

# Mapping Invisible Threads: Unveiling Economic Landscapes through Micro-Level, High-Frequency 19th-Century Postal Data

Mengyue Zhao  
beczhaozmy0@gmail.com

## **Abstract**

This paper aims to measure the geographical extent of social, economic and political activities, that is, the geographic outline of markets in the United States from the year 1849 to the year 1871 by using 191, 928 records of town-level postal revenues. The geographical dimension of economic activities had been understudied for this period, even within the domain of Economic Geography. Economists usually think of economic activities as bounded by artificial borders such as the state borders or the county borders, or vague geographical features such as “the Mississippi Valley” or “the Great Plains”. In this paper, the geographic extent of markets is measured as the spatial clustering of high levels of human activities at a highly refined geographical scale and high frequency, rather a subset of artificial borders. In other words, the methodology proposed in this paper for the explanation of how the markets coalesced, that is, how they came together, and cohere, that is, how they stayed together overtime.

This paper presents a statistically defensible technique for using postal data for generating the geographical outlines of markets. Postal data is particularly useful as an input to that statistical process because it is micro-analytic, carefully created by the government agency, uniform across time and space, at least within the boundary of the United States, made available to be used in computer systems, and unexplored as an analytical tool. A bi-annual, national-scale postmaster compensation dataset is digitized, cleaned, geocoded and analysed in this paper. Postal revenue could be inferred from this postmaster compensation dataset because the Federal Government’s formula directly links local postmaster compensation to postal revenue. The entire dataset covers each town, city and populated place in the country that has a post office during the period from 1813 to 1881.

## Table of Contents

Abstract.....	2
<b>Chapter 1. Introduction</b> .....	<b>4</b>
Section 1.1 Thesis Introduction .....	4
Section 1.2 Literature Review .....	7
Section 1.3 Data source .....	9
<b>Chapter 2. Methodology</b> .....	<b>12</b>
Section 2.1 The Interpretation of Postal Activities.....	12
Section 2.1.1. The Composition of Postal Revenue.....	12
Section 2.1.2 The Relation of Historical Postal Activities to Social, Political and Economic Activities .....	13
Section 2.1.3 The Relation Of Postal Activities to Population.....	14
Section 2.1.4 Summary Statistics .....	19
Section 2.1.6 Changes In the Cost of Postage .....	20
Section 2.1.7 Rural Free Delivery, Urban Free Delivery, Private Express.....	21
Section 2.2 The Interpretation of Hotspot Analysis Maps.....	23
Section 2.2.1 The Interpretation of Clusters .....	25
Section 2.2.2 Geographic and Economic References .....	27
Section 2.3 Matching Errors .....	27
<b>Chapter 3. The Reports On Hotspot Analysis.</b> .....	<b>29</b>
Section 3.1 Southern States .....	29
Section 3.1.1 The Black Belt and Alabama, Mississippi, Georgia.....	29
Section 3.1.2 Report on the Hotspot Analysis for Alabama, Georgia, and Mississippi (1849- 1871) .....	31
Section 3.2 North-eastern States.....	41
Section 3.2.1 Maine .....	41
Section 3.2.2 New York.....	43
Section 3.2.3 Pennsylvania .....	46
Section 3.3 Midwestern States.....	49
Section 3.3.1 Wisconsin.....	49
Section 3.3.2 Illinois .....	50
Section 3.4 Western States .....	53
Section 3.4.1 California.....	53
Section 3.4.2 Oregon .....	55
<b>Chapter 4. Conclusion</b> .....	<b>56</b>
Appendix A. Results for Hotspot Analyses.....	58
Bibliography .....	136

## **Chapter 1. Introduction**

The goals of this dissertation are the following:

- to establish a methodology for the delineation the outline of markets in Nineteenth-Century America by using postal revenue as a proxy for social, economic and political activities;
- to establish the theoretical and the statistical foundations for using postal revenue as a proxy for human activities;
- to infer national-scale, bi-annual frequency postal data from postmaster compensation;
- to match the post offices in this dataset to their geographical location;
- to discover practical problems for the analysis of this large dataset and find out ways to address them;
- to visualize and analyse a selection of states with the proposed methodology, illustrate and discuss the usefulness and practicality of the proposed methodology through these examples.

This dissertation establishes a methodology for explaining how markets coalesced and how this coalescing persisted. It proposes a new way to delineate the extent of the market by using an unexplored bi-annual, national-scale dataset for postal revenues. The second chapter of this paper establishes the connections between postal activities and social, economic and political activities with careful reasoning and researches. The third chapter of this paper applies the proposed methodology on a selection of states. These examples demonstrate the usefulness and the practicality of this methodology, as well as some problems that might arise along with the use of this dataset and solutions to those problems.

### **Section 1.1 Thesis Introduction**

The purpose of this paper is to identify the various separate economic markets that were distinct from one another but exhibited uniformity within themselves with the use of micro-level postal revenue data. The idea is to identify markets not on the basis of artificial political boundaries or geographical features, but on the basis of postal activities. The underlying question that this paper addresses is “Where did everything happen?”. The foundation of this paper is to propose a rational argument for the measurement of the coalescing of markets and the duration of such coalescence for Nineteenth-Century America.

The definition of a market includes the following ideas: it is a place where buyers and sellers can efficiently interact. It is a place where the contract that they come up with could be insured, incorporated and interpreted in a safe and secure way. It is a place where the formation of prices is reasonably uniform within the area of the market and distinct from the kinds of prices and economic activities outside of the market. What we have said already implies that markets have spatial coordinates. This might not be entirely true. An investor in the United Kingdom could buy and sell stocks in New York. On the other hand, the buyers and sellers of an ice-cream have to be close to one another so that the ice cream does not melt before the buyers get to enjoy it. Besides, international income certainly generates a lot of activities that are localised. The stockbroker who sold New York stocks to London would purchase goods and services locally. The grocery store to which he goes, the school to which he sends his kids all generated revenues from the stockbroker, when in fact, a portion of his income comes from New York. It is traditional at least to assume, implicitly and explicitly, that markets have geographical boundaries, at least of some rough kind. But traditionally, when those geographical boundaries have been discussed and described, they turn out to be boundaries established by convention, political accident, or arbitrary convenience.

This paper is interested in understanding whether and how economic, social and political activities are concentrated in one industry or product area and the change of those patterns over time. One underlying assumption for geographic analysis is that transportation cost matters for the Nineteenth-Century U.S. economy. Economic activities concentrated in one area because of locational benefit, geographic constraints, information costs, as well as some political reasons. Hotspot analysis

is performed both at the state to illustrate the formation of clusters of higher-than-average economic activities and the transitions of these clusters overtime. National-scale, local level, biennial postal revenue data allows for a refreshed understanding of economic, social, and political activities at a detailed geographic scale which is not available in the census.

Postal activities are roughly associated with these economic, social and political activities, and above all, information transmission. We cannot directly observe the decision made by a shoemaker to send out advertising materials, to mail his wages to a salaried worker or invite customers to a special opening. Economic, social and political activities that are associated with these activities gave rise to postal activities. Postal activities are the shining light behind a curtain that tells us that there is something going on in one specific area. The edges of the markets are the edges of the hot clusters. These markets are interesting because they are places within which clustering of activities existed and was confined.

Again, the purpose of this paper is to describe markets in a way that is not dependent on arbitrary political or geographical lines. Rivers, mountains, county lines, state lines or city limits do not constitute markets. They are referred to as referential features that give geographical coordinates a context. These lines are secondary to the location of post offices. Hot clusters might incidentally coincide with county edges, but the extent of markets is defined by the clustering of distinctively high postal activities.

This paper presents a statistically defensible technique for using postal data for generating the geographical outlines of markets. Postal data is particularly useful as an input to that statistical process because it is micro-analytic, carefully created by the government agency, uniform across time and space, at least within the boundary of the United States, made available to be used in computer systems, and unexplored as an analytical tool. The contribution of the paper will be to bring to the attention of the scholars of social sciences and history, to whom the geographical spaces were relevant to social science -type analytical missions. In this paper, geographic areas are defined in terms of hotspots and cold spots, which means that these are markets or neighbourhood where economic activity is distinctively more intense or less intense than it is in general in each of the states.

There is a general reason for using a clustering technique. First, this paper looks for a continuous geographical area that had more intensive or less intensive activities than the rest of the geographical areas. Hotspots analysis makes sure that each hot or cold spot that is identified is surrounded by similar spots in its proximity. Practically, absolute postal revenue, when illustrated on a map, do not work very well in summarizing continuous areas because one area often includes both post offices with seemingly large revenues and post offices with seemingly small revenues.

Economists have been talking loosely and easily about markets. This introduction has imperfectly described why it is that we even think that markets have a geographical existence. But in a general way, the idea of market borders is very variable with the type of products. However, most of the activities went on in the New York markets are traditionally thought of as going on along the city line. This paper goes beyond the city line and investigates the neighbourhoods of cities.

## **Section 1.2 Literature Review**

This paper attempts to measure the geographic extent of markets and human activities in Nineteenth-Century America with a spatial clustering technique. Such topics remain understudied in the literature of economic geography and of economics in general.

Theoretical innovations that paid attention to the geographical dimension of economic activities include Marshall's external returns to localization of industrial activities (Marshall 2009), Von Thünen's Isolated State (Thünen 1966), Walter Christaller's Central Place Theorem and Krugman's New Economic Geography (Krugman 1991). Douglas North's Location Theory examines regional growth in Nineteenth-Century America (North 1964). These ideas provide theoretical frameworks for understanding the spatial clustering of economic activities, but they do not examine or conceptualize the actual geographic context of human activities.

One of the most influential papers that concerns similar questions as this paper is "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U. S. Regional Manufacturing Structure, 1860–1987" by Sukkoo Kim. (Kim 1995) Kim examines the Heckscher-

Ohlin model, internal and external economies of scales by using indexes of regional specialization and county-level manufacturing, services and agricultural data. Price convergence among regions is another indicator that such literature concerns. A lot of papers that pay attention to the geographical dimension of economic activities take similar approaches. Rosenbloom (2009), for example, studied the geographic integration of labour markets in the late Nineteenth-Century U.S. by measuring price convergence. In this dissertation, there is no pre-determined economic model to be tested. The geographical details of the postal revenue dataset are maintained as much as possible. This paper pays attention to the one question – “Where did everything happen?”

Again, understanding the geographical context of historical economic activities does not seem to be the priority of the literature of Economic Geography. Papers that do consider the geographical extent and context of economic activities of Nineteenth-Century America are mostly industry-specific studies. Cozen and Cozen, for example, studied the spatial distribution of retailers in Milwaukee. (Cozen and Cozen 1979). Lamoreaux and Sokoloff measured the spatial clustering of innovation for glass industries in Nineteenth-Century America (Lamoreaux & Sokoloff 1997). Foust (1967) studied the westward movement of cotton production by studying the census records of Yoeman farmers.

The most popular source of microdata for studying Nineteenth-Century U.S. is the decadal Census Records, which was made accessible online by researchers at Minnesota Population Centre. (Fitch& Ruggles 2003; Ruggles 2014). The United States had nationwide decadal census records since the year 1790. Full individual census records are available in machine-readable form for the years 1850 and 1880. The decadal census records for other years between 1850 and 1910 are available in 1% samples. The National Historical Geographical Information System provides decadal census records aggregated at the county level, including total manufacturing output and total agricultural output for the years after 1850. These two data sources are useful for population research purposes and for understanding economic activities at an aggregated level, however, they could not be used to measure the extent of economic, social and political activities at a refined geographic scale and high frequency.



## Section 1.3 Data source

### General Information on the Data source

This paper uses biennial, nation-wide and town-level postal revenue data as a proxy for social, economic and political activities as well as communication flows. For the northeastern states, biennial data between the years 1849 and 1871 has been digitized. For the rest of the country, the dataset for every four years between 1849 and 1871, plus the year 1869, has been digitized by the author. This dissertation is the first study that infers postal revenue from postmaster compensation, geocodes the postal revenues, and conducted statistical analysis on town-level, national-scale postal revenue for the Nineteenth Century U.S.

The postal revenue data is calculated from the annual postmaster compensation data from the Official Register of the United States (USCSC & USBC 1849-1871). Since the federal government's formulas link the postal revenue of each post office directly to its postmaster's compensation, postal revenue is calculated directly and uniformly throughout the country. According to the Congressional Records, the postmaster takes a percentage of "the moneys arising from the postage of letters, newspapers[,] and packets..." (13 STAT. 1813, P.220). All formulas that are used in this paper are found in the Congressional Records. They were discovered and initially compiled by Professor Terence Hines, a dedicated philatelist, at Pace University. Velk (2006) and Hines (2006) first discovered and presented this usage of the federal government's formula at the Smithsonian Museum's symposium on postal history. Hines & Velk did a statistical analysis of the postmaster compensation for New Hampshire. The author of this dissertation learnt about the usefulness of these formulas while working with Professor Velk. No study before this dissertation attempted to digitize the full records of postmaster compensation, infer postal revenues from it, geocode the dataset and study it systematically with statistical techniques. No study before this dissertation pinned down the specific relationships between postal revenue and social, economic and political activities with careful reasoning.

The dataset covers every town in the nation that had a post office. The entire dataset spans the period of 1813-1881. In this paper, the author digitizes, cleans and geocodes the thirteen records of

Official Register for the years from 1847-1871. The federal government’s formula changed 2 times during this time period. The formulas for postmaster compensation for the periods of 1847-1851, 1853, 1855-1871 are respectively found in the records of 29<sup>th</sup> and (the Act of March 3 and the Act of June 22 of) the 32<sup>nd</sup> Congress. 191891 records of post offices have been digitized, cleaned and analyzed in this paper.

Table 3.1 A Typical Row From the Official Register

Name of the Post Office	County	State	First and Last Name of the Postmaster	Annual Postmaster compensation (\$)
Clingman	Wikes	North Carolina	J.C.Greene	100.54

The original postmaster compensation records are excerpted and sent to professional typists for mechanical digitization under the instruction of the author. The dataset is then processed by the author according to the following steps: All records have been cleaned by using computer algorithm and then linked to the historical and contemporary place names and geographical coordinates from the U.S. Geological Survey (United States Congress n.d.) by using fuzzy matching algorithms in the programming language Python. <sup>1</sup> 60% of the post office records were successfully matched with their geographical location. The counties where each of the post offices were located are matched separately with the US Census data and geocodes from the National Historical Geographic Information System (NHGIS) (Manson et al. 2018) by the author.

#### The Federal Government’s Formulas

According to the Twenty-Ninth Congressional Record, from the year 1847 to the year 1851, a postmaster takes 40 per cent on an amount not exceeding 100 dollars, 33.3 percent on an amount over 100 dollars and not exceeding 400 dollars, 30 percent on an amount over 400 dollars but not

---

<sup>1</sup> The scripts used in this paper for processing the dataset are based on a fuzzy word matching algorithm called FuzzyWuzzy, which essentially calculates a percentage ratio for linguistic distances between words. (The documentation for this algorithm is available from : [http://anhaidgroup.github.io/py\\_stringmatching/v0.4.x/TokenSort.html](http://anhaidgroup.github.io/py_stringmatching/v0.4.x/TokenSort.html).)

exceeding 2400 dollars and 12.5% on an amount over 2400 dollars. (29 STAT. 1847, p.148)

According to the Thirty- Second Congressional Record (Act of March 1), for the year 1853, a postmaster takes 50 per cent on an amount not exceeding 100 dollars, 40 per cent on an amount over 100 dollars and not exceeding 400 dollars, 35 percent on an amount over 400 dollars but not exceeding 2400 dollars and 15% on an amount over 2400 dollars. (32 STAT. 1852, p.182)

According to the Thirty- Second Congressional Record (Act of June 22), from the year 1855 to the year 1871, a postmaster takes 60 per cent on an amount not exceeding 100 dollars, 50 per cent on an amount over 100 dollars and not exceeding 400 dollars, 40 percent on an amount over 400 dollars but not exceeding 2400 dollars and 15% on an amount over 2400 dollars. (32 STAT. 1852, p.255)

Box rent was included in the calculation of postmaster compensation. Box rent constituted less than 4% of the annual postal revenue in the year 1882, according to the Postmaster General's Report (The Federal Government of the United States 1882; Zhao 2018). Box rent is considered as negligible in this paper.

The calculations below are an example for the years 1849-1851 based on the formula found in the Twenty-Ninth Congressional Record:

$$\textit{Postmaster Compensation} = a\% \times \textit{Annual Postal Revenue} + \textit{Box rents}$$

$$\textit{Where } a = 40 \textit{ when } \$0 < \textit{Annual Postal Revenue} < \textit{or} = \$100,$$

$$a = 33.3 \textit{ when } \$100 < \textit{Annual Postal Revenue} < \textit{or} = \$400,$$

$$a = 30 \textit{ when } \$400 < \textit{Annual Postal Revenue} < \textit{or} = \$2400,$$

$$a = 12.5 \textit{ when } \textit{Annual Revenue} > \$2400$$

The author maps postmaster compensation data to postal revenue according to this formula:

$$\textit{If } m \leq 0.4 \times 100,$$

$$\textit{then } PR = \frac{m}{0.4};$$

If  $100 \times 0.4 < m \leq 100 \times 0.4 + (400 - 100) \times 0.333$ ,

$$\text{then } PR = 100 + \frac{m - (100 \times 0.4)}{0.333};$$

If  $100 \times 0.4 + (400 - 100) \times 0.333 < m$   
 $\leq 100 \times 0.4 + (400 - 100) \times 0.333 + (2400 - 400) \times 0.3$ ,

$$\text{then } PR = 400 + \frac{m - (400 - 100) \times 0.333 - 100 \times 0.4}{0.3};$$

If  $m > 100 \times 0.4 + (400 - 100) \times 0.333 + (2400 - 400) \times 0.3$ ,

$$\text{then } PR = 2400 + \frac{m - (2400 - 400) \times 0.3 - (400 - 100) \times 0.333 - 100 \times 0.4}{0.125}$$

where  $m$  = annual postmaster compensation,

$PR$  = inferred postal revenue

## Chapter 2. Methodology

### Section 2.1 The Interpretation of Postal Activities

#### Section 2.1.1. The Composition of Postal Revenue

As discussed in section 1.3, the postal revenue that is derived from postmaster compensation includes all receipts of the post offices that are associated with the postage of letters and newspapers. This dataset does not neglect any significant source of postal business. The main sources for the gross revenue for the postal department are “letter postage, including foreign postage, and stamps sold”, “newspaper and pamphlet postage”, which constitutes 96% of the post department’s gross income in the year 1850, and 92% of the gross revenue in the year 1871. (“Postmaster General’s Annual

Report” 1850, p. 404). And 92% for the year 1871(“Postmaster General’s Annual Report” 1871, p. 48). The rest of the postal department’s gross revenue was not directly associated with the postal business. The federal government stipulated that letters could be only posted with a stamp on it (Priest 1975, p.79). The revenues from the postage of letters included the sales of the stamp that was attached to each of those letters.

#### Section 2.1.2 The Relation of Historical Postal Activities to Social, Political and Economic Activities

Postmasters received a percentage of postal revenues which arose from the postage of letters and newspapers, not from the postage of goods. The U.S. post system did not deliver packages before 1913. The United States Post Office Department did not establish parcel delivery service, which was called Parcel Post, before 1913. (Kielbowicz 1994) The Thirteenth Congress stipulated that the postmaster took a percentage of “the moneys arising from the postage of letters, newspapers and packets...”. (13 STAT. 1813, P.220) This law was enacted until the Pendleton Civil Service Reform Act was passed in 1882. The legal use of the word “packages” refers to “letters composed of several sheets or to packages containing a number of letters”. The monopoly statues that prohibited the delivery of letters and packets by private expresses did not prohibit the delivery of packages containing merchandises and personal items that were not letters (Johnston 1968, p. 389). Problems regarding the competitions by private expresses is discussed in section 2.1.7.

Postal revenue, therefore, is a direct account for general information transmission. To the extent that information transmission is an integral part of all social, economic and political activities, postal revenue is an account for those activities. A typical merchant in the early Nineteenth Century, for example, needed to use the federal postal system to send his catalogues to his targeted consumers. Any bills, receipts, or letters concerning the details of a particular contract also had to be transmitted through the postal system. Postal revenue also accounts for political activities. Political campaigns in the Nineteenth Century relied heavily on newspapers. Without television and electronic media, newspapers were one of the few sources of political information for the public. Baldasty remarks that third party campaigns such as the Anti-Masonry campaigns in the 1820s, short of other channels to

communicate with the public, relied entirely on newspapers to reach the voters. (Baldasty 1992, p. 15). The postage of newspapers was monopolized by the federal postal system in the Nineteenth Century (Priest 1975, p. 45, 67-68). The relation of postal activities to social activities is self-evident. Any social communication in distance required the postage of letter. However, writing letters requires literacy, therefore, marginalized populations such as women and African Americans that had low literacy due to the lack of access to education is not included in this dataset. This reflects the unfortunate history that minorities were to some extent isolated from the general social, political, and economic activities in the Nineteenth Century.

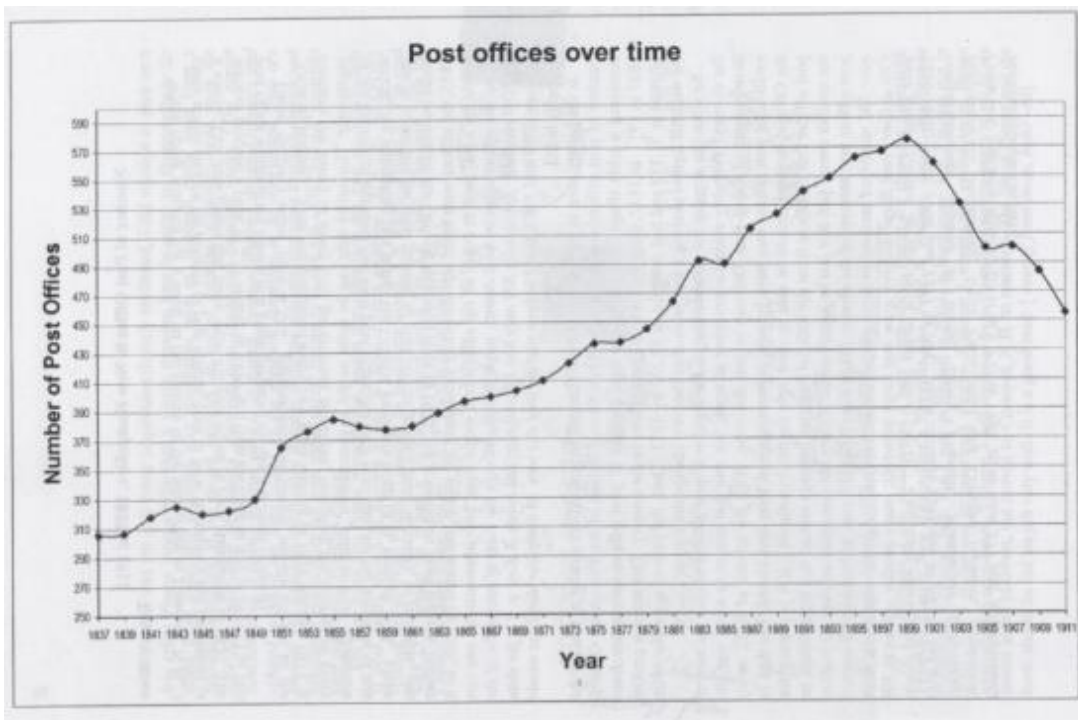
### Section 2.1.3 The Relation Of Postal Activities to Population

This section ponders the question of whether postal activities should be considered as an indicator for population or as an indicator for general economic, social, economic and political activities. In this paper, postal activities are considered as a proxy for human activities, including social, economic and political activities, and most directly, the transmission of information. The “settlement” of post offices refers to the emergence of each dot on the graph. A significant population usually coexist in a location with significant social, economic and political activities. However, the decision to open up a new post office in one area must be based on the combination of the accessibility of transportation, population and political and economic influence of that area. Therefore, the settlement of post offices, as well as the level of postal activities, are good measurements for the reach and the level of human activities in one area, but not as good measurements for population per se.

Hines’ hypothesis is that depopulation is not always associated with the closure of post offices. He supports this hypothesis by using the data for the number of post offices established in New Hampshire in the Nineteenth Century (See Figure 1.). It is a common sense that westward expansion happened in the early Nineteenth Century, yet the number of post offices peaked in the year 1899 in New Hampshire. Such patterns of increase in the number of post offices were not unusual in other New England states. (Hines 2005, p. 7-8) Besides, the general postal volume grew sixteen times

more rapidly than the U.S. population after 1850 because of its falling real cost. (“America's Mailing Industry.” n.d.)

Figure 1. The Number of Post Offices Over Time in New Hampshire, 1837-1911. From *Number of post offices in New Hampshire from 1837 to 1911* by T. Hines (2005).



Although there is no direct proof for the disassociation between the timing of depopulation and the decline of post offices for the entire country, two speculations could still be made from the history of the population and the postal system of New Hampshire -- First, the number of post offices is not a good proxy for population due to the mismatch in the timing of the decline in the number of post offices and the timing of population changes. Second, postal activities are potentially a better proxy for general economic, social and political activities, as well as information transmission, than for population per se.

Table 1.1 shows the result of regressing county-level postal revenue on county-level economic indicators from the Census Records for the years 1850, 1860 and 1870. These economic

indicators include manufacturing output, farm output and total output per capita, which is calculated from the sum of manufacturing output and farm output divided by the population. In the regressions, the postal revenue data for the year 1861 and the year 1871 has been paired with the census records for the year 1860 and the year 1870 respectively. Since this paper only collected data for the north-eastern states for the year 1851, the postal revenue of the year 1849, rather than the year 1851, is paired with the 1850 census records. State fixed effect and year fixed effect has been applied to all regressions. Removing the fixed effects did not change the result.

The result of the regressions shows that postal revenue is statistically significantly correlated with farm output, manufacturing output, total output per capita, as well as population (p-value <0.05 for each of regressions 1-4). The magnitude of this correlation is small. An increase in one-dollar worth of manufacturing output or farm output is associated with less than one penny increase in postal revenue. Each of the four indicators is associated with about 2% of the variations in postal revenue, as indicated by the R-squared value for regression 1-4. Small magnitudes and small R-squared values suggest that postal revenue is associated with economic, social and political activities that the manufacturing output, farm output and population data from the Census Records does not account for.



This result supports the hypothesis that postal revenue is a better indicator for general human activities, which includes economic, social and political activities as well as information transmission, rather than population per se.

Although the absolute value for the coefficient for farm output (-98.04) per capita is much larger in comparison to that for farm output (0.002), the coefficient is negative. Besides that, as shown in regression 6 in Table 1.1, regressing farm output controlling for population gives an insignificant negative coefficient for farm output. These two results suggest that postal revenue is negatively correlated with farm output when controlling for population or when using the per capita value for farm output.

The purpose of this regression practice is to look for correlation, not causality. It shows that postal revenue is sensitive to agricultural and industrial activities, as well as population. It also shows that per capita manufacturing output and per capita, total output has a much larger magnitude in their correlations with postal revenues. This finding implies that postal revenue picks up many activities that are independent of the population. Population and general human activities, of course, are not mutually exclusive. Where there is a significant population tends to be the location of significant human activities, but the location of the population does not necessarily change with all significant changes in the spatial clustering of human activities.

Table 1.1 Regression of Postal Revenue on Manufacturing Output, Farm Output, Population and Total Output Per Capita

	Dependent Variables					
	Revenue					
	(1)	(2)	(3)	(4)	(5)	(6)
Manufacturing Output	0.001*** (0.00004)				0.0002* * (0.0001)	
Farm Output		0.002*** (0.0003)				-0.0003 (0.0003)
Population			0.20*** (0.010)		0.16*** (0.021)	0.204*** (0.011)
Total Output				114.71		

Per Capita ****				*** (6.82)		
Control For Population	No	No	NO	NO	YES	YES
State Fixed Effect	YES	YES	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES	YES	YES
Observatio ns	70,504	71,741	71,863	70,442	70,502	71,714
R2	0.021	0.018	0.022	0.020	0.021	0.023
Residual Standard Error	89,207.98 (df = 70455)	88,605.83 (df = 71691)	88,370.25 (df = 71813)	89,222.83 (df = 70393)	89,171.66 (df = 70452)	88,408.89 (df = 71690)

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

\*\*\*\* Total Output Per Capita = (Manufacturing Output + Farm Output)/ Population

Table 1.2 Regression of Postal Revenue on Manufacturing Output, Farm Output, Population and Total Output Per Capita

	Dependent Variables	
	Revenue	
	(7)	(8)
Manufacturing Output Per capita****	148.74*** (6.886)	
Farm Output Per capita****		-98.04*** (12.129)
Control For Population	No	No
State Fixed Effect	YES	YES
Year Fixed Effect	YES	YES
Observations	70,502	71,741
R2	0.023	0.018
Residual Standard Error	89,106.07 (df = 70455)	88,594.23(d f = 71691)

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

\*\*\*\* Manufacturing Output Per Capita = Manufacturing Output/ Population; Farming Output Per Capita = Manufacturing Output/ Population

### Section 2.1.4 Summary Statistics

There are 191,891 records of post offices in this dataset in total for the period of 1849-1871. The summary statistics in table 2.1 shows that the postal revenue dataset is largely right-skewed. The mean postal revenue is 16755.52 dollars, 13 times larger than the median. In other words, 50% of the post offices had postal revenues that were at least 13 times smaller than the national average. After removing the top 3% records of postal revenues, the national average postal revenue decreased to 4163. The summary statistics show that the mean of the dataset is dominated by the top 3% records of postal revenues. It implies that the majority of this dataset covers locations with small and medium level postal revenues. Only a small portion of this dataset consists of large cities with very large postal revenues.

Table 2.1 Summary Statistics- the Entire Dataset

Number of Records	Mean	Minimum	Maximum	Standard Deviation
191,891	16755.52	0	1146151	77485.72

25 <sup>th</sup> percentile	Median (50 <sup>th</sup> percentile )	75 <sup>th</sup> percentile	90 <sup>th</sup> percentile	99 <sup>th</sup> percentile
424	1266.00	4169	15563.00	66426.80

### Section 2.1.6 Changes In the Cost of Postage

This section evaluates the potential impact of change in the cost of postage on the analysis conducted in this paper. Table 3 shows that the postal fees changed three times during the period examined in this paper. The cost of postage became cheaper both in general and for long-distance services. Before 1863, the cost of postage to locations more than 300 or 3000 miles away was higher than that to closer locations. After 1863, the cost of postage was not differentiated by distance.

Did postal revenue become higher in one area simply because the cost of postage is lower, not because there were more human activities going on? Reductions in the cost of postal services that were uniform across the country made the cost of postage lower in all areas. Hotspot analysis compares the average postal revenues between neighbourhoods and the average postal revenue of the rest of the state for each year. It essentially looks for neighbourhoods with higher-than-average postal revenues for each year. Uniform reduction across time should not impact this analysis because it examines the postal revenues for each year in relative terms.

Reductions in the cost of postal services that were disproportional for longer distances made long-distance communication less costly. The ease of long-distance communication makes certain social, political and economic activities that required such communication less costly. Is there geographical inequality in the effect of these reductions? Potentially, yes. There could be, for example, an urban-rural divide in the excess demand for long-distance postal services. The cost of long-distance postage might be prohibitive to rural, low-income areas. More excess demand would be satisfied in low-income areas after the reduction of postage. For instance, a hot cluster might suddenly appear in rural, low-income areas on the map for the year 1863 because the cost of postage to locations 3000 miles away has decreased from 10 pence to 3 pence. A farmer on holiday in after the year 1863 would like to send a post guard faraway back home describing a day's activities, whereas his grandfather in 1810 might not send a letter to relatives ever except Christmas time. But hot clusters do not suddenly appear in 1863 in urban areas for the same reason because urban citizens could afford the cost of 10 pence.

The income-demand of postal services were altered in significant ways by the reduction of the cost of postage, and to the extent that there is a geographical difference in the pattern of income, such alteration has geographical implication as well. Does such alteration create bias in the analysis of this paper? Not necessarily. The cost of postal services is an integral part of the cost of human activities. If the postal services were cheaper, there will be more economic, social and political activities going on. If a hot cluster suddenly appears on the map in a rural area because of nothing but the cheaper cost of postage, it implies that more social, economic and political activities were enabled in that area.

Table 3. The Changes in Postage Fees. From Rates for Domestic Letters, 1792-1863 and Rates for Domestic Letters Since 1863 by United State Postal Service. (2008, 2019).

Date	Distance	Fees (pence)
1845	not over 300 miles	5
	over 300 miles	10
1851	not over 3000 miles	3 if prepaid, 5 if not prepaid
	over 3000 miles	6 if prepaid, 10 if not prepaid
1855	not over 3000 miles	3 (prepayment required)
	over 3000 miles	10 (prepayment required)
1863	all distance	3

#### Section 2.1.7 Rural Free Delivery, Urban Free Delivery, Private Express

This section discusses three sources of mailing services that would potentially impact postal businesses, including Rural Free Delivery, Urban Free Delivery, and Private Express Services. The

finding of this paper is that competitions to postal service were minor and free delivery services were not important during the period of 1849-1871 which is investigated in this paper.

With the Rural Free Delivery System and urban free delivery, urban citizens and rural citizens no longer had to travel to the post offices to collect their mails. The Rural Free Delivery System (RFD) was created by the Federal Government in the 1890s and was not in operation before then. (Kernell & McDonald 1999, p797). Urban free delivery was enacted before the RFD. In the year 1863, the federal government granted free delivery in cities that had a population greater than 50,000, which resulted in 66 cities which were entitled to free urban delivery service (Henkin 2008, p. 90; Hecker 2018, p. 16-17). Free delivery services reduced the travelling cost of the collection of mails for the citizens. Therefore, it should have encouraged the utilization of postal services in those big cities.

The increase in postal business due to urban free delivery is not reflected in this paper's dataset because there were more than 66 post offices of which the postmaster was receiving the maximum compensation in each year of the dataset. But the impact of urban free delivery on this paper's dataset is minor. First, only 66 post offices in large cities were affected. Second, large cities are not the focus of this paper. This paper concerns **neighbourhoods** of towns, cities, and places that had postal revenues that were higher than the state or average and uses their locations to delineate the outlines of markets. Fluctuations in the postal revenue for big cities do not create an impact on this practice. The effect of maximum postmaster compensation on the inferred postal revenue of big cities is discussed extensively in the example of Philadelphia, Pennsylvania in Section 3.2.3, the example of Chicago, Illinois in Section 3.3.2, and the example of San Francisco, California in Section 3.4.1, as well as in the conclusion.

Private express was not an important source of competition for postal service because of the strengthening of the Monopoly Statues by Congress. The monopoly of the postal service by the federal government coastline part of the U.S. Constitution and had its roots from the Sixteenth Century. In early Nineteenth Century, the monopoly status of the federal postal system suffered from the competition of private expresses, mostly in the eastern part of the country, due partly to the ambiguity of the provision by the Constitution. The congressional statue of 1845, however, reversed

this trend by prohibiting the delivery of all letters and packets by private expresses. There were violations of the 1845 Statue, but they were only incidental. Besides, the exemption of this statue was minimal. Exemptions were allowed, for example, for “carriage by special messenger employed by a particular occasion only”. (Priest 1975, p. 45, 67-68)

## **Section 2.2 The Interpretation of Hotspot Analysis Maps**

Hotspot Analysis looks for spatial clusters of high or low values. For each post office, it calculates whether it was located in a neighbourhood of post offices that had statistically significant higher or lower postal revenues than all post offices outside of this neighbourhood. The distance of a neighbourhood is selected as the minimum distance such that each post office in the neighbourhood had at least one neighbouring post office. A statistical cluster of high postal revenues (hot cluster) is indicated by red, orange and yellow colours and the statistical significance indicated by these colours decreases in this order. A statistical cluster of low postal revenues (cold cluster) is indicated by the colours blue, light blue and green. Small grey dots indicate post offices that do not form clusters of either high or low values. (“How Hotspot analysis Works” n.d.)

The “neighbourhood” in hotspot analysis is defined by an input distance. In this paper, this distance is selected arbitrarily to be 30km. There is an economic explanation to a neighbourhood of 30km – it is a distance that a person could walk in a day at a reasonable pace. Since people living in a neighbourhood defined by a distance of 30km could easily be physically present to each other, it is more likely that people in this neighbourhood are bounded with extensive social, economic and political connections, which are all relevant to the delineation of a contiguous market.

Figure 2 illustrates the hotspot analysis with the example of Utica. The bright red dot at the centre of the map in Figure 2 is Utica, New York. When a hotspot analysis is conducted on all towns in New York, it selects all towns within a sphere of which Utica is the geographical centre, and of which the diameter is 30km. This selection of towns is highlighted in the graph in bright blue. The average postal revenue of all highlighted dots, including Utica, are then compared to the average postal revenue of all non-highlighted dots in New York. If the average revenue of Utica’s neighbourhood is statistically higher than the rest of the state, then Utica will be coloured red, orange,

or yellow, according to the level of statistical significance. In the hotspot analysis for New York, this process is repeated for all dots of post offices that existed in New York in a year.

Figure 2.1 Hotspot Analysis Explained – the Example of Utica, New York (1853)

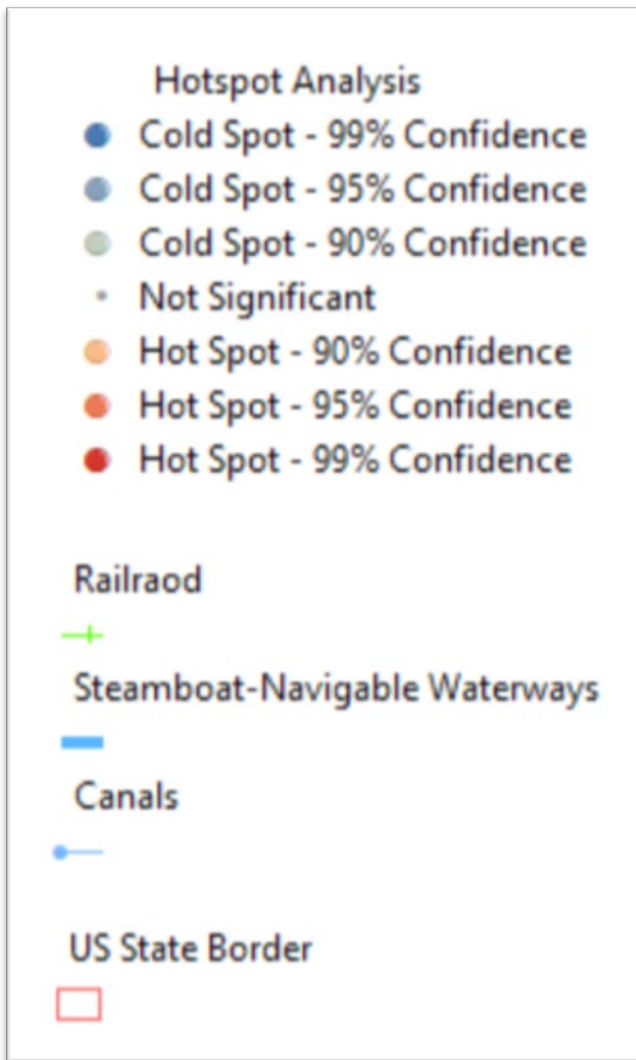


Figure 2.2 Hotspot Analysis Explained – the Result for New York State (1853)





Figure A0. Legend for All Maps In this Dissertation



### Section 2.2.1 The Interpretation of Clusters

The shape of economic clusters could be roughly categorized into three categories: band-shaped, circle-shaped, and single points. The extent and density of the hot and cold clusters are illustrated in a uniform manner in this paper by noting three parameters in a bracket. The parameter “d” represents the distance. The distance parameter measures the extent of a cluster. The distance of a cluster is measured by the distance from one extreme of the cluster to another extreme of it. The parameter “n” measures the density of a cluster. The density of a cluster is represented by the number

of post offices that are included in a cluster. The last parameter has four digits, it represents the year for which the “n” parameter and the “d” parameter measures.

For example, “(d = 50km, n = 40, 1867)” represents a hot or cold cluster that has 40 post offices, and a distance of 50km between its two extremes in the year 1867. If the shape is important for the purpose of analysis it will be either mentioned in the form of “(d = 50km, n= 50, 1867, band-shape)” or mentioned directly in the sentences. The parameters provide a succinct way to summarize and address the extent and the intensity of a cluster in a particular year. For simplicity purpose, if a cluster contains a contemporary county seat or a contemporary large state, this cluster is named after the name of the city.

This paper conducts a state-level hotspot analysis. The sample for the hotspot analysis presented on each map is all post offices in the state. The result of the hotspot analysis for each state is then merged together visually to see whether the hot clusters of neighbouring states bear any proximity. Although the maps for the hotspot analysis for a particular state includes the result for other states visually, the analysis is conducted separately for each state.

The result of hotspot analysis is relativistic, it shows whether there is spatial clustering of statistically higher or lower values than the rest of the state. Hotspot analysis reflects whether economic activities are concentrated in one industry or product area and how that concentration changes over time. One underlying assumption for geographic analysis is that transportation cost matters for the Nineteenth-Century U.S. economy. Economic activities concentrated in one area because of locational benefit, geographic constraints, as well as some political reasons.

Hotspot analysis gives more emphasis to neighbourhoods of towns, populated places, and cities, rather than one or two large cities which only constitute the top 3% of the dataset. It searches for contiguous neighbourhoods that had distinctively higher or lower human activities than locations outside of those neighbourhoods. Single post offices that had very high postal revenues but were surrounded by post offices with very low postal activities are unlikely to be considered to be a hotspot since despite there were substantial human activities going on in that city, the neighbourhood mean

was very low. We already have a substantial amount of information on the history of big cities such as Philadelphia, Chicago and Boston. Rather than telling more stories about big cities, this paper seeks to measure and to illustrate how and where markets coalesced across neighbourhoods of cities.

### Section 2.2.2 Geographic and Economic References

The geographic references in this paper consist of transportation references, state and country borders, topography and county-level economic activities. The hotspot maps are overlaid with the maps for railroad, canals and steamboat-navigable rivers. The development of these transportation methods is traced geographically at biannual frequency. The GIS datasets for the transportation system were created by Jeremy Atack, a Professor of Economics at Vanderbilt University (Atack 2015, 2016, 2017). The topographic maps (Esri 2019) are contemporary, and it might not describe the Nineteenth-Century topography with complete accuracy. However, geographic features such as coastlines and mountains which are important to this paper are relatively stable over time. Besides that, rivers that are likely to have changed course in the time span of 200 years. To the author's knowledge, there is no reliable digital source that describes the historical location of rivers. Therefore, this paper gives more emphasis to steamboat navigable rivers. County-level manufacturing and agricultural output data are from the decadal U.S. census records. These economic data are represented graphically on decadal county maps and compared to the hotspot maps.

## Section 2.3 Matching Errors

60% of the post offices have been matched to their geographic location in this research. 40% of the records are not matched. There are two general sources of matching errors—First, the typos that cannot be fixed by computer algorithms. Second, historical places that neither exist anymore nor recorded by the federal government. Typos are randomly distributed. The second source of unmatched records could either be unimportant places out of where people eventually or places that were initially very important but diminished and forgotten completely in a period of 200 years.

Since the number of unmatched records is relatively stable over time if a cluster has been identified correctly, the comparisons of statistical hot and cold clusters overtime are accurate. Besides that, if unmatched records had even geographic distribution, the identification of clusters are more accurate if no evidence suggests that the unmatched data are substantially different across space. These criteria for accuracy are matched unevenly across the country. Since the county that a post office belongs to is rarely unmatched, the number of post offices that are not matched within each county for each year is counted and illustrated on a map of county borders. This map is overlaid with the hotspot analysis. Figure 2.3 shows an example of such overlaid maps. This practice is conducted throughout the paper to verify whether some unexpected patterns are the result of unmatched records.

Figure 2.3 The Number of Matched Records for Each County in New York (1853)



# Chapter 3. The Reports On Hotspot Analysis

## Section 3.1 Southern States

### Section 3.1.1 The Black Belt and Alabama, Mississippi, Georgia

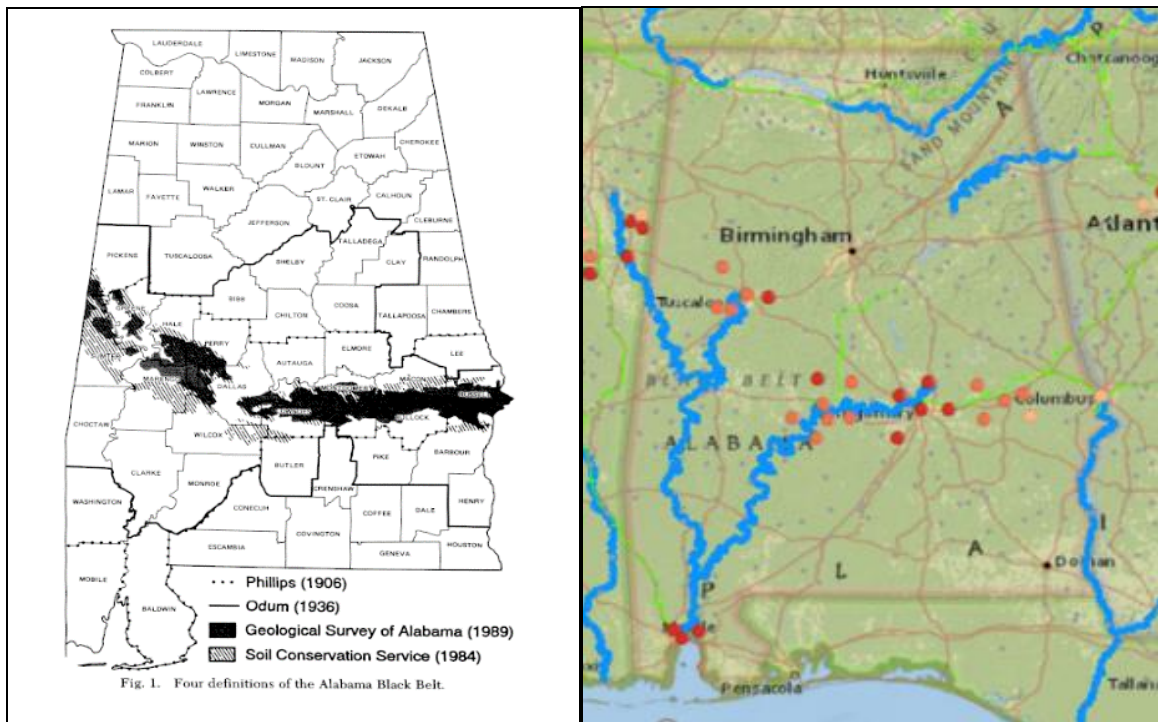


Figure 3.1 and Figure 3.2. The Black Belt in Alabama In the Nineteenth Century

Right Graph (Figure 3.2): The Delineation of The Outline of the “Black Belt” by Town-Level Postal Activities for the Nineteenth Century

Left Graph (Figure 3.1): The Delineations of the Outline of the “Black Belt” by Historical Soil Quality or County-Level Racial Composition. From *On Defining the Alabama Black Belt: Historical Changes and Variations*, recreated by Webster & Samson (1992, p. 165)

The Methodology proposed in this paper for the delineation of the market with the use of postal data prevents vague generality, imprecision and incompleteness of the conventional definitions of the geographical outline of the Black Belt by social sciences. This paper is interested in the question of how areas coalesced and how those coalescing persisted or not persisted. The arbitrary

definition of state lines, county lines do not tell a story of how cities, towns and places become contiguous economic, social and political areas which can be called market areas.

The geological definition of the term “Black Belt” describes “a long crescent of dark, fertile soils that stretches from eastern Alabama to Northern Mississippi” (Barone 2005, p. 170). Since the early 20<sup>th</sup> Century, it has been frequently referred to by scholars as a socio-economic concept that describes an extensive neighbourhood of the disadvantaged population of African Americans. Webster & Samson (1992) remark this confusion in social sciences between the ecological definition of “Black Belt” and its identity implications in social sciences. The Black Belt is well-known as the “Cotton Kingdom” due to the history of the vigorous growth of the slavery-based cotton plantation in the early Nineteenth Century. (Cleland 1920) Researchers of the socio-economic status of the Black Belt population, such as Falk & Rankin (1992) and Webster & Samson (1992), although frequently referring to the history of cotton plantation and slavery, seldom establish a direct connection between the present and the past. For the history of the Black Belt, the question of “Where did everything happen?” thus essentially became an inseparable part of the question of “Who we are?” or “How we came to be?”.

A research finds out that the contemporary population within the Black Belt consists of a group of people which had socio-economic and demographic characteristics that were statistically distinct from people outside of the Black Belt region—they were African Americans who live in the Black Belt region and suffer from higher dependency ratio, higher unemployment rate and higher infant mortality rate than the rest of Alabama. (Webster & Bowman 2008) The delineation of the historical outline of the Black Belt region, however, relies greatly on either the geographical extent of fertile soil or on simple racial compositions. Some research, for example, defines the Black Belt as counties that had a proportion of African American population higher than 50%. (Webster & Samson 1992) To connect the present with the past, it is necessary to understand whether there were distinctive social, economic and political activities within this region which have been vaguely delineated as the Black Belt by the researches illustrated in figure 3.1.

As illustrated in figure 3.2 and discussed in the report in Section 3.1.2, the result of hotspot analysis shows there were clusters of higher economic, social and political activities which stretched across the geographical locations which have been vaguely considered as the historical “Black Belt” by social scientists. The Black Belt in the Nineteenth Century were distinct from the rest of Alabama, Mississippi and Georgia. The analysis of the geographical extent of the market by postal activities confirms that the contemporary impoverishment and devastation of the contemporary Black Belt has deep historical roots, which was very likely to be linked to the aggressive establishment of the “cotton kingdom” before the Civil War.

Historical postal activities conjure a socio-economic context for the historical Black Belt. Hotspot analysis of human activities, however, cannot confirm whether the high activities within the Black Belt in the Nineteenth Century bears a causal relationship to the contemporary socio-economic status of the African American population within the Black Belt. Such causality requires more rigorous statistical investigation and is beyond the scope of this paper. This purpose of this paper is to delineate the outlines of markets. But the results do suggest that the slavery-based cotton plantations within the Black Belt devastated the livelihood of generations of African Americans.

Section 3.1.2 Report on the Hotspot Analysis for Alabama, Georgia, and Mississippi (1849-1871)

*Note: See Figure A1.1- A3.7 in Appendix A for the mapped results of hotspot analysis for Alabama, Georgia, and Mississippi.*

The analysis of postal activities not only confirm that the Black Belt was a contiguous market area with substantial social, political and economic activities but also pins down the timing of the changes of the coalescence of markets. The Black Belt was most extensive between the years 1857 and 1861. For this specific period, the Black Belt identified by hotspot analysis roughly fitted with the vague description that it stretched “from eastern Alabama to Northern Mississippi”, and modifies the east end of Black Belt to Columbus, Georgia and its surroundings.

The core of the Black Belt is the cluster of towns surrounding Montgomery, Alabama. This cluster was not only one of the first clusters that came into existence along with the Starkville, Mississippi cluster and the Columbus, Georgia cluster but was also the only cluster that had substantial activities after the Civil War when other clusters all disappeared. Comparing figure 3.14 and figure 3.16, human activities within Black Belt experienced substantial shrinkage in the size of its extent as well as in terms of the region's relative importance in Alabama, Mississippi and Georgia. This finding suggests that the Civil War had a negative impact on the slavery-based cotton plantation economy, but did not eradicate its core area in Montgomery, Alabama and its surrounding towns.

The Starkville area at the northern end of the Tombigbee River in Mississippi and Montgomery area at the eastern end of the Alabama River were the earliest Black Belt area that became significant in term of economic and social activities as measured by postal activity at the state level in the year 1849 ( $d = 40\text{km}$ ,  $n = 5$ ). The two clusters were 250 km apart from each other. The clusters most resemble a continuous "belt" between the years 1857 and 1871 and the belt-shape first emerged in the year 1857. A belt-shape cluster formed along and in proximity to Alabama River, extending from the early Montgomery cluster towards Columbus, Georgia and westwards along the Alabama River ( $d = 160\text{ km}$ ,  $n = 17$ ). More precisely, the two extremes of the cluster located 72km east to Montgomery, AL and 83km west to Montgomery respectively. 81km away from the extended Montgomery cluster, another cluster emerged at Tuscaloosa, Alabama ( $d = 35\text{km}$ ,  $n = 5$ , 1857). Columbus, Georgia was also a hotspot that had geographical and transportation relevance to the "Black Belt" area. Columbus extended into a cluster ( $d = 56\text{km}$ ,  $n = 6$ , 1861).

These clusters which were roughly located in the "Black Belt" area did not immediately connect with each other geographically. The economic connections among the clusters were unclear because the transportation was costly. Individuals could perhaps travel between these locations in one or two days. However, large-scale movements of goods among the Tuscaloosa cluster, the Montgomery cluster and the Starkville cluster were very costly. The clusters were on average more than 60 km apart from each other and were not connected by railroads. To transport large quantities of goods from the Tuscaloosa cluster to the west end of the Montgomery cluster, for example, required a



detour that was 400-km long over the steamboat-navigable part of Tombigbee River and Alabama River.

All clusters, however, were connected to the portal city, Mobile, which was a hotspot throughout the period of 1849-1871. This suggests that Mobile were an important outlet for the goods produced in the Black Belt area. Perhaps the output of the Black Belt exceeded the capacity of Mobile to ship it, The Florida& Alabama railroad that started operation in 1861 connected Montgomery to Pensacola, a portal city in Florida as well as to Mobile. Pensacola, however, did not become important right away in 1861. No hotspot was observed in 1861 near or at Pensacola. There was no data for the years between 1863 and 1867. In 1869, two hotspots formed in proximity to Pensacola and another hotspot formed at a portal city 30km away. After the Civil war, the Tuscaloosa and the Starkville clusters disappeared in 1869. The Columbus cluster also disappeared in 1871. The level of human activities that existed within these clusters became indistinguishable from that of the rest of the state. The Montgomery cluster persisted into 1871 but significantly shrank in size.

In summary, hot clusters did not distribute across the entire “Black Belt”. The hot or cold clusters are geographical areas that are statistically different from the rest of the state as measured by their postal activities. The clusters located within the span of the “Black Belt”, but they were more likely to form at the northern ends of the steam-navigable parts of the rivers. Temporally, the hot clusters most resembled an extensive belt between 1857 and 1861, starting from Columbus, Georgia, crossing mid-Alabama, ending at Starkville, Mississippi. Remember that each hotspot is associated with a neighbourhood/market that had postal revenues that were above the state average. The extent of the “Black Belt”, as identified by neighbourhoods of higher-than-average social, political and economic activities shrank significantly after the Civil War. This observation suggests that the distribution of economic and social activities of Alabama, Mississippi, and Georgia concentrated less at the “Black Belt” after the Civil War. In the case of Alabama, the concentration activities diffused across the state from the Black Belt in Alabama since no new cluster of high activities emerged in the state after the Civil War.

Figure 3.3. The Black Belt Delineated by the Hotspot Analysis of Postal Revenue Data



Figure 3.4. The Black Belt of Alabama. From *Quantitatively Delineating the Black Belt Geographic Region* by Webster & Bowman (2008).

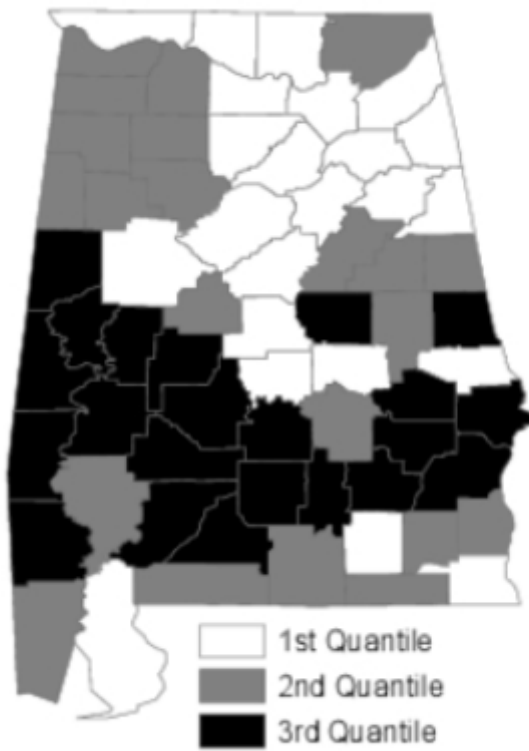


Figure 1b Component 1 Scores

Figure 3.5. Hotspot Analysis for the Black Belt Region (1849)

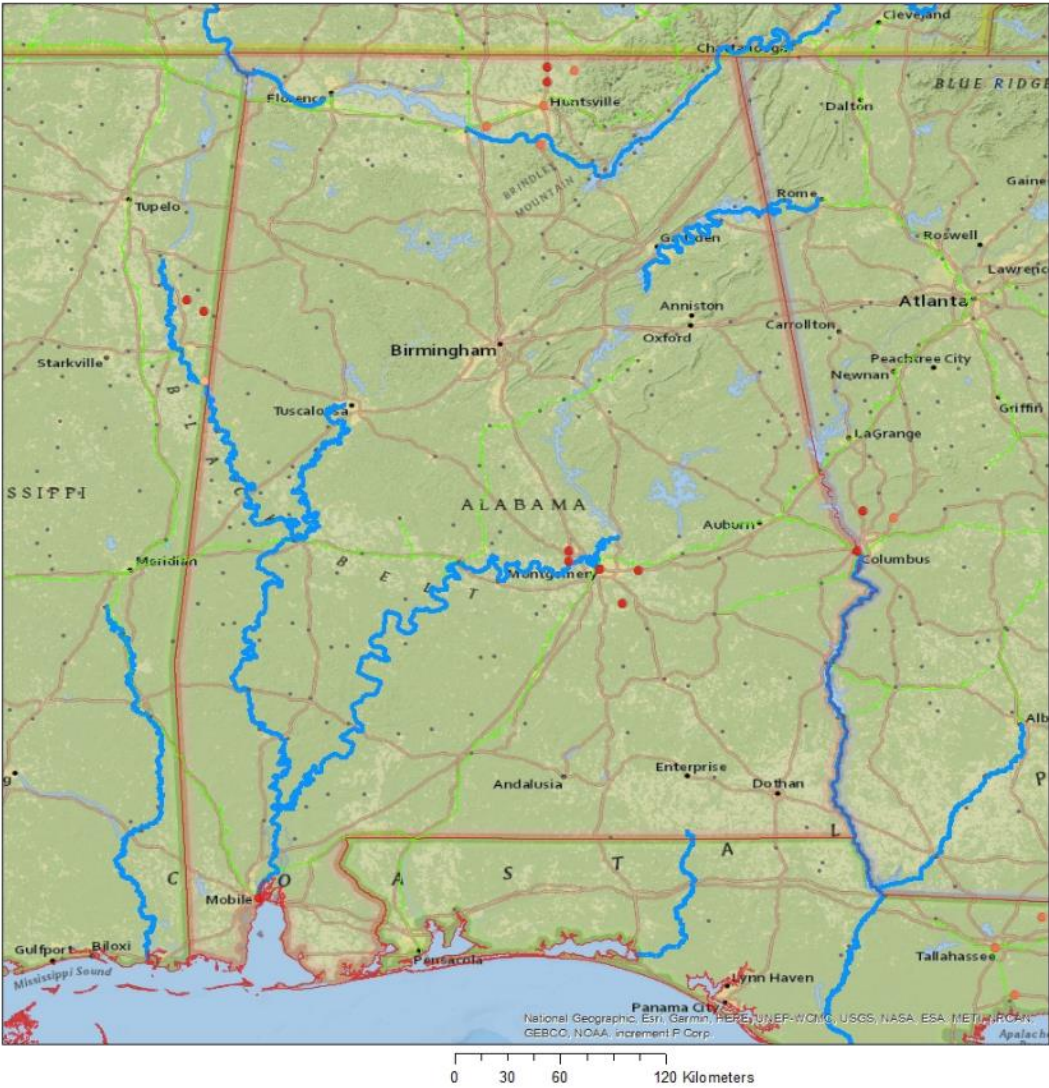




Figure 3.7. Hotspot Analysis for the Black Belt Region (1853)

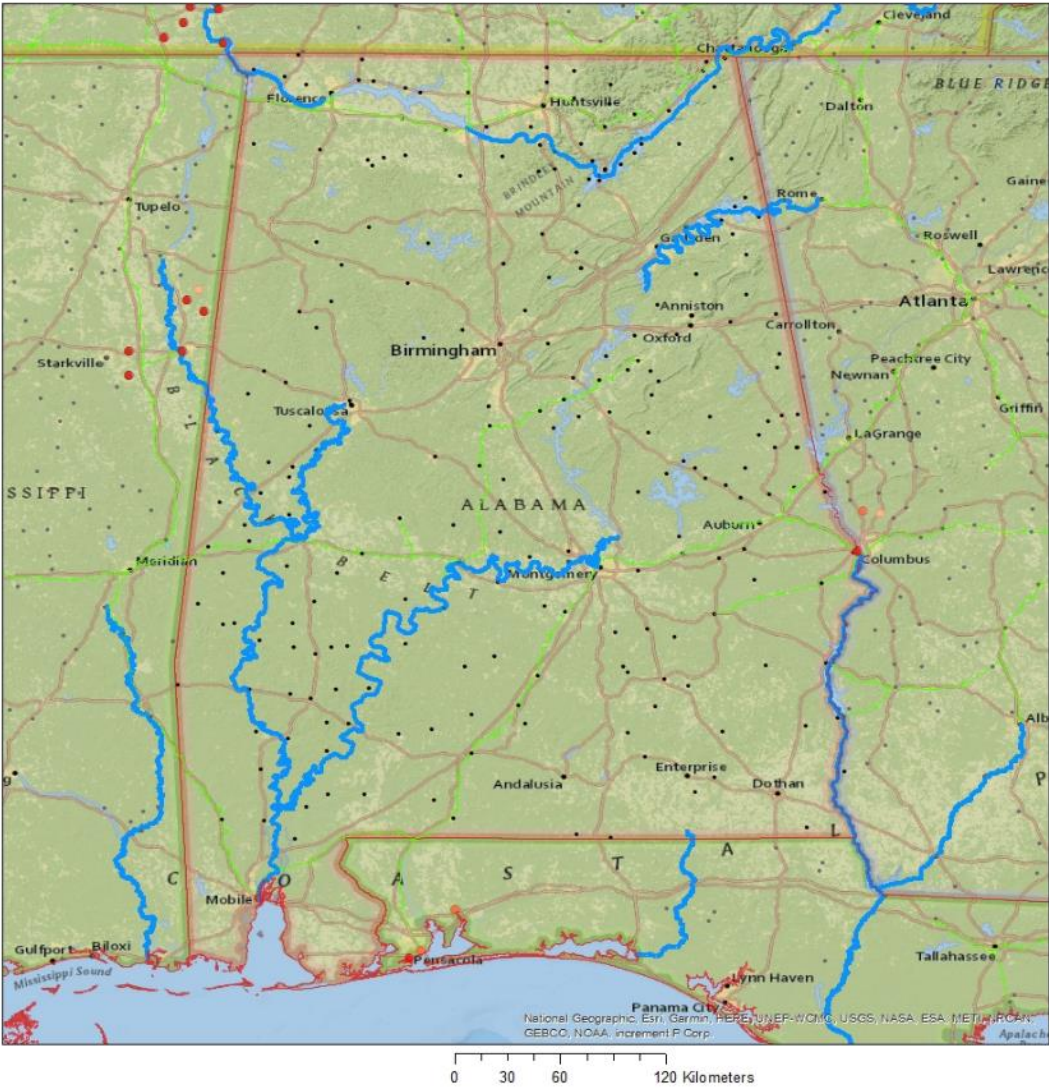


Figure 3.9. Hotspot Analysis for the Black Belt Region (1857)

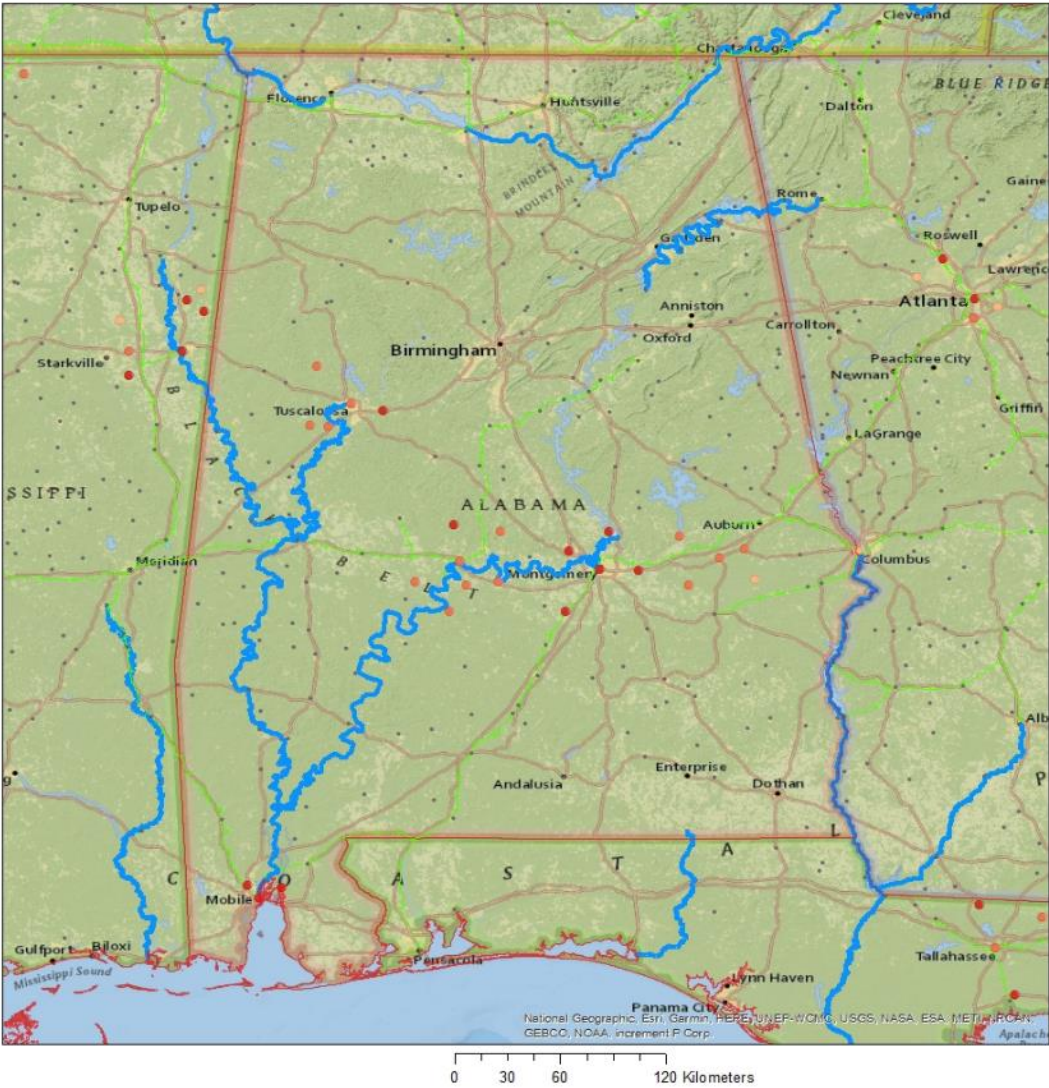




Figure 3.11. Hotspot Analysis for the Black Belt Region (1861)

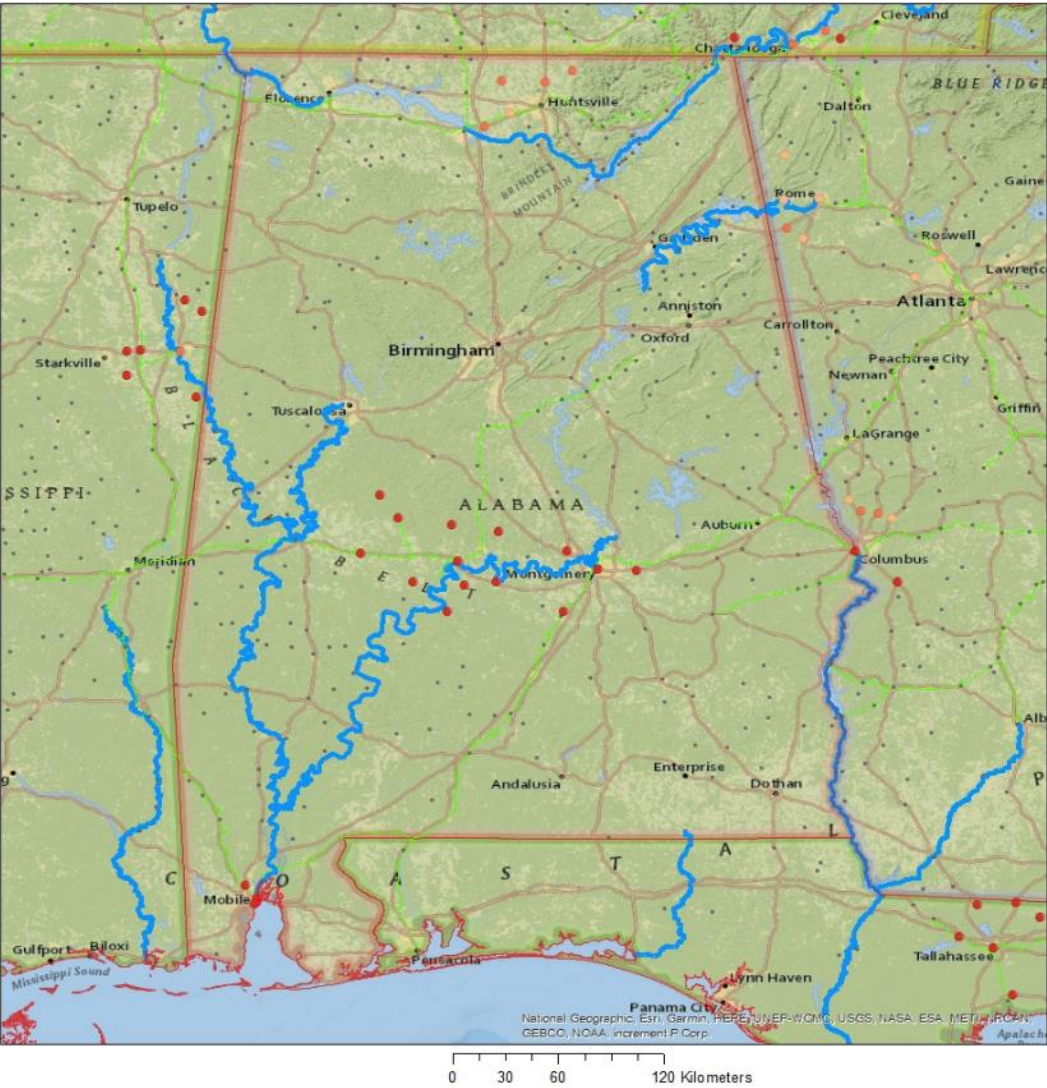


Figure 3.15. Hotspot Analysis for the Black Belt Region (1869)

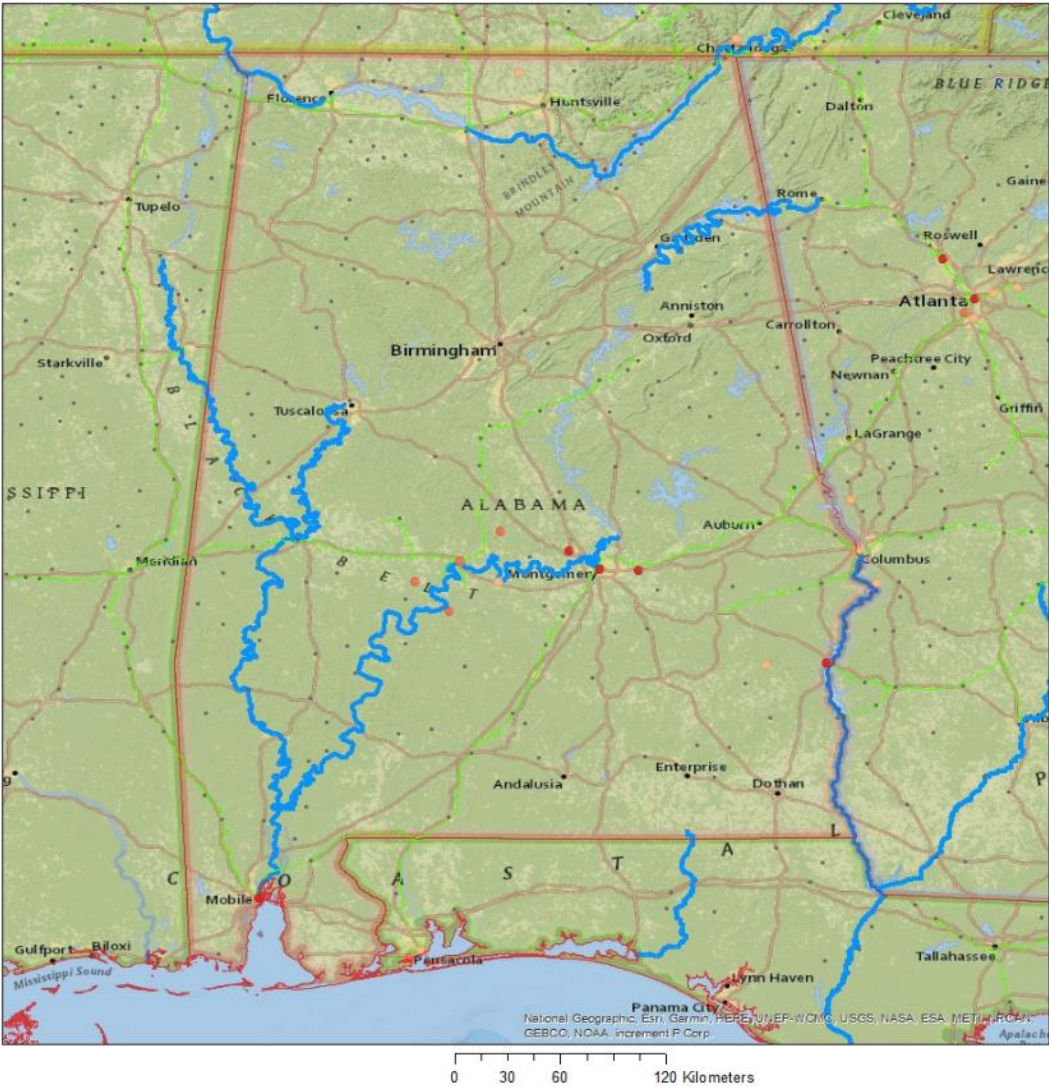
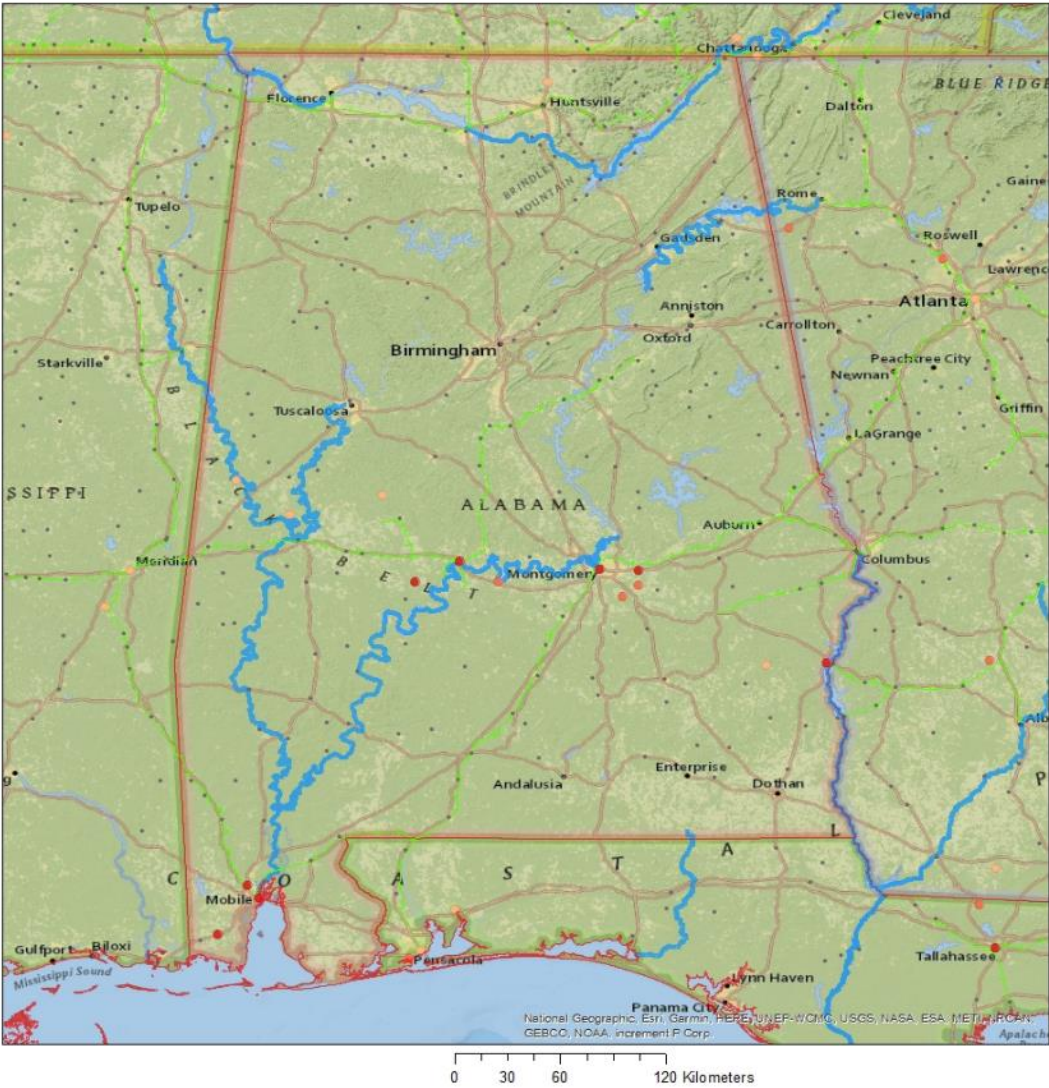




Figure 3.16. Hotspot Analysis for the Black Belt Region (1871)



**Section 3.2 North-eastern States**

**Section 3.2.1 Maine**

*Note: See Figure A4.1- A4.12 in Appendix A for the mapped results of hotspot analysis for Maine.*

In long term, postal activities concentrated near the coastline in the southwest corner of the state and form two clusters – the Augusta cluster ( $d = 56\text{km}$ ,  $n = 31$ , 1849) and the smaller Portland cluster ( $d = 48\text{km}$ ,  $n = 13$ , 1849) in the southwest direction to it. The two clusters are first observed in 1849 and persisted into 1871.

The Augusta cluster was connected to the coastline by the Kennebec River waterway. The Portland cluster was connected by railroad to the high clusters surrounding Boston, MA, and Concord, NH. The Portland cluster situated at the end of the Oxford-Cumberland Canal. The existence of the two clusters predates the completion of the railroad in Maine.

Starting from 1857, the cluster experienced expansion both towards the northeast and the southwest direction. The railway connected the Oxford-Cumberland Canal with the steamboat-navigable waterway on the Kennebec River which ended at Augusta. The Portland cluster extended about 70km towards the northeast direction and formed the Portland-Augusta cluster.

The arrival of the railway was first accompanied by the diminishment in the extent of the two clusters, but the two clusters regained importance in 1857 and merged together. From 1857 to 1871, the centre of the Augusta cluster gradually shifted toward Portland and the coastline ( $d = 120\text{km}$ ,  $n = 54$ , 1871). From 1857 to 1871, the Portland cluster also extended towards Maine's border with New Hampshire which were also close to the Lowell cluster in MA. It seems that the high clusters in New Hampshire and Massachusetts had some geographical attraction for the nearby cluster in Maine.

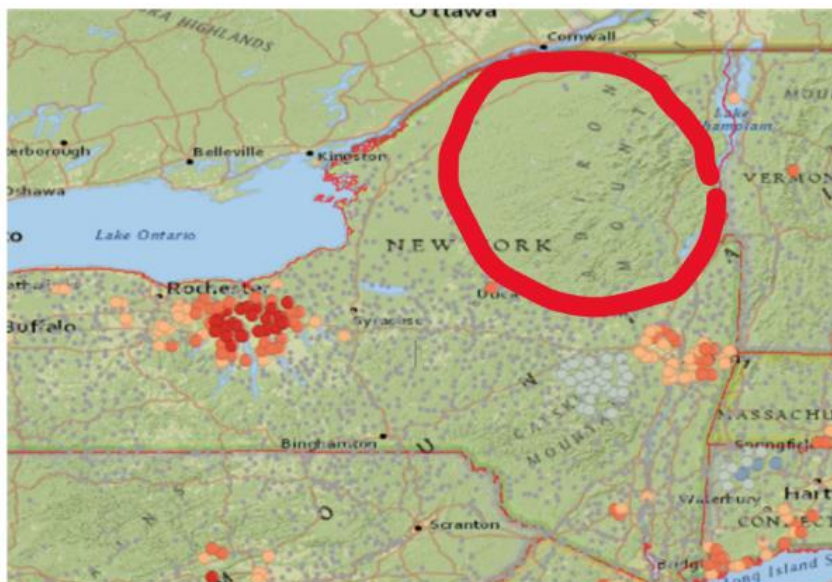
Separated from the transportation facilities in the southwest corner, a third hot cluster formed in the Fundy Bay area bordering Canada ( $d = 42\text{km}$ ,  $n = 10$ , 1851). The cluster first emerged all of a sudden in 1851 and then experienced a big shrinkage during the period of 1849-1871. After the Civil War, only three towns on the coastline remained important. In 1871, All of the hotspots in the cluster disappeared.

### Section 3.2.2 New York

*Note: See Figure A6.1- A6.12 in Appendix A for the mapped results of hotspot analysis for New York.*

The postal activity of New York clustered near Rochester in the upper Finger Lakes area ( $d = 90\text{km}$ ,  $n = 50$ ). Another smaller cluster situated near the Niagara Falls ( $n = 13$ ,  $d = 27\text{km}$ , 1853). A third cluster was centred by Albany city and overlaps the counties of Albany and Rensselaer ( $n = 39$ ,  $d = 50\text{km}$ , 1853). A Fourth cluster, the New York cluster was centred by the New York City ( $n = 74$ ,  $d = 62\text{km}$ , 1853). Three hotspots were found in Utica within a distance of 35km. Each of the clusters either was connected to the Hudson's River by the Erie Canal or situated over the Hudson's River. There was a large blank spot in the Adirondack mountains area ( $d = 190\text{km}$ ) throughout the period, except for some insignificant spots at the edge of the mountains. This observation coincides with the perception that the Adirondack Mountains were a historical recreational area and that settlements and activities were scarce in that area.

Figure 3.17. The Blank Spot in the Adirondack Mountains Area, New York (1867)



Most of the clusters and their extent were persistent, except Albany, which experienced a period of decline starting from 1857. In 1857, a cold cluster ( $d = 58$ ,  $n = 32$ , 1857) emerged 18km west to the Albany cluster in the Catskill Mountains area. Paralleling the emergence of the cold cluster, Albany started to shrink. By 1863, It has shrunk to non-existence whereas the boundary of the cold cluster extended ( $d = 62$ ).

The pattern of new settlements of post offices in one area with initially low postal revenue might be confused with the pattern of decline. Between 1853 and 1871, there was some new settlement in the Albany area. The number of post offices increased from 141 to 156 in the five counties where the Albany and the cold cluster were situated. Among them, Schoharie county experienced an increase from 33 post offices to 41 post offices. However, a decrease by 10% in the number of post offices were unlikely to have suppressed the hot cluster for an entire area for a period of 18 years even if the income of these additional 15 post offices were very low. Therefore, it is more likely that the Albany area experienced a sustained period of decline since 1853. From the 1850s to 1870s, Albany and Rensselaer were two of the most important manufacturing counties in New York. The amount of farm output also increased during this period in those counties. Unlike the counties at which the other clusters were situated, Rensselaer also owned the highest number of spindles in the 1840s. Perhaps the decline of the area was associated with the decline of the spinning industry.

A similar pattern occurred in the Niagara Fall cluster. This small cluster disappeared in 1863, paralleling the emergence of a big cold cluster about 20km away. By 1871, all spots in the Niagara area became non-significant. The beginning of the decline of the clusters paralleled the emergence of a cold cluster nearby ( $n = 30$ ,  $d = 70$ km). This pattern suggests that there were some linkages between a booming or historically important area and the neighbourhoods nearby. Even if the incomes of the nearby neighbourhood were not high, these neighbourhoods experience negative impacts when the hot cluster experienced a major and persistent decline.

The New York City cluster was a good example of the economic centre of two neighbouring states gravitating in one geographical area. Again, intra-state hotspot analysis was performed on each state, and the results of each state were graphically combined. By 1859, the extension of the New



York cluster within New York state remained largely within the Long Island area. However, the New York cluster geographically merged with the New Jersey cluster. The merged cluster grew at roughly the same pace. Besides New Jersey, the hot cluster of Connecticut also shifted from New Haven towards Stamford, between 1853 and 1871. By 1871, the hot cluster of New Jersey, extending from its heartland, had merged with the New York City cluster. All these clusters situated within major manufacturing counties of the states. Within the New York state, the influence of the New York City extended upwards along the Hudson's River into the White Plains only by 1857.

Figure 3.18 The New York Cluster, merged with hot clusters in Connecticut and New Jersey

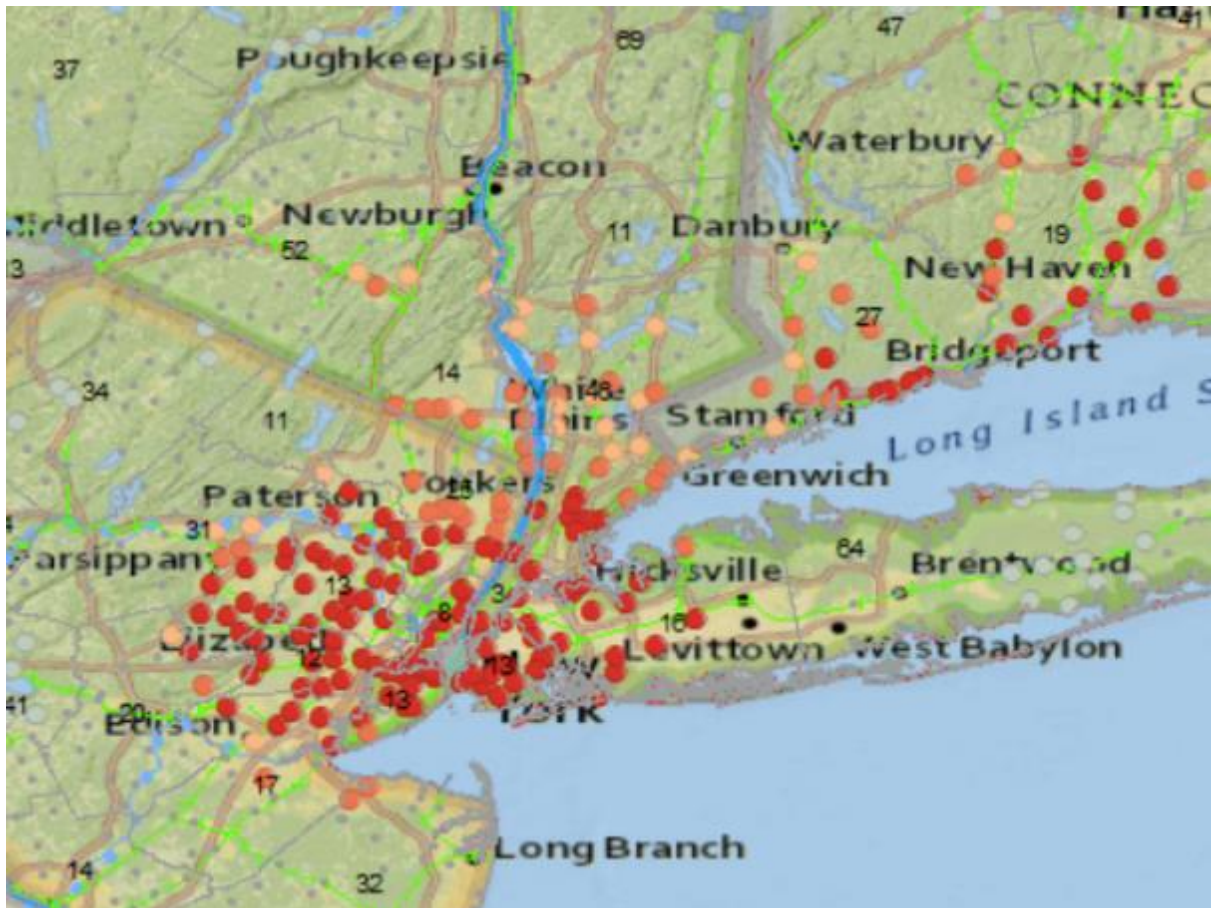
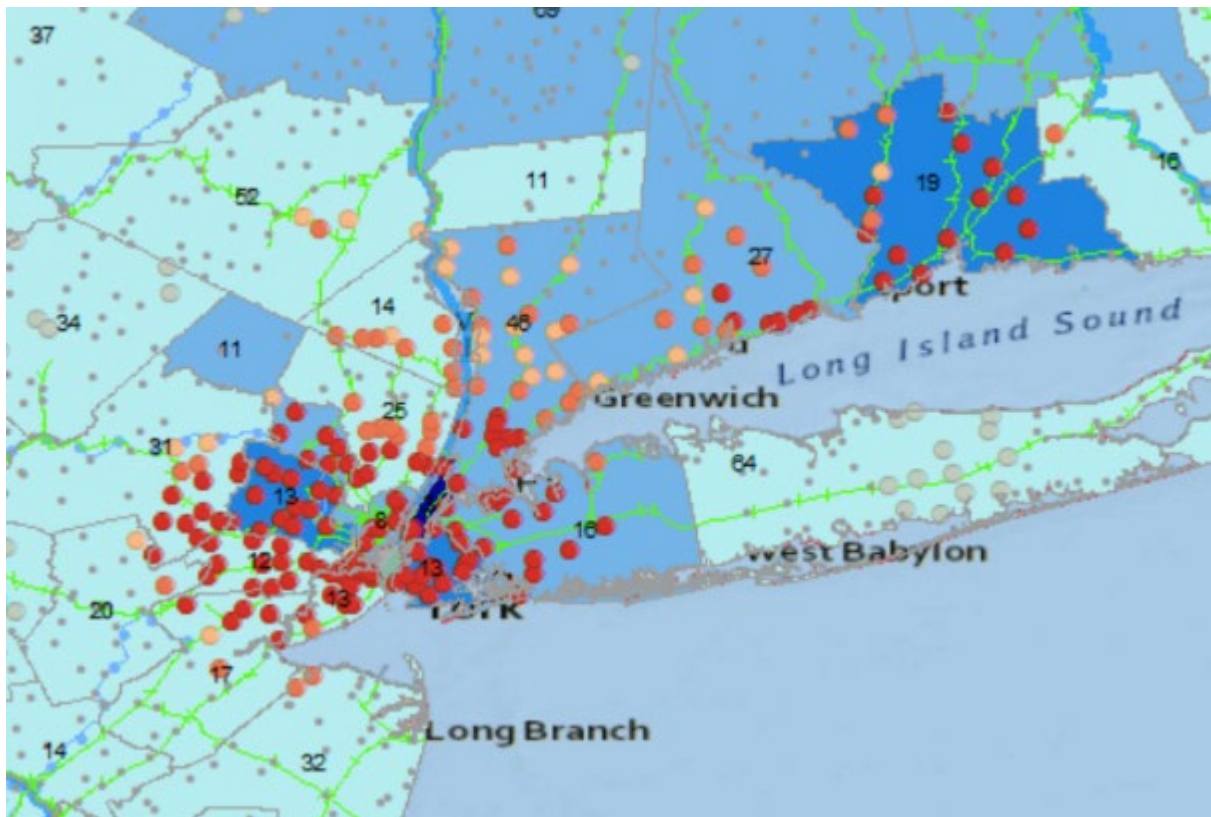


Figure 3.19 The NY-NJ-CT Cluster (1871), Illustrated Over a Heat Map of Manufacturing Output in 1870



### Section 3.2.3 Pennsylvania

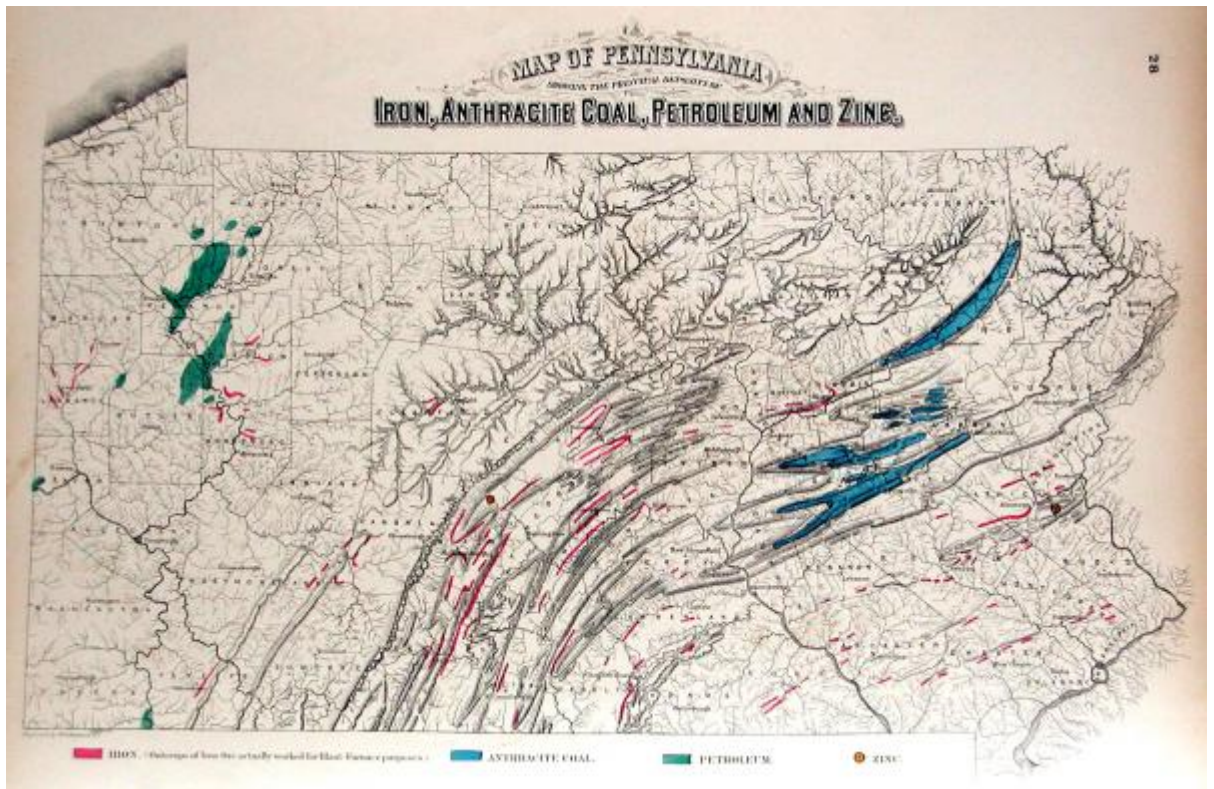
*Note: See Figure A7.1- A7.12 in Appendix A for the mapped results of hotspot analysis for Pennsylvania.*

Three clusters were persistent in Pennsylvania: The Erie cluster (n = 9, d = 55km, 1849; n = 81, d = 102km, 1871), the Reading/Appalachian Mountains Cluster (n = 16, d = 45, 1849) and the Philadelphia cluster (d = 55km, n = 31, 1849 – band shape; n = 47, d = 55 km, 1863- circle). Scranton, Lancaster, Altoona and Pittsburgh as well as some areas in the Appalachian Mountains area exhibited some clustering pattern in this period, but these clusters were neither persistent nor extensive.

The expansion of the Erie cluster towards the Allegheny River was the most noticeable change in Pennsylvania in this period. Before 1861, clustering only existed in a couple of towns at the lakeshore. Starting from 1861, hot spots started to fill in the blank space between Erie and the river. Paralleling the expansion of the Erie cluster was the completion of the Erie Railroad in this specific area.

The Appalachian Mountains experienced a prolonged period of mining boom at the beginning of the Nineteenth Century. Part of the canal system in the Appalachian Mountains area was built by mining companies to facilitate its business. Clustering not only existed among the mining towns in the mountains but also existed in the lower areas near Lancaster, which were connected to the mines by these railroads and canals. The Reading/Appalachian Mountains cluster shifted into the mountains starting from 1853, along with the construction of the coalfields railroads.

Figure 3.20. “Map of Pennsylvania Showing the Principal Deposits of Iron, Anthracite Coal, Petroleum and Zinc” by Walling & Gray (1872)



### *Philadelphia*

The Philadelphia cluster was initially a thick band of 31 hotspots along the Delaware River near the estuary. From 1849 to 1871, the cluster expanded 20 kilometres inland. The cluster experienced a major decline after the war. In 1869, it shrank to about a third of its size in 1863. In 1871, only three towns 25 km away from Philadelphia remained. The observation of the



diminishment of the Philadelphia cluster is confused with the integration of 16 post offices in Philadelphia county into 1 post offices in 1869. Since postmaster compensation was censored by a maximum, the removal of 15 nearby post offices in Philadelphia county and the integration of them into 1 post offices might have caused the disappearance of hotspots in the Philadelphia cluster. A single post office, the Philadelphia post office took over the business of 14 post offices, but this integration of postal business was not reflected by the inferred postal revenue of the Philadelphia post office since postmaster compensation was censored by a maximum.

A closer examination of the average postal revenues of the towns surrounding Philadelphia confirms this speculation. The disappearance of the Philadelphia City and the towns close to it as hotspots were observed for the year 1869. Note that the data for the year 1867 and the year 1865 was missing, therefore, the change in mean postal revenue were investigated for the years 1863 and the year 1869, rather than for the years 1867 and the year 1869. Integration of the businesses of 15 post offices in the Philadelphia county is the major source of confusion. If this integration is the main reason that the mean of postal revenues of the post offices within a 30km distance from Philadelphia City became insignificant spots in the year 1869, then after excluding all post offices in Philadelphia county for both the years 1863 and 1869, the mean for each year should look similar.

A comparison of the means before and after excluding the post offices in Philadelphia counties confirms the hypothesis that the integration of 15 post offices was the main reason that the areas surrounding Philadelphia City became insignificant spots. After excluding the 15 post offices in Philadelphia county from the calculation of the neighbourhood mean for the year 1863, the mean postal revenue of post offices within a 30km distance from Philadelphia City decreased to 25058.79 dollars, only 39% higher than the state average. After excluding the post office of Philadelphia city, which was the only remaining post office after the integration of postal business in the year 1869, the mean postal revenue was 28931.72, 40% higher than the state average. The postal revenues were similarly higher than the state average after excluding the post offices in Philadelphia county, suggesting that the Philadelphia should not have disappeared in the year 1869 if there were no integration of the postal businesses. Including the post offices in Philadelphia county into the



calculation of the neighborhood mean, on the other hand, the average postal revenue of all post offices within 30km from the Philadelphia City become 57354.69 in the year 1863, 217% higher than the state average, and this number decreased to 38783.13 in the year 1869, 88% above the state average. The integration of postal business makes a big difference for the neighbourhood mean when it is compared to the state mean. Therefore, it seems that the integration of the postal business, rather than the decrease in postal revenue was the reason that the Philadelphia cluster suddenly became insignificant in the year 1869.

One deficiency of this dataset is that the salary of the postmasters had a maximum. If the postal revenue of a post office were censored by a maximum, and at the same time, the postal activities in an extensive neighbourhood were integrated into this post office, then the average of the neighbourhood would be lower than the actual level of its postal activities. On the other hand, the maximum of postal revenues is not a problem on its own. **In areas where post offices were evenly distributed geographically, the maximum simply curbed the influence of a large city and makes the influence of being surrounded by small towns more prominent.** This paper attempts to find a way to observe the geographical extent of markets where high activities were located. A big city surrounded by small towns is not the pattern that interests this paper. What is interesting to this paper are continuous neighbourhoods that had towns with high postal revenues that were geographically close to each other.

### Section 3.3 Midwestern States

#### Section 3.3.1 Wisconsin

*Note: See Figure A8.1- A8.12 in Appendix A for the mapped results of hotspot analysis for Wisconsin.*

The settlements of post offices spread from the lowlands in the southeast corner of the state to the Northern Highlands of Wisconsin between 1849 and 1871. Wisconsin River and Lake Winnebago

were markers of the frontier in 1849. In 1849, post offices were only dense in one-quarter of the state in the southeast corner. By 1871, settlements had covered areas beyond the lower half of the state.

Hot clusters of economic, social and political activities also moved along with the frontier of settlement. In 1849, the two clusters, the Kenosha cluster ( $d = 30\text{km}$ ,  $n = 4$ , 1849) and the Madison cluster dominated the postal activities of the state. The Madison cluster disappeared in 1853. Some hotspots remerged 60km below Madison at Wisconsin's border with Illinois, and it geographically merged with a large hot cluster at the Southern part of Rock River. The Kenosha cluster continued to extend and intensify between 1853 and 1861 ( $d = 36\text{km}$ ,  $n = 13$ , 1861). The Kenosha cluster at the lakeshore of Michigan Lake was the most persistent cluster in Wisconsin.

Before 1861, there was only one hotspot, Lacrosse, north to Rock River. The railroads connected the lower part of Wisconsin to Lake Winnebago and Lacrosse in 1859. The operation of these railroads paralleled the emergence of the two hot clusters at Lake Winnebago ( $d = 30\text{m}$ ,  $n = 8$ , 186) and Lacrosse ( $d = 50\text{km}$ ,  $n = 6$ , 1861). Some hotspots also emerged at the nodal point of this railroad system below Rock River. The Lacrosse cluster disappeared after the war. Lake Winnebago disappeared in 1865, but quickly reappeared in 1869 and extended to the Green Bay area in the northeast direction ( $d = 95\text{km}$ ,  $n = 17$ , 1869). The Kenosha cluster started to shrink in 1865 and the trend continued into 1871.

It seems that high economic and social activities continued to expand to the southwest part of the state. Between 1869 and 1871, a new cluster formed and expanded near Eau Claire ( $d = 36\text{km}$ ,  $n = 4$ , 1871) at the northern end of Chippewa River, a tributary of the Mississippi River in upper-west Wisconsin. New settlements above Rock River in lower-west Wisconsin created a neighbourhood of post offices with low activities.

### Section 3.3.2 Illinois

*Note: See Figure A9.1- A9.12 in Appendix A for the mapped results of hotspot analysis for Illinois.*

Illinois had four clusters in 1849 – the Chicago cluster (d = 47km, n =7,1849 ), the Peoria cluster(d = 58km, n = 21, 1849), the Spring field cluster(n = 13, d = 45km ), and the Quincy cluster(n= 8, d = 50km, 1849). In 1863, another three clusters formed respectively at Rockford near Illinois's border with Wisconsin (n=19, d = 56km, 1861), near Saint Louis at lower Illinois over the Mississippi River and at the northwestern corner of the state.

In 1857, the hotspots diffused away from each of the clusters along Mississippi River and the railroads. Only the Rockford cluster, connecting to the hot cluster in Wisconsin, remained and extended. A new cluster emerged at Cairo, in the southern tip of Illinois where Ohio River and Mississippi River converged. It is unclear what happened between 1863 and 1867 due to missing data. After the Civil War, the clusters continued to diffuse in upper Illinois near the rivers.

### *Chicago*

Chicago had a hot cluster only before 1857, but not anymore between 1857-1871. This does not mean that the postal activities in Chicago declined. In fact, the postal activities in Chicago were pretty high after 1857. For example, in 1869, the two postmasters of the Chicago post office were earning a total of 4,000 dollars. In 1857, the postmaster of the Chicago post office was earning 2,000 dollars. Both of the two numbers were the maximum compensation at the time. The postmasters of the Chicago post office had been earning maximum compensations since 1851.

In this period, there was no postal service integration as such in the case of Philadelphia, PA, at least not in the geographical sense. In the years 1869 and 1871, the closest post office to Chicago was 5 km away. Besides, there were at least six post offices within Chicago's 15km radius and there were 16 post offices within Chicago's 22km radius, not including the unmatched post offices. In the years 1853 and 1857, the settlement of post office in Illinois, in general, was less dense and there were fewer post offices within Chicago's 15km radius. However, there were at least 11 post offices within 22 kilometres. During the time for which the disappearance of the Chicago hot cluster is observed, between 1853 and 1857, there was no change in the number of post offices in Chicago's proximity.

This observation renders less plausible one source of error that is generated by the combination of maximum postmaster compensation and political integration. In the case of Philadelphia, 15 post offices were integrated into 3 post offices in the year 1869. At the same time, the Philadelphia hot cluster disappeared. In this case, the integration of postal services could have created three post offices that took up postal services of an extensive neighbourhood which was previously covered by 15 post offices. Such integration should have created postal offices with extremely large postal revenues that were not reflected by the inferred postal revenue since postmaster compensation was censored by a maximum. In the case of Chicago, the possibility that Chicago took up postal services in an extensive geographical area is eliminated, since there is no observation of a sudden decrease in Chicago's proximity. Therefore, the postal revenues of Chicago's surroundings reflect the actual amount of historical postal activities in that area.

What is left is the problem with maximum postmaster compensation. The problem is that if the Chicago City itself had extremely large postal activities going on in this period, well exceeding what's inferred from the postmaster compensation, it is not observed by the statistical technique used in this paper. However, it is beneficial or at least does not make a large impact on the analysis conducted in this paper if the influence of the large cities is curbed. The purpose of this paper is to define market extension at the microlevel. Therefore, the search conducted in this paper is for geographically contiguous neighbourhoods of towns, cities, and populated places that had higher social, political and economic activities than the rest of the state. **Large cities, uncensored by a maximum, might light up an entire neighbourhood as hot spots while in fact it was surrounded by small towns with little to no activities. Towns and places with little to no activities, however, is not considered as the extent of markets.**

In the year 1853, the average postal revenue of all post offices within 30km from Chicago, including Chicago City was 86085.75 dollars. After excluding Chicago, the number became 1366.00 dollars. In the year 1857, the average postal revenue of all post offices was 59133.55 dollars. After excluding Chicago, the number became 4895.90 dollars. The state average increased from 10679.79 to 21753.60 dollars. A simple comparison of averages shows that Chicago dominated the mean postal

revenue in both years. After excluding Chicago, the average postal revenue of post offices Chicago City's surroundings were very low in comparison to the state average. Although the absolute average postal revenue in Chicago's surroundings increased by 3529.90 between 1853 and 1857, it did not keep pace with the increase in the state average. The implication of the comparison of the averages is that Chicago was surrounded by many small towns. Before 1853, since other towns beyond the Chicago area had similar or lower social, economic and political activities, the Chicago was a hot cluster. After 1853, as indicated by the increase of the state mean postal revenue from 10679.79 to 21753.60 dollars, these small towns did not increase faster than the rest of the state even though they did grow. The influence of Chicago City did not change, as controlled by the censor of maximum postmaster compensation, 2000 dollars. Again, this paper investigates the activities of neighbourhoods, not big cities or towns.

### **Section 3.4 Western States**

#### Section 3.4.1 California

*Note: See Figure A11.1- A11.12 in Appendix A for the mapped results of hotspot analysis for California.*

The settlement of post offices in California began with the upper-Sacramento Valley area. More precisely, most of the post offices clustered in the areas of the Cascade Mountains and Coastal Mountains which were right next to the Sacramento Valley. The areas above the San Francisco Bay had much denser settlement than the areas south to the San Francisco Bay. The settlement of post offices did not stretch much beyond the extent of the Sacramento River and the steamboat-navigable part of San Joaquin River. The settlement of post offices quickly densified between 1853 and 1857 in the upper- Coastal Mountains surrounding San Francisco Bay and in the upper Cascade Mountains. By 1861, the web of the postal system started to stretch beyond the initial two areas of settlement into the northwest corner of California and Los Angeles. Between 1861 and 1871, post offices continue to densify in the areas that have been described. Settlement in the upper-Sacramento Valley area remains the densest and extended from the mountains into the Sacramento Valley. The steamboat-navigable

part of Sacramento River and San Joaquin River remained a good reference for the extent of denser settlements in California. The Mojave Desert area in south California remained largely unsettled.

In 1853 there were three hotspots at and near the northern end of the Feather River in the upper-Sacramento Valley area. In 1857, three small clusters emerged in upper-Sacramento Valley. The San Francisco area did not have extensive neighbourhoods of high activities before 1865. Before 1861, it seems that the neighbourhoods by the area convergence of Sacramento River and San Joaquin River about 50km east to San Francisco were more important.

### *San Francisco*

Similar to the Chicago post office and Philadelphia post office, San Francisco had very high postal revenue in as early as 1851. Its postmaster had been earning the maximum compensation throughout the period. However, San Francisco was initially surrounded by many small towns. The average postal revenue of post offices within a 30km distance from the San Francisco city, excluding the San Francisco city itself, was 2774.17 dollars, 95% below the state average, 58336.08 dollars. In 1871, despite the fact that the state average shrank to 27205.07 dollars due perhaps to the war, the average postal revenue of the post offices within a 30km distance from the San Francisco, excluding the San Francisco City itself, had increased to 69864.63 dollars, 61% above the state average in the year 1871.

Despite the well-known fact that the San Francisco city was a centre of population in the Nineteenth Century, it seems that the towns San Francisco were not affected until around 1865 when a hot cluster (n=12, d = 100km, 1865) first emerged in the San Francisco area. Starting from the year 1865, the San Francisco cluster continued to extend its boundary and increase the number of towns that were included as hotspots (n = 27, d = 168km, 1871).

Postal Revenue of Postal Offices Within 30km from San Francisco City (1871)

Postal Revenue of Post Offices Within 30 km of San Francisco City (1871)

Post Office	County	Longitude	Latitude	Postal Revenue
Oakland	Alameda	37.80437	-122.271	486384
San Rafael	Marin	37.97361	-122.531	24673
San Leandro	Alameda	37.72493	-122.156	22673
San Mateo	San Mateo	37.56299	-122.326	11340
San Quentin	Marin	37.94131	-122.485	5003
Alameda	Alameda	37.7641	-122.242	43424
San Pablo	Contra Costa	37.96215	-122.346	4102
Millbrae	San Mateo	37.59855	-122.387	400

Postal Revenue of Post Offices within 30km from San Francisco City (1857)

Post Office	County	Longitude	Latitude	Postal Revenue (\$)
San Leandro	Alameda	37.72493	-122.156	8248
Alameda	Alameda	37.7641	-122.242	4827
San Pablo	Contra Costa	37.96215	-122.346	1855
San Rafael	Marin	37.97361	-122.531	1120
Mill Valley Air Force Station	Calaveras	37.92278	-122.602	595
Oakland	Alameda	37.80437	-122.271	0

Section 3.4.2 Oregon

*Note: See Figure A12.1- A12.12 in Appendix A for the mapped results of hotspot analysis for Oregon.*

The first year that Oregon had records for postmaster compensation is the year 1851. The data for 1851 is not included in this paper. Therefore, the analysis starts with the year 1853. Most of the

post offices located near Willamette River and its confluence with Columbia River in the year 1853. Between the year 1853 and the year 1871, the settlement of post office densified the most in the same area, and also densified in the Blue Mountains area.

In Oregon, three hotspots first emerged at the estuary of the Lower Columbia River in the year 1853. Another two hotspots located on the coastline 102 km away South to the Lower Columbia River and near the convergence of Columbia River and Willamette River respectively. Starting from the year 1857, the centre of social, economic and political activities moved from the coastline to the confluence of the rivers ( $d = 45\text{km}$ ,  $n = 4$ , 1857). Between the year 1857 and the year 1871, the cluster continued to intensify and extended its boundary ( $d = 39\text{km}$ ,  $n = 6$ , 1861;  $d = 52\text{km}$ ,  $n = 14$ , 1871). The confluence of the Columbia River and Willamette River seems to be the most important place in terms of social, economic and political activities in this period of time.

## **Chapter 4. Conclusion**

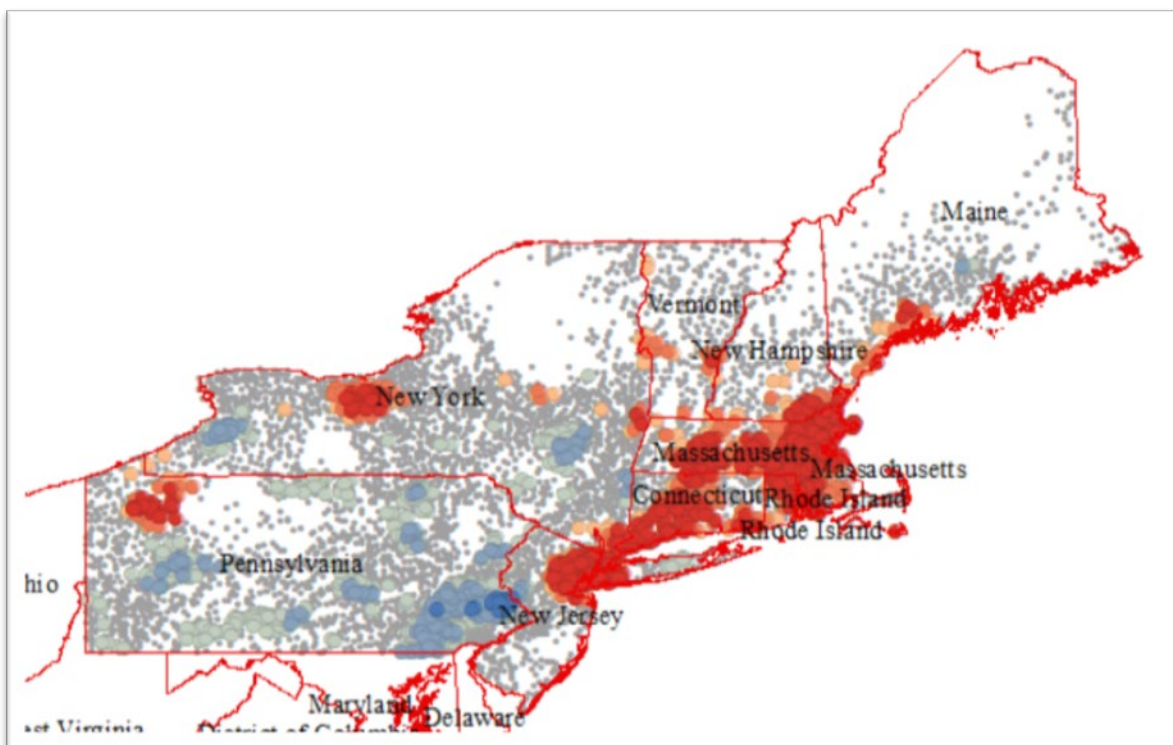
This dissertation establishes a methodology for explaining how markets coalesced and how this coalescing persisted. It proposes a new way to delineate the extent of the market by using an unexplored bi-annual, national-scale dataset for postal revenues. The second chapter of this paper establishes the connections between postal activities and social, economic and political activities with careful reasoning and researches. The third chapter of this paper applies the proposed methodology on a selection of states. These examples demonstrate the usefulness and the practicality of this methodology, as well as some problems that might arise along with the use of this dataset and solutions to those problems.

The example of the delineation of the Black Belt in Section 3.1.1 illustrates how the methodology proposed in this paper could be used to clarify, confirm and investigate vague socio-geographical concepts that are supported by biased and incomplete historical data. The example of Wisconsin illustrates that the methodology proposed in this paper could potentially be useful for identifying the frontier of human activities in the Nineteenth Century. Similarly, the example of the western states in Section 3.4 illustrates that the method proposed in this paper is useful in identifying where the migrants to the west initially settled.



Hotspot analysis also illustrates that state borders and county borders, as well as natural geographical features, are not very useful for the definition of the geographical extent of markets. The outlines of the clusters of high social, political and economic activities, as measured by the postal activities, frequently don't coincide these artificial borders. The hotspot analysis on the human activities in all states finds that county borders do not overlap with the outline of markets that are defined by postal activities. The irrelevance of the state borders is also evident in the illustration of figure 4.1. Figure 4.1 illustrates the result of using the entire north-eastern region as the sample for hotspot analysis (neighbourhood distance = 30km). None of the outlines of the distinctive areas of human activities overlaps with the state border.

Figure 4.1 Regional Hotspot Analysis- the Example of Northeastern States (1871)



It should be reemphasized that the hotspot analysis looks for “neighbourhoods of towns” that were associated with social, economic and political activities which were higher than the average of the towns outside of their neighbourhoods. A spot on the map would be considered to be a “hotspot”

by the analysis only if the average postal revenue of all post offices within its neighbourhood is higher than average. The usefulness of this practice is to identify contiguous areas that had higher-than-average activities. We already know what happened to large cities such as Boston, New York or Chicago in the Nineteenth Century, but we do not have much information about their surroundings. The examples of Philadelphia, Pennsylvania in Section 3.2.3, Chicago, Illinois in Section 3.3.2, and San Francisco, California in Section 3.4.1 illustrate that hotspot analysis is sensitive to the level of activities in the towns surrounding big cities.

## Bibliography

- America's Mailing Industry. (n.d.) *The Mailing Industry and the United States Postal Service*, Smithsonian Postal Museum. Retrieved from:  
<https://postalmuseum.si.edu/americasmailingindustry/United-States-Postal-Service.html>.
- Atack, J. (2015). Historical Geographic Information Systems (GIS) database of Steamboat-Navigated Rivers During the Nineteenth Century in the United States. Retrieved from the author's personal WWW page at Vanderbilt University:  
<https://my.vanderbilt.edu/jeremyatack/data-downloads/>.
- Atack, J. (2016). Historical Geographic Information Systems (GIS) database of US railroads. Retrieved from the author's personal WWW page at Vanderbilt University:  
<https://my.vanderbilt.edu/jeremyatack/data-downloads/>.
- Atack, J. (2017). Historical Geographic Information Systems (GIS) database of Nineteenth-Century U.S. Canals. Retrieved from the author's personal WWW page at Vanderbilt University: <https://my.vanderbilt.edu/jeremyatack/data-downloads/>.
- Baldasty, G. J. (1992). *The commercialization of news in the nineteenth century*. The University of Wisconsin Press.
- Conzen, M. P., & Conzen, K. N. (1979). Geographical structure in nineteenth-century urban retailing: Milwaukee, 1836-90. *Journal of Historical Geography*, 5(1), 45.
- Fitch, C. A., & Ruggles, S. (2003). Building the national historical geographic information system. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 36(1), 41-51.
- Esri. (2019). USA Topo Maps [basemap]. Scale Not Given. Copyright [2014] by the National Geographic Society.

- Foust, J. D. (1967). The Yeoman Farmer and Westward Expansion of US Cotton Production. *The Journal of Economic History*, 27(4), 611-614.
- Falk, W., & Rankin, B. (1992). The Cost of Being Black in the Black Belt. *Social Problems*, 39(3), 299-313. doi:10.2307/3096964
- Henkin, D. M. (2008). *The postal age: The emergence of modern communications in nineteenth-century America*. University of Chicago Press
- Hines, T. (2006). **Statistical Analysis of Postmaster Compensation.**, *Winton M. Blount Symposium On Postal History*. Smithsonian National Postal Museum, Washington D.C.. Retrieved from <https://postalmuseum.si.edu/research/symposiums-and-lectures/november-3-4-2006-symposium/november-3-4-2006-symposium-abstracts.html#Velk>
- Hines, T., & Velk, T. (2010). Explorations in the official register: Statistical analysis of postmaster compensation data from 19th-century New Hampshire. *A Chronology of Middle Missouri Plains Village Sites*, 29.
- How Hot Spot Analysis (Getis-Ord Gi\*) works. (n.d.). *ArcGIS Mapping Clusters toolset concepts*. Retrieved from <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm>
- Hecker, L.B. (2018). The history of Rural Free Mail Delivery in the United States. Retrieved from <https://ir.uiowa.edu/cgi/viewcontent.cgi?article=4164&context=etd>.
- Hines, T. (2005). [The] Number of post offices in New Hampshire [from] 1837 to 1911. *Granite Posts*, 18 (4), 7-8.
- Johnston, J. (1968). The united states postal monopoly. *Business Lawyer (ABA)*, 23(2), 379-406.

- Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, 99(3), 483-499.
- Kielbowicz, R. B. (1994). Rural Ambivalence Toward Mass Society: Evidence from the US Parcel Post Debates, 1900–1913. *Rural History*, 5(1), 81-102.
- Kim, S. (1995). Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860–1987. *The Quarterly Journal of Economics*, 110(4), 881-908.
- Kernell, S., & McDonald, M. (1999). Congress and America's Political Development: The Transformation of the Post Office from Patronage to Service. *American Journal of Political Science*, 43(3), 792-811. doi:10.2307/2991835
- Lamoreaux, N. R., & Sokoloff, K. L. (1997). Location and technological change in the American glass industry during the late nineteenth and early twentieth centuries (No. w5938). National Bureau of Economic Research.
- Marshall, A. (2009). Principles of economics: unabridged eighth edition. Cosimo, Inc.
- Manson, S., & Schroeder, J., & Riper, D.V., & Ruggles, S. (2018) *IPUMS National Historical Geographic Information System: Version 13.0 [Database]. Minneapolis: University of Minnesota.*
- North, D. (1964). Location theory and regional economic development. Regional development and planning. Cambridge: MIT.
- Priest, G. (1975). The History of the Postal Monopoly in the United States. *The Journal of Law & Economics*, 18(1), 33-80. Retrieved from <http://www.jstor.org/stable/725246>
- Rosenbloom, J. L. (1990). One market or many? Labor market integration in the late nineteenth-century United States. *The Journal of Economic History*, 50(1), 85-107.
- Ruggles, S. (2014). Big microdata for population research. *Demography*, 51(1), 287-297.

Thünen, Johann Heinrich von & Hall, Peter Geoffrey (1966). Von Thunen's isolated state : an English edition of Der isolierte Staat (1st ed). Pergamon Press, Oxford; London

The Federal Government of the United States. (1882) Annual Report of the Postmaster-General of the United States for the Fiscal Year Ended June 30, 1882. Washington: Government Printing Office. Retrieved from <https://babel.hathitrust.org/cgi/pt?id=coo.31924092982028;view=1up;seq=9>.

United States Congress, U.S. Department of the Interior. (2008). Domestic and Antarctic Names - State and Topical Gazetteer Download Files, U.S. Geological Survey. [Data file]. Retrieved from <https://www.usgs.gov/core-science-systems/ngp/board-on-geographic-names/download-gnis-data>.

United States Congress. (1846). Statue at Large *and Treaties* of The United States of America, Organization of the Government in 1789, to March 3, 1845 (Vol 3). Boston: Little, Brown and Company. Retrieved from: <https://www.loc.gov/law/help/statutes-at-large/13th-congress/c13.pdf>

United States Congress. (1855). Statue at Large *and Treaties* of The United States of America, December 1, 1851, to March 3, 1855 (Vol 10). Boston: Little, Brown and Company. Retrieved from: <https://www.loc.gov/law/help/statutes-at-large/32nd-congress/c32.pdf>.

United States Congress. (1862). Statue at Large *and Treaties* of The United States of America, December 1, 1845, to March 3, 1851 (Vol 9). Boston: Little, Brown and Company. Retrieved from: <https://www.loc.gov/law/help/statutes-at-large/29th-congress/session-2/c29s2ch33.pdf>

United States Civil Service Commission. [from old catalogue]., United States Bureau of the Census. [from old catalogue]. (1849-1871). Official Register Of The United States for the years 1849, 1851, 1853, 1855, 1857, 1859, 1861, 1863, 1865, 1867, 1869, 1871. Washington: U.S. Govt. Print. off.. Retrieved from <https://catalog.hathitrust.org/Record/009557655/Home>.

United States Postal Service. (2008). Rates for Domestic Letters, 1792-1863. Retrieved from <http://about.usps.com/who-we-are/postal-history/domestic-letter-rates-1792-1863.pdf>.

United States Postal Service. (2019). Rates for Domestic Letters Since 1863. Retrieved from <http://about.usps.com/who-we-are/postal-history/domestic-letter-rates-1792-1863.pdf>.

Velk, T. (2006). **Vital Information for Economists, Political Scientists, and Sociologists.**, *Winton M. Blount Symposium On Postal History*. Smithsonian National Postal Museum, Washington D.C.. Retrieved from <https://postalmuseum.si.edu/research/symposiums-and-lectures/november-3-4-2006-symposium/november-3-4-2006-symposium-abstracts.html#Velk>

Webster, G., & Bowman, J. (2008). Quantitatively Delineating the Black Belt Geographic Region. *Southeastern Geographer*, 48(1), 3-18. Retrieved from <http://www.jstor.org/stable/26225503>

Webster, G. R. & Samson, S. A. (1992). On Defining the Alabama Black Belt: Historical Changes and Variations. *Southeastern Geographer* 32(2), 163-172. The University of North Carolina Press. Retrieved August 27, 2019, from Project MUSE database.

Walling, H.F., Gray, O. W. (1872) Map of Pennsylvania Showing the Principal Deposits of Iron, Anthracite Coal, Petroleum and Zinc. Stedman, Philadelphia, PA: Brown & Lyon, Philadelphia, PA. Retrieved from <https://explorepahistory.com/displayimage.php?imgId=1-2-2077&storyId=1-9-E>.

Zhao, M. et al. (2018) Using Postal Revenue as A Proxy for Nineteenth-Century Local Economic Activities-- Data Validation. Paper presented 2018 World Economic History Congress, Boston.