

UNIVERSITY OF THE WITWATERSRAND
Johannesburg



FACULTY OF SCIENCE

**Identifying the determinants of housing prices in the City of Johannesburg for the year
2012: A spatial analysis approach**

Audrey Mmatebogo Mbatha

Supervisor: Dr Cletah Shoko

School of Geography, Archaeology and Environmental Studies

**A research report submitted to the Faculty of Science, University of the Witwatersrand,
Johannesburg, in partial fulfilment of the requirements for the degree of Master of
Science.**

Year 2021

Declaration

I, Audrey Mmatebogo Mbatha declare that this research report is my own research work. This research report has been submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



(Signature of candidate)

23rd day of June 2021 at Pretoria.

Abstract

An attempt has been made to identify the determinants of housing prices in the City of Johannesburg, Gauteng Province, South Africa. A dataset of 2012 housing prices at sub-place level and variables were categorised into housing demand and housing supply. The regressions; Ordinary Least Squares (OLS), spatial lag and error models and geographical detector technique were employed. The results obtained from Moran's I test suggested that there is spatial dependence present in the data. From the properties that were captured in 2012, the maps indicated that the full title ownership type is the most favoured one in the municipality. The hot spot analyses suggested a significant clustering of housing prices around Sandton and Houghton area in the region B and the low-low spatial cluster was observed in region D, Soweto area and region G the Lenasia area. From spatial regression model, the results showed negative effects exerted by unemployment, number of sales, and renting whereas income, mortgages loans, city services, transfer duty and property ages are positively and significantly associated with the housing prices. The study analysed the direction and strength of association between housing prices and their potential determinants. The influence of the factors on housing prices were ranked by power of determinant (PD) value whereby transfer duty value was 0.7658 and household income was 0.6499. The geographical detector technique results revealed that transfer duty and household income exerted strong influence in relation to housing prices followed by renting and unemployment while city services shown to have weak positive influence with a value of 0.0911.

Dedication

I would like to dedicate this report to my loving family for their support through the study years and hope that this inspires them all.

Acknowledgements

I would like to thank God for the strength and courage to study this programme despite the challenges that one encountered which provided frustration throughout.

I would like to thank Dr Cletah Shoko, my supervisor for the guidance and support provided, her guidance ensured that the work is concluded.

Furthermore I would like to thank my family, my husband Lazarus Mbatha, my children Thando and Nhlanhla Mbatha for all the support. The emotional support carried me through until I completed the work.

I would like to thank my colleague, Sabelo Mnguni from Department of Rural Development and Land Reform for providing the moral studying support.

Lastly I would also like to thank the officials from Centre for Affordable Housing Finance in Africa (CAHF), AFRIGIS, Statistics South Africa and City of Johannesburg municipality who assisted me with the provision of the required research data.

Table of Contents

Declaration.....	ii
Abstract.....	iii
Dedication	iv
Acknowledgements.....	v
List of Figures	3
List of Tables	4
1 Introduction	5
1.1 Problem statement	6
1.2 Rationale of the study.....	7
1.3 Aim	8
1.4 Specific objectives.....	8
1.5 Structure of the report.....	8
2 Literature Review introduction.....	9
2.1 Demand and supply indicator analysis for house prices.....	9
2.2 Spatial and non-spatial factors influencing house prices	10
2.2.1 Macro-economic factors	10
2.2.2 Construction activities	14
2.2.3 Land supply	15
2.2.4 Real Estates' characteristics.....	15
2.2.5 Spatial determinants of housing prices	16
2.2.6 Spatial effects and housing prices dynamics	17
2.3 The role of spatial analysis and GIS approaches in analysing housing prices.....	18
3 Materials and methods.....	21
3.1 Study area	21
3.2 Explanatory variables and their expected effects.....	23
3.3 Data acquisition	25
3.4 Methods.....	28
3.4.1 Explanatory data analysis.....	28
3.4.2 Measuring spatial effects.....	28
3.4.3 Geographical detector technique	30
4 The results.....	32
4.1 Spatial distribution of housing prices in CoJ	32
4.2 Spatial distribution of different housing tenures in CoJ	33

4.3	Descriptive statistics	38
4.3.1	Summary statistics and plots	38
4.3.2	Mapping of the selected variables.....	42
4.4	Modelling Spatial dependence and clustering.....	50
4.4.1	Moran's I statistic test.....	50
4.4.2	Geary's C statistic test.....	51
4.4.3	Local Moran's I.....	52
4.5	Model selection using OLS regression	53
4.5.1	Model fitting	54
4.5.2	Model selection.....	54
4.5.3	Model diagnostics	55
4.6	Fitting the Spatial linear regression	57
4.6.1	Testing for type of spatial dependence present in the data.....	57
4.6.2	Spatial lag model.....	58
4.6.3	Spatial error model	59
4.7	Geographical detector technique	59
5	Discussion.....	61
5.1	Mapping of the housing prices and housing tenures	61
5.2	Measuring spatial auto correlation.....	62
5.3	Factors affecting housing prices from regression models	62
5.4	The estimated impact of factors.....	63
6	Conclusion and recommendation.....	65
7	Limitations of the study	65
8	References	66

List of Figures

Figure 1: Study area, CoJ with city regions (Data Source: City of Johannesburg).....	22
Figure 2: 2012 Average housing transaction prices by market segments in CoJ.....	32
Figure 3: Spatial distribution of the 2012 residential property homeowner tenure and prices by market segments in CoJ: (a) Freehold, (b) Estates and (c) Sectional titles.....	36
Figure 4: Distribution of (a) 2011 rental housing and (b) households mortgage not paid	37
Figure 5: Histogram showing the distribution of property ages.....	39
Figure 6: Histogram of raw housing prices and unemployment observations.....	40
Figure 7: Scatter plot showing relationship between housing prices with average income and transfer duty	40
Figure 8: Histograms of log transformed housing price observations.....	41
Figure 9: Scatterplot of transformed housing prices versus unemployment	42
Figure 10: Distribution of the 2011 unemployment in CoJ.....	43
Figure 11: Distribution of 2011 average household income of CoJ	44
Figure 12: Distribution of 2012 new constructions in CoJ	45
Figure 13: Distribution of 2012 transfer duty of CoJ	46
Figure 14: Distribution of 2012 properties sold in CoJ	47
Figure 15: 2011 Number of years the properties in existence in CoJ	48
Figure 16: Distribution of the total number of the 2012 different city services	49
Figure 17: Moran's I scatterplot for transformed housing prices.....	51
Figure 18: Spatial clustering and spatial outliers' map of housing prices at sub place level.....	52
Figure 19: Histogram of model residuals.....	55
Figure 20: QQ-plots and fitted values.....	56
Figure 21: The power of determinant for the factors guiding the housing price effect.....	59

List of Tables

Table 1: Description of the explanatory variables 27

Table 2: Statistical summary of the variables 38

Table 3: The results of spatial correlation test 50

Table 4: Stepwise procedure results..... 54

Table 5: Estimation results of regressions 60

CHAPTER 1 : GENERAL INTRODUCTION

1 Introduction

Globally governments and individuals have been concerned about the changing housing prices because these changes have influence on the national socio-economic factors (Selim, 2009). The capital gains that are expected from the investments made in housing has an effect on housing prices whereby the demand for housing increases. This high demand in housing in turn triggers high rise and fall in housing prices (Selim, 2009). Housing prices can be influenced by different factors for example macro-economic factors, community characteristics, spatial differences and different amenities (Kim and Park, 2005). In many industrialised countries across the globe their residential real estate industry encountered huge recurring differences in terms of housing prices and housing volumes. Real estate cycles are mostly described by a sudden increase whereby a fall or a crash will follow (Nneji et al., 2013).

In United Kingdom (UK) for example property market during the late 1980s, the financial market liberalisation resulted in a housing prices booming up, however subsequent to interest rates increasing, residential prices suffered a sudden fall during the early 1990s (Nneji et al., 2013). The real estate market in 2008 went through the Great Recession period, which lasted nearly four years. In the United States of America housing prices reached their peak during the early 2006, and began to drop in 2006 till 2007, and then in 2012 the housing prices reached the new lows (Standard and Poor Dow Jones Indices). The biggest price decline was reported in the Case–Shiller 2008 home price index and this was the biggest drop that was ever recorded in history (Mantell, 2009). This biggest price drop resulted in the bursting of the real estate bubble. This bursting resulted in credit crisis which caused the 2007–2009 recession for the United States of America (USA) (Holt, 2009).

The real estate industry in South Africa forms the biggest part of the country's property market with Gauteng province as home to the largest regional housing market, accounting for 34% of the total national housing stock (Centre for Affordable Housing Finance in Africa (CAHF), 2010). Gauteng province has developed more as compared to other provinces, with the total of 1.4 million immigrants between 2001 and 2016 (Statistics South Africa (StatsSA), 2017). The movement of large numbers of migrants to Gauteng (people looking for better

opportunities) along with natural population growth have had an influence on the province's growing and changing population. This province experiences an increase in migration of people who have a variety of housing needs and gains in population numbers.

The 2018 World Bank Doing Business report (Doing Business, 2018), ranked South Africa as the 107th out of 190 countries globally, in terms of how registration of a property is simple and easy to do. The property market in the country is well established with the best cadastral system that affords buyers, sellers and investors for a procedural protection (CAHF, 2010). However, home owners from 2008 to 2018 were negatively impacted because the real house prices in South Africa dropped by 4.8% (Delmendo, 2019). The residential market's growth in South Africa stayed quiet and the real estate specialists are expecting that there would be no movement until quarter four of the year 2020 when supply is anticipated to make even with demand (Shaw, 2019).

1.1 Problem statement

Housing is the main asset of most households' wealth. In South Africa, most people that are working their biggest part of their riches is the worth of their property (Franken, 2007). Therefore home ownership is valuable to most households because this is one of the ways in which they acquire wealth (Friedman et al., 2013).

House prices keep changing with inflation over time, most important when the house is in a neighbourhood that most people are after it (ABSA, 2015). The fluctuations in the housing prices, are concerns to most households and this affects the buying power. These changes, either increasing or decreasing, impact the individual households, property investors, businesses, mortgage loan lenders and policy makers. Rising housing prices increase the riches of property owners but decreases the opportunities for the households to afford to buy their properties (Ge and Williams, 2015). The middle-income households are unable to live in well-located urban areas due to most families still has to fund their housing independently, with savings or non-mortgage credit (Tsatsaronis et al., 2004).

The real estate industry in South Africa is driven by the following factors; economic trends, consumer confidence and household finances (Arnold et al., 2017). These factors have an impact on the affordability of housing and the accessibility of and demand for mortgage loans. It is therefore imperative to comprehend the behaviour of property prices. For all to understand the behaviour of property prices there is a need to better understand the factors that influence increase/decrease in housing prices.

1.2 Rationale of the study

In the beginning of the 2000s, house prices in South Africa had observed uninterrupted four years of continued increases in housing price (Franken, 2007). Following the increases in the house prices this boosted activities redirected investment concentration in the real estate market. It seemed that the flourishing property prices was led and backed by economic factors, such as increase in income, high affordability, low inflation and low interest rates (Franken, 2007). Property prices, property sales and new property construction all significantly increased amid declining inventory and record-low mortgage interest rates.

According to de Guzman (2018), 2000 to 2006 period was South Africa's housing boom but, in the first quarter of 2008, the housing boom came to a standstill because of financial crisis felt globally. South Africa's housing prices dropped to a negative economic expansion in all market sectors in the fourth quarter of 2008 (CAHF, 2010). From the year 2009, the greatest growth was observed in the inexpensive markets (such as low household income and low cost areas). In 2012 the sinking housing market began recovering and the house prices rose higher than previously after the decline trends that were experienced as from 2008 (de Guzman, 2018).

CAHF and South African Cities Network (SACN) (2014) reports that Johannesburg recorded the greatest number of sales, highest percentage of transactions and properties bonded, and second highest sales price than other metropolitan municipalities in 2012. In consideration of the above facts on the housing market performance, this study focused on the year 2012 in order to find the housing prices determining factors in City of Johannesburg municipality since

it's housing market is more active, contains the most number of households and growing compared to other municipalities in the country (CAHF and SACN, 2014).

The study focussed in obtaining an understanding in the strengths and characteristics of factors affecting the housing prices, in order to apply appropriate approaches for future housing needs, which will then enable the property to improve households' economic strength and wellbeing.

1.3 Aim

The aim of this study was to identify the potential housing price determinants in the residential property market in the City of Johannesburg for the year 2012.

1.4 Specific objectives

The objectives of this research were the following:

1. To map how the 2012 housing prices, property ownership, renting and mortgage loan payments are distributed in the City of Johannesburg.
2. To test for spatial correlation and clustering in the housing prices.
3. To identify significant variables associated with housing prices using the spatial regression models.
4. To identify factors affecting housing prices and their relations.
5. To estimate the magnitude of the impact of the factors affecting the housing prices.

1.5 Structure of the report

In addition to introduction section, this research report consists of 7 sections which are set out as follows: Section 1 introduces the study and provides for the aim, objectives and research questions and Section 2 provides the review of the literature on relevant studies followed by Section 3 which describes the methodology adopted, Section 4 presents the results and Section 5 discusses the results of the analysis; Section 6 concludes the report whereas Section 7 presents limitations.

CHAPTER 2 LITERATURE REVIEW

2 Literature Review introduction

The following literature review discusses different factors that affect housing prices as observed from previous studies. The review focuses on factors such as demand and supply of housing, spatial factors, non-spatial factors, and emphasizes on the role spatial effects play and then concludes with the support which statistical analysis, spatial analysis and geographical information system (GIS) technology can provide in studying the housing prices determinants.

2.1 Demand and supply indicator analysis for house prices

It is a fundamental economic principle that when demand exceeds supply, prices tend to rise (Kramer, 2019) and if there is a huge supply of a good or service but there is not sufficient demand for it, the price will fall (Hall, 2018). Changes that have taken place in house prices are due to the economic reasons that cause changes in demand or supply (Franken, 2007). In the ordinary market circumstances, when the prices of housing increase this decreases the demand for housing and consequently a negative relationship between housing demand and house prices is experienced (Peek and Wilcox, 1991).

The economic variables that has effect on the demand for housing consist of an increase in net income, tax system and interest rates (Tsatsaronis and Zhu, 2004). The demand and supply variables employ an effect on the increase of house prices although the demand variables employ larger effect than the supply (Franken, 2007); and there is a positive relation between the supply of housing stock and house prices, which means that when the house price increases then housing supply increases (Tse, Ho and Ganesan, 1999).

Tsatsaronis and Zhu (2004) stated that inflation and nominal interest rates are most important drivers of housing prices as they explain the most variations in housing prices, and shocking outcome was that household income had a small influence over housing price. The association between market demand and market supply for housing goods establishes the market price where the demand equates supply.

2.2 Spatial and non-spatial factors influencing house prices

Residential property prices differ between countries, regions or city because of the differences macroeconomic factors and the performance thereof. The overall financial performance of a country or municipality influences the price stability on property markets (de Bruyne et al., 2013). The house price differences are observed on the housing market because of differences in macro-economic factors, quality of housing and living, land supply, construction activities and geographical location (de Bruyne et al., 2013).

2.2.1 Macro-economic factors

The macro-economic variables cause a substantial effect on the level of housing price as the variables define the conditions of the general economics in any country (Gaspareniene et al., 2016). These factors e.g. income, employment, demographic effects, inflation, Gross Domestic Product (GDP) and policies has an effect on the prices of products and services that are created in different industries. This effect on prices then determine the products and services structure of the general demand.

2.2.1.1 Income

Ge (2009) stated that income as one of the house demand factor is the significant economic variable for describing variations in consumer expenditure and is a proxy of affordability. Muth (1960) study indicates that real income is an important factor when determining the prices of housing because income connects to decisions to be made by household; and concluded further that the demand for housing is largely sensitive to fluctuations in income and price. Maattanan and Tervio (2010) confirm that household income differences make a contribution towards housing price variations whereby high earnings make it possible for potential buyers to spend more money on housing.

Xu and Tang (2014) further state that house purchasing is greatly affected by income. Household disposable income has an important effect on the confidence of the participating in the economy. So changes in property prices affects the consumer spending in the economy. Xu and Tang (2014) analysed eight variables in the UK and the results presented a relationship that is negative between house prices and disposable income and money supply while it is

revealed that there is existence of a positive relationship between unemployment, cost of construction, credit, GDP and interest rate. The residuals indicated that there were real estate market bubbles after 2005 in UK housing market because the prices of the housing were substantially differed from the evenness relationship. Fortura and Kushner (1986) found that income constitutes a major variable in housing need, finding that an increase of 1% in the household income causes an increase of 1.11% in house prices (Wang et al., 2017). Bischoff (2012) further identified that real estate and income were interdependent. In Goodhart and Hofman (2008) study it was found that monetary variables that includes cash money supply and money credit.

2.2.1.2 Employment

Adam and Roland (2009) argued that economic activities are closely associated with house prices. These activities includes employment, GDP, industrial production and real consumption. Employment opportunities also attracts people to the municipality, this increases the local housing prices (Berg, 2002). Brooks and Tsolacos (1999) see unemployment rate as an indicator of the economic condition and it is negatively correlated to house price. According to Baffoe-Bonnie (1998) employment is one of the key basis for residential expansion. Employment serves as the lead indicator of future building activity and as current general economic indicator. Building activities are accelerated by higher employment growth, while lower rates links a decrease in the property market activity (Baffoe-Bonnie, 1998). The results of Baffoe-Bonnie (1998) support the earlier studies that changes that happen in employment make a significant contribution to housing prices. Constant changes in employment lead to constant changes in income, and which means there would changes in the demand for housing.

2.2.1.3 Demographic effects

The demographic variables are another set of factors that influence the housing demand (Hui, Wang and Jia, 2016). The demographic effects might have an effect on the prices, for example population growth and immigration (migration within the country) has an upward influence on municipal level housing prices. Green and Hendershott (1996) discovered that growing changes in demographic factors are expected to increase the property price. Potepan (1994)

study found that the source of variation in the growing population is migration and it is likely to encounter a large impact in the house price performance (Ge, 2009). Low population mobility decreases the demand in property acquisition (Gabriel and Nothaft, 2001). Ge (2009) found that the main determining factors in real estate price in New Zealand were migration in the long term, much expected from the investments, permits issued for building activities, mortgage loans and unemployment rates. Furthermore the substantial variations in the residential property demand generally created by the movements of migration. The demand for houses became less in New Zealand due to more people leaving New Zealand, and then this resulted in more houses being available for sell in the market which then resulted in the decrease in house prices (Ge, 2009). Population growth increases the demand for housing (Tafirenyika and Muzindutsi, 2017). Mankiw and Weil (1989) and Lai (2016) also found when population increases this increase lead to the demand of new houses increasing. The urban areas experience large population immigration because the people seek for improved living circumstances and better available work opportunities (Todaro and Smith, 2011). Urbanization is related to high demand for residential properties in developed areas.

2.2.1.4 Policies

The different kinds of government policies such as planning policies (de Bruyne et al., 2013), tax and regulatory policy and macroeconomic policy proved to affect house prices as they interact to produce changes in real estate investment (Baffoe-Bonnie, 1998). When tax policy and the regulations of the financial institution changes, this change contributes significantly towards the increase in real estate investment (Baffoe-Bonnie, 1998), for example transfer duty in South Africa, monetary policy can influence the mortgage rates. Transfer duty is a tax charged on the value of any property purchased by any person by way of a deal (South African Revenue Service (SARS) 2018). The new property price will vary to its neighbours' price taking into consideration its size because of values increasing end up with more taxes. This high taxes results in contributing towards the problem that is existing of housing affordability (Herbert and Gibler, 2014).

2.2.1.5 Interest rate

The amount of money (interest) that the central bank (South African Reserve Bank) loans the commercial banks known as the repo rate affects how the commercial banks charge the interest on the customers' loans. Baffoe-Bonnie (1998) stated that the changes that are occurring in regular interest rates will have an effect on the capability of working class to buy properties and many other usable goods. Low interest rates mean one can borrow money for cheap. One can borrow large amounts of money to afford more expensive things and this increases demand. The residential property sector always welcomes the cutting of the repo rate because it will help stimulate the housing market and this provides relief to homeowners who are battling to keep up with their monthly repayments. Barot and Yang (2002) have considered the real interest rate in their study and concluded that housing prices are negatively related to interest rates in UK.

2.2.1.6 Inflation

Investors are drawn into property market as inflation increases because of expectation of future appreciation in house prices (Franken, 2007). Gupta et al. (2010) found that in South Africa the property price inflation reacts negatively to monetary policy shocks. Inflation as explained by Hendry (2001) explains inflation as the outcome of an additional demand over supply in the country's economy. The expectation of the consumer on inflation plays an important role in need for houses (Baffoe-Bonnie, 1998). The inflation rate is explained by Amadeo (2019) as the percent rise or drop in commercial prices during a certain period.

2.2.1.7 Growth Domestic Product

Gross Domestic Product (GDP) affects the domestic demand for national outputs and measures the country's overall economic activity (Investopedia.com). Strong economic growth (where the production of goods and services are on an increase) have a tendency to promote demand for houses which in turn propels property prices (investopedia.com).

2.2.1.8 Exchange rate

Foreign trade impacts the price of the currency in relation to the global currency. The exchange rate has an influence on economy. A weak trading rate makes the exports to increase and by extension has an effect on real estate cycle (Cloete, 1990).

The above (section 2.2.1.5 to 2.2.1.8) macro-economic factors are the national (country level) indicators that influence the economic activities e.g. housing prices. These factors will be left out of this study because this study is conducted in a municipality at the residential sub-place level. These indicators are not available at that level.

2.2.2 Construction activities

Costs of construction and new housing play a vital part in studying property prices (Jacobsen et al., 2005). Expenses in construction that are to be taken into consideration consists of construction materials and the labour (Adam and Roland, 2009). These expenses are included when building new houses or maintaining existing houses. The high construction costs in some cases causes decline in the construction of new properties, eventually this will lead to housing prices increasing (Xu and Tang, 2014). This would then mean that the house buyers have to incur this increased construction costs. Karol (2009) claimed that if construction costs increases then this causes the house prices to appreciate more, which will then reveal that real estate developers have strong influence in shifting the construction costs to buyers. Construction costs make up the majority of the price of new houses in most markets, but their study has been relatively neglected (Gyourko et al., 2006).

The housing supply market is mostly modelled perfectly with the new construction investment. New properties always provide for positive spill-over effects on existing neighbourhoods by creating a vibrant neighbourhood because the vacant stands are now populated (Herbert and Gibler, 2014). New construction is greatly valued by those buyers who will want the utmost expensive new characters. Herbert and Gibler (2014) found that in established suburbs the newly constructed properties has a positive influence, however insignificant on the current existing house prices. The new properties are challenging the

existing houses that are of the same size and thus increase the house supply while demand in houses remains constant.

2.2.3 Land supply

The supply of land and construction costs both openly impact on the housing supply and housing prices (Li, 2010). The price charged for the acquiring the land is seen as a major factor that is influencing the property prices and housing supply (Bischoff, 2012). The scarcity of land (one of the house supply variable (Ge, 2009)) increases the prices for houses due to the restrained residential development together with higher housing demand (Peng and Wheaton, 1994). The housing supply and property prices in New Zealand cities are expected to be upwards (Ge, 2009) because of the land that is accessible and available for development is declining. The land availability is then deemed be an essential variable for estimating the housing prices (Ge, 2009).

2.2.4 Real Estates' characteristics

Chau et al. (2001) and So et al. (1995) suggested that real estate housing are goods that are classified by durability, structure and spatial fixation and these characteristics decide the on the house price. Wong et al. (2002) classifies property attributes into three classes, location attributes (access to social and economic facilities) which is about the centrality and accessibility (McMillan et al., 1980; Kockelman, 1997), structural (physical) traits (floor area, floor height, number of rooms, age) of property, and neighbourhood attributes such as how are the land use patterns and neighbourhood quality (Koramaz and Dokmeci, 2012).

The neighbourhood attributes are attributes of the house settings and its surrounding area (Huang et al., 2017) which show the development level of the area. These attributes include the working people's financial levels and population density (most of the time areas that densely populated are likely to be in much competition).

The structural attributes are those attributes concerned with the interior of the property which are vital factors to decide the housing price (Tu et al., 2004). The characteristics like floor area, age of the house, number of rooms can be heavily associated to the house prices

as suggested by Ball (1973) that attributes that most people desire are shown in the higher real estate market price. Several residential market attributes like the age of houses has influence and the quality of life could have a positive impact on housing prices. The price of the property depends on the residential housing's features which are well classified for each property (de Aguiar et al., 2014).

2.2.5 Spatial determinants of housing prices

When determining the property price, the geographical location is identified as the main important determinant variable (Yusof and Ismail, 2012). Gelfand et al. (2004) studied the impact of housing price and the importance of spatial element was highlighted in analysing housing prices. The location of properties is important as this is related to general accessibility and the distance travelled to economic centres (de Bruyne et al. 2013).

Koramaz and Dokmeci (2012) stated that proximity characteristics for example distances to travel to central business district (CBD), centres of employment, transportation nodes and publicly available amenities are the most important spatial determining factor determining the differences in property prices. Koramaz and Dokmeci (2012) investigated housing values in Istanbul. The multiple regression model was used to create the semi-hedonic price model which emphasised characteristics on location as one domain from the three domains in hedonic price model. The spatial determining factors used in the study were distance to central business district, sub-centres, transportation and the coast, which were used in the multiple regression analysis. All variables produced a substantial effect on the differences in property price level in Istanbul. Housing unit resulted as the major factor in explaining the values of housing price in Istanbul and CBD distance has employed negative impact on how the housing values are spatially distributed. Furthermore Koramaz and Dokmeci (2012) applied interpolation technique to model the housing price values in a continuous surface. Kriging interpolation was conducted using ARCGIS 9.2 on actual and predicted data and the it indicated the better locations representing house price values.

The availability of public transportation (Bajic, 1983) and proximity to economic centres (de Bruyne et al. 2013) increases housing prices. De Palma et al. (2005) highlighted that when the

time to travel and the amount costs one to access the employment centres is decreased there is a significantly increase in housing prices. Houses that have access to public transport appeals more to the potential buyers. They are prepared to pay more and buy those properties (So et al., 1996). People like mobility in their lives (de Bruyne, 2013). The location of work place is vital when deciding where to settle in buying a house. The purchase is made by also considering the distance thereof and the time will take to travel. Housing prices differ according to proximity to conveniences (city amenities). Housing units with relative proximity to urban centres and amenities tend to have considerable high market value (Xiao et al., 2017). Property value increases due to being closer to stores amenities because amenities close to area attract residents choose that area to live. Housing is a commodity that have several attributes and property buyers routinely consider the amenities that are available in the area, how easy will it be to access work place, how are properties nearby compares before purchasing (Zainora et al., 2016).

2.2.6 Spatial effects and housing prices dynamics

A number of studies (Clarke and Daniel, 2006; Ge, 2009; Selim, 2009; Xu and Tang, 2014, Olowofeso, 2016) on housing price dynamics have been conducted and most of them failed to take into consideration the geographical effects. Spatial influences which include spatial dependence/autocorrelation, spatial heterogeneity and spatial spill-overs of housing prices among neighbouring regions (Wang et al., 2017).

Anselin (2001) defines spatial auto-correlation as the coincidence of similarities in the location and attributes. This implies that in real estate market, houses which are next to each other locations their prices are likely to be similar. Features which are lying next to each other are mostly correlated than those features that are not close to each other (Cressie, 1989). There are two reasons why the house prices are autocorrelated (Basu, 1998) namely neighbourhoods that are developed and grow stronger at the same time means properties with alike structural features (homeowners tend to follow their neighbours' improvement activities Yu et al (2007)) and location amenities shared by neighbourhood residential properties (Dublin et al., 1999). The Property Assessors appraise the value of the property by referring to the suburb conditions (Yu et al. 2007). With the spatial autocorrelation that exists

in the housing prices this causes violation in the regression model assumptions that the observations are to be independent (Anselin, 1988).

Spatial heterogeneity as described by Huang et al. (2017) is that constant change and variation in distribution of spatial data in the area of interest or non-stationary process over space exists. This affects the operational process whereby the same set of housing attributes could generate different housing prices in different parts of the study area (Yu et al., 2007). Urban neighbourhood structures are diversified and this diversity has an impact on consumer's assessment of housing qualities on housing prices (Can, 1990). Yu et al. (2007) employed two approaches for example spatial regression and geographically weighted regression (GWR) to study the spatial effects namely spatial dependence and spatial heterogeneity. The GWR surface was created using the inversely distance weighting interpolation and the spatial regression performed well in terms of predicting the model accuracy than ordinary least square (OLS) which tends to produce overestimated coefficients.

2.3 The role of spatial analysis and GIS approaches in analysing housing prices

Spatial analysis is the practical use of statistical techniques applied to geographical related research questions in order to provide the solutions (Gatrell, 1987) and the results thereof are explained, visualised and dependent on the locations of the objects being analysed. The evaluations in real estate were performed initially by spatial interaction and statistical analysis which are now included in the spatial regression model. The evaluation was to explain and create principles of spatial econometrics (Koramaz, 2012). The spatial data analysis and available GIS tools interfaced to form a framework that is integrated for spatial decision support system and able to provide visualization (Anselin, 1998) for example visualizing the distribution of housing prices and enrich the analytical analysis.

GIS provides number of functionalities such as ability to manipulate, map and visualise spatial data to enable the explorations of the data. Can (1998) further states that when using GIS, you are able to use data acquired from different sources with disparate formats. GIS is good in querying the data spatially and overlay capabilities which facilitates how the data is organised and managed to fit research needs (Can, 1998). GIS is an appropriate tool for the

analysis of housing markets because residential property is stationary in geographic space (Ismail et al., 2005).

GIS spatial querying capabilities enables a query based on spatial relationships (Can, 1998) e.g. can query the housing transactions from the current listing within the three-kilometer radius that occurred within the last six months. Using the geo-coding function, the residential properties can be geo-coded by using the street address data. The street address of properties provides a unifying element in geo-coding and overlay operations. These operations allow that the address of the property could be used as way to link spatial aggregation of individual transaction data at chosen levels of geographic entities (Can, 1998).

Migozzi (2019) investigated how financialization unfolds in Cape Town metropolitan, South Africa using different methods including fieldwork, interview and spatial analysis, the history of housing policies in South Africa revealed that accessibility to housing financial assistance and home-ownership was restricted to racial criteria which then formed racialized and segregated urban space (Migozzi, 2019). GIS mapping is an important part of visualization of housing prices in the spatial analysis (Koramaz and Dokmeci, 2012). To visualise the structures of the housing market in Cape Town, Migozzi (2019) created the residential areas' zoning and each housing transaction was geo-located, using spatial categories namely white-only, coloured and black residential areas and some post-apartheid areas. During the apartheid Black, Coloured and Indian people because of racial discrimination struggled to access housing finance, and Black people were not allowed to own property in urban areas until 1986 (Migozzi, 2019). Capital housing subsidies continued to be a good policy that facilitated that people get access to housing and that they own homes. South Africa suffered an intensive property boom from 2003 to 2007 before prices declined significantly (Migozzi, 2019). The property boom provided housing finance as the major tool in acquiring property in the property market. The appropriate access to financing on housing conveniently encouraged the Cape Town property boom and this was revealed in the discussions held with estate agents and developers.

Huang et al. (2017) study examined the housing prices changing aspects based on the 12732 pieces of housing data. The data had house attributes and covered all districts of Shanghai.

The housing prices in Shanghai are analysed with GIS, OLS and GWR. Huang et al. (2017) employed the GIS tool that is used to analyse the hot spots in the data to help determine if properties with high or low or prices gather together. Hot Spot Analysis uses Getis-Ord G_i^* statistic to identify the spatial clusters of values that are statistically significant. Huang et al. (2017) found in Shanghai that geographical distribution of low or high housing prices was clustered based on Moran's I statistic and also hot spots were spatially concentrated.

In Huang et al. (2017) study, the value of 0.916 for Moran's I was obtained and this indicated that there is spatial clustering in the distribution of high or low housing prices. The study results indicate that the effects of the geographical attributes of housing differs over the space and these contribute to the final housing price. Geographical weighted regression is the other option to use when dealing with the spatial auto-correlation. The adjusted R^2 in Geographical weighted regression model was 0.826 which shows that GWR has been active in exploring the spatial non-stationary distribution of housing. It was noted that GWR model has limitations and it is not suitable for all data because some variables are held constant throughout the area that is being studied and other variables differ spatially.

Zainora et al. (2016) highlight that with GIS technology, maps are used to represent information of interest and this assist all those that are involved in the real estate market such as developers, buyers, realtors and sellers when looking for appropriate houses to be developed or assessing the market values.

CHAPTER 3 : MATERIALS AND METHODS

3 Materials and methods

3.1 Study area

The study focussed on City of Johannesburg (CoJ) municipality as the study area (Figure 1) and its housing market. CoJ is a metropolitan municipality in the centre of Gauteng province and its covered by an area of 1,645 square kilometres.

The CoJ population of 4.4 million included 76.4% of African, 12.3% of white, 5.6% of coloured, and 4.9% of Indian (Statistics South Africa, 2011). Furthermore, the municipality has 1 434 856 households with an average household size of 2.8 persons per household and unemployment rate of 25% (Statistics South Africa, 2011) and seven regions which they are responsible for service delivery divided the municipality. The current regions with main towns are:

- Region A - Diepsloot, Midrand and Ivory Park
- Region B - Northcliff and parts of Sandton and Rosebank
- Region C - Roodepoort
- Region D - Soweto, Doornkop, Diepkloof and Meadowlands
- Region E - Alexandra and parts of Sandton and Rosebank
- Region F - inner city and Johannesburg South
- Region G - Ennerdale, Orange Farm, Lenasia, Eldorado Park and Protea

CoJ's housing market is more active, affordable and growing more than most cities (CAHF and SACN 2014). Regarding the market size, Johannesburg has the biggest real estate industry and Johannesburg comes second compared to Cape Town when comparing the overall worth of the residential property market (CAHF, 2018). Furthermore the affordability in Johannesburg is better but most households are still experiencing challenges and on average it takes 2.7 times the average income to pay for the average property (CAHF and SACN, 2014).

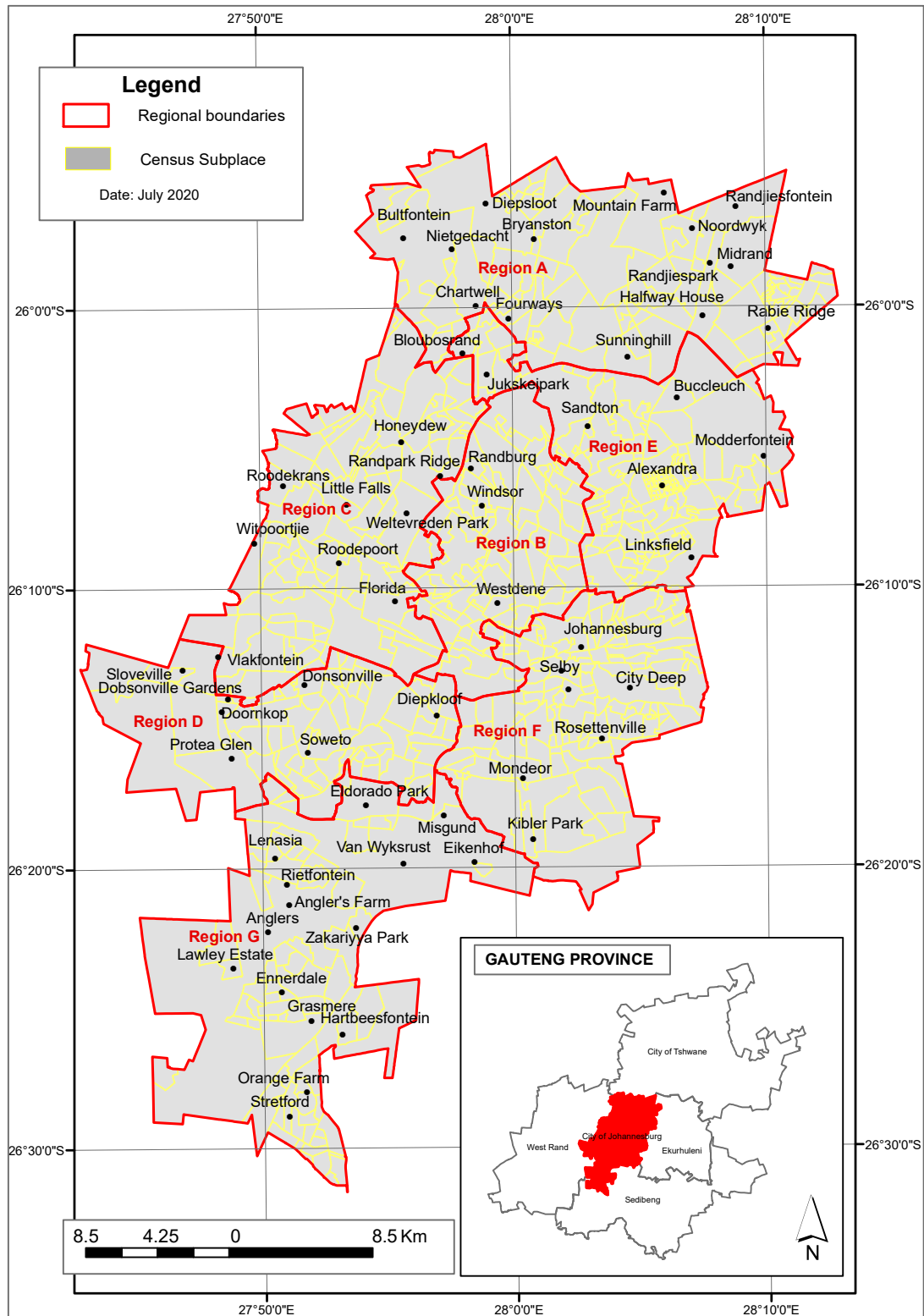


Figure 1: Study area, CoJ with city regions (Data Source: City of Johannesburg)

3.2 Explanatory variables and their expected effects

The study adopted a combination of a housing supply and housing demand perspective (view) factors to explain housing prices as the dependent variable. From the housing *supply* perspective, number of properties sold, new constructions and building permits were used. In support of the housing *demand* perspective the following variables were used: household income, unemployment, property transfer duty and property ownership. The following are the selected variables:

- **Household income.** Hofmann (2004) found that house prices are usually driven by **income growth** and interest rates. Income affects housing demand in relation to the buying power (Mankiw and Weil, 1989). The disposable income is a representation that household being able to afford the housing stock and income is correlated positively with housing prices therefore higher income means higher demand for housing (Xu and Tang, 2014).
- **Unemployment.** Employment opportunities in most municipalities has power to attract people to municipality and this triggered escalations in local housing prices (Berg, 2002) then de Bruyne and van Hove (2013) expect a higher unemployment in municipality to cause housing downwards. **Property ownership** (number of owner occupation properties, renting, mortgage loans payment). Most houses are financed through mortgage loans (Tafirenyika and Muzindutsi, 2017) so when the availability of loans increases, the housing supply will ultimately increase.
- The **proportion of renters** in a given city represents the firm demand for housing and indicates the difficulty of housing purchases (Wang et al., 2017). Number of people renting signifies the demand for urban housing (Goodman, 2003).
- **Property Transfer Duty** different kinds of government regulations such as planning policies impact the market outcome (de Bruyne and van Hove, 2013). It is the required tax from government which is payable by buyers when transferring the house from the seller's name to the buyer, new owner's name (Nolan, 2019). Hamilton (1976) hypothesized the effects of property tax capitalization are to decrease the worth of bigger properties and cause increase in worth of smaller properties that are in the same vicinity.

- Borowiecki (2009) found that activities of any construction are considerate to the mortgage market and are vital in housing prices. The **new constructions** effects on the prices of neighbouring existing properties depends on how big the properties are (Herbert and Gibler, 2014). It is believed that high finances in the cost of construction will decrease the actual construction and housing stock (Xu et al., 2014). The new houses compete with the existing houses with same size and this increases supply while the demand remains the same (Herbert and Gibler, 2014).
- **Property Sales (number of properties sold)** is an indicator for the performance of purchasing activities of residential properties. The property market is often affected by what is happening in the country at large and the big influences are political, economic and social arena.
- **Building plans** (building permits applications). Number of building plans (permits) that are passed serves as a most important indicator for the supply of properties (Clarke and Daniel, 2006). Building plans are the leading indicator for building activity and they present overall attitude about investment into real estate.
- Roback (1982) found that **city amenities** influence the choice that people have in deciding where to live and this contributes to grow the city. City service level is important characteristics of the city's attraction and competitiveness (Wang et al., 2017). The amenity variables are categorised into structural characteristics (age of the property, rooms, floors, types of roof), locational characteristics (accessibility to CBD, availability of public transit and environmental factors (degree of development, living standards, land use pattern). The geographical distribution of urban amenities affect the city-level demand because many people like properties that are in locations that provide a higher quality of life (Huang et al., 2015).
- **Demographic effects** are important in influencing housing demand. According to Maennig and Dust (2008) the increase in population and shrinkage in the population shrinkage respectively increases or decreases housing demand and the increase in population density increases the property prices. The municipalities which experience high growth in population tends to experience increase in housing prices this is due to increased residential housing demand (de Bruyne et al., 2013).

3.3 Data acquisition

The dataset contains the 2012 housing prices data and 2011 South African National Census data obtained from Centre for Affordable Housing Finance (CAHF), Statistics South Africa (StatsSA), AFRIGIS and City of Johannesburg. AFRIGIS is a company that is passionate about location and answering the “WHERE” questions. It provides easy, secure access to quality location information and insights. Most of their operations are producing maps, locate places, people and things, and communicate the results thereof.

CAHF is an independent organisation, pursuing a mission in supporting and growing the residential housing markets in Africa. The CAHF data is based on real estates that are registered formally on the deeds registry supplied by Lightstone Property Company in Microsoft Excel format and merged the suburb-level data from the deeds registry with StatsSA sub-place data from the 2011 Census. The residential property data is separated into four segments by CAHF as follows:

- The entry market where properties that worth R300 000 or less,
- The affordable market, this is the market where properties worth R300 001 - R600 000,
- The conventional market is where the properties worth R600 001 - R1.2 million and
- The high-end market with properties worth over R1.2 million.

The spatial level of analysis of housing prices is identified as the Sub-place name level (the lower level) of the two layers that forms Census Place name layer. The Place name layers consist of a main level of place name, i.e. main places (MP) and a lower level or sections of MPs i.e. sub-places (Statistics South Africa). The Place name spatial layers fit within the boundaries of the administrative boundaries used in South Africa as defined by the Municipal Demarcation Board. The spatial layer, the sub-place name layer, contains the 804 polygons that are determined by Statistics South Africa, which correspond to suburbs, or villages.

The variable of interest for the study is the 2012 housing prices; this includes houses that were only sold and captured in 585 sub places in 2012 at the Deeds office measured in rands. The 2012 housing prices were used because of the housing market that was sinking before

some years and was now bouncing back. The Property transfer duty data is calculated using the value of the property and the rate obtained from SARS website (www.sars.gov.za). The census data describe the socio-economic characteristics of each subplace.

The independent variables at sub-place are counts of individual variables (e.g. number of new constructions, properties sold, approved plan), unemployment rate, transfer duty fees are listed and described in Table 1. The gaps in the data were encountered then the mean of each variable was used to replace the missing data in order to allow for the analysis on complete data.

The city level service (2012 amenity point of interest) layer is sourced from AFRIGIS Company. The amenity point of interest data is categorised as follows education, entertainment, commercial, filling stations, medical, government, place of worship, recreation, stores, tourism attraction and transportation. Examples thereof are schools, tertiary institutions, fast food outlets, restaurants, shopping centres, petrol stations, convenience stores, post offices, public and private hospitals, clinics, churches, sport facilities, parks, laundries, supermarkets, museums, bus stops and railway stations.

Table 1: Description of the explanatory variables

DATA	DESCRIPTION	Unit of measure	DATA SOURCE
Dependent variable			
Unit of analysis : Census sub-place			
Housing prices	Average transaction prices per sub-place	Rands	CAHF, 2012
Independent variables			
Macro-economic factors			
Household income	Average household income - per sub-place (all sources of income including salaries, social grants and any other remittances).	Rands	StatsSA, 2011
Unemployment	Unemployment rate per sub-place - 15 years and older persons	Rate %	StatsSA, 2011
Mortgage loans Renting (Property ownership /tenure)	household that owns and has not paid off their mortgage per sub-place people renting out per sub-place	Number of (total count)	CAHF 2012, StatsSA, 2011
Property transfer duty	Fees payable to SARS when transferring a property It is calculated using the schedule obtained from SARS website.	Rands	SARS, 2012
Demographic factors			
Population density	Number of people per given area	People per sq /km	StatsSA, 2011
Constructions			
New constructions	New property transactions used as proxy for new constructions	Number of (total count)	CAHF 2012,
Sales	residential transactions (properties sold) per sub-place	Number of (total count)	CAHF 2012
Building permits / plans	Number of approved building plans per suburb	Number of (total count)	CoJ, 2012
City service level	The number of city services amenities available in the municipality	Number of (total count)	AFRIGIS, 2012
Structural characteristics			
Age	Average ages of property , grouped ages	no. of years	StatsSA, 2011
Rooms	Average number of rooms of a housing unit, grouped rooms	No. of rooms	StatsSA, 2011

3.4 Methods

3.4.1 Explanatory data analysis

In order to provide summary and describe trends in the data, the descriptive statistics were computed by implementing R packages. The data summary of statistics and its graphical methods were created to describe and understand the data (variables) better. The summary statistics used included minimum, median, mean, maximum, standard deviation, skewness and kurtosis. The calculated kurtosis describes the tail shape of the data distribution and the skewness measures the data symmetry. An obtained value that is greater than 2 for skewness is considered as highly skewed (Kim, 2013). The negative skewness value shows that the data distribution is skewed to the left and the positive skewness shows that the data distribution is right-skewed. Kurtosis for a perfect normal distribution is zero (Kim, 2013). Leptokurtic distributions have positive kurtosis values (Nneji, 2013) with a higher peak and taller tails than a normal distribution. Mesokurtic distribution results when kurtosis is equal to zero, which mean that the kurtosis reached normal distribution (McLeod, 2019). Negative values of kurtosis indicate that a distribution is flat (less peaked) and has thin tails which is platykurtic distribution.

The graphical methods used included the histograms, scatterplots and chloropleth maps. The histograms, scatterplots and chloropleth maps are used to describe and visualise the data distributions, discover patterns (arrangement of objects in an area) of association by clusters or hot-spots and detect spatial data outliers (Anselin, 1998).

3.4.2 Measuring spatial effects

To determine the extent of spatial autocorrelation and heterogeneity in the data, the following steps were implemented by using R procedure packages. The first step was to choose the neighbourhood criterion, to check which areas are linked to each other, then followed by assigning the weights to the areas that are linked. Neighbourhood structure and spatial weight matrix are created for spatial analysis. To create the neighbours and their corresponding spatial weights, the sub-place shapefile is loaded into R this will indicate which sub-place is a neighbour with which sub-place. Then the Queen's spatial weights are created with the row standardisation option which means that the elements of each row sum up to one. Under Queen's criterion, areas are neighbours when they share boundary or vertices

(corner). Information on neighbours is then stored in a connectivity or weight matrix (W) (Can, 1998), which has dimensions of N by N matrix (where N is the number of spatial entities such as houses, sub-place). The spatial weights matrix is used to confirm location similarity and connections (Anselin et al., 2000) defining the neighbour connectivity (which one is the neighbour?) and neighbour weights (how much does each neighbour count?) (Brazil, 2020).

The third step is to run a statistical test to examine spatial autocorrelation. Taking into a consideration that the data might possibly present spatial dependence and heterogeneity, a statistical test (Moran's I and Geary 'C statistics) to examine the spatial autocorrelation is performed. Moran's I identifies whether the house pricing belong to cluster, random or dispersal pattern. Moran's I test is performed to test the hypothesis of absence of spatial clustering.

H_0 =there is no spatial dependence/autocorrelation/Moran's $I=0$ for the housing prices

H_a =there is spatial dependence/autocorrelation/Moran's $I \neq 0$ for the housing prices

Finally to quantify the relationships between housing prices and independent variables OLS regression is applied in order to fit the best model. When fitting the best model the key linear regression assumptions are not to be violated. However OLS does not account for spatial effects but since spatial data analysis was developed; the spatial effects are able to be taken into consideration (Wang et al., 2017) through the spatial models namely spatial error model and spatial lag model.

The spatial autoregressive regression models are fitted to include spatial autocorrelation in construction of model. Error model is fitted as alternative to the lag model. In the spatial lag model (substantive), the effects that are spatially related are incorporated by including a spatially lagged dependent variable as an additional predictor. In the spatial error models (nuisance), the spatial effects are included in the error term spatially weighted.

3.4.2.1 Regression assumptions

The key regression assumptions were as follows:

- Linear relationship – the linearity can be tested using the scatterplots checking for outliers.

- Multivariate normality – The dependent variable should be normally distributed. This assumption can be tested with a histogram plot. The data that is not normally distributed can be transformed using non linear transformation, log transformation.
- Little or no multicollinearity – The predictors should not be strongly correlated with each other (no multicollinearity). Variance Inflation Factor (VIF) is used to test multicollinearity. Where the VIF value is greater than 5 it indicates that multicollinearity may be present and $VIF > 10$ indicates that there is multicollinearity among the data (variables). if there is multicollinearity the solution is to remove the variable with high VIF.
- No auto-correlation – The observations (variable) should be independent of each other. If this assumption is violated, the coefficient estimates could be wrong, www.bisolutions.us. Durbin-Watson's d test is used to test auto correlation in the model. The values of $1.5 < d < 2.5$ would indicate that there is no auto-correlation in the data.
- Homoscedasticity – can be tested with a scatterplot check if the residuals are equal across the regression line. The Goldfield Quandt test can also be applied for heteroscedasticity.

Spatial observations close to each other are related and this violates the OLS model assumptions. These violations are tested by using some of diagnostic tools such as scatter plots, histograms, box plots and various formal statistical tests such as the Moran's Index correlation tests, the Bruesch-Pagan tests of homoscedasticity, the Jarque-Bera test statistics of normality of residuals. The explanatory power of the OLS model is assessed using R^2 value (Getis and Aldstadt, 2004). It is used to judge the goodness of fit of the model. The Akaike Information Criterion (AIC) is commonly used to compare performance between any two models (Akaike, 1974). A model with the least AIC value is the better model.

3.4.3 Geographical detector technique

Lastly the investigation was completed by employing the geographical detector a method suggested by Wang et al. (2010). This method is used when there is a presence of number of important determining factors relating to the housing price. Wang et al. (2010) used geographical detector method to evaluate the power of association between the independent variables used and dependent variable (housing prices) in China. The outcome of Wang et al. (2010) study indicated that the cost of land contributed immensely to housing

prices. The cost of land was followed by proportion of renters and the state of the house. This method is used in analysing the impacts of underlying factors on local disease risk. The magnitude of the significant (impact) factors (variables) was investigated and estimated using the geographical detector method. Wang et al. (2010) analysis investigated the factors that were influencing the housing price variations.

CHAPTER 4 : THE RESULTS

4 The results

4.1 Spatial distribution of housing prices in CoJ

The spatial distribution of City of Johannesburg sub-places' 2012 average housing prices are shown in Figure 2. The 2012 average housing prices shown are divided (by CAHF) into four segments as the entry, affordable, conventional and high-end market. The housing prices shown are for properties only sold and registered in 2012 at the Deeds Office.

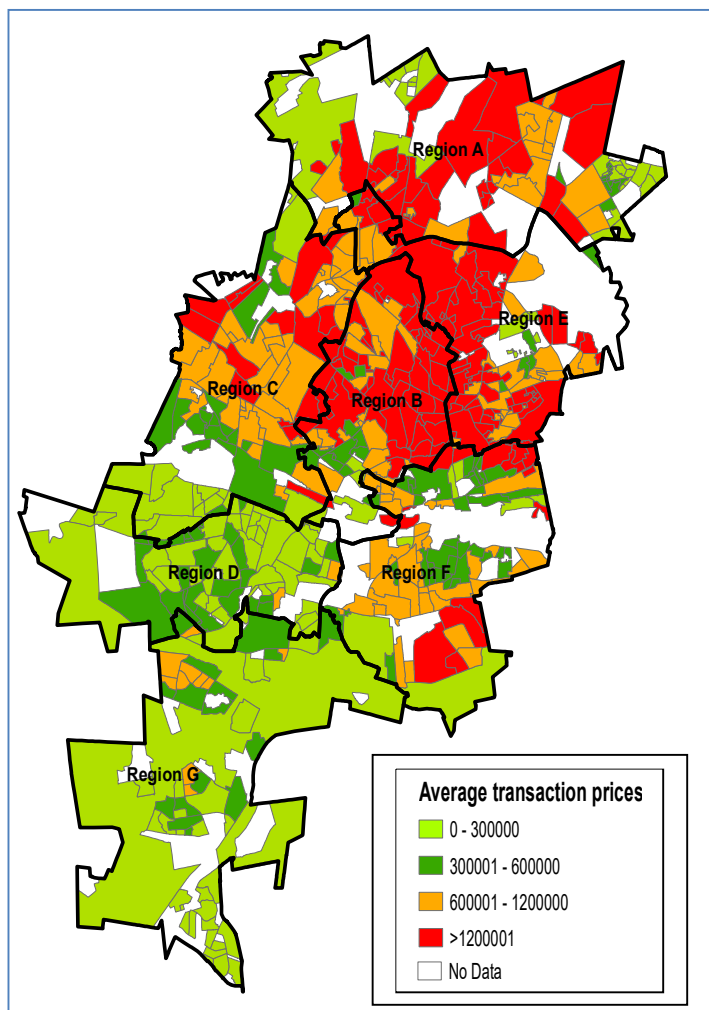


Figure 2: 2012 Average housing transaction prices by market segments in CoJ

Data Source: Centre for Affordable Housing Finance in Africa (CAHF) using 2012 deeds registry data supplied by Lightstone Property

Each demarcated boundary on the map (Figure 2) represents the Census sub-place. The majority of the entry and affordable markets (properties less than and equal to R300 000)

houses are located in Region D and Region G, the south western parts of the municipality. It is mostly in Soweto, Doornkop, Diepkloof and Meadowlands (region D) settlements in the outskirts and Ennerdale, Orange farm (region G). Region C and Region F, (Roodepoort and Johannesburg South respectively) are found to vary with all the markets. Region C comprises largely of farms and small holdings. The high-end market (properties more than R1.2mil) houses are located in Region B, Northcliff, Houghton, Sandton, Rosebank, the upmarket residential areas in the northern sub places of the metropolitan municipality.

4.2 Spatial distribution of different housing tenures in CoJ

The home-ownership and renting are considered to be the most prevalent forms of tenure when coming to people's tenure choices. The home-ownership tenure is explained as where homes are outrightly owned and mortgaged. Tenure is the legal right (status) which people possess for them to occupy the accommodation of their choice (Diaz, 2009), the different ways that one can own a property. The individual home-ownership tenure e.g. freehold, estates and sectional titles shown in Figure 3; and renting tenure were explored Figure 4.

Freehold (full title) properties are properties that are entirely owned and these are single home on a single stand and one can dispose it at will. On the other hand, sectional title describes the units that are separately owned or sections within a complex. Estates type is described as extensive area of land usually with an elaborate house on it. The freehold ownership properties are located in most parts of the municipality (Figure 3.a). This shows that the full title is the most favoured home-ownership tenure. The freehold ownership in the high-end market is mostly in Region B, and E, central place of the municipality and the entry and affordable market is to the west of the municipality in Soweto, Doornkop (region D). Region E has more established suburbs and newer suburbs that was due to developments in and near central Sandton.

Most of sectional titles are located in the northern and eastern parts of the municipality. Sectional title units are more affordable than freehold properties and that provide more people a way to enter the property market. There are no sectional titles in Region G in Ennerdale, Orange farm the southern part of the municipality (Figure 3.c).

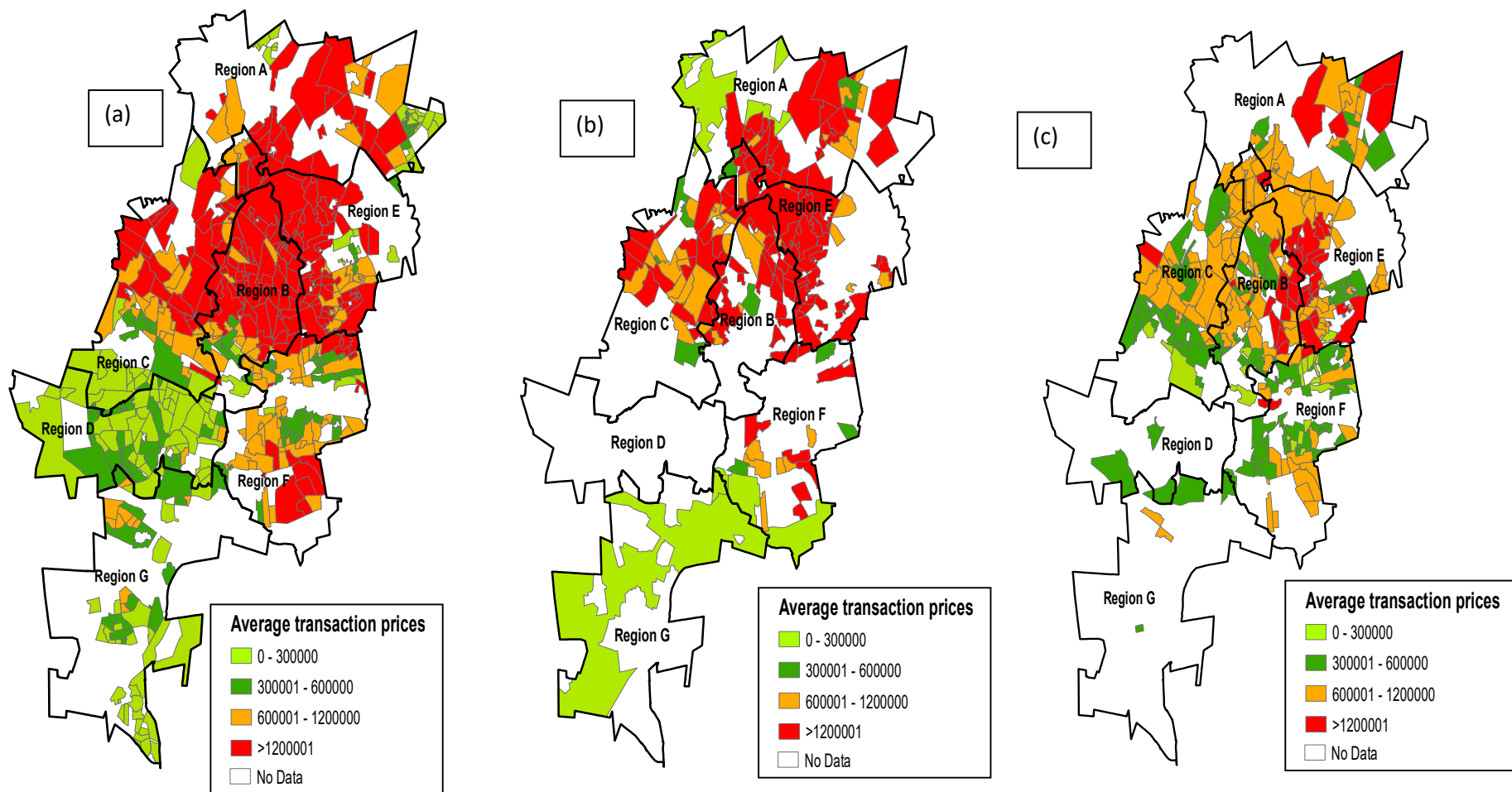


Figure 3: Spatial distribution of the 2012 residential property homeowner tenure and prices by market segments in CoJ: (a) Freehold, (b) Estates and (c) Sectional titles
 Data Source: Centre for Affordable Housing Finance in Africa (CAHF) using 2012 deeds registry data supplied by Lightstone Property.

In Figure 3.b, Estate properties mostly exist in the high-end market. There are no Estates in Region D western part of the municipality. The densification of residential areas are most common as construction of more townhouse complexes is increasing and subdivisions of large single residential properties are preferred. Comparing all three ownerships, Region A (in Diepsloot, Midrand, and Ivory Park) experiences all types of the ownership per all markets. The freehold and estates that are of entry and affordable market are observed in western and southern part of municipality.

The residential property market comprises of owner occupied and rental housing. Figure 4 shows that more households preferred renting housing option than buying through mortgage loans. The decision to buy, rent or invest in a property is depended on the current market factors applicable to residential market participants. The municipality experienced renting housing and home-ownership housing.

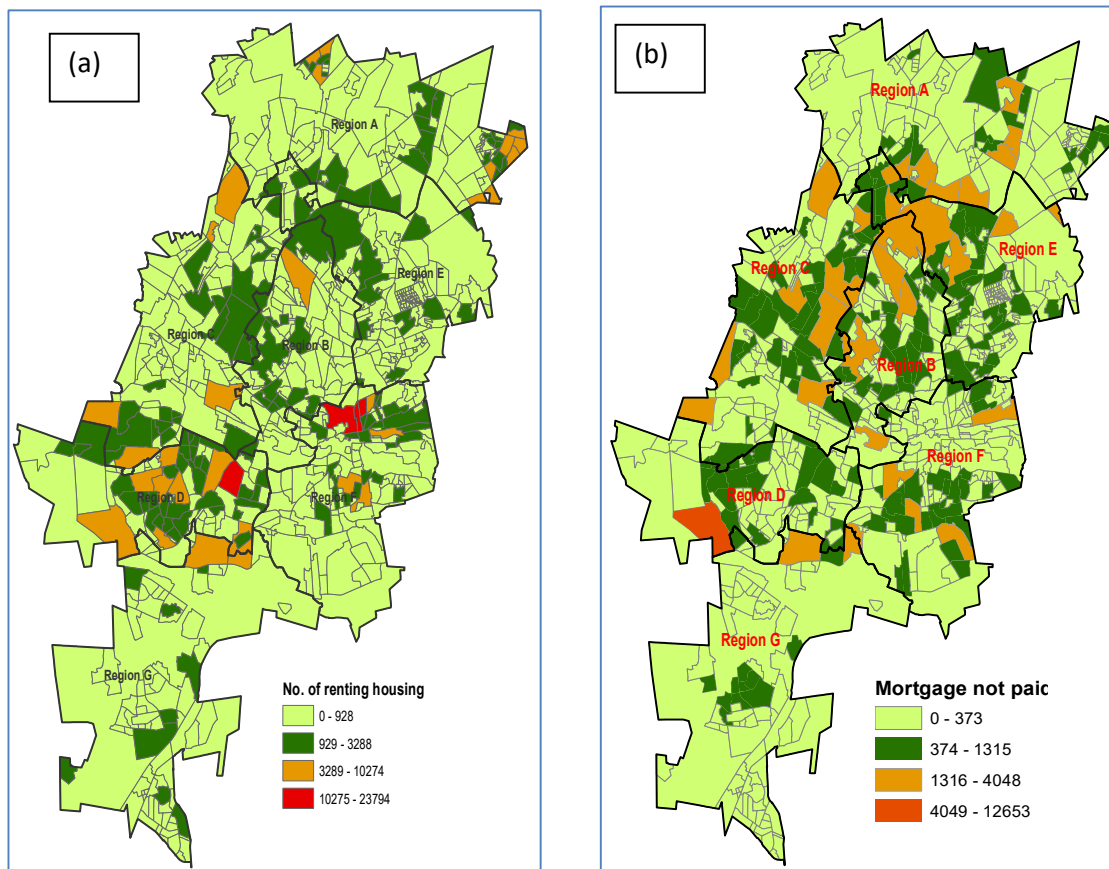


Figure 4: Distribution of (a) 2011 rental housing and (b) households mortgage not paid
Data Source: Census 2011

Orlando East (region D), Hilbrow, Berea and Johannesburg central (region F) sub-places have experienced high number of households living in rented housing. Some of the households have opted for home-ownership by purchasing homes through mortgages loans. The high number of mortgages not paid off are experienced in Roodepoort, Protea Glen, Randburg, Sandton and central part of the municipality.

4.3 Descriptive statistics

4.3.1 Summary statistics and plots

The summary statistics are summarized in Table 2. The summary statistics for the housing prices showed the sub-places' lowest average price of R18 000 and highest average price of R11 400 000 among 804 sub places in CoJ with an average selling price (overall mean) of R991 859 during 2012 (Table 2). The summary statistics of number of sales in 2012 showed sub-places performed in purchasing activities with a maximum of 1208 (Stesa Agricultural Holdings (AH)) properties sold whereas other sub-places have not sold any property. The dependent variable housing prices and independent variable transfer duty indicated no difference between mean and median values. All variables had high variance and standard deviation (above 0). The variable with the highest difference between mean and median values was found in household income.

Table 2: Statistical summary of the variables

Variables	Min	Median	Mean	Max	Std.Dev	Skewness	Kurtosis
Housing Price	18000	991859	991859	11400000	837848.10	3.87	32.98
Household Income (R)	0	110290	174525	795491	163338.1	1.17	0.87
Unemployment (%)	0	13.17	17.94	65.18	14.86	0.57	-0.93
Transfer duty (R)	0	64394	64394	949000	84080.16	3.06	17.97
Mortgage loans not paid off (count)	0	125	294	12653	620.57	11.02	196.45
Renting (count)	0	248	735	23794	1605.26	6.82	70.37
Population density (p/sq km)	0	2988	7991	105516	12449.83	3.00	11.06
New constructions	0	0	12	1205	63.13	12.75	200.28
Sales (count)	0	9	43	1208	100.03	5.62	44.65
Building plans (count)	0	0	6.05	1861	71.85	22.63	558.58
City services (count)	0	2	12.34	429	29.98	6.11	57.21
Rooms (count)	0	4.40	4.426	10.00	1.665	-0.192	0.351
Age (years)	0	12.60	16.27	57.40	16.061	0.509	-1.068

The summary statistics of unemployment showed that sub-place level unemployment (number of persons not employed) varied between the lowest at 0% and highest at 65% with the average of 17% in CoJ. The summary statistics of transfer duty showed that some of the properties did not attract the property duty fee, whereas some attracted the highest fee of R949 000 per property. The summary statistics showed that there is an average household income of R174 525 with some households with no income. Maximum of 23 794 households were renting accommodation as compared to 12 653 of mortgages that have not been paid off. Some of the sub-places did not have new constructions and some sub-places had a high number of 1205 for new constructions.

The housing prices are skewed with the +3.877 value. The value indicates that the distribution of the data is positively skewed with high degree of skewness thus implying non-normality of housing prices. The skewness indicated that the dependent variable should be transformed to conform to a normal distribution. The variable was log transformed.

The excess kurtosis value of raw housing prices was 32.979, which indicates that the distribution of housing prices growth is leptokurtic. This is consistent with the fact that its histogram is not bell-shaped.

The property ages and unemployment kurtosis (Figure 5 and Figure 6) indicates a flatter peak with a negative kurtosis value of -1.06 and -0.93 (Table 2) showing platykurtic distribution.

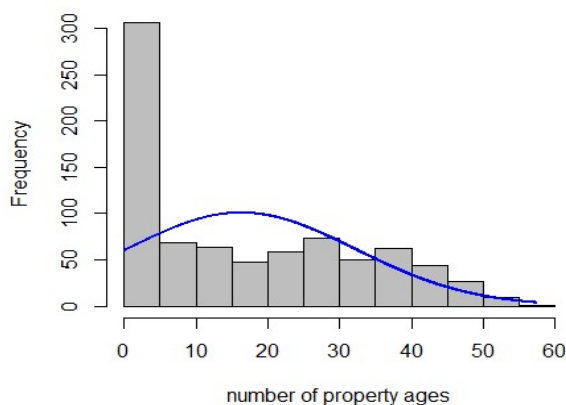


Figure 5: Histogram showing the distribution of property ages

The normality of the raw data distribution was evaluated by obtaining the skewness and the kurtosis values; and the histogram to indicate the distribution characteristics.

The histogram (Figure 6) is showing that the raw housing prices and unemployment data is not normally distributed is skewed to the right with a high degree of skewness, which means that the data needed to be transformed.

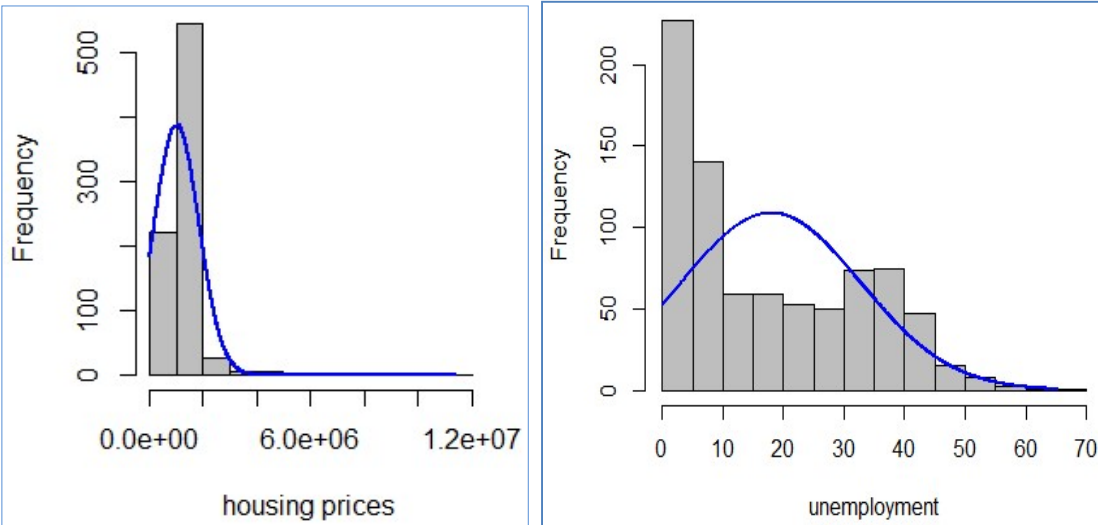


Figure 6: Histogram of raw housing prices and unemployment observations

Parametric tests necessitate that residuals to be normally distributed, the relationship between variables should be linear, no outliers and that the residuals be homoscedastic and it is not the case, the model assumptions are not met, as shown in Figure 7.

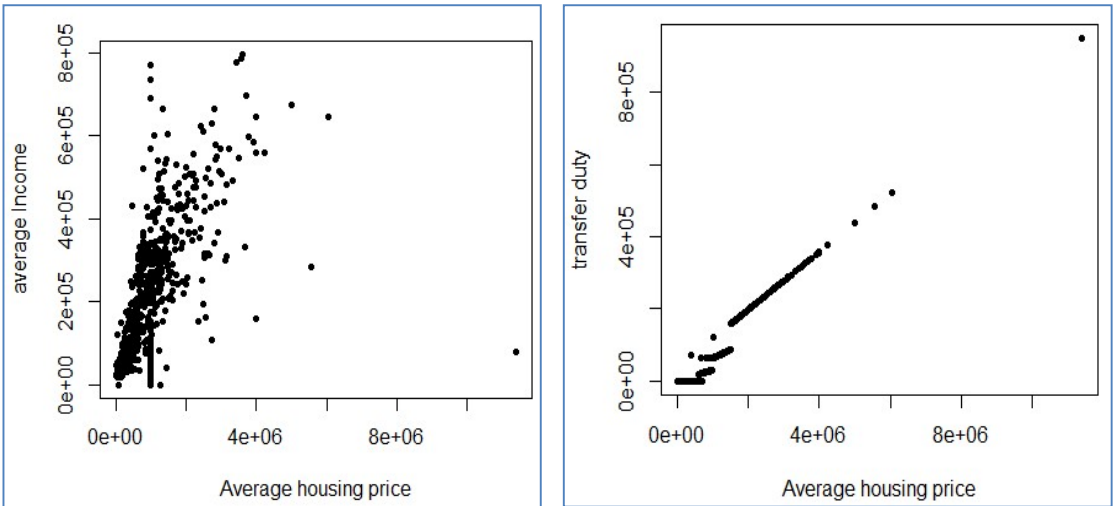


Figure 7: Scatter plot showing relationship between housing prices with average income and transfer duty

When residuals fail to meet these above mentioned assumptions one way to address them is to transform one or more variables to follow a normal distribution (Mangiafico, 2016). For right-skewed data, the most common used methods of transformations include logarithm, square root and cube root. To address data that do not fit model assumptions (Mangiafico, 2016) is then to transform the data using the logarithmic transformation.

The histogram (Figure 8) of the transformed observations shows the normality (normal distribution) in the observation. The log transformed data was used for spatial statistical analysis.

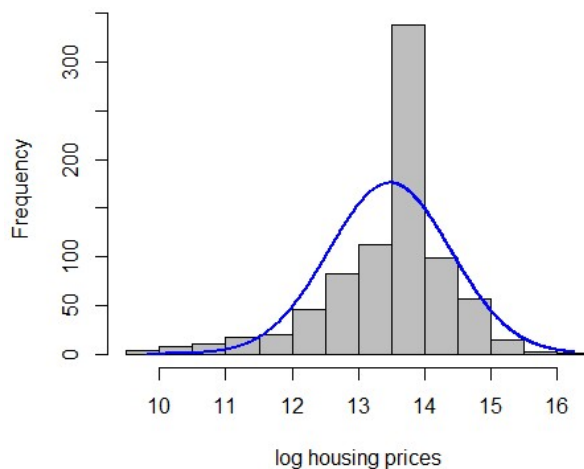


Figure 8: Histograms of log transformed housing price observations

The scatterplot (Figure 9) trendline shows a downhill pattern, this then indicates negative relationship between the housing prices and unemployment. As the X-values increase, (moving in right hand direction) the Y-values are likely to decrease, downward movement.



Figure 9: Scatterplot of transformed housing prices versus unemployment

4.3.2 Mapping of the selected variables

To explore the data further the visual aids like choropleth maps were employed and geographical distribution of different variables was generated based on the Jenks natural breaks classification in ARCGIS ArcMap 10.6. Choropleth maps are helpful in assessing if clustering or randomness of the phenomena under investigation exists or not.

The randomness in the spread of the data is present when there are no patterns of clusters of similar or dissimilar values. The Jenks method used determines the best arrangement of values into different data classes (ESRI, 2017). The spatial distribution is illustrated with series of GIS based maps, Figure 10 to Figure 16.

Figure 10 shows the distribution of 2011 unemployment in sub-places of CoJ municipality. The higher number of those unemployed is concentrated mostly in the south and south-western part of the municipality Region D, F, G and with some parts in region A as well. The sub-places with high unemployment are Soweto, Eldorado, Johannesburg central, parts of Alexandra, Cosmo city, Diepsloot, Ennerdale south, Lawley and Ivory Park, cluster areas of high values. Regions B, C, and some parts of A and E experience low number of unemployment.

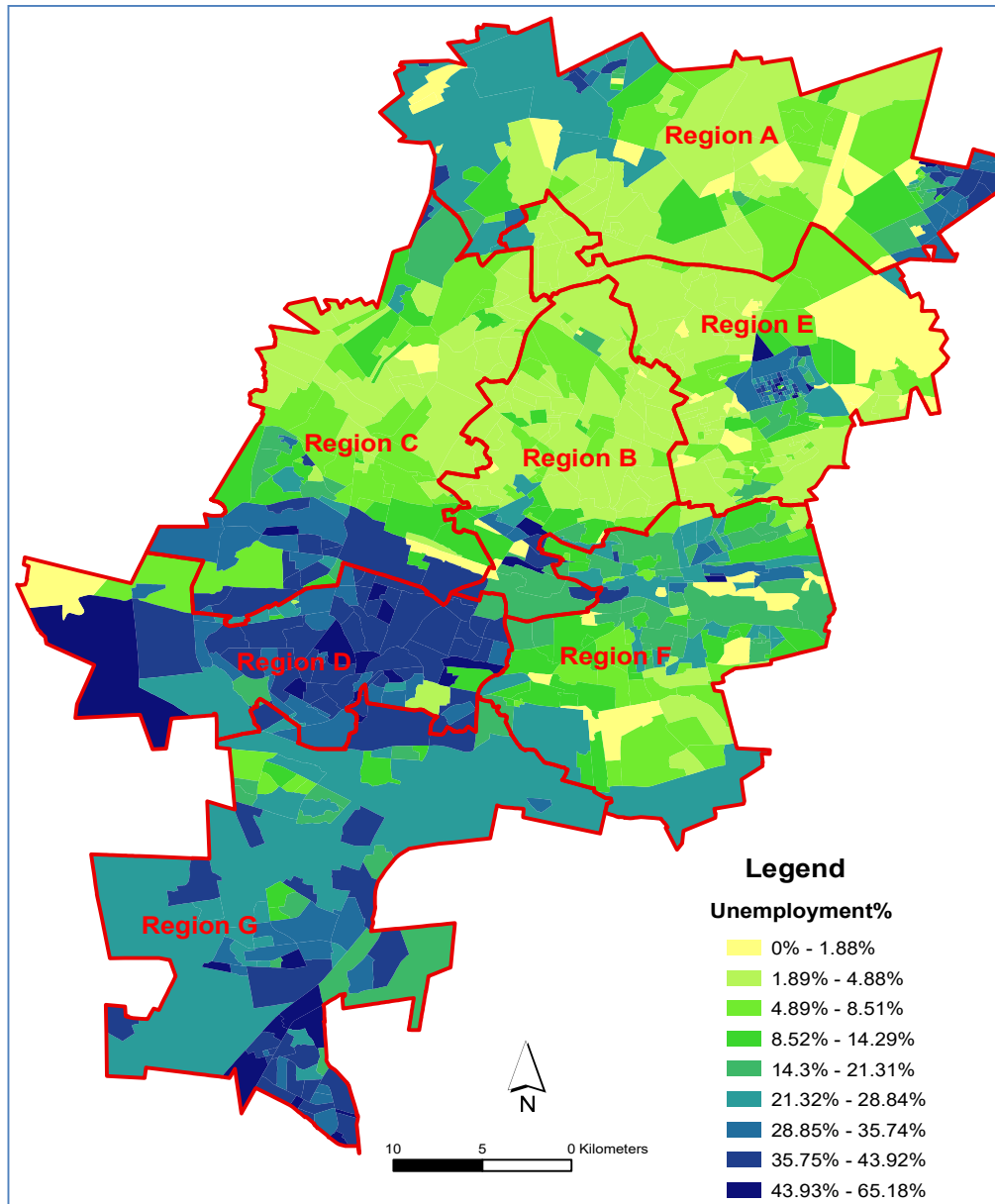


Figure 10: Distribution of the 2011 unemployment in CoJ
Data Source: Census 2011

Figure 11 shows that the households with lowest income (by lighter yellow colour) are mostly in region D and G. Region B (Sandhurst, Saxonwold, Northcliff), Region A (Dainfern Golf Estate, Kyalami Estate, Blue Hills AH, Dainfern Valley Estate) and western side of region E (Morningside, Bryanston, Atholl) experience higher incomes as shown in darker blue. The income includes all sources of income including salaries, social grants and any other remittances. Bryanstan, Lone Hill, Sunninghill and Kyalami Estate are sub-places in the municipality that are experiencing the highest income.

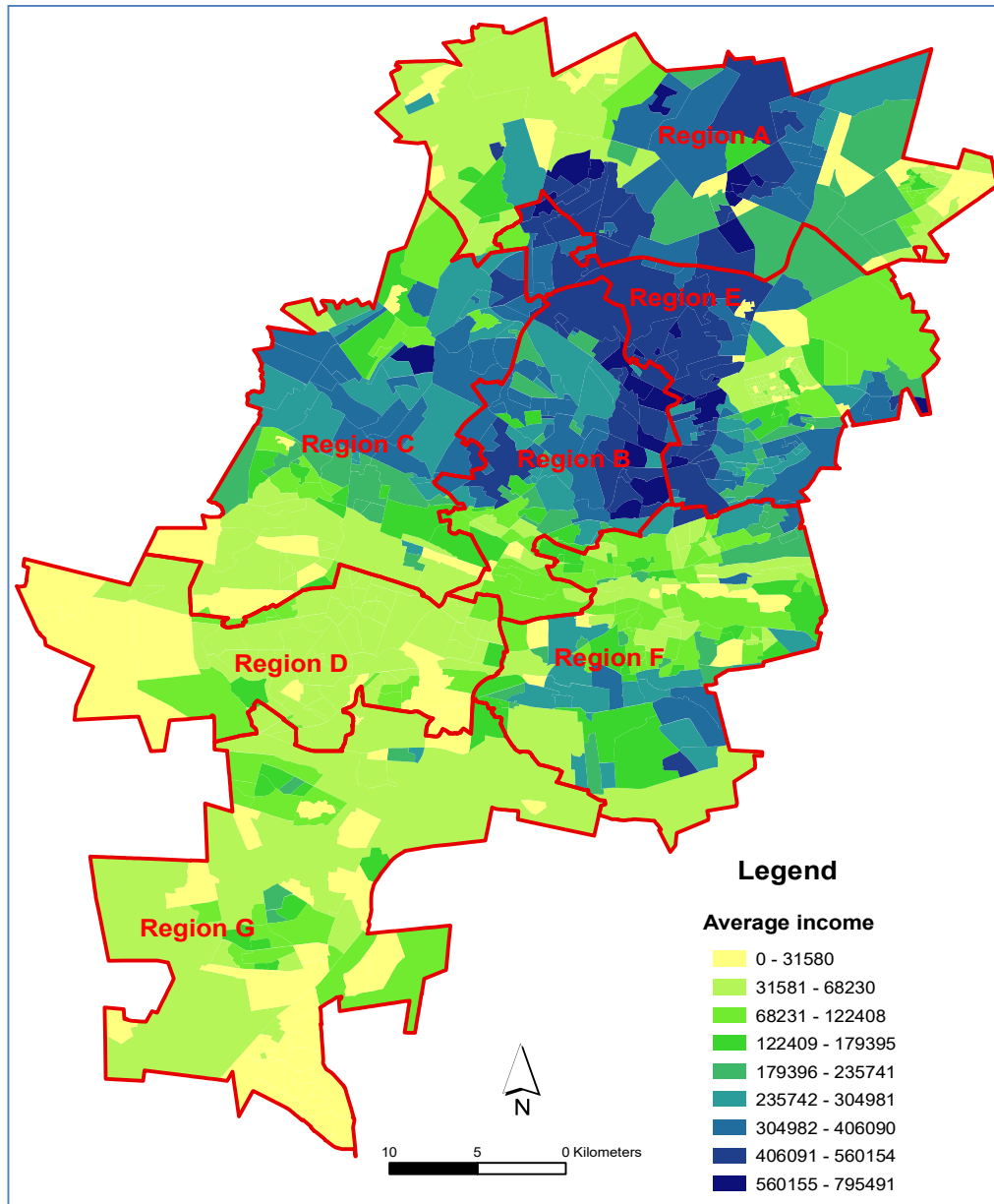


Figure 11: Distribution of 2011 average household income of Col
Data Source: Census 2011

Figure 12 shows that most of the regions in 2012 experienced low number of new constructions. Stesa agricultural holdings in region D, Cosmo city in region C, Bryanstan in region B and Noordwyk in region A for example are experiencing high number of new constructions.

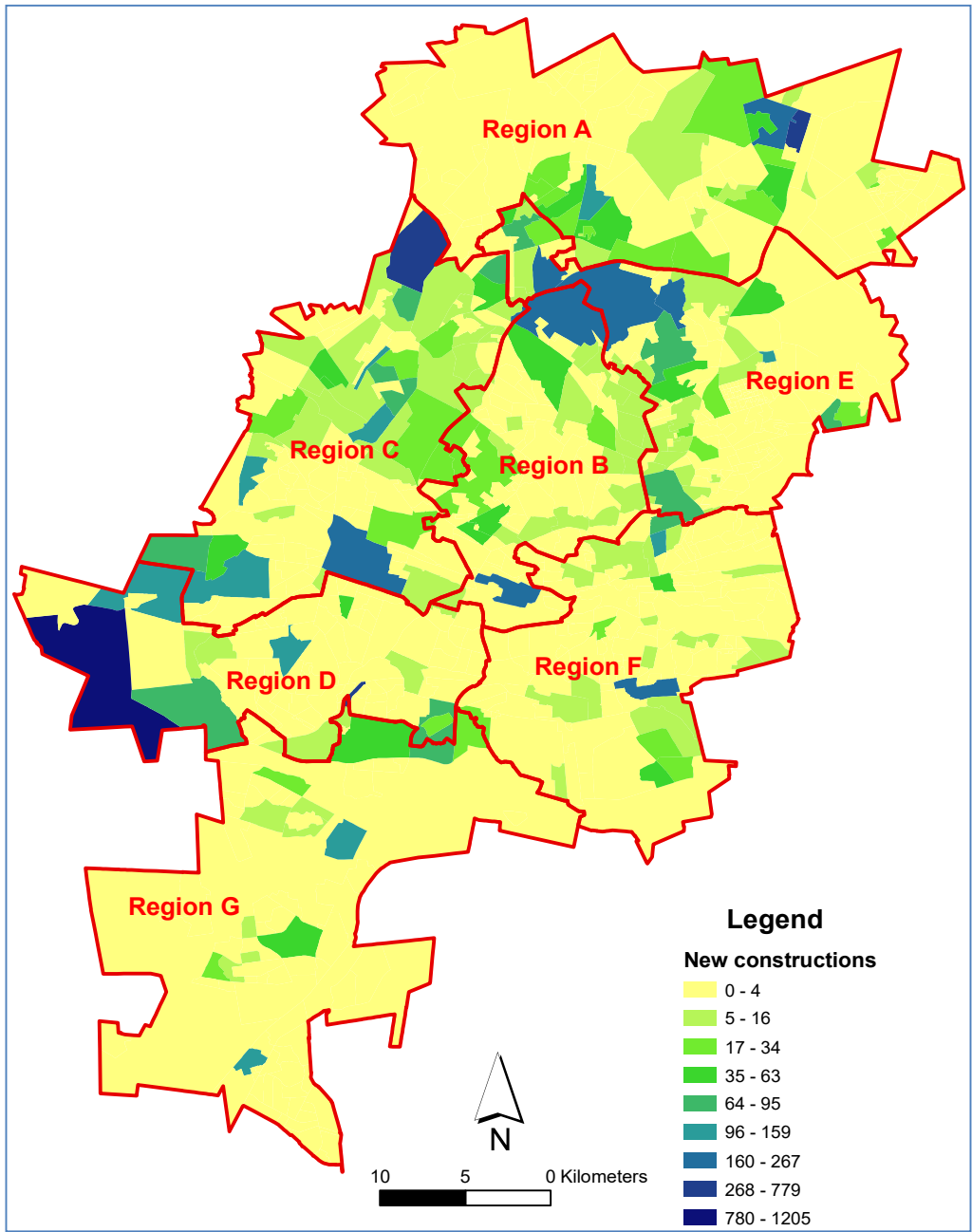


Figure 12: Distribution of 2012 new constructions in Col
Data Source: CAHF 2012

Figure 13 shows that most of properties that are exempted from the 2012 transfer duty fees are depicted in the south and south western part of the municipality. These are the properties that fall in the R0 – R 600 000 price band. The darker blue colour represent the transfer duty fee paid for the properties with more than R 1 500 000 price band. High values cluster with high values in the northern part of the municipality Region B and E and are located e.g. in Northcliff, Kyalami, Bryanston, Parktown, Glen Austin AH, Modderfontein AH, Risipark and Midrand.

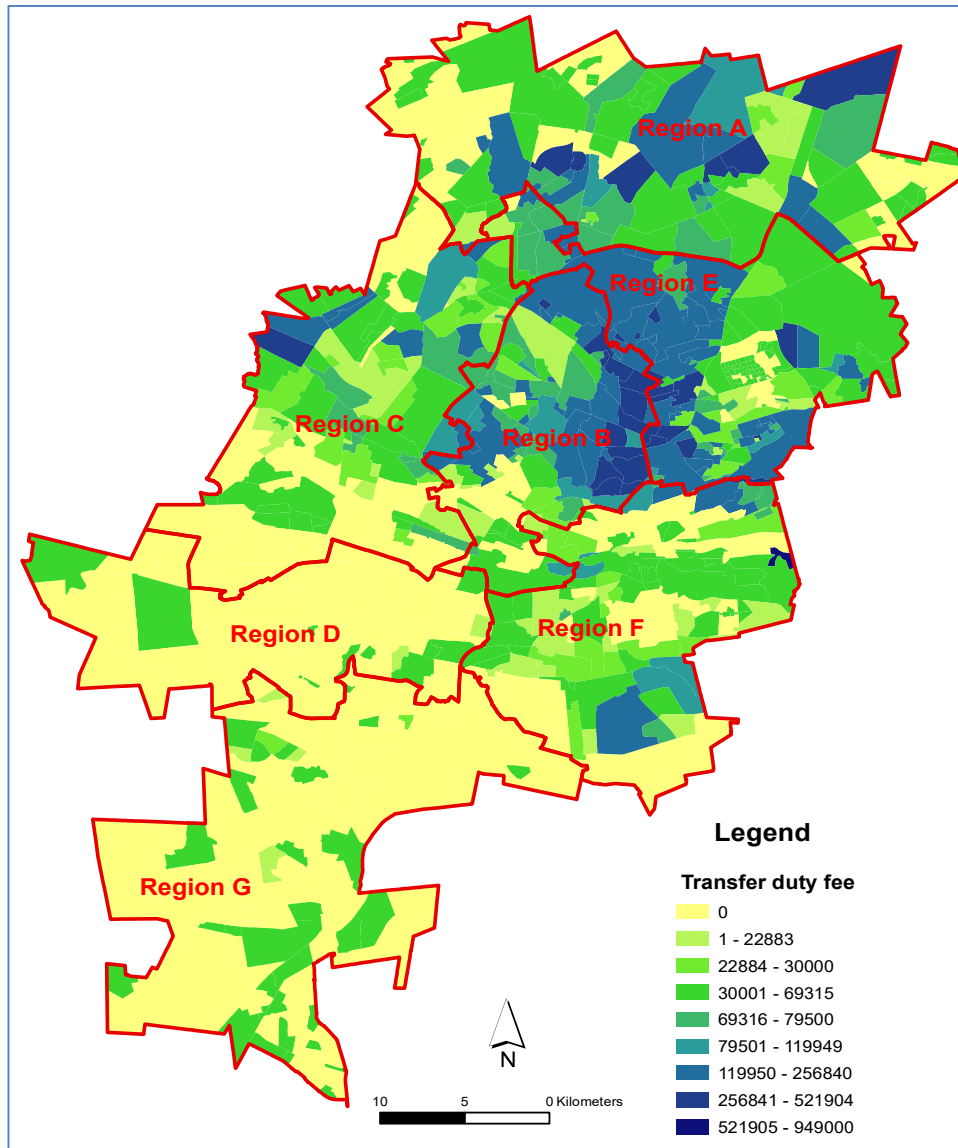


Figure 13: Distribution of 2012 transfer duty of CoJ
Data Source: SARS website

Figure 14 shows that in 2012 more properties were sold in Region D (Stesa AH, Protea Glen), Region C (Cosmo city, Weltevreden park), Region A (Noordwyk), Region B (Sandton, Bryanstan) shown with darker blue. Stesa AH has the highest number, with 1208 properties sold. The southern part of the municipality, Region G experiences low number of sold properties.

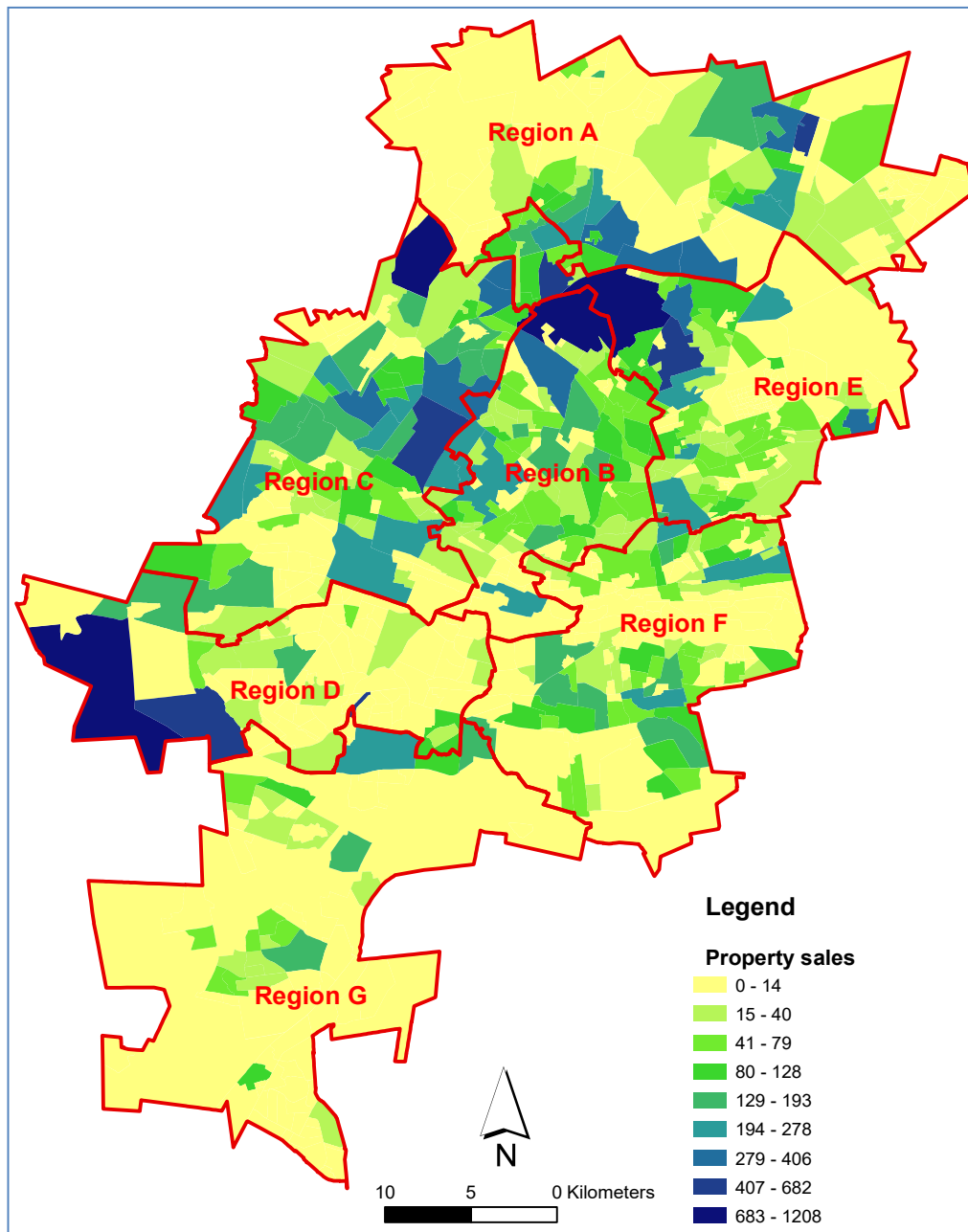


Figure 14: Distribution of 2012 properties sold in CoJ

Data Source: CAHF 2012

Region D (Dube, Mfolo, Orlando West), B (Westcliff, Westdene, Northcliff, Blairgowrie) and F (City Deep Gold Mine, Kensington, South Hills) are the regions with older properties in existence as compared to the other regions. Region A has most properties that are less than twenty-five years in existence.

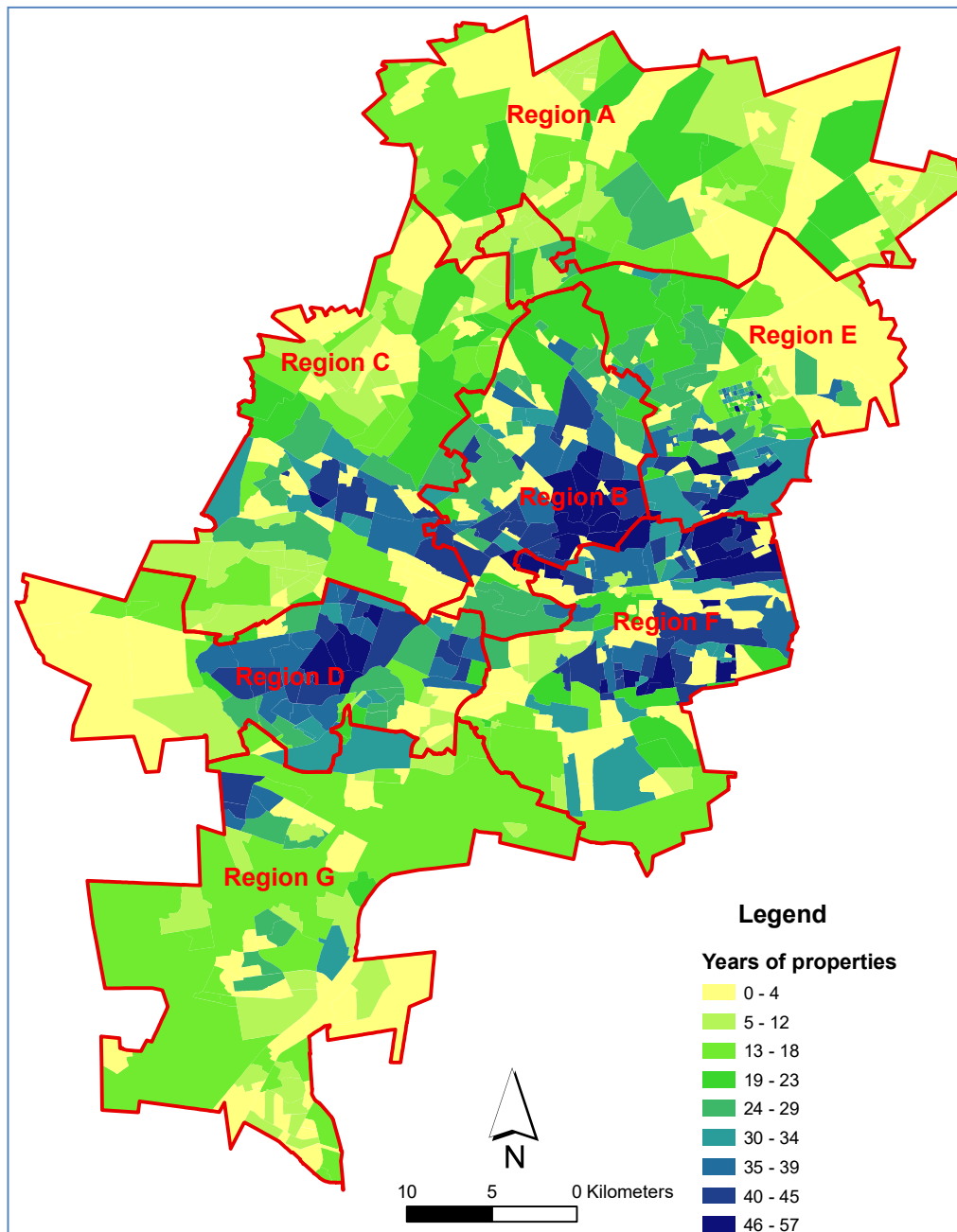


Figure 15: 2011 Number of years the properties in existence in CoJ

Data Source: Census 2011

The amenities have an impact on the city-level service demand. Most available amenities in 2012 for CoJ were education, entertainment, commercial, transportation and stores; and lowest in number was for the tourism attractions, Figure 16.

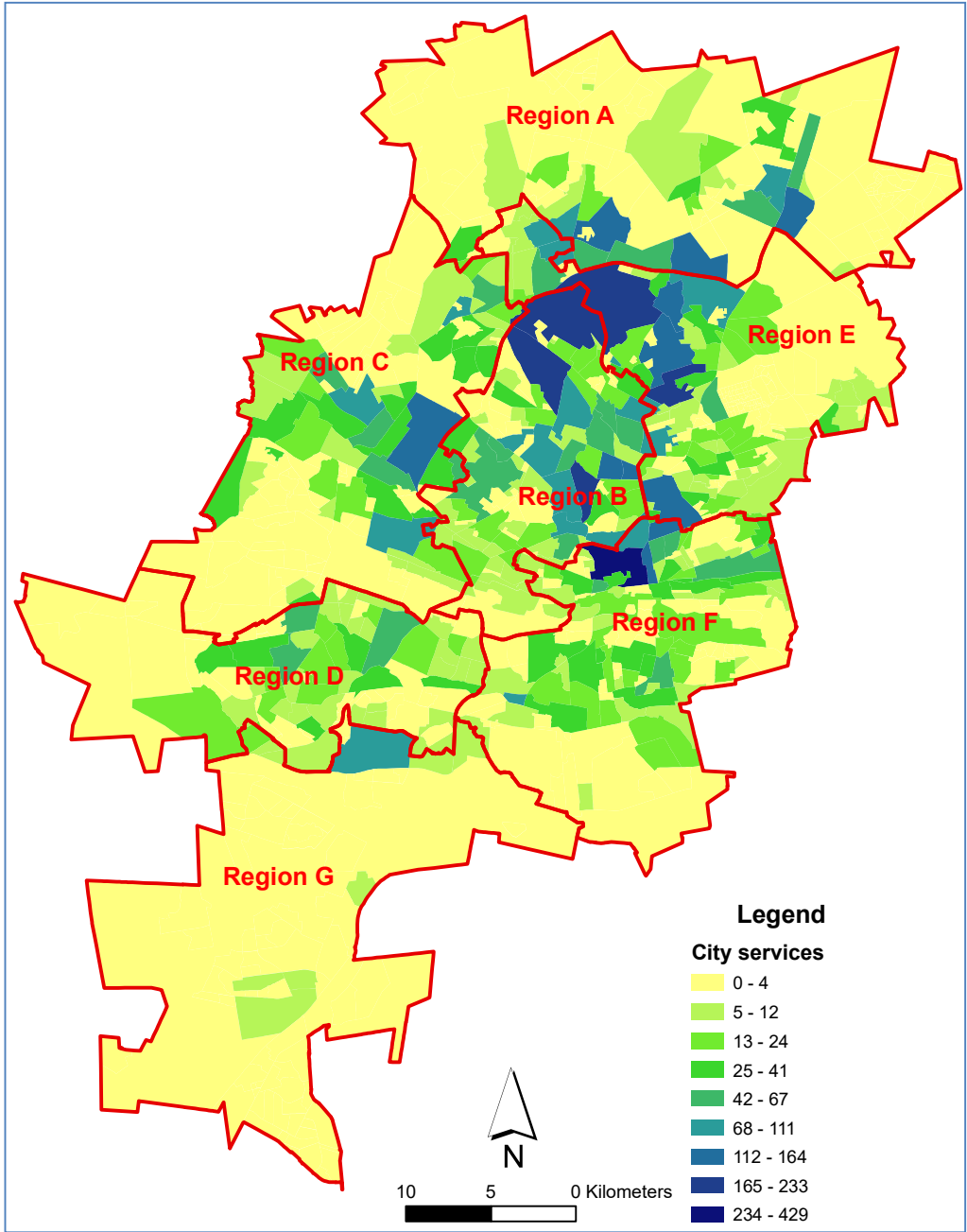


Figure 16: Distribution of the total number of the 2012 different city services
Data Source: AFRGIS 2012

4.4 Modelling Spatial dependence and clustering

4.4.1 Moran's I statistic test

Spatial auto correlation is mostly measured by calculating Moran's I and values ranged from -1 to +1 (Huang et al., 2017). The value of +1 meant perfect positive spatial autocorrelation where high values or low values cluster together, and -1 meant perfect negative spatial autocorrelation, and 0 meant perfect spatial randomness (Tu and Xia, 2008). The p-value is the indicator of whether the spatial autocorrelation is statistically significant and it is associated with z-score (Huang et al., 2017). Mou et al. (2017) states when high Moran's I value is obtained this means that there is much dependency on residuals. When there is dependency on the residuals this means that the change in the dependent variable cannot be explained sufficiently by the model. The smaller Moran's index the better the model would be.

Global Moran's I found to be 0.4608255823, with $p < 0.05$, shows significant positive spatial autocorrelation was observed for response variable, housing prices, indicating the presence of possible spatial patterns in their spatial distribution (Fu et al., 2010), Table 3.

Table 3: The results of spatial correlation test

Moran I test under randomisation	
Moran's I statistic	0.4608255823
Expected index	-0.0012453300
Variance	0.0004676406
p-value	2.2×10^{-16}
z-score (std dev)	21.367

Moran's I scatterplots (Figure 17) uses four quadrant to map the number of observations. Each quadrant shows the type of specific space association:

The **top-right** is quadrant one with positive spatial autocorrelation. This quadrant shows high values above the mean surrounded by high values. This is a high-high (HH) structure.

The **bottom-right** is quadrant two with negative spatial autocorrelation. This quadrant has higher values surrounded by low values. This is a high-low (HL) structure.

The **bottom-left** is quadrant three with positive space autocorrelation. This quadrant shows lower values surrounded by low values. This is low-low (LL) structure.

The **upper-left** is quadrant four with negative spatial autocorrelation. This quadrant shows low values surrounded by high values. This is a low-high (LH) structure.

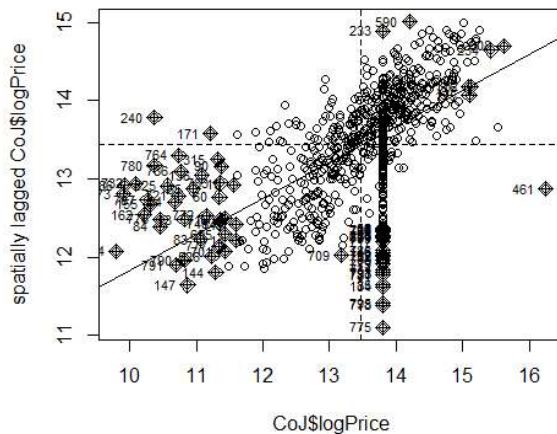


Figure 17: Moran's I scatterplot for transformed housing prices

The Moran's scatterplot plots different patterns such as, outliers and transition regions, (Negreiros et al., 2010). The plot Figure 17 highlights that there is a high-high and a low-low cluster pattern and the low-high outlier. The slope of the line indicate a global positive spatial auto-correlation

4.4.2 Geary's C statistic test

By using the log transformed housing prices, the obtained Geary's C statistic is 0.548992287 with the indication of positive spatial correlation within the housing prices in the neighbouring sub places with similar values together. The values of Geary's C statistic are in the range of 2 to 0, with the value 2 means perfect spatial correlation and 1 means perfect dispersion, and 0 means perfect spatial randomness. 0-1 indicates positive correlation where similar values are together, and 1-2 indicates negative correlation where different values are close together. The spatial dependence is evident in the housing prices across the sub places of City of Johannesburg. We reject the null hypothesis as p-value (0000) is less than 0.05 then we conclude that there is spatial dependence. The Moran's Index and Geary's C tests both reveal that there is positive correlation among the housing prices.

4.4.3 Local Moran's I

The Local Moran statistic was suggested by Anselin (1995) as a means to recognize local clusters (concentration of high values, concentration of low values) and local spatial outliers at each specific location, Figure 18. The hot spot analysis tool was used to identify housing clusters, whether the houses with high or low prices cluster together. A hotspot is known as a cluster of a particular interest (houses with high or low prices) in growth and change analysis (Poudyal et al., 2016) and the example of the hotspot would be a cluster showing a group of adjacent sub places with high-high values.

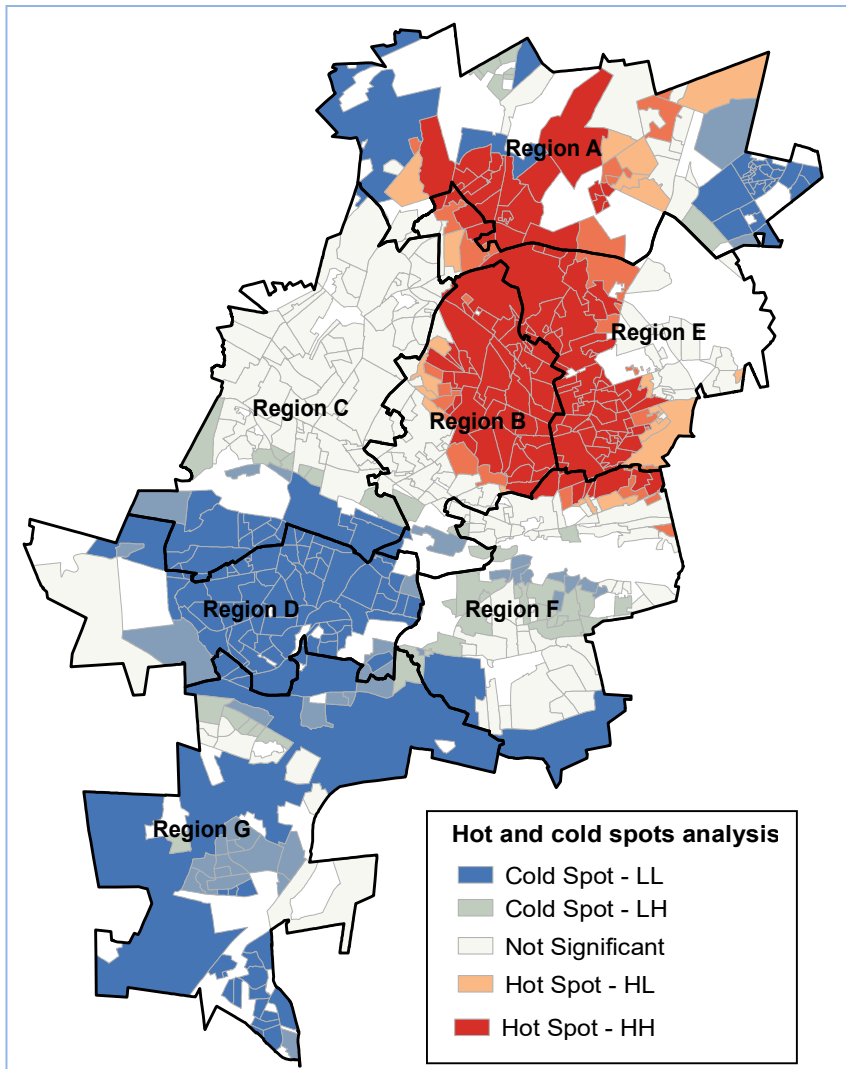


Figure 18: Spatial clustering and spatial outliers' map of housing prices at sub place level

The hot spot analysis tool Getis-Ord G_i^* (in ESRI ARCMAP) was used to identify housing clusters, whether the houses with high prices cluster together or low prices cluster together.

The statistical significance reported in the results is automatically adjusted using False Discovery Rate (FDR). This is a correction method for the multiple testing and spatial dependence in the optimised hot spot analysis tool (ESRI, 2017). The areas with dark blue colour (Figure 18) shows low values are clustering together, and areas in red shows high values clustering together. The no significant areas shown in white, this happens when the values are near zero indicating that there is no clustering of either values high or low.

The local Moran's reveals that the high-high spatial structure is encountered in the Region B (Sandton and Randburg area), as this area is famous for its suburban developments. Very small high-high spatial cluster was indicated in northern direction from Region E in Fourways and Sunninghill area. The very high-high spatial cluster is indicated in red which means that housing prices which have high values are likely to be clustered together. The low-low cluster where the ones with low values are likely to cluster together in region D and G, Soweto and Orange farm area. The low-high outliers were mostly located away from the high-high hot spot in the region F represented in grey colour.

4.5 Model selection using OLS regression

Regression method continues to be the preferred technique for investigating the effects of determining factors of housing prices at various levels (Wang et al., 2017) and OLS is commonly preferred to explore the frequently used to estimate the underlying determining factors. The regression equation is expressed in the following form:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{for } i = 1 \dots N$$

Where y represents a vector of observations of the dependent variable and β_0 is the intercept parameter term and β_1 is the slope parameter for variable X_i , ε_i is the error term (Bernasco and Elfers, 2011).

4.5.1 Model fitting

The association among housing prices (dependent variable) and the explanatory variables is quantified by performing a standard linear regression to find the best fit. Ordinary least squares attempts to minimise the residuals. Residuals are known to be the variance in the observed response and its fitted value by the linear regression (Kamenetsky et al., 2019). The Adjusted R-squared (model coefficient of determination) of the model indicate 75% of variation in the response variable (housing price) can be explained. The following variables renting, unemployment, new construction, number of property sales, number of approved plans, population density, and population ratio are negatively associated to the housing prices. The household income, renting, mortgage not paid off, city services, transfer duty, unemployment, number of property sales, and property age are significantly related to housing prices.

4.5.2 Model selection

In model selection we search for model that contains lesser independent variables that are also significant. The backward elimination (stepwise variable selection) method was applied. This method first minimises the Akaike's Information Criterion (AIC) and obtains lowest AIC which then guide in choosing the model parameters that are contained in the regression model. In the last step of performing the backward elimination, the reference model had an AIC value of 1221.56. The lower the AIC the better the model (Brazil, 2020). The below model (Table 4) will be the final model as the better fit.

Table 4: Stepwise procedure results

Step: AIC=-1221.56

$\text{CoJ\$logPrice} \sim \text{CoJ\$logInc} + \text{CoJ\$logRent} + \text{CoJ\$logNpaid} + \text{CoJ\$logSERV} + \text{CoJ\$logDuty} + \text{CoJ\$logUnemp} + \text{CoJ\$logSales} + \text{CoJ\$logAges}$

	Df	Sum of Sq	RSS	AIC
None>			172.06	-1221.56
CoJ\$logNpaid	1	2.951	175.01	-1209.89
CoJ\$logSERV	1	4.255	176.32	-1203.92
CoJ\$logAges	1	5.044	177.11	-1200.33
CoJ\$logSales	1	5.107	177.17	-1200.04
CoJ\$logRent	1	6.212	178.27	-1195.04
CoJ\$logUnemp	1	7.17	179.23	-1190.74
CoJ\$logInc	1	15.435	187.5	-1154.49
CoJ\$logDuty	1	180.875	352.94	-645.94

4.5.3 Model diagnostics

After the final model was selected, a model diagnostics was performed in order to evaluate the model assumptions, such as normality, linearity, equal variance and independence. To evaluate *linearity and equal variance*, the plot of residuals versus the fitted responses will be used (Kamenetsky et al., 2019). For *normality*, the quantile-quantile (QQ) plot of the standardised residuals will be employed.

4.5.3.1 Test for normality

Diagnostics plots in Figure 19 suggest that the residuals may be accepted to follow a normal distribution. Shapiro statistical test was performed to check this assumption. The hypotheses were formulated as stated below.

H_0 : The residuals do not follow a normal distribution,

H_A : The residuals follow a normal distribution

The p-value for the Shapiro normality test statistic was 0.933 with p value of less than 0.05. we reject the null hypothesis and conclude with alternative hypothesis that residuals follow normal distribution. The histogram of model residuals show a normal distribution.

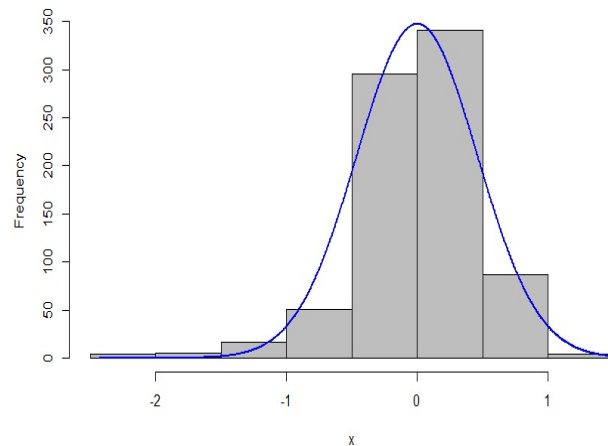


Figure 19: Histogram of model residuals

4.5.3.2 Test for linearity

Figure 20 below presents the commonly used diagnostics plots for residuals in the fitted model.

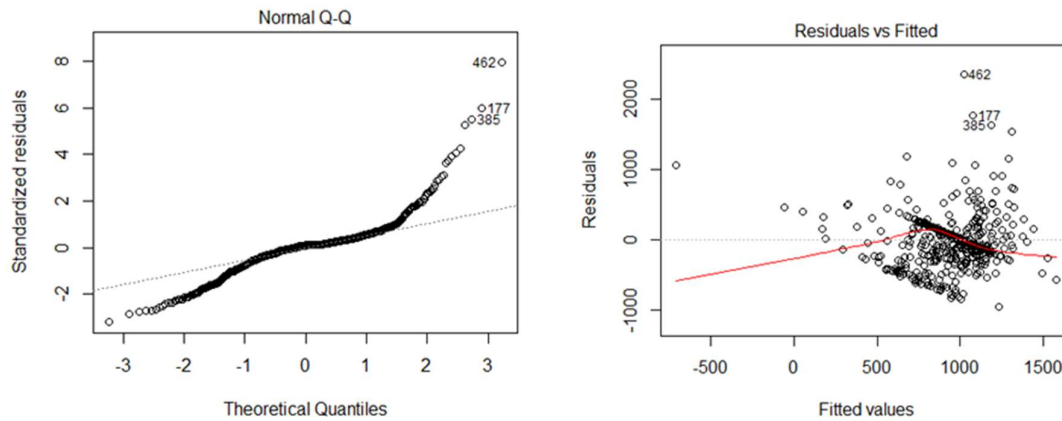


Figure 20: QQ-plots and fitted values

The normal Q-Q plot (Figure 20) showed a departure upward from the regression line on the right hand side end and downward departure at the lower end of the left side of the line, indicating right skewness in the residuals. The residuals versus fits plot is used to test the assumption that the residuals are randomly distributed, have constant (equal) variance along the regression line (i.e. homoscedasticity) and whether the relationship between variables is linear (i.e. linearity). Good residuals vs. fitted plot need to be with no obvious shape whereby there is no clear patterns in the data and no clear outliers (Hagerman, 2017). The residuals appear to be scattered fairly random without being closer to the regression line.

4.5.3.3 Test for multicollinearity in the data

The model's Variance Inflation Factor (VIF) was used to assess the multicollinearity and a value of 4.004 was obtained the value means that the independent variables are moderately correlated; this then violates the multicollinearity assumption that independent variables should not be correlated. VIF quantifies how much the variance is inflated. If the value of VIF is above 7.5 some variables are to be removed as they will be redundant (ESRI, 2017).

4.5.3.4 Test for auto-correlation in the data

To test for autocorrelation in the residuals from OLS a Durbin-Watson test is performed and a value of 1.8237 and p-value of less than 0.05 was obtained. This means that we reject the null hypothesis and conclude with alternative hypothesis that there is spatial dependence. This indicated that there is a positive correlation. The independence assumption violated. The

Durbin Watson test statistic is within 0 to 4, where 0 to less than 2 indicate positive autocorrelation, 2 to 4 indicate negative autocorrelation and 2 there is no autocorrelation.

Ho=there is no spatial dependence/autocorrelation for the housing prices

HA=there is spatial dependence/autocorrelation for the housing prices

The Durbin Watson test statistic is within 0 to 4, where 0 to less than 2 indicate positive autocorrelation, 2 to 4 indicate negative autocorrelation and 2 there is no autocorrelation.

4.5.3.5 Test for homoscedasticity

The Breusch-Pagan (BP) statistic test was conducted to check if the variance of the residuals were constant. The hypothesis for formal test of homoscedasticity were set up as follows;

H₀: The residuals of the OLS model have a constant variance (homoscedasticity)

H_A: The residuals of the OLS model does not have a constant variance (heteroscedasticity)

Breusch-Pagan Test was used to test for heteroscedasticity in a linear regression model and the test statistic of 120.04 with a p-value of less than 0.05 were calculated which then suggest that the null hypothesis was rejected and it was concluded that the residuals do not have a constant variance, heteroscedasticity is present in the model. Evidence of spatial autocorrelation on the residuals indicates that the variance is not constant in the dataset spatial regression is an alternative. When the variance of the error terms in the regression analysis is not constant, then heteroskedasticity is present (Breetzke, 2012). This violates the assumption that the observations are independent of one another.

4.6 Fitting the Spatial linear regression

4.6.1 Testing for type of spatial dependence present in the data

The Moran's I test statistic has high capability against extensive variety of spatial alternatives (Anselin, 2003). However, Moran's I does not provide guidance in the selection of which alternative models would be most appropriate. Lagrange Multiplier (LM) test statistic is applied in order to allow which model to choose between spatial error models (SEM) and spatial lag models (SLM). Moran's I test showed that there was spatial dependence present in the model. Further tests were conducted to check the evidence of spatial dependence. This was to check which kind of spatial dependence present in the data and Lagrange multiplier

diagnostics was conducted and the results shown below. The hypotheses tests were set up as follows:

H₀: The Lagrange multiplier for the spatial lag model is not significant

H_A: The Lagrange multiplier for the spatial lag model is significant

H₀: The Lagrange multiplier for the spatial error model is not significant

H_A: The Lagrange multiplier for the spatial error model is significant

H₀: The Lagrange multiplier for the spatial lag model robust is not significant

H_A: The Lagrange multiplier for the spatial lag model robust is significant

H₀: The Lagrange multiplier for the spatial error model robust is not significant

H_A: The Lagrange multiplier for the spatial error model robust is significant

Lagrange multiplier diagnostics for spatial dependence

Model: $\text{lm}(\text{formula} = \text{CoJ\$logPrice} \sim \text{CoJ\$logInc} + \text{CoJ\$logRent} + \text{CoJ\$logNpaid} + \text{CoJ\$logSERV} + \text{CoJ\$logDuty} + \text{CoJ\$logUnemp} + \text{CoJ\$logSales} + \text{CoJ\$logAges})$
weights: W

LMerr = 22.977, df = 1, p-value = 1.639e-06

LMlag = 19.712, df = 1, p-value = 9.004e-06

RLMerr = 7.6441, df = 1, p-value = 0.005696

RLMlag = 4.3786, df = 1, p-value = 0.03639

4.6.2 Spatial lag model

After identifying if the presence of spatial dependence existed, the spatial lag model was applied using the maximum likelihood approach. The results from the spatial lag model indicated that three predictors (independent), renting accommodation, unemployment, and number of property sales, are negatively associated with the housing prices. However all variables are significant variables influencing the housing prices. The AIC value is 1043.7 which is smaller than the AIC value of 1062.1 for the standard linear regression model.

4.6.3 Spatial error model

Spatial error model, as another option to spatial lag model it was then considered for comparison. A spatial error model incorporates spatial dependence in the errors. From calculated output (Table 5), it is noted that the estimate of the spatial correlation coefficient (Lambda) for error model is 0.2937. The Lambda is more than Rho (0.1465) for the spatial lag model. The AIC value is 1038.6 which is smaller than the AIC value of 1062.1 for the standard linear regression model. The log-likelihood for error model is -508.31 more than -510.84 for lag model.

4.7 Geographical detector technique

After determining the significances and directions of the impacts of the potential determining factors on housing prices, the magnitude of the impact of these factors was estimated using the geographical detectors model as proposed by Wang et al. (2017) which this model is based on spatial variance analysis (SVA) to assess the social economic factors impact. Factor detector (q-statistic) quantifies the impact of an environmental factor on the observed spatial disease pattern using power of determinant (PD) (Wang, 2012). $PD \in [0,1]$. Using the geographical detector technique the magnitude of the impact of these factors was investigated and estimated as shown below:

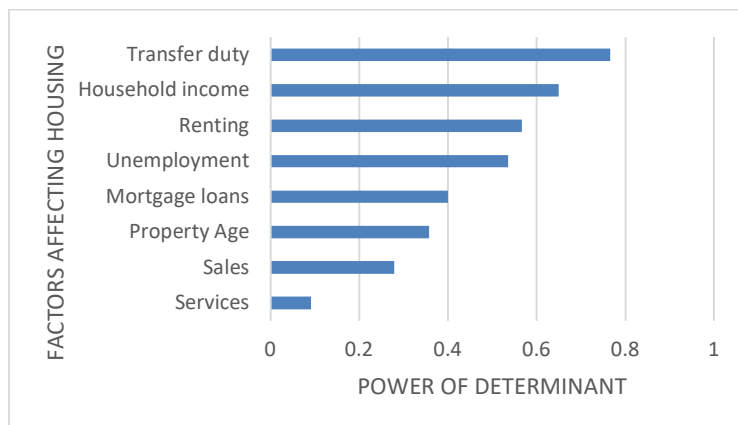


Figure 21: The power of determinant for the factors guiding the housing price effect

Table 5: Estimation results of regressions

OLS	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.213621	0.106780	114.381	< 2e-16 ***
CoJ\$logInc	0.079027	0.009358	8.445	< 2e-16 ***
CoJ\$logRent	-0.085665	0.015989	-5.358	1.11e-07 ***
CoJ\$logNpaid	0.058472	0.015834	3.693	0.000237 ***
CoJ\$logSERV	0.065490	0.014770	4.434	1.06e-05 ***
CoJ\$logDuty	0.117579	0.004067	28.909	< 2e-16 ***
CoJ\$logUnemp	-0.150044	0.026068	-5.756	1.23e-08 ***
CoJ\$logSales	-0.079821	0.016432	-4.858	1.43e-06 ***
CoJ\$logAges	0.064282	0.013316	4.827	1.66e-06 ***
Adjusted R-squared: 0.7465 p-value: < 2.2e-16				

SEM	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	12.2253134	0.1053288	116.0681	< 2.2e-16
CoJ\$logInc	0.0671504	0.0091593	7.3314	2.278e-13
CoJ\$logRent	-0.0845984	0.0164185	-5.1526	2.569e-07
CoJ\$logNpaid	0.0709504	0.0159964	4.4354	9.190e-06
CoJ\$logSERV	0.0619413	0.0152478	4.0623	4.859e-05
CoJ\$logDuty	0.1203706	0.0042004	28.6570	< 2.2e-16
CoJ\$logUnemp	-0.1172415	0.0266365	-4.4015	1.075e-05
CoJ\$logSales	-0.0829175	0.0163735	-5.0641	4.103e-07
CoJ\$logAges	0.0476297	0.0133246	3.5746	0.0003508
Lambda: 0.29379, Log likelihood: -508.306 for error model				
AIC: 1038.6, (AIC for lm: 1062.1)				

SLM	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	10.3349591	0.4096192	25.2307	< 2.2e-16
CoJ\$logInc	0.0670573	0.0093613	7.1633	7.878e-13
CoJ\$logRent	-0.0802953	0.0157219	-5.1072	3.269e-07
CoJ\$logNpaid	0.0661476	0.0155448	4.2553	2.088e-05
CoJ\$logSERV	0.0629837	0.0144865	4.3477	1.376e-05
CoJ\$logDuty	0.1109049	0.0043712	25.3716	< 2.2e-16
CoJ\$logUnemp	-0.1160084	0.0260801	-4.4481	8.661e-06
CoJ\$logSales	-0.0937525	0.0163559	-5.7320	9.923e-09
CoJ\$logAges	0.0560999	0.0130814	4.2885	1.799e-05
Rho: 0.14657, Log likelihood: -510.8459 for lag model				
AIC: 1043.7, (AIC for lm: 1062.1)				

CHAPTER 5 : THE DISCUSSION

5 Discussion

The study aimed at identifying the potential housing price determinants in the residential property market in the City of Johannesburg for the year 2012 with the objectives to explore the spatial patterns of housing prices, housing tenure per sub-place and the association of house prices with their potential determining factors. The approach was to employ both ordinary least square (OLS), spatial regression and geographical detector technique. The data analysis was carried out using spatial statistics in the RStudio statistical open source computing environment (version 1.1.456, <http://www.r-project.org>) and GIS software ARCGIS (version 10.6) for mapping.

5.1 Mapping of the housing prices and housing tenures

Choropleth maps were employed to visualise the spatial distribution of housing prices, different housing tenures and selected study variables. The classification thereof was based on the Jenks natural breaks classification as illustrated by series of maps in Figure 10 to Figure 16. The spatial pattern shown (Figure 3) was that the full title home-ownership type is the most favoured in the municipality. The full title ownership properties are located in most parts of the municipality. Most of sectional titles are located in the northern and eastern parts of the municipality. There are no sectional titles in Region G e.g. in Ennerdale, Orange farm area. Estate properties exist mostly in the high-end market and there are no Estates properties in region D western part of the municipality. The municipality experienced both renting housing and home-ownership housing. There were more number of households that preferred renting than buying through mortgage loans as shown in Figure 4. Households prefer renting because it costs less to rent a home than to own a home (Haury, 2020). This is due to additional expenses for example property taxes, property maintenance involved in owning a home. Housing prices tend to be higher in the sub-places with higher average income.

5.2 Measuring spatial auto correlation

From the results obtained from both the global Moran's I, local Moran's I and Geary's C tests confirm that there is significant positive spatial association within the modelled data. The obtained p-value is almost zero which shows that it is almost impossible for the data to be a result of random spatial process (Huang et al., 2017). The null hypothesis of no spatial autocorrelation is rejected and concludes on the alternative hypothesis that there is spatial dependence. The spatial dependence in the housing prices is evident across the sub places of City of Johannesburg. House prices are spatially autocorrelated for two general reasons because properties developed same time share structural features and location amenities (Basu and Thibodeau, 1998).

The clusters of spatial association was also examined to see sub-places that were hot or cold spots for housing prices. The spatial pattern of the housing prices was found to be characterised by spatial dependence and the hot spots are spatially clustered (Figure 18). The hot spot analyses suggested that significant clustering of housing prices, high-high spatial structure, is observed in the region B of Sandton and Randburg. A very high-high spatial cluster means that housing prices which have high values are likely to be clustering together. The low-low spatial cluster was observed in region D and G.

5.3 Factors affecting housing prices from regression models

OLS and spatial regression was employed to incorporate spatial effect in the model construction, this included error (nuisance) and the lag (substantive), and the spatial error model was the most option. The regression analysis results are shown in Table 5. The goodness-of-fit statistics such as AIC was used to estimate the fitting degree of regressions. After Akaike's Information Criterion was applied as a guidance in selecting the model parameters to be included in the regression model and eight variables were selected as follows: household income, renting, mortgage not paid off, city services, transfer duty, unemployment, sales and property age. The AIC for OLS was found to be 1221.56; which was higher than both SLM and SEM with the value 1043.7 and 1038.6 respectively. The obtained Adjusted R-squared indicated 75% of variation in the response variable of housing price is explained. All the explanatory variables were significant with the p-value less than 0.05 from

this standard linear regression. It has shown that a percent of unemployment is associated with approximately 0.15% decrease in house prices.

In addition, the robust Lagrange multiplier tests were undertaken and the results indicated that the spatial error model was more suitable than the spatial lag model for this study. The comparison shows that the data was fit using the spatial error analysis than OLS. From the Lagrange Multiplier test results, when both the LMerr and LMlag were assessed they were both significant with the p-value less than 0.05. This then necessitated the comparison of the robust forms. Results suggested that the error model is the more likely alternative one with the p-value of 0.005 compared to p-value of 0.036 for lag model. The p-values for all the tests further show that there was significant spatial dependence in the data and the spatial error regression model was fitted.

The SEM regression is appropriate when spatial effects (spatial autocorrelation) are incorporated in the error term while the SLM incorporates the spatial effects in the lag term of the model (Anselin et al., 1996). SLM regressions are appropriate when there is spatial autocorrelation in the dependant (y) term of the OLS regression equations (Voss et al., 2006). The results of the spatial error model indicate that three independent variables, number of sales, unemployment and renting are negatively and significantly associated with the housing prices whereas income, mortgages loans, city services, transfer duty and property ages are positively and significantly associated with the housing prices. The negative sign of the independent variable (regression coefficient) in the regression equation indicates that the presence of a unit increase in that variable will have a reducing effect on the housing price (Olowofeso et al., 2015). It has shown that a percent of unemployment is associated with approximately 0.11% decrease in house prices while other variables are kept constant. Household income had a positive relationship with the housing prices. A unit increase in household income result in increase in housing prices while other variables are held constant.

5.4 The estimated impact of factors

After identifying the significances and the directions of the effects of potential factors on housing, the impact was estimated. The geographical detector technique was used to analyse

the strength of the association between the selected variables and housing prices in CoJ. The geographical detector results showed that of the variables included transfer duty and household income contribute the most, that transfer duty and income can predominantly explain the spatial variability of housing prices followed by renting and unemployment while city services shown to have weak positive influence.

Figure 20 shows that factors influencing housing prices can be ranked by power of determinant value as follows; transfer duty (0.7658), household income (0.6499), renting (0.5670), unemployment (0.5355), mortgage loans (0.3999), property age (0.3572), property sales (0.2785), and city services (0.0911). The transfer duty constitutes major portion of the costs involved when purchasing the property. The greater the value of the property, the higher the duty that needs to be paid. Housing prices are affected by income because income relates to households buying power. The higher the income, the greater the capacity to pay for the potential houses. Income was found to be significant with positive influence in the spatial regression results and it was also found to have a second higher power of determinant (0.6499).

De Bruyne et al. (2013) highlighted that housing prices will be driven downwards by the higher unemployment experienced by the municipality. Spatial regression (spatial error) reveal that unemployment has exerted negative strong influence in relation to housing prices. The higher the unemployment implies less earning for the households which means that there would be less available money for housing (Ge and Williams, 2015). While in CoJ different property ownership is encouraged, renting is considered transitional option and this drives the housing demand. The shortage of housing will put upward pressure on the price of rented accommodation. The results for renting reveal that it is significant and have negative correlation on housing prices. Homeowners with mortgage that is not paid off they are still have a burden on repayments to keep up with their contractual obligation of repaying the mortgage. Mortgage loan has a positive association with housing prices. The results of geographical detector indicate that this property purchasing activity has also exerted a negative weak influence to the housing prices of CoJ. It has been shown that many housing characteristics and neighbourhood/city amenities have a measurable and statistically significant impact on house prices (Duijn and Rouwendal, 2012).

CHAPTER 6 CONCLUSION

6 Conclusion and recommendation

An attempt has been made to identify the determinants of housing prices in the City of Johannesburg. The 2012 housing prices dataset at the sub-place level was used and the variables were categorised into housing demand and housing supply. The regression models and geographical detector technique was employed in order to identify the magnitude of the impact exerted by significant factors on housing. Eight variables were selected to be fitted in the final regression model. The research has identified the factors that affect residential house prices and explored their impact. The developed models suggest that the main drivers of house prices in CoJ property market are property transfer duty, household income, renting and unemployment. The contribution of each variable was quantified and it was possible to determine the significance of each variable. This significance then provides assistance in the decision making process on which factors to be considered when developing and buying a house. In addition this significance could provide the public authorities with assistance in determining the real estate tax rates effectively taking into consideration the characteristics of the nearby neighbourhoods.

Further research on exploring factors like stand (land) prices and transfer costs that influence housing prices is needed. These data are not recorded by the Deeds office and this is where the Property agencies which control housing information would be able to assist in this regard. Land is one of the key determining factors of economic property.

7 Limitations of the study

The study was limited in terms of available data and cost associated with some of the available data. Variables data for example the crime related e.g. burglaries, transfer costs and land/stand prices were not available and also not available in the level of interest. Ideally it would be good for this study if these data were available in Census sub-places in order to assess their impact on housing prices.

8 References

- Amadeo, K. (2019). Inflation, how it's measured and managed?. [Online] Available from: <https://www.thebalance.com/what-is-inflation-how-it-s-measured-and-managed-3306170>
- Anselin, L. (2003). An introduction to spatial regression analysis in R. [Online]. Available from: <http://sal.agecon.uiuc.edu>
- Anselin, L., and Getis, A. (1992). Spatial statistical analysis and geographic information systems. *The Annals of Regional Science*. 26:19-33.
- Arnold K., le Roux, A. and Hattingh, M. (2017). Impact of Gautrain stations on property prices and sales activity in the City of Johannesburg between 2006 and 2015. *South African Journal of Geomatics*, Vol. 6. No. 2, 184-195.
- Baffoe-Bonnie, J. (1998). The dynamic impact of macroeconomic aggregates on housing prices and stock of houses: A National and Regional Analysis. *Journal of Real Estate Finance and Economics*, 17:2, 179-197
- Basu, S. and Thibodeau, T. (1998). Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, Vol 17:, 61-85
- Bourassa, S. C., Cantoni, E., and Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *Journal of Real Estate Finance and Economics*, 35(2), 143–160.
- Brazil, N. (2020). Spatial regression. GEO 200CN: Quantitative Geography. [Online] Available from: <https://geo200cn.github.io/spatialreg.html>
- Brunsdon, C., Fotheringham, A. S. and Charlton, M. (1996). Geographically Weighted Regression: A method for exploring spatial nonstationarity. *Geographical Analysis*. Vol. 28, No. 4.
- Can A. (1998). GIS and Spatial Analysis of Housing and Mortgage Markets. *Journal of Housing Research*: 1998, Vol. 9, No. 1, pp. 61-86.
- Centre for Affordable Housing Finance in Africa (CAHF). (2010). 2010 Yearbook Housing Finance in Africa: A review of some of Africa's housing finance markets. 1st edition. <http://housingfinanceafrica.org/documents/2010-yearbook-housing-finance-in-africa/>
- Centre for Affordable Housing Finance in Africa. (2018). Understanding Housing Markets: using municipal data. <http://housingfinanceafrica.org/app/uploads/Understanding-Municipal-Housing-Markets.pdf>

- Centre for Affordable Housing Finance in Africa (CAHF) and South African Cities Network (SACN). (2014). Findings Report: Understanding Housing Markets in Johannesburg, South Africa. [Online] Available from: <http://housingfinanceafrica.org/app/uploads/File-4.-JHB-City-Report-FINAL.pdf>.
- Clark, A.E., and Daniel, T. (2006) Forecasting South African house prices. *Investment Analysts Journal*, 35:64 27-33.
- Cohen, J.P., Ioannides, Y.M., and Thanapisitikul, W. (2016). Spatial effects and house price dynamics in the USA. *Journal of Housing Economics*.
- De Guzman M. (2018). South Africa's housing market accelerates. [Online]. Available from: <https://www.globalpropertyguide.com/Africa/South-Africa/Price-History> [20 May 2019]
- De Bruyne, K and Van Hove, J. (2013) Explaining the spatial variation in housing prices: an economic geography approach, *Applied Economics*, 45:13, 1673-1689.
- De Wet, T., Plagerson, S., Harpham, T. and Mathee, A., (2011). Poor housing, good health: A comparison of formal and informal housing in Johannesburg, South Africa. *International Journal of Public Health*, 56:6 625-633.
- Franken, M. (2007). Determinants of the South African residential real estate growth cycle. Unpublished dissertation Master of Business Management. Stellenbosch University.
- Friedman, S. Tsao, Hui-shien & Chen, C. (2013). Housing Tenure and Residential Segregation in Metropolitan America. *Demography*, 50:1477–1498.
- Gaspareniene, L., Remeikiene, R. and Skuka, A. (2016). Assessment of the impact of macroeconomic factors on housing price level: Lithuanian case. *Intellectual Economics* 10: 122–127
- Ge, X.J. (2009). Determinants of House Prices in New Zealand. *Pacific Rim Property Research Journal*. 15(1):90-121.
- Goodman, A.C. and Thibodeau, T.G. (2007). The Spatial Proximity of Metropolitan Area Housing Submarkets. *Real Estate Economics*. 35:2, 209–232.
- Hagerman, I. (2017). Residual Plots Part 1 — Residuals vs. Fitted Plot. Data Distilled
- Haury, A. (2020). 10 reasons why renting could be better than buying, owning a home isn't right for everyone. [Online]. Available from: <https://www.investopedia.com/financial-edge/1112/reasons-renting-is-better-than-buying.aspx>
- Huang, Z., Chen, R., Xu, D. and Zhou, W. (2017). Spatial and hedonic analysis of housing prices in Shanghai. *Habitat International*, 67, 69–78.

- Kamenestsky, M., Chi, G., Wang, D. and Zhu, J. (2019). Spatial regression analysis of poverty in R. *Spatial Demography*, 7: 113 - 147
- Koramaz, T.K. and Dokmeci, V. (2012) Spatial Determinants of Housing Price Values in Istanbul, *European Planning Studies*, 20:7, 1221-1237.
- Kramer L. (2019). How Does the Law of Supply and Demand Affect Prices? Available Online <https://www.investopedia.com/>
- LeSage, J. and Pace, R.K. (2009). Introduction to Spatial Econometrics. *CRC Press, Boca Raton*, 33--42, 114--115
- Li, S., Ye, X., Lee, J. and Gong J., and Qin, C. (2016). Spatiotemporal Analysis of Housing Prices in China: A Big Data Perspective. *Applied Spatial Analysis*
- McLeod, S. A. (2019). What is kurtosis? [Online] Available from <https://www.simplypsychology.org/kurtosis.html>
- Moreira de Aguiar, M., Simoes, R. and Golgher, A. (2014). Housing market analysis using hierarchical-spatial approach: the case of Belo Horizonte, Minas Gerais, Brazil. *Regional studies, Regional Science*. 1:1, 116-137.
- Mushongera, D. (2017). Beyond GDP in assessing development in South Africa: The Gauteng City-Region Socio-Economic Barometer. 34(3):330-346.
- Negreiros, J.G., Painho, M.T., Aguiar, F.J. and Aguiar, M.A. (2010). A comprehensive framework for exploratory spatial data analysis: Moran location and variance scatterplots. *International Journal of Digital Earth*, 3:2, 157-186
- Nilsson P. (2017). Are valuations of place-based amenities driven by scale?. *Housing Studies*, 32:4, 449-469,
- Nneji, O., Brooks, C. and Ward, C.W.R. (2013). House price dynamics and their reaction to macroeconomic changes. *Journal of Economic Modelling*. 32, 172-178.
- Olowofeso, E and Oyetunji, A.K. (2016). Assessing the impact of selected macroeconomic variables in the determinants of sustainable residential housing prices in Lagos. *Ethiopian Journal of Environmental studies and Management*, 9(4): 405–411.
- Reibel, M. and Agrawal, A. (2007). Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review*, 26(5–6), 619–633.
- Ruonavaara, H. (2007). Types and forms of housing tenure: Towards solving the comparison/translation problem. [*Scandinavian Housing and Planning Research*](#).

- Ryan, S. and Enderle, B.E. (2012). Examining spatial patterns in affordable housing: the case of California density bonus implementation, *Journal of Housing and the Built Environment*, 27:413–425
- Selim, H. 2009. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*.
- Shaw, B. (2019). SA residential property market remains flat with growth predicted only late in 2020. [Online] Available from: <https://www.iol.co.za/business-report/economy/sa-residential-property-market-remains-flat-with-growth-predicted-only-late-in-2020-38923617>. Accessed on 25 February 2020
- Standish, B., Lowther, B., Morgan-Grenville, R. and Quick, C. (2005). The determinants of residential house prices in South Africa. *Investment Analysts Journal*, 34:61, 41-48.
- Tafirenyika, S. and Muzindutsi, P. (2017). Determinants of house prices and new construction activity: An empirical investigation of the Namibian housing market. *The Journal of Developing Areas*. 389-407
- Todes, A., Weakley, D. and Harrison, P. (2018). Densifying Johannesburg: context, policy and diversity. *Journal of Housing and the Built Environment*. 33:281–299
- Tsatsaronis, K. and Zhu, H. (2004). What drives housing price dynamics: cross-country evidence. BIS Quarterly review, March 2004.
- Tu, Y., Sun, H., and Yu, SM. 2007. Spatial Autocorrelations and Urban Housing Market Segmentation *Journal of Real Estate Finance and Economics*, 34:385–406.
- Turok, I., and Visagie, J., (2018). Does moving to a city mean a better life? New evidence. [Online]. Available from <http://www.econ3x3.org/article/does-moving-city-mean-better-life-new-evidence> accessed 24 February 2020
- Wang, Jin-Feng and Hu, Y. (2012). Environmental health risk detection with GeogDetector. *Environmental Modelling and Software*, 33:114-115.
- Wang, Y., Wang, S., Li, G., Zhang, H., Jin L., Su, Y. And Wu, K. (2017). Identifying the determinants of housing prices in China using spatial regression and geographical detector technique. *Applied Geography*, 79:26–36.
- Wiggin T. (2012). The year the housing market turned the corner. [Online]. Available from: www.aol.com.
- Xu, L. and Tang, B. (2014). On the determinants of UK house prices. *International Journal of Engineering Research*, 57-64.

- Yu, D., Wei, Y.D., and Wu, C. (2007) Modelling spatial dimensions of housing prices in Milwaukee, WI. *Environment and Planning B: Planning and Design*, 34:1085-1102.
- Yusof, A.M. and Ismail, S. (2012). Multiple regressions in analysing house price variations.
- Zahirovich-Herbert, V. and Gibler, K. (2014). The effect of new residential construction on housing prices. *Journal of Housing Economics*. 26:1–18
- Zainora, A., Norzailawati, M. and Tuminah, P. (2016). A spatial analysis on GIS-Hedonic pricing model on the influence of public open space and house price in Klang Valley, Malaysia. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B8