



《地理信息系统原理与开发》实习报告

题目： 共享单车骑行印象

班级序号： 111181

学生姓名： 周麟

任课教师： 杨林

中国地质大学信息工程学院软件工程系

2021 年 1 月

目录

“共享单车骑行印象”实习报告	3
一 轨迹数据清洗	3
1.1 骑行数据提取	3
1.1.1 清除河里点	3
1.1.2 清除存在巨大偏移量的点	3
二 骑行点对应至街道生成路径图	3
2.1 简单的位置查询	3
2.2 网络分析	4
2.2.1 生成连通道路网	4
三 生成骑行热力图	6
3.1 2d 热力图分析	6
3.2 3d 热力图分析	7
四 骑行扩散过程模拟	9
4.1 骑行扩散指标计算及分析	9
4.2 孤立社区检测	11
4.3 社区扩散指数	11
五 可视化	12
5.1 vue	12
5.2 antV	12
5.3 html	13
六 成果展示	14
6.1 主页面	14
6.2 骑行量数据统计	15
6.3 骑行时长数据统计	15
6.4 骑行距离数据统计	15
6.5 骑行动态轨迹	16
6.6 骑行静态轨迹	16
6.7 2d 热力图	16
6.8 3d 热力图	17
6.9 流向图	17
6.10 孤立社区分析图	18
6.11 社区扩散指数分析图	18
七 实习心得	18
八 参考文献	19

“共享单车骑行印象”实习报告

一 轨迹数据清洗

1.1 骑行数据提取

因为某些因素，可能在进行骑行数据的统计时，会造成数据或多或少的偏移，所以在进行骑行数据提取时，会将河里点数据、存在较大偏移的点筛选掉。

1.1.1 清除河里点

通过位置查询将所有与河流面图层相交的骑行点选择出来，通过编辑操作将点删除并且保存编辑结果即可。

1.1.2 清除存在巨大偏移量的点

观察单车轨迹数据 csv 文件发现有大量坐标为（）的点，经过了可视化之后，发现这些点并不在武汉的底图范围内，初步判定为可能是网络或其他原因而造成轨迹坐标偏移，通过 java 编码对于这些点清除，写入新的 csv 文件。对没有清除较大偏移点在 antV 中可视化如下图 1 所示。



图 1 未清除偏移点路径图

二 骑行点对应至街道生成路径图

2.1 简单的位置查询

如果简单的对轨迹点和道路图层进行位置的查询，筛选出距离轨迹点 100 米以内的道路作为某一特定的轨迹点所经过的路径，生成的轨迹图层可能会覆盖到某些不可能经过的道路数据，或者轨迹与轨迹点之间并不连通（通过该轨迹不能到达目的轨迹点）。

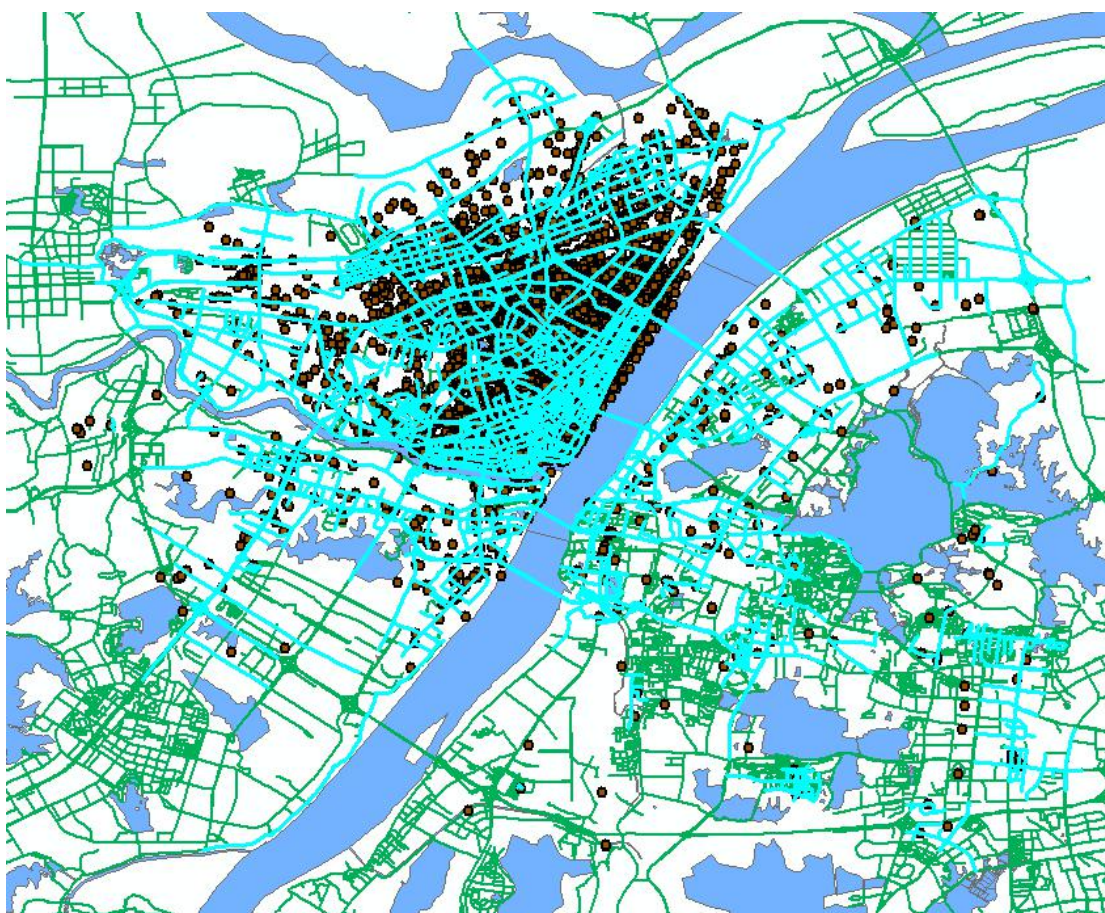


图 2 选择与轨迹点距离 100 米的街道生成的“伪”路径图

2.2 网络分析

利用 arcmap 中的网络分析进行道路生成,具体过程创建要素数据集,创建网络数据集,生成道路网和道路节点,将所有轨迹点转化到道路节点(考虑到可能会将单车停放至小区中,所以对轨迹点和道路节点进行位置查询,查询距离轨迹点 100 米以内的道路节点)进行网络分析,生成的最小路径图,可以近似认为是武汉市共享单车的骑行轨迹。其中进行网络分析的要点就是,要保证道路的连通性,确保生成的道路节点都能到达道路网上的任意节点。

2.2.1 生成连通道路网

2.2.1.1 具体过程

原始的道路数据中,可能会存在线段断开的情况,所以首先使用 arcmap 的线拓展功能,将断开的线段连接。并且还会存在一条路径会存在多条道路的问题,那么进行多路转单路的操作在进行网络分析,就能正确的生成对于所有轨迹点的最小路径的生成,进而近似地代表为共享单车轨道路径。

2.2.1.2 多线转单线

通过 arcmap 测量工具得知对于多线道路数据中宽度在 40~50m 内,所以生成 30 的缓冲区图片,图 3 展示道路缓冲区。

将生成的该缓冲数据导出为 tiff 文件,再导入图层中,在创建一个新的 shp 文件,保存的数据为折线类型。

利用 arcScan 工具对导出的 tiff 数据进行矢量化分析，将分析结果保存到新建的 shp 文件中，图 4 展示矢量化结果。



图 3 道路缓冲区



图 4 缓冲区矢量化结果

经过上述操作没有断点，并且单线的单路数据生成，接下来利用该数据创建网络数据集之后生成相应的道路网进行道路分析，结果如图 5 所示。

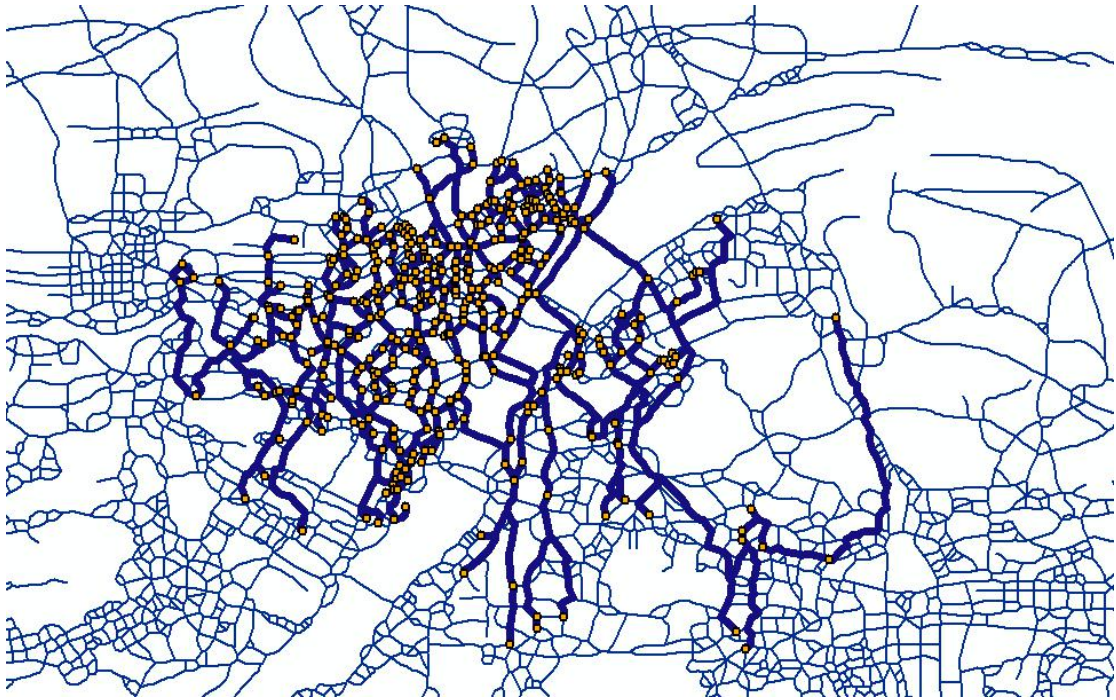


图 5 网络分析结果

三 生成骑行热力图

3.1 2d 热力图分析

生成热力图使用的是核密度分析。在进行分析之前也需要进行准备工作，首先设置工作空间参数，需要修改的部分为工作空间，即为选择创建的地理数据库；处理范围，处理的数据集；栅格分析，即生成栅格的尺寸。单元越小，处理的时间越长，精度越大。

然后就可以进行核密度分析，分析开始还需要选择合适的投影坐标系，不然无法生成正确的高度数据。

国家 2000 坐标系高斯投影下，按 3° 分带，中央经线 114E, EPSG 代码 4547，加带号的话是 38° 带，图 6 为选择的投影坐标系。

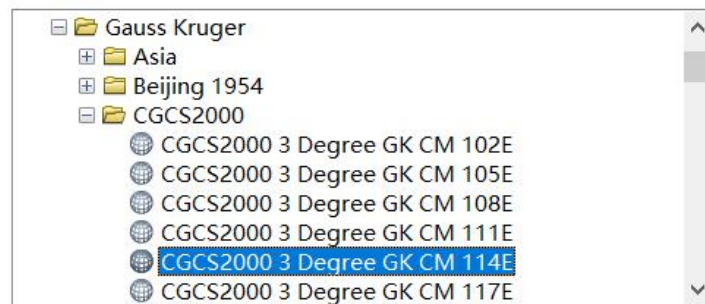


图 6 核密度分析投影坐标系

图 7 为核密度分析结果，该结果当作 2d 热力图分析结果。

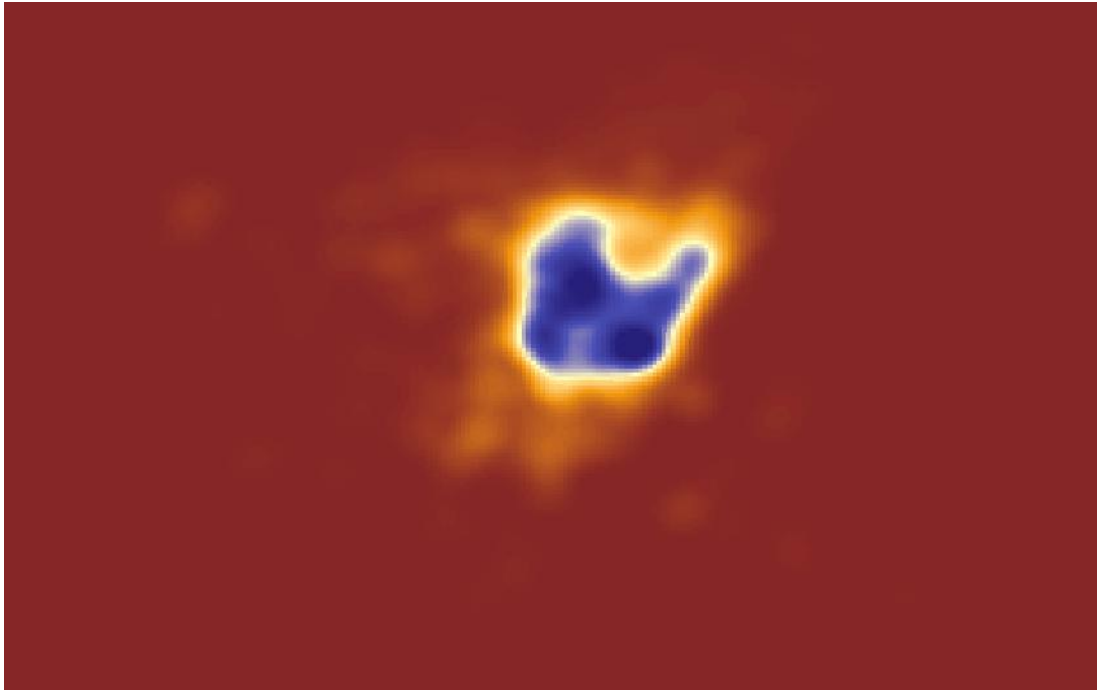


图 7 2d 热力图分析结果

3.2 3d 热力图分析

需要使用到 ArcScene，将第一步生成的 2d 核密度分析生成的栅格数据导入。为了方便进行高度的渲染，这里修改符号系统为已分类，选择合适的分层层次，这里分 10 层，具体设置如图 8 所示，渲染结果如图 9 所示。然后设置基本高度，为了高度显示更明显，设置尽量大的场景单位系数，具体设置如图 10 所示，渲染结果如图 11 所示。

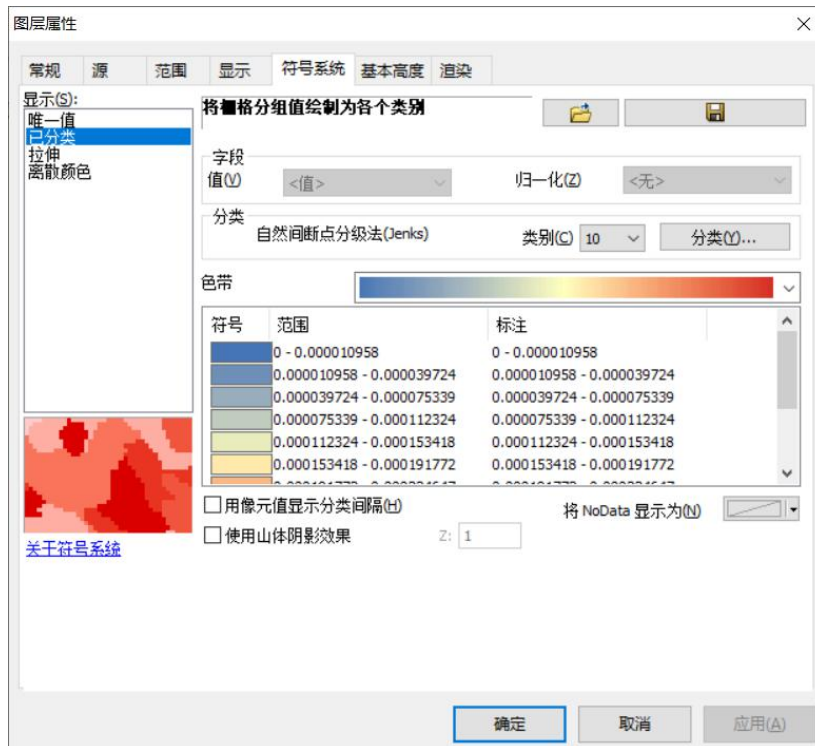


图 8 修改热力分析栅格渲染方式

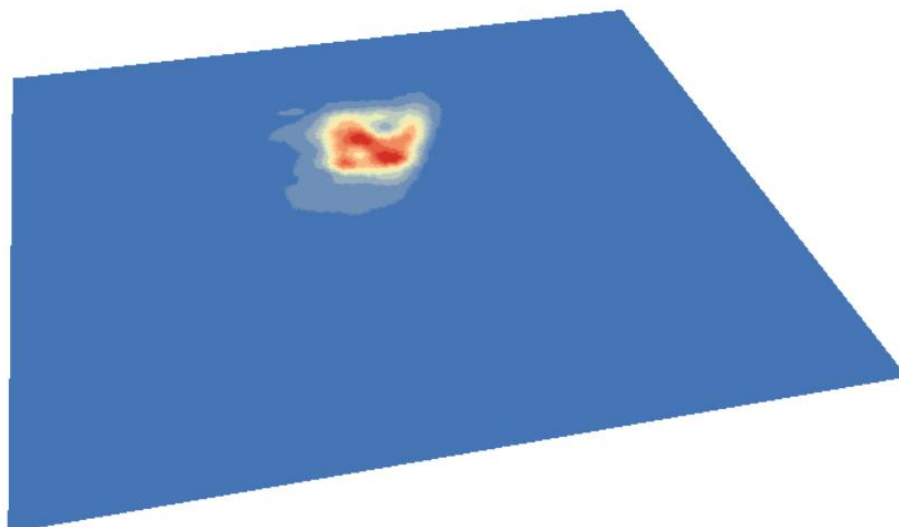


图 9 修改渲染方式后结果

从表面获取的高程

☐ 没有从表面获取的高程值

☒ 在自定义表面上浮动:

C:\Users\76532\Documents\ArcGIS\Default.gdb\KernelD_bic10

栅格分辨率...

从要素获取的高程

☒ 没有基于要素的高度

☐ 使用图层要素中的高程值

用于将图层高程值转换为场景单位的系数: 自定义 10000000.0

图 10 修改高程场景单位系数参数设置

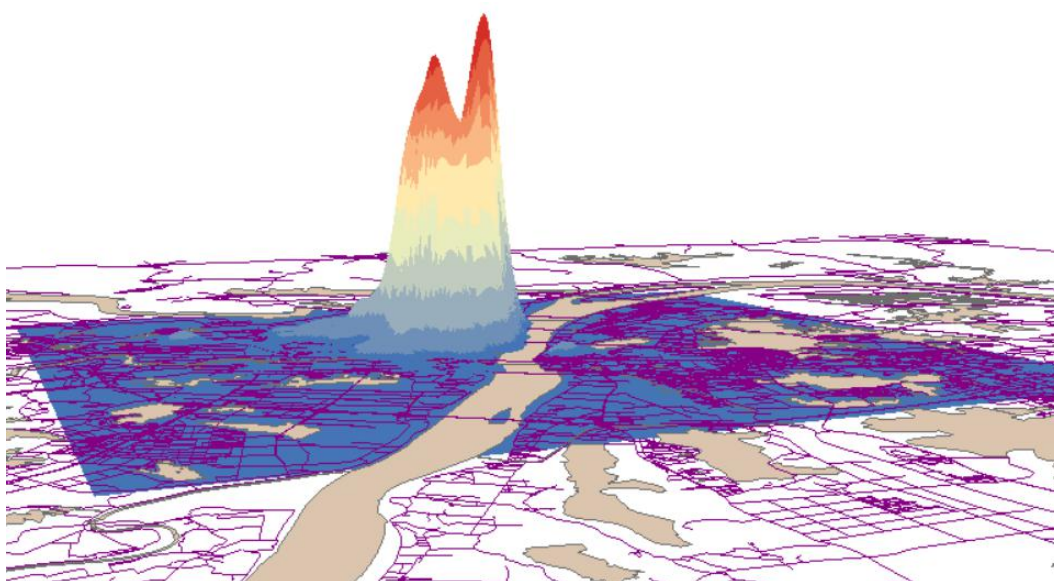


图 11 修改高程系数渲染结果

四 骑行扩散过程模拟

4.1 骑行扩散指标计算及分析

骑行扩散指标一共涉及两个指标 P_i 和 P_c ，其含义分别代表：

P_c 社区扩散程度= 某一时间段停留 **unique** 自行车数/总可使用的自行车数

P_i 单个自行车扩散程度 = **unique** 社区数量/全部的社区数量

必须要解决这些变量如何表示的问题 必须要明确时间段的值，所以不能使用“03-骑行轨迹 0313-0317.csv”，因为里面只有路径坐标，没有时间段信息。必须使用“共享单车 0416”内的 csv 文件中的 9687 个单车数据进行分析。因为每个共享单车记录的路径点不多，那么对应的社区数可能更少，所以在计算 P_i 变量时，特定的时间段就设置为当天。但是在计算 P_c 时还是会使用特定时间进行计算具体原因在 4.3 点会具体解释。

因为“共享单车 0416”数据文件中的轨迹点只分布在洪山区附近，所以只筛选出洪山区的社区数据进行分析，如图 12 所示，否则如果使用全部社区数据进行分析，在洪山区之外的所有社区都是孤立社区，没有参考价值。

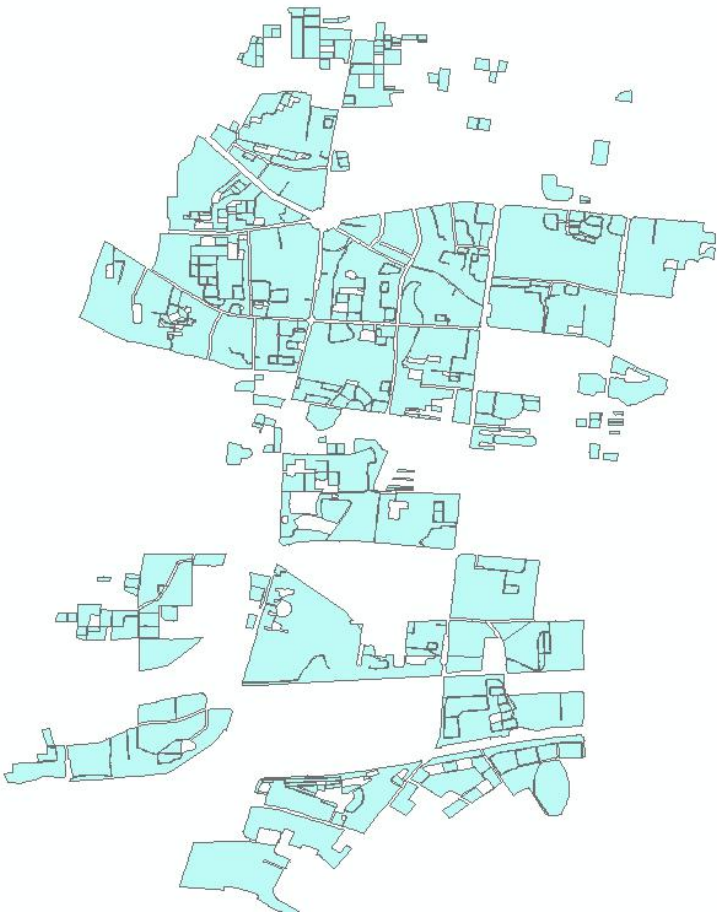


图 12 洪山区社区图层数据

社区数查看属性表，得知有 330 个数据。接下来就是对每辆自行车在一天内访问 **unique**

社区数进行分析。对于访问社区这个概念，个人的理解是进入社区是一个终止的状态，而不是开始的状态，即不是以社区作为起点，而是作为终点，以自行车的终点位置理解为一个社区。

将共享自行车终点坐标进行一个汇总之后，全部显示到社区图层之上，然后对终点和社区面图层进行空间连接，生成一个新的数据集，这个数据集是在终点数据集的基础上新增了社区所代表的 unique id，如图 13 所示，将生成的属性表导出 excel 进行进一步的数据分析。

bicycleNum	startTime	startX	startY	endTime	endX	endY	centerX	centerY	cID
04bqD	2018-04-16T08:33:47	114.399776	30.486383	2018-04-16T20:51:13	114.415405	30.498939	114.414393	30.500809	121
04bZeX	2018-04-16T08:42:51	114.398441	30.476326	2018-04-16T08:43:57	114.411886	30.475851	114.412655	30.47375	308
04bZeX	2018-04-16T09:51:25	114.411886	30.475851	2018-04-16T16:57:19	114.411896	30.464836	114.41165	30.465416	277
04Ea5D	2018-04-16T07:59:34	114.411928	30.493735	2018-04-16T08:00:40	114.405155	30.494523	114.404827	30.493269	64
04Ea5D	2018-04-16T08:40:54	114.405697	30.49412	2018-04-16T08:41:39	114.403062	30.466003	114.409337	30.46841	295
04Ea5D	2018-04-16T19:33:28	114.387963	30.496465	2018-04-16T19:34:33	114.388352	30.500348	114.389201	30.498128	96
04Mbx7	2018-04-16T07:17:21	114.401845	30.504507	2018-04-16T07:35:52	114.395725	30.505612	114.397142	30.503261	140
04c6v7	2018-04-16T21:00:07	114.402279	30.479444	2018-04-16T21:01:26	114.407819	30.502208	114.409086	30.501093	114
04cvYD	2018-04-16T07:36:39	114.390885	30.48524	2018-04-16T07:38:03	114.391763	30.505862	114.3906	30.504488	149
04cvYD	2018-04-16T17:41:38	114.391763	30.505862	2018-04-16T17:42:42	114.400325	30.508493	114.399663	30.50704	165
04cvYD	2018-04-16T18:20:31	114.400325	30.508493	2018-04-16T18:23:01	114.403164	30.502874	114.402887	30.501752	127
052gdD	2018-04-16T12:12:07	114.396616	30.516914	2018-04-16T13:55:26	114.413535	30.507044	114.412263	30.505907	168
052pLX	2018-04-16T07:15:23	114.394518	30.516087	2018-04-16T08:56:23	114.39364	30.510415	114.391483	30.511116	182
056B8x	2018-04-16T07:15:47	114.374181	30.466925	2018-04-16T08:02:40	114.389229	30.477703	114.393132	30.47416	325
056B8x	2018-04-16T11:47:39	114.405324	30.491457	2018-04-16T11:48:34	114.389091	30.477878	114.393132	30.47416	325
05da3r	2018-04-16T18:30:54	114.401416	30.48185	2018-04-16T18:32:13	114.392839	30.504853	114.3906	30.504488	149
05da3r	2018-04-16T18:47:35	114.392839	30.504853	2018-04-16T19:26:25	114.383628	30.499993	114.383254	30.499541	111
05dobx	2018-04-16T13:02:23	114.400592	30.516782	2018-04-16T13:03:23	114.38752	30.514114	114.387891	30.514049	192
05dobx	2018-04-16T15:33:21	114.412291	30.495154	2018-04-16T15:45:05	114.396158	30.517782	114.397218	30.51857	210

图 13 空间连接生成的新的数据集

对于导出的 excel 文件，可以发现对于同一辆自行车在不同的时间段内，会对一个社区访问多次，如图 14 所示，但是对于这种情况只能计数为 1，处理方法是利用 java 数据结构中的 linkedhashmap，该数据结构可以不断地添加数据，并且遇到重复值，不会重复保存，保证了数据的独一性，图 15 为 java 程序分析结果。

Z5mGkP	2018-04-1	114.3988	30.51423	2018-04-1	114.3946	30.51357	114.3947	30.51281	187
nRbxGN	2018-04-1	114.3853	30.5166	2018-04-1	114.3911	30.51042	114.3915	30.51112	182
qRoJYp	2018-04-1	114.392	30.52102	2018-04-1	114.3938	30.51138	114.3915	30.51112	182
6A2O46	2018-04-1	114.3943	30.53117	2018-04-1	114.3942	30.52151	114.3939	30.52165	217

图 14 不同时间段的重复数据

bicycleNumber	spreadPercer
QvPAv	3.030303
85b2Oe	2.7272727
27eRN4	2.4242423
B5Galy	2.4242423
ZADGyB	2.4242423
05kJrV	2.1212122
dX4Aov	2.1212122
Exy4eM	2.1212122
ovR6r7	2.1212122
QKA5kk	2.1212122
wE3EQ7	2.1212122
Y3yo3J	2.1212122
0B7oeX	1.8181818
6Amr7P	1.8181818
6NE4AZ	1.8181818
DxRA2b	1.8181818
evKvzD	1.8181818
Kz54DA	1.8181818
MBwADx	1.8181818
NELmLz	1.8181818

☒ 0.30303028 (3301)
☒ 0.60606056 (1254)
☒ 0.9090909 (545)
☒ 1.2121211 (219)
☒ 1.5151515 (69)
☒ 1.8181818 (17)
☒ 2.1212122 (7)
☒ 2.4242423 (3)
☒ 2.7272727 (1)
☒ 3.030303 (1)

图 15 Pi 分析结果

可知，即使是最活跃的共享单车也没有访问到超过 3% 的社区，说明该地区共享单车流通性并不是很高。

4.2 孤立社区检测

接下来就是选择合适的 P_i 指标进行孤立社区的判断。*We select a group of bikes that have the highest diffusion range ($P_i \geq 65\%$) and track their whereabouts in the region. In total, there are 68 such bikes. We count the total number of visits by these bikes in each community during the year. A community is considered isolated if the count is no more than 68. That is, the average visit time per bike is no more than one in an isolated community.*^[1]

决定选取 2.1212122% 扩散指数以上的共享自行车进行判断，符合该条件的一共有 12 个点，因为社区只有 330 个所以超过 330 的数据集不能使用，并且太大的数据集没有参考价值。利用网络分析得到面上的点数量 Join_Count，即对于每个社区面图层来说上面分布的点的个数，面上点数量如果没有达到 12 个就为孤立社区，图 16 为孤立社区检测结果。

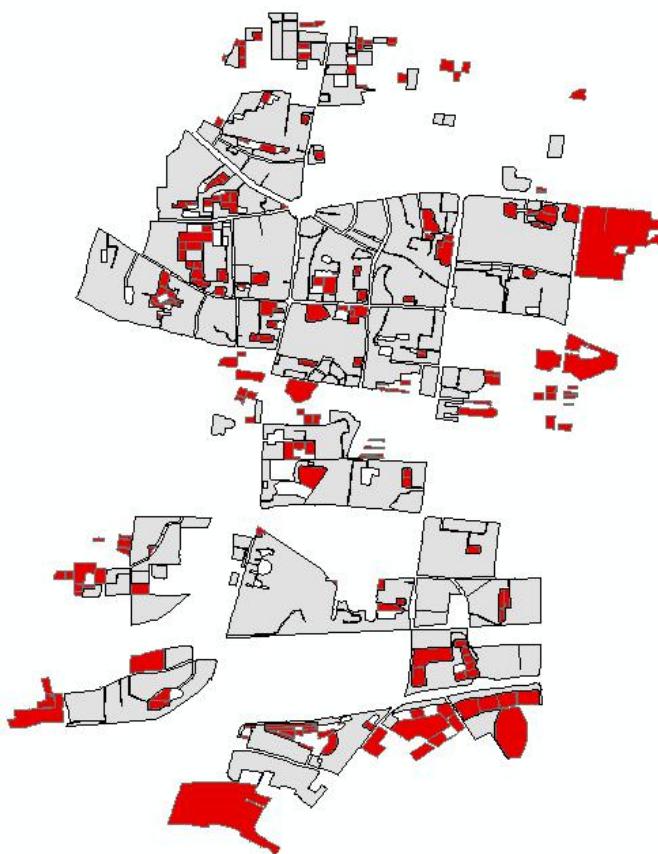


图 16 孤立社区检测结果，红色为孤立社区

4.3 社区扩散指数

对社区扩散指数 $P_c = \text{某一段时间停留 unique 自行车数} / \text{总可使用的自行车数}$ 的公式进行进一步理解。

对于分子，某一段时间停留 unique 自行车数，处理很简单，对单车轨迹数据进行表格合并之后，筛选出某一时段的轨迹终点数据，在图层上显示，并且与社区图层进行空间连接，新增的字段 Join_Count，可以作为分子。

对于分母，总可使用的自行车数，个人认为不能简单的使用自行车的总数进行计算。而是针对于特定的某个社区内可以使用的自行车数量，也就是在一天以内，在该社区停留的自行车数量，可以作为对于该社区可以使用的自行车数量。因为个人理解社区扩散指数，就是对于该社区可停留的自行车的使用频率。如果使用洪山区所有自行车作为分母进行计算，对于偏远的地区，停留的自行车数量很少，并且可能有些自行车在该社区可能使用不到，经过计算依然很少，但是该社区自行车的使用频率是很高的，理论上计算出来的 P_c 结果应该偏大，但是利用所有自行车数量计算就会偏小。

分析过程为提取出 10:00~19:00 时间段的轨迹终点，与社区图层进行空间连接之后新增的数据集，新增 Join_Count 字段，再与一整天的轨迹终点进行空间连接，新增 Join_Count_1 字段，然后在属性表中再新增一个双精度字段 P_c ，利用字段计算器计算 $P_c = \text{Join_Count} / \text{Join_Count_1}$ ，该值就代表社区扩散指数，利用该结果对社区图层进行渲染结果如图 17 所示。

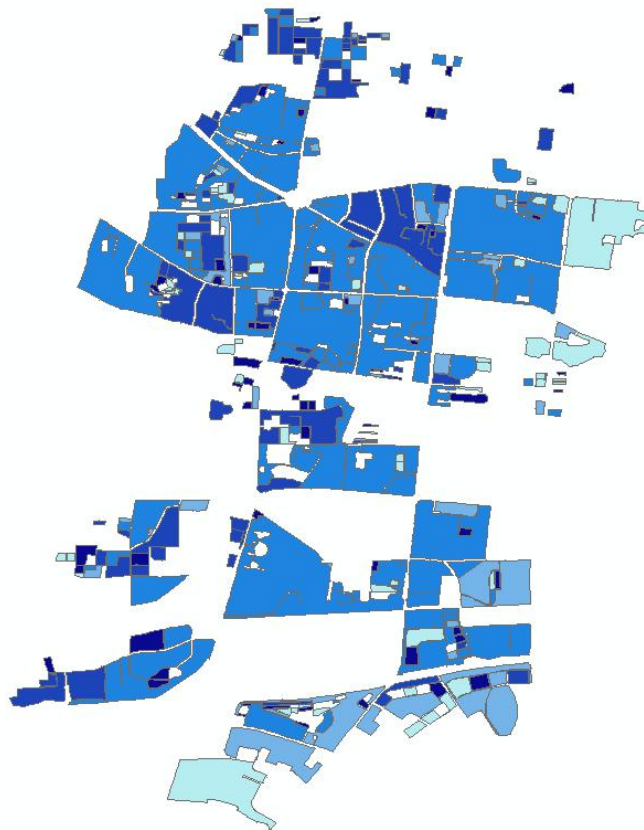


图 17 社区扩散指数分析，颜色越深，说明扩散指数越高，自行车使用频率越高

五 可视化

5.1 vue

考虑到 antV 提供了 vue 语言的案例，为了方便学习和使用，所以该系统搭建了 vue 环境进行开发，集成了 vue 和 html 页面，首页入口页使用 Ant Design Vue 中的布局。

5.2 antV

前期打算所有分析结果利用 antV 接口进行显示，但是发现，如果想要利用接口显示，

就要构造出符合 antV 接口所需要的文件格式，例如前期构造出了自行车轨迹 Json 文件，如图 18 所示，在 antV 中显示如图 19 所示。

```
[
  - {
    - path: [
      - [
        114.25796536341,
        30.574366718337
      ],
      - [
        114.25765467528,
        30.573249885102
      ],
      - [
        114.25893832014,
        30.572876511982
      ],
      - [
        114.25891827439,
        30.572806734127
      ],
    ],
    - color: [
      217,
      220,
      250
    ]
  }
]
```

图 18 自行车轨迹 Json 文件，path 为路径点，color 为随机生成的路径颜色



图 19 antV 接口显示路径图，可以发现，道路数据并不能全部对应上，先前以为是因为坐标系偏移，后来发现是因为没有将河里的点筛选掉，并且轨迹点没有对应到道路上。

所以经过了多次的尝试，放弃了在 antV 显示地理数据的想法，只会利用 antV 进行简单的数据统计显示。

5.3 html

因为不能利用 antV 进行地理数据分析展示，所以使用 qgis 插件 qgis2threejs 对生成的 shp 数据进行网页发布生成 html，然后将页面添加到系统中，通过按钮事件进行跳转。

需要注意的是，其他的数据分析数据可以直接生成，但是对于 3d 热力图还要另外设置参数，需要设置参数如图 20 所示。

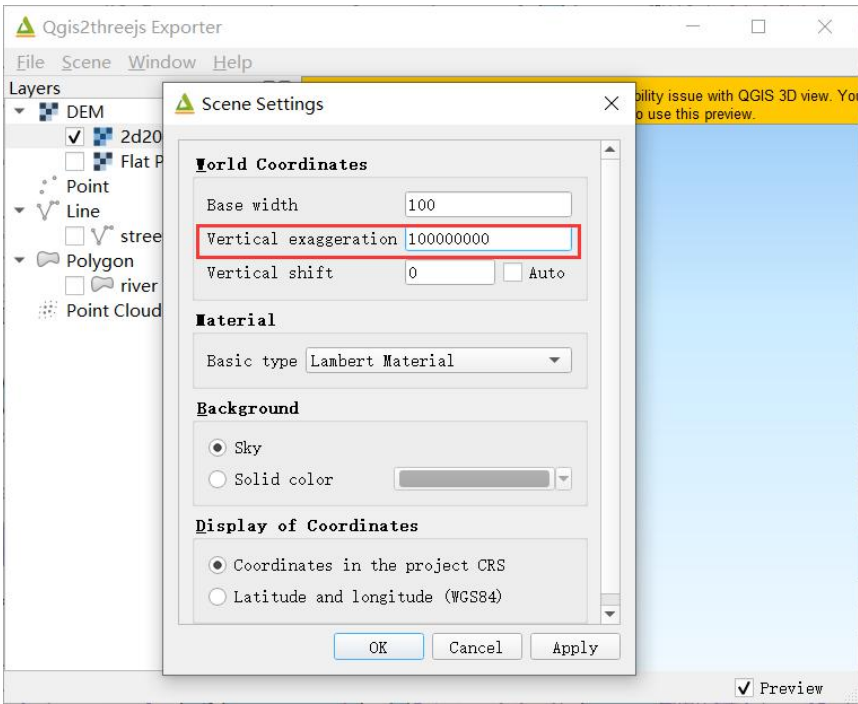
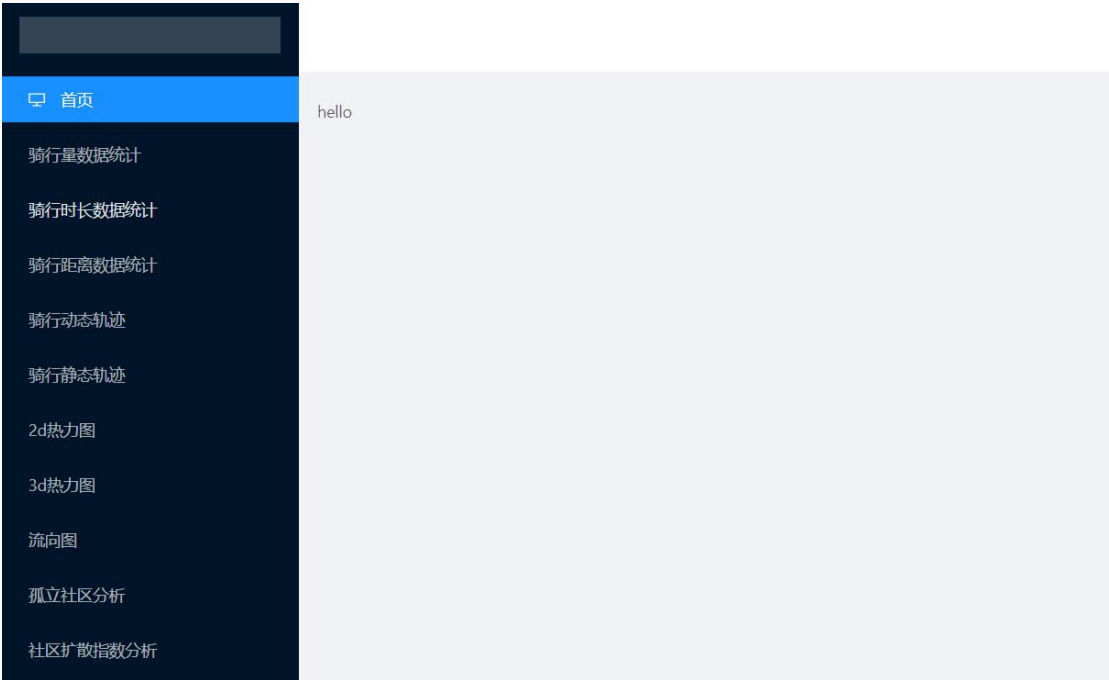


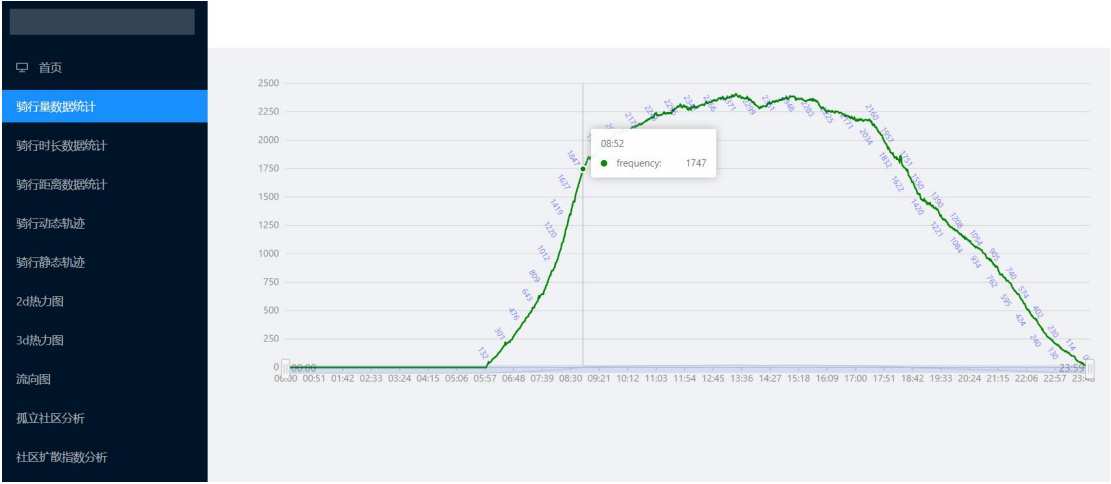
图 20 scene 中的 vertical exaggeration 参数设置，如果不设置该参数，栅格数据的起伏情况就不明显

六 成果展示

6.1 主页面

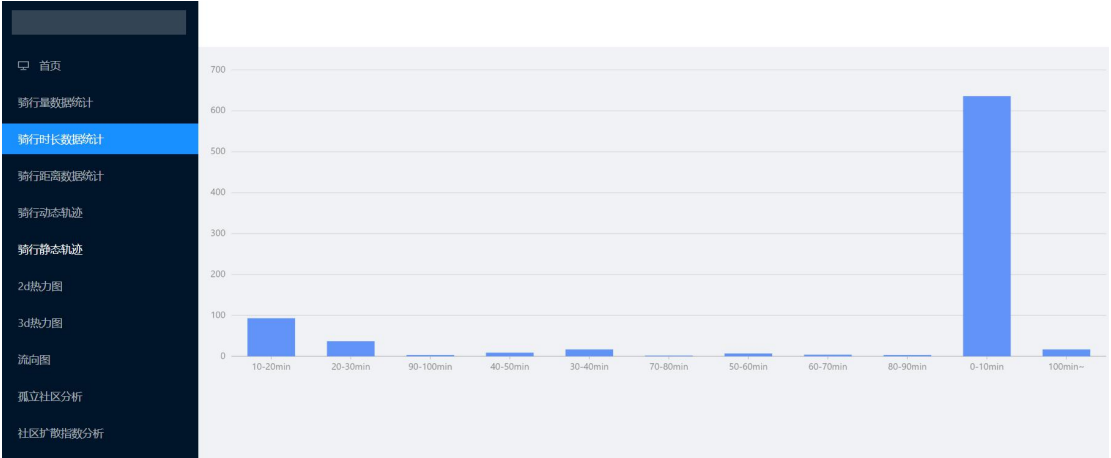


6.2 骑行量数据统计

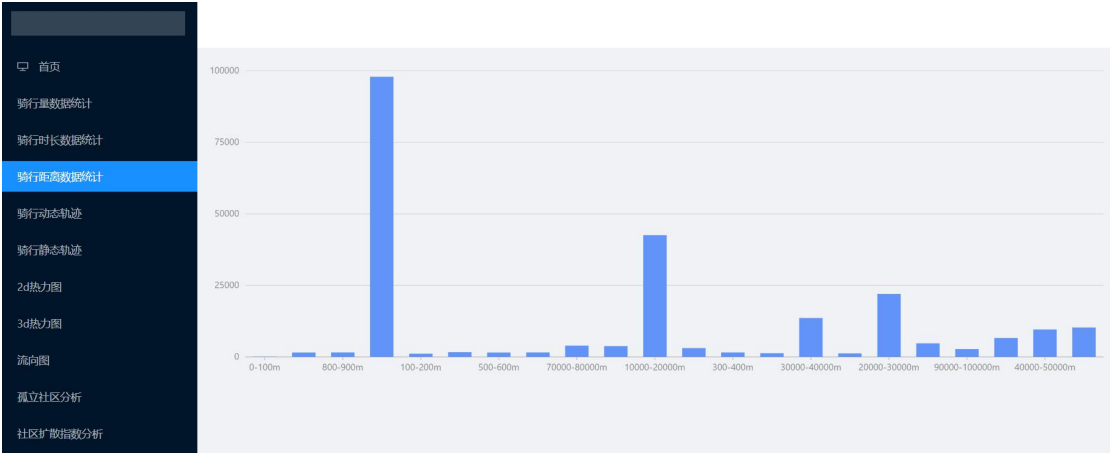


带缩略轴的折线图，可以通过滑动下方的缩略轴，查看任意一段时间内的数据。

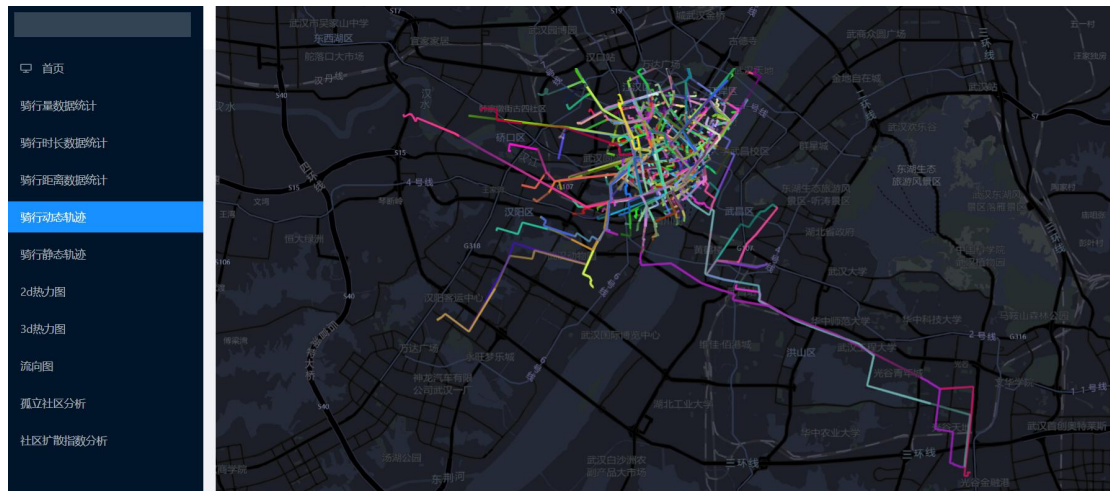
6.3 骑行时长数据统计



6.4 骑行距离数据统计



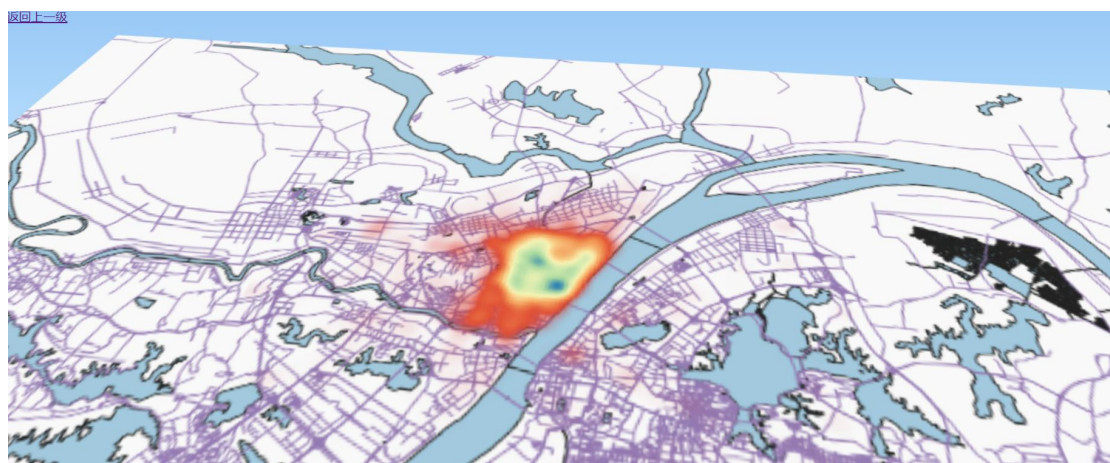
6.5 骑行动态轨迹



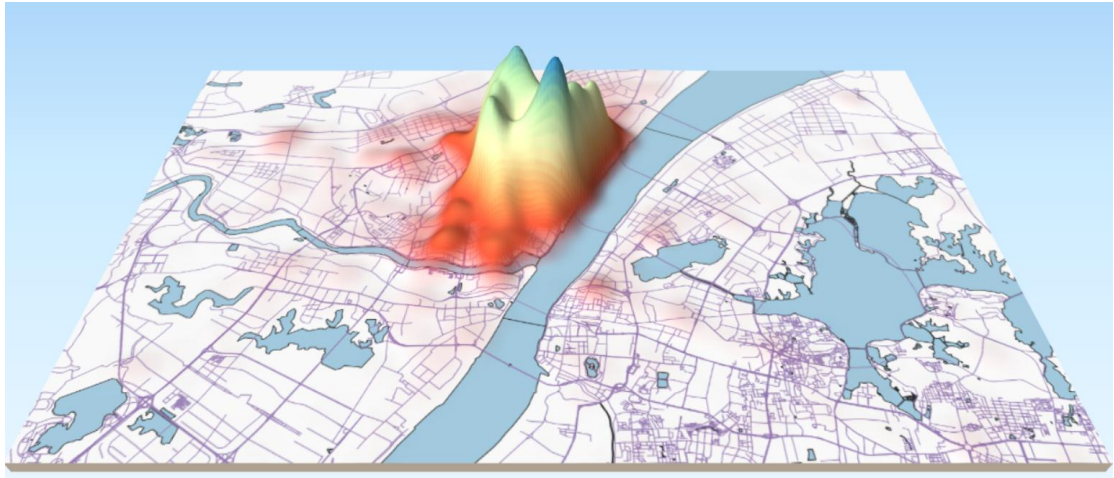
6.6 骑行静态轨迹



6.7 2d 热力图



6.8 3d 热力图



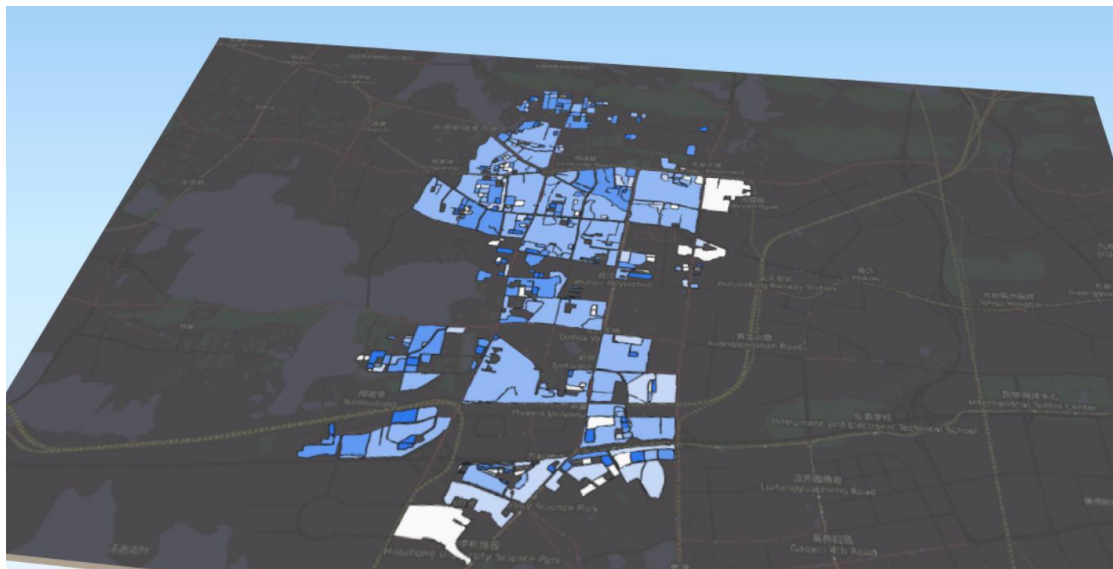
6.9 流向图



6.10 孤立社区分析图



6.11 社区扩散指数分析图



七 实习心得

经过这次实习，学习了如何搭建 vue 环境，并且能够进行简单的开发前端开发，还学会了手动构建 json 数据格式，使用了许多 arcmap 工具进行空间分析。

在进行空间数据分析时，总是会碰到地理空间坐标和投影坐标没有设置好，会发生偏移的情况，体会到了地理空间数据分析对数据的精确性的把握十分重要的，让自己在处理的过程中，尽可能地细心，也再一次地体会到了团队协作中沟通的不易。

八 参考文献

[1] Xueyan Wei,Sida Luo,Yu (Marco) Nie. Diffusion behavior in a docked bike-sharing system[J]. Transportation Research Part C,2019,107.