



Clustering Approach to Generate Pedestrian Traffic Pattern Groups: An Exploratory Analysis*

Carolina Matamoros-Jiménez^a ■ Henry Hernández-Vega^b

Abstract: This study shows the development of patterns of temporal hourly volume distributions in an urban area in Costa Rica, based on a cluster analysis of pedestrian data. This study aims to establish specific pattern groups for the temporal variation of weekday pedestrian volumes applying cluster analysis in the central business district of Guadalupe in San José. For 46 counting sites, vectors with the weekday hourly factors, the proportion of the daily pedestrian traffic, were estimated. A hierarchical cluster method was implemented to group the vectors of hourly factors from the different counting sites. This method groups elements by minimizing the Euclidean distance between elements of the same group and, at the same time, maximizing the distances from elements of other groups. In addition, the groups found through this analysis are related to land use through buffers of different radii. Eight temporal pattern groups were obtained through cluster analysis. Two pattern groups account for more than two-thirds of the sites included in the study. Fisher's exact independence test shows that banks and public services could explain some of the patterns observed. The classification of 46 counting sites based on temporal distribution patterns, and the relation with the establishments in the area, allows a simplification of the information and facilitates an understanding of the pedestrian mobility in the area. Further research is required that leads towards geographical elements that could explain the differences in temporal and mobility patterns.

Keywords: Pedestrian; temporal pattern; cluster analysis; mobility; urban area

Received: 13/12/2019

Accepted: 16/06/2021

Available online: 31/12/2021

* Research article

a BSc Civil Engineering. Laboratorio Nacional de Materiales y Modelos Estructurales, Universidad de Costa Rica, Montes de Oca, Costa Rica. E-mail: caromatamoros08@gmail.com
ORCID: <http://orcid.org/0000-0002-3415-7474>

b MSc Civil Engineering. Laboratorio Nacional de Materiales y Modelos Estructurales, Universidad de Costa Rica Montes de Oca, Costa Rica. E-mail: Henry.hernandezvega@ucr.ac.cr
ORCID: <https://orcid.org/0000-0002-4765-7320>

How to cite: H. Hernández-Vega and C. Matamoros-Jiménez, "Clustering Approach to Generate Pedestrian Traffic Pattern Groups: An Exploratory Analysis", *Cien.Ing.Neogranadina*, vol. 31, no. 2, pp. 41-59, Dec. 2021.

Enfoque de agrupamiento para generar grupos de patrones de tráfico peatonal: un análisis exploratorio

Resumen: el presente estudio muestra el desarrollo de patrones de distribuciones temporales de volumen por hora en un área urbana de Costa Rica con base en un análisis de grupos de datos de peatones. Este estudio tiene como objetivo establecer grupos de patrones específicos para la variación temporal de los volúmenes de peatones entre semana mediante la aplicación del análisis de grupos en el distrito comercial central de Guadalupe en San José. Para 46 sitios de conteo, se estimaron los vectores con los factores horarios del día de la semana y la proporción del tráfico peatonal diario. Se implementó un método de agrupamiento jerárquico para los vectores de factores horarios de los sitios de conteo; este método agrupa elementos minimizando la distancia euclidiana entre elementos del mismo grupo mientras maximiza las distancias con elementos de otros grupos. Los grupos encontrados a través de este análisis están relacionados con el uso del suelo a través de búferes de diferentes radios. Se obtuvieron ocho grupos de patrones temporales mediante análisis de grupos; dos de estos representan más de dos tercios de los sitios incluidos en el estudio. La prueba de independencia exacta de Fisher muestra que los bancos y los servicios públicos podrían dar cuenta de algunos de los patrones observados. Esta clasificación permite una simplificación de la información y facilita la comprensión de la movilidad peatonal en la zona. En este sentido, se requieren más investigaciones que conduzcan a elementos geográficos que podrían explicar diferencias en los patrones temporales y de movilidad.

Palabras clave: peatón; patrón temporal; análisis de grupos; movilidad; área urbana

Introduction

Pedestrian traffic data must be a fundamental base for the decision process related to sustainable mobility. However, according to [1], pedestrian mobility and cycling are the two modes of transportation more understudied.

Pedestrian monitoring is required to generate necessary inputs for an informed decision-making process to benefit the population. For instance, to warrant a pedestrian traffic signal on a specific location, it is necessary to have a minimum number of pedestrians crossing the street during a period. Usually, the warrant refers to the pedestrian traffic peak hours; therefore, it is necessary to understand the temporary variations of pedestrian volumes.

In many jurisdictions, short-duration counts can be applied to reduce monitoring costs, and the information collected can be adjusted using expansion factors from places with similar temporal patterns to obtain daily or annual pedestrian traffic estimates, which could allow jurisdictions, for example, to determine the population who benefited from a specific improvement.

Despite this, there is only one study available regarding the temporal variation of pedestrian volumes in Costa Rica [2] and, the monitoring efforts have been historically focused on motorized traffic that provokes an underestimation of non-motorized traffic [1].

There is minimal knowledge of the geographical and temporal variation of pedestrian traffic [3; pp. 4-31], even though pedestrian data must be the base for the evaluation of active mobility projects, transportation planning, and pedestrian exposure analysis [4]. This lack of information limits the ability of agencies to adequately improve and manage the non-motorized infrastructure and improve pedestrian safety efficiently [5], even though the positive impacts of non-motorized modes of transportation have been recently acknowledged [6].

If information is not collected for a specific road user, there is a probability that it could not be adequately included in the management of a city. Therefore, pedestrian volumes are a valuable source of information; however, they have not been extensively explored in Costa Rica. There is a lack

of pedestrian mobility databases in the country, and most of the traffic monitoring effort is related to motorized traffic.

Suppose a pedestrian study is performed in the country based on peak hour manual counts. In that case, there is no reliability on the expansion factors used to predict average daily pedestrian traffic estimates because there is minimal knowledge regarding the temporal variation of these volumes.

Similarly, [3] indicates that in the United States, there is a predisposition to apply very short duration counts for pedestrian traffic monitoring (pp. 4-1). Furthermore, [7] indicated that the most common source for pedestrian data is manual counts, even though annual estimates from short-term manual counts could generate errors between 30 and 60 percent. Additionally, [8] indicate that the research related to the study of temporal factors for non-motorized traffic is scarce.

This article explores a specific urban area through cluster analysis to develop temporal pattern groups. The methodology proposed could be replicated in other jurisdictions to have a better understanding of pedestrian behavior.

This study aims to establish specific distinct pattern groups for the temporal variation of weekday pedestrian volumes applying cluster analysis in the central business district of Guadalupe in San José, Costa Rica.

Background

In a study related to the use of automatic counters to extrapolate volumes from manual counts, [9] argued that it is crucial to consider aspects like weather, hour, and location to establish hourly factors. These hourly factors can be used to expand short-term counts. In their study, they indicated the importance of characterizing these factors for future application. [4] applied three different methods to expand short-duration counts: assuming that there are no hourly variations in pedestrian traffic, applying temporal factors obtained from motorized traffic, and relating expansion factors from cities with similar characteristics. They concluded that applying expansion factors from other jurisdictions could generate more significant errors than the other analyzed methods.

Typically, cluster analysis has been used as a method to characterize temporal traffic patterns. For example, cluster analysis has different applications: to characterize truck flows [10]-[13] or traffic patterns [14], to classify roads according to their use [15], or to identify unusual patterns or non-recurrent events [16], [17]. Additionally, [18] characterized the stations of a public bicycle system according to entrance and exit. Also, they related their station classification with the geographic location of the stations. [19] applied clustering techniques to identify possible variables that could influence the pedestrian injury severity.

In Costa Rica, clustering analysis has been used to characterize temporal factors for motorized traffic [20]; in this study, routes are classified according to the temporal distribution of traffic. The methodology proposed by [20] has been adopted in this study for non-motorized traffic, using the data collected in a previous study [2]. On the other

hand, [21] found that commercial and service developments and the area of walkways available affect pedestrians' volume.

Study area, data collection, and counting sites

The study area comprises the Central Business District (CBD) of Guadalupe in the Municipality of Goicoechea, in the province of San Jose, Costa Rica. Guadalupe is one of the seven districts of Goicoechea. Statistics regarding Guadalupe and Goicoechea are in Table 1. This information can be helpful to evaluate the applicability of this research in similar cities outside Costa Rica. The district of Guadalupe was selected due to similarities to other urban areas in Costa Rica: a predominant commercial land use, its streets and avenues forming a grid, and the presence of sidewalks.

Table 1. Characteristics of the Canton of Goicoechea

and District of Guadalupe

Characteristic	Goicoechea	Guadalupe	Source
Area (square km)	31.65	2.39	[22]
Population	136,112	22,520	[23]
Percentage of urban population (2011)	98.5	-	[24]
Literacy rate (2011)	99.0	-	[24]
Percentage of population with a disability (2011)	11.7	-	[24]
Average number of years of education received by people aged 25–49 (2011)	10.5	-	[24]
Average number of years of education received by people aged 50 and older (2011)	8.7	-	[24]
Homicide rate per 10,000 habitants (2019)	-	0.4	[25]
Road traffic crashes with victims (injured or dead) (2017)	385	176	[26]

Source: Own elaboration

Additionally, [2] mentioned that in Guadalupe, several transit routes converge; therefore, the study area attracts and generates several pedestrian trips. Expressly, the study area is limited by avenues 27 and 35 and streets 39 and 67, the national routes

218 and 201 shown in Fig. 1. Route 218 has an Average Daily Traffic (ADT) of 36379 vehicles per day, and Route 201 has an ADT of 21,490 vehicles per day. Both ADT estimates are for the year 2015 [27].

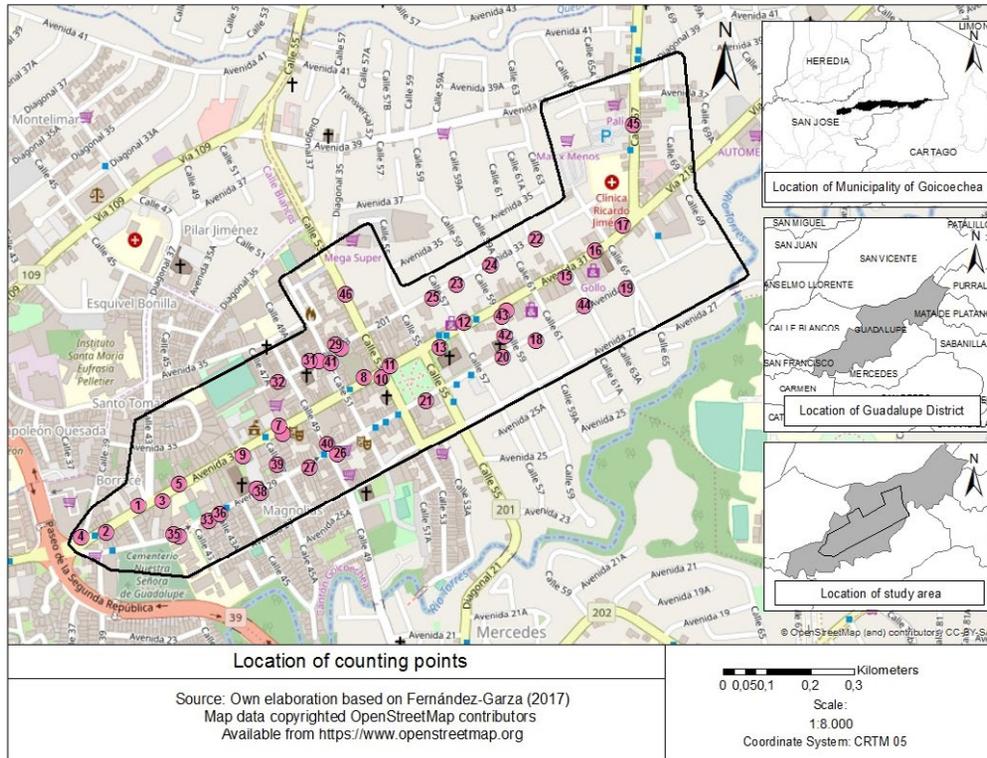


Fig. 1. Research area, Guadalupe Central Business.

Source: Own elaboration based on [32], [33]

Forty-seven bus routes cover Guadalupe–Moravia’s sector, most of them use National Route 218, and more than 200,000 passengers use these routes every day [28].

Materials and methods

The data were collected at 46 different sites, distributed as shown in Fig. 1. The number of sites was determined to have enough spatial coverage in the study area and the number of available sites to place the pedestrian counter. The automatic pedestrian counter must be attached to a pole or a traffic sign, limiting the number of available spots where the counter can be placed. Counts duration was between eight and sixty days between the end of April and October 2016. At one point, the pedestrian data were collected permanently during five months. Additionally, for verification purposes,

manual two-hour counts were performed in seven different points.

Regarding the pedestrian counts, at every counting site, an automatic pedestrian counter was placed. The counter has a passive infrared sensor with a high-precision lens. Generally, the sensor was attached to a utility post, over some time (usually one week), at the border of the sidewalk. The device registered pedestrian volumes at 15-minute intervals.

All 46 counting sites were included in the cluster analysis. The temporal variation of pedestrian traffic at every counting site was scrutinized. Three different sites presented unique characteristics and based on them, and some decisions were made:

The Counting site 16, next to a bus stop that connects different districts between Moravia and Desamparados, presented a different behavior during Mondays and Thursdays; therefore,

for analysis purposes, this site was divided as site 16.1 and site 16.2.

Similarly, site 45 was considered three different sites: 45.1 (May 23rd and 25th), 45.2 (May 24th and 26th), and 45.3 corresponding to May 29th, 30th, and 31st.

Site 22 presented two different patterns, so it was considered two different sites: 22.1 and 22.2.

For every counting site, the hourly factors, which are the proportion of the daily pedestrian traffic, were estimated using eq. 1:

$$FH_i = \frac{(VH_i)_p}{(VD)_p} \quad (1) \text{ Adapted from [27].}$$

Where:

FH_i : Hourly factor, for the hour “i” at counting site “p.”

$(VH_i)_p$: Pedestrian volume, at the hour “i” at the counting site “p.”

$(VD)_p$: Pedestrian daily traffic volume at the counting site “p.”

The cluster method proposed by [29] was implemented to hierarchical grouping the vectors of hourly factors from the different counting sites.

This method groups elements by minimizing the Euclidean distance among elements of the same group and maximizing the distances from other groups. The different elements are hierarchically grouped; the decision to join the elements to a specific group is defined by choosing the minimum Euclidean distance to a specific group. Once the element is included within a group, the Euclidean distances are recalculated, and the process is repeated until only one group, which contains all elements, remains [30]-[31].

This process can be visualized through a dendrogram, where each element is assigned to a group; they are grouped until only one is formed. The vertical axis in the graph represents the Euclidean distance; therefore, the shortest the length of the vertical lines, the closest the grouping, indicating that these are the most similar elements comprising them. To establish the practical number of pattern groups, an adaptation of the hybrid approach proposed by [13] was adopted.

Once the groups have been established through cluster analysis, this classification is related to the location of the different facilities, categorized by land use, that may be sources of pedestrian trips, for which the possible attractors are defined below.

Table 2. Land use categories applied in the study

Category	Description	Number of locations
Municipal and public services	Includes the Municipality, the entities of payments of public services, and the municipal services	7
Health services	A clinic, the red cross headquarters, and a hospital	3
Banks	All financial facilities in the area	14
Recreational	Parks	2
Supermarkets	Small size grocery and convenience stores are not included	7
Churches	Contains churches from different Christian denominations	16
Restaurants	This category includes all the restaurants and “sodas” of the place that open at daylight-hours	43
Night recreational	Includes restaurants, bars, and theaters that are open during the night.	25
Schools	Academic and artistic schools	8

Source: Own elaboration

This is an exploratory analysis because many facilities from different types are in the same block. This study attempts to identify the impact

of each location type on pedestrian traffic. Fig. 2 shows the different facilities considered in the present study.

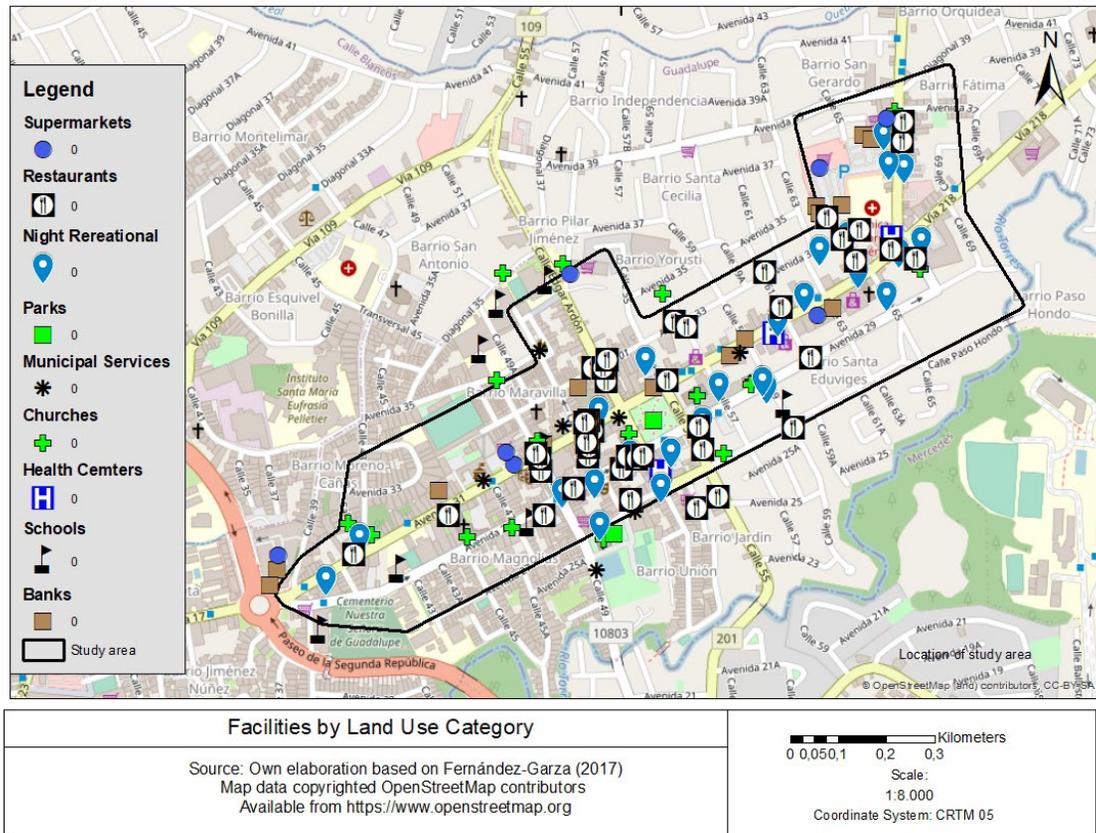


Fig. 2. Facilities in the study area by facility type.
 Source: Own elaboration

The different facilities were located on the map, and influence zones were established for each category at different distances (50, 75, 100, 125, and 150 m), and the counting sites found in each influence area were classified according to their group. The influence zone for each facility was established by the pedestrian travel path from the facility’s front door. Distances longer than 150 m were not considered due to an overlap between many of the counting points, and considering the size of the study area and distances shorter than 50 m would exclude many facilities from the analysis.

A hypothesis test was considered: There is a relationship between the pedestrian temporal

pattern groups and the land use in the surrounding areas.

A test of independence between the clusters and the different areas of influence is performed to prove this hypothesis.

One limitation of this study is that the interaction of different location types in the influence area of a site was not considered. Other aspects not considered in the present study are related to the width or accessibility (for example, presence of ramps for people with limited mobility) of the sidewalks in the study area; further details regarding the characteristics of the sidewalk can be found in [2].

Results

Cluster analysis

The cluster analysis was performed using R-Studio, and six different pedestrian pattern groups were obtained. Fig. 3 shows the dendrogram with the 46 counting sites on the horizontal axis. Additionally, based on the hybrid approach proposed by [13], pedestrian pattern groups C and D were divided into two groups for practical purposes:

- Pedestrian Pattern Group C1 includes sites 21, 26 and 32
- Pedestrian Pattern Group C2 includes sites 2, 5 and 45.2
- Pedestrian Pattern Group D1 includes sites 42 and 45.1
- Pedestrian Pattern Group D2 includes sites 7, 8, 9, 10, 11, 20, 23, 24, 25, 28, 29, 30, 31, 43, and 44.

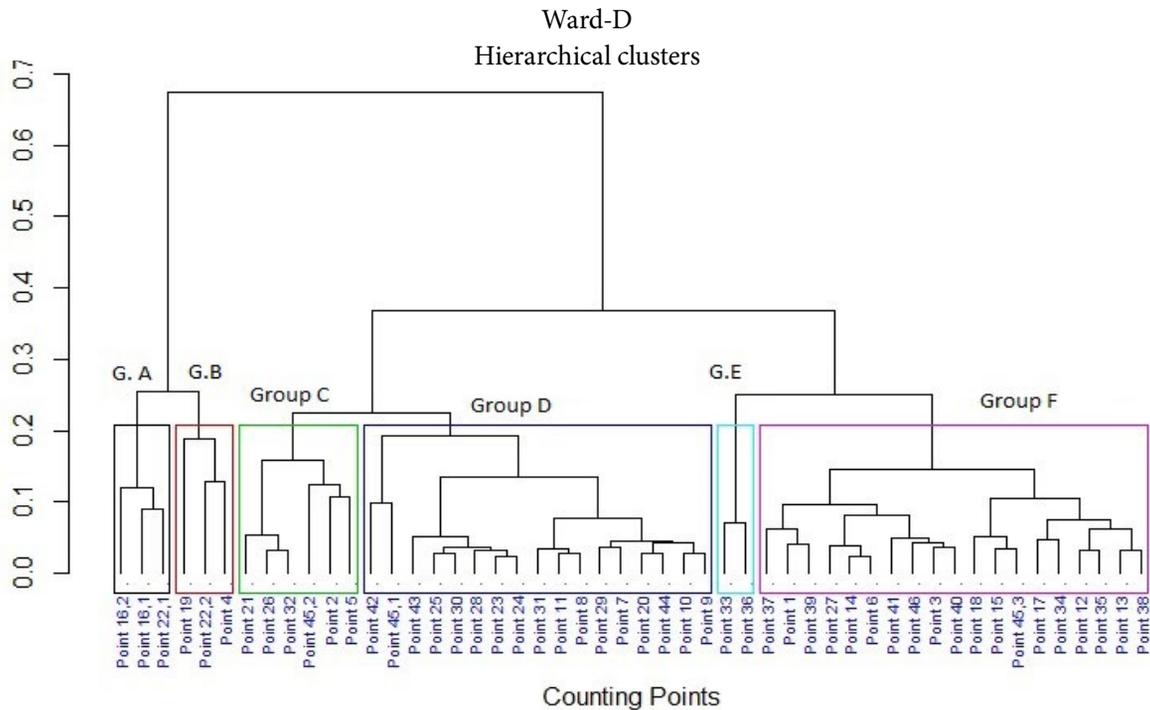


Fig. 3. Dendrogram from the cluster analysis.

Source: Own elaboration

Fig. 4 shows the spatial distribution of the different pattern groups. Groups D2 and F have 15 and 19 counting sites, respectively, comprising

more than two-thirds of the sites considered in the study. Groups A, C2, and E presented fewer counting sites on nearby locations.

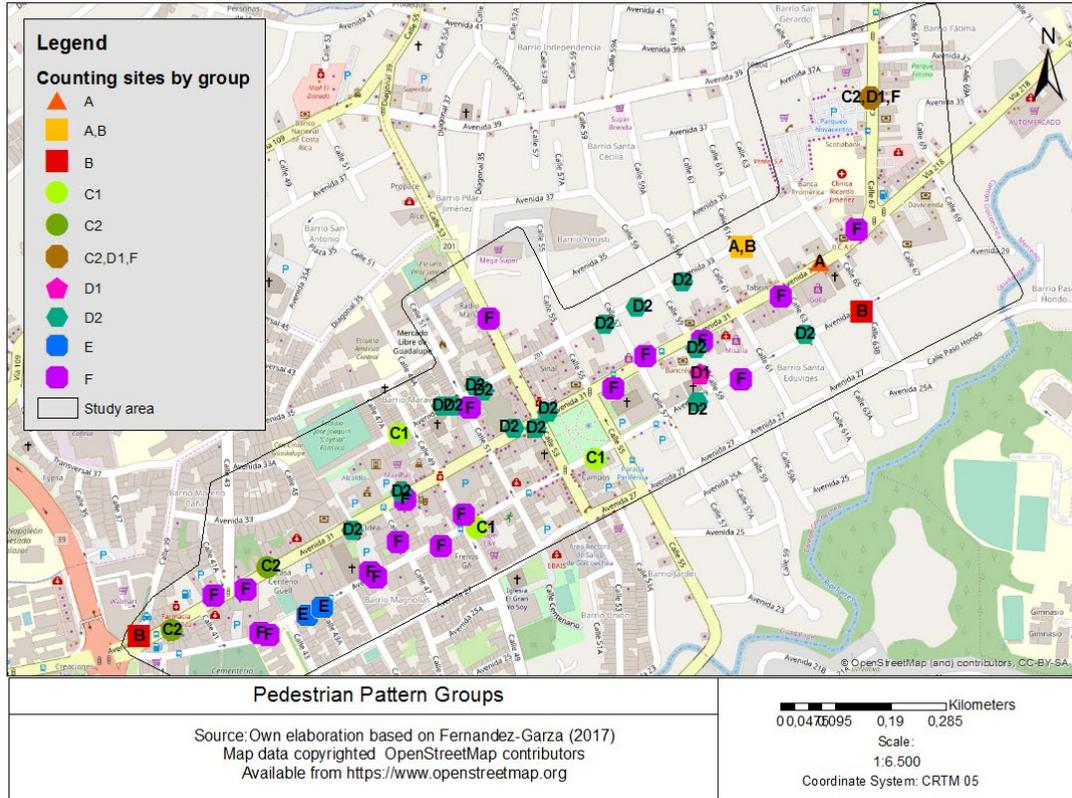


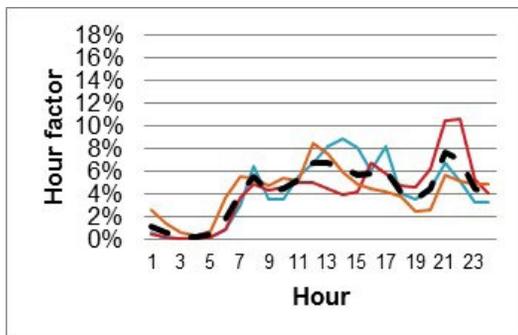
Fig. 4. Location of the counting sites by pedestrian pattern group assigned.

Source: Own elaboration based on [32], [33]

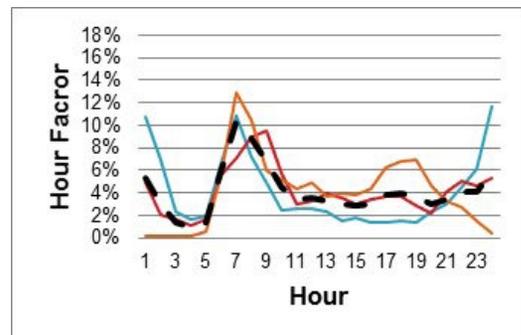
Fig. 5 depicts the eight patterns obtained. Every graph contains the hourly distribution for the counting sites in the group. Additionally, the

dashed lines represent the average value for each group. Table 3 shows the average hourly factors obtained for each group.

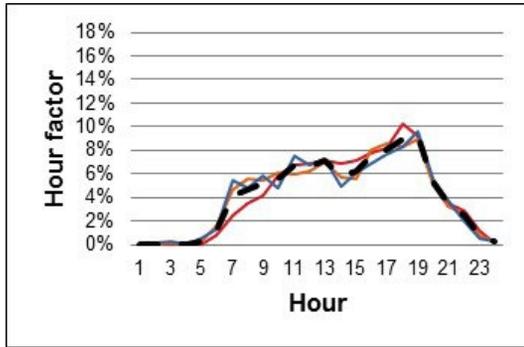
(a) Group A



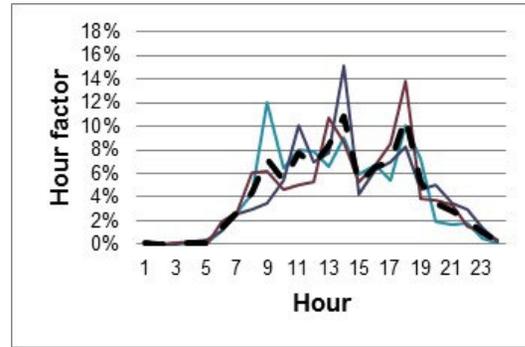
(b) Group B



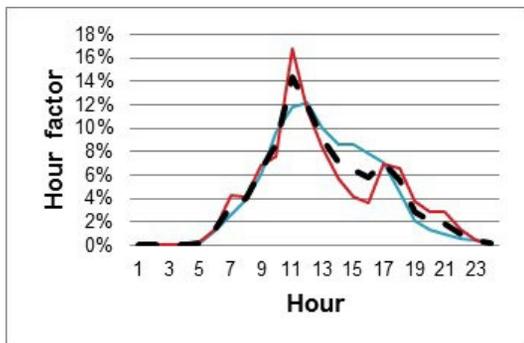
(c) Group C1



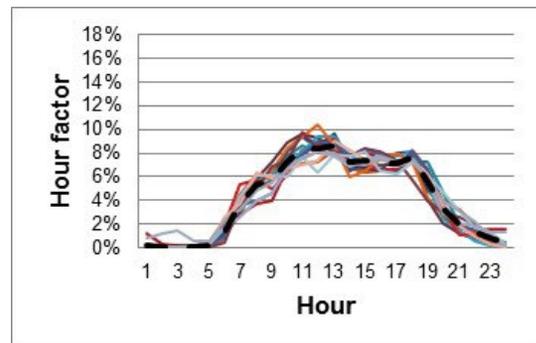
(d) Group C2



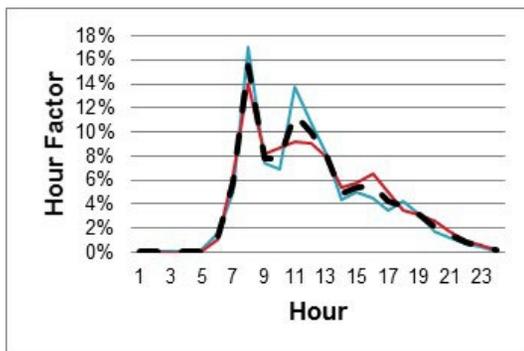
(e) Group D1



(f) Group D2



(g) Group E



(h) Group F

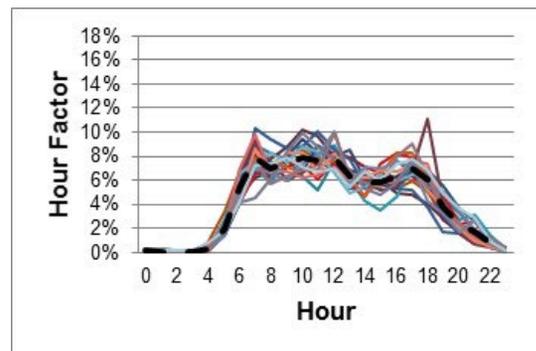


Fig. 5. Pedestrian pattern groups. *Note.* The dashed line shows the average values.

Source: Own elaboration

Table 3. Average hourly factors, in percentage, for each pedestrian pattern group

Hour	Group A	Group B	Group C1	Group C2	Group D1	Group D2	Group E	Group F
0	1.15	5.26	0.04	0.14	0.07	0.25	0.05	0.12
1	0.53	3.07	0.05	0.03	0.04	0.15	0.06	0.10
2	0.20	1.38	0.11	0.11	0.02	0.15	0.03	0.06
3	0.13	0.99	0.02	0.16	0.07	0.09	0.05	0.07
4	0.38	1.40	0.25	0.18	0.25	0.25	0.12	0.36
5	1.81	6.03	1.22	1.40	1.31	1.41	1.37	1.96
6	4.06	10.32	4.20	2.67	3.41	3.76	5.71	5.46
7	5.56	8.87	4.62	4.43	4.01	5.32	15.50	7.85
8	4.19	6.78	5.15	7.25	6.44	5.79	7.80	7.02
9	4.49	4.44	5.54	5.47	8.60	7.29	7.80	7.45
10	5.19	3.36	6.76	7.71	14.31	8.34	11.43	7.83
11	6.71	3.58	6.62	6.71	11.91	8.41	9.91	7.67
12	6.72	3.31	7.13	8.39	9.18	8.64	8.00	7.90
13	6.26	3.02	5.84	10.91	7.13	7.23	4.89	6.31
14	5.69	2.84	6.23	5.16	6.39	7.32	5.40	5.86
15	5.74	3.05	7.57	6.58	5.76	7.18	5.52	5.87
16	6.02	3.78	8.10	6.97	7.03	7.09	4.29	6.39
17	4.19	4.02	8.97	10.77	5.53	7.60	3.86	7.02
18	3.55	3.74	9.26	5.25	2.93	5.67	3.12	6.12
19	4.38	3.06	5.36	3.53	2.13	3.36	2.14	3.74
20	7.60	3.44	3.50	2.83	1.88	2.03	1.47	2.32
21	6.94	4.13	2.44	2.10	0.97	1.40	0.85	1.70
22	4.48	4.14	0.82	0.98	0.45	0.81	0.48	0.76
23	4.01	5.81	0.20	0.26	0.17	0.44	0.15	0.25

Source: Own elaboration

The following paragraphs describe the temporal variations of pedestrian traffic for each group, and some pictures are added to provide a context of some sites; the analysis of location categories and pedestrian flow patterns is explained further below.

Group A: Sites 16.1, 16.2, and 22.1. This group does not present well-defined peak hours. Even

though the counting site 16 was separated in two, the cluster analysis joined them due to the pattern differences between Mondays and Thursdays because these differences are less significant than the differences among other sites. There is a variation of traffic, around four percent of total daily traffic, between 8 am and 11 pm (Fig. 6).



Fig. 6. Avenida 33 between Calles 63 and 65, Guadalupe, Goicoechea.

Source: Taken by the authors

Group B: Sites 4, 19, and 22.2. All counting sites in the group presented a very high morning peak between 7 and 9 am. On the other hand, site 19 presents high pedestrian percentages late at night. This behavior might be explained by the location of a nearby bar with night activity.

Group C1: Sites 21, 26, and 32. These counting sites presented a very similar hourly distribution. Between 1 and 5 am, the pedestrian volumes are

practically null. After 5 am, there is a sharp increase in pedestrian traffic between five and seven. The number of pedestrians increases as the time of the day increases until 7 pm. The counting sites 21 and 26 are close to bus stops for busses coming from the city center that could explain the peak at 7 pm, due to workers commuting back to their homes (Fig. 7).



Fig. 7. Northwestern corner of Parque de Guadalupe, Goicoechea. San José, Costa Rica.

Source: Taken by the author

Group C2: Sites 2, 5, and 45.2. The hourly distribution of traffic shows short peaks. Two prominent peaks are predominant, one at 2 pm and another at 6 pm. These peaks could be explained by the location of a commercial plaza in front of site 45.

Group D1: Sites 42 and 45.1. These two sites presented a very similar hourly distribution of pedestrian traffic with a very high peak at eleven in the morning; however, they are in different locations, and their average weekday pedestrian traffic differs (705 and 2411 pedestrians for sites 42 and 45, respectively).

Group D2: Sites 7, 8, 9, 10, 11, 20, 23, 24, 25, 28, 29, 30, 31, 43, and 44. The hourly distribution for all counting sites is very similar, and they practically follow the average distribution from the group. This hourly distribution presents a constant increase in pedestrian traffic between 6 am until a peak is reached at 11 am, then traffic slowly decreases until 6 pm. After six, there is a sharp decrease in the percentage of daily traffic until 10 pm. Most sites are located on avenues 31 and 33. These sites also present a wide range of daily traffic, from 276 pedestrians per day, on-site 44, to 5215 pedestrians per day, on-site 7 (Fig. 8).



Fig. 8. Avenida 31-Calle 55 intersection, Guadalupe, Goicoechea.

Source: Taken by the authors

Group E: Sites 33 and 36. These two sites practically do not present pedestrian volumes between 1, and 5 am. At 6 am, pedestrian traffic increases at these sites, with a very high peak around 8 am (approximately 16 % of the daily

traffic) and another peak at 11 am. Nearby, the Fernando Centeno Güell School, an educational center for people with disabilities, may influence the hourly variation of pedestrian traffic at these locations (Fig. 9).



Fig. 9. Avenida 29-Calle 43 intersection, Guadalupe, Goicoechea.

Source: Taken by the authors

Group F: Sites 1, 3, 6, 12, 13, 14, 15, 17, 18, 27, 34, 35, 37, 38, 39, 40, 41, 45.3, and 46. Group 6 includes the biggest number of counting locations. There are no abrupt changes in the hourly factors over the day. The pedestrian traffic is practically null between midnight and 4 am; then, the traffic

increases until it reaches a plateau between 7 am and 5 pm when the pedestrian traffic drops. The daily traffic significantly varies among the counting sites from 353 pedestrians per day to 8067 pedestrians per day, at counting sites 34 and 1 respectively (Fig. 10).



Fig. 10. Avenida 31-Calle 61 intersection, Guadalupe, Goicoechea.

Source: Taken by the authors

Relation between groups and land use

Because some groups have very few counting points assigned, it was decided to redefine the pattern groups as D2, F, and others (others include groups A, B, C1, C2, D1, E) to have a more robust hypothesis test. Once this new notation is established, Fisher's exact hypothesis test is performed to analyze the independence of the data.

It also includes an analysis of the average weekday traffic (TPD ES) and its relationship with the established clusters groups, and all points are included.

Relationship between the influence of each facility type and the land use categories for each group

In this approach, the eight groups found are maintained through differences in time factors; for six,

it is impossible to observe significant relationships between zones of influence and the different land uses considered since they have very few points. It is expected that at a greater distance, the number of points per group that the different establishments increase influence; this is only true for groups D2 and F, which have a more significant number of points.

Fig. 11 shows the result obtained for group D2. It is observed that the establishments that influence a higher percentage of the points are in the upper part of the graph in blue tones, that is: banks, restaurants, public services (municipal), bus stops, night recreation and churches and in red the categories that have less influence for this group (schools, parks, and supermarkets).

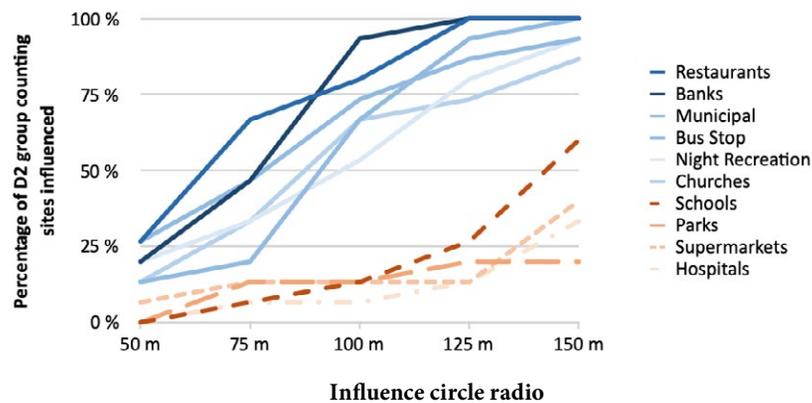


Fig. 11. Percentage of pedestrian pattern group D2's counting sites influenced by different land uses.

Source: Own elaboration

This behavior found for group D2 is repeated in all groups, considering that the land use is mixed and the study area is small.

Relationship between the influence circle radio and the clusters groups for each establishment

As mentioned previously, it was required to join some of the eight groups established through clusters at this study stage to achieve the independence test.

For each land use assigned to each location, the relationship between the distances from the other land developments and the number of counting sites of each category that enter each influence circle is obtained, and a Fisher's exact independence test is performed with a significance of $\alpha = 0.05$.

According to the results shown in Table 4, only for the case of banks and public services, the null hypothesis is rejected; that is, the established groups depend on the distance they are from the banks and public services.

Table 4. Fisher’s exact independence test results

Case	p-Value	Case	p-Value
Banks	0.0011	Hospitals	0.3439
Public Services	0.0052	Supermarkets	0.4506
Schools	0.1136	Parks	0.5632
Night Recreation	0.2117	Restaurants	0.5497
Bus Stop	0.2702	Churches	0.8180

Source: Own elaboration

As shown in Figs. 12 and 13 for both banks and public services, the percentage of points influenced in each group increases concerning

distance. In both cases, group D2 is the most influenced by them.

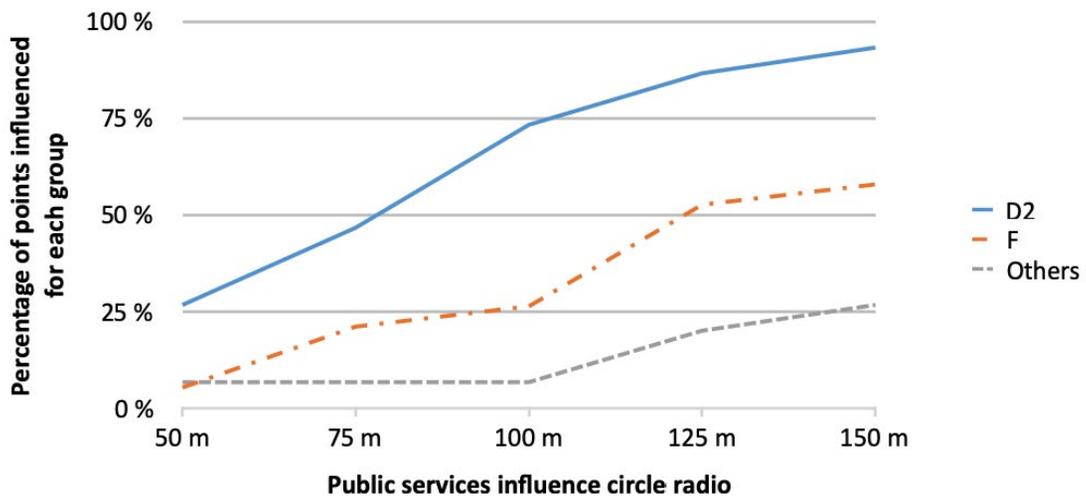


Fig. 12. Percentage of counting sites influenced by public services

Source: Own elaboration

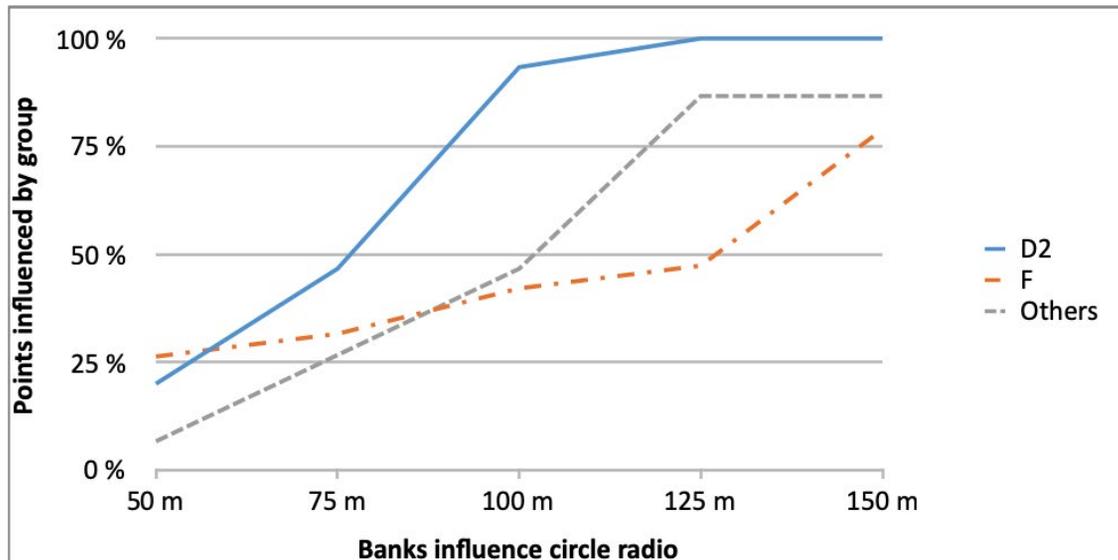


Fig. 13. Percentage of counting sites influenced by banks

Source: Own elaboration

Conclusions

The cluster analysis proposed in this study was appropriate to identify sites with similar temporal distributions of pedestrian traffic. However, this method does not explain the different temporal patterns obtained.

Therefore, a test was performed to relate the groups obtained with the land use in the surrounding areas; however, no relationship between them was found. It appears that the groups have a mixed influence from the different land uses. Further research is required because only the influence of banks and public services is suggested for Group D2.

This study shows how the temporal distribution of pedestrian traffic could vary significantly even in the same CBD. Nonetheless, two groups dominated the study area: Group F with 19 counting sites and Group D2 with 15 sites. These two groups contain more than two-thirds of the total sites included in the study. The variation of pedestrian traffic is similar for both groups. Based on their variations, the pedestrian volume is null, or

very low, between midnight and five in the morning in the study area.

Additionally, the peak pedestrian traffic is around midday. After this period of the day, the number of pedestrians slowly decreases until 6 pm. After sunset, which is around 6 pm, the pedestrian traffic decreases significantly.

A limitation in the study is the number of counting sites, and these are not distributed uniformly as not all sidewalks within the study area have the necessary characteristics to place counters properly. Due to the implementation of exclusive transit lanes during the peak periods in May 2019, additional research is necessary to determine the effect of this measure on pedestrian behavior.

Additional research is recommended to determine explicative variables for the differences in the patterns found. For example, other studies have found relationships between weather and pedestrian behavior [1], [34] or the effect of urban density in walking activity [35]. The diversity of the patterns found indicates the need for further research and analysis better to understand a complex phenomenon such as pedestrian mobility. The

interaction of different variables like transit facilities, services, and the weather should be included in future studies. Extreme precaution is recommended in using expansion factors for short-term counts; due to the heterogeneity of patterns based on the collected data.

This project was developed at LanammeUCR as part of the activities related to Law 8114, as amended.

References

- [1] C. Bongiorno, D. Santucci, F. Kon, P. Santi, and C. Ratti. "Comparing bicycling and pedestrian mobility: Patterns of non-motorized human mobility in Greater Boston," *J. Transp. Geogr.*, vol. 80, p. 102501, 2019, doi: <https://doi.org/10.1016/j.jtrangeo.2019.102501>
- [2] A. G. Fernández-Garza, H. Hernández-Vega. "Estudio peatonal en un centro urbano: un caso en Costa Rica," *Rev. Geogr. Am. Cent.*, vol. 1, no. 62, pp. 267-300, 2018, doi: <https://doi.org/10.15359/rgac.62-1.10>
- [3] US Department of Transportation. *Traffic Monitoring Guide*, 2016. Available: https://www.fhwa.dot.gov/policyinformation/tmguidetmg_fhwa_pl_17_003.pdf
- [4] C. Milligan, R. Poapst, and J. Montufar. "Performance measures and input uncertainty for pedestrian crossing exposure estimates," *Accid. Anal. Prev.*, vol. 50, pp. 490-498, 2013, doi: <https://doi.org/10.1016/j.aap.2012.05.024>
- [5] P. Ryus, E. Ferguson, K. Laustsen, R. Schneider, F. Proulx, T. Hull, et al. *Guidebook on Pedestrian and Bicycle Volume Data Collection*, 2014, doi: <https://doi.org/10.17226/22223>.
- [6] P. Rietveld. "Biking and walking: the position of non-motorized transport modes in transport systems," in *Handbook of transport systems and traffic control*, 2001, pp. 299-319, doi: <https://doi.org/10.1108/9781615832460-019>
- [7] D. Johnstone, K. Nordback, and S. Kothuri. "Annual average non-motorized traffic estimates from manual counts: quantifying error," *Transp. Res. Rec.*, vol. 2672, no. 43, pp. 134-144, 2018, doi: <https://doi.org/10.1177%2F0361198118792338>
- [8] M. El Esawey, C. Lim, T. Sayed, and A. I. Mosa. "Development of daily adjustment factors for bicycle traffic," *J. Transp. Eng.*, vol. 139, no. 8, pp. 859-871, 2013, doi: [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000565](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000565)
- [9] R. Schneider, L. Arnold, and D. Ragland. "Methodology for counting pedestrians at intersections: use of automated counters to extrapolate weekly volumes from short manual counts," *Transp. Res. Rec.*, vol. 2140, no. 1, pp. 1-12, 2009, doi: <https://doi.org/10.3141/2140-01>
- [10] T. Papagiannakis, M. Bracher, and N. Jackson. "Utilizing cluster techniques in estimating traffic data input for pavement design," *J. Transp. Eng.*, vol. 132, no. 11, pp. 872-879, 2006, doi: [https://doi.org/10.1061/\(ASCE\)0733-947X\(2006\)132:11\(872\)](https://doi.org/10.1061/(ASCE)0733-947X(2006)132:11(872))
- [11] F. Sayyady, J. Stone, K. Taylor, F. Jadoun, and R. Kim. "Clustering analysis to characterize mechanistic-empirical pavement design traffic data in North Carolina," *Transp. Res. Rec.*, vol. 2160, no. 1, pp. 118-127, 2010, doi: <https://doi.org/10.3141/2160-13>
- [12] J. Regehr. "Understanding and anticipating truck fleet mix characteristics for mechanical-empirical pavement design," presented at Transportation Research Board 2011 Annu. Meeting. Washington DC: Transportation Research Board, 2011.
- [13] M. Reimer and J. Regehr. "A hybrid approach for clustering vehicle classification data to support regional implementation of the mechanistic-empirical pavement design guide," *Transp. Res. Rec.*, vol. 2339, no. 1, pp. 112-119, 2012, doi: <https://doi.org/10.3141/2339-13>
- [14] E. van Berkum and W. Weijermars. "Analyzing highway flow patterns using cluster analysis," presented at 8th Int. IEEE Conf. Intelligent Transportation Systems, Vienna, Austria, 2005.
- [15] J. Wyatt and S. Sharma. "Classification of Saskatchewan highways according to type of road use," *Can. J. Civ. Eng.*, vol. 13, no. 1, pp. 53-58, 1986, doi: <https://doi.org/10.1139/l86-008>
- [16] F. Soriguera and D. Rosas. "Deriving Traffic Demand Patterns from Historical Data" presented at Transportation Research Board 2012 Annu. Meeting,

- Washington DC: Transportation Research Board, 2011, DOI: [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000456](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000456)
- [17] W. Weijermars. "Analysis of urban traffic patterns using clustering," Ph.D. dissertation. Dept. Civ. Eng., Fac. Eng. Technol., Univ. Twente, 2007.
- [18] P. Vogel, T. Greiser, and D. C. Mattfeld. "Understanding bike-sharing systems using data mining: Exploring activity patterns," *Procedia Soc. Behav. Sci.*, vol. 20, pp. 514-523, 2011, DOI: <https://doi.org/10.1016/j.sbspro.2011.08.058>
- [19] M. G. Mohamed, N. Saunier, L. F. Miranda-Moreno, and S. V. Ukkusuri. "A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal, Canada," *Saf. Sci.*, vol. 54, pp. 27-37, 2013, DOI: <https://doi.org/10.1016/j.ssci.2012.11.001>
- [20] J. Magaña-Cubillo, H. Hernández-Vega, and D. Jiménez-Romero. "Aplicación de análisis de conglomerados a para la caracterización de factores temporales de tránsito para Costa Rica" presented at Congr. Ingeniería Civil, Reto del desarrollo de infraestructura y servicios, San José, Costa Rica, 2014.
- [21] B. Pushkarev and J. M. Zupan. *Pedestrian travel demand*, 1971. Available: <https://onlinepubs.trb.org/Onlinepubs/hrr/1971/355/355-004.pdf>
- [22] Instituto Tecnológico de Costa Rica. *Atlas de Costa Rica 2014*, Cartago, Costa Rica, 2016.
- [23] Instituto Nacional de Estadísticas y Censos. *Estadísticas vitales 2018: población, nacimientos, defunciones, matrimonios*. Dirección General de Estadística y Censos, 2019. Available: https://www.inec.cr/sites/default/files/documentos-biblioteca-virtual/repoblacv2018_0.pdf
- [24] Programa Estado de la Nación en Desarrollo Humano Sostenible. *Indicadores Cantonales*, 2013. Available: <https://www.inec.cr/sites/default/files/documentos/poblacion/estadisticas/resultados/repoblacenso2011-01.pdf.pdf>
- [25] Ministerio de Justicia. *Anexo estadístico. Atlas de ocurrencia de delitos 2019*, 2020. Available: <http://observatorio.mj.go.cr/recurso/anexo-estadistico-atlas-de-ocurrencia-de-delitos-2019>
- [26] Cosevi. *Anuario Estadístico de accidentes de tránsito con víctimas en Costa Rica*, 2019. Available: <https://www.csv.go.cr/documents/20126/50694/Anuario+estad%C3%ADstico+de+accidentes+de+tr%C3%A1nsito+con+v%C3%ADctimas+Costa+Rica+2017.pdf/dfcf2e128-2660-517b-c360-7cc5cfa5cd2a?t=1574094470460>
- [27] Mopt. *Anuario Información del tráfico 2018*, 2018. Available: <https://www.mopt.go.cr/wps/wcm/connect/f9d4084d-6330-4c21-b947-61dabc81cdfd/AnuarioTransito2018.pdf?MOD=AJPERES>
- [28] Epyypsa – Siguma GP. *Apoyo al modelo general de sectorización de transporte público San Jose, Costa Rica*, 2014.
- [29] J. H. Ward. "Hierarchical Grouping to Optimize an Objective Function," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 236-244, 1963, DOI: <https://doi.org/10.1080/01621459.1963.10500845>
- [30] O. Hernández-Rodríguez. *Temas de análisis estadístico multivariante*, San José: Universidad de Costa Rica, 2013.
- [31] J. Trejos-Zelaya, W. Castillo-Elizondo, and J. González-Varela. *Análisis Multivariado de Datos Métodos y Aplicaciones*, San José: Universidad de Costa Rica, 2014.
- [32] A. G. Fernández-Garza. "Análisis de la movilidad peatonal y caracterización de peatones en el centro de Guadalupe como caso de estudio y aplicación," B.S. thesis, Fac. Eng., Univ. Costa Rica, 2017.
- [33] OpenStreetMap contributors. *Planet dump* [Data file], 2015. Available: <https://planet.openstreetmap.org>.
- [34] A. P. Vanky, S. K. Verma, T. K. Courtney, P. Santi, and C. Ratti. "Effect of weather on pedestrian trip count and duration: City-scale evaluations using mobile phone application data," *Prev. Med. Rep.*, vol. 8, pp. 30-37, 2017, DOI: <https://doi.org/10.1016/j.pmedr.2017.07.002>
- [35] A. Forsyth, J. M. Oakes, B. Lee, and K. H. Schmitz. "The built environment, walking, and physical activity: Is the environment more important to some people than others?" *Transp. Res. Part D: Transp. Environ.*, vol. 14, no. 1, pp. 42-49, 2009, DOI: <https://doi.org/10.1016/j.trd.2008.10.003>