Final thesis report

The impact of safety on cycling route choice

Author
Maaike Kuiper
5752507

Supervisor
Dr. C. Maat

Professor
Prof.dr.ir. P.J.M. van Oosterom

June 2, 2021

The impact of safety on cycling route choice

An analysis of the impact of objective and subjective safety aspects on cycling route choice behaviour in the Netherlands

Maaike Kuiper

Supervisor:
Dr. Kees Maat
Faculty of Civil Engineering and Geosciences
TU Delft

Responsible professor:
Prof. dr. ir. Peter van Oosterom
Faculty of Architecture and Built Environment
TU Delft

In partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Geographical Information Management and Applications

June 2, 2021

Preface

In this document I present my master thesis *'The impact of safety on cycling route choice'*. This thesis has been written as part of the master's programme Geographical Information Management and Applications (GIMA), which is a co-operation between Utrecht University, Delft University of Technology, University of Twente and Wageningen University.

I look back on my thesis as a period in which I learned lots of new things, both in theoretical terms as in terms of the research process. Completing this thesis would not have been possible without the help of many people whom I would like to thank.

First of all, I would like to thank my supervisor Kees Maat for his extensive feedback and valuable insights during this research process. Secondly, Danique Ton was of great help in explaining the specifics of cycling route choice modelling by providing new insights into choice set generation and the inclusion of the Path Size in a route choice model. Thirdly, I would like to thank Peter van Oosterom for his feedback.

Furthermore, I am very thankful to Jimme Smit and Harmke Vliek for performing the data collection and data preparation together, and discussing solutions for problems we encountered. Both were essential during this period. Our regular morning meetings helped to keep me going during this worldwide pandemic and were a great start of the day.

In line with this, I would like to thank my other fellow students, friends and family for their endless support and comforting coffee breaks. A special thanks to Johan Kuiper for his detailed proofreading and feedback.

I wish you a pleasant reading experience.

Maaike Kuiper
Utrecht, 2 June 2021

Abstract

This thesis studies the effect of safety on route choice behaviour of cyclists. In order to measure bicycle behaviour, data is collected by tracking respondents with GPS-devices. This resulted in a set of observed routes that all exist of an origin-destination pair. For each observed route a set of alternatives was generated between the origin and destination. For the choice set generation the approach was based on a set of labelled alternatives. This means that each alternative has its own goal on which the trajectory is based. The spatial attributes of the observed route are analysed relative to the spatial factors of the alternative routes with the use of a MultiNomial Logit (MNL) choice model. To account for overlap between the routes a path size factor was used. With the estimated choice model in Stata the impact of objective safety, traffic safety and social safety were studied, based on the following factors: accidents, separated bicycle paths, bicycle lanes, speed limits, street lighting, crime rates and an urban or natural environment. Together with these safety factors, the interaction with personal characteristics such as gender, age and cycling experience was analysed. Results from the choice model indicated that respondents preferred bicycle lanes, safe speed limits, street lighting and nature in their cycling route choice behaviour. In the context of safety, the effects of bicycle lanes, speed limits and street lighting were as expected. With regard to personal characteristics, age was the most influential factor with a significant impact on various safety factors. To conclude, the results of this study show that traffic safety had the largest impact on route choice behaviour of cyclists.

Table of contents

List of tables

List of figures

1 Introduction

1.1 Context and problem statement

Cycling is an important mode of transport in the Netherlands, which is considered to have many benefits: it is a healthy and low cost manner for people to transport themselves and it is associated with low levels of pollution compared to other types of transport (Handy, van Wee, & Kroesen, 2014). On the downside of the health benefits due to its active character, is the health burden resulting from bicycle crashes. The Netherlands is known for its relatively safe cycling environment, partly due to the fact that almost a quarter of all trips and about 10% of the total travel distances of Dutch people is covered by the use of bicycles. This has created traffic circumstances in which people are used to cyclists as road users (Ministerie van Infrastructuur en Waterstaat, 2019). The wide use of bicycles for a longer period of time has created a leading international position in bicycle use and safety for the Netherlands. However, cyclists still seem to make up the majority of severe injuries and fatalities in traffic in the Netherlands (Schepers, Twisk, Fishman, Fyhri, & Jensen, 2017). The share of cyclist fatalities amounts up to almost 25% of all traffic fatalities from 1996 to 2019 (CBS, 2020a). With respect to serious road injuries, cyclists accounted for 64% of all severe road injuries in 2018 (SWOV, 2019). Last year, during lockdown as a consequence of Covid-19, a peak has been reached, with the highest number of cyclist fatalities in 25 years (CBS, 2021a). Cyclists are relatively vulnerable road users and as the numbers show there remains room for progression in cyclist' road safety in the Netherlands (Schepers et al., 2017). Therefore, the Dutch government focuses on tackling the safety problems that exist for cyclists by means of various initiatives from multiple institutions (SWOV, 2018). Some of these initiatives are targeting specific groups, such as the program 'Doortrappen' that is aimed at improving safety for elderly people who tend to be extra vulnerable road users (Ministerie van Infrastructuur en Waterstaat, 2018). Another initiative was recently launched: 'De Meefietslijn' aimed at a different kind of safety, and offered people (mainly women) the possibility to call with someone to enhance their feeling of social security. This idea was created after a man assaulted multiple women in a neighbourhood in Utrecht, and many women were in fear to cycle on their own (Fietsersbond, 2020). Insight into the importance of the variety of safety factors for cyclists is lacking in many cases. Unsafe circumstances can affect the route choices that are made and therewith possibly create more unsafe cycling circumstances (Weijermars & Dijkstra, 2008), which can result in an increase in traffic accidents involving cyclists.

1.2 Research gap

Several studies found that safety aspects seem to affect the route choice cyclists make (Harvey, Krizek, & Collins, 2008; Menghini, Carrasco, Schüssler, & Axhausen, 2010; Segadilha & Sanches, 2014). The other way around, route choice influences the traffic flows as well as traffic safety as Weijermars and Dijkstra (2008) pointed out, therewith showing a mutual interplay between safety and route choice. Many research efforts have revealed that stress, safety and comfort for cyclists can substantially affect route choice behaviour, however the specific factors that cause this vary greatly over space (Caviedes & Figliozzi, 2018). Lawrence and Oxley (2019) also state that the route choices cyclists make can have a significant impact on their safety, because risk factors such as traffic volumes and speed are unevenly distributed over space. For example, the type of bicycle facilities can affect route choice behaviour: in several studies cyclists would choose routes with facilities where they are separated from motorists (Broach, Dill, & Gliebe, 2012; Kaplan & Prato, 2015). Safety and cyclists' perception of safety are thus, vice versa, considered to be a rather important determinant in route choice behaviour (Lawrence & Oxley, 2019). The distinction can be made between two types of safety: objective and subjective safety. Objective safety entails the occurrence of accidents. In subjective (perceived) safety, factors such as traffic volume, speed and facilities come into play (Heinen, van Wee, & Maat, 2010). In cycling, social safety can also be a factor that might change ones perception of safety and cycling behaviour. The danger of crimes in certain areas can cause a cyclist to take a detour. However this effect has not been addressed in many studies (Appleyard & Ferrell, 2017). Furthermore, the perception of safety can differ between gender and age. A research into route choices in London in 2012 shows that cyclists make their route choices primarily based on the degree of safety at certain points as well as the avoidance of traffic. Especially among women safety concerns were an important factor in their route choices (Steer Davies Gleave, 2012).

However, there are not many studies that specifically examine the different aspects of objective and subjective safety in route choices (subjective safety consists of the perceived social safety and traffic safety). Besides, many studies focusing on the relation between safety and route choices are based on people's stated preferences (Song, Ni, & Li, 2017; Steer Davies Gleave, 2012; Winters & Teschke, 2010). This provides a good image of what people prefer in terms of bicycle facilities or traffic volumes for example. There is, however, a possibility that people show different behaviour than they state to prefer. Therefore, an analysis of observed behaviour of route choices and the relation to various safety indicators over space can provide additional insights in route choice behaviour.

1.3 Societal relevance

With a rising number of cyclist accidents and fatalities in 2020, the director of the Dutch Cycling Union (Fietsersbond) calls for action: "it is important to take action, by adopting tangible measures as soon as possible" (Fietsersbond, 2021). Social safety becomes a larger concern for cyclists as well, that can be diminished by creating an environment that enhances the feeling of safety (Fietsersbond, 2020). This study will therefore focus on the influence of safety on cyclists' actual route choice behaviour. In this way, the research can help identifying risk factors where safety for cyclists can be improved, which can be of help for municipalities and provinces to take concrete measures. Additionally, the influence of safety on route choices will be compared between groups with different personal characteristics, such as gender and age.

This thesis therewith contributes to the planning, transportation and public health literature by performing a geographic analysis on cycling patterns and differentiating between personal characteristics. This can demonstrate the difference in patterns of cycling and safety considerations among cyclists. These findings can help identifying specific target groups for policies and campaigns aimed at improving safety in certain areas. Besides that, it can be of help in policy discussions to prepare decisions in infrastructural design, which will have varying impacts on different road users in terms of making the cycling environment safer. Having the knowledge of safety preferences among cyclists can thus help identify goals and develop targeted policies.

1.4 Research objective and research questions

In this thesis it is hypothesized that safety considerations play a role in cyclist' route choice behaviour. The objective is to determine which safety factors- and to what extent these factors influence the route choices that cyclists in the Netherlands make, by analysing GPS data of cyclists. In order to reach this research objective, the following research question will be answered:

How and to what extent does safety affect the route choice behaviour of cyclists?

And the following sub-questions:

- What is the influence of objective safety on route choice behaviour?
- How does subjective safety – ranging from traffic safety to social safety – affect route choice behaviour?
- Which personal characteristics affect the extent to which cyclists find safety factors important in their route choices?

1.5 Research scope

This study focuses on cycling route choices spread out over several regions in the mid-area of the Netherlands. This can be seen in Figure 1.1 where all routes that are used for analysis in this study are visualised. The study specifically examines the route choice behaviour of cyclists and the safety aspects that might explain route choices. This research thus does not take into account all other environmental, social or economic factors that might influence or explain cycling behaviour and route choices. It is important to take in mind that the data collection in this research took place during a lockdown as a consequence of Covid-19, which might have affected respondents' cycling behaviour.

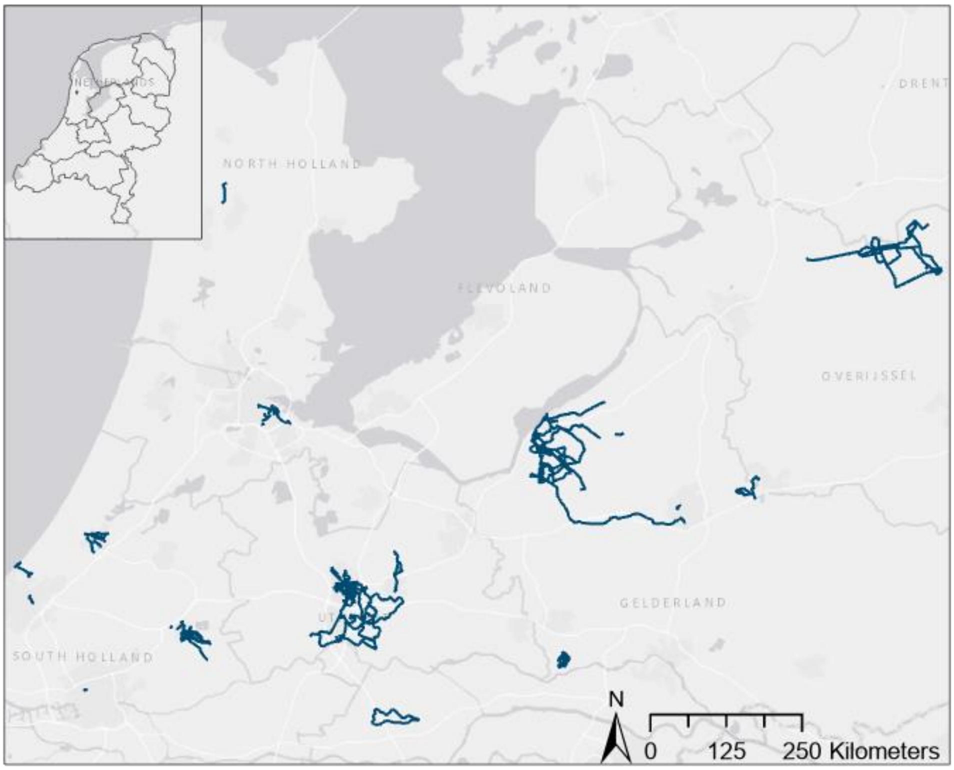


Figure 1.1. Study's cycling routes

1.6 Research approach

In order to answer the main research question, the first step is to identify safety factors that might affect cycling behaviour and perform an open data search to find corresponding datasets that identify the safety of a route. The cycling data is collected by tracking respondents with a GPS-tracker for a week. Subsequently, a survey is filled in by respondents to gather data on their personal characteristics and stated preferences. The central dataset, from which safety data is extracted and which is used for the generation of routes, is the Fietsersbond dataset (Fietsersbond, n.d.). The collected GPS and safety data are used as input for a set of route alternatives for each trip, these are generated with network analyst using ArcMap and ArcGIS Pro. These are the result of a joint data collection and data preparation in this study. Subsequently, the alternative routes are used in the route choice model. Additionally, the varying influence of personal characteristics is analysed. In this way the effect of safety aspects and personal characteristics on route choices are estimated, using a discrete choice model.

1.7 Research outline

In the next Chapter the theoretical framework on the relation between safety and cycling behaviour will be given. Chapter 3 will provide a methodological background on measuring bicycle behaviour and route choice modelling. Subsequently, the research methods of this study will be elaborated in Chapter 4. Chapter 5 will present the results of the analysis. Then, in the discussion in Chapter 6 the results will be related to the outcomes of previous studies. Finally, in Chapter 7 the conclusion of this research will be given.

2 Literature review: safety and cycling behaviour

This Chapter serves as the theoretical framework to identify the factors that affect road safety and are considered to affect route choice behaviour of cyclists.

According to Heinen et al. (2010) two types of safety can be identified: objective and subjective safety. The first refers to 'real' safety measured by the number of cycling accidents that occur. Subjective safety entails how cyclists perceive and experience safety (Heinen et al., 2010). Both types of safety for cyclists can be determined by several factors; variables that come up in existing literature will be discussed in the next Sections.

2.1 Traffic accidents

There are various factors indicating the degree of safety for cyclists. The occurrence of traffic accidents is the most prominent indicator of the safety of a certain road. Sener, Eluru and Bhat (2009) conducted surveys to research bicycle route choice preferences and their results show that 70% of the respondents considered cycling somewhat or very dangerous in light of traffic accidents.

However, other studies found no significant relation between traffic crashes and cycling usage or behaviour. Sun, Mobasheri, Hu and Wang (2017) created a buffer of 300 meters around places where traffic accidents took place, but the researchers did not found a relation between the occurrence of traffic accidents and bicycle usage. Sun, Du, Wang and Zhuang (2017) investigated the influence of several environmental and socio-economic factors on recreational cycling behaviour. Their results are comparable and they find a negative but non-significant relation between traffic accident density and the recreational cycling rate. A possible explanation for the absence of a significant relation between traffic accidents and cycling can be the fact that traffic accidents are mostly a consequence of unsafe situations. Bicycle crashes are more likely to occur at intersections, and the occurrence is related to factors such as speed limits for motor vehicles, car- and bicycle traffic volume, bicycle facilities, weather circumstances and cyclist's personal characteristics (Kondo, Morrison, Guerra, Kaufman, & Wiebe, 2018). Therefore, as cited in Segadilha and Sanches (2014), El-Geneigy (2010) claims that the perception of road safety is important and not the actual number of accidents that occur on the road. Therefore, it is not sufficient to focus solely on traffic accidents in determining the effect of safety on route choice behaviour. In the next Sections the factors existing in the built environment that can affect people's perception of safety and therewith affect their cycling behaviour will be discussed.

2.2 Bicycle facilities

Bicycle facilities entail facilities specifically designed for cyclists such as bicycle lanes and paths. Kaplan and Prato (2015) claim that bicycle paths where cyclists are separated from motorists are considered to create safer situations as well as greater safety perceptions among cyclists in Copenhagen. The number of conflicts and stress due to road-sharing are clearly diminished when bicycle paths are used (Kaplan & Prato, 2015). The route choice model in a study by Broach et al. (2012) also showed that there was a tendency to choose separated bicycle paths, followed by bicycle boulevards. In many other studies it became apparent that cyclists generally prefer separate bike lanes or paths as it creates a more segregated space where they can cycle without interference of other road users. The separation from motorists also creates a larger feeling of safety among cyclists (Casello & Usyukov, 2014; Menghini et al., 2010).

However, a study in Amsterdam found that when cycling is a widely used mode of transport, separate bicycle paths do not necessarily attract cyclists. They therefore suggest that more research is required into the importance of separate bike lanes in a dense cycling network (Ton, Cats, Duives, & Hoogendoorn, 2017). Similarly, another study found that cyclists do not always choose specific bicycle facilities such as lanes and boulevards in areas where cycling is dominantly present. This study was performed in the Netherlands and shows an avoidance of bike lanes and boulevards compared to the shortest path (Bernardi, Geurs, & Puello, 2018). This shows, that even though the majority of researches find a positive relation between route choice and bicycle facilities, it seems to be dependent on the network in the area.

Heinen et al. (2010, p. 63) also state that there is a distinction between gender, age and cycling experience in the preference of bicycle facilities: ”Inexperienced cyclists, women and younger cyclists tend to consider bicycle facilities to be more important”. A study in Melbourne, Australia also found a preference of female cyclists to use off-road paths over roads with no bicycle facilities (Garrard, Rose, & Lo, 2008). This indicates that route choices might differ among groups as they have different preferences in bicycle facilities.

2.3 Traffic volumes

Many roads are shared between cyclists and motorists, therefore roads with high volumes are perceived to create safety problems for cyclists. A stated preference research into cycling behaviour by Parkin, Wardman and Page (2007) concluded that busy roads increase the perceived risk. Bicycle lanes did seem to have a positive effect on this perception, but not enough to compensate the considered risk of large volumes of motorists (Parkin et al., 2007). Misra and Watkins (2018) using gender and age segmented route choice models, found that Annual Average Daily Traffic (AADT) had a significant impact on route choice for various groups. An increase in the AADT with 1,000 vehicles diminished the chance of choosing a route by 55% for females, for older cyclists it meant a reduction of 60% (Misra & Watkins, 2018).

However, Casello and Usyukov (2014) compared two different route choice models, where they found that in the first model the volume did not affect route choices significantly. The second model actually showed that higher car volumes would increase the attractiveness of a road for cyclists. Additionally, Zimmermann, Mai and Frejinger (2017) found that there is no significant difference between a medium and high traffic load in route choices. In addition, a study in San Francisco into route choices found that traffic volume had no effect on the route choices of cyclists (Hood, Sall, & Charlton, 2011). The influence of traffic volume on route choice thus seems to be related with other infrastructural characteristics such as separate bike lanes.

2.4 Traffic speed

A higher speed limit is in many cases considered to create unsafe circumstances for cyclists and can thus affect route choices. A study from Sener, Eluru and Bhat (2009a) in Texas showed a preference towards roads with lower speed limits. Even more experienced cyclists, who took routes with moderate speed limits more often, tended to avoid roads with high speed limits because of safety risks (Sener et al., 2009a). When analysing bicycle volumes along street segments, Jestico, Nelson and Winters (2016) found that volumes decreased when speed limits were above 30 km/h, indicating that cyclists generally avoid these roads. In another study the conclusion was that cycling along quiet, local roads was more important than the presence of any other physical environmental factor (excluding distance). Roads with a speed limit of 30 km/h with few commercial destinations and mixed land use were the most popular among cyclists, even when bicycle lanes were not present (Verhoeven et al., 2018). Bicycle facilities where people are separated from other road users are expected to decrease the impact of higher speed limits on people's perceived safety (Winters, Brauer, Setton, & Teschke, 2013). Comparable as to the influence of traffic volume on cycling route choice, Misra and Watkins (2018) note that an increase of speed with 10 m/h (+- 16 km/h) diminishes the chance of choosing a route with 40% for both female and older cyclists. Furthermore, a relation between bicycle accidents and increasing speed of motorists was found in a study of Stone and Broughton (2003). This shows that a high speed limit can create unsafe situations as well as a perceived insecurity for cyclists that causes them to take a detour.

2.5 Intersections and traffic lights

Almost every bicycle trip generally crosses one or more intersections. The amount of possible traffic crash locations increase as the number of links increases. An increase of links takes place at intersections, where the increase of turning movements affect safety. As a result the intersection characteristics and traffic control mechanisms such as traffic lights affect choices on whether, when and where people cycle (Buehler & Dill, 2016).

Cyclist injuries occur more often at intersections, however mostly the injuries were not severe (Zahabi, Strauss, Manaugh, & Miranda-Moreno, 2011). Besides, signalized intersections with lighting as well as traffic lights are considered to have a positive effect on cyclist safety (Chen & Shen, 2016; Han, Huang, Lee, & Wang, 2018).

Generally, intersections and traffic lights are hard to avoid when cycling and they also cause irritation because of delays. Therefore, it can be expected that cyclists might want to avoid intersections and their traffic lights or stop signs (Heinen et al., 2010). A preference towards routes with fewer intersections is found in many studies, for example, a research of Sarjala (2019) into both pedestrian and cyclist route choice found that roads with high intersection densities are generally avoided. This intersection density was even found to have the most significant correlation with route choice. A route choice study in Amsterdam showed that the amount of intersections per kilometre was relatively low on the observed routes (Ton, Duives, Cats, & Hoogendoorn, 2018).

However, sometimes people even prefer traffic lights or stop signs as it can create more safety (Heinen et al., 2010). Lu, Scott and Dalumpines (2018) concluded from their model outcomes that the majority of the routes their respondents cycled had more intersections, probably because people are willing to take a detour for proper bicycle facilities that avoid high traffic volumes. A study in Portland reports that cyclists generally avoid traffic signs, unsignalized intersections and stop signs. However, the negative effect of stop signs was decreased when the traffic volume increased substantially. Reasoning that increased perceptions of safety or time savings due to the traffic sign were considered, however not proven (Broach et al., 2012). Verhoeven et al. (2018) researched cycling behaviour among adolescents in Gent and found no significant relation between intersections and route choices. From the various studies it seems that the importance of intersections and their characteristics, i.e. the presence of traffic lights, in route choice differs among infrastructural and personal characteristics.

2.6 Street lighting

Chen and Shen (2016) found that there was a significant negative relation between the number of streetlights and the number of cyclist accidents. This shows that better street lighting is associated with more safety for cyclists. Furthermore, cyclist's perception of safety is higher in well-lighted areas. According to their conclusion, an improvement of street lighting decreases the probability of cyclist injuries as well as the severity (Chen & Shen, 2016). Kim, Kim, Ulfarsson and Porrello (2007) also showed the relation between the presence of street lighting and the severity of the injury: darkness with no streetlights was shown to increase incapacitating injuries and fatal injuries even with a change of over 100%. A stated preference research into the perception of safety of cyclists in Dublin found that cyclists prefer roads with street lights. Based on the collected data they compared three different models. From their cyclist-network interaction model it seemed that a share of cyclists would alter their route to use roads with street lights. Cyclists, who tend to alter their route based on this aspect, were also likely to alter their routes for a route that is perceived as safe, contains quiet roads and continuous cycle lanes. This indicates the relation between well-lit streets and perceived safety. This perceived safety with regard to street lighting could have a link with social safety as well, which will be further elaborated in the next Section. Provision of street lights can for example also attract non-cyclists to start cycling according to the authors (Lawson, Pakrashi, Ghosh, & Szeto, 2013).

However, in an exploration of Dessing et al. (2016) into Children's' route choices in Amsterdam, the observed routes contained less street lighting than the shortest routes. This can be explained by the fact that busy roads are avoided in this study, which generally have more street lighting. Osama and Sayed (2017) found a positive association between the density of light poles and bicycle accidents. They assigned this to the higher bicycle volume on streets with better street lighting, where the exposure to risk is also generally higher.

2.7 Social safety

Social safety is described by Zwerts, Allaert, Janssens, Wets and Witlox (2010, p. 702) as follows: "...the protection or feeling of being protected against the dangers caused by human actions in public spaces. Examples of these incidents are aggressive behaviour, public drunkenness, vandalism, drug trading and use, assaults and murder."

In cycling behaviour studies, social safety is not frequently taken into account as an explanatory variable for route choices or preferences. This can possibly be explained by the fact that it is associated with the feeling of insecurity and thus rather subjective and hard to measure. Related to this, Rietveld and Daniel (2004, p. 533) refer to it as personal security: "Personal security: relates to ease of going out at any time of the day and in any sector of a city without being anxious about one's individual safety".

An unsafe feeling in the context of social safety is also related to the previous factor: the amount of street lighting. A shortage of street lighting can create an unsafe feeling. Bohle and Verkehr (2000) found that routes through more remote green areas can be very attractive for cyclists, but at the same time the feeling of safety can decrease here. Women and children would choose other routes during darker daytime due to an unsafe feeling in these areas. This indicates that more abandoned or rural/nature areas can deteriorate the feeling of safety (Bohle & Verkehr, 2000).

Crime rates in neighbourhoods can also be an indicator of the feeling of safety, as this feeling can possibly diminish when one has to cycle through an area with high crime rates or a bad reputation. Many studies on the impact of crime on travel behaviour find a complex interaction between crime rates, the built environment, perceived safety and the mode of travel. However, few studies seem to focus on cycling route choice behaviour in the context of social safety (Appleyard & Ferrell, 2017). Sun, Mobasheri, et al. (2017) researched cycling behaviour of bicycle-sharing users in Chicago. One of the datasets used by the researchers to measure safety was a record of violent crimes that took place. The hourly number of on- and off-street violent crimes within a 300 meter buffer is used. From their model it appeared that these social safety factors have a significant negative association with the number of arrivals at the bike-sharing system. For the number of departures this appears to be non-significant. Violent crimes thus tend to decrease the usage of the bicycle-sharing system. Assuming that this is similar to normal bicycle usage, it is possible that people tend to avoid these unsafe areas in their bicycle trip (Sun, Mobasheri, et al., 2017). A small study in Enschede into social safety for cyclists quote several experiences of people who feel unsafe at certain locations as they are scared to be attacked. One woman claimed that in the dark, she would avoid taking her usual fastest route as she would feel unsafe at certain points in her usual route (Fietsersbond, 2019). Contradictory to these findings, Sener et al. (2009b) concluded from their survey results that only 20% deemed cycling dangerous in the context of crime. Similarly, Hood et al. (2011) found no relation between the number of violent crimes and the route choices made by cyclists.

2.8 Conceptual framework

The factors that, based on the literature review, are assumed to be associated with cyclist safety are presented in the conceptual framework in Figure 2.1. Various studies are performed outside of the Netherlands, therefore it is disputable how comparable these research outcomes are to cycling behaviour in the Netherlands. Nonetheless, the studies provide a proper framework to analyse the resulting safety factors in the Netherlands.

One of the determining factors in cyclist route choice is road safety. Road safety can be broadly divided into objective and subjective safety. Objective safety concerns the number of traffic accidents that took place. Subjective safety involves more factors and can be subdivided into two ’safety categories’: traffic safety and social safety.

One of the indicators belonging to traffic safety is bicycle facilities, where especially the segregation of a bicycle path from the road can be of importance in route choices cyclists make. Furthermore, high traffic volumes and traffic speeds are considered to have a negative effect on safety and busy roads and roads with high speed limits are therefore expected to be avoided. However, this is related with bicycle facilities as a segregated path can benefit the feeling of safety on these type of busy roads. Generally, there is a preference towards few intersections, however together with traffic lights and stop signs these are expected to advance safety. As a consequence, people might choose for roads with intersections and traffic lights on purpose. The last factor is street-lighting, there seems to be a preference towards well-lit streets in previous studies. More street-lighting advances the traffic safety, as well as the feeling of safety, which also makes it part of social safety. Social safety is about the fear of not being protected against criminal assaults. The type of environment one cycles through can have an impact on the feeling of safety, but also the presence of street lighting. Better street-lighting can decrease fear of criminal assaults. Furthermore, higher crime rates generally affect bicycle usage and route choice.

All these factors that are part of subjective safety differentiate among personal characteristics such as gender, age and cycling experience. Where women and older cyclists, for example, are generally more concerned with safety, and expected to adapt their route choices more because of safety concerns.

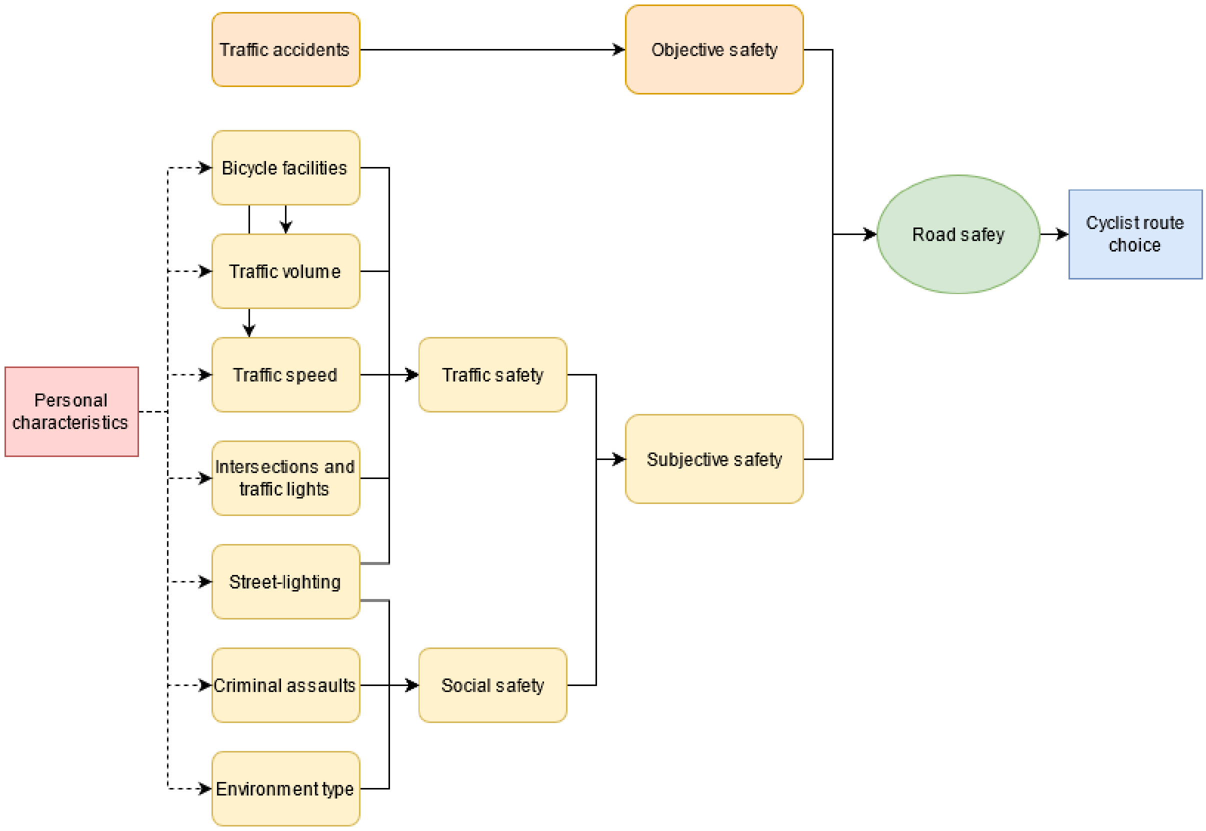


Figure 2.1. Conceptual framework

Based on this conceptual model, two safety approaches can be distinguished concerning route choice behaviour: objective safety and subjective safety. Here, subjective safety consists of traffic safety and social safety. From an open data search it appeared that there are some limitations as not all data are available in the Netherlands. Therefore, traffic volume and intersections and traffic lights are not taken into account in the analysis. This resulted in the following safety factors to be studied in this research:

- Traffic accidents
- Bicycle facilities
- Traffic speed
- Street lighting
- Crime rates
- Type of environment

These safety factors will be used in route choice modelling, specifically for the generation of a set of route alternatives and statistical analysis. A methodological background on route choice modelling will be given in the next Chapter. The exact research steps in route choice modelling will subsequently be elaborated in Chapter 4.

3 Methodological background: modelling route choices

In this Section a methodological background on measuring cycling behaviour and bicycle route choice models will be presented to substantiate the chosen methodology.

3.1 Measuring bicycle behaviour

Studies on route choice behaviour can generally be divided by the data collection method, with a focus on either stated preference (based on an experiment) or revealed preference (based on observed data) (Casello & Usyukov, 2014). Both approaches have their benefits and limitations.

In a stated preference model, data on route choice behaviour is extracted based on what respondents state they prefer in a specific route (Yang & Mesbah, 2013). Respondents have to choose from different route alternatives and make a trade-off between their characteristics (Broach et al., 2012). The stated preference approach is frequently used as it is a low-cost and relatively easy method to collect data, as it is not required to collect actual cycling data (Yang & Mesbah, 2013). Besides, a detailed network dataset and the generation of route alternatives for each origin-destination set are both not required in this case (Broach et al., 2012). A large disadvantage is that there is no proof that the preferences match with the actual choices cyclists make (Casello & Usyukov, 2014). Meaning that the hypothetical alternatives given in this method can give a lacking image of the actual preference of a respondent for certain facilities (Broach et al., 2012). Therefore, a revealed preference approach is deemed more suitable.

A revealed preference approach measures route choice behaviour of cyclists by using actual geographic route data. The use of this method became widespread with the rise of geographic information systems (GIS) (Zimmermann et al., 2017). Early on this was done by hand-drawn routes, which would entail people drawing their route on a map and then subsequently analysing these with the help of GIS. These types of analyses brought some useful insights, however these never lead to the estimation of a complete route choice model. The main reason for this was the absence of a proper set of path alternatives, as they primarily focus on the deviation of an observed path from the shortest path. Moreover, this method can create some accuracy problems, as people might not remember exactly which route they took (Casello & Usyukov, 2014; Menghini et al., 2010; Zimmermann et al., 2017).

In more recent studies that take on a revealed preference approach, the use of GPS technology became popular. People would be tracked with GPS devices to determine their route choice behaviour. This approach is more accurate as it uses recorded data and does not rely on peoples' memory for data (Casello & Usyukov, 2014; Menghini et al., 2010; Zimmermann et al., 2017). Besides, the GPS data contains detailed temporal and spatial information of people's cycling behaviour (Ton et al., 2018). The research of Menghini et al. (2010) was one of the first studies that performed a revealed preference study in cycling route choice on a large scale with the use of GPS trajectories. With the help of these trajectories they estimated a full cycling route choice model. Subsequently, more studies using this method followed. An important part of these types of research is the choice set generation as this has to be a realistic set of alternatives for each origin-destination pair. A frequently used criterion is therefore the coverage of the observed routes, this will be further elaborated in the next Section (Menghini et al., 2010; Zimmermann et al., 2017).

A downside is the complexity of the data analysis, because data will be full of noise, since it is not derived from a controlled experiment. The data collection is considerably more time consuming, which in many studies results in a smaller sample size. However, the benefits of this method exceed these computational problems (Yang & Mesbah, 2013). Therefore, this study will be based on the revealed preference approach.

3.2 Route choice modelling

The first step in route choice modelling is the generation of a set of alternative routes. Subsequently, these routes and their spatial characteristics are used as input in a route choice model.

3.2.1 Choice set generation

For each origin-destination (OD) pair, numerous alternative routes are generally available. As previously mentioned, the generation of a realistic set of alternatives is important. The size as well as the composition of choice sets is important, because when insufficient, it can lead to incorrect specifications of the eventual route choice model (Bovy, 2009). As Zimmermann et al. (2017, p. 185) state that: "...path generation algorithms should be able to reproduce the observed routes for a high proportion of origin-destination pairs". The efficacy of a generation method can thus be tested by the degree to which these cover the observed routes.

There are various methods that can be used to generate this set of alternative paths, these can be divided into deterministic and stochastic shortest path based methods (Bekhor, Ben-Akiva, & Ramming, 2006; Broach et al., 2012). Deterministic and stochastic indicate how the output of a model is determined. In a deterministic model, the output of the model is fully determined by its parameters and input. In stochastic models randomness and uncertainty are taken into account, meaning that the model input and parameters can give a varying output (Bovy, 2009).

The most common methods are the K-shortest paths, link-elimination and link-penalty generation approaches (Bekhor et al., 2006; Prato, 2009). All of these deterministic methods use different algorithms that search for the shortest paths between an OD-pair, based on varying behavioural traveller assumptions (Bekhor et al., 2006; Menghini et al., 2010; Prato, 2009).

Furthermore, path labelling and doubly stochastic generation are common approaches. The labelling approach is based on the assumption that travellers have different goals. Some wish to minimize travel distance or time, some might prefer driving through scenic landscapes and others might want to avoid busy roads (Prato, 2009). This method proposed by Ben-Akiva, Bergman, Daly and Ramaswamy (1984), generates multiple alternative routes with a specific criterion that is optimised for each OD pair. This means an optimal path for each specific label is created to be used for route choice analysis. Prato (2009) mentions various studies that used this labelling approach successfully, for example Ramming's research. He presented a simplified version of the approach where he searches for the shortest path using 16 different labels, varying from the shortest travel time to maximum travel time through safe neighbourhoods. The labelling approach however, does not always give completely adequate choice sets as the specific labels do not always have the ability to represent the range of spatial and functional variety among routes that is required. However, the labelling approach does offer the possibility to take different objectives in consideration and base the route choice analysis on this (Bovy, 2009; Prato, 2009). The doubly stochastic generation of choice sets adapts the formerly discussed method by combining the benefits of both stochastic path generation and labelling. This means the generated choice sets are based on labels and give a probabilistic outcome. This is a successful method as well, however computation times were found to be high (Bovy &

Fiorenzo-Catalano, 2007; Halldórsdóttir, Rieser-Schüssler, Axhausen, Nielsen, & Prato, 2014; Hood et al., 2011; Prato, 2012).

3.2.2 Route choice models

The number of alternatives resulting from the choice set generation influences the estimation of discrete choice models. An extensive range of models is used in existing route choice studies. The general theory for choice models is based on the random utility theory, which is based on the (economic) assumption in which travellers want to optimize their utility by finding the optimal combination of characteristics from a set of route alternatives (Dane, Feng, Luub, & Arentze, 2020). With the use of a route choice model, the chance that a path alternative is chosen can be estimated. Dane et al. (2020) subdivide the variety of models by the manner in which they handle the overlapping problem. This problem involves the sharing of one or more links by alternative paths.

The first group of models that is used in modelling travel behaviour is constituted by models that do not account for overlap. Examples are the MultiNomial Logit (MNL) and Nested Logit (NL), these do not seem highly appropriate for modelling route choice. A MNL model contains the simplest approach in modelling travel behaviour but does not consider similarities among alternatives because of overlapping routes. NL is based on the assumption that each alternative exclusively belongs to one option, while actually it is possible for routes to share many links (Dane et al., 2020; Dhakar & Srinivasan, 2014; Prato, 2009).

The second type of models uses a tree structure and do take overlap into consideration. This group is part of the Generalized Extreme Value (GEV) structures, these models account for similarities in the stochastic (chance variable) part of the utility function. However they do not account for variation or correlation over time or unobserved factors (Prato, 2009). Popular models in this group are the Cross Nested Logit (CNL) and Generalized Nested Logit (GNL). Both of these models overcome the correlation problem between nests, but the result is an extremely large and complicated model when it is applied to real-world networks (Dane et al., 2020; Prato, 2009).

The next group of models accounts for overlap by adding an additional attribute to the deterministic part of the utility function of a model. The Path-Size Logit (PSL) model is a popular practice in this group of models. The PSL model modifies the MNL model by using a path size term in the deterministic component (Dhakar & Srinivasan, 2014; Prato, 2009). The path size factor is thus introduced to diminish the alternative route's disutility when overlap takes place. The path size factor is an estimate of the overlap of an alternative with all other alternatives in the choice set (Ben-Akiva & Bierlaire, 1999). The PSL model has been successfully employed in several studies into GPS-based bicycle route choice models to analyse the effect of several variables on route choice behaviour (Dhakar & Srinivasan, 2014; Hood et al., 2011; Hoogendoorn-Lanser, van Nes, & Bovy, 2005; Menghini et al., 2010; Sobhani, Aliabadi, & Farooq, 2019). One of the reasons for its frequent use is the relative easiness and low computational effort of the model, but nonetheless accurate results (Dane et al., 2020; Prato, 2009). Various formulations for the path size are presented in literature. Ben-Akiva and Bierlaire (1999) present it as follows: the utility U_{in} of alternative i is given by:

$$U_{in} = V_{in} + \beta_{PS} lnPS_{in} + \epsilon_{in} \qquad (1)$$

where V_{in} stands for the explanatory variable(s), and the path size PS is defined as follows:

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \in Cn} \delta_{aj}} \tag{2}$$

where Γ_i is the set of links in path i; L_a is the length of link a and L_i is the length of path i. δ_{aj} is the link-path incidence variable that is 1 if link a is on path j and 0 otherwise (Ben-Akiva & Bierlaire, 1999; Dane et al., 2020, p. 115). The result is the degree of overlap. If a route is fully independent, with no overlap, PS_{in} would be 1. The path size logit is thus one extra variable that stands for the overlap of route alternatives, which is thus specific for each OD-set.

The last group takes overlap in consideration by allowing covariance between the error terms of the alternatives. An example is the Mixed Logit (ML) model, which considers heterogeneity among respondents. This model is, however, rather complex, which results in an extensive run time. Besides, the model has a specific calculation method that requires a precise and different method for specification, this can sometimes result in faulty outcomes, therefore this model is not used (Dane et al., 2020; Prato, 2009).

Furthermore, it is important to consider that choice models can be estimated based on their data structure: models with cross-sectional data or models with a nested data structure. Where the nested structure indicates a panel dataset, which means that decision makers make multiple choices as is the case in this research: cyclists cycle multiple routes (Stata, n.d.).

3.3 Chosen methodology

Measuring bicycle behaviour in this research will be done by collecting GPS data focusing on a revealed preference approach, where a few stated preferences of respondents collected by means of a survey, will be compared to actual route choice behaviour. Based on the existing literature on route choice modelling, the choice set generation will be performed using the labelling approach in order to be able to variate and compare between different types of safety. This means that different traveller goals will be used as a label for the generation of alternative routes. Based on the conceptual framework, a social safe and traffic safe route are included in these labelled alternatives. Furthermore, the following labels are used to create representative routes: a shortest, fastest, most continuous and greenest route. In order to model route choice behaviour, these path alternatives will be included in a MultiNomial Logit choice model with a path size attribute to account for overlap. Besides the nested structure of this study's data must be taken into account by using a model for panel datasets. With this model the effect of spatial variables on route choice probability will be estimated. In the next sections the exact research methods of this study will be elaborated.

4 Research methods

The data on cycling route choice behaviour has been analysed based on empirical data. In Figure 4.1 the complete research design of this thesis is visualised. The outcome of the literature study as described in Section 2.8 (the preparation phase) functions as the input for the next steps where the data of the safety factors will be operationalized and prepared for analysis. But first, an elaboration on the data collection of cycling data and its preparation will be given. Then, the steps in data preparation for the safety factors and the network dataset will be presented. The cycling network of the Fietsersbond (n.d.) functions as the base dataset for route choice analysis in this research as it is a detailed network dataset with many safety data attributed to it. Subsequently, the steps in the choice set generation are presented in Section 4.5. To end, the specifics of the statistical analysis will be given.

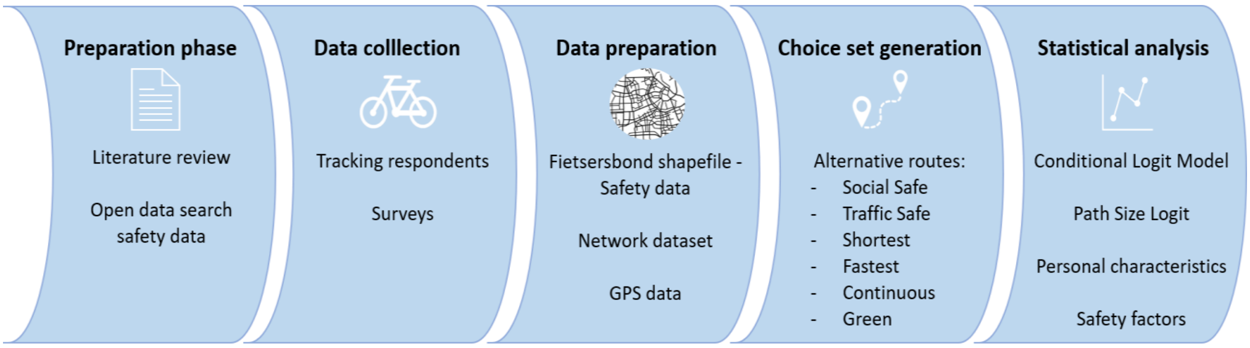


Figure 4.1. Research design

4.1 Data collection

Cycling behaviour data was collected by tracking people with a GPS-device and conducting surveys. The data collection (as well as the data preparation) is part of a joint data collection process with fellow GIMA students. The sample of respondents of the study was composed by the three researcher's own social networks. This resulted in a group of approximately 75 respondents. The research area was rather spread out over various municipalities in the mid-region of the Netherlands. In Figure 4.2 the research areas are visualised in green. The fragmented research area and the non-random research sample are the result of limitations due to Covid-19. The respondents were asked to take a GPS tracker with them for the short period of a week. GPS data is considered to give a significantly more accurate presentation of actual cycling behaviour than traditional measuring methods according to Bohte and Maat (2009). This means that GPS devices are deemed more accurate than smartphone applications. Besides, there is no negative consequence for the respondent's phone battery, making it user-friendly. Therefore the choice was made to use GPS-devices in this study. In order to deliver the GPS-devices to respondents, each student was responsible for bringing the devices to one third of the respondents associated with an explanation on how the process works. Accompanied with the tracker a letter with explanations about the functionalities and privacy measures was given (see Appendix A). At the end of the week the devices were collected again. A limited amount of GPS trackers was available, therefore the tracking took place over a period of approximately one month: from November 24th, 2020 till January 4th, 2021.

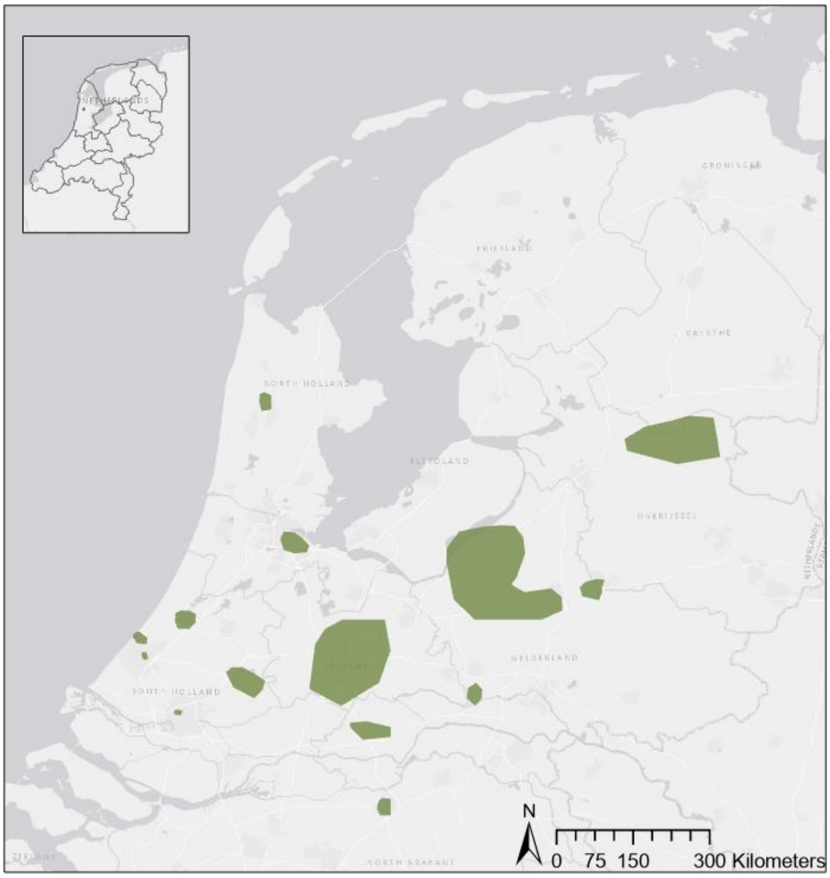


Figure 4.2. Research area

Additionally, the respondents were asked to fill in an online survey, with general questions on personal characteristics and more specific questions for each student's research. The surveys were conducted at the end of each respondent's week of tracking, in this way the questions in the survey cannot affect the cycling behaviour of respondents. The survey results are then used to link and compare personal characteristics and stated preferences of respondents to their cycling behaviour in the analysis. The survey questions can be read in Appendix B. In the survey respondents are asked for their personal preferences with regard to certain safety-related traffic circumstances, such as street lighting and speed limits for example. The responses to these statements give the indicated preference of respondents. By means of a pivot table in Excel their descriptive statistics are calculated, these will be presented in Section 5.1. Since answers to such questions can also incite respondents to give socially desirable answers, the stated preferences are compared to their actual cycling behaviour.

In the end, 73 respondents participated in the GPS tracking of which 70 participants filled in the survey. Solely the routes of respondents that filled in the survey are included in the research. During the GPS tracking it was necessary for the participants to charge the tracker at the end of each day, nevertheless, some respondents forgot this which led to some missing cycling data. Moreover, there were two malfunctioning trackers that lost signal during a certain tracking period, leading to more missed data. The initial GPS tracking resulted in 495 routes. However, in the data preparation phase, it turned out that certain routes had to be removed due to various reasons, such as errors in the map matching process or the failure to generate alternative routes. This resulted in a final research sample consisting of 61 participants and a total of 451 routes. Due to sets of completely overlapping routes (100% overlap for the observed route and all alternatives), 34 more routes are removed from the eventual choice model. This will be further elaborated in Section 4.6. Figure 4.3 gives an overview of the number of participants and routes.

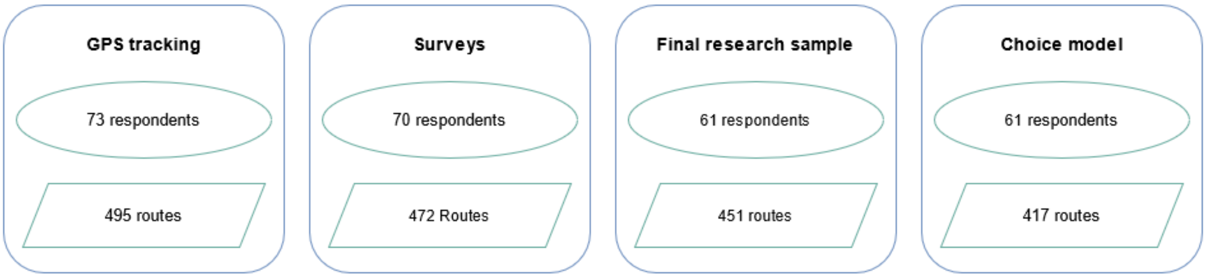


Figure 4.3. Research sample

4.2 Data preparation – GPS data

In Figure 4.4 the research flow of the GPS data preparation is shown, which contained the data collection and preparation of cycling behaviour data.

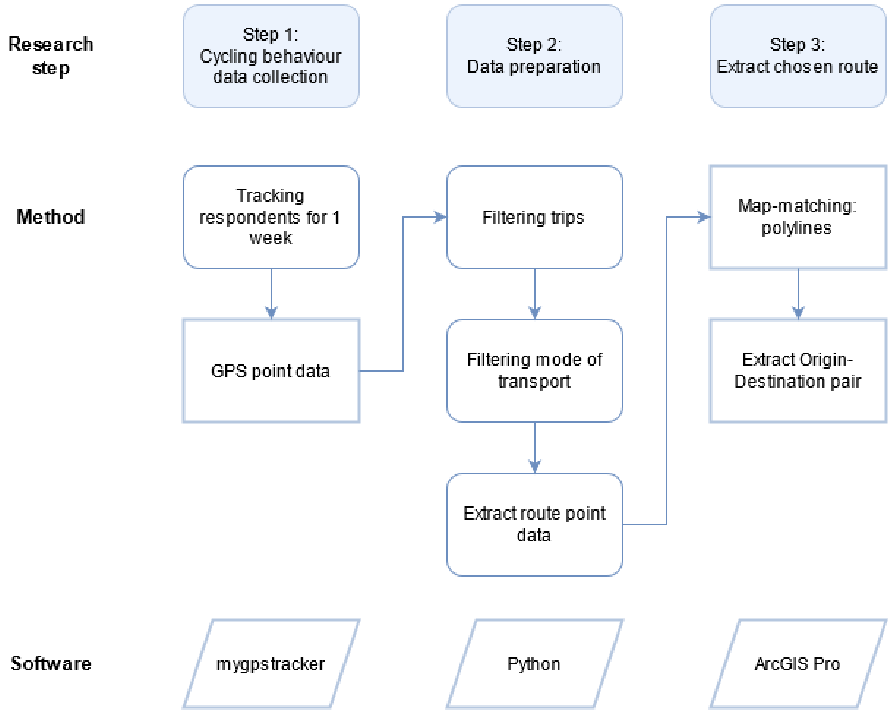


Figure 4.4. Research steps data preparation

The GPS devices store location and time data every few seconds, which resulted in a point dataset. For each point the X, Y, Z data was stored. It is important to validate this GPS data as it is probable to have incorrect data points due to bad satellite connectivity. This can result in either extremely long distances or high speeds (Rewa, 2012). The GPS data was stored continuously over the period of one week in which the respondent was tracked by the GPS device. This means that the cycling trips had to be filtered out. Besides, the data required to be filtered by the mode of transport: bicycles. Finally, map-matching was performed (Menghini et al., 2010). The data preparation process thus required a few steps that are lined out as follows in literature:

1. *Filtering of trips*
   Schuessler and Axhausen (2009) subdivide the GPS points into trips and activities. Trips and activities were detected in a similar manner as in their research. In their study activities are detected based on the following two criteria: selecting activities where the speed is low (0.01 m/s) for a short period, and activities where the point density is high for a certain sequence (of at least 10 points or 300 s). Activities where signal loss turned up were identified by using the time difference between two GPS points that follow each other up. With these criteria, the potential activity start and end points are determined. Subsequently, the ’trip objects’ were identified (Schuessler & Axhausen, 2009).
2. *Filtering of mode of transport: bicycle*
   The mode of transport was filtered out by making assumptions on acceleration and speed thresholds (Schuessler & Axhausen, 2009). This was done for each mode of transport: bicycle, foot, car, bus and train in this way the trips can be attributed with the right mode of transport. Bicycle trips can be identified by the average speed that must be below 25 km/h and a maximum speed below 45 km/h. Then, the cycling trips are selected for further use (Schuessler & Axhausen, 2009).
3. *Map-matching*
   After these two steps, the map-matching process was performed. This means that the GPS points were matched to the network dataset in order to determine the routes that respondents travelled (Schuessler & Axhausen, 2009). The aim was to minimise the discrepancies between the GPS trajectories and the network to eventually retrieve routes that properly match the GPS data (Ton et al., 2018).

Step 1 and 2 were combined in a Python script, where the order of these two steps was inverted and the selection process was somewhat adapted to the data in this study. For each tracker and period the respondent was tracked with a GPS device, the same process was executed. First, the bicycle points were filtered out by only selecting routes where the length was over 20 points, the average speed below 30 km/h and over 10 km/h, and the maximum speed below 40 km/h. Subsequently, these points were divided into the taken routes by splitting them based on a time difference of over 300 seconds where the speed is 0. Lastly, the remaining points were filtered, again, based on the speed requirements formerly mentioned. This can be seen in the Python script in Appendix C. The output of this Python script is a csv file with GPS point data for every route a respondent has taken.

Map-matching was an important step in the process that was required to create line trajectories from the point data. These line trajectories represent the routes that respondents cycled from A to B. From the begin and end points of these routes the origin and destination pair of were extracted. These trajectories were exactly matched to the network of the Fietsersbond dataset. In this way discrepancies between the GPS points and actual network were minimised. Furthermore, the spatial data of the Fietsersbond network could be joined to the trajectories more easily in this way. Map-matching was initially attempted to be done with the use of a Python script implemented in ArcPy developed by Scheider (2018). However, the map-matching program did not generate proper routes in many cases. Due to time constraints it was not possible to alter the map-match code, therefore, the choice was made to manually map-match de GPS data. The points from the csv files were visualised in ArcGIS Pro, by using the tool ‘XY Table To Point’. Subsequently, a line was drawn along the network dataset following the GPS points, starting at the begin point (where FID = 0) and ending at the last known GPS point. In this way, the discrepancies between the GPS points and network dataset are minimised, as shown in Figure 4.5. In some areas the network is dense, which can make it difficult to determine to which road section a string of GPS points belongs. In these cases the choice for a road section was made based on the assumption that a cyclist cycles on the right side of the road and that the outer road sections are the cycling paths. Performing manual map-matching was time-consuming, however, it does ensure the accurate results. And as mentioned, map-matching is an essential process to match GPS data to the network.

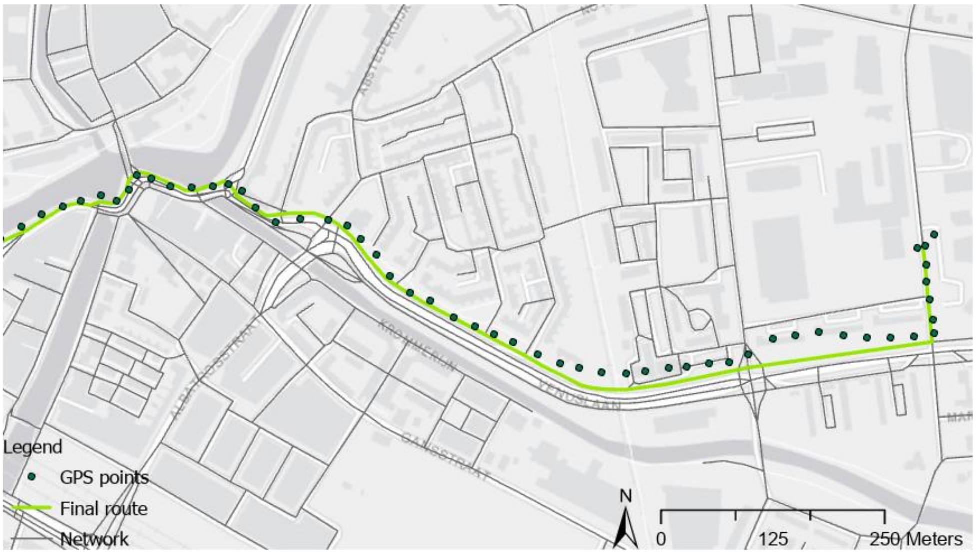


Figure 4.5. Map-matching visualisation

The Origin and Destination (OD) points were subtracted from the begin and endpoints of the newly created polyline. Lastly, the new route had to be joined with the Fietsersbond dataset, to assign all attributes of this dataset to the observed route. An overview of the tools used in the process is shown in Figure 4.6. From a csv file with GPS point data, the points are visualised with 'XY Table to Point', then the trajectories are created with a new feature class. Subsequently, the OD-pairs can be generated with the next tool in the scheme. Finally, a spatial join can be executed, which results in routes that have all the required spatial data for further analysis.

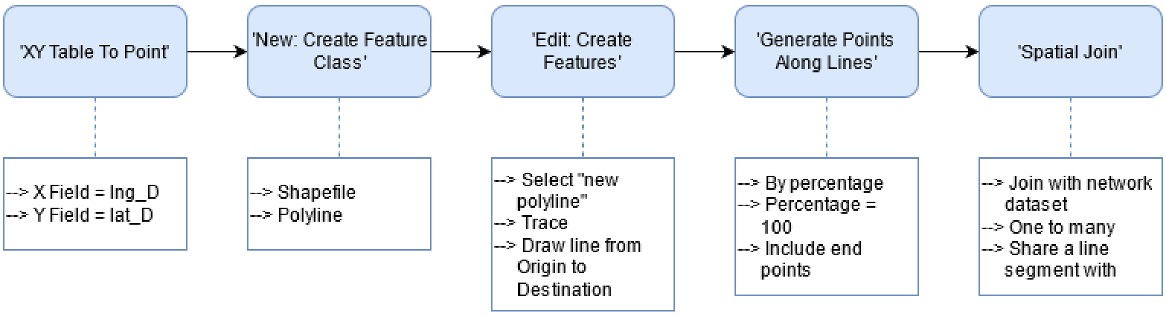


Figure 4.6. Overview map-matching process and tools

4.3 Data preparation – Spatial data

For the eventual spatial analysis all required safety data needed to be attributed to a road section. For each safety aspect different data types, reclassification and aggregation methods are used. First, these different safety aspects and their characteristics will be elaborated. Furthermore, additional datasets were required to generate the greenest, fastest and continuous routes for the choice set. In Section 4.3.2, all methods to enrich the Fietsersbond dataset with the safety data, as well as the other spatial data will be discussed.

4.3.1 Safety data

Accidents

Rijkswaterstaat (The National Department of Waterways and Public Works) keeps track of all traffic accident reports made by the police and links these to the National Road Database (Nationaal WegenBestand, NWB) in the so-called: Bestand geRegistreerde Ongevallen Nederland (Database Registered accidents Netherlands) (Rijkswaterstaat CIV, 2019). The result is a point dataset of the years 2008 till 2017 where each accident represents a point. For each accident many attributes are recorded, such as the involved mode of transport, the severity of the accident or the speed limit at that location. This means that the accidents can be selected based on the relevant mode of transport: bicycle. However, the dataset does not provide exact locations; an accident gets linked to either an intersection or road section. For each road section the midpoint then serves as the accident location in the dataset. Due to the fact that the accuracy of the accident location cannot be guaranteed there are no x/y-coordinates available (Rijkswaterstaat CIV, 2019).

The accident dataset is a point dataset consisting all kinds of accidents, first the accidents involving bicycles were selected. This was done by means of the 'select by attributes' function, where points that have a bicycle or E-bike attributed in the 'Partij object 1' or Partij object 2' field were used to create a new shapefile of bicycle accidents. In order to match these accident points to a road section, a buffer was used. A buffer of 25 meters was used to account for inaccuracy of the GPS data as well as the accident data. With a buffer of 25 meters the roads that are within 25 meters, at for example an intersection, are also considered dangerous. In this way the buffers were tied to specific road sections.

Bicycle facilities: road type

The road type is an indicator of the type of bicycle facility present at a road section. The Fietsersbond dataset contains details on the type of road where a distinction was made between the road types mentioned in Table 4.1 (Fietsersbond, n.d.). These were reclassified as shown in the right column. The main goal of the reclassification was to distinguish road types based on the segregation of cyclists from other road users. The reclassification of several road types to mixed road is important, as it indicates that cyclists are not separated from motorists. These were thus considered less safe than the other road types.

Table 4.1. Road type reclassification

Road type	Reclassification
[null]	[null]
Onbekend	
Ventweg	Mixed Road
Normale weg	
Voetgangersgebied	
Veerpont	
Voetgangersdoorsteekje	Non-cycling path
Solitair bromfietspad	
Bromfietspad (langs weg)	
Fietsstraat	Separated cycling path
Solitair fietspad	
Weg met fiets(suggestie)strook	Bicycle lane
Fietspad (langs weg)	

Speed limits

For each road section, the maximum car speed is given, which has ranging values from 15 km/h to 130 km/h (Fietsersbond, n.d.). Speed limits up to 30 km/h were considered safe. All road sections with a speed limit over 30 km/h were classified unsafe (Jestico et al., 2016; Sener et al., 2009a).

Street lighting

Another attribute that was derived from the Fietsersbond is a valuation of the presence of street lighting, which is classified as follows for each road section: not lit, partially lit or well lit (Fietsersbond, n.d.). In order to determine the effect of the presence of street lighting more accurately, it is taken into account if a route is cycled through the dark or in daylight. This was based on the time that is tied to the point dataset. A route cycled between 8.00 a.m. and 17.00 p.m. was given the value 'light'. All other routes were given the value 'dark'.

Crime rates

The crime rates per neighbourhood form the data behind the 'criminal assault factor', which affects the social safety. The dataset gives the actual number of crime rates per neighbourhood in 2020 in a csv file (CBS, 2021b). This file was joined with the District and Neighbourhood Map of the Netherlands (CBS, 2020b), this was done based on the neighbourhood codes specified in both datasets. This resulted in a polygon dataset, which had to be joined to the Fietsersbond network to tie the number of crimes to a road section. The higher the crime rate, the more unsafe a road is deemed.

Environment type

The type of environment can influence the social safety as well. These are already attributed to the Fietsersbond dataset in the following categories:

- Unknown
- Forest
- Nature (except forest)
- Acres/grasslands
- Rural areas
- Built environment (many green areas)
- Built environment (little to no green areas) (Fietsersbond, n.d.).

Forest and Nature were considered less safe. Both built environment categories were considered safe (Bohle & Verkehr, 2000).

4.3.2 Network enrichment

For each safety dataset, the data was aggregated to the bicycle network shapefile in order to use the data for the generation of alternative routes and statistical analysis. Besides, data on water, environment and lighting were required for the green alternative. Data on stop signs, traffic lights and intersections were required for the fastest and continuous routes (retrieved from: (OpenStreetMap Contributors, 2020)). In order to enrich the Fietsersbond network with the spatial data, various methods were used: either reclassifying the attributes or joining data to the shapefile.

Reclassification

The datasets on road type, speed limits, street lighting, type of environment, water and intersections were derived from the Fietsersbond network dataset and were thus already attributed to the road sections. In order to use these datasets for the generation of alternative routes as restriction fields, reclassification into numerical values was required (this will be further elaborated in Section 4.5). The specific numerical values given in the reclassifications can be read in Appendix D.

Spatial joins

The remaining datasets on accidents, crime rates, stop signs and traffic lights still needed to be matched to the road sections. The crime rate dataset consists of neighbourhood polygons with a number of crimes tied to it. These were spatially joined by taking a mean of the neighbourhood crime rates a road crosses. Figure 4.7 visualises how a route can cross multiple neighbourhoods with different crime rates: the average of these crime rates was extracted and assigned to the complete route.

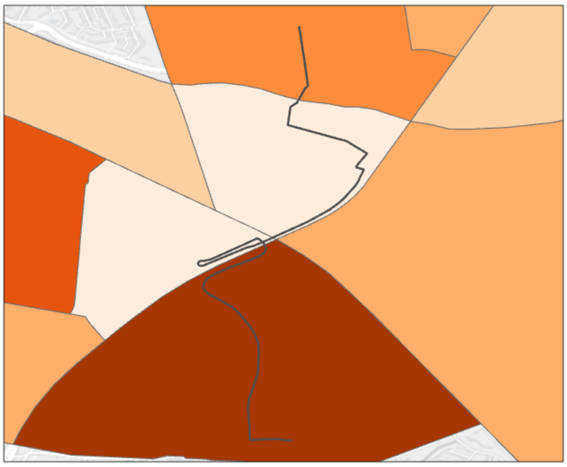


Figure 4.7. Route and neighbourhood crime rates

The accident dataset consists of point features, for which a buffer of 25 meters around the accident was used to then spatially join it to a road section. For both the stop signs and traffic lights a buffer of 15 meters was created. In these shapefiles there are a lot of overlapping buffers, which resulted in a large amount of stop signs/traffic lights for each road section at an intersection, where in fact there was for example only one traffic light at that section. Therefore, the boundaries of these buffers are dissolved. Subsequently, all datasets are spatially joined to the Fietsersbond shapefile. The workflow of this process is shown in Figure 4.8.

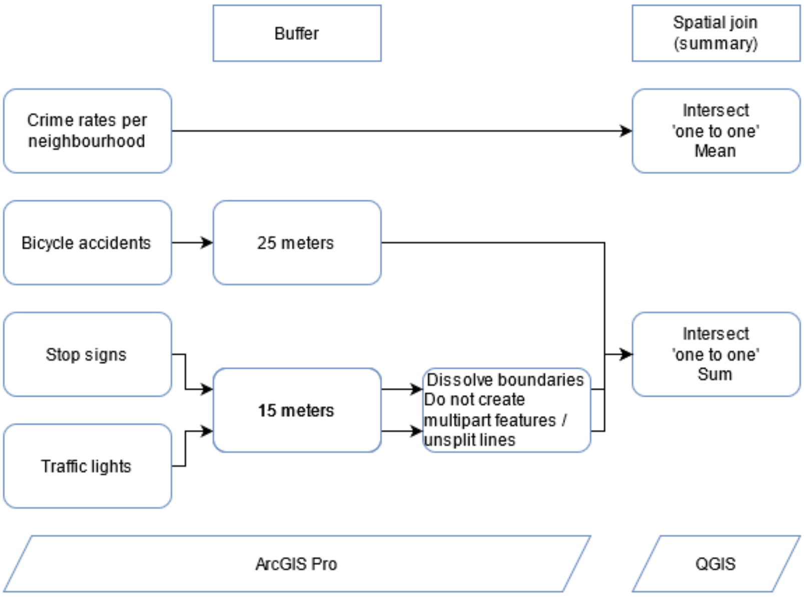


Figure 4.8. Process spatial data joins

4.4 Data preparation – Fietsersbond network

The Fietsersbond shapefile with all safety attributes added was used as input for the network creation. However, before the network dataset could be created, the topology of the shapefile had to be fixed, which will be covered in the next paragraph, whereafter the network requirements will be discussed.

4.4.1 Network topology

A network topology represents the spatial relation between line segments in a network. There are three relationships that are important in network topology: adjacency, connectivity and containment. Blake (2010) defined the following characteristics of a line network topology: the end point of segment can connect to the end point of another segment, this connection is called a node. Secondly, a line segment can cross over a node but not connect to that node or the line segments that are joined by that node. Thirdly, a road segment can intersect other road segments. Finally, a line segment can be terminated along another segment, which is represented with a node (Blake, 2010). In ArcMap these topology characteristics are operationalized with certain rules that a network must adhere to. A few examples of these rules are 'Must not have dangles', 'Must not intersect' and 'Must not have pseudonodes' (ArcGIS Pro, n.d.). The original Fietsersbond shapefile appeared to have some of these topology errors, such as lines that were not properly connected. In Figure 4.9 the problem is visualised. This created a problem in generating accurate alternative routes. In order to solve these errors, the following steps were taken in ArcCatalog and ArcMap:

1. Import the Fietsersbond network in a personal geodatabase and create a new topology in ArcCatalog with the rule: 'Must not have dangles'.
2. Import the topology in ArcMap.
3. Start editing and open the error inspector window to select the errors.
4. Solve the errors by using the trim and snap function with a distance of 0,2 meters.
5. Add a new rule: 'Must not intersect or touch interior'.

6. Select all point errors and split them.
7. Add the rule: 'Must not have pseudo nodes'.
8. Select all errors and use the merge to largest function.

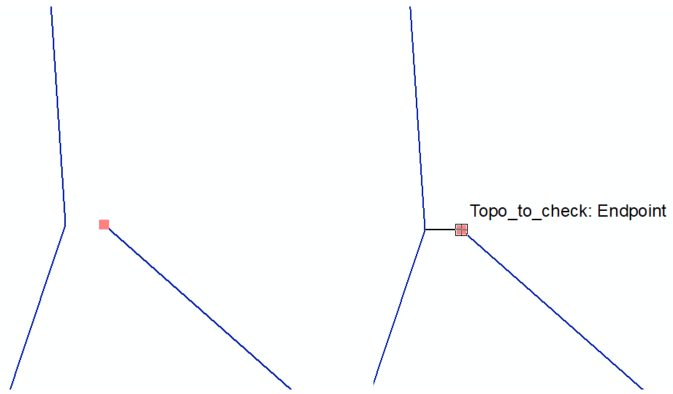


Figure 4.9. Topology error fix

After these steps were taken this network was used to generate alternative routes, assuming that the network was correct. However, after examining the alternatives closely it was found that the generated routes were not always accurate. This became mainly clear since the overall average length of the shortest routes was substantially longer than the average length of most other routes. This can be seen in Table 4.2, where the average length in kilometres of the primarily generated routes is given in the left column (the continuous route was not generated in the first instance).

Table 4.2. Average length alternative routes

Route	Average length	New average length
Observed	3.16	4.29
Social Safe	2.79	4.35
Traffic safe	3.08	4.29
Green	2.93	4.66
Shortest	2.98	3.96
Fastest	2.74	4.04
Continuous	-	4.08

It was found that the inaccurate generation of routes was the result of the last step in solving the errors: 'merge to largest'. As a consequence of this function, multiple segments were merged to one segment as shown in Figure 4.10, where the light blue selected segments in each map are merged to one segment.

Figure 4.10. Segment merging problem

Whenever the routing algorithm is searching for alternative paths, and it wants to use one of these road sections it will have to use the complete segment and thus go around the corner. This resulted in some road sections to become uncrossable in the network analysis. This is shown in Figure 4.11, where the red route is the observed route. The yellow route is the shortest route with the firstly generated network, where you can see that it avoids to cross the road segments visualised in Figure 4.10. The blue route in the right figure is the shortest route created based on the new network shapefile. Here the road sections are split by using the 'split lines' tool in ArcGIS Pro.

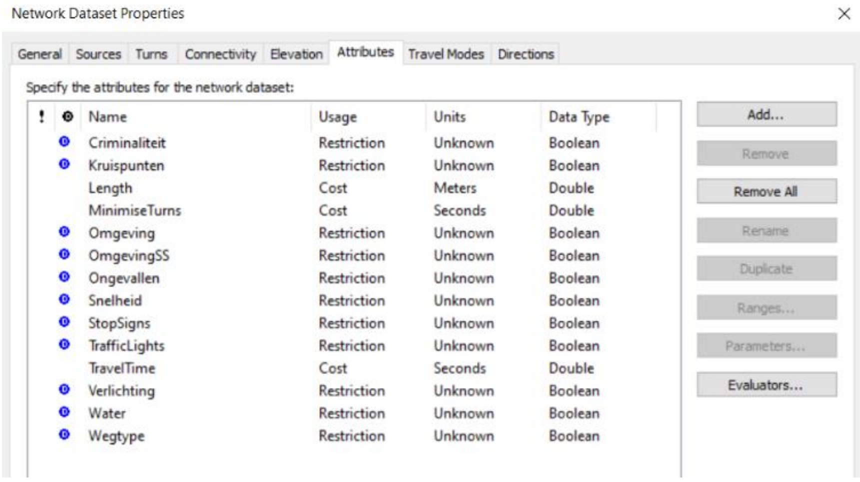
Figure 4.11. Example inaccurate route generation due to topology errors

For some routes the network was a problem in generating proper alternatives, for other routes it was the fact that it concerned round trips. This meant that the origin and destination are closely located to each other, as visualised in the left map of Figure 4.12, where the red route is the observed route. Using these OD-points as input results in the generation of an exremely short route between these two points, as visualised with the yellow route in right map in Figure 4.12. After examination of the route lengths, a selection was made based on the condition that alternative routes were substantially longer than the observed route. For the routes that appeared to be round trips, the choice was made to create extra points on the route: every 10 percent an extra point is added. This was then used as input for the routing model in ArcMap, which will be discussed in 4.6. In this way it is possible to create alternative routes for the recreational round trips as well. The blue route in the right map presents the result. Both fixing the topology errors and adding extra stops to recreational routes resulted in more accurate average lengths for the alternatives as can be seen in Table 4.2.

Figure 4.12. Example inaccurate route generation round trips

4.4.2 Network specifications

After the complete preparation of the Fietsersbond network shapefile, a network dataset was created to use in the network analysis. In order to calculate the alternative routes, different cost factors and restrictions were used for each alternative, these needed to be attributed in the network dataset (Figure 4.13). In order to be able to use multiple factors in the generation of the labelled alternatives, Length, TravelTime and MinimiseTurns were used as cost factors and all other attributes were used as restrictions.



Name	Usage	Units	Data Type
Criminaliteit	Restriction	Unknown	Boolean
Kruispunten	Restriction	Unknown	Boolean
Length	Cost	Meters	Double
MinimiseTurns	Cost	Seconds	Double
Omgeving	Restriction	Unknown	Boolean
OmgevingSS	Restriction	Unknown	Boolean
Ongevallen	Restriction	Unknown	Boolean
Snelheid	Restriction	Unknown	Boolean
StopSigns	Restriction	Unknown	Boolean
TrafficLights	Restriction	Unknown	Boolean
TravelTime	Cost	Seconds	Double
Verlichting	Restriction	Unknown	Boolean
Water	Restriction	Unknown	Boolean
Wegtype	Restriction	Unknown	Boolean

Figure 4.13. Network dataset properties

For each cost- and restriction factor the field on which it is based had to be specified. These fields were derived from the Fietsersbond shapefile. For the restriction attributes it was specified if one wants to avoid certain values or prefer certain values. For example, 'avoid high' means that road sections with high crime rates are completely avoided, whereas 'prefer low' means that all road sections can be taken but with a preference for road sections with low crime rates. Therefore, the choice was made to only use the 'prefer' function, the specific restrictions used for each alternative are discussed in Section 4.5

As for the cost factors, Length was based on the length field in the shapefile, calculated with the 'calculate geometry' tool. Travel time was also added as a new field in the shapefile with the following formula:

$$Travel\ time = \frac{\frac{15}{3.6}}{Length} \tag{3}$$

where an average cycling speed of 15 km/h was assumed. This was converted to m/s by dividing it by 3.6. Subsequently, this was divided by the length of a segment to attribute the travel time in seconds to each road section. In this way, time could be used as a cost factor that adds time penalties for turns in a route. This means that the costs rise with the amount of turns in a route. This field is the source for TravelTime and MinimiseTurns as a cost factor in the network dataset. For both cost factors the costs of a turn ('global turn delay') varied, this will be specified in the next section.

4.5 Choice set generation

In 4.2 it is elaborated how the OD-pairs were subtracted from each route, from this a shapefile consisting of an origin and destination point resulted. In order to use the OD-pairs in the choice set generation in ArcMap, a stops dataset had to be created. For this, two stops datasets are created: one containing the OD-pairs where routes go from A to B, and one containing the round trips that contain additional points as discussed in Section 4.5.1. Figure 4.14 provides a schematic overview of the steps required to generate the stops dataset, which was then used as input for the route model. The resulting routes were then joined to the Fietsersbond dataset to attribute the spatial data to the routes. The scheme also presents the different software that were used for this.

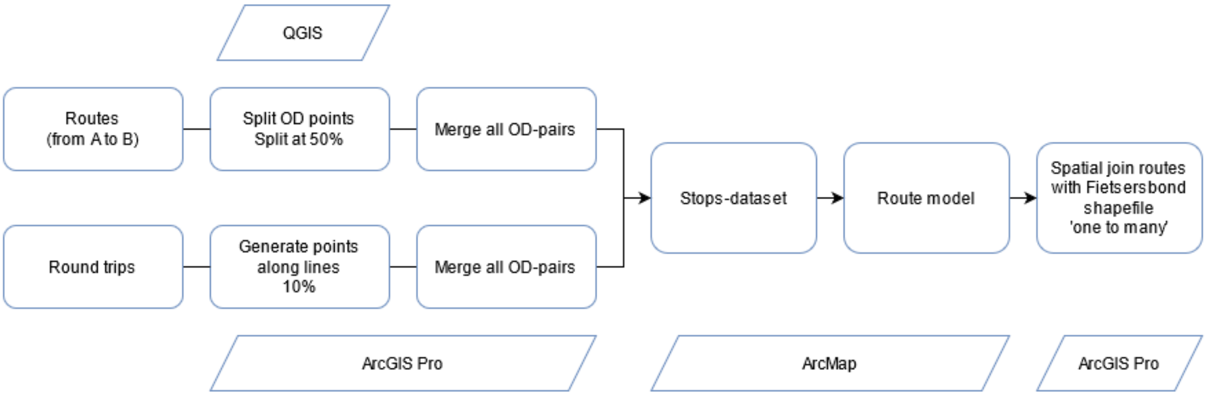


Figure 4.14. Research steps generation route alternatives

The network dataset and the stops dataset were used as input for the model that generates the alternative routes. The model can be seen in Appendix E. A route layer was first created, which used the stops dataset and properties of a certain alternative to generate routes between the origin and destination points. For each generated alternative the same model was run with different properties: the costs and restrictions attributed to the network dataset as presented in Section 4.4.2. In Table 4.3 the costs and restrictions that were used as properties for each alternative can be read. The output was a shapefile with a row for each route, this dataset had to be spatially joined with the Fietsersbond shapefile to attribute all spatial data to the routes.

Table 4.3. Costs and restrictions route alternatives

Alternative	Cost	Global turn delay	Restrictions	Specification
Green	Length		Water	Prefer: High
			Environment	Prefer: High
			Street lighting	Prefer: High
Social Safe	Length		Crime rates	Prefer: Low
			Street lighting	Prefer: High
			Environment	Prefer: Medium
Traffic Safe	Length		Road type	Prefer: High
			Speed limits	Prefer: Low
			Street lighting	Prefer: High
			Accidents	Prefer: Low
Shortest	Length			
Fast	Travel Time	Straight across no road = 0 Straight across road = 4 Reverse = 10 Right turn = 2 Left turn = 6	Stop signs	Prefer: Low
			Traffic lights	Prefer: Low
			Intersections	Prefer: Low
Continuous	Minimise Turns	Straight across no road = 0 Straight across road = 0 Reverse = 30 Right turn = 20 Left turn = 20	Stop signs	Prefer: Low
			Traffic lights	Prefer: Low
			Intersections	Prefer: Low

4.6 Statistical analysis

4.6.1 Spatial indicator preparation: safety data

After the spatial join with the network dataset, each route consists of a set of rows for every segment it crosses. For the eventual data analysis a single value for a specific spatial factor is required. This means that the values attributed to each segment were aggregated to a value for the complete route. This was done by means of the script presented in Appendix F. The result of these lines of codes is a large excel file with all spatial indicators for each route and the alternatives in the columns, as depicted in Table 4.4 with three example variables. Crime rates are the average number of crimes of the neighbourhoods a route crosses. Then street lighting is based on a valuation of the measure of street lighting: ‘not lit’, ‘partially lit’ or ‘well lit’. ‘Partially lit’ and ‘well lit’ are here combined into the percentage of street lighting on a route. Separated bicycle paths also gives a percentage of the amount of separated bicycle paths: the observed route crossed separated bicycle paths for 58,75% of the route and the social safe alternative 67,57%.

Table 4.4. Example excel input file for statistical analysis

Route	Alternative	Crime rates	Street lighting	Separated bicycle path
T03_4_01	Observed	68,49	89,6%	58,75%
T03_4_01	Shortest	88,99	90,07%	66,66%
T03_4_01	Social Safe	96,67	94,35%	67,57%
T03_4_01	Fastest	81,06	53,96%	35,95%
T03_4_01	Traffic Safe	96,27	93,78%	73,25%
T03_4_01	Green	87,38	96,78%	64,48%
T03_4_01	Continuous	66,78	73,4%	36,78%

The spatial variables prepared to be used in the choice model represent the following:

- *Crime rates*: the average number of crime rates of the neighbourhood a route crosses.
- *Accidents*: the number of accidents on the segments of a specific route.
- *Street lighting*: the percentage of the route that is lit.
- *Speed limit*: the percentage of the route that is considered to have a safe speed limit: a maximum of 30 km/h.
- *Bicycle facilities:* the percentage of the route that crosses a segregated bicycle path, bicycle lane or a non-cycling path.
- *Environment*: the percentage of the route that crosses nature or an urban or rural environment.

Some of the spatial indicators from the previous aggregation steps were recalculated in Stata to be taken into account in the statistical analysis, an overview of their valuation is given in Section 4.7.4, Table 4.6.

4.6.2 Spatial indicator preparation: Path Size Logit (PSL)

The PSL is a factor that takes into account the amount of overlap between alternatives: it punishes routes based on the extent to which the route segments overlaps. If routes completely overlapped, these needed to be removed from the set of alternatives. This was required because otherwise completely overlapping routes would have been double counted in the model, which would result in the model to give faulty results. So, in order to analyse route choices, the routes need to be unique, and then the PSL corrects for the overlapping segments within routes. Therefore, the percentage of overlap between a route and its alternatives is calculated with a Python script (Appendix G). The percentage of overlap is based on the overlap between route links. A unique id was generated for each link in the network. In the calculation, a link that is present in the observed route and also in one of the alternatives was given a value of 1, and otherwise a value of 0 was given. This comparison of link id's was performed for each alternative route. These values of 0 and 1 were then used as input to calculate the percentage of overlap. Based on the percentages, completely overlapping routes (100%) were removed, which means that only one row of the overlapping group of alternatives was preserved, others were removed. This means that the number of alternatives for each OD-pair is now varying; ranging from 1 option to 7 options for a certain OD-pair.

In Figure 4.15, in both maps, it is visualised how certain parts of route alternatives from A to B overlap, and how others do not overlap. The left map visualises that the traffic and social safe route completely overlap, therefore these labels are joined together and only one route is contained. From the maps it for example also becomes apparent that the shortest route shows the largest overlap in this case, and that the green route is also similar to both safety routes.

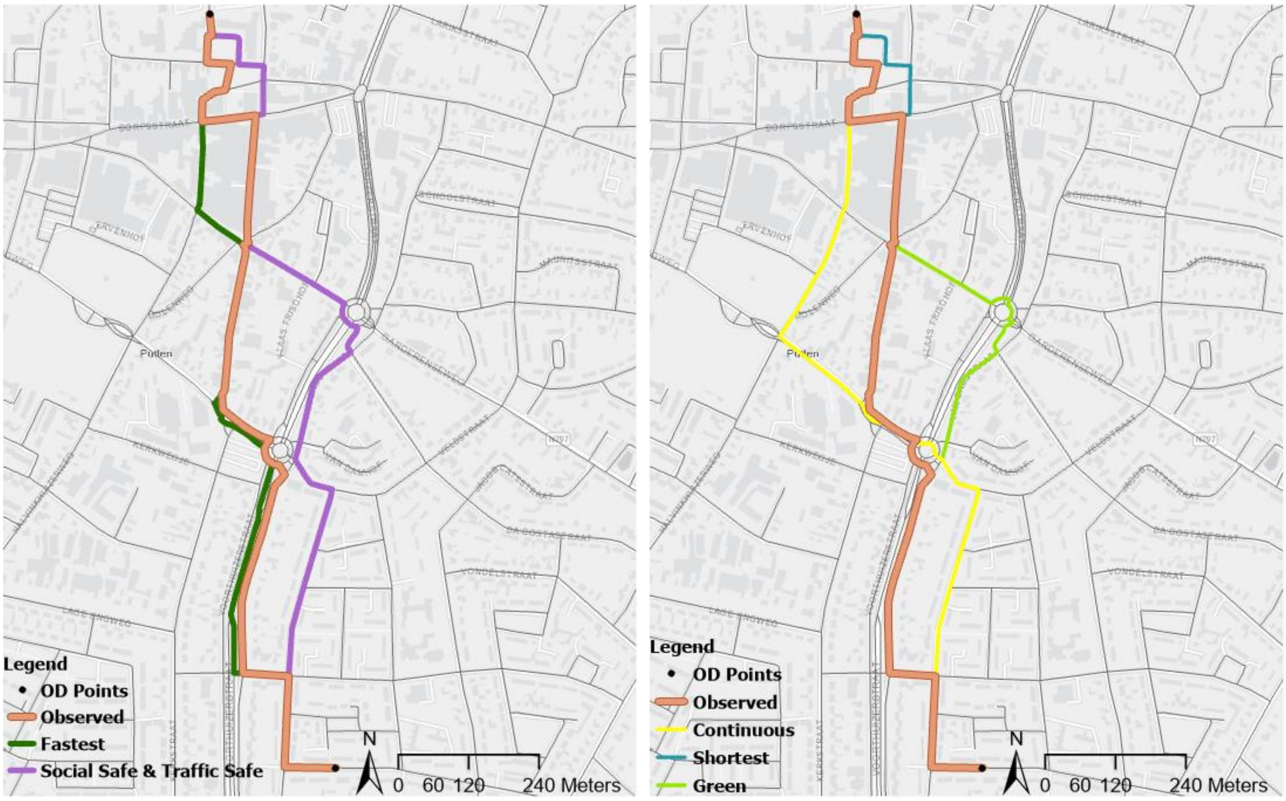


Figure 4.15. Example overlapping route segments

After the removal of all overlapping routes the PSL was calculated using formula 2 (given in Section 3.2). The result is a measure of overlap for each route. If a route is fully independent, with no overlap, PS_{in} would be 1, if routes completely overlap PSL would be 0. This was calculated using the script in Appendix G. In order to adopt the PSL as a variable in the choice model, the natural log was taken of the outcome of the script. This was calculated in Stata using the following function: *generate lnPathSize = ln(PathSize)* (see Appendix J). The resulting factor was adopted in the choice model to correct the model for overlap between routes.

4.6.3 Personal- and trip-characteristics

A few more variables were tied to each route in the dataset. These contain a set of personal characteristics, one temporal characteristic and a characteristic indicating the trip purpose. The personal characteristics were retrieved from the survey. The temporal characteristic indicates whether a person cycled through the dark or in daylight, this was calculated using the script in Appendix H. The trip purpose was tied to each route based on the OD-points of a route. The zip codes of the OD-points were combined with the zip codes of respondents' work place and home address, and the presence of shops in the neighbourhood of the OD-points. Based on these factors the trip purpose of a route was determined (Appendix I). This results in the following additional factors utilized in statistical analysis:

- *Age*
- *Gender*
- *Cycling experience*
- *Daylight*
- *Trip purpose*

4.6.4 Choice model in Stata

As elaborated in Section 3.2, a MultiNomial Logit choice model was used to analyse route choices made by cyclists in Stata. Since cyclists can cycle multiple trips and thus make multiple choices, the 'cmxtmixlogit' command must be used to estimate a panel-data choice model. This command is normally used for a mixed logit model, however the random variation was not specified in this research, which means the estimated model thus resulted in a MNL model.

The variables that were considered to have an effect on route choices can be divided into two groups: the independent variables and the interaction variables. Where the independent variables contain the safety factors and the PSL. The personal characteristics, together with one temporal characteristic and the trip purpose, function as interactions. The interaction variables were deemed to have an effect on how people view the safety factors, which subsequently have an effect on the dependent variables: the route alternatives. This is visualised in Figure 4.16. Besides, the personal- and trip characteristics, two of the independent variables are also presented as interaction variables here. Bicycle facilities and the type of environment are both locational factors that, based on the literature, were deemed to have possible effects on other independent variables.

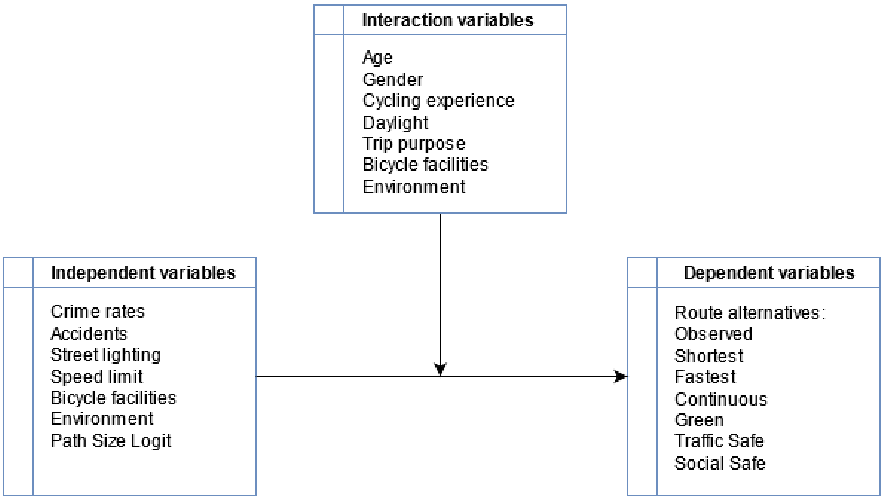


Figure 4.16. Variables included in statistical analysis

In order to estimate the MNL model, a closer look was taken at the descriptive statistics of the variables as well as at the correlation between the variables. These are presented in Table 4.5. Based on the high correlations between the bicycle facilities: non-cycling path and separated bicycle path, non-cycling path was excluded from the model. The same applied to the environment type, where the correlation between a rural and an urban environment was found to be high: rural was removed as a variable from the model. The correlation between urbanity and street lighting was high as well, nevertheless the choice was made to include both variables in the choice model. This choice was made based on the literature review, where both variables turned out to represent different aspects of safety. The remaining variables were adopted in the choice model together with the length and PSL. Additionally, the separated bicycle path and bicycle lane were combined into a new factor to analyse its effect more closely: safe bicycle facilities.

Table 4.5. Correlation matrix

	Crime Rates	Accidents	Street Lighting	Speed Limit	Bicycle Lane	Non Cycling Path	Separated Bicycle Path	Rural	Urban	Nature
Crime Rates	1.0000									
Accidents	-0.0562	1.0000								
Street Lighting	0.1028	-0.0167	1.0000							
Speed Limit	0.1254	-0.2475	0.4305	1.0000						
Bicycle Lane	0.0133	0.1349	0.0755	-0.2866	1.0000					
Non Cycling Path	-0.0131	0.0230	-0.4765	-0.2965	-0.4423	1.0000				
Separated Path	0.0040	-0.1267	0.4523	0.5339	-0.2836	-0.7346	1.0000			
Rural	-0.2575	-0.2518	-0.4449	-0.1199	-0.1291	0.0139	0.0829	1.0000		
Urban	0.1732	-0.0528	0.8710	0.4855	0.1640	-0.4983	0.4087	-0.5805	1.0000	
Nature	-0.0497	-0.0731	-0.2271	-0.0840	-0.0511	0.0443	-0.0087	0.0155	-0.1987	1.0000

An overview of the calculation and valuation of all spatial indicators, personal characteristics and trip- and temporal characteristics used for statistical analysis is given in Table 4.6.

The script in Appendix J shows how the choice model was estimated in Stata. From the removal of overlapping routes, 34 routes ended up with only one option: all labelled alternatives completely overlapped with the observed route in these cases. This means there were no other choices in these 34 cases, therefore, these were removed from the choice model when the command was executed. Furthermore, the script depicts how the relation between single variables and route choice was determined, as well as how the relation between the interaction variables and the spatial indicators was calculated using Stata. In Chapter 5 the results of these calculations will be discussed.

Table 4.6. Explanation, units and values of variables

t	
Value indicating the route and its corresponding alternatives.	
id	
Id value given to each route a participant has taken (and the generated alternatives).	
choice	
Observed route (chosen route)	1
Alternative routes	0
length	
Length of a route in kilometres.	
Crime rates	
Average annual number of crime rates in the neighbourhoods a route crosses.	
Accidents	
Number of accidents on the roads a route traverses.	
Street lighting	
The percentage of a route that is lit: based on the Likert-scale values 'partially lit' and 'well lit'.	
Urban	
The percentage of a route that crosses a built environment, based on the categories: 'built environment (little to no green areas)' and built environment '(many green areas)'.	
Rural	
The percentage of a route that crosses a rural environment, based on the categories: 'rural or village' and 'acres/grasslands'.	
Separated bicycle path	
The percentage of a route that crosses a separated bicycle path, based on the categories: 'bicycle street' and 'solitary bicycle path'.	
Bicycle lane	
The percentage of a route that crosses a bicycle lane, based on the categories: 'road with bicycle suggestion lane' and 'bicycle lane along the road'.	
Non cycling path	
The percentage of a route that crosses no cycling facilities, based on all remaining road type categories: reclassified as 'mixed road', 'non-cycling path' and 'null' in section 4.4.1.	
Safe bicycle facilities	
A combination of separated bicycle paths and bicycle lanes: the percentage of a route that crosses both facilities.	
lnPathSize	
The natural log taken from the PSL, indicating the overlap between route alternatives.	
Age	
The age of the respondent, calculated by subtracting the birth year given by respondents from 2021.	
Gender	
Male	1
Female	0
Cycling experience	
Daily	0
Approximately 1 day per week	1
Approximately 2 day per week	2
Approximately 3 day per week	3
Approximately 4 day per week	4
Approximately 5 day per week	5
Approximately 6 day per week	6

5 Results

This Chapter presents and discusses the results from the spatial and statistical analysis. Firstly, an overview of the descriptive statistics resulting from the spatial data and survey will be given. Secondly, the survey outcomes with regard to the indicated preferences will be given. Thirdly, the choice model output will be presented to indicate the effect of safety factors. Each individual safety factor and possible interactions with other factors will be presented. Then, a comparison will be made between the impact of traffic safety and social safety. Finally, an overview of the influencing variables will be given.

5.1 Descriptive statistics

The descriptive statistics of the different spatial variables of the observed routes (n=451) are presented in Table 5.1. For each observed route a maximum of 7 alternative routes was created, which sums up to a total of 1,836 alternatives. The length of each route is given in kilometres, with the shortest trip having a length of 230 meters and a maximum length of 89.29 kilometres for all trips. On average respondents travelled 4.29 kilometres by bicycle per trip. Table 5.1 also describes the path size variable. As a path size of 0 means complete overlap and 1 indicates a unique route, a mean of 0.59 and standard deviation of 0.19 implies that the generated alternatives perform reasonably well with only some overlap. Bicycle accidents and crime rates for each observed route and the generated alternatives are given in absolute numbers. The average annual number of crimes on routes is rather high, but has a high standard deviation, which indicates that the variation in crime rates between routes was quite large. The number of accidents counts up to 17 on average per route. All other factors are given in percentages of the complete route. More than 60% of the routes is on average cycled through an urban environment. The percentage cycled through rural and natural environments are substantially lower, this is also caused by the fact that there is an average of almost 25% missing data for the type of environment on a route. Bicycle facilities seem to be absent for more than 40% of all routes, however, the majority of the routes crossed separated bicycle paths. For around 70% of the routes street lighting was present. Less than one third of the routes crossed roads with speed limits below 30 km/h. However, the share of missing data for speeds limits was large as can be seen in Table 5.1.

The descriptive statistics of the respondents' trip- and personal characteristics are presented in Table 5.2. The average age of the respondents is 35 years old, ranging from 16 to 73 years old. Besides, women made up the majority of the research group, and they also cycled the vast majority of the trips. The group's cycling experience is high with the majority of the group going for daily bicycle rides. Most of these trips have a utilitarian or 'other' (undefined) purpose, and are primarily cycled in daylight.

Table 5.1. Descriptive statistics spatial indicators

	Mean	**Std. Dev.**	**Min**	**Max**
Length (km)	4.29	6.44	0.23	89.29
Path Size	0.59	0.19	0.21	1.00
ln(PathSize)	-0.58	0.34	-1.58	0.00
Crime rates (nr.)	667.91	853.29	14.78	8,449.15
Accidents (nr.)	17.14	20.24	0.00	104.00
Urban (%)	66.06	28.54	0.00	100.00
Nature (%)	2.13	10.45	0.00	100.00
Rural (%)	8.94	18.40	0.00	100.00
Separated bicycle path (%)	56.54	25.10	0.00	100.00
Bicycle lane (%)	30.69	21.71	0.00	100.00
Non cycling path (%)	42.76	25.24	0.00	100.00
Street lighting (%)	70.24	25.05	0.00	100.00
Speed limits (%)	27.66	24.78	0.00	100.00
Missing data environment (%)	22.88	20.68	0.00	100.00
Missing data speed limits (%)	55.96	26.85	0.00	100.00
n=451				

Table 5.2. Descriptive statistics personal- and trip-characteristics

Continuous				
	Mean	**Std. Dev.**	**Min**	**Max**
Age	35.67	16.51	16	73
Categorical				
	Categories	**Numerical value**	**Frequency**	**Percentage**
Respondents gender	Female	0	40	65.57
	Male	1	21	34.43
Trip gender	Female	0	335	74.28
	Male	1	116	25.72
Cycling experience	1 day (a week)	1	1	1.64
	2 days	2	5	8.20
	3 days	3	5	8.20
	4 days	4	5	8.20
	5 days	5	6	9.84
	6 days	6	7	11.48
	Daily	0	32	52.46
Trip purpose	Commute	1	42	9.31
	Recreational	2	30	6.65
	Utilitarian	3	185	41.02
	Other	4	194	43.02
Day/ night	Dark	0	0	20.84
	Light	1	1	79.16

5.2 Survey results

In Appendix K the exact results from the relevant statements in the survey can be read. From the survey it appeared that most people did not necessarily avoid unsafe routes when riding their bicycle, as can be seen in the histograms in Figure 5.1 (in some cases the question was not filled in, this is indicated with a '-'). Regarding traffic safety there is no clear preference to cycle on roads where speed limits for cars do not exceed 30 km/h, or a wish to avoid busy traffic circumstances. Participants did generally agree to prefer to cycle on separated bicycle paths: 48,33% agreed and 20% totally agreed.

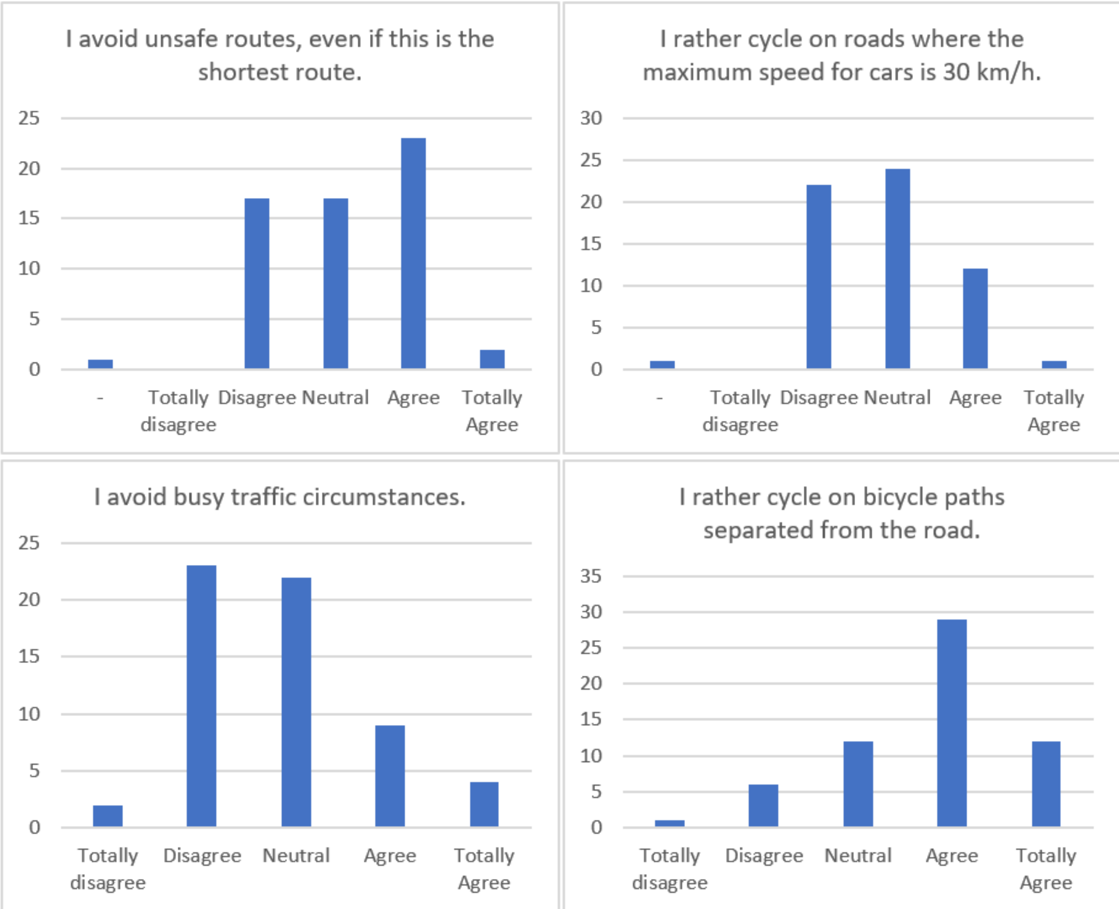


Figure 5.1. Histograms survey statement responses 'Traffic safety'

Concerning social safety, Figure 5.2 visualises that most people stated to take this into account in their route choices. Regarding street lighting, people stated to prefer to take roads with proper street lighting when it is dark. With regard to avoiding abandoned areas when cycling, the opinions seem to be rather divided, since 5.00% totally disagreed, 31.67% disagreed, 21.67% was neutral, 33.33% agreed, 8.33% totally agreed. In social safety preferences, gender appears to play a role. Female respondents tended to agree or totally agree to the statements visualised in Figure 5.2, while it seemed less important for men as they were more likely to disagree (see Appendix K for detailed percentages).

In the next Section, the actual route choice behaviour of respondents will be discussed based on the results from the spatial analysis.

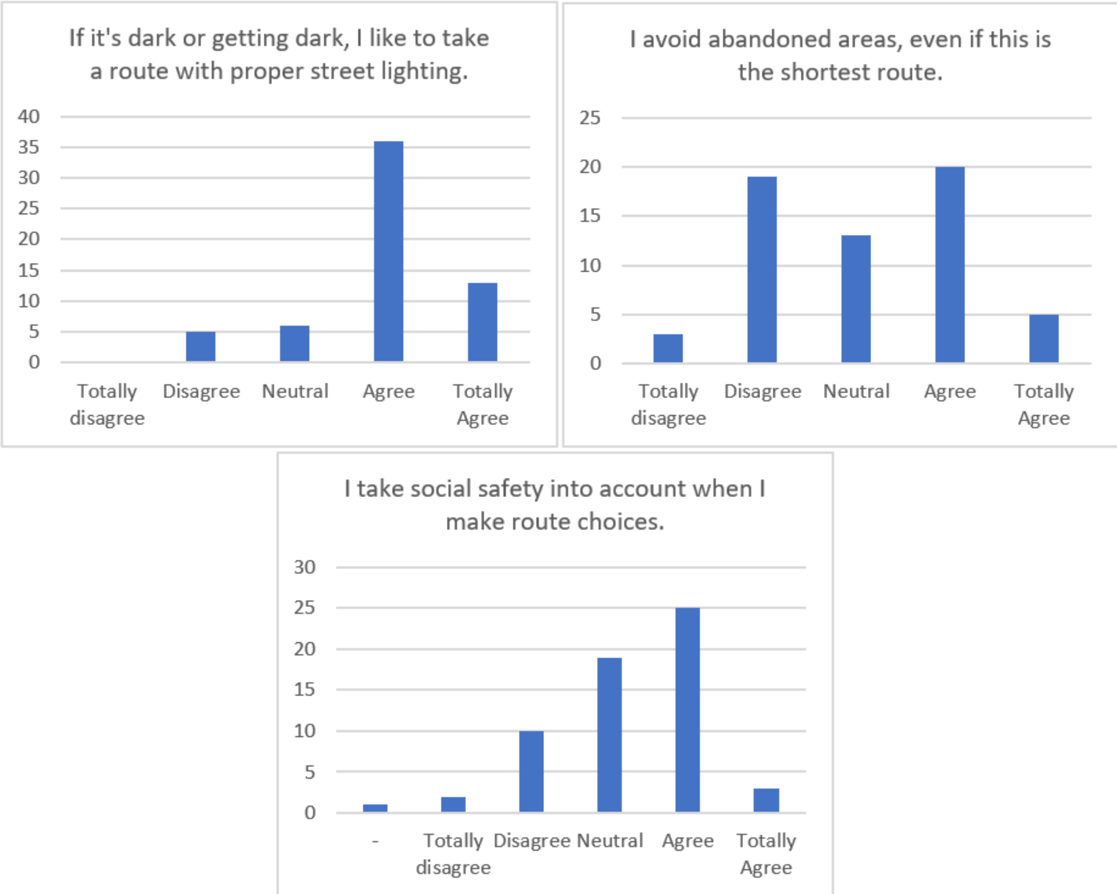


Figure 5.2. Histograms survey statement responses 'Social safety'

5.3 Choice model

The outcomes of the MNL model are presented in Table 5.3. The value of $p<0.01$ for 'Prob > Chi2' indicates that the model is statistically significant and thus has a good fit. The odds ratios indicate the impact of a variable on route choice probability. A value >1 indicates a positive relation and a value <1 indicates a negative relation. The odds ratios can be interpreted as the probability that someone will choose for that specific variable in route choices: in case of an odds ratio of 1.3 for example, a person is 30% more likely to choose the route with every extra unit of that variable. 90%-, 95%- and 99%-confidence levels are applied to determine the significancy of a variables' impact. The results of the model reflect the effect of the variables on the route choices made by respondents. The path size, urbanity, nature, street lighting, speed limits and bicycle lanes showed significant impacts on route choice.

The path size variable was estimated to be positive and significant. In line with the explanation given by Dane et al. (2020) this was as expected, because the factor needs to be positive in order to correct for the overlap between route alternatives. The significance shows that there is indeed correlation between overlapping routes.

With regard to the environment types, an urban environment appeared to have a negative impact on the route choices made by respondents, indicating that respondents tend to avoid urban areas. Nature was found to have a significant positive effect on route choice: with every 1% extra nature a route crosses, a route is 12% more likely to be chosen. For street lighting a significant positive effect is observed, which means people preferred to choose routes with a higher percentage of street lighting. Speed limits up to 30 km/h have a positive impact on route choices of respondents, meaning that with every extra percent of safe speed limits, people were

2% more likely to choose that route. Furthermore, the presence of bicycle lanes was found to be significant, where respondents were more likely to choose for routes with a higher percentage of bicycle lanes. However, effects from separated bicycle paths were not significant. For social safety variables, aside from street lighting, no significant relations with route choices were found. Furthermore, length is not significant, which means that the length of a route had no impact on the probability that a route would be chosen.

In the next Section, the individual effect of each specific safety variable will be further elaborated, together with its interactions.

Table 5.3. Choice model output

Variable	Odds Ratio	Std. Err.	z	P>|z|	[95% Interval]	Conf.
Length	0.964	0.064	-0.55	0.581	0.847	1.098
lnPathSize	5.635	1.126	8.65	0.000***	3.809	8.337
Urban	0.966	0.014	-2.43	0.015**	0.940	0.993
Nature	1.126	0.036	3.75	0.000***	1.058	1.199
Street lighting	1.024	0.014	1.80	0.071*	0.998	1.051
Speed limit	1.021	0.008	2.78	0.005***	1.006	1.037
Bicycle lanes	1.059	0.007	8.57	0.000***	1.045	1.073
Separated bicycle paths	0.999	0.007	-0.12	0.904	0.986	1.013
Crime rates	0.999	0.000	-0.20	0.840	0.999	1.001
Accidents	1.010	0.008	1.22	0.223	0.994	1.027

Wald chi2(10) = 184.82
Prob > chi2 = 0.000
Log likelihood = -552.96831
No. persons = 59
No. Trips = 417
n = 2,246
Significance: * = p<0.1; ** = p<0.05; *** = p<0.01

5.4 Singular effects of safety variables on route choices

In this Section the relation of each singular safety factor to route choices will be presented. In addition, their interaction with personal- and trip characteristics is elaborated. All Stata results underlying the findings are presented in Appendix M.

5.4.1 Accidents

The incidence of accidents thus appeared to have no significant effect on the route choice probability. However, when looking at the singular effect of this variable, the positive relation was found to have a 1% significance level, which means that people are more would rather cycle on a road with more accidents. On average the observed route is found to have 2 more accidents than its alternatives: ± 19 relative to ± 17 accidents on a route.

Of the respondents preferring a route with accidents, men were 13% more likely to choose a route where more accidents occur, relative to women who were 3% more likely. Also age has a positive significant effect on this factor: increasing age leads to a preference for roads with more accidents. Concerning respondents' cycling experience there is no clear relation between their level of experience and the amount of accidents on route. Accidents are, however, an indicator of objective safety, which can explain the fact that personal characteristics do not play a role. Furthermore, the interaction between accidents and urbanity must be considered, as a positive significant interaction between these factors is found. The fact that 66% percent of the observed routes is travelled through an urban environment can explain the positive relation of accidents on route choice probability. A map of the accidents in the research area is given in Figure 5.3. This visualises that accidents occur more in urbanized areas, especially in the north eastern research areas this becomes apparent: accidents occur in the villages and cities, but less in the more rural areas.

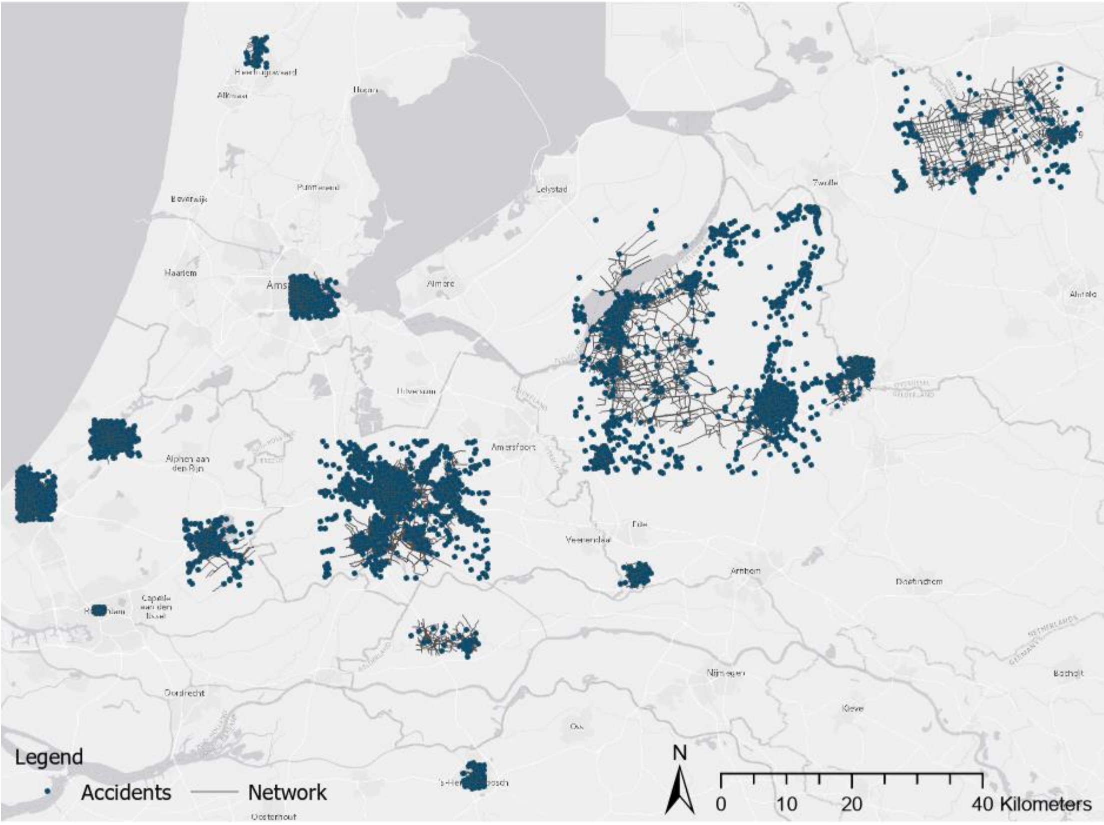


Figure 5.3. Accident location map

5.4.2 Bicycle facilities

As the model indicated, separated bicycle paths do not have a significant influence on the route choices cyclists make. But, when looking individually at this variable, it was found to have a significant negative impact on the route choice probability, indicating that people did not show a preference to cycle on separated bicycle paths. This outcome is also different from what people stated to prefer; which shows a general tendency to agree to rather cycle on bicycle paths separated from the main road. As for the interactions, cycling experience seemed to have varying impacts on preferring separated bicycle paths. Respondents cycling 1 or 2 times a week had a positive relation to choosing for separated bicycle facilities (significant for '2 times a week'). For people cycling more frequently in a week this relation turned negative, showing less preference for separated bicycle paths.

Bicycle lanes did seem to have a positive significant ($p < 0.0001$) effect on route choices: with every 1% bicycle lane extra, a route was 5.9% more likely to be chosen. There was no indicated difference for this factor between the different genders: both men and women show significant positive relations to bicycle lanes, where the odds of choosing for bicycle lanes are slightly higher for men. Concerning respondents' age, it showed a significant positive relation, meaning that the older people are, the more likely they are to choose to cycle on bicycle lanes. Varying levels of cycling experience did not play a role in the choice for bicycle lanes.

Geographically, bicycle lanes are predominantly located in urban areas, whereas separated bicycle paths are mostly located in less urbanized areas. The map in Figure 5.4 visualises this for the more eastern research area around Putten and Apeldoorn. Figure 5.5 shows the bicycle facilities in the city of Amsterdam. In Figure 5.5 it becomes apparent that separated bicycle paths are generally located in more rural areas where the network is less dense. In the dense network of the city of Amsterdam the majority of bicycle facilities consists of bicycle lanes, as can be seen in the map. This geographic relation is also shown by the interaction of both factors. The relation between bicycle lanes and urban environment is positive significant, whereas for separated bicycle paths and urbanity a negative significant relation is indicated.

The combination of both bicycle facilities into one safe facility-factor resulted in a positive significant relation to route choice probability: with every extra percentage of safe bicycle facilities, the probability of choosing that route increased with 2%. For gender, age and cycling experience the resulting relations were similar as for bicycle lanes. The tendency of respondents to choose for safe bicycle facilities was higher in recreational trips: the odds of choosing a route rose with ±8% with every 1% extra of safe facilities ($p = 0.025$).

When comparing the bicycle facilities' descriptive statistics of the observed routes relative to their alternatives, the mean values for separated bicycle paths were extremely close to each other as shown in Table 5.4. While the difference for bicycle lanes (and thus also safe facilities) showed a slightly greater difference in percentage between the observed route and the alternatives: the observed route had 5% more bicycle lanes on average than the alternative routes.

Table 5.4. Descriptive statistics bicycle facilities

Facility type	**Observed (1) / alternative (0)**	**Mean**	**Std. dev.**
Separated bicycle path	1	56.54%	25.10%
	0	56.23%	24.14%
Bicycle lane	1	30.69%	21.71%
	0	25.32%	17.28%
Safe facilities	1	87.23%	27.23%
	0	81.55%	25.56%

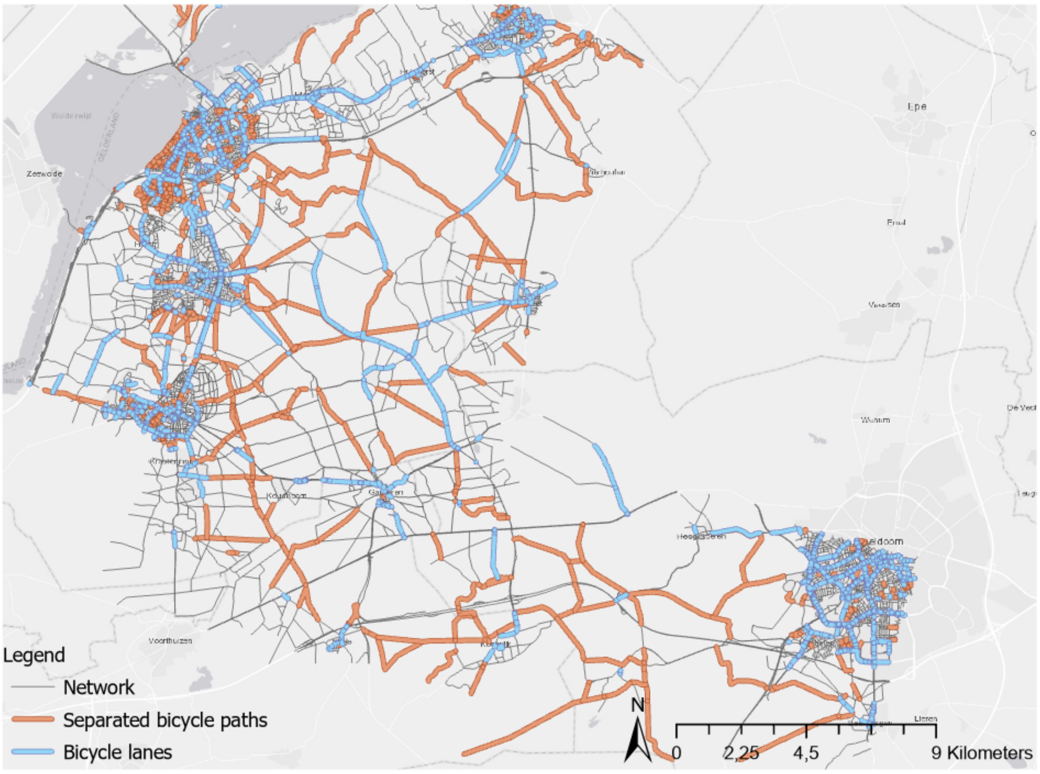


Figure 5.4. Bicycle facilities eastern region

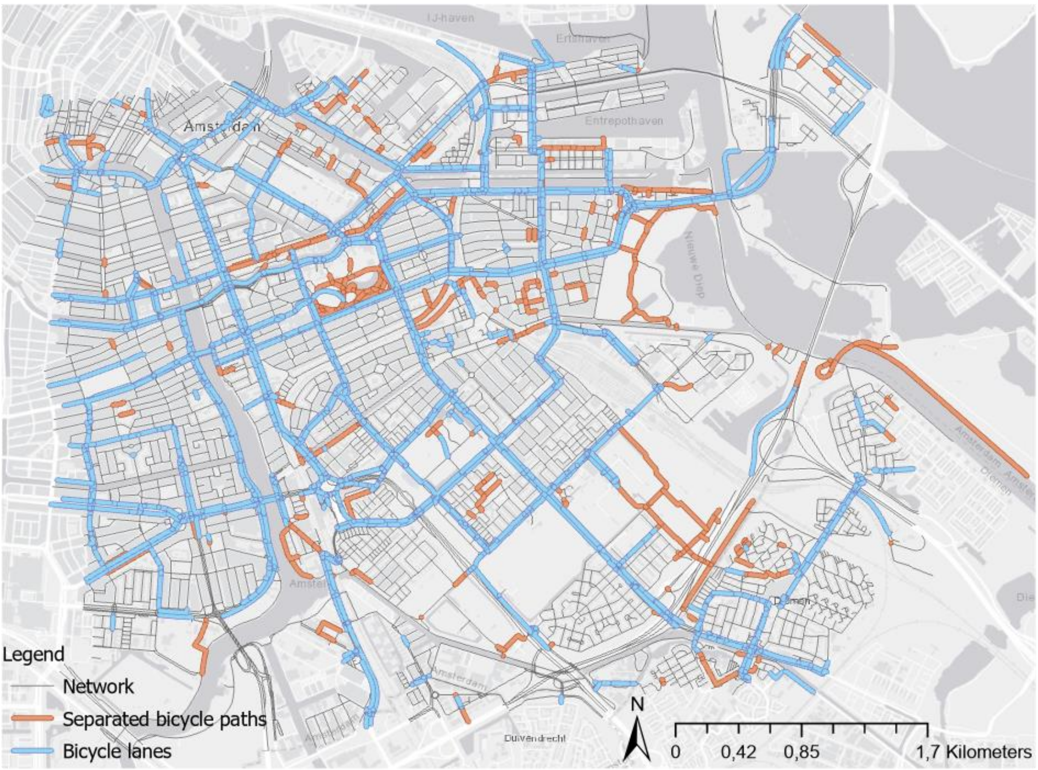


Figure 5.5. Bicycle facilities Amsterdam

5.4.3 Speed limit

As stated in 5.3, safe speed limits (up to 30 km/h) had a significant positive influence on the route choices respondents make. The impact is small though: with every extra percent of roads with safe speed limits, a route is 2% more likely to be chosen. In line with this, the difference in percentages between the observed- and alternative routes was not large, as can be seen in Table 5.5.

Table 5.5. Descriptive statistics speed limits

	Mean	Std. dev.
Observed	27.66%	24.78%
Alternatives	25.55%	22.50%

When looking at this variable's singular effect, it shows a significant negative impact on route choice probability, which may be due to the small differences. Except for age, personal characteristics did not seem to have significant effects on choosing roads with lower speed limits. Concerning age, elderly turned out to be more likely to choose roads with higher speed limits.

Urban and rural environment types had a negative relation to safer speed limits, meaning that people in urban and rural areas showed no preference to lower speed limits in their cycling behaviour. The majority of the routes was cycled in urban environments, which is a possible explanation for the singular negative effect of safe speed limits on route choice. However, speed limits are generally lower in urban areas, so this negative effect was not expected. A possible explanation is the high share of missing data for speed limits in especially urban areas. This is visualised in the map in Figure 5.6 for the city of Utrecht: it becomes apparent that safe speed limits (green road sections) and missing data (yellow road sections) are nearly equally present.

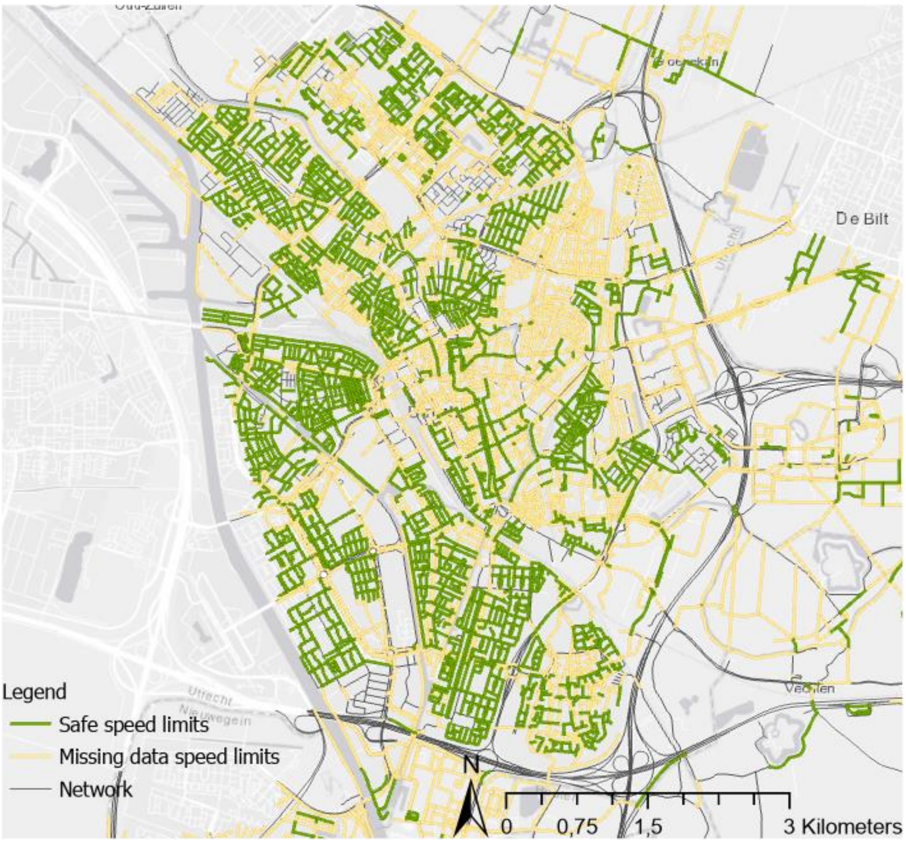


Figure 5.6. Speed limits and missing data in Utrecht

Lastly, there seemed to be an interaction with the type of bicycle facility. Separated bicycle paths show a negative relation to speed limits up to 30 km/h ($p = 0.00$). Whereas for bicycle lanes and non-cycling facilities there is a positive relation, which means that in this case people are more likely to choose for safer speed limits ($p < 0.01$; $p = 0.018$).

5.4.4 Street lighting

Street lighting was found to have a positive relation to route choice probability at a 10% significance level: with every extra percentage of street lighting a route is 2% more likely to be chosen. When looking at the singular effect of street lighting on the odds of choosing a certain route, no statistically significant relation was found. Interactions of personal characteristics to this variable were also non-significant. An explanation for this could be that almost 80% of all routes was cycled in daylight (although no significant relation between cycling in the dark/daylight and the presence of street lighting is found).

Despite these non-significant relations, the factor had a significant positive relation in the choice model. This means that in interaction with all other adopted safety variables, street lighting is estimated to affect route choice probability. Besides, this is in line with respondents' stated preference, where around 80% (of which 55% was female) stated to prefer to take a route with proper street lighting if it is dark or getting dark. Differences between the impact of street lighting in the context of traffic safety and the context of social safety will be discussed in Section 5.5.

5.4.5 Crime rates

Crime rates did not have a significant effect on route choices cyclists made. Also between personal characteristics and crime rates no significant relations were found. The interaction between crime rates and the 'daylight factor' on route choices was non-significant as well.

5.4.6 Type of environment

From the literature review it was expected that an urban environment in the context of safety would have a positive impact on route choices. However, in the choice model an urban environment is found to have a significant negative impact. Looking at urbanity's singular effect the relation is however non-significant.

Concerning the interactional variables, gender and age did not have an impact on the probability of choosing for an urban environment. Furthermore, the time of the day a route is cycled or the presence of street lighting did not have significant relations to choosing for a more urban route. Looking at the differences in the percentage of urban environment between the cycled routes and the alternatives, the observed route generally crossed 3% more urban environment (Table 5.6). Even though a percentage of 66% urban environment on a route is high, relative to its alternative routes it is not substantially higher.

For nature the relation to route choice probability was expected to be negative. In the choice model this was contrary: a positive significant relation on route choices cyclists make was found, indicating that people preferred to cycle through nature. The average percentages of natural environment for the observed and alternative routes did not vary greatly; there is only a 1% difference. But relatively this means that the observed route had twice as much nature as the alternative routes, which explains nature's positive impact on route choices.

With regard to the interactions, both men and women were more likely to choose for a natural environment. For women the chance of choosing for a route with more nature was around 13%, for men this is 7%. Age also appeared to be a significant factor in choosing to cycle through nature, meaning that elderly preferred to cycle through nature. Moreover, the presence of street lighting had a positive significant interaction with nature, although the odds of choosing for a natural environment did not vary between day- or night-time.

Table 5.6. Descriptive statistics environment type

Type of environment	Observed (1) / alternative (0)	Mean	Std. dev.
Urban	1	66.06%	28.54%
	0	63.41%	29.61%
Nature	1	2.13%	10.45%
	0	1.06%	6.44%

5.5 Traffic safety vs. social safety

In this thesis, traffic and social safety are distinguished. Regarding traffic safety, speed limits and bicycle lanes gave a positive significant relation. For social safety the environment type had a contradicting effect on route choice probability: urbanity showed a negative impact and nature a positive impact. Street lighting belongs to both safety types and indicated a significant positive relation in the route choice model.

For traffic safety multiple factors thus show a significant relation. Whereas in social safety only street lighting indicates an impact, which is also limited by the fact that the majority of the route is cycled in daylight. To further compare differences between the effect of traffic safety and social safety, a look can be taken at the percentage of overlap of the traffic safe, and respectively social safe alternative with the observed route. The percentages of overlap for all alternatives are given in Table 5.7. The percentages of overlap lie close to each other (ranging from 45.99% to 53.63%). The shortest routes showed the highest percentage of overlap with the observed routes. However, right after this follow the traffic safe and social safe route. With a percentage of 51.34 the observed route showed a lot of overlap with the traffic safe route. The overlap with the social safe route was also relatively high. Comparing these two percentages, traffic safety thus seemed to have more correspondence with the observed route, indicating that this plays a larger role in cyclist route choice.

Table 5.7. Percentage of overlap with observed route

Alternative	Overlap
Traffic safe	51.34%
Social safe	49.29%
Shortest	53.63%
Fastest	49.08%
Continuous	47.92%
Green	45.99%

5.6 Overview impacting variables

5.6.1 Independent variables

An overview of the relations of all safety variables are presented in Table 5.8. Both the relations resulting from the choice model, as well as singular relations are given. Significant safety variables are thus an urban and natural environment, street lighting, speed limits and bicycle lanes.

*Table 5.8. Overview impact of safety variables (***, **, *: significance at 1%, 5%, 10% level)*

Safety variables	Choice model effect (OR)	Singular effect (OR)
Accidents	1.010275	1.038195***
Separated bicycle paths	0.9991834	0.9853103***
Bicycle lanes	1.059114***	1.052145***
Speed limit	1.021494***	0.98579***
Street lighting	1.024231*	1.000217
Crime rates	0.999937	1.000002
Urban environment	0.9663199**	0.9984555
Nature	1.12629***	1.099068***

5.6.2 Interactions

Table 5.9 gives the possible interactions of the safety variables to the personal- and trip-characteristics. This table only presents an overview of the significant relations between these two types of variables in respect to route choice, for street lighting, crime rates and an urban environment the '-' indicates that no significant relations were found. Age had an impact on the preference for accidents, bicycle lanes and nature with regard to route choice: older respondents preferred to cycle on routes with lots of these variables present. Elderly however did not prefer to cycle on roads with safe speed limits. For gender this applied to the following variables: bicycle lanes and nature, where both men and women showed a preference for these factors. For bicycle lanes the preference was higher among men and for nature the preference was higher among women. An urban environment interacted with the impact of accidents, separated bicycle paths and bicycle lanes. Varying levels of cycling experience only had different and significant impacts on the preference for separated bicycle facilities. Furthermore, relative to other tip purposes, a recreational trip purpose increased the chance of choosing for safe bicycle facilities in route choices. In turn, the type of bicycle facility had an impact on the effect of speed limits on route choice: if respondents were separated from the road, people did not necessarily prefer safer speed limits.

*Table 5.9. Overview safety factors and interaction variables (***, **, *: significance at 1%, 5%, 10% level)*

Safety variables	**Interaction variables**		**Odds Ratio**
Accidents	Age		1.001163***
	Urbanity		1.000297***
Separated bicycle paths	Cycling experience	1	1.210719
		2	1.063134*
		3	0.9417077*
		4	0.9816286
		5	0.9472901***
		6	0.988822
		7	0.9880129**
	Urbanity		0.9998666***
Bicycle lanes	Age		1.001302***
	Gender	Female	1.047609***
		Male	1.070562***
	Urbanity		1.000415***
Safe bicycle facilities	Trip purpose: recreational		1.079349**
Speed limit	Age		0.9996258***
	Separated bicycle paths		0.9997271***
	Bicycle lanes		1.00098***
	Non-cycling facilities		1.000275**
Street lighting	-		-
Crime rates	-		-
Urban environment	-		-
Nature	Age		1.002005***
	Gender	Female	1.134662***
		Male	1.069764**
	Streetlighting		1.001485***

This Chapter presented the results of the study. In Chapter 6 these results will be tied and compared to previous research outcomes (presented in Section 2). Besides, an evaluation of the different aspects of this research process will be given, associated with recommendations for future studies.

6 Discussion

In this Chapter the research results will be discussed and related to findings in previous studies. Subsequently, the limitations of this research and the impact of the worldwide pandemic Covid-19 will be reviewed.

6.1 Research outcome

The results of this study are based on 451 routes and survey responses of 61 respondents. 40 of the respondents are women and 21 respondents are men. The respondents' ages range from 16 to 73 years old. The research area of the study is rather spread out over the mid-region of the Netherlands, and contains urban as well as rural and nature areas. The characteristics of the 451 routes and 1,836 alternatives are analysed in a MultiNomial Logit choice model with a Path Size factor to account for overlap. The outcomes of this choice model indicate the impact of the safety factors on the route choices made by respondents.

Based on literature review, accidents are conceptualized as a factor that constitutes objective safety, therefore its relation to cyclists' route choices was not always obvious in previous studies (Sun, Du, et al., 2017; Sun, Mobasheri, et al., 2017). In this study, the route choice model showed indeed no significant impact for accidents. However, the singular effect of accidents on route choices was found to be significant and the correlation was positive, indicating that more accidents on a road were actually preferred. A possible explanation for this relation is the fact that crashes occur more frequently where traffic volumes are high (Kondo et al., 2018). Relating to cycling behaviour of respondents in this study, it is likely that their chosen routes were already popular and busy, and these choices did not take the number of accidents into account. This is in line with the fact that the majority of the respondents stated to generally not avoid busy traffic circumstances. The interaction with other types of road safety factors explains why accidents did not have a significant impact in the overall route choice model. This indicates that respondents did not take the incidence of accidents into account, probably because in most cases cyclists are unaware of the number of accidents occurring at a certain location. Therefore, the following safety factors are expected to have more impact on route choices.

With regard to bicycle facilities, separated bicycle paths showed no significant relation to route choices. When looking at its singular effect the relation was even significantly negative, indicating that people did not prefer to cycle on separated bicycle paths. Previous studies generally showed a clear preference towards this type of facility, which is why this was an unexpected result. Particularly since the respondents did stated to prefer to cycle on roads where they are separated from the road. The influence of cycling experience on route choices found by Heinen et al. (2010) – who based this conclusion on previous studies – that unexperienced cyclists prefer separated bicycle facilities, is underpinned by the results where more experienced cyclists were less likely to choose to cycle on separated bicycle paths. Furthermore, the findings of Ton et al. (2017) suggesting that separated bicycle paths are not as important in dense networks, can be an explanation for this research where the majority of the routes was cycled in an urban environment. At the same time there is a negative interaction between urbanity and separated bicycle paths, meaning that there are less separated bicycle facilities in urban areas, which could also explain the lower preference for separated bicycle paths since most of the respondents cycled in urban environments. Besides, in rural/natural areas it is more obvious to choose separated cycling facilities than in urban areas where separated bicycle paths are generally situated along large and busy roads, which makes the roads less attractive. The fact that cycling is an important and widely used mode of transport in the Netherlands can also be a reason for the absence of a preference to this type of safe bicycle facilities (Bernardi et al., 2018).

Bicycle lanes, however, did seem to have the expected positive impact on route choice probability in this study. Concerning personal characteristics, age plays a role: the older respondents are, the more probable it becomes that they will choose for a route with a higher share of bicycle lanes. Moreover, the relation between an urban environment and bicycle lanes is positive, which could underpin the preference of bicycle lanes over separated facilities in dense urban areas.

Analysing a combination of both types of facilities into one safe bicycle facility resulted in a positive significant relation. This shows that respondents were in general more likely to choose for safe facilities, as was found in multiple studies (Broach et al., 2012; Casello & Usyukov, 2014; Menghini et al., 2010). The descriptive statistics showed that a single route of a respondent on average consisted of 87% safe bicycle facilities, with such a high share of safe facilities the preference to cycle on safe facilities evidently exists to a certain extent. With regard to the variance of preferences for safe bicycle facilities among personal characteristics found by Garrard et al. (2008) and Heinen et al. (2010): this is found for bicycle lanes and safe facilities where older people showed a greater preference for these facilities. Besides, a relation with cycling experience and the preference for separated bicycle paths was found in this study. However, there are no distinct differences found between gender and the preference for specific bicycle facilities.

Multiple studies show a general tendency of cyclists to choose roads with speed limits below 30 km/h (Jestico et al., 2016; Verhoeven et al., 2018). These findings are confirmed by the results of this study (as presented in 5.4.3): a positive significant relation between lower speed limits and route choice probability is proven. The discrepancy between the percentage of roads with safe speed limits on the observed route and its alternatives is small: ±28% and ±26%, respectively. Besides, there is a large share of missing data in especially urban areas. Both of these factors provide an explanation for the fact that speed limits' singular effect on route choices is negative, suggesting that people are not tending to choose for roads with safe speed limits. However, as mentioned, in the choice model people showed opposite preferences due to its interactions with other safety factors.

In respect to personal characteristics, Misra and Watkins (2018) found that women and older people are less likely to cycle on roads with higher speed limits. The difference between sex cannot be underpinned by this study's results, but the relation between increasing ages and a preference for roads with speed limits up to 30 km/h is proven. The effect of motor vehicle speed limits is expected to vary among different types of bicycle facilities: the impact generally decreases when separated bicycle facilities are available (Winters et al., 2013). For separated bicycle paths this study indeed showed a negative relation to speed limits up to 30 km/h with regard to route choice, meaning that when separated facilities are present there is no clear preference for safe speed limits. While bicycle lanes and non-cycling facilities did have a positive significant relation, showing that in this case respondents were in fact more likely to choose routes with lower speed limits.

In earlier studies the presence of street lighting is mainly studied in relation to bicycle accidents, where less street lighting generally results in more accidents (Chen & Shen, 2016; Kim et al., 2007). As a safety factor the presence of street lighting is thus expected to have a positive impact on route choice probability. The route choice model confirms this: with every 1% extra street lighting on a route, the odds of choosing this route rises with 2%. This contradicts the finding by Dessing et al. (2016) who found that the observed routes of respondents contained less street lighting than alternatives, which they explained by the fact that their respondents preferred quiet roads. Busy roads are generally better lit. In this study, respondents stated to not avoid busy roads, which can be an explanation for the positive impact of street lighting (Osama & Sayed, 2017). The singular effect of street lighting, as well as its interaction with personal characteristics is however non-significant. This can be due to the fact that the majority of the routes was cycled in daylight in this study. Because although respondents did stated to prefer taking routes with proper street lighting when it is dark, they did not show this in their cycling behaviour. The majority of respondents stated to prefer street lighting was female. Street lighting as a safety factor is related to both traffic safety and social safety. Concerning social safety it was indeed expected that women would prefer to cycle on well-lit roads (Bohle & Verkehr, 2000), but the observed cycling behaviour did not indicate this. Further differences between traffic and social safety will be elaborated later on.

Crime rates were conceptualized as an indicator of social safety. The relation between crime rates and cycling route choice behaviour has not been studied frequently. However, it did show a negative impact on bicycle usage in areas where crime rates are high (Sun, Mobasheri, et al., 2017). Therefore, it was deemed to give similar results in bicycle behaviour. However, no significant relation between crime rates and route choices were found, which is in line with findings of a study by Hood et al. (2011). In the context of social safety gender was expected to play a role as well (Fietsersbond, 2019), however this was not the case in the present research. An explanation for the absence of an impact of crime rates may be that it is an objective measure for the feeling of safety. This feeling of personal security (Rietveld & Daniel, 2004) might be unrelated to the actual number of crimes in a certain neighbourhood. Therefore, street lighting and the type of environment are included to measure the impact of social safety in route choice behaviour.

From a social safety point of view it is expected that a natural environment can have a negative impact on route choice probability, the feeling of safety may decrease here because nature areas are usually more abandoned. Especially women and children were likely to choose not to cycle here in the dark. This results in the expectation that less abandoned urban areas are preferred in cycling behaviour and natural areas are not preferred (Bohle & Verkehr, 2000). Results from this present study indicate the opposite: an urban environment has a negative impact on route choices and a natural environment a positive impact. These results did not differ between day or night. For nature, there was however a significant positive interaction with the presence of street lighting, showing that the presence of street lighting would increase the odds of choosing for a route through nature. Gender did not play a very distinct role in this factor, except that women were actually a bit more likely to choose for a natural environment. Age did play a role, showing that older people have higher odds to choose to cycle through nature. However, this does not prove that children are more likely to cycle through less abandoned areas since the minimum age in the research group is 16. Stated preferences of respondents showed that a narrow majority did wish to avoid abandoned areas of which a majority was female.

Disregarding social safety, cycling through nature can actually be very attractive, which can explain the contradicting research results. When looking at the observed routes' descriptive statistics it becomes apparent that the majority of the routes (66%) is in fact cycled through an urban environment, which is somewhat contradicting.

To return to the differences in impact of traffic safety and social safety, the impact of social safety seems to be limited. Only street lighting resulted in the expected output: a positive impact on route choice probability. As mentioned, the type of environment showed contradicting results and crime rates had no significant impact. This is probably due to the fact that a feeling of social safety is hard to measure by means of spatial factors. But, also in stated preferences there was no clear indication that people take social safety into account in their route choices. The survey results did show that women consider social safety more important than men in route choices.

With regard to traffic safety, bicycle lanes, safe speed limits and street lighting show the expected significant relations. In the overall route choice model, traffic safety thus appeared to be more influential than social safety. This is in line with the average percentage of overlap of the observed route with the labelled alternatives: the traffic safe route shows 51% overlap and the social safe alternative 49%.

6.2 Research limitations

There are certain limitations with respect to the data and its collection, preparation and analysis that will be elaborated in this Section, associated with recommendations for further research.

Because of Covid-19 the data collection process had to be altered during our study (a more extensive exploration of Covid-19's implications will be given in Section 6.3). As a result the data was collected among our relatives. The fact that respondents are known to the research obstructs the effort to gather data among a representative research group with varying personal characteristics. The respondents were eager to help us by providing data, which may have resulted in cycling data that is not representative for their usual cycling behaviour. Respondents are tracked with the GPS tracker for a period of one week, which makes it possible that a respondent cycles the same route to a certain location multiple times a week, all of these trips are included in the research.

Two other data limitations that are associated with gathering data through relatives are related to time and space. Due to budgeting only 20 trackers were available during the research, which made it impossible to track all respondents at the same time. As a consequence, respondents are tracked over a period of approximately 4 weeks. This means that temporal variations, such as varying weather circumstances that can affect cycling frequency, are not taken into account. This can play a major role, because of the long data collection period. Furthermore, the collection of data among friends and family resulted in a research area that is rather spread out over the Netherlands. This means that the spatial varieties, such as differences in the network density are high.

One of the options was to make use of existing cycling trip data from the Fietsersbond, however, this would mean that additional data such as personal characteristics and stated preferences would not be available. Therefore, the choice was made to collect the data among relatives. For future research it would therefore be recommended to take a large sample in a smaller research area during the same period. It is also advised to take a random sample in order to retrieve a representative research group with varying personal characteristics and of all sorts of social classes.

Previous studies suggest that the use of GPS trackers is one of the more accurate methods to collect cycling data. However, there are some threats concerning the GPS-devices that were used to track the respondents, as there is still a chance of inaccurate data. During the data collection this became apparent when one of the trackers showed its signal at a certain location in the Netherlands, while it was knowingly on the other side of the country. This resulted in the loss of some data. Furthermore, it is impossible for the GPS points retrieved from the trackers to always give its exact location. During the map-matching process this inaccuracy sometimes proved to be problematic. In such a dense cycling network as in the Netherlands an inaccuracy of a few meters can already create discrepancies with the actual route that was taken. However, by matching it to the closest road section and using the assumption that people generally cycle on the right side of the road, the cycling trajectory is given as accurately as possible. Despite these drawbacks the use of GPS trackers currently remains the most precise method to measure and study cycling behaviour.

Safety data is mainly retrieved from the Fietsersbond dataset. The dataset is very extensive and contains many different factors that are relevant in research into cycling behaviour. Nevertheless, there are certain shortcomings for this dataset, the first one being that the dataset is created and updated by volunteers. As a result the network is not completely up to date and the shapefile's topology is not correct in all areas. With regard to the completeness of the dataset, there were some areas in this research where cyclists frequently cycled, which were also visual on Google StreetView, but not existent in the Fietsersbond shapefile. This made the map-matching process less accurate as well. Topology errors resulted in the faulty generation of alternative routes, which amongst others resulted in a substantial delay in this research. Later, these errors were fixed to improve the generation of alternative routes.

The voluntary character of the dataset also means that a lot of the attributed data is a valuation based on the volunteer's perception. For example street lighting consists of the categories 'not lit', 'partially lit' or 'well lit', this degree of the presence of street lighting can vary among people's perspective. Besides, the dataset includes many road sections where the volunteers attributed no extra data. As a consequence, there is a large share of missing data for certain factors. Especially for speed limits and the environment type there was a large share of missing data, which can affect the choice model outcomes. In this research the missing data is considered in the results of the study. For future research it would be recommended to complement the data, with other existing datasets or by collecting data in the field. Especially if a more concentrated research sample in a specific area is taken it would be possible to supplement the data with self-collected data about for example speed limits. Due to the spread out research area this was impossible in this study for the given time period.

Lastly, the choice set generation method used in this research has certain limitations. The choice was made to generate labelled alternatives for each observed trip. Many other methods exist as discussed in Section 3.2.1, which include algorithms that create a vast amount of alternative routes based on certain assumptions. In a route choice model the spatial factors that are tied to the observed (chosen) route are compared to the spatial factors tied to its corresponding alternatives. This means that in this study, the safety factors of the chosen route are analysed relative to a maximum of 7 alternatives. With more alternative routes it would have been possible to perform a more extensive route choice behaviour analysis. However, given the time for this thesis it was deemed impossible to create such an algorithm that would create alternative routes for each OD-pair. The choice was therefore made to make use of labelled alternatives, which served the objective to compare different types of safety and also showed to perform rather well in previous studies (Ben-Akiva et al., 1984; Prato, 2009).

6.3 Covid-19 implications

The Covid-19 pandemic had a large impact on this research' outcomes and its process. As previously mentioned, the data collection process had to be adapted. The initial plan was to take a random sample in the city of Delft, and to track all respondents at the same time for a period of a week. Stricter measurements of the government that were taken right before the start of the initial data collection made it impossible and irresponsible to collect data according to that plan. This also meant that budgeting for the research dropped, as a consequence there were thus only 25 GPS-devices at our disposal which resulted in the eventual data collection process. If it had been possible to pursue the initial plan, the limitations discussed in the previous Section could have been overcome.

Aside from the implications on the data collection process and research group, Covid-19 also had an impact on the cycling behaviour of respondents. From the survey it appeared that their commuting behaviour changed as well as their recreative cycling behaviour. Recreational cycling increased for 47% of the respondents, 22% cycled less for leisure and for the remaining group nothing changed. In contrast, commuting cycling increased for 5% of the respondents, 65% stated to cycle less to work or school and for 30% of the respondents nothing changed as a consequence of Covid-19 or the question was not applicable. This shows that the cycling behaviour of this research group definitely changed in this period, which may affect the research's representativeness or conflict comparisons with previous studies. Nonetheless, this research can provide valuable insights in cycling behaviour during a pandemic. Recently it was found that a rise in bicycle accidents occurred during the lockdown period in the Netherlands (CBS, 2021a). Insight into cycling behaviour during this period can therefore possibly be of extra value.

In Chapter 7 a conclusion on this research will be given by answering the sub-questions and the main research question.

7 Conclusion

To conclude this research, this Chapter will provide an answer to the main research question: '*How and to what extent does safety affect the route choice behaviour of cyclists?*', and to the sub-questions. The hypothesis was that safety will have implications for cycling route choices. In the next section the sub-questions will be answered. Subsequently, a response for the main question will be provided. To end, recommendations for future research and policy will be given.

7.1 Sub-questions

The first sub-question was: "What is the influence of objective safety on route choice behaviour?". This is answered by looking at the impact of the incidence of accidents on route choice behaviour. The number of accidents' individual impact, showed that cyclists were more likely to choose roads where more accidents occur. However in combination with other safety factors, accidents showed no significant effect on cyclist route choice behaviour. This is in line with previous studies, where people showed to be unaware of the actual number of accidents on a road (Sun, Du, et al., 2017; Sun, Mobasheri, et al., 2017). Therefore, it can be concluded that objective safety plays no role in cyclists' route choice behaviour.

The second sub-question: "How does subjective safety – ranging from traffic safety to social safety – affect route choice behaviour?", is addressed by considering both safety types' corresponding safety factors. Social safety consists of the factors: crime rates, the environment and street lighting. Crime rates had no significant impact on route choices made by respondents. Urban areas were not preferred to cycle through by respondents, whereas a natural environment was preferred. In the context of social safety, the results were expected to be contrary (Bohle & Verkehr, 2000). Street lighting did have a positive impact, which means that respondents preferred to cycle on roads with street lighting. However, the majority of the routes was cycled in daylight, limiting the conclusion. The overall impact of social safety on route choices in this research is thus limited.

Traffic safety includes the influence of bicycle facilities, speed limits and street lighting. Regarding bicycle facilities the impact varies among the type of facility. Separated bicycle paths were generally not preferred by cyclists, which can be explained by the dense urban network in which the majority of the routes is cycled (Ton et al., 2017). The presence of bicycle lanes did however have a positive effect on route choice probability. Combining both factors into one safe facility-factor showed that there was a general tendency among bicyclists to choose to cycle on safe facilities. The bicycle behaviour of respondents also showed that they preferred to cycle on safer roads with speed limits (up to 30 kilometres per hour). In interaction with the presence of separated bicycle paths there was, however, no preference for roads with lower speed limits. In the general route choice model street lighting thus also had a significant positive impact.

Comparing both types of safety, traffic safety thus has a greater influence than social safety on route choices of cyclists. This is confirmed by the average percentage of overlap of the observed routes with the labelled alternatives: 51% of overlap for the traffic safe routes compared to 49% of overlap for the social safe routes.

The last sub-question was: "Which personal characteristics affect the extent to which cyclists find safety factors important in their route choices?". In relation to safety, the impact of gender, age and cycling experience were studied. Age turns out to be the most influential character in this study, in the way that the older people, the more likely they were to choose for routes with more bicycle lanes, safe bicycle facilities and safe speed limits. However, also for roads with more accidents and a natural environment, elderly were more likely to choose these. There were no substantial differences between men and women on the extent to which they find certain safety factors important. Men preferred to cycle on bicycle lanes, and women showed to prefer to cycle through nature. This indicates that women actually showed to find safety less important than men, which is in contrast with previous studies (Bohle & Verkehr, 2000; Heinen et al., 2010). Cycling experience only had a significant impact on the preference for separated bicycle paths in route choices: respondents who cycled less frequently showed a higher preference for separated bicycle facilities than respondents who cycled more often.

7.2 Main research question

To conclude, the main research question will be answered. Various safety factors have been addressed. Objective safety in terms of accidents had no substantial impact on route choices. Traffic safety is the most influential factor in route choice behaviour of cyclists in this research: the presence of bicycle lanes, safe speed limits and street lighting all fostered the route choice probability of cyclists. The presence of bicycle lanes had the largest impact on the probability of a route being chosen, for safe speed limits and street lighting the impact was slightly smaller. The impact of social safety on route choice behaviour was limited. Regarding social safety only street lighting thus showed to have an effect. A socially safer environment actually had a negative impact on route choice behaviour, showing that cyclists did not consider the environment as an important factor in their safety considerations.

Cyclists thus show to be susceptible for subjective safety aspects in their route choices, where the impact of traffic safety is larger than the impact of social safety. Besides, the importance of certain safety aspects in route choice behaviour varies among personal characteristics.

7.3 Research and policy recommendations

For future research it would be recommended to analyse cycling behaviour based on GPS data that is retrieved from a larger and random research sample. Besides, it would be beneficial to concentrate the data collection in a smaller research area to be able to draw conclusions based on cycling data that is cycled in the same type of network. A smaller research area would also enable possibilities to supplement spatial data where data is missing. This would also make it possible to research the impact of safety factors of which no (extensive) datasets were available: traffic volume, and intersections and traffic lights.

Moreover, it would be interesting to create alternative routes with the use of another choice set generation method, and to analyse route choice behaviour based on these alternatives. Analysing the impact of safety with the use of a different route choice model and comparing the results could also be interesting to examine the influence of the type of route choice model one uses.

Furthermore, the study showed that urbanity interacted with, for example, the impact of safe cycling facilities. Therefore, it would be worthwhile to further explore the differences between an urban and natural/rural environment with regard to the impact of safety on cycling behaviour.

Lastly, it could be interesting to analyse the impact of Covid-19 on cycling behaviour by comparing this study's findings with a post-lockdown research into cycling behaviour. The restrictions due to the lockdown caused people to be less likely to go outside in the evening, as a consequence most of the routes were cycled in daylight. This limits the resulting impact of street lighting on cycling behaviour, as well as the impact of other social safety factors. Therefore, it would be interesting to further explore the impact of social safety on cycling behaviour in a post-Covid-19 period.

The preference of cyclists to cycle on roads where traffic safety is higher, indicates a niche for policymakers: improvements in road safety with regard to bicycle facilities, street lighting and low speed limits can create safer circumstances for cyclists, which can incite more people to cycle. For speed limits for example, this means that it would be desirable to lower speed limits below 30 kilometres per hour in busy urban areas. Besides, the interaction of bicycle facilities, speed limits and urbanity creates opportunities for policymakers. The interaction of separated bicycle paths and speed limits (where respondents showed no preference for lower speed limits in the case of separated facilities) shows that the construction of separated bicycle facilities can decrease the impact of higher speed limits on cyclist route choice behaviour. Cyclists did state to prefer to cycle on paths separated from the main road, but their behaviour shows otherwise. This may indicate that the availability of separated bicycle facilities is limited. The limited amount of separated facilities became specifically apparent for urban areas in this research. This indicates that the construction of these types of facilities can improve safe cycling behaviour, and therewith a safe bicycle environment. Besides, the impact of age showed that elderly were more susceptible for various safety factors in cycling. Therefore, it is important to develop targeted campaigns and policies that specifically focus on older cyclists.

However, preferences to safe routes were in general found to be limited, which means that it will not be sufficient to solely focus on building a safe bicycle infrastructure in policymaking. Additional campaigns to raise awareness among all road-users to consider safety will be required to create a more safe environment for cyclists.

References

Appleyard, B. S., & Ferrell, C. E. (2017). The Influence of crime on active & sustainable travel: New geo-statistical methods and theories for understanding crime and mode choice. *Journal of Transport and Health*, *6*, 516–529. https://doi.org/10.1016/j.jth.2017.04.002

ArcGIS Pro. (n.d.). Network topology — Documentation. Retrieved May 26, 2021, from https://pro.arcgis.com/en/pro-app/latest/help/data/utility-network/about-network-topology.htm

Bekhor, S., Ben-Akiva, M. E., & Ramming, M. S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, *144*(1), 235–247. https://doi.org/10.1007/s10479-006-0009-8

Ben-Akiva, M. E., Bergman, M. J., Daly, A. J., & Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour. In *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=nc1GCPRsH-4C&oi=fnd&pg=PA299&dq=ben-akiva+1984+labeling+approach&ots=pW5gYb5Vq-&sig=CpX9NdOlDRwFI1mcNf3K70n2oi8&redir_esc=y#v=onepage&q=ben-akiva 1984 labeling approach&f=false

Ben-Akiva, M. E., & Bierlaire, M. (1999). *Discrete Choice Methods and their Applications to Short Term Travel Decisions*. https://doi.org/10.1007/978-1-4615-5203-1_2

Bernardi, S., Geurs, K., & Puello, L. L. P. (2018). Modelling route choice of dutch cyclists using smartphone data. *Journal of Transport and Land Use*, *11*(1), 883–900. https://doi.org/10.5198/jtlu.2018.1143

Blake, L. (2010). *Spatial Relationships in GIS – Geospatial Topology Basics*. *2*(September), 1–7. Retrieved from http://webhelp.esri.com/arcgisserver/9.3/java/index.htm#geodatabases/topology_in_arcgis.htm

Bohle, W., & Verkehr, P. (2000). Attractiveness of bicycle-facilities for the users and evaluation of measures for the cycle-traffic. *Publication of: Swedish National Road and Transport Research Institute*, 89–94. Retrieved from http://www.velomondial.net/velomondiall2000/PDF/BOHLE.PDF

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, *17*(3), 285–297. https://doi.org/10.1016/j.trc.2008.11.004

Bovy, P. H. L. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews*, *29*(1), 43–68. https://doi.org/10.1080/01441640802078673

Bovy, P. H. L., & Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: Behavioral and probabilistic foundations. *Transportmetrica*, *3*(3), 173–189. https://doi.org/10.1080/18128600708685672

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005

Buehler, R., & Dill, J. (2016). Bikeway Networks: A Review of Effects on Cycling. *Transport Reviews*, *36*(1), 9–27. https://doi.org/10.1080/01441647.2015.1069908

Casello, J. M., & Usyukov, V. (2014). Modeling Cyclists' Route Choice Based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2430*(1), 155–161. https://doi.org/10.3141/2430-16

Caviedes, A., & Figliozzi, M. (2018). Modeling the impact of traffic conditions and bicycle facilities on cyclists' on-road stress levels. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 488–499. https://doi.org/10.1016/j.trf.2018.06.032

CBS. (2020a). CBS: Werkelijk aantal verkeersdoden - Trend. Retrieved September 30, 2020, from https://theseus.swov.nl/single/?appid=5dbac35a-5fbd-401f-b711-682176941688&sheet=ZjpemJ&opt=currsel%2Cctxmenu

CBS. (2020b). Wijk- en buurtkaart 2020. Retrieved March 29, 2021, from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2020

CBS. (2021a). 610 verkeersdoden in 2020. Retrieved May 26, 2021, from https://www.cbs.nl/nl-nl/nieuws/2021/15/610-verkeersdoden-in-2020

CBS. (2021b). Misdrijven per wijk en buurt per jaar. Retrieved March 29, 2021, from https://politieopendata.cbs.nl/portal.html?_la=nl&_catalog=Politie&graphtype=Table&tableId=47018NED&_theme=87

Chen, P., & Shen, Q. (2016). Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accident Analysis and Prevention*, *86*, 239–246. https://doi.org/10.1016/j.aap.2015.11.002

Dane, G., Feng, T., Luub, F., & Arentze, T. (2020). Route choice decisions of E-bike users: Analysis of GPS tracking data in the Netherlands. *Lecture Notes in Geoinformation and Cartography*, 109–124. https://doi.org/10.1007/978-3-030-14745-7_7

Dessing, D., de Vries, S. I., Hegeman, G., Verhagen, E., van Mechelen, W., & Pierik, F. H. (2016). Children's route choice during active transportation to school: Difference between shortest and actual route. *International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 1–11. https://doi.org/10.1186/s12966-016-0373-y

Dhakar, N. S., & Srinivasan, S. (2014). Route Choice Modeling Using GPS-Based Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board*, *2413*(1), 65–73. https://doi.org/10.3141/2413-07

Fietsersbond. (n.d.). Metagegevens database fietsrouteplanner + fietsknooppunten + POI s - PDF Gratis download. Retrieved October 12, 2020, from https://docplayer.nl/7295641-Metagegevens-database-fietsrouteplanner-fietsknooppunten-poi-s.html

Fietsersbond. (2019). *Fiets en Sociale Veiligheid*. Retrieved from https://files.fietsersbond.nl/app/uploads/sites/26/2019/04/16210710/Fiets-en-Sociale-Veiligheid.pdf

Fietsersbond. (2020). De Meefietslijn - Fietsersbond. Retrieved May 2, 2021, from https://www.fietsersbond.nl/nieuws/de-meefietslijn/

Fietsersbond. (2021). *Schokkende toename in fietsdoden vereist onmiddellijke actie - Fietsersbond*. Retrieved from https://www.fietsersbond.nl/nieuws/schokkende-toename-in-fietsdoden-vereist-onmiddellijke-actie/

Garrard, J., Rose, G., & Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine*, *46*(1), 55–59. https://doi.org/10.1016/j.ypmed.2007.07.010

Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K. W., Nielsen, O. A., & Prato, C. G. (2014). Efficiency of choice set generation methods for bicycle routes. *EJTIR Issue*, *14*(4), 332–348. Retrieved from https://journals.open.tudelft.nl/ejtir/article/view/3040

Han, C., Huang, H., Lee, J., & Wang, J. (2018). Investigating varying effect of road-level factors on crash frequency across regions: A Bayesian hierarchical random parameter modeling approach. *Analytic Methods in Accident Research*, *20*, 81–91. https://doi.org/10.1016/j.amar.2018.10.002

Handy, S., van Wee, B., & Kroesen, M. (2014). Promoting Cycling for Transport: Research Needs and Challenges. *Transport Reviews*, *34*(1), 4–24. https://doi.org/10.1080/01441647.2013.860204

Harvey, F., Krizek, K. J., & Collins, R. (2008). *Using GPS Data to Assess Bicycle Commuter Route Choice*.

Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by Bicycle: An Overview of the Literature. *Transport Reviews*, *30*(1), 59–96. https://doi.org/10.1080/01441640903187001

Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters*, *3*(1), 63–75. https://doi.org/10.3328/TL.2011.03.01.63-75

Hoogendoorn-Lanser, S., van Nes, R., & Bovy, P. (2005). Path Size Modeling in Multimodal Route Choice Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *1921*(1), 27–34. https://doi.org/10.1177/0361198105192100104

Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, *52*, 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006

Kaplan, S., & Prato, C. G. (2015). A Spatial Analysis of Land Use and Network Effects on Frequency and Severity of Cyclist–Motorist Crashes in the Copenhagen Region. *Traffic Injury Prevention*, *16*(7), 724–731. https://doi.org/10.1080/15389588.2014.1003818

Kim, J. K., Kim, S., Ulfarsson, G. F., & Porrello, L. A. (2007). Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis and Prevention*, *39*(2), 238–251. https://doi.org/10.1016/j.aap.2006.07.002

Kondo, M. C., Morrison, C., Guerra, E., Kaufman, E. J., & Wiebe, D. J. (2018). Where do bike lanes work best? A Bayesian spatial model of bicycle lanes and bicycle crashes. *Safety Science*, *103*, 225–233. https://doi.org/10.1016/j.ssci.2017.12.002

Lawrence, B. M., & Oxley, J. A. (2019). You say one route, we observe four: Using naturalistic observation to understand route-choices in cyclists. *Safety Science*, *119*, 207–213. https://doi.org/10.1016/j.ssci.2019.01.004

Lawson, A. R., Pakrashi, V., Ghosh, B., & Szeto, W. Y. (2013). Perception of safety of cyclists in Dublin City. *Accident Analysis and Prevention*, *50*, 499–511. https://doi.org/10.1016/j.aap.2012.05.029

Lu, W., Scott, D. M., & Dalumpines, R. (2018). Understanding bike share cyclist route choice using GPS data: Comparing dominant routes and shortest paths. *Journal of Transport Geography*, *71*, 172–181. https://doi.org/10.1016/j.jtrangeo.2018.07.012

Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice*, *44*(9), 754–765. https://doi.org/10.1016/j.tra.2010.07.008

Ministerie van Infrastructuur en Waterstaat. (2018). *Landelijk Actieplan Verkeersveiligheid 2019-2021: Veilig van deur tot deur*. Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2018/12/05/bijlage-2-landelijk-actieplan-verkeersveiligheid-2019-2021

Ministerie van Infrastructuur en Waterstaat. (2019). *Mobiliteitsbeeld 2019*. 204. Retrieved from https://www.kimnet.nl/publicaties/rapporten/2019/11/12/mobiliteitsbeeld-2019-vooral-het-gebruik-van-de-trein-neemt-toe

Misra, A., & Watkins, K. (2018). Modeling Cyclist Route Choice using Revealed Preference Data: An Age and Gender Perspective. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(3), 145–154. https://doi.org/10.1177/0361198118798968

OpenStreetMap Contributors. (2020). Planet dump retrieved from https://planet.osm.org. Retrieved May 19, 2021, from https://planet.openstreetmap.org/

Osama, A., & Sayed, T. (2017). Evaluating the Impact of Socioeconomics, Land Use, Built Environment, and Road Facility on Cyclist Safety. *Transportation Research Record: Journal of the Transportation Research Board*, *2659*(1), 33–42. https://doi.org/10.3141/2659-04

Parkin, J., Wardman, M., & Page, M. (2007). Models of perceived cycling risk and route acceptability. *Accident Analysis and Prevention*, *39*(2), 364–371. https://doi.org/10.1016/j.aap.2006.08.007

Prato, C. G. (2009). Route choice modeling: Past, present and future research directions. *Journal of Choice Modelling*, *2*(1), 65–100. https://doi.org/10.1016/S1755-5345(13)70005-8

Prato, C. G. (2012). Meta-analysis of choice set generation effects on route choice model estimates and predictions. *Transport*, *27*(3), 286–298. https://doi.org/10.3846/16484142.2012.719840

Rewa, K. C. (2012). *An Analysis of Stated and Revealed Preference Cycling Behaviour : A Case Study of the Regional Municipality of Waterloo*. Retrieved from https://uwspace.uwaterloo.ca/handle/10012/6910

Rietveld, P., & Daniel, V. (2004). Determinants of bicycle use: Do municipal policies matter? *Transportation Research Part A: Policy and Practice*, *38*(7), 531–550. https://doi.org/10.1016/j.tra.2004.05.003

Rijkswaterstaat CIV. (2019). *Bestand geRegistreerde Ongevallen Nederland (BRON)*.

Sarjala, S. (2019). Built environment determinants of pedestrians' and bicyclists' route choices on commute trips: Applying a new grid-based method for measuring the built environment along the route. *Journal of Transport Geography*, *78*, 56–69. https://doi.org/10.1016/j.jtrangeo.2019.05.004

Scheider, S. (2018). GitHub - simonscheider/mapmatching: An arcpy based HMM map matching tool. Retrieved October 13, 2020, from https://github.com/simonscheider/mapmatching

Schepers, P., Twisk, D., Fishman, E., Fyhri, A., & Jensen, A. (2017). The Dutch road to a high level of cycling safety. *Safety Science*, *92*, 264–273. https://doi.org/10.1016/j.ssci.2015.06.005

Schuessler, N., & Axhausen, K. W. (2009). Processing Raw Data from Global Positioning Systems without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, *2105*(1), 28–36. https://doi.org/10.3141/2105-04

Segadilha, A. B. P., & Sanches, S. da P. (2014). Identification of Factors that Influence Cyclistś Route Choice. *Procedia - Social and Behavioral Sciences*, *160*(Cit), 372–380. https://doi.org/10.1016/j.sbspro.2014.12.149

Sener, I. N., Eluru, N., & Bhat, C. R. (2009a). An analysis of bicycle route choice preferences in Texas, US. *Transportation*, *36*(5), 511–539. https://doi.org/10.1007/s11116-009-9201-4

Sener, I. N., Eluru, N., & Bhat, C. R. (2009b). An Analysis of Bicycle Route Choice Preferences Using a Web-Based Survey to Examine Bicycle Facilities. *Transportation*, *36*(5), 511–539.

Sobhani, A., Aliabadi, H. A., & Farooq, B. (2019). Metropolis-Hasting based Expanded Path Size Logit model for cyclists' route choice using GPS data. *International Journal of Transportation Science and Technology*, *8*(2), 161–175. https://doi.org/10.1016/j.ijtst.2018.11.002

Song, R., Ni, Y., & Li, K. (2017). Understanding cyclists' risky route choice behavior on urban road sections. *Transportation Research Procedia*, *25*, 4157–4170. https://doi.org/10.1016/j.trpro.2017.05.356

Stata. (n.d.). *Stata guide - cmxtmixlogit*. Retrieved from https://www.stata.com/manuals/cmcmxtmixlogit.pdf

Steer Davies Gleave. (2012). *Cycle route choice - Final survey and model report*. *44*(June). Retrieved from http://content.tfl.gov.uk/understanding-cycle-route-choice.pdf

Stone, M., & Broughton, J. (2003). Getting off your bike: Cycling accidents in Great Britain in 1990-1999. *Accident Analysis and Prevention*, *35*(4), 549–556. https://doi.org/10.1016/S0001-4575(02)00032-5

Sun, Y., Du, Y., Wang, Y., & Zhuang, L. (2017). Examining Associations of Environmental Characteristics with Recreational Cycling Behaviour by Street-Level Strava Data. *International Journal of Environmental Research and Public Health*, *14*(6), 644. https://doi.org/10.3390/ijerph14060644

Sun, Y., Mobasheri, A., Hu, X., & Wang, W. (2017). Investigating Impacts of Environmental Factors on the Cycling Behavior of Bicycle-Sharing Users. *Sustainability*, *9*(6), 1060. https://doi.org/10.3390/su9061060

SWOV. (2018). *Monitor Verkeersveiligheid 2018 - Achtergrondinformatie en onderzoeksverantwoording*.

SWOV. (2019). Ernstig verkeersgewonden in Nederland | SWOV - factsheet. Retrieved September 30, 2020, from https://www.swov.nl/feiten-cijfers/factsheet/ernstig-verkeersgewonden-nederland

Ton, D., Cats, O., Duives, D., & Hoogendoorn, S. (2017). How Do People Cycle in Amsterdam, Netherlands?: Estimating Cyclists' Route Choice Determinants with GPS Data from an Urban Area. *Transportation Research Record: Journal of the Transportation Research Board*, *2662*(1), 75–82. https://doi.org/10.3141/2662-09

Ton, D., Duives, D., Cats, O., & Hoogendoorn, S. (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society*, *13*, 105–117. https://doi.org/10.1016/j.tbs.2018.07.001

Verhoeven, H., Van Hecke, L., Van Dyck, D., Baert, T., Van de Weghe, N., Clarys, P., … Van Cauwenberg, J. (2018). Differences in physical environmental characteristics between adolescents' actual and shortest cycling routes: A study using a Google Street View-based audit. *International Journal of Health Geographics*, *17*(1), 1–15. https://doi.org/10.1186/s12942-018-0136-x

Weijermars, W., & Dijkstra, A. (2008). *Verkeersveiligheid van routes en van routekeuze - Indicatoren om de veiligheid van routes te beschrijven*. (november). Retrieved from http://projectwaalbrug.pbworks.com/f/cvs08_19.pdf

Winters, M., Brauer, M., Setton, E. M., & Teschke, K. (2013). Mapping bikeability: A spatial tool to support sustainable travel. *Environment and Planning B: Planning and Design*, *40*(5), 865–883. https://doi.org/10.1068/b38185

Winters, M., & Teschke, K. (2010). Route preferences among adults in the near market for bicycling: Findings of the cycling in cities study. *American Journal of Health Promotion*, *25*(1), 40–47. https://doi.org/10.4278/ajhp.081006-QUAN-236

Yang, C., & Mesbah, M. (2013). Route choice behaviour of cyclists by stated preference and revealed preference. *Australasian Transport Research Forum, ATRF 2013 - Proceedings*, (October). Retrieved from https://www.australasiantransportresearchforum.org.au/sites/default/files/2013_yang_mesbah.pdf

Zahabi, S. A. H., Strauss, J., Manaugh, K., & Miranda-Moreno, L. F. (2011). Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2247*(1), 81–90. https://doi.org/10.3141/2247-10

Zimmermann, M., Mai, T., & Frejinger, E. (2017). Bike route choice modeling using GPS data without choice sets of paths. *Transportation Research Part C: Emerging Technologies*, *75*, 183–196. https://doi.org/10.1016/j.trc.2016.12.009

Zwerts, E., Allaert, G., Janssens, D., Wets, G., & Witlox, F. (2010). How children view their travel behaviour: A case study from Flanders (Belgium). *Journal of Transport Geography*, *18*(6), 702–710. https://doi.org/10.1016/j.jtrangeo.2009.10.002

Appendices

A. Respondent letter

Geachte heer of mevrouw,

Allereerst hartelijk dank voor het meedoen aan ons onderzoek.

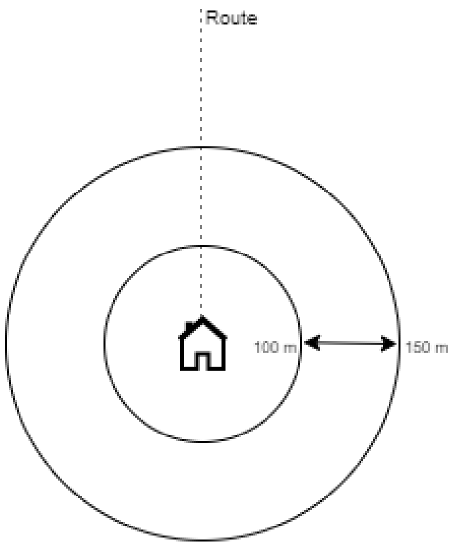
In deze brief vindt u een handleiding voor het gebruik van de GPS-trackers.

Wat we aan u vragen is het volgende:

- De GPS-tracker een week lang bij u te dragen; het maakt niet uit of u hem in uw zak of tas stopt.
- De GPS-tracker iedere dag op te laden; hiervoor ontvangt u een oplaadkabel.
- Achteraf de naar u gestuurde enquête in te vullen.

We benadrukken dat uw privacy goed gewaarborgd wordt:

- Uw data wordt automatisch verstuurd naar een centrale server waartoe alleen onze begeleider toegang tot heeft.
- Onze begeleider anonimiseert de GPS-ritten door de routes op ongeveer 100 tot 150 meter van uw huisadres te laten stoppen. Daardoor is het onmogelijk om uw huisadres te traceren. We weten alleen dat het ergens op 100 tot 150 meter afstand van het eind van de route ligt. Dit is zichtbaar in de afbeelding. Voor ons lijkt het alsof uw huis ergens in de buitenste ring staat terwijl uw huis daadwerkelijk in het midden van de cirkel staat.
- Wij hebben geen toegang tot de originele data; die wordt door onze begeleider verwijderd.
- In lijn met richtlijnen van het rijk, heeft onze begeleider een privacyverklaring getekend. Onderzoekers hebben geen enkel belang bij individuele data, maar kijken alleen naar patronen van de hele groep.
- De data wordt niet doorgegeven aan derden.



Bij vragen kunt u altijd contact opnemen met de persoon van wie u de tracker heeft ontvangen.

- Harmke Vliek:
  g.h.vliek@students.uu.nl
  0611689821
- Jimme Smit:
  j.smit6@students.uu.nl
  0616374277
- Maaike Kuiper:
  m.d.kuiper2@students.uu.nl
  0623407886

Bij voorbaat dank en veel fietsplezier!

B. Survey

Geachte heer of mevrouw,

Wij zijn studenten aan de Universiteit Utrecht en de TU Delft. Wij doen onderzoek naar fietsroutekeuze. Hiervoor vragen wij u om uw fietsverplaatsingen bij te houden met een zogeheten GPS-tracker. Ook vragen wij u of u deze enquête wilt invullen.

In deze enquête stellen wij u een aantal vragen die ons helpen om uw fietsverplaatsingen te begrijpen. De vragen gaan over de frequentie van uw fietsgedrag en uw voorkeuren met betrekking tot de omgeving waarin u fietst tijdens uw route. Tot slot zullen er nog wat persoonskenmerken gevraagd worden. Uw gegevens worden vertrouwelijk behandeld en volledig anoniem verwerkt, zoals we beschreven hebben in de brief. Het invullen van de enquête vraagt ongeveer 15 minuten van uw tijd.

Alvast hartelijk bedankt voor uw medewerking!

---

1. GPS tracker identificatienummer (vermeld op de achterzijde van de tracker, wanneer u deze uit de hoes haalt):

De volgende vragen gaan over uw fietsgewoonten voordat de corona crisis uitbrak.

2. Voor welke doelen gebruikt u een fiets (voor de corona crisis)? (meerdere antwoorden mogelijk)
   - ☐ Woon-werkrit
   - ☐ Zakelijk: ritten tijdens werk
   - ☐ School of studie
   - ☐ Voorzieningen bezoeken (winkels, supermarkten en dergelijke)
   - ☐ Recreatief
   - ☐ Sport
   - ☐ Vrienden of familie bezoeken
   - ☐ Overig

3. Wat voor soort fiets(en) heeft u?
   - ☐ Stadsfiets zonder versnellingen
   - ☐ Stadsfiets met versnellingen
   - ☐ Sportieve fiets (hybride)
   - ☐ Racefiets
   - ☐ Mountainbike
   - ☐ Elektrische fiets (max. 25 km/uur)
   - ☐ Speed pedelec (max. 45 km/uur)
   - ☐ Ligfiets
   - ☐ Vouwfiets
   - ☐ Tandem
   - ☐ OV-fiets
   - ☐ Anders, namelijk

4. Indien u meerdere fietsen heeft, voor welk doel gebruikt u elke fiets? Selecteer de doelen waarvoor u fietst en selecteer het soort fiets dat u daarvoor gebruikt

	Woon-werkrit	Zakelijk	School of studie	Voorzieningen bezoeken	Recreatief	Sport	Vrienden of familie bezoeken	Overig
Stadsfiets zonder versnellingen	○	○	○	○	○	○	○	○
Stadsfiets met versnellingen	○	○	○	○	○	○	○	○
Sportieve fiets (hybride)	○	○	○	○	○	○	○	○
Racefiets	○	○	○	○	○	○	○	○
Mountain bike	○	○	○	○	○	○	○	○
Elektrische fiets (max. 25km/uur)	○	○	○	○	○	○	○	○
Speed pedelec (max. 45km/uur)	○	○	○	○	○	○	○	○
Ligfiets	○	○	○	○	○	○	○	○
Vouwfiets	○	○	○	○	○	○	○	○
Tandem	○	○	○	○	○	○	○	○
Ov-fiets	○	○	○	○	○	○	○	○
Overig	○	○	○	○	○	○	○	○

5. Wat is de onderhoudsstaat van uw meest gebruikte fiets?
- ☐ Uitstekend onderhouden
- ☐ Goed onderhouden
- ☐ Gemiddeld
- ☐ Minder goed onderhouden
- ☐ Niet onderhouden

6. Op welke leeftijd bent u zelf gaan fietsen in de openbare ruimte (Vul alleen het getal in)?

7. Hoe vaak fietst u?
- ☐ Dagelijks
- ☐ Ongeveer ____ dagen per week
- ☐ Ongeveer ____ dagen per maand
- ☐ Zelden
- ☐ Nooit

8. Welke afstand fietst u ongeveer voor de volgende bestemmingen (indien van toepassing)?
- ☐ Werk: … km
- ☐ School/studie: … km
- ☐ Boodschappen: … km
- ☐ Naar openbaar Vervoer: … km

9. Geef uw voorkeuren aan in de volgende uitspraken

Kruis aan in hoeverre u het eens bent (bedenk dat het anoniem is)

	Helemaal oneens		Ne u- tra al		Helemaal eens
9.1 Ik fiets graag langs het *water*, ook als dit niet de kortste route is.	○	○	○	○	○
9.2 Ik fiets graag door het *park*, ook als dit niet de kortste route is.	○	○	○	○	○
9.3 Ik fiets graag door het *bos*, ook als dit niet de kortste route is.	○	○	○	○	○
9.4 Fietsen door een mooie omgeving is voor mij belangrijk, ook als dit niet de kortste route is.	○	○	○	○	○
9.5 Ik fiets graag via een levendige route, ook als dit niet de kortste route is.	○	○	○	○	○
9.6 Ik fiets graag langs herkenningspunten zoals kunstwerken of gebouwen door een interessante stedelijke omgeving.	○	○	○	○	○
9.7 Met slecht weer zoek ik een beschutte route.	○	○	○	○	○
9.8 Ik vermijd het liefst een lawaaiige omgeving, bijvoorbeeld van verkeer.	○	○	○	○	○

	Helemaal oneens		Ne u- tra al		Helemaal eens
9.9 Als het schemert of donker is, neem ik graag een route met goede straatverlichting.	○	○	○	○	○
9.10 Ik vermijd verkeerslichten zo veel mogelijk.	○	○	○	○	○
9.11 Ik vermijd drukke kruispunten zo veel mogelijk.	○	○	○	○	○
9.12 Ik fiets liever op fietspaden die gescheiden zijn van de weg.	○	○	○	○	○
9.13 Ik vermijd verkeersdrukte.	○	○	○	○	○
9.14 Ik fiets liever op wegen waar de maximale snelheid voor auto's 30 km/u is.	○	○	○	○	○
9.15 Ik vermijd onveilige routes, ook als dit de kortste route is.	○	○	○	○	○
9.16 Ik houd rekening met sociale veiligheid in mijn routekeuze.	○	○	○	○	○
9.17 Ik houdt rekening met kans op diefstal als ik mijn fiets parkeer	○	○	○	○	○
9.18 Ik vermijd verlaten gebieden, ook als dit de kortste route is.	○	○	○	○	○
9.19 Ik vermijd wegwerkzaamheden.	○	○	○	○	○

	Helemaal oneens		Ne u- tra al		Helemaal eens
9.20 Als ik fiets dan probeer ik de kortste route te nemen.	○	○	○	○	○
9.21 Als ik fiets dan probeer ik de reistijd zo kort mogelijk te houden.	○	○	○	○	○
9.22 Als ik fiets dan probeer ik zo veel mogelijk te fietsen zonder te hoeven stoppen (verkeerslichten bijvoorbeeld).	○	○	○	○	○
9.23 Als ik fiets dan probeer ik een hoog tempo aan te houden.	○	○	○	○	○
9.24 Als ik fiets neem ik wel eens doorsteekjes om een deel van de route af te snijden.	○	○	○	○	○
9.25 Ik fiets nooit door een rood stoplicht.	○	○	○	○	○

	Helemaal oneens		Ne u- tra al		Helemaal eens
9.26 Ik vind fietsen ontspannend.	○	○	○	○	○
9.27 Ik zie mijzelf als iemand met een goede conditie.	○	○	○	○	○
9.28 Ik vind een e-bike duur.	○	○	○	○	○
9.29 Ik vind een e-bike het geld waard.	○	○	○	○	○

De volgende vragen gaan over de mogelijke verandering van uw fietsgedrag met betrekking tot de corona crisis.

10. Werkt u thuis vanwege de corona crisis?
- ☐ Ja
- ☐ Nee
- ☐ Gedeeltelijk
- ☐ Niet van toepassing

11. Voor welke doelen gebruikt u de fiets (voor de corona crisis)? *(meerdere antwoorden mogelijk)*
- ☐ Woon-werkrit
- ☐ Zakelijk: ritten tijdens het werk
- ☐ School of studie
- ☐ Voorzieningen bezoeken (winkels, supermarkten en dergelijke)
- ☐ Recreatief
- ☐ Sport
- ☐ Vrienden of familie bezoeken
- ☐ Overig

12. Is uw fietsgedrag met betrekking tot **werk of school** veranderd sinds de coronacrisis?
- ☐ Ja, ik fiets meer
- ☐ Ja, ik fiets minder
- ☐ Nee
- ☐ Niet van toepassing

13. Is uw **recreatieve** fietsgedrag veranderd sinds de coronacrisis?
- ☐ Ja, ik fiets meer
- ☐ Ja, ik fiets minder
- ☐ Nee
- ☐ Niet van toepassing

14. Vermijdt u sinds de coronacrisis drukke fietsroutes?
- ☐ Ja
- ☐ Nee
- ☐ Af en toe
- ☐ Niet van toepassing

Ten slotte vragen we u enkele persoonlijke gegevens. Uw gegevens worden anoniem verwerkt.

15. Wat is uw geboortejaar?

16. Wat is uw geslacht?
- ☐ Vrouw
- ☐ Man
- ☐ Anders / wil ik niet zeggen

17. Wat is uw hoogste opleiding?
- ☐ Lbo, mulo, mavo, vmbo of gelijkwaardig
- ☐ Havo, vwo, mms, hbs, mbo of gelijkwaardig
- ☐ Hbo of universiteit
- ☐ Anders

18. Wat is de samenstelling van uw huishouden?
- ☐ Alleenstaand *zonder* thuiswonende kinderen
- ☐ Alleenstaand *met* thuiswonende kinderen op de basisschoolleeftijd of jonger
- ☐ Alleenstaand *met* thuiswonende kinderen op de middelbare schoolleeftijd of ouder
- ☐ Samenwonend/gehuwd *zonder* thuiswonende kinderen
- ☐ Samenwonend/gehuwd *met* thuiswonende kinderen op de basisschoolleeftijd of jonger
- ☐ Samenwonend/gehuwd *met* thuiswonende kinderen op de middelbare schoolleeftijd of ouder
- ☐ Samenwonend met andere volwassenen *(zoals studentenhuis, zorgcentrum of woongroep)*
- ☐ Thuiswonend bij ouder(s) of pleegouder(s)

19. Wat is uw (belangrijkste) dagelijkse bezigheid?
- ☐ Betaald werk, voltijd
- ☐ Betaald werk, deeltijd (minder dan 36 uur/week)
- ☐ School/studie
- ☐ Geen betaald werk, gepensioneerd, vrijwilligerswerk, overig

20. Wat is het netto inkomen van uw huishouden per maand? (als u samenwonend/gehuwd bent, beide inkomens tezamen).
- ☐ Minder dan € 2000
- ☐ Tussen € 2000 en € 4000
- ☐ Tussen € 4000 en € 6000
- ☐ Meer dan € 6000
- ☐ Dat weet ik niet / dat wil ik niet zeggen

21. In welk type woning woont u?
- ☐ Vrijstaand
- ☐ Twee onder een kap
- ☐ Rijwoning
- ☐ Boven- of benedenwoning
- ☐ Portiekwoning, flat, appartement
- ☐ Anders

22. Beschikt u over een auto?
- ☐ Ja, altijd
- ☐ Ja, maar in overleg
- ☐ Nee

We willen graag een indruk van de gebouwde omgeving rond uw woning en werk, daarvoor gebruikten wij de postcode. Dit is niet herleidbaar tot uw woning.

23. Wat is de postcode van uw woning? (XXXX AB)

24. Wat is de postcode van uw werk? (XXXX AB)

---

25. Heeft u nog opmerkingen of suggesties?

C. Script data filtering GPS data

```
import csv
import numpy

a = []
description = ["imei,time,lng,lat,angle,speed,altitude"]
route = [description]
name = "resultsTXX_X_01.csv"
inputname = "ResultsT21_4.csv"

data = list(csv.reader(open(inputname)))
name = name[:8] + inputname [8:12] + name[12:]

#Hier worden de punten met een snelheid lager dan 5 eruit gehaald
for i in range(len(data)):
    if i == 0:
        i += 1
    else:
        if int(data[i][5]) <= 5:
            i += 1
        else:
            a.append(data[i])
            i += 1


for j in range(len(a)):
    if j == len(a)-1:
        route.append(a[j])
        totalspeed = 0
        maxspeed = 0
        for k in range(len(route)-1):
            totalspeed += int(route[k+1][5])
            if int(route[k+1][5]) > maxspeed:
                    maxspeed = int(route[k+1][5])
        avgspeed = totalspeed / (len(route)-1)
        #Hier kiezen hoeveel punten een route zijn
        #Hier kiezen welke avg speed de cutoff is
        if len(route) > 20 and avgspeed < 30 and avgspeed > 10 and maxspeed
        < 40:
            #with open(name, 'w', newline='', encoding='utf-8') as myfile:
                #wr = csv.writer(myfile, quoting=csv.QUOTE_MINIMAL)
                #wr.writerows(route)
            with open(name, 'w', newline='', encoding='utf-8') as csv_file:
                writer = csv.writer(csv_file, delimiter=',')
                writer.writerows(route)

            tempname = int(name[13:15])+1
            if tempname < 10:
                print(name)
                name = name[:14] + str(tempname) + name[15:]

            else:
                name = name[:13] + str(tempname) + name[15:]
        #route = [description]
        print("hoera")
        #route.clear()
    else:
        #print(a[j][1])
        datetime1 = a[j][1]
        date1 = (datetime1[:10])
```

```
time1 = (datetime1[11:])

datetime2 = a[j+1][1]
date2 = (datetime2[:10])
time2 = (datetime2[11:])
#print(date1)
#print(time1)

Y1= int(date1[:4])
m1= int(date1[5:7])
d1= int(date1[8:10])
H1= int(time1[:2])
M1= int(time1[3:5])
S1= int(time1[6:8])

Y2= int(date2[:4])
m2= int(date2[5:7])
d2= int(date2[8:10])
H2= int(time2[:2])
M2= int(time2[3:5])
S2= int(time2[6:8])

timedifference = (S2+(60*M2)+(3600*H2))-(S1+(60*M1)+(3600*H1))
#Hier kiezen na hoeveel minuten pauze een nieuwe route begint
if abs(timedifference) > 300:
    route.append(a[j])
    totalspeed = 0
    maxspeed = 0
    for k in range(len(route)-1):
        totalspeed += int(route[k+1][5])
        if int(route[k+1][5]) > maxspeed:
            maxspeed = int(route[k+1][5])
    avgspeed = totalspeed / (len(route)-1)
    #Hier kiezen hoeveel punten een route zijn
    #Hier kiezen welke avg speed de cutoff is
    if len(route) > 20 and avgspeed < 30 and avgspeed > 10 and
    maxspeed < 40:
        #with open(name, 'w', newline='', encoding='utf-8') as
        myfile:
            #wr = csv.writer(myfile, quoting=csv.QUOTE_MINIMAL)
            #wr.writerows(route)
        with open(name, 'w', newline='', encoding='utf-8') as
        csv_file:
            writer = csv.writer(csv_file, delimiter=',')
            writer.writerows(route)

        tempname = int(name[13:15])+1
        if tempname < 10:
            print(name)
            name = name[:14] + str(tempname) + name[15:]

        else:
            name = name[:13] + str(tempname) + name[15:]
    route = [description]

else:
    route.append(a[j])
```

D: Spatial data reclassification

The reclassification of the data into the values in the table below are designed for the generation of the alternative routes.

Road type	Value
No data	[null]
Mixed Road	1
Non-cycling path	2
Bicycle lane	3
Separated bicycle path	4

Intersection type	Value
No data	[null]
Intersection	1
Intersection with traffic lights	2
Roundabout	3

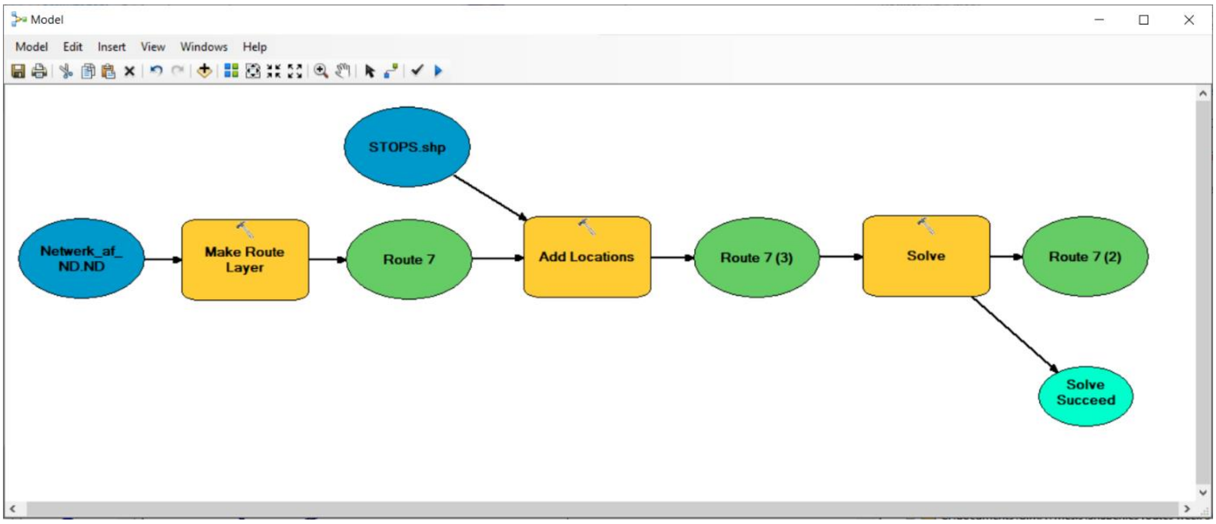
Street lighting	Value
No data	[null]
Not lit	1
Moderately lit	2
Well lit	3

Maximum speed	Value
No data	[null]
15	1
30	1
50	1
60	2
70	2
80	2
100	2
120	2
130	2

Environment	Value
No data	[null]
Built environment	1
Rural environment	2
Nature	3

Water	Value
No data	[null]
No	0
Yes	1

E. Model builder alternative routes



F. Script spatial data preparation

```
import pandas as pd
import numpy as np
import os
filenames = []
print(os.getcwd()) #path waar je nu in zit
for f in os.listdir():
    if f.endswith(".txt"):
        filenames.append(f)
dataframes = [pd.read_csv(f, delimiter=";") for f in filenames]
tracker_df = []
for df, name in zip(dataframes, filenames):
    df.insert(0, "Route", f"{name.split('.')[0]}")
    df.insert(1, "TrackerID", f"{name[:5]}")
    df.insert(2, "numberoflinks", f"{len(df)}")
    tracker_df.append(df)
final_df = pd.concat(tracker_df).reset_index(drop=True)
final_df.iloc[:, -25:-8] = final_df.iloc[:, -25:-8].convert_dtypes()
final_df.iloc[:, -25:-8] = final_df.iloc[:, -25:-8].apply(lambda x:
x.str.replace(",", "."))
final_df.iloc[:, -25:-8] = final_df.iloc[:, -25:-8].apply(lambda x:
x.astype("float").astype("int"))
final_df.iloc[:, -7:-2] = final_df.iloc[:, -7:-2].convert_dtypes()
final_df.iloc[:, -7:-2] = final_df.iloc[:, -7:-2].apply(lambda x:
x.str.replace(",", "."))
final_df.iloc[:, -7:-2] = final_df.iloc[:, -7:-2].apply(lambda x:
x.astype("float").astype("int"))
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].convert_dtypes()
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].str.replace(",",
".").astype("float").apply(lambda x: round(x, 2))

Length = final_df.loc[:, ["Route", "Shape_Le_1"]]
Length = Length.groupby("Route").sum()
Length["Total_Length"] = Length["Shape_Le_1"]
del Length["Shape_Le_1"]
final_df = pd.merge(final_df, Length, on = "Route")
final_df["Routeperc"] = ((final_df["Shape_Le_1"] /
final_df["Total_Length"]) * 100).apply(lambda x: round(x, 2))

pd.get_dummies(final_df, prefix="", prefix_sep="")
x = pd.get_dummies(final_df.loc[:, ["WEGNIVEAU", "WEGTYPE", "WEGDEKSRT",
"OMGEVING", "VERLICHTIN", "MAXSNELHEI", "KRP_TYPE"]])
x.insert(0, "Route", final_df["Route"])
x.insert(1, "TrackerID", final_df["TrackerID"])
x.insert(2, "numberoflinks", final_df["numberoflinks"])
frame1 = x.groupby(["Route", "TrackerID", "numberoflinks"]).sum()
frame1.reset_index(level = 2, inplace = True)
frame1.reset_index(level = 1, inplace = True)

y = final_df.loc[:, ["Route", "TrackerID", "WEGKWAL2", "HINDER2",
"SCHOONH2", "VERLICHT2"]]
y.convert_dtypes()
y["WEGKWAL2"] = y["WEGKWAL2"].astype(str).astype(int)
y["HINDER2"] = y["HINDER2"].astype(str).astype(int)
y["SCHOONH2"] = y["SCHOONH2"].astype(str).astype(int)
y["VERLICHT2"] = y["VERLICHT2"].astype(str).astype(int)
y = y.groupby("Route").mean()
frame2 = y.groupby(["Route"]).mean()
frame2.iloc[:, :4] = frame2.iloc[:, :4].apply(lambda x: round(x, 2))
```

```
z = final_df.loc[:, ["Route", "WATER2", "SNELFIETS2", "Routeperc"]]
z["Waterperc"] = z["WATER2"] * z["Routeperc"]
z["Snelfperc"] = z["SNELFIETS2"] * z["Routeperc"]
z.loc[:, ["Route", "Waterperc", "Snelfperc"]].groupby("Route").sum()
frame3 = z.loc[:, ["Route", "Waterperc",
"Snelfperc"]].groupby("Route").sum()

q = final_df.loc[:, ["Route", "MAXSNELH2", "Routeperc"]]
q["MAXSNELH3"] = np.where(q["MAXSNELH2"] == 1, 1, 0)
q["Maxsnelperc"] = q["MAXSNELH3"] * q["Routeperc"]
frame4 = q.loc[:, ["Route", "Maxsnelperc"]].groupby("Route").sum()
frame4.insert(0, "Alternative", "Observed")

Wegtype= final_df.loc[:, ["Route", "WEGTYPE", "WEGTYPE2", "Routeperc"]]
Wegtype["Mixed_Road"] = np.where(Wegtype["WEGTYPE2"] == 1, 1 *
Wegtype["Routeperc"], 0)
Wegtype["Non_Cyclingpath"] = np.where(Wegtype["WEGTYPE2"] == 2, 1 *
Wegtype["Routeperc"], 0)
Wegtype["Bicyle_lane"] = np.where(Wegtype["WEGTYPE2"] == 3, 1 *
Wegtype["Routeperc"], 0)
Wegtype["Separated_Bicyclepath"] = np.where(Wegtype["WEGTYPE2"] == 4, 1 *
Wegtype["Routeperc"], 0)
Wegtype["Voetgangersdoorsteekje"] = np.where(Wegtype["WEGTYPE"] ==
"voetgangersdoorsteekje", 1 * Wegtype["Routeperc"], 0)
Wegtype2 = Wegtype.loc[:, ["Route", "Mixed_Road", "Non_Cyclingpath",
"Bicyle_lane", "Separated_Bicyclepath",
"Voetgangersdoorsteekje"]].groupby("Route").sum()

Omgeving = final_df.loc[:, ["Route", "OMGEVING2", "Routeperc"]]
Omgeving["Landelijk1"] = np.where(Omgeving["OMGEVING2"] == 1, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Landelijk5"] = np.where(Omgeving["OMGEVING2"] == 5, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Landelijk"] = Omgeving["Landelijk1"] + Omgeving["Landelijk5"]
Omgeving["Bebouwd2"] = np.where(Omgeving["OMGEVING2"] == 2, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Bebouwd3"] =  np.where(Omgeving["OMGEVING2"] == 3, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Bebouwd"] = Omgeving["Bebouwd2"] + Omgeving["Bebouwd3"]
Omgeving["Natuur4"] = np.where(Omgeving["OMGEVING2"] == 4, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Natuur6"] = np.where(Omgeving["OMGEVING2"] == 6, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Natuur"] = Omgeving["Natuur4"] + Omgeving["Natuur6"]
Omgeving2 = Omgeving.loc[:, ["Route", "Landelijk", "Bebouwd",
"Natuur"]].groupby("Route").sum()

NewMerge = pd.merge(frame4, Wegtype2, on = "Route").merge(Omgeving2, on =
"Route")
NewMerge.to_excel("ObservedNewMerge.xlsx")

result = pd.merge(frame1, frame2, on = "Route").merge(frame3, on = "Route")

spatjoin = final_df.iloc[:, :1]
spatjoin["Ongevallen"] = final_df.loc[:, "Count_sum"]
spatjoin["VerkeersLichten"] = final_df.loc[:, "Sign_Count"]
spatjoin["StopBorden"] = final_df.loc[:, "Stop_Count"]
spatjoin["Bruggen"] = final_df.loc[:, "Brug_count"]
spatjoinA = spatjoin.groupby("Route").sum()
spatjoinB = final_df.iloc[:, :1]
spatjoinB["Misdaadcijfers"] = final_df.loc[:, "Criminalit"]
```

```
spatjoinB["Misdaadcijfers"] = spatjoinB["Misdaadcijfers"].convert_dtypes()
spatjoinB["Misdaadcijfers"] =
spatjoinB["Misdaadcijfers"].astype("float")#.astype("int")
spatjoinC= spatjoinB.groupby("Route").mean()
spatjoin2 = pd.merge(Length, spatjoinA, on = "Route").merge(spatjoinC, on =
"Route")

spatjoin2["Ongev/len"] = spatjoin2["Ongevallen"] /
spatjoin2["Total_Length"]
spatjoin2["Verkeerslicht/len"] = spatjoin2["VerkeersLichten"] /
spatjoin2["Total_Length"]
spatjoin2["Stop/len"] = spatjoin2["StopBorden"] / spatjoin2["Total_Length"]
spatjoin2["Brug/len"] = spatjoin2["Bruggen"] / spatjoin2["Total_Length"]
spatjoin2["Misdaad/len"] = spatjoin2["Misdaadcijfers"] /
spatjoin2["Total_Length"]
spatjoin2["Ongev/len"] = spatjoin2["Ongev/len"].apply(lambda x: round(x,
2))
spatjoin2["Verkeerslicht/len"] =
spatjoin2["Verkeerslicht/len"].apply(lambda x: round(x, 2))
spatjoin2["Stop/len"] = spatjoin2["Stop/len"].apply(lambda x: round(x, 2))
spatjoin2["Brug/len"] = spatjoin2["Brug/len"].apply(lambda x: round(x, 2))
spatjoin2["Misdaad/len"] = spatjoin2["Misdaad/len"].apply(lambda x:
round(x, 2))
spatjoin2.reset_index(level = 0, inplace = True)
spatjoin3 = spatjoin2.loc[:, ["Route", "Ongev/len", "Stop/len", "Brug/len",
"Misdaad/len"]]

Misdaad = final_df.loc[:, ["Route", "Criminalit", "Routeperc",
"Total_Length"]]
Misdaad["Misdaad"] = Misdaad["Criminalit"] * Misdaad["Routeperc"]
Misdaad["Misdaad"] = Misdaad["Misdaad"].convert_dtypes()
Misdaad["Misdaad"] = Misdaad["Misdaad"].astype("float")
Misdaad
Misdaadtable = Misdaad.groupby("Route").mean()
Misdaadtable.to_csv("Misdaadtable.csv")

Verlichting = final_df.loc[:, ["Route", "VERLICHT2", "Routeperc"]]
Verlichting["Niet_Verlicht"] = np.where(Verlichting["VERLICHT2"] == 1, 1 *
Verlichting["Routeperc"], 0)
Verlichting["Verlicht"] = np.where(Verlichting["VERLICHT2"] == 2 | 3, 1 *
Verlichting["Routeperc"], 0)
Verlichting2 = Verlichting.loc[:, ["Route", "Niet_Verlicht",
"Verlicht"]].groupby("Route").sum()
Verlichting2.insert(0, "Alternative", "Observed")
Verlichting2.to_excel("ObservedVerlichting.xlsx")

DonkerlichtWeer = pd.read_csv('SpatJoin/Donkerlicht_Weer.txt', sep=",")

result2 = pd.merge(result, spatjoin3, on = "Route").merge(DonkerlichtWeer,
on = "Route")
result2.insert(3, "Alternative", "Observed")

result2.to_excel("MergeObservedRoutes.xlsx")
```

G. Script overlap percentage and Path Size Logit

```
#This is an example of the calculation for the observed route. This script
is altered to perform the same calculation for the alternative routes.

import pandas as pd
import numpy as np

import os
filenames = []
print(os.getcwd()) #path waar je nu in zit
for f in os.listdir():
    if f.endswith(".txt"):
        filenames.append(f)
dataframes = [pd.read_csv(f, delimiter=";") for f in filenames]
tracker_df = []
#create file with required information for PSL: link_id, PS length, Total
length
for df, name in zip(dataframes, filenames):
    df.insert(0, "Route", f"{name.split('.')[0]}")
    df.insert(1, "TrackerID", f"{name[:5]}")
    df.insert(2, "numberoflinks", f"{len(df)}")
    tracker_df.append(df)
final_df = pd.concat(tracker_df).reset_index(drop=True)
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].convert_dtypes()
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].str.replace(",",
".").astype("float").apply(lambda x: round(x, 2))


y = final_df.loc[:, ["Route", "Shape_Le_1", "link_id"]]
y.insert(1, "Alternative", "Observed")

x = y.groupby(["Route"]).sum()
x["Tot_Leng"] = x["Shape_Le_1"]
del x["Shape_Le_1"]

z = pd.merge(y, x, on = "Route")
z["PS_Length"] = z["Shape_Le_1"] / z["Tot_Leng"]
z.to_csv("PS_Observed.csv")

import pandas as pd
import numpy as np

#import preparation files all alternatives
Check = pd.read_csv('Check.csv', sep=";")
Check.set_index("Route", inplace = True)

Obs = pd.read_csv('PS_Observed.csv', sep=",")
del Observed["Unnamed: 0"]
Continuous = pd.read_csv('Continuous/PS_Continuous.csv', sep=",")
del Continuous["Unnamed: 0"]
Fastest = pd.read_csv('Fastest/PS_Fastest.csv', sep=",")
del Fastest["Unnamed: 0"]
Green = pd.read_csv("Groen/PS_Green.csv", sep=",")
del Green["Unnamed: 0"]
Shortest = pd.read_csv("Shortest/PS_Shortest.csv", sep=",")
del Shortest["Unnamed: 0"]
SocSafe = pd.read_csv("SocSafe/PS_SocSafe.csv", sep=",")
del SocSafe["Unnamed: 0"]
TrSafe = pd.read_csv("TrSafe/PS_TrSafe.csv", sep=",")
del TrSafe["Unnamed: 0"]
```

```
#Create routelinks all alternatives
Obs["routelink"] = Obs["Route"].astype(str) + " " +
Obs["link_id"].astype(str)
Continuous["routelink"] = Continuous["Route"].astype(str) + " " +
Continuous["link_id"].astype(str)
Fastest["routelink"] = Fastest["Route"].astype(str) + " " +
Fastest["link_id"].astype(str)
Green["routelink"] = Green["Route"].astype(str) + " " +
Green["link_id"].astype(str)
Shortest["routelink"] = Shortest["Route"].astype(str) + " " +
Shortest["link_id"].astype(str)
SocSafe["routelink"] = SocSafe["Route"].astype(str) + " " +
SocSafe["link_id"].astype(str)
TrSafe["routelink"] = TrSafe["Route"].astype(str) + " " +
TrSafe["link_id"].astype(str)
#Calculation if Observed segment overlaps with alternative
Obs["Continuous"] = np.where(Obs["routelink"].isin(Continuous["routelink"])
& Check["Continuous"] == 1, 1, 0)
Obs["Fastest"] = np.where(Obs["routelink"].isin(Fastest["routelink"]) &
Check["Fastest"] == 1, 1, 0)
Obs["Green"] = np.where(Obs["routelink"].isin(Green["routelink"]) &
Check["Green"] == 1, 1, 0)
Obs["Shortest"] = np.where(Obs["routelink"].isin(Shortest["routelink"]) &
Check["Shortest"] == 1, 1, 0)
Obs["SocSafe"] = np.where(Obs["routelink"].isin(SocSafe["routelink"]) &
Check["SocSafe"] == 1, 1, 0)
Obs["TrSafe"] = np.where(Obs["routelink"].isin(TrSafe["routelink"]) &
Check["TrSafe"] == 1, 1, 0)

#Calculation of overlap percentage
Overlap = Obs.groupby("Route").sum()
Overlap = Overlap.drop("T09_1_04")
Overlap = Overlap.drop("T09_1_05")
Overlap = Overlap.drop("T15_2_03")
Overlap = Overlap.drop("T12_4_05")
Overlap = Overlap.drop("T17_4_02")
Overlap.iloc[:, -6:] = Overlap.iloc[:, -6:].apply(lambda x: round(x, 2))
Overlap.to_excel("Overlap.xlsx")

#Path Size Logit Calculation
#Link sum
Obs["LinkSum"] = 1 + Obs["Continuous"] + Obs["Fastest"] + Obs["Green"] +
Obs["Shortest"] + Obs["SocSafe"] + Obs["TrSafe"]’
#path size for individual segment
Obs["PathSize"] = Obs["PS_Length"] * (1 / Obs["LinkSum"])
#path size for complete route
Pathsize = Obs.groupby("Route").sum()

Pathsize = Pathsize.drop("T09_1_04")
Pathsize = Pathsize.drop("T09_1_05")
Pathsize = Pathsize.drop("T15_2_03")
Pathsize = Pathsize.drop("T12_4_05")
Pathsize = Pathsize.drop("T17_4_02")
Pathsize.insert(0, "Alternative", "Observed")
Pathsize.to_csv("Pathsize_new.csv")

Obs["Continuous_Perc"] = Obs["PS_Length"] * Obs["Continuous"]
Obs["Fastest_Perc"] = Obs["PS_Length"] * Obs["Fastest"]
Obs["Green_Perc"] = Obs["PS_Length"] * Obs["Green"]
Obs["Shortest_Perc"] = Obs["PS_Length"] * Obs["Shortest"]
```

```
Obs["SocSafe_Perc"] = Obs["PS_Length"] * Obs["SocSafe"]
Obs["TrSafe_Perc"] = Obs["PS_Length"] * Obs["TrSafe"]
```

H. Script 'day/light'

```
import pandas as pd
import numpy as np

import os
filenames = []
print(os.getcwd()) #path waar je nu in zit
for f in os.listdir():
    if f.endswith(".csv"):
        filenames.append(f)
dataframes = [pd.read_csv(f, delimiter=',') for f in filenames]
tracker_df = []
for df, name in zip(dataframes, filenames):
    df.insert(1, "Route", f"{name[7:-4]}")
    tracker_df.append(df)
final_df = pd.concat(tracker_df).reset_index(drop=True)
final_df

data = final_df.loc[:, ["Route"]]
data["Datum"] = final_df.time.str[:10]
data["Tijd"] = final_df.time.str[11:]

data2 = data.loc[:, ["Route", "Datum", "Tijd"]]
data2["Hour"] = data.Tijd.str[:2]

data2["darklight"] = np.where((data2["Hour"] >= 8) & (data2["Hour"] <= 17),
"light", "dark")

dum = pd.get_dummies(data2.loc[:, "darklight"])
dum.insert(0, "Route", data2["Route"])

dum.groupby(["Route"]).sum()

route3 = data2.loc[:, ["Route", "Datum", "Tijd"]]
route3["Tijdstip"] = np.where(dum["light"] > dum["dark"], 1, 0)
route4 = route3.groupby("Route").mean()

route5 = data2.groupby(["Route", "Datum"]).mean()
del route5["Hour"]
route5.reset_index(level = 1, inplace = True)

tr = route5.join(route4, how="outer")

tr['Datum'] =  pd.to_datetime(tr['Datum'], infer_datetime_format = False )

tr["Datum"] = tr["Datum"].dt.strftime("%d-%m-%Y")

dr = pd.read_csv("Weer/Weer.csv")

dr["Datum"] = pd.to_datetime(dr["Datum"])
dr["Datum"] = dr["Datum"].dt.strftime("%d-%m-%Y")

x = dr.groupby("Datum").sum()
```

```
result = tr.join(x, on=["Datum"])

result.to_csv("Weer/Donkerlicht_Weer.txt")
```

I. Script trip purpose

```
import pandas as pd
import numpy as np

import os
filenames = []
print(os.getcwd()) #path waar je nu in zit
for f in os.listdir():
    if f.endswith(".txt"):
        filenames.append(f)
dataframes = [pd.read_csv(f, delimiter=";") for f in filenames]
tracker_df = []
for df, name in zip(dataframes, filenames):
    df.insert(0, "Route", f"{name[3:-4]}")
    df.insert(1, "ODPair", f"{name.split('.')[0]}")
    tracker_df.append(df)
final_df = pd.concat(tracker_df).reset_index(drop=True)

final_df["BUFF_DIST"] = final_df["BUFF_DIST"].convert_dtypes()
final_df["BUFF_DIST"] = final_df["BUFF_DIST"].str.replace(",",
".").astype("float").astype("int").apply(lambda x: round(x, 2))

selection = final_df.loc[:, ["Route", "ODPair", "FID", "PC6", "shop",
"BUFF_DIST"]]
selection["OorD"] = selection["FID"] + selection["BUFF_DIST"]
selection["Origin_shop"] = np.where((selection["OorD"] == 50), 1, 0)
selection["Destination_shop2"] = selection["OorD"] - 50
selection["Destination_shop"] = np.where(selection["Destination_shop2"] ==
1, 1, 0)

PC = selection.loc[:, ["Route", "Origin_shop",
"Destination_shop"]].groupby("Route").sum()
origindata = pd.read_csv('Selection/origindata.txt', sep=" ")
originPC6 = pd.read_csv('Selection/originPC6.txt', sep=" ")
destinationPC6 = pd.read_csv('Selection/destinationPC6.txt', sep=" ")
origindata["origin_PC6"] = originPC6.loc[:, "origin_PC6"]
origindata["destination_PC6"] = destinationPC6.loc[:, "destination_PC6"]

ODdata = pd.merge(origindata, PC, on = "Route")
ODdata.to_csv("ODdata.csv")

import pandas as pd
import numpy as np
import os

Fiets = pd.read_csv('Fiets2.csv', delimiter=";")
Fiets.set_index('TrackerID', inplace=True)

OD = pd.read_csv('ODdata.csv', delimiter=",")
del OD["Unnamed: 0"]
OD.set_index('Route', inplace=True)

Round = pd.read_csv('Round.csv', delimiter = ";")
Round.set_index('Route', inplace=True)
```

```python
WorkPC = pd.read_csv('WorkPC2.csv', delimiter = ";")
WorkPC["TrackerID"] = WorkPC["Tracker_ID"]#.replace({"Tracker_ID":
"TrackerID"})
del WorkPC["Tracker_ID"]
WorkPC.set_index('TrackerID', inplace=True)

df = pd.read_csv('End.csv', delimiter = ";", error_bad_lines=False)
df.set_index(['Route', 'TrackerID'], inplace=True)

final_df = df.join(OD, how="outer")
final_df2 = final_df.join(Round, how="outer")
final_df2.reset_index(level = 0, inplace = True)

newdataframe = pd.merge(final_df2, Fiets, on="TrackerID", how="outer")
final = pd.merge(newdataframe, WorkPC, on ="TrackerID", how="outer")
final.set_index('Route', inplace=True)
final.reset_index(level = 0, inplace = True)

final.to_csv("Goal.csv")

import pandas as pd
import numpy as np
import os

Final = pd.read_csv('Goal.csv', delimiter=",")
del Final["Unnamed: 0"]

Final["G_Rec"] = np.where(Final["Round"] == 1, 1, 0)
Final["G_Com"] = np.where(Final["Postcode_Work"] == Final["origin_PC6"], 1,
0)
Final["G_Com2"] = np.where(Final["Postcode_Work"] ==
Final["destination_PC6"], 1, 0)
Final["G_Uti"] = np.where((Final["Origin_shop"] +
Final["Destination_shop"]) > 0, 1, 0)

Conditions = [
     Final['G_Rec'] == 1,
    Final['G_Com'] == 1,
    Final['G_Com2'] == 1,
    Final['G_Uti']  == 1
    ]

outputs = [
     'Recreational', 'Commute', 'Commute', 'Utilitarian'
]

New = np.select(Conditions, outputs, 'Other')
df = pd.DataFrame (New)

df["Purpose"] = df[0]
del df[0]

Result = Final.join(df, how = "outer")

Result.to_excel("Eindfile_real.xlsx")
```

J. Stata script

```
*=====================================================================
* Route choice modelling bicycle use
* Maaike Kuiper, 2021
*=====================================================================

* Standaardmap in voor alle data, do-files, etc
cd "C:\documents\GIMA\Thesis\Stata16\Stata16\thesis_stata"

* Import
clear all
import excel Finaldata_routes.xlsx, sheet("Sheet1") firstrow
save gimathesis.dta, replace

*---------------------------------------------------------------------
* Data preparation
*---------------------------------------------------------------------

** Create id variables
egen t = group(Route)
bysort t: generate gid = _n
generate choice = 1 if Alternative == "Observed"
replace  choice = 0 if Alternative != "Observed"
egen id = group(TrackerID)
order t id gid choice, first

** Rename
rename Bebouwd Urban
rename Natuur Nature
rename Landelijk Rural
rename Misdaad CrimeRates
rename Ong_Count_ Accidents
rename Verlicht StreetLighting

** Create new variables
* Natural log path size
generate lnPathSize = ln(PathSize)

* Length in km
generate Length = Total_Length/1000

* Bicycle facilities
generate AllFacilities = WEGTYPE_ + WEGTYPE_ONBEKEND +
WEGTYPE_bromfietspadlangsweg + WEGTYPE_fietspadlangsweg +
WEGTYPE_fietsstraat + WEGTYPE_normaleweg + WEGTYPE_solitairbromfietspad +
WEGTYPE_solitairfietspad + WEGTYPE_ventweg + WEGTYPE_voetgangersdoorsteekje
+ WEGTYPE_voetgangersgebied + WEGTYPE_wegmetfietssuggestie
generate perc_BicycleLane = (WEGTYPE_wegmetfietssuggestie +
WEGTYPE_fietspadlangsweg) / AllFacilities * 100

generate safe_facilitiesilities = Separated_Bicyclepath + perc_BicycleLane
generate non_cyclingpath = 100 - (Separated_Bicyclepath + perc_BicycleLane)

* Generate routepercentage of 30km/h
generate totals_speedlimit = MAXSNELHEI_ + MAXSNELHEI_100 + MAXSNELHEI_120
+ MAXSNELHEI_130 + MAXSNELHEI_30 + MAXSNELHEI_50 + MAXSNELHEI_60 +
MAXSNELHEI_70 + MAXSNELHEI_80 + MAXSNELHEI_ONBEKEND +
MAXSNELHEI_stapvoets15
generate SpeedLimit = (MAXSNELHEI_stapvoets15 + MAXSNELHEI_30) /
totals_speedlimit * 100
```

```
* Missing values
generate perc_missingSpeed = (MAXSNELHEI_ + MAXSNELHEI_ONBEKEND) /
totals_speedlimit * 100
generate missing_environment = 100 - (Urban + Nature + Rural)


*--------------------------------------------------------------------
* Descriptives
*--------------------------------------------------------------------
** Personal characteristics
*Age
collapse Age, by(id)
summarize Age

* Gender
collapse Gender, by(id)
tab Gender
* Gender trips
tab Gender if choice == 1

* Cycling experience
collapse CyclingExperience, by(id)
tab CyclingExperience

* Trip Purpose
*make purpose numerical
generate PurposeN = 1 if Purpose == "Commute"
replace PurposeN = 2 if Purpose == "Recreational"
replace PurposeN = 3 if Purpose == "Utilitarian"
replace PurposeN = 4 if Purpose == "Other"
*tab
collapse PurposeN, by(t)
tab PurposeN

* dark/light
collapse Tijdstip, by(t)
list TrackerID

*Overlap between routes
tab Alternative, summarize(Observed_perc)
tab Alternative, summarize(SocSafe_Perc)
tab Alternative, summarize(TrSafe_Perc)
tab Alternative, summarize(Fastest_Perc)
tab Alternative, summarize(Shortest_Perc)
tab Alternative, summarize(Continuous_Perc)
tab Alternative, summarize(Green)

** Spatial factors
* Describe spatial factors observed routes
summarize Length if choice == 1
summarize Urban if choice == 1
summarize Nature if choice == 1
summarize Rural if choice == 1
summarize Separated_Bicyclepath if choice == 1
summarize perc_BicycleLane if choice == 1
summarize non_cyclingpath if choice == 1
summarize CrimeRates if choice == 1
summarize Accidents if choice == 1
summarize StreetLighting if choice == 1
summarize SpeedLimit if choice == 1
```

```
summarize PathSize if choice == 1
summarize lnPathSize if choice == 1

* Missing values
summarize perc_missingSpeed if choice == 1
summarize missing_environment if choice == 1

* Spatial factors observed (1) & alternatives relatively (0)
tab choice, summarize(Length)
tab choice, summarize(Urban)
tab choice, summarize(Nature)
tab choice, summarize(Rural)
tab choice, summarize(Separated_Bicyclepath)
tab choice, summarize(perc_BicycleLane)
tab choice, summarize(safe_facilities)
tab choice, summarize(CrimeRates)
tab choice, summarize(Accidents)
tab choice, summarize(StreetLighting)
tab choice, summarize(SpeedLimit)

*---------------------------------------------------------------------
* Correlation
*---------------------------------------------------------------------
correlate CrimeRates Accidents StreetLighting SpeedLimit perc_BicycleLane
non_cyclingpath Separated_Bicyclepath Rural Urban Nature

*---------------------------------------------------------------------
* Model
*---------------------------------------------------------------------
* Set choice model
cmset id t, noalt

** Choice model
cmxtmixlogit choice Length lnPathSize Urban Nature StreetLighting
SpeedLimit perc_BicycleLane Separated_Bicyclepath CrimeRates Accidents,
noconstant nolog or

* traffic safety
cmxtmixlogit choice Length lnPathSize StreetLighting SpeedLimit
perc_BicycleLane Separated_Bicyclepath Accidents, noconstant nolog or

* social safety
cmxtmixlogit choice Length lnPathSize Urban Nature StreetLighting
CrimeRates, noconstant nolog or

** Effect of single spatial factors
cmxtmixlogit choice Accidents, noconstant nolog or

cmxtmixlogit choice perc_BicycleLane, noconstant nolog or
cmxtmixlogit choice Separated_Bicyclepath, noconstant nolog or
cmxtmixlogit choice safe_facilities, noconstant nolog or

cmxtmixlogit choice SpeedLimit, noconstant nolog or

cmxtmixlogit choice StreetLighting, noconstant nolog or

cmxtmixlogit choice CrimeRates, noconstant nolog or

cmxtmixlogit choice Rural, noconstant nolog or
cmxtmixlogit choice Urban, noconstant nolog or
cmxtmixlogit choice Nature, noconstant nolog or
```

```
** Effect of interaction variables safety factors

* Accidents
cmxtmixlogit choice c.Accidents#i.Gender, noconstant nolog or
cmxtmixlogit choice c.Accidents#c.Age, noconstant nolog or
cmxtmixlogit choice c.Accidents#i.CyclingExperience, noconstant nolog or
cmxtmixlogit choice c.Accidents#c.Urban, noconstant nolog or

* Bicycle facilities
* Separated bicycle path
cmxtmixlogit choice c.Separated_Bicyclepath#i.Gender, noconstant nolog or
cmxtmixlogit choice c.Separated_Bicyclepath#c.Age, noconstant nolog or
cmxtmixlogit choice c.Separated_Bicyclepath#i.CyclingExperience, noconstant
nolog or

* Bicycle lane
cmxtmixlogit choice c.perc_BicycleLane#i.Gender, noconstant nolog or
cmxtmixlogit choice c.perc_BicycleLane#c.Age, noconstant nolog or
cmxtmixlogit choice c.perc_BicycleLane#i.CyclingExperience, noconstant
nolog or
cmxtmixlogit choice c.perc_BicycleLane#c.Urban, noconstant nolog or

* Separated bicycle path + bicycle lane: safe facilities
cmxtmixlogit choice c.safe_facilities#i.Gender, noconstant nolog or
cmxtmixlogit choice c.safe_facilities#c.Age, noconstant nolog or
cmxtmixlogit choice c.safe_facilities#i.CyclingExperience, noconstant nolog
or
cmxtmixlogit choice c.safe_facilities#c.Urban, noconstant nolog or
cmxtmixlogit choice c.safe_facilities#i.PurposeN, noconstant nolog or

* Speed limits
cmxtmixlogit choice c.SpeedLimit#i.Gender, noconstant nolog or
cmxtmixlogit choice c.SpeedLimit#c.Age, noconstant nolog or
cmxtmixlogit choice c.SpeedLimit#i.CyclingExperience, noconstant nolog or
cmxtmixlogit choice c.SpeedLimit#c.Urban, noconstant nolog or
cmxtmixlogit choice c.SpeedLimit#c.Rural, noconstant nolog or
cmxtmixlogit choice c.SpeedLimit#c.Separated_Bicyclepath, noconstant nolog
or
cmxtmixlogit choice c.SpeedLimit#c.perc_BicycleLane, noconstant nolog or
cmxtmixlogit choice c.SpeedLimit#c.non_cyclingpath, noconstant nolog or

* Street lighting
cmxtmixlogit choice c.StreetLighting#i.Gender, noconstant nolog or
cmxtmixlogit choice c.StreetLighting#c.Age, noconstant nolog or
cmxtmixlogit choice c.StreetLighting#i.Tijdstip, noconstant nolog or

* Crime rates
cmxtmixlogit choice c.CrimeRates#i.Gender, noconstant nolog or
cmxtmixlogit choice c.CrimeRates#c.Age, noconstant nolog or
cmxtmixlogit choice c.CrimeRates#i.Tijdstip, noconstant nolog or
cmxtmixlogit choice c.CrimeRates#c.StreetLighting, noconstant nolog or

* Environment
* Urban environment
cmxtmixlogit choice c.Urban#i.Gender, noconstant nolog or
cmxtmixlogit choice c.Urban#c.Age, noconstant nolog or
cmxtmixlogit choice c.Urban#i.Tijdstip, noconstant nolog or
cmxtmixlogit choice c.Urban#c.StreetLighting, noconstant nolog or
```

```
* Nature
cmxtmixlogit choice c.Nature#i.Gender, noconstant nolog or
cmxtmixlogit choice c.Nature#c.Age, noconstant nolog or
cmxtmixlogit choice c.Nature#i.Tijdstip, noconstant nolog or
cmxtmixlogit choice c.Nature#c.StreetLighting, noconstant nolog or
```

K. Survey results

– traffic safety:

Statement: 'I rather cycle on bicycle paths separated from the road.'

Response	Percentage
Totally disagree	1.67%
Disagree	10.00%
Neutral	20.00%
Agree	48.33%
Totally agree	20.00%

Statement: 'I avoid busy traffic circumstances.'

Response	Percentage
Totally disagree	3.33%
Disagree	38.33%
Neutral	36.67%
Agree	15.00%
Totally agree	6.67%

Statement: 'I rather cycle on roads where the maximum speed for cars is 30 km/h.'

Response	Percentage
Totally disagree	0.00%
Disagree	36.67%
Neutral	40.00%
Agree	20.00%
Totally agree	1.67%
-	1.67%

Statement: 'I avoid unsafe routes, even if this is the shortest route.'

Response	Percentage
Totally disagree	0.00%
Disagree	28.33%
Neutral	28.33%
Agree	38.33%
Totally agree	3.33%
-	1.67%

- Social safety:

Statement: 'If it's dark or getting dark, I like to take a route with proper street lighting.'

Response	Percentage
Totally disagree	0.00%
Male	0.00%
Female	0.00%
Disagree	8.33%
Male	5.00%
Female	3.33%
Neutral	10.00%
Male	5.00%
Female	5.00%
Agree	60.00%
Male	23.33%
Female	36.67%
Totally agree	21.67%
Male	1.67%
Female	20.00%

Statement: 'I take social safety into account when I make route choices.'

Response	Percentage
Totally disagree	5.00%
Male	3.33%
Female	1.67%
Disagree	31.67%
Male	16.67%
Female	15.00%
Neutral	21.67%
Male	10.00%
Female	11.67%
Agree	33.33%
Male	3.33%
Female	30.00%
Totally agree	8.33%
Male	1.67%
Female	6.67%

Statement: 'I avoid abandoned areas, even if this is the shortest route.'

Response	Percentage
Totally disagree	3.33%
Male	1.67%
Female	1.67%
Disagree	16.67%
Male	8.33%
Female	8.33%
Neutral	31.67%
Male	16.67%
Female	15.00%
Agree	41.67%
Male	8.33%
Female	33.33%
Totally agree	5.00%
Male	0.00%
Female	5.00%
-	1.67%
Male	0.00%
Female	1.67%

M. Stata results

Accidents

```
note: alternatives are unbalanced

Mixed logit choice model                     Number of obs      =      2,257
                                             Number of cases    =        421
Panel variable: id                           Number of panels   =         59

Time variable: t                             Cases per panel: min =        1
                                                              avg =     10.9
                                                              max =       25

                                             Alts per case:   min =        2
                                                              avg =      5.4
                                                              max =        7

Integration points:          0                  Wald chi2(1)    =      24.33
Log likelihood =      -667.05161                Prob > chi2     =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Accidents	1.038195	.0078892	4.93	0.000	1.022847	1.053774

```
Mixed logit choice model                     Number of obs      =      2,257
                                             Number of cases    =        421
Panel variable: id                           Number of panels   =         59

Time variable: t                             Cases per panel: min =        1
                                                              avg =     10.9
                                                              max =       25

                                             Alts per case:   min =        2
                                                              avg =      5.4
                                                              max =        7

Integration points:          0                  Wald chi2(2)    =      33.89
Log likelihood =      -661.9752                 Prob > chi2     =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Gender#c.Accidents						
0	1.031172	.0077498	4.08	0.000	1.016094	1.046474
1	1.133325	.0341914	4.15	0.000	1.068254	1.20236

```
note: alternatives are unbalanced

Mixed logit choice model                     Number of obs      =      2,257
                                             Number of cases    =        421
Panel variable: id                           Number of panels   =         59

Time variable: t                             Cases per panel: min =        1
                                                              avg =     10.9
                                                              max =       25

                                             Alts per case:   min =        2
                                                              avg =      5.4
                                                              max =        7

Integration points:          0                  Wald chi2(1)    =      21.31
Log likelihood =      -669.67301                Prob > chi2     =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.Accidents#c.Age	1.001163	.000252	4.62	0.000	1.000669	1.001657

```
Mixed logit choice model                     Number of obs      =      2,257
                                             Number of cases    =        421
Panel variable: id                           Number of panels   =         59

Time variable: t                             Cases per panel: min =        1
                                                              avg =     10.9
                                                              max =       25

                                             Alts per case:   min =        2
                                                              avg =      5.4
                                                              max =        7

Integration points:          0                  Wald chi2(1)    =      12.45
Log likelihood =      -675.09876                Prob > chi2     =     0.0004
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.Accidents#c.Urban	1.000297	.0000842	3.53	0.000	1.000132	1.000462

```
Mixed logit choice model                     Number of obs      =      2,257
                                             Number of cases    =        421
Panel variable: id                           Number of panels   =         59

Time variable: t                             Cases per panel: min =        1
                                                              avg =     10.9
                                                              max =       25

                                             Alts per case:   min =        2
                                                              avg =      5.4
                                                              max =        7

Integration points:          0                  Wald chi2(6)    =      35.24
Log likelihood =      -655.72344                Prob > chi2     =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
CyclingExperience#c.Accidents						
0	1.032143	.0088292	3.70	0.000	1.014983	1.049594
1	1	(omitted)				
2	1.229011	.1051616	2.41	0.016	1.039254	1.453415
3	1.260788	.1506183	1.94	0.052	.9975954	1.593418
4	1.182945	.0813251	2.44	0.015	1.033823	1.353577
5	1.001443	.0186959	0.08	0.938	.9654616	1.038765
6	1.09763	.0417046	2.45	0.014	1.01886	1.18249

Bicycle facilities

– Separated bicycle path:

```
Mixed logit choice model                        Number of obs      =      2,257
                                                Number of cases    =        421
Panel variable: id                              Number of panels   =         59

Time variable: t                                Cases per panel: min =        1
                                                                 avg =     10.9
                                                                 max =       25

                                                Alts per case:   min =        2
                                                                 avg =      5.4
                                                                 max =        7

Integration points:          0                      Wald chi2(1)   =     11.39
Log likelihood =      -675.80445                    Prob > chi2    =    0.0007
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Separated_Bicyclepath	.9853103	.0043207	-3.37	0.001	.9768781	.9938152

```
Mixed logit choice model                        Number of obs      =      2,257
                                                Number of cases    =        421
Panel variable: id                              Number of panels   =         59

Time variable: t                                Cases per panel: min =        1
                                                                 avg =     10.9
                                                                 max =       25

                                                Alts per case:   min =        2
                                                                 avg =      5.4
                                                                 max =        7

Integration points:          0                      Wald chi2(1)   =     12.22
Log likelihood =      -675.28207                    Prob > chi2    =    0.0005
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.Separated_Bicyclepath#c.Urban	.9998666	.0000382	-3.50	0.000	.9997918	.9999414

```
Mixed logit choice model                        Number of obs      =      2,257
                                                Number of cases    =        421
Panel variable: id                              Number of panels   =         59

Time variable: t                                Cases per panel: min =        1
                                                                 avg =     10.9
                                                                 max =       25

                                                Alts per case:   min =        2
                                                                 avg =      5.4
                                                                 max =        7

Integration points:          0                      Wald chi2(7)   =     24.53
Log likelihood =      -666.72697                    Prob > chi2    =    0.0009
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
CyclingExperience#c.Separated_Bicyclepath						
0	.9880129	.0053647	-2.22	0.026	.977554	.9985837
1	1.210719	.3185879	0.73	0.467	.7228674	2.027813
2	1.063134	.0365831	1.78	0.075	.9937966	1.137308
3	.9417077	.0291722	-1.94	0.053	.8862323	1.000656
4	.9816286	.0164982	-1.10	0.270	.9498195	1.014503
5	.9472901	.0161825	-3.17	0.002	.9160981	.9795442
6	.988822	.0118814	-0.94	0.350	.9658069	1.012386

– Bicycle lane:

```
Mixed logit choice model                        Number of obs      =      2,257
                                                Number of cases    =        421
Panel variable: id                              Number of panels   =         59

Time variable: t                                Cases per panel: min =        1
                                                                 avg =     10.9
                                                                 max =       25

                                                Alts per case:   min =        2
                                                                 avg =      5.4
                                                                 max =        7

Integration points:          0                      Wald chi2(1)   =    100.61
Log likelihood =      -621.05618                    Prob > chi2    =    0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
perc_BicycleLane	1.052145	.0053318	10.03	0.000	1.041747	1.062647

```
Mixed logit choice model                        Number of obs      =      2,257
                                                Number of cases    =        421
Panel variable: id                              Number of panels   =         59

Time variable: t                                Cases per panel: min =        1
                                                                 avg =     10.9
                                                                 max =       25

                                                Alts per case:   min =        2
                                                                 avg =      5.4
                                                                 max =        7

Integration points:          0                      Wald chi2(2)   =    101.25
Log likelihood =      -619.63982                    Prob > chi2    =    0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Gender#c.perc_BicycleLane						
0	1.047608	.0058584	8.32	0.000	1.036188	1.059153
1	1.070562	.012888	5.66	0.000	1.045597	1.096122

```
Mixed logit choice model                        Number of obs      =      2,257
                                                Number of cases    =        421
Panel variable: id                              Number of panels   =         59

Time variable: t                                Cases per panel: min =        1
                                                                 avg =     10.9
                                                                 max =       25

                                                Alts per case:   min =        2
                                                                 avg =      5.4
                                                                 max =        7

Integration points:          0                      Wald chi2(1)   =     52.79
Log likelihood =      -651.80377                    Prob > chi2    =    0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.perc_BicycleLane#c.Urban	1.000415	.0000571	7.27	0.000	1.000303	1.000527

```
Mixed logit choice model                Number of obs       =      2,257
                                        Number of cases     =        421
Panel variable: id                      Number of panels    =         59

Time variable: t                        Cases per panel: min =         1
                                                         avg =      10.9
                                                         max =        25

                                        Alts per case:   min =         2
                                                         avg =       5.4
                                                         max =         7

Integration points:            0           Wald chi2(1)     =      73.51
Log likelihood =        -637.58458         Prob > chi2      =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.perc_BicycleLane#c.Age	1.001302	.0001519	8.57	0.000	1.001004	1.001599

```
Mixed logit choice model                Number of obs       =      2,257
                                        Number of cases     =        421
Panel variable: id                      Number of panels    =         59

Time variable: t                        Cases per panel: min =         1
                                                         avg =      10.9
                                                         max =        25

                                        Alts per case:   min =         2
                                                         avg =       5.4
                                                         max =         7

Integration points:            0           Wald chi2(7)     =     108.47
Log likelihood =        -611.61585         Prob > chi2      =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
CyclingExperience#c.perc_BicycleLane						
0	1.054519	.0066134	8.46	0.000	1.041636	1.067561
1	.6316731	.3982884	-0.73	0.466	.1835628	2.173702
2	1.228759	.0758198	3.34	0.001	1.088789	1.386722
3	1.033006	.0390961	0.86	0.391	.9591525	1.112547
4	1.003864	.0184845	0.21	0.834	.9682807	1.040754
5	1.046916	.0145748	3.29	0.001	1.018736	1.075875
6	1.056084	.0156711	3.68	0.000	1.025811	1.087249

– Safe facilities:

```
Mixed logit choice model                Number of obs       =      2,257
                                        Number of cases     =        421
Panel variable: id                      Number of panels    =         59

Time variable: t                        Cases per panel: min =         1
                                                         avg =      10.9
                                                         max =        25

                                        Alts per case:   min =         2
                                                         avg =       5.4
                                                         max =         7

Integration points:            0           Wald chi2(1)     =      32.61
Log likelihood =        -663.41071         Prob > chi2      =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
safe_facilities	1.02188	.003873	5.71	0.000	1.014318	1.0295

```
Mixed logit choice model                Number of obs       =      2,257
                                        Number of cases     =        421
Panel variable: id                      Number of panels    =         59

Time variable: t                        Cases per panel: min =         1
                                                         avg =      10.9
                                                         max =        25

                                        Alts per case:   min =         2
                                                         avg =       5.4
                                                         max =         7

Integration points:            0           Wald chi2(2)     =      34.15
Log likelihood =        -662.04308         Prob > chi2      =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Gender#c.safe_facilities						
0	1.018516	.0042922	4.35	0.000	1.010138	1.026963
1	1.034461	.0089899	3.90	0.000	1.01699	1.052231

```
Mixed logit choice model                Number of obs       =      2,257
                                        Number of cases     =        421
Panel variable: id                      Number of panels    =         59

Time variable: t                        Cases per panel: min =         1
                                                         avg =      10.9
                                                         max =        25

                                        Alts per case:   min =         2
                                                         avg =       5.4
                                                         max =         7

Integration points:            0           Wald chi2(1)     =      31.07
Log likelihood =         -664.0534         Prob > chi2      =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.safe_facilities#c.Age	1.000679	.0001219	5.57	0.000	1.00044	1.000918

```
Mixed logit choice model                Number of obs       =      2,257
                                        Number of cases     =        421
Panel variable: id                      Number of panels    =         59

Time variable: t                        Cases per panel: min =         1
                                                         avg =      10.9
                                                         max =        25

                                        Alts per case:   min =         2
                                                         avg =       5.4
                                                         max =         7

Integration points:            0           Wald chi2(4)     =      35.88
Log likelihood =        -661.25157         Prob > chi2      =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
PurposeN#c.safe_facilities						
1	1.013369	.0086316	1.56	0.119	.996592	1.030429
2	1.079349	.0367855	2.24	0.025	1.009606	1.15391
3	1.02533	.006371	4.03	0.000	1.012919	1.037894
4	1.020408	.0058961	3.50	0.000	1.008917	1.03203

```
Mixed logit choice model                Number of obs       =      2,257
                                        Number of cases     =        421
Panel variable: id                      Number of panels    =         59

Time variable: t                        Cases per panel: min =         1
                                                         avg =      10.9
                                                         max =        25

                                        Alts per case:   min =         2
                                                         avg =       5.4
                                                         max =         7

Integration points:            0           Wald chi2(7)     =      48.19
Log likelihood =        -650.09651         Prob > chi2      =     0.0000
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
CyclingExperience#c.safe_facilities						
0	1.02414	.0046872	5.21	0.000	1.014994	1.033368
1	1.06159	.0822432	0.77	0.440	.9120379	1.235666
2	1.146545	.047973	3.27	0.001	1.056271	1.244533
3	.9717709	.0207415	-1.34	0.180	.9319569	1.013286
4	.9824955	.016809	-1.03	0.302	.9500967	1.015999
5	1.009062	.0128263	0.71	0.478	.9842339	1.034517
6	1.035373	.0142504	2.53	0.012	1.007816	1.063683

– Descriptives all three variables:

```
. tab choice, summarize(Separated_Bicyclepath)
```

choice	Summary of Separated_Bicyclepath Mean	Std. Dev.	Freq.
0	56.229254	24.135058	1,836
1	56.539335	25.095355	451
Total	56.290402	24.322165	2,287

```
. tab choice, summarize(perc_BicycleLane)
```

choice	Summary of perc_BicycleLane Mean	Std. Dev.	Freq.
0	25.316195	17.284088	1,836
1	30.687998	21.706189	451
Total	26.375523	18.36083	2,287

```
. tab choice, summarize(safe_facilities)
```

choice	Summary of safe_facilities Mean	Std. Dev.	Freq.
0	81.545449	25.558984	1,836
1	87.227332	27.234466	451
Total	82.665925	25.990409	2,287

Speed limits

choice	Summary of SpeedLimit Mean	Std. Dev.	Freq.
0	25.554245	22.495738	1,836
1	27.659128	24.784457	443
Total	25.963399	22.967943	2,279

```
note: alternatives are unbalanced

Mixed logit choice model                     Number of obs       =      2,246
                                             Number of cases     =        417
Panel variable: id                           Number of panels    =         59

Time variable: t                             Cases per panel: min =         1
                                                              avg =      10.9
                                                              max =        25

                                             Alts per case:   min =         2
                                                              avg =       5.4
                                                              max =         7

Integration points:           0                   Wald chi2(1)   =      7.21
Log likelihood =      -674.22483                  Prob > chi2    =    0.0072
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
SpeedLimit	.98579	.0052541	-2.69	0.007	.9755458	.9961418

```
Mixed logit choice model                     Number of obs       =      2,246
                                             Number of cases     =        417
Panel variable: id                           Number of panels    =         59

Time variable: t                             Cases per panel: min =         1
                                                              avg =      10.9
                                                              max =        25

                                             Alts per case:   min =         2
                                                              avg =       5.4
                                                              max =         7

Integration points:           0                   Wald chi2(2)   =      7.40
Log likelihood =      -674.12446                  Prob > chi2    =    0.0247
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Gender#c.SpeedLimit						
0	.9846183	.0058725	-2.60	0.009	.9731753	.9961958
1	.9905106	.0117482	-0.80	0.421	.9677501	1.013806

```
Mixed logit choice model                     Number of obs       =      2,246
                                             Number of cases     =        417
Panel variable: id                           Number of panels    =         59

Time variable: t                             Cases per panel: min =         1
                                                              avg =      10.9
                                                              max =        25

                                             Alts per case:   min =         2
                                                              avg =       5.4
                                                              max =         7

Integration points:           0                   Wald chi2(1)   =      6.59
Log likelihood =      -674.52536                  Prob > chi2    =    0.0102
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.SpeedLimit#c.Age	.9996258	.0001457	-2.57	0.010	.9993403	.9999114

```
Mixed logit choice model                   Number of obs      =      2,246
                                           Number of cases    =        417
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(7)    =      12.64
Log likelihood =       -670.52693             Prob > chi2     =     0.0815

------------------------------------------------------------------------------------------
                       choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------------------+-----------------------------------------------------------
choice                        |
CyclingExperience#c.SpeedLimit|
                            0 |   .9830232    .0065291   -2.58   0.010     .9703092    .9959037
                            1 |    1.13852    .1328553    1.11   0.266     .9057591    1.431094
                            2 |   .9765911    .0258749   -0.89   0.371     .9271716    1.028645
                            3 |   1.081278    .0589744    1.43   0.152      .971654     1.20327
                            4 |    .986448    .0190034   -0.71   0.479     .9498965    1.024406
                            5 |   .9847296    .0167564   -0.90   0.366     .9524293    1.018125
                            6 |    .987991    .0156403   -0.76   0.445     .9578073    1.019126
------------------------------------------------------------------------------------------
```

```
Mixed logit choice model                   Number of obs      =      2,246
                                           Number of cases    =        417
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(1)    =      12.20
Log likelihood =       -671.48389             Prob > chi2     =     0.0005

------------------------------------------------------------------------------------
            choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
choice             |
c.SpeedLimit#c.Urban|  .9998044     .000056   -3.49   0.000     .9996946    .9999142
------------------------------------------------------------------------------------
```

```
Mixed logit choice model                   Number of obs      =      2,246
                                           Number of cases    =        417
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(1)    =      11.77
Log likelihood =        -671.5805             Prob > chi2     =     0.0006

------------------------------------------------------------------------------------
             choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------+---------------------------------------------------------------
choice              |
c.SpeedLimit#c.Rural|   .9987815    .0003549   -3.43   0.001     .9980861    .9994774
------------------------------------------------------------------------------------
```

```
Mixed logit choice model                   Number of obs      =      2,246
                                           Number of cases    =        417
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(1)    =      18.09
Log likelihood =       -668.16026             Prob > chi2     =     0.0000

--------------------------------------------------------------------------------------------------
                           choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------------+---------------------------------------------------------------
choice                            |
c.SpeedLimit#c.Separated_Bicyclepath|   .9997271    .0000642   -4.25   0.000     .9996014    .9998529
--------------------------------------------------------------------------------------------------
```

```
Mixed logit choice model                   Number of obs      =      2,246
                                           Number of cases    =        417
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(1)    =      34.87
Log likelihood =       -658.18556             Prob > chi2     =     0.0000

------------------------------------------------------------------------------------------------
                         choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------------------+---------------------------------------------------------------
choice                          |
c.SpeedLimit#c.perc_BicycleLane |    1.00098    .0001661    5.91   0.000     1.000655    1.001306
```

```
Mixed logit choice model                   Number of obs      =      2,246
                                           Number of cases    =        417
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(1)    =       5.59
Log likelihood =       -675.07054             Prob > chi2     =     0.0180

-----------------------------------------------------------------------------------------------
                        choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------------------+---------------------------------------------------------------
choice                         |
c.SpeedLimit#c.non_cyclingpath |   1.000275    .0001162    2.36   0.018     1.000047    1.000503
-----------------------------------------------------------------------------------------------
```

Street lighting

```
note: alternatives are unbalanced

Mixed logit choice model                   Number of obs      =      2,257
                                           Number of cases    =        421
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(1)    =       0.00
Log likelihood =       -681.58795             Prob > chi2     =     0.9465

--------------------------------------------------------------------------------
        choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
choice         |
StreetLighting |   1.000217    .0032279    0.07   0.946     .9939101    1.006563
--------------------------------------------------------------------------------
```

```
Mixed logit choice model                   Number of obs      =      2,257
                                           Number of cases    =        421
Panel variable: id                         Number of panels   =         59

Time variable: t                           Cases per panel: min =        1
                                                            avg =     10.9
                                                            max =       25

                                           Alts per case:   min =        2
                                                            avg =      5.4
                                                            max =        7

Integration points:             0             Wald chi2(2)    =       0.01
Log likelihood =       -681.58711             Prob > chi2     =     0.9969

---------------------------------------------------------------------------------------
               choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------+----------------------------------------------------------------
choice                |
Gender#c.StreetLighting|
                    0 |   1.000279     .003566    0.08   0.938     .9933139    1.007292
                    1 |   .9999352    .0075951   -0.01   0.993     .9851593    1.014933
---------------------------------------------------------------------------------------
```

```
Mixed logit choice model                        Number of obs      =     2,257
                                                Number of cases    =       421
Panel variable: id                              Number of panels   =        59

Time variable: t                                Cases per panel: min =       1
                                                                 avg =    10.9
                                                                 max =      25

                                                Alts per case:   min =       2
                                                                 avg =     5.4
                                                                 max =       7

Integration points:          0                      Wald chi2(1)   =      0.00
Log likelihood =     -681.5891                      Prob > chi2    =    0.9625
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
choice						
c.StreetLighting#c.Age	.9999951	.0001042	-0.05	0.962	.9997909	1.000199

```
Mixed logit choice model                        Number of obs      =     2,257
                                                Number of cases    =       421
Panel variable: id                              Number of panels   =        59

Time variable: t                                Cases per panel: min =       1
                                                                 avg =    10.9
                                                                 max =      25

                                                Alts per case:   min =       2
                                                                 avg =     5.4
                                                                 max =       7

Integration points:          0                      Wald chi2(2)   =      0.07
Log likelihood =    -681.55463                      Prob > chi2    =    0.9650
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
choice						
Tijdstip#c.StreetLighting						
0	.9983268	.0079791	-0.21	0.834	.9828099	1.014089
1	1.000583	.0035296	0.17	0.869	.9936894	1.007525

Crime rates

```
Mixed logit choice model                        Number of obs      =     2,257
                                                Number of cases    =       421
Panel variable: id                              Number of panels   =        59

Time variable: t                                Cases per panel: min =       1
                                                                 avg =    10.9
                                                                 max =      25

                                                Alts per case:   min =       2
                                                                 avg =     5.4
                                                                 max =       7

Integration points:          0                      Wald chi2(1)   =      0.00
Log likelihood =    -681.59016                      Prob > chi2    =    0.9923
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
choice						
CrimeRates	1.000002	.0002473	0.01	0.992	.9995179	1.000487

```
Mixed logit choice model                        Number of obs      =     2,257
                                                Number of cases    =       421
Panel variable: id                              Number of panels   =        59

Time variable: t                                Cases per panel: min =       1
                                                                 avg =    10.9
                                                                 max =      25

                                                Alts per case:   min =       2
                                                                 avg =     5.4
                                                                 max =       7

Integration points:          0                      Wald chi2(2)   =      2.60
Log likelihood =    -679.65468                      Prob > chi2    =    0.2720
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
choice						
Gender#c.CrimeRates						
0	1.000173	.0002687	0.64	0.519	.9996467	1.0007
1	.9981629	.0012407	-1.48	0.139	.9957341	1.000598

```
Mixed logit choice model                        Number of obs      =     2,257
                                                Number of cases    =       421
Panel variable: id                              Number of panels   =        59

Time variable: t                                Cases per panel: min =       1
                                                                 avg =    10.9
                                                                 max =      25

                                                Alts per case:   min =       2
                                                                 avg =     5.4
                                                                 max =       7

Integration points:          0                      Wald chi2(1)   =      0.63
Log likelihood =    -681.26439                      Prob > chi2    =    0.4262
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
choice						
c.CrimeRates#c.Age	.9999939	7.67e-06	-0.80	0.426	.9999789	1.000009

```
Mixed logit choice model                        Number of obs      =     2,257
                                                Number of cases    =       421
Panel variable: id                              Number of panels   =        59

Time variable: t                                Cases per panel: min =       1
                                                                 avg =    10.9
                                                                 max =      25

                                                Alts per case:   min =       2
                                                                 avg =     5.4
                                                                 max =       7

Integration points:          0                      Wald chi2(2)   =      4.18
Log likelihood =    -679.38707                      Prob > chi2    =    0.1237
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
choice						
Tijdstip#c.CrimeRates						
0	1.00083	.0004815	1.72	0.085	.9998862	1.001774
1	.9996745	.0002958	-1.10	0.271	.9990949	1.000254

Type of environment

– Urban environment:

```
Mixed logit choice model                 Number of obs      =      2,257
                                         Number of cases    =        421
Panel variable: id                       Number of panels   =         59

Time variable: t                         Cases per panel: min =        1
                                                          avg =     10.9
                                                          max =       25

                                         Alts per case:   min =        2
                                                          avg =      5.4
                                                          max =        7

Integration points:          0               Wald chi2(1)   =       0.22
Log likelihood =        -681.48009           Prob > chi2    =     0.6386
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Urban	.9984555	.0032858	-0.47	0.639	.9920361	1.004916

```
Mixed logit choice model                 Number of obs      =      2,257
                                         Number of cases    =        421
Panel variable: id                       Number of panels   =         59

Time variable: t                         Cases per panel: min =        1
                                                          avg =     10.9
                                                          max =       25

                                         Alts per case:   min =        2
                                                          avg =      5.4
                                                          max =        7

Integration points:          0               Wald chi2(1)   =       0.24
Log likelihood =        -681.46832           Prob > chi2    =     0.6211
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.Urban#c.Age	.9999452	.0001108	-0.49	0.621	.9997282	1.000162

```
Mixed logit choice model                 Number of obs      =      2,257
                                         Number of cases    =        421
Panel variable: id                       Number of panels   =         59

Time variable: t                         Cases per panel: min =        1
                                                          avg =     10.9
                                                          max =       25

                                         Alts per case:   min =        2
                                                          avg =      5.4
                                                          max =        7

Integration points:          0               Wald chi2(2)   =       0.38
Log likelihood =        -681.40147           Prob > chi2    =     0.8278
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
Gender#c.Urban						
0	.9979055	.0035635	-0.59	0.557	.9909455	1.004914
1	1.001554	.0085201	0.18	0.855	.9849938	1.018393

```
Mixed logit choice model                 Number of obs      =      2,257
                                         Number of cases    =        421
Panel variable: id                       Number of panels   =         59

Time variable: t                         Cases per panel: min =        1
                                                          avg =     10.9
                                                          max =       25

                                         Alts per case:   min =        2
                                                          avg =      5.4
                                                          max =        7

Integration points:          0               Wald chi2(2)   =       0.35
Log likelihood =        -681.41391           Prob > chi2    =     0.8379
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interva
choice						
Tijdstip#c.Urban						
0	.9957415	.0081271	-0.52	0.601	.9799393	1.0117
1	.9989808	.0035925	-0.28	0.777	.9919644	1.0060

```
Mixed logit choice model                 Number of obs      =      2,257
                                         Number of cases    =        421
Panel variable: id                       Number of panels   =         59

Time variable: t                         Cases per panel: min =        1
                                                          avg =     10.9
                                                          max =       25

                                         Alts per case:   min =        2
                                                          avg =      5.4
                                                          max =        7

Integration points:          0               Wald chi2(1)   =       2.76
Log likelihood =        -680.20715           Prob > chi2    =     0.0969
```

choice	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
choice						
c.Urban#c.StreetLighting	.9999574	.0000257	-1.66	0.097	.999907	1.000008

– Nature:

```
Mixed logit choice model                         Number of obs       =      2,257
                                                 Number of cases     =        421
Panel variable: id                               Number of panels    =         59

Time variable: t                                 Cases per panel: min =         1
                                                                  avg =      10.9
                                                                  max =        25

                                                 Alts per case:   min =         2
                                                                  avg =       5.4
                                                                  max =         7

Integration points:            0                      Wald chi2(1)   =      14.47
Log likelihood =       -673.74205                     Prob > chi2    =     0.0001

------------------------------------------------------------------------------
      choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
choice       |
      Nature |   1.099068    .027292     3.80   0.000     1.046857    1.153882
------------------------------------------------------------------------------
```

```
Mixed logit choice model                         Number of obs       =      2,257
                                                 Number of cases     =        421
Panel variable: id                               Number of panels    =         59

Time variable: t                                 Cases per panel: min =         1
                                                                  avg =      10.9
                                                                  max =        25

                                                 Alts per case:   min =         2
                                                                  avg =       5.4
                                                                  max =         7

Integration points:            0                      Wald chi2(2)   =      14.56
Log likelihood =       -673.06093                     Prob > chi2    =     0.0007

------------------------------------------------------------------------------
      choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
choice       |
Gender#c.Nature |
           0 |   1.134662   .0444073     3.23   0.001     1.050879    1.225124
           1 |   1.069764    .035436     2.04   0.042     1.002517    1.141521
------------------------------------------------------------------------------
```

```
Mixed logit choice model                         Number of obs       =      2,257
                                                 Number of cases     =        421
Panel variable: id                               Number of panels    =         59

Time variable: t                                 Cases per panel: min =         1
                                                                  avg =      10.9
                                                                  max =        25

                                                 Alts per case:   min =         2
                                                                  avg =       5.4
                                                                  max =         7

Integration points:            0                      Wald chi2(1)   =      12.36
Log likelihood =       -674.84854                     Prob > chi2    =     0.0004

------------------------------------------------------------------------------
      choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
choice       |
c.Nature#c.Age |   1.002005    .000571     3.52   0.000     1.000887    1.003125
------------------------------------------------------------------------------
```

```
Mixed logit choice model                         Number of obs       =      2,257
                                                 Number of cases     =        421
Panel variable: id                               Number of panels    =         59

Time variable: t                                 Cases per panel: min =         1
                                                                  avg =      10.9
                                                                  max =        25

                                                 Alts per case:   min =         2
                                                                  avg =       5.4
                                                                  max =         7

Integration points:            0                      Wald chi2(2)   =      14.51
Log likelihood =       -673.35042                     Prob > chi2    =     0.0007

------------------------------------------------------------------------------
      choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
choice       |
Tijdstip#c.Nature |
           0 |   1.160455   .0837288     2.06   0.039     1.007425     1.33673
           1 |   1.088861    .028943     3.20   0.001     1.033586    1.147092
------------------------------------------------------------------------------
```

```
note: alternatives are unbalanced

Mixed logit choice model                         Number of obs       =      2,257
                                                 Number of cases     =        421
Panel variable: id                               Number of panels    =         59

Time variable: t                                 Cases per panel: min =         1
                                                                  avg =      10.9
                                                                  max =        25

                                                 Alts per case:   min =         2
                                                                  avg =       5.4
                                                                  max =         7

Integration points:            0                      Wald chi2(1)   =      10.05
Log likelihood =       -676.43504                     Prob > chi2    =     0.0015

------------------------------------------------------------------------------
      choice | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
choice       |
c.Nature#c.StreetLighting |   1.001485   .0004688     3.17   0.002     1.000567    1.002404
------------------------------------------------------------------------------
```