permanent home address as their residential address instead of their current address. Similarly, student workers may write their parents' address for residence in W-2 for social security records.

In addition, Murakami (2007) points out that there may be a potential problem with wrong addresses provided by undocumented workers with "borrowed" social security numbers. The data from QCEW is inadequate to accurately measure the commuters, as it does not include self-employed people and federal government employees. Nationwide, around 10 percent of workers are self-employed, and approximately one percent of workers are Federal government employees (Murakami, 2007). This confirms that a section of the workforce has been unaccounted for by the LEHD OnTheMap data. The socio-economic variable based clustering gives equal weight to all variables, but some variables may be more important with respect to others in the parish's economy.

Louisiana is bordered by Texas, Mississippi and Arkansas. Hence, there must be commuting flows between parishes in the Louisiana border and these states. However, we have only used commuting flows within Louisiana for the study. This may have led to omission of some proportion of commuters. Hence, the true extent of the labor market might have been overshadowed.

5.3 Future Research

The labor market based on commuting has been formed based only on commuting linkages. For further research, other relationships like trade between goods and services could be taken into account. We can see from the results that the clusters using socio-economic variables have very low goodness-of-fit statistics. This could be due to inclusion of a low number of variables, indicating that further classifications may benefit from including more socio-economic variables. Another beneficial approach in studying the labor market and in regional study might be to observe the “before” and “after” implementation of a certain policy in order to analyze the effectiveness of the policy. A final interesting avenue for future study would be to take minimum population into consideration while defining regions.

References

Anderson, K. E. "An Agricultural Classification of England and Wales." *Tijdschrift Voor Economische en Sociale Geografie.* Vol. 66, Issue 3 (1975), pp. 148–158.

Blien, Uwe, Franziska Hirschenaeur and Phan thi Hong Van. " Model-Based Classification of Regional Labor Markets for Purposes of Labor Market Policy." *Beiträge zum wissenschaftlichen Dialog aus dem Institut für Arbeitsmarkt- und Berufsforschung.* IAB Discussion Paper. No. 29 (2006).Web. July 10, 2011. <http://doku.iab.de/discussionpapers/2006/dp2906.pdf>

Brown, Lawrence A. and John Holmes. "The Delimitation of Functional Regions, Nodal Regions, and Hierarchies by Functional Distance Approaches." *Journal of Regional Science.* Vol. 11, No. 1 (1971), pp. 57-72.

Celik, H. Eray, Sinan Saracli, and Sanem Sahribanoglu. "Analysis of Socio-Economic Similarities of the Cities in East Anatolia Region". *World Applied Sciences Journal*. Vol. 12, No. 10 (2011), pp. 1807-1813.

"Census 2000 Geographic Definitions". US Census Bureau. July 13, 2011. Web. September 15, 2011. < http://www.census.gov/geo/www/geo_defn.html#CensusDivision>.

Cook, Peggy J. and Karen L. Mizer. "The Revised ERS County Typology, An Overview". Economic research Service, United States Department of Agriculture. December, 1994.Web. November 15, 2011. <http://www.ers.usda.gov/publications/rdrr89/rdrr89.pdf>.

Davis, Elizabeth E. and Fantu N. Bachewe. "Employment and Commuting Patterns in Rural Labor Markets." Department of Applied Economics, University of Minnesota. May, 2004. Web. January 15, 2011. <http://purl.umn.edu/19955>.

"Demographic and Economic Profile Louisiana". Rural Policy Research Institute. July, 2006. Web. January 5, 2012. < http://www.rupri.org/Forms/Louisiana.pdf>

Fannin, J. Matthew, Ashok K. Mishra and Raghavendra Kongari. "Initial Labor Market Region Results (Presentation)". Department of Agricultural Economics and Agribusiness, LSU Agricultural Center. August 6, 2010.

Fisher, Lloyd and John W. Van Ness. "Admissible Clustering Procedures." Biometrika. Vol. 58, No. 1 (Apr., 1971), pp. 91-104.

"Fisheries Economics of the United States 2008." National Marine Fisheries Service. April 2010. Web. March 1, 2011. <http://www.st.nmfs.noaa.gov/st5/publication/econ/2008/FEUS 2008 ALL.pdf>

"General Facts about the Gulf of Mexico." Environmental Protection Agency. May 17, 2010. Web. March 15, 2011. < http://www.epa.gov/gmpo/about/facts.html>.

"Geographic areas Reference Manual." U.S. Census Bureau. Sep 16, 2005. Web. April 14, 2011. < http://www.census.gov/geo/www/GARM/Ch4GARM.pdf>.

"Gulf OCS Region Activities." Bureau of Ocean Energy Management Regulation and Enforcement. April 1, 2011. Web. April 1, 2011. <http://www.gomr.boemre.gov/homepg/offshore/gulfocs/gulfocs.html>.

Holmes, John H. "Linkages Between External Commuting and Out-Migration: Evidence from Middle-Eastern Pennsylvania." *Economic Geography*. Vol. 48, No. 4 (1972), pp. 406-420.

Hoover, E.M., and F. Giarratani. An Introduction to Regional Economics. *The Web Book of Regional Science*. ed., Scott Loveridge. Morgantown, WV: Regional Research Institute, West Virginia University. 1999. Web. Nov 1. 2010. <http://www.rri.wvu.edu/regscweb.htm>.

Ilbery, Brian W. "Dorset Agriculture: A Classification of Regional Types." *Transactions of the Institute of British Geographers*, Royal Geographical Society. Vol. 6, No. 2 (1981), pp. 214-227.

Johnson, Thomas G., Daniel M. Otto and Steven C. Deller. *Community Policy Analysis and Modeling*. Iowa: Blackwell Publishing Professional, 2006.

"Journey to Work 2000." Census 2000 Brief. U.S. Census Bureau. March, 2004. Web. March 20, 2011. < http://www.census.gov/prod/2004pubs/c2kbr-33.pdf>.

Kristensen, Kund. "Functional Economic Areas of Denmark: Applying Input-Output Techniques to Commuting." Research Center for Bornholm Forsknings Center, Denmark. August, 1998.

"Local Employment Dynamics." U.S. Census Bureau, Office of Economic Studies, Dec. 11 2009. Web. Nov 15, 2010. <http://lehd.did.census.gov/led/led/led.html>.

Massart, D. Luc and Leonard Kaufman. The Interpretation of Analytical Chemical Data by the use of Cluster Analysis. New York: John Wiley and Sons, 1983. Print.

"Measuring Rurality: 2004 County Typology Codes". Economic Research Service , United States Department of Agriculture. June 26, 2005. Web. December 25, 2011. <http://www.ers.usda.gov/Briefing/Rurality/Typology/Methods/>.

Measuring Rurality: Commuting Zones and Labor Market Areas." Economic Research Service, United States Department of Agriculture. July 21, 2010. Web. December 30, 2011.

Murakami, Elaine. "Understanding LEHD and Synthetic Home to Work Flows in "On The Map." Office of Planning, Environment and Realty (HEP), U.S. Department of Transportaion. May 24, 2007.

"Occupational Employment Statistics (OES) Highlights." The Bureau of Labor Statistics, U.S. Department of Labor. October, 2010. Web. March 3, 2011. <http://www.bls.gov/oes/highlight_gulf.pdf>

"Outer Continental Shelf (OCS) Oil and Gas Assessment 2006." Bureau of Ocean Energy Management Regulation and Enforcement. Nov 12, 2010. Web. April 1, 2011. <http://www.boemre.gov/revaldiv/RedNatAssessment.htm>.

Poole, Richard W. "Implications of Labor Characteristics and Commuting Patterns for Regional Analysis: A Case Study." *Land Economics.* Vol. 40, No. 1 (1964), pp. 110-116.

Prodromidis, Prodromos-Ioannis. "Deriving Labor Market Areas in Greece from Commuting Flows." Center for Planning and Economic Research (KEPE), Athens, Greece. April, 2008.

Renkow, Mitch, Dale M. Hoover and Jon Yoder. "Commuting and Migration in North Carolina: Does Suburbanization Explain Trends?" AER No. 13, Department of Agricultural and Resource Economics, North Carolina State University (1997).

Rovan, Joze and Joze Sambt. "Socio-economic Differences among Slovenian Municipalities: A Cluster Analysis Approach." Developments in Applied Statistics. Metodoloski, 19, Ljubljana: FDV (2003).

"State Totals: 2011 Archives." LSU AgCenter. 2011.

Tolber, Charles M. and Molly Sizer. "US Commuting Zones and Labor Market Areas: A 1990 Update." Staff Paper No. AGES 9614, USDA-ERS (1996).

Ward, Joe H. Jr. "Hierarchical Grouping to Optimize an Objective Function." *Journal of American Statistical Association*, Volume 58, Issue 301(March, 1963), pp. 236-244.

Appendix Data of Socioeconomic Characteristics of Louisiana Parishes

FIPS	Parish Name	Unemployment Rate	Median Household Income	In-commuters /Civilian Labor	Number of Establishments
22001	Acadia	6.3	35,583	0.283	1,114
22003	Allen	9.1	34,506	0.459	338
22005	Ascension	5.7	60,995	0.343	1,962
22007	Assumption	7.8	42,494	0.178	257
22009	Avoyelles	7.3	30,791	0.184	704
22011	Beauregard	7.6	42,167	0.247	589
22013	Bienville	9.3	29,847	0.322	250
22015	Bossier	5.8	49,053	0.489	2,375
22017	Caddo	7.4	38,095	0.371	6,331
22019	Calcasieu	6.1	43,534	0.262	4,283
22021	Caldwell	9.3	35,345	0.209	191
22023	Cameron	5.7	55,117	0.633	159
22025	Catahoula	9.9	29,892	0.226	173
22027	Claiborne	9.1	32,301	0.268	252
22029	Concordia	10.7	28,520	0.323	380
22031	De Soto	8.2	34,958	0.243	382
22033	East Baton Rouge	6	44,720	0.620	12,169
22035	East Carroll	12.6	23,186	0.195	124
22037	East Feliciana	7	38,856	0.324	265
22039	Evangeline	7.7	30,897	0.245	523
22041	Franklin	10.4	30,031	0.224	396
22043	Grant	7.6	38,335	0.107	187
22045	Iberia	6.7	41,272	0.514	1,752
22047	Iberville	9.2	38,703	0.759	536
22049	Jackson	7.4	35,359	0.191	250
22051	Jefferson	6.1	46,428	6.354	11,928
22053	Jefferson Davis	5.6	39,359	0.020	605
22055	Lafayette	4.9	47,901	10.443	7,879
22057	Lafourche	4.4	47,909	0.163	1,938
22059	La Salle	6.6	41,808	0.030	290
22061	Lincoln	7.2	35,111	0.488	980
22063	Livingston	5.9	51,946	0.175	1,647
22065	Madison	9.1	24,485	0.288	210
22067	Morehouse	14.1	28,909	0.220	485
22069	Natchitoches	7.8	31,554	0.315	808
22071	Orleans	7.7	35,243	0.597	8,324

(Table continued)

22073	Ouachita	6.7	37,928	0.339	4,203
22075	Plaquemines	6.3	50,454	1.076	679
22077	Pointe Coupee	6.7	38,944	0.219	390
22079	Rapides	6.2	38,872	0.357	3,268
22081	Red River	9.2	30,285	0.308	135
22083	Richland	9.4	31,557	0.318	399
22085	Sabine	7.9	34,683	0.181	456
22087	St Bernard	6	36,660	0.332	627
22089	St Charles	6.3	56,869	0.647	946
22091	St Helena	10.5	32,014	0.213	117
22093	St James	9	46,774	0.397	309
22095	St John the Baptist	8.4	46,380	0.412	740
22097	St Landry	7.2	32,877	0.274	1,620
22099	St Martin	6.4	39,719	0.310	876
22101	St Mary	7.3	38,437	0.592	1,410
22103	St Tammany	5.3	58,868	0.263	5,856
22105	Tangipahoa	7.4	37,238	0.343	2,293
22107	Tensas	11.7	24,456	0.160	92
22109	Terrebonne	4.7	47,565	0.475	2,950
22111	Union	10.4	35,269	0.207	338
22113	Vermilion	6.4	38,872	0.214	1,037
22115	Vernon	6.5	42,322	0.218	685
22117	Washington	8.5	29,928	0.195	672
22119	Webster	8.4	34,342	0.288	816
22121	West Baton Rouge	6.5	45,167	0.765	499
22123	West Carroll	15.8	38,038	0.190	175
22125	West Feliciana	7.3	49,936	0.689	184
22127	Winn	8.1	32,505	0.408	329

Vita

Deepa Acharya was born and raised in Nepal. She spent a considerable part of her life in Chitwan before moving to Pokhara, to finish her schooling in Gandaki Higher Secondary Boarding School. She obtained Bachelor in Science in Agriculture from Institute of Agriculture and Animal Science (IAAS), Tribhuvan University in 2008. She worked at a non-governmental organization called LI-BIRD (Local Initiatives for Biodiversity, Research and Development) in Pokhara, a place in Western Nepal, for almost a year after the completion of her undergraduate studies. There, she travelled to various parts of Nepal and had an opportunity to work at grassroots level for the welfare of farmers. She joined the Agricultural Economics department of Louisiana State University for Masters degree in the spring of 2010. Her hobbies are travelling, painting and sports.