Bachelor Thesis Geography, Planning and Environment (GPE), Nijmegen School of Management, Radboud University Nijmegen, July 2023

Why housing prices not only matter for your wallet:

A study on the effect of housing prices on the ethnic composition of neighbourhoods in the city of Amsterdam.

by

Jesse Luimes

Bachelor's Thesis GPE

Why housing prices not only matter for your wallet:

A study on the effect of housing prices on the ethnic composition of neighbourhoods in the city of Amsterdam.

Student: Jesse Luimes

Student Number: S1060683

Supervisor: Ainul Fajri

Second Reader: Adam Calo

Date: 27-07-2023

Wordcount: 19535

Preface

Dear reader,

I hereby present to you my bachelor's thesis: "Why housing prices not only matter for your wallet: a study on the effect of housing prices on the ethnic composition in neighbourhoods". This thesis is written as final assignment, before finalising my bachelor's degree in *Geography, Planning and Environment* at Radboud University Nijmegen. The research and writing of this thesis took place from February to July 2023. With this thesis I aimed to expand my skills in data analysis and quantitative methods. This surely happened, since the topic involved techniques, I had little experience in beforehand. Furthermore, I have learnt how to deal better with the struggles within the process of writing a thesis. Hence, I would consider this thesis in general a valuable learning experience and I therefore can present this thesis with due pride. This thesis is written for everyone who is interested in the subject and wants to know more about the topic of ethnic composition and housing prices.

The work of this thesis does not stand by itself. I definitely owe thanks to the following people:

- Ainul Fajri MA. – for her incredible patience and for providing feedback, help and guidance throughout the writing process of this thesis.
- Prof. Dr. Huub Ploegmakers, Prof. Dr. Natasha Wagner, Prof. Dr. Pascal Beckers and Jarno Wevers – for helping out with methodological choices regarding data.
- Paul Meijer for providing excellent feedback regarding the use of English.
- My family, especially my parents, for constantly motivating me to keep going on and help me out with the stress this thesis gave me occasionally.

And finally, I want to thank my Lord and Saviour Jesus Christ, who gave me the strength and concentration to finish this thesis.

I hope you have a good time reading this thesis and that the topic is presented in such a way that it is easy to read, understandable and interesting.

Jesse Luimes

Aalten, July 27th, 2023

Brief Summary

The research objective of this thesis is to describe the effect of the housing prices on the ethnic composition in neighbourhoods among other influential factors. The definition of ethnic composition in this thesis is the following: 'the percentage of people from a non-western migration background (compared to western backgrounds). To determine other factors, a closer look has been taken to the concept of neighbourhood choice. This is defined as 'the amount of people choosing to live in a particular neighbourhood'. Ethnic composition is (except for birth/death rates) mainly influenced by neighbourhood choice (including the move to a particular neighbourhood), as the choice for a new neighbourhood would influence the percentage of people with a non-western migration background (Zwiers et al., 2018). Hence, factors which influence neighbourhood choice influence the ethnic composition as well. This way, neighbourhood choice is used as a lens to determine other potential factors. The other factors were determined from the literature used, and it appeared that next to the housing prices, other factors were income, location, and homogeneity of neighbourhoods. In chapter 2.1 further elaboration on these factors will be given.

In chapter 2.2, background information about relevant relationships between those concepts will be provided. This will be information with regard to migrants. The definitions of people with a migration background and migrant are closely related (Centraal Bureau voor de Statistiek , 2023); therefore this will also apply to people with a migration background. The relationships between housing prices and migrants, migrant's settlement choices and homogeneity and the relationship between income and migration background will be discussed. This background information is crucial to understand the conceptual model, statistical analysis and the discussion chapter. One of the important topics in this section, is the concept of native flight, which means that once many migrants are entering neighbourhoods, the native people who were living there start to move out. They see migrants living in their neighbourhoods as negative amenity.

After the concepts and their relevant mutual relationships, the ethnicity theory is introduced in chapter 2.3. This ethnicity theory consists of three main streams or approaches, being: primordialism, instrumentalism and constructionism. All these streams provide their own way of explaining how ethnicity comes into existence and how ethnic conflicts are formed. Primordialism states that ethnicity is a fixed entity people are born with and it is therefore not changeable. This has as a consequence that ethnic conflict between groups is inevitable. Instrumentalism states that ethnicity is neither fixed, nor acquired by birth (Adlparvar & Tadros, 2016). According to Barth (1998 as cited in Adlparvar and Tadros, 2016, p. 125) "individuals selectively emphasise those forms of cultural differentiation that are important to them". Ethnic conflicts would form because of the politicisation of those differences in groups. Ethnic conflicts are not there because of the ethnic differences, but rather because ethnicity is being used as a tool to achieve higher goals. The third constructionist stream states that ethnicity is socially constructed by the people. Every person is contributing to this social construct by their social interactions. There are three main groups which socially construct ethnicity: individuals, discursive formations (e.g., speeches) and structural forces (e.g., modernisation) according to Adlparvar and Tadros (2016).

After the elaboration of the theory, all the approaches are applied to ethnicity and ethnic conflict in the neighbourhood. Every approach results in a different perspective, varying from inevitable ethnic conflict in heterogeneous neighbourhoods (primordialism), to the formation of ethnic conflict because of politicisation of differences between people living next to each other (instrumentalism), to ethnic conflict created by ethnic violent accidents in neighbourhoods (constructionism).

In chapter 2.4 a conceptual model is constructed, derived from the literature. This model forms the basis of the data analysis, and it also visualises the factors influencing each other.

In the third chapter preparations for the data analysis are conducted. A quantitative data analysis is used in the form of a multiple linear regression analysis. This will be done on the basis of all 99 neighbourhoods in the city of Amsterdam. The independent variables taken into account are the housing price (measured in WOZ values), location (measured in biking distance to Dam Square in kilometres), income (measured in the average income per person per neighbourhood) and homogeneity of a neighbourhood (measured in percentage of people with a non-western migration background in a neighbourhood in 2021). All four independent variables are influencing the dependent variable ethnic composition (measured in percentage of people with a non-western migration background in a neighbourhood in 2022). It turned out that the homogeneity-variable severely disrupted the result of the multiple regression analysis, and that the location-variable was not significant. Therefore, the homogeneity variable as well as the location variable were tested separately in two simple linear regression models.

This resulted in the following: for every 1% increasement in the housing price, the ethnic composition decreased by 0.452%. This mean that for every 1% increase in housing prices, the percentage of people with a non-western migration background will decrease by 0.452%. Hence, there is a negative relationship between the housing price and the ethnic composition. However, the influence of the housing price on the ethnic composition falls into nothingness compared to the influence of homogeneity on the ethnic composition. The wish of homogeneity of people has a significantly larger effect than the housing price on the ethnic composition. The influence of the homogeneity variable was that for every 1% increase in homogeneity, the ethnic composition would increase 0.993%. The effect of the average income in a neighbourhood is more than 1.5 times the effect of the housing price, when comparing with the standardised coefficients. Per 1% increase in the average income of a neighbourhood, the percentage of people with a migration background decreases by 0.764%. The influence of the location of a neighbourhood is as follows: when the biking distance to the city centre increases by 1%, the percentage of people with a non-western migration background decreases by 0.377%. From this it can be concluded that the effect of housing price is relatively small compared to the factors homogeneity, average income and location. Of all four variables, it is the third variable when it comes to influence on the ethnic composition.

An important element of the discussion session is the combination of the concept of native flight and the results of the analysis. This means that, in case of this thesis, people with a non-western migration background enter a cheaper neighbourhood (Centraal Bureau voor de Statistiek et al., 2014), since they generally have lower income (Koot et al., 2019) and the native people start to move out (Kalantaryan & Alessandrini, 2020), a downward spiral might be started. An increasing amount of people with a non-western migration background will move in (homogeneity), because of the lower housing prices (found in the data analysis) and resulting in a drop in housing prices as well, which results in even more people with a non-western migration background moving in. This all starts with the fact that people with a non-western migration background generally have a lower income (Koot et al., 2019). Hence, they look for cheaper neighbourhoods (Centraal Bureau voor de Statistiek et al., 2014), which will eventually become even cheaper as described above. This will potentially make neighbourhoods increasingly homogeneous. This was shown in the data-analysis; lower housing prices lead to a higher percentage of people with a non-western migration background.

Table of Contents

1. Introduction

1.1 Motive

Recently, the Dutch housing market has not performed well. Since 2014, prices have risen to an all-time high in Q2-2022 (NU.nl, 2023). The most important cause for this is a shortage in housing (Verwaaij, 2023). This puts many home seeking families in a difficult position, as they cannot find a home. This applies to both buyers and renters. However, current homeowners are in a favourable situation. They have a house, and they are enjoying financial benefits compared with renters (as they pay of their house and build capital). They also might be in the situation where they can possibly rent out their home. A common problem homeowners could potentially have, is that they cannot move, because they will not be able to find another home (Centraal Bureau voor de Statistiek, 2022). Therefore, people who are looking for a new home need to reduce or lessen their demands and are forced to review them. This also applies for housing prices and the choice of neighbourhood.

This raises the question as to which demands people have for their new neighbourhood and how these changes in demand affect neighbourhoods? Are people for example willing to live in a mixed neighbourhood or deprived area, if they cannot find a house elsewhere? Eventually, as people need reduce their demands because of rising housing prices, neighbourhoods will change (RTL Nieuws, 2021). There are clear examples in the Netherlands of deprived areas which are now developing due to gentrification, like the neighbourhood *Bos en Lommer* in the city of Amsterdam, because wealthier people start to move in (Hansen, 2021). This results in rising housing prices (on top of the already high housing prices due to a shortage in housing). Moreover, research by Hwang and Ding (2020, as cited in Feder, 2020) has shown that “predominately black neighbourhoods minorities are the most affected by gentrification” because these minorities are pushed out of an increasingly expensive neighbourhood. According to Hammel (2009) gentrification is the process of the displacement of working-class residents for wealthier professionals, which can lead to neighbourhood development. This means that wealthier people move into poorer neighbourhoods, which constantly occurs in the city of Amsterdam, due to the housing shortage and the poor being pushed out (NUL20, 2012). Although the study by Hwang and Ding (2020, as cited in Feder, 2020) was conducted in the U.S., it raised questions as to how the rising housing prices would affect ethnic groups in the Netherlands and particularly in the city of Amsterdam. This resulted in the following question: What is the effect of housing prices on the ethnic composition of neighbourhoods? Furthermore, the question can be asked which other influential factors on the ethnic composition of neighbourhoods could be there.

The way ethnic composition is expressed in this thesis is by taking the percentage of people with a non-western migration background in neighbourhoods (compared to western backgrounds). An explanation for this choice will be provided in chapter 2.1.2. Nevertheless, this is where the crux of this research is; seeing how big the effect of housing prices, among other factors, on the ethnic composition of the neighbourhood is. This will be done by a case study on the city of Amsterdam and led to the following research questions:

1.2 Research Questions

Main research question:

<u>What are the effects of housing prices, among other influential factors, on the neighbourhood's ethnic composition in the city of Amsterdam?</u>

Sub research questions:

What are other influential factors on the ethnic composition of a neighbourhood?

What is the effect of the other influential factors on ethnic composition?

1.3 Research Objective

The objectives of this study are to describe the effect of housing prices on the neighbourhood's ethnic composition and to determine the other influential factors on the neighbourhood's ethnic composition. To achieve these objectives, mainly quantitative research methodology, specifically the multiple regression analysis method—with help of literature—is used to answer the research questions in chapter 1.2.

1.4 Relevance

1.4.1 Societal Relevance

This research is societally relevant, because the ethnic composition of a neighbourhood is important to the people who live there. The ethnic composition can, for example, influence the housing price (Kalantaryan & Alessandrini, 2020). Furthermore, if a neighbourhood is ethnically mixed, it will partly characterise how people in this neighbourhood live and get along with each other and define the social cohesion within neighbourhoods (Glas et al., 2019).

Another important matter is, that after this research is completed, it will be known how big the housing price effect is, with regard to other factors on ethnic composition. This allows policymakers to project more accurately what will happen to neighbourhoods if prices increase or decrease and what will happen to the ethnic composition. The Dutch government and municipalities are very much occupied with neighbourhoods and their composition and conditions. There is a long tradition of (regional) governments trying to achieve mixed neighbourhoods and improve vulnerable neighbourhoods (Van der Velden & Can, 2022). It will become easier to steer on these urban developments within neighbourhoods (e.g., an increase of housing prices resulting in a loss of affordable housing or minorities being pushed out due to gentrification (Hwang & Ding, 2020, as cited in Feder, 2020)) and to eventually achieve better situation for the people living in the city of Amsterdam. The fact that inhabitants of Amsterdam can benefit from the improved situation can positively contribute to the image of Amsterdam. Through this, new employees can be attracted (which is beneficial for companies), and it becomes more attractive for developers to invest in the city (Raszkowski, 2012).

Furthermore, this thesis shows the importance of homogeneity with regard to ethnic composition of neighbourhoods. It provides additional proof to the fact that people look for homogeneity when choosing a new neighbourhood (Hedman et al., 2011). This extra information can be used by, for example, (regional) governments to respond to the situation created by the fact that people desire homogeneity. For instance, the mixed neighbourhood thought of the Dutch government and municipalities disappeared from the political agenda ten years ago but has returned, because of concerns about growing differences between neighbourhoods (Van der Velden & Can, 2022). This shows that Dutch (regional) governments are concerned with homogeneous and mixed neighbourhoods. As this mixed neighbourhood thought is coming back, it is vital to think about the consequences of these mixed neighbourhoods in relation to ethnic conflicts. The thoughts provided in this thesis in relation to this matter can be used to deal with or prevent ethnic conflicts within neighbourhoods (depending on the approach from the ethnicity theory), which is potentially beneficial for neighbourhoods.

1.4.2 Scientific Relevance

Although research has been done on some parts of this topic—particularly on “neighbourhood choice”, which affects ethnic composition—this thesis concerns the relationship between housing prices and ethnic composition, which has rarely been quantified. Earlier research was more focused on topics like neighbourhood choice, homogeneity or gentrification. An example is research from Sweden, which tried to find a link between the extended family and neighbourhood choice (Hedman, 2012). Furthermore, there is a Dutch study, by Van Ham (2017) about personal and neighbourhood characteristics.

Other research has been done in Sweden on neighbourhood choice by Hedman et al. (2011). In this research, the researchers mainly looked at the big influential factors on neighbourhood choice. This might seem similar this thesis, but there is no focus on housing prices. Moreover, this research is old, with data until 2006 (if there were a housing problem, it was not as serious as it is now) nor was it conducted in the Netherlands.

Also, a bachelor's thesis with the research question whether housing prices influenced ethnic composition has been written by Maltchev (2021). The focus of this research was on low socio-economic status, gentrification and forced movement (Maltchev, 2021), which again is different from this thesis. This thesis aims to focus more on the housing prices itself and other factors of ethnic composition, like homogeneity, income and location.

As stated above, the factors which influence neighbourhood choice are known. The fact that there is an effect from gentrification on housing prices and ethnic compositions is well known. However, what rarely has been studied is the particular effect of housing prices on the ethnic composition. Knowledge about this is lacking, as is the quantified effect of the other factors which influence neighbourhood choice. Therefore, it can be said that this research is scientifically relevant; there are barely examples with relation to ethnic composition which are relevant today. Hence, there is empirical evidence which needs to be filled. Also, there is not a large number of Dutch research available where the effect of housing prices is studied with help of a data analysis and is compared to other influential factors on ethnic composition.

Another important aspect of this research is the application of the ethnicity theory. The ethnicity theory has rarely been applied as a lens to understand ethnic composition. Therefore, this research attempts to bridge the gap between this existing theory and the concept of ethnic composition. This research also attempts to link homogeneity to ethnic conflict with help of the ethnicity theory.

The scope of this research is different from research in which researchers study the effects of people with a non-western migration background on housing prices. The common question in those types of research is often more along the lines of how much the number of people with a non-western migration background (and thus ethnic composition) needs to change for the housing price to change. Research like this has been conducted by Kalantaryan and Alessandrini (2020). In this thesis, the opposite effect is measured: "How much do housing prices need to change in order to change the percentage of people with a non-western migration background in a neighbourhood?" The current issue is that when we know the influence of people with a non-western migration background on housing prices, the influence of housing prices on the percentage of people with a migration background is not necessarily known.

The aim of this research is to obtain knowledge on what the influence of housing prices is on the ethnic composition of a neighbourhood compared to other influential factors to fill this research gap in the Netherlands. This will provide valuable information for researchers who want to further research the topics of ethnic composition in neighbourhoods in the Netherlands and the determining factors behind this. Although not being the main subject, it will also provide valuable information with regard to neighbourhood choice research; on what the effect of the factors (behind a neighbourhood choice) on the ethnic composition is.

2. Theoretical Framework

In this chapter, concepts used in the thesis will be discussed. These are: ethnic composition, neighbourhood choice, residential mobility, housing prices, income, location, and homogeneity. Furthermore, extra information on various relationships between concepts will be provided and theories will be applied to concepts. The goal of this theoretical framework is to set out how much of the topic has already been explored and what is still open to be studied. Also, theory is set out and applied on the concept of ethnic composition. At the end of this chapter, a conceptual model is proposed, which is derived from the theory and concepts discussed in this chapter. In this conceptual model, the mutual (and potential contradictory) relationships between variables will be visualized and a clear image will be created, which will be used in the statistical analysis which follows after this chapter.

2.1 Concepts

2.1.1 Housing prices

A small part of this theoretical framework will be dedicated to the current housing situation and specifically the housing prices. In the introduction, it has already been set out that housing prices are rising, due to housing shortage and certain other developments within neighbourhoods (gentrification). Ultimately, this research tries to connect this situation to the ethnic composition of neighbourhoods. Therefore, a concept of housing prices needs to be introduced.

The concept of housing prices in this research will be presented in the form of WOZ values *(waardering omroerende zaken)*. This translates to 'valuation surrounding real estate' (usually house + surrounding properties). This value is calculated based on the recent average sales prices of surrounding comparable properties in the neighbourhood and is used as the basis for tax calculations. This is different from the actual market value, for which a property needs to be appraised. Hence, the WOZ value correlates with the actual market value (Van Overbeek Makelaars, 2022). Furthermore, most information on housing values which can be found in public records is expressed in WOZ values.

The housing price is an important factor in itself for neighbourhood choice, as is shown in research of the National Association of Realtors (2011). Table 1 shows what the most important factors are when it comes to neighbourhood choice (of buyers in the US). Affordability of homes is the third-most important factor in Table 1, which is linked to housing price. Affordability in this case has to do with how easy or not it is for a potential buyer to purchase a certain property. One should see affordability as the inexpensiveness of an object (Oxford Learner's Dictionaries, 2023a), and it is therefore linked to housing price.

Another research on China shows that the housing price plays a role in internal migration (within countries/regions) and the choice where to live, since it influences the costs of the move significantly (Zhou & Hui, 2022).

Housing prices are not only among the most influential factors in the US and China, but also in a country like Nigeria, as shown by a study by Princewill et al. (2017). According to Princewill et al. (2017) affordability is a pull-factor for a neighbourhood (people get attracted towards a neighbourhood).

		ADULT COMPOSITION OF HOUSEHOLD					CHILDREN IN HOME	
	All Buyers	Married couple	Single female	Single male	Unmarried couple	Other	Children under 18 in home	No children in home
Quality of the neighborhood	64%	67%	62%	57%	64%	58%	67%	63%
Convenient to job	49	47	51	49	58	41	49	49
Overall affordability of homes	44	42	47	49	47	46	43	45
Convenient to friends/family	39	36	48	37	38	36	34	41
Quality of the school district	25	33	15	11	24	18	48	14
Convenient to shopping	25	25	25	24	25	22	20	27
Design of neighborhood	24	27	20	21	22	22	23	25
Convenient to schools	19	24	12	9	17	16	43	7
Convenient to entertainment/ leisure activities	19	18	20	22	21	18	13	22
Convenient to parks/recre-ational facilities	17	18	14	17	15	11	19	15
Availability of larger lots or acreage	15	18	6	10	17	12	19	13
Convenient to health facilities	9	9	10	8	5	8	5	11
Convenient to public transportation	7	6	8	10	7	12	5	8
Home in a planned community	6	7	6	5	3	4	6	7
Convenient to airport	6	6	6	7	4	7	3	6
Green (environmentally friendly) community features	5	5	4	6	3	6	5	5
Other	5	5	6	6	4	13	5	6

Table 1: *Factors influencing neighbourhood choice* (National Association of Realtors, 2011).

2.1.2 Ethnic composition

While doing research about the effects of various factors like housing prices, homogeneity, location and average income on ethnic composition, it is good to first define what is meant by ethnic composition. The term 'ethnic' comes from the term ethnicity. Isajiw (1974, p. 120 as cited in Bilge et al., 2021) defines it as "the socialisation process by which individuals in involuntary groups inherit and share common culture in their groups". Thus, ethnicity is related to culture. The aforementioned definition of Isajiw (1974) implies that one needs to be part of a group. The decision has been made to group people based on their migration background. This was done to still be able to 'organise' people in groups in which they were born (involuntary groups) and of which they inherited a culture

(Isajiw, 1974, p. 120 as cited in Bilge et al., 2021). The grouping of people suits the quantitative topic of the thesis well, as this is a way to make ethnic composition measurable.

The main concept 'ethnic composition' is defined and quantified as the percentage of people with a non-western migration background (Centraal Bureau voor de Statistiek, 2023). One should see this percentage as compared to the percentage of western backgrounds (including native Dutch people in case of Amsterdam), thus more like a ratio. Both these groups added up, would be the *total* ethnic composition. In the thesis, ethnic composition can be referred to as 'people with a migration background' as these expressions are used interchangeably, since one is the definition of the other.

People with a non-western migration background are people of which at least one parent is born in the continents Africa, Asia or Latin America. People with a Japanese or Indonesian background are excluded from this number, because of their social-economic and social-cultural position, since many of these people are born in the Dutch East Indies or employees of Japanese companies with their families (Centraal Bureau voor de Statistiek, 2023). The definition of people with a non-western migration background implies that those people are migrants themselves, or the children of first-generation migrants (which makes them second-generation migrants). The Dutch government banned the word *allochtoon* (translates to migrant), because it had a negative connotation; the word was substituted for the Dutch translation of 'people with a migration background' (De Ree, 2016). Hence, this shows how close the two expressions (migrant and people with a migration background) are related. In this thesis, it will be specifically people with a non-western migration background. However, in chapter 2.2, background information about the relationships between migrants and various factors will be discussed, more information about this choice will follow in chapter 2.2.

More on ethnicity will be discussed in the theory section of this chapter. There, it is also discussed how ethnicity comes into existence, according to ethnicity theory.

2.1.3 Neighbourhood Choice

A concept which is closely related to ethnic composition, is neighbourhood choice (Zwiers et al., 2018). According to Zwiers et al. (2018, p. 96): "A small body of research highlights a second driver and has argued that ethnic neighbourhood change is the result of both residential mobility and demographic change". In other words: the ethnic composition changes due to either people choosing to move to another neighbourhood (neighbourhood choice), or people being born in a neighbourhood or passing away in a neighbourhood. In this thesis, for the reason of model simplicity, only the concept of neighbourhood choice will be maintained. If people *choose* to live in a neighbourhood, that already implies residential mobility, because one *moves* into that neighbourhood (Hedman et al., 2011). Therefore, choosing a new neighbourhood is connected to ethnic composition.

The following definition of a neighbourhood is given by Mike Jenks and Nicola Dempsey: "'Neighbourhood' is defined as both a district—a physical construct, describing the area in which people live, and a community—a social construct, describing the people who live there " (Briggs, 1997, 208; Galster, 2001 as cited in Jenks & Dempsey 2007, p. 173). If this definition is applied to Amsterdam, it could cause confusion. Amsterdam is divided into seven *stadsdelen* (districts), which are subdivided into *wijken* (neighbourhoods), which are subdivided into *buurten* (translated to neighbourhoods as well). A *buurt* (neighbourhood) is usually one or several streets whereas a *wijk* (neighbourhood) consists of multiple *buurten*. However, the problem is that both *wijk* as well as *buurt* are translated to neighbourhood. For the subject and the data analysis, the *wijk*-definition will be retained. Datasets with data regarding *buurten*, which would result in over 400 observations, had many missing values, which would disturb the analysis. In the dataset used for the data analysis, there are 99 of those *wijken* (neighbourhoods). .

Secondly, a brief definition of choice will be provided: "An act of choosing between two or more possibilities; something that you can choose" (Oxford Learner's Dictionaries, 2023b). Therefore, in this thesis, neighbourhood choice is defined as the amount of people that choose to live in a certain neighbourhood. This implies that they move there. The choice to *not* move to a certain neighbourhood is not taken into account, because this does not affect the ethnic composition of that particular neighbourhood (Hedman et al., 2011). Thus, neighbourhood choice in this thesis only includes the people who actually went to move to and live in the neighbourhood of their choice.

This is connected to ethnic composition: If one *decides* to live in or move to a certain neighbourhood, the demography of a neighbourhood changes, which is a big driver of ethnic neighbourhood change (i.e., ethnic composition) according to Zwiers et al. (2018).

Therefore, it is vital to understand what influences this concept of neighbourhood choice as the exact same things will also influence ethnic composition (Zwiers et al., 2018). Thus, based on the neighbourhood choice concept, we try to understand what affects ethnic composition, since there is little information available on changes in the ethnic composition in itself. Other concepts are all connected to neighbourhood choice and to ethnic composition.

2.1.4 Income

Income is described as influential on neighbourhood choice (Hedman et al., 2011). This research of Hedman et al. (2011) is based on a study in Uppsala, Sweden and is one of the larger research projects done on neighbourhood choice. In the results is visible that income variables explain more than 12% of the variance in neighbourhood choice, which is more than any other variable (Hedman et al., 2011). Hence, it can we said that income is an important driver of neighbourhood choice. With less income, a family tends to buy a cheaper house, usually in a cheaper neighbourhood (Centraal Bureau voor de Statistiek et al., 2014). Hence, income plays a role in choosing the neighbourhood.

That income does not only play a role in Sweden, shows research from Nigeria by Princewill et al. (2017). Income and affordability (I.e., lower housing prices) are among the most important factors when choosing a neighbourhood. However, the most important influence on neighbourhood choice in Nigeria is security (Princewill et al., 2017). Nevertheless, this research still shows that, even in a totally different country, income is important. Income can also form a limitation: a family wants to move to a neighbourhood, for instance, to be closer to friends, but the houses are too expensive in said neighbourhood. Therefore, they cannot afford to choose that neighbourhood to live in, and thus not influence the ethnic composition (Troost et al., 2021). In this way, income is linked to neighbourhood choice and ethnic composition.

2.1.5 Location

Research performed by Chhetri et al. (2006) attempts to make a connection between the attractiveness of a neighbourhood and residential location decision choices. This research was performed in Brisbane, Australia and showed that factors like convenient to shopping areas or schools, close to family or friends and open and green spaces are significantly more influential on residential location decision choices, because these are what makes a neighbourhood attractive.

Other research, mentioned before, done by the National Association of Realtors (2011), gives the following insights in Table 1: The first four factors (quality of neighbourhood, convenient to work place, overall affordability of homes, convenient to friends or family), are standing out compared to the other the factors, meaning that they have a bigger influence on neighbourhood choice and thus are more important. In the table, it is visible that location plays a big role, when it comes to choosing a neighbourhood. With the term ‘convenience to’ is meant that the connection to those places is good (either close by or a quick connection); therefore, it is considered within the concept location in this thesis.

2.1.6 Homogeneity

According to various research, homogeneity is an important concept in relation to neighbourhood choice. These research include the research of Hedman et al. (2011), Ham (2017) and Chhetri et al. (2006), all mentioned before. Homogeneity can be defined as: “The quality or state of being all the same or all of the same kind” (Oxford Learner’s Dictionary, 2023c). When connecting this to the topic, this would mean the following according to Hedman et al. (2011): people want to live in a neighbourhood where people with the same characteristics live. The research of Chhetri et al. (2006) give homogeneity as reason why people can choose a certain neighbourhood. Another research of Hedman (2012) shows that people tend to move to neighbourhoods where their family lives.

Research done by Ham (2017) showed that there is a significant interaction between neighbourhood and personal characteristics". This fact also supports that homogeneity plays a role in looking for a neighbourhood. The advantages of homogeneity will be discussed later in the context chapter.

It works the other way around as well: where homogeneity can be a pull factor, heterogeneity can be a push factor. The research of Zwiers et al. (2018, p.96) talks about "how ethnic heterogeneity stimulates the out-mobility of the native (majority) of population to more white neighbourhoods". According to Zwiers et al., (2018, p. 96) the concept of residential mobility is deeply connected to the ethnic composition and one of the main drivers for *ethnic* neighbourhood change: "The selective moving behaviour of different ethnic groups can affect ethnic neighbourhood change in different ways".

2.2 Relationships between concepts

The purpose of this section is to supply information which makes it easier to understand the topic in general, but also to understand the to be proposed conceptual model, the statistical analysis (and its results) and the discussion at the end of the thesis. Information on people with a migration background in relation to these topics is scarce. Therefore, information on the topics in relation to migrants will be given instead.

Background information about the relationship between housing prices and migrants will be discussed. Furthermore, background information about settlement choices and homogeneity will be addressed, as they form an essential link in a network of factors. Also, the relationship between income and migration background will be discussed.

Since people with a non-western migration background are very close to migrants, by definition; matters which apply to migrants also apply to people with a non-western migration background. As these people with a non-western migration background are either migrants themselves, or one of their parents is a migrant; this would make them second-generation migrants (Centraal Bureau voor de Statistiek , 2023).

About the relationship between housing prices and the percentage of people with a non-western migration background in neighbourhoods:

There has been much research on the relationship between housing prices and migrants. The vast majority of these research have been done on the effect of migrants and migration on the housing prices in the area in which they settle, but rarely in the other way around, as this thesis aims to do. Nevertheless, it is still important to understand this relationship, as this clarifies how those two factors relate to each other.

A study from China, conducted in 2017, found that a one percent increase in migrants in a city in general leads to a 0.701% increase in housing price (Zhang et al. 2017). This would be due to the fact that migrants play a role in economic growth, boosting the local economy (because there is more labour force available) and therefore making that city a more attractive place to live, increasing its housing prices (Birrell and Healy, 2008). The study distinguishes between higher and lower educated migrants, resulting in the observation that higher educated migrants have an even more positive effect on the housing prices. However, there are also research which show that migrants bring exactly the opposite effect to a neighbourhood, because it makes the local labour market more competitive, forcing native people to move out to other areas, resulting in decreasing housing prices (Sa, 2015). This study also mentions the impact of urbanisation on housing prices. An increase of one percent of the urbanisation level results in a 0.343% increase in the housing price. The studies on these topics are conducted from different perspectives, resulting in multiple different explanations for this phenomenon. Firstly, it can be explained by the direct effect urbanisation has on the housing price. There is a demand for housing within a city, therefore prices go up. This leads to another explanation: Because there is a demand due to urbanisation, *people expect* the price to rise further. This expectation would create even more demand, with prices rising even higher than they were already (Luo, 2011, as cited in Zhang et al., 2017).

As is shown, it is highly likely that migration towards a city will cause its housing prices to rise. For the city of Amsterdam, this appears also to be the case. With the rise of housing prices, the number of migrants and expats grew too (Velzel, 2023). For this thesis, it is relevant to acquire knowledge of this topic on a neighbourhood level.

An Italian study finds that on neighbourhood level, the results are mostly the other way around (Kalantaryan & Alessandrini, 2020). According to this research, it is generally the case that the more diverse a neighbourhood is, the lower the housing prices are. However, this relationship is non-linear: the effect is not the same for every percentage of migrants. Up until a percentage of 45%, the relationship between those two variables is positive. If the percentage of migrants is higher than 45%, the relationship turns to a negative relationship. The study gives various explanations as to why this could be. The positive part of the relationship can be explained by the additional demand in housing (hence, the housing prices rise). After many migrants have come to the neighbourhood, and native people living in that neighbourhood consider that as a negative amenity, demand will decrease, and the housing prices

decrease as well. The phenomenon that native people (these people are not necessarily Caucasian) flee when the immigration to a place like a city, rises is called *native flight*. In native flight, there is a certain search for a place with homogeneity by the native people (Kalantaryan & Alessandrini, 2020). People do want to live in a place with similar people and not with totally different people (Hedman et al., 2011), hence if too many different people are moving in 'their' neighbourhood, the native people tend to move away (Stonawski et al., 2019).

The study of Kalantaryan and Alessandrini (2020) also specified which types of migrants had a positive effect on housing prices and which groups had a negative effect. This specification was done because of the earlier discovered non-linear relationship. Migrants from Asia and Eastern Europe were associated with declining housing prices, while migrants from a country like Switzerland were associated with rising housing prices on neighbourhood level.

About the migrants' settlement choices, ethnic enclaves and homogeneity:

Another concept that plays a role is homogeneity. The importance of this concept has already been discussed and the concept has been defined in chapter 2.1.6. Therefore, in this section, more is told about the underlying factors for this homogeneity and its relationship to other factors. This is done on the basis of so-called ethnic enclaves. Why do these so-called ethnic enclaves exist and why do migrants specifically move to these areas, forming essentially homogeneous neighbourhoods? This might not be directly about people with a non-western migration background. But as said earlier, people with a non-western migration background are either migrants themselves or descendants of migrants (Centraal Bureau voor de Statistiek , 2023). Therefore, this information is also applicable to people with a non-western migration background.

Firstly, a definition of an ethnic enclave is given. Although there are many different definitions of an ethnic enclave, the definition of Lim et al. (2017, p. 138) is one of the most comprehensive: "An ethnic enclave is a geographical area where a particular ethnic group is spatially clustered and socially and economically distinct from the majority group". Classic examples of these ethnic enclaves are, for instance, Chinatowns, one can find across many American cities (Brittner, 2022). An ethnic enclave is different from an ethnic neighbourhood in one particular aspect: in ethnic enclaves, there is also an economic aspect (with economic advantages), where a neighbourhood does not necessarily have this aspect (although it could) according to Brittner (2022). In spite of the difference in economic aspect, it is still good to look at the advantages of both, since they are relatively similar structures (Qadeer et al., 2010) and in which homogeneity gives the same advantages. Although many see ethnic enclaves as a negative matter and an integration policy failure, there are clear advantages according to researchers such as Schüller and Chakraborty (2022) and, Martén et al. (2019). Some of these advantages will be discussed in the following section.

Choosing to live in ethnic enclaves has certain advantages over other living areas, particularly with regards to the labour market. An enclave can provide a warm welcome for newly arrived migrants, and it provides the network for quickly finding a new job (Schüller & Chakraborty, 2022). For instance, research conducted by Martén et al. (2019, p. 5) shows that “refugees who live in ethnic clusters are more likely to have full-year employment, and that this effect persists for at least five years after arrival”. Although these are refugees and not voluntary migrants who want to look for a new neighbourhood, it still shows some of the economic benefits for new people in an ethnic enclave or community. However, all of this depends on the quality of the enclave, in which quality is measured in terms of income, education and employment rates. Furthermore, the network within the enclave needs to be important, to benefit from the advantages of the enclave in relation to the labour market. The size of the enclave is of less importance (Schüller & Chakraborty, 2022).

Besides the advantages regarding the labour market, numerous other advantages exist. The ethnic network which was shortly mentioned earlier has potentially a positive impact on integration via the jobs acquired through the network (Martén et al., 2019). According to research done by Qadeer and Kumar (2006, p. 1): “The social benefits outweigh the disadvantages of the predominance of one ethnic group”. Social benefits are potentially the first benefits wherefore people move there. After they have moved there, they might discover other economic benefits. Another big advantage is that there are signs that homogeneity in neighbourhood leads to higher social cohesion within neighbourhoods, which leads to higher life satisfaction. Neighbourhood social cohesion declined as the neighbourhood becomes more heterogeneous (Cheung & Leung, 2011).

About low-income groups and people with a non-western migration background:

In the Netherlands, it is still common for people with a non-western migration background to earn significantly less on a yearly basis. Research done by Koot et al. (2019) shows that in 2019, a household with a non-western migration background earns at least 16% less than a Dutch household. This 16% applies only for households from Suriname. Households from Moroccan descent earn 31% less on average, Turkish households 26% and Antillean households earn 21% less. This differs from generation to generation, later generations with non-western migration background earn more money than their parents' generation, nevertheless they are still falling behind on people from Dutch descent.

This difference in income has various reasons. A large part of this difference in income can be explained by differences in levels of education (especially for women). The numbers from the year 2017 are as follows: The amount of Dutch descent people with a higher education degree is 45% (men) and 54% (women). This can be either a college degree or a university degree. For people with Turkish descent this is respectively 22% and 15%, Moroccan descent 28% and 26%, Surinam descent 35% and 40% and of Antillean descent 35% and 38% have gotten a higher education degree. This explains why later generations of migrants have smaller wage gaps, because the third generation did attend school in the Netherlands. Another reason for the difference in income could be that groups with a non-western descent are more dependent on government benefit programs. Furthermore, the percentage of women with work as the most important source of income is significantly lower under people with a non-western migration background (Koot et al., 2019). Altogether, it is evident that people with a non-western migration background earn significantly less money for various reasons. These differences have consequences, which will be discussed in the next section.

Because of this difference in income, there is also a difference in where these people live. In case a family has less money, they are logically more likely to live in a neighbourhood with, on average, cheaper houses (Centraal Bureau voor de Statistiek et al., 2014). As described earlier, these can be neighbourhoods with a high percentage of people with a non-western migration background, due to native flight. This would imply that homogeneity and housing prices influence each other in both ways. Firstly, homogeneity (for people with a non-western migration background) creates lower housing prices due to native flight (Kalantaryan & Alessandrini, 2020). Secondly, these lower housing prices attract lower income groups, which are often people with a non-western migration background (Koot et al., 2019) and neighbourhoods turn more homogeneous. Then, a neighbourhood would potentially enter a downward spiral, in which housing gets increasingly cheaper. This does not mean that there are no native people with a lower income, but according to concept of homogeneity, they will choose to live somewhere else, as homogeneity is such an important factor.

Research from McMillen and Singell (1992) shows that a wage gradient exists. This is a phenomenon within urban economics which states that wages decrease the further away someone lives from the city, together with the premise that the work location and the location of residence are positively correlated, according to the monocentric city model. This implies that people with a non-western migration background often work and live further from the city centre, since they have lower income, generally speaking (Koot et al., 2019). It also implies that the more expensive parts of the city are closer to the city centre.

The results of the data analysis in chapter 4 will show whether the information above also is the case in Amsterdam and if people with a migration background generally live further from the city centre.

2.3 Theory

After the most important concepts have been introduced and it is shown how they move together in influencing the ethnic composition in a neighbourhood, a theory will be discussed and applied. The theory which will be used is the ethnicity theory, which has not been made by one person but developed through the years because of contributions of various researchers.

2.3.1 Ethnicity theory

The ethnicity theory tries to understand ethnicity from various perspectives. It tries to answer questions on what ethnicity is, how ethnicity is created or how ethnic conflict is formed. This is done from three streams or approaches, being: primordialism, instrumentalism, and constructionism (Adlparvar & Tadros, 2016). Firstly, general information of this theory will be provided, before diving deeper into the three approaches. The differences between those approaches will be discussed, clarified and background information on these approaches will be given.

Over the last 50 years, more and more researchers started to focus on ethnicity, for example Fredick Barth and Thomas Eriksen (Adlparvar & Tadros, 2016). This led to the creation of a network of theory, which eventually developed into multiple 'streams', which are described below. Several smaller streams are excluded because they are not as influential as these three. One of the bigger discussions within the ethnicity debate is how ethnicity comes into existence. Because differences in ethnicities can create ethnic conflicts, these theories are related to topics like groups, nations and states. Also, these three streams (primordialism, instrumentalism and constructionism) are applied on conflict situations and try to explain how ethnic conflict comes into existence (Adlparvar & Tadros, 2016). This fits well to the concept of ethnic composition, because in heterogeneous neighbourhoods, violent conflicts can surface any time (Wo, 2022)

Hence, it should be known first what ethnic conflict exactly is and how it is formed. Every approach has its own perspective on that question. Ethnic conflict in relation to our topic is drawn broader than wars and violence. In this case it also includes minor ethnic violence which can happen on neighbourhood level, which can create tensions (e.g., harassment, small brawls, etc.). Ethnic conflict and tensions might also be a reason for people to leave their neighbourhood (StudySmarter, n.d.). Therefore, this theory is better than other theories (like critical race theory from Derrick Bell (1976, as cited in Delgado & Stefancic, 1998); it provides a good overview of different thoughts about ethnicity and how it is created. In addition to that, it applies this on ethnic conflict. It does this better than other theories: critical race theory talks about transforming and studying the relationship among race, racism, and power according to Delgado and Stefancic (2006). One of the main propositions that critical race theorists agree on would be that racism is ordinary, it is the way society works and it is an everyday experience of black people. Another important proposition is that race is socially constructed and not a product of biology (Delgado & Stefancic, 2006). Delgado and Stefancic are among the more important thinkers within this theory (Alcoff, 2021). It appears that it is not easy to define race; this becomes clear after reading the following passage from an article from Sun (1995, p.44): "It can be argued that the development of the new scientific definition of race and the related research guidelines is implausible without the verification of the new definition and the falsification of other definitions on the basis of the empirical data because it would be dictatorial for any science organization to claim that a concept should be valid". Sun is one of the scientists who tried to define race but failed to produce a definition. Neither could Atkin (2016), in his book *The Oxford Handbook of Philosophy and Race*, in which he tried to find a definition. He does this with ordinary definitions, but he also tries it with existing scientific knowledge, which did not work. According to Zack (2017, p. 147): "if we set about solving the definitional question by finding putative concepts from biological science to ser as scientific race concepts, we do not seem to have any viable candidates". However, there seem to be at least some definitions available. The first definition (Smith et al., 1997, as cited in Templeton 2013, p. 263) is the following: "One definition regards races as geographically circumscribed populations within a species that have sharp boundaries that separate them from the remainder of the species". He also has a second definition: "A second definition defines races as distinct evolutionary lineages within a species" (Templeton, 2013, p. 263). This biological definition is by no means what the definition of race according to the critical race theory is. According to Delgado & Stefancic (2006, p. 3): "Race and races are products of social thought and relations. Not objective, inherent, or fixed, they correspond to no biological or genetic reality; rather, races are categories that society invents, manipulates, or retires when convenient". It becomes clear that there is a debate about race. Various scholars will say it is socially constructed, while other argue it is solely biological. All of these definitions do not really correspond well to the topic of ethnic composition. Especially not the critical race theory part, in which the definition of race is not fixed. The definitions of ethnicity however can also vary, depending on the perspective, as the ethnicity theory argues. A broader definition of ethnicity is given by Isajiw in 1974 (see chapter 2.1.1). Later, in 1993, he argues that ethnicity includes attitudes, values and, preconceptions, but also "has a distinct culture in the sense of a total way of life" (Isajiw, 1993, p.

410). This already fits the topic of ethnicity much better, and therefore, ethnicity theory is chosen over critical race theory.

Primordialism:

Primordialism is generally considered as one of the older approaches within the whole body of theories on ethnicity. According to primordialists, ethnicity/ethnic identities are fixed, they are assigned at birth, and are therefore unchangeable (hence, they are also genetic). This implies the fact that ethnicity is embedded biologically and that everybody is born in a certain ethnic group (as a culturally defined unit) following a study done by Adlparvar and Tadros (2016). According to Adlparvar and Tadros (2016, p.124) “this led to tribes and later ethnic groups being classified by aspects of their material culture in addition to biological and territorial features”. In other words, this led to the fact that ethnic groups were sorted mainly based on the cultural habits and objects they use and secondly based on biological and territorial aspects. And because everyone has an ethnic identity (because every individual has been born with one), “this ethnic identity serves a fundamental human need for belonging and meaning” (Adlparvar & Tadros, 2016, p. 124). Another aspect which results from this statement is that cohesion in ethnic groups is important.

Seeing ethnicities as fixed ‘objects’ bring certain consequences. For example, on the field of ethnic conflict: because ethnic differences are inevitable, ethnic conflicts will arise between groups who hate each other (Adlparvar & Tadros, 2016). Another important aspect which primordialism is right about is that it gives an explanation for the fact why people within ethnic groups get along with each other (Che, 2016).

Where primordialism lacks as an approach is the fact that is generalises everything. There are clear exemptions, which go straight into this theory. These exemptions also applicable at neighbourhood level. In a lot of mixed neighbourhoods in Amsterdam, people live happily together. In other neighbourhoods (like the *Bijlmer*), there are tensions, but not a larger ethnic conflict. The theory dismisses the fact that a lot of the conflict is not because of the differences itself, but because of the political and/or socio-economic structures. It denies the fact that ethnic conflict does not always come from other external conditions (Che, 2016).

Instrumentalism:

The approach of instrumentalism came as reaction to the approach o`f primordialism. It is the opposite approach of primordialism. Instrumentalism states that ethnicities are not fixed (Adlparvar & Tadros, 2016). According to Barth (1998 as cited in Adlparvar & Tadros, 2016, p. 125): “individuals selectively emphasise those forms of cultural differentiation that are important to them”. The ethnic boundaries which are still there (between ethnic groups), are maintained because of 'us' and 'them' thinking between groups (Adlparvar & Tadros, 2016).

This implies one thing: instrumentalism does not assume that conflict comes from the difference in ethnicity itself (because it is not fixed). Rather, it states that conflict comes into existence because of the politicisation of those differences (Che, 2016). Instrumentalism also states that ethnicity does not hold more importance to political issues than ideological beliefs (Reuter, 2023.). This means that neither of the two aspects is a more important cause for conflict. According to instrumentalism, an ethnic conflict arises when two ethnic groups are fighting for the same goal (Reuter, 2023), for example the control over a country. Then ethnicity is being politicised, otherwise it would not play an important role, as said before. According to Reuter (2023), an instrumentalist could see ethnicity as a tool, to unify and organise larger groups, to reach certain goals. Within instrumentalism, there is one particular perspective which clarifies this further: the elite perspective. One of the major contributions on the elite perspective comes from Paul Brass. According to Kataria (2018, p.133): "Brass's work emphasises the role of elites in the formation and persistence of ethnic identity". He argues that "elites and counter-elites within ethnic groups select aspects of the group's culture, attach new value and meaning to them, and use them as symbols to mobilise the group, to defend its interests, and to compete with other groups" (Brass, 1979 as cited in Kataria, 2018, p. 133). In the basis, this goes back to group manipulation.

The downside of instrumentalism is that it does not explain why people within one ethnic group get along with each other better than others, which is something the primordialist-approach does. In the end, this theory seems more nuanced as it is not as one-sided as the primordialist-approach. (Che, 2016)

Constructionism:

The approach of constructionism see ethnicity as a social construct. It is something produced by social interactions between people and groups; every person is contributing to this production. This approach is the 'successor' of the instrumentalist approach. The approach can be sorted into three different groups of literature. This has to do with the question as to who is actually producing the ethnicity as a social construct. There are three entities which are able to produce ethnicity: individuals, discursive formations or broad structural forces (Adlparvar & Tadros, 2016).

1. Individuals who socially construct ethnicity:

This first group looks like an addition to the instrumentalist approach. However, the instrumentalist approach was only focused on the elite (which uses ethnicity as a tool), but this part of the constructivist approach also includes the ordinary people (Adlparvar & Tadros, 2016). The social construction of ethnicity goes as follows according to Fearon and Laitin (2000, para. 2.4 as cited in Adlparvar & Tadros, 2016): “This body of literature sees ethnicity created and recreated through the everyday actions of individuals, who perceiving themselves as associated with a certain ethnic identity act to confirm, contest or propagate that identity”. With ‘this body of literature’ is meant, this part of the constructivist approach. Basically, this says: I am Dutch, so I act Dutch.

When it comes to ethnic conflict, the constructivist approach’s main focus is on how individuals (both normal people and elites) are starting violence (Adlparvar & Tadros, 2016). It tries to answer the question how ordinary people contribute to ethnic conflict (building on the instrumentalist approach, which was discussed earlier). According to Adlparvar & Tadros (2016, p. 126): “It is suggested that marginalised members of ethnic groups contest existing ethnic identities, thereby constructing new ones. This, in turn, can result in retaliatory violence from those elites who benefited from the previous form of ethnic identity”. In other words: The ‘oppressed’ group of people start to fight the ethnic identities (e.g. their beliefs, habits and lifestyle), which results in backfiring from elites who benefitted from that ethnic identity. So, it is in some way a fight of the ordinary people against the elite.

2. Discursive Formations which socially construct ethnicity:

To start off the subject of discursive formations, it would be helpful to define what discursive formations are. The explanation of Adlparvar and Tadros (2016) mentions ‘cultural systems’ as synonym. According to Scott (2001, para 8), Foucault defined discursive formations as follows: “Discursive formations, according to Foucault, are groups of statements which may have any order, correlation, position, or function as determined by this disunity”. This can be, for instance, a group of statements which show any sign of coherence (e.g., a story). The idea with discursive formations is that these result in the construction of ethnic differences. This theory assumes that the cultural systems themselves make the differences and nothing else. This implies that whenever there are cultural differences, this will lead to ethnic differences and that this is inevitable. This leans towards primordialism, because in primordialism, it is argued that ethnic differences are inevitable. This is generally considered as one of the criticisms of seeing discursive formations as an agent of social construction.

With regard to ethnic conflict, discursive formation again leans towards primordialism. According to Adlparvar & Tadros (2016, p.127): “this part of constructionist literature focuses on the capacity of discourse to predispose members of one ethnic group to view members of another as natural targets of violence”. This means that discourse is able to set up two ethnic groups against one another, by giving one group a negative image towards the other group. Besides the critique regarding primordialism, another point of critique is that discursive formation cannot explain different forms of ethnic violence worldwide (Adlparvar & Tadros, 2016).

3. Structural forces which socially construct ethnicity:

The last part is about structural forces, these can be political forces, but also economical. This part makes up for the majority of constructionist ethnicity approach. These forces play the biggest role in socially constructing ethnicity. Because these forces are structural, they involve everyone (Adlparvar & Tadros, 2016). This means that everyone is affected by an ethnic conflict in the end, whether they want to be involved or not. A big subject within this subgroup of constructionism, is ethnogenesis (which is, as the word says, what the origin of ethnicity is).

If it is about conflict, this part of constructionist theory states that the broad structural forces (like modernisation) can lead to ethnic conflict. This can lead to changes in social interactions, such as increased tensions or violent conflicts, and influence the socially constructed nature of ethnicity (Adlparvar & Tadros, 2016). Those increased tensions or violent conflicts can eventually lead to bigger ethnic conflicts, which will, over time, involve more and more people. An example of ethnic conflict caused by a big political structural force can be a leader (of the ethnic group which is in power) who sets for another course in the political landscape of the country, which leads to the marginalisation of one ethnic group. This can lead to increased tension and eventually violent ethnic conflicts (Reuter, 2023).

2.3.2 Application of Ethnicity Theory

All three main streams (primordialism, instrumentalism, constructionism) of this theory will be applied on the topic ethnicity in the neighbourhood and how homogeneity plays a role. These ethnicity theory approaches have all been applied on conflict situations and they all try to give an answer on the same set of questions. These questions are how ethnicity is formed and how ethnic conflict arises from ethnical differences (Adlparvar & Tadros, 2016). With regards to the topic of ethnic composition in neighbourhood, especially the second question is from significant important, because also in neighbourhoods, ethnic conflict can arise. Therefore, the three main streams will be applied on how all of those theories would explain ethnical conflict in neighbourhoods. After the application, a conclusion will be drawn on which of these theories is the closest to reality and which can be used to systematically explain ethnicity better.

Primordialism:

Seeing ethnicity as fixed and unchangeable has its consequences in explaining ethnic conflict. Because this approach states that ethnic conflict is inevitable in a heterogeneous ethnic compound or between groups that hate each other, this will also be the case for neighbourhoods, according to primordialism. When the primordialist approach is applied to neighbourhoods and neighbourhood choice, the conclusion will be that heterogeneous neighbourhoods will not work, because in these neighbourhoods, ethnic conflict is inevitable. This ethnic conflict will happen, and what potentially might happen, is that people start to move out to other neighbourhoods where they feel more comfortable. In this regard, there is a clear link to existing knowledge on native flight, as discussed in chapter 2.2. Native people would start to see this as negative amenity and start to move out to neighbourhoods which fit their own characteristics better (Kalantaryan & Alessandrini, 2020). This neighbourhood will probably be a neighbourhood where already more of their kind of people live, because research has shown that homogeneity appears to be one of the most influential factors in neighbourhood choice (Hedman et al.2011). This potential circle ignites a self-reinforcing process in which the ethnic composition of a neighbourhood will slowly turn more homogeneous.

The point with this application of primordialism is that it assumes things, which are not necessarily true. Heterogeneous ethnic compositions do not always lead to ethnic conflict—this also applies to neighbourhoods (Adlparvar & Tadros, 2016). However, it is because of this false assumption that the conclusion has been drawn that homogeneous neighbourhoods will form by themselves, which might as well *not* be true.

Instrumentalism:

The instrumentalism approach states that ethnic conflict does not emerge from differences in ethnicity, but that ethnicity is politicised, resulting in ethnic conflict. Clearly, this goes beyond neighbourhoods. One way in which ethnicity was being politicised was by the Dutch politician Geert Wilders with his “less Moroccans” statement in 2014. He clearly used ethnicity to set up one group of people against another group. This led to a wave of protests (Omroep Brabant, 2014), and he was indicted by 6400 people (RTL Nieuws et al., 2016). Obviously, there were already tensions, but this could have developed into an ethnic violence in the Netherlands, especially in heterogeneous neighbourhoods with a lot of different ethnicities (Wu et al., 2011). This is a clear sign that these abetments, like the one of Wilders, are still used in western countries, often associated with racism. Because these things can happen in all places, this implies that it does not make a difference which country a neighbourhood is in. The argument that ethnicity can be politicised appears to be valid for many countries. Another way to politicise ethnicity can be by forced segregation, although this is not always necessarily a bad thing, as shown in the following paragraph.

That the instrumentalism approach is also applicable at neighbourhood level shows research of Klašnja and Novta (2016, p. 927):"We find that for highly ethnically polarised societies, increasing ethnic segregation decreases the incidence and intensity of conflict. In contrast, in societies with low ethnic polarization, increasing segregation increases conflict. This is because segregation and polarisation jointly determine the spread of conflict". This can be applied to neighbourhoods. If there are multiple polarized ethnic groups living together in neighbourhoods and there is constant tension, it might be better to just separate these groups (if that does not happen by itself, because people want homogeneity as pointed out in chapter 2.1.6).

Constructionism:

Within the constructionism approach, there are three different 'sub-approaches', which all provide a different view on how ethnicity is created and how ethnic conflicts are formed. Individuals, discursive formations and structural forces all can socially construct ethnicity and can start ethnic conflict. Also, by not viewing ethnicity as fixed, like in primordialism, it is possible to change ethnicity. If ethnicity is socially constructed, one might be able to adopt another ethnicity, by moving to another country for example. For neighbourhoods, one can wonder how big differences between third generation people with a non-western migration background and native Dutch people really are, as much integration might have taken place.

If it is about individuals who socially construct ethnicity and create ethnic conflict, there are some examples. In some ways one can see the aftermath of the George Floyd murder as an ethnic conflict for black people (and everybody who sympathised with them), which started by the killing of George Floyd by the police agent (CNN World, 2020). This can also happen in neighbourhoods, and with tension present, everybody is on high alert.

Discursive formations can work on neighbourhood level too (although it seems unlikely that this 'stays in the neighbourhood'), when one tells stories about a group, and these stories are taking over, while socially constructing ethnicity.

The impact of structural forces on ethnicity are also there for neighbourhoods. Political structural forces also exist on city level, like in the earlier mentioned segregation example. Therefore, it is clear that structural forces can also create an ethnic conflict on city level. Another possibility is that the structural forces are worldwide, but that it has an effect on the ethnic groups in a neighbourhood. In the aftermath of the George Floyd murder, there were not only protests in the United States, but all across the globe, it affected people all around the globe, to go protesting (CNN World, 2020).

The theory which explains best is likely to be a combination of primordialism as well as instrumentalism. This is because ethnicity is unique for every group and if you are born in that group, you will likely adopt that ethnicity. However, this does not result in immediate ethnic conflict, there are enough examples of heterogeneous countries in which many ethnic groups live peacefully together. To prevent such ethnic conflict (in neighbourhoods), most approaches will agree that homogeneity in neighbourhoods is not necessarily a negative matter and that it might even help preventing ethnic conflict (e.g., the primordialist and to a certain degree the instrumentalist approach as well). Following the thought of the primordialist and instrumentalist approach, an ethnic conflict would be harder to form in a homogeneous neighbourhood (Adlparvar & Tadros, 2016).

2.4 Conceptual model

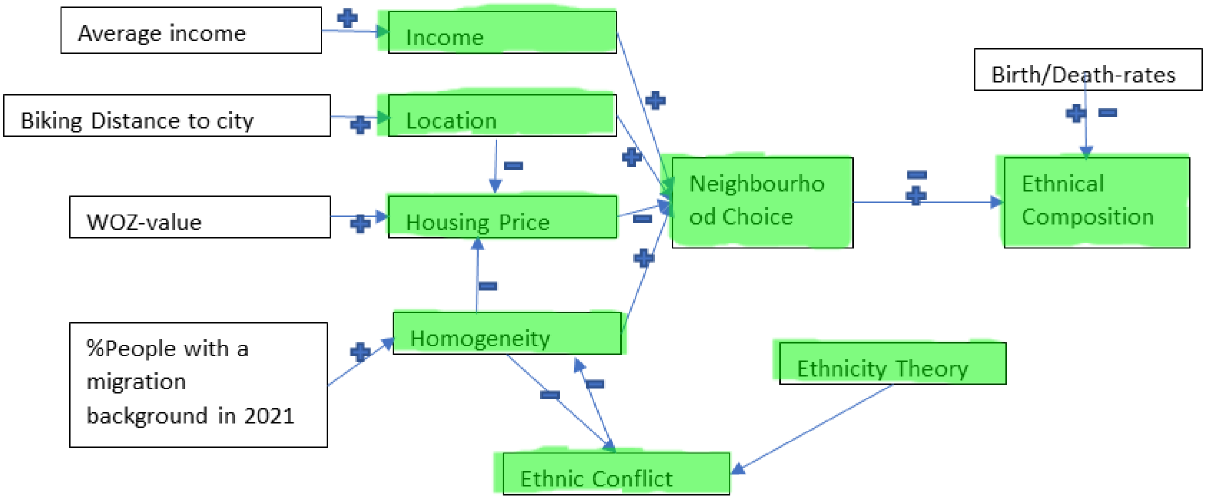

Figure 1: *Conceptual Model*.

The conceptual model as shown in Figure 1 is derived from findings in the literature and the application of the ethnicity theory. Based on those findings, a data-analysis will be set up in chapter 3. These findings are the factors which could have a potential effect on the ethnic composition (besides housing prices: income, location and homogeneity). Furthermore, the ethnicity theory has been put into the conceptual model, and applied to the concept of ethnic composition, to understand them better. These factors came forth from the literature, in which relationships between those factors are shown. For example, the relationship of homogeneity and ethnic composition through neighbourhood choice. This applies to every variable. As pointed out in the theory, the ethnic composition is connected to neighbourhood choice. Following research by Zwiers et al. (2018) the ethnic composition can only be changed when someone *decides* to move to a certain neighbourhood (or birth or death rates, but that is outside the scope of this thesis). There are exemptions in which someone is forced to move to a certain neighbourhood, but that is in fact also a choice (albeit not their own).

It is good to mention that this thesis aims for a better understanding of the middle section (the green marked part and their connecting arrows). The four factors which influence the four factors are not within the scope of this thesis. The box with the factor birth/death rates is outside the scope of this thesis as well. The four boxes on the left of the green part are deemed to be the most influential factors on the ethnic composition. All of these, except housing prices, show a positive correlation to neighbourhood choice. Neighbourhood choice is in this case is defined as the amount of (new) people choosing to live in a certain neighbourhood, thus moving there. This effect of housing price on neighbourhood choice is negative, because the more expensive a house is, the fewer people can afford this (Centraal Bureau voor de Statistiek et al., 2014). Depending on which people are moving, this can either have a positive effect or negative effect on the ethnic composition (I.e., an increasing or decreasing of the percentage of people with a non-western migration background). In the conceptual model, the used theory has been added, to visualise which factors can be clarified or explained by this theory. However, it was decided to not place any plus or minus symbols, because these would not have any meaning, as one cannot really quantify the ethnicity theory.

There are some arrows between the independent variables. Location does influence the housing price negatively, since housing further from the city centre is cheaper (McMillen & Singell, 1992). There is also an arrow from homogeneity to the housing price, because the percentage of people with a non-western migration background does influence the housing price (demand decreases because of native flight and so does the housing price) according to Kalantaryan and Alessandrini (2020).

The following factors are being controlled for: first, income (which was found in the research of Hedman et al. (2011)), and second, location, (which was found in study from Chhetri (2006) and in the research of the National Association of Realtors (2011)), are of importance. Plus, the factor of homogeneity, which was found in the research done by Chhetri et al. (2006), Hedman et al. (2011) and Ham (2017). These factors came out of the literature as essential factors (on neighbourhood choice and eventually ethnic composition). One could argue that crime might be a reason to move out of another neighbourhood and should therefore be a fifth box. However, there was no accurate data on this available at neighbourhood level; the data which is available would be inaccurate (other neighbourhood layout). Furthermore, crime would partly be due to ethnic conflict and tension, especially in heterogeneous neighbourhoods (Gu et al., 2023). Hence, the homogeneity will increase in a neighbourhood, because of native flight (Kalantaryan & Alessandrini, 2020) and according to the primordial approach inevitable conflict decreases as a result from this increasing homogeneity (Adlparvar & Tadros, 2016). This has further effects on the ethnic composition, making crime still partly influencing ethnic composition (through the homogeneity variable).

The intention is to use all of these four independent variables, but in case there is multicollinearity or if it is deemed better to leave them out of the model (if the benefits outweigh the disadvantages), some of the independent variables will be left out and estimated separately. Multicollinearity will be explained in chapter 3.4.4.

The concept of neighbourhood choice is not used as a variable. This concept influences ethnic composition, so the choice has been made to leave it out of the equation for the reason of model simplicity and multicollinearity. Furthermore, it is hard to operationalise neighbourhood choice and residential mobility. Although there is a good unit which is able to express this variable, when the amount of people who moved in a neighbourhood is the definition. The point is that the data needs to be precise; it needs to be per neighbourhood and also the ethnicity needs to be known (of which people moved out or in). This data is not available in big datasets with the CBS neighbourhood layout. If one chooses to express it in a percentage, it would in the end become nothing more than a constant without adding a lot to the model. And last but not least, the type of people who made the choice to live in a certain neighbourhood will be part of the same ethnic composition one wants to measure. So, one ends up measuring a very similar effect, over the same group of people. Nevertheless, it has been proven by this theoretical framework that it is very useful to look at the variable factor ‘neighbourhood choice’ while conducting research like this, to also find what has an effect on ethnic composition.

3. Methodology

The research will mainly be quantitative research, with a quantitative data analysis with data from the city of Amsterdam. Quantitative research tries to find causal relationships between variables by data and statistics, whereas qualitative data tries to understand a certain phenomenon by doing interviews and observations (Stevenson University, 2022). Since this research is about discovering and quantifying relationships between variables, quantitative research is used as method. Using qualitative research is not possible, since it is not known if there is a relationship which can be understood, because there is the possibility that there is no effect as well. An additional advantage which quantitative research has over qualitative research is that results are generalisable to the whole population. This can be done because the number of observations is relatively high, which ensures a higher external validity (Bhandari, 2023a); it can be used in a broader context than the city of Amsterdam alone.

All research questions will mainly be answered by the data-analysis, since it is quantitative research. However, this is done with supplementary existing knowledge from literature research. The second and third research question are there to answer the main research question more easily. The first and third research questions will be answered by the data analysis, further effects of the housing prices and other factors on the ethnic composition will be given through existing knowledge. The preamble of the answer on the second research question will be provided by the literature, but it will be further researched in the data analysis whether these findings were also suitable in case of this thesis. This is because even when being certain of those factors, as assumed in the theoretical framework (logically, since most literature aligns on this topic); a final check, whether these relationships apply in this thesis as well, will ensure an absolute and certain answer on the second research question.

3.1 Data Collection

Firstly, cross-sectional data is compiled from the internet, on which all needed data is available (via websites like CBS.nl, the Central Bureau of Statistics in the Netherlands). Cross-sectional data is data which has been collected during one point in time (for instance, a variable only contains data from the year 2022), but with many observations. The data is collected on the factors which have been discussed towards the end of the theoretical framework. This includes data on the percentage of non-western migrants from the years 2021 and 2022. Also, data regarding the housing values, distance to city centre and the average income per neighbourhood are collected. All the data collected is from the year of 2022 (except for the one year of percentage of non-western migrants in 2021). This is the most recent year possible, which is done to provide an image as actual as possible, without running the risk of drawing conclusions which are not valid anymore today. Then, the dataset was made, combining all collected data in one Excel file, which can be used in the statistical software SPSS, in which the analysis will be done. This Excel file contains one dependent variable and four independent variables. SPSS is the preferred choice over other software applications because it is a comprehensive statistical tool, with many built-in functionalities, but still easy to use and relatively quick in calculations (William, 2022). In Table 2 it is shown how the dataset looks in SPSS, showing 5 of the 99 variables. Explanation on the units of measurement follows later in chapter 3.3.

	Wijkenenbuurten	EthnicComposition	HousingPrice	Homogeneity	Location	AverageIncome
1	Burgwallen-Oude Zijde	19,01%	481,000	18,03%	,45	36,30
2	Burgwallen-Nieuwe Zijde	19,06%	476,000	18,26%	,45	40,70
3	Grachtengordel-West	12,64%	804,000	11,23%	1,10	62,30
4	Grachtengordel-Zuid	14,68%	837,000	12,83%	1,00	77,90
5	Nieuwmarkt/Lastage	17,11%	509,000	15,84%	1,50	42,70

Table 2: *Snapshot of dataset in SPSS.*

All data is on the measurement level of ratio. This means that there are equal intervals between variables. If one has the values of 1, 2, and, 3, the intervals are all 1, and the difference between 1 and 2 is the same as between 2 and 3. Also, ratio data has the characteristic that no negative values are possible (Bhandari, 2023b).

The term reliability (which refers to the consistency of measurements) is not applicable in this scenario. Reliability measures whether the measurement is consistent, so if one conducts the measurement again (by taking surveys for example), will one obtain the same result? That is the question reliability tries to answer (Benders, 2022). However, these results are obtained by CBS. Since this is a trusted agency in the world of statistics, it is assumed that the results of their research, which are used in this research, are reliable.

The data will be taken of all neighbourhoods in one city, which in this case is Amsterdam. In the dataset, the CBS layout of neighbourhoods is used. The data was already collected and put in order by CBS in this layout; the dataset contains 99 neighbourhoods, resulting in 99 observations. The fact that all neighbourhoods are taken can be explained: the city of Amsterdam is a diverse city, with all kinds of people with various backgrounds, which has the potential of being generalised to other parts of the Netherlands if the whole city is taken. However, there are potential issues with generalising results to the whole country, since it is a data analysis done with data from Amsterdam, which is entirely different from the countryside of the province of Drenthe for example. In order to partly counter this issue, all neighbourhoods are taken into consideration to ensure that the effect is valid for different types of neighbourhoods. Because the neighbourhoods are taken so widely, it will still be relatively representable of other parts in the Netherlands, Europe or even the other parts of the world (one can think of other cities similar to Amsterdam). It can be said that the study has a relatively high external validity. External validity is a measure for how well results can be generalised to (other parts of) the population (Andrade, 2018).

One might argue that there is a difference in population in different neighbourhoods. However, this is not a problem, because in this whole research, calculations are done in percentages and in averages. This means that differences in neighbourhood population size already have been taken into account.

3.2 Method Choice

A multiple regression analysis will be conducted, to see how much of the change in ethnic composition can be assigned to a change in housing prices and how much to other variables. A multiple regression analysis is a type of regression analysis in which multiple independent variables all influence one dependent variable. The advantage of this method is that one can see how big the effect of every single independent variable is on the dependent variable. This ensures that one can see the relative influence of certain independent variables (Weedmark, 2018), while maintaining the other factors constant (Smith, 2015). This makes a multiple regression set-up a suitable choice for this type of research. However, the disadvantage of this method is that there could be multicollinearity (which can eventually lead to dropping out variables for a good result). The variables which will potentially be dropped out, will be estimated separately from the other variables in a single regression analysis. The regression analysis will be either linear or non-linear, depending on the best fit.

3.3 Statistical Model

The following factors from the conceptual model were translated to variables and are intended to be used in the model (including the units in which they are measured):

- Housing prices: housing prices will be measured in WOZ values. This is because the WOZ value is determined by the recent average sales prices of comparable houses close by in the neighbourhood (Van Overbeek Makelaars, 2022). Thus, the housing price factor is made measurable by taking the housing values in the data analysis. The fact that this WOZ value is based on sales prices of houses in close proximity implies it could be really close to the actual sales price of the house (at the moment the value is calculated). Furthermore, it is corrected for the size of the house and property size.

  If one takes the average sales prices, that does imply that it only includes the houses which are already sold. For instance, if one takes the average sales price of the fourth quarter in 2022, only the houses sold in the fourth quarter in 2022 are taken into account.

The actual sale prices does not give a good image of the overall average of a neighbourhood, since in some neighbourhoods, more houses might be sold, disturbing the representation of a particular average sales price of a neighbourhood. The WOZ values are made with a computer model for each house; therefore, the average WOZ value is a clear representation of the WOZ values of the whole neighbourhood.

- Ethnic composition: This will be measured by the percentage of people in the neighbourhood with a non-western migration background. This gives a good image of how diverse the neighbourhood is, and how it is ethnically composed. An increase in ethnic composition means that there will be more people with a non-western-migration background. Other ethnic groups (like Dutch) will not be taken into account, as this thesis is specifically about people with a non-western migration background.
- Income: The average income of a person in a certain neighbourhood is taken.
- Location: It is hard to include convenience to job, school and shopping in one variable, since those facilities are all in different locations (these factors came out as important in Table 1). It is therefore best to choose the “biking distance to city centre in kilometres” as unit. The city centre is defined as the Dam Square. Most people are going to live close to their job (hence, this can be crossed out against each other for differences between and within neighbourhoods) and the same applies to schools. That is why this type of unit is chosen, as many people work towards to the city centre, which is the shopping area as well. There is specifically chosen for the biking distance, since much traffic within the city of Amsterdam goes by bike. Within the city centre, 60% of the trips are by bike (Dudley, 2022).
- Homogeneity: This will be measured in the percentage of people with a non-western migration background of a previous year. It should be a self-reinforcing effect. The ethnic composition of earlier years has an effect on the current ethnic composition. Because people choose a neighbourhood for the earlier ethnic composition (homogeneity), according to Hedman et al. (2011), thereby changing the current ethnic composition (Zwiers et al., 2018). By taking a year’s delay for the ethnic composition variable (to create this variable), an attempt is made to express the self-reinforcing effect. This self-reinforcing effect became clearer throughout the theoretical framework (chapter 2).

With regards to the choice of variables, there is an important matter which needs to be taken into account, which is validity. Validity in general considers the question whether a method measures what it wants to measure (Middleton, 2022). External validity has already been explained. According to Bhandari (2022a) internal validity means the following: “Internal validity is the extent to which you can be confident that a cause-and-effect relationship established in a study cannot be explained by other factors”. This can be checked by seeing if the dependent variable changes, when the independent variable is changing, which is the case in the data analysis of this thesis (Bhandari, 2022a). This implies that adding more variables can increase the internal validity. However, adding more variables can also increase the chance of the need to deal with multicollinearity (Wood & Airth, 2023). This is a realistic scenario, since the topic is a complex phenomenon, when looking at the conceptual model. When talking about validity in general, the variables do measure what is needed to be measured.

Because there can be multicollinearity, it is possible that some of these variables will not be taken into account in the final model (since they might disturb the outcome).

3.4 Data preparation

After collection of the data, the data will be prepared for the actual analysis. This will be done in this chapter and on the basis of the four assumptions, which need to be satisfied for linear regression analysis (Statistics Solutions, 2023). One can see this chapter as the methodological process of creating the best model. This means that the tests in this chapter are not part of the analysis, but the preparational work for the analysis itself. Before entering into the assumptions, a look at the descriptive statistics will be taken.

Data-Preparation:

Before all the assumptions were put through the test, it is always good to get a descriptive SPSS output. Table 3 shows the descriptive SPSS output which provides a complete overview of all numbers on every variable. This gives a lot of information on a wide range of topics.

Statistics

		EthnicComposition	HousingPrice	Homogeneity	Location	AverageIncome
N	Valid	99	99	99	99	98
	Missing	0	0	0	0	1
Mean		31,8458%	454,79765	31,1903%	5,2616	36,1051
Median		27,8243%	424,00000	26,8684%	4,7000	34,3500
Std. Deviation		16,24256%	185,642340	16,28847%	2,98498	12,89921
Skewness		,605	1,613	,634	,849	1,349
Std. Error of Skewness		,243	,243	,243	,243	,244
Kurtosis		-,559	3,101	-,492	-,006	2,000
Std. Error of Kurtosis		,481	,481	,481	,481	,483
Minimum		4,14%	115,000	3,89%	,45	17,60
Maximum		70,41%	1157,000	70,41%	12,30	79,10

Table 3*: Descriptive Statistics of All Variables.*

There was no need to account for missing values, since the dataset was almost complete. This means that there will be minor inaccuracies in that regard, but nothing needs to be done to deal with these issues, as it is not problematic for the model. This is partly due to the fact that the data was taken from recent years; the older the datasets were, the more missing values the available datasets for this type of research question had. Part of the reason why the most recent data has been chosen, is due to this fact.

The minimum, maximum, mean and median values give a good image on what the values are and what the actual reality is. For example, the average amount a person has to cycle to the city centre is 5.26 kilometres. This implies that cycling, for example, 6 kilometres to the city centre is normal. There are plenty of examples like this which can be derived from these descriptive statistics.

Assumptions for multiple regression:

These are the assumptions which need to be satisfied, to run a strong multiple regression model (Statistics Solutions, 2023):

1. The relationship can be expressed in a linear way.
2. The variables are normally distributed: a normal distribution without too many outliers or skewness.
3. There should be homoscedasticity, meaning: the variances are equal in the different groups being compared, so every observation has more or less the same error term (Scribbr, 2020). This is visualised in Figure 12. In other words: the error term has a constant variance.

4. There is little or no multicollinearity: because multicollinearity is a big topic within multiple regression analysis, it will be extensively covered and explained later on in this chapter.

All these assumptions will be tested, to eventually ensure that the model is valid and will not depict a biased outcome. Firstly, it is checked whether a linear model is usable, by drawing scatterplots with SPSS to see how the data is distributed. In case a linear model does not fit the data, there is a chance that the transformed data (in case there is no linear relationship or no normal distribution) will fit a linear model. However, transforming is only done when there are no other options, since it is generally considered better to work with untransformed variables, because of interpretation issues in the end (P. Beckers, personal communication, June 8, 2023). Maximum effort will be put into preparing the data ensuring that the assumptions will be met, so the regression model will be as strong as possible.

3.4.1 Assumption 1: Linearity

The first assumption states that the variable can be expressed in linear way. For a linear model to be successful, the data should have a relationship of linear nature. This needs to be satisfied, otherwise it does not make sense to run a linear model and a non-linear model would simply be better (Statistics Solutions, 2023). Normally, one chooses between a linear or non-linear based on how well the model fits the data (by looking to the goodness-of-fit = R-squared), this poses a problem which will be explained first: a non-linear model does not have valid R-squared (Frost, 2022a). In a linear regression model the total variance between groups is a combination of the explained variance and the unexplained variance (total variance = explained variance + unexplained variance). The R-squared is derived from this, it is the percentage explained variance of the total variance. However, for non-linear models this is not the case: total variance ≠ explained variance + unexplained variance. At the same time, this makes the R-squared not valid, SPSS does calculate an R-squared, but it is not trustworthy (Frost, 2022b).

Error term:

For this reason, a look at the standard error is taken. This standard error of the estimate needs to be as low as possible, because then the regression line fits the data (scatterplot) best. Such a scatterplot is visualised in Figure 3, and through these dots would go a regression line. In a scatterplot, the variables are set out against each other, with their respective observations in a graph. So, one observation has a particular value for, in this case, the housing price variable and the ethnic composition variable, and the observation will be put somewhere in the graph. Then a pattern of scatterplot dots will appear, out of which information regarding, for example, linearity can be derived. What happens is that there is only one line through the scatterplot points (which are observations) for which the sum of squares of the error terms is the smallest. This demands more clarification: it starts with the error term, in Figure 2, this is the distance between the actual observed value and the estimated value (which is the regression line). You want this distance to be as small as possible, because this means that your model is closer to the actual truth. To compare different models within datasets, one needs to get a number which tells how far the particular model lies from the actual truth. In a linear model, this is done by quadrating all the error terms, of all observations (every point/observation has its own error term because it has its own distance to the line). Then, all these numbers are summed up, to form the 'sum of squares'. The model which has the smallest sum of squares is the model in which the points are closest to the regression line. Hence, this model has the best fit, since it is the closest it can get to reality, compared to all other regression lines which are calculated. The computer calculates the best-fitting line; numerous lines are tried by the computer, but there is only one line which has the smallest sum of squares (Wagner, 2023).

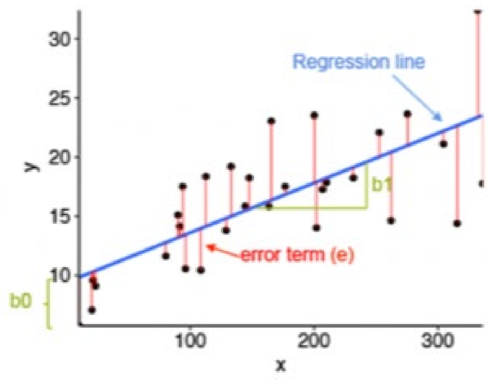

Figure 2: *Visualisation of the error-term* (Lakshmi, 2022)

It is possible to visualise the effect of two independent variables on the independent variable in a 3D-axis model with XYZ-axis. However, there is not a fourth dimension, hence visualising more than three variables is impossible. Therefore, only the variables of interest are taken.

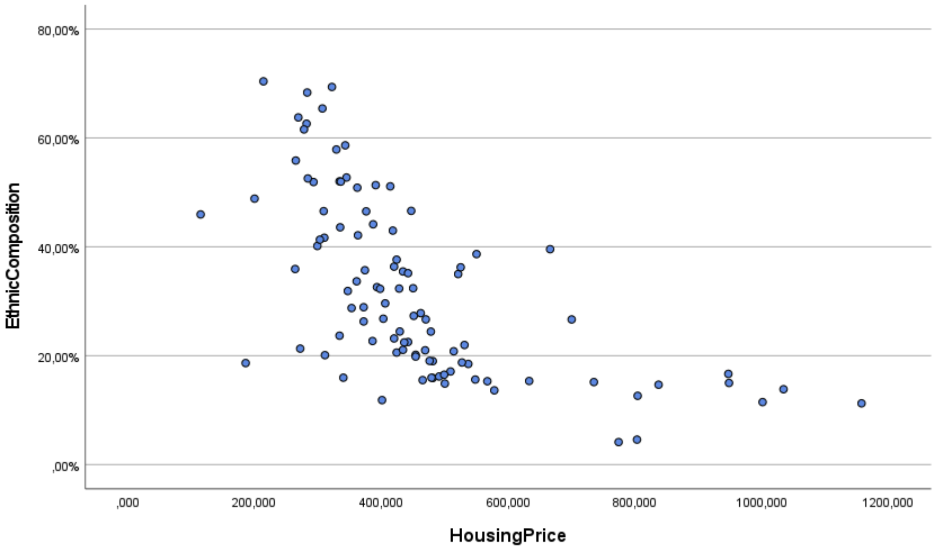

Figure 3: *Scatterplot to illustrate non-linear shape of data.*

As shown in Figure 3, a linear line does not seem to fit the shape of the data. A line that would fit better is some kind of negative exponential graph or half of a parabolic (because that is the shape the dots have). One of the possibilities is transforming the data, this leads to less outliers (extreme observations) and less skewed data. This might lead to better shaped data. However, it is best to transform as least as possible, due to interpretation issues after a regression analysis (P. Beckers, personal communication, June 8, 2023). Therefore, a test for linearity has been conducted for the relationships between the four independent variables and the dependent variable. These tests are visible in the Tables 4, 5, and 6.

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
EthnicComposition * HousingPrice	Between Groups	(Combined)	25028,055	90	278,090	2,692	,067
		Linearity	9758,688	1	9758,688	94,473	<,001
		Deviation from Linearity	15269,367	89	171,566	1,661	,226
	Within Groups		826,364	8	103,296		
	Total		25854,420	98			

Table 4: *Test for linearity between ethnic composition and housing price.*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
EthnicComposition * Location	Between Groups	(Combined)	18406,733	59	311,979	1,634	,053
		Linearity	4877,920	1	4877,920	25,543	<,001
		Deviation from Linearity	13528,813	58	233,255	1,221	,256
	Within Groups		7447,686	39	190,966		
	Total		25854,420	98			

Table 5: *Test for linearity between ethnic composition and location.*

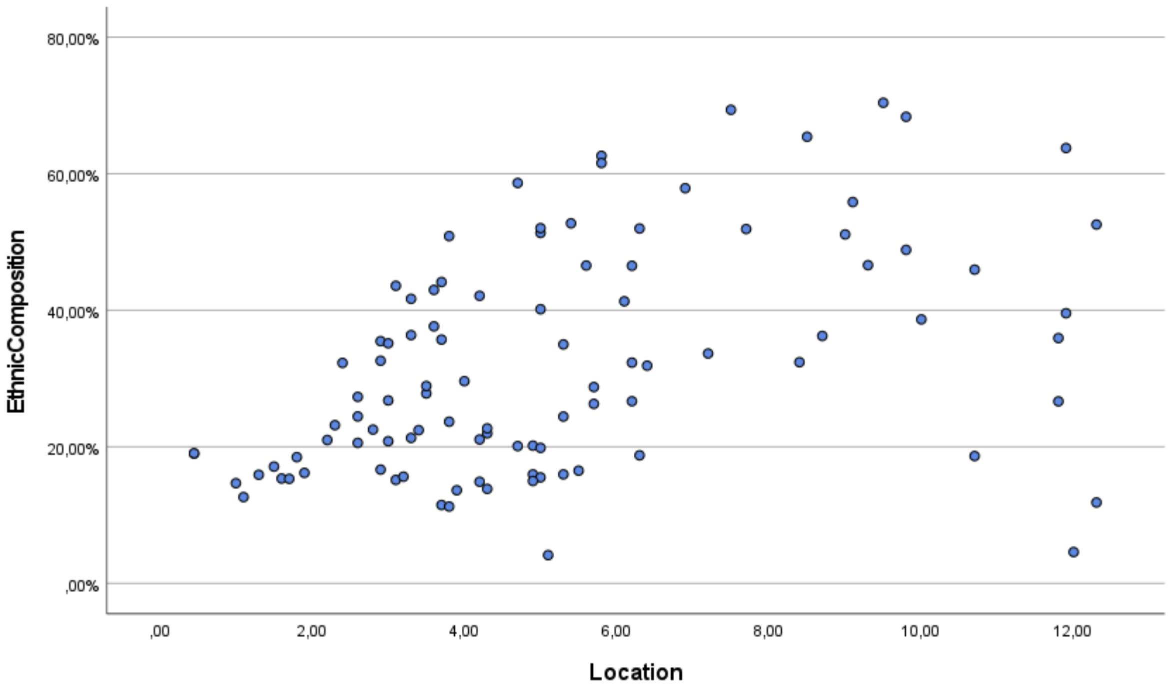

Figure 4: *Scatterplot to illustrate linearity between ethnic composition and location.*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
EthnicComposition * AverageIncome	Between Groups	(Combined)	24197,736	82	295,094	2,716	,016
		Linearity	11848,513	1	11848,513	109,063	<,001
		Deviation from Linearity	12349,223	81	152,460	1,403	,235
	Within Groups		1629,586	15	108,639		
	Total		25827,323	97			

Table 6: *Test for linearity between ethnic composition and average income.*

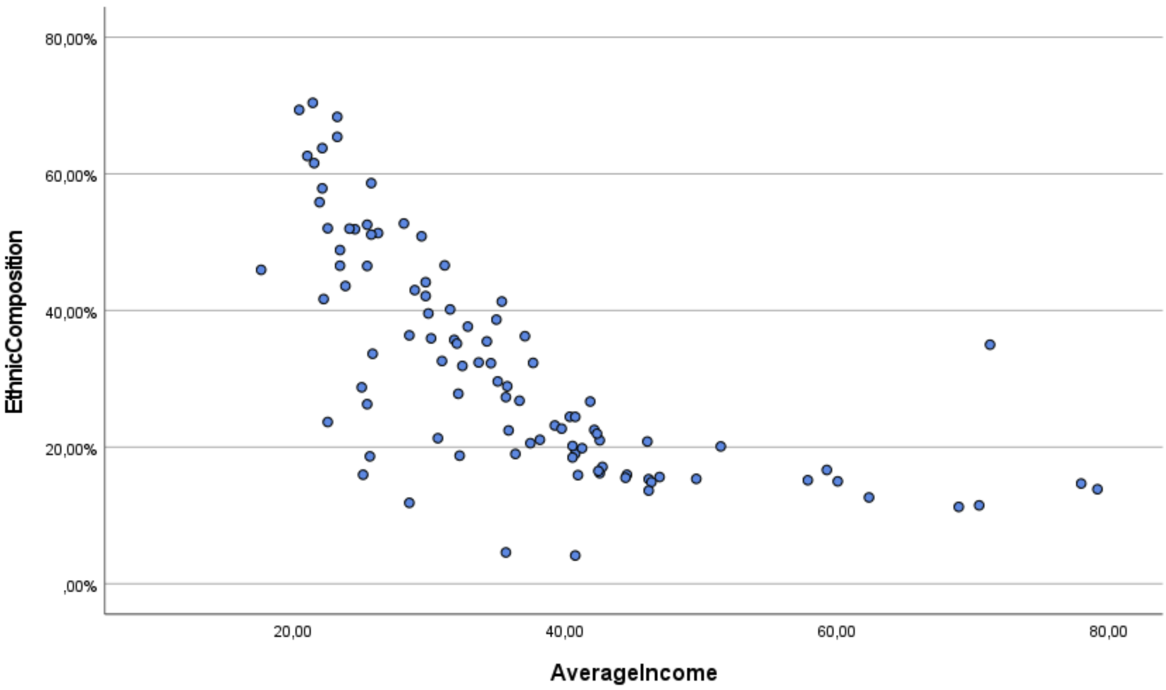

Figure 5: *Scatterplot to illustrate linearity between ethnic composition and average income.*

As visible in the Tables 4, 5, and 6 there is significant linearity and insignificant deviation from linearity. This is also visible in Figures 3, 4, and 5, although the image is a bit distorted due to outliers. This means that according to this test, three of the four independent variables do have a linear relationship. For the variable homogeneity, no ANOVA-table could be created (so no test for linearity has been conducted), therefore, a look at a scatterplot is taken. This scatterplot is shown in Figure 6 and illustrates that there definitely is a linear relationship.

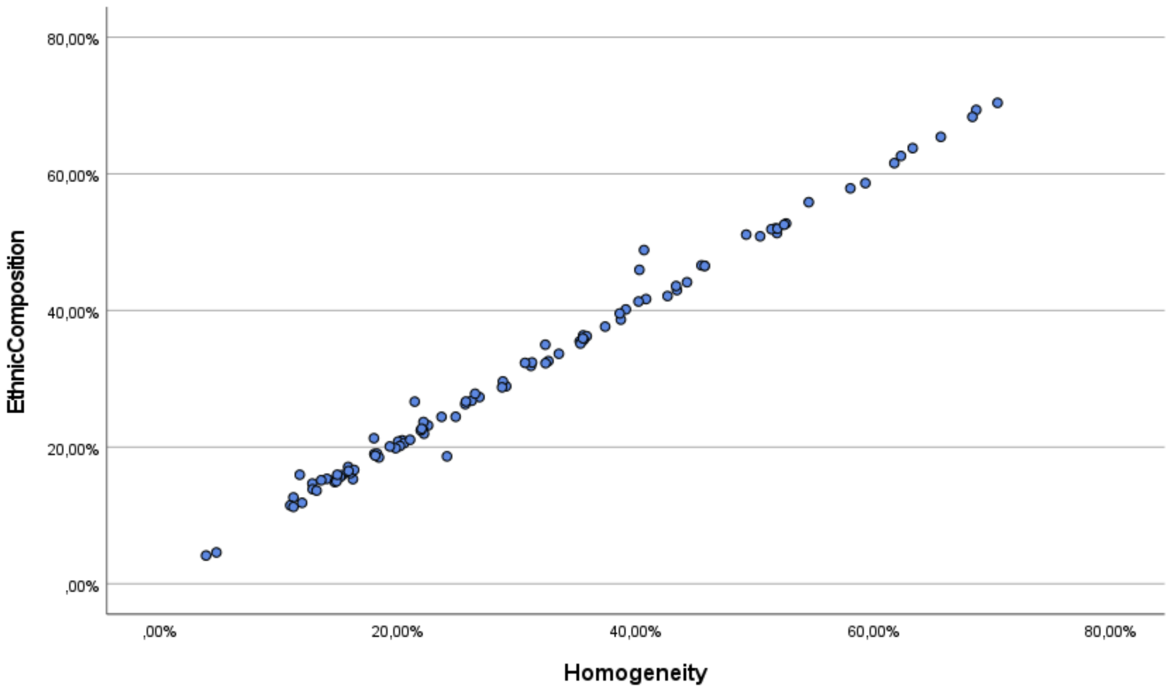

Figure 6: *Scatterplot to visualise linearity between ethnic composition and homogeneity variables.*

As is shown, all relationships are of a linear nature. This means that this linearity is not a reason for any potential transformation and that the first assumption is satisfied.

3.4.2 Assumption 2: Normality

The second assumption states that the data should be normally distributed. Once this is the case, the data analysis can be continued. Normally distributed data means that the observations are normally distributed around its mean. A normal distribution would follow the black line, which roughly follows a bell shape (Statistics Solutions, 2023). A visualisation is given in the Figures 7, 8, 9, 10, and 11, what normal or non-normal distributions could look like. The Figures 7 until 11 are the distributions of the observations of the variables used in this analysis.

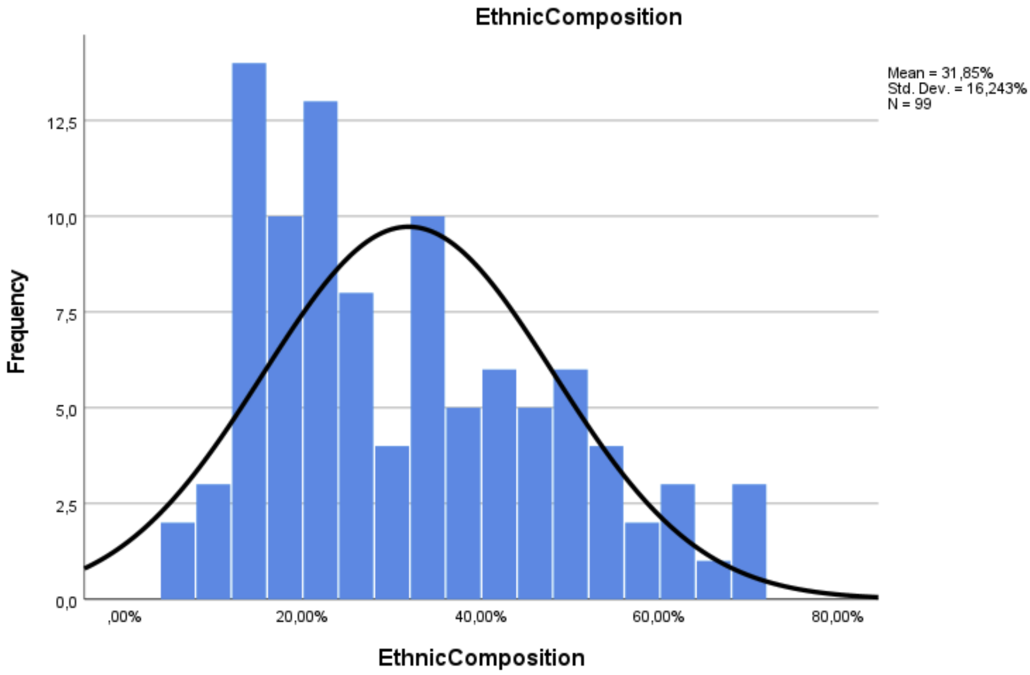

Figure 7: *Distribution of observations of the ethnic composition variable.*

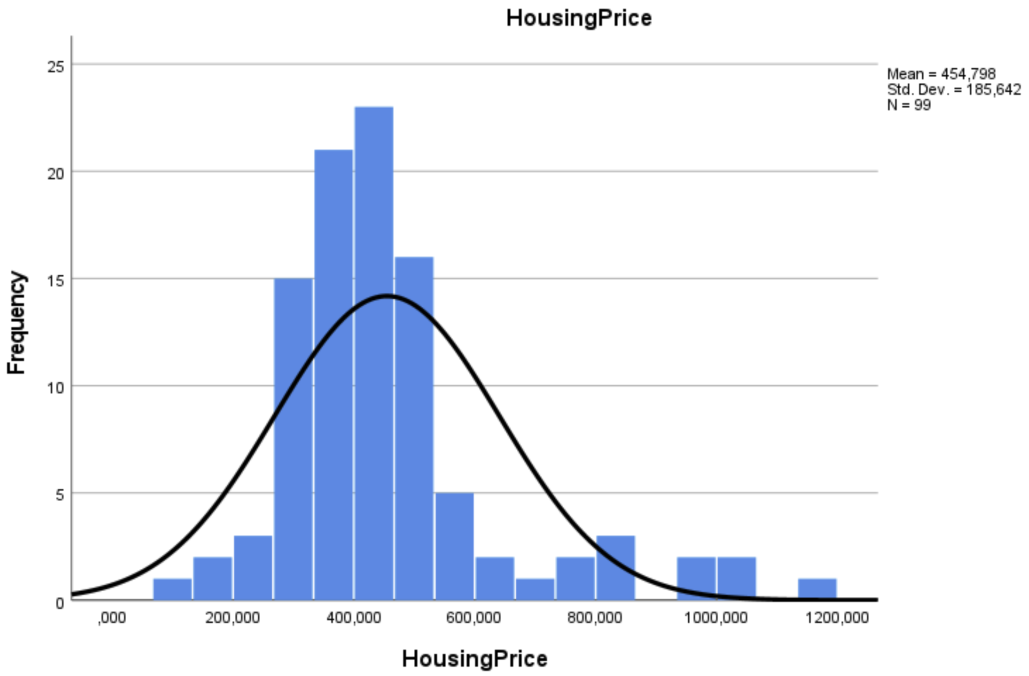

Figure 8: *Distribution of observations of the housing price variable.*

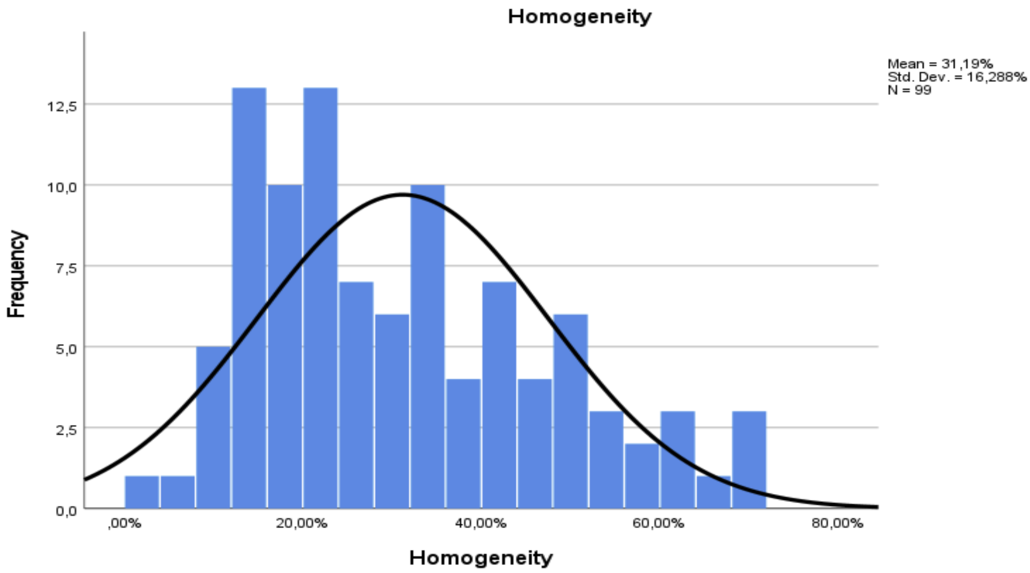

Figure 9: *Distribution of observations of the homogeneity variable.*

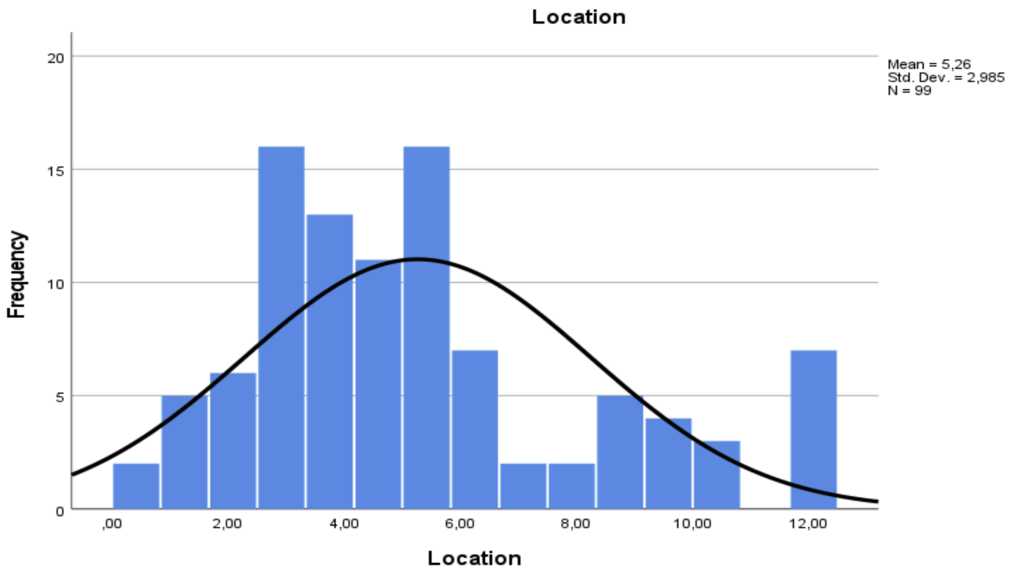

Figure 10: *Distribution of observations of the location variable.*

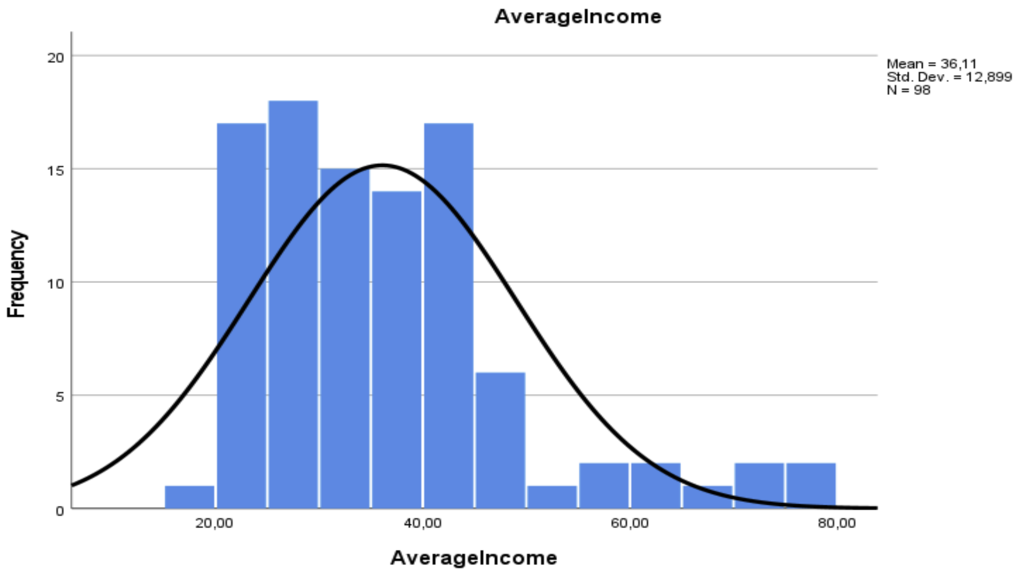

Figure 11: *Distribution of observations of the average income variable.*

In general, it is desirable that the blue bars fit as much within the black line as possible, otherwise it is a non-normal distribution. What a non-normal distribution of observation looks like is shown in Figure 9. If one draws an imaginary line through the blue bars, it becomes visible that this imaginary line would lean to the left. This could be a non-linear distribution.

Although it might not seem the case that the blue bars fit well within the black line, this does not need to be the case; the graph (being the imaginary line through the blue bars) can for example be distorted by outliers. Therefore, a closer look at the actual numbers has been taken. The shape of these bars is captured with the following terms: skewness and kurtosis and are explained in the following section.

3.4.2.1 Skewness

At first, the distribution of the data was not normally distributed. The data was very right-skewed (positively skewed), with skewnesses ranging from 0.605 to 1.613, as can be seen in Table 3. Skewness is when the distribution is asymmetric and leans towards the right (positive skewness) or towards the left (negative skewness). Skewness tells the direction of the outliers (Chen, 2023).

As one can see, there indeed is much skewness in this dataset, when one looks at the description table. This implies that the distribution of the observations is very asymmetrical, which could be an explanation for the non-linear data. However, this type of skewness does not necessarily impose problems, since skewness is acceptable if it is not smaller than -2 or bigger than 2 (Hair et al., 2022 as cited in Smartpls, 2023). None of the variables in the dataset show these larger amounts of skewness, so no transformations are done.

3.4.2.2 Kurtosis

Another matter which is clearly visible in the distributions (Figures 7 until 11), is that the tops are different from the normal distribution. This is called kurtosis and is a measure for the tailedness in a dataset. This further translates into the number of outliers in a dataset (Turney, 2022).

The normal distribution (black graph) has a kurtosis of 3 (Turney, 2022). According to MRC Cognition and Brain Sciences Unit (2009) SPSS has a disturbed definition of kurtosis, as the program subtracts 3 points from the actual number (I.e., all kurtosis numbers should be 3 higher in Table 3). This means that only the variable housing price and the variable income have relatively high kurtoses (mesokurtic). This also means that the other three variables (ethnic composition, homogeneity and location) have a normal kurtosis (platykurtic). This implies that these three last mentioned variables have thinner tails (so the graph is stretched). This means that there are less outliers than in the other two variable (housing price and income), which is generally speaking a good sign.

For kurtosis, the acceptable range lays between -7 and 7 (MRC Cognition and Brain Sciences Unit, 2009). All the variables lay within this range of kurtosis and are therefore acceptable, when looking at the values of the kurtosis in Table 3.

As can be seen in Table 3, the skewness as well as the kurtosis of the various variables are on an acceptable level, although some are on the high side. Therefore, it is assumed the variables are normally (enough) distributed and the second assumption is satisfied.

In the context of transforming as little as possible (due to interpretation issues), a decision has been made to continue with untransformed variables.

3.4.3 Assumption 3: Homoscedasticity

This assumption is about that there should be homoscedasticity. This means that the error terms have a constant variance, which are in fact the distance between the actual observation, visualised in Figure 12.

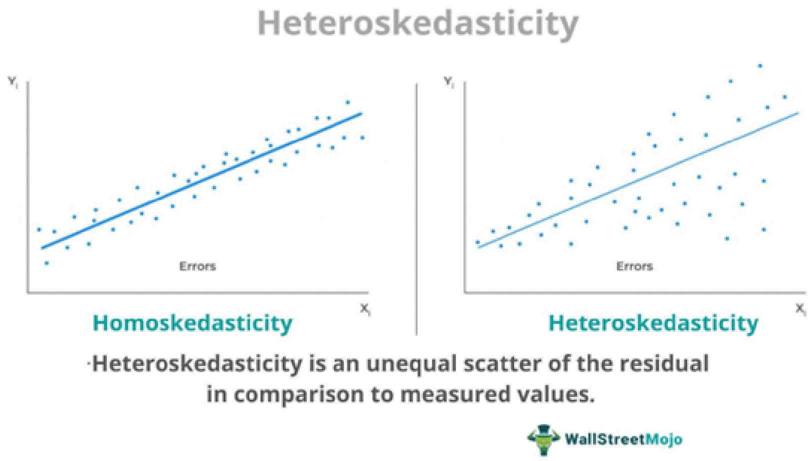

Figure 12: *Visualisation of Homoscedasticity and Heteroscedasticity* (Wallstreetmojo Team, 2022).

The variance is the distance between the predicted value and the observed value. In case of heteroscedasticity, these error terms do not have the same variance. The problem is that SPSS assumes that homoscedasticity is the case, and if not, a model could be invalid. With heteroscedasticity the variance of the estimated regression coefficient could be bigger (so a higher standard error), but the model does not see this, because it assumes that there is no heteroscedasticity. This makes that sometimes regression coefficients are shown as significant, but in reality, they are not (Statology, 2020). What basically happens if a coefficient is not significant, is that it does not mean anything, and that the data does not provide evidence that the relationship exists (Hennig, 2022). Hence, it is very important that coefficients are significant.

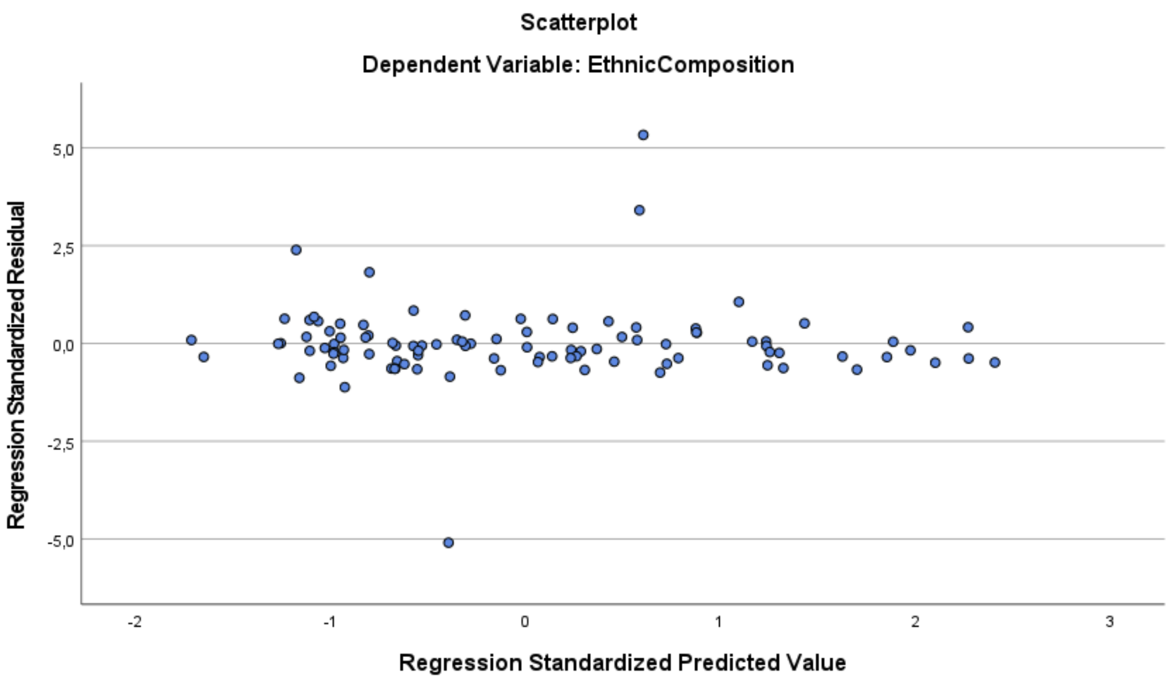

Figure 13: *Scatterplot in which the predicted values have been set out against the actual values.*

As one can see in Figure 13, there is barely heteroscedasticity, and therefore, this assumption is satisfied. Figure 13 has been created by plotting the residuals against each other in a scatterplot. This means that the predicted value (the ones on the regression line which one can draw through the scatterplot) is set out against the actual values from the scatterplot.

3.4.4 Assumption 4: Multicollinearity

The fourth assumption will be tested with in this paragraph. Multicollinearity occurs when the independent variables are correlating with each other, and thus disturbing the end result. The goal of regression is to isolate an independent variable and measure the separate effect of that particular variable on the dependent variable. What happens if the independent variables are (highly) correlated is that a change in independent variable 1 leads to a change in independent variable 2 and also in the dependent variable. This implies that it is unclear how much of that change in the dependent variable is caused by independent variable 1. In other words, it is not clear then, how much of the variance explained by that variable can be assigned to the direct relationship. As independent variable 2 also changed, this too has an effect on the dependent variable. Briefly, this is the problem of multicollinearity (Frost, 2023).

A closer look to the VIF values was taken, before the actual analysis was done, as multicollinearity can severely damage the estimations. To counter this, one often ends up removing one of the independent variables (or in our case, it can happen that one of the four independent variables needs to be removed). Other consequences are that if there is a small uncertainty, the standard error goes up. This makes the t-value go down (t-value = estimated coefficient/standard error), which makes the coefficients not significant anymore (Buijs, 2023).

Multicollinearity typically reveals itself through extremely high R^2-values; it is as if the model 'fits too well'. Usually an R^2-value higher than 0.8 is suspicious. Another way to find potential multicollinearity is to run linear regression on the independent variables, to see if they are correlated between themselves. The last way to find multicollinearity is by using the VIF-value (Variance Inflation Factor). This VIF value is derived from the R^2-value, by the following formula:

VIF = 1/(1-R^2).

With an R^2 not allowed to be higher than 0.8, this means that the VIF should not be higher than 5 (Buijs, 2023).

3.4.4.1 First Attempt

The first test, with the full model, there is potential multicollinearity at hand. While conducting a regression analysis, the Tables 7, 8 and 9 are part of the typical SPSS regression output:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,997[a]	,994	,993	1,33892%	2,436

a. Predictors: (Constant), AverageIncome, Location, Homogeneity, HousingPrice

b. Dependent Variable: EthnicComposition

Table 7: *Model Summary with explained variance (R-square).*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	,371	,877		,423	,674		
	HousingPrice	-,003	,001	-,031	-1,997	,049	,294	3,396
	Homogeneity	,992	,012	,993	84,331	<,001	,500	1,999
	Location	,064	,054	,011	1,182	,240	,740	1,352
	AverageIncome	,038	,022	,030	1,782	,078	,240	4,167

a. Dependent Variable: EthnicComposition

Table 8: *Estimated coefficients table.*

Correlations

		EthnicComposition	HousingPrice	Homogeneity	Location	AverageIncome
Pearson Correlation	EthnicComposition	1,000	-,616	,997	,453	-,677
	HousingPrice	-,616	1,000	-,611	-,309	,833
	Homogeneity	,997	-,611	1,000	,449	-,681
	Location	,453	-,309	,449	1,000	-,450
	AverageIncome	-,677	,833	-,681	-,450	1,000
Sig. (1-tailed)	EthnicComposition	.	<,001	<,001	<,001	<,001
	HousingPrice	,000	.	,000	,001	,000
	Homogeneity	,000	,000	.	,000	,000
	Location	,000	,001	,000	.	,000
	AverageIncome	,000	,000	,000	,000	.
N	EthnicComposition	98	98	98	98	98
	HousingPrice	98	98	98	98	98
	Homogeneity	98	98	98	98	98
	Location	98	98	98	98	98
	AverageIncome	98	98	98	98	98

Table 9: *correlation matrix with mutual correlations.*

This high R^2-value in Table 7 could mean multicollinearity and when looking into the correlation-matrix (Table 9), it becomes clear that this high value comes from the homogeneity variable. This independent variable correlates strongly with every other independent variable, but it also correlates strongly with the dependent variable. However, the VIF-values in Table 8 are all acceptable, so there might not be multicollinearity, although the average income variable has a VIF-value of 4.167, which is on the high side. Moreover, multicollinearity only exists between independent variables, while here it is clear that the high R^2-value comes from the high correlation between the ethnic composition and homogeneity variables. Logically, these variables are highly correlated. Because the values of the percentage of non-western migration background in a neighbourhood do not change abruptly (rather it slowly increases or decreases). Then, the model says that the ethnic composition of year 1 is a good estimator for the ethnic composition of year 2.

The t-values in Table 8 are relatively low (except the homogeneity variable, which is suspiciously high). This results in insignificant coefficients, and only the homogeneity and housing price variable are significant. T-values are often used to check the hypothesis. This thesis does not have a hypothesis, because it is mainly descriptive research and there is no statement to prove or to disprove (Jablonski, 2014). In case there is a H0-hypothesis and a H1-hypothesis, a higher t-value leads to a greater chance to reject the H0-hypothesis. In other words, according to Minitab (2016): "The greater the magnitude of T, the greater the evidence against the null hypothesis. This means there is greater evidence that there is a significant difference. The closer T is to 0, the more likely there is not a significant difference". Hence, a higher t-value leads to stronger evidence the relationship between two variables is actually there. Therefore, high t-values are favourable.

Last but not least, the standard errors are on the high side. What the standard error actually measures, is the uncertainty around a number. This shows the following according to Bhandari (2022b): "A high standard error shows that sample means are widely spread around the population mean—your sample may not closely represent your population. A low standard error shows that sample means are closely distributed around the population mean—your sample is representative of your population."

Concluding that this is a failed attempt with the results not being favourable, with a lot of mutual correlations, low t-values, high standard error, and potential multicollinearity. The homogeneity variable is suspected to be the cause of these poor results and will therefore be dropped out of the model, and the effect will be estimated separately. It is dropped out to improve the model results, to properly estimate the main effect of interest (the effect of housing prices on ethnic composition). One cannot afford that one variable is disturbing the estimation of every other variable. Therefore, another attempt has been undertaken to see if there was any improvement in the model result.

3.4.4.2 Second Attempt

In this model, there was heteroscedasticity as can be seen in Figure 14. This means that the created homoscedasticity came from the homogeneity variable and by dropping this homogeneity variable, the underlying heteroscedasticity comes to the surface. So, now the third assumption is not satisfied.

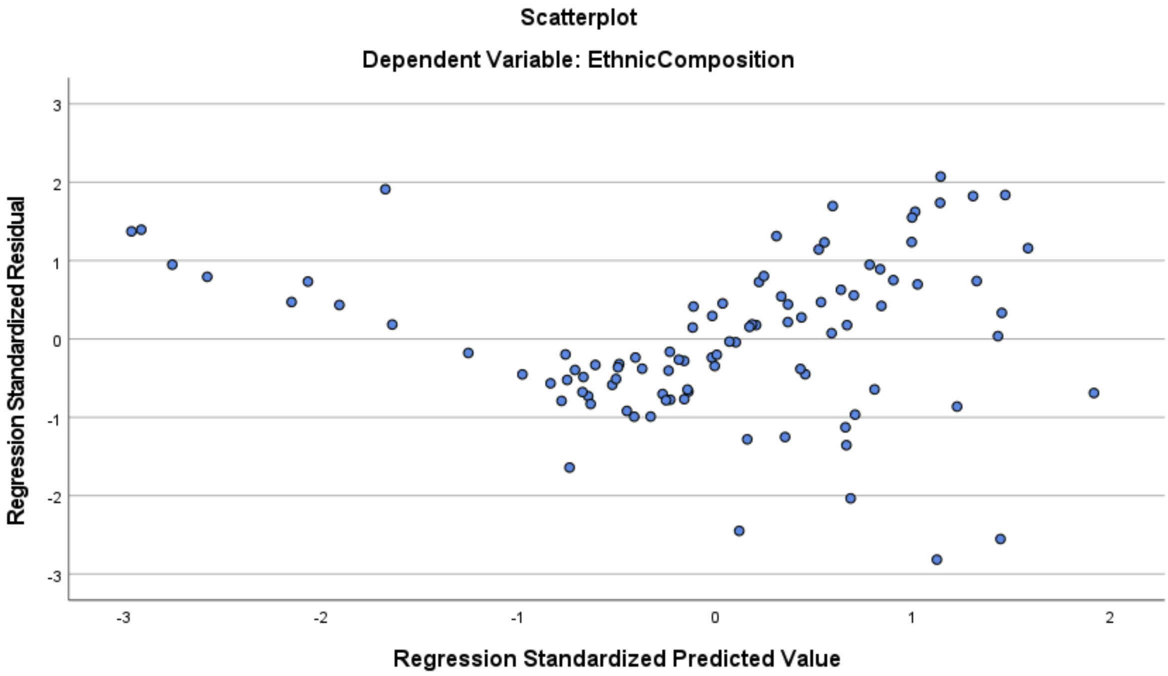

Figure 14: *Actual values and predicted values plotted against each other.*

Because homoscedasticity is one of the assumptions, which is needed to be met, the other results become more or less irrelevant. Firstly, this problem of heteroscedasticity needs to be solved. The solution lies in a log-transformation (Statology, 2020). According to Htoon (2021) this is a powerful transformation tool, in which the logarithms of all values are taken. One transforms a full variable, this means that of that variable, all the values for that particular variable will be transformed.

The new value would be:

$$\text{new value} = {}^{10}\log(\text{old value})$$

Mathematically, this would mean that every value shrinks significantly. For instance, if the old value is 100, this would mean that:

$$\text{new value} = {}^{10}\log(100) = 2, \text{ since } 10^2 = 100$$

This example illustrates how much a log-transformation shrinks the old values. What will happen in a scatterplot like Figure 14 is that the log-transformed variables will be more centred around the x-axis (the line y=0), because the distances between the actual and predicted values will get smaller (all values shrink). Therefore, a log transformation is a solution for heteroscedasticity. This log-transformation was applied, and the data was already more homoscedastic, except a couple of outliers, but the homoscedasticity assumption was satisfied. Also, the other assumptions were satisfied. Therefore, these results deemed to be useful, and this means that the best possible model has been found, ending this methodological process. The results of this test are taken to chapter 4.

3.5 Simple Linear Regression

One of the four independent variables have been thrown out of the model, namely the homogeneity. This variable will be tested separately, with untransformed variables (since the transformation was needed because the homogeneity variable has been thrown out of the model) This will be done with the intention of still getting a good estimation of the effect.

Furthermore, another simple linear regression will be done, for the location variable, because it did not yield a significant coefficient. With the simple linear regression, an attempt will be made to still get significant coefficients. This will be done with log-transformed variables, to get the data as homoscedastic as possible.

3.6 Ethical Considerations

The researcher will take a fully independent and unbiased stand in the data analysis and the interpretation of the results. Also, the researcher will accurately report data outcomes and not selectively report data for a result in his favour. To ensure this, data collection, analysis and reporting will be fully transparent, and everything will be made available for further research. Furthermore, all other professional codes of conduct will be adhered to, ensuring that this research is conducted with full integrity and honesty.

4. Results

In this chapter, the results of the various regression analyses will be set out. As shown in the methodology section, various regression analyses have been run, to ensure the best possible outcome. In the multiple regression analysis, the effect of the independent variable housing prices (made measurable by taking the WOZ values), the variable location (made measurable by calculating the amount in kilometres to the city centre on a bike), the variable homogeneity (made measurable by taking the percentage of every neighbourhood) and the average income variable (measured by the average income of a person in a neighbourhood) are all measured on the dependent variable ethnic composition. Remember that the independent variable homogeneity is not in the main model anymore.

That is the reason why a simple linear regression has been run. In the simple single linear regressions the effect of several of those independent variables is measured separately as well, since they were dropped out, to improve model quality. The independent variable homogeneity (made measurable by taking the ethnic composition of earlier years) will be measured on the ethnic composition.

Firstly, the multiple regression will be shown, thereafter the two simple single linear regressions. Also, some brief explanatory commentary will be provided, showing that the models are good enough, without making any concluding remarks.

4.1 Multiple Regression

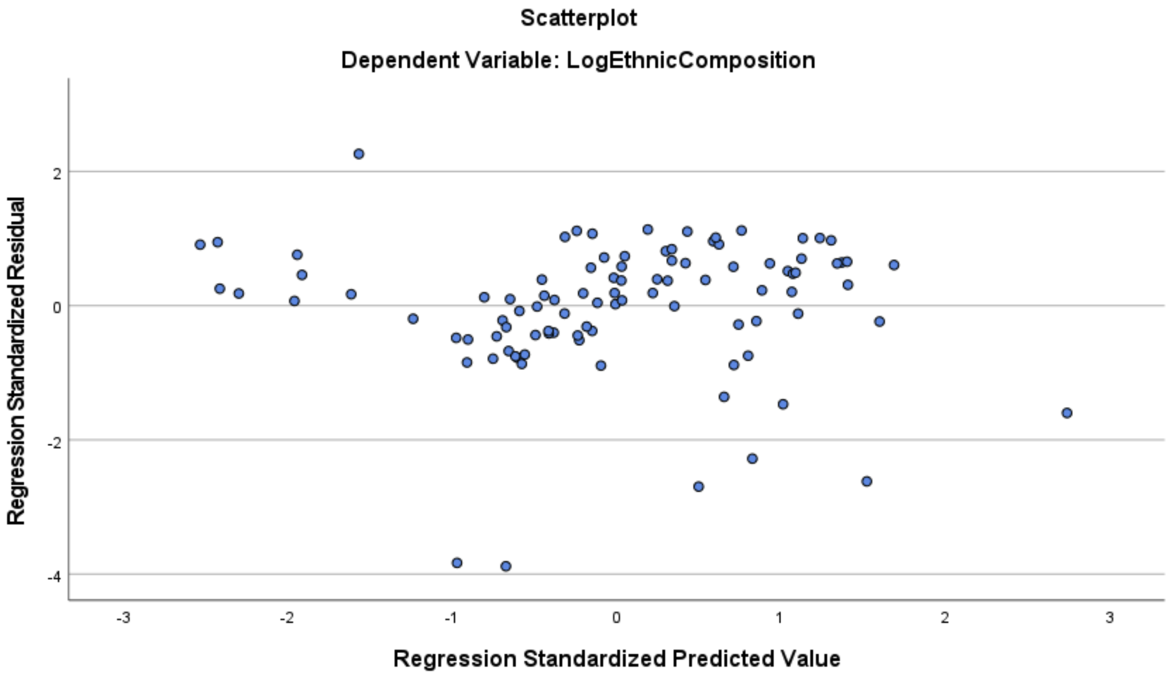

Figure 15: *Scatterplot to show homoscedasticity*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,732[a]	,536	,521	,16968

a. Predictors: (Constant), LogAverageIncome, LogLocation, LogHousingPrice

b. Dependent Variable: LogEthnicComposition

Table 10: *Model summary to show explained variance.*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	3,771	,326		11,562	<,001		
	LogHousingPrice	-,452	,188	-,298	-2,405	,018	,322	3,102
	LogLocation	,043	,070	,050	,617	,538	,764	1,308
	LogAverageIncome	-,764	,225	-,442	-3,403	<,001	,293	3,415

a. Dependent Variable: LogEthnicComposition

Table 11: *Coefficient table to show estimated coefficients, t-values, and VIF-values.*

Correlations

		LogEthnicComposition	LogHousingPrice	LogLocation	LogAverageIncome
Pearson Correlation	LogEthnicComposition	1,000	-,681	,383	-,711
	LogHousingPrice	-,681	1,000	-,398	,823
	LogLocation	,383	-,398	1,000	-,485
	LogAverageIncome	-,711	,823	-,485	1,000
Sig. (1-tailed)	LogEthnicComposition	.	<,001	<,001	<,001
	LogHousingPrice	,000	.	,000	,000
	LogLocation	,000	,000	.	,000
	LogAverageIncome	,000	,000	,000	.
N	LogEthnicComposition	98	98	98	98
	LogHousingPrice	98	98	98	98
	LogLocation	98	98	98	98
	LogAverageIncome	98	98	98	98

Table 12: *Correlation matrix to show mutual correlations.*

In Figure 15 it is visible that there is no real heterogeneity, except for some small outliers (pay attention to the scale of the axis, which could be misleading while interpreting this).

The final explained variance is 0.521 as can be seen in Table 10. So, of all the variance between neighbourhoods (with regards to percentage of people with a non-western migration background), 52.1% of the variance in the ethnic composition variable is explained. When testing with more than one independent variable, one looks at the adjusted R^2. This is because R^2 can be increased by adding more variables, which can lead to overfitting and multicollinearity (Wagner, 2023).

From Table 11, the following information can be derived. The coefficients are all significant on a level of 0.05, except for the location variable. This variable will not become significant, even not by dropping another variable out of the model. This was tested afterwards, but these results are not relevant for this results section. This can be due to a lack of observations, or the actual absence of a relationship (Ebert, 2018). Furthermore, in Table 11, all the VIF-values are lower than 5, so there is no significant multicollinearity.

All variables are log-transformed, and the relationship is of linear nature. According to Ford (2018) this should be interpreted as follows, while looking at the coefficient table in Table 11: 1% increase in the housing price leads to a decrease of 0.452% in the ethnic composition. And a 1% increase in the average income of a neighbourhood leads to a decrease of 0.764% of the ethnic composition.

The standardised coefficient can be used to compare coefficients between themselves, because all variables are in different units of measurement. The standardised coefficient does not have a unit of measurement, but measures in the number of standard deviations. This implies that it is useful for comparing the different independent variables with each other (Goyal, 2023). However, the standardised coefficient is measured in standard deviations, so it also depends on the size of the standard deviation. It is therefore possible that the unstandardised coefficient differs significantly from the standardised coefficient.

The location variable is not significant. Therefore, interpretation would not make sense. It will be presented in the linear regression.

4.2 Simple Linear Regression

4.2.1 Homogeneity variable

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,996[a]	,992	,992	1,42178%

a. Predictors: (Constant), Homogeneity

b. Dependent Variable: EthnicComposition

Table 13: *Model summary of simple linear regression with homogeneity as independent variable.*

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	,862	,310		2,780	,007		
	Homogeneity	,993	,009	,996	112,663	<,001	1,000	1,000

a. Dependent Variable: EthnicComposition

Table 14: *Coefficient table of simple linear regression with homogeneity as independent variable.*

The interpretation of these results is that 99.2% of the variance is explained, as derived from Table 13. The coefficient is 0.993. The model predicts that the same neighbourhood would have an ethnic composition which is 0.993% higher than the other neighbourhood in 2022, concluded from Table 14.

4.2.1 Location variable

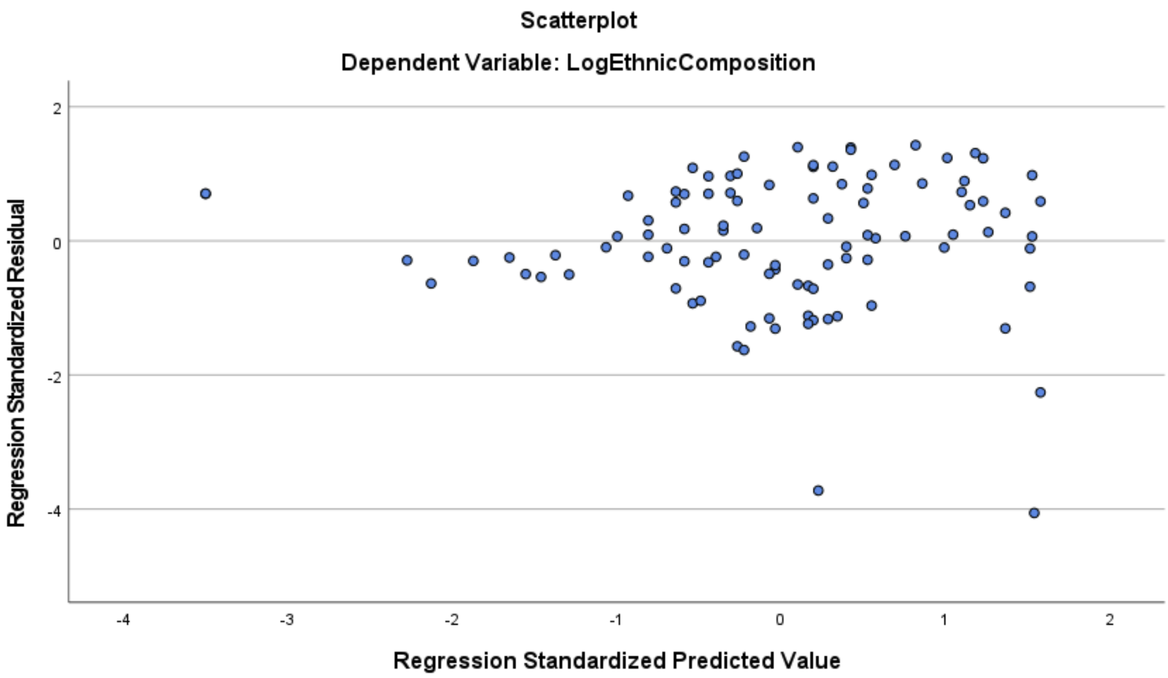

Figure 16: *Scatterplot of predicted values set out against actual values.*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,377[a]	,142	,133	,22709

a. Predictors: (Constant), LogLocation

b. Dependent Variable: LogEthnicComposition

Table 15: *Model summary of simple linear regression with location as independent variable.*

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	1,232	,057		21,635	<,001		
	LogLocation	,325	,081	,377	4,011	<,001	1,000	1,000

a. Dependent Variable: LogEthnicComposition

Table 16: *Coefficients table of simple linear regression with location as independent variable.*

As shown in Figure 16, this regression came with a bit of heteroscedasticity. But it is not as bad as in Figure 12, because the classic cone shape was not there. Therefore, it was deemed to be useful. From Table 15 can be derived that location could explain 13.3% of the variance in the ethnic composition. In Table 16, the coefficient of 0.325 was significant at a level of 0.05. A 1% change in the location, means an increase of 0.325% in the ethnic composition. In other words: if the cycling distance to the city centre (Dam Square) goes up by 1%, the percentage of people with a non-western migration background increases by 0.325%.

5. Discussion & Conclusion

5.1 Discussion

The research question was what the effect of housing prices is on the ethnic composition of a neighbourhood (what is the percentage of people with a non-western migration background in a neighbourhood?) is. The sub-research questions were what the other influential factors on ethnic composition are and, what the effect of those other factors is.

To answer the main research question, a quantitative data-analysis has been done. The result of this analysis is that if the average housing price in a neighbourhood goes up by 1%, the percentage of people with a non-western migration background decreases by 0.452%. Later in the discussion, it becomes clear that the effects of housing price on ethnic composition reaches further than only this percentage.

In order to answer the second and third research questions, which were what other influential variables on ethnic composition could be and what their effect was, a look at neighbourhood choice has been taken. This is because a factor which influences neighbourhood choice, ultimately also influences the ethnic composition (Zwiers et al., 2018). Out of this came that other than the housing prices, homogeneity plays a big role. Also, the location, relative to job and acquaintances (e.g. friends or family) and income played a big role in neighbourhood choice, and therefore also in ethnic composition. The effect of housing prices is significantly smaller than the effect of homogeneity and is also smaller than the effect of the average income.

The effect of homogeneity would be 0.933% in the percentage of people with a non-western migration background in 2022 per 1% increase of this group of people in 2022. This implies that the effect of homogeneity is still there, because the percentage of people with a non-western migration background is still increasing, but the growing rate is declining, hence the self-reinforcing effect becomes less strong (since 0.933% < 1%).

The effect of the average income is the following: every 1% increase in the average income in a neighbourhood, the percentage of people with a non-western migration background decreases by 0.764%. The effect of location was that for every 1% increase in the cycling distance to the city centre, the percentage of people with a non-western migration background goes up by 0.325%.

These effects are not comparable to each other, therefore, a look to the standardised coefficients is taken. Average income has a standardised coefficient of 0.442, locations' standardised regression is 0.377 and housing price has a standardised coefficient 0.298. This means that the effect of average income is more than 1.5 times the size than that of the housing price. The effect of location on ethnic composition lays between the average income variable and the housing price variable.

One could argue that the absolute effect of a 1% increase in the housing price (-0.454% in ethnic composition) is higher than the absolute effect of a 1% increase in the location variable (+0.325% in ethnic composition), while the standardised coefficient of the location variable is bigger than the one of the housing price variable (respectively 0.377 against –0.298). But as mentioned earlier in the results chapter, these standardised coefficients are only for comparing the effect (because of different measuring units or scales) and moreover, this standardised coefficient depends on the standard deviation (Goyal, 2023).

That the effect of housing prices on ethnic composition reaches further than only the percentage from the data analysis, becomes clear when connecting the results of the data analyses to the existing knowledge. People with a non-western migration background generally have less income on average (Koot et al., 2019), which makes they often live in cheaper neighbourhoods (Centraal Bureau voor de Statistiek et al., 2014), as more people start to move in, native flight occurs (Kalantaryan & Alessandrini, 2020). That people with a migration background often live in cheaper neighbourhoods is also visible the effect of income on ethnic composition; the average income variable showed a strong negative relationship with ethnic composition. Meaning that if the average income in a neighbourhood goes up, the percentage of people with a non-western migration background will decrease, which also shows that people with a non-western migration background have a lower income.

The effect of the other factors (location, average income, homogeneity) on ethnic composition reaches further than their percentages from the data analysis as well.

The data analysis showed a positive relationship between location and the ethnic composition. This means that the further a neighbourhood is located from the city, the higher the percentage of people with a non-western migration background. This aligns with the existing knowledge since housing further away from the city centre is cheaper according to research by McMillen and Singell (1992). Because people with a non-western migration background earn less (Koot et al., 2019), they live further from the city, because it is cheaper and this impacts the place they live (Centraal Bureau voor de Statistiek et al., 2014). This came forth from the literature, as well as the data analysis conducted for this thesis. This could well be the reason why housing prices, income and location were correlated with each other in the data-analysis, showing that there are mutual relationships between them. This would line up with the existing knowledge on these topics, provided in chapter 2.2.

Homogeneity influences ethnic composition because people want to live together with people who share the same characteristics. This is supported by the data analysis, which showed that a 1%

increase in the percentage of people with a non-western migration background in a neighbourhood in 2021, would lead to an increase of only 0.993% in that same neighbourhood in 2022.

Furthermore, there is the native flight principle, which states that native people (those are not per definition Caucasian) move away if too many people with a different background are coming to the neighbourhood. Combining this with the fact that the data-analysis, shows that higher housing prices lead to a lower percentage of people with a non-western migration background in a neighbourhood. This fact also goes the other way: lower housing prices would then lead to more people with a non-western migration background in neighbourhoods.

Adding this data and the theory around homogeneity together, will result in a scenario in which more and more native people move out of the neighbourhood (native flight) and more and more people with a non-western migration background will move in (because of homogeneity, there are already many similar people living there). This potential cycle would start as follows:

This whole process starts with low housing prices, which will attract people with a migration background, since they have a lower income on average. Then, native flight occurs and neighbourhoods become increasingly homogeneous, which lowers housing prices even more and attracts even more people with a migration background.

The fact that native people see people with a migration background as a negative amenity and therefore flee, is partly supported by ethnicity theory, depending on the approach one is using. For instance, if one is looking from primordialism, moving out would mean avoiding inevitable ethnic conflict; hence, primordialism would see native flight as something positive.

However, there is also a lot of research available which shows the opposite effect (I.e., rising housing prices through migrants) as shown by Zhang et al. (2017)., namely that there are economic chances available for migrants and potentially also for people with a non-western migration background who are going to live closely together. This could lead to a better network, which creates economic opportunities.

5.2 Conclusion

The first research question was what the effect of the housing prices would be on the ethnic composition, among other influential factors. The answer on the first research question is that a 1% increase in housing price leads to a decrease of in the percentage of people with a non-western migration background of 0.452%, showing a negative relationship. The reason for this is that those people in general have less income(Koot et al., 2019); native people see an increase of people with a non-western migration background as a negative amenity, thus demand and therefore prices within this neighbourhood decrease, this is called native flight (Kalantaryan & Alessandrini, 2020). The answer to the second research question, on which other influential factors on ethnic composition there are, was the following: income, location and homogeneity. The third research question was what the effects of these earlier mentioned others factors were. To answer the third research question: it can be concluded that homogeneity plays by far the biggest role when it comes to ethnic composition. People do want to live together with people with whom they share the same characteristics (Hedman et al., 2011). This also causes native flight, which means the native people move out when too many different people come in. The quantified effect of homogeneity is an increase of 0.993% in the percentage of people with a non-western migration background if the ethnic composition of the previous year would increase by 1%. After homogeneity, the average income of a neighbourhood plays a significant role when it comes to ethnic composition. Namely, if the average income would go up by 1%, then the percentage of people with a non-western migration background would decrease with 0.764%. The effect of the location variable is that if the cycling distance goes up by 1%, the percentage of people with a non-western migration background goes up by 0.325%. Housing price comes only in as fourth most influential variable on the ethnic composition, if one compares the standardised coefficients.

Critical Reflection

Even though writing this thesis did not go without obstacles, in general, there were no major obstacles. Points of improvement for this thesis could be the depth of the theoretical framework, but also the data processing. There might have been other ways in which the data could have been processed better (preparation, transformation, etc.), by using other transformation techniques that the author is not aware of. However, for the level of this thesis, the data was processed well. Also, the depth of the theoretical framework could have been better, by diving even deeper into some of the concepts which were there. Furthermore, there might have been theories which are closer connected to the concept, of which the author was not aware. This would have resulted in a

Another point of improve would be the conceptual model. The author should have made a bigger one right from the start, to give a good and comprehensive overview of the relations which are there (in the network of coherent factors). This would have saved a lot of time and thinking, instead of just writing and drawing it all down in a conceptual model.

The last matter which should have been improved is the choice for the expression 'ethnic composition', as this might cause confusion by the reader. It would be better to more carefully choose another expression, which is more unambiguous.

Recommendations for further research

It would be helpful for other researchers to conduct further research on whether an increase in migrants in neighbourhoods (and potentially people with a non-western migration background), leads to rising or declining housing prices. This in order to see what the real effect is on neighbourhoods, since there are two directions this road can take (increasing or decreasing housing prices with all its consequences). Researching this topic further can lead to significant improvements in policy, as, if it turns out negatively, this will have serious consequences.

A big limitation of this research lies in the way the effect of homogeneity is measured. It is debatable whether the effect of homogeneity is measured correctly, due to the extreme adjusted R^2-values. This is definitely one of the limitations of this thesis. Further research could therefore aim to improve the measurability of this variable and therefore make a better quantification of this very important factor on ethnical composition. Another research recommendation would be to research whether homogeneity is really a self-reinforcing effect like the cycle provided in the discussion session. In case it would be, further research on this would be recommended.

References

Adlparvar, N., & Tadros, M. (2016). The Evolution of Ethnicity Theory: Intersectionality, Geopolitics and Development. IDS Bulletin, 47(2). https://doi.org/10.19088/1968-2016.136

Alcoff, L. (2021, September 15). Critical Philosophy of Race (Stanford Encyclopedia of Philosophy). Stanford Encyclopedia of Philosophy. Retrieved June 21, 2023, from https://plato.stanford.edu/entries/critical-phil-race/

Andrade, C. (2018). Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation. Indian Journal of Psychological Medicine, 40(5), 498–499. https://doi.org/10.4103/ijpsym.ijpsym_334_18

Benders, L. (2022). Validiteit en betrouwbaarheid vaststellen in je scriptie. Scribbr. https://www.scribbr.nl/onderzoeksmethoden/validiteit-en-betrouwbaarheid-vaststellen-scriptie/

Bhandari, P. (2022a). Internal Validity in Research | Definition, Threats, & Examples. Scribbr. https://www.scribbr.com/methodology/internal-validity/

Bhandari, P. (2022b). What Is Standard Error? | How to Calculate (Guide with Examples). Scribbr. https://www.scribbr.com/statistics/standard-error/

Bhandari, P. (2023a). External Validity | Definition, Types, threats & examples. *Scribbr*.

https://www.scribbr.com/methodology/external-validity/

Bhandari, P. (2023b). Levels of Measurement | Nominal, Ordinal, Interval and Ratio. Scribbr.

https://www.scribbr.com/statistics/levels-of-measurement/

Bilge, N., Marino, M., & Webb, L. M. (2021). Definitions of Ethnicity in Communication Scholarship: A New Perspective. Howard Journal of Communications, 32(3), 213–234. https://doi.org/10.1080/10646175.2021.1871869

Birrell, B., & Healy, E. (2008, 1 januari). How Are Skilled Migrants Doing? Informit. https://search.informit.org/doi/abs/10.3316/ielapa.157423695183563

Brittner, T. (2022, 23 april). Ethnic Enclave Overview & Examples | What is an Ethnic Enclave? - Video & Lesson Transcript | Study.com. study.com. https://study.com/academy/lesson/what-is-an-ethnic-enclave-definition-examples-types.html

Buijs, L. (2023, februari 22). Econometrics Lecture 4. Econometrics Regression analysis of time-series data (I), Nijmegen, Gelderland, Nederland.

Centraal Bureau voor de Statistiek, Latten, J., Das, M., Musterd, S., & Van Gent, W. (2014). Bevolkingstrends 2014: Te veel of te weinig inkomen voor de buurt? Dan een grotere kans om te verhuizen. In *cbs.nl*. CBS. Retrieved July 23, 2023, from https://www.cbs.nl/-/media/imported/documents/2014/17/2014-01-04-b15-art.pdf?la=nl-nl

Centraal Bureau voor de Statistiek. (2022, November 23). Meer mensen willen verhuizen maar kunnen geen woning vinden. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/nieuws/2022/47/meer-mensen-willen-verhuizen-maar-kunnen-geen-woning-vinden

Centraal Bureau voor de Statistiek. (2023). Persoon met een niet-westerse migratieachtergrond. cbs.nl. Retrieved June 21, 2023, from https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/persoon-met-een-niet-westerse-migratieachtergrond

Che, A. M. (2016). Linking Instrumentalist and Primordialist Theories of Ethnic Conflict. E-International Relations. https://www.e-ir.info/2016/06/01/linking-instrumentalist-and-primordialist-theories-of-ethnic-conflict/

Chen, J. (2023). Skewness: Positively and Negatively Skewed Defined With Formula. Investopedia. https://www.investopedia.com/terms/s/skewness.asp

Cheung, C., & Leung, K. (2011). Neighborhood homogeneity and cohesion in sustainable community development. Habitat international, 35(4), 564–572. https://doi.org/10.1016/j.habitatint.2011.03.004

Chhetri, P., Stimson, R. J., & Western, J. (2006). *Modelling the Factors of Neighbourhood Attractiveness Reflected in Residential Location Decision Choices*. 2006 by the Japan Section of the Regional Science Association International. https://www.jstage.jst.go.jp/article/srs/36/2/36_2_393/_article

CNN World. (2020, June 13). Protests across the globe after George Floyd's death. edition.cnn.com. Retrieved July 25, 2023, from https://edition.cnn.com/2020/06/06/world/gallery/intl-george-floyd-protests/index.html

Delgado, R., & Stefancic, J. (1998). Critical Race Theory: past, present, and future. *Current Legal Problems*, *51*(1), 467–491. https://doi.org/10.1093/clp/51.1.467

Delgado, R., & Stefancic, J. (2006, 1 mei). Critical Race Theory: An Introduction, Second Edition on JSTOR. Geraadpleegd op 6 juni 2023, van https://www.jstor.org/stable/j.ctt9qg9h2

De Ree, M. (2016, October 25). Termen allochtoon en autochtoon herzien. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/corporate/2016/43/termen-allochtoon-en-autochtoon-herzien

Dudley, D. (2022, October 15). How the Bicycle Conquered Amsterdam. *Bloomberg.com*. https://www.bloomberg.com/news/articles/2022-10-14/how-the-bicycle-conquered-amsterdam

Ebert, T. A. (2018). Reasons: 1) Small sample size relative to the variability in your data. 2) No relationship between dependent and independent variables. [Comment on "Hello, Why regression coefficients are not statistically significant in simple regression ?"]. https://www.researchgate.net/post/Hello_Why_regression_coefficients_are_not_statistically_significant_in_simple_regressionn

Feder, S. (2020, December 1). *Gentrification disproportionately affects minorities | Stanford News*. news.stanford.edu. Retrieved July 15, 2023, from https://news.stanford.edu/2020/12/01/gentrification-disproportionately-affects-minorities/

Ford, C. (2018, August 17). *Interpreting Log Transformations in a Linear Model | University of Virginia Library Research Data Services + Sciences*. data.library.virginia.edu. Retrieved June 19, 2023, from https://data.library.virginia.edu/interpreting-log-transformations-in-a-linear-model/

Frost, J. (2022a). How to Choose Between Linear and Nonlinear Regression. Statistics By Jim. https://statisticsbyjim.com/regression/choose-linear-nonlinear-regression/

Frost, J. (2022b). R-squared Is Not Valid for Nonlinear Regression. Statistics By Jim. https://statisticsbyjim.com/regression/r-squared-invalid-nonlinear-regression/

Frost, J. (2023). Multicollinearity in Regression Analysis: Problems, Detection, and Solutions. Statistics By Jim. https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/

Glas, I., Engbersen, G., & Snel, E. (2019). The street level and beyond: The impact of ethnic diversity on neighborhood cohesion and fear of crime among Dutch natives and nonnatives. Journal of Urban Affairs, 41(6), 737–755. https://doi.org/10.1080/07352166.2018.1533379

Goyal, C. (2023). Understanding Regression Coefficients: Standardized vs Unstandardized. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/03/standardized-vs-unstandardized-regression-coefficient/

Gu, X., Liu, L., Lan, M., & Zhou, H. (2023). Measuring perceived racial heterogeneity and its impact on crime: An ambient population-based approach. Cities, 134, 104188. https://doi.org/10.1016/j.cities.2022.104188

Ham, V. M. (2017). *Incorporating Neighbourhood Choice in a Model of Neighbourhood Effects on Income | TU Delft Repositories*. https://repository.tudelft.nl/islandora/object/uuid:c39efd80-9e16-4011-aa7c-55a66636da0e?collection=research

Hammel, D. (2009). Gentrification. In *Elsevier eBooks* (pp. 360–367). https://doi.org/10.1016/b978-008044910-4.01051-8

Hansen, E. (2021, August 9). *Bos en Lommer: Amsterdam's gentrification frontier – revisited - Centre for Urban Studies*. Centre for Urban Studies - University of Amsterdam. Retrieved July 15, 2023, from https://urbanstudies.uva.nl/content/blog-series/amsterdams-gentrification-frontier-revisited.html

Hedman, L. (2012, January 31). *Moving Near Family? The Influence of Extended Family on Neighbourhood Choice in an Intra-urban Context*. onlinelibrary.wiley.com. Retrieved February 16, 2023, from https://onlinelibrary.wiley.com/doi/10.1002/psp.1703.

Hedman, L., Van Ham, M., & Manley, D. (2011). Neighbourhood Choice and Neighbourhood Reproduction. *Environment and Planning A: Economy and Space*, *43*(6), 1381–1399. https://doi.org/10.1068/a43453

Hennig, C., PhD. (2022, 20 januari). Interpretation of Non statistically significant coefficient. stats.stackexchange.com. https://stats.stackexchange.com/questions/561208/interpretation-of-non-statistically-significant-coefficient

Htoon, K. S. (2021, December 13). Log Transformation: Purpose and Interpretation - Kyaw Saw Htoon - Medium. Medium. https://medium.com/@kyawsawhtoon/log-transformation-purpose-and-interpretation-9444b4b049c9

Isajiw, W. W. (1993, 1 januari). DEFINITION AND DIMENSIONS OF ETHNICITY: A THEORETICAL FRAMEWORK. TSpace Repository. https://tspace.library.utoronto.ca/handle/1807/68

Jablonski, J. (2014, 16 juni). Research and Course Guides: PSC 352: Introduction to Comparative Politics: Develop a Thesis/Hypothesis. libguides.mssu.edu. https://libguides.mssu.edu/c.php?g=185296&p=1223659

Jenks, M., & Dempsey, N. (2007). Defining the neighbourhood: Challenges for empirical research. Town Planning Review, 78(2), 153–177. https://doi.org/10.3828/tpr.78.2.4

Kalantaryan, S., & Alessandrini, A. (2020). Housing Values and the Residential Settlement of Migrants: Zooming in on Neighbourhoods in Italian Provincial Capitals. Spatial demography, 8(3), 293–350. https://doi.org/10.1007/s40980-020-00068-1

Kataria, S. (2018). Explaining Ethnicity: Primordialism vs. Instrumentalism. Advances in Social Sciences Research Journal, 5(4). https://doi.org/10.14738/assrj.54.4394

Kim, B. (2015, September 15). Should I Always Transform My Variables to Make Them Normal? | University of Virginia Library Research Data Services + Sciences. data.library.virginia.edu. Retrieved June 23, 2023, from https://data.library.virginia.edu/normality-assumption/

Klašnja, M., & Novta, N. (2016). Segregation, Polarization, and Ethnic Conflict. The Journal of Conflict Resolution, 60(5), 927–955. https://doi.org/10.1177/0022002714550084

Koot, P., van Elk, R., & Jongen, E. (2019). Inkomensongelijkheid naar migratieachtergrond in kaart. In CPB.nl. https://www.cpb.nl/sites/default/files/omnidownload/cpb-achtergronddocument-inkomensongelijkheid-naar-migratieachtergrond-in-kaart.pdf

Lakshmi, M. (2022, May 2). *Understanding Linear Regression with all Statistical Terms* - Data Science Blog. Data Science Blog. https://data-science-blog.com/blog/2022/05/02/understanding-linear-regression-with-all-statistical-terms/

Lim, S., Yi, S. S., De La Cruz, N. L., & Trinh-Shevrin, C. (2017). Defining Ethnic Enclave and Its Associations with Self-Reported Health Outcomes Among Asian American Adults in New York City. Journal of Immigrant and Minority Health, 19(1), 138–146. https://doi.org/10.1007/s10903-015-0334-6

Maltchev, S. (2021, August 5). *Rotterdam's rising housing prices: A panel study on the effects of gentrification on the ethnic makeup of the city's neighbourhoods.* Retrieved February 16, 2023, from https://thesis.eur.nl/pub/59723/Maltchev-S-495645-thesis-.pdf

Martén, L., Hainmueller, J., & Hangartner, D. (2019). Ethnic networks can foster the economic integration of refugees. Proceedings of the National Academy of Sciences of the United States of America, 116(33), 16280–16285. https://doi.org/10.1073/pnas.1820345116

McMillen, D. P., & Singell, L. D. (1992). Work location, residence location, and the intraurban wage gradient. Journal of Urban Economics, 32(2), 195–213. https://doi.org/10.1016/0094-1190(92)90005-6

Middleton, F. (2022). The 4 Types of Validity in Research | Definitions & Examples. Scribbr. https://www.scribbr.com/methodology/types-of-validity/

Minitab (Red.). (2016, 4 november). What are T Values and P Values in Statistics? blog.minitab.com. https://blog.minitab.com/en/statistics-and-quality-data-analysis/what-are-t-values-and-p-values-in-statistics

MRC Cognition and Brain Sciences Unit. (2009). Testing normality including skewness and kurtosis. MRC CBU Wiki. Retrieved June 19, 2023, from https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/Simon

National Association of Realtors. (2011, April 7). *Factors Influencing Neighborhood Choice*. www.nar.realtor. Retrieved February 21, 2023, from https://www.nar.realtor/blogs/economists-outlook/factors-influencing-neighborhood-choice

NU.nl. (2023, February 22). *Huizenprijzen stijgen weer na vijf maanden van daling*. https://www.nu.nl/economie/6252236/huizenprijzen-stijgen-weer-na-vijf-maanden-van-daling.html

NUL20. (2012, July 4). *Ruim kwart corporatiehuurders verdient meer dan 33.000 euro | NUL20*. www.nul20.nl. Retrieved July 16, 2023, from https://www.nul20.nl/ruim-kwart-corporatiehuurders-verdient-meer-dan-33000-euro

Omroep Brabant. (2014, 22 maart). Honderden demonstranten verwacht bij protest in Eindhoven tegen uitspraken Geert Wilders.

https://www.omroepbrabant.nl/nieuws/1738885/honderden-demonstranten-verwacht-bij-protest-in-eindhoven-tegen-uitspraken-geert-wilders

Oxford Learner's Dictionaries. (2023a). *affordability noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com*. https://www.oxfordlearnersdictionaries.com/. Retrieved July 22, 2023, from https://www.oxfordlearnersdictionaries.com/definition/english/affordability

Oxford Learner's Dictionaries. (2023b). Definition of choice noun from the Oxford Advanced Learner's Dictionary. oxfordlearnersdictionaries.com. https://www.oxfordlearnersdictionaries.com/definition/english/choice_1?q=choice

Oxford Learner's Dictionaries. (2023c). Definition of homogeneity noun from the Oxford Advanced Learner's Dictionary. oxfordlearnersdictionary.com. https://www.oxfordlearnersdictionaries.com/definition/english/homogeneity?q=homogeneity

Princewill, U., Davidson, A., & Vesta, U. (2017). Determinants of Residential Neighbourhood Choice in a Nigerian Metropolis. *IOSR Journal of Humanities and Social Sciences*, *22*(7), e-ISSN: 2279-0837. https://doi.org/10.9790/0837-2207110111

Qadeer, M., & Kumar, S. (2006). Ethnic Enclaves and Social Cohesion. Canadian Journal of Urban Research, 15(2), 1–17. https://www.jstor.org/stable/26192499

Qadeer, M. A., Agrawal, S., & Lovell, A. (2010). Evolution of Ethnic Enclaves in the Toronto Metropolitan Area, 2001–2006. Journal of International Migration and Integration, 11(3), 315–339. https://doi.org/10.1007/s12134-010-0142-8

Raszkowski, A. (2012). The importance of city image in Socio-Economic development. *ResearchGate*. https://www.researchgate.net/publication/292134040_The_Importance_of_City_Image_in_Socio-Economic_Development

Reuter, T. K. (2023, 29 mei). Ethnic conflict | Definition, Causes, Types, Examples, & Facts. Encyclopedia Britannica. https://www.britannica.com/topic/ethnic-conflict#ref1226117

RTL Nieuws. (2021, November 15). *Achterstandswijk populairder bij starters: "Door hoge huizenprijzen."* rtlnieuws.nl. Retrieved July 15, 2023, from https://www.rtlnieuws.nl/editienl/artikel/5267236/achterstandswijk-populair-hoogopgeleiden-gentrificatie-woningnood

RTL Nieuws, Geeraedts, R., & de Goede, A. (2016, 25 mei). PVV-getuigen: Minder Marokkanen-uitspraak Wilders nauwkeurig voorbereid. RTL Nieuws. https://www.rtlnieuws.nl/nederland/politiek/artikel/569081/pvv-getuigen-minder-marokkanen-uitspraak-wilders-nauwkeurig

Sa, F. (2015). Immigration and House Prices in the UK. The Economic Journal, 125(587), 1393–1424. https://doi.org/10.1111/ecoj.12158

Schüller, S., & Chakraborty, T. (2016). Ethnic enclaves and immigrant economic integration. IZA world of labor, 287. https://doi.org/10.15185/izawol.287

Scott, A. (2001). Michel Foucault's The Archaeology of Knowledge. angelfire.com. Retrieved June 9, 2023, from https://www.angelfire.com/md2/timewarp/foucault.html

Scribbr. (2020, 24 september). What is homoscedasticity? https://www.scribbr.com/frequently-asked-questions/what-is-homoscedasticity/

Smartpls. (2023). How to Interpret Excess Kurtosis and Skewness - SmartPLS. Smartpls.com. Retrieved June 19, 2023, from https://www.smartpls.com/documentation/functionalities/excess-kurtosis-and-skewness/

Smith, G. A. (2015). Multiple Regression. Elsevier eBooks, 301–337. https://doi.org/10.1016/b978-0-12-803459-0.00010-8

Statistics Solutions. (2023). Assumptions of Linear Regression - Statistics Solutions. Retrieved June 19, 2023, from https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-linear-regression/

Statology. (2020, November 18). Understanding Heteroscedasticity in Regression Analysis. Statology.org. Retrieved June 19, 2023, from https://www.statology.org/heteroscedasticity-regression/

Steele, K., & Orri Stefánsson, H. (2020, 9 oktober). Decision Theory (Stanford Encyclopedia of Philosophy). https://plato.stanford.edu/entries/decision-theory/

Stevenson University. (2022, September 30). *SU Library: Qualitative vs. Quantitative Research: Overview*. Retrieved April 12, 2023, from https://stevenson.libguides.com/c.php?g=236343

Stonawski, M. J., Rogne, A. F., Bang, H., Christensen, H., & Lyngstad, T. H. (2019, January 23). Ethnic Segregation and Native Out-Migration in Copenhagen. https://doi.org/10.31235/osf.io/tx7b6

StudySmarter. (n.d.). Push Factors of Migration. StudySmarter.co.uk. Retrieved July 26, 2023, from https://www.studysmarter.co.uk/explanations/human-geography/population-geography/push-factors-of-migration/

Sun, K. (1995). APA PsycNet DoiLanding page. American Psychologist, 50(1), 44. https://doi.org/10.1037/0003-066X.50.1.43

Templeton, A. R. (2013). Biological races in humans. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44(3), 262–271. https://doi.org/10.1016/j.shpsc.2013.04.010

Troost, A. A., Van Ham, M., & Janssen, H. J. (2021). Modelling Neighbourhood Effects in Three Dutch Cities Controlling for Selection. Applied Spatial Analysis and Policy, 15(2), 455–482. https://doi.org/10.1007/s12061-021-09411-5

Turney, S. (2022). What Is Kurtosis? | Definition, Examples & Formula. Scribbr. https://www.scribbr.com/statistics/kurtosis/

Van der Velden, J., & Can, E. (2022, September 28). *Gemengde wijken herladen - Platform31 | Kennis en netwerk voor stad en regio*. Platform31 | Kennis En Netwerk Voor Stad En Regio. Retrieved July 15, 2023, from https://www.platform31.nl/artikelen/gemengde-wijken-herladen/

Van Overbeek Makelaars. (2022, 31 oktober). Is er verschil tussen de taxatiewaarde en WOZ-waarde? Geraadpleegd op 24 maart 2023, van https://www.vanoverbeek.nl/is-er-verschil-tussen-taxatiewaarde-en-woz-waarde/

Velzel, B. (2023, 17 april). Amsterdammers verlaten de stad door hoge huizenprijs en ontevredenheid over buurt. AT5.nl. Geraadpleegd op 8 juni 2023, van https://www.at5.nl/artikelen/220014/amsterdammers-vertrekken-door-dure-huizenprijs-en-ontevredenheid

Verwaaij, A. (2023, February 23). Waarom zijn huizen in Nederland zo duur? *NPO Kennis*. https://npokennis.nl/longread/7757/waarom-zijn-huizen-in-nederland-zo-duur

Wagner, N. (2023, 6 februari). Econometrics Lecture 2. Econometrics: Regression analysis of cross-sectional data, Nijmegen, Gelderland, Nederland.

Wallstreetmojo Team. (2022, 17 september). Heteroscedasticity. WallStreetMojo. https://www.wallstreetmojo.com/heteroscedasticity/

Weedmark, D. (2018, 13 maart). The Advantages & Disadvantages of a Multiple Regression Model. Sciencing. https://sciencing.com/advantages-disadvantages-multiple-regression-model-12070171.html

William, K. (2022, 23 september). What is SPSS? Definition, Features, Types, and Use Cases. SurveySparrow. https://surveysparrow.com/blog/what-is-spss/#section7

Wo, J. C. (2022). Neighborhood Effects on Crime: The Concentration of Racial/Ethnic Groups and the Heterogeneity Among Such Groups. Crime & Delinquency, 001112872211345. https://doi.org/10.1177/00111287221134592

Wood, D., & Airth, M. (2023, 27 mei). How to Improve Validity of a Scientific Investigation - Video & Lesson Transcript | Study.com. study.com. https://study.com/academy/lesson/how-to-improve-validity-of-a-scientific-investigation.html

Wu, Z., Hou, F., & Schimmele, C. M. (2011). Racial Diversity and Sense of Belonging in Urban Neighborhoods. City & Community, 10(4), 373–392. https://doi.org/10.1111/j.1540-6040.2011.01374.x

Zack, N. (Red.). (2017). The Oxford Handbook of Philosophy and Race [Ebook]. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190236953.001.0001

Zhang, J., Hui, E. C., & Sun, J. (2017). Population migration, urbanization and housing prices: Evidence from the cities in China. Habitat international, 66, 49–56. https://doi.org/10.1016/j.habitatint.2017.05.010

Zhou, J., & Hui, E. C. (2022). Housing prices, migration, and self-selection of migrants in China. Habitat international, 119, 102479. https://doi.org/10.1016/j.habitatint.2021.102479

Zwiers, M., Van Ham, M., & Manley, D. (2018, December 20). *Trajectories of ethnic neighborhood change: Spatial patterns of increasing ethnic diversity | A+BE | Architecture and the Built Environment*. https://journals.open.tudelft.nl/abe/article/view/3528

Appendix
Figures

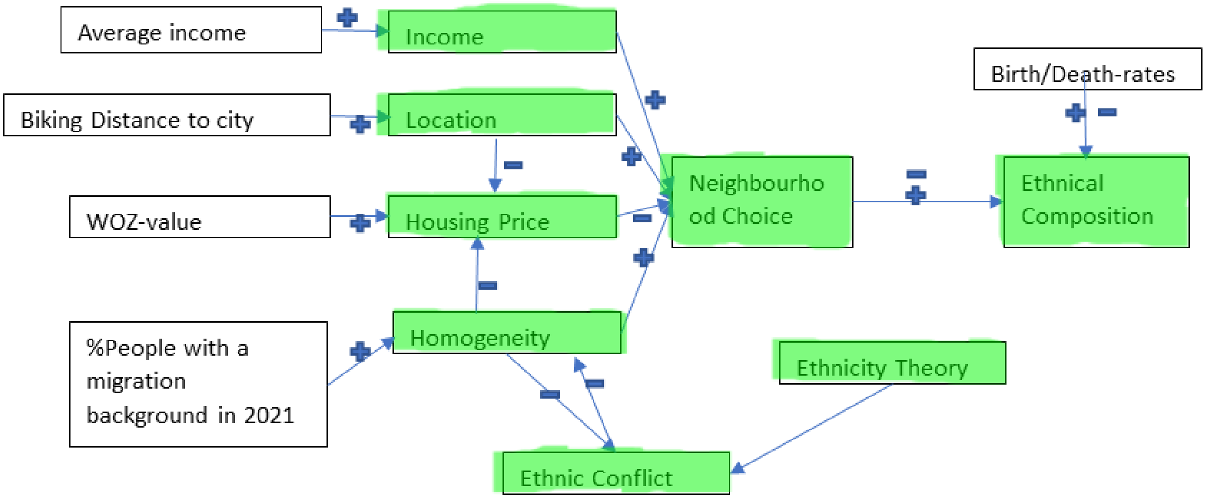

Figure 1: *Conceptual Model.*

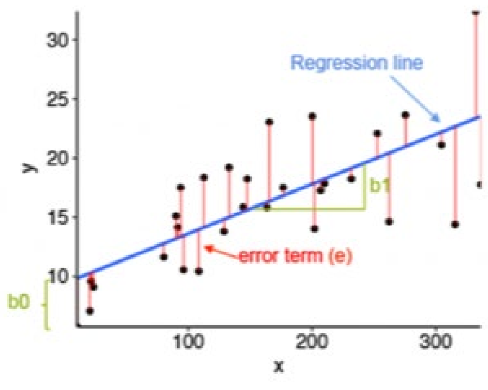

Figure 2: *Visualisation of the error-term* (Lakshmi, 2022)

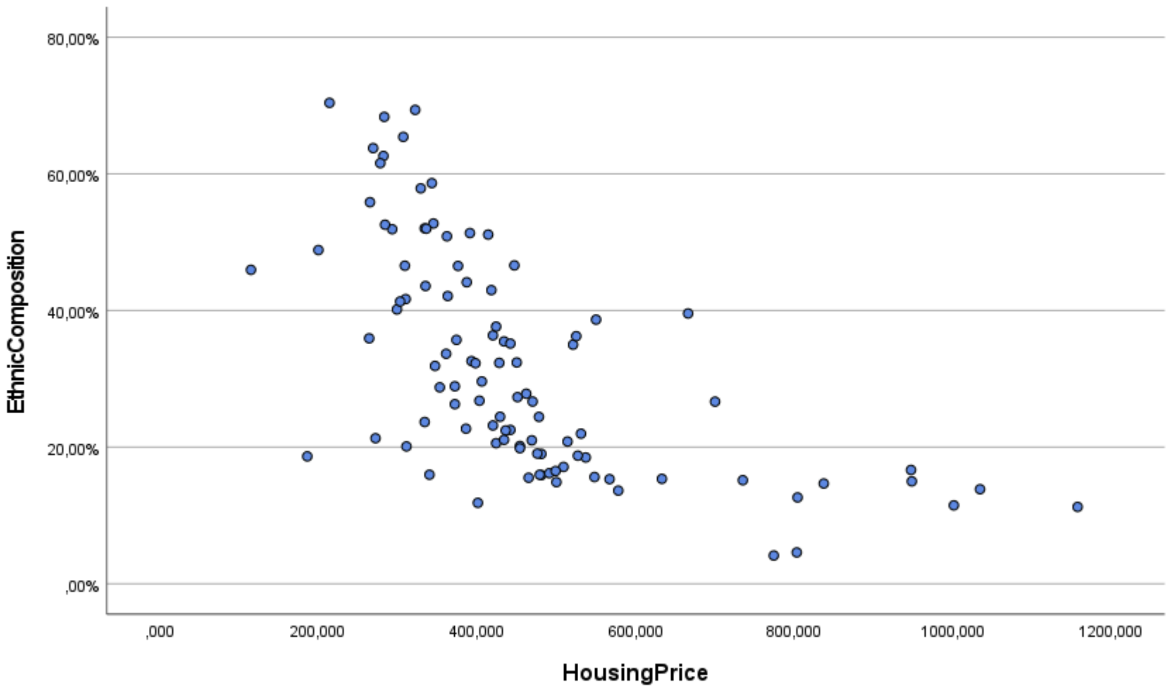

Figure 3: *Scatterplot to illustrate non-linear shape of data.*

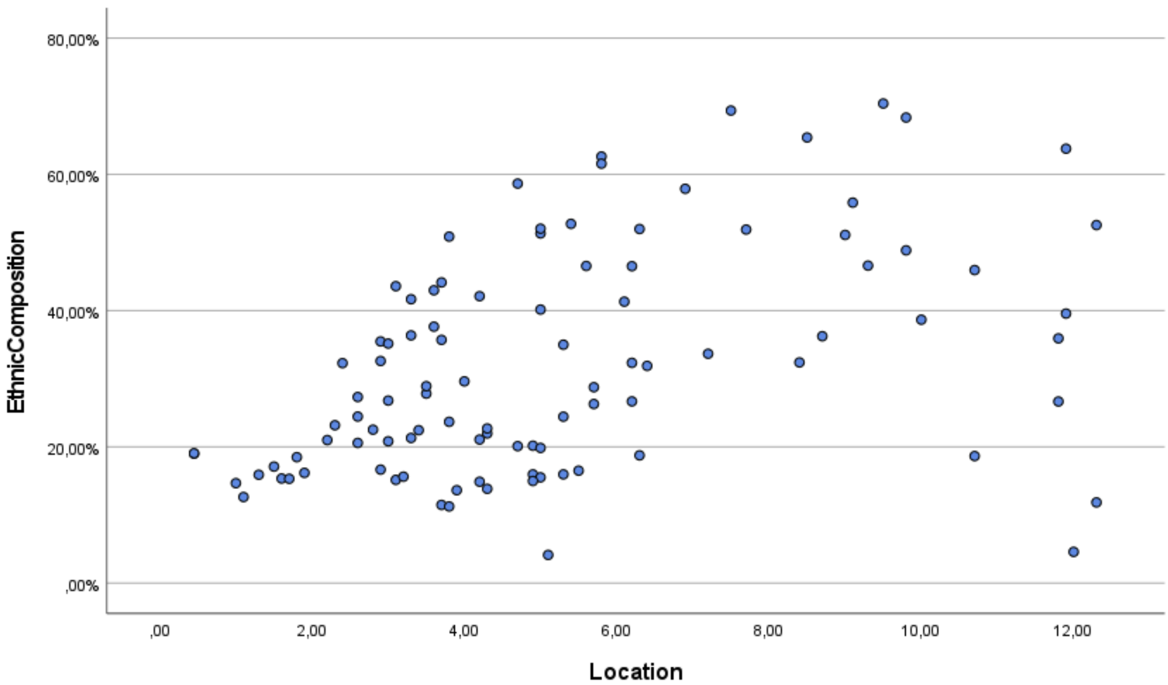

Figure 4: *Scatterplot to illustrate linearity between ethnic composition and location.*

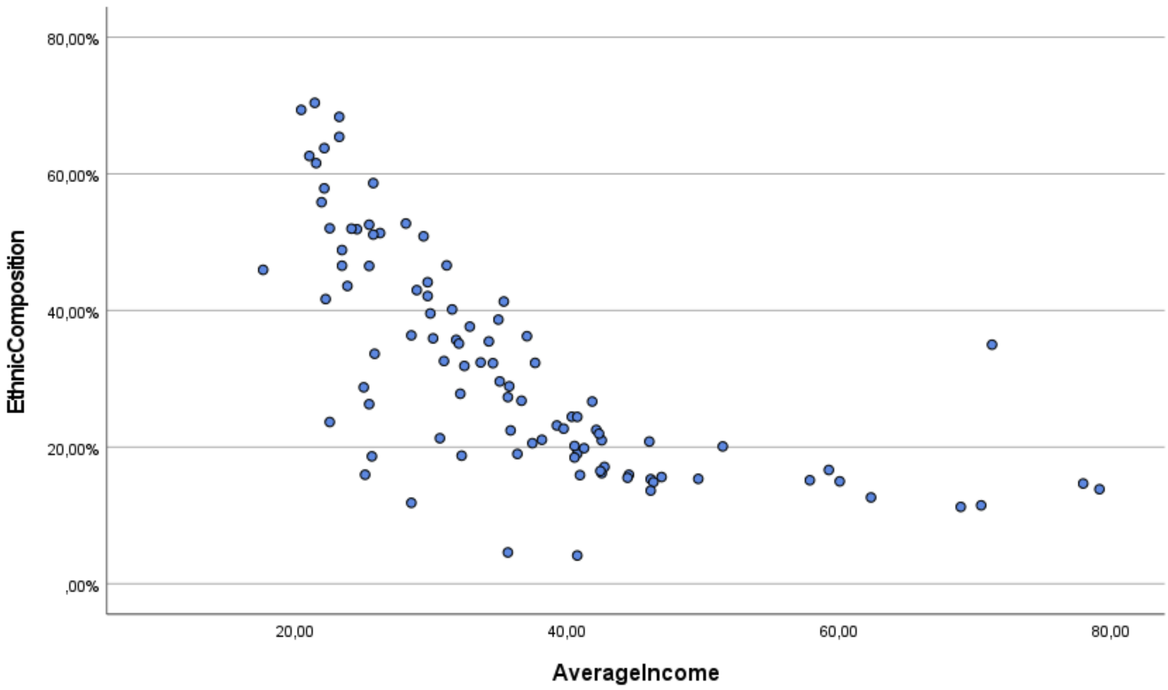

Figure 5: *Scatterplot to illustrate linearity between ethnic composition and average income.*

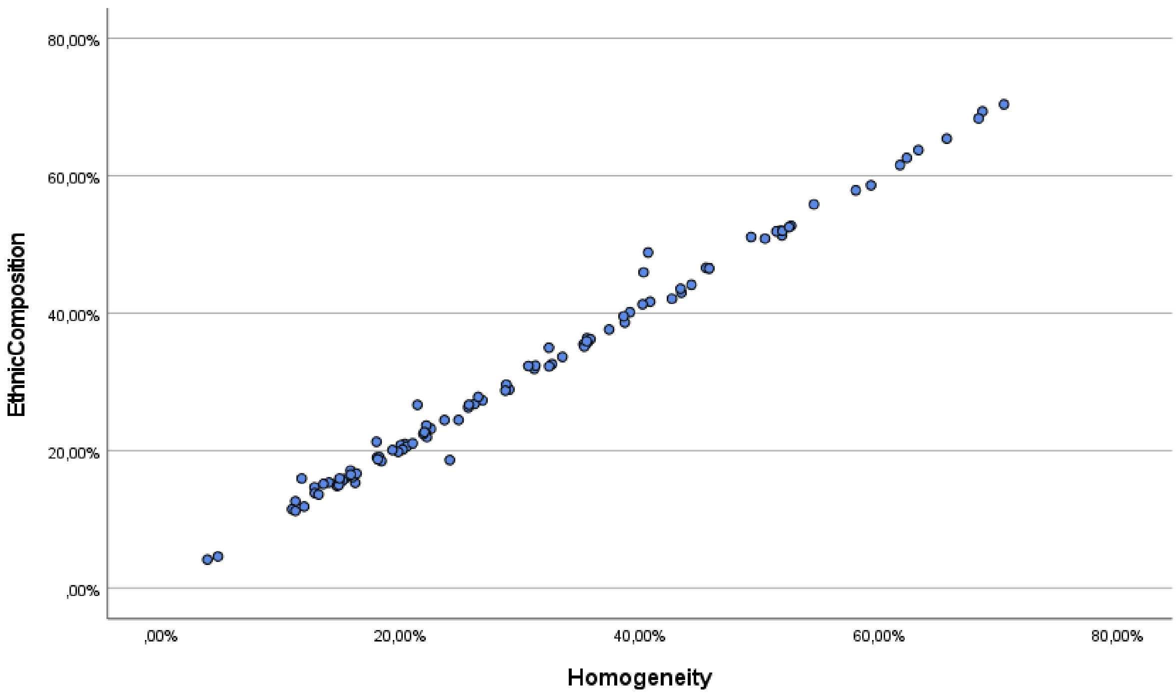

Figure 6: *scatterplot to visualise linearity between ethnic composition and homogeneity variables.*

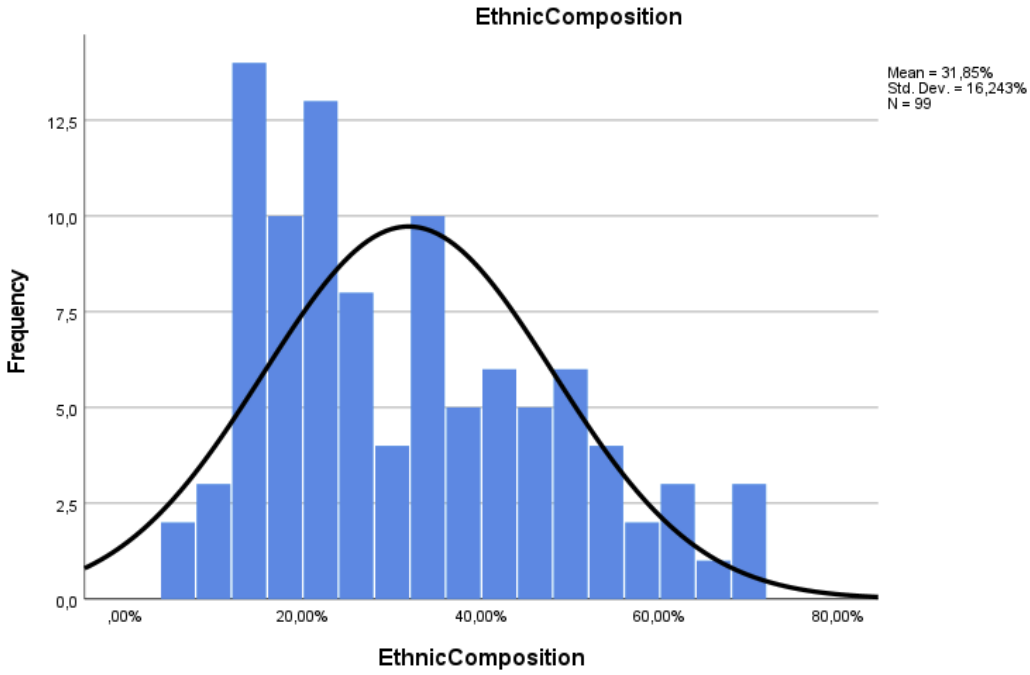

Figure 7: *Distribution of observations of the ethnic composition variable.*

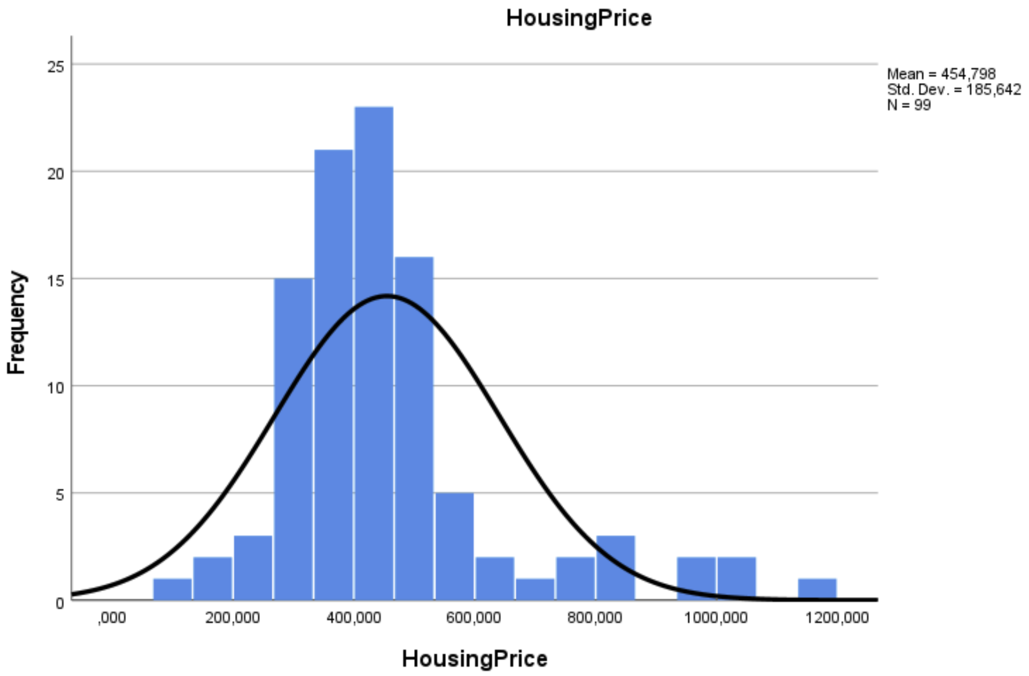

Figure 8: *Distribution of observations of the housing price variable.*

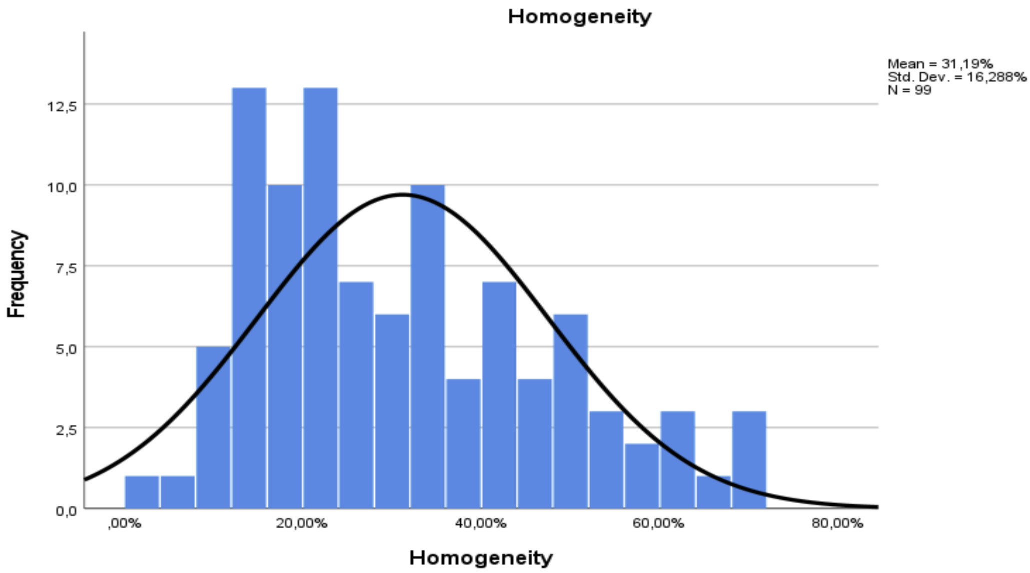

Figure 9: *Distribution of observations of the homogeneity variable.*

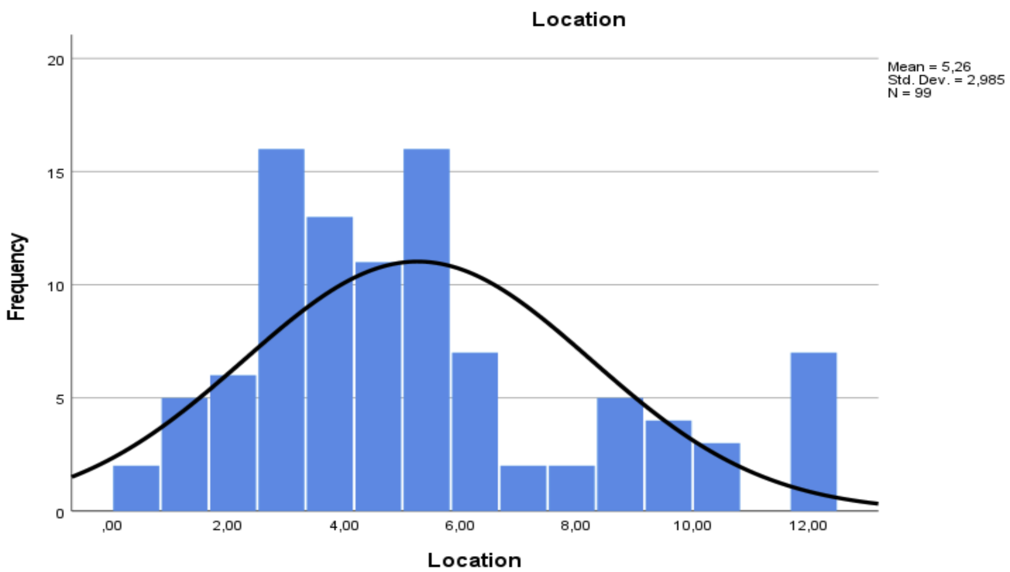

Figure 10: *Distribution of observations of the location variable.*

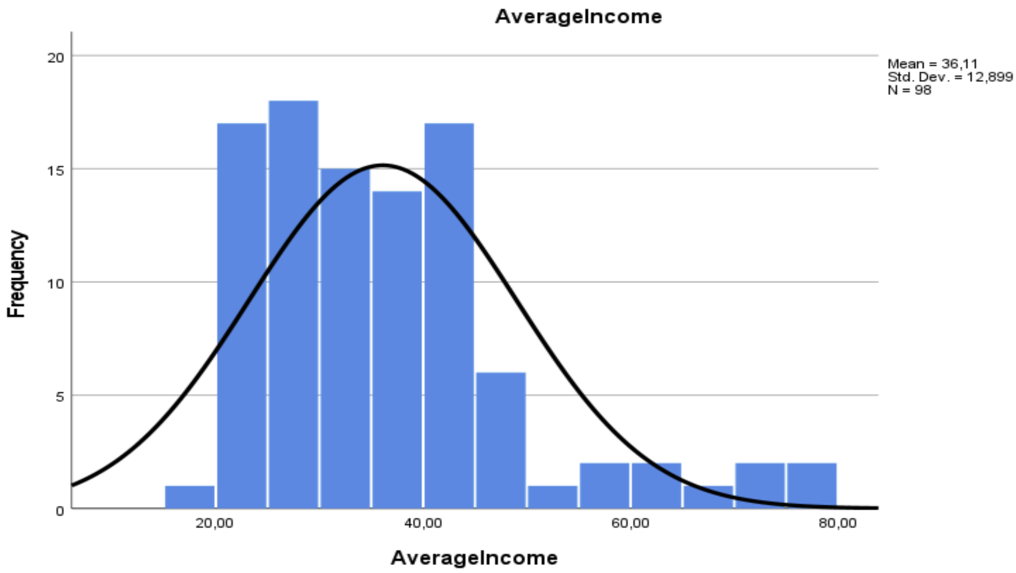

Figure 11: *Distribution of observations of the average income variable.*

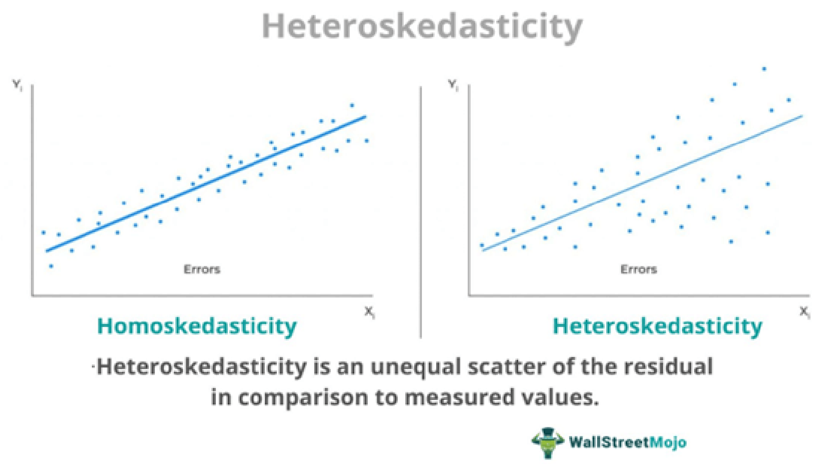

Figure 12: *Visualisation of Homoscedasticity and Heteroscedasticity* (Wallstreetmojo Team, 2022).

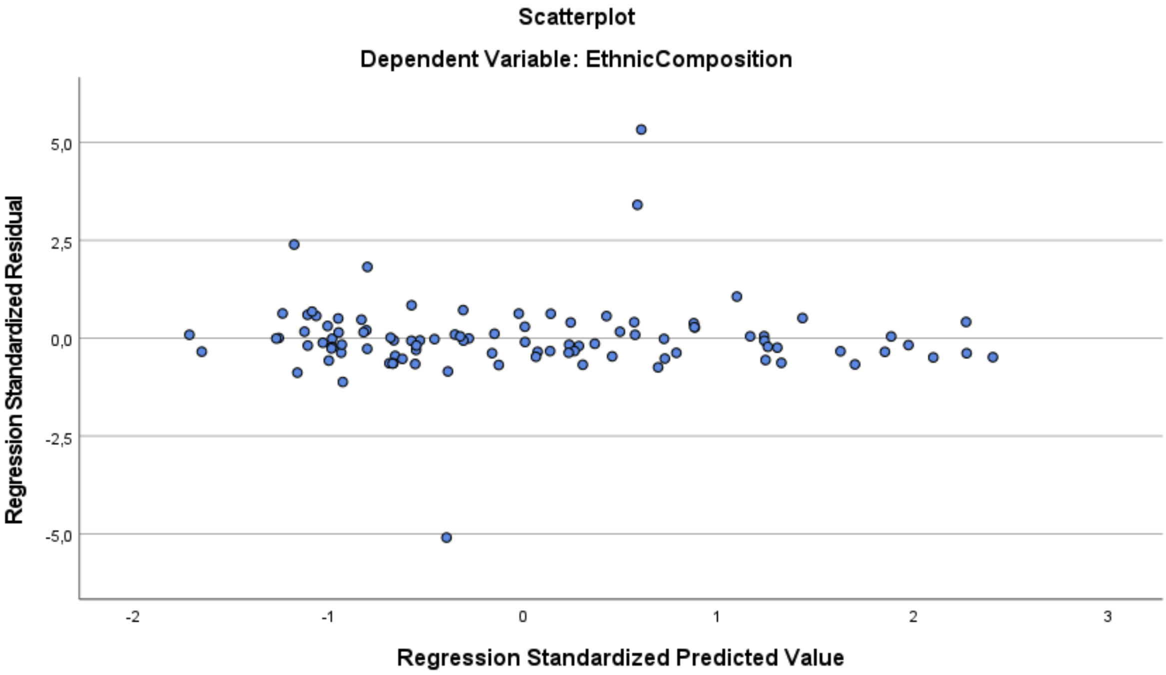

Figure 13: *Scatterplot in which the predicted values have been set out against the actual values.*

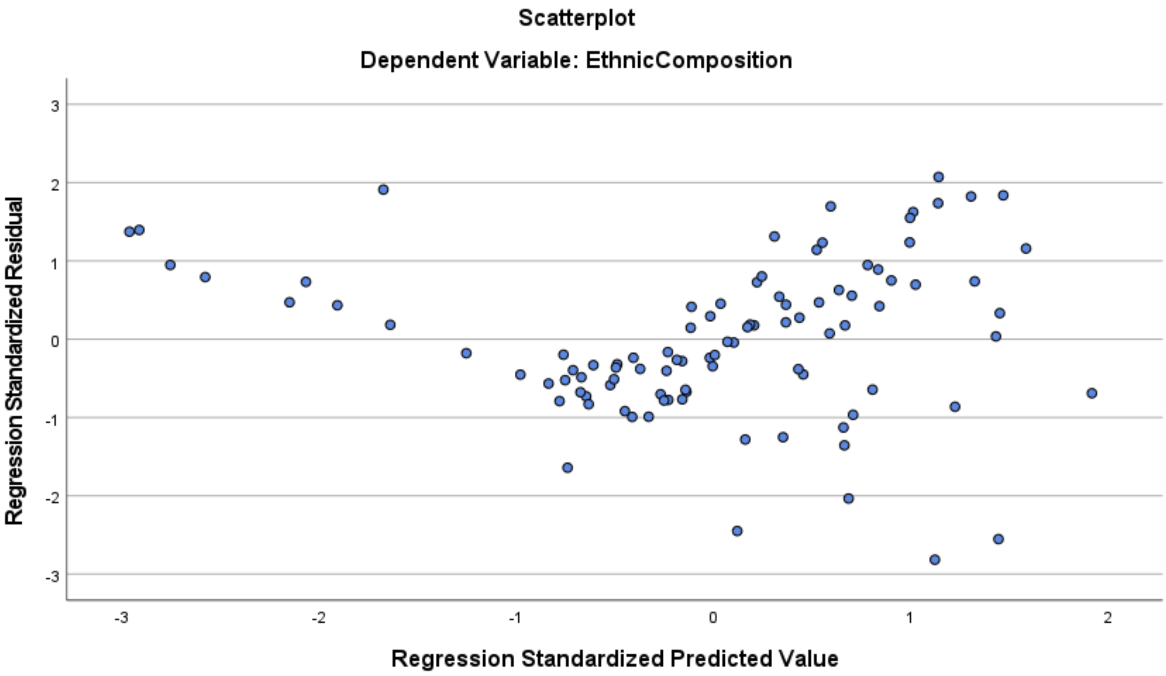

Figure 14: *Actual values and predicted values plotted against each other.*

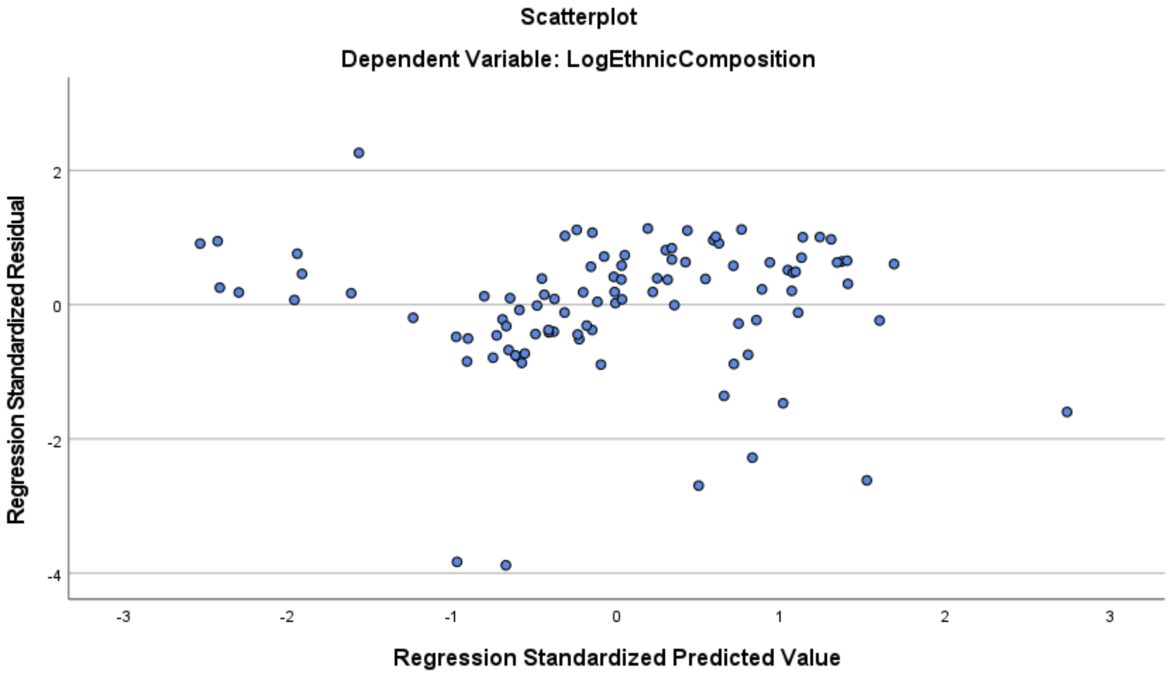

Figure 15: *Scatterplot to show homoscedasticity*

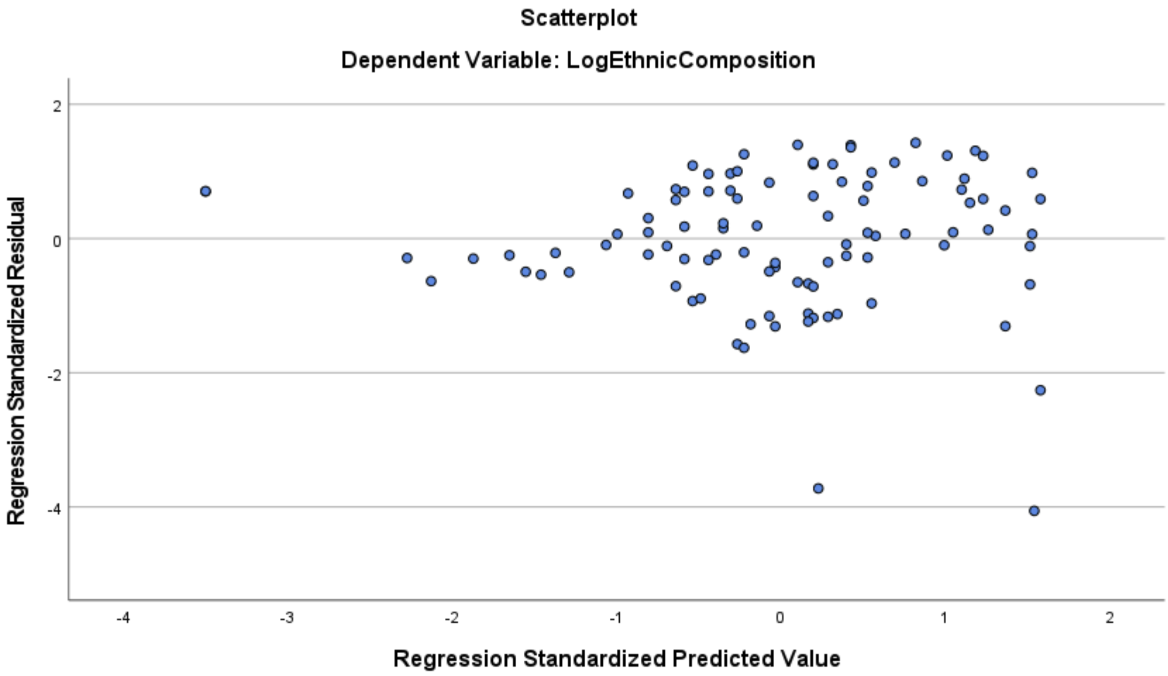

Figure 16: *Scatterplot of predicted values set out against actual values.*

Tables

		ADULT COMPOSITION OF HOUSEHOLD					CHILDREN IN HOME	
	All Buyers	Married couple	Single female	Single male	Unmarried couple	Other	Children under 18 in home	No children in home
Quality of the neighborhood	64%	67%	62%	57%	64%	58%	67%	63%
Convenient to job	49	47	51	49	58	41	49	49
Overall affordability of homes	44	42	47	49	47	46	43	45
Convenient to friends/family	39	36	48	37	38	36	34	41
Quality of the school district	25	33	15	11	24	18	48	14
Convenient to shopping	25	25	25	24	25	22	20	27
Design of neighborhood	24	27	20	21	22	22	23	25
Convenient to schools	19	24	12	9	17	16	43	7
Convenient to entertainment/ leisure activities	19	18	20	22	21	18	13	22
Convenient to parks/recre- ational facilities	17	18	14	17	15	11	19	15
Availability of larger lots or acreage	15	18	6	10	17	12	19	13
Convenient to health facilities	9	9	10	8	5	8	5	11
Convenient to public transportation	7	6	8	10	7	12	5	8
Home in a planned community	6	7	6	5	3	4	6	7
Convenient to airport	6	6	6	7	4	7	3	6
Green (environmentally friendly) community features	5	5	4	6	3	6	5	5
Other	5	5	6	6	4	13	5	6

Table 1: *Factors influencing neighbourhood choice* (National Association of Realtors, 2011).

	Wijkenenbuurten	EthnicComposition	HousingPrice	Homogeneity	Location	AverageIncome
1	Burgwallen-Oude Zijde	19,01%	481,000	18,03%	,45	36,30
2	Burgwallen-Nieuwe Zijde	19,06%	476,000	18,26%	,45	40,70
3	Grachtengordel-West	12,64%	804,000	11,23%	1,10	62,30
4	Grachtengordel-Zuid	14,68%	837,000	12,83%	1,00	77,90
5	Nieuwmarkt/Lastage	17,11%	509,000	15,84%	1,50	42,70

Table 2: *Snapshot of dataset in SPSS.*

Statistics

		EthnicComposition	HousingPrice	Homogeneity	Location	AverageIncome
N	Valid	99	99	99	99	98
	Missing	0	0	0	0	1
Mean		31,8458%	454,79765	31,1903%	5,2616	36,1051
Median		27,8243%	424,00000	26,8684%	4,7000	34,3500
Std. Deviation		16,24256%	185,642340	16,28847%	2,98498	12,89921
Skewness		,605	1,613	,634	,849	1,349
Std. Error of Skewness		,243	,243	,243	,243	,244
Kurtosis		-,559	3,101	-,492	-,006	2,000
Std. Error of Kurtosis		,481	,481	,481	,481	,483
Minimum		4,14%	115,000	3,89%	,45	17,60
Maximum		70,41%	1157,000	70,41%	12,30	79,10

Table 3: *Descriptive Statistics of All Variables.*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
EthnicComposition * HousingPrice	Between Groups	(Combined)	25028,055	90	278,090	2,692	,067
		Linearity	9758,688	1	9758,688	94,473	<,001
		Deviation from Linearity	15269,367	89	171,566	1,661	,226
	Within Groups		826,364	8	103,296		
	Total		25854,420	98			

Table 4: *Test for linearity between ethnic composition and housing price.*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
EthnicComposition * Location	Between Groups	(Combined)	18406,733	59	311,979	1,634	,053
		Linearity	4877,920	1	4877,920	25,543	<,001
		Deviation from Linearity	13528,813	58	233,255	1,221	,256
	Within Groups		7447,686	39	190,966		
	Total		25854,420	98			

Table 5: *Test for linearity between ethnic composition and location.*

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
EthnicComposition * AverageIncome	Between Groups	(Combined)	24197,736	82	295,094	2,716	,016
		Linearity	11848,513	1	11848,513	109,063	<,001
		Deviation from Linearity	12349,223	81	152,460	1,403	,235
	Within Groups		1629,586	15	108,639		
	Total		25827,323	97			

Table 6: *Test for linearity between ethnic composition and average income.*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,997[a]	,994	,993	1,33892%	2,436

a. Predictors: (Constant), AverageIncome, Location, Homogeneity, HousingPrice

b. Dependent Variable: EthnicComposition

Table 7: *Model Summary with explained variance (R-square).*

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	,371	,877		,423	,674		
	HousingPrice	-,003	,001	-,031	-1,997	,049	,294	3,396
	Homogeneity	,992	,012	,993	84,331	<,001	,500	1,999
	Location	,064	,054	,011	1,182	,240	,740	1,352
	AverageIncome	,038	,022	,030	1,782	,078	,240	4,167

a. Dependent Variable: EthnicComposition

Table 8: *Estimated coefficients table.*

Correlations

		EthnicComposition	HousingPrice	Homogeneity	Location	AverageIncome
Pearson Correlation	EthnicComposition	1,000	-,616	,997	,453	-,677
	HousingPrice	-,616	1,000	-,611	-,309	,833
	Homogeneity	,997	-,611	1,000	,449	-,681
	Location	,453	-,309	,449	1,000	-,450
	AverageIncome	-,677	,833	-,681	-,450	1,000
Sig. (1-tailed)	EthnicComposition	.	<,001	<,001	<,001	<,001
	HousingPrice	,000	.	,000	,001	,000
	Homogeneity	,000	,000	.	,000	,000
	Location	,000	,001	,000	.	,000
	AverageIncome	,000	,000	,000	,000	.
N	EthnicComposition	98	98	98	98	98
	HousingPrice	98	98	98	98	98
	Homogeneity	98	98	98	98	98
	Location	98	98	98	98	98
	AverageIncome	98	98	98	98	98

Table 9: *correlation matrix with mutual correlations.*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,732[a]	,536	,521	,16968

a. Predictors: (Constant), LogAverageIncome, LogLocation, LogHousingPrice

b. Dependent Variable: LogEthnicComposition

Table 10: *Model summary to show explained variance.*

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	3,771	,326		11,562	<,001		
	LogHousingPrice	-,452	,188	-,298	-2,405	,018	,322	3,102
	LogLocation	,043	,070	,050	,617	,538	,764	1,308
	LogAverageIncome	-,764	,225	-,442	-3,403	<,001	,293	3,415

a. Dependent Variable: LogEthnicComposition

Table 11: *Coefficient table to show estimated coefficients, t-values, and VIF-values.*

Correlations

		LogEthnicComposition	LogHousingPrice	LogLocation	LogAverageIncome
Pearson Correlation	LogEthnicComposition	1,000	-,681	,383	-,711
	LogHousingPrice	-,681	1,000	-,398	,823
	LogLocation	,383	-,398	1,000	-,485
	LogAverageIncome	-,711	,823	-,485	1,000
Sig. (1-tailed)	LogEthnicComposition	.	<,001	<,001	<,001
	LogHousingPrice	,000	.	,000	,000
	LogLocation	,000	,000	.	,000
	LogAverageIncome	,000	,000	,000	.
N	LogEthnicComposition	98	98	98	98
	LogHousingPrice	98	98	98	98
	LogLocation	98	98	98	98
	LogAverageIncome	98	98	98	98

Table 12: *Correlation matrix to show mutual correlations.*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,996[a]	,992	,992	1,42178%

a. Predictors: (Constant), Homogeneity

b. Dependent Variable: EthnicComposition

Table 13: *Model summary of simple linear regression with homogeneity as independent variable.*

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	,862	,310		2,780	,007		
	Homogeneity	,993	,009	,996	112,663	<,001	1,000	1,000

a. Dependent Variable: EthnicComposition

Table 14: *Coefficient table of simple linear regression with homogeneity as independent variable.*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,377[a]	,142	,133	,22709

a. Predictors: (Constant), LogLocation

b. Dependent Variable: LogEthnicComposition

Table 15: *Model summary of simple linear regression with location as independent variable.*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	1,232	,057		21,635	<,001		
	LogLocation	,325	,081	,377	4,011	<,001	1,000	1,000

a. Dependent Variable: LogEthnicComposition

Table 16: *Coefficients table of simple linear regression with location as independent variable.*