

Geographic Information Dataset of Mobile Phone User Portrait in Inner-city Nanjing (2020)

Huang, Q. S.^{1,2,3} Chen, J. Y.¹ Song, W. X.^{4*}

1. School of Civil Engineering and Architectural, Zhejiang University of Science and Technology, Hangzhou 3100232, China;

2. School of Architecture and Urban Planning, Tongji University, Shanghai 200092, China;

3. Zhejiang Jiaxing Digital City Laboratory Co. Ltd., Jiaxing 314050, China;

4. Key Laboratory of Watershed Geography, Nanjing Institute of Geography and Lakes, Chinese Academy of Sciences, Nanjing 210008, China

Abstract: The spatial distribution of urban residents can vary significantly over time. Using mobile phone user profile data and geographic location information, we extracted data on the spatial distribution of urban residents in inner-city Nanjing, focusing on their daytime activities and nighttime residences, categorized by time period. We classified each indicator variable of residents' mobile user profiles into seven levels—highest, high, medium-high, medium, medium-low, low, and lowest—using natural breaks (Jenks) and equal interval methods for both day and night, with communities as the spatial units. This was compiled into a Geographic Information System (GIS) dataset of mobile phone user portrait data for the inner-city of Nanjing in 2020. The dataset covers 125 communities and includes mobile phone user portrait data across eighteen typical dimensions, such as the proportion of young people, middle-aged people, older adults, males, females, married people, single people, parents of infants, primary school parents, middle school parents, high, medium, and low consumption levels, car owners, housing owners, employee workers, local residents, and non-local residents during the day and night in inner-city Nanjing. The data are archived in the form of .shp, .tif, and .txt, totaling 311 files with a data size of 1.32 MB (compressed into one file, 481 KB).

Keywords: Nanjing's inner-city; mobile phone portraits; daytime activities; nighttime residence

DOI: <https://doi.org/10.3974/geodp.2024.02.08>

CSTR: <https://cstr.science.org.cn/CSTR:20146.14.2024.02.08>

Dataset Availability Statement:

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: <https://doi.org/10.3974/geodb.2024.07.09.V1> or <https://cstr.science.org.cn/CSTR:20146.11.2024.07.09.V1>.

Received: 04-03-2024; **Accepted:** 10-06-2024; **Published:** 25-06-2024

Foundations: National Natural Science Foundation of China (42201251, 42171234); Young Scientist Foundation of Zhejiang University of Science and Technology (2023QN013)

***Corresponding Author:** Song, W. X., Nanjing Institute of Geography and Lakes, Chinese Academy of Sciences, wxsong@niglas.ac.cn

Data Citation: [1] Huang, Q. S., Chen, J. Y., Song, W. X. Geographic information dataset of mobile phone user profiling data in inner-city of Nanjing (2020) [J]. *Journal of Global Change Data & Discovery*, 2024, 8(2): 178–186. <https://doi.org/10.3974/geodp.2024.02.08>. <https://cstr.science.org.cn/CSTR:20146.14.2024.02.08>.

[2] Huang, Q. S., Chen, J. Y., Song, W. X. Geographic information dataset of mobile phone user profiling data in inner-city of Nanjing (2020) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. <https://doi.org/10.3974/geodb.2024.07.09.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2024.07.09.V1>.

1 Introduction

The mobility of urban populations and the diversification of social structures are key topics in current urban geography research^[1]. With the widespread adoption of new-generation information technologies and the advancement of a digital China, the interconnectedness of territorial space across the three dimensions—information, physical, and social—is becoming increasingly tighter due to networked, ubiquitous connections and intelligent sensing computing^[2]. Big data from mobile devices offers numerous potential applications in urban development research. For instance, Wang *et al.* compared traditional travel survey data with mobile signaling data in their research and highlighted the trends and challenges in enhancing the recognition accuracy of mobile signaling data in practical applications^[3]. Similarly, Niu *et al.* explored the heterogeneous characteristics of peri-urban migration for work and life in Wuhan city using cell phone signaling data^[4]. Mobile user profile data, as an emerging data source, provides deeper insights into residents' socioeconomic attributes and behavioral preferences. By analyzing the duration and types of APP usage on mobile devices combined with offline visitation scenarios, researchers can better understand patterns of population mobility and aggregation. This approach provides valuable insights into patterns of population mobility and aggregation, which have important theoretical and practical significance for studying the spatio-temporal behavior of residents' daily activities and for spatial optimization.

The varied activity patterns of residents at different times of the day create a diverse socio-spatial structure. During working hours, the population density of urban offices and leisure spaces increases significantly, while at night, residents mainly gather in residential areas^[5]. Mobile phone user profile data provides multi-dimensional spatial and temporal information on user behavior through GPS location, wireless networks (WIFI), IP addresses, internet logs, APP usage, and other sources, all collected under user authorization. This data, combined with the detection of users' stopping points in their offline daily activities, allows for the analysis of residents' social and economic attributes and behavioral patterns using model training and other methods. Economic attributes and behavioral patterns are labeled and clustered accordingly. Even though there are still some limitations in terms of information completeness and group coverage, mobile user profile data has the advantages of strong timeliness, high geographic spatial accuracy, and relatively easy accessibility. When traditional official statistical data is challenging to collect in real time, mobile user profile data can provide effective data support for relevant research in human geography and urban planning. In this study, we constructed a dataset of day and night spatial distribution characteristics for various types of users in the inner-city of Nanjing, based on mobile phone user portrait data and geographic location information.

2 Metadata of the Dataset

Table 1 summarizes the metadata of a geographic information dataset on mobile phone user portrait data in the inner-city of Nanjing (2020)^[6]. It includes the dataset's full name, short name, authors, year of the dataset, spatial resolution, data format, data size, data files, data publisher, and data sharing policy, etc.

3 Methods

This study utilizes mobile user profile data customized by the daily interactive “Getui” data intelligence service platform¹. By applying geographic spatial intelligence technology, it captures and deeply analyzes the online and offline behaviors of mobile users. It also constructs a multidimensional mobile user profile based on the user's geographic location,

¹ The daily interactive “Getui” data intelligence service platform. www.getui.com.

Table 1 Metadata summary of the geographic information dataset of mobile phone user portrait in inner-city Nanjing (2020)

Items	Description
Dataset full name	Geographic information dataset of mobile phone user portrait in inner-city Nanjing (2020)
Dataset short name	NJday_night2020
Authors	Huang, Q. S., Zhejiang University of Science and Technology, huangqinshi@zust.edu.cn Chen, J. Y., Zhejiang University of Science and Technology, 212302833008@zust.edu.cn Song, W. X. Nanjing Institute of Geography and Lakes, Chinese Academy of Sciences, wxsong@niglas.ac.cn
Geographical region	Nanjing inner-city (32°00'96"N–32°09'85"N, 118°73'91"E–118°82'64"E)
Year	2020
Spatial resolution	152 m×152 m
Data format	.shp, .tif, .txt
Data size	1.32 MB (before compression), 481 KB (after compression)
Data files	(1) StudyArea: Nanjing inner-city community-scale study area, NJcommunity.shp (2) TypicalAttribute: Nanjing inner-city community-scale mobile phone user portrait dataset with different attributes, naming rule "attribute + time".tif (3) Readme: description file on data attributes.txt
Foundations	National Natural Science Foundation of China (42201251, 42171234); Young Scientist Foundation of Zhejiang University of Science and Technology (2023QN013)
Data publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	(1) <i>Data</i> are openly available and can be free downloaded via the Internet. (2) End users are encouraged to use <i>Data</i> subject to citation. (3) Users, who are by definition also value-added service providers, are welcome to redistribute <i>Data</i> subject to written permission from the GCdataPR Editorial Office and the issuance of a <i>Data</i> redistribution license. (4) If <i>Data</i> are used to compile new datasets, the 'ten percent principal' should be followed such that <i>Data</i> records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset ^[7]
Communication and searchable system	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

personal attributes, behavioral characteristics, interest preferences, and application scenarios. The data is derived from user characteristics such as unique mobile device identification code ID (Identity Document, GID), timestamps, GPS location information, IP Internet logs, APP usage, and offline scene preferences, formed through model training and other methods. To address the current situation of residents using multiple cell phones from different carriers simultaneously, users with one GID corresponding to multiple International Mobile Equipment Identifiers (IMEIs) over an extended period of time are aggregated and processed. Compared to other mobile device data, this dataset’s spatial recognition accuracy is not constrained by the density of base stations or the type of operators. To some extent, it can provide more precise GPS location information for each user.

3.1 Study Area

Nanjing is a major center city in the Yangtze River Delta region, with its urban core area located within the inner-city area enclosed by the Ming City Wall. This study uses the community as the spatial unit and focuses on data collected in November 2020. It examines the spatial distribution data of urban residents in Nanjing’s inner-city during daytime activities and nighttime residences across different time periods. Based on the 2020 administrative divisions of Nanjing, the study area encompasses 125 community spatial units. However, data from the Taipingmen community showed abnormalities in several variables related to mobile phone user portraits and the day-night gap, which may not reflect the actual situation. Therefore, these were excluded from the analysis. The year 2020 was chosen as the year for data collection to facilitate comparisons with related studies using data from “The Seventh

Population Census.” November was specifically chosen as the month for data collection because it is free of major festivals that could disrupt daily activities, and the COVID-19 epidemic was relatively well-controlled, allowing Nanjing residents to resume normal life. During the data collection process, locations with a cumulative occurrence of 30 days in a month were chosen as either daytime activity areas or nighttime residences. In addition, peak hours for commuting and dining were excluded to avoid the influence of random events on the residents’ daily behavioral patterns during these times.

3.2 Data Collection and Processing

Mobile user profile data is obtained by providing mobile application developers with efficient messaging push technology services through SDKs. This allows information to be pushed to users, and with their authorization, data is collected from applications with location services enabled. This includes LBS data such as latitude, longitude, WIFI, and IP addresses, among others. By utilizing big data analysis tools, we can segment and describe the basic portrait of mobile phone users, categorizing them based on age group, gender, consumption level, house ownership, vehicle ownership, and other attributes. Based on the real-time location information of each mobile user, collected in November 2020, daytime activity is defined as occurring between 10:00 and 17:00, while nighttime residence is defined as occurring between 21:00 and 06:00 the next day. A user’s daytime activity location and nighttime residence location are determined by identifying the places where the user has the longest cumulative presence during these periods. On this basis, the spatial differentiation patterns and differences between daytime and nighttime social spaces in the city are analyzed.

The social characteristics of the group are determined through model training using data from three dimensions: the percentage of various types of APPs installed and active by users, offline scene preferences, and living environment. Age attributes are derived by training the model on users’ APP installations, activity, cell phone brand and model, and offline behavioral preferences. Local users are inferred from the identity information provided during APP registration and the city where users spent the most time during the Chinese New Year. The identification of parents of elementary school students is based on active APP installations and offline scene preferences related to tutorials. Consumption levels are determined through model training that analyzes APP installation activities, offline behavior, and living conditions. By combining this data with real credit card consumption samples, users’ consumption characteristics are inferred. High consumption level represents the top 20% of the population with the highest scores, while medium consumption level represents the 40% of the population in the middle range of scores. Scores ranked in the bottom 40% are categorized as low consumption levels. The employee group includes teachers, doctors, programmers, and other similar occupational combinations, which are identified primarily through APP usage behavior and offline scene preferences. In addition, the percentage of this group may be underestimated due to the lower penetration of smartphones among older age groups. Using the basic user profile data and geographic location information, Jenks natural breaks and classification methods for day and night are applied, with the community as the spatial unit. Statistics are analyzed by time period, and differences between day and night are calculated. This ends in the creation of a geographic information dataset for mobile phone user portraits in Nanjing’s inner-city (2020) (Figure 1).

4 Data Results and Validation

4.1 Data Composition

The geographic information dataset for mobile phone user portrait data for Nanjing’s inner-city (2020) includes community maps, street maps, and grid maps of the area. It features

data on mobile phone user portraits across 125 communities in Nanjing’s inner-city, encompassing six typical dimensions: daytime and nighttime population sizes, proportion of youth groups, proportion of locals, proportion of elementary school parents, proportion of high consumption, and proportion of employee groups. These dataset is archived in .shp, .tif, and .txt formats, totaling 311 files. The specifics of the data are shown in Figure 2.

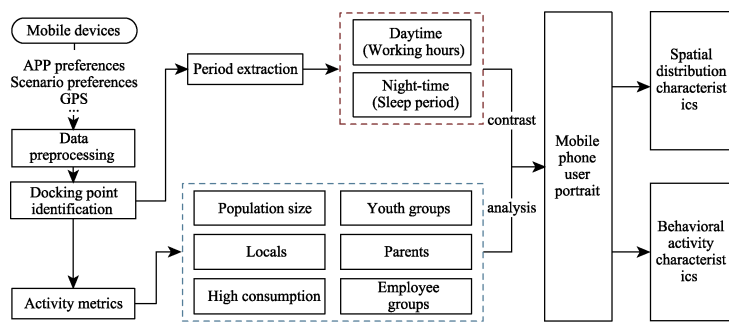


Figure 1 Mobile phone user portrait diagram of dataset development

4.2 Data Products

Figures 2 and 3 illustrate the mobile user profile data, showcasing the spatial distribution of various attribute groups in this study. These figures depict the daytime and nighttime spatial distribution of 18 different mobile user groups in each community unit, respectively. The groups include young, middle-aged, elderly, male, female, married, single, parents of infants and toddlers, elementary school parents, middle school parents, high-consumption, medium-consumption, low-consumption, vehicle-owning, homeowner, employee, local, and non-local groups. The title in the bottom left of each visualization represents the population attribute, while the legend in each cell represents the proportion of that group within the respective community’s population. The legend is categorized into seven levels using the natural break method and follows the same classification standards for day and night. By comparing the data distribution of different types of people at different time periods, these visualizations reveal the spatial characteristics of various social groups in Nanjing, providing insights into the needs and preferences of different socioeconomic groups.

4.2.1 Spatial Distribution Characteristics of Different Attribute Groups at Daytime

Figure 2 illustrates the daytime spatial distribution of mobile phone user profile data across various attribute groups in Nanjing’s inner-city area in 2020. Regarding age, young individuals are mainly concentrated in commercial areas such as Xinjiekou and Gulín, while middle-aged and older adults are more prevalent in the city’s peripheral areas. From a gender perspective, the spatial distribution of males and females is relatively balanced. However, there is a higher proportion of males in areas such as Yihe Road and Xianxia Road, whereas females are more concentrated in Xinjiekou. In terms of family structure, the spatial distribution of single and married groups shows minimal variation, with singles more common in peripheral areas like Suzhou Road South and Nanjing University. Families with primary and secondary school students are predominantly found in Xianxia Road and Gongjiao Village in the eastern part of the city, while families with infants and toddlers are mainly situated in Huguanguan and Yihe Road. Concerning consumption levels, high-consumption groups tend to be more active in commercial areas like Yihe Road in the city center during the day, whereas medium- and low-consumption groups are distributed across the city’s peripheral areas. The proportion of car owners is fairly evenly distributed across different areas, while the proportion of homeowners is notably lower in central areas of the city, such as Xinjiekou,

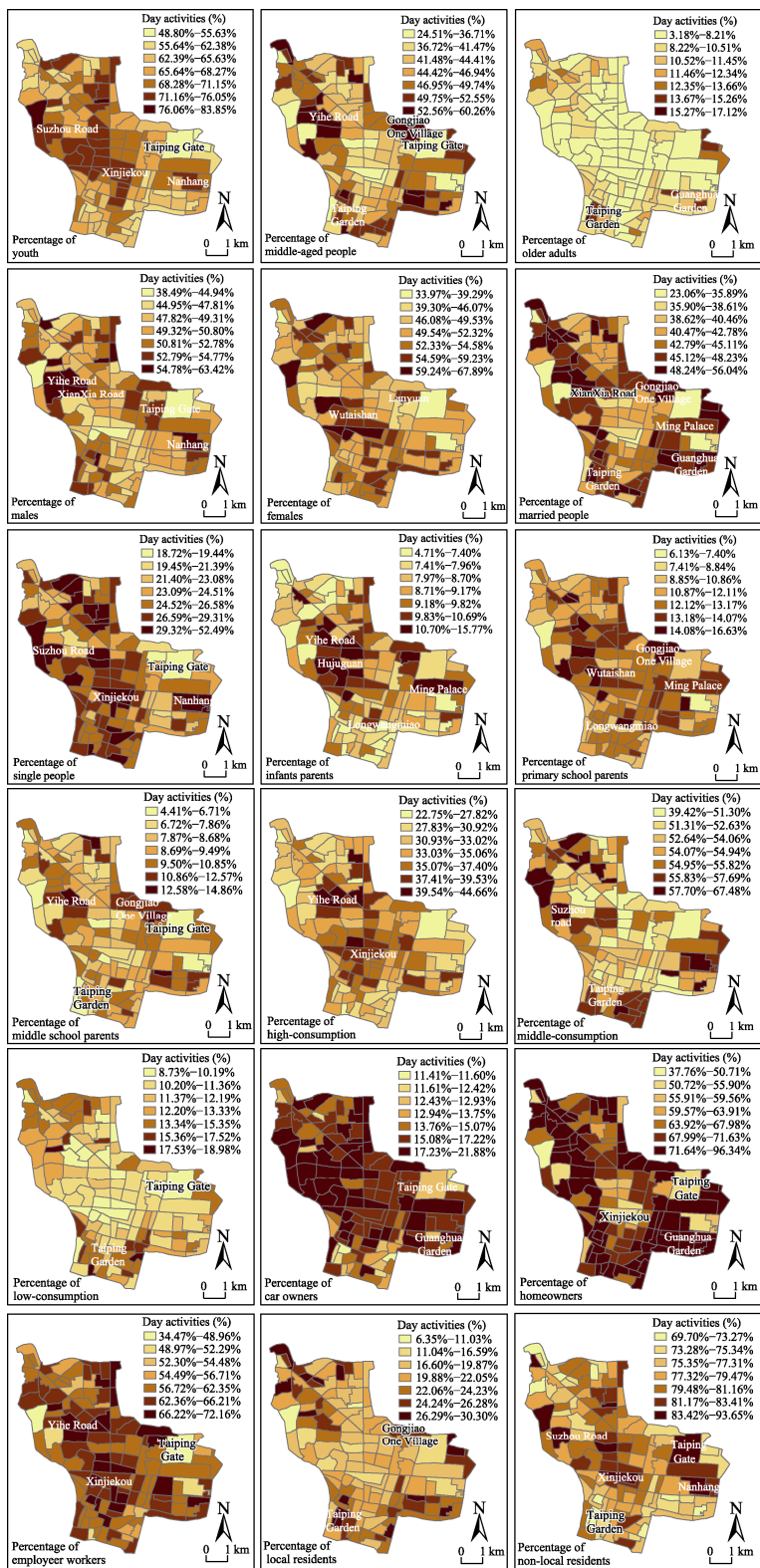


Figure 2 Visualization maps of the daytime spatial distribution of mobile user profile data for different attribute groups in Nanjing's inner-city area in 2020

compared to surrounding areas. Occupationally, there is minimal spatial differentiation among employee groups, and the proportion of people active during the day is relatively high overall. The population proportion is notably higher in areas such as Yihe Road in the city center and Gongjiao Village in the eastern part of the city, while it is comparatively lower in places such as Taiping Gate and Gulin. In terms of household registration status, local residents are predominantly located in the peripheral areas of the inner-city. Consequently, the proportion of local residents engaging in daytime activities in the city's central areas like Xinjiekou is relatively low.

4.2.2 Spatial Distribution Characteristics of Different Attribute Groups at Nighttime

Figure 3 illustrates the nighttime spatial distribution of mobile phone user profile data across various attribute groups within Nanjing's inner-city area in 2020. In terms of age distribution, young people tend to concentrate their nighttime activities in areas like Suzhou Road, Qingdao Road, and Nanjing University of Aeronautics and Astronautics, whereas the distribution of nighttime residential populations in other parts of the city is relatively even. Conversely, the middle-aged and elderly groups are more likely to be found in urban fringe areas such as Gongjiao Village and Guanghua Park. In terms of gender, the spatial distribution of males and females is relatively balanced overall. However, the proportion of males is higher in the Suzhou Road area, while females are more concentrated in Gulin and Xinjiekou. From the family structure perspective, there is a notable difference in the nighttime spatial distribution between the married group and the single group. The married group is predominantly located in the northeastern part of the city, including areas like Yihe Road and Gongjiao Village. On the other hand, the single group is more likely to cluster in the western part of the city, such as Suzhou Road and Gulin. Families with infants/toddlers, primary school students, and middle school students have a higher degree of nighttime spatial clustering, particularly in areas such as Lanyuan, Xianxia Road, Gongjiao Village, and Wutaishan, which are closely related to the high-quality basic education resources. From a consumption level perspective, the high-consumption group is mainly concentrated in the city center and commercial areas, with minimal significant nighttime residential spatial differentiation. In contrast, the medium and low-consumption groups are more densely clustered in the urban fringe areas, such as Gulin and Nanjing University of Aeronautics and Astronautics. The nighttime spatial differentiation of car owners and homeowners is relatively similar, with noticeable differences. The proportion of car owners and homeowners is relatively high in the Gongjiao Village and Taipingmen areas in the northeastern part of the city, as well as in the Gulin area in the western part of the city. In terms of occupations, the spatial distribution of the nighttime residential population of the employee group is relatively balanced, with a higher proportion residing in the northern part of the city, particularly the Gongjiao Village community. Regarding household registration status, local residents exhibit more pronounced spatial differentiation in nighttime residential areas, with a higher proportion in the older urban areas of the north, south, and east. Non-local residents, on the other hand, have a higher proportion in the city's center areas such as Xinjiekou and the northwestern areas such as Gulin.

5 Discussion and Conclusion

Leveraging mobile user profiles, this new data source has obvious advantages, including broader sample coverage, enhanced spatio-temporal data accuracy, and enriched socioeconomic dimensions. This study outlines the comprehensive mobile user profile characteristics of various social groups in Nanjing's inner-city in 2020. It does so by analyzing WIFI, GPS location information, IP browsing logs, and user APP usage data from mobile phone users in Nanjing's inner-city in 2020 and integrating these with users' offline preferences and location service data. In addition, it has compiled the Geographic Information Dataset of Mobile

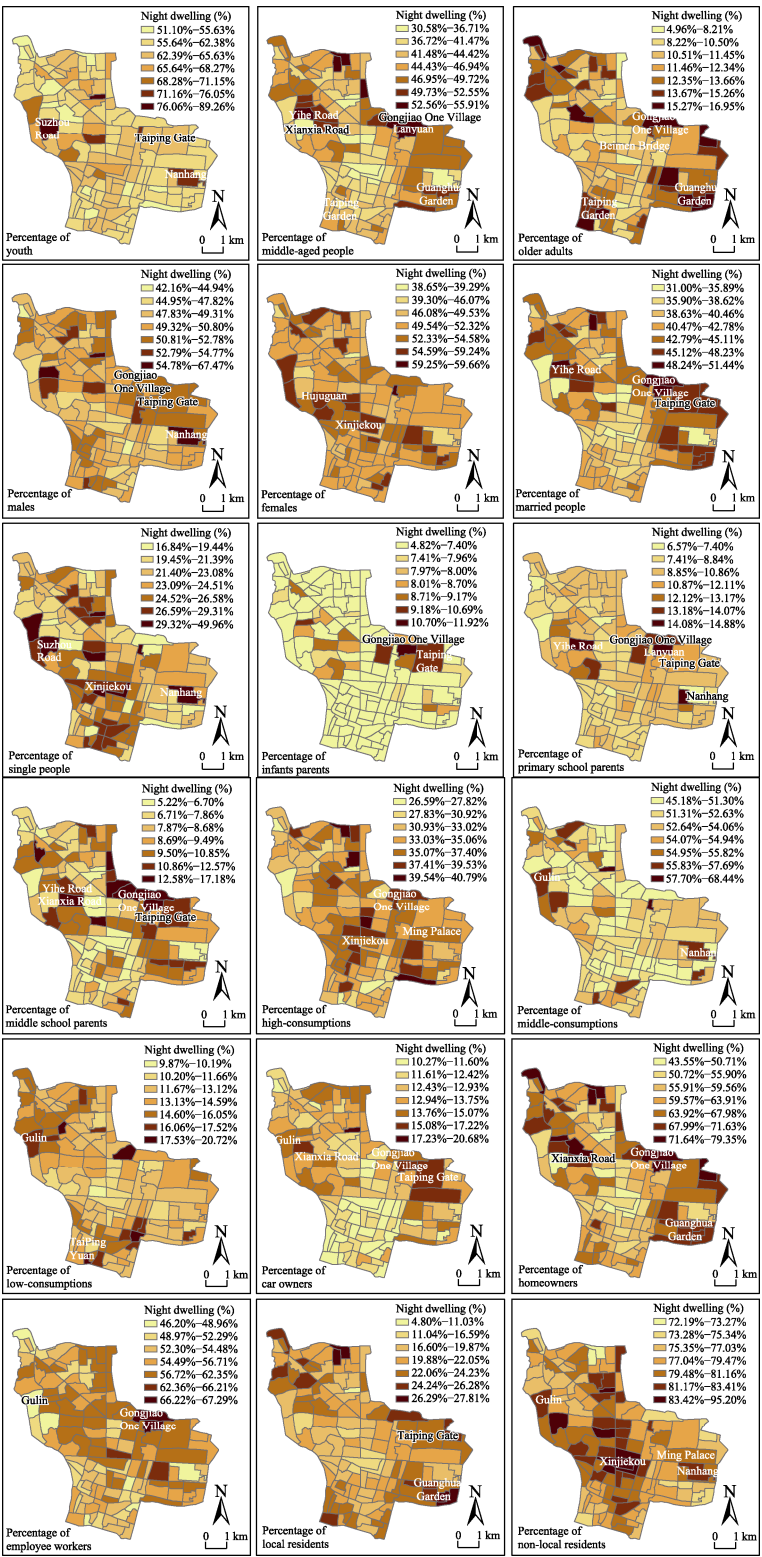


Figure 3 Visualization maps of the nighttime spatial distribution of mobile user profile data for different attribute groups in Nanjing's inner-city area in 2020

Phone User Portraits in Nanjing's inner-city (2020). Specifically, the dataset comprises 18 different cell phone user groups, including youth, middle-aged, elderly, male, female, married, single, parents of infants and toddlers, parents of elementary school students, parents of secondary school students, high-consumption, middle-consumption, low-consumption, vehicle-owning, house-owning, employee, local, and non-local groups, along with their spatial distributions during both daytime and nighttime in each community. The geographic spatial analysis of this dataset shows that Nanjing's inner-city area features a "core-periphery" spatial structure, with Xinjiekou acting as the urban core. Xinjiekou and its surrounding areas are central hubs for business and commercial activities, attracting a large number of high-consumption groups, youth, employee workers, and non-local residents. In contrast, the primary residential areas are concentrated in the northern, southern, and eastern peripheries, with a notably high proportion of local residents in the urban center and increased population movement at night. The northern, southern, and eastern peripheral areas are the main residential zones of Nanjing's inner-city. This dataset provides valuable insights into the spatial activity patterns of different social groups within Nanjing's inner-city, enhancing our understanding of the current status and future development trends of urban spatial development. This can provide valuable insights for urban social space and behavioral planning research. Due to current limitations in analysis techniques and research capabilities, there is still room for improvement in the accuracy and reliability of the results obtained from training the dataset model. However, in the absence of real-time official statistical data, this dataset provides valuable multidimensional references and data support for urban social geography research.

Author Contributions

Song, W. X. designed the algorithms for the dataset. Huang, Q. S. and Chen, J. Y. contributed to the data processing and analysis. Huang, Q. S. and Chen, J. Y. wrote the data paper.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Gitelson, A. A., Kaufman, Y. J., Stark, R., *et al.* Novel algorithms for remote estimation of vegetation fraction [J]. *Remote Sensing of Environment*, 2002, 80(1): 76–87.
- [1] Huang, Q. S., Zhou, Q., Song, W. X. Multidimensional steering and scale response in the study of urban residential differentiation in the new era [J]. *Progress in Geography*, 2023, 42(3): 573–586.
- [2] Zhen, F., Yuan, C., Zhang, S. Q., *et al.* A study on the path of smart land enabling high-quality development of cities: a case study of Chongqing [J]. *Journal Spatio-temporal Information*, 2024, 31(2): 1–13
- [3] Wang, D., Han, B. L., Zhang, T. R., *et al.* Accuracy analysis of mobile signaling data in measuring travel indices: based on the comparison with household travel survey [J]. *Progress in Geography*, 2024, 43(5): 854–869.
- [4] Niu, Q., Wu, L., Sheng, F. B., *et al.* Measuring the dynamic balance of jobs and housing in Wuhan's suburban new towns based on individual job and housing migration [J]. *Acta Geographica Sinica*, 2023, 78(12): 3095–3108.
- [5] Song W. X., Xu, D., Wang, J. K., *et al.* Spatial differentiation of daytime and nighttime society in Nanjing inner-city based on cell phone portrait data [J]. *Acta Geographica Sinica*, 2024, 79(2): 421–438.
- [6] Huang, Q. S., Chen, J. Y., Song, W. X. Geographic information dataset of mobile phone user portrait data in the inner-city of Nanjing (2020) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. <https://doi.org/10.3974/geodb.2024.07.09.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2024.07.09.V1>.
- [7] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (Updated 2017).