Housing conditions and crime rates in London 2021

Ana Cuba Varas

October 2024

Introduction

How are **housing conditions** related to **crime**?

Social disorganization theory

More residential turnover
Less social control

+

Socio economic factors

→

More crime

Housing problems in London:
- Affordability
- Quality
- Inequality
- Gentrification

Variables

- Accommodation types
- Tenure
- House price per square meter
- Vacant dwellings

+

Deprivation

→

- Total crime
- Violence and sexual offences
- Anti-social behaviour crimes
- Burglaries

How are housing conditions related to crime?

1. **Is crime correlated to housing conditions?** → correlations with each variable and regressions (3 α comparisons: regression with tenure, accommodation type and deprivation)
2. **Can a combination of housing and deprivation conditions help explain crime hotspots?** → regression model (1 comparison)
3. **If we work with clusters to categorize the data, does it improve the model?** → regression model (2 comparisons)
4. **Is geography (space) relevant for the variables considered?** → spatial autocorrelation
5. **Are the models showing signs of these spatial autocorrelations?** → errors check
6. **Does adding space improve the model?** → spatial regression models (3 comparisons)

Bonferroni correction: $\frac{\alpha}{n} = \frac{0.05}{9} = 0.00556$

Variables

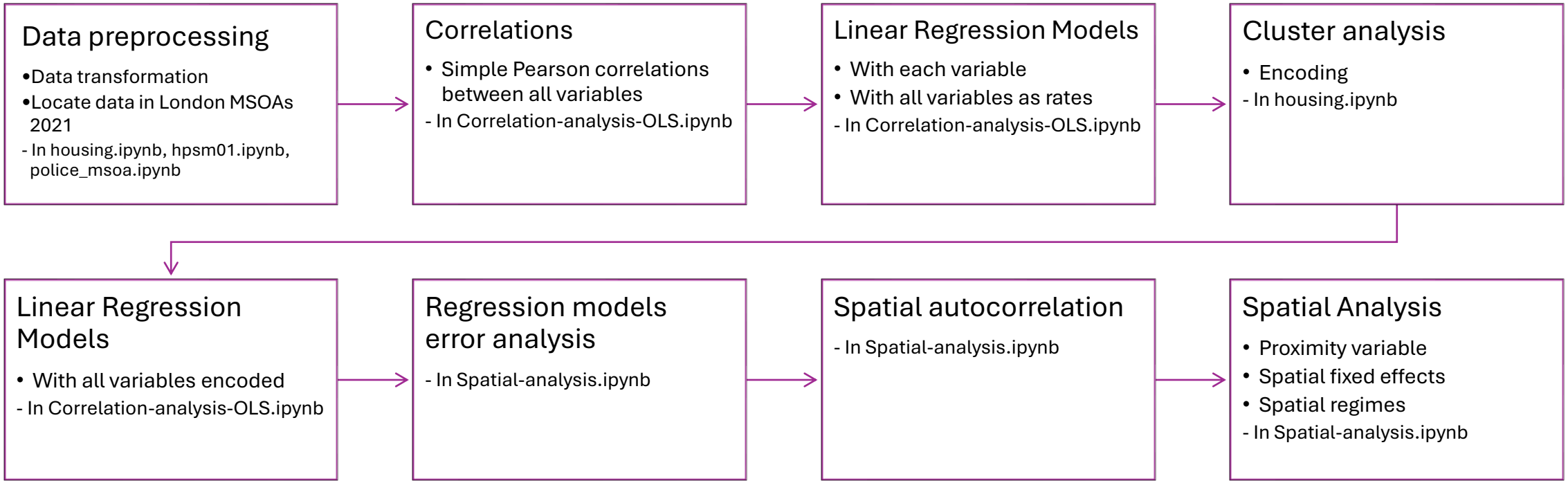
Datasets	Variables	Columns
- Police data crimes reported - House price per square metre (hpsm) - Nomis TS011: Households by deprivation dimensions - Nomis TS044: Accommodation type - Nomis TS054: Tenure - Nomis RM204: Number of dwellings - ONS, Open geography portal: Postcode lookup (2023) - ONS, MSOA population estimates 2021 - ONS, Open geography portal: MSOA 2021 boundaries	**Crime**	- Total crime - Violence and sexual offences - Anti-social behaviour crimes - Burglaries
	House price	Average house price per square metre for each MSOA, 2021
	Tenure	- Owned (with or without mortgage) - Private rent - Social rent (Omitted: own share and rent free)
	Accommodation type	- Houses (detached, semi-detached, terraced) - Flats (purpose-built) - Other (part of a converted house, commercial or other building, caravans and temporary structures)
	Vacant dwellings	Dwellings with no households living in
	Deprivation	- Not deprived in employment, education, health or overcrowding - Deprived in one dimension - Deprived in 2 dimensions - Deprived in 3 dimensions - Deprived in 4 dimensions

Variables

Dependent variables	
Total crime	$\log_{10} \frac{Total\ crimes\ reported}{Population/1000}$
Violence and sexual offences	$\log_{10} \frac{VSO}{Population/1000}$
Anti-social behaviour crimes	$\log_{10} \frac{ASB}{Population/1000}$
Burglaries	$\log_{10} \frac{Burglaries\ reported}{Number\ of\ dwellings/1000}$

Independent variables	
House price per square metre	$\log_{10}(Average\ hpsm)$
Vacant dwellings	% Empty dwellings
Tenure	% Own % Private rent % Social rent
Accommodation type	% Houses % Flats % Other
Deprivation	% not deprived % deprived in one dimension % deprived in 2 dimensions % deprived in 3 dimensions % deprived in 4 dimensions

Implementation



Is crime correlated to housing conditions?

Pearson correlation each variable with crime

Accommodation type

	%_Houses	%_Flats	%_Other
%_Houses	1		
%_Flats	-0.86477	1	
%_Other	-0.47340	0.02199	1

Tenancy

	%_Owned	%_Social-rent	%_Private-rent
%_Owned	1		
%_Social-rent	-0.71857	1	
%_Private-rent	-0.55332	-0.09975	1

Deprivation

	%_not deprived	%_deprived-1	%_deprived-2	%_deprived-3
%_not deprived	1			
%_deprived-1	-0.55062	1		
%_deprived-2	-0.84615	0.77843	1	
%_deprived-3	-0.83521	0.50319	0.88177	1
%_deprived-4	-0.63513	0.23571	0.56568	0.70002

Correlations with crime

Variables	Total crime	Violence and sexual offences	Anti-social Behaviour crimes	Burglaries
%_not deprived	-0.44656	-0.59238	-0.46017	-0.23348
%_deprived-1	-0.19627	0.05959	-0.18271	-0.12459
%_deprived-2	0.15942	0.39149	0.18678	0.11364
%_deprived-3	0.33134	0.48728	0.37445	0.25147
%_deprived-4	0.38619	0.43794	0.43293	0.32042
Log10-price	0.36259	0.08537	0.38627	0.27957
%_Owned	-0.69459	-0.64234	-0.74105	-0.47956
%_Social-rent	0.37200	0.46057	0.41529	0.26683
%_Private-rent	0.44285	0.32428	0.48307	0.36827
Empty %	0.55777	0.36577	0.53215	0.25542
%_Houses	-0.61079	-0.45297	-0.66401	-0.43511
%_Flats	0.48357	0.41132	0.54931	0.33669
%_Other	0.26198	0.12126	0.28265	0.28362

Is crime correlated to housing conditions?

Regression models for each variable with crime

Adjusted R-squared	Total crimes	Violence and Sexual Offences	Anti-social Behaviour	Burglaries
Deprivation	0.4834	**0.4513**	0.4938	0.1767
Tenure	**0.5431**	0.4165	**0.5887**	**0.2431**
Accommodation type	0.4299	0.2264	0.4689	0.1947

Significance of coefficients (no constant)	Total crimes	Violence and Sexual Offences	Anti-social Behaviour	Burglaries
Deprivation	2 / 5	2 / 5	3 / 5	3 / 5
Tenure	**3 / 3**	2 / 3	**3 / 3**	1 / 3
Accommodation type	**3 / 3**	**3 / 3**	**3 / 3**	1 / 3

Can a combination of housing and deprivation conditions help explain crime?

Regression model all variables with crime

Adjusted R-squared	Total crimes	Violence and Sexual Offences	Anti-social Behaviour	Burglaries
Deprivation	0.4834	0.4513	0.4938	0.1767
Tenure	0.5431	0.4165	0.5887	0.2431
Accommodation type	0.4299	0.2264	0.4689	0.1947
All variables together (including vacant dwellings and house price)	**0.5598**	**0.5199**	**0.6018**	0.2669

Significance of coefficients (no constant)	Total crimes	Violence and Sexual Offences	Anti-social Behaviour	Burglaries
Deprivation	0 / 5	0 / 5	0 / 5	0 / 5
Tenure	1 / 3	1 / 3	2 / 3	2 / 3
Accommodation type	0 / 3	0 / 3	0 / 3	0 / 3
Vacant dwellings	0	0	0	0
House price	0	1	0	1
Total independent variables	**1/13**	**2/13**	**2/13**	**3/13**

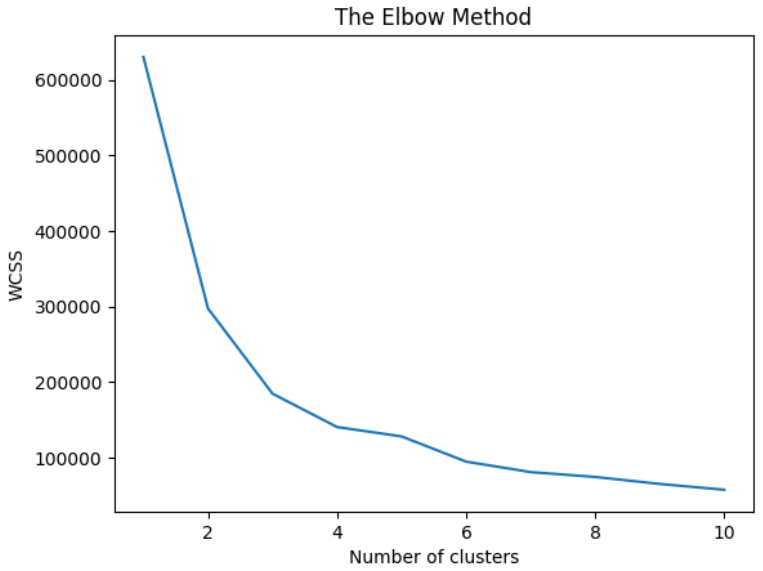
Can cluster data improve the model?

Variables	As rates (percentages)	Encoded
Tenure	- Own - Private rent - Social rent	- Mixed tenures - Owned & Private - More private rent - More owned - More social rent
Accommodation type	- Houses - Flats - Other	- Mainly flats - Mainly houses - Mixed types
Deprivation	- not deprived, - deprived in one dimension - deprived in 2 dimensions - deprived in 3 dimensions, - deprived in 4 dimensions	- Most deprived - Mixed extremes - Mixed less deprived - Mixed most deprived - Less deprived

Can cluster data improve the model?

For each compound variable

1. Get the elbow graph
2. Get the silhouette coefficient
3. Select the number of clusters
4. Filter and get the statistics for each cluster
5. Analyse them in Excel to infer the cluster name
6. One hot encoder and assign names



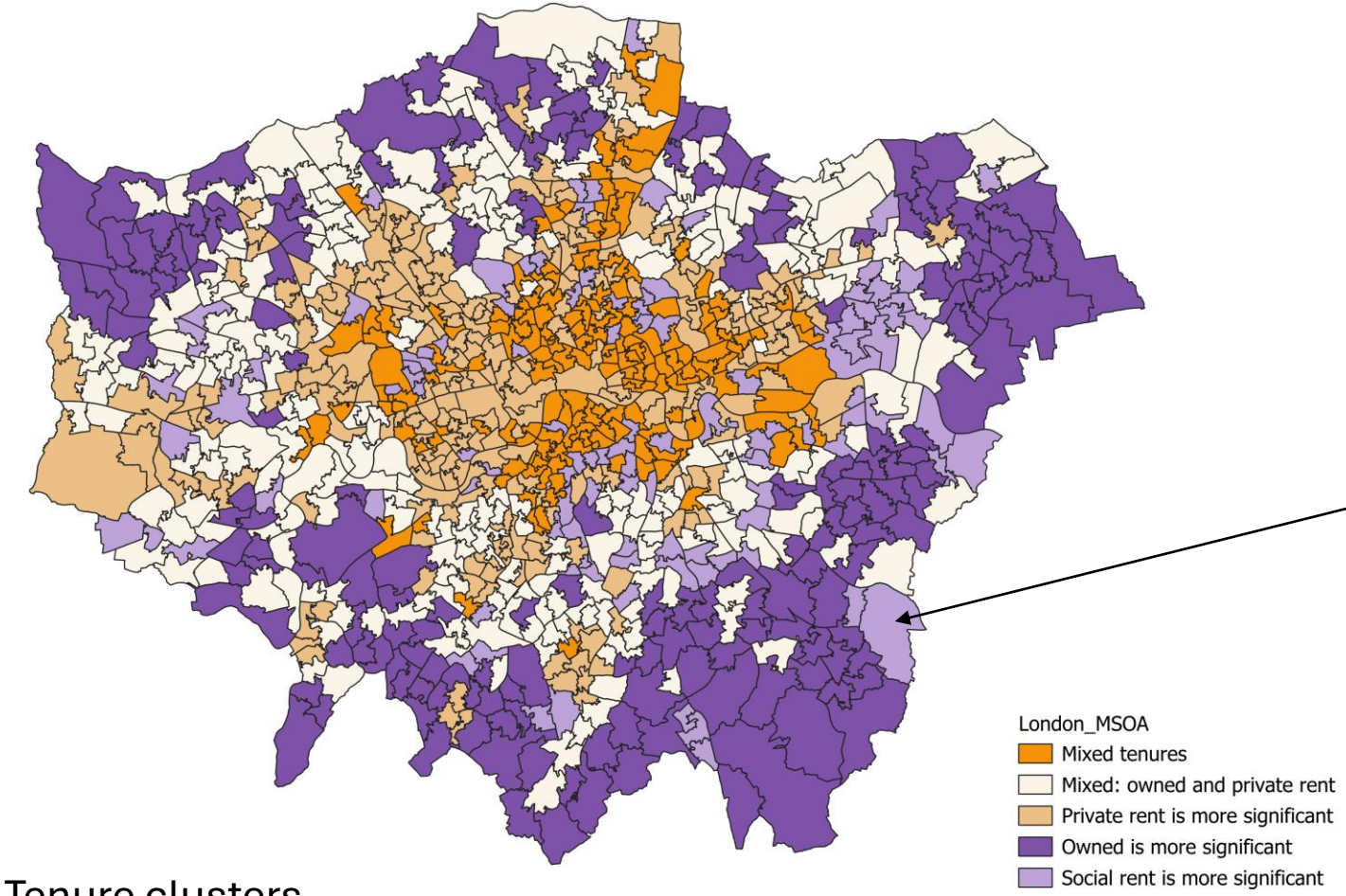

```
For cluster= 2, Silhouette Coefficient is 0.43973556345986037
For cluster= 3, Silhouette Coefficient is 0.4120297705222142
For cluster= 4, Silhouette Coefficient is 0.3716442583534864
For cluster= 5, Silhouette Coefficient is 0.39015234158711526
For cluster= 6, Silhouette Coefficient is 0.3717761769931299
For cluster= 7, Silhouette Coefficient is 0.3524758435764228
```

Tenure clusters

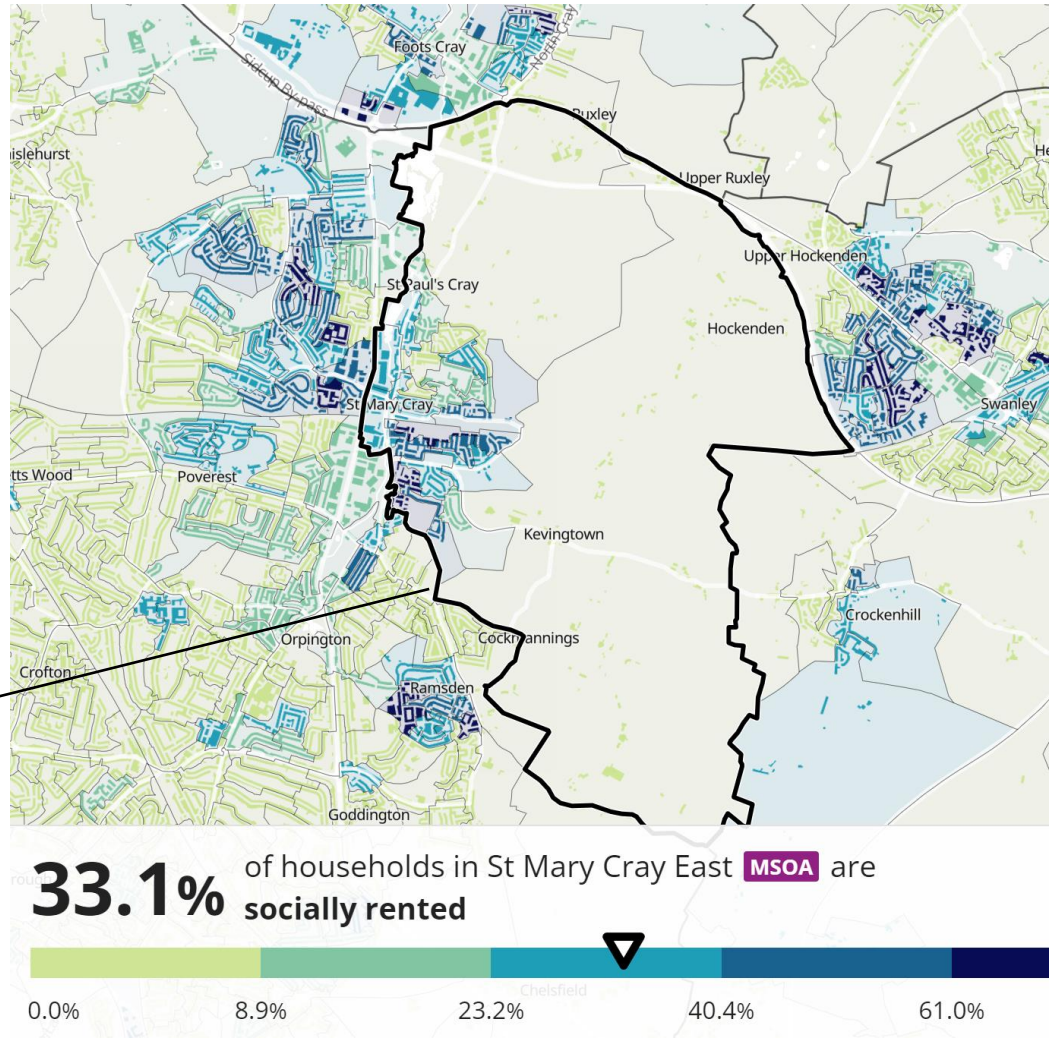
	Cluster 0			Cluster 1			Cluster 2			Cluster 3			Cluster 4		
	%_Owned	%_Social-rent	%_Private-rent	%_Owned	%_Social-rent	%_Private-rent	%_Owned	%_Social-rent	%_Private-rent	%_Owned	%_Social-rent	%_Private-rent	%_Owned	%_Social-rent	%_Private-rent
mean	21.6034	37.0625	29.5653	51.6893	15.3573	26.6155	33.3229	16.5473	37.3836	72.2392	7.6162	15.6231	31.1677	42.4531	18.8918
std	5.2428	7.5425	4.5137	5.6547	6.9983	6.1390	7.7913	6.4863	5.7914	7.4058	4.7741	5.2583	9.9137	10.0470	4.0713
min	10.3808	21.0032	18.9710	39.7248	2.6364	8.4794	11.3618	1.4824	22.2014	61.5476	1.2733	4.8583	9.9576	27.1470	8.8804
max	34.7289	56.2926	42.8969	62.4314	31.5414	43.1637	47.4684	29.7027	61.2553	89.4548	22.7618	29.7795	50.4009	67.6079	26.5105
Inf.	Mixed tenures			Mixed: owned and private rent			Private rent is more significant			Owned is more significant			Social rent is more significant (with owned)		

Can cluster data improve the model?

- Visualising encoded data in maps (QGIS)
- Check if it made sense in reality



Tenure clusters



Source: https://www.ons.gov.uk/census/maps/choropleth/housing/tenure-of-household/hh-tenure-5a/rented-social-rented?msoa=E02000145

Can cluster data improve the model?

Regression model all variables after encoding with crime

Adjusted R-squared	Total crimes	Violence and Sexual Offences	Anti-social Behaviour	Burglaries
All variables together (including vacant dwellings and house price)	**0.5598**	**0.5199**	**0.6018**	0.2669
All variables together after encoding	0.5179	0.4796	0.5631	0.2438

Significance of coefficients (no constant)	Total crimes	Violence and Sexual Offences	Anti-social Behaviour	Burglaries
Deprivation	2 /4	3 / 4	1 / 4	3 / 4
Tenure	2 / 4	3 / 4	3 / 4	3 / 4
Accommodation type	0 / 2	0 / 2	1 / 2	0 / 2
Vacant dwellings	1	1	1	0
House price	0	1	0	1
Total independent variables:	**5/12**	**8/12**	**6/12**	**7/12**

Is geography relevant for the variables?

Global spatial autocorrelation

Variables as rates

Variable	Moran's I	p-value
Log10-tot	0.468748	0.001
Log10-VSO	0.375409	0.001
Log10-ASB	0.494389	0.001
Log10-Bur	0.401564	0.001
%_not deprived	**0.620349**	0.001
%_deprived-1	**0.714593**	0.001
%_deprived-2	0.569703	0.001
%_deprived-3	0.502256	0.001
%_deprived-4	0.371753	0.001
Empty %	**0.783216**	0.001
%_Houses	**0.741416**	0.001
%_Flats	0.548163	0.001
%_Other	**0.609899**	0.001
%_Owned	**0.735682**	0.001
%_Social-rent	0.492993	0.001
%_Private-rent	0.558733	0.001
Log10-price	**0.886571**	0.001

Encoded variables

Variable	Moran's I	p-value
Log10-tot	0.468748	0.001
Log10-VSO	0.375409	0.001
Log10-ASB	0.494389	0.001
Log10-Bur	0.401564	0.001
Most deprived	0.422333	0.001
→ Mixed extremes	**0.614855**	0.001
↔ **Mixed less deprived**	0.153224	0.001
Mixed most deprived	0.164473	0.001
Less deprived	0.377653	0.001
Log10-vac	0.716266	0.001
Mainly flats	0.505061	0.001
Mainly houses	0.506354	0.001
Mixed types	0.260787	0.001
→ Mixed tenures	0.378670	0.001
Owned & Private	0.229879	0.001
More private rent	0.275576	0.001
More owned	**0.467926**	0.001
↔ **More social rent**	0.156956	0.001
Log10-price	0.886571	0.001

Is Geography relevant in the model?

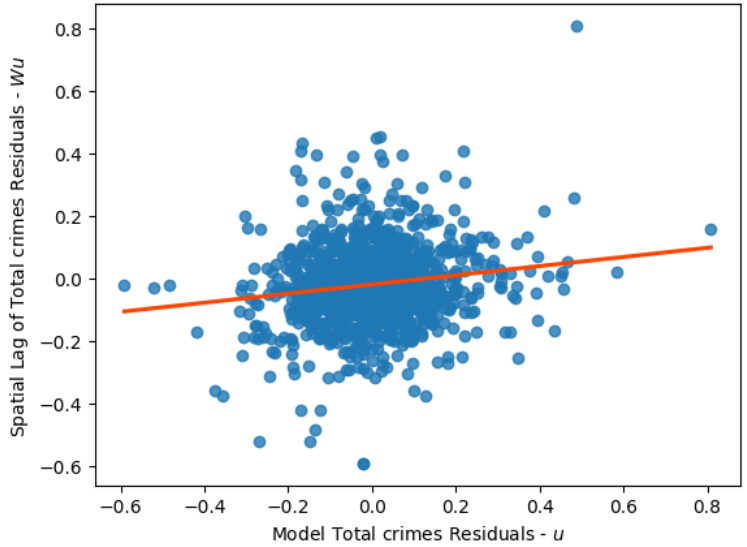
2nd Regression model all variables after encoding with crime

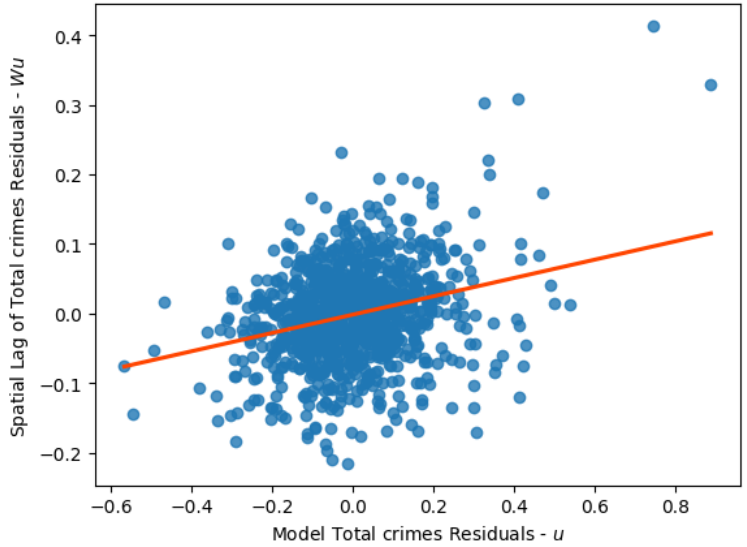
Significance of coefficients (change from 1st model)	Total crimes	Violence and Sexual Offences	Anti-social Behaviour	Burglaries
Deprivation	4/4 (+2)	4/4 (+1)	3/4 (+2)	2/4 (-1)
Tenure	2/4	2/4 (-1)	2/4(-1)	2/4(-1)
Accommodation type	0/2	0/2	1/2	0/2
Vacant dwellings	1	1	1	0
House price	0	1	0	1
Total independent variables:	**7/12 (+2)**	**8/12**	**7/12 (+1)**	**5/12 (-2)**

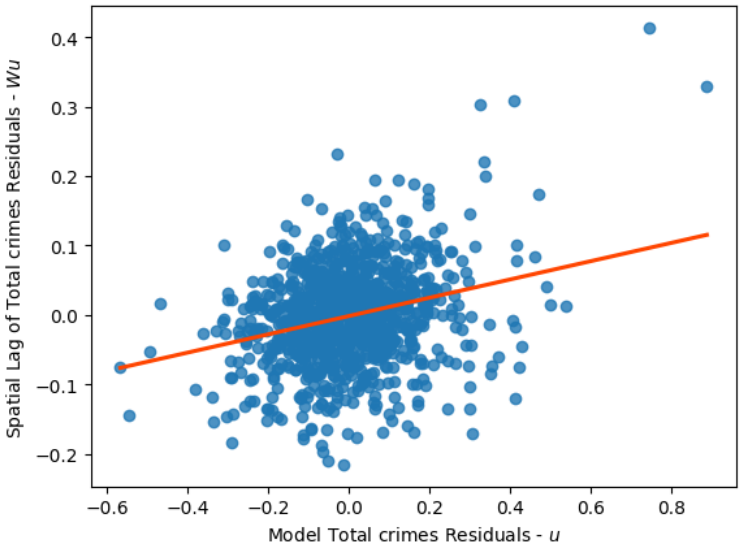
Are the models showing these spatial autocorrelations?

Prediction error at each MSOA and the prediction error at the MSOA nearest to it

Total Crimes

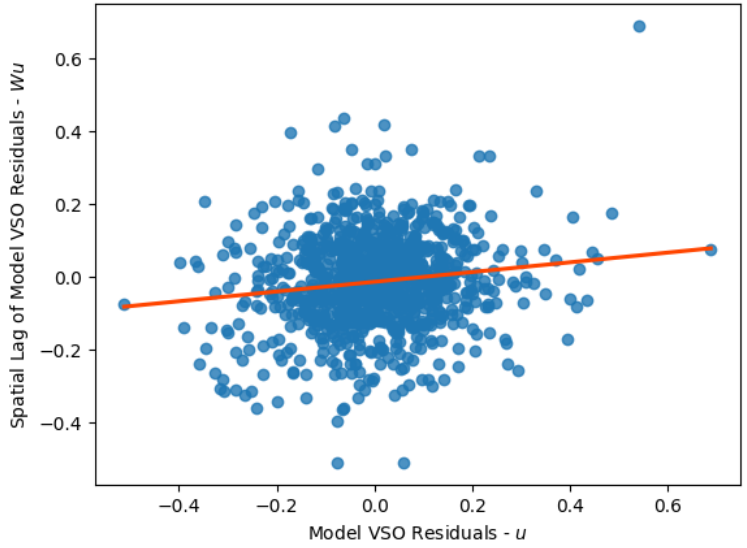




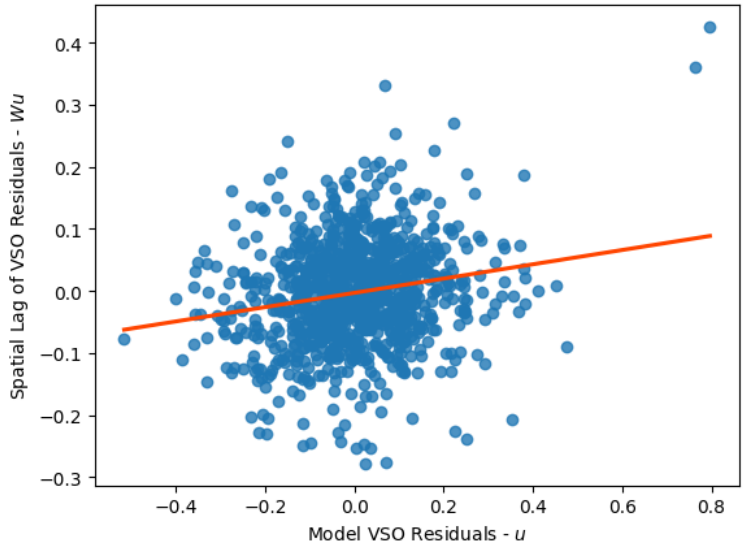


Violence and Sexual Offences

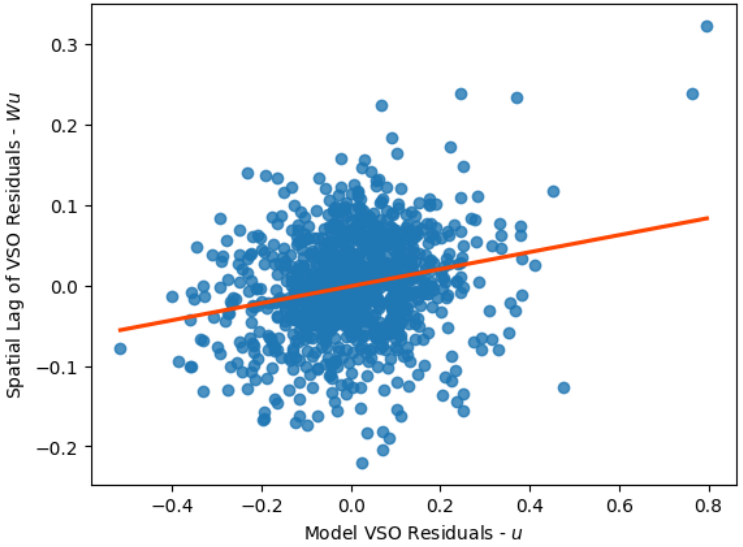
Variables rates



Variables encoded 1



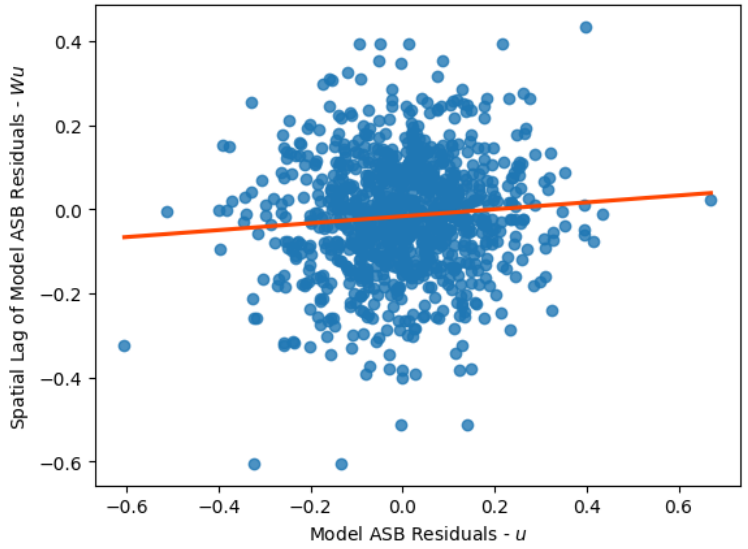
Variables encoded 2



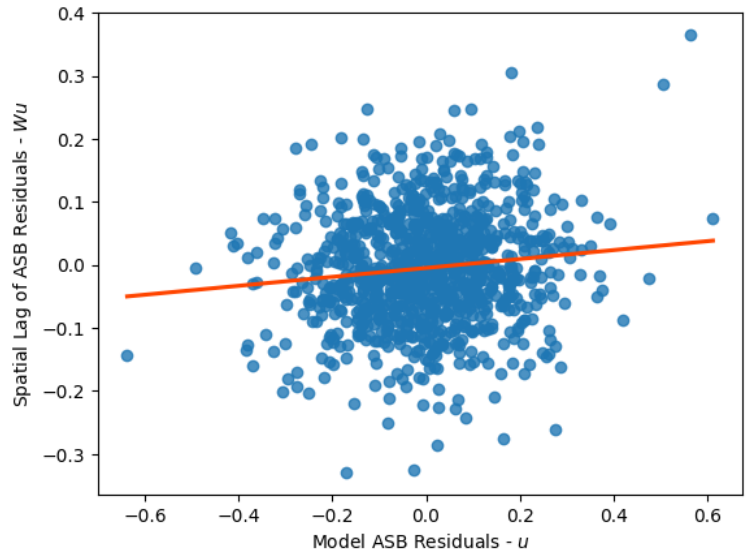
Are the models showing these spatial autocorrelations?

Prediction error at each MSOA and the prediction error at the MSOA nearest to it

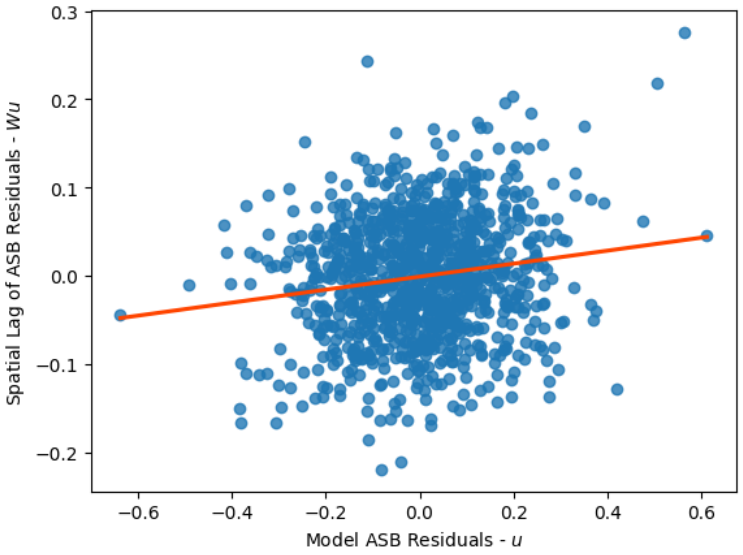
Anti-social Behaviour crimes



Variables rates

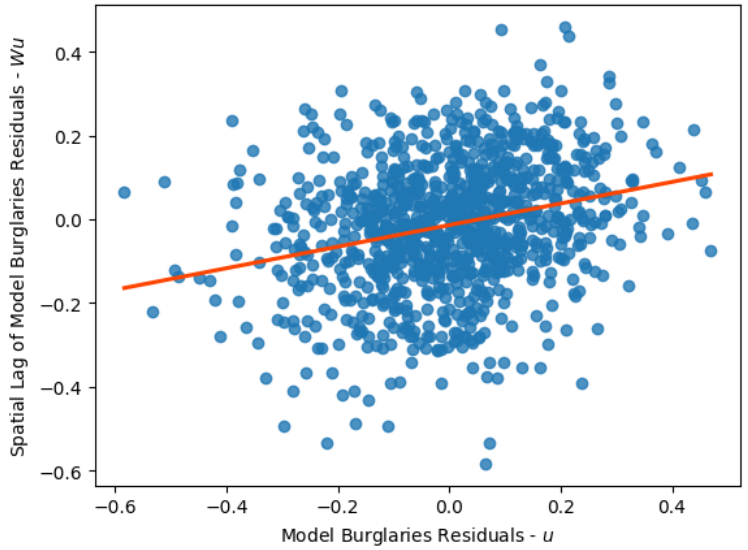


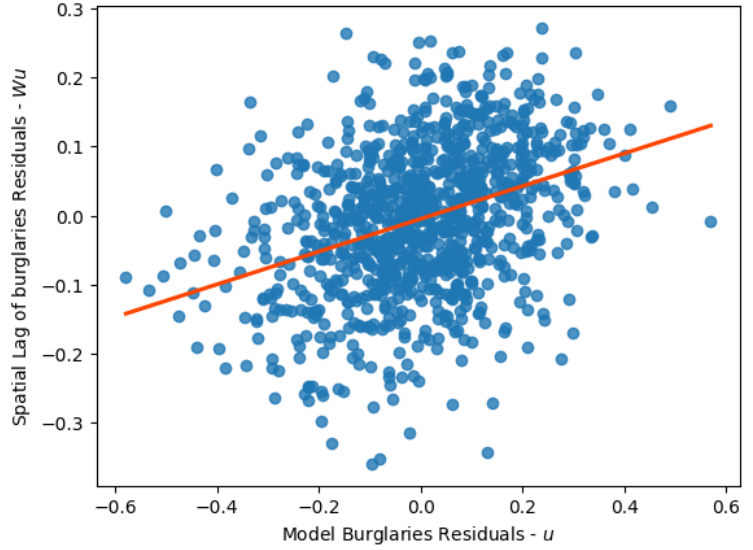
Variables encoded 1

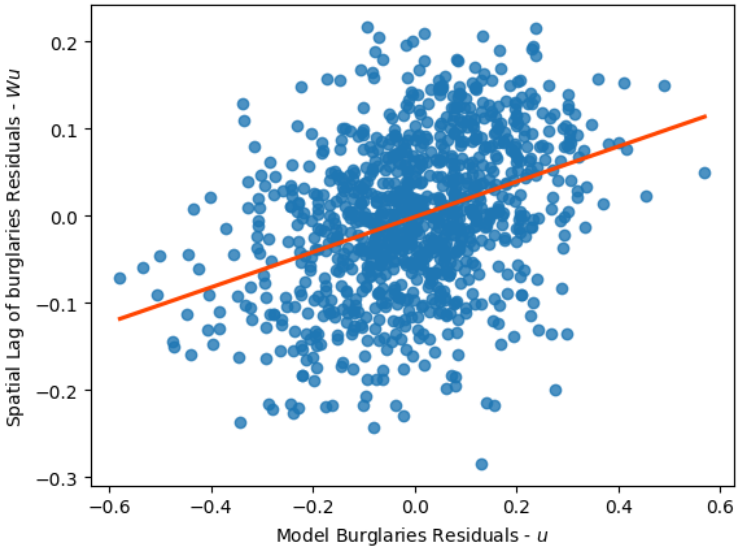


Variables encoded 2

Burglaries





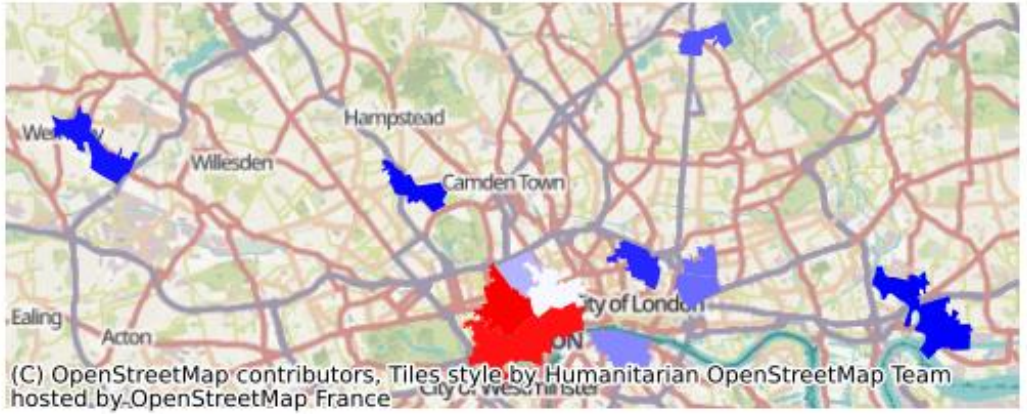


Where are the errors more significant?

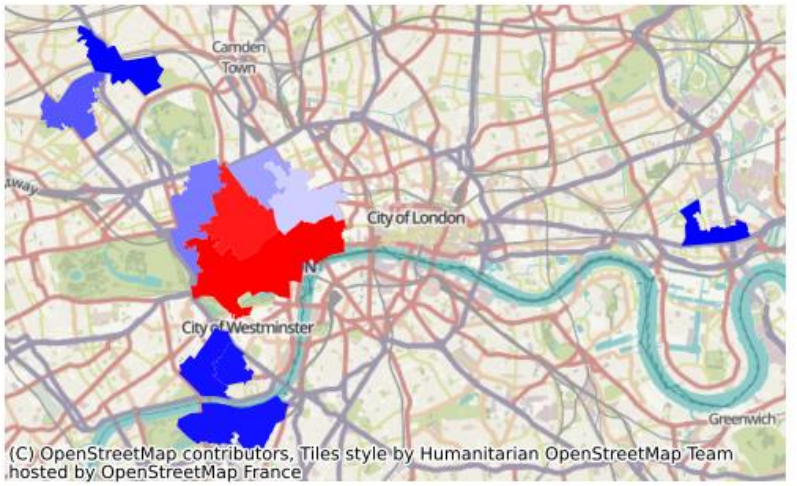
Areas where the model significantly under predicts

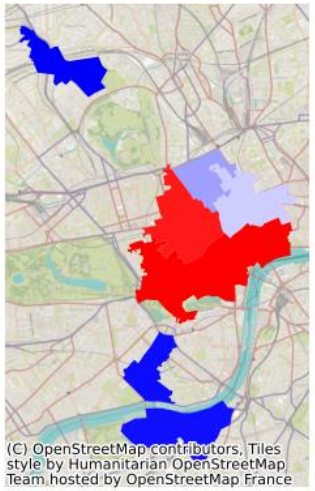
Local spatial autocorrelation of errors with 6 neighbours

Total Crimes



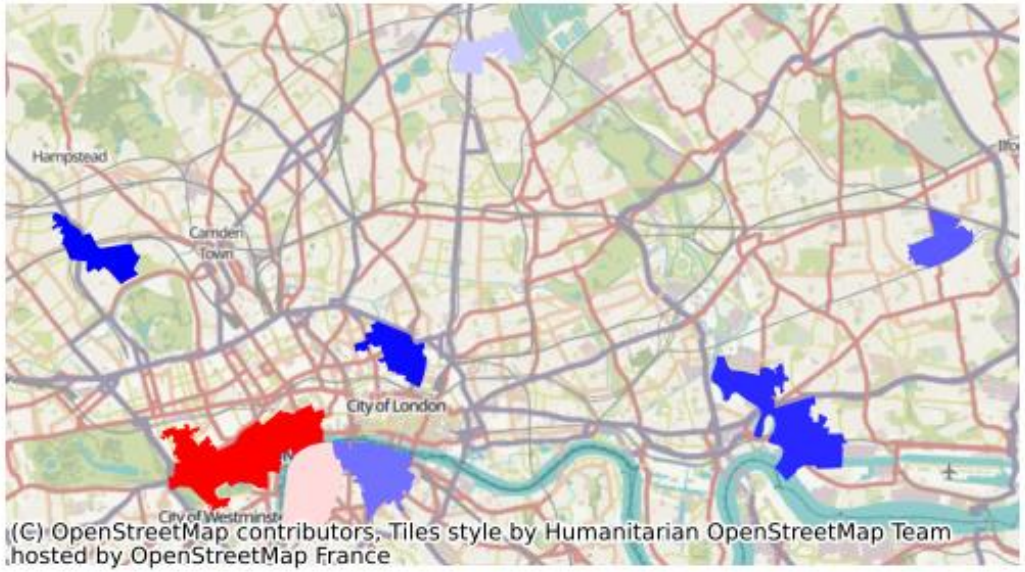
Variables as rates

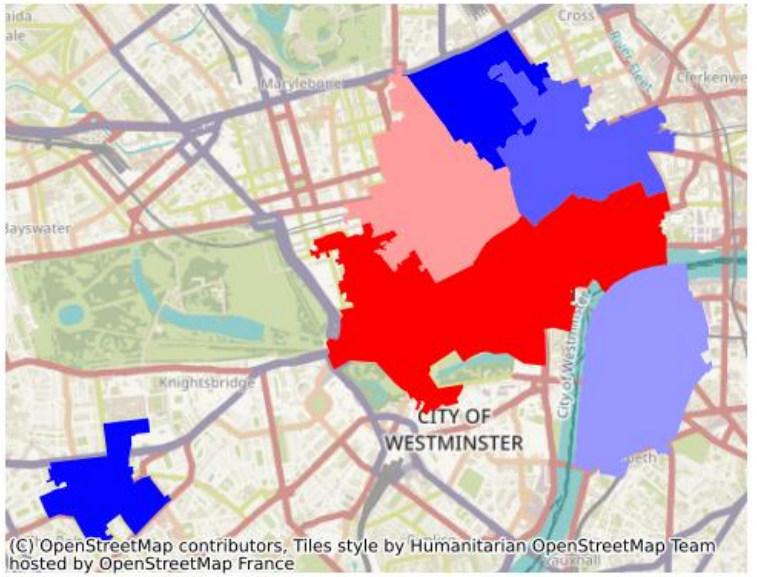


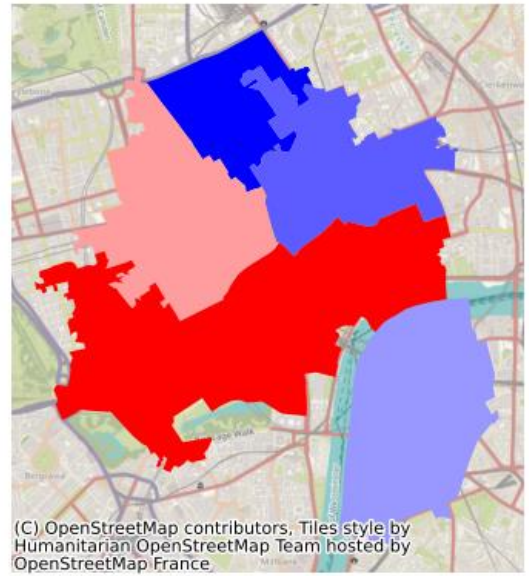


Variables encoded, 1 & 2

Violence and Sexual Offences





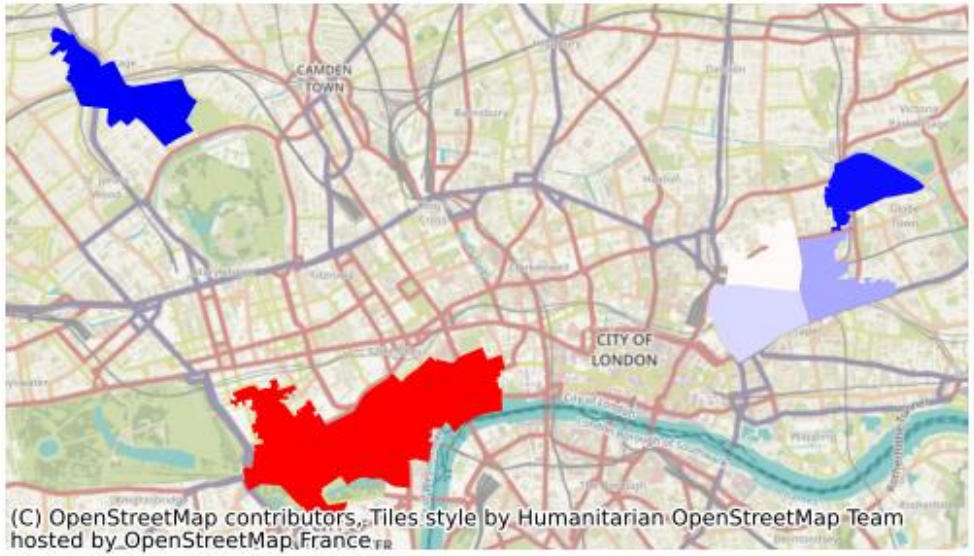


Where are the errors more significant?

Areas where the model significantly under predicts

Local spatial autocorrelation of errors with 6 neighbours

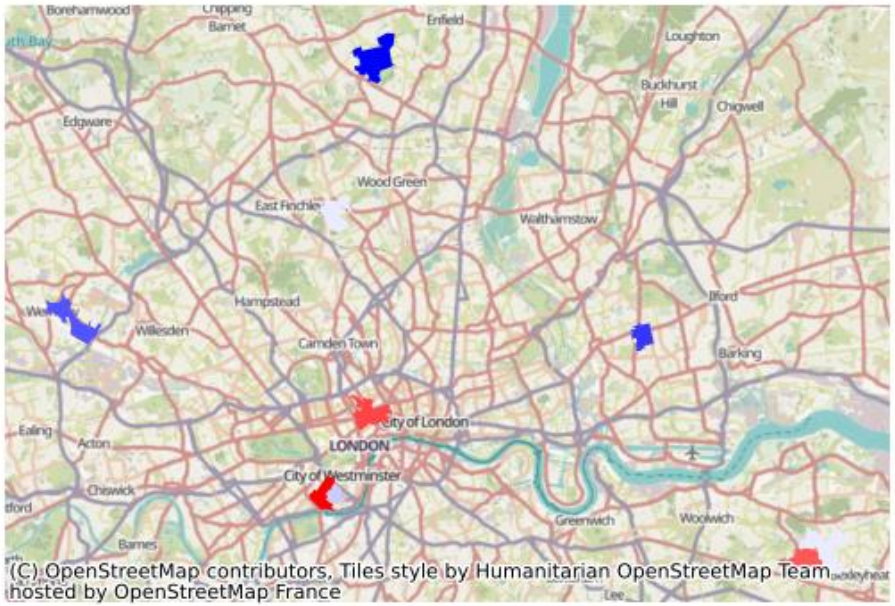
Anti-social behaviour crimes

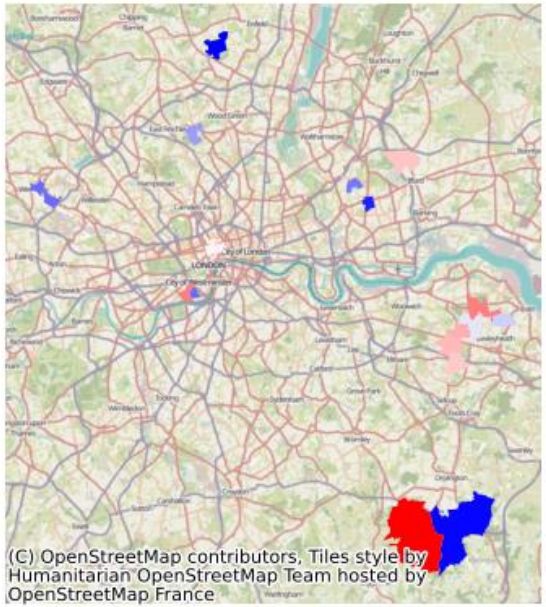


Variables as rates

Variables encoded, 1 & 2

Burglaries



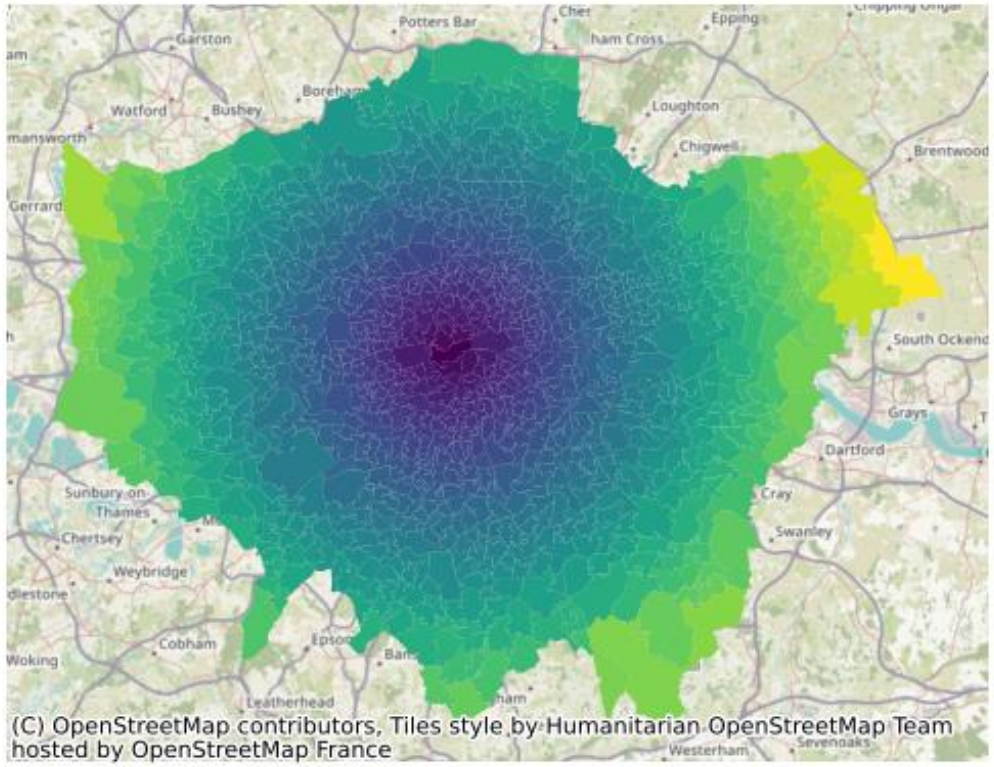


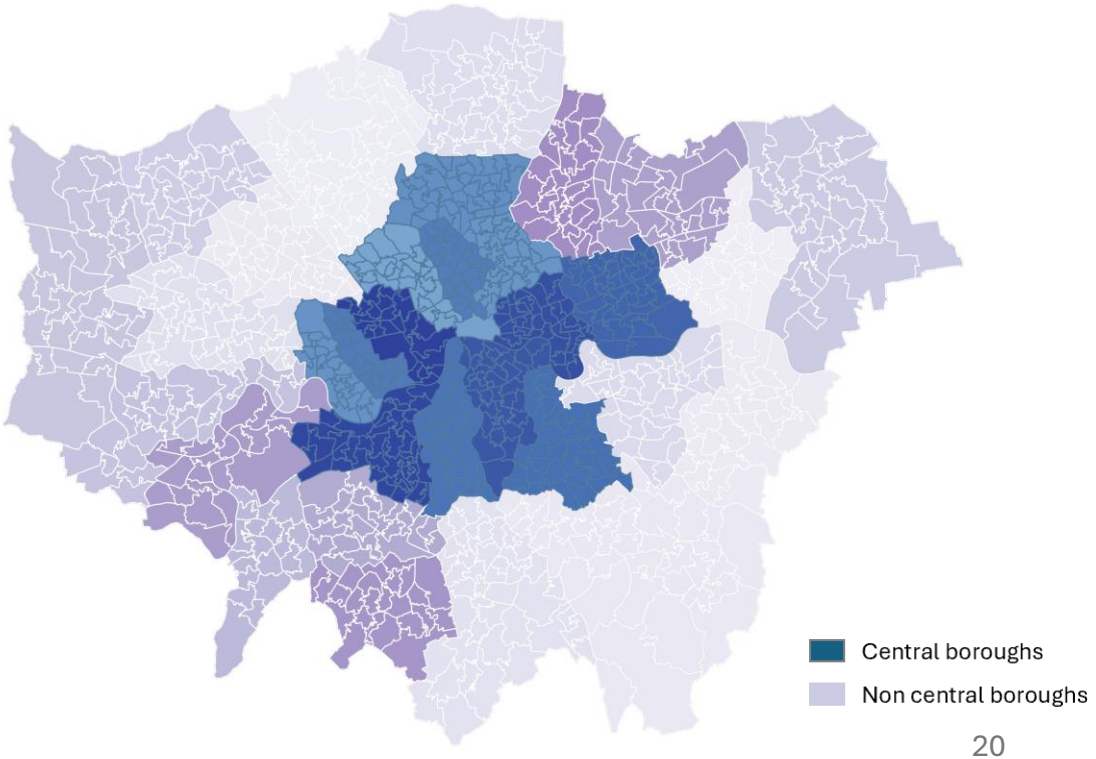
Does adding space improve the model?

1. Adding proximity to Westminster018 to the equation

2. Spatial fixed effects: Central or non central

3. Spatial regimes: 2 equations





Does adding space improve the model?

For Total crimes as target

1. Adding proximity to Westminster018 to the equation
2. Spatial fixed effects: Central or non central
3. Spatial regimes: 2 equations

	Proximity variable		**Spatial fixed effect**	
R-squared	0.52379		0.5266	
Adj. R-squared	**0.51752**		**0.5203**	
	Coeff.	**P-Value**	**Coeff.**	**P-Value**
CONSTANT 0	1.82128	0.00000	1.53421	0.00000
CONSTANT 1			1.50032	0.00000
Mixed tenures	0.04553	0.01935	0.05089	0.00851
Owned & Private	-0.05617	**0.00466**	-0.06696	**0.00086**
More private rent	0.01092	0.55689	0.00753	0.68548
More owned	-0.13852	**0.00000**	-0.15447	**0.00000**
Log10-vac	0.27159	**0.00000**	0.27833	**0.00000**
Mixed extremes	0.12589	**0.00002**	0.11518	**0.00012**
Most deprived	0.06648	**0.00022**	0.06721	**0.00018**
Mixed most deprived	0.05322	**0.00014**	0.05205	**0.00019**
Less deprived	-0.04129	**0.00519**	-0.04324	**0.00335**
Mainly flats	0.01360	0.36970	0.01932	0.19374
Mainly houses	-0.01498	0.30720	-0.01674	0.24386
Log10-price	0.00188	0.98030	0.07814	0.22081
Dist-Westminster018	-0.00085	0.60002		

Spatial regimes					
0.5363					
0.5244					
Non central		**Central**		**Chow test**	
Coeff.	**P-Value**	**Coeff.**	**P-Value**	**Statistic**	**P-value**
2.30505	0.00000	0.96966	**0.00715**	6.92029	**0.00852**
0.06628	0.07624	0.04644	0.05504	0.19883	0.65567
-0.04955	0.05668	-0.07738	0.03029	0.39764	0.52831
0.03914	0.13824	-0.00603	0.83314	1.34636	0.24592
-0.14629	**0.00000**	-0.00352	0.97022	2.06892	0.15033
0.28687	**0.00000**	0.23870	**0.00014**	0.39216	0.53117
-0.13373	0.37570	0.11239	**0.00168**	2.51948	0.11245
0.05926	0.01912	0.06624	0.01078	0.03726	0.84693
0.03606	0.04443	0.05827	0.01272	0.56981	0.45033
-0.04070	0.03136	-0.02704	0.29249	0.18370	0.66821
0.00363	0.90769	0.02773	0.11687	0.44915	0.50274
-0.02036	0.23697	-0.00951	0.73135	0.11081	0.73922
-0.12998	0.17278	0.22295	0.02212	6.71875	0.00954

Conclusions

Is crime correlated to housing conditions?

The correlations between tenure, accommodation type, house price and vacant dwellings are not strong enough independently. Deprivation shows a clearer correlation.

Can a combination of housing and deprivation conditions help explain crime hotspots?

It helps to explain in some measure, above the 50% for total crime, VSO and ASB

It doesn't work for burglaries

If we work with clusters to categorize the data, does it improve the model?

The model explains less, but some of the coefficients of the categories become more significant.

Is geography (space) relevant for the variables considered?

It is definitely an important factor in all the chosen variables.

Are the models showing signs of these spatial autocorrelations?

The models errors cluster in specific parts of the city

Does adding space improve the model?

It is not obvious, but it needs further exploration.

Further questions

- What would happen if we would encode (rank) the crime variables, the house price and the vacant dwellings?
  - What would the Spearman correlation say?
  - How the regression models with each group of variables would change?

- How would the regression models for each variable independently change if we add the spatial components (proximity, spatial fixed effect, spatial regimes)?

- How would the models improve if we add spatial dependence (the effects of the dependent and independent variables in the neighbours)?

- The effects of changing just the housing variables and not the structural one (deprivation) in the regression models