

# 城市大数据的分析与统计

Analysis and statistics for urban big data

龙瀛，博士  
清华大学建筑学院  
2017年10月27日

Approaching the Human City: Beijing Studio  
September 11 - 23, 2016  
COURSE SYLLABUS

*A joint workshop between the Human Cities Initiative at Stanford University  
& Tsinghua University Academy of Art and Design and the School of Engineering*

We invite you to participate in this experiment while we are holding class. Allow yourself to be fully present in the room, so you can listen to your classmates and what they have to say and share with you. Experience the freedom of not having to have your attention diverted or your mood instantly altered by whatever email or text message should come your way. You may take a phone call if it seems particularly urgent. But for most circumstances, we urge that you give yourself permission to be in control of your own time and energy— to actively choose where you want to direct your attention, as opposed to a portable device making that decision for you.

**Please turn your mobile devices to “silent” or “do not disturb” mode, and do not take them out for the duration of the class.** We strongly encourage you to take notes using pencil and paper— as research shows that this helps with memory retention— but If you must have a laptop to take notes, **please do not check your e-mail or browse the internet at any time.**

There will be plenty of opportunities to plug in once you leave the classroom. Let’s treat our classroom as a sacred space to enjoy the moment.

# 《大数据与城市规划》教学大纲

1. 大数据与城市规划概论（9月22日） W1
2. 变化中的中国城市与未来城市（9月29日） W2
3. 城市大数据类型与典型数据介绍（10月13日） W4（分组完成）
4. 城市大数据的获取（10月20日） W5（分发第一版本的北京老城数据、第一次课课外沙龙）
5. **城市大数据的统计与分析（10月27日） W6（介绍中期汇报要求）**
6. 城市大数据的可视化（11月3日） W7
7. 学生作业思路汇报与点评（11月10日） W8（预计占用第二大节部分时间）
  
8. 基于图片大数据的城市空间研究（11月17日） W9
9. 城市大数据挖掘：空间句法（11月24日） W10
10. 城市大数据挖掘：城市网络分析（12月1日） W11
11. 数据增强设计（12月8日） W12
12. 战略、总体及详细规划中的大数据应用（12月15日） W13
13. 城市设计中的大数据应用（12月22日） W14
14. 大模型：跨越城市内与城市间尺度的大数据应用（12月29日） W15
15. 学生作业终期汇报与点评（1月5日） W16（一次或分两次）

# 中期讨论

- 大作业的主题/题目11月8日前上报给助教
  - List: 人名/学号、作业题目、组长及其联系方式
- 欢迎提前与授课老师交流、讨论
- 11月10日 (W8) 与11月17日 (W9) 课程合并, 中期汇报 (每组10分钟汇报/5分钟讨论)
  - 8:00-9:35, 前半部分
  - 9:50-中午, 后半部分
  - 助教根据第二大节是否有课, 安排汇报顺序
- “基于图片大数据的城市空间研究” 内容额外由助教协助安排时间课外讲授
- 建议本节课后开始大作业的现场调研

第一大节 (s115, 8:00-9:35)				
组长		汇报时间	院系	中期汇报题目
曹哲静	BYJ	8:00-8:15	建筑学院	北京旧城地铁站最后一公里建成环境对步行行为的影响研究
沙焯星	SYX	8:15-8:30	建筑学院	基于房租数据模型的存量规划研究——以北京二环内为例
温祺	WQ	8:30-8:45	水利水电工程系	基于城市设计尺度的步行体验研究
许宁婧	XNJ	8:45-9:00	建筑学院	公共空间品质对个体情绪的影响: 基于位置微博和现场照片的大规模模测
徐特	XT	9:00-9:15	环境学院	基于街景数据的交叉口活力指标构建
周旭影	YXY	9:15-9:30	中规院	旧城声音景观
第二大节 (中203, 9:50-12:00)				
组长		汇报时间	院系	中期汇报题目
封蓉	FR	9:50-10:05	建筑学院	北京二环内城市气味
裴昱	PY	10:05-10:20	建筑学院	基于空间数据的北京二环内夜间城市活力测度相关性分析
秦岭	QL	10:20-10:35	建筑学院	基于大数据的“大院”街区开放性研究
万耀璿	WYL	10:35-10:50	土木工程系	基于空间视角的北京二环内影响街道安全性因素与评价
谢湘雅	XXY	10:50-11:05	建筑学院	北京旧城街道功能变化
许阳	XY	11:05-11:20	中规院	二环内新建轨道交通站点对邻近用地状况的影响研究
詹旭强	YXQ	11:20-11:35	建筑学院	空气质量对旅游景点活力的影响调研——以北京二环内为例
周宜笑	YYX	11:35-11:50	建筑学院	城市物质空间与社会空间国际化匹配程度探讨——以北京旧城区为例

## 上一堂课的回顾

- 结构化网页的抓取
- 利用API抓取开放数据
- 共享的第一个版本的北京旧城数据
  
- 第一次课外沙龙（李栋博士）

# 运用互联网数据进行城市规划与研究所使用的主要工具及流程

➤ STEP1：定位查找数据源的网络地址



Chrome (谷歌浏览器)

➤ STEP2：将获取的网络开放数据保存在本地



火车采集器 (V8)

➤ STEP3：数据的清洗、预处理



Microsoft Excel 2007

➤ STEP4：坐标系统转换 (地理坐标→平面坐标)



万能坐标转换器

➤ STEP5：空间分析、空间统计及可视化表达



IBM SPSS 19.0



ArcGIS 10.x软件包

# 互联网眼中的“城市”

**大众点评** 找商户 享优惠 写点评

搜索:

按拼音查找 国内城市 | 国外城市

<b>A</b>	阿巴嘎旗   阿坝   阿坝县   阿比斯库   阿伯特   阿布扎比   阿德莱德   阿尔巴尼亚   更多
<b>B</b>	八宿县   巴巴多斯   巴楚县   巴淡岛   巴德岗   巴德依舍   巴登-巴登   巴东县   巴尔的摩   更多
<b>C</b>	伯克利   沧县   沧源傣族自治县   沧州   苍南   苍梧县   苍溪县   曹县   册亨县   策勒县   更多
<b>D</b>	达尔罕茂明安联合旗   达拉斯   达拉特旗   达勒姆   达日县   达州   达孜县   大安市   大阪   更多
<b>E</b>	埃德蒙顿   埃及   埃塞俄比亚   埃森   埃因霍温   爱丁堡   爱沙尼亚   赤道几内亚   东帝汶   更多
<b>F</b>	法国   法库县   法兰克福   法罗群岛   法属波利尼西亚   法属圭亚那   繁昌县   繁峙县   范县   更多
<b>G</b>	大堡礁   大提顿国家公园   大峡谷国家公园   大洋路   德国   噶尔县   改则县   盖州市   更多
<b>H</b>	北海道   兵庫   飞弹高山   广岛   哈巴河县   哈尔滨   哈尔施塔特   哈密地区   哈密市   更多
<b>I</b>	爱达荷州   爱尔兰   爱荷华州   冰岛   茨城   怀特岛   谦卑市   石垣岛   天空之城   更多
<b>J</b>	抚州   伽师县   鸡东县   鸡西   鸡泽县   积石山保安族东乡族撒拉族自治县   基督城   基拉尼   更多
<b>K</b>	仓敷市   驯鹿町   高雄   高知   古晋   哈萨克斯坦   韩国   基拉尼   基里巴斯   基隆   更多
<b>L</b>	巴宣区   长岛   抚顺县   吉县   拉萨   拉斯佩齐亚   拉斯维加斯   拉脱维亚   拉孜县   来安县   更多
<b>M</b>	澳门   泌阳县   宫古岛   宫崎   麻城市   麻江县   麻栗坡县   麻阳苗族自治县   马鞍山市   更多
<b>N</b>	北卡罗来纳州   北马里亚纳   长崎   长野   朝鲜   登别   荷兰   名古屋   拿骚   内华达州   更多
<b>O</b>	阿曼   安大略省   奥克兰   奥兰多   奥林匹亚   奥马鲁   奥斯特   冲绳   大阪   大分   更多
<b>P</b>	巴布亚新几内亚   巴基斯坦   巴拉圭   巴拉望岛   巴勒斯坦   巴黎   巴拿马   巴色   芭堤雅   更多
<b>Q</b>	皇后镇   卡塔尔   魁北克   昆士兰   七台河   栖霞市   祁东县   祁连县   祁门县   祁县   更多
<b>R</b>	俄罗斯   刚果共和国   拉贾斯坦邦   兰斯   雷恩   雷根斯堡   雷克雅未克   雷诺   里加   更多
<b>S</b>	埃文河畔斯特拉福德   白浜町   硅谷   集贤县   静冈   旧金山   柯桥区   濠州内海   海池县   更多
<b>T</b>	碧江区   大溪地   德黑兰   德克萨斯州   登嘉楼   地拉那   蒂马鲁   东京   洞爷湖町   都灵   更多
<b>U</b>	美国   美属维尔京群岛   蔚山市   翁布里亚大区   乌代布尔   乌得勒支   乌尔姆   乌干达   更多
<b>V</b>	巴拉德罗   梵蒂冈   弗吉尼亚州   弗拉基米尔   福伦丹   海参崴   瓦杜兹   瓦拉纳西   瓦莱塔   更多
<b>W</b>	北莱茵-威斯特法伦   富宁县   渭源县   和歌山   华沙   华盛顿哥伦比亚特区   华盛顿州   更多
<b>X</b>	泸溪县   莘县   威县   浚县   西安   西昌市   西充县   西畴县   西递   西丰县   西哈努克   更多
<b>Y</b>	横滨   黄石国家公园   临猗县   龙仁   芦山县   日惹   沙巴   山梨   四日市市   乐清   更多
<b>Z</b>	采尔马特   津巴布韦   萨格勒布   萨拉戈萨   苏黎世   绥宁县   台州市   播州区   枞阳县   更多

如果没有您要找的城市, 请点击 [联系我们](#)

**地标** TOP

**渝中区** 日光广场 | 大都会广场 | 解放碑纪念碑 | 罗汉寺 | 纽约大厦 | 王府井百货解放碑店 | 国贸豪生大酒店 | 金鹰财富中心 | 时代广场 | 创汇首座 | 解放碑英利国际 | 得意世界 | 重庆大礼堂 | 三峡博物馆 | 鹅岭公园 | 人民广场 | 希尔顿酒店 | 朝天门广场 | 龙湖时代天街 | 大坪英利国际 | 重庆天地

**沙坪坝区** 炫地购物广场 | 王府井 | 重庆大学 | 耀华新纪元购物广场 | 重庆师范大学 | 凯德 | ARC广场 | 凤天路美食街 | 磁器口古镇 | 西南政法大学沙坪坝校区

**江北区** 协信星光 | 金源时代购物广场 | 茂业百货 | 北城天街广场 | 未来国际大厦 | 大融城购物中心 | 朗晴广场 | 新世界百货 | 星天广场 | 金源不夜城 | 洋河体育场

**渝北区** 水晶郛城 | 财富中心 | 鲁能星城 | 西南政法大学 | 金岛花园 | 康庄美地 | 龙湖西苑 | 西南政法大学渝北校区

**南岸区** 国际社区 | 亚木山谷 | 南坪步行街 | 协信星光时代广场 | 五小区 | 万达广场 | 上海城 | 融侨半岛 | 盛汇广场 | 东原1891

**九龙坡区** 西城天街 | 奥体中心 | 佰腾数码广场 | 申泰赛菲特大酒店

**大渡口区** 大渡口公园正门

**北碚区** 天奇广场 | 嘉陵风情步行街

**江津区** 重百江津商场

**潼南区** 潼南县人民医院 | 外滩国际城 | 潼南油菜花景区 | 巴渝大道 | 站前大道 | 华堂购物中心 | 潼南江南医院 | 潼柏路 | 大佛寺景区

**地铁沿线** TOP

**1号线** 较场口 | 七星岗 | 两路口 | 鹅岭 | 大坪 | 石油路 | 杨公桥 | 歇台子 | 烈士墓 | 磁器口 | 石桥铺 | 石井坡 | 双碑 | 高庙村 | 赖家桥 | 微电园 | 马家岩 | 陈家桥 | 大学城 | 沙坪坝 | 尖顶坡 | 小龙坎 | 小什字

**2号线** 天堂堡 | 建桥 | 金家湾 | 刘家坝 | 白居寺 | 大江 | 鱼洞 | 新山村 | 大渡口 | 平安 | 马王场 | 大堰村 | 动物园 | 杨家坪 | 谢家湾 | 袁家岗 | 大坪 | 佛图关 | 李子坝 | 牛角沱 | 曾家岩 | 大溪沟 | 黄花园 | 临江门 | 较场口

**3号线** 六公里 | 五公里 | 四公里 | 南坪 | 两路口 | 重庆北站 | 牛角沱 | 华新街 | 郑家院子 | 观音桥 | 红旗河沟 | 嘉州路 | 唐家院子 | 铜元局 | 工贸 | 狮子坪 | 龙头寺 | 八公里 | 童家院子 | 麒龙 | 九公里 | 金渝 | 岔路口 | 花溪 | 金童路 | 大山村 | 学堂湾 | 鸳鸯 | 鱼胡路 | 金竹 | 园博园 | 鱼洞 | 翠云 | 长福路 | 回兴 | 双龙 | 碧津 | 江北机场 | 二塘

**6号线** 北碚 | 天生 | 状元碑 | 龙凤溪 | 蔡家 | 向家岗 | 金山寺 | 礼嘉 | 曹家湾 | 康庄 | 大竹林 | 光电园 | 九曲河 | 冉家坝 | 大龙山 | 花卉园 | 黄泥磅 | 红土地 | 五里店 | 红旗河沟 | 大剧院 | 小什字 | 上新街 | 刘家坪 | 长生桥 | 邱家湾 | 茶园

# 互联网眼中的“城市”

## 商区 ▲ TOP

- 渝中区 解放碑 | 两路口 | 朝天门 | 大坪 | 上清寺 | 七星岗 | 洪崖洞 | 观音岩 | 较场口 | 化龙桥 | 菜园坝 | 临江门
- 沙坪坝区 沙坪坝 | 小龙坎 | 天星桥 | 磁器口 | 歌乐山 | 大学城 | 三峡广场 | 龙湖U城天街 | 美丽熙街 | 陈家桥 | 汉渝路 | 沙坪坝区政府 | 杨公桥 | 欣阳广场 | 烈士墓 | 双碑 | 莲花湖风景区
- 江北区 南桥寺 | 北滨二路 | 观音桥 | 洋河花园 | 大石坝 | 五里店 | 北城天街 | 北滨路 | 九街 | 龙湖星悦荟
- 渝北区 龙溪 | 松树桥 | 新牌坊 | 黄泥磅 | 两路 | 冉家坝 | 北部新区 | 人和 | 龙头寺 | 回兴 | 空港新城 | 加州 | 花卉园 | 汽博中心 | 大竹林 | 南桥寺
- 南岸区 南坪 | 南滨路 | 四公里 | 弹子石 | **南山** | 福利社 | 上新街 | 洋人街 | 茶园新区 | 回龙湾
- 九龙坡区 杨家坪 | 石桥铺 | 开发区 | 黄桷坪 | 九滨路 | 袁家岗 | 巴国城 | 谢家湾 | 动物园 | 陈家坪 | 二郎 | 西彭 | 白市驿
- 大渡口区 大渡口公园 | 新天泽国际广场
- 巴南区 龙洲湾 | 鱼洞 | 李家沱
- 北碚区 龙凤桥 | 施家梁 | 三溪口 | 水土 | 澄江 | 北碚老城 | 北碚城南 | 北碚城北 | 北碚正码头 | 天生丽街 | 文星湾 | 北温泉 | 缙云山 | 歇马 | 蔡家
- 江津区 德感 | 双福 | 鼎山 | 大同路步行街
- 黔江区
- 綦江区 大学城
- 大足区
- 潼南区 潼南县其他 | 桂林街道 | 梓潼街道 | 上和镇 | 龙形镇 | 古溪镇 | 宝龙镇 | 玉溪镇 | 米心镇 | 群力镇 | 双江镇 | 花岩镇 | 柏梓镇 | 崇龛镇 | 塘坝镇 | 新胜镇 | 太安镇 | 小渡镇 | 卧佛镇 | 五桂镇 | 田家镇 | 别口镇 | 寿桥镇
- 铜梁区 铜梁县其他 | 巴川街道 | 东城街道 | 南城街道 | 土桥镇 | 二坪镇 | 水口镇 | 安居镇 | 白羊镇 | 平滩镇 | 石鱼镇 | 福果镇 | 维新镇 | 高楼镇 | 大庙镇 | 围龙镇 | 华兴镇 | 永嘉镇 | 安溪镇 | 安溪镇 | 太平镇 | 旧县镇 | 虎峰镇 | 少云镇 | 蒲吕镇 | 侣俸镇 | 小林镇 | 双山镇 | 庆隆镇
- 荣昌区 荣昌县其他 | 昌元街道 | 昌洲街道 | 广顺街道 | 双河街道 | 安富街道 | 峰高街道 | 直升镇 | 路孔镇 | 清江镇 | 仁义镇 | 河包镇 | 古昌镇 | 吴家镇 | 观胜镇 | 铜鼓镇 | 清流镇 | 盘龙镇 | 远觉镇 | 清升镇 | 荣隆镇 | 龙集镇
- 璧山区 璧山县其他 | 璧城街道 | 璧泉街道 | 青杠街道 | 来凤街道 | 丁家街道 | 大路街道 | 八塘镇 | 七塘镇 | 河边镇 | 福祿镇 | 大兴镇 | 正兴镇 | 广普镇 | 三合镇 | 健龙镇
- 垫江县 垫江县其他 | 桂溪镇 | 新民镇 | 沙坪镇 | 周嘉镇 | 普顺镇 | 永安镇 | 高安镇 | 高峰镇 | 五河镇 | 澄溪镇 | 太平镇 | 鹤游镇 | 坪山镇 | 砚台镇 | 曹回镇 | 杠家镇 | 包家镇 | 白家镇 | 永平镇 | 三溪镇 | 裴兴镇 | 长龙乡 | 沙河乡 | 大石乡 | 黄沙乡

## 分类 ▲ TOP

- 榛果民宿
- 美食 私房菜 | 家常菜 | 农家菜 | 创意菜 | 东北菜 | 川菜 | 粤菜 | 面馆 | 小龙虾 | 茶馆 | 火锅 | 西北菜 | 烧烤 | 海鲜 | 小吃快餐 | 日本料理 | 湘菜 | 韩国料理 | 西餐 | 自助餐 | 咖啡厅 | 东南亚菜 | 面包甜点 | 其他
- 购物 水果生鲜 | 免税店 | 综合商场 | 食品茶酒 | 化妆品 | 服饰鞋包 | 珠宝首饰 | 花店 | 运动户外 | 药店 | 办公/文化用品 | 亲子购物 | 数码产品 | 书店 | 家居建材 | 眼镜店 | 超市/便利店 | 更多购物场所
- 周边游 景点 | 旅游其他 | 温泉 | 滑雪 | 采摘/农家乐 | 展馆展览 | 动植物园 | 水上娱乐
- 休闲娱乐 VR | 咖啡厅 | 酒吧 | KTV | 茶馆 | 采摘/农家乐 | 棋牌室 | 私人影院 | 文化艺术 | 真人CS | 轰趴馆 | 运动健身 | 足疗按摩 | 洗浴/汗蒸 | DIY手工坊 | 网吧网咖 | 游乐游艺 | 密室 | 台球馆 | 桌面游戏 | 更多休闲娱乐
- 丽人 祛痘 | 美甲美睫 | 美发 | 化妆品 | 美容/SPA | 瘦身纤体 | 整形 | 纹绣 | 瑜伽 | 舞蹈 | 纹身 | 产后塑形
- 结婚 婚房装修 | 旅拍 | 婚纱摄影 | 婚宴 | 婚戒首饰 | 婚纱礼服 | 婚庆公司 | 彩妆造型 | 婚礼跟拍 | 个性写真 | 司仪主持 | 更多婚礼服务 | 西服定制 | 婚车租赁 | 婚礼小商品
- 亲子 月子会所 | 托班/托儿所 | 宝宝派对 | 亲子玩乐 | 亲子旅游 | 婴儿游泳 | 幼儿教育 | 幼儿园 | 亲子摄影 | 亲子游乐 | 亲子购物 | 孕妇写真 | 孕产护理 | 早教中心 | 幼儿外语 | 更多亲子服务 | 幼儿才艺
- 学习培训 外语培训 | 音乐培训 | 职业技术 | 升学辅导 | 美术培训 | 驾校 | 留学 | 兴趣生活 | 教育院校 | 更多教育培训
- 家装 装修设计 | 建材 | 家具家居 | 家装卖场 | 家用电器
- 运动健身 健身中心 | 壁球馆 | 游泳馆 | 攀岩馆 | 羽毛球馆 | 射箭馆 | 马术场 | 溜冰场 | 瑜伽 | 舞蹈 | 漂流 | 武术场馆 | 篮球场 | 网球场 | 足球场 | 高尔夫场 | 滑雪 | 保龄球馆 | 台球馆 | 乒乓球馆 | 体育场馆 | 更多运动场馆
- 爱车 年检站 | 交警队 | 4S店/汽车销售 | 洗车 | 维修保养 | 配件/车饰 | 汽车租赁 | 停车场 | 加油站 | 汽车美容 | 汽车保险 | 更多汽车服务
- 酒店 高端酒店 | 快捷酒店 | 商务酒店 | 主题酒店 | 情侣酒店 | 民宿 | 度假酒店 | 别墅 | 五星级/豪华型 | 四星级/高档型 | 三星级/舒适型 | 经济型 | 公寓式酒店 | 精品酒店 | 度假村 | 更多酒店住宿 | 客栈旅舍 | 青年旅舍
- 生活服务 洗涤护理 | 商务服务 | 文化传媒 | 便民服务 | 家电维修 | 搬家运输 | 丧葬 | 金融 | 旅行社 | 快照摄影 | 文印图文 | 生活配送 | 回收 | 演出票务 | 政府机构 | 居家维修 | 网站 | 商圈 | 交通 | 房屋地产 | 公司企业 | 家政 | 售票点 | 更多生活服务
- 医疗健康 医院 | 中医 | 齿科 | 药店 | 体检中心 | 整形 | 妇幼医院 | 宠物医院 | 其他医疗
- 宠物 宠物店 | 宠物医院
- 电影演出赛事 剧场/剧院 | 音乐厅/礼堂 | 艺术中心/文化广场 | 热门演出 | 赛事展览 | 其他电影演出赛事 | 电影院 | 演出场馆



# 本讲大纲

1. 数据分析与统计概述
2. 基于ArcGIS的数据分析
3. 基于SPSS的数据统计
4. 案例介绍：基于街景图片的街道绿化研究

# 一、数据分析与统计概述

本节部分幻灯片来自西安建筑科技大学的郑晓伟老师，在此表示感谢

# 为什么要做数据分析与统计

- 数据→信息→知识→智慧

# 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

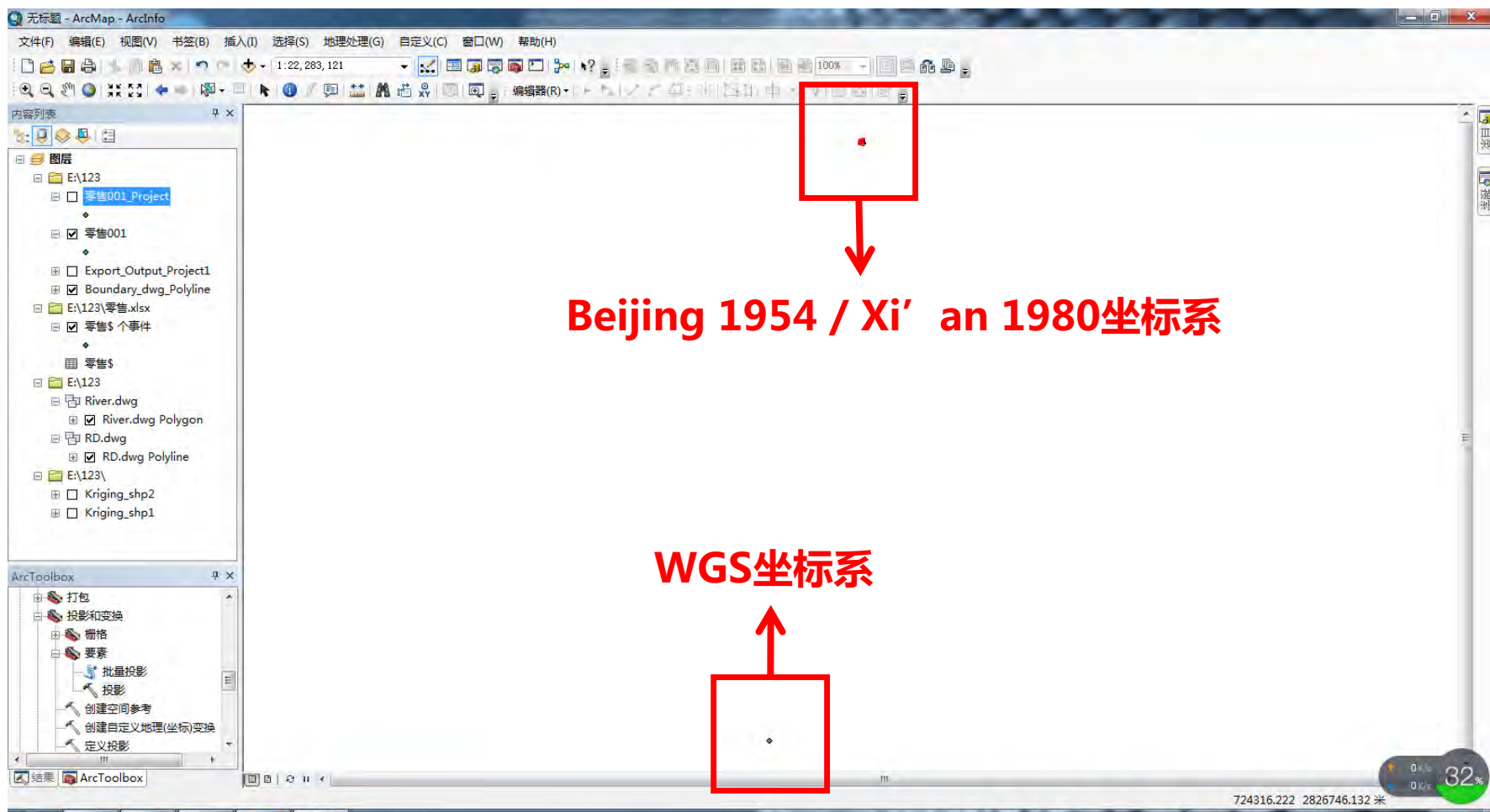
- Excel、Access、ArcMap、SPSS无论哪个软件，都需要进行数据预处理

- 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）

- 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



# 位置纠偏



- 腾讯，谷歌，高德是火星坐标系、百度是百度坐标
- 坐标系统的不同，会导致空间位置无法完全匹配，需要纠正到地球坐标
  - 保证一个项目的所有图层的坐标处于同一坐标体系（地球与火星）

# 位置纠偏：未来交通实验室的工具



	A	B	C
1	1	108.9476	34.24587
2	2	108.9476	34.25782
3	3	108.9485	34.24925
4	4	108.938	34.25243
5	5	108.934	34.25051
6	6	108.9232	34.23627
7	7	108.9473	34.25789
8	8	108.9486	34.25013
9	9	108.9551	34.25435
10	10	108.9489	34.25114
11	11	108.9392	34.2558

```
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
1,108.942977537453, 34.2474656784029
2,108.942982355667, 34.2594063777044
3,108.943816694545, 34.2508402191483
4,108.933362489567, 34.2540299722615
5,108.929354541146, 34.252120020051
6,108.918577355224, 34.2378962093459
7,108.942653270121, 34.2594769410353
8,108.943971782149, 34.2517137948719
9,108.950401159916, 34.2559277029528
10,108.944228366813, 34.2527287260284
11,108.934582343306, 34.2573997981637
```

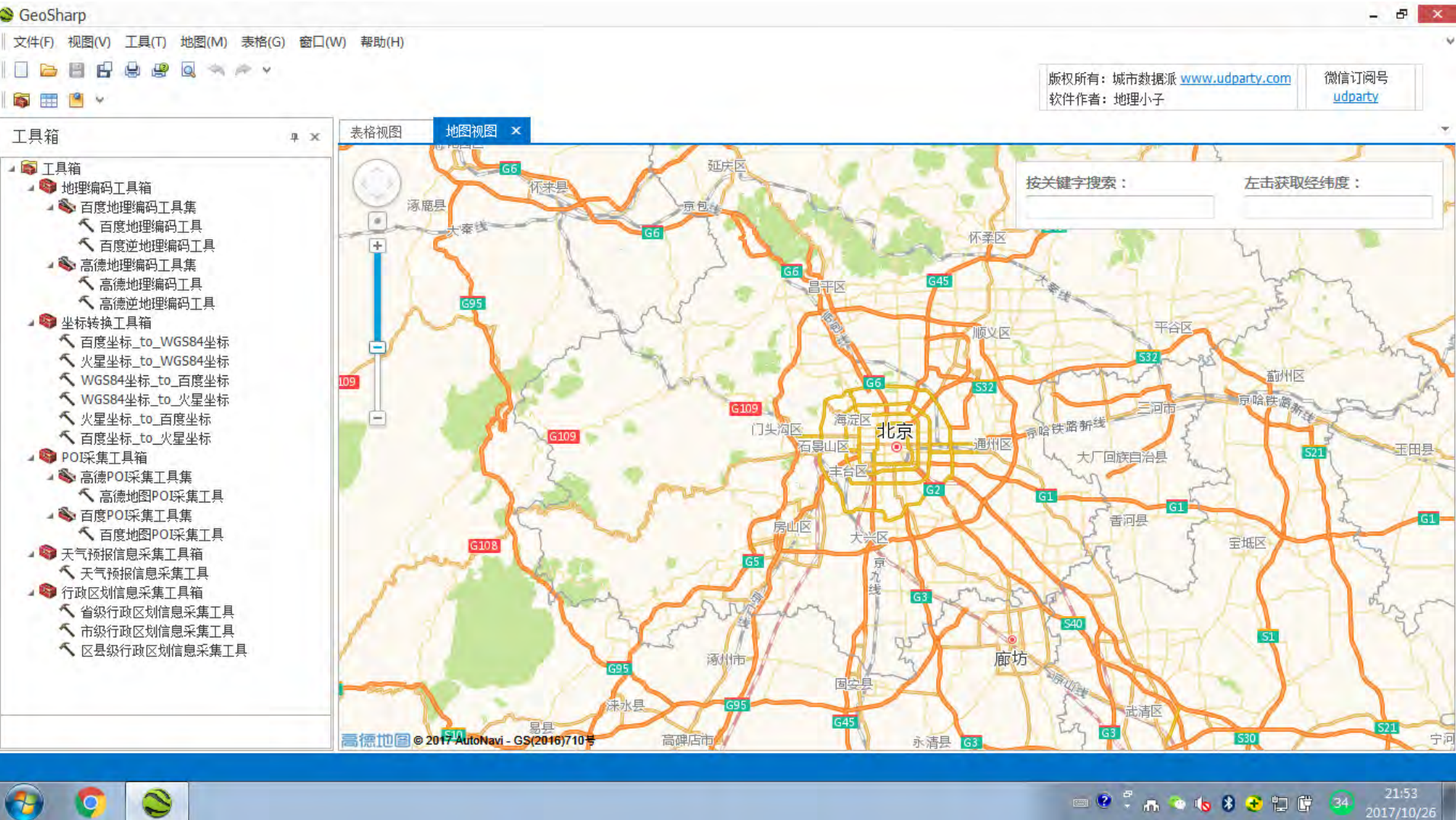
- 推荐利用未来交通实验室的万能坐标转换工具进行纠偏
  - 百度等也提供了纠偏/加偏的API
- Win7以上系统以管理员模式打开
- .net Framework版本需在4.0以上

# 数据空间化 (GeoCoding)



• 地理编码GeoCoding = 由地址到经纬度

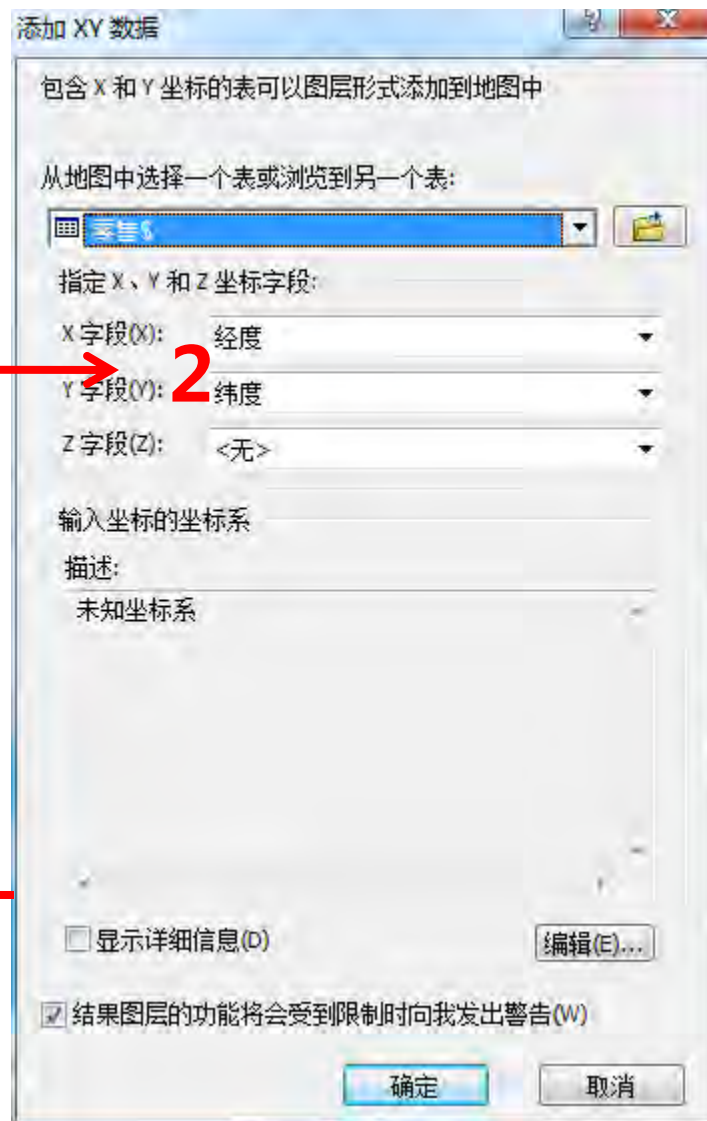
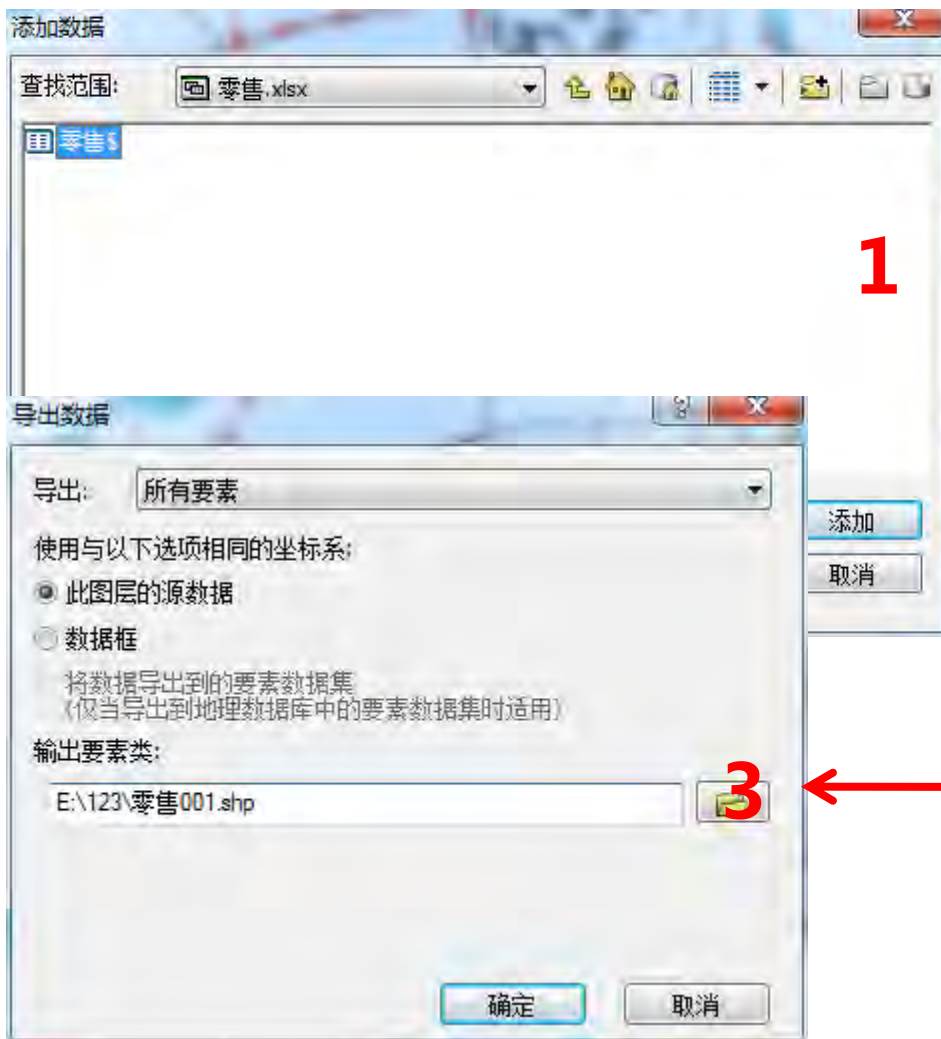
# 数据空间化 (GeoCoding)



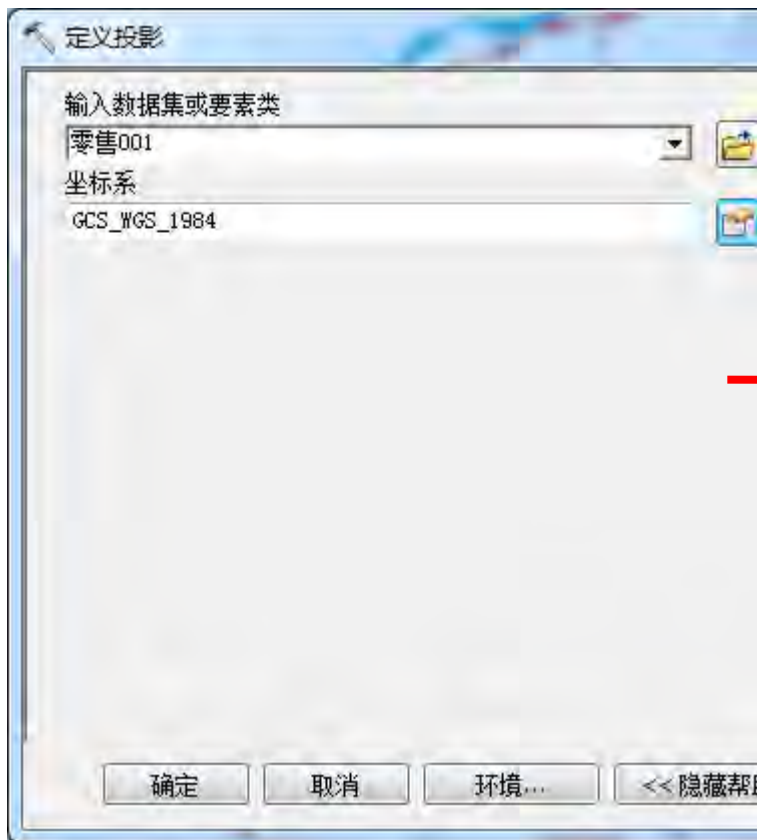
- GeoSharp可以实现位置纠偏（加偏），也可以实现空间化

# 数据空间化 (GeoCoding)

## ➤ 添加XY数据——导出数据

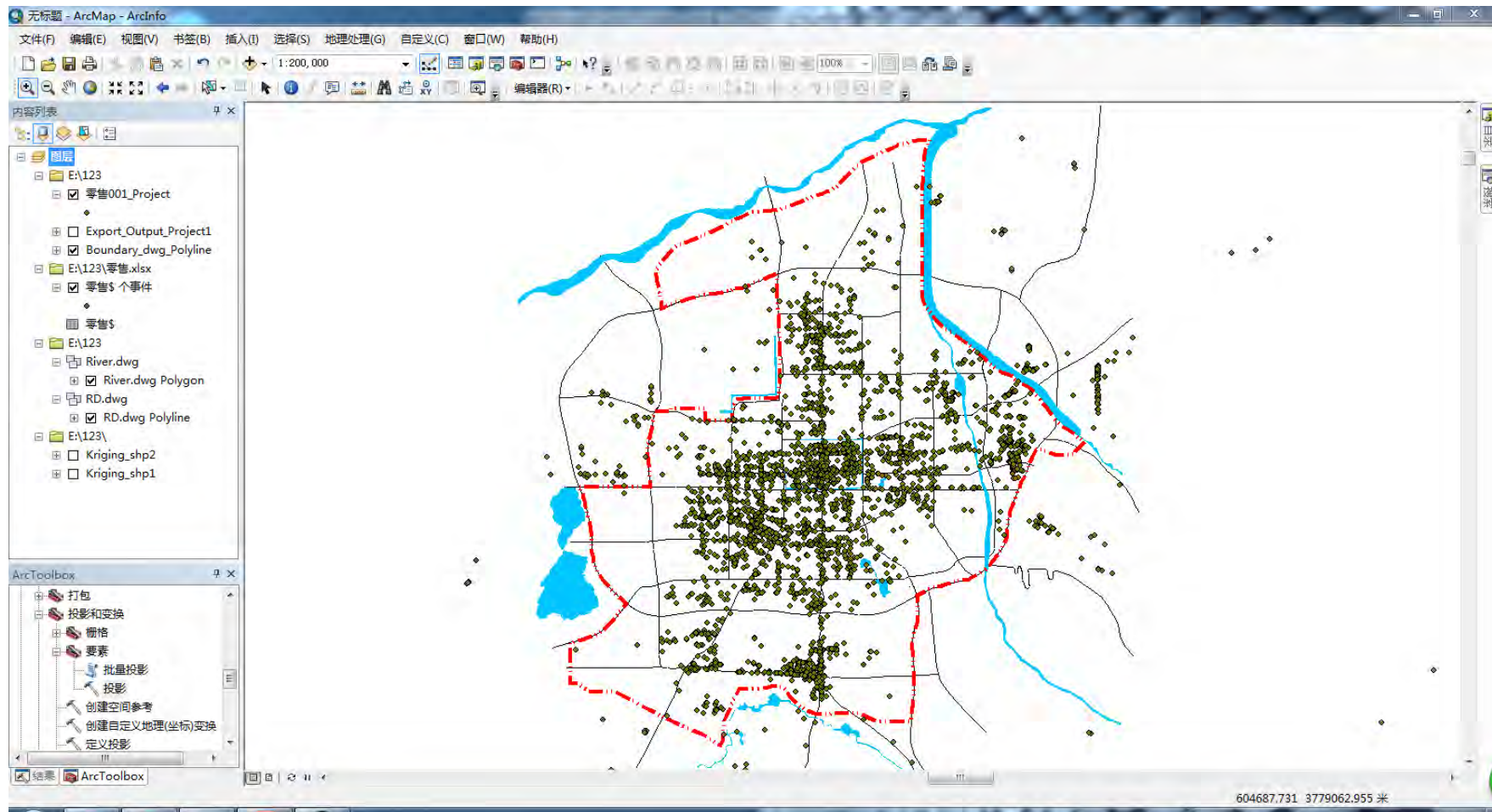


# 数据空间化 (GeoCoding)



- 定义投影系统

# 数据空间化 (GeoCoding)



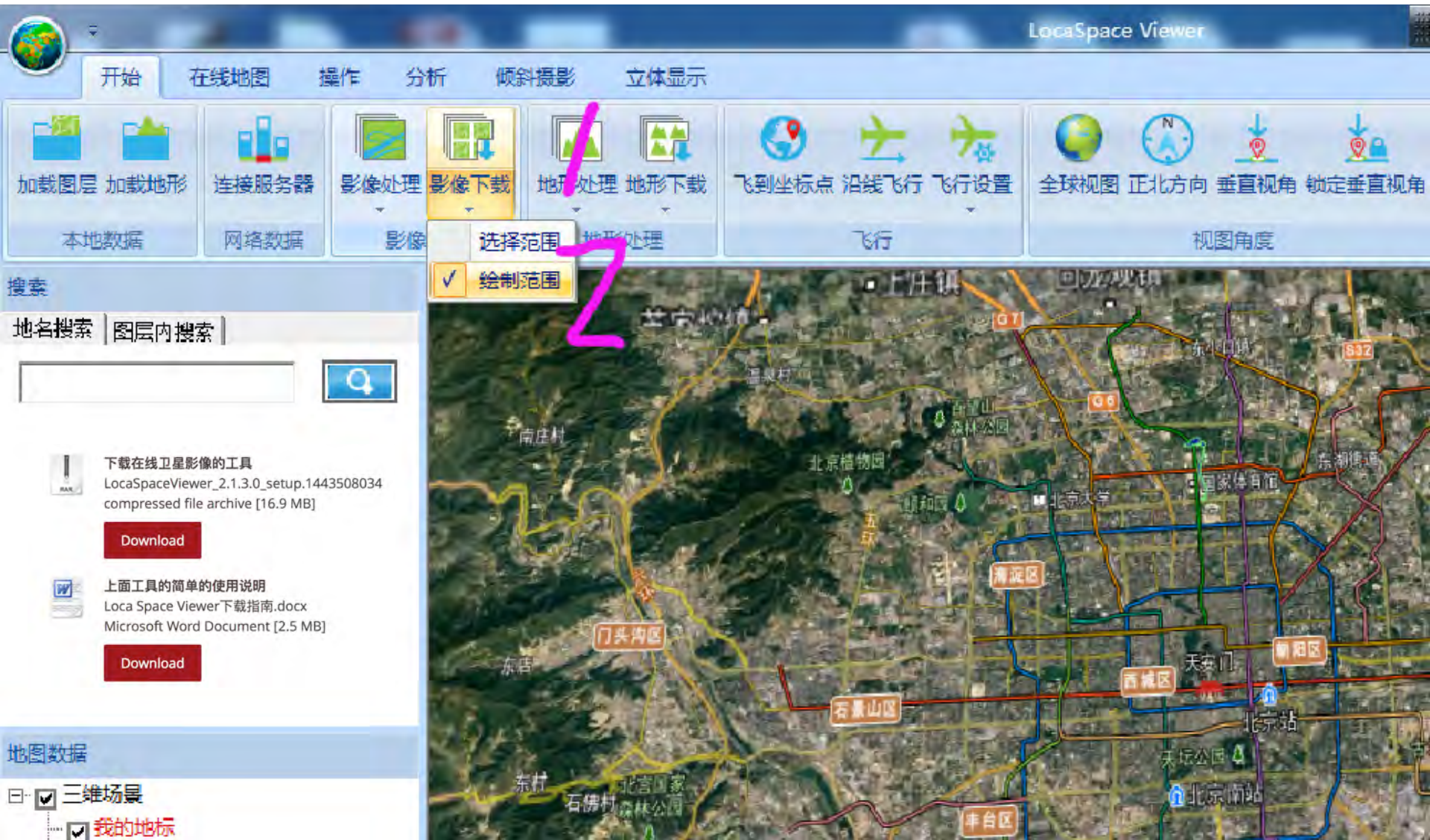
- 空间化后的数据一览
- 点状地物可以用于后续的不同空间单元的分析与统计

# 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）！



# 自动下载影像地图的工具 Loca Space Viewer



- <https://www.beijingcitylab.com/courses/structural-urban-design/>

# 二、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

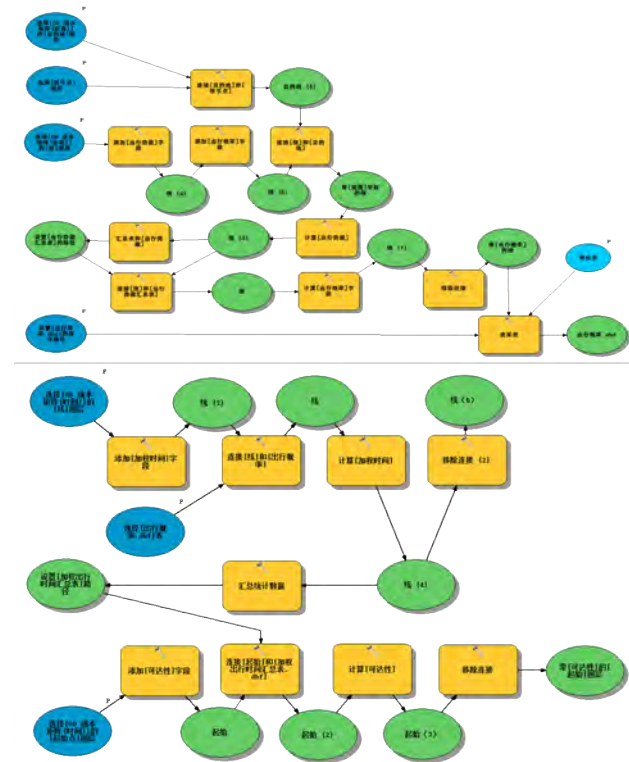
- 显示与管理
  - 加载、增加/删除字段、选择一部分对象
- 计算
  - 计算字段（属性与空间）、关联join、空间关联spatial join
- 分析与统计
  - 一般需要调用Toolbox
  - 字段的summarize功能
- 可视化
  - 符号化、layout视图

## • 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

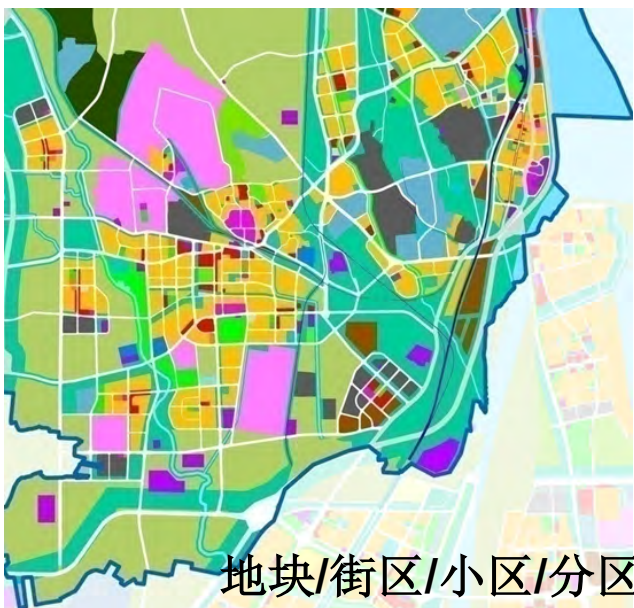
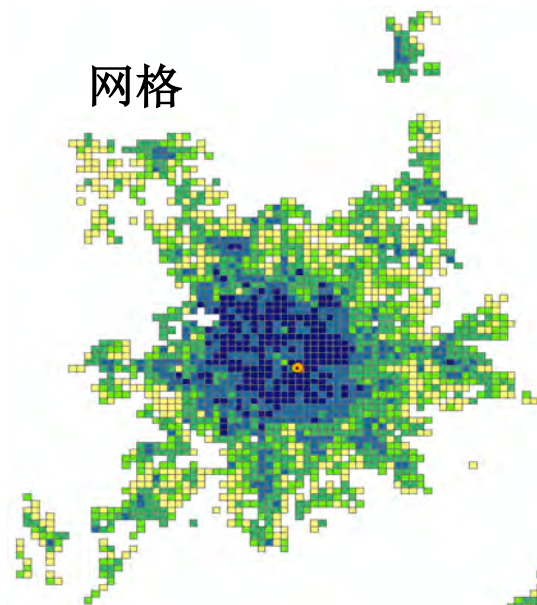
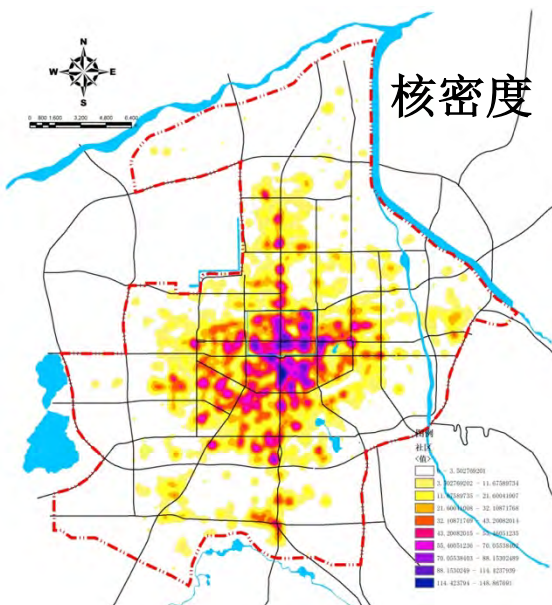


- 会了Python与不会所做的工作的巨大差异（个人经历的分享）
- 具体演示：

# 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

# 三、基于SPSS的数据统计

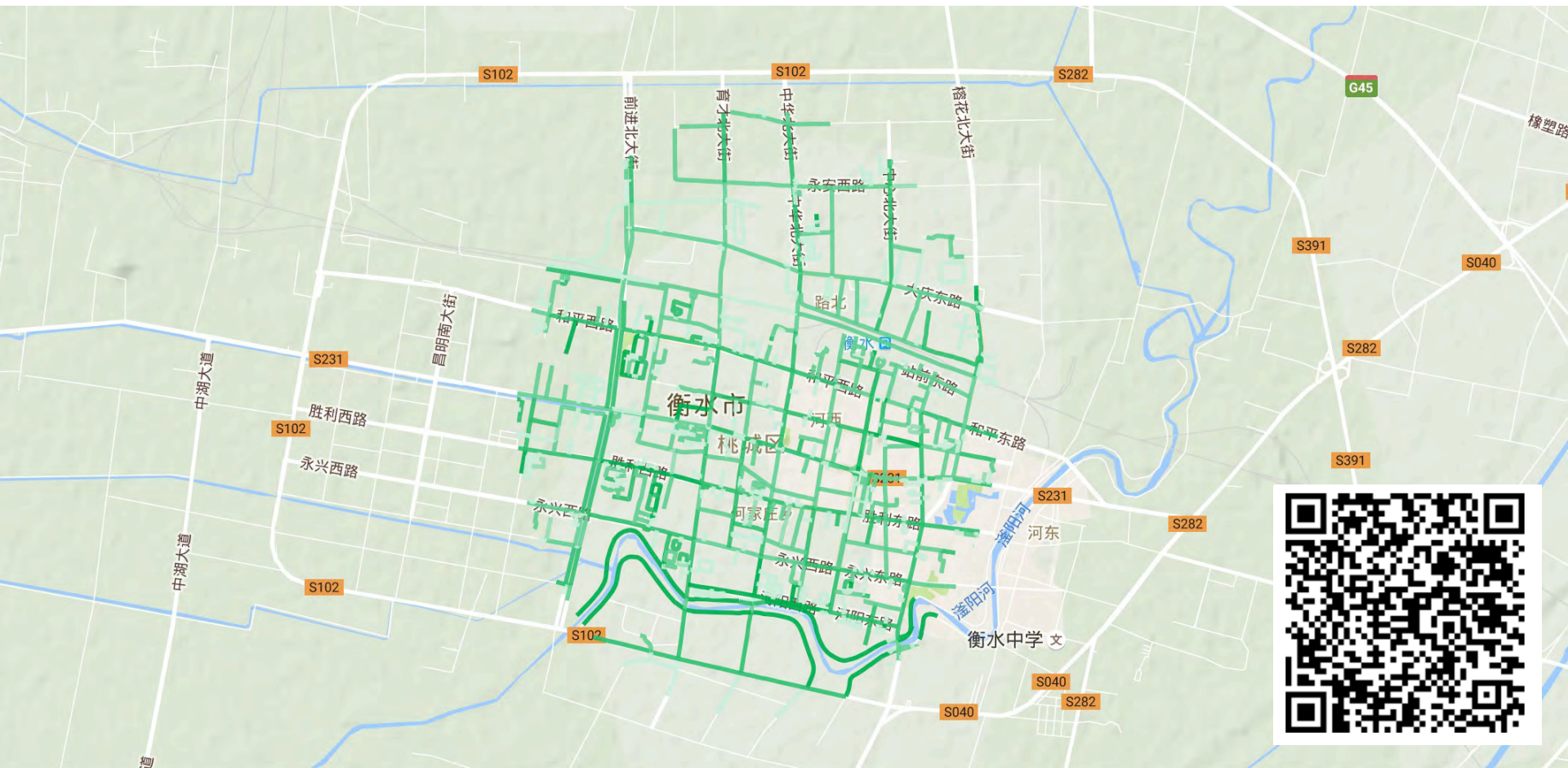
SPSS的很多功能可以用Excel实现

- 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

# 四、案例介绍

基于街景图片的街道绿化研究

# 成果展示：GeoHey平台



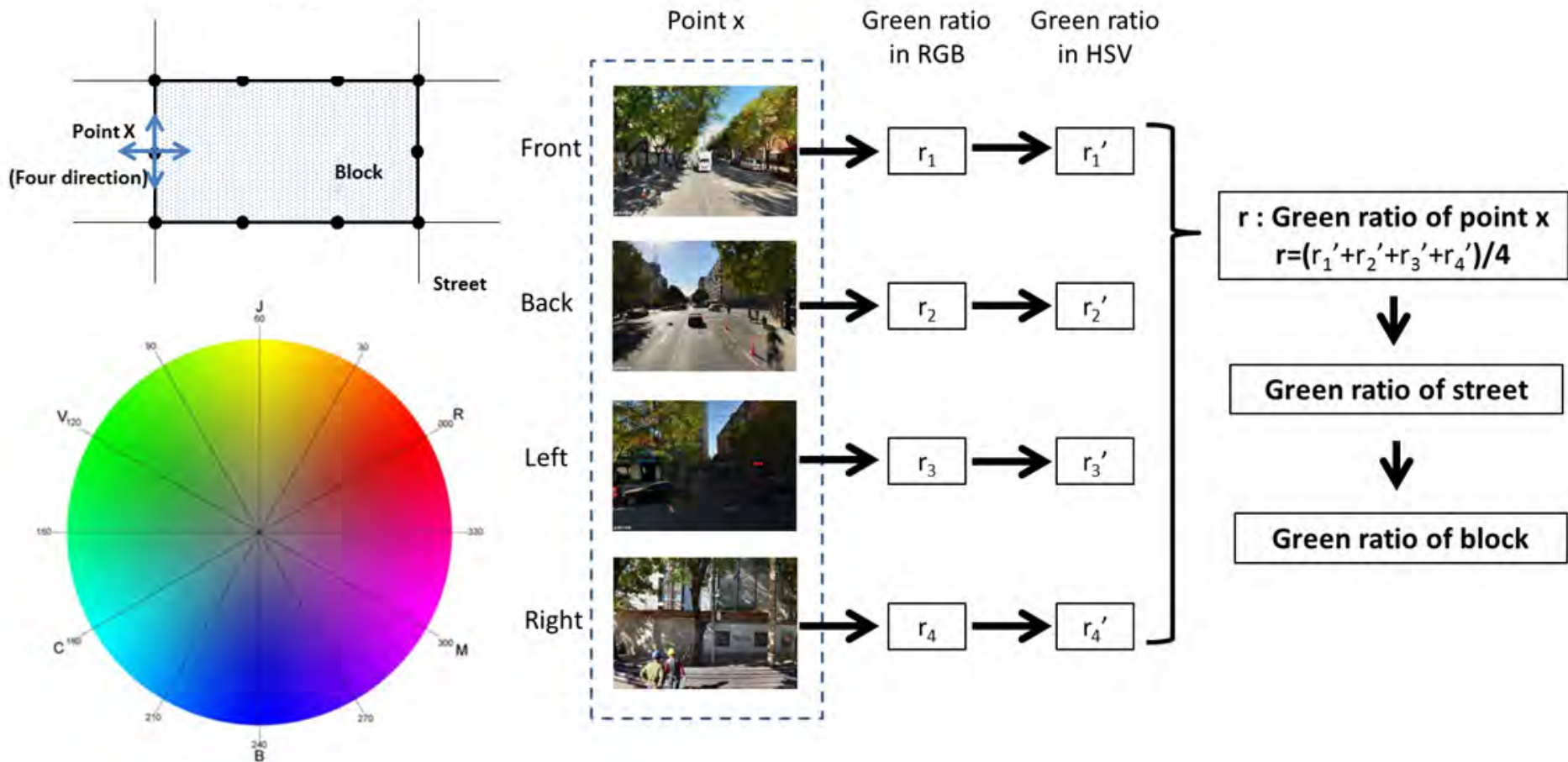
- <https://geohey.com/apps/dataviz/357b07615c4b4e25b76dcdd1ca9cd8f2/share?ak=ZmYzNmY0ZWJhYjcwNGU2ZGExNDgxMWUxNmZiOWNhNGY>

# 绿地率 vs 绿视率



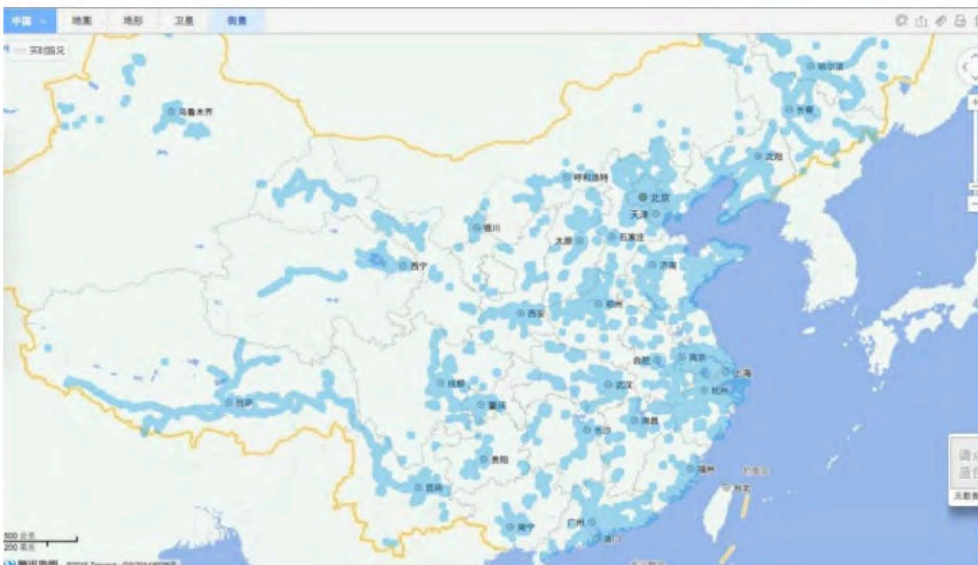
- 绿化是建成环境的重要要素，具有净化空气、缓解紧张情绪等作用，是空间规划关注的重要对象（如长久以来对田园城市的追求）
- 国家住房与城乡建设部的国家园林城市多批名单
- 平面的绿化（绿地率）与立体的绿化（绿视率）

# 基于街景图片评价街道绿视率的技术路线



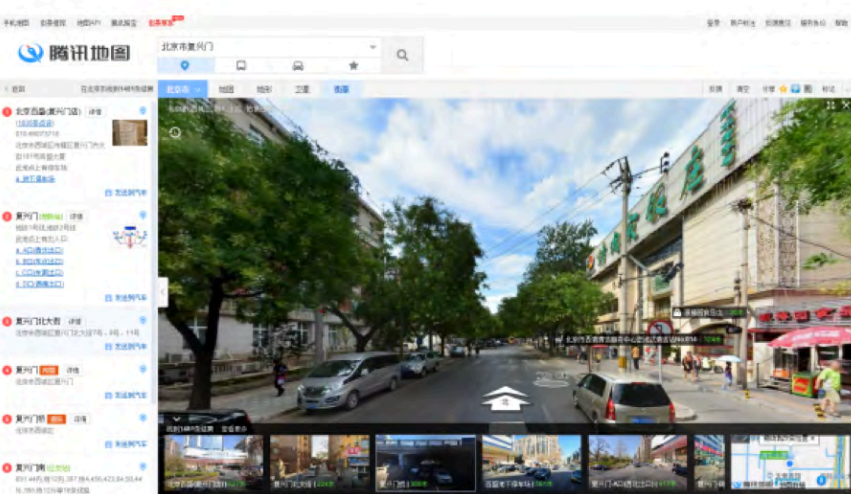
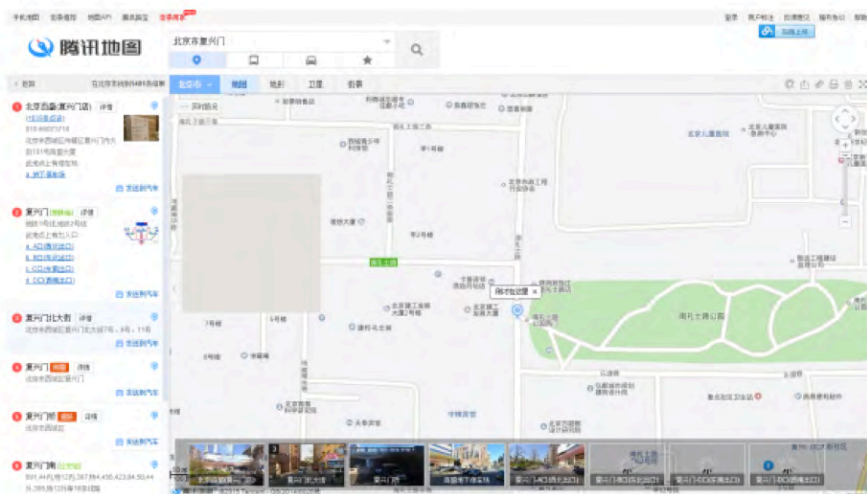
- 基于街景图片对街道绿视率进行评价，主要包括街景点提取（街道50m间隔）、街景图片抓取、街景图片识别以及绿视率统计分析四个环节

# 腾讯街景地图 (中国最大的街景服务提供平台, 具有时光机功能)



(a)

(b)



(c)

(d)



# 腾讯提供的街景图片抓取的API

<http://apis.map.qq.com/ws/streetview/v1/image?size=600x480&pano=10011022120723095812200&pitch=0&heading=0&key=OB4BZ-D4W3U-B7VVO-4PJWW-6TKDJ-WPB77>

Parameter	Mandatory item or not	Description	Examples
size	Yes	Picture size in pixel, maximum width 960 px and height 640 px	size=138x187
location	One in location or pano	Coordinates or place name for confirming the street view location	location=Tsinghua University or location=39.12,116.83
pano		Street view ID for confirming the street view location	pano=10011022120723095812200
heading	No	The value of heading represents the angle the forward direction making with the north, which is measured in clockwise with a range from 0 to 360 degree (0 as the default value)	North: heading=0 East: heading=90 South: heading=180 West: heading=270
pitch	No	The vertical angle of the camera covers -20 to 90 degree, in which a positive number stands for the level of looking up and vice versa (0 as the default value)	pitch=0
key	Yes	Developer's key (can be retrieved by online application)	key=OB4BZ-D4W3U-7BVVO-4PJWW-6TKDJ-WPB77

[http://lbs.qq.com/panostatic\\_v1/guide-getImage.html](http://lbs.qq.com/panostatic_v1/guide-getImage.html)

清华大学



# 基于API抓取街景的代码示意

```
def save_file(path, file_name, data):
    if data == None:
        return

    mkdir(path)
    if(not path.endswith("/")):
        path=path+"/"
    file=open(path+file_name, "wb")
    file.write(data)
    file.flush()
    file.close()

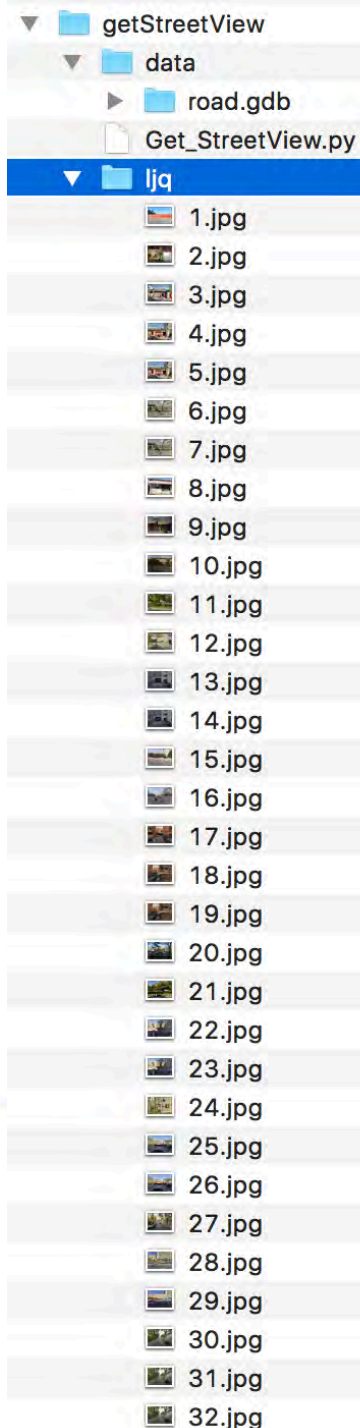
#读取坐标
#获取经度坐标
def get_Long(shp):
    env.workspace="d:/data/road.gdb"
    #print 'Processing'
    #读取坐标
    cur=arcpy.SearchCursor(shp)
    #经度(x)
    point_Long=[]
    for row in cur:
        point_Long.append(row.POINT_X)
    return point_Long
#获取纬度坐标
def get_Lat(shp):
    env.workspace="d:/data/road.gdb"
    #print 'Processing'
    #读取坐标
    cur=arcpy.SearchCursor(shp)
    #经度(x)
    point_Lat=[]
    for row in cur:
        point_Lat.append(row.POINT_Y)
    return point_Lat

#获取街景所对应的点的ID
def get_FID(shp):
    env.workspace="d:/data/road.gdb"
    cur=arcpy.SearchCursor(shp)
    point_FID=[]
    for row in cur:
        point_FID.append(row.FID)
    return point_FID

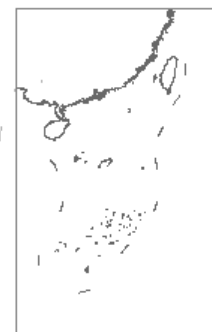
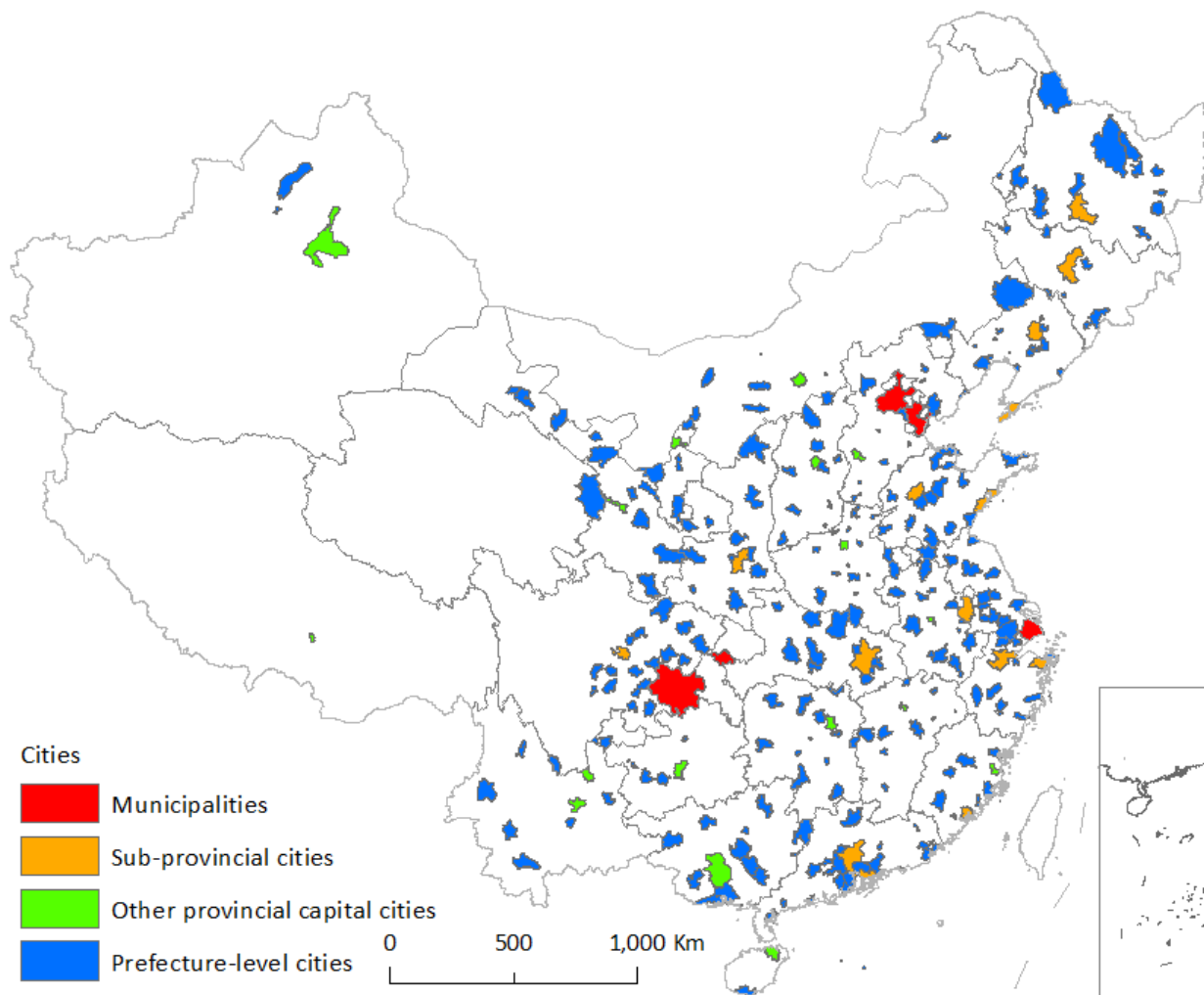
#获取ID
def get_id(data):
    match1=re.search(r'id',data)
    if match1 is not None:
        ID=re.findall(r'id": "(\\w*)',data)
        return ID

#开始, 获取街道节点坐标及其FID
print "start"
shp="shp_FeatureVerticesToPoints1.shp"
point_Long=get_Long(shp) #经度
point_Lat=get_Lat(shp) #纬度
FID=get_FID(shp) #FID

i=0
for i in range(0,len(point_Lat)):
    print "Print Picture"+str(i+1)
    #print i
    #根据街道节点坐标获取街景ID
    url_point="http://apis.map.qq.com/ws/streetview/v1/getpano?location="+str(point_Lat[i])+","+str(point_Long[i])+
"&radius=200&key=ULTBZ-VZ7WD-YLJ4Y-P4W7P-O2LLS-54F3J"
    ID_1=get_id(get_file(url_point))
    #ID存在, 获取该ID对应的街景
    if ID_1 is not None:
        #除去读取出的街景ID前后多余字符
        ID_2=re.sub(r'\\\'',',',str(ID_1))
        ID=re.sub(r'\\\'',',',str(ID_2))
        #print ID
        #获取相应街景
        url="http://apis.map.qq.com/ws/streetview/v1/image?size=900x640&pano="+str(ID)+"&pitch=0&heading=0&
key=ULTBZ-VZ7WD-YLJ4Y-P4W7P-O2LLS-54F3J";
        #print url
        #以该街景对应点的FID命名, 保存该街景
        save_file("d:/ljq/",str(FID[i])+".jpg", get_file(url))
    i+=1
```

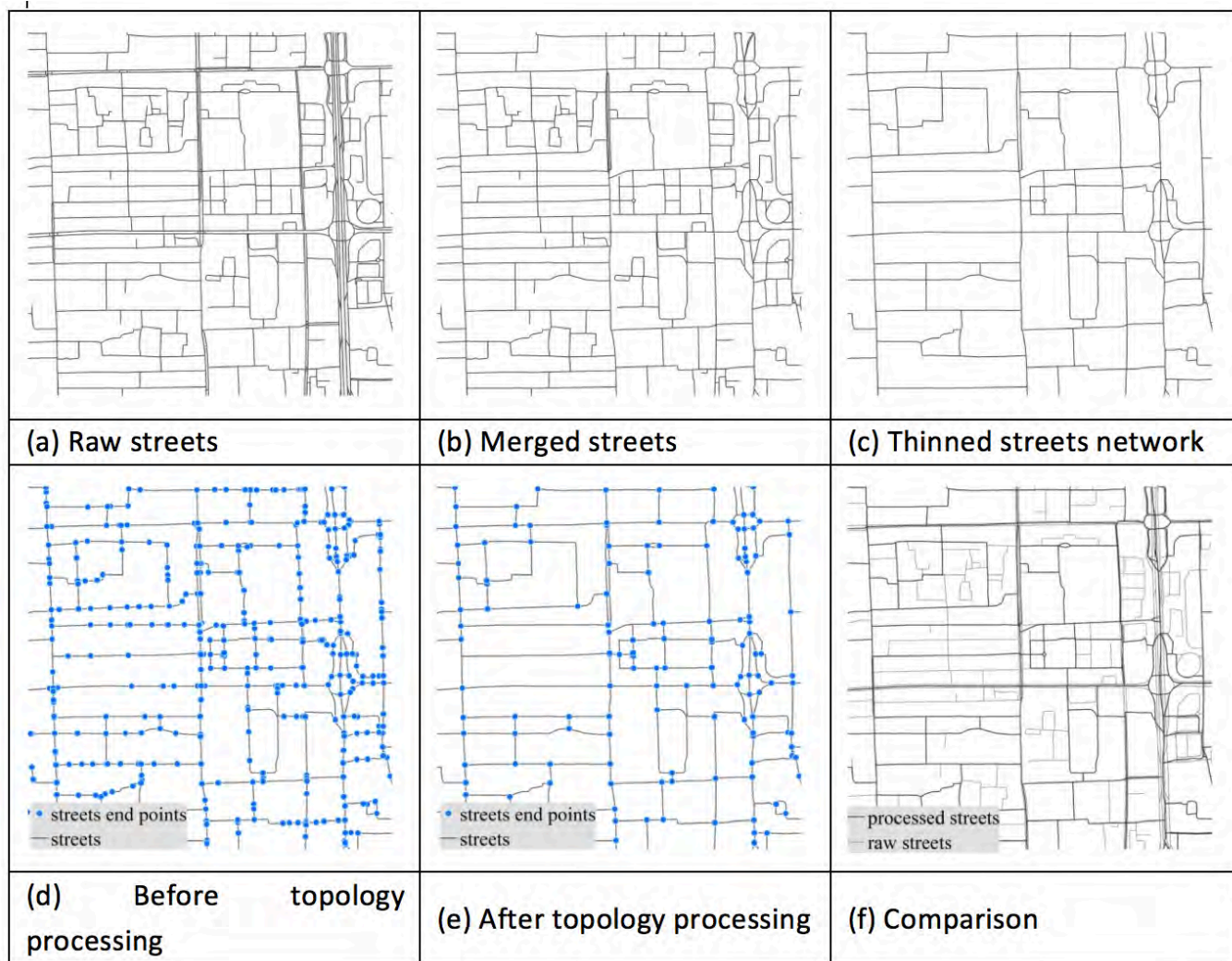


# 研究范围：中国288个地级及以上城市的中心区



- 4个直辖市，15个副省级城市，17个一般省会城市，252个地级市
- 考虑到有限的抓取时间和计算能力，选取每个城市中心的3km半径范围作为每个城市的研究范围（对应74.8万条街道）

# 街道数据预处理流程



- 为了更便捷地抓取街道上不同位置的街景图片（每隔50m），需要对街道数据进行预处理，需要合并街道、瘦化街道和拓扑处理三个步骤（ArcGIS中完成）

# 不同绿视率水平对应的街景图片示意



Figure 6 Street view pictures with various green percentages (only 127 locations/sites with the green ratio greater than 0.8)

not green  $\leq 0.2$   
 somehow green (0.2-0.4]  
 green (0.4-0.5]  
 very green  $> 0.5$

# 336,990个位置的平均绿视率为0.248



Figure 7 Typical SVPs with a green ratio of 0.248

# 北京中心城西北地区的街道绿化评价



- 基于各个点的街道绿视率计算结果，对街道层次的绿视率进行评价，较高的街道可以作为步行系统规划的参考

# 131个有效城市的结果一览

Type	# features	Min	Max	Mean	Green ratio			
					<0.2	0.2-0.4	0.4-0.5	>0.5
Locations	173,425	0.000	0.913	0.277	55,962 (32.3%)	85,702 (49.4%)	21,224 (12.2%)	10,537 (6.1%)
Street segments with over 13 locations per km)*	23,917	0.002	0.840	0.261	8,188 (34.2%)	12,619 (52.8%)	2,258 (9.4%)	852 (3.6%)
Blocks greater than 1 ha and with over 1 location per ha**	9,424	0.002	0.737	0.265	2,583 (27.5%)	5,931 (62.9%)	718 (7.6%)	192 (2.0%)

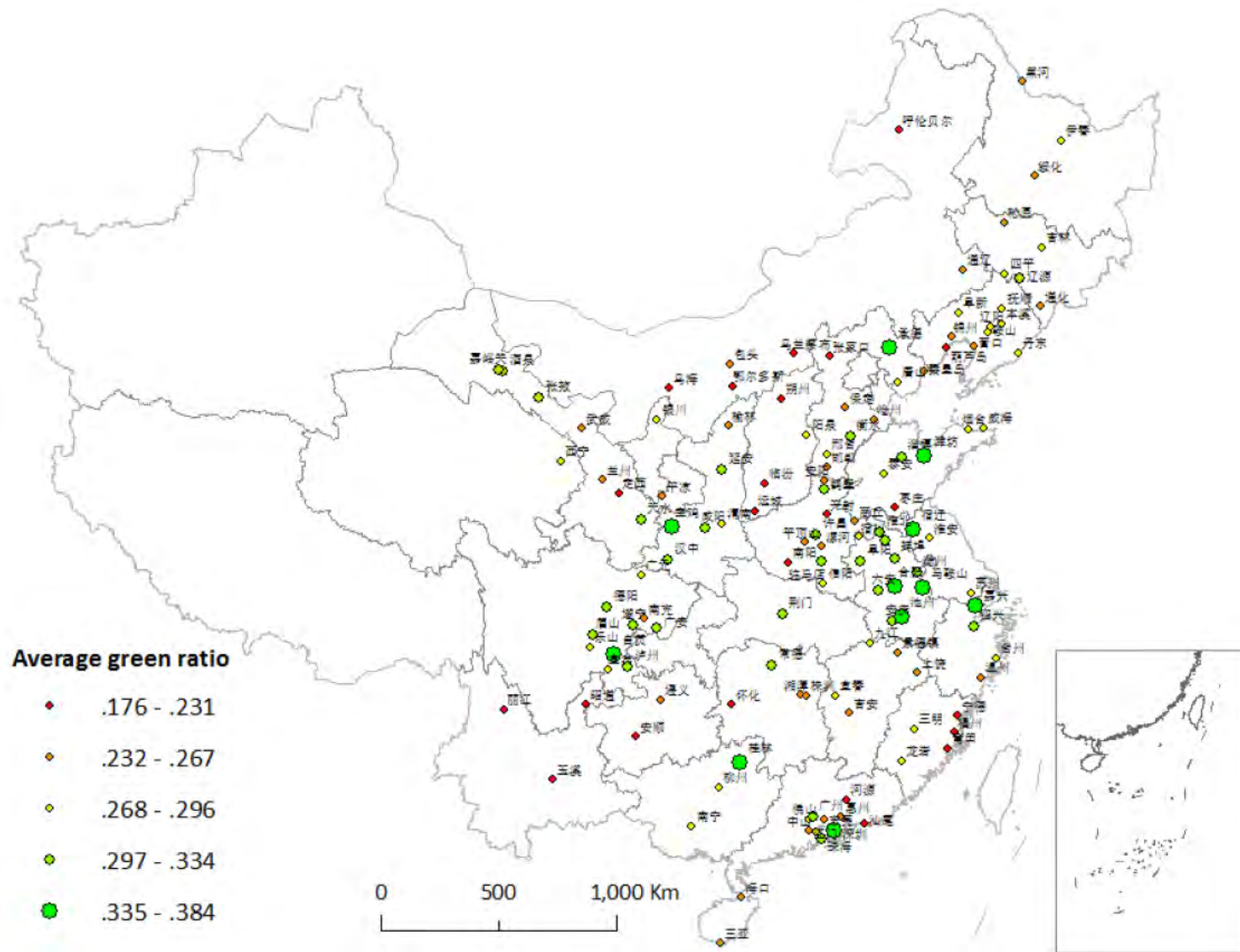
\* “13” is the average value of location density for all street segments.

\*\* “1” is the average value of location density for all blocks greater than 1 ha

- 部分城市的街景拍摄日期不适合评价街道绿视率（如秋冬季节）
- 131个有效城市的平均街道绿视率范围为 0.132-0.384



# 131个有效城市的平均街道绿视率

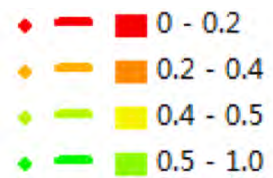
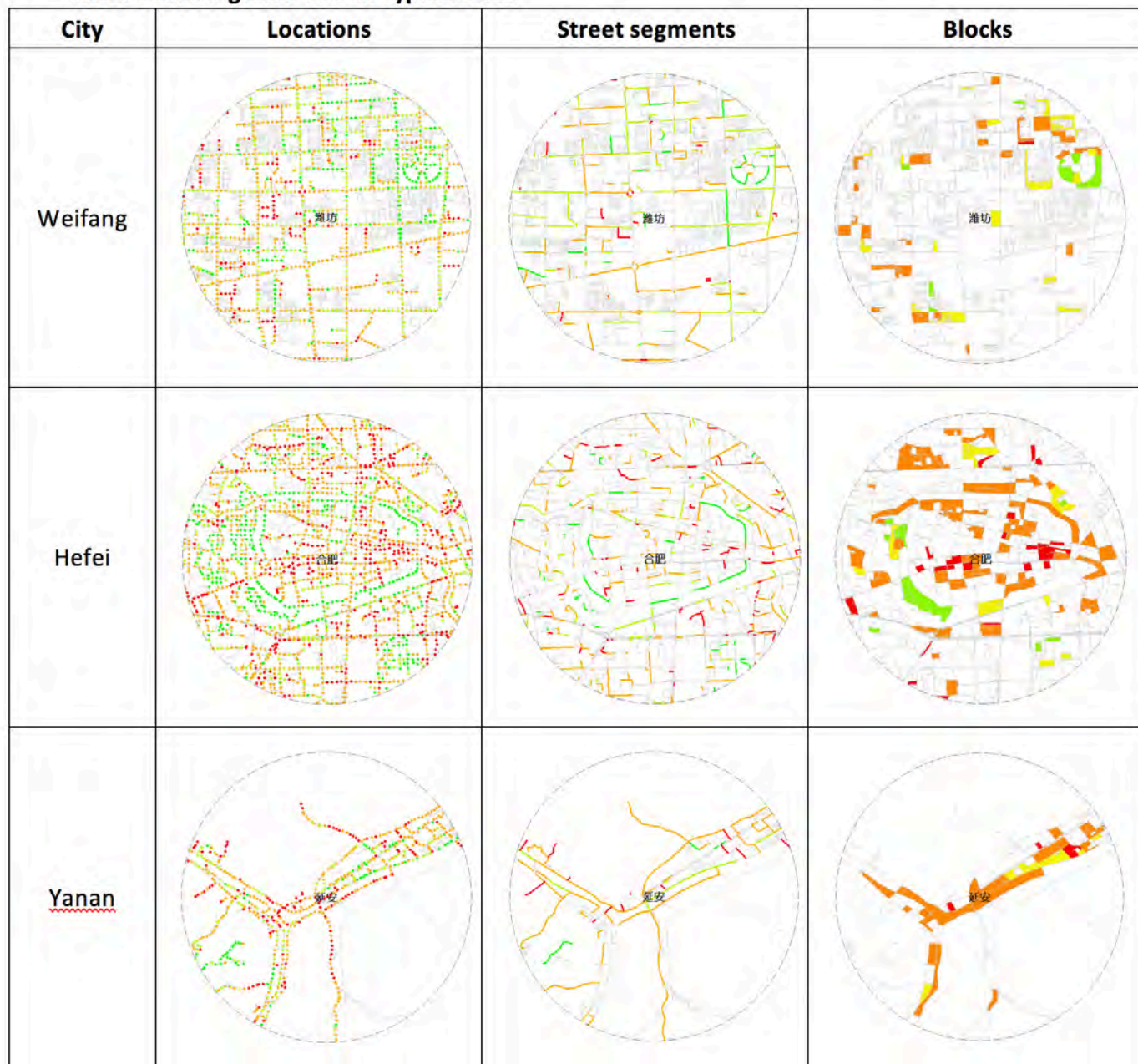


- 前五城市均为国家园林城市（潍坊、自贡、宝鸡、马鞍山和承德）



# 典型城市的街道绿视率结果

Table 3 Street greenness for typical cities



# 街道绿视率的解释模型

Variables	Model1		Model2		Model3	
	Coefficients	<i>p</i> values	Coefficients	<i>p</i> values	Coefficients	<i>p</i> values
(Constant)						
CENTER	<b>0.061</b>	0.000	<b>0.053</b>	0.000	<b>0.052</b>	0.000
LENGTH	<b>0.080</b>	0.000	<b>0.093</b>	0.000	<b>0.093</b>	0.000
SIZE			-0.014	0.191	<b>0.025</b>	0.006
LEVEL			<b>-0.017</b>	0.041	-0.010	0.242
DENSITY			<b>-0.070</b>	0.000	<b>-0.088</b>	0.000
DESIGN			<b>0.083</b>	0.000	<b>0.108</b>	0.000
ECONOMY			<b>0.047</b>	0.000	<b>0.066</b>	0.000
MIDDLE					<b>0.017</b>	0.019
WEST					<b>0.060</b>	0.000
Adjusted R <sup>2</sup>	0.010		0.027		0.029	

Note: coefficients in bold indicate being significant at the 0.05 level

- 在点的层面，街道绿化率与多个因素的回归模型显示，距离城市中心越远，街道倾向于更绿；街道长度更长，城市经济越发达、等级越高、人口密度越低，街道倾向于更绿；

# 关于做研究的小窍门

- 第一讲/第二讲：参考文献的重要性
  - 外国人的姓名写法、认真与否、文献等级
- 第三讲：
  - 论文与报告的区别（是否有科学问题）
  - 问题：Problem vs Question
- 第四讲：两类论文
  - 方法：证明方法优于已有的其他方法（效率/科学性、规划师/公众/同行评价？）
  - 实证：证明发现，与其他人发现的异同，对理论的贡献
- 第五讲：善用Google Scholar
  - <https://scholar.google.com/citations?user=4KAatI4AAAAJ>
  - 文献检索、跟踪某个学者的新发表或新引用、查看自己领域的更新updates、参考文献格式生成

- 本课件待放到课程网站

- <https://www.beijingcitylab.com/courses/bdup2017/>

龙瀛 周垠 Long Ying Zhou Yin

- 阅读材料：

- Liu and Long 2016 EPB（地块）
  - 龙瀛和周垠 2016 新建筑（街道）

街道活力的量化评价及影响因素  
分析

——以成都为例

Quantitative Evaluation on Street Vibrancy and Its Impact Factors: A Case Study of Chengdu

- 答疑

- [ylong@tsinghua.edu.cn](mailto:ylong@tsinghua.edu.cn)
  - 建筑学院新501办公室（默认每周下课后-11AM、Open Office Hour）
    - 建议提前两天通过info系统进行预约
  - 助教：徐婉庭，[wantingsu1031@qq.com](mailto:wantingsu1031@qq.com)

Article

## Automated identification and characterization of parcels with OpenStreetMap and points of interest

Xingjian Liu

The University of Hong Kong, Hong Kong

Ying Long

Tsinghua University and Beijing Institute of City Planning, China

B Planning and Design

Environment and Planning B:

Planning and Design

2016, Vol. 43(2) 341–360

© The Author(s) 2015

Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/0265813515604767

[epb.sagepub.com](http://epb.sagepub.com)





龙瀛, [ylong@tsinghua.edu.cn](mailto:ylong@tsinghua.edu.cn), 新建筑馆501, 13661386623



北京城市实验室  
Beijing City Lab

<http://www.beijingcitylab.com>



新浪微博: 龙瀛a1\_b2 北京城市实验室BCL

微信公众号: beijingcitylab

清华大学

