每日出行节律可视化

作业选题：每日出行节律

时间范围：2017.10.17-2017.10.23

# 数据来源

## 计步数据

来自手机计步数据，如表格1，包括步数和距离(Km)。考虑到骑车，并不确定此处的距离是否仅为步行距离，于是用距离/步数\*1000得到单步距离，即字段DISTANCE/STEPS，其标准差为0.0015m，均值为0.7135m，可以认为此处的距离不包括骑行距离。因此采用DISTANCE字段代表每日步行距离。

Table 1 计步数据

|  |  |  |  |
| --- | --- | --- | --- |
| DATE | STEPS | DISTANCE | DISTANCE/STEPS |
| 10/17 | 3914 | 2.79 | 0.712825754 |
| 10/18 | 8148 | 5.82 | 0.714285714 |
| 10/19 | 3751 | 2.68 | 0.71447614 |
| 10/20 | 9287 | 6.63 | 0.713901152 |
| 10/21 | 7196 | 5.14 | 0.714285714 |
| 10/22 | 8273 | 5.91 | 0.714372054 |
| 10/23 | 8068 | 5.73 | 0.710213188 |

## 骑行数据

来自共享单车OFO和MOBIKE，如表格2，每条记录为一次骑行，包括日期(DATE)、时间(TIME)、骑行时长(DURATION: minute)、骑行距离(DISTANCE: Km)、单车品牌(BRAND)。

Table 2 骑行数据

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DATE | TIME | DURATION | DISTANCE | BRAND |
| 10/17 | 7:19 | 7.4 | 1.1 | OFO |
| 10/17 | 12:19 | 4.9 | 0.7 | OFO |
| 10/17 | 14:22 | 7.9 | 1.1 | OFO |
| 10/17 | 18:38 | 2.4 | 0.6 | OFO |
| 10/17 | 19:21 | 3 | 0.7 | OFO |
| 10/17 | 20:18 | 4.5 | 0.6 | OFO |
| 10/17 | 22:30 | 7.1 | 0.8 | OFO |
| 10/18 | 11:10 | 2 | 0.314 | MOBIKE |
| 10/18 | 17:36 | 7 | 1 | MOBIKE |
| 10/18 | 19:22 | 4 | 0.659 | MOBIKE |
| 10/18 | 21:28 | 6.7 | 1.8 | OFO |
| 10/19 | 8:10 | 7.5 | 1.4 | OFO |
| 10/19 | 11:32 | 7 | 2.4 | OFO |
| 10/19 | 12:52 | 5.6 | 1.9 | OFO |
| 10/19 | 22:12 | 6 | 1.1 | OFO |
| 10/20 | 8:31 | 3 | 0.597 | MOBIKE |
| 10/20 | 14:12 | 5.5 | 3.7 | OFO |
| 10/20 | 21:17 | 5.7 | 1.2 | OFO |
| 10/21 | 12:54 | 3 | 0.9 | OFO |
| 10/21 | 13:01 | 4 | 1 | OFO |
| 10/21 | 21:39 | 15.6 | 2.8 | OFO |
| 10/22 | 9:01 | 5.6 | 0.8 | OFO |
| 10/22 | 14:35 | 6.8 | 4.3 | OFO |
| 10/22 | 20:38 | 5 | 0.812 | MOBIKE |
| 10/22 | 22:38 | 7 | 1 | MOBIKE |
| 10/23 | 7:57 | 8 | 1 | MOBIKE |
| 10/23 | 9:50 | 2 | 0.359 | MOBIKE |
| 10/23 | 11:59 | 4.3 | 0.7 | OFO |
| 10/23 | 12:06 | 12.8 | 1.7 | OFO |
| 10/23 | 14:59 | 3 | 0.4 | OFO |
| 10/23 | 17:33 | 5 | 1.7 | OFO |
| 10/23 | 17:53 | 6.9 | 3.1 | OFO |
| 10/23 | 18:22 | 3.1 | 0.7 | OFO |
| 10/23 | 22:02 | 7 | 1.5 | OFO |

# 可视化

根据已有数据，提出以下问题进行探究：

1. **每日总出行量是否存在一定的模式？是否接近？**

每日总出行量=每日步行距离+每日骑行距离。首先是程序自动生成的结果，如图1（a），每日出行公里数在8KM-18KM之间摆动，差别还是较大的，但自动生成的纵坐标轴从0开始，坐标范围太大，不能突出差别，于是更换纵坐标范围为[8,18]，如图1（a），较之图1（b），更能反映每日总出行量的差别。

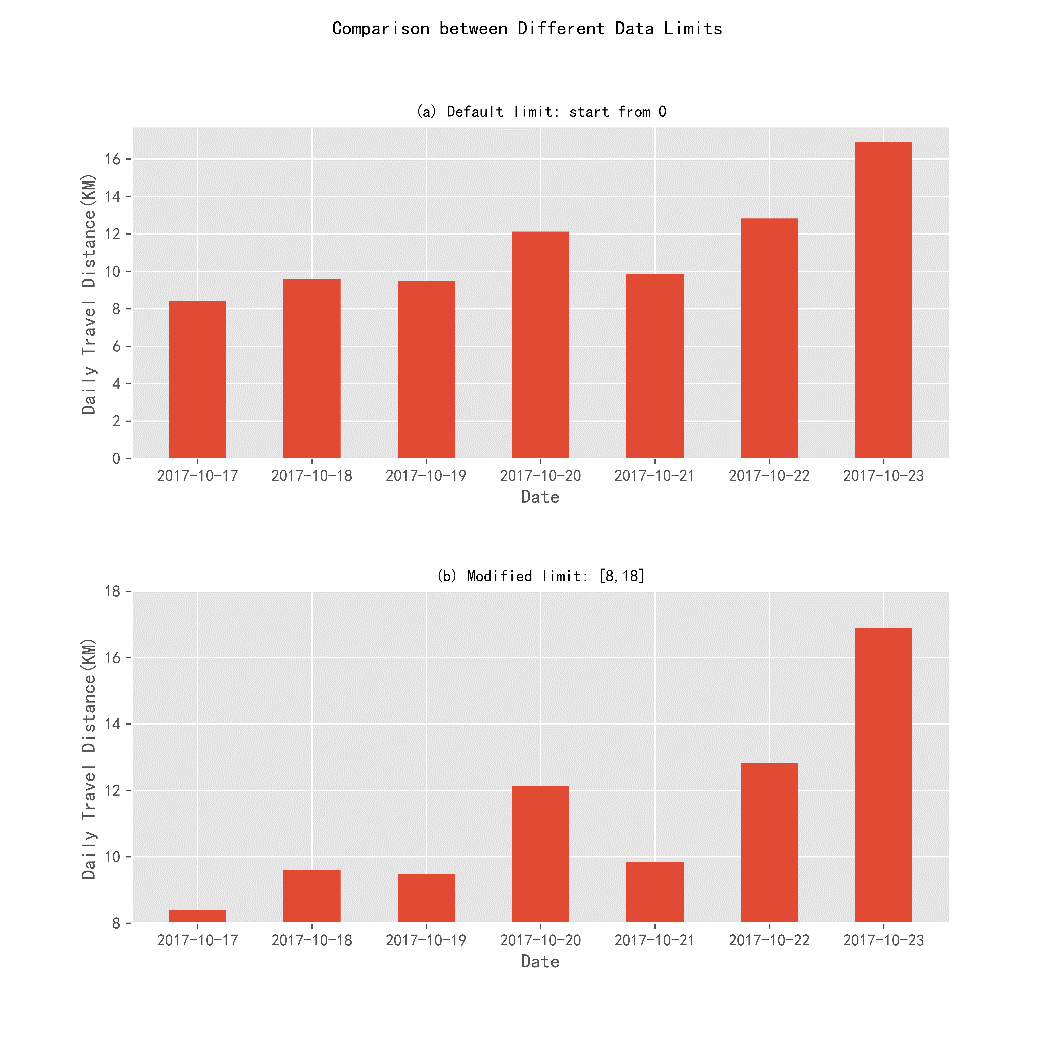


Figure 1 Comparison between Different Data Limits

1. **每日骑行时间段分布和是否存在一定的模式？是否接近？**

首先，按照每小时一个时间段，把每天划分为24个时间段。统计每个时间段内骑行次数。结果如图2，可以发现1.骑行一般从上午7:00开始，晚上23:00前结束。2.上午10:00-11:00和下午15:00-17:00没有骑行，说明这段时间不会有长距离出行。

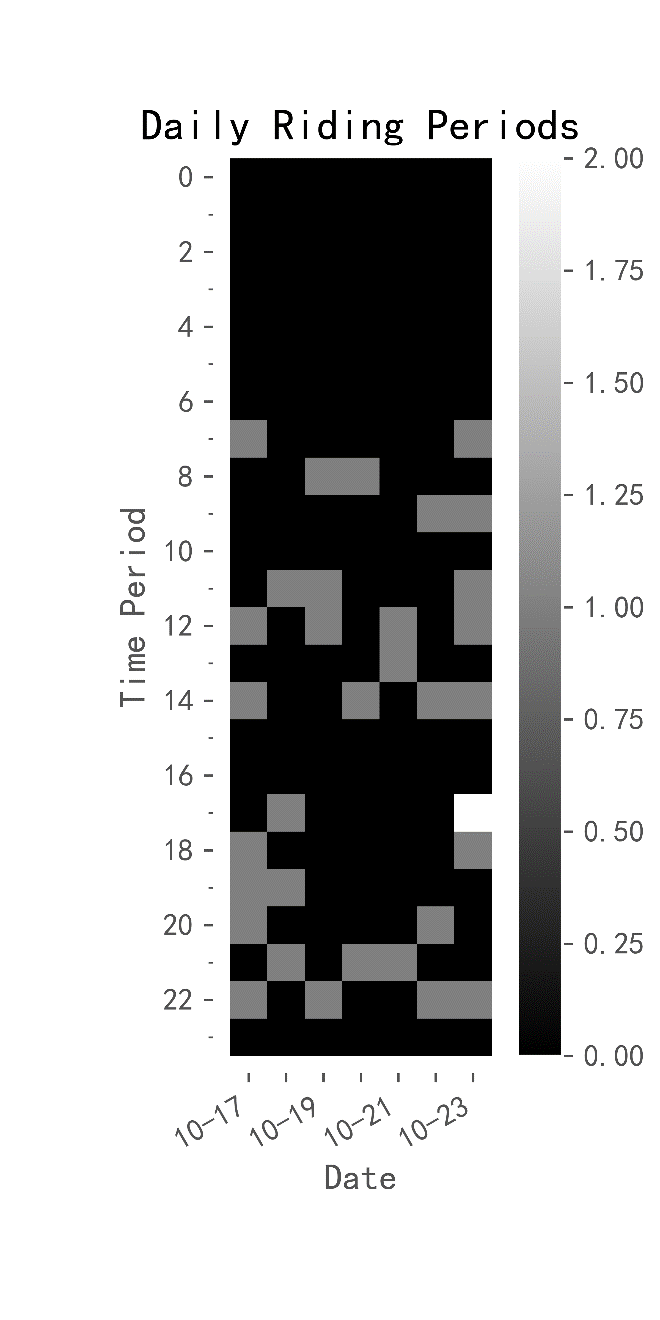


Figure 2 Daily Riding Periods

# 讨论

原始数据中未用到的数据：

1. **共享单车的品牌数据**

我考虑过用两者使用次数的比值反映校内OFO和MOBIKE数量之比（因为就我个人而言，在OFO和MOBIKE存在的情况下，选择哪种车骑行的概率是一样的），但该统计数据只有一个值，故没有进行可视化

1. **骑行数据中的骑行时间**

骑行数据用到了骑行次数和骑行距离，骑行时间实际上蕴含其中，反应信息不多，故没有进行可视化。

# 原始数据和代码

见github