

# Centripetal Cities: Using Big Data to Measure Job-Housing Separation Based on Employment-Residence-Commuting Trade-off

Ming Lu (✉ [luming1973@sjtu.edu.cn](mailto:luming1973@sjtu.edu.cn))

Shanghai Jiao Tong University

Chong Peng

Nanjing Audit university

---

## Article

**Keywords:** big data, urban spatial structure, job–housing separation, commuting

**Posted Date:** July 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-37632/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

With the advent of postindustrial society, populations are becoming increasingly concentrated in large cities, especially in urban centers. Here we study the “centripetal city” phenomenon. With many new service-industry jobs concentrated in central cities, people face the trade-off between employment, residence, and commuting. Using multisource big data from Shanghai, China, we develop a new job–housing separation index to reflect the trade-off between employment, housing price and commuting. We demonstrate that residents in central urban areas within a radius of approximately 20 km from the central business district tolerate job–housing separation in exchange for lower housing prices. Recent data indicate that job–housing separation accounts for 20% of housing prices. Our framework outperforms previous metrics, which not only provides a basis for understanding the formation and evolution of spatial structure in large cities, but can also guide wise planning and managing interventions for the United Nations Sustainable Development Goals.

## Introduction

Urban development trends indicate that urban economic activities, especially the complex ones, are concentrated in certain metropolises<sup>1</sup>. Over the past 30 years, several major metropolises around the world, such as New York City, London, and Tokyo, have witnessed increasing population re-concentration in central cities (See Supplementary Figs.1 and 2). This study refers to such cities as “centripetal cities.” The population re-concentration in big cities in the world also brings the costs of agglomeration to the urban development<sup>2</sup>, especially the challenges to commuting efficiency due to job-housing separation. Therefore, it is very important to understand the causes and consequences of the job-housing separation in big cities for the public policies and the sustainable development goals.

However, the existing job-housing separation indices are defined as resident workers/jobs ratio<sup>3</sup>, excessive commuting and commuting time<sup>4-6</sup>, etc., which have not fully considered the balance between employment, housing price and commuting. Earlier studies focused on the relationship between commuting and other variables, such as commuting and housing price<sup>7, 8</sup>, and commuting and wage<sup>9, 10</sup>. After that, scholars tried to analyze the relationship among wages, housing prices and commuting costs<sup>11</sup>, but, due to the limitations of data and methods, the trade-off between the three is difficult to be directly verified. In recent years, some economic theoretical models have considered the trade-off among wages, housing prices and commuting in cities<sup>12</sup>. Some empirical studies also used micro individual data to find that residents choose employment and residence by trading off between wages, housing prices and commuting time<sup>13</sup>, and that the wage premium is greatly associated with high housing prices and long commuting time<sup>14</sup>. However, none of the existing studies shows the job-housing separation in different positions in the spatial layout of the city. The empirical study considering the tradeoff between employment, housing price and commuting using household travel survey and population census<sup>15</sup> did not construct an index to illustrate intra-city spatial structure. Fortunately, the emergence and availability of big data, especially cellular phone data, provide larger and higher resolution and refined samples for the above research<sup>6, 16, 17</sup>. By using big data to identify residence, employment place and commuting

information, we can accurately measure the degree of job-housing separation in the city, and the demographic heterogeneity of different age and gender groups can also be further analyzed. Combined with other big data, we can better understand the characteristics of "centripetal city" and show how to balance urban wages, housing prices and commuting costs in family location decision.

The job-housing separation defined in this paper refers to the spatial distance between employment and residence. When deciding on a residence, people need to consider trade-offs between employment, residence, and commuting cost. Three behavioral combinations arise as a result: (1) residing in the city center, close to work but paying high housing prices along with better job opportunities; (2) residing in the suburbs, with low housing prices but longer commutes associated with better job opportunities in the city center; and (3) residing in the suburbs, paying lower housing prices but close to work with lower incomes, or less opportunities. It is expected that under the first two combinations, housing prices should be negatively correlated with job–housing separation<sup>18</sup>. While mobile phone data can be converted into our three indicators of interest—workplace, residence, and commuting distance—a better indicator can be constructed by the following formula:

$$\text{Job-housing separation index} = \ln\left(\frac{Jobs_{i,out}}{Jobs_{i,local}}\right) \times \sum (N_{ij} \times distance_{ij}), \quad (i \neq j) \quad , \quad (1)$$

Here,  $Jobs_{i,out}$  and  $Jobs_{i,local}$  represent, respectively, the number of employed individuals who reside in the  $i$ -grid area but work in other grids and the number of employed individuals who reside in the  $i$ -grid area and work locally.  $N_{ij}$  is the number of commuter trips from residence-grid  $i$  to employment-grid  $j$ , and  $distance_{ij}$  takes the Euclidean distance between residence-grid  $i$  and employment-grid  $j$  as the commuting distance per resident.

Here, we integrate multisource big data such as mobile phone tracking, housing prices, and restaurant consumption in Shanghai to analyze the formation mechanism of the centripetal city phenomenon (See Supplementary Fig 3 and Introduction on study area for details). When compared with the conventional indices, our job–housing separation index is a new one to capture the employment–commuting trade-off using big data and its relationship with housing prices, as well as the effects of subway construction. Specifically, we show that residents in urban areas within a radius of 20 km tolerated job–housing separation in exchange for lower housing prices. The phenomenon of exchanging longer commutes for lower housing prices is more obvious in the blocks with high proportion of males and young people. We further find that the centripetal city phenomenon in large global cities is driven by the large number of new service jobs that require interaction in central urban areas.

## Results

**More service-industry jobs are concentrated in the city center.** As noted by Balland et al., urban economic activities, especially complex ones, tend to be concentrated in a few large cities<sup>1</sup>. These activities require a deeper division of labor, knowledge, and specialization to ensure low coordination costs by creating

multiple mixing-and-matching opportunities<sup>1, 19</sup>. The same logic applies to inner urban areas. Complex economic activities are concentrated within the central city, which is conducive to efficient interaction and coordination<sup>20</sup>. Because of resource allocation through market mechanisms, prime city-center locations are acquired by industries and enterprises with the greatest willingness to pay. Furthermore, considering the service industry's increasing dominance in the urban economy, more service jobs are concentrated in the city centers. Indeed, Fig.1 shows the spatial distribution of new jobs created in producer services between 2000 and 2008 in Shanghai, China. It is evident that over time, new jobs in producer services became concentrated in city centers, especially the inner city—the centripetal force behind centripetal cities.

Besides producer services, most local services (e.g., gyms, financial services, restaurants, and theaters) require face-to-face interaction and are therefore local goods concentrating in central urban areas. These areas offer the advantage of location to conveniently provide goods and services to residents and enterprises alike. Furthermore, focusing on the consumption side, as people's income levels continue to increase, they increasingly demand premium, high-quality, diversified goods and services. Since most of these goods and services are not easily transported or stored, they tend to be concentrated in central urban areas with high-density populations, along with the corresponding employed population, thereby promoting the transformation of big cities into “consumer cities”<sup>21-23</sup>. Increasing employment concentration has been accompanied by increasingly concentrated consumption. A big city eventually becomes a consumer city given the requirements of variety, higher quality, and more diversified services. In this regard, city centers typically offer better consumption amenities.

Restaurant data are strongly predictive of spatial distribution of consumption activities<sup>24</sup>. Therefore, we used accessible and timely updated restaurant data from China's Dianping.com (Details in supplementary Note 1) to collect the number of good reviews, which were used to represent consumption quality (Robust metric details in supplementary Note 2). For the measurement of consumption diversity, we use Simpson's diversity index<sup>25, 26</sup> to measure the number of catering categories and the uniformity of distribution, as well as the measurement index of consumption diversity welfare. The formula is given as follows:

$$D=1-\sum_{i=0}^n(N_i/N)^2, \quad (2)$$

where  $D$  is the diversity of consumption,  $N_i$  is the amount of  $i$ -type cuisine in a grid,  $N$  is the total amount of all cuisines in a grid, and  $n$  is the total types of cuisines in a grid (38 types in this study). The value range of  $D$  is (0, 1). The greater the value of  $D$ , the higher the consumption diversity of each grid cell.

Fig. 2 shows that regardless of the consumption quality and diversity in Shanghai, high-value  $D$ s are mainly distributed in the central urban area, especially in the inner city (the area within the inner ring road).

Furthermore, as explained above, many new jobs in services are concentrated in central urban areas. This leads to these areas having goods and services that are more varied, of higher quality, and more diverse, which in turn drives a rich set of consumer activities and employment in consumer service industries. Therefore, many people (especially the employed) are concentrated in central urban areas during the day. However, due to high housing prices in central urban areas, residents face a job–housing price–commuting trade-off.

We used big data from anonymized mobile phones users as the main data source (See Methods for details), combined with census data, to depict the spatial characteristics of residents' employment, residence, commuting, and other behaviors. Mobile phone data include personal spatial information and changes throughout a day and can therefore depict urban spatial structures and residents' behaviors with finer granular geographic and temporal scales<sup>16, 27-31</sup>. We collect the gridded signaling dataset of mobile phone users in Shanghai in June 2019, provided by the telecom operator China Unicom. Each signaling observation included the user ID, time stamp, and cellphone tower coordinates. A significant advantage of using population census and mobile phone data is that they show spatial distribution changes in the population in small geographical scales.

**The population has seen a recentralizing trend during the past decade.** We observe the spatial population distribution characteristics on a 250-m grid. Interestingly, within a short span of 20 years, Shanghai has undergone a trend shifting from population decentralization to recentralization. From 2000 to 2010, the population density of most parts of the city center declined significantly (See Figs.3a, b). From 2010 to 2019, this trend reversed, and the areas with high population density growth were concentrated in the inner city. By superposing Shanghai's subway tracks, where the population is increasing in the suburbs is also closer to the subway. According to the classical urban economics theory, in a monocentric city, population density shows a decreasing trend as the distance to the central business district, or CBD, increases<sup>32,33</sup>. With the population moving closer to downtown over time, the population density gradient becomes steeper. (See Supplementary Fig.4 for details).

**The recentralization of population is associated with job–housing separation.** Next, mobile phone data from June 2019 was used to construct the commute flow data of employees in Shanghai within 250 m grids to calculate the job–housing separation index for each grid cell. We should first point out that the traditional job–housing separation index does not adequately capture the job–housing price–commuting trade-off in the formation of job–housing separation. Traditionally, the job–housing separation index has focused on the balance between quantity<sup>34</sup> and quality<sup>35</sup>. First, the balancing quantity has been measured by the traditional job–housing ratio, which refers to the number of jobs divided by the number of residents in a given area. Second, for balancing quality, previous studies have often used Thomas's independent index—the ratio of the number of people residing and working in an area to the number of people residing in the area but working outside of it. However, neither index includes information about the severity of job–housing separation.

Fig.4a shows that areas experiencing severe job–housing separation are mainly concentrated in the suburbs. The developed public transportation systems in large cities (especially subways) can alleviate the time and psychological costs of job–housing separation. Thus, although it is difficult to measure the exact toll, we can presume that under the same level of job–housing separation, housing prices located closer to subways are higher. By superposing Shanghai’s subway tracks, we further found that these areas often coincide with the outer suburbs covered by subways. This indicates that employed individuals who choose to reside in the outer suburbs and work in the city center reduce commute time and psychological costs by using the subway. The degree of job–housing separation in central urban areas within the outer ring road, especially in the inner city, is relatively low, reflecting the high degree of job–housing balance in those regions. To verify the accuracy of the improved index, we also calculated and visualized a separation index based on Thomas’s algorithm above mentioned as a comparison. The results showed that the spatial regularity reflected by the traditional index, including the coincidence effect in terms of subway tracks, is obviously inferior to the reflections of our improved indicator (See Fig4.b).

**Centripetal Commuting of Shanghai Residents.** Job–housing separation and commuting can also be determined using the urban commuting network system. We referred to Taylor and Derudder’s world city network method<sup>36</sup>. Furthermore, as above, we used mobile phone signaling data from June 2019 in Shanghai to construct a 250-m grid dataset of journey-to-work commuting flow to depict the commuting characteristics and modes of Shanghai employees (See Fig. 5). In Figs. 5a-b, numerous residents employed in the inner (within the inner ring road) and central (between the inner and outer ring road) city travel from various areas in the city. Fig. 5c shows that most residents employed in the suburbs (outside the outer ring road) also reside there, and therefore work nearby, which is considered suburb–suburb commuting.

Next, we took the total employed population as a sample and calculated the number and proportion of employees in different circles according to their residence in various locations. We further analyzed the spatial sources of employees from these locations in the city (See Fig.5). The results show that, first, 53.88% of employed individuals in the entire city are concentrated in the central urban area, of which 75.24% travel from local residences and 24.76% travel from the suburb. Second, the number of jobs in the inner, central, and suburban regions is relatively high, accounting for 39.14%, 63.61%, and 88.88%, respectively, of the total number of jobs in the respective regions. This indicates that the degree of self-containment (Details in Supplementary Note 3) in the suburb is relatively high. Third, taking the entire city into account, the number of individuals from the central city working in the suburb accounts for only 5.13% of total employment in the city (i.e., city–suburb commuters). Meanwhile, the number of employees from the suburbs working in the central city accounts for 13.34% (i.e., suburb–city commuters). Apart from the centralizing force of service employment and consumption in the central urban area, the rapid development of the subways has alleviated the costs of job–housing separation.

Since we considered employment, residence, and commuting as a set of variables determined simultaneously, the degree of job–housing separation should be related to housing prices. In particular,

residents who work in the central urban area accept a trade-off between long commuting distances and low housing prices. Therefore, we further overlaid the spatial distribution map of the subway tracks and housing prices for analysis. Housing price data are collected from *Lianjia beike* (<https://sh.ke.com/>), the largest, the most detailed and reliable real-estate website in China. This dataset contains the transaction price, transaction time, detailed address, and other housing attributes of more than 150,000 commercial houses between January 1, 2015, and October 17, 2019. For this, we used a web crawler and geocoded and vectorized the corresponding detailed addresses. Finally, housing prices were averaged in a grid with a side length of 250 m. Fig. 6a shows that housing prices were the highest in the inner and central city, which experience lower levels of job–housing separation. This shows that residents choose to reside and work in the central urban area to enjoy lower commuting costs at the expense of paying more for housing. Meanwhile, residents in the outer suburban areas who use the subway system face relatively low housing prices and receive higher incomes by working in the city center, but they pay higher commuting costs. Although subways reduce the time and psychological costs of long-distance commuting, the prices of properties located near subways are higher than those farther away.

As explained above, the degree of job–housing separation should correlate negatively with the prices of properties within a short distance of the central urban area. Fig. 6b shows the change of housing prices and the job–housing separation gradient for the period 2015–2019. Since the outer suburbs can have a combination of low housing prices, nearby employment, and low incomes, we did not include subjects located more than 20 km from the CBD. As indicated by the Fig.6b, the degree of job–housing separation in each year shows a negative relationship with housing prices. OLS regression analysis using the latest data from 2019 shows that job–housing separation accounts for 20% of housing prices.

We present the gradient fitting curve between the job–housing separation and housing prices in 2019, using two subsamples data of 500 m and 500–1000 m from the subway in Shanghai’s suburbs. We then investigated the gradient relationship between the two using different locations in relation to the subway. The gradient of job–housing separation and housing prices in the area 500 m from the subway is clearly lower than that in the area 500–1000 m from the subway (See Fig. 6c). Thus, as indicated earlier, we confirmed that a developed public transportation system can alleviate the time and psychological costs of long-distance commuting for employees, though at the expense of higher housing prices.

Based on the above analysis, we can predict that if employment continues to concentrate in central urban areas, and people remain reluctant to pay high commuting costs, the gradient curve between the degree of job–housing separation and housing prices will show an increasingly steeper trend over time. However, whether this trend is realized will be influenced by residents’ preferences, improved traffic flows, and the housing supply in central urban areas. In other words, if residents prefer long commutes to high housing prices, or traffic conditions significantly improve and decrease commuting costs, or the housing supply in central urban areas increases, the gradient curve between the job–housing separation index and housing prices will not change significantly. The results shown in Fig.6b indicate that the relationship between the job–housing separation index and housing prices remained relatively stable between 2015 and 2019.

However, in the three years since 2017, the gradient curve has become steeper. Whether this trend will continue should be investigated over a longer period.

**Demographic heterogeneity analysis.** The agglomeration of population to the central city leads to the trade-off between employment, residence and housing price, but there are also significant differences among individuals with different demographic characteristics. We further use the mobile phone signaling data to count the gender and age of each 250 meter grid working population. According to the demographic attributes of more than half of the grid, we take them as the main attributes of the grid, so as to investigate the gradient of job-housing separation and housing price under different gender and age groups (see Fig. 7a and b). It can be seen from figure 7a that the gradient between the job-housing separation and the housing price of male dominating grids is lower than that of female dominating grids. As Thomas et al. 37 pointed out, the gender difference of commuting distance is huge. When women are re-employed, their wages are lower and their commuting time is shorter, so the indifference curve of female wage and commuting is steeper. About 10% of the gender wage gap can be explained by the gender gap in the most acceptable commuting. Our study shows that male employees prefer to commute farther in exchange for lower housing prices. Similarly, compared with the older (35 years old and above) employees, the younger (35 years old below) employees have a lower gradient of job-housing separation and housing prices. In a nutshell, men and young people are more tolerant of the cost of long-distance commuting.

## Discussion

With the rise of the Internet, mobile terminals and data sensors, and big data mining technology, it is possible to capture the distribution of complex spatial economic activities and the evolutionary laws of big cities more precisely than ever that will support policy needs<sup>38, 39</sup>. We have analyzed the formation mechanism of the centripetal city phenomenon by integrating multisource big data from Shanghai, China, such as mobile phone tracking and housing transactions. We develop a job–housing separation index to estimate the balance between employment and commuting and found that residents in urban areas within a radius of 20 km tolerated job–housing separation in exchange for lower housing prices. The existing indices of job-housing separation can not reflect the trade-off among employment, commuting and housing price. The index proposed in this paper can show the above-mentioned relationship in space. The demographic heterogeneity reflected by the gradient between the job-housing separation and the house price shows that men and young people are more tolerant of the costs of long-distance commuting. Furthermore, our results indicate that the centripetal city phenomenon in large global cities is driven by the large number of new service jobs that require interaction in central urban areas. One contribution of our work is its innovative use of big data and network science to construct a more accurate job–housing separation index based on behavioral analysis that can capture the employment–commuting trade-off and its relationship with housing prices, as well as the effects of subway construction. Approximately 95% of urban expansion in the future is expected to take place in developing countries. Therefore, in terms of the integration of economics and natural science, this study contributes to the literature by summarizing the spatial structure characteristics and evolution of



consumption, population, and commuting in a leading city in the world's largest developing country. From a public-policy point of view, this study provides not only a basis for understanding the formation and evolution of spatial structure in large cities but also a guide for wise planning and managing interventions during the era of inclusive and sustainable development. It is worth further pointing out that if administrative forces are used to restrict the development of central urban areas, the result may be loss of welfare, either high housing prices, or long commutes. And there are gender and age differences in this effect. Young people and men are more likely to take on long commutes, while older people and women are more likely to take on high housing prices.

## Methods

**Population Data Processing Methods.** We process the population data included three steps: selecting the period, cleaning the data, and constructing the model and calculating the indicators.

First, regarding the study period, the spatial distribution of population varies between different periods, especially in population distributions with instantaneous changes. Seasons, holidays, weather, and other factors all affect human activity, which is reflected on the diversity of the population's spatial distribution. Based on this, the spatial distribution of the urban population can be accurately reflected only when an appropriate period is selected. China's population census has been held on June 30, 1953; June 30, 1964; July 1, 1982; July 1, 1990; November 1, 2000; and November 1, 2010. Therefore, the months of June and November could be considered. Based on the 2018 data, we compared monthly population changes in the first- and second-tier cities (Beijing, Shenzhen, Chengdu, Wuhan, Wulumuqi, and Changchun) and third- and fourth-tier cities (Hengyang, Nanchong, Nanyang, Siping, Weifang, and Weinan). Next, we compared the population in June and November with the annual mean and median. The comparison showed that June was the ideal month. Finally, we selected 30 consecutive days of data in June 2019 to ensure the basic data source would be more accurate and of higher quality than sources used in previous studies.

Second, regarding data cleaning, big data can accept inaccuracies, but many improvements are still required in basic data processing using mobile big data. To ensure the reliability and scientific validity of the results, the original data were carefully cleaned. First, we ruled out abnormalities. Some of the mobile signaling data lacked latitude and longitude coordinates, affecting the analysis. Thus, we directly eliminated these specific data and signaling records. Second, we eliminated the "ping-pong effect." If the signal strength of two base stations changes dramatically in a certain area within the mobile communication system, the phone pings will switch back and forth between the two base stations, while the corresponding signaling records will continue, creating a so-called ping-pong effect. The specific process flow was as follows: the daily signaling records were counted, and the number and distance of switching between each pair was investigated for base station pairs. Simultaneously, the time duration was considered. Generally, if there are multiple switches in a short time and within short distances, the ping-pong effect is assumed to exist, which is then eliminated. However, long-distance signaling switching is considered valid data to be retained.

Third, regarding model construction and indicator calculation, equipment remaining in Shanghai for more than 10 days in a month as processing samples was selected, and the number of employed individuals in the daytime and the number of residents at night in Shanghai were calculated. The daytime distribution of the equipment was used to select the grid with the most accumulated working time in the day from 9:00 a.m. to 12:00 p.m. and from 1:00 p.m. to 6:00 p.m. The data for the night distribution were selected from 9:00 p.m. to 7:00 a.m. the next day, which is the highest accumulated time in a month. It is assumed that the nighttime location is the residence location of the equipment owners while the daytime location is their employment location. Previous studies have indicated that one of the biggest challenges and uncertainties when using cellular data to conduct urban research is the difference in market penetration rates between various regions and mobile phone companies. Concurrently, when mobile phone user data expand, the mobile phone penetration rate of a city will affect the results<sup>1</sup>. To address these issues, and on the basis of identifying effective users, we used the market share at the district-county level of Unicom to carry out sample expansion and then conducted a second check with reference to the mobile phone penetration rate in the city. Ultimately, we obtained the total population of the final 250 m grid cell sample expansion during the day and at night. The relevant population augmentation methods and processes are detailed below. Considering the maximum deviation caused by possible errors, we carried out a 0.5% tail reduction in the data.

The method and process for population augmentation consisted of three steps: (1) Exclude non-people number cards (p1). Unicom's number cards are not only sent to actual people but also include Internet-of-Things equipment. These users were excluded. (2) Calculate Unicom's market share (p2). It mainly goes through the following three steps—First, Use Unicom users as seeds to identify their true location, which serves as the benchmark of the calculation. Second, based on the above assumptions, the weighted average method is used to estimate the location of users on different networks (including China Mobile Communications Corporation and China Telecom users) with a call connection with Unicom users. Based on the location of Unicom users and the estimated location of users on different networks, the district-county-level market share of Unicom users is obtained. The last but not least, Check the district-county-level market share of Unicom users using the estimations above with the provincial-level market share of Unicom; use this as the final district-county-level market share. (3) Calculate the mobile phone penetration rate (p3). Use the number of all valid mobile phone numbers per 100 residents. This solves the problem of counting one person with more than one card multiple times and of excluding users without mobile phones. The mobile phone penetration rate in Shanghai was found to be 1.46/100.

## Declarations

### Data availability

All data supporting the study's findings are available from the corresponding author upon reasonable request.

### Code availability

The code that supports this study's findings is available from the corresponding author upon request.

## Acknowledgments

## Author contributions

contributed equally to this work and jointly oversaw it.

## Additional information

**Supplementary Information** accompanies this paper at

**Competing interests:** The authors have no competing interests to declare.

**Peer review information:**

**Publisher's note:**

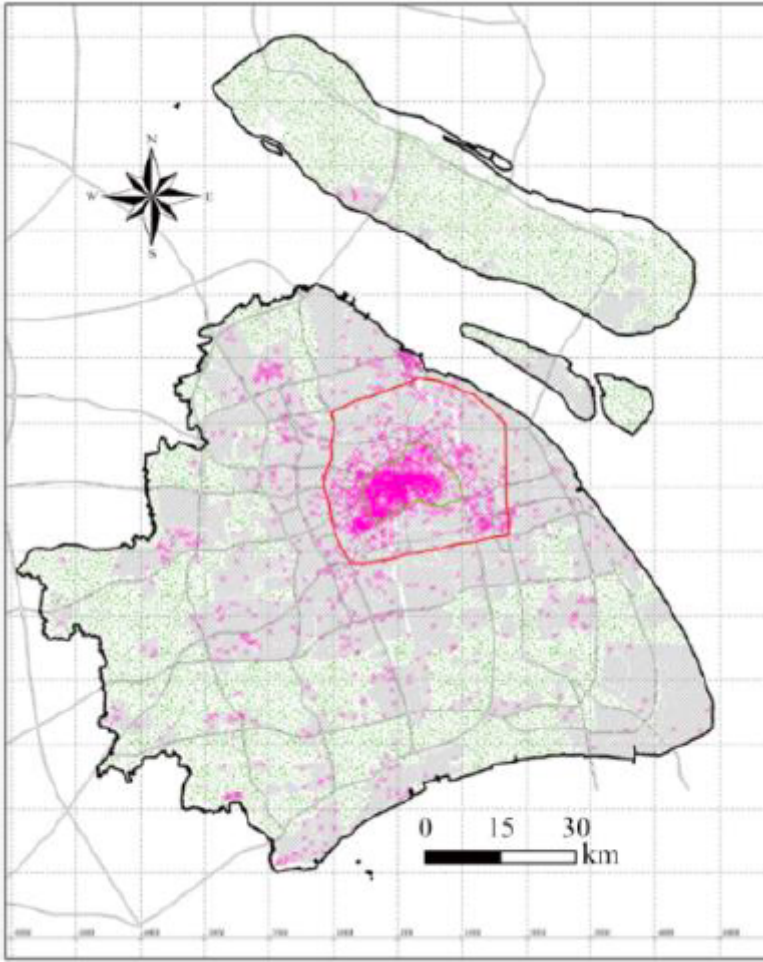
## References

1. Balland, et al. Complex economic activities concentrate in large cities. *Nature Human Behaviour*. <http://dx.doi.org/10.2139/ssrn.3219155> (2020).
2. Glaeser, E.L. Are Cities Dying. *Econ. Perspec.* **12**, 139-160 (1998).
3. Peng, Z. The jobs-housing balance and urban commuting. *Stud.* **34**, 1215-1235 (1997).
4. Horner, M. & Murray, A. Excess commuting and the modifiable areal unit Problem. *Stud.* **39**, 131-139 (2002).
5. Loo, B.P.Y. & Chow, A.S.Y. Jobs-housing balance in an era of population decentralization: An analytical framework and a case study. *Transp. Geogr.* **19**, 552-562 (2011).
6. Zhang, , Zhou, J. & Zhang, T. Quantifying and visualizing jobs-housing balance with big data: A case study of Shanghai. *Cities* **66**,10-22 (2017).
7. Zax, J.S. Compensation for commutes in labor and housing markets. *J. Urban.* **30**, 192-207 (1991).
8. Tse, C.Y. & Chan, A.W.H. Estimating the commuting cost and commuting time property price gradients. *Reg. Sci Urban. Econ.* **33**, 745-767 (2003).
9. Petitte, R.A. & Ross, S.L. Commutes, neighborhood effects, and compensating differentials: revisited. *J. Urban.* **46**, 1-24 (1999).
10. Timothy, D. & Wheaton, W.C. Intra-Urban wage variation, employment location, and commuting times. *J. Urban.* **50**, 338-366 (2001).
11. Lucas, R.E. & Rossi-Hansberg, E. On the internal structure of cities. *Econometrica* **70**, 1445-1476 (2002).

12. Manning, A. The real thin theory: monopsony in modern labour markets. *Labour. Econ.***10**, 105- 131 (2003).
13. So, S., Orazem, P.F. & Otto, D.M. The effects of housing prices, wages, and commuting time on joint residential and job location choices. *Am. J Agric. Econ.* **83**, 1036-1048 (2001).
14. Mayock, T. Wages, Housing prices and commutes. *Real. Estate. Econ.* **44**, 258-300 (2016).
15. Suzuki, & Lee, S. Jobs – housing imbalance, spatial correlation, and excess commuting. *Transport. Res A-PO.* **46**, 322-336 (2012).
16. Zhou, X., Yeh, A.G.O. & Yue, Y. Spatial variation of self-containment and jobs-housing balance in Shenzhen using cellphone big data. *Transp. Geogr.* **68**, 102-108 (2018).
17. Ahas, R., Aasa, A., Mark, Ü., Pae, T. & Kull, A. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management* **28**, 898-910 (2007).
18. Huiran, H., Chengfeng, Y. & Enru, W. Evolution of jobs-housing spatial relationship in Beijing metropolitan area: A job accessibility perspective. *Geogr. Sci.* **25**, 115-128 (2015).
19. Bettencourt, L.M. The origins of scaling in cities. *Science* **340**, 1438-1441 (2013).
20. Du, R. & Zheng, S. Agglomeration, housing affordability, and new firm formation: The role of subway network. *J. Hous. Econ.* **48**, 101668 (2020).
21. Glaeser, E.L., Kolko, J. & Saiz, A. Consumer city. *J. Econ. Geogr.* **1**, 27-50 (2001).
22. Glaeser, L. & Gottlieb, J.D. Urban resurgence and the consumer city. *Urban. Stud.* **43**, 1275-1299 (2006).
23. Schiff, N. Cities and product variety: evidence from restaurants. *J. Econ. Geogr.* **15**, 1085-1123 (2015).
24. Dong, , Ratti, C. & Zheng, S. Predicting neighborhoods' socioeconomic attributes using restaurant data. *Proc. Natl Acad. Sci. USA* **116**, 15447-15452 (2019).
25. Simpson, E.H. Measurement of diversity. *Nature* **4148**, 688 (1949).
26. Schaub, , Buchmann, N., Lüscher, A. & Finger, R. Economic benefits from plant species diversity in intensively managed grasslands. *Ecol. Econ.* **168**, 106488 (2020).
27. Blumenstock, J. Don't forget people in the use of big data for development. *Nature* **561**, 170-172 (2018).
28. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350**, 1073-1076 (2015).
29. Louail, T. et al. From mobile phone data to the spatial structure of cities. *SCI REP-UK* **4**, 1-11 (2015).
30. Chen, M.K. & Pope, D.G. Geographic mobility in America: Evidence from Cell Phone *NBERw27072*(2020).
31. Deville, et al. Dynamic population mapping using mobile phone data. *Proc. Natl Acad. Sci. USA* **111**, 15888-15893 (2014).

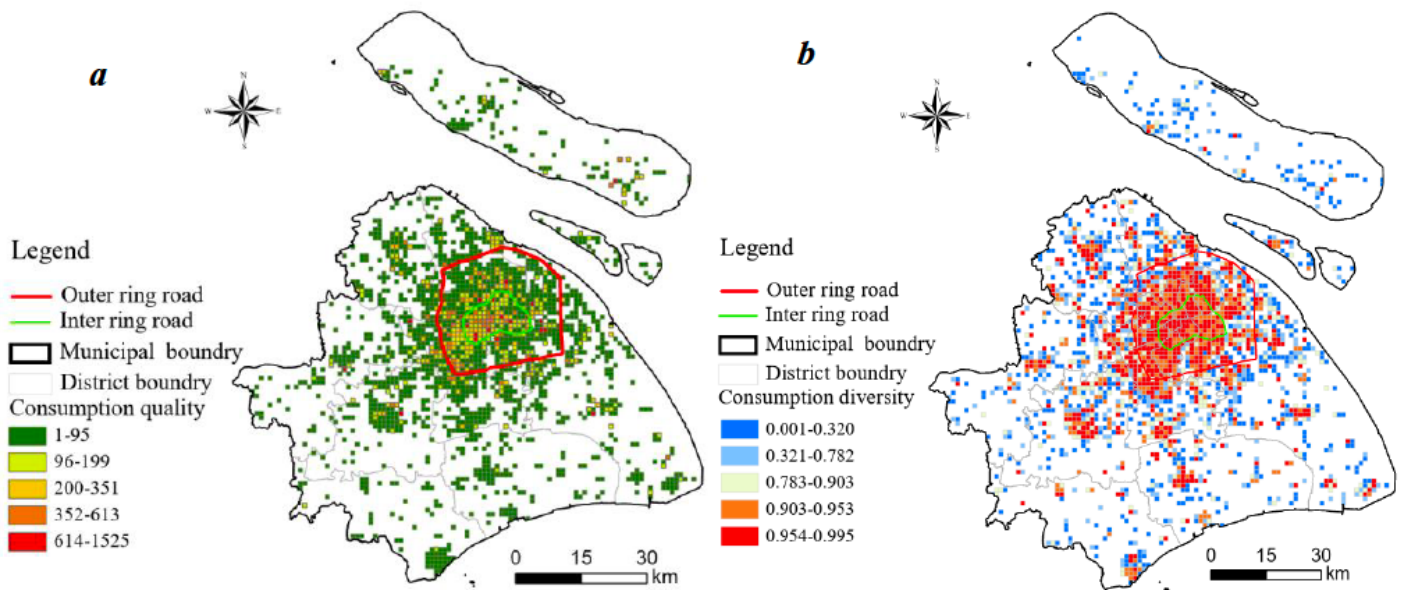
32. Duranton, G. & Puga, D. *Urban land use. Handbook of Regional and Urban Economics, Volume 5A*.<http://dx.doi.org/10.1016/B978-0-444-59517-1.00008-8> (Elsevier, Netherlands, 2015).
33. Clark, C. Urban population densities. *Royal Stat. Soci.* **114**, 490-496 (1951).
34. Horner, Spatial dimensions of urban commuting: A review of major issues and their implications for future geographic research. *Prof. Geogr.* **2**, 160-173 (2004).
35. Thomas, R. *London's New Towns: A study of self-contained and balanced* <http://worldcat.org/isbn/0853740089> (Political and Economic Planning (PEP): London, 1969).
36. Taylor, P.J. & Derudder, B. *World city network: A global urban analysis*. (Routledge, New York, 2016).
37. Le Barbanchon, T., Rathelot, R. & Roulet, A. Gender differences in job search: Trading off commute against wage. Working Paper,
38. Glaeser, E.L., Kominers, S.D., Luca, M. & Naik, N. Big data and big cities: the promises and limitations of improved measures of urban life. *Econ. Inq.* **56**, 114-137 (2018).
39. Athey, S. Beyond prediction: Using big data for policy problems. *Science* **355**, 483-485 (2017).

## Figures



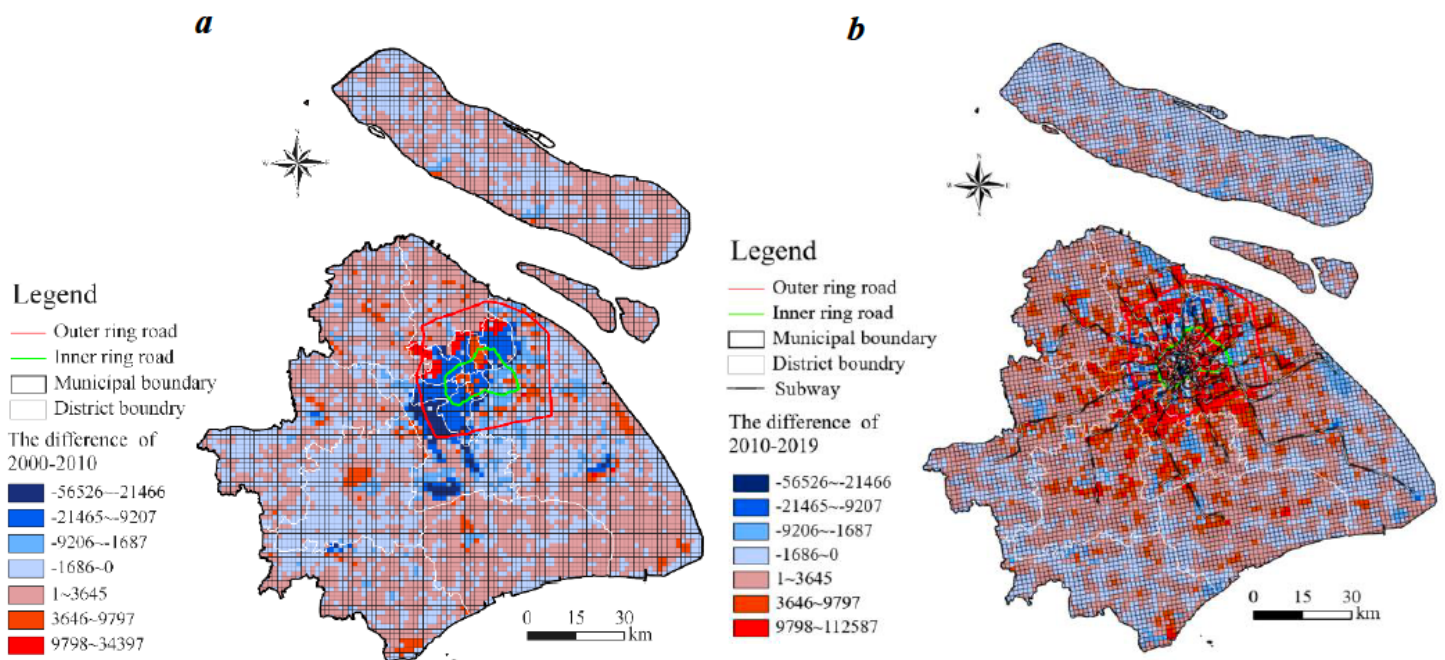
**Figure 1**

spatial distribution of new jobs in producer services in Shanghai. "X" in pink color represents the number of new jobs in service industry. Each unit of "X" is 1,000 job. Calculated based on 2000 and 2008 economic census data. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 2**

Spatial pattern of consumption quality (Fig.2a) and consumption diversity (Fig.2b) (using the restaurant industry as an example). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

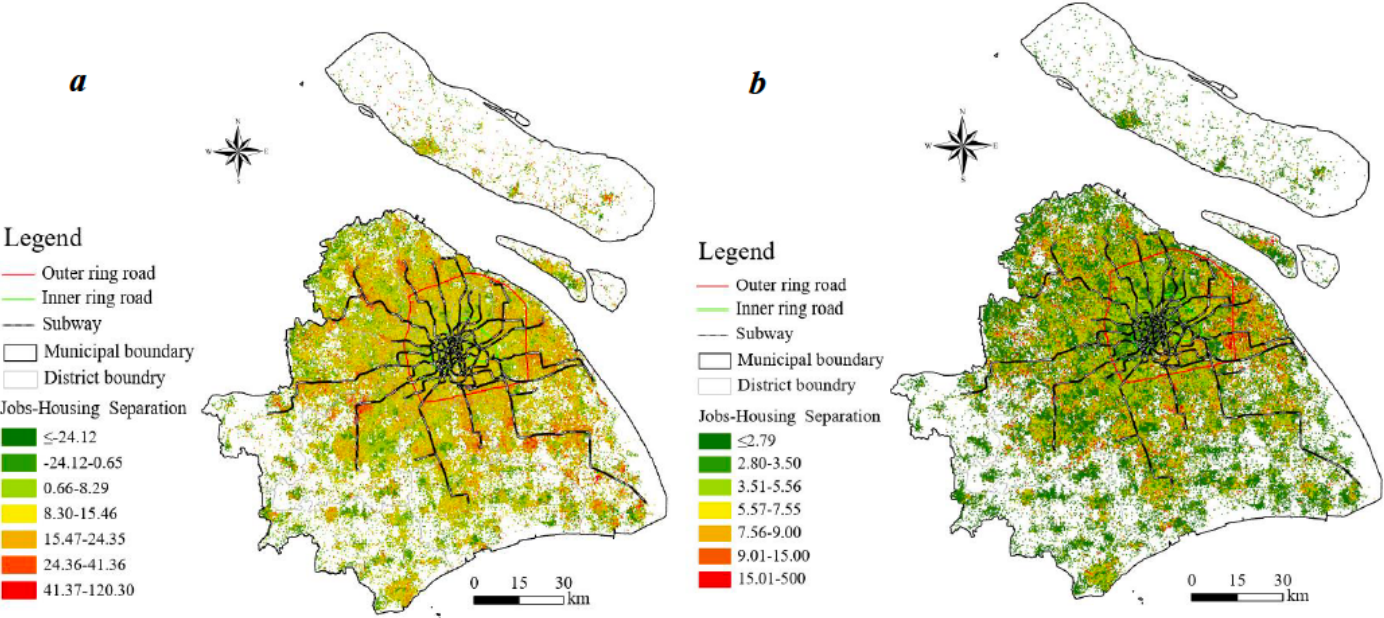


**Figure 3**

a and b represent population changes in Shanghai from 2000 to 2019. (1) The 1-km grid population data for 2000 and 2010 are from the Chinese Academy of Sciences (<http://www.resdc.cn/DOI>). The 1 km grid



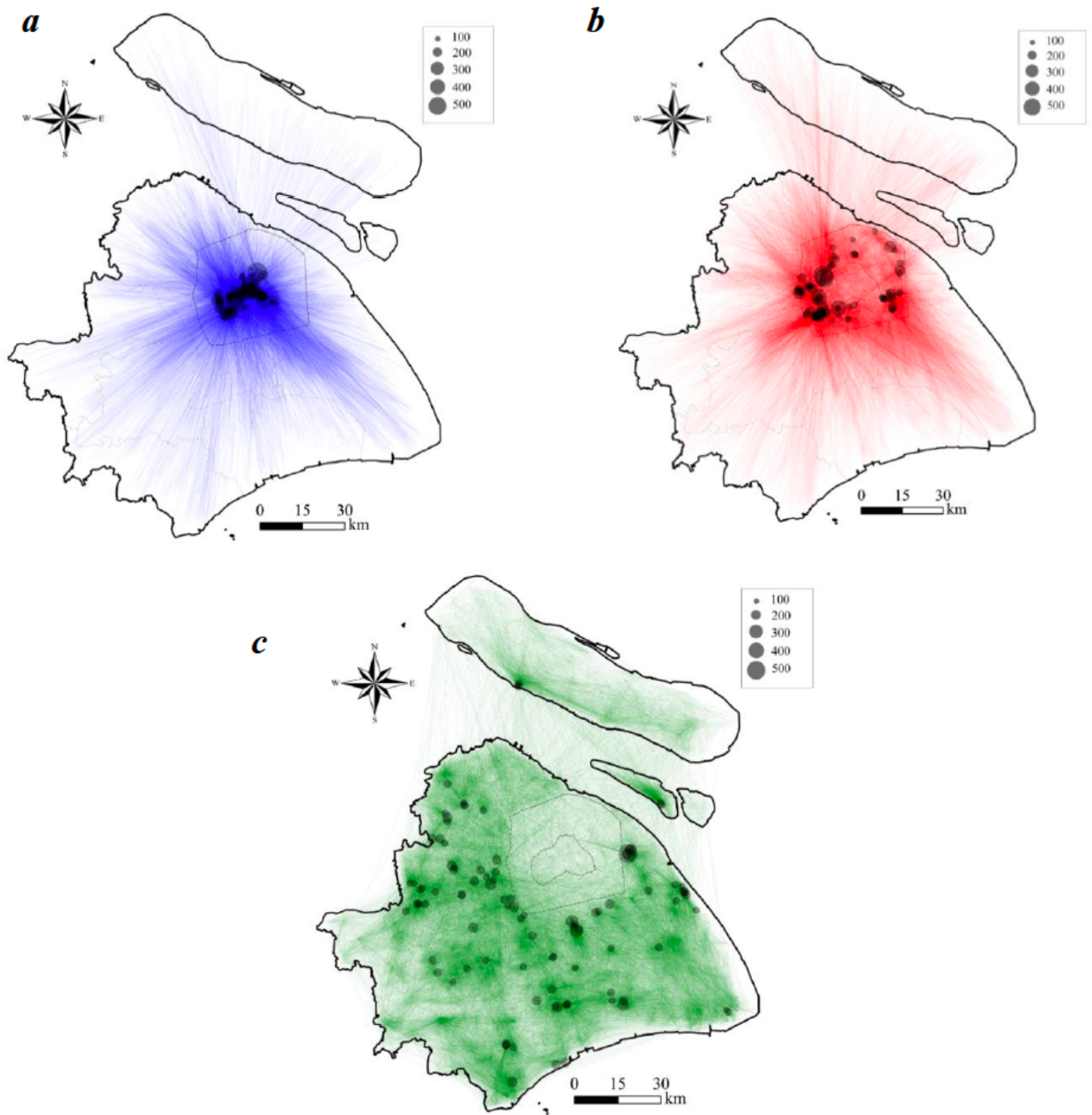
dataset for the spatial distribution of population in China comes from the Data Registration and Publishing System of Resource and Environmental Science Data Center, Chinese Academy of Sciences (<http://www.resdc.cn/DOI>), 2017. DOI:10.12078/2017121101. (2) To compare the population data in 2000 and 2010, the nighttime population and daytime population in 2019 on 250-m grid cell are aggregated into 1- km grid cell. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 4**

a is spatial distribution of job–housing separation in Shanghai by new methods. b is spatial distribution chart of the Job–Housing Separation Index calculated by Thomas’s Algorithm. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

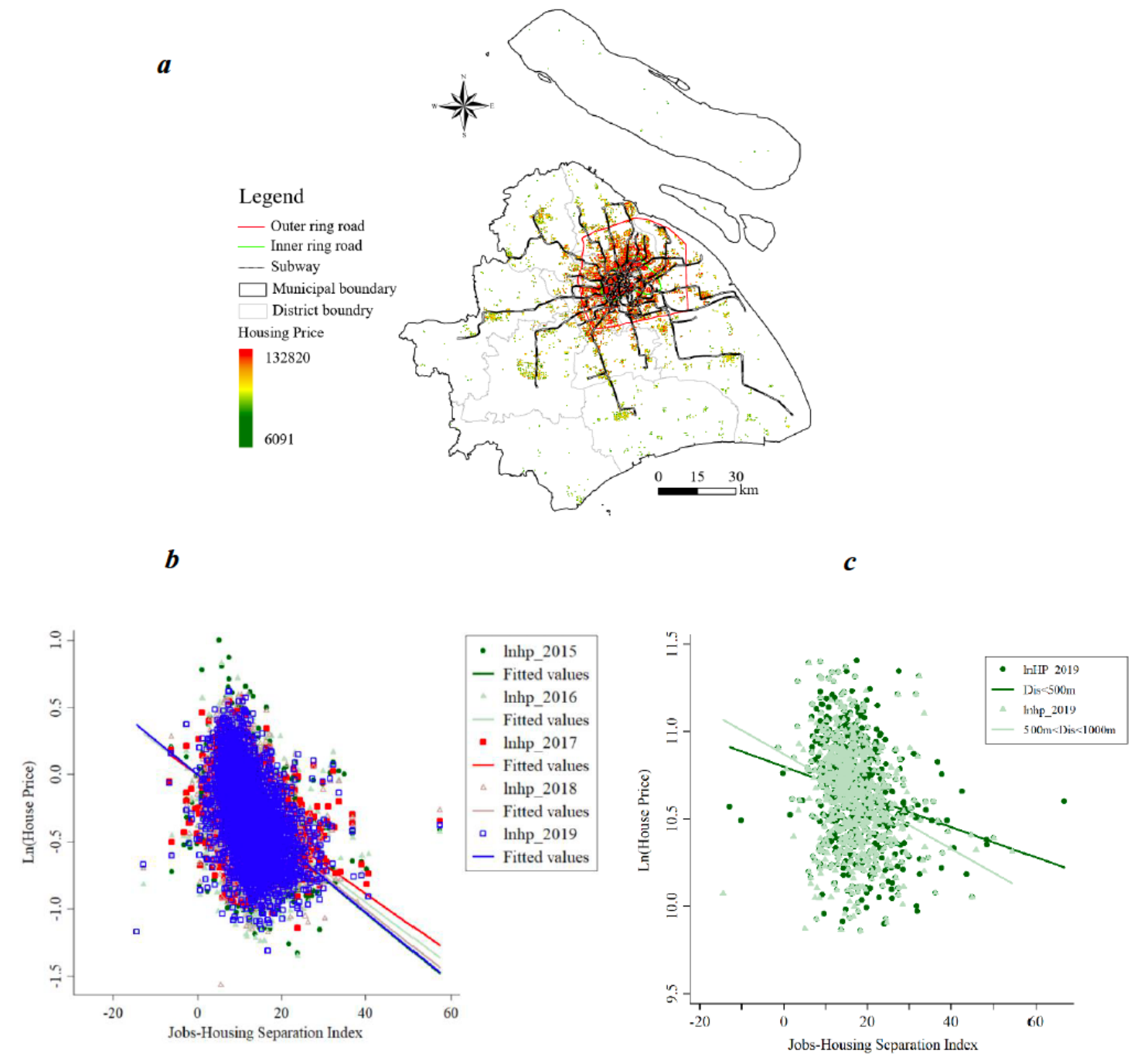




**Figure 5**

Commuting Direction of Employed People in Different Regions of Shanghai. We only selected the top three lines of each departure grid to the destination. The place of employment aggregates the number of employees flowing in from all places of residence. For representation, the first 100 places were selected in the form of circles in the order of size. Circle size indicates the number of places where the workplace originates. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal

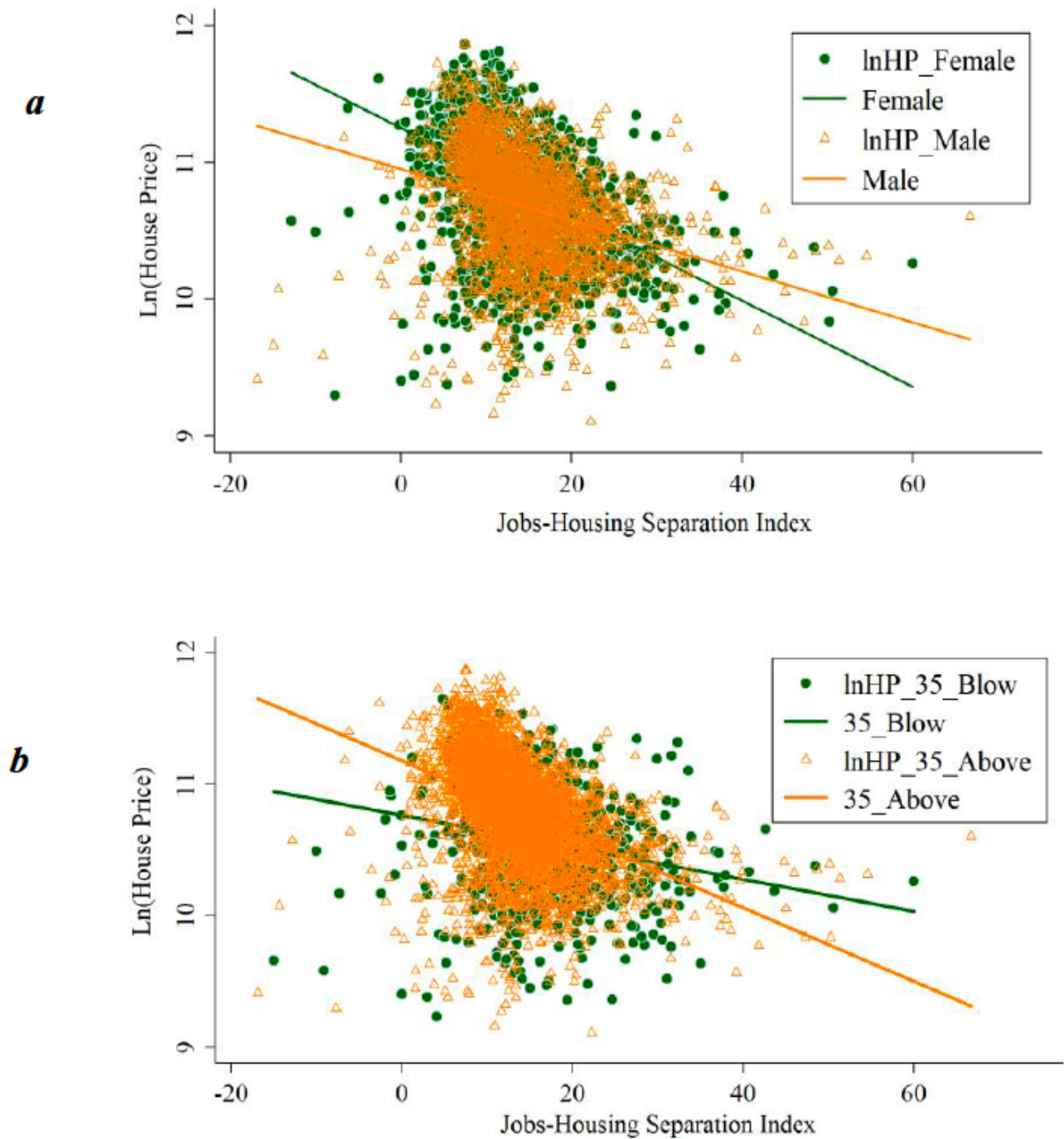
status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 6**

Relationship between housing prices and job–housing separation. Fig.6a is the spatial distribution of housing prices in Shanghai. Note: The unit of house prices is Yuan. In Fig. 6b, the curves in the figure are the fitting results for different years, with the intercept terms by ordinary least-squares (OLS) estimation removed for cross-time comparison. Fig. 6c refers to the relationship between job–housing separation and housing prices in the suburbs. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square

concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 7**

The demographic heterogeneity of the relationship between the job-housing separation and housing prices (different gender and age). The sample of Fig.7a is divided into two groups according to gender, and Fig.7b is divided into two groups according to age, respectively under 35 years old, and 35 years old and above.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [CentripetalCitiesSupplementary.pdf](#)