Titre: Title:	Gender differences in urban recreational running: a data-driven approach
Auteurs: Authors:	Grant McKenzie, Daniel Romm, Clara Féré, & Maria Laura Guerrero Balarezo
Date:	2025
Type:	Article de revue / Article
Référence: Citation:	McKenzie, G., Romm, D., Féré, C., & Guerrero Balarezo, M. L. (2025). Gender differences in urban recreational running: a data-driven approach. Journal of Transport Geography, 124, 104171 (11 pages). https://doi.org/10.1016/j.jtrangeo.2025.104171

Journal of Transport Geography 124 (2025) 104171

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Gender differences in urban recreational running: A data-driven approach

Grant McKenzie [a,*], Daniel Romm [a], Clara Féré [a], Maria Laura Guerrero Balarezo [b]

[a] Platial Analysis Lab, McGill University, Montréal, Canada
[b] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal and CIRRELT, Montréal, Canada

A R T I C L E I N F O

Keywords:
Running
Gender
Exercise
Spatial analysis

A B S T R A C T

Exploring the dynamics of urban recreational running, this study examines the spatial and temporal patterns of running activities among men and women in two major North American cities, Montréal, Canada and Washington, DC, USA. A total of 20,446 running trajectories from a geosocial fitness tracking application were analyzed, revealing significant gender differences. These gender preferences differ in terms of location and time, highlighting significant variations between the two cities and shifts between day and night running habits. We further investigate the influence of socio-economic, demographic, and built environment factors on these different spatiotemporal patterns. Regression models show that proximity to bike lanes and parks strongly influenced running locations in both cities, with a preference for lower population density and lower median household income areas. Insights from this work are important for urban planners and public health officials, providing a data-driven foundation for developing more inclusive and safe public spaces for recreational activities. The study not only contributes to our understanding of urban recreational behaviors but also addresses broader societal concerns about gender and public space utilization.

1. Introduction

Recreational running, or jogging, ranks among the world's most popular fitness activities. In the United States and Canada, it is estimated that over 15 % of individuals aged 15 and older engage in running (Running USA, 2022; Statistics Canada, 2023). Similarly, a global Nielsen (2021) reveals that around 12 % of people in developed countries run at least once a week. Running is also notable for its relatively balanced gender participation. In Canada, for instance, women make up 57 % of runners (Statistics Canada, 2023), while in the United Kingdom, they account for 51 % (Sport England, 2016), and in the United States, 36 % (Running USA, 2022).

Despite these statistics, societal narratives often highlight significant gender-based differences in the patterns of running, specifically regarding the time and location of the activity. For instance, women are less likely to run at night than men (Visontay, 2024). Various factors have been proposed to explain these disparities, ranging from physiological limitations (Joyner, 1993) to concerns about physical safety. Indeed, substantial research corroborates the latter, particularly in relation to women's perception of physical safety (Roper, 2016; Butryn and Furst, 2003). Understanding the differences in when and where women and men run is important for designing cities that value equitable access to public space as well as for promoting physical activity, active transportation, and encouraging health and well-being.

The last two decades have seen a surge in the popularity of geosocial fitness tracking applications, such as *MapMyFitness* and *Strava*, coinciding with the widespread adoption of wearable fitness trackers equipped with location sensors. Strava, the data source for this study, in particular has gained traction. It provides a platform for users to upload their fitness activities, predominantly running and cycling, from their tracking devices and compare their performance with others. The application promotes competition by creating segments along roads or trails, where athletes vie for top spots on leader boards or titles like *Queen/King of the Mountain*. As of 2022, Strava claims more than 7 billion uploaded activities from over 100 million users across 195 countries (Strava, 2022). While the population that uses Strava does not represent younger or older portions of the general public, it does represent a wide swath of athletes from novice and recreational runners to elite athletes (Venter et al., 2023). For researchers, Strava's API offers an unparalleled opportunity to access and analyze this wealth of data.

This study aims to uncover spatial and temporal gender differences in recreational running, along with the contributing factors, by analyzing running patterns in Washington, District of Columbia in the United States of America, and Montréal in Québec, Canada. It explores

* Corresponding author.
E-mail address: grant.mckenzie@mcgill.ca (G. McKenzie).

https://doi.org/10.1016/j.jtrangeo.2025.104171
Received 25 July 2024; Received in revised form 13 February 2025; Accepted 24 February 2025
Available online 4 March 2025
0966-6923/

how these patterns vary throughout the day and differ between genders. Specifically, the study will address four research questions (RQs) focused on these aspects:

RQ1 *Does a preference for running locations differ based on gender?* To address this question, we access running trajectories from a social fitness tracking application, split them based on the self-identified gender of the athlete, and assess the dominance of activity by region.

RQ2 *Does a gender difference in running locations vary between cities?* We investigate gender-based differences in running locations between two major cities in North America, namely Montréal, Canada, and Washington D.C., USA.

RQ3 *How does the gender difference in running locations vary temporally?* We assess gender-based differences in running locations by hour of the day, and day vs. night to assess the effect of time on the geographical difference in gender-based running preferences.

RQ4 *What factors influence differences in gender-based running location preferences and does the importance of these factors vary by region and temporally?* To address this question we construct a series of regression models using socio-economic, demographic, and built-environment data as explanatory variables.

2. Related work

Academic research has long focused on the spatial and temporal dimensions of recreational running. The majority of this research has traditionally used qualitative methods (Qviström et al., 2020; Ettema, 2016; Parks et al., 2003). However, a growing trend is the use of data-driven approaches in this field (Lee and Sener, 2021; Best and Braun, 2017; Yang et al., 2023). Despite the increasing volume of studies, there remains a gap in research specifically addressing gender differences in running, particularly from a data-driven perspective.

A recent study by Battiston et al. (2023) on gender inequality in urban cycling shares similarities with our approach, analyzing Strava activities across various cities. Although their focus is primarily at the city level, they also delve into the gender-cycling gap on a street level in New York City. Their findings suggest that women would benefit more than men from enhanced cycling infrastructure, such as bike paths, in cities. This aligns with other studies that highlight the gender-cycling gap and report that women are more likely to walk or use public transport for commuting than men, but less likely to cycle (Goel et al., 2023). Zhong et al. (2022) also identified gender differences in active mobility, but in this case urban footpath selection for pedestrian travel. They found that footpath selection was influenced by factors like population and arterial road density. However, their study was limited to a narrow range of built environment factors within a single city. While there have been studies exploring gender differences in walking, Pollard and Wagnild (2017) argue that walking and other physical activities are not comparable to running. Runners, they suggest, are more sensitive to disruptions due to their focus on maintaining speed and stability. This highlights a unique aspect of running that distinguishes it from other forms of physical activity.

Although research on gender differences in running route selection is limited, existing studies highlight significant disparities (Schuurman et al., 2021). As a public activity, urban running exposes runners to environmental influences. Rana (2022) points out that public spaces are often gendered in design and perceived as predominantly male domains, compelling women to carefully consider their presence in these spaces. Decisions about where, when, and how to run are often dominated by safety concerns (Roper, 2016). Runners may face harassment and may alter their routes to avoid high-risk areas (Hockey and Collinson, 2006; Roper, 2016). Polko and Kimic (2022) find that security-related factors play a more important role for women, particularly in urban parks. Factors like lighting, visibility, seclusion, and park maintenance impact women's sense of safety more than men's, making some parks less attractive for women runners, especially at night. Ghani et al. (2019) observes that neighborhood environments influence walking habits differently across genders, with women being more responsive to aspects like the built environment, safety, and cleanliness. Running-specific research aligns with these findings, emphasizing the importance of lighting, clear sight-lines, and the presence of others, particularly fellow runners and women, for women runners to feel safe (Schuurman et al., 2021; Roper, 2016). The threat of violence, including harassment or assault, heavily influences decisions around running (Collinson, 2008; Rana, 2022). Additionally, gender roles, such as family obligations and childcare, can also impact running habits (Qviström et al., 2020).

These insights provide an overview of how recreational activities, particularly those in the public domain, are substantially different experiences depending on gender. The spatiotemporal aspect of these activities, seen in route preferences, may be especially sensitive to gender-based differences. This underscores the need for more comprehensive research in this area, particularly studies supported by robust data and quantitative analysis.

3. Study area & data

3.1. Study area

Two cities in North America were chosen for comparative analysis, namely Montréal, Québec in Canada (MTL) and Washington, District of Columbia in the United States of America (DC). Both are large cities with populations of 4.3 million people and 6.4 million people in their greater metropolitan areas, respectively (2022). Further reasons for their selection include the fact that they are in different countries, with differing political and social make-ups and both with diverse populations ranging in age, race, and income. Despite the large populations in the greater regions of the cities, they do differ in a number of ways. For instance, the Washington metropolitan area has a population density of 972 people per mile (U.S. Census Bureau, 2020) whereas the Montréal metropolitan area has a population density of just 355 people per mile (Statistics Canada, 2022). DC also has a more racially diverse population with a larger income disparity (Busette and Elizondo, 2022) and a unique East-West racial divide between the black and white population (Logan, 2017). The availability of a large amount of *Strava* data for both of these regions also allowed for analysis of gender differences.

3.2. Data

Running trajectories were accessed from Strava between October 2011 and October 2022 using the Strava application programming interface (API).[1] Our data collection strategy involved finding the top 10 *segments*[2] in DC and MTL and then identifying the most recent activities that traversed the segment. Through the API and related data collection scripts, a total of 33,482 running trajectories, along with metadata, were accessed. The actual dates of the activities range between December 2005 and October 2022. These data included the GPS fixes for each trajectory along with the start time and date. The gender of the runner is self-reported on the Strava application with users asked to select from *Man*, *Woman*, *Prefer not to say*, or *Non-binary*. We restricted our analysis to only those trajectories where athletes self-reported a gender of *Man* or *Woman*. Notably, during the data collection period of this project, Strava's search functionality limited search by gender to either women or men. The data were cleaned to remove invalid or unrealistic trajectories[3] and limit trajectories to one per athlete. We set this limit so that a single athlete did not skew our analysis. After cleaning, a total of 20,446 activities were used for further analysis. A breakdown of these activities

[1] https://developers.strava.com/docs/reference/

[2] Parts of a road or trail over which athletes compete for fastest travel time.

[3] We removed activities with unrealistic speeds (e.g., faster than 3:40 min mile), where consecutive GPS fixes spanned more than 100 m, and activities that lasted more than 6 h.

by region and gender are presented in Table 1 along with average distance and duration.

In examining the factors that contribute to spatial and temporal running preferences between men and women, we accessed a range of socio-economic, demographic, and built-environment variables. An overview of these data are provided in Table 2 with full details on data sets used at [repository blinded for review]. These independent variables were chosen based on existing literature and what we perceived to be important factors that might contribute to someone choosing to run in a location, as well as the availability of data in both DC and MTL. The socio-economic and demographic data were accessed from either the Canadian or American censuses at the level of dissemination area or block group, respectively. Since the geographic regions of reporting differed from the hexagonal regions used in our analysis, we applied areal interpolation[4] to reassign values to the hexagon geometries. Built environment data were collected from each city's open data web portals. Finally, prior to analysis, we normalized all data to ensure that values were constrained between 0 and 1. This was done to allow for comparison between cities, as the magnitude of raw values varied between cities in some cases.

4. Methodology

In this section, we outline the methodological approach used to address the research questions introduced in the introduction (Fig. 1). The process begins with data pre-processing, during which the data is aggregated into a standardized set of hexagonal cells. Following this, we conduct three types of analysis. First, we perform a spatial analysis to examine regional gender differences, calculating the percentage distribution of genders across regions and conducting a spatial autocorrelation analysis to assess patterns of clustering. Second, we analyze running patterns temporally to uncover time-based differences in running behavior between genders. Lastly, we apply spatial regression models to explore the relationship between gender-specific running patterns and various socio-economic, demographic, and built environment factors.

4.1. Spatial analysis

Two hexagon grids were generated to cover the extents of the running activities in DC and MTL. The radii of the hexagons were set at 200 m. A number of hexagon grids at different resolutions were tested, but a 200 m radius was selected as it was large enough to include runs along wide routes and trajectories with minor GPS multipathing errors, but small enough so as not to over generalize. The motivation for representing running activities through intersection with a regular grid rather than snapping to a road network is that many running trajectories do not follow traditional roads, with runners choosing to run through parks, along beaches, or making their own trails. Hexagons are well-suited to represent curves of geographic features which are not often perpendicular in shape, such as roads and running trajectories. Fig. 2 shows raw running activities overlaid on top of our 200 m resolution hexagon grid. By selecting a regular grid, we could intersect our line geometries and count the number of runs that traversed each of the grid cells. In order to ensure robust results, we removed all grid cells that contained less than 10 unique running activities from further analysis. This was done to reduce the influence of a single run on our overall analysis results. (See Fig. 3.)

Table 1
Number of activities, average distance (km), and average duration (min:*sec*) of running activities split by gender and region. The number of activities run after sunset are included in square brackets.

Gender		Montreal, QC	Washington, DC
Women	Number of Activities [after sunset]	5174 [415]	5052 [491]
	Mean Distance (median)	10.17 (8.53)	11.42 (9.66)
	Mean Duration (median)	53:46 (45:49)	57:42 (48:54)
Men	Number of Activities [after sunset]	5333 [619]	4887 [564]
	Mean Distance (median)	11.25 (9.98)	12.12 (10.30)
	Mean Duration (median)	58:22 (45:01)	53:52 (47:51)

[4] Areal interpolation calculated using the Tobler PySAL module: https://pysal.org/tobler/

Table 2
Independent variables for regression analysis. These include socio-economic, demographic, and built environment variables.

Category	Data	Description
Built Env.	Land/water Ratio	Percent of hexagon that is land.
	Parks & Rec. Space	Percent of hexagon overlapping a 400 m buffer of a park or recreational space.
	Bike Lanes	Percent of hexagon overlapping a 2 m radius buffer of bike lanes.
	Sidewalks	Binary of hexagon containing a sidewalk.
	Road Density	Sum of the kernel density of road overlapping hexagon.
	Walkability	Walkability index value at center of hexagon. Notably, 'walkability' in MTL is represented as Active Living Potential (ALP).
	Proximity to Water	Percent of hexagon overlapping a 400 m buffer of a lake, river, ocean.
	Tree Canopy	Percent of hexagon overlapping tree canopy layer.
	Traffic Volume	Median vehicle traffic volume per hexagon.
	Slope	Median slope value per hexagon.
	Points of Interest	Number of POI in each hexagon, split into categories: Public, Health, Leisure, Catering, Accommodation, Shopping, Tourism.
Socio-economic	Population Density	Median population density per hexagon.
	Median Income	Median household income per hexagon.
	Mode of Commuting	Percent of hexagon by mode of commuting: Walking, Public Transit, Biking, Car, Other.
	Crime Density	Prevalence of crime type in hexagon: Theft, Car Theft, Vandalism, Violent Crime.
Demographic	Race	Percent of hexagon white or Caucasian.
	Visible Minority	Percent of hexagon of visible minority (Black or African American, Asian, First Nation, Hawaiian or Pacific Islander).
	Age Group	Percent of hexagon by age group: Under 15, 15–65, 65 + .

Next, for each hexagon cell, we calculated the number of running activities completed by women and by men, separately. Since we had a different number of running activities for each gender, the value in each cell was normalized by dividing by the total number of running activities in the data set. We then calculated the percentage of runs in that cell run by women. For example, if a single cell in MTL contained 35 running activities by women and 42 running activities by men, we calculated the normalized values as $\widetilde{F} = 35/5174$ and $\widetilde{M} = 42/5052$. The percentage of runs in that cell run by women was then calculated as $\widetilde{F} / (\widetilde{F} + \widetilde{M}) \times 100$. This was calculated for all cells in each of our regional data sets. Given these percentage running values by women, we then used global Moran's I to calculate the spatial autocorrelation in the gender-split running data. This was done to determine if higher percentages of female running, and inversely male running, clustered spatially and to what degree. Moran's I was calculated for both regional data sets independently in order to address `RQ1`, namely to identify if there is spatial clustering by gender and `RQ2` if that spatial clustering varies between regions.

4.2. Temporal analysis

In addressing `RQ3`, we calculated the temporal differences in running

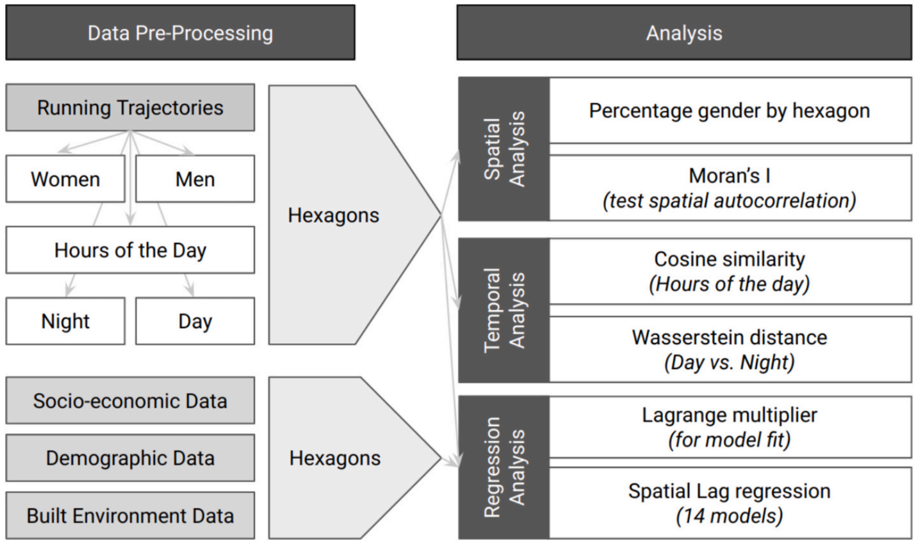


Fig. 1. An overview for our approach.

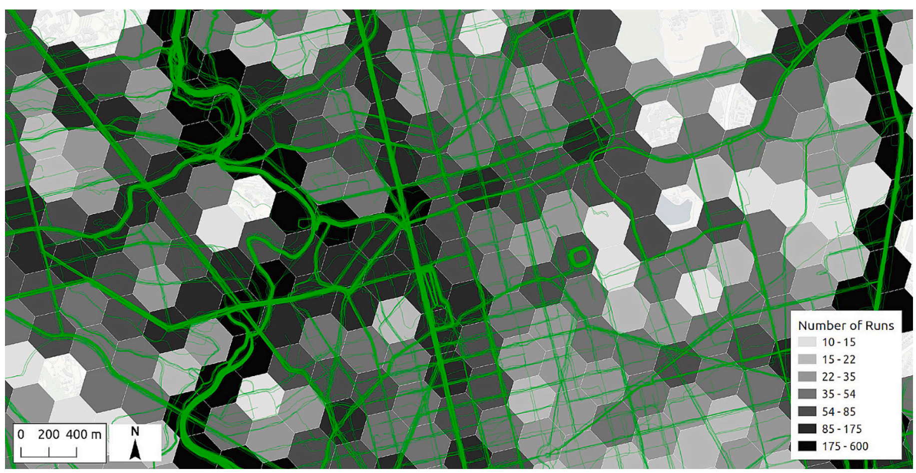


Fig. 2. Example region in DC showing raw running trajectories (green) on top of our hexagon grid.

activities split by gender and geographic region. First, all runs were grouped by the hour of the day in which the activity occurred. These were separated by region and gender and normalized by dividing each count of runs per hour, by the total number of runs. This allowed us to identify what times of day are more popular for running, split by gender and region. Given the normalized data, we then subtract hourly runs by men from hourly runs by women for each region independently. Negative values indicate times of day that are more popular among men with positive values indicating higher popularity among women. We then calculated the cosine similarity of the temporal patterns for each pair of cities and genders. Cosine similarity is a measure used to calculate the similarity between two vectors in an inner product space. In our case, we use the temporal patterns as our vectors and compare each possible pair of vectors. The analysis results in a value bounded between 0 and 1, with 1 indicating the two temporal patterns are identical.

Next, we accessed historical daily sunrise and sunset data from the National Oceanic and Atmospheric Administration[5] for MTL and DC. Using these data and the activity start times, we labeled each run as either a *day* or *night* run (see counts in Table 1. We then generated new hexagon data sets for daylight and night, for each gender and in each region. After normalizing the data, we calculated *Wasserstien* distance[6] between our two gendered data sets for each city. Wasserstein distance (Vallender, 1974), is a mathematical approach used to measure the dissimilarity between two probability distributions. In our case, these are two dimensional probability distributions with our weights being the relative percentage of runs that occurred in each hexagon cell. This two-dimensional analysis allows us to compare the spatial dissimilarities of regional running patterns between women and men with respect to day or night.

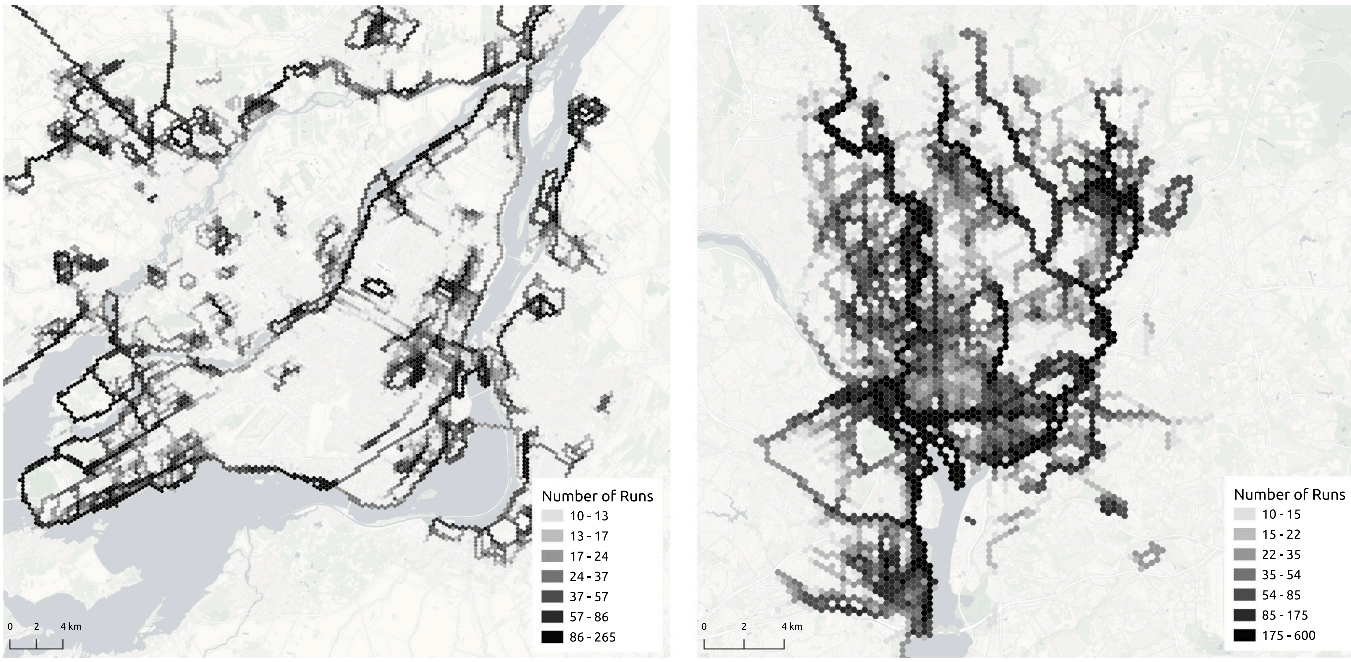
4.3. Spatial regression analysis

With the objective of identifying which socio-economic, demographic, or built environment factors contributed to increased running by women and men (RQ4), we ran a series of regression models. For our regression analysis, we restricted the spatial extents of our data sets to the political boundary of Washington D.C.[7] and the island of

[5] https://gml.noaa.gov/grad/solcalc/

[6] https://cran.r-project.org/web/packages/emdist/emdist.pdf

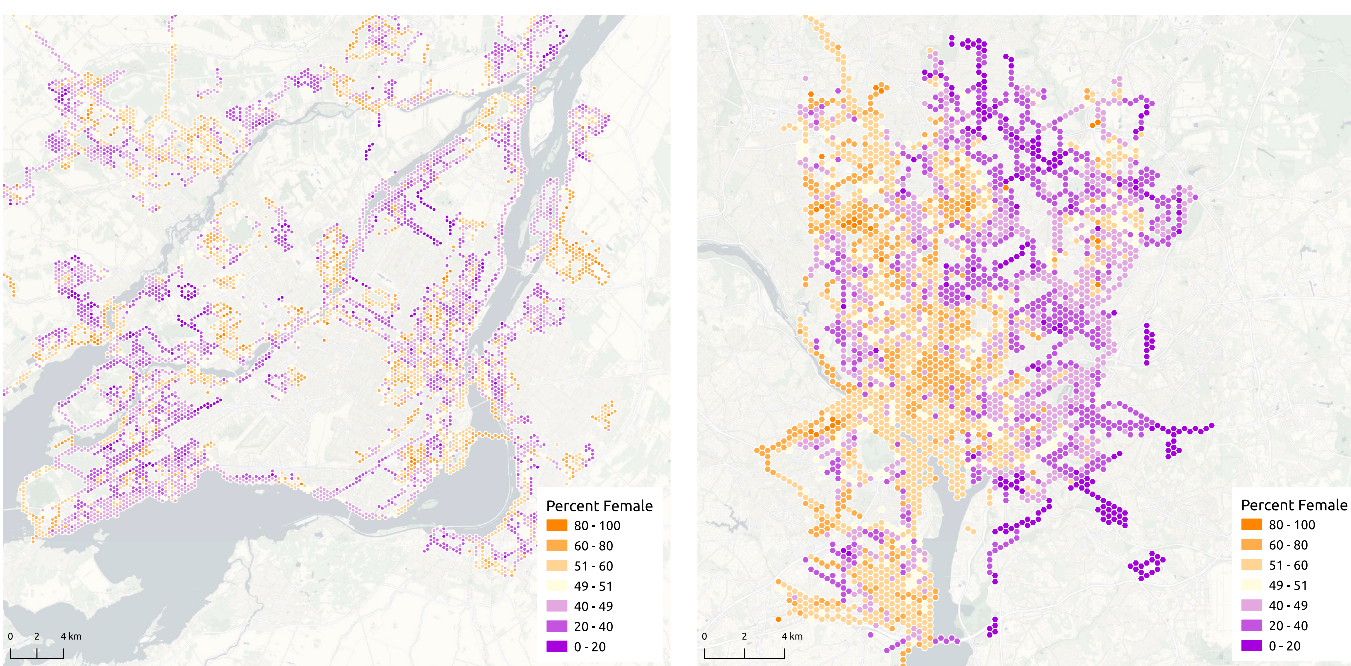
[7] This is smaller than the area of all of our activities, which range over parts of Virginia and Maryland.



(a) Montréal, QC (b) Washington, DC

Fig. 3. Number of unique runs that start, stop, or cross each hexagon. Only those hexagons that include at least ten runs are included in this analysis. Notably, Montréal, QC has a lower density of runs than Washington, DC.

Montréal, QC. This is due to the limitation of the data sources that typically only report values for within specific political boundaries. To start, we ran a simple linear regression model with all of our independent variables and the *total number of runners* per hexagon as our dependent variable. This was done for both MTL and DC, independently. Observing the variance inflation factors (VIF) for all our predictive variables, we removed those with VIF values above 5, those indicating a high degree of multicollinearity. We then used global Moran's I to measure the degree of spatial clustering within our data. The results indicated a high degree of spatial clustering across all linear models. We next computed the Lagrange multiplier statistics which indicated that a spatial lag regression model was a better fit for our data. Provided these outcomes we constructed a series of spatial lag models for each of our cities using a spatial weights matrix based on a location's six nearest



(a) Montréal, QC (b) Washington, DC

Fig. 4. Gender percentages assigned to 200 m hexagons based on activity traversal. The color orange represents female dominated regions while the color purple represents male dominated regions.

neighbors (borders of hexagon). We removed variables that did not show significance in any of our models and re-ran our analyses. After exploratory analysis, we ended up with a total of 14 different spatial lag regression models, seven for each city. All models included the same independent variables and assessed one of the following dependent variables:

- Total number of runners;
- Number of female runners;
- Number of male runners;
- Number of female runners at night;
- Number of male runners at night;
- Number of female runners during the day;
- Number of male runners during the day.

5. Results

5.1. Spatial analysis

The results of aggregating runs into hexagons and splitting by percentage of female running activities per hexagon are shown in Fig. 4. Darker saturations of orange indicate a higher percentage of female runners with darker purples indicating male-dominated regions. Visibly, one can see clear spatial clustering across both MTL and DC, though the clustering patterns differ between the two cities. Most notably, we observe smaller clusters distributed across MTL compared to the large and contiguous clusters visible in the DC data. We can also clearly see an east-west divide in the DC data with areas to the west dominated by female runners and areas to the east dominated by men. A global Moran's *I* analysis on percent women results in a significant statistic of 0.492 (Expected: −0.0003) for MTL and significant value of 0.647 (Expected −0.0007) for DC, meaning substantial spatial clustering in the two data sets but a greater degree of clustering in DC. Focusing on the raw count of women and men running, the Moran's I statistic indicated a slightly higher degree of significant clustering for women (0.600) than men (0.558) in DC. The same pattern was observed, to a lesser degree, in MTL with women (0.523) clustering more than men (0.511).

We also developed density plots to examine the differences in gender percentages between the two cities. Fig. 5 shows density plots for percent female running activities in each hexagon. These density plots show DC with higher kurtosis than MTL, indicating that there is a larger number of hexagons with lower and higher percentages of female runners in DC as compared to MTL. MTL has a higher number of hexagons with 40–50 % female runners, whereas the dominant percent of female running hexagons in DC is in the 50–60 % range.

The results of our Wasserstien distance calculations reveal that in DC and MTL, the differences in the spatial distribution of running activities between women and men increases after sunset. For DC, comparing day running between women and men produces a Wasserstien distance of 0.0062. The same comparison for night time activities produces a distance of 0.0081. For MTL, the Wasserstien distance is 0.0072 during the day, and 0.0079 at night. As absolute values, the differences are small, but relative to one another the difference is substantial.

5.2. Temporal analysis

A basic descriptive analysis of the running activities shows that 13.1 % of all running activities conducted by men, in both MTL and DC, were done at night, compared to 8.8 % and 10.8 %, respectively, for women. Furthermore, we explored the running activities of both men and women, split by hour of the day in which the run started. Fig. 6 shows which hours of the day were *relatively* more popular by which gender. We say relatively, here, as we normalized across all activities to ensure even comparison, so that the sum of each temporal distribution is one. We then subtracted the temporal popularity distribution for men from women. The results indicate that morning runs are relatively more popular with women when compared to men, with men demonstrating a relatively larger interest in running in the evenings. We can see this visually in the figure, but these findings are also supported through our cosine similarity analysis. The results of this analysis indicate that genders are most similar *within* cities with MTL = 0.989 and DC = 0.977. Comparing the same gender *across* cities resulted in lower similarities with women = 0.921 and men = 0.944.

Switching our focus to day versus night, Fig. 7 shows the percent female running per hexagon with night time subtracted from day time. This allows us to visually assess the differences in where women and men choose to run after sunset. In this figure, high saturations of red indicate regions where we measured the greatest decrease in percent female runners from day time to night. Green shows the areas where female running percentages increased from day to night. Similar to Fig. 4, we can see spatial clustering, though the clustering is at a finer spatial scale. We can also clearly identify a number of running paths, regions where multiple runners have chosen to run.

In Fig. 7b, we see the largest decrease in percentage of female runners at night in areas like Alexandria, Virginia and the University of Maryland campus in College Park, Maryland.[8] The largest increase in percentage of female runners at night is the northern-most end of Rock Creek Park in Maryland. In MTL, (Fig. 7a) we see the largest decrease in percentage of female runners at night in areas like Frédéric-Back Park and L'Île Bizard. The largest increase in percentage of female runners at night is the northeast side of Mount Royal. Fig. 8 shows us a quintessential case, that of Frédéric-Back Park in northwest Montréal. This is a park that surrounds an old landfill and has limited public lighting with a large amount of tree canopy cover. During daylight hours we see that the majority of runners in the region are women, but at night only running activities by men are reported around the park.

5.3. Regression analysis

Our initial analysis of factors contributing to an increase in *all runners* are provided in Tables 3 and 4. In general, we do find some consistency across our two cities of interest. Land to water ratio is significant in both cities and across genders, reporting that a decrease in land area leads to a slight increase in running activities, indicating a preference for running by water amenities. Runners in both cities prefer to run in areas of lower population density and lower median household income, though the magnitude of population density is higher in DC than in MTL. The reverse is true for the impact of median household income for each city. We also found that runners in both cities prefer to run in areas with some elevation change and near bike lanes, though the latter is much more influential in MTL than DC. Our comparative analysis found that runners in DC had a preference for running in regions with some level of traffic and a slight preference for areas where people tended not to commute to work through walking. In contrast, runners in MTL preferred running in areas low in Active Living Potential, not in or near parks, but with some degree of tree canopy coverage. The importance of the factors varied substantially between cities, with bike lanes and park proximity being very important for runners in MTL, but much less so in DC. Comparatively, population density and traffic were more important for runners in DC.

In comparing the difference between women and men, we find that in MTL the significance of our variables stayed mostly the same as the *all runners* model, with a few exceptions. Running activities by men increased as the percentage of the population that was white in a region increased. This was not found for female running activities in MTL. A decrease in traffic led to an increase in female running activities, but was not significant for men. Bike lanes remained the most important contributing factor to running for both genders, with running decreasing

[8] While in neighboring US states, these regions are part of the greater Washington, DC region.

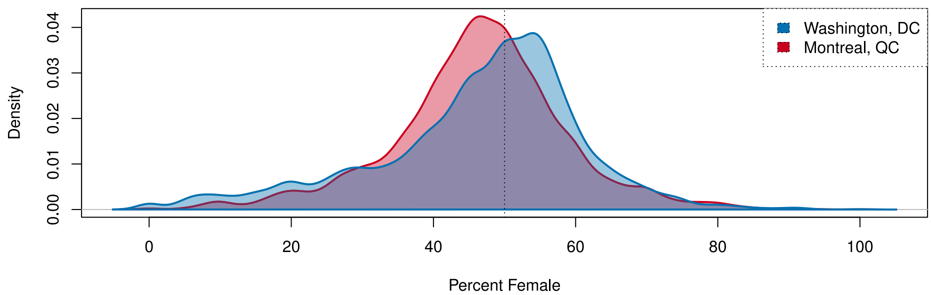


Fig. 5. Density plots of hexagons based on percent female runners in Washington, DC and Montréal, QC.

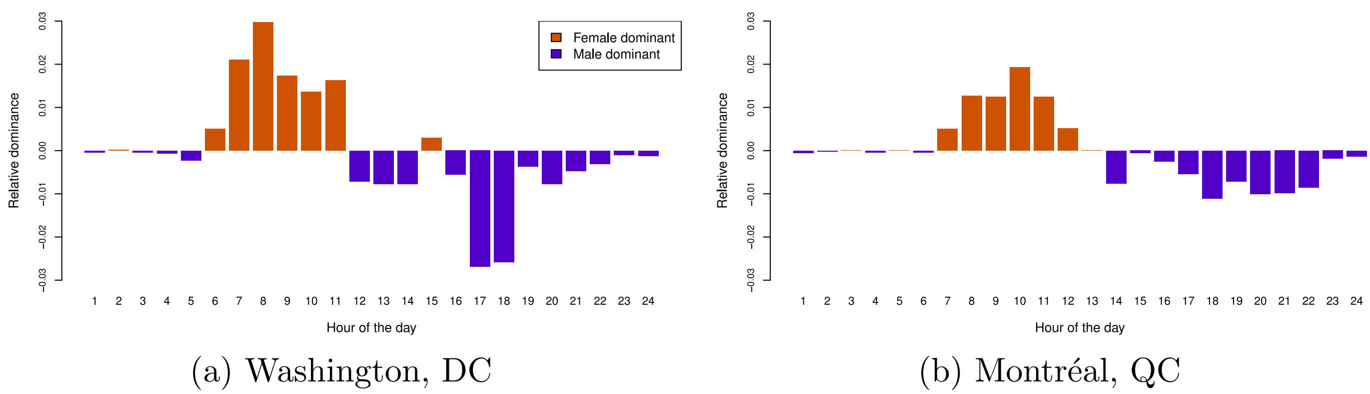


Fig. 6. Run start times aggregated to hours of the day split by gender and city. The data for each gender and city were normalized to sum to one, and then male runs were subtracted from female runs. The resulting figures show which genders were dominant at which time of day.

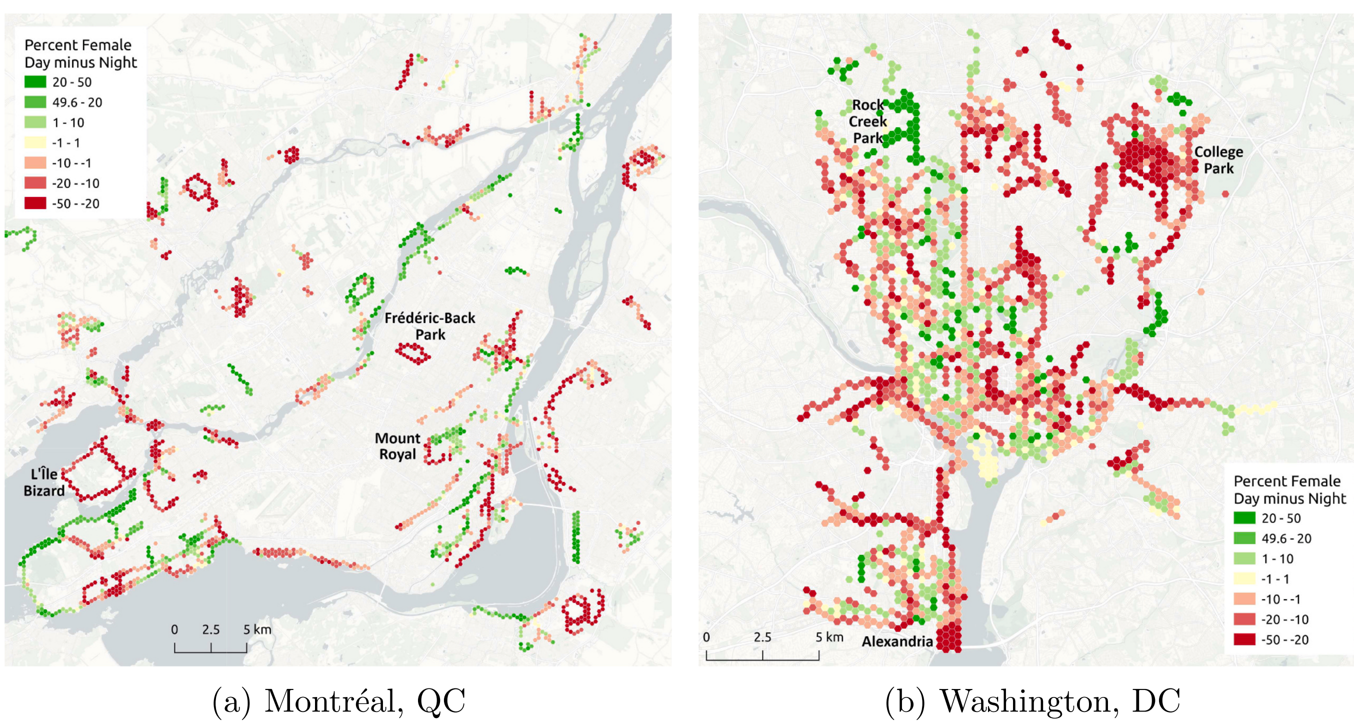
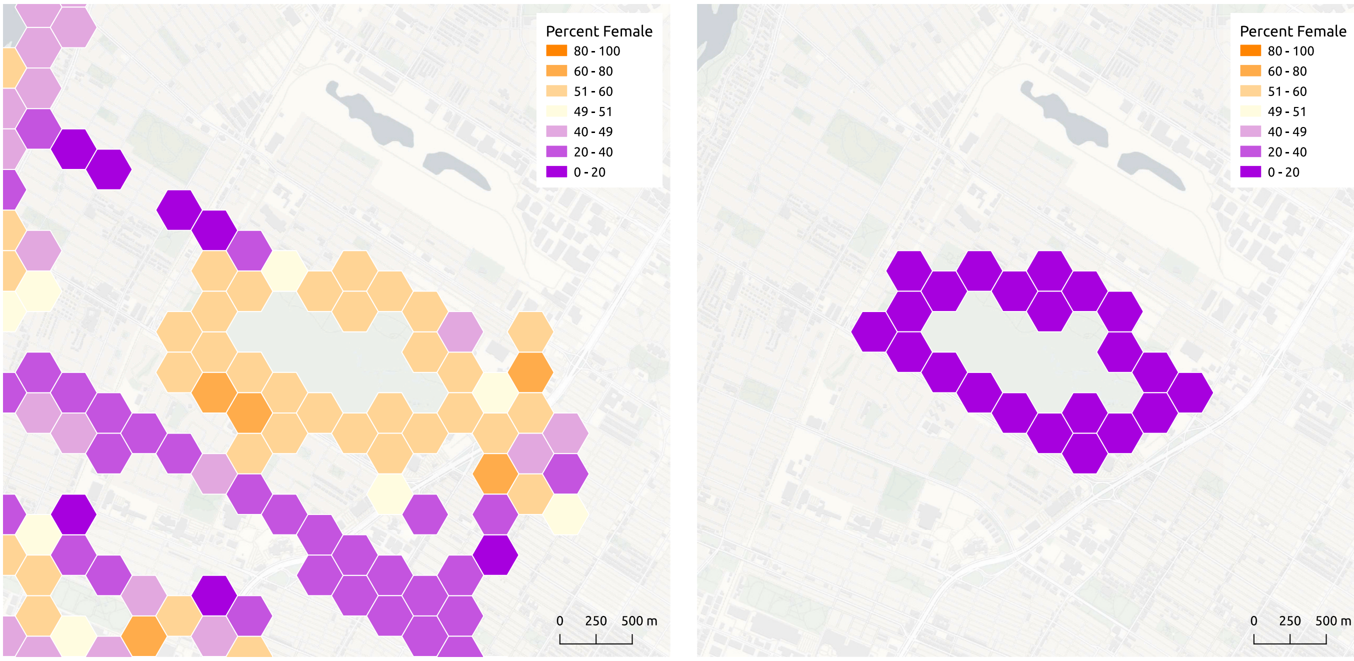


Fig. 7. Percent women per hexagon calculated for runs in day time versus night. The figure shows night time percent subtracted from day time. Red areas are regions with the greatest decrease in percent of female runners from day to night while green areas are regions with the greatest increase in percent of female runners.

in areas near parks. In both cases, the absolute values of these coefficients were higher for women than men. DC showed a similar pattern to MTL, with both genders exhibiting similar relationships with the independent variables. The only notable difference, while small, was that women were less likely to run in areas containing high numbers of people that walked to work, where as this variable was not significant for men.

The results of our day versus night running analysis in MTL (Table 5)



(a) Runs during the day (b) Runs at night

Fig. 8. Frédéric-Back Park in northwest Montréal.

Table 3
Regression results for All Runners, Women, and Men in Montréal, QC. Shaded regions indicate significance of $p < 0.1$.

Coefficients	All Runners	Women	Men
(Intercept)	0.10931***	0.12342***	0.09691***
Land Ratio	−0.07063***	−0.08606***	−0.05847***
Walkability / ALP	−0.02833*	−0.02972*	−0.02683*
Race White	0.01733	0.00996	0.02277*
Commute Public Transit	−0.00127	−0.00082	−0.00134
Commute Bike	0.00470	0.01278	−0.00125
Commute Walk	0.00764	0.01248	0.00382
Median HH Income	−0.03802**	−0.03628*	−0.03894**
Population Density	−0.06321***	−0.06907***	−0.05879***
Total Traffic	−0.04094	−0.05701*	−0.02885
Crime Violent	0.01062	0.01635	0.00658
Crime Car Theft	−0.01854	−0.02662	−0.01248
Slope	0.16605***	0.20816***	0.13392***
Bike Lane Buffer	1.35809***	1.48275***	1.26359***
Tree Canopy	0.04233***	0.04840***	0.03730***
Parks & Recreation Buffer	−1.07299***	−1.16166***	−1.00592***
Road Density	0.03975	0.03816	0.04144
POI - Public	0.01621	0.02565	0.00928
Rho	0.7071***	0.71042***	0.7097***
AIC	−4524	−4053	−4711
AIC for LM	−3221	−2724	−3392
Log Likelihood	2282.003	2046.691	2375.506
Nagelkerke pseudo-R^2	0.45	0.46	0.44

Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4
Regression results for All Runners, Women, and Men in Washington, DC. Shaded regions indicate significance of $p < 0.1$.

Coefficients	All Runners	Women	Men
(Intercept)	0.11581***	0.10669***	0.12343***
Land Ratio	−0.11241***	−0.11256***	−0.10986***
Walkability / ALP	−0.00085	−0.00013	−0.00276
Race White	−0.00870	−0.00354	−0.01886
Commute Public Transit	0.00063	0.00051	0.00071
Commute Bike	−0.00377	−0.00325	−0.00390
Commute Walk	−0.00105**	−0.00096*	−0.00088
Median HH Income	−0.00003***	−0.00003***	−0.00003***
Population Density	−0.31536***	−0.29534***	−0.29009***
Total Traffic	0.16859***	0.15218***	0.16607***
Crime Violent	−0.04103	−0.03860	−0.03659
Crime Car Theft	0.03857	0.04665	0.03026
Slope	0.10027***	0.10182***	0.10012***
Bike Lane Buffer	0.03391*	0.03029*	0.03306*
Tree Canopy	0.09612	0.09029	0.07132
Parks & Recreation Buffer	0.05291	0.05017	0.04072
Road Density	0.00287	0.00950	−0.02387
POI – Public	0.00040	−0.00068	−0.01462
Rho	0.7106***	0.71206***	0.7126***
AIC	−1809	−2019	−1839
AIC for LM	−1214	−1413	−1234
Log Likelihood	924.253	1029.405	939.68
Nagelkerke pseudo-R^2	0.53	0.56	0.52

Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

highlight some interesting differences. Active Living Potential significantly (negatively) contributes to running in a region but only during the day. Similar patterns were also observed for median household income and presence of tree canopy, both being important for both genders, but only during daylight hours. In contrast, nearby *public* points of interest (e.g., police stations, hospitals) are positively associated with running for both genders, but only at night. In examining the magnitude of the coefficients, we see that both women and men prefer running near water and in areas with more slope during the day. This magnitude change between day and night continues for bike lanes and parks & recreational areas. Comparing female to male running activities we see that running activities by men were significantly associated with an increase in racially white populations, but only during the day. Our modelling also identified populations with increased public transit commuters as being important for male runners at night. By comparison, females running at night preferred regions with a high percentage of people that walked to work.

In DC (Table 6), we find a much bigger difference in factors

Table 5
Regression results for female and male runners during the day and night in Montréal, QC. Shaded regions indicate significance of $p < 0.1$.

	Women		Men	
	Day	Night	Day	Night
(Intercept)	0.12403***	0.05122***	0.09545***	0.02603**
Land Ratio	−0.08383***	−0.06348***	−0.05684***	−0.02216
Walkability / ALP	−0.02992*	−0.01182	−0.02663**	−0.00618
Race White	0.01033	0.00276	0.02163*	0.01175
Commute Public Transit	−0.00185	0.01025	−0.00397	0.01719*
Commute Bike	0.01458	−0.00790	0.00119	−0.01479
Commute Walk	0.00690	0.05321**	0.00129	0.01561
Median HH Income	−0.04067**	0.01999	−0.03938**	−0.00482
Population Density	−0.06575***	−0.068937***	−0.05138**	−0.05879***
Total Traffic	−0.05421*	−0.05628*	−0.02523	−0.02886
Crime Violent	0.02554	−0.06898	0.01096	−0.02475
Crime Car Theft	−0.02621	−0.01798	−0.01104	−0.01177
Slope	0.20394***	0.14916***	0.12882***	0.06147***
Bike Lane Buffer	1.45902***	1.05938***	1.17351***	0.84329***
Tree Canopy	0.04824***	0.07512	0.03728***	0.00685
Parks & Rec. Buffer	−1.12931***	−0.95620***	−0.92977***	−0.69824***
Road Density	0.03362	0.06704	0.03774	0.03456
POI – Public	0.01915	0.07181**	0.00364	0.03770*
Rho	0.71436***	0.71836***	0.71089***	0.72897***
AIC	−4122	−3712	−5040	−5985
AIC for LM	−2759	−2330	−3703	−4527
Log Likelihood	2080.776	1875.858	2540.003	3012.632
Nagelkerke pseudo-R^2	0.47	0.41	0.45	0.40

Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6
Regression results for female and male runners during the day and night in Washington, DC. Shaded regions indicate significance of $p < 0.1$.

	Women		Men	
	Day	Night	Day	Night
(Intercept)	0.10669***	0.02615	0.12343***	0.02161
Land Ratio	−0.11256***	−0.02906	−0.10986***	−0.02068
Walkability / ALP	−0.00013	−0.00363	−0.00276	0.01467
Race White	−0.00354	0.02061*	−0.01886	0.00866
Commute Public Transit	0.00051	0.00049	0.00071	0.00032
Commute Bike	−0.00325	−0.00155	−0.00390	−0.00465
Commute Walk	−0.00096*	−0.00155***	−0.00088	−0.00119*
Median HH Income	−0.00003***	0.01999	−0.00003***	−0.00002***
Population Density	−0.29534***	−0.15177	−0.29009***	−0.36102***
Total Traffic	0.15218***	0.11622***	0.16607***	0.12041***
Crime Violent	−0.03860	−0.02836	−0.03659	−0.04622
Crime Car Theft	0.04665	0.00024	0.03026	0.00407
Slope	0.10182***	0.02360	0.10012***	−0.00270
Bike Lane Buffer	0.03029	0.03790**	0.03306*	0.01427
Tree Canopy	0.09029	0.07512	0.07132	0.22854***
Parks & Rec. Buffer	0.05017	0.08717*	0.04072	0.06220
Road Density	0.00950	0.03323	−0.02387	0.14031**
POI – Public	−0.00068	0.08587***	−0.01462	0.05240
Rho	0.71206***	0.73397***	0.7126***	0.72608***
AIC	−2019	−2257	−1839	−1825
AIC for LM	−141	−1620	−1234	−1245
Log Likelihood	1029.405	1148.463	939.68	932.346
Nagelkerke pseudo-R^2	0.56	0.48	0.52	0.42

Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

contributing to running between the day and night, as compared to MTL. Running near water was a preference for both women and men during the day, but not the night. Similarly, slope was an important factor during the day, but not the night. The percentage of the population that walked to work was negatively correlated with running for both women and men at night, but only for women (and much less so) during the day. Median household income and population density were significant contributors to running for both women and men during the day, but only for men at night. In comparing genders, we find that women have a preference for running in regions with greater percent white populations, near bike lanes, parks, and public points of interest, but only at night. Men, on the other hand, prefer tree canopy and increased road density, but again, only at night. Running activities by men were only significant near bike lanes during the day.

6. Discussion

Based on the results of the analyses conducted in this work, we can revisit and address the four original research questions presented in the introduction. In RQ1 we asked if preference in running locations differs between genders. Based on our analysis of running in our two focal cities, we can definitively say that there is a gender-based preference towards running locations. The results indicated significant spatial clustering with different patterns for male and female runners. In DC, areas to the West are dominated by female runners, and areas to the East by male runners. This mirrors the East-West racial divide of the city (Logan, 2017). In MTL, smaller clusters are distributed across the region. The results of our regression analysis also indicate that there are preferences for different aspects of the built environment that are location specific, such as access to water or demographic factors like population density. In addressing RQ2, namely, if gender differences in running locations vary between cities, our results indicate that they vary significantly between MTL and DC. This is reflected in our spatial, temporal, and regression-based analysis. The results of our analysis of RQ3 suggests that gender differences in running location also vary temporally. Women are relatively more likely to run in the mornings, whereas men show a greater relative preference for evening runs. Women in DC exhibited the most unique temporal running patterns in our analysis with respect to spatial clustering and temporal patterns. For both cities, the day versus night analysis also reveals significant differences in the locations where women and men choose to run, with notable decreases in female running activities in certain areas at night. Our analysis addressing RQ4 investigates this further by examining the built environment, socio-economic, and demographic factors that influence gender-based running location preferences. These factors include access to water, population density, median household income, elevation change, proximity to bike lanes, and traffic levels. In MTL, bike lanes and park proximity were more influential, whereas in DC, population density and traffic were more important. Temporally, factors like median household income, proximity to water, and slope have different influences on running activities during the day versus at night. There was a notable difference between day and night running between our two focal cities, with DC showing a larger difference in significant contributing variables than MTL. This difference continued when gender was included in our analysis. There are a number of potential explanations for this, not least of which are the social, political, and historical differences between the two countries and cities. For both cities, built environment variables like access to public points of interest and bike lanes were more important for women at night than men.

Our findings are in line with other research in this domain that study fitness activity preferences through surveys, interviews or fitness tracking applications. For example, recent work by Zhang et al. (2023) found that both urban and rural runners preferred running near large bodies of water and in areas with fewer streets. Other work has found that runners prefer green spaces and comfortable running surfaces and tend to avoid areas with high population densities (Deelen et al., 2019). These mirror our findings that runners in both DC and MTL prefer low population density areas and to run on or near bike lanes. While not significant in DC, runners in MTL did prefer regions with more tree canopy.

Notably, slope was a significant factor contributing to running in both of our study cities. This does reflect most recreational runners preference to have some variety to their running activities (Norman and Pickering, 2019). While DC does not have a wide range of gradients on which to run, there is some variation in elevation across the city and many of the parks boasting high running volume (e.g., Rock Creek Park) do include elevation change. In MTL, the large mount (Mount Royal) in the middle of the city is heavily used for recreational running, so it makes sense that slope would be an important contributing factor. Given that many of the running trails in large parks, both in MTL and DC, do not have adequate public lighting infrastructure, it follows that slope would be less of a factor at night.

The threat of violence to women is a pervasive part of our culture and has a significant effect on the decisions women make about where, when, how, and with whom they occupy public space (Roper, 2016; Coble et al., 2003). Prior to analyzing our data we hypothesized that reported criminal activity would significantly correlate with preference towards running locations. Surprisingly, the results of our analysis found that while crime demonstrated a slight (significant negative) correlation with increased running activities on its own, when included in our regression models, the results were not significant. There are a number of possible reasons for this but we would suggest that reported crimes are not an exact proxy for *sense of safety*. This is in line with research in this domain showing that perception of safety does not always correlate with criminal activity (Amiruzzaman et al., 2021; Ogneva-Himmelberger et al., 2019). Other factors such as the length of time that an individual has lived in an area may be important in such a calculation. Unfortunately, access to other data that would be important, such as public lighting (Svechkina et al., 2020), incidents of catcalling (Martinez et al., 2023), etc. were not accessible for analysis. Future efforts in this domain should focus on collecting or accessing such data for analysis.

It is important to acknowledge that the data collected and analyzed in this work only represent a fraction of recreational runners. Most recreational runners do not record their activities, let alone publicly share them through a geosocial fitness tracking application. While we argue that some level of generalization of these findings is reasonable, the results presented in this work really reflect the population in our sample and use of these findings for policy implementation should consider this.

Further work on this topic should explore additional built environment variables as well as compare these results to cities outside of North America. Analysis of additional data from different sources could also decrease the single-source bias of such work, as would alternative research methods such as interviewing runners and conducting surveys, perhaps in tandem with geospatial analysis of a large dataset. Our next step in this domain is to further explore gender differences across fitness activity types, e.g., cycling, and compare our results.

7. Conclusions

The findings presented in this work have insightful implications for urban planning, public health, gender studies, and community safety. Specifically, our results identifying gender-based preferences highlight the need for more inclusive planning and consideration of preferences that cater to the needs of all genders. The importance of water bodies, bike lanes, and green spaces in both Montréal and Washington, DC suggest that planners should consider these elements of city landscapes in order to promote physical activity and good health. This aligns with the United Nations Sustainable Development Goals (SDG) (United Nations Department of Economic and Social Affairs, 2015) of Good Health and Well-being (SDG 3) and would promote increased active transport within a city (SDG 11). Our observed temporal differences in running patterns between genders may inform scheduling of community events and the allocation of resources for public health initiatives. The unique differences in spatial and temporal running patterns between women and men also contribute to the continually growing body of literature on changing gender roles and cultural dynamics in urban settings and reinforce similar conclusions found in the literature on the gendered dimensions of running and other public activities. In identifying and acknowledging these differences, governments and industry leaders can continue to develop policies that reduce inequalities (SDG 5 & 10).

Author statement

We, the authors of the manuscript titled “*Gender differences in urban recreational running: A data-driven approach*,” confirm that this work is original and has not been published previously, nor is it under consideration for publication elsewhere. All authors have significantly contributed to the research and writing of this manuscript, have reviewed the final version, and approved its submission to the *Journal of Transport Geography*.

Furthermore, we declare that there are no conflicts of interest related to this work.

CRediT authorship contribution statement

Grant McKenzie: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Daniel Romm:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Clara Féré:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Maria Laura Guerrero Balarezo:** Writing – review & editing, Writing – original draft.

Data availability

The authors do not have permission to share data.

References

Amiruzzaman, M., Curtis, A., Zhao, Y., Jamonnak, S., Ye, X., 2021. Classifying crime places by neighborhood visual appearance and police geonarratives: a machine learning approach. J. Comput. Soc. Sci. 1–25.

Battiston, A., Napoli, L., Bajardi, P., Panisson, A., Perotti, A., Szell, M., Schifanella, R., 2023. Revealing the determinants of gender inequality in urban cycling with large-scale data. EPJ Data Sci. 12 (1), 9.

Best, A., Braun, B., 2017. Using a novel data resource to explore heart rate during mountain and road running. Phys. Rep. 5 (8), e13256.

Busette, C., Elizondo, S., 2022. Economic disparities in the washington, dc metro region provide opportunities for policy action.

Butryn, T.M., Furst, D.M., 2003. The effects of park and urban settings on the moods and cognitive strategies of female runners. J. Sport Behav. 26 (4), 335.

Coble, T.G., Selin, S.W., Erickson, B.B., 2003. Hiking alone: understanding fear, negotiation strategies and leisure experience. J. Leis. Res. 35 (1), 1–22.

Collinson, J.A., 2008, February. Running the routes together: corunning and knowledge in action. J. Contemp. Ethnogr. 37 (1), 38–61.

Deelen, I., Janssen, M., Vos, S., Kamphuis, C.B., Ettema, D., 2019. Attractive running environments for all? A cross-sectional study on physical environmental characteristics and runners' motives and attitudes, in relation to the experience of the running environment. BMC Public Health 19, 1–15.

Ettema, D., 2016, November. Runnable cities: how does the Running environment influence perceived attractiveness, Restorativeness, and Running frequency? Environ. Behav. 48 (9), 1127–1147.

Ghani, F., Rachele, J.N., Loh, V.H., Washington, S., Turrell, G., 2019, January. Do differences in social environments explain gender differences in recreational walking across Neighbourhoods? Int. J. Environ. Res. Public Health 16 (11), 1980.
Goel, R., Oyebode, O., Foley, L., Tatah, L., Millett, C., Woodcock, J., 2023. Gender differences in active travel in major cities across the world. Transportation 50 (2), 733–749.
Hockey, J., Collinson, J.A., 2006, April. Seeing the way: visual sociology and the distance runner's perspective. Vis. Stud. 21 (1), 70–81.
Joyner, M.J., 1993. Physiological limiting factors and distance running: influence of gender and age on record performances. Exerc. Sport Sci. Rev. 21 (1), 103–134.
Lee, K., Sener, I.N., 2021. Strava metro data for bicycle monitoring: a literature review. Transp. Rev. 41 (1), 27–47.
Logan, J.R., 2017. Racial segregation in postbellum southern cities: the case of Washington, DC. Demogr. Res. 36, 1759.
Martinez, C.T., Tappan, C.S., Baker, H., Edwards, M., Martinez, J., 2023. Sexual harassment and assault across trail and ultrarunning communities: a mixed-method study exploring gender differences. Sociol. Sci. 12 (6), 359.
Nielsen, 2021. Recreational running consumer research study.
Norman, P., Pickering, C.M., 2019. Factors influencing park popularity for mountain bikers, walkers and runners as indicated by social media route data. J. Environ. Manag. 249, 109413.
Ogneva-Himmelberger, Y., Ross, L., Caywood, T., Khananayev, M., Starr, C., 2019. Analyzing the relationship between perception of safety and reported crime in an urban neighborhood using gis and sketch maps. ISPRS Int. J. Geo Inf. 8 (12), 531.
Parks, S.E., Housemann, R.A., Brownson, R.C., 2003, January. Differential correlates of physical activity in urban and rural adults of various socioeconomic backgrounds in the United States. J. Epidemiol. Community Health 57 (1), 29–35.
Polko, P., Kimic, K., 2022, May. Gender as a factor differentiating the perceptions of safety in urban parks. Ain Shams Eng. J. 13 (3), 101608.
Pollard, T.M., Wagnild, J.M., 2017, April. Gender differences in walking (for leisure, transport and in total) across adult life: a systematic review. BMC Public Health 17 (1), 341.
Qviström, M., Fridell, L., Kärrholm, M., 2020, July. Differentiating the time-geography of recreational running. Mobilities 15 (4), 575–587.
Rana, J., 2022, September. Gendered enskilment: becoming women through recreational running. Senses Soc. 17 (3), 290–302.
Roper, E.A., 2016. Concerns for personal safety among female recreational runners. Women Sport Phys. Activity J. 24 (2), 91–98.
Running USA, 2022. 2022 Global Runner Survey – North America Edition.
Schuurman, N., Rosenkrantz, L., Lear, S.A., 2021. Environmental preferences and concerns of recreational road runners. Int. J. Environ. Res. Public Health 18 (12), 6268.
Sport England, 2016. Active People Survey.
Statistics Canada, 2022. Population and dwelling counts: Census metropolitan areas, census agglomerations and census subdivisions (municipalities). https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810000301 (Accessed on 12/18/2023).
Statistics Canada, 2023. Participation in canadian society through sport and work.
Strava, 2022. New strava initiative strives to get more women to the starting line in professional sports.
Svechkina, A., Trop, T., Portnov, B.A., 2020. How much lighting is required to feel safe when walking through the streets at night? Sustainability 12 (8), 3133.
U.S. Census Bureau, 2020. U.S. census bureau quickfacts: District of columbia. https://www.census.gov/quickfacts/fact/table/DC/PST045222 (Accessed on 12/18/2023).
United Nations Department of Economic and Social Affairs, 2015. The 17 Sustainable Development Goals. https://sdgs.un.org/goals (Accessed on 01/17/2024).
Vallender, S., 1974. Calculation of the wasserstein distance between probability distributions on the line. Theory Prob. Appl. 18 (4), 784–786.
Venter, Z.S., Gundersen, V., Scott, S.L., Barton, D.N., 2023. Bias and precision of crowdsourced recreational activity data from strava. Landsc. Urban Plan. 232, 104686.
Visontay, E., 2024. 'Not Safe to Run in the Dark': How Inadequate Lighting in Public Spaces Is Creating Barriers for Women. The Guardian.
Yang, W., Hu, J., Liu, Y., Guo, W., 2023. Examining the influence of neighborhood and street-level built environment on fitness jogging in Chengdu, China: a massive gps trajectory data analysis. J. Transp. Geogr. 108, 103575.
Zhang, S., Wang, Z., Helbich, M., Ettema, D., 2023. Assessing runners' exposure to natural and built environments in the Netherlands: a descriptive assessment based on gps tracking. Int. J. Environ. Health Res. 1–20.
Zhong, Q., Li, B., Chen, Y., 2022. How do different urban footpath environments affect the jogging preferences of residents of different genders? Empirical research based on trajectory data. Int. J. Environ. Res. Public Health 19 (21), 14372.