Urban and socio-economic correlates of property prices in Dublin's area

Damien Dupré
Business School
Dublin City University
Glasnevin, Dublin 09
damien.dupre@dcu.ie

***Abstract*—Understanding the characteristics of the housing market is essential for both sellers and buyers. However, the housing market is influenced by multiple factors. In this paper, the urban and socio-economic structure of an area is used to predict the price of 10387 properties sold in 2018 in the city of Dublin. More precisely, the direct distance from each property to 160 urban features taken from OpenStreetMap is calculated, and an extreme gradient boosting linear regression performed. Using these features, the model explains 45% of the housing price variance. The most important features in this model are the proximity to an embassy and to a grassland. In addition, the results of a population census from 2016 are also used to correlate with the price of properties. From this census, 48 features are used as the input of a gradient boost linear regression model. In all, the socio-economic features are explaining 43% of the housing price variance as well. The density of individuals reporting that they are not providing unpaid personal help for a friend or family member as well as individuals reporting that they have no religion are the most important socio-economic feratures. By taking into account either urban or socio-economic features, it is possible to accurately estimate housing prices and to predict their evolution.**

***Index Terms*—Property Price; Housing Market; Feature Analysis; Machine Learning; Geocoding**

I. Introduction

First time property acquisition is an important achievement in individuals' lifetime. It provides not only housing security but also the feeling of being a landowner. However, access to the status of landowner is a challenge for low to middle income earners as it corresponds to significant spending which will potentially have budget impacts for years (Savage et al. 2014). For this reason, understanding property price markets as well as factors driving price is essential to identify investment opportunities.

Beside their inherent characteristics such as size, design and materials (Bourassa, Cantoni, and Hoesli 2007; Liu 2013), property valuation is dictated by multiple external factors explaining their spatial autocorrelation (Basu and Thibodeau 1998). Some of these factors are macroeconomic (Antonakakis and Floros 2016), related to the evolution of population general wealth while others are related to the characteristics of the neighborhood (Dubin 1998). These characteristics include the urban and socio economic structure of the neighborhood (Goodman and Thibodeau 1998). However, it is difficult to evaluate the influence of such external factors on property valuation (Clapp, Kim, and Gelfand 2002). By using machine learning algorithms, prediction of property prices has become more and more accurate. Techniques used usually include artificial neural networks due to the volatility of the market (Limsombunchai 2004; Feng and Jones 2015; Selim 2009). Indeed, artificial neural network techniques can quickly adjust to volatility changes, therefore incorporating sharp increases or decreases in their predictions (Aragonés, Blanco, and Estévez 2007). However, the model obtained from most artificial neural networks are difficult to interpret (Rudin 2019).

In this paper, an alternative machine learning algorithm is used to identify urban and socio-economic correlates of property price evolution in the area of Dublin, Ireland in 2018. Due to the high number of potential features, an eXtreme Gradient Boosting (XGBoost) regression model is believed to obtain the best results with low parameter customization and high speed (Chen and Guestrin 2016). Contrary to most artificial neural network techniques, XGBoost is an interpretation-focused method which automatically provides estimates of feature importance from a trained predictive model.

II. Method

Since 2010, under the Irish Property Services (Regulation) Act, all individuals acquiring a property in Ireland have to declare it to Property Services Regulatory Authority (PSRA). They must provide details such as the date of sale, the price and the address of all residential properties purchased in Ireland as declared to the Revenue Commissioners for stamp duty purposes (Property Services Regulatory Authority 2020). Data is filed electronically by persons doing the conveyancing of the property on behalf of the purchaser, and it must be noted that errors may occur when the data is being filed. Because of the evolution of urban planning and the evolution of socio-economic features measured by the 2016 census, only properties sold in 2018 were processed to avoid potential temporal incoherence. In 2018, the Property Price Register references 18665 properties sold in Dublin's area. Among these properties, those corresponding to apartments have been removed from the original dataset in order to evaluate only houses (10.3% of the properties). Finally, some properties have been unsuccessfully geocoded because of mistakes on their address or because they have not been found (38.0% of the properties). These properties have also been removed from the original dataset to obtain a final dataset of 10387 properties.

In order to evaluate the spacial distribution of the property sold, property addresses were geocoded (i.e., converted to latitude and longitude) using the Nominatim API (Hoffmann 2020) to OpenStreetMap data. OpenStreetMap is a collaborative project which aims to create and provide access to free editable maps of the world (Haklay and Weber 2008).

To estimate the geographical distribution of the price density, multiple methods such as multiple regression (McCluskey et al. 2000) or Bayesian smoothing (Clapp, Kim, and Gelfand 2002) have been employed. While these methods are efficient in a non-restricted space, they have limitations when they are used in coastal areas. This problem is known as "finite area smoothing" and occurs when predictions from a model are approximated across geographical barriers, such as irregular coastal shapes, which can lead to poorly predicted values (Ramsay 2002). To deal with the influence of geographical boundaries on price estimation, a generalized additive model using soap film smoothing has been used (Wood, Bravington, and Hedley 2008).

A. *Distance to urban features*

OpenStreetMap combines information about more than 400 urban features including road information and building information to categorize features such as amenities, leisure or tourism structures (OpenStreetMap 2020a). Among these features, only 160 are available or relevant to the Dublin Area. Urban features are obtained using the R package osmdata (Padgham et al. 2017) which queries the Overpass API (OpenStreetMap 2020b). The distance between each property and the closest point corresponds to each of the 160 urban features.

B. *Density of socio-economic features*

In addition to the distance to every urban feature, each property sold in 2018 was related to 48 socio-economic features of the small area including the property, as measured by the Irish census 2016. The results of Irish 2016 census consultation is accessible through both the Central Statistics Office (Central Statistics Office 2020) and the All-Island Research Observatory (All-Island Research Observatory 2020) websites. The data obtained can be mapped over small area boundaries which are fractions of Irish Electoral Division map. The social features extracted are corresponding to population information, religion, carers and health. Economic features correspond to the characteristics of each small area; including the proportion of housing types, number of rooms, occupancy and tenure per small area. Each property is then associated to the value corresponding to its small area. For anonymisation purposes, the results from small areas having less than five respondents to the census were converted to a proportion of five respondents.

C. *eXtreme Gradient Boosting*

Lastly, an eXtreme Gradient Boosting (XGBoost) model was calculated to examine the extent to which property prices rely on urban and on socio-economic features. The XGBoost algorithm optimizes the prediction accuracy by performing boosting iterations (set to 100 iterations as the best trade-off between accuracy and speed), thereby minimizing the root mean squared error when the model's objective is a continuous output (Friedman 2001). The R package "xgboost" (v.1.1.1.1) is used to implement the XGBoost algorithm (Chen and Guestrin 2016). With this version, two types of booster are available: tree based and linear regression based. Different types of learning task can also be defined according to the distribution of the output. In the case of a continuous output, they include regression with squared loss, regression with squared log loss, gamma regression with log-link, and tweedie regression with log-link.

The original dataset has been randomly split to 80% for training and 20% for testing the models. The prediction accuracy is estimated using the coefficient of determination (R^2), the Root Mean Square Error ($RMSE$) and the Mean Absolute Error (MAE) of the predicted property price values of the test dataset.

Both tree and linear regression based boosting model are used for every of the continuous output learning task in order to compare their accuracy with urban and socio-economic features. As no specific weight or bias in the models have been hypothesized, parameters of the XGBoost algorithms have been kept as default values. Next, the most accurate model is used to identify the features that are contributing most to the prediction of property prices. This contribution (also called "importance") is a measure of the improvement in accuracy brought by a feature. It is based on the absolute magnitude of feature's coefficient.

III. Results

In 2018, the average price of a sold property in Dublin was €394,093 with a standard deviation of €170,219 (Figure 1).

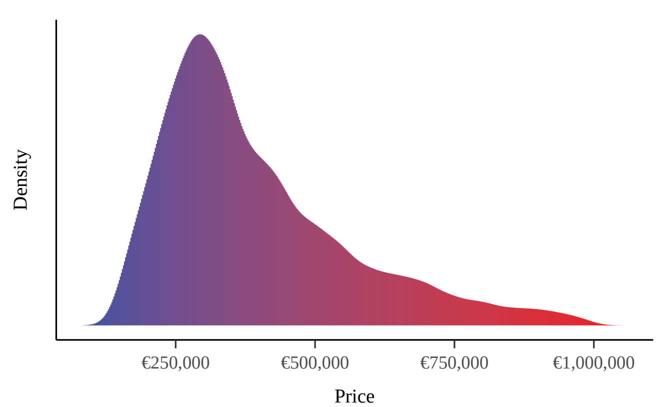

Fig. 1. Overall distribution of the property prices density in the city of Dublin in 2018

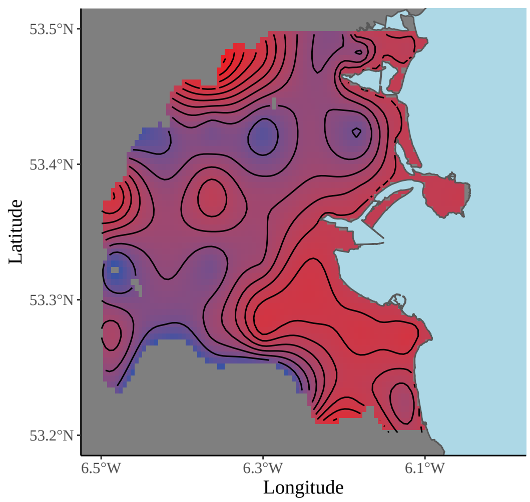

Fig. 2. Geographic distribution of the property prices density in the city of Dublin in 2018 using a generalized additive model with soap film smooth parameter to take into account the influence of coastal boundaries. Estimates outside of the 95% Confidence Interval were removed.

The density of housing prices reveals a unimodal distribution of property prices. In addition, the geographical distribution of property prices indicates a higher valuation in the South-West of the city as well as on the coast line (Figure 2).

A. *Correlates with urban features*

In order to compare the possibilities of the different type of parameters in the XGBoost models, all combinations of tree and linear regression based boosters with all possible output objectives (i.e squared error, squared log error, gamma, and tweedie) are displayed Table I.

TABLE I
ACCURACY OF XGBOOST MODELS USING URBAN FEATURES ACCORDING THE TYPE OF GRADIENT BOOSTING METHOD AND OUTPUT OBJECTIVE.

Objective	$RMSE$	R^2	MAE
Tree based			
squarederror	€127,783	0.44	€85,115
squaredlogerror	€426,989	-	€391,670
gamma	€129,052	0.43	€87,039
tweedie	€126,626	0.45	€85,073
Linear regression based			
squarederror	€150,847	0.21	€111,786
squaredlogerror	€406,939	0.00	€369,708
gamma	€177,238	0.07	€133,973
tweedie	€171,480	0.08	€129,138

Note:
Missing value ares due to incompatibilities between data and methods.

The model comparison reveals that tree based boosting models are more accurate than linear regression based model. In addition, within the possible objectives of the tree based boosting models, the tweedie regression with log-link appears to obtain the highest R^2 value as well as the lowest $RMSE$ and MAE prediction errors.

Using the XGBoost algorithm with a tree based booster and a tweedie regression with log-link learning objective, the 160 urban features are explaining 44.9% of the property price variance ($F(1, 2076) = 1,690.80$, $p < .001$) with a $RMSE$ of €126,626 and a MAE of €85,073.

The overall correlation of the predicted property prices with urban features is shown in Figure 3A. It can be noticed that property prices situated at the low and the high end of the range are the most difficult to predict (Figure 3B). A possible explanation is the absence of distinctive and recurrent patterns in urban features for these properties. However, the prices higher than €300,000 are potentially driving down the prediction accuracy due to outliers.

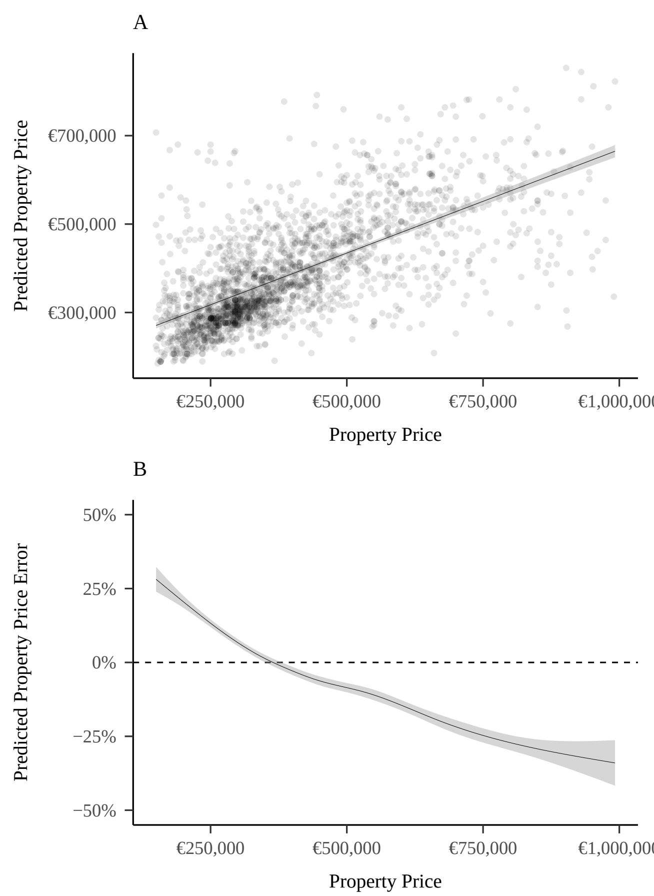

Fig. 3. Property price prediction accuracy (A) and prediction error (B) using urban features with XGBoost.

By analyzing their importance (Table II), the most relevant geographical features to predict housing prices are the distance to an embassy (15.6%) and the distance to natural grasslands such as parks and gardens (4.8%).

TABLE II
TOP 10 MOST IMPORTANT URBAN FEATURES CONTRIBUTING TO THE PROPERTY PRICE PREDICTION.

Feature Category	Feature Type	Importance
amenity	embassy	15.6%
natural	grassland	4.8%
power	line	2.4%
route	bus	2.4%
amenity	bar	2.0%
boundary	political	1.9%
place	island	1.6%
boundary	administrative	1.5%
highway	residential	1.4%
barrier	wall	1.4%

B. Correlates with socio-economic features

Similarly to urban features, the accuracy of different XGBoost models using socio-economic features to predict property prices are displayed Table III.

TABLE III
ACCURACY OF XGBOOST MODELS USING SOCIO-ECONOMIC FEATURES ACCORDING THE TYPE OF GRADIENT BOOSTING METHOD AND OUTPUT OBJECTIVE.

Objective	$RMSE$	R^2	MAE
Tree based			
squarederror	€131,270	0.42	€86,892
squaredlogerror	€426,989	-	€391,670
gamma	€129,221	0.43	€86,949
tweedie	€130,927	0.42	€86,707
Linear regression based			
squarederror	€146,963	0.25	€108,151
squaredlogerror	€406,213	0.02	€369,023
gamma	€183,008	0.08	€136,445
tweedie	€174,731	0.10	€130,483

Note:
Missing value ares due to incompatibilities between data and methods.

The comparison of accuracy metrics reveals that tree based boosting models obtain better results than linear regression based boosting models. Within these tree based boosting models, the gamma regression with log-link is the best model as it obtains the highest R^2 and lowest $RMSE$ values.

Using the XGBoost algorithm with a tree based booster and a gamma regression with log-link learning objective, the 48 socio-economic features explain 42.8% of the property price variance ($F(1, 2076) = 1,554.37$, $p < .001$) with a $RMSE$ of €129,221 and a MAE of €86,949.

The overall correlation of the predicted property prices with socio-economic features is shown in Figure 4A. Similar to the model used for urban features, the model based on socio-economic features reveals that prices lower than €300,000 lead to the highest errors in over-valuation while prices higher than €300,000 lead to systematic under-valuation errors (Figure 4B).

According to the analysis of the relative importance of socio-economic features (Table IV), the most important socio-economic features is the proportion individual not providing regular unpaid personal help for a friend or family member with a long-term illness, health problem or disability (32.8%). It also appears that the proportion of large properties (i.e., houses with eight or more rooms) in the small area containing the property (19.1%) as well as the proportion of people reporting that they have no religion (4.2%), influence the model.

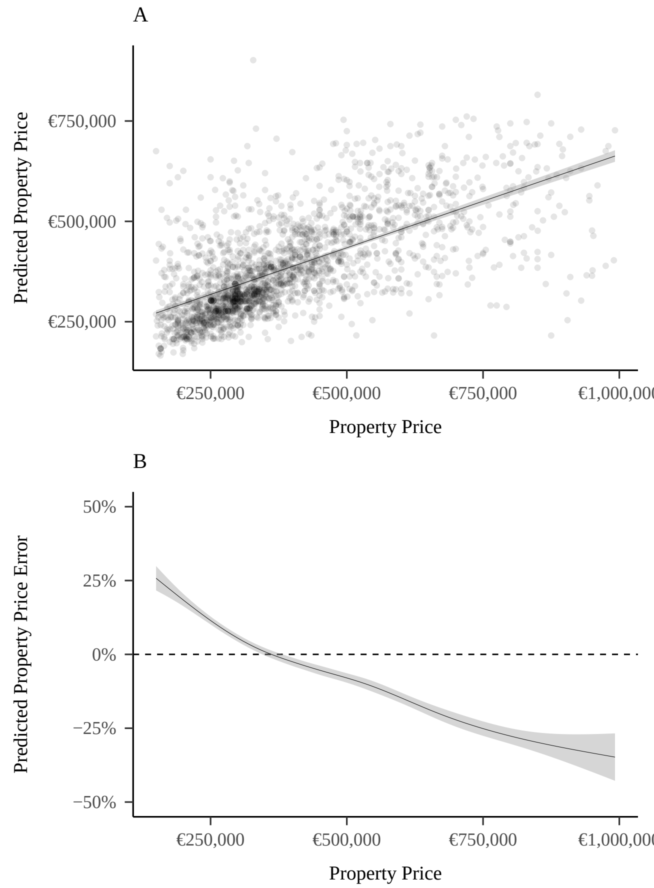

Fig. 4. Property price prediction accuracy (A) and prediction error (B) using socio-economic features with XGBoost.

TABLE IV
TOP 10 MOST IMPORTANT SOCIO-ECONOMIC FEATURES CONTRIBUTING TO THE PROPERTY PRICE PREDICTION.

Feature Category	Feature Type	Importance
carers	Provides No Care	32.8%
housing rooms	8 or more Rooms	19.1%
religion	No Religion	4.2%
population	Age 0 - 14	3.2%
housing rooms	6 Rooms	2.1%
carers	1-19 hours unpaid PW	1.9%
housing tenure	Social Rented	1.8%
general health	Very Good	1.7%
religion	Other Catholic	1.6%
general health	Good	1.5%

IV. Discussion

While Dublin is a very specific city due to its low population density spread on a large surface, its housing market is one of the most expensive in Europe. The main reason is the structure of the market, mainly made of residential properties and few apartment buildings. In this study, the prices of 10387 properties sold in 2018 in the area of Dublin from the Property Services Regulatory Authority dataset were analyzed in order to identify the potential factors correlating with these prices. The enforcement of a housing price registry has been shown as being an essential tool to regulate the housing market (Kitchin 2013). However, a major limitation of this study is the absence of characteristics for the property itself. Indeed, prices are mainly determined by factors such as the size of the property or the number of rooms. In addition, while some features appear to be more important than others in their contribution to property prices, these results are not causal. Thus, a feature's importance can be either the cause or the consequence of property prices in a neighborhood

Despite the absence of property characteristics, urban and socio-economic features were used to identify spatial correlates to the housing prices. Results revealed that proximity to embassies and grasslands is a driver of property prices. Embassies are in general located in the most expensive areas of cities which are also the most aesthetically pleasing and the most secure part of cities. In addition, the density of embassies and properties allocated to embassy staff reduces the density of property available on the market and drives their prices up. The distance to a park or a natural area is also a very important factor for property prices. Results also revealed that the density of individual not providing regular unpaid personal help in the area is the most important socio-economic feature to predict property prices. This feature is related to the multiple factors such as the average age of the household or the economic status of the household as they are related to general health (Smith 1999). Large properties in the area and the proportion of individuals reporting having no religion are also relatively important. Again, the size of the property sold was not included in the dataset; however, the higher the density of large properties in the area, the higher the probability for the property to be a large property. Finally, the relation with the expression of religious belief, or more precisely its absence, is a very interesting feature. Again, the relationship between religion and income is not confirmed by the literature. Whereas some studies reveal a positive correlation between religion and income (Guiso, Sapienza, and Zingales 2003; Elgin et al. 2013) others reveals no evidence of such relationship (De La O and Rodden 2008). Here, it appears that prices of properties are higher in areas with a high density of people indicating that they have no religion.

V. Conclusion

The evolution of housing prices is a pressing issue in most European capitals and specially in Dublin. Given their significant increase, properties are less and less affordable for individuals. By preforming a feature analysis with urban and socio-economic features, it is possible to evaluate and predict the potential price of a property. Indeed features such as the presence of embassies or parks are criteria that influence significantly the price of properties. Similarly, the characteristics of inhabitants in the area such as religion, health and age is correlated to the evolution of housing prices. These results allow an understanding of why some areas have higher prices than others which is relevant information not only for real estate agents in charge of property valuations but also for buyers in order to estimate the real value of a property.

VI. Acknowledgement

The authors would like to thank the developers of the following R packages used to process, analyze, display and report data: R (Version 4.0.2; R Core Team 2020) and the R-packages *dplyr* (Version 1.0.2; Wickham, François, et al. 2020), *forcats* (Version 0.5.0; Wickham 2020), *ggplot2* (Version 3.3.2; Wickham 2016), *here* (Version 0.1; Müller 2017), *kableExtra* (Version 1.1.0; Zhu 2019), *latex2exp* (Version 0.4.0; Meschiari 2015), *lubridate* (Version 1.7.9; Grolemund and Wickham 2011), *lwgeom* (Version 0.2.5; Pebesma 2020), *magrittr* (Version 1.5; Bache and Wickham 2014), *mgcv* (Version 1.8.31; Wood 2011, 2004, 2003; Wood et al. 2016), *nlme* (Version 3.1.148; Pinheiro et al. 2020), *OpenStreetMap* (Version 0.3.4; Fellows and JMapViewer library by Jan Peter Stotz 2019), *osmdata* (Version 0.1.3; Padgham et al. 2017), *papaja* (Version 0.1.0.9997; Aust and Barth 2020), *patchwork* (Version 1.0.1; Pedersen 2019), *purrr* (Version 0.3.4; Henry and Wickham 2020), *readr* (Version 1.3.1; Wickham, Hester, and Francois 2018), *sf* (Version 0.9.5; Pebesma 2018), *stringr* (Version 1.4.0; Wickham 2019), *tibble* (Version 3.0.3; Müller and Wickham 2020), *tidyr* (Version 1.1.1; Wickham and Henry 2020), *tidyverse* (Version 1.3.0; Wickham, Averick, et al. 2019), *tmap* (Version 3.1; Tennekes 2018), *xgboost* (Version 1.1.1.1; Chen et al. 2020), and *yardstick* (Version 0.0.7; Kuhn and Vaughan 2020).

VII. Data availability

The R code and relevant data for statistical computing are available at https://github.com/damien-dupre/DSAA_2020.

References

All-Island Research Observatory. 2020. "AIRO Data Store." June 1, 2020. http://airo.maynoothuniversity.ie/datastore.

Antonakakis, Nikolaos, and Christos Floros. 2016. "Dynamic Interdependencies Among the Housing Market, Stock Market, Policy Uncertainty and the Macroeconomy in the United Kingdom." *International Review of Financial Analysis* 44: 111–22.

Aragonés, José R, Carlos Blanco, and Pablo García Estévez. 2007. "Neural Network Volatility Forecasts." *Intelligent Systems in Accounting, Finance & Management: International Journal* 15 (3-4): 107–21.

Aust, Frederik, and Marius Barth. 2020. *papaja: Create APA Manuscripts with R Markdown*. https://github.com/crsh/papaja.

Bache, Stefan Milton, and Hadley Wickham. 2014. *Magrittr: A Forward-Pipe Operator for R*. https://CRAN.R-project.org/package=magrittr.

Basu, Sabyasachi, and Thomas G Thibodeau. 1998. "Analysis of Spatial Autocorrelation in House Prices." *The Journal of Real Estate Finance and Economics* 17 (1): 61–85.

Bourassa, Steven C, Eva Cantoni, and Martin Hoesli. 2007. "Spatial Dependence, Housing Submarkets, and House Price Prediction." *The Journal of Real Estate Finance and Economics* 35 (2): 143–60.

Central Statistics Office. 2020. "Census 2016 Small Area Population Statistics." June 1, 2020. https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–94.

Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. 2020. *Xgboost: Extreme Gradient Boosting*. https://CRAN.R-project.org/package=xgboost.

Clapp, John M, Hyon–Jung Kim, and Alan E Gelfand. 2002. "Predicting Spatial Patterns of House Prices Using Lpr and Bayesian Smoothing." *Real Estate Economics* 30 (4): 505–32.

De La O, Ana L, and Jonathan A Rodden. 2008. "Does Religion Distract the Poor? Income and Issue Voting Around the World." *Comparative Political Studies* 41 (4-5): 437–76.

Dubin, Robin A. 1998. "Predicting House Prices Using Multiple Listings Data." *The Journal of Real Estate Finance and Economics* 17 (1): 35–59.

Elgin, Ceyhun, Turkmen Goksel, Mehmet Y Gurdal, and Cuneyt Orman. 2013. "Religion, Income Inequality, and the Size of the Government." *Economic Modelling* 30: 225–34.

Fellows, Ian, and using the JMapViewer library by Jan Peter Stotz. 2019. *OpenStreetMap: Access to Open Street Map Raster Images*. https://CRAN.R-project.org/package=OpenStreetMap.

Feng, Yingyu, and Kelvyn Jones. 2015. "Comparing Multilevel Modelling and Artificial Neural Networks in House Price Prediction." In *2015 2nd Ieee International Conference on Spatial Data Mining and Geographical Knowledge Services (Icsdm)*, 108–14. IEEE.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics*, 1189–1232.

Goodman, Allen C, and Thomas G Thibodeau. 1998. "Housing Market Segmentation." *Journal of Housing Economics* 7 (2): 121–43.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. http://www.jstatsoft.org/v40/i03/.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2003. "People's Opium? Religion and Economic Attitudes." *Journal of Monetary Economics* 50 (1): 225–82.

Haklay, Mordechai, and Patrick Weber. 2008. "Openstreetmap: User-Generated Street Maps." *IEEE Pervasive Computing* 7 (4): 12–18.

Henry, Lionel, and Hadley Wickham. 2020. *Purrr: Functional Programming Tools*. https://CRAN.R-project.org/package=purrr.

Hoffmann, Sarah. 2020. "Nominatim: Open-Source Geocoding with Openstreetmap Data." June 1, 2020. https://nominatim.org/.

Kitchin, Rob. 2013. "Making Informed Decisions on Future Housing Policy." *Housing Ireland: A Journal for Irish Housing Professionals*, no. 3: 36–41.

Kuhn, Max, and Davis Vaughan. 2020. *Yardstick: Tidy Characterizations of Model Performance*. https://CRAN.R-project.org/package=yardstick.

Limsombunchai, Visit. 2004. "House Price Prediction: Hedonic Price Model Vs. Artificial Neural Network." In *New Zealand Agricultural and Resource Economics Society Conference*, 25–26.

Liu, Xiaolong. 2013. "Spatial and Temporal Dependence in House Price Prediction." *The Journal of Real Estate Finance and Economics* 47 (2): 341–69.

McCluskey, William J, William G Deddis, Ian G Lamont, and Richard A Borst. 2000. "The Application of Surface Generated Interpolation Models for the Prediction of Residential Property Values." *Journal of Property Investment & Finance*.

Meschiari, Stefano. 2015. *Latex2exp: Use Latex Expressions in Plots*. https://CRAN.R-project.org/package=latex2exp.

Müller, Kirill. 2017. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2020. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

OpenStreetMap. 2020a. "Map Features (Wiki)." June 1, 2020. https://wiki.openstreetmap.org/wiki/Map_Features.

———. 2020b. "Overpass API." June 1, 2020. https://wiki.openstreetmap.org/wiki/Overpass_API.

Padgham, Mark, Bob Rudis, Robin Lovelace, and Maëlle Salmon. 2017. "Osmdata." *The Journal of Open Source Software* 2 (14). https://doi.org/10.21105/joss.00305.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

———. 2020. *Lwgeom: Bindings to Selected 'Liblwgeom' Functions for Simple Features*. https://CRAN.R-project.org/package=lwgeom.

Pedersen, Thomas Lin. 2019. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2020. *nlme: Linear and Nonlinear Mixed Effects Models*. https://CRAN.R-project.org/package=nlme.

Property Services Regulatory Authority. 2020. "Residential Property Price Register." June 1, 2020. https://propertypriceregister.ie.

Ramsay, Tim. 2002. "Spline Smoothing over Difficult Regions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (2): 307–19.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rudin, Cynthia. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1 (5): 206–15.

Savage, Michael, James Barlow, Peter Dickens, and Tom Fielding. 2014. *Property Bureaucracy & Culture*. Routledge.

Selim, Hasan. 2009. "Determinants of House Prices in Turkey: Hedonic Regression Versus Artificial Neural Network." *Expert Systems with Applications* 36 (2): 2843–52.

Smith, James P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status." *Journal of Economic Perspectives* 13 (2): 145–66.

Tennekes, Martijn. 2018. "tmap: Thematic Maps in R." *Journal of Statistical Software* 84 (6): 1–39. https://doi.org/10.18637/jss.v084.i06.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr.

———. 2020. *Forcats: Tools for Working with Categorical Variables (Factors)*. https://CRAN.R-project.org/package=

———. 2004. "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models."

forcats.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Lionel Henry. 2020. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Wickham, Hadley, Jim Hester, and Romain Francois. 2018. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wood, Simon N, Mark V Bravington, and Sharon L Hedley. 2008. "Soap Film Smoothing." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (5): 931–55.

Wood, S. N. 2003. "Thin-Plate Regression Splines." *Journal of the Royal Statistical Society (B)* 65 (1): 95–114.
Journal of the American Statistical Association 99 (467): 673–86.

———. 2011. "Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models." *Journal of the Royal Statistical Society (B)* 73 (1): 3–36.

Wood, S. N., N., Pya, and B. S"afken. 2016. "Smoothing Parameter and Model Selection for General Smooth Models (with Discussion)." *Journal of the American Statistical Association* 111: 1548–75.

Zhu, Hao. 2019. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.