Università degli studi di Milano-Bicocca

Data Science Lab for Smart Cities

Final Essay

The Role of Gender in Urban Transportation Patterns: A Comprehensive Study Based on the MoTiV Dataset

Authors:
Davide Prando - 886694 - d.prando1@campus.unimib.it
Eleonora Rossi - 886756 - e.rossi100@campus.unimib.it

June 12, 2023

Abstract

This paper explores the gendered dimensions of public space, with a specific focus on mobility, and emphasizes the importance of integrating feminist urban design principles into city planning. Gender plays a critical role in shaping individuals' access to public spaces, yet gender considerations in urban planning have often been neglected. By examining the intersection of gender and mobility, this study aims to shed light on the disparities and challenges faced by different genders in navigating urban spaces. The research utilizes the MoTiV (Mobility and Time Value) dataset, encompassing various indicators related to passenger mobility patterns. The dataset was collected through the Woorti mobile application and covers users from ten different countries. By analyzing indicators within the dataset, such as means of transport usage, trip purposes, and time of trips, the study aims to uncover insights into how gender influences transportation choices and travel patterns. By addressing the gender data gap and considering a comprehensive set of indicators, policymakers and urban planners can work towards creating more inclusive and equitable transportation systems that cater to the diverse needs of all individuals. The integration of feminist urban design principles can contribute to fostering sustainable urban mobility ecosystems and promoting gender equality in cities.

1 Problem Description and Indicators

Gender is a very relevant variable in understanding how someone can access public spaces. Despite this being universally known, in everyday city planning gender remains a neglected theme. Since the 1970s, many scholars (mostly females) started pointing out that urban spaces tend to be constructed mostly for men or heteronormative families[1] and they all tried to propose different city models, built around the concept of inclusivity but all these results were often overlooked by policymakers.

Leslie Kern, one of the most influential scholars in the field of inclusive urban design developed the concept of a more inclusive city and started addressing the issue of how most of the cities we live in are built around men's needs. The bodies of females, but also more in general the bodies of who is not an able bodied working male, have to fight for their place in the urban space. In her book "Feminist City, claiming space in a man made world", Kern highlights how the public space isn't neutral but shapes the possibilities of who lives in the city, according to the needs and the ideas of who built them and since the cities we live in have been built primarily by men urbanists, architects and so on, they are designed according to their needs. In the last decades, feminist urban design started to take more and more space into academia, defining a feminist city as a city for everyone, built around everyone's needs, built around the idea of care.

One of the fields where gender plays a very important role is the field of mobility, where mobility means moving around the city safely, easily, and affordably.

1.1 Gender and Mobility

The exploration of gender disparities in commuting behaviors is of paramount importance due to its potential implications for various aspects of urban life, including transportation planning, infrastructure development, and societal gender equality. By identifying the underlying factors that shape gender-based mobility disparities, policymakers and urban planners can develop targeted interventions to foster more inclusive and sustainable transportation systems.

There are many issues related to gender and mobility. First of all, many transportation systems are built around commuting patterns, useful for someone who simply has to reach the place of employment (usually the men), less useful for someone with more complex mobility needs, related for example to caregiving or domestic tasks (usually the women) [2]. This kind of transportation system is built on the idea that one has to move once in the morning from the house to the workplace and then back in the evening (in the famous peak hours). Studies show[3] that women are more likely to work informal jobs, with different time schedules than the standard 9 to 17, they are less likely to own or have access to a car and they have more complex travel needs, moving many times at different times of the day, usually of peak hours.

In this we can see the harm done by traditional gender roles often assign men as breadwinners and women as caregivers and the profounds gender pay gaps that penalize female work: if in a family one of the parents have to renounce to his job, it's hardly ever the man who does such thing, also because men tend to earn more.

Previous research[4] has identified several key factors influencing gender-based differences in commuting behaviors. Socioeconomic factors, such as income disparities, occupational segregation, and the availability of flexible work arrangements, contribute significantly to gendered mobility variations. Additionally, societal expectations, cultural norms, and safety concerns can also influence the travel choices of women who do not commute to work. Understanding these factors and their impact on commuting behaviors is crucial to address the unique challenges faced by different genders in accessing transportation resources and opportunities.

Another important issue is related to safety. Women are more likely to consider not going out after dark [5], not walking alone [6], not using public transit [7], and not choosing specific routes [8] in order to avoid putting themselves at risk of violence.

This project aims to contribute to the existing body of knowledge by conducting an in-depth investigation of the commuting habits of men and women. By employing rigorous data analysis techniques, we seek to identify and understand the underlying factors that shape the mobility patterns of these distinct gender groups. The findings of this study will help unravel the complexities of gender-based mobility disparities and provide insights for policymakers, urban planners, and transportation practitioners to develop gender-sensitive strategies that promote inclusivity and equality in transportation planning and design.

1.2 Gender Data Gap

Once we chose the topic for this project, the first issue we encountered was the lack of data in this field of analysis. We tried contacting the local transportation company, ATM, but they couldn't provide us with data of mobility patterns broken down by gender. Of course the issue of missing data is not limited to the city of Milan.

Scholars and organizations [9] that aim to fill this data gap point out that all studies on mobility tend to focus on the quantitative dimension of the phenomena and policy makers just try to optimize the issue of moving the largest amounts of people from point A to point B in the shortest time but there is much more to that. Modern approaches are more focused on each single experience, analyzing all the factors that influence one's mobility choices, going against the idea that all genders experience the same issues. The aim of this kind of analysis is related to building a better and more inclusive transportation environment for everyone, not only for women.

This lack of data is even more surprising knowing that women tend to use public transportation more than men: for example, in France 2 out of 3 of public transport passengers are women[10]. Of course, considering the division by gender is only one of the possible analysis to perform on mobility data in order to build a more accessible transportation system. One possible interesting development would be integrating this data with socio economic characteristics of each person in order to better understand how these factors could influence one's relation with the city.

1.3 Indicators

There are numerous indicators available for studying mobility through a gender lens. The International Transport Forum (ITF), a supranational institution comprising 66 member countries, part of the broader Organisation for Economic Co-operation and Development (OECD), has outlined a comprehensive set of indicators on its website[11]. These indicators are particularly valuable for applying gender analysis to transportation projects or policies, emphasizing the importance of considering them during the early stages of project development.

The ITF identifies eight distinct categories of indicators that offer insights into various aspects of gender-related mobility:

- Passenger Transport: These indicators provide information on individuals' travel patterns disaggregated by gender. They encompass details such as the modes of transport used, travel distances, travel times, trip purposes, and more.
- Access to Vehicles: This category includes indicators related to vehicle ownership, access to different types of vehicles, possession of driver's licenses, and similar factors.
- Transport Costs: Indicators within this category focus on household expenditure on transportation, shedding light on the financial aspects of mobility.

- Road Safety: These indicators encompass gender-disaggregated data on fatalities and injuries, gender-based prevention measures, and other relevant factors related to road safety.
- Gender-Based Violence: Indicators in this category examine the perception of safety during travel, public awareness of gender-based violence, legal protections in place, and the number of reported assaults or incidents.
- Transport and Land Use Planning: This category pertains to indicators related to urban planning strategies that consider gender-specific needs and interests, ensuring that transportation planning is inclusive and accommodating.
- Accessibility: Indicators within this category evaluate the accessibility of services, including the family-friendliness of transportation stations, vehicles, and overall services available.
- Transport Labor: These indicators focus on gender-specific participation within the transportation workforce, including workforce composition, gender pay gap analysis, and related labor market dynamics.

For the purpose of this paper, our analysis will concentrate on the indicators falling within the first category: Passenger Transport. By studying various aspects of individuals' transportation patterns, we aim to understand the relevance of gender as a significant variable in shaping mobility behaviors and preferences.

Specifically, we will examine the following indicators:

- Normalized Means of Transport Usage: We will analyze the proportion of different modes of transportation used by individuals, taking into account the covered distance. This normalization will provide insights into the relative preference for each mode of transport, considering gender and trip purpose.
- Purpose of Trips: We will investigate the purpose behind individuals' trips, taking into consideration different genders. This analysis will help us uncover any gender-related differences in travel motivations and preferences.
- Time of the Trips: Additionally, we will consider the temporal dimension by examining the time of the trips. This analysis may reveal variations in travel patterns based on gender, trip purpose, and the time of day or day of the week.

To conduct our analysis, we will utilize a dataset comprising various trips undertaken by different individuals. By examining these indicators within the dataset, we aim to gain insights into how gender influences transportation choices, travel distances, and trip purposes.

2 Data Analytics, Optimization and Policy Suggestions

2.1 Dataset

The dataset we used for this project is the MoTiV (Mobility and Time Value) dataset[12], that contains information about passengers and their trips with many different indicators. The data contained in this dataset are collected through a dedicated mobile application, called Woorti, used by users in 10 different countries: Belgium, Croatia, Finland, France, Italy, Norway, Portugal, Slovakia, Spain, and Switzerland. Data contained in the dataset spans from May 1st 2019 to December 13th 2019.

This app when downloaded requested some socio demographic information to the users and then recorded their trips (dividing each trip in multiple legs according to the different means of transports used) and after the trip requested some insights (like the purpose, activities performed during the trip, and other observations).

The dataset is constituted by 13 tables and we selected only the tables that contained variables of interest for our study. So, we considered the two datasets containing the trips and the legs info (trips.csv and legs.csv) and the legs coordinates (legs_coordinates.csv,), the one containing users' info (user_details.csv), the one with purpose information (purpose.csv) and the one with means of transports' informations (mots.csv).

This dataset is composed of various trips made by 3330 different users, 1880 men, 1435 women and 15 that didn't specify their gender. We decided to remove the records with gender labeled "Other" as, although representing an interesting phenomenon for future and more detailed studies, their numerosity and lack of further details make the category non-relevant in this case. The dataset is composed of 64,098 trips and 158,897 legs, resulting in, on average, 4.6 trips for user a day that are composed of, as average, 1.99 legs for females and 2.08 for males.

2.2 Data Exploration

2.2.1 Geographical Distribution

First of all we proceeded in analyzing the legs' coordinates dataset in order to plot a graphical representation of the records. Specifically we chose to plot the starting coordinates of each leg on an Europe map, plotted using the European Coastline shapefile[13]. (fig. 1). Each blue dot represents a man's leg and a red dot represents a woman's.

Observing geographical trip distribution, we can notice its dis-homogeneity inside each country, for example the Italian records are mainly distributed in the North part of the country, with almost no records from the islands and the South, so we must acknowledge that our analysis will be biased by this and each conclusion on a country shall be interpreted in relation with the regions where data were acquired.

Table 1 and fig. 2 indicates the number of records in the dataset for each different country and the ratio between females and males in all countries (with more than 50 users):

Country	Number of users	Number of Females	Number of Males	Number of trips
Belgium	321	103	218	6190
Spain	430	213	217	9285
Finland	341	258	83	4450
France	295	92	203	4168
Italy	274	75	199	4666
Norway	354	153	211	6262
Portugal	298	146	152	5490
Slovakia	520	199	321	10123

Countries with less than 100 different users weren't inserted in the table.

We can notice that the global female to male ratio is hardly recovered in each country, as the ratio varies greatly ranging from 0.12 to 2.43, this bias is caused by the modality with which the data are collected (voluntary installation and usage of an APP), this problem can be addressed by a more centralized, organized and detailed transport data acquisition by the institutions and transport agencies. This bias is taken into account

by the normalization applied to the data, thus although it may create artifacts in the analysis, its effect should be mitigated.

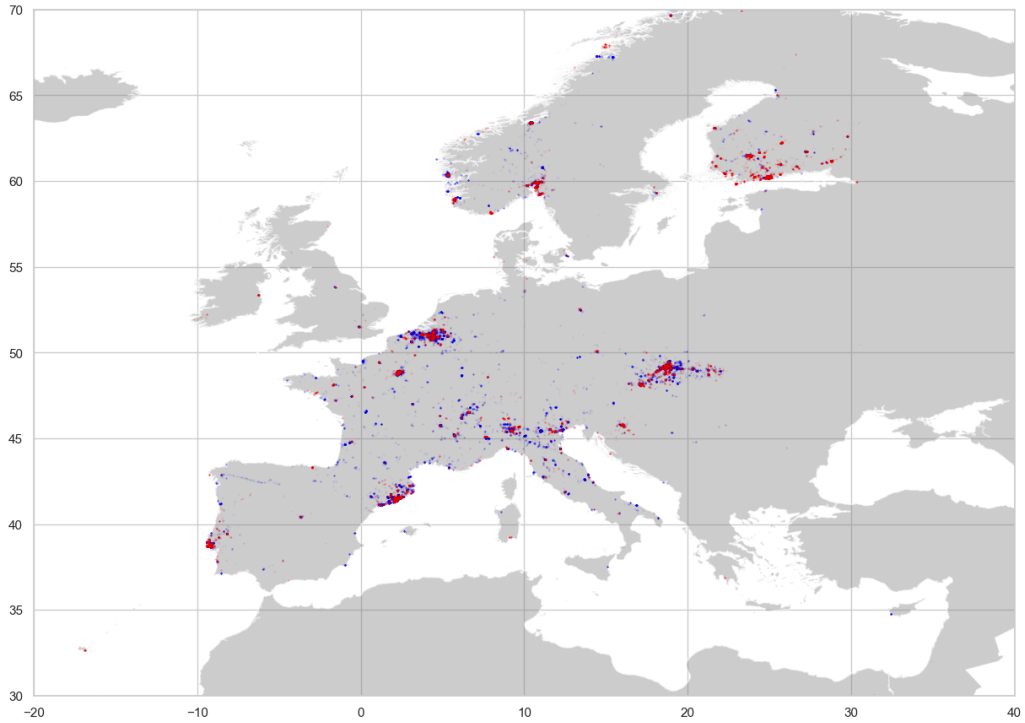

Figure 1: Map of all starting point of the observations

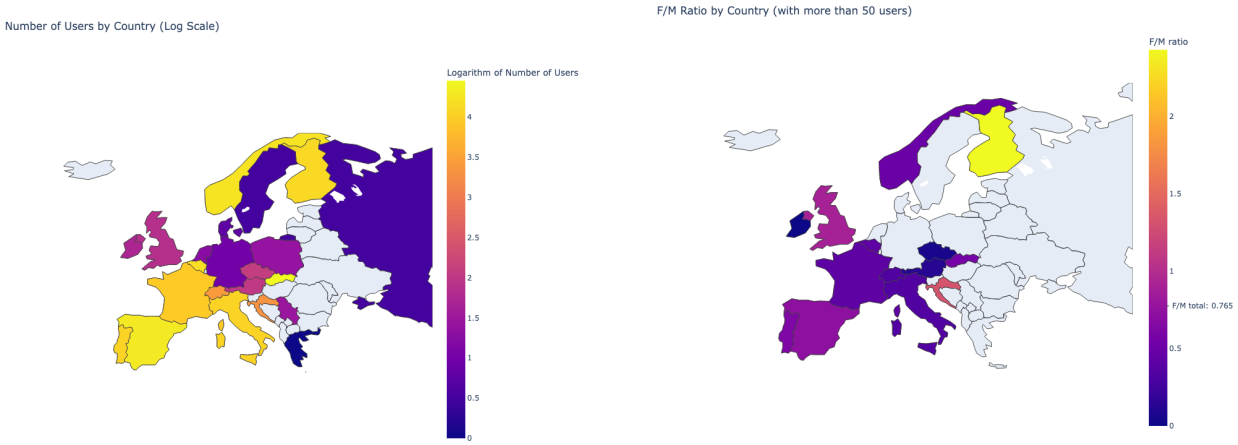

Figure 2: Choropleth of number of users (on the left) and of female to male ratio (on the right)

2.2.2 Purposes and means of transportation

To highlight possible transportation behavior differences among the genders, two variables of great interest are the means of transportation category and the purposes. This dataset contains 5 different transport categories: walking (55,246 records), private motorized (30,537 records), cycling and emerging micromobility (19,234 records), public transport short distance (14,510) and public transport long distance (1104) and 8 different purposes: *business trip* (6937 records), *everyday shopping* (17266), *home* (36,429),*leisure and hobby* (35,805), *personal tasks and errands* (30,646), *pick up and drop off* (14,237), *work* (45,914) and *other* (16,268). One of the critiques directed at this kind of surveys is that all the care tasks tend to be splitted in multiple different

categories, for example, errands, shopping, pick up and drop off, leading to the underreporting of this group of tasks [14]. For this reason, we decided to merge the categories of *home*, *everyday shopping*, *personal tasks*, *errands*, *pick up* and *drop off* in a new category named *care tasks*.

We studied the combined effect of the purpose of the trip and of the means of transportation. Notice that the normalization of data was done considering weighting each leg for its distance and separating the two genders. In this way all the data can be interpreted as an answer to the question: "what percentage of their daily movement used to go to work men/women perform with a private motorized vehicle?". The data normalized in this way don't allow to extract consideration about the total amount of movement performed by the different genders but how they use their movement regardless of its total length.

Then we proceeded in computing the same relative frequencies for each purpose and for clarity we decided to plot (fig. 3) the difference between the genders for each transport category divided by purpose.

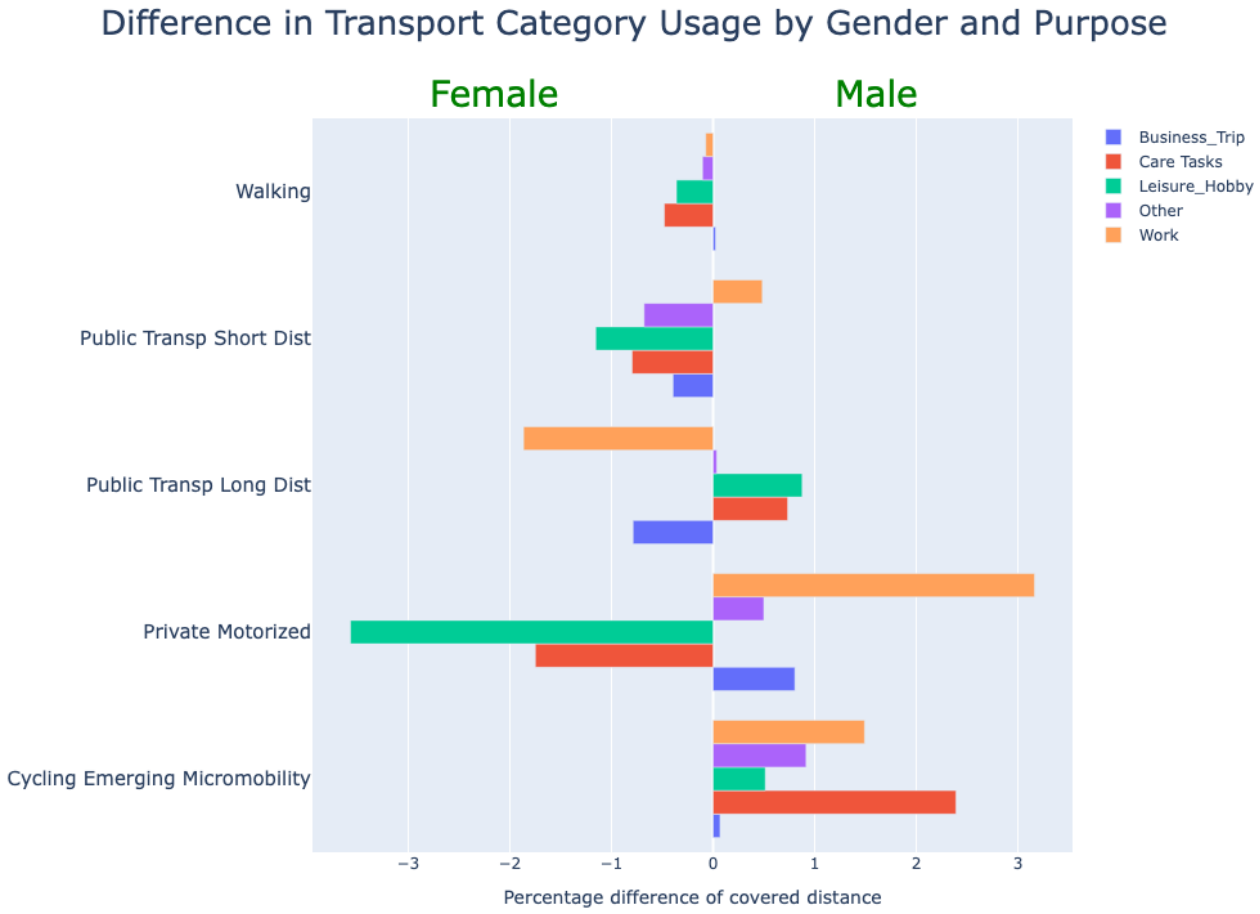

Figure 3: Bar plot of differences between percentage usage of different means of transport for different purposes for men and women

Notice that this plot represents the difference among the percentage of movement spent on a certain transport mean and with a certain purpose between males and females so, if we had a situation of symmetry of the distributions all the values would have been aligned on the 0 axis. The bars on the left represent the fact that women use more that mean of transportation for the specific purpose and the length of the bar represents the size of this difference. For example, we can see that men use more cycling and emerging micro mobility for all the purposes while women tend to walk for a higher fraction of their commutes. Men use cars or other forms of private motorized vehicles for work while women use them more for care tasks and hobbies. This kind of analysis and results is a key step in revealing the different needs of the two considered genders and the different

patterns in their car usage. Thus, we can expect a more intricate private transport mean usage pattern from women and a more predictable pattern, home-work commute, for men.

The reported difference in mobility patterns between genders indicates a state disequality that is revealed by considering the whole dataset, thus on all countries together. To study the geographical distribution of this phenomenon, we proceeded with the analysis per country, using the computed differences for each transport category to build a disparity index for each country by summing the absolute value of these differences, in other words the length of each bar reported in fig. 3.

In fig. 4 we show the logarithm of the just defined disparity index on a choropleth.

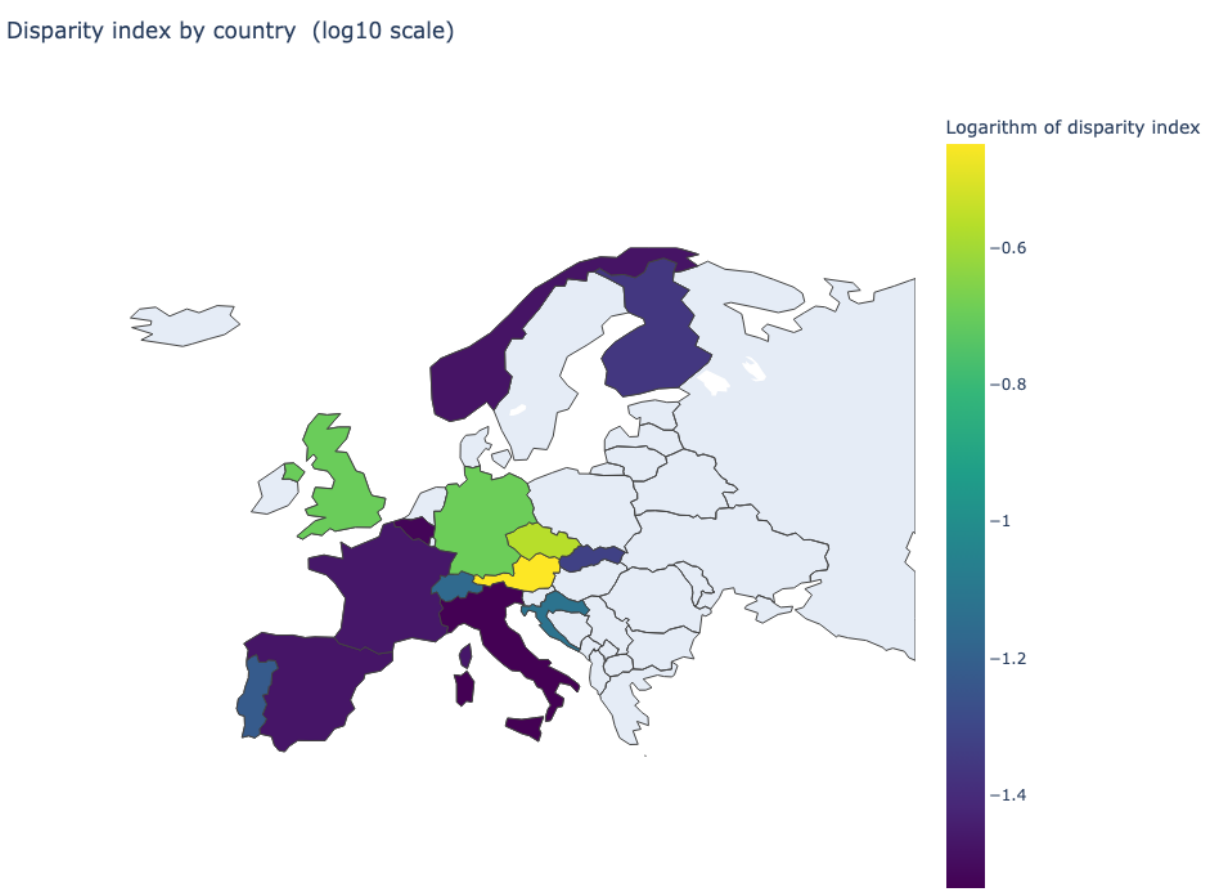

Figure 4: Choropleth of the logaritm of the disparity index

A log scale is used in order to better visualize the differences in the index.
We can notice a higher value of disparity index for Slovakia followed by Germany and Great Britain. However, crossing this data with plot reported in fig 2 we can notice that both these countries have very low user count, thus this index cannot be considered reliable. It is interesting to notice a lower index for Italy compared with Switzerland as both countries have an user count high enough to be considered representative. A broad and capillary investigation and data collection would help in building a more robust indicator of disparity to be used for policy making inside each country and at European level.

2.2.3 Time period analysis

Another meaningful indicator to study is the distribution of the time of the trips as it can highlight common patterns, for example the tendency to move in the morning and evening for work commutes, and the difference

in these common patterns between the genders. To perform this analysis, we based perform two trip hours aggregation accordingly to:, 24 hours division of the day and common time periods that share the same human activities

- Early Morning (5 AM - 8 AM)
- Morning (8 AM - 11 AM)
- Lunch (11 AM - 2 PM)
- Early Afternoon (2 PM - 4 PM)
- Afternoon (4 PM - 6 PM)
- Early Evening (6 PM - 8 PM)
- Evening (8 PM - 10 PM)
- Late Evening (10 PM - 12 AM)
- Night (12 AM - 5 AM)

The plot (fig. 5) shows the relative frequency of the different time periods for males and females and it shows only slight differences in the distribution, with a higher peak for women in the evening.

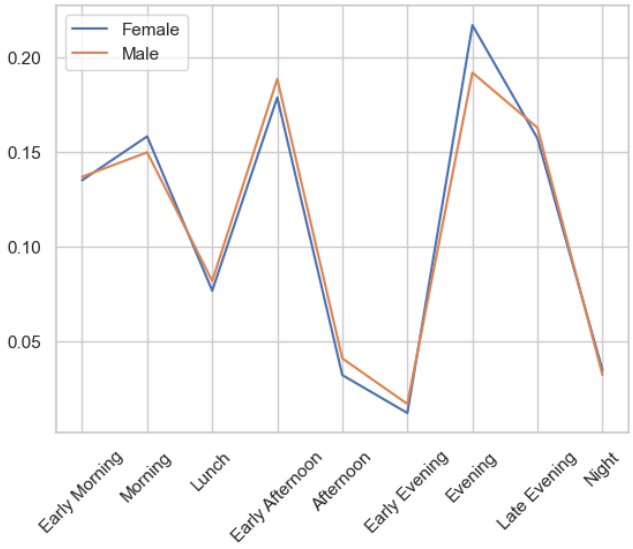

Figure 5: Line Plot of time period distribution divided by gender

A more detailed observation (based on the 24-hour breakdown) reveals a more pronounced difference between the two distributions (fig. 6)

We then choose two different approaches to try to better understand if the differences we can detect in the graph are statistically significant differences.

As different countries showed different "disparity index" we analyzed each country separately. In fig. 7 two examples are reported: in the first country considered (Italy) we can notice higher peaks in early morning and late afternoon for men and a flatter distribution in the morning for women, while Portugal (second plot), showed less evident difference between males and females movements during the day.

2.2.4 Statistical validation

To assess the real difference between males and females movement hour distribution it is important to know the variability behind the phenomenon, that means the difference between two, apparently different, distributions have, in reality, stochastic origin.

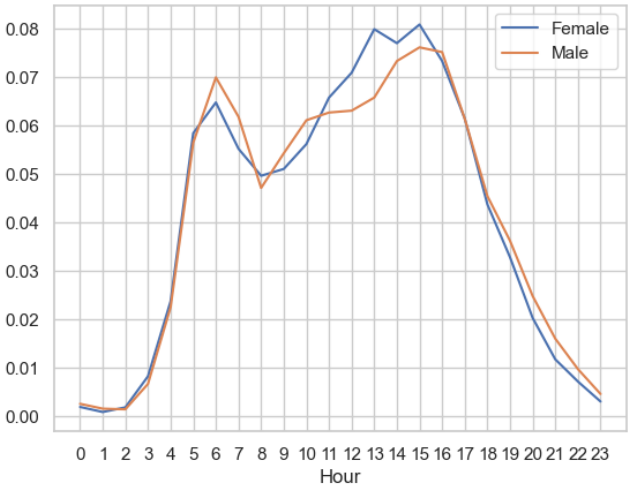

Figure 6: Line Plot of hour distribution divided by gender

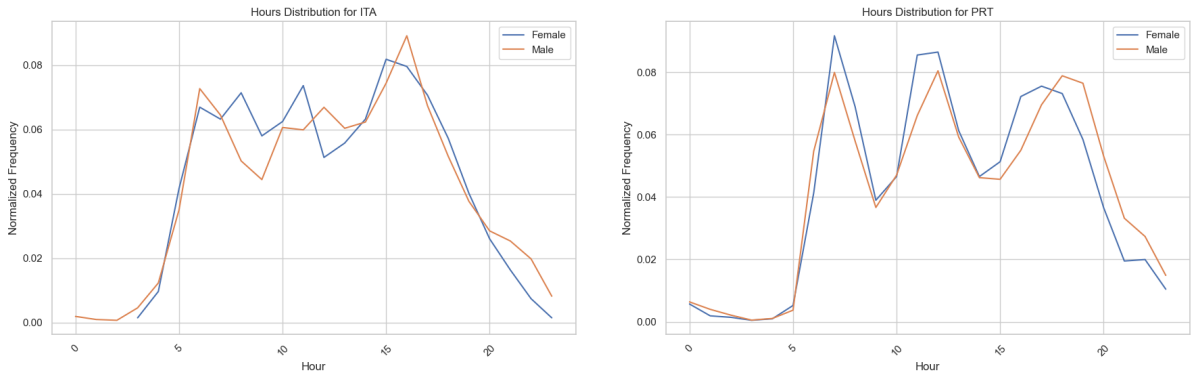

Figure 7: Line plot of hour distribution for Italy and Portugal

To validate the statistical difference between two distributions we firstly used a Shapiro-Wilk test, to confirm or reject the normality hypothesis. As expected, being generated from a sociocultural phenomenon and being time based, the distributions are not gaussians. For this region, we choose the Mann-Whitney test to accept or reject the null hypothesis stating that the two distributions are actually generated from the same phenomenon with no difference between genders.

This test was applied for all countries and with the sole exception of countries with only one gender represented, the results indicate a very very strong acceptance of the null hypothesis, thus that there are no differences between males and females in transport habits. However, although these conclusions could be used to paint an egalitarian scenario, they are spoiled by an insidious problem: the two distributions are constrained in range, between 0 and 23, moreover people have usually the same waking/sleeping habits, regardless of the gender and this should not affect this analysis.

For these reasons, a more robust approach was used:
Data are firstly divided by gender, creating males and females subsamples, then each subsample was divided in two 1000 times, computing each time the euclidean distance between the two resulting distributions. In this way a robust estimator of the intrinsic stochastic difference between homogeneous subsamples was built. The confidence interval at 95% was calculated using quantile methods on the bootstrapping just described. Finally, the real difference between males and females distributions is computed as euclidean distance, resulting in **0.0259**. The results indicate that the difference between the two genders are much bigger than the one explainable with the sole stochasticity (table 2), proving beyond any reasonable doubt the importance of this variable.

	Average value	95% interval
Male	0.0067	0.0045 - 0.0089
Female	0.0086	0.0057 - 0.012

2.3 Conclusions

Our study reaffirms the profound significance of accurate, representative data in comprehending the nuances of mobility dynamics, especially as they intersect with gender disparities. The noticeable biases of the dataset, such as certain countries being underrepresented, illustrate how incomplete data can distort our understanding of mobility patterns and can inadvertently steer policy decisions off course.

The present sample provided an unbalanced representation of gender, focusing predominantly on the binary genders and largely neglecting non-binary or other gender identities due to their limited representation in the dataset. This oversight stems from the data collection process itself, underlining the urgent need to broaden our data acquisition strategies to ensure inclusive representation of all gender identities. The geographic distribution of the collected data is equally critical. Our analysis showed an unequal distribution of data across regions, which highlights the necessity of well-balanced geographical data collection. The lack of sufficient data in certain regions creates knowledge gaps and biases, reinforcing the importance of comprehensive data collection efforts. Policy implications drawn from our analysis, while contingent on comprehensive and representative data, present important considerations for future transport planning and policymaking. For instance, the evident differences in mobility patterns between genders suggest that transport policies should be tailored to address the specific needs of different gender groups.

Women, for example, are often burdened with the 'double shift' of formal employment and unpaid domestic work, which includes tasks like shopping, child care, and elder care. Their transportation needs are therefore complex and multifaceted. Policies could be implemented to support these needs better, such as ensuring safe, affordable, and efficient public transportation that connects residential areas to essential amenities like childcare facilities, supermarkets, and healthcare centers. Men, on the other hand, typically exhibited more predictable patterns centered around work commutes. Transport policies could harness this pattern to encourage more sustainable commuting practices, such as promoting carpooling, cycling, or public transportation over private vehicle use.

The data also indicate a significant need to accommodate non-binary and other gender identities in transportation planning. Policymakers could leverage these insights to make transportation systems more inclusive and responsive to the needs of all genders. This might include designing safety measures that cater to the specific vulnerabilities of these groups or ensuring that public facilities are gender-inclusive.

In conclusion, comprehensive, balanced, and inclusive data is the bedrock of effective transport planning and policy. By harnessing such data, we can challenge entrenched gender roles, promote inclusivity, and strive towards a more equitable transport system that respects and caters to the diversity of all its users.

Code availability: The code used in this project is available at `https://github.com/zavidos/SmartCitiesLab`.

References

[1] L. Kern, "Feminist City: Claiming Space in a Man-made World," 2020.

[2] "WomenMobilizeWomen," https://womenmobilize.org/.

[3] F. Bonnet, "Women and Men in the Informal Economy – A Statistical Brief."

[4] H. Terraza, M. B. Orlando, C. Lakovits, V. Lopes Janik, and A. Kalashyan, "Handbook for Gender-Inclusive Urban Planning and Design," Feb. 2020.

[5] H.-P. Hsu, "How Does Fear of Sexual Harassment on Transit Affect Women's Use of Transit?" in *Transportation Research Board Conference Proceedings*, vol. 2, 2011.

[6] C. Keane, "Evaluating the Influence of Fear of Crime as an Environmental Mobility Restrictor on Women's Routine Activities," *Environment and Behavior*, vol. 30, no. 1, pp. 60–74, Jan. 1998.

[7] G. Lynch and S. Atkins, "The influence of personal security fears on women's travel patterns," *Transportation*, vol. 15, no. 3, pp. 257–277, Sep. 1988.

[8] K. Lucas and J. Machin, "Integrating crime and fear of crime initiatives with transport and accessibility improvements in deprived areas," in *Annual European Transport Conference 2004*, Strasbourg, France, 2004.

[9] "Why Gender Data Matters: A brief examination of the gender data gap in transport."

[10] V. Ceccato, "Women's victimisation and safety in transit environments," *Crime Prevention and Community Safety*, vol. 19, no. 3, pp. 163–167, Sep. 2017.

[11] Carl, "The Gender Indicators," https://www.itf-oecd.org/gender-indicators, Apr. 2023.

[12] C. Consonni, S. Basile, M. Manca, L. Boratto, G. Pourhashem, T. Kovacikova, A. Freitas, and Y. Cornet, "MoTiV: A Dataset of European User Mobility for Behavioral-Data," Sep. 2020.

[13] "Europe coastline shapefile — European Environment Agency," https://www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis-1/gis-data/europe-coastline-shapefile.

[14] I. S. de Madariaga, "Vivienda, movilidad y urbanismo para la igualdad en la diversidad: ciudades, género y dependencia," *Ciudad y Territorio Estudios Territoriales*, pp. 581–597, Dec. 2009.