Analysis of Economic Development Factors Affecting Carbon Emissions in London by Borough Based on GWR Model

Mengqing Zhao

Programme: MSc Spatial Data Science and Visualisation

Department: Centre for Advanced Spatial Analysis

https://github.com/MengqingZhao/CASA0005_Assignment.git

- ## Introduction

The Intergovernmental Panel on Climate Change (2007, para.4) observed that the global average temperature has risen by 0.74°C in the past 100 years. It means that greenhouse effect has become a social problem cannot be ignored. Releasing greenhouse gases constantly may lead to long-term, serious, and irreversible damages to the environment and human health. IPCC concluded that greenhouse gas emissions caused by human activity are the main reason of rising global temperatures. The impact of carbon emissions on the environment is becoming more and more prominent in the current economy. From the United Nations Framework Convention on Climate Change (UNFCCC) in 1992 to the Kyoto Accord in 1997, and then to the Copenhagen conference in 2009, a series of issues including energy, economics, and environment, have become the focus of policy makers around the world.

Therefore, while the economy is growing rapidly, we should also consider protecting the environment. People should keep a balance between economics development and carbon emissions. The concept of low-carbon economy was first proposed by the UK in an Energy White Paper *Our energy future - creating a low carbon economy* (2003). It is aims to apply low-carbon power sources, like renewable energy resources, to minimize the output of greenhouse gas emissions into the atmosphere, especially carbon dioxide, to protect ecosystems and keep the sustainable development of human beings.

As a result, in order to develop appropriate policies, strategies and standards, it is necessary to analyze the impact of economic development on the carbon emissions and find the important factors that affect carbon emissions. In this report, several factors related to economic development are applied which are average income, employment rate, active enterprises, and population. The aim of this assignment is to analyze of the influences of these economic development factors on carbon emissions in London by borough.

- **Literature review**

At present, relevant scholars around the world have carried out a large number of analyses in different regions. In general, these studies mainly focus on the factor decomposition of carbon emission changes and the prediction of future carbon emissions by establishing the prediction model.

The causality and cointegration between economic growth and carbon emissions in India were analyzed from 1971 to 2006, though ARDL bounds testing combining with incorporating energy supply, investment and employment (Sajal Ghosh, 2010). However, there is only a bi-directional short-run causality rather than a long-run equilibrium relationship. There is a paper examining long-term causal relationship between economic growth, electricity consumption, and carbon emissions in Nigeria from 1970 to 2008 by using a Multivariate Vector Error Correction (VECM) framework (Godwin and Usenobong, 2012). The results indicated that increase of economic is related to growing carbon emissions but as electricity consumption grows, carbon emissions also rise in the long term.

In terms of factor decomposition of carbon emission, Tao, Tingguo, and Lianjun (2007) believe that the relationship between China's per capita CO2 emissions and per capita GDP presents an inverted U-shaped environmental Kuznets curve. Considering spatial effect, Bangying and Fanglin (2010) adopted geographically weighted regression (GWR) technology to introduce spatial model and found that there is an endogenous economic relationship between provincial carbon emissions and economic development factors, containing industrial structure, GDP, population, foreign investment and energy price.

Most scholars have basically covered all the factors affecting carbon emissions in the decomposition of carbon emission change factors and the establishment of carbon emission prediction model. However, the independent factors selected by each scholar are different in factor decomposition and model establishment. In the process of carbon emission prediction, the influence degree of each influencing factor on future carbon emission prediction is different. In addition, some studies have a large time span in the selection of sample data, and the sample data affected by accidental factors are not excluded, so the selected sample data cannot well explain the future trend of carbon emissions. In addition, the carbon emission factors adopted by different scholars in the calculation process of energy conversion are also different, which makes the predicted value of future carbon emission of selected regions vary greatly.

In this assignment, average income, employment rate, active enterprises, and population are selected to analyze that how these factors affect carbon emissions in London by borough. To observe and understand data, descriptive statistics will be implemented first. Because of multiple predictor variables and spatial autocorrelation, not only Ordinary Least Squares (OLS) Multiple Linear Regression should be used but also geographically weighted regression (GWR) model is also supposed to be applied to study the impacts of these factors on carbon emissions in different borough of London.

- ## Methodology

Step1: Data collection and processing

Multiple datasets are applied to complete this analysis and each dataset has been extracted with related key part for research purposes by using Microsoft Excel. Here is the list of the extracted part in each dataset:

Total carbon dioxide emissions in London by borough in 2017 (Greater London Authority, 2017), unit: kilo tonnes.

Average personal incomes of tax payers by tax year in London by borough in 2016-2017 (HM Revenue & Customs, 2018), unit: £ /year

Employment rate in working age in London by borough in 2017 (Office for National Statistics, 2019), unit: %

The number of active enterprises in London by borough in 2017 (Office for

National Statistics, 2017).

Population in London by borough in 2011 (Census Information Scheme, 2011). It should be noticed that because the last population census in the UK was in 2011, there is only precise data of population for 2011.

Then, merge the edited data and London borough shape file together with a common ID by left joining London_Borough_Excluding_MHW.shp and data_edit.csv in R. All the datasets and the shape file can be found in the shared folder.

Step2: Data analysis

To observe and understand data, descriptive statistics was implemented first. Plot the values of these variables on map. Data distribution with histogram should be plotted to check normal distribution.

Because there are multiple predictor variables, Ordinary Least Squares (OLS) Multiple Linear Regression should be used. Examining multicoliniarity between predictor variables with correlation coefficient and Variance Inflation Factor (VIF) is necessary. If the variables are highly correlated with large correlation coefficient value exceeding 0.8 or large VIF exceeding 10, the influence of these variables are effectively double counted and their explanatory power are overstated. Then some variables should be removed and repeat the regression model again. Moreover, the residuals in the model are supposed to be normally distributed. The residuals can be plotted as a histogram and check if there is a normal distribution. Homoscedasticity should also be considered. If the errors do not have constant variance, the parameter estimates might be wrong.

Because this dataset is spatially referenced, it is necessary to check for spatial autocorrelation by mapping the residuals to see if there are any apparent obvious patterns. If yes, test for spatial autocorrelation more systematically by using Moran's I.

Considering spatial autocorrelation, geographically weighted regression (GWR) model is also supposed to be applied. The GWR coefficients for different variables can be plotted on map to explore the coefficient ranges. Then the significance of these results should be tested. Generally, if a coefficient estimate is more than 2 standard errors away from zero, then it is statistically significant. Here is the flowchart of the whole analysis procedure.

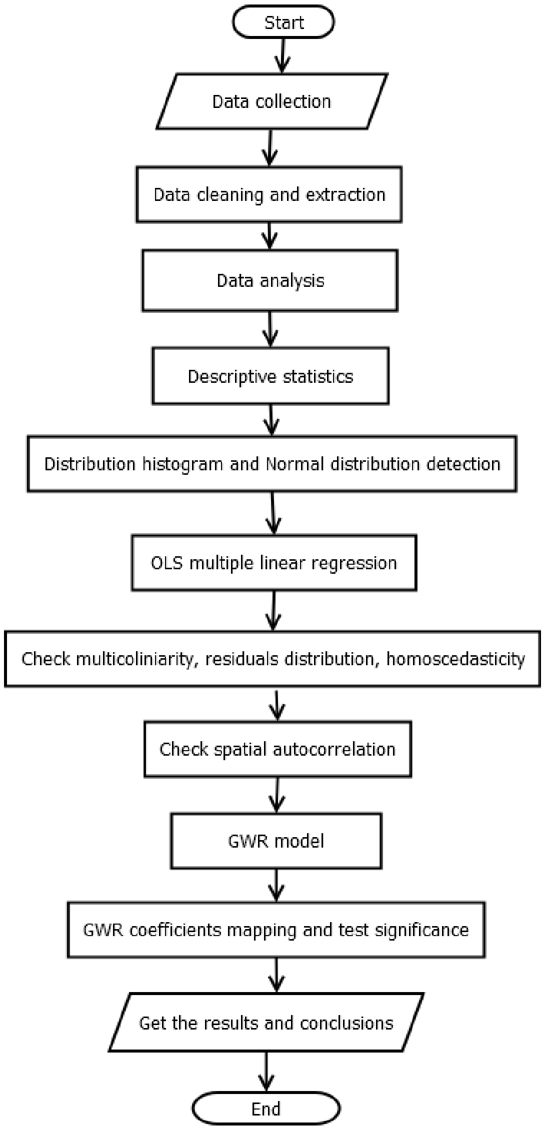


Figure 1: Flowchart of the analysis procedure

- **Results**

Descriptive statistics:

Here are five variables including four independent variables including average income, employment rate, active enterprises, and population and one dependent variable which is carbon emission. At first, plot the values of these variables on map:

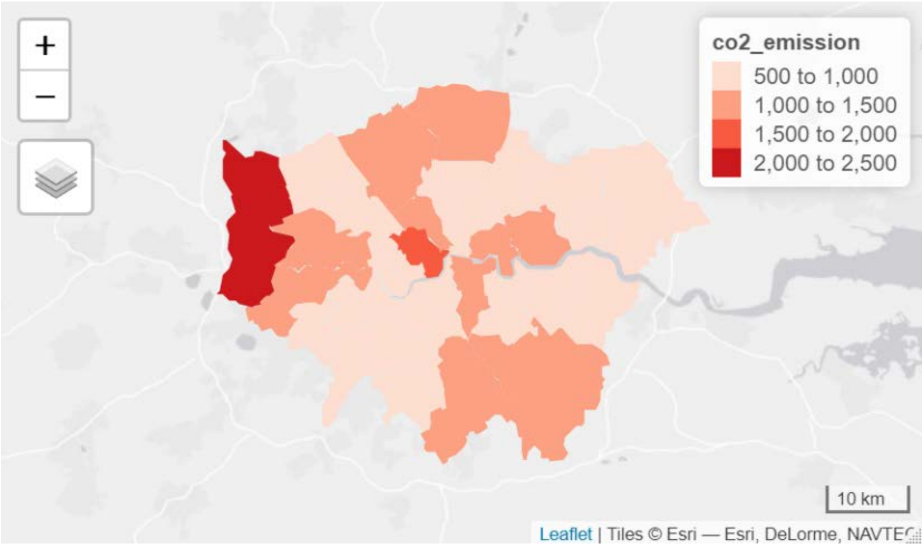
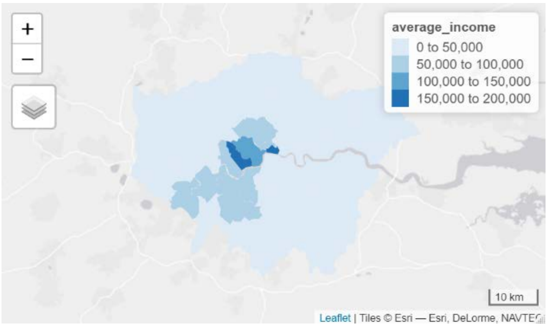


Figure 2: Carbon emissions on map



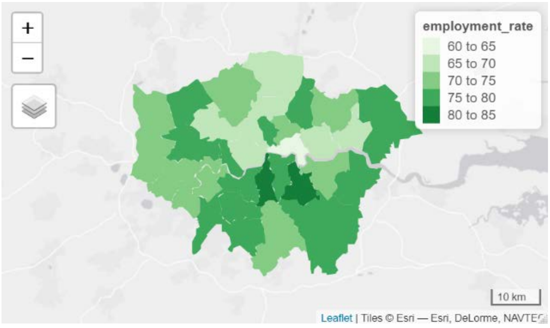
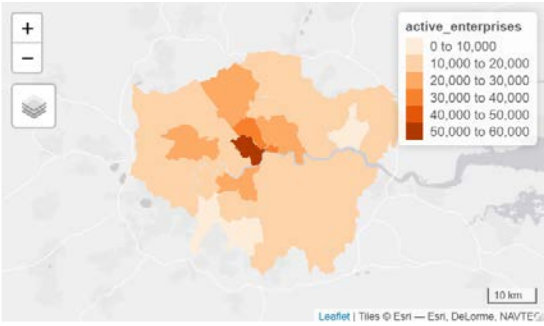


Figure 3: Average incomes on map

Figure 4: Employment rates on map



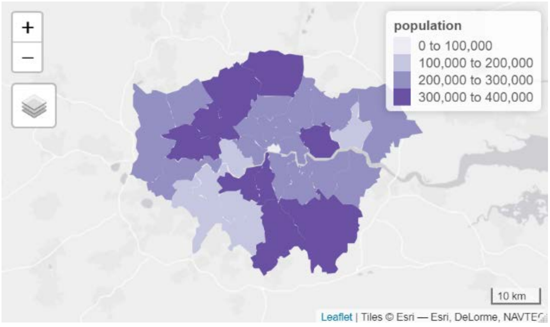


Figure 5: Active enterprises on map

Figure 6: Population on map

According to the above five figures, it is clear that West London has the highest carbon emissions, followed by the middle area. Average incomes are higher in central and south-west London. Central London has the highest employment rate and the most active enterprises. There are more people in north and south London.

Then the data distribution should be plotted to check normal distribution:

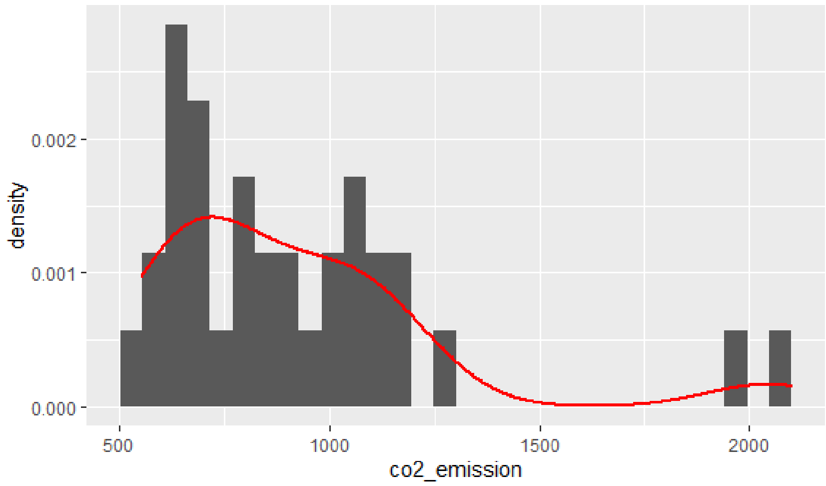
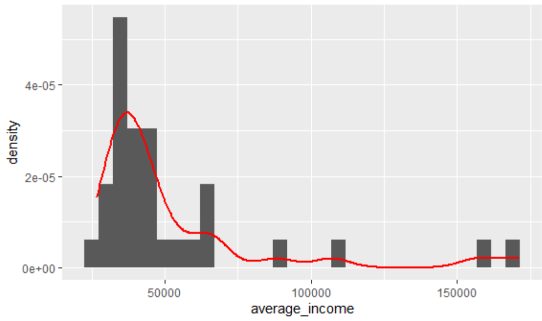


Figure 7: The distribution histogram of carbon emissions



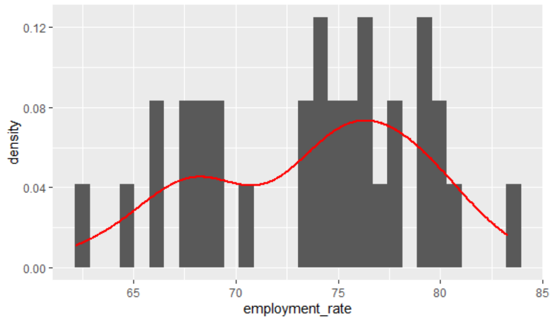
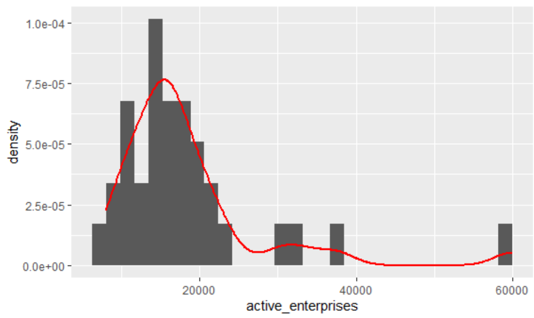


Figure 8: The distribution histogram of average incomes

Figure 9: The distribution histogram of employments



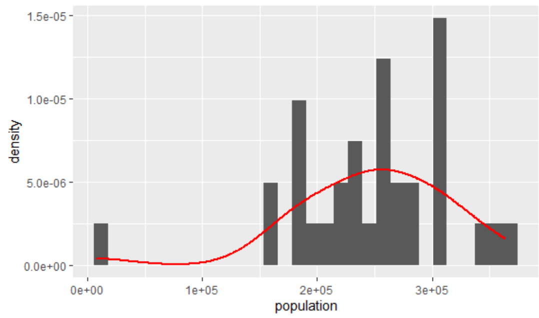


Figure 10: The distribution histogram of active enterprises

Figure 11: The distribution histogram of population

According to the above five figures, because there are only 33 items in the dataset. This amount might be a little bit small and the above data distributions can be approximated as normally distributed. There will be a linear relationship.

OLS Regression model:

The next part is OLS Regression model with multiple factors. Here are the results of multiple regression:

term	estimate	std.error	statistic	p.value
(Intercept)	896.4768867	832.079466	1.077393354	0.290499747
average_income	0.000246482	0.002280427	0.108085742	0.9146985
employment_rate	-11.74129887	10.27262528	-1.142969646	0.262730237
active_enterprises	0.018659181	0.006264452	2.978581278	0.005922674
population	0.002143731	0.000949629	2.257439806	0.031968795

Table 1: Result of OLS multiple regression

According to Table 1, employment_rate is the only one with a negative coefficient and other factors are positively related to carbon emissions. It should be noticed that p-value of population is lower than 0.05 so population is statistically significant and p-value of active_enterprises is lower than 0.01 so population is statistically highly significant.

The R-Squared value of this regression model is 0.484743226, which means that this model explains about 48.5% of variance and leave a lot of variation unaccounted for.

Then examine the multicoliniarity between predictor variables with correlation coefficient and (VIF)

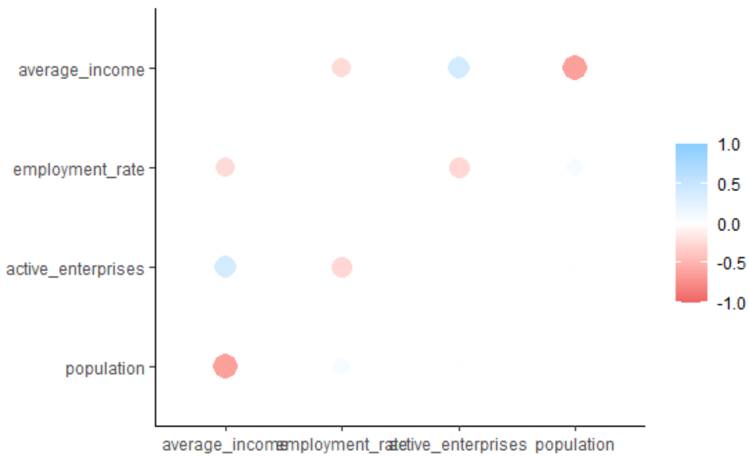


Figure 12: Correlation coefficients between the predictor variables

term	average_income	employment_rate	active_enterprises	population
average_income	NA	-0.390	0.510	-0.627
employment_rate	-0.390	NA	-0.406	0.248
active_enterprises	0.510	-0.406	NA	-0.0858
population	-0.627	0.248	-0.0858	NA

Table 2: Correlation coefficients between the predictor variables

According to Figure 12 and Table 2, all the values of correlation coefficients are less than a 0.8 correlation.

average_income	employment_rate	active_enterprises	population
2.542972	1.280125	1.668702	1.896808

Table 3: VIF between the predictor variables

According to Table 3, all the values of VIFs are less than 10.

Both correlation coefficients and VIFs indicate that these predictor variables are ideal for the multiple regression model.

The residuals can be plotted as a histogram and check if there is a normal distribution.

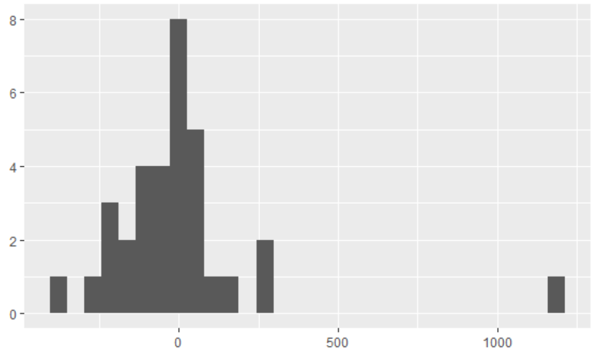


Figure 13: The histogram of residuals

According to Figure 13, the above data distribution can be approximated as normally distributed.

Then check homoscedasticity by plotting the residuals in the model against the predicted values.

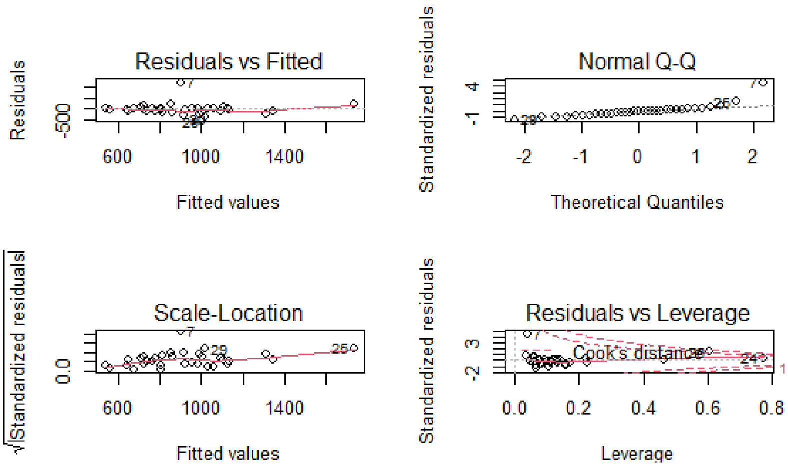


Figure 14: The residuals in the model against the predicted values

According to Figure 14, in the first plot (residuals vs fitted), it is obvious that there is a random cloud of points with a straight horizontal red line, which means that the parameter estimates is correct.

Because this dataset is spatially referenced, it is necessary to check for spatial autocorrelation by mapping the residuals to see if there are any apparent obvious patterns.

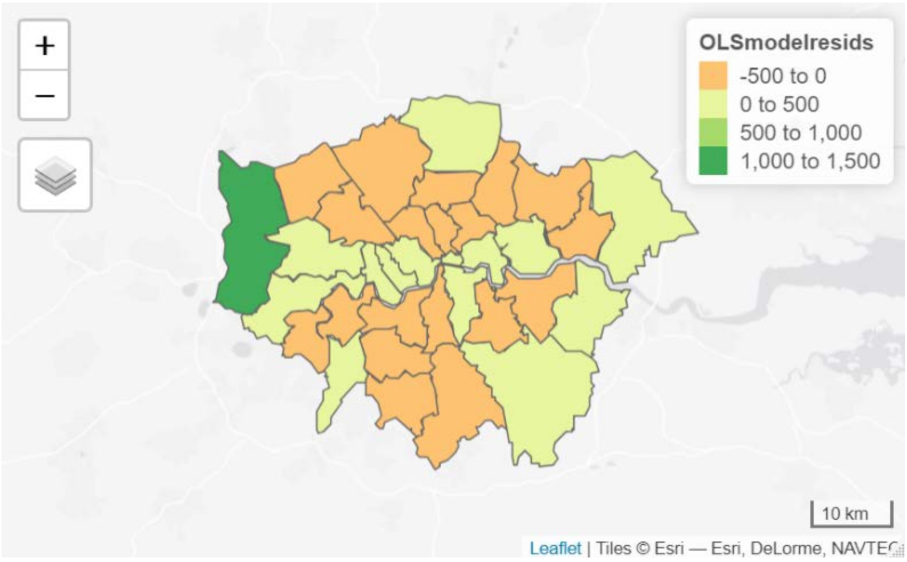


Figure 15: The residuals map

According to Figure 15, there look to be some areas next to other areas with the same colors, which means that there are some spatial autocorrelation biasing this model. Then, test for spatial autocorrelation more systematically by using Moran's I.

```
estimate1 estimate2 estimate3 statistic p.value method            alternative
    <dbl>     <dbl>     <dbl>     <dbl>   <dbl> <chr>             <chr>
   0.0941   -0.0312   0.00584      1.64  0.0504 Moran I test un~  greater
```

Figure 16: Moran's I statistic for k-nearest neighbours of 4

According to the result in Figure 16, it is clear that there is some weak spatial autocorrelation in the residuals.

GWR model:

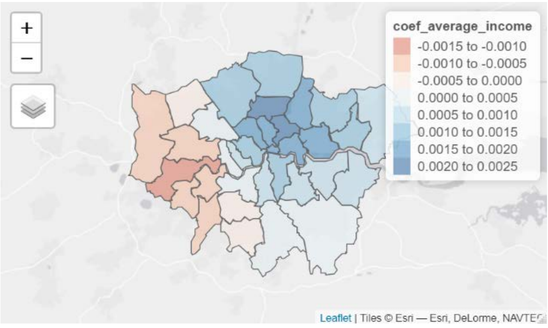
Considering spatial autocorrelation, geographically weighted regression (GWR) model is also supposed to be applied.

```
                        Min.      1st Qu.      Median      3rd Qu.
X.Intercept.        4.8568e+02  7.1038e+02  8.2760e+02  9.6984e+02
average_income     -1.1992e-03  1.3705e-04  8.8861e-04  1.4776e-03
employment_rate    -2.1495e+01 -1.3969e+01 -1.2690e+01 -1.1676e+01
active_enterprises  1.7417e-02  1.8416e-02  1.9463e-02  2.0601e-02
population          1.9010e-03  2.1190e-03  2.2074e-03  2.4425e-03
                        Max.    Global
X.Intercept.        1.7831e+03 896.4769
average_income      2.1089e-03   0.0002
employment_rate    -1.0177e+01 -11.7413
active_enterprises  2.1699e-02   0.0187
population          2.7946e-03   0.0021
Number of data points: 33
Effective number of parameters (residual: 2traceS - traceS'S): 10.86092
Effective degrees of freedom (residual: 2traceS - traceS'S): 22.13908
Sigma (residual: 2traceS - traceS'S): 275.7021
Effective number of parameters (model: traceS): 8.645226
Effective degrees of freedom (model: traceS): 24.35477
Sigma (model: traceS): 262.862
Sigma (ML): 225.8203
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 479.8292
AIC (GWR p. 96, eq. 4.22): 459.998
Residual sum of squares: 1682828
Quasi-global R2: 0.5784918
```

Figure 17: Summary of GWR coefficient estimates at data points

According to the result in Figure 17, it shous that how the coefficients vary across the 33 boroughs in London. The global coefficients are the same as the coefficients in the above OLS linear regression model. It can be found that R-Squared value of this GWR model is 0.5784918, larger than R-Squared value of the above linear model which is 0.484743226.

To explore the coefficient ranges, plot the GWR coefficients for different variables on map.



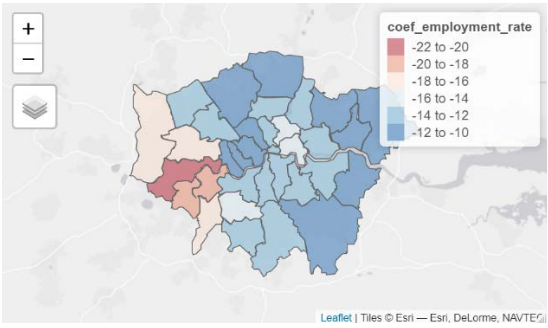
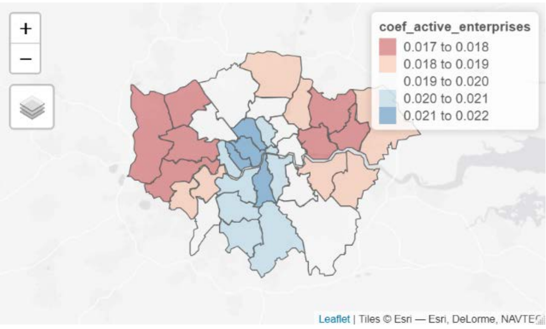


Figure 18: GWR coefficients for average incomes
Figure 19: GWR coefficients for employment rates



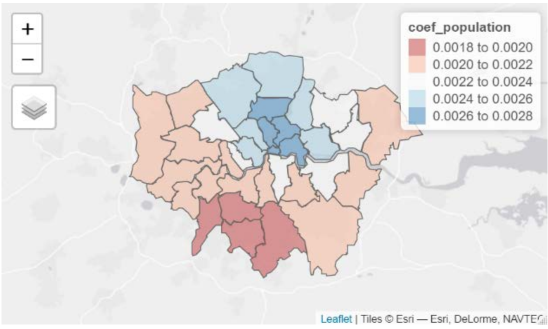
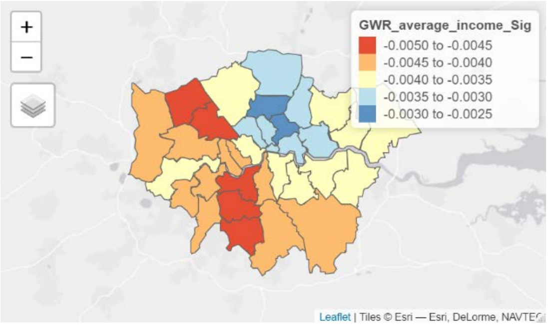


Figure 20: GWR coefficients for active enterprises
Figure 21: GWR coefficients for population

According to the above four figures, for average incomes, there is a negative relationship in west London and a positive relationship in other areas. For employment rates, there is a negative relationship in west London and a positive relationship in other areas. For active enterprises, there is a negative relationship in south London and a positive relationship in other areas. For population, there is a negative relationship in north London and a positive relationship in other areas.

Then to test the significance of these results, calculate "coefficient estimate - 2*standard errors" and then plot the calculation results on map.



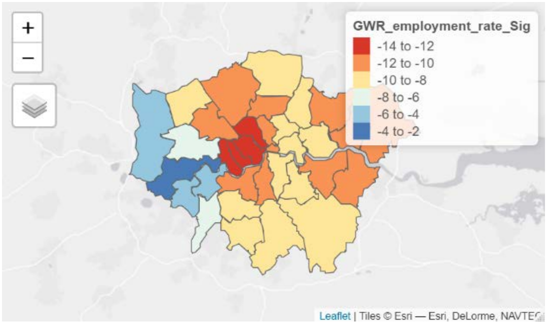
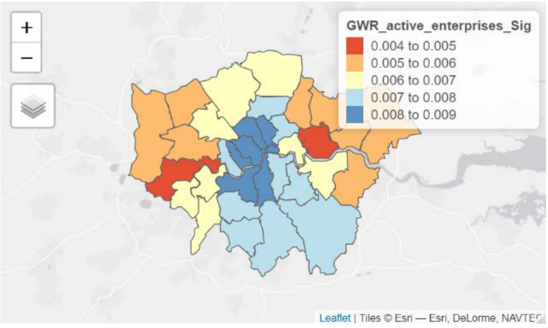


Figure 22: The calculation results for average incomes
Figure 23: The calculation results for employment rates



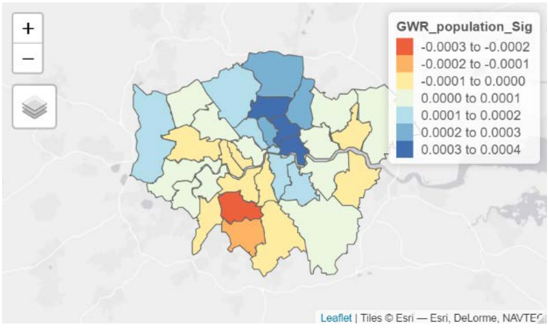


Figure 24: The calculation results for active enterprises
Figure 25: The calculation results for population

If the value is larger than 0, which means that the coefficient estimate is more than two standard errors away from zero, it will be statistically significant with the 95% confidence level.

According to the above four figures, it is obvious that the coefficient estimates for average incomes and employment rates are not significant with all negative values. The coefficient estimate for active enterprises is statistically significant with all positive values. The coefficient estimate for population is locally significant with both positive values and negative values.

- ## Discussion

From the above analysis, the number of active enterprises has a significant influence on carbon emissions and population has a local significant influence on carbon emissions. Active enterprises and population are positively correlated in all the boroughs. Employment rate is negatively correlated in all the boroughs. Average income is negatively correlated in west London and positively correlated in east London. Therefore, in order to reduce carbon emissions, it is suggested that improve the employment rate and keep a balance in average incomes between different boroughs. Therefore, the local government of London can develop the economy with establish and perform carbon emissions reduction policy and conservative energy policy with more renewable resources. At the same time, it should also be paid attention to reducing the differences in economic levels between different regions and adjusting the industrial structure of west London.

In addition, it would be better to add time dimension to observe the trends of multiple factors and carbon emissions in the long run. The analysis result in only one year is not persuasive enough, as it is kind of unstable and unreliable.

- ## Conclusion

In this assignment, the influences of the economic development factors, including average income, employment rate, active enterprises, and population, on carbon emissions in London by borough have been analyzed. Active enterprises and population are positively correlated to carbon emissions in all the boroughs. Employment rate is negatively correlated to carbon emissions in all the boroughs. Average income is negatively correlated in west London and positively correlated in east London. It reminds that while developing the economy, people should also pay attention to energy conservation and emission reduction. It is supposed to adjust industrial structure, keep a balance in the difference in economic level. Establishing and performing eco-friendly policies, strategies and standards, developing and promoting the use of renewable energy sources should be put into practice without delay.

(Word count: 2531 words)

References

Pachauri, R.K and Reisinger, A. (2007) ‘Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change’. *IPCC*. pp. 104. Available from: https://www.ipcc.ch/report/ar4/syr/ (Accessed: 10 January 2021)

Energy White Paper (2003) *Our energy future - creating a low carbon economy*. UK: DTI.

Sajal. G. (2010) ‘Examining carbon emissions economic growth nexus for India: A multivariate cointegration approach’. *Energy Policy*. 38 (6), pp. 3008-3014.

Godwin. E. A and Usenobong. F. A. (2012) ‘Electricity Consumption, Carbon Emissions and Economic Growth in Nigeria’, *International Journal of Energy Economics and Policy*. 2 (4), pp. 292-306. [Online] Available from: https://www.researchgate.net/profile/Usenobong_Akpan/publication/265285198_Electricity_Consumption_Carbon_Emissions_and_Economic_Growth_in_Nigeria/links/54d47fa50cf2970e4e63498a/Electricity-Consumption-Carbon-Emissions-and-Economic-Growth-in-Nigeria.pdf (Accessed: 10 January 2021)

Tao. S, Tingguo. Z, and Lianjun. T. (2007) ‘Theoretical analysis and econometric testing on relationship of environmental pollution and economic growth’. *Scientia Geographica Sinica*. 27(2), pp. 156-162

Bangying.S and Fanglin.S. (2010) ‘Local spatial pattern analysis of carbon emissions and economic development—An empirical study of provincial low-carbon economy based on GWR’. *Finance Economics*. 265, pp.41-49

Greater London Authority (GLA) (2017) *London Energy and Greenhouse Gas Inventory (LEGGI)* [Online] Available from: https://data.london.gov.uk/dataset/leggi (Accessed: 10 January 2021)

HM Revenue & Customs (2018) *Average Income of Tax Payers, Borough* [Online] Available from: https://data.london.gov.uk/dataset/average-income-tax-payers-borough (Accessed: 10 January 2021)

Office for National Statistics (ONS) (2019) *Employment Rates by Gender and*

Age or Disability, Borough [Online] Available from: https://data.london.gov.uk/dataset/employment-rates-by-gender-and-age (Accessed: 10 January 2021)

Office for National Statistics (ONS) (2018) *Business Demographics and Survival Rates, Borough* [Online] Available from: https://data.london.gov.uk/dataset/business-demographics-and-survival-rates-borough (Accessed: 10 January 2021)

Census Information Scheme (2011) *Historical Census Tables* [Online] Available from: https://data.london.gov.uk/dataset/historical-census-tables (Accessed: 10 January 2021)

Greater London Authority (GLA) (2014) *Statistical GIS Boundary Files for London* [Online] Available from: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london (Accessed: 10 January 2021)

Declaration of Authorship

I, Mengqing Zhao, confirm that the work presented in this assessment is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Mengqing Zhao

Date of signature: 11 Jan 2021

Assessment due date: 11 Jan 2021