



北京城市实验室
Beijing City Lab

ID of the slides

34



Slides of BCL

www.beijingscitylab.com

How to cite

Author(s), Year, Title, Slides at Beijing City Lab, <http://www.beijingscitylab.com>

E.g. Long Y, 2014, Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest, Slides at Beijing City Lab, <http://www.beijingscitylab.com>

开放数据与定量城市研究

李栋 博士
中国城市规划设计研究院

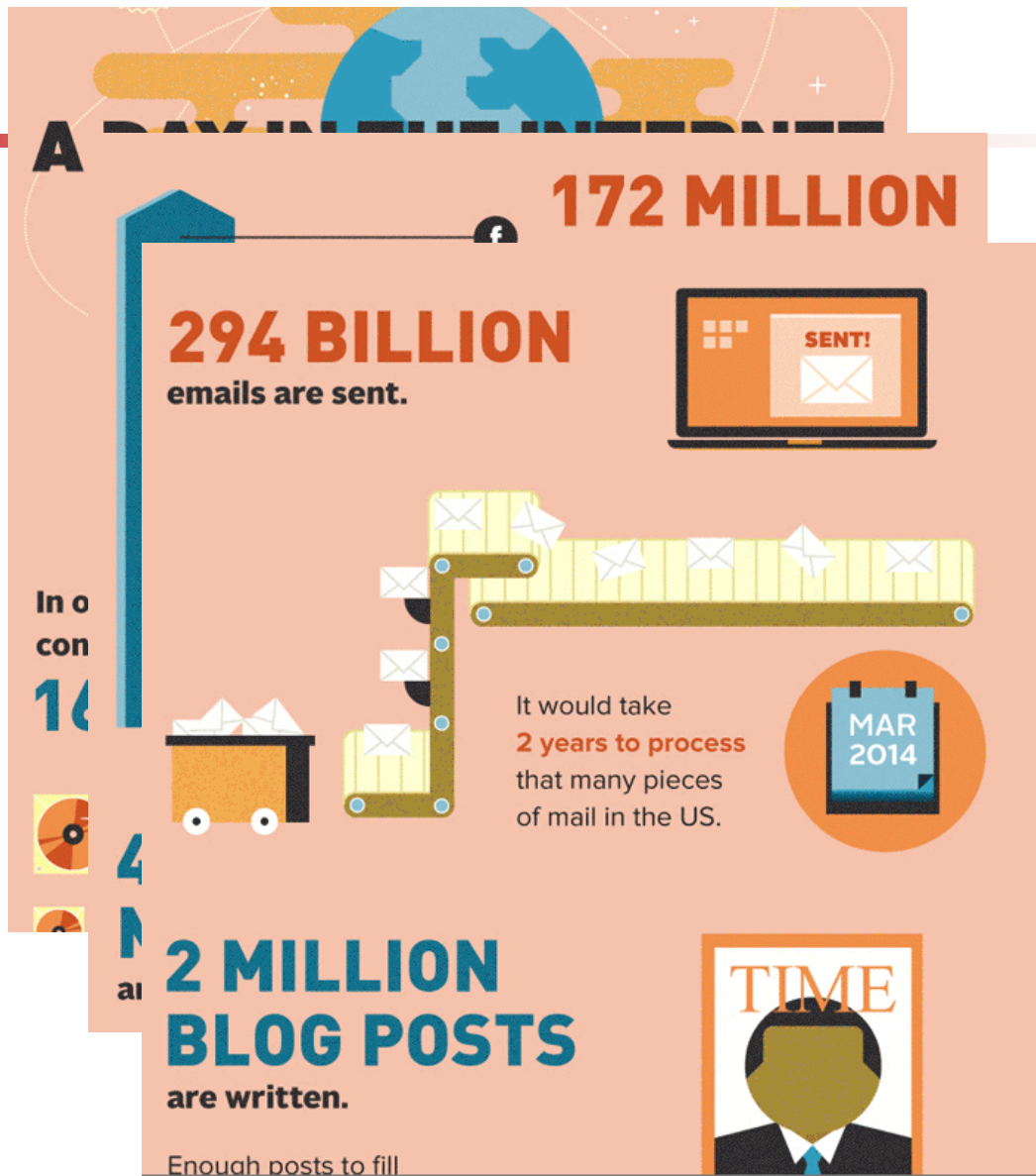
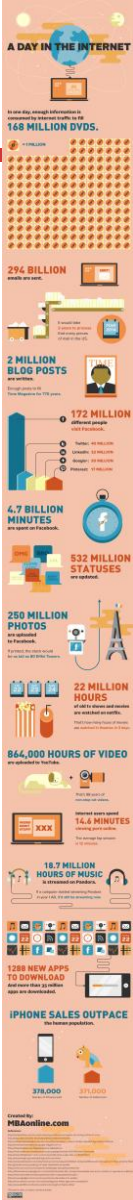
2014年10月

交流内容

1. 概念
2. 方向
3. 应用
4. 思考

1. 概念

- 大数据的样子：因特网的一天
 - 互联网一天信息量等于1.68亿张DVD
 - 电子邮件：1940亿封
 - 博客文章：200万篇
 - 社交网络：1.72亿人访问Facebook，4000万人访问Twitter.....
 - 6000万活跃用户、5000万次手机访问微博



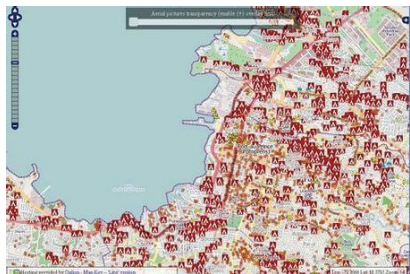
1. 概念

• 技术背景

– 2000年以来IT及互联网技术的发展：

- 搜索引擎——社交网站（UGC）——数据中心
- 智能手机——APP——传感器（物联网）

} 云计算——大数据



OpenStreetMap, OSM



公众环境研究中心
中国水污染地图



墨迹空气果

1. 概念

- 大数据的4V特征

- **Volume**（大量）

- Variety（多样）

- Velocity（高速）

- **Value**（价值）

- Veracity（真实性）

- Visualization（可视化）

- Validity（有效性）

- Volatility（易变性）

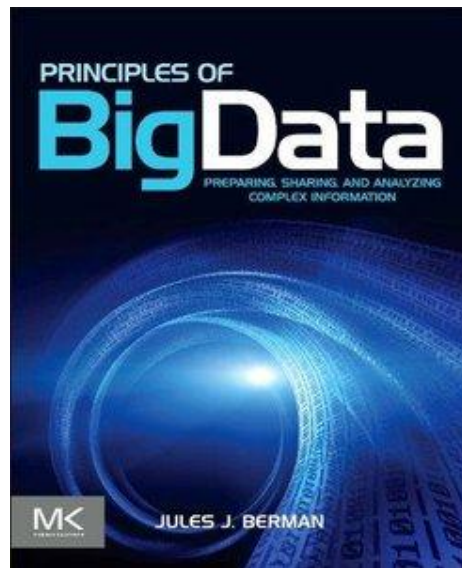
- Variability（变异性）

- Vision（想象力）

- Verbalizer（描述性）

1. 概念

- 另一种定义：与小数据（传统数据）的差异
 - 目标（Goals）
 - 位置（Location）
 - 数据结构及内容（Data structure and content）
 - 数据准备（Data preparation）
 - 数据生命周期（Longevity）
 - 测度（Measurements）
 - 可重复性（Reproducibility）
 - 成本（Stakes）
 - 内省（Introspection）
 - 分析（Analysis）



Dr. Jules J. Berman
作者背景

1. 概念

- 开放数据：可获取数据的粒度、精度在不断提高

- 传统数据：

- 各类普查、统计年鉴、公报、现场问卷调查等
 - 时间粒度：年
 - 空间粒度：行政单元
 - 样本规模：1%抽样（人口普查）或更低

- 开放数据、新型数据：

- 时间粒度：实时（秒）；空间粒度：经纬度（米）
 - 样本规模 >> 传统调查和普查



1. 概念

• 政府开放数据

- 信息公开：文档
- 统计数据
- 部门数据
- 开放数据



1. 概念

• 商业开放数据

- 电子地图
- 社交网络
- 电子商务
-



JIOS应用接入



1. 概念

• 科研开放数据

- 科学数据共享网站
- 数据堂



1. 概念

- Application Programming Interface (API) , 应用程序编程接口
 - 访问特定网址返回格式化的数据
 - 初衷: 一定程度共享网站拥有的数据, 鼓励开发者贡献内容、繁荣应用生态圈、实现商业价值
 - 促使: 新的科研数据来源

```
{
  - statuses: [
    - {
      created_at: "Sun Jan 05 23:49:20 +0800 2014",
      id: "3663529498434330",
      mid: "366351",
      idstr: "366351",
      text: "雾霾",
      in_reply_to: "http://t.cn/R...",
      source: "<a href='\"#\"'>...",
      favorited: false,
      truncated: false,
      in_reply_to_status_id: null,
      in_reply_to_status_id_str: null,
      in_reply_to_user_id: null,
      in_reply_to_user_id_str: null,
      pic_ids: [
        "a6133a...",
      ],
      thumbnail_pic: null,
      bmiddle_pic: null,
      original_pic: null,
      geo: {
        type: "P",
        - coordin
          25.
          110
        ],
      },
      user: {
        id: "278...",
      },
    },
  ],
  - {
    aqi: 270,
    area: "天津",
    co: 1.402,
    co_24h: 1.072,
    no2: 58,
    no2_24h: 45,
    o3: 1,
    o3_24h: 134,
    o3_8h: 9,
    o3_8h_24h: 108,
    pm10: 306,
    pm10_24h: 273,
    pm2_5: 220,
    pm2_5_24h: 179,
    position_name: "市监测中心",
    primary_pollutant: "细颗粒物(PM2.5)",
    quality: "重度污染",
    so2: 31,
    so2_24h: 23,
    station_code: "1013A",
    time_point: "2014-10-08T08:00:00Z"
  },
  - {
    aqi: 263,
    area: "天津",
    co: 1.696,
    co_24h: 1.631,
    no2: 64,
    no2_24h: 60,
    o3: 10,
    o3_24h: 187,
  },
}
```

www.pm25.in

1. 概念

- Application Programming Interface (API) , 应用程序编程接口

- 访问特定网址返回格式化的数据
- 初衷：一定程度共享网站拥有的数据，鼓励开发者贡献内容、繁荣应用生态圈、实现商业价值
- 促使：新的科研数据来源

数据源

- 政府
- 企业
- 科研
-

API



用户

- 采集
- 分析
- APP
- 反馈
-

2. 方向

- 数据科学 (Emerging Data Sciences) : 定量研究新趋势
 - 经典: 流体力学 (Fluid Dynamics)
 - 传统: 计算流体力学 (**Computational** Fluid Dynamics)
 - 新兴: 计算社会学 (**Computational** Sociology)、计算考古学 (**Computational** Archaeology)、计算语言学 (**Computational** Linguistics)、计算生物学 (**Computational** Biology)、计算化学 (**Computational** Chemistry)、计算经济学 (**Computational** Economics)、计算流行病学 (**Computational** Epidemiology)

2. 方向

- 数据科学：采用计算机科学方法和思路去解决相关学科的问题
 - 缺少计算机运算能力就无法实现分析的方法： *...methods that could not realistically be performed without the aid of a computer...*
 - 算法与模型：模拟、预测、人工智能、复杂统计.....
 - 学科数据，行业应用

2. 方向

• 定量城市研究

- 北京: Beijing City Lab (BCL)
- MIT: SENSEable City Lab
- ETH: Future Cities Laboratory (FCL) 新加坡

• 城市化地区

- 人类活动
- 数据密度
- 城市系统分析和调控的复杂性



Home

Projects

Members

Working papers

Slides

Data released

Ranking

Blogs

Links&Partners

About

The Beijing City Lab (BCL) is a virtual research community, dedicated to studying, but not limited to, China's capital Beijing. The Lab focuses on employing interdisciplinary methods to quantify urban dynamics, generating new insights for urban planning and governance, and ultimately producing the science of cities required for sustainable urban development. The lab's current mix of planners, architects, geographers, economists, and policy analysts lends unique research strength.

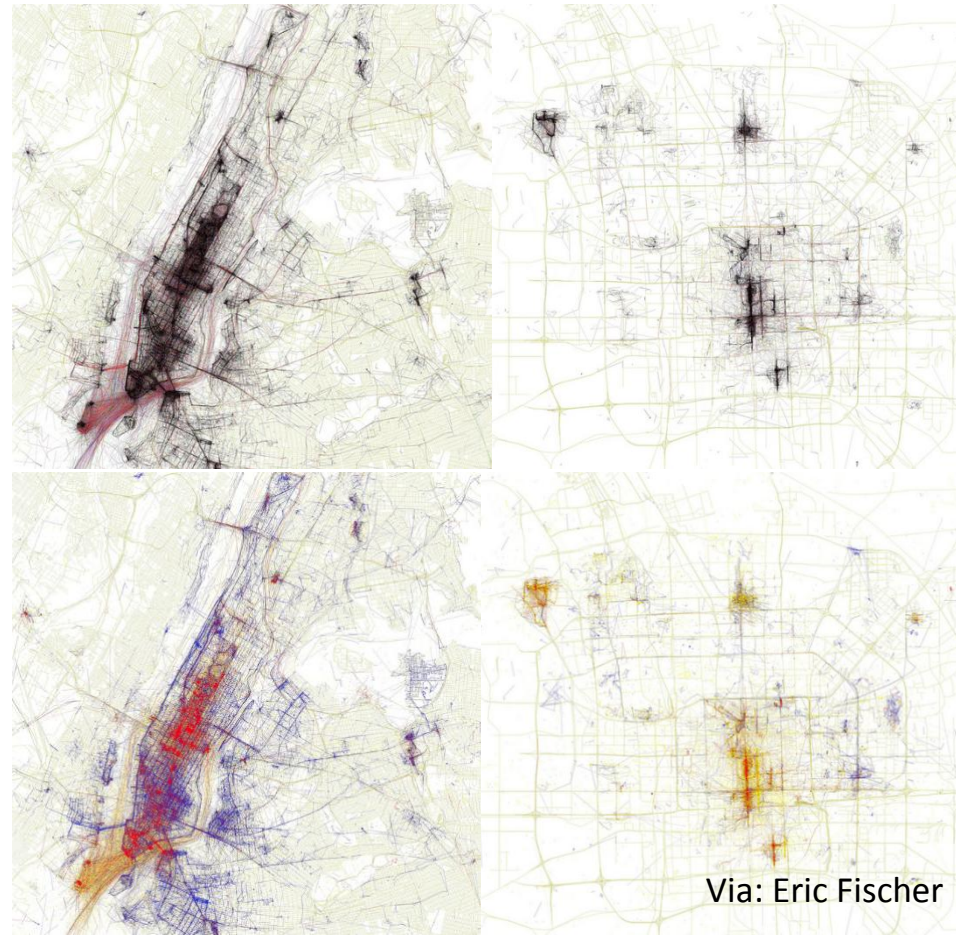


www.beijingscitylab.com



2. 方向

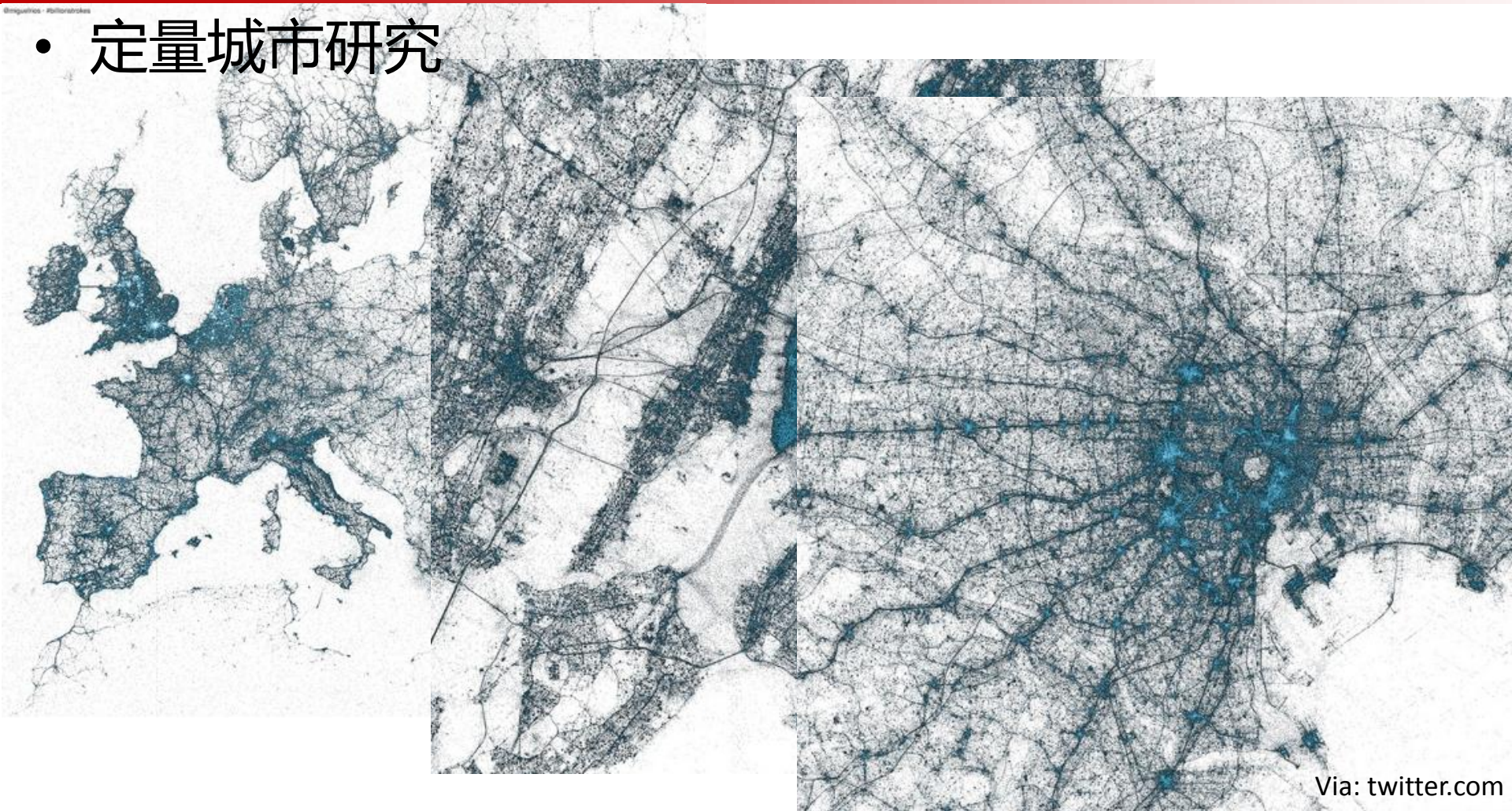
- 定量城市研究



Via: Eric Fischer

2. 方向

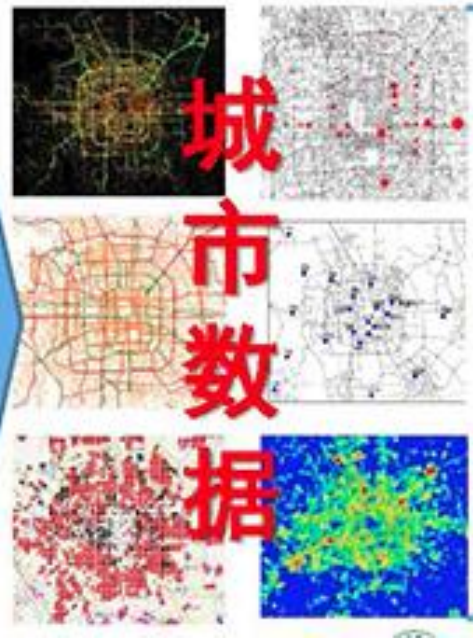
- 定量城市研究



2. 方向

- 定量城市研究
 - 城市数据类型

信息基础设施

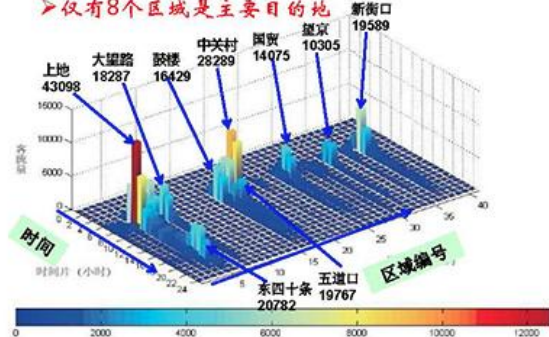


多源市民活动数据



每天天通苑住户都去哪儿了？

- 住户主要通往市38个区域
- 仅有8个区域是主要目的地



2. 方向

• BCL的几项工作

- Projects
- 1 BUDEM
- 2 Urban Growth Boundaries
- 3 Bus Landscapes
- 4 Population China
- 5 Planning Support Systems
- 6 Urban Form
- 7 Population Synthesis
- 8 Social Network Mining
- 9 Big Model
- 10 Beijing Parking
- 11 Urban Network Analysis
- 12 AM10:00

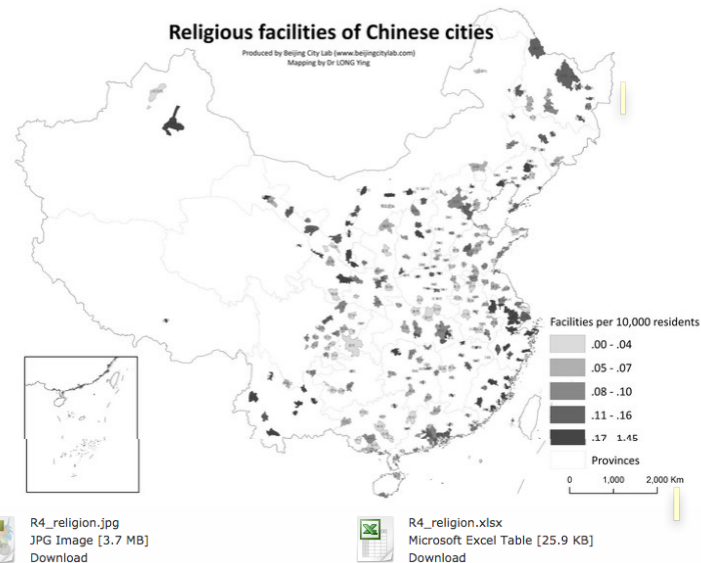


- 覆盖全国所有城市的地块尺度的城市增长模型
- 利用约束性元胞自动机方法 (vector CA)
- 654个城市, 76万个城市地块
- Simulating urban expansion at the parcel level for a
- <http://arxiv.org/abs/1402.3718>

4 Religious facilities in Chinese cities

LONG Ying, 2014

From a dataset of national POIs, we gathered 5,237 religious facilities (like temples, churches and mosques) for various religions of 287 prefectural or above cities in China. We then ranked these cities in terms of facility number per 10,000 people.



3 Mobile penetration of Chinese cities

LI Dong, 2014

We inferred the activity level of mobile internet users in Chinese cities, based on information about geotagged micro-blog messages (aka Weibo) for the first half of 2014. Data were collected and counted via Weibo API. Cities are ranked based on per capita geotagged weibos, reflecting the mobile penetration. Overall, the amount of geotagged weibos are positively correlated with total population in individual cities. We also notice large numbers of weibos per capita for tourist towns (e.g., Lijiang) and cities with relatively high livability (e.g., Xiamen and Chengdu).



2. 方向

• 新城市科学

– 驱动力：技术工具的革新

“老”城市科学	“新”城市科学
粗略的“区域科学”：城市经济学、社会物理学、中心地理论、交通工程学	城市复杂性理论和网络科学
静态的、截面的、系统论的视角	演进的、复杂科学的视角：离散性、“自下而上”
50年前创立：集合层面的模型（aggregate）	20年以来的新技术和新工具：分解的、离散的、个体层面的模型
调查、统计数据；较低计算需求	数据可获得性与计算机计算能力不再是主要问题
城市发展的概况	城市发展的细节

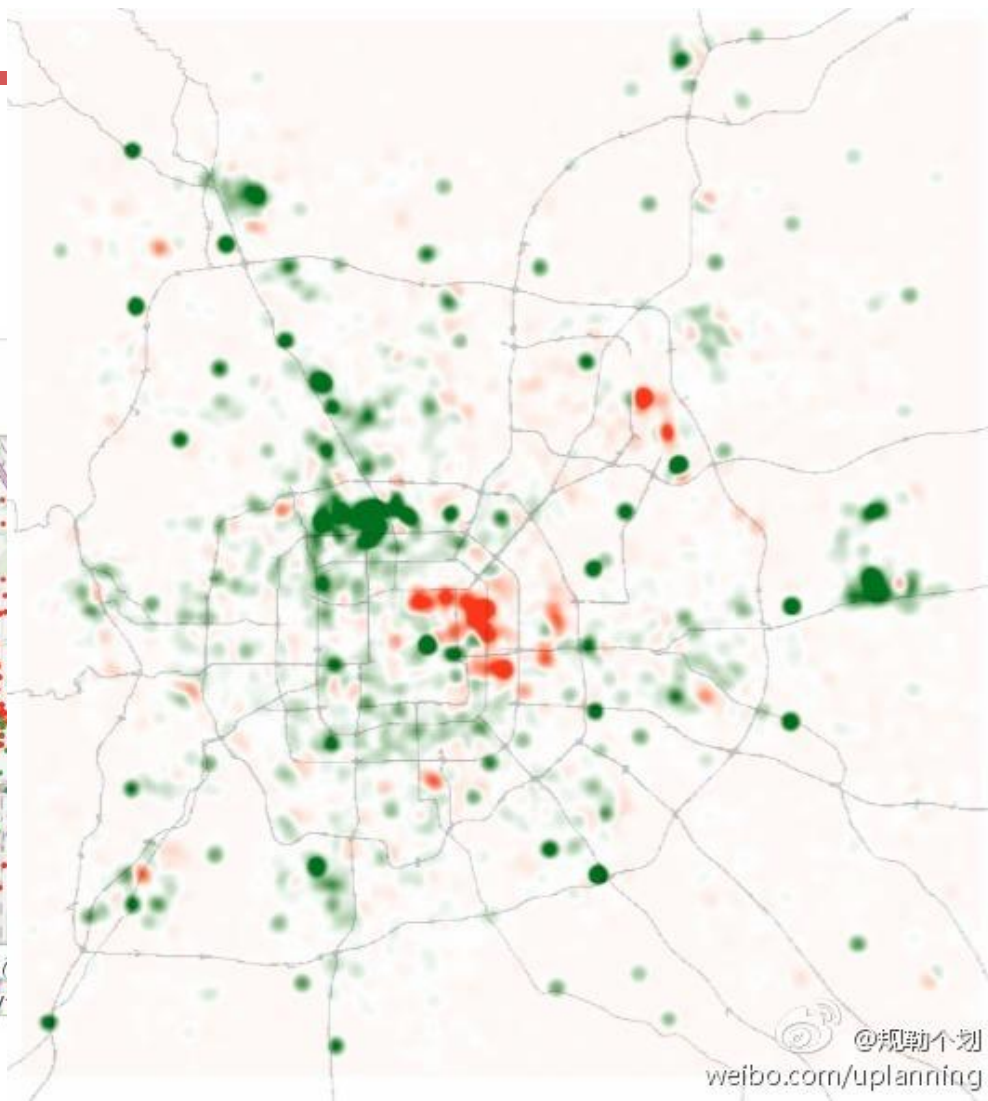


THE NEW SCIENCE
OF CITIES

MICHAEL BATTY

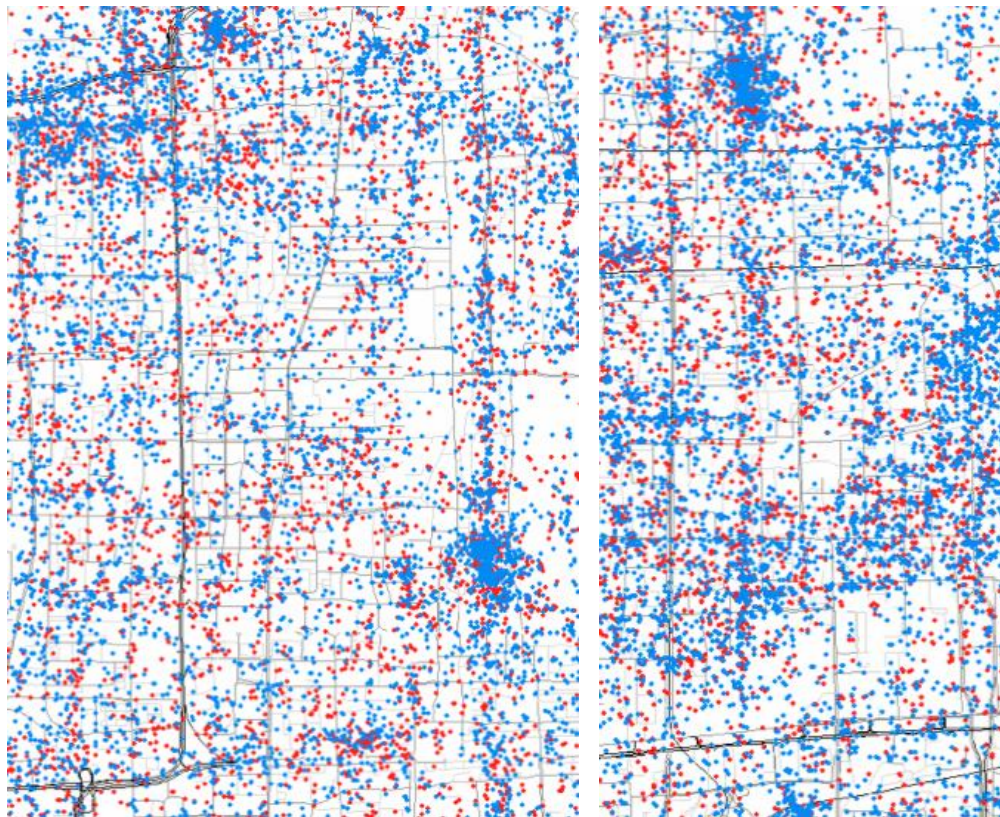
3. 应用

- LBS数据 (e.g.微博)



3. 应用

- LBS数据 (e.g.微博)



```
- statuses: [
  - {
    created_at: "Wed Oct 08 10:03:17 +0800 2014",
    id: "3763340666590129",
    mid: "3763340666590129",
    idstr: "3763340666590129",
    text: "→_→ http://t.cn/z82PWU5",
    source: "<a href='\"http://app.weibo.com/t/feed/c66T5g\"' rel='\"nofollow\"'>Android</a>",
    source_type: 1,
    favorited: false,
    truncated: false,
    in_reply_to_status_id: "",
    in_reply_to_user_id: "",
    in_reply_to_screen_name: "",
    pic_ids: [],
    - geo: {
      type: "Point",
      - coordinates: [
        39.99542,
        116.30307
      ]
    },
    - user: {
      id: "5070466383",
      idstr: "5070466383",
      class: 1,
      screen_name: "薛怀璧杉杉",
      name: "薛怀璧杉杉",
      province: "11",
      city: "1000",
      location: "北京",
      description: "",
      url: "",
      profile_image_url: "http://tp4.sinaimg.cn/5070466383/50/5704712371/0",
      cover_image_phone:
      "http://ww1.sinaimg.cn/crop.0.0.640.640/6cf8d7ebjw1ehfr4xa8psj20hs0hsgp",
      profile_url: "u/5070466383",
      domain: "",
```


3. 应用

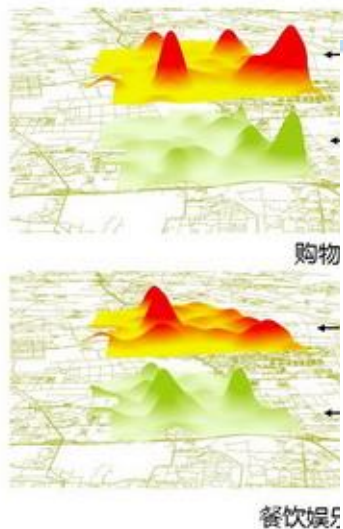
- LBS数据 (e.g.微博)



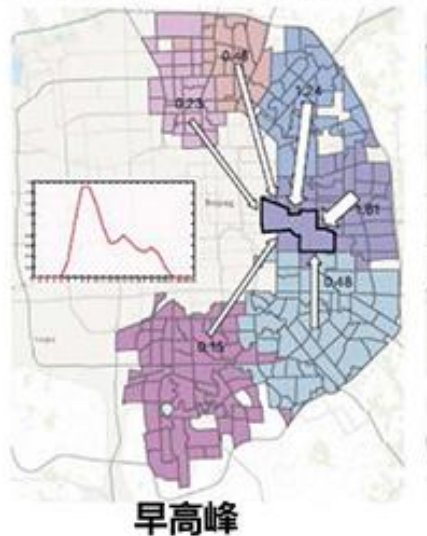
3. 应用

- 城市活动与联系：活动、交通、信息、.....

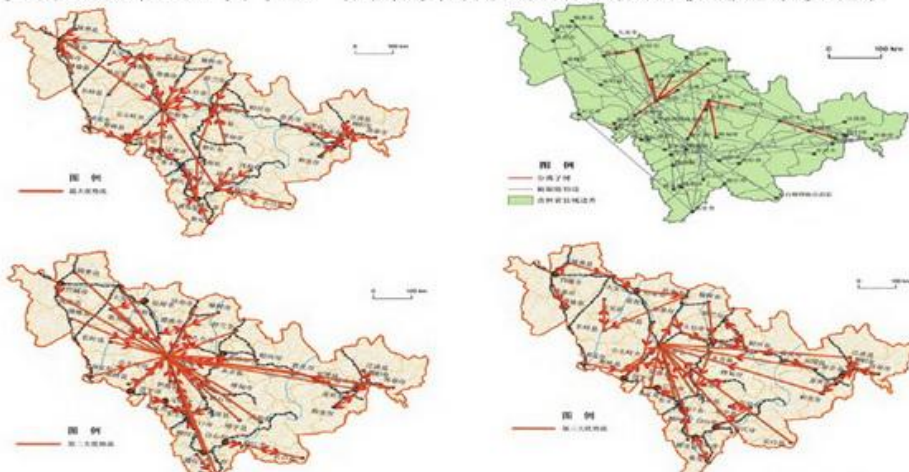
居民活动、城市设施的核密度分析（基于GPS、POI数据）



时空主成分的相关性分析（以国贸为例）

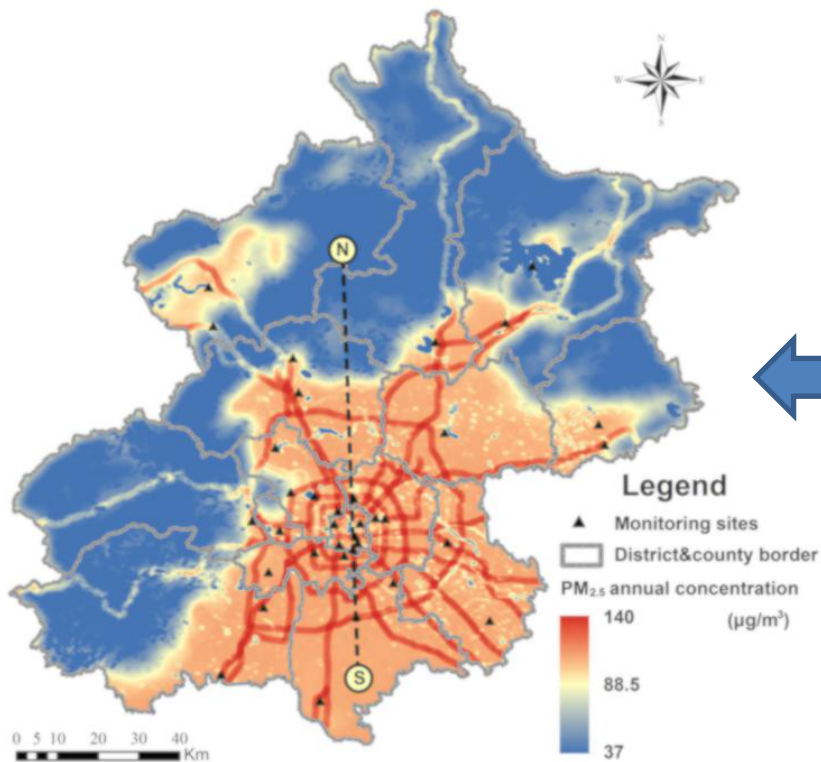
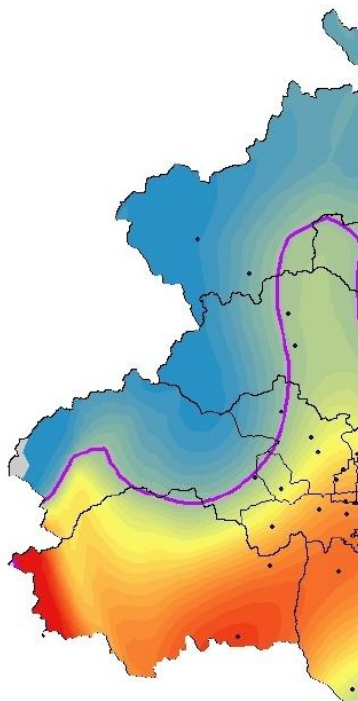
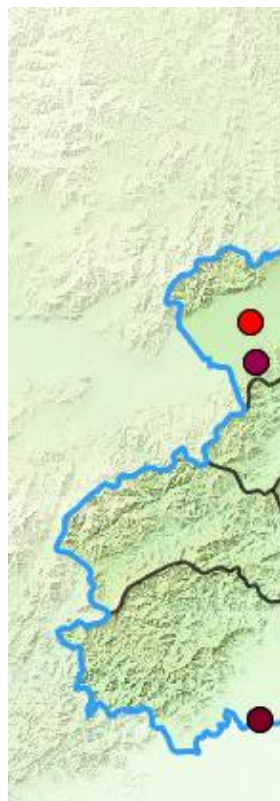


基于通信流的城市关联系分析，把握城市网络格局变化过程和特征，为城镇体系规划提供技术支持



3. 应用

- 基于土地利用回归 (Land Use Regression, LUR) 的大气模型及其精细化



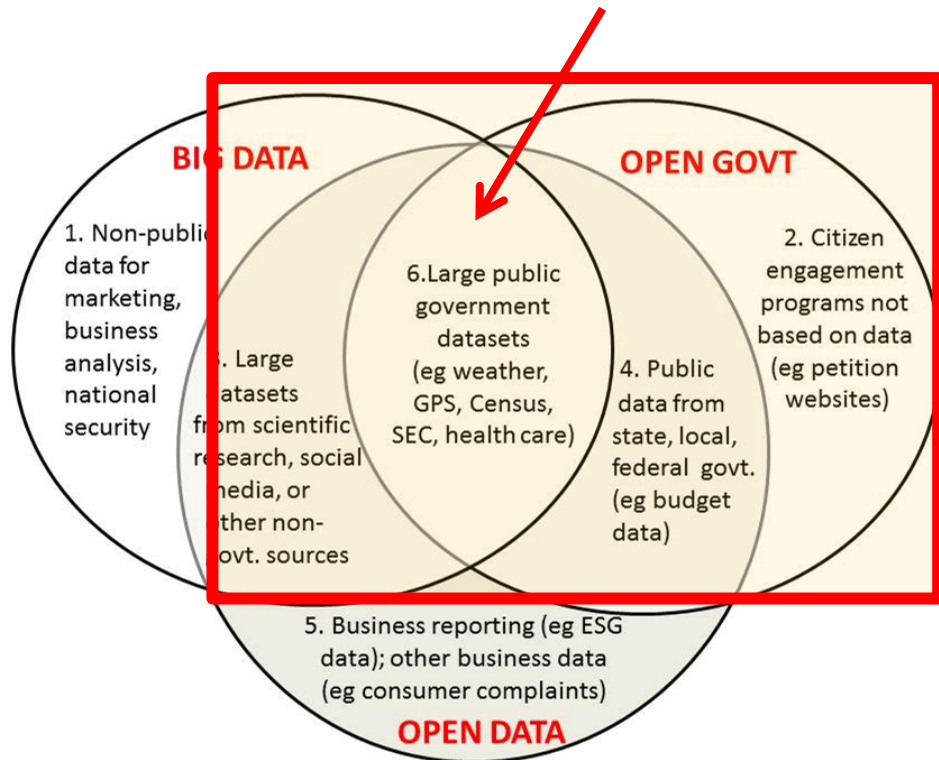
- 气象
- 交通
- 居民活动
-

4. 思考

- 开放数据等新型数据的一个重要特征：源于大众、为了大众

- 新型城镇化倡导以人为本，但如何理解“人”
- 规划中合理使用新型数据，有助精细化的分析城市，使方案制定更加科学，基础设施、公共服务设施的布局更加有效、公众参与更加便捷

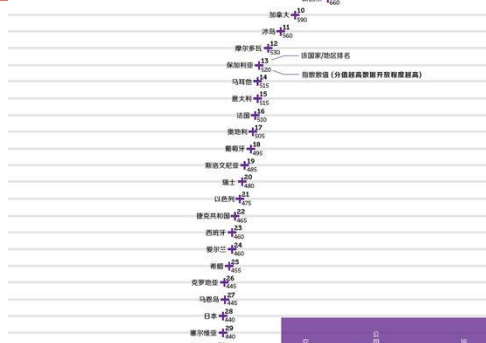
规划行业可挖掘的领域：大数据与开放型政府的交集



4. 思考

2013各国/地区政府 开放数据指数

網易新聞 DATA 数读
数据来源: OPEN DATA INDEX



• 数据开放：开放型政府的重要特征

— 数据开放的标准化、系统化

- 美国政府: data.gov
- 英国政府: data.gov.uk
- 世界银行: data.worldbank.org

— 世行开放数据研究成果：概念、案例、政策、应用.....

1. 交通时刻表
2. 政府预算
3. 政府支出
4. 选举结果

data.worldbank.org/about/open-government-data-toolkit/knowledge-repository

THE WORLD BANK
Working for a World Free of Poverty

English Español Français العربية Русский 中文

Home About Data Research Learning News Projects & Operations Publications Countries Topics

Data

By Country By Topic Indicators Data Catalog Microdata Initiatives What's New Support Products

Open Data Essentials

SHARE

This section of the toolkit provides a summary of the essential elements of open government data, starting with its definition and covering issues from open data portals, applications that use open data, the benefits of open data, open data policy declarations, learning resources and technical assistance resources for open data. The toolkit provides links to many examples of these.

What is Open Data?

Data is open if it satisfies both conditions below:

- **Technically open:** available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application
- **Legally open:** explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions. Below you can find some examples of terms of use in data licensing:

- UK Open Government License
- The World Bank Terms of Use

Open Government Data Toolkit

- Open Data Essentials
- What is Open Data?
- Examples of Open Data Portals
- Open Data Benefits
- Open Data Applications
- Open Data Policies
- Open Data Learning Resources
- Technical Assistance for Open Data
- Technology Options
- Demand and Engagement

4. 思考

• 小结

1. **IT等部分产业：**大数据已经成为基础业务
2. **全社会：**数据意识和数据思维催生数据科学
3. **科学研究：**通过API等方式产生的标准化开放数据，促使数据科学、可计算科学等定量研究快速发展
4. **规划咨询行业：**精细化的开放数据完全契合“以人为本”的发展思路
 - 理念与工具的准备，迎接开放数据、大数据的到来

致谢

- 请多提宝贵意见！
 - 规勒个划@weibo