UCL DEPARTMENT OF GEOGRAPHY

YEAR 2018-19

EXAM CANDIDATE ID:	YBCG6
MODULE CODE:	GEOG0149
MODULE NAME:	Housing and Urban Analytics
COURSE PAPER TITLE:	What factors can explain the uneven geography of house prices in London
WORD COUNT:	2400

What factors can explain the uneven geography of house prices in London

Introduction:

House prices have always been a classic and hot issue in geospatial analysis. For homebuyers and homeowners, a clear concept in the price level of local houses is necessary to help them make a shrewd judgment on the transaction amount before reaching a deal; for investors and property developers, a good insight into the house prices trend can help them maximize the profits; for the government, it is critical to clarify the specific factors behind the current distribution pattern of house prices to develop targeted policies on housing market. In other words, by exploring the distribution pattern of housing prices, potential geospatial information can be found to back up the policy. Recently, the general housing price of London was slowly declining (Hyun and Milcheva, 2019), but once the geographical scale is specified into a certain ward, there was no dataset valuable for public to directly refer to. Besides, due to the uncertainty of housing price changing, investors and governments are unlikely to make housing decisions correctly without a clear prediction model of house prices. Therefore, finding the variables affecting housing prices is quite important. Some researchers proved that the house price is positively proportional to the health of the local residents through cancer cluster data(Davis, 2004); The analysis carried out in Hong Kong through the Hedonic price model also indicated that transport can also be regarded as an important impact on the distribution pattern of house prices (So, Tse and Ganesan, 1997). Another group of researchers found the age structure can impact the housing market as well (Green and Hendershott, 1996). Other potential factors such as crime rate (Tita, Petras and Greenbaum, 2006), unemployment (Evans and McCormick, 1994) are also found relative to the house price distribution. Though there are so many factors behind the distribution pattern of house prices, they have different correlations with house prices. Only with the best combination of factors, the house price distribution can be well-interpreted.

To deal with the problems above, this essay presents the distribution pattern of house prices in London behind which the education level, life expectancy and public transport are the three most significant factors.

Data and Methodology

Housing Transaction Records (https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads) is the main dataset in this analysis, which contains a total of 78,100 housing transaction records in London; By using Ordnance S urvey's Postcode Transfer (https://www.ordnancesurvey.co.uk/opendatadownload/products.html) this essay can convert the postcodes of more than 90% transaction records into geographic coordinates; London Boundary Shapefile (https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london) is an important dataset f or presenting the median house price for each ward; Census statistics (https://d ata.london.gov.uk/dataset/ward-profiles-and-atlas) is used for the further explorati on on the factors that affect the house prices. It provides an overview on thes e districts in terms to the population, diversity, households and other aspects.

In term to shapefile, the spatial scale is determined as ward rather than borough or LOSA. Compared to borough, GWR model perform a better spatial analysis in ward scale, because only in ward scale the number of neighbors of each ward object are large enough to construct a complete spatial weights matrix. Compared to LOSA object, each ward object contains more transection records which excludes the contingency in the analysis. The small wards placed in the city of London are analyzed as a whole rather than in individual because of the same reason.

The workflow is shown below (Fig, 1). Based on the content of the thematic map, a rough spatial distribution pattern can be confirmed; Hotspot is a commonly used spatial autocorrelation methods which can amplify the differences or similarities among adjacent wards to make the distribution pattern much clear; After a clear understanding of the distribution of housing prices, key factors are identified by correlation analysis. Geographically weighted regression (GWR) is a regression model that contains spatial attributes, of which the impact bandwidth is determined by k-nearest neighbors in this essay. It is used to explore the spatial distribution patterns of quantified impact of identified factors on the house prices. In the end, the GWR outputs of factors are interpreted based on proven distribution pattern of house price, correlations and relative reading materials.

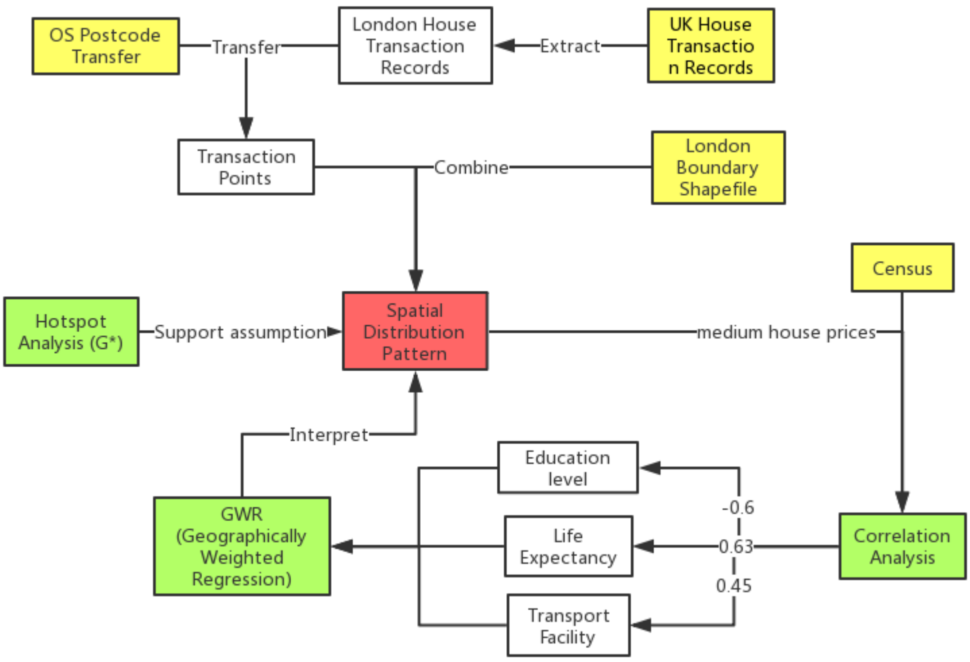

Fig 1 the workflow of the analysis

Results and Analysis

Distribution pattern

About 40% wards falling into the lowest price range of 230 to 445 thousand can be identified roughly. In other words, the overall level of house prices in London is around 445 thousand. However, as such low-cost wards are scattered throughout the London, except for the large-scale cluster in the east can be clearly defined as a low-cost area, other low-cost wards seem not closely clustered. Conversely, the high-cost wards only occupy one small area in Westminster borough. The information gained from the thematic map that the low-cost area in the east London and the high-cost area in the Westminster borough are both highly consistent with the actual situation that east London is relatively poorer while Westminster is one of the richest boroughs in London.

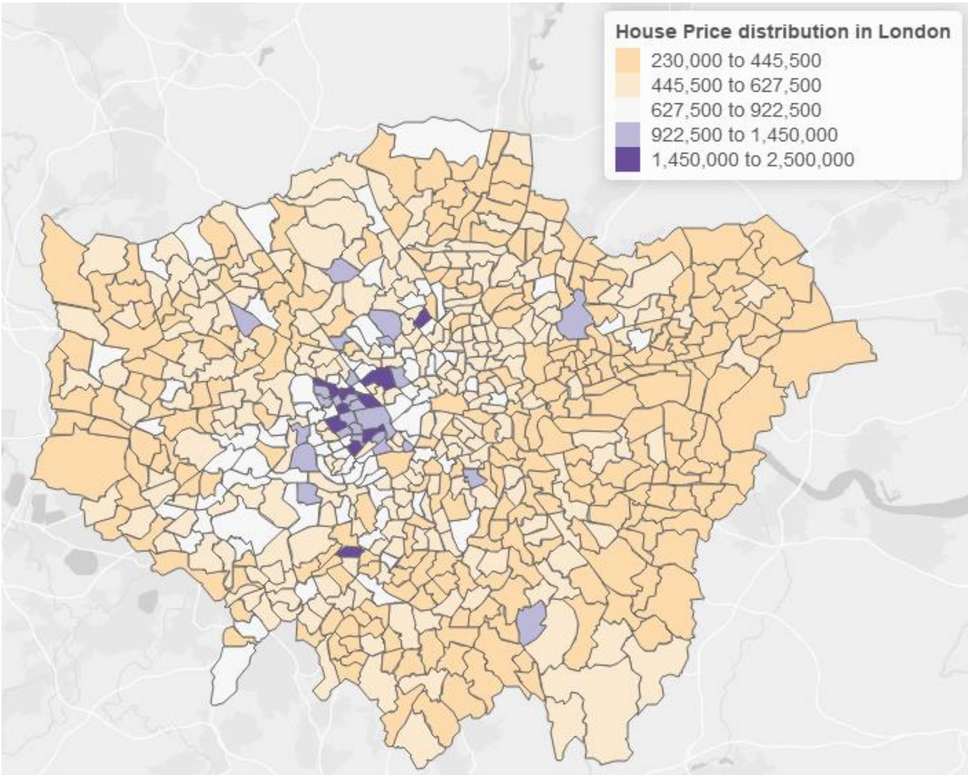

Fig 2 distribution pattern of house prices in London

Hotspot analysis (G*)

The conclusion above made subjectively seems not persuasive enough. Therefore, g* value is used to confirmed it. Its output (Fig, 3) shows that the locations of the clusters basically match to the assumption above. The high-value cluster on the left side covers most wards in Westminster and Kensington and Chelsea boroughs and several wards in Camden and Hammersmith borough, while the low-valued cluster on the other side covers most wards in Barking borough and part of wards in Baxley borough. Additionally, the figure shows that from the cluster center to its edge, its feature on high-cost or low-cost house prices gradually fades.

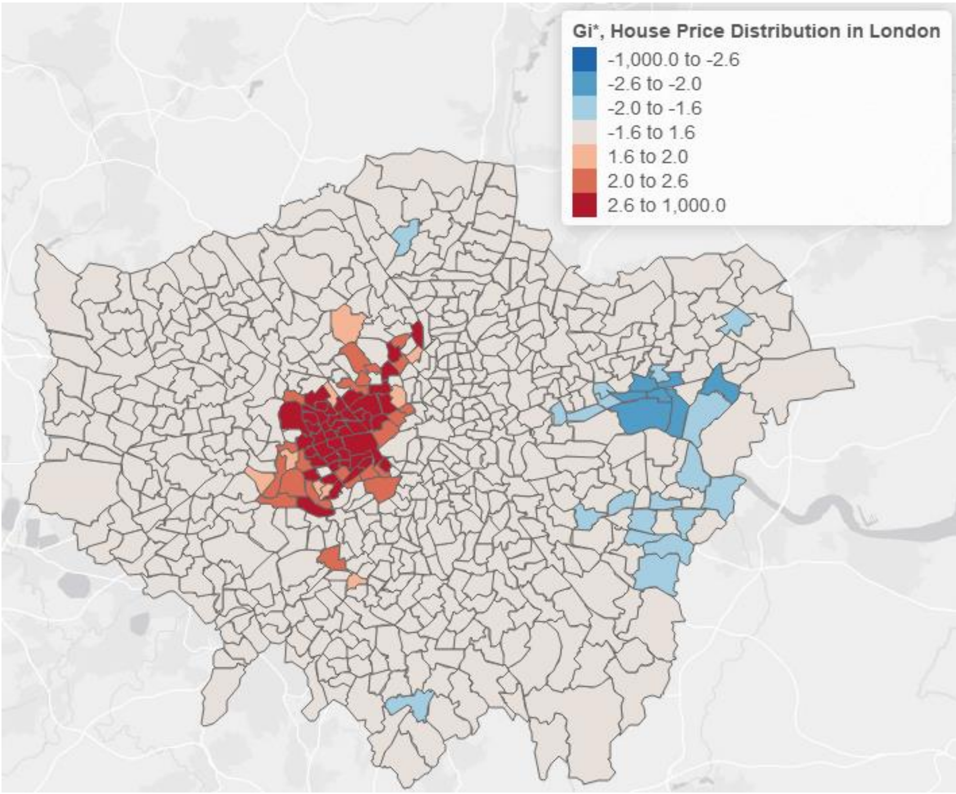

Fig 3 G* of house prices in London

Predictors identification

Through correlation analysis, a total of 14 potential factors are identified as correlated and sorted in descending order of the correlations (List 1). Their absolute correlations with the median house prices – "x" generated in this essay are all larger than 0.4. To avoid the inconvenience brought by the long variable name, these factors will be discussed by using Fac[ranking] in the rest part of this essay.

```
 [1] "x"
 [3] "X..dwellings.in.council.tax.bands.F..G.or.H...2015"
 [5] "X..with.Level.4.qualifications.and.above...2011"
 [7] "Average.Public.Transport.Accessibility.score...2014"
 [9] "X..of.households.with.no.adults.in.employment.with.dependent.children...2011"
[11] "General.Fertility.Rate...2013"
[13] "X..All.Children.aged.0.15...2015"
[15] "X..with.no.qualifications...2011"


 [2] "Median.House.Price..ï......20"
 [4] "Median.Household.income.estimate..2012.13."
 [6] "Male.life.expectancy..2009.13"
 [8] "Female.life.expectancy..2009.13"
[10] "Claimant.rate.of.key.out.of.work.benefits..working.age.client.group...2014."
[12] "Children.aged.0.15...2015"
[14] "X..dwellings.in.council.tax.bands.C..D.or.E...2015"
```

List 1 potential factors

To identify the valuable factors, some obviously inappropriate variables need to be discarded. The Fac[2] extracted from Census dataset also presents median house prices. Though it proves the reliability of "x" generated in this essay with a high correlation at 0.93, it is of no use in factor identification. For those with respect to tax band – Fac[3] and Fac[14], the causality between them and the house price is totally opposite to factor identification task in this section. It is the house prices that directly lead to the houses fall into the exact tax bands. In other words, the house price is “the factor” for the variables rather than the result. Therefore, they should not be regarded as factors either. Another variable, Fac[11] – the fertility rate does not directly affect housing prices. Even if it can impact the age structure, the impact on housing prices is lagging behind. Conversely, several studies have proved the decline in fertility rate caused by high housing prices (Yi and Zhang, 2010). Therefore, Fac[11] and Fac[12] and Fac[13] which are highly related with it are also excluded.

Other those highly correlated variables need to be dropped as well. According to the figure 4, a feasible factor identification plan can be formulated among the rest variables whose correlations with each other are in green boxes which refers to the absolute correlation below 0.4. For example, the correlations of Fac[5] with Fac[6] (0.34) and Fac[8] (0.3) are not so high, but as Fac[6] and Fac[8] are too highly correlated (0.71) to exist simultaneously as individual factors, only one of them can be preserved. Then, to reduce the residual of regression model in linear fitting, factors that are weakly correlated with each other should be involved as much as possible. The only combination meeting both requirements above contains Fac[6] – Male’s life expectancy, Fac[7] –Transport and population accessibility and Fac[15] – population ratio of no qualifications.

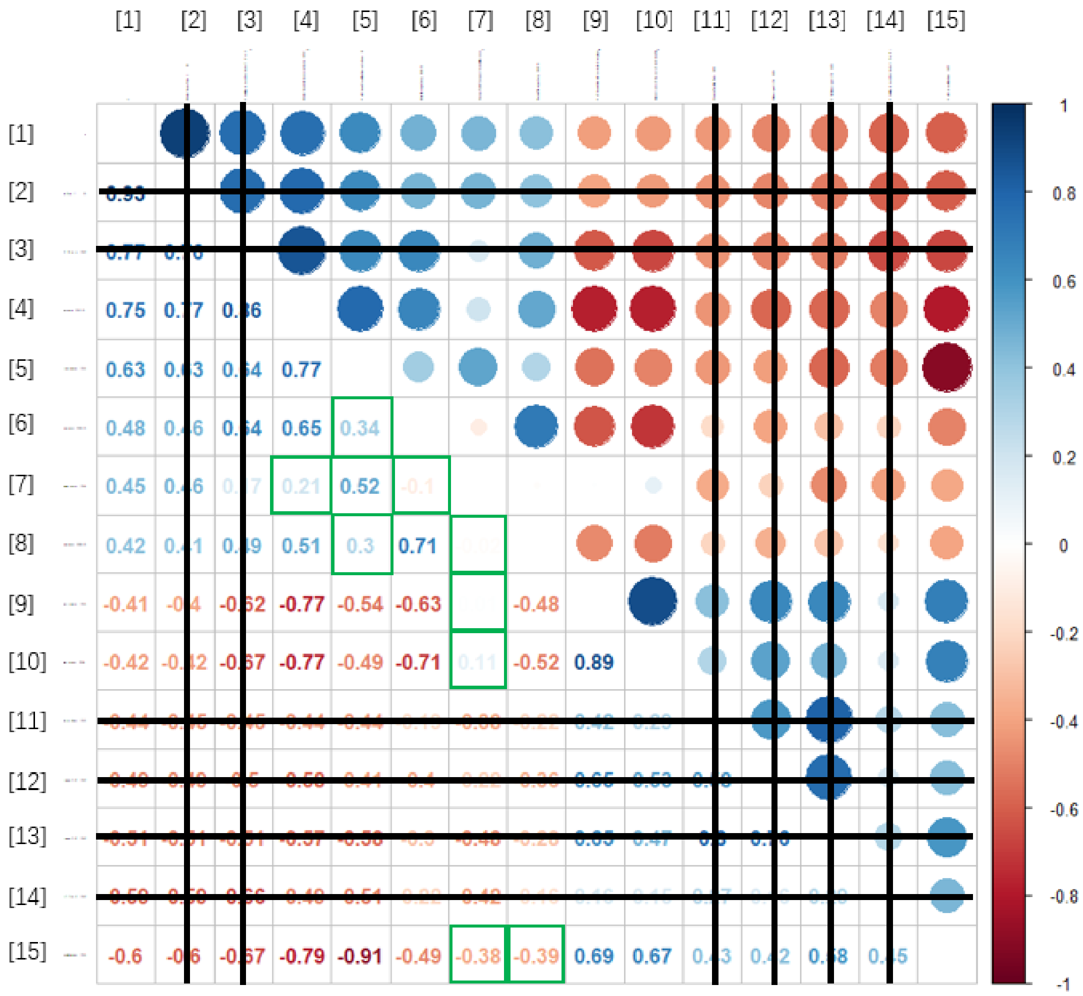

Fig 4 the correlation among potential factors

Geographically weighted regression model (GWR)

The global output of GWR model constructed by factors obtained above reveals the quantified effect of three factors on house prices in London. For example, as the the male's life expectancy increase by one year, the local house prices will rise up by maximum – 91987 pounds or drop down by minimum – 13998.6. Similarly, the quantized ranges of the other two factors can also be obtained. However, the information brought by the global GWR coefficients are limited. To further explain their roles in the house price distribution pattern, more detailed information derived from the local GWR coefficients need to be got.

```
> gwr.model
Call:
gwr(formula = Cor_make$x ~ Cor_make$Male.life.expectancy..2009.13 +
    Cor_make$Average.Public.Transport.Accessibility.score...2014 +
    Cor_make$X..with.no.qualifications...2011, data = Cor_make,
    coords = cbind(Cor_make$X1, Cor_make$X2), adapt = GWRbandwidth,
    hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.02454243 (about 15 of 625 data points)
Summary of GWR coefficient estimates at data points:
                                                                Min.   1st Qu.     Median   3rd Qu.      Max.   Global
X.Intercept.                                              -7126861.5 -2470776.9 -1115946.4 -470054.2 1746404.5 -2674587
Cor_make.Male.life.expectancy..2009.13                      -13998.6    14843.4    21391.0   37279.2   91987.0    39133
Cor_make.Average.Public.Transport.Accessibility.score...2014 -40773.5     8578.7    27195.3   55915.7  111536.0    71018
Cor_make.X..with.no.qualifications...2011                   -29718.4   -14252.5   -11614.9   -9758.9    6740.3   -10506
Number of data points: 625
Effective number of parameters (residual: 2traceS - traceS'S): 101.9122
Effective degrees of freedom (residual: 2traceS - traceS'S): 523.0878
Sigma (residual: 2traceS - traceS'S): 149176.5
Effective number of parameters (model: traceS): 70.58516
Effective degrees of freedom (model: traceS): 554.4148
Sigma (model: traceS): 144900.7
Sigma (ML): 136473.3
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 16715.51
AIC (GWR p. 96, eq. 4.22): 16624.11
Residual sum of squares: 1.164061e+13
Quasi-global R2: 0.7052632
```

Fig 5 GWR output

life expectancy

As house prices is positively correlated to life expectancy, it can be considered that the life expectancy in the high-cost area is much higher compared to that in the low-cost area. Besides, life expectancy is also an indication to the living condition and physical and mental health of people (Lubitz *et al.*, 2003). Therefore, it can be seen that in the high-cost area, which is also called the wealthy area, the safety, greening, medical institution services and other aspects of living standard inside are at the highest level.

When everyone wants to possess limited resources with high quality, the most prominent one will become the target of people's struggle, even if its transaction value far exceeds its actual value (Himmelberg, Mayer and Sinai, 2005). According to the figure 6, the life expectancy has a strongest impact on the high-cost area as expected. The rich are willing to invest several times more money for these housing resources with the best quality. Through high housing prices, they monopolize the best living resources and then form a so-called wealthy neighborhood (Fisher, 2008). Therefore, the essence of the output that high-cost area is affected most by the life expectancy is actually the competition among the people for the best living qualities. The low-cost area is hardly affected by the life expectancy, because the relatively poor living quality here are far less attractive than that of other regions.

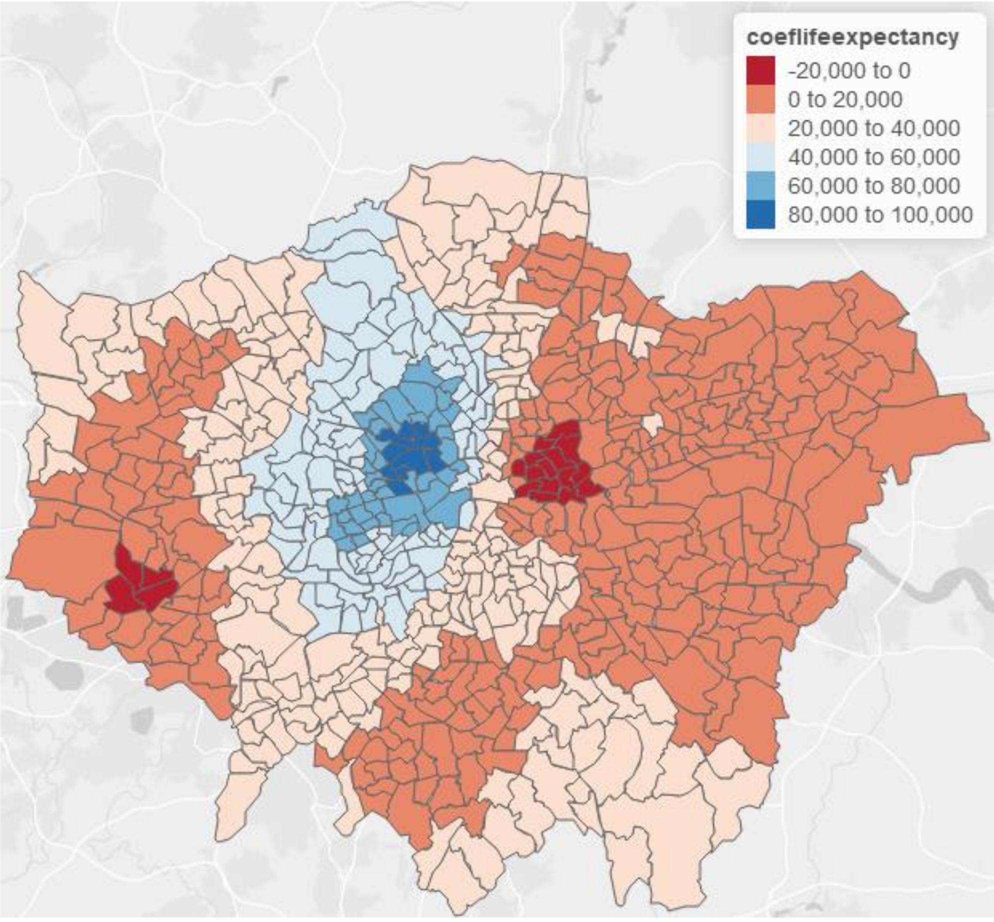

Fig 6 GWR coefficient of male's life expectancy

public transport accessibility score

Public transport accessibility score (or level) is a common method for tfl to compute local transport access which hinges on the distance from any point to the nearest public transport stop, and service frequency at those stops (Currie, 2010). Figure 7 indicates the public transport has a maximum impact mainly on the west London border and the west of the high-cost area while has a minimum impact mainly on the central London and the west of the low-cost area. There are two reason for the output. The first reason is the public transport services itself. According to the positive correlation between public transport accessibility score and house prices, it can be determined that the public transportation in the high-cost area is more convenient than that in the low-cost area. Therefore, part of the vicinity of the high-cost area are highly affected because of the convenient public transport. The second reason is the dependence of residence on the public transport. As almost all people living in the wealthy area have more convenient private transport, the public transport service has little use to them. So, part of wealthy area even falls into the low-impact area. People living further away from the city center cared more about the public transport accessibility nearby. Therefore, the impact from public transport scores on housing prices of part London suburbs are greater than that of central London which possesses more convenient public transport facilities.

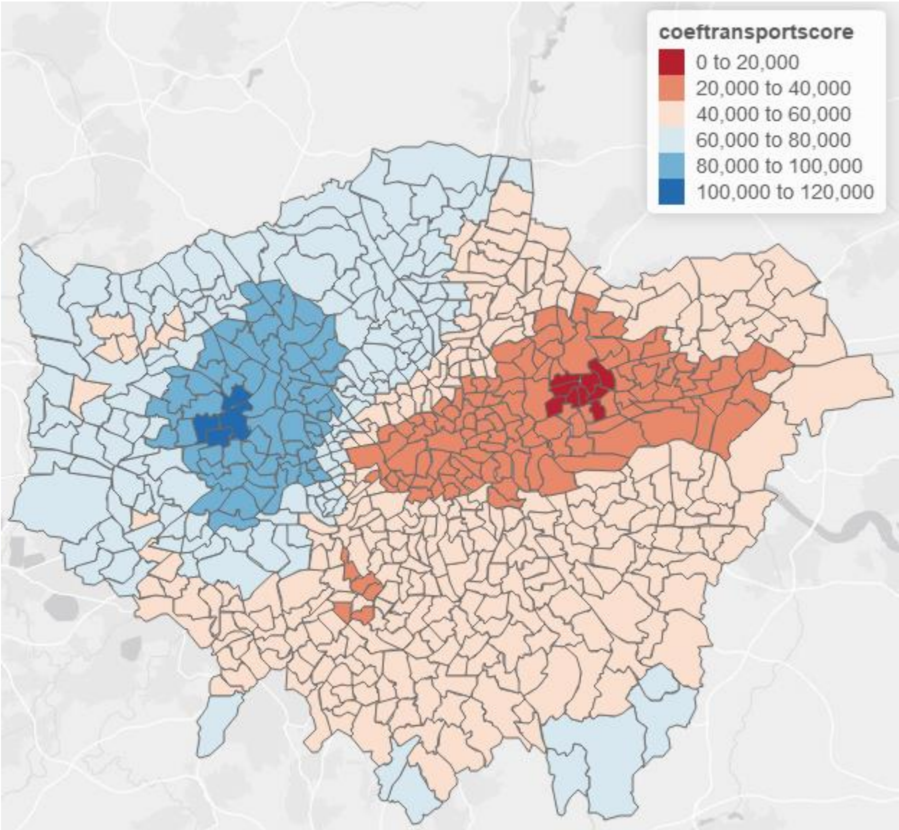

Fig 7 coefficient of transport accessibility score

education level

In the traditional concept, the poor level of education will produce more unschooled people, whose income are usually lower compared with those well-educated (Griliches and Mason, 1972). In that case, they cannot afford house prices as much as that of well-educated people. Therefore, a cluster of unschooled people always means the lower house prices in the area. Figure 8 indicates that house prices in almost all regions will decline, as there is an increase in the population ratio of unschooled people. Because what they can afford determines local house prices when the cluster is formed (Sani, 2013). However, only that of the Westminster borough in the wealthy area keep constant. Because the ratio of unschooled people and the house prices are negatively correlated, the former is actually at a very low level in the Westminster borough. The assumption based on the fact is that before the ratio of unschooled people exceed a threshold and form the cluster, it has little impact on the house prices.

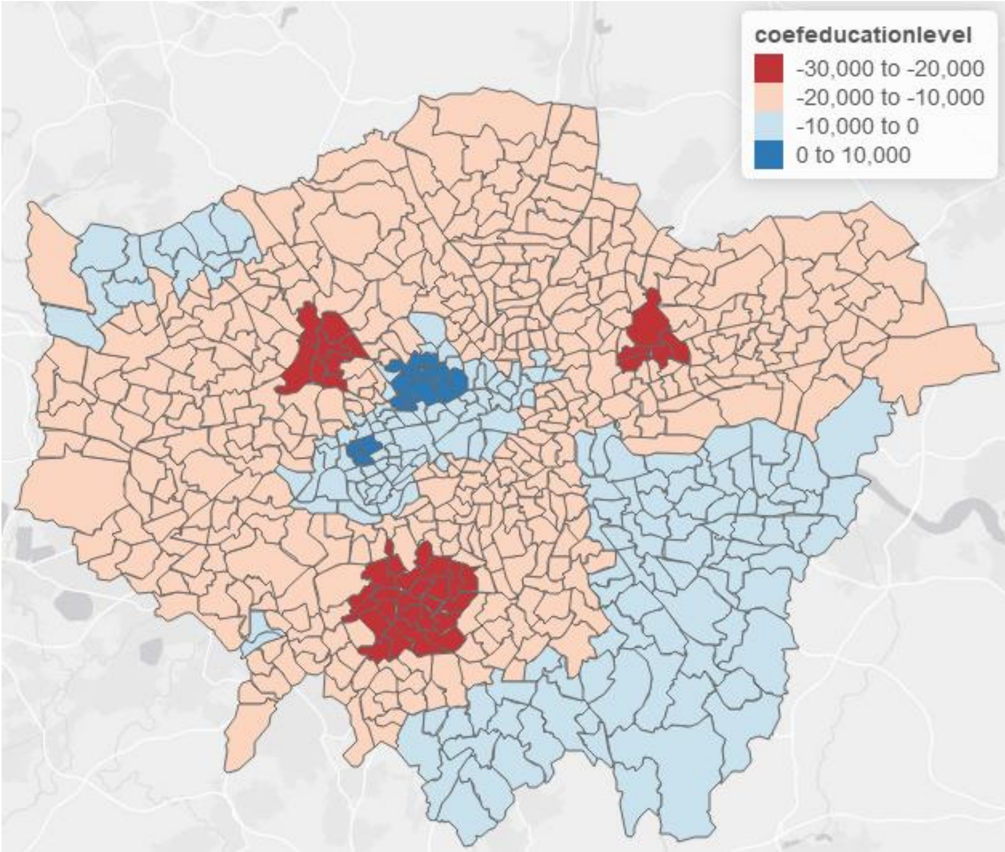

Fig 8 coefficient of education level

Conclusions and recommendations

London has high-cost cluster of house prices in its Midwest and a low-cost cluster in its East, which can be explained with respect to the local life expectancy, the public transport accessibility and the education level. More specifically, a good living standard can enhance life expectancy, which then leads to the monopolization by rich people through high house prices. The district with more effictive public transport service always means the higher house prices inside. However, due to the geographical advantage of the city center and the preference of private transport for people living in wealthy area, the house prices most affected by the public transport are those inside the districts with high public transport accessibility but far from the city center. If the cluster of unschooled people is formed, the local house price will be decreased, which is determined by their low-income level. However, if their ratio is at a low level, like that in Westminster borough, its impact on house prices is very limited.

Therefore, for buyers and housing developers, in order to maximize their profits, they should always pay attention to the improvement of the living resources and facilities nearby their properties in the future such as newly built hospitals, the newly implemented environmental protection plan, and the newly opened subway lines outside the city center. For the government, it is necessary to carefully consider the potential influence on housing market when formulating environmental, medical, transportation, and educational policies. By improve living qualities in other areas, the extremely high house prices in some place can be effectively controlled. Local housing market can be stimulated by the common transportation service level especially in the suburbs of London. In addition, improve the local level of education is also an effective stimulus for the housing market.

References

Currie, G. (2010) 'Quantifying spatial gaps in public transport supply based on social needs', *Journal of Transport Geography*, 18(1), pp. 31–41. doi: 10.1016/j.jtrangeo.2008.12.002.

Davis, L. W. (2004) 'The effect of health risk on housing values: Evidence from a cancer cluster', *American Economic Review*, 94(5), pp. 1693–1704.

Evans, P. and McCormick, B. (1994) 'The New Pattern of Regional Unemployment: Causes and Policy Significance', *The Economic Journal*, 104(424), pp. 633–647. doi: 10.2307/2234638.

Fisher, E. M. (2008) 'USA stratified monopoly: A simulation game about social class stratification', *Teaching Sociology*, 36(3), pp. 272–282.

Green, R. and Hendershott, P. H. (1996) 'Age, housing demand, and real house prices', *Regional Science and Urban Economics*, 26(5), pp. 465–480.

Griliches, Z. and Mason, W. M. (1972) 'Education, Income, and Ability', *Journal of Political Economy*, 80(3, Part 2), pp. S74–S103. doi: 10.1086/259988.

Himmelberg, C., Mayer, C. and Sinai, T. (2005) 'Assessing high house prices: Bubbles, fundamentals and misperceptions', *Journal of Economic Perspectives*, 19(4), pp. 67–92.

Hyun, D. and Milcheva, S. (2019) 'Spatio-temporal effects of an urban development announcement and its cancellation on house prices: A quasi-natural experiment', *Journal of Housing Economics*, 43, pp. 23–36.

Lubitz, J. *et al.* (2003) 'Health, life expectancy, and health care spending among the elderly', *New England Journal of Medicine*, 349(11), pp. 1048–1055.

Sani, N. M. (2013) 'Residual income measure of housing affordability', *International Journal of Advances in Engineering & Technology*, 5(2), p. 1.

So, H. M., Tse, R. Y. and Ganesan, S. (1997) 'Estimating the influence of transport on house prices: evidence from Hong Kong', *Journal of Property Valuation and Investment*, 15(1), pp. 40–47.

Tita, G. E., Petras, T. L. and Greenbaum, R. T. (2006) 'Crime and Residential Choice: A Neighborhood Level Analysis of the Impact of Crime on Housing Prices', *Journal of Quantitative Criminology*, 22(4), p. 299. doi: 10.1007/s10940-006-9013-z.

Yi, J. and Zhang, J. (2010) 'The effect of house price on fertility: evidence from Hong Kong', *Economic Inquiry*, 48(3), pp. 635–650.