

南京内城区手机用户画像地理信息数据集（2020）

黄琴诗^{1,2,3}, 陈家裕¹, 宋伟轩^{4*}

1. 浙江科技大学土木与建筑工程学院, 杭州 3100232; 2. 同济大学建筑与城市规划学院, 上海 200092;
3 浙江嘉兴数字城市实验室有限公司, 嘉兴 314050;
4. 中国科学院南京地理与湖泊研究所流域地理学重点实验室, 南京 210008

摘要: 城市居民的空间分布随时间变化可能存在显著差异, 基于手机用户画像数据与地理位置信息, 分时段提取了南京内城区居民在日间活动与夜间居住的空间分布数据。采用 Jenks 自然断点和日夜相同分级方法, 以社区为空间单元, 将居民手机用户画像各项指标变量分别划分为最高、高、中高、中、中低、低和最低 7 个等级, 整编形成南京内城区手机用户画像的地理信息数据集 (2020)。该地理信息数据集包括: 南京内城区日间和夜间的青年群体占比、中年群体占比、老年群体占比、男性群体占比、女性群体占比、已婚群体占比、单身群体占比、婴幼儿群体家长占比、小学家长占比、中学生家长群体占比、高消费占比、中消费群体占比、低消费群体占比、有车群体占比、有房群体占比、职员群体占比、本地群体占比、外地群体占比 18 个典型维度在 125 个社区的手机用户画像数据。数据存储为 .shp、.tif、.txt 数据, 共 311 个文件, 数据量为 1.32 MB (压缩为一个文件, 481 KB)。

关键词: 南京内城; 手机画像; 日间活动; 夜间居住

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

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

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2024.07.09.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2024.07.09.V1>.

1 前言

城市人口的流动性和社会结构的多元化是当前城市地理学研究的热点之一^[1], 随着新一代信息技术的深度应用和数字中国的推进, 网络化的泛在关联和智能化的感知计算使得国土空间在“信息—物理—社会”三个维度的耦合度日益紧密^[2]。移动设备大数据在城市发展研究中具有广泛的应用潜力, 王德等通过与传统出行调查数据的比较研究, 提出在实践过程中提高手机信令数据识别精度的趋势与挑战^[3], 牛强等基于手机信令数据揭示了武

收稿日期: 2024-03-04; 修订日期: 2024-06-10; 出版日期: 2024-06-25

基金项目: 国家自然科学基金 (42201251, 42171234); 浙江科技学院青年科学基金 (2023QN013)

*通讯作者: 宋伟轩, 中国科学院南京地理与湖泊研究所, wxsong@niglas.ac.cn

数据引用方式: [1] 黄琴诗, 陈家裕, 宋伟轩. 南京内城区手机用户画像地理信息数据集 (2020) [J]. 全球变化数据学报, 2024, 8(2): 178–186. <https://doi.org/10.3974/geodp.2024.02.08>. <https://cstr.escience.org.cn/CSTR:20146.14.2024.02.08>.

[2] 黄琴诗, 陈家裕, 宋伟轩. 南京内城区手机用户画像地理信息数据集 (2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2024. <https://doi.org/10.3974/geodb.2024.07.09.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2024.07.09.V1>.

汉市职住迁移的城郊异质性特征^[4]。手机用户画像数据作为一种新兴的数据来源,通过挖掘移动设备用户的 APP 使用时长与类型等特征,结合线下到访空间场景,深入刻画居民多维度的社会经济属性与行为偏好特征,有助于深入理解人口的流动和集聚模式,对研究居民日常活动的时空行为特征与空间优化具有重要的理论和实践意义。

居民在不同时段差异化的活动模式使得社会空间结构呈现出多元化的特征,工作时段城市办公空间和休闲空间的人口密度显著增加,而夜间时段居民主要集聚在居住空间^[5]。手机用户画像数据是在用户授权的情况下,通过用户的 GPS 位置信息、无线网络 WIFI、互联网协议 IP、上网日志、APP 使用情况等数据获取多维立体的用户时空行为信息,结合用户线下日常活动中的停驻点探测,通过模型训练等方式分析对居民的社会经济属性和行为模式进行标签与聚类。虽然在信息完整度和群体覆盖度等方面尚存在一些不足,但手机用户画像数据具有时效性强、地理空间精度高且相对容易获取等优势,在传统官方统计数据难以实时采集的情况下,可为人文地理与城市规划等相关研究提供有效的数据支撑。本研究基于手机用户画像数据与地理位置信息,构建了南京市内城区不同类型用户日夜间空间分布特征数据集。

2 数据集元数据简介

《南京内城区手机用户画像地理信息数据集(2020)》^[6]的名称、作者、地理区域、数据年代、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表1。

3 数据研发方法

3.1 数据来源

本研究采用每日互动“个推”数据智能服务平台¹定制的手机用户画像数据,运用地理空间智能技术捕捉并深度分析手机用户线上和线下行为,结合用户地理位置、个人属性、行为特征、兴趣偏好和应用场景等构建多维度的手机用户画像。该数据基于移动设备唯一识别码 ID (Identity Document, Gid)、时间戳、GPS 位置信息、IP 上网日志、APP 使用情况等用户特征数据以及用户线下场景偏好等数据,通过模型训练等手段所形成的数据。针对当前居民同时使用多个运营商多部手机的情况,将较长时间段内一个 Gid 对应多个国际移动设备识别码(IMEI)的用户进行聚合处理。与其他移动设备数据相比,该数据集的空间识别精度不受基站密度和运营商类型限制,一定程度能够较为准确地定位到每个用户的 GPS 位置信息。

南京是长三角地区的重要中心城市,其城市核心区为以明城墙围合的内城区域。研究以社区为空间单元,选取2020年11月作为数据采集时期,按照不同活动时段提取南京内城区居民在日间活动与夜间居住的空间分布数据。按照2020年南京行政区划,研究区域内包含125个社区空间单元。其中,太平门社区手机用户画像中多项变量及其日夜差距异常,可能与真实情况存在偏差,故予以剔除。选择2020年作为数据采集年份,有助于与采用“七

¹ 每日互动“个推”数据智能服务平台. www.getui.com.

表 1 《南京内城区手机用户画像地理信息数据集（2020）》元数据简表

条 目	描 述
数据集名称	南京内城区手机用户画像地理信息数据集（2020）
数据集短名	NJday_night2020
作者信息	黄琴诗, 浙江科技大学, huangqinshi@zust.edu.cn 陈家裕, 浙江科技大学, 212302833008@zust.edu.cn 宋伟轩, 中国科学院南京地理与湖泊研究所, wxsong@niglas.ac.cn
地理区域	南京市内城区（32°00'96"N-32°09'85"N, 118°73'91"E-118°82'64"E）
数据年代	2020
空间分辨率	152 m×152 m
数据格式	.shp、.tif、.txt
数据文件	311 个（压缩前），1 个（压缩后）
数据量	1.32 MB（压缩前），481 KB（压缩后）
数据集组成	（1）StudyArea：南京市内城区社区尺度研究范围，NJcommunity.shp （2）TypicalAttribute：南京市内城区社区尺度不同属性手机用户画像数据集，命名规则“属性+时间”.tif （3）Readme：关于数据属性的说明文件.txt
基金项目	国家自然科学基金（42201251，42171234）；浙江科技学院青年科学基金（2023QN013）
数据计算环境	ArcGIS
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10%引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 ^[7]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

普”数据开展的相关研究进行比较。选择 11 月为采集期月份，因为该月没有干扰日常活动的重大节庆，且疫情相对得到有效控制，南京居民基本恢复正常生活。在数据采集过程中，选择 1 个月内 30 天累计出现最久的位置作为日间活动或夜间居住地，并剔除了上下班交通和早晚用餐高峰时段，以避免这些时段随机偶发事件对居民日常行为规律的干扰。

3.2 研发流程

手机用户画像数据是通过为移动应用开发者提供高效的消息推送技术服务,采用 SDK 方式完成信息推送同时，在用户授权的情况下，依赖于开启位置服务的应用程序签到和上报，获取用户经纬度、WIFI、IP 等 LBS 数据。借助大数据分析工具对手机用户基本画像进行细分与描述，可以将手机用户根据年龄段、用户性别、用户消费水平、是否拥有房屋、是否拥有机动车等标签进行分类。根据每位手机用户的实时位置信息，以 2020 年 11 月为数据采集时段，以 10:00 至 17:00 为日间活动时段，21:00 至次日 06:00 为夜间居住时段，基于用户在这些时段内累计出现时间最长的位置，分别确定其日间活动地点和夜间居住地点，在此基础上分析形成城市日夜间社会空间分异格局及差异。

通过用户安装的不同 APP 类型占比及活跃情况、线下场景偏好与居住环境三个维度的数据进行模型训练获得群体社会特征。其中，年龄属性通过用户 APP 安装、活跃、手机品牌型号、线下行为偏好等进行模型训练获得。本地用户根据注册应用程序时提供的身份信息，以及春节期间居留时间最长的城市推断。小、中学生家长通过教辅类 APP 安装活跃和线下场景偏好计算。消费水平通过 APP 安装活跃、线下行为、居住环境等进行模型训练打分，结合用户真实刷卡消费标签样本推断用户的消费特征，高消费水平代表分值最高的 20% 人群，中消费水平代表分值中间段的 40% 人群，分值排在后 40% 的为低消费水平。职员群体包括教师、医生、程序员等职业组合，主要通过 APP 使用行为和线下场景偏好获得。有车有房群体通过 APP 安装活跃、线下行为、线上搜索记录等行为识别获取。此外，由于智能手机在老年群体中的普及度较低，因此这部分人群的占比可能会低于实际情况。根据用户基本画像数据与地理位置信息，采用 Jenks 自然断点和日夜相同分级方法，以社区为空间单元，分时间段进行统计并计算日夜间差值，最终形成南京内城区手机用户画像地理信息数据集（图 1）。

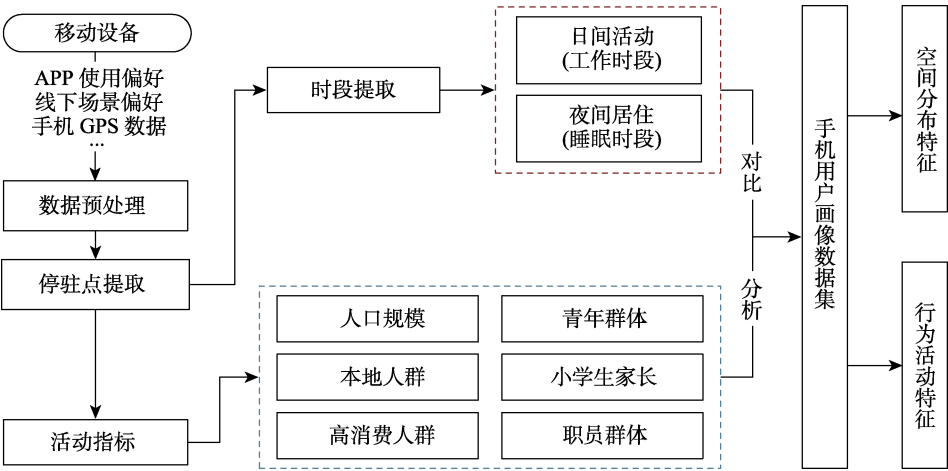


图 1 手机用户画像数据获取流程示意图

4 数据结果与验证

4.1 数据集组成

《南京内城区手机用户画像地理信息数据集（2020）》中数据底图包括南京市内城区社区尺度的矢量图。数据内容包括南京内城区日间和夜间的青年群体、中年群体、老年群体、男性群体、女性群体、已婚群体、单身群体、婴幼儿群体家长、小学家长、中学生家长群体、高消费、中消费群体、低消费群体、有车群体、有房群体、职员群体、本地群体、外地群体等 18 个典型维度在 125 个社区的手机用户画像数据。数据存储为.shp、.tif、.txt 数据，共 311 个文件，数据结果见图 2、图 3。

4.2 数据结果

图 2、图 3 为本次研究所展示的不同属性群体空间分布的手机用户画像数据可视化图，

分别展示了 2020 年包括青年群体、中年群体、老年群体、男性群体、女性群体、已婚群体、单身群体、婴幼儿群体家长、小学家长、中学生家长群体、高消费群体、中消费群体、低消费群体、有车群体、有房群体、职员群体、本地群体、外地群体等 18 类不同手机用户群体在各个社区单元中的日间与夜间空间分布情况。可视化图左下角标题代表人群属性，单元格对应的图例代表该类型群体在所属社区人口中的占比，采用自然断点法和日夜相同分级标准将图例划分成为七级。可视化图通过比较不同类型群体在不同时间段的数据分布，呈现南京市不同类型的社会群体的空间特征，有助于探测不同社会经济群体的需求和偏好。

4.2.1 不同属性群体日间空间分布特征

图 2 展示了 2020 年南京市内城区不同属性群体手机用户画像数据的日间空间分布情况。从年龄分布来看，青年群体主要聚集于新街口和古林等商业化程度较高的区域，在城区东部的太平门区域青年群体的占比较少，相比较而言，中老年群体更倾向于集聚在城市

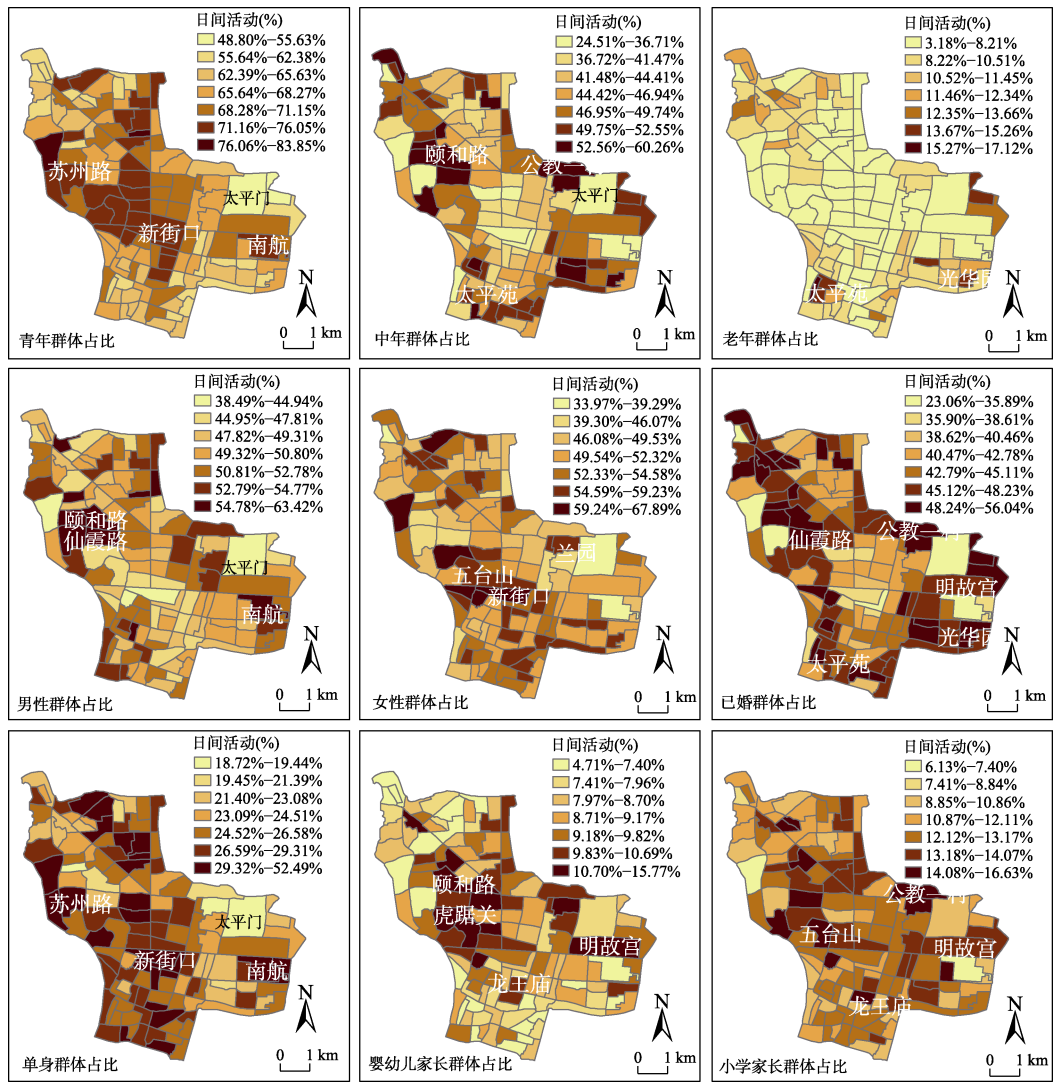


图 2 2020 年南京市内城区不同属性群体手机用户画像数据日间空间分布的可视化图

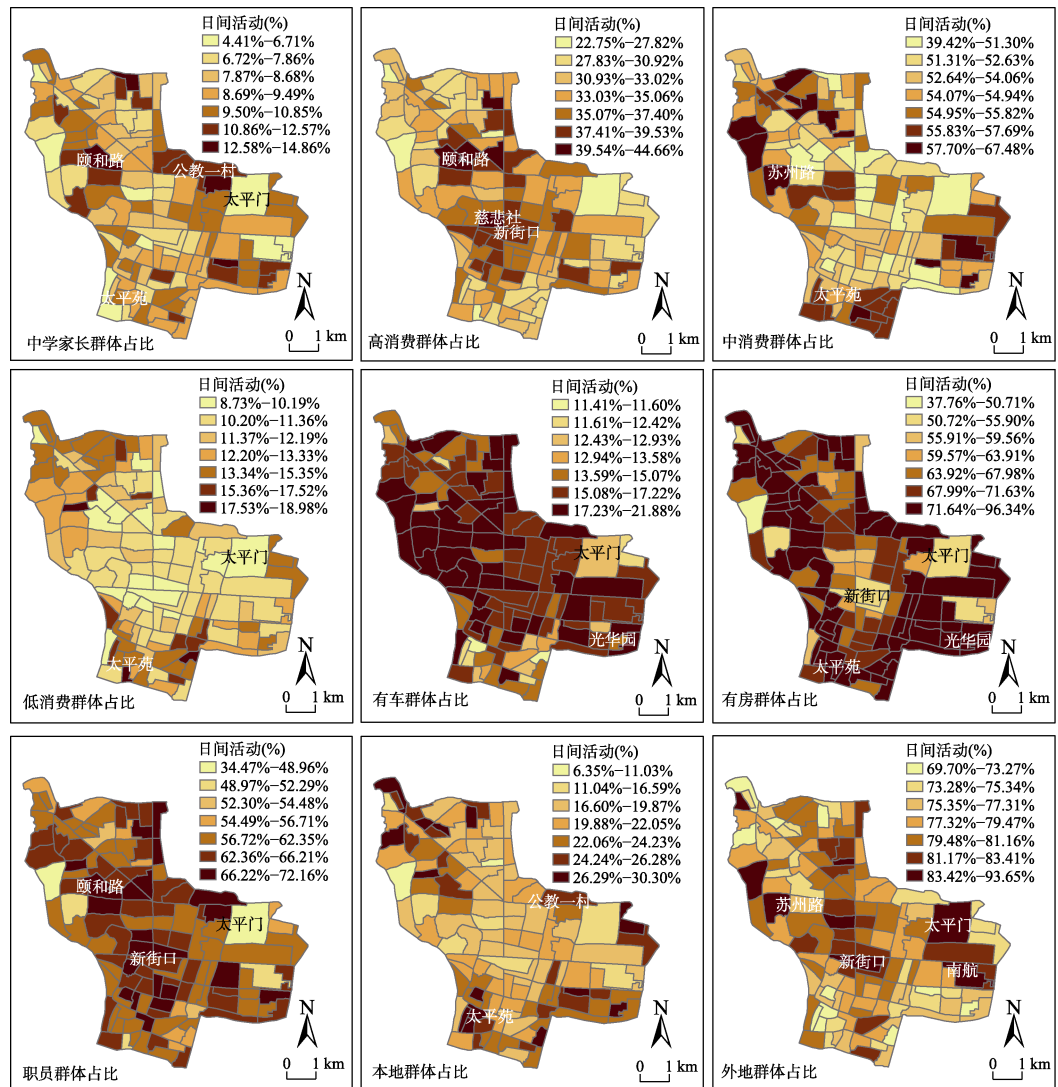


图 2 2020 年南京市内城区不同属性群体手机用户画像数据日间空间分布的可视化图（续）

周边地区。从性别差异来看，男性和女性在空间分布上相对均衡，其中颐和路和仙霞路区域男性占比较高，而新街口女性占比较高。从家庭结构来看，单身与已婚群体空间分布差异不大，在苏州路南航等城区边缘地区单身群体占比较高。家中有小学和中学生的家庭在城中仙霞村与城市东侧以北的公教一村区域较为集中，而家中有婴幼儿的家庭则主要集中于虎踞关与颐和路区域。从消费水平来看，消费水平对人群空间分布有显著影响，高消费群体日间时段更倾向于在城市中心的颐和路等商业区域活动，而中低消费群体则更多分布在城市周边地区。有车群体占比较为均匀，有房群体则在新街口等城市中心地区的占比显著低于周边地区。从职业特征来看，职员群体空间分异情况相对较不显著，日间活动人口在整体人群占比较高。在城中颐和路以及城市东侧区域的公教一村等区域人口占比相对较高，而太平门与古林等区域的人口占比则相对较低。从户籍状况来看，本地人主要分布于

内城区的周边地区，新街口等城市中心区域的本地人日间活动人口占比相对较低。

4.2.2 不同属性群体夜间空间分布特征

图 3 展示了 2020 年南京市内城区不同属性群体手机用户画像数据的夜间空间分布情况，从年龄分布来看，青年群体的夜间活动主要集中在苏州路、青岛路和南航等区域，而城区内部其他区域夜间居住人口空间分布则相对均匀，相比之下，中老年群体更倾向于集聚在公教一村、光华园等城市边缘区域。从性别差异来看，男性与女性空间分布均较为均匀，但男性在苏州路的人口占比较高，女性则在古林和新街口地区更为集中。从家庭结构来看，已婚群体与单身群体夜间空间分布差异显著，已婚群体更倾向于居住在颐和路及公教一村等城市东北侧，而单身群体则更倾向于集聚在城市西侧苏州路和古林地区。家中有婴幼儿、小学生、中学生的家庭夜间空间分布的集聚度均较高，主要集中于兰园、仙霞路、公教一村、五台山等区域，与这些区域周边优质的基础教育资源密切相关。从消费水平来

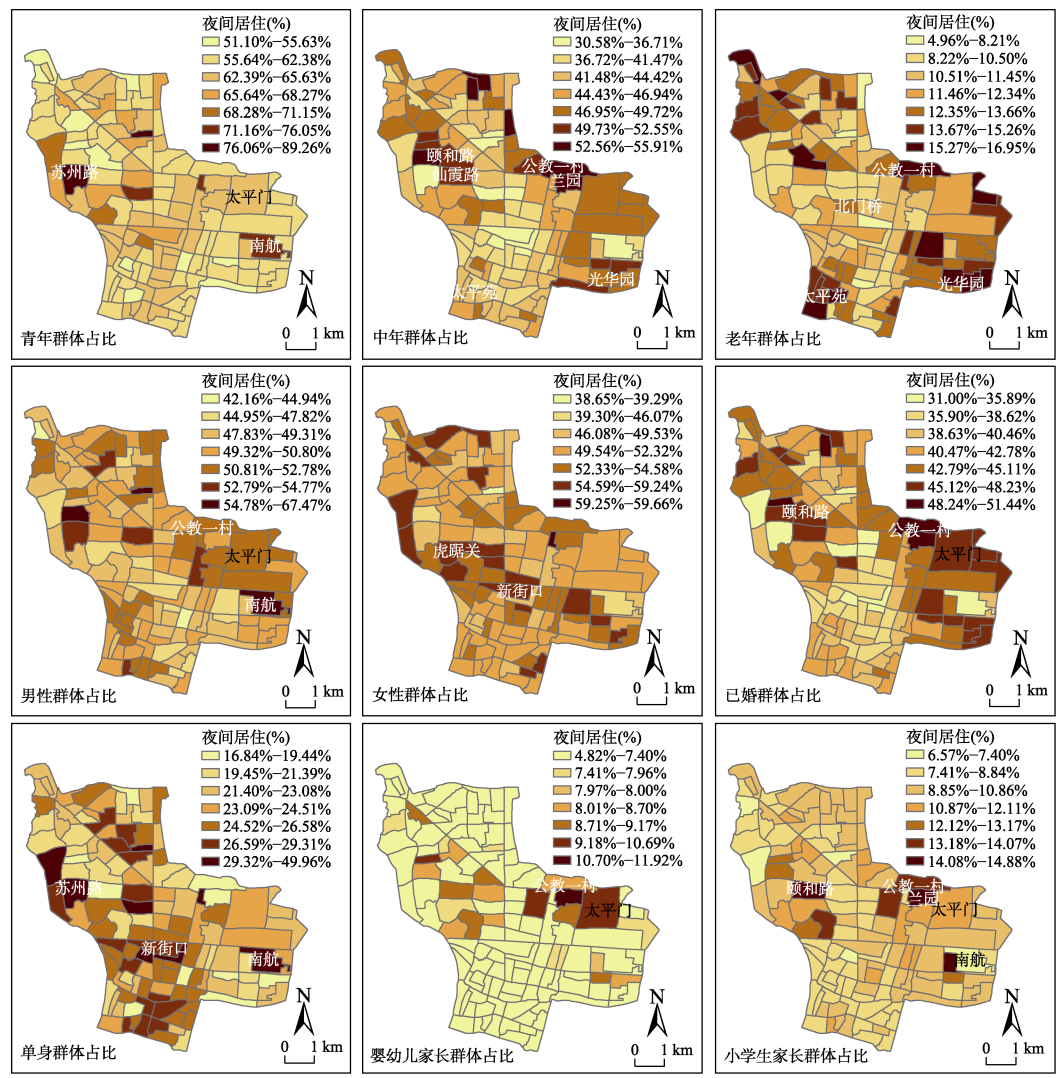


图 3 2020 年南京市内城区不同属性群体手机用户画像数据夜间空间分布的可视化图

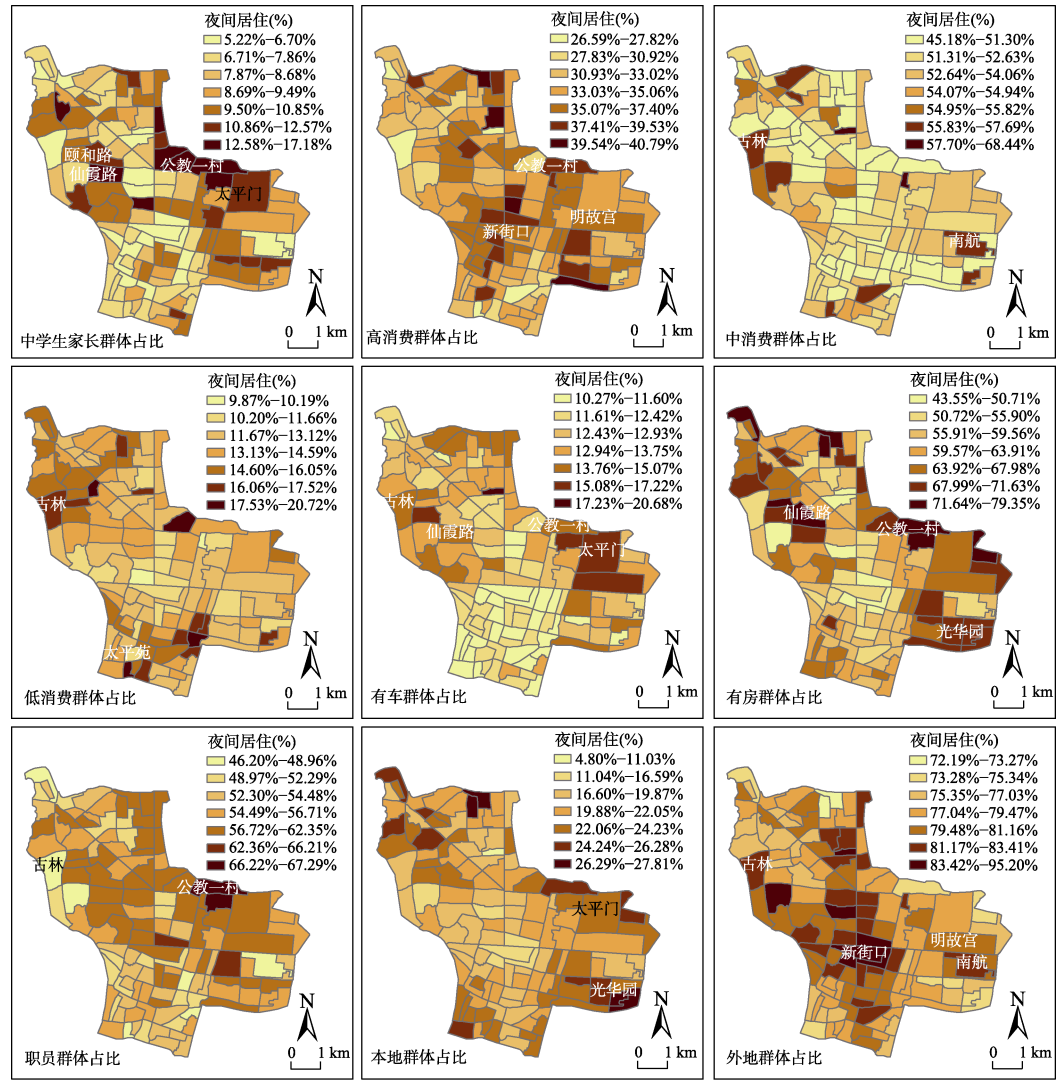


图3 2020年南京市内城区不同属性群体手机用户画像数据夜间空间分布的可视化图（续）

看，高消费群体主要分布在城市中心区和商业区，夜间居住空间分异情况不显著，而中低消费群体则更多集聚在吉林或南航等城市边缘区域。有车群体与有房群体在夜间空间分异情况较为相近且较为显著，城区东北侧的公教一村与太平门地区以及城市西侧的吉林等地区有车有房群体占比均较高。从职业特征来看，职员群体夜间居住空间分布较为均衡，城市北侧的夜间居住人群空间占比较高，公教一村社区内夜间居住人口占比最高。从户籍状况来看，本地人夜间居住地区空间分异较为显著，城北、城南、城东等老城区区域人口占比较高，外地人则在新街口等城市中心地区以及吉林等城市西北侧地区人口占比较高。

5 讨论和总结

以手机用户画像为依托的新型数据来源，在扩大样本覆盖范围、提升时空数据精度和

丰富社会经济维度等方面具有明显优势。通过分析 2020 年南京市内城区手机用户的 WIFI、GPS 位置信息、IP 上网日志、用户 APP 使用情况数据,结合用户的线下偏好与位置服务数据,刻画了 2020 年的南京内城区不同社会群体全方位的手机用户画像特征,整编形成南京内城区手机用户画像地理信息数据集(2020)。数据集具体包括 2020 年南京市内城区青年群体、中年群体、老年群体、男性群体、女性群体、已婚群体、单身群体、婴幼儿群体家长、小学家长、中学生家长群体、高消费、中消费群体、低消费群体、有车群体、有房群体、职员群体、本地群体、外地群体 18 类不同手机用户群体以及其在各个社区的日夜间空间分布情况。通过对数据集的地理空间分析,发现南京市内城呈现以新街口为城市核心的“核-缘”空间结构。新街口及其周边区域是商业和商务活动的聚集地,吸引了大量高消费群体、青年群体、职员群体、外地人群体。同时,城市的居住功能主要集中分布于城北、城南、城东等周边区域,城市中心本地人占比相对较高,于夜间存在较多的人口流动情况。城北、城南、城东等周边区域是承载南京市内城区主要居住功能的区域。该数据集通过分析南京市内城区不同社会群体的手机用户画像数据,揭示不同社会群体的日夜间空间活动规律,有助于深入理解城市空间发展的现状和未来发展趋势,为城市社会空间和行为规划等研究提供了重要依据。鉴于分析技术和研究能力有限,该数据集模型训练结果的准确度和可靠性尚有提升潜力,但在缺乏实时官方统计数据的背景下,为城市社会地理学研究提供了多维度的参考和数据支撑。

作者分工:黄琴诗对数据集进行了数据分析,并撰写了数据论文;陈家裕对数据集进行了数据分析,并撰写了数据论文;宋伟轩对数据集进行了总体设计。

利益冲突声明:本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] 黄琴诗,周强,宋伟轩.新时期城市居住分异研究的多维转向与尺度响应[J].地理科学进展,2023,42(3): 573-586.
- [2] 甄峰,袁超,张姗姗等.智慧国土赋能城市高质量发展的路径研究——以重庆为例[J].时空信息学报,2024,31(2): 1-13
- [3] 王德,韩滨鹂,张天然等.手机信令数据的出行测度准确性分析——基于与居民出行调查数据的比较[J].地理科学进展,2024,43(5): 854-869.
- [4] 牛强,伍磊,盛富斌等.基于个体职住迁移的武汉郊区新城职住动态平衡测度方法[J].地理学报,2023,78(12): 3095-3108.
- [5] 宋伟轩,徐昀,王捷凯等.基于手机画像数据的南京内城日夜间社会空间分异[J].地理学报,2024,79(2): 421-438.
- [6] 黄琴诗,陈家裕,宋伟轩.南京内城区手机用户画像地理信息数据集(2020)[J/DB/OL].全球变化数据仓储电子杂志,2024. <https://doi.org/10.3974/geodb.2024.07.09.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2024.07.09.V1>.
- [7] 全球变化科学研究数据出版系统.全球变化科学研究数据共享政策[OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).