Point Patterns Analysis

Angélica Portocarrero and Atenea Rojas

September 3, 2024

1 Study Context

There are different studies on traffic accidents, starting from some already established and complementing them with other methods that are usually used for the analysis of pollution or diseases, among others. The analysis by Spatial Punctual Patterns is implemented, which is composed by several tests that give as a result the distribution of the points in the space, being understood by this that they can be: Random, Uniform or Aggregate. This means that depending on the results and the characteristics of the points where accidents occurred, the data may present a pattern, may be totally dispersed or may have some intensity. The study by Point Spatial Patterns is complemented with basic statistics and the purification of the database used, which in this case is obtained through the Secretary of Traffic and Transportation, the entity in charge of filling out the forms for each of the accidents that occur in the city.

2 Data Dictionary

The data presented in this analysis is the result of a comprehensive cleaning process. Initially, two datasets from the years 2009 and 2010 were provided, each containing various discrepancies and incomplete data. These datasets were merged into a single dataset, with an additional column indicating the year, to facilitate a uniform analysis. After merging, the data was thoroughly cleaned to remove any inconsistencies and to ensure that all variables were properly formatted and accurately represented. The following table describes the variables in the final cleaned dataset used for analysis.

Variable	**Description**
mes_fallecimiento	Month in which the fatality occurred.
mes_accidente	Month in which the accident occurred.
comuna	The commune or district where the accident occurred.
sexo	Gender of the deceased (M for Male, F for Female).
edad	Age of the deceased at the time of the accident.
edad_agrupada	Age group category of the deceased (e.g., 20-29, 30-39, etc.).
fecha_accidente	Date when the accident occurred.
fecha_fallecimiento	Date when the death occurred.
hora_fallecimiento	Time of the day when the death occurred.
hora_accidente	Time of the day when the accident occurred.
dia_semana_fallecimiento	Day of the week when the death occurred.
dia_semana_accidente	Day of the week when the accident occurred.
condicion	Condition of the deceased (e.g., pedestrian, driver, passenger, etc.).
vehiculos	Type of vehicles involved in the accident.
coordenada_x_metros	X-coordinate of the accident location in meters.
coordenada_y_metros	Y-coordinate of the accident location in meters.
coordenada_x_km	X-coordinate of the accident location in kilometers.
coordenada_y_km	Y-coordinate of the accident location in kilometers.
ano	Year in which the accident and fatality occurred (2009 or 2010).
periodo_dia_accidente	Part of the day when the accident occurred (e.g., morning, afternoon, night).

Table 1: Descriptions of variables used in the final cleaned dataset.

It is important to note that the variables are in Spanish because the dataset focuses on Cali, Colombia, a city where Spanish is the primary language. In an effort to preserve the authenticity and context of the data, it was decided to manage the dataset in its original language.

3 Analysis Objective

The objective of this study is to perform a spatial data analysis to characterize and model traffic homicides in Cali. For this, a working group is needed to develop the following activities:

1) According to the information provided, perform an exploratory data analysis that allows characterizing homicides by traffic accidents in Cali.

2) Perform an analysis of punctual patterns for each year of homicides due to traffic accidents in Cali; by sex, age groups and type of vehicle.

3) Fit a model that allows explaining the spatial behavior of homicides by traffic accidents in Cali.

4 Exploratory Data Analysis

Before starting any kind of analysis, it is important to conduct an initial process to explore the data comprehensively. This exploratory data analysis (EDA) aims to uncover insights and identify patterns and relationships within the data. The information gathered during this phase is crucial for guiding the subsequent data cleaning and preparation steps, ensuring that the final analysis is based on accurate and well-understood data.

The first operations performed with the data were directed towards discovering the shape of the dataset. In this case, as shown in **Table 2** it was determined that the dataset consists of 599 rows and 20 columns. Additionally, there are no duplicate or null values, which is a result of the prior data cleaning process as mentioned earlier.

Rows	Columns	Null values	Duplicate values
599	20	0	0

Table 2: Dataset shape

Next, **Table 3** presents the descriptive statistics for the most relevant variables that will be used in the subsequent analysis. These statistics provide a crucial overview of the main characteristics of the data, allowing for the identification of trends, patterns, and potential anomalies. Since the dataset has been previously cleaned, ensuring no null or duplicate values, the reliability of the results in the following stages of analysis is well-supported.

It should be noted that, the table specifically focuses on data related to "accidents" rather than "fatalities," as the primary aim is to analyze the flow of accidents. The key variables examined are those directly related to the occurrences of accidents, providing insights into the patterns and characteristics of these incidents. Notably, the "hora_accidente" variable was not included, which contains 24 distinct categories, as this granularity could complicate the analysis. Instead, the "periodo_dia_accidente" variable was chosen, which offers a more manageable and meaningful grouping of the data into relevant time periods. This approach allows for a clearer understanding of when accidents are most likely to occur during the day, facilitating more actionable insights.

	2009 (N = 322)	2010 (N = 277)
Condición		
Ciclista	44 (14%)	42 (15%)
Moto	143 (44%)	102 (37%)
Peatón	112 (35%)	114 (41%)
Vehículo	23 (7.1%)	19 (6.9%)
Día de la semana del accidente		
Lunes	40 (12%)	34 (12%)
Martes	31 (9.6%)	34 (12%)
Miércoles	38 (12%)	27 (9.7%)
Jueves	50 (16%)	30 (11%)
Viernes	40 (12%)	42 (15%)
Sábado	58 (18%)	56 (20%)
Domingo	65 (20%)	54 (19%)
Sexo		
Femenino	64 (20%)	49 (18%)
Masculino	258 (80%)	228 (82%)
Edad agrupada		
0-9	4 (1.2%)	1 (0.4%)
10-19	22 (6.8%)	12 (4.3%)
20-29	72 (22%)	67 (24%)
30-39	52 (16%)	34 (12%)
40-49	49 (15%)	47 (17%)
50-59	42 (13%)	31 (11%)
60-69	32 (9.9%)	28 (10%)
70-79	33 (10%)	32 (12%)
80Y+	16 (5.0%)	25 (9.0%)
Periodo del día del accidente		
Madrugada	61 (19%)	55 (20%)
Mañana	67 (21%)	69 (25%)
Tarde	109 (34%)	71 (26%)
Noche	85 (26%)	82 (30%)
Mes del accidente		
Enero	23 (7.1%)	25 (9.0%)
Febrero	24 (7.5%)	26 (9.4%)
Marzo	24 (7.5%)	19 (6.9%)
Abril	28 (8.7%)	28 (10%)
Mayo	30 (9.3%)	24 (8.7%)
Junio	33 (10%)	29 (10%)
Julio	25 (7.8%)	17 (6.1%)
Agosto	27 (8.4%)	23 (8.3%)
Septiembre	24 (7.5%)	25 (9.0%)
Octubre	21 (6.5%)	14 (5.1%)
Noviembre	31 (9.6%)	20 (7.2%)
Diciembre	32 (9.9%)	27 (9.7%)

Table 3: Estadísticas descriptivas de las variables de accidentes para 2009 y 2010

4.1 Condición (Condition)

For **Ciclista**, the proportion of cyclists involved in accidents remains relatively stable between the two years, with 14% in 2009 and 15% in 2010. This consistency suggests that the risk factors affecting cyclists may not have changed significantly year over year. However, given that cyclists represent a notable portion of the total accidents, interventions targeting cyclist safety, such as better infrastructure or awareness campaigns, could be beneficial.

For **Moto**, the percentage of motorcyclists involved in accidents decreased from 44% in 2009 to 37% in 2010. This decrease might indicate an improvement in road safety for motorcyclists or possibly changes in the number of motorcyclists on the road. Despite this decrease, motorcyclists still represent the largest group affected, highlighting the need for continued focus on motorcycle safety measures.

For **Peatón**, pedestrian accidents slightly increased from 35% in 2009 to 41% in 2010, which could suggest growing risks for pedestrians. This increase underscores the importance of pedestrian-focused safety measures, such as improved crosswalks, traffic signals, and public awareness campaigns about pedestrian safety.

For **Vehículo**, the proportion of vehicle-related accidents remains relatively low and consistent, with 7.1% in 2009 and 6.9% in 2010. Although these figures are smaller compared to other groups, the consistency suggests a stable risk profile for vehicle accidents, potentially due to effective traffic regulations or the nature of vehicle accidents.

4.2 Día de la Semana del Accidente (Day of the Week of Accident)

Domingo (Sunday) consistently sees the highest proportion of accidents, with 20% in 2009 and 19% in 2010. This could be linked to factors such as increased recreational activities or less traffic enforcement on weekends. Targeted safety campaigns or increased law enforcement on Sundays could help reduce accidents.

Jueves y Sábado (Thursday and Saturday) also see a high number of accidents, maintaining similar proportions between the two years. This might reflect patterns in weekday activities or traffic flows. Understanding these patterns could help in developing specific strategies to reduce accidents on these days.

The distribution across other days of the week remains relatively even, suggesting no specific day stands out as particularly high-risk, apart from the previously mentioned days.

4.3 Sexo (Gender)

For **Femenino (Female)**, the percentage of females involved in accidents decreased slightly from 20% in 2009 to 18% in 2010. This minor change could be influenced by various factors such as shifts in traffic patterns or vehicle usage among women.

For **Masculino (Male)**, males consistently represent a significant majority of accident victims, at 80% in 2009 and 82% in 2010. This might reflect higher exposure to road risks, possibly due to higher driving frequencies or involvement in riskier behaviors. Tailored road safety campaigns focusing on male drivers could help in reducing these figures.

4.4 Edad Agrupada (Age Group)

The **20-29** age group consistently has the highest proportion of accidents in both years (22% in 2009 and 24% in 2010). This trend could be linked to higher levels of risky driving behaviors among younger drivers or increased exposure due to higher mobility. Interventions targeting young drivers, such as educational programs or stricter enforcement of traffic laws, could help reduce accidents in this age group.

40-49 and **50-59** also show a relatively high percentage of involvement in accidents, which might indicate that middle-aged individuals are at a consistent risk. This could be due to factors such as frequent commuting or distractions while driving.

The slight increase in accident involvement in older age groups (**60-69** and **70-79**) could reflect issues related to declining physical or cognitive abilities, suggesting a need for targeted safety measures or vehicle design adaptations to support older drivers.

4.5 Periodo del Día del Accidente (Time of Day of Accident)

The **Tarde (Afternoon)** is consistently the period with the highest proportion of accidents, with 34% in 2009 and 26% in 2010. This could be related to higher traffic volumes during these hours, as people return from work or school. Traffic management strategies such as staggered work hours or enhanced traffic control during these times could help reduce the number of accidents.

Noche (Night) accidents also represent a significant proportion, particularly in 2010 (30%). This increase suggests a need for better lighting, enhanced visibility for vehicles, and possibly more traffic law enforcement during nighttime hours.

The **Madrugada y Mañana (Early Morning and Morning)** periods show lower but consistent accident rates, highlighting the need for continued vigilance and safety measures during these times.

4.6 Mes del Accidente (Month of Accident)

Abril (April) and **Junio (June)** show the highest number of accidents in both years, with slight variations. This pattern could be influenced by specific factors in these months, such as local events, festivals, or changes in traffic dynamics that increase road usage.

The distribution of accidents across other months is relatively even, suggesting no strong seasonal pattern outside of the mentioned peaks. This even distribution indicates that road safety campaigns should be consistently maintained throughout the year to address the ongoing risk on the roads.

4.7 Visualizations

4.7.1 Age Distribution

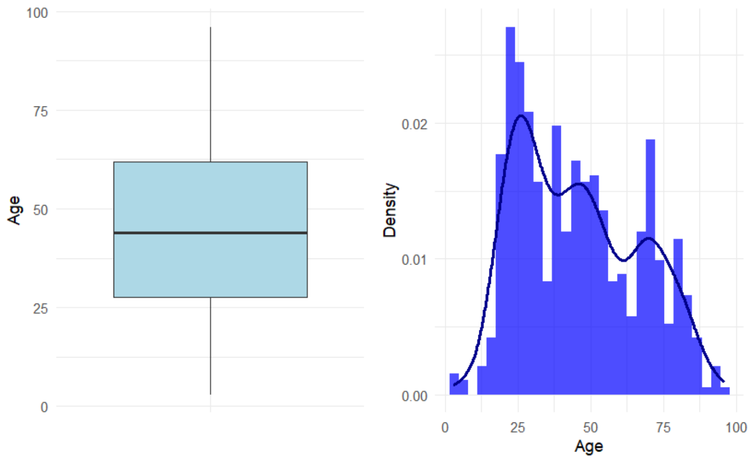

Figure 1: Age Distribution

The visualization in **Figure 1** provides insights into the age distribution of the individuals involved in accidents. The boxplot on the left shows that the median age is slightly above 50 years, with the interquartile range (IQR) spanning from approximately 30 to 65 years. This indicates that half of the accidents involve individuals within this age range. The whiskers extend further, capturing a broader range of ages, although there are no significant outliers.

The density plot on the right complements this by illustrating a multimodal distribution with peaks around the 20-30, 40-50, and 60-70 age ranges. This suggests that different age groups are involved in accidents, possibly indicating distinct behaviors or circumstances leading to accidents in these age brackets. As noted in the summary statistics.

It is possible to say that the data indicates that while young adults and middle-aged individuals are frequently involved in accidents, there is also a notable representation of older individuals. This calls for diverse strategies targeting these specific age groups to improve road safety across all demographics.

4.7.2 Distribution of accidents by comuna with year comparison

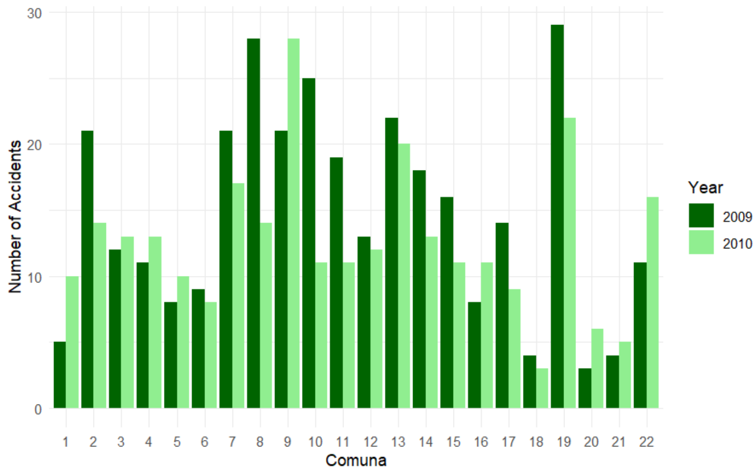

Figure 2: Distribution of accidents by comuna with year comparison

The visualization in **Figure 2** reveals significant variation in the distribution of traffic accidents across different comunas in Cali, Colombia, for the years 2009 and 2010. Notably, Comuna 10, which is a central area known for its commercial and residential zones, shows one of the highest frequencies of accidents. This is likely due to the high density of traffic and pedestrian activity in this area, given its proximity to major commercial centers and heavily trafficked roads.

Comuna 9, another area with a mix of residential, commercial, and educational institutions, also exhibits a high number of accidents. This can be attributed to the bustling activity in this comuna, where the convergence of different types of traffic, including public transportation and private vehicles, increases the likelihood of accidents. The presence of schools and universities could also contribute to the elevated risk, as these institutions attract significant foot traffic and young drivers.

Comuna 19, characterized by its residential neighborhoods and important commercial areas, particularly in the south of Cali, shows a notable increase in accidents in 2010 compared to 2009. This may reflect growing urbanization and increased vehicle use in this area, which traditionally has been less congested than the central comunas. The rise in accidents here underscores the need for enhanced road safety measures as the area continues to develop.

The data indicates that tailored interventions focusing on the specific needs and characteristics of each comuna could be crucial in effectively reducing traffic accidents. For instance, in more commercially active areas like Comunas 9 and 10, strategies could focus on improving pedestrian safety and managing heavy traffic flows, while in residential areas like Comuna 19, efforts might target speed control and public awareness campaigns. This nuanced approach is vital for addressing the unique challenges of each area in Cali.

4.7.3 Distribution of accidents by month with year comparison

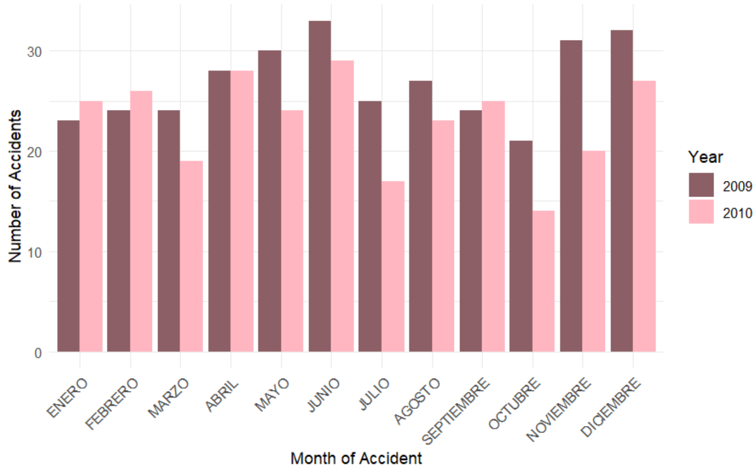

Figure 3: Distribution of accidents by month with year comparison

The visualization in **Figure 3** illustrates the distribution of traffic accidents by month for the years 2009 and 2010. Junio (June) consistently exhibits one of the highest numbers of accidents in both years, indicating a potential pattern that might be linked to increased road usage during this period, possibly due to mid-year activities or specific weather conditions in Cali, Colombia.

Other months like Abril (April) and Diciembre (December) also show relatively high accident rates, particularly in 2009, where December marks a significant rise. This could be associated with holiday-related travel and festivities, leading to increased road activity. It's important to note that while Mayo (May) had a high number of accidents in 2009, this trend did not continue in 2010, where it saw a relative decline. This variation highlights the need for continuous monitoring and adaptive road safety strategies that respond to dynamic patterns in accident occurrences throughout the year.

4.7.4 Distribution of accidents by gender with year comparison

Figure 4: Distribution of accidents by gender with year comparison

The visualization in **Figure 4** clearly shows a significant gender disparity in the distribution of traffic accidents for both 2009 and 2010. In both years, males are overwhelmingly more involved in accidents than females. In 2009, there were 258 accidents involving males compared to 64 involving females, and in 2010, the numbers were 228 for males and 49 for females. This consistent trend suggests that males are at a higher risk of being involved in traffic accidents in Cali, Colombia, possibly due to differences in driving behavior, exposure, or other socio-cultural factors. These insights emphasize the importance of developing gender-specific road safety interventions, particularly targeting the behaviors and conditions that contribute to the higher incidence of accidents among males.

4.7.5 Distribution of accidents by age group with year comparison

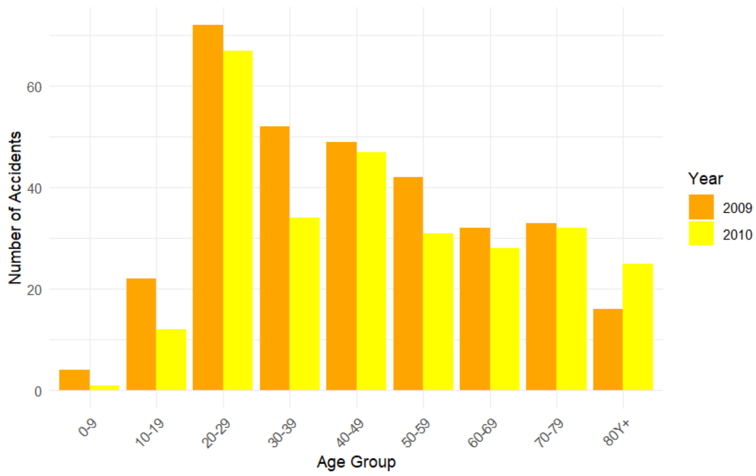

Figure 5: Distribution of accidents by age group with year comparison

The visualization in **Figure 5** illustrates the distribution of traffic accidents across various age groups for the years 2009 and 2010. The 20-29 age group consistently shows the highest number of accidents in both years, indicating that young adults are particularly vulnerable to traffic incidents in Cali, Colombia. Interestingly, while the 20-29 age group remains stable, the 30-39 age group shows a noticeable decrease in accidents from 2009 to 2010. This decline could suggest improved road safety measures or changes in behavior among this demographic. However, the persistent risk for the 20-29 age group may be linked to specific factors such as the type of transportation commonly used by this demographic, which will be further explored in subsequent analyses (refer to **Figure 12**). Additionally, the data reveals lower accident frequencies for the youngest (0-9 and 10-19) and oldest (80+) age groups, reflecting their reduced exposure to road traffic. These findings emphasize the importance of tailored road safety strategies, particularly focusing on young adults, to effectively address the most at-risk populations.

4.7.6 Distribution of accidents by day of the week with year comparison

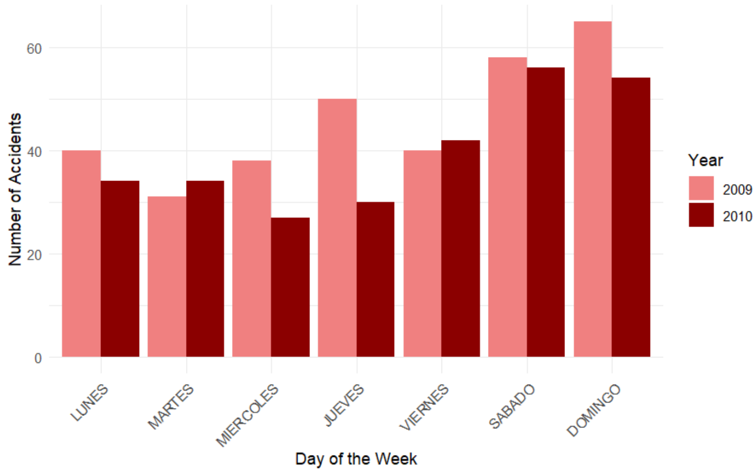

Figure 6: Distribution of accidents by day of the week with year comparison

The visualization in **Figure 6** presents the distribution of traffic accidents across the days of the week for 2009 and 2010. The data reveals that Saturdays and Sundays are the days with the highest number of accidents in both years. On the other hand, weekdays like Tuesday and Wednesday have the lowest accident frequencies, indicating that routine workdays might involve less risky travel behavior or lower traffic volumes related to social activities. The consistency in the pattern, with weekends still showing the highest numbers, underscores the need for continued focus on road safety campaigns or law enforcement during weekends to further reduce accidents.

4.7.7 Distribution of accidents by hour of the day with year comparison

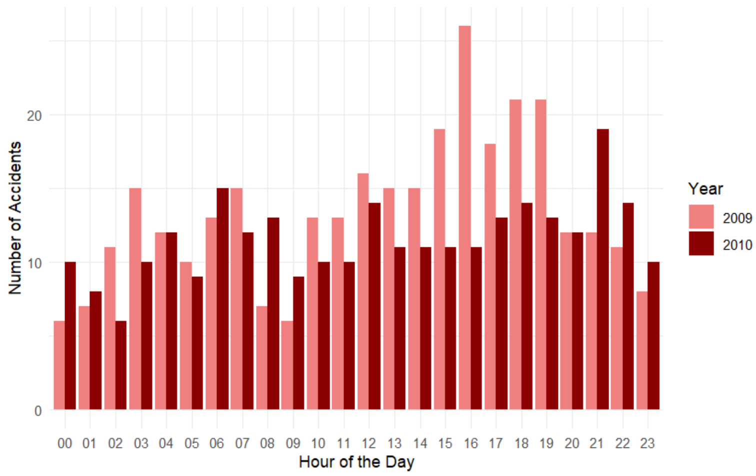

Figure 7: Distribution of accidents by hour of the day with year comparison

The visualization in **Figure 7** presents the distribution of traffic accidents throughout the hours of the day for the years 2009 and 2010. The data reveals a clear pattern where the frequency of accidents increases significantly during the late afternoon, particularly around 16:00 (4 PM). This peak could be attributed to the combination of high traffic volumes during the evening rush hour and possibly fatigued drivers after a full day of work. A secondary peak occurs in the early morning hours, particularly between 6:00 and 8:00, likely coinciding with the morning rush hour when traffic volumes are again high.

Comparing the two years, 2010 shows a slight increase in accidents during certain afternoon and evening hours, suggesting that the risk factors contributing to these peak times may have persisted or even intensified. Additionally, the distribution of accidents during the late-night hours (around 22:00) in 2010 shows a slight uptick compared to 2009, which could point to increased night-time activities or higher alcohol consumption during this time frame.

Interestingly, the early hours of the morning (from midnight to around 4:00) show a relatively lower number of accidents, which gradually rise as the morning progresses, highlighting the transition from quieter roads to busier traffic periods.

The following graph will provide a more aggregated view of this distribution by grouping the hours into broader periods of the day (morning, afternoon, evening, and night), making it easier to visualize and interpret the key patterns and risk periods for traffic accidents. The next analysis will focus on these periods, offering further insights into when the roads are most dangerous.

4.7.8 Distribution of accidents by period of the day with year comparison

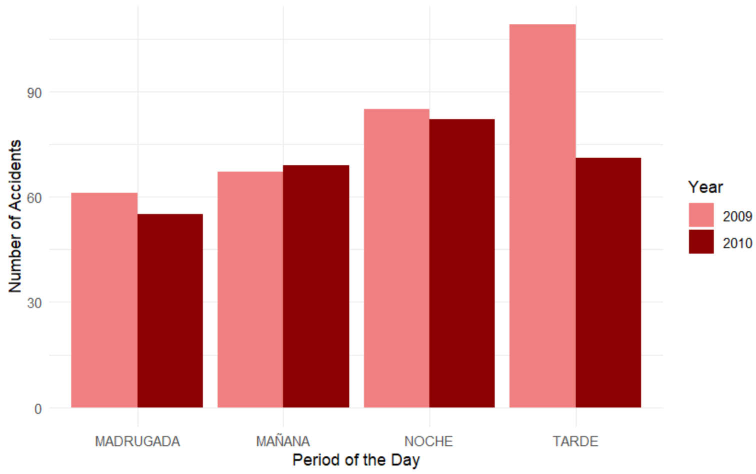

Figure 8: Distribution of accidents by period of the day with year comparison

Figure 8, as mentioned before, summarizes the distribution of traffic accidents across different periods of the day for the years 2009 and 2010. The data reveals that the "Tarde" (afternoon) period experiences the highest number of accidents in 2009, but this trend does not continue into 2010, where the number of accidents during the afternoon decreases significantly. This suggests that while the afternoon was previously the most dangerous period, improvements or changes in traffic patterns might have reduced the risk in 2010.

Accidents during the "Noche" (night) period remain relatively stable between the two years, indicating that risk factors during nighttime, such as reduced visibility and possibly impaired driving, are consistent. The "Mañana" (morning) period also shows consistency, reflecting a steady risk during morning commutes.

Interestingly, the "Madrugada" (early morning) period actually shows a slight decrease in accidents in 2010 compared to 2009, indicating a possible reduction in risk during these early hours. This decrease could be attributed to fewer late-night activities or improved measures to address fatigue-related incidents during this period. A similar trend is observed in the data for Saturday and Sunday (**Figure 6**), where the number of accidents decreased in 2010 compared to 2009. This reduction on weekends could also be linked to a decline in nocturnal activities or more effective enforcement of road safety measures during nights, particularly when the risk of accidents due to factors like alcohol consumption or fatigue tends to be higher.

4.7.9 Distribution of accidents by condition of the person with year comparison

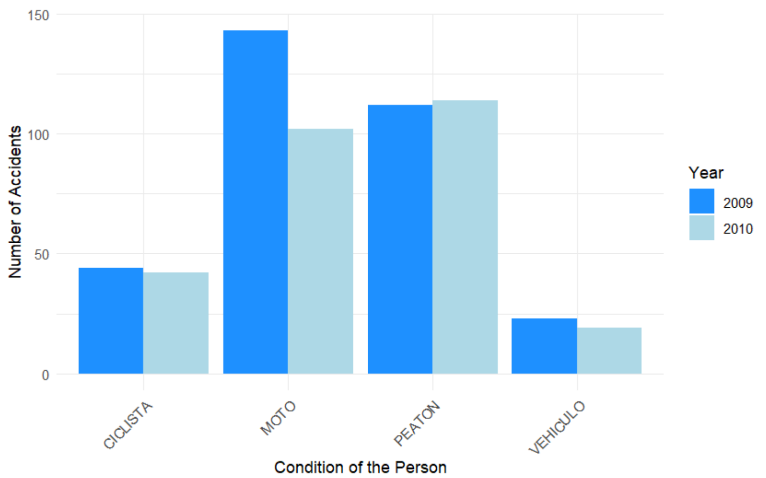

Figure 9: Distribution of accidents by condition of the person with year comparison

The distribution of accidents by the condition of the person, as illustrated in **Figure 9**, reveals a shift in the pattern of accident involvement between 2009 and 2010. While motorcyclists ("MOTO") were the most frequently involved in accidents in 2009, there was a significant decrease in 2010. In contrast, pedestrians ("PEATON") became the group with the highest number of accidents in 2010, surpassing motorcyclists. The number of accidents involving cyclists ("CICLISTA") and vehicle drivers ("VEHICULO") remained relatively low in both years, with minimal variation. This shift highlights a growing concern for pedestrian safety in 2010, indicating that targeted interventions might be needed to address this rising trend.

4.7.10 Homicides by traffic accidents, broken down by month and gender with year comparison

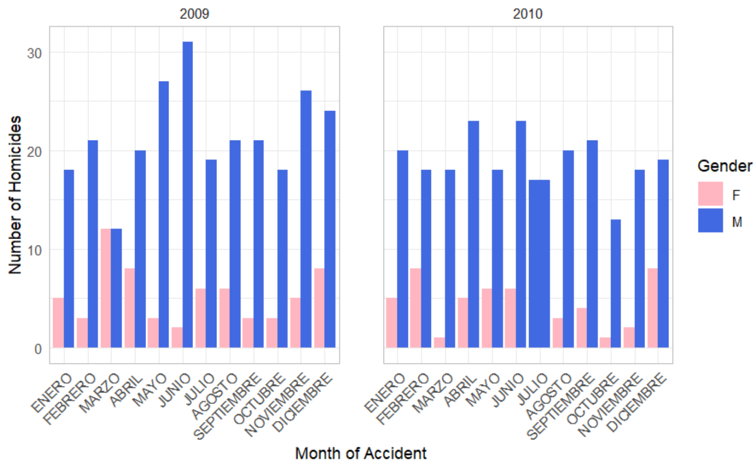

Figure 10: Homicides by traffic accidents, broken down by month and gender with year comparison

Figure 10 provides a detailed comparison of homicides by traffic accidents, categorized by month, gender, and year (2009 and 2010). Across both years, a clear pattern emerges where males ("M") are significantly more represented in traffic-related homicides compared to females ("F"), with male fatalities consistently surpassing female fatalities each month.

In 2009, the months of June, November and May show particularly high numbers of male fatalities, with June reaching nearly 30 incidents, marking it as the most dangerous month for males in that year, followed closely by November and May. Female fatalities, although much lower overall, show a noticeable peak in March, suggesting that this month posed an increased risk for this gender.

In contrast, 2010 exhibits a little different pattern. The distribution of male fatalities becomes more even throughout the year, with April and June standing out as the months with the highest numbers, but without the sharp peaks observed in 2009. For females, fatalities are relatively stable throughout the year, with no single month showing a drastic increase, although February and December have slightly higher numbers compared to other months.

An interesting observation in the 2010 data is that in July, all recorded traffic-related homicides involve males, with no female fatalities reported. This suggests a potential gender-specific risk factor during this month that disproportionately affected men. It could be linked to particular activities or events that predominantly involved males, warranting further investigation.

While males consistently represent a higher proportion of traffic-related homicides throughout both years, the absence of female fatalities in July 2010 highlights the importance of considering gender-specific trends in road safety analysis. Understanding these nuances can guide more targeted and effective interventions.

4.7.11 Homicides by traffic accidents, broken down by condition and gender with year comparison

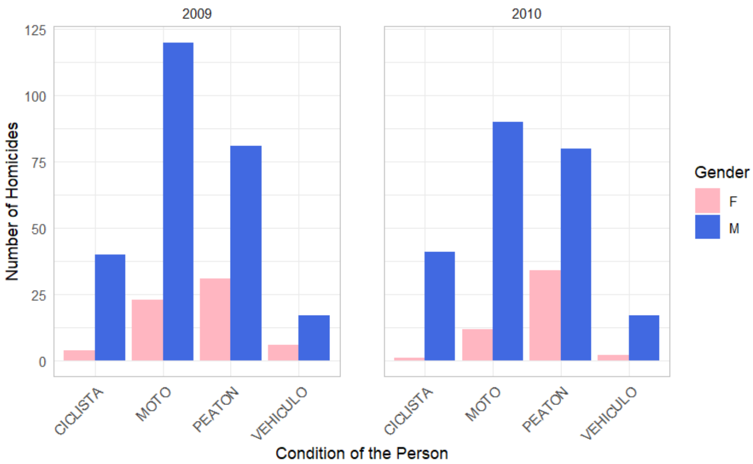

Figure 11: Homicides by traffic accidents, broken down by condition and gender with year comparison

It can be seen in **Figure 11** the distribution of traffic-related homicides by the condition of the person, gender, and year. In both 2009 and 2010, motorcyclists (”MOTO”) remain the leading category for fatalities, particularly among males, despite a decrease in the total number of motorcyclist fatalities in 2010 compared to 2009. Pedestrians (”PEATON”) follow as the second most affected group, with a slight increase in female fatalities in 2010 compared to 2009.

Relating this to the previous insight from the monthly distribution by gender (**Figure 10**), the consistent dominance of male fatalities among motorcyclists aligns with the observed peaks in male fatalities during certain months, such as June and November. These periods could coincide with higher motorcycle usage or riskier conditions for motorcyclists, suggesting a continued vulnerability for this group across different times of the year.

4.7.12 Homicides by traffic accidents, broken down by age group and condition with year comparison

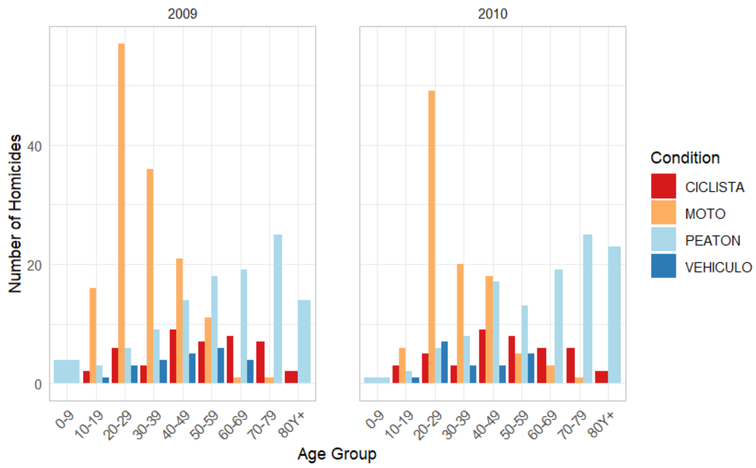

Figure 12: Homicides by traffic accidents, broken down by age group and condition with year comparison

The analysis of homicides by traffic accidents, broken down by age group and condition (**Figure 12**), reveals distinct patterns across different age categories for the years 2009 and 2010.

In 2009, the highest number of fatalities occurred in the 20-29 age group, with motorcyclists being the most affected. This age group accounted for a significant proportion of the fatalities, suggesting a high risk associated with young adults, particularly those riding motorcycles. Pedestrians also saw a notable number of fatalities, particularly in the older age groups (60-69 and 70-79), which may indicate vulnerabilities among the elderly when it comes to road safety.

In 2010, the pattern shifted slightly, but the 20-29 age group still remained the most affected, again with motorcyclists at the forefront. However, there was a reduction in fatalities among motorcyclists compared to 2009, while pedestrian fatalities remained consistent, particularly in the 70-79 and increasing in 80+ age groups. This consistency among pedestrians, especially in the older age groups, could suggest an ongoing vulnerability for these individuals, reinforcing the need for targeted safety interventions for both young motorcyclists and elderly pedestrians.

These findings, when compared to the previous analysis of monthly fatalities by gender, show that the high risk for motorcyclists in the 20-29 age group corresponds with the overall high male fatality rates observed in months like June and November 2009. This correlation suggests that young male motorcyclists are a particularly vulnerable group, and targeted measures during these peak months could be beneficial in reducing fatalities.

4.7.13 Boxplot of age by condition with year comparison

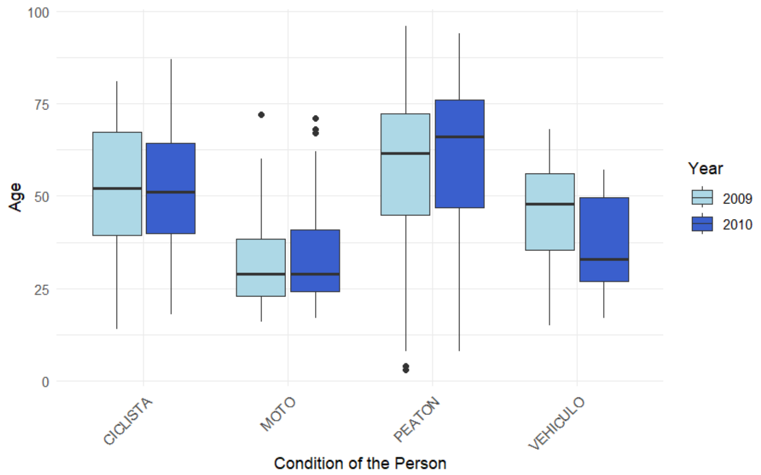

Figure 13: Boxplot of age by condition with year comparison

The boxplot in **Figure 13** provides a comparison of age distributions across different conditions (Cyclist, Motorcyclist, Pedestrian, and Vehicle occupants) for the years 2009 and 2010.

The median age for cyclists remains consistent across both years, hovering around 50 years, with a wide interquartile range indicating a diverse age group involved in these accidents. For motorcyclists, the median age is lower, around 30 years, with little variation between 2009 and 2010. However, there is a noticeable increase in the range of ages involved in accidents in 2010, indicating that not only young adults but also older individuals were increasingly involved in motorcycle-related incidents.

Pedestrians show a higher median age, especially in 2010, where it approaches 60 years, suggesting that older pedestrians are at a significant risk. This aligns with previous observations that older age groups are more vulnerable, particularly in pedestrian-related accidents.

For vehicle occupants, the median age decreases in 2010 compared to 2009, indicating a younger demographic becoming more involved in vehicular accidents. This could suggest a shift in the age groups at risk or perhaps changes in driving behavior or patterns.

The presence of outliers in some categories, particularly among pedestrians and motorcyclists, further emphasizes the variability in age groups involved in these accidents, pointing to the need for diverse safety strategies that account for these differences.

5 Analysis of point patterns by year (gender, age group and type of vehicle)

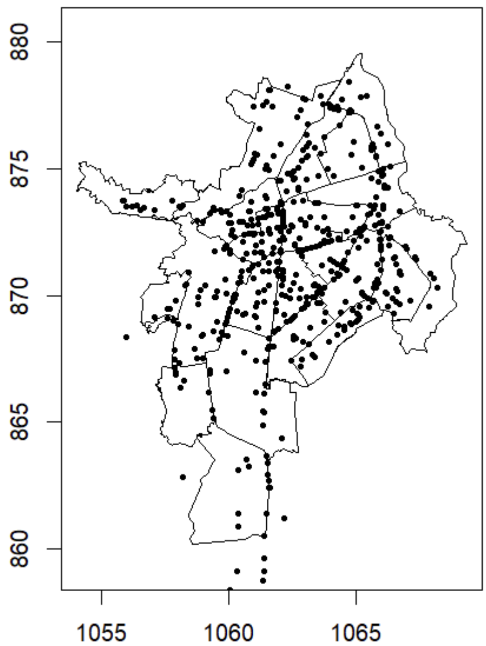

Figure 14: Map of fatal traffic accidents in Cali

For the analysis of point patterns, initially, a map of all traffic accidents was plotted as the start of an exploratory analysis, as shown in **Figure 14**. This figure shows the accidents for both 2009 and 2010 years, it can be seen from this figure that most incidents were grouped around the center of the city. This makes sense as this is the busiest area of the city. Whereas the south of the city, wasn't as developed during those years, thus, having a lower presence of roads and lower traffic resulted in lower accident rate.

The study area was segmented into 4 by 5 quadrats, resulting in 17 total quadrats as this was the split that left the least amount of areas without data points. This can be seen on **Figure 15**

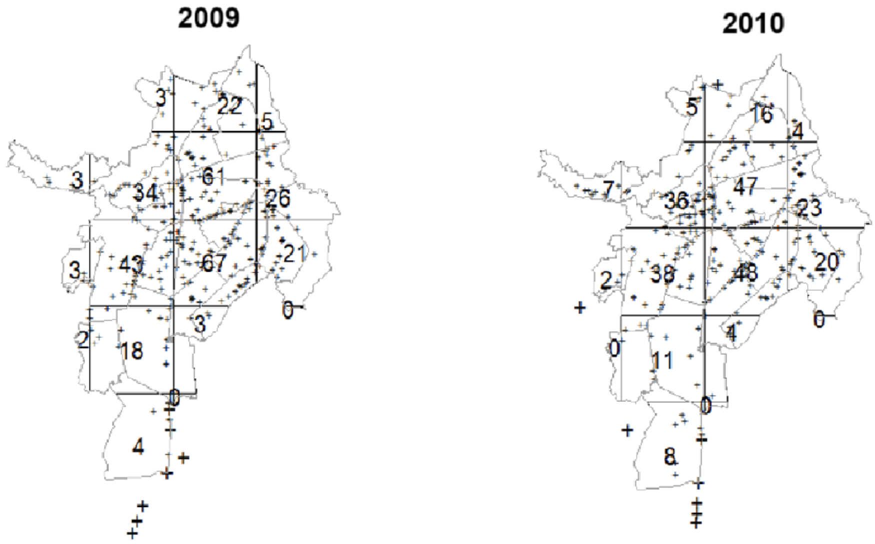

Figure 15: Plot of frequency of fatal traffic accidents per quadrat

After that, a Chi-squared test of CSR was performed on data of both years, using quadrat counts. Its results displayed on **Table 4**, show that With p-values of 2.9055e-13 for 2009 and 4.9393e-07 for 2010, the null hypothesis that the spatial pattern of points follows a completely random distribution is strongly rejected. This indicates that there is very strong evidence that the spatial pattern observed in the data is not random, which indicates that there are is a significant variability in the density across the study area. This means that the fatal traffic accidents are not randomly distributed around the city of Cali and instead, show places where fatal accidents were more present. This is was also supported by the initial visualization of accidents in **Figure 14**.

Year	**2009**	**2010**
p-value	2.9055e-13	4.9393e-07

Table 4: Chi-squared test of CSR for 2009 and 2010

5.1 Gender

The analysis of the Gender variable started by plotting the presence of accidents per year, categorized by gender, with ”M” corresponding to ”Male” and ”F” corresponding to ”Female”. This plot can be seen on **Figure 16**, where it is apparent that the male population seems more vulnerable to homicides by traffic accidents.

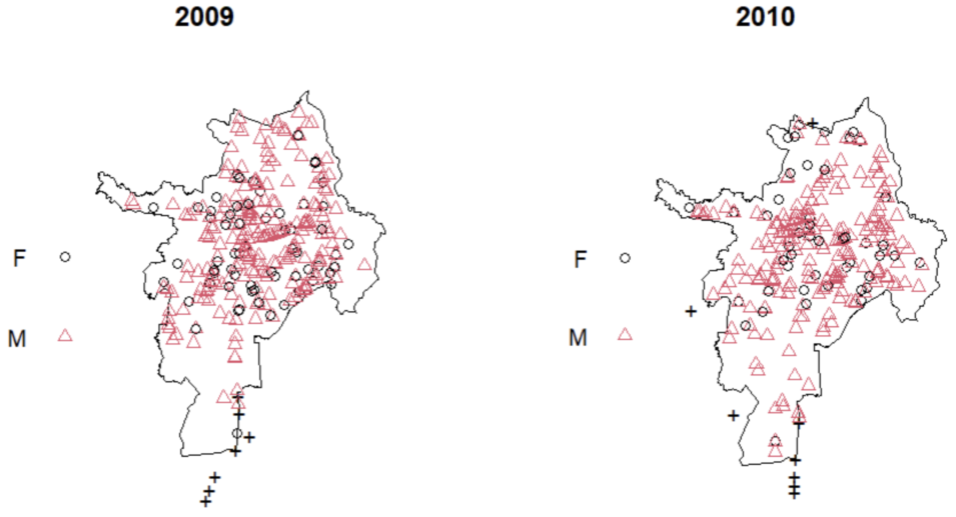

Figure 16: Fatal Traffic Accidents per Gender

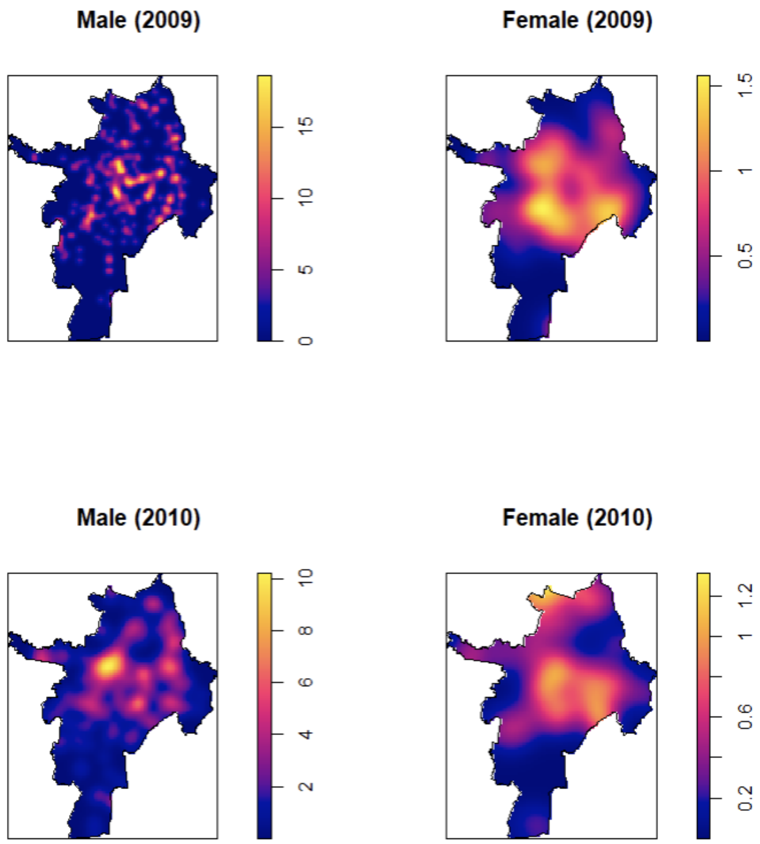

Figure 17: Kernel Density estimation of Homicides by Traffic Accidents per Gender and Year

The Kernel Density Estimation (KDE) method is used as it seems to be an effective tool for analyzing accident patterns in Cali, Colombia. KDE creates a smooth, continuous map of accident density, identifying high-risk "hotspots" without assuming any specific data distribution. This flexibility is especially valuable in complex urban environments where traffic patterns are unpredictable. KDE uses a kernel function to smooth accident locations, providing a clear visual representation of risk areas. Its non-parametric nature ensures that the analysis remains robust even in the absence of prior assumptions about data distribution. This is crucial for maintaining the integrity of the analysis in varied and irregular urban settings.

Figure 17 shows the estimated density of fatal traffic accidents per gender for both 2009 and 2010. It can be seen that overall, there is a higher density around the center of the city, as well as a lower presence of female victims, whereas male victims seem to be strongly aggregated in some spots. This shows that the density is not constant, denoting an inhomogeneous process.

In order to further explore the nature of the point pattern, Ripley's K function was opted for use. This method compares the observed spatial pattern in the sample with what would be expected under spatial randomness, giving an insight of the behavior of the observed pattern.

Ripley's K function is the correct application for stationary and isotropic processes [4], however, as seen on **Figure 17**, is not the current case for the analysis point pattern. Given the non-stationary process of the point pattern, the inhomogeneous K function ($K_{\text{inhom}}(r)$)was used. This function is used when the density of points varies in space, as it adjusts the calculation of $K(r)$ to account for this variation in density, allowing a more accurate assessment of spatial structure.

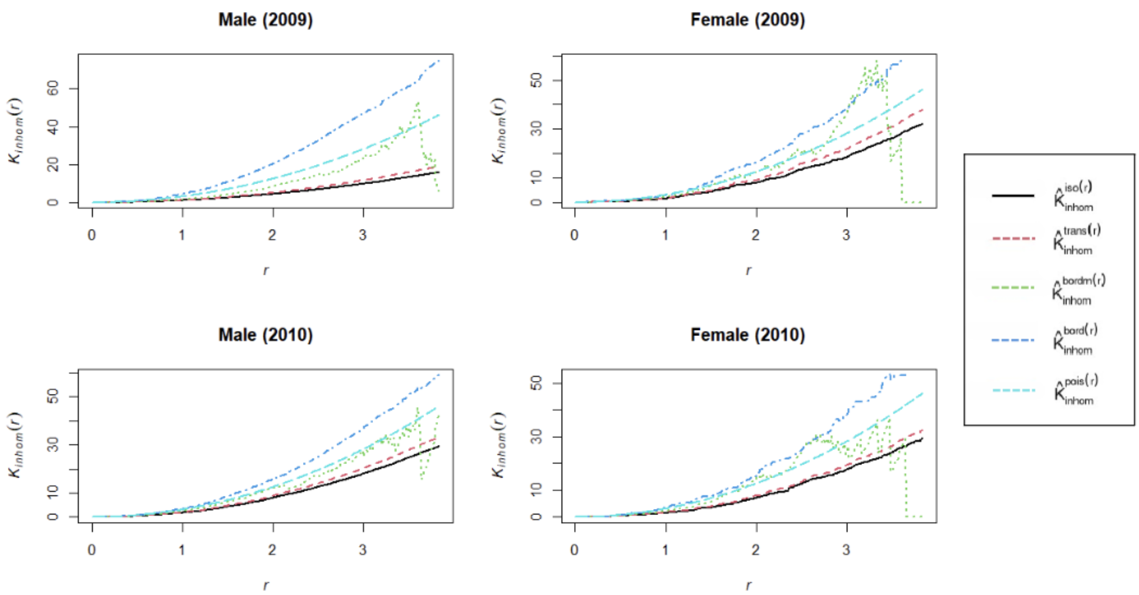

Figure 18: Inhomogeneous K Function for Gender

According to what can be seen on **Figure 18**, for both genders and years, the pattern seems to be notably more spatially disperse than that expected of a Complete Spatial Randomness (CSR), as for all four cases, the $K_{\text{inhom}}(r)$ curve represented in black is below the CSR line (represented as the dashed blue line). This suggests that the fatal victims of traffic accidents are not necesarilly clustered in one specific area but are more dispersed. It is important to note that for female victims during 2009, the pattern seems to be closer to a random process.

5.2 Age group

For Age Groups, the variable was recoded to get similar age groups as those stipulated by the Colombian Ministry of Health and Social Protection [5]. Getting in result five age groups; 0-11 years (Childhood), 12-18 (Adolescence), 19-26 (Young adulthood), 27-59 (Adulthood), and 60+ (Senior Citizens). **Figure 19** shows the amount of fatal traffic accidents per age group, where it can be seen that adults and young adults compose the vast majority of the victims.

It is important to note that for the childhood age period (0-11) in 2010, there was only one victim, therefore neither the density, nor the $K_{\text{inhom}}(r)$ could be estimated.

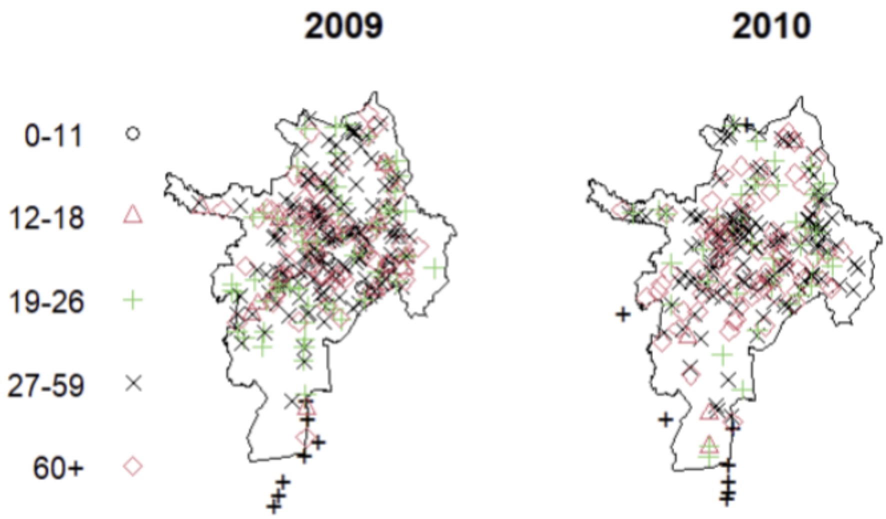

Figure 19: Fatal Traffic Accidents per Age Group

Figure 20 shows how the spatial concentration of traffic fatalities varies by age group and year in the city of Cali. It is observed that the density of points varies both between age groups and between years, suggesting changes in the spatial distribution of the population over time.

The increased crash density in 2010 for middle-aged adults suggests a possible deterioration in road safety in those areas or an increase in vehicle or human traffic in those sectors. The concentration of accidents in specific areas could be related to factors such as road infrastructure, traffic volume, or risk behaviors associated with different ages. By this figure, it can be concluded that the spatial pattern of fatal crashes is not homogeneous and suggests that certain areas of the city are more prone to fatal crashes, especially for certain age groups.

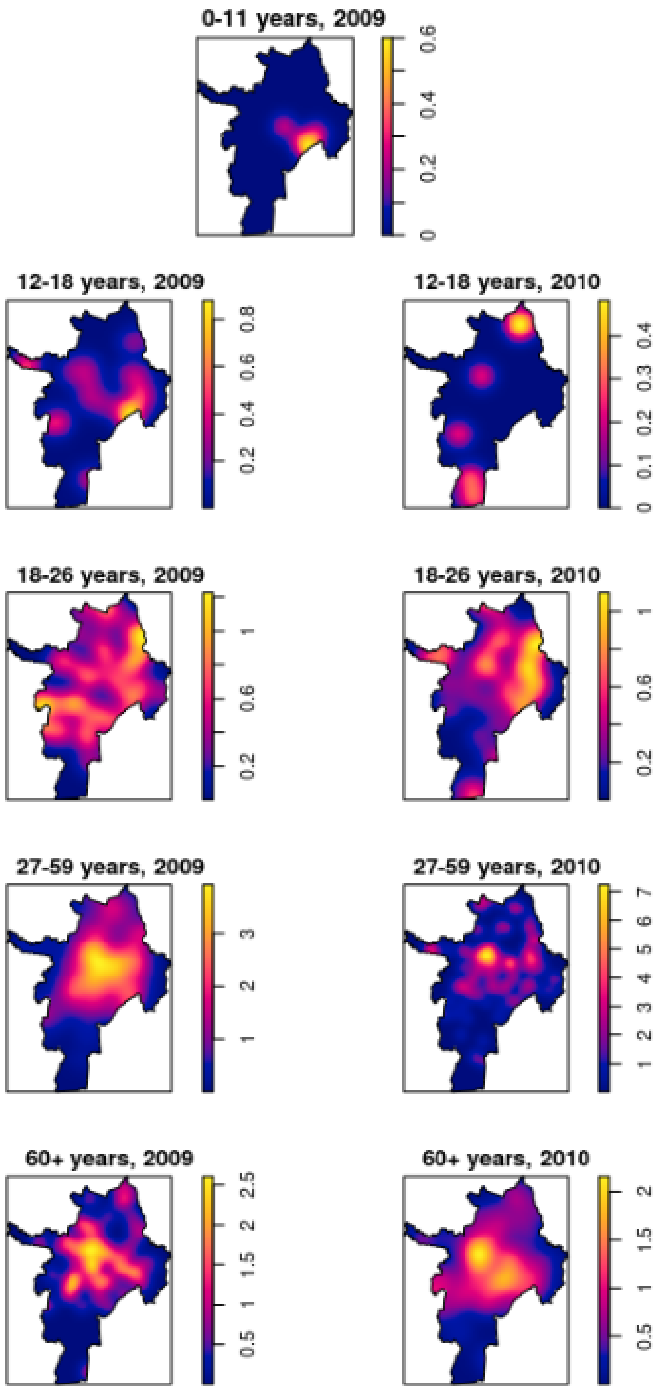

Figure 20: Kernel Density estimation of Homicides by Traffic Accidents per Age Group and Year

In the graphs of the Inhomogeneous K function seen in **Figure 21** and **Figure 22**, it can be seen that in most cases, the $K_{\text{inhom}}(r)$ curve represented in black is below the CSR line (represented as the dashed blue line). This indicates a regular pattern in the distribution of fatal traffic accidents, suggesting that accidents tend to be more spatially dispersed than would be expected under a model of complete randomness (CSR). This suggests that, although the accidents are not completely randomized, they are also not all strongly clustered in small areas, reflecting a more dispersed pattern overall. It is important to note that for 2010, the pattern for the 60+ age group seems to be close to CSR, whereas and the 12-18 age group that drifts away from the CSR line and suggests a notably disperse pattern for this age group in particular.

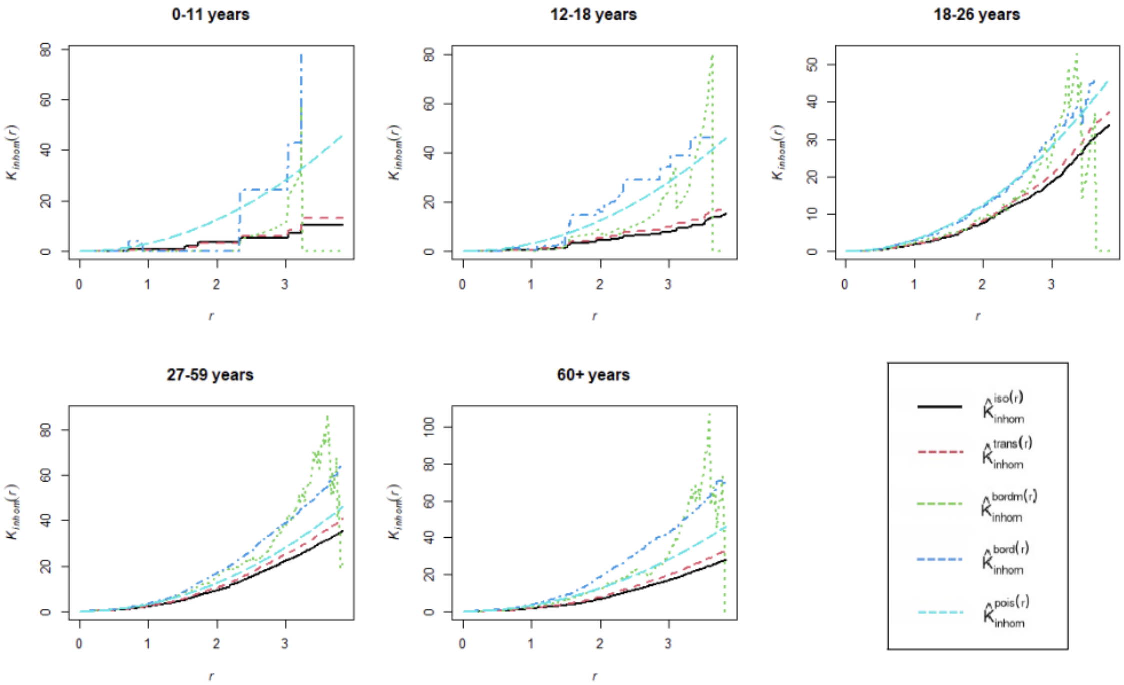

Figure 21: Inhomogeneous K Function for Age Groups, 2009

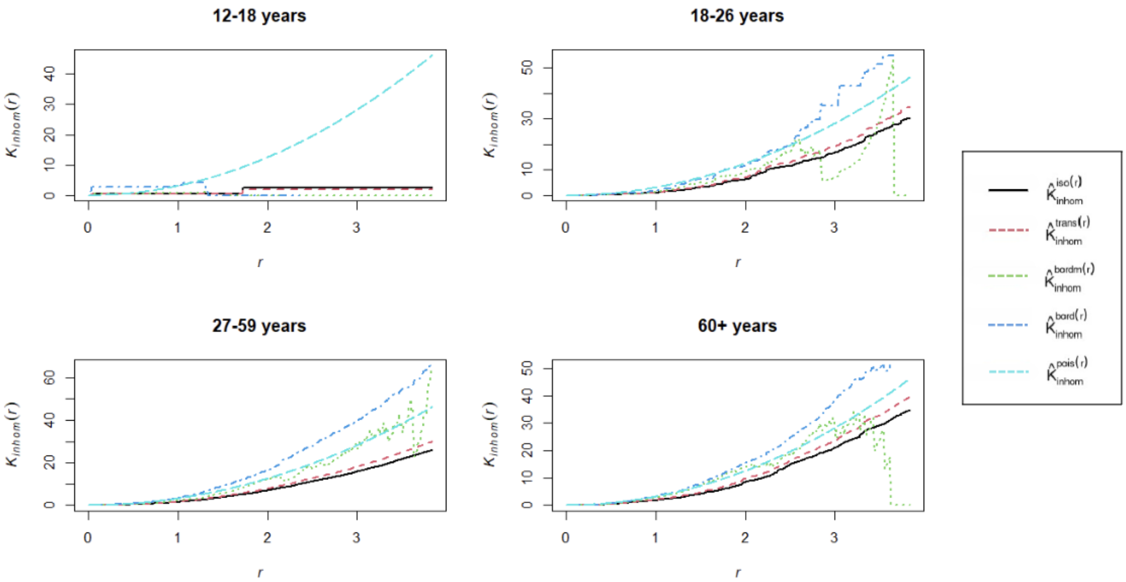

Figure 22: Inhomogeneous K Function for Age Groups, 2010

5.3 Condition

Figure 23 shows the amount of fatal victims of traffic accidents per condition. This plot shows a prevalence of motorcycle driving traffic accident fatal victims followed by pedestrian fatal victims for 2009, whereas for 2010, the prevalence is based on pedestrians, followed then by motorcycles.

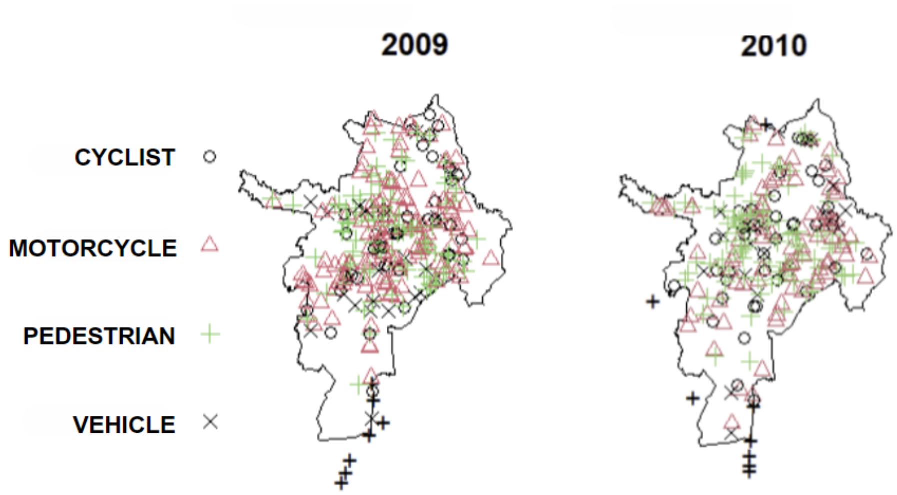

Figure 23: Fatal Traffic Accidents per Condition

Figure 24 shows the kernel estimated density of fatal traffic accidents per condition, in general, a high concentration of accidents is observed in certain areas of the city, with variations by type of user. For motorcyclists, pedestrians and cyclists, the areas of high density appear to be mostly dispersed between the center and north-east of the city in both years, suggesting that certain parts of the city are systematically dangerous for these groups. Crash density for pedestrians shows a considerable increase in 2010 compared to 2009, which could indicate an increase in pedestrian vulnerability in that year.

For other vehicles, the density is lower compared to the other categories, but is also concentrated in specific areas. These patterns imply that interventions to improve road safety should be targeted to areas where a high concentration of accidents is observed, especially for the most vulnerable users such as motorcyclists and pedestrians. This also, once again highlights the non constant density of the pattern, leading towards further exploration of the pattern through the $K_{\text{inhom}}(r)$ function.

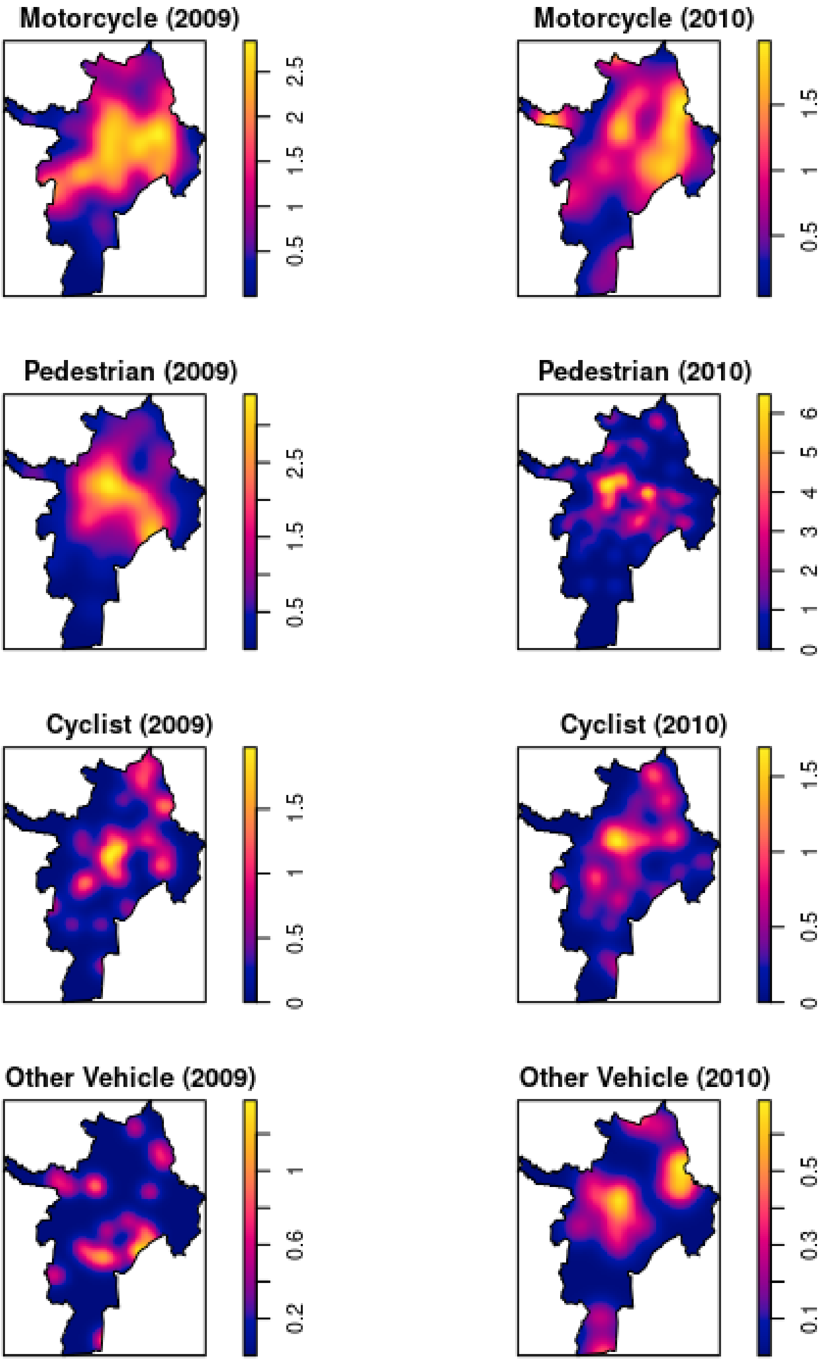

Figure 24: Kernel Density estimation of Homicides by Traffic Accidents per Condition and Year

Figure 25 and **Figure 26** show the graphs of the Inhomogeneous K function per condition for 2009 and 2010 respectively. In both of these, it is apparent that the pattern stayed more dispersed than a CSR, given by the lower values than the reference line. Notably, 'other vehicle' was the most dispersed for both years, followed by cyclists. Whereas 'motorcycle' and 'pedestrian' for 2009 get very close to a spatial randomness but still present more dispersion than a CSR.

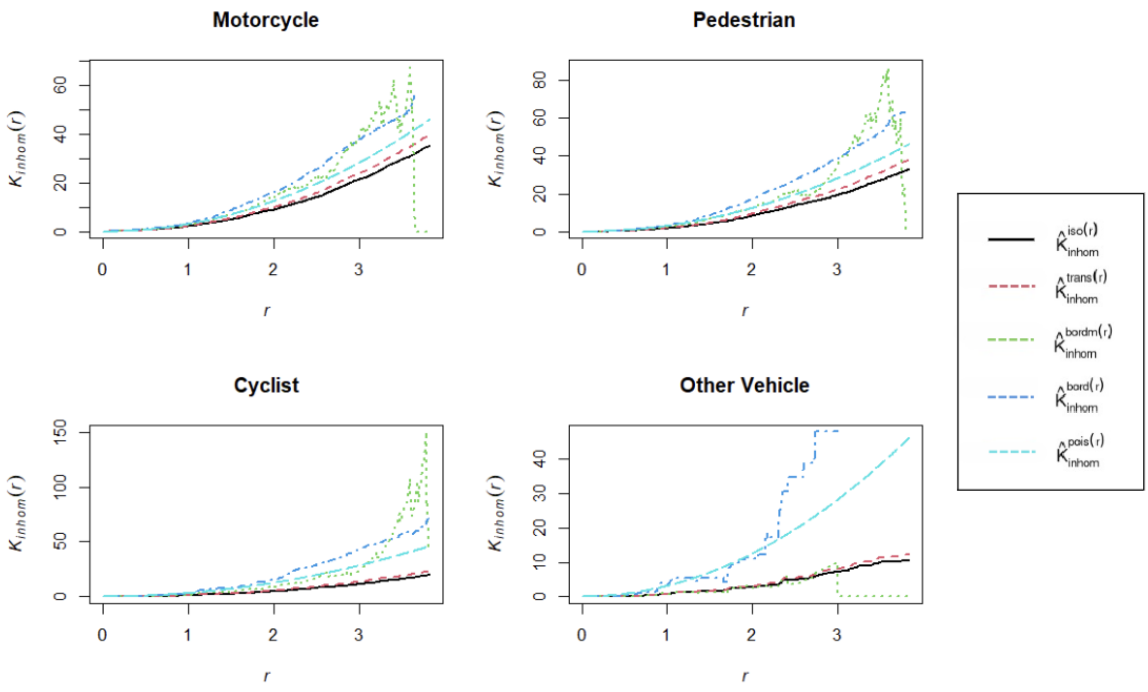

Figure 25: Inhomogeneous K Function for Condition, 2009

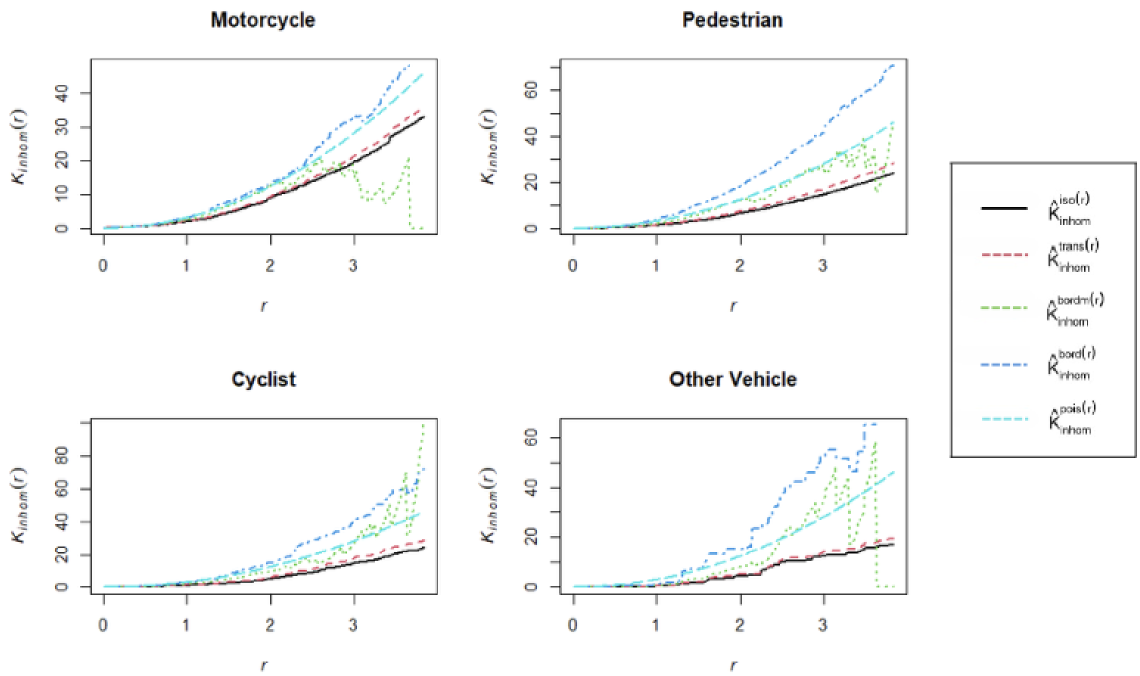

Figure 26: Inhomogeneous K Function for Condition, 2010

6 Model adjustment

The following models analyze the spatial distribution of homicides in Cali based on gender, condition, and age group. Initially, an inhomogeneous Poisson cluster model was considered for its ability to detect spatial clusters with varying intensities. However, further analysis (detailed in the “Analysis of Point Patterns by Year” section) revealed that the data exhibited a regular distribution rather than distinct clusters. As a result, we adopted a non-homogeneous Poisson model, which better suits the non-stationary nature of our data, capturing spatial variations in event intensity without assuming a cluster structure. Each model was applied separately for each variable and year, covering data from 2009 and 2010 for a robust analysis.

To implement this, we used the **ppm** function from the *spatstat* package in R. This approach allows for modeling event intensity as a function of categorical variables (marks) and spatial coordinates, providing a comprehensive view of how homicide patterns vary with these factors and their geographic distribution. By incorporating variables such as sex, condition, and age group along with spatial coordinates (x and y), the **ppm** function enables us to fit a Poisson point process where spatial intensity depends on both marks and location. This consistent approach across all variables and years ensures that our analysis thoroughly captures the relationship between the variables and the spatial intensity of events.

6.1 Gender 2009

To explore how the spatial distribution of homicides in Cali during 2009 is influenced by gender, a non-homogeneous Poisson model was implemented. The model incorporates the gender of the victims as a categorical variable (with levels "F" from Female and "M" from Male), along with the spatial coordinates (x and y), to allow the intensity of homicides to vary across different locations and between genders.

6.1.1 Coefficients of the Fitted Model

The coefficients obtained from the fitted model are summarized below:

Parameter	Estimate	Standard Error	CI 95% (Lower)	CI 95% (Upper)
Intercept	-89.8872	20.7932	-130.6412	-49.1332
marksM	1.3863	0.1409	1.1102	1.6624
x	0.0448	0.0191	0.0073	0.0823
y	0.0477	0.0147	0.0190	0.0765

Table 5: Coefficients of the non-homogeneous Poisson model for gender in 2009.

The fitted model reveals a significant influence of both gender and spatial location on homicide intensity in Cali during 2009. The negative intercept (-89.8872) indicates a very low baseline intensity when not accounting for gender or location, suggesting that in the absence of these factors, the expected intensity of homicides is extremely low. The standard error for the intercept (20.7932) indicates a moderate level of uncertainty, and the 95% confidence interval (CI) for the intercept, ranging from -130.6412 to -49.1332, does not include zero, reinforcing the statistical significance of this baseline estimate.

The positive coefficient for *marksM* (1.3863) suggests that homicides are more frequent among males compared to females. Specifically, the exponentiated coefficient $\exp(1.3863) \approx 4$ implies that the intensity of homicides for males is approximately four times higher than for females. The standard error for *marksM* (0.1409) is relatively small, indicating a precise estimate, and the 95% CI (1.1102 to 1.6624) further supports the significance of this effect, with the interval well above zero.

The coefficients for the spatial coordinates x (0.0448) and y (0.0477) indicate slight but statistically significant increases in homicide intensity across these dimensions. The standard errors for x (0.0191) and y (0.0147) are small, suggesting precise estimates. The 95% CI for

x (0.0073 to 0.0823) and y (0.0190 to 0.0765) do not include zero, confirming the positive association of spatial location with homicide intensity.

6.1.2 Predicted Intensity Maps

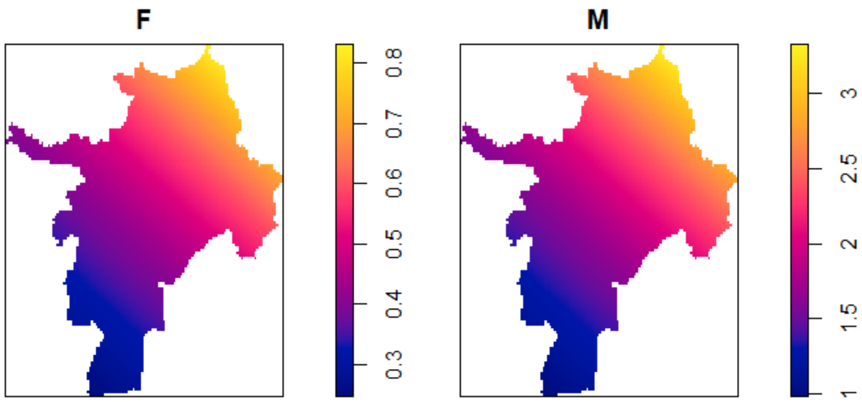

Figure 27: Predicted Homicide Intensity by Gender in Cali, 2009

The maps in Figure 27 depict the predicted intensity of homicides for females (left) and males (right) across Cali in 2009. The color gradient from blue to yellow indicates increasing intensity, with yellow areas representing the highest predicted intensity.

For females, the intensity varies from approximately 0.3 to 0.8, with the highest intensities concentrated in the northern regions of the city. For males, the intensity ranges from about 1 to 3, also showing higher concentrations in the northern areas, but with an overall greater intensity compared to females.

These maps further illustrate the gender-based differences in homicide intensity identified by the model. The broader range and higher maximum intensity for males reinforce the earlier finding that the risk of homicide is significantly greater for males. The geographic gradient also suggests that certain areas, particularly in the north, are more prone to higher homicide rates, and these trends are more pronounced among males.

6.2 Gender 2010

The analysis of homicide distribution by gender in Cali for the year 2010 follows the same methodology as in 2009. A non-homogeneous Poisson model was used, incorporating gender as a categorical variable, now for the year 2010, along with spatial coordinates (x and y) to allow the intensity of homicides to vary across different locations and between genders.

6.2.1 Coefficients of the Fitted Model

The coefficients obtained from the fitted model are summarized below:

Parameter	Estimate	Standard Error	CI 95% (Lower)	CI 95% (Upper)
Intercept	-73.4957	22.4874	-117.5701	-29.4213
marksM	1.5525	0.1606	1.2378	1.8672
x	0.0230	0.0207	-0.0175	0.0635
y	0.0552	0.0158	0.0241	0.0862

Table 6: Coefficients of the non-homogeneous Poisson model for gender in 2010.

The fitted model for 2010 indicates a significant influence of both gender and spatial location on homicide intensity in Cali. The negative intercept (-73.4957) suggests a very low baseline intensity when not considering gender or location, highlighting that without these factors, the expected intensity of homicides is extremely low.

The positive coefficient for *marksM* (1.5525) indicates that homicides are more frequent among males compared to females. Specifically, the exponentiated coefficient $\exp(1.5525) \approx 4.72$ implies that the intensity of homicides for males is nearly five times higher than for females. This transformation from the log scale to the original intensity scale is essential for meaningful interpretation. The standard error for *marksM* is 0.1606, which is relatively small compared to the estimate, indicating a precise estimate of the coefficient. The 95% confidence interval for *marksM* (1.2378 to 1.8672) does not cross zero, further confirming the significance of this effect, with a high degree of confidence that the true effect of being male on homicide intensity is positive.

The coefficients for the spatial coordinates x (0.0230) and y (0.0552) also indicate slight but significant increases in homicide intensity across these dimensions, particularly along the y-axis, which shows a stronger effect. The standard errors for x (0.0207) and y (0.0158) suggest that these estimates are reasonably precise, although the effect of x is less certain compared to y, as indicated by the wider confidence interval for x (-0.0175 to 0.0635) which barely includes zero. This suggests that while there is a tendency for the intensity to increase with higher x-values, this effect is not as robust as the effect along the y-axis, which has a narrower confidence interval (0.0241 to 0.0862), clearly indicating a positive association.

6.2.2 Predicted Intensity Maps

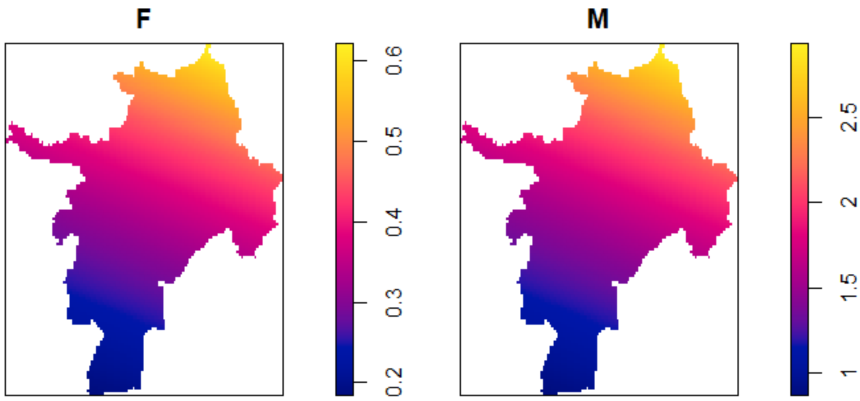

Figure 28: Predicted Homicide Intensity by Gender in Cali, 2010

The maps in Figure 28 depict the predicted intensity of homicides for females (left) and males (right) across Cali in 2010. The color gradient from blue to yellow represents increasing intensity, with yellow areas indicating the highest predicted intensity.

For females, the intensity ranges from approximately 0.2 to 0.6, with higher intensities concentrated in the northern regions. For males, the intensity ranges from about 1 to 2.5, also showing higher concentrations in the northern areas, but with an overall greater intensity compared to females.

These maps illustrate the pronounced difference in homicide intensity between genders, in 2010, with males facing a significantly higher risk, particularly in specific northern areas of the city. The spatial gradient further confirms that geographic location plays a crucial role in the distribution of homicides, with certain regions experiencing consistently higher rates.

6.3 Age Group 2009

As with the previous models, a non-homogeneous Poisson model was applied to analyze the spatial distribution of homicides in Cali for 2009, this time focusing on different age groups. The model incorporates the age group of the victims as a categorical variable, along with spatial coordinates (x and y), to account for variations in homicide intensity across different locations.

6.3.1 Coefficients of the Fitted Model

The coefficients obtained from the fitted model are summarized below:

Parameter	Estimate	Standard Error	CI 95% (Lower)	CI 95% (Upper)
Intercept	-92.6441	20.7989	-133.4091	-51.8791
marks12-18	1.4469	0.5557	0.3577	2.5361
marks19-26	2.6912	0.5167	1.6786	3.7039
marks27-59	3.6700	0.5063	2.6776	4.6623
marks60+	2.9704	0.5127	1.9656	3.9752
x	0.0448	0.0191	0.0073	0.0823
y	0.0477	0.0147	0.0190	0.0765

Table 7: Coefficients of the non-homogeneous Poisson model for age groups in 2009.

The fitted model for 2009 shows a significant influence of age group and spatial location on homicide intensity in Cali. The negative intercept (-92.6441) indicates a very low baseline intensity when not considering age or location, which suggests that in the absence of these factors, the expected intensity of homicides is extremely low. The standard error for the intercept (20.7989) is moderate, and the 95% confidence interval (CI) for the intercept, ranging from -133.4091 to -51.8791, confirms the significance of this baseline estimate.

The positive coefficients for the age groups indicate that the intensity of homicides increases with age, particularly for the groups 27-59 and 60+. For instance, the coefficient for the 27-59 age group (3.6700) is the highest, implying that homicides are most frequent among this group. The exponentiated coefficient $\exp(3.6700) \approx 39.25$ suggests that the intensity of homicides for the 27-59 age group is nearly 40 times higher than that for the reference group (0-11 years). The standard errors for these coefficients are relatively small, indicating precise estimates, and the confidence intervals confirm the significance of these effects, as they do not include zero.

The spatial coefficients for x (0.0448) and y (0.0477) show slight increases in homicide intensity across these dimensions. These coefficients are statistically significant, with small standard errors and confidence intervals that do not include zero, suggesting that geographic location also plays a role in the distribution of homicides across age groups.

6.3.2 Predicted Intensity Maps

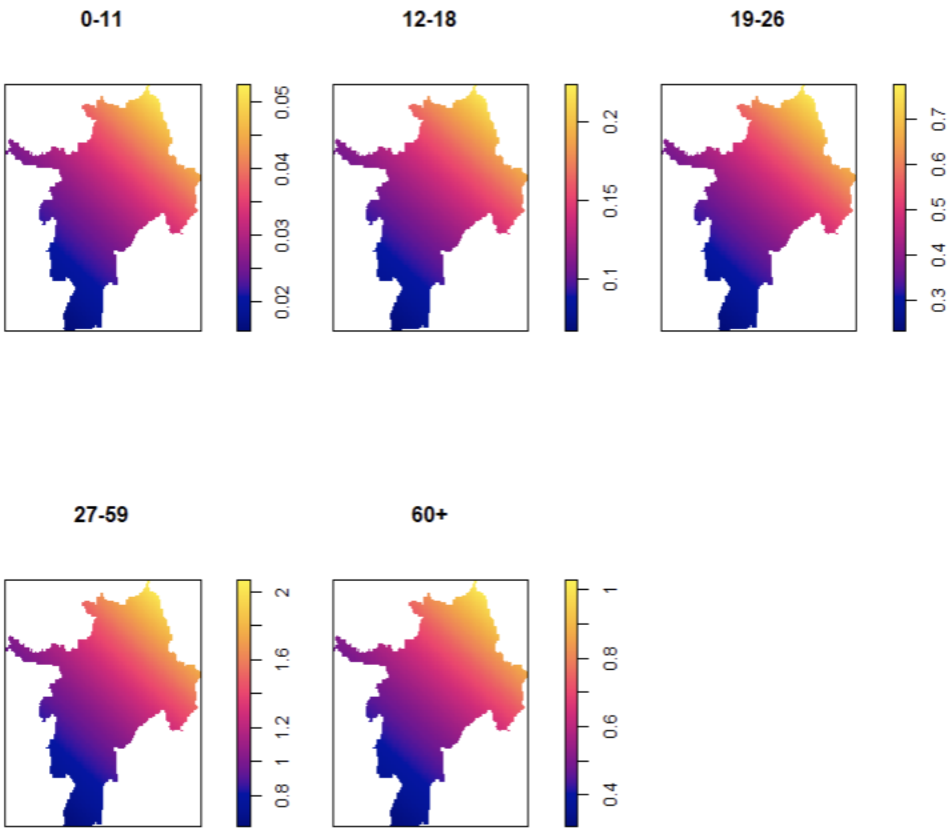

Figure 29: Predicted Homicide Intensity by Age Group in Cali, 2009

The maps in Figure 29 illustrate the predicted intensity of homicides across different age groups in Cali for 2009. The color gradient from blue to yellow indicates increasing intensity, with yellow areas representing the highest predicted intensities.

For the youngest age group (0-11), the intensity is very low, ranging from 0.02 to 0.05, indicating minimal homicide risk. As age increases, particularly for the 27-59 age group, the intensity rises significantly, with values ranging from 0.8 to 2.0, concentrated mostly in the northern regions of the city. The 60+ age group also shows elevated intensity levels, though slightly lower than the 27-59 group, ranging from 0.4 to 1.0.

These maps clearly demonstrate the age-related differences in homicide risk, with older age groups, particularly 27-59, facing the highest risk. The geographic distribution also highlights certain areas, particularly in the north, where the intensity of homicides is consistently higher across all age groups.

6.4 Age Group 2010

As with the previous models, a non-homogeneous Poisson model was applied to analyze the spatial distribution of homicides in Cali for 2010, focusing on different age groups. The model incorporates the age group of the victims as a categorical variable, along with spatial coordinates (x and y), to account for variations in homicide intensity across different locations.

6.4.1 Coefficients of the Fitted Model

The coefficients obtained from the fitted model are summarized below:

Parameter	**Estimate**	**Standard Error**	**CI 95% (Lower)**	**CI 95% (Upper)**
Intercept	-77.3459	22.5091	-121.4629	-33.2289
marks12-18	1.7918	1.0801	-0.3252	3.9088
marks19-26	3.8067	1.0111	1.8250	5.7883
marks27-59	4.9053	1.0037	2.9381	6.8725
marks60+	4.4067	1.0061	2.4348	6.3786
x	0.0230	0.0207	-0.0175	0.0635
y	0.0552	0.0158	0.0241	0.0862

Table 8: Coefficients of the non-homogeneous Poisson model for age groups in 2010.

The fitted model for 2010 reveals a significant influence of age group and spatial location on homicide intensity in Cali. The negative intercept (-77.3459) suggests a very low baseline intensity when not considering age or location, with a moderate standard error (22.5091). The 95% confidence interval for the intercept ranges from -121.4629 to -33.2289, confirming the significance of this baseline estimate.

The coefficients for the age groups demonstrate that the intensity of homicides increases with age, with the 27-59 group having the highest coefficient (4.9053). This implies that homicides are most frequent among individuals in this age group. The exponentiated coefficient $\exp(4.9053) \approx 135$ suggests that the intensity of homicides for the 27-59 age group is approximately 135 times higher than for the reference group (0-11 years). The standard errors for these coefficients are relatively moderate, and the confidence intervals indicate that the effects for the 19-26, 27-59, and 60+ age groups are significant, as they do not include zero. The 12-18 age group has a wider confidence interval that includes zero, suggesting a less certain effect.

The spatial coefficients for x (0.0230) and y (0.0552) indicate slight increases in homicide intensity across these dimensions, particularly along the y-axis, which has a stronger effect. The confidence intervals for x include zero, suggesting some uncertainty, whereas the interval for y does not, indicating a significant positive association with homicide intensity.

6.4.2 Predicted Intensity Maps

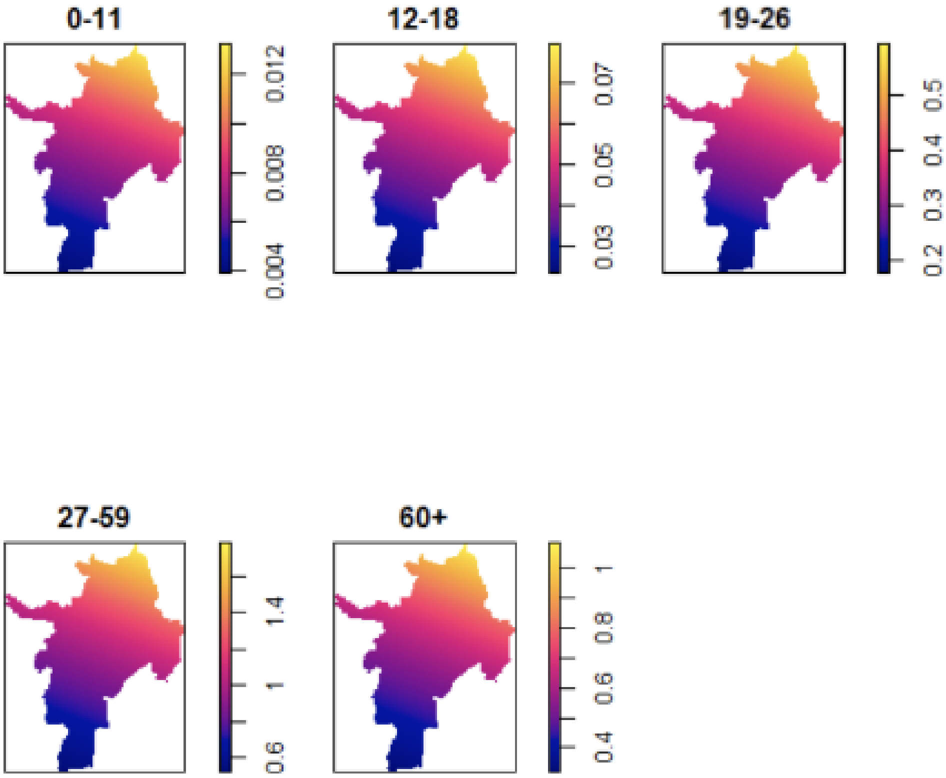

Figure 30: Predicted Homicide Intensity by Age Group in Cali, 2010

The maps in Figure 30 show the predicted intensity of homicides across different age groups in Cali for 2010. The color gradient from blue to yellow indicates increasing intensity, with yellow areas representing the highest predicted intensities.

For the youngest age group (0-11), the intensity is extremely low, ranging from 0.004 to 0.012, indicating minimal homicide risk. As age increases, particularly for the 27-59 age group, the intensity rises significantly, with values ranging from 0.6 to 1.4, predominantly in the northern regions of the city. The 60+ age group also shows elevated intensity levels, though slightly lower than the 27-59 group, ranging from 0.4 to 1.0.

These maps clearly illustrate the age-related differences in homicide risk, with the 27-59 age group facing the highest risk. The geographic distribution also highlights certain areas, particularly in the north, where the intensity of homicides is consistently higher across all age groups.

6.5 Condition 2009

To explore the spatial distribution of homicides in Cali during 2009, a non-homogeneous Poisson model was employed with a focus on the condition of the individuals involved, such as motorcyclists, pedestrians, cyclists, and vehicle occupants. By incorporating condition as a categorical variable along with spatial coordinates (x and y), the model allows us to assess how homicide intensity varies across different locations depending on the individual's condition.

6.5.1 Coefficients of the Fitted Model

The coefficients obtained from the fitted model are summarized below:

Parameter	Estimate	Standard Error	CI 95% (Lower)	CI 95% (Upper)
Intercept	-89.0887	20.7930	-129.8423	-48.3351
marksPEATON	-0.2231	0.1268	-0.4716	0.0253
marksCICLISTA	-1.2040	0.1759	-1.5488	-0.8592
marksVEHICULO	-1.8971	0.2340	-2.3558	-1.4385
x	0.0448	0.0191	0.0073	0.0823
y	0.0477	0.0147	0.0190	0.0765

Table 9: Coefficients of the non-homogeneous Poisson model for condition in 2009.

The fitted model for 2009 reveals a significant influence of condition and spatial location on homicide intensity in Cali. The negative intercept (-89.0887) suggests a very low baseline intensity when not considering condition or location, with a moderate standard error (20.7930). The 95% confidence interval for the intercept ranges from -129.8423 to -48.3351, confirming the significance of this baseline estimate.

The coefficients for the conditions show that the intensity of homicides varies depending on the condition. For example, motorcyclists (MOTO) have the highest baseline intensity (reference group), while the coefficients for pedestrians (marksPEATON: -0.2231), cyclists (marksCICLISTA: -1.2040), and vehicle occupants (marksVEHICULO: -1.8971) are all negative. This suggests that compared to motorcyclists, these other groups have lower intensities. Specifically, the exponentiated coefficients indicate that the intensity for pedestrians is approximately 0.8 times that of motorcyclists, for cyclists it is 0.3 times, and for vehicle occupants, it is 0.15 times. The standard errors for these coefficients are relatively small, indicating precise estimates, and the confidence intervals for cyclists and vehicle occupants do not include zero, confirming the significance of these effects. The pedestrian group, however, has a confidence interval that includes zero, suggesting a less certain effect.

The spatial coefficients for x (0.0448) and y (0.0477) indicate slight increases in homicide intensity across these dimensions. These coefficients are statistically significant, with small standard errors and confidence intervals that do not include zero, suggesting that geographic location plays a role in the distribution of homicides across different conditions.

6.5.2 Predicted Intensity Maps

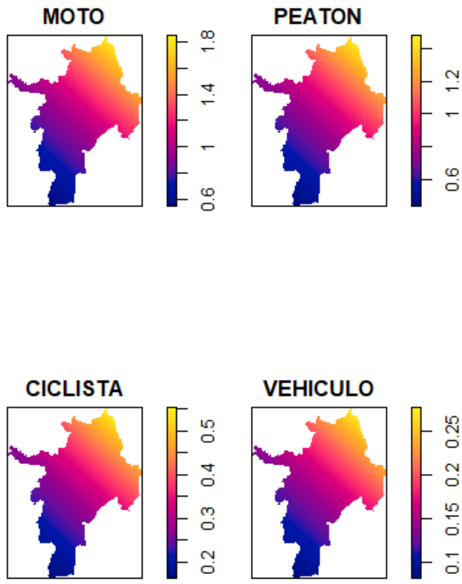

Figure 31: Predicted Homicide Intensity by Condition in Cali, 2009

The maps in Figure 31 show the predicted intensity of homicides across different conditions in Cali for 2009. The color gradient from blue to yellow indicates increasing intensity, with yellow areas representing the highest predicted intensities.

For motorcyclists (MOTO), the intensity is highest, ranging from 0.6 to 1.8, indicating a relatively high risk of homicide, particularly in the northern parts of the city. Pedestrians (PEATON) also face a considerable risk, with intensity values ranging from 0.6 to 1.2. Cyclists (CICLISTA) and vehicle occupants (VEHICULO) have much lower predicted intensities, with ranges from 0.2 to 0.5 and 0.1 to 0.25, respectively.

These maps clearly illustrate the differences in homicide risk based on condition, with motorcyclists facing the highest risk. The geographic distribution also highlights certain areas, particularly in the north, where the intensity of homicides is consistently higher across all conditions.

6.6 Condition 2010

To further analyze the spatial distribution of homicides in Cali during 2010, a non-homogeneous Poisson model was employed with a focus on the condition of individuals involved, such as motorcyclists, pedestrians, cyclists, and vehicle occupants. This model incorporates condition as a categorical variable alongside spatial coordinates (x and y) to explore how homicide intensity varies based on these factors across different locations.

6.6.1 Coefficients of the Fitted Model

The coefficients obtained from the fitted model are summarized below:

Parameter	Estimate	Standard Error	CI 95% (Lower)	CI 95% (Upper)
Intercept	-72.7609	22.4871	-116.8348	-28.6870
marksPEATON	0.1246	0.1386	-0.1471	0.3962
marksCICLISTA	-0.8714	0.1860	-1.2359	-0.5069
marksVEHICULO	-1.6405	0.2507	-2.1318	-1.1492
x	0.0230	0.0207	-0.0175	0.0635
y	0.0552	0.0158	0.0241	0.0862

Table 10: Coefficients of the non-homogeneous Poisson model for condition in 2010.

The fitted model for 2010 shows a significant influence of condition and spatial location on homicide intensity in Cali. The negative intercept (-72.7609) indicates a low baseline intensity when not considering condition or location, with a moderate standard error (22.4871). The 95% confidence interval for the intercept ranges from -116.8348 to -28.6870, confirming the significance of this baseline estimate.

The coefficients for the conditions highlight that the intensity of homicides varies depending on the condition. Motorcyclists (MOTO) are used as the reference group, with the highest baseline intensity. Pedestrians (marksPEATON: 0.1246) have a slightly higher intensity than motorcyclists, though the effect is not statistically significant (CI includes zero). In contrast, cyclists (marksCICLISTA: -0.8714) and vehicle occupants (marksVEHICULO: -1.6405) have significantly lower intensities, as indicated by the negative coefficients. The exponentiated coefficients suggest that the intensity for cyclists is approximately 0.42 times that of motorcyclists, and for vehicle occupants, it is about 0.19 times. The confidence intervals for cyclists and vehicle occupants do not include zero, confirming the significance of these effects.

The spatial coefficients for x (0.0230) and y (0.0552) indicate slight increases in homicide intensity across these dimensions, particularly along the y-axis, which shows a stronger effect. The confidence intervals for x include zero, suggesting some uncertainty, while the interval for y does not, indicating a significant positive association with homicide intensity.

6.6.2 Predicted Intensity Maps

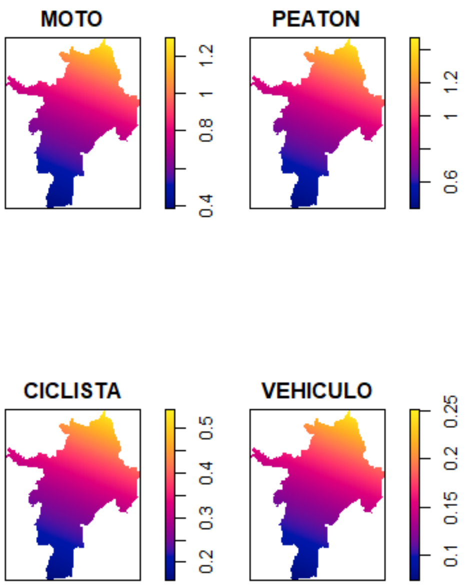

Figure 32: Predicted Homicide Intensity by Condition in Cali, 2010

The maps in Figure 32 depict the predicted intensity of homicides across different conditions in Cali for 2010. The color gradient from blue to yellow indicates increasing intensity, with yellow areas representing the highest predicted intensities.

For motorcyclists (MOTO), the intensity remains high, ranging from 0.6 to 1.2, indicating a relatively high risk of homicide, particularly in certain areas of the city. Pedestrians (PEATON) also face a similar level of risk, with intensity values ranging from 0.6 to 1.2. Cyclists (CICLISTA) and vehicle occupants (VEHICULO), however, show much lower predicted intensities, with ranges from 0.2 to 0.5 and 0.1 to 0.25, respectively.

These maps clearly illustrate the differences in homicide risk based on condition, with motorcyclists and pedestrians facing the highest risks. The geographic distribution also highlights certain areas where the intensity of homicides is consistently higher across these conditions.

7 Conclusion

This study provides a comprehensive analysis of the spatial distribution of homicides due to traffic accidents in Cali, Colombia, for the years 2009 and 2010. By employing non-homogeneous Poisson models, we explored how factors such as gender, age group, and the condition of individuals involved (e.g., motorcyclists, pedestrians, cyclists, vehicle occupants) influence the intensity of these events across different locations in the city. The results consistently showed that certain demographics, particularly young adults and middle-aged individuals, as well as motorcyclists and pedestrians, are at a significantly higher risk of being involved in fatal traffic accidents. The spatial analysis further highlighted that these high-risk groups tend to spread across specific areas of Cali, particularly in the center to northern regions, indicating that these locations are more prone to deadly incidents.

The application of Kernel Density Estimation (KDE) and the use of inhomogeneous K function $K_{\text{inhom}}(r)$ allowed for a nuanced understanding of the non-random nature of these accidents. The KDE maps illustrated clear "hotspots" of high accident intensity, aligning with areas of dense traffic and significant pedestrian activity. The inhomogeneous K functions confirmed the non-stationary and non-isotropic nature of the point patterns, reinforcing the importance of spatial considerations in traffic accident analysis. The findings suggest that interventions should be tailored to these high-risk areas and groups, potentially focusing on improving road infrastructure, enforcing traffic laws, and implementing targeted safety campaigns.

This study highlights the critical need for location-specific and demographic-specific strategies to reduce traffic-related homicides in Cali. The observed patterns indicate that while the overall risk varies across the city, certain populations and areas are consistently at higher risk. Future efforts should prioritize these vulnerable groups and regions to effectively mitigate the incidence of fatal traffic accidents. Continued monitoring and adaptive safety measures are essential to address the dynamic nature of urban traffic environments and ensure the safety of all road users in Cali.

References

[1] A. Baddeley and R. Turner. Modelling spatial point patterns in r, 2019.

[2] X. J. Eskindir Ayele Atumo and C. Fu. Spatial point pattern analysis of traffic violations in luzhou city, china. *Transportation Letters*, 14(10):1162–1171, 2022.

[3] Esri. Cómo funciona el análisis cluster espacial de distancia múltiple (función k de ripley). `https://pro.arcgis.com/es/pro-app/latest/tool-reference/spatial-statistics/h-how-multi-distance-spatial-cluster-analysis-ripl.htm`. Accessed: 2024-09-01.

[4] R. G. Henao. Notas de clase: Estadística espacial, 2011.

[5] Ministerio de Salud y Protección Social de Colombia. Ciclo de vida. `https://www.minsalud.gov.co/proteccionsocial/Paginas/cicloVida.aspx`, 2024.

[6] Spatstat Team. Solution 5: Marked point patterns. `https://www.spatstat.org/SSAI2017/solutions/solution05.html`, 2017.

[7] L. D. R. Sánchez and W. P. Rios. Modelos de procesos puntuales para la identificación espacial de los sismos ocurridos en colombia, 2019.