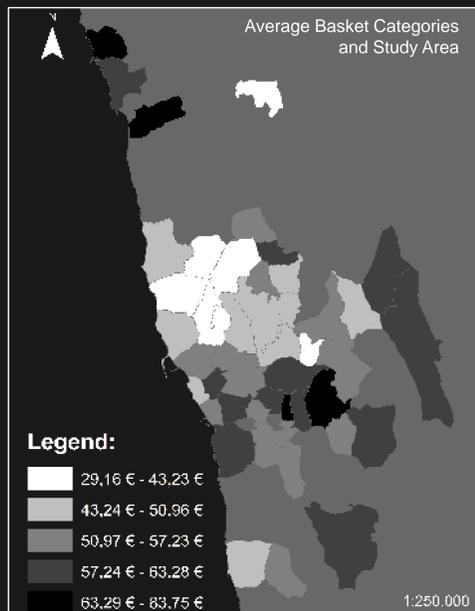


# AVERAGE BASKET

Modeling the "average basket" of a retail store using Geographically Weighted Regression



This study aimed at investigating a linear regression model to explain the variability of the "average basket" of a retail store located in the north of Portugal. 37 potential explanatory variables' data were collected.

Ordinary Least Squares (OLS) was applied to many alternative regression models, which were diagnosed for all OLS assumptions. The explanatory variables of the model that passed the diagnosis criteria were then included in a Geographically Weighted Regression (GWR) model, namely the "Number of family rented houses per km<sup>2</sup>", the "Direct distance to the point of sale" and the "Number of main residences in buildings with two dwellings per km<sup>2</sup>".

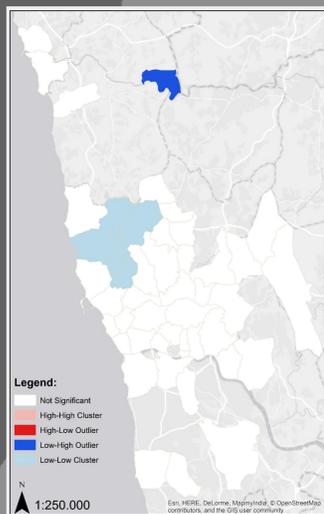
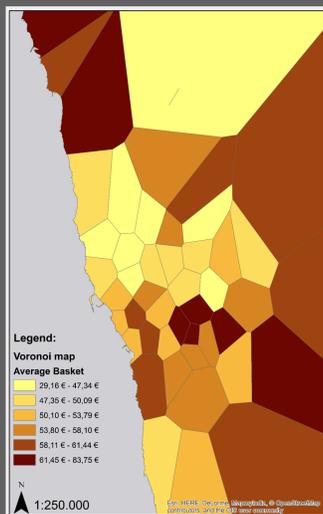
The GWR model proved to have a better fit to the data, having a higher Adjusted R-Squared (58%) and a smaller AICc (308.08) than the OLS model, without evidencing problems through the possible diagnostic methods.

The GWR approach provided additional insights about the regional variation of the explanatory variables, even though their coefficients can be unreliable because each local regression was based on few observations. In the extreme northwest of the study area, the average basket was better explained by the "Number of family rented houses per km<sup>2</sup>" and the "Direct distance to the point of sale". This region corresponds to a fishing and bathing area, thus customers living there have distinct characteristics.

It is important to point out that it is unclear what statistical tests can reliably diagnose GWR models' problems, and that the coefficients can be correlated even when there is no collinearity among explanatory variables.

## EXPLORATORY DATA ANALYSIS

- Summary Statistics
- Data posting
- Regional histogram
- Indicator maps
- Voronoi maps
- Local Moran's I statistics
- Global Moran's I statistic



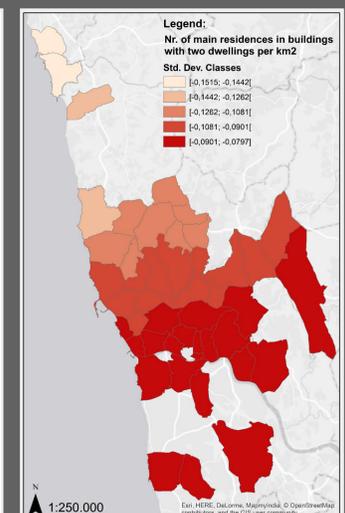
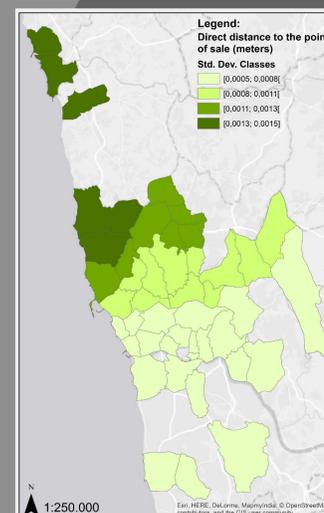
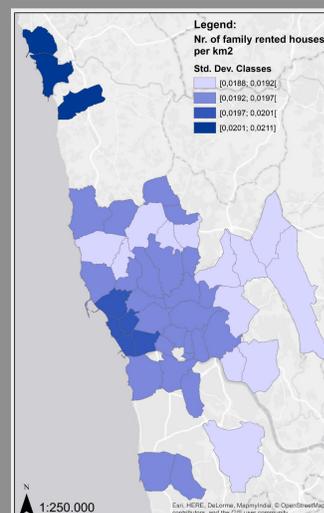
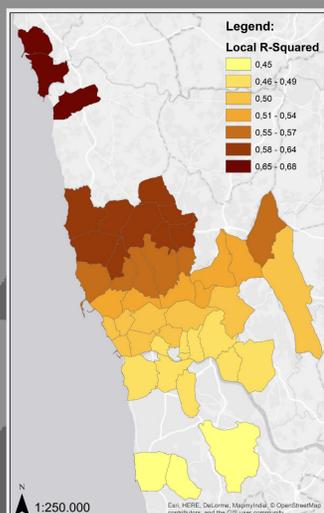
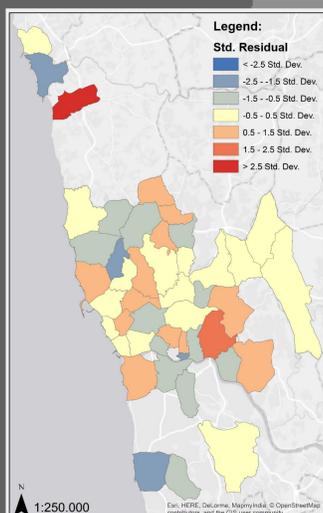
## OLS MODELS DIAGNOSTICS

- Multicollinearity
- Robust t-tests
- Jarque-Bera test
- Koenker test
- Global Moran's I statistic
- Joint F-test
- Adjusted R-Squared
- Adjusted Akaike's Information Criterion (AICc)

Diagnostics statistics	
Adjusted R-Squared	0.44
AICc	319.88
Joint F-Statistic	12.74*
Variance Inflation Factor (VIF)	2.12
Jarque-Bera statistic	6.68
Koenker statistic	5.22
Global Moran's I statistic	0.11**
Model parameters	
Intercept	44.5916
Nr. of family rented houses per km <sup>2</sup>	+0.0192*
Direct distance to the point of sale (meters)	+ 0.0009*
Nr. of main residences in buildings with two dwellings per km <sup>2</sup>	-0.0931*
* Significant at the 0.1% level	
** p-value	

## GWR MODEL DIAGNOSTICS

- Local multicollinearity
- Global Moran's I statistic
- Local R-Squared
- Adjusted Akaike's Information Criterion (AICc)



## References

Anderson, P. (2004) *Business Economics and Finance with MATLAB, GIS, and Simulation Models*. Chapman & Hall/CRC Press LLC, Florida (USA), Ch.13.

Cheng, E. W. L., & Li, H. (2004) Exploring quantitative methods for project location selection. *Building and Environment*, 39(12), 1467–1476.

Griffith, D. (2008) Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, 40(11), 2751–2769.

Hernandez, T. (2007) Enhancing retail location decision support: The development and application of geovisualization. *Journal of Retailing and Consumer Services*, 14(4), 249–258.

Krause-Traudes, M., Scheider, S., Rüping, S., Meßner, H. (2008) *Spatial data mining for retail sales forecasting*. 11th AGILE International Conference on Geographic Information Science 2008, University of Girona, Spain.

Mendes, A. (2005) *Modelação de vendas de novas superfícies comerciais*. PhD Thesis, Technical University of Lisbon, Portugal.

Owen, S. H., & Daskin, M. S. (1998) Strategic facility location: A review. *European Journal of Operational Research*, 111, 423–447.

Páez, A., Farber, S., & Wheeler, D. (2011) A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43(12), 2992–3010.

Reigadinha, T., Godinho, P., Dias, J. (2015) Portuguese food retailers – Exploring three classic theories of retail location. *Journal of Retailing and Consumer Services*, 34, 102–116.

Wheeler, D., & Tiefelsdorf, M. (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2), 161–187.