

GEOGRAPHIC DISTRIBUTION OF URBAN RETAILS AND ITS SPATIAL RELATIONSHIP WITH SUBWAY NETWORK:

A Case Study of Retail POI Data in Shanghai

**A Thesis Presented to the Faculty of Architecture and Planning
COLUMBIA UNIVERSITY**

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Urban Planning

by Yuan Hua

May 2018

Advisor: Weiping Wu

Reader: Leah Meisterlin

ACKNOWLEDGMENTS

I would like to thank my thesis advisor, Prof. Weiping Wu, for the guidance, patience and understanding throughout the research process in the past year. In a bigger scope, she encouraged me to concentrate on the big data spatial research and helped me refining the analysis logic. On the other hands, I'm also greatly thankful for Weiping's patient help and detailed comments in great detail for the writing part. I enjoyed the whole research process with Weiping.

Besides my advisor, I would also like to thank Leah Meisterlin as my reader. Leah provided me huge support on the construction of my methodology. The methodology part could not be completed without the technical guidance Leah offered specifically for this thesis and analytical skills I learnt from Leah's Advanced Spatial Analysis course. Her broad knowledge and skills in urban analytics not only helped me with my research but also encouraged me to dive deeper into this field for my future career.

I also want to send my gratitude to my friend Yuanchen Hua, who provided great assistance for the data collection part of my thesis in terms of web scraping. Yuanchen also gave me suggestions and inspirations when conducting the research throughout the whole year.

Last but not least, I am also very grateful to my classmates, families and friends who offered help, support and company during my research process.

ABSTRACT After over two decades of development, Subway has become one of the main transit modes for people's everyday commute, and is considered to attract and form retail clusters around subway stations. Based on a point of interest (POI) dataset of retail stores in Shanghai, this study looks into the geographic distribution of retail stores in Shanghai and examines its spatial relationship with subway network. The retail stores are classified into six types: department stores, specialty stores, convenience stores, restaurants and groceries, recreational facilities and personal service facilities. A series of spatial statistical analysis for measuring geographic distribution, analyzing geographic patterns and mapping clusters are conducted. The spatial pattern shows consistency to the polycentric urban structure of Shanghai and subway stations. In addition, six retail types reveal different extent of spatial clustering at different analysis scales. Last but not least, to qualify the spatial relationship between retail density and other influence factors, including density of population and local bus service, as well as the accessibility to subway station, Ordinary Least Squares (OLS) linear regression is conducted. It turns out that the Kernel Density estimation of retail stores is positively related to population density, bus stop density and existence of subway service, while negatively affected by distance from subway stations. Bus stop density tend to be the most significant influential factors, while department stores enjoy a relatively stronger correlation with distance from subway stations compared to other retail types.

KEYWORDS geographic distribution, retail, subway, kernel density, points of interest data

TABLE OF CONTENTS

INTRODUCTION	1
BACKGROUND	4
LITERATURE REVIEW	8
Retail Spatial Distribution Pattern	8
Effect of Subway on Retail Business	9
Measurement of Spatial Pattern of Retail Business	9
Case of Shanghai	11
METHODOLOGY	12
Data Sources	12
Web Scraping	14
Dataset Cleaning and Processing	16
Data Analysis and Assumptions	18
GEOGRAPHIC DISTRIBUTION OF RETAIL STORES	29
Global Pattern Analysis	29
Local Pattern Analysis	36
SPATIAL RELATIONSHIP OF RETAIL SPATIAL DENSITY AND SUBWAY NETWORKS	44
CONCLUSION AND DISCUSSION	50
Conclusion	50
Interaction between Retail Development and Subway Expansion	52
Limitations	55
BIBLIOGRAPHY	57

TABLES

Table 1 Overview of Original Datasets	13
Table 2 Retail Category	17
Table 3 Variables for OLS Regression	26
Table 4 Results of Average Nearest Neighbor Analysis of Retail Stores	31
Table 5 Distance Band Calculation from Neighbor Count of Retail Stores	32
Table 6 Area within Subway Station and Bus Stop Buffers	42
Table 7 Number of Retail Stores within Subway Station Buffers	42
Table 8 OLS Regression Result	48

FIGURES

Figure 1 Administrative Region of Shanghai	5
Figure 2 Shanghai Subway Development	6
Figure 3 Commercial Centers in Shanghai	7
Figure 4 Urban Transportation Network	14
Figure 5 Map of the Six Types of Retail Stores	18
Figure 6 Sensitivity to Study Area of Average Nearest Neighbor	20
Figure 7 Mean Centers and Directional Distribution Ellipses of the Six types of Retail Stores	30
Figure 8 Spatial Autocorrelation by Distance Yielded by Global Moran's I	35
Figure 9 Hot Spot Map of All Types of Retail Stores in Shanghai	35
Figure 10 Cluster and Outlier Map of All Types of Retail Stores in Shanghai	37
Figure 11 Overlapping Hot Spot Map with (a) High Value Outliers (b) Low Value Outliers	38
Figure 12 Kernel Density Estimation of All Types of Retailers	39
Figure 13 Kernel Density Estimation the Six Types of Retail Stores	41
Figure 14 Subway Station Service Buffer	43
Figure 15 Visualization of Explanatory Variables	44
Figure 16 New Town Development Led by Subway	53

INTRODUCTION

“No matter how good its offering, merchandising, or customer service, every retail company still has to contend with three critical elements of success: location, location, and location” (Taneja, 1999, p. 136). “Being central” is a major principle that can be traced back to the Hotelling’s (1929) classic location problem of ice-cream shops. But how to define “being central”? Different cities adopted different urban structures, with the first ones being quite simple considering some circular and concentric forms. What’s more, new transportations networks have enabled cities to spread out from their original historic cores, generating the new polycentric forms of urban areas, as well as the retail developments (Buliung, 2011; Zhong, et al, 2015). Understanding these polycentric structures is crucial for development of planning strategies, given the strong influence of spatial structure on people’s daily life and economic growth (Anas et al., 1998; Zhong, et. al, 2015).

The opinions vary when it comes to identifying the influential transportation network. Hill et al. (1993) stated that it is the urban street network configuration that shapes the economic and service activities instead of certain attractors such as subway stations or particular destinations. However, subway network, as a major component of the mass transit system in many cities, conveys a large number of commuters and thus supports efficient flow of people between different parts of the city. This flow indicates potential consumers for retailers. For example, urban railway stations in Tokyo, Japan are meanwhile high-density retail clusters due to the long history of transit-oriented developments (TOD) (Calimente, 2012). In this case, proximity to subway stations indicates an attractive location for retailers. The discussion raises the question whether subway network configuration contributes to retail distribution in Shanghai under decades of development.

There is a rich body of literature on the spatial pattern of retail business in Shanghai before and at the beginning stage of the development of subway network, indicating that retail was clustering linearly along major commercial streets and in the old city center, People Square, and would be attracted by local bus service (Li, 2006; Hao, 2007; Wang et al., 2009; Song, Lin, Kong, 2010). But it's not clear how subway may have altered this. In addition, although there is no issued transit-oriented development plan in Shanghai, I still observed that there emerged quite a few large retail store clusters around the subway stations in the past few years. Therefore, this research is to figure out the current geographic distribution of retail stores in Shanghai is after years of development of subway network and to examine how the pattern correlates with the subway network. It is also worthy to see if the effect of subway network has outweighed that of local bus services.

The paper starts with a background overview of subway and retail development in Shanghai, as well as my original observation and hypothesis. Then several spatial statistical methods for measuring geographic distribution, analyzing geographic patterns and mapping clusters are conducted to describe the overall distribution pattern of retail stores. The retail dataset is obtained through web scraping from DIANPING.COM¹ (a counterpart of Yelp in China). To further identify the pattern, the Kernel Density² of retail stores is used as an indicator to demonstrate the current spatial distribution. To further address the effect, I model the spatial

¹ DIANPING.COM (<https://www.dianping.com>) is a website providing information, including location, consumption levels and tripartite reviews, for catering, stores, leisure, local services and entertainment businesses throughout China.

² In statistics, Kernel Density Estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. In spatial analysis, Kernel Density calculates the density of features in a surrounding neighborhood.

relationship to explain retail density with distance from subway stations, with consideration of other influential factors, including population density and local bus service by adopting Ordinary Least Squares (OLS) linear regression. Moreover, whether different types of retail have different distribution patterns and correlation on the given factors will also be examined.

BACKGROUND

Shanghai, located in the Yangtze River Delta along the eastern coastline of China, is one of the four municipalities directly under the control of the central government and one of the most populous cities in China, with a population of more than 24.2 million by the end of 2016 (2017 Shanghai Statistical Yearbook). The total area is 6,341 square kilometers (2,448 sq. mi). As shown in Figure 1 (a), there are three ring expressways in Shanghai. The area inside the middle ring is considered as the inner city area, whereas the area outside the outer ring is defined as suburb. The total area inside the outer ring is 1546.8 sq. km, equaling to 24.4% of the total area, which is generally considered as the urban area of Shanghai. According to the 2017 Shanghai Statistical Yearbook, the total population of the inner city area as of 2016 is approximately 15.67 million (roughly 65% of the total population), with 11.16 million living in the inner city area (within the middle ring). The ring road system, as well as the population concentration, indicates the global single-center urban structure of Shanghai.

In terms of administrative division, Figure 1 (a) shows the 16 municipal districts in Shanghai. The 16 districts are generally divided into three parts. The first two parts refer to the two parts of main city area divided by the Huangpu River: Puxi (“West Bank”) and Pudong (“East Bank”). Puxi usually refers to the historic city center of which Shanghai originally urbanized and developed, consisting of seven denser and more developed municipal districts, while Pudong is administered as the Pudong New Area, merged by the original districts of Pudong and Nanhui in 2009. The other eight districts are considered as suburbs with less development and population.

Despite the global monocentric structure discussed above, as shown by the city centers system in Figure 1 (b) according to Shanghai Master Plan 2017-2035, there emerge several sub-

centers and district sub-centers within the main city area and in the suburban areas along the subway lines, which contributes to the current polycentric urban structure of Shanghai. Shanghai is to draw on the key transportation corridors, especially subway networks to form a “*networked spatial system characterized by multi-centers, clusters, and compactness*” (Shanghai Master Plan 2017-2035, p. 37).

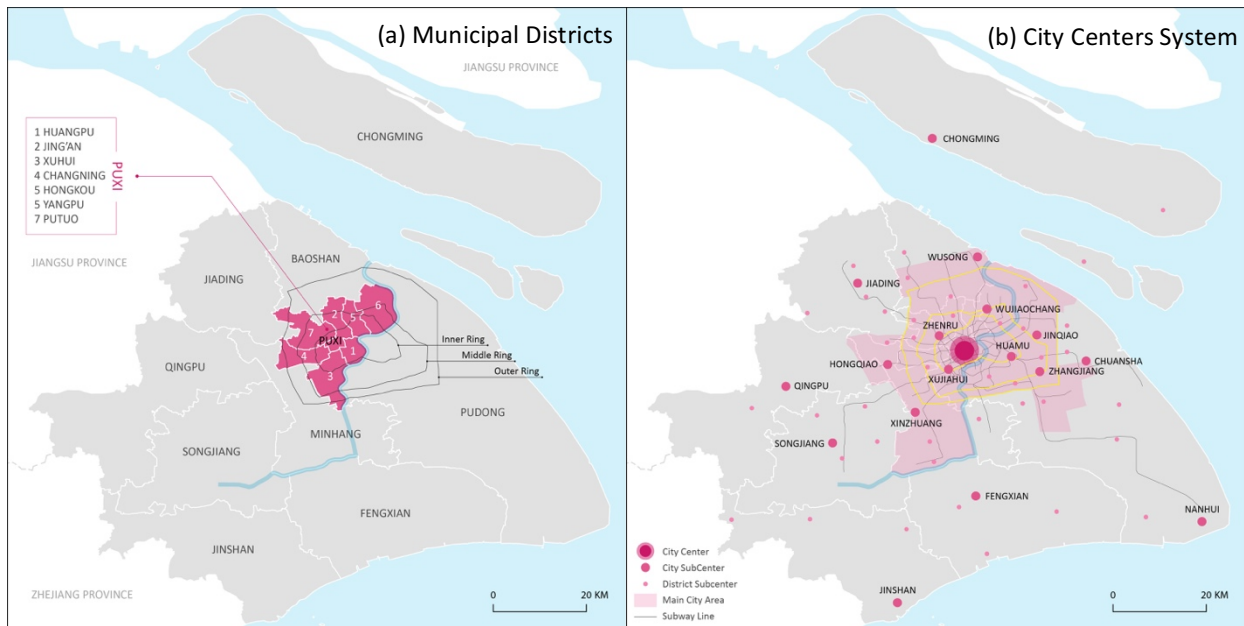


Figure 1 Administrative Region of Shanghai
(Data Source: Shanghai Master Plan 2017-2035)

This city is selected because the long history of Shanghai Metro and the retail prosperity. The Shanghai Metro, opening in 1993 with full-scale construction tracing back to 1986, is the third-oldest rapid transit system in China (“Shanghai Metro”, 2018). Figure 2 shows the development of Shanghai subway network. The network originates in the city center in 1993 and expands radically to the outer city area, with one ring route (Line 4). The structure of the subway network, adding to the ring road system shown in Figure 1, indicates the original monocentric urban structure of Shanghai. As the urban development and subway expand, new city sub-centers or new towns were developed and connected by subways to the city center.

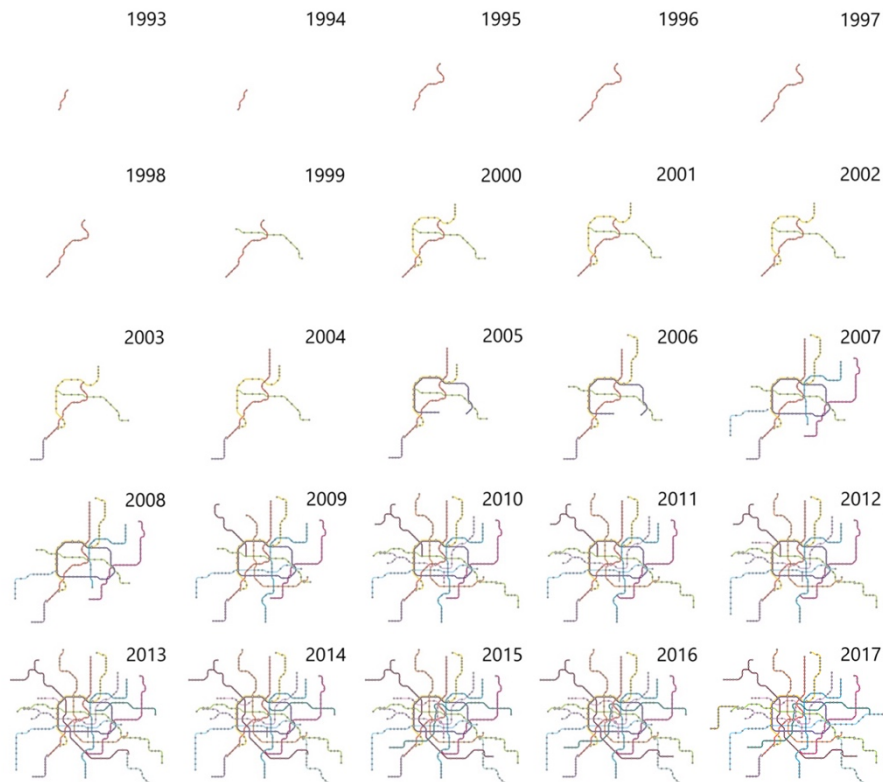


Figure 2 Shanghai Subway Development
 (Source: “Shanghai Metro”, Wikipedia)

The subway network in Shanghai serves 13 of its 16 municipal districts and has extended to Jiangsu Province. Shanghai has the longest distance of coverage and second largest by number of stations with 14 lines and 364 stations totaling 588 kilometers (365 mi) by 2015, surpassing London and New York (Riedel, 2014; CityMetric, 2015). In addition, subway service in Shanghai ranked second to Beijing in the *World Ranking of the Annual Ridership of Urban Subway Transit* in 2016 provided by Wikipedia (2018). According to the *Ranking*, the annual passenger volume reached 3,401 million in 2016, equaling to 9.29 million per day. On March 9th, 2018, the per day ridership peaked at 12.23 million (Shanghai Metro, 2018). Given the fact that the population of Shanghai is 24.20 million in 2016 and 24.19 million in 2017 (“Demographics of Shanghai”, 2018), there is large and ever-growing dependence of everyday modes of commute on subway service. On the other hand, the three largest service industries are financial services,

retail, and real estate in Shanghai since 2011 (“Shanghai”, 2018), which means that retail is one of the supporting industries of Shanghai, which lays the foundation of this study.

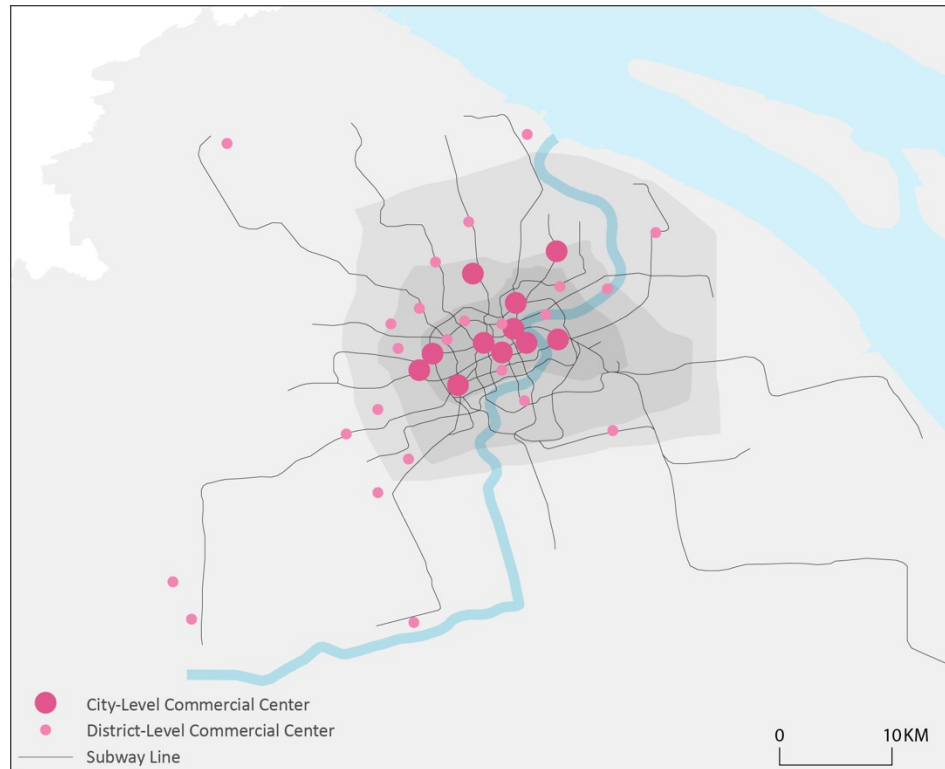


Figure 3 Commercial Centers in Shanghai
(Data Source: Shanghai Master Plan 2017-2035)

The large number of commuters brought by the subway may lay the foundation of retail growth, which indicates the potential correlation between subway service and retail development. Based on my personal observation, along with the development of subway network, quite a lot of transit-oriented retail spaces have clustered and grown around the subway stations. Figure 3 shows the 36 commercial centers in the inner city of Shanghai, 11 of which serve at city-level, overlaid upon the subway network. It is obvious that most of the commercial centers are located near and thus served by subway lines. Previous studies have looked into the influence and the changing pattern; however, little research is based on big data and quantitative analysis that explicitly examine the retail clusters and their correlation with subway network.

LITERATURE REVIEW

Since the thesis is looking into the spatial patterns of retail stores and their clustering phenomenon around subway stations, this literature review briefly summarizes the classical theories of retail spatial patterns, and then highlights how subway has affected distribution of retail business. Methods adopted in previous research to measure geographic distribution patterns are also discussed. Finally, similar research focusing on Shanghai are discussed.

Retail Spatial Distribution Pattern

Early theories understand and summarize the geographic distribution patterns of retails in different ways. In 1931, William Reilly was inspired by Newton's Law of Gravity and created an application of the gravity model, Law of Retail Gravitation, to measure the retail trade between two cities. In Reilly's formulation, the "attractiveness" of the two retail centers is postulated to be proportional to their size and inversely proportional to the square of the distance to them. Each retail center will thus form a certain trade area of attracting consumers (Rosenberg, 2017).

Walter Christaller's Central Place Theory shares the similar idea of identifying retail distribution. This theory was given on the basis of Christaller's study of settlement patterns in southern Germany, pointing out the central place exists primarily to provide goods and services to its surrounding population. The higher the level of the central places are, the less those central places are needed indicates and the longer distance from each other (Briney, 2017).

Looking specifically into urban retail spatial distribution, Gertis (1961) and B. J. Garner (1966) stated that the retail volume decreases as the distance from city center increases, and the location of retails significantly determines its profitability (Wang, Zhang & Ma, 2009). As Hillier, et al. (1993) state, urban non-residential economic activities tend to be determined by the structure of

the urban networks rather than by the presence of specific attractors or magnets, meaning that the retails tend to be linear or networked distribution instead of centric.

Nowadays the development of marketing geography (geomarketing) and retail geography further look into where to place retail stores based on where their customers are through use of Geographic Information System (GIS), which further indicate the meaning of integrating geographical intelligence into retail organization process.

Effect of Subway on Retail Business

Subway network tends to be one of the key attractions for retail location selection. Since the subway network developed throughout the world, researchers have started to figure out the effect of subway on retail business. Hu (2007) conducts qualitative analysis and concludes three types of effect: first, the enhancement of traditional trade areas with service of subway lines; second, establishment of new trade areas at transit hubs; and third, recession of original trade areas without service of subway lines. Tsou & Cheng (2013) propose a retail spatial integrated model (RSIM) to examine the relationship between retail patterns and urban network structure, including street configuration, bus network and metro network in Taipei, China. In addition to the overall influence on retail business, Hou (2014) analyzes the radiation and interaction pattern of neighboring trade areas: independent radiation, interconnection radiation and embedded radiation.

Measurement of Spatial Pattern of Retail Business

In terms of the measurement of spatial pattern of retail business, density and diversity in retail categories are the two most common ways. Jiao et al. (2015) conduct Kernel Density analysis on a Point of Interest (POI) dataset scraped from Baidu Map to understand the spatial pattern of

retail business in Guangzhou, China, and conclude that public transportation stations are the key element influencing retails. Strano et al. (2007), based on, similarly, Kernel Density evaluation, to testify how much higher street centrality statistically determines a higher presence of activities in terms of shops and services in Bologna, Italy. The street centrality is also measured through Multiple Centrality Assessment (MCA). Regression is run between centrality and activities. Yu et al. (2015) also based their analysis of CBD in Shenzhen and Guangzhou on Kernel Density Estimation. Wang et al. (2014) examine the relationship between the density of retail stores and the street centrality in Changchun, China and reach the conclusion that the street centrality captures location advantage in a city and plays a crucial role in shaping commercial land use intensity.

Besides density, diversity is also adopted as an indicator of spatial pattern. Huang & Chen (2014) used the Shannon information entropy model, equilibrium degree model and dominance degree model to discuss the link between urban railway system and retail business structure of metro sites in Guangzhou, and to evaluate the diversity of retail business. The results showed: firstly, the retail stores in the near suburbs are of the most diversity; secondly, the transfer stations show the most balanced retail business structure; and finally the later the metro site open, the more balanced the retail business structure is.

In addition, other research methods are also investigated. Ertekin et al. (2008) investigate the spatial distribution of shopping malls in Istanbul through a regression analysis between the shopping mall space and income, population and distance to CBD. The results reveal that income is the only factor affecting the location of shopping malls.

Case of Shanghai

In the past few years, quite a few researches discussed how subway has influenced the retail spaces specifically in Shanghai, most of which depended on qualitative or basic spatial analysis. Development of the subway network in Shanghai is widely considered as one of the key elements influencing urban retail distribution; more precisely, the spatial pattern of retail spaces is transforming from punctuate or linear clusters into multi-central and reticulate commercial circles around subway stations (Li, 2006; Hao, 2007; Wang et al., 2009; Song et al., 2010). Wang et al. (2009) further states that not only the spatial form of the retail changes, but also its component. The new retail circles are transforming into multi-functional consumption complexes, providing retails, restaurants, recreations and hotels. Song et al. (2010) point out that the retail spaces tend to cluster more around subway hub stations, and the extension of subway lines also lead the expansion of retail development to suburban areas. However, the existing researches that specifically focus on Shanghai lack quantitative analysis, which can better demonstrate and visualize the spatial pattern of retail business.

METHODOLOGY

To understand the spatial pattern of retail stores and its correlation with transit nodes, I collected data of 353,882 retail stores, boundaries of 16 districts and 230 sub-districts, population, subway network and bus network in Shanghai. The data was cleaned and reorganized for further spatial statistical analysis. To better examine the correlation, I conducted Ordinary Least Square (OLS) regression analysis to qualify to what extent retail density is influenced by other factors. The data processing and analysis was mainly carried out through ArcGIS and Python in Jupyter Notebook.

Data Sources

There are six main datasets required for this research (see Table 1). The first is the Point of Interest (POI) dataset including the information of retail stores extracted from DIANPING.COM updated by January 2018. This is because there is no public data source providing retail data. It is noteworthy that the DIANPING database isn't a complete list of all retails because it does not include some newly-opened and seldom-searched stores; thus these stores are excluded in this research. The coordinates of the retail stores were obtained through the Geocoder API provided by AMap (a counterpart of Google Map in China). The second and third dataset of the polygon shapefiles of 16 districts and 230 sub-districts¹ in Shanghai are from Urban Data Party². The 2010 population is provided by 2010 Chinese Census, with spatial unit at sub-district level. The 2016 population at district level is provided by 2017 Shanghai Statistical Yearbook. These two

¹ Sub-districts (Chinese: 街道; pinyin: jiē-dào, literally "Street") is one of the smaller political divisions of PRC. In general, an urban district is further divided into several sub-districts, consisting several residential communities or neighborhoods.

² Urban Data Party (www.udparty.com) is a public online platform dedicating to urban big data sharing and analytics.

datasets are used to calculate the 2016 population at sub-district level. The other two datasets, also provided by Urban Data Party, consisting of 17 subway routes and 368 stations, and 1,388 bus routes and 29,912 stops in Shanghai updated till October 2017. These two datasets are updated to October 2017. It is worth noting that, due to unavailability of the street network data, the bus network, which covers most of the main roads, was adopted to be a proxy for street network. Moreover, the bus stops were considered as an indicator of urban activity and local transportation service because bus services are densely located in the city and in proportional to the amount of human traffic (not only residents, but also tourists, passerby, workers or users of facilities) of the specific serving area. In general, the more bus service, the busier and more populated the area. Since the coordinates provided by AMap are under GCJ-02 geodetic system¹ (colloquially Mars Coordinates), all the four datasets were projected to GCJ-02. It is a geodetic datum formulated by the Chinese State Bureau of Surveying and Mapping (Chinese: 国测局; pinyin: guó-cè-jú) based on WGS-84 (Kunpeng Web Scraping).

Table 1 Overview of Original Datasets

Dataset		Source	Count	Attributes	Data Format
Retail	Store	DIANPING.COM	353,882	Name, number of customer reviews, consumption level, retail category, trade area, address	Table
	Coordinate	AMap API	324,660	Coordinate	Table
District		Urban Data Party	16	District name, area	Shapefile (polygon)
Sub-district		Urban Data Party	230	Sub-district name, area	Shapefile (polygon)

¹ GCJ-02 uses an obfuscation algorithm which adds apparently random offsets to both the latitude and longitude, with the alleged goal of improving national security.

Table 1 (continued)

Dataset		Source	Count	Attributes	Data Format
Population	2010	2010 Chinese Census	230	2010 population by sub-district	Table
	2016	2017 Shanghai Statistical Yearbook	16	2016 population by district	Table
Subway	Station	Urban Data Party	368	Name, id, coordinate, route no.	Shapefile (point)
	Route	Urban Data Party	17	Route no., id, coordinates for origin and destination stations, total number of stations, total length, operating time	Shapefile (polyline)
Bus	Stop	Urban Data Party	29,912	Name, id, coordinate, route no.	Shapefile (point)
	Route	Urban Data Party	1,388	Route no., id, coordinates for origin and destination stations, total number of stations, total length, operating time	Shapefile (polyline)

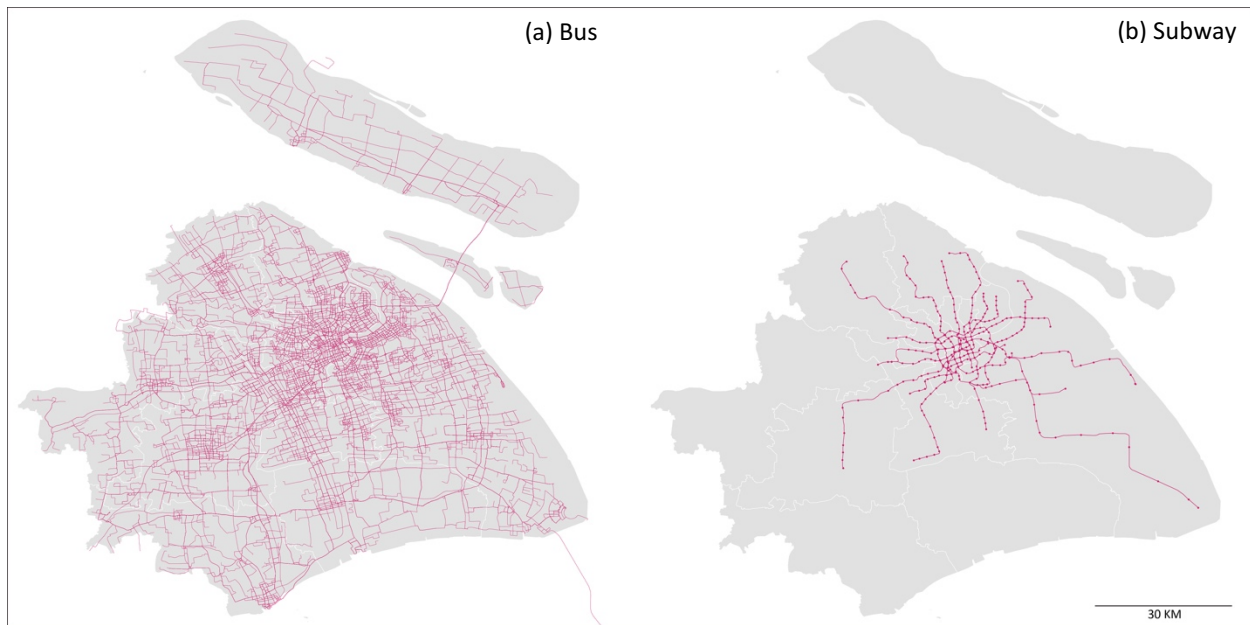


Figure 4 Urban Transportation Network

Web Scraping

Web scraping is a method that allows the computer automatically to extract the data from a certain website content. A web scraper simulates a human's manual examination and copy-and-paste to access web pages, turn pages, find specific data elements on the page, extract the data and save the data as a structured dataset. In this project, a web scraper was built to automatically

collect information shown on the web pages of DIANPING.COM. The scraper was built through Bazhuayu¹. The scraper works as following:

STEP 1: Visit the publicly available index page of DIANPING.COM.

STEP 2: Click the link of one retail category and visit the content page showing information of each of specific category (i.e., If label restaurant clicked, the scraper will visit the page showing a list of restaurants and the related information including name, location, etc.).

STEP 3: Receive HTML data back from the web server that defines the content of web pages.

STEP 4: Extract the required data elements from the HTML using the XPath² query language.

STEP 5: Save the data to a structured dataset on a hard drive.

STEP 6: Click *NEXT PAGE* till the last page to loop through all pages of that specific category.

STEP 7: Click the link of the next retail category and execute STEP 3-6.

During the data gathering, I gathered information of 353,882 thousands retail businesses across the 16 districts of Shanghai. This dataset covers every retails updated by January 2018 (the period the web scraping processed) on DIANPING.COM. The extracted dataset was used for subsequent analysis after cleaning and reorganizing.

¹ Bazhuayu (Chinese: 八爪鱼) (www.bazhuayu.com) is a web scraper builder that provides code packages for non-specialists.

² XPath (XML Path Language) is a query language for selecting nodes from an XML document. It may be used to compute values (e.g., strings, numbers, or Boolean values) from the content of an XML document. (<https://en.wikipedia.org/wiki/XPath>)

Dataset Cleaning and Processing

The 353,882 records of raw data retrieved from DIANPING.COM and AMap requires substantial filtering, cleaning and reclassification. The first step was the identification and flagging of duplicate items in Python. The data was extracted by retail category classified by DIANPING.COM, which allows duplicates in different categories. For example, the bakery stores were included in both *Retail Stores* and *Restaurants*. Considering that there might be retail stores with same name across Shanghai; thus, I removed the records with both same name and same address. Finally, 324,660 rows of data (91.7% of the original dataset) were retained for further processing and analysis.

Next, the retails were geocoded through the Geocoder API provided by AMap. The Geocoder API requires either retail name or retail address and returns coordinates. Given the fact that the description of address provided by DIANPING.COM is not always normative – some addresses are described in terms of “*near the entrance*” or “*at the opposite side*”, I request coordinates by providing retail names. After displaying the coordinates of the retails geocoded by AMap, some retails were observed locating outside the boundary of Shanghai. These retails were re-geocoded manually in the Coordinate Picker¹ tool provided by AMap.

¹ The Coordinate Picker tool of AMap (<https://lbs.amap.com/console/show/picker>) works similarly as the Geocoder API, which requires name or address and returns coordinate. The Picker provides user interface with the actual map, allowing me to manually compare and match the coordinates picked with the location shown on DIANPING.COM. So this tool was adopted for correcting the falsely geocoded locations.

Table 2 Retail Category

Category	Count	Percent	Description
Department Store	845	0.26%	Comprehensive shopping centers
Specialty Store	121,328	37.37%	Store catering to one specific retail market; for example, clothing, furniture, home appliances, book stores, cosmetics, houseware, sporting goods. (These stores might sit inside the department stores.)
Convenience Store	18,623	5.74%	Small retail businesses that provides everyday items.
Restaurant/Grocery	139,890	43.09%	Restaurants including bakeries, fast-food restaurants, cafes, etc. and other retail stores providing fresh food or fruits.
Recreational Facility	29,364	9.04%	Cinemas, karaoke, archery stadiums, etc.
Personal service Facility	13,389	4.12%	Laundries, drugstores, pet shops, barbershops, daycare centers, parent-child centers, etc.
Total	324,660	100%	

In order to examine if different retail categories yield different spatial distribution patterns, the retail businesses were reclassified into six categories referring to *Classification of Retail Formats (GB/T18106-2004)* issued by Ministry of Commerce of PRC in 2004, including *Department Store, Specialty Store, Convenience Store, Restaurant/Grocery, Recreational Facility and Personal service Facility*, as specified in Table 2. The *Classification of Retail Format* originally includes 17 categories. Due to data inaccessibility, the categories, such as *Vending Machine, Online Shop, Teleshopping*, were not included in this project. What is noticeable here is that The *Classification* revised and included the retails categories related to online shopping in order to better reveal the changing trends of consumption custom and retail formats starting from twentieth century in China. Exclusion of these categories lays one of the limitations of this research, for I anticipated a different dependence on location and urban transportation network for these categories and thus a different spatial distribution patterns from the conventional retails (e.g. department stores and convenience stores) included in this research.

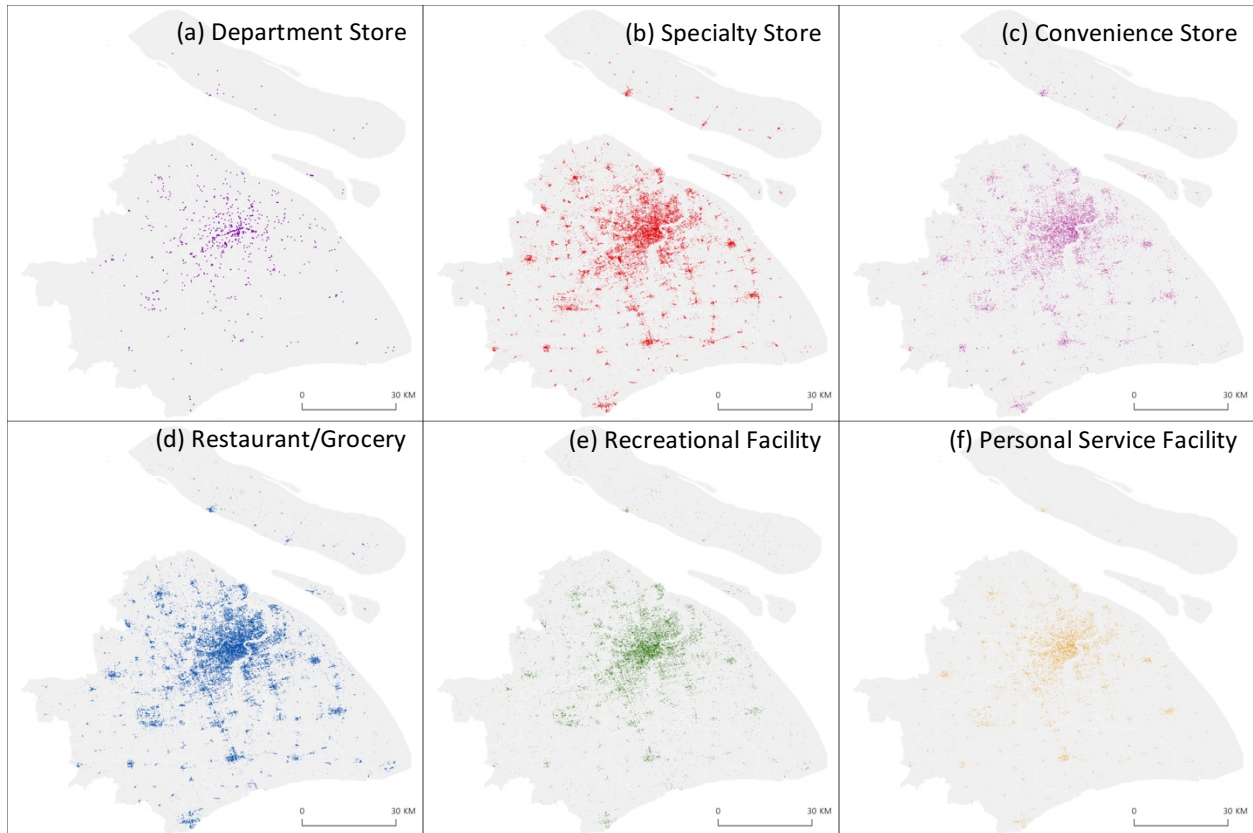


Figure 5 Map of the Six Types of Retail Stores

Data Analysis and Assumptions

To understand the geographic distribution pattern of retail stores, three types of analysis were conducted. Firstly, global spatial statistical analyses were conducted to provide an overview of the spatial pattern, followed by local analyses to observe smaller-scale distribution pattern. Finally, a linear regression analysis was conducted looking deeper into the spatial correlation between retails and other influential factors. The spatial analysis was conducted in ArcGIS using the toolsets of *Measuring Distribution*, *Analyzing Patterns* and *Mapping Clusters in Spatial*

Statistical toolbox and toolsets of *Density* and *Distance* in *Spatial Analyst* toolbox¹ and the statistical analysis was performed via Python in Jupyter Notebook.

Global Pattern Analysis

The descriptive analysis begins with two geographic distribution measures – “*Mean Center*” and “*Standard Distance*” – in ArcGIS. The “*Mean Center*” tool identifies the geographic center (or the center of concentration) for a set of spatial features, which can be considered as the average location of retail stores. The “*Standard Distance*” tool measures the degree to which features are concentrated or dispersed around the geometric mean center, as a radius equal to one standard deviation of the distances of retail stores from the mean center as calculated by the “*Mean Center*” tool. Next, the *Directional Distribution (Standard Deviational Ellipse)* method advances the concepts of “*Mean Center*” and “*Standard Distance*” by identifying the central tendency, dispersion and directional trends captured by the standard, so that the directional variation of spatial spread can be observed, e.g., the stores are more spread along the south-north axis than others. These tools are used to examine the geographic distribution of each store type as specified previously. Given the global scale monocentric urban structure discussed previously, I anticipated the mean centers to locate around the city center, People Square, or around the geographic center of Shanghai. In addition, the ellipses were expected to spread out evenly along each direction. In particular, the distribution of convenience stores and personal service facilities were expected to be most dispersed from their center (i.e., biggest ellipses) because they are assigned according to residential estates, which makes them less sensitive to the central location.

¹ The description and explanation of the spatial analysis tools is mainly referred to the tool reference provided by ESRI.

The other four types of retail stores, especially restaurants and specialty stores, were expected to be more compact around their center given that they tend to benefit from their collective offering of diverse and highly specialized commodities.

However, these patterns are merely descriptive and reveal limited information of stores' spatial pattern. As a result, I looked into the distance between retail stores to further identify the distribution pattern using "Average Nearest Neighbor". The "Average Nearest Neighbor" index measures the average distance from each feature of its nearest neighboring feature. The smaller the distance, the closer the stores. This index reveals whether individual stores are within proximity (considered as "clustered") or far apart from each other (considered as "dispersed"). If the index ("Nearest Neighbor Ratio") is less than 1, the pattern exhibits clustering. If the index is greater than 1, the trend is toward dispersion. A smaller average distance of a specific type of retails (i.e., "Observed Distance") in comparison with a random distribution of the same number of stores (i.e., "Expected distance") indicates a clustered pattern, and a larger distance represents a dispersed pattern (Mitchell, 2005). In addition, the Z-score provided by the "Average Nearest Neighbor" index refers to whether the identified cluster or dispersed pattern is statistically significant instead of the result of random chance.

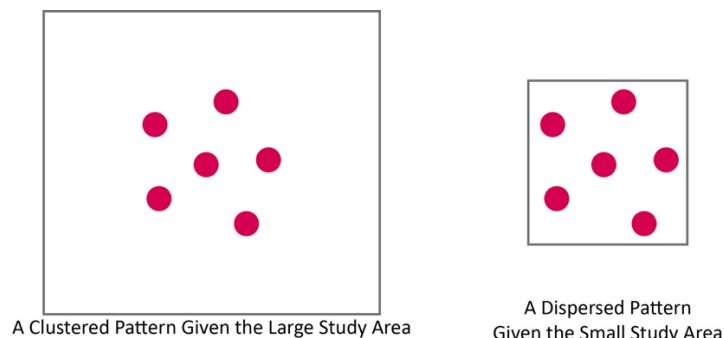


Figure 6 Sensitivity to Study Area of Average Nearest Neighbor

Since the result of this method is very sensitive to the *Area* value (see Figure 6) – small changes in the *Area* parameter value can result in considerable changes in the z-score results, two *Area* values were set. First, I set the *Area* value to cover the whole area of Shanghai in order to see at city-scale if the retail stores are clustered or dispersed. In this test, I anticipated all retail types to reveal a spatial clustering, given the large study area. For the same reason discussed above, specialty stores and restaurants were anticipated to yield the largest negative Z-scores, indicating the most clustered pattern, for the same reason discussed above, while the spatial pattern of convenience stores and personal service facilities tend to be less statistically significant. Then another test is run to observe the spatial pattern of the area inside the outer ring (the urban area of Shanghai), excluding the retail stores in suburban areas that were observed to be more dispersed in Figure 5. All six types of retail stores are observed to geographically concentrate within the outer ring, so I assumed that some types of retail stores would reveal a dispersed pattern in this test. Although they are clustering at city-scale, they are actually evenly distributed in the main city area.

To further identify the geographic patterns of retail stores as distance changes, the tool “*Incremental Spatial Autocorrelation*” was adopted¹. The definition of Spatial Autocorrelation

¹ Here provides the parameter setting when I adopted the “*Incremental Spatial Autocorrelation*” analysis. The parameters of Beginning Distance and *Distance Increment* required for this analysis were derived from the “*Calculate Distance Band from Neighbor Count*” tool, which calculates the minimum distance, average distance (same as calculated in the “*Average Nearest Neighbor*” index) and maximum distance between all input retail stores. The *Beginning Distance* was set at the maximum distance (if smaller than this distance, there will be at least one retail stores with no neighbor) as the starting point for the “*Incremental Spatial Autocorrelation*” analysis, and the *Distance Increment* was set at the average distance as the processing intervals.

provided by ESRI is “a measure of the degree to which a set of spatial features and their associated data values tend to be clustered together in space (positive spatial autocorrelation) or dispersed (negative spatial autocorrelation)” (“Spatial Autocorrelation”, 2018). This tool measures the spatial autocorrelation for a series of distances and calculates the Z-scores reflecting the intensity of spatial clustering as in the “Average Nearest Neighbor” index. A peak in Z-scores indicates the distance at which the spatial clustering is most pronounced. In this project, this tool helps to testify how the clustering pattern at different scales (indicated by the distance values), or to say, if retail stores are clustering within a specific range of distance or are continuously clustering as distance goes up. The statistical significance of the clustering pattern of the six types of retail stores were expected to reveal different trends. Here a few assumptions were made. On the one hand, department stores, specialty stores and recreational facilities were anticipated to show a stronger clustering pattern at a smaller scale, meaning that they tend to cluster within a smaller area to provide diverse commodities while as the study scale grows, the clustering pattern become less significant because the clusters themselves tend to sit away from each other to prevent competition. Particularly, the anticipation about department stores was yield given the limited quantity – there are only over 800 department stores, and according to Figure 5, apparently they are not limited within the city center area. On the other hand, convenience stores, restaurants/groceries and personal service facilities were anticipated to show a continuously significant clustering pattern as the distance increases.

After answering the question “Is retail stores clustering?”, I identified and visualized the locations and extents of the statistically significant hot spots and cold spots. In this project, the

clusters are mapped based on number of retail stores¹. *Hot Spot Analysis (Getis-Ord Gi*)* was conducted to identify statistically significant spatial clusters of retail stores. Clusters of high values (aggregated number of retail stores) are considered as hot spots, whereas those of low values are identified as cold spots. This tool calculates the Getis-Ord Gi* statistic² (a z-score), showing whether the observed spatial clustering of high or low values is more pronounced than one would expect in a random distribution of those same values. The larger the Gi* statistic is, the more intense the clustering of retail stores (hot spot). The result of Hot Spot Analysis is then visualized through Inverse distance weighted (IDW) interpolation to generate a continuous surface based on the sampled data at retail store points.

Local Pattern Analysis

Because the *Hot Spot analysis* cannot identify statistically significant spatial outliers (a high value surrounded by low values or a low value surrounded by high values), I then conducted the *Cluster and Outlier Analysis (Anselin Local Moran's I)*³, which distinguishes between a

¹ Here specifies the calculation of the required input parameters of *Hot Spot Analysis (Getis-Ord Gi*)* and *Cluster and Outlier Analysis (Anselin Local Moran's I)*. Since these two tools require Input Field containing different values, the incident data of the POI dataset has to be aggregated based on location (in the original dataset value of each point is 1 because there is only one retail store in one location). Points of retail stores within a x, y tolerance at 100 meters were considered as identical and the aggregated number of retail stores was calculated using Integrate with Collect Events tool in ArcGIS.

² The calculation of Gi* statistic is explained in <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm>.

³ The calculation of Local Moran's I statistic is explained in <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm>.

statistically significant cluster of high values, cluster of low values, outlier in which a high value is surrounded primarily by low values, and outlier in which a low value is surrounded primarily by high values at 95 percent confidence level. With this tool, I'm able to identify the outliers (high-density retail areas) in suburban areas where there is in general no cluster of retails.

To evaluate the actual density of retail stores, the *Kernel Density* method in ArcGIS is used to calculate the magnitude per unit area (i.e., density) from point features of retail stores using a kernel function to fit a smoothly tapered surface to each point. The kernel function allows the measure to weigh nearby objects more than far ones (Wang et al., 2014). I employed a cell size of 100 m × 100m to create the raster layer covering a rectangular area over the municipal area of Shanghai consisting of 1,401,700 grid cells (1,070 columns by 1,310 rows). The rectangular raster layer was later clipped to the boundary of Shanghai to get rid of the grids with no data. The output raster layer of *Kernel Density* converts the original point layer of discrete events (store locations) to a continuous surface (Wang et al., 2014). The main difference between the hot spots and the high-density areas here is that the high-density areas only indicating the area with high values themselves, while to be identified as a statistically significant hot spot, the area will have a high value and be surrounded by areas with high values as well. Density was calculated for all types of retails and for each type to testify if one type of stores tends to have a specific distribution pattern.

Last but not least, since this research is trying to identify the spatial relationship of retails with subway stations, distance from subway stations is also an important reference in analysis. A 500-meter and a 1,000-meter buffer were created from each subway stations, where the 500-meter buffer was considered as the core area with subway accessibility and the 1,000-meter buffer the threshold area people want to walk away from subway stations. Total number of retails stores

was calculated within the buffer areas to provide an overview on how many retail stores are located within an endurable walking distance from subway stations; in other words, how many retail stores are clustering around subway stations. To further examine the clustering effect of subway stations of different types and/or of different locations on retail stores, the average distance of the nearest 50 and 100 retail stores from each subway stations was calculated by the “*Near Table*” tool provided by ArcGIS. A new buffer area was created in accordance to the average distance calculated, which can be interpreted as another way illustrating retail density – the smaller the buffer area, the denser the retails are around that subway station.

Modeling Spatial Relationship

In this section, a correlation analysis and several Ordinary Least Squares (OLS) regressions were conducted to quantify the correlation between retail density and urban network. I carried out several OLS linear regressions to quantify the correlation of retail density on the explanatory variables. It is worth noting here that it is likely statistical overkill to sample every single cell of the raster layers. As a result, instead of using all raster cells, I create 2,000 random sample points in Shanghai for the regression¹. The dependent variable is the *Kernel Density* estimations as explained previously with the search radius at 1,000 meters – a distance that I considered as an endurable walking distance for people to search for retail store. Thus the retail density within

¹ Here explains how the 2,000 sample points are created. First, a random point layer of 2,000 points covering the whole area of Shanghai was created with the *Create Random Points* tool in ArcGIS. Then the value of the raster layers was assigned to the point features using the *Extract Multi Values to Points* tool.

1,000 meters is perceivable for people as a retail cluster, and is used as dependent variable for the regression.

Table 3 Variables for OLS Regression

Variable	Description & Calculation
Population Density 2016 by sub-district	Population density by sub-district for 2016 was projected as following: population by sub-district for 2010 * population percentage change by district from 2010 to 2016 / area of each sub-district. The density is measured in thousands of people per square kilometer.
Euclidean Distance from a Nearest Subway Station	The distance from subway stations was generated by the <i>Euclidean Distance</i> tool in ArcGIS. The value was then reclassified to 1-5 using walking increments of 500 meters, with ranked scores from one (within 500 meters from a subway) to five (greater than 2,000 meters from a subway). The distance variable was converted into an interval variable because what I'm not trying to obtain the actual slight distance difference. A distance within 500 meters was considered as most walkable, whereas farther than 2,000 meters away was considered as absolutely unwalkable.
Availability of Subway Service	Binary variable generated by the previous distance variable, showing if the random point is within 1,000 buffer of a subway station. If so, it is assigned as 1, indicating that it is covered by subway service; if not, it is valued as 0.
Kernel Density of Bus Stops	The Kernel Density of bus stops was considered as a proxy for urban activity and local transportation service. The density was also calculated under cell size at 100 m × 100m and search radius at 1,000 meters as other previous density variables.
Line Density of Bus Routes	The density of bus routes, as a proxy for street density, was calculated by the <i>Line Density</i> tool in ArcGIS under cell size at 100 m × 100m and a search radius at 1,000 meters.

The independent variables are listed in Table 3, including the population density, availability of subway service, straight-line distance from subway station, bus stop density and bus route density. Population is taken into consideration in the regression as an proxy for potential consumer basis of retail stores, and the indicators of the subway and bus service are included to testify how the retail development is influenced by transportation network. Moreover, this is also to examine which network has a stronger effect as discussed before.

All variables except availability of subway service were first generated as raster maps and then the values were extracted by the random points, whereas the binary variable for subway service was generated afterwards based on Euclidean Distance. Here I have a few hypotheses.

Population density was anticipated to have a significant positive correlation with retail stores.

The two indicators of subway were included to detect the dependence of retail development on

subway network. The distance variable was expected to have a negative effect, while the availability variable was assumed to positively affect retail. Bus stop density was considered as a proxy for local transportation, and as explained previously, revealed the concentration of urban activity. This variable was expected to have a positive coefficient, and the extent of influence of local bus service was to be compared with the kind-of-new-coming subway service, so as to see after years of development, if the effect of subway has outweighed that of bus. Last but not least, bus route density was adopted as a proxy for main street density, indicating the street connectivity of a certain area. And this was anticipated to positively influence retail density.

Before conducting the regression, I ran the multicollinearity test using Variance Inflation Factors (VIF)¹. Then two models were created for the linear regression. One only takes into consideration the population density and the existence of subway service, while the other includes all other independent variables. This difference of the two models is to examine whether it is the binary situation of the existence of the subway service that affect retail density, or it is the actual distance from the subway that matters. If both matters, which one has the stronger influence. Then, for each retail type and model, a regression was carried out on original (unstandardized) variables to generate unstandardized coefficients, and another regression was conducted on standardized variables to produce standardized coefficient to compare the strength

¹ Variance Inflation Factor (VIF) is the ratio of variance in a model with multiple terms, divided by the variance of a model with one term alone. It quantifies the severity of multicollinearity in an ordinary least squares regression analysis. It provides an index that measures how much the variance (the square of the estimate's standard deviation) of an estimated regression coefficient is increased because of collinearity.

of correlation of different explanatory variables. The regression was conducted using the *Statsmodels* module¹ in Python, and was run for both all kinds of retails.

¹ Statsmodels (<https://www.statsmodels.org/stable/index.html>) is a Python module that provides classes and functions for estimating many different statistical models, as well as for conducting statistical tests, and statistical data exploration.

GEOGRAPHIC DISTRIBUTION OF RETAIL STORES

This section provides an exploratory analysis of the geographic distribution of retail stores by category in Shanghai. Then I looked into the retail stores within the neighboring area around subway stations and provided a series of Kernel Density estimation that will be used in the correlation analysis.

Global Pattern Analysis

The mean centers and distribution ellipses of the six types of retail stores are shown in Figure 7. The biggest red dot near Huangpu River represents the publicly-perceived city center of Shanghai, People Square. It is noticeable and interesting that all the six mean centers are located neither around the city center, People Square, nor around the geometric center of Shanghai, which is out of assumption. Instead, they sit to the southwest of the city center around Xujiahui. Xujiahui is a historic area of commerce and culture since 1988 administratively within Xuhui District, one of the seven inner city districts as depicted in Figure 1. It is now designated as a sub-center of Shanghai, well-known for shopping and entertainment. It is served by three subway lines: Line 1, Line 9 and Line 11. Here are two possible explanations of why the centers of retails are located around Xujiahui instead of People Square. The first reason is that Xujiahui is located more adjacent to the geometric center of Shanghai, which is to the south-west of People Square. Secondly, Puxi (“West Bank”) was more developed and dense than the other part of Shanghai, thus enjoys a more prosperous retail development. Pudong was originally referred to as the less-developed area in Shanghai and wasn’t urbanized and developed until 1990s as a Special Economic Zone dedicating to finance, trade and high-technology development. Among the six mean centers, the center of department stores is the closest to the city center. Compared to the

other four types, the centers of department store and specialty store are located nearer to People Square, showing that these two types concentrate more in the city center area.

Among the six directional distribution ellipses, the distribution of convenience stores is most disperse, whereas that of recreational facilities, restaurants and groceries. That is to say, convenience stores are less likely to benefit from being near each other around the city center than other stores and recreational facilities, restaurants and groceries tend to cluster together in order to attract customers by their collective offering of diverse commodities (Wang et al., 2014). All the six ellipses are mainly elongated along the Northeast-Southwest direction, which caters to the shape of the main city area defined in Shanghai Master Plan 2017-2035 illustrated in Figure 1 (b).

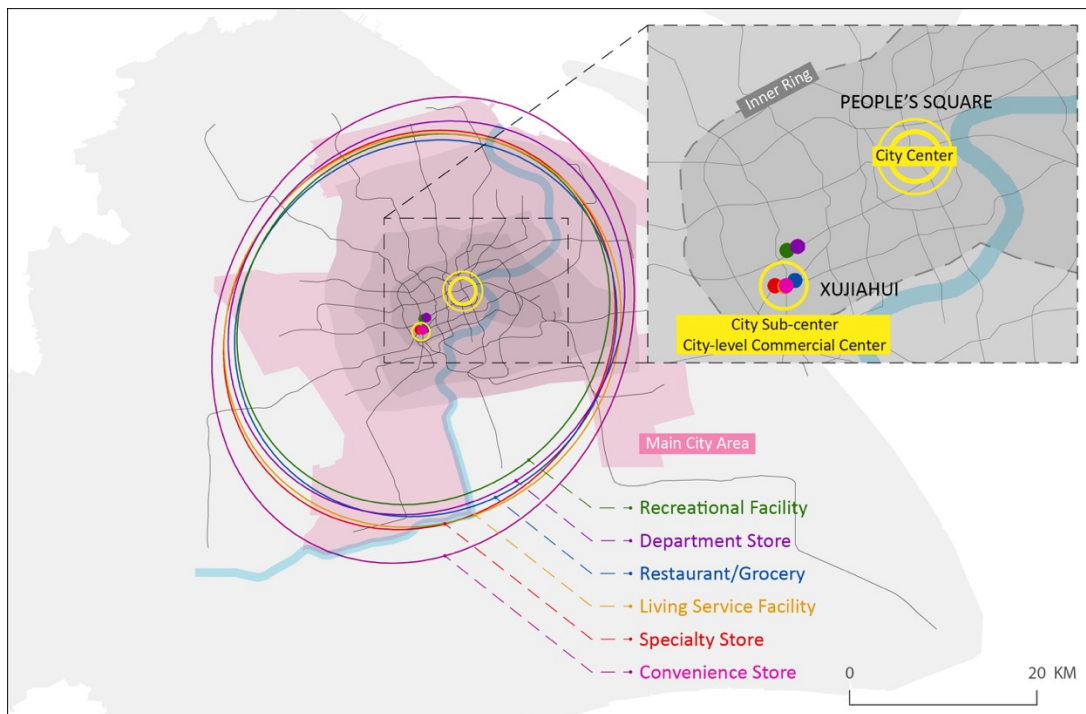


Figure 7 Mean Centers and Directional Distribution Ellipses of the Six types of Retail Stores

The “Average Nearest Neighbor” index is summarized in Table 4. The retail categories were reorganized according to the Z-scores from most statistically significant to least. Table 4 (a)

shows the result yield in the study area covering the whole area of Shanghai while Table 4 (b) looks only into the area within the outer ring of Shanghai.

Table 4 Results of Average Nearest Neighbor Analysis of Retail Stores

(a) Study Area: The Whole Area of Shanghai

Category	Nearest Neighbor Ratio	Expected Distance (m)	Observed Distance (m)	Z-score
Restaurant/Grocery	0.16	146.38	22.75	-604.30
Specialty Store	0.15	156.49	23.98	-564.25
Recreational Facility	0.28	318.68	89.72	-235.53
Convenience Store	0.31	395.91	124.07	-179.29
Personal Service Facilities	0.27	455.23	123.20	-161.45
Department Store	0.40	1638.72	660.22	-33.20
All Stores	0.15	97.23	14.16	-931.26

Note: Given the Z-scores, index of all categories are at significant at 0.01, meaning that there is a less than 1% likelihood that the clustered pattern could be the result of random chance.

(b) Study Area: The Area within the Outer Ring of Shanghai (Urban Area)

Category	Nearest Neighbor Ratio	Expected Distance (m)	Observed Distance (m)	Z-score
Restaurant/Grocery	0.43	52.58	22.75	-405.88
Specialty Store	0.42	56.46	23.98	-383.31
Recreational Facility	0.78	114.76	89.72	-71.51
Personal Service Facilities	0.72	169.95	124.2	-60.89
Convenience Store	0.86	144.1	124.07	-36.37
Department Store	0.98	676.49	660.22	-1.34
All Stores	0.4	34.51	14.16	-642.7

Note: Given the Z-scores, index of all categories, except Department Store, are at significant at 0.01.

In Table 4 (a), the ratio is all less than 1, indicating that the stores altogether (without differentiation of retail categories) as well as the six retail categories are highly clustered at city-scale. And the clustered pattern is statistically significant at 0.01 which is far from random. However, according to Table 4 (b), the result differs. The statistical significance of the clustering pattern of the top five types of retail stores decrease; however, department stores show no statistically significant spatial pattern (neither clustered nor dispersed). This is true and understandable if we imagine that the retail stores concentrate in the urban area of Shanghai, while when we zoom in, these stores are not actually sitting near each other and or it is possible that they form different clusters, but the clusters are not adjacent to each other in order to avoid

competition. Restaurant/grocery shows the most statistically significant spatial cluster in both tests, followed by specialty store.

Table 5 Distance Band Calculation from Neighbor Count of Retail Stores

Category	Minimum Distance (m)	Average Distance (m)	Maximum Distance (m)
Restaurant/Grocery	0.00	22.75	4831.52
Specialty Store	0.00	23.98	7937.19
Recreational Facility	35.32	89.72	6069.27
Personal Service Facilities	53.86	123.20	6880.78
Convenience Store	32.22	124.03	3443.36
Department Store	233.05	660.22	9686.22
All Stores	0.00	14.16	3522.28

Note: The minimum distance for all retail stores, restaurant/grocery and specialty store is zero. This is because some of the stores are located on different floors of a department store. In this case, they share same coordinates.

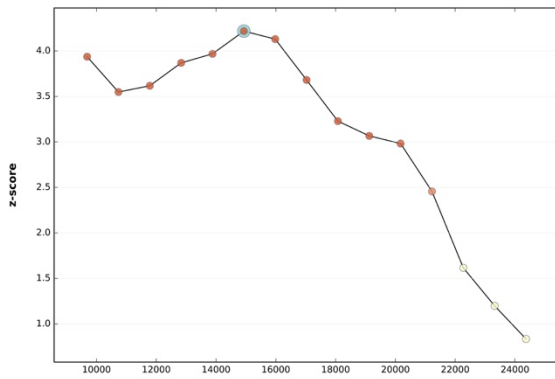
In terms of the distance from each other, according to Table 4, restaurant/grocery enjoys the shortest average distance (“*Observed Distance*”) from other stores of the same category, followed by specialty store. This reflects that these two types of stores may benefit from locating near each other and thus create a cluster of similar retail stores to cater to the customers’ multi-store selective shopping behavior. The specialty stores are always observed to cluster within department stores adding to the reason of the strong cluster pattern. Nevertheless, the maximum distance for specialty store is the second largest and the sharp contrast to the shortest minimum distance among all types, meaning that despite the overall clustering pattern, some of the specialty in fact locates far from others.

Recreational facilities also have a tendency of being near each other, but not as strong as the previous two. The average distances of convenience stores and personal service facilities, as well as the ratio, are about the same. These two are usually provided in accordance to residential communities, so the cluster pattern is not strong and might be mindful of avoiding competition with similar stores. However, convenience stores enjoy the shortest maximum distance, indicating that the distance ensuring each convenient store to have at least one neighbor is the

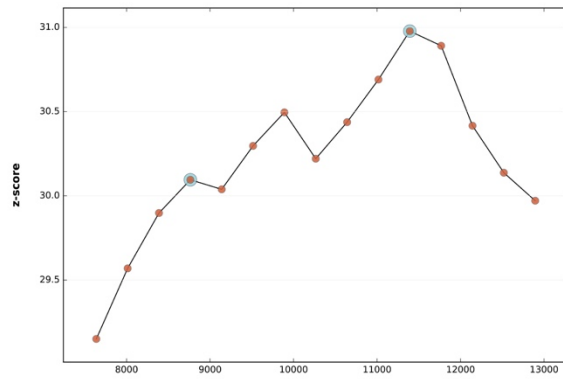
shortest. In sum, though do not sit near each other, enjoys an overall close distance from another facility of the same type and thus a more clustering pattern at a larger scale. The department stores locate the farthest from each other, as well as in terms of minimum and maximum distance presented in Table 5. There are two possible explanations. First, the department stores usually enjoys large footprints which spontaneously increase the distance in between. Second reason lies in its high service level. Department stores usually provide comprehensive commodities of different retail categories, so apart from those city-level commercial centers such as People Square or West Nanjing Road (similar to Times Square and Fifth Avenue in New York city), most commercial areas don't possess that large consumer capacity to allow a cluster of several department stores.

The result of the previous analyses has already revealed the spatial pattern would change over distance and scale, here I looked into the result of "*Incremental Spatial Autocorrelation*" for a further understanding of how the trend actually changes over distance. As shown in Figure 8, the change of statistical significance of the retail spatial clustering over distance (i.e., analysis area) can be apparently classified into two distinct groups: continuously increases as for convenience store, restaurant/grocery and recreational facility, and sharply decreases after peak as for department store, specialty store and personal service facility. Compared to the previous assumption, the analysis results for recreational facility and personal service facility turn out to be the of the different trend. For the first group, the significance of spatial clustering kept growing up as analysis area increases. No statistically significant peak was observed. This trend shows that these three types of retail stores aren't located quite near each other at a smaller scale but still form statistically significant clusters. And these clusters themselves sit closely at larger scales. This indicates that these retail types suffer competition of similar kinds, but still need to

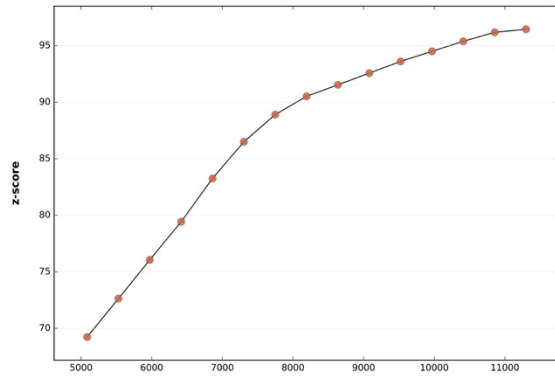
cluster in order to meet the large demands. This pattern accord with my observation in Shanghai. For example, convenience stores usually don't sit right near each other but still form clusters based on residential estates. On the other hand, for the second group, the intensity of spatial clustering drops dramatically after increasing to a peak significance level. These three retail types enjoy this clustering trend because they tend to cluster together in smaller scale to provide diverse commodities catering to customers' multi-store selective shopping behavior, while as scale increase, the small-scale clusters don't sit near each other to avoid competition.



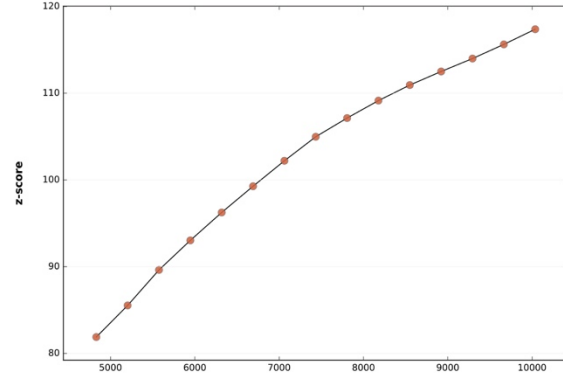
(a) Department Store



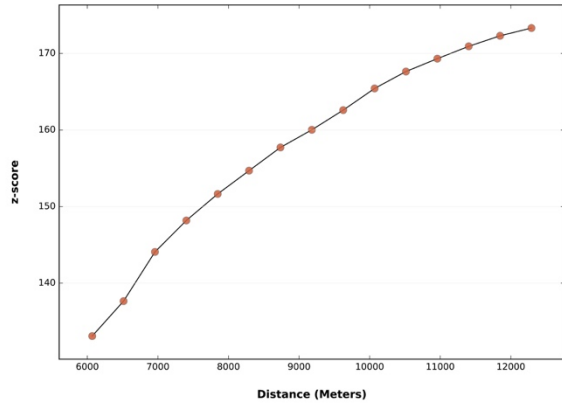
(b) Specialty Store



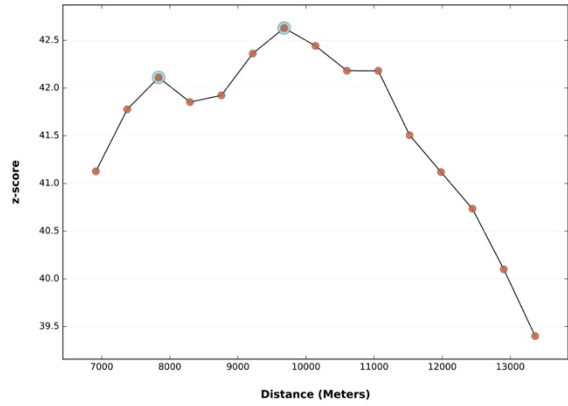
(c) Convenience Store



(d) Restaurant/Grocery



(e) Recreational Facility



(f) Personal Service Facility

Figure 8 Spatial Autocorrelation by Distance Yielded by Global Moran's I

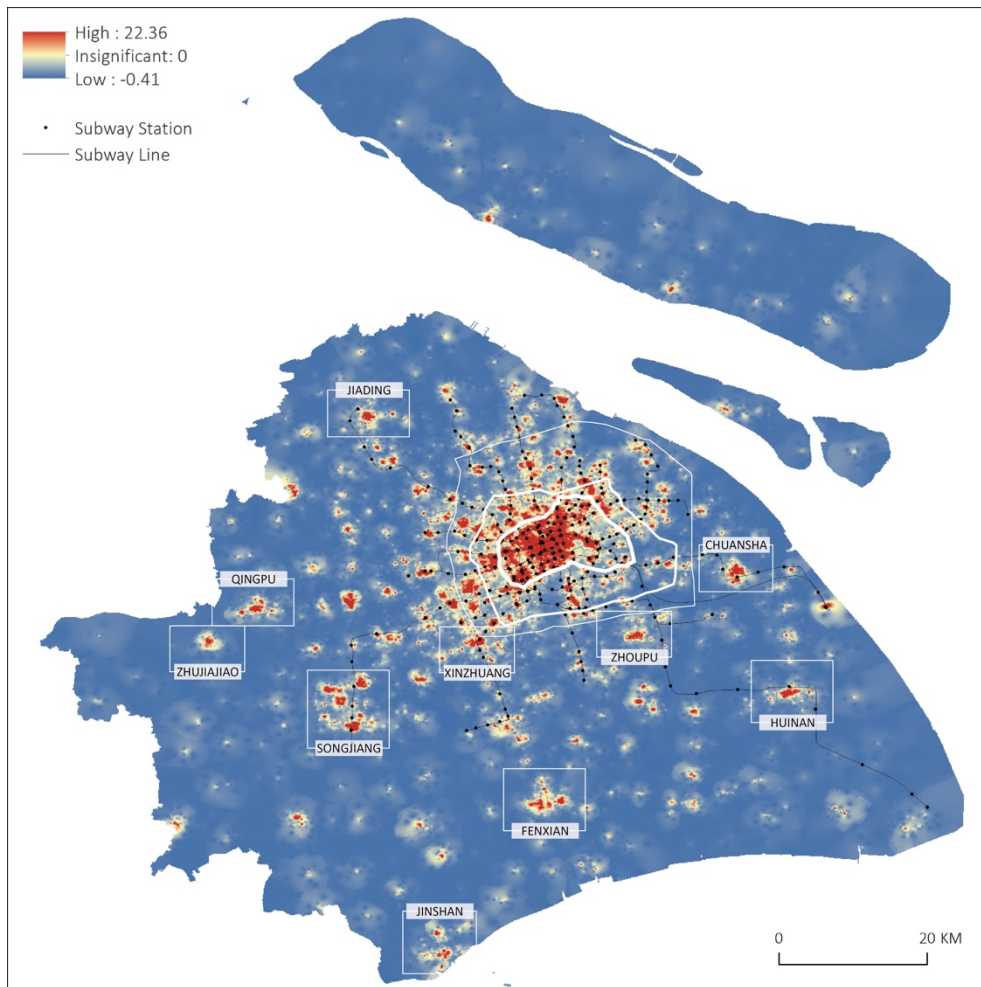


Figure 9 Hot Spot Map of All Types of Retail Stores in Shanghai

Figure 9 mapped the clusters of all types of retails in Shanghai: red for hot spots, blue for cold spots and yellow for no statistically significant pattern areas. The polycentric retail cluster

pattern coincides the multi-level city centers and urban structure illustrated in Figure 1 (b) in the BACKGROUND section. The spatial overlap of retail hotspots and subway stations is quite apparent. The six subway lines extending to the west part of Shanghai (Puxi, “West Bank”), from the city center coincide the linear configuration of hot spots around the subway stations, while in the east part of the city, the overlap occurs occasionally. The existence of hot spots beyond the service area of subway network shows that there are other factors influencing concentration of retail stores other than subway stations; for example, populated new sub-centers or new towns of suburban area as labeled such as Jiading, Qingpu, Zhujiajiao, Songjiang, Huinan, Fenxian, and so on. Since the retail development occurred long before the construction of subway network, the question if it is subway network that leads to retail development or vice versa is indeed a “chicken or egg dilemma”. More importantly, this causality is a theoretical preposition. Without time-series dataset, it is hard to identify the casual relationship, which is one of the limitations of this research. However, the spatial relationship between retail clusters and subway stations can still be observed and quantified through spatial statistical analysis.

Local Pattern Analysis

Looking into local scale, Figure 10 illustrates not only cluster of high/low value clusters but also high/low value outliers based on *Anselin Local Moran's I*. Besides the cluster of high values in the city center, several clusters are observed in the suburban area along the subway lines.

Compared to the result yield by *Hot Spot Analysis*, a few different conclusions can be made.

First, within the belt surrounding the inner city area, the area of insignificant spatial pattern in the Hot Spot Map is identified as low value outliers in the Cluster and Outlier Map as shown in detail in Figure 11 (a), meaning that they are the low value areas surrounded by high values., whereas the cold spots are classified as areas with insignificant values. Second, not all hot spots

in the peripheral area are identified as high values clusters. Four examples were shown in Figure 10. Chuansha, Xinzhuang and Songjiang were designated as sub-centers in Shanghai Master Plan 2017-2035 as depicted in Figure 1 (a), while Fenxian is a peripheral sub-center without subway service. Figure 11 (b) illustrates that the other smaller hot spots and insignificant areas are plotted as high value outliers (high value surrounded primarily by low values). Compared to the hot spots, those outliers don't that strongly coincide with the subway stations, indicating that in the area that the clustering pattern of retail stores isn't statistically significant, the occurrence of retail concentration in a small scale (such as a commercial center of a housing estate) is quite frequent with no necessary correlation with subway stations.

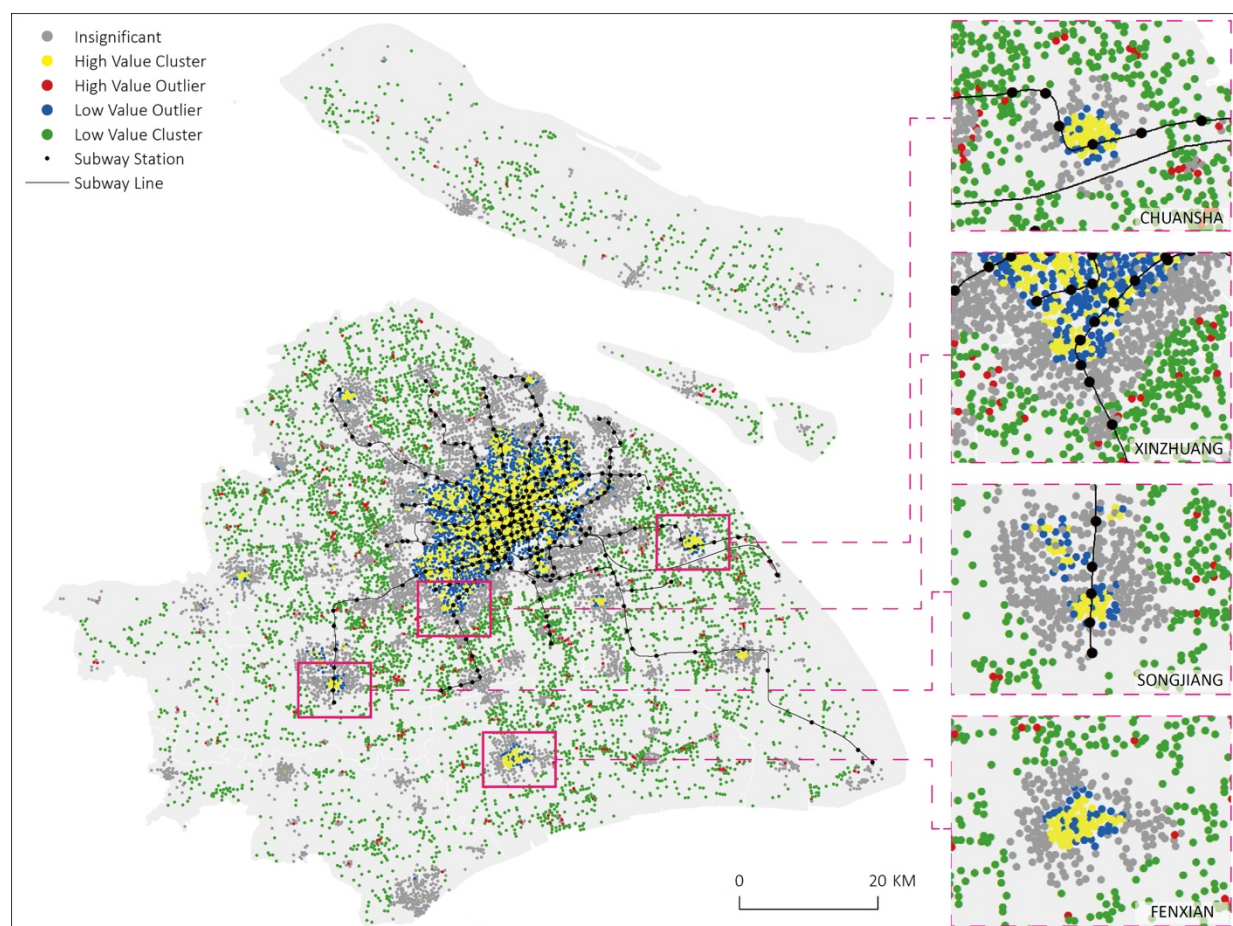


Figure 10 Cluster and Outlier Map of All Types of Retail Stores in Shanghai

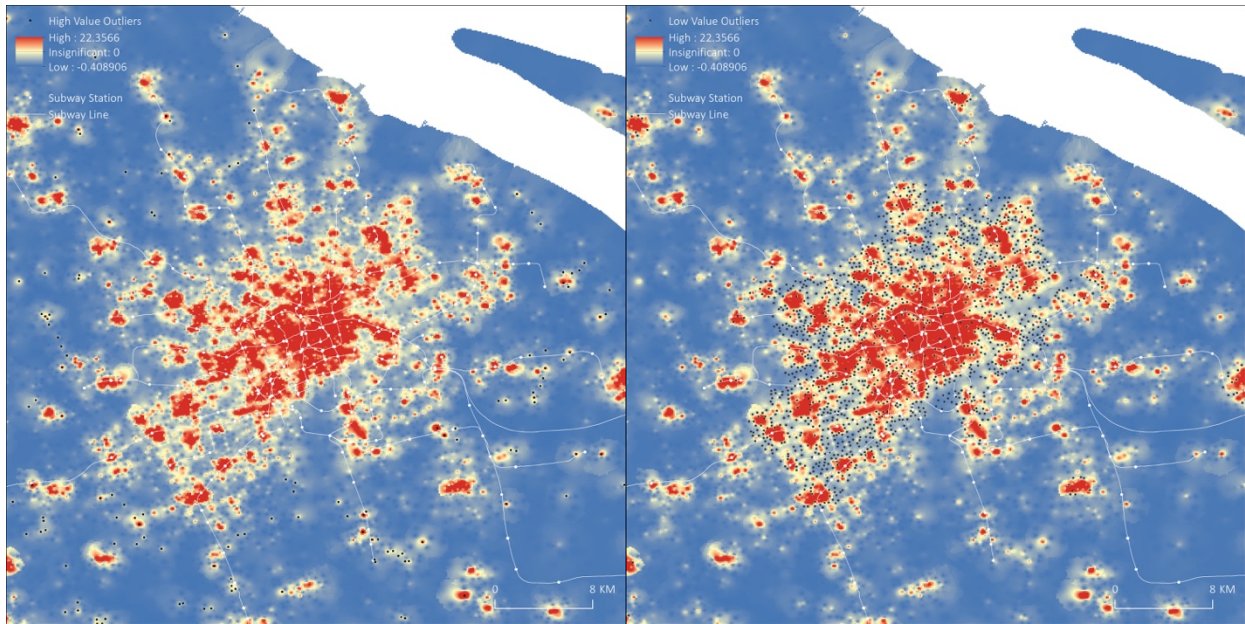


Figure 11 Overlapping Hot Spot Map with (a) High Value Outliers (b) Low Value Outliers

The Kernel Density estimation reveals the density of retails within each geographic unit (100m * 100m) without taking into consideration the value of surrounding cells. Thus the high-density area is obviously larger than the hot spot areas. It is adopted as a visualization of the density of retail stores throughout Shanghai and later as the dependent variable of the linear regression.

Figure 12 mapped the Kernel Density values of all types of retail stores classified by Natural Break. The high-density area concentrates in the city center area and distribute in plaques as expanding to the suburban areas. In addition, subway stations apparently coincide with but fail to cover all the high-density area of retail stores. For example, as discussed previously, sub-center areas such as Fengxian and Jinshan in the southern part of Shanghai are not served by subway.

The linear expanding pattern along some subway lines is explicit, especially in Puxi. The high-density area explicitly grows from city center to sub-centers Jiading, Wusong, Songjiang and Fengxian. This result is consistent with the result yielded from the Hot Spot Map, implying other influential factors for retail density.

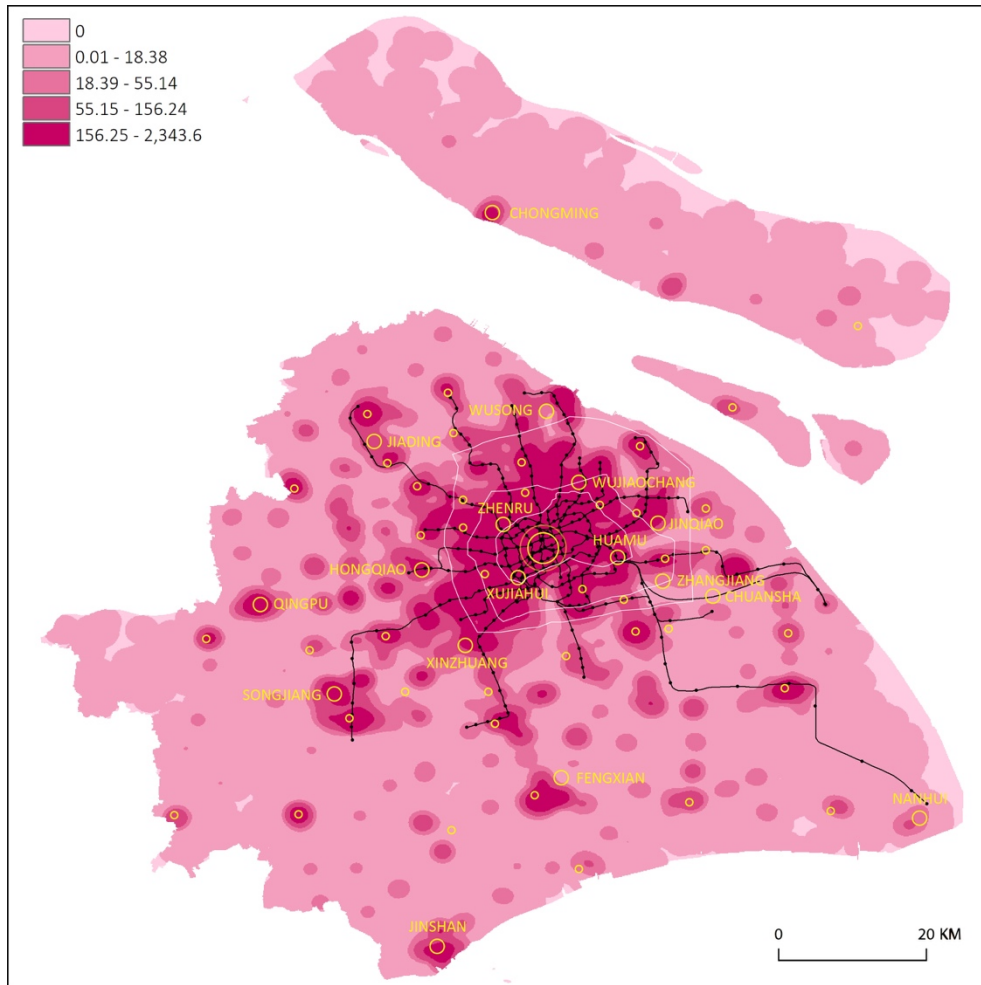


Figure 12 Kernel Density Estimation of All Types of Retails
(overlaid with city centers)

Figure 13 maps the Kernel Density values of each type of retail stores classified by Natural Break. Overall, the density of each type shows a similar trend of decreasing from the city center to outskirt area, as well as the coincidence with city center, sub-centers and subway stations. In addition, the linear expansion of high-density areas along retail stores is obvious in the inner city area, and in suburban areas the high-density areas tend to show discrete plaques; for example, in Songjiang and Fengxian to the south and Wusong to the north. As far as I'm concerned, the similarity of density distribution patterns of different types of retail lies in the overall trend of comprehensive development of retail spaces in Shanghai. New retail stores area usually designed as a complex providing diverse goods and services and can thus become a center of a certain

surrounding area. It is worth noting that Nanhui, a sub-center to the south-east corner of Shanghai, fail to enjoy a high density of specialty store, convenience store, recreational facility and personal service facility. The new town in Nanhui is called Dishuihu (Chinese: 滴水湖), constructed since 2000. It turns out that after years of development, the new town still fail to enjoy a prosperous retail development. Compared among the density of different types of retail stores, some areas with low density of department stores enjoy a medium (or sometimes high) density of other type of retail stores. This is because in suburban areas, the demand for department and the volume of potential consumers are not sufficient for developing department stores, but other retail types like convenience stores or personal service facilities that meet basic living demands are still required. This also provides explanation for the fact that these two types of retail stores enjoy the most expansion into suburban areas and diffuses more continuously.

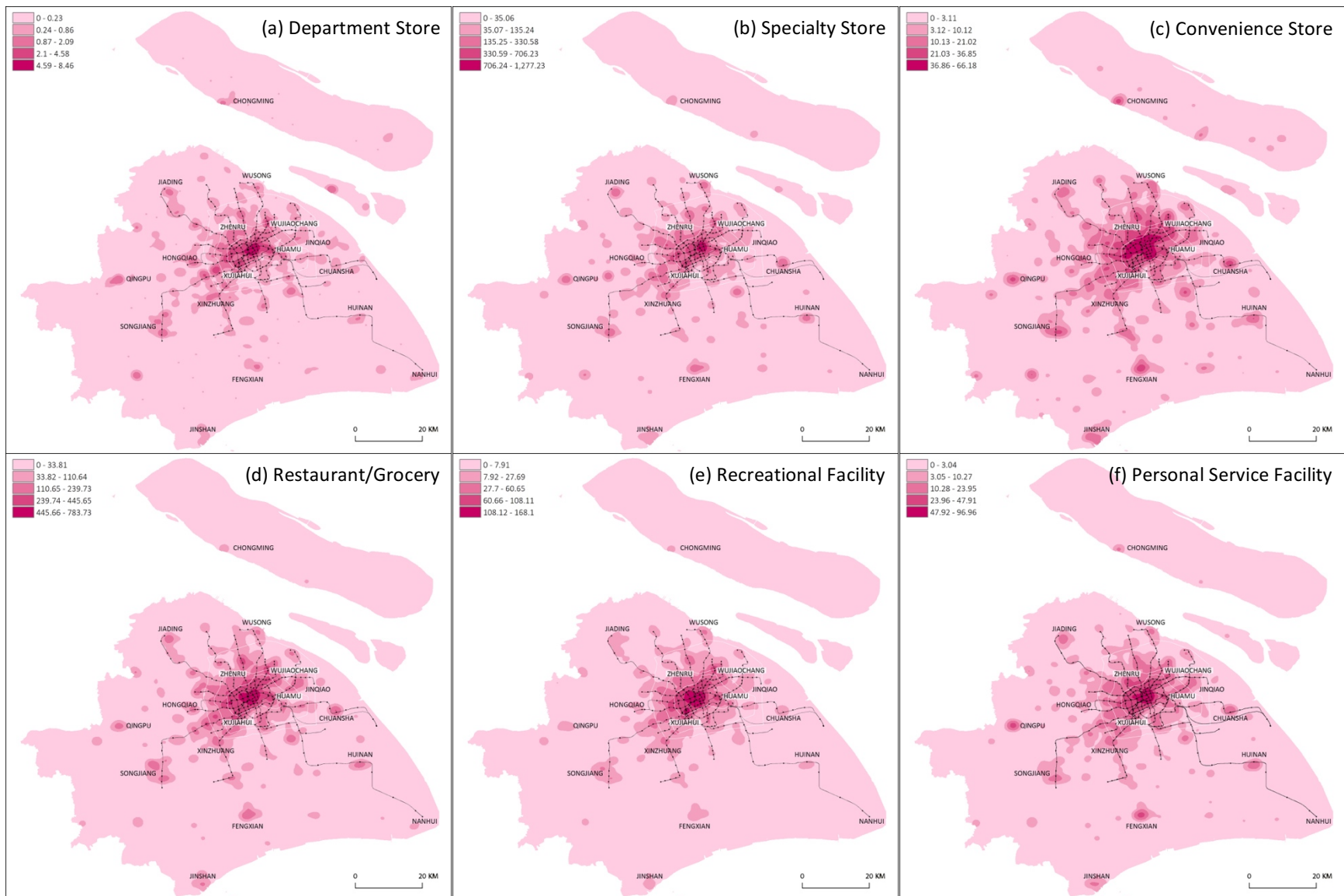


Figure 13 Kernel Density Estimation the Six Types of Retail Stores

Looking more specifically into the area around subway stations, I counted the number of retail stores by category. The result is listed in Table 7. Compared to the area summarized in Table 6, for the 500-meter buffer of subway stations, approximately 1/3 of the retail stores (32.20 percent) clustered in the 3.32 percent of the total area of Shanghai. When the buffer area was enlarged to 1,000 meters away from subway stations, the proportion of retail stores only slightly increase by 0.9 percent to 33.10 percent, indicating that the clustering effect of stores around subway stations tend to cover with a radius of 500 meters. Beyond this radius, the retail stores tend to distribute in a more disperse pattern (e.g. along streets). It is also worth noting that 98.40 percent of retail stores cluster on the 54.67 percent of the total area of Shanghai within 500 meters from bus stops. This proves my assumption that the bus stops can, to some extent, act as a proxy for concentration of urban activity and local transportation service in urban areas.

Table 6 Area within Subway Station and Bus Stop Buffers

Buffer	Area (sq. km)	% of Total Area
Subway	500M	3.32%
	1,000M	8.98%

Table 7 Number of Retail Stores within Subway Station Buffers

Category	Within 1,000-m Buffer (a)	% of Total Number	Within 500-m Buffer (b)	% of Total Number	% of the Number within 1,000-m Buffer (b/a)
Department Store	350	41.42%	345	40.83%	98.57%
Specialty Store	42,491	35.02%	41,630	34.31%	97.97%
Convenience Store	4,416	23.71%	4,238	22.76%	95.97%
Restaurant/Grocery	46,065	32.93%	44,728	31.97%	97.10%
Recreational Facility	9,702	33.04%	9,323	31.75%	96.09%
Personal Service Facility	4,010	29.95%	3,844	28.71%	95.86%
All Stores	107,470	33.10%	104,537	32.20%	97.27%

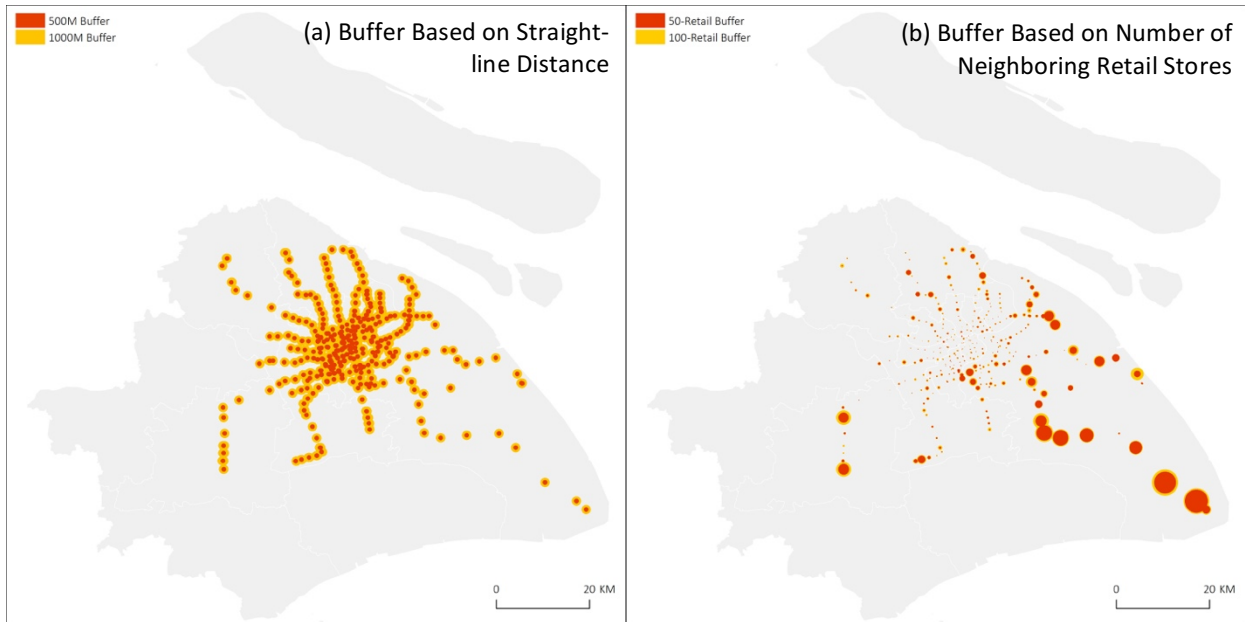


Figure 14 Subway Station Service Buffer

Figure 11 (b) shows the subway station buffer defined by calculating the average distance from the subway stations to the nearest 50 and 100 retail stores, which is apparently different from the straight-line distance buffer illustrated in Figure 14 (a). They are smaller in city center compared to the distance buffer, whereas they are much bigger in suburban areas. The size of the buffer area increases from city center to outer-skirts of urban area. The smaller the circle, the denser the retail stores are around that station. In addition, the radius doesn't increase much to cover 50 and 100 retail stores, indicating that the nearest 100 retail stores distribute relatively compact around subway stations.

SPATIAL RELATIONSHIP OF RETAIL SPATIAL DENSITY AND SUBWAY NETWORKS

This section takes a step further to examine and qualify how the distribution of retail stores correlate with subway network and other influential factors. The explanatory variables are visualized in Figure 15.

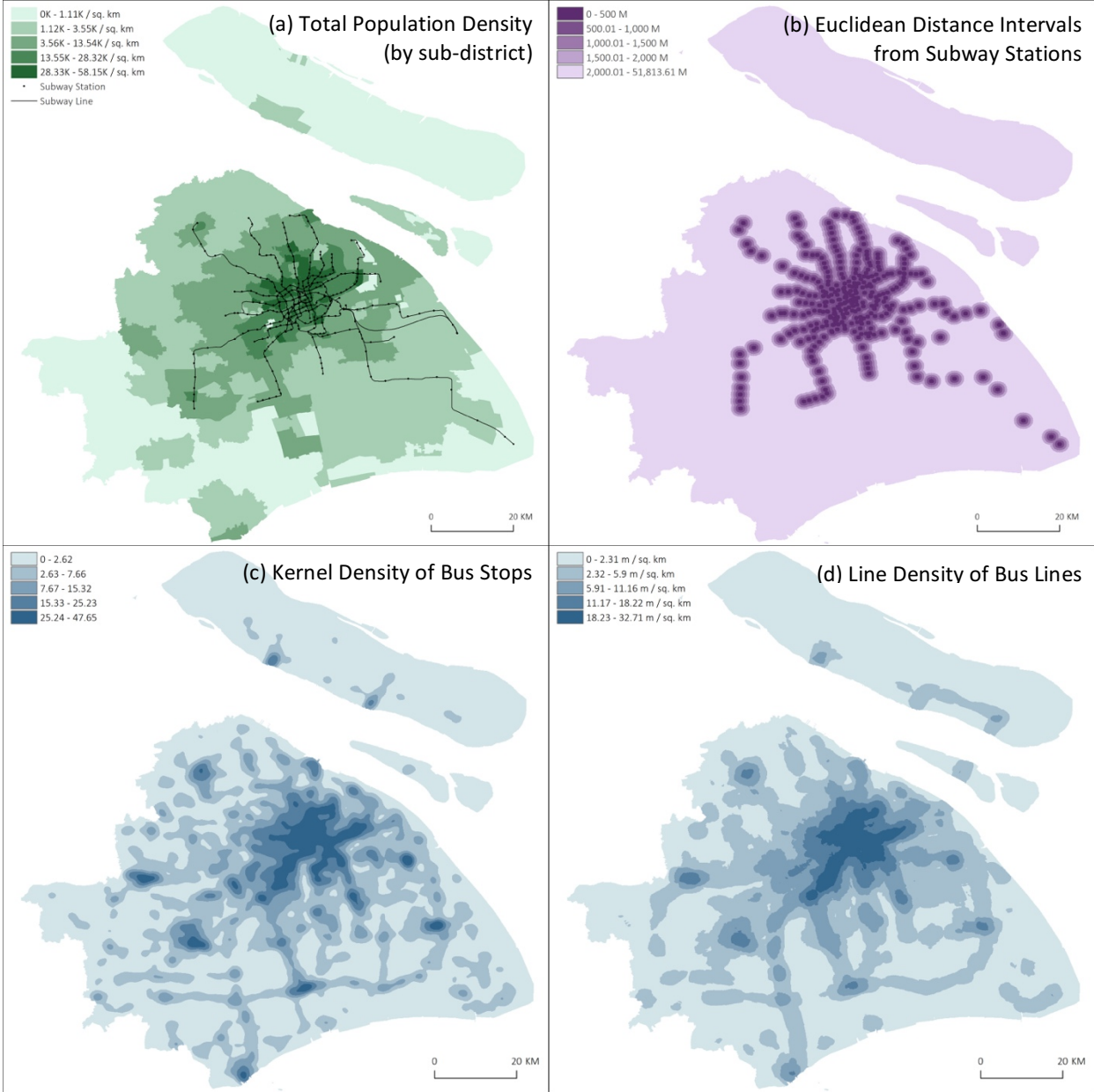


Figure 15 Visualization of Explanatory Variables

The explanatory variables show a similar spatial pattern of high-density in city center with expansion of denser segments either linearly or in discrete plaques in suburban areas. Meanwhile, the density distribution of these variables show high coincidence. It is understandable that due to the agglomeration effect on socio-economic elements (like population) and the corresponding physical elements (like retails and transportation network) in urban area – the interrelation tend to be unavoidable. As a result, before the OLS linear regression, I ran multicollinearity test using Variance Inflation Factors (VIF). According to the result of VIF test, line density of bus lines shows a high multicollinearity and was thus removed from the regression model, leaving only three explanatory variables in Model 2. The result of the two models of OLS linear regression was summarized the result in Table 8 for each of the six retail types as well as all retails as a whole. Since 2000 random points were generated, the number of observation for the regression was 2000.

Overall, according to Table 8, Model 2 enjoys a stronger explanatory power (Adj. R-squared) than Model 1. Amongst the six retail types, the density of convenience store is best explained by both models – Model 1 explains over 65% of the variance of store density and Models 2 explains over 76%. The coefficients in both models show similar trend for each retail type and all retail stores and coincide with the previous assumption. The result reveals that population density, availability of subway service, as well as the density of local bus service and urban activity positively influence the retail density, while the Euclidean distance from subway station has a negative impact. In addition, it's noticeable that the effect of distance from subway station on density of specialty store and personal service facility isn't statistically significant, while the coefficient of the binary criterion indicating the coverage of subway station is significant in both models. This indicates that for these two types of retail stores, what matters is if it is available via

subway network rather than the exact difference of distance from subway stations. If available, the density increases, and vice versa.

The unstandardized coefficient implies the absolute power of influence by showing one unit change in the explanatory variable will the amount of change in retail density equaling to the value of the coefficient. For example, the result for convenience store can be interpreted in following equation:

$$\begin{aligned} \text{KDen (CS)} &= \text{intercept} \\ &+ 3.4434 * \text{Den (Population)} \\ &- 5.3101 * \text{EDist (Subway Station)} \\ &+ 4.0184 * \text{KDen (Bus Stop)} \end{aligned}$$

This interpretation allows me to understand which type of retail is more influenced by a specific independent variable. As marked red in Table 8, in both models, the coefficient of all four explanatory variables is higher than that of all other five retail types, meaning that one unit increase in those variables will yield the sharpest increase in density of restaurant and grocery. The department store is the least sensitive to the four independent variables as marked blue. However, this might be explained by the least quantity of department stores among the six types. The overall density of department store is comparatively low and thus the variation of density is also low.

On the other hand, the standardized coefficient allows comparison among various explanatory variables for each dependent variable. As marked as bold font, in model 1, population density enjoys a stronger correlation with retail density for all six types as well as for all retail stores than the availability of subway service, while in model 2, the impact of density of bus stop overweigh population density for all types as well as all for retail stores except for recreational facility. In

addition, the distance from subway stations tend to have the weakest spatial correlation with retail density. This answers the question raised at the beginning of this research – after decades of development, the effect of subway service on retail density fails to outweigh that of local bus service. This weak effect might be resulted from the fact that the subway lines haven't fully extended throughout Shanghai, while there already exist retail clusters and town centers providing retails for basic living demands (like convenience goods and personal service) in those areas without subway service.

However, comparing the six retail types, the difference between impact strength of distance from subway stations and other two dependent variables is the smallest for department store. This means a relatively strong spatial relationship between department stores and subway stations. In addition, based on my own observation, the department stores always act as a pioneer facility designated by the city government along with the extension of subway service when a new town or a new commercial center is to be established.

Table 8 OLS Regression Result

Dependent Variable	MODEL1						MODEL2					
	R ²	Adj. R ²	Independent Variable	Coefficient		Std. Coef.	R ²	Adj. R ²	Independent Variable	Coefficient		Std. Coef.
KDen (DS)	0.302	0.301	Den (Population)	0.0274 (0.002)	***	0.3699 (0.024)	0.370	0.369	Den (Population)	0.0120 (0.002)	***	0.1625 (0.027)
			Subway Service	0.3737 (0.039)	***	0.2338 (0.024)			EDist (Subway Station)	-0.0662 (0.011)	***	-0.1521 (0.026)
									KDen (Bus Stop)	0.0222 (0.002)	***	0.3664 (0.026)
KDen (SS)	0.383	0.383	Den (Population)	4.8670 (0.221)	***	0.5010 (0.023)	0.488	0.488	Den (Population)	2.6793 (0.214)	***	0.2758 (0.022)
			Subway Service	34.5460 (4.780)	***	0.1647 (0.023)			KDen (Bus Stop)	3.8198 (0.175)	***	0.4799 (0.022)
KDen (CS)	0.652	0.652	Den (Population)	0.7684 (0.019)	***	0.6954 (0.017)	0.761	0.761	Den (Population)	0.4926 (0.019)	***	0.4458 (0.017)
			Subway Service	3.8476 (0.408)	***	0.1613 (0.017)			EDist (Subway Station)	-0.2706 (0.103)	***	-0.0417 (0.016)
									KDen (Bus Stop)	0.4261 (0.014)	***	0.4706 (0.016)
KDen (RG)	0.599	0.598	Den (Population)	5.9963 (0.181)	***	0.6076 (0.018)	0.717	0.717	Den (Population)	3.4434 (0.181)	***	0.3489 (0.018)
			Subway Service	48.7951 (3.917)	***	0.2290 (0.018)			EDist (Subway Station)	-5.3101 (0.996)	***	-0.0915 (0.017)
									KDen (Bus Stop)	4.0184 (0.139)	***	0.4969 (0.017)
KDen (RF)	0.578	0.578	Den (Population)	1.3683 (0.042)	***	0.6169 (0.019)	0.638	0.637	Den (Population)	0.9579 (0.046)	***	0.4318 (0.021)
			Subway Service	9.6168 (0.902)	***	0.2008 (0.019)			EDist (Subway Station)	-1.3204 (0.253)	***	-0.1013 (0.019)
									KDen (Bus Stop)	0.6435 (0.035)	***	0.3540 (0.020)
KDen (PSF)	0.527	0.527	Den (Population)	0.5996 (0.020)	***	0.6017 (0.020)	0.660	0.660	Den (Population)	0.3438 (0.018)	***	0.3450 (0.018)
			Subway Service	3.7775 (0.429)	***	0.1755 (0.020)			KDen (Bus Stop)	0.4375 (0.015)	***	0.5357 (0.018)

Table 8 (continued)

Dependent Variable	MODEL1					MODEL2					
	R ²	Adj. R ²	Independent Variable	Coefficient	Std. Coef.	R ²	Adj. R ²	Independent Variable	Coefficient	Std. Coef.	
KDen (ALL)	0.543	0.543	Den (Population)	13.6813 (0.454)	*** 0.5911 (0.020)	0.659	0.658	Den (Population)	7.8054 (0.465)	***	0.3372 (0.020)
			Subway Service	101.2976 (9.805)	*** 0.2027 (0.020)			EDist (Subway Station)	-8.8878 (2.567)	***	-0.0653 (0.019)
								KDen (Bus Stop)	9.4329 (0.359)	***	0.4916 (0.019)

Note: *** p<0.01, ** p<0.05, * p<0.1

CONCLUSION AND DISCUSSION

Conclusion

In conclusion, the retail stores are distributed in an overall clustered pattern in Shanghai with concentration of high-density area in city center and plaques of clusters expanding into the peripheral of urban area and suburbs. The 1,000-meter buffer of subway stations covers only 9% of the total area of Shanghai, but it includes about one-third of retail stores. Moreover, the density of retail stores around subway stations also decrease from city center to suburban areas. The mean center of the retail stores extends to the area near Xujiahui, a historic commercial center and sub-center of Shanghai, rather than the designated city center, People Square.

According to the previous literature, there were originally only three commercial streets and one commercial center area in the city center before the construction of first subway line, Line 1 (Li, 2006; Hao, 2007; Wang et al., 2009; Song et al., 2010), while nowadays various hot spot of retail stores emerges throughout Shanghai, with apparent spatial coincidence to the location of subway stations. In downtown Shanghai, the overlaying pattern is not that obvious because the retail density is generally high. In the west part of Shanghai (i.e., Puxi), which is more developed, the linear expansion pattern along subway lines are most obvious. While in the less-developed area of Pudong and south part of Shanghai, most area is out of the service area of subways, the retail hot spots occur in sub-district centers or town centers. This also matches the polycentric urban structure of Shanghai. Admittedly, this coincidence is quite a “chicken-and-egg” causality dilemma. Subway stations tend to be located in the existing commercial centers to serve more people; on the other hand, new retail stores tend to grow and cluster around subway stations to obtain a better accessibility for the public. Because of the lack of time-series data of retail stores, I cannot identify the causality relationship in-between.

Specifically, different retail types share similar common spatial distribution patterns to all retail stores, but meanwhile reveal difference. The spatial clustering of department stores is the least statistically significant, while that of restaurants and groceries is the most, followed by specialty stores. More importantly, the spatial clustering of department store, specialty store and personal service facilities dramatically decrease after reaching to its peak at a specific distance range, meaning that though clustering in smaller scale, the clusters tend to locate apart from each other. On the contrary, convenience store, restaurant/grocery and recreational facility enjoy a continuously increasing spatial cluster as distance goes up. In addition, the high-density area of department stores and recreational facilities tend to concentrate in the city center area, whereas clusters of convenience stores and personal service facilities expand more into the outskirt area because they are the retails that meet the basic living demands and are usually assigned as supporting facilities of residential estates. The average distance of specialty stores, restaurants and groceries is the shortest, meaning that these two types of retail stores tend to locate near each other to cater to the customers' multi-store selective shopping behavior. That between convenient stores and living service facilities is also quite large than expected, meaning that these two types of retails might be mindful of avoiding competition with similar stores. However, the convenience stores enjoy the shortest maximum distance, indicating a small distance range.

According to the spatial relationship analysis, retail density turns out to be positively related to population density, availability of subway service, and density of local bus service. Distance from the nearest subway station turns out to negatively influence retail density and the correlation is the weakest among all explanatory variables. Moreover, the correlation isn't statistically significant for specialty store and personal service facilities, indicating the

development of these two particular types of facilities isn't dependent on subway service.

Density of local bus service is the most influential factor of retail density, indicating that after decades of subway development, the effect of subway on retail development hasn't overweigh bus service yet. However, this might be explained by the fact that some existing sub-centers and new towns are still without the reachable area via subway. It is noticeable that the impact of density of bus stop overweigh population density for all types as well as all for retail stores except for recreational facility. Last but not least, department store enjoys a relatively stronger correlation with distance from subway stations compared to other retail types.

Interaction between Retail Development and Subway Expansion

My analysis shows that there is high spatial coincidence of retail clusters and subway stations, as well as the quantitative positive relationship between retail density and proximity to subway distance. As discussed previously, the causal analysis of these two development is indeed a "chicken or egg dilemma". Lots of retail clusters were formed decades before the development of subway network, and the expanding process of the two occurs alternatively and spontaneously with interaction nowadays.

In terms of the influence of existing retail clusters on subway network expansion, the location of new subway stations and the extending direction of existing subway lines should take into consideration the dense retail clusters. This is in fact the location strategy of subway stations in dense city center area. For example, the designated city center, People Square, is served by three subway lines, Line 1, Line 2 and Line 8, and the mean center of all retail stores, Xujiahui, is served by served by three subway lines Line 1, Line 9 and Line 11. What's more, in Shanghai, the expansion of subway network is adopted more as an approach of developing new centers.

With service of subway, new retail clusters and residential estates can thus be placed farther into

the peripheral area so as to lower the density of the city center. This is also a strategy of revealing the polycentric urban structure in Shanghai. It seems reasonable to anticipate that the retail clusters will thus expand in plaques around subway stations along the subway routes. Figure 16 provides two examples. Line 5 opened in 2003 as an extension towards south-west of Line 1, connecting Xinzhuang (terminal station of Line 1), a sub-center as mentioned before, with the Minhang Economic & Techonological Development Zone in Minhang District. Minhang Development Zone is a satellite new town designated in 1986, located approximately 15 kilometers outside the outer ring of Shanghai. The opening of Line 5 dramatically improved access to downtown Shanghai. Line 9 also extends farther towards the south-west part of Shanghai, connecting Songjiang New Town with the city center.

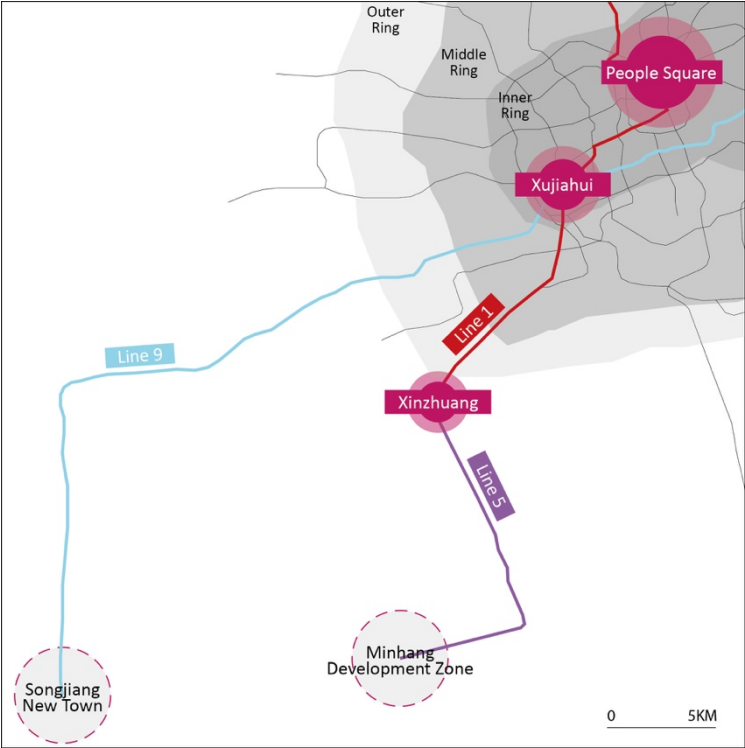


Figure 16 New Town Development Led by Subway

However, despite the attractiveness of retails around stations, subway service also largely expands people’s commute range. As a result, there also lies the possibility that with subway

stations improving connectivity in urban areas, the retail stores tend to concentrate more to the center within a proper distance – it's still time-consuming to travel from suburban areas to city centers even by subway. This indicates the difference between distance-based and time-based measure when evaluating accessibility and attractiveness. The dominant way planners and developers currently adopted to locate new retail complexes is based on the service area of subway stations based on linear distance (“as the crow flies”) or better on street-oriented distance instead of the actual time required to arrive at the destination, but the willingness to travel to one place is determined by the potential cost of travel, which is better measured and simulated by time-based distance, according to a migration research in New Zealand (Poot, Jacques, et al., 2016). Porta & Romice (2014) also argue that urban disciplines should seriously consider the factor of time in urbanism and a deeper “time-conscious” approach should be adopted. In the context of this research, it might actually take more time for you to walk to a store that is located at the outskirts of the serving area of a subway station than go to a store a few stations away.

Based on this bidirectional consideration, I anticipated that there are two hierarchy of effects of subway network on retail developments. Firstly, the expansion of subway network indicates a large number of potential consumers, so it will cause retail clusters around the subway stations compared to other areas without subway service. On the other hand, the subway network will strengthen the hierarchy of retail clusters and further improve the service level of those retail clusters in the city center. For example, the existing commercial centers located in downtown Shanghai enjoy better accessibility via subway network, and thus the further retail development is promoted by the large commuter flows brought by subway, which forms a virtuous cycle of retail development. However, the agglomeration effect of city center is not likely to fully outweigh the attractiveness of subway stations in outskirts areas given the hot spot and kernel

density in the previous section, especially for those retail clusters meeting basic living demands, such as convenience stores and personal service facilities.

Limitations

While comprehensive and in-depth, this research is limited in terms of datasets, data processing and analysis. First of all, the POI dataset of retail stores is an incident dataset without information of consumer volume or employment scale, so it can only yield count of retail stores instead of revealing the true service level of each specific store. For example, when calculating retail density, a restaurant with only a few tables should not be given the same weight as a restaurant holding banquets. DIANPING.COM does provide the number of customer reviews. However, I considered this information not reliable enough to be adopted as a proxy for the scale of retail stores and thus didn't weigh the retail stores with this attribute for the following two reasons: first, the review counts vary markedly for different retail types (for example, lots of people review restaurants but hardly any review convenience stores); second, the review counts largely depend on how long the store has opened (imagine a newly-opened store and a store with decades of history).

Second, the geographic unit of the population data is at sub-district level, which is unable to reveal the population distribution in detail within each sub-district. Compared to the raster map calculated with cell size at 100m * 100m, the population data is not as accurate. This inaccuracy harms the explanatory power of the population density variable that generated from the population data. For example, the high-density area of retail stores might be actually located in a small populated area in a sub-district with overall low density. The retail density is actually highly correlated with population density in microscale, but this cannot be revealed with population data at sub-district level. In addition, the 2016 population data is only available at

district level, so it is projected from 2010 population data assuming that all sub-districts within one district enjoys a same percentage change from 2010 to 2016. This projection further harms the accuracy and validation of the population density variable.

Last but not least, the limitation also exists in the measure of distance from subway station. Straight-line distance (Euclidean Distance) is adopted instead of network distance based on urban street network because the data of street network is not available to the public. As a result, I decided to use distance intervals rather than observe the correlation of retail density with slight change of exact distance in this research. However, if the network distance is available, the regression model is able to explain more about the spatial relationship of retail and subway.

BIBLIOGRAPHY

- Anas, Alex, Arnott, Richard & Small, Kenneth A. (1998) Urban Spatial Structure. *Journal of Economic Literature* 36: 1426-1464.
- ArcGIS Desktop Tool Reference. Cluster and Outlier Analysis (Anselin Local Moran's I). <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm> [Accessed on Feb 12th, 2018].
- ArcGIS Desktop Tool Reference. Incremental Spatial Autocorrelation. <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/incremental-spatial-autocorrelation.htm> [Accessed on Feb 12th, 2018].
- ArcGIS Desktop Tool Reference. Hot Spot Analysis (Getis-Ord Gi*) <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/hot-spot-analysis.htm> [Accessed on Feb 12th, 2018].
- Briney, Amanda (2017). Central Place Theory. <https://www.thoughtco.com/central-place-theory-1435773> [Accessed on Nov 28th, 2017].
- Buliung, Ron N. (2011) Wired People in Wired Places: Stories about Machines and the Geography of Activity, *Annals of the Association of American Geographers*, 101:6, 1365-1381, DOI: 10.1080/00045608.2011.583568.
- BusinessDictionary. Trade Area. <http://www.businessdictionary.com/definition/trade-area.html> [Accessed on Feb 25th, 2018].
- Calimente, John (2012) Rail Integrated Communities in Tokyo. *The Journal of Transport and Land Use* 5(1): 19:32.
- CityMetric (2015). Which City Has the Longest Metro System? <http://www.citymetric.com/transport/which-city-has-longest-metro-system-1359> [Accessed on Dec 5th, 2017].
- Crucitti, Paolo, Latora, Vito, Porta, Sergio (2006). Centrality Measures in Spatial Networks of Urban Streets. *Physical Review*, 73, 036125:1-5.
- Ertekin, Ozhan, Dokmeci, Vedia, Unlukara, Tuba & Ozus Evren (2008). Spatial Distribution of Shopping Malls and Analysis of their Trade Areas in Istanbul. *European Planning Studies*, 16(1): 143-155.
- ESRI GIS Dictionary. Spatial Autocorrelation. <https://support.esri.com/en/other-resources/gis-dictionary/term/spatial%20autocorrelation> [Accessed on Feb 12th, 2018].
- Giannopoulou, Maria, Roukounis, Yiannis & Stefanis, Vasilis (2012). Traffic Network and the Urban Environment: An Adapted Space Syntax Approach. *Social and Behavioral Sciences*, 48: 1887-1896.

- Hao, Lijun (2007). Research of the Impact of Rail Transit on the Development of Commercial Space in Shanghai. *Market Modernization*, 6(507): 166.
- Hiller, B., Penn, A., Hanson, J., Grajewski, T. & Xu, J. (1992). Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement. *Environment and Planning B*, 20: 29-66.
- Hotelling, Harold (1929). Stability in competition. *Economic Journal*, 39(153): 41–57.
- Hou, Chenbo (2014). Influence of Urban Rail Transit Station on Business District's Radiation Pattern. *Urban Trail Transit*, 27(6): 59-66.
- Hu, Min (2007). The Effects of Rail Transit on Urban Spatial Patterns. *Urban Transportation*, 11: 34-39.
- Huang, Xiaobing, Chen, Zhongnuan (2014). The Retail Business Structure of the Metro Site Based on the Information Entropy: A Case Study of the 15 Subway Sites in Guangzhou. *Economic Geography*, 34(3): 38-44.
- Jiao, Yao, Liu, Wangbao, Shi, Enming (2015). Research on the Spatial Distribution and Mechanism of Commercial Activities of Guangzhou Based upon Multi-Source POI Data. *Urban Insight*, 6: 86-96.
- Kunpeng Web Scraping. Introduction of common electronic map coordinates in China. <http://www.site-digger.com/html/articles/20140115/72.html>. [Accessed on Jan 25th, 2018]
- Li, Chengyu (2006). The Vicissitudes of Business Design in Shanghai. *Economy Shanghai*, 1: 72-73.
- Mitchell, Andy (2005). In *The ESRI Guide to GIS Analysis Volume 2: Spatial Measurements & Statistics*. Redlands, CA: ESRI Press.
- Poot, Jacques, Alimi, Omoniyi, Cameron, Michael & Maré, David (2016). The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science. Discussion Paper. <http://ftp.iza.org/dp10329.pdf> [Accessed on Sep 25th, 2017]
- Porta, Sergio & Romice, Ombretta (2014). Plot-Based Urbanism: Towards Time-Consciousness in Place-Making. Working Paper. <https://strathprints.strath.ac.uk/35640> [Accessed on Nov 28th, 2017]
- Riedel, Hans-Ulrich (2014) Chinese Metro Boom Shows No Sign of Abating. *International Railway Journey*. <http://www.railjournal.com/index.php/metros/chinese-metro-boom-shows-no-sign-of-abating.html?channel=525> [Accessed on Dec 5th, 2017]
- Rosenberg, Matt (2017) Reilly's Law of Retail Gravitation. <https://www.thoughtco.com/reillys-law-of-retail-gravitation-1433438> [Accessed on Nov 28th, 2017].
- Seaborn. Visualizing Linear Relationships. <https://seaborn.pydata.org/tutorial/regression.html#regression-tutorial> [Accessed on Feb 10th, 2018].

- Shanghai Metro. Hit! A New Peak of Per Day Ridership of Shanghai Metro!
<http://www.shmetro.com/node49/201803/con115037.htm> [Accessed on Jan 4th, 2018].
- Shanghai Municipal Bureau of Statistics. Shanghai Statistical Yearbook 2017.
<https://www.chinayearbooks.com/shanghai-statistical-yearbook-2017.html> [Accessed on Jan 5th, 2018].
- Shanghai Urban Planning and Land Resource Administration Bureau (2018). Shanghai Master Plan 2017-2035. <http://www.shanghai.gov.cn/newshanghai/xxgkfj/2035004.pdf> [Accessed on Jan 5th, 2018].
- Strano, Emanuele, Cardillo, Alessio, et al. (2007) Street centrality vs. commerce and service locations in cities: A Kernel Density Correlation case study in Bologna, Italy.
https://www.researchgate.net/publication/2178106_Street_centrality_vs_commerce_and_service_locations_in_cities_a_Kernel_Density_Correlation_case_study_in_Bologna_Italy [Accessed on Dec 2nd, 2017].
- Song, Peichen, Lin, Tao & Kong, Weiqiang (2010) Impact of Shanghai Urban Mass Transit on Spatial Distribution of Retail Business. *Urban Mass Transit*, 13(4): 25-28.
- Taneja, Sunil (1999). Technology moves in. *Chain Store Age*, 75(5): 136–138.
- Tsou, Kowan, Cheng, Haoteng (2013) The effect of multiple urban network structures on retail patterns: A case study in Taipei, Taiwan. *Cities*, 32:13-23.
- Wang, Fahui, Chen, Chen, Xiu, Chunliang, Zhang, Pingyu (2014) Location Analysis of Retail Stores in Changchun, China: A Street Centrality Perspective. *Cities*, 41: 54-63.
- Wang, Nengzhou, Zhang, Ji & Ma, Renfeng (2009) The Influence of Subway Economics on Spatial Pattern of Urban Retails. *City*, 12: 68-74.
- Wikipedia. Central Place Theory. https://en.wikipedia.org/wiki/Central_place_theory [Accessed on Nov 15th, 2017]
- Wikipedia. Demographics of Shanghai.
https://en.wikipedia.org/wiki/Demographics_of_Shanghai [Accessed on Jan 4th, 2018].
- Wikipedia. Reilly's law of retail gravitation.
https://en.wikipedia.org/wiki/Reilly%27s_law_of_retail_gravitation [Accessed on Nov 28th, 2017].
- Wikipedia. Shanghai Metro. https://en.wikipedia.org/wiki/Shanghai_Metro [Accessed on Jan 4th, 2018].
- Wikipedia. World Ranking of the Annual Ridership of Urban Subway Transit.
<https://zh.wikipedia.org/wiki/%E4%B8%96%E7%95%8C%E5%9F%8E%E5%B8%82%E8%BB%8C%E9%81%93%E4%BA%A4%E9%80%9A%E5%B9%B4%E4%B9%98%E5%AE%A2%E9%87%8F%E5%88%97%E8%A1%A8> [Accessed on Jan 4th, 2018].

Yu, Wenhao, Ai, Tinghua & Shao, Shiwei (2015) The Analysis and Delimitation of Central Business District Using Network Kernel Density Estimation. *Journal of Transport Geography*, 45:32-47.

Zhong, Chen, et al. (2015) Revealing Centrality in the Spatial Structure of Cities from Human Activity Patterns. *Urban Studies*, 1-19, DOI: 10.1177/0042098015601599.