

# 《大数据与城市规划》课程概论

## 课程共享的基础数据

Urban Data of Beijing Old City

龙瀛, 清华大学建筑学院

ylong@tsinghua.edu.cn

清华大学



# 一、概论

课程整体情况

代表性研究机构与个人

课程共享的基础数据

# 本讲重点

---

北京旧城数据介绍

基本设置

相关阅读材料

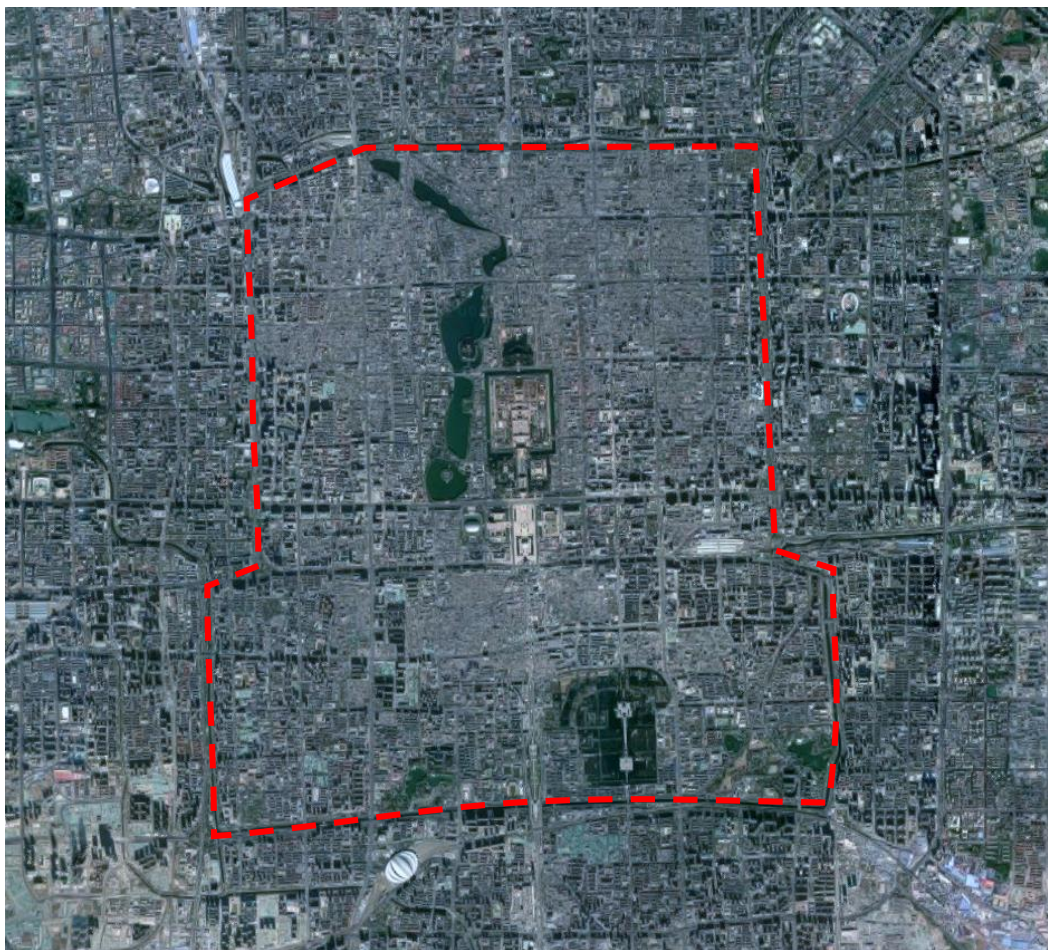
# 北京旧城数据介绍

---

北京旧城

# 北京旧城

---



# 北京旧城数据 分类目录

p.s. 本页可直接点击字段进行查阅

## 点位数据

### 交通点位

**BusStations2014**

公交站点2014

**Junction2009/2014**

街道交叉口2009/2014

**SubwayStations**

地铁站

### 经济点位

**HousingPrice**

房价

**Jobs\_economicsurvey3**

第三次经济普查(社区尺度)

**LandTransaction2013**

土地交易情形2013

### 街景图片

**StreetViewPictures\_point\_MS\_API**

街景图片位置点(使用微软识别点)

**StreetViewPictures\_point\_SegNet**

街景图片位置点(使用SegNet计算)

### 设施点位

**Firms2010**

企业机构2010

**Office201609**

写字楼2016年09月

**POI2009/2011/2014/2016**

兴趣点2009/2011/2014/2016

**Tiananmen**

天安门

### 活动点位

**Dianping2014jped**

大众点评点2014

**Weibo\_POI\_Checkins**

微博签到点

**Weibos2016**

微博发文位置2016

## 线状数据

### 交通联系

**Commute2016alpha**

通勤OD2016

**Road2009/2014/2016**

道路2009/2014/2016

### 街道状态

**Streets\_Harvey**

街道形态指标

**Streets\_WalkScore**

街道步行指数

## 面状数据

### 旧城范围

**Ext**

研究范围

### 实体环境

**Buildings2017**

建筑物2017

**Parcels2011AICP**

地块2011(依据路网划分)

**GeoGrids**

地理编码

(北京城市规划设计院划分)

### 行政边界

**FunctionalZones**

功能分区

**Jiedaos**

街道办事处

**TownshipPOP**

街道办事处 人口信息

**Juweihui**

居委会

**ResidentialCommunity**

居住小区

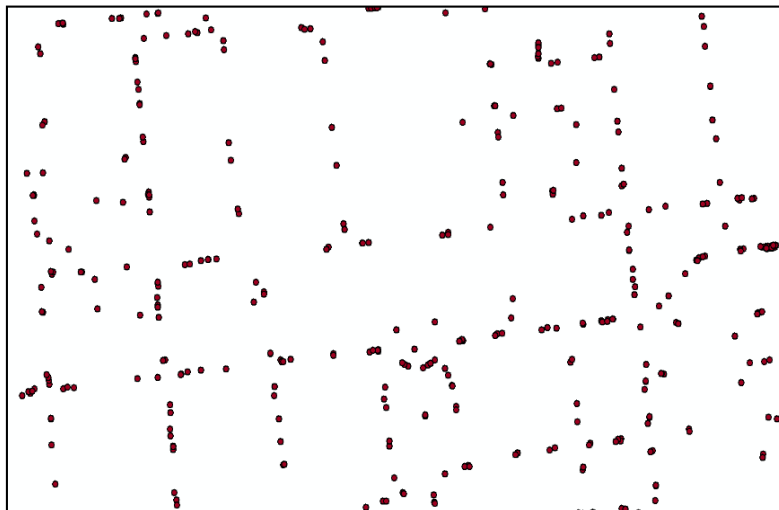
**TAZs**

交通分析地块块

(北京市交通管理局划分)

# BusStation2014

## 公交站点2014



图层类型File Type : Point

数据量Amount : 5,125

核心字段Key Word :

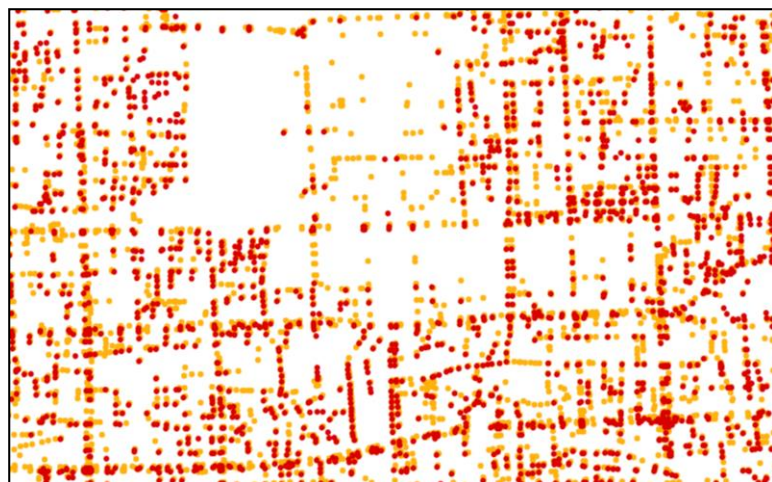
#name: 公交站名

#num : 到达该站的公交总数

图层意义Meanings : 说明各公交站点的分布 (年份 : 2014)、公交站名以及会抵达该站的公交车数量。

# Junction2009 / 2014

## 道路交岔口 2009 / 2014



● 2009  
● 2014

图层类型File Type : Point

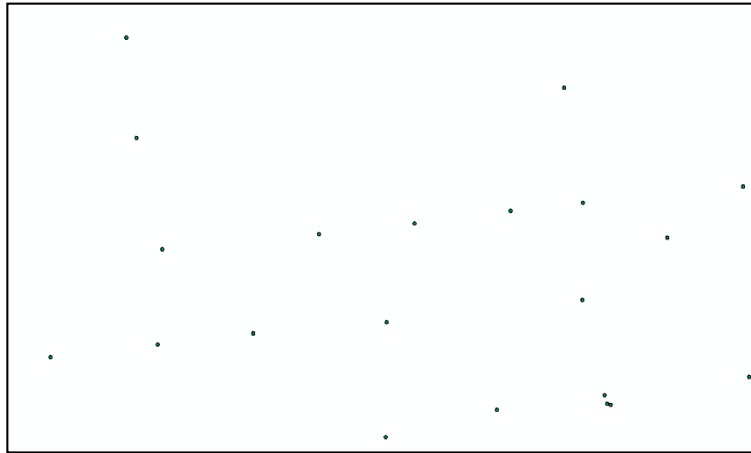
数据量Amount : 11,224 / 21,573

图层意义Meanings : 显示2009、2014年的道路交岔口的空间分布



# SubwayStations

## 地铁站点



图层类型File Type : Point

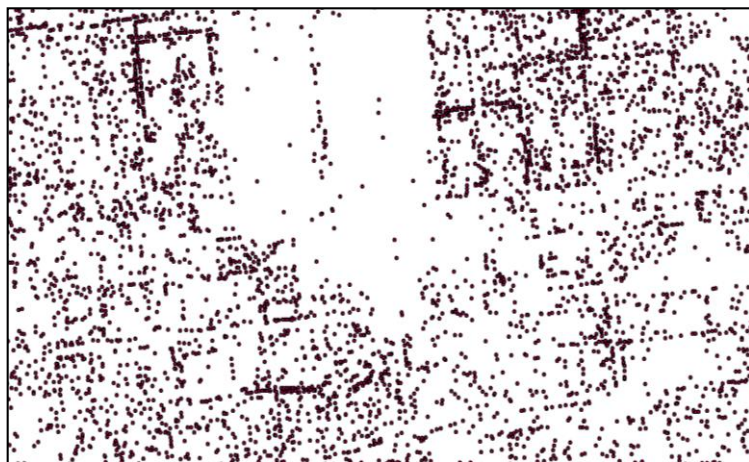
数据量Amount : 123

核心字段Key Word :

#name : 地铁站名

图层意义Meanings : 显示北京各地铁站点站名与位置。

## 企业机构2010



图层类型File Type : Point

数据量Amount : 107,492

核心字段Key Word :

#机构名称 : 企业名称

#行业(类别) : 企业所属类别

#员工人数 : 企业员工总人数

#成立时间 : 企业成立时间

#大/中/小类代码 : 与字段“行业”相对应

#地址 : 企业所在地址

#年营业额 : 企业年营业额

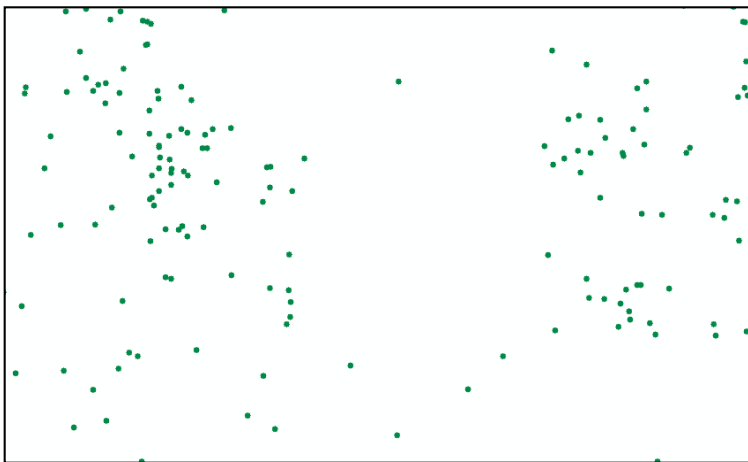
#企业性质 : 国有、私营、合作经营等

#注册币种/资本 : 企业注册相关信息

图层意义Meanings : 说明企业机构位置与相关信息。

# Office201609

## 写字楼 2016年09月



图层类型File Type : Point

数据量Amount : 483

核心字段Key Word :

#名称 : 写字楼名称

#在租房源数 : 当前可供租用房源数量

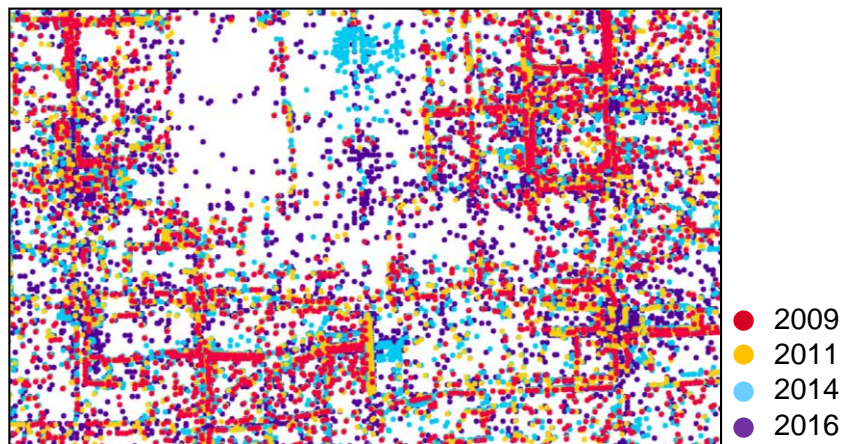
#日租金(元\_天\_平米) : 该写字楼内平均租金

#商圈 : 写字楼所属商圈

图层意义Meanings : 显示写字楼的名称和当时租金 (数据时间为 : 2016年09月)。

# POI2009/2011/ 2014 / 2016

## 兴趣点2009/2011 / 2014 / 2016



图层类型File Type : Point

数据量Amount(各年) :

13,118 / 25,210 / 47,764 / 223,206

核心字段Key Word :

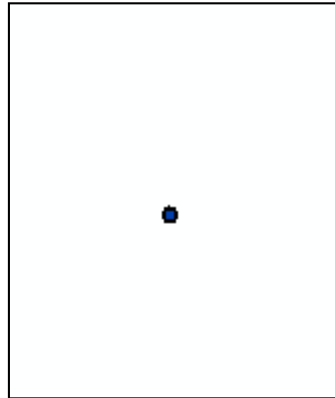
#NAME : 兴趣点名称

#CTYPE : 类型

图层意义Meanings : 分别显示2009、2011、2014、2016年的兴趣点POI位置、类别与相关信息。

# Tiananmen

## 天安门



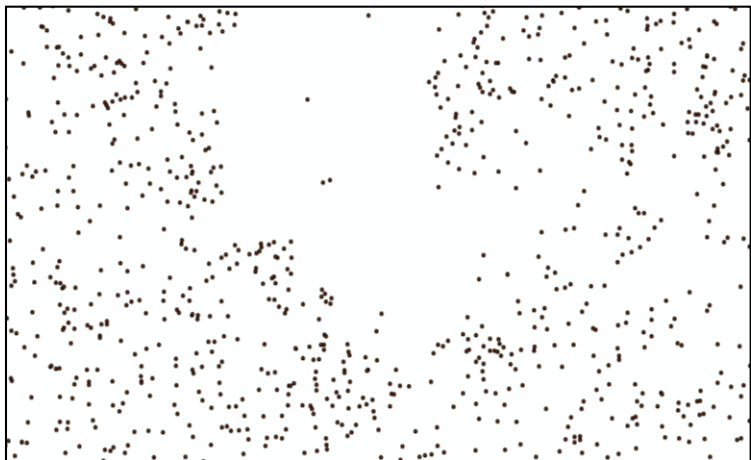
图层类型File Type : Point

数据量Amount : 1

图层意义Meanings : 显示天安门位置。

# HousingPrice

## 房价



图层类型File Type : Point

数据量Amount : 3,138

核心字段Key Word :

#小区名 : 小区名称

#均价 : 小区内住房平均售价

#楼栋总数 : 小区内住房楼栋数

#建筑类型 : 住房的建筑类型

#周遭服务指标 , 如 : 餐饮、公共设施、科教文化等服务指标

#建筑年 : 建成年份

#物业费用 : 小区内住房平均物业费

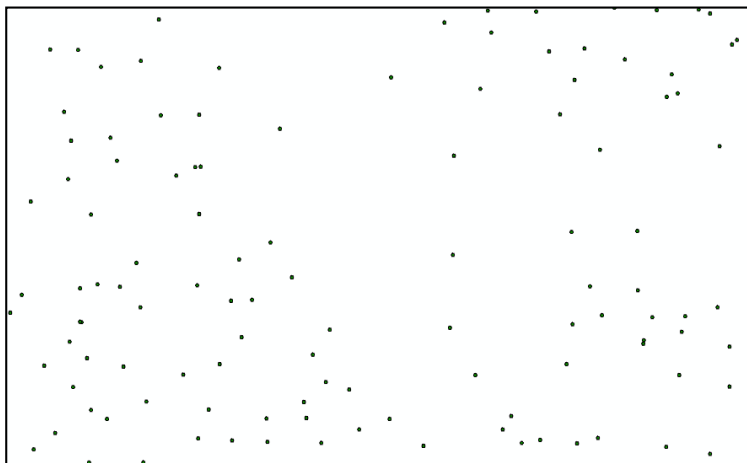
#房屋总数 : 小区内房屋总数

#在售房源数 : 目前在售的房源数量

图层意义Meanings : 显示2016年的住宅小区位置、特征和房价信息。

# Jobs\_economicsurvey3

## 第三次经济普查(社区尺度)



图层类型File Type : Point (社区中心点)

数据量Amount : 485

核心字段Key Word :

#OO\_法人从业人员期末人数 : 各行业期末之从业人员数量

#OO\_法人单位数 : 各行业已注册之法人单位数量

#OO\_营业收入 : 各行业营业总收入

图层意义Meanings : 显示以社区为单元所统计的各行业从业人口、单位数和营业收入  
(注意 : 此图层为社区尺度 , 以社区中心点代表社区)。

# LandTransaction2013

## 土地交易情况2013

图层类型File Type : Point

数据量Amount : 2,444

核心字段Key Word :

#投资项目名称 : 开发项目内容

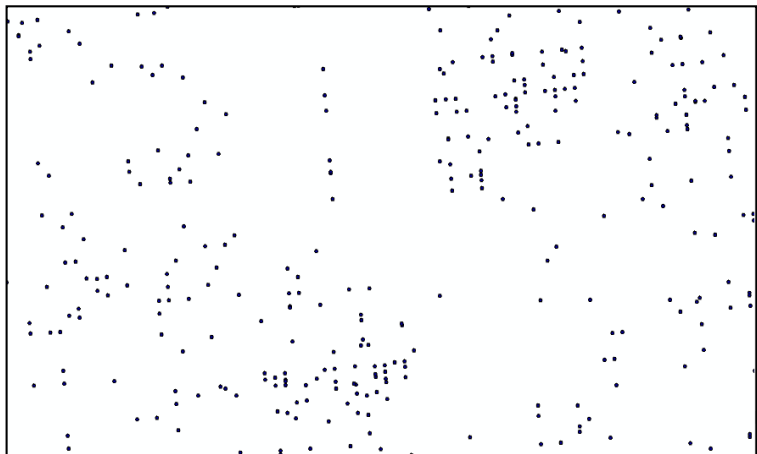
#土地用途YT\_SM : 该土地交易之主要用途，如仓储用地、高档住宅等。

#供应面积 : 该投资项目所提供之总土地面积

#建筑面积 : 该投资项目所提供之可建筑面积

#容积率 : 该投资项目之容积率

#土地均价 : 该投资项目之土地均价

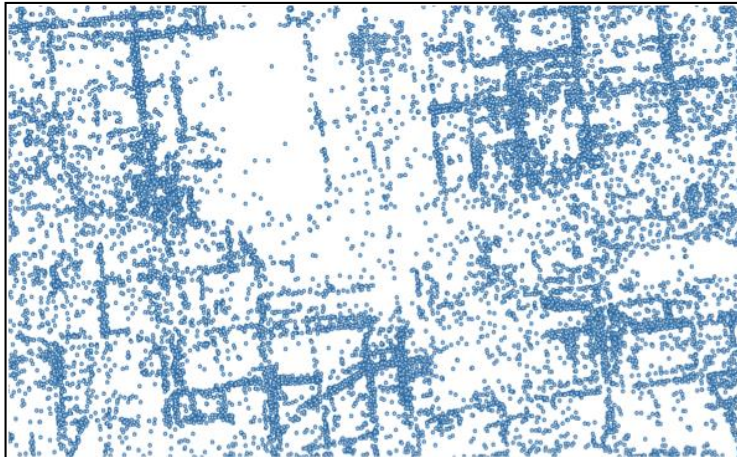


图层意义Meanings : 显示截至2013年的每笔土地交易之名称、类别等具体信息。



# Dianping2014jped

## 大众点评点2014



图层类型File Type : Point

数据量Amount : 108,995

核心字段Key Word :

#Name : 店家或设施名称

#big\_cate / small\_cate : 大小层级类别，为店家或设施分类

#all\_remarks/very\_good\_remarks to very\_bad\_remarks : 大众点评上评价状况

#avg\_price : 于该店家或设施平均消费金额

#tags : 与该店家或设施相关之标签

图层意义Meanings : 显示饮食休闲娱乐消费等相关设施点的分布、价位，及点评情况。

# Weibo\_POI\_Checkins

## 微博签到点

图层类型File Type : Point

数据量Amount : 51,185

核心字段Key Word :

#title : 签到点名称

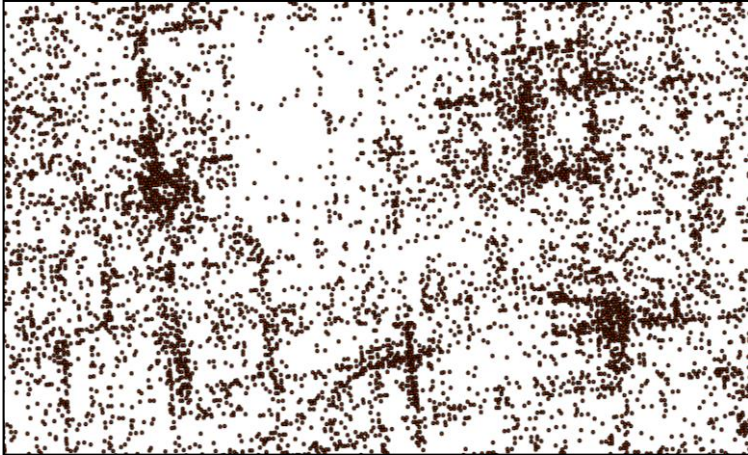
#checkin\_num : 于该签到点签到之次数总量

#checkin\_user\_num : 于该签到点签到之用户总数

#photo\_num : 于该签到点发出的照片总数

#categorys : 与分类名称相对应之编号

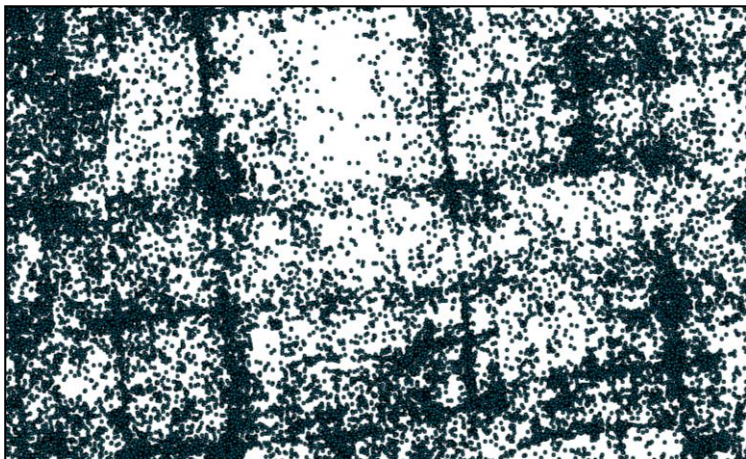
#category\_name : 签到点属性类别，如高等院校、餐饮美食、生活娱乐等等



图层意义Meanings : 显示各微博签到点的类别、名称、签到数量等信息。

# Weibos2016

## 微博发文地点2016



图层类型File Type : Point

数据量Amount : 759,805

核心字段Key Word :

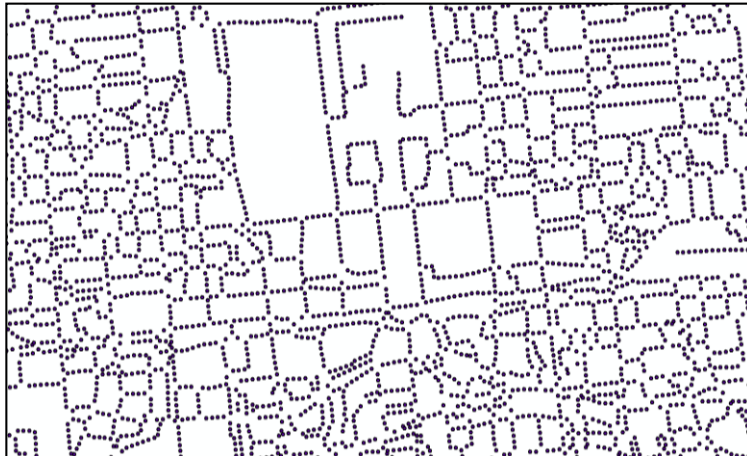
#TEXT : 用户发博的内容文本

#SOURCE : 用户所使用的设备，如iphone6s

图层意义Meanings : 此图层记录某时间段内，所有北京二环内的用户发博信息，其中包含发文时的位置、发文文本内容等信息。

# StreetViewPictures\_point\_MS\_API

## 街景图片位置点(使用微软API识别)



图层类型File Type : Point

数据量Amount : 13,013

核心字段Key Word :

#pano\_id : 该点所对应的街景图片编号

#CV\_R\* : 基于微软API所识别出的结果 (以代码方式呈现)

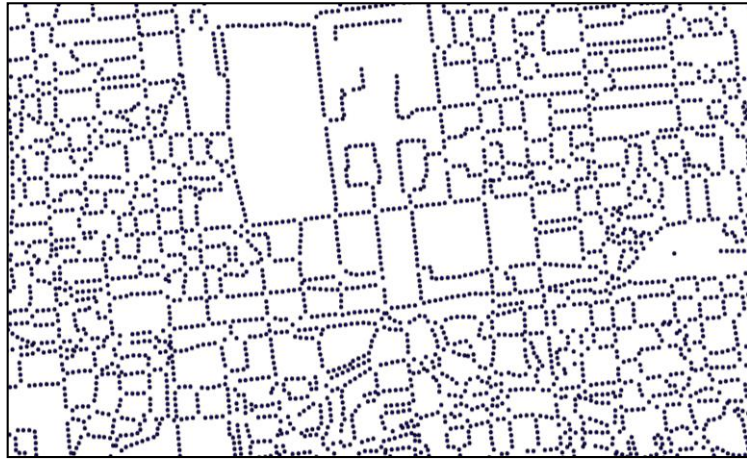
\*CV\_R :  
(example)

```
{
  "categories": [
    { "name": "abstract_", "score": 0.0078125 },
    { "name": "building_street", "score": 0.36328125 },
    { "name": "others_", "score": 0.00390625 },
    { "name": "outdoor_", "score": 0.015625 }
  ],
  "tags": [
    { "name": "outdoor", "confidence": 0.95521062612533569 },
    { "name": "road", "confidence": 0.95403581857681274 }
  ],
  "description": {
    "tags": [
      "outdoor", "road", "building", "car", "street", "front", "bus", "truck", "city", "parked", "large", "driving", "lot", "traffic", "riding", "parking", "people", "bench", "group", "man", "train", "standing", "white", "tall"
    ],
    "captions": [
      {
        "text": "a car parked in front of a building",
        "confidence": 0.60742662899963862
      }
    ],
    "requestId": "381f7ed1-8074-4678-9a5a-ce3b16e58242",
    "metadata": {
      "width": 900, "height": 640, "format": "Jpeg"
    },
    "faces": [],
    "color": {
      "dominantColorForeground": "Black",
      "dominantColorBackground": "Grey",
      "dominantColors": [ "Grey", "Black" ],
      "accentColor": "4C667F",
      "isBWImg": false
    }
  }
}
```

图层意义Meanings : 说明各街景图片位置，以及基于微软API的识别结果 (街景图片来源：腾讯)

# StreetViewPictures\_point\_SegNet

## 街景图片位置点(使用SegNet计算)



图层类型File Type : Point

数据量Amount : 12,919

核心字段Key Word :

#pano\_id :该点所对应的街景图片编号

#图像识别的各项指标，数值越大代表其指标占整体图片比例越高

识别的指标如：天空Sky, 建物Building, 竿Pole, 路面标示RoadMarking,

道路Road, 行人道Pavement, 树Tree, 篱笆Fence, 标示牌SignSymbol,

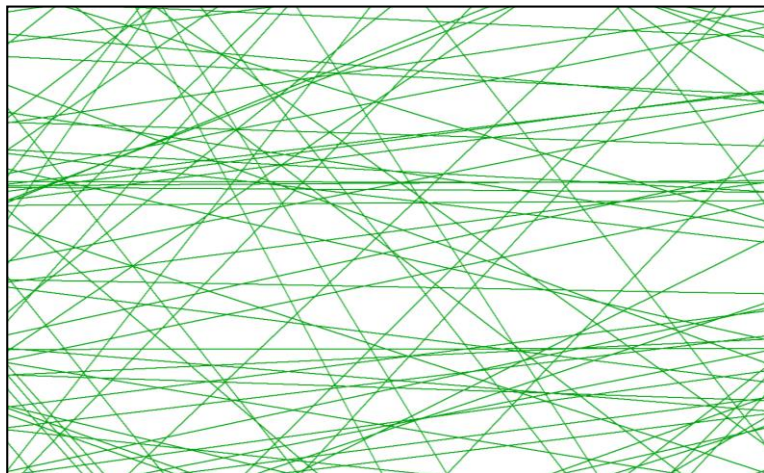
车Car, 行人Pedestrian, 骑行自行车者Bicyclist

图层意义Meanings : 此图层利用SegNet解码器识别街景图象，将街景综合环境根据周边要素和图形本身特征进行物体识别，实现场景客观要素的智能分割（详见唐婧娴和龙瀛(2017)）

唐婧娴，龙瀛·特大城市中心区街道空间品质的测度——以北京二三环和上海内环为例[J]·规划师，2017(2)：68-73.

# Commute2016alpha

## 通勤OD(直线距离)2016



图层类型File Type : Polyline

数据量Amount : 220,084

核心字段Key Word :

#people : 行经相同起讫路径的总数量(以“人”为单位)

图层意义Meanings : 本图层以“联通手机信令数据”推导出各通勤路径(起讫点的直线距离), 并将相同起讫路径进行人数统计作为该路径的流量。

# Road2009/2014/2016

## 道路2009/2014/2016



— 2009  
— 2014  
— 2016

图层类型File Type : Polyline

数据量Amount : 10,650 / 27,632 / 27,061

核心字段Key Word :

#NAME : 道路名称

#LY / ctype : 道路所属层级(高速、铁路、市区一级等)

图层意义Meanings : 各级道路的空间分布与基本信息。

# Street\_Harvey

## 街道形态指标

图层类型File Type : Polyline

数据量Amount : 5,109



核心字段Key Word :

#width : 道路宽度

#length : 道路长度

#var\_height : 建筑高度标准差

#var\_HbyA : 建筑高度标准差 ( 根据建筑面积做了加权 )

#continue\_left : 左侧连续度

#continue\_right : 右侧连续度

#bld\_den : 建筑个数密度

#KGB : 道路高宽比

#cross\_section\_right : 右侧高度与宽度比

#cross\_section\_left : 左侧高度与宽度比

#ave\_height\_left : 左侧建筑平均高度 ( 建筑面积进行了加权 )

#ave\_height\_right : 右侧建筑平均高度 ( 建筑面积进行了加权 )

图层意义Meanings : 基于Harvey等(2017)提出的方法, 利用街道两侧建筑物计算的街道形态指标



# Street\_WalkScore

## 街道步行指数



图层类型File Type : Polyline

数据量Amount : 5,050

核心字段Key Word : (详细计算过程参见下页)

#WalkScore : 步行指数, 该段街道之步行指数总得分

#function\_density : 功能密度, 该段街道之功能密度得分

#function\_mix : 功能混合度, 该段街道之功能混合度得分

#junction\_density : 道路交叉口密度, 该段街道之道路交叉口密度得分

图层意义Meanings : 利用街道网络和POI计算的各个街道的步行指数(Walk Score) · 详见周垠和龙瀛(2017)

周垠, 龙瀛. 街道步行指数的大规模评价——方法改进及其成都应用[J]. 上海城市规划, 2017 (1): 88-93.

# Street\_WalkScore

## 街道步行指数

核心字段计算公式 Equations :

1. WalkScore : 街道步行指数

$$Walk\ Score = \sum_{i=1, j=1}^{m, n} (w_i \times S_{i,j} \times DD_j) \times \frac{100}{15}$$

$w_i$ : 某类设施的影响权重  
 $i$ : 表示不同类型的设施  
 $j$ : 表示不同的步行距离  
 $S_{i,j}$ : 表示某类设施和步行距离的服务范围  
 $DD_j$ : 表示设施的距离衰减系数

2. function\_density : 街道功能(POI)密度

$$FUNCTION\_DENSITY = \frac{\text{number of POIs}}{\text{area (km}^2\text{)}}$$

3. function\_mix : 街道功能混合度，一个区域内功能设施的多样性

$$FUNCTION\_MIX = -\sum_{i=1}^n (p_i \times \ln p_i)$$

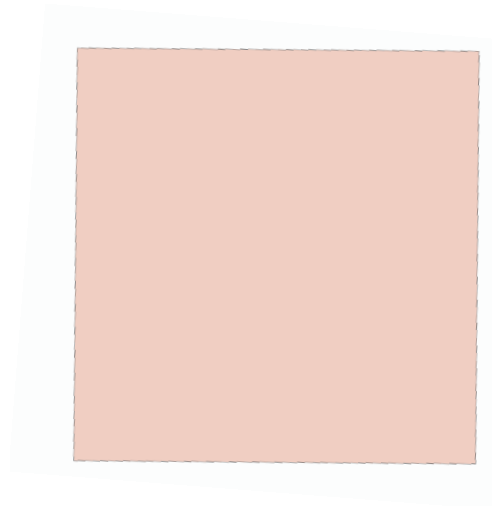
$p_i$ : 每个街道沿途 i 类型的设施占总体设施的比例

4. junction\_density : 道路交叉路口密度，一个区域内交叉路口数量，单位：#/km<sup>2</sup>

$$JUNCTION\_DEN = \frac{\text{number of junctions}}{0.5\text{km search radius}}$$

# Ext

## 研究范围



图层类型File Type : Polygon

数据量Amount : 1

图层意义Meanings : 北京二环内范围(旧城)

# Buildings2017

## 建筑物2017



图层类型File Type : Polygon

数据量Amount : 38,437

核心字段Key Word :

#FootPrint\_m2 : 建物底层面积

#FloorNumber : 建物层数

# FloorSpace\_m2 : 建物总楼面面积 (一般为底层面积\*层数)

图层意义Meanings : 说明北京内各个建筑物的位置、基底面积(底层面积)、层数与楼面面积 (原则上为底层面积乘以层数)。

# Parcels2011AICP

## 地块 2011(依据路网划分)



图层类型File Type : Polygon

数据量Amount : 5,533

图层意义Meanings : 此图层显示利用道路网络划分现状地块 ( 具体方法详见Liu and Long 2016 EPB\_AICP ) 。

# GeoGrids

## 地理编码



图层类型File Type : Polygon

数据量Amount : 2,122

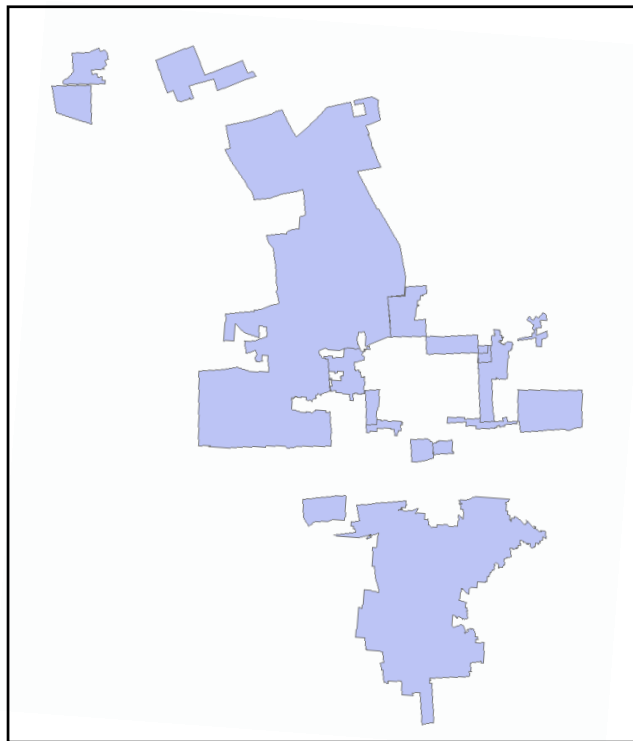
核心字段Key Word :

#GEOCODE : 地块编码

图层意义Meanings : 此为依照自然要素或道路划分之地理编码。

# FunctionalZones

## 功能园区



图层类型File Type : Polygon

数据量Amount : 15

核心字段Key Word :

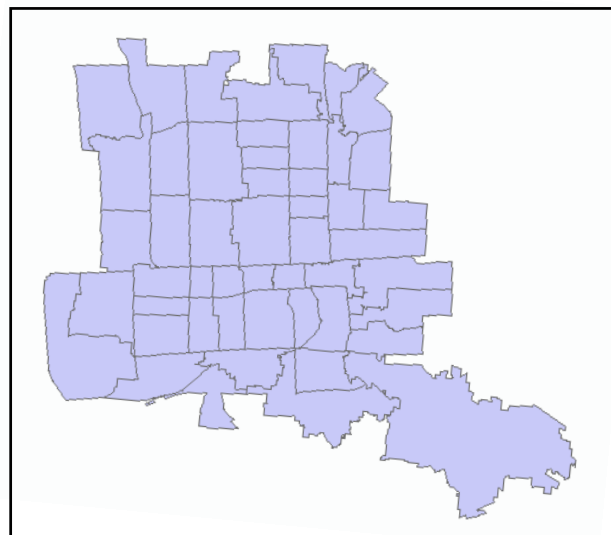
#园区名称 : 产业园区名称

#园区类别 : 园区所属类别 , 分为文化创意、商务金融、开发区

图层意义Meanings : 说明北京内各产业功能园区信息。

# Jiedaos

## 街道办事处



图层类型File Type : Polygon

数据量Amount : 50

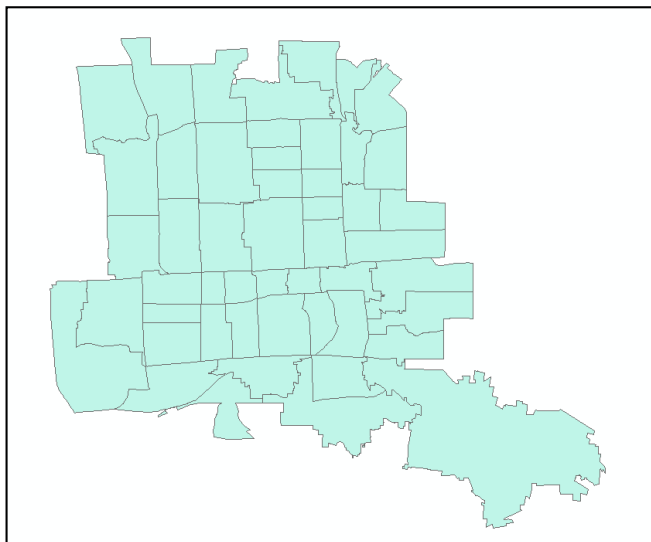
核心字段Key Word :  
#Name : 街道办事处名称

图层意义Meanings : 显示各街道办事处的名称和空间分布。



# TownshipPOP

## 街道办事处 人口信息



图层类型File Type : Polygon

数据量Amount : 50

核心字段Key Word :

#名称 : 街道办事处之名称

#各种人口统计数据 , 如 : 常住人口、户籍人口、外来人口等

图层意义Meanings : 记录不同年份各街道办事处的人口统计信息  
( 如 : 常住人口、户籍人口、外来人口等。 )

# Juweihui

## 居委会



图层类型File Type : Polygon

数据量Amount : 2,256

核心字段Key Word :

#XZQMC : 居委会的行政区名称

图层意义Meanings : 显示居委会的名称和空间分布。

# ResidentialCommunity

## 居住小区

图层类型File Type : Point

数据量Amount : 4,279

核心字段Key Word :

#NAME : 小区名称

#Area (km2) : 小区占地面积

#building\_a : 小区内建筑面积，每幢建物底层面积总和

#floor\_area : 小区内建筑楼面面积总和

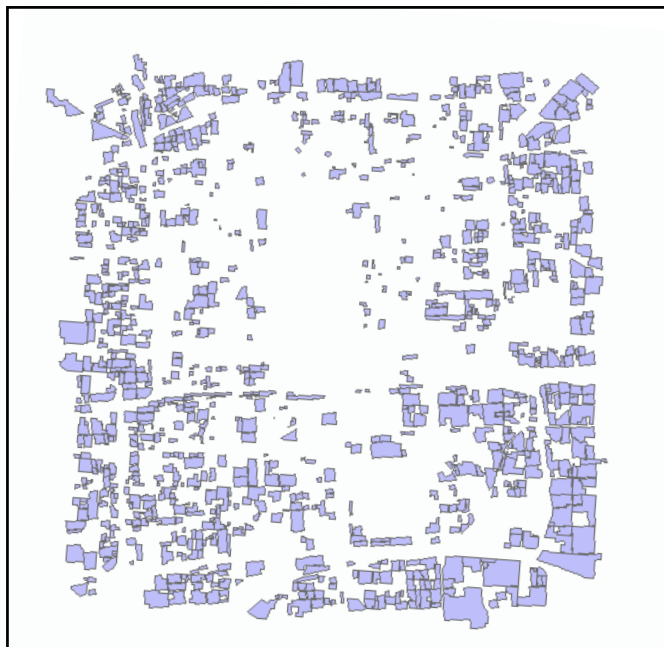
#plot\_area : 小区内容积率

#households : 小区内可容纳之户数

#price : 小区内住房平均价格

#property\_f : 小区内住房平均物价费用

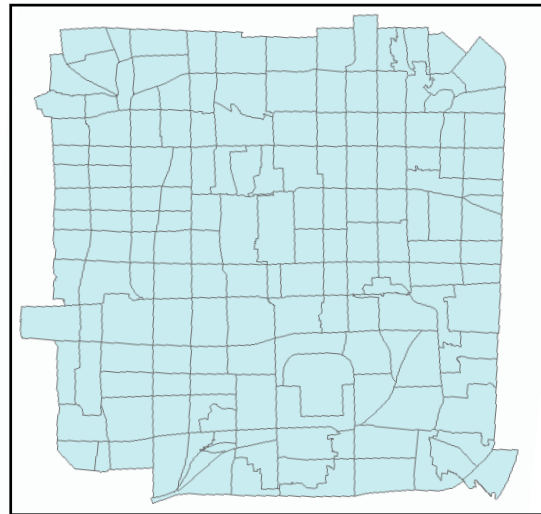
#临近设施状态，如：school, postoffice, mall等，说明邻近的设施名称或是当前使用状况



图层意义Meanings : 显示居住小区的具体边界、基本物质空间与周遭临近设施。

# TAZs

交通分析地块 (依据北京市交通管理局所划分)



图层类型File Type : Polygon

数据量Amount : 167

核心字段Key Word :

#District : 所属行政区

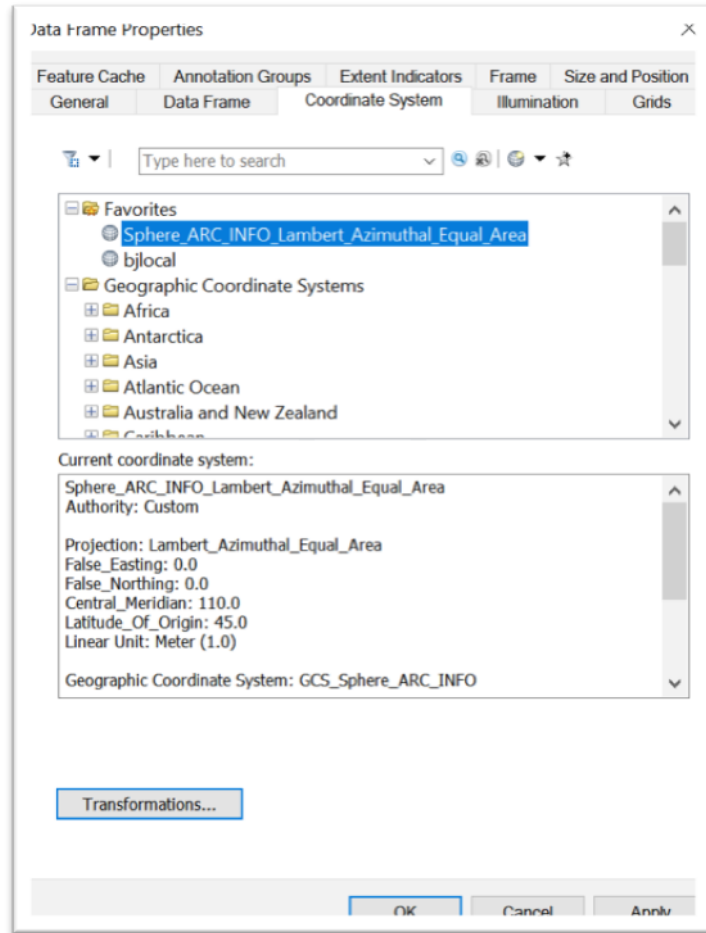
#Street : 所属街道办

图层意义Meanings : 北京市交通管理局所划分的交通分析地块。

# 基本设置

---

# 基本设置



ArcMAP 坐标系统建议使用：GCS\_WGS\_1984  
(设置路径：View / Data Frame Properties / Coordinate System)

使用 Sphere\_ARC\_INFO\_Lambert\_Azimuthal\_Equal\_Area.prj  
可得出长度与面积单位

# 相关阅读材料

---

# 相关阅读材料

---

若使用本讲数据进行学术研究，建议参考、引用下列文章：

- 龙瀛, 郎崑. 新数据环境下的中国人居环境研究[J]. 城市与区域规划研究, 2016, 8(1):10-32.
- 龙瀛, 刘伦. 新数据环境下定量城市研究的四个变革[J]. 国际城市规划, 2017, 32(1):64-73.
- 龙瀛, 沈尧. 数据增强设计——新数据环境下的规划设计回应与改变[J]. 上海城市规划, 2015(2):81-87.



---

# 本讲结束

---