

UNIVERSIDADE FEDERAL DO PARANÁ

PEDRO AUGUSTO BORGES DOS SANTOS

THE IMPACT OF BUILT ENVIRONMENT ON ROAD SAFETY IN CITIES

CURITIBA

2021

PEDRO AUGUSTO BORGES DOS SANTOS

THE IMPACT OF BUILT ENVIRONMENT ON ROAD SAFETY IN CITIES

Document presented as the qualifying paper  
for the obtaining of a degree in Master of Sciences  
by the Programa de Pós-Graduação em  
Planejamento Urbano of Setor de Tecnologia of  
Universidade Federal do Paraná .

Supervisor: Prof. Dr. Jorge Tiago Bastos  
Co-supervisor: Prof. Dr. Oscar Oviedo-Trespalacios

CURITIBA

2021

## ABSTRACT

Road traffic crashes causes more than 1.35 million deaths and 50 million in injuries per year, being the eighth leading cause of death in the world. Road traffic deaths involving pedestrians, cyclists and motorcyclists represents more than half of the total value of global road traffic deaths. The excess of speed performed by motorized vehicles is one of the main risk factors of road crashes, influencing the severity and chance of occurrence of road crashes. Most of traffic crashes and conflicts happens in cities. City development characteristics, including the built environment, can influence the road safety performance in urban areas. The main objective of this research is to investigate the influence of the built environment in road safety, using the city of Curitiba, Brazil, as scenario of the study. The built environment consists of physical features inside the city, including development patterns and roadway designs, and can be split into six categories: density, diversity, design, destination accessibility, distance to transit and demographics. Speeding was used as a road safety performance indicator. The Geographically Weighted Regression (GWR) statistical model was used to explore the correlation between built environment (BE) variables and the occurrence of speeding. Previous works used GWR to explore the correlation between BE and road crashes, displaying it's ability to analyze scenarios that are spatially nonstationary. The speeding data was collected using the Naturalistic Driving Study (NDS) method. NDS consists in monitoring drivers behavior in their own cars, both in normal and safety critical conditions, with the use of GPS sensors and cameras equipped in the volunteer's cars. In this work, data from 16 drivers was used, containing 491 trips, 238.85 hours of driving and 5,362.75 km of distance travelled in Curitiba and it's metropolitan area, ranging between 2019 and 2021. The GWR model was applied using Curitiba's traffic analysis zones as zonal level. Multiple kernel types and bandwidths sizes were tested. The model diagnostics showed that GWR performed better than the global regression, but none of the models could validly predict the spatial heterogeneity present in the occurrence of speeding. Only the variable "density of speed cameras", included in the design category, showed a correlation to speeding, at a statistical significance of 95%, with 97.2% of TAZs presenting a negative coefficient, therefore, a inverted correlation to speeding. The current sample is relative small to conduct a more significant analysis. It is important to test new models with a bigger sample before reaching any further conclusions.

**Key-words:** Built environment. Speeding. Road safety.

## RESUMO

Os sinistros de trânsito são causa de mais de 1,35 milhão de mortes e 50 milhões de feridos por ano, sendo a oitava causa de morte no mundo. As mortes no trânsito envolvendo pedestres, ciclistas e motociclistas representam mais da metade do valor total das mortes no trânsito global. O excesso de velocidade praticado por veículos motorizados é um dos principais fatores de risco dos sinistros de trânsito, influenciando na gravidade e na chance de ocorrência dos sinistros. A maioria dos sinistros de trânsito e conflitos acontecem nas cidades. As características do desenvolvimento espacial de uma cidade, incluindo o ambiente construído, podem influenciar o desempenho da segurança viária em áreas urbanas. O objetivo principal desta pesquisa é investigar a influência do ambiente construído na segurança viária, utilizando a cidade de Curitiba, Brasil, como cenário do estudo. O ambiente construído consiste em características físicas dentro da cidade, incluindo padrões de desenvolvimento e características das vias, e pode ser dividido em seis categorias: densidade, diversidade, design, acessibilidade ao destino, distância ao trânsito e dados demográficos. O excesso de velocidade foi usado como um indicador de desempenho de segurança viária. O modelo estatístico de Regressão Geograficamente Ponderada (RGP) foi utilizado para explorar a correlação entre as variáveis do ambiente construído (AC) e a ocorrência de excesso de velocidade. Trabalhos anteriores utilizaram o RGP para explorar a correlação entre o AC e os sinistros de trânsito, exibindo sua capacidade de analisar cenários que são espacialmente não estacionários. Os dados de velocidade foram coletados usando o método do Estudo Naturalístico de Condução (ENC). O ENC consiste em monitorar o comportamento dos condutores em seus próprios carros, tanto em condições normais quanto em condições críticas de segurança, com o uso de sensores GPS e câmeras equipadas nos carros dos voluntários. Neste trabalho, foram utilizados dados de 16 motoristas, contendo 491 viagens, 238,85 horas de condução e 5.362,75 km de distância percorrida em Curitiba e sua região metropolitana, no período entre 2019 e 2021. O modelo RGP foi aplicado usando as zonas de trânsito (ZTs) de Curitiba como unidade de análise. Vários tipos de kernel e tamanhos de largura de banda foram testados. Os diagnósticos do modelo mostraram que o RGP teve um desempenho melhor do que a regressão global, mas nenhum dos modelos pôde predizer validamente a heterogeneidade espacial presente na ocorrência do excesso de velocidade. Apenas a variável “densidade de radares”, incluído na categoria design, apresentou correlação com o excesso de velocidade, com significância estatística de 95% e com 97,2% das ZTs apresentando coeficiente negativo, portanto, correlacionado invertidamente com o excesso de velocidade. A amostra atual é relativamente pequena para conduzir uma

análise mais significativa. É importante testar novos modelos com uma amostra maior antes de chegar a mais conclusões.

**Palavras-chaves:** Ambiente construído. Excesso de velocidade. Segurança viária.

## LIST OF FIGURES

FIGURE 1 – ROAD TRAFFIC RATES IN OECD COUNTRIES . . . . .	13
FIGURE 2 – ROAD TRAFFIC DEATHS ON BRAZIL . . . . .	14
FIGURE 3 – ROAD TRAFFIC MORTALITY RATE ON BRAZIL . . . . .	15
FIGURE 4 – ROAD TRAFFIC FATALITY RATE ON BRAZIL . . . . .	16
FIGURE 5 – MOTORIZATION RATE ON BRAZIL . . . . .	16
FIGURE 6 – ROAD TRAFFIC DEATHS ON CURITIBA . . . . .	17
FIGURE 7 – MOTORIZATION RATES ON BRAZILIAN STATES CAPITAL CITIES, IN 2019 . . . . .	18
FIGURE 8 – MORTALITY RATES ON BRAZILIAN STATES CAPITAL CITIES, IN 2019 . . . . .	19
FIGURE 9 – FATALITY RATES ON BRAZILIAN STATES CAPITAL CITIES, IN 2019 . . . . .	19
FIGURE 10 – HADDON MATRIX . . . . .	20
FIGURE 11 – RELATIONSHIP BETWEEN SPEED AND BRAKING DISTANCE . . . . .	22
FIGURE 12 – RELATIONSHIP BETWEEN SPEED AND KINETIC ENERGY . . . . .	23
FIGURE 13 – PROBABILITY OF PEDESTRIAN FATALITY AT DIFFERENT IM- PACT SPEEDS . . . . .	24
FIGURE 14 – RELATIONSHIP BETWEEN TRAFFIC SPEED, VOLUME AND DENSITY . . . . .	26
FIGURE 15 – EXTRACTION OF FREE-FLOW AND SPEEDING EPISODES FROM TRIPS . . . . .	28
FIGURE 16 – BUILT ENVIRONMENT AND TRAFFIC SAFETY . . . . .	29
FIGURE 17 – STRUCTURAL AXES IN CURITIBA . . . . .	37
FIGURE 18 – LAND USE DENSITIES IN CURITIBA . . . . .	37
FIGURE 19 – BEHAVIORAL TRAFFIC SAFETY STUDIES . . . . .	38
FIGURE 20 – KERNEL FUNCTIONS PLOT . . . . .	45
FIGURE 21 – CAMERA POSITIONING INSIDE THE VEHICLE . . . . .	47
FIGURE 22 – CAMERA IMAGES . . . . .	48
FIGURE 23 – NATURALISTIC DATA COLLECTION AND PROCESSING . . . . .	48
FIGURE 24 – VALID TIME SELECTION . . . . .	49
FIGURE 25 – VARIABLE DATA FLOWCHART . . . . .	51
FIGURE 26 – ROAD AXIS BUFFER . . . . .	51
FIGURE 27 – SPEEDING OCCURENCE AND SPEEDING OPPORTUNITY . . . . .	52
FIGURE 28 – PD AND LDI . . . . .	53

FIGURE 29 – DIS AND DSC . . . . .	54
FIGURE 30 – TSD AND PAR . . . . .	55
FIGURE 31 – SND AND DCSU . . . . .	56
FIGURE 32 – BSD AND AVI . . . . .	56
FIGURE 33 – GWR MODELLING PROCESS . . . . .	57
FIGURE 34 – DISTANCE TRAVELED PER DAY OF THE WEEK . . . . .	60
FIGURE 35 – TRIPS PER DAY OF THE WEEK . . . . .	61
FIGURE 36 – DISTANCE TRAVELED PER HOUR OF THE DAY . . . . .	61
FIGURE 37 – TRIP START PER HOUR OF THE DAY . . . . .	62
FIGURE 38 – TRAVELED DISTANCE PER TAZ . . . . .	62
FIGURE 39 – SPEEDING PER DAY OF THE WEEK . . . . .	64
FIGURE 40 – SPEEDING PER HOUR OF THE DAY . . . . .	64
FIGURE 41 – TYPES OF SPEEDING BEHAVIOR OF DRIVERS . . . . .	65
FIGURE 42 – REMAINING TAZS AFTER DATA ARRANGEMENT . . . . .	66
FIGURE 43 – SPEEDING PER TAZ . . . . .	67
FIGURE 44 – SPEARMAN CORRELOGRAM OF INDEPENDENT VARIABLES	67
FIGURE 45 – GW LOCAL $R^2$ . . . . .	69
FIGURE 46 – SPEEDING GW LOCAL MEAN AND STANDARD DEVIATION . .	70
FIGURE 47 – DSC AND PAR COEFFICIENT ESTIMATES . . . . .	70
FIGURE 48 – DIS AND DCSU COEFFICIENT ESTIMATES . . . . .	71
FIGURE 49 – AVI AND PD COEFFICIENT ESTIMATES . . . . .	72
FIGURE 50 – LDI AND BSD COEFFICIENT ESTIMATES . . . . .	72

## LIST OF TABLES

TABLE 1 – FACTORS THAT AFFECTS SPEEDING AND SPEED CHOICE . . . . .	25
TABLE 2 – BUILT ENVIRONMENT VARIABLES . . . . .	30
TABLE 3 – CORRELATION BETWEEN DENSITY AND ROAD SAFETY OUTCOMES . . . . .	31
TABLE 4 – CORRELATION BETWEEN DIVERSITY AND ROAD SAFETY OUTCOMES . . . . .	32
TABLE 5 – CORRELATION BETWEEN DESIGN AND ROAD SAFETY OUTCOMES . . . . .	33
TABLE 6 – CORRELATION BETWEEN DESTINATION ACCESSIBILITY AND ROAD SAFETY OUTCOMES . . . . .	35
TABLE 7 – CORRELATION BETWEEN DISTANCE TO TRANSIT AND ROAD SAFETY OUTCOMES . . . . .	35
TABLE 8 – CORRELATION BETWEEN DEMOGRAPHICS AND ROAD SAFETY OUTCOMES . . . . .	36
TABLE 9 – NATURALISTIC DRIVING STUDIES PROGRAMS . . . . .	40
TABLE 10 – KERNEL FUNCTIONS . . . . .	44
TABLE 11 – GWR AND ROAD SAFETY STUDIES . . . . .	46
TABLE 12 – SELECTED PARTICIPANTS INFORMATION . . . . .	50
TABLE 13 – BUILT ENVIRONMENT VARIABLES . . . . .	53
TABLE 14 – DESCRIPTIVE STATISTICS OF TRIP DISTANCES . . . . .	59
TABLE 15 – DESCRIPTIVE STATISTICS OF SPEEDING PER TRIP . . . . .	63
TABLE 16 – DESCRIPTIVE STATISTICS OF VARIABLES . . . . .	66
TABLE 17 – MODEL DIAGNOSTICS . . . . .	68
TABLE 18 – SUMMARY OF LINEAR MODEL ESTIMATES . . . . .	68
TABLE 19 – DESCRIPTIVE STATISTICS OF GWR COEFFICIENT ESTIMATES	69
TABLE 20 – PERCENTAGE OF TAZS WITH POSITIVE/NEGATIVE COEFFICIENTS . . . . .	73

## SUMMARY

<b>1</b>	<b>INTRODUCTION</b>	<b>9</b>
1.1	OBJECTIVES	9
1.2	JUSTIFICATION	10
1.3	THESIS STRUCTURE	11
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>12</b>
2.1	ROAD SAFETY SCENARIO	12
2.2	SPEEDING AS A RISK FACTOR	22
2.3	ROAD SAFETY PERFORMANCE AND THE BUILT ENVIRONMENT	29
2.4	NATURALISTIC DRIVING STUDIES	38
2.5	GEOGRAPHICALLY WEIGHTED REGRESSION	43
<b>3</b>	<b>METHODS</b>	<b>47</b>
3.1	NATURALISTIC DRIVING STUDY - BRAZIL	47
3.2	DATA PROCESSING AND VARIABLE CONSTRUCTION	50
3.2.1	Speeding	50
3.2.2	Built Environment	52
3.3	GEOGRAPHICALLY WEIGHTED REGRESSION MODELLING	55
<b>4</b>	<b>ANALYSIS AND RESULTS</b>	<b>59</b>
4.1	NATURALISTIC DATA	59
4.2	SPEEDING DATA	60
4.3	GEOGRAPHICALLY WEIGHTED REGRESSION RESULTS	65
<b>5</b>	<b>CONCLUSIONS</b>	<b>74</b>
	<b>BIBLIOGRAPHY</b>	<b>76</b>
	<b>APPENDIX 1 - POINTS INTO LINESTRING CONVERSION</b>	<b>88</b>
	<b>APPENDIX 2 - GWR R SCRIPT</b>	<b>89</b>
	<b>ANNEX 1 - DATA ACQUISITION AND CONVERSION SCRIPTS</b>	<b>98</b>

## 1 INTRODUCTION

In 2010, the United Nations (UN) declared the Decade of Action for Road Safety 2011-2020, calling the signatory countries to reduce in half the predicted number of deaths and injuries in traffic until 2020 (WHO, 2011). In Brazil, the National Plan of Reduction of Deaths and Injuries (PNATRANS) was created in 2018, aiming to reduce the national fatality rate (deaths per 10,000 vehicles) and mortality rate (deaths per 100,000 population) in half, within the period of 2019-2028 (MINISTRY OF CITIES, 2018). Following the Decade of Action 2011-2020, the UN declare another Decade of Action for Road Safety with the same objectives of the last one, but considering a time frame of 2021-2030 (WHO, 2020).

One of the main risk factors of road crashes is speeding, which increases the severity and occurrence of crashes (MOHAN, 2016b). In urban areas, the severity of crashes has a higher impact to more vulnerable users, like pedestrians and cyclists (WELLE et al., 2016). The characteristics inside the urban environment, including land use codes and the built environment can mitigate the occurrence of speeding and improve the road safety performance (KNOFLACHER, 2016). Therefore, this work aims to better explore the relationship between the built environment inside cities and the level of road safety performance, considering the occurrence of speeding as a intermediate outcome.

### 1.1 OBJECTIVES

In the context of urban planning and road safety management, the main objective of this research is to investigate the influence of demographic, land use and constructed environment on the speeding practices as a risk factor on the frequency and severity of road crashes. The scenario of the study is the city of Curitiba, capital of the state of Paraná, Brazil. The correlation will be analyzed with the use of the Geographically Weighted Regression statistical model.

With the use of speeding data collected from drivers that participated in a Naturalistic Driving Study (NDS) performed in Curitiba, this investigation aims to contribute to the knowledge of its relationship with variables from the built environment (BE). These BE variables are categorized by Ewing and Dumbaugh (2009) in five groups called “5D”: density, diversity, design, destination accessibility and distance to transit. In addition to these variables, this study aims to relate the income, as a demographic variable, to the speeding as well.

As co-benefits in the investigation, this thesis aims to offer some level of insight regarding the development and update of speed control in cities, considering the operational and structural planning guidelines for the road systems and land use. With these factors in mind, it could be possible to present new ideas to the planning practices in Curitiba, regarding mobility plans, master plans and zoning laws.

## 1.2 JUSTIFICATION

The excess of traffic speed performed by vehicles on urban environments are a main risk factor in the chance and severity of road crashes, considering there is a great volume of interaction between motorized and vulnerable users (ELVIK et al., 2009). Considering the direct relationship between the mobility, the land use patterns and the built environment (DE VOS; WITLOX, 2013), it is important to investigate these factors in search of improvements in the management process of the road safety. Speeding, which is on the primary causes for traffic crashes (WHO, 2013) can be better studied and analyzed by the collection of naturalistic data.

Having in mind that most of the traffic crashes and conflicts happens in cities (WHO, 2018), as a consequence of the fast process of urbanization in the last decades, it is crucial to create safer conditions in these localities. In Brazil, the management of the mobility is attributed to the municipalities (BRASIL, 1997). Therefore, the management of road safety, as a task inherent to traffic management, is also an attribution of the municipalities. It is necessary to observe the issues of road safety in cities when conducting the process of urban planning. Some national laws foresees this integration between the planning process by the municipalities and the sustainable mobility are clearly defined: The Statute of the Cities (BRASIL, 2001) and The Law of the Urban Mobility (BRASIL, 2012).

The built environment can affect the traffic safety through three main mediators: the traffic volumes, traffic conflicts and traffic speeds - which can directly affect the crash occurrence and severity in urban environments (EWING; DUMBAUGH, 2009). Traffic crashes are a final outcome of the problems involving the road safety. Therefore, in order to have a method to analyzed the operational conditions of the traffic safety in cities, it can be beneficial to use a intermediate outcome - or a safety performance indicator (SPI) (BASTOS, 2014) - which can be the amount of speeding that occurs across its territory, in addition to the studies that uses the traffic crashes as dependent variables.

### 1.3 THESIS STRUCTURE

This document is divided into five chapters. Next to this chapter, Chapter 2 includes the literature review, Chapter 3 contains the methodological proceedings used in this work and Chapter 4 includes the data that resulted from the method and it's analysis. Chapter 5 includes a brief conclusion of the thesis.

## 2 LITERATURE REVIEW

This Chapter is divided into 5 sections, each presenting a relevant subject to the study. Section 2.1 has discussions involving the road safety problems that occurs and how they can be managed. Section 2.2 presents the main risk factor investigated in this document: speeding. It is discussed how it is related to the road safety, the factors that can influence speeding and how its occurrence can be investigated. Section 2.3 includes the discussion about the built environment and it's variables, relating to the road safety performance and urban planning practices. Section 2.4 discusses naturalistic driving studies as a method of speeding data collection, and how it was performed in previous works. Section 2.5 discusses about the geographically weighted regression statistical method, and how it can be used to investigate built environment variables and road safety outcomes.

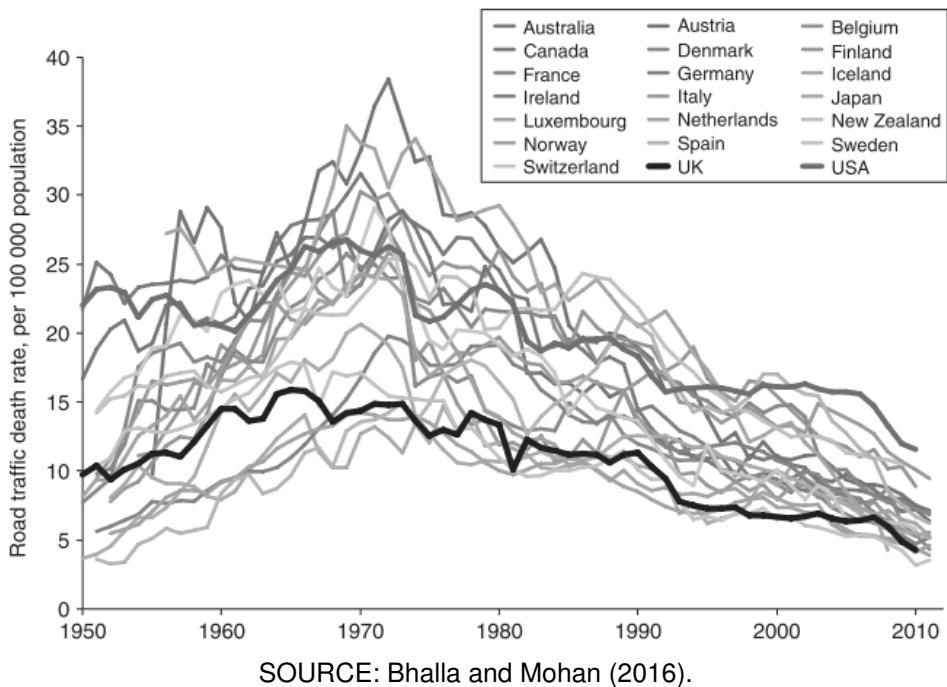
### 2.1 ROAD SAFETY SCENARIO

Road traffic crashes around the world claims more than 1.3 million lives in each year, representing the eighth leading cause of deaths and causing up to 50 million injuries. Low and middle income countries (LMICs), including Brazil, suffers from traffic crashes death rates three times higher when compared to developed countries (WHO, 2018). The increasing number of traffic crashes deaths on LMICs is a consequence of a intense process of motorization that has been occurring in the last decades. To Bhalla and Mohan (2016), it is important to understand the evolution of road safety management on OECD developed countries in order to overcome the road safety problems in LMICs.

Over the last century, the road safety performance of the OECD countries showed a consistent pattern. The road traffic death rate (per 100,000 population) in these countries were rising until the 1960s, as seen in FIGURE 1. After this period, all countries showed a declining pattern, whilst the LMICs still had a rising pattern due to the rapid growth in their motor vehicle fleets. This behavior of rising and declining trend on the road traffic death rate could be explained by three phenomenons: economic determinism, risk substitution and a political shift in the road safety paradigm (BHALLA; MOHAN, 2016).

In the scope of economic determinism, the road traffic deaths are defined as a process related to the country development. The rising pattern is associated with the increase in motorization, and the falling happens after a certain level of development,

FIGURE 1 – ROAD TRAFFIC RATES IN OECD COUNTRIES



SOURCE: Bhalla and Mohan (2016).

where the countries have the means to start investing in road safety. But this hypothesis has some flaws. This creates an impression that LMICs are not able to invest in road safety before becoming full developed, which is inaccurate. Also, this idea shifts the focus on investment in direct interventions, encouraging the countries to focus on income growth as a strategy for road safety (BHALLA; MOHAN, 2016).

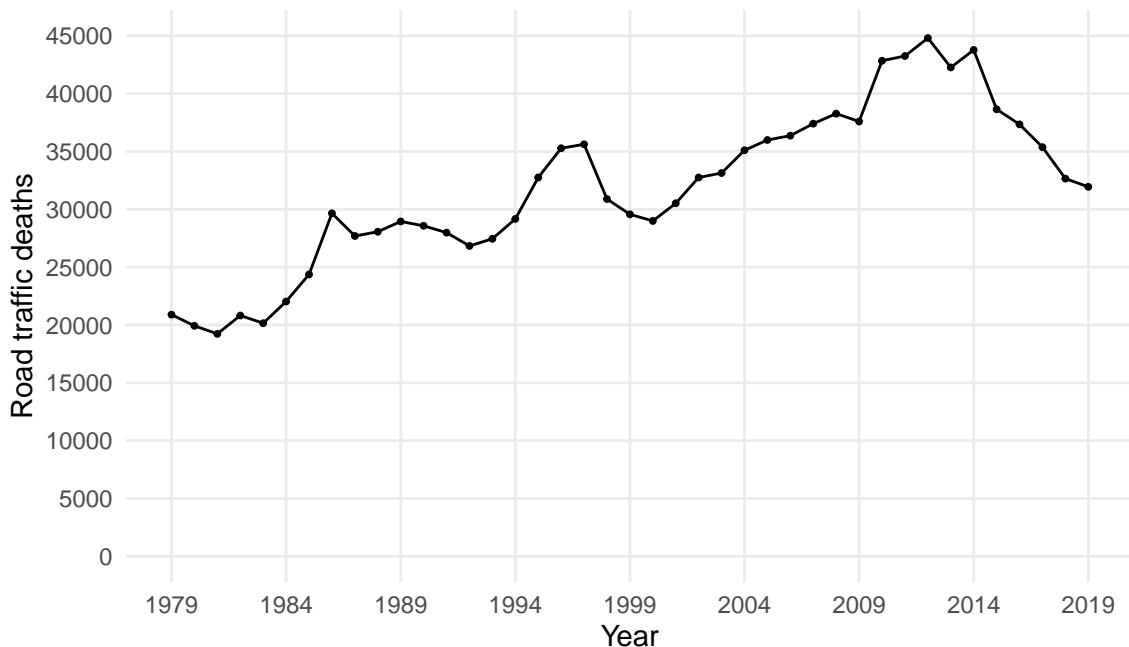
As the mobility in general becomes more motorized towards the use of the car, pedestrians starts transitioning into car users. This circumstance is known as risk substitution, where the increase in car users occurs at the same time as the number of pedestrians declines, lowering the exposure to more severe road traffic injuries like the collision between cars and pedestrians and lowering the number of road traffic deaths. But this phenomenon can be imprecise when considering the motorization of LMICs, in which mass transit, motorcycles and other non motorized transports have a greater role on the vehicle fleet (BHALLA; MOHAN, 2016).

A factor that explains this pattern in a more precise way is the political shift in the road safety paradigm that happened on the OECD countries. Before the 1950s, the belief that drivers were the only responsible for the road crashes lead the discussions regarding road safety. Therefore, most of the interventions in road safety management ignored the design and development of the build environment and the vehicles. Between the 1960s and the 1970s, OECD countries started to regulate transport in order to tackle the road safety problems that were rising, establishing new laws and road safety

management institution on national e local levels. (BHALLA; MOHAN, 2016). Analyzing the road safety data from Brazil, it is possible to correlate the actual stage with what the OECD countries passed during the 1970s.

According to WHO (2018), it was predicted that Brazil would had a road traffic mortality rate (number of deaths per 100,000 inhabitants) of 22.5 in 2016, the highest rate between the south american countries. In 2019 (the last entry available by Ministry of Health (2020)), road crashes were responsible for 31,945 deaths. Considering the Decade of Action for Road Safety 2011-2020 (WHO, 2011), Brazil will reach the goal of reducing the road traffic deaths in half (comparing to the number projected for 2020, in case of a rising trend in deaths) by the end of the 2010s. FIGURE 2 has the time series of road traffic deaths in the last decades.

FIGURE 2 – ROAD TRAFFIC DEATHS ON BRAZIL

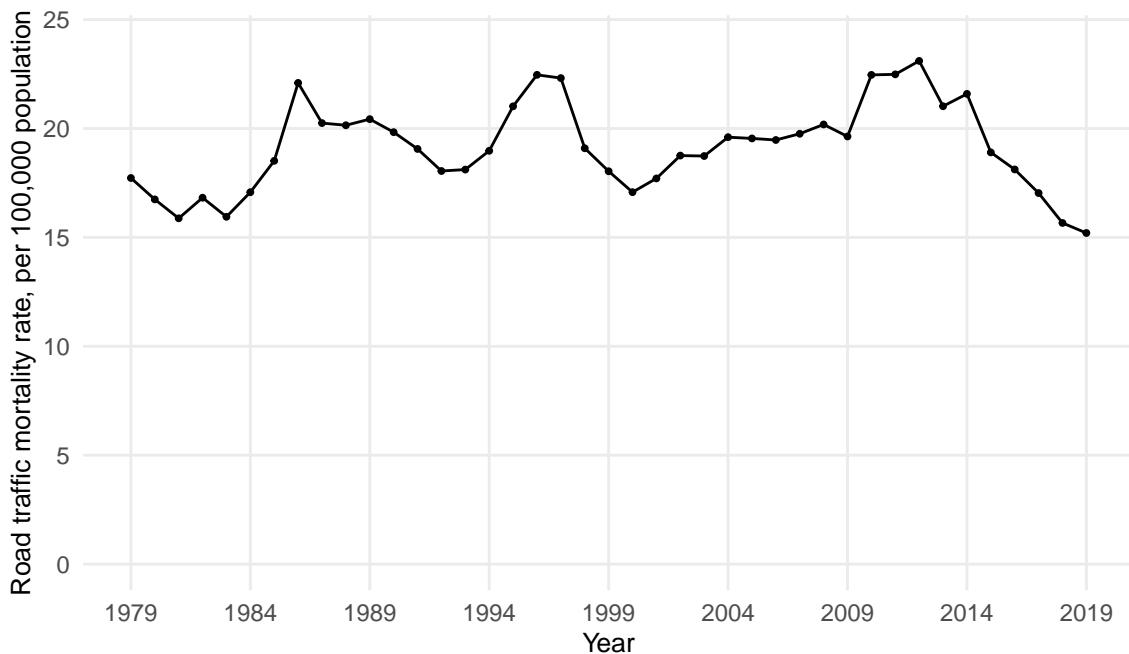


SOURCE: The Author, based on Ministry of Health (2020).

The year of 1979 is the earliest official data entry available. Starting in the 1980s, there is a overall rising trend in the numbers of road traffic deaths that reach its peak in 2012, then it starts declining. In 2010, the year before the Decade of Action, Brazil had 42,844 deaths and reached the max value in 2012: 44,812 road traffic deaths. Almost at the end of the decade the number of deaths declined, reaching 31,945. The next plot (FIGURE 3) shows the mortality rate in Brazil, considering the same time span. In general, it follows the same pattern of the absolute road traffic deaths.

The maximum registered mortality rate is 23.1 deaths per 100,000 inhabitants in 2012, the same year which Brazil reached its maximum value for absolute road traffic

FIGURE 3 – ROAD TRAFFIC MORTALITY RATE ON BRAZIL



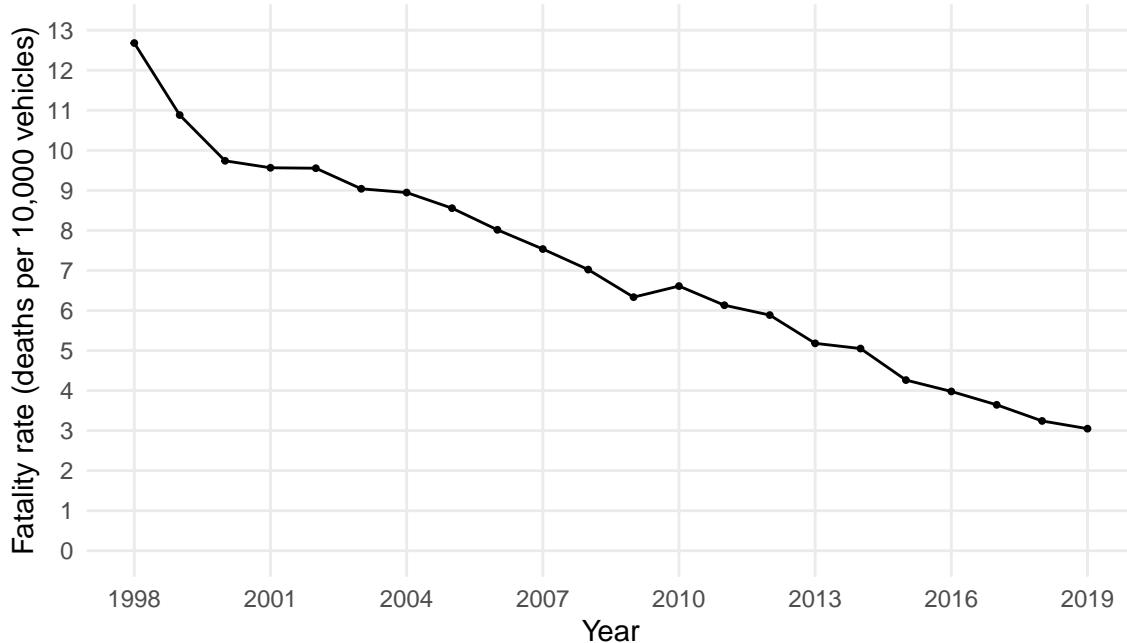
SOURCE: The Author (2021), based on Ministry of Health (2020) and Ministry of Health (2021).

deaths. Since then the mortality rate has been declining, reaching a rate of 15.2 deaths per 100,000 inhabitants in 2019. In comparison to the OECD countries (FIGURE 1), Brazil presented a strong declining pattern a few years later, after the 2010s. Another indicator that considers the exposition to traffic hazards is the fatality rate - number of deaths per 10,000 vehicles (FIGURE 4). The oldest entry of fleet size in the DENATRAN (2020) database is 1998.

Between 1998 and 2019 there is a declining pattern in the fatality rate, starting with 12.7 deaths per 10,000 vehicles in 1998 and ending with 3.0 in 2019, representing a reduction of 76%. This declining pattern can be explained by the continuous rise of the motorization rate (FIGURE 5). In 1998, the motorization rate was 151 vehicles per 1,000 population, increasing to the rate of 499 in 2019. The motorization can be defined as one indicator of the development un the country. Considering the motorization stages defined by Jørgensen (2005), Brazil can be classified nowadays with a exploding motorization.

The first stage of motorization - developing motorization - occurs when a country have a motorization rate between 50 and 100 vehicles per 1,000 inhabitants. This first stage happened in Brazil before 1998. The explosion of the motorization is the second stage, and it happens with a motorization rate of between 300 and 400. The third and last stage is the saturation, and occurs when the motorization rates reaches more than 400 and its tendency stops rising (JØRGENSEN, 2005). Although Brazil has a present

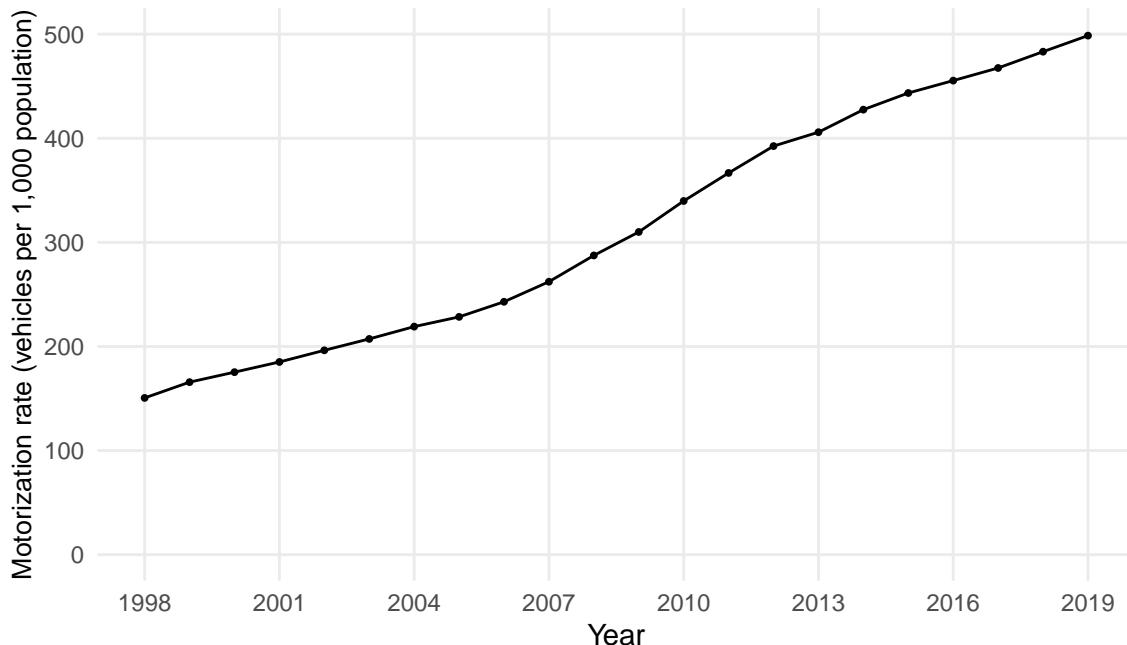
FIGURE 4 – ROAD TRAFFIC FATALITY RATE ON BRAZIL



SOURCE: The Author (2021), based on Ministry of Health (2020) and DENATRAN (2020).

motorization rate above 400, it is plausible to affirm that the country still hasn't reached the stage of saturation, considering that the rates are still in a rising pattern.

FIGURE 5 – MOTORIZATION RATE ON BRAZIL



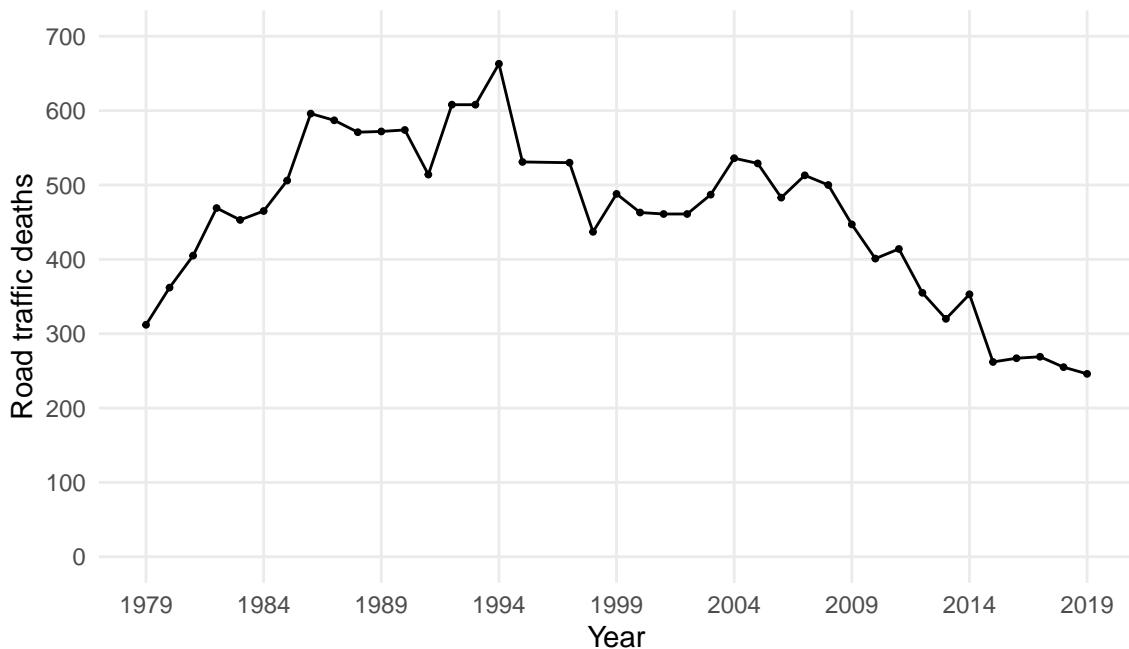
SOURCE: The Author (2021), based on Ministry of Health (2021) and DENATRAN (2020).

In the 1970s, Brazil have undergone through a intense development in the car industry, lowering the price of the vehicles and increasing the development of road

infrastructure in cities and rural areas, impacting directly in the rise of the motorization in the country (VASCONCELLOS, 2013). To Harvey (1982), this incentive to motorization created a urban environment that looked to answer the increasing demand of car use. This process lead to worse safety conditions in brazilian cities, specially to non motorized users in the road system.

The variation in the number of road traffic deaths and road traffic mortality heavily influenced by the socioeconomic development and political landscape (FERRAZ et al., 2012). The country had different economical situations between the 2000s and the 2010s. In the period of 2000-2010, Brazil had a continuous growth in the economy, which reflected on the rising pattern of the road traffic deaths and road traffic mortality rates. After the 2010, the country entered in a economic recession, leading to the reduction of these numbers (BASTOS; GARRONCE, et al., 2020). Focusing on the area of study, the road traffic deaths in Curitiba are presented in FIGURE 6.

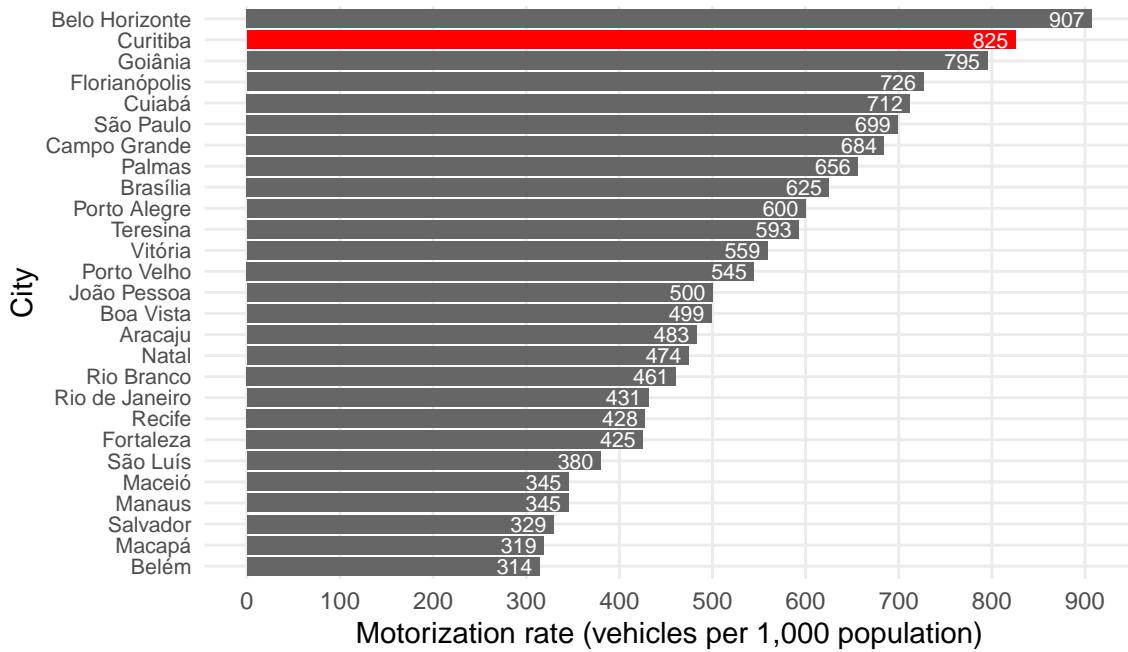
FIGURE 6 – ROAD TRAFFIC DEATHS ON CURITIBA



SOURCE: The Author (2021), based on Ministry of Health (2020).

The tendency of road traffic deaths presented a rising pattern between 1979 and 1994, starting with 312 deaths and reaching a maximum of 663 deaths. Overall, the pattern of the time series started declining after 1995, ending with a minimum value of 246 in 2019. Following the brazilian trend, Curitiba's motorization rate kept rising in the last decades. The increase in motorized transit can lead to more exposition to traffic crashes and injuries, in case road safety interventions are not implanted correctly. The plot in FIGURE 7 compares the motorization rate between all brazilian states capital cities.

FIGURE 7 – MOTORIZATION RATES ON BRAZILIAN STATES CAPITAL CITIES, IN 2019

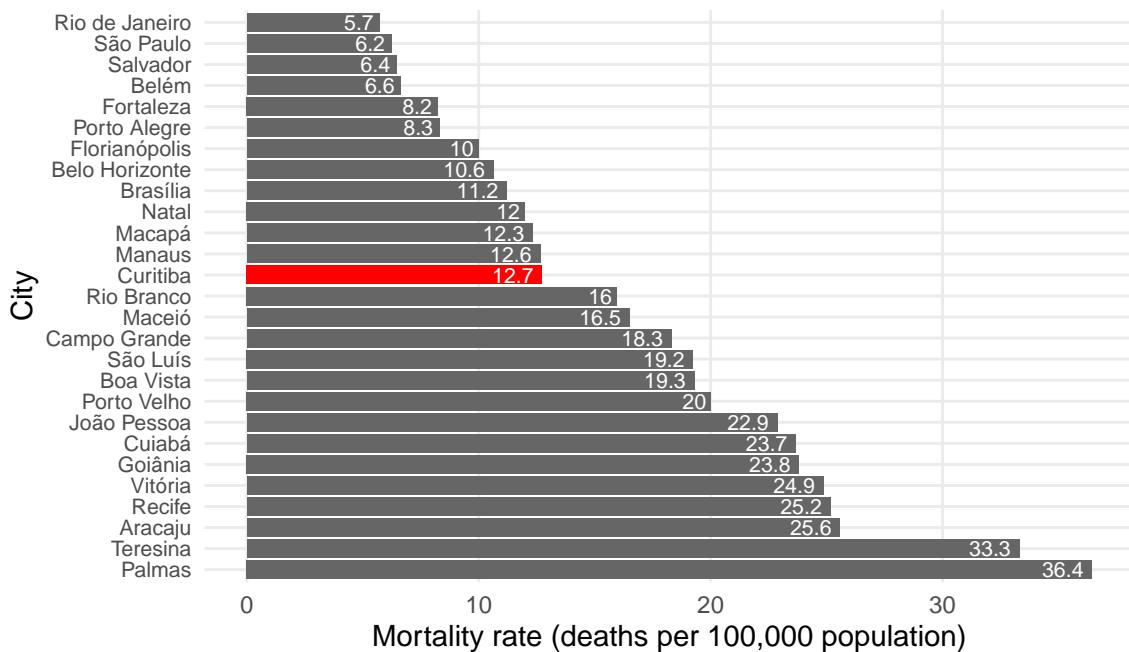


SOURCE: The Author (2021), based on Ministry of Health (2021) and DENATRAN (2020).

Curitiba was, in 2019, the second most motorized capital in Brazil, with a motorization rate of 825 vehicles per 1,000 population, higher than the brazilian average, losing only to Belo Horizonte, which presented a motorization rate of 907 vehicles per 1,000 population. When comparing the road traffic mortality rate (FIGURE 8), Curitiba stays in the 13th place with 12.7 deaths per 100,000 inhabitants, below the brazilian average of 15.2 in 2019. The lowest value of road traffic mortality rate belongs to Rio de Janeiro, presenting a rate of 5.7 deaths per 100,000 population, and the highest belongs to Palmas, with a alarming rate of 36.4. In 2015, Curitiba had a mortality rate of 13.94 deaths per 100,000 inhabitants, a higher value when comparing to more developed countries, like Sweden (2.8), UK (3.1) and Germany (4.1). Curitiba's rate for 2015 has a similar value to other latin american countries: Mexico (13.1), Uruguay (13.4), Peru (13.5) and Argentina (14.00) (WHO, 2018).

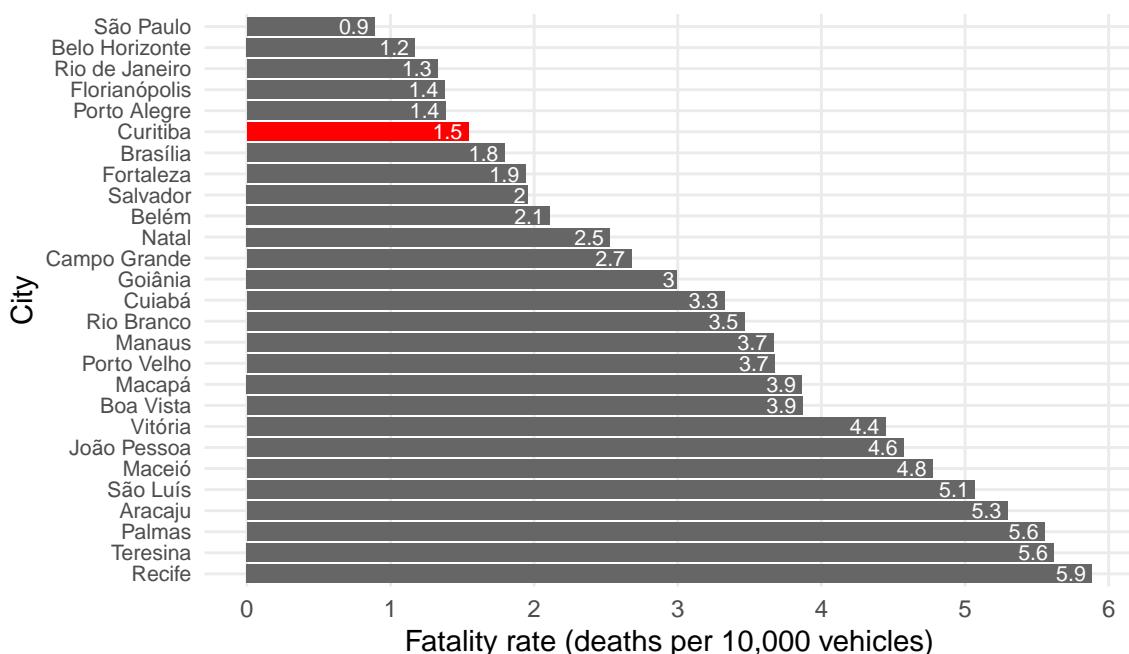
Considering the fatality rate as a road safety performance indicator, FIGURE 9 plot shows that Curitiba have the 6th best performance, with a fatality rate of 1.5 deaths per 10,000 vehicles. São Paulo is the capital with the lowest fatality rate (0.9) and Recife has the highest one (5.9). Observing the fatality rate in Brazil in the same period (3.0), Curitiba is below the country average, with exactly 50% less deaths per 10,000 vehicles. Road traffic crashes and its related injuries have become a serious health problem across brazilian cities. As motorization increases, the number of conflicts and possible crashes increases as well.

FIGURE 8 – MORTALITY RATES ON BRAZILIAN STATES CAPITAL CITIES, IN 2019



SOURCE: The Author (2021), based on Ministry of Health (2020) and Ministry of Health (2021).

FIGURE 9 – FATALITY RATES ON BRAZILIAN STATES CAPITAL CITIES, IN 2019



SOURCE: The Author (2021), based on Ministry of Health (2020) and DENATRAN (2020).

Capable interventions in road safety needs to consider traffic deaths and injuries as serious public health problems. Therefore, its control must follow the principle as control of any other public health problem. Road traffic injuries are the result of a complex interaction of multiple factors. These factors can involve human, environmental and vehicle factors, in addition to sociological, psychological, physical and technological factors present the process of traffic crashes. As other health problems, it is possible to analyze road traffic injuries considering three different phases in time: pre-crash, during the crash and post-crash (MOHAN, 2016a).

These distinct phases of time and factors of road traffic crashes can be arranged into a matrix created by Haddon (1980) in order to map all events related to a injury. Named after its author, the Haddon matrix (FIGURE 10) consists of two dimensions: one for the time aspect of the event and a second one representing three main discrete factors: human, vehicle and environment. Crossing each step of both dimensions sets nine cells, in which can be made a list of countermeasures to control the damage or to prevent a possible incident, associating these measures to each pair of factors.

FIGURE 10 – HADDON MATRIX

		FACTORS		
		Human	Vehicle	Environment
PHASES	Pre-crash	1	2	3
	During the crash	4	5	6
	Post-crash	7	8	9

SOURCE: The Author (2021), based on Haddon (1980).

Cells 1, 2 and 3 contains measures towards the prevention of crashes. On a road traffic crash scenario, cell 1 considers the behavior (aggressive driving, driving under influence, distraction) and training of the road users (pedestrians, drivers, motorcyclists and cyclists), cell 2 presents safety interventions related to the vehicles (use of daylight headlights, speed control systems, etc.) and cell 3 contains all elements of the road infrastructure the and build environment that influences directly into the occurrence of this event. Cells 4, 5 and 6 comprehends measures that can reduce the severity during the occurrence of a crash event.

In cell 4 there are measures that include the use of seat belts, helmets and protective clothing. Cell 5 can consider the crashworthiness and safety design of the vehicles and cell 6 includes elements (or lack of elements) that can influence the severity in case of a collision (guard rails, concrete barriers, street furniture, etc.). In the last row are included the measures related to the control and treatment of injuries after the event. Cell 7 presents the treatments related to the victims (hospital care and rehabilitation), cell 8 contain measures related to the safety systems of a vehicle and cell 9 can consider the general management of a crash scene (MOHAN, 2016a).

The road safety problems that happens on urban environments may differ from the ones that happens in major roadways. Even with lower operating speeds, the quantity of conflicts in urban road systems can be considerable higher when comparing to rural roadways, considering the quantity of different motorized and non-motorized transport modes that uses the infrastructure in the same time. Consequently, it is essential to follow requirements for a safe infrastructure. Three main ones are functionality, homogeneity and recognition (SWOV, 2003). Each road on a network needs to have a specific function (balance between mobility and access), with traffic distribution working as intended. Homogeneity consists in reducing the points of conflict between transport modes with great difference of mass and speed. Finally, the situations on traffic should have a level of predictability, passing what behavior should be expected from the road users.

In order to reduce road traffic crashes and possible consequent injuries or deaths, it is relevant to consider the main risk factors that causes and intensify these events. According to WHO (2004), risk in road traffic can be classified into four elements which are directly influenced: exposure, crash involvement, crash severity and post-crash severity. The risk factors that can be classified in these elements also can be distributed within the Haddon matrix as well FIGURE 10. The main risk factors related to the protection of the user consists in the usage of seat belts, helmets, air bags, child restraints and helmets. Focusing on the behavioral factors, the main ones are driving under influence of alcohol and another drugs (DUI), distraction and inattention (SHINAR, 2017).

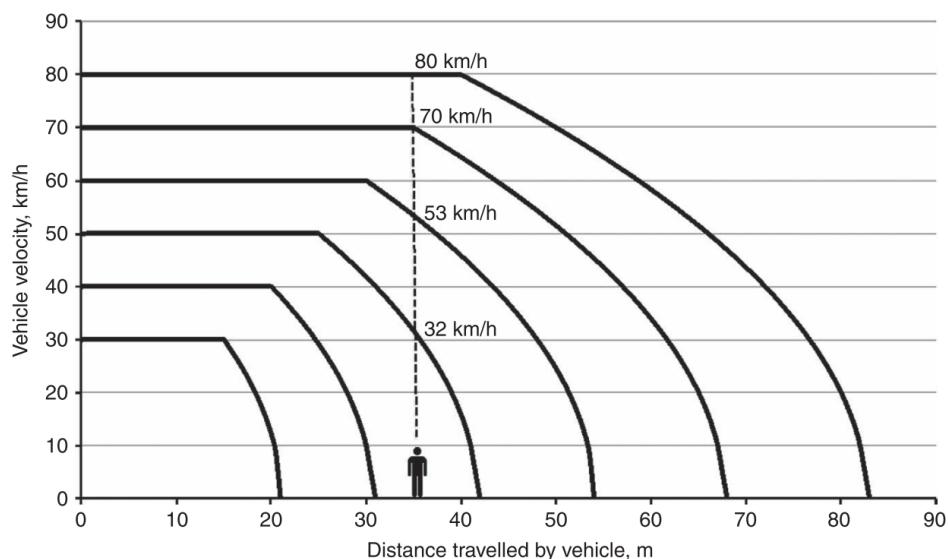
The excess of speed and the speed differential in urban environments are two risk factors that affects the chance of traffic crashes occurrence and the severity of these crashes. Speeding is a complex phenomenon with multiple causes, consequences and methods of prevention. Given its complexity and spotlight in this thesis, this risk factor will be discussed in the next section (2.2).

## 2.2 SPEEDING AS A RISK FACTOR

Speeding is one of the primary global causes of road traffic fatalities, affecting two main dimensions: probability and severity (WHO, 2013). The increase in speeds of vehicles is directly correlated to the increase in the occurrence of crashes and its average severity, hence, strongly related to the road safety (MOHAN, 2016b). Another speed-related risk factor is the speed differential, which is the speed deviation from the average operating speed (SHINAR, 2017). Ferraz et al. (2012) defines speeding as an inappropriate speed, leading to more conflict and crashes in certain traffic conditions.

The excess of speed can influence three main factors in the occurrence of a traffic conflict: reaction time, braking distance and force of impact, which is directly correlated to the severity of injuries (MOHAN, 2016b). Reducing the speed helps the driver to increase its reaction time, also helping the driver to take a correctional action to anticipate crashes (ELVIK et al., 2009). Regarding the braking distance, lower speeds reduces the distance necessary to fully stop the vehicle. FIGURE 11 shows the reaction time of a vehicle added to the braking distance, resulting in the distance travelled by the vehicle. The dashed line represents a pedestrian standing 35 meters from the traveling vehicle.

FIGURE 11 – RELATIONSHIP BETWEEN SPEED AND BRAKING DISTANCE



SOURCE: Mohan (2016b).

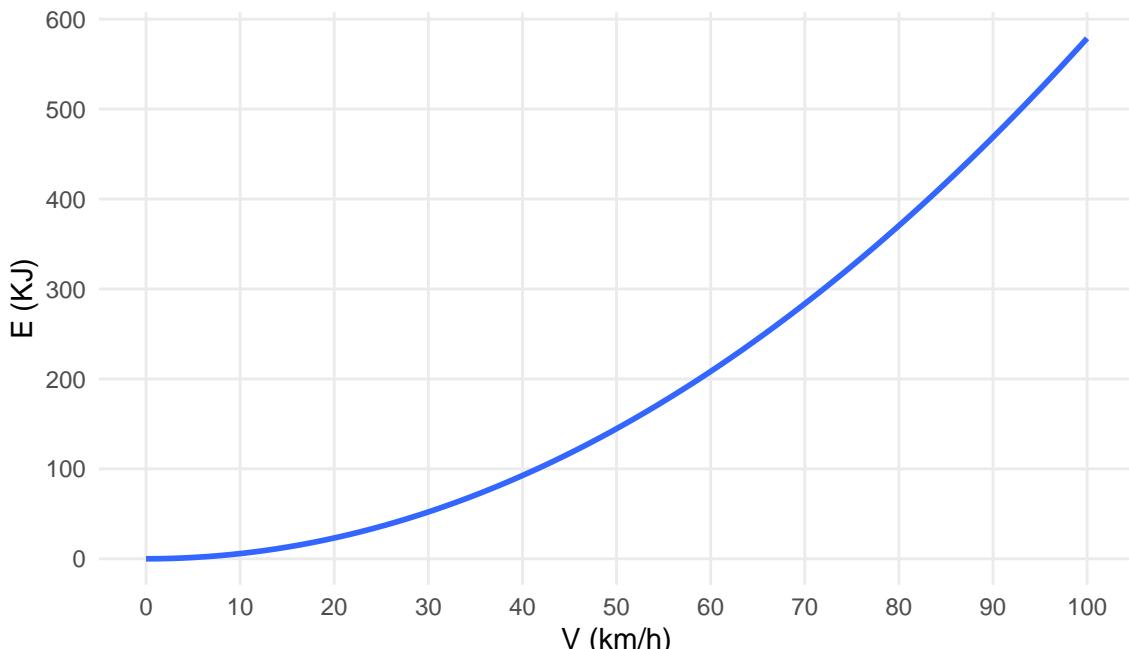
The horizontal lines from the plot shows the cruising speeds of the vehicle before they start breaking, longer lines at higher speeds represents greater reaction times. In this scenario, only speeds equal or below 40 km/h are able to avoid a collision with the pedestrian. As the vehicle speed rises, the impact speed rises as well. This leads to the increase in force of impact between vehicles and pedestrians. The severity

of injuries sustained by pedestrians depends on the energy of impact - the kinetic energy transferred to the human body. This energy of an object (the vehicle) is directly related to its velocity and mass, detailed in the following equation:

$$E = 0.5 \times MV^2; \quad (2.1)$$

where  $E$  is the kinetic energy,  $M$  is the object's mass and  $V$  is the velocity of the object. The plot in FIGURE 12 shows the increase in kinetic energy, considering an average car mass of 1,500 kg (ZERVAS; LAZAROU, 2008) and speeds varying between 0 and 100 km/h. The quantity of energy doubles when the speed changes from 40 km/h to 60 km/h, and almost triples at 70 km/h. This variation shows how the reduction of speed limits can greatly change the energy of impact.

FIGURE 12 – RELATIONSHIP BETWEEN SPEED AND KINETIC ENERGY

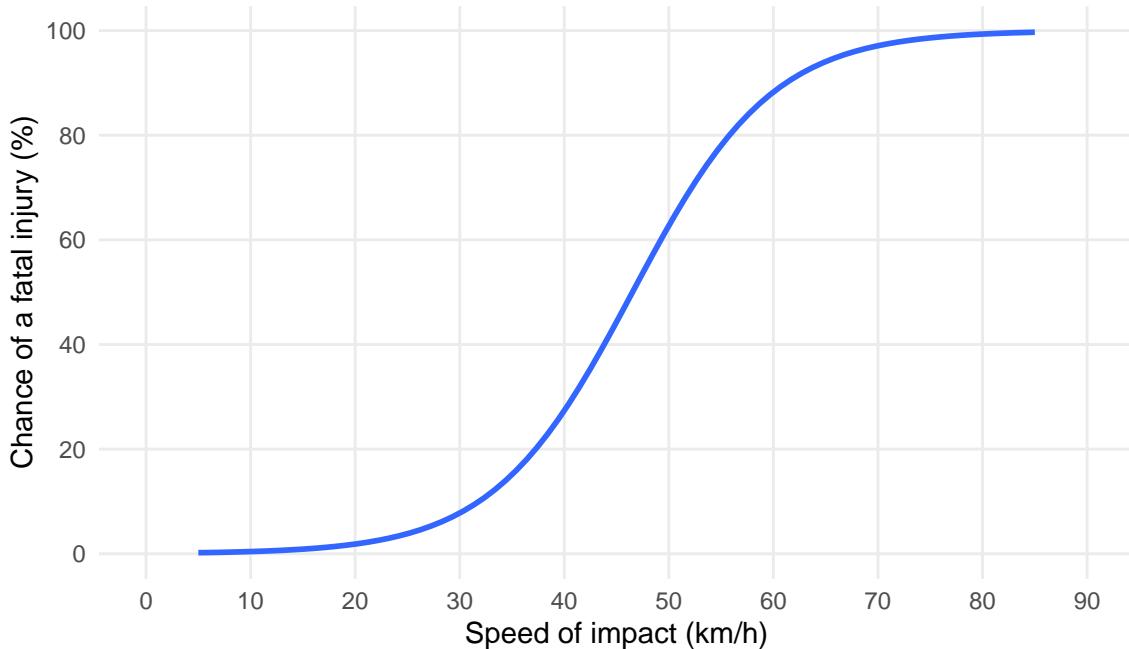


SOURCE: The Author (2021).

Considering the event of an impact between a front of a car and a pedestrian, Ashton (1980) presented the relationship between the impact speed and the percentage of fatally injured pedestrians per speed window, represented in the plot of FIGURE 13. The curve shows how a reduction from 60 km/h to 40 km/h on the speed of impact can reduce the chance of a fatal injury from 90% to 30%, approximately. Reducing the speed of impact to 30 km/h will reduce the chance of a fatal injury to approximately 10%. In general, as the speed of impact rises, the chance of a fatal injury rises too, with a greater variation between speeds of 30 km/h and 60 km/h.

The risks highlights the importance of proper speed management in road traffic, specially in urban environments. Roadways have higher operational speeds, but cities

FIGURE 13 – PROBABILITY OF PEDESTRIAN FATALITY AT DIFFERENT IMPACT SPEEDS



SOURCE: The Author (2021), based on data from Ashton (1980).

contains a more expressive interaction between motorized and vulnerable users, in which lower speeds can still represent a risk to pedestrians and cyclists. One method to manage the operating speed is the enforcement of speed limits. It is recommended a speed limit of 50 km/h in urban arterial roads, and a limit of 30 km/h in roads with a high flow of pedestrians and cyclists (WHO, 2008). Reducing the mean speeds can greatly favor the decline in fatal, non fatal and property damage only (PDO) road crashes (ELVIK, 2013).

The operating and mean speed of the road traffic depends on how the drivers chooses the speed. This choice is related to the power and stability of the user's vehicle, to road and traffic conditions, to driver's perception of safety, to the level of enforcement, travel motivations, personal characteristics and behavior of other drivers (MOHAN, 2016b; SHINAR, 2017). Considering all theses factors, it is not effective to rely only on traffic limits to prevent speeding. If the design speed of a road is higher than the speed limit and the road traffic has a low density, it is more difficult to enforce a the desired limit. The difficulty to enforce the speed limits may be higher to LMICs, which have less resources to adopt the proper road designs policies and enforcement levels required to this task (MOHAN, 2016b).

In TABLE 1 is presented a few groups of factors affecting speeding and speed choice, with its categorization based on the three main discrete factors of road crashes established by Haddon (1980): human, vehicle and environment. Analyzing the human

(or driver) factors, the background characteristics includes the level of experience, education and training of the drivers. Demographic characteristics considers the groups of age, sex and income. The general health of the driver, including the conditions of vision, hearing and sleep patterns are included in the physiological factors. The last factor in the human group - attitudes, beliefs and motivations considers how the road user perceives the control and norms present in traffic situations (RICHARD; CAMPBELL; LICHTY, et al., 2013).

TABLE 1 – FACTORS THAT AFFECTS SPEEDING AND SPEED CHOICE

<b>Human</b>	<b>Vehicle</b>	<b>Environment</b>
Background characteristics	Type/Size	Road elements
Demographic characteristics	Engine power	Weather
Physiological factors	Comfort	Traffic conditions
Attitudes, beliefs and motivations	Field of view	Surroundings
	Age	Speed Enforcement

SOURCE: The Author (2021), based on Richard, Campbell, Lichty, et al. (2013), Shinar (2017) and WHO (2008).

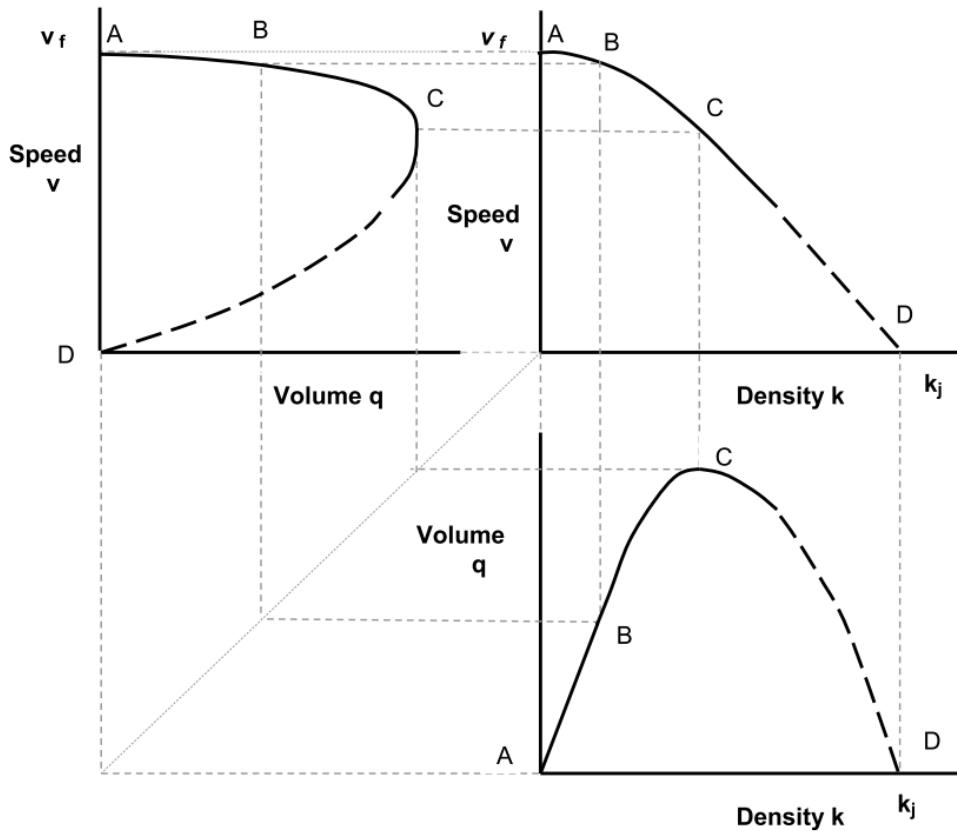
All these human factors are related to the general driver profile and its performance when driving the car. As an example of the influence in the individual differences present in the demographic characteristics, younger drivers are more likely to speed than mature and older drivers, and men are more likely to speed than women (SHINAR, 2017). The vehicle characteristics also affects the speed choice of the user. The type, engine power, comfort, field of view and age of the vehicle can affect the possibility to the driver reach the desired speed.

The environmental factors affecting speed choice consists of all the elements external to the vehicles and are greatly related to how the driver behaves and to how the vehicle is able to operate. Road elements can include the general characteristics of a road layout, quantity and type of intersections, design and level of maintenance, based on its alignment, gradient, width, surface condition, geometric pattern and road lightning. The presence of intersections and the high density of traffic controls the operating speeds in urban areas (MOHAN, 2016b). Weather can also affect how the driver will choose his speed, considering its effects on surface conditions (dry, wet, ice) and natural light.

The speed and speed variance are to a certain extent determined by the traffic flow regime (SHINAR, 2017). Depending on the level of flow and density or any other operational condition of the traffic, the driver cannot be able to reach a higher desired speed, removing his ability to have an opportunity to speed (RICHARD; CAMPBELL; LICHTY, et al., 2013; BASTOS; SANTOS, et al., 2021). The plot in FIGURE 14 illustrates the relationship between speed, density and volume (or flow) in uninterrupted flow

conditions of a certain road. This type of flow occurs in areas without interference from external factors, like intersections or crossings (GREEN; LEWIS, 2020).

FIGURE 14 – RELATIONSHIP BETWEEN TRAFFIC SPEED, VOLUME AND DENSITY



SOURCE: Green and Lewis (2020).

When density and volume are at its lowest value, the road user can reach higher speeds without having traffic to prevent it (point A). Between points A and B, volume and density starts to rise and speeds decline a little, but still is high enough to be characterized as a free flow speed. In this situation, the driver starts to experience a lack of manoeuvre freedom, but it still have the opportunity to speed. At the highest level of density (point C), speed declines as density rises. Between C and D, density rises until it reaches its maximum value, reducing overall speeds and volume (GREEN; LEWIS, 2020). In this last stage, the drivers have less flexibility to choose a desired speed, having to operate in a forced flow condition.

Established as an environmental factor affecting speeding choice (TABLE 1), the surroundings of the road, including equipments and some traffic calming techniques like gateway treatments, can be perceived by the driver as elements that gives hints to avoid higher speeds (WHO, 2008). Overall, the built environment and its five elements - density, design, diversity, destination accessibility and distance to transit - can be perceived as environmental factors which affects the road safety, through the relationship to the

numbers of road crashes and occurrence of speeding. These characteristics are better explored in the next section (2.3).

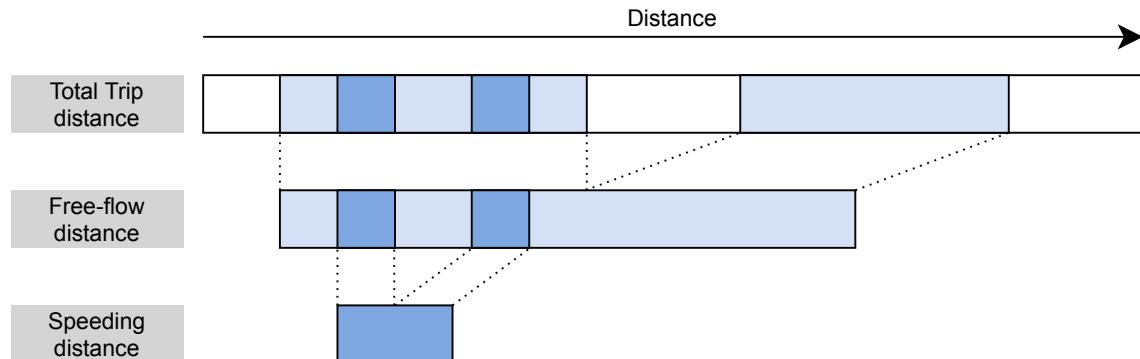
In order to collect speeding data, it is important to establish what defines speeding and to identify a measure of exposure to speeding. To Richard, Campbell, Brown, et al. (2013), defining speeding and its constitution is one of the key challenges of speeding studies. The authors presented five distinct approaches for defining speeding: ad hoc, analytical, kinematic, psychological and behavioral. The ad hoc approach assumes that the fastest driving in a sample is a proxy for speeding and considers the top X% of speeds above the posted speed. Although this approach can provide sufficient data for analysis, the speeding collected may be not related to behavior or safety.

The analytical approach fixes a criterion relative to the speed limit (+ X% or + X km/h). It is a simple method to implement, but can miss some factor related to road type and overall speed level. The analytical approach is used in this work and its process of implementation is detailed in Section 3.2. The kinematic approach considers the driving environment and its operational conditions, in order to connect the speed behavior to specific situations. This approach requires a certain level of quantity and quality of GIS data to support it. The psychological approach is based on what the driver considers to be speeding, providing insight into factors that are related to individuals perception of speeding. The necessity of additional previous information about the speeding beliefs and attitudes of the drivers can be a disadvantage of this approach. At last, the behavioral approach sets four speed bands to describe different behavior in each one, related to the chosen speed by the driver (RICHARD; CAMPBELL; BROWN, et al., 2013).

The operational aspects of urban road traffic (mostly in interrupted flow conditions) arbitrary reduces the amount of speeding being measured. To address this problem, it is necessary to extract the free-flow situations of the trip from the total trip distance, removing sections in which drivers had no opportunity to speed. This process of extracting parts of the trip in which the driver is exposed to speeding is presented in FIGURE 15. Therefore, the actual measure of speeding is distance performed in speeds above the posted speed compared to the distance performed in free-flow speed. This process was applied in this work and it is described in Section 3.2.

The data collected also depends on the method applied. The collection of speed data can be done with the use of speed traps - fixed or mobile (HIDALGO-SOLÓRZANO et al., 2020; WHO, 2008), the analysis of road crashes reports (WATSON et al., 2015), roadside observational studies (SHINAR, 2017), questionnaires (DINH; KUBOTA, 2013),

FIGURE 15 – EXTRACTION OF FREE-FLOW AND SPEEDING EPISODES FROM TRIPS



SOURCE: The Author (2021), based on Richard, Campbell, Brown, et al. (2013).

smartphone data (WARREN; LIPKOWITZ; SOKOLOV, 2019) and GPS data collection (MORENO; GARCÍA, 2013; WANG et al., 2018). Studies that uses naturalistic data includes the capture video data from the drivers and GPS can relate the speeding action registered with a certain behavioral action from drivers (BASTOS; SANTOS, et al., 2020) and environmental factors (MORENO; GARCÍA, 2013). The speeding data collected from driving simulation studies also can be correlated with other environmental or behavioral factors (YADAV; VELAGA, 2020). The aspects of naturalistic driving studies and its comparison to other methods will be better explored in the Section 2.4.

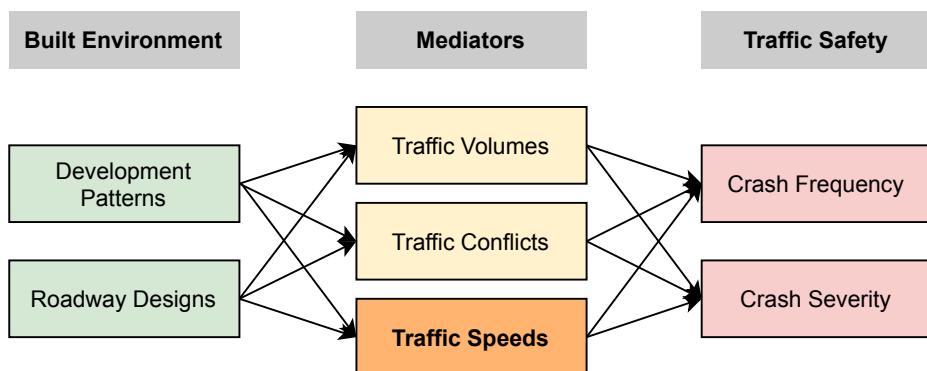
Knowing the factors that affects speeding and the methods to collect and analyse speeding data, it is possible to create countermeasures to this risk factor in road traffic crashes. Following the categories established by Haddon (1980) and presented in TABLE 1, these countermeasures can be classified into behavioral (human), vehicular and environmental approaches. Behavioral approaches includes the education and training to teach speed awareness to drivers. Enforcement is also a behavioral approach. Enforcement by the police or other traffic authority is related to a greater rate of speed limit compliance, but it is a highly localized measure (SHINAR, 2017).

The vehicle approach to intervene in speeding practices includes speed limits and advisory systems in cars. Environmental approaches are consisted of direct changes in infrastructure, including traffic calming measures, and policy approaches, including the establishment of speed limits and administrative actions. Speed bumps, lane narrowing, roundabouts and rumble strips are some of the traffic calming techniques that are capable to reduce speeds in urban environments (WELLE et al., 2016). Other key elements of urban design also can intervene in the practice of speeding, including street block sizes, road network connectivity and road lane width.

## 2.3 ROAD SAFETY PERFORMANCE AND THE BUILT ENVIRONMENT

The built environment consists of the physical elements and features, including the development patterns and the roadway designs of a city. Built environment affects the physical activity inside the city, including the overall mobility, road safety and health. According to Ewing and Dumbaugh (2009), the frequency and severity of road traffic crashes are related to the built environment through three mediators: traffic volumes, traffic conflicts and traffic speeds, as shown in FIGURE 16. Therefore, the land use and transport plans influences the choice of destination, travel models, travel distances and routes, impacting the road traffic safety (TIWARI, 2016).

FIGURE 16 – BUILT ENVIRONMENT AND TRAFFIC SAFETY



SOURCE: The Author (2021), based on Ewing and Dumbaugh (2009).

Traffic volume is directly related to crash frequency and the increase of vehicle miles travelled (VMT), which is a exposure variable to the occurrence of road traffic crashes. Traffic speed (and its excess), the main factor investigated in this paper, is affected by the operating and encouraged speed defined by the roadway design and development pattern. Traffic conflicts are related to the increase in vehicle flow and traffic speed (EWING; DUMBAUGH, 2009). Overall, the built environment can be characterized into five “D” variables: (i) density, (ii) diversity, (iii) design, (iv) destination accessibility and (v) distance to transit, being defined as the 5D (EWING; CERVERO, 2010). Demographics of residents, while not being part of the built environment, can be considered as the sixth “D”.

TABLE 2 contains the categories previously described and its variables. It is important to note that it is not presented all the possible built environment variables, only the main ones found in previous papers. In addition, these are rough categories, divided by boundaries that are subject to change (EWING; CERVERO, 2010). Density measures a variable of interest per unit of area and can include the quantity of population, housing or number of jobs, showing the “hotspots” of the overall activity in a city. Diversity refers to different land use types within a given area, ranging from single-use environments

to more varied land uses. The street network and its characteristics is included in the design category, and considers elements like block size, type / quantity of intersections, street widths, quantity of pedestrian crossings and street network density.

TABLE 2 – BUILT ENVIRONMENT VARIABLES

Category	Variable
Density	Population
	Housing
	Employment
Diversity	Land use
	Block size
	Type / Quantity of intersections
	Street widths
	Quantity of pedestrian crossings
Design	Road hierarchy
	Street network density
	Quantity of commercial and service units
	Distance to city center
	Quantity of jobs
Destination Accessibility	Quantity of bus stops
	Distance to nearest transit stop
Demographics	Income
	Age
	Sex

SOURCE: The Author (2021), based on Ewing and Dumbaugh (2009), Ewing and Cervero (2010) and Obelheiro, Silva, Nodari, et al. (2020).

Destination accessibility measures how easy is to access places of attraction, including jobs, services, commerce and districts of overall interest, considering business districts and city downtown. Distance to transit usually is measured by the level of offer of transport services, being measured by the quantity of transit stops and network density. Network density also can be included in the design category. Demographics, while not being a characteristic of the build environment, can influence the occurrence of speeding in urban environments, including income, age and sex. In the following tables (TABLE 3 to TABLE 8) are presented the built environment categories and its variables considered in this work, including the authors that investigated them related to the road safety performance in urban areas, with the focus on one final outcome indicator - road crashes and one mediator - occurrence of speeding.

TABLE 3 displays a list of authors which investigated the relationship between population density and road safety in urban areas. Regarding road traffic crashes, Dumbaugh and Rae (2009) and Obelheiro, Silva, Nodari, et al. (2020) found a inverted correlation between number of habitants per area and occurrence of road crashes. Higher densities environments causes people to drive less (DUMBAUGH; RAE, 2009), reducing the exposure to road traffic accidents. A research considering road safety and BEs about another brazilian city - Porto Alegre - also presented a result of a inverted

relationship between injury crashes and population density in most of the city's area.

TABLE 3 – CORRELATION BETWEEN DENSITY AND ROAD SAFETY OUTCOMES

Category	Variable	Road safety outcomes	
		Crashes	Speeding
Density	Population	(–) Dumbaugh and Rae (2009)	
		(+) Dumbaugh, Li, and Joh (2013)	
		(+) Lee, Abdel-Aty, and Jiang (2015)	
		(–) Obelheiro, Silva, Nodari, et al. (2020)	—
		(+) Pirdavani et al. (2014)	
		(±) Welle et al. (2016)	

SOURCE: The Author (2021).

NOTE: (+): positive correlation between variable and outcome; (–): inverted correlation between variable and outcome; (±): correlation not clear / depends on other factors.

Dumbaugh, Li, and Joh (2013), Lee, Abdel-Aty, and Jiang (2015) and Pirdavani et al. (2014) found a direct relationship between population density and the occurrence of road crashes. There is a positive relationship between population density and total pedestrian crashes (DUMBAUGH; LI; JOH, 2013), in addition to motor vehicle crashes and bicycle crashes (LEE; ABDEL-ATY; JIANG, 2015) and injury crashes (PIRDAVANI et al., 2014). To Welle et al. (2016), higher density areas contains less motorized trips, which can reduce crashes involving cars, but can increase the number of conflicts within the area, therefore increasing the risk of overall crashes. In regard to the chance of speeding occurrence, a study relating population density as a BE variable to the speeding practice could not be found. Most authors cites the speeding occurrence as a risk factor for these crashes. The statistical relationship between BE variables and the occurence of motor vehicle speeding still needs to be better explored. It can be assumed that higher density areas can lead to a higher density of traffic vehicles, inducing a interrupted flow and reducing the capacity to reach free-flow speeds and speeds above the posted limits.

TABLE 4 contains previous works that discussed the relationship between BE diversity and road safety outcomes, focusing on the diversity of the land use in urban areas. In relation to road crashes, all listed authors found a direct correlation between diversity and the safety outcome. To Obelheiro, Silva, Nodari, et al. (2020) and Obelheiro, Silva, and Nodari (2019), more diverse environments can lead to higher traffic volumes and conflicts, due to the attraction of distinct type of users, leading to higher chance of accidents. A higher level of mixed land use is correlated with more crashes from multiple categories: pedestrian, bicycles, injuries and fatal crashes (OYANG; BEJLERI, 2014). The same relation was found by Rhee et al. (2016) considering injury and fatal crashes in the city of Seoul, in Korea.

Regarding the occurrence of speeding, it could not be found previous papers

TABLE 4 – CORRELATION BETWEEN DIVERSITY AND ROAD SAFETY OUTCOMES

Category	Variable	Road safety outcomes	
		Crashes	Speeding
Diversity	Land use	(+) Amoh-Gyimah, Saberi, and Sarvi (2017)	
		(+) Obelheiro, Silva, and Nodari (2019)	
		(+) Obelheiro, Silva, Nodari, et al. (2020)	—
		(+) Ouyang and Bejleri (2014)	
		(+) Rhee et al. (2016)	

SOURCE: The Author (2021).

NOTE: (+): positive correlation between variable and outcome; (-): inverted correlation between variable and outcome; (±): correlation not clear / depends on other factors.

that have a statistical investigation relating this BE category and speeding data. A hypothesis that can be explored is the idea that areas with higher level of mixed land use can lead to more attraction and generation of motorized and non-motorized trips. Therefore, the increase in conflicts and traffic control can induce the reduction of average speeds. Equal to the population density, this correlation still needs to be better explored.

TABLE 5 contains previous works that discussed the relationship between road safety outcomes and five design variables: intersections, speed cameras, signalized intersections, arterial roads and street network density. With respect to the density of intersections, Dumbaugh and Li (2011), Dumbaugh, Li, and Joh (2013), Elvik et al. (2009) and Huang, Wang, and Patton (2018) correlates this variable directly with the occurrence of road traffic crashes. 4-way intersections were found to have a direct correlation to pedestrian and bicycle crashes (DUMBAUGH; LI, 2011; DUMBAUGH; LI; JOH, 2013) and to all vehicle categories (HUANG; WANG; PATTON, 2018).

Diversely, three other authors discuss about a inverse correlation between intersections and number of road traffic crashes. The density of intersections is correlated to less road crashes in all levels of severity (MARSHALL; GARRICK, 2011; OUYANG; BEJLERİ, 2014) and non-motorized crashes (ZHANG, Y. et al., 2015). Considering the type of intersection, 4-way intersections can lead to more crashes, but 3-way intersections leads to the reduction of crashes (EWING; DUMBAUGH, 2009). Intersections can have mixed effects on crash incidence. The increase in conflicts caused by intersections can increase the incidence of road crashes, but reduces the quantity of more severe crashes, considering the reduction of speeds.

Regarding speeding outcomes, all listed authors correlated the quantity of intersections with less occurrence of speeding. Higher number of intersections decreases traffic flow and average speeds (ELVIK et al., 2009; EWING; DUMBAUGH, 2009). The quantity of intersections can be related to block sizes. More intersections mean shorter length of road sections between nodes, reducing the opportunity for the drivers to reach

TABLE 5 – CORRELATION BETWEEN DESIGN AND ROAD SAFETY OUTCOMES

<b>Variable</b>	<b>Road safety outcomes</b>	
	<b>Crashes</b>	<b>Speeding</b>
Density of intersections	(±) Dumbaugh and Rae (2009)	(–) Dumbaugh and Rae (2009)
	(+) Dumbaugh and Li (2011)	(–) Elvik et al. (2009)
	(+) Dumbaugh, Li, and Joh (2013)	(–) Ewing and Dumbaugh (2009)
	(+) Elvik et al. (2009)	(–) Huang, Wang, and Patton (2018)
	(±) Ewing and Dumbaugh (2009)	(–) Obelheiro, Silva, and Nodari (2019)
	(+) Huang, Wang, and Patton (2018)	(–) Obelheiro, Silva, Nodari, et al. (2020)
	(–) Marshall and Garrick (2011)	
	(–) Ouyang and Bejleri (2014)	
Density of speed cameras	(–) Y. Zhang et al. (2015)	
	(–) Høye (2015)	(±) Li, Zhang, and Ren (2020)
	(–) Li, Graham, and Majumdar (2013)	(–) Li, Graham, and Majumdar (2013)
	(–) Li, Zhang, and Ren (2020)	(–) Oliveira et al. (2015)
Traffic signal density	(+) Park, Park, and Kwon (2019)	
	(+) Obelheiro, Silva, Nodari, et al. (2020)	(±) Elvik et al. (2009)
	(+) Lovegrove and Sayed (2006)	(±) Furth et al. (2018)
Proportion of arterial roads	(+) Lee, Abdel-Aty, and Jiang (2015)	
	(+) Dumbaugh and Rae (2009)	(+) Dumbaugh and Li (2011)
	(+) Dumbaugh and Li (2011)	(+) Dumbaugh, Li, and Joh (2013)
	(+) Dumbaugh, Li, and Joh (2013)	(+) Ewing and Dumbaugh (2009)
	(+) Ewing and Dumbaugh (2009)	(+) Huang, Wang, and Patton (2018)
	(+) Huang, Wang, and Patton (2018)	(+) Obelheiro, Silva, Nodari, et al. (2020)
	(+) Obelheiro, Silva, Nodari, et al. (2020)	(+) Welle et al. (2016)
	(+) Ukkusuri et al. (2012)	
Network density	(+) Welle et al. (2016)	
	(+) Yu and Xu (2017)	
Network density	(–) Marshall and Garrick (2010)	—
	(–) Marshall and Garrick (2011)	

SOURCE: The Author (2021).

NOTE: (+): positive correlation between variable and outcome; (–): inverted correlation between variable and outcome; (±): correlation not clear / depends on other factors.

higher desired speeds.

In regard to the correlation between speed cameras and road crashes, Park, Park, and Kwon (2019) observed a direct correlation between the implementation of speed cameras and overall number of road crashes, but the severity of these crashes reduced. To Høye (2015) and Li, Graham, and Majumdar (2013), speed cameras reduced the occurrence of road crashes inside a certain buffer, with its effect decreasing as distance from cameras increase. In addition, the crash reducing effect of speed cameras decrease after a certain time of installation (LI; ZHANG; REN, 2020). Li, Graham, and Majumdar (2013) and Oliveira et al. (2015) detected a speed reduction inside a 200m buffer from the camera. To Li, Zhang, and Ren (2020), the speed before and near the camera is reduced, but the average speeds can increase after the speed camera, due to the “kangaroo effect”, where the road users tends to compensate the “wasted” time caused by the reduction of speeds when passing through the speed cameras.

The quantity of traffic signals can have mixed effects on speeding and presented a positive correlation in all listed works. Lovegrove and Sayed (2006) and Lee, Abdel-Aty, and Jiang (2015) observed a positive correlation between the density of traffic signals and crashes from motorized and non-motorized vehicles. To Obelheiro, Silva, Nodari, et al. (2020), traffic signals are usually installed in areas with more traffic conflicts, therefore it can act as a proxy for higher traffic volumes and exposure to crashes. In respect to speeding, the effects of traffic signals depends on the cycle length, progression speed and space between signalized intersections. Longer cycles creates better speeding opportunities, in addition to closer intersection spacing (ELVIK et al., 2009; FURTH et al., 2018).

Across all the listed works, it is unanimous the identification of a positive correlation between arterial roads and both road safety outcomes: occurrence of road crashes and speeding. Arterial roads favors mobility and higher operating speeds, regularly containing a higher number of lanes and larger lanes, in comparison to collector and local roads. This increase in the occurrence of crashes is directly related to the increase in overall speed and the occurrence of speeding episodes. As discussed in the last section, higher speeds can increase the number and the severity of road traffic crashes.

Finally, the last variable listed on TABLE 5 is street network density. To Marshall and Garrick (2010), higher risk of fatal or severe crashes occurs in areas with very low street network density, and the occurrence of crashes is higher in these areas (MARSHALL; GARRICK, 2010). Regarding the occurrence of speeding, studies correlating the two variables could not be found. It can be possible to relate the street network density to the intersection density, therefore make an hypothesis of a inverted correlation between network density and speeding, but further investigation it is still needed.

TABLE 6 contains a couple of studies that discuss the correlation between density of commercial and services units, as a variable that measures the destination accessibility, and road traffic crashes, a road safety outcome. To Ouyang and Bejleri (2014) and Welle et al. (2016), areas with easy access to destination generates less vehicle-kilometers travelled by motorized vehicles, therefore reducing exposure to road traffic crashes. The correlation between density of commercial and services units and occurrence of speeding still needs further investigation. The reduction in use of cars in areas with better accessibility can be related to road infrastructure less focused on motorized vehicles and more focused on the well-being of pedestrians and cyclists, leading to the reduce in speeding.

The listed studies in TABLE 7 shows a direct correlation between bus stop

TABLE 6 – CORRELATION BETWEEN DESTINATION ACCESSIBILITY AND ROAD SAFETY OUTCOMES

<b>Category</b>	<b>Variable</b>	<b>Road safety outcomes</b>	
		<b>Crashes</b>	<b>Speeding</b>
Destination Accessibility	Density of commercial and services units	(–) Ouyang and Bejleri (2014) (–) Welle et al. (2016)	—

SOURCE: The Author (2021).

NOTE: (+): positive correlation between variable and outcome; (–): inverted correlation between variable and outcome; (±): correlation not clear / depends on other factors.

density, as a measure of distance to transit, and road crashes. Bus stops leads to more traffic conflicts, being a increased activity generator, both as origin and destination to pedestrians (KIM; PANT; YAMASHITA, 2010). In regard to speeding, the presence of bus stops in a road can reduce the free-flow speed of other motorized vehicles (BANSAL; AGRAWAL; TIWARI, 2014; KOSHY; ARASAN, 2005), therefore reducing the opportunity of speeding.

TABLE 7 – CORRELATION BETWEEN DISTANCE TO TRANSIT AND ROAD SAFETY OUTCOMES

<b>Variable</b>	<b>Road safety outcomes</b>	
	<b>Crashes</b>	<b>Speeding</b>
Bus stop density	(+) Obelheiro, Silva, Nodari, et al. (2020) (+) Ouyang and Bejleri (2014) (+) Wei and Lovegrove (2013) (+) Kim, Pant, and Yamashita (2010)	(–) Bansal, Agrawal, and Tiwari (2014) (–) Koshy and Arasan (2005)

SOURCE: The Author (2021).

NOTE: (+): positive correlation between variable and outcome; (–): inverted correlation between variable and outcome; (±): correlation not clear / depends on other factors.

TABLE 8 presents studies which discussed the correlation between road crashes and level of income. Higher levels of income is related to less road traffic crashes. To Obelheiro, Silva, and Nodari (2019), the income level of a region can directly represent the overall quality of its road system infrastructure. The age and maintenance level of vehicles can be related with the level of income, which as a risk factor as well. In consideration to the occurrence of speeding, it could not be found a study which discusses its correlation to the level of income in an area.

Most of the built environment variables discussed are influenced by the planning process of a city. Master plans establishes built area densities and type of activities that should be permitted in specific areas. Therefore, the exposure to traffic risks is influenced by these policies. The travel demand is a function of the level of economic activities and income level, and its direction depends on the local of the economic activities, representing the land use patterns. Different types of travel and flows can

TABLE 8 – CORRELATION BETWEEN DEMOGRAPHICS AND ROAD SAFETY OUTCOMES

<b>Category</b>	<b>Variable</b>	<b>Road safety outcomes</b>	
		<b>Crashes</b>	<b>Speeding</b>
Demographics	Income	(–) Obelheiro, Silva, and Nodari (2019) (–) Marshall and Ferenchak (2017)	—

SOURCE: The Author (2021).

NOTE: (+): positive correlation between variable and outcome; (–): inverted correlation between variable and outcome; (±): correlation not clear / depends on other factors.

have different and conflicting requirements, a scenario of conflict which is common in cities (TIWARI, 2016).

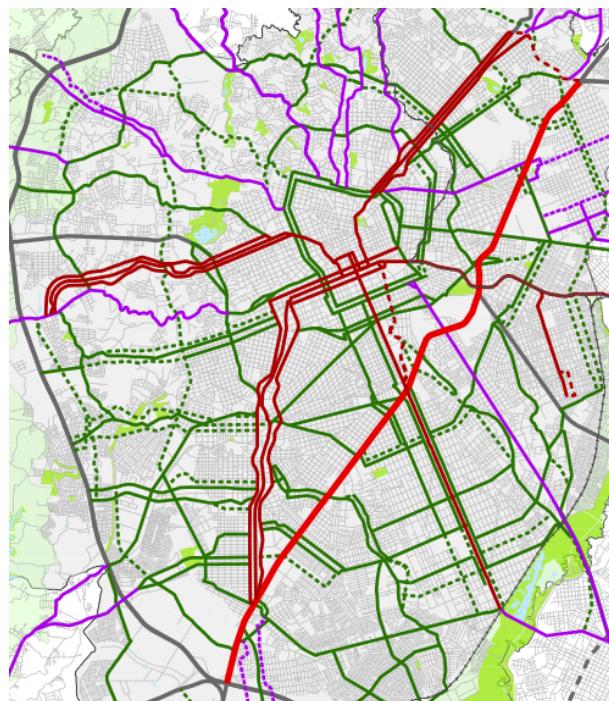
In 1942, the Agache Plan established a radial expansion to the city of Curitiba's territory. In the 1960s happened an unprecedented growth of the city's population, unforeseen by the Agache Plan, creating new housing and mobility problems to the city. Having these problems in mind, in 1965 was created the Instituto de Pesquisa e Planejamento de Curitiba (Institute of Research and Planning of Curitiba - IPPUC), in order to develop a new and formal master plan. The main cornerstone of the plan was the integration between mobility and land use, consisting of 5 main corridors of growth. These corridors contains mixed-use, high-rise developments and bus rapid systems (BRT) to attend this new demand (ROSÁRIO, 2016).

Curitiba's master plan promotes the integration between the road system, transport and land use, as a urban development policy (CURITIBA, 2015). The land use integration with mobility is lead by the road hierarchy and the distribution of population and economic activities into areas with required infrastructure. Zoning in Curitiba considering a macro scale is driven by the structural axes and densification axes. Structural axes are the main axes of city growth, containing areas that expands the traditional city center and creating corridors of mixed land use of high density, supported by the mobility and public transport infrastructure. Theses axes are shown in red inside the map from FIGURE 17.

In addition, the Densification Axes acts as a complement to this urban structure, including a mixed land use and medium levels of density. Regarding the density, Curitiba (2015) establishes a low density as 80 residences per hectare, medium density between 80 and 200 residences per hectare and high density between 200 and 400 residences per hectare. FIGURE 18 shows how these higher densities are set near the structural axes. Areas in red represent the higher density areas.

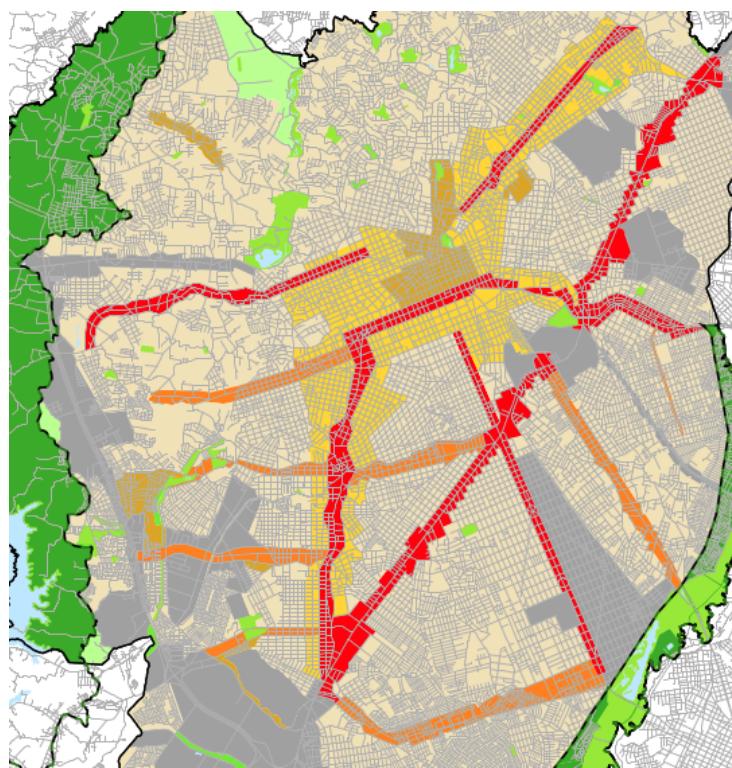
Curitiba's land zoning plan splits the micro zoning system into four categories: central, residential, mixed use and specific destination (CURITIBA, 2019). Central zoning involves the traditional downtown of the city, characterized by a high concentration of

FIGURE 17 – STRUCTURAL AXES IN CURITIBA



SOURCE: Curitiba (2015).

FIGURE 18 – LAND USE DENSITIES IN CURITIBA



SOURCE: Curitiba (2015).

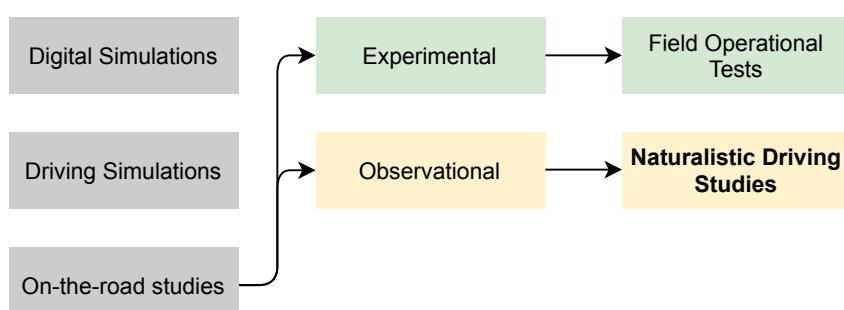
services, commerces and jobs. Residential zones compasses different types, depending on the density. It can include commercial units and limited occupancy rate, depending on the environment characteristics. Mixed use zones encourages the coexistence of residential and non-residential uses, mixing residences, services and commercial units of medium density. Specific destination zoning can include educational, military, historic and industrial uses, among others.

From decades ago until today, the majority of the urban environment was built for and around the needs of cars. Building and land use codes contribute to the causes of road traffic accidents, including speeding (KNOFLACHER, 2016). Cities and its decision makers needs to mind the traffic safety problems that are generated by the planning practices, as well the vantages of creating new spaces and land patterns that favors the safety inside the city.

## 2.4 NATURALISTIC DRIVING STUDIES

Human behavior researches that focus on road traffic safety have multiple methods, including digital simulations, driving simulators and on-the-road studies. Digital simulations studies consists on the modelling of a hypothetical situation based on statistical and microssimulations functions which can be run multiple times on a software. Driving simulation studies contains a physical mockup of a real vehicle, where the subjects drives in accord to the computer-generated scenario. On-the-road studies consists of scenarios with real cars and traffic, and can be categorized into two types: experimental and observational studies (SHINAR, 2017).

FIGURE 19 – BEHAVIORAL TRAFFIC SAFETY STUDIES



SOURCE: The Author (2021), based on Shinar (2017).

Experimental studies contains some level of manipulation or control of the situation, where the independent variable is manipulated. This allows to better analyze the involved variables in a individual matter, but creates a artificial scenario, different from everyday situations. One popular type of experimental method is the field operational test (FOT). In FOT, volunteers are provided with cars equipped with multiple unobtrusive

sensors, in order to investigate how they react to a specific system installed by the researchers. Diversely, observational studies simply observes the behavior of drivers in natural conditions, without the control of dependent and independent variables (SHINAR, 2017).

A type of observational study is the naturalistic driving study (NDS). NDS consists in monitoring drivers behavior in their own cars, both in normal and safety critical conditions. Volunteer's cars are usually equipped with video cameras, GPS systems and other sensors. The objective of the NDS is to record the driver's behavior during a everyday scenario. The cameras allows the recording of the driver, the car panel, instruments and the recording of the surrounding areas, including other vehicles, pedestrians and environmental elements. GPS or similar equipments allows register of the driver's instant speed and position (SHINAR, 2017). Therefore, NDS is a method that can collect information about the environment, the vehicle and the human behavior.

The main difference between FOT and NDS is the element of study. FOT's usual objective is to evaluate a particular system or instrument inside or outside the vehicle. NDS' main objective is to record typical driving behaviors in multiple situations, including ones that presents a risk for the occurrence of road traffic crashes, like speeding occurrence, use of mobile phones, etc. NDS generates a great number of data in comparison to other methods like FOT and driving simulators, but the method has its disadvantages as well.

NDS in behavioral studies could be considered as an "improvement" in comparison to driving simulation studies, but its advantages and disadvantages makes them complementary methods, not competitive. In a driving simulation, the sample size is smaller, and the immersion is not perfect, given the possibility of the occurrence of motion sickness and adaptation period of the volunteers. But the main strength of driving simulators is the fact that it creates and controls the events and stimulus occurred in the scenario, making possible to better isolate variables and replicate crashes, near crashes and critical events (SHINAR, 2017).

In comparison, NDS have a typically large sample size and gives practically total immersion to the volunteers' everyday trips. The events occurred in traffic are totally uncontrolled and real. This gives a higher "credibility" to the data collected, but makes it harder to isolate the desired variables in study. The exposure to the experiment is usually higher in NDS, where the volunteers have the data registered over multiple trips (CARSTEN; KIRCHER; JAMSON, 2013). Overall, the NDS characteristics and advantages are valuable to the present study, considering its capability of collecting a comparably large dataset of speeds and location throughout the city's territory.

TABLE 9 presents some of the main NDS researches that were established in the last two decades. The 100-Car Naturalistic Driving Study, conducted by the Virginia Tech Transportation Institute, was the first large scale NDS and was performed in the United States. Its instrumentation was composed of 5 cameras, 1 accelerometer and 1 GPS sensor, among others. Some of the vehicles were owned by the researchers and others by the volunteers. The main location of study was the metropolitan of northern Virginia and Washington, DC. The main goal of the research was to maximize the potential to record crash and near-crash events. In total, 100 vehicles were used, with 241 drivers among primary and secondary volunteers. The total trip time of the data collected was above 40,000 hours, with a distance above 3 million kilometers (NEALE et al., 2005).

TABLE 9 – NATURALISTIC DRIVING STUDIES PROGRAMS

<b>Program</b>	<b>Countries</b>	<b>Vehicles</b>	<b>Drivers</b>	<b>Hours</b>	<b>Distance [km]</b>
100 Car NDS	USA	100	241	42,300	3,218,688
SHRP2 NDS	USA	3,500 <sup>(2)</sup>	3,500 <sup>(2)</sup>	1,000,000 <sup>(2)</sup>	56,327,040
ANDS	Australia	346	409	716,320	1,512,630
UDRIVE	EU <sup>(1)</sup>	192	287	87,870	4,000,000 <sup>(2)</sup>
CNDS	Canada	140	149	53,000 <sup>(2)</sup>	1,800,000 <sup>(2)</sup>

SOURCE: The Author (2021).

NOTES: <sup>(1)</sup>: UK, Netherlands, France, Spain, Germany and Poland; <sup>(2)</sup>: Approximate values.

Another NDS study performed in the United States was the SHRP2 (Strategic Highway Research Program-2) NDS, gathering a massive quantity of data. The study was conducted in 6 states, with more than 3,500 vehicles, million hours of video and trip data collected and 56 million kilometers travelled among all the trips. More than 1,500 crashes were registered, establishing a valuable sample for road safety studies (NJORD; STEUDLE, 2015). The Australian Naturalistic Driving Study (ANDS) launched in 2015, aiming to collect road safety data from 360 vehicles - 180 in Sydney and 180 in Melbourne. The vehicles are equipped with 4 cameras, GPS sensors and accelerometer (ANDS, 2017). Accordingly to Larue et al. (2019), the ANDS was able to collect data from 346 vehicles and 409 participants. The last report related a collection of more than 700,000 hours of travelled period and more than 1,500,000 kilometers of travelled distance (ANDS, 2017).

In the European Union, the UDRIVE naturalistic driving study was large scale NDS program installed in 6 EU countries: UK, Netherlands, France, Spain, Germany and Poland. It lasted for 5 years, between 2012 and 2017, and its objective was to collect behavioral data from three transport modes: cars, trucks and powered two-wheelers. 192 vehicles and 287 drivers were part of the research. The data collected reached almost 90,000 hours of driving data and approximately 4,000,000 kilometers of

travelled distance (NES et al., 2019). The Canadian Naturalistic Driving Study (CNDS) instrumented a total of 140 vehicles, including cars, pickup trucks and minivans, with 4 cameras, a frontal radar sensor and one GPS radar. 149 drivers participated in the study, with approximately 53,000 hours of data collected and 1,800,000 kilometers of travelled distance (CNDS, 2021).

In Brazil, a naturalistic driving study was established in 2019 in the city of Curitiba-PR, conducted by the Universidade Federal do Paraná. The Naturalistic Driving Study - Brazil (NDS-BR) has already collect data from 16 instrumented vehicles, each with a kit of 3 cameras and 1 GPS sensor. Currently, NDS-BR collected almost 240 hours of driven time and 5,000 km of travelled distance in Curitiba and metropolitan area. This research its still ongoing, and further details about the process of data collection will be discussed in Section 3.1. In all the previous NDS researches described, the collection of speed data was a possible due to the GPS sensors or accelerometers.

NDS allows the investigation of speeding episodes and how other factors can influence the occurrence of speeding. These factors can include environmental elements (road infrastructure, built environment, weather, traffic conditions, etc.), human and behavior elements (age, gender, secondary tasks engagement, education, personality) and vehicle characteristics (engine power, level of conservation). Richard, Campbell, Brown, et al. (2013), Richard, Brown, et al. (2017) and Richard, Lee, et al. (2020) utilized NDS speed data to better understand different types of speeding and how the drivers behaves towards them. Richard, Campbell, Brown, et al. (2013) established a classification called behavioral speeding, splitting it into 4 types. The first type consists of speed below the posted limit. In the second type, drivers are speeding but they feel like they are driving at a safe speed. On the third type, speeds reaches above the enforcement limit, where the driver starts to accept a higher level of risk. The 4th type consists of reckless speeding, where the driver reaches speeds way above the posted speed, without any consideration to the risks.

Richard, Brown, et al. (2017) clustered speeding behavior into 6 categories, from a NDS data collected from 88 drivers in Seattle. The authors used the k-means method on the sample, resulting in 6 clusters: (i) Speeding up, (ii) speed drop, (iii) incidental speed, (iv) casual speeding, (v) cruising speeding and (vi) aggressive speeding. Speeding up behavior consists on reaching a high exceeding speed in a short period of time. with a high variability. Speed drop is complementary to speeding up, representing characteristics that are consistent with a braking behavior when entering a lower posted speed zone.

Incidental speeding was the most common speeding type detected in the

sample. It consists of mean speeds barely reaching above the posted speed, occurring in longer periods of time. Casual speeding was the second most common type, and has a similar behavior to incidental speeding, but has a higher level of mean speed above the posted speed, with similar lengths of occurrence. Cruising speeding is a deliberate type of speeding, most common on freeways and state highways - places with more opportunity for speeding. Aggressive speeding presents higher levels of mean speed, with a high variability in longer periods of time. In this cluster, drivers engage in repeated speed adjustments, always trying to reach maximum speed.

Richard, Lee, et al. (2020) used a sample of 100 trips from the SHRP2 NDS database in order to examine how the occurrence of speeding episodes are affected by basic factors. The authors found drivers in the age group between 16 to 24 had more speeding episodes than other age categories. Considering the days of the week, the occurrence of speeding was lower between Monday and Thursday. Highways present higher free-flow and speeding episodes in comparison to other type of roads.

Ellison and Greaves (2015) analyzed speeding data from 106 drivers in Sydney, collected with the use of NDS, in order to establish how much time drivers are saving with speeding in urban environments. The speeding stretch of a trip was compared to a simulated stretch with the posted speed traveling the same length. Then, the time delta between the two scenarios was calculated to create the time saved during the speeding episode. The results showed an average time saving of 26 seconds per day, or 3 minutes per week. Considering the median driver, less than 2 hours per year is saved, and 75% of drivers never save more than 3 min in a single day.

Hamzeie, Savolainen, and Gates (2017) investigated how the average speed chosen by drivers was influenced by the traffic conditions and how it affected the risk of crashes. The authors used the SHRP2 NDS data and concluded that higher travel speeds were directly correlated to higher speed limits, and high speed variance is correlated to higher risk of crash. Kong et al. (2020) also investigated speeding behavior patterns related to the traffic and road conditions, from a sample of naturalistic data collected from 92 vehicles. The authors concluded that speeding duration is related to driving on roadways with higher function classes, to longer trips and to speed loss causes by congestion, causing a desire of needing to "recover the lost time" by the drivers.

Using the NDS-BR data, Bastos, Santos, et al. (2021) compared traffic risks behavior between carpooling and non carpooling drivers, including speeding. The authors concluded that carpooling drivers presented a safer behavior towards speeding, having less episodes in comparison to non carpooling drivers. Bastos, Santos, et al. (2020) also investigated the occurrence of mobile phone use (MPU) while driving and

concluded that MPU occurs in higher frequencies when the average speed is lower. To analyze the relationship between speeding and speed cameras, Amancio (2021) also used the BR-NDS dataset.

Overall, the papers cited in this Section are just a small sample of the existing studies including speeding behavior and NDS data. The NDS projects that were concluded or are ongoing generates an ample database of speeds and speeding episodes, making possible to continually investigate this risk behavior.

## 2.5 GEOGRAPHICALLY WEIGHTED REGRESSION

Regression analysis is a tool to investigate the dependency between variables and to predict future parameters based on previous ones. This type of statistical analysis can show the relationship between dependent variables (LINDLEY, 1987). In investigation of the effects of the built environment on the occurrence of crashes in urban areas, negative binomial (NB) regression is a traditional model used (WEI; LOVEGROVE, 2013; ZHANG, C. et al., 2014). But the quality of this model is limited by the incapacity of analyzing the spatial dependency and heterogeneity expected to happen on road safety factors related to the urban environment (OBELHEIRO; SILVA; NODARI, 2019).

To overcome this limitation, the Geographically Weighted Regression (GWR) allows the exploration of the relationship between variables on a spatial nonstationarity context. The spatial nonstationarity is a scenario where it is assumed that parameters are not constant across the space. A global regression model may be incapable of explaining the relationship between sets of variables with a acceptable level of precision in this scenario. The GWR model considers that the nature of the model must alter over space to reflect the structure within the data, allowing the actual parameters for each location in space to be modeled and mapped (BRUNSDON; FOTHERINGHAM; CHARLTION, 1996).

The basic form of GWR is detailed on the following equation:

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \epsilon_i; \quad (2.2)$$

where  $y_i$  is the dependent variable at location  $i$ ,  $x_{ik}$  is the value of the  $k$ th independent variable at location  $i$ ,  $m$  is the number of independent variables,  $\beta_{i0}$  is the intercept parameter at location  $i$ ,  $\beta_{ik}$  is the local regression coefficient for the  $k$ th parameter at location  $i$  and  $\epsilon_i$  is the random error at location  $i$ . The GWR model depends on a spatial weighting function called  $w_{ij}$  that controls the contribution of the point  $j$  on the calibration of a model for point  $i$ . The spatial weighting function represents the idea that for each point  $i$  there is a bump of influence around it. Considering this

"bump", observations closer to  $i$  have more influence in the estimation of  $i$ 's parameters (BRUNSDON; FOTHERINGHAM; CHARLTON, 1996; GOLLINI et al., 2013).

In the context of a multivariate GW model, these influences are calculated by a weighted least squares approach, described on the following equation:

$$\hat{\beta}_i = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y ; \quad (2.3)$$

where  $X$  is the matrix of the independent variables with a column of 1s (ones) for the intercept,  $y$  is the dependent variable vector,  $\hat{\beta} = (\beta_{i0}, \dots, \beta_{im})^T$  is the vector of  $m+1$  local regression coefficients and  $W_i$  is the diagonal matrix denoting the spatial weighting ( $w_{ij}$ ) of each observed data for regression point  $i$  at location  $(u_i, v_i)$  (defined by the selected spatial weighting function) (GOLLINI et al., 2013). The spatial weighting function is also known as the kernel function. This function can have multiple configurations, including Gaussian, Exponential, Boxcar, Bisquare and Tricube configurations, detailed on TABLE 10.

TABLE 10 – KERNEL FUNCTIONS

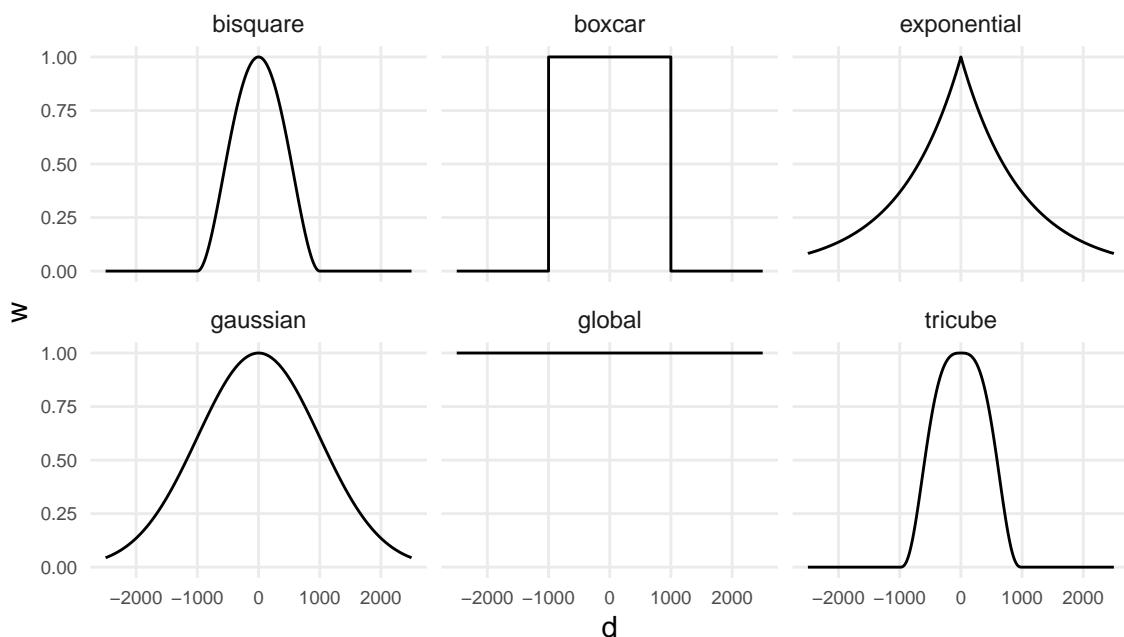
Kernel	Equation
Global Model	$w_{ij} = 1$
Gaussian	$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right)$
Exponential	$w_{ij} = \exp\left(-\frac{ d_{ij} }{b}\right)$
Boxcar	$w_{ij} = \begin{cases} 1 & \text{if }  d_{ij}  < b, \\ 0 & \text{otherwise} \end{cases}$
Bisquare	$w_{ij} = \begin{cases} \left(1 - (d_{ij}/b)^2\right)^2 & \text{if }  d_{ij}  < b, \\ 0 & \text{otherwise} \end{cases}$
Tricube	$w_{ij} = \begin{cases} \left(1 - ( d_{ij} /b)^3\right)^3 & \text{if }  d_{ij}  < b, \\ 0 & \text{otherwise} \end{cases}$

SOURCE: Gollini et al. (2013)

The distance between observations  $i$  and  $j$  is defined by  $d_{ij}$ , inside a chosen  $b$  bandwidth, which is the key controlling parameter in all kernel functions. FIGURE 20 presents the plot of the kernel functions, for a generic bandwidth  $b = 1000$ , where  $w$  is the weight and  $d$  is the distance between two observations. As the plot shows, the global kernel function represents a global linear regression, where all the variables are constant across the space. The bandwidth has two configurations: fixed or adaptive. Fixed bandwidth consists of a absolute value of distance, presenting better outcomes for uniform sized zonal levels (HUANG; WANG; PATTON, 2018), including grids with equal areas. Adaptive bandwidth consists in choosing a fixed quantity of neighbors for each regression point  $i$ , with variable sizes. This configuration works better for zonal

levels with irregular sizes, including traffic analysis zones (TAZ) and census blocks (YU; XU, 2017). The size of bandwidth can be chosen with the use of a cross-validation function or testing multiple values to reach a lower value of corrected Akaike information criterion (AICc). These processes will be better explored in Section 3.3.

FIGURE 20 – KERNEL FUNCTIONS PLOT



SOURCE: The Author (2021), based on Gollini et al. (2013).

The GWR model presented on equation 2.2 has a limitation: it can be used only when the distribution of the dependent variable is Gaussian (SILVA; RODRIGUES, 2013). Therefore, count data needs other GWR configurations in order to fit in the model, including Poisson and Negative Binomial distributions. The Geographically Weighted Poisson Regression (GWPR) is described by Obelheiro, Silva, Nodari, et al. (2020) and Nakaya et al. (2005) with the following equation:

$$y_i \sim \text{Poisson} \left[ t_i \exp \left( \sum_k \beta_{ik} (u_i, v_i) x_{ik} \right) \right]; \quad (2.4)$$

where  $t_i$  is an offset variable and remaining variables are defined in equations 2.2 and 2.3. The Geographically Weighted Negative Binomial Regression (GWNBR) described by Silva and Rodrigues (2013) with the following equation:

$$y_i \sim \text{NB} \left[ t_i \exp \left( \sum_k \beta_{ik} (u_i, v_i) x_{ik} \right), \alpha (u_i, v_i) \right]; \quad (2.5)$$

where NB represents negative binomial and  $\alpha$  is the parameter of overdispersion with spatial variation  $(u_i, v_i)$ .

TABLE 11 contains studies in which GWR models are used to investigate the relationship between BE variables and road safety performance, measured by the occurrence of road crashes. In regard to zonal level, most investigations used TAZ as unit of analysis. Obelheiro, Silva, Nodari, et al. (2020) created a traffic safety zonal level (TSAZ) based on census blocks. All authors but one used an area as zonal level. Arvin, Kamrani, and Khattak (2019) performed GWR models using intersection points as location of regression.

TABLE 11 – GWR AND ROAD SAFETY STUDIES

<b>Author</b>	<b>Zonal level</b>	<b>Distribution</b>	<b>Kernel</b>	<b>Bandwidth</b>
Amoh-Gyimah, Saberi, and Sarvi (2017)	TAZ, grid	Poisson	Bisquare, gaussian	Adaptive
Arvin, Kamrani, and Khattak (2019)	Intersections	Poisson	Bisquare, gaussian	Adaptive
Hadayeghi, Shalaby, and Persaud (2010)	TAZ	Poisson	Bisquare, gaussian	-
Huang, Wang, and Patton (2018)	Census block	Gaussian	-	Fixed
Obelheiro, Silva, and Nodari (2019)	TAZ	Poisson, NB	-	Adaptive
Obelheiro, Silva, Nodari, et al. (2020)	TSAZ	Poisson, NB	-	Adaptive
Pirdavani et al. (2014)	TAZ	Gaussian	Bisquare, gaussian	Adaptive
Rhee et al. (2016)	TAZ	Gaussian	Gaussian	Fixed
Yu and Xu (2017)	Census block	NB	-	Adaptive
Y. Zhang et al. (2015)	TAZ	Poisson	-	-

SOURCE: The Author (2021)

Regarding the type of GW model, most authors chose GWNBR or GWPR, considering that road crash count distribution fits better in these models. In relation to kernel type, it only varied between two options: Bisquare and Gaussian. Most studies contained the use of adaptive kernels, which is a better option for zonal level with irregular sizes. In conclusion, GWR is a method that enables the exploration of findings that might otherwise be missed if only a global regression method is applied. It is always important to really check if the GWR model will describe the data better than a global regression model comparing the results from both. All the process of constructing a GWR model and its parameters for this work is described in Section 3.3.

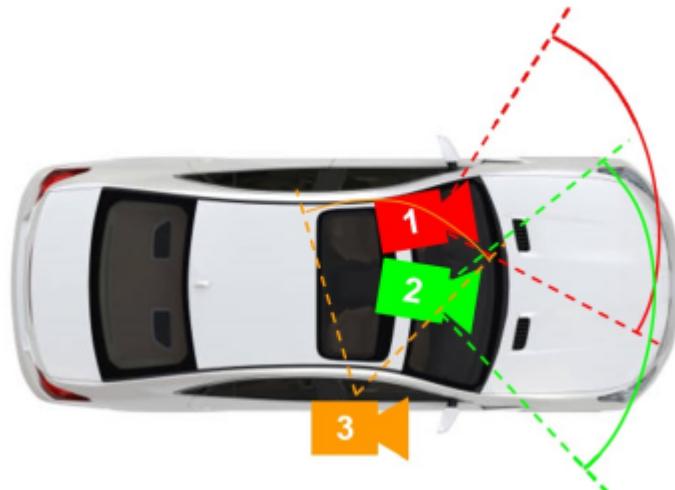
### 3 METHODS

This chapter is divided into 3 sections. Section 3.1 contains the steps of collecting and processing naturalistic driving data. Section 3.2 presents the process of creating the dependent and independent variables. Section 3.3 includes the process of elaborating a GWR model.

#### 3.1 NATURALISTIC DRIVING STUDY - BRAZIL

The Naturalistic Driving Study - Brazil (NDS-BR) utilized a non-intrusive instrumentation of vehicles, with two data acquisition system (DAS). Each DAS contains three high definition cameras, one GPS sensor, one laptop with a GNU/Linux operational system distribution and one power supply. One camera was positioned on the right window of the car, towards the inside of the vehicle facing the driver. Two other cameras were positioned on the windshield of the car, facing towards the front outside. Considering these two, one camera is facing slightly to the left and another one to the right. The visual scheme of camera positioning is presented on FIGURE 21.

FIGURE 21 – CAMERA POSITIONING INSIDE THE VEHICLE



SOURCE: Amancio (2021).

The GPS sensor was positioned near the car panel. The cameras and GPS sensor were connected to the laptop, which controlled and synchronized the data acquisition. The laptop was installed inside a box, positioned on the floor by the passenger's seat. The DAS was installed in the personal vehicle of each volunteer. There was no audio collected from inside the vehicle, in order to not inhibit possible conversation of the driver during the study. Cameras 1 and 2 permitted the collection of visual data about

the environment outside the vehicle - traffic conditions, traffic control, weather, conflicts with other vehicles, etc. Camera 3 permitted the collection of visual data about the behavior of the driver, including the use of seatbelt, mobile phone use, etc. FIGURE 22 shows images collected by the cameras.

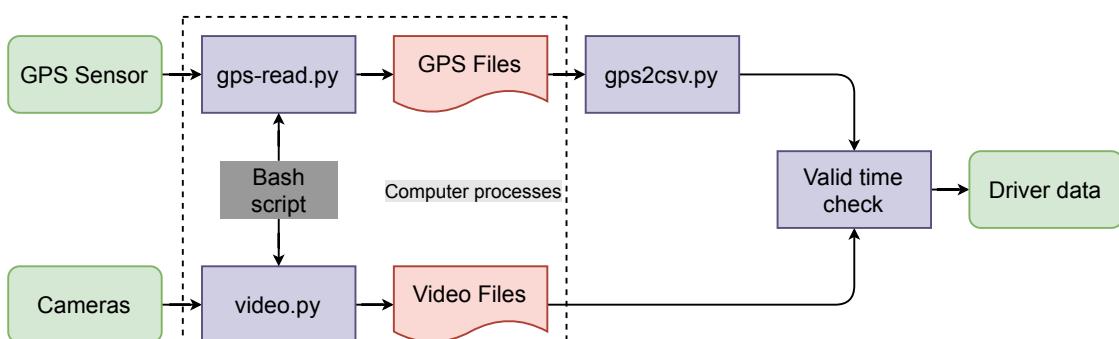
FIGURE 22 – CAMERA IMAGES



SOURCE: The Author (2021).

The DAS kit allowed the acquisition of GPS and video data in synchronization, collecting information in a frequency of one second. The process of acquisition is illustrated in FIGURE 23. The main controller of the data collection was the laptop. In the beginning of each trip, the driver needed to manually turn on the laptop after starting the car engine and before it started driving. A Bash script was written to run each time the laptop was turned on, with the objective to start two main Python scripts used in the data acquisition: `gps-read.py` and `video.py` (BORGUEZANI et al., 2020).

FIGURE 23 – NATURALISTIC DATA COLLECTION AND PROCESSING

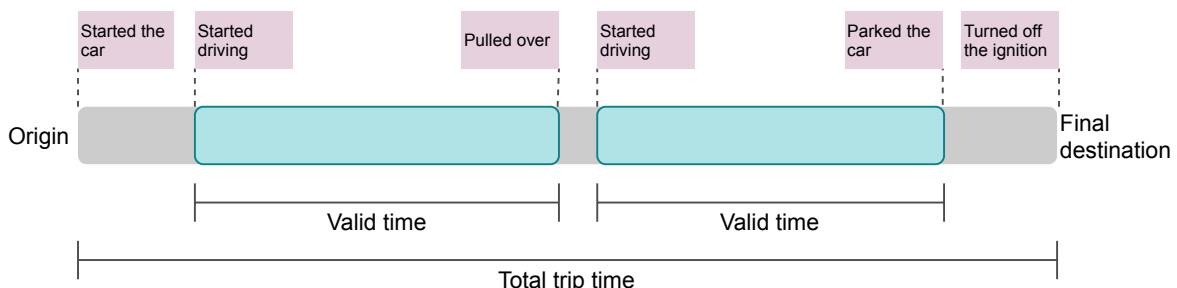


SOURCE: The Author (2021).

The GPS sensor signals were translated into a `.nmea` format file, with the execution of `gps-read.py`. In order to analyze the collected data, after the NDS runs another Python script was used by the NDS-BR researchers to translate the `.nmea` into a `.csv` file. The GPS data contain information including latitude and longitude coordinates, date, time, speed, heading and altitude. The video files from all cameras were processed by the `video.py` script, having the timestamp and date printed on the screen. All the programs written in Python are presented on Annex 1.

After data was extracted from the laptop, it was necessary to make a valid time check, in order to exclude times from the final sample where the driver wasn't really driving the vehicle. FIGURE 24 presents how these invalid times were removed from the sample. For each trip, the period between the start of the car and when the volunteer started driving was considered as invalid. During the trip, if the driver pulled over briefly and then started driving again, this period between the actions was considered as invalid as well. Finally, the gap between parking the car and turning off the DAS was discarded from the sample.

FIGURE 24 – VALID TIME SELECTION



SOURCE: The Author (2021).

The first trip of each driver was defined invalid. This trip was considered to be, to the driver, a time of getting familiar with the monitoring system. Moments with failure on the DAS, where the GPS or video data weren't properly recorded, also were considered as invalid time. The process of identifying invalid times was completely manual. A group of researchers identified the invalid times using the video data as reference.

The NDS-BR started collecting data on August, 2019. The local of the study is Curitiba and its metropolitan region. Curitiba is the capital city of the State of Paraná, in Brazil. Until March of 2021, the project collected data from 16 volunteers, which composed the sample of naturalistic data used in this thesis. The drivers volunteered through a survey that was divulged in social networks. TABLE 12 contains information about the participants and its vehicles, including the duration of data collection, age, gender, age of driver license, vehicle model year and horsepower.

The age of the drivers varied between 21 and 62 years old, with 7 females and 9 males. The duration of data collection from each driver varied between 5 and 17 days. Regarding the volunteers' vehicles, the model year varied between 2001 and 2019, with horsepower varying between 74 and 163 HP. All volunteers received a briefing on how to operate the equipment and signed a term of agreement, allowing the use of their data to academic research purposes. Overall, a total of 491 trips were made, resulting in 238.85 hours of driving and 5,362.75 km of distance travelled.

Only driver  $D_{14}$  had a car with automatic gearbox. All other drivers had vehicles

TABLE 12 – SELECTED PARTICIPANTS INFORMATION

ID	Duration	Age	Gender	License Age	Vehicle year	Horsepower
$D_1$	13	33	F	12	2012	96
$D_2$	12	40	M	3	2010	113
$D_3$	8	22	M	3	2011	103
$D_4$	16	25	M	6	2003	114
$D_5$	13	41	F	23	2014	75
$D_6$	13	28	M	9	2013	163
$D_7$	17	45	M	18	2011	130
$D_8$	12	32	F	10	2015	79
$D_9$	11	29	M	9	2019	74
$D_{10}$	14	62	M	37	2018	69
$D_{11}$	6	28	F	10	2012	75
$D_{12}$	5	45	F	26	2001	101
$D_{13}$	14	21	F	2	2006	81
$D_{14}$	6	34	M	16	2014	124
$D_{15}$	8	49	F	24	2018	84
$D_{16}$	11	27	M	3	2014	109

SOURCE: The Author (2021)

NOTES: Duration in days; age and license age in years; M-male; F-female

with a manual gearbox. Drivers  $D_1$ ,  $D_2$ ,  $D_3$  e  $D_4$  casually made trips in carpooling situation, offering rides to strangers with the use of carpooling apps. Drivers  $D_{14}$ ,  $D_{15}$  and  $D_{16}$  worked as mobility app drivers. This caused a greater participation from them in the overall sample, which will be observed in the results.

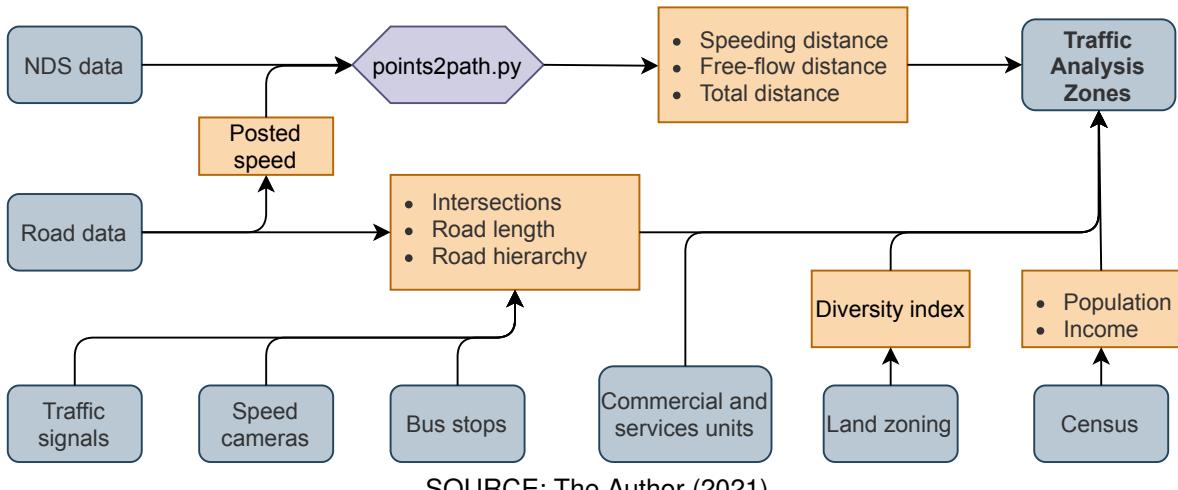
### 3.2 DATA PROCESSING AND VARIABLE CONSTRUCTION

This section includes the methods utilized to construct all the utilized variables, divided into two subsections: the speeding variable (dependent variable) and built environment variables (independent variables). FIGURE 25 presents a brief flowchart on the process of transforming the data sources into a complete dataset included in the traffic analysis zones.

#### 3.2.1 Speeding

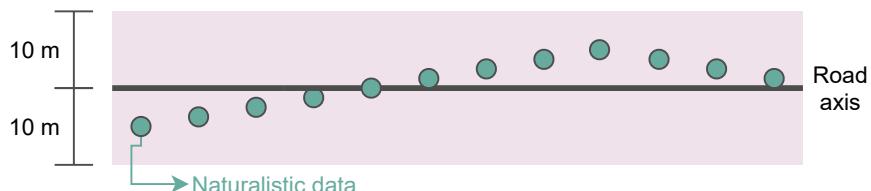
In order to extract speeding information from the NDS sample, it was necessary to compare the performed speeds from drivers with the respective posted speed. Posted speed data was collected from the road axis shape file obtained from OpenStreetMap (OPENSTREETMAP CONTRIBUTORS, 2017), completing some missing information with posted speeds from IPPUC road axis shape files (IPPUC, 2021). Using the QGIS software, it was created a 10 meter buffer around the road axes. In sequence, a intersection between the NDS points and the road buffer was performed, in order to associate the posted speed and the performed speed to the same point in space. This process is illustrated in FIGURE 26.

FIGURE 25 – VARIABLE DATA FLOWCHART



SOURCE: The Author (2021).

FIGURE 26 – ROAD AXIS BUFFER



SOURCE: The Author (2021).

Points with coordinates outside the buffers (in the middle of street blocks, for example) were removed from the sample. Also, points without posted speed and without performed speed data were removed from the sample. The extraction of free-flow and speeding events followed a method defined by Richard, Campbell, Brown, et al. (2013), presented in FIGURE 15. The occurrence of speeding is represented by the proportion of distance travelled in speeds above the posted speed, divided by the distance performed in free-flow speeds. It is necessary to remove sections of the sample where drivers had no opportunity to speed. Therefore, the occurrence of speeding is defined the following equation:

$$SP = \frac{D_{Vsp}}{D_{Vf}} ; \quad (3.1)$$

where  $D_{Vsp}$  is the distance performed in speeds exceeding 5 km/h above the posted speed (analytical approach) and  $D_{Vf}$  is the distance performed in free-flow speeds. Free-flow speeds includes performed speeds higher than the posted speed minus 10 km/h, and lower than speeding. FIGURE 27 presents a section of a trip, illustrating the thresholds previously described.

To extract distance data from the gps coordinates, it was made a conversion from the coordinates into a linestring, a geometry type from the Well-known text (WKT)

FIGURE 27 – SPEEDING OCCURENCE AND SPEEDING OPPORTUNITY



SOURCE: The Author (2021).

format. QGIS is able to read the WKT format and calculate the distances. Only points with a 1 second gap from each other (observed from the time variable) were connected with a linestring. The conversion process was made with a python script, included on Appendix 1. Total distances, distances performed in free-flow speeds and distances in speeding were calculated for each TAZ in Curitiba, in order to calculate the occurrence of speeding.

### 3.2.2 Built Environment

All variables (speeding and built environment) were calculated for each traffic analysis zone (TAZ) inside Curitiba. 135 TAZs, established in Curitiba's origin/destiny research made by IPPUC in 2018, were used for this work (IPPUC, 2018). The BE variables that were calculated are listed in TABLE 13. Inside the density category, population data was obtained from the last Brazilian Census in 2010 (IBGE, 2021). Using the available census blocks, it was possible to calculate the population density (PD) data into the TAZs.

Inside the diversity category, the land use diversity index (LDI) used information from Curitiba's zoning plan (CURITIBA, 2019) as input for its calculation. For each TAZ, the diversity index was calculated based on the following equation (HUANG; WANG;

TABLE 13 – BUILT ENVIRONMENT VARIABLES

Category	Variable	Description [unit]
Density	PD	Population Density [inhab./km <sup>2</sup> ]
Diversity	LDI	Land use diversity index
Design	DIS	Density of intersections [no./km]
	DSC	Density of speed cameras [no./km]
	TSD	Traffic signal density [no./km]
	PAR	Proportion of arterial roads
	SND	Street network density [km/km <sup>2</sup> ]
Destination accessibility	DCSU	Density of commercial and services units [no./km <sup>2</sup> ]
Distance to transit	BSD	Bus stop density [no./km]
Demographics	AVI	Average income [BRL]

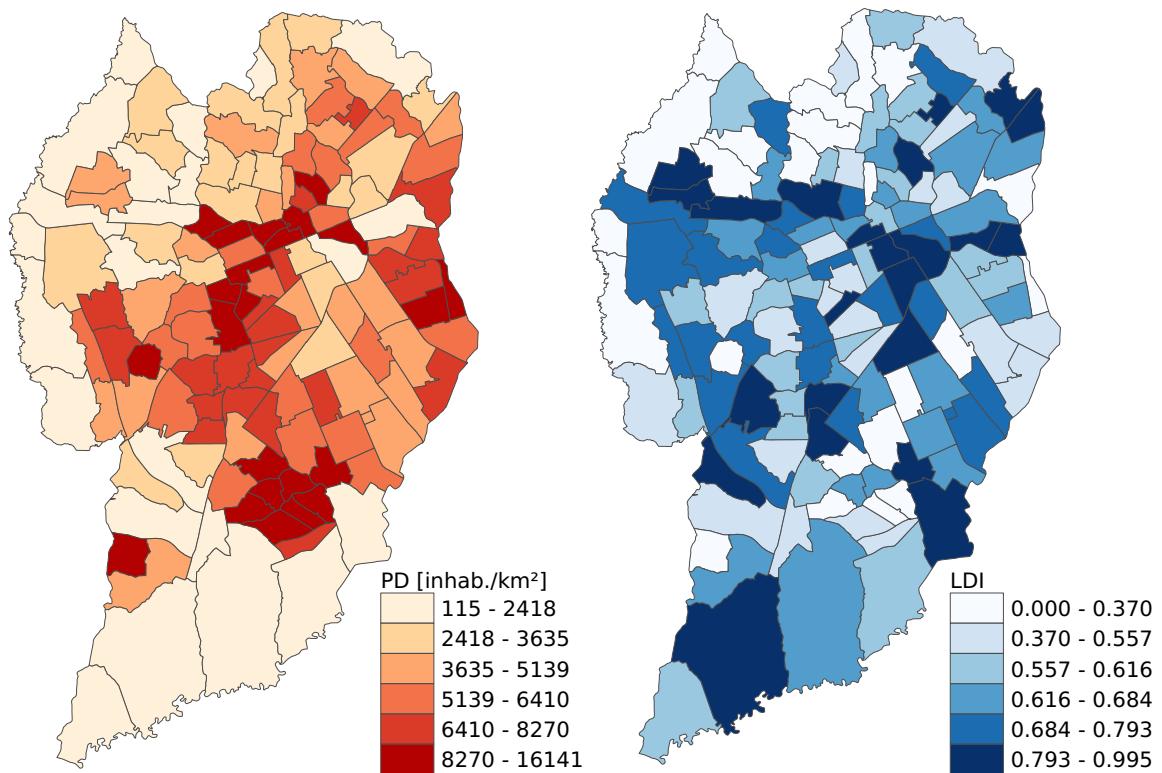
SOURCE: The Author (2021)

PATTON, 2018):

$$LDI = \frac{-\sum_i^n P_i \times \ln(P_i)}{\ln(n)} ; \quad (3.2)$$

where  $i$  is the type of land use,  $n$  is the total number of land use types in a TAZ and  $P_i$  is the proportional area of land use type  $i$ . LDI varies between 0 and 1, where 0 represents a single land use zone and 1 represents a perfect distribution of land uses. The GIS data of each zoning type was obtained from IPPUC (2021). PD and LDI variables are plotted in FIGURE 28.

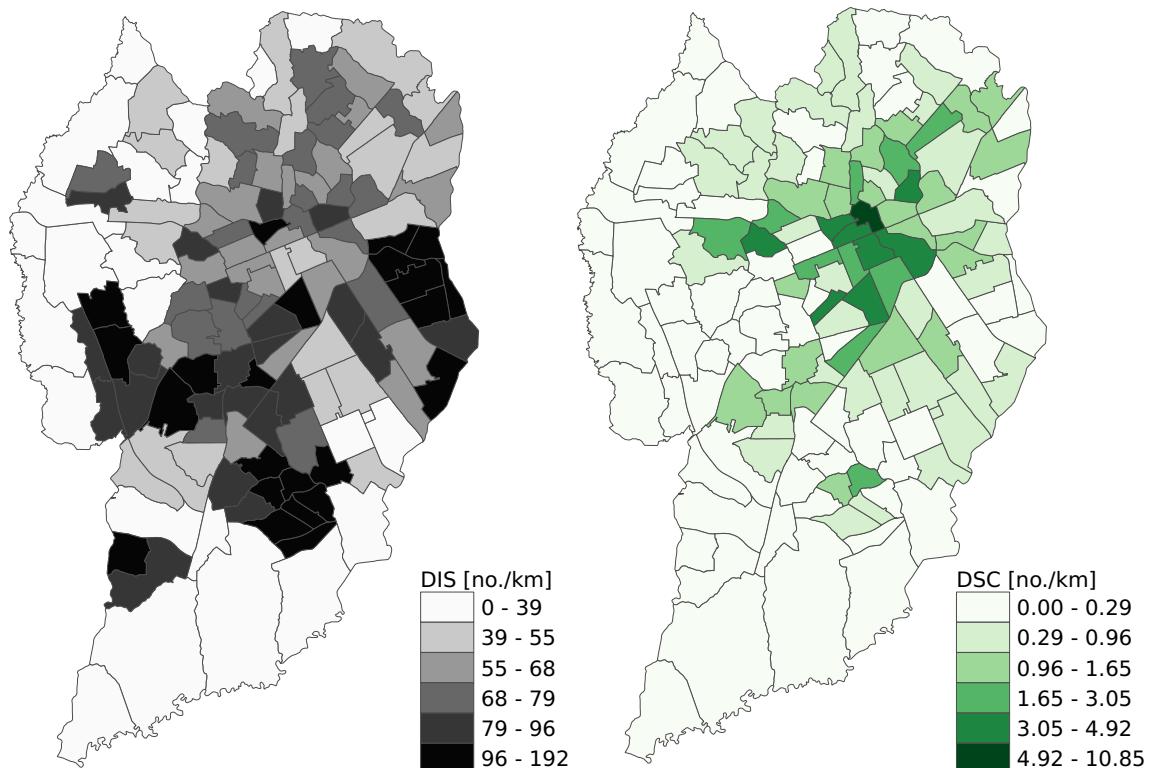
FIGURE 28 – PD AND LDI



SOURCE: The Author (2021), based on IBGE (2021) and IPPUC (2018, 2021).

In regard to the design category, the density of intersections (DIS) was calculated based on the number of intersections divided by the length of roads. The intersection points and road lengths were extracted from the IPPUC (2021) road axis data. Intersection points were extracted using the "line intersection" function included in QGIS. Density of speed cameras (DSC) and traffic signal density (TSD) follows the same method applied on density of intersections. Speed camera points were obtained from the Municipal Office of Social Defense and Transit of Curitiba (SETRAN, 2020) and traffic signal points were obtained from IPPUC (2021). DIS and DSC variables are plotted in FIGURE 29.

FIGURE 29 – DIS AND DSC

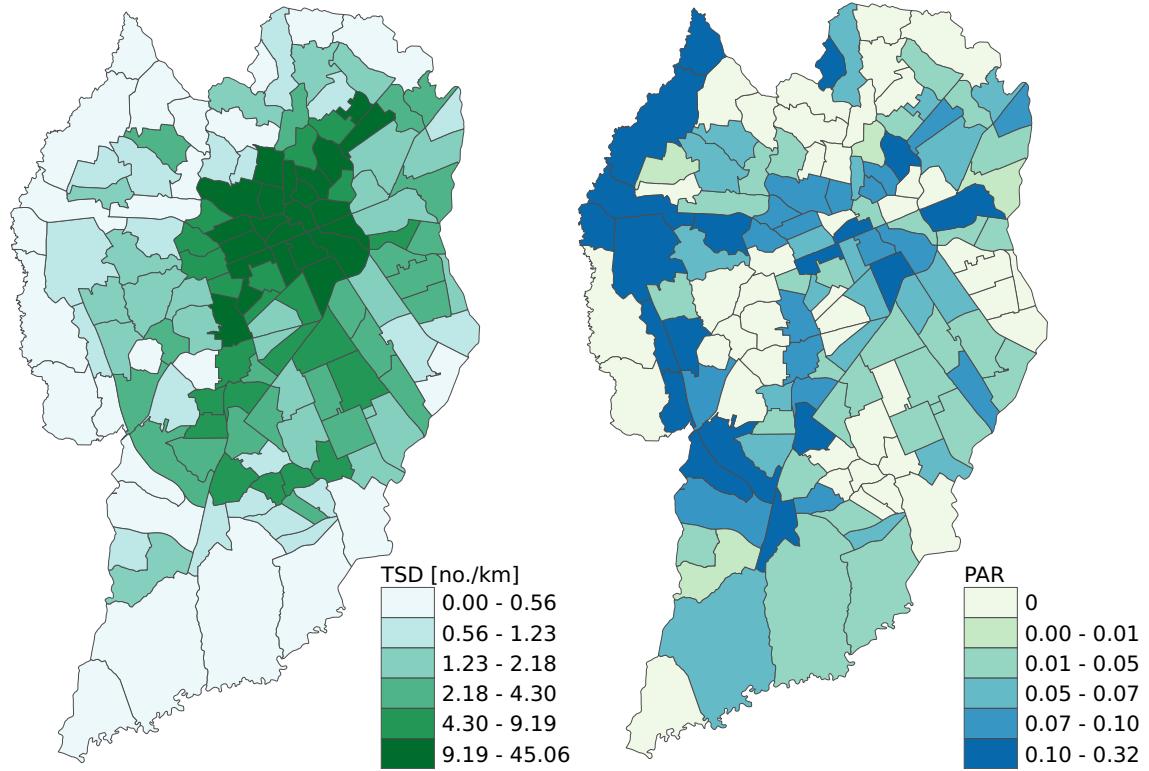


SOURCE: The Author (2021), based on IPPUC (2018, 2021) and SETRAN (2020)

The proportion of arterials (PAR) used the IPPUC road axis data as well. The road hierarchy data follows the categories established by the Brazilian Traffic Code (BRASIL, 1997): rapid transit, arterial, collector and local roads. In this variable, rapid transit and arterial roads were combined into one category representing arterial roads. Hence, the proportion of arterial roads was based on the division between arterial roads length and total road length. TSD and PAR variables are plotted in FIGURE 30

Street network density (SND) is the total road length divided by the area in squared kilometers of the respective TAZ. Regarding the destination accessibility, the location of commercial and services units from 2019 was provided by IPPUC (2021),

FIGURE 30 – TSD AND PAR



SOURCE: The Author (2021), based on IPPUC (2018, 2021)

making possible the calculation of the density of commercial and services units (DCSU) per TAZ. SND and DCSU variables are plotted in FIGURE 31.

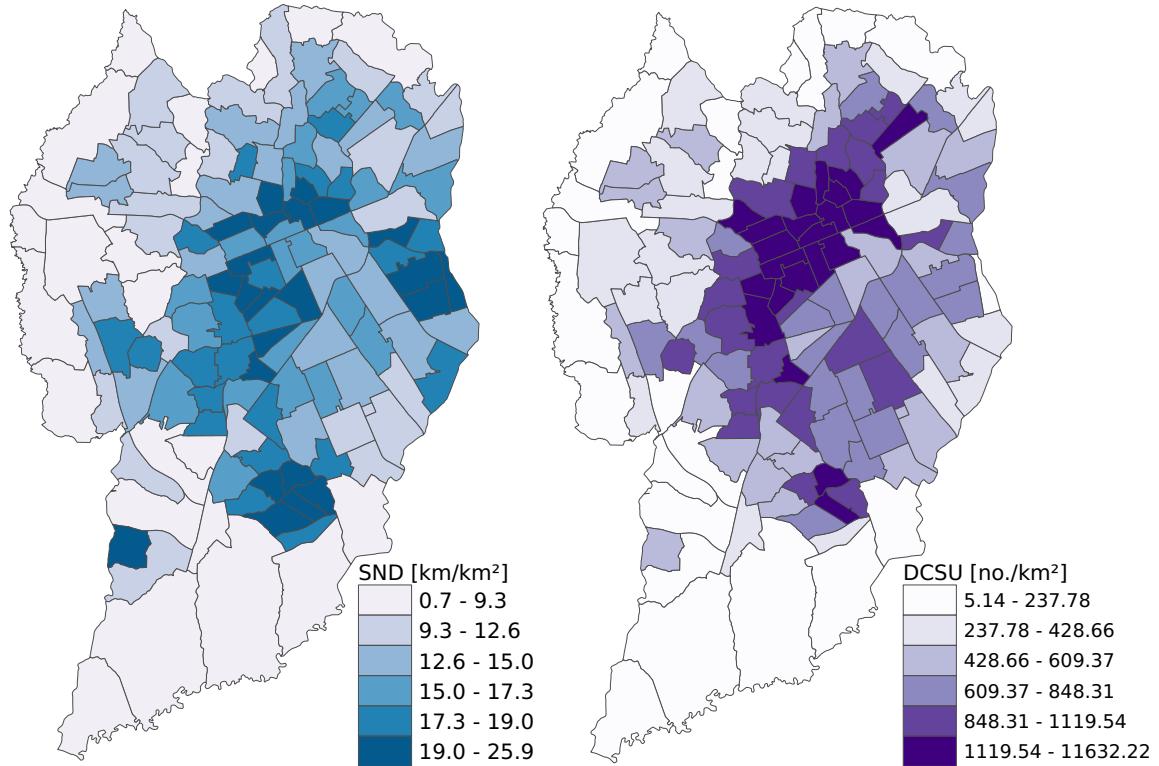
The distance to transit is represented by the bus stop density (BSD) per road length. The location of bus stops across the city was provided by IPPUC and URBS (2020), with data from 2020. The average income (AVI), representing the demographics category, was obtained from the Brazilian Census of 2010 (IBGE, 2021). BSD and AVI variables are plotted in FIGURE 32.

### 3.3 GEOGRAPHICALLY WEIGHTED REGRESSION MODELLING

This section contains the process of modelling a GWR. The steps performed in this section are presented in FIGURE 33. They were implemented using the R programming language as the main tool, with the aid of functions extracted from the GWmodel (GOLLINI et al., 2013) and spdep (BIVAND; PEBESMA; GOMEZ-RUBIO, 2013) packages. The TAZs data passed through five main steps in order to construct and make a diagnostic of the GWR model: data arrangement, bandwidth size selection, kernel type selection, calculation of a global linear model and Moran's I application on the model's residuals.

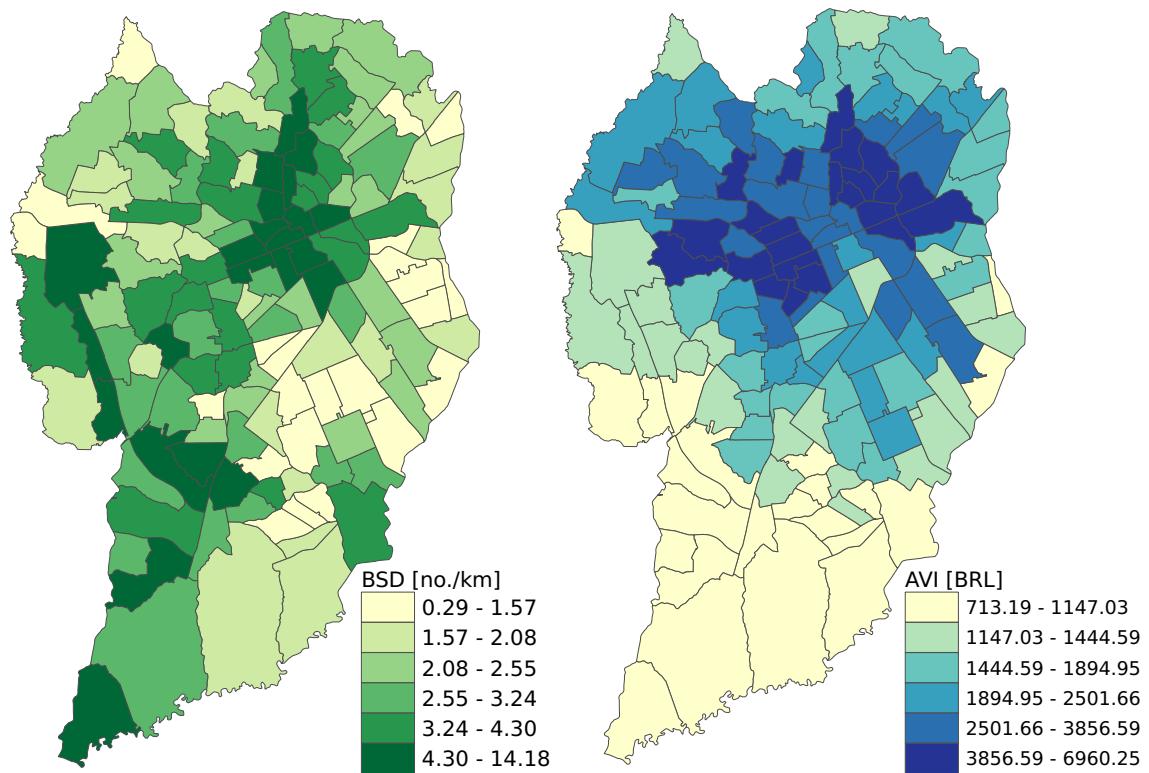
The first two steps of the data arrangement process are related to how the

FIGURE 31 – SND AND DCSU



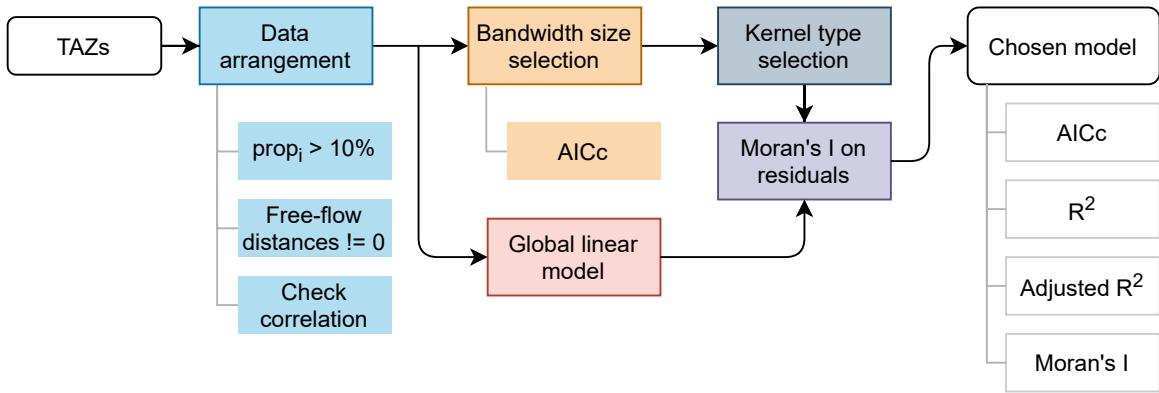
SOURCE: The Author (2021), based on IPPUC (2018, 2021)

FIGURE 32 – BSD AND AVI



SOURCE: The Author (2021), based on IPPUC (2018), IPPUC and URBS (2020) and IBGE (2021)

FIGURE 33 – GWR MODELLING PROCESS



SOURCE: The Author (2021).

naturalistic data of travelled distances are distributed across the TAZs. In order to check these distances, a new variable was created to represent the proportion of travelled distances per TAZ total road length ( $prop_i$ ). This variable is defined in the following equation:

$$prop_i = \frac{D_i}{R_i} ; \quad (3.3)$$

where  $D_i$  is the distance travelled in the  $i$ -th TAZ and  $R_i$  is the total road length inside the  $i$ -th TAZ. Hence, all TAZs with a value above 10% were kept in the sample. This process was made in order to keep TAZs with a significant amount of travel data.

The second step of data arrangement was the exclusion of TAZs without free-flow episodes. Distances performed in free-flow speeds is the exposure variable of the speeding variable. TAZs with zero free-flow distances could not be used in the process of calculating speeding, therefore they were removed from the sample. Finally, a Spearman correlation was performed between all the independent variables. This correlation is based on the following equation (DODGE, 2010):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} ; \quad (3.4)$$

where  $d_i$  is the difference between two ranks of  $i$ -th observations from different variables and  $n$  is the total number of observations. The inclusion of independent variables with correlations above 0.8 can influence negatively in the quality of the GWR model (GOLLINI et al., 2013). Therefore, variables with a Spearman correlation above 0.8 were removed from the sample.

Considering the variability of TAZs sizes in the sample, an adaptive bandwidth has better performance than a fixed bandwidth (HUANG; WANG; PATTON, 2018), therefore an adaptive bandwidth was used in this work. In order to choose the best kernel type, it was necessary to calculate the adaptive bandwidth size (number of

neighbors) for each kernel type (gaussian, bisquare, tricube, boxcar and exponential). The GWmodel package in R has a function (`bw.gwr()`) for automatic bandwidth selection. This function runs multiple GWR models with multiple bandwidth values and extracts the bandwidth size for the GWR model with the lowest value of corrected Akaike Information Criterion (AICc) (GOLLINI et al., 2013).

Minimizing AICc in order to select the bandwidth provides a trade-off between goodness of fit and degrees of freedom. (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002). For GWR, AICc is defined as:

$$AIC_c = 2n \log_e (\hat{\sigma}) + n \log_e (2\pi) + n \left\{ \frac{n + \text{Tr}(S)}{n - 2 - \text{Tr}(S)} \right\} ; \quad (3.5)$$

where  $n$  is the sample size,  $\hat{\sigma}$  is the estimated standard deviation of error term and  $\text{Tr}(S)$  is the trace of the hat matrix, which depends on the bandwidth (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002). With the discovered optimal bandwidths, the basic form of GWR, presented previously in EQUATION 2.2, was performed five times, one for each type of kernel (TABLE 10). The application of the GWR method utilized the `gwr.basic()` function of the GWmodel package. In addition, a global linear regression was performed in order to compare the results with GWR. This is necessary because its important to verify if the GWR model has any advantage over a ordinary least squares (OLS) approach (BRUNSDON; FOTHERINGHAM; CHARLTON, 1996).

Moran's  $I$  was applied on the models residuals to examine the presence of spatial autocorrelation. A regression model with a good spatial analysis performance shows a lack of spatial autocorrelation on it's residuals. The results of Moran's  $I$  test varies between -1 and 1, where -1 represents a perfect dispersion pattern and 1 represents a perfect clustering pattern of the data. Results close to 0 and/or without the desired statistical significance ( $p$ -value) represents a lack of spatial autocorrelation. Moran's  $I$  consists on the following equation (GETIS; ORD, 2010):

$$I = \frac{n}{W} \times \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} ; \quad (3.6)$$

where  $n$  is the sample size,  $w_{ij}$  is the matrix of spatial weights,  $W$  is the sum of all  $w_{ij}$ ,  $x$  is the variable of interest, indexed by  $i$  and  $j$ . and  $\bar{x}$  is the mean of  $x$ . This test was applied using the `moran.test()` function included in the spdep package (BIVAND; PEBESMA; GOMEZ-RUBIO, 2013).

The final step is to analyse the performance between models. This analysis was based on four performance indicators: AICc,  $R^2$ , adjusted  $R^2$  and the value of Moran's  $I$  on residuals. Higher values of  $R^2$  and adjusted  $R^2$  indicates a better performance. The code with all steps described in this section is included in Appendix 2.

## 4 ANALYSIS AND RESULTS

This chapter focus on the results from the NDS-BR (distances, speeding) and the results from the GWR (correlation between speeding and built environment).

### 4.1 NATURALISTIC DATA

The process of removing invalid times from the total sample and choosing the travelled data that happened in Curitiba resulted in 3,459.92 kilometers of travelled distance and a total of 410 trips, representing a travelled time of 134.79 hours. Incomplete trips - where a portion of the trip happened inside Curitiba - were kept in the sample, without sections the happened outside the city territory. TABLE 14 contains the summary of all trips per driver. Overall, all trips in the sample had a mean distance of 8.43 and a median of 4.76 kilometers travelled. Driver  $D_5$  had 56 trips, the highest amount, and driver  $D_{12}$  had the least trips performed, with a value of 6 trips.

TABLE 14 – DESCRIPTIVE STATISTICS OF TRIP DISTANCES

<b>Driver</b>	<b>n</b>	<b>Mean</b>	<b>SD</b>	<b>Min.</b>	<b>1Q</b>	<b>Median</b>	<b>3Q</b>	<b>Max.</b>	<b>Total distance</b>
$D_1$	29	5.49	2.36	1.17	4.70	5.84	6.13	12.60	159.33
$D_2$	14	14.09	7.81	2.41	7.14	14.40	21.94	22.70	197.31
$D_3$	15	11.58	4.41	4.16	7.89	11.86	14.82	19.14	173.72
$D_4$	48	4.26	2.47	0.24	2.30	3.50	5.88	9.36	204.67
$D_5$	56	2.71	2.86	0.01	0.66	2.14	3.72	15.69	152.01
$D_6$	37	3.21	1.44	0.90	1.99	2.82	4.45	6.85	118.83
$D_7$	29	10.84	5.60	0.48	5.12	12.88	15.67	17.04	314.44
$D_8$	31	6.40	3.55	0.67	2.46	7.00	9.15	13.14	198.55
$D_9$	11	4.63	2.04	1.69	3.24	4.51	5.87	7.63	50.92
$D_{10}$	40	5.65	4.02	0.41	2.04	5.05	8.80	14.87	225.97
$D_{11}$	12	1.72	1.63	0.17	0.66	1.07	2.30	5.90	20.62
$D_{12}$	6	1.71	2.38	0.07	0.21	0.46	2.51	5.97	10.28
$D_{13}$	23	3.97	2.01	0.10	2.96	3.94	4.94	7.51	91.26
$D_{14}$	31	21.06	17.28	0.06	7.70	17.68	30.05	72.93	652.71
$D_{15}$	21	38.33	40.40	1.33	10.73	25.57	47.75	168.26	805.02
$D_{16}$	7	12.04	15.21	0.02	3.69	3.92	14.99	42.97	84.27
<b>Total</b>	<b>410</b>	<b>8.43</b>	<b>13.76</b>	<b>0.01</b>	<b>2.27</b>	<b>4.76</b>	<b>9.22</b>	<b>168.26</b>	<b>3,459.92</b>

SOURCE: The Author (2021)

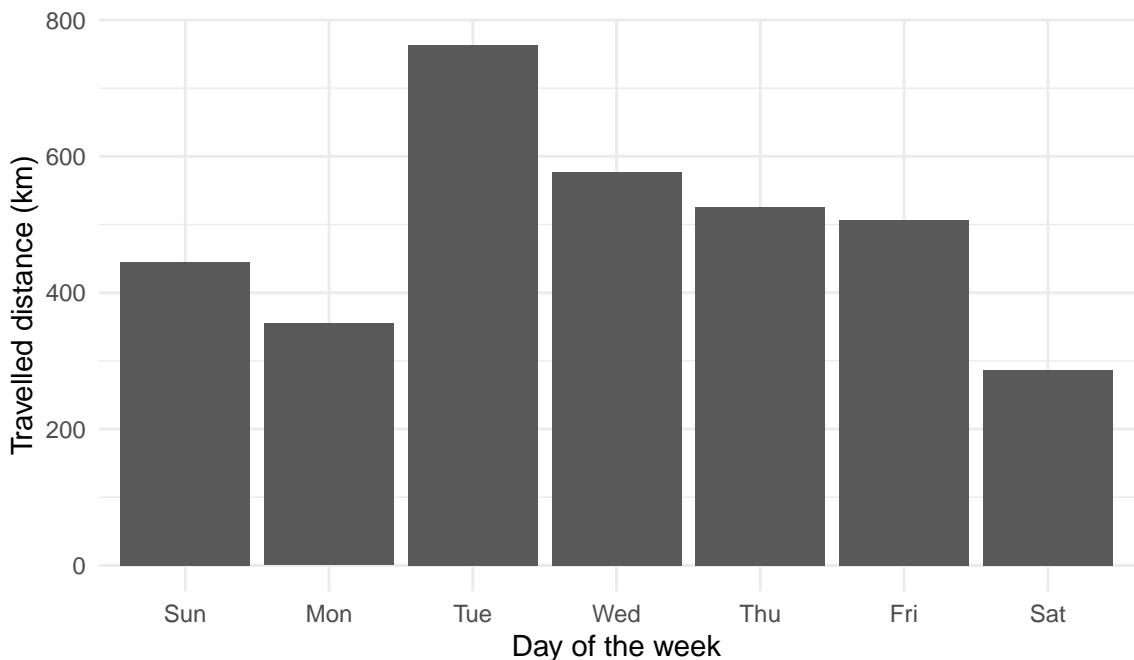
NOTES: Distances in kilometers.

In regard to the mean distance of trips, driver  $D_{15}$  had the highest mean value of 38.33 kilometers and driver  $D_{12}$  had the lowest mean of 1.71 kilometers. The maximum distance of a trip was 168.26 kilometers, performed by driver  $D_{15}$  and minimum distance was 0.01 kilometers, by the driver  $D_5$ . In total, drivers  $D_7$ ,  $D_{14}$  and  $D_{15}$  had the highest travelled distance in the sample, with distances of 314.44, 652.71 and 805.02 kilometers,

respectively. Drivers  $D_{14}$  and  $D_{15}$  were mobility app drivers during the process of the NDS, leading to this higher amount of total distance travelled and bigger trips.

FIGURE 34 shows the distribution of distance travelled per day of the week. Tuesdays had most of the travelled distance and Saturdays were the days with least travelled distance. Regarding the trips, FIGURE 35 also shows its quantity per day of the week. Tuesdays had the highest amount of trips and Sundays the lowest. Overall, most travelled distance and trips happened in higher amounts during weekdays. FIGURE 36 and FIGURE 37 shows travelled distance and trip start distribution per hour of the day, respectively.

FIGURE 34 – DISTANCE TRAVELED PER DAY OF THE WEEK



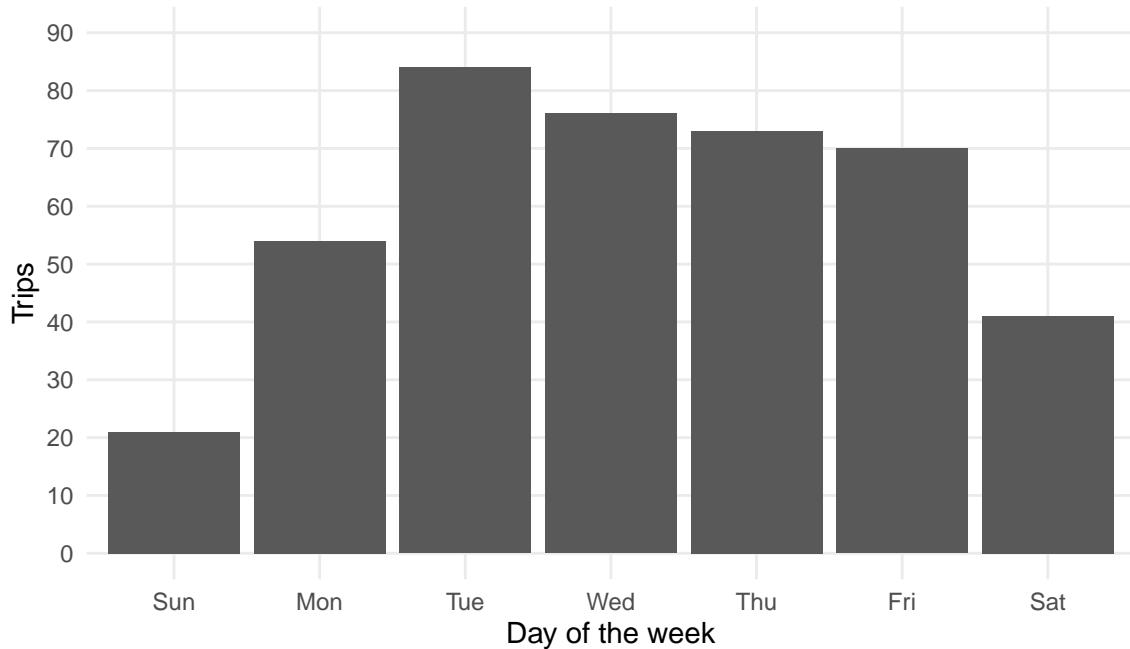
SOURCE: The Author (2021)

Most of the travelled distance happened between 7:00 - 8:00, 13:00 - 14:00 and 19:00 - 20:00. Between 17:00 - 18:00 there were most of the trip starts, followed by 19:00 - 20:00 and 7:00 - 8:00. FIGURE 38 contains the travelled distance per TAZ in Curitiba. The TAZ with highest travelled distance had 165.56 kilometers, and the lowest had 0.04 kilometers. 7 TAZs had zero travelled distance. Most of the trips happened in the central area of the city.

## 4.2 SPEEDING DATA

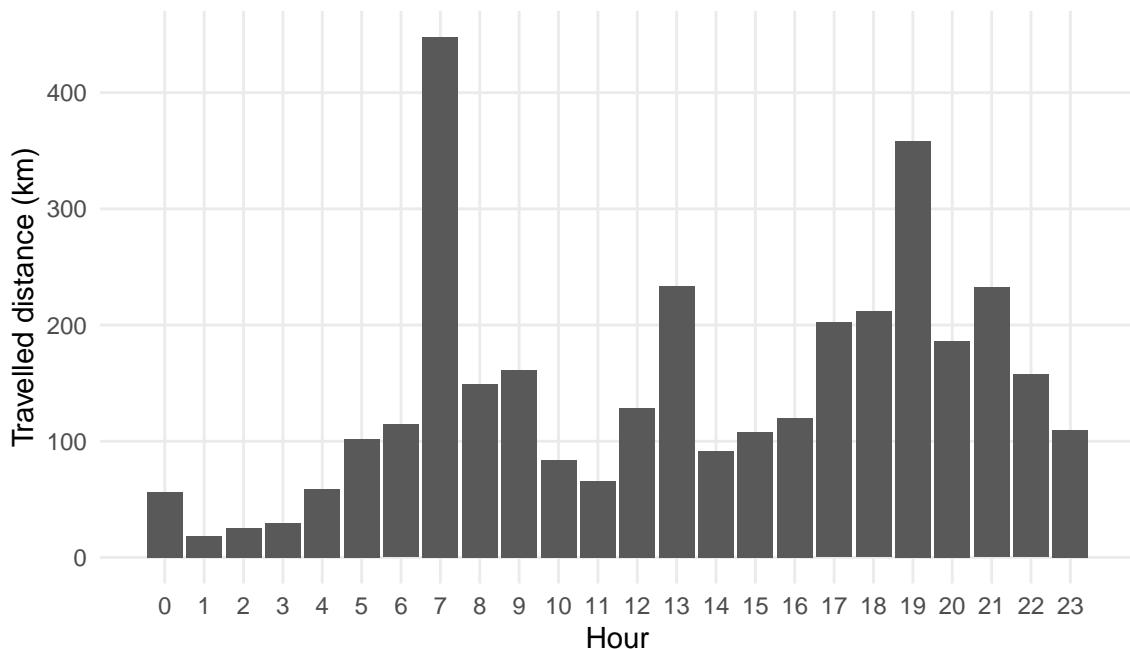
TABLE 15 shows the summary of the speeding data per trip for each driver. Driver  $D_{16}$  had the highest mean speeding value of 0.56, while  $D_5$  had the lowest mean value of 0.25. Drivers  $D_4$ ,  $D_5$ ,  $D_8$ ,  $D_{10}$ ,  $D_{11}$ ,  $D_{12}$ ,  $D_{13}$  and  $D_{14}$  had trips without any

FIGURE 35 – TRIPS PER DAY OF THE WEEK



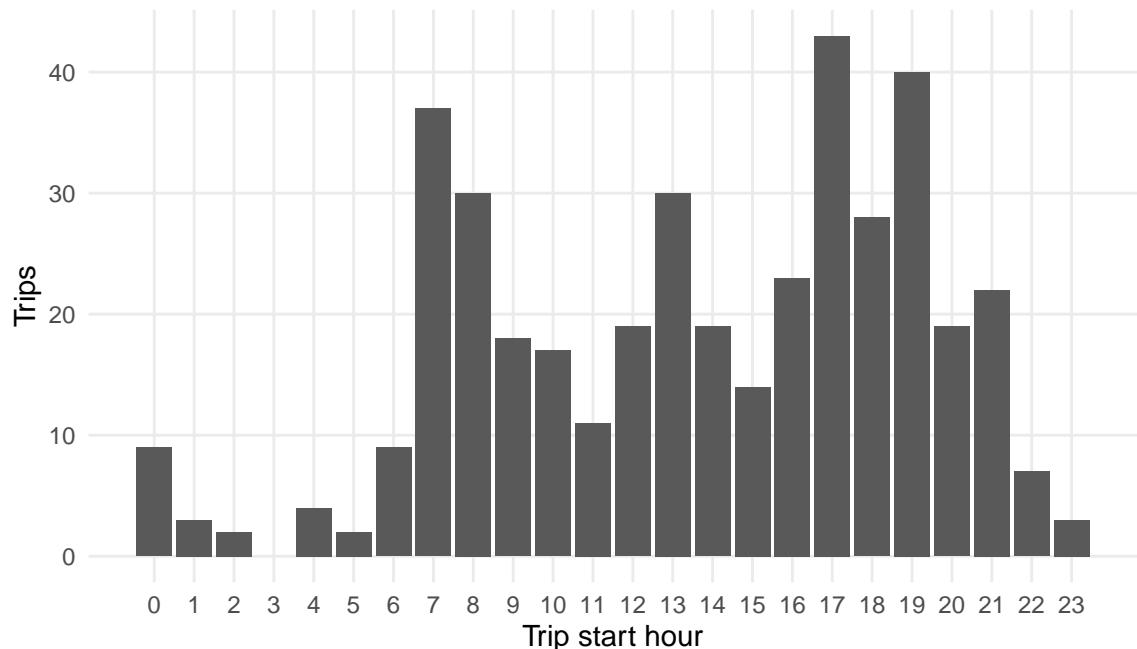
SOURCE: The Author (2021)

FIGURE 36 – DISTANCE TRAVELED PER HOUR OF THE DAY



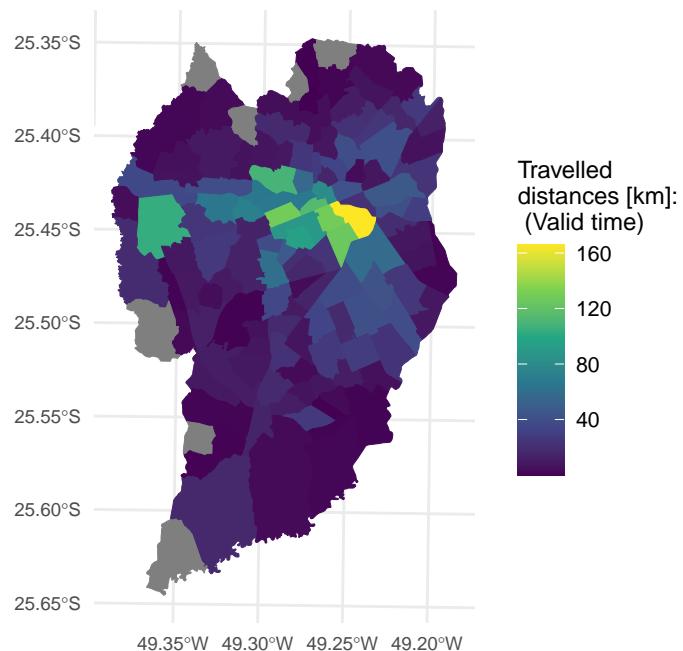
SOURCE: The Author (2021)

FIGURE 37 – TRIP START PER HOUR OF THE DAY



SOURCE: The Author (2021)

FIGURE 38 – TRAVELED DISTANCE PER TAZ



SOURCE: The Author (2021)

speeding. Drivers  $D_5$ ,  $D_{10}$ ,  $D_{13}$ ,  $D_{14}$  and  $D_{16}$  had the highest speeding value of all trips, reaching 0.79. Considering the sum of all trips, the highest value of speeding was performed by driver  $D_{12}$ . The average speeding value of the overall trip sample was 0.35, and 0.41 considering all the travelled distance from the sample.

TABLE 15 – DESCRIPTIVE STATISTICS OF SPEEDING PER TRIP

<b>Driver</b>	<b>Mean</b>	<b>SD</b>	<b>Min.</b>	<b>1Q</b>	<b>Median</b>	<b>3Q</b>	<b>Max.</b>	<b>Total SP</b>
$D_1$	0.29	0.17	0.07	0.16	0.27	0.39	0.63	0.35
$D_2$	0.35	0.16	0.09	0.20	0.43	0.47	0.53	0.42
$D_3$	0.48	0.17	0.23	0.32	0.48	0.60	0.78	0.46
$D_4$	0.32	0.14	0.00	0.26	0.31	0.40	0.66	0.33
$D_5$	0.25	0.24	0.00	0.03	0.18	0.40	0.79	0.31
$D_6$	0.48	0.15	0.10	0.39	0.47	0.59	0.74	0.52
$D_7$	0.26	0.14	0.03	0.18	0.25	0.33	0.76	0.26
$D_8$	0.33	0.21	0.00	0.18	0.26	0.52	0.70	0.38
$D_9$	0.46	0.13	0.23	0.37	0.49	0.58	0.62	0.48
$D_{10}$	0.35	0.22	0.00	0.17	0.35	0.46	0.79	0.43
$D_{11}$	0.26	0.19	0.00	0.13	0.28	0.36	0.64	0.30
$D_{12}$	0.30	0.30	0.00	0.04	0.27	0.47	0.76	0.66
$D_{13}$	0.26	0.18	0.00	0.15	0.22	0.34	0.79	0.26
$D_{14}$	0.46	0.17	0.00	0.40	0.49	0.58	0.79	0.50
$D_{15}$	0.38	0.12	0.07	0.33	0.39	0.47	0.57	0.39
$D_{16}$	0.56	0.19	0.30	0.45	0.57	0.70	0.79	0.55
<b>Total</b>	<b>0.35</b>	<b>0.20</b>	<b>0.00</b>	<b>0.19</b>	<b>0.34</b>	<b>0.49</b>	<b>0.79</b>	<b>0.41</b>

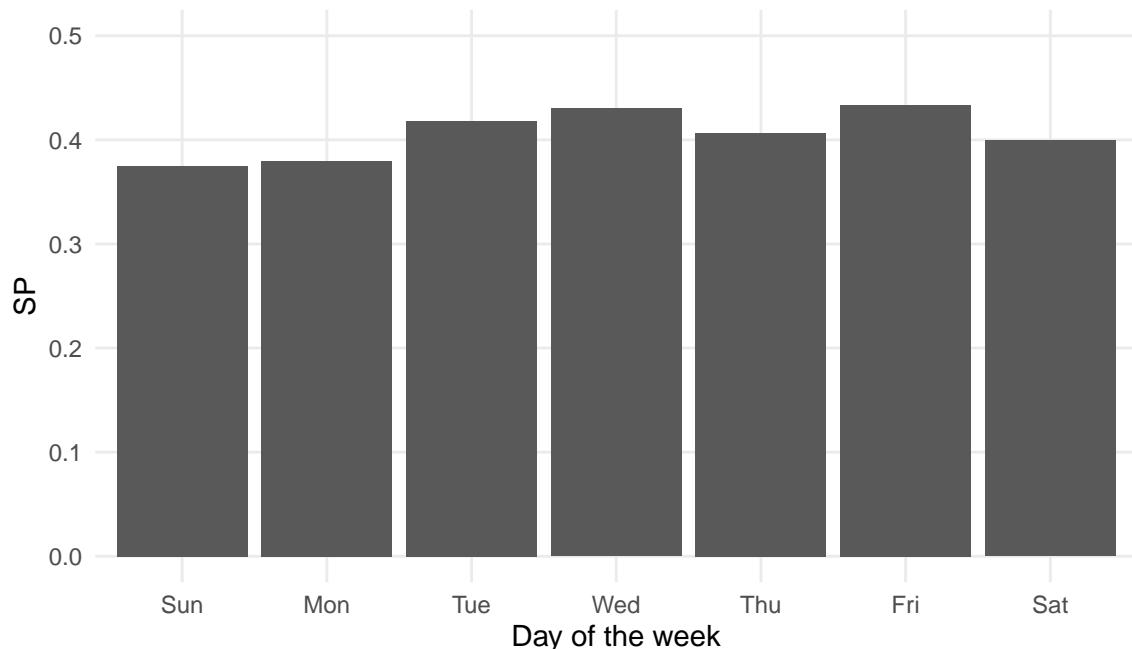
SOURCE: The Author (2021).

Across all days of the week, speeding followed a similar pattern (FIGURE 39). Speeding in Wednesdays and Fridays were a bit higher than other days, and Sundays and Mondays presented lower values of speeding. Regarding the hour of the day (FIGURE 40), the highest value of speeding happened between 0:00 - 1:00 and the lowest between 1:00 and 2:00.

Richard, Campbell, Lichy, et al. (2013) establishes four types of speeding behavior - situational speeding, incidental speeding, habitual speeding and casual speeding - based on two variables: percentage of trips with any speeding and average speeding per trip, of each driver. Plotting these variables and setting a zone boundary of 20% in each variable (arbitrarily defined by Richard, Campbell, Lichy, et al. (2013)) creates 4 zones, one for each type of speeding behavior. FIGURE 41 is the result of this method application on the speeding sample.

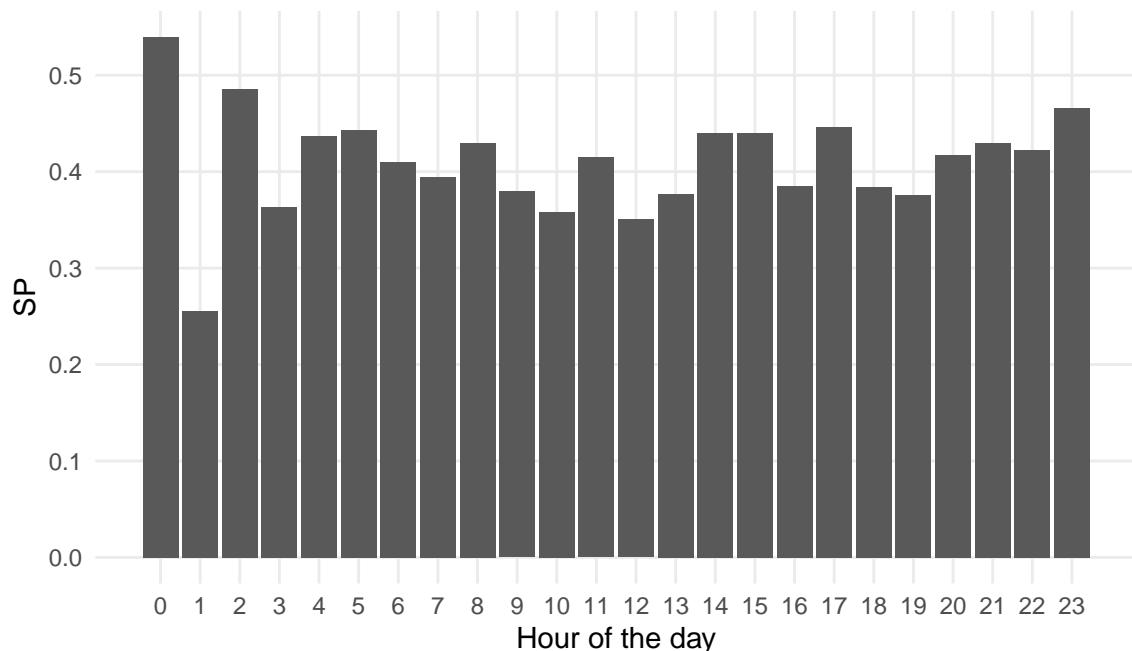
All 16 drivers are classified into habitual speeders, in which they regularly speed and have large proportions of speeding in their trips. Situational speeding happens when a high level of speeding per trip happens a few times, indicating that drivers classified in this category usually do not speed. Incidental speeding happens in small proportions of trips and in a few trips, representing unintentional speeding events. Casual speeding consists of higher percentage of trips with any speeding, but with low average speeding

FIGURE 39 – SPEEDING PER DAY OF THE WEEK



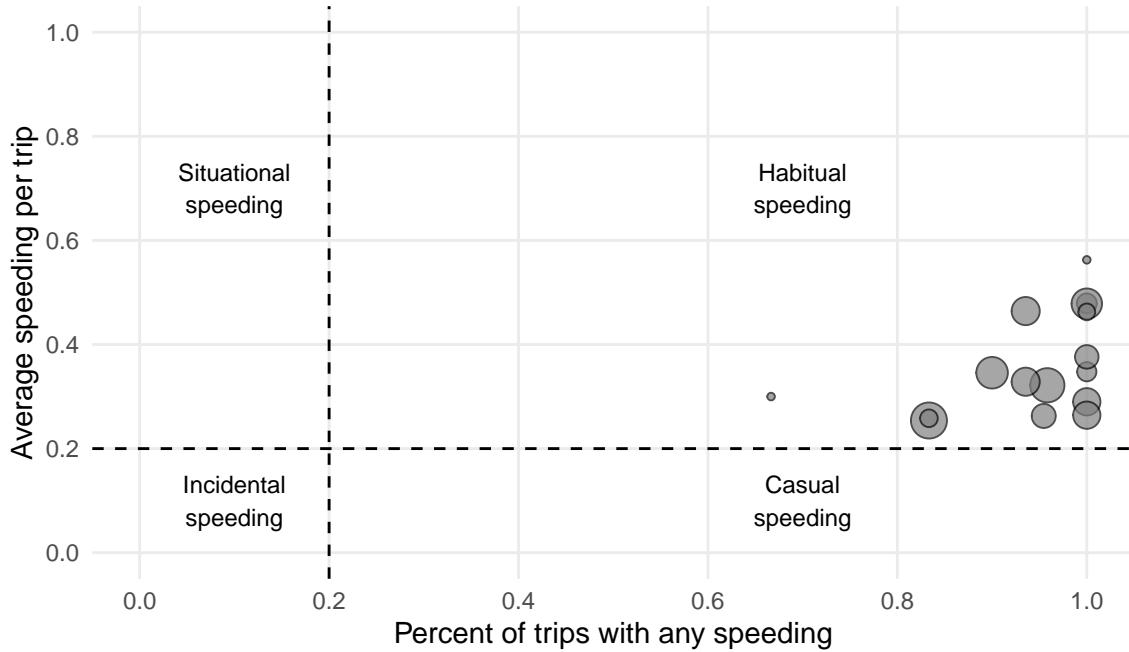
SOURCE: The Author (2021)

FIGURE 40 – SPEEDING PER HOUR OF THE DAY



SOURCE: The Author (2021)

FIGURE 41 – TYPES OF SPEEDING BEHAVIOR OF DRIVERS



SOURCE: The Author (2021), based on Richard, Campbell, Lichtry, et al. (2013)

NOTE: Dot sizes indicates quantity of trips.

per trip.

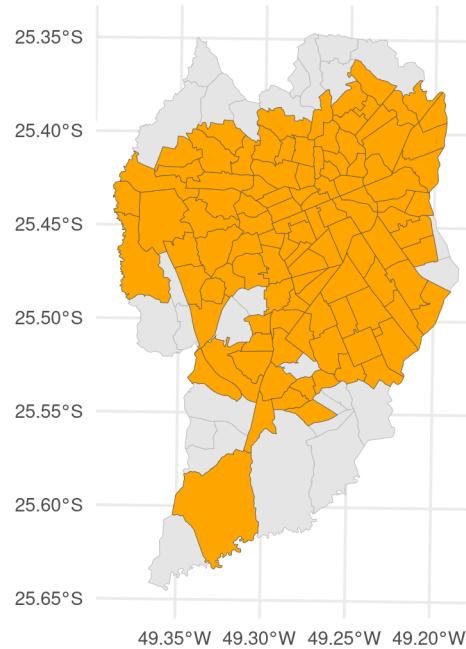
#### 4.3 GEOGRAPHICALLY WEIGHTED REGRESSION RESULTS

The data arrangement process reduced the total sample of 135 TAZs into 107 TAZs. This resulted into a remaining sample of 3,425.04 kilometers of total travelled distance, which includes 2060.31 kilometers performed in free-flow speeds and 840.88 kilometers performed in speeding. FIGURE 42 shows the remaining TAZs of the sample.

TABLE 16 includes the descriptive statistics of the 11 variables, including speeding, and considering the 107 TAZs. The value of speeding for each TAZ varied between 0.14 and 0.84, with a mean value of 0.43. The map in FIGURE 43 shows the speeding results for each TAZ. DSC, PAR and TSD values resulted in zero for some TAZs, indicating a lack of speed cameras, arterial roads and traffic signals in these zones.

The final step of the data arrangement process was the check of correlation between all independent variables, using the Spearman method. The results of this process are displayed in the correlogram of FIGURE 44. Two Spearman coefficients resulted in values above 0.8: in the correlation between street network density (SND) and density of intersections (DIS); and between traffic signal density (TSD) and density of commercial and services units (DCSU). In order to resolve this issue, it was decided

FIGURE 42 – REMAINING TAZS AFTER DATA ARRANGEMENT



SOURCE: The Author (2021)

TABLE 16 – DESCRIPTIVE STATISTICS OF VARIABLES

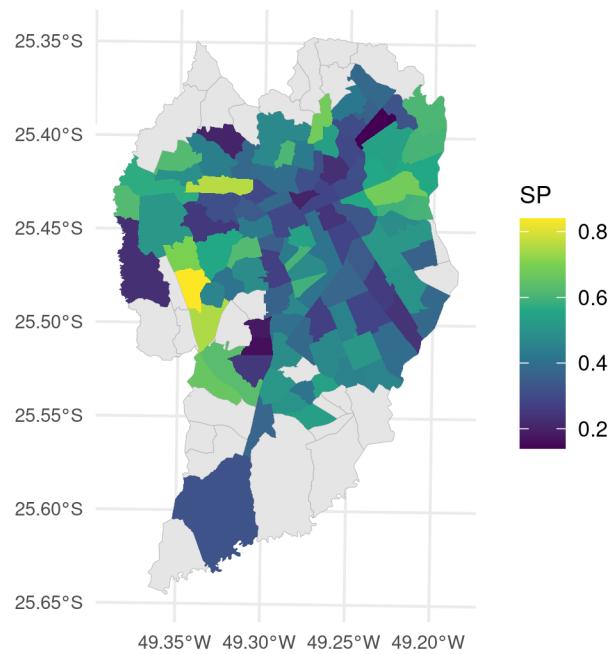
Variable [unit]	Mean	SD	Min.	1Q	Median	3Q	Max.
AVI [BRL]	2678.82	1521.43	836.63	1462.05	2163.80	3650.05	6960.25
BSD [no./km]	3.28	2.30	0.29	1.96	2.63	3.91	14.18
DCSU [no./km <sup>2</sup> ]	1079.68	1468.19	15.08	436.20	694.81	1048.59	11632.22
DIS [no./km]	69.97	27.49	3.57	52.34	68.40	82.98	170.64
DSC [no./km]	1.08	1.55	0.00	0.00	0.61	1.42	10.85
LDI	0.63	0.21	0.00	0.55	0.66	0.77	0.99
PAR	0.06	0.06	0.00	0.00	0.05	0.08	0.32
PD [inhab./km <sup>2</sup> ]	5650.32	2997.25	114.66	3286.92	5387.50	7225.75	16140.99
SND [km/km <sup>2</sup> ]	15.16	4.04	3.68	12.61	15.72	18.12	22.36
SP	0.43	0.15	0.14	0.31	0.42	0.54	0.84
TSD [no./km]	6.77	8.93	0.00	1.50	3.47	7.61	45.06

SOURCE: The Author (2021).

to remove SND and TSD from the analysis.

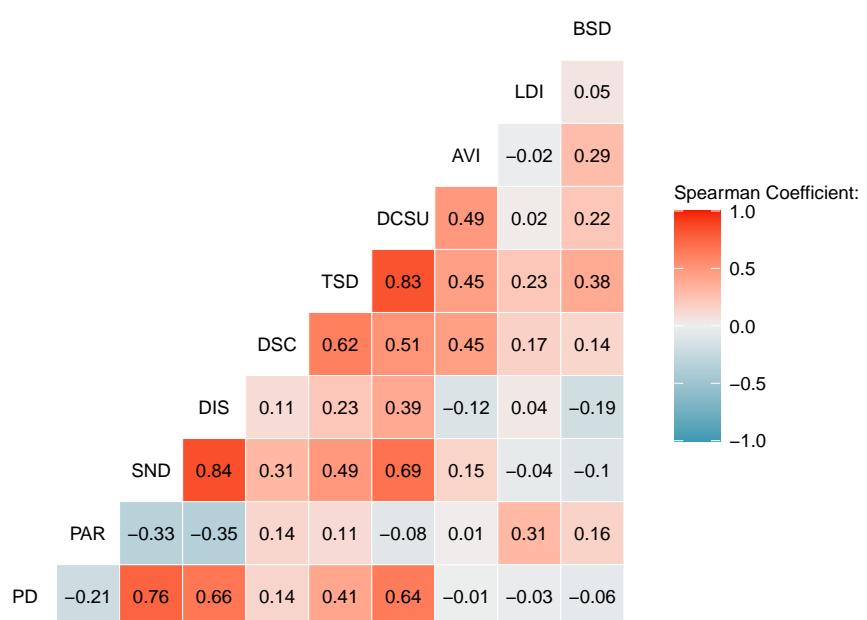
TABLE 17 shows the model diagnostics of five GWR and one global linear model, based on 4 performance indicators: AICc,  $R^2$ , Adjusted  $R^2$  and Moran's  $I$  on residuals. The  $p$ -value indicates the statistical significance of Moran's  $I$  results. The GWR model using a Boxcar kernel with a bandwidth of 39 neighbors had the best performance, with a AICc value of -119.306,  $R^2$  of 0.465, adjusted  $R^2$  of 0.329 and Moran's  $I$  on residuals of 0.166. Moran's  $I$  results shows a spatial autocorrelation in the model's residuals, therefore this model couldn't reach a satisfactory result in estimating the coefficients, considering the spatial dependency and nonstationarity of the speeding variable. The GWR model with the boxcar kernel was selected, considering it's relative

FIGURE 43 – SPEEDING PER TAZ



SOURCE: The Author (2021)

FIGURE 44 – SPEARMAN CORRELOGRAM OF INDEPENDENT VARIABLES



SOURCE: The Author (2021)

performance.

TABLE 17 – MODEL DIAGNOSTICS

<b>Model</b>	<b>Kernel</b>	<b>AICc</b>	<b>R<sup>2</sup></b>	<b>Adj. R<sup>2</sup></b>	<b>Moran's I</b>	<b>p-value</b>	<b>Bandwidth</b>
GWR	Gaussian	-116.050	0.348	0.199	0.195	0.001	37
	Bisquare	-117.757	0.406	0.243	0.174	0.004	94
	Tricube	-118.337	0.389	0.243	0.176	0.003	94
	Boxcar	-119.306	0.465	0.329	0.166	0.002	39
	Exponential	-113.918	0.354	0.178	0.201	0.002	60
	Global linear	-	-109.915	0.188	0.121	0.214	0.001

SOURCE: The Author (2021).

All GWR models presents a better performance in comparison to the global linear model, indicating that GWR is a better method to explore the relationship between speeding and the built environment. When applying Moran's *I* on the speeding variable, it resulted in a value of 0.318 (*p*-value of 0.001), revealing that speeding is spatially autocorrelated in a clustered aspect. TABLE 18 shows the coefficient estimates of the linear regression model, in addition to the standard error, *t*-values and *p*-values. The only coefficient statistical significant at the 95% level is the density of speed cameras (DSC), which express a inverted correlation to speeding.

TABLE 18 – SUMMARY OF LINEAR MODEL ESTIMATES

<b>Variable</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t-value</b>	<b>p-value</b>
Intercept	4.296e-01	6.904e-02	6.222	1.21e-08
DSC	-2.347e-02	1.087e-02	-2.160	0.0332 *
PAR	1.954e-02	2.458e-01	0.079	0.9368
DIS	6.502e-04	6.996e-04	0.929	0.3550
DCSU	-1.807e-05	1.541e-05	-1.173	0.2438
AVI	-7.446e-06	9.671e-06	-0.770	0.4432
PD	-2.393e-06	6.689e-06	-0.358	0.7213
LDI	5.063e-02	6.752e-02	0.750	0.4551
BSD	2.809e-04	7.577e-03	0.037	0.9705

SOURCE: The Author (2021).

NOTE: \* Significance at 95%

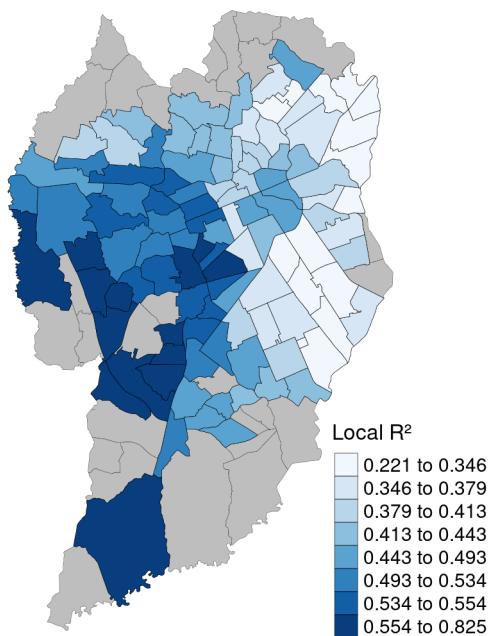
TABLE 19 contains the descriptive statistics of the chosen GWR model coefficient estimates. Maps in FIGURES 47 to 50 shows these coefficients in each TAZ. FIGURE 45 shows the variation of local geographically weighted *R*<sup>2</sup>. The western part of the city displays TAZs with higher values of *R*<sup>2</sup>, contrary to the east side of Curitiba, which displays lower values. Local *R*<sup>2</sup> values varied between 0.221 and 0.825.

FIGURE 46 displays the geographically weighted local mean and local standard deviation of the speeding variable across Curitiba. The western ou southern parts of the city presented a higher values, while the central area presented lowe values of speeding. The east side of Curitiba also has higher values of speeding, indicating a pattern where more peripheral TAZs has higher values of speeding than central areas. Regarding the

TABLE 19 – DESCRIPTIVE STATISTICS OF GWR COEFFICIENT ESTIMATES

Variable	Min.	1Q	Median	3Q	Max.
Intercept	7.0981e-02	3.0645e-01	4.5490e-01	5.5708e-01	0.7088
DSC	-6.4562e-02	-2.6972e-02	-1.5533e-02	-9.6683e-03	0.0088
PAR	-2.2749e+00	-1.6285e+00	-1.2634e+00	2.3534e-01	0.8133
DIS	-2.8407e-03	-1.5710e-03	-4.2615e-04	1.3861e-03	0.0030
DCSU	-4.5773e-04	-1.2688e-04	-2.5941e-05	-1.7180e-05	0.0000
AVI	-4.0573e-05	-1.4678e-05	-3.8717e-06	3.8475e-05	0.0002
PD	-1.0546e-05	7.5243e-07	6.7442e-06	1.2547e-05	0.0000
LDI	-3.2753e-02	4.2652e-02	9.0970e-02	1.4980e-01	0.2512
BSD	-2.1257e-02	-2.8852e-03	9.1676e-04	5.8549e-03	0.0254

SOURCE: The Author (2021).

FIGURE 45 – GW LOCAL  $R^2$ 

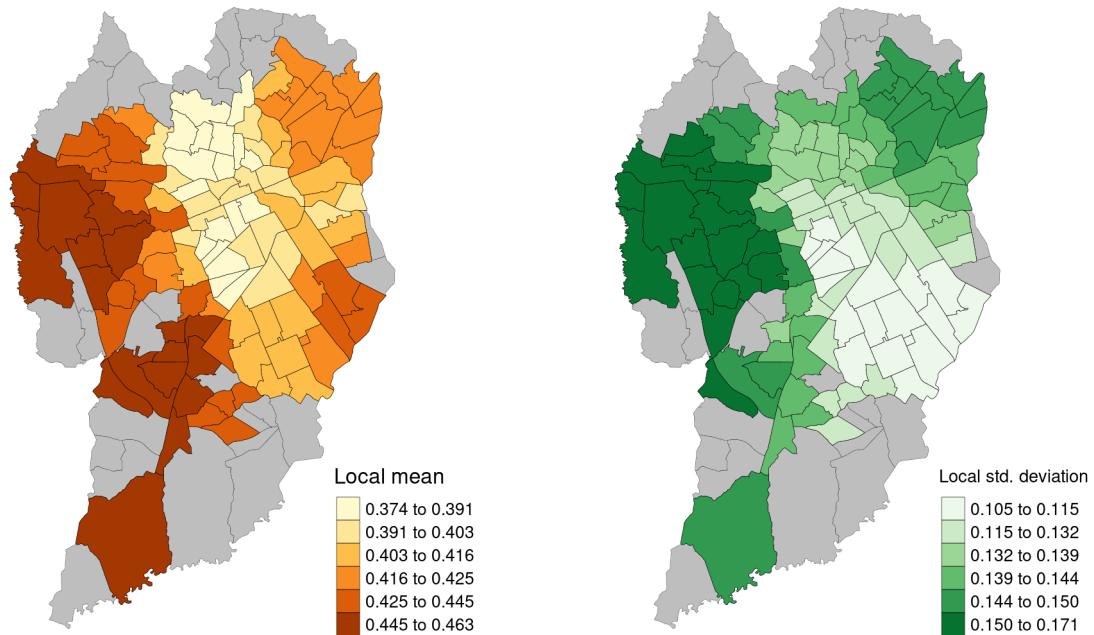
SOURCE: The Author (2021)

standard deviation (SD) of the speeding data, the western side of the city has TAZs with higher values, while the southeastern side has lower values. SD had a variation between 0.105 and 0.171.

FIGURE 47 displays the coefficient estimates of the density of speed cameras (DSC) and proportion of arterial roads (PAR). DSC presented a inverted correlation to speeding in most of the TAZs. PAR had mixed results, presenting direct and inverted correlations to speeding in different areas of the city.

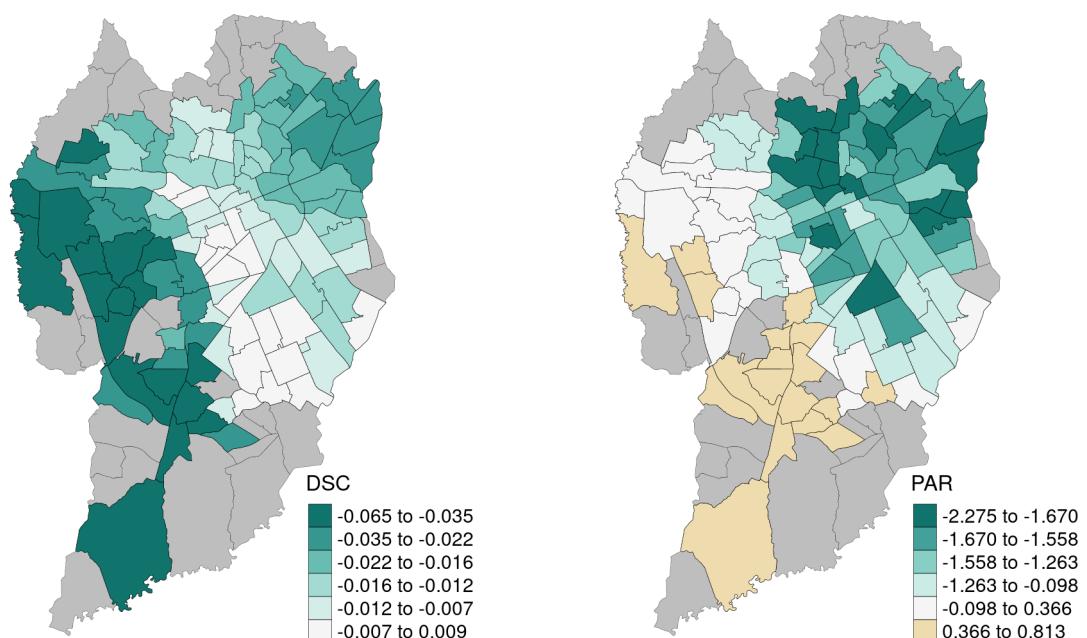
The density of intersection (DIS) and density of commercial and services units coefficient estimates are displayed in FIGURE 48. DIS has a inverted correlation to speeding in the northeast part of the city, while the west and south sides shows TAZs with a direct correlation between speeding and DIS. Regarding DCSU, almost all of the

FIGURE 46 – SPEEDING GW LOCAL MEAN AND STANDARD DEVIATION



SOURCE: The Author (2021)

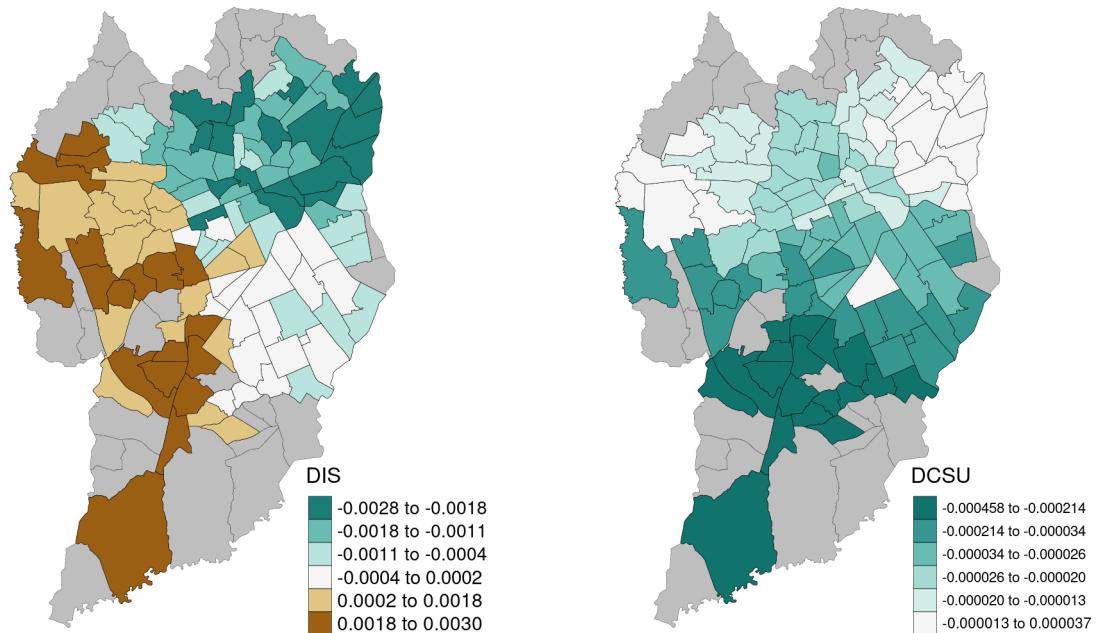
FIGURE 47 – DSC AND PAR COEFFICIENT ESTIMATES



SOURCE: The Author (2021)

TAZs shows a inverted correlation to the dependent variable.

FIGURE 48 – DIS AND DCSU COEFFICIENT ESTIMATES



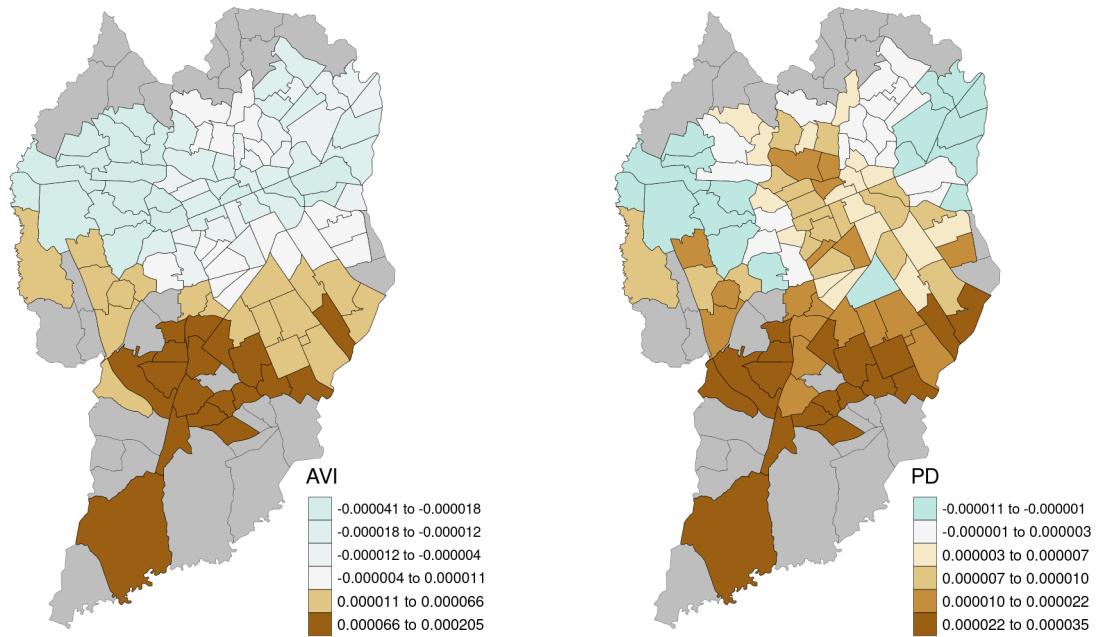
SOURCE: The Author (2021)

FIGURE 49 displays the coefficient estimates of average income (AVI) and population density (PD). AVI coefficients showed a inverted correlation to speeding in TAZs closer to the north and center of the city, while the southern side contains TAZs with a direct correlation between AVI and speeding. PD is directly correlated to speeding in most of the area, with higher values in the southern part of the city.

Land use diversity index (LDI) and bus stop density (BSD) coefficient estimates are plotted in FIGURE 50. LDI coefficients shows a direct correlation to speeding in most of the TAZs, with higher values in the northeast part of Curitiba. BSD shows mixed results, with similar quantity of TAZs with direct and inverted correlation to speeding.

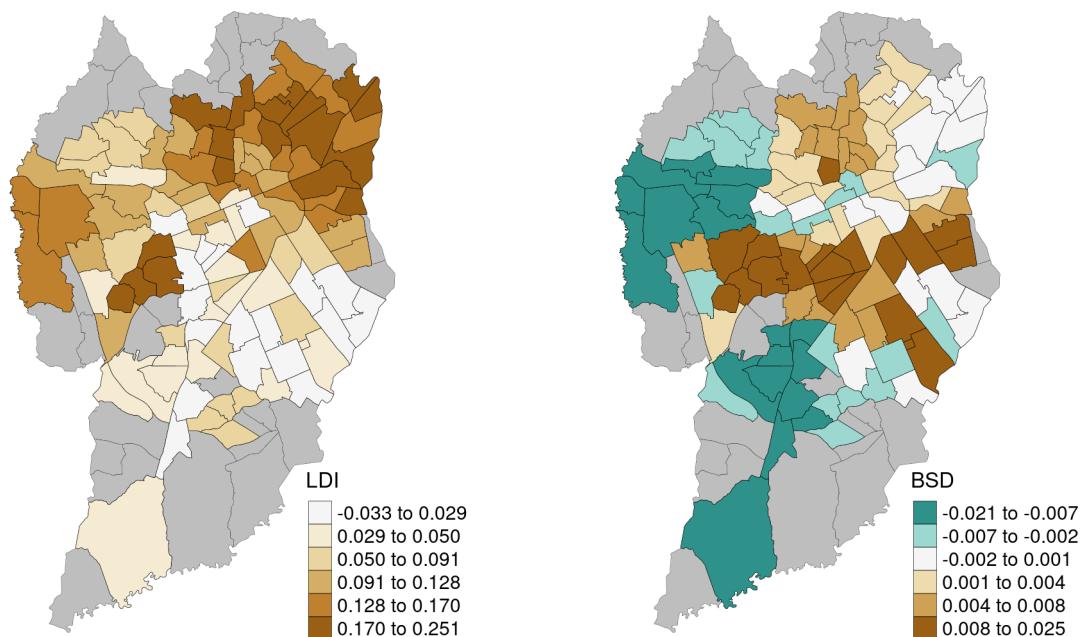
Finally, TABLE 20 presents a list with the percentage of TAZs with positive and negative coefficients, for each variable presented in the maps.

FIGURE 49 – AVI AND PD COEFFICIENT ESTIMATES



SOURCE: The Author (2021)

FIGURE 50 – LDI AND BSD COEFFICIENT ESTIMATES



SOURCE: The Author (2021)

TABLE 20 – PERCENTAGE OF TAZS WITH POSITIVE/NEGATIVE COEFFICIENTS

<b>Variable</b>	<b>Negative</b>	<b>Positive</b>
DSC	97.2%	2.8%
PAR	69.2%	30.8%
DIS	62.6%	37.4%
DCSU	99.1%	0.9%
AVI	57.9%	42.1%
PD	19.6%	80.4%
LDI	4.7%	95.3%
BSD	46.7%	53.3%

SOURCE: The Author (2021).

## 5 CONCLUSIONS

The Naturalistic Driving Study enabled the collection of speed data from a sample of 16 drivers in the city of Curitiba, Brazil. Comparing to the described NDS in Section 2.4, 16 volunteers is a small sample to conduct a more significant analysis. Some volunteers had more participation in the sample than others. Drivers  $D_{14}$  and  $D_{15}$  had the highest proportion in the total travelled distance: 19% and 23%, respectively. Regarding the spatial coverage, most of the travelled distance of the sample was performed in central areas of the city, in comparison to peripheral areas. The quantity of travelled distance wasn't enough to cover all the present TAZs in Curitiba (reducing 135 into 107).

In addition to the built environment, other attributes can influence resultant speeding values. Characteristics from the drivers (age, experience, license age, education) and from the vehicles (age, engine power) also can influence the speeding results. Using Richard, Campbell, Lichy, et al. (2013) method of speeding behavior classification, all drivers from the current sample were classified as "habitual speeders". The study would be more complete if the sample contained drivers with other speeding behaviors as well.

The Geographically Weighted Regression with a Boxcar kernel had the best performance, comparing to other kernel types and the linear regression. But none of the models presented a satisfactory performance, considering that the residuals from all models are spatially autocorrelated. This indicates that they couldn't validly predict the spatial heterogeneity present in the occurrence of speeding. All GWR models showed a better performance than the linear regression. To Brunsdon, Fotheringham, and Charlton (1996), calibration of the GWR model for regions close to the edge of the study area are subject to greater sampling error. The "holes" present in the current sample of TAZs could have caused a negative impact on the GWR performance.

Based on the results from the linear regression (TABLE 18), only the density of speed cameras (DSC) coefficient showed statistical significance. Its inverted correlation to speeding is similar to previous results found by Li, Zhang, and Ren (2020), Li, Graham, and Majumdar (2013) and Oliveira et al. (2015). The GWR model indicates a inverted correlation between DSC and speeding on 97.2% of the current sample of TAZs (FIGURE 47, TABLE 20). It is important to remind that this correlation does not imply causation, therefore further investigation is needed.

Before reaching to any further conclusions, it is important to perform new models

based on a sample with a higher number of drivers. NDS-BR is still ongoing, making able to collect more speeding data from new volunteers. More testing with the independent variables is also needed, in order to identify key variables that works better within the model.

## BIBLIOGRAPHY

- AMANCIO, Eduardo Cesar. **Influência de Dispositivos de Fiscalização Eletrônica de Velocidade no Comportamento do Condutor em Cenário Urbano**. 2021. Master Thesis – Universidade Tecnológica Federal do Paraná. Cit. on pp. 43, 47.
- AMOH-GYIMAH, Richard; SABERI, Meead; SARVI, Majid. The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models. **Analytic Methods in Accident Research**, v. 13, p. 28–51, Mar. 2017. ISSN 22136657. DOI: 10.1016/j.amar.2016.11.001. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2213665716300471>. Cit. on pp. 32, 46.
- ANDS. **ANDS Progress report for May 2017**. [S.l.: s.n.], 2017. Available from: <http://www.andsw.unsw.edu.au/news-events/ands-progress-report-may-2017>. Cit. on p. 40.
- ARVIN, Ramin; KAMRANI, Mohsen; KHATTAK, Asad J. How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data. **Accident Analysis & Prevention**, v. 127, p. 118–133, June 2019. ISSN 00014575. DOI: 10.1016/j.aap.2019.01.014. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457518304950>. Cit. on p. 46.
- ASHTON, S J. A Preliminary Assessment of the Potential for Pedestrian Injury Reduction Through Vehicle Design. **SAE Technical Papers**, Sept. 1980. DOI: 10.4271/801315. Available from: <https://www.sae.org/content/801315/>. Cit. on pp. 23, 24.
- BANSAL, Prateek; AGRAWAL, Rishabh; TIWARI, Geetam. Impacts of Bus-stops on the Speed of Motorized Vehicles under Heterogeneous Traffic Conditions: A Case-Study of Delhi, India. **International Journal of Transportation Science and Technology**, v. 3, n. 2, p. 167–178, June 2014. ISSN 20460430. DOI: 10.1260/2046-0430.3.2.167. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2046043016300946>. Cit. on p. 35.
- BASTOS, Jorge Tiago. **Análise Estratégica da Segurança viária no Brasil: Pesquisa de índices e Indicadores**. 2014. S. 290. PhD thesis – Universidade de São Paulo / Hasselt University. Cit. on p. 10.

BASTOS, Jorge Tiago; GARRONCE, Francisco Vieira, et al. **Desempenho brasileiro na década de ação pela segurança no trânsito: análise, perspectivas e indicadores 2011-2020**. Brasilia: Viva editora, 2020. P. 120. Cit. on p. 17.

BASTOS, Jorge Tiago; SANTOS, Pedro Augusto B. dos, et al. Is organized carpooling safer? Speeding and distracted driving behaviors from a naturalistic driving study in Brazil. **Accident Analysis & Prevention**, v. 152, p. 105992, Mar. 2021. ISSN 00014575. DOI: 10.1016/j.aap.2021.105992. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457521000233>. Cit. on pp. 25, 42.

\_\_\_\_\_. Naturalistic Driving Study in Brazil: An Analysis of Mobile Phone Use Behavior while Driving. **International Journal of Environmental Research and Public Health**, v. 17, n. 17, p. 6412, Sept. 2020. ISSN 1660-4601. DOI: 10.3390/ijerph17176412. Available from: <https://www.mdpi.com/1660-4601/17/17/6412>. Cit. on pp. 28, 42.

BHALLA, Kavi; MOHAN, Dinesh. Understanding the Road Safety Performance of OECD Countries. In: TIWARI, Geetam; MOHAN, Dinesh (Eds.). **Transport Planning and Traffic Safety - Making Cities, Roads and Vehicles Safer**. Boca Raton: CRC Press, 2016. P. 1–15. DOI: 10.1201/b19737-2. Cit. on pp. 12–14.

BIVAND, Roger S.; PEBESMA, Edzer; GOMEZ-RUBIO, Virgilio. **Applied spatial data analysis with R, Second edition**. [S.I.]: Springer, NY, 2013. Available from: <https://asdar-book.org/>. Cit. on pp. 55, 58.

BORGUEZANI, Jackson Rossi et al. Plataforma de coleta de dados naturalísticos de segurança viária. In: 34º Congresso de Pesquisa e Ensino em Transporte da ANPET. [S.I.: s.n.], 2020. P. 2610–2617. Cit. on pp. 48, 103.

BRASIL. **Lei nº 10.257, de 10 de julho de 2001 - Estatuto das Cidades**. Brasília: Presidência da República, 2001. Cit. on p. 10.

\_\_\_\_\_. **Lei nº 12.587, de 3 de janeiro de 2012 - Política Nacional de Mobilidade Urbana**. Brasília: Presidência da República, 2012. Cit. on p. 10.

\_\_\_\_\_. Lei nº 9.503, de 23 de setembro de 1997. **Código de trânsito Brasileiro**, 1997. Cit. on pp. 10, 54.

BRUNSDON, Chris; FOTHERINGHAM, A. Stewart; CHARLTON, Martin E. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. **Geographical Analysis**, v. 28, n. 4, p. 281–298, Sept. 1996. ISSN 00167363. DOI:

- 10.1111/j.1538-4632.1996.tb00936.x. Available from:  
<http://doi.wiley.com/10.1111/j.1538-4632.1996.tb00936.x>. Cit. on pp. 43, 44, 58, 74.
- CARSTEN, Oliver; KIRCHER, Katja; JAMSON, Samantha. Vehicle-based studies of driving in the real world: The hard truth? **Accident Analysis & Prevention**, v. 58, p. 162–174, Sept. 2013. ISSN 00014575. DOI: 10.1016/j.aap.2013.06.006. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457513002340>. Cit. on p. 39.
- CNDS. **Canada Naturalistic Driving Study (CNDS)**. [S.l.: s.n.], 2021. Available from: <https://www.canada-nds.net/index.html>. Cit. on p. 41.
- CURITIBA. Lei Nº 14.771. **Revisão do Plano Diretor de Curitiba**, 2015. Cit. on pp. 36, 37.
- \_\_\_\_\_. Lei Nº 15.511. **Zoneamento, Uso e Ocupação do Solo no Município de Curitiba**, 2019. Cit. on pp. 36, 52.
- DE VOS, Jonas; WITLOX, Frank. Transportation policy as spatial planning tool; Reducing urban sprawl by increasing travel costs and clustering infrastructure and public transportation. **Journal of Transport Geography**, Elsevier Ltd, v. 33, p. 117–125, 2013. ISSN 09666923. DOI: 10.1016/j.jtrangeo.2013.09.014. Available from: <http://dx.doi.org/10.1016/j.jtrangeo.2013.09.014>. Cit. on p. 10.
- DENATRAN. **Estatísticas - Frota de veículos**. [S.l.: s.n.], 2020. Available from: <https://www.infraestrutura.gov.br/component/content/article/115-portal-denatran/8552-estat%C7Bi%D7D%C7icas-frota-de-ve%C7B%C7Bi%D7D%C7culos-denatran.html>. Cit. on pp. 15, 16, 18, 19.
- DINH, Do Duy; KUBOTA, Hisashi. Speeding behavior on urban residential streets with a 30km/h speed limit under the framework of the theory of planned behavior. **Transport Policy**, v. 29, p. 199–208, Sept. 2013. ISSN 0967070X. DOI: 10.1016/j.tranpol.2013.06.003. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0967070X13000905>. Cit. on p. 27.
- DODGE, Yadolah. **The Concise Encyclopedia of Statistics**. New York: Springer-Verlag, 2010. P. 502. ISBN 978-0-387-31742-7. Cit. on p. 57.
- DUMBAUGH, Eric; LI, Wenhao. Designing for the safety of pedestrians, cyclists, and motorists in urban environments. **Journal of the American Planning Association**, v. 77, n. 1, p. 69–88, 2011. ISSN 01944363. DOI: 10.1080/01944363.2011.536101. Cit. on pp. 32, 33.

- DUMBAUGH, Eric; LI, Wenhao; JOH, Kenneth. The built environment and the incidence of pedestrian and cyclist crashes. **URBAN DESIGN International**, v. 18, n. 3, p. 217–228, Aug. 2013. ISSN 1357-5317. DOI: 10.1057/udi.2013.2. Available from: <http://link.springer.com/10.1057/udi.2013.2>. Cit. on pp. 31–33.
- DUMBAUGH, Eric; RAE, R. Safe Urban Form: Revisiting The Relationship Between Community Design and Traffic Safety. **Journal of the American Planning Association**, v. 75, n. 3, p. 309–329, 2009. Cit. on pp. 30, 31, 33.
- ELLISON, Adrian B.; GREAVES, Stephen P. Speeding in urban environments: Are the time savings worth the risk? **Accident Analysis & Prevention**, v. 85, p. 239–247, Dec. 2015. ISSN 00014575. DOI: 10.1016/j.aap.2015.09.018. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457515300750>. Cit. on p. 42.
- ELVIK, Rune. A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. **Accident Analysis and Prevention**, v. 50, p. 854–860, Jan. 2013. ISSN 00014575. DOI: 10.1016/j.aap.2012.07.012. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457512002667>. Cit. on p. 24.
- ELVIK, Rune et al. **The Handbook of Road Safety Measures**. Second edi. Bingley: Emerald Group Publishing Limited, 2009. ISBN 9781848552500. Cit. on pp. 10, 22, 32–34.
- EWING, Reid; CERVERO, Robert. Travel and the built environment: A meta analysis. **Journal of the American Planning Association**, v. 76, n. 3, p. 265–294, 2010. ISSN 01944363. DOI: 10.1080/01944361003766766. Cit. on pp. 29, 30.
- EWING, Reid; DUMBAUGH, Eric. The built environment and traffic safety: A review of empirical evidence. **Journal of Planning Literature**, v. 23, n. 4, p. 347–367, 2009. ISSN 08854122. DOI: 10.1177/0885412209335553. Cit. on pp. 9, 10, 29, 30, 32, 33.
- FERRAZ, Antonio Clóvis Pinto "Coca" et al. **Segurança viária**. São Carlos: [s.n.], 2012. ISBN 9788598156699. Cit. on pp. 17, 22.
- FOTHERINGHAM, A. Stewart; BRUNSDON, Christopher; CHARLTON, Martin. **Geographically Weighted Regression - the analysis of spatially varying relationships**. Chichester: John Wiley & Sons, Ltd, 2002. ISBN 0-471-49616-2. Cit. on p. 58.
- FURTH, Peter G. et al. Using Traffic Signal Control to Limit Speeding Opportunities on Bidirectional Urban Arterials. **Transportation Research Record: Journal of the**

- Transportation Research Board**, v. 2672, n. 18, p. 107–116, Dec. 2018. ISSN 0361-1981. DOI: 10.1177/0361198118790638. Available from: <http://journals.sagepub.com/doi/10.1177/0361198118790638>. Cit. on pp. 33, 34.
- GETIS, Arthur; ORD, J. K. The Analysis of Spatial Association by Use of Distance Statistics. **Geographical Analysis**, v. 24, n. 3, p. 189–206, Sept. 2010. ISSN 00167363. DOI: 10.1111/j.1538-4632.1992.tb00261.x. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.1992.tb00261.x>. Cit. on p. 58.
- GOLLINI, Isabella et al. GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models, June 2013. arXiv: 1306.0413. Available from: <http://arxiv.org/abs/1306.0413>. Cit. on pp. 44, 45, 55, 57, 58.
- GREEN, David; LEWIS, Kenneth. **Guide to Traffic Management Part 2: Traffic Theory Concepts**. 3.0. ed. Sydney: Austroads, 2020. P. 122. ISBN 978-1-925854-76-3. Cit. on p. 26.
- HADAYEGHI, Alireza; SHALABY, Amer S.; PERSAUD, Bhagwant N. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. **Accident Analysis & Prevention**, v. 42, n. 2, p. 676–688, Mar. 2010. ISSN 00014575. DOI: 10.1016/j.aap.2009.10.016. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457509002954>. Cit. on p. 46.
- HADDON, W. Advances in the epidemiology of injuries as a basis for public policy. **Public Health Reports**, v. 95, n. 5, p. 411–21, 1980. ISSN 0033-3549. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7422807>. Cit. on pp. 20, 24, 28.
- HAMZEIE, Raha; SAVOLAINEN, Peter T.; GATES, Timothy J. Driver speed selection and crash risk: Insights from the naturalistic driving study. **Journal of Safety Research**, v. 63, p. 187–194, Dec. 2017. ISSN 00224375. DOI: 10.1016/j.jsr.2017.10.007. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022437517300178>. Cit. on p. 42.
- HARVEY, D. O trabalho, o capital e o conflito de classes em torno do ambiente construído nas sociedades capitalistas avançadas. **Espaços e Debates**, Jun-Set, n. 6, p. 6–32, 1982. Cit. on p. 17.
- HIDALGO-SOLÓRZANO, Elisa et al. Prevalence of speeding and associated factors in four Mexican cities. **Journal of Epidemiology and Community Health**, jech-2019-213352, Apr. 2020. ISSN 0143-005X. DOI: 10.1136/jech-2019-213352. Available from: <http://jech.bmjjournals.org/lookup/doi/10.1136/jech-2019-213352>. Cit. on p. 27.

- HØYE, Alena. Safety effects of fixed speed cameras—An empirical Bayes evaluation. **Accident Analysis & Prevention**, v. 82, p. 263–269, Sept. 2015. ISSN 00014575. DOI: 10.1016/j.aap.2015.06.001. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457515002225>. Cit. on p. 33.
- HUANG, Yuan; WANG, Xiaoguang; PATTON, David. Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach. **Journal of Transport Geography**, v. 69, p. 221–233, May 2018. ISSN 09666923. DOI: 10.1016/j.jtrangeo.2018.04.027. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0966692317306373>. Cit. on pp. 32, 33, 44, 46, 52, 57.
- IBGE. **Censo 2010**. [S.I.: s.n.], 2021. Available from: <https://censo2010.ibge.gov.br/>. Cit. on pp. 52, 53, 55, 56.
- IPPUC. **Apresentação dos Resultados da Pesquisa Origem Destino**. [S.I.: s.n.], 2018. Available from: <https://www.ippuc.org.br/>. Cit. on pp. 52–56.
- \_\_\_\_\_. **Dados Geográficos**. [S.I.: s.n.], 2021. Available from: <https://ippuc.org.br/geodownloads/geo.htm>. Visited on: 10 July 2021. Cit. on pp. 50, 53–56.
- IPPUC; URBS. **Bus Stops in Curitiba - PR**. Curitiba: [s.n.], 2020. Cit. on pp. 55, 56.
- JØRGENSEN, N.O. Traffic Safety. In: THAGESEN, Bent (Ed.). **Highway and Traffic Engineering in Developing Countries**. London: Chapman and Hall, 2005. chap. 6, p. 121–137. ISBN 0-203-22367-5. Cit. on p. 15.
- KIM, Karl; PANT, Pradip; YAMASHITA, Eric. Accidents and Accessibility: Measuring Influences of Demographic and Land Use Variables in Honolulu, Hawaii. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2147, n. 1, p. 9–17, Jan. 2010. ISSN 0361-1981. DOI: 10.3141/2147-02. Available from: <http://journals.sagepub.com/doi/10.3141/2147-02>. Cit. on p. 35.
- KNOFLACHER, Hermann. Road Safety in Urban Areas. In: TIWARI, Geetam; MOHAN, Dinesh (Eds.). **Transport Planning & Traffic Safety - Making Cities, Roads & Vehicles Safer**. Boca Raton: CRC Press, 2016. chap. 17, p. 223–240. Cit. on pp. 9, 38.
- KONG, Xiaoqiang et al. Understanding speeding behavior from naturalistic driving data: Applying classification based association rule mining. **Accident Analysis & Prevention**, v. 144, p. 105620, Sept. 2020. ISSN 00014575. DOI:

- 10.1016/j.aap.2020.105620. Available from:  
<https://linkinghub.elsevier.com/retrieve/pii/S0001457519315593>. Cit. on p. 42.
- KOSHY, Reebu Zachariah; ARASAN, V. Thamizh. Influence of Bus Stops on Flow Characteristics of Mixed Traffic. **Journal of Transportation Engineering**, v. 131, n. 8, p. 640–643, Aug. 2005. ISSN 0733-947X. DOI: 10.1061/(ASCE)0733-947X(2005)131:8(640). Cit. on p. 35.
- LARUE, Gregoire S. et al. Visualising data of the Australian Naturalistic Driving Study. In: 8TH International Symposium on Naturalistic Driving Research. [S.l.: s.n.], 2019. Cit. on p. 40.
- LEE, Jaeyoung; ABDEL-ATY, Mohamed; JIANG, Ximiao. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. **Accident Analysis & Prevention**, v. 78, p. 146–154, May 2015. ISSN 00014575. DOI: 10.1016/j.aap.2015.03.003. Available from:  
<https://linkinghub.elsevier.com/retrieve/pii/S0001457515000779>. Cit. on pp. 31, 33, 34.
- LI, Haojie; GRAHAM, Daniel J.; MAJUMDAR, Arnab. The impacts of speed cameras on road accidents: An application of propensity score matching methods. **Accident Analysis & Prevention**, v. 60, p. 148–157, Nov. 2013. ISSN 00014575. DOI: 10.1016/j.aap.2013.08.003. Available from:  
<https://linkinghub.elsevier.com/retrieve/pii/S000145751300314X>. Cit. on pp. 33, 74.
- LI, Haojie; ZHANG, Yingheng; REN, Gang. A causal analysis of time-varying speed camera safety effects based on the propensity score method. **Journal of Safety Research**, v. 75, p. 119–127, Dec. 2020. ISSN 00224375. DOI: 10.1016/j.jsr.2020.08.007. Available from:  
<https://linkinghub.elsevier.com/retrieve/pii/S0022437520300967>. Cit. on pp. 33, 74.
- LINDLEY, D. V. Regression and Correlation Analysis. In: THE New Palgrave Dictionary of Economics. London: Palgrave Macmillan UK, 1987. P. 1–6. DOI: 10.1057/978-1-349-95121-5\_1873-1. Available from:  
[http://link.springer.com/10.1057/978-1-349-95121-5\\_1873-1](http://link.springer.com/10.1057/978-1-349-95121-5_1873-1). Cit. on p. 43.
- LOVEGROVE, Gordon R; SAYED, Tarek. Macro-level collision prediction models for evaluating neighbourhood traffic safety. **Canadian Journal of Civil Engineering**, v. 33, n. 5, p. 609–621, May 2006. ISSN 0315-1468. DOI: 10.1139/l06-013. Available from:  
<http://www.nrcresearchpress.com/doi/10.1139/l06-013>. Cit. on pp. 33, 34.

- MARSHALL, Wesley; FERENCHAK, Nicholas. Assessing equity and urban/rural road safety disparities in the US. **Journal of Urbanism: International Research on Placemaking and Urban Sustainability**, v. 10, n. 4, p. 422–441, Oct. 2017. ISSN 1754-9175. DOI: 10.1080/17549175.2017.1310748. Available from: <https://www.tandfonline.com/doi/full/10.1080/17549175.2017.1310748>. Cit. on p. 36.
- MARSHALL, Wesley; GARRICK, Norman. Does street network design affect traffic safety? **Accident Analysis and Prevention**, Elsevier Ltd, v. 43, n. 3, p. 769–781, 2011. ISSN 00014575. DOI: 10.1016/j.aap.2010.10.024. Available from: <http://dx.doi.org/10.1016/j.aap.2010.10.024>. Cit. on pp. 32, 33.
- \_\_\_\_\_. Street network types and road safety: A study of 24 California cities. **URBAN DESIGN International**, v. 15, n. 3, p. 133–147, Sept. 2010. ISSN 1357-5317. DOI: 10.1057/udi.2009.31. Available from: <http://link.springer.com/10.1057/udi.2009.31>. Cit. on pp. 33, 34.
- MINISTRY OF CITIES. **National Plan of Reduction of Deaths and Injuries**. [S.I.: s.n.], 2018. Cit. on p. 9.
- MINISTRY OF HEALTH. **Number of traffic fatalities in Brazil**. [S.I.]: Ministry of Health, 2020. Available from: <http://www2.datasus.gov.br/DATASUS/index.php?area=02>. Cit. on pp. 14–17, 19.
- \_\_\_\_\_. **Resident population in Brazil**. [S.I.: s.n.], 2021. Available from: <http://www2.datasus.gov.br/DATASUS/index.php?area=02>. Cit. on pp. 15, 16, 18, 19.
- MOHAN, D. Road Traffic Injury as a Public Health Problem. In: TIWARI, Geetam; MOHAN, Dinesh (Eds.). **Transport Planning and Traffic Safety - Making Cities, Roads and Vehicles Safer**. Boca Raton: CRC Press, 2016. chap. 2, p. 18–28. ISBN 978-1-4987-5147-6. Cit. on pp. 20, 21.
- \_\_\_\_\_. Speed and its Effects on Road Traffic Crashes. In: TIWARI, Geetam; MOHAN, Dinesh (Eds.). **Transport Planning and Traffic Safety - Making Cities, Roads and Vehicles Safer**. Boca Raton: CRC Press, 2016. chap. 9, p. 127–137. ISBN 978-1-4987-5147-6. Cit. on pp. 9, 22, 24, 25.
- MORENO, Ana Tsui; GARCÍA, Alfredo. Use of speed profile as surrogate measure: Effect of traffic calming devices on crosstown road safety performance. **Accident Analysis & Prevention**, v. 61, p. 23–32, Dec. 2013. ISSN 00014575. DOI: 10.1016/j.aap.2012.10.013. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457512003697>. Cit. on p. 28.

NAKAYA, T. et al. Geographically weighted Poisson regression for disease association mapping. **Statistics in Medicine**, v. 24, n. 17, p. 2695–2717, Sept. 2005. ISSN 02776715. DOI: 10.1002/sim.2129. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/sim.2129>. Cit. on p. 45.

NEALE, Vicki L. et al. **An Overview of the 100-Car Naturalistic Driving Study and Findings**. Washington, D.C., 2005. Available from: [https://dot.alaska.gov/highwaysafety/assets/Occ\\_Prot-cellphone\\_National\\_study-Virginia\\_Tech\\_Transp\\_Institute.pdf](https://dot.alaska.gov/highwaysafety/assets/Occ_Prot-cellphone_National_study-Virginia_Tech_Transp_Institute.pdf). Cit. on p. 40.

NES, Nicole van et al. The potential of naturalistic driving for in-depth understanding of driver behavior: UDRIVE results and beyond. **Safety Science**, v. 119, p. 11–20, Nov. 2019. ISSN 09257535. DOI: 10.1016/j.ssci.2018.12.029. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0925753517320945>. Cit. on p. 41.

NJORD, John; STEUDLE, Kirk. Big Data Hit the Road - The First Year of Use of the SHRP 2 Safety Databases. **TR News**, p. 2–8, 2015. Available from: <http://onlinepubs.trb.org/onlinepubs/trnews/trnews300BigData.pdf>. Cit. on p. 40.

OBELHEIRO, Marta Rodrigues; SILVA, Alan Ricardo da; NODARI, Christine Tessele, et al. A new zone system to analyze the spatial relationships between the built environment and traffic safety. **Journal of Transport Geography**, v. 84, p. 102699, Apr. 2020. ISSN 09666923. DOI: 10.1016/j.jtrangeo.2020.102699. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S096669231930225X>. Cit. on pp. 30–35, 45, 46.

OBELHEIRO, Marta Rodrigues; SILVA, Alan Ricardo; NODARI, Christine Tessele. UMA ANÁLISE DA RELAÇÃO ENTRE AMBIENTE CONSTRUÍDO E ACIDENTES DE TRÂNSITO EM ZONAS DE TRÁFEGO. **33º Congresso de Pesquisa e Ensino em Transporte da ANPET**, p. 3696–3707, 2019. Cit. on pp. 31–33, 35, 36, 43, 46.

OLIVEIRA, Daniele Falci de et al. Do speed cameras reduce speeding in urban areas? **Cadernos de Saúde Pública**, v. 31, suppl 1, p. 208–218, Nov. 2015. ISSN 0102-311X. DOI: 10.1590/0102-311X00101914. Available from: [http://www.scielo.br/scielo.php?script=sci%7B%5C\\_%7Darttext%7B%5C&%7Dpid=S0102-311X2015001300208%7B%5C&%7Dlng=en%7B%5C&%7Dtlng=en](http://www.scielo.br/scielo.php?script=sci%7B%5C_%7Darttext%7B%5C&%7Dpid=S0102-311X2015001300208%7B%5C&%7Dlng=en%7B%5C&%7Dtlng=en). Cit. on pp. 33, 74.

OPENSTREETMAP CONTRIBUTORS. **Planet dump retrieved from https://planet.osm.org**. [S.l.: s.n.], 2017. <https://www.openstreetmap.org>. Cit. on p. 50.

- OUYANG, Yiqiang; BEJLERI, Ilir. Geographic Information System-Based Community-Level Method to Evaluate the Influence of Built Environment on Traffic Crashes. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2432, n. 1, p. 124–132, Jan. 2014. ISSN 0361-1981. DOI: 10.3141/2432-15. Available from: <http://journals.sagepub.com/doi/10.3141/2432-15>. Cit. on pp. 31–35.
- PARK, Shin Hyung; PARK, Shin Hyoung; KWON, Oh Hoon. K-Means and CRP-Based Characteristic Investigating Method of Traffic Accidents with Automated Speed Enforcement Cameras. In: PARK, J. et al. (Eds.). **Advanced Multimedia and Ubiquitous Engineering**. Singapore: Springer, 2019. P. 631–637. DOI: 10.1007/978-981-13-1328-8\_81. Available from: [http://link.springer.com/10.1007/978-981-13-1328-8%7B%5C\\_%7D81](http://link.springer.com/10.1007/978-981-13-1328-8%7B%5C_%7D81). Cit. on p. 33.
- PEREIRA, Djenifer Renata; BORGUEZANI, Jackson Rossi. **gps2csv**. [S.I.]: GitHub, 2020. Available from: <https://github.com/djeni98/gps2csv>. Cit. on p. 99.
- PIRDAVANI, Ali et al. Application of Geographically Weighted Regression Technique in Spatial Analysis of Fatal and Injury Crashes. **Journal of Transportation Engineering**, v. 140, n. 8, p. 04014032, Aug. 2014. ISSN 0733-947X. DOI: 10.1061/(ASCE)TE.1943-5436.0000680. Available from: <http://ascelibrary.org/doi/10.1061/%7B%5C%7D28ASCE%7B%5C%7D29TE.1943-5436.0000680>. Cit. on pp. 31, 46.
- RHEE, Kyoung-Ah et al. Spatial regression analysis of traffic crashes in Seoul. **Accident Analysis & Prevention**, v. 91, p. 190–199, June 2016. ISSN 00014575. DOI: 10.1016/j.aap.2016.02.023. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457516300562>. Cit. on pp. 31, 32, 46.
- RICHARD, Christian M.; BROWN, James L., et al. Using Naturalistic Driving Data to Develop a Typology of Speeding Episodes. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2659, n. 1, p. 91–97, Jan. 2017. ISSN 0361-1981. DOI: 10.3141/2659-10. Available from: <http://journals.sagepub.com/doi/10.3141/2659-10>. Cit. on p. 41.
- RICHARD, Christian M.; CAMPBELL, John; BROWN, James, et al. Investigating speeding behavior with naturalistic approaches. **Transportation Research Record**, v. 2365, n. 2365, p. 58–65, Jan. 2013. ISSN 03611981. DOI: 10.3141/2365-08. Available from: <http://journals.sagepub.com/doi/10.3141/2365-08>. Cit. on pp. 27, 28, 41, 51.

- RICHARD, Christian M.; CAMPBELL, John L.; LICHTY, Monica G., et al. **Motivations for Speeding, Volume II: Findings Report.** Washington, D.C., 2013. Available from: <https://www.nhtsa.gov/sites/nhtsa.gov/files/811818.pdf>. Cit. on pp. 25, 63, 65, 74.
- RICHARD, Christian M.; LEE, Joonbum, et al. Using SHRP2 naturalistic driving data to examine driver speeding behavior. **Journal of Safety Research**, v. 73, p. 271–281, June 2020. ISSN 00224375. DOI: 10.1016/j.jsr.2020.03.008. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022437520300384>. Cit. on pp. 41, 42.
- ROSÁRIO, Maria do Rocio. Curitiba Revisited: Five Decades of Transformation. **Architectural Design**, v. 86, n. 3, p. 112–117, May 2016. ISSN 00038504. DOI: 10.1002/ad.2053. Available from: <http://doi.wiley.com/10.1002/ad.2053>. Cit. on p. 36.
- SETRAN. **Speed Cameras in Curitiba - PR.** Curitiba: Secretaria Municipal de Defesa Social e Transito de Curitiba, 2020. Cit. on p. 54.
- SHINAR, D. **Traffic Safety and Human Behavior.** Ed. by David Shinar. [S.I.]: Emerald Group Publishing Limited, Oct. 2017. ISBN 978-0-08-045029-2. DOI: 10.1108/9780080555874. Available from: <https://www.emerald.com/insight/publication/doi/10.1108/9780080555874>. Cit. on pp. 21, 22, 24, 25, 27, 28, 38, 39.
- SILVA, Alan Ricardo da; RODRIGUES, Thais Carvalho Valadares. Geographically Weighted Negative Binomial Regression—incorporating overdispersion. **Statistics and Computing**, May 2013. ISSN 0960-3174. DOI: 10.1007/s11222-013-9401-9. Available from: <http://link.springer.com/10.1007/s11222-013-9401-9>. Cit. on p. 45.
- SWOV. **Sustainably safe road infrastructure: main requirements.** [S.I.], 2003. Cit. on p. 21.
- TIWARI, Geetam. Land Use-Transportation Planning, Mobility and Safety. In: TIWARI, Geetam; MOHAN, Dinesh (Eds.). **Transport Planning & Traffic Safety - Making Cities, Roads & Vehicles Safer.** [S.I.]: CRC Press, 2016. chap. 4, p. 45–58. ISBN 978-1-4987-5147-6. Cit. on pp. 29, 36.
- UKKUSURI, Satish et al. The role of built environment on pedestrian crash frequency. **Safety Science**, v. 50, n. 4, p. 1141–1151, Apr. 2012. ISSN 09257535. DOI: 10.1016/j.ssci.2011.09.012. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0925753511002578>. Cit. on p. 33.
- VASCONCELLOS, Eduardo Alcântara de. **Políticas de Transporte no Brasil - A Construção da Mobilidade Excludente.** [S.I.]: Editora Manole, 2013. Cit. on p. 17.

- WANG, Xuesong et al. Speed, speed variation and crash relationships for urban arterials. **Accident Analysis & Prevention**, v. 113, p. 236–243, Apr. 2018. ISSN 00014575. DOI: 10.1016/j.aap.2018.01.032. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457518300381>. Cit. on p. 28.
- WARREN, Josh; LIPKOWITZ, Jeff; SOKOLOV, Vadim. Clusters of Driving Behavior From Observational Smartphone Data. **IEEE Intelligent Transportation Systems Magazine**, v. 11, n. 3, p. 171–180, 2019. ISSN 1939-1390. DOI: 10.1109/MITS.2019.2919516. Available from: <https://ieeexplore.ieee.org/document/8743351/>. Cit. on p. 28.
- WATSON, B. et al. Profiling high-range speeding offenders: Investigating criminal history, personal characteristics, traffic offences, and crash history. **Accident Analysis & Prevention**, v. 74, p. 87–96, Jan. 2015. ISSN 00014575. DOI: 10.1016/j.aap.2014.10.013. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457514003108>. Cit. on p. 27.
- WEI, Feng; LOVEGROVE, Gordon. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. **Accident Analysis & Prevention**, v. 61, p. 129–137, Dec. 2013. ISSN 00014575. DOI: 10.1016/j.aap.2012.05.018. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0001457512002023>. Cit. on pp. 35, 43.
- WELLE, Ben et al. **O Desenho de Cidades Seguras**. [S.I.], 2016. Cit. on pp. 9, 28, 31, 33–35.
- WHO. **Global Plan for the Decade of Action for Road Safety 2011-2020**. [S.I.], 2011. Cit. on pp. 9, 14.
- \_\_\_\_\_. **Global status report on road safety 2013 - supporting a decade of action**. [S.I.], 2013. P. 1–318. Available from: [http://www.who.int/about/licensing/copyright\\_form/en/index.html](http://www.who.int/about/licensing/copyright_form/en/index.html). Cit. on pp. 10, 22.
- \_\_\_\_\_. **Global status report on road safety 2018**. Geneva, 2018. Cit. on pp. 10, 12, 14, 18.
- \_\_\_\_\_. **Speed management: a road safety manual for decision-makers and practitioners**. Geneva: [s.n.], 2008. ISBN 978-2-940395-04-0. Cit. on pp. 24–27.
- \_\_\_\_\_. **Stockholm Declaration - Third Global Ministerial Conference on Global Safety: Achieving Global Goals 2030**. Stockholm, Feb. 2020. --undefined. Cit. on p. 9.

- WHO. **World Report on Road Traffic Injury Prevention.** [S.I.], 2004. P. 238. ISBN 9789241562607. Available from:  
<http://whqlibdoc.who.int/publications/2004/9241562609.pdf>. Cit. on p. 21.
- YADAV, Ankit Kumar; VELAGA, Nagendra R. Alcohol-impaired driving in rural and urban road environments: Effect on speeding behaviour and crash probabilities. **Accident Analysis & Prevention**, v. 140, p. 105512, June 2020. ISSN 00014575. DOI: 10.1016/j.aap.2020.105512. Available from:  
<https://linkinghub.elsevier.com/retrieve/pii/S0001457519315866>. Cit. on p. 28.
- YU, Chia-Yuan; XU, Minjie. Local Variations in the Impacts of Built Environments on Traffic Safety. **Journal of Planning Education and Research**, p. 0739456x1769603, Mar. 2017. ISSN 0739-456X. DOI: 10.1177/0739456X17696035. Available from:  
<http://journals.sagepub.com/doi/10.1177/0739456X17696035>. Cit. on pp. 33, 45, 46.
- ZERVAS, Efthimios; LAZAROU, Christos. Influence of European passenger cars weight to exhaust CO<sub>2</sub> emissions. **Energy Policy**, v. 36, n. 1, p. 248–257, Jan. 2008. ISSN 03014215. DOI: 10.1016/j.enpol.2007.09.009. Available from:  
<https://linkinghub.elsevier.com/retrieve/pii/S0301421507003849>. Cit. on p. 23.
- ZHANG, Cuiping et al. Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones. **Mathematical Problems in Engineering**, v. 2014, p. 1–9, 2014. ISSN 1024-123X. DOI: 10.1155/2014/987978. Available from:  
<http://www.hindawi.com/journals/mpe/2014/987978/>. Cit. on p. 43.
- ZHANG, Y. et al. Investigating the associations between road network structure and non-motorist accidents. **Journal of Transport Geography**, v. 42, p. 34–47, Jan. 2015. ISSN 09666923. DOI: 10.1016/j.jtrangeo.2014.10.010. Available from:  
<https://linkinghub.elsevier.com/retrieve/pii/S0966692314002336>. Cit. on pp. 32, 33, 46.

## APPENDIX 1 - POINTS INTO LINESTRING CONVERSION

points2path.py:

```

1 import pandas as pd
2
3 coord = pd.read_csv('output/validtime_speeding.csv', dtype = {'id': str,
4                     'time_acum': int, 'lat': float, 'long': float})
5
6 coord2 = coord[['id', 'long', 'lat', 'time_acum', 'date', 'time', '',
7                 'spd_kmh', 'limite_vel', 'rel_spd']]
8
9 coord2['WKT'] = None
10
11 for i in range(1, len(coord2)):
12     if (coord2['time_acum'][i] - coord2['time_acum'][i-1]) == 1:
13         coord2['WKT'][i] = 'LINESTRING (' + str(coord2['long'][i]) + ', '
14         + str(coord2['lat'][i]) + ',' + str(coord2['long'][i-1]) + ',' + str(
15             coord2['lat'][i-1]) + ')'
16         coord2['rel_spd'][i] = (coord2['rel_spd'][i] + coord2['rel_spd'][
17             i-1])/2
18     else:
19         coord2['WKT'][i] = 0
20
21 linhas = coord2.query('WKT != 0')
22
23 print(linhas.head())
24
25 linhas.to_csv('output/validtime_speeding_line.csv', sep = ';', columns =
26                 ['WKT', 'id', 'long', 'lat', 'time_acum', 'date', 'time', 'spd_kmh',
27                  'limite_vel', 'rel_spd'])

```

## APPENDIX 2 - GWR R SCRIPT

```

1 library(GWmodel)
2 library(rgdal)
3 library(tidyverse)
4 library(spdep)
5 library(tmap)
6 library(xtable)
7
8 # Taz import and tidy ----
9
10 taz_gwr <- readOGR("input", "taz_quali")
11
12 taz_data_tidy <- function(taz_shape){
13   taz_shape@data <- taz_shape@data %>%
14     mutate(prop = total_leng / road_l)
15
16   taz_shape <- taz_shape[!is.na(taz_shape$freeflow_l),]
17
18   taz_shape <- taz_shape[taz_shape$prop > 0.1,]
19
20   taz_shape@data <- taz_shape@data %>%
21     select(-fid, -BAIRRO, -area, -total_leng, -freeflow_l, -speeding_l,
22           -road_l,
23             -prop, -pop)
24
25   return(taz_shape)
26 }
27
28 taz_gwr <- taz_data_tidy(taz_gwr)
29
30 # Sample size ----
31
32 sample_taz <- readOGR("input", "taz_quali")
33
34 calc_sample_size <- function(sample){
35   sample_taz@data <- sample_taz@data %>%
36     mutate(prop = total_leng / road_l)
37
38   sample_taz <- sample_taz[!is.na(sample_taz$freeflow_l),]
39
40   sample_taz <- sample_taz[sample_taz$prop > 0.1,]

```

```

41 total <- sample_taz@data %>%
42   select(total_leng) %>%
43   sum()
44
45 ff <- sample_taz@data %>%
46   select(freeflow_l) %>%
47   sum()
48
49 sp <- sample_taz@data %>%
50   select(speeding_l) %>%
51   sum()
52
53 results <- c(total, ff, sp)
54
55 names(results) <- c("Total", "Free-flow", "Speeding")
56
57 return(results)
58 }
59
60 sample_size <- calc_sample_size(sample_taz)
61
62 # Variable global summary ----
63
64 make_global_summary <- function(taz_shape){
65   df <- taz_shape@data %>%
66     select(-cod_taz) %>%
67     pivot_longer(PD:BSD, names_to = "var", values_to = "value") %>%
68     group_by(var) %>%
69     summarise(mean = mean(value),
70               sd = sd(value),
71               min = min(value),
72               `1q` = quantile(value, 0.25),
73               median = median(value),
74               `3q` = quantile(value, 0.75),
75               max = max(value))
76
77   xtable(df, type = "latex")
78 }
79
80 global_summary <- make_global_summary(taz_gwr)
81
82 print(global_summary)
83
84 # GWR function ----
85

```

```

86 gwr <- function(taz_shape){
87   # Sort independent variables for best fit
88   ind_var <- c("PD", "PAR", "DIS", "DSC", "DCSU", "AVI", "LDI", "BSD")
89
90   sort_var <- function(taz){
91     model_select <- model.selection.gwr(
92       DeVar = "SP",
93       InDeVars = ind_var,
94       data = taz,
95       bw = 30,
96       approach = "AIC",
97       kernel = "gaussian",
98       adaptive = TRUE
99     )
100    return(as.formula(model_select[[1]][[36]][[1]]))
101  }
102
103  sorted_formula <- sort_var(taz_shape)
104
105  # Extracting bw size for each kernel type
106  kernel_type <- c("gaussian", "bisquare", "tricube", "boxcar", "exponential")
107
108  bw_sizes <- vector(mode = "integer", length = 5)
109
110  calc_bw_sizes <- function(kernel){
111    bw.gwr(formula = sorted_formula,
112            data = taz_shape,
113            approach = "AIC",
114            kernel = kernel,
115            adaptive = TRUE)
116  }
117
118  bw_sizes <- map_dbl(kernel_type, calc_bw_sizes)
119
120  names(bw_sizes) <- kernel_type
121
122  # Running a GWR for each kernel type
123  gwr_calc <- function(bw, kernel){
124    gwr.basic(formula = sorted_formula,
125              data = taz_shape,
126              bw = bw,
127              kernel = kernel,
128              adaptive = TRUE)
129  }

```

```

130
131 gwr_results <- vector(mode = "list", length = 5)
132
133 gwr_results <- map2(bw_sizes, kernel_type, gwr_calc)
134
135 return(gwr_results)
136 }
137
138 gwr_model_results <- gwr(taz_gwr)
139
140 # Extracting diagnostic for each GWR model ----
141
142 extract_diagnostic <- function(gwr){
143   df <- tibble(test = c("RSS.gw", "AIC", "AICc", "enp", "edf", "gw.R2",
144     "gwR2.adj", "BIC",
145     "bandwidth"),
146     gaussian = c(unlist(gwr[[1]][["GW.diagnostic"]]),
147       gwr[["gaussian"]][["GW.arguments"]][["bw"]])
148   ,
149   bisquare = c(unlist(gwr[[2]][["GW.diagnostic"]]),
150     gwr[["bisquare"]][["GW.arguments"]][["bw"]])
151   ,
152   tricube = c(unlist(gwr[[3]][["GW.diagnostic"]]),
153     gwr[["tricube"]][["GW.arguments"]][["bw"]]),
154   boxcar = c(unlist(gwr[[4]][["GW.diagnostic"]]),
155     gwr[["boxcar"]][["GW.arguments"]][["bw"]]),
156   exponential = c(unlist(gwr[[5]][["GW.diagnostic"]]),
157     gwr[["exponential"]][["GW.arguments"]][["bw"]]))
158
159   df %>%
160     pivot_longer(-test, names_to = "kernel", values_to = "value") %>%
161     pivot_wider(names_from = test, values_from = value)
162 }
163
164 diagnostic_table <- extract_diagnostic(gwr_model_results)
165
166 # Moran's I on residuals ----
167
168 calc_moran <- function(gwr_model_data){
169   # Extracting neighbors
170   nb <- poly2nb(gwr_model_data[[1]][["SDF"]], queen = TRUE)
171
172   # Setting weights for each neighbor
173   lw <- nb2listw(nb, style = "W", zero.policy = TRUE)

```

```

171
172 # Moran's I with Monte Carlo Simulation for gwr
173 gwr_mmc <- function(gwr_model_data){
174
175   moran.mc(gwr_model_data[["SDF"]]$residual, lw, nsim = 999,
176   alternative = "greater")
177 }
178
179 gwr_mmc_results <- map(gwr_model_data, gwr_mmc)
180
181 global_mmc <- moran.mc(gwr_model_data[[1]][["lm"]]$residuals, lw, nsim
182 = 999,
183   alternative = "greater")
184
185 # Results table
186 results <- tibble(
187   model = c("gaussian", "bisquare", "tricube", "boxcar", "exponential"
188   , "global"),
189   i = c(gwr_mmc_results[[1]][["statistic"]], gwr_mmc_results[[2]][["statistic"]],
190   gwr_mmc_results[[3]][["statistic"]], gwr_mmc_results
191   [[4]][["statistic"]],
192   gwr_mmc_results[[5]][["statistic"]], global_mmc[["statistic"]]),
193   p_value = c(gwr_mmc_results[[1]][["p.value"]], gwr_mmc_results
194   [[2]][["p.value"]],
195   gwr_mmc_results[[3]][["p.value"]], gwr_mmc_results
196   [[4]][["p.value"]],
197   gwr_mmc_results[[5]][["p.value"]], global_mmc[["p.value"]]
198   ))
199 )
200
201 return(results)
202 }
203
204 morans_i <- calc_moran(gwr_model_results)
205
206 # Moran's I on SP ----
207
208 calc_moran_sp <- function(taz_shape){
209   nb <- poly2nb(taz_shape, queen = TRUE)
210   lw <- nb2listw(nb, style = "W", zero.policy = TRUE)
211   moran.mc(taz_shape$SP, lw, nsim = 999, alternative = "greater")
212 }
213
214
215 
```

```

207 sp_moran_results <- calc_moran_sp(taz_gwr)
208
209 # Selecting best model ----
210 # manual process, check diagnostic (code to be implemented ...)
211
212 gwr_chosen_model <- gwr_model_results[[4]]
213
214 # GW summary on speeding (mean and SD maps) ----
215
216 summary_maps <- function(taz_shape){
217   vars = c("SP", "PD", "PAR", "DIS", "DSC", "DCSU", "AVI", "LDI", "BSD")
218
219   # Summary calc
220   summary <- gwss(taz_shape,
221                     vars = vars,
222                     bw = gwr_chosen_model[["GW.arguments"]][["bw"]],
223                     kernel = gwr_chosen_model[["GW.arguments"]][["kernel"]]
224                   ],
225                     adaptive = TRUE,
226                     quantile = TRUE)
227
228   # Base map import
229   taz <- st_read("input", "taz_quali")
230
231   # Mean map
232   sp_lm <- tm_shape(taz) +
233     tm_fill(col="grey") +
234     tm_borders(col="black", lwd=0.1) +
235     tm_shape(summary[["SDF"]]) +
236     tm_fill(col="SP_LM", n = 6, style="quantile", title = "Local mean")
237     +
238     tm_borders(col="black", lwd=0.2) +
239     tm_layout(frame=FALSE) +
240     tm_legend(legend.position = c(0.82,0.00),
241               legend.title.size = 0.8,
242               legend.text.size = 0.6)
243
244   # SD map
245   sp_lsd <- tm_shape(taz) +
246     tm_fill(col="grey") +
247     tm_borders(col="black", lwd=0.1) +
248     tm_shape(summary[["SDF"]]) +
249     tm_fill(col="SP_LSD", n = 6, style="quantile", title = "Local std.
250     deviation",
251               palette = "Greens") +

```

```

249   tm_borders(col="black", lwd=0.2) +
250   tm_layout(frame=FALSE) +
251   tm_legend(legend.position = c(0.82,0.00),
252             legend.title.size = 0.8,
253             legend.text.size = 0.6)
254
255 results <- list(sp_lm, sp_lsd)
256
257 names(results) <- c("sp_lm", "sp_lsd")
258
259 return(results)
260 }
261
262 gwr_summary <- summary_maps(taz_gwr)
263
264 ## Save summary plots ----
265
266 tmap_save(tm = gwr_summary[[1]],
267            filename = "plots/mean_map.png",
268            height = 3.5,
269            width = 3,
270            units = "in",
271            dpi = 300)
272
273 tmap_save(tm = gwr_summary[[2]],
274            filename = "plots/sd_map.png",
275            height = 3.5,
276            width = 3,
277            units = "in",
278            dpi = 300)
279
280
281 # Plot GWR results ----
282
283 plot_results <- function(gwr_model_data){
284   # Import base map
285   taz <- st_read("input", "taz_quali")
286
287   # Variables
288   vars <- c(colnames(gwr_model_data[["SDF"]])@data[2:9])
289
290   make_maps <- function(var){
291     maps <- tm_shape(taz) +
292       tm_fill(col="grey") +
293       tm_borders(col="black", lwd=0.1) +

```

```

294     tm_shape(gwr_model_data[["SDF"]]) +
295     tm_fill(col = var, n = 6, style = "quantile", palette = "-BrBG") +
296     tm_borders(col="black", lwd = 0.2) +
297     tm_layout(frame=FALSE, legend.width = 0.5) +
298     tm_legend(legend.position = c(0.75,0.00),
299                 legend.title.size = 0.8,
300                 legend.text.size = 0.6)
301   }
302
303 result_maps <- map(vars, make_maps)
304
305 return(result_maps)
306 }
307
308 gwr_results_maps <- plot_results(gwr_chosen_model)
309
310 ## Save gwr results maps ----
311
312 names <- colnames(gwr_chosen_model[["SDF"]][data[2:9]])
313
314 save_result_maps <- function(tm, names){
315   tmap_save(tm = tm,
316             filename = paste("plots/", names, ".png", sep = ""),
317             height = 3.5,
318             width = 3,
319             units = "in",
320             dpi = 300)
321 }
322
323 map2(gwr_results_maps, names, save_result_maps)
324
325 # Plot R-squared ----
326
327 plot_r2 <- function(gwr_model_data){
328   # Import base map
329   taz <- st_read("input", "taz_quali")
330
331   maps <- tm_shape(taz) +
332     tm_fill(col="grey") +
333     tm_borders(col="black", lwd=0.1) +
334     tm_shape(gwr_model_data[["SDF"]]) +
335     tm_fill(col = "Local_R2", n = 8, style = "quantile", palette =
336               "Blues",
337               title = "Local R2 ") +
338     tm_borders(col="black", lwd = 0.2) +

```

```

338     tm_layout(frame=FALSE, legend.width = 0.6) +
339     tm_legend(legend.position = c(0.82,0.00),
340                 legend.title.size = 0.8,
341                 legend.text.size = 0.6)
342 }
343
344 r2_map <- plot_r2(gwr_chosen_model)
345
346 tmap_save(tm = r2_map,
347             filename = "plots/r2_map.png",
348             height = 3.5,
349             width = 3,
350             units = "in",
351             dpi = 300)
352
353 # Count positive and negative coefficients per TAZ and variable ----
354
355 make_results_table <- function(gwr_model_data){
356   gwr_results <- gwr_model_data[["SDF"]]\@data %>%
357     select(Intercept:BSD)
358
359   gwr_results %>%
360     mutate(across(everything(), ~ case_when(
361       . >= 0 ~ "pos",
362       . < 0 ~ "neg",
363       TRUE ~ NA_character_
364     ))) %>%
365     pivot_longer(Intercept:BSD, names_to = "variables", values_to = "count") %>%
366     mutate(n = 1) %>%
367     group_by(variables, count) %>%
368     summarise(n = sum(n)) %>%
369     pivot_wider(names_from = count, values_from = n) %>%
370     replace(is.na(.), 0) %>%
371     mutate(prop_neg = neg / (neg+pos),
372           prop_pos = pos / (neg+pos)) %>%
373     arrange(-prop_neg)
374 }
375
376 results_table <- make_results_table(gwr_chosen_model)
377
378 results_table %>%
379   mutate(prop_pos = prop_pos * 100,
380         prop_neg = prop_neg * 100)

```

## ANNEX 1 - DATA ACQUISITION AND CONVERSION SCRIPTS

Code written by Pereira and Borguezani (2020)

gps2csv.py:

```

1 #!/usr/bin/python3
2 import pynmea2, sys, getopt
3 import pandas as pd
4 import arrow
5
6 import functions as fct
7
8 def addRowInDataFrame(raw_row, df):
9     df_concat = []
10    for ignore, msg in enumerate(raw_row):
11        mtype = type(msg).__name__
12
13        if mtype in fct.functions:
14            getDataFunction = fct.functions[mtype]
15
16            rows = getDataFunction(msg)
17            cols = fct.columns[mtype]
18
19            df_concat.append(pd.DataFrame([rows], columns=cols))
20        else:
21            continue
22
23    try:
24        df_row = pd.concat(df_concat, axis=1)
25        valid = df_row.get('Valid?')[0]
26
27        # Concat row in DataFrame if it is valid
28        if valid:
29            df = pd.concat([df, df_row])
30
31    return df
32 except:
33     return df
34
35 help_str = ("Usage: gps2csv.py -i <input> -o <output>\n"
36             "Convert .nmea to .csv\n"
37             "\n"
38             "  -h, --help   Print this help message and exit\n"

```

```

39         " -i, --input  Read data from <input> file\n"
40         " -o, --output Write processed data in <output> file\n"
41     )
42
43 def getCommandLineArgs(argv):
44     if len(argv) == 0:
45         return None
46
47     cmdline = {}
48
49     try:
50         opts, args = getopt.getopt(argv, "hi:o:", ["help", "input=", "output="])
51     except getopt.GetoptError:
52         print(help_str)
53         sys.exit(2)
54
55     for opt, arg in opts:
56         if opt in ('-h', '--help'):
57             print(help_str)
58             sys.exit()
59         elif opt in ('-i', '--input'):
60             cmdline['i'] = arg
61         elif opt in ('-o', '--output'):
62             cmdline['o'] = arg
63
64     if not ('i' in cmdline and 'o' in cmdline):
65         print(help_str)
66         sys.exit(2)
67
68     return cmdline['i'], cmdline['o']
69
70 def main(filename='test.nmea', output='out.csv'):
71     e = []
72     row = []
73     first = None
74     data = pd.DataFrame()
75
76     total_lines = 0
77     with open(filename, 'r') as f:
78         for line in f:
79             total_lines += 1
80
81     print('Input file: "{}" ({} lines)'.format(filename, total_lines))
82     print('Output file: "{}"'.format(output))

```

```

83
84     print(',')
85
86     with open(filename, 'r') as f:
87         i = 0
88         for line in f:
89             i += 1
90             print('Line {}/{}'.format(i, total_lines), end='\r')
91             try:
92                 msg = pynmea2.parse(line)
93                 if type(msg).__name__ != 'TXT':
94                     if first == None:
95                         first = msg
96                     elif type(msg) == type(first):
97                         data = addRowInDataFrame(row, data)
98                         row = []
99
100                    row.append(msg)
101            except:
102                e.append('{}: {}'.format(i, sys.exc_info()[1]))
103
104        if len(row) > 1:
105            data = addRowInDataFrame(row, data)
106
107        print(',')
108        print(',')
109        print('Saving into "{}".format(output))')
110        print(',')
111
112        ## Variation of Time - 'S'
113        time = pd.to_timedelta(data['Time'])
114        S_value = time.diff().fillna(pd.Timedelta(0.0))
115        data['S'] = S_value.apply(lambda x: (x.to_pytimedelta().seconds))
116
117        ## Cumulative sum of Time - 'TIME_ACUM'
118        data['TIME_ACUM'] = S_value.cumsum().apply(
119            lambda x: (x.to_pytimedelta().seconds)
120        )
121
122        ## Filename
123        data['GPS_FILE'] = filename[:-5]
124
125        data.to_csv(output, index=False)
126
127        if len(e) == 0:

```

```

128     print('Ok')
129 else:
130     print('There were some errors')
131     for err in e:
132         print(err)
133
134 if __name__ == '__main__':
135     cmdline = getCommandLineArgs(sys.argv[1:])
136     if cmdline:
137         main(cmdline[0], cmdline[1])
138     else:
139         main()

```

### functions.py:

```

1 import datetime
2
3 columns = {
4     'RMC': ['Date', 'Time', 'X', 'Y',
5              'Latitude', 'Longitude'],
6     'VTG': ['Heading', 'Knots', 'M/H', 'KM/H'],
7     'GGA': ['Altitude (m)', 'Altitude (ft)'],
8     'GLL': ['Valid?']
9 }
10
11 def RMCDData(msg):
12     '',
13     Date, Time, X, Y, Latitude, Longitude
14     Brasilia Time (BRT), UTC -3
15     '',
16     try:
17         UTC = datetime.datetime.combine(msg.datestamp, msg.timestamp)
18         BR = UTC - datetime.timedelta(hours=3)
19
20         return [BR.date().isoformat(), BR.time().isoformat(), msg.
21                 longitude,
22                         msg.latitude, msg.latitude, msg.longitude]
23     except:
24         return ['', '', msg.longitude, msg.latitude, msg.latitude, msg.
25                 longitude]
26
27 def VTGData(msg):
28     '',
29     Heading, Knots, M/H, KM/H
30     '',
31     try:

```

```

30         knot = float(msg.spd_over_grnd_kts)
31         knot2mh = 1.15078
32         miles = knot2mh * knot
33         heading = msg.true_track
34         kmhr = msg.spd_over_grnd_kmph
35
36         return [heading, knot, miles, kmhr]
37     except:
38         return ['', '', '', '']
39
40 def GGAData(msg):
41     '',
42     Altitude (m), Altitude (ft)
43     '',
44     meter2feet = 3.28084
45     try:
46         feet = meter2feet * msg.altitude
47         return [msg.altitude, feet]
48     except:
49         return ['', '']
50
51 def GLLData(msg):
52     '',
53     Valid?
54     '',
55     status = False
56     if msg.status == 'A':
57         status = True
58
59     return [status]
60
61 functions = {
62     'RMC': RMCData,
63     'VTG': VTGData,
64     'GGA': GGAData,
65     'GLL': GLLData,
66 }

```

Code written by Borguezani et al. (2020).

gps-read.py:

```

1 import serial
2 import string
3 import math
4 import cv2

```

```

5
6 import os
7 import shutil
8
9 # Hora e Data
10 import datetime
11 from time import time
12 from time import sleep
13
14 if not os.path.exists("output-gps"): os.makedirs('~/output-gps')
15
16 #dmesg |grep tty
17 ser = serial.Serial('/dev/ttyACM0',baudrate=115200, timeout=1)
18
19 filename1 = "GPS"+str(datetime.datetime.now().strftime('%Y-%m-%d_%H-%M-%S'))+".nmea"
20 filename2 = "GPS"+str(datetime.datetime.now().strftime('%Y-%m-%d_%H-%M-%S'))+".txt"
21
22 raw_data = open(filename1, 'w')
23 timeline = open(filename2, 'w')
24
25 #raw_data = open("GPS"+str(time())+".nmea", 'w')
26 #timeline = open("GPS"+str(time())+".txt", 'w')
27
28
29 while True:
30     line = ser.readline()
31     current_time = datetime.datetime.now().time()
32     print(current_time, ', ', line) # Write on console
33     linetime = '{:%H:%M:%S:%f}'.format(current_time) + ',' + line.decode('ascii')
34     timeline.write(linetime+'\n')
35     raw_data.write(line) # Write to the output log
            file
36
37     # frame=cv2.imread('HereGPS.jpg')
38     # cv2.imshow('Input', frame)
39     # c = cv2.waitKey(1)
40     # if c == 27:
41     #     print ('Finishing...')
42     #     break
43
44 ser.close
45 f.close

```

```

46 #if not os.path.exists("output-gps"): os.makedirs('./output-gps')
47
48 #shutil.move(filename1 , '/home/onsv/output-gps')
49 #shutil.move(filename2 , '/home/onsv/output-gps')

```

### video.py:

```

1 import numpy as np
2 import cv2
3
4 import os
5 import shutil
6
7 #Hora e Data
8 import datetime
9
10 #from datetime import date
11 from time import time
12 from time import sleep
13
14 filename1 = format(datetime.datetime.now()) +(1).avi'
15
16 current_time = datetime.datetime.now().strftime('%Y-%m-%d_%H-%M-%S')
17 #current_time = date.today()
18 filename1 = format(datetime.datetime.now()) +(1).avi'
19 filename2 = format(datetime.datetime.now()) +(2).avi'
20 filename3 = format(datetime.datetime.now()) +(3).avi'
21 frames_per_seconds = 20.0
22 # numero da camera
23
24 cam1 = cv2.VideoCapture(1)
25 cam2 = cv2.VideoCapture(0)
26 cam3 = cv2.VideoCapture(3)
27
28 writer1 = cv2.VideoWriter_fourcc(*'XVID')
29 writer2 = cv2.VideoWriter_fourcc(*'XVID')
30 writer3 = cv2.VideoWriter_fourcc(*'XVID')
31
32 out1 = cv2.VideoWriter(filename1, writer1, frames_per_seconds, (640,480)
33 )
33 out2 = cv2.VideoWriter(filename2, writer2, frames_per_seconds, (640,480)
34 )
34 out3 = cv2.VideoWriter(filename3, writer3, frames_per_seconds, (640,480)
35 )
35 while (True):

```

```

37     ret, frame1 = cam1.read()
38     ret, frame2 = cam2.read()
39     ret, frame3 = cam3.read()
40     current_time = datetime.datetime.now()
41     frame1 = cv2.putText(frame1, format(current_time), (10,50), 0, 1,
42                           (0, 255, 255), 2, cv2.LINE_AA)
43     frame2 = cv2.putText(frame2, format(current_time), (10,50), 0, 1,
44                           (0, 255, 255), 2, cv2.LINE_AA)
45     frame3 = cv2.putText(frame3, format(current_time), (10,50), 0, 1,
46                           (0, 255, 255), 2, cv2.LINE_AA)
47     if ret==True:
48         #flip inverte a imagem de cabeca para baixo
49         # frame = cv2.flip(frame, 0)
50         out1.write(frame1)
51         out2.write(frame2)
52         out3.write(frame3)
53         cv2.imshow('cam1', frame1)
54         # cv2.moveWindow('cam1', 0,0)
55         cv2.imshow('cam2', frame2)
56         # cv2.moveWindow('cam2', 630,0)
57         cv2.imshow('cam3', frame3)
58         # cv2.resizeWindow('cam3', 100,100)
59         # cv2.moveWindow('cam3', 630,630)
60     else:
61         break
62     #aperta q para sair:::
63     if (cv2.waitKey(1) & 0xFF == ord('q')):
64         break
65
66 # Move os arquivos para o diretorio
67 if not os.path.exists("output-video"): os.makedirs('./output-video')
68
69 shutil.move(filename1,'/home/onsv/output-video')
70 shutil.move(filename2,'/home/onsv/output-video')
71 shutil.move(filename3,'/home/onsv/output-video')
72
73 cam1.release()
74 cam2.release()
75 cam3.release()
76 out1.release()
77 out2.release()
78 out3.release()
79 cv2.destroyAllWindows()

```