Contents

List of Figures

List of Tables

Chapter 1

Literature Review and Related Works

1.1 What is institutional Investment

Institutional investors are defined in the Securities Exchange Act of 1934 (henceforth referred to as the SEA of 1934) as investors (natural or legal entities[1]) with investment discretion (or beneficial ownership) over a pool of funds greater than one hundred million dollars[2]. The theory is that by pooling capital, investors are in a better position to manage investment risk, and thus achieve a better risk-adjusted return (Davis and Steil, 2001). Those more familiar with the investment literature will see

[1] Institutional investors can organize under different corporate structures, such as banks, insurance companies, defined benefit pension fund, investor broker-dealer, hedge fund and incorporated company.

[2] The statute allows for the Securities and Exchange Commission to lower the threshold to a number no small than ten million dollars. However, this discretion has not been exercised as the date of publication (Davis and Steil, 2001). US Code. Title 15 - Commerce and Trade, Chapter 2B - Securities Exchanges, 78m. Available online at `www.law.cornell.edu/uscode/pdf/uscode15/lii_usc_TI_15_CH_2B_SE_78m.pdf`

the obvious hand of the efficient frontier hypothesis, in which larger pools of capital can more efficiently manage negatively correlated investment positions (Markowitz, 1952).

1.1.1 History of Institutional Investment

Blume and Keim (2012) trace the history of institutional investors to the first decade of the twentieth century, where they accounted for approximately five percent of the U.S. stock market, and about two thirds of the US Stock market in 2010. Commenting on this growth, Friedman (1996) notes that the share of institutional money in the US stock market grew fastest in the decades after the second world war, going from approximately 10 percent in 1950 to just under 50 percent in 1994. Similarly, the Institutional Investor Study by the U.S. Securities and Exchange Commission (1971) found, with a strict definition of institutional ownership of all outstanding stock in the stock-market at seven percent in 1900, and 19 percent in 1952. Using a broader definition of institutional investor, the study found ownership of 24 percent of outstanding stock in 1952 and 26 percent in 1958. Regardless of the definition used by the report, institutional ownership favoured positions that invested disproportionately in large publicly traded companies. Also cited in the congressional report was a census of stock ownership done by the New York Stock Exchange. This study found institutional ownership of all outstanding stock on its exchange showed growth from 31.1 percent in 1962 to 35.5 percent in 1965 and to 39.4 percent in 1970.

There's a similar growth trend within the subset of institutional investors called hedge funds. Using their own proprietary research and government supplied data, the research firm BarkleyHedge publicizes a count of hedge funds operating in the universe of US securities. Figure 1.1 demonstrates the evolution of hedge fund assets under management from a rapid recovery and growth in assets under management in the early 2000s stock market boom, followed by a precipitous drop after the Great Financial Crisis of 2008, superseded by a slow and steady rise during the Obama recovery.

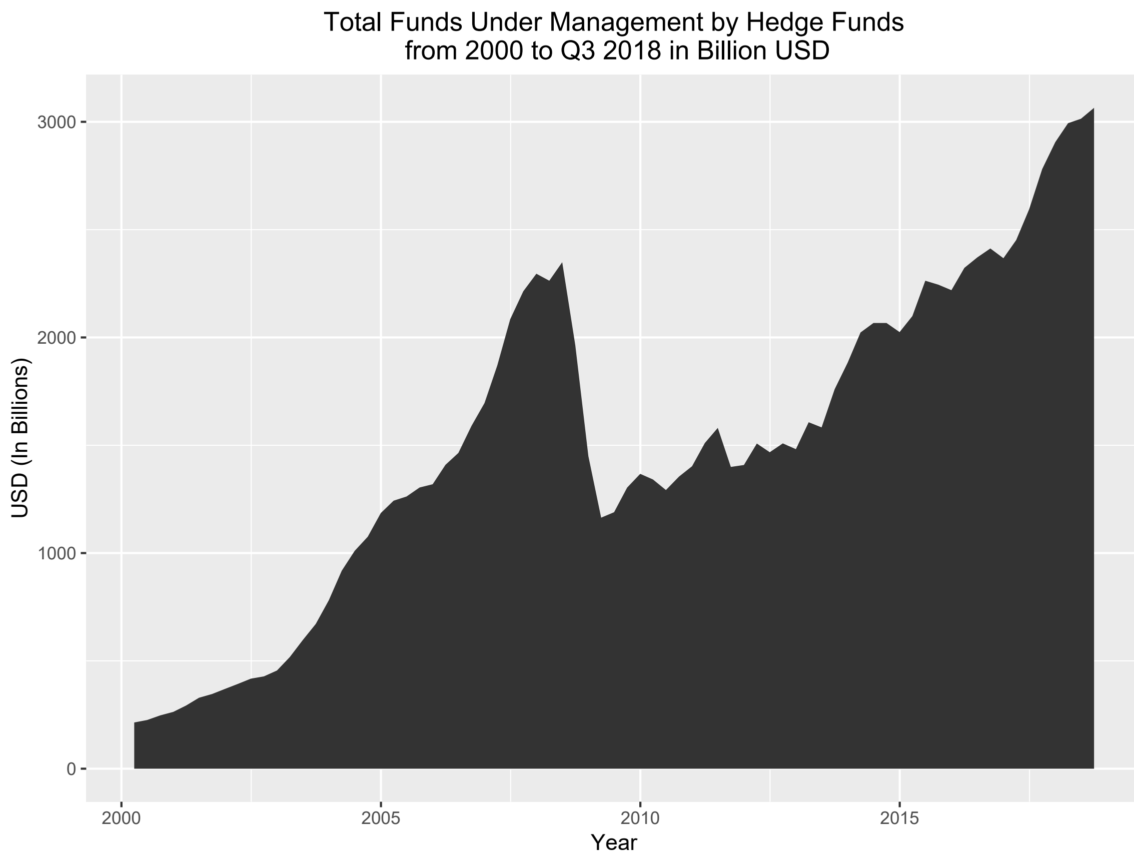


Figure 1.1: Total funds under management as measured by the research firm BarkleyHedge in December 2018.

Therefore, there's a consilience from these authors showing the gradual rise in importance of institutional investors in the US stock market across the twentieth century.

1.1.2 Fears and Questions about Institutional Investment in the 1960s

While the definition of investor capitalist can become quite broad – anybody who engages in market activity for profit can be defined as a capitalist – most people and institutions of modest means have a marginal impact on the market as a whole. At the other extreme, many fear that the concentration of substantial pools of capital can have a distorting effect on markets in a manner akin to how stellar objects gain influence over their peers via gravity as they accrue mass. To continue with the Newtonian gravity analogy, it was hoped that periodic disclosure of investments by the largest investors would shine a light on their stock movements and thus level the playing field with investors of more modest means. This periodic reporting would chart the distortions caused by large pools of capital, just like how gravitational distortions on other planets were used to predict and find the orbit Neptune.

The legal mechanism that mandates the periodic disclosure of institutional capital is Section 13F of the SEA of 1934. This section of law was signed by President Gerald Ford in January of 1975 and took effect in 1978 [3]. Yet, the passing of this bill was a long and tortuous affair spread over the better part of a decade and spanned four different congresses as well as the presidencies of Lyndon Banes Johnson and Richard Milhous Nixon. A look at the bill's legislative history, the rational, as well as what was discarded during the sausage-making process of getting legislation passed, can

[3] US Code, Title 15, Chapter 2b, 78m

provide insight on what the bill was meant to cover, what it wasn't meant to cover, in addition to the intended use of the tools created by the bill.

During much of the 1960s, there was fear that some shadowy cabal of investors were manipulating the stock market - seen as a key driver of American success in the cold war - to their own ends and to the public's expense via underhanded techniques such as front-running and manipulating who could serve on the board of directors. In order to allay fears, and find remedies if such action were warranted, the 91st Congress (January 3, 1969 to January 3, 1971) commissioned a study, which was completed and presented in front of the 92nd.

While the 1971 report could not prove extant manipulation by institutional investors, the report did suggest that a periodic disclosure of investment positions would help allay fears by increasing transparency in the market and thus reduce the perception of corruption. Furthermore, the report shows that investors – across different lines of investment, be it insurance, banks, pension funds among others – were increasingly conscious of "performance", and thus were willing to increase the risk of their portfolio in exchange for higher yields (U.S. Securities and Exchange Commission, 1971). However, the commission found in interviews with investors that they were unaware of the nature of the risk they were running by chasing higher yields. In order to protect investors, the report suggests that periodic disclosure of investment risk would help investors balance risk and reward in their investment decisions, and looked for regulatory tools to make this a reality. The report also found that the SEC had the pre-exiting statutory authority to require increased risk reporting for mutual funds under the Investment Company Act of 1940, but that institutional investors were not covered by this Act since by their very nature, institutional investors were not a public facing investment provider. As a consequence, the SEC asked the Congress for tools to mandate regular disclosure of stock holdings for institutional investors. One more problem uncovered by the report was the disparate treatment of domestic and offshore investment funds. It was found that in practice, funds that operated outside of the territorial jurisdiction of the United States had a competitive advantage since they operated under a more permissive regulatory and taxation regime. The report suggests that by equalizing the playing-field by forcing foreigners to register with the Securities and Exchange Commission, foreign investors would also receive stronger consumer protections.

Senator Harrison Williams (D-NJ) [4] shepherded the 13F amendment through multiple reform minded Congresses (Shaw, 1981). The first pieces of legislation that can be recognized as the ancestors of the current Section 13F are a pair of bills called Senate Bill 2234 and Senate Bill 2683. The more ambitious bill (Senate Bill 2234) had a more inclusive definition of who is an institutional investor, a reporting threshold of 10 million dollars rather than 100 million dollars in S.2683, as well as mandating reporting of a broader basket of holdings, such as real estate, art, bonds, cash deposits, and commodities in addition to securities. By contrast, Senate Bill 2683 is the more modest of the two bills that Senator Williams presented concurrently to the Senate Banking committee, and is substantially similar to the present section 13F of the Securities and Exchanges Act of 1934.

[4] Ironically Senator Williams is the only Senator successfully convicted during the "ABSCAM" investigation into Congressional corruption in the early 1980s. Gershman (1982)

Senate Bill 2234 was deemed to be too invasive and impractical by the Republican ranking member Bill Bennett, since the broader basket of disclosure wasn't as easily priced as securities that are openly and regularly traded on various exchanges. As a compromise, language was added to Senate Bill 2683 to give the SEC discretion to ratchet down the reporting threshold to 10 million dollar should they feel it necessary.

Senate Bill 2683 sailed out of the banking committee and passed in the Senate with little opposition. However, the bill did not make it to the House of Representatives. Journalists covering this story attribute the failure in the lower house to the intrepid lobbying by Wall Street agents, upset by the lowering of brokerage rates that was recommended by the Congressional report (Zimmerman, 1971). During the lame duck session between the 93rd and 94th Congresses, Senator Harrison Williams went on a publicity tour in order to drum up support for the bill in the face of the New York based opposition (Dallos, 1974b,a). His efforts were rewarded when the language to create section 13F of the SEA of 1934 was passed by congress early in the 94th Congress and was signed into law by President Gerald Ford on June 4th, 1975 (Library of Congress, 1975).

1.2 Why Geography and not Economics

1.2.1 Trading in Aspatial Random Walks

In 1900, French mathematician Louis Bachellier submitted his thesis called *"Théorie de Spéculation"*, in which Bachellier formulated that the long-run expected value of speculation on a market experiencing a random walk process was zero. In other words, if one were to assume that the stock market was truly random and thus had a long term trend of zero, it would be impossible to gain money off the stock market by buying and selling stocks only at the opportune time over a sufficiently long time period. While the mathematical proofs in Bachellier's work was more intuitive than rigorous, often hinting mathematical concepts that would shape the field of Mathematics in the twentieth century such as Brownian motion and Markov chains, this work was an important stepping stone to Eugene Fama's Efficient Market Hypothesis (Courtault et al., 2000).

The Efficient Market Hypothesis (EMH) (Fama, 1970, 1991) posits that asset prices fully reflect all available information. As such, it follows that it's impossible for the average investor to continuously outperform the market average performance on a risk adjusted basis, since any information is updated and baked-into the price of the security. Eugene Fama offers the theory in three related variants: Weak, Semi-Strong, and Strong. The Weak variant posits that it is impossible to derive future prices from past information, the Semi-Strong variant posits that current prices reflect all known public information and the Strong version that all information (private and public) is reflected in the price (Fama, 1970). Graves (2003) argues that this seminal paper cast a long shadow on the field of investment research, to the point that many papers fail to consider geography as a plausible explanation for sustained trading advantage, since it would violate the Semi-Strong and Strong version of the EMH. For example, Easley and O'Hara (2011) find that hedge funds survive on information asymmetry, private knowledge and price ambiguity, but fail to inquire about possible sources for these sustained advantages. Similarly, Cohen et al. (2008) find that mutual funds overweight stocks of firms in which the directors of the mutual funds have a board of directors connection with a shared educational network (*alma matter*), but fail to consider current social networks and geographical proximity as confounding variables.

That being said, the literature is rife with studies that appear to conciliate on the point that there is some geographic bias in investment returns, and that these abnormal returns stem mostly from local information asymmetry. However, it does appear that this phenomenon was stronger prior to the information technology and telecommunications revolution that was ushered in during the 1980s.

1.2.2 Big role for geography

From the first market towns to Marshallian industrial districts, commerce and other economic activities are the *sine qua non* of its existence. An inherent advantage of being located at a trade nexus is the ability to easily compile information on market conditions. Westaway (1974) finds that as firms grow, management functions aggregate towards larger urban centres, since these places have greater access to necessary specialized information. This serves as a foundation for Pred (1977), where they theorizes that the location of information-intensive activities is a positive feedback process. Furthermore, Wheeler (1988) as well as Wheeler and Mitchelson (1989a) show that urban centres see benefits proportional to their relative importance in corporate decision making. This fits nicely with Quaternary Location Theory (QLT)(Semple, 1985), which emphasises that command and control functions will naturally aggregate to large urban centres.(Wheeler and Mitchelson, 1989b).

In an interesting parallel to the debate on the importance of Marshallian agglomeration with regards to footloose industries, that is to say those that do not necessitate large fixed upfront costs such as factories, Mitchell (2019) looks at the productivity of literary authors of the 18^{th} and 19^{th} century. This study found that when controlling for a multitude of factors, authors were most productive when located in London UK, and that there's a statistically robust relationship between time spent in London and increased productivity. Furthermore, the results of this paper suggest that there was a benefit to being located in London that was not present in other UK and Irish literary cities such as Edinburgh or Dublin. The paper posits that geographic concentration fosters thicker social networks with their peers, individuals of influence (agents, editors, publishers) and patrons, thus facilitating the ease of getting published.

While the initial flurry of Quaternary Location Theory papers focused mostly on corporate locational preferences – specifically command and control centres, it wasn't long before the field turned its attention towards banking and investment. An early paper looking at the geography of institutional investment is Green (1993). This paper looks at inter-city ownership of American institutional investors by using a sample of 395 institutional investors that held stocks in Fortune 500 companies for the year 1980. In this sample, New York City is the only city in the first tier of urban hierarchy, followed by a set of four second tier cities and a steep decline thereafter. The ranking in between city population and financial ownership is not correlated, and the ordinary least-squares (OLS) spatial gravity model explains about 6 to 9 percent of the local bias in holdings. In a follow-up paper, (Green, 1995), the author adds an additional time window (1990) and compares the new data with the data from 1980. The OLS spatial gravity model for 1990 is quite different than the model for 1980, showing a more diffused spatial process, which the author ascribes to the increased role of telecommunications. Green also notes the absolute increase of investors in New York City, but that its role is less dominant in the urban hierarchy in 1990 than it was in 1980.

Meyer and Green (1996) examines the spatial distribution of mutual funds from 1940 to 1985. They find that most mutual funds are managed out of 3 main cities: New York, Boston and Chicago (in that order). Using log-linear analysis on three explanatory variables (location tier [5], year and mutual fund type), the researchers find that they can rule out a 3 way interaction, but can't rule out a 2-way interaction in the data. Closer examination shows that the most profitable funds are located in core cities.

Graves (1998) examines the location of mutual fund companies for the year 1996. The author posits that the size of a fund is a function of the fund's past performance, and that the past performance is somewhat dependant on the amount and quality of information available.

Graves (1998) gives three reasons why mutual funds have different spatial patterns than banks. The first reason is that mutual funds and banks have a different history of spatially-based regulations. More specifically, mutual funds did not expe-

[5] Core, Semi-Core and Periphery

rience the State banking era regulatory regime. Secondly, unlike banks which need to interact with customers on a regular basis to perform banking functions such as check cashing and bill payment, mutual funds can conduct their business by mail and other methods of communication. Lastly, banks and mutual funds have different economies/diseconomies of scale curves with regards to personnel and investment positions. This is mostly due to the fact that investment positions do not scale well, as they become more illiquid with size.

While Graves (1998) hypothesizes that the control nexus of investment funds will coalesce into the cities at the top of the urban hierarchy, the opposite seems to be happening, for smaller centres are growing faster than larger cities. A possible explanation for drop in the growth rate of funds in New York City, is that modern telecommunications have reduced benefit of co-location to the point that the higher rent no longer commensurate with the locational advantage. According to Graves, this result calls into question the ability of Quaternary Location Theory to explain the contemporary pattern of investment locations. Graves offers as an explanation that the theory was written during an era with highly aggregated data, and inferior communications technology – lacking fax, internet and low cost wireless communication.

Outside of Geography and located mainly in Finance, there exists a parallel literature examining the influence of locational choice and investment returns. Furthermore, this literature is highly steeped in empirical examinations over fitting evidence into established geographical theories. Hau (2001) finds that traders on the Frankfurt Stock Exchange who are located in Frankfurt outperform traders located outside of Frankfurt on a intra-day basis, suggesting that there is an information distance decay function. Similarly, Dvořák (2005) reports that foreign traders fare worse than domestic traders at the Jakarta stock exchange, and Choe et al. (2005) discovers that foreign-born traders pay on average 21 basis points more than domestic traders when buying stocks, and received 16 basis points less than domestic traders when selling. Meanwhile, Teo (2009) found that hedge funds with offices in the same country as their investments outperform hedge funds without an office in the same country as their investment.

Following this trend, Zhu (2002) used data from a discount brokerage firm and found that individual investors show a propensity to invest in companies that are local to them, and that this propensity cannot be explained by fundamentals-based investment strategies. Since these individual investors are also more likely to invest in firms that advertise heavily, the author suggests that this is a results of investors being biased by firms they find familiar. This finding is similar to the findings of Huberman (2001), who found that owners of Regional Bell Operating Companies tended to live in areas that were served by the company. That being said, Monk (2009) states that while investing in firms in which the investor has a high level of familiarity may represent a sub-optimal strategy from the point of view of traditional portfolio theory. In some cases, it can provide for those willing to look beyond the efficient market hypothesis a source of information overlooked by the market, and thus a way to profit from information asymmetry. That being said, well publicised investment flops in which State pension funds are used to prop-up failing local champions leading to large losses, such as the 80 percent haircut the State of Connecticut experienced on its loan to Colt Industries in the early 1990s, can make this type of strategy politically difficult to execute.

Bradley et al. (2016) reports that, in a sample of 16 internally managed state pension funds, they are over-weighted in local companies by 26 percent relative to the average portfolio. Furthermore, these investments occur predominately in companies that are active in local politics, as measured by both political donations and active lobbying. The authors explore three non-mutually exclusive explanations for this over-investment:

1. **Information advantages due to local effect:** This theory posits that political connections lead to better information flow to the pension fund trustees, and this can be used for trading advantage.

2. **Familiarity:** This theory posits that managers are more familiar with local firms and over-estimate the quality of their information, but is otherwise a neutral position.

3. **Pay to Play:** This theory posits that political bias and influence peddling leads to malinvestment of State pension funds into politically connected firms. These conflicted motives lead to worse performance.

In total, the evidence (that the effect is stronger in States with a larger share of politically appointed pension fund board trustees as well as States with more powerful members of congress) points towards solution 3 as being the most likely.

Malloy (2005) reports that geographically proximate analysts outperform distant analysts in their buy and sell recommendations. The author posits that analysts who make house calls rather than conference calls can obtain more valuable and actionable private information via face to face communication, direct view of the operations floor, talk to floor employees as well as being better positioned to talk to suppliers. The effect is stronger in smaller locals. Similarly, Farooq (2013) studied the buy and sell recommendations by foreign and local stock analysis covering Thailand, Indonesia, Malaysia and South Korea during the Asian Financial crisis (1997-1999). This study found that foreign-based analysts had more accurate buy recommendations, whereas local analysts had more accurate sell recommendations. Furthermore, Eckel et al. (2011) found, via spatial regression analysis, larger returns than what would be expected for investment firms that invested in companies within a headquarter with 50 miles of their location compared to a random portfolio of companies with similar attributes.

Continuing on the theme of information decaying over distance providing real investment advantages, Cashman et al. (2017) use the cost of borrowing capital for publicly traded real estate companies in Asia-Pacific as a proxy for the cost of information opacity. The authors conclude that more diffused firms (those operating in more than one country), have higher capital costs than firms that only operate in one country, and thus they posit that companies pay an opacity tax.

1.2.3 Moderate Role for Geography

There exists other literature that believe that locational advantages were quite measurable prior to the telecommunications revolution of the 1970s and 1980s, and accept limited role at best for locational advantages to accrue in the face of modern telecommunications technology.

Looking at the time period between 1925 and 1978, Rhoades (1982) looked at the distribution of deposits in commercial banks and found that due to bank consolidation that were mostly driven by mergers and acquisitions, the distribution of bank deposits were increasingly concentrated towards the top end of the top 100 largest banks list. Furthermore, while this period saw important demographic changes in the USA with the increasing population in the Southern and South-Western United States, changes in the location of the top 100 largest banks were less reflective of the demographic shift than would be expected in a naive model in which bank size is a function of population. This suggests that large urban centres with preexisting banking centres have an innate pull factor that make banks less footloose than would otherwise be assumed.

With a more expansive look at locational preferences, Bodenman (1998) examines the exodus of Finance, Insurance and Real Estate (FIRE) sector firms in downtown Philadelphia, Pennsylvania. During the period between 1983 and 1993, the concentration of FIRE firms located in downtown Central Business District (CBD) fell from 61.9 percent to 24.9 percent. Examining why firms were leaving the Central Business District, the author asked FIRE sector businesses for factors that were at the heart of the locational preferences. Personal preference and quality of life were given as top answers, whereas assess to information was not given as a priority. In a related study, Bodenman (2000) looks at how the information technology revolution permits institutional investors broader choice of location without sacrificing access to high quality and quantity of data/information. Bodenman finds that not all actions taken by institutional investors require face to face contact, such as accounting and regulatory compliance, portfolio management, and trading. In contrast, activities that do require face to face contact, such as finding and/or managing clients as well as researching investment opportunities do not require a constant downtown presence. As a consequence, Bodenman (2000) posits that active traders will have a propensity to locate in the CBD, whereas passive investors and quantitative traders will locate in suburban office parks where rent is less expensive.

Gong and Keenan (2012) examine the geographical dispersion and return on the island of Manhattan shortly before and in the aftermath of the 9/11 terrorist attack. Of the 79 firms surveyed, fifty-four did not change location, while ten moved on a temporary basis (one month to a year and a half), and fifteen changed locations permanently. Of the ten who changed locations toward the periphery of the New York area, the most common reason for returning to Manhattan is the ability to meet with clients. Most of the firms that moved were located in Downtown and Midtown, in contrast, those that returned were located exclusively in Downtown Manhattan. Furthermore, the survey says that prior to the 9/11 attack, most firm managers were reporting that their locational preferences were shaped by maximising the prestige of the building, adjacency to the New York Subway system, as well as being conveniently located in order to meet with clients. After the attack, the location preference was dominated by an emphasis on office space, building infrastructure and rental costs, while keeping in mind that high prestige buildings would be more susceptible to terrorism in the future.

And while we may not have seen the death of distance as predicted by O'Brien (1992), it can be argued that there is a role for space and place, as well as telecommunications reducing the benefits of co-location. This is the heart of Moriset and Malecki's argument (Moriset and Malecki, 2009), where modern telecommunications re-arrange spatial forces of agglomeration. Better communications reduces the need for vertical hierarchies and remove the premiums of co-location.

Chapter 2

The Data Pipeline

2.1 Introduction

The 13F-HR report is the cornerstone of this study, for it offers a very detailed peek into the stock holdings of all institutional investors with holdings above 100 million dollars USD in fair market value, as well as voluntary reports for firms with smaller holdings[1]. Understanding the data pipeline, that is to say how the data went from the SEC's Edgar server, wrangled into the databases, and then cleaned prior to use in statistical models is important in understanding the strengths and limitations of these models. Otherwise it's garbage in, garbage out (GIGO) research.

2.2 The 13F-Holding Report

??

[1]Some institutional investors with holdings under 100 million USD are compulsory rather than voluntary in nature due to having exceeded the 100 million USD reporting threshold in the previous 4 quarters.

There are countless news articles that use 13F-Holding Reports (13F-HR) data as a basis for "whale watching", that is to say, poring over the 13F reports of successful investors such as, but not limited to Warren Buffet, and imitating their strategies and/or replicating their holdings on a smaller scale (Brody, 2012). While some may debate the wisdom of buying and selling stocks based on what experts were holding 45 days in the past[2], other say that these reports allow smaller investors to gain insights based on the research departments of larger investors (WilmerHale, 2013).

The data for this thesis was collected from the SEC's Edgar database between 2015 and February 18$^{\text{th}}$, 2019. The Edgar database provides 13F filings in two different formats. The first of these formats is the ".txt" format, which covers the period of March 31, 1999 to March 31, 2013. It should be noted that despite the existence of older filings on the Edgar server prior to March 31, 1999, these filings covering the time period of 1990 to 1998 only exist for a handful of filers each quarter and thus would provide an incomplete and biased sample. This era of filings contain holding information in an unstructured format that are easily human readable, but unreliable when parsed by computers. The second era of filing formats covers the periods of June 30$^{\text{th}}$, 2013 to December 31$^{\text{st}}$, 2018. These filings are in the newer "XBRL" file format, which is a derivative of the popular "XML" file structure. This file format has the benefit of being easily machine readable. Furthermore, all 13F-HR/A files represent amendments to previous filings were integrated in to the database.

Due to the difficulties in parsing the older ".txt" file formats, this lead to the cre-

[2] 13F-HR reports are due to the SEC for public access no more than 45 days after the end of a quarter. For example, reports for the period ending March 31st would be due no later than May 15th (or the next Monday if that date would fall on a Saturday or Sunday)

ation of two different databases of institutional investors. One piece of information that was easily extracted from the ".txt" files was the business address of the investor. This leads to the creation of a database containing what is essentially "phone book" information for all institutional investors that filed at least 1 quarterly report during the 20 year period covered by this research ($n = 242084$). The second database is derived from the "XBRL" encoded files, and contains a list of all positions reported by the filer to the SEC. Since some filers chose to disclose more information than required, and in the interest of maintaining a fair comparison across firms, only positions containing securities were kept in the database ($n = 92539$).

When plotting the duration of how long different filers (as defined by unique Central Index Key (CIK)) exist in the database, as seen in Figure 2.1, one notices a pattern in the data where peeks can be found at $n + 1$ quarters where n is zero or an integer divisible by 4. The most likely explanation for this reporting artifact is the requirement to report for the next four quarters after which they have fallen back under the 100 million dollar reporting threshold.

2.2.1 Investors by Country

While these filings are filed under pain of perjury, there is no guarantee that these filings are a true and accurate reflection of the investor's books. In fact, the SEC's EDGAR server warns users that they are not responsible for any damages caused by acting on incorrect information. In line with this warning, it is obvious that some filings are incorrect. In a few cases, one quarter's filings were orders of magnitude

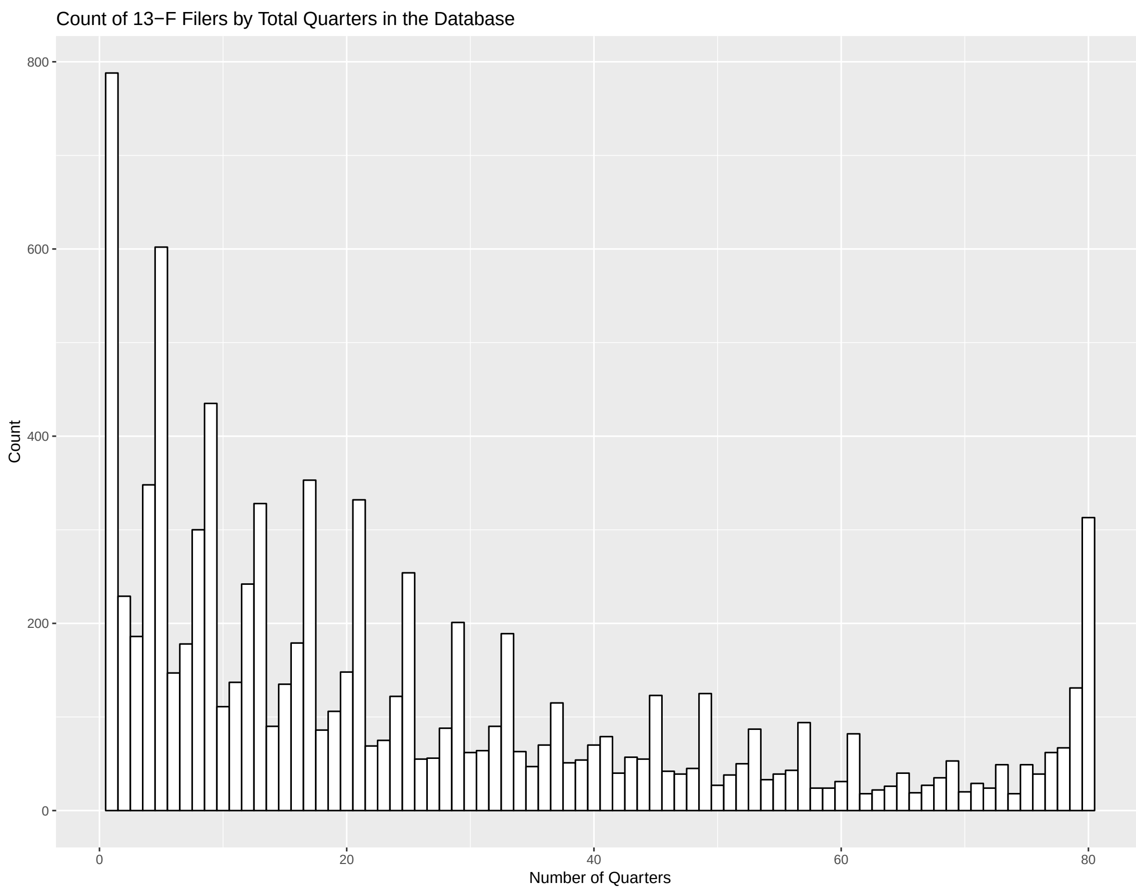


Figure 2.1: Count of 13-F filers by quarter in the EDGAR 13-F Database. One should note the regular pattern of $n + 1$ quarters, where n is zero or an integer divisible by 4.

larger than all other filings reported by that filer. For example, Firm 0000863748's filing for March 31st, 2016 reported a total fund value of 5,632,710,967,874.14 USD. This value is more than twice the value recorded for BlackRock family of funds, as well as being orders magnitude larger than the neighbouring filings. While there is no absolute guarantee that all filings are accurate, the yearly totals were verified for anomalous values using the Rosner Test as found in the EnvStats R package (Millard, 2013). This is further complicated by the fact that the legal basis for 13F-HR disclosure mandates only the disclosure of securities, and thus the conversion of an investment position to a non-reportable position has a warping effect on the top-line value for each fund. For example, if an investor were to convert a million dollar position in a company into a million dollars worth of real estate, the 13F-HR filing would show a drop of 1 million dollars in the subsequent filing, however the fund's true bottom line did not change. Furthermore, research conducted by Griffin and Xu (2009) looked at the difference between institutional investors and mutual funds, and how they organize their respective short and long positions. As a matter of law, mutual funds can't short stocks, and thus are forced to make their profit off of their long positions. By contrast, the hedge fund's more permissive regulatory regime allows for short-selling, and thus allows for the set-up of using long positions for hedges, and short-selling as a profit-generator. That being said, the researchers found that there is no statistical difference between the long position profitability between hedge funds and mutual funds. As a consequence, the long positions as reported in the 13F-HR filings should still hold valuable insights in corporate command and control functions, especially since many firms have a waiting period before the power to vote on board of directors vest.

During the period of June 2013 to December 2018, there were 570 filings with anomalous top-line values. These suspect filings were extracted from the database, and replaced with a synthetic entry using a weighted average of the surrounding 4 quarters[3].

However, not all abrupt changes in top-line valuation are due to erroneous filings. One such example is BlackRock which underwent a change of reporting scheme for 2017 onwards, where it decided to consolidate more reports under one filing (BlackRock Advisors, LLC, BlackRock Fund Advisors, BlackRock Investment Management, LLC, BlackRock Group Limited, BlackRock Institutional Trust Company, N.A. and BlackRock Japan Co., Ltd.), and thus went from reporting 70.6 billion USD to 1.8 trillion USD.

Interestingly, Bernard L. Madoff Investment Securities LLC (CIK number 00001386924) exists within the database from June 2006 to September 2008. However, as was revealed in December 2008, Bernie Madoff was at the centre of a 50 billion dollar Ponzi scheme (Appelbaum et al., 2008) in which instead of investing his client's money, he would deposit investments into his personal bank account, as well as pay redemption from this account. As Harry Markopolos detailed in his testimony to the House Financial Services Committee in the aftermath of the Bernie Madoff scheme's unravelling, use of 13F-HR should have uncovered the scheme years earlier, since what he reported on the disclosure form did not match what he was telling clients (Markopo-

[3]The main weighting is a (0.2/0.3/suspect entry/0.3/0.2), however, June 2013 and first company filings are treated with a suspect entry/0.6/0.4 (opposite weights for last filing and December 2018), the filing for September 2013 and filers with suspect second entry is 0.4/suspect entry/0.4/0.2. (Inverse weights for September 2018 and penultimate filings)

los, 2009). Due to being a known fraud, Bernard L. Madoff Investment Securities LLC (CIK number 00001386924) was censored from the database. While it's unknown how many other fraudulent investment funds exist, there is no other choice than to believe that all the filings are done in good faith, and that the 570 anomalous filings were based on human error.

2.3 Tying Capital to Physical Space

Financial Capital is inherently global while money often acts on the local scale (Clark, 2005). An apt metaphor according to Clark is that money will flow like mercury due to the following properties:

> Characteristically, mercury tends to (1) run together at speed, (2) form in pools, (3) re-form in pools if disturbed, (4) follow the rivulets and channels of any surface however smooth it may appear to be, and (5) is poisonous in small and large doses if poorly managed. (Clark, 2005)[p105]

These characteristics can make mapping global finance difficult. With the information available in the form 13F-HR, the best one can do to tie the command and control functions of the decision makers is to use the business address in which investors deal with the US regulatory system, and the Securities and Exchange Commission in particular.

2.4 The Time Period

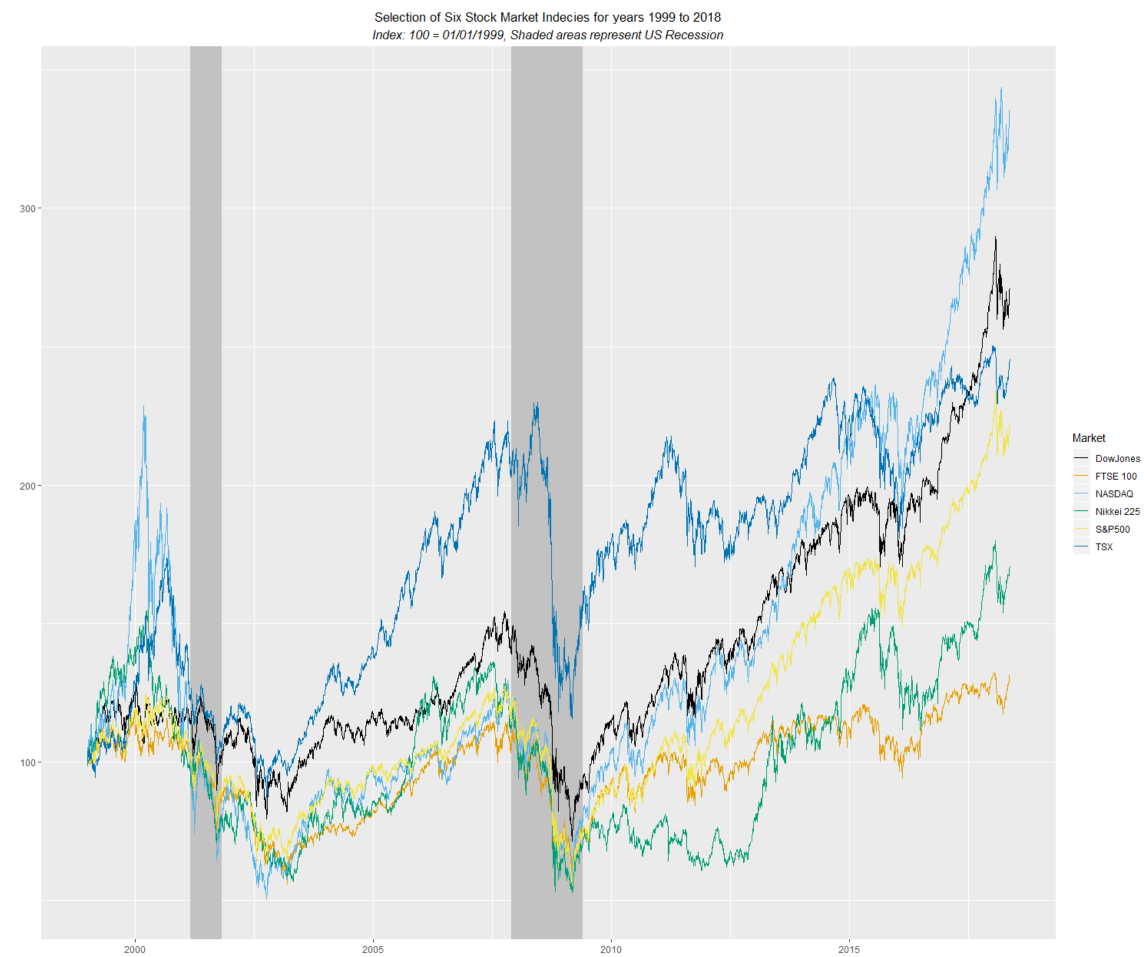


Figure 2.2: Collection of six stock indices for the years 1999 to 2018. The information was collected from Yahoo! Finance API on December 28, 2018. Shaded Areas represent recessions as defined by the National Bureau of Economic Research's Business Cycle Dating Committee `https://www.nber.org/cycles/cyclesmain.html`. The first recession dates from March 2001 to November 2001 and the second dates from December 2007 to June 2009.

Stock market indices provide a general guideline on the overall health of the stock market(Lo, 2016). From the investor's point of view, this is often used as a performance benchmark in which to evaluate their return *vis-a-vis* their peers. Figure 2.2 shows a collection of six stock indices. Three of these indices are used as bellweathers of the US Stock-Market: The Dow Jones Industrial Average (DJIA/DOW)[4], The Standard and Poors 500 (S&P 500)[5] and the National Association of Securities Dealers Automated Quotations Composite (NASDAQ Composite)[6]. The three other indices give insights to the national stock markets of various important regions for this study. The first is the UK's FTSE 100, Japan's Nikkei 225 and Canada's TSX.

Examining the correlations over time of various stock index is beyond the scope of this thesis, one would be remiss to forget to draw attention to the correlated nature of the various stock indices. That being said, being aware of the general nature of the stock market (Bear vs Bull market) gives context to whether growth in an investor's position can be partially explained by capital gains rather than attracting new clients and capital. More specifically, the 20 year period of 1999 to 2018 is an era that can be characterised as having strong overall growth, punctuated by 2 rather large financial crises - the DotCom crash of 2000, and the Great Financial Crisis of 2008-2009. As a consequence, this time period contains 2 powerful bull markets in which the market recovers powerfully from crash. The first being the mid-aughts economic boom, and the other the Obama recovery.

While stock markets are somewhat useful in determining the scope and duration of a recession, Samuelson (1966) oft-quoted quip of "the stock market has forecast

[4] The Dow Jones Industrial Average is an index of 30 blue chip US stocks covering the US economy except for transportation and utilities. The mix of 30 stocks has changed over time to reflect changes in the economy (S&P Dow Jones Indecies, 2020a).

[5] The S&P 500 is an index of 500 large-cap stocks that tries to be representative of the US economy (S&P Dow Jones Indecies, 2020b)

[6] The NASDAQ is a broad-based index of over 3000 stocks listed on the NASDAQ stock exchange. This index is heavily weighted towards the tech sector, and as such the "irrational exuberance" of the DotCom era cast a large shadow over this index, taking 15 years to surpass to the record highs that were recorded during this era (NASDAQ, 2018).

nine of the last five recessions" has a certain amount of truth to it. This is why the significant stock market correction that took place in 2016 isn't shaded as a recession in figure 2.2, since this did not have a significantly negative impact on the broader economy. This is why the National Bureau of Economic Research (NBER) does not have a fixed definition of what exactly constitutes a recession, going for an approach similar to Justice Potter Stuart's definition of obscenity - "You know it when you see it" (Jacobellis v. Ohio, 378 U.S. 184. 1964). As such, the NBER's Business Cycle Dating Committee is charged at taking a holistic view of the economy when determining the length and breath of a recession such as changes in employment, housing starts, payroll numbers, manufacturing output and aggregate hours worked in the economy rather than fixate of certain metrics such as stock market contractions or changes in Gross Domestic Product(Robert et al., 2020).

2.5 Conclusion

The section explores the data collection pipeline from the SEC's Edgar server to the decision to create two separate databases of 13F investors. As seen from the examples of inconceivable wealth declared in certain 13F-HR files due to various clerical errors, the data cleaning was an important factor in being able to trust the outputs of the models. Furthermore, the inability to trust the semi-structured text format lead to the creation of the "phone book" database and the more machine readable "XBRL" based database. The first database covers the time period of 1999 to 2018 and contains what is essentially phone book information such as years active and locations. The more detailed "XBRL" based database covers the time period of June 2013 to December 2018. This second database contains a detailed stock listing of their end of quarter holdings. Both databases were then geocoded using Google Maps API. Next, these databases were contextualized by exploring the time period in which they were active.

Chapter 3

Exploring the Data

3.1 Introduction

Statistician John Tukey is a strong advocate for exploratory data analysis (EDA). Collectively, EDA is a series of graphical and quantitative techniques used to explore novel data in order to examine its data structure, and thus generate insights that can be used as a springboard for hypothesis and model generation (Tukey, 1977; Hoaglin et al., 1983).

This chapter performs EDA on the data at various ground scales (country, state, core-based statistical area (CBSA), county and point) using a variety of techniques such as simple counts to more elaborate techniques such as Ripley's K and the gravity model of trade.

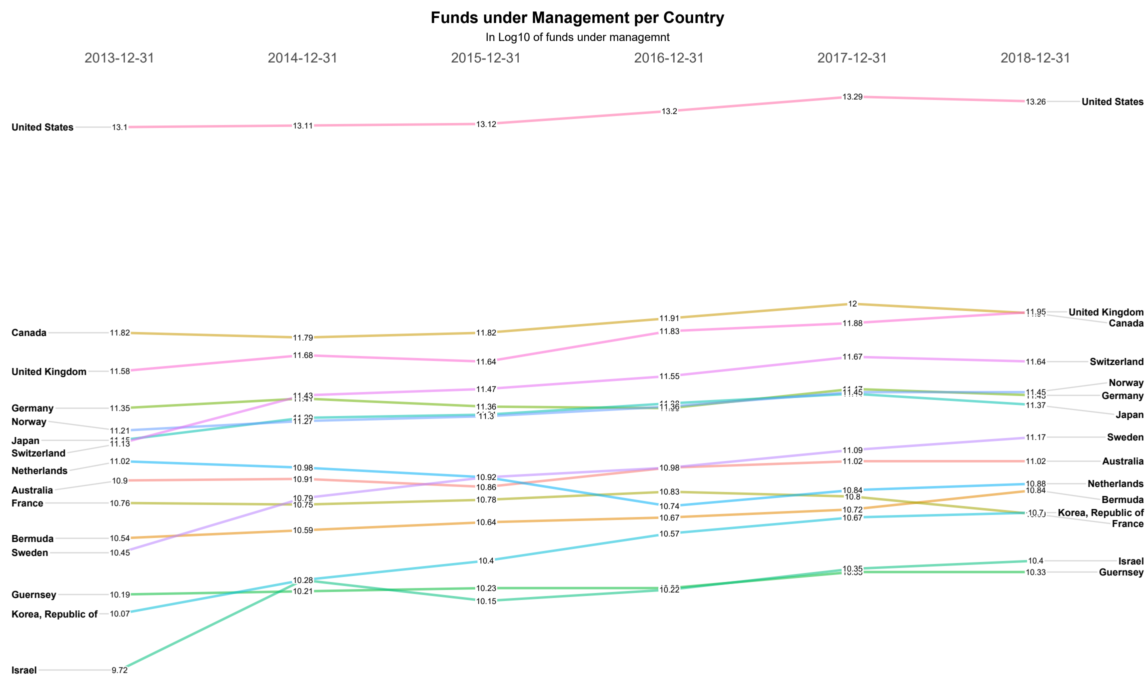


Figure 3.1: Funds under management by country/political entity for top 15 countries in the world by funds under management. Due to the very large gap between the USA and all other countries, the dollar value is represented in log10 form.

3.2 Count and Percentage by Region

Figure 3.1 is a slope graph showing the sum of funds under management for all firms headquartered in each country. As explored earlier, the 13F holdings report is a US legal instrument primarily interested in reporting the holdings of shares of US headquartered companies. It is no surprise that the United States of America is over-represented in this database. Furthermore, since many of the other countries on this list have their own robust domestic stock markets, one should take caution before making direct comparison between the US-based investors and foreign investors. Secondly, it is interesting to note that Canada, despite being a smaller economy than the United Kingdom, is home to more investors as measured by funds under management than UK based investors[1]. Finally, it is not surprising that the list of countries in Figure 3.1 are mostly populated by advanced economies and countries/political entities that specialise in financial services, such as Switzerland, Guernsey, and Bermuda.

3.2.1 Investors By State

While the 13F data has global reach with regards to foreign investors using US investment system, the use of domestic stock market is a significant confounding variable. Therefore, for practical purposes, the focus of this research will be centred to a greater extent on the United States of America, its commonwealths and oversees territories.

There exists institutional investors in every US State, however there is a very unequal distribution when it comes to their location, by both number of investors and funds under management. Wheeler and Mitchelson (1989a); Green (1995); Bodenman (2000); Graves (2003) have seen and forecasted the continued relative decline of New York, and specifically it's namesake city. And yet, despite the continual relative decline of New York State's position at the centre of the United States's financial system (Figure 3.2), New York State is still home to the largest growth in institutional investors in absolute terms for this time period (Figure 3.3). It should be noted that the renewal of New York's relative decline resumes on or around the first quarter of 2007. This will be discussed in further detail at the county level (Section 3.2.3) and in point pattern analysis using Ripley's K (Section 3.3).

[1] This is strictly true provided that Crown Dependencies (Guernsey, Jersey, and the Isle of Mann) and British Overseas Territory (Gibraltar and Bermuda being the most prominent) are excluded from the UK's total. With respect to the law, the Crown Dependencies are not part of the UK legislative and legal apparatus, and are autonomous with regard to their legal system, however the Crown is ultimately responsible for maintain good governance of these territories. (Ministry of Justice, 2018)

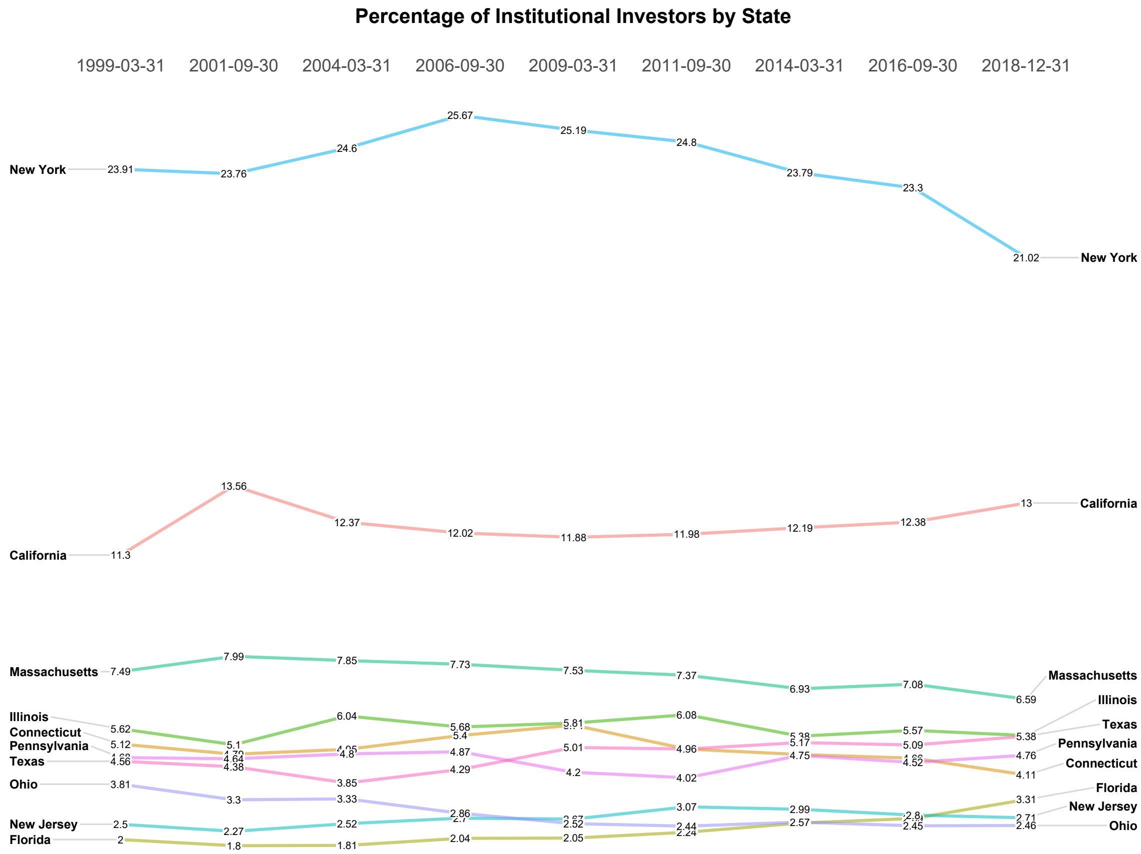


Figure 3.2: Percentage of Institutional Investors locational preference by share of investors by State.

While the region that contained the former industrial heart of the United States of America is experiencing a rather severe relative decline, these regions still manage to grow their number of firms in absolute terms. This suggests that the cause of relative decline is a slower genesis of new firms rather than a migration of footloose firms. This is consistent with the findings of Gong and Keenan (2012), which shows that despite large shocks, a firm's geographical preferences are sticky.

Further evidence for the point that firms are sticky can be found in figure 3.4, where the great circle distance was measured between the locations of the first and second, second and third, third and fourth, ect... locations of firms in the "phone-

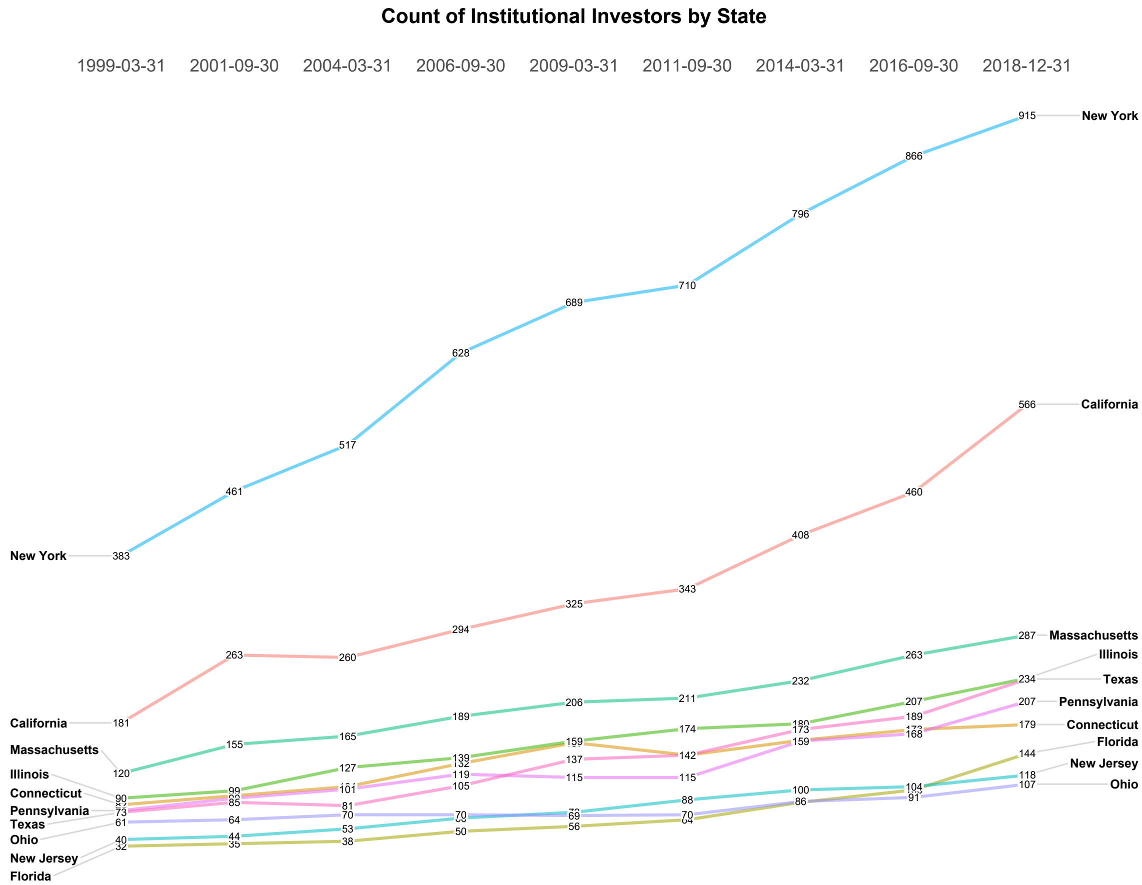


Figure 3.3: Count of Institutional Investors by State for the period 1999 to 2018.

book" database of 13F filers created by the author. In the database there are 14 922 unique location and CIK combinations, of which 5 603 firms (CIK) stay in the same location for the duration. For the remainder of 3 649 firms (CIKs), the database show them making 5 190 moves, for a total of 9 319 unique CIK/locations. While this 9 319 unique locations may make the moves to appear very footloose, one must remember that a move implies two distinct locations. In this database, the most footloose firm has a total of 7 moves, but this is the far end of the distribution as seen in Table 3.1.

During this time period, 4 917 of 5 190 (94.7%) of location changes have been what can be considered intra-city/intra-metro area (less than 150 km) rather than 47

Number of Moves	1	2	3	4	5	6	7
Count	2,477	886	221	52	9	3	1

Table 3.1: Number of Moves by an Institutional Investor between the years 1999 and 2018

inter-city. This lack of long-distance movement makes attracting firms to a new locale a near-non factor in location changes over time, suggesting some costs in movement, or that rent isn't a top-line deciding factor in location. Even more important for how sticky firms are in their locational preference are that 2 903 of 5 190 (55.9%) of firm locational changes are of less than 1 km in distance.

One would be remiss to not point out that movement can evade capture in this data set by closing down firm A in location Alpha and creating firm B in location Beta. However, since this would necessitate a non-negligible amount of paperwork, it is doubtful that this would occur only for the purpose of concealing changes in location.

A further cause for the widespread distribution of institutional investors in the United States is the historical legacy of US banking regulations. The 10th Amendment of the US Constitution reserved banking regulations to the States, whereas the commerce clause gave the Federal government jurisdiction over interstate commerce. This division in jurisdiction lead to the creation of a regime of regional banks rather than a small clique of national banks (Calomiris, 2000). Furthermore, the proliferation of State-managed employee pension funds ensures the existence of institutional investors outside of financially centred metropolitan areas such as New York, Boston, Chicago or San Francisco. This remains the case despite the recent trend of out-

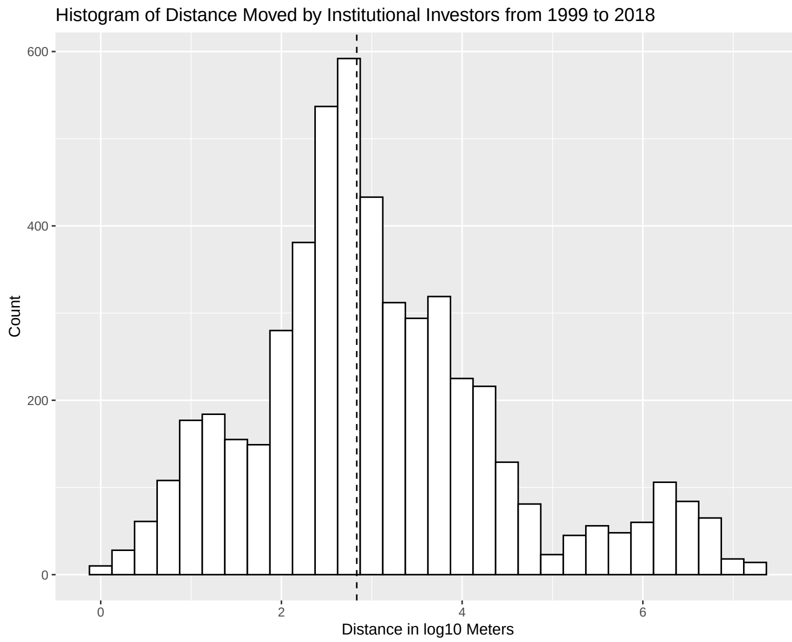


Figure 3.4: Distance Moved by firms during the time period of 1999 to 2018 in Log10 meters. The dashed vertical bar represents the median distance traveled of 680 meters and a mean distance of 269 040 meters (rounded to the nearest 10m).

sourcing a sizable portions of pension funds into more opaque (and thus outside of the purview of 13F disclosure) and hopefully high yielding private placement deals **?**.

3.2.2 Investors by Core-Based Statistical Area

Core-Based Statistical Areas (CBSA) are a relatively recent geographical construct by the U.S. Office of Management and Budget with the goal of creating a set of nationally consistent geographies that are useful for tabulating and comparing statistics. These areas consist of at least 1 core county with a population greater than 10 000 inhabitants, as well as all adjacent counties with substantial economic and social integration (US Census Bureau, 2016). The CBSA is a useful construct for comparing urban areas since it creates a more homogeneous unit of comparison between different urban areas in the United States, particularly since the USA has a disparate mix of regional sub-units such as New England townships and Louisiana parishes. Furthermore, the CBSA is subdivided into either a Metropolitan Statistical Area (population greater than 50 000) or a Micropolitan Statistical Area (population less than 50 000).

Figure 3.5 illustrates the absolute count of institutional investors by CBSA. As previously mentioned in the State breakdown of institutional investors in the previous section, the New York - Newark - Jersey City CBSA gains the largest absolute amount of new institutional investors by a considerable margin, and Figure 3.6 shows a similar picture to Figure 3.2 in which New York sees a relative decline. Due to the presence of a few investors in non-CBSA counties, the investors located outside of CBSA were added to figures 3.5 and 3.6. Of particular note is the rapid rise of investment firms outside of the USA during this time period. Figure 3.7 is similar to figure 3.6, but with the absence of foreign investment firms. When comparing these graphs, the difference in slope trajectory when the number of foreign firms is removed from the baseline is remarkable. At this scale, the relative density of investment firms still follows the same inverted U shape, with a peak on or about the first quarter of 2007.

3.2.3 Investors by County

Diving further down the building blocks of US territorial systems, the next level down is that of the county. There are 3 242 counties and county equivalents in the USA, and its territories, of which 2 707 do not have institutional investors during the entire period. In March of 1999, 2 972 counties do not host an institutional investor, however, by December 2018, the number of counties devoid of institutional investors

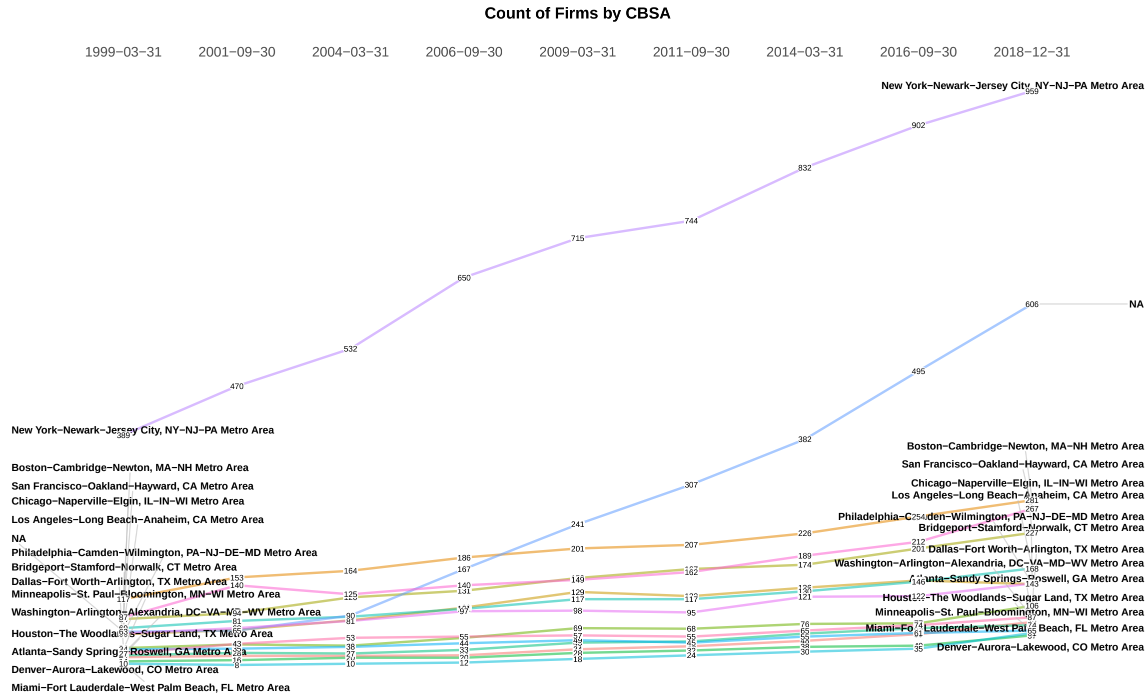


Figure 3.5: Count of Institutional investors by Core-Based Statistical Areas for the period 1999 to 2018

falls to 2 786. Considering that the USA added over 2 500 institutional investors during this period, this suggests that new institutional investors are attracted to counties with a pre-existing institutional investor population rather than filling-out empty counties.

This larger number of counties permits a different sort of analysis to be used: that of comparing Gini coefficients over time. The Gini coefficient is a common descriptive statistic of inequality, with a value of 1 describing perfect inequality (one case having all of the measured variable) and 0 describing perfect equality (all cases having equal amounts of the variable).

The Chow test is a statistical test developed by econometrician Gregory Chow for determining if two regression lines are equal. Within the field of time series analysis,

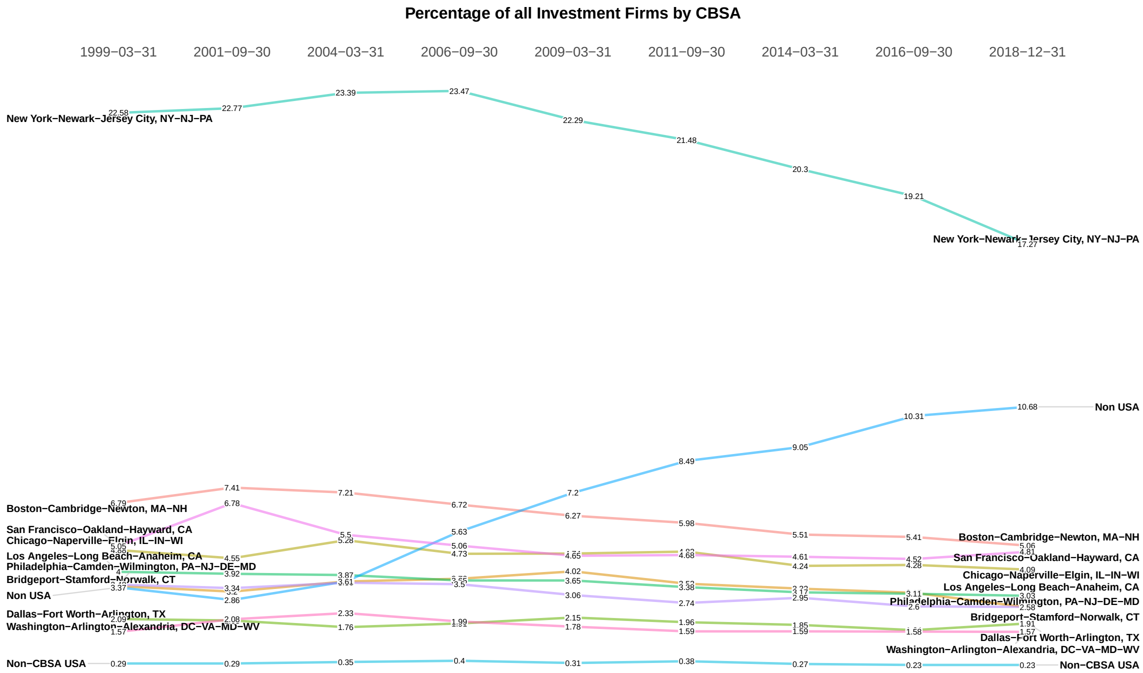


Figure 3.6: Share of Institutional investors by Core-Based Statistical Areas for the period 1999 to 2018

this is useful for determining if there is the presence of a structural break in the data. A look at figure 3.10 shows an increase in spatial dispersion over time. Using a Chow test (Figure 3.11) to find the change in linear trend of the Gini coefficient indicates that there is a breakpoint in trend on June 30th, 2011 (Chow, 1960). This is much later than the breakpoints mentioned earlier when looking at the concentration of firms in States and CBSAs. This can be somewhat explained by the Gini coefficient being more sensitive to areas going from 0 to 1 than say 15 to 16.

3.2.4 By County Urban Intensity Index

Counties (and their equivalents) are important building blocks in the American territorial administration. However, not all counties are created equal. For example,

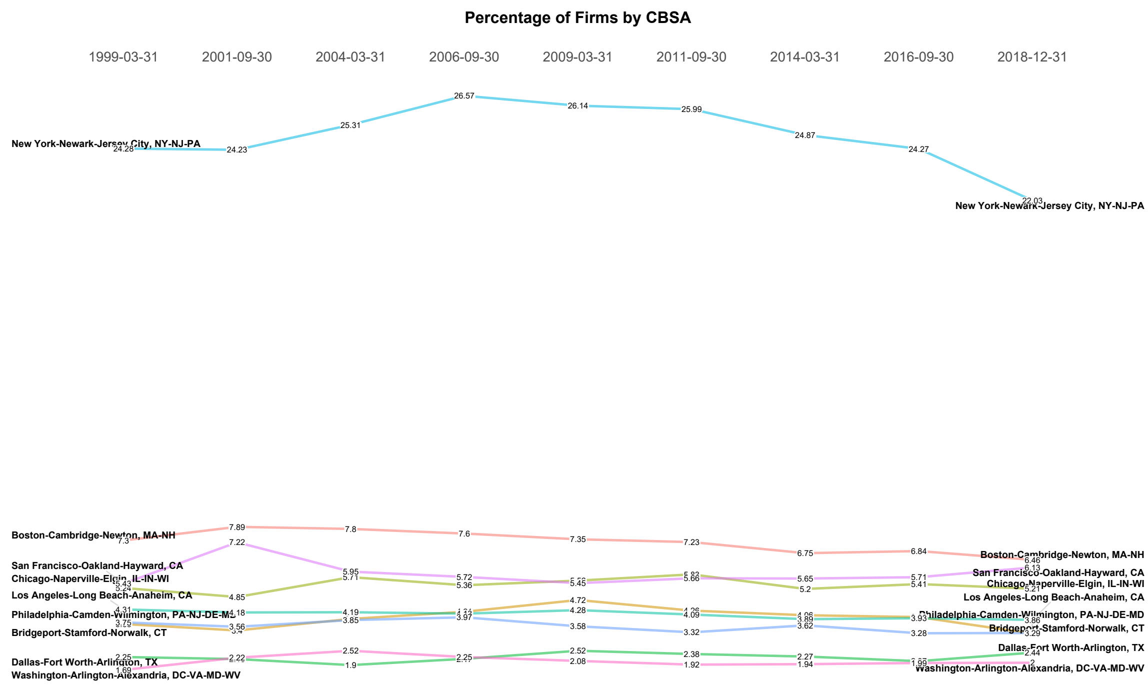


Figure 3.7: Percent by Share of Institutional investors by Core-Based Statistical Areas for the period 1999 to 2018

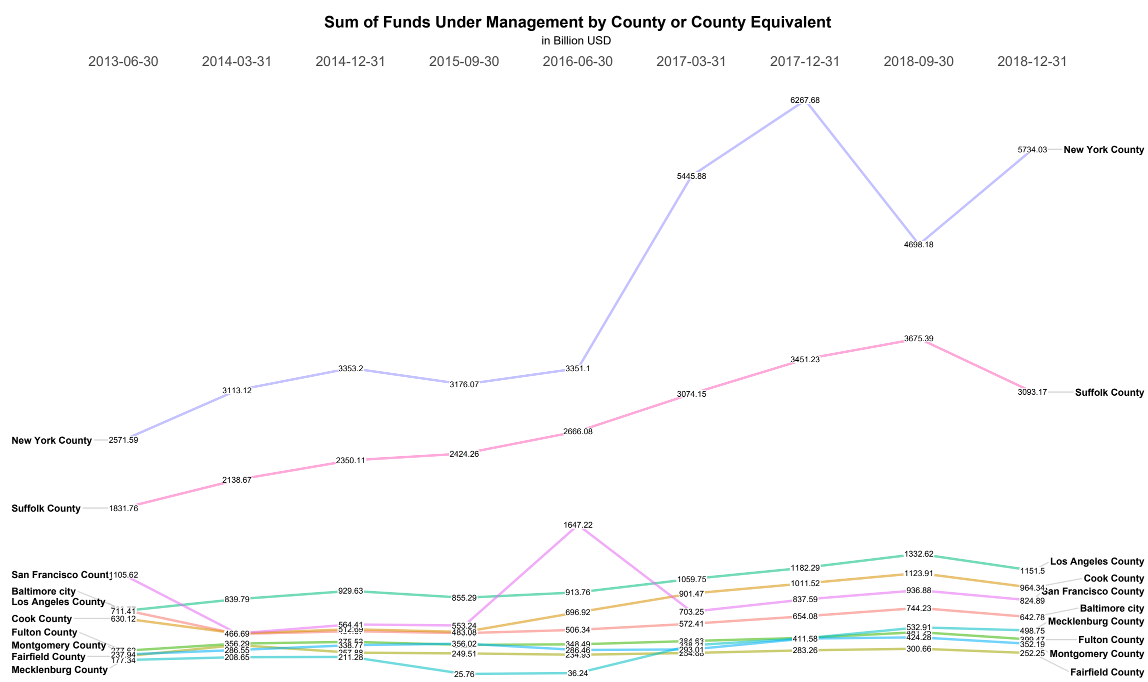


Figure 3.8: Share of Institutional investors by county for the period 1999 to 2018

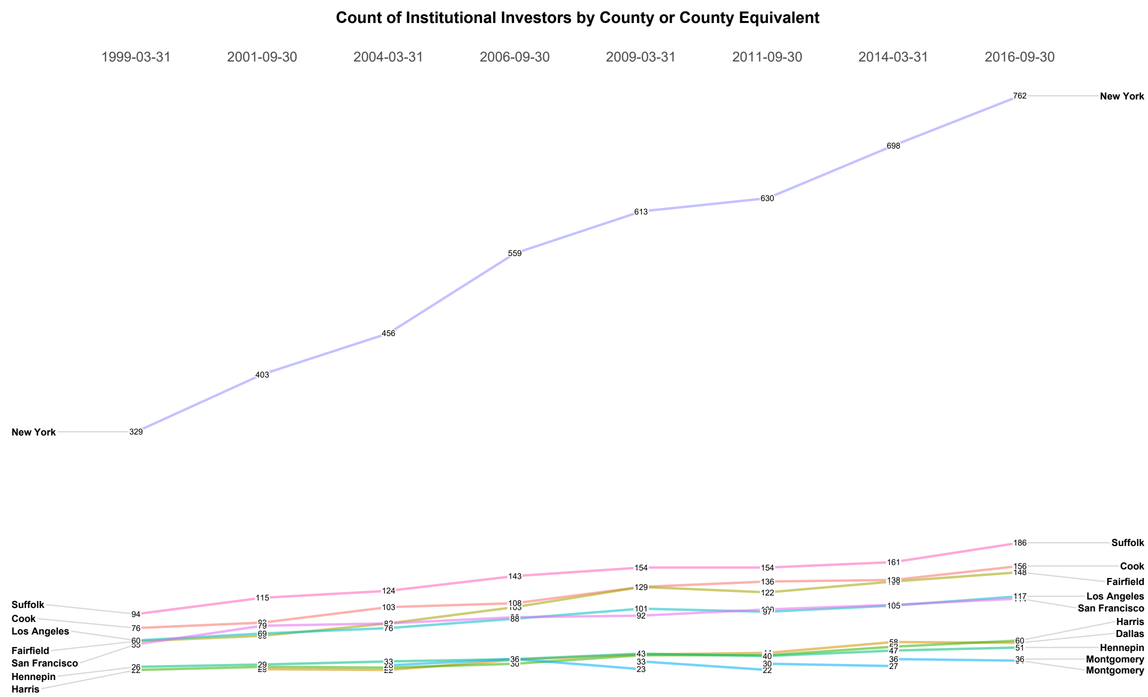


Figure 3.9: Count Institutional investors by county for the period 1999 to 2018

Los Angeles County in California has a population approaching 10 million people, whereas rural counties such as Loving County in Texas contains less than 200 inhabitants (U.S. Census Bureau, 2013). In the field of health geography and epidemiology, rural-urban divide can play a role in predicting health outcomes. The National Center for Health Statistics devised a classification scheme for all US counties that can be used as a proxy for the degree of urban surface area in each county (Ingram and Franco, 2014). This classifies counties into one of 6 different categories.

1. Metropolitan Categories:

   (a) **Large Central Metropolitan counties** (Category 1) are counties in Metropolitan Statistical Areas (MSAs) with at least 1 million inhabitants, and one of the following characteristics:

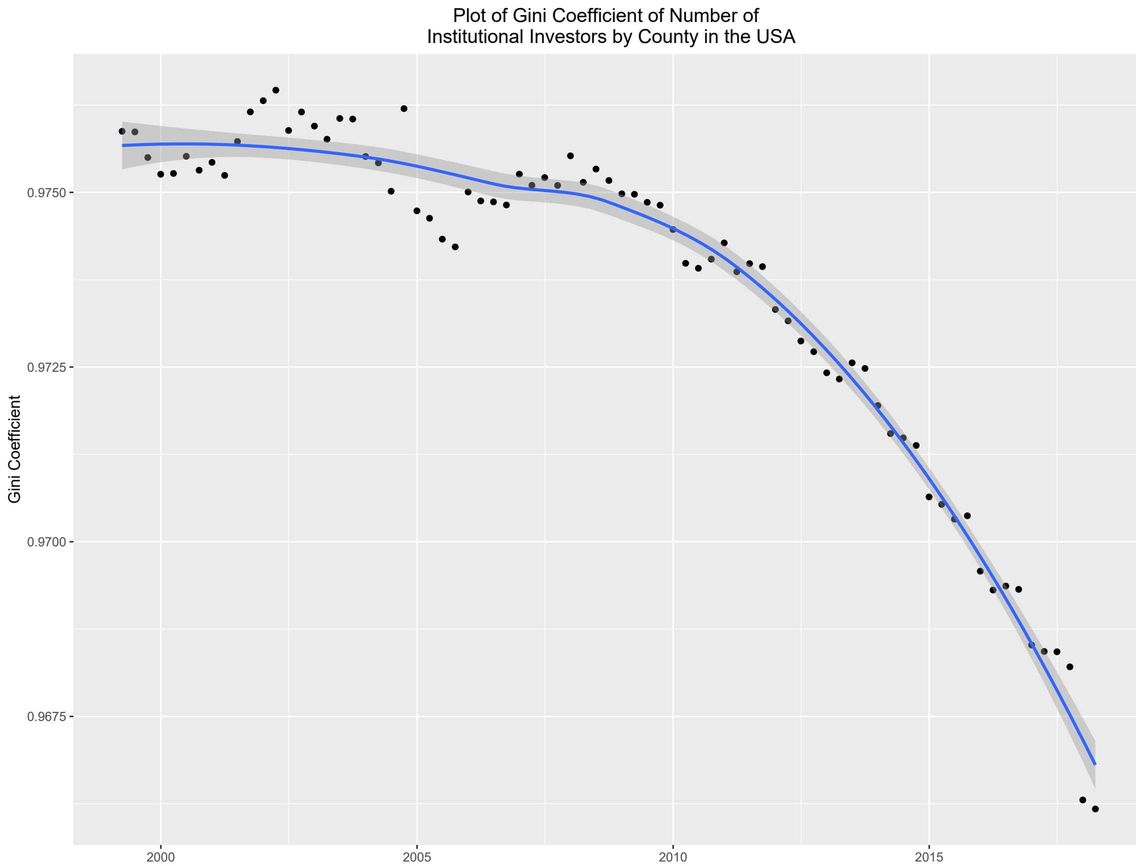


Figure 3.10: Gini Coefficent of US County Count

i. contain the entire population of the largest principal city of the MSA, or

ii. are completely contained within the largest principal city of the MSA, or

iii. contain at least 250,000 residents of any principal city in the MSA.

Examples: New York County New York[2], Bronx County New York, Los Angeles County California, Cook County Illinois.

(b) **Large peripheral metro counties** (Category 2) are counties in a MSA

[2]Coterminous with Manhattan Borough in the City of New York

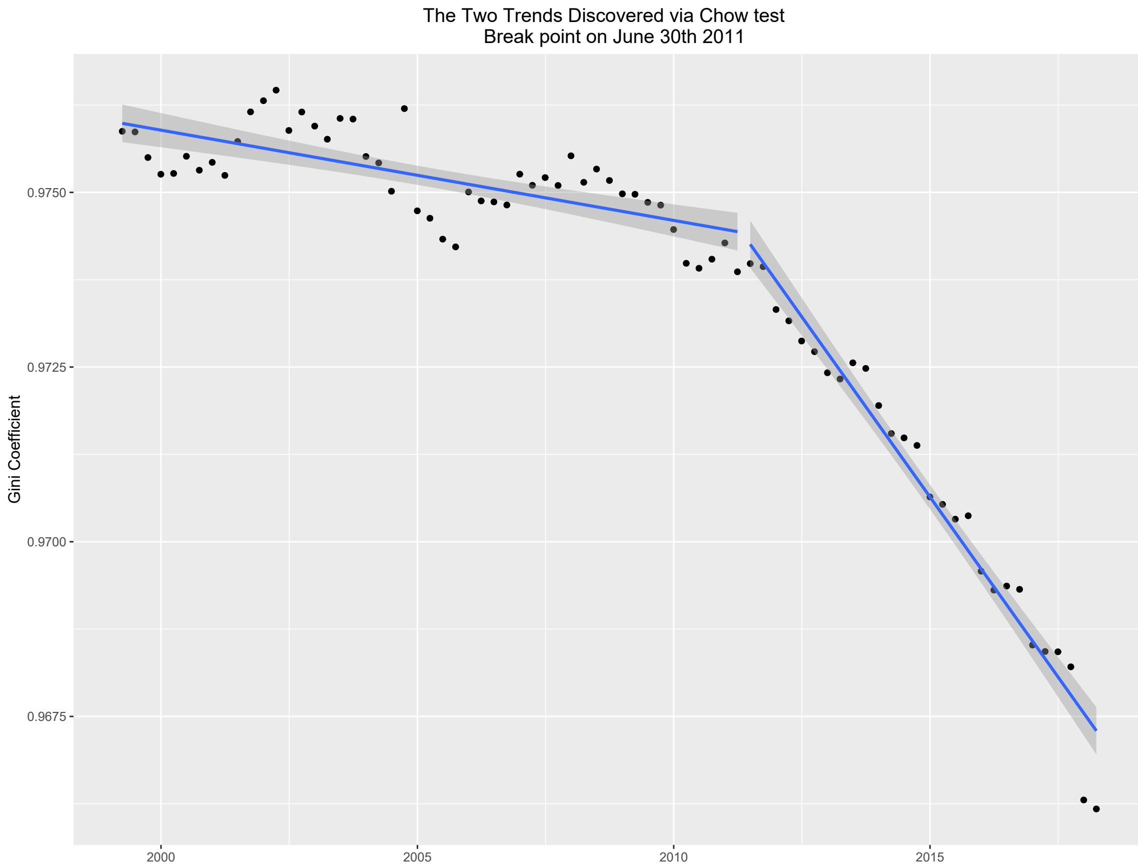


Figure 3.11: Result of Chow test. Breakpoint on June 30th 2011

with a population greater than or equal to 1 million, but do not qualify as category 1 county.

Examples: Orange County New York, San Mateo County California

(c) **Medium metro counties** (Category 3)are counties in MSA with a population greater than 250,000 but less than one million in population.

Example: Fresno County California, New London County Connecticut

(d) **Small metro counties** (Category 4)are counties in MSAs with populations greater than 50,000 but less than 250,000 in population.

Example: Yuma County Arizona, Franklin County Vermont

2. Non-metropolitan Categories:

   (a) **Micropolitan counties** (Category 5) are counties in a micropolitan statistical area

       Example: Juneau City and Borough Alaska, Talladega County Alabama

   (b) **Noncore counties** (Category 6) are counties that do not contain a micropolitan statistical areas

       Example: Loving County Texas, Denali Borough Alaska

This categorisation of counties gives insight into the type of region the new institutional investors prefer. As predicted by Quaternary Location Theory, it is hardly surprising that institutional investors are primarily found in large urban areas. This was also hinted in Figures 3.5, 3.6, and 3.7, where it shows that the majority of investors are clustered around the topmost cities in the USA urban hierarchy. Therefore it should be of no surprise that Figure 3.12 indicates that 95 percent of institutional investors are located in Metropolitan counties, and that the share of investors in Micropolitan counties is quite stable over time. The largest change is that category 2 counties see an increase in market share, which mostly comes at the expense of category 1 counties. This provides evidence that while downtown areas are slightly less attractive to investors, going for bargain basement land costs is also not a preferred strategy, or else we would see an uptick over time in the counts of category 5 or category 6 counties. While the relative gains of category 2 counties are impressive, one should not loose sight of the fact that the largest absolute growth in the number of institutional investors occurs in category 1 counties (Figure 3.13).

It should be noted that the drop in number of firms in the aftermath of the 2008 great financial crisis is of nearly equal proportion in all categories of counties (Figure 3.12). Yet it is quite evident when looking in absolute numbers of extant institutional investors (Figure 3.13) that category 1 counties take a longer period of time to reestablish their number of investors.

This growth in secondary counties in a conurbation may also hint a second phenomenon, such as an increase preference/and or availability of suburban office space in response to the expense of downtown offices. Pohl (2004) examines the remaining stock of real-estate in Manhattan after the terrorist attack on the World Trade Center and concludes that the destruction of World Trade Center Buildings 1, 2 and 7, as well as the damage on the other buildings essentially removed nearly a quarter of Manhattan's tier 1 and 2 office space from the market, and that the resulting scramble for office space tightened Manhattans' office market, spilling over into the other 4 boroughs as well as suburban New York, New Jersey, and Connecticut.

3.2.5 Investors By Region

A way to reconcile the decline in market share of the New York region in Figures 3.2, 3.6, and 3.7 with Figures 3.12 and 3.13 is to ask if the traditional definition of State or CBSA is too narrow, and that the declines may be partially explained by the modifiable area problem (MAP). The MAP is a source of statistical bias in geography-based data aggregation, since boundaries on reporting areas can have an outsized influence (Fotheringham and Wong, 1991). A common extreme case of the

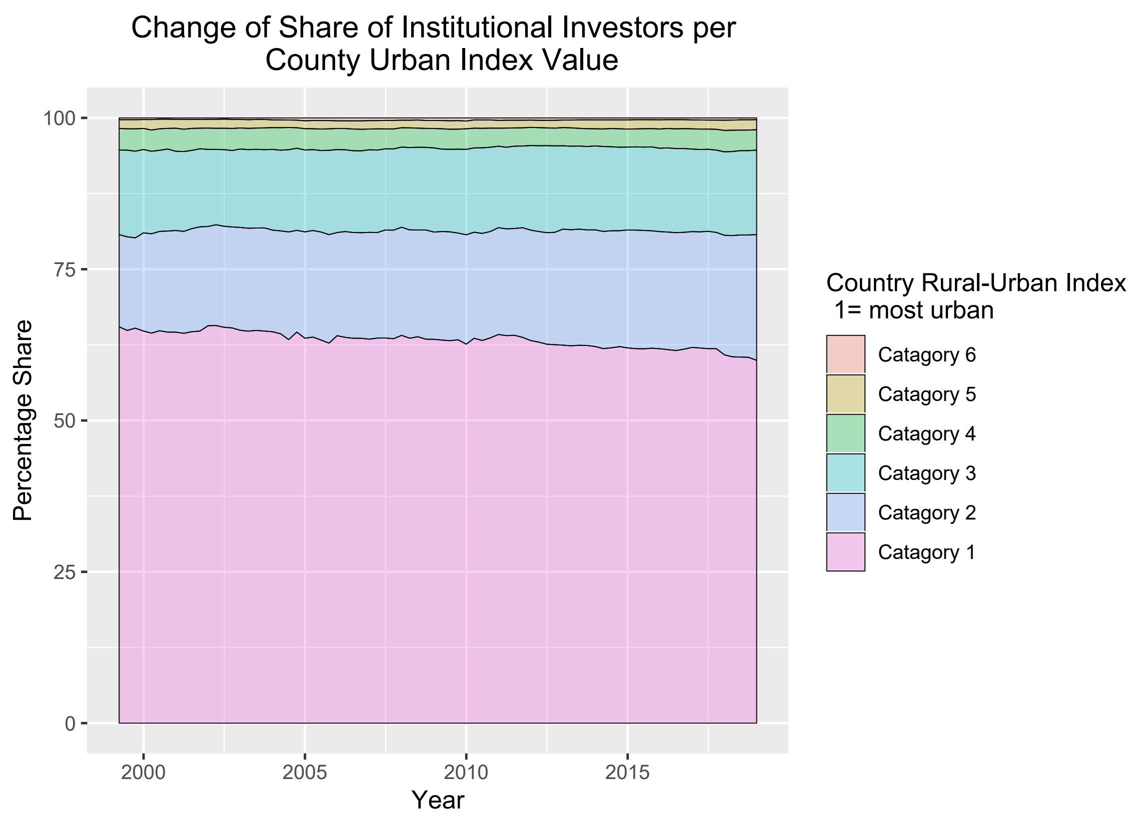


Figure 3.12: Percentage Share of firms by County Urban Index Value from 1999 to 2018

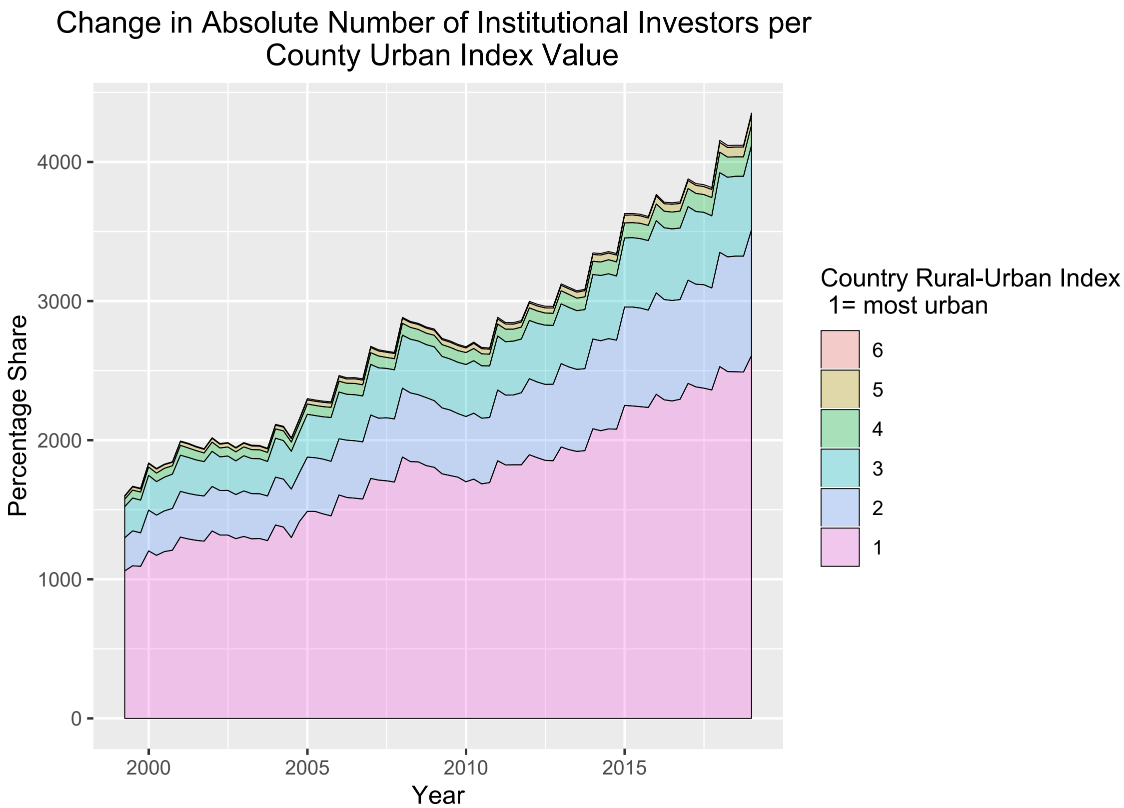


Figure 3.13: Count of firms by County Urban Index Value from 1999 to 2018

MAP is gerrymandering, in which a political party can gain more seats relative to its vote share by controlling how the votes are aggregated into different districts. In this case, all of the levels of aggregation seen so far (State, CBSA and County) fail to holistically capture Megaregions in the USA, and in particular the Boston-New York-Washington (Bos-Ny-Wash) megaregion (Lang and Nelson, 2007). While it may not fully encompass the Bos-Ny-Wash, the US Census Bureau's Region[3] does a good approximation of this.

In the case of Figure 3.15, the decline of the North East is much slower than one would expect from previous graphs, mostly due to the inclusion of the south shore of Connecticut and the North shore of New Jersey.

Increases in the number of Southern-based investors lies mainly in the growth of firms located in the DC/Arlington Virginia region, as well as Atlanta. With regards to the decline of the Mid-West, as mentioned previously, this is more of a relative decline than an absolute decline, for while it started the study period with 321 (20%) institutional investors and ended with 728 (17.6%).

3.3 The K-function

One of the earliest uses of point pattern analysis is the famous cholera map by Dr. John Snow. Although he knew nothing about the cause of the bacterial outbreak, he did discover that the cases of cholera were clustered around a particular water pump on Broad Street. Although scholarship such as Brody et al. (2000) call into

[3] `https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf` for a map listing the geographies encompassed by the different regions

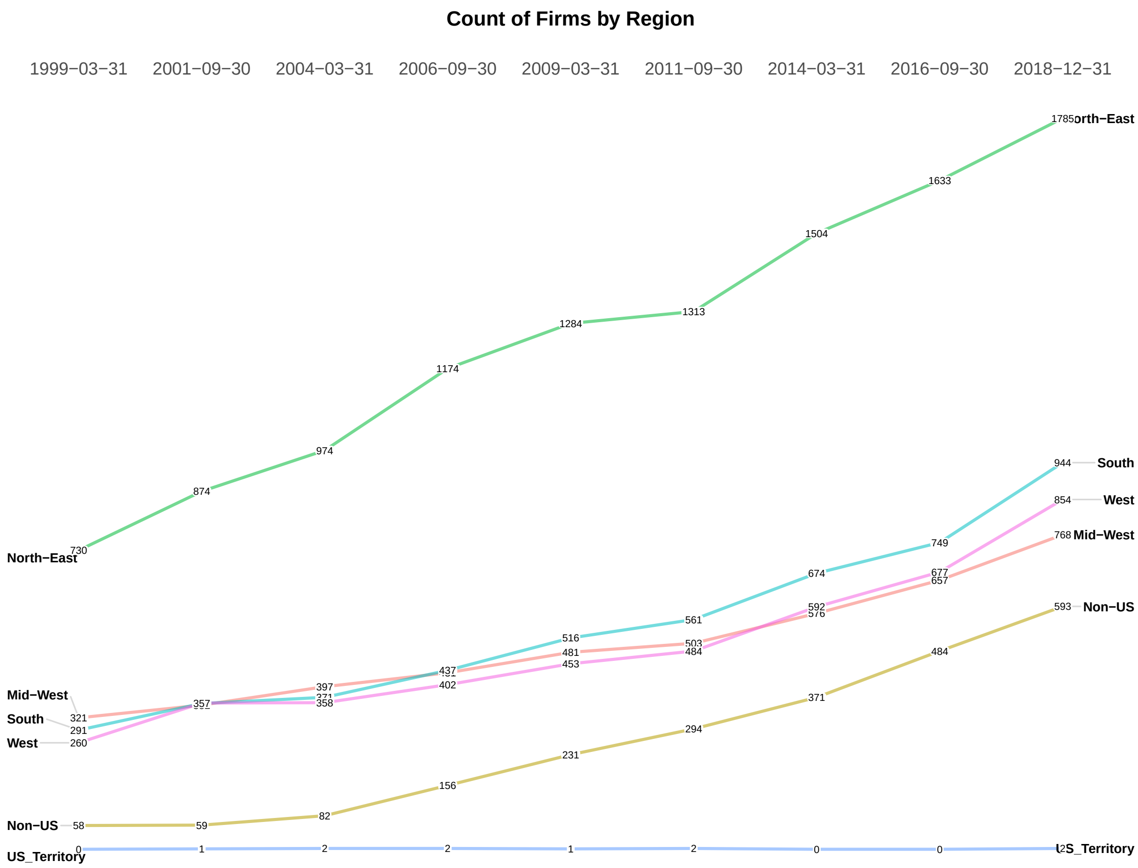


Figure 3.14: Relative percentage of institutional investors by region during the study period (March 1999 to December 2018)

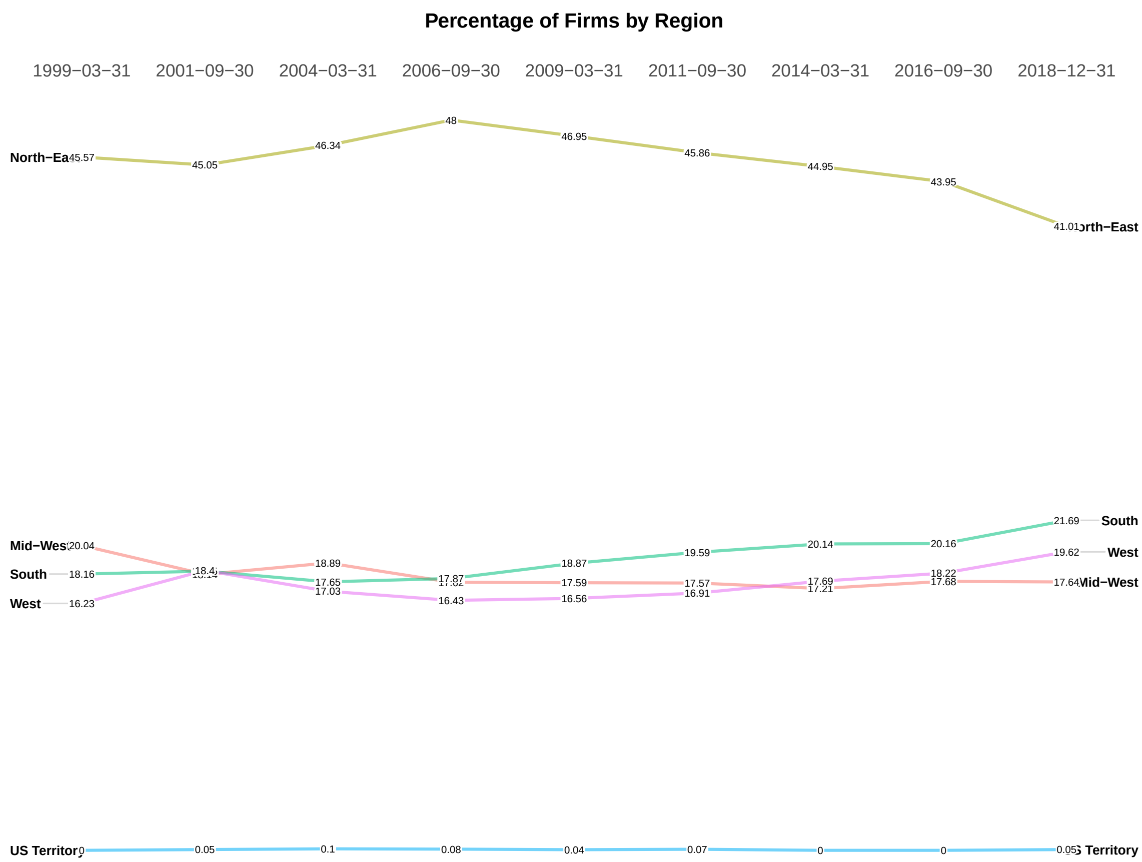


Figure 3.15: Number of Firms by Country
[Total Number of firms by region (as defined by the US Census Bureau) during the study period (March 1999 to December 2018)]

question whether Dr.Snow's map was more confirmatory than exploratory since the insights into the cause of the cholera epidemic requires an understanding of germ theory. That is to say, that Dr. Snow's maps would not be able to create their historic insights without subject matter expertise. Regardless of whether Dr. Snow used his point density mapping technique as a starting point or only for confirmation of his hypothesis, a common method of quantifying points in space, is measuring the intensity of the point pattern per unit of area. Old staples used for measuring point patterns are quadrat analysis and nearest neighbour index. However, these techniques have well known limitations such as the undue influence caused by border selection as well as the inability to determine whether points cluster or disperse at different ground scales (Baddeley et al., 2015).

The examination of various ground scales is important since firms may exhibit different clustering tendencies at various scales. The mirroring of population maps and geographic profile maps at a national scale is humorously examined in XKCD comic 1138 (Figure 3.16) (Munroe, 2012). However, firms may behave differently at different scales. For example, a national maps of firms such as coffee shops, fast food chains, banks, automated teller machines, gas stations and grocery stores may mirror the national population map, yet they would appear diffuse on a local map, for each operates their own local catchment areas. However, other sectors such as software development have a tendency to cluster at the local and regional level .

At its most basic form, the K-function calculates using a Poisson process of actual vs expected counts of points within distance h of each point in the data set(Dixon, 2014). This yields a density function, which can be compared to the expected point

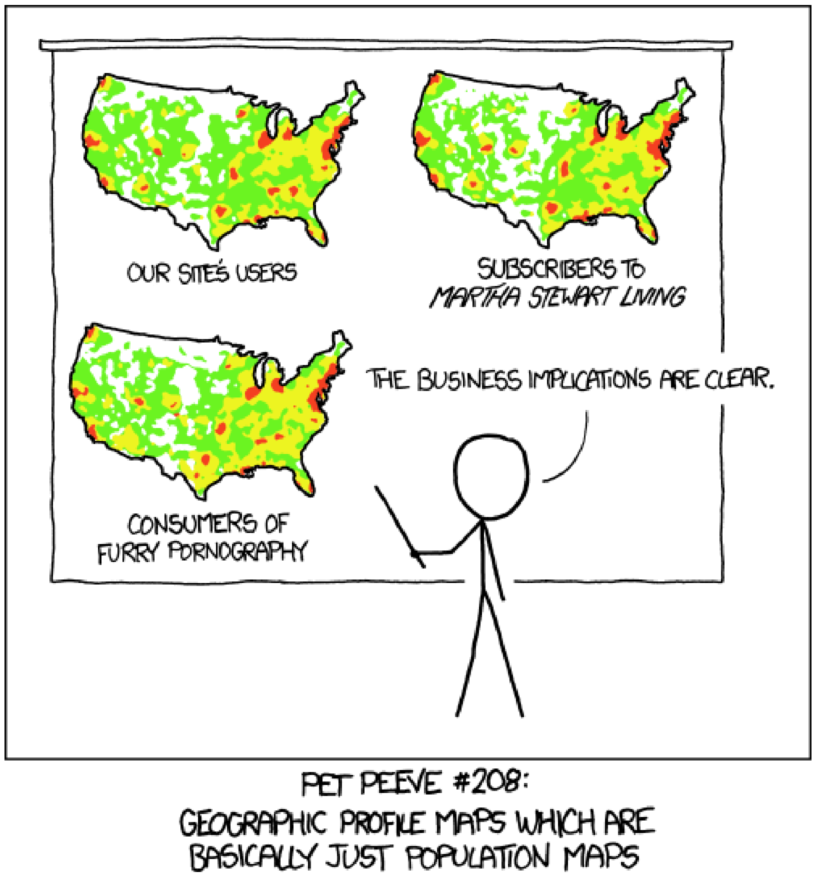


Figure 3.16: XKCD #1138 - Heatmaps by Randall Monroe. This illustrates the point that many paterns can be aproximated by human density. Used with Permission (Creative Commons Attribution-NonCommercial 2.5 License)

pattern intensity under the conditions of complete spatial randomness at different distances. For more information about *Ripley's K*, see Ripley (1976); Fischer and Getis (2009); Baddeley et al. (2015).

With regards to examining the clustering behaviour of institutional investors at various scales, the inherent ability to be used at various ground scales makes Ripley's K well suited for examining the clustering behaviour of institutional investors. This facilitates the examination of spatial clustering of institutional investors and determines if they exhibit locational preferences closer to that of ATMs or software developers.

The K-function in it's most basic form can be written as follows:

$$K(d) = \lambda^{-1} E(Nd) \tag{3.1}$$

Where Nd is the number of events Xi within distance d of a randomly chosen event from all points $\{X_i, ..., X_j\}$. When working with a sample of data points $\{X_j\}$, the K-function for the underlying distribution isn't usually known. However, it can be estimated by using a sample. If d_{ij}

$$\hat{K}(d) = \hat{\lambda}^{-1} \sum_{i} \sum_{i \neq j} \frac{(d_{ij} < d)}{n(n-1)} \tag{3.2}$$

$$\hat{\lambda} = \frac{n}{|A|} \tag{3.3}$$

The CSR equation

$$K_{csr}(d) = \pi d^2 \tag{3.4}$$

Brunsdon and Comber (2015)

3.3.1 Spherical K-function

The basic implementation of *Ripley's K* technique assumes that the point pattern exists on a Euclidean surface. While it may be justifiable to assume a Euclidean plain for regions of less than a few hundred kilometres (Lynch and Moorcroft, 2008; Wilschut et al., 2015), the use of Euclidean space becomes problematic above such distances, and the global distribution of institutional investors is certainly more than a few hundred kilometers, and thus spherical geometry becomes a better option. Furthermore, Tobler (2002) demonstrates that while the Earth is technically an oblate spheroid, most statistical techniques on a continental scale can be done adequately on a sphere.

The K-function displayed in Figures 3.17, 3.18, 3.19, and 3.20 were performed in statistical language R using Robeson's implementation of spherical geometry (Robeson et al., 2015). This analysis was conducted with a 99-fold cross-validation, in which for each time step, the 1/99 of the data was randomly reserved from the data set[4]. This creates an envelope of possible K-functions. Particular care should be noted for the third and fourth quarters of 2004. These quarters were run a second time with a similar result, suggesting that the problem may lie with the data pipeline from Edgar

[4]The calculation of the K function for the 80 quarters involved in this study was performed on 3 different computers for a duration of 3 months for a total of 9-computer/months calculation time

rather than a sudden and reversible shift in locations preference. A similar, but less extreme discontinuity exists between the fourth quarter of 2013 and the first quarter of 2014.

As with the other forms of measuring the concentration and dispersion at various scales seen earlier, the overall trend of initial concentration and dispersion on or after 2007 continues with the K-function. In greater detail, Figure 3.17 looks at the 1 km scale, where there is an initial concentration followed by a gradual diffusion starting on or around 2003. Figure 3.18 shows a slightly different picture, more akin to the County and CBSA graphs of concentration from 1999 to on or about 2007 and an increased diffusion afterwards. Figure 3.19 shows a similar pattern - just not as starkly. Finally, Figure 3.20 shows that the continental scales resemble the shape seen in Figure 3.17, since there is a continual diffusion of firms from a earlier peak.

This is an important confirmation of the trend, since Ripley's K is a point pattern analysis, and is thus immune to the modifiable areal unit problem. This suggests that something fundamental in the business world occurred in the time-frame of the pivot that changed the calculus in terms of benefits of the forces of agglomeration and desegregation. There is precedence in the location preferences shifting in the past, with a substantial amount of dispersion occurring in the 1970s and 1980s when the first telecommunication revolution occurred (Bodenman, 2000).

While it is beyond the scope of this research, it would be interesting to examine if the rise of so called business-oriented "smartphones" by Blackberry (formerly known as Research In Motion), touchscreen smartphones such as "iPhone" and "Android" devices, in addition to widespread wifi-enabled cafes have reduced the productivity tax of conducting business away from the office, and thus reduce the costs of locating outside of the central business district.

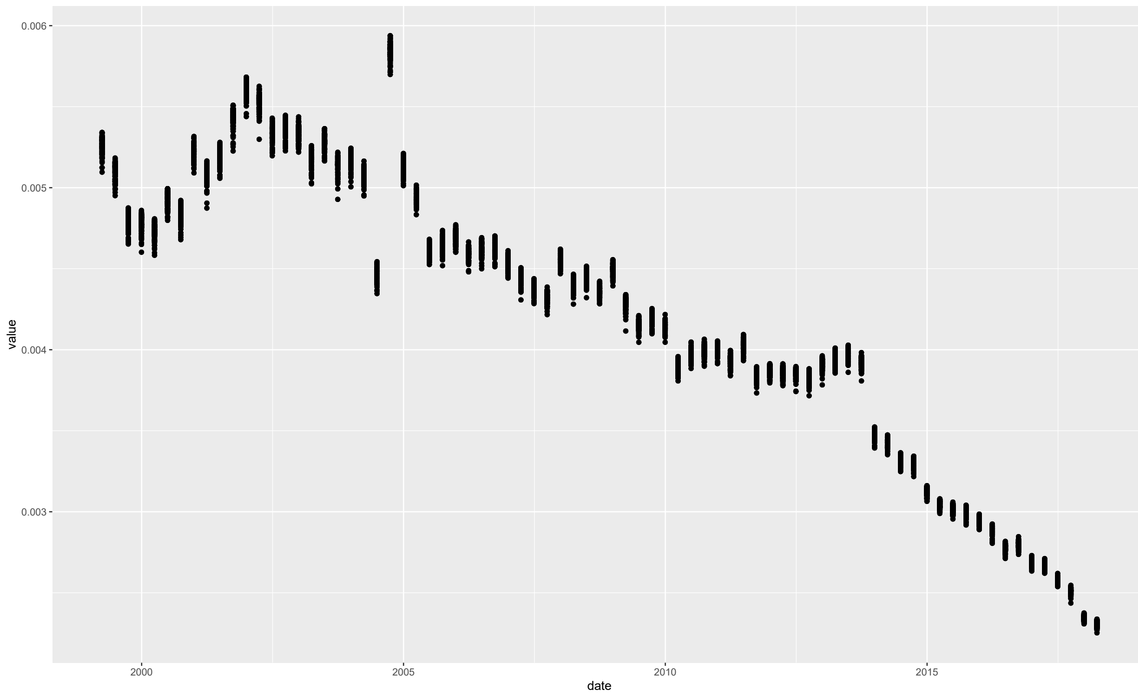


Figure 3.17: Spherical K-function for the range band of 1 km for the years 1999 to 2018. Each quarter consist of 99 points representing a cross-validated K-function.

3.4 Standard Deviation Ellipsis

The standard deviation ellipse is a useful tool in measuring the dispersion of a point pattern, and comparing the same region at different points in time in order to draw insights about a particular phenomenon (Yuill, 1971). The standard deviation ellipse, and its simpler cousin the standard deviation circle, create a line that enclose one standard deviation of all points from the centre of all points. The surface area contained by this line allows researchers to characterise the concentration or diffusion of a phenomena, and multiple such ellipses allow for the examining of a trend. Figure

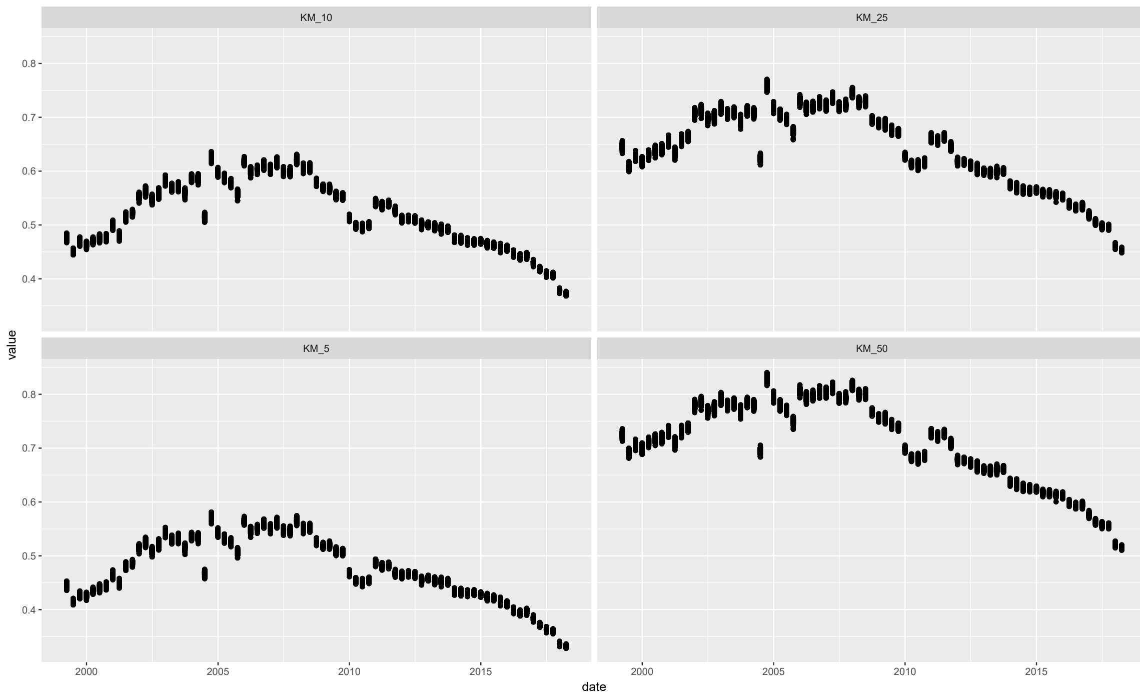


Figure 3.18: Spherical K-function for range bands 5km, 10km, 50km, 100km for the years 1999 to 2018.

3.21 shows the evolution in time of the standard deviation ellipse for the 5 cities that contain the largest amount of institutional investors. As is consistent with the other measures of clustering examined so far, these 5 cities show a long-term trend towards diffusion as investors start to show up in suburban office parks. Furthermore, while it can be somewhat imprecise to directly compare densities across cities due to the vagaries of urban planning, it is almost impossible for Los Angeles County to appear denser or more concentrated in a particular metric than New York County (Manhattan). It is immediately apparent that Los Angeles' famous urban sprawl and lack of a proper CBD create a rather diffuse concentration of institutional investors - a point that will be visited in more detail in the following chapter.

That being said, the advantage of the standard deviation ellipse over the standard

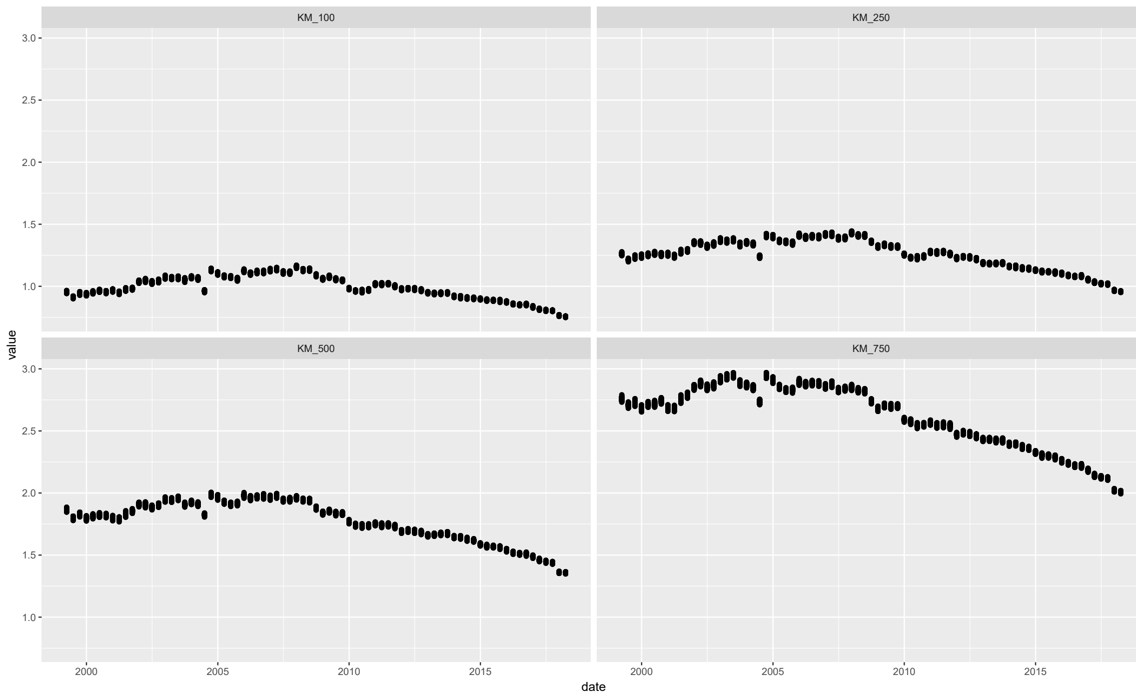


Figure 3.19: Spherical K-function for range bands 100 km, 250 km, 500 km, 750 km for the years 1999 to 2018.

deviation circle is the addition of orientation and eccentricity. Eccentricity is measured on a scale of 0 (perfect circle) to 1 (perfect line). An increase in eccentricity with a commensurate increase in surface area is suggestive of a new cluster being created near the perimeter of the ellipse. This is the case with regards to Boston in the mid-aughts in the Route 128 corridor (Figure 3.22). This will be examined in further detail Chapter 4.4.

For the detailed metrics of the one standard deviation ellipse, see Appendices B.1 for Boston, B.2 for Chicago, B.3 for Lost Angeles, B.4 New York City and, B.5 San Francisco.

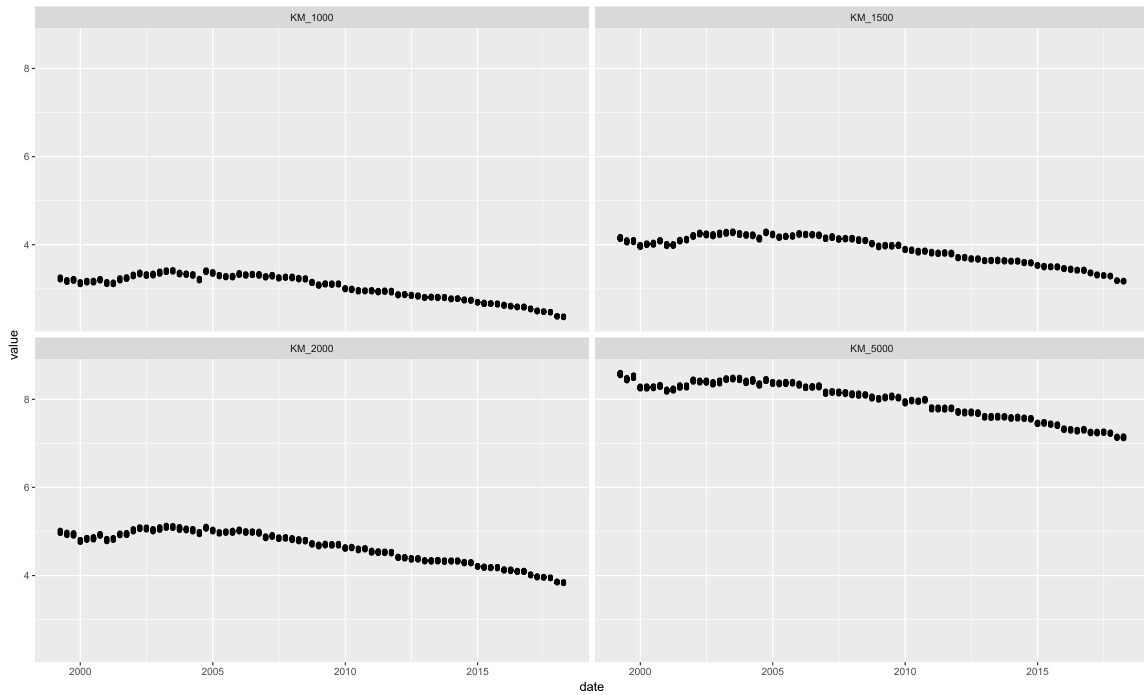


Figure 3.20: Spherical K-function for range bands 1000 km, 1500 km, 2000 km, 5000 km for the years 1999 to 2018.

3.5 The Gravity Model of Trade

Gravity model of trade is an empirically derived technique to describe and predict flows from a variety of origins to destinations. One of the first researchers to propose a model for explaining flows of population across space is Ravenstein (1885). He identified a series of "laws" of migration, while not explicitly referencing Newtonian gravity, identified the key variables of distance as well as push and pull factors (Tobler, 1995).

The most naive way of allocating flows across a land mass is to assume a uniform distribution. However, this is questionable at best, for this disregards a myriad of variables that can be used to account for differences in trade. Nobody would seriously expect that trade between New York County, New York (Manhattan) and Loving

Figure 3.21: Standard Deviation Ellipse over time for the top 5 CBSAs by number of institutional investors for the time period of 1999 to 2018.

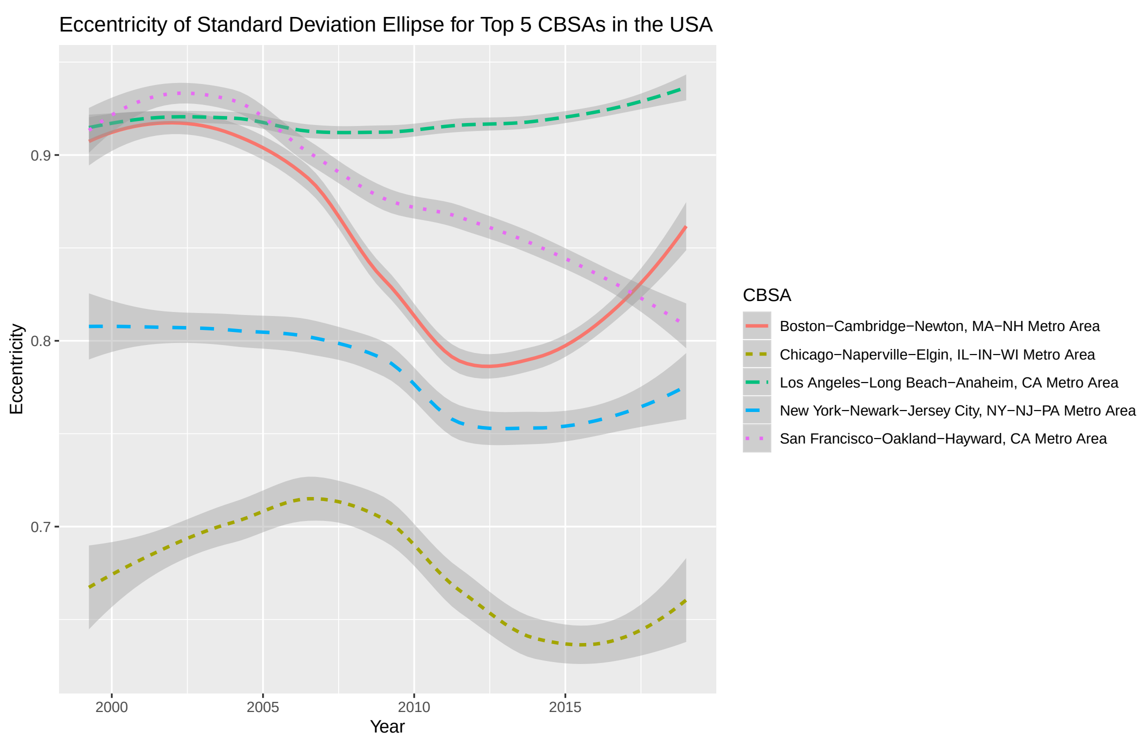


Figure 3.22: Standard Deviation Ellipse Eccentricity over time for the top 5 CBSAs by number of institutional investors for the time period of 1999 to 2018.

County, Texas to be on the same level as that between New York County, New York and Los Angeles County, California. Standardizing the flow by a variable such as population might help, but there's no guarantee that the flow scales solely with population (Crymble, 2019).

The most naive version of the gravity model is as follows:

$$F = G\frac{M_1 M_2}{r^2} \tag{3.5}$$

Equation 3.5 is inspired by Sir Issac Newton's gravity equation. As with the gravity equation, F represent the trade in goods from points M_1 to M_2. M_1 and M_2 represents the aggregate push and pull factors and is traditionally measured as the size of each's market. r^2 is the square of the distance between these points and G is a constant representing the friction of trade, such as the conditions of the roads, the productivity of the longshorepeople or tariff regimes. Unlike the theoretical apple falling from a tree (or the spherical cow thrown by a frictionless trebuchet in a vacuum), human endeavours are plagued by free will and the myriad of uncertainty that follows. However, in an insight that could have come from Isaac Asimov's Hari Seldon, populations are easier to model than individuals, since the vagaries of human existence averages out in the aggregate.

A gravity model's goal is to tell the user: Given a number of influencing forces (distance, costs of living, desirability, access to services, access to markets) affecting the movements of a large number of entities of the same type (fungible commodities or similarly situated people) between a set number of points, what is the most probable distribution? Furthermore, comparing real-world flows to the model's prediction can be used to find anomalies, and these can be useful starting points for future research(Crymble, 2019).

With respect to the gravity model, one must make sure that the data is either complete or a representative sample of the underlying flows, else the model will be hopelessly biased. In this case, the model will be using the universe of 13F holdings for the period of June 2013 to December 2018 to create flows between investors and to the company in which the stocks belong. The destination information is drawn from the COMPUSAT database of stock information filings, and more specifically, the address of their headquarters which was subsequently geocoded using Google Maps (Capital IQ Compustat, 2019). The push and pull factors were calculated as the total stock ownership in the 13F database for each quarter in each CBSA. CBSAs were used for this analysis rather than States ($n = 50$) or counties and their equivalents($n = 3,142$) due to the CBSA's occupation of a "sweet spot" with regards to detail and manageability ($n = 935$, of which there are 465 CBSAs which contain at least one flow).

The resultant flows matrix was quite porous, with X of Y cells being otherwise empty. This poses a problem for the model, since zero is undefined when transformed by logarithm. A quick and dirty remedy for this is to add a dummy transaction of 1/10 000 of USD to each CBSA. For each cell that would otherwise reported zero flow now reports 0.005 USD in flows. While the value of 0.005 USD is too small to be represented in hard currency, this value will give a defined value when transformed.

Table 3.2: Gravity model of trade as applied to investment flows between US CBSAs for the period of 2013Q2

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.70^{***}	-7.46^{***}	-6.41^{***}	-29.94^{***}	-6.20^{***}	-29.19^{***}
	(0.10)	(0.10)	(0.10)	(0.18)	(0.10)	(0.18)
Distance(log)	-0.09^{***}	-0.13^{***}	-0.23^{***}	-0.31^{***}	-0.26^{***}	-0.33^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Invest. at Origin(log)	0.23^{***}	0.22^{***}	0.18^{***}	0.14^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Invest. at Destination(log)	0.20^{***}	0.20^{***}	0.15^{***}	0.12^{***}	0.15^{***}	0.12^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin Is State Capital		1.86^{***}			1.82^{***}	1.45^{***}
		(0.04)			(0.04)	(0.04)
Dest. is State Capital		0.93^{***}			0.70^{***}	0.18^{***}
		(0.04)			(0.03)	(0.04)
Origin Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin Population(log)				2.50^{***}		2.38^{***}
				(0.02)		(0.02)
Destination Population(log)				2.40^{***}		2.38^{***}
				(0.02)		(0.02)
R^2	0.24	0.25	0.33	0.31	0.34	0.31
Adj. R^2	0.24	0.25	0.33	0.31	0.34	0.31
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.17	5.14	4.85	4.93	4.82	4.91

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

3.5.1 Discussion

In total, 6 models were run for each quarter for a total of 114 total models. Since there is very little quarter to quarter variation between model runs, only the model for the second quarter of 2013 (June 30th, 2013) will be discussed here. The results of the other quarters are available in Appendix C.

The first model is the most naive model possible where only the distance between CBSAs, as measured CBSA centroid to CBSA centroid, as well as the investment capital available in each origin and destination are considered. Consistent with previous literature such as Green (1995); Graves (1998); Coval and Moskowitz (1999, 2001); Dvořák (2005), model 1 (Table 3.2) shows a significant distance decay function in the flows between different CBSAs. Furthermore, this naive model can explain 24 percent of the variance seen in the network of flows.

Examining the residuals of the naive model, the largest outliers are where the model drastically underestimated the flows between large cities with robust financial centres, such as Boston to New York, San-Francisco to New York, New York to San-Francisco, New York to Boston. At the other side of the outliers the model has trouble factoring eccentric portfolio choices, such as foundations being bequeathed large amounts of a single stock. One such notable example is the Kellogg W. K. Foundation Trust, for it is a holder of a large amount of Kellogg stock located in the relatively rural city of Battle Creek Michigan, the historical home of the Kellogg Corporation, yet shows no ties to nearby large financial centres such as Chicago or New York, as well as mid to lower tier financial centres such as Detroit or Minneapolis-

Saint Paul.

Models 2 through 4 build on the naive model by adding an extra explanatory variable. In the case of model 2, binary variables were added to the model representing if the CBSA contained a State capital. This was added in order to control for the observation that many State pension funds are located in their Capitol city (at least from an administrative capacity) rather than in a nearby financial centre, such as the various New York State employees and teachers pension funds being controlled out of Albany NY rather than New York City. Similarly, one can point to the California Public Employees' Retirement System (CalPERS) and the California State Teachers' Retirement System (CalSTRS) being run out of Sacramento California rather than San Francisco or Los Angeles. Unsurprisingly, model 2 shows this to be a significant factor in predicting monetary flows. This is consistent with the literature such as Bradley et al. (2016) that examines the role of State-level power brokers in fostering a suitable business environment.

Model 3 adds the human population of the CBSA as a variable, while model 4 adds this population transformed by the logarithm of the population. Here the untransformed population count is a better predictor variable of flows than the log of the population when looking at the adjusted r^2 and Root Mean Square Error (RMSE).

Models 5 and 6 are kitchen sink approaches, where all of the explored explanatory variables are included in the model. It should be noted that the human population of the origins and destinations are not examined at the same time as the log of human population since this would be in effect measuring the same thing twice, and thus unbalancing the model by adding covariates.

Taken as an ensemble, Model 5 has the lowest residual mean square error and hightest r^2. This model suggests that there is definitely is a distance decay function with regards to investing.

3.6 Conclusion

This chapter performs an exploratory treatment of the data from various geographic scales and using simple geographic techniques. Across the different scales of analysis (state, CBSA, county, and point), and technique from simple counts to more computer intensive techniques such as the K-function and standard ellipse, there is a broad agreement that overtime the locational preferences of investors steer toward slightly less concentration, while still maintaining a decidedly major metro area preference. This time period shows a continued relative decline of New York City within the American hierarchy of financial cities. However, it is important to note that this decline is only relative, and that New York City is still the number one location for new institutional investors in the absolute sense.

Lastly, the gravity model of trade as applied to institutional investors suggests that distance plays a part in investment flows, and that distance decay can be measured. Furthermore, the less naive models continue to show the importance of State Capitals and large metro areas with regards to locating institutional investors, suggesting that institutional investment continues to play a strong command and control function within the American and world economy.

Chapter 4

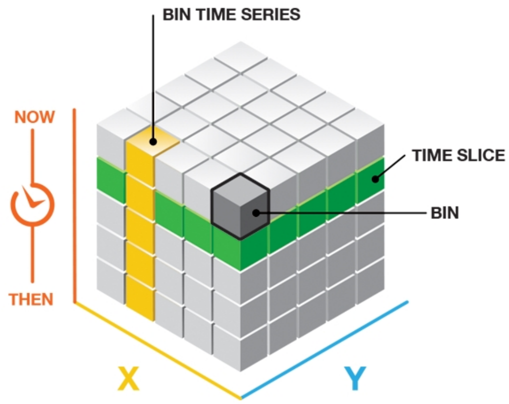
Space Time

4.1 Introduction

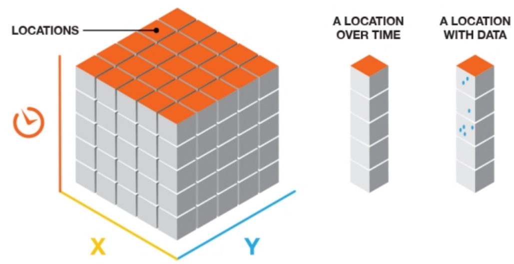
The previous chapter shows that institutional investment is mostly an urban phenomenon. This chapter examines the evolution of institutional investors across space and time. Furthermore, for ease of statistical analysis, both databases will only draw from investors located in the continental United States (CONUS), as well as for the top 5 core-based statistical areas (CBSA) in terms of total institutional investment. In alphabetical order, these 5 metro regions are Boston, Chicago, Los Angeles, New York City and San Francisco.

4.2 Space-Time Cube

The space-time cube is a space-time analytical technique that bins point objects into a space-time grid in order to examine the relationship between points not only in



(a) A schematic explanation of the time-cube



(b) Examination of a timecube cell

Figure 4.1: Schematic Illustration of a TimeCube. It should be noted that unlike this schematic representation of the time cube, the analysis in this paper uses a hexagonal bin rather than a square bin for spatial data. Image from: `http://desktop.arcgis.com/en/arcmap/10.3/tools/space-time-pattern-mining-toolbox/visualizing-cube-data.htm`

space, but across time ESRI (2019). Two types of space-time cubes are created, the first one aggregates the total number of institutional investors for the time period of March 1999 to December 2018. The second space-time cube aggregates the total number of funds under management for the period of June 2013 to December 2018.

The first step in creating a space-time cube is the creation of a Network Common Data Form (NetCDF) file. This file format permits ArcGIS to store multidimensional information with a defined geographical position (x and y) alongside a defined time period as well as any additional relevant information such as count data, sum, average, median and standard deviation. This creates a data-structure in which further analysis can be performed, such as emerging hotspot analysis and local outlier analysis. Figure 4.1 provides two perspectives on the data aggregation process.

It should be noted that unlike Figure 4.1a and 4.1b, this analysis was run using hexagonal bins. Unlike the traditional square bins (or in Esri's parlance, a fishnet grid), the hexagons have multiple advantages over squares, such as: of the three geometric forms that can tessellate (repeat a shape over and over without overlap), the square, the hexagon and the equilateral triangle, hexagons have the lowest perimeter to area ratio. This is due to hexagons being the closest of the three tessellating shapes to a circle. As such, this reduces the border effect when binning points, since the hexagon has the shortest average distance between perimeter and centroid. Furthermore, the centroids of hexagons are equidistant from each other when tessellated. This cannot be said about squares in a grid using the queen's movement, for the distances between centroids in square bins are shorter along the rook's movement than the bishop's movement due the the Pythagorean theorem. Lastly, at larger distances hexagons suffer less distortion than squares. Unfortunately for square bins, the implementation of spatial bins in this project does not play to its strengths, such as ease of use when conducting matrix algebra and having an orthogonal coordinate system (Birch et al., 2007).

With regards to the time dimension of the data, the dates are aligned such that bins coincide with the last date in the datasets (December 31, 2018), and work backwards from there in 3 month intervals. As such, each temporal bin covers one filing period for 13F-HR disclosures. (Figure 4.1b)

4.2.1 Emerging Hot Spot Analysis

Emerging hot spot analysis is the space-time implementation of the Getis-Ord Gi* statistic (Getis and Ord, 2010), and examines whether high or low values cluster geographically. High g values are created when the local sum and that of its neighbours are significantly larger than their proportion to the global sum, with low values in the reverse case. The ArcGIS implementation of Emerging Hot Spot Analysis performs the False Discovery Rate (FDR) correction. FDR accounts for multiple testing, and therefore compensates for the possibility that certain features would be classified as hot or cold by chance alone (ESRI, 2019).

The next step is to perform Mann-Kendall trend test to detect temporal trends at each spatial location. Depending on the results of the Getis-Ord Gi* statistic and the trend direction from the Mann-Kendall test, there is a total of 17 possible answers, and their definitions are listed in Table D.1 in Appendix D (ESRI, 2019).

4.2.2 Local Outlier Analysis

Local outlier analysis is the space-time implementation of the Anselin Local Moran's I statistic. This tool identifies concentrations of high values (high-high), low values (low-low) in addition to spatial ouliers in which high values are surrounded by low values (high-low), and low values that are surrounded by high values (low-high). Unlike traditional Anselin Local Moran's I statistic, the local outlier analysis variant offers a 5th category, in which it flags bins that have different Anselin Local Moran's I statistic values during the timeframe.

4.3 United States of America

The first use of space-time analysis will focus on the United States as a whole, after which the basic analysis will be repeated on the five largest metro areas.

When creating the NetCDF file for the United States of America, the size of spatial bins was set at 50 km. This value was chosen since this permitted a local window with a radius of 300 km according to the ESRI implementations of Emerging Hotspot Analysis and Local Outlier Analysis. This latter figure is important since it would represent the longest possible day trip during a business day (Fritsch and Schilder, 2006). Furthermore, we should keep in mind that the 50 km range band showed one of the highest level of change over time with regards to the K-function.

4.3.1 Count Data

Figure 4.2 shows the results of the emerging hotspot analysis using the address book database. These results should come as no surprise after reading the previous chapter, in which the vast majority of institutional investors are located in the New York, Boston, Chicago, Los Angeles and San Francisco regions. After all, institutional investment is a decidedly urban phenomenon despite being a theoretically footloose industry in an era of wireless telecommunications and computerized stock trading. In addition to these regions, there is some strong, but inconsistent growth in the Texas Triangle (a megaregion that encompasses San Antonio, Dallas-Fort Worth and Houston), the Miami-Dade region of South Florida, the Ohio Valley and the Raleigh Triangle (Raleigh, Durham and Chapel Hill, North Carolina).

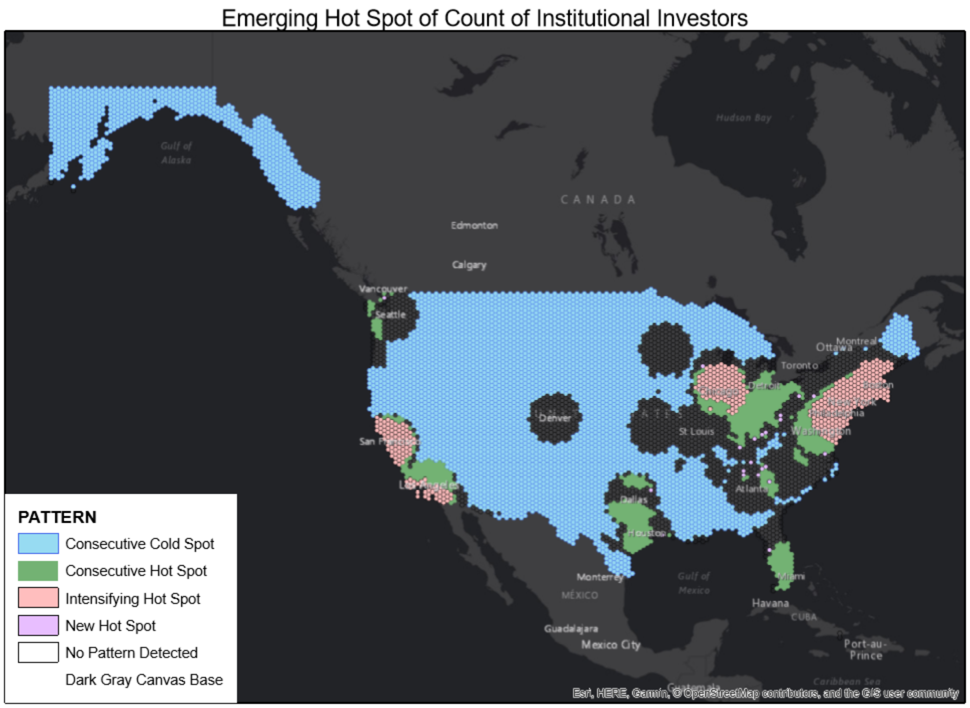


Figure 4.2: Emerging Hot Spot Analysis of locations of Institutional Investors in the United States of America for the period of March 1999 to December 2018.

Painting a similar picture than Figure 4.2, the local outlier analysis (Figure 4.3) indicates that the cities of New York, Boston, Chicago, Los Angeles and San Francisco are high-high clusters.

What is also of interest, is the light sprinkling of high-low clusters in Figure 4.3. These light blue dots coincide with secondary and tertiary financial centres as well as State capitals where State-employee pension funds are managed. Low-high clusters appear to be confined to bridging the gaps between nearby high-high clusters, such as the peripheral areas of the North-East mega-region. These low-high clusters are not unexpected, since they are definitionally low areas surrounded on multiple sides

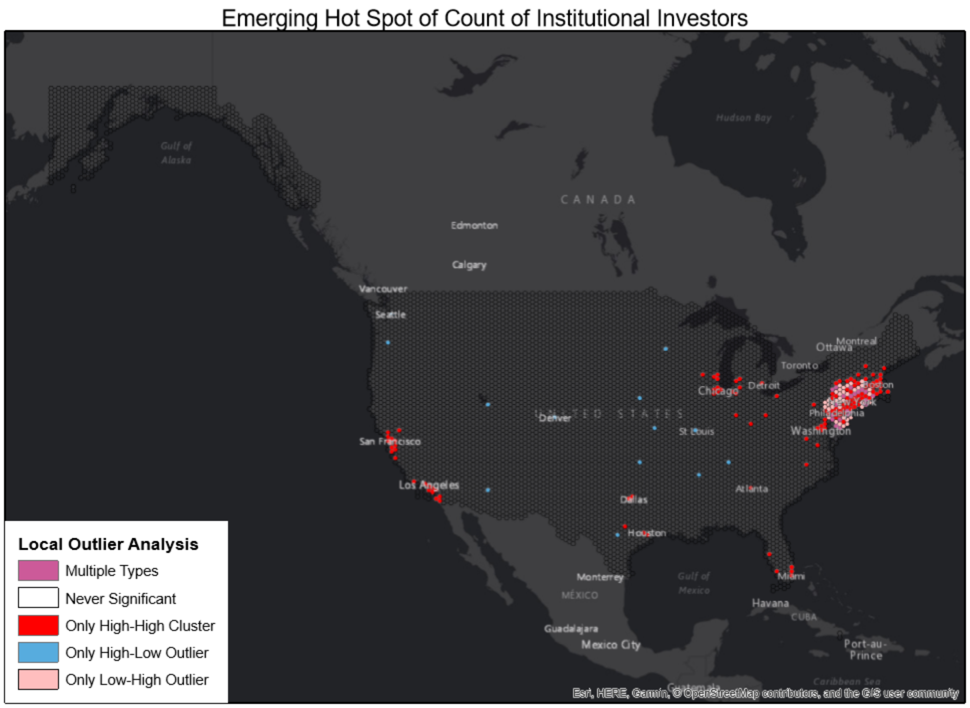


Figure 4.3: Local Outlier Analysis for Number of Institutional Investors in the USA for the time period March 1999 to December 2018

by high areas.

4.3.2 Funds under Management

Using the same technique on the holdings database presents a slightly different outcome as seen in Figure 4.4. Using money under management rather than count data puts more emphasis on New York and San Francisco, while at the same time removing all of the consecutive cold spot areas and turning them into regions with no detectable patterns.

As with Figure 4.4, which is based on the holdings database, the local outlier

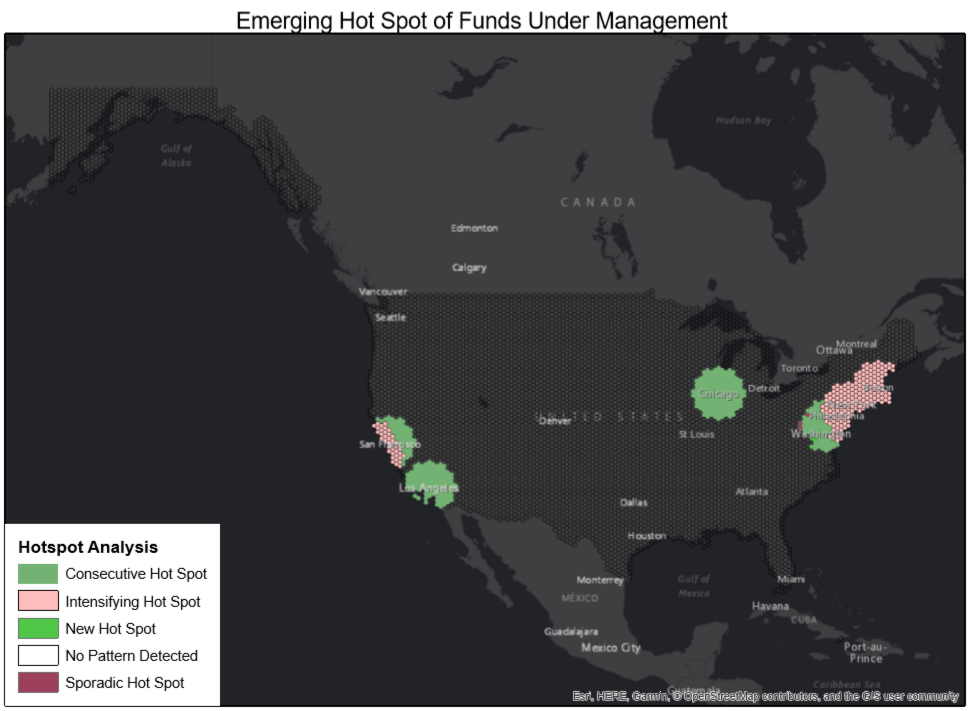


Figure 4.4: Local Outlier Analysis of USA-based Institutional Investors located in the United States of America for the period of March 1999 to December 2018.

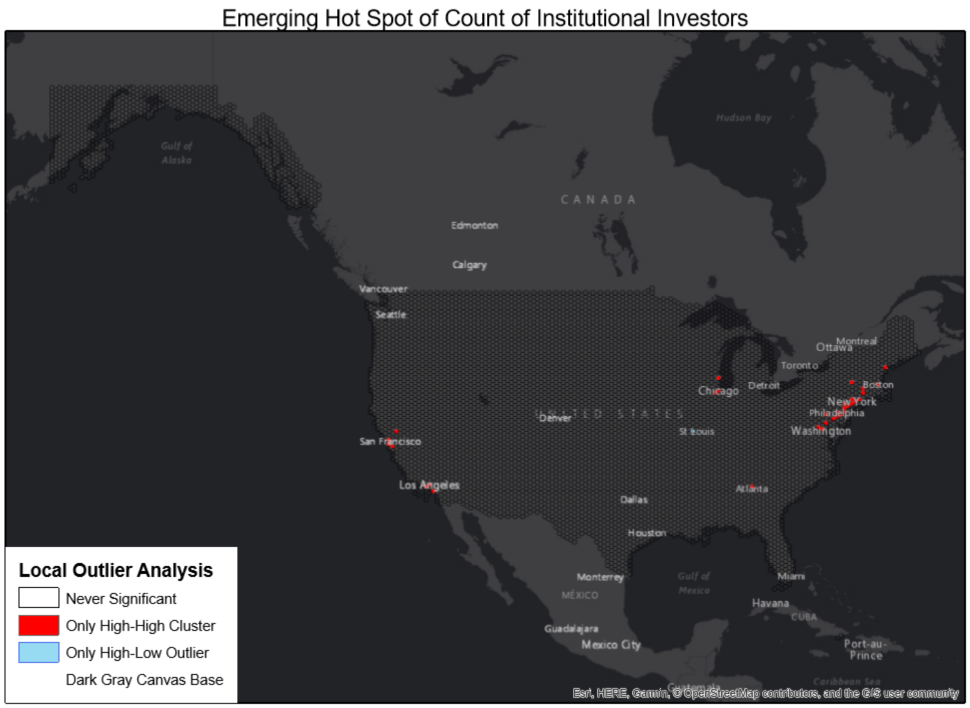


Figure 4.5: Local outlier analysis for funds under management in the United States for the Time Period of June 2013 to December 2018.

analysis (Figure 4.5) is much more restrained than the analysis done on the address book database. Immediately noticeable is the absence of the high-low hexes dotting the capitals of fly-over states, as well as the more restrained presence of low-high clusters in the Bos-NY-Wash. Lastly, as a lone bright spot in a sea of nothingness, Atlanta is the only place outside of the 5 largest US cities for institutional investment that is a high-high hex. This is consistent with the trend seen in Chapter 3.2.5 where Atlanta was becoming the financial centre of the US South-East.

4.4 Boston

As seen in the various tables and analysis in Chapter 3, Boston consistently ranks at the second most important metro area in terms of count of institutional investors and funds under management. The hex bins for the Boston analysis measure 1 km between horizontal parallels and use a local window radius of 8 km. In order to make the comparisons between cities meaningful, this scheme of hexagonal grid and local window size was kept across different metro areas (Chicago, Los Angeles, New York City, and San Francisco).

4.4.1 Count Data

Figure 4.6 identifies a large cluster covering the areas of Central Boston as well as the southern tip of the Massachusetts Route 128 corridor between the suburban cities of Dedham, Needham and Wellsley. This cluster essentially contains 3 different types of hot spots. The first area of central Boston is classified as an intensifying hot spot. This indicates a very high rate of increase in density of institutional investors by hex bin in the area around Boston Commons in downtown Boston. The second type of hot spot covers the outer periphery of central Boston, as well as the southern arc of Highway 128. Lastly, the southern part of the community of Dedham contains a sporadic hot spot indicating that this zone sees intermittent changes in institutional investor count over time. The inclusion of the southern part of the route 128 high tech corridor in the investment cluster isn't surprising considering the long history of partnership between high tech research and development and finance capital (Kenney and Von Burg, 1999).

Figure 4.8 displays of local outlier analysis confirms the importance of both central Boston as well as the southern arch of the route 128 corridor.

4.4.2 Funds under Management

Unlike Figure 4.6's larger cluster, the emerging hot spot analysis in Figure 4.10 using funds under management as a criteria is more exclusionary since it only contains central Boston, and ignores the Massachusetts Route 128 corridor. A partial explanation for this is the high collection of bank and insurance based institutional investors located in Boston's financial district that abuts Boston Common .

Following in a similar theme to Figure 4.10, the local outlier analysis only finds high-high clusters in central Boston. Interestingly, the model accurately picks out Boston Common as a non-cluster. A look at the region shows that many institutional investors surround this 25 hectare urban park, and this creates a discontinuity.

4.5 Chicago

4.5.1 Count Data

As displayed in Figure 4.14, Chicago contains one intensifying hot spot in the Chicago Loop neighbourhood, including satellite hot spots in the Napierville-Aurora suburb to the West, as well as Evanston and Highland Park to the North.

Using local outlier analysis, only the Loop district contains high-high hexagons.

Figure 4.6: Hot Spot Analysis of Number of Firms in Boston for the time period March 1999 to December 2018

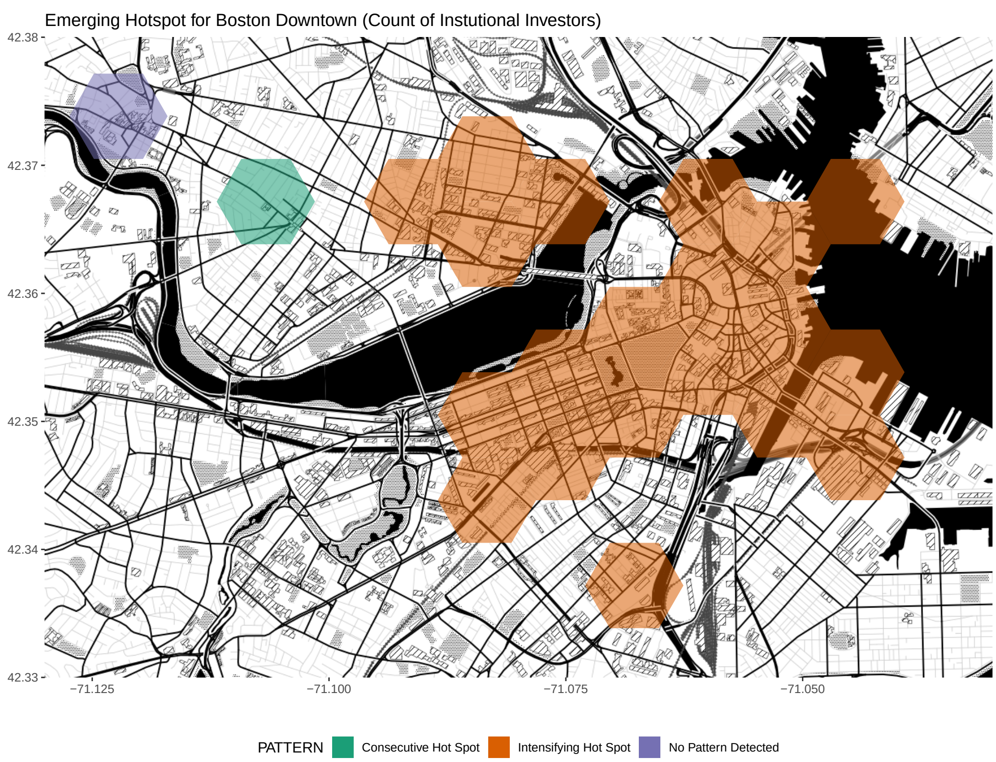


Figure 4.7: Hot Spot Analysis of Number of Firms in Downtown Boston for the time period March 1999 to December 2018

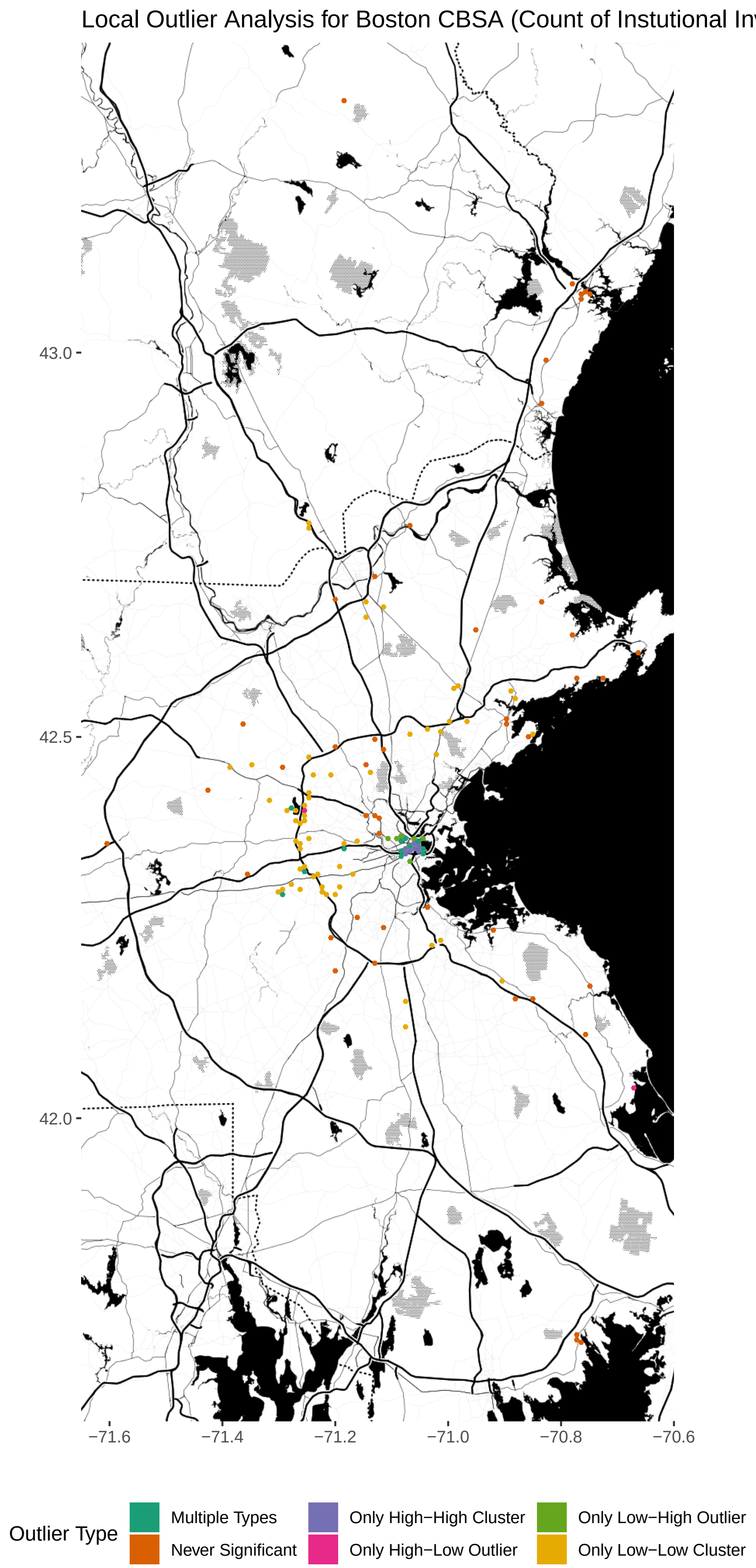


Figure 4.8: Boston Local Outlier Analysis - Count of Institutional Investors

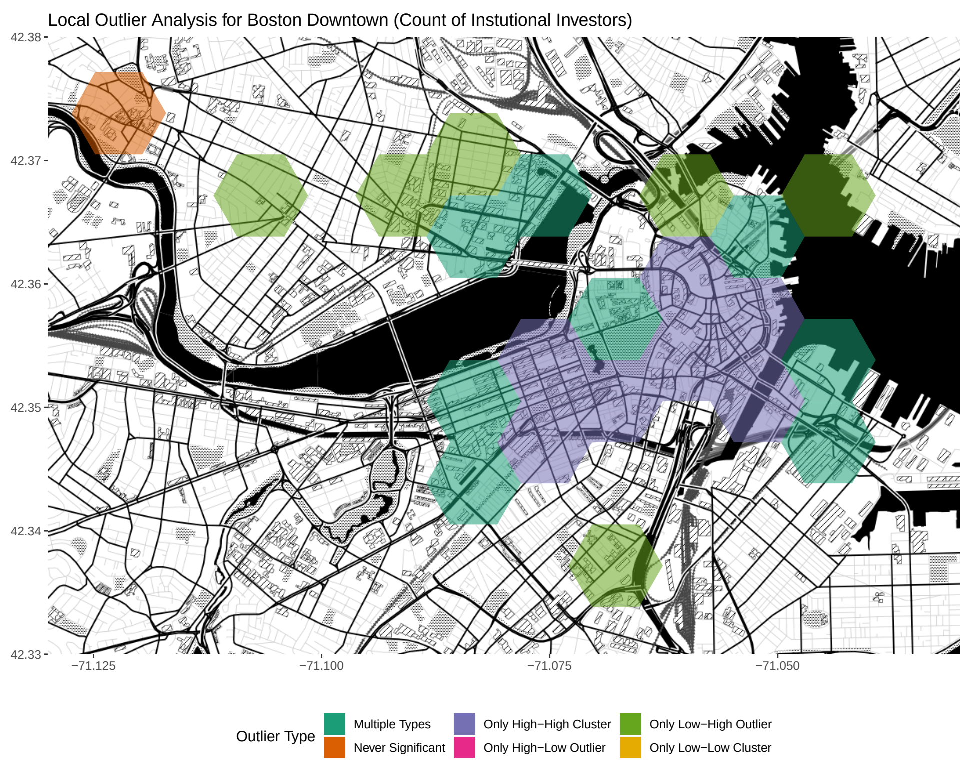


Figure 4.9: Downtown Boston Local Outlier Analysis - Count of Institutional Investors

Figure 4.10: Emerging Hot Spot Analysis of Funds under Management for Boston for period June 2013 to December 2018

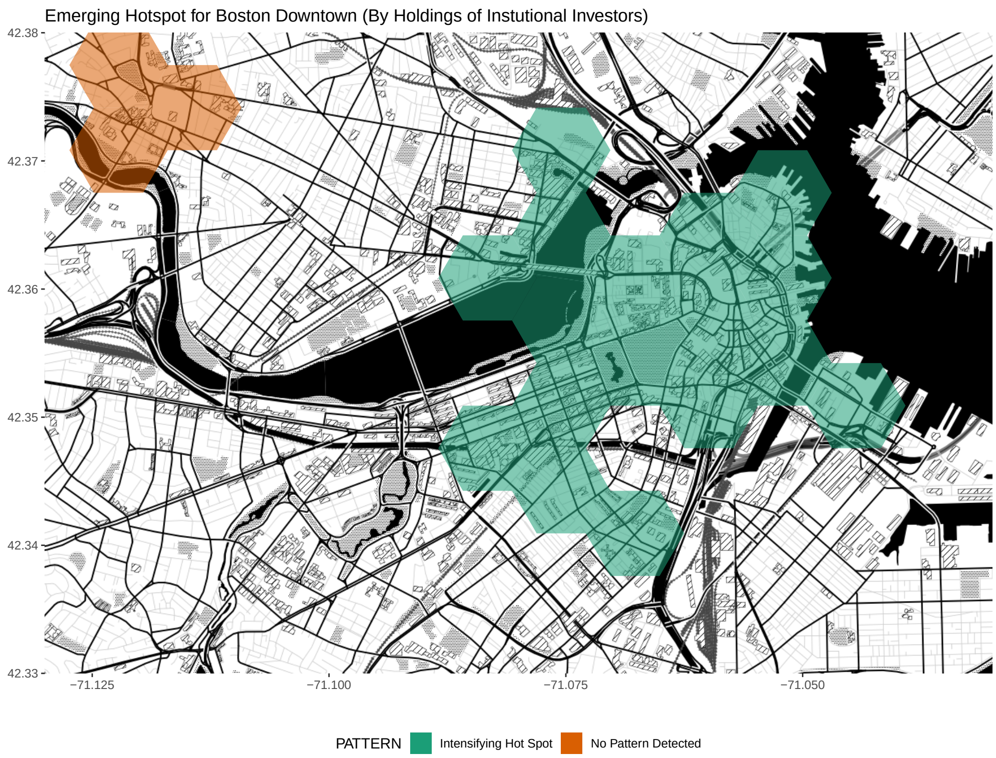


Figure 4.11: Emerging Hot Spot Analysis of Funds under Management for Downtown Boston for period June 2013 to December 2018

Figure 4.12: Boston Local Outlier Analysis - Funds under Management

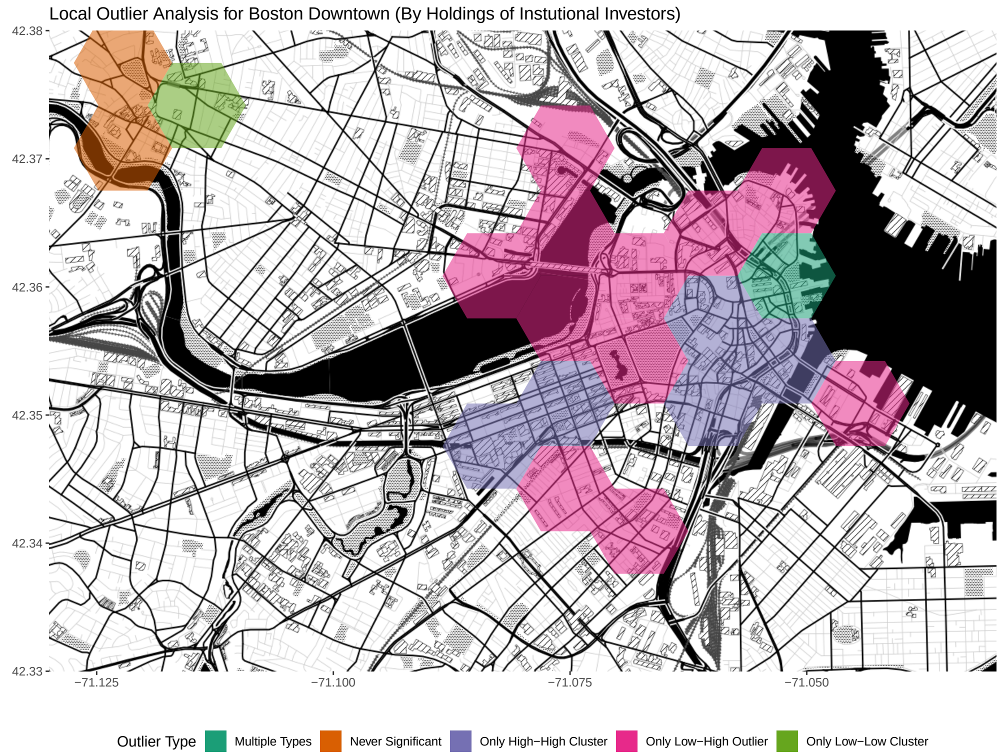


Figure 4.13: Boston Downtown Local Outlier Analysis - Funds under Management

This is consistent with institutional investors perfering CBDs. Futhermore, there is a conspicuous absence of investors on the South Side of Chicago, however this is not a region of Chicago known for having much financial capital.

4.5.2 Funds under Management

Figure 4.18 suggests a similar picture to the other emerging hot spot analysis maps where the key variable is funds under management, for there are less regions defined as a hot spot. In this case, the hot spots in Evanston and Highland Park disappear, and the Napierville-Aurora cluster is much smaller in size.

Figure 4.20 paints a similar story than Figure 4.18, for the main cluster of high-high hexagons is located in the Chicago Loop district. A secondary cluster of a single

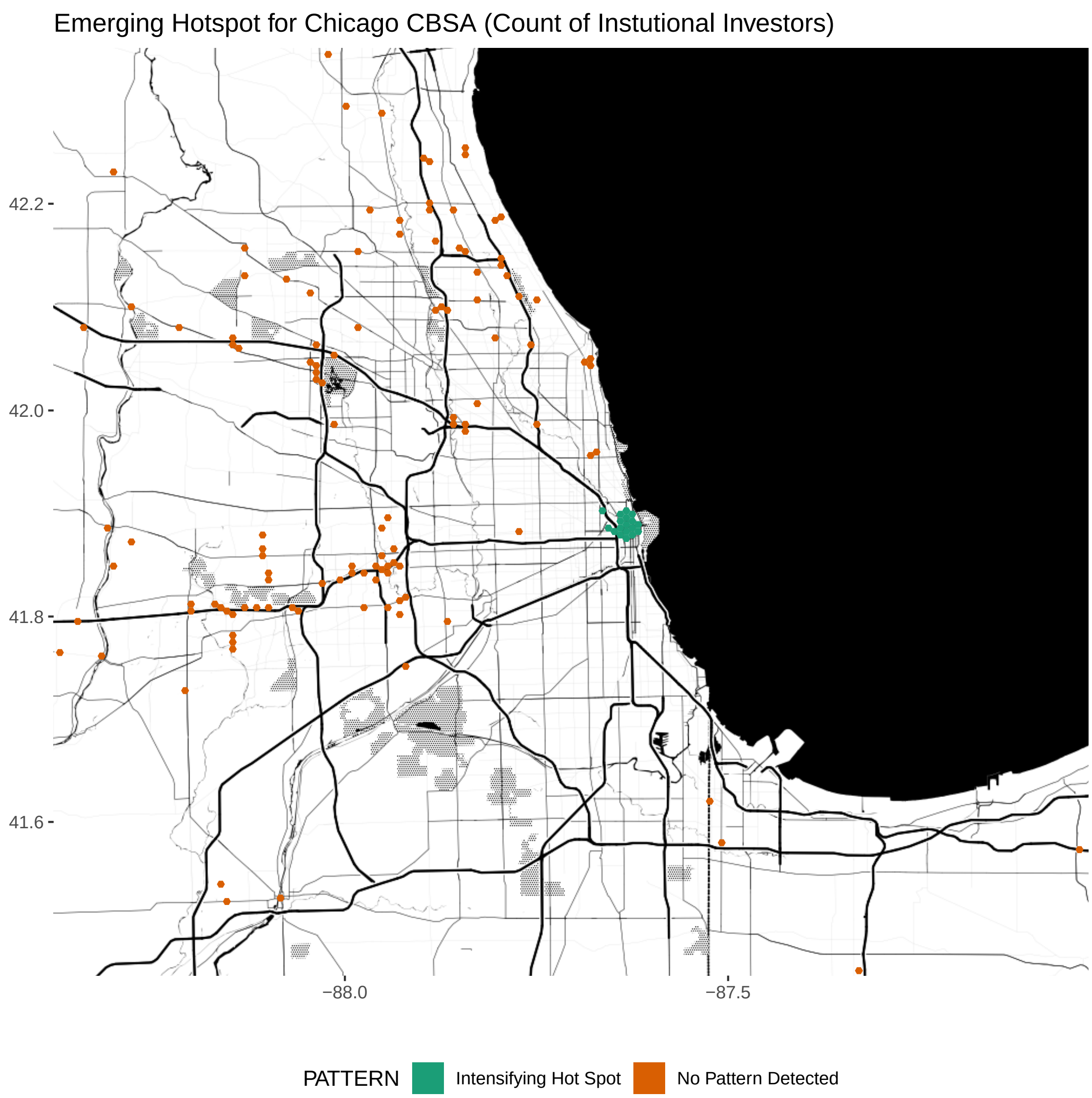


Figure 4.14: Hot Spot Analysis of Number of Firms in Chicago for the time period March 1999 to December 2018

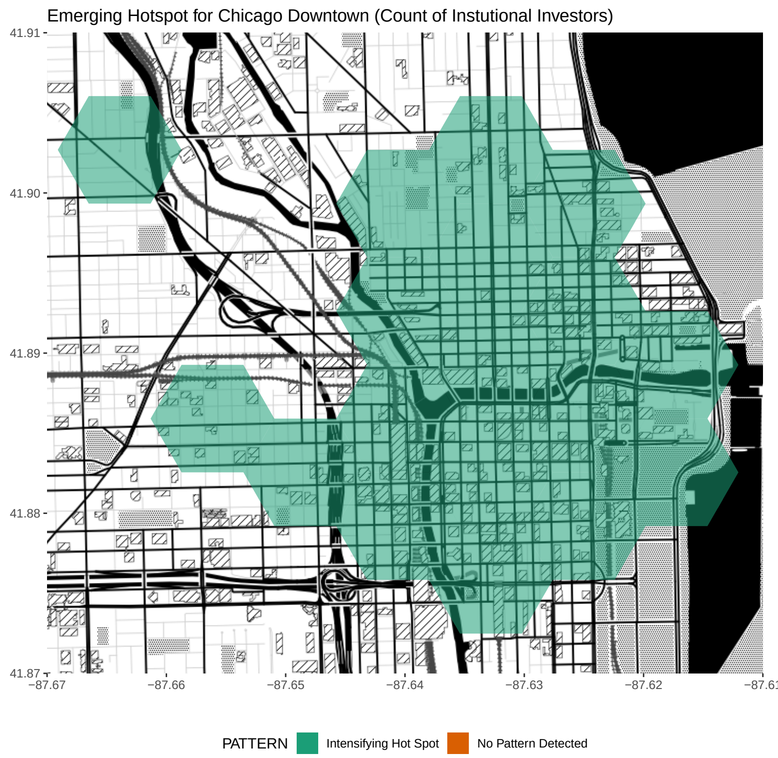


Figure 4.15: Hot Spot Analysis of Number of Firms in Chicago for the time period March 1999 to December 2018

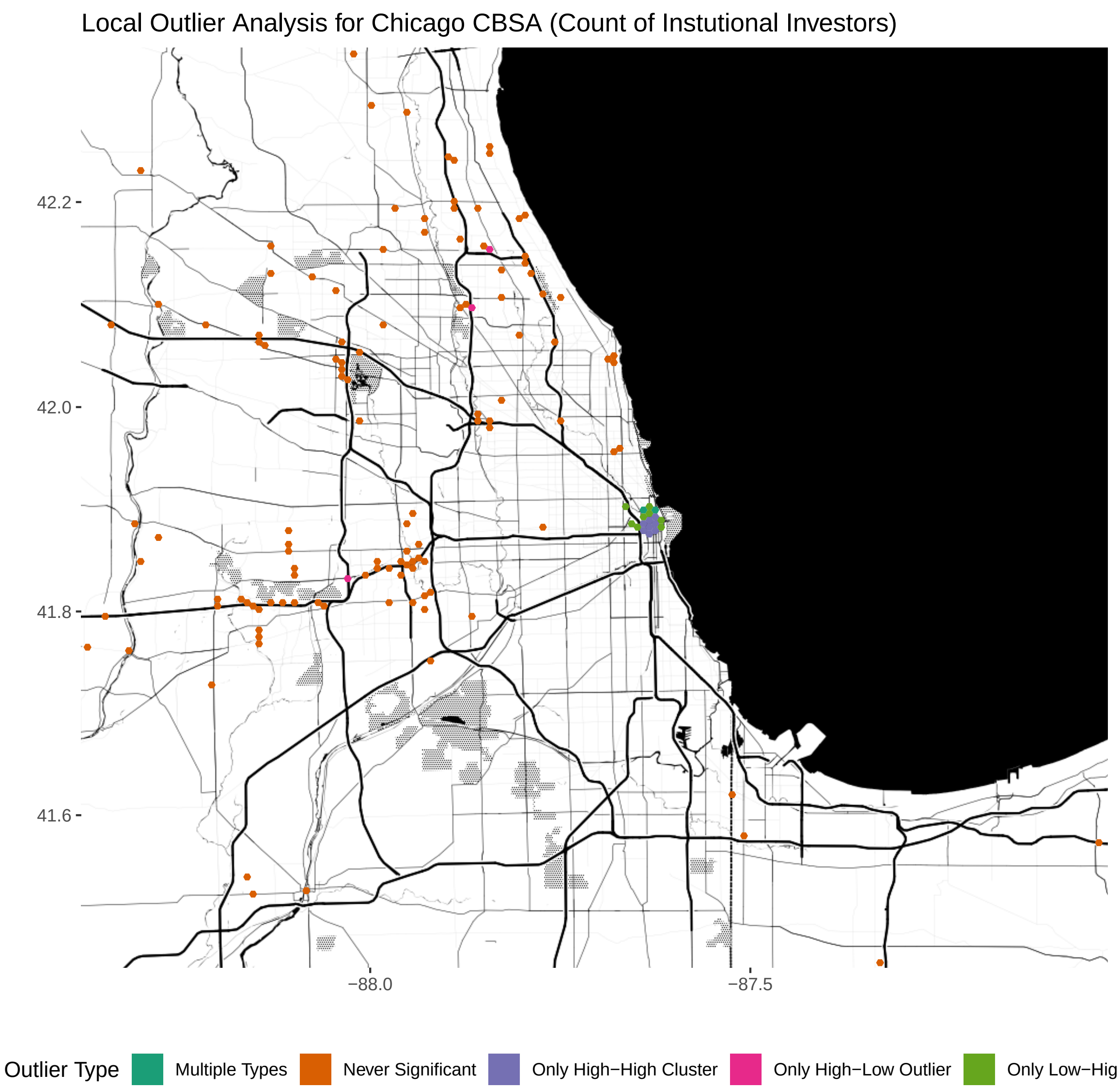


Figure 4.16: Chicago Local Outlier Analysis - Count of Institutional Investors

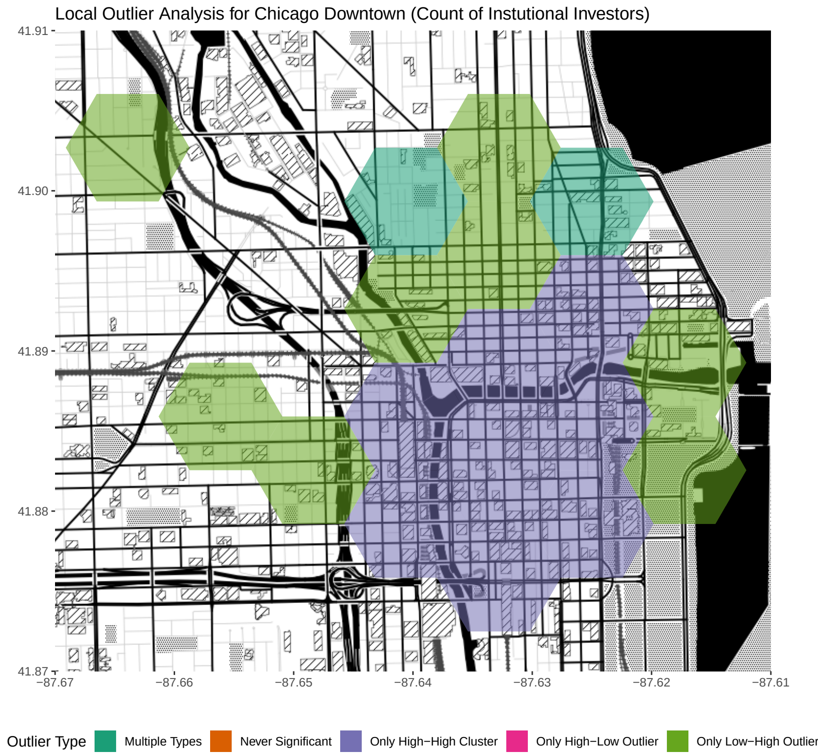


Figure 4.17: Chicago Local Outlier Analysis - Count of Institutional Investors

high-high hexagon exists in the Napierville-Aurora region. Furthermore, the cluster in the Loop neighbourhood of Chicago is much more defined in this analysis compared to the count map. This sharper cluster is not surprising considering the presence of the Chicago financial district, anchored by the Chicago Mercantile Exchange, at the centre of the Loop.

4.6 Los Angeles

4.6.1 Count Data

Figure 4.22 indicates that there is an absence of a central financial district and that investors are more diffused. As such, unlike Boston and Chicago, the emerging hot

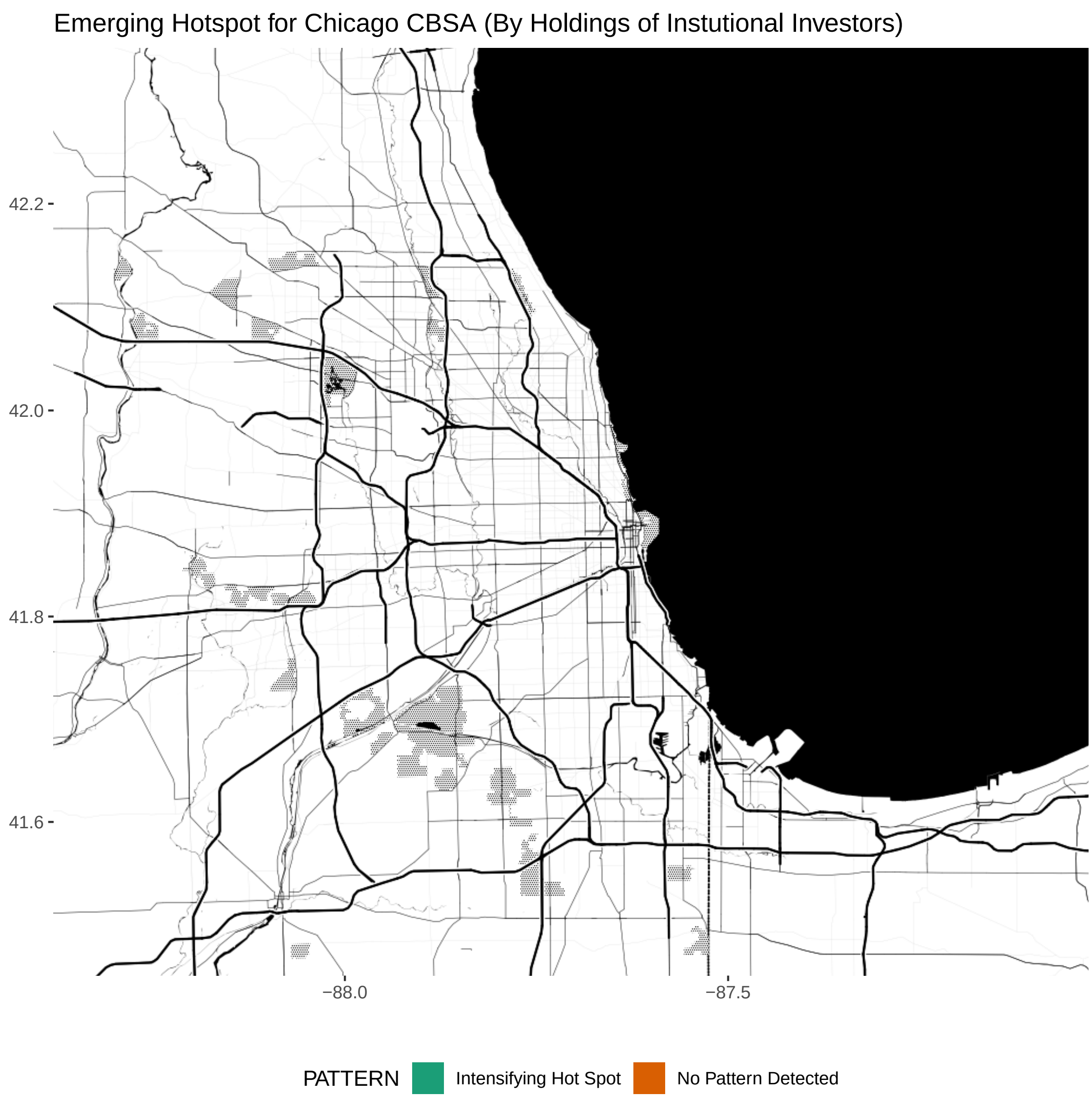


Figure 4.18: Emerging Hot Spot Analysis of Funds under Management for Chicago for period June 2013 to December 2018

Figure 4.19: Emerging Hot Spot Analysis of Funds under Management for Chicago for period June 2013 to December 2018

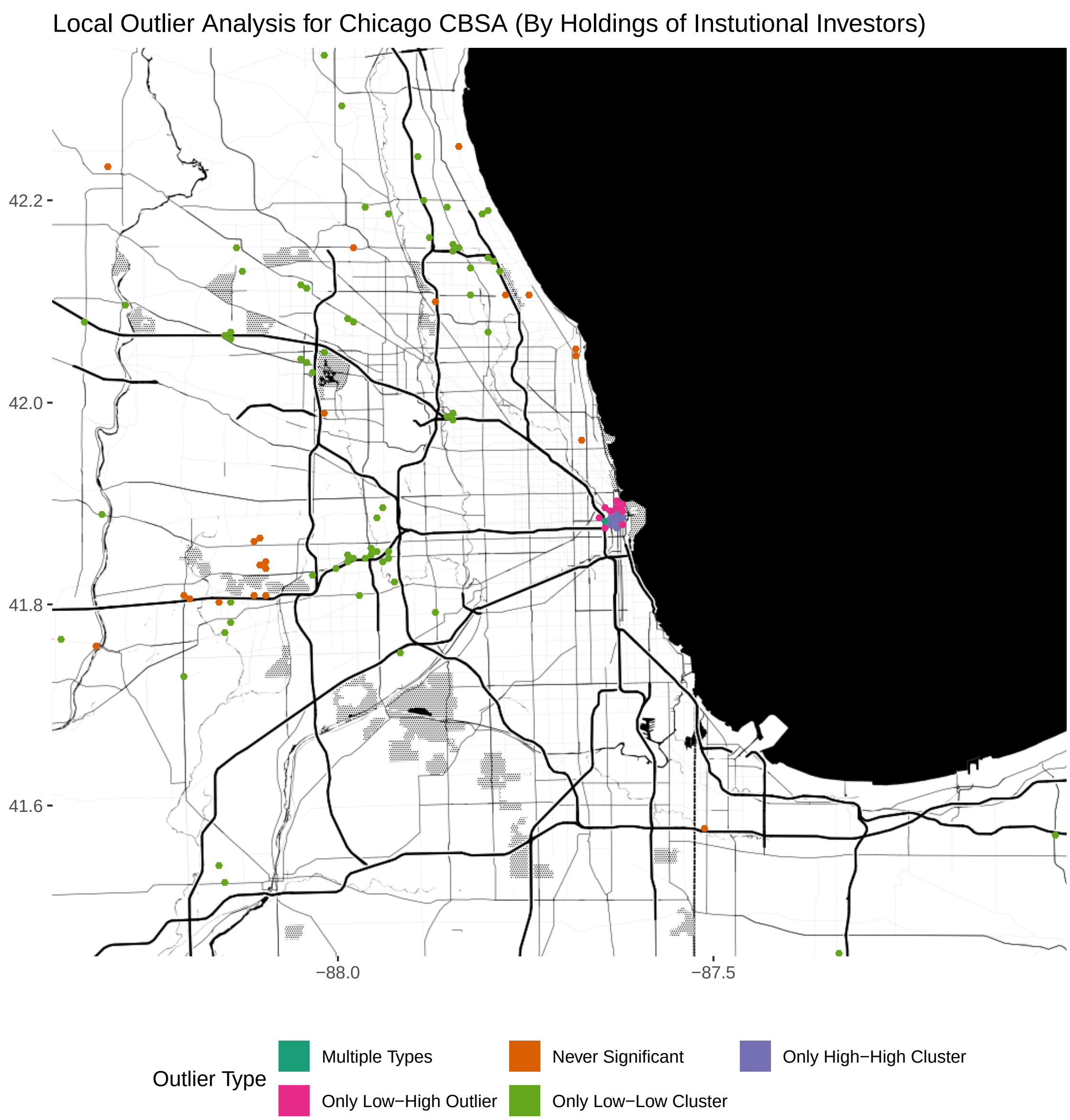


Figure 4.20: Chicago Local Outlier Analysis - Funds under Management

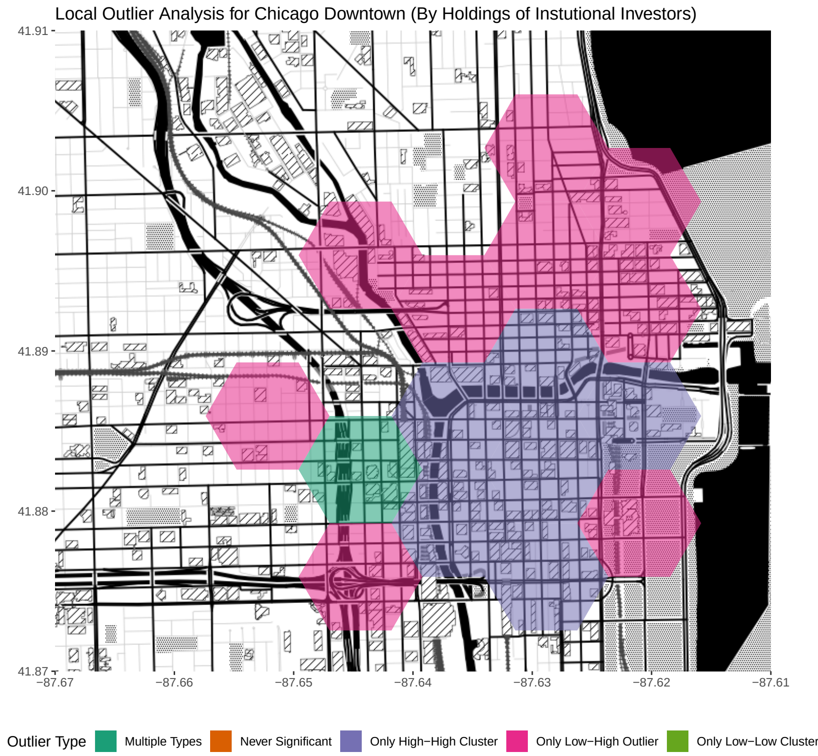


Figure 4.21: Chicago Local Outlier Analysis - Funds under Management

spot analysis map for Los Angeles offers more categories. This broad spread of hot spots is not really surprising considering Los Angeles's history and reputation for urban sprawl and suburban office parks (Dear and Flusty, 1998; Harris and Lewis, 1998). The lack of a historic CBD comprised of skyscrapers on the scale of New York's Wall Street and Midtown or Chicago's Loop district and decentralized city administration certainly help in creating multiple small intensifying hot spots around the city such as Downtown, Santa Monica, Beverly Hills, Costa Mesa and Irvine.

These hot spot locations also show up in Figure 4.24 as local outliers. However, there is a large amount of hexagons displaying the mixed outlier type in Santa Monica. This can be partially explained by the diffuse nature of locations in Santa Monica compared to other clusters such that across time they might appear as high-highs or high-lows due to neighbourhood effects.

4.6.2 Funds Under Management

Continuing the theme seen in all previous maps with regards to analysing funds under management, the map that is weighted by money rather then the mere presence of an investor reduces the importance of suburban investors. This suggests that while suburban investors are becoming more common, their portfolio of holdings are smaller than CBD-based investors.

The emerging hot spot analysis for Figure 4.26 as well as the local outlier analysis in Figure 4.28 drops the Costa Mesa and Irvine hot spots. Furthermore, the Downtown Los Angeles hot spot remains the only one that is still an intensifying hot spot. This can be explained by the recent construction boom in high grade office towers being built in the Downtown after an influx of foreign capital and a planning mandate towards densification (Marino, 2019).

4.7 New York City

4.7.1 Count Data

Figure 4.30 displays the singular emerging hot spot cluster for the New York region. Unsurprisingly, this hot spot covers the heart of the US financial universe: the Financial District and Midtown on Manhattan Island, and extending somewhat into the Bronx, Brooklyn and Hudson County, New Jersey. Furthermore, the intensifying

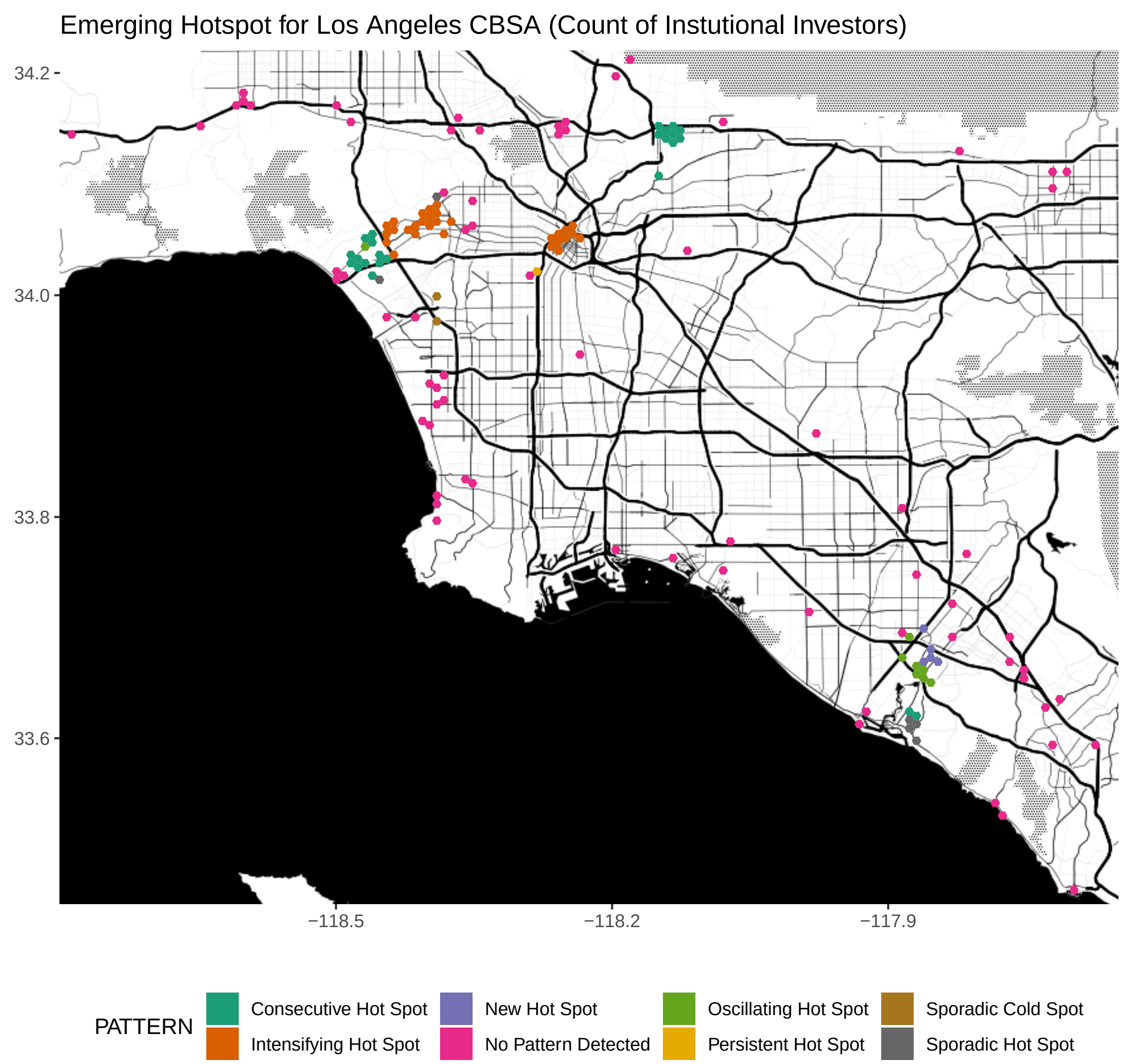


Figure 4.22: Hot Spot Analysis of Number of Firms in Los Angeles for the time period March 1999 to December 2018

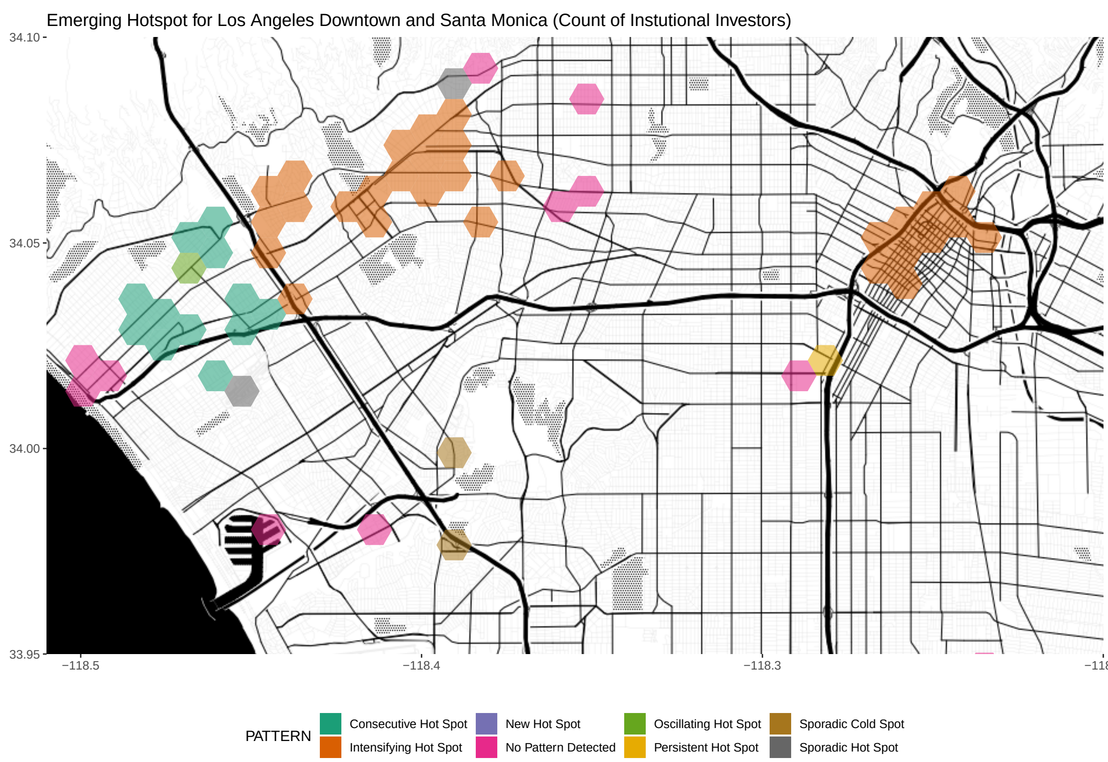


Figure 4.23: Hot Spot Analysis of Number of Firms in Downtown Los Angeles and Santa Monica for the time period March 1999 to December 2018

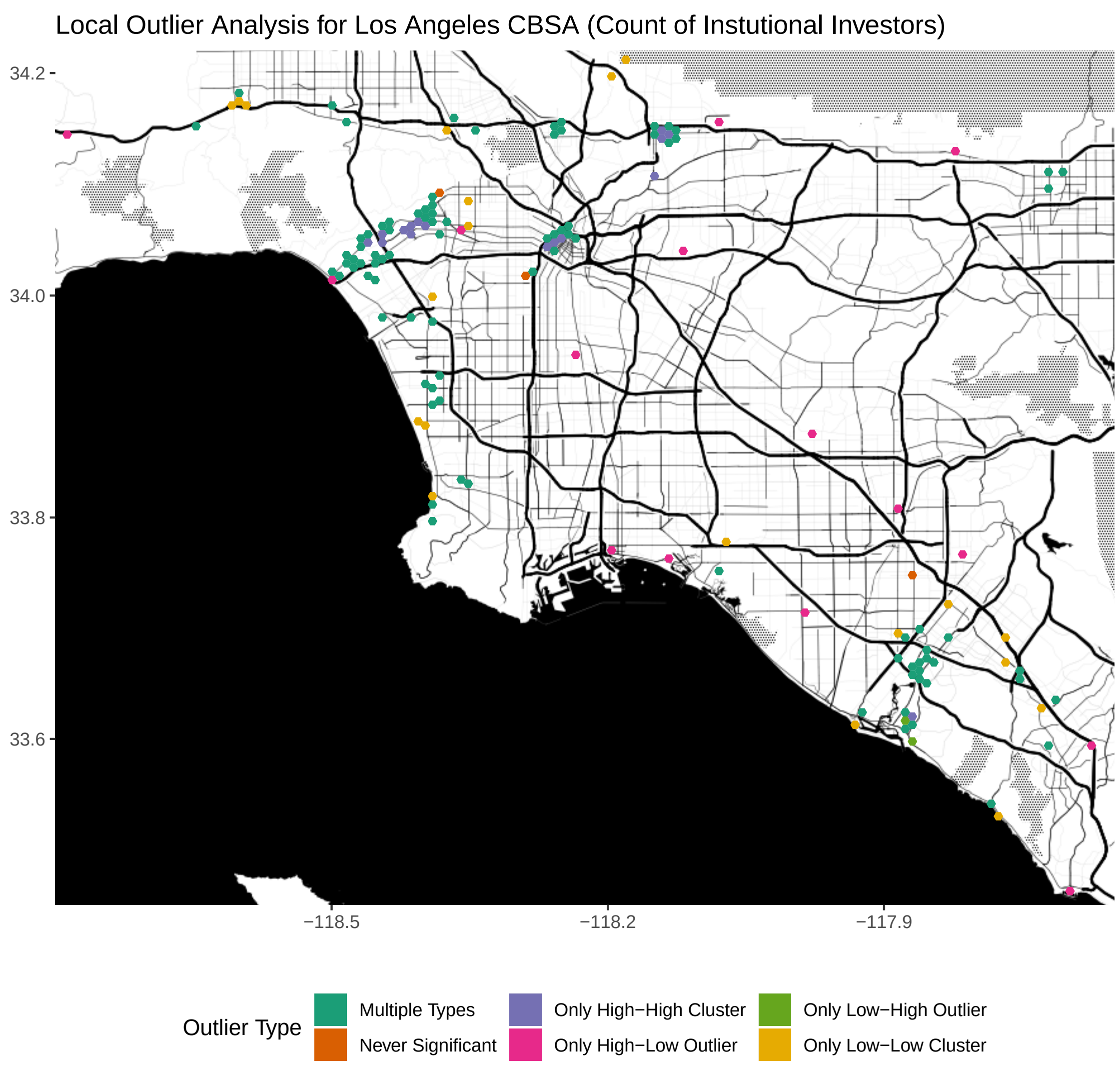


Figure 4.24: Los Angeles Local Outlier Analysis - Count of Institutional Investors

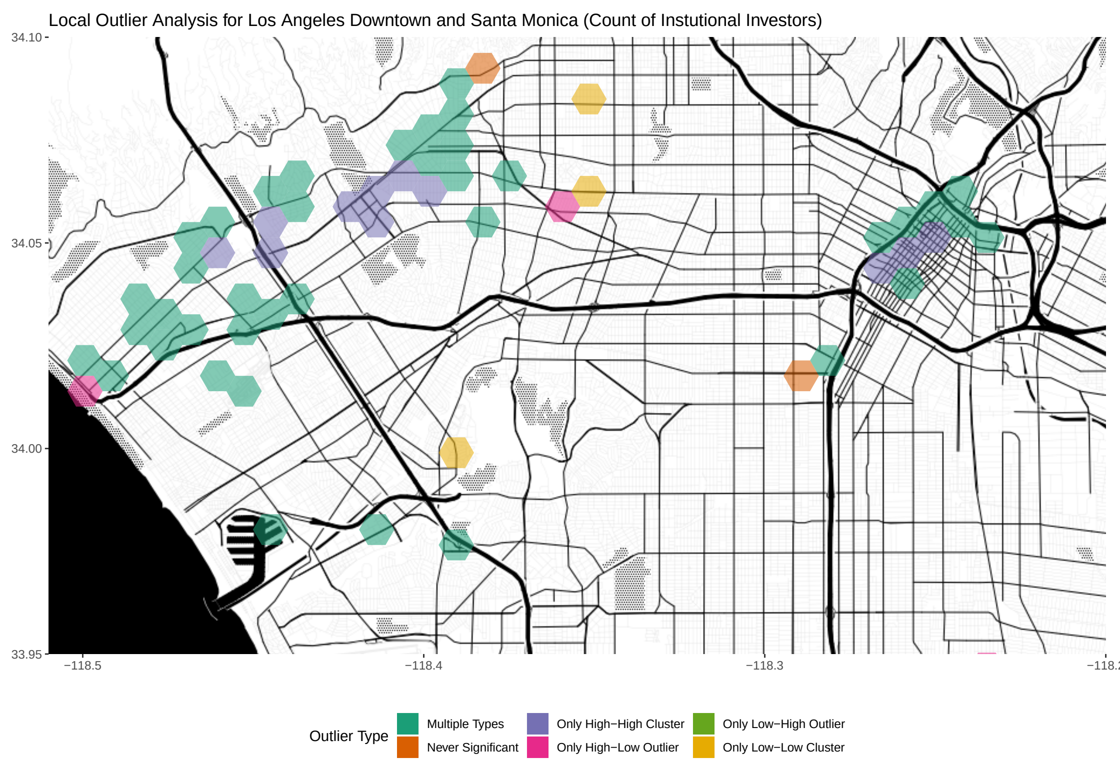


Figure 4.25: Downtown Los Angeles and Santa Monica Local Outlier Analysis - Count of Institutional Investors

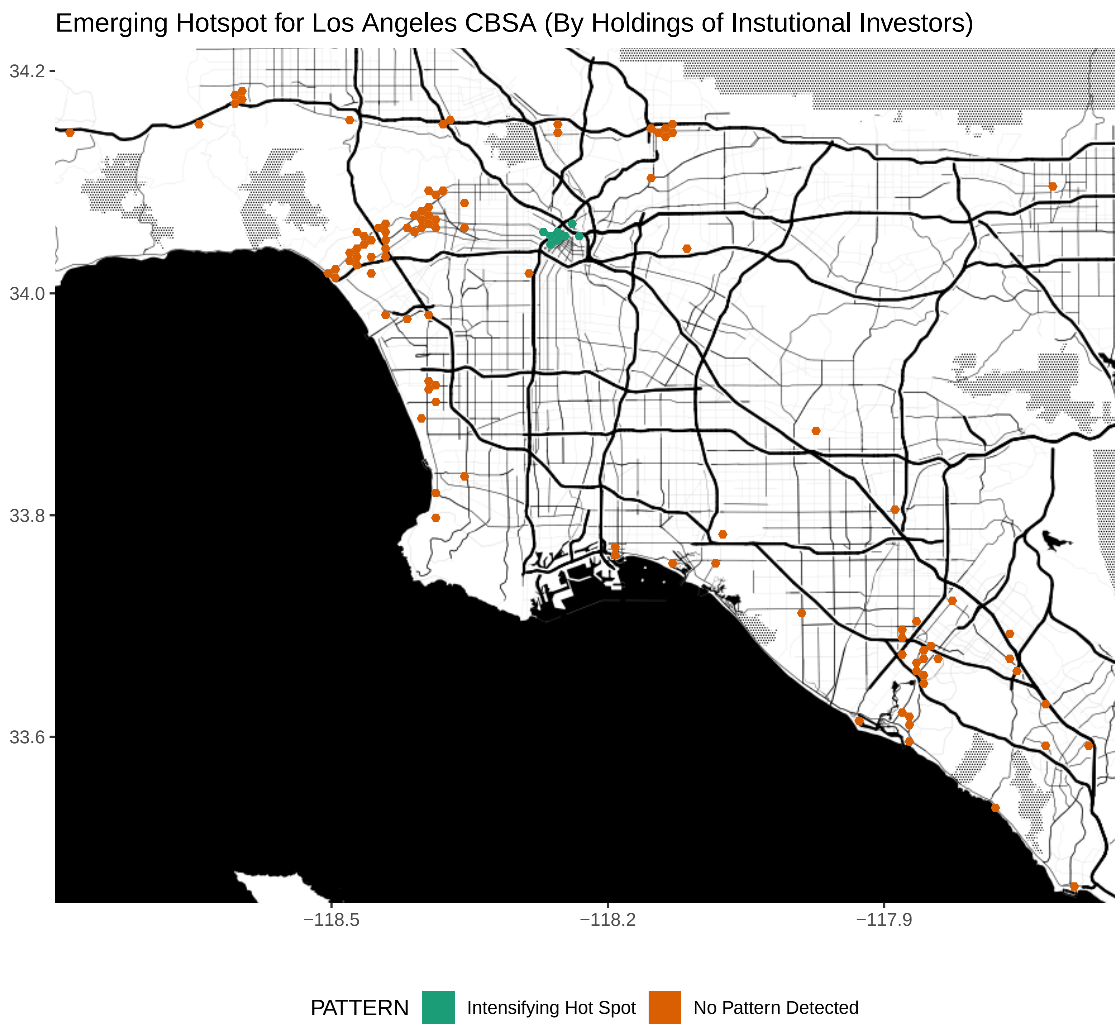


Figure 4.26: Emerging Hot Spot Analysis of Funds under Monument for Los Angeles for period June 2013 to December 2018

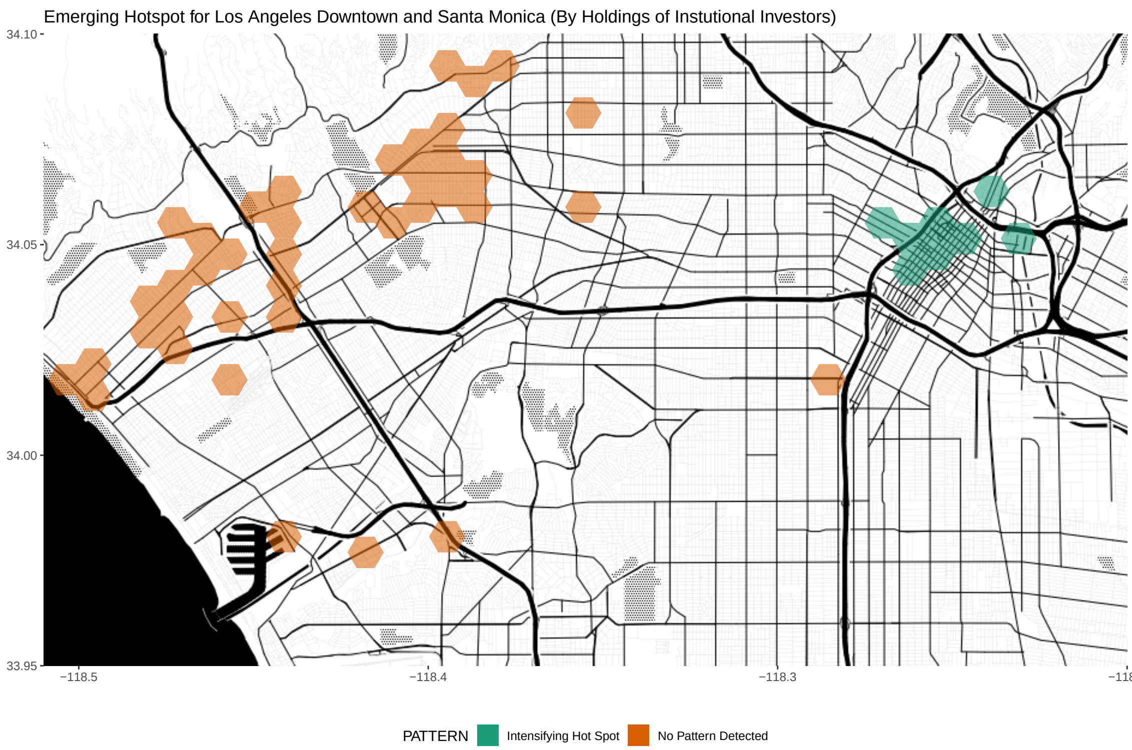


Figure 4.27: Emerging Hot Spot Analysis of Funds under Monument for Downtown Los Angeles and Santa Monica for period June 2013 to December 2018

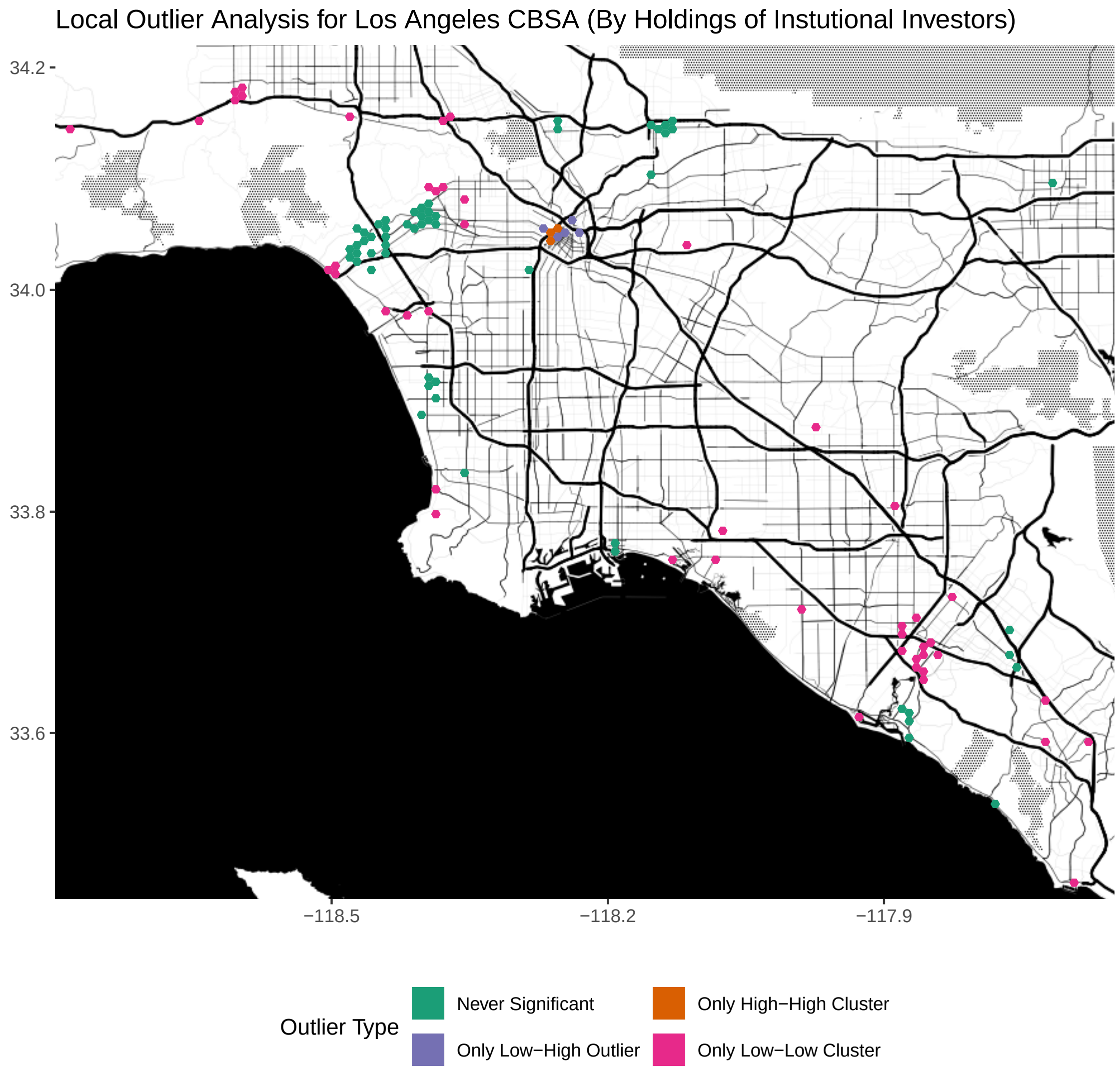


Figure 4.28: Los Angeles Local Outlier Analysis - Funds under Management

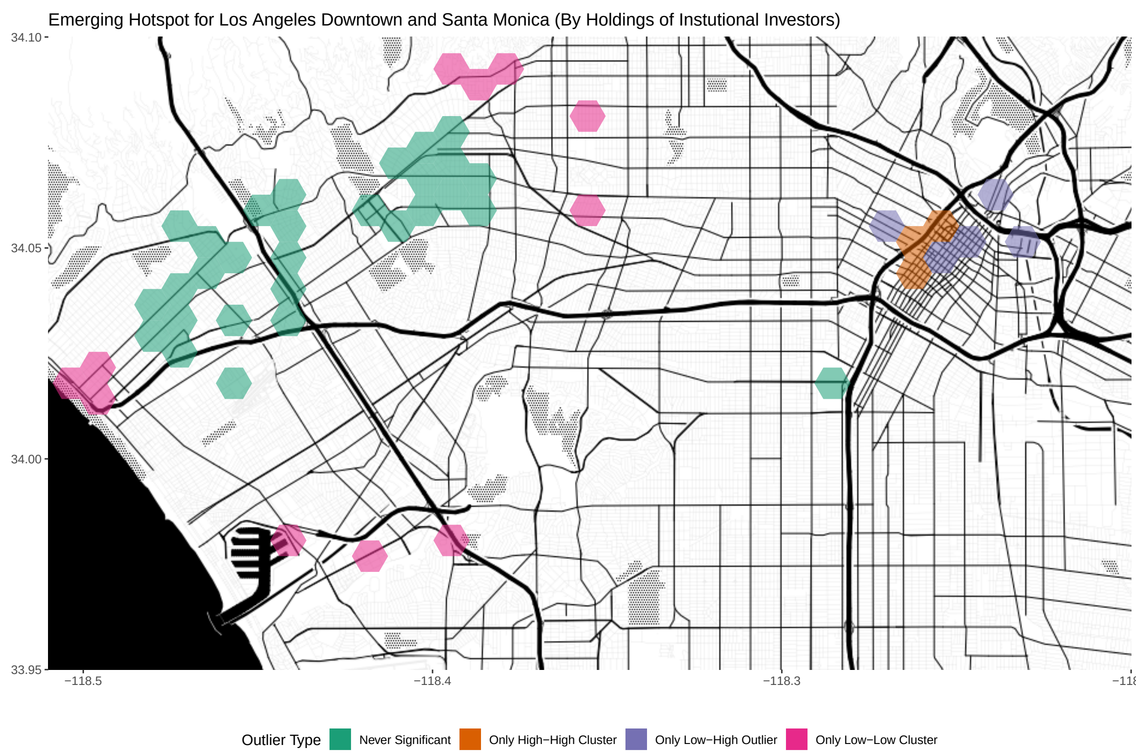


Figure 4.29: Downtown Los Angeles and Santa Monica Local Outlier Analysis - Funds under Management

hotspot over Manhattan and the constant hot spot to the south of it is evidence in the shift northwards towards Midtown Manhattan due to the desire to be near the intercontinental exchange - that is to say where transatlantic fiber optic cables come to shore in North America.

Providing more detailed spatial resolution on high-high hot spots, Figure 4.32 finds that most of the high-high hexes are located in Manhattan, and a few isolated hexes are located in Brooklyn, Bronx and Hudson Counties. Notable by its absence, the highly residential Stuyvesant Town neighbourhood on the east side of Manhattan is largely devoid of institutional investors.

4.7.2 Funds Under Management

Once again, the use of funds under management as the unit of measure for emerging hot spot analysis shows a more restrictive hot spot. In fact, Figure 4.34 is simply a more restrictive version of Figure 4.30. The same can be said of Figure 4.36 treatment of local outlier analysis when compared to Figure 4.32. That being said, this more restrictive criteria removes most of the high-high clusters in Hudson County and Brooklyn County, suggesting once again that these investors located outside of the CBD have a smaller bankroll than the investors located in the CBD.

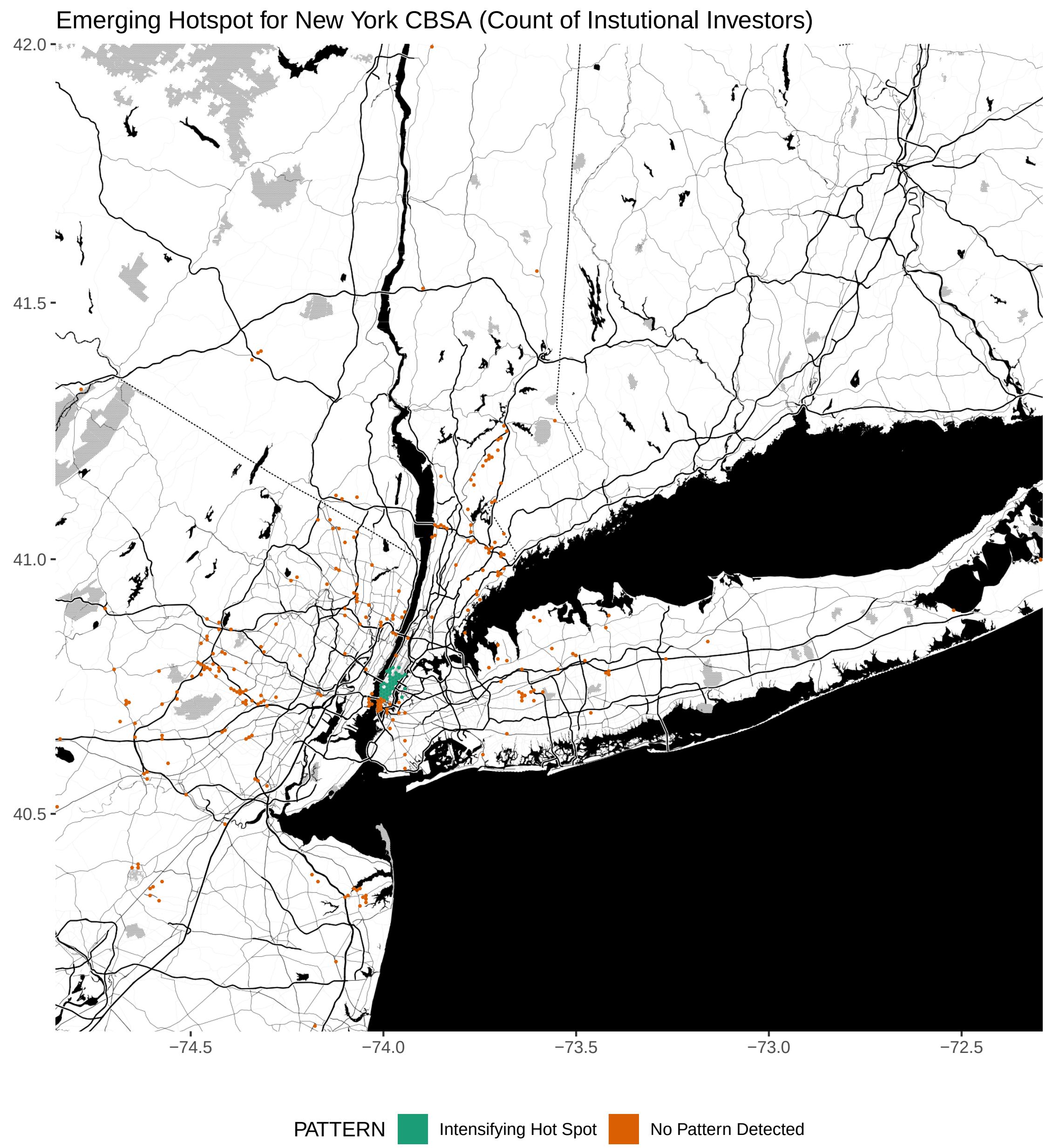


Figure 4.30: Hot Spot Analysis of Number of Firms in New York for the time period March 1999 to December 2018

Figure 4.31: Hot Spot Analysis of Number of Firms in New York for the time period March 1999 to December 2018

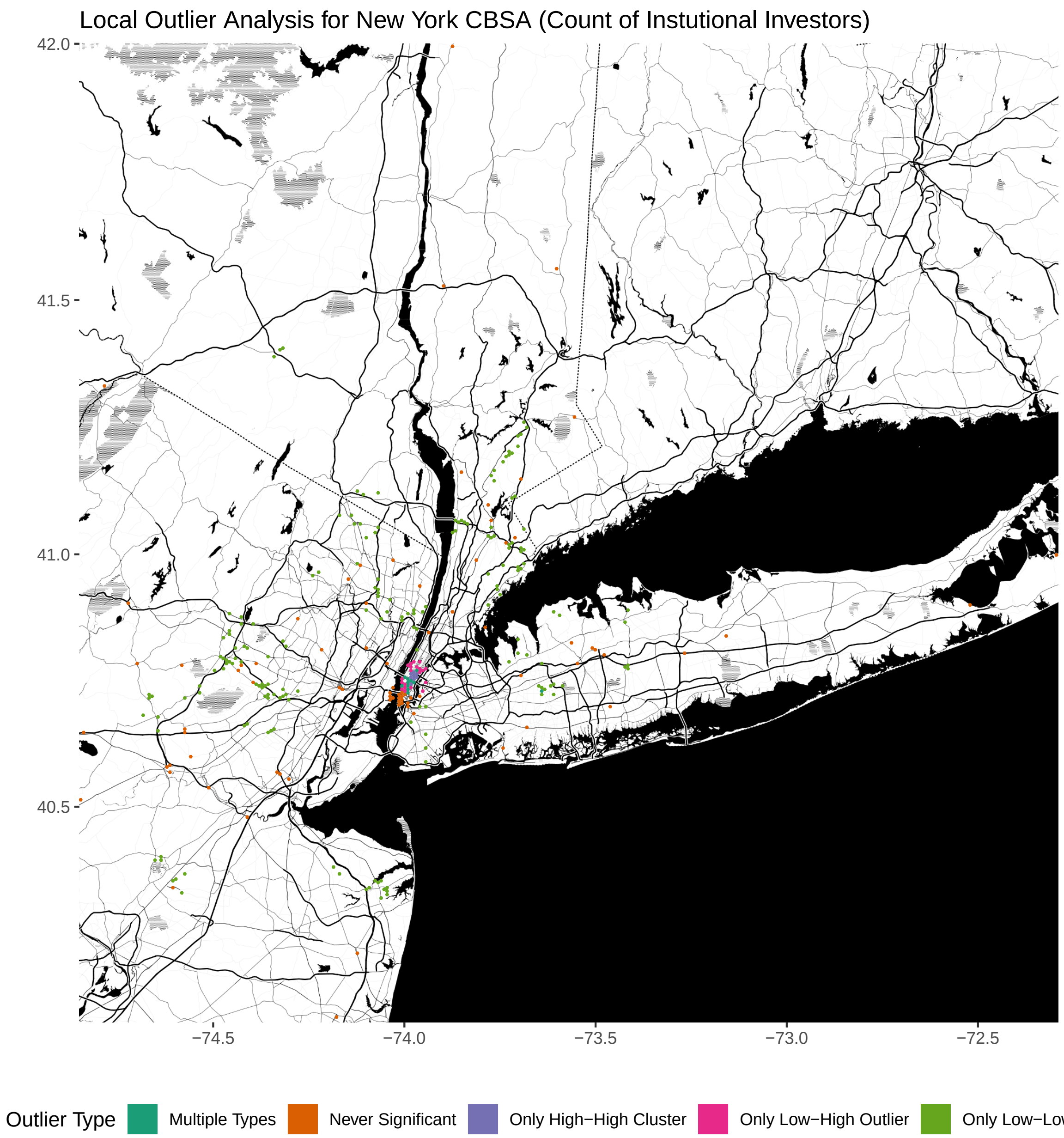


Figure 4.32: New York Local Outlier Analysis - Count of Institutional Investors

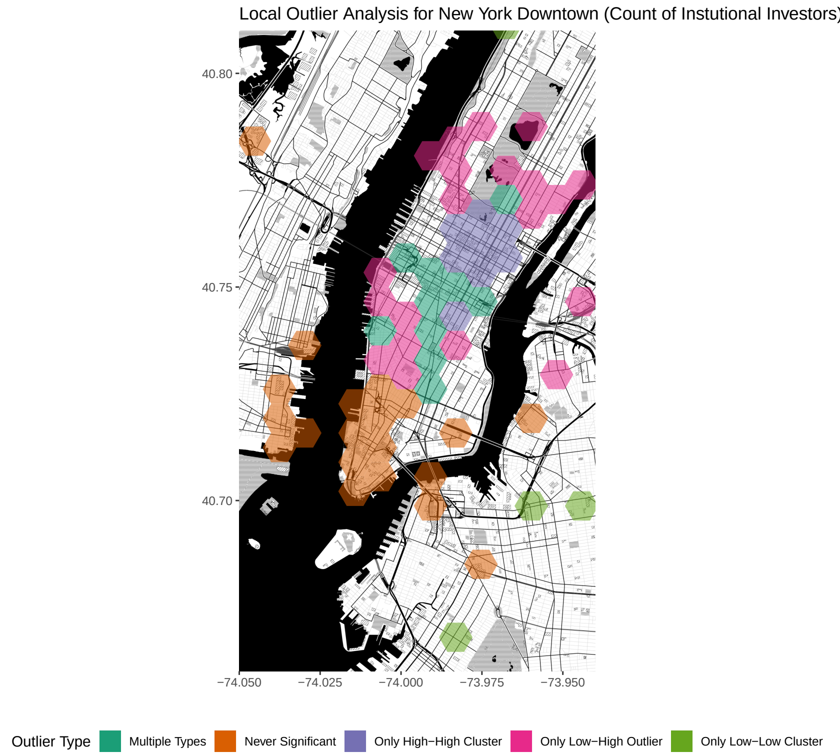


Figure 4.33: New York Local Outlier Analysis - Count of Institutional Investors

4.8 San Francisco

4.8.1 Count Data

Figure 4.38 displays five hot spots: an emerging hot spot in San Francisco's central business district, San Mateo, a small emerging centre north of the Golden Gate Bridge along with consecutive hot spots in Palo Alto and Walnut Creek.

Figure 4.40 displays the results of the local outlier analysis and finds the same five clusters.

Figure 4.34: Emerging Hot Spot Analysis of Funds under Management for New York for period June 2013 to December 2018

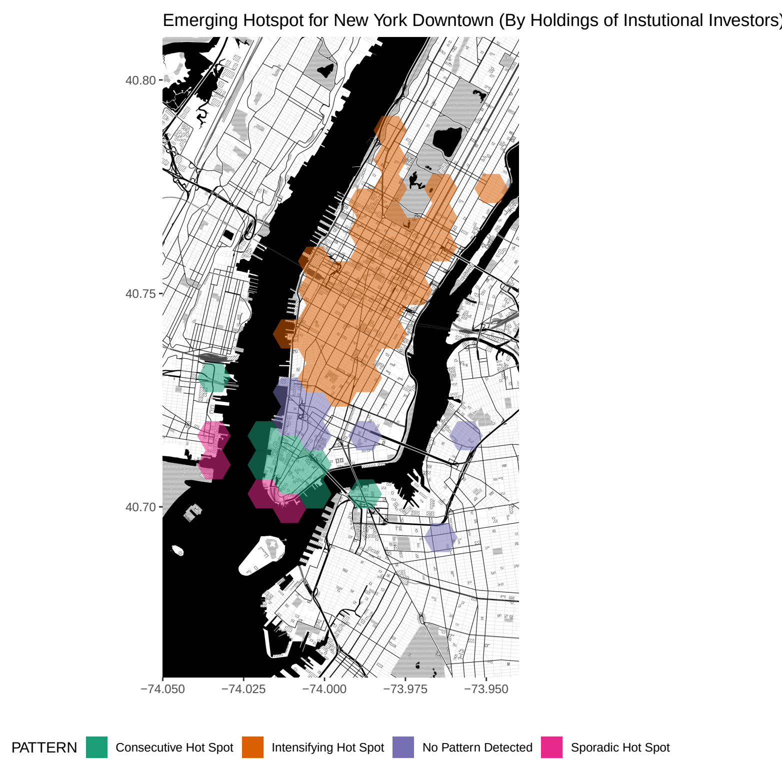


Figure 4.35: Emerging Hot Spot Analysis of Funds under Management for New York for period June 2013 to December 2018

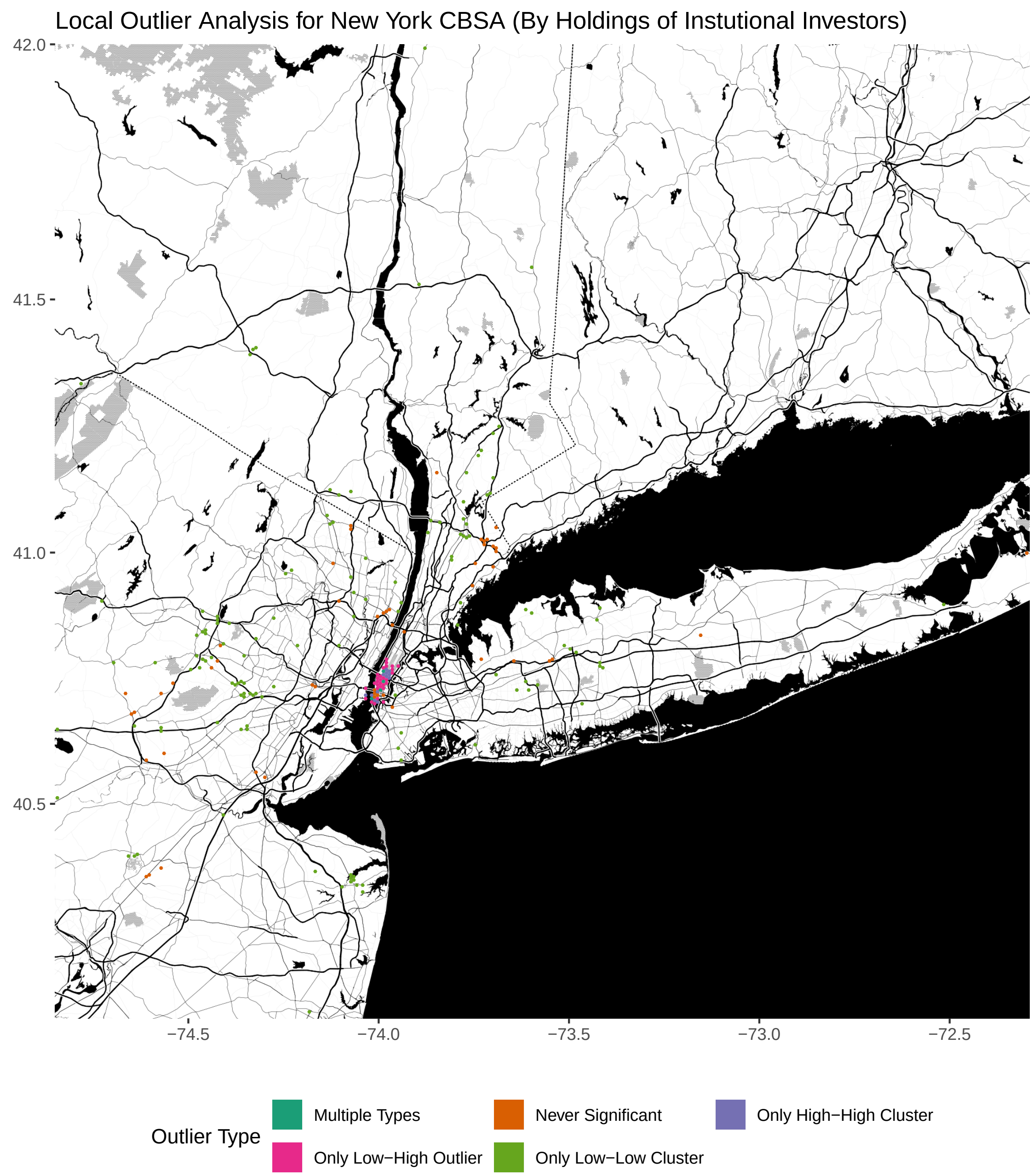


Figure 4.36: New York Local Outlier Analysis - Funds under Management

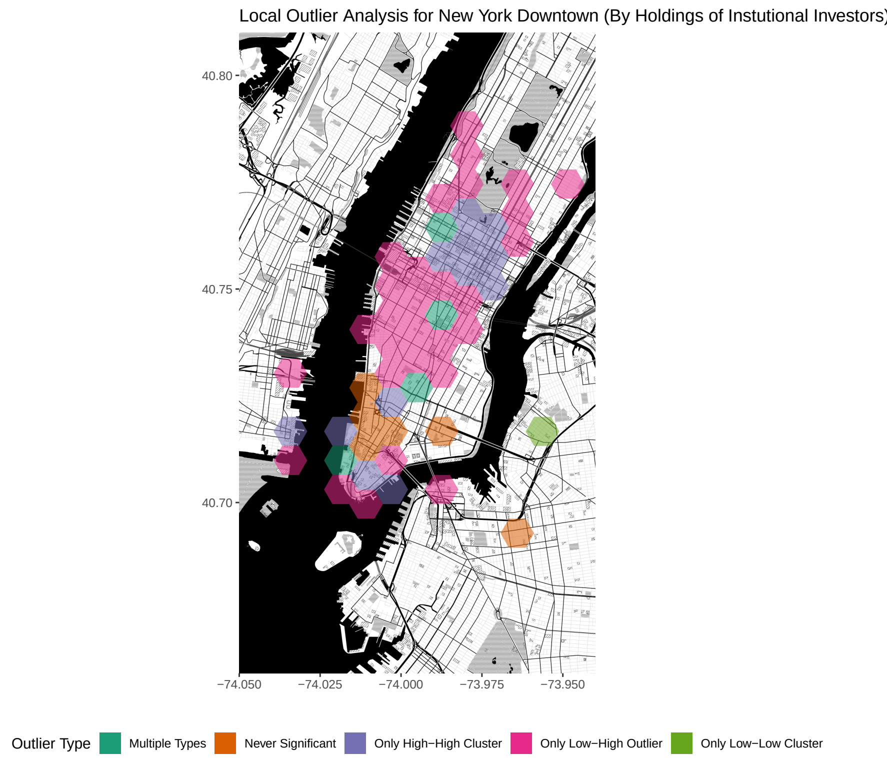


Figure 4.37: New York Local Outlier Analysis - Funds under Management

4.8.2 Funds Under Management

In a continuing theme of having the funds under management Figures 4.42 and 4.44 show fewer hot spots than count data. These hot spots are located in San Francisco's CBD and in San Mateo.

4.9 Conclusion

When looking at the Continental United States, it appears that institutional investors are not evenly distributed across its vast surface. As a matter of fact, other than a few outlying homesteads of institutional investors located in State capitals, most of the investors are located in the major metro areas of Boston, Chicago, Los Angeles, New

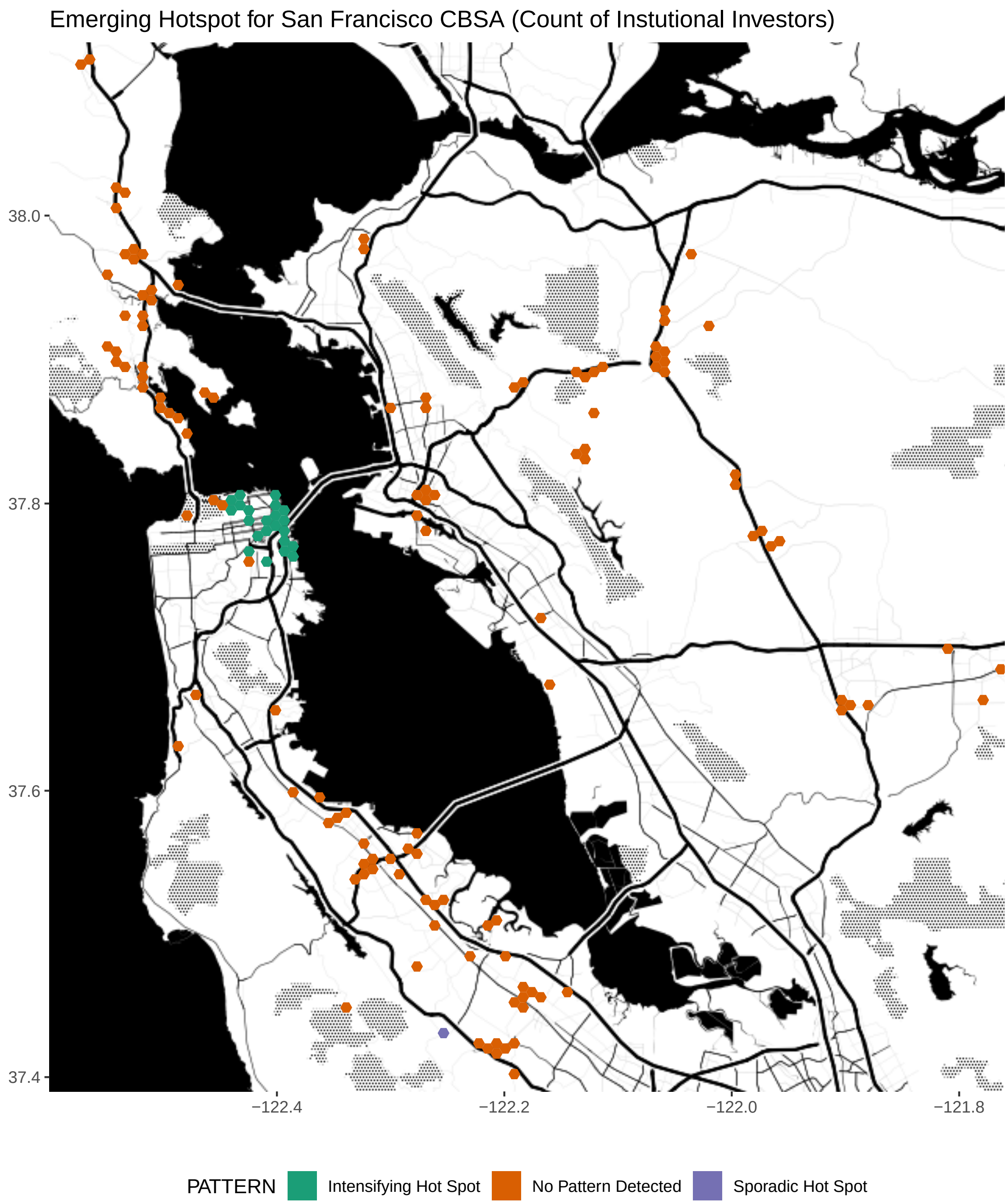


Figure 4.38: Hot Spot Analysis of Number of Firms in San Francisco for the time period March 1999 to December 2018

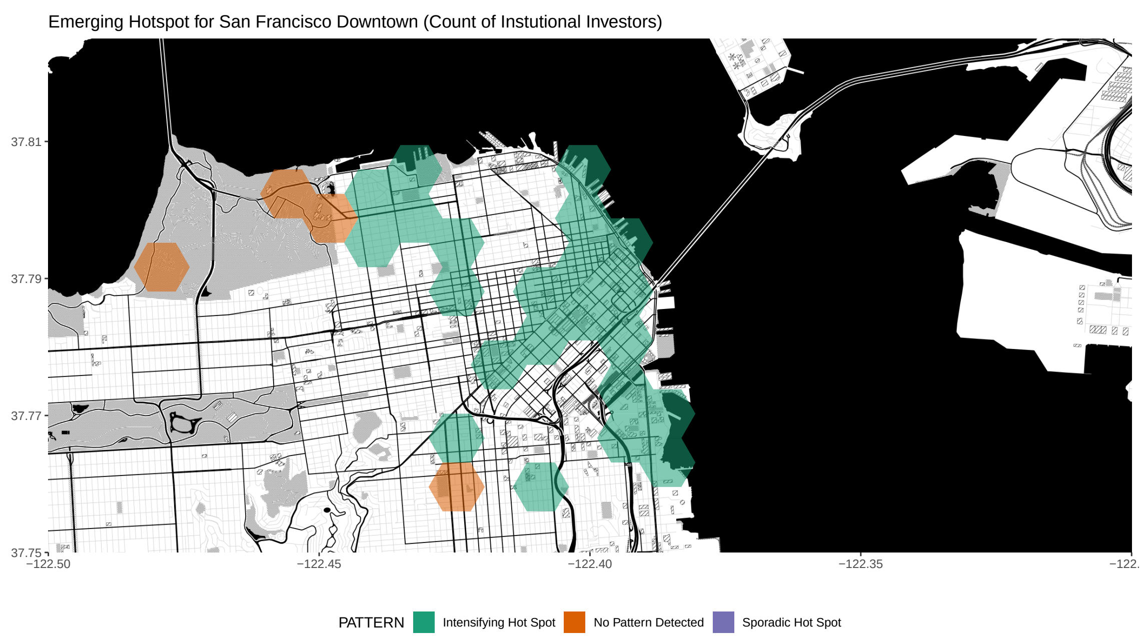


Figure 4.39: Hot Spot Analysis of Number of Firms in San Francisco for the time period March 1999 to December 2018

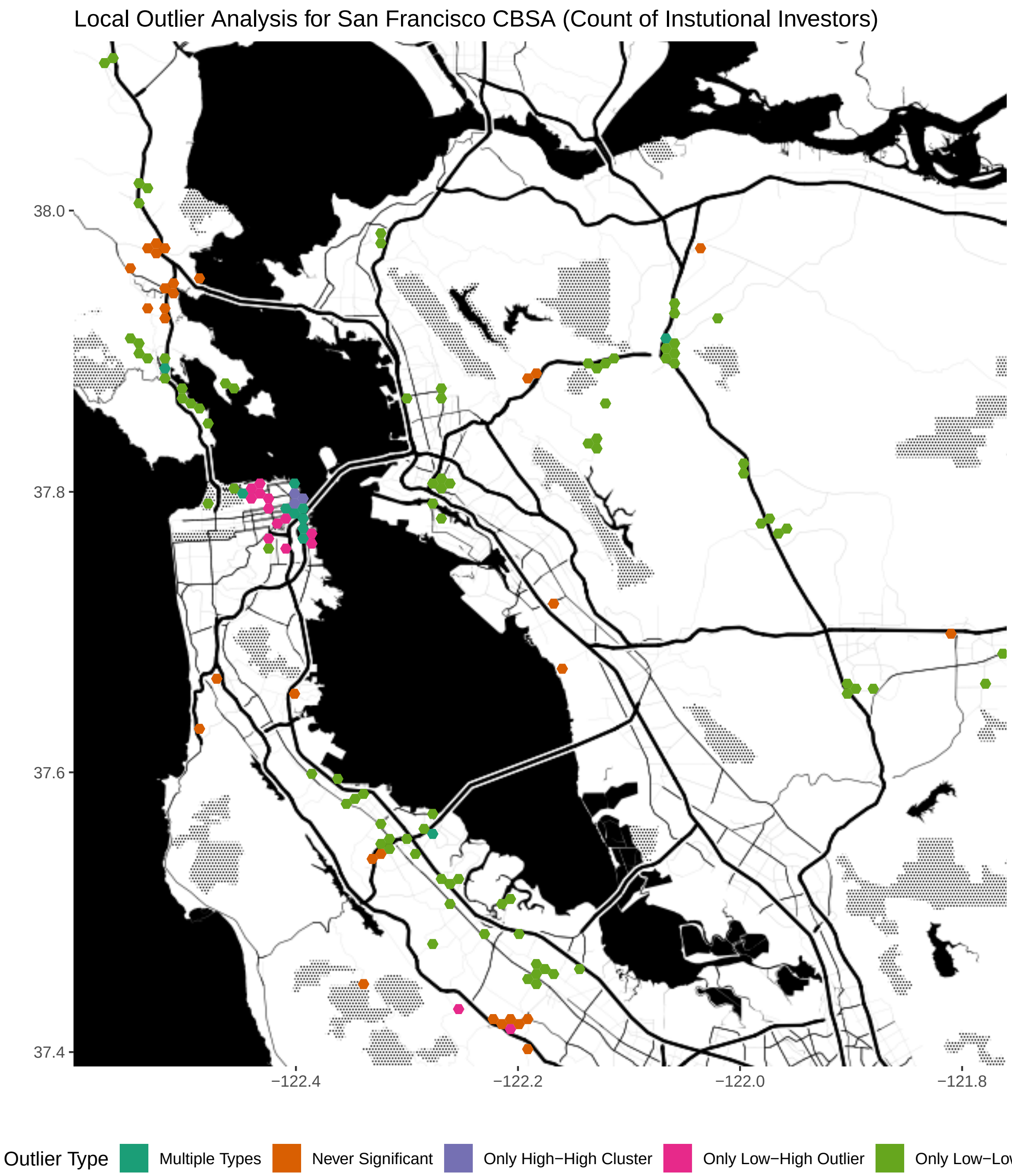


Figure 4.40: San Francisco Local Outlier Analysis - Count of Institutional Investors

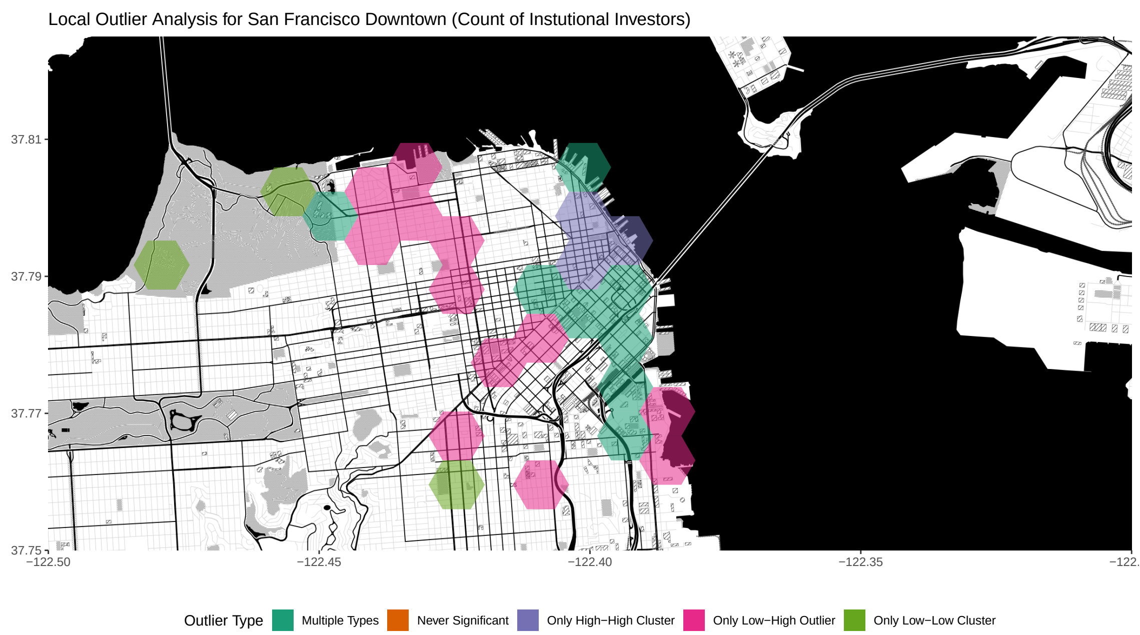


Figure 4.41: San Francisco Local Outlier Analysis - Count of Institutional Investors

York and San Francisco. While one might be tempted to think that this is merely a collection of large US metro areas, the absence of population rich regions such as Dallas-Fort Worth, Houston, Philadelphia, Washington DC, Miami and Atlanta from the ranks of top cities is reassuring that the top 5 cities isn't simply a replication of XKCD Comic 1138 (Figure 3.16) using institutional investors rather than subscribers to Martha Stewart.

Across these five cities, institutional investors exhibit a strong propensity to cluster, and more often than not these clusters are located in the downtown cores of cities. Even with Los Angeles lack of a highly developed CBD for a city of it's size, its sprawling nature and deliberately decentralized history, the existence of investor clusters somewhat pushes back against Graves's assertion that the benefits of co-

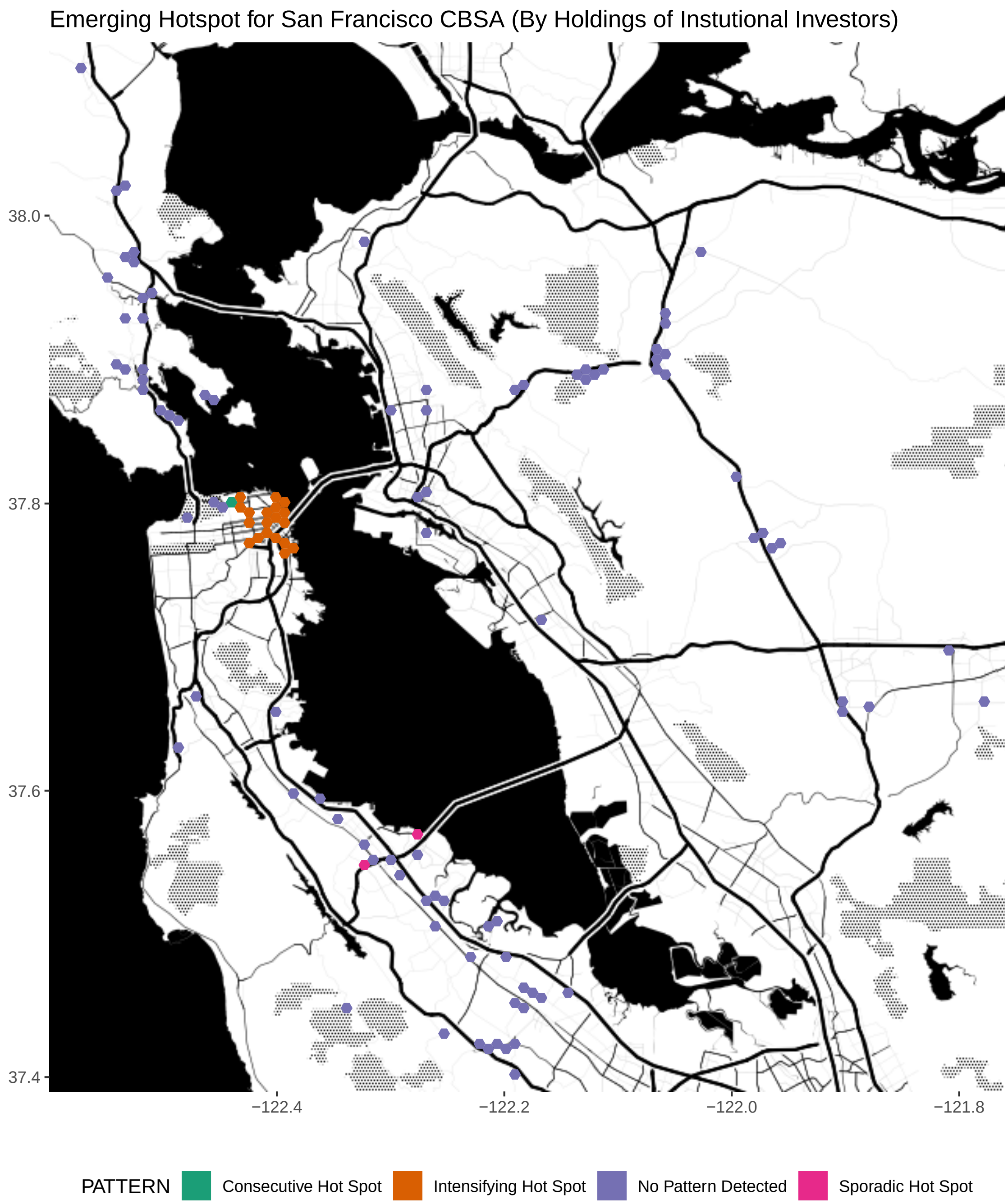


Figure 4.42: Emerging Hot Spot Analysis of Funds under Management for San Francisco for period June 2013 to December 2018

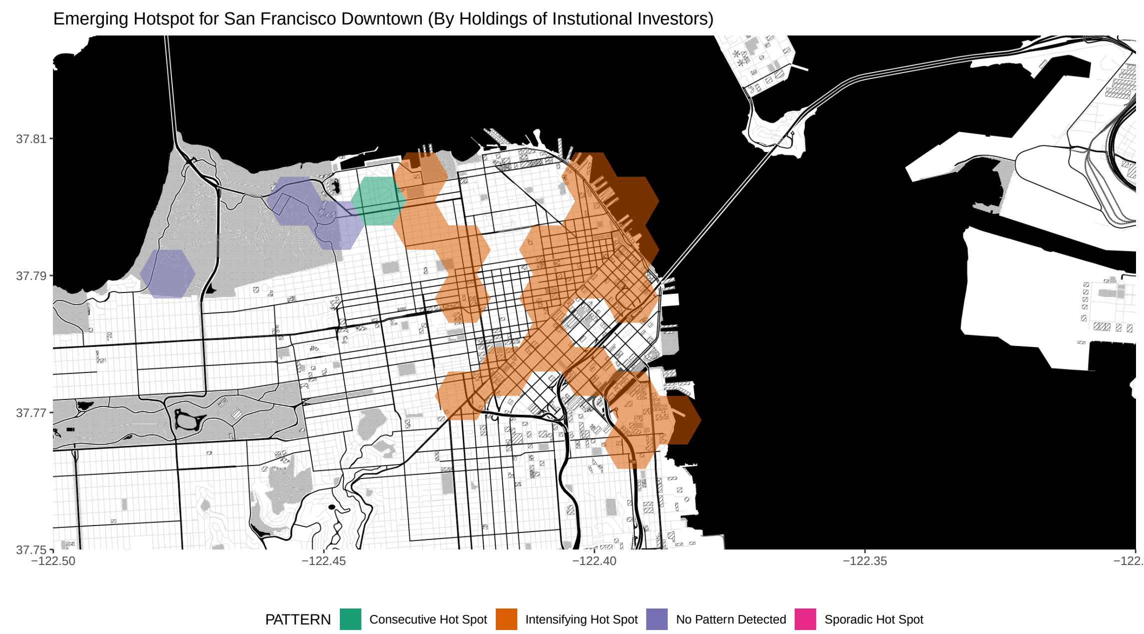


Figure 4.43: Emerging Hot Spot Analysis of Funds under Management for San Francisco for period June 2013 to December 2018

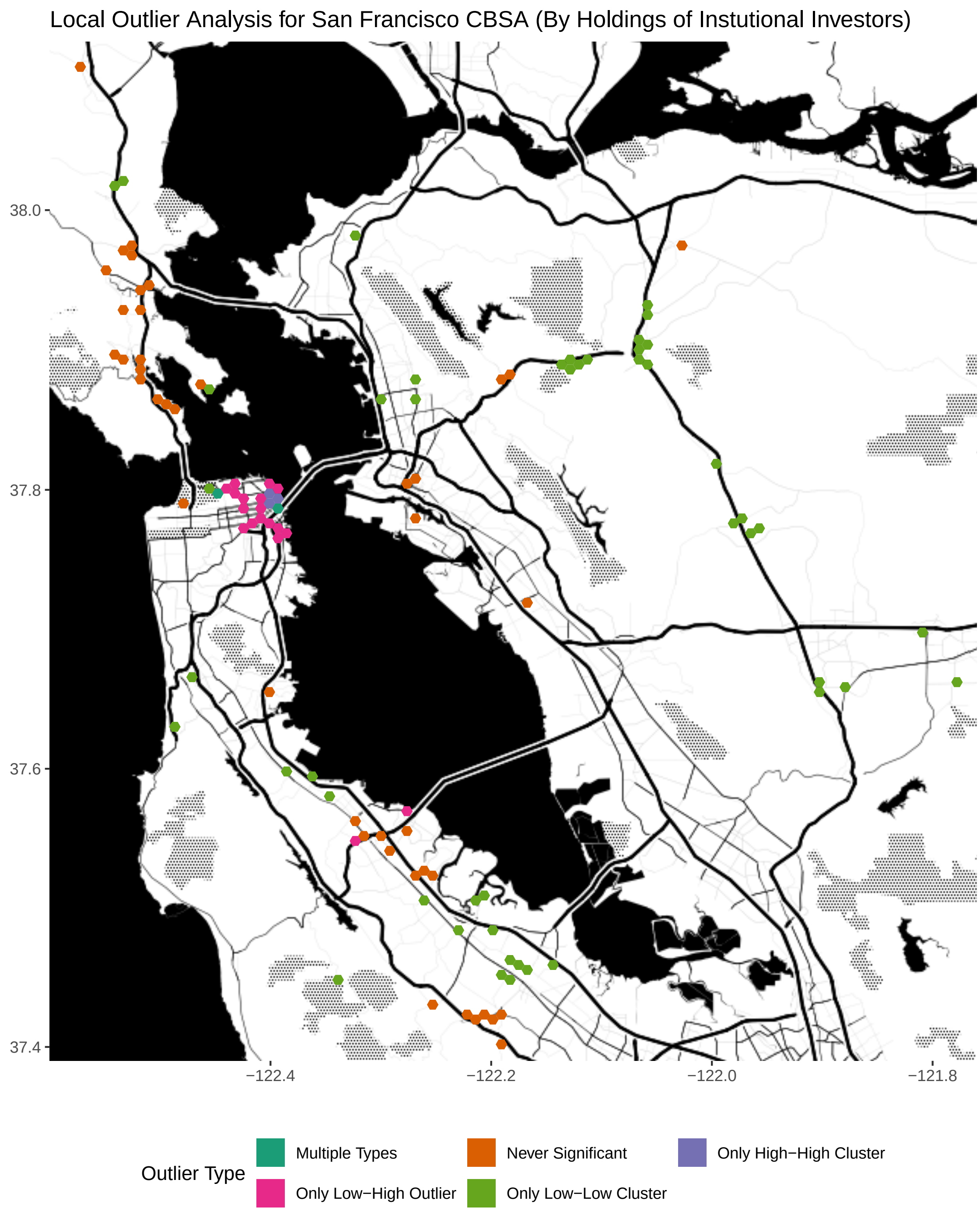


Figure 4.44: San Francisco Local Outlier Analysis - Funds under Management

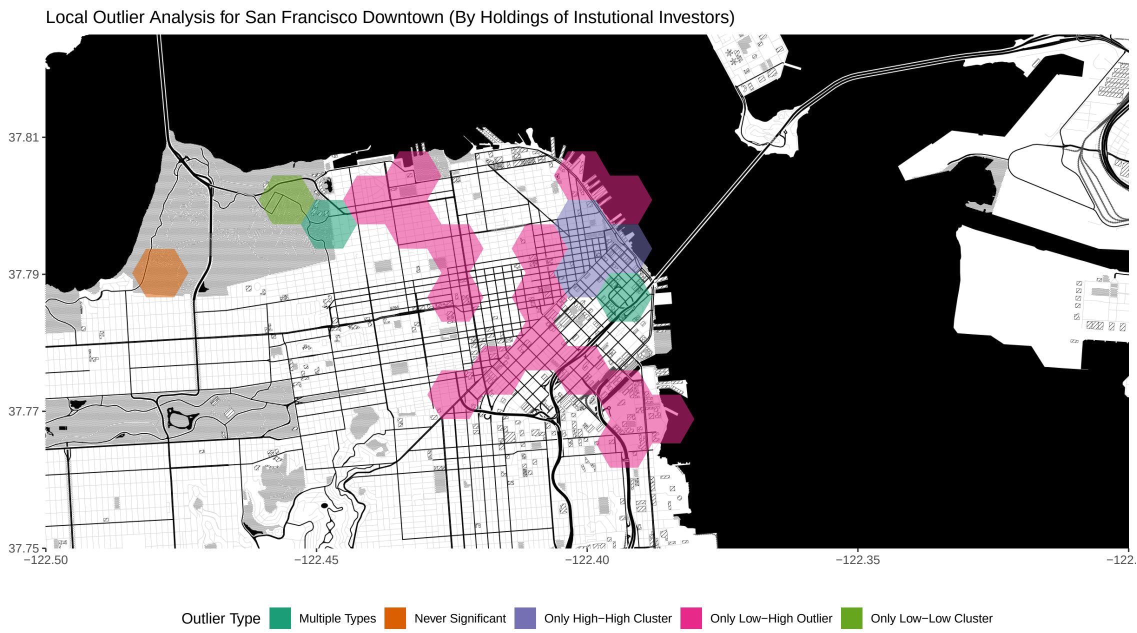


Figure 4.45: San Francisco Local Outlier Analysis - Funds under Management

location in an urban core were more than offset by the ever increasing cost of rents, and that investors of the future might seek more peripheral locations (Graves, 2003).

That being said, one should not forget that identifying clusters can be problematic. It is possible that new firms showing up on the periphery of a metro area's suburbia might not have the required density to show up a cluster, even as the total ratio between CBD and suburbs may tilt evermore into the suburban office park's favour. This is probably the most likely explanation for reconciling this chapter with Chapter 3. There are some hints at suburban centres being centres of clustering, notably the Route 128 in Boston, Evenston and Highland Park in Chicago, Irvine CA, and Walnut Creek in San Francisco. However, it should be noted that these areas have a historically smaller bankroll than the investors that tend to aggregate into CBD, suggesting that there might be a size threshold where being in the CBD becomes more worthwhile than in suburban office parks.

The buyer's remorse over choosing low land costs over a central location can been seen in the saga of the Swiss bank UBS. This Swiss-headquartered multinational bank was attracted by Stamford Connecticut's low land prices and generous tax incentives. However, this out of the way location became a severe hindrance in attracting top tier talent from New York's financial sector due to long commutes, as well as chronic difficulties in meeting with Manhattan-based clients (Bagli, 2011).

Chapter 5

LDA of Investments in the United States

5.1 Introduction

While chapter 4 explored the locational preferences of institutional investors in the US as a whole and in the five largest American metropolitan areas by total funds under management, this chapter will explore whether geography can play a role in individual investors portfolio choices.

While Modern Portfolio Theory (MPT), as established by Markowitz (1952) advocates for holding a broad and negatively correlated portfolio, the notion of ”not putting all of one’s eggs in a single basket” is an old one, for Lofthouse (1997) finds that such advice was formally practised by the British investment firm Investment Registry as far back as 1904.

In concert with MPT’s emphasis on diversification, the reaction to the Crash of

October 1987 placed renewed emphasis on risk management and the rise of "Value at Risk" (VAR) based investing in which firms would try to maximise returns while minimizing risk. This led to a homogenizing effect in investment strategies, as explained by Andrew G. Haldane, executive director of the Financial Stability at the Bank of England at a conference on risk management:

> Within the financial sector, diversity appears to have been reduced for two separate, but related, reasons: the pursuit of return; and the management of risk. The pursuit of yield resulted in a return on equity race among all types of financial firm. As they collectively migrated to high-yield activities, business strategies came to be replicated across the financial sector. Imitation became the sincerest form of flattery.
>
> So savings cooperatives transformed themselves into private commercial banks. Commercial banks ventured into investment banking. Investment banks developed in-house hedge funds through large proprietary trading desks. Funds of hedge funds competed with traditional investment funds. And investment funds - pension, money market mutual, insurance - imported the risk the others were shedding. (Haldane, 2009)[p.18]

As explored in Chapter 1, there is a substantial literature showing that stock pickers are biased towards industries in which they are knowledgeable or have personal connections. In particular, Coval and Moskowitz (2001) find that investors can draw abnormally high returns from local knowledge, and another study by Cohen et al. (2008) makes a compelling case that stock pickers are biased towards selecting stocks of companies that their board of directors contain shared alumni networks.

Rather than looking at geographic differences of investors based on the type of institution they belong to such as but not limited to banks, hedge funds, pension funds, and insurance companies, this study will attempt to create a functional portfolio archetypes using machine learning and aggregate these archetypes by geography in order to look for regional patterns.

5.2 Latent Dirichlet allocation

Latent Dirichlet allocation (LDA) is a generative statistical technique developed by David Blei to find themes that are common across a corpus of texts (Blei et al., 2003). This technique is a derivation and refinement of Papadimitriou et al. (1998) and Papadimitriou et al. (2000) work on Latent Semantic Indexing.

LDA has made certain classification tasks feasible to conduct in a short time, such as analysing a large sample of digitized 18th century American newspapers for the topics of the day that would otherwise be unfeasible for any individual to read (Newman and Block, 2006). Another well known use of LDA is for finding in near-realtime the topics of controversy and/or debate at an academic conference via Twitter usage by the participants of the conference (Marwick, 2014).

In addition to text analysis, LDA has been used in multiple different fields such as finding latent patterns in biodiversity data (Valle et al., 2014), genetic data, images, social networks (Blei, 2012) as well as remote sensing data Lienou et al. (2010).

5.2.1 How does LDA work?

Ted Underwood, who studies the intersection of Information Science and English Literature, contends in his academic blog post entitled "Topic modeling made just simple enough[sic]" that academic papers make LDA look much harder than it is in practice, since their main goal is to show how and why their underlying formulas work and the mathematical proofs rely on highly advanced mathematics. If we take the algorithms to work as intended, the practice of LDA can be easily explained in practice (Underwood, 2012).

LDA assumes that each document being analyzed contains a multitude of different topics, and each of these topics are latent, that is to say they can't be directly observed, but can be defined indirectly. Edwin Chen's classic introduction to LDA example is quite straight forward (Chen, 2011). Take the following five sentences:

1. I like to eat broccoli and bananas.
2. I ate a banana and spinach smoothie for breakfast.
3. Chinchillas and kittens are cute.
4. My sister adopted a kitten yesterday.
5. Look at this cute hamster munching on a piece of broccoli.

If we treat each sentence as a document for LDA purposes, and we were to limit ourselves to two topics, we would see something to the effect of the following:

- **Sentences 1 and 2:** 100% Topic A
- **Sentences 3 and 4:** 100% Topic B
- **Sentence 5:** 60% Topic A, 40% Topic B

At this point, we see that the topics consists of:

- **Topic A:** 30% broccoli, 15% bananas, 10% breakfast, 10% munching, etc...
- **Topic B:** 20% chinchillas, 20% kittens, 20% cute, 15% hamster, etc...

At which point, we can see that Topic A consists mostly of food and food adjacent activities, whereas Topic B is about animals and their general cuteness.

At this point, it is important to state that LDA assumes that language is a "bag of words". That is to say that for the purpose of the model, the order of words and punctuation isn't considered important information. While this may cause some miss-coding of information in a limited context, since grammar, punctuation and word order can relay important information, larger corpora smooth-out these ambiguities. For example, an LDA model would treat the following two sentences as being identical:

- Have you eaten, my child?
- Have you eaten my child‽

This study will be using LDA on Stock unique identifiers (CUSPI), the "bag of words" methodology works to our advantage, since the presented order of stocks in an institutional investor's portfolio will not influence the sorting algorithm. This relative location agnosticism is useful in this case, since unlike earth movers' distance classification, this method of classification is dependant on the initial relative distribution within the input variables, and therefore there is no need for a special ordering of stock positions in the input file.

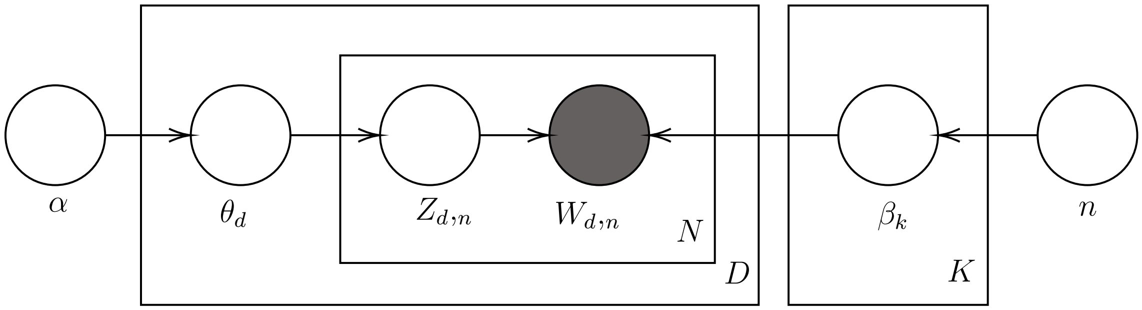


Figure 5.1: Graphical Model of Latent Dirichlet Allocation replicated from the graphic in Blei (2012), where K is the total number of topics, β_k is the topic, a distribution over the vocabulary, D is the total number of documents, Θ_d is the per-document topic proportions, N is the total number of words in a document, $Z_{d,n}$ is the per-word topic assignment, $W_{d,n}$ observed word, and finally α and n as dirichlet parameters.

The LDA process is mapped out graphically in Figure 5.1 and written out in Equation 5.1. For each possible topic (Z),

$$P(Z|W,D) = \frac{\#ofwordsWintopicZ + \beta_w}{totaltokensinZ + \beta} * (\#ofwordsinDthatbelongtoZ + \alpha) \quad (5.1)$$

That being said, for the purpose of this study, the

Considering that we are training the model on For our purpose of using LDA on stock portfolio, the relative ordering of the stocks has zero effect on the output, just the amount of each stock.

A closer analogue to using LDA is using this technique to classifying card selection in games such as Magic:The Gathering (Hlynsson, 2017). This collectible card game uses 60 cards decks that are selected ahead of time. Due the game's complex resource system and multiple different strategies for attacking one's opponent, cards are not fungible, and thus the game consolidates towards certain discreet collection of cards.

Similarly, the use LDA can be used to aggregate different stock portfolios into different investment strategies strategies.

5.3 Preparing the Data

In order to conduct an LDA analysis, the data was taken from the XBRL database of 13-F HR database for the period of the second quarter of 2013 to the end of 2018. The process used in collecting and cleaning this data was explained in Chapter **??**.

Unfortunately the database had to be pruned of all holdings of less than 1 million dollars so that the matrix operations conducted by the LDA package would fit within the computer's available RAM (Random Access Memory). At the time, these computers contained 32gb of RAM. This value of 1 million dollars was achieved in an iterative manner, with one computer starting with all transactions above 10 million dollars and reducing this threshold by 1 million dollars every time the LDA converged on a solution and a second computer starting with all transactions and pruning by increments of 100 000 USD until the algorithm converged rather than crash the program due to overwhelming the available RAM. Furthermore, due to the nature of the LDA algorithm (needing full matrix operations), it was unfeasible to spread the workload across multiple computers, nor to slice the program into year-long slices and perform 5 LDA analyses, since this would give us the worst of both worlds - no time continuity, and the multiple testing problem.

In practice, this reduces the size of the database from X to Y filers, and the value Xhat to Yhat. That being said, the pruning of the database focuses the analysis on stock positions that have substantial, if theoretical, corporate power rather than holdings that are simply intended passively to accrue in value and render dividends as part of a diversification strategy under the modern portfolio theory.

Furthermore, in this LDA analysis each filer-quarter is treated as independent filers in the LDA model. Stock positions do shift over time to the point that acting on information 45 days old can be ruinous, a fact that many whale watchers repeat in their newsletters and news reports (Brody, 2012; Brodie, 2013). Since stock positions shift over time to newer strategies, this should not pose a problem; for example this would treat a caterpillar and a butterfly differently. While indubitably the same creature, the caterpillar and the butterfly look, act, and occupy different ecological niches. This returns to the lumper-splitter problem. In this case, do we value tracing the metamorphosis, or the different niches both ends occupy? This treatment of investors and filing periods as discrete periods allows for the tracing of an investor's strategy shifting from predominantly X to predominantly Y. However, since the follow-on analysis will take time into effect, not having it in the original training model is simply a nod to feasibility.

Literary-based LDA suggests removing stop words. These words are command grammatical words such but not limited to pronouns, common adjectives and articles that make text understandable, but don't necessarily convey the latent topic. For example, any LDA analysis that uses English language prose would be overwhelmed by articles such as "the" and as such the most common word, and would thus saturate any analysis of say Sherlock Holmes books by Arthur Conan Doyle (Silge, 2018). That being said, there are no "words" - that is to say stock - that are as common as the word "the" in this analysis. In fact the most common CUSIP in the training database is CUSIP 037833100 (Apple Inc.), accounting for approximately one percent of all positions in the pruned database. While this popularity should not be surprising considering Apple's status in the investing world during the late aughts and the early to mid twenty-tens, this is nowhere as common as "the" or "they" in English prose.

Another practice that is common in literary-based uses of of LDA is stemming words. This removes prefixes and suffixes of words such that only their roots are used. For example, faster and fastest relate the same idea – fast. However, since the words used in this analysis are in-fact CUSIP numbers, there is no need for stemming. A case can be made that various class of stocks could have been stemmed since they are related to the same company, however this was not chosen since different class of stocks can be held for different reasons, such as using preferred stocks in a manner similar to bonds with the reduced voting rights exchanged for higher dividends and seniority. In other words, while different classes of stocks may be tied to the same company, they operate in different segments of portfolio allocation. For example, due to their promise to never force a stock split on their shareholders, Berkshire Hathaway was finding that their stock was getting into unwieldy large stock price, for investors would have to liquidate more stock value than they would usually need by selling one share. As such, partly to offer a more manageable stock denomination in order to ease buying into the fund by smaller investors, as well as scare-off index funds that would coast on Berkshire Hathaway's 13-F HR reports which chairperson Warren Buffet mused would lead to loss of goodwill due to the lower performance of these imitation index funds, Berkshire Hathaway renamed their existing stocks into

Berkshire Hathaway A and offered a newer stock with 1/30 the face value of Berkshire Hathaway A and lessor voting rights as Berkshire Hathaway B. (?) The class B stock was further split at a 1/50 ratio in 2010 to make the Berkshire Hathaway Class B stock to be equivalent to 1/1500th of a Berkshire Hathaway Class A stock (Crippen, Crippen)

5.4 Number of topics

LDA requires the user to determine *a priori* the number of topics used in the Topic Model. This leads to the lumper vs splitter problem. Where one has to classify n objects, the optimal number of categories will exist between 1 and n, for 1 category encompasses the ensemble of things to be classified, and n categories will have perfect fit, but is utterly meaningless since it does not reduce data into a meaningful form. As such, classification is an art as well as a science, since many categories can exist as part of a continuum.

In this case, the optimal number of topics selected was facilitated by the R package LDAtuning (Nikita, 2019). This package takes the Document-Term matrix and runs an ensemble of 4 different information criteria in order to find the optimal number of topics. These methods are Arun et al. (2010) Cao et al. (2009) Griffiths and Steyvers (2004) and Deveaud et al. (2014). From these four information criterion techniques, the suggested number of topics occurs where differences between these methods are minimized. Figure 5.2 displays the results of LDAtunings' estimates for the number of topics. This resultant plot shows that the numbers of topics where the differences

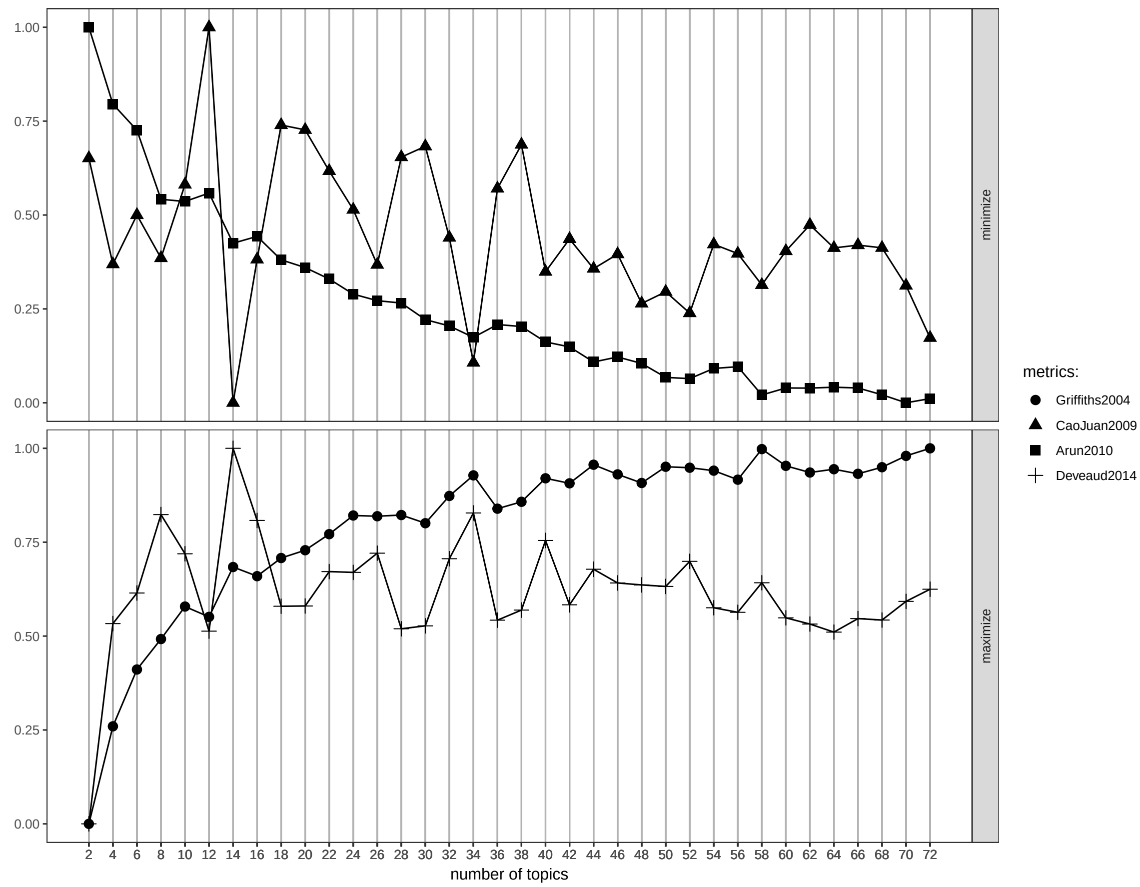


Figure 5.2: LDAtuning Ensemble for Determining the Number of Topics in LDA. As can be seen from the short distance between Deveaud(2014) and CaoJuan(2009) around 14 topics and the close agreement between the Griffiths(2004) and Arun(2010) measure as the number of topics increases - especially after 58. This suggests that a number of topics should be between 14 and 58. within this band, all 4 metrics are in closest agreement at 34 topics, therefore 34 topics will be used in the LDA analysis.

are minimized occur at 8, 14, 34 and 72 topics. However, we can further refine this for a better fit. A n of 8 and 14 offer a poor fit under Griffiths and Steyvers (2004), and thus this method suggest a much larger optimal number. By contrast, Cao et al. (2009) and Deveaud et al. (2014) suggest topics at 8, 14 and 34, with Deveaud et al. (2014) offering poorer solutions as the number of topics increases. As such, 34 topics offers the best compromise between the different tuning methods and was chosen.

5.5 Applying the Model to the Data

After the model is trained, the LDA provides two tables: beta table and gamma table. The first table, beta table, gives the probability of each stock belonging to each topic, whereas the second table, gamma table, contains the probability of each investor belonging to each topic.

5.5.1 Beta Table

Figures 5.3 to 5.6 display the 10 stocks with the highest probability of being assigned to each Topic. It should be noted that the order of each topic number is purely arbitrary, and nothing should be read in the rank-order of the different topics, nor the relative distance between topic numbers (Silge, 2018).

Within these topics, some are easier to label than others. For example, Topic 7 appears to be concentrated in Canadian banks as well as energy companies, Topic 9 suggests to be a smorgasbord of various ETF and indexed securities, whereas Topic 25 appears to be a strong collection of bluechip staples.

On the other hand, this 34 topic LDA gives us topics that would appear superficially similar, but are treated as different topics. For example, Topics 10 and 13 are anchored by Berkshire Hathaway stock, but the main difference between the two is that Topic 13 puts a much larger importance on the acumen of Warren Buffet than Topic 10's more diversified approach.

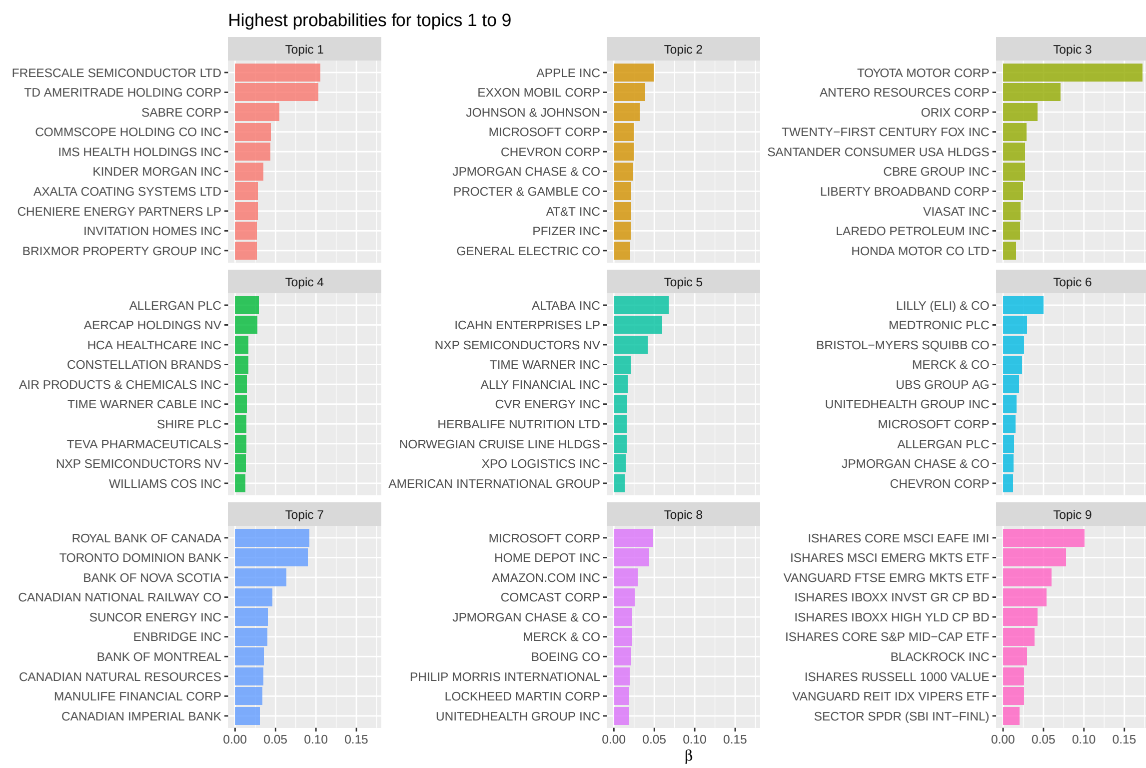


Figure 5.3: Topic Model with 34 Topics, Topics 1 thought 9. This represents the 10 most likely stocks being associated to a particular portfolio archetype.

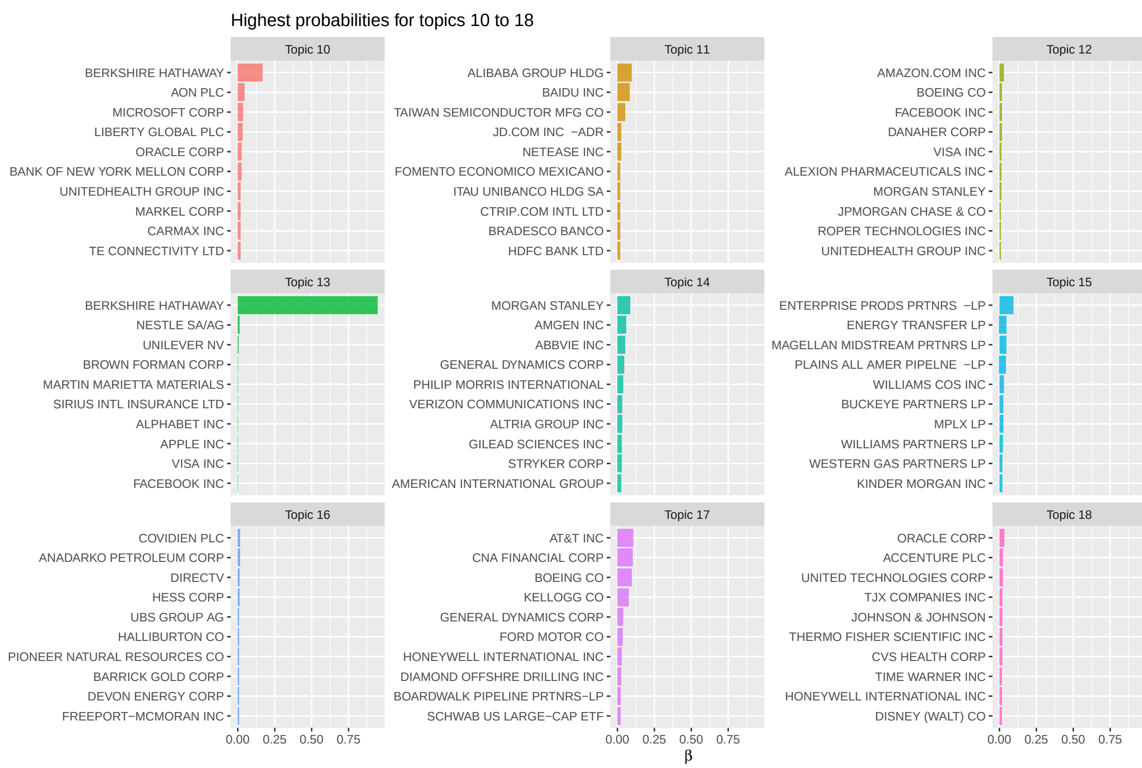


Figure 5.4: Topic Model with 34 Topics, Topics 10 thought 19. This represents the 10 most likely stocks being associated to a particular portfolio archetype.

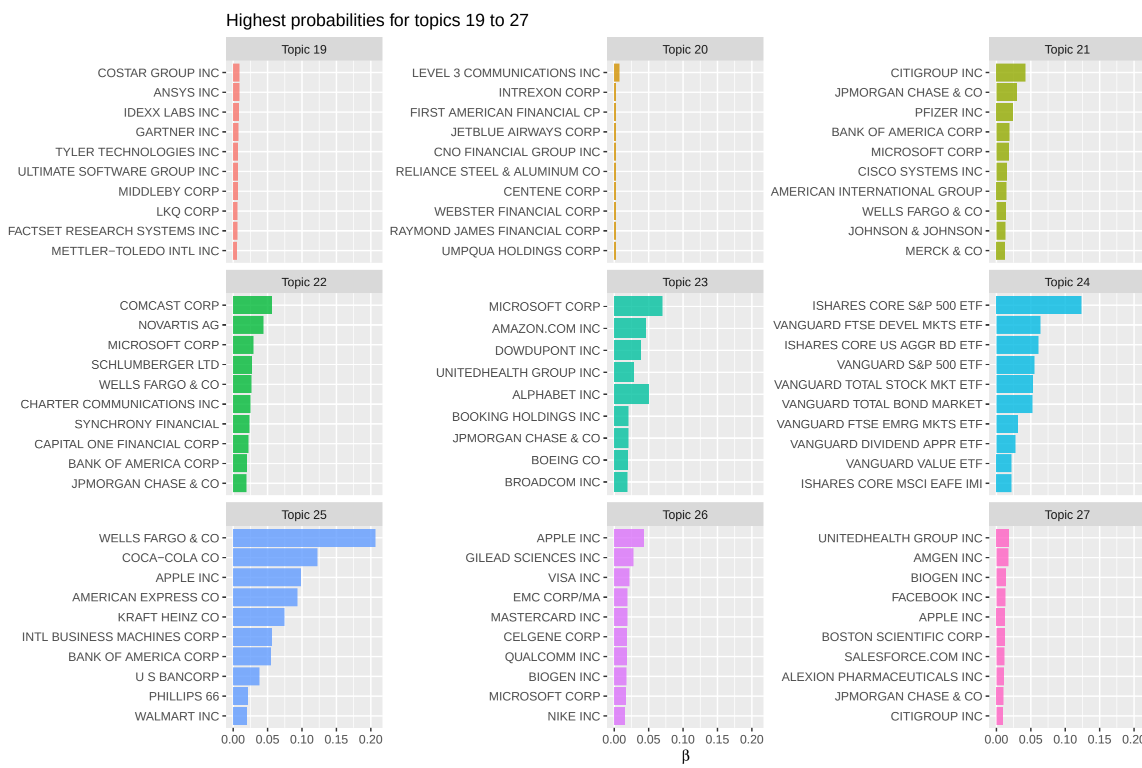


Figure 5.5: Topic Model with 34 Topics, Topics 19 thought 27. This represents the 10 most likely stocks being associated to a particular portfolio archetype.

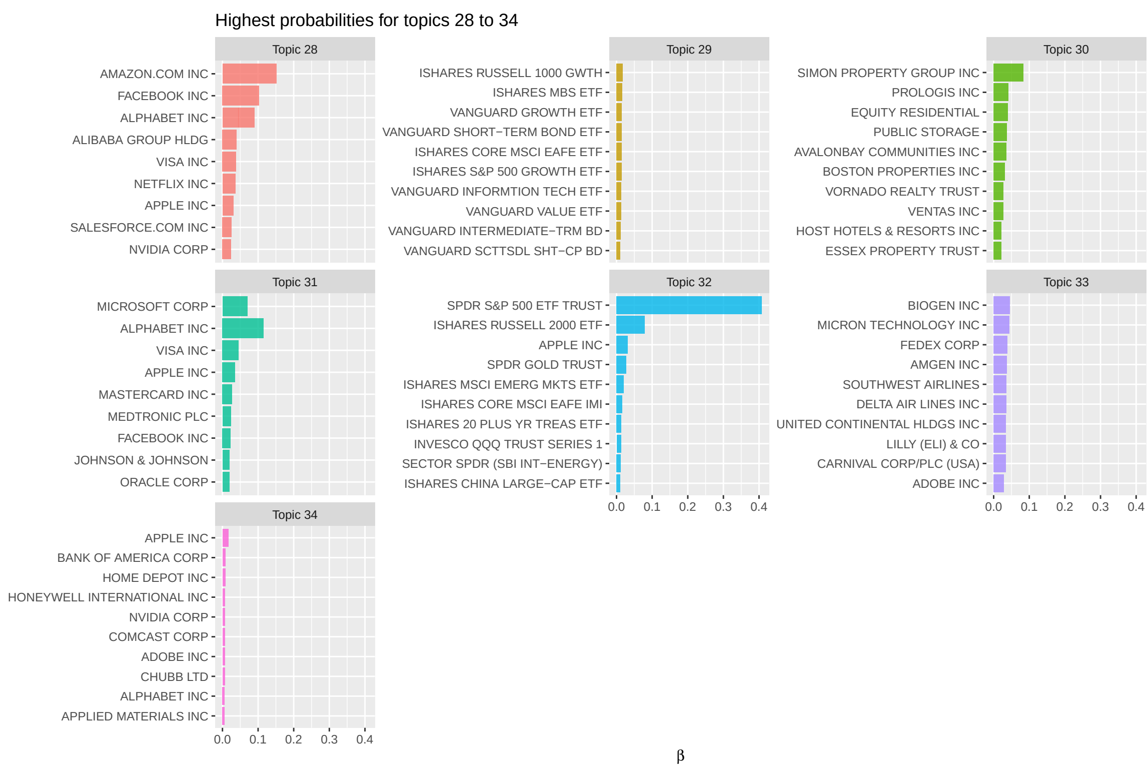


Figure 5.6: Topic Model with 34 Topics, Topics 28 thought 34. This represents the 10 most likely stocks being associated to a particular portfolio archetype.

5.5.2 Gama table

The per-document-per-topic probabilities is found in gamma table of the output. This table aggregates each stock's probability of belonging to a topic for each investor and thus gives the probability of each investor of belonging to each topic. The aggregate probability of each topic is displayed in Tables E.1 to E.3, giving us an idea of how the popularity of each topic fares over time. For example, Topic 26 saw a precipitous decline from 172.40 to 14.15 aggregate investor probability of belonging to this topic, conversely Topic 23 grew from 3.58 to 198.21 in this same metric.

Given that the investors were already geocoded in a previous chapter, the investors' topic probability was aggregated by State, and figures E.1 to E.34 were created using the geofacet package in R. These geofacet maps allows for the thematic representation of line graphs in a geometric patters that resembles the adjacency of US States, offering easier insights into the evolution of a topic over time than a series of choropleth maps.

Looking deeply at the aggregate investor probability tables gives hints at why certain seemingly related topics, such as Topics 10 and 13 – high concentrations of ETFs – as mentioned earlier might have a high thematic similarity, however these investors are given high probability classification to one topic have and a correspondingly low probability classification for the other topic. Going back to the fundamentals of Modern Portfolio Theory (MPT) might give insights into this outcome, and we are simply seeing two broadly similar strategies that are conceptually similar, but use different securities in the process. Furthermore, a look at the tables E.1 to E.3 indicates that these topics are getting more followers over time, however figures E.10 and E.13 show that this growth is geographically uneven, given that Topic 13 has most of its growth coming from investors located outside of New York State than is the case with Topic 10.

-As mentioned earlier, New York State the predominant location for investors, and this appears to be the case for most of the topics under study.

except Topic 29, 13 diffused 15, more in Texas than Cali, NY still king 16 Fades out over time

5.6 By State Results

-shift share was calculated as follows: Each portfolio was weighted using the beta table This weighted portfolio was aggrigated by State These state totals were used to perform Shift-Share.

Since each state had at least one investors for the duration of the time period, there was no need to add the fudge factor in order to prevent the divide by zero problem. This was nessessary for county level analysis, but since the principle county is highly reflective of the State.

- California-New York are often at odds

Topic 13 is very strong in Missouri, Texas and Illinois, while weak in NY and CA.

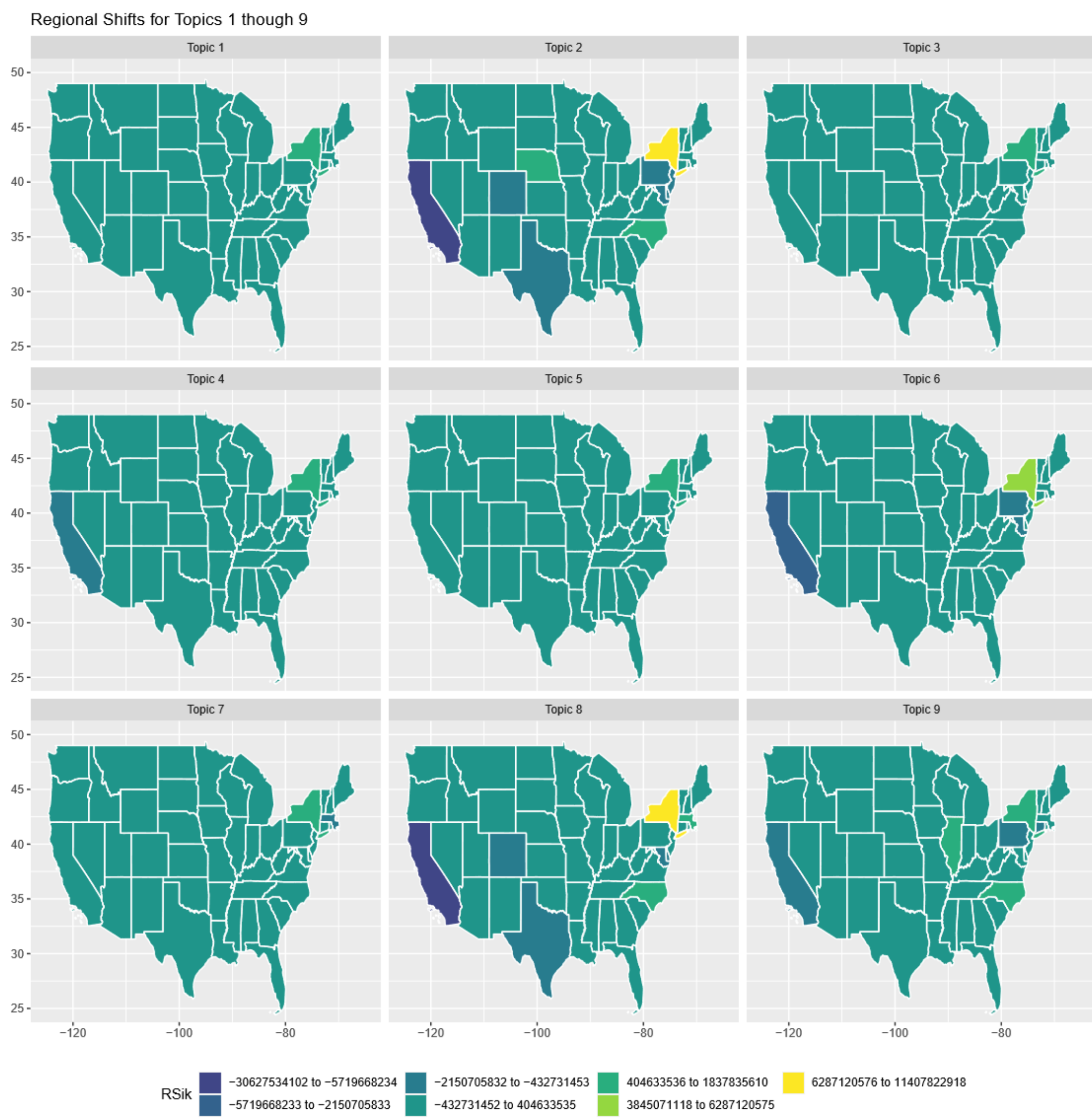


Figure 5.7: Regional shifts for topics 1 though 9 of the 34 topic LDA for the Continental USA.

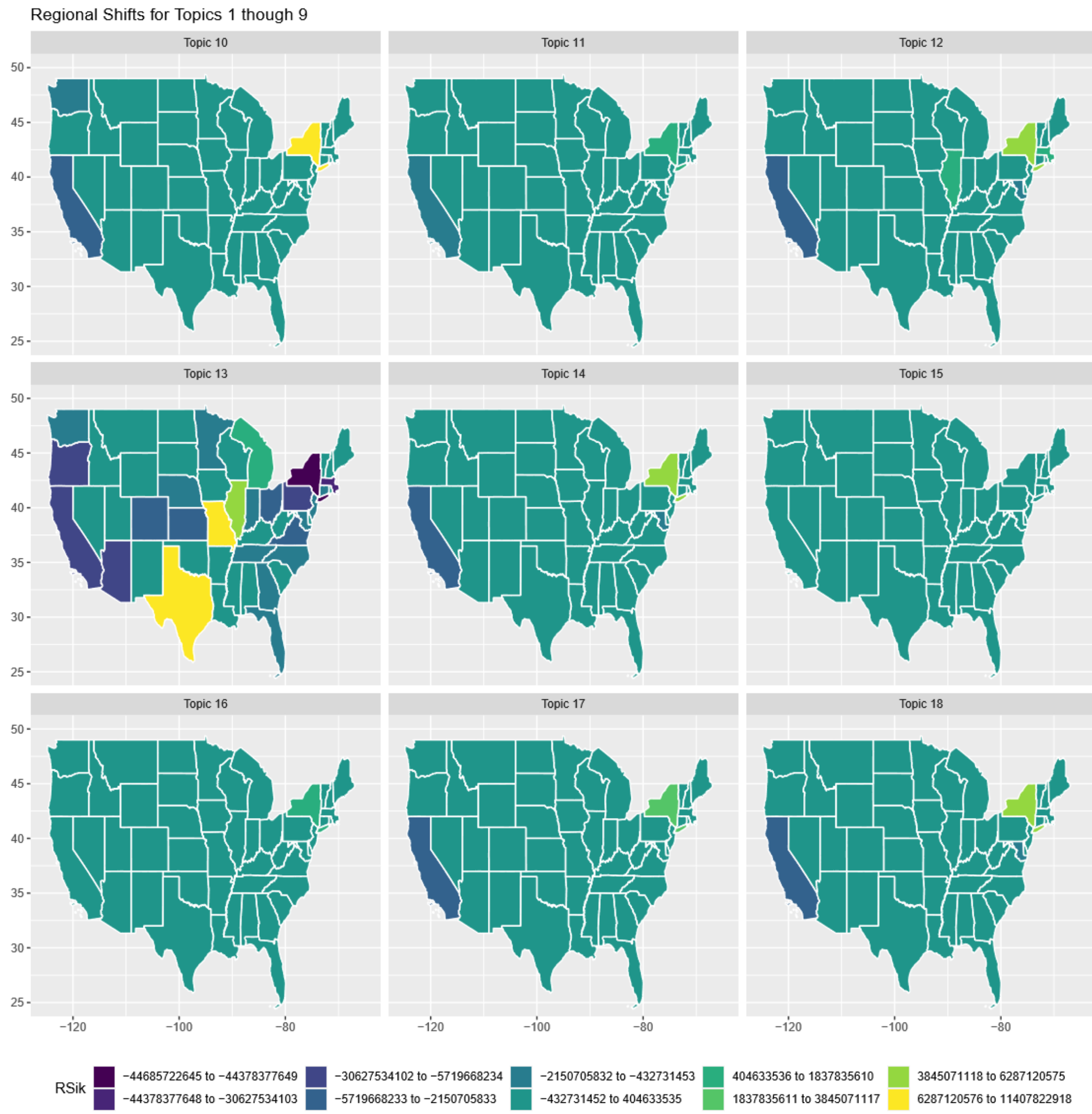


Figure 5.8: Regional shifts for topics 10 though 18 of the 34 topic LDA for the Continental USA.

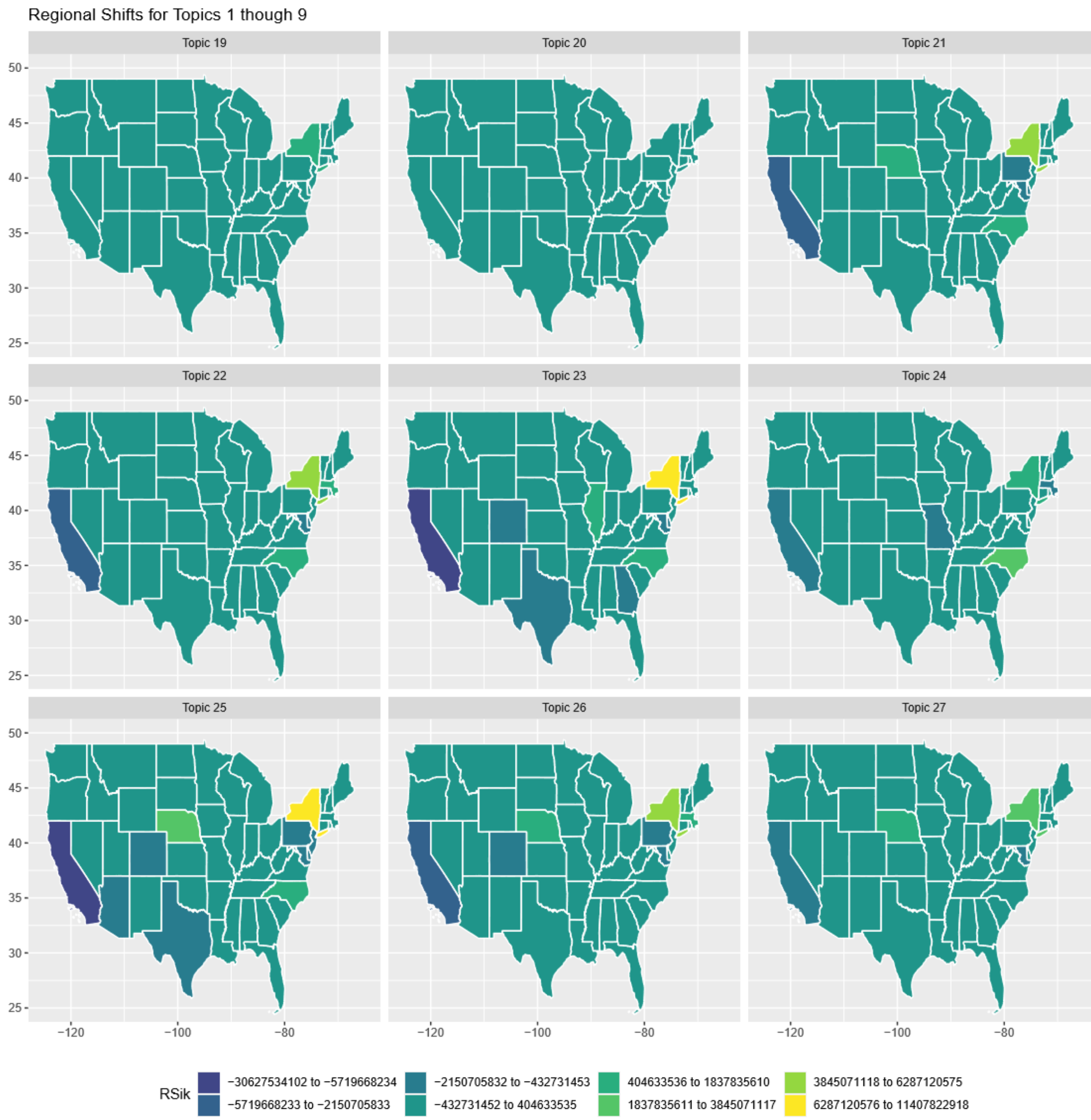


Figure 5.9: Regional shifts for topics 19 though 27 of the 34 topic LDA for the Continental USA.

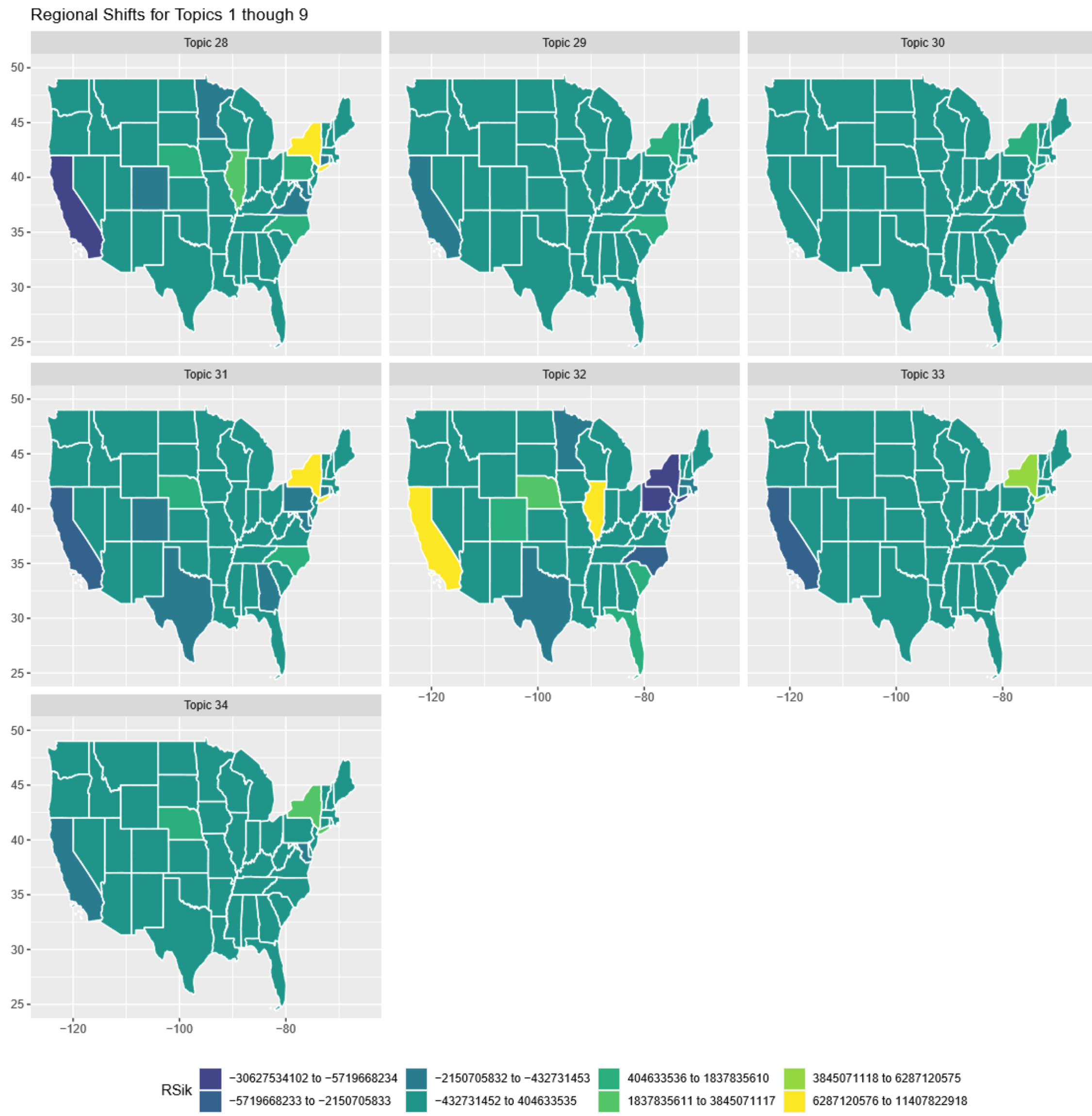


Figure 5.10: Regional shifts for topics 28 though 34 of the 34 topic LDA for the Continental USA.

Appendices

Appendix A

Total of Investors in US Counties by Year

County	1999Q1	1999Q2	1999Q3	1999Q4	2000Q1	2000Q2	2000Q3	2000Q4	2001Q1	2001Q2	2001Q3	2001Q4	2002Q1	2002Q2	2002Q3	2002Q4
New York County, New York	329	328	333	365	360	371	375	414	400	402	403	436	429	422	418	435
Suffolk County, Massachusetts	94	105	101	106	108	106	110	113	113	118	115	118	120	122	119	121
Cook County, Illinois	76	77	78	82	80	83	86	83	87	90	83	99	91	90	89	95
Fairfield County, Connecticut	59	63	61	66	65	63	62	65	66	68	66	70	66	66	67	72
San Francisco County, California	55	54	56	73	73	72	73	82	81	79	79	84	83	82	83	80
Los Angeles County, California	60	69	67	72	66	69	71	71	73	71	69	71	68	73	67	73
Harris County, Texas	22	21	24	30	27	27	27	29	28	26	26	23	24	23	22	23
Dallas County, Texas	18	17	16	19	17	21	18	23	24	23	23	20	19	18	18	16
Hennepin County, Minnesota	26	26	26	25	26	24	25	32	32	29	29	30	27	26	26	26
King County, Washington	16	15	15	18	18	17	18	21	21	21	21	20	21	21	19	20
Montgomery County, Pennsylvania	13	13	12	13	14	16	14	18	18	18	16	18	17	18	19	20
San Mateo County, California	17	17	16	28	28	28	29	40	38	37	36	26	25	25	25	21
Westchester County, New York	18	17	17	21	21	21	22	21	20	22	25	25	27	24	24	25
San Diego County, California	12	15	13	18	17	18	17	21	19	19	19	20	19	17	19	17
Fulton County, Georgia	17	16	15	15	12	18	14	15	17	17	18	18	17	19	17	16
Middlesex County, Massachusetts	10	10	9	10	10	9	10	13	12	13	13	15	16	15	14	16
Hamilton County, Ohio	15	16	15	16	16	15	14	16	18	19	17	19	20	20	20	20
Milwaukee County, Wisconsin	15	14	14	15	14	15	14	14	14	12	14	15	14	16	15	15
St. Louis County, Missouri	8	8	8	10	12	13	13	16	16	17	17	18	17	16	17	16
Montgomery County, Maryland	6	7	8	9	10	10	12	11	11	11	12	13	14	14	13	14
Denver County, Colorado	10	11	11	13	13	14	13	15	12	10	10	10	9	11	10	9
Santa Clara County, California	7	7	7	11	11	13	12	18	18	17	17	18	16	16	17	14
Baltimore County, Maryland	20	21	21	22	18	20	20	20	21	21	21	21	21	20	20	17
Cuyahoga County, Ohio	20	19	17	17	17	19	21	24	26	22	20	21	20	20	18	18
Orange County, California	9	11	12	12	13	13	14	13	13	13	12	14	15	14	14	14
Marin County, California	7	8	8	12	13	12	11	17	17	18	16	14	15	15	15	15
Chester County, Pennsylvania	10	12	12	13	15	15	15	15	13	13	15	14	14	14	14	15
Oakland County, Michigan	8	9	8	9	10	10	10	9	11	10	10	9	9	9	9	9
Tarrant County, Texas	18	18	18	19	18	18	18	19	17	17	17	21	21	22	22	18
Allegheny County, Pennsylvania	13	13	12	14	13	13	13	12	13	14	14	16	15	15	13	14

New Castle County, Delaware	12	13	13	14	12	9	12	11	13	11	12	12	12	12	10	11
Philadelphia County, Pennsylvania	9	10	11	12	12	13	13	12	13	14	14	14	14	16	15	16
Mecklenburg County, North Carolina	8	8	7	8	8	8	9	9	9	9	9	10	10	10	10	10
Hartford County, Connecticut	14	12	14	15	14	14	15	16	16	16	17	15	15	13	14	15
DuPage County, Illinois	4	4	4	6	5	5	6	6	5	5	5	5	5	5	4	6
Delaware County, Pennsylvania	15	14	12	12	11	12	9	13	12	11	11	11	11	12	12	11
Shelby County, Tennessee	9	7	9	9	9	9	9	9	8	8	9	10	10	10	10	10
Travis County, Texas	6	7	7	6	7	7	6	7	8	7	9	7	7	7	7	6
Morris County, New Jersey	8	8	8	7	8	8	7	7	8	8	8	10	10	12	11	12
District of Columbia	9	10	10	15	14	13	13	14	15	13	14	14	15	15	15	16
Bergen County, New Jersey	11	12	11	12	11	11	10	10	10	9	8	9	8	10	9	10
Douglas County, Nebraska	7	6	7	8	8	8	8	9	9	8	8	7	7	7	7	8
Johnson County, Kansas	3	3	3	4	4	3	4	4	4	4	4	5	5	5	5	5
Palm Beach County, Florida	6	6	6	6	6	3	3	5	4	5	5	6	5	4	4	4
Multnomah County, Oregon	11	11	11	13	13	14	13	13	10	10	10	10	10	10	10	10
Contra Costa County, California	4	5	5	7	5	5	6	7	6	6	6	6	5	5	5	5
Jefferson County, Kentucky	8	8	8	9	9	9	9	9	9	8	8	10	9	9	9	8
Richmond County, Virginia	7	9	9	8	7	6	7	7	7	9	9	8	8	8	8	8
Mercer County, New Jersey	5	5	5	8	8	9	10	8	10	8	8	6	5	6	5	4
Providence County, Rhode Island	9	9	7	9	7	6	6	6	6	6	6	6	6	6	6	5
Jackson County, Missouri	10	8	10	10	9	10	10	12	10	10	10	9	9	10	10	10
New Haven County, Connecticut	8	9	9	9	7	8	8	8	9	9	9	11	10	9	9	9
Norfolk County, Massachusetts	6	7	6	7	6	8	9	10	11	9	10	10	9	7	10	10
Maricopa County, Arizona	3	3	3	5	7	4	6	8	9	7	7	7	7	7	7	6
Monroe County, New York	6	5	6	6	6	6	6	6	5	7	6	6	8	6	6	5
Essex County, Massachusetts	2	2	2	5	5	6	6	8	7	7	8	7	7	6	6	5
Pinellas County, Florida	5	5	6	6	5	6	5	7	7	7	7	7	7	6	6	6
Hudson County, New Jersey	5	4	4	4	4	3	3	3	3	4	4	5	5	6	6	5
Jefferson County, Alabama	5	7	6	7	7	7	6	6	6	7	7	7	6	6	6	6
Henrico County, Virginia	3	3	3	3	3	4	4	4	4	6	6	5	5	5	5	5
Franklin County, Ohio	9	11	10	7	7	7	8	7	6	7	7	7	6	8	6	7

Cumberland County, Maine	7	6	6	7	7	7	7	6	4	3	4	5	5	5	4	7
Nassau County, New York	4	3	4	4	4	4	4	3	3	3	3	3	4	3	5	4
Baltimore County, Maryland	2	1	2	2	2	3	2	2	3	3	3	3	3	3	3	3
Arlington County, Virginia	7	7	7	6	7	7	8	6	6	6	7	9	8	8	7	8
Dane County, Wisconsin	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7
Davidson County, Tennessee	4	5	5	3	3	2	2	3	3	4	3	4	4	4	4	5
Duval County, Florida	2	3	2	4	4	4	4	5	5	5	5	6	6	6	5	6
Pulaski County, Arkansas	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4
Fairfax County, Virginia	3	5	3	4	3	4	4	4	3	3	3	3	3	4	2	3
Bexar County, Texas	4	3	3	3	3	3	3	3	3	5	5	3	3	3	3	3
Somerset County, New Jersey	3	4	4	4	3	4	4	4	5	5	5	5	4	5	5	5
Union County, New Jersey	0	0	0	0	0	0	0	0	0	0	1	2	2	2	2	3
Salt Lake County, Utah	4	15	9	9	9	9	9	9	9	5	5	5	5	5	5	5
Wake County, North Carolina	0	2	1	1	1	0	2	1	2	2	3	5	5	5	5	5
Butler County, Ohio	1	2	1	3	2	2	3	2	4	3	3	3	3	3	3	3
Miami-Dade County, Florida	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
Lake County, Illinois	2	2	2	2	2	2	2	2	2	1	3	3	3	3	3	3
Arapahoe County, Colorado	4	5	4	4	4	5	4	5	6	6	6	6	5	6	6	6
Tulsa County, Oklahoma	3	4	3	3	3	3	3	3	3	3	3	4	4	4	4	4
Essex County, New Jersey	4	4	4	7	7	7	7	7	7	6	5	7	6	6	6	6
Marion County, Indiana	4	4	4	3	3	3	4	6	5	7	6	7	7	8	8	8
Forsyth County, North Carolina	2	6	5	6	6	6	7	7	7	7	6	5	6	5	4	5
Wayne County, Michigan	3	3	3	5	4	4	4	5	5	5	5	5	4	4	4	4
Alameda County, California	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3
Lancaster County, Pennsylvania	3	3	3	3	3	3	3	3	3	4	4	4	4	3	4	4
Ramsey County, Minnesota	7	6	5	6	6	6	6	5	5	4	5	5	5	5	4	5
Charlottesville County, Virginia	1	2	2	3	3	3	3	5	5	5	4	4	5	5	5	5
Ozaukee County, Wisconsin	4	5	4	5	5	5	5	5	6	6	6	5	6	6	6	5
Polk County, Iowa	7	6	6	5	5	5	5	5	4	2	3	4	4	4	4	4
Cobb County, Georgia	4	5	5	6	6	6	6	5	4	4	4	5	5	5	4	6
Albany County, New York	3	3	3	1	2	2	2	2	2	2	2	2	2	2	2	2

Collier County, Florida	1	2	1	1	1	1	1	1	1	1	1	2	2	3	3	3
Kent County, Michigan	4	5	5	5	5	5	5	6	6	5	5	4	4	4	4	4
Rockingham County, New Hampshire	3	3	3	4	4	4	5	7	4	4	4	5	5	5	5	5
Erie County, New York	4	4	4	4	4	4	4	5	4	4	4	4	5	5	5	4
Plymouth County, Massachusetts	2	2	2	3	3	1	4	3	3	3	3	1	2	2	2	2
Montgomery County, Ohio	5	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5
Merrimack County, New Hampshire	3	3	3	4	4	4	4	4	6	6	6	5	6	5	5	5
Orange County, Florida	4	5	5	4	5	5	5	6	6	5	5	5	5	4	4	4

County	2003Q1	2003Q2	2003Q3	2003Q4	2004Q1	2004Q2	2004Q3	2004Q4	2005Q1	2005Q2	2005Q3	2005Q4	2006Q1	2006Q2	2006Q3	2006Q4
New York County, New York	425	425	413	461	456	406	492	512	507	500	489	562	554	555	559	611
Suffolk County, Massachusetts	123	123	124	126	124	119	125	133	128	126	130	144	143	143	143	155
Cook County, Illinois	95	96	96	105	103	96	106	104	103	101	99	111	109	111	108	117
Fairfield County, Connecticut	74	75	71	83	82	75	87	92	89	91	91	105	103	105	103	123
San Francisco County, California	74	71	72	81	82	79	84	94	91	91	90	96	96	94	90	89
Los Angeles County, California	70	73	70	78	76	75	74	79	81	80	77	87	86	87	88	101
Harris County, Texas	23	21	22	22	25	24	24	27	28	29	29	30	31	31	30	36
Dallas County, Texas	15	16	17	22	22	20	23	32	33	33	32	34	35	34	35	40
Hennepin County, Minnesota	27	27	27	32	33	33	32	35	35	34	35	36	36	36	36	40
King County, Washington	20	20	21	25	24	24	25	23	25	24	23	23	25	24	24	25
Montgomery County, Pennsylvania	23	23	23	27	28	28	27	31	32	29	34	36	34	34	36	37
San Mateo County, California	20	19	22	19	19	19	19	18	18	17	18	21	21	20	20	27
Westchester County, New York	25	25	23	22	21	22	21	29	28	29	30	28	28	27	26	28
San Diego County, California	17	17	17	19	19	20	20	21	20	21	21	21	23	22	22	28
Fulton County, Georgia	16	17	17	17	17	17	16	15	16	14	17	16	17	17	17	19
Middlesex County, Massachusetts	16	16	16	16	17	17	17	19	19	20	18	18	19	19	18	21
Hamilton County, Ohio	20	20	20	19	19	17	19	19	22	22	23	23	23	23	22	24
Milwaukee County, Wisconsin	16	18	16	18	17	17	23	23	23	23	23	22	21	21	20	24
St. Louis County, Missouri	16	16	16	15	16	15	15	16	14	14	14	16	15	14	17	17
Montgomery County, Maryland	14	14	14	16	17	17	17	19	18	18	18	18	18	19	20	21
Denver County, Colorado	9	10	9	13	14	14	13	15	16	14	14	16	16	15	15	19
Santa Clara County, California	15	15	16	16	15	15	14	18	16	15	16	17	16	16	17	17
Baltimore County, Maryland	18	18	17	17	16	17	17	16	16	16	16	18	18	16	16	19
Cuyahoga County, Ohio	17	18	17	19	19	19	19	18	18	19	19	19	18	18	17	17
Orange County, California	13	13	13	13	12	10	13	14	14	15	13	15	13	13	13	19
Marin County, California	13	13	13	15	16	16	17	18	17	17	16	16	18	18	18	22
Chester County, Pennsylvania	15	15	15	17	16	18	18	17	18	17	17	18	21	21	21	24
Oakland County, Michigan	9	9	9	10	10	10	11	11	11	12	12	12	11	12	12	17
Tarrant County, Texas	16	16	16	15	15	15	15	16	16	16	16	15	15	15	15	18
Allegheny County, Pennsylvania	14	14	14	16	14	15	15	15	15	14	15	14	14	14	15	15

New Castle County, Delaware	10	10	10	11	10	10	10	9	10	14	10	11	12	13	13	10
Philadelphia County, Pennsylvania	15	15	15	15	15	15	13	14	12	11	11	12	12	12	13	14
Mecklenburg County, North Carolina	12	12	12	13	12	13	11	12	14	14	13	15	14	14	14	14
Hartford County, Connecticut	14	14	13	13	13	13	11	12	12	11	10	14	13	14	14	15
DuPage County, Illinois	6	6	6	9	9	9	9	12	11	11	12	14	13	13	13	14
Delaware County, Pennsylvania	9	10	9	11	11	12	12	11	11	12	10	10	11	11	10	9
Shelby County, Tennessee	11	10	9	11	11	11	11	10	10	10	10	11	11	11	11	11
Travis County, Texas	7	7	7	7	7	7	7	7	7	7	7	8	8	8	8	9
Morris County, New Jersey	13	13	12	14	13	12	13	12	12	13	13	13	13	13	13	14
District of Columbia	16	16	16	17	16	14	14	15	15	14	14	16	15	15	13	17
Bergen County, New Jersey	10	10	10	9	9	6	8	11	10	11	12	12	11	12	11	15
Douglas County, Nebraska	8	8	8	8	8	8	10	10	10	10	10	12	11	11	11	13
Johnson County, Kansas	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	7
Palm Beach County, Florida	4	4	5	5	6	7	7	7	8	8	6	7	8	8	9	10
Multnomah County, Oregon	10	10	9	7	7	7	8	8	8	8	8	8	8	8	8	10
Contra Costa County, California	5	5	5	6	6	6	6	6	7	7	7	9	8	9	9	11
Jefferson County, Kentucky	9	9	9	9	8	9	9	11	12	12	12	14	13	13	12	13
Richmond County, Virginia	8	7	7	8	9	9	9	10	10	10	10	12	11	10	10	9
Mercer County, New Jersey	4	5	5	5	6	6	6	8	8	8	8	8	8	8	7	8
Providence County, Rhode Island	4	5	5	7	7	7	7	8	8	7	8	9	9	9	9	11
Jackson County, Missouri	10	10	10	10	10	10	9	9	9	9	9	10	10	10	10	10
New Haven County, Connecticut	9	9	10	10	9	11	10	11	8	7	16	12	11	11	12	12
Norfolk County, Massachusetts	10	10	10	10	10	10	10	11	11	11	11	12	12	10	10	8
Maricopa County, Arizona	7	7	7	3	3	3	3	5	5	8	6	7	6	6	6	7
Monroe County, New York	6	6	6	8	8	8	8	8	8	9	8	9	9	9	8	8
Essex County, Massachusetts	5	5	6	6	6	6	6	7	7	7	7	6	6	7	7	8
Pinellas County, Florida	6	6	6	6	5	6	6	8	9	10	10	10	10	9	9	8
Hudson County, New Jersey	4	5	6	7	7	7	7	6	6	7	8	5	7	7	7	7
Jefferson County, Alabama	6	6	6	6	6	6	6	5	6	6	6	6	6	6	6	7
Henrico County, Virginia	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	8
Franklin County, Ohio	7	6	6	7	7	7	7	7	6	7	5	7	8	7	7	8

Cumberland County, Maine	5	6	6	6	6	6	6	6	6	6	6	7	7	8	7	7
Nassau County, New York	4	4	4	5	5	6	6	8	8	8	8	7	7	8	6	8
Baltimore County, Maryland	3	3	3	4	4	5	5	5	5	5	5	5	6	6	6	6
Arlington County, Virginia	8	8	8	10	9	9	9	9	11	10	10	11	11	10	10	10
Dane County, Wisconsin	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8
Davidson County, Tennessee	5	5	5	6	6	6	7	7	9	8	8	8	7	8	8	8
Duval County, Florida	7	7	8	8	8	6	8	7	7	7	7	10	9	10	11	10
Pulaski County, Arkansas	4	4	4	4	6	4	4	3	5	5	5	5	5	5	5	8
Fairfax County, Virginia	3	3	3	2	3	4	4	4	4	4	4	4	5	4	4	6
Bexar County, Texas	3	3	3	4	4	4	4	5	5	5	5	5	5	5	5	5
Somerset County, New Jersey	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4	5
Union County, New Jersey	2	2	2	2	3	3	3	5	5	4	6	5	5	5	5	7
Salt Lake County, Utah	5	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5
Wake County, North Carolina	4	4	4	6	5	5	5	5	5	5	5	5	5	4	4	5
Butler County, Ohio	3	3	3	8	8	8	8	8	8	8	8	8	8	9	9	8
Miami-Dade County, Florida	2	2	2	2	2	1	3	1	1	1	1	1	2	2	2	5
Lake County, Illinois	2	2	3	4	4	5	5	6	6	6	6	6	6	6	6	6
Arapahoe County, Colorado	7	7	7	7	6	6	6	7	7	7	7	7	7	5	5	7
Tulsa County, Oklahoma	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4
Essex County, New Jersey	6	6	5	4	4	4	4	5	5	5	5	6	6	6	5	5
Marion County, Indiana	8	8	8	6	6	6	6	6	6	5	5	5	4	4	4	4
Forsyth County, North Carolina	4	4	4	5	4	4	4	6	6	6	6	6	6	5	5	6
Wayne County, Michigan	4	4	4	4	4	4	4	5	4	4	4	5	5	5	5	5
Alameda County, California	3	2	2	2	2	2	3	3	3	3	3	3	3	3	3	5
Lancaster County, Pennsylvania	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
Ramsey County, Minnesota	5	5	5	6	5	6	6	6	6	6	6	8	7	7	7	7
Charlottesville County, Virginia	5	5	5	4	4	4	4	4	3	3	3	4	4	4	4	4
Ozaukee County, Wisconsin	5	5	5	5	5	5	5	4	4	4	4	4	4	4	4	5
Polk County, Iowa	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5
Cobb County, Georgia	5	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4
Albany County, New York	3	3	3	3	4	4	4	5	5	6	5	7	6	6	6	6

Collier County, Florida	3	3	2	2	3	3	3	3	2	2	2	2	2	2	2	3
Kent County, Michigan	4	4	4	4	5	4	5	5	5	5	4	5	4	4	4	4
Rockingham County, New Hampshire	5	5	4	5	5	5	5	5	5	5	5	4	5	6	5	5
Erie County, New York	5	5	5	5	5	5	5	6	6	5	5	5	5	5	5	5
Plymouth County, Massachusetts	3	2	1	2	2	2	2	2	2	2	2	2	2	3	3	4
Montgomery County, Ohio	5	5	5	5	5	4	4	5	5	5	5	5	5	5	5	5
Merrimack County, New Hampshire	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Orange County, Florida	4	4	4	4	4	3	3	3	3	3	3	5	4	5	5	4

County	2007Q1	2007Q2	2007Q3	2007Q4	2008Q1	2008Q2	2008Q3	2008Q4	2009Q1	2009Q2	2009Q3	2009Q4	2010Q1	2010Q2	2010Q3	2010Q4
New York County, New York	612	603	602	676	660	657	635	628	613	603	596	569	569	560	562	645
Suffolk County, Massachusetts	154	152	154	164	159	163	162	160	154	158	159	156	154	152	150	160
Cook County, Illinois	114	117	114	128	132	130	125	137	129	134	130	136	132	133	132	137
Fairfield County, Connecticut	120	116	115	130	132	131	129	127	129	122	123	118	114	115	114	120
San Francisco County, California	89	87	89	97	97	98	96	95	92	92	93	97	98	100	100	106
Los Angeles County, California	100	103	99	110	110	109	110	108	101	94	91	91	92	88	88	98
Harris County, Texas	36	36	38	44	42	42	40	41	41	41	41	39	41	39	40	41
Dallas County, Texas	40	40	40	44	44	45	46	44	42	43	44	38	37	37	38	41
Hennepin County, Minnesota	39	38	39	43	43	43	43	43	43	42	41	39	39	39	40	39
King County, Washington	25	25	24	28	28	28	27	32	31	29	29	30	30	31	29	32
Montgomery County, Pennsylvania	35	36	34	34	33	32	34	33	33	32	31	31	30	30	28	30
San Mateo County, California	26	26	26	27	27	26	28	25	23	21	22	22	23	24	23	25
Westchester County, New York	28	28	29	32	31	31	29	28	27	30	28	27	30	29	30	32
San Diego County, California	27	27	25	28	28	29	30	27	26	26	25	26	26	25	25	27
Fulton County, Georgia	19	20	20	24	22	25	24	26	26	27	26	28	28	28	28	29
Middlesex County, Massachusetts	19	20	20	25	27	24	26	23	22	22	21	23	23	23	23	25
Hamilton County, Ohio	23	21	22	24	24	24	25	24	24	23	23	23	25	24	24	24
Milwaukee County, Wisconsin	21	23	24	23	23	23	24	25	26	25	25	26	25	25	25	25
St. Louis County, Missouri	16	17	17	20	20	20	20	20	19	21	20	23	23	23	22	28
Montgomery County, Maryland	21	20	21	21	24	22	22	23	23	24	24	23	25	23	21	22
Denver County, Colorado	19	19	19	19	19	19	19	19	19	20	20	22	21	22	21	22
Santa Clara County, California	17	17	17	19	17	18	17	19	17	17	17	17	15	15	16	18
Baltimore County, Maryland	17	17	17	23	23	23	22	21	21	20	21	15	26	14	20	21
Cuyahoga County, Ohio	15	14	15	16	15	13	14	13	13	12	13	14	16	15	16	17
Orange County, California	18	18	17	20	18	19	17	17	16	17	18	17	17	16	15	21
Marin County, California	20	20	19	19	19	18	20	19	18	17	15	18	16	16	16	19
Chester County, Pennsylvania	24	24	23	25	24	23	23	23	21	21	21	21	19	17	16	17
Oakland County, Michigan	17	17	17	18	18	20	19	22	22	22	21	21	22	22	21	23
Tarrant County, Texas	18	18	18	18	18	18	19	21	20	20	20	16	17	18	18	20
Allegheny County, Pennsylvania	15	15	13	14	13	12	14	13	13	12	14	13	15	15	15	15

New Castle County, Delaware	10	12	13	14	13	13	14	16	16	15	14	18	18	18	19	20
Philadelphia County, Pennsylvania	14	14	14	14	13	13	11	11	13	14	14	12	16	14	15	15
Mecklenburg County, North Carolina	14	14	14	16	16	15	16	15	16	15	14	16	16	16	16	18
Hartford County, Connecticut	16	16	16	16	17	17	17	15	15	14	14	13	13	13	13	12
DuPage County, Illinois	14	14	15	16	15	15	16	15	15	15	15	13	14	13	14	16
Delaware County, Pennsylvania	10	10	12	10	12	11	11	10	9	9	9	10	10	10	10	11
Shelby County, Tennessee	11	11	10	13	11	12	12	12	12	11	11	14	15	15	15	17
Travis County, Texas	9	9	9	13	13	12	14	12	12	12	12	12	13	13	14	14
Morris County, New Jersey	14	14	14	14	14	12	13	13	13	13	12	12	12	11	11	12
District of Columbia	18	19	18	19	18	18	18	16	15	16	14	14	13	12	12	15
Bergen County, New Jersey	12	11	9	13	13	13	13	7	9	9	9	11	11	11	11	13
Douglas County, Nebraska	12	11	11	11	11	10	12	12	10	11	11	12	12	12	12	12
Johnson County, Kansas	7	7	7	9	9	10	10	10	11	11	10	12	12	12	12	13
Palm Beach County, Florida	11	13	11	11	11	12	12	12	12	11	9	12	13	13	13	14
Multnomah County, Oregon	10	10	10	11	11	11	11	11	11	11	11	12	12	10	10	10
Contra Costa County, California	11	11	11	10	10	10	9	11	10	9	8	8	9	10	10	10
Jefferson County, Kentucky	13	13	13	13	13	12	12	13	11	10	10	11	9	9	9	11
Richmond County, Virginia	10	10	10	10	10	10	10	10	8	8	8	8	8	8	8	8
Mercer County, New Jersey	8	8	7	9	9	9	10	8	7	8	10	15	14	14	14	14
Providence County, Rhode Island	11	11	11	11	11	11	10	10	10	10	10	10	11	10	10	10
Jackson County, Missouri	10	10	10	10	10	10	10	10	10	10	10	9	10	10	10	10
New Haven County, Connecticut	12	11	11	11	11	13	11	11	11	12	12	12	11	11	11	8
Norfolk County, Massachusetts	8	8	8	7	6	6	6	5	5	4	5	5	5	5	5	5
Maricopa County, Arizona	7	6	6	8	7	7	7	6	5	4	3	4	4	4	4	5
Monroe County, New York	8	8	8	9	9	10	10	8	9	8	8	8	8	8	7	8
Essex County, Massachusetts	8	8	8	10	9	9	8	8	9	10	10	10	10	11	10	10
Pinellas County, Florida	9	9	9	9	9	9	8	8	8	7	7	7	7	7	9	9
Hudson County, New Jersey	8	8	9	9	9	9	9	10	10	10	10	10	10	10	10	12
Jefferson County, Alabama	7	6	6	6	5	7	7	7	7	9	9	8	8	8	8	9
Henrico County, Virginia	8	8	8	9	8	9	9	10	11	11	11	10	10	10	10	10
Franklin County, Ohio	8	8	7	8	7	7	8	7	9	9	9	9	9	8	8	9

Cumberland County, Maine	7	8	8	10	10	10	10	10	9	9	9	9	9	9	8	10
Nassau County, New York	10	10	10	9	9	9	10	10	11	11	10	8	9	9	9	8
Baltimore County, Maryland	7	7	7	8	8	8	8	8	8	8	8	10	10	10	9	11
Arlington County, Virginia	10	10	9	9	8	8	7	5	5	6	6	6	6	6	6	6
Dane County, Wisconsin	8	8	8	9	9	9	9	9	9	8	8	7	7	9	8	7
Davidson County, Tennessee	7	7	7	9	9	8	8	8	8	8	8	7	7	8	8	8
Duval County, Florida	9	9	9	8	8	7	7	6	6	6	5	4	5	5	5	5
Pulaski County, Arkansas	7	7	7	7	7	7	7	7	7	7	7	7	8	8	8	8
Fairfax County, Virginia	6	7	7	9	9	9	8	8	7	7	7	6	6	6	6	5
Bexar County, Texas	5	5	5	6	6	6	6	6	6	6	7	7	7	7	7	7
Somerset County, New Jersey	6	6	6	6	6	6	6	7	7	7	7	8	8	8	8	8
Union County, New Jersey	6	6	6	9	8	8	8	7	7	7	7	5	5	5	5	6
Salt Lake County, Utah	4	5	6	4	4	4	4	4	4	4	4	4	4	4	4	5
Wake County, North Carolina	5	5	6	6	7	7	7	7	7	8	8	8	8	8	8	8
Butler County, Ohio	8	8	7	9	9	9	9	9	8	7	7	7	8	8	8	7
Miami-Dade County, Florida	6	6	6	6	7	7	6	5	5	5	5	3	3	4	4	8
Lake County, Illinois	6	6	6	6	4	4	4	4	4	3	3	4	4	4	5	7
Arapahoe County, Colorado	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	9
Tulsa County, Oklahoma	4	4	4	5	5	5	5	6	6	8	8	8	8	8	8	8
Essex County, New Jersey	4	4	3	5	3	3	3	3	3	3	4	5	5	5	5	5
Marion County, Indiana	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4
Forsyth County, North Carolina	6	6	6	5	5	5	5	5	5	5	5	5	5	5	5	6
Wayne County, Michigan	5	6	6	6	6	6	5	5	4	4	4	4	5	4	4	5
Alameda County, California	5	5	5	5	6	6	6	6	6	6	6	7	7	7	7	7
Lancaster County, Pennsylvania	6	6	6	5	5	5	5	5	5	5	5	6	6	6	6	6
Ramsey County, Minnesota	7	7	6	8	8	8	7	5	5	5	5	4	5	5	5	4
Charlottesville County, Virginia	4	4	4	6	6	6	6	6	6	6	6	7	7	7	7	7
Ozaukee County, Wisconsin	5	5	5	6	6	6	6	6	5	5	5	5	5	5	5	5
Polk County, Iowa	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Cobb County, Georgia	4	4	4	4	4	4	4	4	5	5	4	5	5	5	5	6
Albany County, New York	5	5	5	6	6	6	5	7	6	6	6	6	6	6	6	5

Collier County, Florida	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	4
Kent County, Michigan	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4
Rockingham County, New Hampshire	5	5	5	6	6	5	4	4	4	4	4	3	3	3	3	4
Erie County, New York	5	5	5	5	5	5	5	5	5	5	5	5	4	4	4	4
Plymouth County, Massachusetts	4	4	4	4	4	4	4	5	7	5	6	5	6	6	6	8
Montgomery County, Ohio	5	5	5	4	4	4	4	5	4	4	4	4	4	4	4	4
Merrimack County, New Hampshire	3	4	5	3	3	4	4	4	4	4	4	4	4	4	4	4
Orange County, Florida	5	5	4	5	5	5	5	6	6	6	6	5	5	4	4	4

County	2011Q1	2011Q2	2011Q3	2011Q4	2012Q1	2012Q2	2012Q3	2012Q4	2013Q1	2013Q2	2013Q3	2013Q4	2014Q1	2014Q2	2014Q3	2014Q4
New York County, New York	635	637	630	649	645	639	636	672	664	663	665	703	698	699	695	762
Suffolk County, Massachusetts	154	153	154	155	156	155	155	162	163	158	157	162	161	163	164	178
Cook County, Illinois	138	138	136	136	134	130	130	135	135	131	132	139	138	141	144	150
Fairfield County, Connecticut	119	120	122	132	133	136	134	128	127	122	123	135	136	135	133	140
San Francisco County, California	103	102	100	103	98	96	96	102	103	99	103	107	106	106	105	118
Los Angeles County, California	96	96	97	100	98	96	97	103	104	103	102	108	105	105	104	108
Harris County, Texas	41	40	41	45	45	45	44	48	48	50	51	51	52	53	52	58
Dallas County, Texas	39	42	44	48	47	48	47	50	52	51	52	58	58	58	56	59
Hennepin County, Minnesota	39	39	40	41	40	41	41	43	41	40	40	48	47	47	46	53
King County, Washington	32	35	34	38	38	40	40	43	44	43	43	42	42	43	44	47
Montgomery County, Pennsylvania	30	30	30	30	29	29	29	32	33	33	33	36	36	36	36	36
San Mateo County, California	27	26	28	29	29	29	29	32	32	31	31	36	36	38	38	47
Westchester County, New York	31	30	30	35	36	36	36	38	38	37	37	34	32	32	32	34
San Diego County, California	26	26	26	29	29	29	28	34	33	34	34	39	39	38	38	42
Fulton County, Georgia	30	30	30	32	32	31	31	31	31	31	30	35	36	37	37	43
Middlesex County, Massachusetts	20	25	25	28	28	29	29	34	32	31	29	33	34	34	34	38
Hamilton County, Ohio	24	24	24	24	24	23	24	24	24	25	26	29	29	29	29	30
Milwaukee County, Wisconsin	24	25	25	26	26	26	26	26	26	27	27	26	26	26	26	25
St. Louis County, Missouri	27	27	26	28	28	28	28	28	28	27	27	31	31	31	31	31
Montgomery County, Maryland	22	22	22	24	24	24	24	24	24	25	25	28	27	27	25	34
Denver County, Colorado	21	22	22	30	28	25	25	26	25	25	25	29	30	31	31	33
Santa Clara County, California	19	19	19	19	20	20	19	21	20	19	19	26	26	26	26	31
Baltimore County, Maryland	20	19	24	19	19	20	19	20	20	19	19	21	20	19	20	18
Cuyahoga County, Ohio	16	15	16	16	16	16	16	17	17	16	17	24	24	25	25	26
Orange County, California	22	22	20	20	22	19	20	21	20	20	20	25	25	25	25	29
Marin County, California	18	18	19	18	18	18	17	20	20	20	18	22	23	23	23	23
Chester County, Pennsylvania	17	15	15	16	15	14	15	17	17	18	17	18	17	16	17	20
Oakland County, Michigan	22	22	23	24	24	24	24	26	26	26	26	28	28	27	27	26
Tarrant County, Texas	20	19	17	22	22	22	22	18	16	14	14	14	14	14	15	16
Allegheny County, Pennsylvania	14	14	14	14	14	14	16	18	18	18	18	22	22	22	22	23

New Castle County, Delaware	19	19	20	20	20	20	20	27	27	26	26	23	21	23	19	20
Philadelphia County, Pennsylvania	13	14	15	18	18	18	18	17	17	18	18	20	20	21	21	21
Mecklenburg County, North Carolina	17	17	17	17	16	16	16	18	16	16	16	18	18	18	18	20
Hartford County, Connecticut	12	12	13	12	11	11	11	11	11	11	13	15	16	16	16	16
DuPage County, Illinois	16	18	18	17	17	17	17	20	20	19	20	23	23	23	23	24
Delaware County, Pennsylvania	11	11	11	13	13	13	14	17	18	19	19	21	21	21	20	21
Shelby County, Tennessee	17	17	17	17	17	18	17	17	16	16	15	18	18	19	18	19
Travis County, Texas	15	15	17	16	16	18	18	18	17	18	18	21	20	20	20	20
Morris County, New Jersey	12	12	13	13	13	12	12	11	11	11	12	15	17	17	17	17
District of Columbia	15	13	13	13	13	11	11	11	11	9	7	8	8	7	7	8
Bergen County, New Jersey	13	12	12	13	12	12	12	11	10	12	12	13	14	14	15	16
Douglas County, Nebraska	11	11	11	13	13	13	12	11	11	12	13	15	15	15	15	18
Johnson County, Kansas	13	13	13	17	16	17	17	17	17	17	18	20	20	20	20	23
Palm Beach County, Florida	14	14	15	16	15	15	17	20	21	19	19	19	18	18	16	16
Multnomah County, Oregon	10	10	10	11	11	11	11	12	11	12	12	12	13	13	13	16
Contra Costa County, California	10	10	9	10	11	12	14	13	13	14	15	16	17	16	16	18
Jefferson County, Kentucky	10	10	10	11	9	9	10	11	10	11	11	11	10	10	10	11
Richmond County, Virginia	8	8	9	9	8	7	8	7	7	8	8	11	12	12	12	14
Mercer County, New Jersey	14	14	14	12	13	13	13	13	14	15	15	16	16	14	14	14
Providence County, Rhode Island	9	9	9	10	10	10	10	11	11	11	11	14	14	14	14	14
Jackson County, Missouri	10	9	9	9	9	8	8	8	8	8	8	9	9	10	10	10
New Haven County, Connecticut	7	7	6	7	7	7	6	6	6	7	6	6	6	6	6	7
Norfolk County, Massachusetts	5	5	6	7	7	7	7	9	9	8	9	11	11	11	11	13
Maricopa County, Arizona	5	6	5	7	7	7	7	7	8	9	9	13	11	11	11	15
Monroe County, New York	8	8	8	8	8	8	8	8	9	10	10	10	9	9	10	11
Essex County, Massachusetts	10	10	10	11	11	11	11	11	11	9	9	9	9	9	9	11
Pinellas County, Florida	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
Hudson County, New Jersey	12	12	12	11	11	11	11	11	11	10	11	12	12	11	10	9
Jefferson County, Alabama	9	8	10	10	10	10	10	12	12	10	10	12	12	12	12	11
Henrico County, Virginia	10	10	10	11	11	11	10	12	12	12	11	12	12	12	12	10
Franklin County, Ohio	9	9	9	9	9	9	9	9	9	9	8	8	7	7	7	9

Cumberland County, Maine	10	11	11	11	12	10	10	10	10	10	10	9	9	9	9	9
Nassau County, New York	9	8	8	9	10	10	10	11	11	11	11	11	11	11	11	11
Baltimore County, Maryland	11	11	11	11	11	10	10	10	9	10	10	12	13	13	11	12
Arlington County, Virginia	6	5	6	7	7	6	6	8	7	7	7	8	7	7	7	7
Dane County, Wisconsin	7	7	7	6	6	6	6	7	7	8	9	8	7	7	7	9
Davidson County, Tennessee	8	7	7	8	7	8	7	8	7	9	9	10	10	10	9	10
Duval County, Florida	5	5	5	7	7	7	7	7	7	7	7	8	9	9	9	9
Pulaski County, Arkansas	8	8	8	8	7	8	8	9	9	9	9	11	11	11	11	11
Fairfax County, Virginia	5	5	5	5	5	5	5	6	8	8	8	11	11	11	12	14
Bexar County, Texas	6	6	6	8	7	7	7	8	8	8	9	12	11	11	11	11
Somerset County, New Jersey	8	8	9	10	10	10	10	12	11	11	11	11	10	10	10	9
Union County, New Jersey	6	6	6	6	7	8	8	10	10	10	10	12	12	12	12	12
Salt Lake County, Utah	5	5	5	5	5	6	6	8	8	9	9	9	8	8	8	8
Wake County, North Carolina	9	9	9	10	10	9	9	8	8	8	8	8	8	8	8	8
Butler County, Ohio	7	7	8	8	8	8	8	8	8	7	7	7	7	7	7	7
Miami-Dade County, Florida	8	8	8	10	10	10	10	10	10	10	10	9	9	9	9	10
Lake County, Illinois	7	6	8	8	7	7	7	9	9	8	8	8	8	8	8	11
Arapahoe County, Colorado	9	8	8	4	4	4	4	4	4	5	4	4	4	3	3	3
Tulsa County, Oklahoma	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	9
Essex County, New Jersey	6	6	6	5	5	5	5	6	5	5	5	5	5	6	6	9
Marion County, Indiana	4	4	4	4	5	5	4	4	4	4	4	6	6	6	6	8
Forsyth County, North Carolina	6	6	6	6	6	6	6	6	6	6	6	5	5	5	5	6
Wayne County, Michigan	5	5	3	4	4	3	4	4	4	4	4	7	7	7	7	7
Alameda County, California	5	6	6	6	6	6	6	5	5	5	5	7	7	8	8	9
Lancaster County, Pennsylvania	6	6	6	6	6	6	6	6	6	6	6	7	8	8	8	8
Ramsey County, Minnesota	4	4	4	5	5	5	5	5	4	4	4	4	4	4	4	4
Charlottesville County, Virginia	7	7	6	5	4	4	4	5	5	5	5	7	7	7	7	7
Ozaukee County, Wisconsin	5	5	5	5	5	5	6	7	7	6	6	5	5	5	5	5
Polk County, Iowa	5	5	5	5	6	6	6	6	6	7	7	7	7	7	6	6
Cobb County, Georgia	5	5	5	5	5	5	5	6	5	7	7	6	6	7	6	7
Albany County, New York	5	5	5	5	5	5	5	6	7	6	6	6	6	6	6	7

Collier County, Florida	4	4	4	7	7	7	7	9	7	8	8	9	8	8	8	10
Kent County, Michigan	4	4	4	5	5	5	5	5	6	6	6	7	7	7	7	7
Rockingham County, New Hampshire	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
Erie County, New York	4	5	5	5	4	4	4	4	4	4	4	4	4	4	4	5
Plymouth County, Massachusetts	8	8	8	8	7	8	7	8	5	5	5	5	6	6	6	6
Montgomery County, Ohio	4	4	4	3	3	3	3	4	3	4	4	6	6	6	6	6
Merrimack County, New Hampshire	4	4	4	4	5	5	5	5	5	5	5	5	6	5	5	6
Orange County, Florida	5	5	4	3	3	3	3	3	3	3	3	5	5	6	6	6

County	2015Q1	2015Q2	2015Q3	2015Q4	2016Q1	20126Q2	2016Q3	2016Q4	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1
New York County, New York	760	757	752	783	760	755	762	782	764	755	751	783	769
Suffolk County, Massachusetts	178	179	179	188	186	182	186	188	184	184	183	190	186
Cook County, Illinois	151	151	152	155	152	157	156	169	169	167	165	171	172
Fairfield County, Connecticut	141	140	143	152	149	148	148	147	145	142	142	145	141
San Francisco County, California	114	116	114	119	115	117	114	131	129	128	127	136	138
Los Angeles County, California	109	108	109	117	117	117	117	121	119	117	117	122	119
Harris County, Texas	58	57	56	57	57	58	60	63	63	65	64	66	61
Dallas County, Texas	59	60	57	58	56	56	57	63	65	67	63	69	68
Hennepin County, Minnesota	53	53	53	53	54	52	51	53	53	54	53	52	55
King County, Washington	48	48	48	55	53	54	52	54	52	50	50	52	52
Montgomery County, Pennsylvania	35	36	35	40	35	36	36	39	39	40	38	40	42
San Mateo County, California	44	44	43	45	43	44	43	42	43	42	41	48	52
Westchester County, New York	34	34	32	32	32	31	30	32	31	31	31	37	37
San Diego County, California	43	43	42	46	43	42	42	44	42	41	41	50	49
Fulton County, Georgia	44	44	44	45	45	46	45	48	48	48	47	53	53
Middlesex County, Massachusetts	36	35	32	35	34	35	32	37	34	35	35	41	42
Hamilton County, Ohio	30	29	29	29	30	29	29	29	29	29	29	29	29
Milwaukee County, Wisconsin	26	26	26	27	27	26	26	24	24	24	25	25	25
St. Louis County, Missouri	31	29	30	33	32	32	32	33	33	35	35	35	35
Montgomery County, Maryland	34	35	35	34	34	34	34	34	33	33	33	37	36
Denver County, Colorado	33	33	32	31	28	28	27	28	29	29	30	34	34
Santa Clara County, California	29	28	29	30	33	32	32	29	30	30	30	36	31
Baltimore County, Maryland	18	18	18	19	18	18	18	20	20	19	19	19	17
Cuyahoga County, Ohio	26	26	26	26	26	26	27	24	23	23	23	24	23
Orange County, California	28	28	28	28	27	28	29	31	31	32	32	39	38
Marin County, California	24	24	23	24	25	25	26	26	27	27	27	33	33
Chester County, Pennsylvania	23	23	23	23	23	23	23	25	24	23	22	21	21
Oakland County, Michigan	28	28	28	28	29	28	28	26	25	25	25	25	25
Tarrant County, Texas	16	16	15	15	15	15	15	17	19	18	18	22	22
Allegheny County, Pennsylvania	24	24	24	25	24	23	23	24	24	24	23	29	28

New Castle County, Delaware	20	20	20	20	19	20	19	19	18	17	17	18	18
Philadelphia County, Pennsylvania	20	19	19	17	17	17	17	21	21	21	21	23	22
Mecklenburg County, North Carolina	21	21	20	20	20	20	20	21	21	22	22	24	22
Hartford County, Connecticut	15	15	15	16	16	16	16	21	21	21	21	20	21
DuPage County, Illinois	25	25	25	24	24	24	24	24	23	22	21	24	24
Delaware County, Pennsylvania	22	22	22	21	21	21	21	24	24	24	25	24	26
Shelby County, Tennessee	20	20	20	20	21	19	20	19	19	19	19	21	19
Travis County, Texas	21	21	21	23	23	23	23	24	24	24	24	26	27
Morris County, New Jersey	18	18	18	17	16	16	16	20	20	19	18	20	20
District of Columbia	7	7	7	8	8	8	9	10	9	9	9	11	10
Bergen County, New Jersey	16	16	16	16	17	17	17	16	17	18	18	20	20
Douglas County, Nebraska	19	19	17	19	18	19	19	18	19	19	19	22	21
Johnson County, Kansas	24	23	24	23	22	22	23	21	23	24	24	30	29
Palm Beach County, Florida	17	17	17	17	17	18	19	20	20	20	18	20	20
Multnomah County, Oregon	16	15	16	16	16	16	15	15	15	15	15	16	16
Contra Costa County, California	19	19	19	21	19	19	19	21	22	22	22	25	25
Jefferson County, Kentucky	10	10	11	12	11	11	12	13	13	13	13	13	13
Richmond County, Virginia	14	14	15	16	16	17	17	18	18	18	17	19	19
Mercer County, New Jersey	14	13	13	15	15	16	16	14	12	12	11	12	13
Providence County, Rhode Island	14	14	14	15	15	14	15	16	16	16	16	15	15
Jackson County, Missouri	10	11	11	11	11	11	10	11	10	10	10	10	11
New Haven County, Connecticut	6	7	7	7	7	7	7	8	9	9	9	9	10
Norfolk County, Massachusetts	13	13	12	14	14	14	12	10	9	10	9	14	13
Maricopa County, Arizona	14	14	15	18	18	18	18	21	22	22	23	26	27
Monroe County, New York	11	11	11	13	13	13	13	13	13	13	13	14	14
Essex County, Massachusetts	11	11	11	11	12	12	11	11	11	13	13	15	14
Pinellas County, Florida	9	9	9	9	9	10	10	11	12	12	13	16	15
Hudson County, New Jersey	9	10	10	11	11	11	10	10	11	11	10	10	9
Jefferson County, Alabama	11	10	10	10	10	10	10	12	12	12	11	14	13
Henrico County, Virginia	10	10	9	11	11	11	11	11	11	11	11	12	12
Franklin County, Ohio	9	9	9	8	9	8	8	9	9	9	9	14	14

Cumberland County, Maine	9	9	9	9	9	9	9	9	9	8	9	12	12
Nassau County, New York	11	10	11	10	10	10	10	9	9	9	9	11	11
Baltimore County, Maryland	12	12	12	12	13	12	12	13	14	14	13	13	13
Arlington County, Virginia	7	7	7	8	8	8	8	8	8	8	8	8	8
Dane County, Wisconsin	8	8	8	8	8	8	9	9	9	9	9	10	10
Davidson County, Tennessee	10	9	9	9	9	9	10	10	11	11	11	12	11
Duval County, Florida	9	9	9	9	9	9	9	11	11	11	11	12	12
Pulaski County, Arkansas	11	11	11	11	11	11	11	11	11	11	11	12	12
Fairfax County, Virginia	13	13	13	12	12	13	13	14	14	14	14	17	17
Bexar County, Texas	11	11	11	12	12	12	12	13	13	13	13	14	14
Somerset County, New Jersey	9	8	8	8	7	7	7	8	9	8	8	9	9
Union County, New Jersey	12	13	12	12	13	12	13	15	14	15	15	16	16
Salt Lake County, Utah	8	8	8	9	9	9	9	10	10	11	11	13	13
Wake County, North Carolina	8	8	8	9	9	9	10	11	11	11	10	14	14
Butler County, Ohio	7	7	7	7	7	7	7	7	7	7	7	7	7
Miami-Dade County, Florida	10	10	10	13	12	12	12	19	19	19	21	23	24
Lake County, Illinois	11	11	11	13	13	13	13	13	14	15	13	14	15
Arapahoe County, Colorado	4	4	5	6	6	7	7	8	8	8	8	9	9
Tulsa County, Oklahoma	9	9	9	10	10	10	10	10	10	10	10	11	11
Essex County, New Jersey	8	8	9	8	8	8	8	10	10	11	11	12	11
Marion County, Indiana	8	8	8	8	8	8	10	9	9	9	10	12	12
Forsyth County, North Carolina	6	6	6	6	6	6	6	6	6	6	6	6	6
Wayne County, Michigan	8	8	8	10	10	10	10	10	9	10	10	10	10
Alameda County, California	9	9	9	9	9	9	9	9	9	8	8	10	11
Lancaster County, Pennsylvania	9	8	7	7	7	7	7	7	7	7	7	7	7
Ramsey County, Minnesota	4	4	5	5	5	5	5	5	5	5	5	6	6
Charlottesville County, Virginia	7	7	7	8	8	8	8	8	7	7	8	8	8
Ozaukee County, Wisconsin	5	6	6	6	6	6	6	6	6	5	5	6	6
Polk County, Iowa	6	6	6	6	6	6	6	7	7	7	7	8	8
Cobb County, Georgia	6	6	6	7	7	7	7	8	8	8	8	10	10
Albany County, New York	7	7	8	7	7	7	7	7	7	7	7	7	7

Collier County, Florida	10	10	10	11	11	11	11	11	11	11	11	13	13
Kent County, Michigan	6	6	7	7	7	7	6	6	6	6	6	6	5
Rockingham County, New Hampshire	6	6	6	6	6	6	5	5	5	5	6	7	7
Erie County, New York	5	5	5	5	5	5	5	5	5	5	4	7	7
Plymouth County, Massachusetts	7	7	7	7	7	7	8	8	8	8	8	8	8
Montgomery County, Ohio	6	6	6	6	6	5	5	5	5	5	5	5	5
Merrimack County, New Hampshire	6	6	6	6	5	6	6	6	6	6	6	6	5
Orange County, Florida	7	5	4	5	5	5	5	5	5	5	5	5	6

Table A.6

Statistic	N	Mean	St. Dev.	Min	Pctl(25)
2013-06-30	37	361,724,599,086.000	1,895,074,090,837.000	5,262,489.000	508,346,100.000
2013-09-30	36	471,786,382,800.000	2,469,232,987,050.000	6,938,906.000	666,285,358.000
2013-12-31	38	387,872,845,153.000	2,048,164,709,854.000	6,410,356.000	506,257,186.000
2014-03-31	36	393,452,554,491.000	1,999,895,653,327.000	6,003,780.000	715,848,108.000
2014-06-30	36	409,884,028,525.000	2,111,515,536,315.000	6,517,894.000	836,577,738.000
2014-09-30	36	416,210,625,582.000	2,105,702,195,361.000	0.000	756,239,791.000
2014-12-31	36	429,802,525,820.000	2,159,087,137,130.000	3,370,399.000	1,284,989,917.000
2015-03-31	35	441,553,408,839.000	2,199,717,803,638.000	4,350,822.000	1,258,095,020.000
2015-06-30	36	458,439,328,282.000	2,332,537,222,491.000	2,050,956.000	1,479,575,955.000
2015-09-30	35	424,452,295,009.000	2,093,120,903,709.000	2,202,779.000	1,503,822,986.000
2015-12-31	38	409,865,192,159.000	2,115,137,423,473.000	6,082,505.000	930,208,495.000
2016-03-31	37	568,498,343,229.000	2,303,107,049,805.000	13,506,315.000	1,421,005,233.000
2016-06-30	37	466,721,844,528.000	2,353,517,694,736.000	21,398,550.000	1,348,027,680.000
2016-09-30	37	497,999,695,980.000	2,530,238,141,435.000	21,490,771.000	1,382,778,565.000
2016-12-31	41	461,389,130,696.000	2,473,701,776,177.000	27,399,132.000	652,373,977.000
2017-03-31	42	473,690,668,475.000	2,628,482,894,763.000	21,584,298.000	764,742,336.000
2017-06-30	41	464,779,342,875.000	2,470,420,098,231.000	5,422,220.000	1,292,703,473.000
2017-09-30	40	499,719,108,823.000	2,614,704,069,783.000	5,401,694.000	1,248,087,476.000
2017-12-31	37	626,497,690,160.000	3,193,364,218,034.000	12,389,701.000	2,480,364,150.000
2018-03-31	38	609,496,322,391.000	3,198,970,993,146.000	14,660,470.000	1,952,852,726.000
2018-06-30	39	614,266,799,706.000	3,256,499,600,178.000	20,797,646.000	1,609,488,442.000
2018-09-30	38	638,507,176,794.000	3,114,353,951,398.000	5,130,178.000	1,692,093,317.000
2018-12-31	38	603,993,931,418.000	2,943,453,939,244.000	2,866,856.000	1,877,447,283.000

Appendix B

Standard Ellipse for Top 5 CBSAs

Table B.1: Boston-Cambridge-Newton, MA-NH Metro Area

Date	Centre.x	Centre.y	Sigma.x	Sigma.y	Theta	Eccentricity	Area.sde	TanTheta	SinTheta
1999-03-31	2017284.03	2418332.10	9203.99	21864.11	177.20	0.91	632.20	-0.05	0.05
1999-06-30	2017338.74	2418075.13	8604.81	20909.38	177.43	0.91	565.24	-0.04	0.04
1999-09-30	2017440.34	2418162.84	8724.14	21398.07	177.33	0.91	586.47	-0.05	0.05
1999-12-31	2017339.10	2419136.75	9645.27	23714.33	176.49	0.91	718.58	-0.06	0.06
2000-03-31	2017464.67	2419331.66	9412.29	23544.74	175.47	0.92	696.21	-0.08	0.08
2000-06-30	2016882.18	2420073.45	7962.14	22323.71	2.67	0.93	558.40	0.05	0.05
2000-09-30	2017671.02	2419302.14	10630.91	26389.12	173.85	0.92	881.34	-0.11	0.11
2000-12-31	2017340.35	2421010.94	9760.17	27855.89	179.18	0.94	854.13	-0.01	0.01
2001-03-31	2017216.10	2419096.32	9863.02	22802.96	177.82	0.90	706.56	-0.04	0.04
2001-06-30	2017433.45	2419236.30	9817.06	22674.11	178.41	0.90	699.30	-0.03	0.03
2001-09-30	2017420.37	2419346.16	10106.31	22862.42	179.09	0.90	725.88	-0.02	0.02
2001-12-31	2017136.91	2419831.46	9375.40	24103.73	1.47	0.92	709.94	0.03	0.03
2002-03-31	2017282.68	2419722.81	9737.25	23950.33	0.58	0.91	732.65	0.01	0.01
2002-06-30	2017481.43	2419825.01	9265.43	23902.16	179.39	0.92	695.75	-0.01	0.01
2002-09-30	2017204.20	2419623.86	9798.96	24126.13	0.43	0.91	742.71	0.01	0.01
2002-12-31	2017154.22	2419307.58	9822.89	23704.85	1.57	0.91	731.52	0.03	0.03
2003-03-31	2017307.55	2419159.94	10083.29	23580.02	0.86	0.90	746.96	0.01	0.01
2003-06-30	2017168.18	2419284.56	9762.77	23556.96	1.56	0.91	722.51	0.03	0.03
2003-09-30	2017125.99	2419061.52	9558.14	21797.97	3.12	0.90	654.54	0.05	0.05
2003-12-31	2017297.17	2419418.82	9812.25	23473.34	2.26	0.91	723.59	0.04	0.04
2004-03-31	2017260.17	2419380.99	9939.07	23493.41	2.39	0.91	733.57	0.04	0.04
2004-06-30	2017248.36	2419507.24	10051.36	23858.74	2.27	0.91	753.39	0.04	0.04
2004-09-30	2017230.11	2419386.92	9827.34	23360.57	2.00	0.91	721.22	0.03	0.03
2004-12-31	2016958.32	2419338.08	10021.83	22812.13	1.48	0.90	718.23	0.03	0.03
2005-03-31	2016927.61	2419397.36	10142.36	23141.39	1.49	0.90	737.36	0.03	0.03
2005-06-30	2016910.01	2419414.80	10171.40	23207.57	1.50	0.90	741.58	0.03	0.03
2005-09-30	2016968.29	2419379.48	10116.56	23057.83	1.83	0.90	732.83	0.03	0.03
2005-12-31	2016869.60	2418538.75	9777.40	20508.90	1.58	0.88	629.96	0.03	0.03

Table B.1: Boston-Cambridge-Newton, MA-NH Metro Area

2006-03-31	2016772.31	2418980.38	9990.21	22189.63	1.77	0.89	696.43	0.03	0.03
2006-06-30	2017018.44	2419512.84	10163.34	23810.62	1.45	0.90	760.25	0.03	0.03
2006-09-30	2017034.22	2419067.24	10197.99	22349.95	1.29	0.89	716.05	0.02	0.02
2006-12-31	2017429.84	2419102.92	10246.04	21658.06	0.79	0.88	697.15	0.01	0.01
2007-03-31	2017484.21	2419174.89	10285.07	21789.00	0.61	0.88	704.04	0.01	0.01
2007-06-30	2017434.57	2419143.62	10348.84	21874.70	0.78	0.88	711.19	0.01	0.01
2007-09-30	2017444.96	2419129.62	10297.46	21767.21	0.77	0.88	704.18	0.01	0.01
2007-12-31	2017232.33	2419460.47	10253.49	22128.34	0.42	0.89	712.80	0.01	0.01
2008-03-31	2017108.19	2419452.06	10697.74	22319.79	0.34	0.88	750.12	0.01	0.01
2008-06-30	2017304.50	2419081.89	10234.82	20799.45	179.63	0.87	668.78	-0.01	0.01
2008-09-30	2017184.26	2418585.76	10455.96	19404.29	0.48	0.84	637.40	0.01	0.01
2008-12-31	2017497.67	2418613.16	11433.90	19909.64	175.71	0.82	715.17	-0.08	0.07
2009-03-31	2018099.75	2418706.57	11591.70	20615.97	174.80	0.83	750.76	-0.09	0.09
2009-06-30	2017958.95	2419083.49	11291.14	20394.37	177.40	0.83	723.43	-0.05	0.05
2009-09-30	2018016.74	2418938.17	11512.29	20415.68	176.78	0.83	738.37	-0.06	0.06
2009-12-31	2017668.44	2418608.09	11565.10	18778.34	176.52	0.79	682.27	-0.06	0.06
2010-03-31	2017816.62	2418553.83	11949.13	18895.55	175.26	0.77	709.33	-0.08	0.08
2010-06-30	2017912.14	2418715.10	12181.40	19140.48	176.87	0.77	732.49	-0.05	0.05
2010-09-30	2017813.05	2418587.68	12066.10	19078.30	175.22	0.77	723.20	-0.08	0.08
2010-12-31	2017754.06	2418326.26	12815.26	20218.24	166.01	0.77	813.99	-0.25	0.24
2011-03-31	2018074.17	2418390.78	12638.56	20751.67	166.61	0.79	823.95	-0.24	0.23
2011-06-30	2017829.94	2418379.13	12903.91	20556.09	166.17	0.78	833.32	-0.25	0.24
2011-09-30	2017776.51	2418283.41	12895.20	20453.79	166.91	0.78	828.61	-0.23	0.23
2011-12-31	2017474.17	2418396.13	13056.44	20536.18	165.96	0.77	842.35	-0.25	0.24
2012-03-31	2017463.08	2418441.58	12579.16	20432.07	167.41	0.79	807.45	-0.22	0.22
2012-06-30	2017516.88	2418377.33	12919.50	20484.84	166.14	0.78	831.43	-0.25	0.24
2012-09-30	2017383.49	2418450.70	12694.28	20435.06	167.28	0.78	814.96	-0.23	0.22
2012-12-31	2017076.84	2418242.34	13124.20	20086.21	167.15	0.76	828.17	-0.23	0.22
2013-03-31	2016605.73	2418547.33	12030.90	19670.72	173.57	0.79	743.48	-0.11	0.11
2013-06-30	2016519.77	2418389.71	11763.01	19825.87	172.54	0.80	732.66	-0.13	0.13

Table B.1: Boston-Cambridge-Newton, MA-NH Metro Area

2013-09-30	2016522.01	2418397.91	11791.45	19920.57	172.45	0.81	737.94	-0.13	0.13
2013-12-31	2016485.51	2418400.95	11649.20	19914.22	170.37	0.81	728.80	-0.17	0.17
2014-03-31	2016554.11	2418344.01	12023.19	19954.94	169.19	0.80	753.74	-0.19	0.19
2014-06-30	2016583.06	2418324.34	11944.61	19876.37	169.22	0.80	745.86	-0.19	0.19
2014-09-30	2016605.71	2418268.21	11906.73	19810.16	169.56	0.80	741.02	-0.18	0.18
2014-12-31	2016754.43	2418527.96	11544.07	20683.14	172.04	0.83	750.11	-0.14	0.14
2015-03-31	2016975.73	2418484.99	11739.78	20790.75	171.26	0.83	766.80	-0.15	0.15
2015-06-30	2017005.45	2418474.42	11728.58	20790.91	171.26	0.83	766.07	-0.15	0.15
2015-09-30	2017215.11	2418559.45	11557.69	20878.66	171.18	0.83	758.09	-0.16	0.15
2015-12-31	2016991.62	2418222.18	11772.66	21267.59	168.97	0.83	786.58	-0.19	0.19
2016-03-31	2017053.00	2418341.28	11894.29	21415.41	169.16	0.83	800.23	-0.19	0.19
2016-06-30	2016992.61	2418380.90	12041.08	21541.36	169.09	0.83	814.87	-0.19	0.19
2016-09-30	2017195.22	2417993.38	11952.27	20319.40	166.16	0.81	762.98	-0.25	0.24
2016-12-31	2017099.68	2418136.92	11795.47	19934.61	163.76	0.81	738.71	-0.29	0.28
2017-03-31	2017311.89	2418230.82	11548.23	20227.18	163.94	0.82	733.84	-0.29	0.28
2017-06-30	2016971.03	2418381.48	11995.83	20756.13	164.56	0.82	782.22	-0.28	0.27
2017-09-30	2016898.73	2418720.91	11824.28	21789.07	165.85	0.84	809.40	-0.25	0.24
2017-12-31	2016577.02	2418878.66	12428.23	22453.41	170.24	0.83	876.68	-0.17	0.17
2018-03-31	2016510.18	2418814.92	12160.65	22503.44	168.98	0.84	859.72	-0.19	0.19
2018-06-30	2016632.09	2419094.40	12720.97	23701.48	171.05	0.84	947.21	-0.16	0.16
2018-09-30	2016582.38	2418822.21	12660.46	22713.01	169.35	0.83	903.39	-0.19	0.18
2018-12-31	2016634.02	2419482.82	12586.75	24936.52	168.33	0.86	986.05	-0.21	0.20

Table B.2: Chicago-Naperville-Elgin, IL-IN-WI Metro Area

Date	Centre.x	Centre.y	Sigma.x	Sigma.y	Theta	Eccentricity	Area.sde	TanTheta	SinTheta
1999-03-31	683188.54	2129480.69	13489.79	17695.75	102.44	0.65	749.94	-4.53	0.98
1999-06-30	682814.48	2129776.21	13576.48	18920.19	104.38	0.70	806.98	-3.90	0.97
1999-09-30	682886.67	2129763.45	13495.40	18831.73	104.36	0.70	798.41	-3.91	0.97
1999-12-31	682490.54	2129396.83	13587.71	19121.54	97.27	0.70	816.24	-7.84	0.99
2000-03-31	682684.61	2129565.96	13571.80	18968.73	100.62	0.70	808.77	-5.34	0.98
2000-06-30	682886.70	2129510.34	13344.16	18719.97	100.69	0.70	784.78	-5.30	0.98
2000-09-30	682935.17	2129040.79	13775.27	18974.25	101.43	0.69	821.14	-4.95	0.98
2000-12-31	682557.20	2129371.44	13512.88	19039.08	97.34	0.70	808.25	-7.76	0.99
2001-03-31	683822.99	2128978.54	13476.78	17317.45	102.75	0.63	733.20	-4.42	0.98
2001-06-30	684111.49	2128556.58	12299.94	16880.96	95.31	0.68	652.30	-10.77	1.00
2001-09-30	684654.74	2128105.55	12493.83	16404.81	86.06	0.65	643.90	14.53	1.00
2001-12-31	683260.59	2130021.24	13354.17	18079.06	117.76	0.67	758.48	-1.90	0.88
2002-03-31	683686.10	2129302.64	13428.88	17530.06	108.58	0.64	739.56	-2.97	0.95
2002-06-30	684080.08	2128880.46	13194.75	16731.20	102.50	0.61	693.55	-4.51	0.98
2002-09-30	683630.77	2129565.70	12852.20	17355.18	104.58	0.67	700.74	-3.84	0.97
2002-12-31	683206.37	2129094.91	13468.06	18393.75	99.03	0.68	778.26	-6.29	0.99
2003-03-31	683348.98	2128734.81	12548.84	18280.28	94.31	0.73	720.67	-13.26	1.00
2003-06-30	683400.41	2128731.27	12447.96	18201.26	94.28	0.73	711.79	-13.35	1.00
2003-09-30	683255.33	2129094.89	13380.22	18313.77	99.00	0.68	769.82	-6.31	0.99
2003-12-31	682854.91	2129065.54	13234.74	18525.38	98.37	0.70	770.25	-6.80	0.99
2004-03-31	682762.38	2129076.25	13346.82	18653.10	98.38	0.70	782.13	-6.79	0.99
2004-06-30	682315.34	2129425.90	14241.89	19138.67	101.02	0.67	856.31	-5.13	0.98
2004-09-30	682657.40	2129274.23	13751.38	18925.60	98.61	0.69	817.61	-6.60	0.99
2004-12-31	681111.06	2129932.84	13829.28	20731.90	99.14	0.75	900.72	-6.22	0.99
2005-03-31	681238.79	2129914.58	13891.82	20634.21	98.90	0.74	900.53	-6.39	0.99
2005-06-30	681087.55	2130107.36	13872.97	20931.92	100.13	0.75	912.28	-5.59	0.98
2005-09-30	681057.60	2129820.46	14002.14	20705.10	95.79	0.74	910.80	-9.86	0.99
2005-12-31	681192.57	2129987.90	13786.96	20244.39	94.91	0.73	876.85	-11.63	1.00
2006-03-31	681538.25	2129875.71	13936.85	19627.24	90.79	0.70	859.36	-72.56	1.00

Table B.2: Chicago-Naperville-Elgin, IL-IN-WI Metro Area

2006-06-30	681432.28	2129861.54	13886.75	20337.69	92.42	0.73	887.26	-23.68	1.00
2006-09-30	681267.75	2129905.77	14042.77	20511.71	92.27	0.73	904.91	-25.18	1.00
2006-12-31	681418.45	2129876.70	13973.37	19999.08	92.15	0.72	877.93	-26.67	1.00
2007-03-31	681253.30	2129935.52	14118.99	20152.45	91.98	0.71	893.88	-28.88	1.00
2007-06-30	681500.35	2129778.62	13809.13	20004.34	91.70	0.72	867.84	-33.67	1.00
2007-09-30	681064.48	2130055.11	14255.16	20173.85	91.58	0.71	903.46	-36.26	1.00
2007-12-31	681849.27	2129616.46	14419.05	19216.55	95.73	0.66	870.49	-9.96	0.99
2008-03-31	682430.57	2129287.05	13559.91	18654.56	92.99	0.69	794.68	-19.12	1.00
2008-06-30	682304.44	2129402.76	13781.78	18730.68	93.01	0.68	810.98	-19.00	1.00
2008-09-30	681961.45	2129306.01	13809.47	19177.49	92.61	0.69	831.99	-21.96	1.00
2008-12-31	682698.29	2129264.78	13207.74	18206.97	95.57	0.69	755.47	-10.25	1.00
2009-03-31	682402.55	2129122.69	13274.54	18748.82	92.77	0.71	781.89	-20.66	1.00
2009-06-30	682549.16	2128978.39	12588.54	18556.01	91.75	0.73	733.85	-32.77	1.00
2009-09-30	682533.98	2128846.62	12671.22	18707.72	91.04	0.74	744.71	-55.04	1.00
2009-12-31	682872.79	2129396.43	13133.42	18286.46	103.33	0.70	754.50	-4.22	0.97
2010-03-31	682519.93	2129372.55	13392.23	18725.87	101.16	0.70	787.85	-5.07	0.98
2010-06-30	682737.66	2129410.16	13402.73	18484.00	101.55	0.69	778.29	-4.89	0.98
2010-09-30	682406.12	2129551.39	13693.38	18820.09	101.31	0.69	809.62	-5.00	0.98
2010-12-31	681954.38	2129992.86	14560.49	19050.35	103.57	0.64	871.42	-4.14	0.97
2011-03-31	682215.96	2129837.28	14386.13	18774.78	100.67	0.64	848.53	-5.31	0.98
2011-06-30	681741.92	2129828.89	14109.25	19808.09	101.86	0.70	878.00	-4.76	0.98
2011-09-30	681446.68	2130273.66	14879.03	20216.66	107.84	0.68	945.00	-3.11	0.95
2011-12-31	681555.33	2130352.81	14799.91	20157.25	107.55	0.68	937.22	-3.16	0.95
2012-03-31	681609.12	2130253.47	14855.22	20090.75	105.86	0.67	937.62	-3.52	0.96
2012-06-30	681558.02	2130162.24	14939.34	20169.99	104.70	0.67	946.64	-3.81	0.97
2012-09-30	681560.60	2130166.32	14939.23	20171.04	104.68	0.67	946.69	-3.82	0.97
2012-12-31	681157.69	2130463.63	15791.37	20320.87	106.18	0.63	1008.12	-3.45	0.96
2013-03-31	681281.23	2130419.03	15787.06	20284.18	105.85	0.63	1006.02	-3.52	0.96
2013-06-30	681292.63	2130531.13	15801.76	20274.39	106.10	0.63	1006.47	-3.46	0.96
2013-09-30	681162.05	2130600.41	15938.81	20211.81	104.89	0.61	1012.07	-3.76	0.97

Table B.2: Chicago-Naperville-Elgin, IL-IN-WI Metro Area

2013-12-31	680255.84	2131002.20	16537.25	20843.95	104.70	0.61	1082.91	-3.81	0.97
2014-03-31	680209.76	2131015.86	16585.20	20884.89	104.69	0.61	1088.18	-3.81	0.97
2014-06-30	680339.72	2130960.92	16443.82	20760.21	104.81	0.61	1072.47	-3.78	0.97
2014-09-30	680295.63	2131206.98	16590.95	20722.78	107.28	0.60	1080.11	-3.22	0.95
2014-12-31	680099.41	2131408.71	16630.00	21204.36	113.27	0.62	1107.81	-2.33	0.92
2015-03-31	679853.01	2131347.45	17146.68	21589.73	112.33	0.61	1162.99	-2.44	0.93
2015-06-30	679846.64	2131281.24	17076.52	21566.66	111.55	0.61	1157.00	-2.53	0.93
2015-09-30	680188.59	2131019.27	16886.07	21266.52	109.69	0.61	1128.17	-2.79	0.94
2015-12-31	680293.78	2131179.78	17213.37	22110.91	120.02	0.63	1195.70	-1.73	0.87
2016-03-31	680158.74	2131210.43	17339.85	22229.60	120.23	0.63	1210.95	-1.72	0.86
2016-06-30	680043.39	2131354.20	17133.14	22534.86	118.00	0.65	1212.95	-1.88	0.88
2016-09-30	680094.32	2131406.81	17197.11	23752.42	118.65	0.69	1283.26	-1.83	0.88
2016-12-31	680123.61	2131440.60	17208.01	23597.73	116.43	0.68	1275.71	-2.01	0.90
2017-03-31	679965.15	2131669.78	17287.67	23504.89	118.49	0.68	1276.57	-1.84	0.88
2017-06-30	679715.01	2132006.22	17506.74	22797.02	120.88	0.64	1253.81	-1.67	0.86
2017-09-30	679997.57	2131508.98	17186.54	22565.73	116.22	0.65	1218.39	-2.03	0.90
2017-12-31	679163.80	2131897.57	17535.34	23129.23	116.59	0.65	1274.16	-2.00	0.89
2018-03-31	679034.26	2131819.27	17474.69	23374.96	115.47	0.66	1283.25	-2.10	0.90
2018-06-30	679627.97	2131829.71	17110.25	22832.96	118.87	0.66	1227.35	-1.81	0.88
2018-09-30	679571.85	2131905.76	17247.92	23079.54	118.45	0.66	1250.59	-1.85	0.88
2018-12-31	679220.07	2131677.84	18023.82	23018.52	116.32	0.62	1303.39	-2.02	0.90

Table B.3: Los Angeles-Long Beach-Anaheim, CA Metro Area

Date	Centre.x	Centre.y	Sigma.x	Sigma.y	Theta	Eccentricity	Area.sde	TanTheta	SinTheta
1999-03-31	-2023593.19	1454511.02	13042.32	31548.43	146.13	0.91	1292.66	-0.67	0.56
1999-06-30	-2022911.69	1453732.65	14419.24	33153.61	142.56	0.90	1501.84	-0.77	0.61
1999-09-30	-2021810.87	1453055.39	14687.26	35317.81	142.62	0.91	1629.61	-0.76	0.61
1999-12-31	-2022703.09	1452789.83	15039.88	34496.32	143.75	0.90	1629.92	-0.73	0.59
2000-03-31	-2022530.45	1451787.22	13231.22	35710.72	145.87	0.93	1484.39	-0.68	0.56
2000-06-30	-2022510.73	1452292.70	13470.70	33814.03	146.87	0.92	1430.99	-0.65	0.55
2000-09-30	-2022849.11	1451899.80	13055.67	34507.76	146.83	0.93	1415.36	-0.65	0.55
2000-12-31	-2022629.58	1452778.35	14429.23	34746.82	145.21	0.91	1575.10	-0.69	0.57
2001-03-31	-2022972.01	1452960.58	13520.52	34210.89	146.18	0.92	1453.14	-0.67	0.56
2001-06-30	-2023088.66	1452681.32	13043.43	34576.56	146.00	0.93	1416.85	-0.67	0.56
2001-09-30	-2023406.69	1453234.04	13503.64	34664.90	146.42	0.92	1470.59	-0.66	0.55
2001-12-31	-2022502.19	1451862.91	13131.79	36059.51	146.84	0.93	1487.63	-0.65	0.55
2002-03-31	-2021782.29	1451259.97	13207.61	36616.93	146.16	0.93	1519.34	-0.67	0.56
2002-06-30	-2022719.70	1452481.52	12867.46	35691.99	146.21	0.93	1442.82	-0.67	0.56
2002-09-30	-2022026.10	1452231.28	12924.97	36463.48	146.37	0.94	1480.60	-0.67	0.55
2002-12-31	-2022887.30	1453142.14	12541.84	35936.62	145.81	0.94	1415.95	-0.68	0.56
2003-03-31	-2022964.00	1453201.51	12986.63	35525.58	145.48	0.93	1449.40	-0.69	0.57
2003-06-30	-2022867.74	1453397.03	12901.07	34972.92	145.36	0.93	1417.45	-0.69	0.57
2003-09-30	-2022530.18	1453124.88	13060.80	35435.29	145.62	0.93	1453.97	-0.68	0.56
2003-12-31	-2023156.71	1453754.56	13103.91	34135.54	145.20	0.92	1405.26	-0.70	0.57
2004-03-31	-2023315.44	1454357.84	13051.63	33645.44	144.60	0.92	1379.56	-0.71	0.58
2004-06-30	-2023892.93	1455126.92	13304.40	32512.61	145.17	0.91	1358.93	-0.70	0.57
2004-09-30	-2022580.84	1453464.64	13136.04	34802.17	146.06	0.93	1436.22	-0.67	0.56
2004-12-31	-2022989.87	1453280.38	13521.08	34930.77	145.92	0.92	1483.78	-0.68	0.56
2005-03-31	-2023071.90	1453426.83	13380.15	34605.72	145.90	0.92	1454.65	-0.68	0.56
2005-06-30	-2022682.11	1452936.57	13686.57	35545.25	146.38	0.92	1528.36	-0.66	0.55
2005-09-30	-2023266.03	1453736.39	13778.89	34245.58	147.04	0.92	1482.41	-0.65	0.54
2005-12-31	-2022846.85	1453818.64	13829.11	34564.79	144.39	0.92	1501.68	-0.72	0.58
2006-03-31	-2023188.39	1454500.96	14028.24	33230.67	143.84	0.91	1464.51	-0.73	0.59

Table B.3: Los Angeles-Long Beach-Anaheim, CA Metro Area

2006-06-30	-2023292.85	1454671.17	14344.61	32948.97	143.25	0.90	1484.84	-0.75	0.60
2006-09-30	-2023544.07	1454661.88	14056.58	32836.33	143.15	0.90	1450.05	-0.75	0.60
2006-12-31	-2022851.59	1452900.53	14634.99	35086.93	144.85	0.91	1613.20	-0.70	0.58
2007-03-31	-2023112.48	1453229.74	14539.25	34814.03	144.90	0.91	1590.18	-0.70	0.57
2007-06-30	-2023550.09	1453434.57	14420.34	34676.88	145.49	0.91	1570.96	-0.69	0.57
2007-09-30	-2023532.05	1453601.16	14686.64	34409.41	144.82	0.90	1587.63	-0.71	0.58
2007-12-31	-2024498.15	1453334.62	14090.49	35286.31	144.38	0.92	1562.00	-0.72	0.58
2008-03-31	-2024301.11	1453804.01	15197.16	34315.30	142.75	0.90	1638.33	-0.76	0.61
2008-06-30	-2023848.53	1453427.70	15101.69	34868.31	143.17	0.90	1654.27	-0.75	0.60
2008-09-30	-2024426.59	1454184.39	15142.13	33927.09	143.43	0.89	1613.92	-0.74	0.60
2008-12-31	-2024207.78	1454153.42	14831.54	34095.94	142.84	0.90	1588.69	-0.76	0.60
2009-03-31	-2024204.92	1454564.29	13743.45	34155.63	144.68	0.92	1474.71	-0.71	0.58
2009-06-30	-2023327.54	1453647.41	13983.34	35589.89	145.61	0.92	1563.46	-0.68	0.56
2009-09-30	-2023490.65	1452853.32	13453.34	36503.82	145.70	0.93	1542.83	-0.68	0.56
2009-12-31	-2023140.76	1453076.89	13994.83	36158.49	145.55	0.92	1589.75	-0.69	0.57
2010-03-31	-2023046.02	1453207.67	14156.65	36012.59	145.61	0.92	1601.64	-0.68	0.56
2010-06-30	-2023736.06	1453141.89	13455.64	35981.75	145.02	0.93	1521.03	-0.70	0.57
2010-09-30	-2023896.70	1453663.02	13776.36	35354.65	144.87	0.92	1530.14	-0.70	0.58
2010-12-31	-2022934.97	1451965.34	14292.27	37076.87	143.98	0.92	1664.77	-0.73	0.59
2011-03-31	-2022755.35	1451545.62	14369.86	37889.29	144.31	0.93	1710.48	-0.72	0.58
2011-06-30	-2022755.35	1451545.62	14369.86	37889.29	144.31	0.93	1710.48	-0.72	0.58
2011-09-30	-2022868.98	1452132.47	14726.09	37101.70	143.78	0.92	1716.45	-0.73	0.59
2011-12-31	-2023550.09	1451796.88	14683.16	36950.27	143.38	0.92	1704.46	-0.74	0.60
2012-03-31	-2022907.48	1451185.56	14573.15	37789.89	143.55	0.92	1730.13	-0.74	0.59
2012-06-30	-2023280.07	1452013.05	14652.84	36495.24	143.27	0.92	1679.99	-0.75	0.60
2012-09-30	-2023173.83	1451804.37	14612.86	36385.72	143.53	0.92	1670.38	-0.74	0.59
2012-12-31	-2022962.64	1451970.30	14873.14	35501.12	143.25	0.91	1658.80	-0.75	0.60
2013-03-31	-2023143.86	1452270.19	14917.88	34980.16	143.23	0.90	1639.38	-0.75	0.60
2013-06-30	-2023142.59	1452123.89	14939.68	34995.56	143.11	0.90	1642.49	-0.75	0.60
2013-09-30	-2022995.54	1452035.88	14968.79	35089.39	143.35	0.90	1650.11	-0.74	0.60

Table B.3: Los Angeles-Long Beach-Anaheim, CA Metro Area

2013-12-31	-2022435.16	1450575.25	14925.76	36918.40	144.80	0.91	1731.13	-0.71	0.58
2014-03-31	-2022107.07	1450373.76	15022.41	37186.40	145.03	0.91	1754.99	-0.70	0.57
2014-06-30	-2022113.18	1450400.05	15011.02	37212.01	145.03	0.92	1754.86	-0.70	0.57
2014-09-30	-2021901.10	1450256.54	15004.00	37163.31	145.24	0.91	1751.75	-0.69	0.57
2014-12-31	-2021593.22	1448993.68	14822.69	38283.51	145.54	0.92	1782.74	-0.69	0.57
2015-03-31	-2021837.04	1449244.90	14781.51	38189.12	145.38	0.92	1773.41	-0.69	0.57
2015-06-30	-2021826.81	1449190.65	14799.02	38323.43	145.38	0.92	1781.75	-0.69	0.57
2015-09-30	-2021966.89	1449131.51	14512.30	38651.25	145.28	0.93	1762.18	-0.69	0.57
2015-12-31	-2022583.27	1450342.59	14800.40	37733.94	144.61	0.92	1754.51	-0.71	0.58
2016-03-31	-2022632.48	1450531.52	14823.84	37702.23	144.47	0.92	1755.81	-0.71	0.58
2016-06-30	-2022535.78	1450350.08	14771.02	37706.22	144.72	0.92	1749.74	-0.71	0.58
2016-09-30	-2022315.93	1449985.14	14721.71	38033.47	145.05	0.92	1759.03	-0.70	0.57
2016-12-31	-2021555.18	1449583.41	14821.36	38814.23	145.79	0.92	1807.30	-0.68	0.56
2017-03-31	-2021358.41	1449409.45	14898.94	38960.49	145.92	0.92	1823.60	-0.68	0.56
2017-06-30	-2021088.00	1448938.84	14713.95	39525.81	145.81	0.93	1827.09	-0.68	0.56
2017-09-30	-2021211.19	1449043.91	14754.60	39766.99	145.61	0.93	1843.32	-0.68	0.56
2017-12-31	-2020714.55	1447689.31	14664.76	41519.60	145.08	0.94	1912.84	-0.70	0.57
2018-03-31	-2020877.32	1447791.08	14525.85	41585.42	145.16	0.94	1897.72	-0.70	0.57
2018-06-30	-2020657.98	1447694.91	14388.31	41820.51	145.14	0.94	1890.38	-0.70	0.57
2018-09-30	-2021186.23	1448444.59	14524.31	41055.09	145.27	0.94	1873.32	-0.69	0.57
2018-12-31	-2020665.85	1446998.10	14296.57	42132.20	145.41	0.94	1892.33	-0.69	0.57

Table B.4: New York-Newark-Jersey City, NY-NJ-PA Metro Area

Date	Centre.x	Centre.y	Sigma.x	Sigma.y	Theta	Eccentricity	Area.sde	TanTheta	SinTheta
1999-03-31	1826489.60	2185498.50	10716.28	18070.34	28.32	0.81	608.36	0.54	0.47
1999-06-30	1826285.94	2185108.81	10871.93	18067.84	29.76	0.80	617.11	0.57	0.50
1999-09-30	1826395.64	2185074.51	10851.36	17995.06	30.44	0.80	613.46	0.59	0.51
1999-12-31	1826488.75	2185436.63	10696.88	18500.20	27.75	0.82	621.70	0.53	0.47
2000-03-31	1826510.63	2185456.18	10641.77	18674.77	26.58	0.82	624.34	0.50	0.45
2000-06-30	1826604.46	2185401.00	11106.52	18791.64	31.28	0.81	655.68	0.61	0.52
2000-09-30	1826801.62	2185613.57	10937.95	19091.76	28.97	0.82	656.04	0.55	0.48
2000-12-31	1826770.56	2185335.68	10355.00	17985.13	30.52	0.82	585.08	0.59	0.51
2001-03-31	1826552.68	2185345.98	10939.91	18030.14	32.93	0.79	619.67	0.65	0.54
2001-06-30	1826598.86	2185552.82	10981.44	18794.12	32.04	0.81	648.38	0.63	0.53
2001-09-30	1826708.26	2185903.20	10743.48	19297.05	31.59	0.83	651.31	0.61	0.52
2001-12-31	1826418.28	2185623.48	11141.30	18980.70	32.62	0.81	664.35	0.64	0.54
2002-03-31	1826660.02	2185882.27	11186.69	19415.53	31.12	0.82	682.34	0.60	0.52
2002-06-30	1826255.66	2185559.89	11392.53	19014.93	34.77	0.80	680.56	0.69	0.57
2002-09-30	1826482.54	2185670.00	11474.98	19200.35	35.88	0.80	692.17	0.72	0.59
2002-12-31	1826354.93	2185647.47	11347.66	19097.65	36.78	0.80	680.83	0.75	0.60
2003-03-31	1826293.41	2185677.63	11598.79	19209.08	36.89	0.80	699.95	0.75	0.60
2003-06-30	1826300.42	2185692.94	11630.19	19202.57	36.66	0.80	701.61	0.74	0.60
2003-09-30	1826368.37	2185488.14	11633.25	19114.14	35.51	0.79	698.56	0.71	0.58
2003-12-31	1826331.29	2185236.76	11073.22	17787.03	38.42	0.78	618.77	0.79	0.62
2004-03-31	1826414.50	2185263.70	10945.70	17480.32	37.21	0.78	601.09	0.76	0.60
2004-06-30	1826476.68	2185400.76	11494.38	18457.38	37.31	0.78	666.51	0.76	0.61
2004-09-30	1826539.88	2185146.00	10567.74	16840.21	39.04	0.78	559.09	0.81	0.63
2004-12-31	1826675.16	2185607.29	10981.47	18526.30	33.44	0.81	639.14	0.66	0.55
2005-03-31	1826666.87	2185483.96	10906.79	18294.07	35.19	0.80	626.84	0.71	0.58
2005-06-30	1826657.39	2185577.66	11036.54	18454.19	35.42	0.80	639.85	0.71	0.58
2005-09-30	1826505.86	2185636.77	11302.68	18740.52	35.44	0.80	665.45	0.71	0.58
2005-12-31	1826538.25	2185400.81	10345.33	17415.98	37.92	0.80	566.03	0.78	0.61
2006-03-31	1826312.22	2185011.84	10170.23	19204.86	37.23	0.85	613.61	0.76	0.60

Table B.4: New York-Newark-Jersey City, NY-NJ-PA Metro Area

2006-06-30	1826346.65	2184954.17	10072.02	18823.38	38.93	0.84	595.61	0.81	0.63
2006-09-30	1826327.74	2184903.25	9795.09	17931.89	39.25	0.84	551.80	0.82	0.63
2006-12-31	1826304.58	2185021.17	10659.54	17469.66	38.86	0.79	585.02	0.81	0.63
2007-03-31	1826607.44	2185209.22	10679.95	16352.37	41.25	0.76	548.66	0.88	0.66
2007-06-30	1826611.56	2185191.38	10710.91	16470.06	41.25	0.76	554.21	0.88	0.66
2007-09-30	1826493.14	2184860.01	10555.15	18119.27	39.05	0.81	600.83	0.81	0.63
2007-12-31	1826324.42	2184862.38	10562.43	17727.64	38.33	0.80	588.25	0.79	0.62
2008-03-31	1826419.79	2184909.40	10634.68	17633.24	38.19	0.80	589.12	0.79	0.62
2008-06-30	1826535.85	2184960.01	10466.79	17309.36	37.20	0.80	569.17	0.76	0.60
2008-09-30	1826464.01	2184888.01	10678.59	17438.25	39.74	0.79	585.01	0.83	0.64
2008-12-31	1826491.79	2184872.47	10220.55	17105.32	43.56	0.80	549.23	0.95	0.69
2009-03-31	1826456.35	2184853.37	10458.26	17277.58	42.51	0.80	567.67	0.92	0.68
2009-06-30	1826573.89	2185032.72	10752.09	17896.67	42.28	0.80	604.53	0.91	0.67
2009-09-30	1826482.27	2184956.12	10744.83	17784.16	42.53	0.80	600.32	0.92	0.68
2009-12-31	1826343.91	2184975.37	11257.92	18108.82	40.97	0.78	640.47	0.87	0.66
2010-03-31	1826428.81	2185101.31	11336.57	18348.25	40.58	0.79	653.47	0.86	0.65
2010-06-30	1826484.36	2185204.43	11638.78	18887.81	37.17	0.79	690.62	0.76	0.60
2010-09-30	1826501.01	2185243.94	11624.66	18906.44	36.96	0.79	690.46	0.75	0.60
2010-12-31	1826433.51	2185260.64	11723.86	18009.29	35.68	0.76	663.31	0.72	0.58
2011-03-31	1826312.82	2185266.27	12070.26	18150.38	33.00	0.75	688.26	0.65	0.54
2011-06-30	1826262.13	2185069.99	11759.53	17533.77	36.19	0.74	647.76	0.73	0.59
2011-09-30	1826176.54	2184985.59	12045.62	17917.91	37.29	0.74	678.06	0.76	0.61
2011-12-31	1826345.19	2185427.42	12928.59	19066.88	30.12	0.74	774.43	0.58	0.50
2012-03-31	1826375.84	2185454.96	12997.07	19209.39	30.87	0.74	784.35	0.60	0.51
2012-06-30	1826383.38	2185556.80	12956.88	19187.72	31.92	0.74	781.04	0.62	0.53
2012-09-30	1826356.41	2185556.53	12929.22	19237.52	31.52	0.74	781.40	0.61	0.52
2012-12-31	1826326.33	2185697.50	12801.93	19215.72	33.32	0.75	772.83	0.66	0.55
2013-03-31	1826408.26	2185713.50	12766.20	19230.61	32.73	0.75	771.27	0.64	0.54
2013-06-30	1826427.04	2185719.39	12857.06	18643.28	32.05	0.72	753.03	0.63	0.53
2013-09-30	1826368.18	2185707.71	12940.92	18676.21	32.11	0.72	759.28	0.63	0.53

Table B.4: New York-Newark-Jersey City, NY-NJ-PA Metro Area

2013-12-31	1826354.18	2185594.42	13357.01	19243.52	42.77	0.72	807.50	0.93	0.68
2014-03-31	1826410.02	2185634.89	13687.82	19964.60	47.56	0.73	858.51	1.09	0.74
2014-06-30	1826508.43	2185707.21	13901.16	20747.69	50.89	0.74	906.09	1.23	0.78
2014-09-30	1826525.78	2185563.76	12675.68	20563.29	55.08	0.79	818.87	1.43	0.82
2014-12-31	1826458.70	2185677.13	13683.31	20327.89	49.81	0.74	873.84	1.18	0.76
2015-03-31	1826520.46	2185607.01	12731.49	20378.87	51.91	0.78	815.10	1.28	0.79
2015-06-30	1826542.75	2185653.61	12801.28	20232.99	52.75	0.77	813.70	1.32	0.80
2015-09-30	1826553.37	2185574.78	12719.33	20140.51	54.13	0.78	804.79	1.38	0.81
2015-12-31	1826554.43	2185321.54	12493.21	20206.44	52.38	0.79	793.07	1.30	0.79
2016-03-31	1826571.02	2185307.01	12210.16	20194.86	53.49	0.80	774.66	1.35	0.80
2016-06-30	1826639.77	2185284.30	12126.23	20272.68	54.06	0.80	772.30	1.38	0.81
2016-09-30	1826522.30	2185194.36	12067.99	20457.26	55.01	0.81	775.59	1.43	0.82
2016-12-31	1826059.87	2184908.51	12618.39	20300.95	52.13	0.78	804.77	1.29	0.79
2017-03-31	1825994.80	2185060.48	12547.64	20294.67	54.31	0.79	800.01	1.39	0.81
2017-06-30	1825981.35	2185240.61	13816.22	20558.71	49.87	0.74	892.35	1.19	0.76
2017-09-30	1826038.63	2185090.96	12615.87	20341.79	53.70	0.78	806.23	1.36	0.81
2017-12-31	1826028.54	2185382.99	13135.30	20407.78	53.06	0.77	842.14	1.33	0.80
2018-03-31	1826043.98	2185373.17	13211.14	20536.14	53.05	0.77	852.33	1.33	0.80
2018-06-30	1825933.07	2185357.94	13758.33	20620.32	55.17	0.74	891.27	1.44	0.82
2018-09-30	1825883.23	2185327.09	13835.64	20723.12	55.00	0.74	900.75	1.43	0.82
2018-12-31	1825904.61	2185266.52	14098.91	20771.99	55.06	0.73	920.05	1.43	0.82

Table B.5: San Francisco-Oakland-Hayward, CA Metro Area

Date	Centre.x	Centre.y	Sigma.x	Sigma.y	Theta	Eccentricity	Area.sde	TanTheta	SinTheta
1999-03-31	-2271098.50	1951744.25	11619.49	25357.42	174.57	0.89	925.64	-0.10	0.09
1999-06-30	-2270950.28	1952069.72	11936.19	25511.46	174.67	0.88	956.65	-0.09	0.09
1999-09-30	-2271031.81	1952555.24	11893.69	24576.70	174.35	0.88	918.31	-0.10	0.10
1999-12-31	-2271110.60	1950865.24	11207.77	26723.12	174.19	0.91	940.93	-0.10	0.10
2000-03-31	-2271575.47	1950912.82	10400.03	26928.92	173.60	0.92	879.84	-0.11	0.11
2000-06-30	-2271481.98	1950708.92	10444.54	27030.83	173.86	0.92	886.95	-0.11	0.11
2000-09-30	-2271807.14	1950614.42	8595.84	26676.71	176.16	0.95	720.39	-0.07	0.07
2000-12-31	-2271654.37	1949185.33	8937.33	29561.79	175.30	0.95	830.02	-0.08	0.08
2001-03-31	-2271832.37	1949372.84	8270.71	29518.50	174.97	0.96	766.99	-0.09	0.09
2001-06-30	-2271864.14	1949656.94	8397.29	29647.30	175.02	0.96	782.12	-0.09	0.09
2001-09-30	-2271783.30	1949399.75	8470.62	29459.26	175.01	0.96	783.95	-0.09	0.09
2001-12-31	-2272124.20	1952351.67	8684.80	25448.38	175.64	0.94	694.34	-0.08	0.08
2002-03-31	-2272414.59	1952578.33	7529.38	25737.52	175.27	0.96	608.80	-0.08	0.08
2002-06-30	-2272404.01	1952540.92	7570.95	25829.93	175.27	0.96	614.36	-0.08	0.08
2002-09-30	-2272414.24	1952576.51	7547.07	25734.42	175.25	0.96	610.16	-0.08	0.08
2002-12-31	-2272497.49	1953616.77	7708.72	24749.30	175.52	0.95	599.37	-0.08	0.08
2003-03-31	-2272374.71	1953450.34	7963.14	24852.76	175.81	0.95	621.74	-0.07	0.07
2003-06-30	-2272470.45	1953692.61	7999.02	24790.68	175.86	0.95	622.98	-0.07	0.07
2003-09-30	-2272416.31	1952797.96	7859.96	25730.81	175.81	0.95	635.37	-0.07	0.07
2003-12-31	-2272424.48	1954572.25	8749.87	23370.84	175.82	0.93	642.43	-0.07	0.07
2004-03-31	-2272484.41	1954762.69	8702.26	23329.57	175.61	0.93	637.81	-0.08	0.08
2004-06-30	-2272469.56	1954703.82	8805.56	23615.45	175.64	0.93	653.29	-0.08	0.08
2004-09-30	-2272463.17	1954903.33	8698.30	23007.81	175.49	0.93	628.72	-0.08	0.08
2004-12-31	-2272557.83	1955271.11	8497.22	22009.18	173.89	0.92	587.53	-0.11	0.11
2005-03-31	-2272268.54	1955112.02	9386.47	22092.99	174.29	0.91	651.49	-0.10	0.10
2005-06-30	-2272245.09	1955207.50	9397.79	22131.61	174.03	0.91	653.41	-0.10	0.10
2005-09-30	-2272225.76	1955012.92	9444.09	22217.22	174.32	0.91	659.17	-0.10	0.10
2005-12-31	-2271889.19	1954218.18	9914.82	22902.63	174.66	0.90	713.38	-0.09	0.09
2006-03-31	-2272139.26	1954471.69	9470.11	23126.57	174.11	0.91	688.04	-0.10	0.10

Table B.5: San Francisco-Oakland-Hayward, CA Metro Area

2006-06-30	-2271970.72	1954706.46	9977.96	22672.30	174.23	0.90	710.70	-0.10	0.10
2006-09-30	-2271942.92	1954621.20	10123.40	22986.20	174.25	0.90	731.04	-0.10	0.10
2006-12-31	-2271217.33	1953343.89	11899.69	25536.77	172.70	0.88	954.67	-0.13	0.13
2007-03-31	-2271072.73	1953143.04	11950.39	25509.60	172.97	0.88	957.71	-0.12	0.12
2007-06-30	-2271027.46	1953105.95	12018.90	25711.53	173.01	0.88	970.83	-0.12	0.12
2007-09-30	-2271017.14	1953022.54	11971.75	25505.34	173.15	0.88	959.26	-0.12	0.12
2007-12-31	-2271322.73	1953146.54	11439.45	24802.00	172.86	0.89	891.34	-0.13	0.12
2008-03-31	-2271289.92	1953124.40	11458.20	24684.20	172.83	0.89	888.56	-0.13	0.12
2008-06-30	-2271280.65	1953270.29	11479.23	24208.19	172.89	0.88	873.02	-0.12	0.12
2008-09-30	-2271414.03	1952915.38	11241.28	25219.94	172.74	0.90	890.66	-0.13	0.13
2008-12-31	-2271100.38	1953542.99	11695.53	24579.52	173.09	0.88	903.11	-0.12	0.12
2009-03-31	-2271147.16	1953840.71	11711.55	24104.94	172.77	0.87	886.89	-0.13	0.13
2009-06-30	-2271304.28	1954187.95	11517.79	23275.84	172.30	0.87	842.22	-0.14	0.13
2009-09-30	-2271311.65	1953546.83	11301.69	23444.27	172.36	0.88	832.40	-0.13	0.13
2009-12-31	-2271493.50	1954268.71	11152.37	23242.84	172.86	0.88	814.34	-0.13	0.12
2010-03-31	-2271227.14	1953992.63	11690.19	22661.79	174.06	0.86	832.27	-0.10	0.10
2010-06-30	-2271021.82	1953619.83	12080.21	23233.25	173.44	0.85	881.73	-0.11	0.11
2010-09-30	-2271033.69	1953866.47	12115.00	22846.96	173.20	0.85	869.56	-0.12	0.12
2010-12-31	-2271151.27	1954029.42	11754.80	23219.11	171.89	0.86	857.45	-0.14	0.14
2011-03-31	-2271189.43	1953165.44	11696.47	24181.83	172.08	0.88	888.57	-0.14	0.14
2011-06-30	-2271143.29	1953466.00	11823.20	23843.99	171.99	0.87	885.65	-0.14	0.14
2011-09-30	-2271310.60	1953132.45	11533.17	24500.93	172.04	0.88	887.73	-0.14	0.14
2011-12-31	-2270997.62	1952845.98	11859.40	24354.92	171.58	0.87	907.40	-0.15	0.15
2012-03-31	-2270780.23	1952750.23	12302.58	24630.31	171.84	0.87	951.95	-0.14	0.14
2012-06-30	-2270716.45	1952949.57	12722.28	24392.07	172.28	0.85	974.91	-0.14	0.13
2012-09-30	-2270383.10	1952910.62	13159.60	24189.79	173.04	0.84	1000.06	-0.12	0.12
2012-12-31	-2271054.87	1952706.84	11812.84	24956.65	173.72	0.88	926.17	-0.11	0.11
2013-03-31	-2271017.10	1952765.61	11976.75	24868.09	173.43	0.88	935.69	-0.12	0.11
2013-06-30	-2270829.01	1952936.50	12410.50	24759.18	173.59	0.87	965.33	-0.11	0.11
2013-09-30	-2270603.73	1952886.95	12842.01	24128.04	174.13	0.85	973.43	-0.10	0.10

Table B.5: San Francisco-Oakland-Hayward, CA Metro Area

2013-12-31	-2270519.53	1952711.59	12865.85	24935.00	174.26	0.86	1007.85	-0.10	0.10
2014-03-31	-2270401.17	1952813.12	13210.79	24914.42	174.37	0.85	1034.02	-0.10	0.10
2014-06-30	-2270479.75	1952472.46	12972.72	25192.38	174.19	0.86	1026.72	-0.10	0.10
2014-09-30	-2270493.17	1952440.30	13047.18	25265.86	174.27	0.86	1035.62	-0.10	0.10
2014-12-31	-2270143.86	1950992.31	13637.89	26246.54	172.99	0.85	1124.52	-0.12	0.12
2015-03-31	-2270029.49	1951565.37	14091.91	25973.67	172.69	0.84	1149.88	-0.13	0.13
2015-06-30	-2270095.67	1951887.93	14071.34	25479.41	172.33	0.83	1126.35	-0.13	0.13
2015-09-30	-2269993.25	1951864.96	14163.67	25555.42	172.36	0.83	1137.13	-0.13	0.13
2015-12-31	-2269875.61	1951730.90	14045.14	25520.52	172.86	0.83	1126.07	-0.13	0.12
2016-03-31	-2270100.12	1952102.05	13819.53	25350.16	172.59	0.84	1100.59	-0.13	0.13
2016-06-30	-2270102.37	1951973.87	13723.77	25438.91	172.51	0.84	1096.79	-0.13	0.13
2016-09-30	-2270056.09	1952158.46	13997.93	25368.20	172.35	0.83	1115.59	-0.13	0.13
2016-12-31	-2270153.71	1953025.85	13991.28	24242.83	171.06	0.82	1065.59	-0.16	0.16
2017-03-31	-2270034.26	1952921.64	14271.78	24493.59	170.77	0.81	1098.20	-0.16	0.16
2017-06-30	-2269924.13	1952981.34	14538.30	24343.69	170.26	0.80	1111.86	-0.17	0.17
2017-09-30	-2269892.17	1953127.04	14561.53	24320.32	170.72	0.80	1112.57	-0.16	0.16
2017-12-31	-2269889.83	1952871.58	14691.40	24949.93	170.99	0.81	1151.55	-0.16	0.16
2018-03-31	-2269802.52	1952382.22	14429.27	25513.97	170.67	0.82	1156.57	-0.16	0.16
2018-06-30	-2269730.55	1952266.81	14553.57	25767.13	170.78	0.83	1178.11	-0.16	0.16
2018-09-30	-2269509.49	1952268.26	14776.42	25965.08	170.94	0.82	1205.34	-0.16	0.16
2018-12-31	-2269552.63	1952321.38	14898.39	25787.05	171.21	0.82	1206.95	-0.15	0.15

Appendix C

Gravity Model

Table C.1: Gravity Model of Trade applied to Institutional Investment for the third quarter of 2013

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.65^{***}	-7.40^{***}	-6.31^{***}	-29.51^{***}	-6.10^{***}	-28.74^{***}
	(0.10)	(0.10)	(0.10)	(0.18)	(0.10)	(0.18)
dist_log	-0.10^{***}	-0.14^{***}	-0.24^{***}	-0.33^{***}	-0.28^{***}	-0.34^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.23^{***}	0.22^{***}	0.17^{***}	0.15^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.20^{***}	0.20^{***}	0.15^{***}	0.12^{***}	0.15^{***}	0.12^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.88^{***}			1.82^{***}	1.48^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		0.96^{***}			0.71^{***}	0.19^{***}
		(0.04)			(0.03)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.43^{***}		2.31^{***}
				(0.02)		(0.02)
Destination_Population_log				2.40^{***}		2.38^{***}
				(0.02)		(0.02)
R^2	0.24	0.25	0.33	0.31	0.34	0.32
Adj. R^2	0.24	0.25	0.33	0.31	0.34	0.32
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.18	5.15	4.86	4.94	4.82	4.92

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.2: Gravity Model of Trade applied to Institutional Investment for the fourth quarter of 2013

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.49^{***}	-7.25^{***}	-6.13^{***}	-29.90^{***}	-5.93^{***}	-29.09^{***}
	(0.10)	(0.10)	(0.10)	(0.18)	(0.10)	(0.18)
dist_log	-0.14^{***}	-0.17^{***}	-0.28^{***}	-0.37^{***}	-0.31^{***}	-0.38^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.23^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.21^{***}	0.20^{***}	0.16^{***}	0.12^{***}	0.15^{***}	0.12^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.92^{***}			1.85^{***}	1.49^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.02^{***}			0.76^{***}	0.22^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.47^{***}		2.36^{***}
				(0.02)		(0.02)
Destination_Population_log				2.48^{***}		2.45^{***}
				(0.02)		(0.02)
R^2	0.24	0.25	0.33	0.31	0.34	0.32
Adj. R^2	0.24	0.25	0.33	0.31	0.34	0.32
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.23	5.19	4.90	4.97	4.86	4.96

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.3: Gravity Model of Trade applied to Institutional Investment for the first quarter of 2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.47^{***}	-7.22^{***}	-6.13^{***}	-29.99^{***}	-5.92^{***}	-29.16^{***}
	(0.10)	(0.10)	(0.10)	(0.18)	(0.10)	(0.18)
dist_log	-0.15^{***}	-0.19^{***}	-0.29^{***}	-0.37^{***}	-0.32^{***}	-0.39^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.21^{***}	0.20^{***}	0.16^{***}	0.12^{***}	0.16^{***}	0.12^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.94^{***}			1.88^{***}	1.52^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.05^{***}			0.79^{***}	0.25^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.48^{***}		2.36^{***}
				(0.02)		(0.02)
Destination_Population_log				2.50^{***}		2.47^{***}
				(0.02)		(0.02)
R^2	0.24	0.26	0.34	0.31	0.34	0.32
Adj. R^2	0.24	0.26	0.34	0.31	0.34	0.32
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.25	5.21	4.92	5.00	4.89	4.98

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.4: Gravity Model of Trade applied to Institutional Investment for the second quarter of 2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.50^{***}	-7.25^{***}	-6.14^{***}	-30.26^{***}	-5.93^{***}	-29.42^{***}
	(0.11)	(0.10)	(0.10)	(0.18)	(0.10)	(0.18)
dist_log	-0.15^{***}	-0.19^{***}	-0.29^{***}	-0.38^{***}	-0.33^{***}	-0.40^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.21^{***}	0.21^{***}	0.16^{***}	0.12^{***}	0.16^{***}	0.12^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.90^{***}			1.83^{***}	1.48^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.09^{***}			0.83^{***}	0.29^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.50^{***}		2.38^{***}
				(0.02)		(0.02)
Destination_Population_log				2.54^{***}		2.51^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.32	0.35	0.32
Adj. R^2	0.25	0.26	0.34	0.32	0.35	0.32
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.28	5.24	4.95	5.02	4.92	5.01

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.5: Gravity Model of Trade applied to Institutional Investment for the third quarter of 2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.48^{***}	-7.23^{***}	-6.13^{***}	-29.73^{***}	-5.93^{***}	-28.93^{***}
	(0.11)	(0.10)	(0.10)	(0.18)	(0.10)	(0.18)
dist_log	-0.15^{***}	-0.18^{***}	-0.29^{***}	-0.38^{***}	-0.32^{***}	-0.39^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.21^{***}	0.21^{***}	0.16^{***}	0.13^{***}	0.16^{***}	0.12^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.84^{***}			1.78^{***}	1.44^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.06^{***}			0.81^{***}	0.28^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.42^{***}		2.31^{***}
				(0.02)		(0.02)
Destination_Population_log				2.51^{***}		2.48^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.32	0.35	0.32
Adj. R^2	0.25	0.26	0.34	0.32	0.35	0.32
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.26	5.22	4.94	5.01	4.91	4.99

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.6: Gravity Model of Trade applied to Institutional Investment for the fourth quarter of 2014

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.38^{***}	-7.14^{***}	-6.03^{***}	-30.82^{***}	-5.83^{***}	-29.92^{***}
	(0.11)	(0.11)	(0.10)	(0.18)	(0.10)	(0.19)
dist_log	-0.18^{***}	-0.22^{***}	-0.32^{***}	-0.41^{***}	-0.35^{***}	-0.42^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.22^{***}	0.21^{***}	0.17^{***}	0.13^{***}	0.16^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.97^{***}			1.88^{***}	1.50^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.13^{***}			0.87^{***}	0.30^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.51^{***}		2.38^{***}
				(0.02)		(0.02)
Destination_Population_log				2.66^{***}		2.63^{***}
				(0.02)		(0.02)
R^2	0.24	0.26	0.34	0.32	0.35	0.32
Adj. R^2	0.24	0.26	0.34	0.32	0.35	0.32
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.35	5.30	5.01	5.07	4.97	5.06

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.7: Gravity Model of Trade applied to Institutional Investment for the first quarter of 2015

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.35^{***}	-7.12^{***}	-5.99^{***}	-30.93^{***}	-5.80^{***}	-30.10^{***}
	(0.11)	(0.11)	(0.10)	(0.18)	(0.10)	(0.19)
dist_log	-0.18^{***}	-0.22^{***}	-0.32^{***}	-0.41^{***}	-0.35^{***}	-0.43^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.23^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.22^{***}	0.21^{***}	0.16^{***}	0.12^{***}	0.16^{***}	0.12^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.85^{***}			1.74^{***}	1.36^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.11^{***}			0.84^{***}	0.27^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.51^{***}		2.39^{***}
				(0.02)		(0.02)
Destination_Population_log				2.69^{***}		2.66^{***}
				(0.02)		(0.02)
R^2	0.24	0.25	0.34	0.32	0.34	0.32
Adj. R^2	0.24	0.25	0.34	0.32	0.34	0.32
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.35	5.31	5.01	5.08	4.98	5.06

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.8: Gravity Model of Trade applied to Institutional Investment for the second quarter of 2015

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.55^{***}	-7.31^{***}	-6.15^{***}	-31.62^{***}	-5.95^{***}	-30.72^{***}
	(0.11)	(0.11)	(0.10)	(0.18)	(0.10)	(0.19)
dist_log	-0.17^{***}	-0.21^{***}	-0.31^{***}	-0.40^{***}	-0.34^{***}	-0.42^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.22^{***}	0.22^{***}	0.17^{***}	0.13^{***}	0.16^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		1.97^{***}			1.86^{***}	1.46^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.15^{***}			0.87^{***}	0.28^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.57^{***}		2.44^{***}
				(0.02)		(0.02)
Destination_Population_log				2.74^{***}		2.71^{***}
				(0.02)		(0.02)
R^2	0.24	0.26	0.34	0.32	0.35	0.33
Adj. R^2	0.24	0.26	0.34	0.32	0.35	0.33
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.38	5.34	5.04	5.10	5.00	5.08

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.9: Gravity Model of Trade applied to Institutional Investment for the third quarter of 2015

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.64^{***}	-7.39^{***}	-6.22^{***}	-31.74^{***}	-6.01^{***}	-30.79^{***}
	(0.11)	(0.11)	(0.10)	(0.18)	(0.10)	(0.19)
dist_log	-0.16^{***}	-0.20^{***}	-0.30^{***}	-0.40^{***}	-0.34^{***}	-0.41^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.22^{***}	0.22^{***}	0.17^{***}	0.13^{***}	0.16^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.02^{***}			1.91^{***}	1.52^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.20^{***}			0.92^{***}	0.32^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.59^{***}		2.45^{***}
				(0.02)		(0.02)
Destination_Population_log				2.75^{***}		2.71^{***}
				(0.02)		(0.02)
R^2	0.24	0.26	0.34	0.32	0.35	0.33
Adj. R^2	0.24	0.26	0.34	0.32	0.35	0.33
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.38	5.33	5.03	5.09	4.99	5.07

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.10: Gravity Model of Trade applied to Institutional Investment for the fourth quarter of 2015

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.57^{***}	-7.30^{***}	-6.12^{***}	-32.13^{***}	-5.90^{***}	-31.20^{***}
	(0.11)	(0.11)	(0.10)	(0.18)	(0.10)	(0.19)
dist_log	-0.18^{***}	-0.22^{***}	-0.32^{***}	-0.42^{***}	-0.36^{***}	-0.44^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.23^{***}	0.22^{***}	0.17^{***}	0.13^{***}	0.17^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.07^{***}			1.97^{***}	1.58^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.20^{***}			0.91^{***}	0.30^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.63^{***}		2.50^{***}
				(0.02)		(0.02)
Destination_Population_log				2.81^{***}		2.77^{***}
				(0.02)		(0.02)
R^2	0.24	0.26	0.34	0.32	0.35	0.33
Adj. R^2	0.24	0.26	0.34	0.32	0.35	0.33
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.42	5.37	5.06	5.12	5.02	5.10

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.11: Gravity Model of Trade applied to Institutional Investment for the first quarter of 2016

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.36^{***}	-7.08^{***}	-5.92^{***}	-31.88^{***}	-5.69^{***}	-31.01^{***}
	(0.11)	(0.11)	(0.10)	(0.19)	(0.10)	(0.19)
dist_log	-0.21^{***}	-0.25^{***}	-0.35^{***}	-0.45^{***}	-0.39^{***}	-0.46^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.22^{***}	0.18^{***}	0.15^{***}	0.16^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.23^{***}	0.22^{***}	0.17^{***}	0.13^{***}	0.17^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.02^{***}			1.95^{***}	1.56^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.15^{***}			0.87^{***}	0.29^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.66^{***}		2.54^{***}
				(0.02)		(0.02)
Destination_Population_log				2.77^{***}		2.74^{***}
				(0.02)		(0.02)
R^2	0.24	0.26	0.34	0.32	0.35	0.33
Adj. R^2	0.24	0.26	0.34	0.32	0.35	0.33
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.43	5.39	5.07	5.14	5.03	5.12

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.12: Gravity Model of Trade applied to Institutional Investment for the second quarter of 2016

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.54^{***}	-7.25^{***}	-6.10^{***}	-32.09^{***}	-5.86^{***}	-31.12^{***}
	(0.11)	(0.11)	(0.10)	(0.19)	(0.10)	(0.19)
dist_log	-0.19^{***}	-0.24^{***}	-0.34^{***}	-0.44^{***}	-0.38^{***}	-0.46^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.24^{***}	0.23^{***}	0.19^{***}	0.15^{***}	0.17^{***}	0.15^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.23^{***}	0.22^{***}	0.18^{***}	0.13^{***}	0.17^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.21^{***}			2.10^{***}	1.70^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.13^{***}			0.86^{***}	0.28^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.65^{***}		2.51^{***}
				(0.02)		(0.02)
Destination_Population_log				2.79^{***}		2.75^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.33	0.35	0.33
Adj. R^2	0.25	0.26	0.34	0.33	0.35	0.33
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.47	5.42	5.12	5.18	5.08	5.15

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.13: Gravity Model of Trade applied to Institutional Investment for the third quarter of 2016

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.56^{***}	-7.26^{***}	-6.14^{***}	-32.41^{***}	-5.89^{***}	-31.43^{***}
	(0.11)	(0.11)	(0.10)	(0.19)	(0.10)	(0.19)
dist_log	-0.20^{***}	-0.24^{***}	-0.34^{***}	-0.44^{***}	-0.38^{***}	-0.46^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.25^{***}	0.23^{***}	0.19^{***}	0.16^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.23^{***}	0.23^{***}	0.18^{***}	0.13^{***}	0.17^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.25^{***}			2.18^{***}	1.78^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.18^{***}			0.89^{***}	0.30^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.71^{***}		2.57^{***}
				(0.02)		(0.02)
Destination_Population_log				2.79^{***}		2.75^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.33	0.36	0.34
Adj. R^2	0.25	0.26	0.34	0.33	0.36	0.34
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.52	5.46	5.16	5.22	5.11	5.19

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.14: Gravity Model of Trade applied to Institutional Investment for the fourth quarter of 2016

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.48^{***}	-7.17^{***}	-6.04^{***}	-33.28^{***}	-5.78^{***}	-32.25^{***}
	(0.11)	(0.11)	(0.10)	(0.19)	(0.10)	(0.19)
dist_log	-0.24^{***}	-0.28^{***}	-0.37^{***}	-0.47^{***}	-0.41^{***}	-0.49^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.25^{***}	0.23^{***}	0.19^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.24^{***}	0.23^{***}	0.18^{***}	0.14^{***}	0.18^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.41^{***}			2.34^{***}	1.95^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.20^{***}			0.91^{***}	0.30^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.77^{***}		2.62^{***}
				(0.03)		(0.03)
Destination_Population_log				2.93^{***}		2.89^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.33	0.36	0.34
Adj. R^2	0.25	0.26	0.34	0.33	0.36	0.34
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.61	5.55	5.24	5.30	5.18	5.26

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.15: Gravity Model of Trade applied to Institutional Investment for the first quarter of 2017

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.50^{***}	-7.19^{***}	-6.06^{***}	-33.03^{***}	-5.80^{***}	-31.99^{***}
	(0.11)	(0.11)	(0.10)	(0.19)	(0.10)	(0.19)
dist_log	-0.23^{***}	-0.27^{***}	-0.37^{***}	-0.46^{***}	-0.40^{***}	-0.48^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.25^{***}	0.23^{***}	0.19^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.24^{***}	0.23^{***}	0.18^{***}	0.13^{***}	0.18^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.38^{***}			2.31^{***}	1.93^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.24^{***}			0.94^{***}	0.33^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.68^{***}		2.54^{***}
				(0.03)		(0.03)
Destination_Population_log				2.95^{***}		2.91^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.33	0.36	0.34
Adj. R^2	0.25	0.26	0.34	0.33	0.36	0.34
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.61	5.55	5.23	5.30	5.18	5.27

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.16: Gravity Model of Trade applied to Institutional Investment for the second quarter of 2017

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.62^{***}	-7.30^{***}	-6.17^{***}	-33.34^{***}	-5.90^{***}	-32.29^{***}
	(0.11)	(0.11)	(0.10)	(0.19)	(0.10)	(0.19)
dist_log	-0.21^{***}	-0.25^{***}	-0.35^{***}	-0.45^{***}	-0.39^{***}	-0.46^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.25^{***}	0.23^{***}	0.19^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.24^{***}	0.23^{***}	0.18^{***}	0.13^{***}	0.17^{***}	0.13^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.44^{***}			2.38^{***}	1.99^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.24^{***}			0.94^{***}	0.31^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.73^{***}		2.58^{***}
				(0.03)		(0.03)
Destination_Population_log				2.95^{***}		2.91^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.33	0.36	0.34
Adj. R^2	0.25	0.26	0.34	0.33	0.36	0.34
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.61	5.55	5.24	5.30	5.18	5.26

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.17: Gravity Model of Trade applied to Institutional Investment for the third quarter of 2017

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.50^{***}	-7.17^{***}	-6.03^{***}	-33.19^{***}	-5.75^{***}	-32.14^{***}
	(0.11)	(0.11)	(0.11)	(0.19)	(0.10)	(0.19)
dist_log	-0.22^{***}	-0.27^{***}	-0.37^{***}	-0.47^{***}	-0.41^{***}	-0.49^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.25^{***}	0.23^{***}	0.19^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.24^{***}	0.23^{***}	0.18^{***}	0.14^{***}	0.18^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.51^{***}			2.45^{***}	2.06^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.19^{***}			0.89^{***}	0.28^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.79^{***}		2.64^{***}
				(0.03)		(0.03)
Destination_Population_log				2.90^{***}		2.86^{***}
				(0.02)		(0.02)
R^2	0.25	0.26	0.34	0.33	0.36	0.34
Adj. R^2	0.25	0.26	0.34	0.33	0.36	0.34
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.62	5.56	5.25	5.31	5.20	5.27

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.18: Gravity Model of Trade applied to Institutional Investment for the fourth quarter of 2017

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.69^{***}	-7.36^{***}	-6.15^{***}	-35.00^{***}	-5.87^{***}	-33.88^{***}
	(0.11)	(0.11)	(0.11)	(0.19)	(0.11)	(0.19)
dist_log	-0.24^{***}	-0.29^{***}	-0.39^{***}	-0.49^{***}	-0.43^{***}	-0.51^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.25^{***}	0.23^{***}	0.19^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.25^{***}	0.24^{***}	0.19^{***}	0.14^{***}	0.18^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.60^{***}			2.49^{***}	2.05^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.23^{***}			0.92^{***}	0.28^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.99^{***}		2.82^{***}
				(0.03)		(0.03)
Destination_Population_log				3.05^{***}		3.02^{***}
				(0.02)		(0.03)
R^2	0.25	0.26	0.35	0.34	0.36	0.35
Adj. R^2	0.25	0.26	0.35	0.34	0.36	0.35
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.73	5.67	5.33	5.38	5.27	5.35

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.19: Gravity Model of Trade applied to Institutional Investment for the first quarter of 2018

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.59^{***}	-7.26^{***}	-6.05^{***}	-34.49^{***}	-5.78^{***}	-33.32^{***}
	(0.11)	(0.11)	(0.11)	(0.19)	(0.11)	(0.19)
dist_log	-0.25^{***}	-0.30^{***}	-0.40^{***}	-0.51^{***}	-0.44^{***}	-0.52^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.25^{***}	0.23^{***}	0.19^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.25^{***}	0.24^{***}	0.19^{***}	0.15^{***}	0.19^{***}	0.15^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.66^{***}			2.53^{***}	2.10^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.26^{***}			0.93^{***}	0.30^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.95^{***}		2.77^{***}
				(0.03)		(0.03)
Destination_Population_log				3.01^{***}		2.97^{***}
				(0.02)		(0.02)
R^2	0.25	0.27	0.35	0.34	0.37	0.35
Adj. R^2	0.25	0.27	0.35	0.34	0.37	0.35
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.73	5.66	5.33	5.38	5.27	5.35

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.20: Gravity Model of Trade applied to Institutional Investment for the second quarter of 2018

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.61^{***}	-7.26^{***}	-6.10^{***}	-35.08^{***}	-5.80^{***}	-33.88^{***}
	(0.11)	(0.11)	(0.11)	(0.19)	(0.11)	(0.19)
dist_log	-0.25^{***}	-0.30^{***}	-0.40^{***}	-0.50^{***}	-0.44^{***}	-0.52^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.26^{***}	0.23^{***}	0.20^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.25^{***}	0.24^{***}	0.19^{***}	0.15^{***}	0.19^{***}	0.15^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.72^{***}			2.63^{***}	2.18^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.31^{***}			1.00^{***}	0.36^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				3.05^{***}		2.88^{***}
				(0.03)		(0.03)
Destination_Population_log				3.01^{***}		2.97^{***}
				(0.02)		(0.02)
R^2	0.25	0.27	0.35	0.34	0.37	0.35
Adj. R^2	0.25	0.27	0.35	0.34	0.37	0.35
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.76	5.69	5.36	5.40	5.29	5.36

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.21: Gravity Model of Trade applied to Institutional Investment for the third quarter of 2018

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-7.68^{***}	-7.32^{***}	-6.21^{***}	-34.60^{***}	-5.91^{***}	-33.33^{***}
	(0.11)	(0.11)	(0.11)	(0.19)	(0.11)	(0.19)
dist_log	-0.21^{***}	-0.26^{***}	-0.36^{***}	-0.47^{***}	-0.40^{***}	-0.49^{***}
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.26^{***}	0.23^{***}	0.20^{***}	0.15^{***}	0.17^{***}	0.14^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.24^{***}	0.24^{***}	0.19^{***}	0.15^{***}	0.19^{***}	0.15^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		2.80^{***}			2.72^{***}	2.28^{***}
		(0.04)			(0.04)	(0.04)
Destination_Is_Capital		1.36^{***}			1.02^{***}	0.41^{***}
		(0.04)			(0.04)	(0.04)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				2.99^{***}		2.82^{***}
				(0.03)		(0.03)
Destination_Population_log				2.94^{***}		2.89^{***}
				(0.02)		(0.03)
R^2	0.26	0.28	0.36	0.35	0.37	0.36
Adj. R^2	0.26	0.28	0.36	0.35	0.37	0.36
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	5.73	5.66	5.35	5.39	5.28	5.35

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.22: Gravity Model of Trade applied to Institutional Investment for the fourth quarter of 2018

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-5.39^{***}	-5.35^{***}	-5.24^{***}	-8.20^{***}	-5.20^{***}	-8.04^{***}
	(0.05)	(0.05)	(0.05)	(0.09)	(0.05)	(0.09)
dist_log	-0.03^{***}	-0.03^{***}	-0.04^{***}	-0.05^{***}	-0.05^{***}	-0.05^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inc_o_log	0.05^{***}	0.05^{***}	0.04^{***}	0.03^{***}	0.04^{***}	0.03^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
inc_d_log	0.21^{***}	0.21^{***}	0.20^{***}	0.21^{***}	0.20^{***}	0.21^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Origin_Is_Capital		0.44^{***}			0.44^{***}	0.37^{***}
		(0.02)			(0.02)	(0.02)
Destination_Is_Capital		0.06^{***}			0.03	0.02
		(0.02)			(0.02)	(0.02)
Origin_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Destination_Population			0.00^{***}		0.00^{***}	
			(0.00)		(0.00)	
Origin_Population_log				0.49^{***}		0.46^{***}
				(0.01)		(0.01)
Destination_Population_log				0.12^{***}		0.12^{***}
				(0.01)		(0.01)
R^2	0.24	0.24	0.25	0.25	0.25	0.25
Adj. R^2	0.24	0.24	0.25	0.25	0.25	0.25
Num. obs.	214832	214832	214832	214832	214832	214832
RMSE	2.53	2.53	2.52	2.52	2.51	2.52

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Appendix D

Definitions

The Definition for the Emerging Hot Spots are taken from ESRI (2019) on the help page detailing Emerging Hotspot analysis[1].

Table D.1: Emerging Hot Spot Analysis

Pattern Name	Definition
No Pattern Detected	Does not fall into any of the hot or cold spot patterns defined below.
New Hot Spot	A location that is a statistically significant hot spot for the final time step and has never been a statistically significant hot spot before.

[1] https://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/learnmoreemerging.htm

Table D.1: Emerging Hot Spot Analysis Definitions

Consecutive Hot Spot	A location with a single uninterrupted run of statistically significant hot spot bins in the final time-step intervals. The location has never been a statistically significant hot spot prior to the final hot spot run and less than ninety percent of all bins are statistically significant hot spots.
Intensifying Hot Spot	A location that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of high counts in each time step is increasing overall and that increase is statistically significant.
Persistent Hot Spot	A location that has been a statistically significant hot spot for ninety percent of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering over time.
Diminishing Hot Spot	A location that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant.

Table D.1: Emerging Hot Spot Analysis Definitions

Sporadic Hot Spot	A location that is an on-again then off-again hot spot. Less than ninety percent of the time-step intervals have been statistically significant hot spots and none of the time-step intervals have been statistically significant cold spots.
Oscillating Hot Spot	A statistically significant hot spot for the final time-step interval that has a history of also being a statistically significant cold spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant hot spots.
Historical Hot Spot	The most recent time period is not hot, but at least ninety percent of the time-step intervals have been statistically significant hot spots.
New Cold Spot	A location that is a statistically significant cold spot for the final time step and has never been a statistically significant cold spot before.

Table D.1: Emerging Hot Spot Analysis Definitions

Consecutive Cold Spot	A location with a single uninterrupted run of statistically significant cold spot bins in the final time-step intervals. The location has never been a statistically significant cold spot prior to the final cold spot run and less than ninety percent of all bins are statistically significant cold spots.
Intensifying Cold Spot	A location that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of low counts in each time step is increasing overall and that increase is statistically significant.
Persistent Cold Spot	A location that has been a statistically significant cold spot for ninety percent of the time-step intervals with no discernible trend, indicating an increase or decrease in the intensity of clustering of counts over time.
Diminishing Cold Spot	A location that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of low counts in each time step is decreasing overall and that decrease is statistically significant.

Table D.1: Emerging Hot Spot Analysis Definitions

Sporadic Cold Spot	A location that is an on-again then off-again cold spot. Less than ninety percent of the time-step intervals have been statistically significant cold spots and none of the time-step intervals have been statistically significant hot spots.
Oscillating Cold Spot	A statistically significant cold spot for the final time-step interval that has a history of also being a statistically significant hot spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant cold spots.
Historical Cold Spot	The most recent time period is not cold, but at least ninety percent of the time-step intervals have been statistically significant cold spots.

Appendix E

LDA Gamma Table Counts

E.0.1 Tables

E.0.2 Maps

Topic	Q2 2013	Q3 2013	Q4 2013	Q1 2014	Q2 2014	Q3 2014	Q4 2014
1	23.28	25.19	26.70	22.95	21.42	20.61	24.40
2	524.14	498.34	531.56	521.16	523.32	529.29	540.23
3	25.33	24.56	25.92	32.59	30.87	29.15	40.45
4	100.68	103.79	127.26	135.15	141.69	149.83	171.77
5	36.42	39.78	44.60	49.04	56.87	57.24	80.39
6	9.78	8.92	9.20	10.78	10.60	10.55	13.38
7	54.09	52.64	50.46	52.10	55.12	56.98	54.73
8	33.46	31.23	33.02	32.55	32.72	35.74	39.68
9	141.96	151.32	165.15	163.62	156.84	143.17	162.96
10	61.79	61.33	65.02	67.02	68.78	72.84	79.48
11	62.61	66.60	75.55	77.50	80.50	82.38	95.93
12	16.50	16.65	18.58	19.40	19.52	19.89	25.48
13	21.13	18.48	23.50	24.51	27.70	29.80	34.57
14	38.39	37.11	36.74	42.38	41.08	45.31	49.11
15	68.72	69.16	82.88	86.14	87.46	90.61	86.85
16	314.44	306.63	304.31	278.40	270.54	222.80	195.87
17	28.82	29.63	33.21	32.24	32.65	32.64	41.71
18	201.16	201.06	207.73	211.25	200.58	190.57	203.84
19	116.77	117.30	129.18	129.30	125.47	129.02	144.79
20	225.17	234.10	258.97	261.05	261.07	262.57	297.43
21	152.19	140.96	155.17	153.09	147.40	150.96	158.72
22	55.67	58.09	63.41	65.36	65.55	62.08	60.02
23	3.58	3.98	3.26	4.58	4.49	4.43	5.02
24	148.99	148.63	179.08	184.07	186.73	194.47	233.87
25	57.60	54.75	63.14	63.35	60.36	61.09	67.52
26	172.40	191.17	201.08	192.10	193.55	196.29	195.40
27	29.79	32.62	46.62	48.24	51.02	54.32	75.91
28	38.87	42.97	46.61	40.85	44.96	50.41	62.17
29	126.87	132.25	145.09	152.00	156.54	158.16	186.48
30	51.39	50.32	47.21	52.13	51.82	50.47	57.45
31	52.70	49.37	58.75	53.43	51.97	59.30	79.55
32	171.44	171.29	192.07	188.81	200.29	198.54	219.59
33	32.34	33.35	38.56	41.29	42.78	42.60	55.69
34	4.54	4.44	5.40	6.58	6.73	7.88	12.54

Table E.1: Topics by Quarter, 2013-2014, All Investors

Topic	Q1 2015	Q2 2015	Q3 2015	Q4 2015	Q1 2016	Q2 2016	Q3 2016	Q4 2016
1	25.13	26.48	26.49	26.54	26.95	28.13	30.19	28.33
2	516.30	506.85	507.28	523.89	546.08	554.90	531.17	543.47
3	40.56	39.22	38.62	40.69	40.92	44.19	48.77	55.10
4	171.88	191.36	187.00	185.08	178.55	160.10	149.76	133.41
5	87.58	85.86	88.22	97.16	97.66	101.11	108.55	151.89
6	20.32	22.68	23.12	25.03	20.82	25.11	21.92	21.07
7	50.90	49.83	47.13	44.43	49.01	48.57	48.61	51.43
8	39.02	35.36	40.03	44.28	46.07	47.61	47.69	53.99
9	164.33	156.35	149.01	160.29	157.56	154.78	157.18	172.26
10	83.29	82.72	85.98	85.62	89.65	84.94	86.48	93.87
11	99.89	94.25	91.63	106.29	103.10	97.14	101.98	107.83
12	28.80	28.89	25.88	22.06	18.83	20.95	16.23	14.71
13	31.41	32.32	28.29	28.73	29.35	27.62	31.14	30.50
14	47.98	48.81	47.90	46.53	43.53	45.84	43.58	43.71
15	87.00	87.96	81.17	83.66	82.54	90.57	87.94	91.83
16	169.74	135.30	93.49	72.15	79.01	75.34	74.03	75.51
17	43.43	45.45	45.75	48.42	48.97	49.52	47.34	51.45
18	197.96	190.39	178.98	164.27	166.64	168.18	156.14	155.66
19	147.56	151.33	153.79	152.52	146.15	151.20	149.44	145.89
20	293.93	298.63	293.38	283.69	275.65	264.32	262.36	288.36
21	158.67	155.56	146.09	142.56	122.91	112.74	112.66	123.31
22	59.43	60.27	59.94	58.63	54.05	60.04	60.76	64.28
23	5.16	5.21	5.81	7.35	8.82	8.43	11.32	15.15
24	238.37	246.98	253.88	295.12	301.59	312.74	318.78	372.18
25	65.41	68.86	73.60	70.74	66.61	59.62	61.13	69.52
26	173.34	178.81	150.48	129.11	91.12	75.38	58.07	36.58
27	80.25	82.27	75.56	86.01	78.39	79.13	87.30	86.31
28	62.31	75.42	89.68	131.34	120.93	129.94	148.62	146.65
29	203.02	204.06	219.40	235.16	231.49	232.18	232.93	252.08
30	59.76	56.31	60.65	63.52	62.56	66.39	61.51	59.17
31	95.24	108.48	156.89	228.83	240.11	222.59	233.94	228.07
32	210.18	205.96	202.40	201.25	212.73	219.33	215.00	231.19
33	51.17	41.58	46.17	48.70	41.19	35.86	39.59	42.21
34	16.67	19.18	28.32	41.33	50.48	66.51	82.86	98.05

Table E.2: Topics by Quarter, 2015-2016, All Investors

Topic	Q1 2017	Q2 2017	Q3 2017	Q4 2017	Q1 2018	Q2 2018	Q3 2018	Q4 2018
1	27.30	25.32	24.35	26.40	23.24	28.09	27.07	26.59
2	516.69	490.06	474.48	525.98	480.04	464.32	471.88	508.19
3	55.84	59.95	62.50	73.66	74.86	81.66	83.73	94.92
4	121.44	96.19	82.03	80.20	71.90	66.18	66.38	69.51
5	176.62	222.77	238.15	259.60	267.49	260.67	260.97	256.51
6	21.96	23.24	22.44	21.42	22.02	21.59	22.04	24.72
7	55.38	53.28	55.31	59.92	55.16	55.66	53.22	57.53
8	55.81	56.19	42.63	42.17	39.32	38.37	35.12	35.07
9	167.90	173.77	174.70	182.46	182.08	162.74	150.53	152.56
10	96.29	95.63	96.78	100.97	99.12	95.22	97.28	108.95
11	114.22	115.64	116.82	127.19	128.91	120.49	115.29	122.39
12	15.12	13.60	12.59	14.24	13.35	13.80	13.21	17.49
13	30.80	29.40	30.85	33.59	32.89	29.60	33.17	39.17
14	42.89	42.69	39.30	39.86	37.75	31.90	32.40	34.15
15	88.69	85.27	83.70	82.24	77.72	76.29	78.18	79.92
16	69.98	59.84	58.90	60.33	58.14	56.98	53.38	55.93
17	47.48	46.09	49.04	55.66	55.90	56.49	53.74	57.51
18	145.68	139.61	124.85	130.97	125.08	111.38	112.29	114.10
19	143.82	144.71	143.91	155.47	154.01	157.54	154.36	161.59
20	265.34	252.92	248.15	251.82	235.83	226.80	220.00	227.90
21	111.46	107.90	101.63	104.71	101.60	90.67	89.19	83.76
22	62.09	57.84	57.89	61.54	57.80	52.81	51.35	56.24
23	16.36	20.22	55.64	86.46	125.99	152.78	172.37	198.21
24	384.19	398.24	404.35	483.80	477.37	455.02	455.27	528.56
25	68.77	64.23	61.75	66.34	60.22	61.59	55.23	59.91
26	30.51	26.37	17.83	19.98	18.43	14.82	13.49	14.15
27	88.36	90.44	94.44	98.13	97.82	96.57	98.16	109.14
28	164.97	179.88	183.62	207.48	221.05	244.46	236.75	235.23
29	249.24	247.97	247.60	292.55	311.47	359.03	369.76	449.72
30	58.49	57.42	55.08	51.36	54.26	56.41	50.57	60.80
31	239.51	245.86	236.46	253.31	247.29	251.41	260.68	299.54
32	215.92	202.15	205.74	231.47	239.90	249.66	241.78	253.67
33	40.13	42.39	42.95	46.24	44.00	40.19	42.33	44.43
34	118.76	127.90	139.54	151.45	154.00	148.79	144.84	135.91

Table E.3: Topics by Quarter, 2017-2018, All Investors

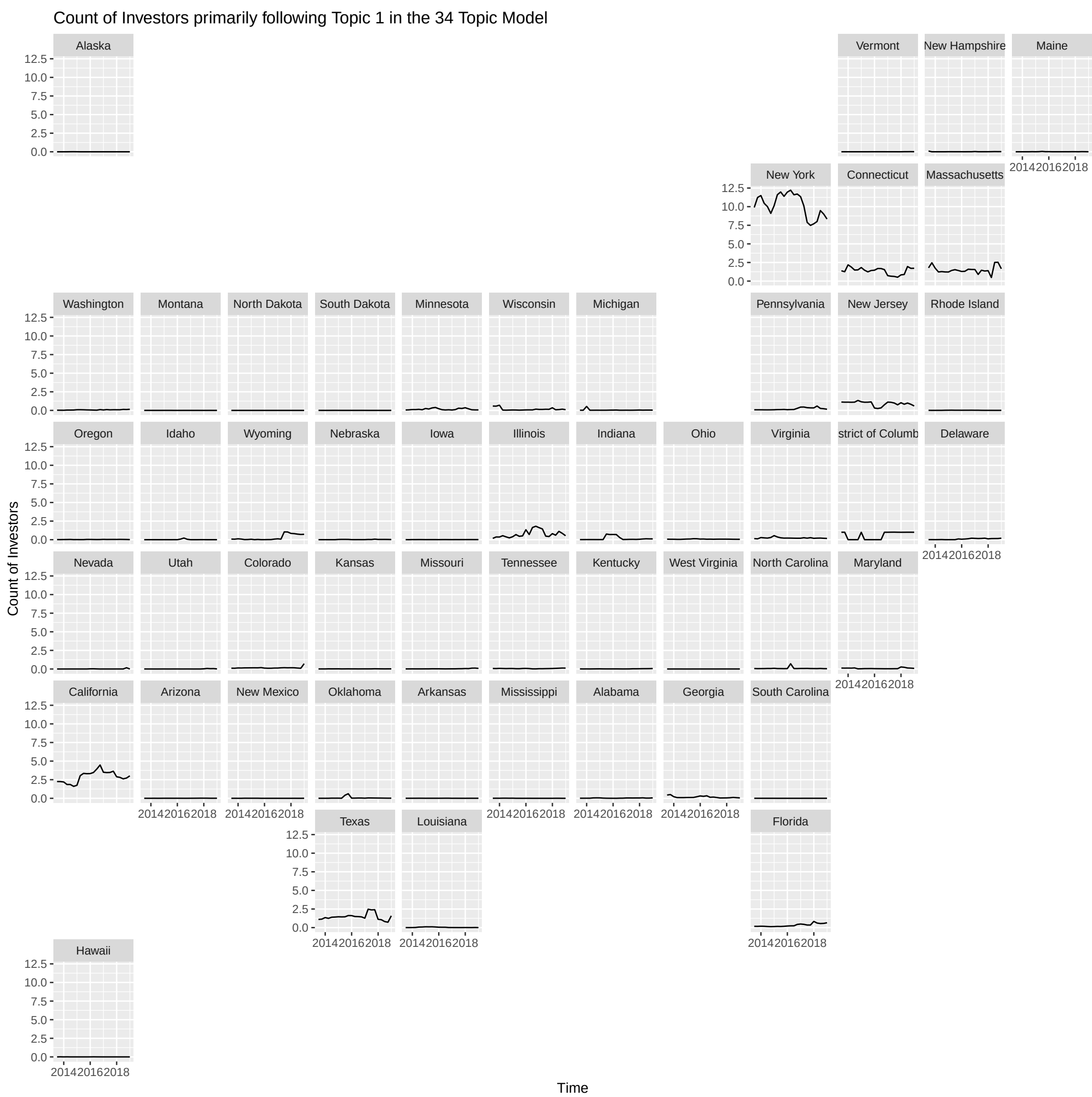


Figure E.1: Count of firms by highest likely topic in the 34 topic LDA for Topic 1

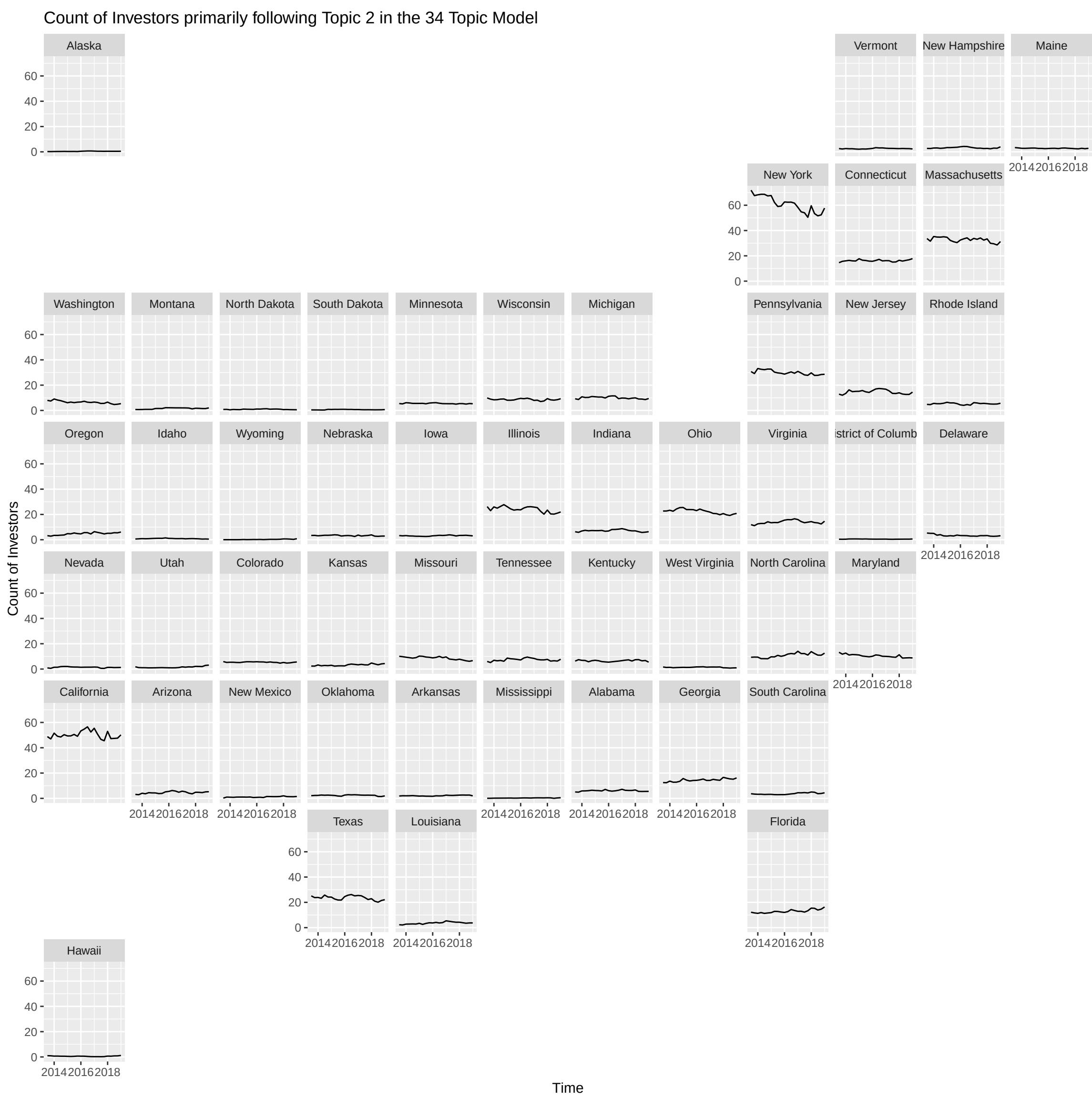


Figure E.2: Count of firms by highest likely topic in the 34 topic LDA for Topic 2

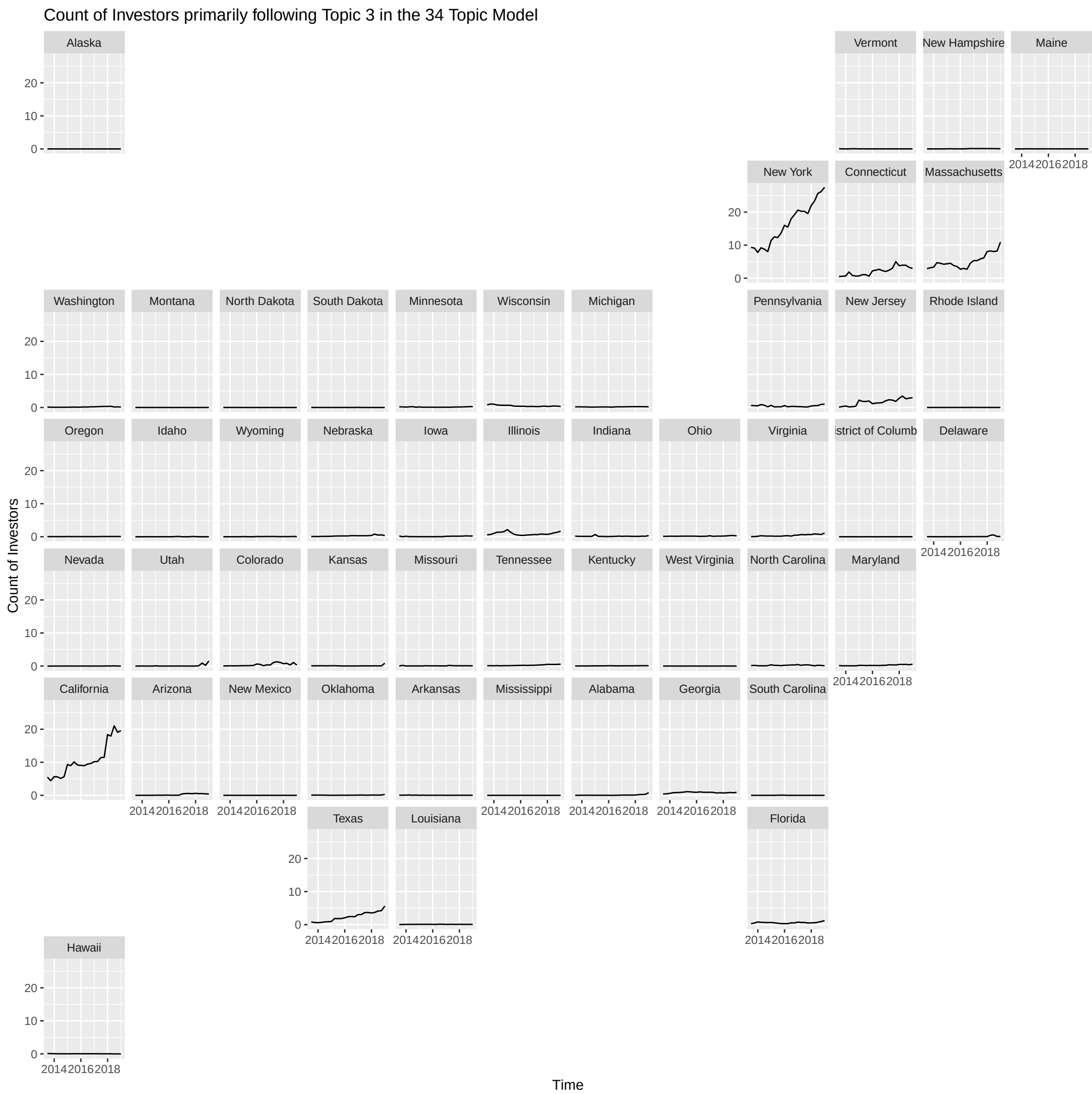


Figure E.3: Count of firms by highest likely topic in the 34 topic LDA for Topic 3

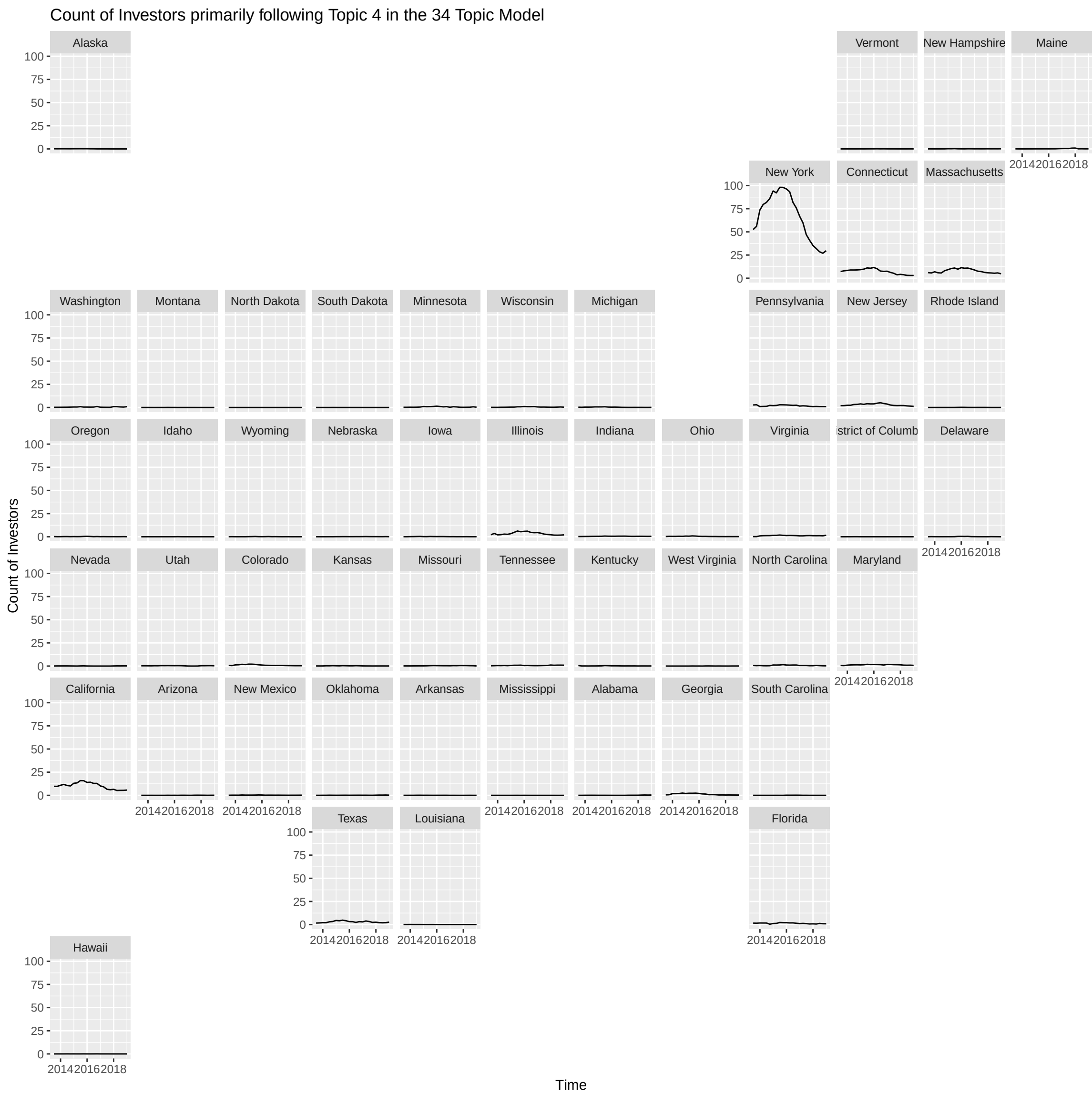


Figure E.4: Count of firms by highest likely topic in the 34 topic LDA for Topic 4

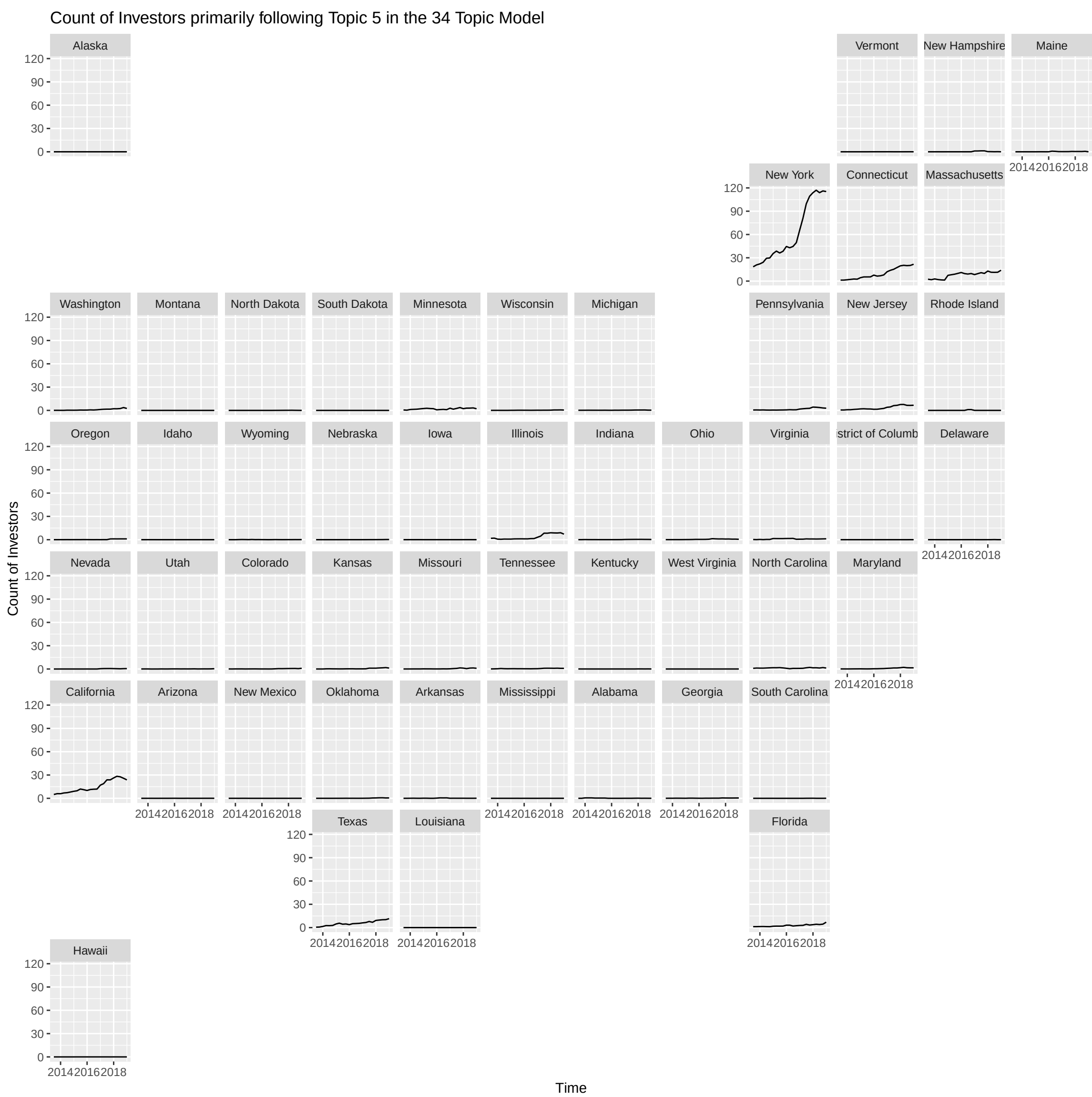


Figure E.5: Count of firms by highest likely topic in the 34 topic LDA for Topic 5

Figure E.6: Count of firms by highest likely topic in the 34 topic LDA for Topic 6

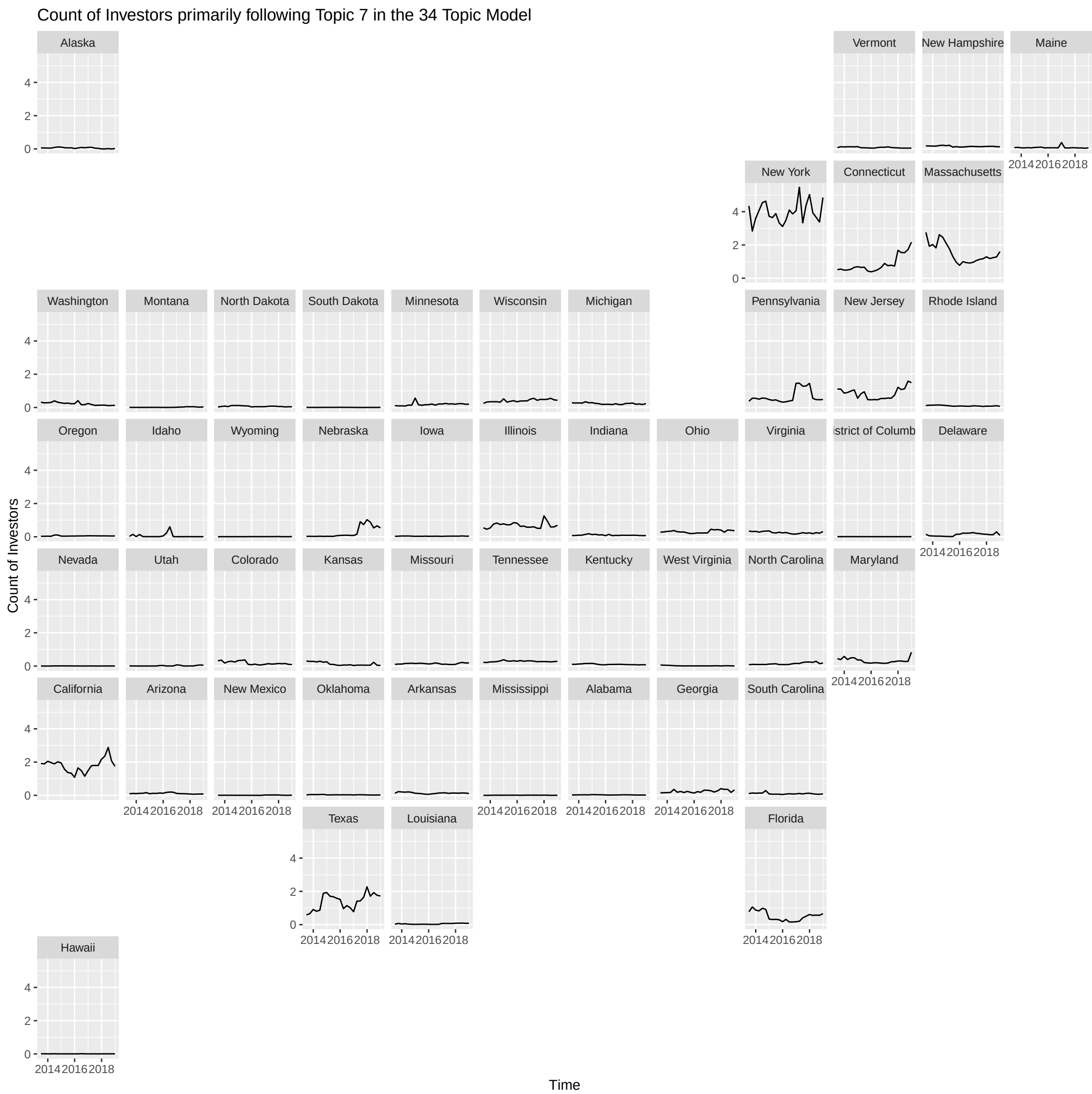


Figure E.7: Count of firms by highest likely topic in the 34 topic LDA for Topic 7

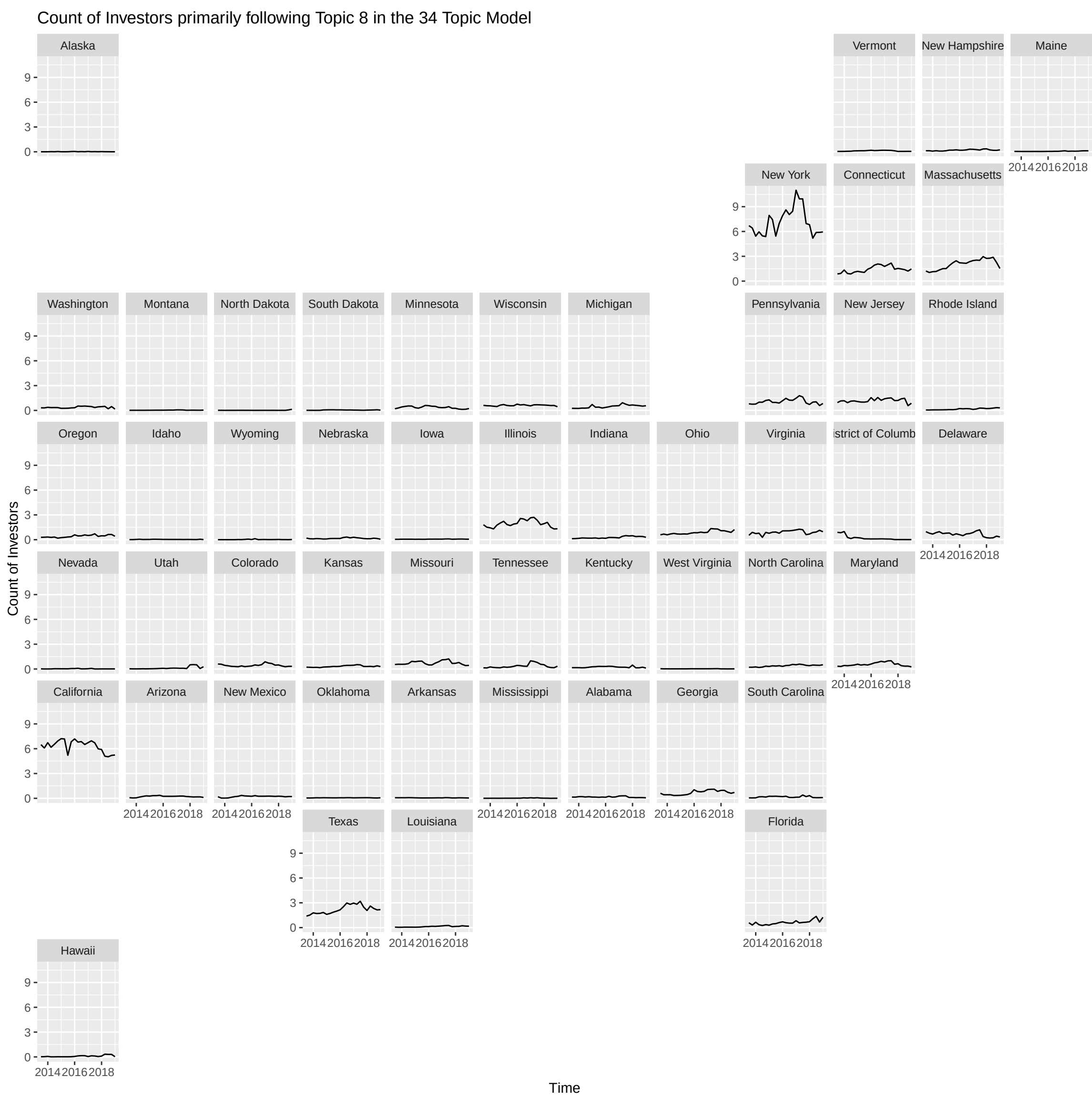


Figure E.8: Count of firms by highest likely topic in the 34 topic LDA for Topic 8

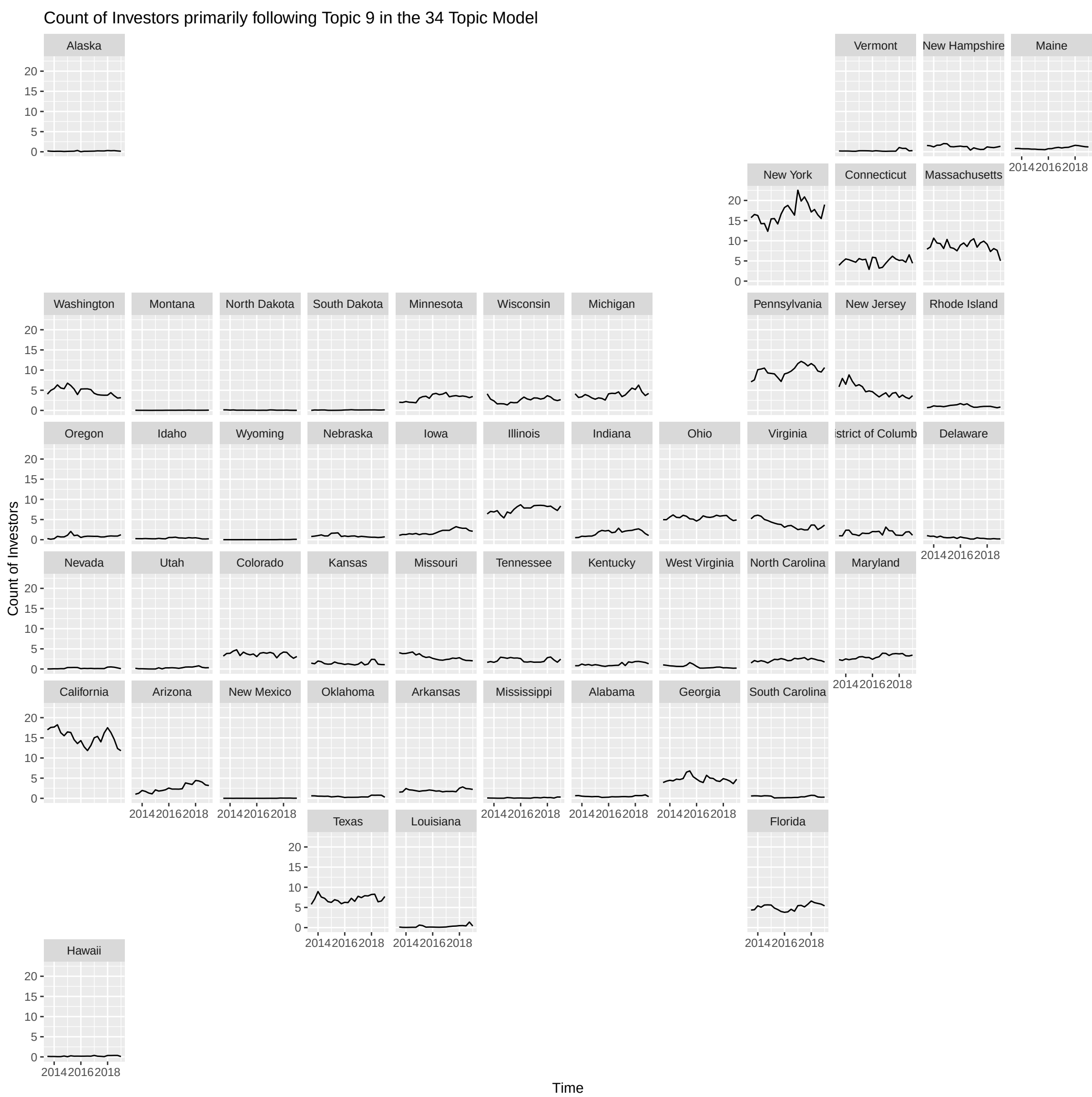


Figure E.9: Count of firms by highest likely topic in the 34 topic LDA for Topic 9

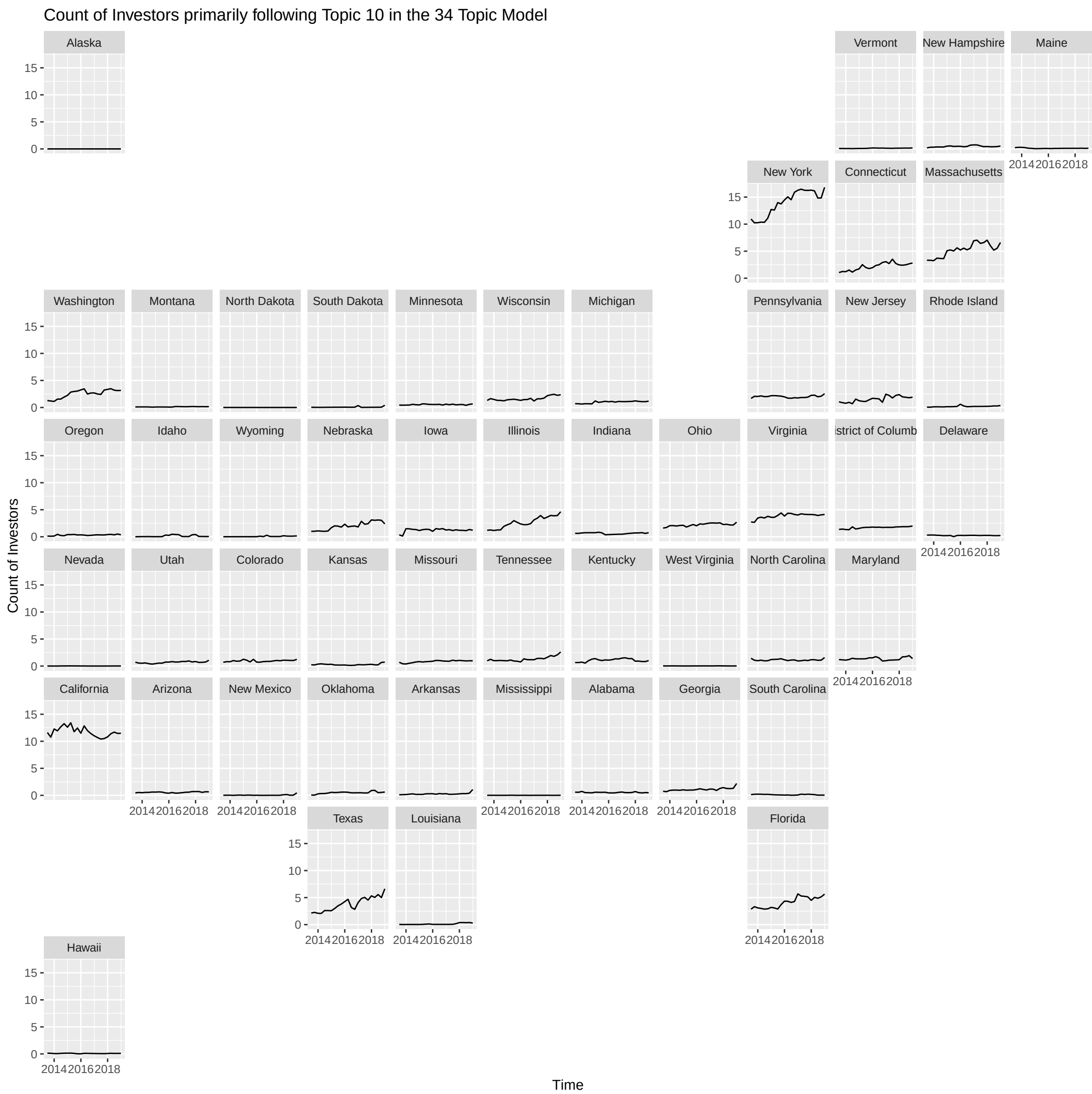


Figure E.10: Count of firms by highest likely topic in the 34 topic LDA for Topic 10

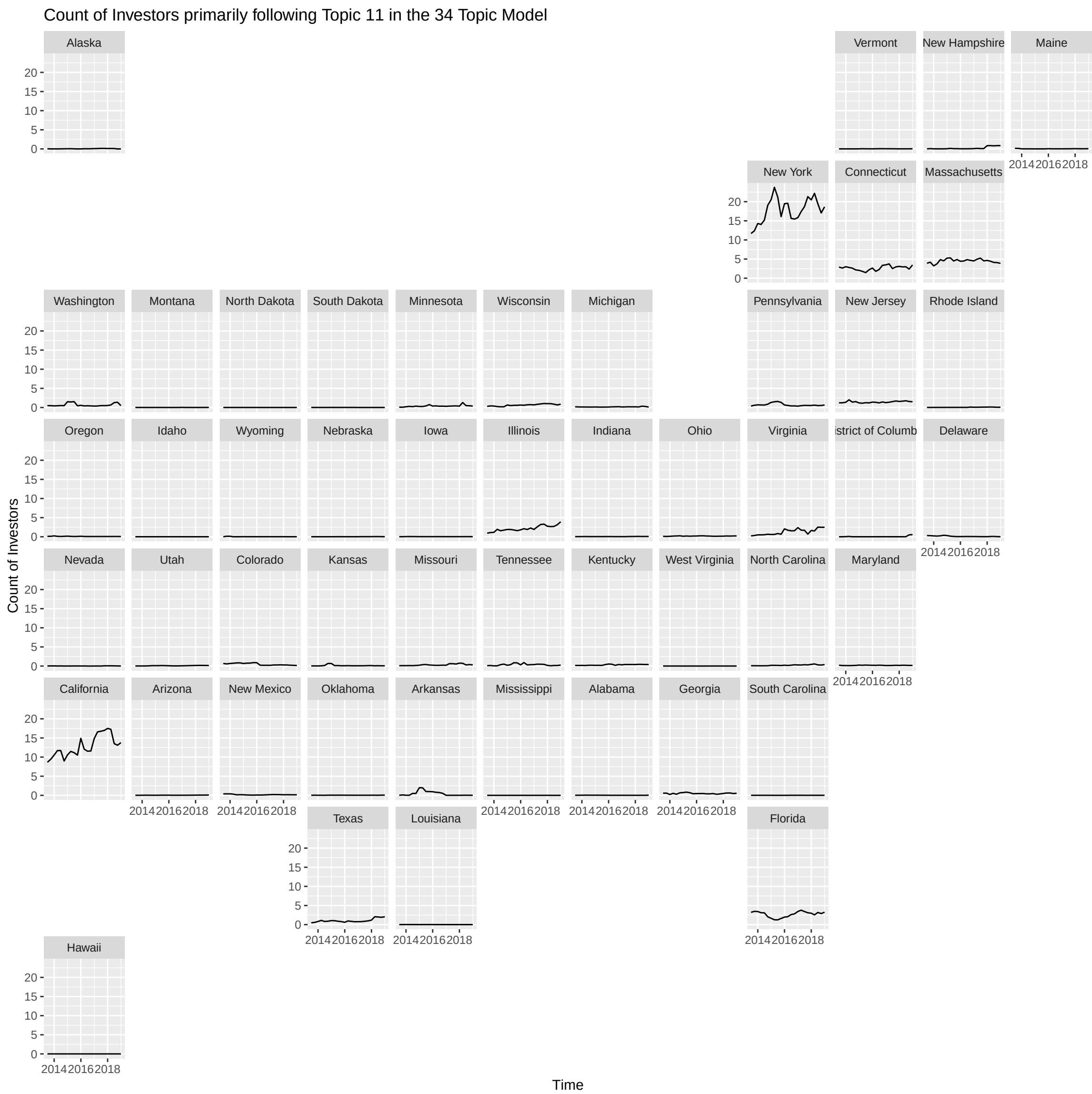


Figure E.11: Count of firms by highest likely topic in the 34 topic LDA for Topic 11

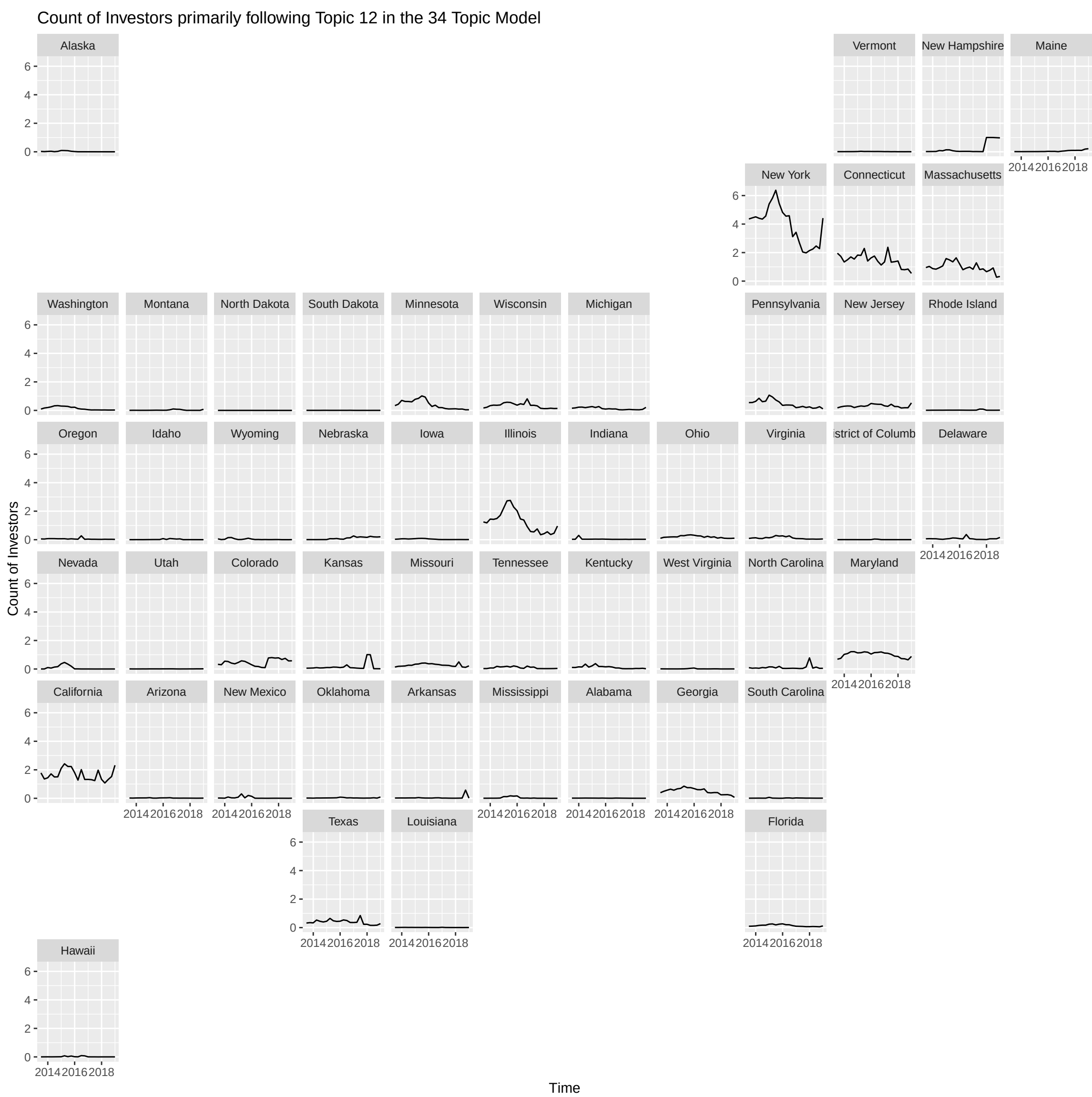


Figure E.12: Count of firms by highest likely topic in the 34 topic LDA for Topic 12

Figure E.13: Count of firms by highest likely topic in the 34 topic LDA for Topic 13

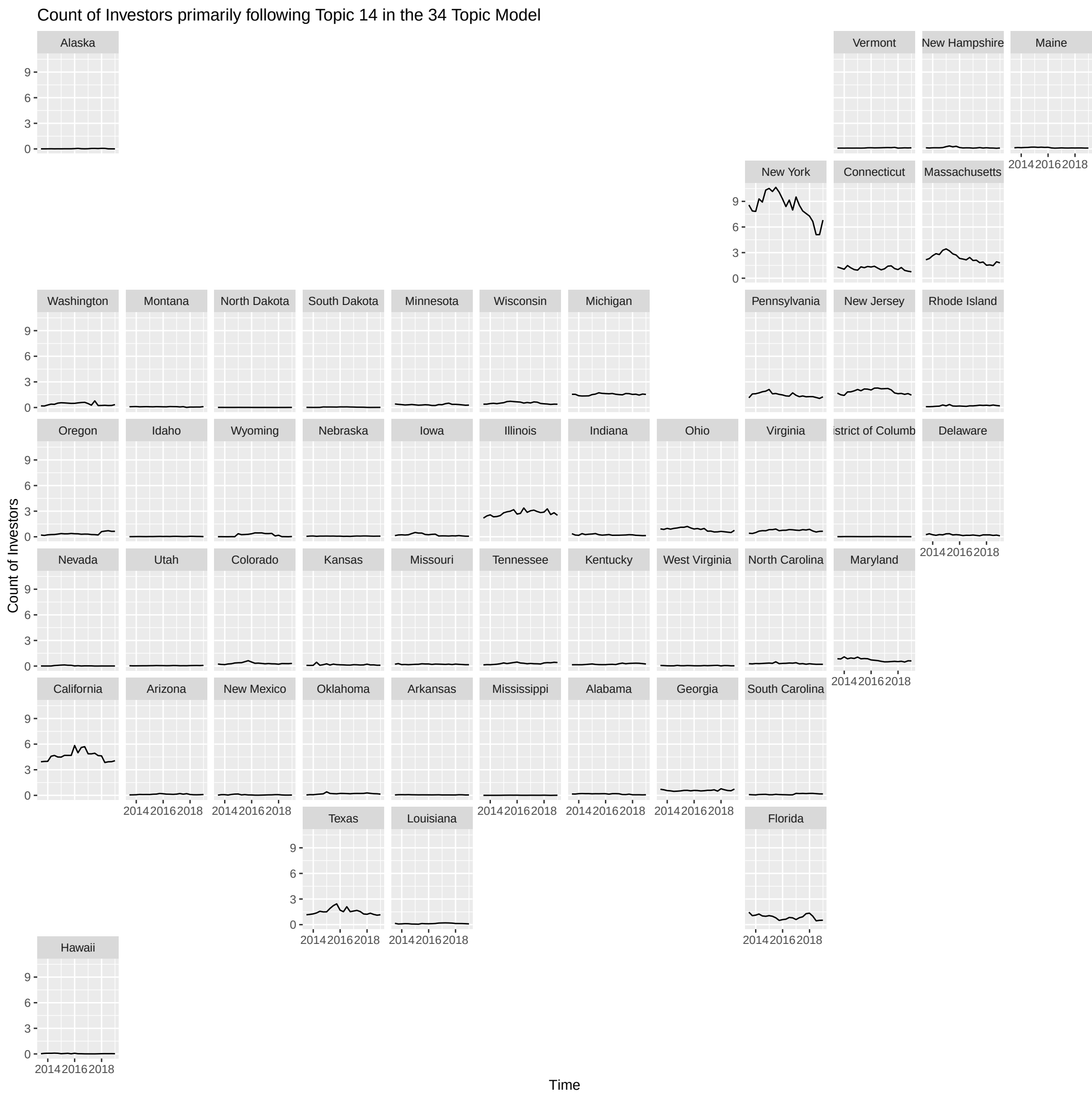


Figure E.14: Count of firms by highest likely topic in the 34 topic LDA for Topic 14

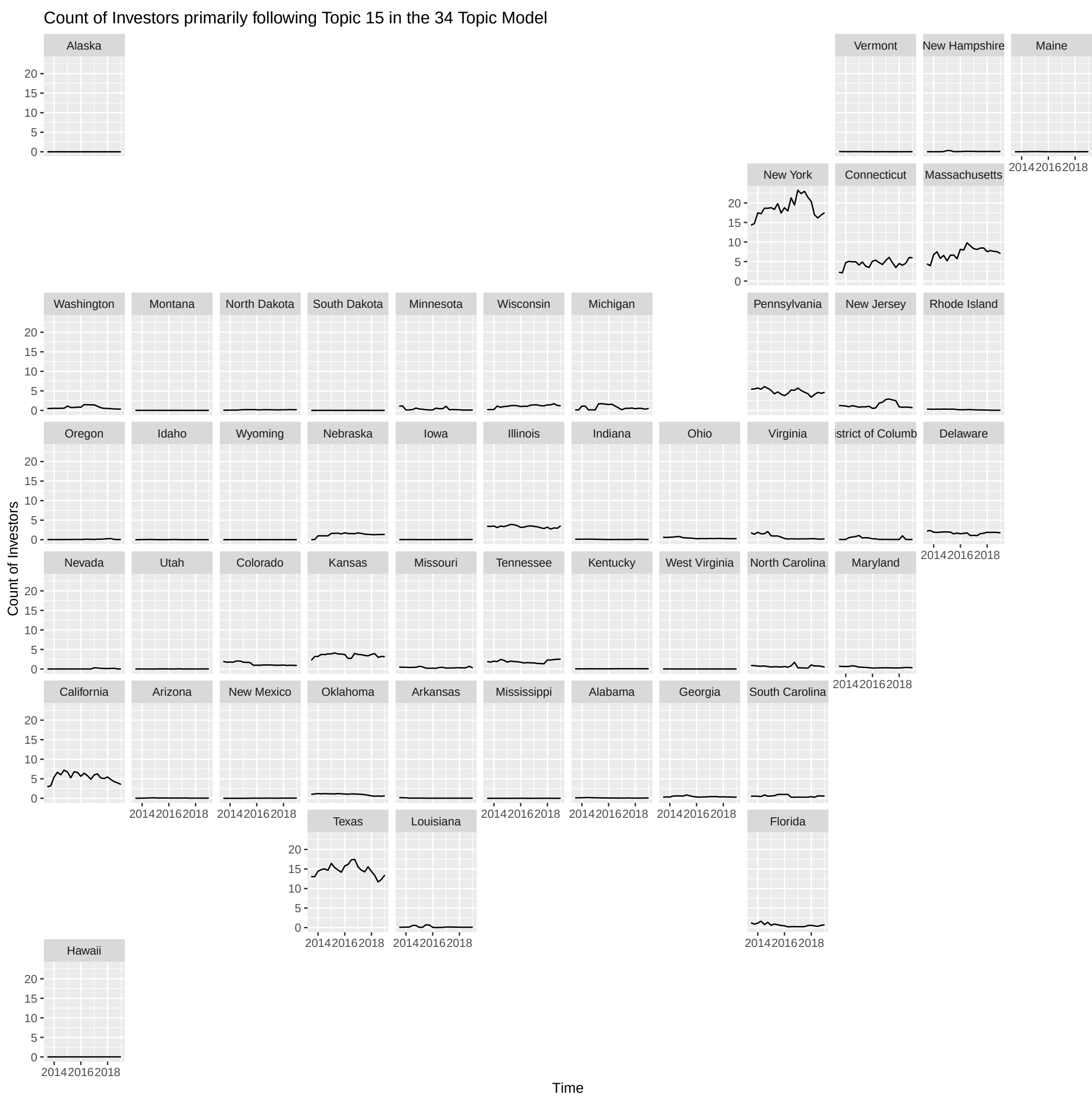


Figure E.15: Count of firms by highest likely topic in the 34 topic LDA for Topic 15

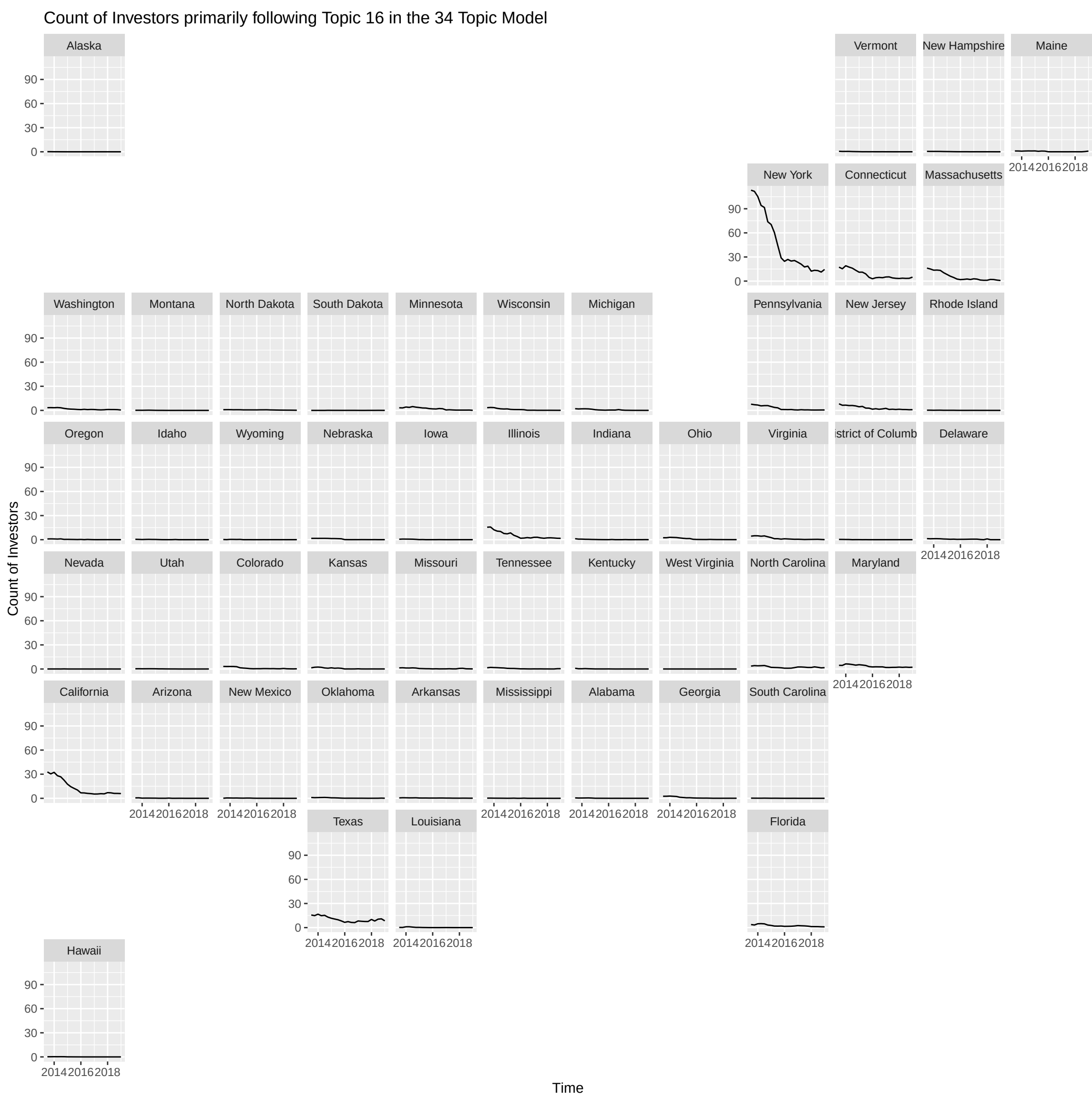


Figure E.16: Count of firms by highest likely topic in the 34 topic LDA for Topic 16

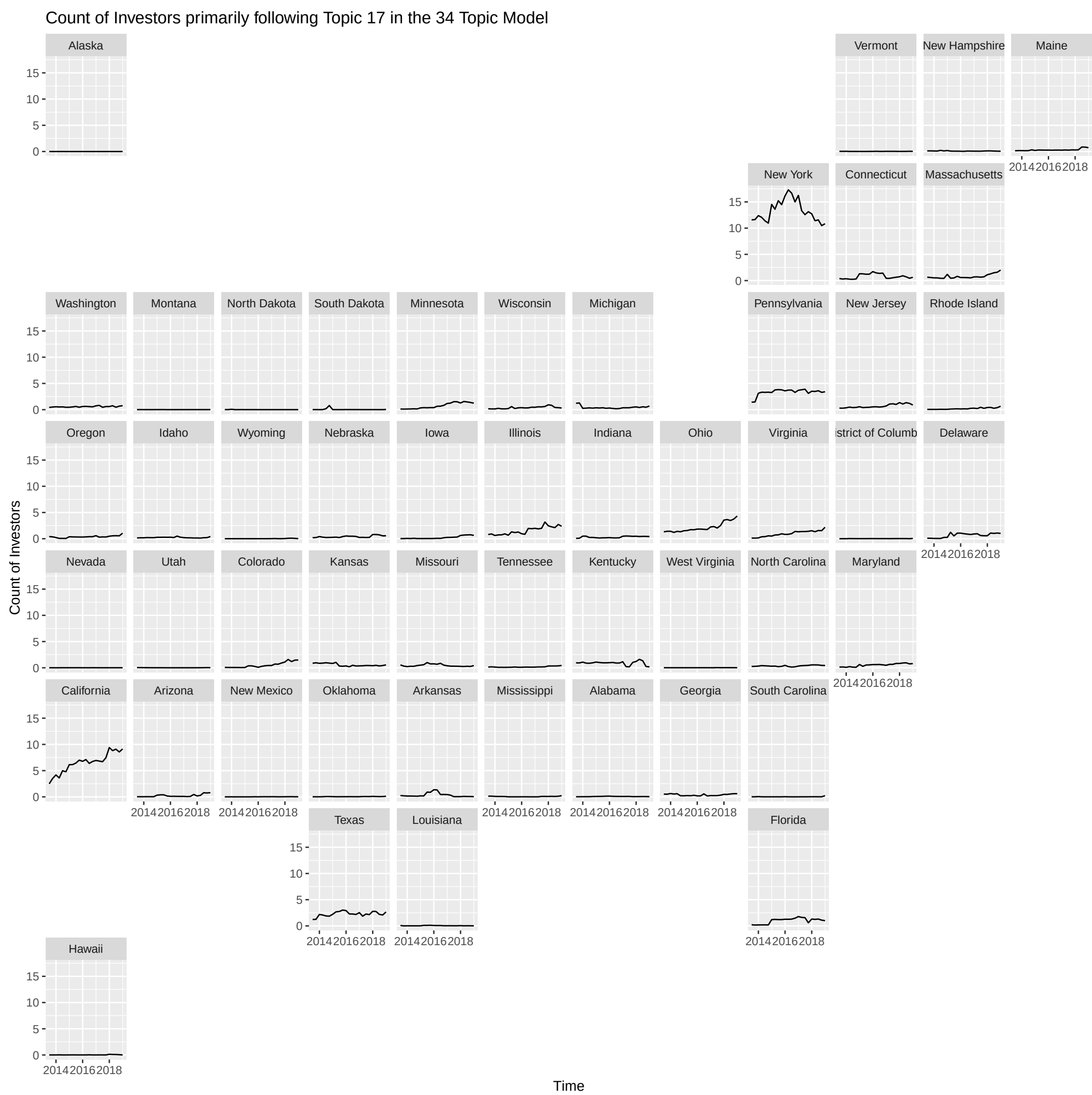


Figure E.17: Count of firms by highest likely topic in the 34 topic LDA for Topic 17

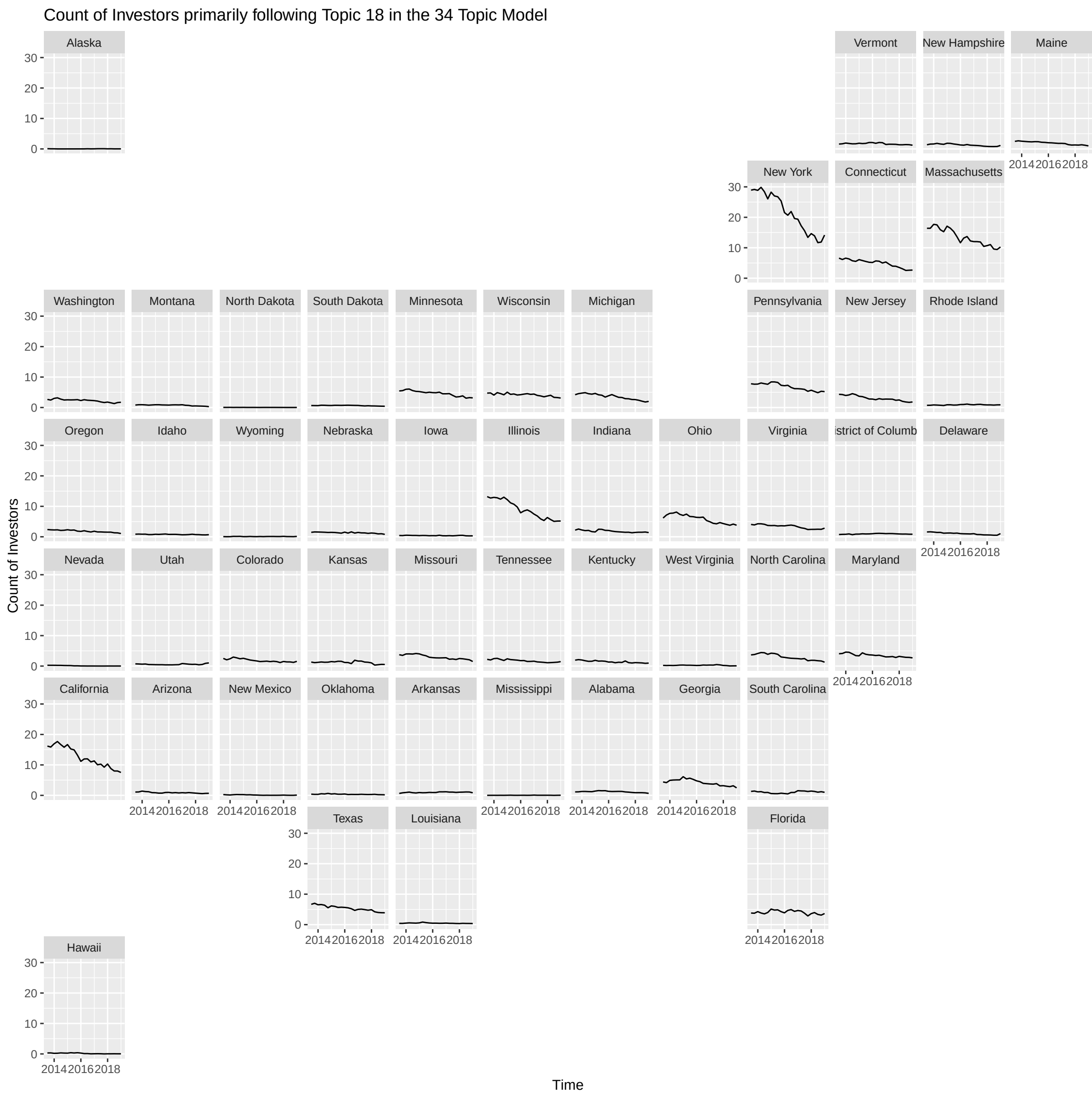


Figure E.18: Count of firms by highest likely topic in the 34 topic LDA for Topic 18

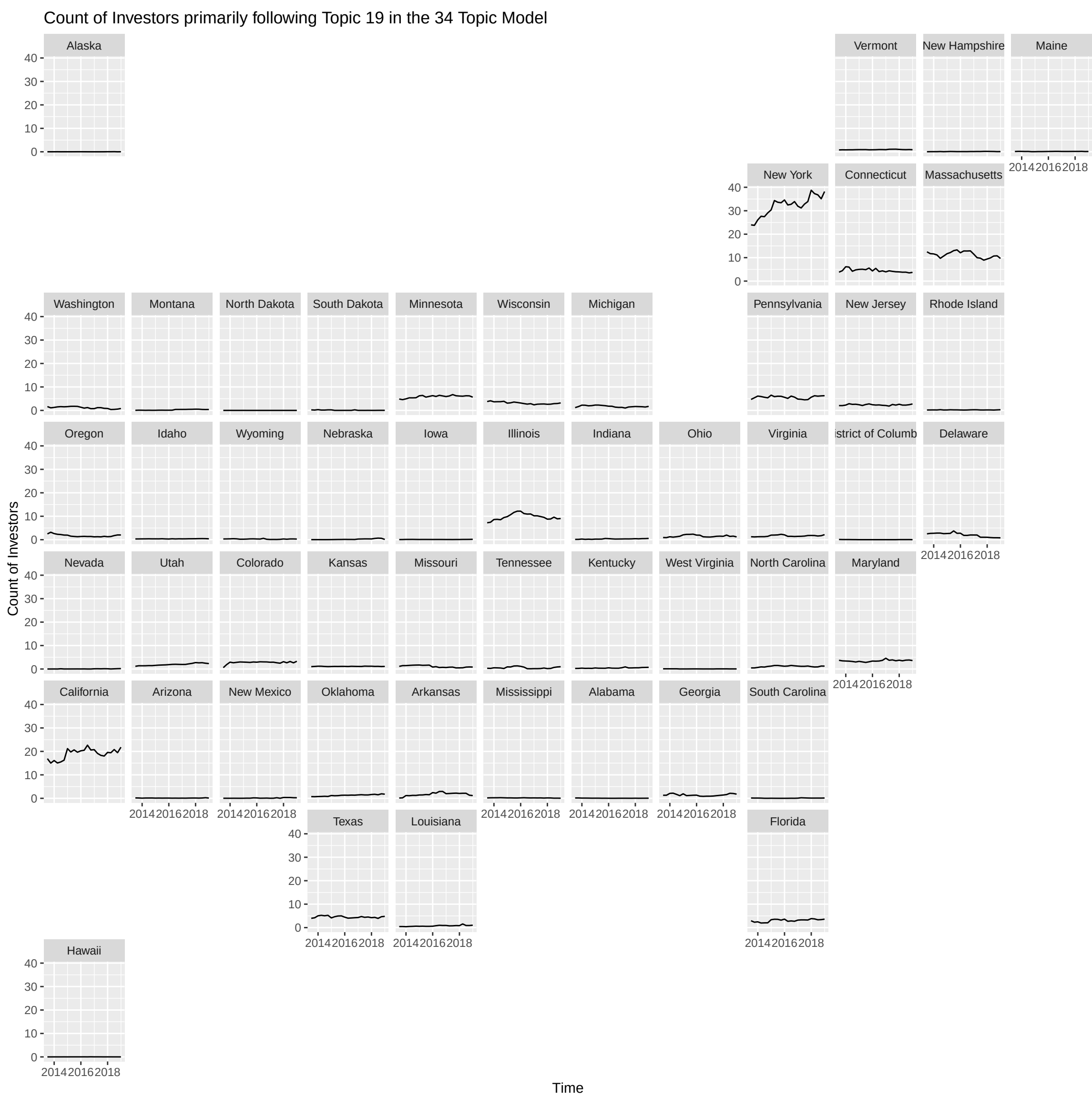


Figure E.19: Count of firms by highest likely topic in the 34 topic LDA for Topic 19

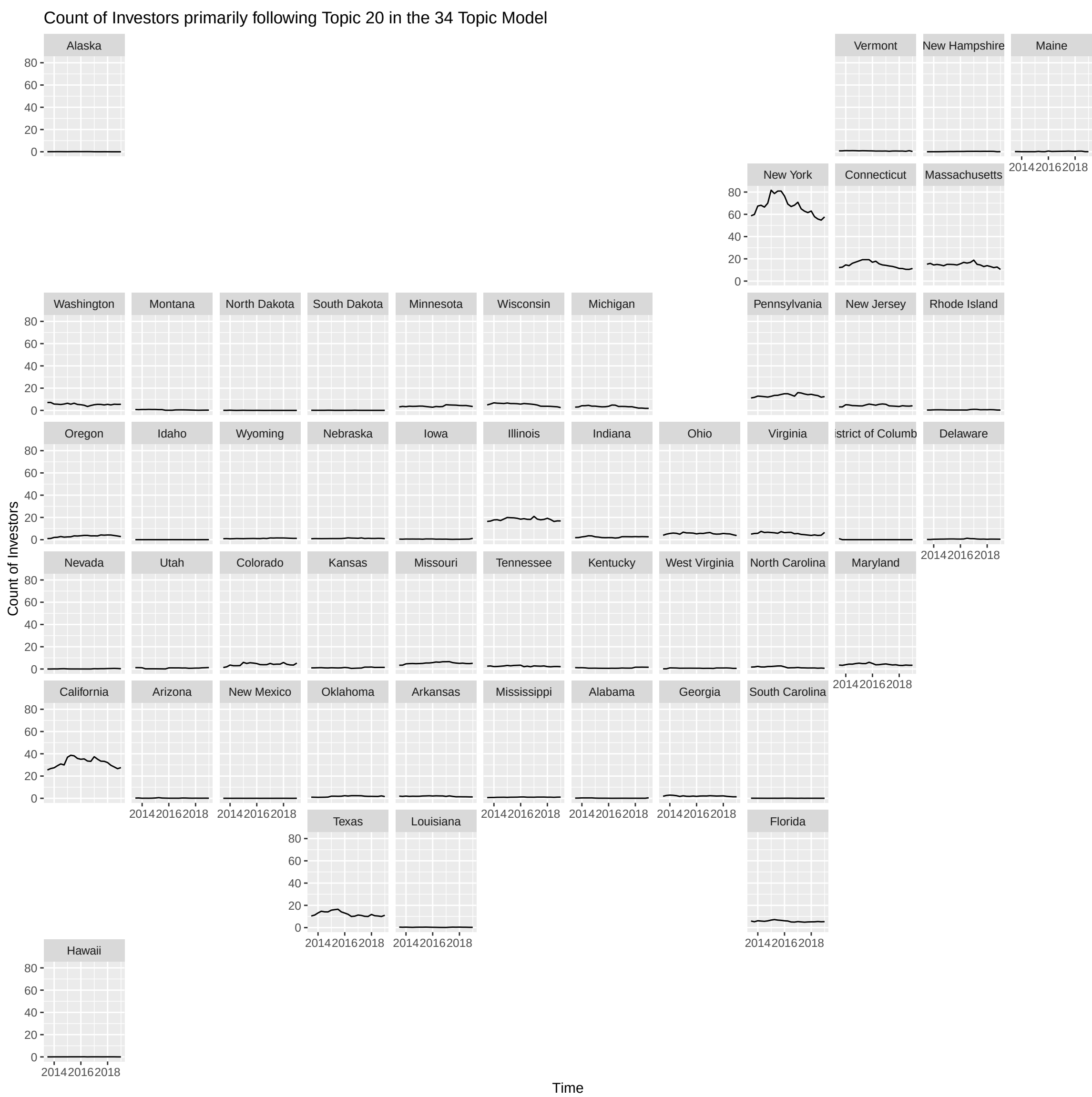


Figure E.20: Count of firms by highest likely topic in the 34 topic LDA for Topic 20

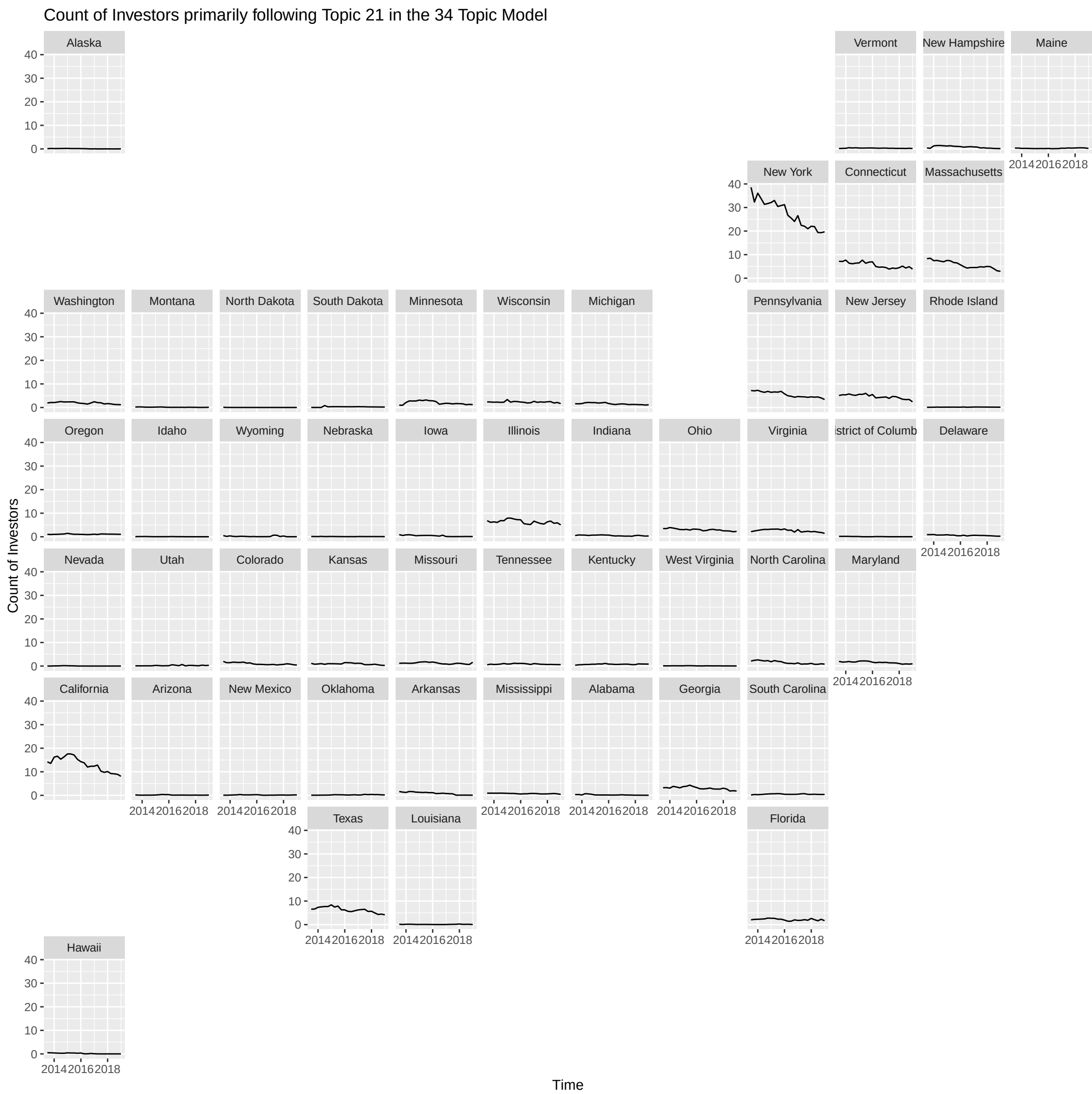


Figure E.21: Count of firms by highest likely topic in the 34 topic LDA for Topic 21

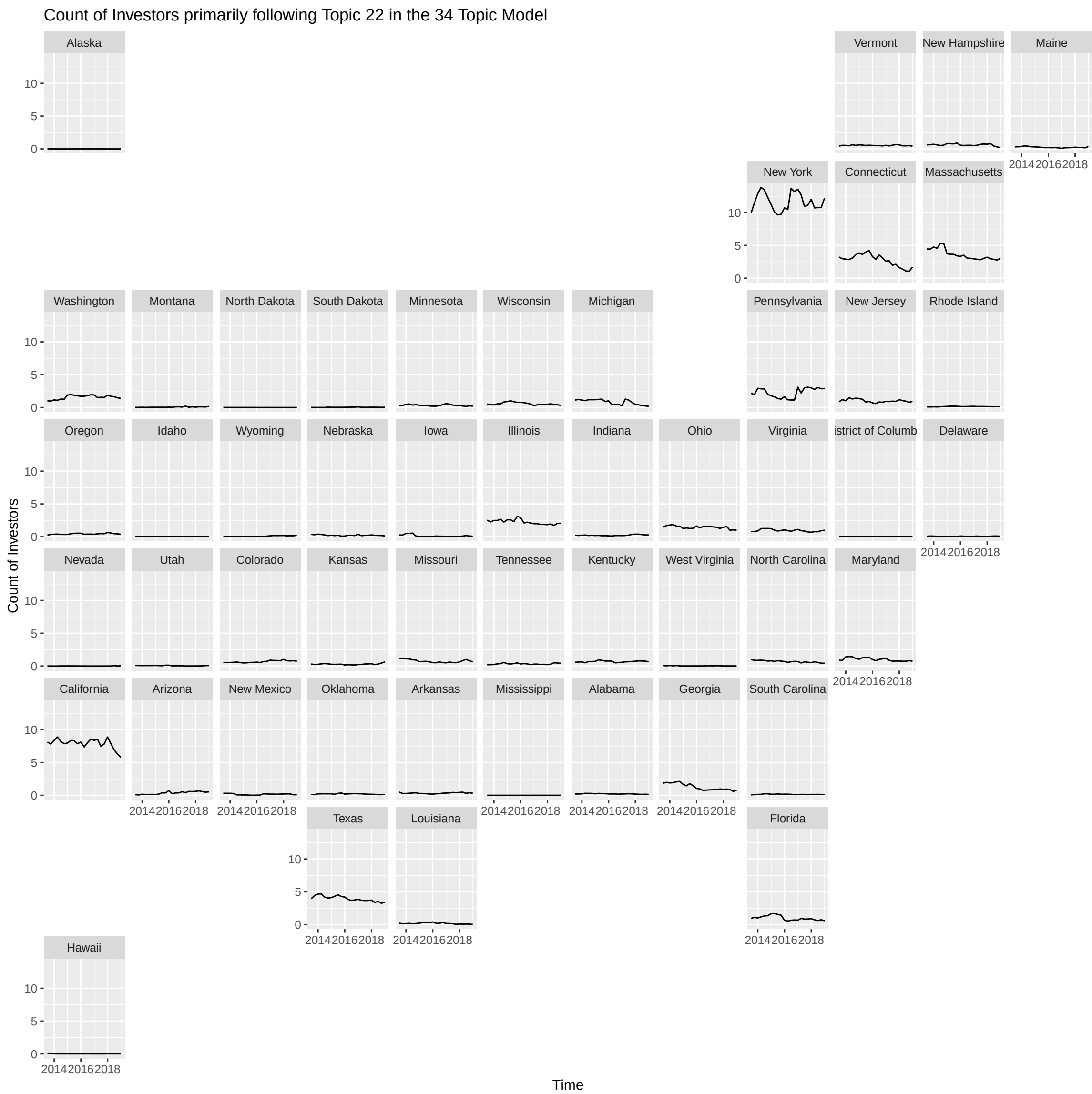


Figure E.22: Count of firms by highest likely topic in the 34 topic LDA for Topic 22

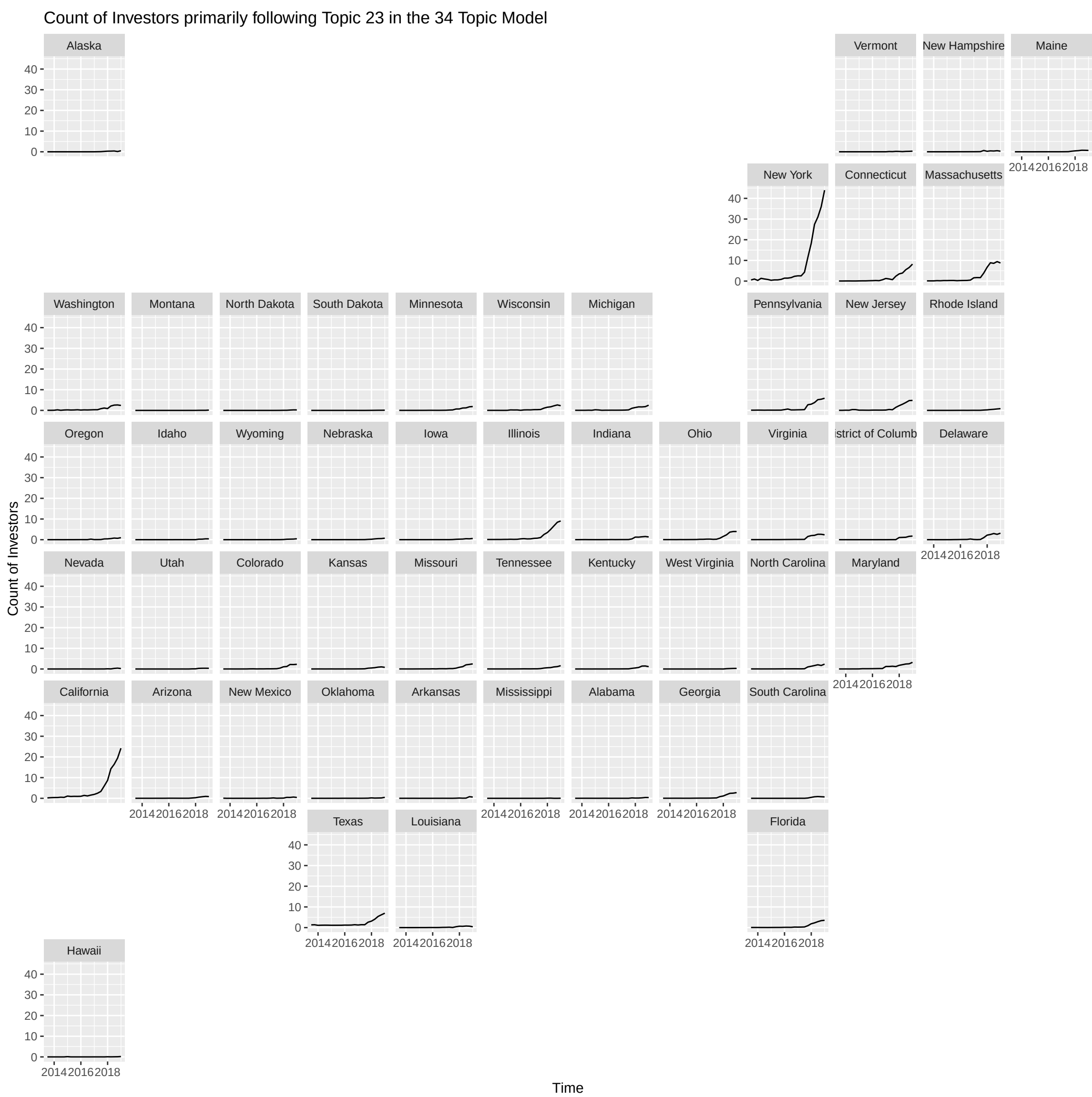


Figure E.23: Count of firms by highest likely topic in the 34 topic LDA for Topic 23

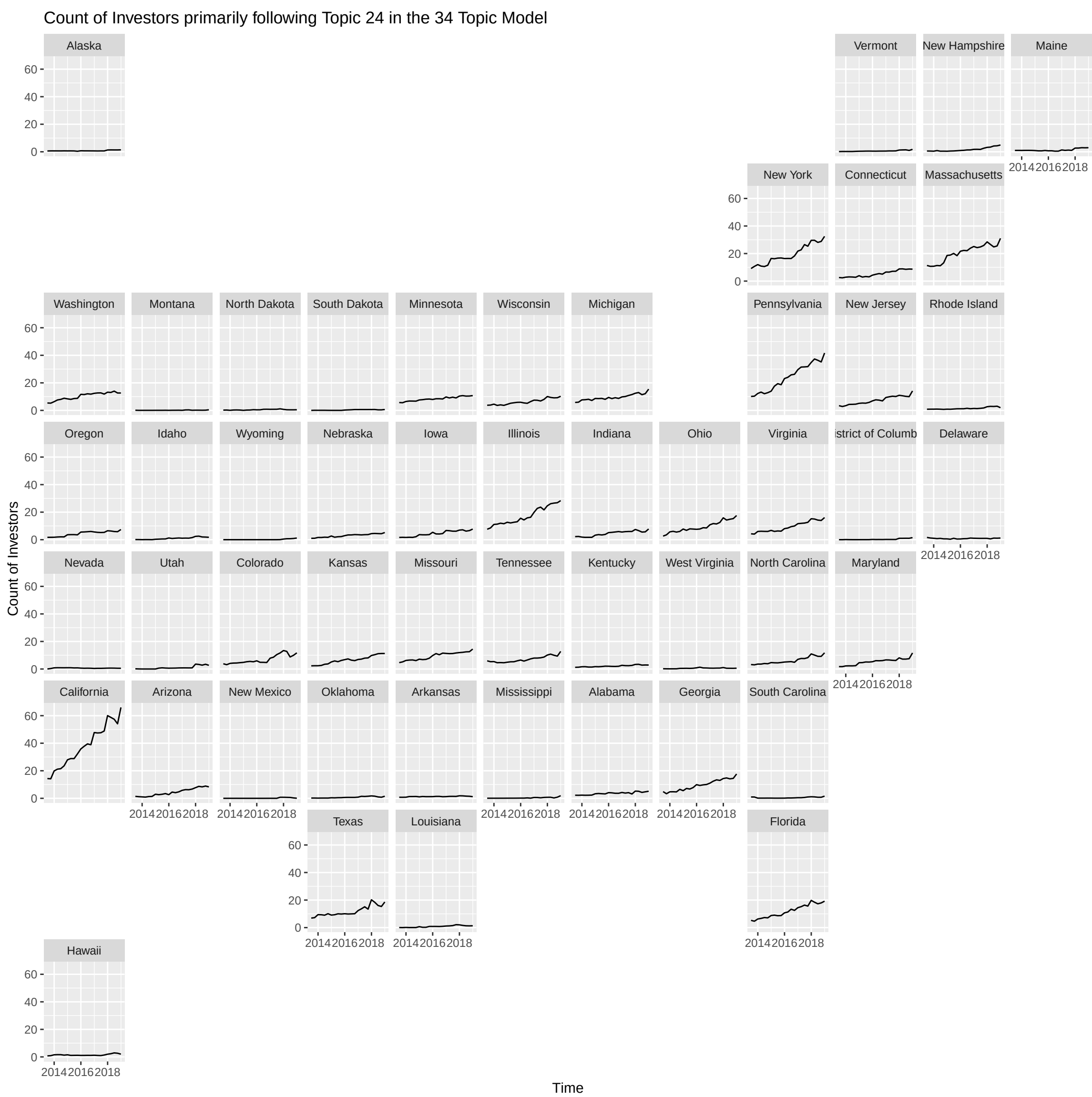


Figure E.24: Count of firms by highest likely topic in the 34 topic LDA for Topic 24

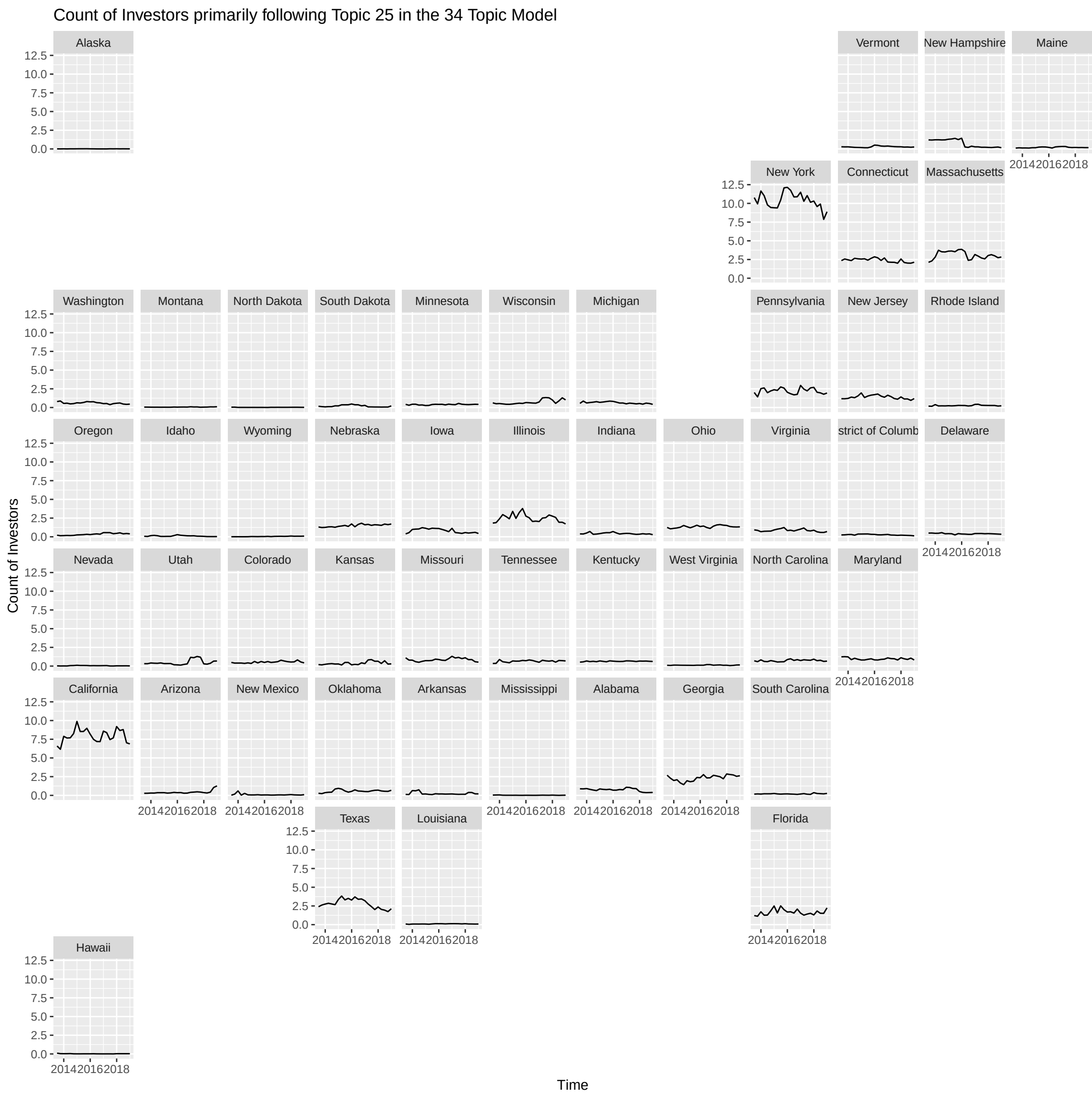


Figure E.25: Count of firms by highest likely topic in the 34 topic LDA for Topic 25

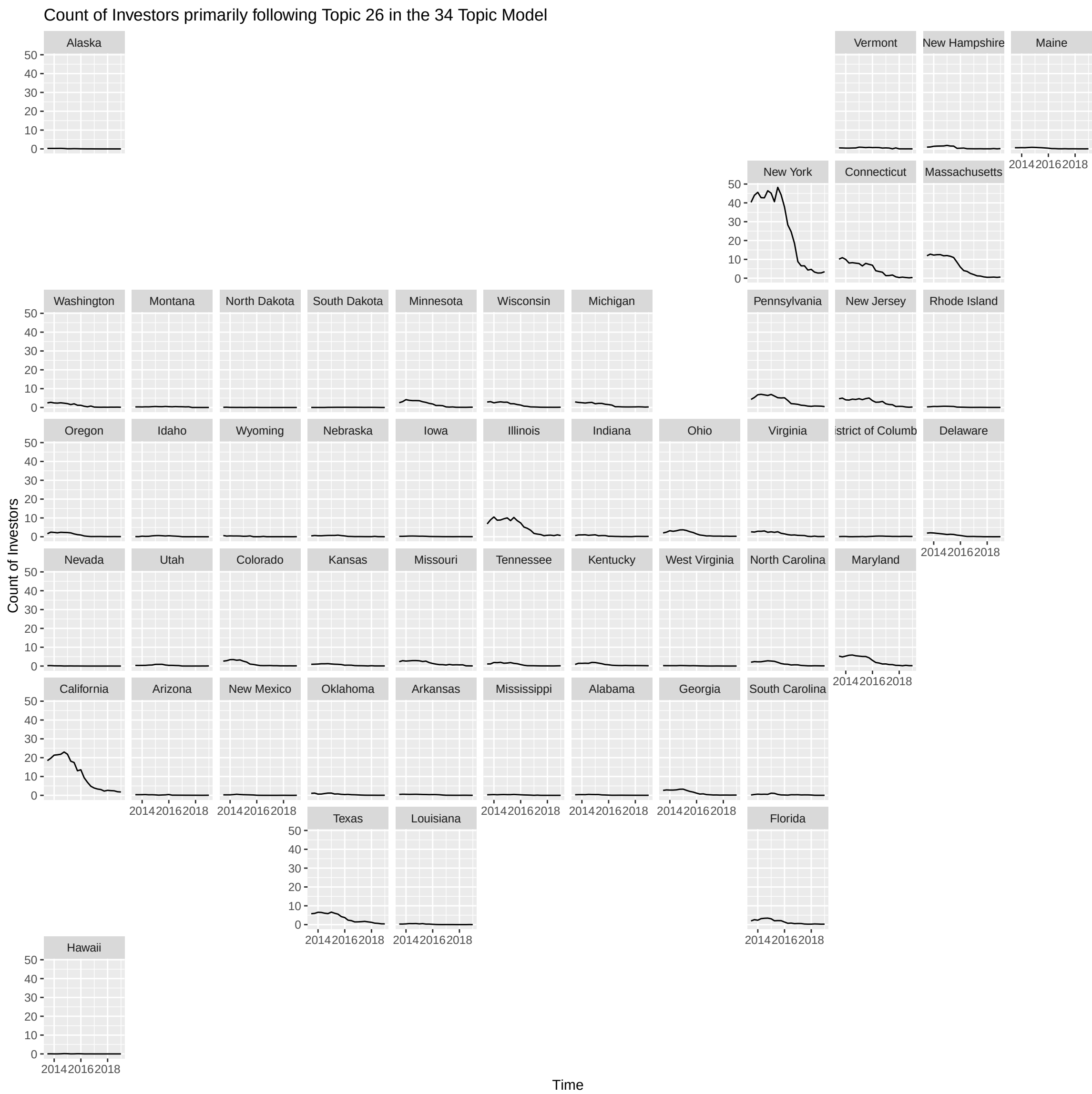


Figure E.26: Count of firms by highest likely topic in the 34 topic LDA for Topic 26

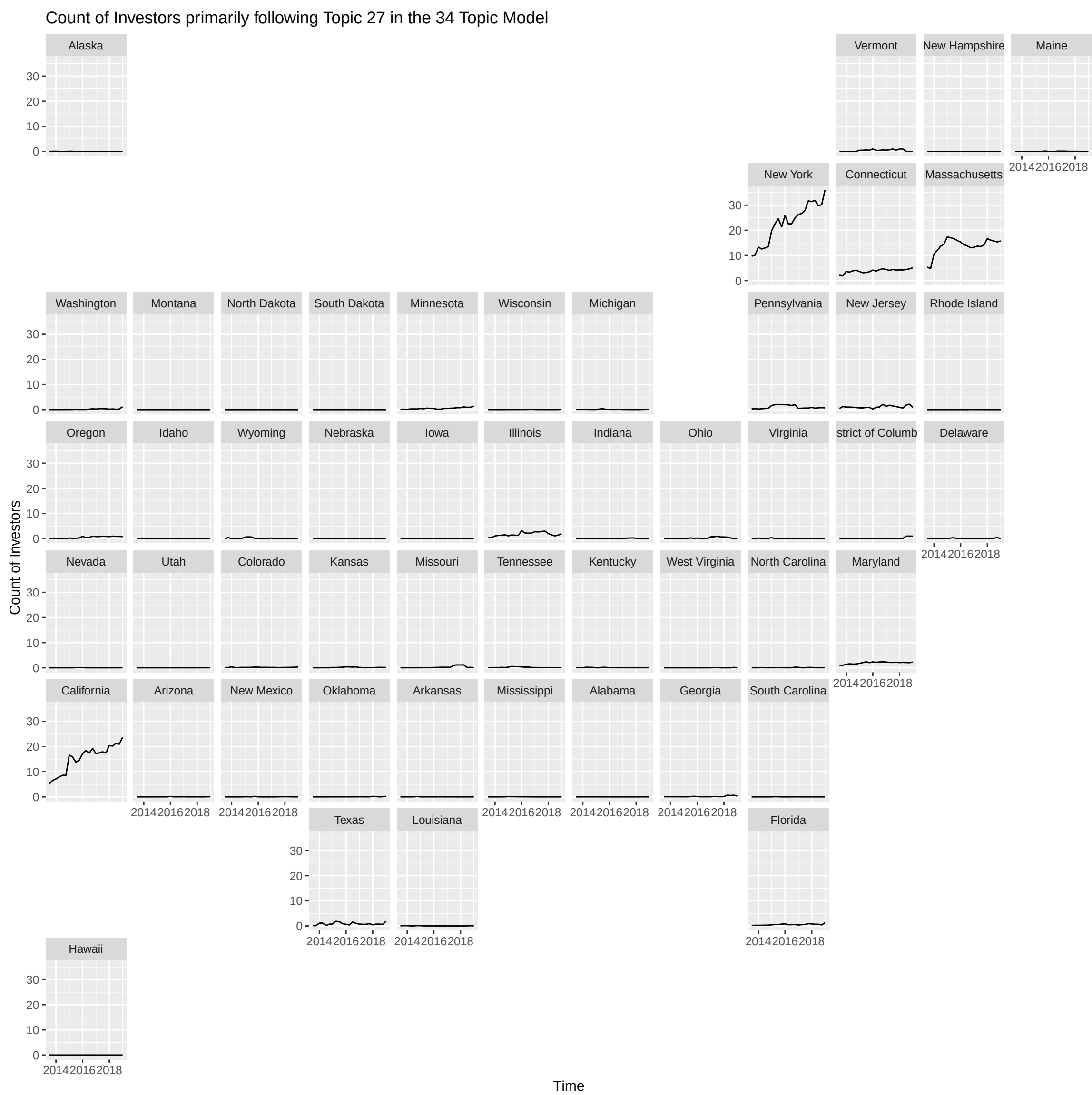


Figure E.27: Count of firms by highest likely topic in the 34 topic LDA for Topic 27

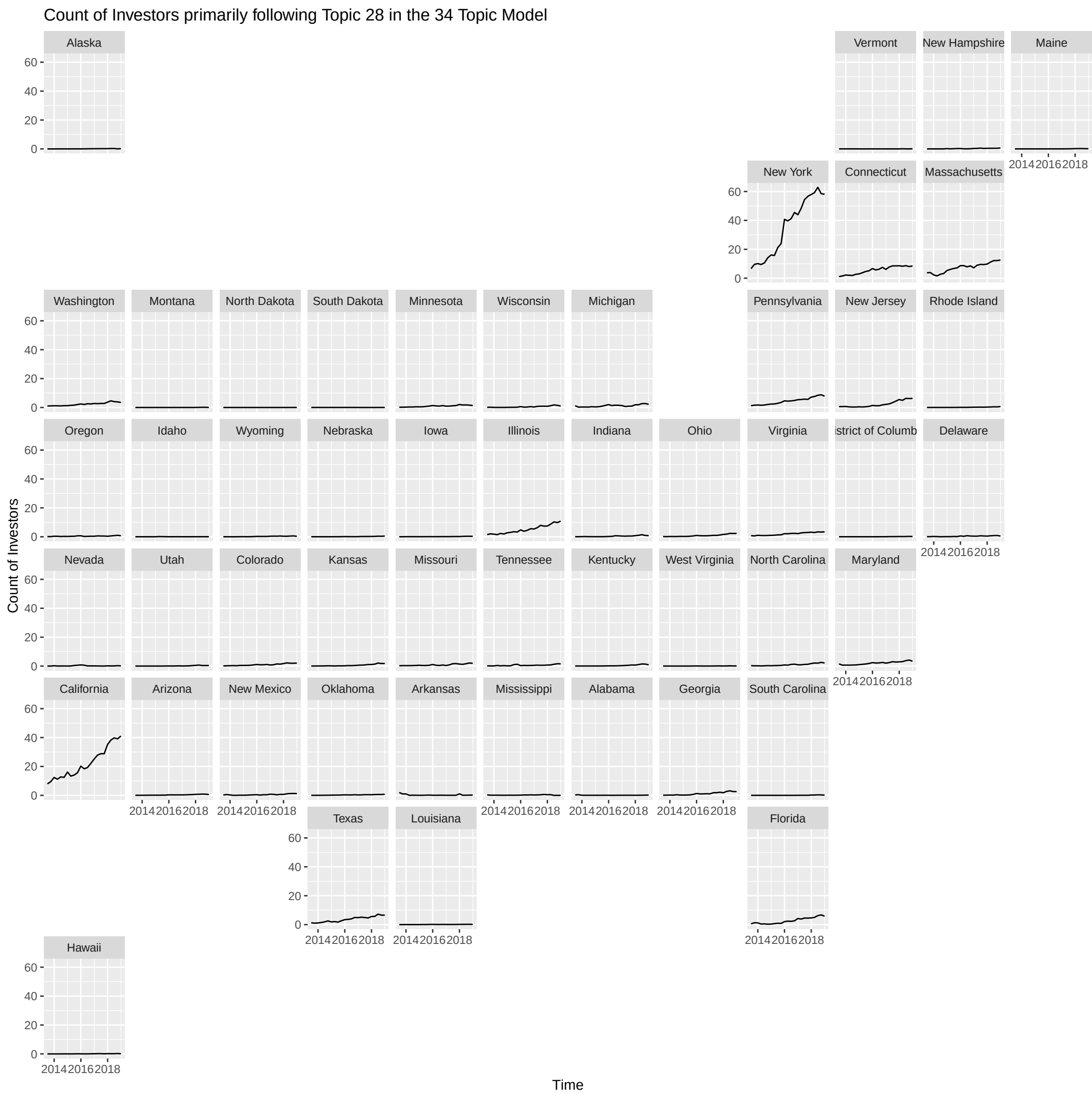


Figure E.28: Count of firms by highest likely topic in the 34 topic LDA for Topic 28

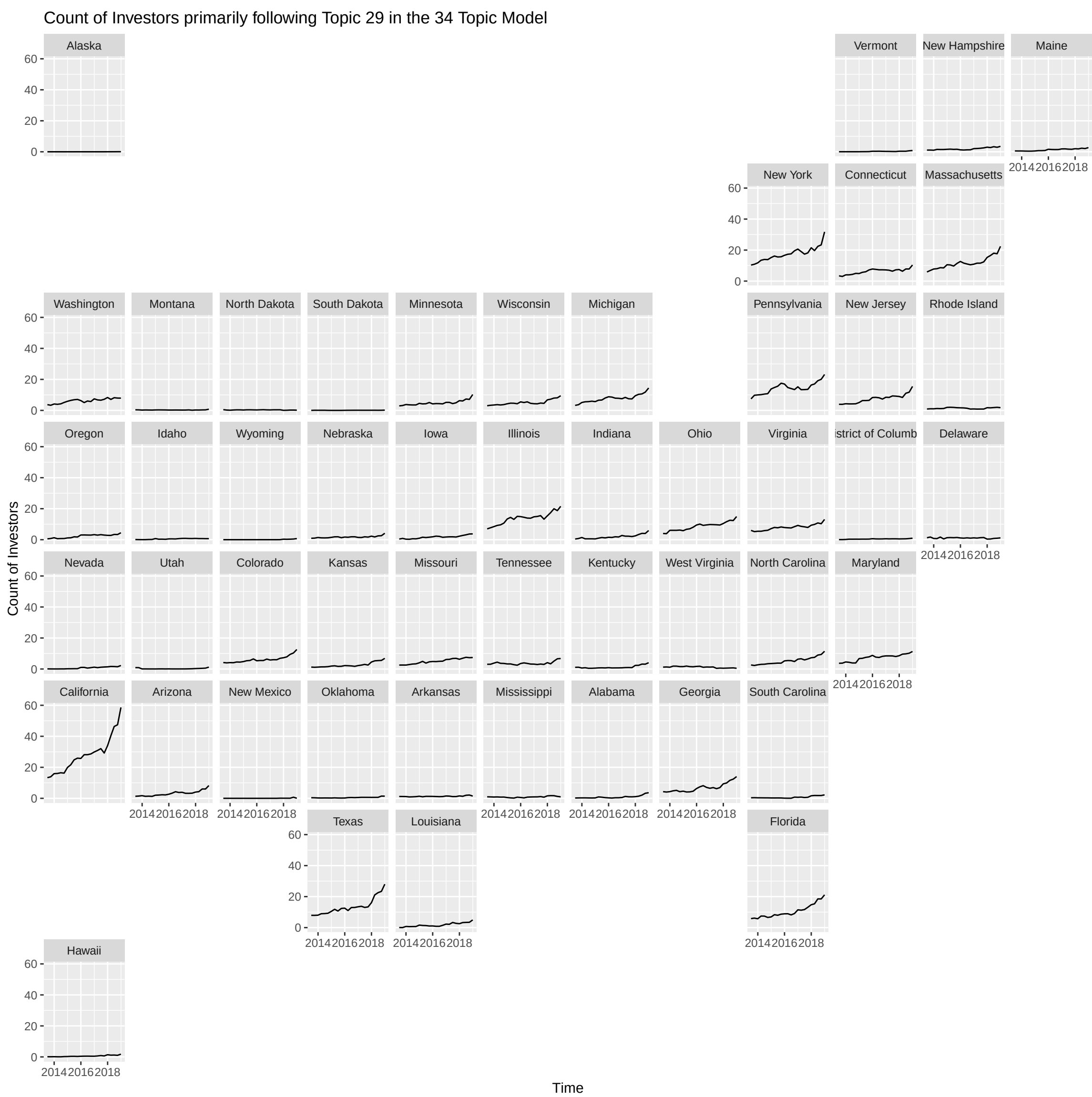


Figure E.29: Count of firms by highest likely topic in the 34 topic LDA for Topic 29

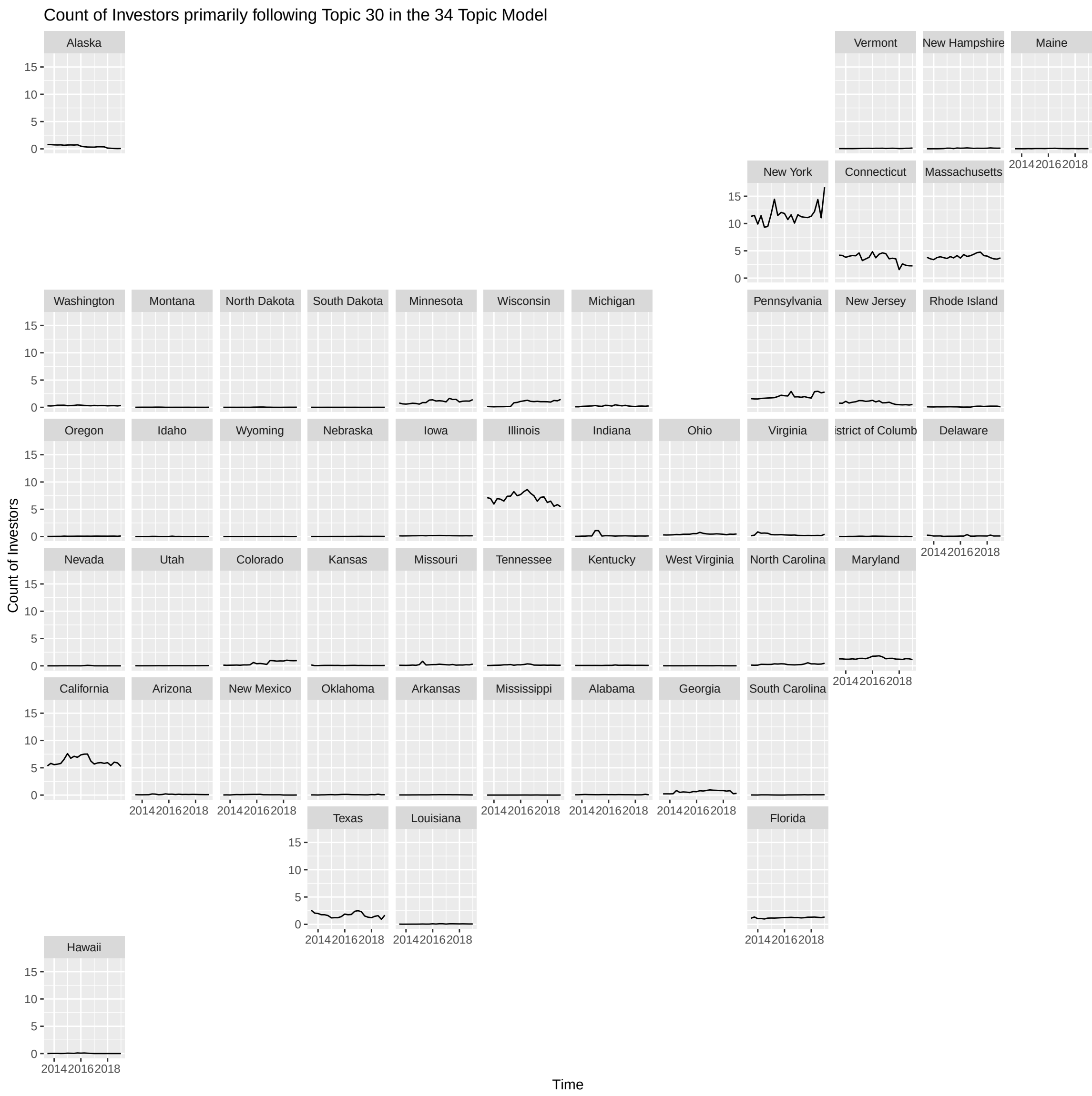


Figure E.30: Count of firms by highest likely topic in the 34 topic LDA for Topic 30

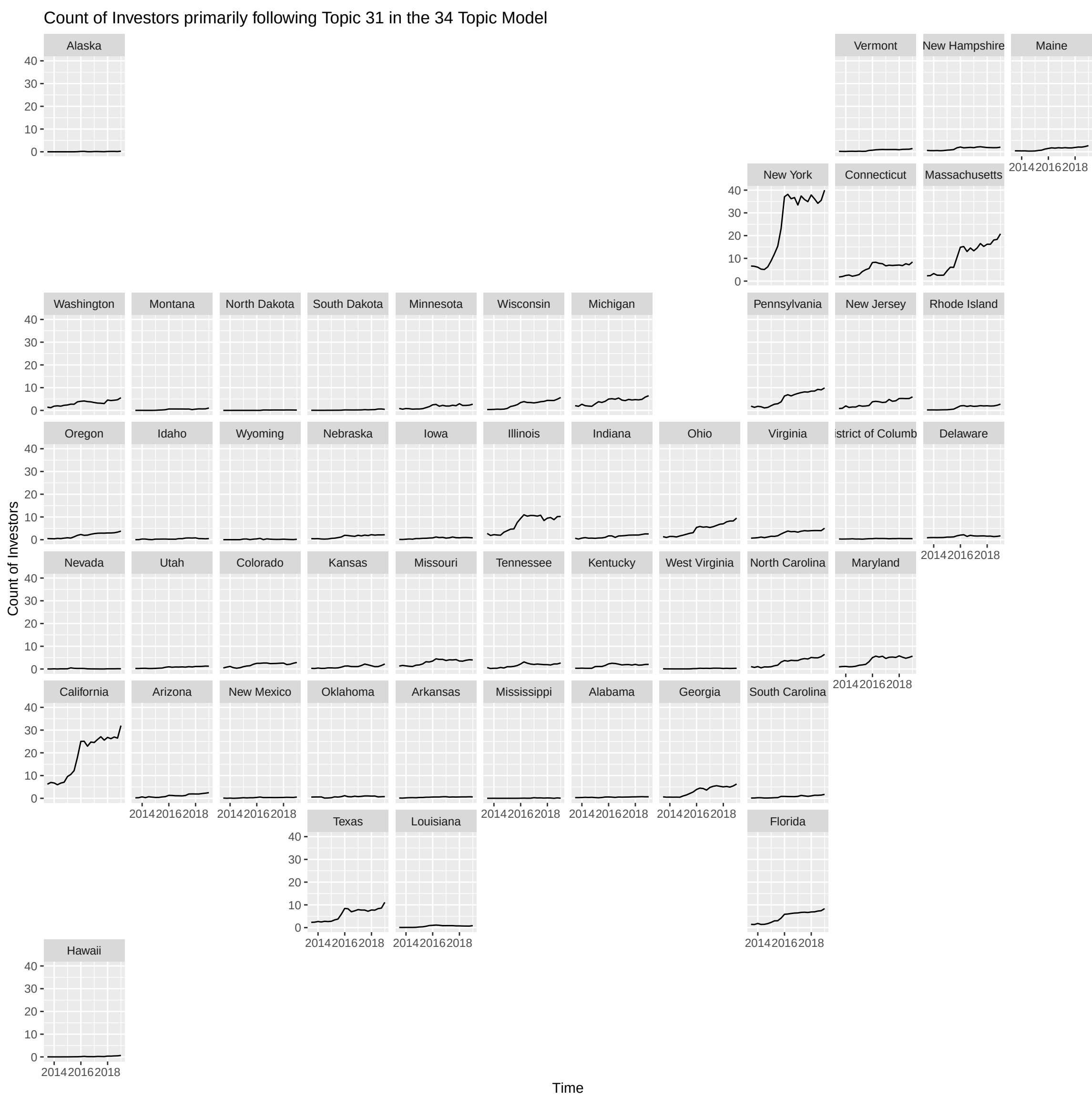


Figure E.31: Count of firms by highest likely topic in the 34 topic LDA for Topic 31

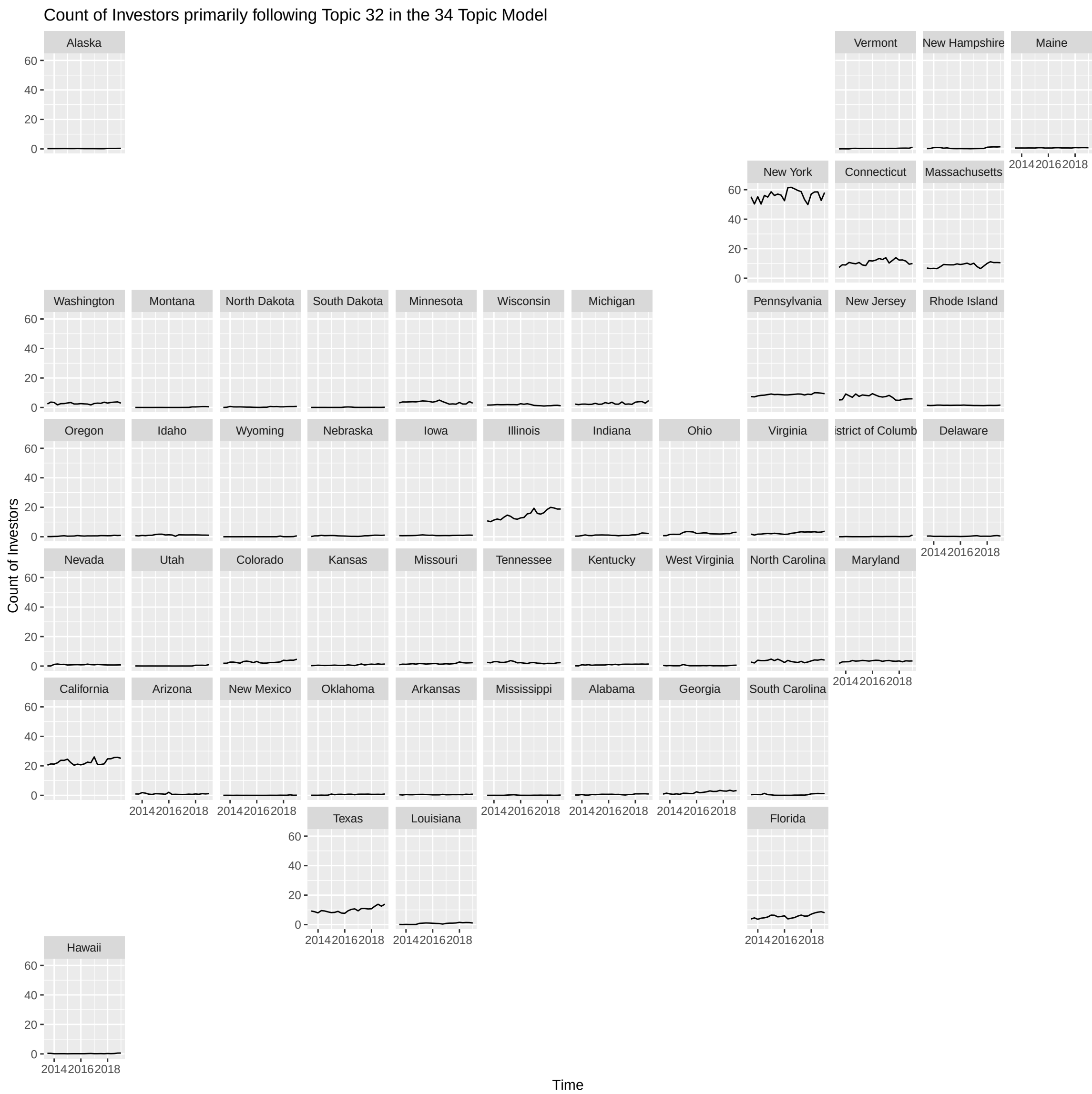


Figure E.32: Count of firms by highest likely topic in the 34 topic LDA for Topic 32

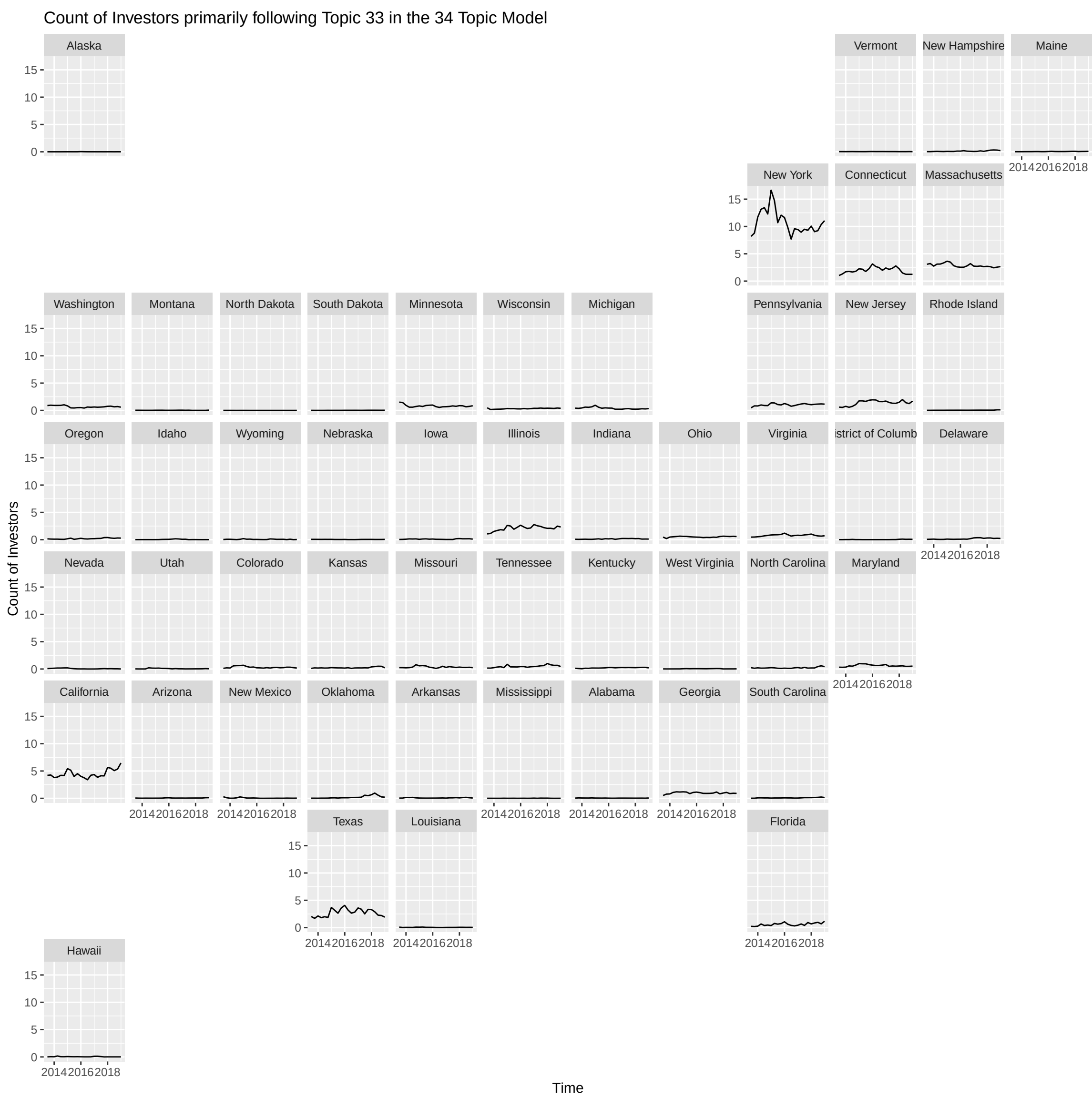


Figure E.33: Count of firms by highest likely topic in the 34 topic LDA for Topic 33

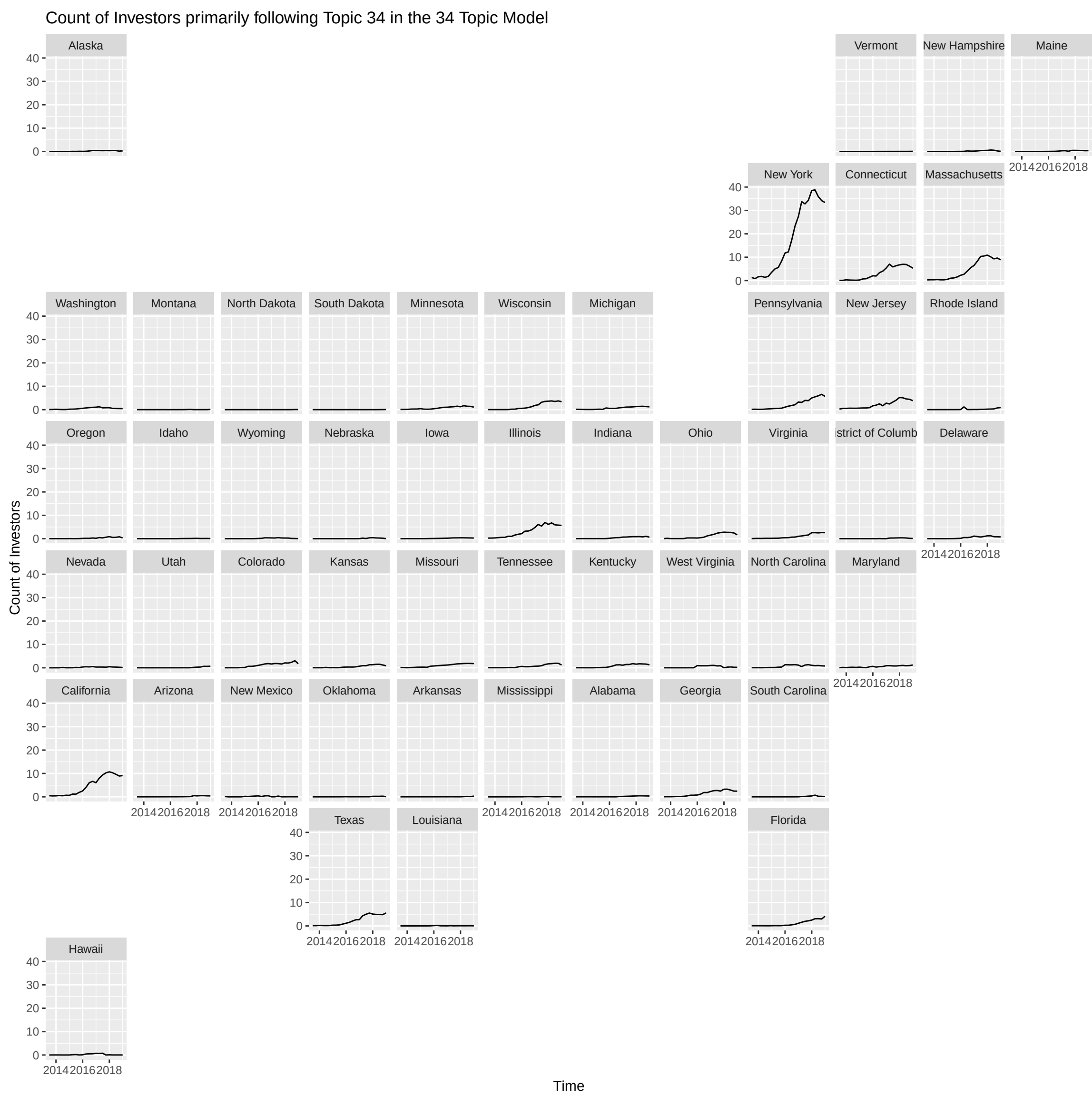


Figure E.34: Count of firms by highest likely topic in the 34 topic LDA for Topic 34

Bibliography

Appelbaum, B., D. S. Hilzenrath, and A. R. Paley (2008). 'All Just One Big Lie'.

Arun, R., V. Suresh, C. E. Veni Madhavan, and M. N. Narasimha Murthy (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi (Eds.), *Advances in Knowledge Discovery and Data Mining*, Berlin, Heidelberg, pp. 391–402. Springer Berlin Heidelberg.

Bachelier, L. (1900). Théorie the la spéculation. *Annales scientifiques de l'É.N.S. 3e série tome 17*, p. 21–86.

Baddeley, A., E. Rubak, and R. Turner (2015). *Spatial Point Patterns: Methodology and Applications with R.* London: Chapman and Hall/CRC.

Bagli, C. V. (2011). Regretting Move, Bank May Return to Manhattan. New York Times `http://www.nytimes.com/2011/06/09/nyregion/ubs-may-move-back-to-manhattan-from-stamford.html`.

BarclayHedge (2018). Hedge fund assets under management. `https://www.barclayhedge.com/solutions/assets-under-management/hedge-fund-assets-under-management/hedge-fund-industry/` (Acessed on 2018-12-22).

Birch, C. P., S. P. Oom, and J. A. Beecham (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological modelling 206*(3-4), 347–359.

Bivand, R., T. Keitt, and B. Rowlingson (2015). *rgdal: Bindings for the Geospatial Data Abstraction Library.* R package version 0.9-2.

Blei, D., A. Ng, and M. Jordan (2003). Latent dirichlet allocation journal of machine learning research (3).

Blei, D. M. (2012, April). Probabilistic topic models. *Commun. ACM 55*(4), 77–84.

Blume, M. E. and D. B. Keim (2012). Institutional Investors and Stock Market Liquidity: Trends and Relationships.

Bodenman, J. E. (1998). The Suburbanization of the Institutional Investment Advisory Industry: Metropolitan Philadelphia, 1983–1993. *The Professional Geographer 50*(1), 112–126.

Bodenman, J. E. (2000). Firm Characteristics and Location: The Case of the Institutional Investment Advisory Industry in the United States, 1983 – 1996. *Papers in Regional Science 79*(1), 33–56.

Bradley, D., C. Pantzalis, and X. Yuan (2016). The influence of political bias in state pension funds. *Journal of Financial Economics 119*(1), 69–91.

Brodie, L. (2013, May). Cramer: Don't Fall for Whale Watching 'Nonsense'! `http://www.cnbc.com/id/100744278`.

Brody, H., M. R. Rip, P. Vinten-Johansen, N. Paneth, and S. Rachman (2000). Mapmaking and myth-making in broad street: the london cholera epidemic, 1854. *The Lancet 356*(9223), 64–68.

Brody, L. (2012). Wall street whale watching: 13f filings flood street.

Brunsdon, C. and L. Comber (2015). *An introduction to R for spatial analysis and mapping*. Sage.

Calomiris, C. W. (2000). *U.S. bank deregulation in historical perspective*. New York: Cambridge University Press. 359 pages.

Cao, J., T. Xia, J. Li, Y. Zhang, and S. Tang (2009). A density-based method for adaptive lda model selection. *Neurocomputing 72*(7), 1775 – 1781. Advances in Machine Learning and Computational Intelligence.

Capital IQ Compustat (2019). Compustat data.

Cashman, G. D., D. M. Harrison, M. J. Seiler, and H. Sheng (2017). The Impact of Geographic and Cultural Dispersion on Information Opacity. *Journal of Real Estate Finance and Economics*, 1–43.

Chen, E. (2011). Introduction to latent dirichlet allocation.

Choe, H., B.-C. Kho, and R. M. Stulz (2005). Do Domestic Investors Have an Edge? The Trading Experience of Foreign Investors in Korea. *Review of Financial Studies 18*(3), 795–829.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*(3), 591–605.

Clark, G. L. (2005). Money Flows Like Mercury: The Geography of Global Finance. *Geografiska Annaler: Series B, Human Geography 87*(2), 99–112.

Cohen, L., A. Frazzini, and C. Malloy (2008). The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy 116*(5), 951–979.

Courtault, J.-M., Y. Kabanov, B. Bru, P. Crépel, I. Lebon, and A. Le Marchand (2000). Louis bachelier on the centenary of théoriethÂ´théorie de la spéculationspÂ´spéculation. *Mathematical Finance 10*(3), 341–353.

Coval, J. D. and T. J. Moskowitz (1999). Home Bias at Home: Local Equity Preference in Domestic Portfolios. *The Journal of Finance 54*(6), 2045–2073.

Coval, J. D. and T. J. Moskowitz (2001). The Geography of Investment: Informed Trading and Asset Prices. *Journal of Political Economy 109*(4), pp. 811–841.

Crippen, A. Berkshire hathaway 'b' shares soar after split. `https://www.cnbc.com/2010/01/21/berkshire-hathaway-b-shares-soar-after-split.html`.

Crymble, A. (2019). Introduction to gravity models of migration & trade. *The Programming Historian 8*.

Dallos, R. E. (1974a, Dec 07). Bad times thin herd at broker convention. Copyright - Copyright Times Mirror Company Dec 7, 1974; Last updated - 2010-05-31.

Dallos, R. E. (1974b, Dec 06). Securities leaders rapped for killing industry overhaul. Copyright - Copyright Times Mirror Company Dec 6, 1974; Last updated - 2010-05-31.

Davis, E. P. and B. Steil (2001). *Institutional Investors*. Cambridge, Mass.: MIT Press. 524 pages.

Dear, M. and S. Flusty (1998). Postmodern Urbanism. *Annals of the Association of American Geographers 88*(1), 50–72.

Deveaud, R., E. SanJuan, and P. Bellot (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique 17*(1), 61–84.

Dixon, P. M. (2014). R ipley's k function. *Wiley StatsRef: Statistics Reference Online*.

Dvořák, T. (2005). Do Domestic Investors Have an Information Advantage? Evidence from Indonesia. *The Journal of Finance 60*(2), 817–839.

Easley, D. and M. O'Hara (2011). Opaque Trading and Asset Prices. *papers.ssrn.com*.

Eckel, S., G. Löffler, A. Maurer, and V. Schmidt (2011). Measuring the effects of geographical distance on stock market correlation. *Journal of Empirical Finance 18*(2), 237–247.

ESRI (2019). *ArcGIS Desktop: Release 10.6.* Redlands, CA.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance 25*(2), 383.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance 46*(5), 1575.

Farooq, O. (2013). Who was informative? Performance of foreign and local analysts' stock recommendations during the Asian financial crisis. *Research in International Business and Finance 29*(1), 61–76.

Feinerer, I., K. Hornik, and D. Meyer (2008, March). Text mining infrastructure in r. *Journal of Statistical Software 25*(5), 1–54.

Fischer, M. M. and A. Getis (2009). *Handbook of applied spatial analysis: software tools, methods and applications.* Springer Science & Business Media.

Fotheringham, A. S. and D. W. S. Wong (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A: Economy and Space 23*(7), 1025–1044.

Friedman, B. (1996). Economic Implications of Changing Share Ownership. *Journal of Portfolio Management 22*(3), 59–70.

Fritsch, M. and D. Schilder (2006, 11). Is venture capital a regional business? - the role of syndication.

Gershman, B. (1982). ABSCAM, the Judiciary, and the Ethics of Entrapment. *Yale Law Journal 91*, 1565–1591.

Getis, A. and J. K. Ord (2010). The analysis of spatial association by use of distance statistics. In *Perspectives on Spatial Data Analysis*, pp. 127–145. Springer.

Gong, H. and K. Keenan (2012). The Impact of 9/11 on the Geography of Financial Services in New York: A Few Years Later. *The Professional Geographer 64*(3), 370–388.

Graves, W. (1998). The Geography of Mutual Fund Assets. *The Professional Geographer 50*(2), 243–255.

Graves, W. (2003). Financing Flexibility in a Global Market: The Metropolitan Distribution of Equity Investment ¹. *Urban Geography 24*(7), 611–635.

Green, M. B. (1993). A Geography of Institutional Stock Ownership in the United States. *Annals of the Association of American Geographers 83*(1), 66–89.

Green, M. B. (1995). A Geography of U.S. Institutional Investment, 1990. *Urban Geography 16*(1), 46–69.

Griffin, J. M. and J. Xu (2009). How Smart Are the Smart Guys? A Unique View from Hedge Fund Stock Holdings. *Review of Financial Studies 22*(7), 2531–2570.

Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences 101*(suppl 1), 5228–5235.

Haldane, A. (2009). Rethinking the financial network. Speech delivered at the Financial Student Association, Amsterdam. `http://www.bis.org/review/r100909e.pdf`.

Harris, R. and R. Lewis (1998). Constructing a Fault(y) Zone: Misrepresentations of American Cities and Suburbs 1900 - 1950. *Annals of the Association of American Geographers 88*(4), 622–639.

Hau, H. (2001). Location Matters: An Examination of Trading Profits. *The Journal of Finance 56*(5), 1959–1983.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). R package version 5.2.2.

Hlynsson, H. D. (2017). Finding magic: The gathering archetypes with latent dirichlet allocation.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey (1983). *Understanding Robust and Exploratory Data Analysis*. New York, Toronto: Wiley. 447 pages.

Huberman, G. (2001). Familiarity breeds investment. *Review of Financial Studies 14*(3), 659–680.

Ingram, D. and S. Franco (2014). Urban-rural classification scheme for counties.

Kenney, M. and U. Von Burg (1999). Technology, entrepreneurship and path dependence: industrial clustering in silicon valley and route 128. *Industrial and corporate change 8*(1), 67–103.

Lang, R. E. and A. C. Nelson (2007). America 2040: The rise of the megapolitans. *Planning 73*(1), 7–12.

Library of Congress (1975). S.249 - An Act to amend the Securities Exchange Act of 1934 to remove barriers to competition, to foster the development of a national securities market system and a national clearance and settlement system, to make uniform the Securities and Exchange Commission's authority over self-regulatory organizations, to provide for the regulation of brokers, dealers and banks trading in municipal securities, to facilitate the collection and public dissemination of information concerning the holdings of and transactions in securities by institutional investment managers, and for other purposes. https://www.congress.gov/bill/94th-congress/senate-bill/249.

Lienou, M., H. Maitre, and M. Datcu (2010). Semantic annotation of satellite images using latent dirichlet allocation. *IEEE Geoscience and Remote Sensing Letters 7*(1), 28–32.

Lo, A. W. (2016). What is an index? *The Journal of Portfolio Management 42*(2), 21–36.

Lofthouse, S. (1997). International diversification. *The Journal of Portfolio Management 24*(1), 53–56.

Lynch, H. J. and P. R. Moorcroft (2008). A spatiotemporal ripley's k-function to analyze interactions between spruce budworm and fire in british columbia, canada. *Canadian Journal of Forest Research 38*(12), 3112–3119.

Malloy, C. J. (2005). The geography of equity analysis. *The Journal of Finance 60*(2), 719–755.

Marino, V. (2019). Revitalization projects reawaken downtown los angeles. *New York Times*.

Markopolos, H. (2009). Testimony of harry markopolos, cfa, cfe, before the u.s. house of representatives committee on financial services wednesday february 4, 2009. Note

= Available online at http://online.wsj.com/public/resources/documents/MarkopolosTestimony20090203.pdf.

Markowitz, H. (1952). Portfolio selection*. *The Journal of Finance 7*(1), 77–91.

Marwick, B. (2014). Chapter 3 - discovery of emergent issues and controversies in anthropology using text mining, topic modeling, and social network analysis of microblog content. In Y. Zhao and Y. Cen (Eds.), *Data Mining Applications with R*, pp. 63 – 93. Boston: Academic Press.

Meyer, S. P. and M. B. Green (1996). Foreign Direct Investment From Canada: An Overview. *Canadian Geographer / Le GÃ©ographe canadien 40*(3), 219–237.

Millard, S. P. (2013). *EnvStats: An R Package for Environmental Statistics.* New York: Springer.

Ministry of Justice (2018). Fact sheet on the uk's relationship with the crown dependencies.

Mitchell, S. (2019). London calling? agglomeration economies in literature since 1700. *J Urban Econ 112*, 16–32.

Monk, A. H. (2009). The geography of pension liabilities and fund governance in the United States. *Environment and Planning A 41*(4), 859–878.

Moriset, B. and E. J. Malecki (2009). Organization versus Space: The Paradoxical Geographies of the Digital Economy. *Geography Compass 3*(1), 256–274.

Munroe, R. (2012). Xkcd 1138.

NASDAQ. Nasdaq composite eligibility criteria.

Newman, D. J. and S. Block (2006). Probabilistic topic decomposition of an eighteenth-century american newspaper. *Journal of the American Society for Information Science and Technology 57*(6), 753–767.

Nikita, M. (2019). *ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*. R package version 1.0.0.

O'Brien, R. (1992). *Global Financial Integration: The End of Geography.* Council of Foreign Relations. 120 pages.

Papadimitriou, C. H., P. Raghavan, H. Tamaki, and S. Vempala (2000). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences 61*(2), 217 – 235.

Papadimitriou, C. H., H. Tamaki, P. Raghavan, and S. Vempala (1998). Latent semantic indexing: A probabilistic analysis. In *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, PODS '98, New York, NY, USA, pp. 159–168. Association for Computing Machinery.

Pohl, N. (2004). Where is Wall Street? Financial Geography after 09/11. *Industrial Geographer 2*(1), 72–93.

Pred, A. (1977). *City-systems in Advanced Economies. Past Growth, Present Processes and Future Options.* Hutchinson.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the statistical society of London 48*(2), 167–235.

Rhoades, S. A. (1982). Size and rank stability of the 100 largest commercial banks, 1925–1978. *Journal of Economics and Business 34*(2), 123–128.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, 255–266.

Robert, H., G. Robert J., P. James, R. Valerie, R. Christina, R. David, S. James, and W. Mark W. (2020). The nber's business cycle dating committee.

Robeson, S., A. Li, and C. Huang (2015). *SphericalK: Spherical K-Function*. R package version 1.2.

Samuelson, P. (1966). Science and stocks. *Newsweek*.

Semple, R. K. (1985). Toward a quaternary place theory. *Urban Geography 6*(4), 285–296.

Shaw, M. (1981). Congress in the 1970s: A Decade of Reform. *Parliamentary Affairs XXXIV*(3), 272–290.

Sievert, C. and K. Shirley (2015). *LDAvis: Interactive Visualization of Topic Models*. R package version 0.3.2.

Silge, J. (2018). The game is afoot! topic modeling of sherlock holmes stories.

S&P Dow Jones Indecies (2020a). Dow jones industrial average.

S&P Dow Jones Indecies (2020b). Dow jones industrial average.

Teo, M. (2009). The Geography of Hedge Funds. *Review of Financial Studies 22*(9), 3531–3561.

Tobler, W. (1995). Migration: Ravenstein, thornthwaite, and beyond. *Urban Geography 16*(4), 327–343.

Tobler, W. (2002). Global spatial analysis. *26*(6), 493–500.

Tukey, J. W. (1977). *Exploratory Data Analysis* (First Edition ed.). Addison-Wesley Publishing Company. 688 pages.

Underwood, T. (2012). Topic modeling made just simple enough.

U.S. Census Bureau (2013). *2013 TIGER/Line Shapefiles [machine-readable data files].*

US Census Bureau (2016).

U.S. Securities and Exchange Commission (1971). Institutional Investor Study Report of the Securities and Exchange Commission, Volumes 1-3. *19*(92), 1–83.

Valle, D., B. Baiser, C. W. Woodall, and R. Chazdon (2014). Decomposing biodiversity data using the latent dirichlet allocation model, a probabilistic multivariate statistical method. *Ecology Letters 17*(12), 1591–1601.

Westaway, J. (1974). The spatial hierarchy of business organizations and its implications for the british urban system. *Regional Studies 8*(2), 145–155.

Wheeler, J. (1988). The corperate role of large metropolitan areas in the united states. *Growth and Change 19*(2), 75–86.

Wheeler, J. O. and R. L. Mitchelson (1989a). Atlanta's Role as an Information Center: Intermetropolitan Spatial Links. *The Professional Geographer 41*(2), 162–172.

Wheeler, J. O. and R. L. Mitchelson (1989b). Information Flows among Major Metropolitan Areas in the United States. *Annals of the Association of American Geographers 79*(4), 523–543.

WilmerHale (2013, January). *Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses.* Available online at `http://files.shareholder.com/downloads/ONE/2272984969x0x628656/4cb574a0-0bf5-4728-9582-625e4519b5ab/Task_Force_Report.pdf`.

Wilschut, L. I., A. Laudisoit, N. K. Hughes, E. A. Addink, S. M. de Jong, H. A. Heesterbeek, J. Reijniers, S. Eagle, V. M. Dubyanskiy, and M. Begon (2015). Spatial distribution patterns of plague hosts: point pattern analysis of the burrows of great gerbils in kazakhstan. *Journal of biogeography 42*(7), 1281–1292.

Yuill, R. S. (1971). The standard deviational ellipse; an updated tool for spatial description. *Geografiska Annaler. Series B, Human Geography 53*(1), 28–39.

Zhu, N. (2002, oct). The Local Bias of Individual Investors. *SSRN Electronic Journal*, 1–59.

Zimmerman, F. L. (1971). House Panel Set To Mull Curbing Bank Trust Units: Patman Bill to Include Rules On Ownership. *Wall Streeet Journal February 8*, 2.