Thesis Final Report

The effect of the built environment on bicycle use as travel mode

J. Kelderman

5 June 2019
Thesis Final Report

Joep Kelderman
UU student number: 6307221
J.Kelderman@students.uu.nl

Supervisor: Dr. C. Maat
Professor: Prof.dr.ir. P.J.M. van Oosterom

Abstract

The vitality of human's living environments, especially in cities, is currently under pressure in terms of traffic density and traffic emission. Dutch National and local governments have initiated working programmes to reach sustainable mobility goals. Cycling in this sense is seen as a means of sustainable transport, which may contribute to less urban congestion, and safer and nicer living environments. Therefore it is important to know which factors influence bicycle use. In travel behaviour research, the built environment is seen as one of the main factors influencing travel mode choice. The density, diversity and design determine the distribution of an individual's activities (residence, work, shops, schools etc.), but also determine how a person's activity space and its infrastructure is constructed. This research aims to address the importance of the built environment in terms of travel mode choice, specifically for bicycle use. The design dimension of the built environment perceives special attention, since this dimension is not extensively researched yet. Multiple measures of the built environment's density, diversity and design are developed and calculated using GIS to indicate the cycling friendliness of a built environment. The results show that higher bicycle use rates are found in dense and diverse built environments. The design dimension poses more challenges to quantify, but the results carefully show that the infrastructural design does influence bicycle use. In general, bicycle use tends to be slightly higher in built environment's where the infrastructure provides a relatively shorter cycling route than motorized travel route.

Preface

This thesis in front of you has been written as part of the Master Programme Geographical Information Management and Applications of the University of Utrecht, Technical University of Delft, Wageningen University and the University of Twente.

The research took place from September 2018 till May 2019 under the supervision of Dr. Kees Maat, University of Delft. As with many thesis researches, unexpected challenges in the research process needed to be tackled. Working with (geographical) data poses challenges in terms of availability, quality and extensity of the data. In the end I am glad that I have been able to expand my knowledge in the field of travel behaviour and have enriched research in this field by means of using GIS.

I would like to thank Kees Maat and Peter van Oosterom for their supervision. Kees Maat has provided good insight into the needs of travel behaviour research, has guided me to perform a valuable analysis to examine bicycle us and delivered necessary feedback. I would also like to thank Mathijs de Haas, Netherlands Institute for Transport Policy Analysis for providing me the option to combine my generated spatial data with their travel dataset. Lastly, thanks to my family and friends who kept me motivated to at this moment hand over a thesis research to be proud of.

I hope you enjoy reading!

Joep Kelderman.

Abbreviations

BAG	Basisregistratie Adressen en Gebouwen	*Key Register Addresses and Buildings*
CBS	Centraal Bureau voor de Statistiek	*Statistics Netherlands*
KM^2	Vierkante Kilometer	*Square kilometre*
MPN	Mobiliteits Panel Nederland	*Netherlands Mobility Panel*
OAD	Omgevingsadressendichtheid	*Adress Density*
OSM		*OpenStreetMap*
OV	Overijssel	
PC4	Postcode 4	*4-digit postal codes*
$\chi 2$		*Chi-squared*
ZH	Zuid-Holland	
-2LL		*-2 Log Likelihood (deviance)*

Table of contents

1 Introduction 7

1.1 Context 7

1.2 Problem statement 9

1.3 Relevance 10

1.3.1 Research gap 10

1.3.2 Social relevance 10

1.4 Research questions 10

1.5 Research steps 11

1.6 Research scope 11

2 Literature review 13

2.1 Travel behaviour 13

2.2.1 What is travel behaviour? 13

2.2.2 Travel mode choice 13

2.2 Built environment 14

2.2.1 Density of a built environment 15

2.2.2 Diversity of a built environment 15

2.2.3 Design of a built environment 15

2.2.4 Previous research on bicycle use and built environment 16

2.3 Built environment measures 16

2.4 Conceptual model 17

3 Methodology 18

3.1 Methodology introduction 18

3.2 Research design scheme 19

3.3 Software 19

3.3.1 SPSS 19

3.3.2 ArcMap 10.5 20

3.4 Datasets 20

3.4.1 Travel behaviour dataset 20

3.4.2 Built environment datasets 20

3.5 Variable operationalization 22

3.5.1 Bicycle use operationalization 22

3.5.2 Control variables operationalization 22

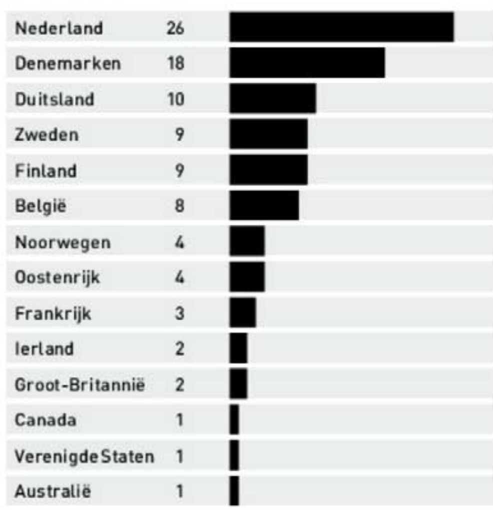
3.5.3 Built environment operationalisation 23
3.6 Variable calculations 25
3.6.1 Built environment measures 25
3.6.2 Bicycle use measure 31
3.7 Statistical analysis 32
3.7.1 Variables 32
3.7.1 Univariate analysis 33
3.7.2 Bivariate analysis 33
3.7.3 Multivariate analysis 33
4 Analysis 36
4.1 Analysis and data sample introduction 36
4.2 Univariate analysis 36
4.3 Bivariate analysis 48
4.3.1 Control variables 48
4.3.2 Built environment variables 49
4.3.3 Correlation matrix 51
4.4 Multivariate analysis: logistic regression 52
4.4.1 Regression without predictors 52
4.4.2 Regression model 1 52
4.4.3 Regression model 2 52
4.4.4 Regression model 3 54
4.4.5 Model performance 55
5 Conclusion 56
6 Discussion 58
7 Acknowledgements 59
8 References 60
9 Appendices 63
Appendix A: Operationalisation formulas of each shortest route ratio variable: 63
Appendix B: Model Builder explanation 63
Appendix C: SPSS syntaxes 66
Appendix D: Extended methodology for multivariate analysis 67
Appendix E: Bivariate analysis 71

1 Introduction

1.1 Context

In the Netherlands, cycling is a very important means of transport in the daily lives of its citizens. For some countries, cycling is mainly considered as a tool for leisure purposes, whereas in the Netherlands cycling also is an important travel mode for daily trips (Rietveld & Daniel, 2004). Figure 1 shows that 26% of all the trips in the Netherlands are made by bicycle. In comparison with other countries such as France (3%), the United Kingdom (2%) or the United States (1%), this is remarkably high. Even though it might sound like cycling numbers are sufficient, cycling could still provide opportunities in the Netherlands as Klinkenberg & Bertolini (2013) argue. They have identified several challenges where bicycle policies could provide a solution to: congestion in cities and a decline of bicycle use in rural areas. For cities it is important to keep being accessible and reachable, whereas for both cities and rural areas it is important to keep cycling rates at a certain level to avoid a growth of motorized travel. For both challenges cycling supports sustainable mobility opportunities, especially taking into account the space for improvement when it comes to short distance trips. According to Harms & Kansen (2018), more than half of all trips by car in the Netherlands cover less than 7.5 kilometres. However, this distance could be covered by bicycle in approximately 30 minutes. With the introduction of electric bicycles for the wider public several years ago, the potential range of cyclists in time and distance has even increased.

Figure 1: Share of trips made by bicycle (Pucher & Buehler, 2012).

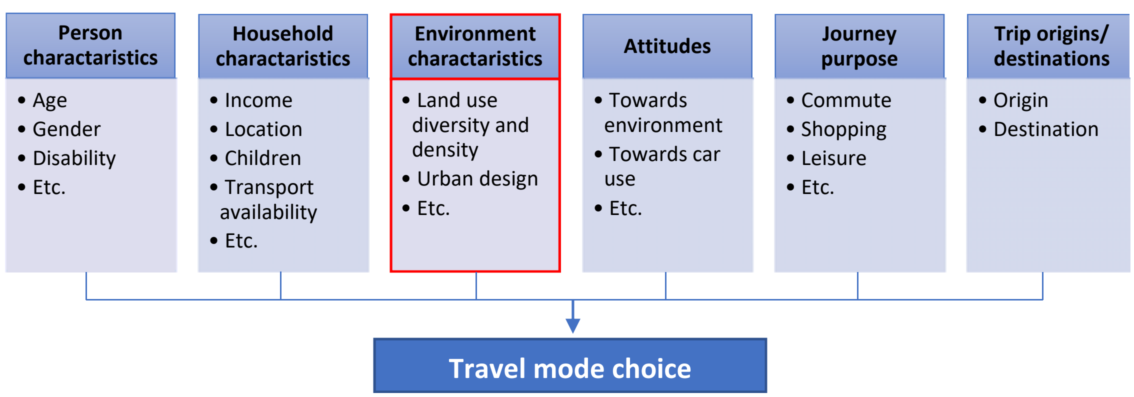


For several years now, the Dutch Government is working on making the Dutch infrastructure more bicycle friendly, since they are aiming to increase bicycle use rates in the next decade. They have identified three main reasons why bicycling as travel mode is beneficial in comparison with car use (Rijksoverheid, 2018). First of all, cycling provides many health benefits to its citizens. Secondly, an increasing rate of bicycle use in commute transport will reduce the amount of cars during rush hours, thus will help reduce traffic jams. Thirdly, bicycles do not require any fuel and therefore do not produce emission while using. It is a sustainable form of transportation.

Mainly these three reasons has the Dutch Government made set a specific mobility goal: in the period 2017-2027, the total number of cycling kilometres in the Netherlands should grow with 20% (Rijksoverheid, 2016). Government organizations, private companies, social organizations and research institutions have combined forces to reach this goal in an alliance named ‘Tour de Force’. Together, they have developed societal ambitions on three scale levels: nationwide, for cities and for rural areas. Vitality is the main focus of this mobility goal for all three scale levels. First of all, cycling has many health benefits for citizens, which will keep them healthy and vital. It has been proven that it has positive effects on obesity, heart- and vascular diseases and depressions (Rijksoverheid, 2016). The vitality of cities and rural areas itself will also increase, because cycling keeps cities sustainable and rural areas accessible (Rijksoverheid, 2016). Cycling is often considered as the key to healthier, calmer, cleaner, and greener cities, offering a more attractive living environment.

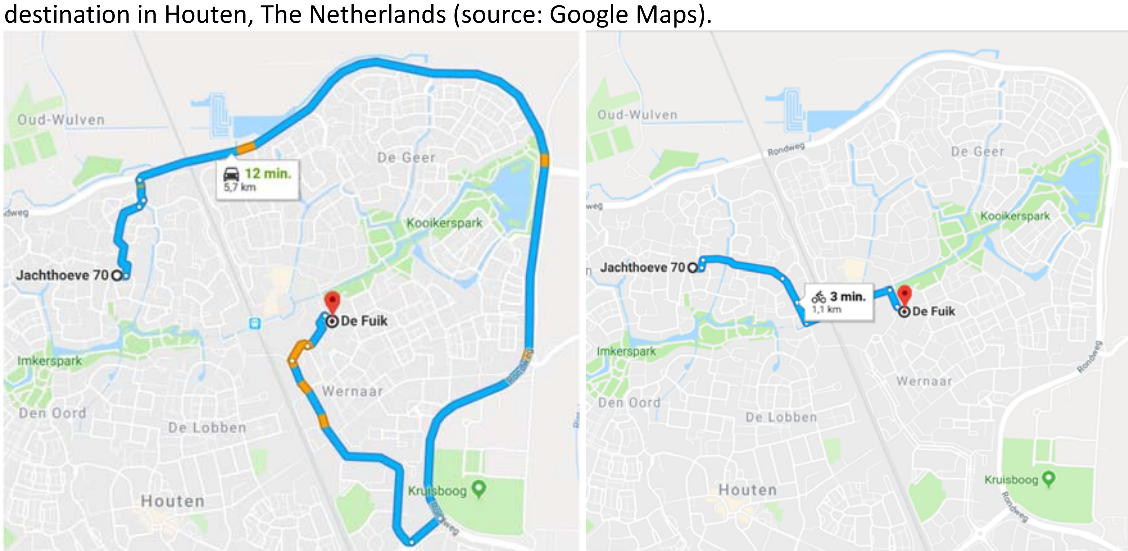
Whether an individual takes a bicycle is determined by several factors which are researched in travel mode choice behaviour studies. Travel mode choice behaviour can be explained through several determinants. Stradling (2011) explains travel mode choice behaviour through personal characteristics, household characteristics, environment characteristics, attitudes, journey purposes and trip origins/destinations (Figure 2). Environment characteristics in this regard are considered as the physical environment where the trips take place, in other articles regularly referred to as the built environment (Cervero & Kockelman, 1997; Ewing & Cervero, 2010; Handy, Boarnet, Ewing, & Killingsworth, 2002; Lu, Xiao, & Ye, 2017; Mccormack & Shiell, 2011).

Figure 2: Determinants of travel mode choice behaviour according to Stradling (2011).



The built environment can simply be defined as the physical environment created for human activities, including the designed buildings, infrastructure and land uses (Handy et al., 2002). In many studies on the subject of the built environment and travel behaviour, the built environment is explained through three dimensions: density, diversity and design (Cervero & Kockelman, 1997; Ewing & Cervero, 2010; Lu et al., 2017). To a certain extent, characteristics of these dimensions tend influence travel mode choices. Cervero & Kockelman (1997) for example proved that compact, mixed-use and pedestrian-friendly designs to a certain degree degenerate trips made by car. Previous research on the built environment and travel behaviour mainly has involved attributes such as travel time, cost and utility functions, but Rodriguez & Joo (2004) have indicated the relevance of including the local built environment in travel behaviour research. Higher percentages of sidewalk available for example is correlated with a higher propensity of walking as travel mode choice in their research. Mccormack & Shiell (2011) also have proven a correlation between cycle-friendly urban designs and bicycle use. One of the best Dutch examples of the influence of cycle-friendly urban designs is the municipality of Houten. This city is planned with the intention to ensure the cycling-friendliness (Maat, 2010). It consist of a small city centre, which is encircled by separate neighbourhoods. A ring road around the city connects the neighbourhoods for motorized travel, since the infrastructure does not offer direct connections by cars between neighbourhoods. A dense network of bicycle paths however does provide cyclists inter-neighbourhood travel. Figure 3 perfectly illustrates the difference in both distance and time between the suggested route by car and by bicycle for a trip from neighbourhood 'De Weerwolf' to neighbourhood 'Wernaar'.

Figure 3: Difference between car (12 min.) and bicycle (3 min.) travel for the same trip origin and destination in Houten, The Netherlands (source: Google Maps).



1.2 Problem statement

The choice to take a bicycle as travel mode is expected to be associated with the built environment. The physical infrastructure and design of the built environment in this case is of high importance, since it determines an individual's activity space (Cervero & Kockelman, 1997). Several researchers have argued that a compact, thus dense and mixed, built environment provides shorter distances to work, schools, services etc.. High density as well tends to encompass a dense public transport network and higher rates of traffic congestion and parking problems. The density of an individual's built environment therefore is expected to affect their travel mode choice. Until now, the majority of scientific research has focused on the relationship between the built environment and mode choice in general, with much emphasis on car use. Several researches on bicycle use and to a certain extent built environment can be identified, but these researches do not aim specifically at the relationship between the built environment and bicycle mode choice, except for two. Winters, Brauer, Setton, & Teschke (2010) found that the built environment does influence healthy travel choices (e.g. bicycle). Moudon et al. (2005) researched the bicycle use and built environment relationship as well, but did not find significant effects. Both findings however were based on telephone interviews where respondents had to provide origin, destination and mode choice of two of their most common non-recreative trips. Other studies focussed on a slightly other topic such as Mccormack & Shiell (2011) for example, who examined the relationship between the built environment and physical activity, of which one is cycling. Hull & O'Holleran (2014) and Caulfield, Brick, & McCarthy (2012) assessed the infrastructure preferences of cyclists and its effect on bicycle use. However, these researches did not validate their findings with real observed travel data. Therefore a valuable addition to these already existing theories on travel mode choice would be to examine the relationship between the built environment and bicycle use specifically by making use of actual travel observations. It will provide insight in to what extent built environment characteristics are of influence on bicycle use. Embracing this knowledge, local policy efforts could be more effectively adjusted to assist encouraging bicycle use.

1.3 Relevance

1.3.1 Research gap

Quite extensive research has already been carried out on the relationship between the built environment and travel behaviour. As already mentioned, many of these researches have put their focus on the infrastructure and mobility in general, proving a lower degree of car use in more dense areas for example. These results are very valuable, but do not provide specific conclusions or recommendations on how to reduce car use for example. In scientific literature, the particular relationship between bicycle use and the built environment is slightly researched, but research findings have not been validated using actual travel data. This research will also incorporate various variables indicating whether a built environment provides actual shorter cycling routes than routes for individual motorized travel modes. Proving this relationship might be very relevant. Take for example urban development policies, which could be designed taking into account the relationship between the built environment and bicycle use. In this way, (local) governments could plan cities in such a way that it encourages its citizens to make use of a bicycle.

1.3.2 Social relevance

In principle, this research aims to examine the relationship between the built environment and bicycle use so that the results can be taken into account in local policies to promote cycling. The research in this way has several societal benefits. First of all, promoting bicycle use is healthy for citizens. As explained in the context section, cycling has positive effects on obesity, heart- and vascular diseases, and depressions and therefore is beneficial for the health of citizens (Rijksoverheid, 2016). Cycling also provides a space-efficient mode of travel in comparison with car use. Increasing bicycle use at the same time degenerates other modes of travel aiming at motorized travel. Reducing the share of motorized travel modes is beneficial for the environment, reducing side effects such as air and noise pollution and traffic congestion (Vandenbulcke et al., 2009). In addition to this, calmer, cleaner and greener cities offer a more attractive urban environment to live in as explained by Rietveld & Daniel (2004) and Hellmann (2016).

1.4 Research questions

As explained, a lack of research on specifically the relationship between the built environment and bicycle use supports further research on this topic. The goal of the research therefore is to provide insight in the built environment characteristics that influence bicycle use, so that local governments are able to take this into account when creating policies to promote cycling. In order to meet this goal, the main research question has been established:

To what extent is bicycle use affected by the built environment?

To be able to carry out this research in a structured manner the main research question has been split up into several sub research questions:

1. How can bicycle use be defined in this research?
2. How can the built environment be defined in this research and what measures can represent cycling friendliness?
3. How to derive built environment measures from spatial datasets using Geographic Information Systems (GIS)?
4. How to integrate both bicycle use and built environment data in a statistical model?
5. What is the outcome of the model and what does it say about the built environment and bicycle use?

1.5 Research steps

The main objective of this research is to find out what the specific relation is between bicycle use and the built environment. The steps to reach this objective that have been identified are:

- to quantify the cycling friendliness of the built environment, resulting in built environment measures
- to derive built environment variables from spatial datasets using GIS
- to perform a statistical (regression) analysis on bicycle use and built environment variables
- to clarify the influence of built environment factors on bicycle use

In short, to be able to perform a statistical regression analysis, variation in built environment measures and variation in bicycle behaviour need to be calculated. The travel behaviour data is aggregated to the scale of four digit postal codes (in short PC4) in the Netherlands. Therefore the built environment measures will be measured for every postal code area in the research.

1.6 Research scope

This research will focus on exploring the relationship between the built environment and bicycle use. Focusing specifically on the mode choice for cycling, this research will not explain specific use of other modes of travel (e.g. walking, public transport, car, etc.). Even though motorized travel routes are included in the research, this is purely to explain the difference in motorized travel and bicycle routes. Aiming at travel mode choices explicitly, this research also does not examine other types of travel behaviour, such as journey purpose, route choice, etc. Even though built environment characteristics are researched, not all built environment characteristics can be included in the research. A limited set of characteristics will be chosen by means of a literature review to keep the research within boundaries.

The geographic scope of this research is limited to two provinces of the Netherlands: Zuid-

Holland and Overijssel. These provinces are taken as a sample of the Netherlands. Zuid-Holland is a relatively dense province, with multiple big cities as Rotterdam, Den Haag and Leiden. On the other hand there is Overijssel, with an almost four times lower population density as Zuid-Holland (Centraal Bureau voor de Statistiek, 2019)

For this research a deliberate decision has been made to relate a journey to the residential built environment of a respondent. This because the majority of journeys is made either from or to the home address of the respondent, thus a respondent's activity space. The Dutch travel behaviour dataset used in this research is available on the scale of two digit, four digit and six digit postal codes. Two digit postal code areas can be substantially large, where the mix in dense and not dense areas is too significant to show sufficient spatial variation in built environment measures. On the other hand, six digit postal codes areas are too small to calculate reasonable built environment measures for. In such small postal code areas could only be one road for example, making road network calculations for the built environment measures useless. The intermediate scale, four digit postal code areas, therefore are taken as the scale to work with in this research. As Rietveld & Daniel (2004) state, a viable spatial variation in bicycle use between municipalities does exist.

To conclude, bicycle use is location specific as explained (through built environment characteristics). Since this research is conducted on Dutch data and cultures of cycling in most countries are not comparable to the Netherlands, this research is assumed to be less representative for other countries.

2 Literature review

This chapter introduces and defines the most important theoretical concepts that are related to bicycle use as travel mode. This theoretical framework is designed to narrow down the relevant concepts to establish a set of variables that will be part of the actual analysis. It will result in a conceptual model that describes relation between variables in the analysis as well.

2.1 Travel behaviour

2.2.1 What is travel behaviour?

Locations of activities such as living, working, shopping and recreating are often spatially separated (Van Acker, Van Wee, & Witlox, 2010). To be able to carry out these activities, individuals need to cover a distance to get from one activity to another. This physical movement of people outside of their reference locations (home) for any purpose is researched in travel behaviour research studies (Axhausen, 2007). This research field consists of a wide array of subjects that are in principle related to travel behaviour. Studying travel behaviour gives insight into choices that individuals and households make as regards to their daily travel (Clifton & Handy, 2001). The choices that people make are displayed in their travel behaviour, think about trip frequency, journey purpose, route choice, trip accompany and travel mode choice for example. Gärling (2004) for example demonstrated the travel choices that people make as follows:

1) activity choice (what shall I do?)
2) destination choice (where shall I do it?)
3) mode choice (how shall I get there?)
4) departure time choice (when shall I go?)

These choices collectively determine an individual's travel behaviour. Since this research will solely focus on the travel mode choice, this dimension specifically will be further explored in the next section.

2.2.2 Travel mode choice

As one of the travel behaviour dimensions, travel mode choice presents the type of travel mode that is chosen to travel from an origin to a destination. A simple division of travel modes can be made by distinguishing motorized and non-motorized travel modes. Take for example car, motorcycle, bus, train, metro, tram, plane as motorized travel and walking, cycling, skating, stepping as non-motorized travel modes. As Stradling (2011) states, the direct impacts of motorized transport puts pressure on the environment through for example global warming and production waste. To preserve the planet's liveability, "*travel choices need to be smarter choices*" (Stradling, 2011, p. 487). By encouraging environmentally friendly, sustainable transport, the harm to the environment could be reduced. Not only do people's attitudes towards certain modes of transport influence their travel choices. In general, travel mode choices of individuals are considered to be influenced by mainly the personal characteristics, household characteristics, environment characteristics, attitudes, journey purposes and trip origins/destinations, as shown in Figure 2 (Stradling, 2011).

The personal (socioeconomic) characteristics, such as age, gender, income, and disability determine a person's travel mode possibilities (Stradling, 2011). Suffering from a disability may make it impossible to walk or cycle, obliging a person to make use of other convenient types of transport. Next, household characteristics affect travel mode choice as well. Think about the household income, number and age of children, the location and transport availability nearby. Van Acker et al. (2010) for example illustrate the situation where individuals prefer to bring and pick up their children from school by car, because this is more convenient than with public transport. This may have to do with the location of the household and the transport availability nearby. Attitudes refer to a positive, negative or mixed evaluative response to some issue, in this case as regards to travel modes (Van Acker et al., 2010). Attitudes indirectly influence behaviour. As an example, they give an individual's perception of cycling as healthy, environment-friendly, etc. which adapts a positive attitude of that individual toward cycling. Next, journey purposes may limit the choice of travel mode, for instance if big or heavy objects need to be carried during the trip. Stradling (2011) also considers trip origins and destinations as factors influencing travel mode. In principal do the trip origin and destination determine the distance that need to be covered and the available types of travel that are located at these locations. Lastly, travel mode choices are influenced by environment characteristics. In travel behaviour research, this is often referred to as the built environment, which can in short be defined as the physical environment created for human activities, including the designed buildings, infrastructure and land uses (Handy et al., 2002). The built environment influences travel mode choice in the sense that some built environment characteristics provide more amenities to a certain travel mode than other travel modes. High density areas, where the mix of land uses is significant for example are more accommodating to travel by foot or bicycle than low density areas (Handy et al., 2002).

Of all these determinants this research will specifically focus on the influence of built environment characteristics on bicycle use, therefore the next section (2.2) will elaborate more on built environment characteristics theories and concepts.

2.2 Built environment

The built environment can be considered as a concept that is open for interpretation. The previous section introduced a simplified definition of the built environment mentioned by (Handy et al., 2002), but almost every article on the topic of built environment specifies the built environment a little different. In general in most researches density, diversity and design are included as important aspects of the built environment, often referred to as the 3d's (Cervero, 2002; Cervero & Kockelman, 1997; Lu et al., 2017). Density represents the density of an area in terms of population or jobs for example, where diversity is illustrated by the diversity of land uses. The design relates to how the physical infrastructure (streets, paths, etc.) is designed. These three aspects of the built environment will be further explained later in this section. Some authors do mention other aspects as well, such as Chen, Gong, & Paaswell (2008) who included access to mass transit stations in their analysis which turned out to be less important than density for example. There are also some frameworks that have a rather micro-economical perspective, where they assume that travel behaviour is influenced by the built environment through utility maximization and travel cost minimalization of individuals (Boarnet & Crane, 2001; Cao, Mokhtarian, & Handy, 2009). However, Van Acker et al. (2010) argue that travel (mode) choices are not fully well-reasoned but are also subject to for example the earlier mentions attitudes towards modes of travel.

2.2.1 Density of a built environment

Density represents the compactness of the built environment, in other words how dense an area is. Density can for example be displayed in terms of population density: how many people per square kilometre do live in an area? Density can also be presented by the household density, dwelling density, job density or building floor area (Ewing & Cervero, 2010). In general it is assumed that density is an important aspect in regards to walking and cycling, since areas with higher neighbourhood density have more possible destinations within reach (e.g. retail, parks, jobs, etc.) (Lu et al., 2017). This has to do with a wider mix of land uses, meaning a wider mix of functions or services within reach. Once distances are relatively small, individuals have more opportunities to travel by foot or by bicycle. Higher densities also are related to less parking facilities and a better quality of public transport, indirectly discouraging car use and therefore encouraging walking or cycling (Cervero & Kockelman, 1997).

2.2.2 Diversity of a built environment

Diversity is directly related to the variety or mix of land uses, which measures the types of land uses within an area and thus the diversity of this area. As (Handy et al., 2002, p. 65) state: '*"Land use" typically refers to the distribution of activities across space, including the location and density of different activities, where activities are grouped into relatively coarse categories, such as residential, commercial, office, industrial, and other activities'.* Land can for example be used for living, working, shopping, and recreational purposes (Lu et al., 2017). As mentioned above, a greater mix of land uses makes the availability of services within reach higher and therefore reduces distances. In addition to this, Cervero & Kockelman (1997) mention the possibility of combining commute travel with for example grocery shopping, as a result of mixed land uses within a neighbourhood. This affects travel behaviour, since it could decrease trip numbers and influence travel mode choices. Diversity can be measured in several simple or more complex ways. A straightforward way is to break down the total land in an area into shares of each type of land use. A more sophisticated way is for example the 'dissimilarity index', which is introduced by Cervero & Kockelman (1997). They divided an area into grid cells and counted for every grid cell the number of neighbouring grid cells having another land use type. They also measured the land use entropy index. This entropy index is a commonly used index for representing the land-use mix by quantifying the homogeneity of land use within an area (Sarkar, Mallikarjuna, Bordoloi, & Mote, 2013). The distance from each house in an area to the closest store is also used to indicate land use diversity. The closer the average distance of houses in an area to certain services, indicates to what extent a built environment is mixed (Handy et al., 2002). Lastly, the jobs-to-population balance compared to the countrywide average could also indicate the diversity of an area (Cervero, 2002).

2.2.3 Design of a built environment

Design is related to how the built environment is set up with for example the design of the street network, affecting the street connectivity and accessibility of a neighbourhood for example (Ewing & Cervero, 2010; Handy et al., 2002). The street connectivity can be defined as the directness and the availability of (alternative) routes from an origin to a destination within a street network (Handy et al., 2002). More crossings indicate a denser road network, thus providing more route alternatives and probably shorter/faster routes. Cervero & Kockelman (1997) mention both the general objective to increase accessibility as well as using design to provide certain mobility groups (such as pedestrians and cyclists) amenities to encourage non-motorized modes of travel. This can for example be done by creating fast crossings, separate bicycle lanes and green-rich bicycle environments. Making one-way streets for cars accessible for cyclists in both directions also improves the connectivity of the cycling network, thus providing relatively shorter routes for cyclists in comparison to motorized travel routes. Design variables in particular are very useful to distinguish for example pedestrian-oriented or bicycle-oriented environments from auto-oriented environments. Ewing & Cervero (2010) for example mention measures for pedestrian-oriented environments as average street widths, sidewalk coverage, numbers of pedestrian crossings, street trees etc.. As regards to cycling, Hull & O'Holleran (2014) argue that the perception of safety, comfort and continuity of the cycling network in the end influences a person's perception of cycling. Measures such as the street connectivity (crossings per square kilometre), share of bike lanes in total roads, street lighting, and type of cycling lane pavement can indicate the cycling friendliness of an area. The difference in shortest cycling route compared to shortest motorized travel route also indicates how cycling friendly a built environment is.

2.2.4 Previous research on bicycle use and built environment

Built environment influences on bicycle use have for a certain extent been subject to research before. Two studies focussed on this exact same relationship, however both drawing different conclusions. Winters et al. (2010) found that built environment factors do have an influence on bicycle use in their study area Vancouver (Canada). Another study also specifically examined the influence of the built environment on cycling in King County, Washington (US), but it turned out to be insignificant (Moudon et al. 2005). Both findings however were based on telephone interviews where respondents had to respond to questions regarding their travel behaviour and actual trips made. The data used in their researches therefore was not based on actual cycling behaviour, but on travel habits perceived by individuals themselves. Moudon et al. (2005) also mention the constraint that the study area had limited bicycling infrastructure, making it difficult to prove a relationship. Other studies focus specifically on the design dimension, trying to examine bicycle infrastructure preferences which can encourage bicycling (Akar & Clifton, 2009; Caulfield et al., 2012; Hull & O'Holleran, 2014; Mccormack & Shiell, 2011). Their findings however have not been validated with actual cycling data. Wardman, Tight, & Page (2007) have researched factors influencing the propensity to cycle to work, but have only incorporated one infrastructural variable in their model (type of cycle lane). Based on interviews, they concluded that if the bicycle infrastructure would improve, the proportion of people choosing bicycle as mode choice would slightly increase.

2.3 Built environment measures

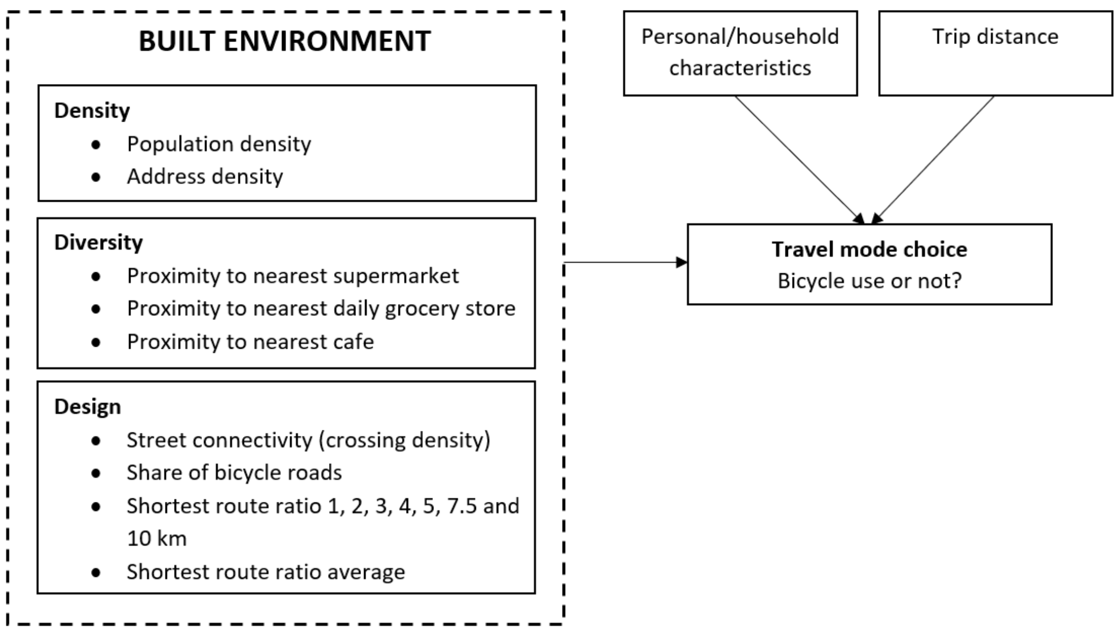
The concept of the 3Ds (density, diversity and design) has originated in the United States (US), where urban planning practices are very different than in Europe and in the Netherlands. Since being introduced by Cervero & Kockelman (1997), the 3Ds concept has provided a framework for further research on the built environment for many researches (e.g. Chen et al., 2008; Ewing & Cervero, 2001, 2010; Lu et al., 2017). From nature, distances between functions (working, living, shopping, recreating) in the US are larger than in the Netherlands, influencing the urban planning practices. Especially taking into account the space and a more extensive separation of functions in the US. Think about urban sprawl in US cities, with distant, private-car oriented suburbs lacking a mix of functions.

However, the 3Ds concept also does have components that are suitable for using the concept as a framework for the built environment in the Netherlands. Both density, diversity and design components can also be applied for the Netherlands. Since this research focusses on bicycle use, the design dimension specifically is an important measure of the cycling friendliness of a built environment. Therefore the density and diversity dimension cover two and three measures, while the design dimension will cover multiple measures.

2.4 Conceptual model

From the theories and concepts discussed in this chapter, a conceptual model has been developed. The conceptual model (Figure 4) shows how the variables and dimensions of the built environment and cycling behaviour in this research are related to each other. In the analysis, these relationships will be tested and assessed.

Figure 4: Conceptual model of the research subject.



3 Methodology

This chapter discusses the data that will be used to analyse the relationship between bicycle use and the built environment. It will discuss the datasets that are used as input and the software, tools and methods that are used for creating the built environment measures in GIS. A simplified research design scheme is included to visualize the complete research process. To conclude, methodology for the statistical analysis and the regression estimation will be explained.

3.1 Methodology introduction

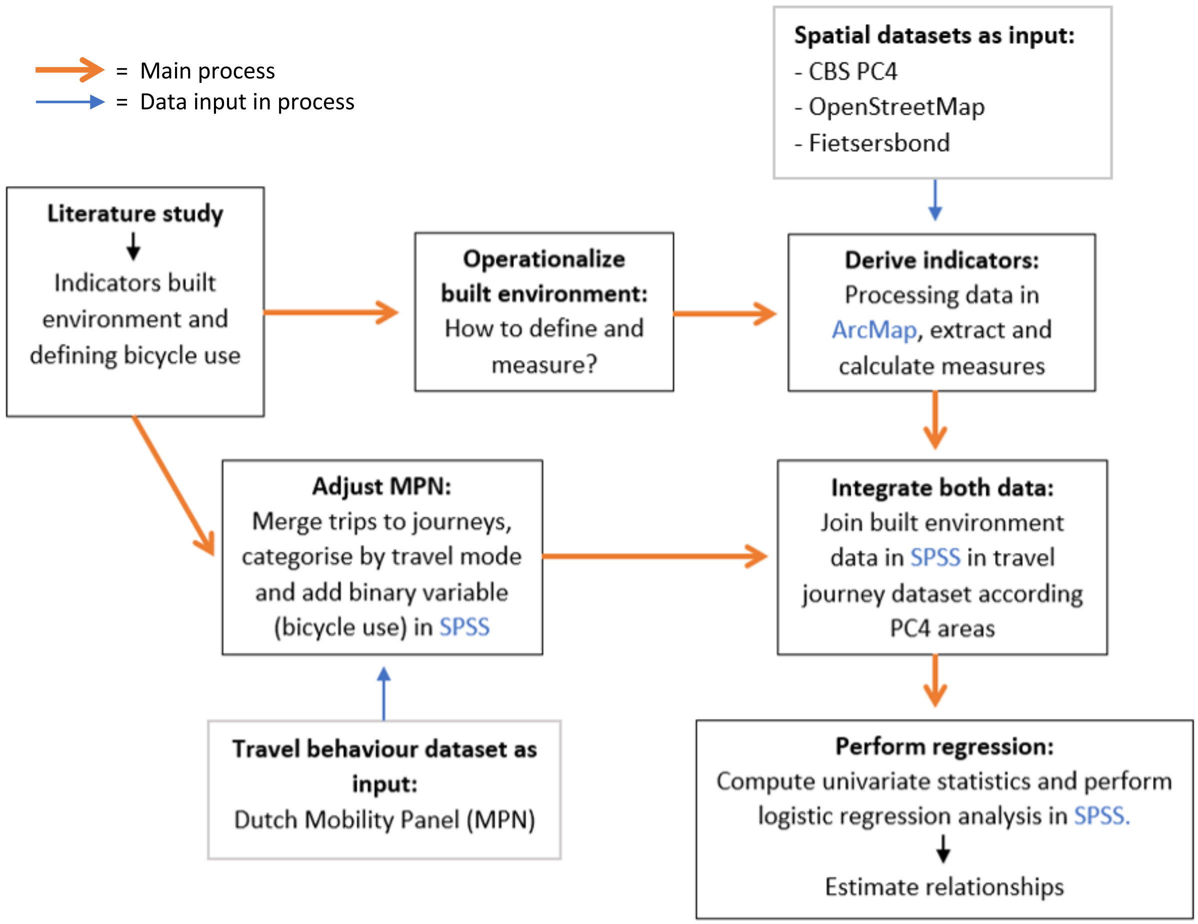
This research comprises of several research phases that need to be taken to reach its goal: providing insight in built environment influences on bicycle use as travel mode. Two main phases can be identified:
1) Processing travel data and calculating built environment measures in GIS.
2) Performing a regression analysis to determine built environment influences on bicycle use as travel mode.

The first step is to explore datasets, software and methodology for the empirical part of this research. On the next page in section 3.2, a research design scheme is presented to visualize the research process and the used methods in short. Section 3.3 explains the software that is used to explore and process the data and to perform the regression analysis. Next, section 3.4 describes the datasets that are used in the research. The travel data needs to be modified and extended, to create a variable indicating whether a trip is made by bicycle or not. Measures on the variables of the built environment need to be derived from spatial datasets. Section 3.5 therefore describes the operationalization of every variable, after which in section 3.6 the data, software and methodology used to derive these measures will be explained. Section 3.7 comprises of the methodology for the regression analysis. This research design scheme visualizes the research process and the used methods in short.

3.2 Research design scheme

To visualize the research process and the steps that need to be taken to successfully carry out this research, the next research design scheme (Figure 5) has been created.

Figure 5: Research design scheme of this thesis project.



3.3 Software

To process the (spatial) data and perform the regression analysis two software programmes will be used. IBM's SPSS, a programme used for analysing and visualising statistical data, is used for the regression analysis. ESRI's ArcMap is used to derive built environment measures, representing the cycling friendliness.

3.3.1 SPSS

As Bryman (2012) mentions, SPSS is probably the most widely used statistical software programme for quantitative data analysis in social sciences. The programme provides a computer environment where datasets can be uploaded, created, transformed, added with new variables and analysed in tables, charts, histograms etc. SPSS also has many statistical analyses integrated into the programme. One of these integrated analyses is the logistic regression model, which will be used to analyse built environment influences on bicycle use as travel mode. After the built environment measures have been derived from the spatial datasets, this data will be joined with the travel behaviour data in SPSS. Finally, the actual statistical analysis will be carried out in SPSS.

3.3.2 ArcMap 10.5

Geographic Information Systems are very useful for handing spatial data. GIS provide many tools to generate, adapt, analyse and visualize spatial information (ITC, 2013). It therefore is very convenient to use for deriving measures of the built environment from spatial datasets. ESRI's ArcMap is one of the GIS software programmes providing tools that can be used to derive these data. Take for example geoprocessing tools such as buffer, clip, merge, dissolve, intersect, union and erase. These tools help to process geospatial information and to create or derive new information from existing data. Section 3.5 'Variable operationalization' and 3.6 'Variable calculations', describe for every built environment variable how it is measured and which geoprocessing tools are used.

3.4 Datasets

One dataset is used to obtain travel behaviour data, while multiple datasets are used to derive built environment measures from. To assign measures of the built environment to every four numerical postal code area, the PC4 map of the CBS PC4 shapefile will be used. All built environment measures will be assigned individually to all postal code areas. The rest of the used datasets will now be introduced, beginning with travel behaviour data.

3.4.1 Travel behaviour dataset

Data on travel mode choices will be derived from a travel behaviour dataset (Hoogendoorn Lanser, 2015). From 2013 onwards, The KiM Netherlands Institute for Transport Policy Analysis has initiated 'The Netherlands Mobility Panel' (MPN). The MPN is used to determine the short-run and long-run dynamics in travel behaviour from both Dutch individuals and households and to determine possible correlation between changes in individual and household characteristics and changes in travel behaviour (S. Hoogendoorn-Lanser, Schaap, & Oldekalter, 2015). The panel consists of approximately 4000 respondents from 2500 households, who filled in questionnaires about their personal characteristics and household characteristics and have held a travel diary for three consecutive days. The travel diary consists of information on addresses of visited locations and the main activities, trips in terms of departure and arrival times, order in which transport modes were used, distances covered, parking costs, delays and travel companion (S. Hoogendoorn-Lanser et al., 2015). For this research, the travel diary, personal and household datasets will be used. The most recent travel behaviour data that is freely accessible is from 2014 and therefore this dataset will be used in this research.

3.4.2 Built environment datasets

For the set of built environment measures, various datasets will be used. Some datasets provide data for multiple built environment measures, therefore the datasets used are introduced per built environment dimension (density, diversity and design).

Density dimension

For measures on density of the built environment, one dataset containing a lot of information will be used: CBS on PC4 scale.

CBS PC4 2015 (Centraal Bureau voor de Statistiek, 2015)
The Centraal Bureau voor de Statistiek, (Statistics Netherlands) publishes statistics aggregated to the PC4 level every subsequent year. Statistics back to 2015 are freely available, therefore the statistics of 2015 are chosen taking into account that the MPN travel data is from 2014. This dataset contains a shapefile with values for a lot of variables per PC4 area. It for example includes numbers on population, households, dwellings, and address density. This data is very useful for quantifying the measures for the density dimension.

Diversity dimension
In order to quantify the diversity of the built environment, several datasets will be used. For the land use diversity share and indices, a dataset containing land use will be used: BGT. For data on distance to shops and jobs-to-population, the CBS PC4 dataset used.

CBS PC4 2015
As explained, this dataset contains statistical information for a lot of variables on PC4 level. It includes a large number of measures of the proximity to services such as supermarkets, grocery shops, healthcare, cafes, restaurants etc.. These measures can be used to indicate how diverse a built environment is in terms of land uses.

Design dimension
To quantify the design of the built environment, two datasets will mainly be used: OpenStreetMap and the Fietsersbond.

OpenStreetMap (OpenStreetMap Contributors, 2019)
The platform OpenStreetMap consists of a community of volunteers who collectively publish the OpenStreetmap, which basically is a world map containing information on for example paths, roads, sidewalks, cafés, transit stations etc (OpenStreetMap, 2018). With its open source principle, this world map is continuously getting improved by using local data and local knowledge. The results, the map itself and the data that it contains, are freely accessible under terms of the OpenStreetMap License. The data may be used for any purpose as long as OpenStreetMap is referred to as source. It contains lots of information on the infrastructural network for motorized travel, but for cyclists and pedestrians as well. Each feature, for example an edge representing a road or path, has an unlimited amount of attributes. The attributes of a road/path feature may provide information on its design: is the road accessible for cyclists? If so, is it a shared road or is there a separate cycle lane? This kind of information may be useful for the design dimension of the built environment. It can serve to quantify the cycling network and therefore indicate the cycling friendliness of a built environment.

Fietsersbond (Fietsersbond, 2019)
De Fietsersbond (Dutch Cyclist Union) has created their own cycling network map making use of volunteered geographic information (VGI). A large number of volunteers help to map cycling paths and routes and to keep their cycling network dataset up to date. With this cycling network map, they have created a route planner for cyclists where users can specify their preferences and desired route type (e.g. shortest, green, race bicycle) to create a cycling route. The underlying dataset containing the cycling network includes many road variables on which the route planner is based on. It for example includes variables such as road type, pavement

type, sufficiency of street lighting, type of environment. This data may be used to quantify the cycling-friendliness of the design dimension as part of the built environment.

3.5 Variable operationalization

3.5.1 Bicycle use operationalization

In the MPN dataset, respondents filled in a complete travel diary for three subsequent days. One real important distinction has been made in the panel data between the definition of a journey and a trip. A person moving from an origin to a destination (reaching the goal of the movement) is called a journey. A trip however is defined as the part of the journey made by using one specific travel mode. A journey therefore can be made by making use of multiple travel modes and thus consist of multiple trips. Every trip within a journey is noted in the diary. Only a respondent's walk to their own travel mode (car, bike or moped) are not taken as a trip.

To clarify, here a short example. A person who bikes to the train station, takes the train to another city and then walks to his work, combines this all in one journey with the purpose of reaching his work. The origin here is his or her own house, the destination is the working location. This journey consists of three separate trips: bicycle – train – foot.

In this research, bicycle use will be taken as the dependent variable. But how is bicycle use in this research precisely defined? This research considers bicycle use as the choice to use a bicycle for a complete journey. Only journeys which are mainly completed by a bicycle will be regarded as bicycle use. A cycling trip to a train station to go to work from there is in essence not a bicycle journey. Such a trip to the train station should be considered as pre-transit travel and therefore does not count as bicycle use. Therefore all journeys mainly made by bicycle are grouped into bicycle use. On the other hand is non-bicycle use in this research determined as individual motorized travel. All journeys that are mainly made by an individual motorized travel mode (car, motorbike, moped or electric moped), are grouped into non-bicycle use. The dataset already contains a variable indicating the main travel mode. This dataset will be extended with a dichotomous variable indicating whether the journey was made by bicycle or by motorized travel.

BicycleUse = if main travel mode of journey by bicycle, then 1 (=Yes)
if main travel mode of journey by individual motorized travel mode, then 0 (=No)

3.5.2 Control variables operationalization

The control variables Gender, age and household income are either personal or household characteristics of respondents. The data for these variables therefore is extracted from the personal and household characteristics datasets. The trip distance is included in the travel diary dataset.

Gender

A respondent's gender is a dichotomous variable coded as either male or female.

GENDER = if male, then 1
if female, then 2

Age

A respondent's age is presented in a range of 10 age groups. This ordinal variable is coded as follows:

AGE = *1: if <12*
2: if 12-17
3: if 18-24
4: if 25-29
5: if 30-39
6: if 40-49
7: if 50-59
8: if 60-69
9: if 70-79
10: if >80

Household income

A respondent's household income is presented in a range of 6 Income groups, presented in income in Euros. This ordinal variable is coded as follows:

HH_INCOME = *1: if <12.500*
2: if 12.500 – 26.200
3: if 26.200 – 38.800
4: if 38.800 – 65.000
5: if 65.000 – 77.500
6: if >77.500

Trip distance

The distance of each trip is presented as a ratio variable and indicates the distance covered in kilometres, without decimals.

TRIP_DIST = Distance of a trip in kilometres

3.5.3 Built environment operationalisation

All built environment measures will be measured on PC4 level using GIS, making use of ArcMap. Since a bicycle is mainly used for relatively short distances, the most reasonable choice is to relate built environment measures of the living area of the respondent to a respondent's journeys.

Population density

Using the CBS dataset, the total population per PC4 area will be extracted. Subsequently, the average population per square kilometre is calculated for every PC4 area.

Pop_Dens = average population / km²

Address density

The CBS dataset already contains a measure indicating the density of addresses, called the OAD (Omgevingsadressendichtheid). For every address is calculated how many addresses are located in a radius of 1 kilometre, divided by the area of the circle (Centraal Bureau voor de Statistiek, 2015). The OAD thus represents the number of addresses per square kilometre. From the OAD measure of all addresses in a PC4 area, the average is calculated. This average

OAD per PC4 area is already included in the CBS dataset.
OAD = average number of surrounding addresses / km²

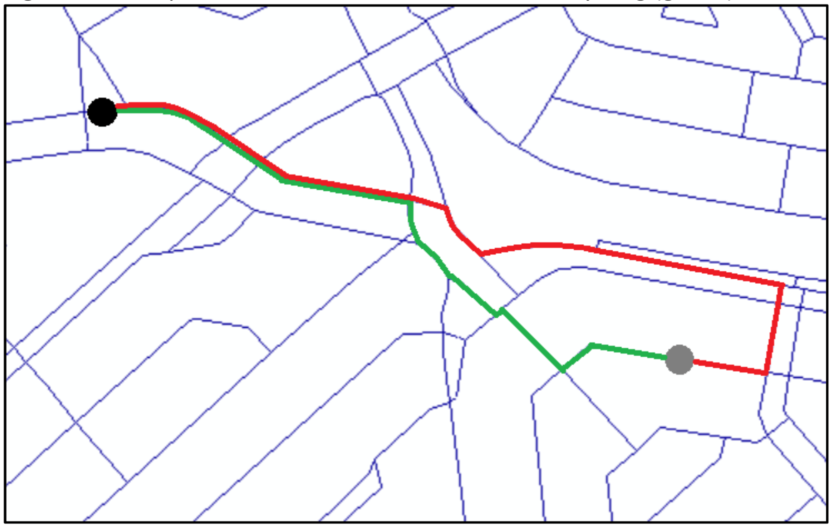
Proximity services
The CBS dataset contains many measures of average distances to services given in kilometres. The distances present the actual travel distance by road, not by a straight line. A few of these measures present the distance from a house to the nearest supermarket, nearest grocery store (for daily groceries: bakery, butcher, greengrocer etc.) and the nearest cafe. In this dataset this distance is already calculated for every inhabitant of a PC4 area and is presented as the average distance for every person living in a PC4 area to the nearest kind of service.
DIST_SUPER = average distance of all addresses to closest grocery shop in km
DIST_DAILY = average distance of all addresses to closest grocery shop in km
DIST_CAFE = average distance of all addresses to closest cafe in km

Street connectivity
Using the OSM and Fietsersbond datasets, measures of the average number of crossings can be calculated. Here, a crossing is a junction where roads meet each other or where roads end. The number of crossings will be calculated per PC4 area, divided by the total square kilometre of the area. This way it will be presented as the density of crossings per square kilometre.
CROSS_DENS = average number of crossings / km²

Share of bicycle roads
Both the OSM and Fietsersbond dataset contain the road network in the Netherlands, indicating the type of road and the type of cycling road (shared, separate lane), if present. By reclassifying the types of road into either cycling road and not solely cycling road and calculating the lengths of road segments, the share of cycling roads in the total amount of roads can be calculated. This measure indicates to what extent the road infrastructure is suited for cyclists, given in the percentage of the total amount of roads .
*PERC_C_ROAD = (total length of cycling roads / total length of all roads) * 100.*

Shortest route ratios
Using the OSM and Fietsersbond datasets, containing the road and bicycle network in the Netherlands, the shortest routes for the same origin and destination can be calculated for both cyclists and motorists. This will give the shortest route and distance for the cycling infrastructure and for motorized travel infrastructure. Dividing the distance of the cycling route by the distance of the motorized travel route gives a ratio indicating whether and to what extent the cycling infrastructure actually provides a shorter distance. Figure 6 shows an example of such a route calculation. The motorized route is longer (e.g. 1000 meters) than the cycling route (e.g. 920 meters). This would result in a shortest route ratio of 0,92, indicating that the cycling route is 0,92 times the distance of the motorized travel route. The route origin for every distance variable is fixed, located on the centroid of a PC4 area. The route destinations will be randomly determined for every (straight line) distance from the centroid: 1, 2, 3, 4, 5, 7.5 and 10 kilometre. After calculating all shortest route ratios, for every PC4 area the average shortest route ratio is calculated. The next standard formula is used to calculate every shortest route ratio, while all formulas can be found in Appendix A.
*SR*km = distance shortest cycling route / distance shortest motorized travel route*

Figure 6: Example of shortest route calculation for cycling (green) and motorized travel (red).



3.6 Variable calculations

All introduced variables that are part of the research need to be derived from the datasets of the MPN, CBS PC4, OpenStreetMap and Fietsersbond. This section will elaborate on the process of calculating these variables by describing the data, software, tools and calculations used. First, for every built environment variable will be explained how the measures are calculated. Next, the process of creating the bicycle use variable will be explained.

3.6.1 Built environment measures

Population density

The CBS 2014 PC4 shapefile contains lots of information and data, aggregated on four digit postal code areas. This file includes the total population number per PC4, which can be used to calculate the population density. The CBS has used ESRI's postal code areas, which are deduced from BAG-addresses (the Dutch Addresses and Buildings Key Register). By using the *calculate geometry tool* in ArcMap, the total area of each postal code area is calculated. Subsequently, a new field is added where a simple calculation for the population density per square kilometre is performed. The result is a value for every PC4 area, indicating the average population number per square kilometre.

Address density

The CBS 2014 shapefile contains a measure for the density of addresses, the OAD. This measure represents the density of addresses in a PC4 area. The CBS already has calculated this measure by counting for every single address, the total number of surrounding addresses within a radius of 1 kilometre. Subsequently, they have divided this number by the area of the 1 kilometre radius. Finally, the average of this surrounding addresses value per single address is calculated per PC4 area. The result is the OAD value per square kilometre for every PC4 area.

Proximity to nearest supermarket, grocery store and cafes

The CBS 2014 shapefile has information included on the proximity to a comprehensive set of services. In this research, the diversity of a built environment is measured by the average proximity to the nearest supermarket, grocery store and café determined by road distance. These measures are already calculated by the CBS for every single inhabitant, after which they have aggregated the average on the PC4 level. The result is the average proximity to a service in kilometres by road distance for every PC4 area.

Street connectivity

A built environment's street connectivity is measured by the density of crossings of the bicycle infrastructure. For this variable, the Fietsersbond dataset and the PC4 data from the CBS is used as input. The calculation takes place in ArcMap. The next model (Figure 7) in short explains which steps are taken to calculate the density of crossings in ArcMap.

Figure 7: Calculation of crossing density per postal code area in ArcMap.

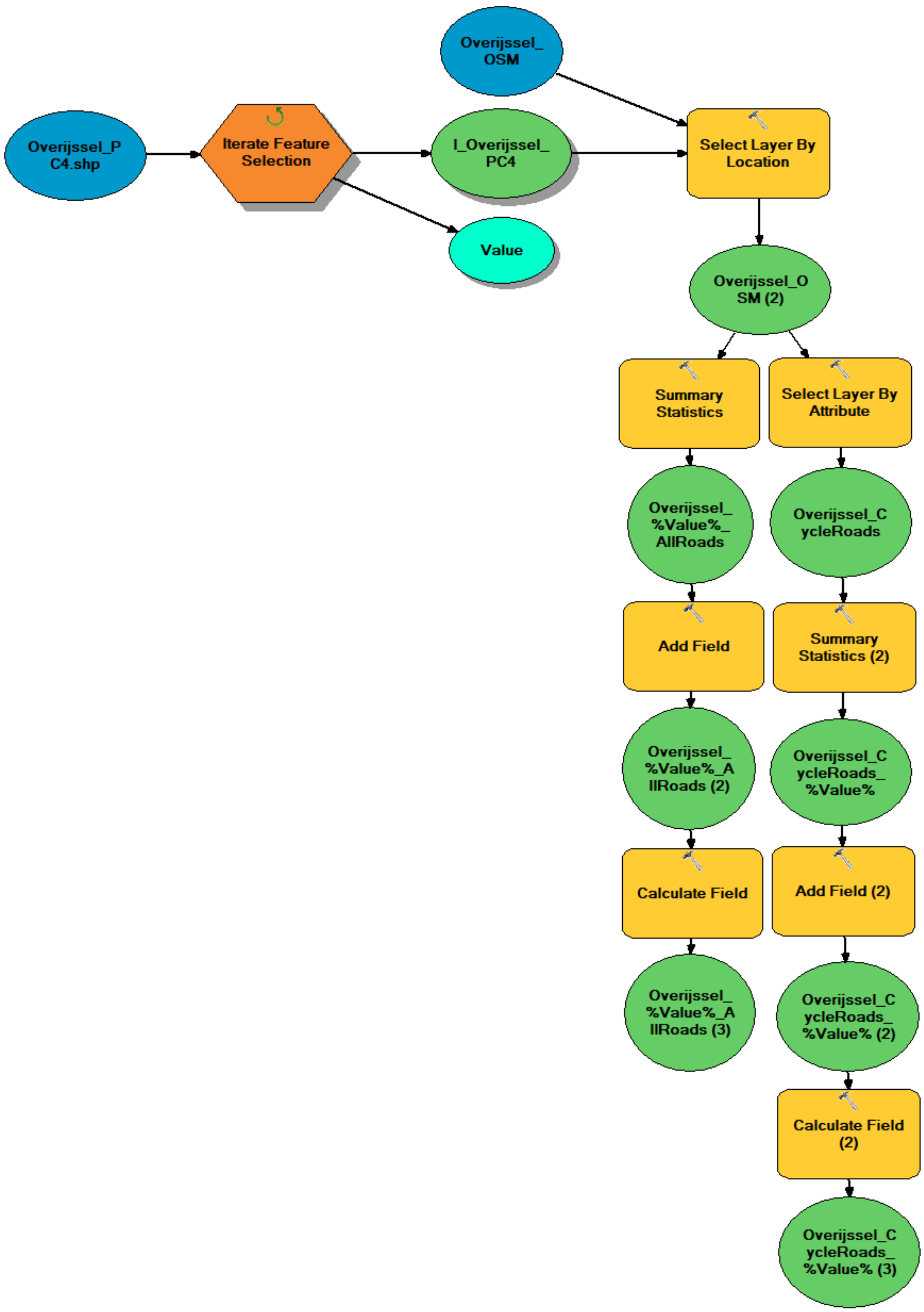
The Fietsersbond road infrastructure shapefile includes all road segments in the Netherlands, also those which are not accessible for cyclists. Cyclists can only use crossings which are cyclist-accessible, therefore all non-accessible roads for cyclist are deleted. One variable in the Fietsersbond dataset indicates the accessibility of road segments for cyclists. A new shapefile is created by selecting all cycling-accessible road segments, resulting in a cycling infrastructure shapefile for the whole Netherlands. Since this research only focusses on Zuid-Holland and Overijssel, all road segments that are not located within these provinces are deleted using the *Clip* Tool.

Using the *Network Analyst Extension*, a Network Dataset is created from the cycling infrastructure shapefile. The results are a file containing a network where all lines are connected and a file with all junctions, locations where road segments cross each other. The junctions file is a point shapefile which is used to calculate the crossing density. Using the *Join by Location* tool, the attributes from the junction point file are assigned to the CBS PC4 file (containing polygons of the postal code areas). The result of the tool is a new shapefile of the PC4 polygons including a new variable, the count of crossings that intersect with that specific PC4 area. The density of crossings in a PC4 area is calculated by adding a new field where the average number of crossings per square kilometre is calculated (junction count / PC4 area in square kilometres).

Share of bicycle roads

To calculate the share of bicycle roads for every postal code area, the OpenStreetMap shapefile is used in combination with the CBS PC4 shapefile. The OSM file contains a variable indicating the type of road segment: road, footway, cycleway, bridleway, etc.. This variable is used to filter in road types and calculating the road segment lengths. Multiple tools and calculations are used for each postal code area. Therefore Model Builder in ArcMap is used, to efficiently follow this procedure. Model Builder makes it possible to create a model of all the tools, calculations and datasets, to be able to perform all commands at once. By iterating through a list of features, the complete model will be applied to every individual feature one by one. It means that the model will automatically iterate through the postal code areas one at a time. The model used is presented in Figure 8 and the explanation of the model is to be found in Appendix B.

Figure 8: Model builder calculating the share of bicycle roads in the total amount of roads (in this case for Overijssel)



Shortest route ratios

To calculate the shortest route ratios for every postal code area, data from the Fietsersbond shapefile and the OpenStreetMap shapefile is used. These datasets contain information on the road network in the Netherlands. Since a large number of tools and calculations is required, Model Builder is used again including an iteration through the postal code areas.

To calculate the difference in shortest route distance for bicycles and motorized travel, the same origin and destination is used for the cycling route and motorized travel route. The origins and destinations are determined as follows:

- Origins: all route distance variables are calculated from the centroid of each postal code area. Therefore first the centroid is calculated by means of the *Calculate Geometry* tool. This tool calculates the X coordinate and Y coordinate of the centroids. By using *Display XY Data*, a point file is created representing the centroids of all postal code areas. This centroid point file is used as input for the *Iterate Feature Selection* in the route model.
- Destinations: The destination of all route distance variables are randomly generated in the bicycle route model using a few tools. These destination points are saved and merged in a folder, to use as input for the destinations in the motorized travel route model.

<u>Cycling route model</u>

The cycling route model is presented in on the next page in Figure 9. First, the model iterates through the postal code areas using the *Iterate Feature Selection* tool. The destination points for all shortest route distance variables (1km, 2km… and 10km) are created by means of placing random points on buffer lines for each distance. These destination points are then saved, since they need to be used in the motorized route model as well. A *Network Dataset* is created after the road network typology is improved, to be able to make route calculations. The route calculations are made by using the *Closest Facility Analysis* tool, where the postal code centroids are added as route origins and the random points as route destinations. All route results are saved in one folder, to be merged later to one file. A more comprehensive explanation of this model is to be found in Appendix B.

<u>Motorized travel route model</u>

The model used is presented on page 30 in Figure 10. For this model, the same *Feature Iterate Selection* tool is used to iterate through all postal code areas. Since motorized travel routes use motorized-specific road segments, a *Network Dataset* is created from the OSM file that only contains road segments that are motorized travel-accessible. The Network Dataset typology is improved by means of the *Planarize* and *Integrate* tools. A *Closest Facility Analysis* is performed to calculate the motorized travel routes, where the postal code centroids are added as route origins and the random points as route destinations again. The network typology includes traffic direction, which is taken as restriction in the route calculation. All route results are saved in one folder again, to be merged later to one file. A more comprehensive explanation of this model is to be found in Appendix B.

Figure 9: Model builder to calculate the shortest bicycle routes (in this case for Overijssel).

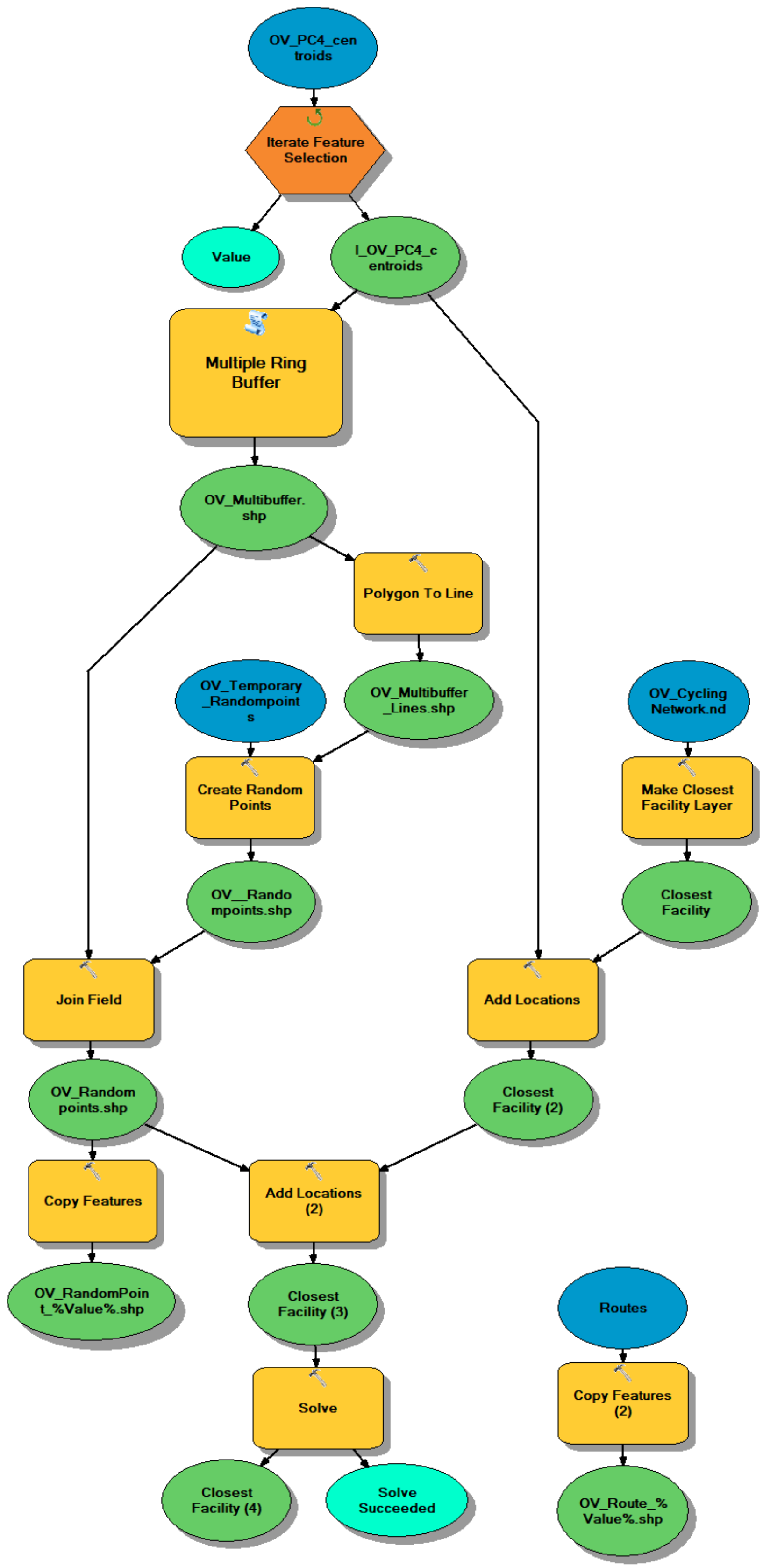


Figure 10: Model builder to calculate the shortest routes for motorized travel (in this case for Overijssel).

<u>Shortest route ratio calculation</u>

The resulting routes of the bicycle route model and motorized route model are all saved in two folders. Using the *merge* tool, all bicycle routes are merged into one file and all motorized routes in another file. Both resulting files are joined together based on the PC4 value. The shortest route ratio is calculated in a new field by dividing cycling route distance by the motorized travel route distance.

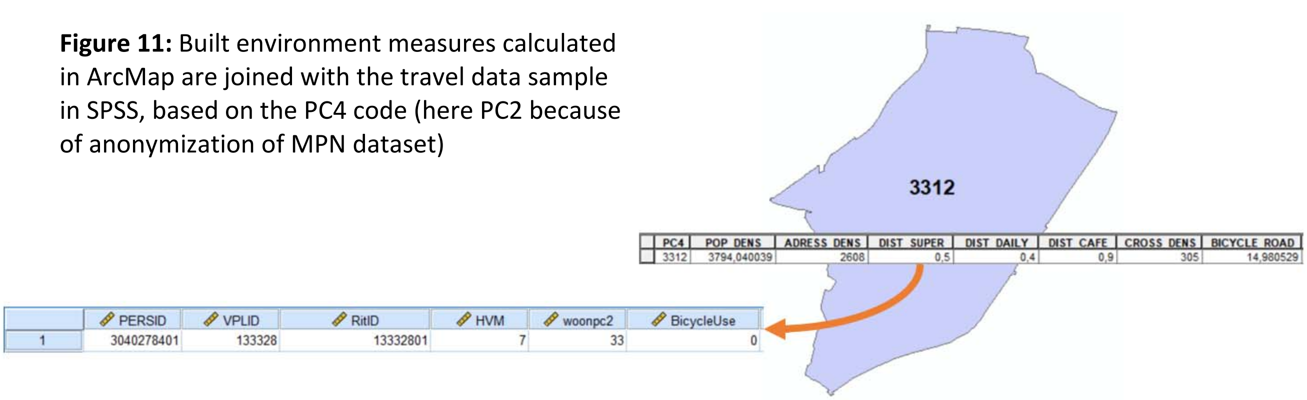
For a small number of postal code areas, not for every distance a route can be calculated. This results in missing values in the dataset. To ensure that the resulting shortest route ratio variable data is as complete as possible, some data will be estimated based the other shortest route ratios within the same postal code. Only in cases where there is one missing value in all shortest route ratio variables, the missing value is estimated by taking the average of the other shortest route ratio variables. For postal code areas with two or more missing values in the shortest route ratio variables, no estimations are made and missing values are kept in the dataset.

To calculate the average shortest route ratio, all distance route ratios are used. The average shortest route ratio is only calculated for all postal code areas where there is there is a maximum of one missing value. Thus for every postal code area where two or more shortest route ratios have missing values, no average is calculated. The average shortest route ratio (AVG_RATIO) in these cases also get assigned a missing value.

3.6.2 Bicycle use measure

The Dutch Mobility Panel (MPN) dataset contains a large set of trips made by all its respondents (approximately 4000) who held a diary for three subsequent days. All the previously explained built environment variable measures will be joined to every trip made in the dataset, based on the residential PC4 of the respondent of the trip. Figure 11 shows how the built environment measures in ArcMap are joined with the travel data in SPSS. Since the MPN data is not publicly accessible on the PC4 level due to privacy laws, the Netherlands Institute for Transport Policy Analysis has offered to join the built environment measures in the travel dataset, based on the residential PC4 area of each respondent. To not mess with privacy regulations and making it impossible to trace back journeys to an individual, they applied a noise in the built environment measures. This noise applied has a margin of 2%, which means that every measure's value is randomly increased or decreased with a maximum of 1%. A crossing density of 305 for example, could change to a value between 302 and 308. As Mivule (2013) explains, this anonymization method does preserve the mean and covariance of the original data. The correlation coefficients and variances however might not be sustained. Unfortunately, this is the only option available to be able to make use of the MPN dataset.

Figure 11: Built environment measures calculated in ArcMap are joined with the travel data sample in SPSS, based on the PC4 code (here PC2 because of anonymization of MPN dataset)



Since every respondent in the MPN dataset have held a travel diary for three subsequent days, variation in travel activities are present within the dataset. Some respondents are less active than other respondents, what leads to over representation of trips made by active respondents. To tackle this issue, a sample of journeys will be drawn so that from every respondent one journey is included. Using SPSS, this one journey per respondent will be picked randomly.

The complete SPSS datafile of all trips made by the respondents in Zuid-Holland and Overijssel is cleaned first. Not every respondent has filled in all the necessary questions regarding each trip made, what has led to incomplete trips. To give an example, some trips do not have a trip ID or main journey travel mode used, but have blank fields for these variables. These trips are removed to create a dataset of complete trips. Each trip that has a blank field for the trip ID and/or the main journey travel mode is removed from the file. Next, a dichotomous variable indicating whether the trip was made by bicycle or by motorized travel is added by means of recoding the main journey travel mode. The trips made by bicycle, or race or electric bicycle are given the value of 1 (=yes), whereas trips made by individual motorized travel modes are given the value of 0 (=no). Motorized travel modalities here are a car, motorbike, scooter or electric scooter. This dichotomous value will serve as the dependent variable in the regression analysis.

Next, for every respondent one trip is randomly selected. The result is a dataset containing one complete trip for each respondent in Zuid-Holland and Overijssel that is made either by bicycle or motorized travel. Figure 12 in short describes the that are made. Having this bicycle use measure included, the dataset is ready for the analysis.

Figure 12: SPSS process to draw sample of the travel data file.

The syntax of the explained SPSS procedure can be found in Appendix C.

3.7 Statistical analysis

3.7.1 Variables

The statistical analysis includes three components: univariate, bivariate and multivariate analysis. In short this means exploring one variable, exploring the relationship between two variables and exploring the relationship between more than two variables. The variables included in the research can be distinguished in dependent, control and independent variables.

Dependent variable

The dependent variable in the research is bicycle use. Bicycle use is expected to be influenced by the control variables and the independent variables (built environment). In the multivariate analysis, bicycle use is predicted by the control and independent variables and thus is the outcome variable.

Controlling variables
Variables that you are not particularly interested in, but do influence the dependent and independent variables are so called confounding variables. Not including confounding variables may lead to wrong conclusions when assessing relationships between the independent and dependent variables (Field, 2013). These variables therefore are controlled for in the bivariate and univariate statistics. The controlling variables are gender, age, household income and trip distance.

Independent variables
The variables that are expected to influence bicycle use are the built environment variables. In the multivariate analysis, these variables are used to predict the outcome variable (dependent) and therefore are the predictor variables.

3.7.1 Univariate analysis
First, all variables in the research are examined individually. Bicycle use is a dichotomous variable and thus is presented in a cross table and pie chart (Bryman, 2012). The control variables and built environment variables are ratio variables (except for gender) and therefore are presented in histograms. For each built environment variable a map is included, showing how its values are spatially distributed.

3.7.2 Bivariate analysis
Second, the relationships between two variables at the time are examined. Here is examined whether variation in one variable coincides with the variation of another variable (Bryman, 2012). The method to examine relationships between a dichotomous variable (bicycle use) and interval/ratio variables (built environment) is calculating Spearman's Rho. This method provides for every relationship between two variables a coefficient between 0 and 1. The closer the coefficient to 0, the weaker the relationship and the closer to 1, the stronger the relationship is. Whether the relationship is positive or negative (+ or -), indicates the relationship's nature.

Statistical significance in quantitative research indicates the degree of coincidence that a relationship found in the research sample does not exist in the population. To express the statistical significance, probability levels are used: the *p*-value. This indicates the probability of rejecting the null hypothesis, where the null hypothesis imposes that two variables are not related in the complete population. In general, under researchers the assumption is made that a probability of 5% is the maximum, a *p*-value of 0,05. In other words, there are fewer than 5 chances in 100 that the research sample shows a relationship between two variables that does not exist in the population. A more strict maximum probability level is 1%, which means a *p*-value of 0,01. Both probability levels in this research are included in the bivariate and multivariate analyses

3.7.3 Multivariate analysis
Regression analysis
The relationships between all variables are examined by means of estimating a regression model. Regression models are used to predict a variable from another variable, or from multiple variables. By fitting a model to data, it is possible to predict an outcome (dependent) variable from one or multiple predicting (independent) variables (Field, 2013). Appendix D

explains the methods for regression analyses more in depth, as this section explains this in short.

Different kinds of regression models can be carried out. When having one predicting variable, a simple linear regression model is performed. When taking several predictor variables, a multiple linear regression model is performed. The simple and multiple regression models assume that the dependent variable has either an interval or ratio measure, thus is continuous. This makes it possible to derive a linear relationship between the dependent and independent variable(s). The general linear regression function is:

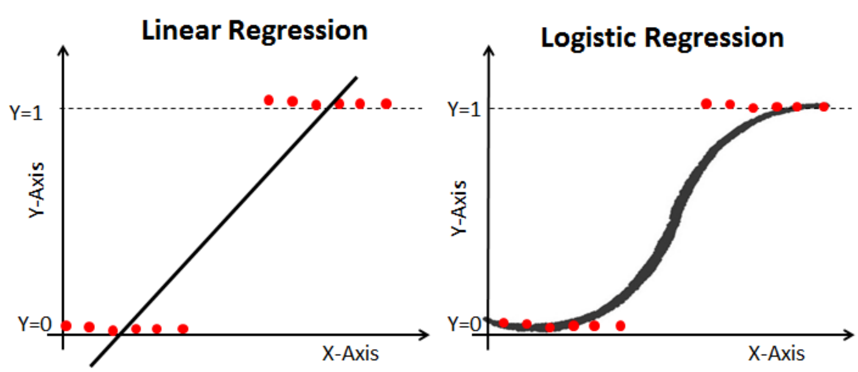
$$Y = (b_0 + b_1X_{1i} + b_2X_{2i} + \cdots + b_nX_{ni}) + \varepsilon$$

Where Y is the outcome variable, b_1 the coefficient of the first predictor (X_1), b_2 the coefficient of the first predictor (X_2), etcetera. Each predictor coefficient is estimated using the method of least squares, of which the actual curve can be drawn. The ε in the model represents the standard error of estimate of the regression curve.

Logistic regression analysis

When the dependent variable is not continuous but dichotomous, linear relationships are not present (Figure 13). Logistic regression in this case is the regression method to use (Sieben, n.d.). This research considers bicycle use as a dichotomous variable: a trip is made by bicycle or not by bicycle, therefore logistic regression is used to examine the relationships. In contrast to linear regression, logistic regression does not predict the exact outcome value of the dependent variable based on the values of the independent variables. It predicts the possibility of an outcome value, since the outcome value can only be either 0 (not bicycle use) or 1 (bicycle use). A logistic regression therefore predicts the probability of an outcome value.

Figure 13: Difference in regression line between linear and logistic regression in case of a dichotomous dependent variable. (Navlani, 2018)



Basically, the multiple linear regression function is transformed in logarithmic terms for the logistic regression:

$$(P)Y = \frac{1}{1+e^{-(b_0+ b_1X_{1i}+b_2X_{2i}+\cdots+b_nX_{ni})}}$$

P(Y) here is the probability of Y occurring, given in the range of 0 and 1. The closer to 1, the more likely Y has occurred. Since the regression coefficients are presented in logit transformed context and are difficult to interpret quickly, the regression coefficients are also presented in odd ratios (OR). Odds ratios are calculated by the plotting e to the power b, the coefficient (Field, 2013):

$$Odds\ ratio\ (OR) = e^{(b)}$$

$$Odds\ ratio\ (OR) = \frac{Odds\ after\ a\ unit\ change\ in\ the\ predictor}{Original\ odds}$$

In short, the OR indicates the ratio between the odds of Y occurring and Y not occurring (Y=1 or Y=0). The ratio has a value between 0 and infinity, where an OR between 0 and 1 indicates a lower odds on Y occurring when a predictor variable's value increases. A value >1 indicates a higher odds of Y occurring when a predictor variable's value increases.

Regression methods

Regression models are built by adding and/or deleting variables into the equation. The Appendix D explains several methods for doing this. This research uses two entry methods: enter and forward selection (conditional). One model is built by means of the forward selection (conditional) method. A computer algorithm builds up the model stepwise by adding variables one at a time. It evaluates whether a variable should be added by assessing the score statistic. Important predictors introduced in theory however might not be included. Therefore using the enter method, the researcher fully decides upon the inclusivity of the variables. To avoid multicollinearity, two independent variables that are strongly correlated are not added in the same model (Field, 2013). One model thus is estimated by means theories and logic thinking, using the Enter method

Model fit and predictive power

The fit of the regression models is mainly measured by the Nagelkerke R Squared, which indicates their predictive power. The higher the proportion of the variance in the dependent variable explained through the predicting variables, the higher the Nagelkerke R^2. The Chi-squared (-2LL) test and Hosmer Lemeshow-test are assessed to indicate if the model fits to the data.

4 Analysis

In this chapter, a statistical analysis is carried out. From the travel data research sample, several analyses are performed. An introduction of the analysis and data sample is given, followed by the univariate statistics of each variable. A closer look into the correlation between the variables will be taken, after which a logistic regression model will be estimated. This model will be used to predict bicycle use from built environment measures.

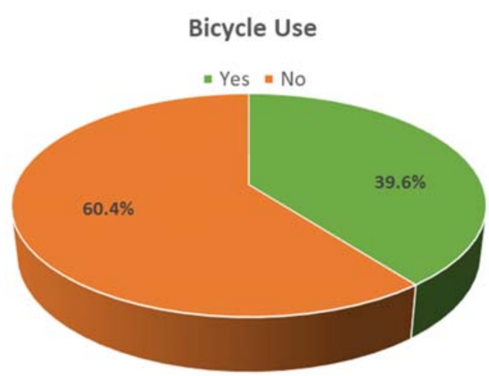
4.1 Analysis and data sample introduction

The analysis part of this research is conducted on the sample that is drawn according to the methodology in section 3.6.2. Univariate analysis of gender, age, household income and trip distance (control variables) is performed over the research sample. The univariate statistics for the built environment variables however will be calculated and presented over the complete set of for measures for all postal code areas in Zuid-Holland and Overijssel. Next, a bivariate analysis will be performed on the research sample, where first the correlation between built environment variables and bicycle use is examined. Subsequently, a logistic regression is estimated to relate bicycle use to the predicting built environment variables controlled for age, gender, household income and trip distance.

First, an introduction into the sample that is used for the research. From the total travel dataset where roughly 15.000 trips are made by inhabitants of Zuid-Holland and Overijssel, a research sample is drawn. The research sample consists of 1324 cases, thus 1324 travel journeys. These journeys are made either by bicycle or by a motorized travel mode. As can be seen in Figure 14, roughly 40% of the trips are made by bicycle, where 60% is made by a motorized travel modality.

Figure 14: Frequency distribution of bicycle use in the research sample.

Bicycle Use	Frequency	Percent
0 (no)	800	60,4
1 (yes)	524	39,6
Total	1324	100,0



4.2 Univariate analysis

This section explores the univariate statistics of both the control variables and the predicting (built environment) variables. The control variable statistics are presented in short, whereas the built environment statistics are presented and visualized more thoroughly to identify patterns in the spatial distribution of the variables. These built environment statistics are calculated over the complete dataset of all the PC4 areas in Zuid-Holland and Overijssel, comprising a total of 846 postal code areas.

Gender

The research sample consist of 1324 trips, all made by a unique individual. 630 trips are made by men (47,6%) and 694 trips of women (52,4%). Thus slightly more women are part of the research sample.

Age

The ordinal variable age is grouped in 10 categories (N=1324). Figure 15 shows the distribution of respondents in age groups. The labels of each group can be found in section 3.5.2 (variable operationalisation). The highest frequencies are in age groups 40-49 an 50-59.

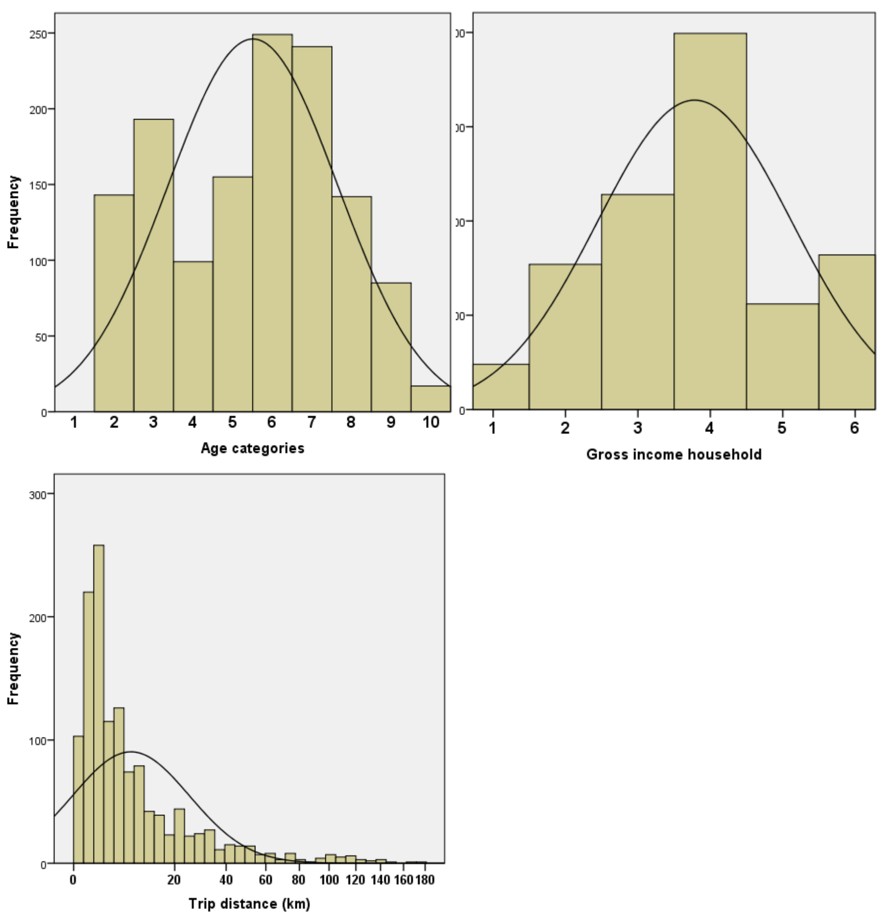
Household income

This ordinal variable is grouped in 6 categories, to be found in section 3.5.2 as well. Not every respondent provided an answer to this question, therefore the total cases is N=1105. Figure 15 shows the frequencies of each income group. There you can see that the most frequent household income group is €*38.800 – €65.000.*

Trip distance

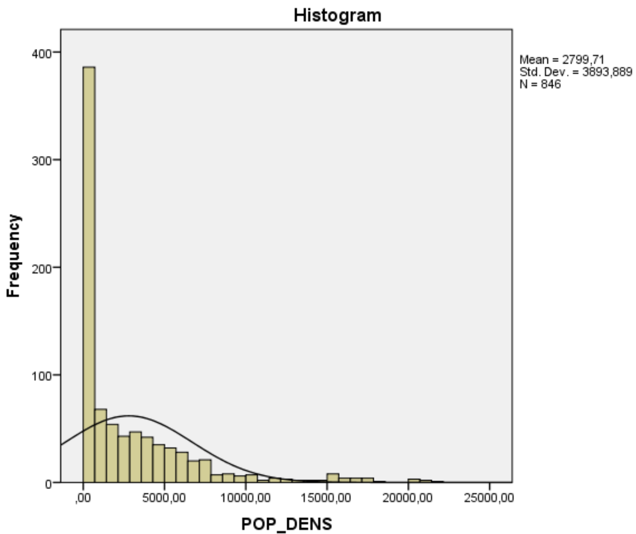
The ratio variable trip distance is given in kilometres (N=1313). The average trip distance is 12,46 kilometres, where 80% of all trips in the research sample are under 16 kilometres. There are a few outliers in the range of longer distance, however only 5% of all trips are longer than 50 kilometres. Figure 15 shows the frequency histogram of the trip distance variable.

Figure 15: Frequency histograms of control variables age, household income and trip distance.



Population density
The variable population density displays a high proportion of postal code areas with a population density lower than 500 inhabitants per km² (42,8%). The mean population density of the total 846 postal code areas is 2800, whereas the median is 1080 inhabitants per km². The difference between the mean and the median already indicates a skewness. The population density has a positive skewness value of 2,1 which means a right-skewed distribution. The frequency histogram in Figure 16 quickly shows that the data is not normally distributed.

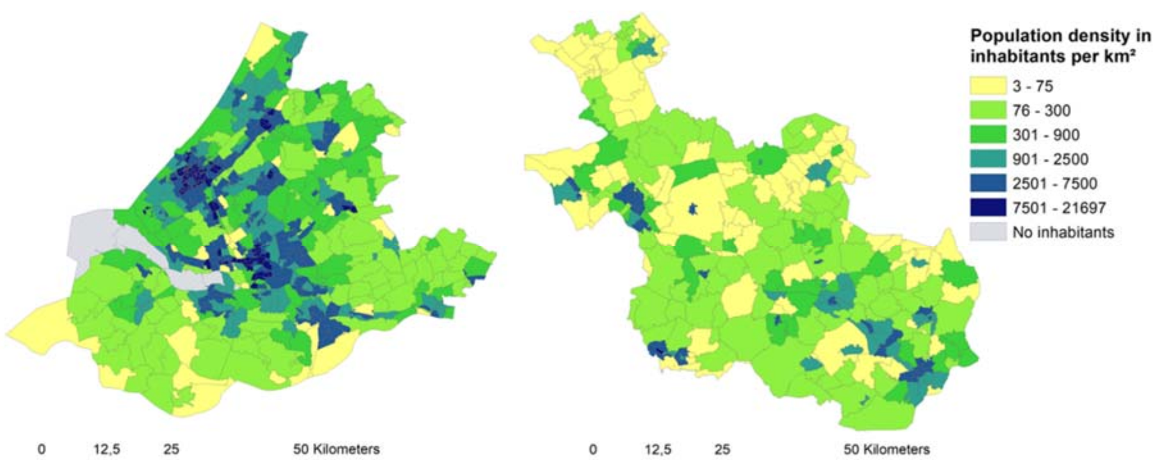
Figure 16: Frequency histogram and descriptive statistics for the Population Density (POP_DENS).



N	Valid	846
	Missing	0
Minimum		0
Maximum		2800
Mean		1605,93
Median		1211,00
Skewness		2,155
Std. Error of Skewness		,084

The spatial distribution of the population density variable is visualized in Figure 17. As expected, the higher population densities are clustered in and around cities. From the map of Zuid-Holland for example, Den Haag, Rotterdam and Leiden are easy to locate. Overijssel is obviously a less denser province, but it is still possible to locate cities as Zwolle, Deventer and Enschede. This map therefore also proves why Zuid-Holland and Overijssel are taken as cases, to ensure sufficient variety in dense and less dense postal code areas.

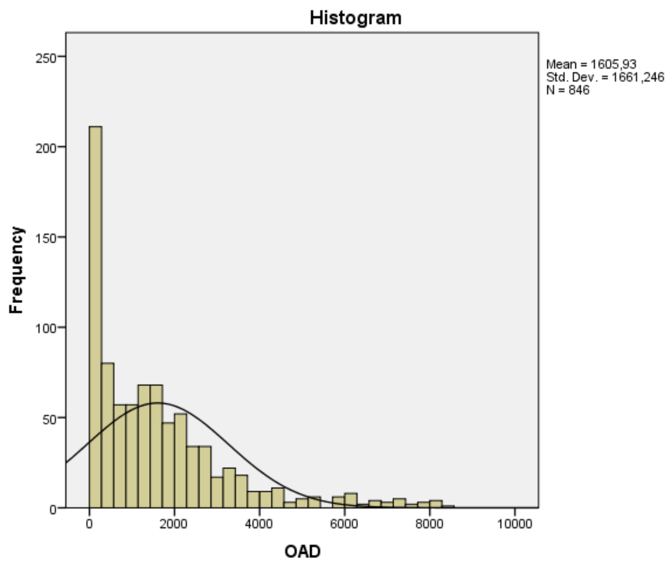
Figure 17: Spatial distribution of the population density for Zuid-Holland (left) and Overijssel (right).



Address density

The address density variable is presented by the OAD, which indicates the average number of surrounding addresses per km² for every postal code area. The total 846 postal code areas have a mean address density of 1606. For this variable the median also is substantially lower than the mean and the skewness value is positive again. However, the skewness value is lower in comparison with this value for the population density, which is also visible in the frequency histogram (Figure 18).

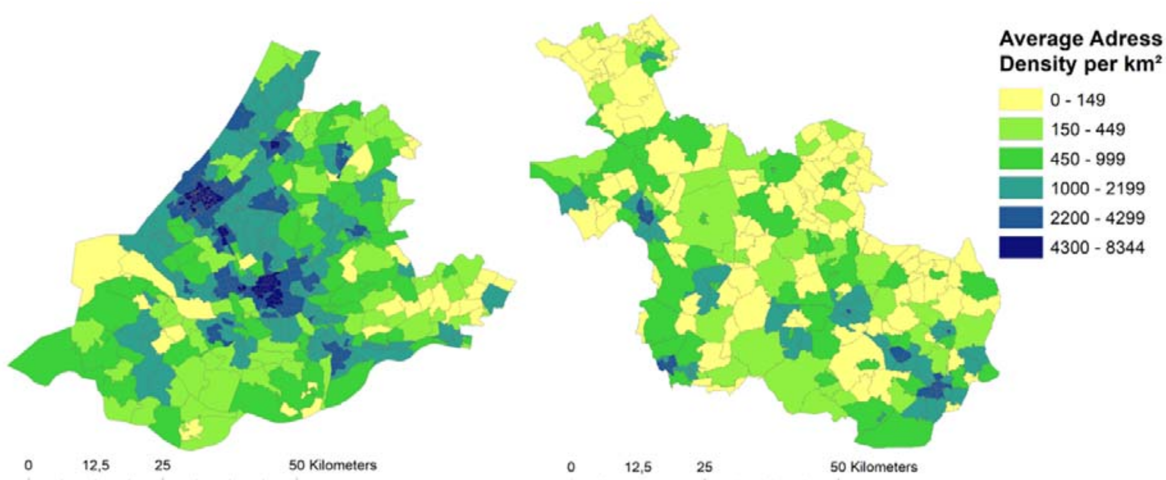
Figure 18: Frequency histogram and descriptive statistics for the address density (OAD).



N	Valid	846
	Missing	0
Minimum		0
Maximum		8344
Mean		1605,93
Median		1211,00
Skewness		1,650
Std. Error of Skewness		,084

The spatial distribution of the address density is presented in Figure 19. The map is relatively similar to the map showing the population density. In reality this is quite logical. More inhabitants means more residential locations, thus more unique addresses. The high density hotspots in these maps also reveal where the cities and their outskirts are located within both provinces. The higher spread of low address density postal code areas in Overijssel is also visible.

Figure 19: Spatial distribution of the population density for Zuid-Holland (left) and Overijssel (right).



Proximity services

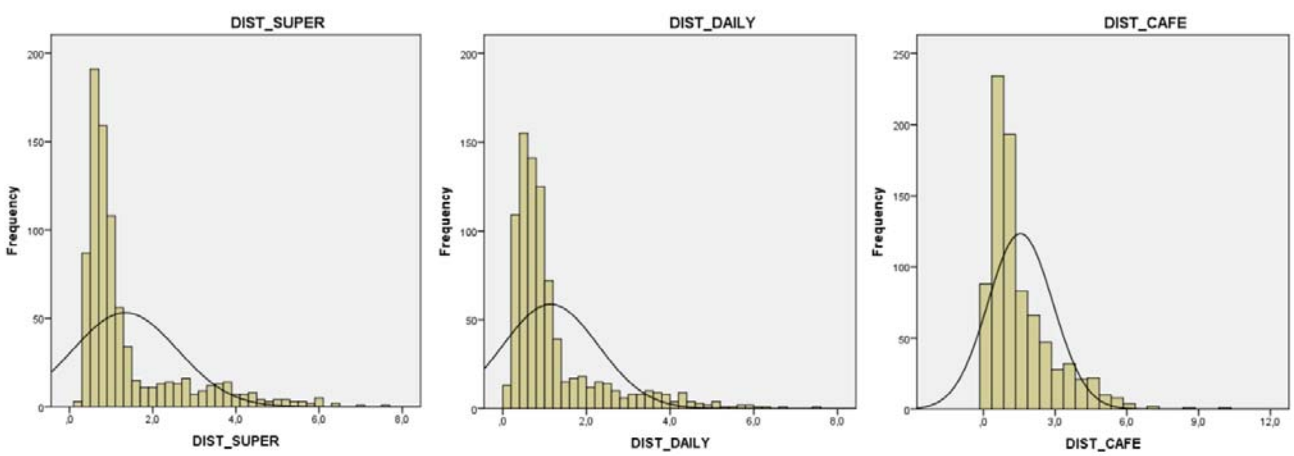
The proximity to services variable is measured for three kinds of services: supermarket, daily grocery store and cafes. Table 1 shows the descriptive statistics for each proximity variable, where for every variable there are 6 missing values in a total N of 846. The mean and maximum values of the three variables are relatively similar. Only the proximity to a café tends to have a little less skewness.

Table 1: Descriptive statistics for each proximity variable, measured in kilometres.

		Supermarket	Daily gr. store	Cafe
N	Valid	840	840	840
	Missing	6	6	6
Minimum		0,2	0,1	0,1
Maximum		7,6	7,4	10,1
Mean		1,338	1,143	1,549
Median		0,800	0,800	1,000
Skewness		1,972	2,165	1,740
Std. Error of Skewness		0,084	0,084	0,084

The table shows for each of the variables a higher mean than the median, which indicates a skewness in the data. The positive skewness value for each variable does prove this assumption. This is also evident when looking at the frequency histograms of the variables, which are presented in Figure 20.

Figure 20: Frequency histograms of proximity to supermarket, daily grocery store and cafes.



For each of the proximity variables, the spatial distribution in Zuid-Holland and Overijssel is presented in three maps on the next page (Figure 21, 22 and 23). The overall higher proximity values in Overijssel are noticeable in comparison with Zuid-Holland. As a less dens province, in most postal code areas the average proximity to a certain kind of service is evidently higher.

Figure 21: Spatial distribution of the average proximity of big supermarkets in kilometres.

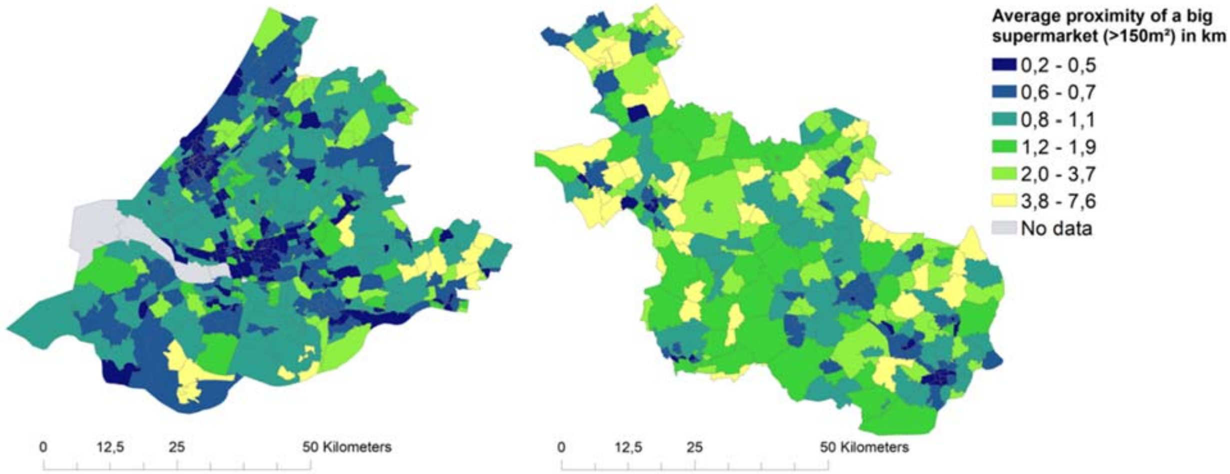


Figure 22: Spatial distribution of the average proximity of a daily grocery store in kilometres

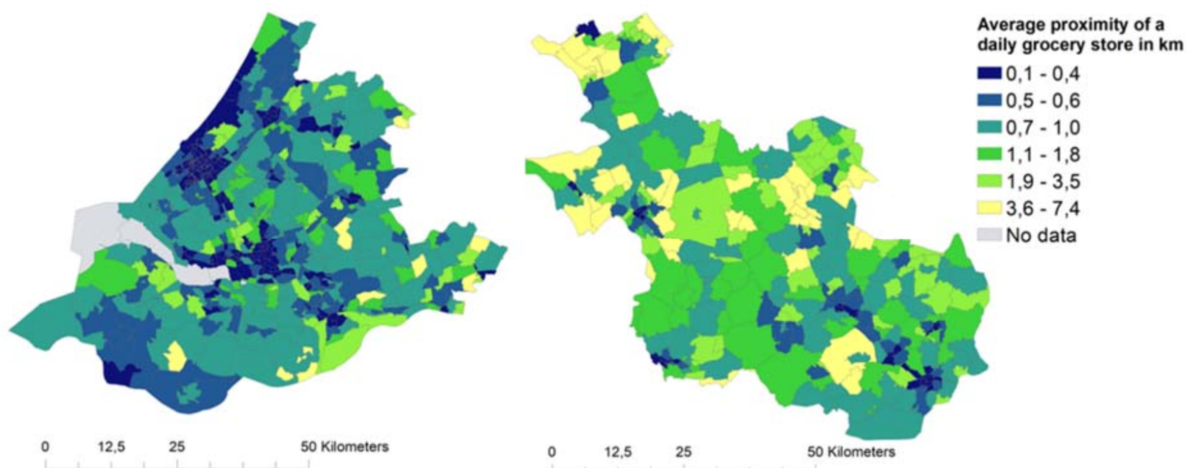
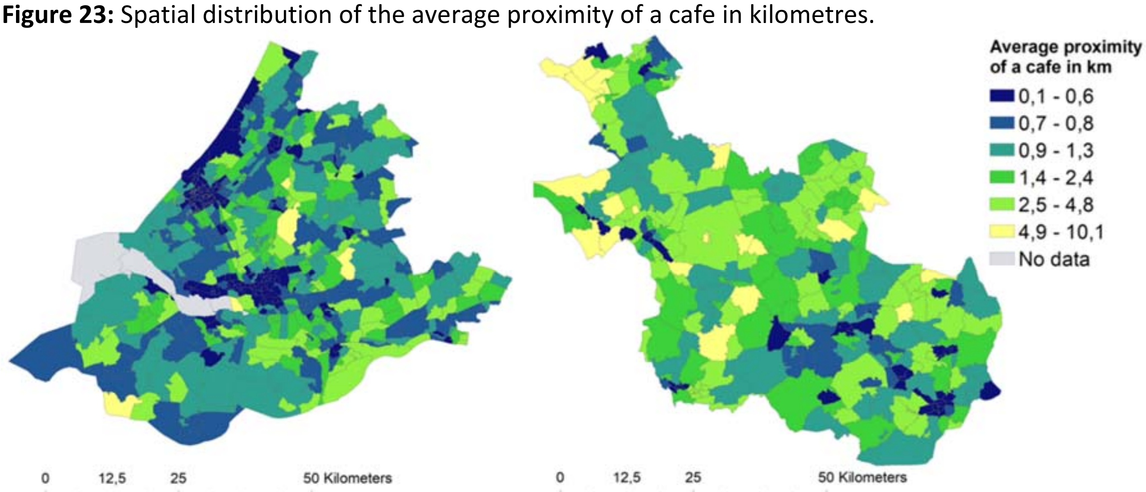
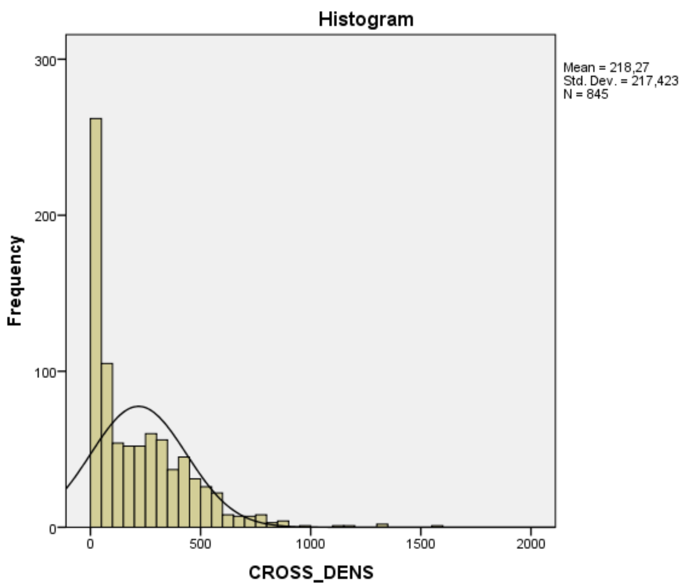


Figure 23: Spatial distribution of the average proximity of a cafe in kilometres.



Street connectivity
The variable street connectivity in this research is presented by the density of crossings in a postal code area, given in number of crossings per km². There are no missing values and the mean number of junctions is 218, whereas the median is 151. A positive skewness value is also noticeable in the frequency histogram in Figure 24, where the highest frequencies can be noticed in the lower crossing density values.

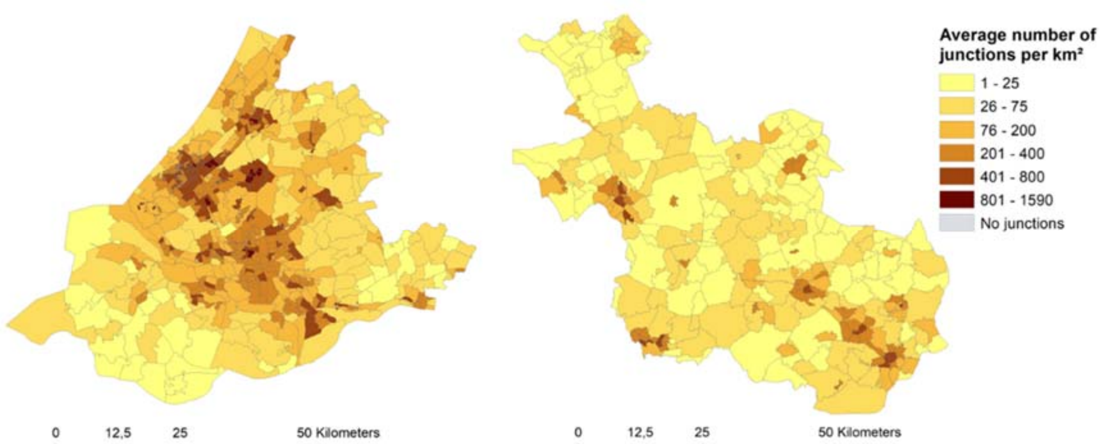
Figure 24: Frequency histogram and descriptive statistics for the crossing density (CROSS_DENS).



N	Valid	846
	Missing	0
Minimum		0
Maximum		1590
Mean		218,01
Median		151
Skewness		1,524
Std. Error of Skewness		,084

The spatial distribution of the crossing density is presented in Figure 25. As expected, the higher crossing densities can be found in city regions for both provinces. What is remarkable is the relatively higher crossing density in the regions of The Hague, Dordrecht and Zoetermeer in comparison with the second biggest city in the Netherlands, Rotterdam. The spread of low crossing density areas in Overijssel is also apparent.

Figure 25: Spatial distribution of the crossing density (CROSS_DENS).

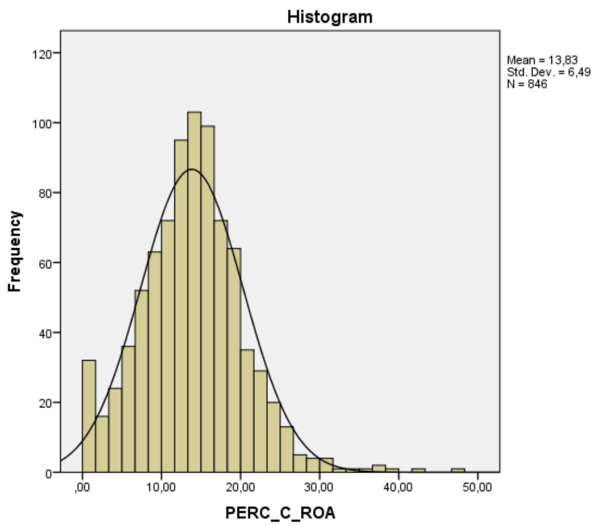


-

Share of bicycle roads

The share of bicycle roads indicates to what extent a postal code area provides a bicycle specific infrastructure. This bicycle road share is given in the percentage of the total bicycle roads length in the total roads length of a postal code area. On average the share of bicycle roads in the postal code areas in Zuid-Holland and Overijssel is 13,83%, where the maximum of this share is 47,83%. In comparison with the previous variables, the data of this variable tends to be more normally distributed. This is noticeable in the frequency histogram and can also be concluded from the relatively low positive skewness value of 0,442 (Figure 26).

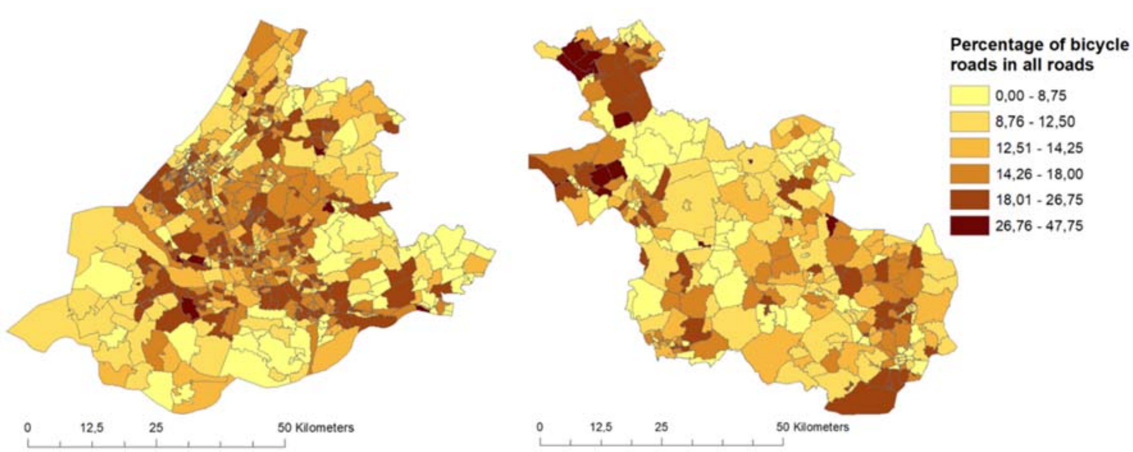
Figure 26: Frequency histogram and descriptive statistics for the share of bicycle roads (PERC_C_ROAD).



N	Valid	846
	Missing	0
Minimum		0
Maximum		47,75
Mean		13,83
Median		13,89
Skewness		0,442
Std. Error of Skewness		,084

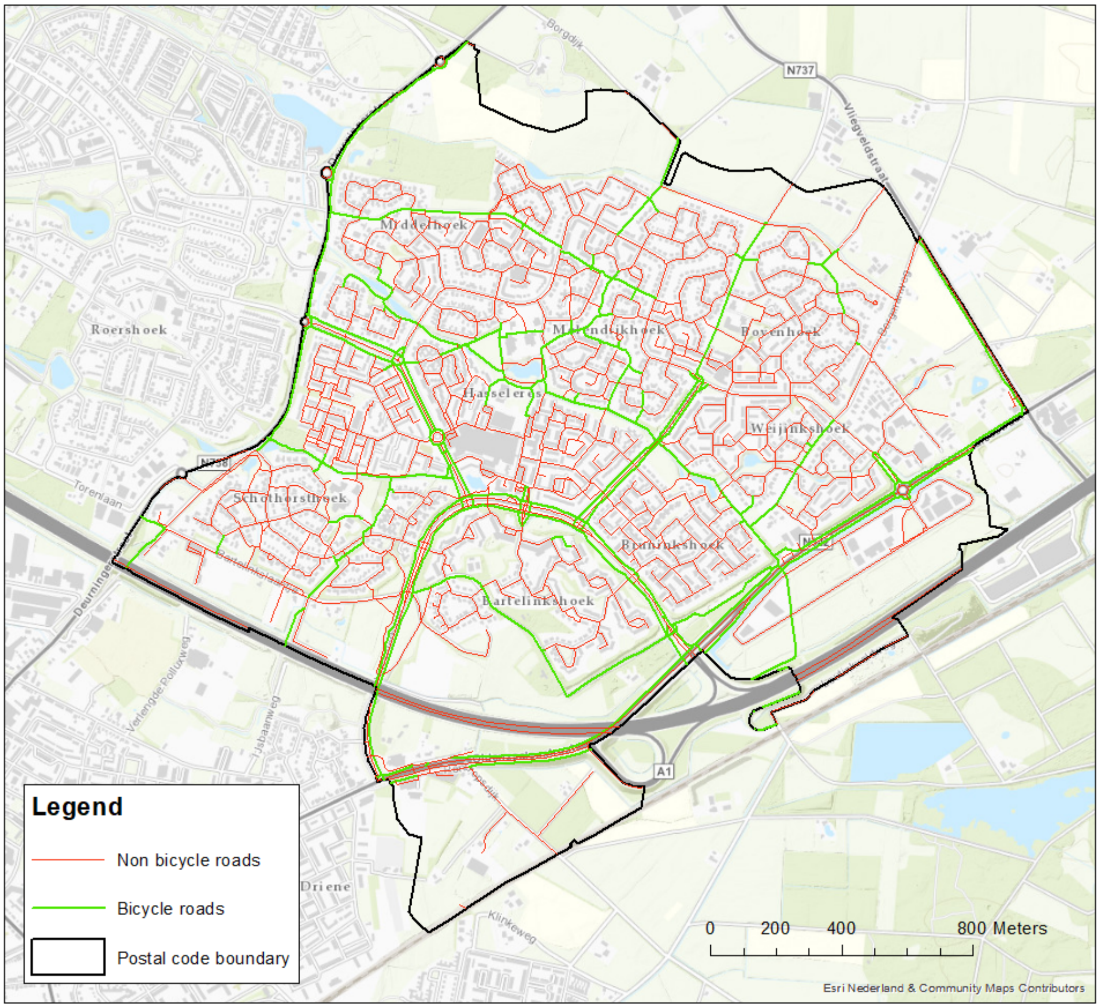
How the share of bicycle roads is spatially distributed is shown in Figure 27. The map shows that cities to a certain extent have a substantially higher share of bicycle roads than rural areas. The spread of high bicycle road share however is not centred mainly in the big cities but tends to be more spread.

Figure 27: Spatial distribution of the share of bicycle roads given in percentage (PERC_C_ROAD).



The map on this page shows an example of the division of bicycle and non-bicycle specific road segments in the postal code area 7558 (Figure 28). The map makes it visible where bicycle road segments are located in the area. They are mainly located next to major roads on a separate bicycle lane and in between neighbourhoods, where bicycle paths connect the neighbourhoods.

Figure 28: Bicycle and non-bicycle roads in the OSM network in postal code 7558 (Overijssel), where the bicycle road share is 25,66%.



Shortest route ratios

The shortest route ratio variables indicate how bicycle friendly a postal code area is by quantifying the route difference in bicycle routes and routes for motorized travel modes. The descriptive statistics for each route ratio is presented in Table 2. Every shortest route ratio variable has a number of missing values, but this is limited to a maximum missing percentage of 2,2% (for the average shortest route ratio variable). For all ratio variables the mean is around 0,90, which indicates that in general cycling routes are slightly shorter than routes for motorized travel modes. There is no general tendency observable in the skewness of the shortest route variables, with all having different skewness values. Remarkable however is the relatively high maximum value of the 1 km shortest route ratio variable, which is a major outlier compared to the maximum value of all the other shortest route ratio variables. The frequency histograms of the variables are presented on the next page in Figure 30.

Table 2: Descriptive statistics for all shortest route ratio variables.

		1 km	2 km	3 km	4 km	5 km	7,5 km	10 km	AVG km
N	Valid	836	835	834	833	833	831	834	827
	Missing	10	11	12	13	13	15	12	19
Minimum		0,03	0,06	0,08	0,10	0,10	0,21	0,25	0,41
Maximum		19,48	3,28	4,61	2,06	1,85	2,58	2,05	4,09
Mean		0,93	0,90	0,90	0,90	0,91	0,91	0,92	0,91
Median		0,92	0,93	0,94	0,95	0,95	0,95	0,96	0,91
Skewness		18,19	1,96	4,01	-0,71	-1,16	0,51	-1,20	8,86
Std. Error of Skewness		0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85

When presenting the spatial distribution of the route ratios for a single distance variable (for example the 1 km ratio), outliers may give an incorrect impression. Therefore here, only the spatial distribution of the average shortest route ratio is visualized. This variable is deducted from all distance variables combined, which gives a more reliable impression of the distribution of shortest route ratios. This map is presented in Figure 29.

Figure 29: Spatial distribution of the average shortest route ratio of bicycle route length in comparison with motorized route length.

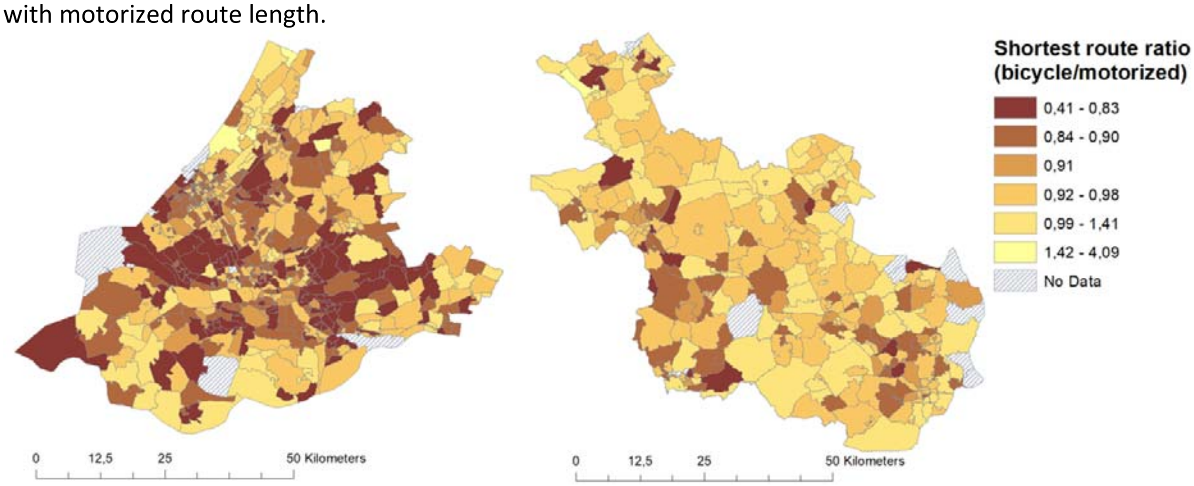
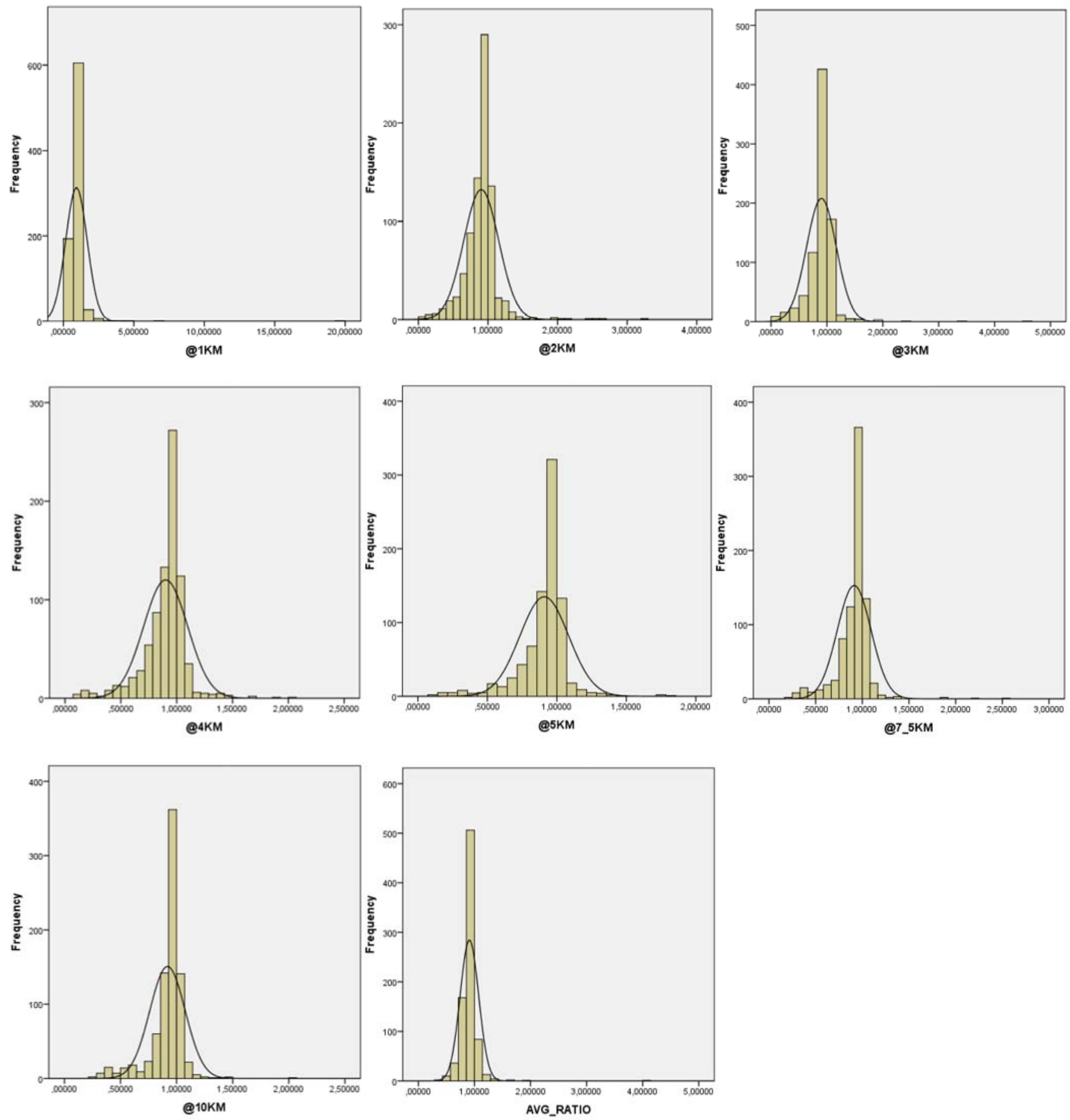


Figure 30: Frequency tables of all shortest route ratio variables.



To give an impression, some example route calculations and the associated route ratios are presented on the next page. Figure 31 shows the calculations for the 4 kilometre ratio in the postal code 8021. The shortest route for motorized travel modes is slightly longer than the shortest route for cyclists, which is approximately 500 meters shorter. The associated route ratio therefore is 0,903 (4752/5261). The other example of postal code 7433 shows the opposite (Figure 32). For the calculation of the 3 kilometre route ratio, the cycling route shows a roughly 400 meters longer route because of the motorway providing motorists a shorter route. The cyclists here need to take a detour to cross the canal, hence resulting in a route ratio of 1,082 (5272/4872).

Figure 31: Calculated routes for the 4 kilometre distance for postal code area 8021 (Overijssel), where the 4 km route ratio is 0,903.

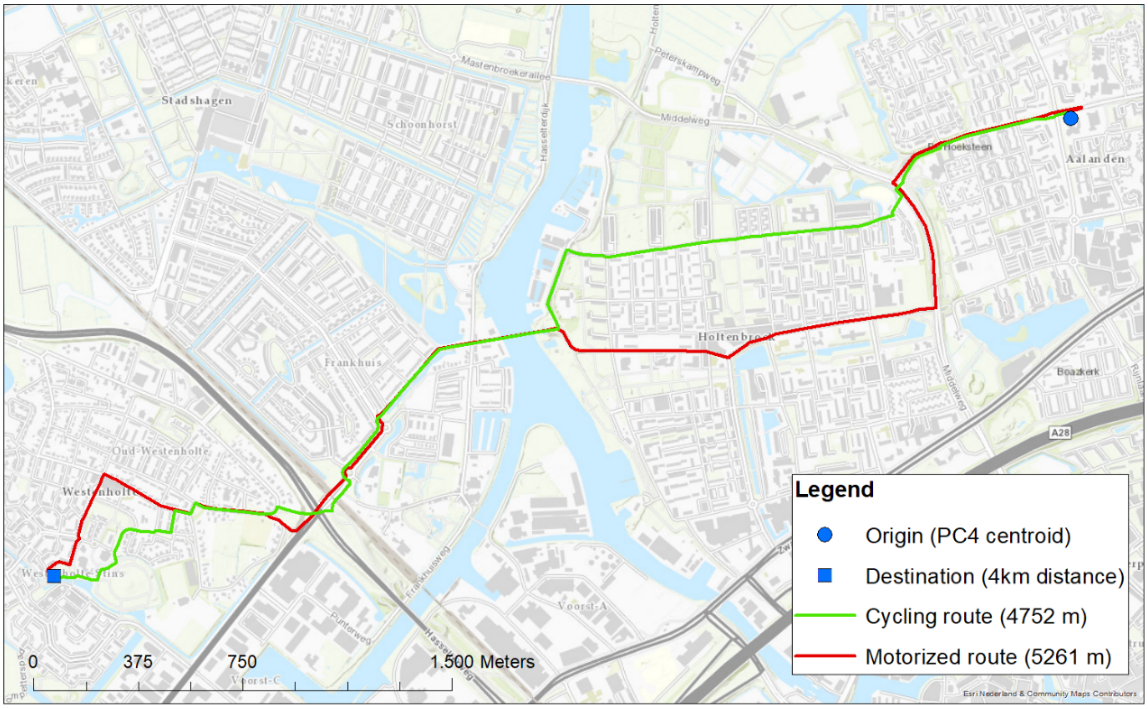
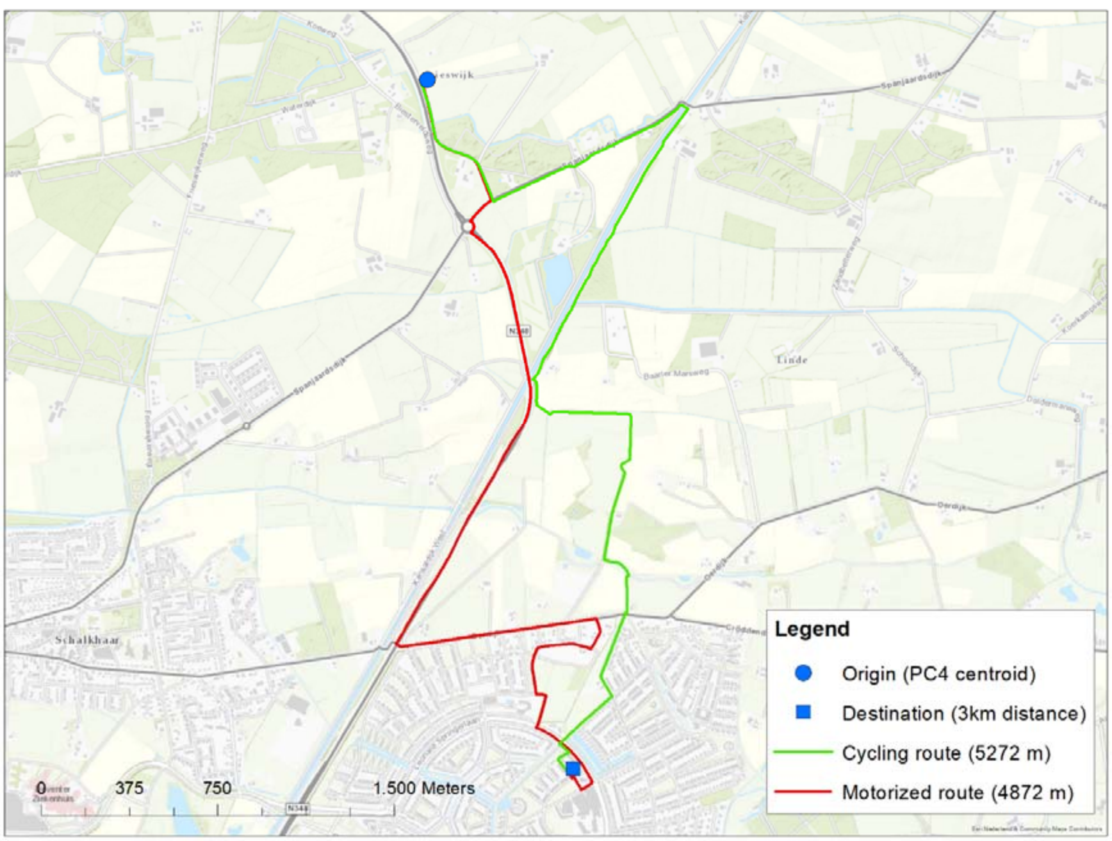


Figure 32: Calculated routes for the 3 kilometre distance for postal code area 7433 (Overijssel), where the 3 km route ratio is 1,082.

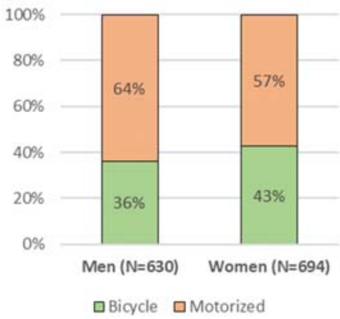


4.3 Bivariate analysis

4.3.1 Control variables

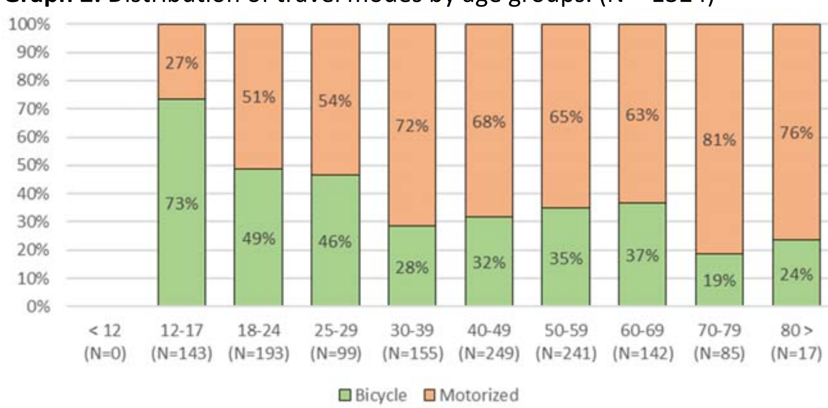
Since the control variables gender, age and household income are nominal and ordinal variables, descriptive statistics do not provide much information. For these variables therefore the distribution between bicycle trips and motorized trips is presented in stacked columns. When separating trips by gender, it turns out that of all trips made by women 43% is made by bicycle (Graph 1). For all trips made by men, only 36% is made by bicycle. Gender thus does tend to influence mode choice.

Graph 1: Distribution of travel modes by gender (N = 1324)



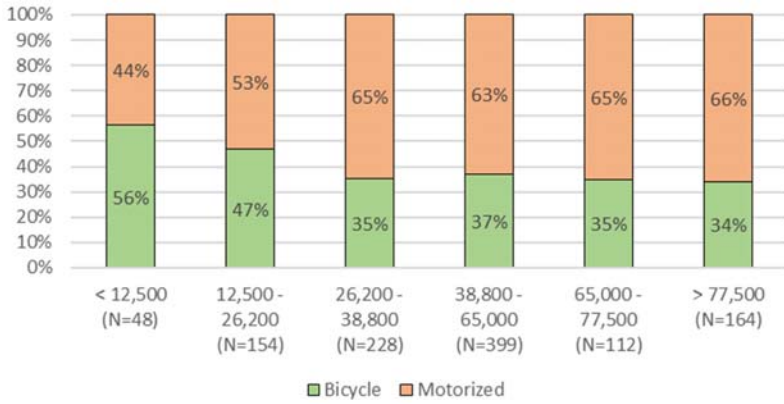
The same accounts for age as controlling variable. Graph 2 shows a trend that the older a person is, the less trips are made by bicycle. For 18-24 year old's, 49% of the trips are made by bicycle compared to only 28% of trips made by 30-39 year old's for example.

Graph 2: Distribution of travel modes by age groups. (N = 1324)



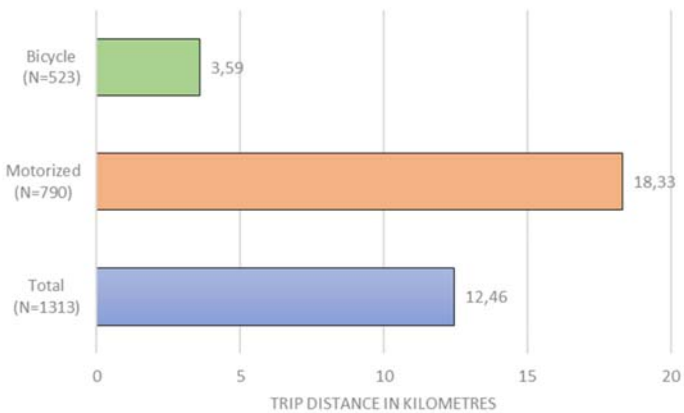
For the control variable household income, the total number of cases (N = 1105) here is lower, because not every respondent has provided their household income. Graph 3 shows a lower percentage of bicycle trips when the household income gets larger For the lower income group (<€12.500), 56% of trips are made by bicycle whereas in the highest income group (>€77.500) only 34% of trips are made by bicycle. Cycling is a relatively affordable mode of travel and therefore accessible for almost everyone, whereas motorized travel modes (in particular cars) are quite costly. This may declare why in the lower income groups, relatively more trips are made by bicycle in comparison with higher income groups.

Graph 3: Distribution of travel modes by household income groups (N = 1105)



Graph 4 shows how the trip distance varies between trips made by bicycle or by a motorized travel mode. As expected, the mean trip distance is considerably higher for trips made by a motorized travel mode. This is no surprise, since long distance trips are less suitable for cycling.

Graph 4: Mean trip distance by travel modes (N = 1313)



4.3.2 Built environment variables

A comparison of the built environment characteristics for bicycle trips and motorized trips is presented in Table 3. It shows the descriptive statistics per built environment variable, both for cycling trips and motorized trips. Most built environment measures differed between cycling and motorized trips in the expected direction. For the density measures for example, the mean population and address density turns out to be reasonably higher for cycling trips in contrast to motorized trips. The same accounts for the mean distances to services, which are all higher for motorized trips than for cycling trips. Crossing density is higher for cycling trips compared to motorized trips. Remarkable is the slightly lower percentage of cycling roads in cycling trips than in motorized trips. The shortest route ratios show to what extent a built environment provides a shorter route for cyclists than for motorized travel. It turns out that for the 1, 2, 3, 4, 5 kilometre distances and the average, the mean ratios are indeed lower in the observed cycling trips than in motorized trips. Contrary are the route ratios for the 7,5 and 10 kilometre distance, where the mean of ratios are higher in the observed cycling trips. This might have to do with shorter direct motorized routes compared to cycling routes when the distance gets bigger (e.g. main roads and highways)

Table 3: Descriptive statistics of built environment characteristics by trip mode for the research sample.

	Motorized trips (N=800)		Cycling trips (N=524)		Total trips (N=1324)	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
POP_DENS	3403	3174	4014	3500	3645	3319
OAD	1913	1273	2235	1411	2041	1338
DIST_SUPER	,81	,53	,76	,33	,79	,47
DIST_DAILY	,72	,48	,64	,34	,69	,43
DIST_CAFE	1,23	,91	1,14	,92	1,20	,92
CROSS_DENS	270	180	306	200	284	189
PERC_C_ROAD	14,40	4,72	14,13	5,00	14,29	4,83
X.1KM	,86	,41	,84	,25	,85	,36
X.2KM	,86	,21	,86	,22	,86	,21
X.3KM	,89	,24	,86	,22	,88	,23
X.4KM	,90	,21	,88	,21	,89	,21
X.5KM	,89	,18	,87	,18	,89	,18
X.7_5KM	,89	,17	,92	,17	,90	,17
X.10KM	,91	,15	,92	,14	,91	,15
AVG_RATIO	,89	,12	,88	,11	,88	,11

In Tables 4, 5 and 6, the correlation between bicycle use, and the control and built environment variables is examined. The strength and significance of each relationship indicates whether a relationship exists or not, whether this relationship is positive or negative and how strong the relationship tends to be (Field, 2013). The Spearman's Rho correlation coefficients have been calculated, because of the dichotomous nature of the dependent variable bicycle use. The associated significance values are included as well, indicating whether or not a relationship is statistically significant.

Table 4: Correlation between bicycle use and control variables gender, age, household income and trip distance.

	GENDER	AGE	HH_INCOME	TRIP_DIST
Spearman's Rho Correlation (rs)	0.066**	-0,221**	-0,080**	-0.477*
Sig. (2-tailed)	0.016	0.000	0.008	0.000
N	1324	1324	1105	1313

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

This table (4) shows that bicycle use is quite strongly correlated to age and trip distance, while not so strong with gender and the household income. All four correlation coefficients however are significant at $p<0,01$. Of these control variables, trip distance is most correlated with bicycle use, where rs=-0,477.

Table 5: Correlation between bicycle use and built environment variables population density, address density, distance supermarket, distance daily grocery store, distance cafe, percentage cycling roads and crossing density.

	POP_DENS	OAD	DIST_SUPER	DIST_DAILY	DIST_CAFE	CROSS_DENS	PERC_C_ROAD
Spearman's Rho Correlation (rs)	0.081**	0.089**	0.014	-0.042	-0.020	0.069*	-0.004
Sig. (2-tailed)	0.008	0.003	0.652	0.166	0.503	0.023	0.896
N	1085	1085	1085	1085	1085	1085	1085

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

From this second correlation table (5), similar conclusions can be drawn as from Table 3. A positive correlation between bicycle use and population, address and crossing density can be found. Being significant at *p*<0,01 for the address density and population and crossing density for *p*<0,05. The strength of the relationships with coefficients between 0,069 and 0,089 tend to be not that strong. A weak positive relationship between bicycle use and the distance to a supermarket is statistically insignificant (rs=0,014). Negative relationships are to be found with the distance to a daily grocery store and café and the percentage cycling roads. This means that if the distance to services or percentage cycling roads increases, bicycle use tends to be lower. Only the coefficient of the distance to a daily grocery store and cafe turns out to be significant, for *p*<0,05.

Table 6: Correlation between bicycle use and all shortest route ratio variables, including the average shortest route ratio.

	SR_1KM	SR_2KM	SR_3KM	SR_4KM	SR_5KM	SR_7.5KM	SR_10KM	AVG_RATIO
Spearman's Rho Correlation (rs)	0.002	-0.006	-0.050	-0.043	-0.091**	0.052	-0.011	-0.010
Sig. (2-tailed)	0.948	0.842	0.100	0.156	0.003	0.088	0.719	0.734
N	1085	1085	1085	1085	1085	1085	1085	1085

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

As noticeable here in Table 6, same as for Table 3, the relationships for most of the shortest route ratios is very slightly negative (2, 3, 4, 5, 10 kilometre and average). A higher ratio thus imply a very slightly lower bicycle use. From these distance variables, only the 5km distance is significant for *p*<0,01, whereas none of the other negative relationships are significant at this level or even at *p*<0,05. The shortest route ratios for the distances of 1 and 7,5 kilometre on the other hand show a positive relationship, where both of the relationships are statistically insignificant.

4.3.3 Correlation matrix

Appendix E contains a correlation matrix of all variables included (dependent, control and predicting variables). Some of the independent variables show a significant and high correlation value. To avoid multicollinearity, independent variables having a highly correlated relationship are better both included in a regression model. Address and population density for example show a strong positive relationship where Spearman's Rho=0,821 for *p*<0,01. Both population density and address density also show a strong positive correlation with the design measure crossing density (e.g. 0,893 and 0,742 for *p*<0,01). Between the diversity variables which indicate the proximity to services (supermarket, daily grocery store and cafe), rather strong relationships are also found (0,512, 0,672 and 0,787 for $p<0,01$)

4.4 Multivariate analysis: logistic regression

4.4.1 Regression without predictors

A logistic regression model is performed to estimate the outcome variable bicycle use by means of including predicting variables. The predicting variables are the control variables and built environment variables, all expected to influence an individual's probability to use a bicycle as transport mode instead of using an individual motorized travel mode. Depending on the model fit and predictive power of a model, certain variables will be included while others might be excluded. As explained in section 3.7.3, the Enter method is used to determine which variables are included and which are not. Looking at the significant change of the deviance of each variable, variables will be added or excluded from the regression model.

Since population density, address density and crossing density are ratio measures on a relatively large scale, these variables are aggregated on thousands (POP_DENS and OAD) and hundreds (CROSS_DENS). This has no effect on the measures of fit or predictability of the models. This only ensures that the resulting odd ratios are not very close to 1,000, making it impossible to show the effect of the variables.

The simplest model is a regression function where only the intercept is included (constant B). No predicting variables at this stage are included in the regression model. The percentage of correctly predicted outcomes in this simple model fully depends on the distribution of the dependent variable bicycle use. This simple model is able to predict 62,1% of the outcomes correctly.

4.4.2 Regression model 1

First, the regression function is expanded by adding just the control variables gender, age, household income and trip distance (Table 7). Compared to the simple model with only intercept B, including the control variables gives a Chi-squared ($\chi 2$) of 335,55. This indicates that the deviance (-2LL) has decreased with that number, indicating that the model including the control variables better fits to the data than the simple model. The percentage correctly predicted outcomes has increased from 61,6% to 71,7%. Looking at the proportion of the variance in bicycle use explained through gender, age, household income and trip distance, this proportion is approximately 35,8% (Nagelkerke $R^2 = 0,358$). The Hosmer and Lemeshow-test is insignificant with a p-value of 0,116, thus de model does fit to the data. The coefficients of the variables gender and household income however are insignificant, thus their effect on bicycle use cannot be interpreted.

4.4.3 Regression model 2

The second regression model is built by means of adding variables based on statistical tests (Table 7). For every step, the variable with the highest significant change in deviance (-2 Log Likelihood) is added until the step where the coefficient of any variable in the model turns insignificant. This model has included the control variables age and trip distance, as a logic result of these being highly significant in model 1 ($p<0,01$). Of the built environment variables the address density, shortest route ratio 5 kilometre and shortest route ratio 7,5 kilometre are included. The proportion of variance in bicycle use explained through the predicting variables in this model is 36,9%, which is higher than the simple model with 35,8%. Where the percentage correctly predicted outcomes is 71,1% in model with only control variables, this is 72,8% in model 2. The Chi-squared test has a value of 416,98, indicating that this model fits better to the data than model 1 (where $\chi 2$=335,55). The Hosmer and Lemeshow-test is insignificant for p>0,05, which means that the model fits to the data.

All variables included in the model are significant, SR7,5km for p<0,05 and all other variables for p<0,01. The coefficients in the model can be interpreted as follows:

- AGE: For each age group you go up, the odds of using a bicycle are 29,9% lower than using a motorized travel mode. ($\frac{1}{Exp(B)}$ = 1/0,770 = 1,299).
- TRIP_DIST: For each extra kilometre in trip distance, the odds of using a bicycle compared to the odds of using a motorized travel mode decrease by 17,2% (1/0,853 = 1,172).
- OAD: For every 1000 addresses increase in address density, the odds of using a bicycle increase by 17%.
- SR5KM: For every 1 unit increase in the 5km shortest route ratio, the odds of using a bicycle decrease by 321,1% (1/0,302 = 3,211).
- SR7,5KM: For every 1 unit increase in the 5km shortest route ratio, the odds of using a bicycle increase by 294,9%.

The last two relatively large odds ratios for the 5km and 7,5km shortest route ratios sound remarkable, but taking into account the small scale of measure of the route ratios puts this number in perspective. A route ratio of 1 indicates the same route distance for bicycle as for motorized travel. A route ratio of 2 (e.g. a one unit increase) indicates a two times longer bicycle route than motorized travel route in that postal code area (shortest route ratio = distance bicycle route/distance motorized route). A 321,1% decrease in odds to take a bicycle as travel mode in that perspective is not surprising. The 294,9% increase in odds for the 7,5km route ratio however is very remarkable. This means that when the ratio gets higher (indicating bike route is getting longer than motorized route), increases the odds of using a bicycle as mode of transport.

Table 7: Regression model coefficients for model 1 (only control variables) and model 2 (OAD, SR5km and SR7,5km controlled for age and trip distance)

		Model 1 (N=1096)			**Model 2** (N=1308)		
	Variables	**B**	***p*-value**	**Exp(B)**	**B**	***p*-value**	**Exp(B)**
Constant	BICYCLE USE	2,363**	,000	10,619	2,828**	,000	5,967
Control Variables	GENDER[1]	,044	,765	1,045			
	AGE	-,270**	,000	,764	-,261**	,000	,770
	HH_INCOME	-,085	,118	,918			
	TRIP_DIST	-,174**	,000	,839	-,159**	,000	,853
Built environment variables	OAD				,157**	,002	1,170
	SR5km				-1,197**	,002	,302
	SR7,5km				1,081*	,010	2,949
Nagelkerke R^2		**R^2 = 0,358**			**R^2 = 0,369**		

**. Significant at the 0.01 level (2-tailed).
*. Significant at the 0.05 level (2-tailed).
1) Male as reference group

4.4.4 Regression model 3

For this regression model, variables are included based on theoretical concepts (from the literature review) and logic thinking. The literature review has shown the importance of all three dimensions of the built environment: density, diversity and design. Design here is the most challenging, but also less researched aspect when it comes to actual bicycle use and therefore the most important. The correlation matrix presented in the bivariate analysis (section 4.3.2) has shown which independent variables strongly correlate to each other and thus should not be included both to avoid multicollinearity. Then regression model 1 has shown that the coefficients of gender and household income are insignificant and thus can be excluded from a model.

Therefore, in this last model, the control variables age and trip distance are included (Table 8). For the density dimension, address density is included. Address density is expected to better represent a built environment's density compared to the population density since it indicates the density of the actual infrastructure. For the diversity dimension, the average distance to the nearest daily grocery store is included. A low distance to daily grocery shops indicates a diverse built environment. Because of the importance of the design variable, two variables are included: Percentage cycling roads and the average shortest route ratio. Theory supports the expectation that a transport infrastructure providing amenities to cyclists makes people more willing to use a bicycle as travel mode. The average shortest route ratio is included because of the more generalized measurement compared to the single distance shortest route ratio's (average of all shortest route ratio variables),

Compared to model 2, this model has a slightly lower Chi-squared value ($\chi 2$=403,79) and the percentage correctly predicted values has decreased to 71,9% (compared to 72,8% in model 2). Nagelkerke R Squared is the same as for model 2, for both models 36,9%. The Hosmer and Lemeshow-test on the other hand is significant at p<0,01, indicating that the model does not fit to the data well. For large sample sizes however, the Hosmer and Lemeshow-test should be treated with care since the test is known to indicate significance in cases where the model in fact does fit to the data.

From the variables included, only AGE, TRIP_DIST and OAD are significant. This means that the coefficients for DIST_DAILY, PERC_C_ROAD and AVG_RATIO should not be interpreted. The next relationships are found in the model:

- AGE: For each age group you go up, the odds of using a bicycle are 29,9% lower than using a motorized travel mode. (1/0,770 = 1,299).
- TRIP_DIST: For each extra kilometre in trip distance, the odds of using a bicycle compared to the odds of using a motorized travel mode decrease by 17% (1/0,855 = 1,170).
- OAD: For every 1000 addresses increase in address density, the odds of using a bicycle increase by 15,5%.

Table 8: Regression model 3 (OAD, DIST_DAILY, PERC_C_ROAD and AVG_RATIO controlled for age and trip distance)

		Model 3 (N=1306)		
	Variables	**B**	***p*-value**	**Exp(B)**
Constant	BICYCLE USE	2,596**	,000	13,409
Control Variables	AGE	-,261**	,000	,770
	TRIP_DIST	-,157**	,000	,855
Built environment variables	OAD	,144*	,015	1,155
	DIST_DAILY	-,129	,500	,879
	PERC_C_ROAD	-,015	,296	,986
	AVG_RATIO	-,662	,279	,516
Nagelkerke R^2		**R^2 = 0,360**		

**. Significant at the 0.01 level (2-tailed).
*. Significant at the 0.05 level (2-tailed).

4.4.5 Model performance

Looking at the performance and fit of each model, it has to be concluded that regression model 2 performed best. Model 2 however lacks inclusivity of built environment variables from each dimension (density, diversity, design). Model 3 aimed to include those built environment variables from each dimension, but failed to improve the performance and fit of the model. Three of the four included built environment variables turned out to be insignificant, making it irrelevant to interpret their coefficients. The overall tendency is that adding more variables to a model increased the level of insignificance of the variables which makes it impossible to statistically interpret the coefficients.

5 Conclusion

This research has aimed to provide insight in the effect of the built environment on bicycle use. Many researches have indicated the importance of built environment characteristics on travel mode choice. Revealed preference studies have proven the importance of bicycle friendly infrastructures on mode choice. In this study, actual travel data and GIS methods are used to analyse the effect of the built environment on bicycle use.

The sub research questions developed in the introduction have acted as a guidance for the research process. First the research subject and variables were defined and operationalized, followed by the measurements of the variables in GIS. Subsequently the built environment characteristics have been joined with the travel data. The actual effect of the built environment on bicycle use in the end is analysed in detail. This complete research process has contributed to answering the main research question. Both the sub and main research questions now will be addressed.

1. How can bicycle use be defined in this research?
Bicycle use is considered as one of the many travel mode choices. This research has made use of travel data obtained by the Netherlands Institute for Transport Analysis, containing movements of respondents. All trips in the research sample were coded as bicycle use, only if a complete journey (from origin to destination) was made by bicycle. Since bicycle use is opposed to travel by motorized travel modes, non-bicycle use in this research is coded for all trips made by car, motorbike or scooter.

2. How can the built environment be defined in this research and what measures can represent cycling friendliness?
The built environment in simple terms is defined as the physical environment created for human activities, including the designed buildings, infrastructure and land uses. It is most clearly explained by differentiating a built environment's density, diversity and design (3D's). Features of the three dimensions indicate differences in the built environment. A low density and highly homogenous land use area represents a total different built environment compared to a very dense and mixed land use area. Measures representing these features are population and address density for the density dimension. The diversity dimension is measured by the average distance to the nearest supermarket, daily grocery store and café. A built environment's design is measured by the crossing density, percentage of cycling road length and shortest route ratios which indicate to what extent the road infrastructure provides a shorter route for cycling than for individual motorized travel modes. Individually and collectively, these measures indicate the cycling friendliness of an area.

3. How to derive built environment measures from spatial datasets using Geographic Information Systems (GIS)?
Collectively, datasets from Statistics Netherlands, the Cyclists Union and OpenStreetMap contain a lot of spatial information, useful for obtaining measures of the built environment. ArcMap's GIS environment with Model Builder is used to extract and calculate the exact measures for each built environment. The Model Builder creates workflows for input data, geoprocessing tools and output data to ease the process where multiple tools are used.

4. How to integrate both bicycle use and built environment data in a statistical model?
The built environment measures derived in ArcMap are added to the travel dataset by the MPN, based on the 4 digit postal code of the respondent. Due to privacy principles, the MPN adds these built environment measures to the travel data by introducing a noise on the measure's values of maximum 1% from their original value. A sample of the resulting dataset is analysed in SPSS. By means a univariate analysis, bivariate analysis and multivariate analysis (logistic regression), the effect of the built environment on bicycle use is examined.

5. What is the outcome of the model and what does it say about the built environment and bicycle use?
The regression analysis has proved that the built environment in fact does influence bicycle use. Even though personal characteristics (age) and trip characteristics (trip distance) showed a significantly higher improvement of model predictability than the built environment variables, the models still did improve after including certain built environment variables. In the model based on theory, the address density, distance to a daily grocery store, cycling roads and average route ratio were included. All variables however turned out to be insignificant, except for address density. A regression model with built environment variables address density and shortest route ratios 5km and 7,5km on the other hand did show significant effects. A higher address density increases the odds of bicycle use, same as for a higher shortest route ratio 5km. However unexpectedly, a higher shortest route ratio 7,5km decreases the odds of bicycle use.

Main research question: *To what extent is bicycle use affected by the built environment?*
Even though control variables age and trip distance had a significantly larger effect on bicycle use, this research has proven that the inclusion of built environment variables did improve the predictability of bicycle use. Therefore the built environment does, in this research slightly, affect bicycle use. The approximate variance in bicycle use explained by the regression models increased from 35,8% for the model with only the control variables to 36,9% for the model including built environment variables. Same accounts for the percentage correctly predicted values by the models: 71,1% for the model only including control variables and 72,8% for the model including built environment variables controlled for age and trip distance. Especially density demonstrated a significant effect on bicycle use, as well as the density's shortest route ratio 5 kilometre. In addition to this, the bivariate analyses for the 1, 2, 3, 4, 5 kilometre ratio and average ratio, the mean ratio was lower for the cycling trips than for motorized trips. However, overall the shortest route ratios turned out to be not consistent enough to give a convincing indication of their effects on bicycle use in the regression models.

6 Discussion

This research has aimed to examine the effect of built environment features on bicycle use. Researches in the cycling behaviour study field currently to some agree used GIS to address the built environment, but on the other hand lacked the use of actual travel data. Alternatively they used methods of revealed preferences of respondents in terms of a built environment's infrastructure and the corresponding (possible) preference for travel mode choices. This research has furtherly explored the relationship by trying to predict bicycle use based on quantified built environment measures and actual travel data. This shows the potential of using GIS in terms of examining which factors are important determinants in travel mode choices.

Even though GIS does have many benefits, it is still considered as a challenge to quantify the built environment in precise measures (Winters et al., 2010). It is known that inconsistencies in research results may stem from methodological issues in GIS practices. The inclusivity of the shortest route ratios for example provide a rather new method of quantifying a built environments cycling infrastructure, but also poses challenges. The long geo processing time of creating and calculating the route ratios has limited the number of routes per distance variables. The ratios are calculated from one (in fact two: motorized route and cycling route), route per postal code area which gives a certain space for errors in the form of outliers. Therefore it is recommended for future research to calculate these shortest route ratios by means of an average of an x number of routes. This will improve the consistency and accuracy of the difference in shortest routes for cyclists and motorized travel.

Consensus on how to define the area of interest for a respondent's trip has not been reached yet in the travel behaviour research field. The geographic scale of an individual's activity space can vary considerably from person to person (Winters et al., 2010). Even though it is argued that individuals mostly travel from or to their residential location, it does not imply that their activity space mainly takes place in this area. Adding to this that the 4 digit postal code areas, determined as respondent's residential location, do vary in area size. Thus for some respondents, a relatively small part of their travel takes place within their residential area. A solution to overcome this issue is to use buffers (fixed distance) to indicate a respondent's residential area.

By the Netherlands Institute for Transport Policy Analysis, a noise has been applied to the calculated built environment measures meaning that the original values were changed with maximum 1%. The exact effect on the precision of the analysis results however stay unclear.

Future travel mode choice research should built further upon the importance of the built environment by focussing on the design. This research has proven again that dense and diverse environments are correlated to a higher propensity of bicycle use. A built environment's design and road infrastructure also tends to be related to bicycle use, but this relationship could not fully revealed. Quantifying the design of a built environment and examining its effect on the actual use of bicycles as travel mode does still provide a challenge. It is recommended to aim attention to route differences and infrastructural amenities between travel modes, since these factors mainly determine the cycling friendliness of a built environment.

7 Acknowledgements

In this publication we made use of data from the Netherlands Mobility Panel, administered by KiM Netherlands Institute for Transport Policy Analysis.

Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org

8 References

Akar, G., & Clifton, K. (2009). Influence of Individual Perceptions and Bicycle Infrastructure on Decision to Bike. *Transportation Research Record: Journal of the Transportation Research Board*, *2140*(October), 165–172. https://doi.org/10.3141/2140-18

Axhausen, K. W. (2007). *Concepts of Travel Behavior Research*. Zurich.

Boarnet, M., & Crane, R. (2001). The influence of land use on travel behavior: specification and estimation strategies. *Transportation Research Part A*, *35*, 823–845. Retrieved from www.elsevier.com/locate/tra

Bryman, A. (2012). *Social Research methods*. *Journal of Experimental Psychology: General* (4th Editio). Oxford University Press.

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research Part A*, *43*, 548–559. https://doi.org/10.1016/j.tra.2009.02.001

Caulfield, B., Brick, E., & McCarthy, O. T. (2012). Determining bicycle infrastructure preferences - A case study of Dublin. *Transportation Research Part D: Transport and Environment*, *17*(5), 413–417. https://doi.org/10.1016/j.trd.2012.04.001

Centraal Bureau voor de Statistiek. (2015). Statistische gegevens per vierkant en postcode. Den Haag: Centraal Bureau voor de Statistiek, Esri Nederland. Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische data/kaart-met-statistieken-per-vierkant-van-100-bij-100-meter

Centraal Bureau voor de Statistiek. (2019). CBS StatLine - Regionale kerncijfers Nederland. Retrieved April 1, 2019, from https://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=70072NED&D1=0,56&D2=0,5-16,103,111,126,143,212,225-226,308,315,327,360,365,405,420,428,474,506,566,572,584,608,621-622,689,728,791&D3=7,12,l&HDR=T&STB=G1,G2&VW=T

Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D*, *7*(4), 265–284. Retrieved from www.elsevier.com/locate/trd

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Chen, C., Gong, H., & Paaswell, R. (2008). Role of the built environment on mode choice decisions: additional evidence on the impact of density. *Transportation*, *35*, 285–299. https://doi.org/10.1007/s11116-007-9153-5

Clifton, K. J., & Handy, S. L. (2001). *Qualitative Methods in Travel Behaviour Research*. Retrieved from https://pdfs.semanticscholar.org/884f/1f779104dfea685c76be61309a2f9fe116d4.pdf

Ewing, R., & Cervero, R. (2001). Travel and the Built Environmental: a Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, (1780), 87–114. Retrieved from http://filecenter.santa-clarita.com/Planning/VistaCanyon/DraftEIR/Volume3/Apx_4_3i_Travel_Built_Environment.pdf

Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta analysis. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Field, A. (2013). *Discovering Statistics Using SPSS* (4th Editio). London: SAGE Publications Ltd. https://doi.org/10.1111/j.1365-2648.2007.04270_1.x

Fietsersbond, De (2019). Network Dataset Fietsersbond. Retrieved January 29, 2019, from http://www.bikeprint.nl

Gärling, T. (2004). *Changes of Private Car Use in Response to Travel Demand Management*. Retrieved from https://pdfs.semanticscholar.org/c021/5aa661f53599ab0ac29903d18c533506cf31.pdf

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine*, *23*(2 SUPPL. 1), 64–73. https://doi.org/10.1016/S0749-3797(02)00475-0

Harms, L., & Kansen, M. (2018). *Fietsfeiten*. Kennisinstituut voor Mobiliteitsbeleid | KiM: Den Haag. Retrieved from https://www.kimnet.nl/publicaties/brochures/2018/03/16/fietsfeiten

Hellmann, H. (2016). *Opportunities for Dutch cycling enterprises in Germany*. Delft. Retrieved from www.dutchcycling.nl

Hoogendoorn-Lanser, S. (KiM N. I. for T. P. A. (2015). Mobiliteitspanel Nederland (MPN 2014). https://doi.org/http://dx.doi.org/10.17026/dans-zyc-7qfv

Hoogendoorn-Lanser, S., Schaap, N. T. W., & Oldekalter, M.-J. (2015). The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. *10th International Conference on Transport Survey Methods, Transportation Research Procedia*, *11*, 311–329. https://doi.org/10.1016/j.trpro.2015.12.027

Hull, A., & O'Holleran, C. (2014). Bicycle infrastructure: can good design encourage cycling? *Urban, Planning and Transport Research*, *2*(1), 369–406. https://doi.org/10.1080/21650020.2014.955210

ITC. (2013). *The core of GIScience: a process-based approach*. Enschede: Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente.

John, R. (2018). Linear Regression with TensorFlow Canned Estimators. Retrieved April 29, 2019, from https://medium.com/coinmonks/linear-regression-with-tensorflow-canned-estimators-6cc4ffddd14f

Klinkenberg, J., & Bertolini, L. (2013). There are still opportunities for Dutch cycling, (3), 1–9. Retrieved from http://www.verdus.nl/upload/documents/There_are_still_opportunities_for_Dutch_cycling_Klinkenberg_Bertolini.pdf

Lu, Y., Xiao, Y., & Ye, Y. (2017). Urban density, diversity and design: Is more always better for walking? A study from Hong Kong. *Preventive Medicine*, *103*, S99–S103. https://doi.org/10.1016/j.ypmed.2016.08.042

Maat, K. (2010). Verstedelijking en mobiliteit: de VS versus Nederland (pp. 1–15). Roermond. Retrieved from https://www.cvs-congres.nl/cvspdfdocs/cvs10_028.pdf

Mccormack, G. R., & Shiell, A. (2011). *In search of causality: a systematic review of the relationship between the built environment and physical activity among adults*. https://doi.org/10.1186/1479-5868-8-125

Mivule, K. (2013). *Utilizing Noise Addition for Data Privacy, an Overview*. Bowie. Retrieved from https://arxiv.org/ftp/arxiv/papers/1309/1309.3958.pdf

Moudon, A. V., Lee, C., Cheadle, A. D., Collier, C. W., Johnson, D., Schmid, T. L., & Weather, R. D. (2005). Cycling and the built environment, a US perspective. *Transportation Research Part D*, *10*, 245–261. https://doi.org/10.1016/j.trd.2005.04.001

Navlani, A. (2018). Understanding Logistic Regression in Python. Retrieved December 2, 2018, from https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python

OpenStreetMap. (2018). About OpenStreetMap. Retrieved December 1, 2018, from https://www.openstreetmap.org/about

OpenStreetMap Contributors. (2019). Planet dump [Data file from 29-01-2019]. Retrieved January 29, 2019, from http://download.geofabrik.de/europe.html

Pucher, J., & Buehler, R. (2012). *City Cycling*. MIT Press. Retrieved from http://dannyreviews.com/h/City_Cycling.html

Rietveld, P., & Daniel, V. (2004). Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A*, *38*, 531–550. https://doi.org/10.1016/j.tra.2004.05.003

Rijksoverheid. (2016). Agenda Fiets 2017-2020, 1–19. Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2017/02/16/agenda-fiets-2017-2020

Rijksoverheid. (2018). Fietsbeleid. Retrieved October 4, 2018, from https://www.rijksoverheid.nl/onderwerpen/fiets/fietsbeleid

Rodriguez, D. A., & Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D*, *9*, 151–173. https://doi.org/10.1016/j.trd.2003.11.001

Sarkar, P. P., Mallikarjuna, C., Bordoloi, R., & Mote, A. (2013). Quantification of Land Use Diversity in The Context of Mixed Land Use. *Procedia - Social and Behavioral Sciences*, *104*, 563–572. https://doi.org/10.1016/j.sbspro.2013.11.150

Sieben, I. (n.d.). *Logistische regressie analyse : een handleiding*.

Stradling, S. G. (2011). Travel mode choice. *Handbook of Traffic Psychology*, (December 2011), 485–502. https://doi.org/10.1016/B978-0-12-381984-0.10034-7

Van Acker, V., Van Wee, B., & Witlox, F. (2010). When Transport Geography Meets Social Psychology: Toward a Conceptual Model of Travel Behaviour. *Transport Reviews*, *30*(2), 219–240. https://doi.org/10.1080/01441640902943453

Vandenbulcke, G., Dujardin, C., Thomas, I., Geus, B. DE, Degraeuwe, B., Meeusen, R., & Int Panis, L. (2009). *Cycle commuting in Belgium: Spatial determinants and "re-cycling" strategies*. Brussels. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.645.4139&rep=rep1&type=pdf

Wardman, M., Tight, M., & Page, M. (2007). Factors influencing the propensity to cycle to work. *Transportation Research Part A*, *41*, 339–350. https://doi.org/10.1016/j.tra.2006.09.011

Winters, M., Brauer, M., Setton, E. M., & Teschke, K. (2010). Built Environment Influences on Healthy Transportation Choices: Bicycling versus Driving. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *87*(6), 969–993. https://doi.org/10.1007/s11524-010-9509-6

9 Appendices

Appendix A: Operationalisation formulas of each shortest route ratio variable:

SR1km = distance shortest cycling route / distance shortest motorized travel route
SR2km = distance shortest cycling route / distance shortest motorized travel route
SR3km = distance shortest cycling route / distance shortest motorized travel route
SR4km = distance shortest cycling route / distance shortest motorized travel route
SR5km = distance shortest cycling route / distance shortest motorized travel route
SR7.5km = distance shortest cycling route / distance shortest motorized travel route
SR10km = distance shortest cycling route / distance shortest motorized travel route
SR_AVG = (SUM(all shortest route ratio's)) / 7

Appendix B: Model Builder explanation

Share of bicycle roads model (Figure 8)

The first step in the model is the iteration through all PC4 areas. In the example of Figure 8, the model will iterate through all postal code areas in Overijssel. The output of the iteration is the polygon of the feature selected, for the first iteration for example postal code area 7216. The postal code number of an iteration is also displayed in the Value box.

To make road segment calculations specifically for one postal code area, the road network within that postal code area needs to be extracted from the OSM road infrastructure shapefile. The *Select Layer By Location tool* selects all road segments of the Overijssel_OSM file that have their centre within the polygon of the postal code area iteration. The result is a road network file of a postal code area.

Subsequently, from that road network file, both the total length of all roads and the total length of the cycling roads need to be calculated. For both calculations, there is a separate set of actions necessary.

The initial OSM road network shapefile that is used as input in this model is modified beforehand. A variable is added and calculated using *Calculate Geometry*, to end up with an OSM road network file that includes the length of each road segment. To calculate the total length of all road segments in a postal code area, the *Summary Statistics* tool is used. It provides an option to calculate multiple measures of statistics, being one of these the sum. The result is a table where the measure of the sum of the length of all road segments is included. Since the postal code number is lost after using this tool, a new field is added and using the *Field Calculator*, the variable is given the postal code area in the *Value* box in Figure 8, indicating the postal code of the iteration.

The same process is performed to calculate the total length of bicycle roads in a postal code area. There is one extra tool included, so that the cycling roads are separated from the rest of the roads first. Using the *Select Layer By Attribute* tool, all bicycle roads are selected from the road network file of a postal code area. Subsequently, the same steps as for the previous calculation for the total length of all roads is performed for the bicycle roads specifically: *Summary Statistics* for the sum of bicycle road length, *Add Field* and *Field calculator* to add a field containing the postal code in the resulting table.

By using the *Merge* tool, all resulting tables (two for each PC4 area) for both the total bicycle road length and the total length of all roads are merged into two tables: a table with total road lengths for all PC4 areas and a table with total bicycle road lengths for all PC4 areas. Next, the tables are joined based on the PC4 number. A field is added, where the share total bicycle road length is calculated using *Field Calculator*: (Total kilometres of bicycle roads/ total kilometres of roads) * 100. The result is the share of bicycle roads variable, given in percentage.

Bicycle route model (Figure 9)
As explained, the route model iterates through all postal code areas by making use of the *Feature Selection Iterator*. The next step in the model is creating destination points for a postal code area. First a *Multiple Ring Buffer* is created for all route distance variables. This means that the ring buffer consists of rings with a distance of 1, 2, 3, 4, 5, 7.5 and 10 kilometres from a postal code's centroid. To be able to generate random points, the buffer polygons need to be transformed into lines by using the *Polygon To Line* tool. Next, the *Create Random Points* tool generates random points on the Multiple Buffer Line file, with a condition that one point needs to be generated on every line. For the reason that the resulting point file has lost the information about its postal code number, the file is joined with the Multiple Buffer file based on their fields indicating the distance. The result is a set of random points for a postal code area, representing the destinations points for all route distance variables. These destination points are also copied to a folder using *Copy Features* so that they can be used as input in the motorized travel route model.

The bicycle route model needs a *Network Dataset* as input to find the shortest routes in this network. The cycling infrastructure shapefile created for the street connectivity measure is used, but is not perfectly routable due to network errors. The *Planarize* tool is used to split line segments at intersections so that turns can be made correctly at each junction. The *Integrate* tool is used to connect lines that are not yet correctly connected by transforming the line segments that fall within a specified X,Y tolerance (20 centimetres). Finally, the network dataset is created with the *Create Network Dataset* tool in ArcCatalog.

The next step is to create and execute the actual routing part of the model. To calculate the shortest routes, a *Closest Facility Analysis* is performed. This analysis makes it possible to search for one or more facilities closest to an origin, based on a cost. In this case, the distance is taken as the cost. The centroid of a postal code area can be added as the location of the origin, while the previously generated random points can be added as the locations of the facilities. The locations are snapped to the cycling *Network Dataset.* The output is set to be the lines of routes with measures (distance). The *Solve Network Analysis* tool solves the analysis, after which the *Copy Features* tool writes the results to a folder where all routes for all postal code areas will be saved.

Motorized travel route model (Figure 10)
The route model for motorized travel uses the same iteration as the cycling route model. Using the *Feature Selection Iterator*, the model iterates through all the postal code centroids which are later in the model used as origin input for the routes.

As this model is created to calculate routes for motorized travel, a routable *Network Dataset* for motorized travel needs to be created first. The OpenStreetMap road infrastructure file provides the complete road network in the Netherlands and is used as input. The OSM file however is not perfectly routable yet, still includes bicycle lanes and has a different coordinate system. Therefore, first the OSM file is transformed to the RD New coordinate system and adjusted to the cycling network using the *Spatial Adjustment Toolbar*. These procedures ensure that the OSM network lays exactly on the Fietsersbond network, so that major differences in the snapping of locations to the network do not occur. Now the road infrastructure file is reduced to only the roads that are accessible for motorized travel by removing the road segments which are not accessible. Next, by *Planarizing* the remaining road segments, lines are split at intersections so that turns can be made correctly at each junction. The *Integrate* tool makes sure that line segments get connected at places where there are minor errors in the network. Finally, the network dataset is created with the *Create Network Dataset* tool in ArcCatalog.

This model also makes use of a *Closest Facility Analysis* to calculate the shortest routes. The centroid of the postal code in an iteration is added as a location for the origin. To add the locations of the destinations, the shapefile of all random points generated in the previous model is used as input. This file contains all destination points for all postal code areas. By using the *Select Layer by Attribute*, all points with the postal code number of the current iteration is selected, which is displayed in the *Value* box. All locations added are again snapped to the *Network Dataset* of the road infrastructure. Since some roads are one-way accessible, the *Closest Facility Analysis* includes an important restriction. The OSM file contains a variable indicating whether a road segment is meant for one-way or two-way traffic. The *Closest Facility Analysis* settings therefore include the restriction of one-way traffic. The output is again set to be the lines of routes with measures (distance). The *Solve Network Analysis* tool solves the analysis, after which the *Copy Features* tool writes the results to a folder where all motorized travel routes for all postal code areas will be saved.

Appendix C: SPSS syntaxes

2.1 Drawing research sample

```
FILTER OFF.
USE ALL.
SELECT IF (PERSID >= 0 AND HVM >= 0 AND POP_DENS >= 0).
EXECUTE.
RECODE HVM (14=1) (15=1) (11=0) (13=0) (1 thru 6=0) (ELSE=SYSMIS) INTO
BicycleUse.
EXECUTE.
FILTER OFF.
USE ALL.
SELECT IF (BicycleUse = 0 OR BicycleUse = 1).
EXECUTE.
SORT CASES  BY PERSID.
SPLIT FILE LAYERED BY PERSID.
COMPUTE ran1 = uniform(1).
RANK VARIABLES = ran1 by PERSID / rank into ran2.
SELECT IF (ran2 le 1).
SPLIT FILE OFF
EXCUTE.
```

2.2 Bivariate analysis: Spearman's Rho correlation

```
NONPAR CORR
  /VARIABLES=Bicycle_Motorized SEX AGE HHINCOME TRIPDIST POP_DENS OAD
DIST_SUPER DIST_DAILY DIST_CAFE CROSS_DENS PERC_C_ROAD SR1KM SR2KM SR3KM
SR4KM SR5KM SR7_5KM SR10KM AVG_RATIO
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

2.3 Multivariate analysis: regression models 1, 2 and 3

Model 1:

```
LOGISTIC REGRESSION VARIABLES Bicycle_Motorized
  /METHOD=ENTER SEX AGE HHINCOME TRIPDIST
  /CONTRAST (SEX)=Indicator
  /PRINT=GOODFIT
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Model 2:

```
LOGISTIC REGRESSION VARIABLES Bicycle_Motorized
  /METHOD=FSTEP(COND) SEX AGE HH_INCOME TRIPDIST POP_DENS OAD DIST_SUPER
DIST_DAILY DIST_CAFE CROSS_DENS PERC_C_ROAD SR1KM SR2KM SR3KM SR4KM SR5KM
SR7_5KM SR10KM AVG_RATIO
  /PRINT=GOODFIT
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Model 3:

```
LOGISTIC REGRESSION VARIABLES Bicycle_Motorized
  /METHOD=ENTER AGE TRIPDIST OAD DIST_DAILY PERC_C_ROAD AVG_RATIO
  /PRINT=GOODFIT
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Appendix D: Extended methodology for multivariate analysis

Regression analysis

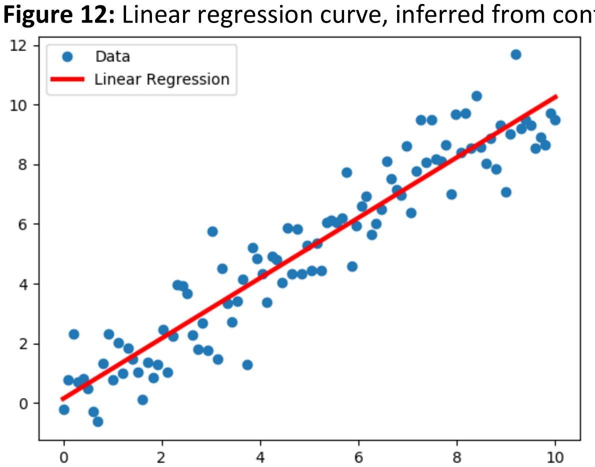
Regression models are used to predict a variable from another variable, or from multiple variables. By fitting a model to data, it is possible to predict an outcome (dependent) variable from one or multiple predicting (independent) variables (Field, 2013).

Different kinds of regression models can be carried out. When having one predicting variable, a simple linear regression model is performed. When taking several predictor variables, a multiple linear regression model is performed. The simple and multiple regression models assume that the dependent variable has either an interval or ratio measure, thus is continuous. This makes it possible to derive a linear relationship between the dependent and independent variable(s). The general linear regression function is:

$$Y = (b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_n X_{ni}) + \varepsilon$$

Where Y is the outcome variable, b_1 the coefficient of the first predictor (X_1), b_2 the coefficient of the first predictor (X_2), etcetera. Each predictor coefficient is estimated using the method of least squares, of which the actual curve can be drawn. The ε in the model represents the standard error of estimate of the regression curve. The error indicates the difference between what the model predicts and the value that is actually observed. An example of a linear regression curve deduced from all data points is presented in Figure 12.

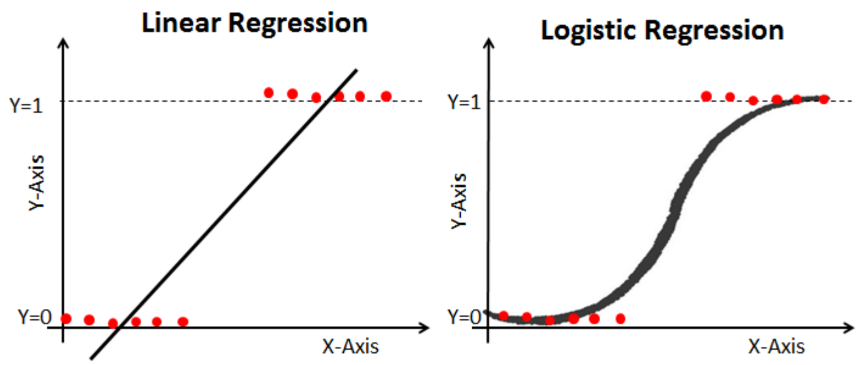
Figure 12: Linear regression curve, inferred from continuous data. (Source: John, 2018)



Logistic regression analysis

One of the main assumptions to use a linear regression model that the dependent variable is continuous (i.e. age or income). In some cases however the dependent variable is not continuous, but is dichotomous. Sieben (n.d.) gives the example of predicting whether a person is going to vote or not based on a person's characteristics (e.g. age, income, number of children). In this case the dependent variable is dichotomous, because this variable's values are either voting or not voting. In the case of a dichotomous relationship, a linear regression cannot be applied but a logistic regression should be applied (Sieben, n.d.). This is clearly shown in Figure 13, where both a linear regression curve and a logistic regression curve are drawn for a dichotomous dependent variable.

Figure 13: Difference in regression line between linear and logistic regression in case of a dichotomous dependent variable. (Navlani, 2018)



In contrast to linear regression, logistic regression does not predict the exact outcome value of the dependent variable based on the values of the independent variables. It predicts the possibility of an outcome value, since the outcome value can only be either or 1. A logistic regression therefore predicts the probability of an outcome value. Basically, the multiple linear regression function is transformed in logarithmic terms for the logistic regression:

$$(P)Y = \frac{1}{1+e^{-(b_0+\,b_1X_{1i}+b_2X_{2i}+\cdots+b_nX_{ni})}}$$

P(Y) here is the probability of Y occurring, given in the range of 0 and 1. This means that the closer the probability to 0, the lower the chance of Y being occurred. A probability close to 1, means that it is very likely that Y has occurred. The variable's coefficients (b_0, $b_{1,...,}$ b_n) in the logistic regression model have been estimated using the maximum likelihood method. Since the regression coefficients are presented in logit transformed context which are difficult to interpret quickly, logistic regression also provides odd ratios (OR) which are deduced from the logit regression coefficient. To calculate the odds ratio, the next formula is used (Field, 2013):

$$Odds\ ratio\ (OR) = e^{(b)}$$

$$Odds\ ratio\ (OR) = \frac{Odds\ after\ a\ unit\ change\ in\ the\ predictor}{Original\ odds}$$

Therefore in short, the OR indicates the ratio between the odds of Y occurring and Y not occurring (Y=1 or Y=0). The ratio has a value between 0 and infinity, where an OR between 0 and 1 indicates a lower odds on Y occurring when a predictor variable's value increases. A value >1 indicates a bigger odds of Y occurring when a predictor variable's value increases.

Regression methods

Depending on the variables that are included in a regression model, the model shows differences in performance. Performance of a model can be expressed in terms of predictive power and measures of fit (Field, 2013). More on model performance and measures of fit later in this section. Model performance can be improved by either adding or deleting certain variables. In other words, a regression estimates the best model to fit to the data that is available. There are multiple methods for estimating which model fits best, and which variables therefore should be included or excluded:

- Hierarchical: starting with only the known predictors based on past work. Subsequently the other predictors are added by the researcher in order of importance of predicting the outcome.
- Forward stepwise: starting with only the constant B (dependent variable) and adding variables one by one based on the significantly improvement of the model.
- Backward stepwise: starting with a complete model, where all predicting variables are included. One by one, predicting variables are being removed if they do not meet certain criteria.
- Enter: all predictors are forced into the model at once. The researcher does not have to make decisions on the order of entering the predictors.

The forward and backward stepwise methods are generally not recommended, since it may undermine the theoretical importance of some predicting variables and the fit of a variable is being made based on other variables in the model (Field, 2013). The risk here is that some predicting variables are excluded from the regression model, while it could be theoretically important to have them included. The hierarchical method and enter method thus are best used to avoid exclusion of important variables. One thing to take in consideration when creating a model is possible multicollinearity. That is when two independent variables are highly correlated, which may influence the probability values of the independent variables in a regression model (Field, 2013).

To determine which predictors are best included when adding variables with the Enter method, several practices can be applied. One of them is assessing the Wald statistic of a variable. The Wald statistic (*z*-statistic) indicates whether the B coefficient of the variable is significantly different from zero. If this statistic is indeed significantly different from zero, you can assume that the predicting variable is making a significant contribution to the prediction of the outcome. The Wald statistic however can be unreliable in cases where the B coefficient is large and the standard error gets inflated. This may lead to a higher probability of rejecting a predicting variable as being significant, when the variable in fact does have a significant contribution to the model. Field (2013) therefore argues that it is more accurate to enter variables into the model based on the difference they make in the deviance of each model. By comparing the change in deviance after entering each predicting variable separately, you can interpret their significance on the ability to predict the outcome variable.

Model fit and predictive power

Measures of fit and predictive power are useful to indicate the performance of a model and thus also indicate if the inclusiveness of variables improves a model's performance. Two important pseudo R-squared measures of predictive power of a model are Cox & Snell R^2 and Nagelkerke R^2, which are similar to R^2 in linear regression (coefficient of determination) (Sieben, n.d.). The Cox & Snell R^2 value can never reach the value of 1, which would mean the predictors fully determine the outcome variable. Therefore it is better to use the Nagelkerke R^2 measure, which can reach the value of 1. The higher the Nagelkerke R^2 value, the higher the proportion of the variance in the dependent variable is explained through the predicting variables. Thus when fitting a model to the data, the highest Nagelkerke R^2 value should be aimed for.

The most important measure of fit is the Chi-squared test, which compares the deviance (-2 log likelihood) of a predicting model with deviance of a model with only the constant B (Field, 2013). A higher Chi-squared value indicates a bigger difference between the model with only the constant B and the model including the predictors. The Chi-squared value also shows the significance, determining whether adding the predictor variables has significantly improved the ability to predict the outcome variable. Another measure of fit is the Hosmer and Lemeshow-test, which tests if predictions made by the model fit perfectly with the observed values based on group memberships. Here, the differences between predicted and observed values should be insignificant to conclude that the model fits to the data (Field, 2013). For larger sample sizes however, this test tends to indicate significance while in fact it is insignificant.

Appendix E: Bivariate analysis

Correlation matrix of all variables (1/2)

	BICYCLE_USE	GENDER	AGE	HH_INCOME	TRIP_DIST	POP_DENS	OAD	DIST_SUPER	DIST_DAILY	DIST_CAFE	CROSS_DENS	PERC_C_ROAD
BICYCLE_USE	1.000											
GENDER	,066*	1.000										
AGE	-,221**	-,060*	1.000									
HH_INCOME	-,080**	-,117**	-,087**	1.000								
TRIP_DIST	-,477**	-,161**	0.012	,114**	1.000							
POP_DENS	,097**	0.004	0.025	-,079**	-,061*	1.000						
OAD	,125**	0.014	0.023	-,151**	-,098**	,821**	1.000					
DIST_SUPER	-0.022	-0.011	-0.006	,095**	0.022	-,582**	-,631**	1.000				
DIST_DAILY	-,085**	-0.019	0.016	,128**	,081**	-,558**	-,667**	,787**	1.000			
DIST_CAFE	-,075**	-0.022	0.000	,148**	,103**	-,328**	-,503**	,512**	,672**	1.000		
CROSS_DENS	,078**	-0.005	0.028	-,065*	-0.051	,893**	,742**	-,498**	-,451**	-,177**	1.000	
PERC_C_ROAD	-0.019	-0.025	0.022	0.008	0.011	,146**	,162**	-,059*	0.024	,093**	,121**	1.000
SR1KM	-0.012	0.006	0.039	-0.009	0.037	-,193**	-,227**	,104**	,102**	0.039	-,207**	-,187**
SR2KM	-0.014	0.029	0.024	0.009	0.010	-,149**	-,116**	,066*	0.051	0.031	-,151**	-,203**
SR3KM	-0.044	0.002	-0.015	0.051	0.019	-,100**	-,093**	0.040	-,068*	-,082**	-,106**	-,122**
SR4KM	-0.036	-0.029	0.033	-0.044	-0.029	-,103**	-,084**	,082**	-0.003	-,090**	-,115**	-,108**
SR5KM	-,079**	-0.020	0.005	0.027	-0.011	-,114**	-,092**	-0.008	-0.041	-0.031	-,097**	-,114**
SR7_5KM	,075**	0.034	-,055*	-0.018	-0.018	-0.036	0.011	,062*	0.005	-0.048	-,059*	-,104**
SR10KM	0.006	0.035	-0.011	-0.021	-0.040	-0.044	-0.025	0.002	-0.001	-0.010	-0.018	-,077**
AVG_RATIO	-0.012	0.020	0.023	-0.035	-0.010	-,173**	-,128**	,098**	0.029	-0.041	-,167**	-,261**

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation matrix of all variables (2/2)

	SR1KM	SR2KM	SR3KM	SR4KM	SR5KM	SR7_5KM	SR10KM	AVG_RATIO
SR1KM	1.000							
SR2KM	**,292****	1.000						
SR3KM	**,100****	**,182****	1.000					
SR4KM	**,090****	**,128****	**,239****	1.000				
SR5KM	**,058***	**,062***	**,164****	**,260****	1.000			
SR7_5KM	**,133****	**,127****	**,109****	**,103****	**,084****	1.000		
SR10KM	0.053	0.037	**,233****	**,098****	**,118****	**,151****	1.000	
AVG_RATIO	**,490****	**,508****	**,481****	**,523****	**,396****	**,400****	**,382****	1.000

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).