

# 基于网约车数据和 POI 数据的城市居民出行规律研究

**摘要：**城市居民的出行与交通问题密切相关，研究居民出行规律特征可以为城市规划和交通资源的合理分配提供支持。网约车作为重要的居民出行方式之一，其行驶的起止点反映了乘客的出发点和目的地。通过分析和挖掘网约车数据能够很好的反映居民的出行特点。本文基于武汉市的网约车数据，结合 GIS 技术，从数据分析的角度对武汉市城市居民的出行行为在时间与空间上的规律与模式进行了研究。研究结果有助于理解武汉市城市居民的出行特点，为居民提供有效的出行参考，为网约车司机提供客源分布信息，为城市交通建设与管理提供有价值的技术支撑。

**关键词：**GPS 出租车数据；居民出行特征；时空分析；热点提取；聚类；网约车；POI；

## 1 引言

近年来，随着城市化的快速发展与互联网和移动终端的普及，城市居民越来越倾向于使用线上的方式满足出行需求，持续优化居民出行服务也逐渐成为现阶段城市管理所面临的重要课题。网约车是满足城市居民日常出行需求的重要途径<sup>[1-2]</sup>。网约车的服务模式是精准的点对点服务，通过平台实现运输车辆和出行需求的匹配，因此能最好的反映出居民出行的时空特征。

随着移动定位技术的快速发展，海量、高精度出租车运营数据为研究人群移动性以及城市空间结构的联系<sup>[3-4]</sup>提供了数据支撑。通过各种数据分析与挖掘手段对其轨迹数据进行深入研究可以帮助乘客了解出行信息<sup>[5-6]</sup>、为司机推荐导航路线<sup>[7]</sup>、改善出租车的运营管理<sup>[8-9]</sup>。目前已有许多学者利用网约车数据对不同地区进行了研究，但由于地理与城市文化差异会对人群的出行模式产生影响。因此，本文利用武汉市的网约车数据及 POI 数据对城市人群的出行规律进行研究，并在此基础上为改善武汉市城市交通建设提供相关建议。主要研究内容包括以下：基于网约车轨迹数据挖掘城市热点出行区域<sup>[10-11]</sup>；利用网约车数据探寻居民出行时空分布<sup>[15-17]</sup>。

## 2 研究区与数据

### 2.1 研究区域

本文研究区域为武汉市，中国中部地区的中心城市，全国重要的工业基地、科教基地和综合交通枢纽。截至 2021 年末，全市下辖 13 个区，总面积 8569.15 平方千米，常住人口 1364.89 万人。作为中国 14 座特大城市之一，有较高的网约车出行需求。

### 2.2 数据及预处理

本文分析的数据为从滴滴等网约车公司购得经过脱敏处理的武汉市网约车轨迹数据，数据产生于 2018 年 11 月 12 日到 2018 年 11 月 25 日期间，具体的数据项包括：精确到秒的数据上传时间、网约车所处经纬度、订单的编号等。由于传感器设备故障或后台系统异常会造成部分数据丢失，位置点偏移等现象，导致采集到的数据不符合客观逻辑。因此，本文对原始数据进行清洗，主要通过对特定数据项设置约束，插值滤波，路网匹配等方法完成了对数据的预处理，最终得到共计 1652867 条数据。最终清洗后网约车的订单数据字段见表 1。

表 1 清洗后订单数据字段

ID	getonDate	getonLon	getonLat	getoffDate	getoffLon	getoffLat
601**3	1541995**9	114.280**2	30.589**1	1541995**0	114.28**0	30.589**3
601**8	1541995**9	114.260**3	30.628**8	1541995**0	114.29**0	30.630**5
601**3	1542056**0	114.385**3	30.517**1	1542060**0	114.31**6	30.707**7
601**4	1541995**1	114.314**3	30.515**5	1541996**0	114.38**4	30.507**2

### 3 基于出租车 GPS 数据的热点区域挖掘

为研究城市人群的出行热点区域，本文对网约车轨迹数据进行挖掘，从中提取出每一条轨迹数据的 OD 点，即乘客的上车点和下车点，将复杂的出行线轨迹抽象为简单的上下车点数据。通过对上下车点数据进行进一步的分析，可以从上下车点中提取出城市人群出行的需求分许以及趋向规律，进而总结出城市人群出行规律，也可以通过计算，区分出包含大量上下车点的空间位置，即热点区域<sup>[12-14]</sup>。

#### 3.1 基于 K-Means 聚类算法挖掘热点区域

本文认为，在热点区域的周围，上下车点将会大量聚集，如果对上下车点按照距离进行聚类，被聚集为同一类的上下车点的中心位置可以近似的看成热点区域。聚类结果如图 1 所示：

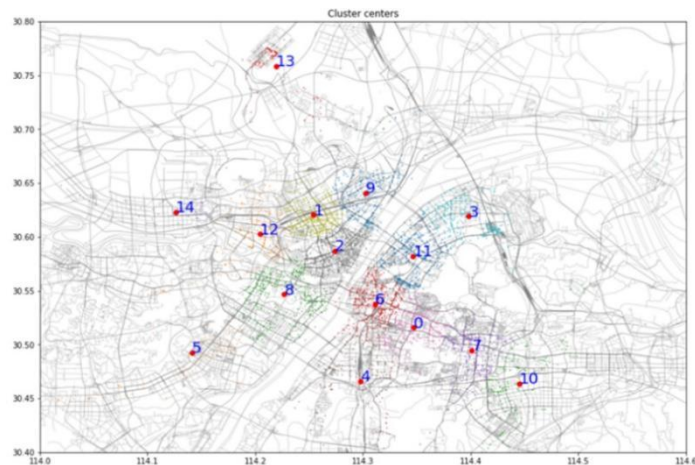


图 1 基于聚类算法提取的热点区域

由于聚类的类数需要人为指定，本文通过实验发现，当类数过低时，聚类的热点区域没有代表性，而当类数过高时，聚类的热点区域失去意义且误差较大。通过对比，选择聚类簇数为 15 时，得出的结果较好。可以看出，热点区域主要分布在武汉市的中心区域，沿长江分布，分布相对集中。其中较为特殊的热点区域为 13 号聚类中心附近区域，该区域为天河机场所处位置，尽管地理位置远离市区中心上下车点密集处，但是还是有较密集的上下车点。

#### 3.2 基于核密度算法的热点区域挖掘

本文进一步通过核密度的算法原理上下车点的聚集程度：视每个上下车点周围创建一个数值随距离点的远近递减的矩阵，通过所有点对应矩阵的叠加，得出的数值可以正比于点的聚集程度，进而提取上下车点热点区域。具体的基于核密度算法提取的结果如图 2 所示：

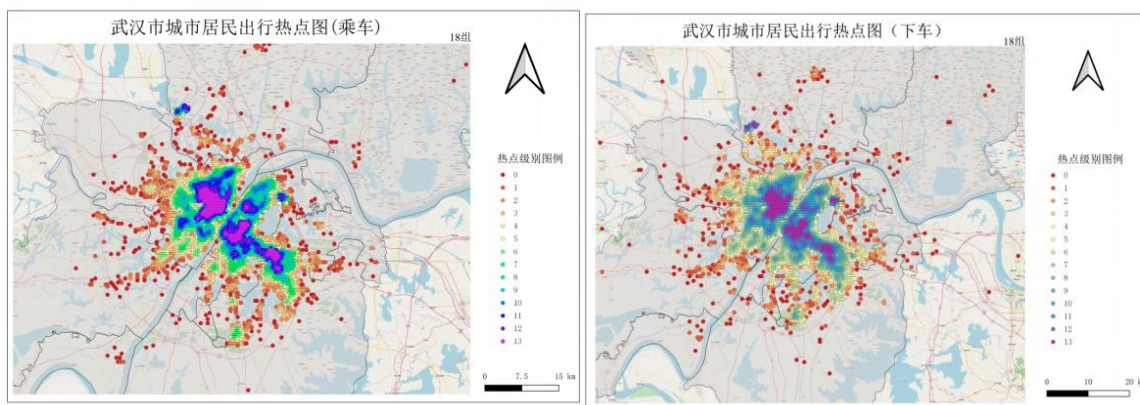


图 2 基于核密度算法提取的热点区域图

从图 2 可以看出，通过得到的核密度分析结果进一步佐证了通过聚类算法得出的结论：上下车热点区域集中在武汉市区中心，且沿长江分布。进一步的通过核密度的值，可以看出，位于江汉区，洪山区，武昌区的热点区域拥有更高的聚集程度，对应于聚类算法中的第 1, 2, 6, 0, 11, 13 号聚类中心，通过经纬度的逆编码以及地图的对比查看，发现这些聚类中心对应的兴趣点为 1 号对应汉口站附近区域，2 号对应 ICC 武汉环贸中心附近住宅区，6 号对应武昌站附近区域，11 号对应武汉站附近区域，0 号对应华中师范大学与武汉大学交界处区域，是更符合热点区域定义的位置。可以看出交通枢纽，例如武昌站，天河机场等，商业中心，以及消费人员密集的大学附近，都是网约车出行的热点区域。

### 3.3 基于缓冲区算法的交通要道的挖掘

除研究面状的热点区域外，本文进一步研究车辆在线状要素即道路上的热点区域。通过建立一定大小的缓冲区，统计每条道路对应的缓冲区包含的上下车点数目，包含数目多的，可以在一定程度上视为交通量大的热点道路，即交通要道。基于上述理论，交通要道提取结果如图 3 所示：



图 3 20 条交通要道

从图 3 的结果可以看出，热点道路有珞狮路、武珞路、青年路、新华路、中山路、文化大道、白沙洲大道、建设大道、二环线等。其中几乎全部道路穿过二环线以内的区域，与提取出的热点区域重叠度较高，此外，通过与上下班高峰期的实时路况情况对比发现，提取出的多数热点道路如：珞狮路、武珞路、中山路、二环线等道路经常出现拥堵的情况，证明这些道路实际上确实有较大的交通流量。最终得出以下结论：交通量较大的热点道路通常具有下述两个特点的至少一个：①道路本身长度较长；②道路与热点区域相交部分较多。

### 3.4 热点区域流动趋势研究

为进一步分析热点区域与时间变化的关系，本文将各个时间段的基于核密度算法提取的热点区域的结果图集合为 GIF 图，使之动态变化，更能看出热点区域随时间变化的关系。某一时间段的截图如图 4 所示：

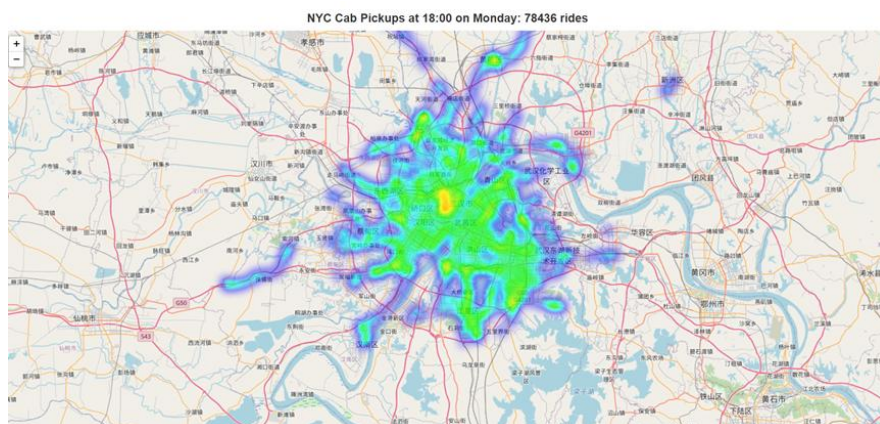


图 4 某日下午 6 点热点分布

总体来说，上车点以及下车点的热点区域具有相当程度的共性，上车的热点区域通常也为下车的热点区域且两者的热点区域都分布在长江沿岸的武昌区，洪山区等地。而如果进行更加细致的划分，可以从趋势图中看出，在深夜以及凌晨，整个市区的出行密度都较低，热点区域较少且变换相对较大，具有一定程度的随机性。而在上下班的时间段，尤其是在工作日期间，如早晨 7 点左右，中午 12 点左右，晚上 6 点左右，热点区域有明显的偏向住宅区的趋势。而其他热点区域的分布主要在医疗设施，商业中心，交通枢纽等地，得出的变化趋势符合人们的主观认知。

## 4 基于出租车 GPS 数据人群出行行为时空分析

对出租车的运营特征的分析主要包括对出租车的载客次数、载客时长、载客人数、订单数、载客里程等数据的统计分析。本文选取两周的数据从时间与空间上进行分析研究，在时间上进行出行时间、出行量、出行人数的分析研究，在空间上进行出行距离，出行速度的分析统计<sup>[18-19]</sup>。

### 4.1 出行时空特征提取

本文利用数据分析的方法从出租车轨迹数据中提取出行时空特征，包括出行距离，出行时长，出租车平均速度。结果如图 5、图 6 所示。

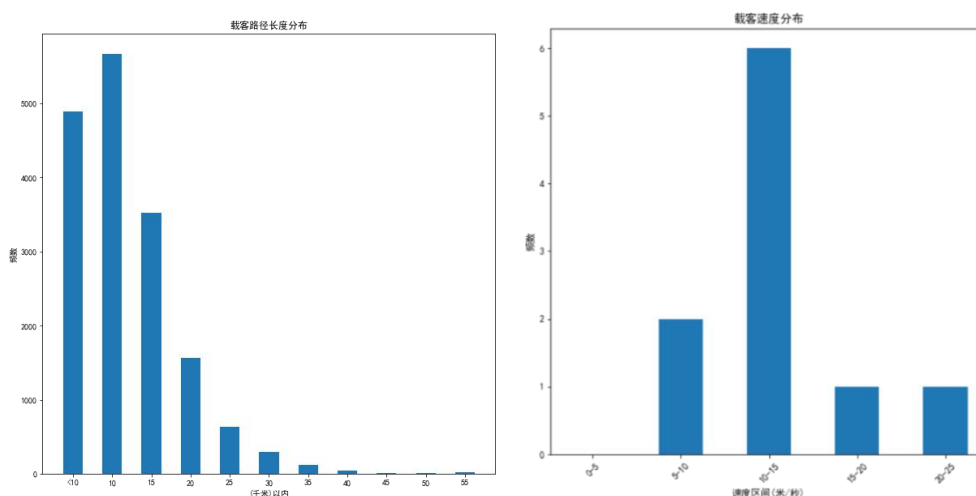


图 5 出行距离/速度分布图



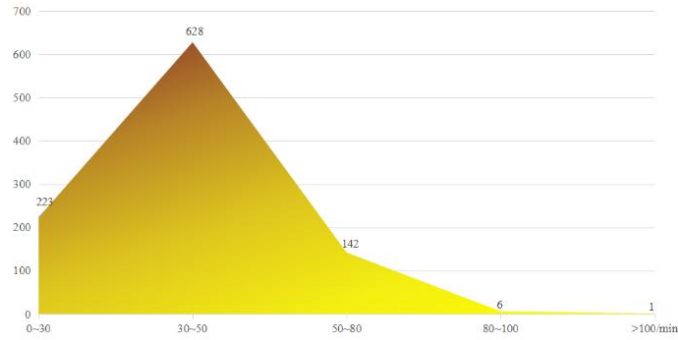


图6 居民出行耗时统计图

图5左图显示周一武汉市居民乘车出行距离所占的比例，可以看出出行距离在0-10公里的比例最大，其次10-15公里占比也较高。大于30公里的几乎没有。图5右图显示出行速度主要分布在10-15m/s区间，说明武汉市交通拥堵状况正常。通过图6可以得出居民出行所用时间主要集中在0-50分钟内，30-50分钟占比最高，极少数花费超过100分钟。

## 4.2 出行时长的时间分布

为了研究居民出行规律的时间分布，将数据重新处理，提取每条订单的详细信息，并添加关联属性如上下车星期、小时，行程时间等。

利用处理后的数据分别计算每天的出行时长总和、平均值；每星期的出行时长总和、平均值；每小时的出行时长总和、平均值。并绘制出行时长与日期、星期的折线图如图7。绘制出行时长与星期、小时的散点图如图8。绘制出行平均时间与小时的箱式图如图9。

由图7可以看到17日、23日、24日出行时长总和最长，12日与25日最低。平均时长25日最长。从日折线图来看，整体订单数量随日期变化呈现规律波动，周期为7天，据此推测订单数量可能与星期几有关。此外，可以看出周一打车时间总和最小，周五与周六时长总和最大，说明周一工作日使用网约车出行的较少，而周五、周六居民外出娱乐，因此出行时长总和增加。周日平均打车时间最长，说明这一天的订单多为长途订单。

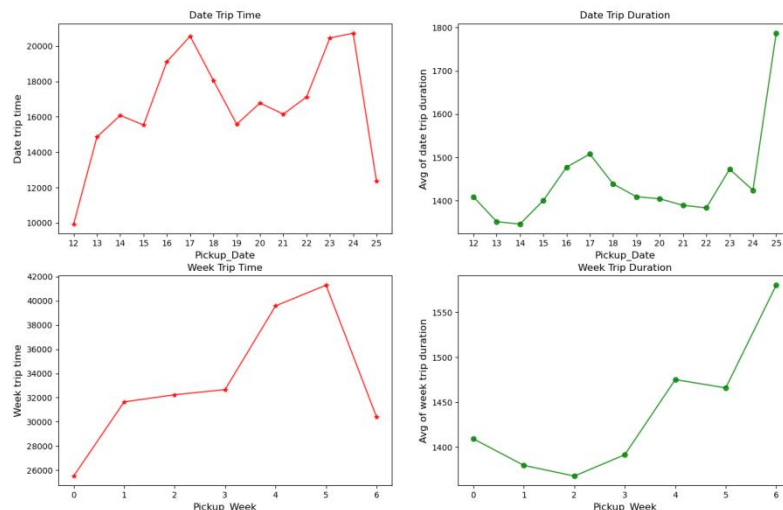


图7 出行时长与日期\星期折线图

由图8左图可以看出周日周一打车次数最低，周五至周六打车次数最高。可能周四至周六市民出行意愿更高。由图8右图可以看出：凌晨一点开始至凌晨五点，订单数量急剧下降，符合人群休息规律。从早6点开始，订单量由谷底回升，早7点至早9点有一个早高峰。在早8点到下午4点之间有轻微波动不明显。

而早 10 点有一个轻微的走低点。下午 5 点 6 点进入打车高峰，从晚 8 点开始平均打车量逐渐降低。推测相比于晚上，市民在白天外出活动时更偏向于网约车出行方式。

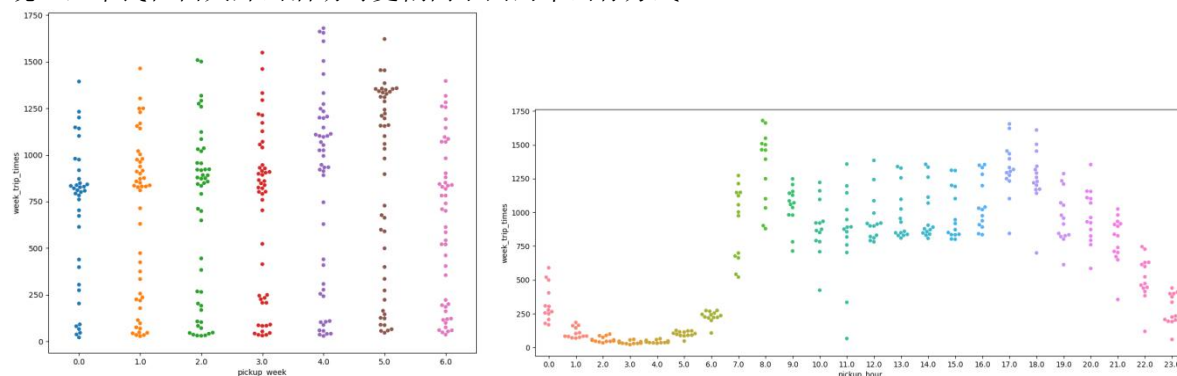


图 8 出行时长与星期\小时散点图

由图 9 可以看出凌晨 3 点至 6 点这一时间段出行平均时长较高。排除交通拥堵影响，行程时间长短可近似等同于距离长短。这段时间的长行车时间较多，据此推测：凌晨 3 点至 6 点这个时间段接到长距离行程单的机率远高于其他时间段。

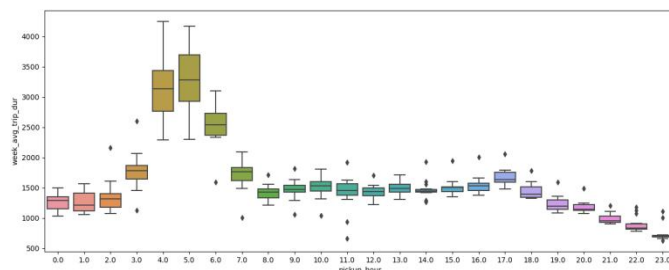


图 9 出行平均时长与小时箱式图

#### 4.3 出行人数的时间分布

分别计算每星期、每小时的出行人数平均值。绘制出行人数与星期的折线图、箱式图如图 10。绘制出行人数与小时的散点图如图 11。

图 10 展示了从星期一到星期日的每日订单总量、每日平均乘车人数随星期的波动。可以看出周四与周日的平均出行人数最多，周日的平均出行人数较多，但相反的周日的出行时间很少，说明周日人们偏向于结伴出行。

由图 11 可以看出凌晨 1 至 6 点订单乘车人数较为分散，既有多人（约伴），也有单人的情况。早 7 点-9 点之间，每个订单乘车人数全日最低，恰好又是上班时间，预计是单人上班打车情况较多。

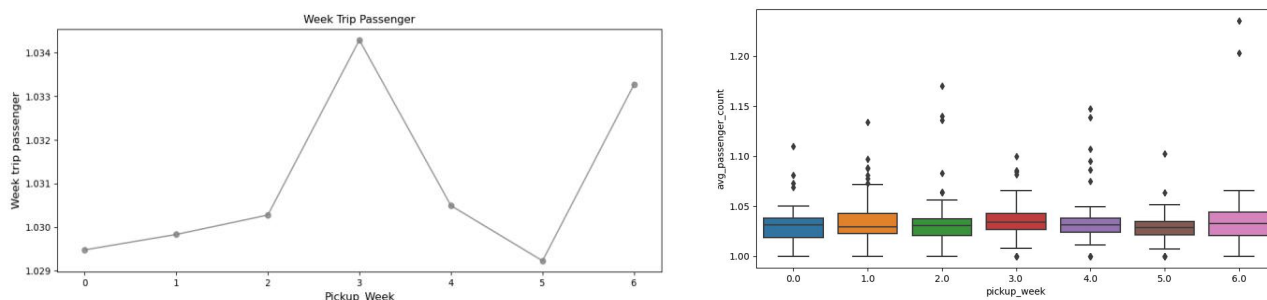


图 10 出行人数与星期关系图

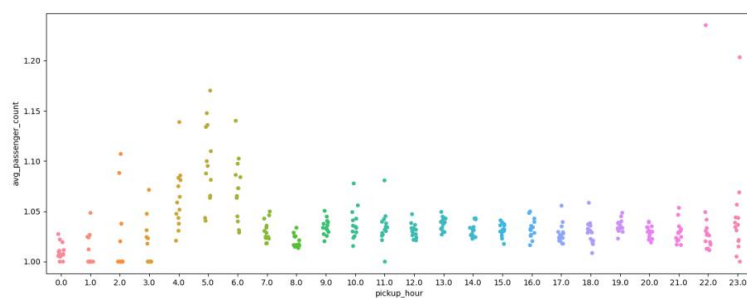


图 11 出行人数与小时散点图

#### 4.4 出行量的时空分布

分别计算并统计每小时的订单数量，结果如图 12。根据上车点位置为订单数据添加“行政区”这一属性，统计各个行政区内订单数量随时间的分布，并添加是否为工作日这一属性，分别统计平均出行量，结果如图 13。

图 12 中蓝色折线为上车点统计量、橙线为下车点统计量。可以看到上车点与下车点保持较好的统一，也与实际的载客规律相符。而上下车点的总数也代表订单数积累的程度。明显的，图中很好地揭示了人群出行的早晚高峰与夜间的低峰时段。即在早八点左右与晚六点左右上下班交通高峰时期与凌晨十二点过后的低出行时期。

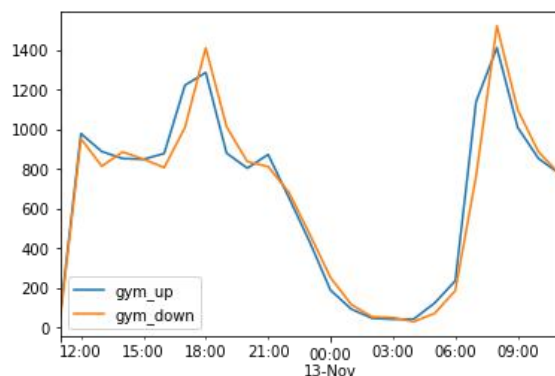


图 12 出行量与小时折线图

图 13 中分别展示了武汉市洪山区、青山区、江夏区的出行量随时间分布。可以看出不同行政区的出行时间分布较为类似，出行高峰集中在周末。但相对来说，洪山区的出行量要远高于江夏区、青山区。说明洪山区的网约车需求量更高。此外对比工作日与非工作日的平均出行量发现，非工作日出行量要高于工作日出行量。

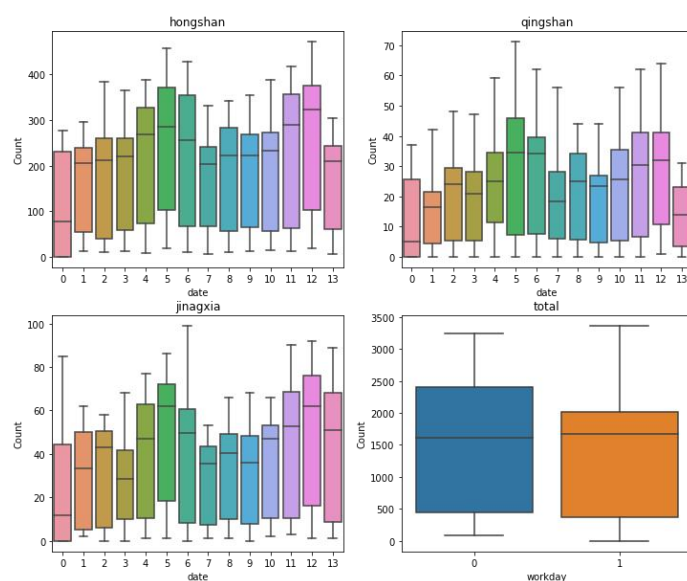


图 13 各行政区出行量与小时散点图

## 5 讨论与结论

本文基于武汉市出租汽车轨迹数据，利用多种方法挖掘了居民出行热点区域，并分别从时间与空间两个维度对居民出行特征进行了统计与分析。研究结论客观真实，可以为城市规划与管理，居民出行选择及司机判断客源分布提供参考<sup>[20-21]</sup>。

本文仅采用了武汉市两个星期的网约车轨迹数据进行研究，后续可获取武汉市疫情管控期间的轨迹数据进行对比分析，研究疫情对居民出行行为的影响。其次，由于居民出行方式的多样性，仅仅利用出租车GPS 数据进行分析难免存在局限，在以后的研究中可以将公交、地铁、共享单车等出行方式一同考虑，从而获得更为客观全面的居民出行规律。

## 参考文献：

- [1] Shi C, Li Q, Lu S, et al. Exploring Temporal Intra-Urban Travel Patterns: An Online Car-Hailing Trajectory Data Perspective[J]. Remote Sensing, 2021,13(9),1825.
- [2] 吴华意,黄蕊,游兰,等.出租车轨迹数据挖掘进展[J].测绘学报,2019,48(11):1341-1356. [ Wu H Y, Huang R, You L, et al. Recent progress in taxi trajectory data mining[J]. Acta Geodaetica et Cartographica Sinica, 2019,48(11):1341-1356. ]
- [3] Liu X, Gong L, Gong Y, et al. Revealing travel patterns and city structure with taxi trip data[J]. Journal of Transport Geography, 2015,43:78-90.
- [4] Zhong C, Schlöpfer M, Müller Arisona S, et al. Revealing centrality in the spatial structure of cities from human activity patterns[J]. Urban Studies, 2017,54(2):437-455.
- [5] Tang K, Chen S, Liu Z. Citywide spatial-temporal travel time estimation using big and sparse trajectories[J]. IEEE Transactions on Intelligent Transportation Systems, 2018,19(12):4023-4034.
- [6] Shi C, Li Q, Lu S, et al. Modeling the Distribution of Human Mobility Metrics with Online Car-Hailing Data—An Empirical Study in Xi'an, China[J]. ISPRS International Journal of Geo-Information, 2021,10(4):268.
- [7] Wu L, Hu S, Yin L, et al. Optimizing cruising routes for taxi drivers using a spatio-temporal trajectory model[J]. ISPRS International Journal of Geo-Information, 2017,6(11):373.
- [8] Rong H, Wang Z, Zheng H, et al. Mining efficient taxi operation strategies from large scale geo-location data[J]. IEEE Access, 2017,5:25623-25634.



- 
- [9] Chen D, Zhang Y, Gao L, et al. The impact of rainfall on the temporal and spatial distribution of taxi passengers[J]. PloS one, 2017,12(9):e0183574.
- [10]杨树亮,毕硕本,Nkunuzimana A,黄铜,万蕾.一种出租车载客轨迹空间聚类方法[J].计算机工程与应用,2018,54(14):249-255.
- [11]姜海林.基于出租车数据的城市居民出行时空特征研究[D].武汉大学,2018.
- [12]刘入嘉.基于出租车 GPS 数据的城市出行空间结构研究[D].东南大学,2018.
- [13]周勃,秦昆,陈一祥,李志鑫.基于数据场的出租车轨迹热点区域探测方法[J].地理与地理信息科学,2016,32(6):51-56.
- [14]秦昆,王玉龙,赵鹏祥,徐雯婷,徐源泉.行为轨迹时空聚类与分析[J].自然杂志,2018,40(3):177-182.
- [15]陈志明.基于大数据的出租车乘客出行特征提取和分析方法研究[D].长安大学,2017.
- [16]刘惠宇,刘鹏,张婷.基于出租车 GPS 数据的城市人群出行模式研究[J].江苏科技信息,2019,36(17):48-51.
- [17]童晓君,向南平,朱定局.基于出租车 GPS 数据的城市居民出行行为分析[J].电脑与电信,2012(1):56-59.
- [18]林基艳,张雅琼,张慧.基于出租车 GPS 轨迹数据挖掘的居民出行特征研究[J].计算机时代,2017(5):37-39.
- [19]方琪,王山东,于大超,朱鸿博,李贺.基于出租车轨迹的居民出行特征分析[J].地理空间信息,2019,17(5):128-130.
- [20]李婷,裴韬,袁烨城,宋辞,王维一,杨格格.人类活动轨迹的分类、模式和应用研究综述[J].地理科学进展,2014(7):938-948.
- [21]贾冲,冯慧芳,杨振娟.基于出租车 GPS 轨迹和 POI 数据的商业选址推荐[J].计算机与现代化,2020(2):21-25.