

# \$patial: Interactive Tool to Model HDB Resale Prices using Geographically Weighted Regression

Guided by Associate Professor of Information Systems: Dr. Kam Tin Seong

Shawn CHUA Jun Yong  
School of Information Systems  
Singapore Management University (SMU)  
shaw.n.chua.2018@sis.smu.edu.sg

Olivia GOO Yu Ya  
School Of Social Sciences  
Singapore Management University (SMU)  
olivia.goo.2018@socsc.smu.edu.sg

## ABSTRACT

Valuing housing prices has seen a surge in popularity over the years amongst policy planners and economist due to the significant impact that properties have on the economy and society. The current hedonic pricing models used in the market fails to take into account the effect of local spatial features on the housing prices which however there has been increasing interest and studies carried out on Geographically Weighted Regression (GWR) model, a more precise regression model to stimulate the spatial distribution of housing prices. There are several GWR models available which poses a challenge to casual users who would like to conduct a simple and understandable analysis of the spatial distribution of housing prices. Therefore to address this issue, we developed \$patial, an interactive and accessible application with an user friendly interface to aid economists and policy planners to effortlessly explore how variations of spatial features such as number of shopping malls surrounding the housing estate in different locations affects the price of the housing flats.

## 1. INTRODUCTION

With rapid urbanization and population growth, the demand for housing around the world is increasing. The volume of housing transactions globally has seen a boom and global housing markets have been steadily climbing up the price index (IMF Global Housing Watch, n.d.). Home ownership being one of the universal signs of success and prosperity due to it being a long term investment places housing prices as an emerging subject of interest amongst policy planners, economists and home owners.

The existing hedonic housing pricing models are linear and does not account for spatial effects on the price of housing flats as such there is emerging interest in using GWR to accurately model housing prices (Chin et al., n.d.). In order to solve the current gaps in the existing pricing models, for the purpose of our project we will be making use of R Shiny

an open-source tool to build user-friendly GWR models for analysis to show the correlation between spatial attributes and resale flat prices.

We have selected Singapore's housing market as our case study in this project as Singapore is known globally as the "tiny red dot" with limited amount of land space thus geographically constrained and albeit that, there has been an increase in construction of public housing in Singapore as a result of the rising number of population which makes it relevant to explore the effects of spatial variations on the housing prices.

This research paper documents the research and methods used in the process of implementing the GWR model in our application in relation to public housing prices. Section 1 provides a general introduction of the paper and motivations and objectives of the research. Section 2 provides a review of related papers, section 3 details the review of the analytical techniques used for visualizing and analyzing data. The results and future improvements are discussed in section 4 and the paper is concluded by highlighting the future direction of the research in section 5.

## 2. REVIEW OF RELATED WORK

The development of \$patial is inspired by two published papers.

The first paper being EzModel (Chin et al., n.d.) where two models, GWR and mixed (semiparametric) GWR is used in the analysis of the datasets. The use of mixed GWR is effective in allowing users to decide which variables they would like to set as global or local variables to create an optimal model for comparison to the GWR model however it seems that there has not been a formal way to identify the two types of variables. Additionally, the use of isoline mapping via inverse weighted interpolation allows for more information to be shown whereby it highlights the difference in coefficient estimates to reflect the different effect of the number of spatial features on the resale price of the flats. Overall, the application is informative and clear allowing us to draw reference from it. The second paper is the Simple Geo-Spatial Analysis using R-shiny (Li Junyi et al., n.d.) where

## 3. DATA COLLECTION & DATA PREPARATION

### 3.1 Data Collection

The 2 main types of data used in our application is:

1. HDB Flat Resale Prices Data
2. Data on features provided to user which is used as independent variables in the GWR model

The fixed spatial features provided by \$patial are:

- (hr /> 1. Locations of MRT stations
2. Primary School Locations
3. Secondary School Locations
4. Community Centre Locations
5. Supermarkets Locations
6. Sports Facilities Locations
7. Preschools Locations
8. Hawkers Locations
9. Shopping Malls Locations

With the exception of:

- Shopping Malls Location – Wikipedia
- MRT and LRT Location – mytransport.sg

The rest of the datasets are obtained from data.gov.sg.

### 3.2 Data Preparation

The datasets identified above came in different file types containing differing data types, some processing must be done for all the imported data to be integrated into the analysis.

The school dataset we obtained contains all schools ranging from primary to tertiary education (specifically junior colleges), therefore we decided to extract data only for primary and secondary schools and keep them as separate datasets for further analysis. Primary and Secondary schools are being focused on for the purpose of this study as we felt that, homeowners will be more concerned for the distance of primary and secondary schools seeing those that attend them are of the younger age group and parents may not want them to travel far. For the datasets containing data on shopping malls and resale flat information, they do not contain geographical coordinate's data which is needed for our analysis, therefore we have tapped on the existing Geocode tool provided by Google Sheets to obtain coordinate data (in longitude and latitude).

After ensuring that all datasets contain coordinate data. We have to convert those that are in coordinates (longitude and latitude) which are calculated in degrees into coordinates (X, Y) which are calculated in meters. This is to ensure proper integration for further analysis when proximity around resale flats need to be computed and distance metrics uses distance in meters. Data preparation is mainly done in Rmarkdown then exported into new csv files for usage in building of Shiny application.

## 4. METHODS

The following section reveals the techniques and algorithms used in the process of designing the application.

### 4.1 Application Architecture

The application was developed using Shiny, an R program package. R shiny is a simple package that is used to build

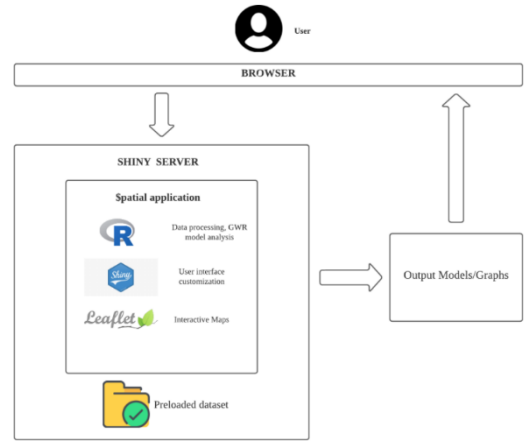


Figure 1: Application Architecture

interactive web applications and dashboards. It runs on a Shiny server hosted by Shinyapps.io, the datasets mentioned in Section 3.1 are imported and stored in the server. At the backend, the CSV and Shapefile datasets are cleaned and used for geocoding, projection and GWR. Whenever the application runs the datasets are automatically loaded for use. The interactive maps featured in the application calls on the Leaflet package for it to be displayed.