School of Social and Political Sciences

Urban Studies

Cycling out of deprivation: what Strava Metro data reveal about the relationship between cycling and deprivation in Glasgow

Word count: 14,624

Presented in partial fulfilment of the requirements for the degree of M.Sc. in Urban Planning and Regeneration

Abstract

The Scottish Government has a bold vision for cycling: that 10% of everyday journeys will be made by bicycle by 2020. Despite policy and spending commitment from all levels of government, the vision is extremely unlikely to be achieved. This raises some key questions: who is not cycling and why? Residents of the most deprived areas of Glasgow have been found to be less likely to commute by bicycle, yet they might have the greatest need to experience the health, economic and connectivity benefits that cycling can bring. This study therefore sets out to explore the relationship between the number of bicycle journeys originating in an area of Glasgow and the level of deprivation of that area. It does so using Strava Metro and Scottish Index Of Multiple Deprivation data, allowing for detailed temporal and spatial analysis. It finds that there does appear to be a relationship between cycling and deprivation in the Glasgow City Council area: the number of bicycle journeys increases, as the level of overall deprivation decreases. This positive association is present for all journeys over the course of 2016, and it is especially significant for morning commute journeys. Several actions are identified to increase levels of cycling among residents of deprived areas, such as focusing on cycling for leisure, given the low of levels of commuting by bicycle, as well as conducting qualitative research to clarify the barriers to cycling and the ways to overcome them.

Contents

Figures

Tables

Acknowledgements

First and foremost, I would like to acknowledge the support of my supervisor, Dr. Katarzyna Sila Nowicka. Thank you for guiding me through a completely new field of study, for introducing me to a completely new set of programmes, for helping me to see cities and cycling in a completely new way. I would have been lost without you. Secondly, I would like to thank my parents, Barbara and Nicholas, for giving me a love of learning and a love of cycling, both of which have inspired this study.

1. Introduction

1.1. Cycling in Scotland

Eight years ago, the Scottish Government launched the Cycling Action Plan For Scotland with a clear vision: that by 2020, 10% of everyday journeys will be made by bicycle (2010). It was claimed that this vision, while undoubtedly bold, was eminently achievable: one third of journeys in Scotland are under two miles (Ibid). The Cycling Action Plan has been reviewed every three years since its launch in 2010, and the commitment to achieve the vision by 2020 has been maintained. This commitment is manifest from Scottish Government spending: the active travel budget was doubled in 2017 from £40 million per year to £80 million per year, which equates to £13.50 per person of the population, significantly higher than the £6.50 per person in England (Cycling Weekly, 2018. Scottish Government, 2017A). Furthermore, the policy and spending commitment to increase levels of cycling is not only manifest at the national level, but also regionally and locally. For Glasgow, at a regional level, the Strategic Development Plan or Clydeplan aims to "increase levels of active travel through the provision of safe and convenient opportunities for walking and cycling" (Glasgow & The Clyde Valley Strategic Development Planning Authority, 2017: 82). At a local level, the City Development Plan seeks to discourage non-essential car journeys and encourage opportunities for active travel, in order to make Glasgow a more connected city (Glasgow City Council, 2016). Indeed, Glasgow's Strategic Plan For Cycling pledges to continue spending on cycling above the Scottish national average, envisioning the city as a place where "cycling is accessible, safe and attractive to all" (Glasgow City Council, 2015: 18).

The policy to increase levels of cycling also accords with wider policies, due to the clear benefits of cycling. With the Cleaner Air Strategy and the Climate Change Bill, the Scottish Government set out the objectives for Scotland's air to be the best in Europe and for emissions to be reduced by 90% by 2050 (2015.

2017B). The shift to active travel will make an important contribution towards achieving these objectives, as well as meeting health targets. The recently launched Get Active, Stay Active plan aims to cut physical inactivity in Scottish adults by 15% by 2030, in response to the high levels of obesity: two thirds of adults are overweight and almost a third are obese (Scottish Government, 2018). Again, not only nationally but also regionally and locally, there is an acknowledgement that increasing levels of cycling can contribute towards delivering wider policy objectives. It can improve health and reduce inequalities, according to the regional Clydeplan (Glasgow & The Clyde Valley Strategic Development Planning Authority, 2017). The local City Development Plan emphasises that cycling is not only healthy and sustainable, but it can also “facilitate social interaction and cost effective access to services, facilities and jobs” (Glasgow City Council, 2016: 109). Glasgow’s Strategic Plan For Cycling highlights that cycling can help achieve various objectives: boosting the local economy; providing cheap and quick access to employment; improving physical health and wellbeing; even boosting self-esteem (2015).

Increasing levels of cycling, therefore, appears to be a priority for national, regional and local government in Scotland. There is an awareness of the benefits of cycling and an acknowledgment of its important role in delivering wider policy objectives: from health and wellbeing, to the economy and the environment. There is also a higher level of government spending for cycling in Scotland than elsewhere in the UK. Despite this policy and spending commitment, however, the vision of 10% of all journeys being made by bicycle by 2020 is extremely unlikely to be achieved: with just eighteen months remaining, the latest transport figures reveal that still only 1% of journeys in Scotland are being made by bicycle (Transport Scotland, 2018). The gap between envisioned and actual levels of cycling is vast and it raises key questions. This introduction will now explore these questions, and ultimately this study will aim to answer them.

1.2. Cyclists and non-cyclists

The main key question is why, despite the policy and spending commitment from all levels of government, are more people in Scotland not cycling? First, however, another key question must be asked: who in Scotland is not cycling? The answer will in turn reveal an answer to the main question: why are they not cycling? Are there geographical factors, with levels of cycling being affected by where people live: the lack of cycling infrastructure; the condition of the roads; the distance to important destinations such as shops, schools and places of work? Are there social factors, with cycling being especially unpopular among particular groups of people? If the reasons for people not cycling are understood, then they can be addressed with concerted and targeted action, such as developing cycling infrastructure in particular places or promoting cycling among particular groups. This type of action is arguably the only way of ensuring that 10% of everyday journeys in Scotland are made by bicycle: if not by the year 2020, then by the earliest possible time. The focus of this study is thus narrowing, towards answering the key question mentioned above: who in Scotland is not cycling?

It will not be the first study to attempt to answer this question. The Bike Life report, conducted by the walking and cycling charity Sustrans, observed lower levels of cycling among women and ethnic minority communities in Edinburgh: cyclists are 63% male and 37% female, 97% white and 3% black and ethnic minority ethnic (2017). This accords with several other studies, which found that women are less likely to cycle for various reasons: environmental factors related to concerns over road safety; personal factors related to complexity of journeys, which need to incorporate travel for commuting, shopping and childcare (Ogilvie et al., 2012. Steinbach et al., 2011). Another study of Glasgow analysed levels of commuting by bicycle, focusing on differences not in terms of gender or ethnicity but rather deprivation. It found that residents of the least deprived areas of the city are nearly three times more likely to cycle to their place of work or study than those of the most deprived areas (Glasgow Centre For Population Health, 2017). However, this study had severe limitations: it solely used data from Scotland's

Census in 2011 to gauge the number of people in an area who cycled to their place of work or study, and it did not take into account other factors in an area that might affect bicycle use. Accordingly, there is need for the relationship between cycling and deprivation to be studied in more detail. There is also potential for such a study to be especially valuable in Glasgow, for several reasons.

Firstly, there are a significant number of people in Glasgow living in deprivation. Almost 50% of the city's population, 283,000 people, reside in the 20% most deprived areas in Scotland (Glasgow City Council, 2016). Secondly, according to a body of studies, cycling can bring various benefits. In terms of health, cycling reduces rates of obesity, diabetes and hypertension: cyclists have higher levels of cardiorespiratory fitness and lower risk of mortality, and they have even been shown to live for six months longer (Glasgow Centre For Population Health, 2017. Panter et al., 2011. Pistoll et al., 2014). In terms of the environment, cycling reduces levels of emissions, both directly by reducing the number of cars and indirectly by easing the flow of traffic (Brand et al., 2014). In terms of the economy and connectivity, cycling is also widely acknowledged to be beneficial. Scotland's Third National Planning Framework emphasises the importance of cycling in achieving the key objective of becoming a connected, successful and sustainable place (Scottish Government, 2014). The City Development Plan and Glasgow's Strategic Plan For Cycling both highlight that cycling can provide cheap and fast access to essential services, facilities and places of work (Glasgow City Council, 2015. Glasgow City Council, 2016).

Thirdly and most significantly, it is precisely the people who live in the most deprived areas of Glasgow who might have the greatest need to experience the benefits of cycling. In terms of health, Scotland is a markedly unequal country, with the highest levels of obesity among both children and adults to be found in the most deprived areas (Information Services Division, 2018. Law et al., 2011). Glasgow has one of the lowest life expectancies in the UK, with the highest levels of illness and mortality concentrated in the most deprived areas of the city (Leyland et al., 2007. Walsh et al., 2013). Residents of deprived areas of Glasgow are also less physically active: in 2014, only 54% met government guidelines that adults should be moderately active for at least 150 minutes every week, with levels being especially low among women, compared with 70% of residents of the least deprived areas (Glasgow Centre For Population Health, 2017. Lamb et al., 2012). Residents of deprived areas of the city also have a poorer diet, consuming less high fibre bread, potatoes, pasta and rice, as well as cereals and green vegetables (Gray et al., 2008). Finally, they are more likely to report a lack of safe spaces for children to play and a lack of amenities conducive to physical activity (Ellaway et al., 2001. Glasgow Centre For Population Health, 2017). It seems clear, therefore, that residents of deprived areas have a need to experience the health benefits that cycling can bring. In terms of economic benefits, cycling can provide access to places of work: a priority for people living in employment deprivation. It is significantly cheaper to buy, maintain, use and park a bicycle than a car: a benefit for those living in income deprivation. In terms of connectivity, cycling can also bring benefits: 51% of people in Glasgow do not have access to a car, much higher than other cities in Scotland, and car ownership is lower in deprived areas of the city (Glasgow Centre For Population Health, 2017). Furthermore, cuts in local government spending can force cuts in public transport services, leaving people who are without a car isolated and unable to access essential facilities, services and places of work (Ibid).

The focus of this study is thus narrowing further, towards examining the relationship between cycling and deprivation in Glasgow. Cycling can bring many health, economic and connectivity benefits. It is the significant number of people living in the deprived areas of Glasgow, where levels of health, employment, income and connectivity are generally lower, who might have the greatest need to experience these benefits. These are the people, however, who have been found to be significantly less likely to cycle to their place of work or study (Glasgow Centre For Population Health, 2017). If the study can corroborate this finding, or at least analyse the relationship between cycling and deprivation in more detail, it could contribute towards increasing the understanding of who is not cycling at the moment and how they can be encouraged to cycle in the future. It could also contribute towards meeting an objective set out in Glasgow's Strategic Plan For Cycling: to conduct "research to identify specific, local actions to increase cycling and target particular groups" (Glasgow City Council, 2015: 31). The potential value of this type of targeted research is clear: gradually enabling more people to experience the benefits of cycling, and ultimately achieving the vision of the Cycling Action Plan for 10% of everyday journeys in Scotland to be made by bicycle.

1.3. Approach

Detailed data are required on both cycling and deprivation, in order to study the relationship between the two. Data on the latter can be obtained from the Scottish Index Of Multiple Deprivation (SIMD), which in 2016 ranked every area in the country according to its deprivation. The overall deprivation ranking is based on several separate aspects of deprivation, such as health, crime, employment, education, housing and access. Data on cycling are more challenging to obtain. These are several conventional methods of analysing bicycle use, including surveys and counts. As discussed in chapter 2.4, however, surveys are limited in the size of samples and the level of detail, while both automated and manual counts do not gather any information on the origin of a bicycle journey. This information is essential for a study of the relationship between cycling and deprivation: as the data on deprivation inherently pertain to residents of a specific area, the data on cycling must also pertain to residents of a specific area, and an accurate method of measuring the bicycle use by residents of a specific area is to measure the number of bicycle journeys starting in that area. There is one source of information on where bicycle journeys start and end, and it is provided by the app Strava. While also having limitations, discussed in detail in chapter 2.5, Strava Metro data are a rich and vast source of information: revealing exactly where and when Strava users cycle. These data are made available under sublicense to the Urban Big Data Centre at the University of Glasgow, and they include one dataset that reports the origins and destinations of every bicycle journey in the Glasgow area over the course of a year. These data, therefore, enable the focus of the study to become narrower and clearer: using Strava Metro data to analyse the relationship between the number of bicycle journeys originating in an area of Glasgow, and the level of deprivation of that area.

The study will begin with a literature review, discussing previous studies that have both examined the relationship between deprivation and cycling, and employed Strava Metro data to analyse bicycle use. It will then provide a detailed description of the data related to cycling, deprivation and geography, before giving a clear explanation of the methodology used to analyse this data: examining whether the number of bicycle journeys fluctuates according to changes in the level of deprivation and other variables. The results of the analysis will then be displayed and discussed: scrutinising what they reveal about the relationship between cycling and deprivation, and exploring whether the relationship changes over a different period. The study will conclude with observations on bicycle use in Glasgow at the moment and recommendations for increasing it in the future, so that cycling can gradually become an everyday mode of transport for everyone.

2. Literature Review

This chapter comprises two parts: the first focusing on studies that examine the relationship between cycling and deprivation, and the second looking at those that employ Strava data to analyse bicycle use. As will be discussed, there appears to be an absence of research that uses Strava Metro data to analyse the relationship between cycling and deprivation: an absence that this study aims to fill.

2.1. Cycling and deprivation: Scottish studies

Only two years remain until 2020, the date when the Cycling Action Plan For Scotland envisioned that 10% of everyday journeys would be made by bicycle, but much progress needs to be made before this vision is actually achieved: nationally only 1% of all journeys are made by bicycle, and 1.2% of people cycle as their main mode of transport (Transport Scotland, 2018. Cycling Scotland, 2017). The latter figure is the same for Glasgow, yet only 1.6% of people in the city commute to their place of work or study by bicycle: much lower than the commuting figures for Scotland of 2.2% and for Edinburgh of 4.3% (Cycling Scotland, 2017. Glasgow Centre For Population Health, 2017). As mentioned in chapter 1.2, 283,000 people in Glasgow are living in deprivation, almost 50% of the population, yet there does not appear to be a significant amount of research into bicycle use in deprived areas of the city (Glasgow City Council, 2016). In general, relevant research on the city analyses ‘active travel’: encompassing both walking and cycling, and often failing to differentiate between the two.

The most recent study, which has significant limitations as set out in chapter 1.2, found that residents of the least deprived areas of the city are nearly three times more likely to cycle to their place of work or study than those of the most deprived areas (Glasgow Centre For Population Health, 2017). Other studies, however, point to a complex or unclear relationship between cycling and deprivation. Bicycle use in Glasgow was found to relate not to different levels of deprivation, but rather to different “sectors” of the city (McCartney et al., 2012).

Using census data and cordon counts, it was observed that levels of cycling are highest in the west sector of Glasgow, across both the more deprived and less deprived areas (Ibid). Levels of walking are also high in the “affluent west sector” and “deprived east sector”, where there are “the most pleasant and direct active travel routes” (Ibid: 124). Levels of active travel are lower in the north sector of Glasgow, however, despite having similar levels of deprivation to the east sector (Ibid). These findings led to the conclusion that “there was no clear pattern in walking or cycling across deprivation deciles” (Ibid: 122). Instead, the highest levels of cycling were observed in the sector with the best cycling infrastructure (Ibid). This conclusion corroborates two earlier studies, which emphasise the importance of infrastructure in encouraging people to walk or cycle (Ogilvie et al., 2007. Ogilvie et al., 2004).

A recent study reached a different conclusion: that there are indeed differences in levels of active travel between socioeconomic groups in Scotland (Olsen et al., 2017). Using data from the Scottish Household Survey, it was found that residents of the most deprived areas are more likely to travel using active modes than those living in the least deprived areas (Ibid). However, as noted above, the definition of active travel encompasses both walking and cycling: the authors claim that “due to a low frequency of cycling, we were unable to distinguish between walking and cycling journey stages in our analyses” (Ibid: 133). Only one other study appears to have recognised a similar relationship in Scotland: that active travel generally, and cycling specifically, is more popular among people with lower incomes. Transport Scotland reported that people from households with an annual income of less than £15,000 are more likely to cycle to work than those from households with a higher income (2017). A study of the UK as a whole found that people from the lowest income households have greater odds of active travel, yet it failed to distinguish between walking and cycling (Rind et al., 2015). Similarly in Brisbane, Australia, residents of the most deprived areas are more likely to walk than those living in the more affluent areas, although no evidence was found of a relationship between cycling and deprivation specifically (Rachele et al., 2015).

As stated above, it appears to be rare for studies of areas outside Scotland to analyse the relationship between levels of cycling specifically and levels of deprivation. More common is to examine the relationship between cycling and affluence or income: one aspect of an area's deprivation, alongside health, crime, employment, education, housing and access and crime. It is to these studies that the literature review now turns.

2.2. Cycling and deprivation: international studies

The majority of studies appear to have found that levels of cycling are lower among people with lower incomes. In the UK, this is true both locally and nationally. In Bristol, people from middle class households show high levels of active travel, yet those from deprived areas of the city rely on cars even for short journeys (Bird, 2010). In London, cycling is dominated by the affluent. Only 1.5% of those living in households earning under £15,000 make at least one trip by bicycle on any given day in the capital, compared with 2.2% of those living in households earning over £35,000 (Green et al., 2010). Furthermore, only 9% of people with an annual household income of less than £20,000 cycle at least once a week, contrasting with 15% of those with an income of more than £50,000 (Transport For London, 2011). Similarly, only 7% and 4% of the people who use Cycle Superhighways 7 and 3 respectively have an annual household income of less than £20,000 (Ibid). Accordingly the situation is summarised thus: "in London… cycling is disproportionately an activity of affluent, white, men" (Steinbach et al., 2011: 1123). In the 2011 English and Welsh Census, there was also an association between greater affluence and higher levels of commuting by bicycle in Bristol, Cambridge, Oxford and Greater London (Goodman et al., 2013). Crucially, these are cities where levels of cycling are high (Ibid). Examining England and Wales more broadly, a different association was observed: "cycling was fairly equal across the socioeconomic gradient but was also slightly more common in deprived areas" (Ibid: 6). Affluent households are significantly more likely to commute by driving, significantly less likely to commute by walking or using public transport, and marginally less likely to commute by cycling (Ibid).

However, other studies of England and Wales found that affluent people are more likely to commute by cycling, and that more people from households with one car commute to work by cycling, than those from households with no car (Parkin et al., 2007. Steer Davies Gleave, 2010). According to the *Strategic Review Of Health Inequalities In England*, 38% of people from the highest 'social grade' use a bicycle in any given week, compared to 12% from the lowest (Marmot et al., 2010). In the UK as a whole, the proportion of people who have access to a bicycle and who cycled in the previous 12 months increases with household income (Department for Transport, 2017). A similar relationship has been observed in internationally. In Stockholm, Sweden, cycling is most popular among people from high income households (Bastian et al., 2017). In San Francisco, USA, deprived areas are associated with low bicycle use (Cervero et al., 2003). In Johannesburg, South Africa, levels of cycling are significantly higher in more affluent areas of the city (Musawka et al., 2016). In Melbourne, Australia, levels of recreational cycling are lower in deprived areas of the city (Kamphuis et al., 2008). In Brisbane, Australia, a similar finding was made about levels of utility cycling, i.e. using a bicycle for commuting or for purposes other than recreation: "being male, younger, employed full time, or university educated increased the likelihood of utility cycling" (Sahlqvist et al., 2012: 818).

A minority of studies, however, reached a different conclusion: that levels of cycling are actually higher among people with lower incomes. In England, it was observed that children from households with lower socioeconomic status are more likely to cycle to primary school (Panter et al., 2013). In London, it was found that the public bicycle sharing scheme is used more by residents of deprived areas of the city (Ogilvie et al., 2012). In the USA, people in poorest household income quartile conduct the highest share of cycling journeys (Pucher et al., 2011). Across the USA, it was observed that 39% of bicycle commuting is conducted by the poorest quartile of household incomes, almost double the 20% conducted by the richest quartile (Flanagan et al., 2016). Furthermore, several other studies found that levels of utility cycling are higher among people from less affluent households, while levels of recreational cycling are higher among people from more affluent households. This finding was observed in international literature reviews, and in a study of active travel across Europe (Beenackers et al., 2012. Heinen et al.. 2010. Krizek et al., 2009). In Melbourne, Australia, “commuters in the most affluent areas were in fact less likely to cycle to work than those in less affluent areas” (Pistoll et al., 2014). However, the factor with the greatest influence on utility cycling levels is not affluence per se but rather infrastructure: more precisely, it is the absence of infrastructure appropriate for utility cycling (Ibid). Although cycling infrastructure is more common in affluent areas of Melbourne, this infrastructure is not always appropriate for commuting: it might be indirect routes through parks, for example, rather than direct routes to the inner city (Ibid).

In Brisbane, Australia, “neighbourhood disadvantage” is not associated with levels of recreational cycling in Brisbane, yet it is associated with levels of utility cycling (Heesch et al., 2015). The association is more nuanced still: “those in Quartiles 1 and 4 (living in the most and least disadvantaged neighbourhoods, respectively) are more likely to cycle for transport than those living in Quartile 3” (Ibid: 158). This association is explained by better cycling infrastructure: “the findings strongly suggest that government investments that provide bicycle infrastructure within inner Brisbane appear to have resulted in more transport cycling than in outer areas and to appeal to residents of the most and least disadvantaged neighbourhoods” (Ibid: 160). In both Brisbane and Melbourne, therefore, the greatest influence on cycling levels appears to be infrastructure rather than affluence per se: utility cycling is encouraged by the presence of infrastructure in Brisbane, and discouraged by its absence in Melbourne (Ibid. Pistoll et al., 2014).

2.3. Cycling and deprivation: conclusion

These studies have contested McBeth's claim that "cycling transcends class, ethnicity, gender, sexuality and age" (2009: 165). The clear majority concluded that levels of cycling are lower among people with lower incomes. This conclusion was reached internationally, nationally for the UK and locally for Scotland and Glasgow. A minority of studies – international, national and local – reached the opposite conclusion, that levels of cycling were higher among people with lower incomes. Several found a more nuanced relationship between cycling and deprivation: highlighting the differences between recreational and utility cycling; noting the influence and importance of infrastructure.

As noted above, there appears to be a focus on analysing the relationship between cycling and affluence, which is only one factor in the deprivation of an area. There also appears to be an absence of research that analyses the relationship between deprivation or indeed affluence and cycling with Strava Metro data. Such data have been employed to study various other aspects of bicycle use, and it is these studies that the literature review will now discuss.

2.4. Strava and cycling: advantages

There are a significant number of studies that analyse bicycle use with data from Strava. Established in 2009 in San Francisco, USA, the Strava app enables cyclists, runners and hikers to track and thereby upload their routes on a smartphone or GPS device (Sun et al., 2017B). It also allows users to analyse their performance, providing information on distance, speed, elevation, cadence, heart rate and calories burned (Christou, 2016. Dunleavy, 2015). Every week, 100,000 new people start using Strava and 2.5 million routes are uploaded to the app, contributing to its database of over 300 billion data points (Dunleavy, 2016. Jestico et al., 2016). The data have both academic and commercial potential. In

2014, the company launched Strava Metro, which provides data services to local authorities and research institutions (Romanillos et al., 2016).

Strava Metro data are comprised of GPS “traces”, which are uploaded by users when they track their routes (Sun et al., 2017B: 2). The traces are aggregated to streets and anonymised to protect users’ privacy (Ibid). Demographic information about gender and age is also provided, but it too is aggregated (Ibid. Romanillos et al., 2016). The data for an area comprise three subsets: Streets, providing minute-by-minute counts of cyclists at every street; Nodes, offering cyclist counts and waiting times at every intersection; Origin / Destination, reporting the start and end points of every journey (Strava, 2016). A full description of the data used in this study is given in chapter 3.1. Thus Strava Metro data are a rich and vast source of information for research into bicycle use: showing how many users are cycling on specific streets per hour, per day and per year; revealing where users are cycling to and from; displaying the actual routes that users take (Conrow et al., 2018. Macklon et al., 2018. Sun et al., 2017B.).

As stated above, the literature review will focus on the numerous studies that employ Strava Metro data to analyse bicycle use. The studies clearly express the advantages of using this data over conventional methods of gathering data on cycling: methods such as “manual bicycle counts, automated bicycle counts, regional travel surveys and direct questionnaires” (Conrow et al., 2018: 22). Manual bicycle counts do not gather any information on the demographic of cyclists, the reason for travelling, or the origin, destination or route of their journey (Ibid). Such counts have severe spatial and temporal limitations, being only situated in certain places and only conducted at certain times (Ibid. Kuzmyak et al., 2014. Ryus et al., 2014). Furthermore, they are expensive, cumbersome and time consuming to conduct (Jestico et al., 2016. Musakwa et al., 2016). Automated bicycle counts are being increasingly used, with the advantage of not being temporally limited: data are collected continuously as cyclists pass (Griffin et al., 2014). Yet these counts still have other limitations: gathering no information other than the number and time of journeys; being situated only in certain places (Jestico et al., 2016. Kuzmyak et al, 2014). Indeed, both manual and automated bicycle counts have “a high spatial granularity but a low spatial coverage”, tending to be located on major rather than minor roads (Sun, 2017A: 1357).

Travel surveys, meanwhile, are limited in the size of samples and the level of detail (Conrow et al., 2018). Furthermore, although cyclists might be asked about the origin and destination points of their journeys, many travel surveys then assume that the route is the most direct path between these two points (Van Heeswijck et al., 2015). Cyclists often choose not to take the most direct path, however, opting instead to avoid traffic or stay on cycling infrastructure (Conrow et al., 2018. Dill, 2009). Using direct questionnaires, it is possible to glean more detailed information about cyclists’ demographics, motivations, perceptions and routes (Conrow et al., 2018). Yet there remain some limitations, with restricted spatial coverage, small sample sizes and expensive costs (Ibid. Griffin et al., 2015. Jestico et al., 2016). Faced with these limitations in the conventional methods of gathering data, studies clearly express the advantages of using Strava Metro data. They enable bicycle use to be analysed in high resolution spatial and temporal detail, as well as on an extensive spatial scale: revealing the precise times, origins, destinations and routes of journeys; allowing researchers to “sample movement across a city” (Boss et al., 2018: 7. Conrow et al., 2018. Sun, 2017A).

It is noticeable that several studies use the word “unprecedented”: Strava Metro data “include unprecedented spatial and temporal resolution”, and can allow for “the effect and impact of interventions to be explored with an unprecedented level of detail and accuracy” (Boss et al., 2018: 7. Macklon et al., 2018: 11). Strava Metro data also have the potential to be a valuable tool for both researchers and authorities: revealing how the behaviour of cyclists is shaped by the local environment; identifying areas where cycling infrastructure is in high demand and where it is needed; revealing preferences between proposed routes for new cycling infrastructure, by enabling underused routes to be discounted (Conrow et al., 2018. Figliozzi et al., 2015. Macklon et al., 2018. Norman et al., 2015. Sun, 2017A. Sun et al., 2017B).

2.5. Strava and cycling: limitations

Strava Metro data also have clear limitations, which are acknowledged and discussed by the studies employing it to analyse bicycle use. Because GPS traces and demographic information are aggregated to safeguard users' privacy, it is not possible to analyse the length of journey, the reasons for travelling or the choice of route on an individual level (Romanillos et al., 2016). Neither it is possible to analyse the role of factors such as age or gender in bicycle use, yet such analysis is likely to be important in planning, designing and managing inclusive cycling infrastructure (Ibid). Another limitation is noted in a study that examines whether Strava Metro data are useful for evaluating how cyclists' behaviour is affected by changes in cycling infrastructure (Heesch et al., 2016). It concludes that Strava Metro data are not especially useful for this purpose, as the number of Strava users is increasingly rapidly: 100,000 new people start using the app every week (Ibid). Over the long term, therefore, it is challenging to assess the impact of any change in cycling infrastructure: the impact is likely to be overshadowed or obscured by the increase in Strava users (Ibid).

Furthermore, another study speculates that, when choosing routes, Strava users are influenced by different factors (Macklon et al., 2018). One factor is competition: because users are encouraged to compete for 'records' over sections of particular routes, they might be more likely to "cluster on particular routes rather than distributing themselves evenly on routes that are representative of general ridership" (Ibid: 4). This relates to a broader limitation of Strava Metro data, which is widely highlighted and discussed: its representativeness. How closely do Strava users represent overall cyclists? How accurately do the bicycle journeys tracked by the app correlate to all the bicycle journeys made in an area?

There are acknowledged to be several issues with the representativeness of Strava Metro data. Firstly, only cyclists with smartphones or GPS devices can track their journeys on the Strava app. This indicates potential for demographic or socioeconomic bias, as only those cyclists who have certain technological and thus financial resources will be represented in the data (Corney, 2016. Goodchild, 2007. Gould, 2013. Heipke, 2010). There is also potential for demographic and attitudinal bias because samples are self-selected: active and enthusiastic cyclists, wishing to showcase their athletic achievements, might be more likely to hear about and engage with Strava (Corney, 2016. Heesch et al., 2016. Macklon et al., 2018. Romanillos et al., 2016). Accordingly, several groups might be underrepresented in Strava Metro data: people with mobility impairments, elderly people, children, students and casual or recreational cyclists (Conrow et al., 2018. Romanillos et al., 2016. Sun, 2017A). Several studies have set out to ascertain the representativeness of Strava Metro data. The literature review will now examine the findings of these studies, as well as those that have employed the data to analyse various other aspects of bicycle use.

2.6. Strava and cycling: outcomes

As mentioned above, the studies that analyse bicycle use with Strava Metro data acknowledge that they have limitations: primarily that Strava users are not representative of overall cyclists. This specific issue has been closely examined. Strava Metro data were compared with demographic information from a household survey in Queensland, Australia, and the conclusion was clear: men, specifically those aged 35–44, are overrepresented in the Strava Metro data (Heesch et al., 2016). 80.1% of Strava users are men and 29.2% were men aged 35–44, compared with 72.1% and 17.7% of overall cyclists (Ibid). A similar conclusion was reached in Austin, USA: Strava users were “heavily skewed toward the male sex, and most are between 25 and 54 years of age” (Griffin et al., 2015: 9).

Other studies have focused on cycling volumes, examining the association in the number of cycle journeys between Strava Metro data and data gathered with conventional methods. In Glasgow, the annual number of journeys on specific streets in Strava Metro data was compared with the annual average daily flow data on those streets provided by the Department of Transport, and an 83% correlation was found (Sun et al., 2017B). This finding indicates that "the spatial distribution of Strava cycling volume is fairly proportional to real cycling volume" (Ibid: 5). In London, Cycle Census data and Strava Metro data were analysed and a 70% correlation was found: "the initial results indicate that data collected using Strava are a promising data source for traffic managers" (Haworth, 2016: 1). In Portland, USA, data from automated cycle counters on a city centre bridge were compared with data on cycling volumes along the same route from Strava Metro, and a 91% correlation was found (Herrero, 2016). In Ottawa-Gatineau, Canada, the data from 11 automated cycle counters were compared with data from Strava Metro and a similar finding was made: "the linear correlations between the Strava sampled ridership and official counts of all bicyclists were high and ranged from 0.76 to 0.96" (Boss et al., 2018: 4). Furthermore, data from manual cycle counters in Victoria, Canada, were compared with data from Strava Metro and a 40%– 56% correlation was found, which indicates that "crowdsourced data may be a good proxy for estimating daily, categorical cycling volumes" (Jestico et al., 2016: 94). A comparison between data from manual cycling counters and data from Strava Metro in Sydney, Australia, found a 79% correlation, showing "a relatively strong positive correspondence in bicycling volumes across the study area" (Conrow et al. 2018: 26). The study area comprised 14 areas across Sydney and the association was found to be lowest in the northern suburbs, where the cycling volumes in Strava Metro data are much higher than in the manual cycling counter data (Ibid). Despite being popular with Strava users, these routes do not have or are not located near any cycling infrastructure, leading to the conclusion that "since many Strava users are focused on fitness, it is possible that despite not having infrastructure, the roads in this area in some way support riding for purposes other than commuting" (Ibid: 26).

A similar conclusion was reached when Strava Metro data were used to analyse the environmental factors on cycling volumes in Austin, USA (Griffin et al., 2015). There was not found to be a clear association between cycling volumes and cycling infrastructure, but instead it was observed that Strava users tend to use routes with steep slopes and challenging terrain (Ibid). It was therefore inferred that fitness focused Strava users, “seeking a route for the purpose of training goals”, might avoid routes with high volumes of traffic and high numbers of traffic lights, such as routes through the centre of Austin, even though that is where most cycling infrastructure is located (Ibid: 17). This does not mean, however, that cycling infrastructure is not deemed desirable by fitness-focused cyclists, but rather that infrastructure is “generally not provided in more rural areas that more likely suit their training desires” (Ibid).

In Glasgow, UK, Strava users were found to be more likely to cycle to green spaces, as well as to routes along the river Clyde (Sun, 2017A). They also tend towards streets with a low volume of traffic, streets surrounded by residential land rather than by commercial or industrial land, and short streets connected with longer and busier roads (Sun et al., 2017B). Furthermore, because Strava users were more likely to cycle at the outskirts of Glasgow, they were potentially exposed to less air pollution (Sun et al., 2017C). In the study of Victoria, Canada, mentioned prevously, Strava Metro data were employed to analyse the influence of environmental factors on bicycle use (Jestico et al., 2016). It made several findings. The first was that cyclists tend to avoid routes with higher traffic speeds and more on-street parking. This corroborates previous studies, which found that higher speeds of traffic and greater amounts of on-street parking deter cyclists (Hood et al., 2011. Stinson et al., 2003). The second finding was that steeper slopes are also deterrents, with a 1% increase in slope resulting in 72 fewer cyclists on average. This is contrary to the study of Austin, USA, mentioned above, which observed that slopes are in fact preferred by Strava users (Griffin et al., 2015). The third finding, however, was similar to one from Austin: “the presence of bike facilities was not significant in predicting cycling volumes” (2016: 95).

Several studies, however, reached a different conclusion. In Ottawa-Gatineau, Canada, for the study mentioned above, Strava Metro data were analysed and it was found that Strava users prefer to use routes with separated cycle paths (Boss et al.; 2018). Indeed, once new cycling infrastructure had been constructed in Ottawa-Gatineau, Strava users actually changed their routes to make use of it (Ibid). Similarly, GPS trackers were attached to cyclists in Portland, USA, and used to glean their preferences: first separated cycle paths; then bicycle boulevards; then streets with low traffic; then streets with painted bicycle lanes; lastly streets without painted bicycle lanes (Broach et al., 2012).

2.7. Strava and cycling: conclusion

Strava Metro has clearly had a significant impact on the way that cycling data are gathered and analysed. Its data provide a high level of spatial and temporal detail over an extensive spatial scale and offer potential for a range of important uses: understanding cycling volumes and routes; examining the influence of environmental factors on bicycle use, thus helping to create better infrastructure; even evaluating cyclists' exposure to air pollution, thus helping to improve health.

The majority of studies appear to agree that Strava Metro data can be correlated with data gathered by other methods, yet there are differing opinions about the other findings that they yield, such as the preferences of Strava users for steep slopes, or cycling infrastructure. There are also doubts over the representativeness of the data, which are widely acknowledged and discussed. As several studies emphasise, however, such doubts do not mean that Strava Metro data should be completely discounted. Instead, it should be used to supplement conventional methods of gathering data: enabling research to be conducted in greater detail on a greater scale, thus leading to a greater understanding of bicycle use (Conrow et al., 2018. Jestico et al., 2016. Macklon et al., 2018).

3. Data and Methodology

Quantitative methods are inherent to the nature of the focus of this study: using Strava Metro data to analyse the relationship between cycling and deprivation in Glasgow. As this analysis examines if one variable fluctuates according to changes in other variables, it requires a regression model (Wooldridge, 2009). The 'explained' or dependent variable is the number of bicycle journeys; the 'explanatory' or independent variables are the level of deprivation and other factors that might affect bicycle use. This chapter will begin by describing the data and the variables in detail, before explaining the methodology used to determine the appropriate regression model and analyse the relationship between the variables.

3.1. Data: Strava Metro

The data on cycling are obtained from Strava Metro, available under sublicense to the Urban Big Data Centre at the University of Glasgow. Strava Metro data for an area comprise three subsets: Streets; Nodes; Origin / Destination. For the reason explained in chapter 1.3, this study used the Origin / Destination dataset: as the data on deprivation inherently pertain to residents of a specific area, the data on cycling must also pertain to residents of a specific area, and an accurate method of measuring the bicycle use by residents of a specific area is to measure the number of bicycle journeys starting in that area. The Origin / Destination dataset contains geographical information for each bicycle journey uploaded to Strava over the course of a year, such as the Polygon where it starts and ends (Strava Metro, 2016). Each Polygon relates to an Output Area, the smallest unit for which census data are provided, with an average size of 50 households (Office Of National Statistics, 2018). The dataset also contains temporal information for each bicycle journey, such as the minute, hour and day when it starts. This study used the dataset for the Glasgow area and for the year 2016, the most recent available: enabling the analysis to be detailed and up-to-date.

3.2. Data: deprivation and geographical

The data on deprivation were obtained from the Scottish Index Of Multiple Deprivation (SIMD), which in 2016 ranked every Data Zone in the country according to its deprivation. Data Zones are larger than Output Areas, with an average population of 500–1000 residents (Scottish Government, 2013). These Data Zones received individual scores for overall deprivation: from 1 (most deprived) to 6976 (least deprived). They were also ranked for overall deprivation in Quintiles, Deciles and Vigintiles, with individual scores categorised into groups of five, ten and twenty respectively. The overall deprivation ranking is based on several separate aspects of deprivation, such as health, crime, employment, education, housing and access. Data Zones received individual scores for each of the separate aspects. In addition, SIMD data contain information on which these individual scores are based: the score for housing, for example, is based on the percentage and the number of houses in a Data Zone that are overcrowded and without heating. For this study, the independent variables related to the deprivation of Data Zones were obtained from SIMD data, while the independent variables related to the geography of Data Zones were obtained from other sources, listed in table 1 below.

3.3. Data: variables

Using the programme RStudio, version 1.1.453, the cycling, deprivation and geographical data were compiled into one comprehensive dataset. Each Polygon in the Strava Metro data was converted into an Output Area, and then matched with the corresponding Data Zone in the SIMD and geographical data. Data Zones outwith the city of Glasgow were then removed from the comprehensive dataset. The final dataset comprised the **746** Data Zones within the Glasgow City Council area: enabling a focused and detailed study of cycling in the city. All the variables in this dataset are listed and explained in table 1 below.

Name	Category	Source	Description
1.1. All Journeys	Dependent / numerical: general	Strava Metro	All bicycle journeys uploaded to Strava in 2016, starting within the Glasgow City Council area.
1.2. Commute Journeys	Dependent / numerical: temporal	Strava Metro	Bicycle journeys uploaded to Strava in 2016, starting within the Glasgow City Council area, between 6am–10am on Monday–Friday: showing where morning commutes start, i.e. where residents live.
2. SIMD Quintile	Independent / categorical: deprivation	SIMD	Overall deprivation ranking of Data Zone, categorised from 1 (most deprived) to 5 (least deprived).
3. Health Quintile	Independent / categorical: deprivation	SIMD	Overall health deprivation of Data Zone, categorised from 1 (most health deprived) to 5 (least health deprived): calculated for this study by categorising the individual health deprivation score into five groups, where each group represents 20% of Data Zones, i.e. 1–1499 = ‘1’, 1500–2999 = ‘2’ etc.
4. Crime Quintile	Independent / categorical: deprivation	SIMD	Overall crime deprivation of Data Zone, categorised from 1 (most crime deprived) to 5 (least crime deprived): calculated for this study using the method described for variable 3.
5. Employment Quintile	Independent / categorical: deprivation	SIMD	Overall employment deprivation of Data Zone, categorised from 1 (most employment deprived) to 5 (least employment deprived): calculated for this study using the method described for variable 3.
6. Education Quintile	Independent / categorical: deprivation	SIMD	Overall education deprivation of Data Zone, categorised from 1 (most education deprived) to 5 (least education deprived): calculated for this study using the method described for variable 3.
7. Working Age Population Density	Independent / numerical: deprivation	SIMD / Open Street Map (OSM)	Density of working age population in Data Zone: calculated for this study by dividing the number of residents of working age (men aged 16–64 and women aged 16–60), by the area in km^2 (Scottish Government, 2016B).

8. Overcrowded Percentage	Independent / numerical: deprivation	SIMD	Percentage of residents in Data Zone living in overcrowded housing: defined as having at least one room fewer than required for number of occupants (Ibid).
9. No Heating Percentage	Independent / numerical: deprivation	SIMD	Percentage of residents in Data Zone living in housing without central heating (Ibid).
10. Distance To Primary School	Independent / numerical: deprivation	SIMD	Average distance in Data Zone to the nearest primary school in metres: calculated for this study from the average drive time and an assumed average speed of 30mph.
11. Distance To Retail Centre	Independent / numerical: deprivation	SIMD	Average distance in Data Zone to the nearest retail centre in metres: calculated for this study from the average drive time and an assumed average speed of 30mph.
12. Distance To City Centre	Independent / numerical: deprivation	N/A	Distance in metres from centre of Data Zone to George Square: calculated for this study by sourcing the coordinates of the Data Zone with the programme QGIS, version 3.2.0, and calculating the distance to the coordinates of George Square using the Euclidean distance formula.
13. Distance To GP Surgery	Independent / numerical: deprivation	SIMD	Average distance in Data Zone to the nearest GP surgery in metres: calculated for this study from the average drive time and an assumed average speed of 30mph.
14. Distance To Petrol Station	Independent / numerical: deprivation	SIMD	Average distance in Data Zone to the nearest petrol station in metres: calculated for this study from the average drive time and an assumed average speed of 30mph.
15. Distance To Post Office	Independent / numerical: deprivation	SIMD	Average distance in Data Zone to the nearest post office in metres: calculated for this study from the average drive time and an assumed average speed of 30mph.
16. Distance To Secondary School	Independent / numerical: deprivation	N/A	Average distance in Data Zone to the nearest secondary school in metres: calculated for this study from the average drive time and an assumed average speed of 30mph.

17. Road Density	Independent / numerical: deprivation	OSM	Density of road network in Data Zone: calculated for this study by dividing the length of roads in metres by the area in km^2.
18. Pothole Complaints Density	Independent / numerical: geographical	FixMy Street. com / UBDC	Density of complaints made about potholes in roads of Data Zone: calculated for this study by dividing the number of complaints by the area in km^2.
19. Transport Points	Independent / numerical: geographical	OSM	Number of public transport stops in Data Zone: including bus, train and subway.
20. Green Space Density	Independent / numerical: geographical	OSM	Density of green space in Data Zone: calculated for this study by dividing the area in km^2 of green space by the total area in km^2.

Table 1: detailed description of all the variables in the final dataset

Name	**Min**	**Max**	**Mean**	**Standard Deviation**
1.1. All Journeys	0	15106	253.98	714.08
1.2. Commute Journeys	0	551	63.29	88.401
7. Working Age Population Density	0	40080	5065	4404.83
8. Overcrowded Percentage	1	51	17.21	11.73
9. No Heating Percentage	1	18	9.436	5.09
10. Distance To Primary School	563.3	3781.9	1616.9	588.09
11. Distance To Retail Centre	643.7	6356.9	2627.7	963.68
12. Distance To City Centre	127.7	10297.6	4941.2	2241.78
13. Distance To GP Surgery	482.8	5471.8	1733.7	740.04
14. Distance To Petrol Station	724.2	4747.6	2232.9	754.79
15. Distance To Post Office	482.8	3862.4	1725	593.16
16. Distance To Secondary School	965.6	6678.8	3323.4	1011.49
17. Road Density	2824	51922	21146	7152.21
18. Pothole Complaints Density	0	383.539	28.328	34.31
19. Transport Points	0	44	3.188	3.23
20. Green Space Density	0	804963	137429	158136

Table 2: descriptive statistics for the numerical variables in the final dataset

3.4. Data: initial analysis

From an initial analysis using RStudio, several key details and patterns emerged from the dataset. There were **189468** bicycle journeys uploaded to Strava in 2016, starting in **731** Data Zones within the Glasgow City Council area. There were no bicycle journeys that started in the remaining **15** Data Zones. Bar charts were used to reveal the spread of bicycle journeys over the day, week and year, as well as the spread of deprivation quintiles throughout the city. In addition, maps were created with QGIS, to illustrate the spatial distribution of variables in the dataset.

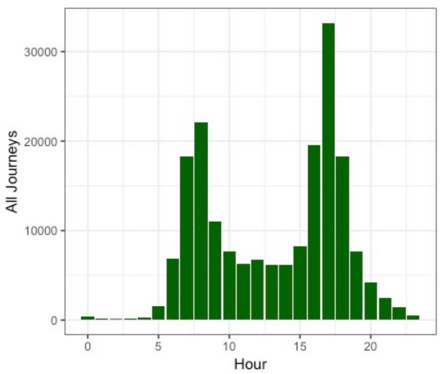

Figure 1: times of all bicycle journeys (variable 1.1), throughout the day

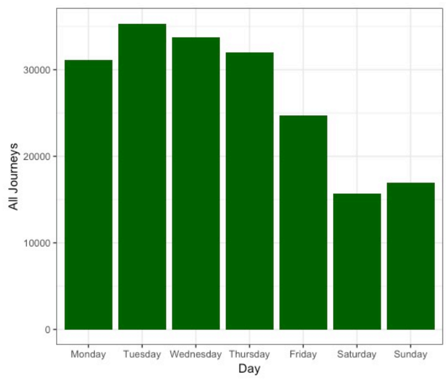

Figure 2: dates of all bicycle journeys (variable 1.1), throughout the week

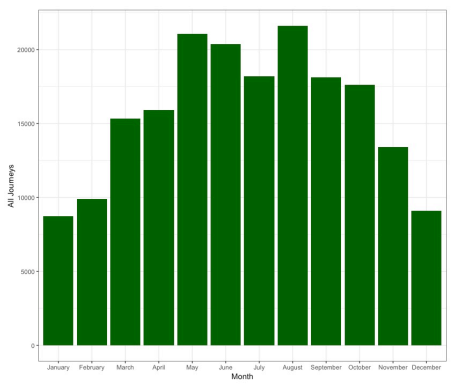

Figure 3: months of all bicycle journeys (variable 1.1), throughout the year

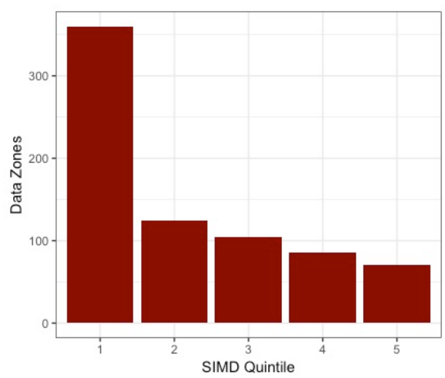

Figure 4: number of Data Zones per SIMD Quintile (variable 2)

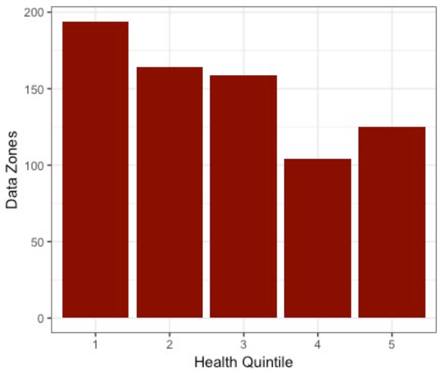

Figure 5: number of Data Zones per Health Quintile (variable 3)

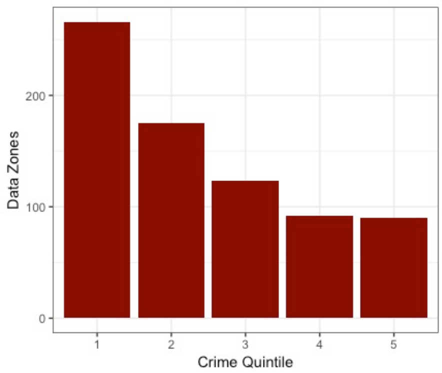

Figure 6: number of Data Zones per Crime Quintile (variable 4)

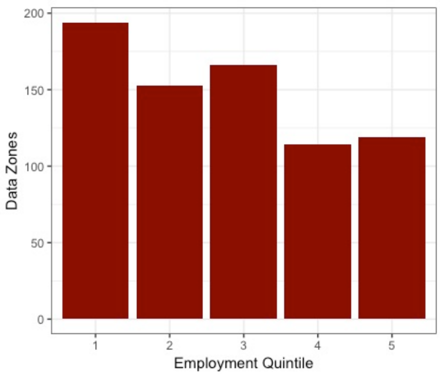

Figure 7: number of Data Zones per Employment Quintile (variable 5)

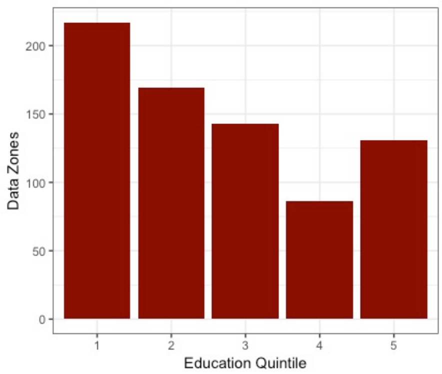

Figure 8: number of Data Zones per Education Quintile (variable 6)

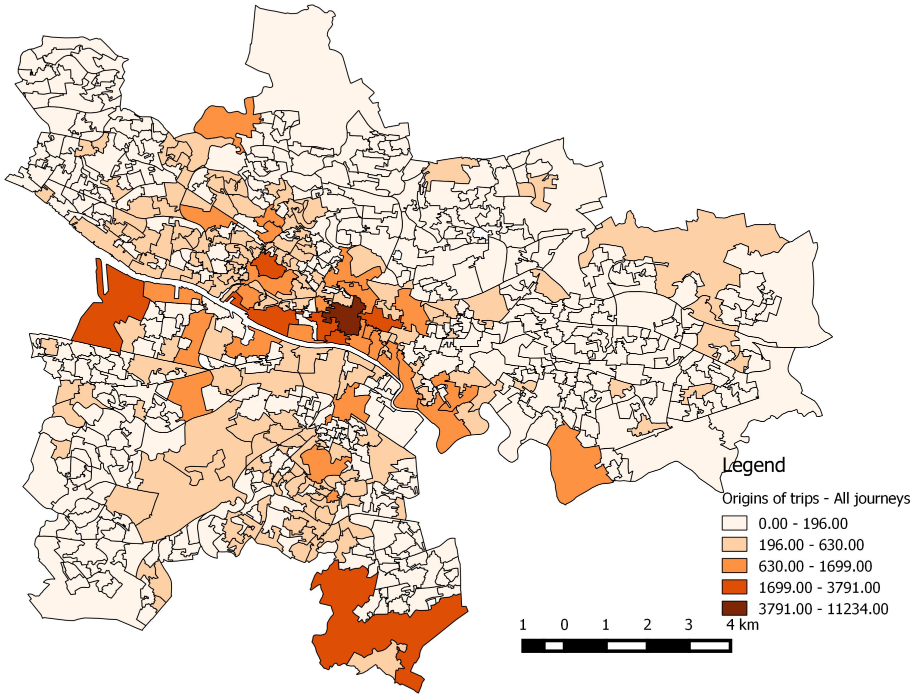

Figure 9: spatial distribution of starting points of all bicycle journeys (variable 1.1)

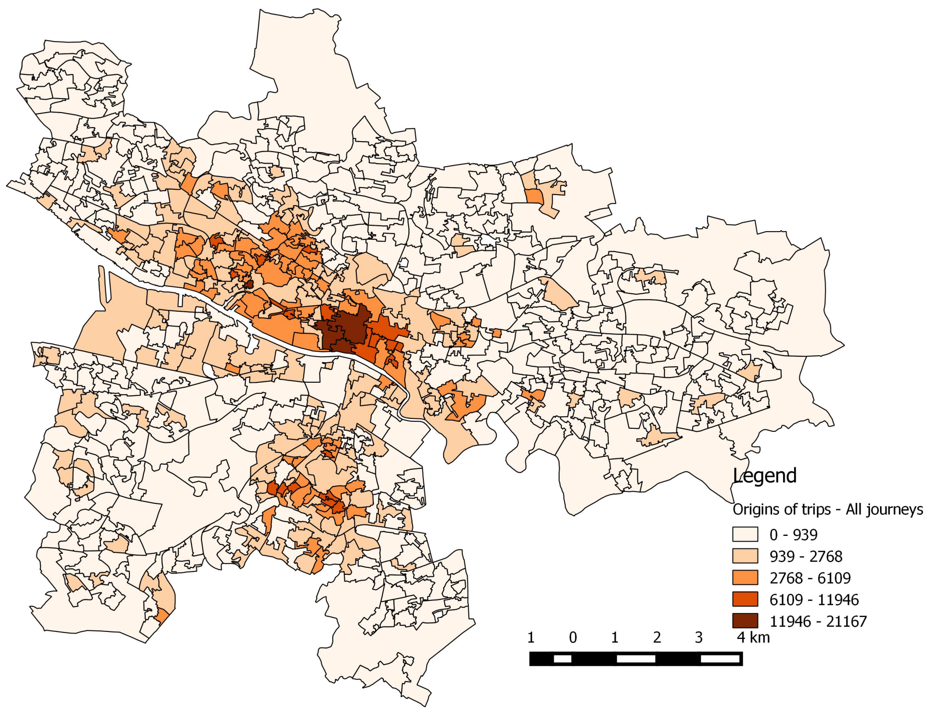

Figure 10: spatial distribution of the density of starting points of all bicycle journeys, calculated by dividing the number of journeys in a Data Zone (variable 1.1) by the area in km^2

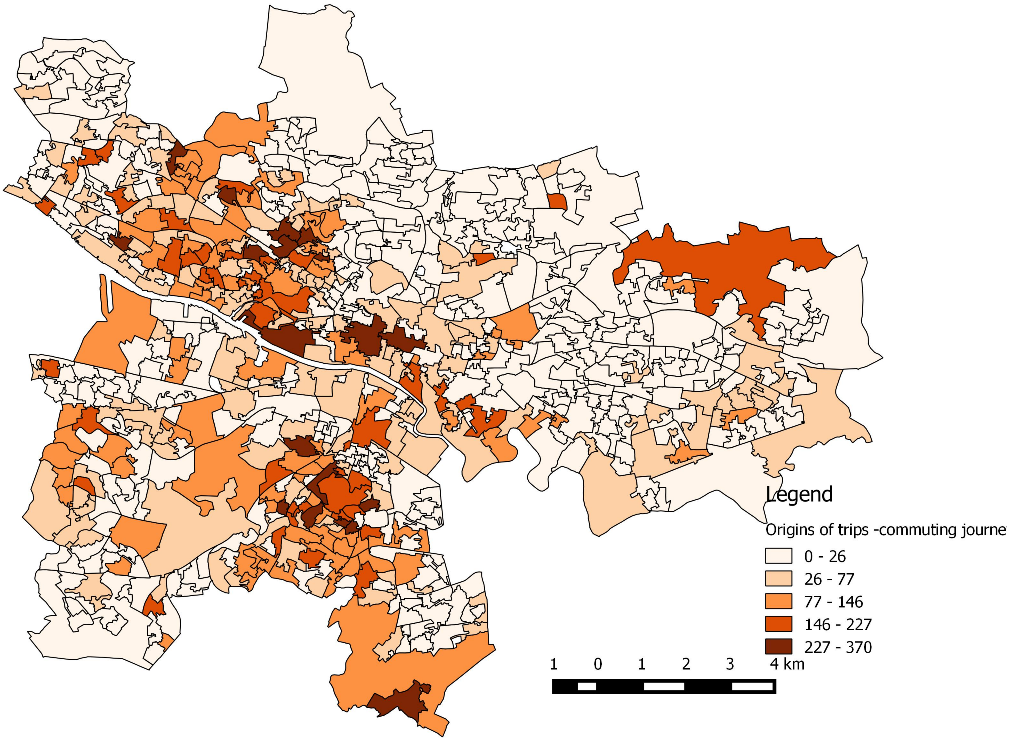

Figure 11: spatial distribution of starting points of commute journeys by bicycle (variable 1.2)

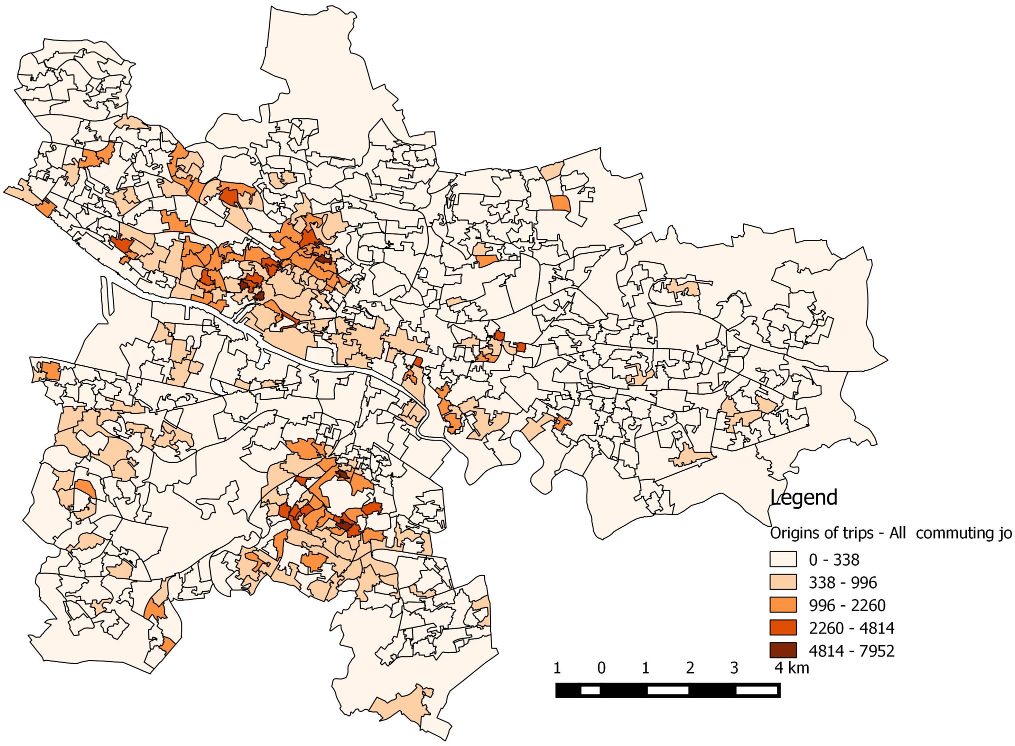

Figure 12: spatial distribution of the density of starting points of commute journeys, calculated by dividing the number of commute journeys in a Data Zone (variable 1.2) by the area in km^2

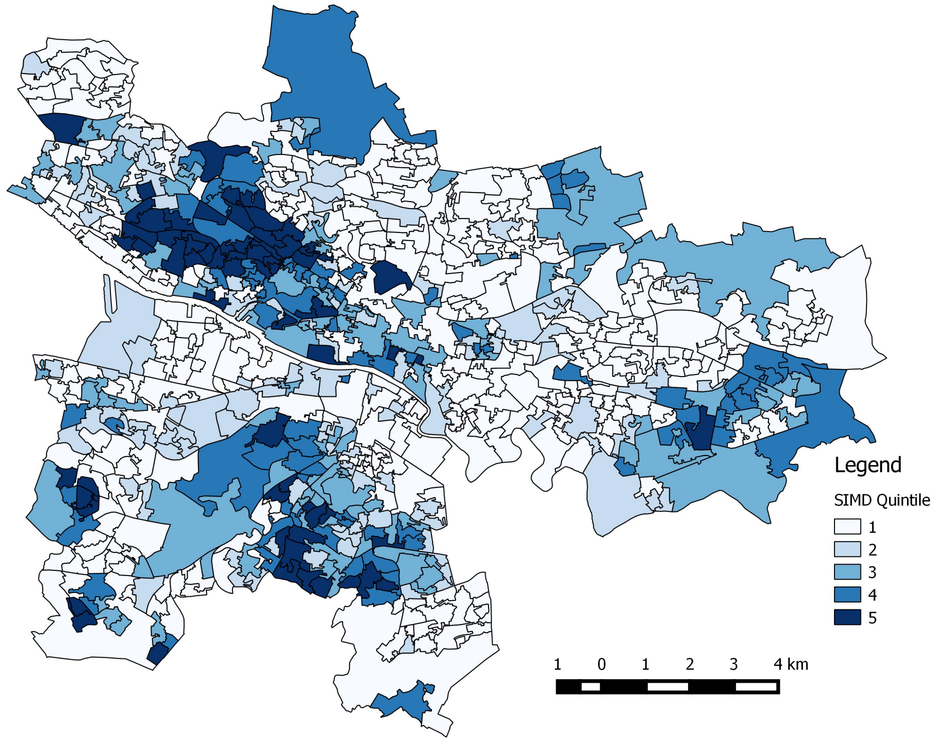

Figure 13: spatial distribution of SIMD Quintiles (variable 2)

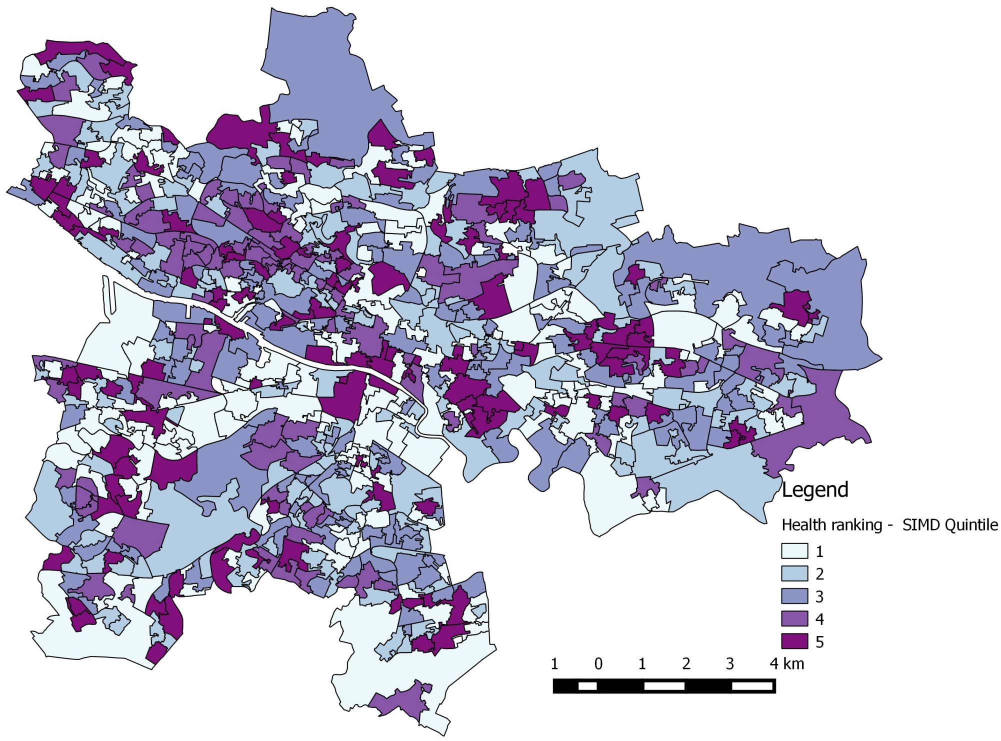

Figure 14: spatial distribution of Health Quintiles (variable 3)

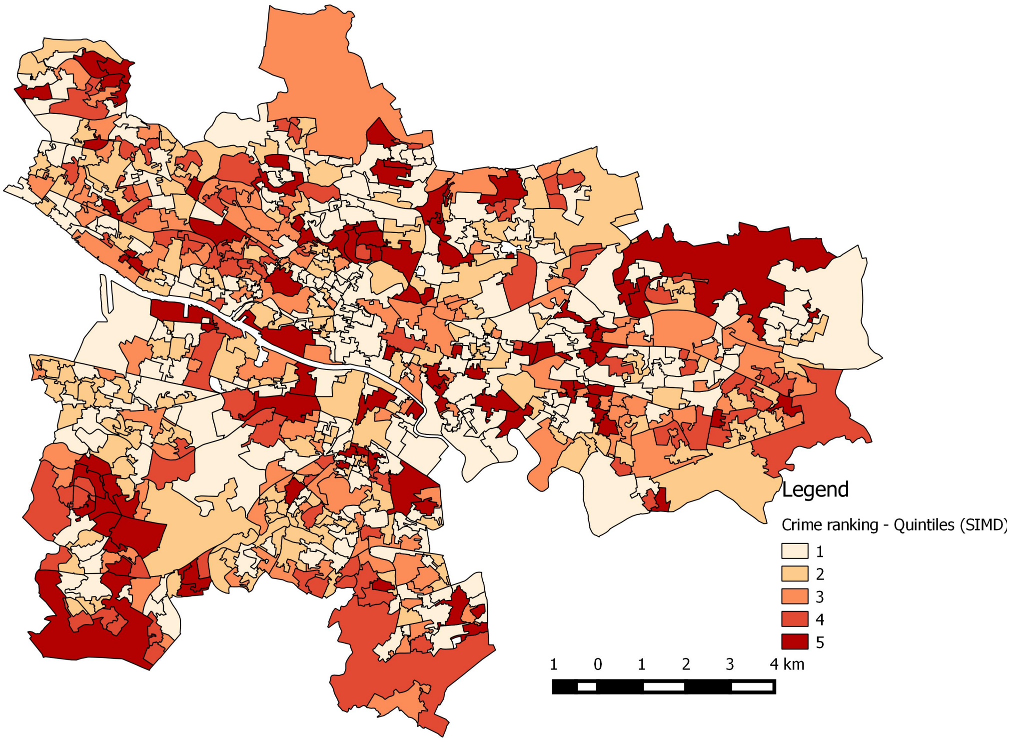

Figure 15: spatial distribution of Crime Quintiles (variable 4)

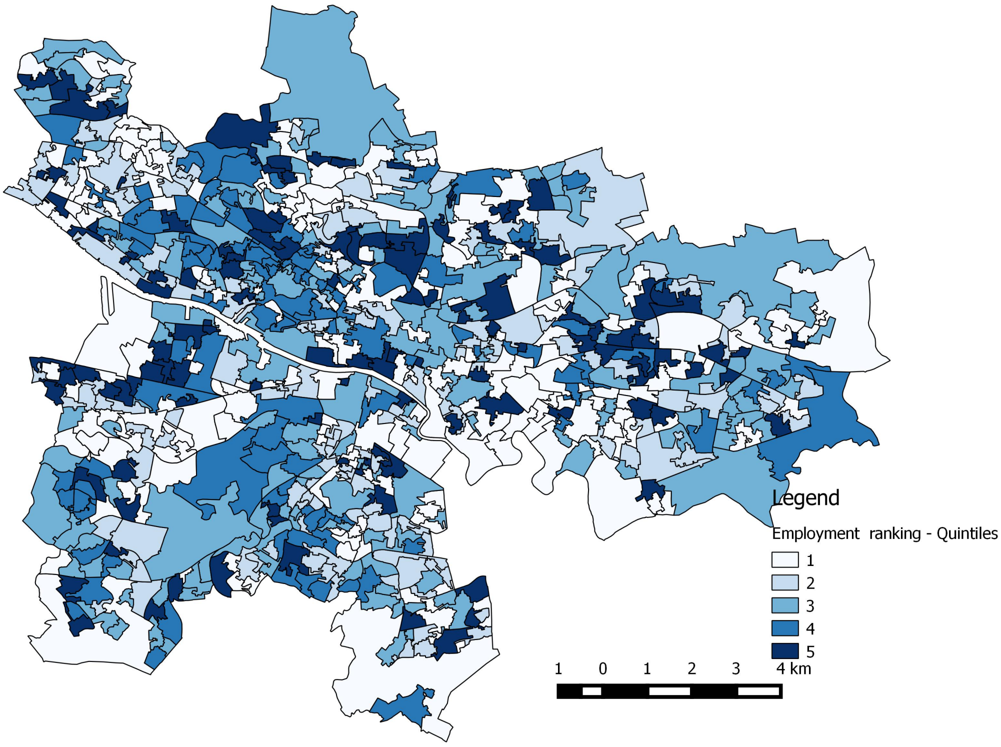

Figure 16: spatial distribution of Employment Quintiles (variable 5)

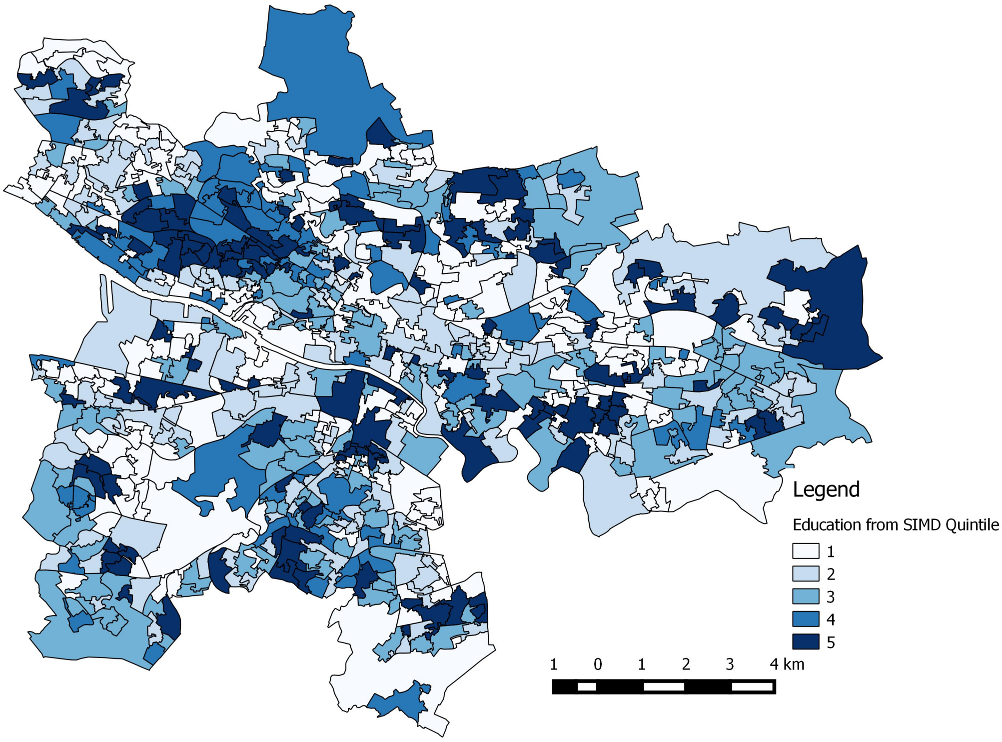

Figure 17: spatial distribution of Education Quintiles (variable 6)

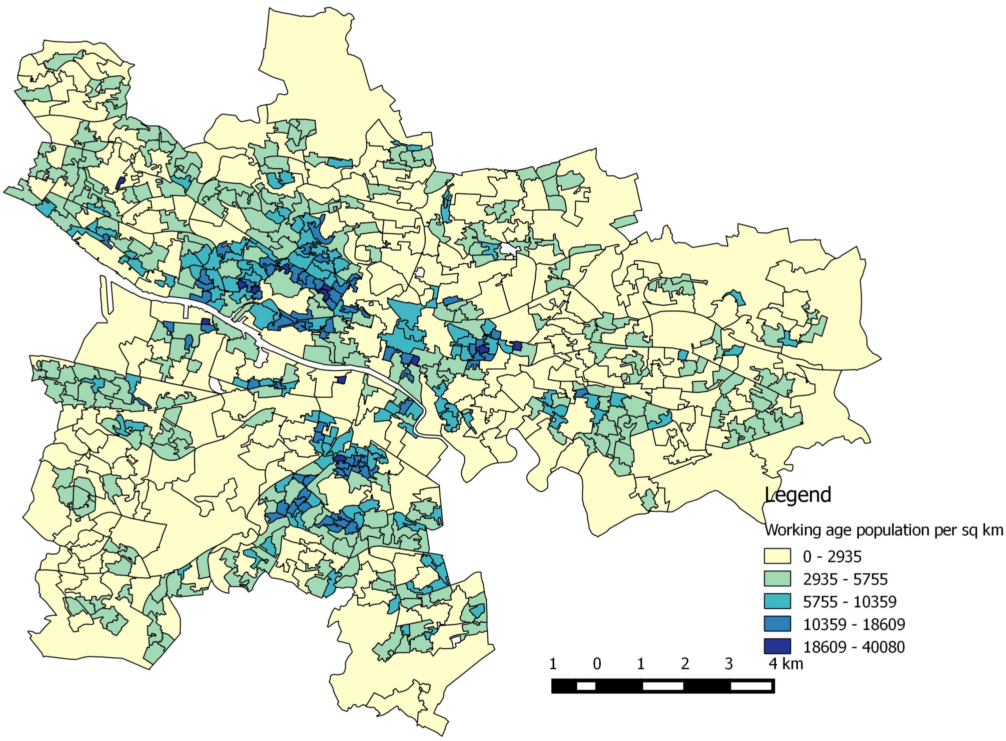

Figure 18: spatial distribution of the density of working age population (variable 7)

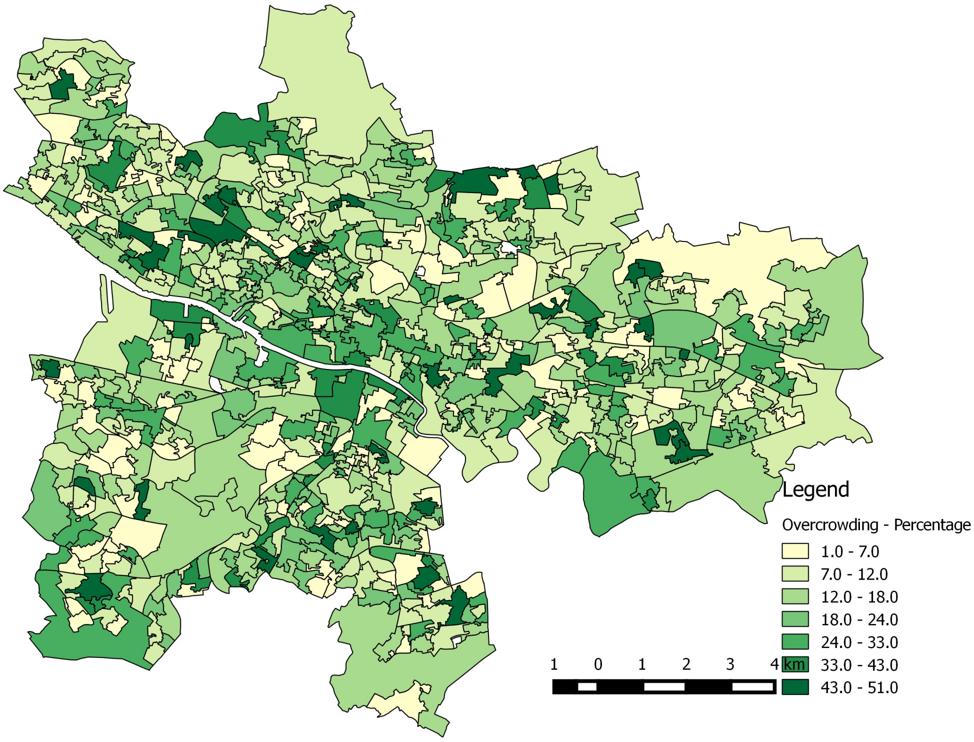

Figure 19: spatial distribution of the percentage of overcrowded housing (variable 8)

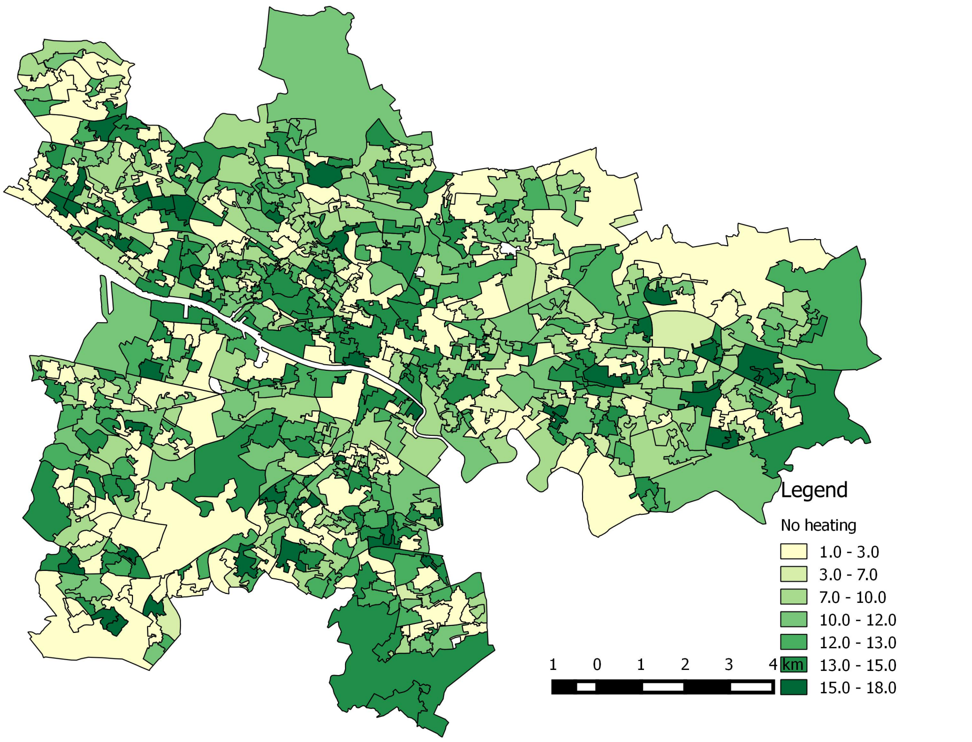

Figure 20: spatial distribution of the percentage of housing with no heating (variable 9)

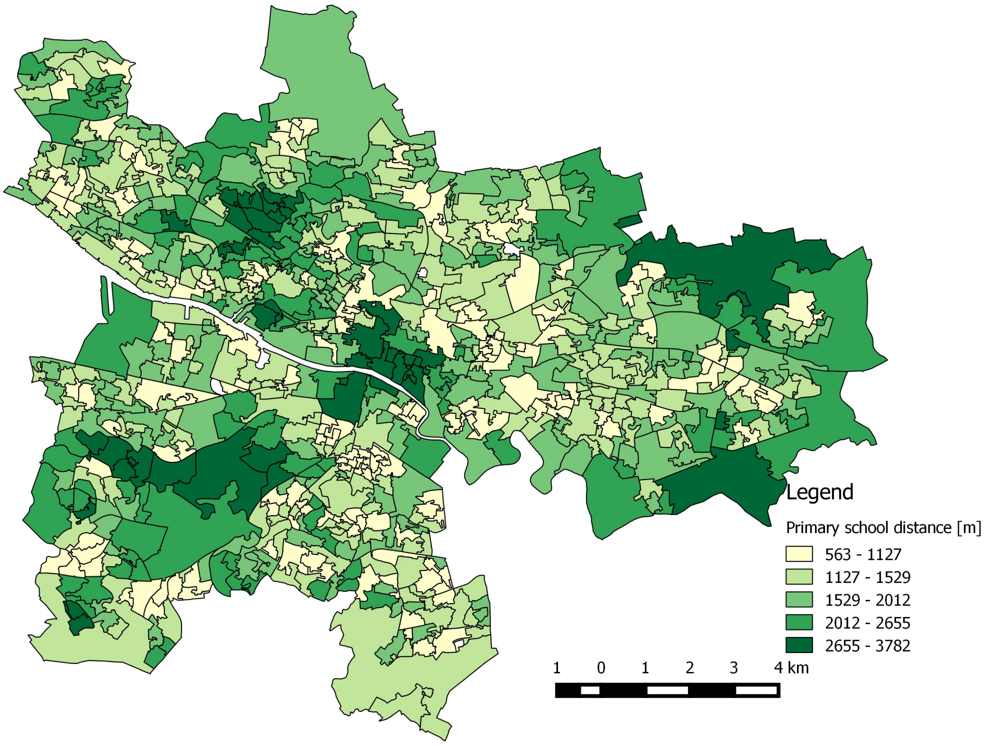

Figure 21: spatial distribution of the average distance to the nearest primary school (variable 10)

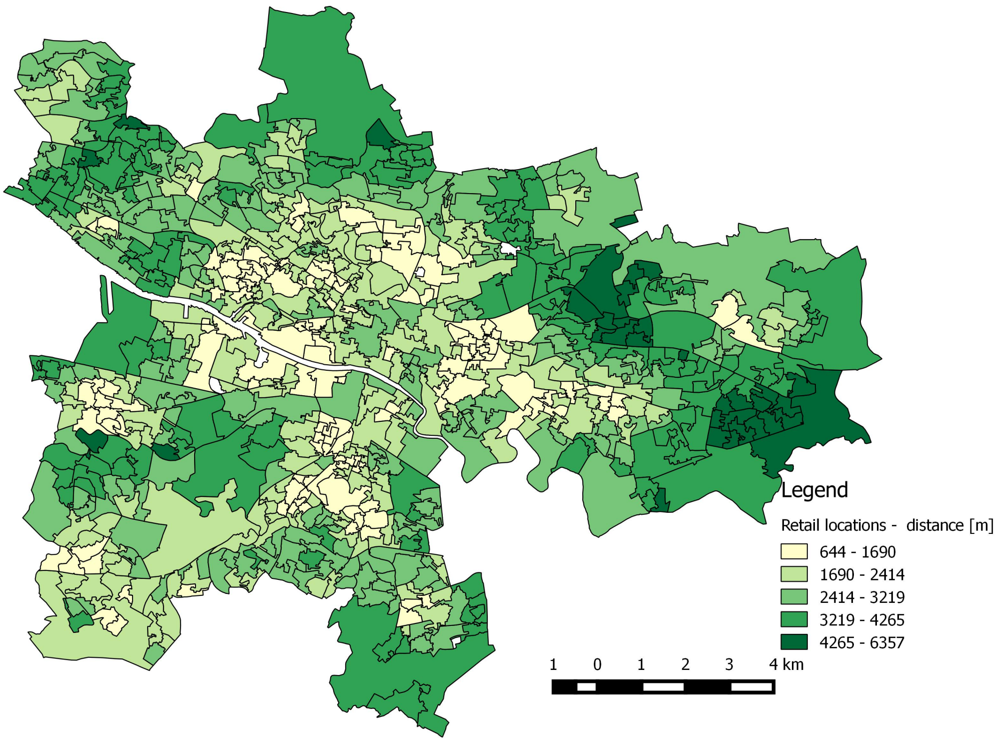

Figure 22: spatial distribution of the average distance to the nearest retail centre (variable 11)

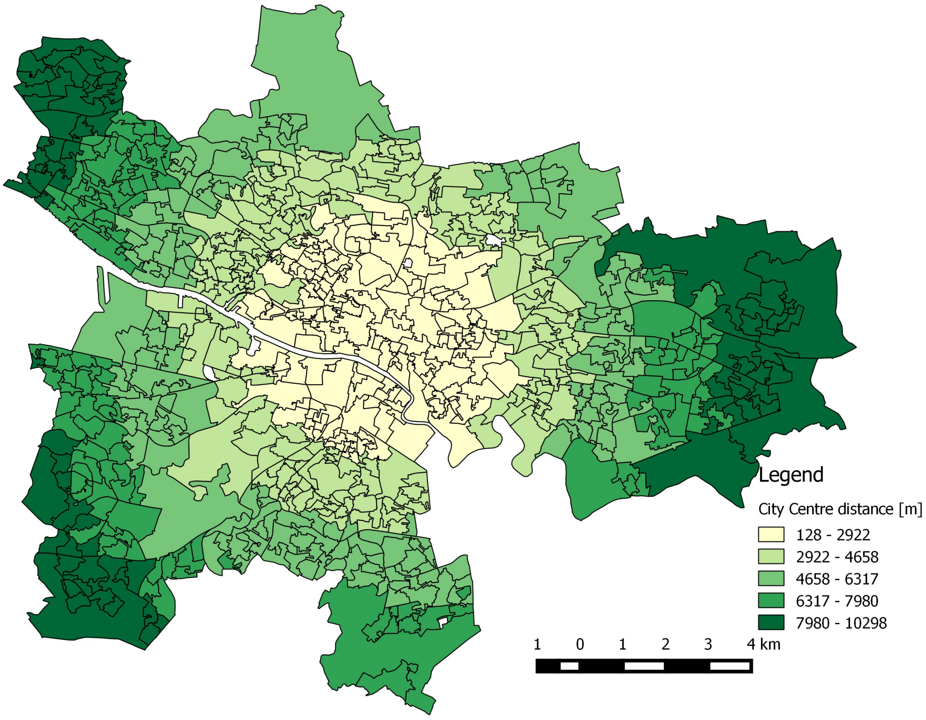

Figure 23: spatial distribution of the average distance to the city centre (variable 12)

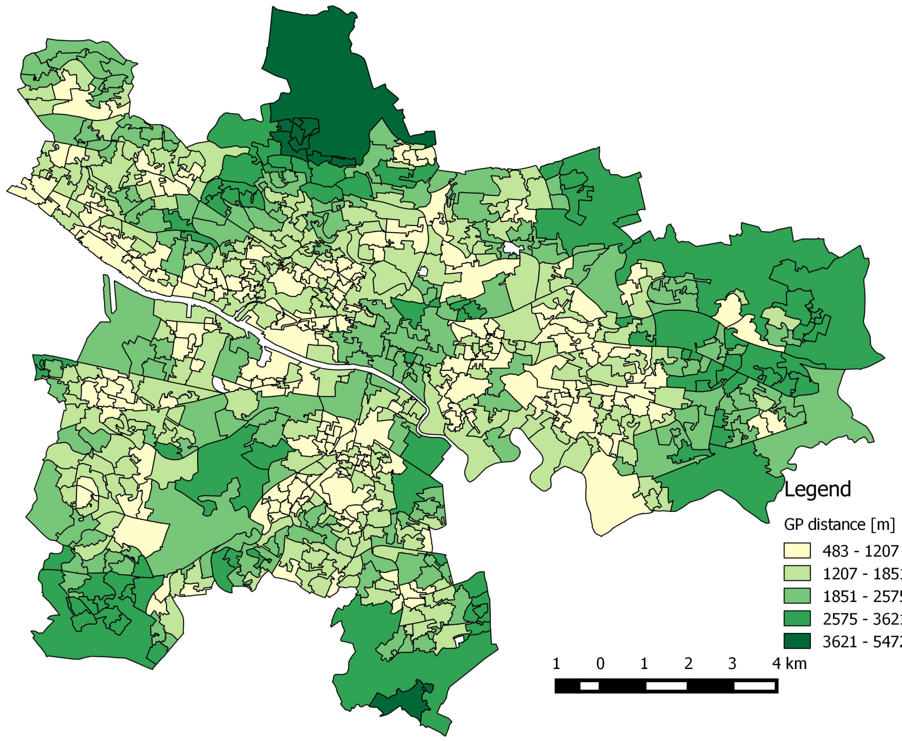

Figure 24: spatial distribution of the average distance to the nearest GP surgery (variable 13)

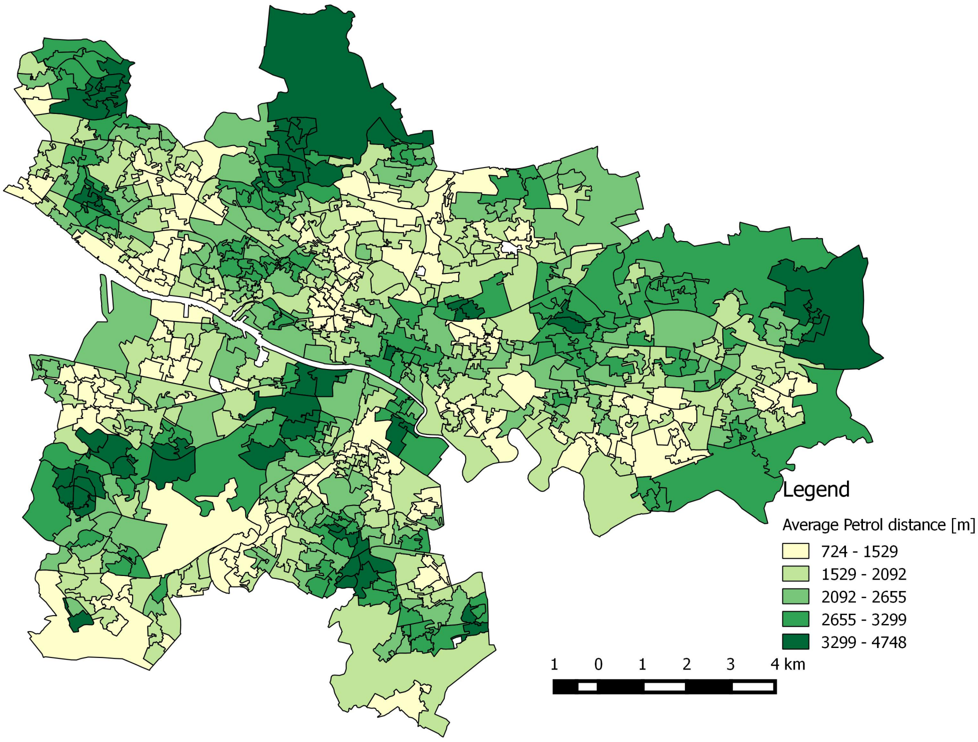

Figure 25: spatial distribution of the average distance to the nearest petrol station (variable 14)

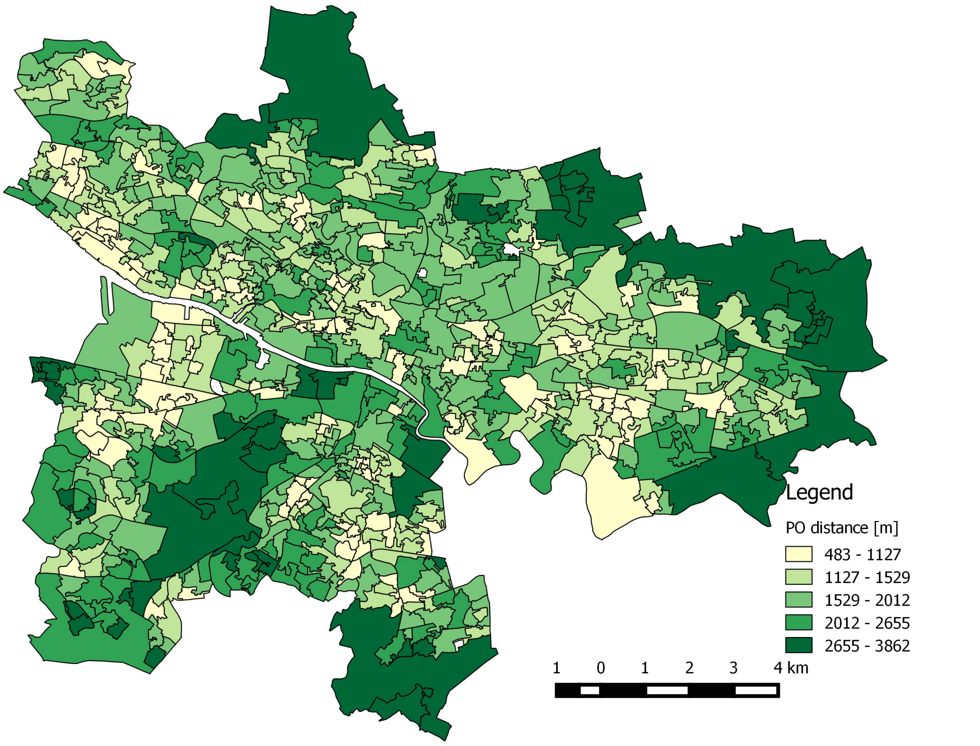

Figure 26: spatial distribution of the average distance to the nearest post office (variable 15)

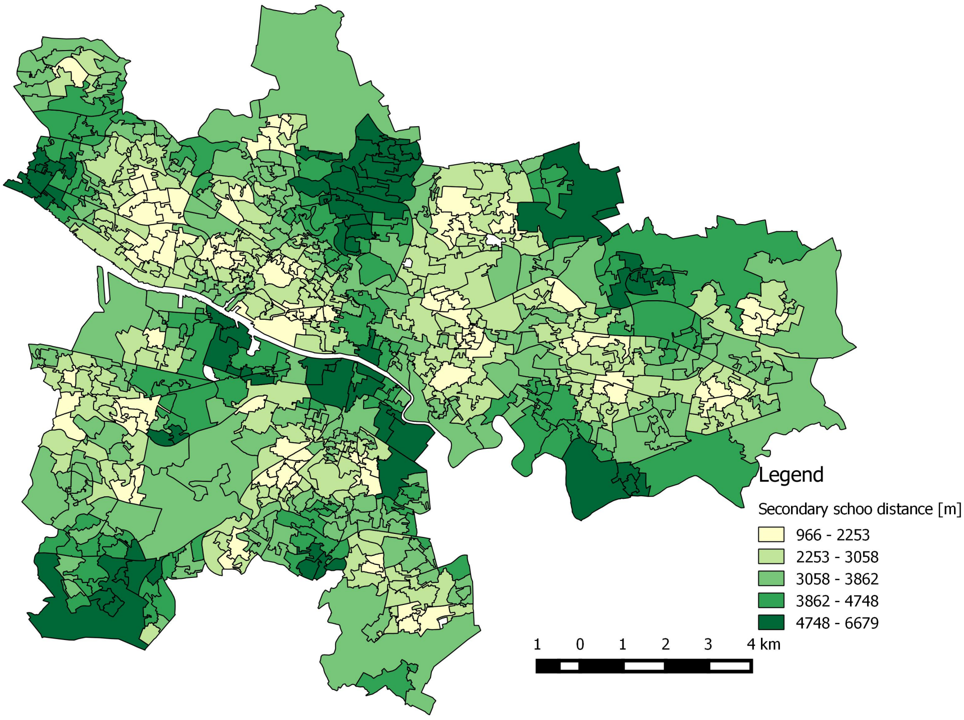

Figure 27: spatial distribution of the average distance to the nearest secondary school (variable 16)

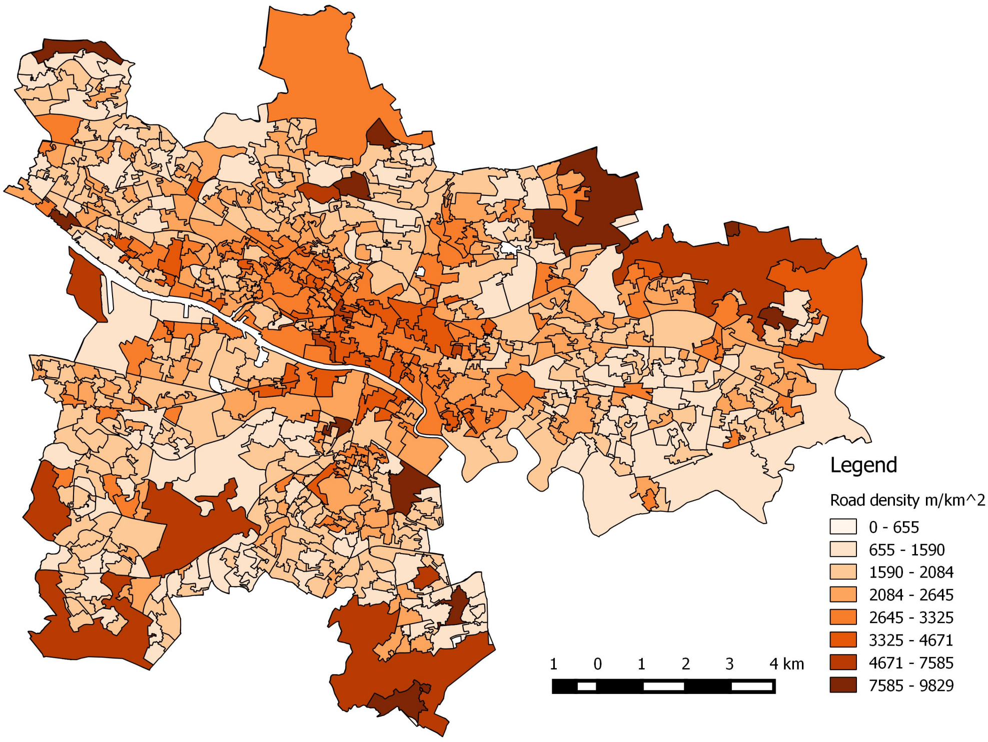

Figure 28: spatial distribution of the density of road network (variable 17)

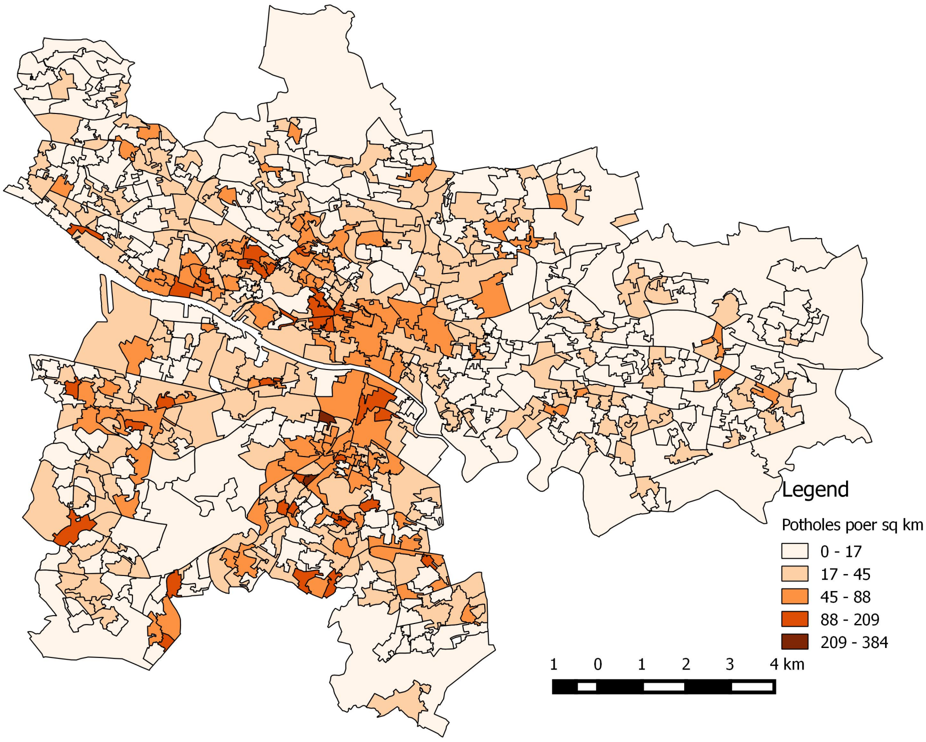

Figure 29: spatial distribution of the density of complaints made about potholes (variable 18)

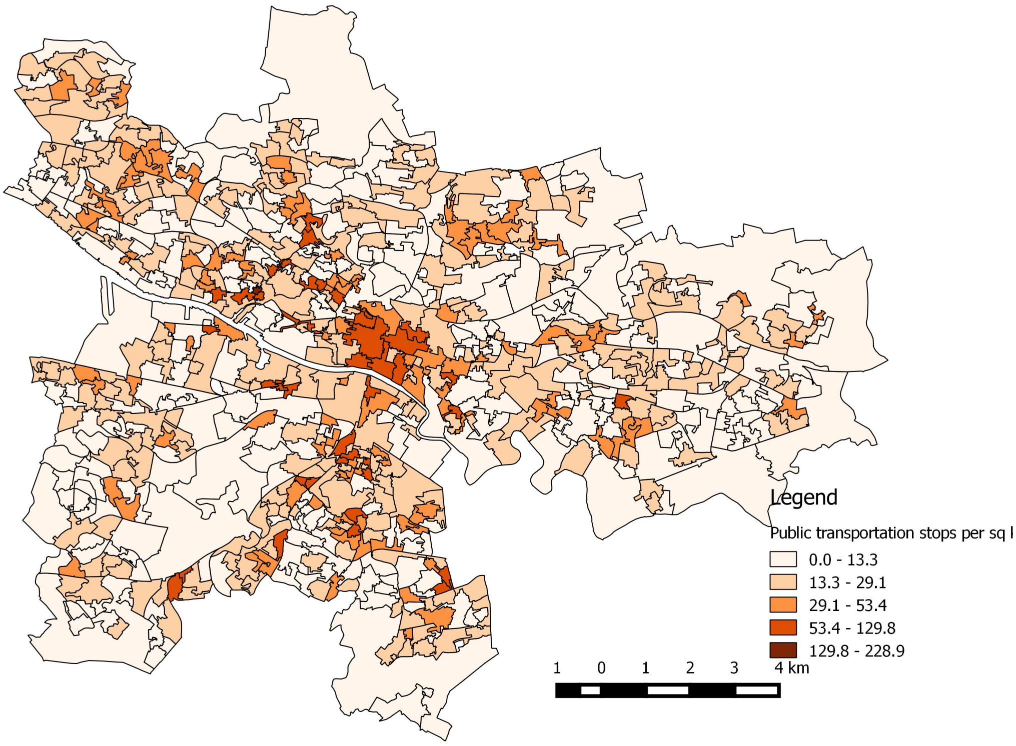

Figure 30: spatial distribution of the number of public transport stops (variable 19)

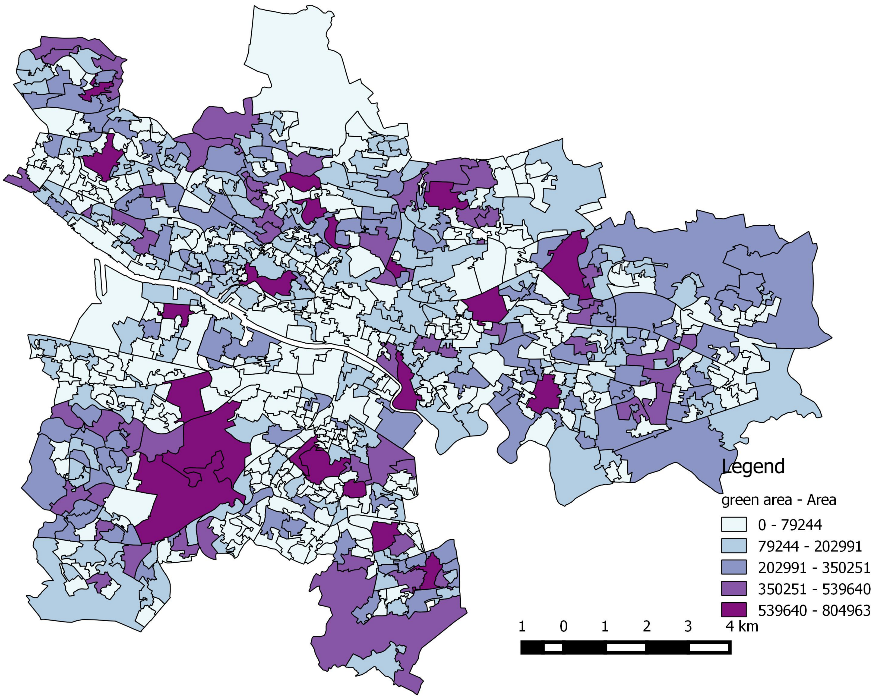

Figure 31: spatial distribution of the density of green space (variable 20)

From this initial analysis, several clear patterns emerge. A higher number of bicycle journeys are made during the summer than in the winter, which suggests that the weather affects levels of cycling. Furthermore, a significantly higher number of bicycle journeys are made between 7am–9am and 4pm–7pm, and on Monday–Friday, indicating that Strava users are cycling to commute. This indication is also given by the maps. In figure 9, which illustrates the spatial distribution of all bicycle journeys, the starting points are concentrated in the city centre of Glasgow. However, in figure 11, which only illustrates the spatial distribution of morning commute journeys, the starting points are spread more evenly throughout the city. This suggests that Strava users are living in areas throughout the city, and commuting to the city centre by bicycle. In chapter 1.2, it was mentioned that a significant number of people in Glasgow are living in deprivation. Figures 4–8 confirm this, showing that a higher number of areas in Glasgow are ranked in quintile 1 or 2: the most deprived. Figures 13–17 show that deprivation is spread throughout the city. The maps reveal several other patterns. In the city centre, the average distance to retail centres is lower, the density of road network and the number of complaints about potholes is higher. In the residential areas outwith the city centre, it is the opposite: lower density of road network, fewer public transport stops, higher density of green space, with retail centres further but schools closer.

3.5. Methodology: choice of model

As explained above, in order to analyse if one variable fluctuates according to changes in other variables, this study requires a regression model. Specifically, it requires a multiple regression model, making it possible to "explicitly control for many other factors that simultaneously affect the dependent variable" (Wooldridge, 2009: 68). As shown in table 1, the dependent variable for this study is the number of bicycle journeys; the independent variables are the level of deprivation and the other factors that might affect bicycle use.

This study involves the modelling of data that can be defined as 'count variables', according to the definition of Beaujean et al.: the lowest possible value is zero, so they can never be negative; the values appear to be positively skewed, with most values being low and few values being high (2016). Increasing the complexity further, count variables often comprise a significant number of zero values (Ibid). The data in this study also meet such a definition. As shown below in the histogram and density plot of the main dependent variable, the data are clearly not symmetrical, much less normally distributed.

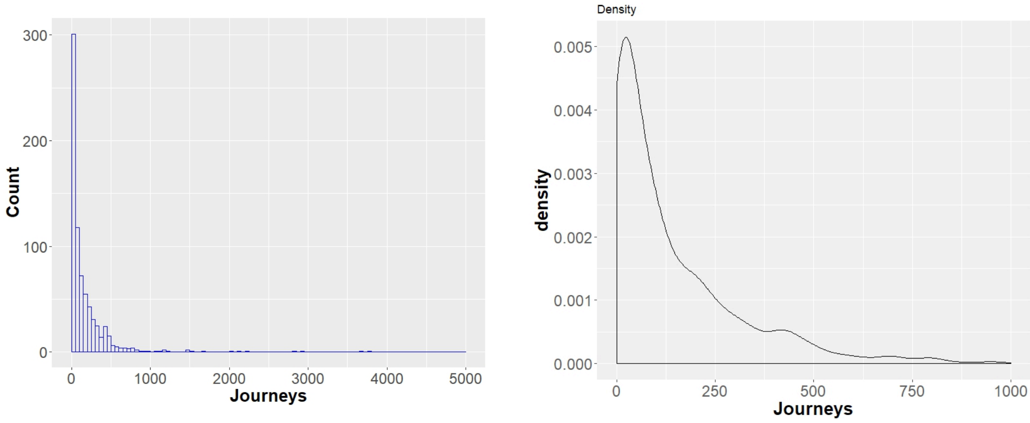

Figure 32: histogram and density plot (with eliminated outliers) of the main dependent variable: All Journeys (variable 1.1 in table 1)

This violates a major assumption of multiple regression models: that the dependent variable is normally distributed, as are the residuals, which typically follow the distribution of the dependent variable (Ibid). As shown in the histogram and density plot above, neither the dependent variable and nor, therefore, the residuals are symmetrically or normally distributed. There are clear limitations with modelling such skewed data: it can lead to inaccurate standard errors, as well as "invalid inferences and poor decisions" (Beaujean et al., 2016: 2. Karazsia et al., 2008). The skewing can be overcome through transforming the data by using the square root, but this transformation also has limitations: it makes the results harder to interpret and it does not address the significant number of zero values, predicting meaningless values such as negative ones, even though count variables can only be positive (Karazsia et al., 2008).

There are, however, several types of regression models for count variables, which can handle the dependent variable being abnormally distributed and do not require it to be transformed (Karazsia et al., 2008). The first type to consider uses the Poisson distribution. This model is relatively simple and easy to interpret, being similar to the linear regression (Ibid). In the Poisson distribution, a single parameter, λ, estimates the variance and mean of the dependent variable: as λ increases, the distribution becomes more normal; as λ decreases to zero, the distribution becomes increasingly positively skewed (Ibid. Beaujean et al., 2016). This skewing is a common feature of count variables, as shown in figure 32. However, the single parameter λ assumes that the variance and mean of the dependent variable are equal (Beaujean et al., 2016). As shown in table 2 above, however, the variance of the dependent variable in this study (714.08^2 = 509912.53) exceeds the mean (253.98). The data are therefore overdispersed, another feature of data with a significant number of zero values (Ibid). Using a Poisson distribution for overdispersed data can result in a model that doesn't have a good fit: that doesn't explain and describe the data well (Ford, 2016). Instead, it is widely acknowledged that the most appropriate type of model for overdispersed data uses the negative binomial distribution, as it does not assume that the variance and the mean of the dependent variable are equal (Beaujean et al, 2016. Cameron, 2009. Ford, 2016. Karazsia et al. 2008. Zeilis et al, 2008).

In the negative binomial distribution, the variance, υ, is modelled with an additional overdispersion parameter as θ, and the mean as λ (Beaujean et al, 2016. Ford, 2016):

$$\boldsymbol{\upsilon = \lambda + (\lambda^2 / \theta)}$$

As the overdispersion parameter increases, the variance reaches the same value as the mean and accordingly the negative binomial distribution becomes a Poisson distribution, when the variance and mean become equal (Beaujean et al, 2016. Ford, 2016. Karazsia et al. 2008). As a result, overdispersion no longer presents an issue. In comparisons between the two, it was found that negative binomial models represent count variables better than Poisson models (Karazsia et al. 2008). The conclusion was reached, therefore, that the most appropriate type of regression model to use for this study was the negative binomial.

3.6. Methodology: different models

The full dataset shown in table 1 was first cleaned: as any * or # values in the independent variables indicate missing values, they were transformed to NA (Scottish Government, 2016B). Using RStudio, the negative binomial regression models were then created with the glm.nb function of the MASS package. Two different models were created: the SIMD Quintile Model and the Separate Quintile Model. The first used the SIMD Quintile (variable 2 in table 1) as an independent deprivation variable, alongside the other independent geographical variables: Pothole Complaints Density; Transport Points; Green Spaces Density (variables 18–20). The second replaced the SIMD Quintile with the other independent deprivation variables: Health, Crime, Employment and Education Quintile; Working Age Population Density; Overcrowded Percentage; No Heating Percentage, Distance To Primary School, Secondary School, Retail Centre, City Centre, GP Surgery, Petrol Station and Post Office; Road Density (variables 3–17). The second model also featured the other geographical variables: Pothole Complaints Density; Transport Points; Green Spaces Density (variables 18–20).

These two different models were created because the SIMD Quintile in the first model is an overall ranking, based on all the separate deprivation variables in the second model: by splitting the SIMD Quintile and the Separate Quintiles between the two different models, it avoids accounting for the same variables twice. Following initial testing of the two models, several variables yielded p values significantly above 0.05 and were discounted due to low significance: No Heating Percentage, Distance to Petrol Station, Post Office and Secondary School (variables 9, 13–16). The two final models are described and analysed in chapter 4 below. In order to analyse whether the number of bicycle journeys fluctuates according to changes in the level of deprivation and other variables, the two models were initially run with the main dependent variable: All Journeys (variable 1.1). For an analysis of whether the relationship between cycling and deprivation changes over a different period, the models were then run with another dependent variable: Commute Journeys (variable 1.2).

3.7. Methodology: limitations

One significant limitation of this study is the fit of the regression models. For negative binomial regression models, there are two measures to assess the goodness of fit: how well the model explains and describes the data. The first is Akaike's Information Criterion (AIC), which balances the model's goodness of fit to the data and imposes a penalty for the model's complexity (Beaujean et al, 2016). Models with a lower AIC value have a better fit, but only relative to other models: AIC values cannot be directly interpreted in isolation (Ibid). The second measure is residual deviance, which shows how well the dependent variable is predicted by the full regression model, including all the independent variables (Lillis, 2014). The ratio of residual deviance to degrees of freedom should be approximately 1, to indicate that the model has a good fit (UCLA: Statistical Consulting Group, 2018). As described in chapter 4 below, however, both measures indicate that some regression models in this study could have a better fit. This is an issue with analysing such abnormally distributed and overdispersed data. As explained in chapter 3.5, however, careful consideration has been given to choosing the most appropriate type of regression model, for delivering the most robust and reliable set of results.

More generally, this study has the same limitations as any study employing Strava Metro data to analyse bicycle use, as explained in chapter 2.5. Although several studies have found a significant association in the number of cycle journeys between Strava Metro data and data gathered with conventional methods, questions persist over its representativeness. Such questions are especially pertinent in this study, which focuses on deprivation. Journeys can only be tracked on the Strava app by cyclists with certain technological resources, such as smartphones and GPS devices, and thus with certain financial resources, as the technology is not inexpensive. If the assumption is made that there are fewer of these well-resourced cyclists in deprived areas of Glasgow, then the limitations of this study are clear. It can explore levels of Strava use, but the number of Strava users in deprived areas might not have a significant association with the overall number of cyclists in these areas, given the assumption that many of them will lack the resources to use the Strava app.

However, this study is certainly more extensive than the previous study that analysed the relationship between cycling and deprivation in Glasgow, which solely used data from Scotland's Census in 2011 and did not take into account other factors in an area that might affect bicycle use (Glasgow Centre For Population Health, 2017). Furthermore, just as this study shares the limitations of other studies that employ Strava Metro data, it also shares their strengths: analysing bicycle use with a high level of spatial and temporal detail. For this study, the spatial detail is essential: making it possible for the number of bicycle journeys starting in an area to be related to other data on that area, including the level of deprivation. In this respect, despite its limitations, the study is, to use a word that features in several studies employing Strava Metro data, "unprecedented".

4. Results and Analysis

4.1. All Journeys

As explained in chapter 3.6, two negative binomial regressions models were created, the first using the SIMD Quintile as the independent deprivation variable and the second using the Separate Quintiles. They were initially run with the dependent variable as All Journeys: all bicycle journeys uploaded to Strava in 2016, starting within the Glasgow City Council area. The two models produced the results below:

Independent Variable	**Estimate**	**Standard Error**	**Pr(>|z|)**	**Significance**
(Intercept)	3.920	0.094	0.000	***
Reference SIMD Quintile 1				
SIMD Quintile 2	0.541	0.125	0.000	***
SIMD Quintile 3	0.917	0.135	0.000	***
SIMD Quintile 4	1.232	0.146	0.000	***
SIMD Quintile 5	1.569	0.159	0.000	***
Pothole Comps Density	0.004	0.001	0.003	**
Transport Points	0.198	0.014	0.000	***
Green Area Density	0.000	0.000	0.372	

AIC: **9115.6 /** Residual Deviation Ratio: 894.72 on 738 Degrees Of Freedom = **1.21236**

Table 3: results of the SIMD Quintile model, with dependent variable 1.1: All Journeys

Independent Variable	Estimate	Standard Error	Pr(>|z|)	Significance
(Intercept)	4.442	0.316	0.000	***
Reference Health Quintile 1				
Health Quintile 2	0.018	0.136	0.892	
Health Quintile 3	-0.153	0.139	0.270	
Health Quintile 4	0.095	0.170	0.574	
Health Quintile 5	0.022	0.144	0.881	
Reference Crime Quintile 1				
Crime Quintile 2	0.209	0.116	0.072	.
Crime Quintile 3	0.078	0.130	0.550	
Crime Quintile 4	0.194	0.149	0.191	
Crime Quintile 5	-0.288	0.145	0.047	*
Reference Employment Quintile 1				
Employment Quintile 2	0.079	0.138	0.565	
Employment Quintile 3	0.179	0.140	0.200	
Employment Quintile 4	0.473	0.169	0.005	**
Employment Quintile 5	-0.287	0.147	0.051	.
Reference Education Quintile 1				
Education Quintile 2	0.538	0.126	0.000	***
Education Quintile 3	0.466	0.132	0.000	***
Education Quintile 4	0.591	0.164	0.000	***
Education Quintile 5	0.212	0.136	0.117	
Working Age Pop Density	0.000	0.000	0.036	*
Overcrowded Percentage	0.003	0.004	0.406	
Distance To Primary Sch	0.000	0.000	0.000	***
Distance To Retail Centre	0.000	0.000	0.002	**
Distance To City Centre	0.000	0.000	0.000	***
Road Density	0.000	0.000	0.002	**
Pothole Comps Density	0.003	0.001	0.040	*
Transport Points	0.133	0.015	0.000	***
Green Area Density	0.000	0.000	0.437	

AIC: **9097.9 /** Residual Deviation Ratio: 889.29 on 720 Degrees Of Freedom = **1.23512**

Table 4: results of the Separate Quintile model, with dependent variable 1.1: All Journeys

In the SIMD Quintile model, the only statistically insignificant variable is the density of green area. All other variables in the first model are statistically significant and positively associated with levels of cycling, including Transport Points: as the number of bus, train and subway stops increases, so too does the number of all journeys. At first this seems surprising. More bicycle journeys might be expected in areas with fewer public transport stops and fewer transport options for residents, who are accordingly prompted to cycle. However, the maps in chapter 3.2 might offer an explanation for this: figure 9 shows that the starting points of all journeys are concentrated in the city centre, where there is also, according to figure 30, the highest concentration of public transport stops. Also surprising is the positive association between levels of cycling and numbers of complaints about potholes: cyclists might be expected to use better quality roads. However, it should be noted that the number of complaints per se might not provide a clear indication about the quality of roads: it might relate to other factors, such the number of road users and the level of their engagement with their environment. It should also be noted that, given the impact of potholes on bicycles, a significant number of the complaints might be coming from cyclists themselves.

In the SIMD Quintile Model, the overall deprivation variables are all statistically significant and all positively associated with levels of cycling: the number of all journeys increases, as the level of deprivation decreases. The highest level of increase occurs in Quintile 5, the least deprived areas. The Separate Quintile Model reveals more complex relationships between cycling and deprivation. There is a positive association with the number of all journeys throughout the Health and Crime Quintiles, apart from Health Quintile 3 and Crime Quintile 5. However, in the Health and Crime Quintiles, there is no clear pattern in the level of increase in all journeys. This is unlike in the SIMD Model, where the level of increase rises (from 0.541, to 0.917, to 1.232) as the level of deprivation decreases. It should also be noted that the Health and Crime Quintiles are not statistically significant. The Employment Quintile, however, is more significant and it displays a clearer pattern. Employment Quintiles 2–4 have a positive association with levels of cycling: as in the SIMD Model, the level of increase in all journeys rises (from 0.079, to 0.179, to 0.473) as the level of employment deprivation decreases. In Employment Quintile 5, however, there is a negative association: in the least employment deprived areas, the number of all journeys decreases. It should be noted, however, that only Employment Quintile 4 is statistically significant. Education Quintiles 2–4 are also statistically significant, making it the most significant of all the deprivation variables. In the Education Quintiles, there is a positive association with levels of cycling, but the level of increase in all journeys stays broadly the same across Quintiles 2, 3 and 4 (at 0.538, 0.466 and 0.591). Showing the same pattern as the Employment Quintile, the level of increase then falls in Education Quintile 5 (to 0.212): in the least education deprived areas, the number of all journeys still increases, but at a much lower level. Potential explanations for these patterns are explored below.

In the Separate Quintile Model, the percentage of overcrowded housing and the density of green area are statistically insignificant. The number of complaints about potholes and the number of public transport stops are positively associated with levels of cycling: as in the SIMD Model, and with the same potential explanations. Furthermore, the distance to primary school, retail centre and city centre are all statistically significant, but they do not have a positive or negative association with levels of cycling.

4.2. Commute Journeys

The two models were then run with the dependent variable as Commute Journeys: bicycle journeys uploaded to Strava in 2016, starting within the Glasgow City Council area, between 6am–10am on Monday–Friday. These results were produced:

Independent Variable	Estimate	Standard Error	Pr(>|z|)	Significance
(Intercept)	2.997	0.110	0.000	***
Reference SIMD Quintile 1				
SIMD Quintile 2	0.799	0.147	0.000	***
SIMD Quintile 3	1.143	0.158	0.000	***
SIMD Quintile 4	1.433	0.171	0.000	***
SIMD Quintile 5	1.808	0.186	0.000	***
Pothole Comps Density	0.003	0.002	0.069	.
Transport Points	0.068	0.016	0.000	***
Green Area Density	0.000	0.000	0.822	

AIC: **7146.6 /** Residual Deviation Ratio: 899.45 on 738 Degrees Of Freedom = **1.21877**

Table 5: results of the SIMD Quintile model, with dependent variable 1.2: Commute Journeys

Independent Variable	Estimate	Standard Error	Pr(>|z|)	Significance
(Intercept)	3.234	0.384	0.000	***
Reference Health Quintile 1				
Health Quintile 2	0.132	0.165	0.426	
Health Quintile 3	0.114	0.169	0.502	
Health Quintile 4	0.508	0.206	0.014	*
Health Quintile 5	0.189	0.175	0.282	
Reference Crime Quintile 1				
Crime Quintile 2	0.417	0.141	0.003	**
Crime Quintile 3	0.180	0.158	0.255	
Crime Quintile 4	0.314	0.181	0.082	.
Crime Quintile 5	-0.342	0.177	0.053	.
Reference Employment Quintile 1				
Employment Quintile 2	0.134	0.168	0.426	
Employment Quintile 3	0.102	0.170	0.549	
Employment Quintile 4	0.268	0.206	0.192	
Employment Quintile 5	-0.506	0.179	0.005	**
Reference Education Quintile 1				
Education Quintile 2	0.402	0.154	0.009	**
Education Quintile 3	0.372	0.161	0.021	*
Education Quintile 4	0.513	0.199	0.010	*
Education Quintile 5	0.066	0.165	0.688	
Working Age Pop Density	0.000	0.000	0.892	
Overcrowded Percentage	0.003	0.005	0.505	
Distance To Primary Sch	0.000	0.000	0.001	**
Distance To Retail Centre	0.000	0.000	0.000	***
Distance To City Centre	0.000	0.000	0.074	.
Road Density	0.000	0.000	0.038	*
Pothole Comps Density	0.003	0.002	0.118	
Transport Points	0.028	0.018	0.117	
Green Area Density	0.000	0.000	0.098	.

AIC: **7198.3 /** Residual Deviation Ratio: 900.4 on 720 Degrees Of Freedom = **1.2505**

Table 6: results of the Separate Quintile model, with dependent variable 1.2: Commute Journeys

Compared to the models for all journeys, these two models have a better fit: the AIC values are lower, and the Residual Deviation Ratios are closer to 1. This shows that the models might offer a better and clearer explanation of the relationship between deprivation and cycling. Broadly, the models for commute journeys show similar patterns to those for all journeys. In both the SIMD Quintile Model and the Separate Quintiles Model, the density of green area is statistically insignificant. The density of working age population is statistically insignificant for commute journeys, compared with being moderately significant for all journeys. This is surprising: a clearer relationship might be expected between the number of residents of working age, and the number of residents cycling to a place of work. The number of pothole complaints is also less statistically significant for commute journeys than all journeys, although there is still a positive association: the number of commute journeys increases, albeit at a lower level than for all journeys, as the number of complaints increases. For commute journeys, the number of public transport stops has a positive association with levels of cycling but it is only statistically significant for the SIMD Quintile Model, not the Separate Quintile Model: a difference to all journeys, when it is statistically significant for both models. Another difference between commute and all journeys is the level of increase in journeys as the public transport stops increase: commute journeys increase at a lower level than all journeys (0.068 and 0.28 compared to 0.198 and 0.133). This might suggest that if there are more public transport options available for residents, they are more likely to use these options and commute via public transport, than commute by bicycle. It is also interesting that the distance to the city centre has a lower statistical significance for commute journeys than for all journeys (. compared with ***). This could be explained, as shown in figures 9 and 11, by the higher number of all journeys actually starting in the city centre.

In terms of deprivation, the SIMD Quintile Model for commute journeys is similar to all journeys. The overall deprivation variables are positively associated with levels of cycling: the number of commute journeys increases, as the level of deprivation decreases. Indeed, the level of this increase is noticeably higher than for all journeys: 1.808 and 1.433 for SIMD Quintiles 4 and 5 for commute journeys; 1.569 and 1.232 for SIMD Quintiles 4 and 5 for all journeys. This is a different conclusion to the one reached by several studies described in chapter 2.2: that levels of commuting by bicycle are higher among people from less affluent households (Beenackers et al., 2012. Heinen et al.. 2010. Krizek et al., 2009. Pistoll et al., 2014). For the Separate Quintile Model for commute journeys, again the patterns are similar to the model for all journeys. Aside from Health Quintile 4, it is noticeable that the Health Quintile is not statistically significant for commute journeys: neither was it significant for all journeys, as explained above. This is surprising, given the widely acknowledged health benefits of cycling, as described in chapter 1.2. An explanation might be offered by the information on which the health deprivation ranking is based. It does not take into account levels of obesity, diabetes and hypertension and cardiorespiratory fitness, which cycling can help to reduce. Rather, the health deprivation ranking is solely based on information such as the number of hospital stays in area related to alcohol and drug misuse, the proportion of people being prescribed drugs for anxiety, depression and psychosis and the proportion of births of low weight (Scottish Government, 2016A): none of which, it is believed, cycling can help to reduce.

For commute journeys, as with all journeys, the most statistically significant deprivation variable is the Education Quintile. Despite this significance, it is difficult to interpret. For both commute journeys and all journeys, the pattern is the same. There is a positive association between levels of cycling and Education Quintiles 2–4, and the level of increase in bicycle journeys stays broadly the same across these quintiles: unlike the overall SIMD Quintile, it does not rise as the level of education deprivation decreases. It is therefore difficult to assess the exact relationship between these Education Quintiles and the number of journeys. Again, an explanation for this might be offered by the information on which the education deprivation ranking is based. Aside from the proportion of working age population with no qualifications, it only takes into account information related to young people: attendance and attainment of school leavers in an area; the number of young people aged 17–21 in higher education; the number of young people aged 16–19 not in education, employment or training. Several studies have found that young people are underrepresented in Strava Metro data, however, as described in chapter 2.6 (Griffin et al., 2015. Heesch et al., 2016). This might explain the difficulty with analysing the relationship between the number of journeys and level of education deprivation: the aspect of deprivation is based on information about young people, who might be underrepresented in the cycling data. Indeed, the representativeness of Strava Metro data is one of the limitations of this study, as discussed in more detail in chapters 2.5 and 3.7.

Yet despite questions over the representativeness of the data used, this study does reach similar conclusions to several others discussed in chapter 2, which do not use Strava Metro data. One study found that across the USA, levels of cycling are significantly lower in the richest quartile than in the poorest (Flanagan et al., 2016). In this study, for both all journeys and commute journeys, there is a negative association between the number of bicycle journeys and Crime and Employment Quintiles 5, yet a positive association in all the other Crime and Employment Quintiles: i.e. in the least crime and employment deprived areas, the number of journeys decreases. One potential explanation might be the higher age of residents of these areas. The employment deprivation ranking is based solely on information related to working age: the number of working age people in an area claiming unemployment, incapacity or disability benefits (Scottish Government, 2016A). It might be assumed, therefore, that areas with fewer residents of working age have lower levels of employment deprivation: as residents who are not of working age cannot be defined as employment deprived. It might also be assumed that the age of residents is higher in areas with lower levels of crime deprivation. As older people might be less likely to cycle, this offers a potential explanation for the negative association between the number of bicycle journeys and Crime and Employment Quintiles 5, but further analysis is clearly needed to explore the assumptions.

More broadly, the results point to the same conclusion as the majority of other national and international research discussed in the literature review. For both all journeys and commute journeys, the overall deprivation variables in the SIMD Quintile Models are positively associated with levels of cycling: the number of bicycle journeys increases, as the level of deprivation decreases. The Separate Quintile Models reveal the complexity of relationships between cycling and deprivation, leading to the conclusion that further research is needed. Yet even such a limited conclusion could not be drawn without analysing the relationship between cycling and separate aspects of deprivation, as this study has done: unlike the majority of other studies in the literature review, which have focused solely on the relationship between affluence and cycling. It is hoped, therefore, that this study has made a small contribution to increasing the understanding of cycling in Glasgow.

5. Conclusion

5.1. Summary

This study set out to analyse the relationship between the number of bicycle journeys originating in an area of Glasgow and the level of deprivation of that area. It did so using Strava Metro and SIMD data, which raised questions over representativeness but allowed for detailed temporal and spatial analysis. It found that there does appear to be a relationship between cycling and deprivation in the Glasgow City Council area: the number of bicycle journeys increases, as the level of overall deprivation decreases. This positive association is present for all journeys over the course of 2016, and it especially significant for morning commute journeys: starting between 6am–10am, on Monday–Friday. The comprehensive data also made it possible for this study to analyse the relationships between levels of cycling and separate aspects of deprivation. These relationships were found to be complex. Health deprivation is statistically insignificant, perhaps because it is ranked using information that does not relate to the health benefits of cycling. Education deprivation, by contrast, is statistically significant, but it is difficult to assess the exact relationship, perhaps because it is ranked using information that only relates to young people: who might be underrepresented in the Strava Metro data. For most aspects of deprivation, the number of bicycle journeys increases as the level of deprivation decreases, but there are two noticeable exceptions: in the least crime and employment deprived areas, the number of journeys actually decreases. This might be explained by the higher age of the residents of these areas, but further analysis is clearly needed. Despite its limitations, as explained in detail in chapters 3.7 and 4.2, it is therefore hoped that this study has made a small contribution towards answering the question posed in the introduction: who in Glasgow is not cycling? It has also identified some steps that could be taken to encourage them to start cycling.

5.2. Targeted actions

As shown in chapter 4.2, there is an especially clear relationship between deprivation and commuting by bicycle: as the level of deprivation decreases, the number of commute journeys increases. Furthermore, the level of this increase is higher than for all bicycle journeys. As set out in Cycling Action Plan For Scotland and Glasgow's Strategic Plan For Cycling, concerted actions are being taken to encourage people in the city to commute by bicycle: developing cycling infrastructure in the city centre; requiring new office developments to provide showers, changing facilities and bicycle parking; promoting the Cycle To Work scheme among companies (Glasgow City Council, 2015. Scottish Government, 2010). This study found that levels of commuting by bicycle are significantly lower in the most deprived areas: perhaps because residents of these areas are less likely to be employed per se, or less likely to be employed by the type of companies that offer the Cycle To Work scheme for employees and have offices with facilities for cyclists. Wider actions are therefore required to increase levels of cycling in deprived areas of Glasgow.

It is encouraging that cycling infrastructure is being incorporated into the plans to regenerate several of these areas: the City Development Plan notes that the regeneration of Sighthill in Springburn, one of the most deprived areas of the city, offers "a perfect opportunity to make the site fully accessible by bike from the outset for both commuting and leisure" (Glasgow City Council. 2016: 19). This study found that the distance to the city centre and the retail centre has a statistical significance to the number of bicycle journeys: indicating the importance of providing a network of cycling infrastructure that connects the places where people live, to the places where they shop and work. Furthermore, given that levels of commuting by bicycle are significantly lower among residents of the most deprived areas, perhaps because they are less likely to be employed per se or less likely to be employed by companies that promote cycling, it might be more effective to focus on encouraging these residents to cycle for leisure. In deprived areas of Hackney in London, organised bicycle rides and family cycling events during the weekends and school holidays have successfully increased levels of cycling, by presenting it as a fun family activity (Transport For London, 2011). The findings of this study suggest that similar actions might be successful in Glasgow.

5.3. Further research

As shown in this study, the relationship between cycling and deprivation in Glasgow is highly complex and it clearly requires further research. It might be useful to analyse the use of Nextbike, the public bicycle sharing scheme, in the context of deprivation. As mentioned in chapter 2.2, it was found that the public bicycle sharing scheme in London is used more by residents of deprived areas of the city (Ogilvie et al., 2012). As with Strava Metro data, the data on Nextbike docking stations might pertain to a specific area: making it possible to gauge the number of bicycles hired in a specific area, and then analyse the relationship between that number and the level of deprivation of the area. Indeed, given that Nextbike enables people to cycle without the financial commitment of buying and maintaining a bicycle, it might be assumed that residents of deprived areas are more likely to use Nextbike than the Strava app. This type of research could help to ensure that docking stations and cycling infrastructure are located in places where they would better encourage residents of deprived areas to cycle. However, it is important to note that developing cycling infrastructure is not sufficient in isolation. Residents of deprived areas must feel that they are able to use it: that it is 'theirs'. This is a significant barrier to the use of facilities and amenities, as other studies have found. In Glasgow, residents of a deprived area, who lived within half a mile of a public green space, asserted the belief that the nearest public green space was actually much further away: "some respondents did not feel that the local park was culturally suitable or available to them" (Macintyre et al., 2008: 912).

Detailed qualitative research is needed to better understand these barriers, and to determine the appropriate actions to overcome them. This type of research has found that residents of a deprived area in Bristol prefer to drive rather than use active travel because they see cars as providing “security, convenience and social approbation”, and they fear being harassed or attacked especially at night (Bird, 2010, p. 4). It has also found that young people in deprived areas of Liverpool avoid using a bicycle because they fear having it stolen, and that people are discouraged from cycling by their perception that it requires lycra clothing and expensive equipment (Cavill et al., 2007. Daley et al., 2011). These barriers need to be fully understood before they can be completely overcome. Only then will cycling become an everyday mode of transport for everyone in Scotland, with 10% of journeys being made by bicycle.

References

Bastian, A., Börjesson, M. 2017. *The City As A Driver Of New Mobility Patterns, Cycling And Gender Equality: Travel Behaviour Trends In Stockholm 1985–2015.* Centre For Transport Studies. Stockholm, Sweden.

Beaujean, A., Morgan, G. 2016. *Tutorial On Using Regression Models With Count Outcomes Using R.* Practical Assessment, Research & Evaluation. Volume 21, Number 2: 1–19.

Beenackers, M., Kamphuis, C., Giskes, K., Brug, J., Kunst, A. 2012. *Socioeconomic Inequalities In Occupational, Leisure-Time, And Transport Related Physical Activity Among European Adults: A Systematic Review*. International Journal Of Behavioural Nutrition & Physical Activity. Volume 9. 10.1186/1479-5868-9-116. Accessed: 13.08.18.

Bird, S. 2010. *Active Transport In Deprived Communities: Why The Car Is King.* International Nonprofit & Social Marketing Conference. Brisbane, Australia. http://eprints.uwe.ac.uk/14654. Accessed: 13.08.18.

Boss, D., Nelson, T., Winters, M., Fersterd, C. 2018. *Using Crowdsourced Data To Monitor Change In Spatial Patterns Of Bicycle Ridership.* Journal Of Transport & Health. 10.1016/j.jth.2018.02.008. Accessed: 13.08.18.

Brand, C., Goodman, A., Ogilvie, D. 2014. *Evaluating The Impacts Of New Walking And Cycling Infrastructure On Carbon Dioxide Emissions From Motorized Travel: A Controlled Longitudinal Study.* Applied Energy. Number 128: 284–295.

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. Volume 46: 1730–1740.

Cameron, C. 2009. *Count Data Regression Made Simple.* University Of California Davis. Davis, USA.

Cavill, N., Watkins, F. 2007. *Cycling And Health: An Exploratory Study Of Views About Cycling In An Area Of North Liverpool, UK.* Health Education. Volume 107, Number 5: 404–420.

Cervero, R., Duncan, M. 2003. *Walking, Bicycling, And Urban Landscapes: Evidence From The San Francisco Bay Area.* American Journal Of Public Health. Volume 93, Issue 9: 1478–1483.

Christou, G. 2016. *Assessing The Potential Of Big Mobility Data To Improve Transport Modelling For Cyclists.* European Transport Conference. University College London. London, UK.

Conrow, L., Wentz, E., Nelson, T., Pettit, C. 2018. *Comparing Spatial Patterns Of Crowdsourced And Conventional Bicycling Datasets.* Applied Geography. Volume 92: 21–30.

Corney, M. 2016. *Preparation And Analysis Of Crowdsourced GPS Bicycling Data: A Study Of Skåne, Sweden.* Masters Degree Thesis. Department Of Physical Geography And Ecosystem Science, Lund University. Lund, Sweden.

Cycling Scotland. 2017. *Annual Cycling Monitor Report.* Cycling Scotland. Edinburgh, UK.

Cycling Scotland, Living Streets Scotland, Paths For All, Sustrans Scotland, Transform Scotland. 2012. *Active Travel, Active Scotland: Our Journey To A Sustainable Future*. Cycling Scotland et al. Edinburgh, UK.

Cycling Weekly. 2018. *Increase In Scotland's Bike Use Is Start of Cycling Revolution.* https://www.cyclingweekly.com/news/latest-news/increase-scotlands-bike-use-start-cycling-revolution-371156. Accessed: 13.08.18.

Daley, M. Rissel, C. 2011. Perspectives And Images Of Cycling As A Barrier Or Facilitator Of Cycling. Transport Policy. Volume 18, Number 1: 211–216.

Department For Transport. 2017. *Public Social Attitudes Survey 2015: Attitudes Towards Public Transport.* Statistical Release. Department For Transport. London, UK.

Dill, J. 2009. *Bicycling For Transportation And Health: The Role Of Infrastructure.* Journal Of Public Health Policy. Volume 30, Supplement 1: 95–110.

Dunleavy, M. 2015. *Crowd Cycling: Understanding Cyclist Behaviour Using The Mobile Tracking App Strava.* Research Paper. University Of Dublin. Dublin, Ireland.

Ellaway, A., Macintyre, S., Kearns, A. 2001. *Perceptions Of Place And Health In Socially Contrasting Neighbourhoods.* Urban Studies. Volume 38, Number 12: 2299–2316.

Figliozzi, M., Blanc, M. 2015. *Evaluating The Use Of Crowdsourcing As A Data Collection Method For Bicycle Performance Measures And Identification Of Facility Improvement Needs.* Report FHWA-OR-RD-16-04. Oregon Department Of Transportation. Oregon, USA.

Flanagan, E., Lachapelle, U., El-Geneidy, A. 2016. *Riding Tandem: Does Cycling Infrastructure Investment Mirror Gentrification And Privilege In Portland, OR And Chicago, IL?* Research In Transportation Economics. Volume 60, Issue C: 14–24.

Ford, C. 2016. *Getting Started With Negative Binomial Regression Modeling.* University Of Virginia. https://data.library.virginia.edu/getting-started-with-negative-binomial-regression-modeling/. Accessed: 13.08.18.

Glasgow & The Clyde Valley Strategic Development Planning Authority. 2017. *Clydeplan: Strategic Development Plan.* Glasgow & The Clyde Valley Strategic Development Planning Authority. Glasgow, UK.

Glasgow Centre For Population Health. 2017. *Active Travel in Glasgow: What We've Learned So Far.* Glasgow Centre For Population Health. Glasgow, UK.

Glasgow City Council. 2016. *City Development Plan.* Glasgow City Council. Glasgow, UK.

Glasgow City Council. 2015. *Glasgow's Strategic Plan For Cycling 2016 – 2025.* Glasgow City Council. Glasgow, UK.

Goodchild, M. 2007. *Citizens As Sensors: The World Of Volunteered Geography.* GeoJournal. Volume 69, Issue 4: 211–221.

Goodman, A. 2013. *Walking, Cycling And Driving To Work In The English And Welsh 2011 Census: Trends, Socioeconomic Patterning And Relevance To Travel Behaviour In General.* PLOS One. Volume 8, Issue 8. 10.1371/journal.pone.0071790. Accessed: 13.08.18.

Gould, J. 2013. *Cell Phone Enabled Travel Surveys: The Medium Moves The Message*. In *Transport Survey Methods: Best Practise For Decision Making*.

Zmud, J., Lee-Gosselin, M., Munizaga, M., Carrasco, J (eds). Emerald Group Publishing. Bingley, UK.

Gray, L., Leyland, A. 2008. *A Multilevel Analysis Of Diet And Socio-Economic Status In Scotland: Investigating The 'Glasgow Effect'.* Public Health Nutrition. Volume 12, Number 9: 1351–1358.

Green, J., Steinbach, R., Datta, J., Edwards, P. 2010. *Cycling In London: A Study Of Social And Cultural Factors In Transport Mode Choice: A Final Report To Smarter Travel Unit, Transport For London.* London School Of Health & Tropical Medicine. London, UK.

Griffin, G., Jiao, J. 2015. *Where Does Bicycling For Health Happen? Analysing Volunteered Geographic Information Through Place And Plexus.* Journal Of Transport & Health. Volume 2, Issue 2: 238–247.

Griffin, G., Nordback, K., Götschi, T., Stolz, E., Kothuri, S. 2014. *Monitoring Bicyclist And Pedestrian Travel And Behaviour, Current Research And Practice.* Transportation Research Circular. Number E-C183.

Haworth, J., 2016. *Investigating The Potential Of Activity Tracking App Data To Estimate Cycle Flows In Urban Areas.* The International Archives Of The Photogrammetry, Remote Sensing & Spatial Information Sciences, Volume XLI-B2. 2016 XXIII ISPRS Congress. Prague, Czech Republic.

Heesch, K., Langdon, M. 2016. *The Usefulness Of GPS Bicycle Tracking Data For Evaluating The Impact Of Infrastructure Change On Cycling Behavior.* Health Promotion Journal Of Australia. Volume 27: 222–229.

Heesch, K., Giles-Corti, B., Turrell, G. 2015. *Cycling For Transport And Recreation: Associations With The Socioeconomic, Natural And Built Environment*. Health & Place, Volume 36: 152–161.

Van Heeswijck, T., Paquet, C., Kestens, Y., Thierry, B., Morency, C., Daniel, M. 2015. *Differences In Associations Between Active Transportation And Built Environmental Exposures When Expressed Using Different Components Of Individual Activity Spaces.* Health & Place. Volume 33: 195–202.

Heinen, E., Van Wee, B., Maat, K. 2010. *Bicycle Use For Commuting: A Literature Review.* Transport Reviews. Volume 30, Issue 1: 105–132.

Heipke, C. 2010. *Crowdsourcing Geospatial Data.* International Society For Photogrammetry & Remote Sensing Journal Of Photogrammetry &Remote Sensing. Volume 65, Issue 6: 550–557.

Herrero, J. 2016. *Using Big Data To Understand Trail Use: Three Strava Tools.* TRAFx Research. https://www.trafx.net/insights.htm. Accessed: 13.08.18.

Hood, J., Sall, E., Charlton, B. 2011. *A GPS-Based Bicycle Route Choice Model For San Francisco, California.* Transportation Letters: The International Of Transportation Research. Volume 3: 63–75.

Information Services Division. 2007. *Childhood BMI Statistics.* http://www.isdscotland.org. Accessed 14.08.18.

Jestico, B., Nelson, T., M. Winters. 2016. *Mapping Ridership Using Crowdsourced Cycling Data.* Journal Of Transport Geography. Volume 52: 90–97.

Kamphuis, C., Giskes, K., Kavanagh, A., Thornton, L., Thomas, L., Lenthe, V., Mackenbach, F., Turrell, G. 2008. *Area Variation In Recreational Cycling In Melbourne: A Compositional Or Contextual Effect?* Journal Of Epidemiology & Community Health. Volume 62, Issue 10: 890–898.

Karazsia, B., Van Dulmen, M. 2008. *Regression Models For Count Data: Illustrations Using Longitudinal Predictors Of Childhood Injury.* Journal Of Pediatric Psychology. Volume 33. Number 10: 1076–1084.

Krizek, K., Forsyth, A., Baum, L. 2009. *Walking And Cycling: An International Literature Review.* State Of Victoria Department Of Transport. Melbourne, Australia.

Kuzmyak, R., Dill, J. 2014. *Walking And Bicycling In The United States.* Accident Analysis & Prevention. Volume 65: 63–71.

Lamb, K., Ogilvie, D., Ferguson, N., Murray, J. Wang, Y., Ellaway, A. 2012. *Sociospatial Distribution Of Access To Facilities For Moderate And Vigorous Intensity Physical Activity In Scotland By Different Modes Of Transport.* International Journal of Behavioral Nutrition and Physical Activity. Volume 9, Number 55: 1–10.

Law, C., Power, C., Graham, H., Merrick, D. 2007. *Obesity And Health Inequalities. Obesity Reviews.* Volume 8, Number 1: 19–22.

Leyland, A., Dundas, R., McLoone, P., Boddy, A. 2007. *Inequalities In Mortality In Scotland 1981–2001.* Glasgow: MRC Social and Public Health Sciences Unit. Glasgow, UK.

Lillis, D. 2014. *Generalized Linear Models In R, Part 2: Understanding Model Fit In Logistic Regression Output.* https://www.theanalysisfactor.com/r-glm-model-fit/. Accessed: 13.08.18.

Macintyre, S., Macdonald, L., Ellaway, A. 2008. *Do Poorer People Have Poorer Access To Local Resources And Facilities? The Distribution Of Local Resources By Area Deprivation In Glasgow, Scotland.* Social Science & Medicine. Volume 67: 900–914

Macklon, G., Burns, N. 2018. *Exploring Strava Cycling Data: Scottish Transport Applications Research.* Sustrans. Edinburgh, UK.

McBeth, M. 2009. *Gender & Cycling*. In *Gendered Journeys, Mobile Emotions.* Letherby, G., Reynolds, G. Ashgate Publishing Company. Aldershot, UK.

Marmot, M., Allen, J., Goldblatt, P., Boyce, T., McNeish, D., Grady, M., Geddes, I. 2010. *Fair Society, Healthy Lives: Strategic Review Of Health Inequalities In England Post-2010.* The Marmot Review. London, UK.

McCartney, G., Whyte, B., Livingston, M., Crawford, F. 2012. *Building A Bridge, Transport Infrastructure And Population Characteristics: Explaining Active Travel Into Glasgow.* Transport Policy. Volume 21: 119–125.

Musakwa, W., Selala, K. 2016. *Mapping Cycling Patterns And Trends Using Strava Metro Data In The City Of Johannesburg, South Africa.* Data In Brief. Volume 9: 898–905.

Norman, G., Kesha, N. 2015. *Using Smartphones For Cycle Planning.* IPENZ Transportation Group Conference. Christchurch, New Zealand.

Office Of National Statistics. 2018. *Census Geography.* https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography. Accessed: 13.08.18.

Ogilvie, F., Goodman, A. 2012. *Inequalities In Usage Of A Public Bicycle Sharing Scheme: Socio Demographic Predictors Of Uptake And Usage Of The London (UK) Cycle Hire Scheme.* Preventive Medicine. Volume 55, Issue 1: 40–45.

Ogilvie, D., Mitchell, R., Mutrie, N., Petticrew, M., Platt, S. 2008. *Personal And Environmental Correlates Of Active Travel And Physical Activity In A Deprived*

Urban Population. International Journal Of Behavioural Nutrition & Physical Activity. Volume 5, Issue 43. 10.1186/1479-5868-5-43. Accessed: 13.08.18.

Ogilvie, D., Foster, C., Rothnie, H. 2007. *Interventions To Promote Walking: Systematic Review.* British Medical Journal. Volume 334, Issue 7605. 10.1136/bmj.39198.722720.BE. Accessed: 13.08.18.

Ogilvie, D., Egan, M., Hamilton, V. 2004. *Promoting Walking And Cycling As An Alternative To Using Cars: Systematic Review.* British Medical Journal. Volume 329, Issue 763. 10.1136/bmj.38216.714560.55. Accessed: 13.08.18.

Olsen, J., Mitchell, R., Mutriec, N., Foley, L., Ogilvie, D. 2017. *Population Levels Of, And Inequalities In, Active Travel: A National, Crosssectional Study Of Adults In Scotland.* Preventive Medicine Reports. Volume 8: 129–134.

Panter, J., Corder, K., Griffin, S., Jones, A., Van Sluijs, E. 2013. *Individual, Sociocultural And Environmental Predictors Of Uptake And Maintenance Of Active Commuting In Children: Longitudinal Results From The SPEEDY Study*. International Journal Of Behavioural Nutrition & Physical Activity. Volume 10, Issue 83. http://www.ijbnpa.org/content/10/1/83. Accessed: 13.08.18.

Panter, J., Griffin, S., Jones, A., Mackett, R., Ogilvie, D. 2011. *Correlates Of Time Spent Walking And Cycling To And From Work: Baseline Results From The Commuting And Health In Cambridge Study.* International Journal of Behavioral Nutrition & Physical Activity. Volume 28, Number 124: 1–13.

Parkin, J., Ryley, T., Jones, T. 2007. *Barriers To Cycling; An Exploration Of Quantitative Analysis.* In *Cycling And Society*. Horton, D., Rosen, P., Cox, P. Ashgate Publishing Company. Aldershot, UK.

Pistoll, C., Goodman, A. 2014. *The Link Between Socioeconomic Position, Access To Cycling Infrastructure And Cycling Participation Rates: An Ecological Study*

In Melbourne, Australia. Journal Of Transport & Health. Volume 1, Issue 4: 251–259.

Pucher, J., Buehler, R., Seinen, M. 2011. *Bicycling Renaissance In North America? An Update And Re-Appraisal Of Cycling Trends And Policies.* Transportation Research Part A. Volume 45: 451–475.

Rachele, J., Kavanagh, A., Badland, H., Giles-Corti, B., Washington, S., Turrell, G. 2015. *Associations Between Individual Socioeconomic Position, Neighbourhood Disadvantage And Transport Mode: Baseline Results From The HABITAT Multilevel Study.* Journal Of Epidemiology & Community Health. Volume 69, Issue 12: 1217–1223.

Rind, E., Shortt, N., Mitchell, R., Richardson, E., Pearce, J. 2015. *Are Income-Related Differences In Active Travel Associated With Physical Environmental Characteristics? A Multi-Level Ecological Approach*. International Journal Of Behavioural Nutrition & Physical Activity. Volume 12, Issue 73. 10.1186/s12966-015-0217-1. Accessed: 13.08.18.

Romanillos, G., Zaltz Austwick, M., Ettema, D., De Kruijf, Z. 2016. *Big Data And Cycling*. Transport Reviews. Volume 36, Issue 1: 114–133.

Ryus, P., Ferguson, E., Laustsen, K., Schneider, R., Proulx, F., Hull, T. 2014. *Guidebook On Pedestrian And Bicycle Volume Data Collection.* UC Berkeley: Safe Transportation Research & Education Center. https://escholarship.org/uc/item/11q5p33w. Accessed: 13.08.18.

Sahlqvist, S., Heesch, K. 2012. *Characteristics Of Utility Cyclists In Queensland, Australia: An Examination Of The Associations Between Individual, Social, And Environmental Factors And Utility Cycling*. Journal Of Physical Activity & Health. Volume 9, Issue 6: 818–828.

Scottish Government 2018. *A More Active Scotland: Scotland's Physical Activity Delivery Plan.* Scottish Government. Edinburgh, UK.

Scottish Government. 2017A. *A Nation With Ambition: The Government's Programme for Scotland 2017-18.* Scottish Government. Edinburgh, UK.

Scottish Government. 2017B. *Climate Change Bill Consultation.* https://www.gov.scot/Topics/Environment/climatechange/Newclimatechangebill. Accessed: 13.08.18.

Scottish Government. 2016A. *Cycling Action Plan For Scotland 2017–2020.* Scottish Government. Edinburgh, UK.

Scottish Government. 2016B. *SIMD 16 Technical Notes.* Scottish Government. Edinburgh, UK.

Scottish Government. 2015. *Cleaner Air For Scotland: The Road To A Healthier Future.* Scottish Government. Edinburgh, UK.

Scottish Government. 2014. *Scotland's Third National Planning Framework.* Scottish Government. Edinburgh, UK.

Scottish Government. 2013. *Using The SIMD: Frequently Asked Questions.* https://www.gov.scot/Topics/Statistics/SIMD/FAQUsingSIMD. Accessed: 13.08.18.

Scottish Government. 2010. *Cycling Action Plan For Scotland.* Scottish Government. Edinburgh, UK.

Scottish Government. 2009. *Scottish Transport Statistics*. Volume 28. Scottish Government. Edinburgh, UK.

Steinbach, R., Green, J., Datta, J., Edwards, P. 2011. *Cycling And The City: A Case Study Of How Gendered, Ethnic And Class Identities Can Shape Healthy*

Transport Choices. Social Science & Medicine. Volume 72, Issue 7: 1123–1130.

Steer Davies Gleave. 2010. *Cycling Potential Index.* http://www.steerdaviesgleave.com/sites/default/files/newsandinsights/cycling_potential.pdf. Accessed: 13.08.18.

Steinbach, R., Green, J., Datta, J., Edwards, P. 2011. *Cycling And The City: A Case Study Of How Gendered, Ethnic And Class Identities Can Shape Healthy Transport Choices.* Social Science & Medicine. Volume 72, Number 7: 1123–1130.

Stinson, M., Bhat, C. 2003. *Commuter Bicyclist Route Choice: Analysis Using A Stated Preference Survey.* Transportation Research Record: Journal Of The Transportation Research Board. Number 1828: 107–115.

Strava Metro. 2016. *Comprehensive User Guide: Version 3.0 For 2016.* Strava LLC. San Francisco, USA.

Sun, Y. 2017A. *Exploring Potential Of Crowdsourced Geographic Information In Studies Of Active Travel And Health: Strava Data And Cycling Behaviour.* The International Archives Of The Photogrammetry, Remote Sensing & Spatial Information Sciences. Volume XLII–2/W7: 1357–1361.

Sun, Y., Du, Y., Wang, Y., Zhuang, L. 2017B. *Examining Associations Of Environmental Characteristics With Recreational Cycling Behaviour By Street-Level Strava Data.* International Journal Of Environmental Research & Public Health. Volume 14, Issue 644. 10.3390/ijerph14060644. Accessed: 13.08.18.

Sun, Y., Mobasheri, A. 2017C. *Utilizing Crowdsourced Data For Studies Of Cycling And Air Pollution Exposure: A Case Study Using Strava Data.* International Journal Of Environmental Research & Public Health. Volume 14, Number 274. 10.3390/ijerph14030274. Accessed: 13.08.18.

Sustrans. 2017. *Bike Life: Edinburgh.* Sustrans. Bristol, UK.

Transport For London. 2011. *What Are The Barriers To Cycling Amongst Ethnic Minority Groups And People From Deprived Backgrounds?* Policy Analysis Research Summary. Transport For London. London, UK.

Transport Scotland. 2018. *Scottish Transport Statistics, No 36, 2017 Edition.* Transport Scotland. Edinburgh, UK.

Transport Scotland. 2017. *Transport & Travel In Scotland 2016.* Transport Scotland. Edinburgh, UK.

UCLA: Statistical Consulting Group. 2018. *Negative Binomial Regression: SAS Annotated Output.* https://stats.idre.ucla.edu/sas/output/negative-binomial-regression/. Accessed: 13.08.18.

Walsh, D., McCartney, G., McCullough, S., Van Der Pol, M., Buchanan, D., Jones, R. 2013. *Exploring Potential Reasons For Glasgow's 'Excess' Mortality: Results Of A Three-City Survey Of Glasgow, Liverpool And Manchester.* Glasgow Centre For Population Health, NHS Health Scotland, University Of Aberdeen. Glasgow, UK.

Wooldridge, J. 2009. *Introductory Econometrics: A Modern Approach.* South-Western. Mason, USA.

Zeileis, A., Kleiber, C., Jackman, S. 2008. *Regression Models For Count Data In R.* Universität Innsbruck. Innsbruck, Austria.