TECHNICAL ESSAY #11 - ANALYSIS USING OPEN DATA

Zixiao Tang (zt232)

Link to GitHub repository: https://github.com/tangzixiao/TE-11_Zixiao

1. Provide a short summary (2-3 paragraphs) on the data sourcing and cleaning strategy you used. Which datasets did you use and why? How did you select your geographies and variables of interest?

Geographies and variables of interest:

The analysis of the living circle reflects the living utility that a place can provide for residents, which is a vital tool for planners to examine the livability of regions. However, the livability of the living place often corresponds to the high living cost, which leads to the potential spatial injustice, so compare livability with housing price is also crucial.

Based on the existing big data, this paper analyzes the spatial distribution structure of the supply of traffic, green space, business, and education in Beijing to reveal the unbalanced distribution in the city, since these variables are key data layers to the definition of 'living circle'. This paper compares the convenience of the living circle with the housing price of each community and observes some trends of spatial injustice in Beijing to provide a reference for the further construction of the living circle of the city.

Data sourcing and cleaning strategy:

For the data selection of this study, the first requirement is that the degree of spatial refinement can support the analysis at the community level. In addition, different data contents need to cover similar times and the same spatial scope. The

classification of service facilities is difficult because the POI data in the public data set is not comprehensive, and the classification is messy, especially for Beijing.

Based on the above reasons, this paper selected the following data for research on the open platform. See **Table 1**.

i) **Residential communities with the unit price of housing in RMB (point shapefile).**

Housing price for 7832 housing projects in Beijing, as of Oct 2013. The data is crawled from Ganjiwang (a real estate agency website) by Beijing City Lab member WANG Jianghao. (link: https://www.beijingcitylab.com/data-released/data1-20/)

ii) **Beijing land parcels with function types (polygon shapefile).**

Long and Liu[7] generated parcel maps for 297 Chinese cities in 2013. They identified the characteristics of city parcels with OpenStreetMap data and POI (points of interest) data. Each parcel is associated with urban function, density, and land use mix degree. In the FUNCTION variable, they classified the parcels into government (GOV), Transport facilities (TRA), Commercial establishments (COM), Education (EDU), Firms (FIR), Greenspace (GRE), Residential land (RES), and Others(OTH). (link: https://www.beijingcitylab.com/data-released/data1-20/)

Long and Liu's parcel analysis is the most comprehensive and detailed data on the classification of functional areas in Beijing in the existing research so far. In addition, the 'functions' variable in their study can meet the needs of functional analysis of the living circle. Therefore, this paper uses the data conclusions of Long and Liu's study as the underlying data of the analysis. **By selecting data using this 'FUNCTION'**

variable, we can derive the green space dataset (GRE), commercial dataset(COM), and education dataset(EDU) in the following analysis.

In ArcGIS, selecting the data by attribute "CITY_ID = 1" to get the Beijing parcels from the national scale dataset.

iii) Beijing Bus stops (point shapefile).

Data of 1543 bus routes and 42,161 stops of Beijing. The data were crawled by the BCL research fellow Dr. Jianghao Wang in 2013. Set the coordinate system as "GCS WGS 1984". This layer of data can be used to measure the transportation convenience of neighborhoods. (link: https://www.beijingcitylab.com/data-released/data1-20/) This data is used for the bus stop dataset analysis.

iv) Base map: Beijing OpenStreetMap data (line shapefile).

The analysis used OpenStreetMap data (mainly the road network data) to locate the datasets mentioned above and visualize the maps.

Table 1 Data Resources and Content

Data Content	Data Type	Data Resource
Communities with housing price	Points	Ganjiwang, BCL
Function Parcels	Polygons	Long and Liu
Bus Stops	Points	BCL
Beijing Base Map	Lines	OpenStreetMap

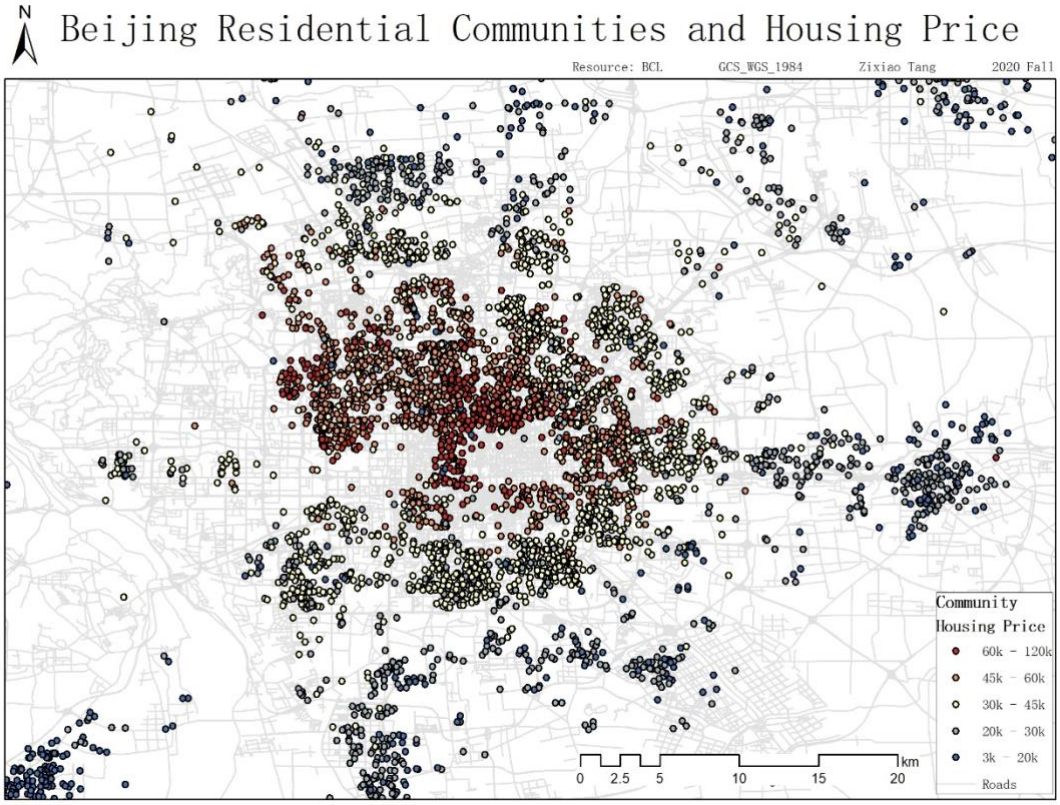
2. What do you understand by metadata? What recording practices did you observe in the repositories you sourced from? How do these data management practices compare to

other data sources that you might have come across (you can explore https://www.opendatanetwork.com/ if you want to compare) and what do you think are the implications for your work?

What do you understand by metadata?- Description Analysis

Through metadata, conclusions can be drawn from the analysis as follows.

First, visualize the communities using 'Communities with Housing Price' data to see the residential space distribution and price intervals.

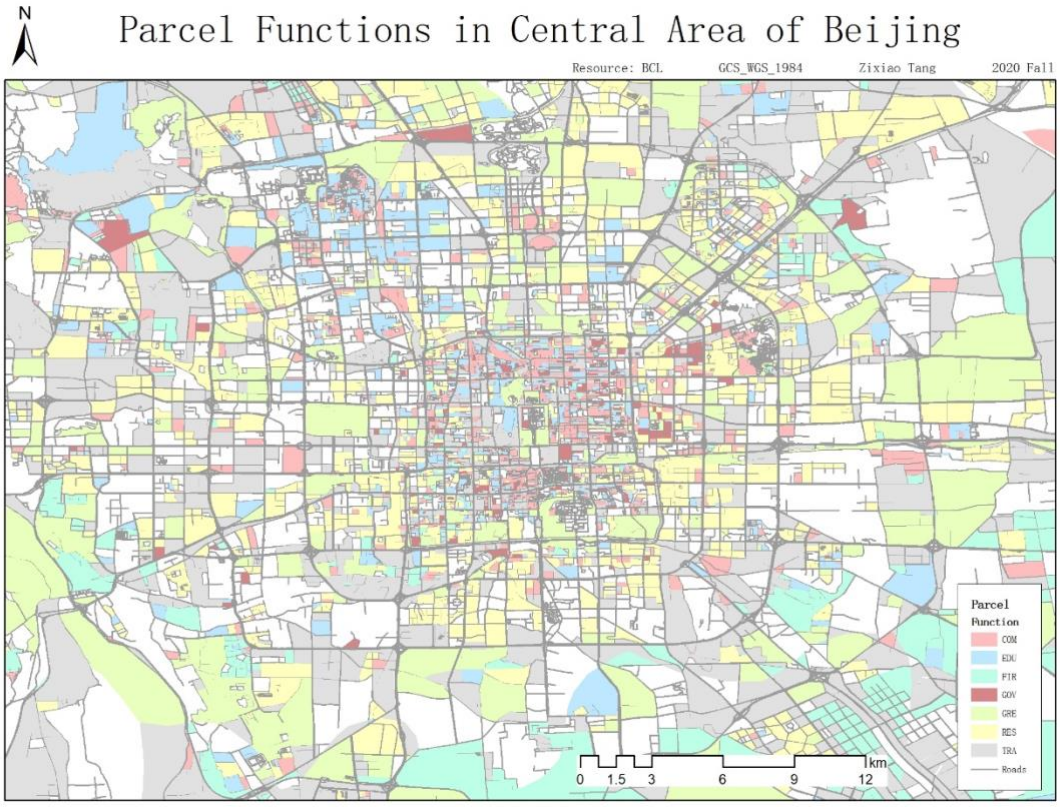


Map 1 Beijing Residential Communities and Housing Price

The communities are mainly concentrated within the Fourth Ring Road of Beijing, which is also the city's central area. Based on this distribution, further analysis will focus on the core area within the Fourth Ring Road. Housing prices generally are

higher in the middle and lower in the margin areas. Besides, the west-north part of the central area has a more expensive housing price than other areas.

Second, visualize the parcels by function types.



Map 2 Parcel Functions in Central Area of Beijing

The ample green space surrounds the city in the margin, while smaller parcels of green space scatter in the inner city. Commercial parcels are more centralized in the inner city and located in the east-north in the core area, where locates the CBD and other main business circles. The education parcels are mainly in the west-north direction. The District of Haidian in the west-north is Beijing's traditional education functional district, which includes Peking University, Tsinghua University, and other nation's top colleges.

Greenspace, commercial, and education parcels are more related to the daily convenience of residential neighborhoods. Thus, the following analysis will focus on these types of parcels.

What recording practices did you observe in the repositories you sourced from?

First, as an essential urban economic data, there is still no official database for Beijing housing price data, so scholars can only obtain data from real estate intermediary websites in a fragmented way through web crawlers. This is only a piecemeal way of capturing the data stored on the site at a given time and in a given region, but the actual house price data is shifting. Therefore, this kind of housing price data is limited in time and space.

Second, the most common storage method of spatial data, such as land use data and but stop data, is stored as a shapefile suitable for ArcGIS and QGIS and stored in the form of layers with attributes.

How do these data management practices compare to other data sources that you might have come across (you can explore https://www.opendatanetwork.com/ if you want to compare) and what do you think are the implications for your work?

At present, some geospatial data that can be seen in the database are stored in shapefile format, some are stored in CSV format, and use geo-codes to mark the space that the data refers to, and some databases provide both formats at the same time.

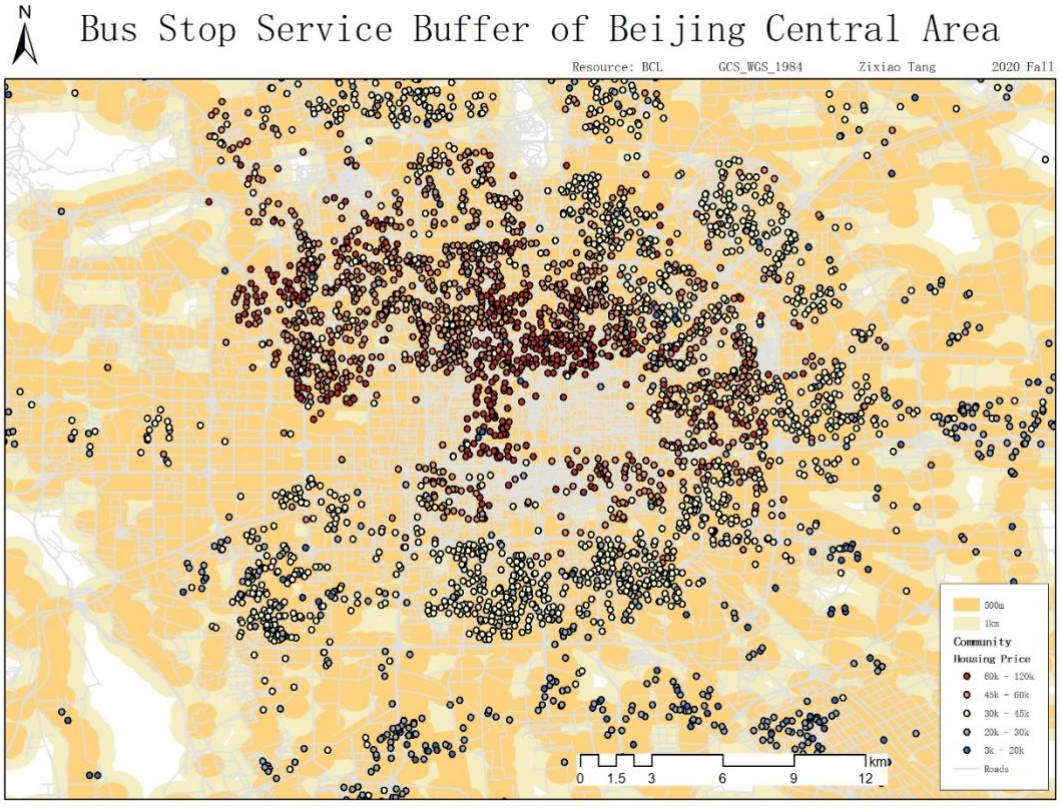
Shapefile is still the most convenient way to store spatial data. Compared with forms with geographic codes, shapefiles can be opened and processed directly in

geographic information system software without the data transformation and projection process of geocoding, which is the most suitable storage method for spatial data.

The implication for my work is that if I want to study geospatial data, it is better to store the data in shapefiles during work. If there are situations where people cannot view the data using GIS software, or if there are other limitations to document delivery, then consider using spreadsheets to store and transfer the data.

3. In not more than 1.5 pages and 5 figures (3 pages and 10 figures for group work), conduct an analysis following the steps covered in the slides from the second in-class workshop. Please note that this page limit is in addition to the previous two questions.

i) Bus Stop Service Supply of Beijing Central Area



The bus stop service covered most of the central urban area and the communities, except for some outliers. The map shows that the coverage of Beijing's public transport system is relatively comprehensive on the scale of 500 meters distance.

However, qualified public transportation coverage does not furthermore differentiate convenience between locations. To investigate more in the transportation convenience of each community, **use the tool "Near" in ArcGIS to calculate the distance to the nearest bus stop of each community point.** Figure shows the statistical result of distance. Most of the communities can get to at least one bus stop within 500 meters.

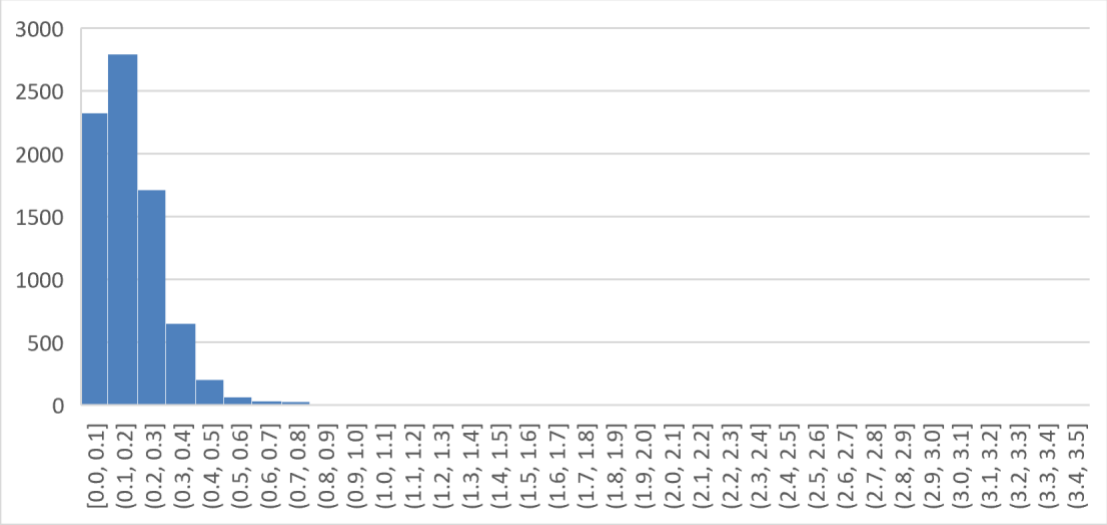
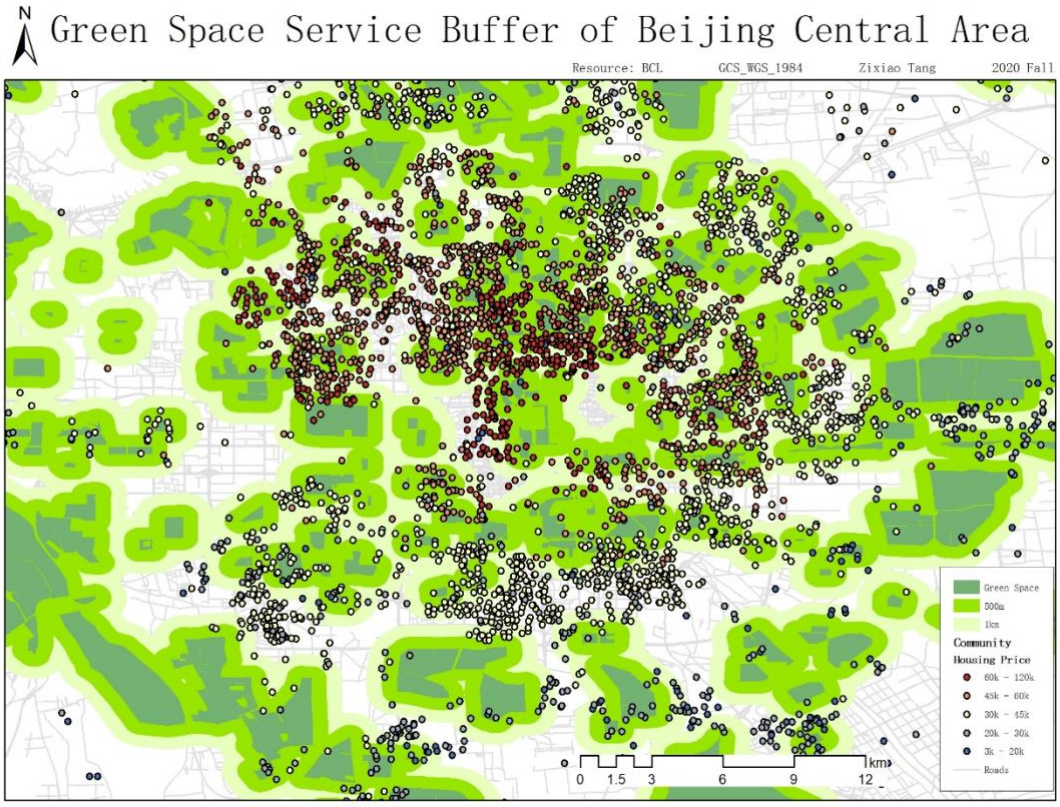


Figure: Nearest Bus Stop Distance (km)

ii) Green Space Service Supply of Beijing Central Area

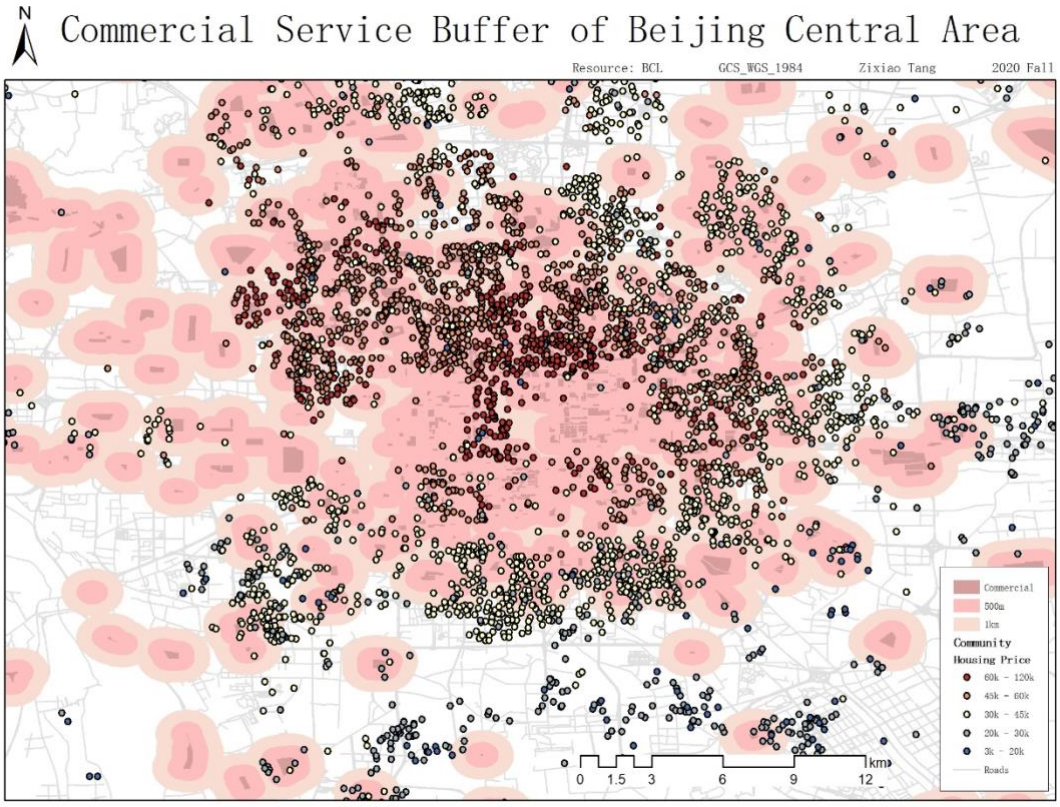


Map 4 Green Space Service Buffer of Beijing Central Area

Greenspace services show significant clusters in the middle north, where the expensive housing also cluster. More areas are out of green space service in the west and south near-suburb outlier, especially the southern belt of the outlier, where also shows a belt distribution of middle-priced housing communities. However, the low-priced communities in the remoter outlier are well-covered by green infrastructure since they are in the suburban area where the built-up space of the city becomes sparse.

To conclude, the green space in Beijing presents the spatial structure of the vacuum circle in the near suburbs, where the middle-priced communities locate.

iii) Commercial Service Supply of Beijing Central Area

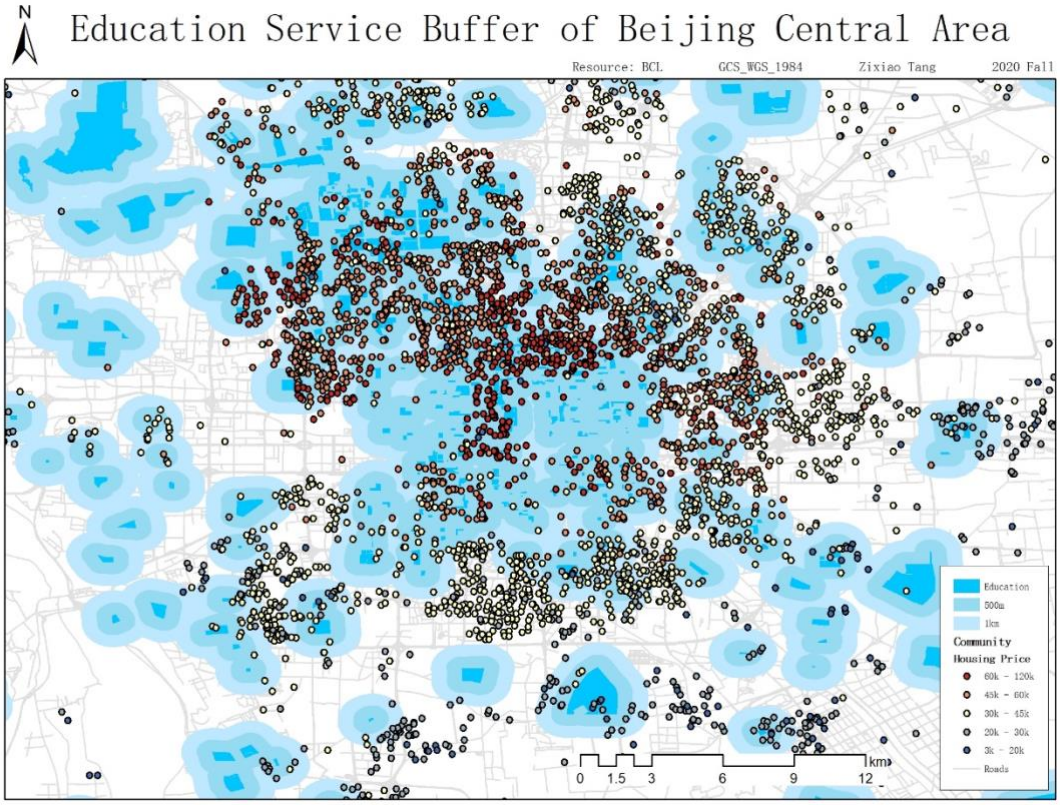


Map 5 Commercial Service Buffer of Beijing Central Area

The commercial service perfectly covered the city's core districts while formed regional centers in the outlier area. The gaps in the south and east suburban area locate the belt of low-priced communities.

In other words, commercial space in Beijing is well-covering residents within the Fourth Ring Road, but residents outside rely on small regional centers nearby, and there is often a lack of services within the convenient living circle.

iv) Education Service Supply of Beijing Central Area



Map 6 Education Service Buffer of Beijing Central Area

The Education service covered the city's core central and northwest areas and formed regional centers in the outliers. The communities out of education service are mainly in the far north, south, and north-east margin areas. The west direction is also a vacuum, but fewer residential communities locate there.

Surprisingly, in the near suburbs due north, due east, and in the south near suburb belt of Beijing, the median housing price neighborhoods show inadequate coverage of educational facilities. This shows that education in Beijing is a scarcer service than green space, commerce, and transport, which is likely to drive up further the price of education highlands in the northwest. Because education is often a rigid need for a family, it is also one of the main drivers for many families to choose where to live.

REFERENCES

[1] Chai, Y.W., Zhang, X. and Sun, D.S. (2015) A Study on Life Circle Planning Based on Space Time Behavioural Analysis: A Case Study of Beijing. Urban Planning Fo- rum, 3, 61-69.

[2] Zhao, Y.Y., Zhang, B. and Zhou, F. (2018) Study on the Spatial Measurement of "15-Minute Community Living Circle" in Beijing Based on POI. The World of Survey and Research, 5, 17-24.

[3] Xiong, W. and Xu, Y.L. (2010) The Urban Settlement Research in the Perspective of Public Facilities: A Case Study of Nanjing. Modern Urban Research, 12, 35-42.

[4] Wu, Q.Q. (2015) The Exploration on the Dynamic Programming of Community in Megacities from the Living Circle Perspective. Shanghai Urban Planning Review, 4, 13-19.

[5] Xiao, J.H., Zhou, D.L. and Hu, J.P. (2018) Measurement And Evaluation Method of Community Life-Cycle Based on Decision Tree Theory: Panyu District of Guangzhou. Planners, 3, 91-96.

[6] Li, Z. , Zheng, J. and Zhang, Y. (2019) Study on the Layout of 15-Minute Community-Life Circle in Third-Tier Cities Based on POI: Baoding City of Hebei Province. Engineering, 11, 592-603. doi: 10.4236/eng.2019.119041.

[7] Long, Y., & Liu, X. J. (2013). Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest. arXiv preprint arXiv:1311.6165.