

城市空间分析方法

Urban Spatial Analysis



龙瀛

清华大学建筑学院

2018年3月20日

《城市模型概论》教学大纲

1. 2月27日W1: 城市模型概论之概论
2. 3月06日W2: 模型基础数据
3. 3月13日W3: 基于规则建模
4. 3月20日W4: 城市空间分析方法
5. 3月27日W5: 模型开发语言
6. 4月03日W6: 元胞自动机模型（栅格）
7. 4月10日W7: 元胞自动机模型（矢量）
8. 4月17日W8: 大数据时代的城市模型展望

2017-2018 学年度春季学期和夏季学期

周次	日 星期	月						
		一	二	三	四	五	六	日
0	2018	19	20	21	22	23	24	25
1	二	26	27	28				
2	三	5	6	7	8	9	10	11
3		12	13	14	15	16	17	18
4		19	20	21	22	23	24	25
5		26	27	28	29	30	31	
6	四							1
7		2	3	4	5	6	7	8
8		9	10	11	12	13	14	15
9		16	17	18	19	20	21	22
10		23	24	25	26	27	28	29
11	五	30						
12		1	2	3	4	5	6	
		7	8	9	10	11	12	13
		14	15	16	17	18	19	20

清华大学

2017-2018 学年度校历

春季学期(2018年)

1. 2月24日、25日教职工照常上班，本科生、研究生2月25日前完成注册。
2. 2月26日全校本科生、研究生开始上课。
3. 妇女节：3月8日正常上课，女教工放假半天。
4. 清明节：4月5日-7日放假调休，共3天。
5. 校庆及“五一”：4月28日、29日（校庆日）教职工照常上班；4月30日-5月4日放假调休，共5天。
6. 端午节：6月18日放假，与周末连休。
7. 第8周期中测验。第17周、18周末考试。

考核方式：考查

- 成绩构成：出勤及过程（30分）+大作业（70分）
- 大作业（任选一种形式）：
 - 形式1：每人撰写某一类城市模型综述的课程论文
 - 形式2：每人撰写城市模型发展趋势与未来展望的课程论文
 - 形式3：利用课程发放的北京五环内数据，开发一个地块尺度的轻量级城市模型（straight forward and light-weight），并附模型介绍（建议2-3人一组）
 - 建议选题：城市开发密度模拟
 - 欢迎与任课教师讨论（建议OPEN OFFICE HOUR时间）
- 提交方式：W13周末（5月27日）前提交给助教陈婧佳
 - W8结课后也同样欢迎约任课教师讨论大作业

零、北京五环内数据介绍



图层一览

[Building2004 建筑物2004](#)

[building2017 建筑物2017](#)

[CBD 中央商务区](#)

[chanye 产业功能区](#)

[date 2015_06_12 Hour 00/10/22 宜出行](#)

[ditiezhan 地铁站](#)

[ecolines 生态红线](#)

[existing 现状地块](#)

[plan2011all 规划地块2011](#)

[POI2016gaode 兴趣点2016](#)

[R5 五环边界](#)

[road2016gaode 道路2016](#)

[subdistricts 乡镇街道办事处](#)

[taz2010 交通分析小区](#)

[tiananmen 天安门](#)

[ww_LSWH 历史文化保护区](#)

[zhongguancun 中关村](#)



B

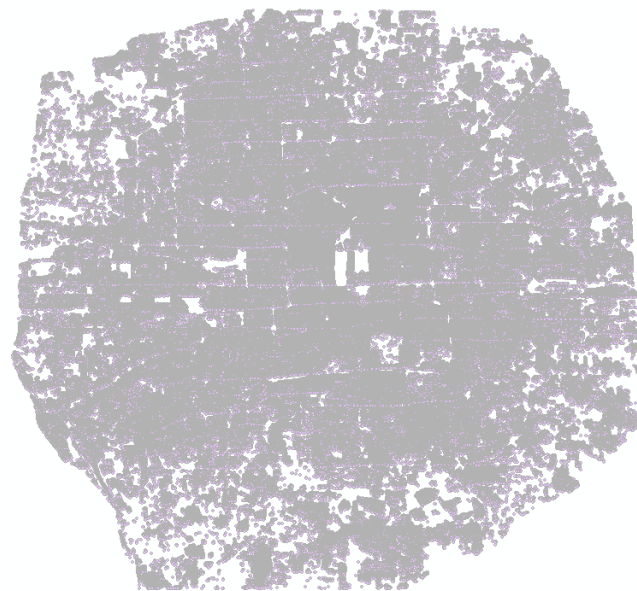
1. building2004

建筑物2004

图层类型File Type: Point

数据量Amount: 1101814

图层意义Meanings: 2004年北京五环内建筑物位置与面积



2. building2017

建筑物2017

图层类型File Type: Polygon

数据量Amount: 267065

图层重点Key Word: #建筑物面积 #建筑物楼层数

图层意义Meanings: 2017年北京五环内建筑物位置与面积





C

3. CBD

中央商务区

图层类型File Type: Polygon



数据量Amount: 14

图层意义Meanings: 中央商务区位置





C

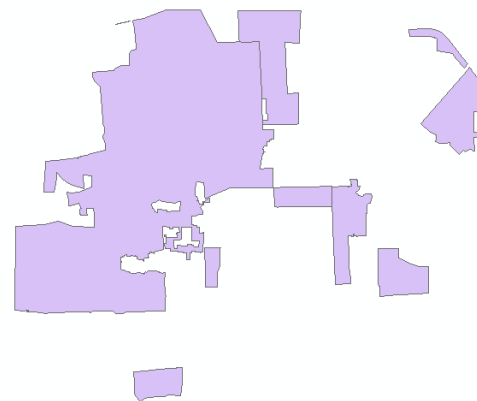
4. chanye

产业功能区

图层类型File Type: Polygon

数据量Amount: 10

图层意义Meanings: 北京五环内产业功能区分布与范围





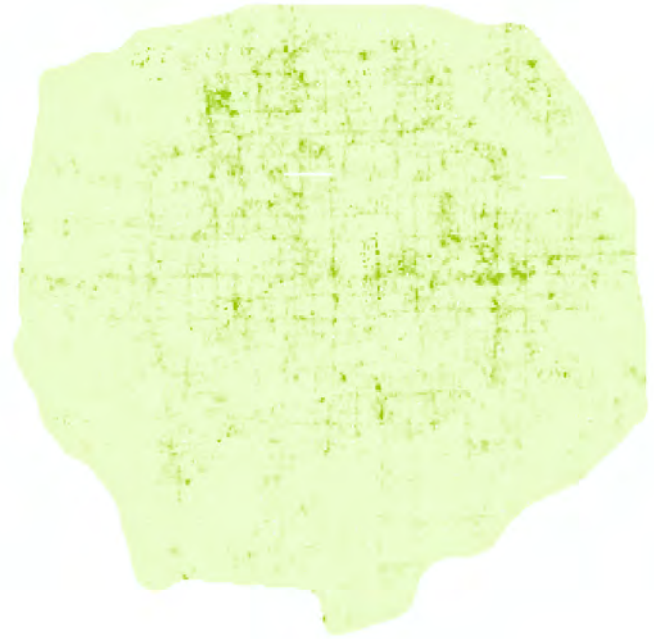
D

5. Date_2015_06_12_Hour_00/10/22

宜出行

图层意义Meanings:

2015年某一天三个时间点的北京五环内宜出行数据





D

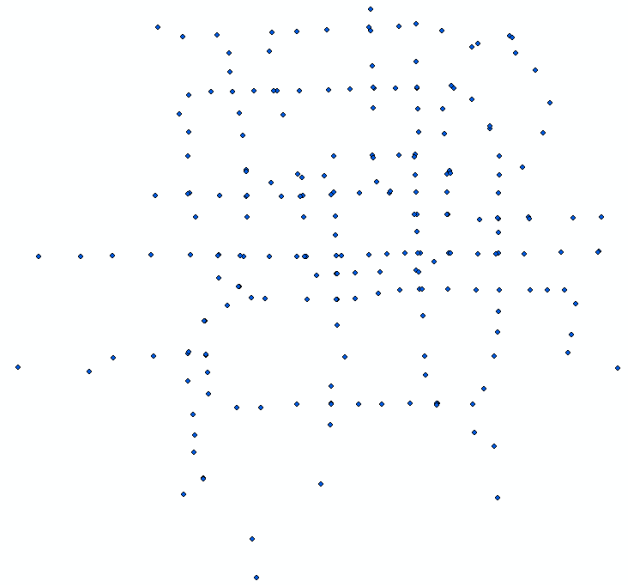
6. ditiezhan 地铁站

图层类型File Type: Point

数据量Amount: 222

图层重点Key Word: #地铁站位置
#站名 #所在线路

图层意义Meanings: 北京五环内地铁站站点分布





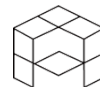
7. Ecolines

生态红线

图层类型File Type: Polygon

数据量Amount: 8284

图层意义Meanings: 北京五环内生态红线范围



8. existing

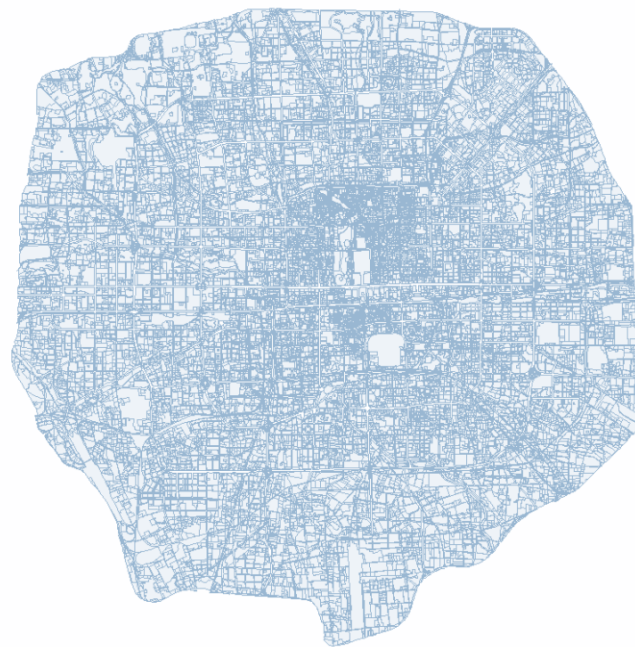
现状地块

图层类型File Type: Polygon

数据量Amount: 31836

图层重点Key Word: #地块面积
#地块属性

图层意义Meanings: 北京五环内现状地块用地性质





9. Plan2011all 规划地块2011

图层类型File Type: Polygon

数据量Amount: 100912

图层意义Meanings: 北京五环内规划地块位置与范围





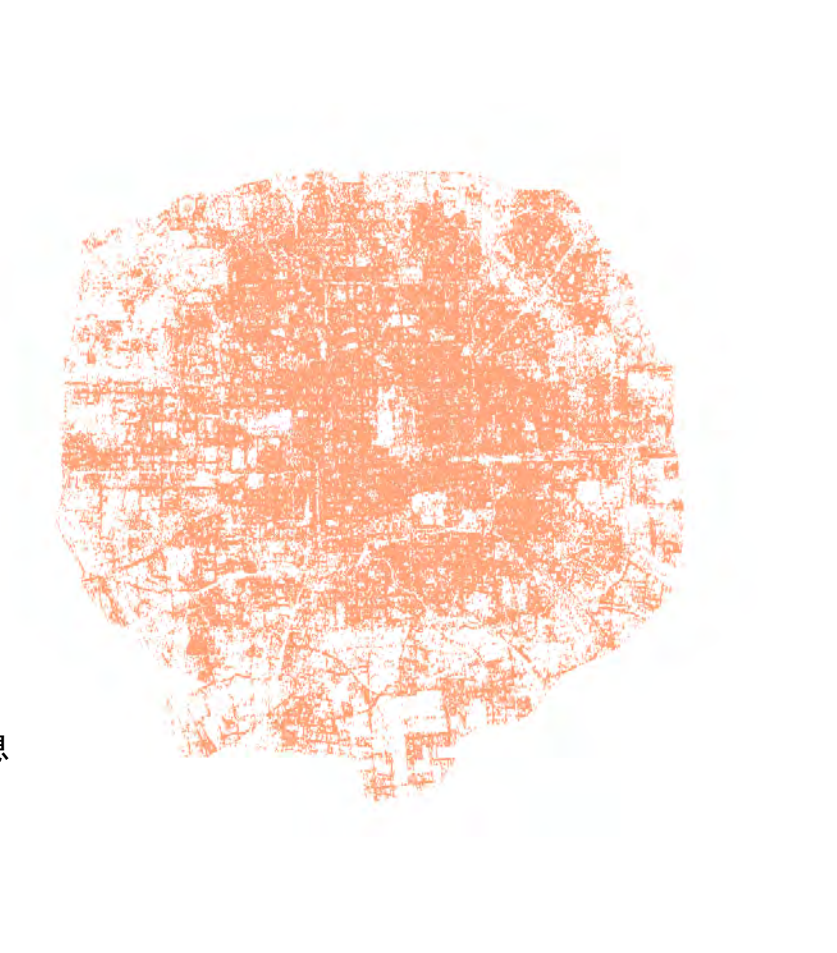
10. POI2016gaode 兴趣点2016

图层类型File Type: Point

数据量Amount: 783658

图层重点Key Word: #兴趣点位置 #兴趣点类别

图层意义Meanings: 2016年北京五环内兴趣点位置与内容信息





R

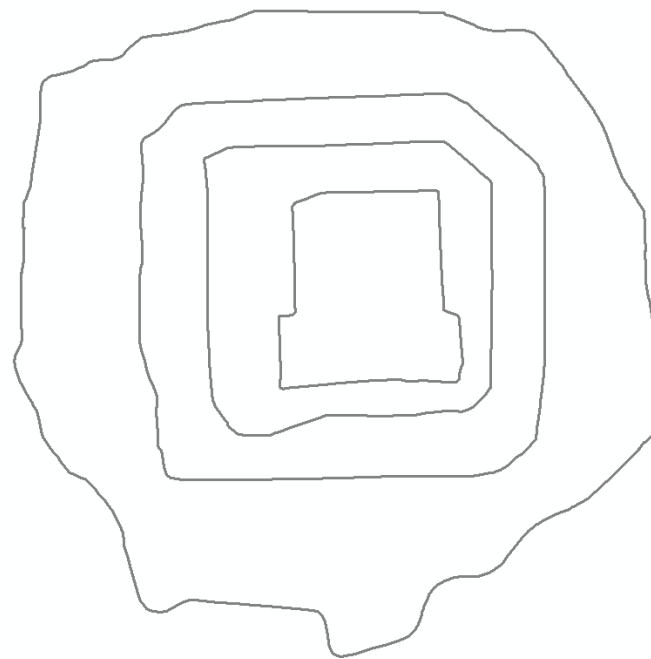
11. R5

环路边界

图层类型File Type: Polygon

数据量Amount: 4

图层意义Meanings: 北京五环环路边界





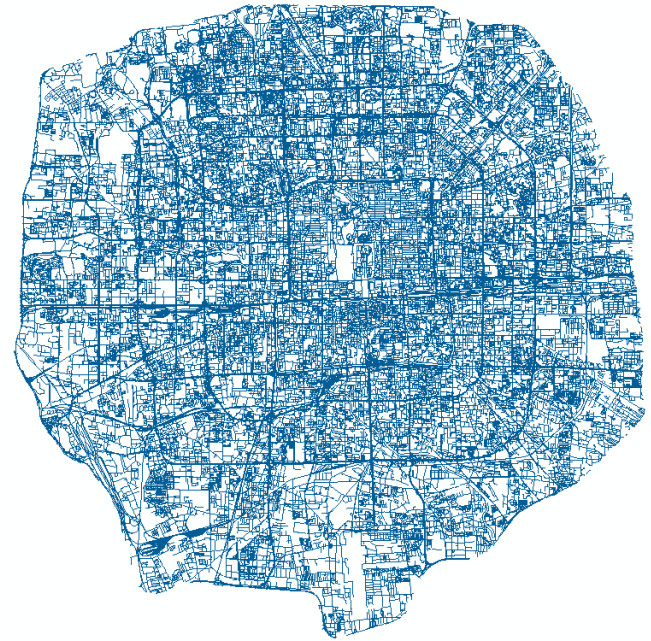
12. road2016gaode 道路2016

图层类型File Type: Polyline

数据量Amount: 136561

图层重点Key Word: #道路名称 #道路级别

图层意义Meanings: 2016年北京五环内各级道路相关信息





13. subdistricts

乡镇街道办事处

图层类型File Type: Polygon

数据量Amount: 307

图层重点Key Word: #行政区名称

图层意义Meanings: 北京五环内各乡镇街道办事处行政范围





T

14. taz2010

交通分析小区

图层类型File Type: Polygon

数据量Amount: 1911

图层重点Key Word: #地块内区域数量 #街道数量

图层意义Meanings: 2010年北京五环内交通分析小区





T

15. tiananmen

天安门

图层类型File Type: Point

数据量Amount: 1

图层意义Meanings: 天安门位置





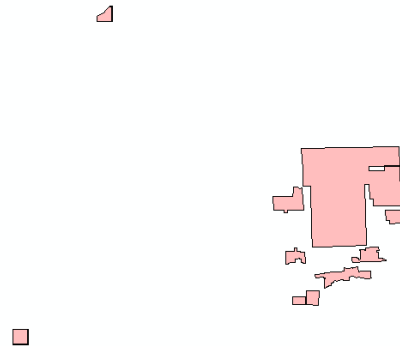
W

16. ww_LSWH

历史文化保护区

图层类型File Type: Polygon
数据量Amount: 16

图层意义Meanings: 历史文化保护区位置与范围





Z

17. zhongguancun

中关村

图层类型File Type: Polygon

数据量Amount: 1

图层意义Meanings: 中关村所在位置与范围



一、数据分析与统计概述

为什么要做数据分析与统计

- 数据→信息→知识→智慧
- 为城市模型提供输入数据，以及分析模型的输出数据

□ 不同类型的数据，需要有不同的分析与统计方法

• 数据类型

- 空间数据：点、线、面、块
- 轨迹数据：若干点的集合、起始点构成的线段
- 文字数据：自然语言处理（NLP）→ 标签/情感
- 图片数据：计算机视觉（CV）→ 标签
- 视频数据：大量图片的集合
- 其他数据：声音、气味。。。



不同类型的数据，需要有不同的分析与统计方法

• 数据类型

- 空间数据：点、线、面、块
- 轨迹数据：若干点的集合、起始点构成的线段
- 文字数据：自然语言处理（NLP）→标签/情感
- 图片数据：计算机视觉（CV）→标签
- 视频数据：大量图片的集合
- 其他数据：声音、气味。。。

• 殊途同归，最终回归到：

- 点线面的空间分析与统计
- 有标签（属性）的点线面的空间分析与统计



数据预处理

- Excel、Access、ArcMap、SPSS无论哪个软件，都需要进行数据预处理



数据预处理

- Excel、Access、ArcMap、SPSS无论哪个软件，都需要进行数据预处理
- 数据预处理做什么（个人经验、先后顺序）？
 1. 是否有明显的数据缺失（是否完整）
 2. 冗余字段的删除
 3. 保留字段改名称
 4. 增加唯一ID字段并计算其数值（可追溯）
 5. 各个字段中的异常或者空值的处理
 6. 空间对象是否重叠/重复（如天安门的例子）

数据预处理

- Excel、Access、ArcMap、SPSS无论哪个软件，都需要进行数据预处理
- 数据预处理做什么（个人经验、先后顺序）？
 1. 是否有明显的数据缺失（是否完整）
 2. 冗余字段的删除
 3. 保留字段改名称
 4. 增加唯一ID字段并计算其数值（可追溯）
 5. 各个字段中的异常或者空值的处理
 6. 空间对象是否重叠/重复（如天安门的例子）
- 需要有强迫症的思维

常用的数据分析与统计的软件

- 办公自动化软件
 - Access、Excel
- 数据库平台
 - Access、SQL Server、ORACLE
- 地理信息系统软件
 - ESRI ArcGIS、GeoDA
- 统计分析工具
 - SPSS、STATA
- 语言类
 - Python、R
- 大数据相关软件
 - 火车采集器（也能进行数据预处理！）、Tableau
- 自然语言处理
 - Python/R的包、ROST（武汉大学）
- 各类API
 - 如微软的计算机视觉API可用于分析图片数据

- 多数工作都能在Excel中完成（辅以插件和VBA）！

二、基于SPSS和Excel的数据统计分析

SPSS的很多功能可以用Excel实现



常用功能

- 数据处理
 - 计算字段
 - 统计性描述
 - 相关分析
 - 0.8
 - 主成分分析 (PCA)
 - 回归分析
 - 回归前的自相关 (autocorrelation) 检查 (VIF)
 - 二元/多元回归、线性/非线性回归
 - 部分数据的回归分析
 - 对数ln (如房价)
 - 聚类分析
 - K-means
-
- 软件展示

Excel中的统计分析功能

Baidu 知道 excel 回归分析 搜索答案 我要提问

最佳答案

用EXCEL做回归分析主要有图表法和函数法:

1、图表法:

选择参与一元线性回归两列数据(自变量x应在应变量y的左侧),插入图表,选择散点图。

选择图表中的数据系列,右击,添加趋势线,点击“选项”选项卡,勾选“显示公式”。显示R平方值。

注意显示出的R2值为R的平方,需要用SQRT()函数,计算出R值。

2、函数法

若X值序列在A1:A100单元格, Y值序列在B1:B100单元格,

则线性公式的截距b

```
=INTERCEPT(B1:B100,A1:A100)
```

斜率k

```
=SLOPE(B1:B100,A1:A100)
```

相关系数R

```
=CORREL(A1:A100,B1:B100)
```

或

```
=CORREL(B1:B100,A1:A100)
```

上述两种方法都可以做回归分析,同时结合图表和函数会取得更满意的效果。

- <https://zhidao.baidu.com/question/46033141.html> (左图)
- <http://blog.csdn.net/dashenghuahua/article/details/53308691> (右图)

CSDN 博客 学院 下载 GitChat 论坛 问答 商城 VIP 活动 招聘 ITeye 码云 CSTO

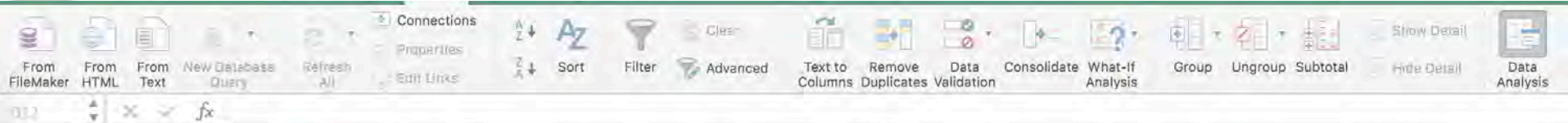
数据分析常用的Excel函数都在这里!

原创 2016年11月23日 18:08:54 标签: 数据分析 数据 excel函数 excel 2405

excel是我们工作中经常使用的一种工具,对于数据分析的朋友来说,这也是处理数据最最基础的工具,然并不是每个人都掌握得很好,玩得很溜哦,主要是excel功能强大,我们平时处理的时候很多函数功能根本不清楚怎么用。今天我们大圣众包小编根据平时工作需要及以往遇到的问题给大家总结些EXCEL中的函数。



Excel中的统计分析功能（高级）

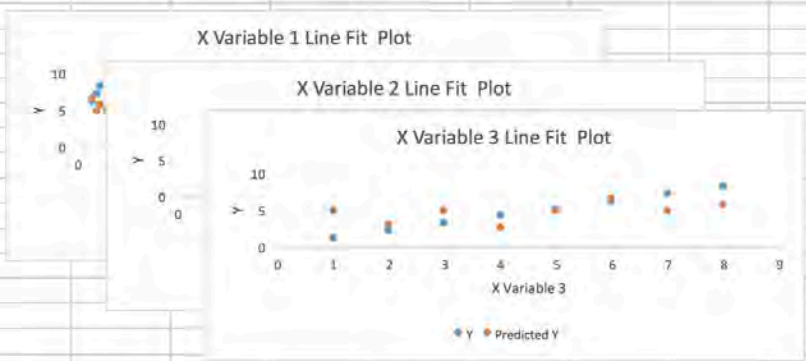


SUMMARY OUTPUT

Regression Statistics				
Multiple R	0.515609			
R Square	0.26585264			
Adjusted R Square	-0.2278063			
Standard Error	2.48331186			
Observations	8			

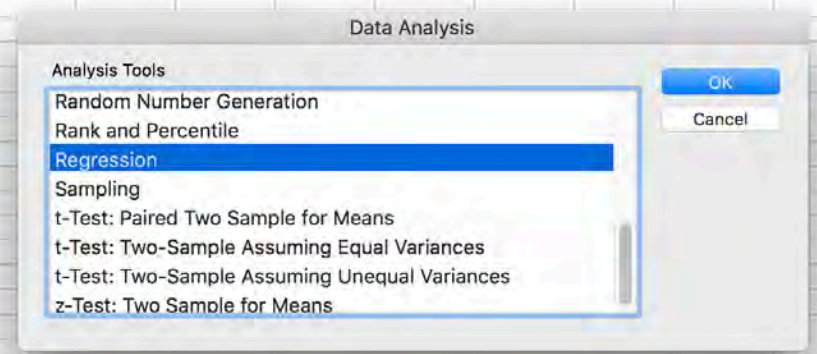
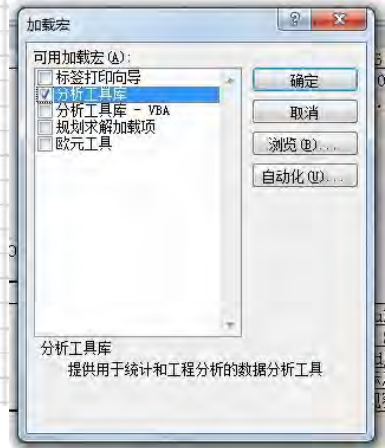
ANOVA					
	df	SS	MS	F	Significance F
Regression	3	11.1658111	3.72193702	0.90531091	0.51307632
Residual	5	30.8341889	6.16683779		
Total	8	42			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.83392285	1.27127736	3.80241401	0.01259734	1.56600035	8.10184535	1.56600035	8.10184535
X Variable 1	0	0	65535	#NUM!	0	0	0	0
X Variable 2	-0.0092319	0.00737461	-1.2518473	#NUM!	-0.0281889	0.00972515	-0.0281889	0.00972515
X Variable 3	0.00358991	0.00541941	0.66241679	0.53701226	-0.0103411	0.01752095	-0.0103411	0.01752095



RESIDUAL OUTPUT

Observation	Predicted Y	Residuals	Standard Residuals
1	4.67799487	-3.6779949	-1.8734403
2	2.91565683	-0.9156568	-0.4664031
3	4.65184667	-1.6518467	-0.8413922
4	2.47774065	1.52225935	0.77538497
5	4.63235927	0.36764073	0.18726316
6	6.35654924	-0.3565492	-0.1816136
7	4.81391163	2.18608837	1.11351595
8	5.47394083	2.52605917	1.28668503



- <http://www.officezhushou.com/excel2010/2854.html>
- <https://www.zhihu.com/question/19600307> ()

三、基于GeoHey的数据可视化

应用 最近使用

[显示全部](#) →



数据上图



打标签



地图DIY



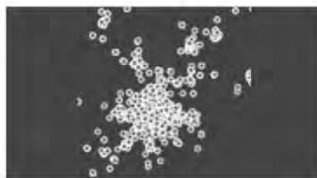
地图涂鸦

我的数据 最近浏览

[显示全部](#) →



2
1257条记录
2016-09-19 22:12



街景停留意愿及客观变化打分
1974条记录
2016-09-07 21:49



busbuffer500_2
5182条记录
2015-11-09 02:13



walkscore
64357条记录
2016-07-11 14:20

公共数据 最近浏览

[显示全部](#) →



晚清中国社会经济分区
10条记录

2015-07-24 02:16

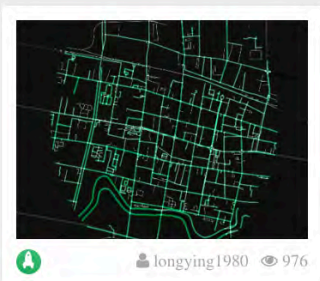
- Carto/MapBox的中国版本，提供便捷的空间数据在线可视化与分享
- www.geohey.com
- （国外的还有MapBox在线可视化平台）



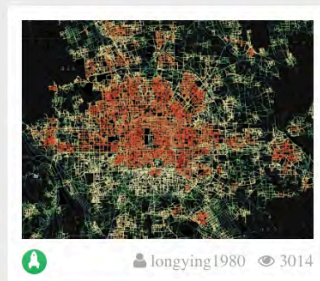
可视化 数据

最新 ▾

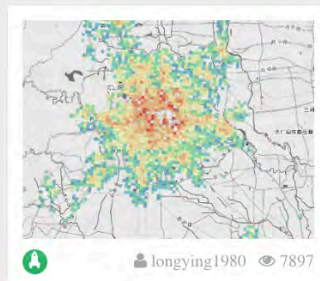
搜索



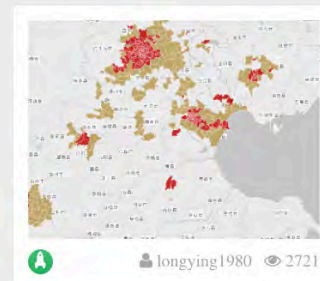
中国城市街道绿视率评估 Street gree...



中国主要城市街道步行性评价 (W...



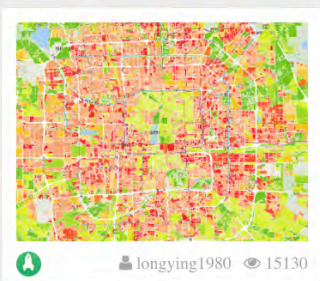
Urban Vitality for Chinese Cities



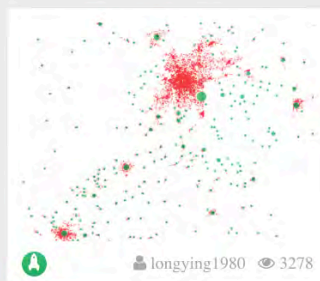
功能性城市地域



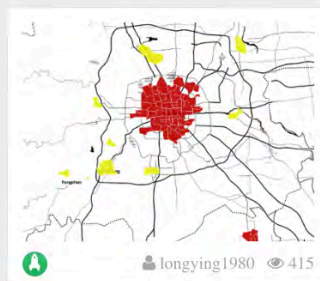
Bus Landscapes in China



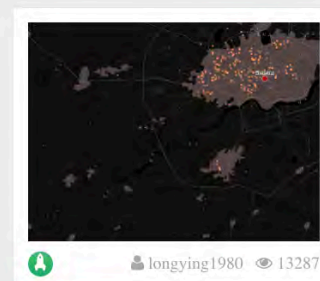
中国精细化尺度城市形态



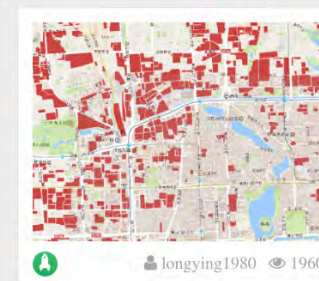
The redefined Chinese city system



Subcenters of China at Township Level



中国居住项目活力及鬼城景观(点越...



北京居住小区项目一览

• <https://geohey.com/gallery/bcl>

利用GeoHey进行数据可视化的基本步骤

• 数据上传

- 空间数据支持zip格式（注意需要在文件目录下对几个ShapeFiles文件进行压缩，不要在文件夹上压缩，也不要rar格式）
- 属性数据csv等格式，可以在GeoHey上空间化（geocoding），但是有数量限制
- 选择数据来源为“GPS设备”

• 数据表达（类似ArcMap）

- 数据筛选
- 符号化

• 工程设置

- 名称、简介
- 是否公开

• 数据分享

- 链接形式
- 嵌入网页模式

上传数据



PlanParcel_v3.zip

数据来源: 百度地图 高德等互联网地图 GPS设备 [查看帮助](#)

[重新选择](#) [上传文件](#)

GeoHey可视化的实际操作

- 鼓励GIS和编程基础较少的同学们：
 - 利用Excel建模
 - 在GeoHey中可视化模拟结果
- 鼓励GIS和编程基础较好的同学们：
 - 利用ArcGIS+Python建模
 - 在ArcGIS或GeoHey中可视化模拟结果

四、基于ArcGIS的数据分析

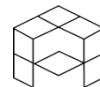
GIS vs CAD

GIS vs Big Data

点、线与面



- **建立GeoDatabase (GDB)**
 - 如果工程不大，建议Personal GDB，物理格式为mdb，与微软Office的Access通用，便于属性数据预处理、分析与统计
 - 栅格图层、Toolbox、网络分析和拓扑检查等也可以存入并在GDB中操作
- **数据管理（类似Windows的资源管理器或Mac的Finder）**
 - 导入图层、过程图层存储、删除图层
- **调用ArcToolbox**
 - 便于操作图层





ArcMap

- 显示与管理
 - 加载、增加/删除字段、选择一部分对象
- 计算
 - 计算字段（属性与空间）、关联join、空间关联spatial join
- 分析与统计
 - 一般需要调用Toolbox
 - 字段的summarize功能
- 可视化
 - 符号化、layout视图



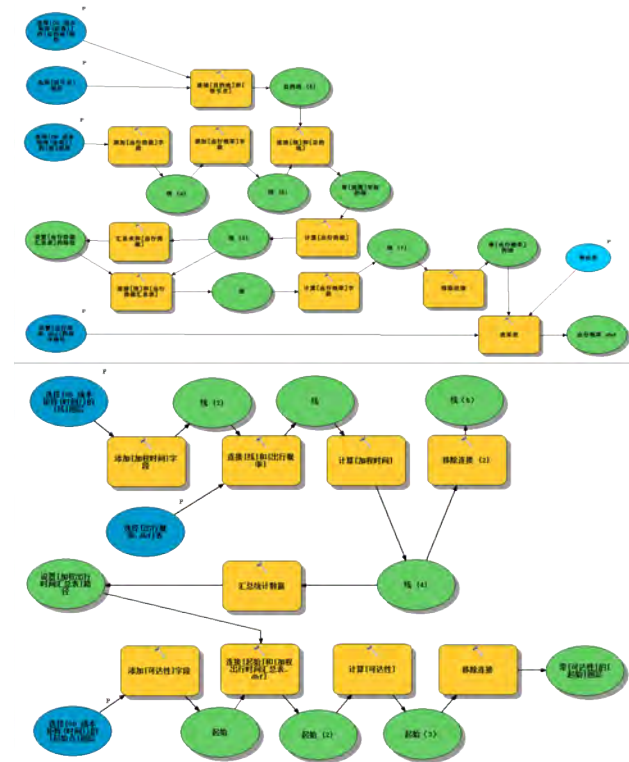


• 个人常用的工具箱展示

- 空间分析 (buffer、overlay、union)
- 空间统计 (全局自相关Moran's I、局部子相关LISA、最小二乘回归OLS、地理加权回归GWR)
- 空间数据处理 (repair)

• Model Builder

- | | |
|------------------------------------|------------------------------------|
| Aggregate Polygons | Geographically Weighted Regression |
| Append | Identity |
| Batch Project | Intersect |
| Buffer | Kernel Density |
| Clip | Multipart To Singlepart |
| Create Personal GDB | Ordinary Least Squares |
| Define Projection | Point Statistics |
| Dissolve | Project |
| Eliminate | Raster to Polygon |
| Eliminate Polygon Part | Reclassify |
| Erase | Repair Geometry |
| Euclidean Distance | Sample |
| Extend Line | Simplify Line |
| Feature To Line | Sort |
| Feature To Point | Spatial Join |
| Feature To Polygon | Summary Statistics |
| Feature to Raster | Trim Line |
| Focal Statistics | |
| Generalize | |
| Generate Near Table | |
| Geographically Weighted Regression | |



ArcGIS Scripting using Python

- 每个Toolbox都可以利用脚本实现（具体参见每个工具箱的help）
- 推荐利用ArcGIS安装的默认Python版本和编译器（2.7.x而不是3.x）
- 课外参考资料待放到网站
 - ExtendingArcGISWithPython
 - Programming ArcGIS 10.1 with Python Cookbook

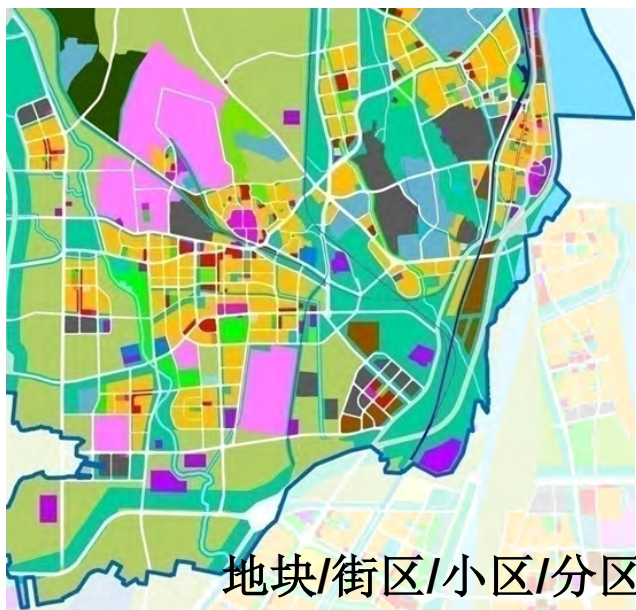
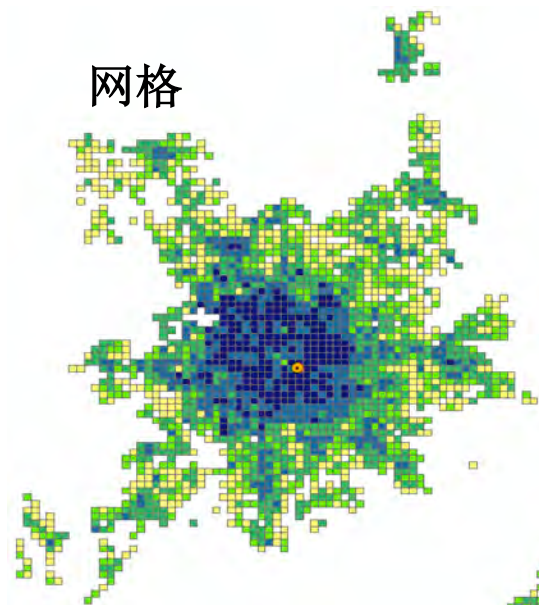
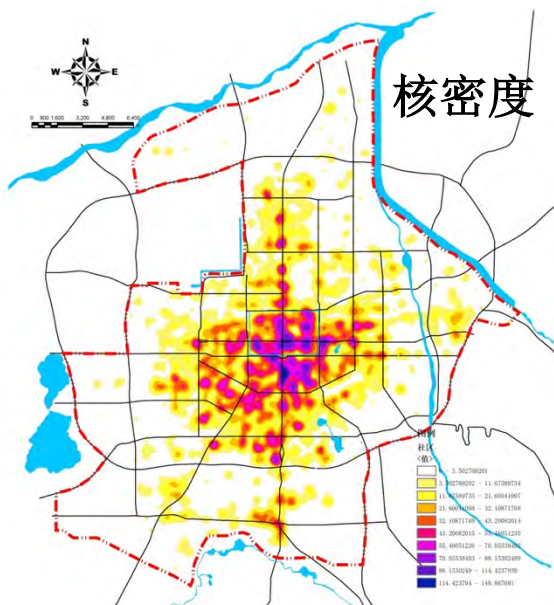


- 会了Python与不会所做的工作的巨大差异（个人经历的分享）
- 具体演示：

ArcGIS的几个常用技巧

- 推荐使用ArcGIS 10.x的英文版本
- 选择合适的分析单元，生成图层作为日后常用
- GeoDatabase（中的OBJECTID/OID会随着操作而变化，建议单独建立一个字段如BLOCK_ID/STREET_ID表示唯一的空间对象ID
- 能用属性来表示，就不用额外生成新的图层
- 尽可能地利用GeoDatabase（gdb）来管理空间数据，而不是ShapeFiles（如果工程不大，建议mdb）
- 不要随意删除mdb中空间图层的属性对象（行）
- Toolbox中的Repair Geometry是个好工具
- 数据库释放空间的方法（Compact database，在数据库上右键）

分析单元的选择 (!!! 我们看待城市的视角)

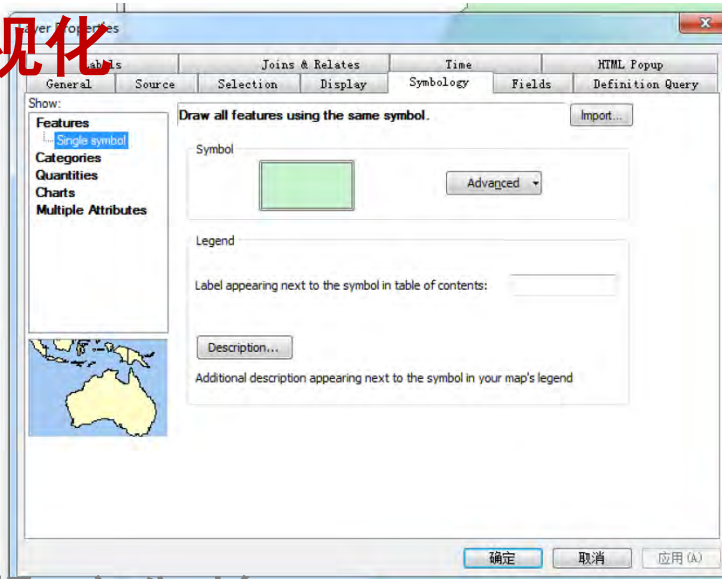


不同分析单元的数据分析

- **核密度：**
 - kernel density工具箱
 - point density是另一种选择
- **网格：**
 - Create fishnet生成不同尺度的网格
 - ArcMap中的spatial join
- **地块：**
 - 利用道路网生成地块（单线与双线），详见所提供的Liu and Long 2016 EPB参考资料（利用兴趣点和道路数据推导地块主导功能、功能密度和功能混合度）
 - 所提供的Parcels2011AICP为2011年利用路网生成的扣除了道路空间的地块/街区
 - ArcMap中的spatial join
- **街道：**
 - 路网数据经过必要的预处理（多线变单线、拓扑处理、细枝末节道路的删除等）
 - 所提供的Road_all_attributes经过了数据预处理
 - ArcMap中的spatial join

五、基于ArcGIS的数据可视化

基于ArcMap的数据可视化



- 显示与管理

- 加载、增加/删除字段、选择一部分对象

- 计算

- 计算字段（属性与空间）、关联join、空间关联spatial join

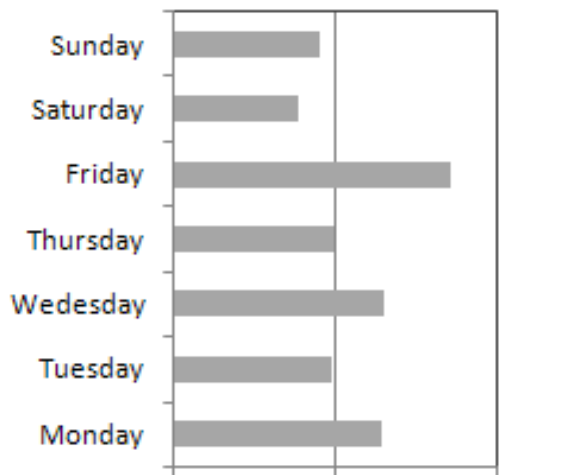
- 分析与统计

- 一般需要调用Toolbox
- 字段的summarize功能

- 可视化

- Symbology符号化（几种方法，features、categories、quantities）
 - 调整分级别显示时的上下顺序（高级设置）
- Definition Query（仅显示部分数据）
- Display（透明度设置）
- Layout视图（设置地图要素如指北针、比例尺、图例等！！！！）

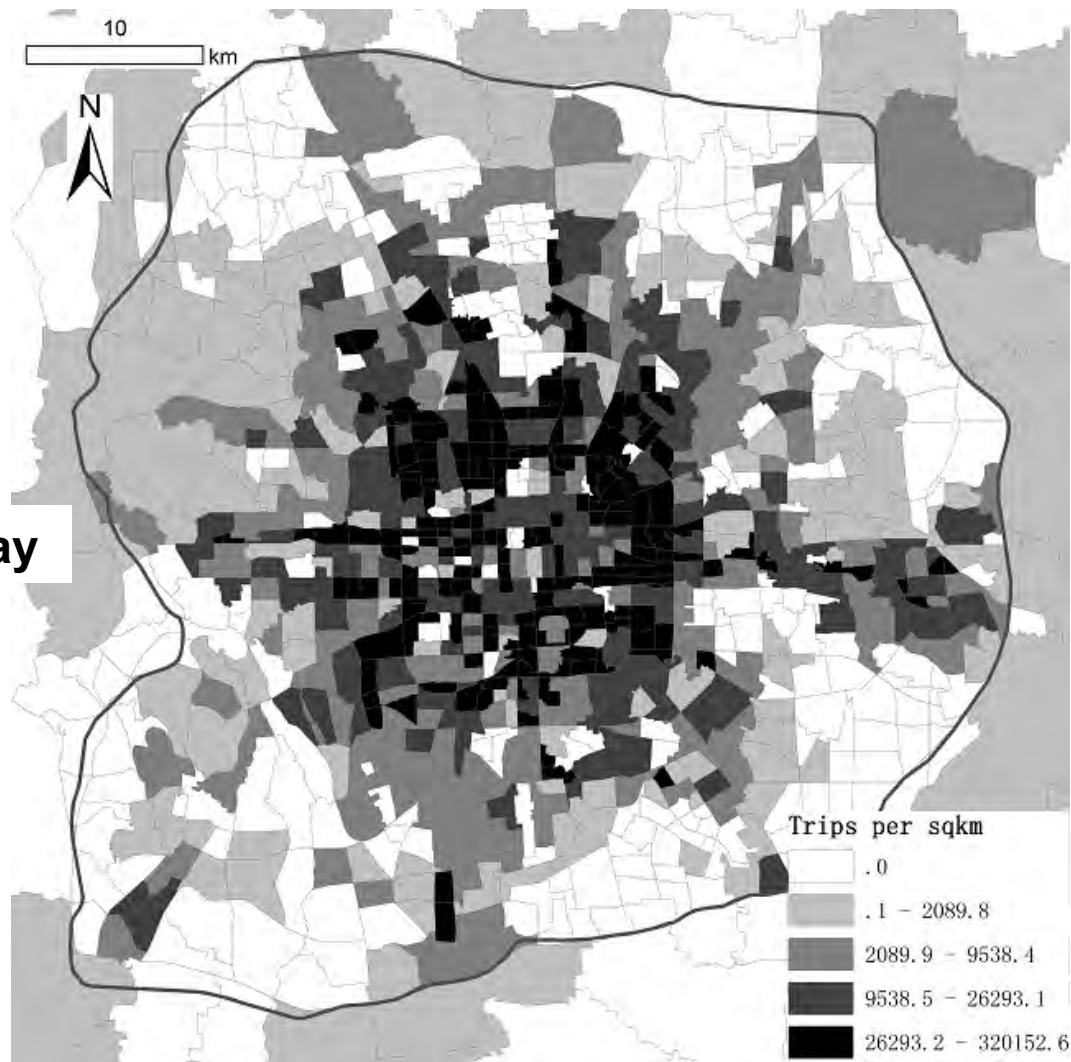
公共交通刷卡数据的总体情况可视化



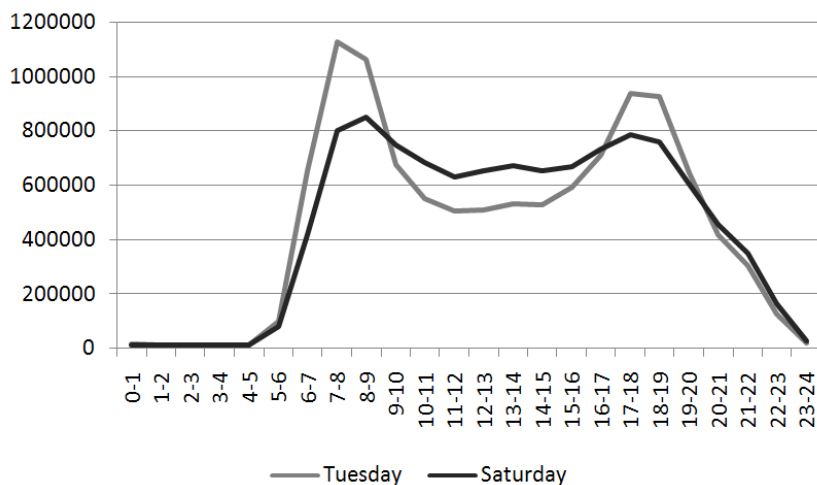
10000000 11000000 12000000

■ Trip count

Trip count (bus riding) for each day



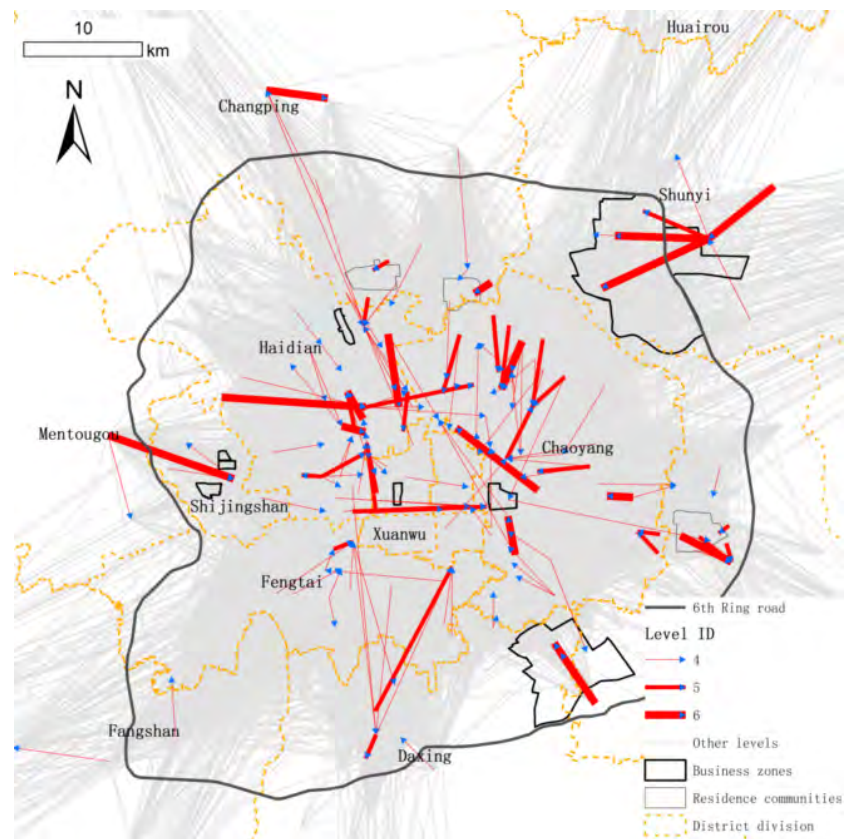
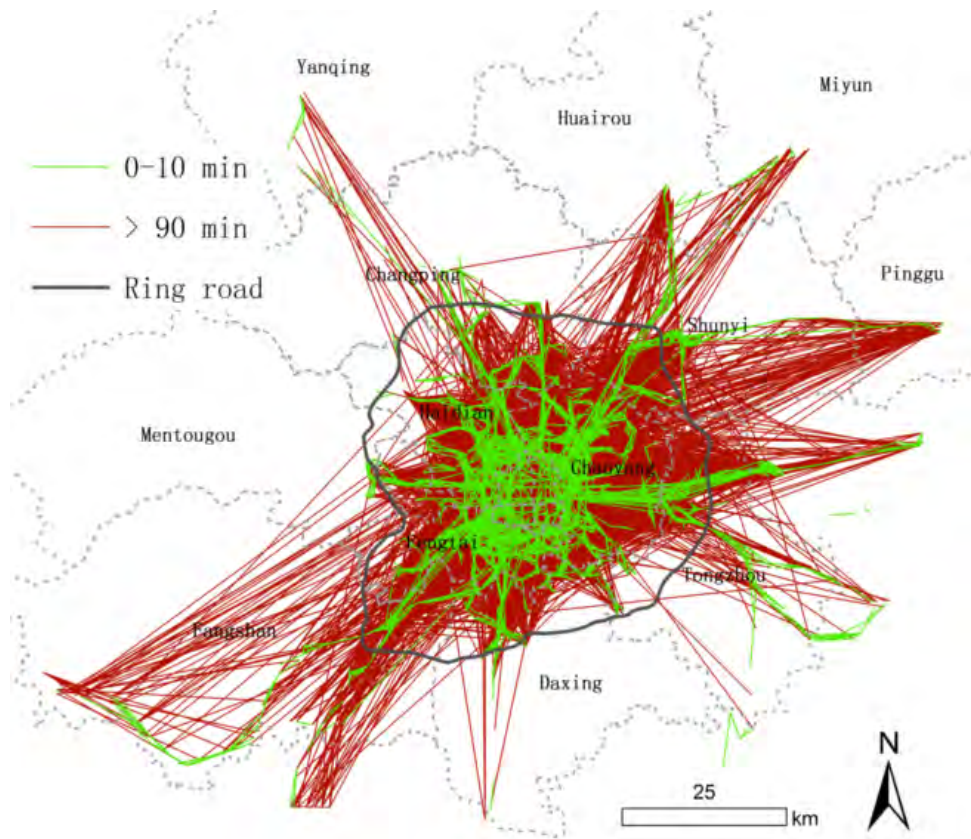
Trip density in the TAZ level in the whole week



— Tuesday — Saturday

Trip count in each hour on Tue and Sat

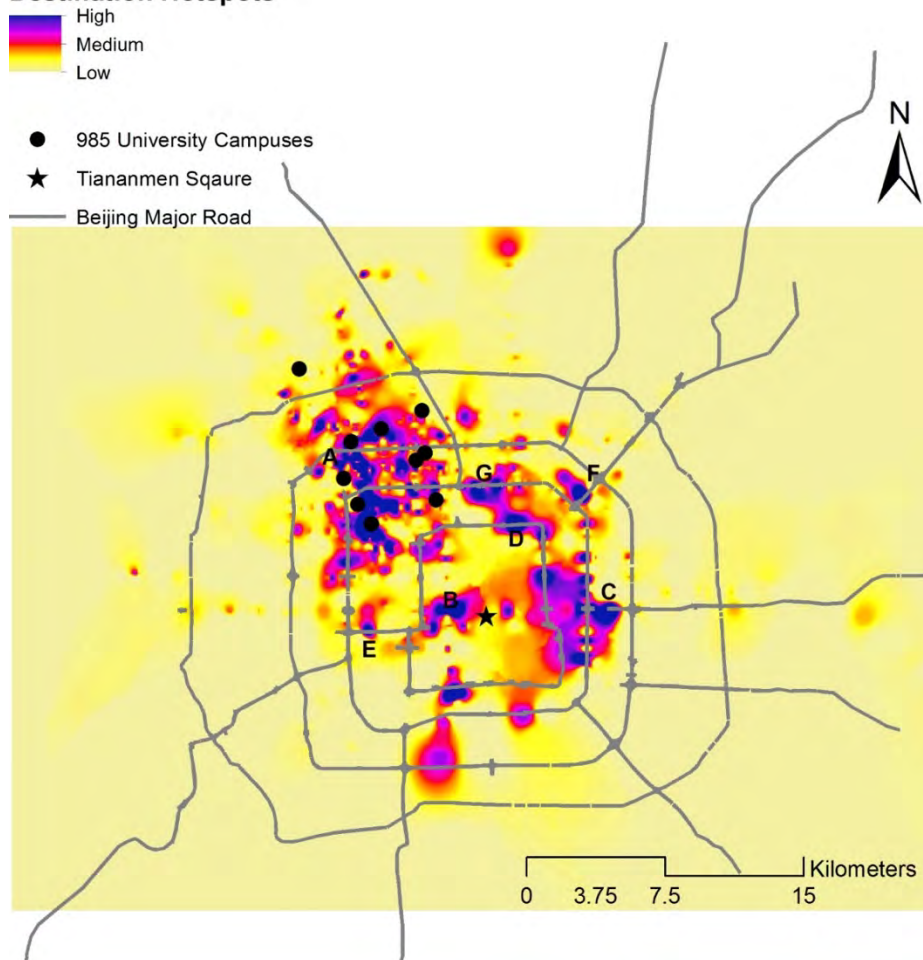
持卡人轨迹的可视化



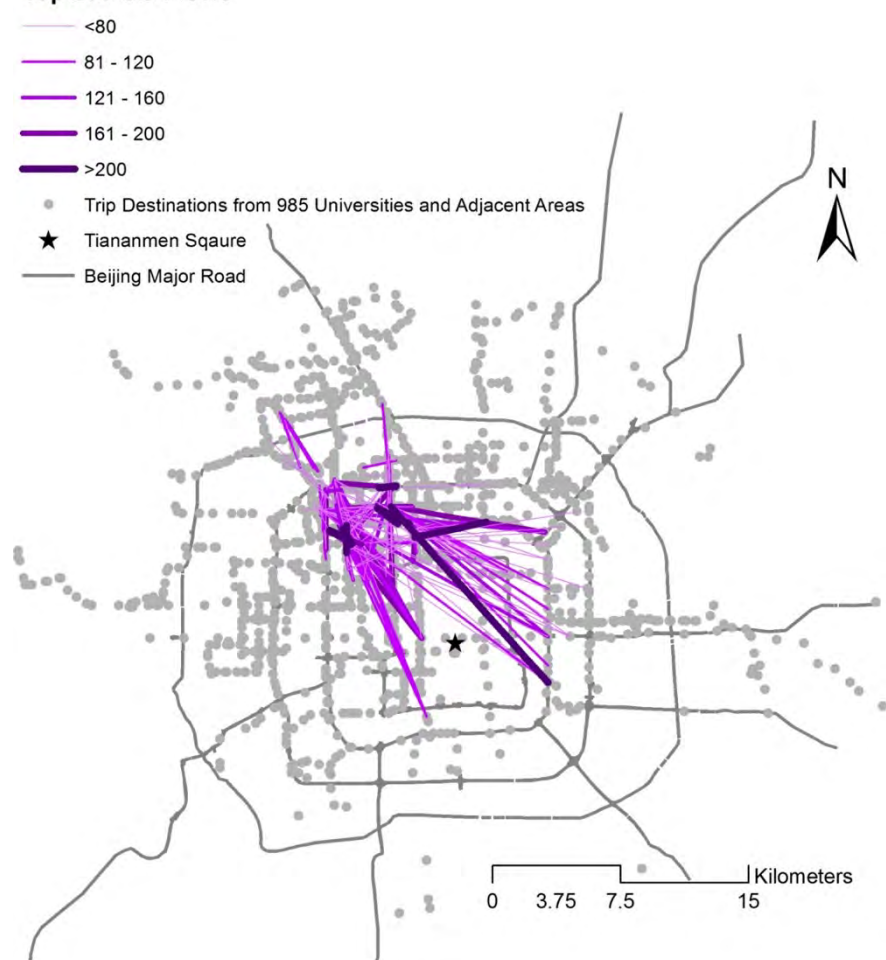
- 北京的极端通勤行为与主导通勤方向

学生卡出行的可视化

Destination Hotspots

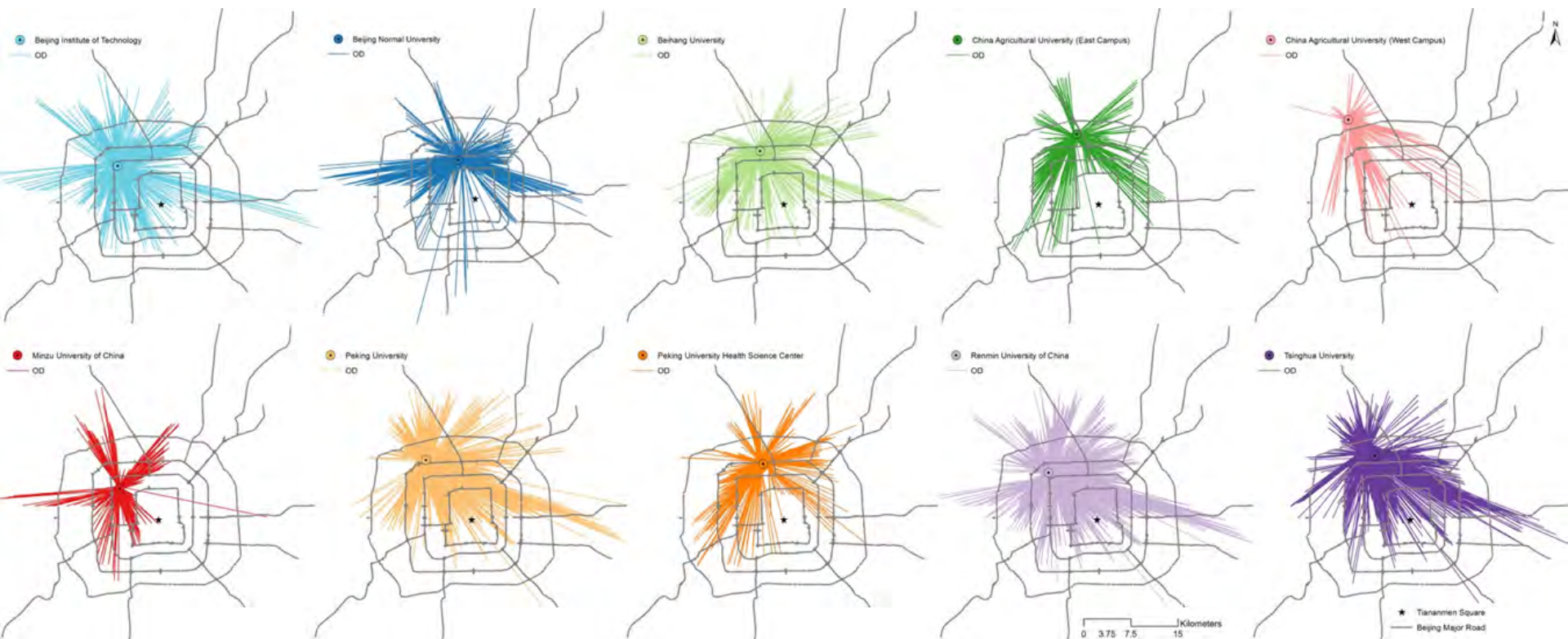


Top 200 OD Flows



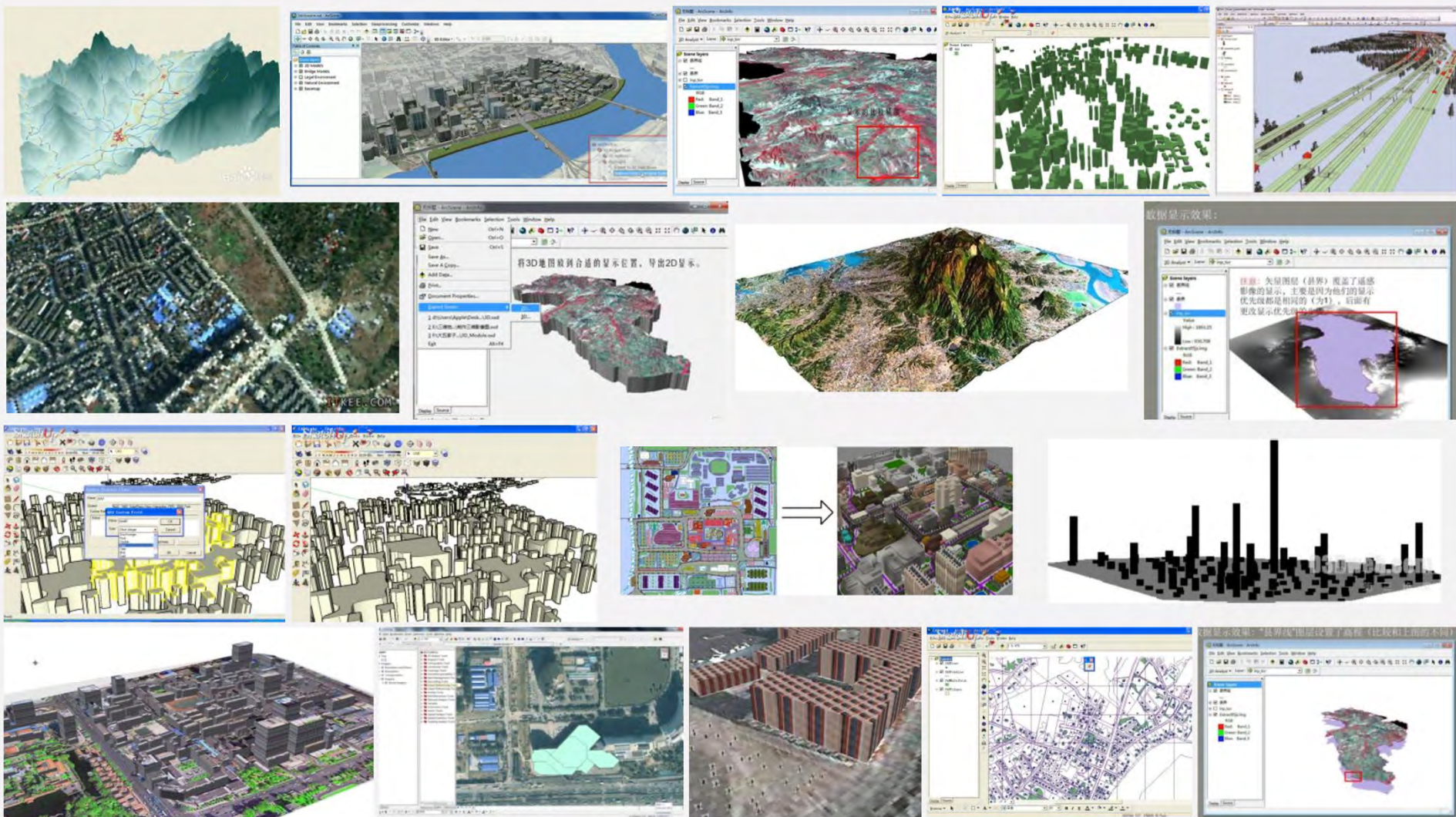
- 显示了大学校园和“最受欢迎的地方”的联系强度：大多数比较强的联系都在三环以内，同时也是北京高收入人群、高知名度团体和高收入公司的所在地。但令人惊讶的是各个大学之间的联系却不怎么强。

学生卡出行的可视化



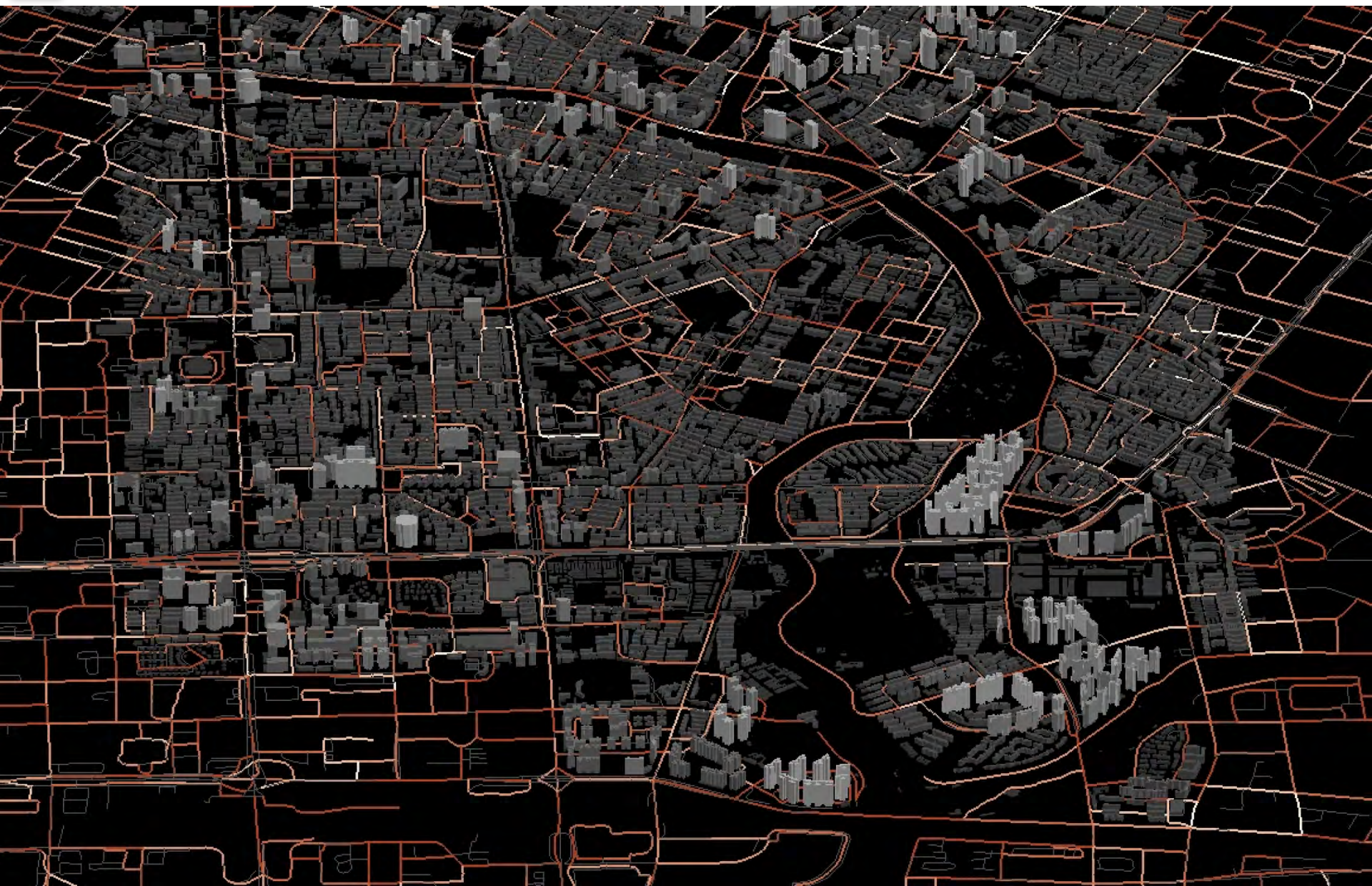
• 北京主要高校的大学生出行

基于ArcScene的数据可视化

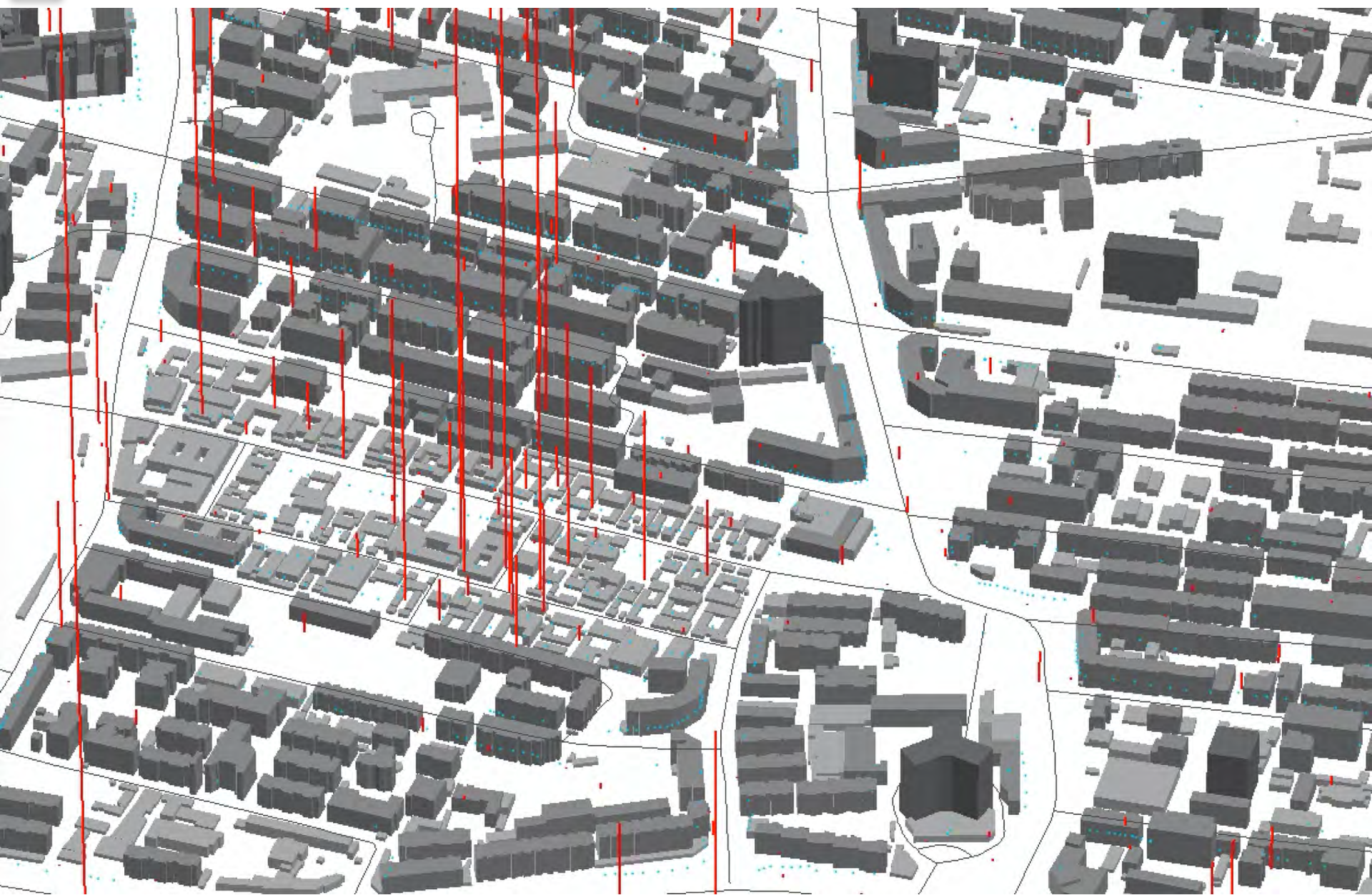


- 三维表达、拔升高度、基准高度（绝对值或某个图层如数字高程模型）
- <http://jingyan.baidu.com/article/4d58d5412969cb9dd4e9c017.html>

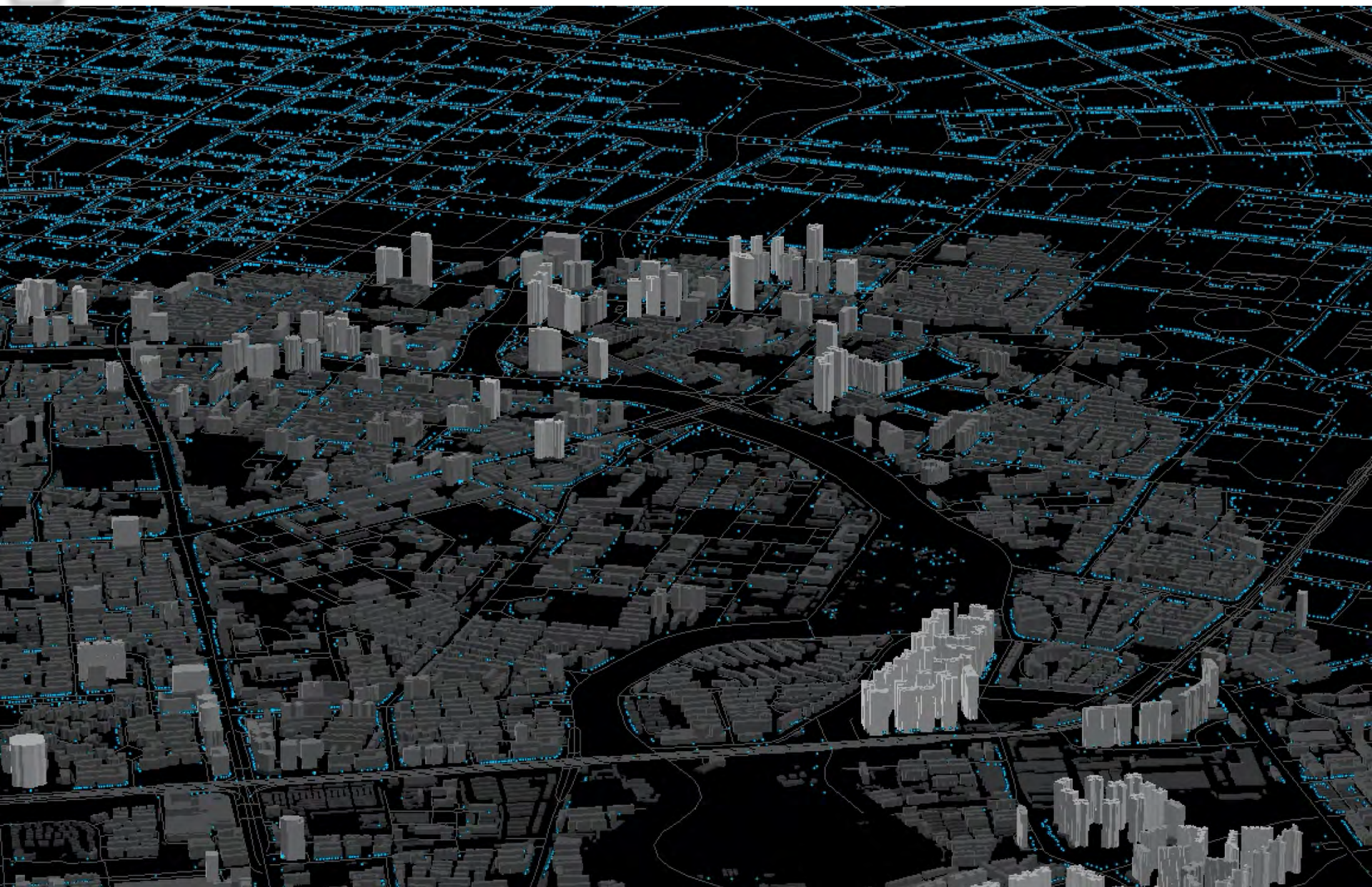
ArcScene可视化案例：街道活力



ArcScene可视化案例：社交媒体活力

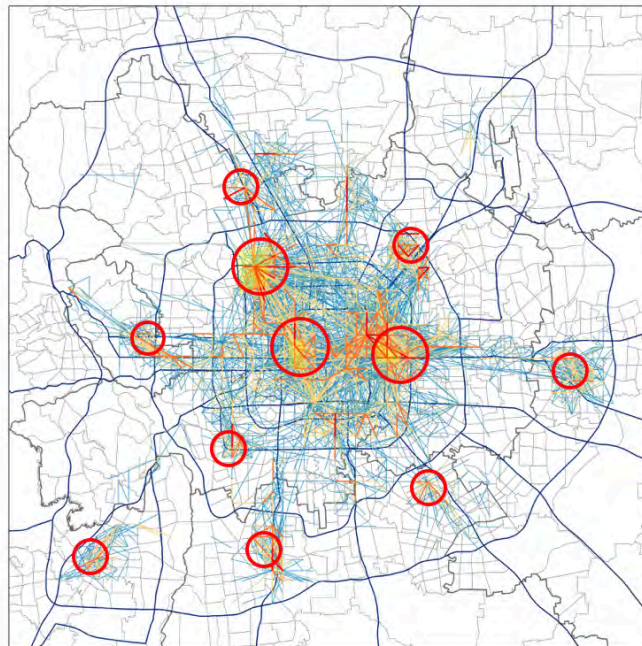
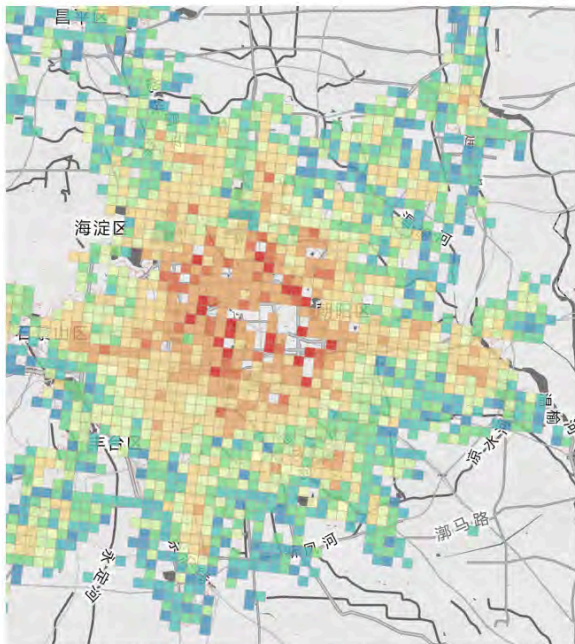


ArcScene可视化案例：城市功能



ArcGIS可视化的实际操作

从空间形式角度：基于位置的和基于联系的



- 基于位置（左）：体现了城市要素的位置或密度（空间分布）
- 基于联系（右）：体现了两两空间发生的联系（点、线和面）（空间联系）
- 两类数据对应了不同的分析、统计与可视化方法
- 两种类型共存的数据：位置微博记录，既有位置，又可以推测用户移动轨迹

清华大学研究生课程《大数据与城市规划》，龙 瀛，ylong@tsinghua.edu.cn 15

清华大学 BCL

• 密度数据的可视化和流数据的可视化

课后安排

- 阅读材料待放到课程网站
 - <https://www.beijingcitylab.com/courses/aium2018/>
- OPEN OFFICE HOUR
 - 每周二下午12:30-13:30
 - 需要提前通过info预约
 - ylong@tsinghua.edu.cn, 新建筑馆501, 13661386623
- 答疑邮箱
 - ylong@tsinghua.edu.cn



北京城市实验室
Beijing City Lab

<http://www.beijingcitylab.com>

