Title of project

Spatial Analysis of China School District Housing Market and Education
——take Beijing as an example

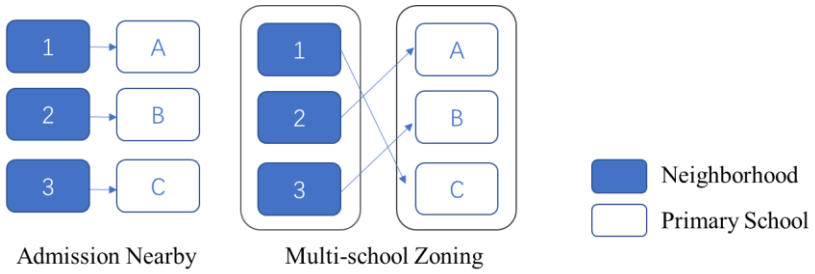
Introduction

Background

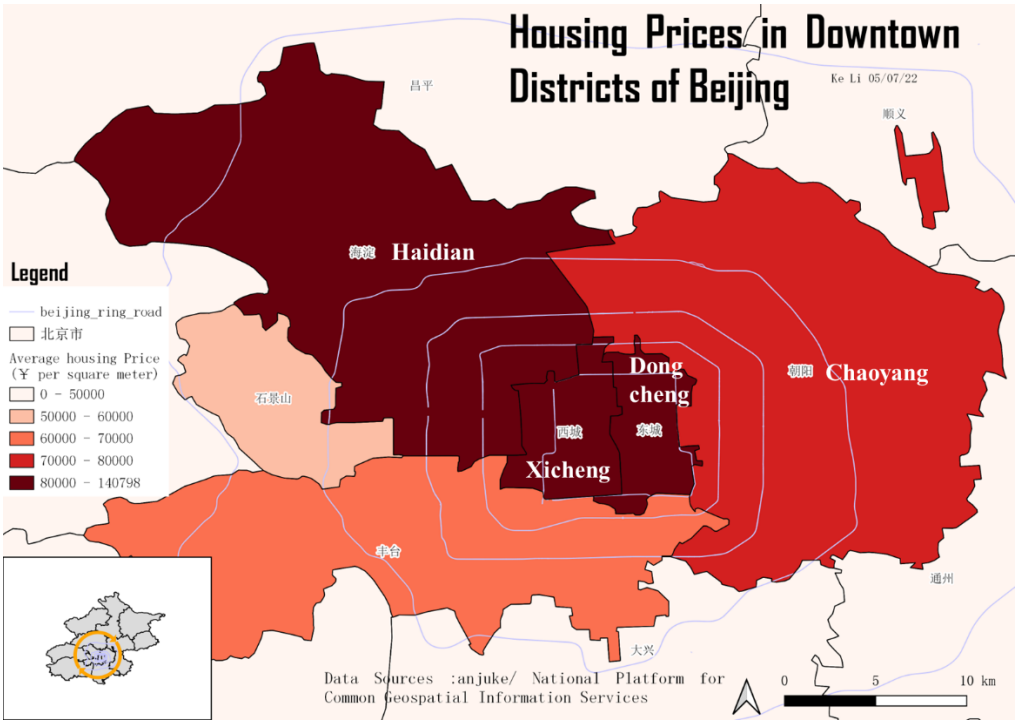
Due to the uneven quality of education in China and the link between school enrollment and home ownership, the term "school district housing" was coined. The only way for parents to send their children to a certain primary school is buying a house in that school district, and parents are eager to do so for the future development of their children. It can be said that the value of education is additionally attached to the value of the housing itself, causing the price of school district housing to soar. Even some old, worn, small housings get fetch an incredible high price.


School district housing is so expensive that China government has been distressed by this matter for a long time. Market itself is fine, but the previous literature has proven that those parents who own the good school district housing are more affluent and educated than who are not. It can be said that school district policy has already led to a segregation to some extent and will increase the inequality in the society for the upcoming years since it influences the education of next generation.

One of the methods that China government has come up with is called “multi-school zoning” which refers to the policy that instead of pointing one neighborhood to a certain primary school they build a large school district (maybe contains dozens of neighborhoods) and children in that school district will be assigned randomly to one of the schools affiliated to the school district.



As the capital of China, Beijing is the pioneer of policy implementation. For the four downtown districts of Beijing——Dongcheng, Xicheng, Chaoyang and Haidian, three of them have already taken the policy, only Chaoyang district, though takes a multi-school zoning at surface, still adopts the single school allocation in most of its areas.



Spatial Questions

Previous literature has proven that the quality of education of the school is a powerful factor of housing price. But I would like to find out which extent it influences the price. Does this influence outweigh other possible factors so that could be visualized spatially?

That is to say, if neighborhood A is in one good school district and its counterpart B is near but not inside the school district, can we tell that in most cases, the average price of A will be higher than that of B regardless of their year of construction, residential structure and property management?

If yes for the above questions, then is the multi-school zoning an effective method to control the housing price?

Significance

This project can be taken as a monitor, if we see a pattern of high premium of housing in school districts, then stricter policy would be in place. Besides, it can check the effectiveness of multi-school zoning policy. The spatial analysis can be used to monitor the market and make future policy plan in the benefit of citizens.

Since the problem of school district housing is a common problem throughout China, same spatial analysis can be done for other cities in China.

Literature review

Previous literature has proven that education quality of the school has effect on the school district. However, no studies have focused on their spatial distribution or aggregation.

For the multi-school zoning, although there is now a lot of literature exploring the disadvantages and advantages of multiple school districts. However, all of them are analyzed from a theoretical perspective or with some individual examples as illustrations, and there is no literature that analyzes them from the perspective of a large amount of actual data.

Data

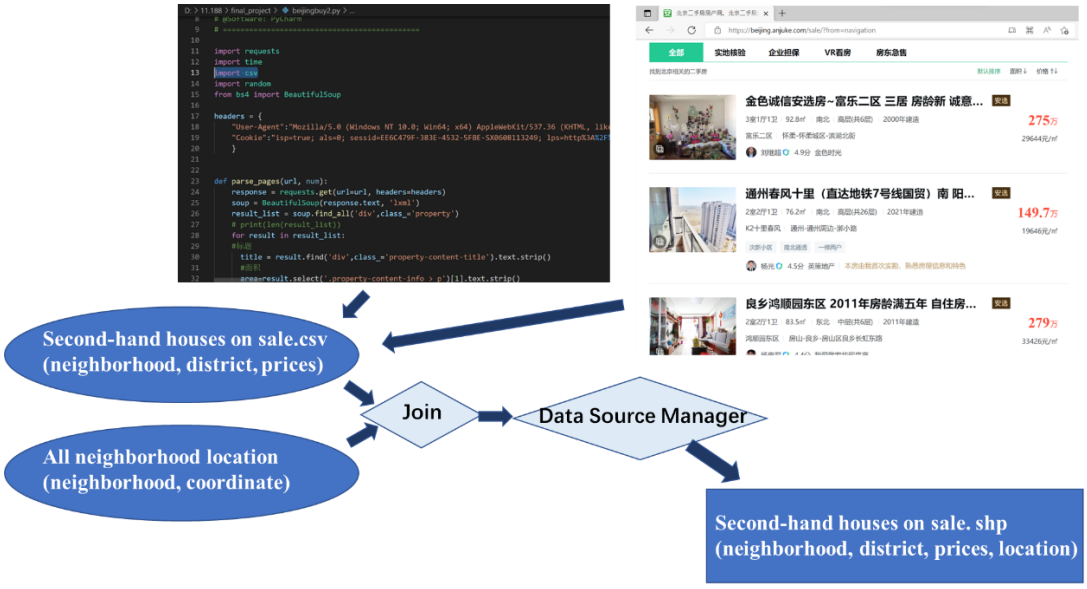
To verify the relation between school education quality and housing price, I need the following data. Also, since I will both look into Chaoyang(single school allocation system) and Xicheng (multi-school zoning) the data will be collected for both of the two districts.

	Data Source	**Explanation of the data sources**	**Data format**
Housing Price			
1. Neighborhood Name	anjuke.com	A famous housing trading website in China	csv
2. Price (per square)	anjuke.com		csv
3. Neighborhood Location	amap.com	Gaode Map (similar to Google map in the U.S.)	csv
4. Which school(district) it belongs to	For Xicheng, from Zhihu.com	An information platform in China	jpg
	For Chaoyang, from Local Education Beau		txt
School Education			
1. School List	bj.bendibao.com	(an information platform used by Beijing locals)	txt
2. Primary School Location	amap.com		csv
3. Evaluation of the schools	A good school nomination list	It was proposed by a Beijing education expert a few years ago and is very well accepted by Beijingers	txt

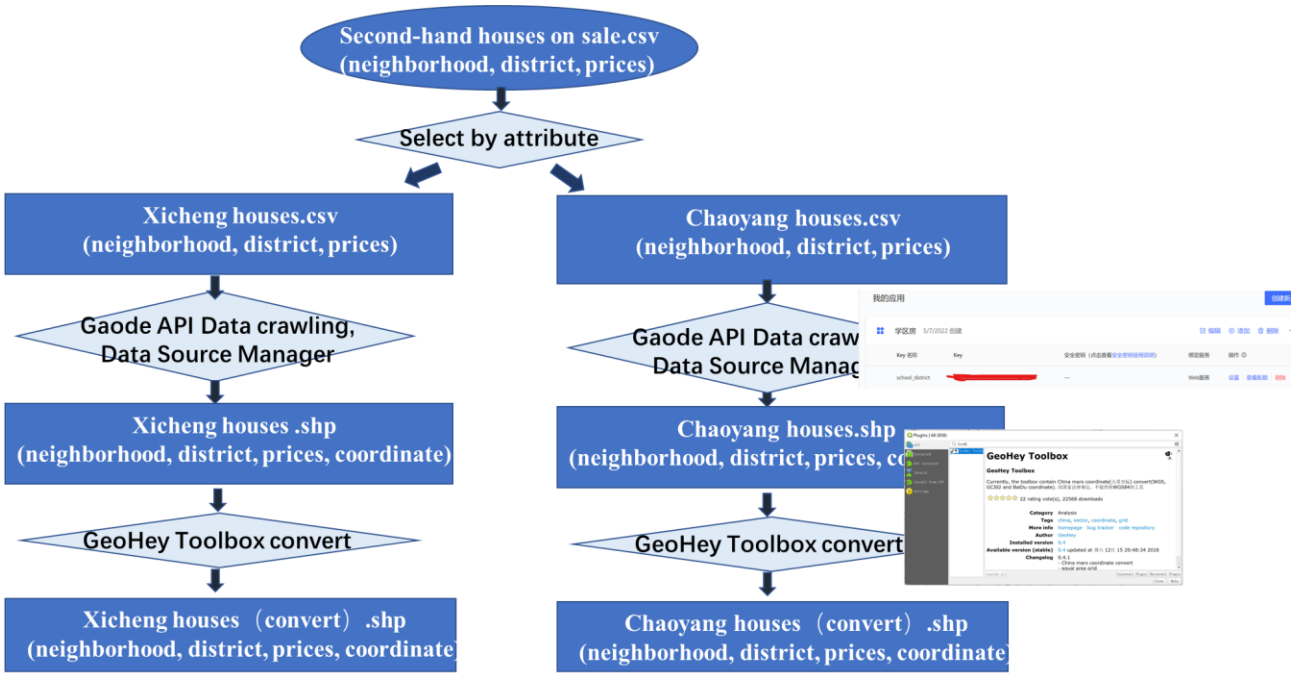
Methodology

For housing price data, I get 1. neighborhood name data and 2. Price from the housing trading website. 3,061 data were obtained by crawling.

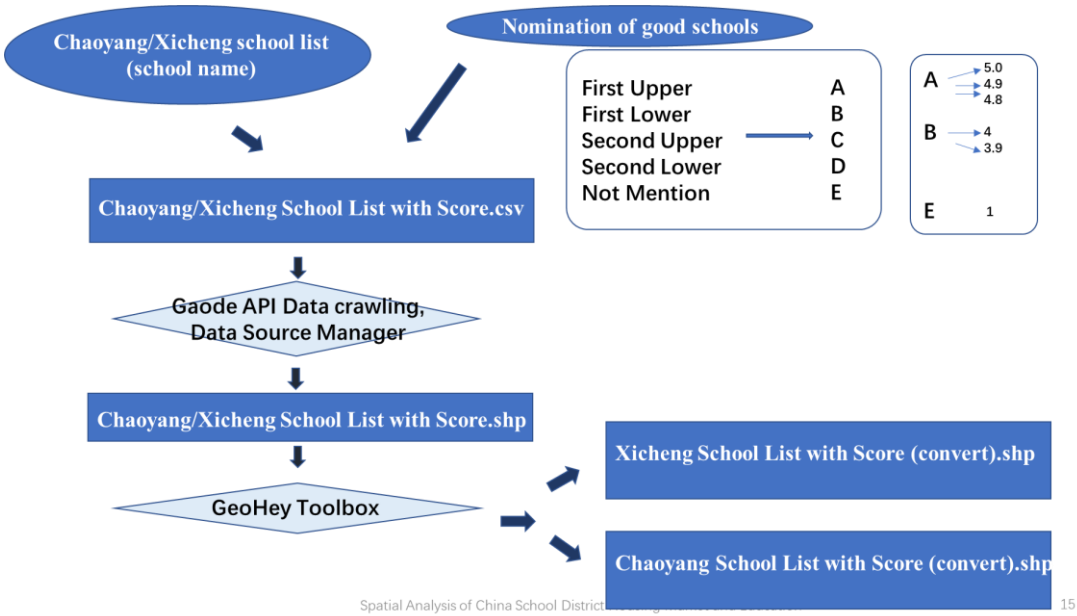
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	标题	面积	户型	朝向	楼层	建造时间	行政区	区域	小区	总价	总价单位	单价	price	
2	10号线石	62.8㎡	2 室 1 厅	1南北	中层(共6层	1995年建	丰台	石榴庄	石榴园北里	353	万	56229元/n	56229	
3	回龙观 新	97.6㎡	2 室 1 厅	1西北	中层(共20	2007年建	昌平	回龙观	新龙城	568	万	58197元/n	58197	
4	近地铁419	72.3㎡	2 室 1 厅	1南北	高层(共6层	1999年建	大兴	西红门	同兴园小区	280	万	38728元/n	38728	
5	南北通透	170.8㎡	3 室 2 厅	2南北	高层(共8层	亚运新新	朝阳	亚运村	亚运新新	1958	万	114638元/	114638	
6	长阳半岛	166.7㎡	3 室 2 厅	2南北	低层(共17	2014年建	房山	长阳	紫云家园6	700	万	41990元/n	41990	
7	大气装修，86㎡		2 室 1 厅	1南	低层(共21	2003年建	丰台	宋家庄	嘉和人家	545	万	63380元/n	63380	
8	和平里安乡	39.4㎡	1 室 1 厅	1南	中层(共12	1995年建	东城	和平里	安外花园	550	万	139453元/	139453	
9	密云区(密	86.9㎡	2 室 2 厅	1南北	中层(共21	2012年建	密云	密云城区	绿地国际	235	万	27040元/n	27040	
10	美景东方2	153.5㎡	3 室 2 厅	2南北	低层(共17	2008年建	朝阳	华威桥	美景东方	1035	万	67445元/n	67445	
11	2室1厅 百	98㎡	2 室 2 厅	1南北	中层(共16	2007年建	朝阳	百子湾	百子湾家	600	万	61225元/n	61225	
12	密云区(密	61㎡	1 室 2 厅	1南北	中层(共6层	2003年建	密云	密云城区	果园西里小	149	万	24423元/n	24423	
13	满五 南北	254.7㎡	4 室 2 厅	2西南	低层(共26	2007年建	朝阳	望京	上京新航	2150	万	84420元/n	84420	
14	上京新航	115.7㎡	2 室 2 厅	2南北	高层(共26	2007年建	朝阳	望京	上京新航	1180	万	102024元/	102024	



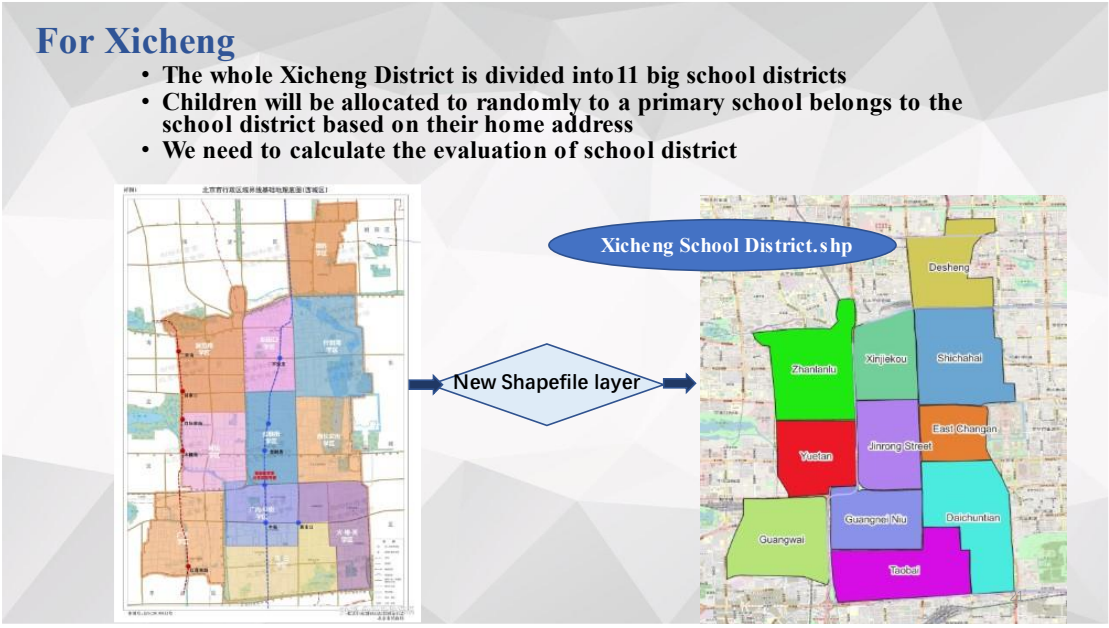
In order to get the data of 3. Location of neighborhoods, I use the Gaode API, and then convert the coordinates I get into the standard WGS 84 coordinates.

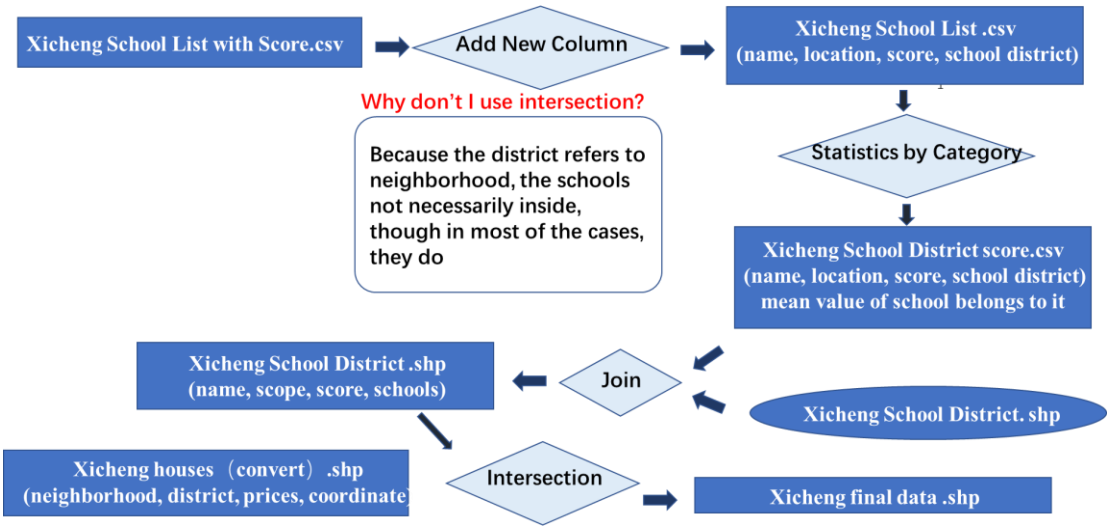


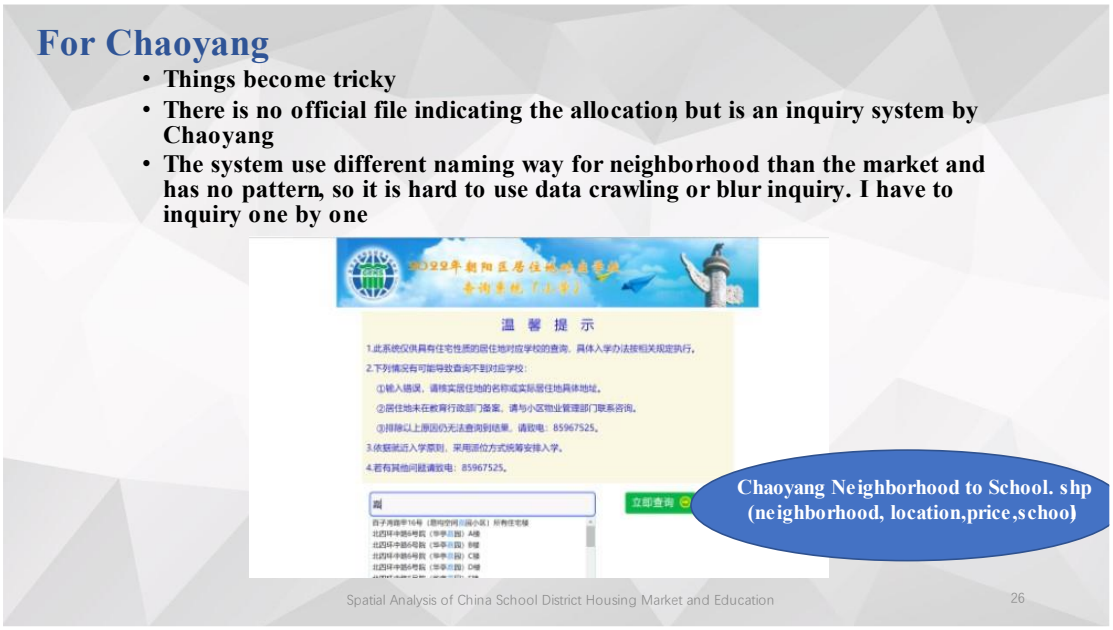
For the data of school education: 1. School List 2. Primary School Location 3. Evaluation of the schools. I search on a local website to get a list with all the primary school in Beijing. I found a list of good school nomination proposed an education expert, to refine parents' preferences for schools, I have made a secondary classification and scoring of the criteria for this nomination.



There is only one data left to be collected, and it is the most important one – 4. the school or school district a certain neighborhood belongs to. And the methods are different between Chaoyang and Xicheng due to their different policy.







Special notes on some steps

There are some interesting tools and steps I'd like to specify. I think some of them will be necessary and helpful if you would like to do an analysis in the context of China.

1. **Data Crawling of the housing price**

   I use Python to crawl all the data needed as to housing marked price from anjuke.com (an Chinese housing trade website).

2. **Gaode Map API data crawling to get the location**

   In the class, we have learnt to get the location of some address used the Google Map. However, suppose you will enter some address in China, it will no longer feasible. A similar tool is Gaode Map and here I use Python to get the data of longitude and latitude.

3. **GeoHey Toolbox**

   The coordinate data we get from the Gaode Map is, unfortunately, not the correct WGS84 data.

It does not mean that it is wrong, it just that due to China's information protection policy requirements, the coordinates given by major map platforms are obtained by encrypting and coding the coordinates of WGS84. So if you directly use the data obtained from Gaode Map for mapping, then there will be a point shift.
Then you need GeoHey Toolbox, which is an QGIS plugin. It can convert your shapefile with GCJ02 or BD09 (which are two mainstream coordinate systems used by Chinese map service platforms) to WGS84 coordinates and display the points in the map directly. It can also realize the revert.

4. **Create a new shapefile**

   We didn't have much chance to dig into it in the class, but it is a fun tool. Sometimes, shapefiles are not available, the only resource you can get is just a map, then you have to create the polygon based the map. I did this work based on the Open Street Map, since the school district will take the road as a division.

   This is a very stress relieving step! But you need to be careful, this method is not feasible for maps with complex divided areas or ambiguous boundaries.

5. **Use the inquiry**

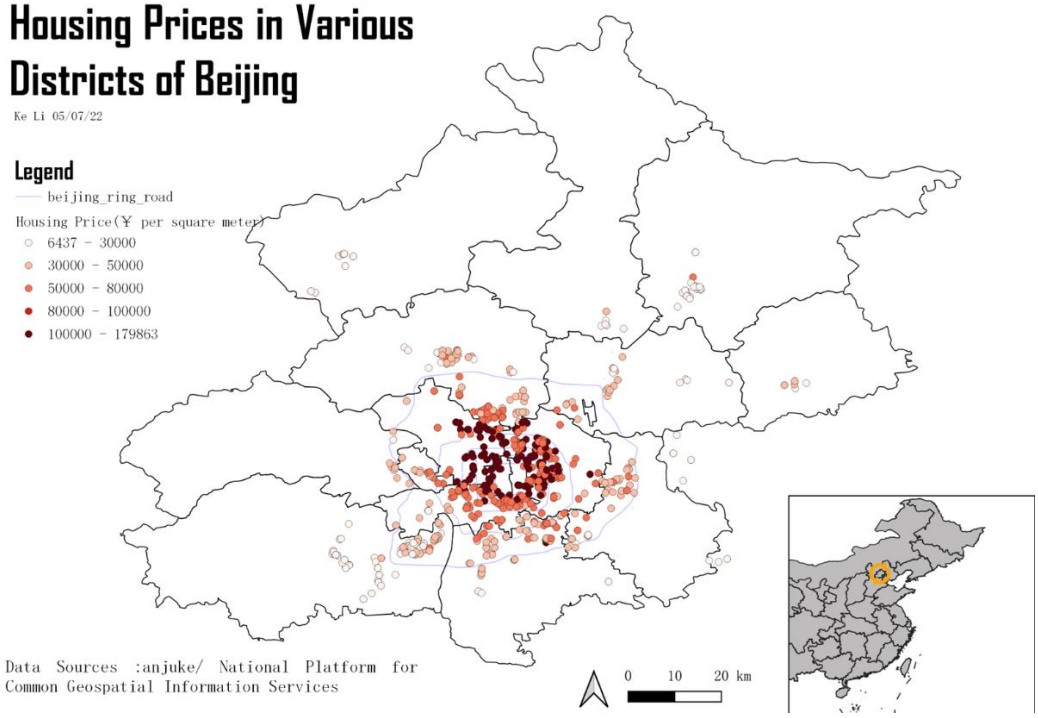
   The lesson I learnt from this step is that sometimes, we do has no method to collect data directly. The only way we can get is manually, but if the data is really needed, then the efforts are worth. But doing that also has a limitation, I inquired for all the data in Chaoyang (near 600 data) and it took me 9 hours to do so. If the amount of the data is 6,000 then I have no idea how I can finish.

Results

As a whole

Housing Prices in Various Districts of Beijing

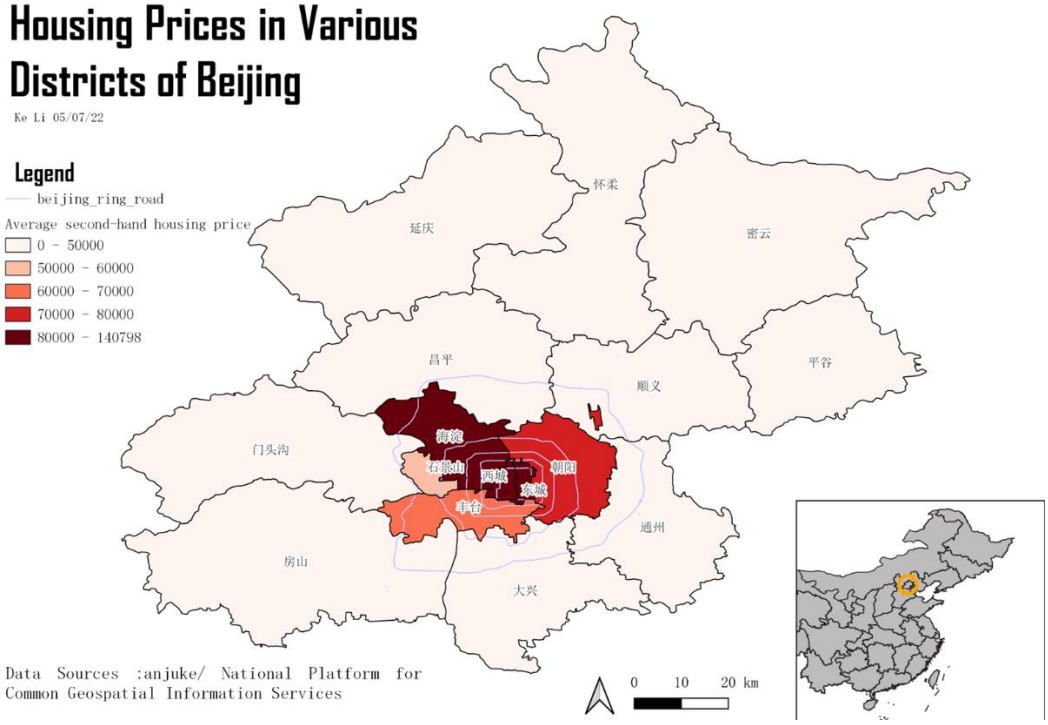
Ke Li 05/07/22



Data Sources :anjuke/ National Platform for Common Geospatial Information Services

Housing Prices in Various Districts of Beijing

Ke Li 05/07/22



Data Sources :anjuke/ National Platform for Common Geospatial Information Services

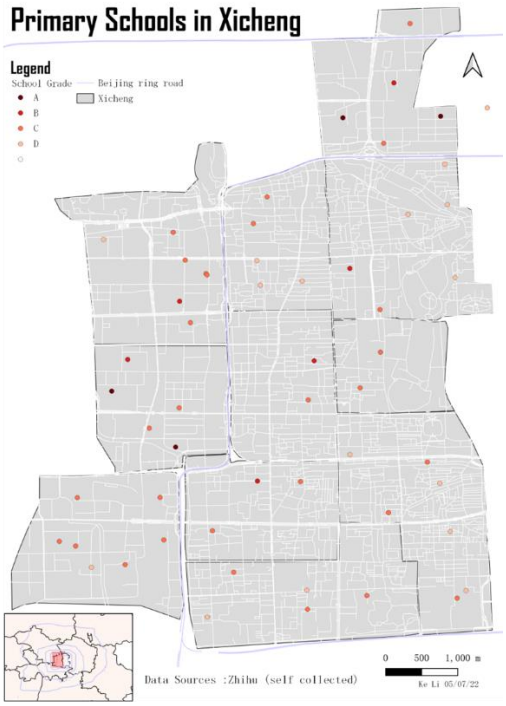
District	count	min	max	range	mean	median	majority
丰台		244	23994	109930	85936	61110.37295	60857	79191
昌平		304	22123	80972	58849	46940.66447	45031	70691
大兴		256	19860	105556	85696	45505.92969	42777.5	29440
朝阳 (Chaoyang)		578	29857	157192	127335	77512.79066	75124.5	63951
房山		481	12189	71275	59086	29076.38877	28366	35517
东城 (Dongcheng)		86	64648	176667	112019	118680.5116	113432	82979
密云		273	9010	63057	54047	23830.22711	24184	26137
西城(Xicheng)		67	74599	179863	105264	140797.6418	147728	111165
海淀(Haidian)		183	46377	179246	132869	93407.24044	86814	71655
通州		210	17750	71366	53616	30955.94286	27318.5	19231
怀柔		48	22195	55506	33311	31441.16667	31150.5	33351
石景山		78	32747	81129	48382	51179.33333	49723	41605
平谷		28	16667	32205	15538	23623.28571	22584	19000
顺义		42	18500	89709	71209	39234.80952	36383	34421
北京周边		28	6437	19968	13531	13708.92857	14033	16630
延庆		49	10914	30157	19243	21461.32653	22434	11895
门头沟		23	22128	49401	27273	37425.26087	37972	44022

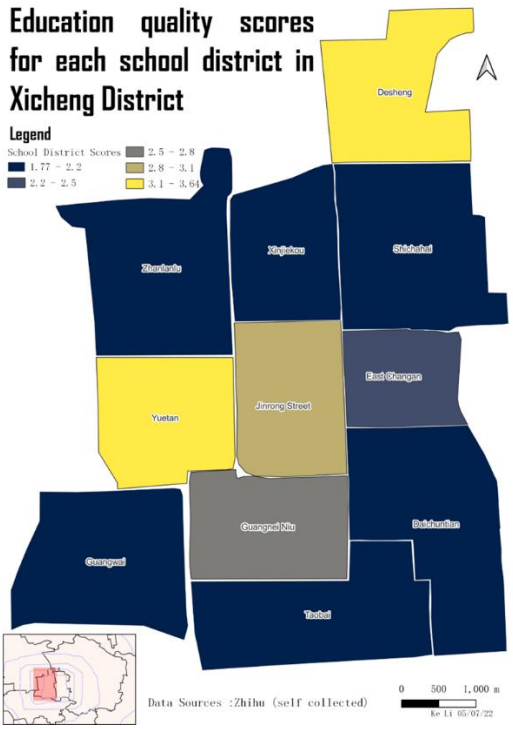
You can see the pattern of housing price distribution in Beijing—the closer to the central, the higher the price are. The four downtown districts still have the most highest price.

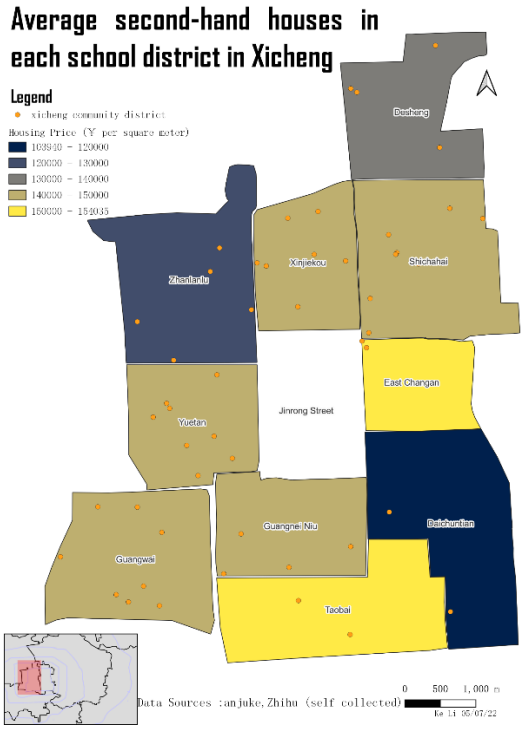
For Xicheng

Grade	Count
A	5
B	8
C	35
D	16

Analyzed field: score
Count: 64
Unique values: 33
NULL (missing) values: 0
Minimum value: 1.1
Maximum value: 5.0
Range: 3.9
Mean value: 2.3891
Median value: 2.1
Standard deviation: 0.9806
Coefficient of Variation: 0.4105
Majority (most frequently occurring value): 2.1
First quartile: 2.05
Third quartile: 2.75
Interquartile Range (IQR): 0.7000

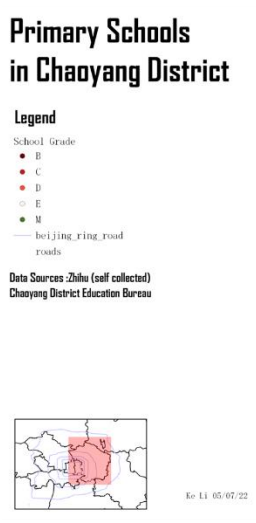


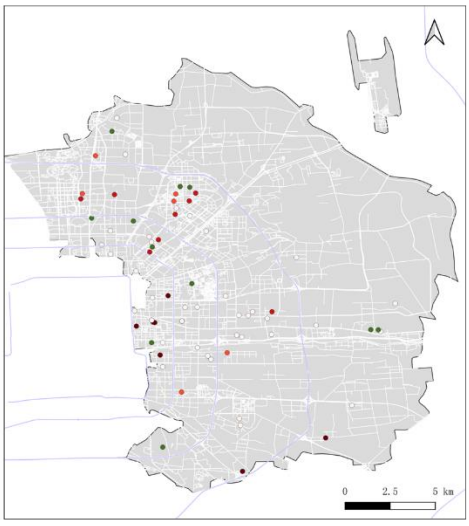




It can be seen directly from the comparison of the two map that the good school districts not necessarily fetch a high price. But we need to be cautious about of this conclusion since the data is limited and maybe not so representative.

For Chaoyang





Grade	Count
B	8
C	10
D	8
E	43
M(middle school)	15

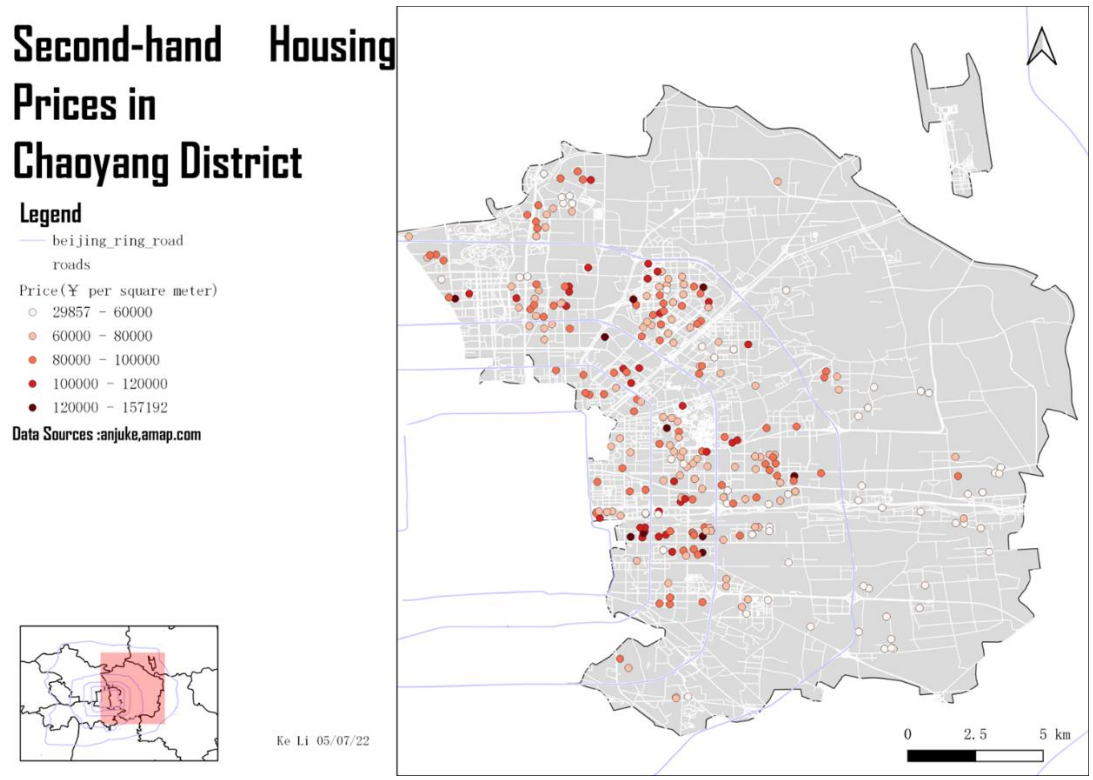
Analyzed field: score
Count: 69
Unique values: 16
NULL (missing) values: 15
Minimum value: 1.0
Maximum value: 4.0
Range: 3.0
Sum: 116.2
Mean value: 1.6841
Median value: 1.0
Standard deviation: 1.0090
Coefficient of Variation: 0.5992
Majority : 1.0
First quartile: 1.0
Third quartile: 2.3
Interquartile Range (IQR): 1.3000

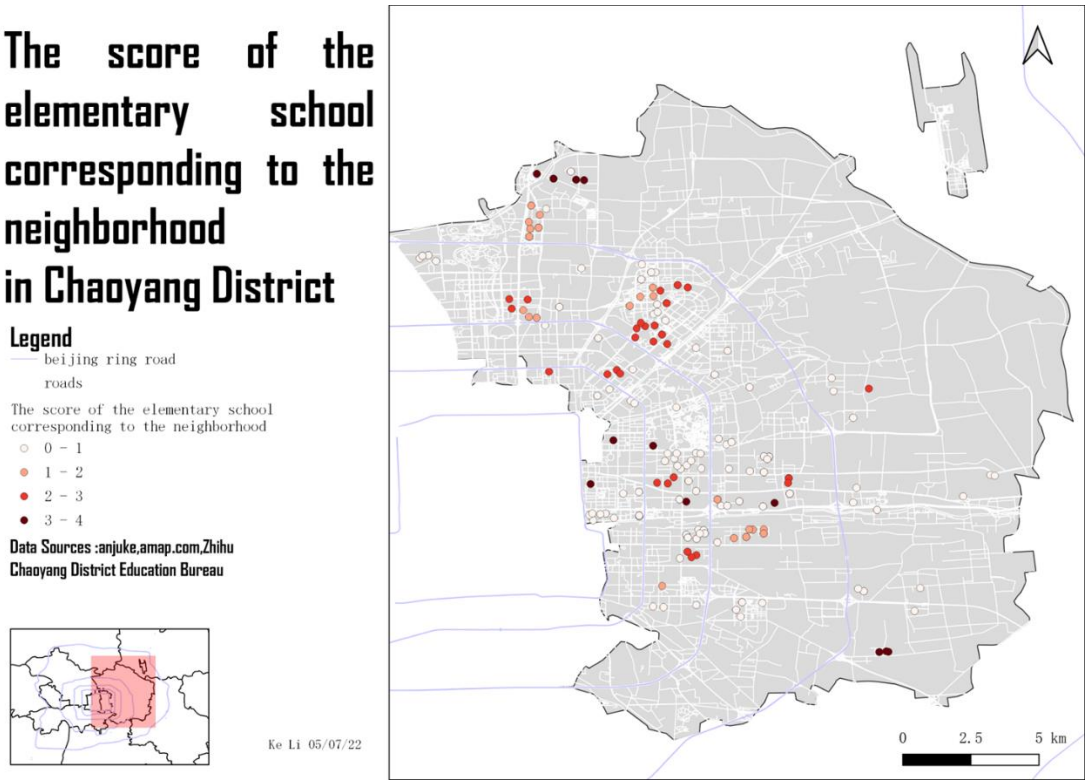
Before look into the relation between housing price and school education in Chaoyang, let we utilize our education data to the most by comparing the education quality of Chaoyang and Xicheng.

attributes	Chaoyang	Xicheng
COUNT	69	64
Unqiue	16	33
NULL	15	0
Minimum value	1	1.1
Maximum value	4	5
Range	3	3.9
Mean	1.6841	2.3891
Median	1	2.1
Standard deviation	1.009	0.9806
Coefficient of Variation:	0.5992	0.4105
Majority	1	2.1
First quartile	1	2.05
Third quartile	2.3	2.75
Interquartile Range (IQR)	1.3	0.7

- **Xicheng has less education resources than Chaoyang(count)**
- **The overall education quality of Xicheng is better than Chaoyang(median and mean)**
- **The educational level of each school in the Xicheng is more concentrated**

Then let's see the correlation part.





Some points have lost in the second map since I omit three categories of neighborhood 1. With no school allocation 2. Allocated to multi-school 3. Allocated to a primary school which is combined with a middle school so get no information of its education quality from the good primary school nomination (it is obvious not mentioned by the list but is unfair to give it an E). But you can still see that there is no concentration of high prices near good schools

Using the Pearson Test, I found that in Chaoyang there is no correlation between the school score and the housing price.

Correlation			
		price	edu_score
price	Correlation coefficient	1	0.043
	Sig.		0.442
	cases	578	328
edu_score	Correlation coefficient	0.043	1
	Sig.	0.442	
	cases	328	328

Conclusion

Though education quality has impacts on housing prices. The impact of school education quality on secondary home prices is not spatially significant both for Xicheng and Chaoyang.

Since this spatial aggregation effect is not even evident in single-school sunrise, the question of whether multi-school zoning helps reduce such aggregation is not valid.

Discussion

Limitations

1. The school ratings used are from a few years ago, and the quality of education at the school may have changed significantly in those years.
2. Due to education reform, many schools have disappeared or merged, so there are missing data and the ratings are not accurate enough.
3. The data comes from second-hand home transaction website ,and is the marked price rather than the transaction price, which does not necessarily reflect the value of the home accurately. And Those houses on sale are unevenly distributed, resulting in data concentration in some areas and lack of data in others.
4. The final correlation test performed should be preceded by a statistical test of the normality of the two variables.

Future Plans

1. Finding more accurate evaluation methods for primary schools.
2. I also crawling down the rent price from anjuke and would like to see the impacts of education quality on rent price
3. I would like to see, besides admission, would proximity to schools has influence on housing price. And whether this impact, if any, will vary depending on the quality of education at the school.

References:

[1]卢为民 & 张琳薇.(2015).学区房问题的根源与破解路径探析. 教育发展研究(Z2),13-17. doi:10.14121/j.cnki.1008-3855.2015.z2.003.
[2]张俊友.(2016).“就近入学”的局限及“大学区制”探索. 中国教育学刊(02),32-36.
[3]胡姗.(2017).“多校划片”政策实施存在的问题及原因探究. 农村经济与科技(16),204-205.
[4] 胡巧玲.(2019).“多校划片”政策对学区房价格的影响分析——以南京名校学区为例. 河北企业(07),68-69.
[5]Chaoyang Education Bureau 2022年朝阳区住宅区对应入学服务片查询（小学）(bjchyedu.cn)
[6]anjuke.com(housing trade website)安居客买卖房网站北京二手房房产网，北京二手房出售，北京买房购房交易信息 - 58 安居客 (anjuke.com)
[7]Amap.com 高德地图 API
[8]CSDN
[9]北京西城学区划分，街道地图+学区划分 - 知乎 (zhihu.com)
[10] Beijing City Lab https://www.beijingcitylab.com/