Concepts, Measures and Indicators of Urban Value in the City of Johannesburg: A GIS Study Using 2018 Data

MSc in GIS and Remote Sensing

School of Geography, Archaeology and Environmental Studies,
University of the Witwatersrand

March 2022

Dylan Weakley
0311337G

Supervisors:

Prof Elhadi Adam
School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand

&

Prof Philip Harrison
School of Architecture and Planning, University of the Witwatersrand

Declaration

I, Dylan Weakley (student Number 0311337G), am a student registered for the degree MSc in GIS and Remote Sensing at the University of the Witwatersrand. I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that all the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if it is found that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: ______________________ Date: 1 March 2022

Table of Contents

Declaration 2

Table of Contents 3

Abstract 5

Acknowledgements 6

List of Figures 7

List of Tables 9

List of Abbreviations 10

Chapter 1: Introduction 11

1.1 Problem Statement 13

1.2 Research Question 14

1.2.1 Research Sub-Questions 14

1.3 Aims and Objectives 15

1.3.1 Aim 15

1.3.2 Objectives 15

1.4 Outline of the research report 15

Chapter 2: Literature review 17

2.1 Urban Value 17

2.2 Location Theory and Hedonic Pricing 18

2.3 Spatial Justice 20

2.4 Spatial Analysis and GIS for Measuring Urban Value 21

Chapter 3: Methods and Data 23

3.1 Study Area 24

3.1.1 Reference map for the CoJ 25

3.2 Data 26

3.2.1 Data Sources 26

3.2.2 Data Projections and Coordinate System 28

3.3 Integrative literature review 28

3.4 Hedonic Regression: Determining Spatial Factors that Influence Residential Property Price and Quality of Life in the CoJ 28

3.5 Spatial Comparisons of different Measures of Urban Value 31

Chapter 4: Spatial Analysis: Approach and Results 33

4.1 Two Spatial Regressions: Spatial and Quality of Life Survey Factors that influence Sectional Scheme property price in the CoJ; and Spatial Factors that influence Quality of Life Index Score 33

4.1.1 Preliminary Analysis and Data Preparation 34

4.1.2 Spatial Regression Approach and Results 41

4.2 Comparison of three measures of Urban Value in the CoJ 49

4.2.1 Data Preparation 49

4.2.2 Visual Comparison of the three Measures of Value 56

4.2.3 Clustering of Measures of Urban Value 59

4.2.4 Tjøstheim's coefficient to test the correlation between the three measures of urban value 61

4.2.5 Intersecting high urban potential with low property price to indicate areas of housing development potential 61

Chapter 5: Discussion and Conclusion 65

5.1 Hedonic Regression Results: Population Density as an Indicator of Urban Value 65

5.2 Comparison of the three measures of Urban Value 67

5.2.1 Thiessen Polygons for Mapping Rateable Value 67

5.2.2 Differences in the Clustering of and Correlation between Different Measures of Urban Value 67

5.3 Limitations of the Study 70

5.4 Considerations for further research 70

5.5 Conclusion 71

References 73

Annexure: Tables, Data, Geo-Processing Models, Extra Maps 80

Abstract

The value that cities provide to their citizens is not equally enjoyed by all, and this inequality is spatially represented. Urban value is a useful lens through which to analyse cities, towards making them more liveable and equal. Urban value in this work considers the spatial factors that attract people to live in different parts of cities, and from which they derive some form of value. Value, and urban value are complex concepts, however. Value may be social, environmental, experiential and many more, but is most commonly expressed in financial terms (exchange value). This can lead to negative outcomes in cities, especially in the context of the inequality many, such as Johannesburg face.

This work assesses and compares three datasets from 2018 representing different types of or approaches to urban value in the City of Johannesburg (CoJ). The CoJ Valuation Roll represents exchange value, while the Gauteng City-Region Observatory's Quality of Life (QoL) Survey and the City of Johannesburg's Nodal Review Urban Potential Model represent social, environmental and place value.

The work uses a literature review to define the theoretical framework and approach and spatial statistics and analysis to compare the datasets or representations of urban value. Hedonic spatial regression is used to determine the spatial factors that influence sectional scheme residential property price and quality of life scores in the CoJ. Visual analysis, statistical comparisons for correlation and clustering and an overlay analysis is then used to compare the three sources of data.

The work finds that as population density increases, sectional scheme property price decreases, a positive finding for the City of Johannesburg's spatial policies which revolve around the idea of densification. Less positive is that as population density increases, Quality of Life index scores decrease, highlighting the importance of integrated and considered approaches to urban densification.

The qualitative analysis comparing the three datasets shows how they do not corelate perfectly and are differently distributed across space. It shows the complexity of a concept such as urban value, that can differ substantially when looked at from different perspectives. The work puts forward a novel way to map the CoJ valuation roll using Thiessen Polygons, that would be useful to the CoJ administration. Lastly, the work creates two overlay maps that are useful to housing developers and could lead to improved development outcomes for the CoJ. They show areas of high housing potential but with low property values that could promote social returns, financial feasibility and more affordable housing in the City.

Acknowledgements

I extend my gratitude to my supervisors, Prof Elhadi Adam and Prof Philip Harrison, and to my previous supervisor, Dr Stefania Merlo. I am grateful to my family for everything they do, especially my wife, Annabel. I am grateful to the City of Johannesburg Metropolitan Municipality (my employer at the time of writing this) who paid the tuition fees for this degree through their subsidised education for employees' scheme[1].

[1] The City of Johannesburg was not involved in the drafting of this research, and the work can be in no way attributed to them.

List of Figures

Figure 1: World Urban vs Rural Population 1960 to 2017 (World Bank, 2019) 11
Figure 2: Method Summary 23
Figure 3: Study area, City of Johannesburg Metropolitan Municipality, South Africa (Map by Author) Boundary Data: (Municipal Demarcation Board, 2016) 24
Figure 4: Reference Map, City of Johannesburg (map by author, data: © OpenStreetMap, Style, © MapBox) 25
Figure 5: Spatial Regression Decision Process (Anselin L. , 2005, p. 199) 31
Figure 6: Average rateable value per square metre for residential sectional scheme properties as per the CoJ Valuation roll 2018 for nodal review hexagons that contain sectional scheme developments (map by author) 35
Figure 7: GCRO Quality of Life Index 2018 Joined to Nodal Review hexagons, with hexagons containing null values (Nodal Review) removed (map by author) 36
Figure 8: Histogram and box and whisker plot for Average Sectional Scheme Value Per Square Metre 39
Figure 9: Histogram and box and whisker plot for Log Transformed Average Value Per Square Metre 39
Figure 10: Histogram of QoL index (left) and Log10 QoL Index (right) 40
Figure 11: Histogram of Transformed QoL Index (Ordered Quantile Technique) 41
Figure 12: Measures of model fit of best subsets for models with 1 to 17 independent variables generated using RSS Selection. Horizontal axis in all cases is number of variables in the model. 42
Figure 13: GCRO Quality of Life Index 2018 average values per CoJ ward (map by author) 50
Figure 14: Nodal Review Densification Index at its original 400m Hexagon level (left) and aggregated to a ward level (right) (map by author) 51
Figure 15:Anomaly in the way value from the Valuation Roll 2018 is in some cases Related to Property, causing a concentration of property price. Transparent areas (where aerial photographs are seen) indicate areas with zero or null values. 52
Figure 16: Valuation Roll 2018- Gross Rateable Value per Square Metre at Generated Thiessen Polygon Level (Map by Author) 55
Figure 17: Average Rand Value per square metre across the City of Johannesburg (2018 CoJ Valuation Roll) at 400m Hexagon level (left) and aggregated to a ward level (right) (map by author) 56
Figure 18: Visual Comparison of three Measures of Urban Value at a Ward Level using Equal Intervals between value range categories (maps by author) 57

Figure 19: Visual Comparison of three Measures of Urban Value at a Ward Level using Jenks' Natural Breaks between value range categories (maps by author) 58
Figure 20: Spatial Concentration of Rateable Value in the CoJ. Percentage of value vs percentage of land 60
Figure 21: Rateable Value in the CoJ in areas that score at least 0.5 in the Nodal Review Densification Potential Index 63
Figure 22: Rateable Value in the CoJ in areas, from the Nodal Review 2020, that are assigned as "no maximum density" 64
Figure 23: ArcMap Spatial Processing Model developed by Author to aggregate values from smaller to larger polygons (e.g., from Nodal Review Hexagons to Wards) 86
Figure 24: Geo-processing model for creating Thiessen Polygon Layer of Average Value Per Square Metre (Valuation Roll 2018), across the CoJ 87
Figure 25: Detail: Valuation Roll 2018- Gross Rateable Value per Square Metre at Generated Thiessen Polygon Level (Map by Author) 88
Figure 26: Nodes and Development Zones from the CoJ Nodal Review (City of Johannesburg, 2020, p. 20) 89

List of Tables

Table 1: Data Used in the Research 27
Table 2: Data Projection and Coordinate System Used 28
Table 3: Results of Cronbach's Alpha test for Internal Consistency 37
Table 4: OLS Results, 8 variables in each model 43
Table 5: Moran's I test for Spatial Autocorrelation of Residuals on OLS models 44
Table 6: Lagrange Multiplier Diagnostics on OLS Model to Determine whether to use a Spatial Lag or Spatial Error Model 45
Table 7: Summarised Results of the Spatial Lag Regression Models 46
Table 8: Moran's I test for Spatial Autocorrelation of the three measures of urban value being compared, at a ward level 59
Table 9: Table interpreting Figure 20: Spatial Concentration of Rateable Value in the CoJ. Percentage of value vs percentage of land 60
Table 10: Results of Tjøstheim's coefficient to test the correlation between the three measures of urban value 61
Table 11: Sample of dataset used. Only 8 records shown (out of 3018) and data transposed (changed columns to rows) to ensure that it fits. 80
Table 12: Explanation of all variables and units 81
Table 13: Test of Cronbach's Alpha when Single Variables are Deleted 84
Table 14: Simplified correlation matrix showing only rows and columns that contain correlations greater than 0.8, with all values converted to absolute values 85

List of Abbreviations

CoJ – City of Johannesburg Metropolitan Municipality

GCRO – Gauteng City-Region Observatory

GIS – Geographic Information System

LIS - Land Information System at the City of Johannesburg

LM - Lagrange Multiplier

MAUP - Modifiable Areal Unit Problem

OLS – Ordinary Least Squares Regression

QoL – Quality of Life

RP - Revealed Preference

SDF – Spatial Development Framework

SP - Stated Preference

SPLUMA - Spatial Planning and Land Use Management Act (no. 16 of 2013)

SS – Sectional Scheme as per the Sectional Titles Schemes Management Act (No. 8 of 2011)

Wits - University of the Witwatersrand

Chapter 1: Introduction

The world has seen extraordinary growth in its cities over the past century, with the urban population more than quadrupling from 1960 to 2017 (see Figure 1). Many theories are put forward for why people live in cities and why cities grow. This research focuses on the broad argument put forward in urban economics, that cities create value. A major reason why people live in cities is for the value (or potential value) they provide, and the growth of cities has clear links with value creation at many scales (Brueckner, 2011; McDonald & McMillan, 2011). The value of cities is usually measured in economic terms, but takes many more forms, from cultural to spiritual, social, academic, artistic, scientific, environmental and more (Lainga, Wheeler, Reeves, & Frost, 2014). This research focuses on the broad concept of urban value[2] as a theoretical frame.

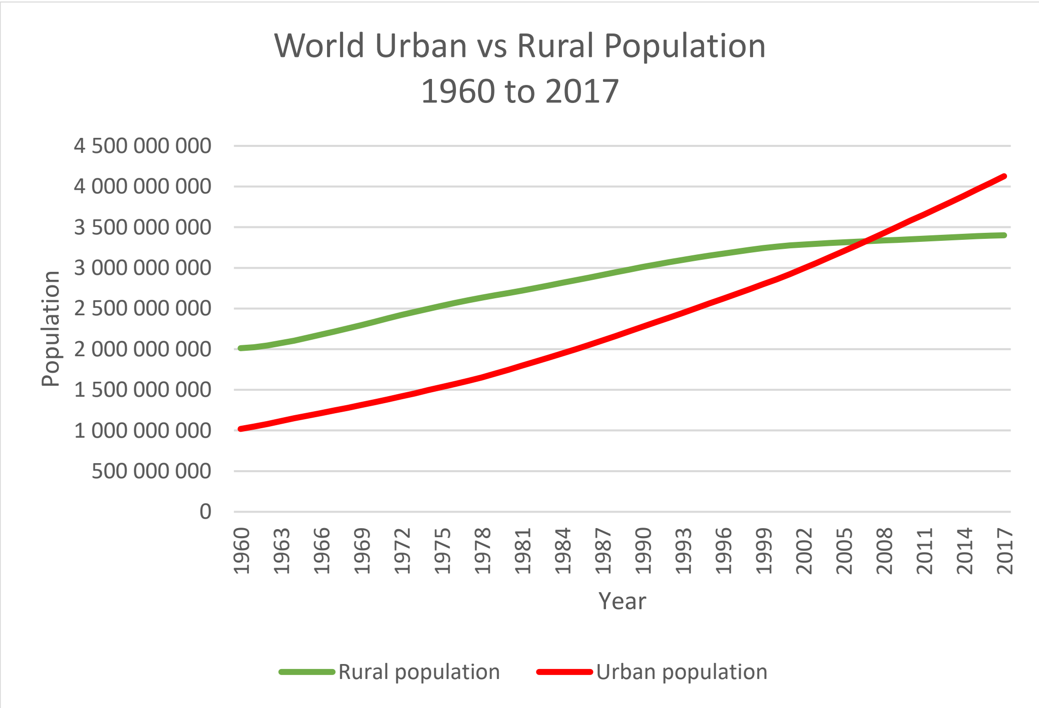

Figure 1: World Urban vs Rural Population 1960 to 2017 (World Bank, 2019)

While value (and urban value) can take many forms, it is also perceived differently by different people and organisations. People (planners, residents, developers. etc.) make spatial decisions based on trade-offs between different types of urban value. For example, when looking for a home, all urban residents look to balance affordability with amenity, convenience and quality of life, also called 'utility

[2] This even though the term 'urban value' is not broadly used in the literature. The term is borrowed from a bid for research funding in 2018 titled "Urban value in African cities" by Dr Colin Marx (University College London), Dr Nicole Zimmerman (University College London) Prof Philip Harrison (University of the Witwatersrand, co supervisor of this work) and Prof Ahmed Adamu (Ahmadu Bello University).

maximisation’ (Alonso, 1964; Burnett, Brooks, & Bassett, 2017). Different types of organisations may see value differently (for example banks vs social housing institutions) and even within organisations, perceptions of urban value may differ. In a city administration for example, the rates collection department will likely view urban value quite differently to the parks department.

Urban value is variable across space. Different cities provide diverse sets of urban values for their citizens, but also produce and share value differently. Some cities influence worldwide economies, while others operate at more local economic scales (Turok & McGranahan, 2013). This work considers how value differs across space at local scales in the City of Johannesburg (CoJ), at sometimes fine grains.

The spatial variability of urban value is important as value is not necessarily enjoyed by all who move to and live-in cities. South Africa is home to some of the most unequal cities in the world, including Johannesburg (UN-Habitat, 2016). For example, while the average household income in Johannesburg is approximately R15,000 per month, over half of households earn below a quarter of that, R3,500 per month (City of Johannesburg, 2016). This inequality is starkly spatially distributed, with poorer people living on the outskirts of the City, far from jobs and amenities, and the wealthy living in central parts of cities with better access (City of Johannesburg, 2016). For spatial policy to direct development in a way that will reduce spatial inequality and make cities more liveable for all, a detailed spatial understanding of urban value is needed (McDonald & McMillan, 2011). This research attempts to contribute to that understanding in the City of Johannesburg.

The research does not attempt to uncover all the complexities of urban value in Johannesburg. It is an exploratory study to assess- using spatial quantitative datasets- what some of the drivers of urban value in Johannesburg are and how these differ across space. It considers how this information may be used to inform better urban planning towards a more spatially just city. It also aims to identify different types or conceptions of urban value used in Johannesburg, represented by different quantitative datasets. The work combines desktop research with spatial and statistical analysis of three primary datasets, being The CoJ Valuation Roll 2018, the Gauteng City Region Observatory’s (GCRO) Quality of Life Survey (QoL) 2018, and the CoJ’s Urban Potential Model that was developed in 2017/18 for the Nodal Review Policy[3] (City of Johannesburg, 2020). The spatial analysis includes two sets of spatial hedonic regression to analyse the spatial factors (from the Nodal Review Urban Potential Model and the QoL survey) that influence property value (from the CoJ valuation Roll) and Quality of life scores (from the QoL) respectively. It then includes comparisons of the three datasets

[3] I was largely responsible for developing this model in my capacity as a Senior Specialist: Strategic Urban Planner, Development Planning Department, at the City of Johannesburg.

representing urban value in the City including a visual comparison through mapping, measurement and comparison of spatial clustering (using Moran's I coefficient), a measure and comparison of spatial correlation between the datasets (using Tjøstheim's coefficient) and lastly the creation of a spatial index showing areas of high potential for residential development (as per the Nodal Review) with low property prices (as per the Valuation Roll).

While the term 'urban value' is broad, the following preliminary working definition is used: *The spatial factors that attract people to live in different parts of cities, and from which they derive some form of value*. As such, the work focusses on urban value from a residential perspective, as opposed to business or industry (primarily to contain the scope of the work).

1.1 Problem Statement

For South African Cities to transform and grow in the manner envisioned in the country's legislation and spatial policies[4] effective planning is needed based on a good understanding of the dynamics that affect city development. A detailed spatial understanding of urban value is needed to develop and implement meaningful spatial plans (McDonald & McMillan, 2011). Added to this, there are two broad problems in the context of South Africa, and more specifically to Johannesburg, that this research seeks to address.

First is that there is a narrow approach to measuring urban value (Glumac, Herrera-Gomez, & Licherona, 2019) that focusses on property price, economic returns on investment, and finances, broadly known as 'exchange value' (Smith, 1776). While this information is undoubtably useful as a strong and real driver of city transformation and growth, if too narrowly followed it can become limiting and ignore some of the many complexities of urban value. The financial-only approach can also be exclusionary and ignore arguably the main stated spatial goal of urban planning policy in South Africa; to transform cities into better places for all. This relates to discrepancies that may exist between different datasets that describe urban value. The data from the City's Valuation Roll, the QoL and the Nodal Review Urban Potential Model do not correlate perfectly, and this work investigates some of those discrepancies. For example, two areas with similar access to amenity according to the Nodal Review may have different property valuations. This may indicate an area for potential development, where property values are relatively low, but theoretical urban potential is high, important information from an urban planning point of view.

[4] Including but not limited to, the Constitution of South Africa, the Spatial Planning and Land use Management Act 2013 and the National Development Plan 2030, the Integrated Urban Development Framework and the Johannesburg Spatial Development Framework 2040.

This leads onto the second problem related to urban planning. In Johannesburg, there are two main tools that urban planners use to guide development. The first is capital and operational investment in space (building and maintaining infrastructure and providing services) and the second is land use policy that guides applications for and decisions on the awarding of land use rights (City of Johannesburg, 2016). While land use policies such as the Spatial Development Framework 2040 (SDF) and the Nodal Review 2021 provide the potential for increased land use rights in certain areas, the market does not always respond with investment and development as intended. As investors are naturally risk averse, development tends to be concentrated in certain areas with Rosebank and Sandton being good examples in Johannesburg (SA Commercial Prop News, 2019; City of Johannesburg, 2021). It is hoped that the analysis comparing different types of urban value may assist in highlighting the potential urban value across the city, to guide investment and development in a more equitable manner.

1.2 Research Question

a) What are the different quantitative concepts, measures and indicators of urban value in the City of Johannesburg, and how may this information inform better urban planning, towards a more spatially just city?

1.2.1 Research Sub-Questions

a) What is urban value, and what are the predominant concepts of urban value according to local and international literature?
b) What approaches and methods can be used to measure urban value, especially looking beyond exchange value (monetary value)?
c) What spatial or GIS methods (and relevant datasets) may be used to quantitatively measure, map and analyse different conceptions of urban value?
d) What are the spatial drivers (from the QoL and the Nodal Review) of residential property price in the CoJ?
e) What are the spatial drivers (from the Nodal Review) of quality of life in the City of Johannesburg (as per the QoL survey 2018)?
f) How can the intersection of the three chosen datasets (property price, QoL survey and access to amenities) be used broaden our understanding of urban value in the City of Johannesburg?

1.3 Aims and Objectives

1.3.1 Aim

The aim of this study is to interrogate the concepts, measures and indicators of urban value in the City of Johannesburg, with the goal of informing better urban planning towards a more spatially just city.

1.3.2 Objectives

The specific objectives of the research are:

a) To investigate different theoretical conceptions of urban value, spatial drivers of urban value and how they relate to urban planning.
b) To determine appropriate methods to quantitatively measure urban value from more than just an economic or exchange value perspective.
c) To determine the spatial factors (from the CoJ Nodal Review data, and the GCRO QoL 2018) that influence residential property price in the CoJ (from the CoJ 2018 valuation roll).
d) To determine the spatial factors (from the CoJ Nodal Review data) that influence Quality of Life in the CoJ (from the 2018 GCRO QoL survey).
e) To compare the three datasets chosen to represent different types of urban value in the CoJ, being the valuation roll, QoL and Nodal Review Urban Potential model.

1.4 Outline of the research report

Summarised, the research report is outlined as follows.

<u>Chapter 1: Introduction</u>

The introduction chapter provides the problem statement, research question aims and objectives of the research. It introduces the concept of urban value and how it is approached using the three datasets chosen.

<u>Chapter 2: Literature Review</u>

The literature review outlines the main threads of literature that guide the work. It starts by introducing the concept of spatial justice as the main normative guiding principle for the work. It then looks at the concepts of value and urban value referencing the fields of urban economics, real estate, economic geography and location theory. Lastly it discusses spatial analysis and GIS based approaches to measuring and analysing urban value.

Chapter 3: Methods and Data

Chapter 3 outlines the methods and data used in the work. The method is made up of a literature review and spatial analysis sections. The spatial analysis includes two parts. First is hedonic spatial regression to test the spatial factors from the chosen datasets that affect Quality of Life scores and residential property price in the CoJ. Second is a comparison of the three datasets representing urban value, using visual comparisons, a measure of spatial clustering, a measure of spatial correlation, and spatial intersection.

Chapter 4: Spatial Analysis, Approach and Results

This chapter presents the steps taken in performing the spatial analysis and presents the results of the analysis but does not give detailed interpretations of the results, which are contained in the final chapter.

Chapter 5: Discussion and Conclusion

This, the final chapter of the work, analyses, interprets and discusses the results from the spatial analysis conducted, and ties them back to the literature review and problem statement of the research. It also discusses the limitations of the work and proposes some considerations for further research.

Chapter 2: Literature review

2.1 Urban Value

"Since value depends on economic rent, and rent on location, and location on convenience, and convenience on nearness, we may eliminate the intermediate steps and say that value depends on nearness".

(Hurd, 1903) in (Alonso, 1964, p. 6).

For thousands of years, philosophers and economists have grappled with the concept of economic value (Fogarty, 1996). This work does not attempt to contribute to theories of value broadly, but it borrows some basic principles from the literature to help in describing and assessing urban value. Adam Smith defines two broad notions of value, being 'value in use' and 'value in exchange' (1776, p. 26). Respectively, these describe "[a] the utility of some particular object [or good], and… [b] the power of purchasing other goods" that an object provides (ibid.), with exchange value generally representing monetary value. This distinction provides the starting point for describing types of urban value that may be found in Johannesburg. The second basic distinction is made between intrinsic (objective) and subjective theories of value. Intrinsic value refers to the value something has in and of itself (Sandler, 2012) while subjective value is inferred on objects, goods and services by people and markets in the context of myriad economic forces such as supply and demand (Fogarty, 1996). This work, while not necessarily 'choosing' a type of value theory to assess, leans more towards subjective theories of value, based on the data used for analysis and the intention to explore urban value from different perspectives.

When determining or measuring value (urban value in the case of this work) two broad approaches are considered, being revealed preference (RP) and stated preference (SP). Revealed preference approaches determine value using "empirical data from economic transactions and other behaviours to determine, ex post, people's willingness to pay for those goods" (Bartholomew & Ewing, 2011, p. 19). Stated preference approaches use mainly surveys to gauge willingness to pay for a good, and thus value (ibid.). Both approaches (or types of data) are used in this work, with the CoJ valuation Roll representing a revealed preference approach to urban value, and the GCRO QoL survey representing a stated preference approach (although not directly referring to property value). While traditionally the approaches refer to exchange value, they are applied to broader notions of value in this work.

2.2 Location Theory and Hedonic Pricing

When considering *urban* value, the broad field of urban economics is the point of departure. The research lends itself to economics, which Begg et al. describe as "the study of how societies make choices under conditions of scarcity" (2014, p. 2). Urban economics may be described as the intersection of economics and geography, or the study of economics across space (O'Sullivan, 2012). According to McDonald and McMillan (2011, pp. 3-4) urban economics studies urban areas using two main approaches: the study of location decisions (of households, businesses, industries) and the study of the urban economy in aggregate (such as economic growth or employment trends of whole cities or city regions). This work focuses on the former: how spatial variation across the City of Johannesburg affects urban value, and how it may affect location decisions of households (or other decision makers), also referred to as location theory (Burnett, Brooks, & Bassett, 2017), discussed later.

Added to the broad conceptions of value mentioned above, there are more specific notions of value that may be applied to urban value. These include economic value (as has already been mentioned) social value, environmental value, experiential value (Lainga, Wheeler, Reeves, & Frost, 2014) and positional value (D'Acci L. , 2014). Social value refers broadly to positive social impacts that can be derived from goods, objects and places (Supply Chain Sustainability School, 2017). Environmental value is similar, and refers to the intrinsic value that natural resources and biodiversity have either in and of themselves (Sandler, 2012) or in terms of services they provide to society (being provisioning, regulating, cultural or supporting services) (Schäffler, et al., 2015). Both social and environmental value can be assigned economic, or monetary values using various methods. This approach is popular when making the argument that social or environmental value is given more weight in public policy or budgeting processes when their effects on a financial bottom line are overt (ibid.) Lastly, place value is put forward as a measure that brings together different notions of value (Brown & Weber, 2011; Carmona, 2019), closely equating to the approach to urban value this work explores. These different types of value are comparable and related to the three legs of sustainability, being social, environmental and economic (IUCN, UNEP & WWF, 1991).

At times different notions of value are used interchangeably, for example where property price is seen as an indicator of social value or quality of life, because economic actors are willing to pay more for what they consider more socially valuable (D'Acci L. , 2014; Burnett, Brooks, & Bassett, 2017). Hurd (1903) summarises this regarding social value, noting that "the basis of residence values is social and not economic- even though the land goes to the highest bidder" (Alonso, 1964, p. 6).

As mentioned earlier there is a strong general economic argument for why cities exist. Essentially it is that cities form because economic activity benefits from scale (producing and trading in large quantities) and agglomeration (locating close to other economic activity) which leads to a concentration of jobs and a resulting concentration of nearby houses (Brueckner, 2011, p. 1).

While access to jobs provides a broad economic explanation for why people are attracted to cities, people can't live in cities for jobs alone. Other services are needed for survival and quality of life. Indeed, the possible negative aspects of living in cities (pollution, congestion, crime, high cost of living etc.) sometimes referred to as the 'urban penalty' (Freudenberg, Galea, & Vlahov, 2005) must be outweighed by the benefits of living there to make cities attractive[5] (O'Sullivan, 2012). Local goods and services may be provided by different levels of government (called spheres of government in South Africa) and include services such as healthcare, education, public transit, safety services (policing, firefighting, emergency services) and refuse removal/public cleaning (Brueckner, 2011). Hard infrastructure also provides essential services, and includes water and electricity supply, sanitation, stormwater management, streets, parks and more. There are then services and goods that are provided by the private sector, and include anything from shops, banks and religious facilities to private healthcare and education. There are also less-concrete benefits of living in a city or in a certain part of a city, that may include the location of family, cultural groupings or even climatic conditions. The point of this is not to list all of the local goods, services and attractions of cities, but to note that access or lack of access to them strongly influences urban value across and between urban areas (Alonso, 1964; Burnett, Brooks, & Bassett, 2017).

Location theory is the study of how access to amenity, goods and services affects the distribution of different land uses (Burnett, Brooks, & Bassett, 2017) and the distribution of urban value. Location theory focusses mainly on economic and exchange value, but it provides useful concepts and methods that can be applied to the wider approach to urban value. An early example of location theory, the von Thünen model, proposes that with all other things assumed equal, agricultural land rent (or value) will decrease with distance from a market due to transport costs (von Thünen, 1826) in (McCain, 2013) and with that, the intensity of land use will also decrease (Bartholomew & Ewing, 2011). This and other early works were built on by William Alonso (1964) in "Location and Land Use: Toward a General Theory of Land Rent". The book introduces a more detailed analysis called 'bid-rent theory' attempting

[5] Of course, it is not always guaranteed that a good quality of life will be found in cities. The relative benefits of cities *should* outweigh living in rural areas but the reality is that many people move to cities to escape poverty in rural areas, often moving into poverty in urban areas instead (Parsons, 2010; Saunders, 2010). The fact that this choice is so frequently made however indicates that there is a net benefit, albeit sometimes small of moving to cities.

to create a theory for how urban land may be valued for different land uses (not only agricultural). The theory still adopts access to a focal point (central business district - CBD) as the main driving factor of land rent but does provide for some variation such as pollution and other environmental factors. While the theory is popular and still in use today[6] two limitations are: a) its focus on mono-centricity (cities with one CBD) (Bartholomew & Ewing, 2011), and b) the attempt to build a theory for *all* cities, ignoring local variances (McCain, 2013).

An evolution of the bid-rent approach is hedonic pricing. Hedonic pricing is used broadly in economics to measure the factors that influence the value of a certain good (D'Acci L. , 2014). These can include any goods, but the approach is well suited to and thus commonly used in land value analysis (Nanda, 2019). Hedonic property pricing uses multiple regression analysis and/or spatial regression to determine the physical characteristics (particular to a property such as number of bedrooms etc.) and locational characteristics (access to amenities, environmental quality, crime rates etc.) that affect the value of a property (D'Acci L. , 2014; Hill, 2013). These are also referred to as intrinsic and extrinsic and attributes respectively (Hill, 2013). This work looks at locational characteristics that affect urban value. While hedonic pricing is mostly used for calculating exchange value, it certainly has potential and has been used to measure other conceptions of urban value, such as Quality of Life (D'Acci L. , 2014). Hedonic pricing is discussed further in Chapter 3: "Methods and Data".

2.3 Spatial Justice

Spatial justice is briefly mentioned here as the main normative guiding principle of this work.[7] As mentioned earlier, a narrow exchange value dominant approach urban value can lead to unequal outcomes. The lens of spatial justice then gives the consideration of urban value in this work a normative direction. Spatial justice is well engrained in spatial planning theory and policy in South Africa, with the Spatial Planning and Land Use Management Act 16 of 2013 (SPLUMA) formally adopting spatial justice an outcome for planning. It recognises that social injustice is embedded in space, and that spatial justice can be employed as a guiding principle to address this injustice (Van Wyk, 2015). As mentioned earlier, this work assesses urban value towards creating better planning and urban outcomes in Johannesburg, and a more spatially just city.

[6] For example, the CoJ built a version of a bid-rent model for its 2016 Spatial Development Framework (City of Johannesburg, 2016).

[7] As was the case with previous research reports I have submitted (Weakley D. , 2009; Weakley D. , 2013).

2.4 Spatial Analysis and GIS for Measuring Urban Value

Location theory, hedonic pricing and land valuation generally lend themselves well to spatial analysis using geographic information systems and spatial statistics (Powe, Garrod, Brunsdon, & Willis, 1997). Geographic Information System (GIS) analysis allows for precise measuring of distances between phenomena (for comparing accessibility) using an array of measures, most prominently for this work being network distance (e.g., distance by road or rail) and Euclidean distance (as the crow flies) (La Rosa, 2014). GIS also allows for visualisation of results, which helps with gaining a more concrete spatial understanding than numeric statistical results (Hamilton, et al., 2016). Lastly, and most importantly for this work, GIS allows for different datasets to be compared to one another using location as the common feature. This is very beneficial, as data from different sources and data collected or designed for different purposes can be compared (Graham, Carlton, Gaede, & Jamison, 2011), as long as the spatial scale and grain of the data are comparable (Beale, Abellan, Hodgson, & Jarup, 2008; Goodchild, 2011).

Many approaches to GIS based property valuation have been developed. Below is a brief overview of some of these, which does not aim to exhaustively list all of the approaches in existence, but to illustrate the breadth of approaches in terms of different conceptions of urban value.

GIS approaches/systems for property valuation have existed for some time, with Peter Wyatt proposing such a system in 1997 (Wyatt, 1997). Wyatt's goal was to expand valuations beyond a human valuers' local knowledge of an area to a more empirical foundation. Using Horsham, UK as the study area, Wyatt focussed on the spatial factors that affect property value, similar to what this research report tries to achieve. Wyatt used network analysis (distance by road) to determine the distance from each property to a spatial database of urban amenities, using the data to create a value index for the study area. A literature review and interview with a valuer in the area were used to determine the spatial factors that affect property price, after which multiple regression analysis was used to measure the effect of different spatial attributes on value (again, a similar approach to what will be used in this research) (Wyatt, 1997).

What Wyatt achieved is broadly similar to hedonic regression or pricing. Importantly however, the approach (using GIS systems) can take different views on the spatial factors that indicate or drive urban value. Powe et al. (1997) for example built a model which tests the affect that access to woodlands has on property prices in New Forest, UK. The work used both distance to and size of forests to create a forest index value for each property in the area. The effect of that index on property price was then statistically tested. Similarly, La Rosa (2014) used access to green spaces as an indicator of urban value (not economic value). Using Catania, south Italy as the study area, La Rosa created accessibility indicators for open space, specifically intended to improve planning and provision of open space to create a more liveable urban environment. In their comprehensive study, Bartholomew & Ewing (2011) trace literature that uses hedonic pricing to measure the effect of pedestrian and transit-oriented development on land prices. They conclude that “the market-shifts in favour of pedestrian and transit-designed development indicated by survey data and demographic analyses are, indeed, being capitalized into real estate prices” (Bartholomew & Ewing, 2011, p. 30).

Similar La Rosa’s approach above, other GIS approaches to measuring urban value don’t necessarily look to relate to economic value. The first example is ‘Walk Score’ which is a patented GIS based system that measures walkability of different parts of cities based on access to amenities and road networks (Bereitschaft, 2018) (Walk Score, 2020). The approach uses network analysis to create an index of walking accessibility to various amenities, scoring different parts of cities ranging from 100 (“Walker's Paradise - Daily errands do not require a car”) to 0 (“Car-Dependent - Almost all errands require a car”) (Walk Score, 2020, p. 1). It is based on a regular grid of latitude/longitude points (500 feet apart), where weightings of access to amenities decrease with distance (Bereitschaft, 2018). The approach is similar to that used by the City of Johannesburg in its Urban Potential Model developed for the Nodal Review, and used in this work as a data source (City of Johannesburg, 2020).

The last examples highlight GIS approaches to measuring quality of life (considered another approach to measuring urban value) in cities. Burnett et al. (2017, p. 456) for example propose an alternative to stated preference (e.g. surveys) and revealed preference (e.g. property price data) measures of urban quality of life, that uses chain store data to build comparative indexes of “quality of life and place”. They propose a metric based on location theory using the positions of two food and beverage chain stores in the United States creating what they call a “McBucks Index” that purportedly indicates quality of life. They then compare it to other measures of QoL in order to test the approach. While they cite limitations of the index, their results “strongly correlated with both traditional objective and subjective QOL techniques” (Burnett, Brooks, & Bassett, 2017, p. 463). Lastly, D’Acci (2014) who provides a thorough analysis of measures of urban quality of life (many of which have been mentioned earlier in this paper) describes ‘isobenefit lines’ as a method of mapping areas of similar benefit or urban value in the City. These are similar to contour lines which show altitude on maps and allow representations of urban value to be mapped in a way that shows the relative benefit, or urban value of different parts of cities (D'Acci L. , 2012) (D'Acci L. , 2014).

Chapter 3: Methods and Data

This section describes the methods and data used in the work. Both are influenced by the preceding problem statement and literature review. The method consists of two parts. The first is the desktop research (literature review above) to assess concepts, measures and indicators of urban value guiding the quantitative spatial analysis. Second is three broad sets of spatial analysis. The spatial analysis consists of, first, an investigation into the spatial attributes that influence residential property price in the CoJ, second, the spatial factors that that influence quality of life scores in the CoJ and third, a quantitative comparison of the three datasets under study. The first two use hedonic spatial regression with the spatial factors data obtained from the CoJ Nodal Review, the Quality of Life scores from the QoL survey 2018, and property price from the CoJ Valuation roll 2018. The third set of analysis uses visual comparisons, a measure of spatial clustering, a measure of spatial correlation, and spatial intersection to compare the three sets of data representing urban value. The method is summarised in Figure 2 below.

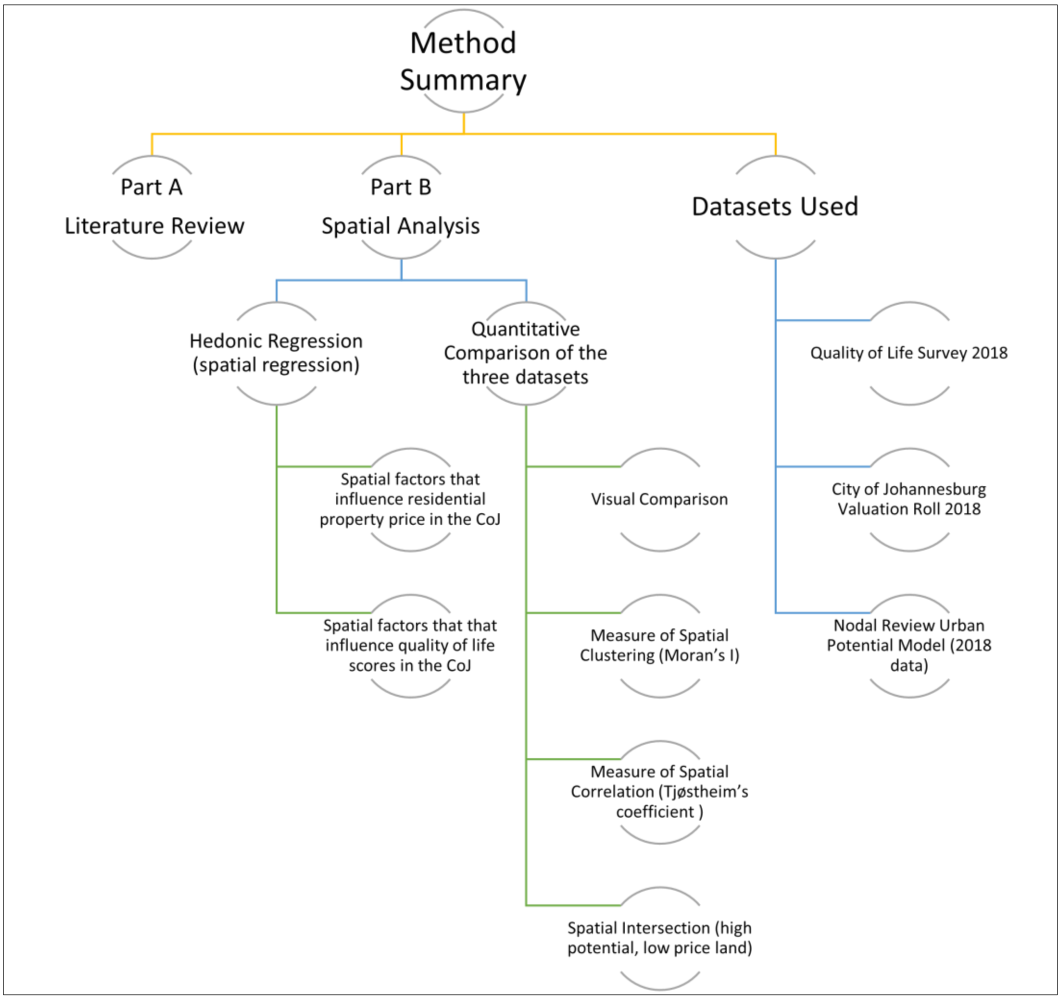

Figure 2: Method Summary

3.1 Study Area

The study area for is the Johannesburg Metropolitan municipality in Gauteng Province, South Africa. The municipality is home to approximately 5.8 million people (Stats SA, 2020) and, as mentioned earlier, is typified by spatial inequality. The Johannesburg Spatial Development Framework reports that more than three hundred thousand households live in informal dwellings, and that well engrained patterns of spatial inequality exist (City of Johannesburg, 2016). The city is sprawled with an average population density of 3000 people per square km and there is a spatial mismatch between where the majority of people live, and the location of jobs and other amenities (City of Johannesburg, 2016). The municipality was chosen as a case study due to the unequal spatial distribution of value it portrays, but also for practical reasons, that the author (at time of writing) works at the municipality, and thus has a special interest in the study area and access to the data needed. The study area is shown in Figure 3 below.

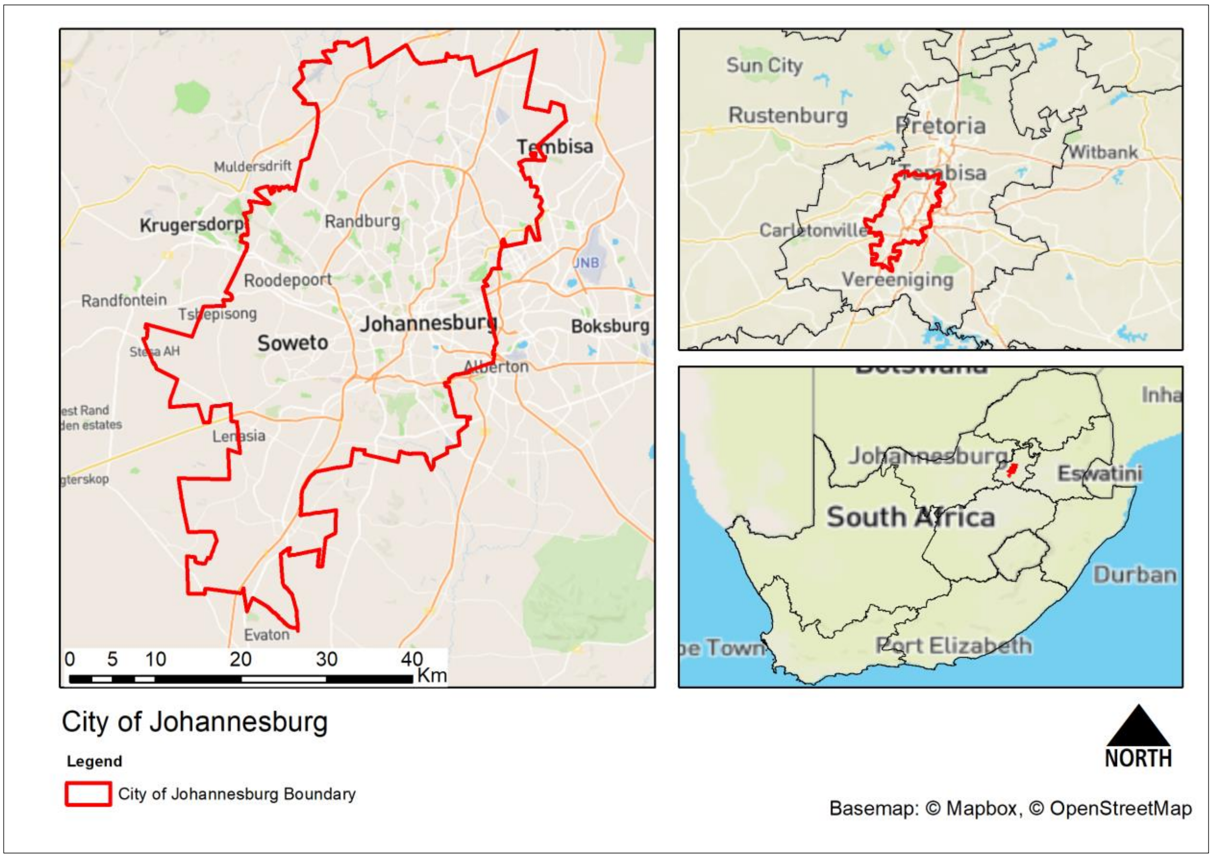

Figure 3: Study area, City of Johannesburg Metropolitan Municipality, South Africa (Map by Author) Boundary Data: (Municipal Demarcation Board, 2016)

3.1.1 Reference map for the CoJ

Many of the maps presented in this work exclude base maps that give the reader spatial reference. This is done so that the data can clearly be seen and not be obscured by base map overlays. As such, a reference map is included here that the reader can turn to as a spatial reference for areas in the City.

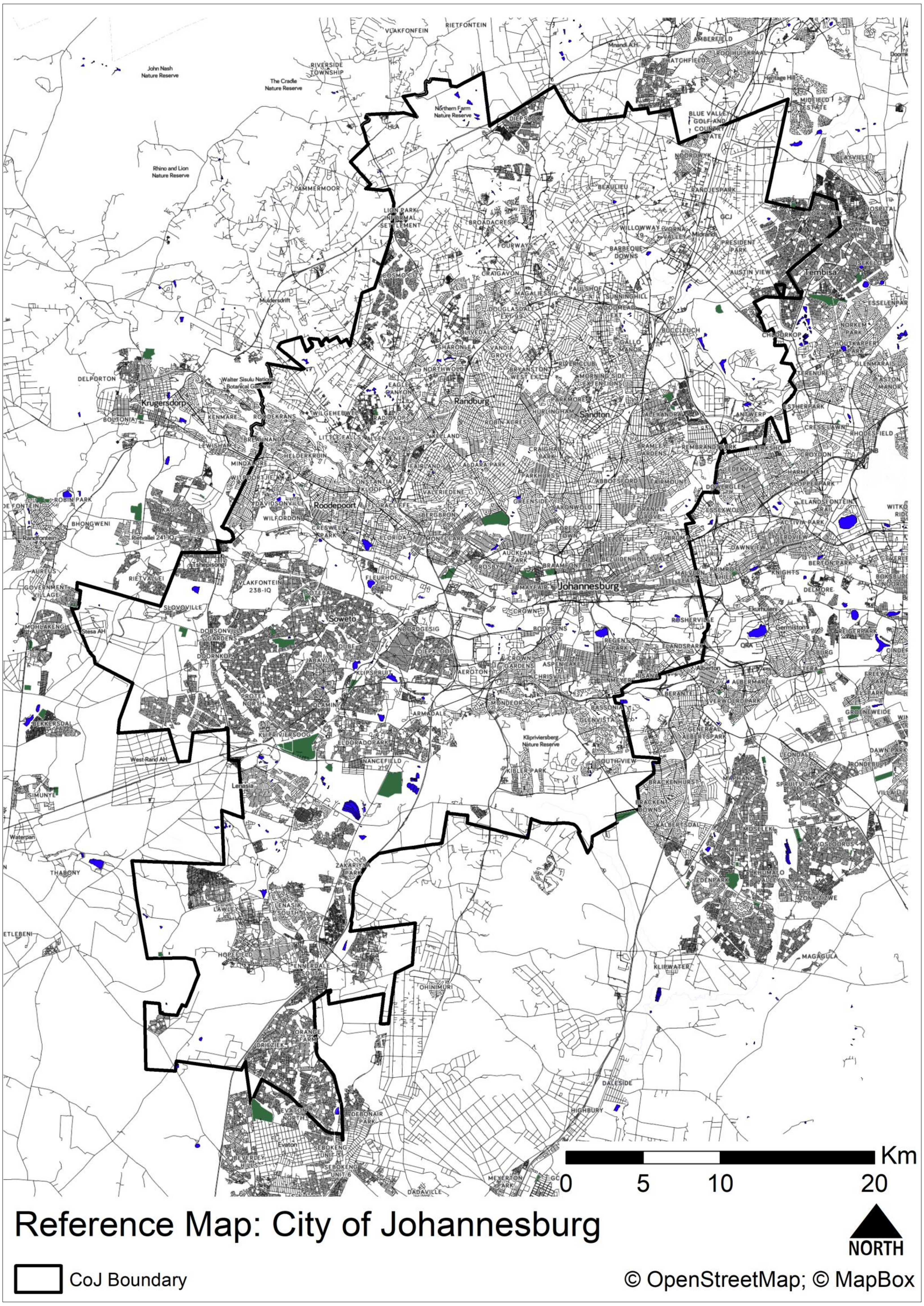

Figure 4: Reference Map, City of Johannesburg (map by author, data: © OpenStreetMap, Style, © MapBox)

3.2 Data

3.2.1 Data Sources

The table below lists the three data sources that are used. The data are based on what is readily available, but also on the desire to compare data that reflects different types of value discussed in the literature review. As such the table includes reference to different conceptions and measures of urban value from the literature review. The data were identified and collected based on the intention of the work; to analyse urban value in the City of Johannesburg. The three datasets were chosen based on the different approaches to urban value outlined in the literature review and were also practically informed by what datasets are available for the chosen time period, 2018, in the City of Johannesburg. The time period, 2018, was chosen so that there is a consistent temporal scale for the data, to make it comparable, and remove any errors in comparison that may occur if using data from different years. It was also done for practical reasons. A single year of analysis was chosen to limit the scope of the research, as is needed for research reports of this nature.

Table 1: Data Used in the Research

Data Name and Source	Description	Spatial Coverage	Value Types and method for determining urban value	Scale	Data Format
CoJ Property Valuation Roll 2018 **Source: City of Johannesburg Metropolitan Municipality**	The official valuation of each property in the City of Johannesburg, used for rates purposes.	City of Johannesburg Metropolitan Municipality	• Property price • Economic/exchange value • Revealed Preference	Erf level (Sub erf level in the case of sectional title units, but with the erf used for location)	Provided as a comma separated table (.csv file) and spatially joined to erf polygon vector layer (shapefile)
GCRO Quality of Life Survey 2018 **Source: Gauteng City Region Observatory**	"The biennial Quality of Life (QoL) survey, first conducted in 2009, measures the quality of life, socio-economic circumstances, attitudes to service delivery, psycho-social attitudes, value-base and other characteristics of the GCR." The data includes an overall quality of life index". (GCRO, 2018)	Gauteng City Region	• Quality of Life • Social Value, Environmental Value • Stated Preference	Ward Level	Provided as an SPSS data file, and spatially joined to Ward polygons (shapefile)
CoJ Nodal Review Urban Potential Model Data 2018 **Source: City of Johannesburg Metropolitan Municipality**	Data compiled for the CoJ Nodal Review 2019. It includes access to amenities in the City (based on the street network) and includes a walkability score for each hexagon. The Data includes two indexes, being a residential densification potential index, and a nodal index for determining the location of economic nodes (City of Johannesburg, 2020)	City of Johannesburg Metropolitan Municipality	• Access to jobs and amenities, street walkability • Social Value, Environmental Value • Revealed Preference	Regular grid of sampling Hexagons (400m between opposite corners)	Provided as a polygon vector layer (shapefile)

3.2.2 Data Projections and Coordinate System

All of the data used in this work was projected to the coordinate system shown in Table 2 below and all of the maps included are in the same coordinate reference system. This is as per the guidelines South African Department of Rural Development and Land Reform (Rural Development and Land Reform, 2011; Rural Development and Land Reform, 2013) and the projections used by the CoJ.

Table 2: Data Projection and Coordinate System Used

Projected Coordinate System:	TM_29_Hartebeesthoek
Projection:	Transverse_Mercator
False Easting:	0.00000000
False Northing:	0.00000000
Central Meridian:	29.00000000
Scale Factor:	1.00000000
Latitude Of Origin:	0.00000000
Linear Unit:	Meter

3.3 Integrative literature review

The literature review forms part of the substantive results of the work (under objectives a and b) and is used to create the theoretical framework for the study. The method used is an integrative literature review (Torraco, 2016). Torraco (2016, p. 404) describes the integrative literature review as "…a distinctive form of research that generates new knowledge about the topic reviewed. It reviews, critiques, and synthesizes representative literature on a topic in an integrated way such that new frameworks and perspectives on the topic are generated." The literature review considers local and international literature on urban value, both as part of theorising different conceptions on urban value (as introduced in the literature review for this proposal) but also to identify drivers and indicators of urban value and different spatial approaches to measuring them.

3.4 Hedonic Regression: Determining Spatial Factors that Influence Residential Property Price and Quality of Life in the CoJ

Hedonic regression is used to determine the spatial factors that influence property price and QoL scores in the CoJ (objectives c and d). For the hedonic regression, the method outlined by Glumac, Herrera-Gomez, & Licherona (2019) is adopted, but for a static time period- 2018. The dependent variables are residential property value (using sectional schemes[8] only) from the CoJ's General Valuation Roll 2018, and QoL score from the QoL 2018. The independent variables tested are factors from the GCRO Quality of Life Survey 2018 (for property price) and from the Nodal Review Urban Potential Model 2019 (for both sets of analysis).

The formula for an ordinary least squares (OLS) regression model is:

$$y_n = \beta_0 + \sum_{k=1}^{K} \beta_{\mathrm{k}} x_{\mathrm{nk}} + \varepsilon_n$$

Where y_n is the independent variable property price for location n (1,…,N), β_0 is the y intercept, β_{k} is the coefficient estimate of independent variable k, x_{nk} is the independent variable k for area n (1,…,N) and ε_n is the error term component of area n (1,…,N) (Glumac, Herrera-Gomez, & Licherona, 2019).

One of the required assumptions of multiple regression analysis is that residuals should not be autocorrelated (Glumac, Herrera-Gomez, & Licherona, 2019). Spatial data is often spatially autocorrelated however due to Tobler's first law of geography which states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). To test spatial autocorrelation of residuals, a Moran's I test (Moran, 1950) is run on the OLS model, checking the p value and observed Moran's I statistic. The hypotheses to be tested are H_0: The spatial pattern of the residuals is random and H_a: The Spatial Pattern of the residuals is correlated.

To account for the likely spatial autocorrelation of residuals, one of two forms of spatial regression is used: spatial lag regression, or spatial error regression (Anselin & Bera, 1998). In a spatial lag model, the autocorrelation is seen as substantive (an influence from a known variable, not an error) and is modelled in the independent variable. In a spatial error model spatial effects are treated as errors (an influence from an unknown variable) and added to the error term of the regression. (Glumac, Herrera-Gomez, & Licherona, 2019; Anselin & Bera, 1998)

A spatial error model adopts the same equation as for OLS above, but where the error term is defined as:

$$\varepsilon_n = \sum_{j \neq n}^{N} w_j \varepsilon_j \rho + v_n$$

[8] Sectional scheme or sectional title properties, as per the Sectional Titles Schemes Management Act, 2011, are properties where multiple owners may own sections of one property. These commonly include blocks of flats or townhouse complexes (each flat or house is a section), with multiple owners, and management by a body corporate made up of owners.

Where w_j is a spatial weight not equal zero if area j $(j \neq n)$ is a neighbour of area *n,* from 1 to N. Where ρ is a spatial coefficient that captures the unobserved average neighbourhood influence that impacts on land plot *n (1,…,N),* and v_n is an error term component of land plot *n (1, …,N)* (Glumac, Herrera-Gomez, & Licherona, 2019, p. 804).

A spatial lag model introduces a spatially lagged variable to the regression procedure, with the equation being:

$$Y = \beta_0 + \beta_x + \rho wy + \varepsilon$$

Where Y is the is the independent variable property price, "x is a matrix of observations on the explanatory variables and ρwy is the spatial lag variable. ρ measures the degree of spatial correlation and w is the spatial weights matrix. The value ρ ranges between −1 and 1. A ρ of zero means there is no spatial correlation [….] ε is a vector of the independent and identically distributed error terms" (Wittowsky, Hoekveld, Welsch, & Steier, 2020, p. 47).

In order to decide whether OLS, spatial lag, or spatial error regression is used, a decision tree proposed by Anselin (2005) is used. The decision tree is depicted in shown in Figure 5 below. Summarised, the process involves: a) fitting an ordinary least squares regression, b) testing for spatial autocorrelation of residuals and, if found, c) performing a Lagrange Multiplier test to decide whether a spatial lag regression or spatial error regression model should be used, d) fitting the chosen model, e) confirming the removal of a spatial autocorrelation of residuals.

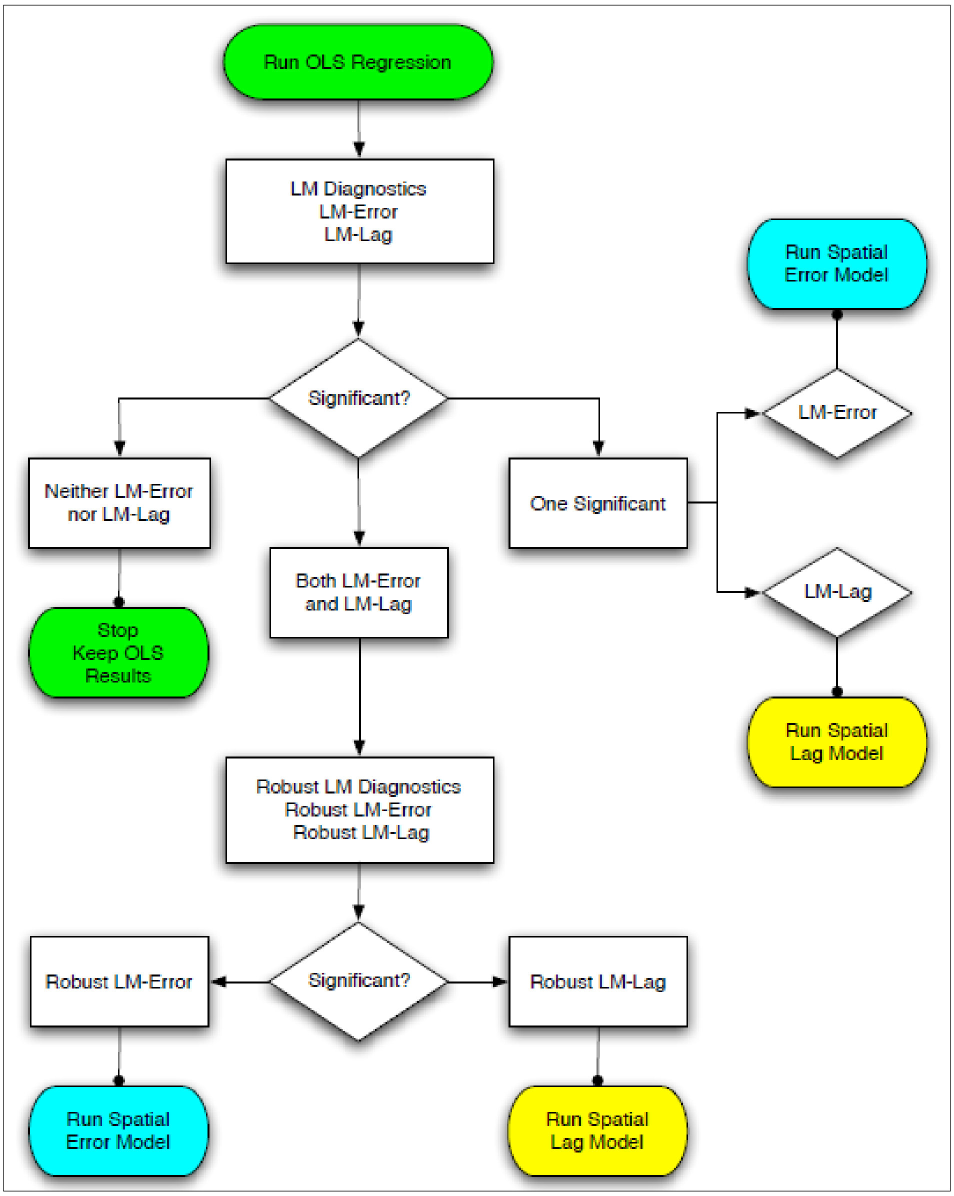

Figure 5: Spatial Regression Decision Process (Anselin L. , 2005, p. 199)

The Lagrange Multiplier test to determine whether a spatial lag, or spatial error model should be used indicates test statistics and probability estimates for each model (Anselin L. , 2003). Generally the model with the smallest p value is used (Anselin L. , 2005).

3.5 Spatial Comparisons of different Measures of Urban Value

The methods to achieve objective e, quantitatively comparing the three datasets on urban value, are discussed here. The first method is a simple visual analysis. This is done at the ward level, and is presented in side by side choropleth maps, allowing the reader to compare areas of high and low urban value, represented by the three datasets. Visualisation is useful to give the viewer an intuitive understanding of the data before any other spatial analysis is done (Słomska-Przech & Golebiowska, 2021).

The visual comparison is followed by a measure and comparison of the spatial clustering of three datasets representing urban value, using a Moran's I test, as discussed in section 3.4 above.

Next, Tjøstheim's (1978) coefficient of spatial association is used to test the relationship for each pair of the variables. Tjøstheim's coefficient is a variation on common correlation coefficient's (Pearson's r, Spearman's ρ and Kendall's τ), all of which perform poorly when applied to spatially dependent data (Hubert & Golledge, 1982). It is a non-parametric coefficient designed specifically for use on spatial data, taking into account the spatial relationship and dependence of the data (Vallejos, Osorio, & Bevilacqua, 2020; Osorio, Vallejos, Cuevas, & Mancilla, 2020).

The final part of spatial analysis done for the research is an intersection visual analysis of areas in the city that have high potential for residential densification (as per the Nodal Review) and have low land prices. The analysis is done at the 400m hexagon scale from the Nodal Review, as the finest grain at which both datasets are available. As mentioned in the problem statement for the work, this comparison is an attempt to highlight areas that are often overlooked for formal residential development in the City, and to possibly guide more equitable investment across the City. The precise method is further detailed in section 4.2.5.

Chapter 4: Spatial Analysis: Approach and Results

This chapter presents the steps taken in performing the spatial analysis and presents the results of the analysis but does not give detailed interpretations of the results, which are contained in the final chapter.

4.1 Two Spatial Regressions: Spatial and Quality of Life Survey Factors that influence Sectional Scheme property price in the CoJ; and Spatial Factors that influence Quality of Life Index Score

This section interrogates, using spatial statistical modelling, a) what the spatial or locational factors are that affect property price in the City of Johannesburg (CoJ) and b) what the spatial or locational factors are that affect QoL index scores in the City. The section describes the data pre-processing done for both sets of regression, and then the spatial regression approaches and results of each.

Due to the complexity and size of the data available (there are almost 1 million valued properties in the CoJ) the section only considers residential sectional scheme property values. This small selection was also made as sectional scheme units, in the City of Johannesburg Land Information System (LIS) data, are the only clear example of discrete dwelling units[9]. Zoning data would not give the same detail, as even Residential 1 zoning (generally one dwelling house per erf) can have up to 3 dwelling units, one main unit and two subsidiary units, with the possibility to divide the subsidiary units into 3 small units each, effectively giving a possibility of 7 dwelling units per erf (City of Johannesburg, 2019).

The locational factors (or independent variables) are obtained from two data sources being the City of Johannesburg's Nodal Review Urban Potential Model (City of Johannesburg, 2020), and the Gauteng City Region Observatory's Quality of Life Survey 2018 (QoL) (GCRO, 2018). The Dependent variables are derived from the City of Johannesburg's General Valuation Roll (property price), 2018 for the Sectional Scheme Price analysis, and the QoL index score is obtained from the QoL data.

This section presents the process and results of the two spatial regression models. The results presented here are similar (but with an extra spatial model added – the QoL model -) to coursework done by the author for the same degree as this dissertation (Weakley D. , 2019; Weakley D. , 2019a).

[9] Even in the case of "Residential 1" zoning in the CoJ Land Use Scheme, although the primary right is "Dwelling house", two subsidiary dwellings are allowed, each divisible into 3 rentable rooms with their own ablutions and kitchens, meaning that a Residential 1 stand can effectively house up to 7 units (City of Johannesburg, 2019).

4.1.1 Preliminary Analysis and Data Preparation

Data Joining and Pre-processing

Much GIS work was done to clean and prepare the data for statistical modelling. A full report of this is available to view at the following link (http://bit.ly/dw_gis_notes) however, the summary of the process is as follows:

a) Sectional Scheme Analysis

   a. Extract residential sectional scheme data from the valuation roll table and join the extracted data to a spatial polygon layer representing sectional scheme units in the CoJ.

   b. Remove properties (sectional scheme units) with unrealistic sizes, including properties that are smaller than or equal to 6m^2 and one property which was 4977m^2.

   c. Calculate the sum of value and the sum of floor area (of all the units) per each sectional scheme development. Using those summed values calculate the average value per square metre per scheme. Join that value to the sectional scheme polygon layer (representing each scheme's parent property).

   d. Convert the sectional scheme layer to a raster dataset using the average value per square metre as the pixel value (5m pixels).

   e. Using the "Zonal Statistics" tool in ArcMap, calculate the average value per square metre for each hexagon from the Nodal Review (see map in Figure 6)

   f. Join Nodal Review data and QoL data (including indexes) to the above dataset, using a spatial join. QoL data was joined from the ward into which the centre of each hexagon falls. This final dataset for the y variable 'average sectional scheme property price per square metre' is shown in Figure 6 below.

b) QoL Score Analysis

   a. Join the QoL index scores to the Nodal Review data, using a spatial join. QoL data was joined from the ward into which the centre of each hexagon falls. This final dataset for the y variable 'average QoL Score' is shown in Figure 7 below.

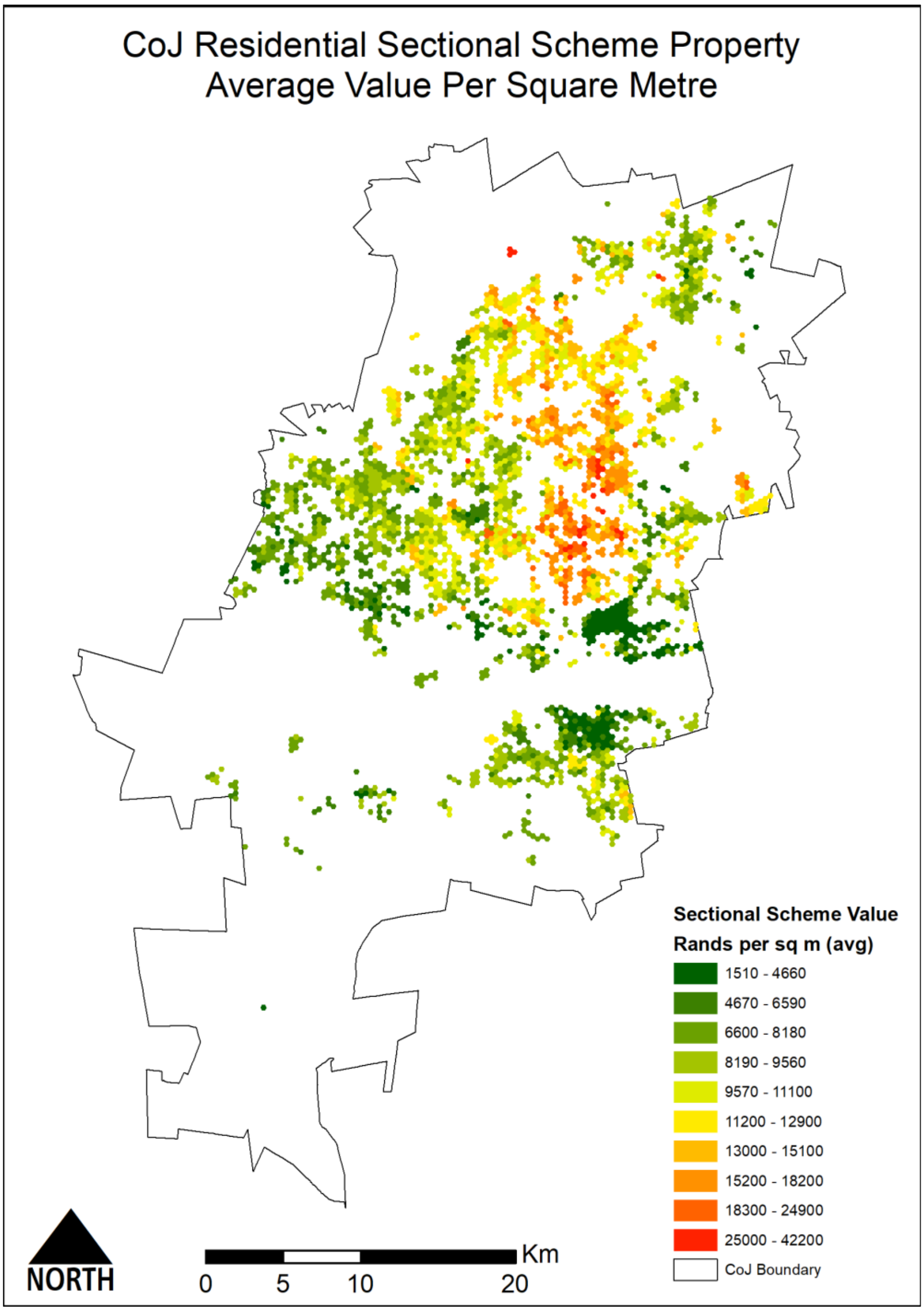

Figure 6: Average rateable value per square metre for residential sectional scheme properties as per the CoJ Valuation roll 2018 for nodal review hexagons that contain sectional scheme developments (map by author)

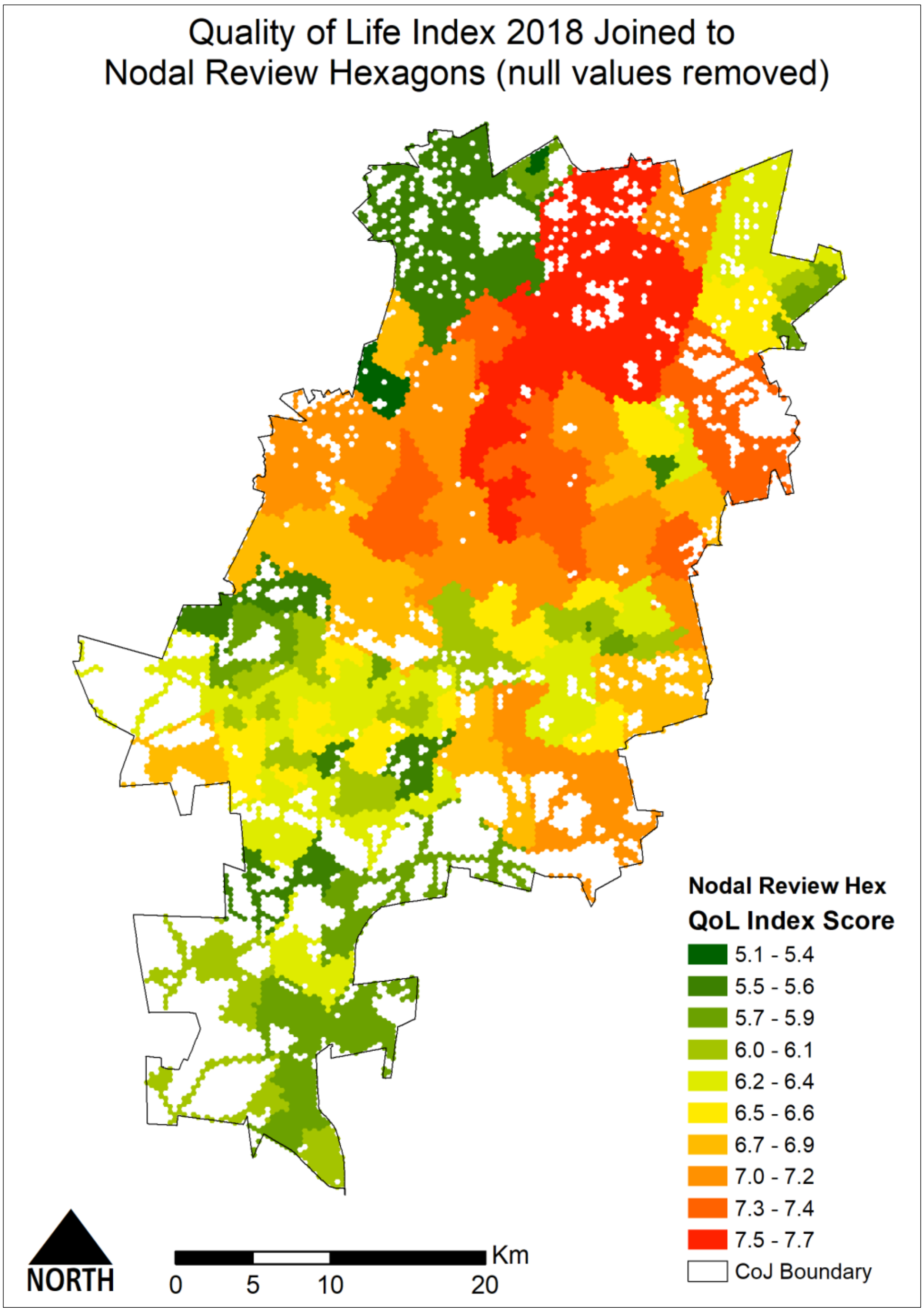

Figure 7: GCRO Quality of Life Index 2018 Joined to Nodal Review hexagons, with hexagons containing null values (Nodal Review) removed (map by author)

Initial Variable Selection

As the work seeks to define the locational attributes that determine property price and QoL scores, most demographic considerations (household income levels, race distribution, employment, age, etc.) were omitted. Population density was kept, however, as it is considered a locational characteristic of an area (affecting the built form, for example). While in the South African context racial and income distribution are likely to correlate with property price and QoL scores (due to the history of racially based segregation and economic exploitation) they are not considered spatial determinants, although they may be indicators. The variables chosen from the Nodal Review and QoL are included in the annexure (pg. 80), with an extract of the data (Table 11 pg. 80) and an explanation of all of the variables, units and data sources (Table 12 pg. 81).

Testing Internal Consistency of Likert Scale Data

Some of the variables chosen from the QoL are Likert scale[10] type questions (listed in Table 13 pg. 46). The QoL is representative at a ward level, and hence responses had to be aggregated to ward level (using mean values) to give an average value per ward for each question. Thus a test was needed to assess if the variables could be treated as continuous for use in the regression. A test for internal consistency was done on the raw data from the QoL (pre averaging at ward level), using the Cronbach's Alpha Coefficient test (Tavakol & Dennick, 2011). The test was done for the study area (Johannesburg Municipality). The results are as follows.

Table 3: Results of Cronbach's Alpha test for Internal Consistency

Cronbach Coefficient Alpha	
Variables	Alpha
Raw	0.848511
Standardized	0.854685

Table 3 above shows that the Cronbach's Alpha for the Likert Scale variables chosen from the QoL is 0.849. This indicates that there is internal consistency, and that the variables can be used as continuous variables.

Variable Reduction

As the Nodal Review and the QoL contained so many variables, some pre-elimination was done before any modelling. The pre-elimination of variables was to remove variables that may cause multi-collinearity (a condition for ordinary least squares regression). The correlation between all independent variables was measured using the principal component analysis tool in SAS. Table 14 pg. 85 is a simplified version of the correlation matrix, including only rows and columns that include correlations of 0.8 and higher, with correlation numbers converted to absolute (positive) values.

Any correlation of 0.8 and higher (highlighted in red in Table 14 pg. 85) was used as a cut off to indicate variables that should be removed from the analysis to prevent multi-collinearity. The variables

[10] Likert Scale data refers mainly to survey data, where respondents are asked to rate some phenomenon on a numeric scale, for example where a respondent may be asked to rate the safety of their neighbourhood on a scale of 1 to 5, with 1 being very unsafe and 5 being very safe.

removed are, from the QoL: 'satisfaction with living standard', 'satisfaction with healthcare services', 'QoL Marginalisation Index', 'satisfaction with dwelling', 'satisfaction with neighbourhood', 'satisfaction with roads', 'satisfaction with drainage', 'satisfaction with sanitation', 'satisfaction with waste', 'satisfaction with water supply', 'satisfaction with energy sources', and 'satisfaction with streetlights', and from the Nodal Review: 'Industrial and commercial buildings within 2km'. All of the variables removed form part of other variables in the dataset: the QoL indicators form part of the overall QoL index, the marginalisation index is very closely negatively correlated to the QoL index (being based on the same dataset), and the Nodal Review variable removed is the sum of two other variables, being 'industrial buildings in 2km' and 'commercial buildings in 2km'.

Sampling

No sampling was done, the full final datasets were used, being average price per square metre for all 400m hexagons (from the Nodal Review) which contain sectional scheme properties, and all hexagons from the Nodal review (with QoL index joined), with hexagons containing null values omitted. The final number of hexagons representing the sectional scheme price dependent variable is 3018 (Figure 6 pg. 35), and the number of hexagons in the QoL index dependent variable model is 12 456 (Figure 7 pg. 36).

Dependent variable transformations

A condition for linear regression is that the dependent variable (y variable) should be normally distributed. A histogram and box-and-whisker-plot were used to investigate whether the condition was met, and transformations done on the two y variables used.

a) Y Variable transformation: Average Sectional Scheme Value per square Metre

Figure 4 shows the histogram and box and whisker charts for Average Sectional Scheme Rateable value per square metre in the CoJ.

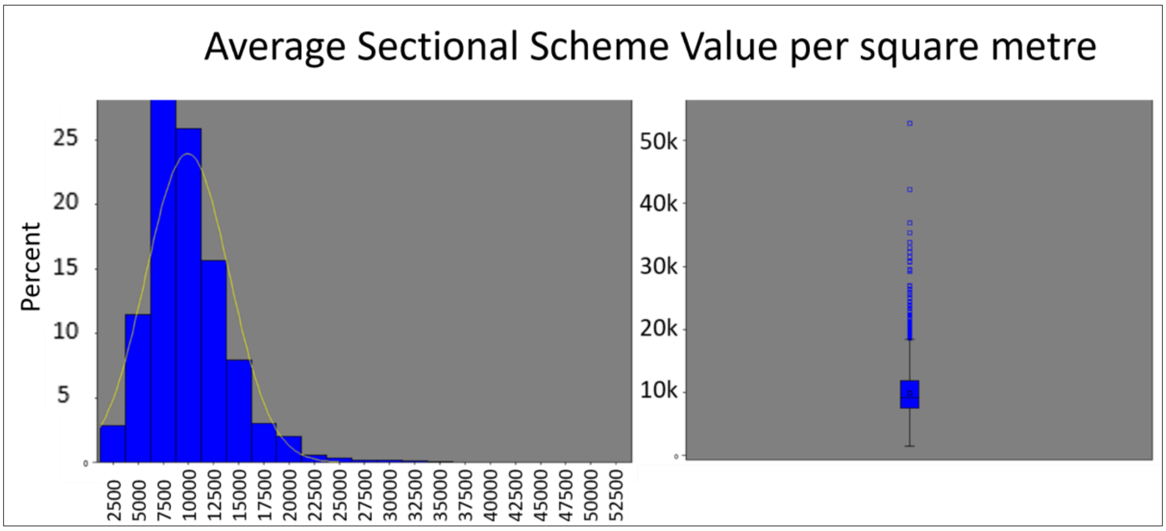

Figure 8: Histogram and box and whisker plot for Average Sectional Scheme Value Per Square Metre

In Figure 8, the histogram shows that the data is positively skewed with a long tail to the right and does not follow a normal distribution. This is confirmed by the box-and-whisker-plot, which shows that the mean is higher than the median, and that there is a long positive tail of values, and many positive outliers. As such, a log transformation was done to try and normalise the data. The results of the log transformation are as follows.

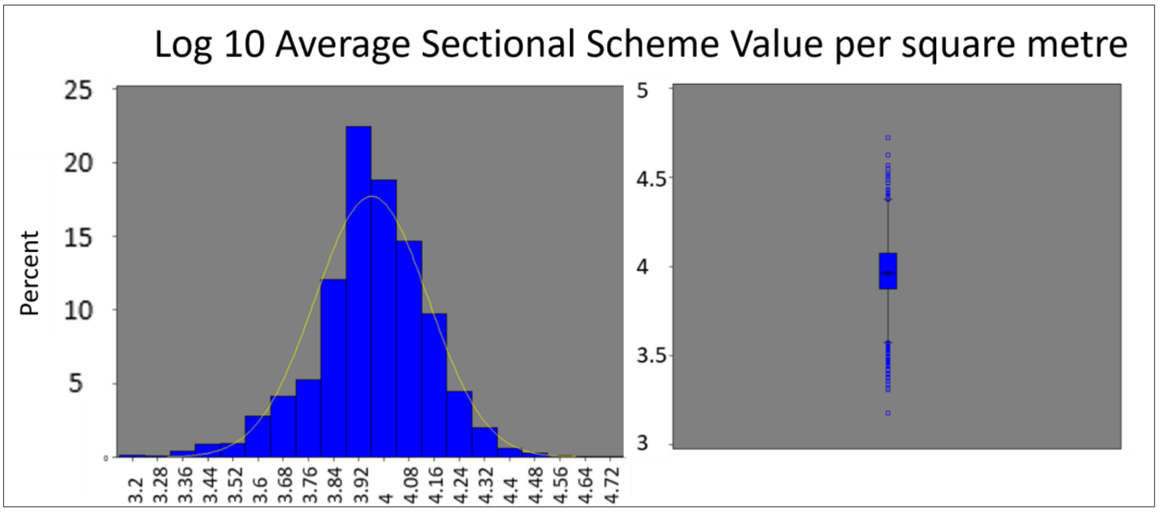

Figure 9: Histogram and box and whisker plot for Log Transformed Average Value Per Square Metre

The log transformed data shows what can be assumed to be a normal distribution. This is seen in Figure 9, the histogram, where the distribution looks normal, and now possibly slightly negatively skewed. In Figure 9, the box-and-whisker-plot shows that the mean and median values are very close to one another, and that the data looks normally distributed around the mean. Notable in the box and whisker plot in Figure 9 however is the large number of outliers which may affect the ordinary least squares model.

b) Y Variable transformation: QoL Index Score

Figure 10 shows the histogram of the QoL index variable on the left, and the Log 10 transformed QoL index variable on the right. The histograms do not give the impression that the data is normally distributed. As such, another transformation (other than log 10) was sought.

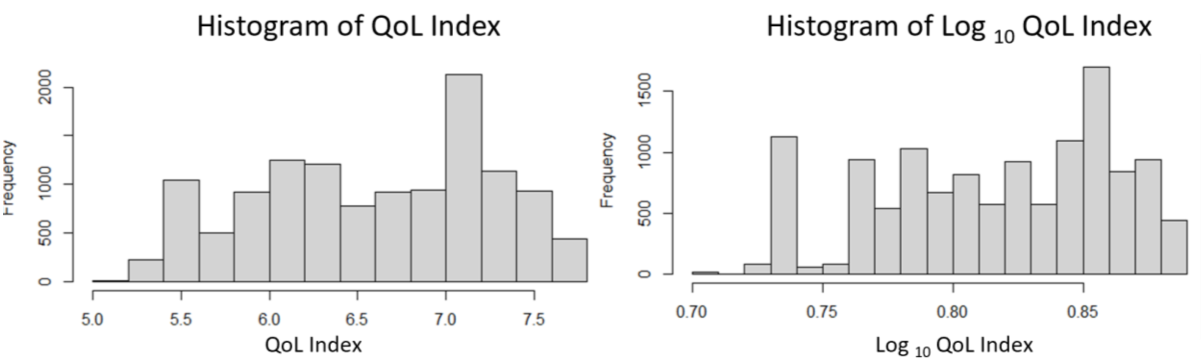

Figure 10: Histogram of QoL index (left) and Log10 QoL Index (right)

In order to perform the transformation, the tool "bestNormalize" (Peterson, 2021) in the software package R was used. The tool performs tests on a variable in order to recommend a transformation that will result in the most normally distributed data. After performing the test, Ordered Quantile Technique (OQT) transformation (van der Waerden, 1952) in (Peterson, 2021) was recommended, and thus performed on the y variable. The distribution of the transformed variable is shown in Figure 11, and shows what is assumed to be a normal distribution for this work. While a significant improvement is seen in terms of the normal distribution of the data, a drawback of the OQT technique is that no meaningful interpretations of beta coefficients is possible (as with log transformations). As such, only the variables that have a significant influence on QoL index (according to the methods used) are reported along with their sign (positive or negative). This then only shows if an independent variable has a significant positive or negative impact on the dependent variable, or not (still useful in answering the research questions).

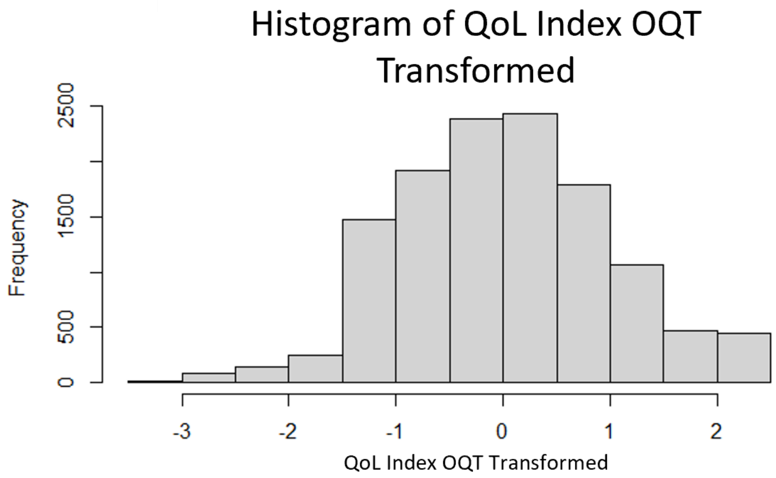

Figure 11: Histogram of Transformed QoL Index (Ordered Quantile Technique)

4.1.2 Spatial Regression Approach and Results

Multiple Regression Model Fitting (Ordinary Least Squares)

While principal component analysis was considered for further dimension reduction, it was not used as the process only considers the variability explained by variables, and not their effect on the dependent variable (Rawlings, Pantula, & Dickey, 1998), so is not useful in answering the research questions. As such, a long list of variables (Table 11 pg. 80 excluding variables removed in section 4.1.1) was used to fit the models. The model for y variable: QoL Index score, excluded all variables from the QoL, as those variables went into the creation of the QoL index and the research question relates only to the spatial factors from the Nodal Review. Model fitting was done by using best subset selection in R, using the tool "regsubsets". The tool selects the best subset model (based on residual sum of squares – RSS) for each number of variables (1 to n) in the model (Lumley, 2020). The goal of the process was to achieve the best model fit with the fewest variables. The process was run with all the x variables included for each of the two y variables, with the Adjusted-R^2, R^2, Mallows' CP and BIC (all measures of model fit) measured for each best subset (1 to n). The results from the tests (for each of the two models created) are shown in Figure 12 below.

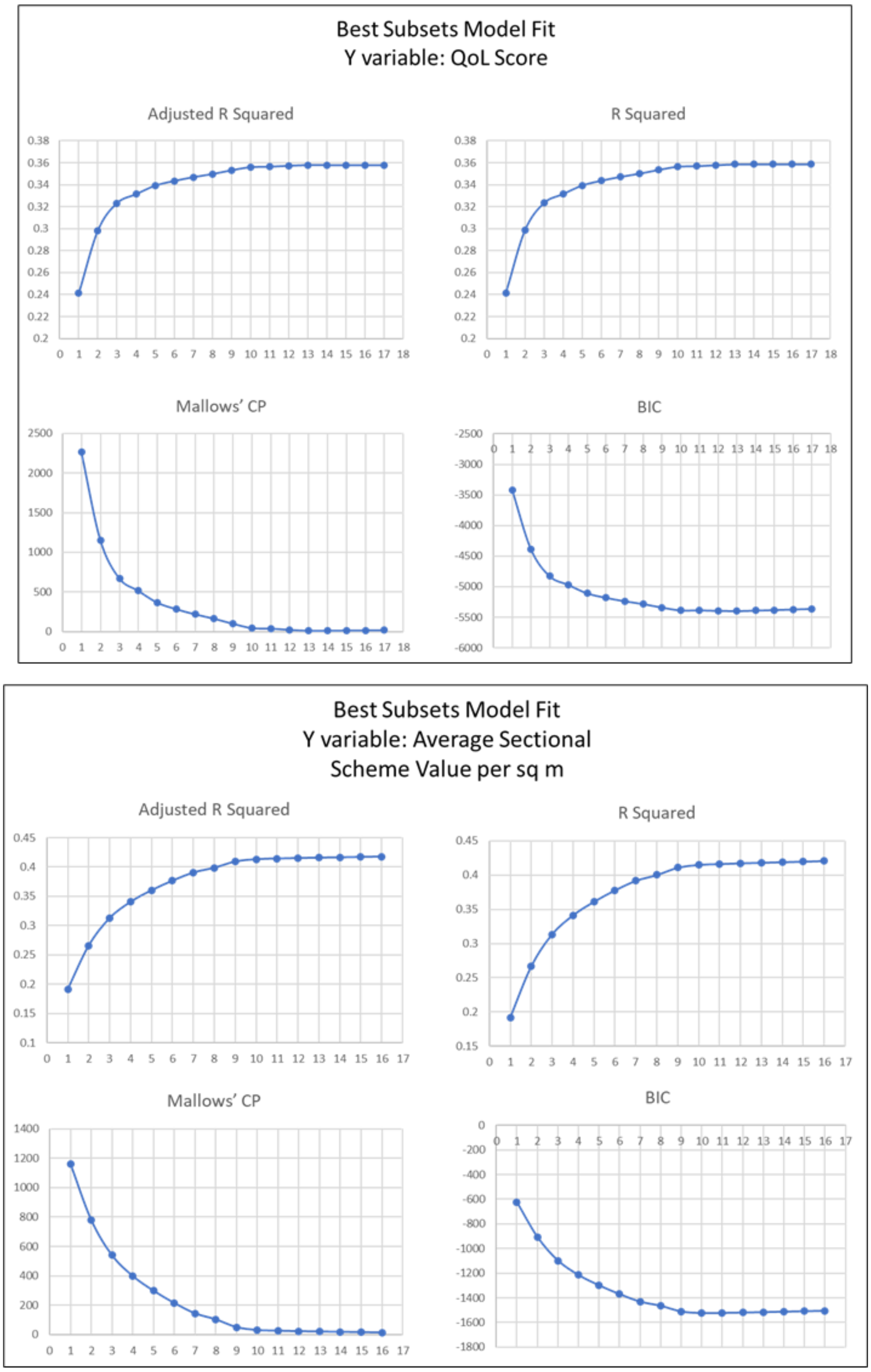

Figure 12: Measures of model fit of best subsets for models with 1 to 17 independent variables generated using RSS Selection. Horizontal axis in all cases is number of variables in the model.

For both models, in trying to minimise the number of variables, and maximise model fit, the number of 8 variables was chosen. This was because in all plots in Figure 12, very little improvement was seen in model fit beyond 8 variables, coupled with a desire to keep the number of variables low.

All 8 variables from each of the two 8-variable subset models were then fitted in ordinary least squares regression models for their corresponding y variables. The results are shown in Table 4 below.

OLS Regression Results and Interpretation

The Ordinary Least Squares regression results for each model are shown in Table 4 below.

Table 4: OLS Results, 8 variables in each model

Sectional Scheme Price per Square Metre		
Dependent Variable Log_{10}(Average Sectional Scheme Value per Square Metre)		
	Beta Coefficient Estimate	P(*t*)
(Intercept)	3.765000	<2e-16
Population Density	-0.000012	<2e-16
Public Schools within 2km	-0.012890	<2e-16
QoL Index Score	0.056870	<2e-16
Industrial Buildings within 2km	-0.000174	<2e-16
Closest Private Tertiary Education (m)	-0.000006	<2e-16
Closest Private Primary School (m)	0.000023	4.E-15
Closest Formal Taxi Rank (m)	-0.000012	<2e-16
Closest Private Hospital (m)	-0.000013	<2e-16

Model Statistics	Model P(f)	< 2.2e-16
	R-squared	0.4003

QoL Index Score		
Dependent Variable: OQT Transformed QoL Index Score		
	Beta Coefficient Estimate	P(*t*)
(Intercept)	0.907900	<2e-16
Population Density	-0.000042	<2e-16
Industrial Buildings within 2km	-0.001139	<2e-16
Closest Train Station (m)	-0.000019	<2e-16
Closest Private Tertiary Education (m)	-0.000013	<2e-16
Closest Registered Tertiary Education (m)	-0.000056	<2e-16
Closest Private Primary School (m)	-0.000077	<2e-16
Closest Formal Taxi Rank (m)	0.000029	<2e-16
Closest Taxi Route (m)	0.000079	<2e-16

Model Statistics	Model P(f)	< 2.2e-16
	R-squared	0.3502

Key	Variable Removed in spatial regression model due to unexpected Symbol (+ -)

Table 4 above summarises the results from the OLS regressions. Noteworthy results are that both models are significant indicated by their P(f) values being well below 0.05 at < 2.2e-16. Both models are similarly strong in terms of R-Square, with the Sectional Scheme price model and QoL index score's R-squared being 0.4 and .035 respectively, showing that the 40% and 35% of variation in the models is explained by the data. All beta coefficients are significant (well below 0.05) however, highlighted in red are variables that show theoretically unexpected or nonsensical values. The sectional scheme price model's beta coefficients for "Public Schools within 2km" is negative, while the beta coefficient for "Closest Private Primary School" is positive, indicating that an *increase* in public schools in 2km *reduces* property price, and that as distance *increases* to the closest private primary school, property price *decreases*. This is similar for "closest taxi rank" and "closest taxi route" in the QoL score model, with positive symbols for beta coefficients indicating that QoL Index score would increase as distance to these amenities increases (or decrease the closer you are to them).

Moran's I tests for Spatial Autocorrelation of Residuals

Next, a Moran's I test was done on the residuals of the OLS models to test for spatial autocorrelation of the residuals, with results shown in Table 5. The method for measuring neighbours and creating the weight matrix was 6 nearest neighbours as a hexagon has 6 sides - the polygons in which the data was calculated. A two-sided test was done that tests if the data are spatially dispersed, randomly distributed, or clustered. With Moran's I statistic, -1 represents perfect dispersion, 0 represents perfect random spatial distribution and 1 represents perfect clustering (Anselin & Bera, 1998). This is important as a regular hexagon grid was used in the Nodal Review (and the same grid used in this work), so there was a possibility that the data could be regularly spread out.

Table 5: Moran's I test for Spatial Autocorrelation of Residuals on OLS models

Moran's I test for Spatial Autocorrelation of Residuals			
Sectional Scheme Price per Square Metre OLS Model		QoL Index Score OLS Model	
p-value	< 2.2e-16	p-value	< 2.2e-16
Observed Moran I	0.57	Observed Moran I	0.88

For both OLS models, spatial autocorrelation of residuals was found, indicating (as suspected) that a spatial regression model would be needed. This in in that both p-values < 2.2e-16 < 0.05, thus the null hypothesis is rejected, and there is sufficient evidence to conclude in favour of the alternative hypothesis, which states that the spatial pattern of the residuals is correlated. The observed Moran's I values of 0.567 (Sectional Scheme Price per square metre model) and 0.88 (QoL index Score Model) indicates that the residuals are spatially clustered for both models, as would be expected with spatial data.

Spatial Regression Model Method Selection

As mentioned earlier, when choosing a spatial regression model, it is necessary to choose between a spatial lag model (where the error is seen as substantive and modelled in the dependent variable) and a spatial error model (where residuals are seen as errors and added to the error term of the regression). The test to determine whether a spatial lag or error model is appropriate is a Lagrange Multiplier or LM test (Anselin L. , 2003). The test was run in R using the OLS models created above, and the same spatial weight matrix created for testing spatial autocorrelation. The results from the tests are shown in Table 6.

Table 6: Lagrange Multiplier Diagnostics on OLS Model to Determine whether to use a Spatial Lag or Spatial Error Model

Lagrange Multiplier Test				
	Sectional Scheme Price OLS Model		QoL Index Score OLS Model	
Model	Statistic	P-value	Statistic	P-value
LM Error Model	3164.84	<2.2e-16	30032.34	<2.2e-16
LM Lag Model	3171.98	<2.2e-16	30091.03	<2.2e-16
Robust LM Error Model	120.01	<2.2e-16	191.41	<2.2e-16
Robust LM Lag Model	127.15	<2.2e-16	250.10	<2.2e-16

Based on the table above, it is seen that the spatial lag models and the spatial error models (including the robust versions of each) were significant, indicated by their p-values all being smaller than 2.2e-16. At this stage then, the spatial lag model was chosen for both Sectional Scheme Price and QoL index Score Models, as it is better equipped at answering the questions of the research. This is in that the spatial lag model assesses local attributes around the data (in this case, for each data points 6 nearest neighbours) rather than just including the autocorrelation found into the error term, as is done in the spatial error model (Anselin & Bera, 1998).

Maximum Likelihood Estimation of the Spatial Lag Models

In order to build the final spatial lag models, the variables indicating unexpected signs (shown in red in Table 4 pg. 43) were removed. Also, after running the spatial Lag once, the variables "Closest Taxi Route (m)" and "Closest Formal Taxi Rank (m)" were removed from the QoL Index score model, as they were not showing significant beta coefficients in the spatial lag model. The final spatial lag model results are shown in Table 7 below.

Table 7: Summarised Results of the Spatial Lag Regression Models

Sectional Scheme Price per Square Metre		
Dependent Variable Log_{10}(Average Sectional Scheme Value per Square Metre)		
	Beta Coefficient Estimate	P(*t*)
(Intercept)	0.809430	< 2.2e-16
Population Density	-0.000004	< 2.2e-16
QoL Index Score	0.015169	0.000308
Industrial Buildings within 2km	-0.000040	0.002066
Closest Private Tertiary Education (m)	-0.000002	3.41E-06
Closest Formal Taxi Rank (m)	-0.000002	0.017339
Closest Private Hospital (m)	-0.000002	0.014611

Model Statistics	Model P(f)	< 2.2e-16
	Pseudo R-squared	0.69546

Autocorrelation of Residuals Test: Moran's I	P-value	0.691
	Observed Moran I	-0.00581

QoL Index Score		
Dependent Variable: OQT Transformed QoL Index Score		
	Beta Coefficient Estimate	P(*t*)
(Intercept)	0.05134600	<2.2e-16
Population Density	-0.00000409	<2.2e-16
Industrial Buildings within 2km	-0.00005741	0.000061
Closest Train Station (m)	-0.00000077	0.081150
Closest Registered Tertiary Education (m)	-0.00000268	<2.2e-16
Closest Private Primary School (m)	-0.00000216	0.07581

Model Statistics	Model P(f)	< 2.2e-16
	Pseudo R-squared	0.91817

Autocorrelation of Residuals Test: Moran's I	P-value	0.999
	Observed Moran I	-0.04997

Key	Non-Significant Variable (95% confidence level)

The model statistics indicate that both models are significant, with P(f) values well below 0.05, at < 2.2e-16. The Nagelkerke Pseudo R squared (Nagelkerke, 1991) for the models both indicate significant improvements on the R-squared values for the OLS models. For the Sectional Scheme Price model, the value improved from 0.4003 to 0.695, and for the QoL index model, it improved from 0.35 to 0.918. AIC also improved from the OLS to Spatial Lag model for both the Sectional scheme and QoL models (-3075.6 to -5332.2 and 29967 to 3852.4 respectively). In terms of autocorrelation of residuals, both spatial lag models resulted in residuals not being spatially autocorrelated (unlike the OLS models). This is seen in the Moran's I scores of the sectional scheme and QoL models being very close to zero (-0.00581 and -0.04997 respectively), and more importantly the p-values (0.691 and 0.999 respectively) being significantly higher than 0.05, allowing the conclusion that the spatial pattern of the residuals is random. As such, it is concluded that both of the spatial lag models show significant improvements on the OLS models.

Reporting and initial interpretation of the Spatial Lag Model Results

The results reported here do not give detailed interpretations of all of the models. They rather give the broad regression formulas, and interpretation of the positive or negative affects the independent variables have on the dependent ones. This is because the research questions for each model were to interrogate the main spatial or QoL factors that affect sectional scheme property price, and the main spatial factors that affect QoL index scores. As such, the models were not built to make predictions on the dependent variables. It is also because two different transformations were used in the two models, making detailed interpretation impractical.

Having said the above, the models give the following results:

a) Sectional Scheme Property Price

The first model, the main spatial factors from the Nodal Review and QoL that influence sectional scheme property price in the CoJ, can be summarised in the following formula:

$$
\begin{aligned}
\widehat{\log_{10}(y)} = \; & 0.809430 \\
& -0.000004 \times \text{Population Density} \\
& \;\; 0.015169 \times \text{QoL Index Score} \\
& -0.000040 \times \text{Industrial Buildings within 2km} \\
& -0.000002 \times \text{Closest Private Tertiary Education (m)} \\
& -0.000002 \times \text{Closest Formal Taxi Rank (m)} \\
& -0.000002 \times \text{Closest Private Hospital (m)}
\end{aligned}
$$

Where y is "Average Sectional Scheme Value per Square Metre" and the hat (^) indicates a predicted value.

This means that, with all factors remaining equal, and based on the methods and data used in this work:

a) As population density increases, average sectional scheme property price decreases.
b) As QoL Index Scores increase, average sectional scheme property price increases.
c) As Industrial Buildings within 2km increases, average sectional scheme property price decreases.
d) As distance to the Closest Private Tertiary Education increases, sectional scheme property price decreases (closer to the amenity means higher property price).
e) As distance to the Closest Formal Taxi Rank increases, sectional scheme property price decreases (closer to the amenity means higher property price).
f) As distance to the Closest Private Hospital increases, sectional scheme property price decreases (closer to the amenity means higher property price).

b) QoL Index Score

The second model, the main spatial factors from the Nodal Review that influence QoL score in the CoJ, can be summarised in the following formula:

$$
\begin{array}{rrcl}
OQT\ \widehat{Transformed}\ (y) = & 0.05134600 & & \\
 & -0.00000409 & \text{x} & \text{Population Density} \\
 & -0.00005741 & \text{x} & \text{Industrial Buildings within 2km} \\
 & -0.00000268 & \text{x} & \text{Closest Registered Tertiary Education (m)}
\end{array}
$$

Where y is "QoL Index score" and the hat (^) indicates a predicted value.

This means that, with all factors remaining equal, and based on the methods and data used in this work:

a) As Population Density increases, QoL Index Score decreases.
b) As Industrial Buildings within 2km increases, QoL Index Score decreases.
c) As distance to the Closest Registered Tertiary Education increases, QoL Index Score decreases (closer to the amenity means higher QoL Index score).

4.2 Comparison of three measures of Urban Value in the CoJ

This section outlines the processes and results of comparisons between the three measures of Urban Value used in this work, being Urban Potential from the CoJ Nodal Review, Rateable Land Value (from the 2018 CoJ valuation Roll) and Quality of Life index. For the Nodal review, the Residential Densification Index is used as opposed to the Economic Nodes Index (City of Johannesburg , 2020a). This is to keep the scope of the work contained, but also to keep with the decision to consider urban value from a residential perspective (see Chapter 1: Introduction). However, in the case of rateable value, all property in the City is used, and not just residential. This is so that a full coverage of property value is seen across the City. The section starts by outlining the data preparation methods, before presenting simple visual comparisons and clustering tests on the data. Tjøstheim's coefficient (1978) is then used to analyse the relationships between the three representations of urban value. Lastly, an intersection analysis is done between areas of high potential for residential densification (Nodal Review) and property value (valuation roll) in an attempt to show low price and high potential land, intended as an indicator for housing development potential.

4.2.1 Data Preparation

Quality of Life Survey QoL index

Data for the QoL Index was used as per the regression analyses presented earlier in section 4.1. The QoL however is representative at a ward level, so the datasets used are at two scales (for different analyses) being at the 400m hexagon level of the Nodal Review (Figure 7 pg. 36), and the ward level (Figure 13 below). As mentioned earlier, the QoL data were joined to the 400m hexagons using the ward into which the centre of each hexagon falls. The ward level data was obtained by averaging the raw QoL scores by ward, and then joining that data to the ward shapefiles.

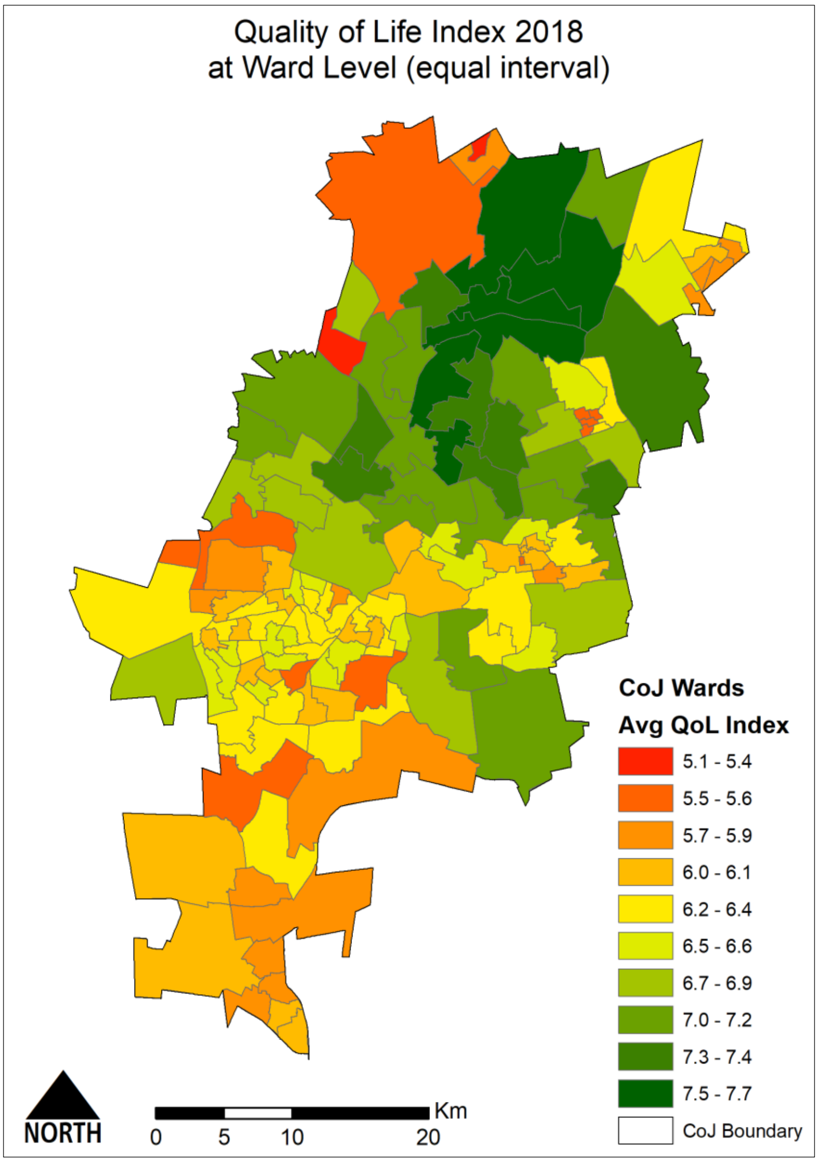

Figure 13: GCRO Quality of Life Index 2018 average values per CoJ ward (map by author)

Nodal Review Residential Densification Potential Index

For the Nodal review, the data from the Nodal Review Indexes was joined to the 400m hexagon shapefiles provided by the City. Then, using zonal statistics in ArcMap, the values were aggregated to a ward level. The full geo-processing model (developed by the author) for aggregating small to larger polygons is seen in the annexure, Figure 23 pg. 86. The data at a 400m hexagon and ward level are shown in Figure 14 below.

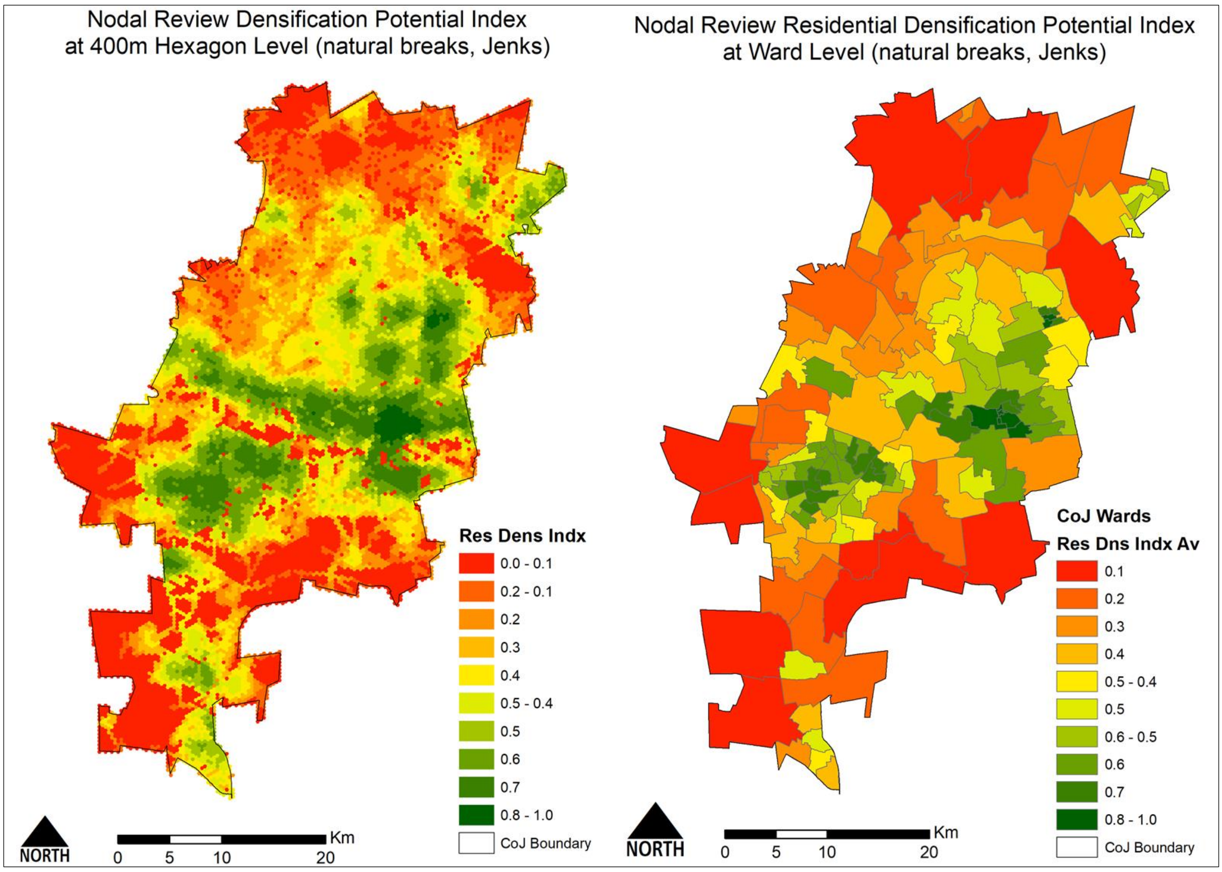

Figure 14: Nodal Review Densification Index at its original 400m Hexagon level (left) and aggregated to a ward level (right) (map by author)

Rateable Value per Square Metre

The least straightforward of the measures of value to map at the 400m hexagon and ward level, was rateable value per square metre, from the CoJ Valuation Roll 2018. This is because of inconsistency in the property shapefiles (property layer from the CoJ Land Information System - LIS) to which the valuation roll data is joined. The major anomaly is that *some* (albeit a vast minority) properties seem to be made up of more than one polygon in the LIS system, although the valuation roll assigns value to just one polygon (through the attribute 'Property ID'). Thus, if one simply joins the data to the property layer, and calculates value per area, sometimes very high numbers are returned when all of the value of a property is allocated to just a small portion of it. Numerous examples of this were seen across the City, with one such shown in Figure 15 below.

Figure 15:Anomaly in the way value from the Valuation Roll 2018 is in some cases Related to Property, causing a concentration of property price. Transparent areas (where aerial photographs are seen) indicate areas with zero or null values.

Figure 15 shows the example of the University of the Witwatersrand (Wits), in Braamfontein, Johannesburg. Coloured-in properties (blue to orange scale) represent erven that have a property value associated with them, with transparent properties (where the aerial photograph can be seen) indicating properties with a zero or null value associated with them. The orange and yellow blocks near the corner of Jan Smuts Avenue and Jorrison Street, are assumed to be associated with the whole value of Wits' campus (where the large part of the aerial photograph can be seen). The orange block alone has a value, according to the valuation roll, of some 2.6 billion rand. When the rand per square metre then is calculated, an unrealistic value of 5.3 million rand per square metre is returned for that block. These anomalies were only discovered when aggregating the values to 400m hexagons, and some very large and seemingly unrealistic outliers were seen.

To solve the problem above, Thiessen polygons were used. “Thiessen polygons, also known as Voronoi polygons, are generated around a set of points in a given space by assigning all locations in that space to the closest member of the point set. Any location in a Thiessen polygon is closer to the corresponding point inside it than to any other member of the point set” (Yamada, 2016, p. 1). Thiessen polygons were developed as a method to delineate the areas represented by weather stations (points) that collect precipitation data (Thiessen, 1911). In this work, they are used to help calculate rateable value per area. Due to the problem with some of the property polygons mentioned above, new polygons needed to be generated that represent the area of each valued property in the City. The method developed creates polygons that represent *gross* rateable value per square metre, as any properties that may not contain an official value (public roads, parks, etc.) are incorporated into the Thiessen polygons created by property value points. As such, the method is an aggregation, and cannot be used to calculate the value of individual properties very accurately. It is useful however for calculating average property price per square metre for larger polygons, such as suburbs, wards or 400m hexagons, as is done in this work. The method for creating the layer representing rateable value per square metre is summarised below, and the geo-processing model (ArcMap) created is shown in Figure 24 pg. 87.

Summarised Method:

a) From the CoJ property layer, remove “Remainder of Township” Erven (mainly roads), and Erven where zoning is “null”.
b) Non-Sectional-Scheme Property
    a. Select all non-sectional scheme (SS) property from the Valuation Table.
    b. Join valuation roll (non-SS) to the property layer using Property_ID as the common field.
    c. Remove properties where the value is zero or NULL.
c) Sectional Scheme Property
    a. Select all sectional scheme (SS) property from the Valuation Table.
    b. Perform a many to one join of sectional scheme unit values, to the property layer using Property_ID as the common field.
    c. Convert properties to points, using the centroid of the property.
    d. Spatially join the points above to the Sectional Scheme layer, summing the values of individual units (points created above, each representing a dwelling unit) to give a

total value per sectional scheme (e.g. for a block of flats, the summed value of all individual flats, assigned to the property containing the building).

d) Merge the Sectional Scheme, and non-sectional scheme layers.
e) Convert the merged layer to points, using centroids. This leaves a point layer with the value of the property it represents.
f) Create Thiessen Polygons using the point layer above and clip the layer to the CoJ boundary.
g) Calculate rand per square metre, by dividing value by area of each polygon.

The resulting layer, representing gross rateable value per square metre across the CoJ is shown in Figure 16 below with a detail example shown in the annexure, Figure 25 pg. 88.

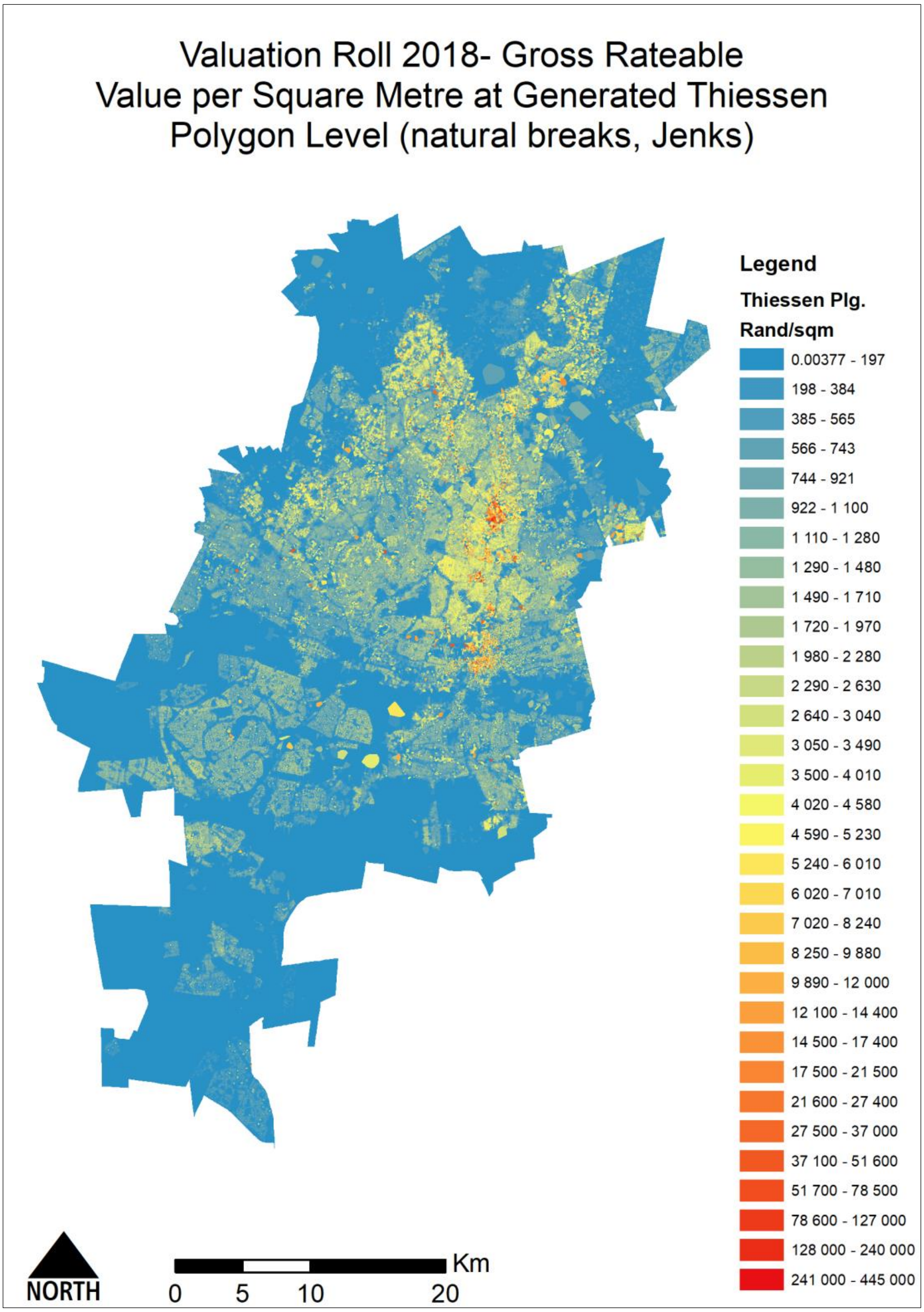

Figure 16: Valuation Roll 2018- Gross Rateable Value per Square Metre at Generated Thiessen Polygon Level (Map by Author)

The data were then aggregated using the same aggregation method described earlier (Figure 23 pg. 86) to 400m hexagon, and ward level, with the results shown in Figure 17 below.

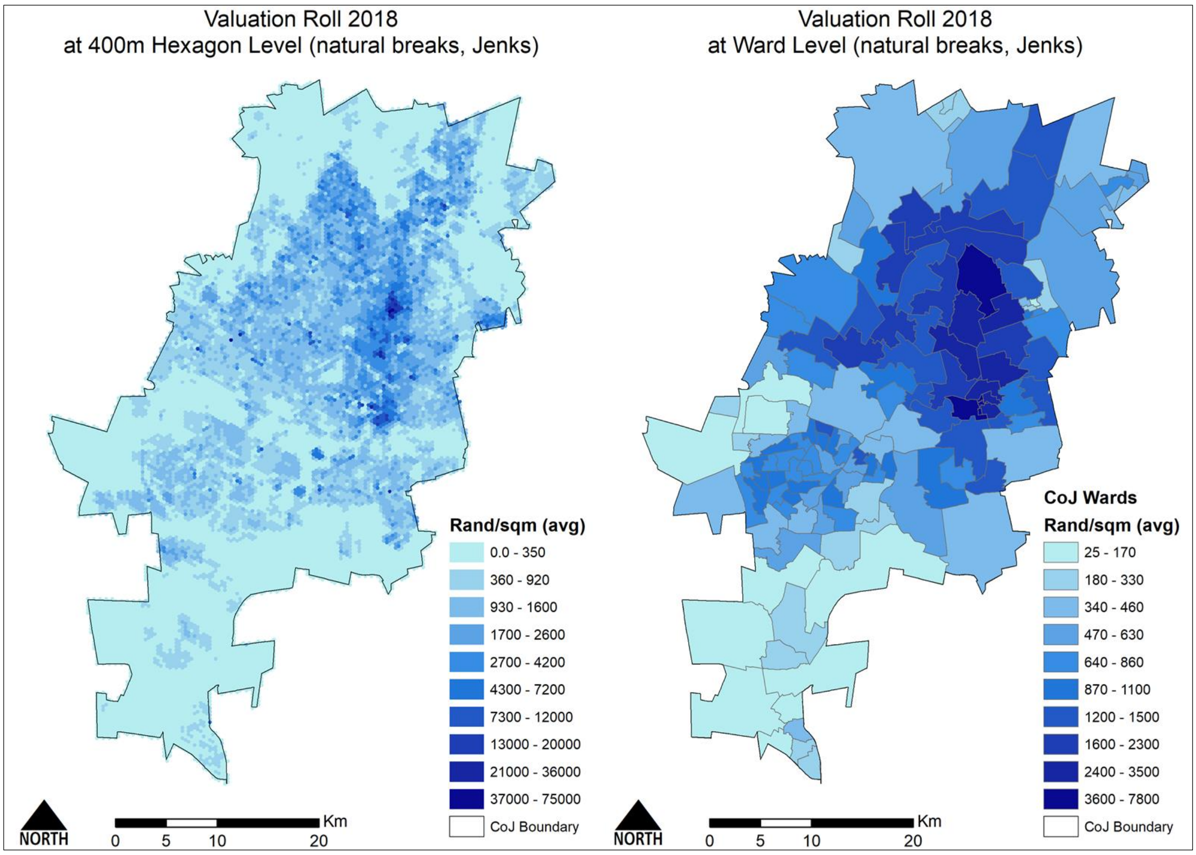

Figure 17: Average Rand Value per square metre across the City of Johannesburg (2018 CoJ Valuation Roll) at 400m Hexagon level (left) and aggregated to a ward level (right) (map by author)

4.2.2 Visual Comparison of the three Measures of Value

The first analysis in comparing the three datasets representing different measures of Urban Value in the City, is a simple visual analysis. This is done at the ward level and is presented in two sets of maps. Both sets of maps contain the same data but use different methods to divide the data into ten subset intervals each. Figure 18 below shows the three datasets using equal intervals between categories (indicated by different colours in the legend). Figure 19 below then presents the same data, but with Jenks' Natural Breaks (Jenks, 1967) to divide the categories, and colours in the legend. The first of the two images is intended to show how the concentration of urban value differs across the different measures, and the second to help compare areas of high and low urban value for the three measures across the City. In all cases, green is used for high values, and red for low values. No interpretation of the results is done here. That is contained in the concluding chapter of the work.

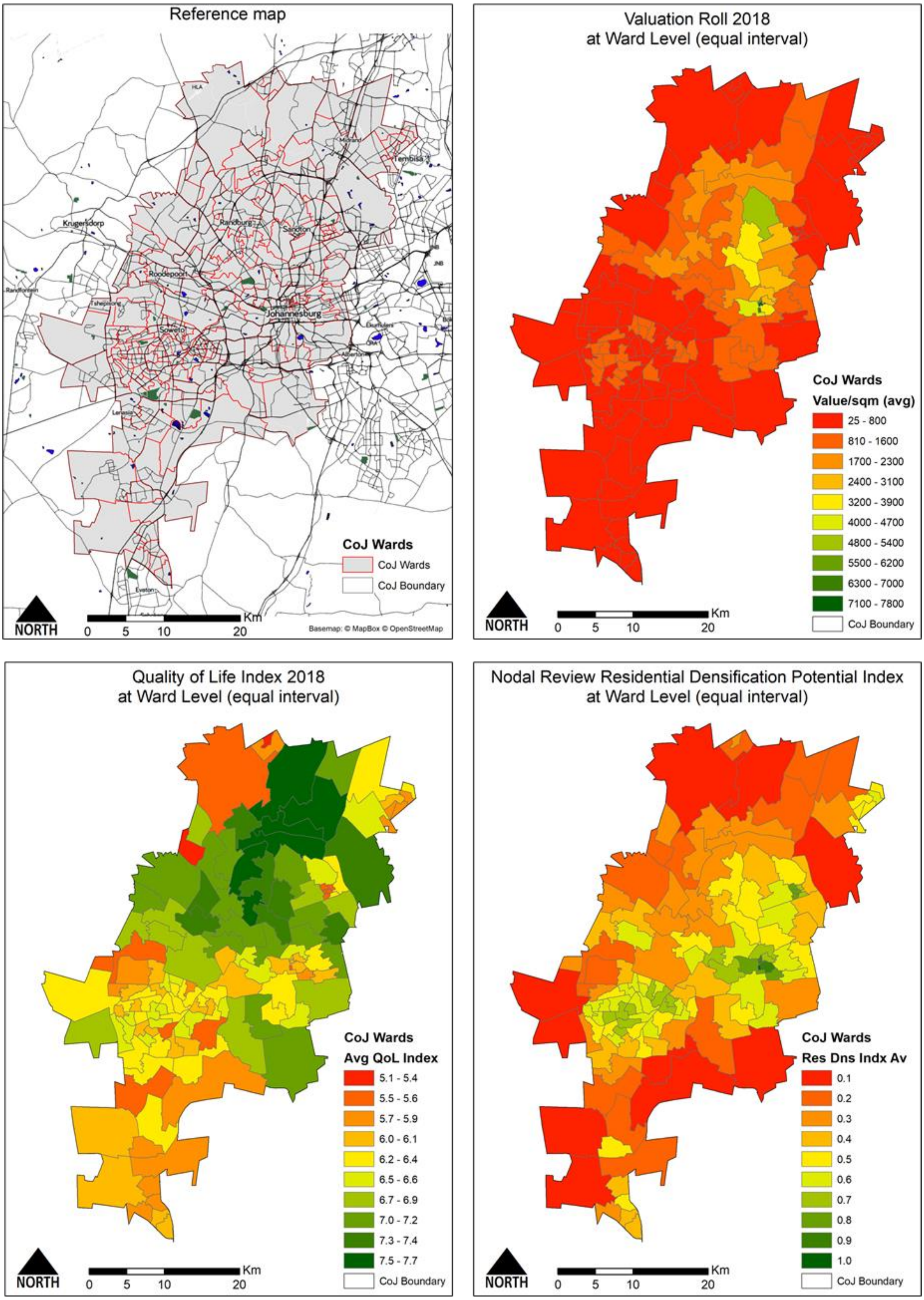

Figure 18: Visual Comparison of three Measures of Urban Value at a Ward Level using Equal Intervals between value range categories (maps by author)

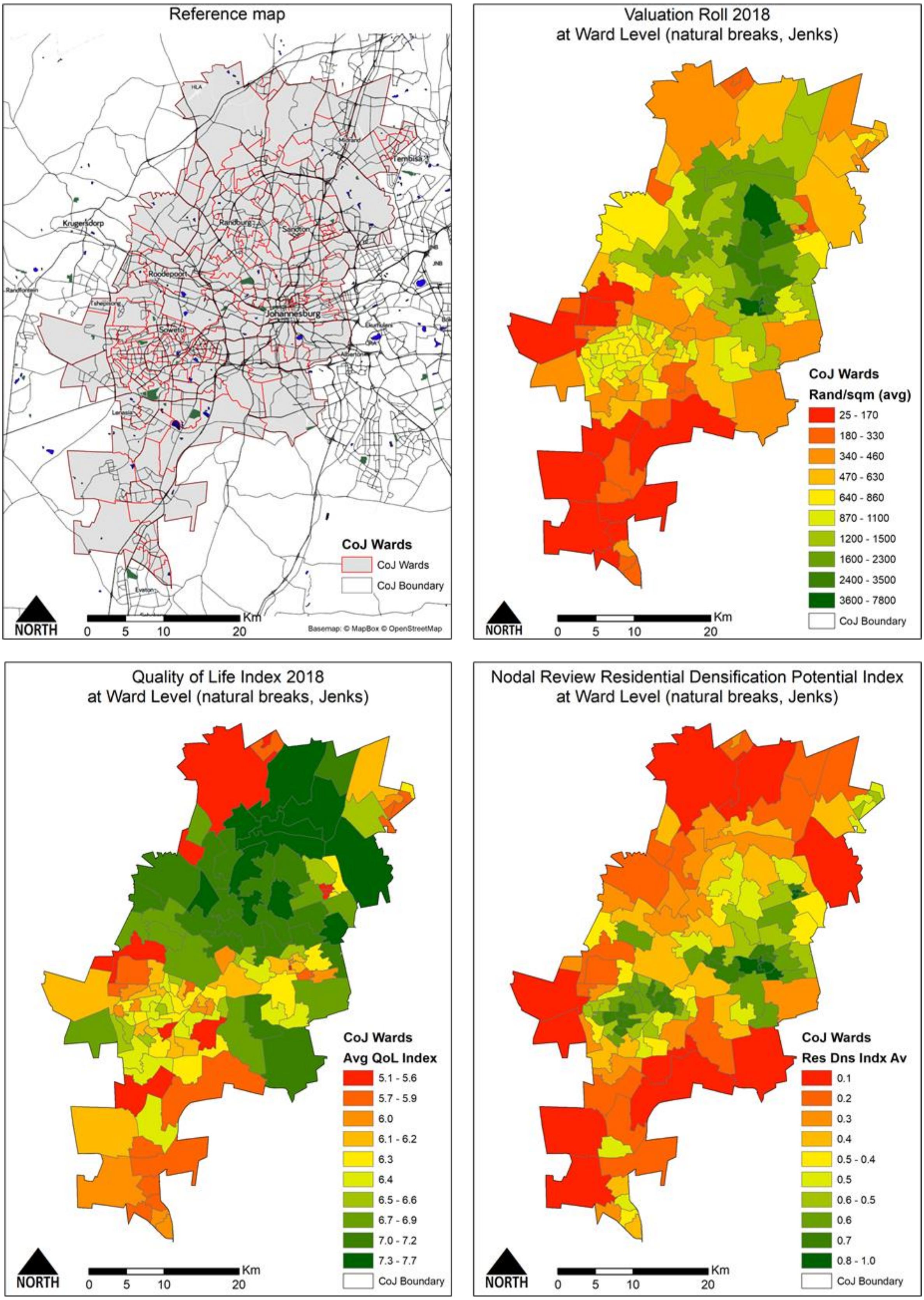

Figure 19: Visual Comparison of three Measures of Urban Value at a Ward Level using Jenks' Natural Breaks between value range categories (maps by author)

4.2.3 Clustering of Measures of Urban Value

Moran’s I test for Spatial Autocorrelation

To add to the visual comparisons above, a statistical test was done to assess the difference in clustering in the three measures of urban value. As comparisons at the same scale was sought, the finest grain available for all three datasets was used, being ward level (as the QoL data is only representative at Ward level). A Moran’s I test was performed on each of the datasets, using 8 nearest neighbours (wards) to create the weights matrix. The results are shown in Table 8 below.

Table 8: Moran’s I test for Spatial Autocorrelation of the three measures of urban value being compared, at a ward level

Moran’s I Test For Spatial Autocorrelation (ward level)			
Variable	**Moran's I Statistic**	**P-Value**	**Indicating**
Rateable Value	0.596	0.001	clustering
Residential Dens. Index	0.549	0.001	clustering
QoL Index	0.348	0.001	clustering
	Alternative Hypothesis = Greater		

From Table 8 above, the p-values of 0.001 (significantly below 0.05), allow us to conclude in favour of the alternative hypothesis, which states that the spatial pattern is correlated, for all three sets of data. The Moran’s I statistics, all significantly above 0, all indicate that the datasets are spatially clustered with Rateable Value being the most spatially concentrated (0.596), Residential Density Index second most (0.549) and QoL index scores least spatially clustered of the three (0.348). This does seem to correlate with the visual results shown in Figure 18 pg. 57, and are discussed in the concluding chapter.

Clustering of Rateable Value: Valuation Roll 2018

As the Moran’s I test above showed such significant clustering of value for the valuation roll data, along with the stark visual clustering shown in Figure 16 pg. 55, a graph was created to quantify this clustering, comparing the percentage of land area to the percentage of value it contains. To do this, the Thiessen polygons and associated rand value per square metre (created in section 4.2.1) was used. This table was exported to spreadsheet format and sorted from smallest to largest in terms of value per square metre. In that order, accumulative value and accumulative area (and associated percentages) were calculated, and plotted on a scatter plot, shown in Figure 20 below. The same analysis was not done for the other two measures of urban value, as their values come in the form of indexes, that are not accumulative in the same way that property price is. One could not say for example that x% of the Quality of Life in the City is concentrated on y% of its land mass, as the indexes are not quantities, but scales.

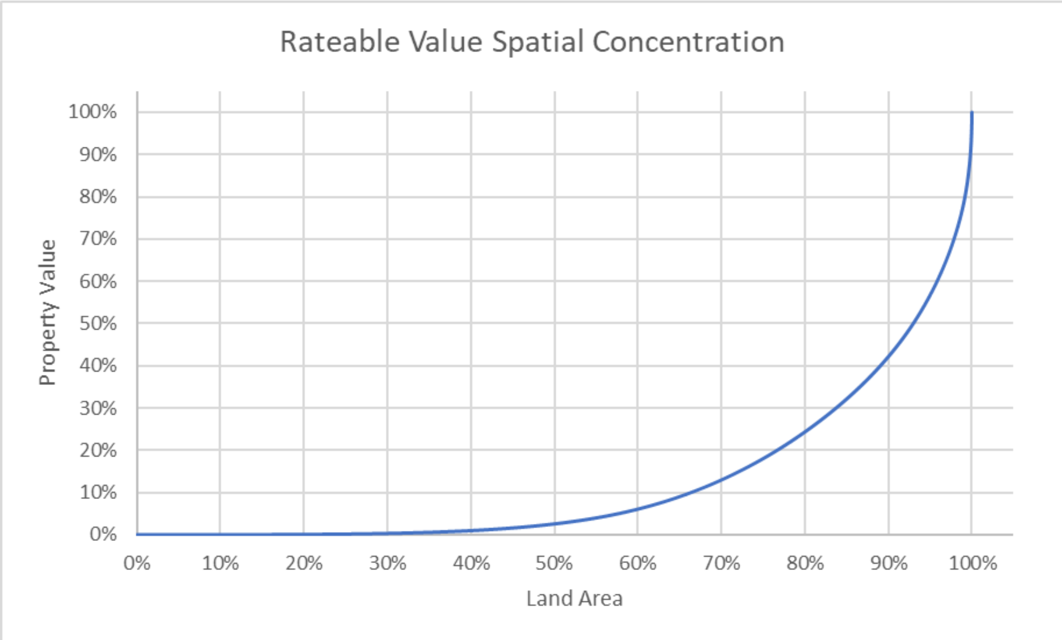

Figure 20: Spatial Concentration of Rateable Value in the CoJ. Percentage of value vs percentage of land

While the key interpretations of Figure 20 are included in Table 9 below, the graph shows just how spatially concentrated rateable value is in the City. The first thing to note is that the least valuable 40% of land in the city makes up only 1% of all rateable value, with the remaining 60% of land containing 99% of the City's Rateable value. This then gets more concentrated as smaller portions of the City's most valuable land are measured, until it is seen that the most expensive 2% of land contains almost 30% of the City's total value, while the most expensive 1% of land, contains over 20% of the total value.

Table 9: Table interpreting Figure 20: Spatial Concentration of Rateable Value in the CoJ. Percentage of value vs percentage of land

40% of the CoJ's land area contains **1.0%** of its rateable value
60% of the CoJ's land area contains **99.0%** of its rateable value
50% of the CoJ's land area contains **97.4%** of its rateable value
40% of the CoJ's land area contains **93.9%** of its rateable value
30% of the CoJ's land area contains **87.0%** of its rateable value
20% of the CoJ's land area contains **75.7%** of its rateable value
10% of the CoJ's land area contains **57.8%** of its rateable value
5% of the CoJ's land area contains **43.2%** of its rateable value
2% of the CoJ's land area contains **29.1%** of its rateable value
1% of the CoJ's land area contains **21.7%** of its rateable value

4.2.4 Tjøstheim's coefficient to test the correlation between the three measures of urban value

The intention of the spatial correlation analysis was only to test if the relationship between the three is positive or negative, and the strength of the relationship. Again, this was done at ward level as the only common scale between the three. The test was run in the software package R using the 'cor.spatial' function (Osorio, Vallejos, Cuevas, & Mancilla, 2020).

Table 10: Results of Tjøstheim's coefficient to test the correlation between the three measures of urban value

Rateable Value	**vs**	**Quality of Life Index**
Coefficient		Variance
0.2633		0.0075

Rateable Value	**vs**	**Nodal Res. Dens. Index**
Coefficient		Variance
0.2793		0.0075

Quality of Life Index	**vs**	**Nodal Res. Dens. Index**
Coefficient		Variance
0.0575		0.0075

The results from the simple spatial association test above show that both the Quality of Life Index Score and the Nodal Review Densification potential index are positively correlated to Rateable value, with values of 0.263 and 0.279 respectively. In light of these results, it is somewhat surprising that the test shows little spatial association between QoL score and the Nodal Review Index, however these results are further discussed and analysed in the concluding chapter.

4.2.5 Intersecting high urban potential with low property price to indicate areas of housing development potential

Two maps were created, each with a slightly different approach to defining areas of high potential for urban development. Both maps use the same source of data for land value (the CoJ valuation roll at hexagon level, Figure 17 pg. 56). Figure 21 below shows land value in all areas that score 0.5 or higher for the residential densification potential index from the Nodal Review (City of Johannesburg , 2020a). Figure 22 below shows all areas from the Nodal Review policy (City of Johannesburg, 2020), (Excluding the LED Zone[11]) where only minimum and no maximum residential densities are promoted. The Zones included for the second map are the Inner City, Metropolitan Nodes, Regional Nodes and the General Urban Zone. These zones are shown in the annexure in Figure 26 pg. 89 for reference. The results are shown below, and discussed in the concluding chapter.

[11] The Local Economic Development Zone was excluded because, while it promotes no maximum densities, the intention of the Zone is not densification, but rather the introduction of economic uses to support already high density, mainly residential parts of the City (City of Johannesburg, 2020).

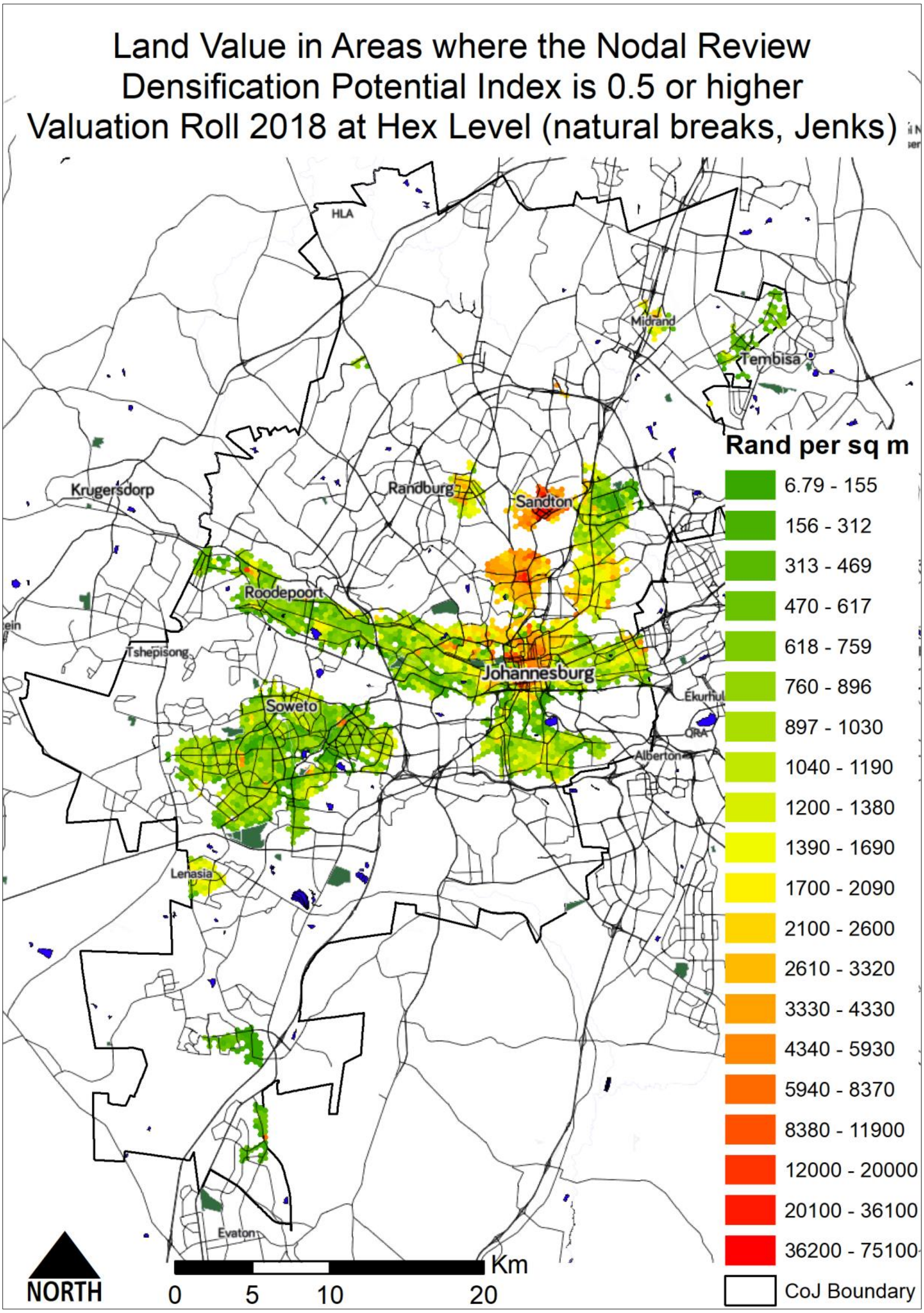

Figure 21: Rateable Value in the CoJ in areas that score at least 0.5 in the Nodal Review Densification Potential Index

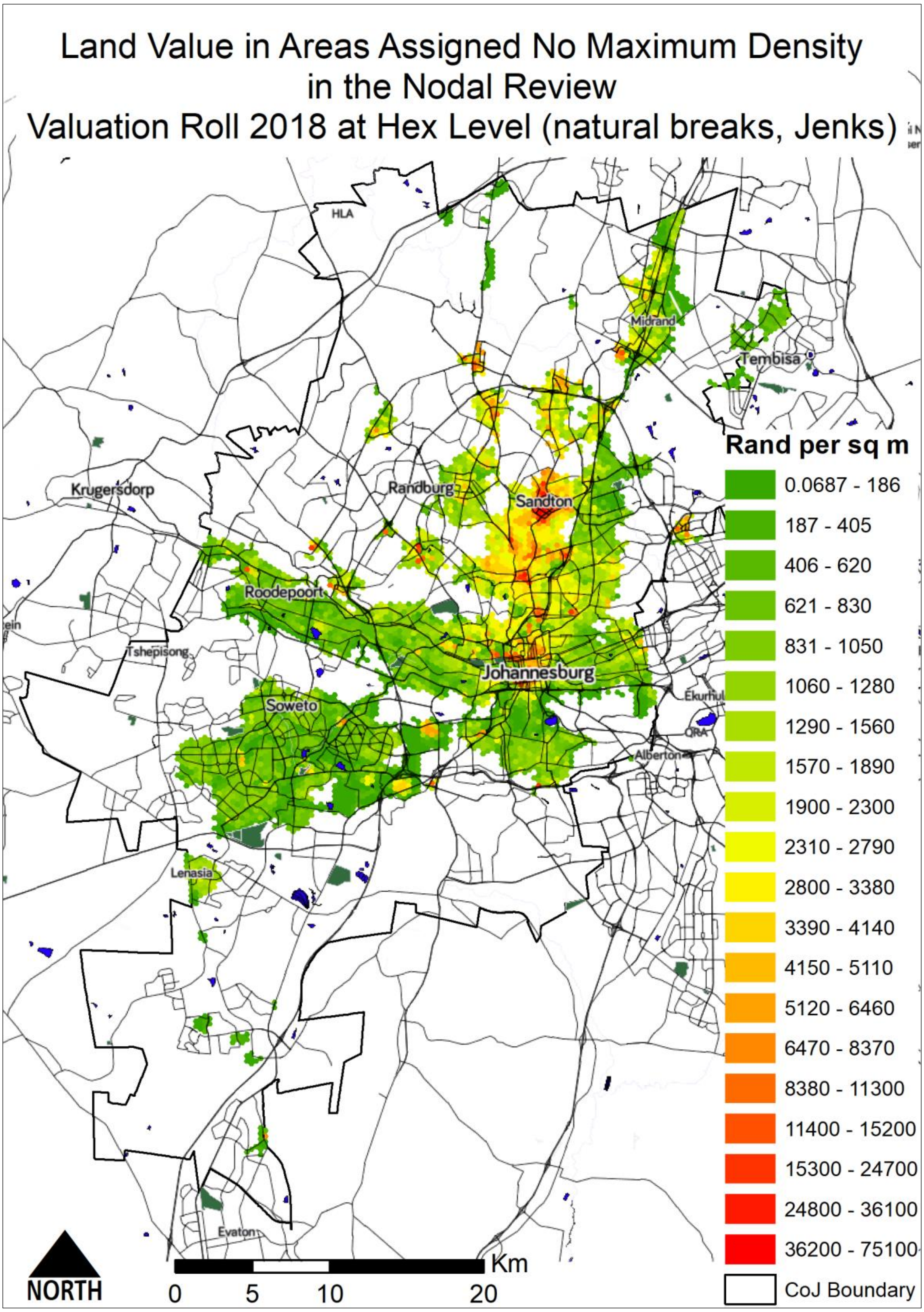

Figure 22: Rateable Value in the CoJ in areas, from the Nodal Review 2020, that are assigned as "no maximum density"

Chapter 5: Discussion and Conclusion

Most of the findings thus far have been presented with little interpretation or discussion. This concluding chapter considers the results of the various spatial analyses conducted and attempts to tie them back to the literature review and explore the implications of the findings from an urban planning perspective. It also relates the work back to the aim of the study; to interrogate the concepts, measures and indicators of urban value in the City of Johannesburg, with the goal of informing better urban planning towards a more spatially just city.

5.1 Hedonic Regression Results: Population Density as an Indicator of Urban Value

The hedonic regression conducted successfully identified (based on the method used) the factors from the Nodal Review and the QoL index that most affect average sectional scheme value, and the factors from the Nodal Review that most affect QoL index scores.

The main notable finding in that for both sectional scheme value, and QoL score, population density is a strong influencing factor. The analysis shows that average sectional scheme value and QoL score both decrease with increased population density. This is significant because of the emphasis placed on densifying Johannesburg in much of its spatial policy (e.g. the Spatial Development framework and the Nodal Review itself) (City of Johannesburg, 2016; City of Johannesburg, 2020), which is influenced by and mirrored in much contemporary planning literature and international city policy (Harrison, Klein, & Todes, 2021). From a residential property point of view, the sectional scheme result may be seen by some (possibly landlords or property developers) as a negative finding. For Johannesburg and this work however, it is seen as a positive as one goal of densification policies is to make housing more affordable in well located parts of the City. The City's Inclusionary housing policy (City of Johannesburg, 2019) promotes smaller units in higher density, well located developments as a way to bring down housing prices, while the Nodal Review states, regarding allowable residential densities:

> *Minimum [density] does however mean that there is no maximum, as long as dwelling units conform to building regulations and the relevant land use scheme, and that development controls are adhered to (height, coverage, FAR, unit size, etc.). **This is to allow developers to build smaller, and thus more affordable units in well located areas (Inner City Node, Metropolitan Nodes, District Nodes and the General Urban Zone)** [emphasis added]*
>
> *(City of Johannesburg, 2020, p. 51)*

The results from this work show that not only are smaller units more affordable (as is common knowledge), but that in higher density areas, average sectional scheme price per square metre is lower than in lower density areas.

Less positive for densification policy and theory, is that the findings show that QoL index scores decrease as population density increases. It is believed however that this is related, at least in part, to two factors. First is that that the QoL index seems to score low density parts of the City very high (even including some rural residential areas). Second is the historic context, that many of the City's high density areas are located in mostly residential areas on the periphery of the City, away from access to amenities and jobs (Harrison, Gotz, Todes, & Wray, 2014; City of Johannesburg, 2016). The first of these factors is seen in Figure 13 pg. 50 where the highest scoring parts of the City (QoL Index) are clustered in the northern half of the City, and most notably around the very low density areas of Carlswald, Bealeau, Kyalami, Blue Hills and surrounds, with many properties being rural-residential smallholdings and many being beyond the Urban Development Boundary (as shown in Figure 26 pg. 89). The second factor is clear that apartheid planning and spatial inequality possibly skew the findings. Many of the City's high-density areas are on the periphery of the City, far from urban amenities and job opportunities (City of Johannesburg, 2016), and score low on the QoL index. Despite this, it is an important finding highlighting that residential densification on its own does not necessarily improve living conditions and quality of life in cities. As the CoJ SDF and Nodal Review emphasise, residential densification must take place in areas with good access to jobs and other amenities and be supported by upgrades in urban infrastructure if the transformative outcomes sought by the policies are to be achieved (City of Johannesburg, 2016; City of Johannesburg, 2020).

The other findings are mainly expected with some surprising results, and some possibly being indicators that relate to some other factors in the urban environment. The fact that sectional scheme price increases with QoL score is expected in the context of Johannesburg discussed immediately above. Interestingly here, a stated preference measure of urban value, the QoL index, correlates to a revealed preference, Valuation Roll, measure of urban value. This also relates to Hurd's (1903) land value is primarily social, even though it is financially costed and traded (see section 2.1).

Interestingly proximity to industrial buildings reduces sectional scheme price and QoL score, which also intuitively makes sense. This is especially related to the specific measure of industrial buildings within 2km which, in the spatial context of the city, is close proximity. The result is also echoed in other similar studies (for residential property price at least) (Wittowsky, Hoekveld, Welsch, & Steier, 2020; Glumac, Herrera-Gomez, & Licherona, 2019). Interesting too, and not necessarily expected, is that sectional scheme average price increases the closer it is to a formal taxi rank (similar to results discussed in the literature review (Bartholomew & Ewing, 2011)). This is positive, showing the importance of access to public transport. As for access to the Closest Private Tertiary Education (which was also a factor for QoL score) and closest Private Hospitals positively impacting sectional scheme property price, it is interesting and logical that these amenities have a positive affect on price. However, it is not clear why they score higher than other amenities, like parks, schools or other forms of public transport like Gautrain. It is possible that these private facilities locate themselves close to higher income residents, rather than the other way around, but more investigation would be needed to make any such conclusions.

5.2 Comparison of the three measures of Urban Value

Section 4.2 presented the results of the various comparisons between the three measures of urban value selected, being property price (CoJ valuation roll 2018) QoL Score and Nodal Review Residential Densification Potential Index. The results are discussed below.

5.2.1 Thiessen Polygons for Mapping Rateable Value

The first notable finding or result is the confirmation of the usefulness of using Thiessen Polygons to map the data, in the context of the peculiarity where (in the datasets) value of some properties in the City is attributed to only a portion of the actual property's extent. As the erven data for the city is largely managed by a central national authority, the Surveyor General's office, it is possible that the problem is not unique to Johannesburg, so the method developed may be useful in mapping value in other municipalities in South Africa too. Whether this is the case or not, the method outlined in Figure 24 pg. 87 is certainly useful for the City of Johannesburg. As the author works in the CoJ administration (at the time of writing) the method will be shared in the administration for use. It would also be useful to apply to method to different valuation roll's (as the valuation roll is updated every 5 years) to compare how concentration of rateable value changes over time. The geo-processing model can be directly used by merely changing the valuation roll and property layer inputs.

5.2.2 Differences in the Clustering of and Correlation between Different Measures of Urban Value

The next analysis, comparing how the datasets differ in terms of correlation and clustering, provided conclusive results. The visual analyses show both aspects (clustering and correlation). Figure 18 on pg. 57 gives the first indication that property price is the most concentrated of the three, followed by Residential Urban Potential and QoL score. This is confirmed by the Moran's I tests which show that all three measures are indeed spatially concentrated (with p values well below 0.05, at 0.01 for all three) with Moran's I statistics of 0.6, 0.59 and 0.35 respectively[12]. Figure 19 on page 58 which uses natural breaks to divide the data into classes for visual representation, better shows how the three are correlated. It shows that land price and residential densification potential are relatively well correlated, while QoL score appears to be less well correlated with the two, while seemingly more similar to the property price. This too is confirmed by the spatial statistic test, Tjøstheim's coefficient. Tjøstheim's coefficient shows the strongest correlations between the pairs rateable value/QoL score, and rateable value/densification potential index, and very little statistical correlation between QoL score and densification potential index.

The reason for the differences is not certain, but the lack of correlation between QoL score and densification potential index is assumed to relate to the different ways in which the data were created. This finding however does warrant further research. As mentioned earlier, the QoL index seems to favour low density, high income areas of the City with, and scores *current* living conditions according to how residents responded to the survey. It thus rates certain parts of the city relatively low, such as greater Soweto, the Inner City and surrounds, Greater Alexandra, and Ivory and Ebony Park, part of larger Thembisa. The Nodal Review urban potential index on the other hand does not consider current living conditions, but rather the *potential* of parts of the City to be developed into "walkable, mixed use neighbourhoods" (City of Johannesburg, 2020, p. 15). The focus on walkability is achieved by measuring access to amenities using 2km walking areas (network distance, by road), that are strongly affected by the density of street networks and intersection density (City of Johannesburg , 2020a). Thus areas with walkable street networks (including those mentioned immediately above) will score relatively higher in the Nodal Review indexes, while low density, high income small holdings that may score high in the QoL index, do not perform well in the Nodal Review indexes due to their with poor walkability in terms of road networks.

The approach taken in the Nodal Review also likely accounts for the high level of clustering in the data, but very notable is just how concentrated rateable value is in the City. This was first illustrated in Figure 16 pg. 55 and in the equal interval visual analysis in Figure 18 pg. 57, and prompted the analysis to test just how concentrated rateable value in the City is. The results in Figure 20 and Table 9 largely speak for themselves, with the headlines being that the cheapest 40% of land in the City contains just 1% of the total rateable value, while more than 20% of the City's value is concentrated on just 1% of the land area. This is expected in cities, where high intensity developments (like office towers or large blocks of flats) concentrate value onto small pieces of land, as higher land use rights increase the value

[12] It is noted that these measures were done at ward level for comparison, and that at their finest spatial grain, the property value (Thiessen polygon's level) and the Nodal Review Index (400m hexagon level) would very likely show much higher levels of spatial clustering.

of land (Smolka & Maleronka, 2018). The detail is shown in Figure 25 pg. 88 highlights the trends, where commercial nodes such as the Inner City, Braamfontein, Rosebank, Melrose Arch, Illovo and Sandton all show relatively high rateable value per square metre. High density residential areas such as Hillbrow, Killarney, Illovo and the flats on the Houghton Golf Course, all also show higher rateable value per square metre than lower density residential suburbs. This is even seen when comparing a suburb like Parkhurst with ~500m^2 residential stands, to Houghton estate, with ~2000m^2 residential properties. While properties in Houghton Estate are known to be much more expensive than those in Parkhurst, the analysis shows that the higher density area (Parkhurst) displays higher rateable value per area. This analysis has important implications for the City's financial operating model. Higher density areas can generate greater income from property rates and, as infrastructure is more efficiently used (shared by more people), can be more financially feasible to provide and maintain (City of Johannesburg, 2020). High Potential Lower Value Land for Residential Development

A major influencing factor for where developers choose to build housing, is land price. This has arguably been one of the reasons why the City of Johannesburg has seen outward sprawl and expansion, for both government and private sector led housing developments, over the past 20 years or so (van Niekerk, 2018). While the Johannesburg SDF and the Nodal Review promote densification in well located areas, formal densification takes place mainly in higher value areas (in and around nodes in the north of the City) or through township establishments in more peripheral areas (City of Johannesburg, 2021). Some lower value areas with high potential for residential densification, and where high density is supported by policy (for example greater Soweto) do not see the same levels of rezoning activity[13] (ibid.). The maps created and shown in Figure 21 pg. 63 and Figure 22 pg. 64 are intended to be used as a resource for developers and the CoJ to highlight possibly overlooked areas for residential densification in the City. The two maps, viewed together, show the potential for investment in residential densification in different categories. First is some formerly black only townships, that when built were on the periphery of the City, but are now relatively well located such as Soweto and Thembisa. Second is the great potential north of the mining belt, following the east-west railway and the Main Reef Road, broadly connecting Roodepoort in the west through Florida, Mayfair and the Inner City to Kensington in the east. Next is some older suburbs south of the Inner City, such as Turffontein and Rosettenville, largely contained in the Turffontein Strategic Area Framework one of the City's 'Corridors of Freedom' (City of Johannesburg, 2014). Fourth is in and

[13] This refers to formal development, for which the City receives land use applications, and not informal or less formal densification such as back yarding. This is not to discount the value of processes like back yarding, but to note that formal investment that contributes to infrastructure improvement, increases in the rates base, results in improved quality of life and security of tenure and possibly greater returns for investors, is also desirable in the City.

around Greater Alexandra, a low income but well located and connected part of the City. And last is in and around some existing economic nodes of the City, with notable examples being Midland, Randburg, and Lenasia. While government has a large role to play in providing housing, the private sector has huge potential to address the housing backlog too, and it is hoped the information contained here can provide some guidance to the City and developers, and diversifying areas of housing investment.

5.3 Limitations of the Study

In terms of the two hedonic spatial regression models, while the approach is considered successful and produced useful results, the main limitation of the approach is the data used. This is in that the different sources of data used were obtained at different spatial scales and converted to the 400m hexagon level. This, especially for the QoL data which was obtained at Ward level and converted to 400m hexagon level. This is seen in Figure 7 pg. 36 where the QoL Scores, although they are at hexagon level, are not variable at the hexagon scale, but rather at the ward level. This may introduce errors in the analysis and is known as the Modifiable Areal Unit Problem (MAUP) (Flowerdew, 2011). This relates to the second limitation, that the three datasets used were not designed to be comparable to one another. They were created for very different uses, and indeed represent quite different phenomena, however this is a general limitation of different types of spatial data, and in the context of this work that actively seeks to compare different types of data, and representations of urban value, it is a necessary limitation. The data is also from one time period only (2018) and thus cannot be generalised over time. Again, this is a necessary practical limitation of the work, in terms of time and availability of data.

5.4 Considerations for further research

Considerations for further research would be firstly to test the effect of the main limitation above, being the MAUP. The first test may be to re-run the analysis, but with all of the data aggregated to a ward level and compare the results. The second suggestion would be to perform interpolation on the input datasets, and then use those modelled surfaces (with a sampling method to create the input data) to run the analysis (Wang, Stein, Gao, & Ge, 2012). This however would be very time consuming, as the interpolation surfaces would need to be created for all variables in the analysis. As mentioned earlier, further research would be needed to investigate some of the less expected results of the spatial regression analyses, including the relationship noted between private tertiary education facilities and private hospitals to QoL Index scores.

It is hoped that the comparison analysis on the three different measures of urban value, along with additional possible future analysis, showing changes to rateable value per square metre over time, and comparisons between infrastructure provision and rateable value, can be used to strengthen arguments for compaction, infill and diversification of land use in the CoJ, as opposed to outward sprawl. It is also hoped, as mentioned earlier, that the Nodal Review policy and land use potential index to property price can help guide housing developers (both private and public sector) to areas that will have high social yields, at more affordable land prices.

In terms to the Thiessen polygon approach to mapping the Valuation Roll Data, a useful exercise to add would be to apply the method to the actual revenue collection data in the City to be able to compare the theoretical rates base (shown by the valuation roll and zoning information) with the actual revenue collected. This could then be applied to future growth models in the City to make revenue projections based on different growth scenarios, such as the SDF and Nodal Review (City of Johannesburg, 2016; City of Johannesburg, 2020). The author intends to do this in his continued work at the City.

5.5 Conclusion

This research has explored, using quantitative spatial datasets, Concepts, Measures and Indicators of Urban Value in the City of Johannesburg. The three datasets used represent different approaches to theoretical concepts of value, with the Valuation Roll representing economic or exchange value obtained using a revealed preference approach, the QoL representing social and environmental value generated using stated preference (survey) and the Nodal Review urban potential model, representing social and environmental value, and generated through spatial modelling and a revealed preference approach. This was an important starting point, choosing different spatial datasets, generated by different groups and for different purposes, that show different types or approaches to urban value.

The research then successfully performed spatial analysis (using GIS methods and spatial statistics) on the three datasets, to meet all of the research objectives. First, using spatial hedonic regression, it was show that, based on the methods used, Population Density, QoL Index Score, Industrial Buildings within 2km, Closest Private Tertiary Education (m), Closest Formal Taxi Rank (m) and Closest Private Hospital (m) are the spatial factors that, together, have the biggest influence on property price in the CoJ. Also using spatial hedonic regression, the research shows that, based on the method used, Population Density, Industrial Buildings within 2km and Closest Registered Tertiary Education (m) are the are the spatial factors that, together, have the biggest influence on QoL index scores in the CoJ. Some of the results are expected, and some less so, but the most significant, in the context of the focus on densification in CoJ and contemporary urban planning theory, is the relationship between population density and both property price and QoL index score. Higher population densities relate to lower property price and lower QoL scores generally, however this is likely related (at least in part) to the apartheid spatial form of the City (where higher densities are historically on the periphery of the City, in areas far from economic activity, jobs and urban amenities) and differences in how the QoL and Nodal Review data were collected and measured. Importantly, it highlights that higher densities do generally result in more affordable housing, even per square metre. But, regarding quality of life, it also raises the important consideration for urban planning that densification on its own, or in the wrong places, will not lead to better urban outcomes. Densification must be a measured process and take place in an integrated manner in areas that have high access to jobs and urban amenities and be accompanied by investment in infrastructure, hard and soft.

The next set of analysis, the comparison between the three datasets representing urban value, showing where they are similar, but importantly that they don’t correlate perfectly. The clustering and correlation analysis shows property price is the most concentrated of the three, followed by Residential Urban Potential and QoL score. The research shows what property price is relatively well spatially correlated to the Nodal Review Urban Potential Densification Index and the Quality of Life Index, but interestingly, the QoL and Nodal Review indexes do not correlate well. This shows that there is a complex interaction of different measures and approaches to urban value in the CoJ, and that the approaches to generating different datasets need to be considered when using them for policy making, investment or other purposes.

Lastly, the research successfully created an index of areas with low property price, and high urban potential, intended as a guide for developers (private and public sector) to develop housing. It shows areas that would have a high social return on investment (in terms of location) and be affordable (in terms of land price).

References

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

Anselin, L. (2003). *An Introduction to Spatial Regression Analysis in R.* Retrieved 25 September, 2019, from University of Illinois: http://labs.bio.unc.edu/buckley/documents/anselinintrospatregres.pdf

Anselin, L. (2005). *Exploring Spatial Data with GeoDa : A Workbook.* Retrieved from Center for Spatially Integrated Social Science: http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf

Anselin, L., & Bera, A. K. (1998). Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics. In A. Ullah, & D. Giles (Eds.), *Handbook of Applied Economic Statistics* (pp. 237-289). New York: Dekker.

Bartholomew, K., & Ewing, R. (2011). Hedonic Price Effects of Pedestrian- and Transit-Oriented Development. *Journal of Planning Literature, 26*(1), 18-34.

Beale, L., Abellan, J. J., Hodgson, S., & Jarup, L. (2008). Methodologic Issues and Approaches to Spatial Epidemiology. *Environmental Health Perspectives, 116*(8), 1105-1110.

Begg, D., Vernasca, G., Fischer, S., & Dornbusch, R. (2014). *Economics* (11th ed.). New York: McGraw-Hill Education.

Bereitschaft, B. (2018). Walk Score® versus residents' perceptions ofwalkability in Omaha, NE. *Journal of Urbanism: International Research onPlacemaking and Urban Sustainability, 11*(4), 412-435.

Brown, G., & Weber, D. (2011). Measuring change in place values using public participation GIS (PPGIS). *Applied Geography, 34*, 316-324.

Brueckner, J. K. (2011). *Lectures on Urban Economics.* Cambridge, MA: MIT Press.

Burnett, P., Brooks, T., & Bassett, P. (2017). Estimating Quality of Life and Place with Location Theory: The McBucks Index. *Social Indicators Research, 134*(2), 455-472.

Carmona, M. (2019). Place value: place quality and its impact on health, social, economic and environmental outcomes. *Journal of Urban Design, 24*(1), 1469-9664.

Charlton, S., & Kihato, C. (2006). Reaching the poor? An analysis of the influences on the evolution of South Africa’s housing programme. In U. Pillay, R. Tomlinson, & J. du Toit (Eds.), *Democracy and Delivery: Urban Policy in South Africa.* Pretoria: HSRC Press.

City of Johannesburg . (2020a). Annexure 2: Urban Potential Modelling Method. In C. o. Johannesburg, *Nodal Review.* Johannesburg: City of Johannesburg.

City of Johannesburg. (2014). *Turffontein Development Corridor Strategic Area Framework.* Johannesburg: City of Johannesburg.

City of Johannesburg. (2016). *Spatial Development Framework 2040.* Retrieved October 21, 2016, from City of Johannesburg: http://bit.ly/cojSDF2040

City of Johannesburg. (2019). *City of Johannesburg Land Use Scheme 2018.* Johannesburg: Gauteng Provincial Gazette.

City of Johannesburg. (2019). *Inclusionary Housing Incentives, Regulations and Mechanisms.* Retrieved from City of Johannesburg: http://bit.ly/cojIH

City of Johannesburg. (2020). *Nodal Review.* Retrieved May 1, 2019, from City of Johannesburg: Citywide Spatial Policies: http://bit.ly/cojcitywide

City of Johannesburg. (2021). *Development Report Johannesburg: 2020/21.* Retrieved May 26, 2021, from Google Data Studio: https://bit.ly/cojtasreport

D'Acci, L. (2012). *Isobenefit Lines, Breaking Point of equal attraction, Uniformity Benefit, Variety Value and Proximity Value, Preference Gap Gain.* Retrieved March 13, 2020, from Cornell University: https://arxiv.org/abs/1210.7510

D'Acci, L. (2014). Monetary, Subjective and Quantitative Approaches to Assess Urban Quality of Life and Pleasantness in Cities. *Social Indicators Research, 115*(2), 531-559.

Flowerdew, R. (2011). How serious is the Modifiable Areal Unit Problem for analysis of English census data? *Population Trends*(145), 106-118.

Fogarty, M. (1996). *A History of Value Theory.* Retrieved July 3, 2019, from Student Economic Review: Trinity College Dublin: https://www.tcd.ie/Economics/assets/pdf/SER/1996/Martin_Fogarty.html

Freudenberg, N., Galea, S., & Vlahov, D. (2005). Beyond urban penalty and urban sprawl: back to living conditions as the focus of urban health. *Community Health, 30*(1), 1-11.

GCRO. (2018). *Quality of Life survey V (2017/18)*. Retrieved July 18, 2019, from Gauteng City Region Observatory: https://gcro.ac.za/research/project/detail/quality-of-life-survey-v-201718/

Glumac, B., Herrera-Gomez, M., & Licherona, J. (2019). A Hedonic Urban Land Price Index. *Land Use Policy, 81*, 802-812.

Goodchild, M. F. (2011). Scale in GIS: An overview. *Geomorphology, 130*, 5-9.

Graham, S. R., Carlton, C., Gaede, D., & Jamison, B. (2011). The Benefits of Using Geographic Information Systems as a Community Assessment Tool. *Public Health Reports, 126*(2), 298-303.

Hamilton, M. C., Nedza, J. A., Doody, P., Bates, M. E., Bauer, N. L., Voyadgis, D. E., & Fox-Lent, C. (2016). Web-based geospatial multiple criteria decision analysis using open software and standards. *International Journal of Geographical Information Science, 30*(8), 1362-3087.

Harrison, P., Gotz, G., Todes, A., & Wray, C. (2014). Materialities, subjectivities, and spatial transformation in Johannesburg. In P. Harrison, G. Gotz, A. Todes, & C. Wray (Eds.), *Changing Space Changing City: Johannesburg After Apartheid* (pp. 2-41). Johannesburg: Wits Press.

Harrison, P., Klein, G., & Todes, A. (2021). Scholarship and policy on urban densification: perspectives from city experiences. *International Development Planning Review, 43*(2), 151-173.

Hill, R. J. (2013). Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy. *Journal of Economic Surveys, 27*(5), 879–914.

Hubert, L. J., & Golledge, R. G. (1982). Measuring Association between Spatially Defined Variables: Tjøstheim's Index and Some Extensions. *Geographical Analysis, 14*(3), 273-278.

Hurd, R. M. (1903). *Principles of City Land Values.* New York: The Record and Guide.

IUCN, UNEP & WWF. (1991). *Caring for the Earth: A Strategy for Sustainable Living.* London: Earthscan.

Jenks, G. F. (1967). The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography, 7*, 186–190.

La Rosa, D. (2014). Accessibility to greenspaces: GIS based indicators for sustainable planning in a dense urban context. *Ecological Indicators, 42*, 122-134.

Lainga, J., Wheeler, F., Reeves, K., & Frost, W. (2014). Assessing the experiential value of heritage assets: A case study of aChinese heritage precinct, Bendigo, Australia. *Tourism Management, 40*, 180-192.

Lumley, T. (2020). *Package 'leaps'.* Retrieved March 20, 2021, from Cran: https://cran.r-project.org/web/packages/leaps/leaps.pdf

McCain, P. (2013). *Modern Urban and Regional Economics* (2nd ed.). Oxford: Oxford University Press.

McDonald, J. F., & McMillan, D. P. (2011). *Urban Economics and Real Estate: Theory and Policy* (2nd ed.). Danvers: Wiley.

Moran, P. A. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika, 37*(1), 17-23.

Municipal Demarcation Board. (2016). *2016 Boundaries – District Municipalities.* Retrieved March 2019, 1, from Municipal Demarcation Board: http://www.demarcation.org.za/site/documents/2016-boundaries-district-municipalities/

Nagelkerke, N. J. (1991). A Note on a General Definition of the Coefficient of Determination. *Biometrika, 78*(3), 691-692.

Nanda, A. (2019). *Residential Real Estate: Urban and Regional Economic Analysis.* Routledge: New York.

Osorio, F., Vallejos, R., Cuevas, F., & Mancilla, D. (2020). *Tools for Assessment the Association Between Two Spatial.* Retrieved from CRAN: https://cran.csiro.au/web/packages/SpatialPack/SpatialPack.pdf

O'Sullivan, A. (2012). *Urban Economics* (8th ed.). New York: McGraw-Hill Education.

Parsons, A. (2010). *The seven myths of slums.* London: Share the Worlds Resources.

Peterson, R. A. (2021). *Using the bestNormalize Package*. Retrieved April 21, 2021, from https://cran.r-project.org/web/packages/bestNormalize/vignettes/bestNormalize.html#the-ordered-quantile-technique

Powe, N. A., Garrod, G. D., Brunsdon, C. F., & Willis, K. G. (1997). Using a geographic information system to estimate an hedonic price model of the benefits of woodland access. *Forrestry, 40*(2), 139-149.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool* (2nd ed.). New York: Springer.

Rural Development and Land Reform. (2011). *The South African Coordinate Reference System.* Retrieved April 2, 2019, from Rural Development and Land Reform: http://www.ngi.gov.za/index.php/resources?download=37:the-south-african-coordinate-reference-system

Rural Development and Land Reform. (2013). *Datums and Coordinate Systems.* Retrieved April 19, 2019, from Rural Development and Land Reform: http://www.ngi.gov.za/index.php/technical-information/geodesy-and-gps/datum-s-and-coordinate-systems

SA Commercial Prop News. (2019, February 8). *Trendy Rosebank gets lion's share of property investment.* Retrieved July 17, 2019, from SA Commercial Prop News: http://bit.ly/2YfHdHn

Sandler, R. (2012). Intrinsic Value, Ecology, and Conservation. *Nature Education Knowledge, 3*(10), 4.

Saunders, D. (2010). *Arrival City.* London: William Heinemann.

Schäffler, A., Christopher, N., Bobbins, K., Otto, E., Nhlozi, M. W., de Wit, M., . . . Phasha, P. (2015). *State of Green Infrastructure in the Gauteng City Region.* Johannesburg: Gauteng City Region Observatory.

Słomska-Przech, K., & Golebiowska, I. M. (2021). Do Different Map Types Support Map Reading Equally? Comparing Choropleth, Graduated Symbols, and Isoline Maps for Map Use Tasks. *International Journal of Geo-Information, 10*(69), 1-20.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations* (MetaLibri Digital Library 2007 ed.). (S. M. Soares, Ed.) London: W. Strahan and T. Cadell.

Smolka, M. O., & Maleronka, C. (2018). Assessing the monetary relevance of land value capture: the case for charges for additional building rights in São Paulo, Brazil. *International Journal of Real Estate and Land Planning, 1*, 124-133.

Stats SA. (2020). *Mid-year population estimates, 2020*. Retrieved Feb 10, 2021, from Statistics SA: http://www.statssa.gov.za/?page_id=1854&PPN=P0302

Supply Chain Sustainability School. (2017). *Social Value and Design of the Built Environment.* London: Social Value and Design of the Built Environment. Retrieved July 17, 2019, from http://bit.ly/2JAozkZ

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55.

Thiessen, A. H. (1911). Precipitation Averages for Large Areas. *Monthly Weather Review*, 1082–1089.

Tjosteim, D. (1978). A measure of association for spatial variables. *BionuUrika, 65*(1), 109-114.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography, 46*, 234-240.

Torraco, R. J. (2016). Writing Integrative Literature Reviews: Using the Past and Present to Explore the Future. *Human Resource Development Review, 15*(4), 404-428.

Turok, I., & McGranahan, G. (2013). Urbanization and economic growth: the arguments and evidence for Africa and Asia. *Environment & Urbanization, 25*(2), 465–482.

UN-Habitat. (2016). *World Cities Report: Urbanization and Development - Emerging Futures.* Nairobi: UN-Habitat.

Vallejos, R., Osorio, F., & Bevilacqua, M. (2020). *Spatial Relationships Between Two Georeferenced Variables.* Cham: Springer.

van der Waerden, B. (1952). Order tests for the two-sample problem and their power. *Indagationes Mathematicae (Proceedings), 55*, 453-458.

van Niekerk, B. (2018). Housing as urbanism: A policy to discourage urban sprawl and provide well-located and affordable housing in South Africa. *Town and Regional Planning, 73*, 68-82.

Van Wyk, J. (2015). Can SPLUMA play a role in transforming spatial injustice to spatial justice in housing in South Africa? *Southern African Public Law, 30*(1), 26-41.

von Thünen, J. H. (1826). *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie.* Hamburg: Perthes.

Walk Score. (2020). *Walk Score Methodology.* Retrieved March 13, 2020, from Walk Score: https://www.walkscore.com/methodology.shtm

Wang, J.-F., Stein, A., Gao, B.-B., & Ge, Y. (2012). A review of spatial sampling. *Spatial Statistics, 2*, 1-14.

Weakley, D. (2009). Water Demand Management in the City of Johannesburg. Johannesburg: University of the Witwatersrand.

Weakley, D. (2013). Recognising Vulnerability and Resilience in Informal Settlements: The Case of Kya Sands, Johannesburg, South Africa. *MSc. Dissertation (Town and Regional Planning)*. Johannesburg: University of the Witwatersrand.

Weakley, D. (2019). *Locational Factors Affecting Sectional-Scheme-Residential-Property-Price in the City of Johannesburg*. Retrieved from Google Drive, Assignment Submitted for the Course STAT7064A: http://bit.ly/dw_stats_proj

Weakley, D. (2019a). *Predicting Residential Sectional Scheme Price in the City of Johannesburg using Spatial Lag Hedonic Regression*. Retrieved from Google Drive: Submitted assignment as part of the course STAT7006A: https://bit.ly/dw_spatial_proj

Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: impact of housing characteristics, accessibility and neighbouring apartments – a case study of Dortmund, Germany. *Urban, Planning and Transport Research, 8*(1), 44-70.

World Bank. (2019). *World Bank Open Data.* Retrieved May 15, 2019, from World Bank: https://data.worldbank.org/

Wyatt, P. J. (1997). The development of a GIS-based propertyinformation system for real estate valuation. *International Journal of Geographical Information Science, 11*(5), 435-450.

Yamada, I. (2016). Thiessen polygons. In *The International Encyclopedia of Geography, online.* Hoboken: John Wiley & Sons, Ltd.

Annexure: Tables, Data, Geo-Processing Models, Extra Maps

Table 11: Sample of dataset used. Only 8 records shown (out of 3018) and data transposed (changed columns to rows) to ensure that it fits.

FID	0	1	2	3	4	5	6	7	8
Avg_SS_val	7831.908	11094.905	14700.037	7115.826	8129.080	8686.771	7995.331	7501.622	8125.011
Log_SS_Val	3.894	4.045	4.167	3.852	3.910	3.939	3.903	3.875	3.910
X	27.908	27.911	27.914	27.916	27.907	27.910	27.916	27.913	27.919
Y	-26.086	-26.084	-26.083	-26.094	-26.092	-26.091	-26.097	-26.092	-26.099
Commrc_2km	24.000	24.000	23.000	13.000	16.000	10.000	52.000	8.000	61.000
F2km_Srv_a	5.777	4.810	5.049	3.921	5.314	5.054	6.129	4.454	6.273
Indust_2km	238.000	267.000	270.000	26.000	81.000	36.000	0.000	0.000	0.000
Ind_Com_2k	262.000	291.000	293.000	39.000	97.000	46.000	52.000	8.000	61.000
School_2km	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Clst_Scl_m	2503.408	2340.725	2398.434	3108.457	3457.685	3557.761	3011.239	3863.694	2698.340
ppl_km_11	1521.087	1158.897	638.839	1075.911	1492.149	1214.829	1318.626	1458.995	2777.936
Cl_Pr_s_m	2503.408	2340.725	2398.434	3108.457	3457.685	3557.761	3011.239	3863.694	2698.340
Cl_H_Sc_m	5459.743	6145.941	5820.797	4053.632	4712.742	4405.390	3292.383	4144.838	3454.529
Cl_Tr_S_m	11824.392	12534.984	12592.692	11026.078	11165.654	11377.836	10264.829	11117.283	10426.974
Cls_Clinic	4268.564	4105.882	4163.590	4266.893	4931.819	4624.467	3511.460	4363.914	3662.109
Clnc_in_2k	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Clst_Hsptl	15974.538	16685.130	16458.329	15098.369	15315.800	15455.943	14342.936	15195.390	14493.585
All_P_2k_m	17200.000	11500.000	11500.000	24200.000	10200.000	29700.000	22600.000	25900.000	25300.000
Clst_dv_pk	2977.945	2815.263	2872.971	2998.040	3111.015	3395.909	2508.434	3360.889	2393.256
Cls_unD_pk	766.819	1270.358	1328.066	1173.652	1075.410	1175.486	594.834	1447.288	568.868
Com_in_Hex	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ind_in_hex	0.000	0.000	33.000	0.000	0.000	0.000	0.000	0.000	0.000
Prv_sc_2km	2.000	2.000	2.000	1.000	1.000	1.000	0.000	1.000	1.000
Cls_prv_hs	1076.170	1177.670	1235.378	3018.548	1384.761	1484.836	2366.247	1933.538	2413.764
Cls_prv_ps	2875.860	2713.178	2770.886	2892.096	2352.719	2637.613	2130.847	2635.149	2292.993

Cs_Fml_Txi	9840.762	9678.080	9735.788	10609.204	10795.039	10895.115	10317.319	11169.774	10004.420
Cls_polc_s	2307.796	2145.114	2202.822	4196.908	3262.074	3362.149	4285.458	3810.851	3972.559
Cl_txi_rt	826.637	781.652	839.360	703.803	801.159	1108.511	161.641	721.564	112.425
Cl_pvt_hpl	3292.096	3503.968	3178.823	1416.620	2545.094	2237.743	1124.735	1977.190	811.836
Cl_Pub_Uni	18330.103	17956.272	17631.127	16635.684	17672.913	17365.561	16252.554	17105.009	16030.900
Cl_Reg_ter	6430.132	6267.449	6325.158	8119.079	6834.374	7141.726	7363.645	7590.427	7514.294
Cl_prv_ter	14084.728	13922.046	13979.754	14821.494	15039.005	15139.081	14529.608	15382.063	14216.709
s_health	1.549	1.549	1.549	1.549	1.549	1.549	1.549	1.549	1.549
s_safety	1.373	1.373	1.373	1.373	1.373	1.373	1.373	1.373	1.373
s_neighbou	1.863	1.863	1.863	1.863	1.863	1.863	1.863	1.863	1.863
s_living_s	2.196	2.196	2.196	2.196	2.196	2.196	2.196	2.196	2.196
s_travel	1.725	1.725	1.725	1.725	1.725	1.725	1.725	1.725	1.725
s_parks	2.686	2.686	2.686	2.686	2.686	2.686	2.686	2.686	2.686
s_librarie	2.922	2.922	2.922	2.922	2.922	2.922	2.922	2.922	2.922
s_drainage	2.333	2.333	2.333	2.333	2.333	2.333	2.333	2.333	2.333
s_pub_heal	2.980	2.980	2.980	2.980	2.980	2.980	2.980	2.980	2.980
s_schools	3.098	3.098	3.098	3.098	3.098	3.098	3.098	3.098	3.098
s_str_lght	2.255	2.255	2.255	2.255	2.255	2.255	2.255	2.255	2.255
s_roads	2.255	2.255	2.255	2.255	2.255	2.255	2.255	2.255	2.255
s_energy	2.137	2.137	2.137	2.137	2.137	2.137	2.137	2.137	2.137
s_waste	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
s_sanitati	1.863	1.863	1.863	1.863	1.863	1.863	1.863	1.863	1.863
s_water	1.902	1.902	1.902	1.902	1.902	1.902	1.902	1.902	1.902
s_dwelling	1.882	1.882	1.882	1.882	1.882	1.882	1.882	1.882	1.882
qol_index	7.131	7.131	7.131	7.131	7.131	7.131	7.131	7.131	7.131
marg_index	1.504	1.504	1.504	1.504	1.504	1.504	1.504	1.504	1.504

Table 12: Explanation of all variables and units

Variable	Full Name	Unit	Source
Avg_SS_val	Average sectional scheme value per square metre	Rand per square metre	CoJ Valuation Roll 2018
Log_SS_Val	Log Average sectional scheme value per square metre	Log Rand per square metre	CoJ Valuation Roll 2019
X	X coordinate of Centroid of 400m Nodal Review Hexagon	Decimal coordinates	Nodal Review 2018

Variable	Full Name	Unit	Source
Y	Y coordinate of Centroid of 400m Nodal Review Hexagon	Decimal Coordinates	Nodal Review 2018
Commrc_2km	Commercial Buildings within 2km	Number of buildings	Nodal Review 2018
F2km_Srv_a	2k service area	square km	Nodal Review 2018
Indust_2km	Industrial buildings within 2km	Number of buildings	Nodal Review 2018
Ind_Com_2k	Industrial and commercial buildings within 2km	Number of buildings	Nodal Review 2018
School_2km	Public schools within 2km	Number of schools	Nodal Review 2018
Clst_Scl_m	Closest public school	m	Nodal Review 2018
ppl_km_11	Population density 2011 per nodal review hexagon	people per square km	Nodal Review 2018
Cl_Pr_s_m	Closest public primary school	m	Nodal Review 2018
Cl_H_Sc_m	Closest public high school	m	Nodal Review 2018
Cl_Tr_S_m	Closest train station	m	Nodal Review 2018
Cls_Clinic	Closest public clinic	m	Nodal Review 2018
Clnc_in_2k	Public clinics within 2km	Number of clinics	Nodal Review 2018
Clst_Hsptl	Closest public hospital	m	Nodal Review 2018
All_P_2k_m	Area of all parks within 2km	square metres	Nodal Review 2018
Clst_dv_pk	Closest developed park	m	Nodal Review 2018
Cls_unD_pk	Closest undeveloped park	m	Nodal Review 2018
Com_in_Hex	Commercial buildings in hexagon	Number of buildings	Nodal Review 2018
Ind_in_hex	Industrial buildings in hexagon	Number of buildings	Nodal Review 2018
Prv_sc_2km	Private schools in 2km	Number of schools	Nodal Review 2018
Cls_prv_hs	Closest private high school	m	Nodal Review 2018
Cls_prv_ps	Closest private primary school	m	Nodal Review 2018
Cs_Fml_Txi	Closest formal taxi rank	m	Nodal Review 2018
Cls_polc_s	Closest police station	m	Nodal Review 2018
Cl_txi_rt	Closest taxi route	m	Nodal Review 2018
Cl_pvt_hpl	Closest private hospital	m	Nodal Review 2018
Cl_Pub_Uni	Closest public university	m	Nodal Review 2018
Cl_Reg_ter	Closest registered tertiary education facility	m	Nodal Review 2018
Cl_prv_ter	Closest private tertiary education facility	m	Nodal Review 2018
s_health	Satisfaction with healthcare services	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018

Variable	Full Name	Unit	Source
s_safety	Satisfaction with safety in your area	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_neighbou	Satisfaction with neighbourhood	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_living_s	Satisfaction with living standards	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_travel	Satisfaction with longest model of travel	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_parks	Satisfaction with parks	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_librarie	Satisfaction with libraries	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_drainage	Satisfaction with drainage	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_pub_heal	Satisfaction with public health services	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_schools	Satisfaction with public schools	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_str_lght	Satisfaction with streetlights	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_roads	Satisfaction with roads	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_energy	Satisfaction with energy sources	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_waste	Satisfaction with waste services	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_sanitati	Satisfaction with Sanitation	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
s_water	Satisfaction with water supply	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018

Variable	Full Name	Unit	Source
s_dwelling	Satisfaction with dwelling	Likert scale, 1 very satisfied, 5 very dissatisfied	QoL 2018
qol_index	Quality of Life Index	Index, 0-5. High good quality of life, low poor quality of life	QoL 2018
marg_index	QoL Marginalisation Index	Index, high number is high marginalisation, low number is low marginalisation	QoL 2018

Table 13: Test of Cronbach's Alpha when Single Variables are Deleted

Cronbach Coefficient Alpha with Deleted Variable				
	Raw Variables		Standardized Variables	
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha
Q2.01_dwelling	0.546641	0.836345	0.567194	0.842180
Q2.03_water	0.427662	0.843374	0.451546	0.847704
Q2.04_sanitation	0.569090	0.836354	0.586854	0.841227
Q2.05_waste	0.553890	0.837571	0.571395	0.841977
Q2.06_energy	0.516083	0.838887	0.534530	0.843754
Q2.09_roads	0.550048	0.835884	0.539962	0.843493
Q2.11_street_lights	0.596217	0.832664	0.589979	0.841076
Q2.12_schools	0.281126	0.849614	0.267999	0.856200
Q2.13_Health	0.418989	0.842831	0.399218	0.850160
Q2.14_Drainage	0.572640	0.834261	0.559406	0.842556
Q2.15_Libraries	0.466245	0.841740	0.439606	0.848267
Q2.19_Parks	0.554472	0.835432	0.531747	0.843887
Q5.09_longest_travel	0.276319	0.847971	0.284986	0.855427
Q10.06_living	0.421504	0.842276	0.434105	0.848525
Q10.09_neighbourhood	0.521921	0.838363	0.538207	0.843577
Q12.05_home	0.284665	0.847876	0.290799	0.855162
Q14.03_health_services	0.360038	0.845295	0.366683	0.851673

Table 14: Simplified correlation matrix showing only rows and columns that contain correlations greater than 0.8, with all values converted to absolute values

	s_living_st	s_health	marg_index	s_dwelling	s_neighbourhood	Indust_2km	s_roads	Commrc_2km	s_drainage	s_sanitation	s_waste	s_water	s_energy	s_str_light
qol_index	0.9168	0.884	0.938	0.9014	0.895	0.3468	0.81	0.2976	0.7849	0.8364	0.821	0.81	0.776	0.81
s_living_st	-	0.87	0.8994	0.8768	0.8626	0.3455	0.82	0.2707	0.7537	0.8245	0.805	0.77	0.753	0.727
s_health		-	0.8862	0.8683	0.8862	0.2861	0.8	0.2331	0.7604	0.8117	0.805	0.78	0.737	0.786
Ind_Com_2k			0.3313	0.3057	0.3709	0.8127	0.15	0.9062	0.1276	0.205	0.188	0.17	0.132	0.124
marg_index			-	0.9316	0.9142	0.3294	0.83	0.2566	0.808	0.8886	0.856	0.85	0.844	0.852
s_dwelling				-	0.897	0.3382	0.86	0.212	0.8248	0.9052	0.892	0.86	0.864	0.871
Prv_sc_2km					0.2248	0.3408	0.09	0.8269	0.0345	0.1058	0.1	0.09	0.051	0.049
s_roads							-	0.081	0.8926	0.8808	0.844	0.82	0.883	0.825
s_drainage									-	0.8833	0.826	0.8	0.836	0.861
s_sanitation										-	0.917	0.89	0.909	0.893
s_waste											-	0.92	0.913	0.885
s_water												-	0.897	0.868
s_energy													-	0.863

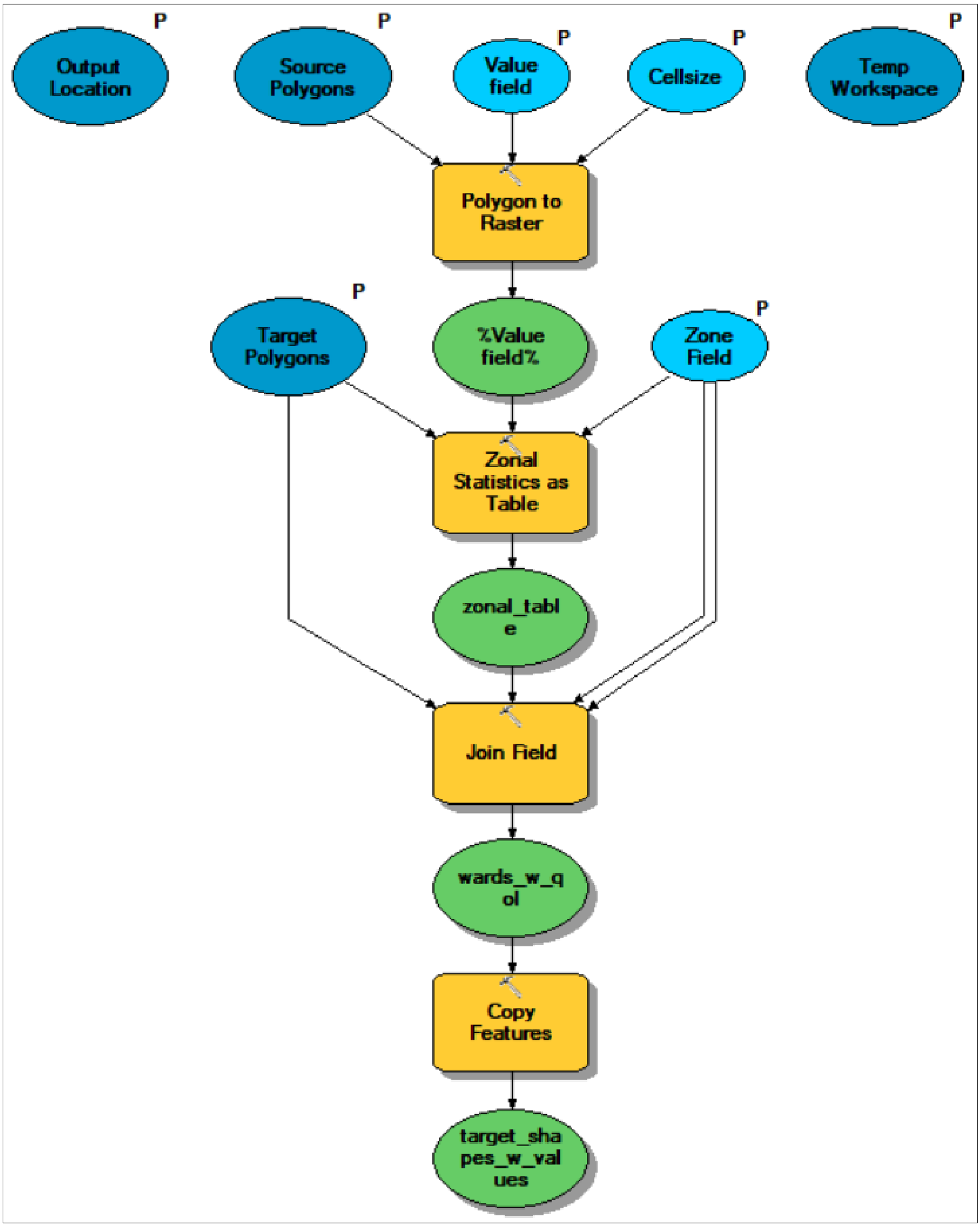

Figure 23: ArcMap Spatial Processing Model developed by Author to aggregate values from smaller to larger polygons (e.g., from Nodal Review Hexagons to Wards)

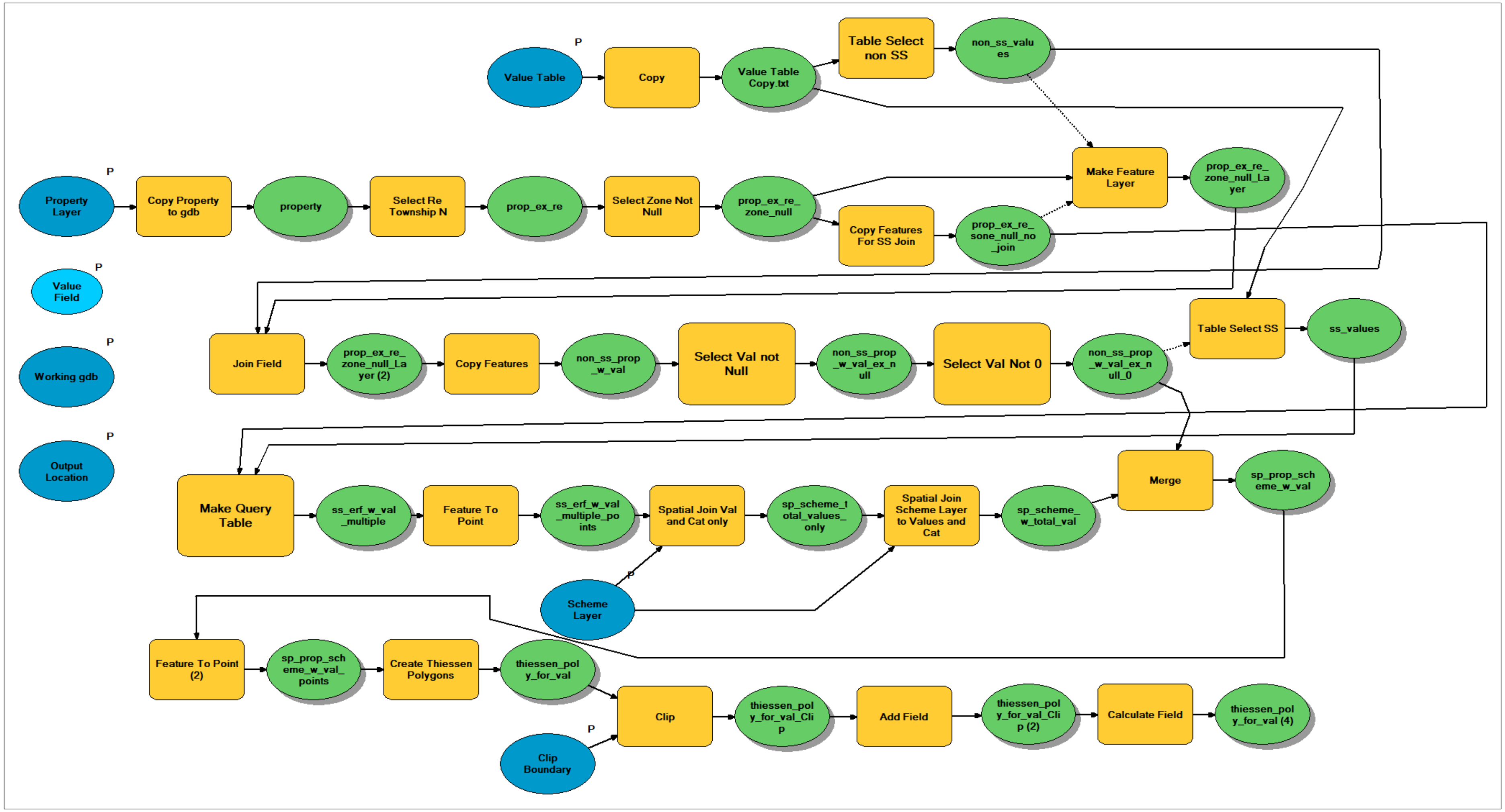

Figure 24: Geo-processing model for creating Thiessen Polygon Layer of Average Value Per Square Metre (Valuation Roll 2018), across the CoJ

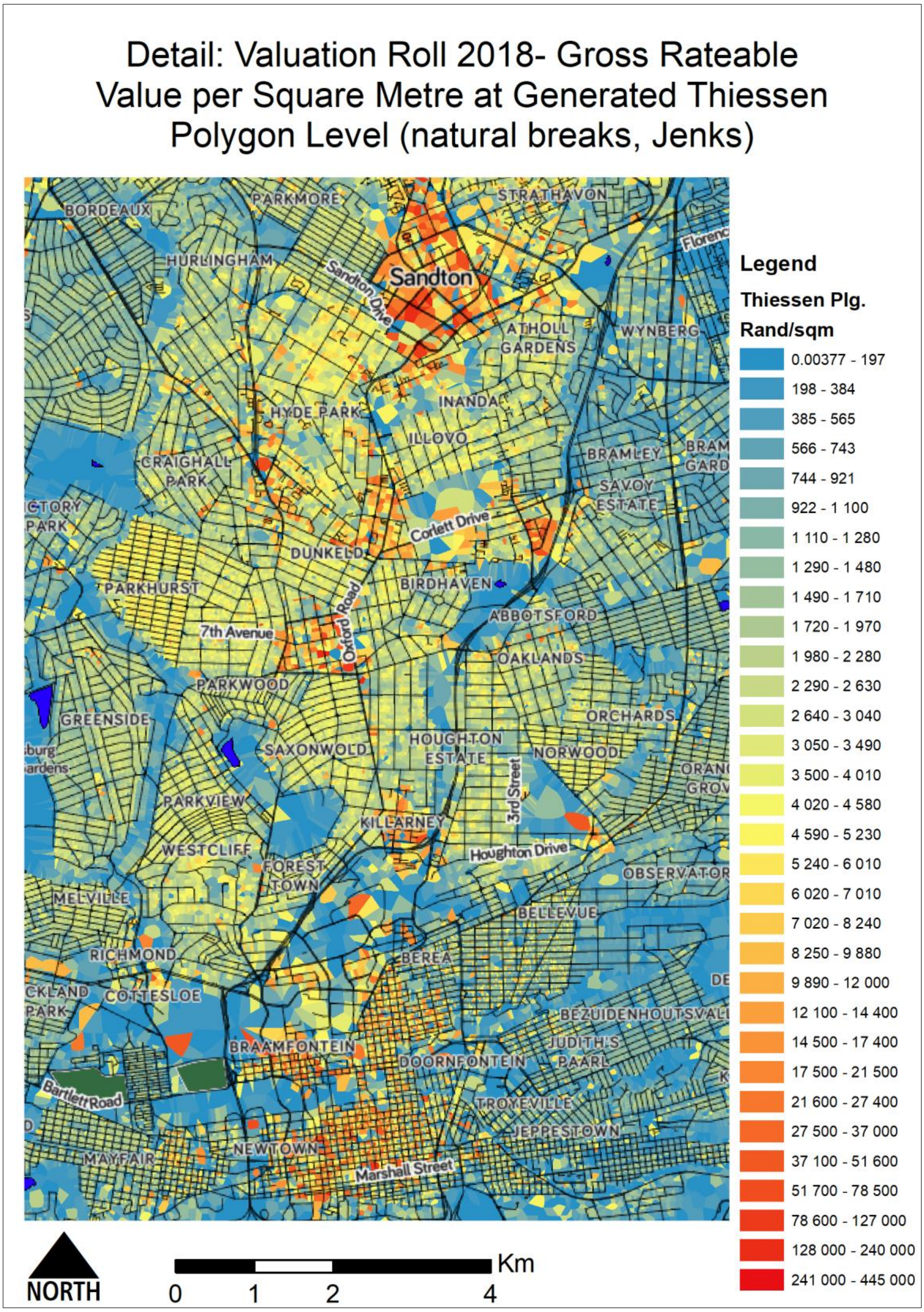

Figure 25: Detail: Valuation Roll 2018- Gross Rateable Value per Square Metre at Generated Thiessen Polygon Level (Map by Author)

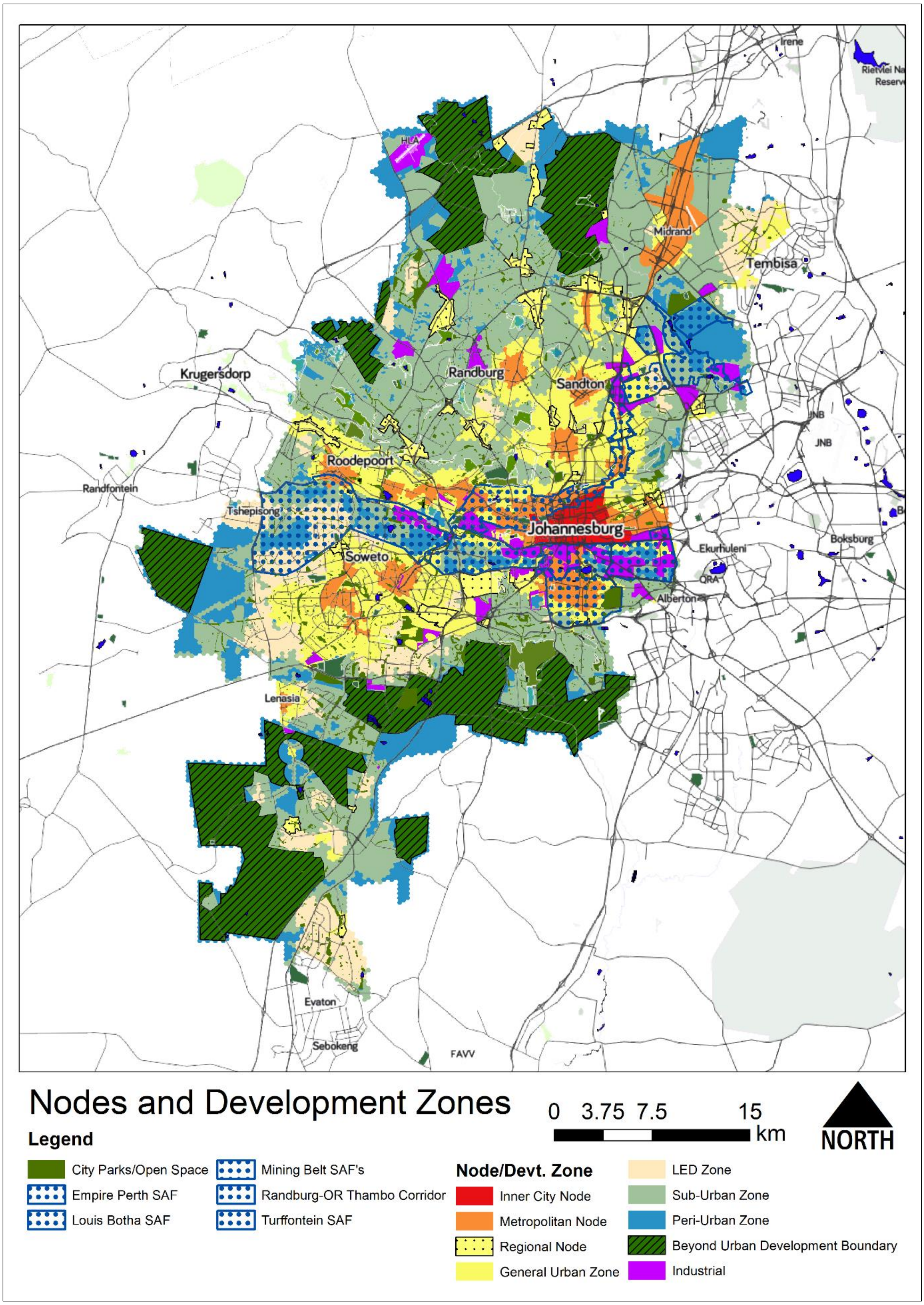

Figure 26: Nodes and Development Zones from the CoJ Nodal Review (City of Johannesburg, 2020, p. 20)