Agglomeration near and far:
San Francisco and
Los Angeles metropolitan areas

Peter Gordon*

John Cho**

Presented at the 58th Annual Meeting of the

Western Regional Science Association

Napa Valley, California February 10-13, 2019

*University of Southern California

**Southern California Association of Governments

ABSTRACT

People (and businesses) in cities want two things, access and space. This presents difficult trade-offs. There are many circumstances and many trade-offs and many resulting choices. This explains commuting near and far. When it comes to the spatial arrangements prompted by agglomerative forces, less is known. Despite the label, the many suggestions concerning “clustering” strategies are unclear on suggested spatial layouts. This research relies on firm-level location data for the Los Angeles and San Francisco areas for selected industries and plant sizes. We find that there is agglomeration near and far -- no matter the region, the industry or the plant size. “Clustering” strategies and discussions must be wary of suggesting (as the name seems to imply) tightly packed spatial arrangements.

I. INTRODUCTION

The basic ideas of this paper are straightforward. They are (1) economic growth is a fundamental[1] objective; (2) economic growth requires innovation; (3) among the prerequisites for innovation are opportunities for interacting and exchanging ideas; (4) interaction opportunities involve a spatial settlement dimension. In short, our astonishing well-offness has, in large part, a spatial explanation.

Beyond this, things are less clear. Agglomeration economies are part of the story and it is agreed that they have a spatial dimension. But is there such a thing as a preferred spatial arrangement? There is much talk about the importance of "clusters" and "density" but specifics, degree of, dimensionality, layout, are left unclear. Map 1 shows the locations of software establishments in the San Francisco Bay area. Whereas there is much discussion of Silicon Valley, it appears that these firms are all over the Bay Area – at all sorts of densities. There are also numerous assertions about "sprawl" and "urban containment" but these are often vague and/or contentious.

There is also the hard fact that human design capabilities can be scaled up only so far. Designers (subject to competitive forces) create all of the products we can put our hands on, even the buildings and facilities that house all these items. Designers also have a hand in suggesting the best layout of major facilities, including commercial and industrial centers. But there are limits to what human design can accomplish. Scaling-up capabilities only go so far. Jane Jacobs told her readers that "a city cannot be a work of art". Monumental buildings have been important for a long time. But at some scale, human action, not human design, must take over. *Emergent orders* – those based on voluntary interactions -- denote the arrangements that result. Beyond some threshold, knowledge is so complex and so dispersed that trial-and-error learning (usually best in a competitive setting, with many inevitable errors made along the way) are the only way to bring about beneficial designs and outcomes.

There are many examples. The ones most widely cited are language, science, common law, art, and culture. Cities also. Jane Jacobs had famously noted that

[1] Some prefer *sustainable* economic growth but are often unclear by what they mean.

neighborhoods and cities are also examples of what sounds like an emergent order. In her words, “Their intricate order – a manifestation of the freedom of countless numbers of people to make and carry out countless plans – is in many ways a wonder” (Jacobs, 1961).[2] Agglomeration opportunities are *realized* if spatial arrangements are congenial. This requires a measure of open-endedness by land use planners.[3][4]

Everyone tries to make the best of the situation they face. Everyone plans. Large numbers of individual plans coalesce into orders that may *appear* to be somehow fashioned top-down but we now know that such top-down success is an impossible dream (for some).

Everyone can see Jacobs’ “intricate order” but what do they make of it? Looking down at a city from an airplane, we claim, one actually sees the spatial layout of a large number of overlapping supply chains, *including supply chains for things and supply chains for ideas*.

Ronald Coase famously observed that plant managers decide what to make vs. what to buy. They must also decide what to buy (and sell) *where*. Supply chains have a spatial layout. Consider that everyone in cities, people or businesses, wants two things: space and accessibility(ies). This describes the choice problem faced by all locators as well as the mediation problem that we expect land markets to handle. Add to all of this the important fact that most locators participate in supply chains for things as well as supply chains for ideas. We make it our business to seek and find *useful* knowledge[5] – useful to our enterprise or our mission. Looking to purposeful entrepreneurial action in this way addresses the textbook worries over free-riding on ideas that are “in the air.”

Supply chains for ideas can involve physical access as well as electronic access. Many people work remotely as well as at a workplace. Most of us are keen to find the *blend* of access modes that works for our enterprise. Death-of-distance dreams were premature. Establishing and maintaining trust relationships requires some physical presence. “The problem with the internet is that he

---

[2] “Death and Life is probably the most sensible book ever written about cities” (Watson, 2001, p. 521)
[3] A similar argument as ours with empirical results for the UK is in Huggins and Thompson (2017)
[4] Considerable empirical literature shows that land use planning is so heavily regulated as to cause housing shortage and housing “affordability” problems. See, for example, Pollakowski and Wachter (1990).
[5] Mokyr (2007)

cannot look her in the eye through a screen, and she cannot ‘feel’ or ‘touch’ him. It is a medium that may help to sustain relationships, but it does not establish deep and complex contacts.” (Leamer and Storper, 2001)

Access, reduced access costs, distance and place enter this discussion. Knowledge is exchanged for money (or for other tangible rewards) *via carefully cultivated networks*. Within these, reputations for careful and truthful attribution are established, honed, and maintained. This applies especially to non-codified (tacit and also not easily patented) knowledge which may require extended conversation, even face-to-face interaction, acquaintanceship and geographic proximity. Knowledge involves learning. Learning is hard. Conversation establishes context which can make all the difference.

Networking is a popular idea but is *fashioning and managing supply chains for ideas* more apt (Figure 1)? We know that new ideas are new combinations of old ideas. This refers to forming new neural connections in our cortex, the “aha” moments.[6] The number of possible combinations is almost uncountable but as our brains become embedded and part of a networked network, as we network with others, the combinatorial possibilities expand dramatically. The bigger, the better. We can describe innovation and the benefits of accessibility, connectivity and agglomeration in this way.

II. THIS RESEARCH

Spatial agglomerations have been explained via a variety of impulses. There are plausibly gains from locating near firms of the same industry (specialization; Marshall, 1890); there also plausible gains from locating near firms that represent complementary (diversification; Jacobs, 1969) sectors. Are the worthy spillover ideas the ones that are highly specific or specialized?

There are also likely gains from the sheer nearby availability of possible substitutes should exiting relationships of any sort be disrupted or deemed inadequate. Locating near diverse labor pools also confers benefits -- for the present or the future.

[6] Romer, Paul (1994)

We studied the co-location of firms for selected industries in the San Francisco and Los Angeles metropolitan areas.[7] The sectors chosen represented entertainment (NAICS 512: Southern California's presumed current growth engine); various sectors that represent "tech" and engineering (the Bay Area's growth engine NAICS 4147); a combined information sector including computer programming, software, and information retrieval services (NAICSS 541511, 511210, 517210, 517919); a finance sector that presumably includes venture capital groups (NAICS 523910); a large technology-based manufacturing sector that includes 43 (six-digit NAICS) sectors.[8]

Ideas and capital have a natural synergy; either one without the other is almost useless. We also combined all five of our sectors. Table 1 presents summary data on these sectors. Maps 2a -2d show firm locations in the two regions.

A recent study that also compared these two regions is by Storper, et al (2016). These authors tried to explain the how and why the Bay Area outpaced LA's growth in recent years. They focused on what they saw as better regional coordination by major actors. But it may be simpler. The Bay Area includes the world's premier "tech" center. Moreover, the dominance of Silicon Valley has been shown to have had little to do with top-down regional planning choices (Saxenian, 1998).

We began by looking at the pairwise co-locations of firms, here defined as observed correlations of jobs at the census block group level.[9] There were 8,248 and 3,978 CBGs in the Los Angeles and Bay Area regions, respectively. The first step was to test the extent to which pairwise co-locations are explained by sales and purchase coefficients from U.S. input-output models.[10] Four regression results are shown in Tables 2a-2d. The two estimated coefficients are significant with the correct signs in three of them. But in all four regressions the proportion of dependent variable variation explained was very low, usually less than two percent. Our candidate for other explanations for co-location and agglomeration

---

[7] We use the Los Angeles MSA, defined as Los Angeles plus Orange counties. The San Francisco MSA excludes Santa Clara county which we add to it to create a six-county area; leaving out the "Silicon Valley" growth engine would make no sense.

[8] See Paytas and Berglund (2004)

[9] From InfoUSA 2016.

[10] US Implan, 2013, 536 sectors. Using regional model coefficients would have introduced two-way causation problems.

is the importance of information exchange. Depending on the nature of the exchange, this activity can involve co-location near or far. Near and far commuting is well known (Figure 2). Cities continue to spread outward. Why not near and far agglomeration? In both cases, there are a variety of accommodations to a variety of situations.

III. FINDINGS

Our data on firms and jobs by sector, location, and size of firms enabled us to estimate Ripley functions for our five sectors in each of the two regions. There were separate estimations for all firms as well as the largest firms. The Ripley plots are shown in Appendix 2.

In this research, we try to learn from the spatial arrangements that we see on the ground. To avoid the problems of identifying centers and subcenters – and leaving out all the other jobs -- we relied on the Ripley K-functions estimated on location data for *all* the firms. To give "clustering" some meaning, we investigated the degree of observed co-location of firms via Ripley's K-function results. The K function is

$$K(d) = \frac{1}{\lambda n}\sum_i \sum_{j \neq i} \frac{I_d(m_{ij})}{w_{ij}} \qquad (2)$$

Where λ is the density of firms in the study area, n is observed number of firms, $I_d(m_{ij})$ is 1 when $m_{ij} \leq d$ or 0 when $m_{ij} > 0$. m_{ij} is the distance between firm i and j. w_{ij} is a weight function for edge correction (Dixon, P, 2002). Simulated outer boundary method is applied for edge correction (ArcGIS Pro, 2017).

The estimation results denote the extent to which *encountering another firm of the same (or designated) sector is greater than random, as distances from sector firms increase.* Greater than random is denoted by displacement of the function above the 45-degree line. Our Ripley estimation results are summarized in Table 3. The estimated functions are shown in Appendix 2.

The Ripley estimations help us address this question: as distances from firms increase, are the odds of encountering a co-locating (same "sector" as per our definitions) firm greater than random? Are estimated Ripley functions above the rising 45-degree plot? A rising 45-degree line shows the expected increasing odds of an encounter with co-locating firm as distances from sector firms increase. Are the actual encounters greater than random? Measuring the areas of the bulge above the 45-degree line summarizes all this.

Table 3 reports the results. All of the percentages are above zero indicating strong clustering. Now fast-forward to the plots. They are all outside the 95-percent confidence bands. For all of the 24 estimations, we see "clustering" all the way to the edge of observed plant locations (30-40 km). This is much greater than a 5km edge of a large downtown cluster. It is the case for both regions, all five sectors, all plant sizes. In most cases, less clustering by the largest (perhaps more independent) firms.

Back to Table 3. According to bulges above the 45-degree line, most clustering is among LA-area entertainment firms (Row 1) – where smiles and handshakes matter most. In both regions, there is least clustering among computer programming (Row 4) firms – where considerable electronic data exchange most likely. Row 2 involves more "tech" firms; the results are broadly similar for the two regions, with less clustering for the largest firms – that are expected to be less dependent on outside help. Row 3, venture capital firms shows the same type of clustering, but the largest firms cluster more. The biggest regional differences are for the largest sector studied (Row 5). This may be the outlier. Combining all five sectors (Row 6) shows that overall spatial layouts are remarkably similar in the two regions.

The final row combines data for all the sectors studied. These results address clustering across sectors. As the Jacobs hypothesis suggests, there is *not* less clustering when we go beyond a single industry. Complementary matters.

IV. DISCUSSION

" … [A] central paradox of our times is that in cities, industrial agglomerations remain remarkably vital despite ever easier movement of goods and knowledge over space" (Glaeser, 2010). We say no paradox. We have shown that there are agglomeration benefits *near and far*; more than one kind of information is communicated and exchanged, the codified and the tacit. Locators pick a suitable *blend* of interaction channels which helps them select their preferred location-networking choices. Over a half-century ago, Mel Webber noted “community without propinquity”. In pre-internet days, he saw that people chose to link in a variety of ways. The choices that individuals make in their personal spheres obviate the many grand plans, including the many “cluster strategies.” Networking has replaced hierarchy.[11]

The co-location of workers and employers is well known (Table 4). Both sides have an interest in avoiding high commuting costs.[12] Firms, likewise, have reasons to be strategic about choosing a location vis a vis other firms. We have tried to be more precise than simply alluding to “density”, “clustering”, “agglomerating.” We have argued and shown that agglomerating near as well as far makes great sense.

Cities are defined by their peculiar land use arrangements. These are emergent even though constrained by topography, history, and development rules. The city will only grow and prosper (and contribute to general human advancement) if the emergent spatial patterns are congenial to the formation and functioning of very large numbers of supply chains – for goods as well as for ideas. “Emergent” is quite distinct from all the discussions focused on “cluster strategies”. “Efficient” may be a fraught term but growth is clear enough.

[11] Gurri (2018) elaborates. He speculates that networks will replace hierarchies.

[12] “The data on the largest U.S. MSAs show that commute times increase only slightly with city size: the elasticity of the average commute time with respect to the number of workers was about 0.1 in 1990 and 2000” (Anas, 2012, p.146)

REFERENCES

ArcGIS Pro. (n.d.). Retrieved August 25, 2017, from http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/multi-distance-spatial-cluster-analysis.htm

Anas, Alex (2012) Discovering the efficiency of urban sprawl. in *The Oxford Handbook of Urban Economics and Planning*, edited by N. Brooks, K. Donaghy, and G. J. Knaap. New York: Oxford University Press.

Glaeser, Edward (2010) "Introduction" *Agglomeration Economics*, E. Glaeser (ed) NBER: The University of Chicago Press.

Gurri, Martin (2018) *The Revolt of the Public*. San Francisco: Stripe Press.

Huggins, Robert and Piers Thompson (2017) "Networks and regional economic development of knowledge ties" *Environment and Planning A*, 1-19.

Jacobs, Jane (1961) *The Death and Life of American Cities*. New York: Random House.

Jacobs, Jane (1969) *The Economy of Cities*. New York: Random House.

Leamer, Edward E. and Michael Storper (2001) "The Economic Geography of the Internet Age" NBER Working Paper 8450.

Marshall, Alfred (1890) Principles of Economics. London: Macmillan.

Mokyr, Joel (2007) "The Market for Ideas and the Origins of Economic Growth in Eighteenth Century Europe" *Tijdschrift Voor Sociale en Economische Geshiedenis*, 4

Paytas, J and D. Berglund (2004) " Technology Industries and Occupations for NAICS Industry Data" Center for Economic Development and State Science & Technology Institute

Pollakowski, Henry O. and Susan M. Wachter (1990) "The Effects of Land Use Constraints on Housing Prices" *Land Economics*, 66:3, 315-324.

Romer, Paul M. (1994) "Economic growth and investment in children." *Daedalus* 123(4):141–154

Rosenthal, Stuart S and William C Strange (2004) "Evidence on the Nature and Sources of Agglomeration Economies" in *Handbook of Regional and Urban Economics,* vol 4, p. 2119-2171.

Saxenian, A. (1998), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Harvard University Press, Cambridge, Ma.

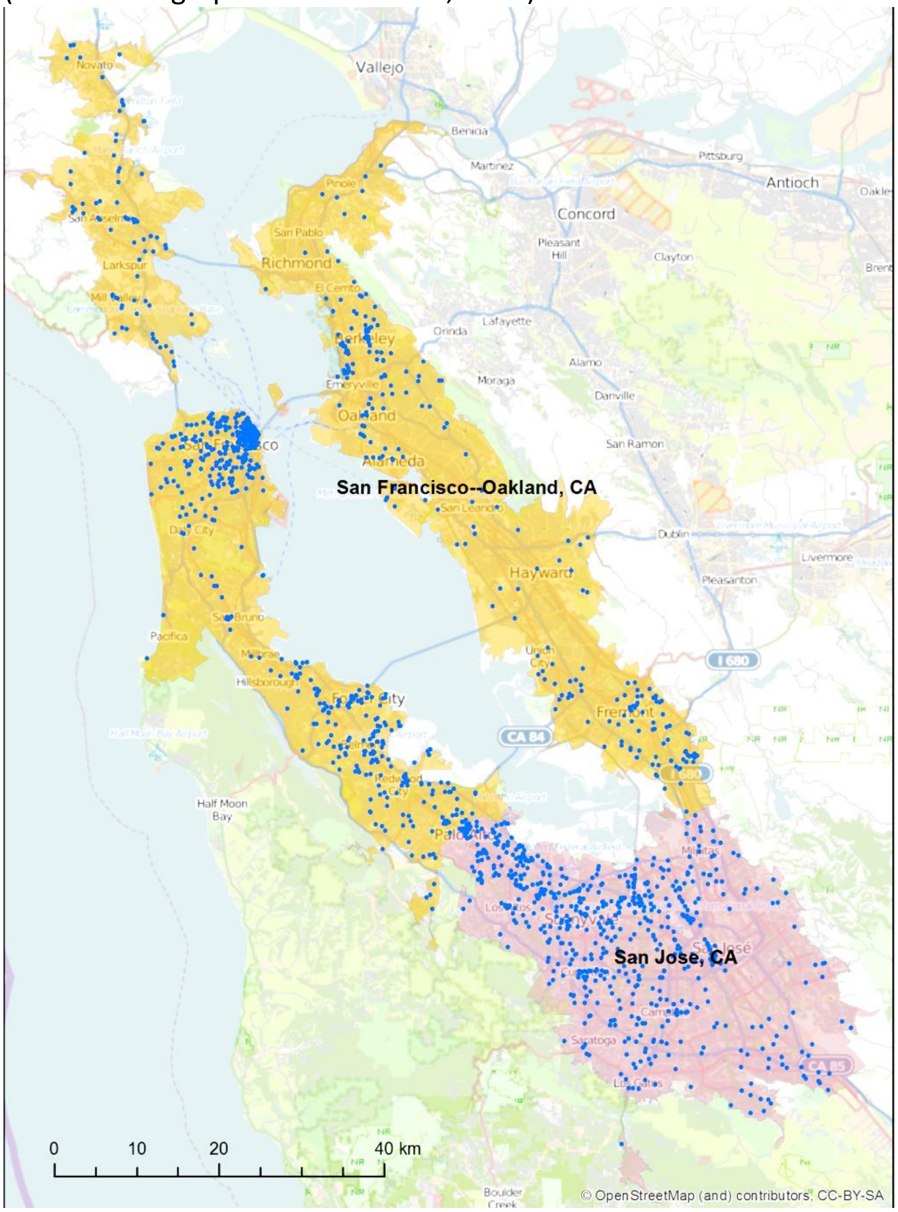
Storper, M. (2016) *The Rise and Fall of Urban Economies: Lessons from San Francisco and Los Angeles*. Stanford: Stanford University Press.

Watson, Peter (2001) *The Modern Mind*. New York: Harper Collins.

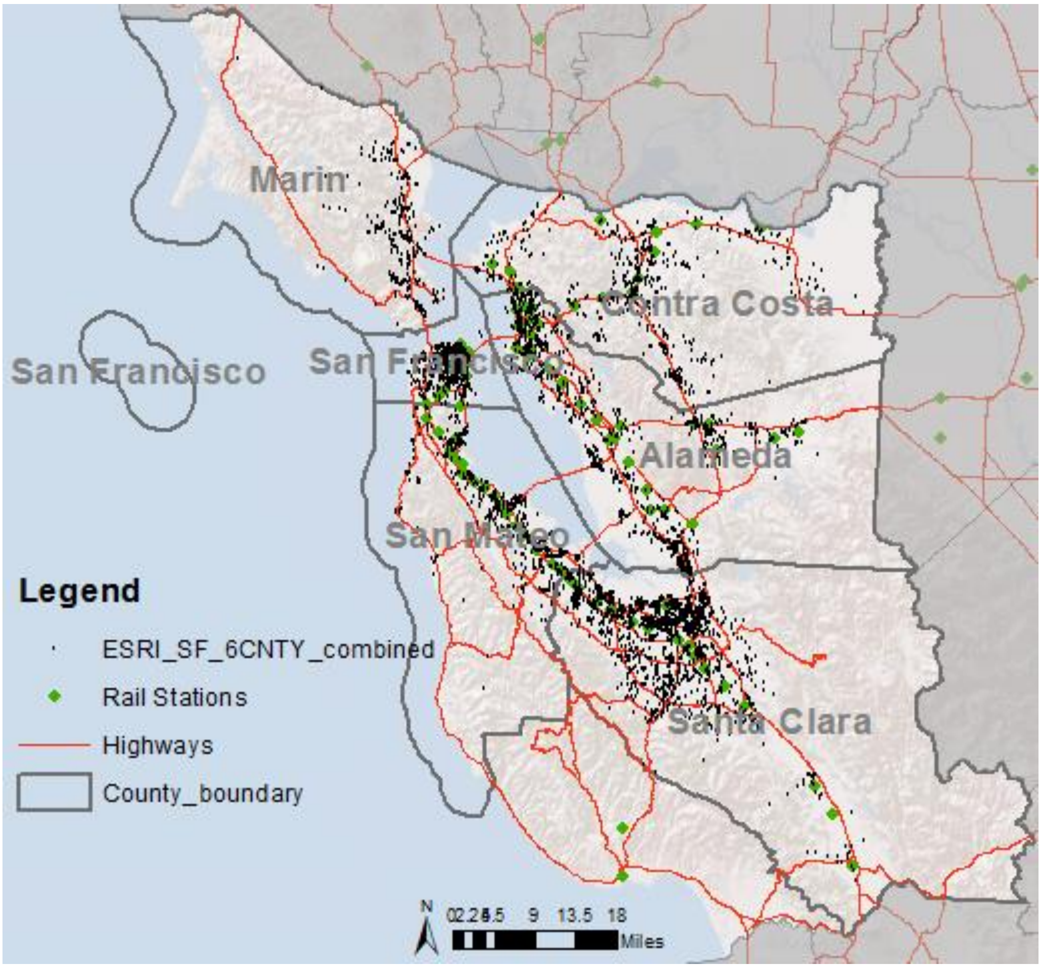
Webber, M.M. (1963) "Order in Diversity: Community without Propinquity" in *Cities and Space*, L. Wingo (ed.). Baltimore: Johns Hopkins University Press.

MAPS, FIGURES, TABLES

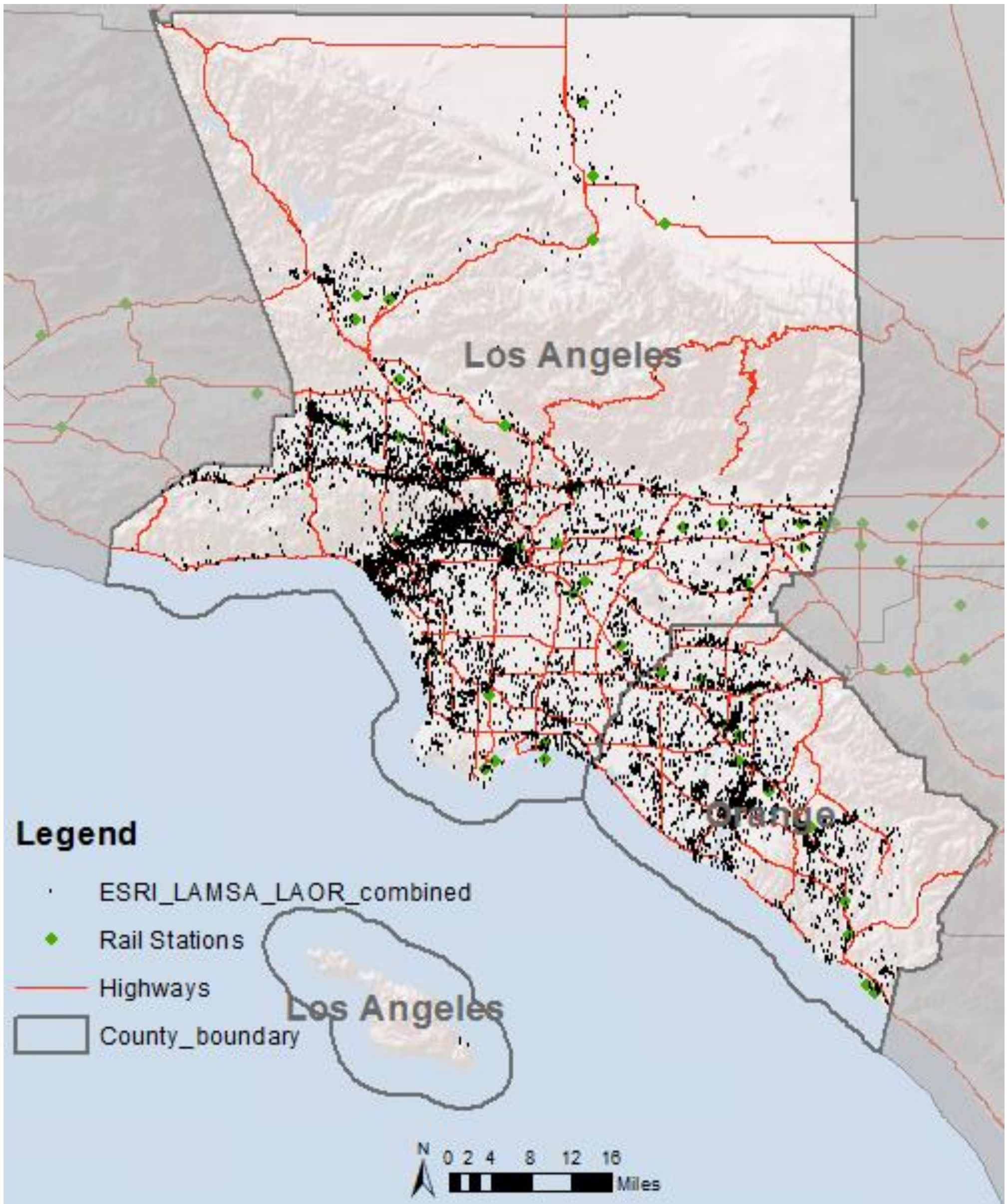
MAP 1: Location of all software establishments in the San Francisco Bay Area, 2013: We see the dots but not the interconnecting networks; firms network near and far. (Source: Geographic Research Inc., 2014)



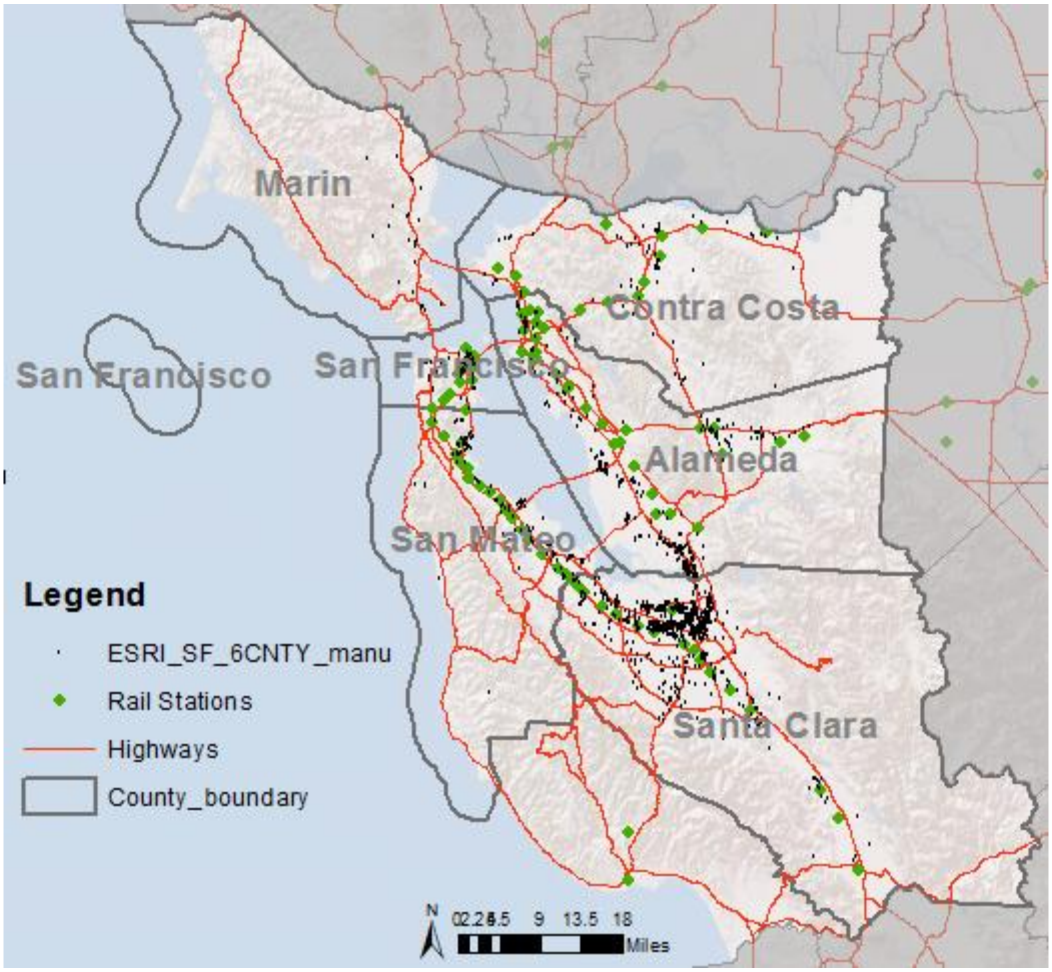
MAP 2a: San Francisco MSA plus Santa Clara County, Locations of combined sectors firms



MAP 2b: Los Angeles MSA (LA + Orange counties), Locations of combined sectors firms



MAP 2c: San Francisco MSA plus Santa Clara County, Locations of technology based manufacturing sectors firms



MAP 2d: Los Angeles MSA (LA + Orange counties), Locations of technology based manufacturing sectors firms

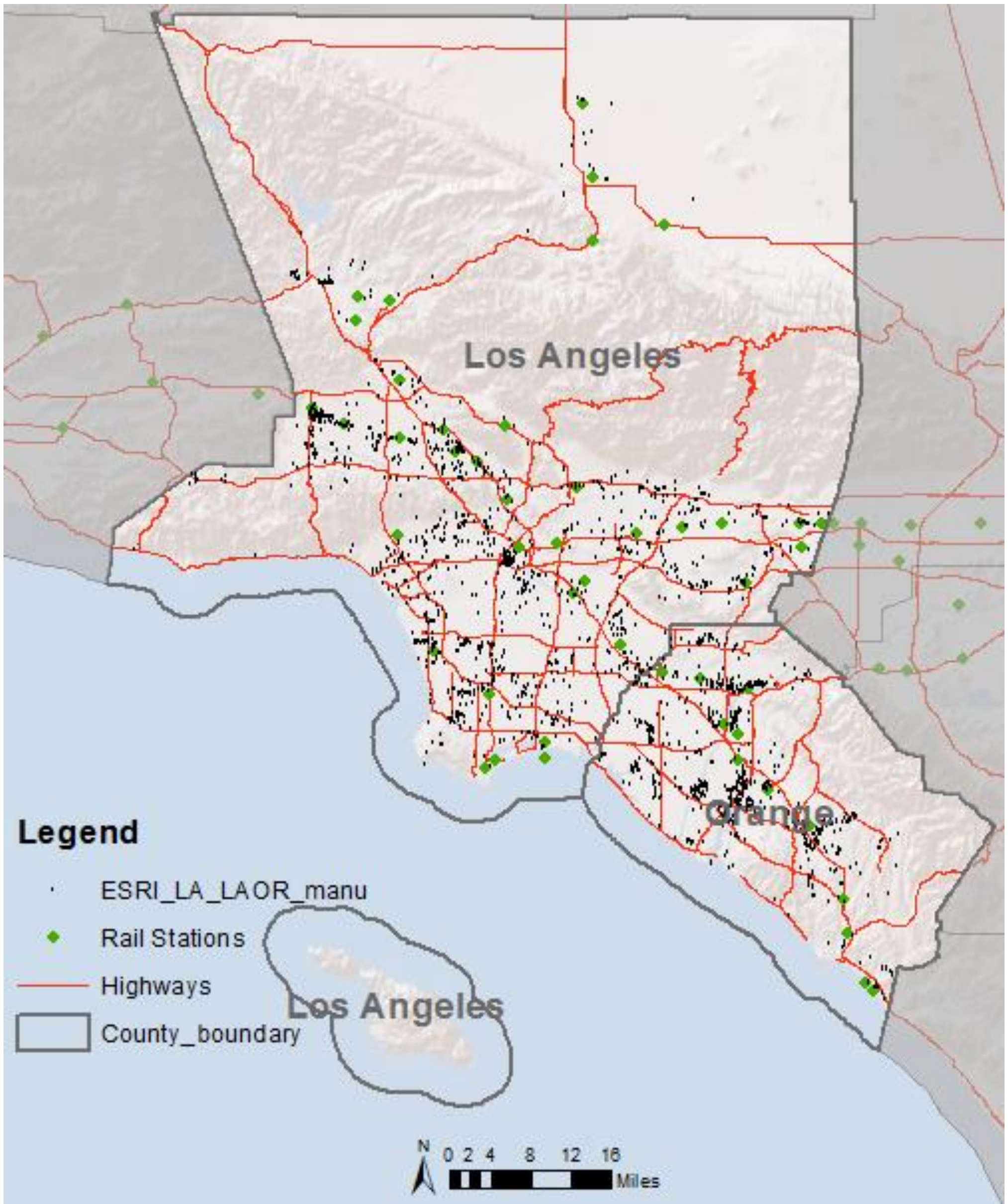
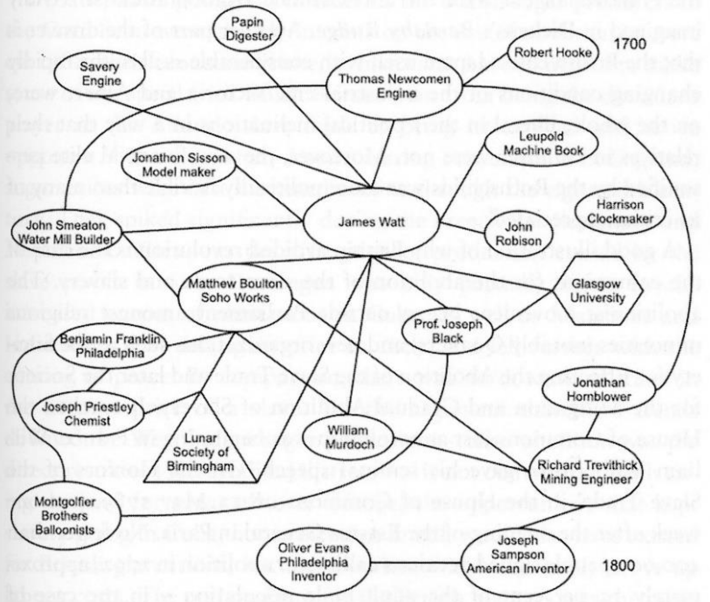


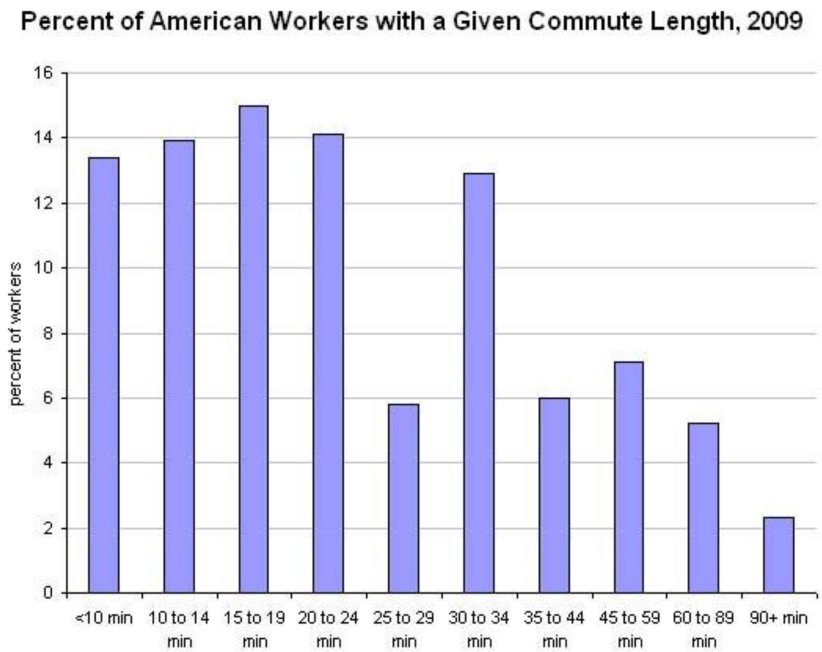
FIGURE 1: Network? Web of social relations? Supply chain for ideas?



18. The 'steam network': James Watt, Matthew Boulton and the social network of steam engine technology, *c.* 1700–1800.

Source: Niall Ferguson, *The Square and the Tower: Networks and Power from the Freemasons to Facebook* (2017), p. 147; includes more examples

FIGURE 2: Cities spread outward; commuting near and far; why not agglomeration/clustering near and far?



Source: U.S. Census Bureau, American Community Survey, 2009

TABLE 1: Firms and Jobs, LA and SF areas, selected sectors

Region	Sector (NAICS)		Firms	% of total firms	Jobs	% of total jobs
Los Angeles County/ Orange County	NAICS 512	Motion Picture and Sound Recording Industries	4,501	0.8%	72,847	1.2%
	NAICS 5417	Scientific Research and Development Services	1,832	0.3%	28,316	0.5%
	NAICS 523910	Venture Capital Companies	1,680	0.3%	11,953	0.2%
	NAICS 541511, 511210, 517210, 517919[13]	Computer Programming, Software, and information Retrieval Services	1,761	0.3%	21,080	0.3%
	NAICS 32-33[14]	Technology-based manufacturing industries	2,114	0.4%	109,113	1.8%
		Five sectors combined	11,888	2.1%	243,309	4.0%
		All Sectors	**563,326**	**100.0%**	**6,034,164**	**100.0%**
SF 6 County	NAICS 512	Motion Picture and Sound Recording Industries	1,158	0.4%	10,678	0.3%
	NAICS 5417	Scientific Research and Development Services	1,629	0.6%	41,990	1.3%
	NAICS 523910	Venture Capital Companies	1,261	0.4%	11,365	0.3%
	NAICS 541511, 511210, 517210, 517919	Computer Programming, Software, and Information Retrieval Services	1,738	0.6%	55,039	1.7%
	NAICS 32-33[11]	Technology-based manufacturing industries	1,605	0.6%	158,574	4.9%
		Five sectors combined	7,391	2.6%	277,646	8.5%
		All Sectors	**280,332**	**100.0%**	**3,264,760**	**100.0%**

[13] Suggested by Rosenthal and Strange (2004)
[14] The sector includes 43 technology-based manufacturing sectors defined by Paytas and Berglund (2004)

TABLE 2a: OLS Estimations: Pairwise co-location correlations explained by national IO coefficients, sales and purchases: Los Angeles county

Variable	Coefficient	Standard Error	Pr > |t|	R-Squared
Intercept	0.1233	0.0047	<.0001*	0.0174
Tech_coef_ij	0.4407	0.1877	0.0190*	
Tech_coef_ji	2.1427	0.3120	<.0001*	

* Statistically significant

N = 2,980

TABLE 2b: OLS Estimations: Pairwise co-location correlations explained by national IO coefficients, sales and purchases: Orange county

Variable	Coefficient	Standard Error	Pr > |t|	R-Squared
Intercept	0.1962	0.0068	<.0001*	0.0054
Tech_coef_ij	-0.0932	0.2382	0.6956	
Tech_coef_ji	1.7637	0.4477	<.0001*	

* Statistically significant

N = 2,902

TABLE 2c OLS Estimations: Pairwise co-location correlations explained by national IO coefficients, sales and purchases: Los Angeles and Orange

Variable	Coefficient	Standard Error	Pr > |t|	R-Squared
Intercept	0.1239	0.0042	<.0001*	0.0182
Tech_coef_ij	-0.0749	0.1482	0.6132	
Tech_coef_ji	2.0550	0.2769	<.0001*	

* Statistically significant

N = 2,988

TABLE 2d: OLS Estimations: Pairwise co-location correlations explained by national IO coefficients, sales and purchases: San Francisco 6-county area

Variable	Coefficient	Standard Error	Pr > |t|	R-Squared
Intercept	0.1275	0.0053	<.0001*	0.0169
Tech_coef_ij	0.5456	0.2114	0.0099*	
Tech_coef_ji	2.3452	0.3534	<.0001*	

* Statistically significant

N = 2,951

TABLE 3: Summary of Ripley function estimations, degrees of observed clustering

Sector (NAICS, SIC)	Description	Degree of observed clustering			
		LA/OR Counties		6 Bay Area Counties	
		All firms	Largest firms	All firms	Largest firms
NAICS 512	Motion Picture and Sound Recording Industries	104%	90%	71%	63%
NAICS 5417	Scientific Research and Development Services	50%	32%	62%	44%
NAICS 523910	Venture Capital Companies	60%	71%	76%	87%
NAICS 541511, 511210, 517210, 517919	Computer Programming, Software, and information Retrieval Services	44%	33%	56%	53%
NAICS 32-33	Technology-based manufacturing industries	55%	28%	95%	88%
	Five Sectors Combined	77%	60%	73%	61%

TABLE 4: U.S. Commuting, 2009, Solo-driver auto trips, minutes: Urban, suburban, second-city means near U.S. average

Area	Means	Variances
Metro	25.2	338.7
Urban	*22.8*	*214.7*
Suburban	*24.5*	*292.4*
Second City	*23.6*	*308.5*
Town and Country	28.4	457.0

Source: An, Gordon, Moore (2015)

APPENDIX 1

COMPARISONS, TWO URBANIZED AREAS (UZAs), 1950-2010

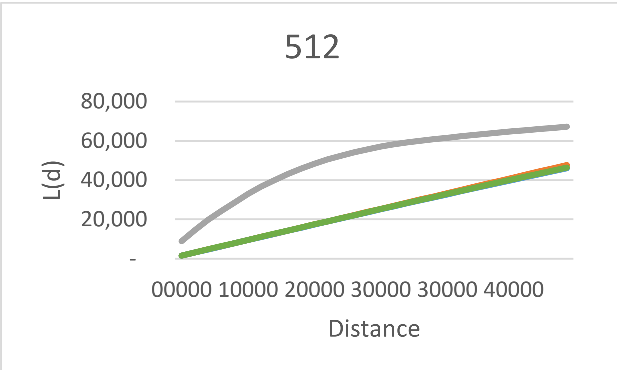
	1950	1960	1970	1980	1990	2000	2010
POPULATION (1,000s)							
Los Angeles, CA	3,997	6,489	8,351	9,479	11,402	11,789	12,151
San Francisco--Oakland, CA	2,022	2,431	2,988	3,191	3,630	3,229	3,281
URBANIZED LAND AREA (sq mi)							
Los Angeles, CA	871	1,370	1,572	1,827	1,966	1,668	1,736
San Francisco--Oakland, CA	287	572	681	796	874	527	524
DENSITY (pop/ sq mi)							
Los Angeles, CA*	4,589	4,736	5,312	5,188	5,800	7,068	6,999
San Francisco--Oakland, CA	7,045	4,250	4,388	4,009	4,153	6,127	6,266

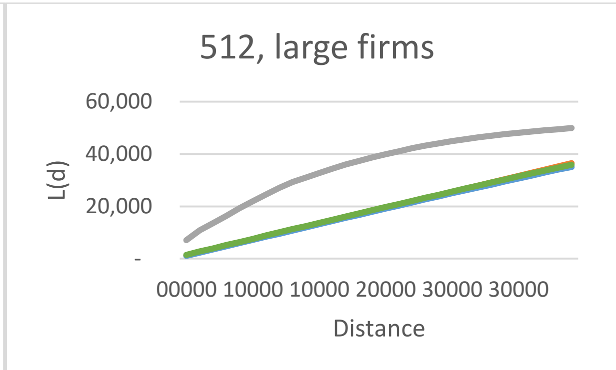
* LA densest in U.S. since 1990

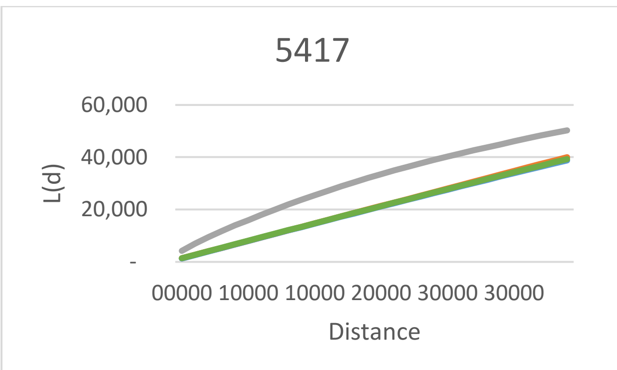
Source: http://www.demographia.com/db-uza2000.htm

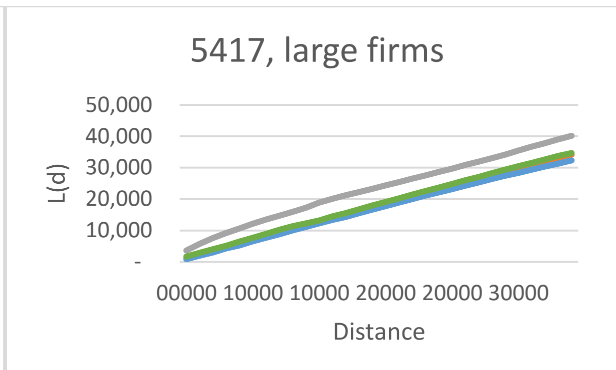
APPENDIX 2: Ripley-K function plots

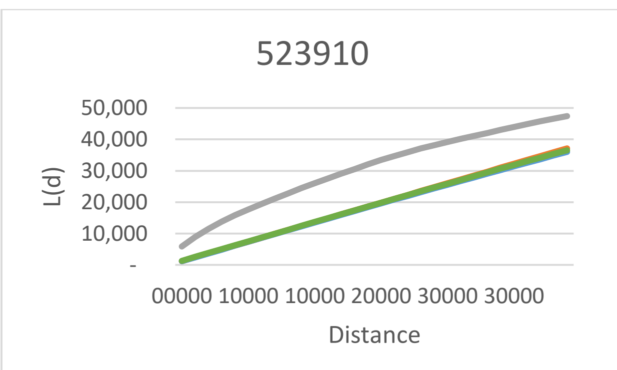
San Francisco Bay Area

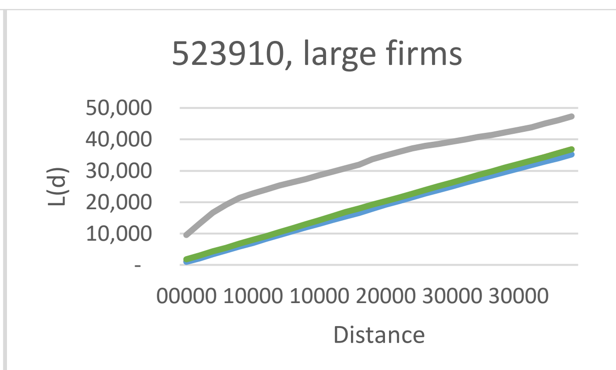


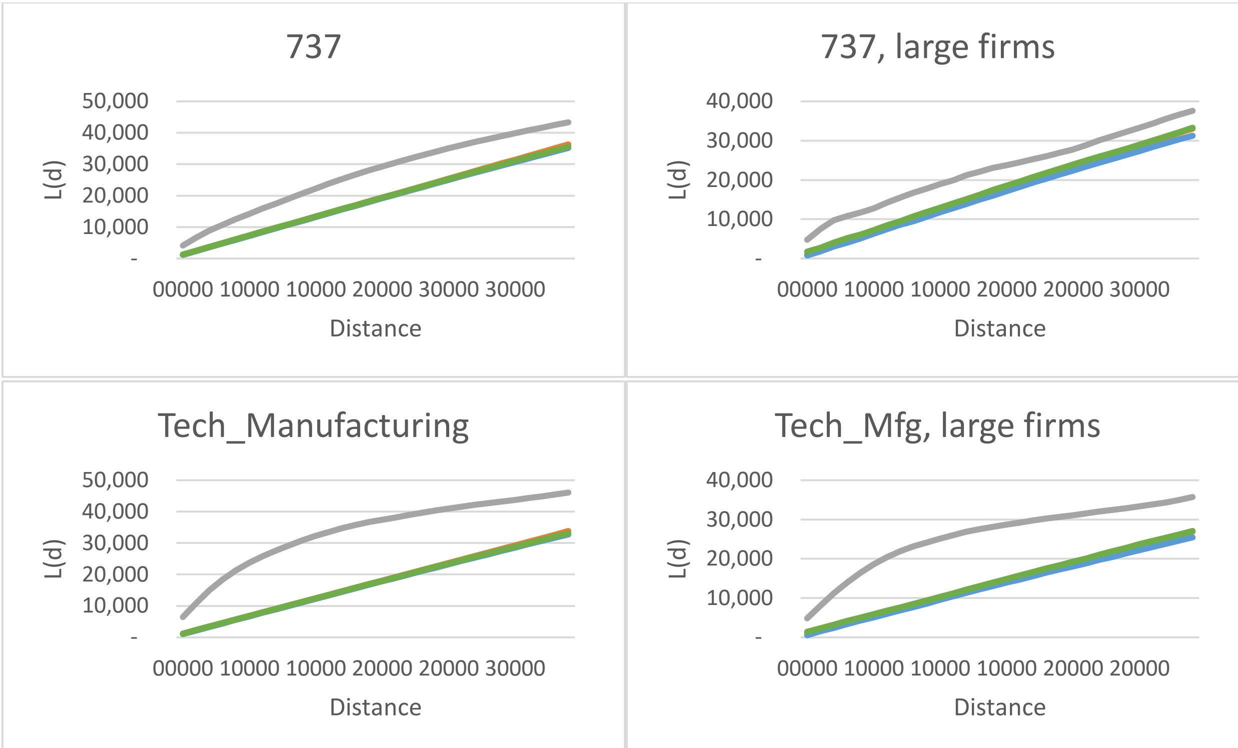




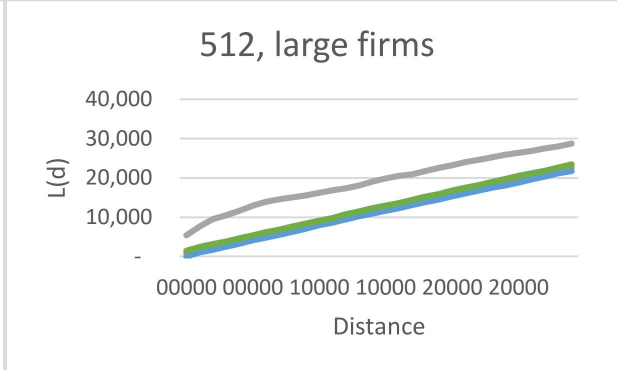


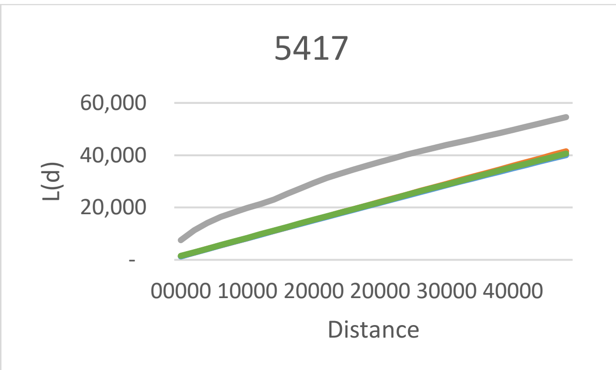


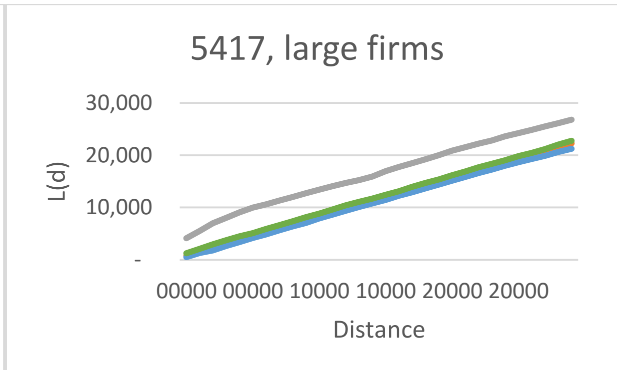


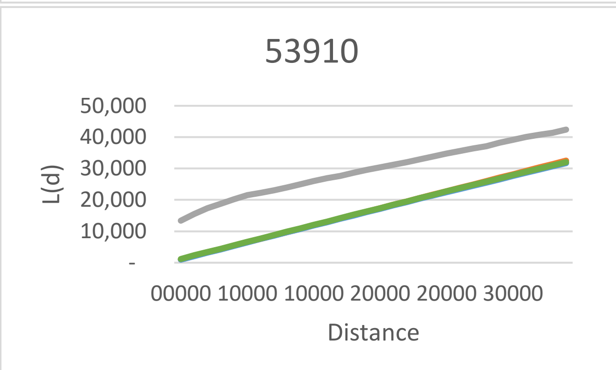

737
50,000
40,000
30,000
20,000
10,000
-
L(d)
00000 10000 10000 20000 30000 30000
Distance
737, large firms
40,000
30,000
20,000
10,000
-
L(d)
00000 10000 10000 20000 20000 30000
Distance
Tech_Manufacturing
50,000
40,000
30,000
20,000
10,000
-
L(d)
00000 10000 10000 20000 20000 30000
Distance
Tech_Mfg, large firms
40,000
30,000
20,000
10,000
-
L(d)
00000 10000 10000 10000 20000 20000
Distance

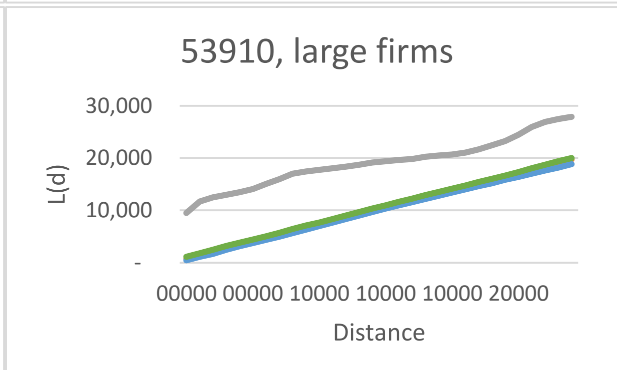
Los Angeles/Orange Counties

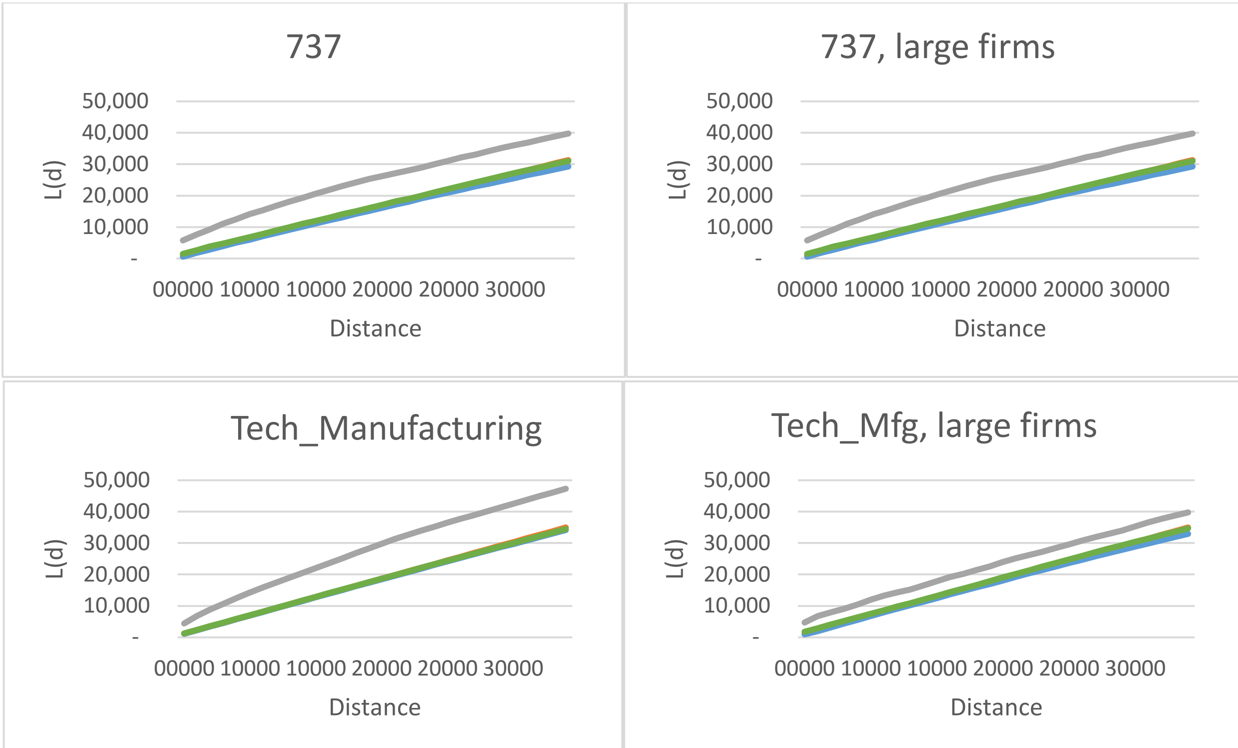


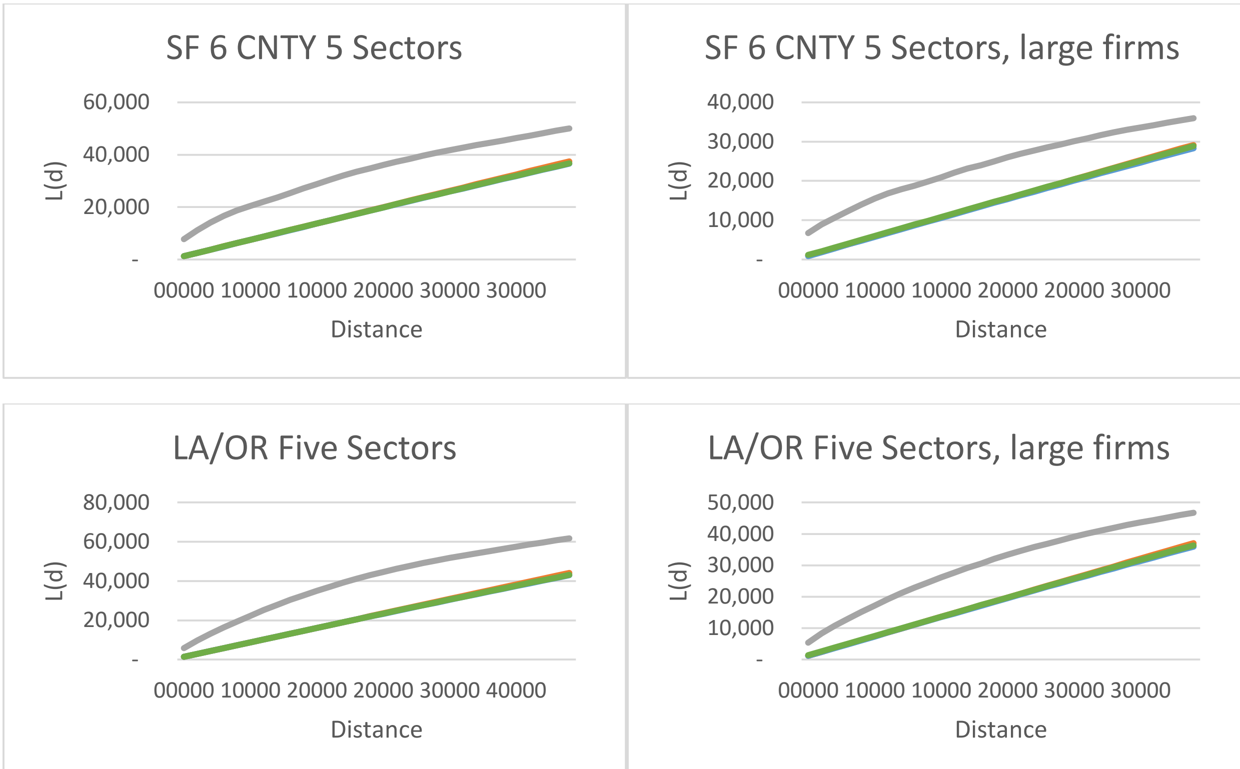









737
737, large firms
Tech_Manufacturing
Tech_Mfg, large firms
50,000
40,000
30,000
20,000
10,000
-
L(d)
00000 10000 10000 20000 20000 30000
Distance


SF 6 CNTY 5 Sectors
60,000
40,000
20,000
-
L(d)
00000 10000 10000 20000 30000 30000
Distance
SF 6 CNTY 5 Sectors, large firms
40,000
30,000
20,000
10,000
-
L(d)
00000 10000 10000 20000 20000 30000
Distance
LA/OR Five Sectors
80,000
60,000
40,000
20,000
-
L(d)
00000 10000 20000 20000 30000 40000
Distance
LA/OR Five Sectors, large firms
50,000
40,000
30,000
20,000
10,000
-
L(d)
00000 10000 10000 20000 30000 30000
Distance