9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Exploring urban income segregation with big geolocation data on human mobility

Keywords: mobile phone geolocations, income and ethnic segregation, mobility patterns

Extended Abstract

Cities face growing socio-spatial segregation between different income and ethnic groups. Innovative data-driven perspectives are crucial to understanding and reducing segregation and inequality. The study of social segregation has mainly focused on residential segregation [1] or the potential opportunities for interactions between groups [2].

Until recently, it has been challenging to capture the interactions/encounters between subpopulations at fine-grained spatio-temporal scales. The availability of big human mobility data [3] has offered new opportunities to describe the dynamics of socio-spatial segregation, accounting for individuals' daily activities from a large population, and thus going beyond the static view of housing segregation [2].

The rising research paradigm of socio-spatial segregation calls for developing new fine-grained segregation measures of urban spaces. However, the existing literature mainly focuses on the phenomenon itself, i.e., defining the concept and measuring it [2]. Urban spaces are visited by subpopulations of different income levels and ethnic backgrounds [3]. It remains elusive how the income segregation of different urban spaces shaped by its visitors is associated with the income levels and ethnic backgrounds of these visitors.

In this study, we seek answers to the below research questions (RQs):

1) How is *visiting* income segregation different from residential income segregation? Here, we define *visiting* segregation of a given place as the unevenness of income levels of its visitors.

2) What are the impacts of income level and ethnic background of visitors on visiting income segregation across different urban spaces?

Data. Our study is based on mobile application data, a source of anonymized population mobility data, consisting of GPS records collected through location-enabled applications installed on people's smartphones. Here, we use a dataset sourced from a diverse set of mobile apps used by adult smartphone users in Sweden[1] (Figures 1a-b). The dataset covers seven months in 2019 (June-Dec) with about 25 million daily GPS time records from a large number of devices. Assuming a device is a person, the population covered by this dataset is equivalent to c.a. 10% of the Swedish residents. After preprocessing, we use a subset that contains 215 active days with 22 million stay points from 136,050 devices. Around 23 days are recorded for an average individual (device) with two stays per recorded day.

Methodology. We first add weight to each recorded stay to reduce sampling bias, e.g., daytime has more locations than nighttime (Figure 1d). Next, we use temporal rules to detect the individuals' home areas [3] and find their corresponding census zones, according to which individuals are assigned socioeconomic attributes such as income distribution. We also give weight to the mobile phone users to make them more representative of the true population (Figures 1e-f).

We measure income segregation as the income unevenness [3] for each census zone, ranging between 0 and 1. Zero unevenness indicates an equal share of the population in the four net

[1] http://www.pickwell.co/

income quantile groups (defined for the Swedish population). And unevenness of one corresponds to a single-income group in the area. The income segregation (residential) is calculated using census statistics and assigned to the phone users by their home zone. Finally, we compute visiting income segregation by 1) extracting the co-existing visitors in hexagons at each half-hour interval and 2) calculating the weighted median values of these visitors' residential income group distribution (four quantiles) and the corresponding income unevenness, together with the shares of lowest income group and foreign background.

Results. We focus on an average non-holiday weekday in regions where the three biggest Swedish cities are located, e.g., the Stockholm area (Figure 2a). Income unevenness among visitors shows similar patterns to residential segregation. However, the areas with higher visiting segregation are more spread in Stockholm (red hexagons in Figure 2a). We define Wealthy regions as having the share of the lowest income group below 25%, while Less wealthy areas have the lowest income group above 25%. Visiting income segregation moderately correlates with residential segregation with a Spearman coefficient of 0.40 for wealthy regions and 0.30 for less wealthy areas ($p < 0.001$). The less wealthy areas have slightly but significantly higher visiting income unevenness than the residential value (Figure 2b, Kolmogorov-Smirniv test statistic = 0.20, $p < 0.001$).

Share of foreign background, as a proxy of ethnic groups, is associated with visiting income unevenness. In less wealthy regions, a moderate correlation exists between high income unevenness and a high proportion of the foreign-background population (0.31, $p < 0.001$). In wealthy areas, there is minimal correlation between these two dimensions. After the share of foreign background reaches 20%, the income segregation gap between the wealthy and less wealthy is deepened, i.e., less wealthy areas see a higher level of income segregation than the wealthy.

Discussion. The disparity between visiting and residential income segregation could be due to different mobility patterns among subpopulations, especially the spatial separation between their residence and daytime activities, highlighting the importance of using mobility data to measure socio-spatial segregation. Areas in Sweden with more foreign-background visitors tend to have higher income unevenness, especially in less wealthy areas. And the income segregation gap between the wealthy and less wealthy is deepened along with the increasing level of foreign background. This study will continue to refine the data processing and debiasing and develop robust models to better understand the relationship between socio-spatial segregation, mobility, built environment, and housing.

References

[1] Otis Dudley Duncan and Beverly Duncan. A methodological analysis of segregation indexes. *American sociological review*, 20(2):210–217, 1955.

[2] Qing-Quan Li, Yang Yue, Qi-Li Gao, Chen Zhong, and Joana Barros. Towards a new paradigm for segregation measurement in an age of big data. *Urban Informatics*, 1(1):5, 2022.

[3] Esteban Moro, Dan Calacci, Xiaowen Dong, and Alex Pentland. Mobility patterns are associated with experienced income segregation in large us cities. *Nature communications*, 12(1):4633, 2021.

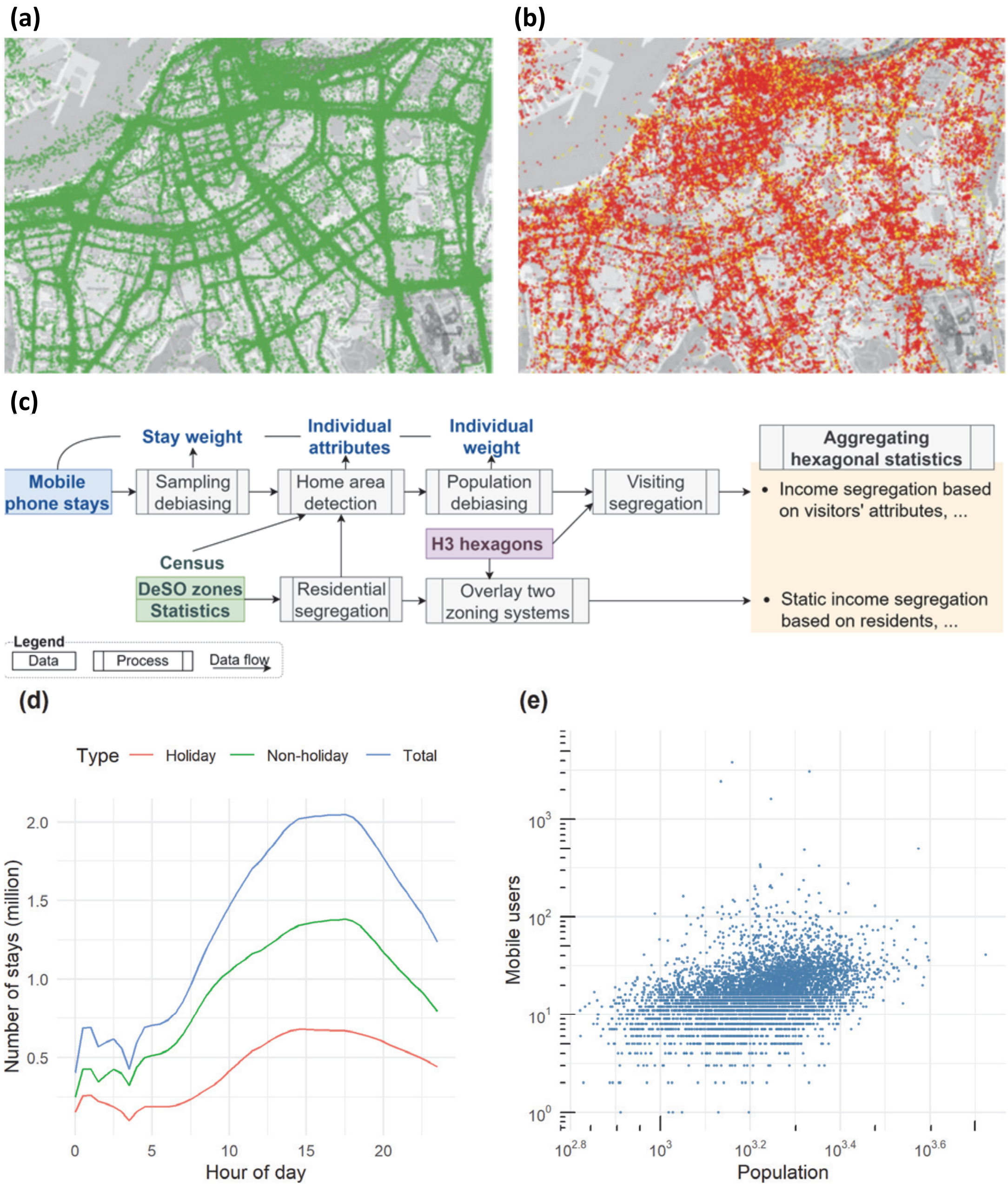

Figure 1: Data and methods. **(a)** Movement points and **(b)** Stay points of the dataset. Mobile application geolocations come from cell towers, GPS, and Wi-Fi sources. *Stays* are where individuals spend time in various activities while not moving, detected by aggregating devices' consecutive mobility traces if they are within 100 m and 1 h apart and have a displacement speed of less than 3 km/h. We filter out those devices with insufficient or abnormal data. Truncated GPS coordinates that point to unreliable locations are also removed. **(c)** Methodological framework. Census data are used with demographic zones (DeSO zones). H3 hexagons have a resolution of 6 with an area of 36.1 km^2. **(d)** Temporal distribution of mobile phone stays. Stay weight is calculated per user as the number of observed locations at a stay's half-hour time interval divided by the maximum number of observed locations among all time intervals over 24 hours. The stay weights mimic the even sampling of a person's whereabouts like we know where this person is all the time. **(e)** Number of users vs. true population size. Each point is a census zone. Individual weight is the reverse of the phone users' count ratio to the DeSO zone's true population.

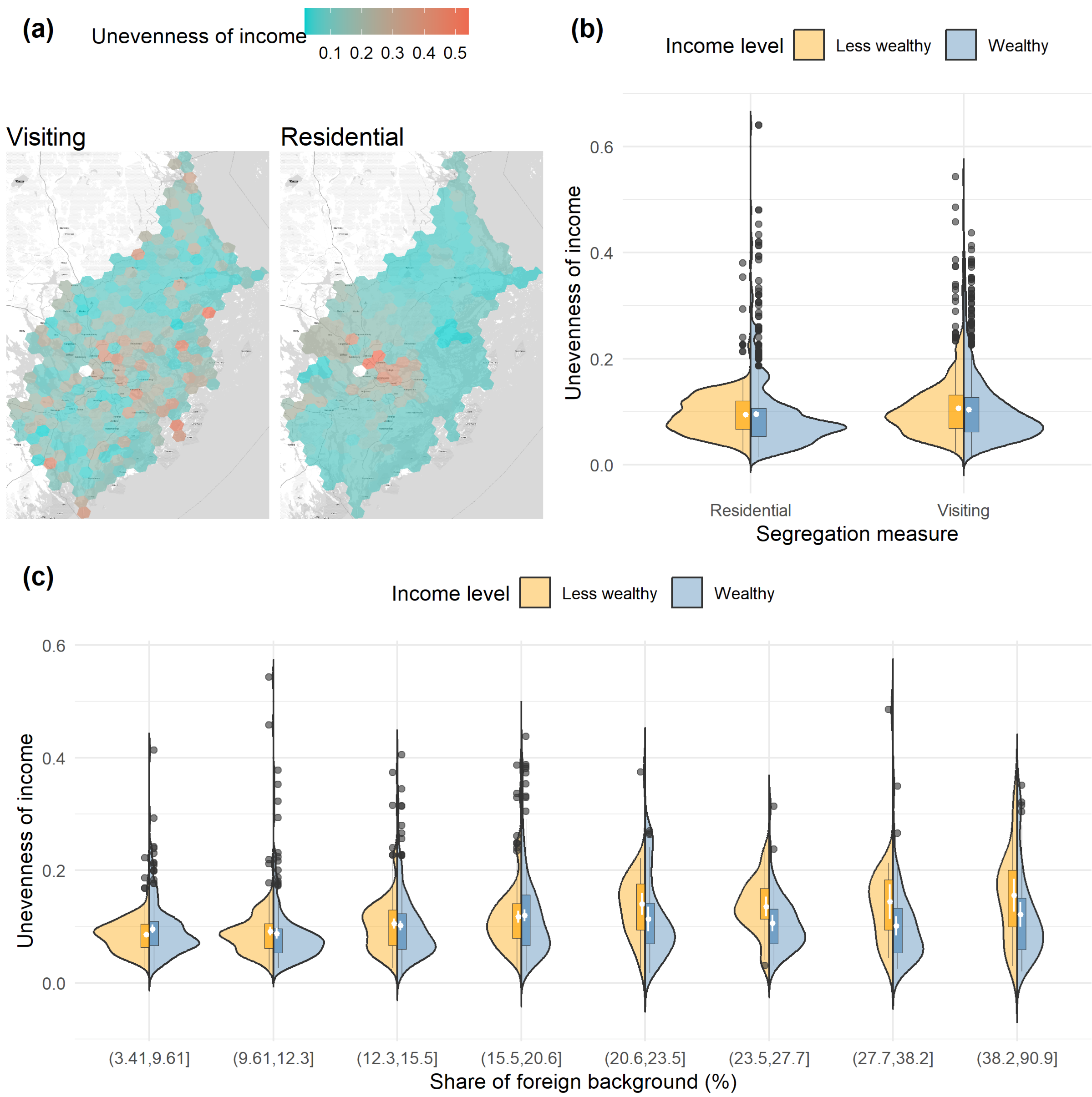

Figure 2: Income segregation at the hexagonal zone level. **(a)** Visiting income unevenness in Stockholm region. **(b)** Residential vs. Visiting income unevenness grouped by income level. White points with ranges indicate the mean value with a confidence interval of 95%. **(c)** Visiting income unevenness as a function of the share of foreign background, grouped by income level. Persons with a foreign background are defined as persons born abroad or born in Denmark with two foreign-born parents.