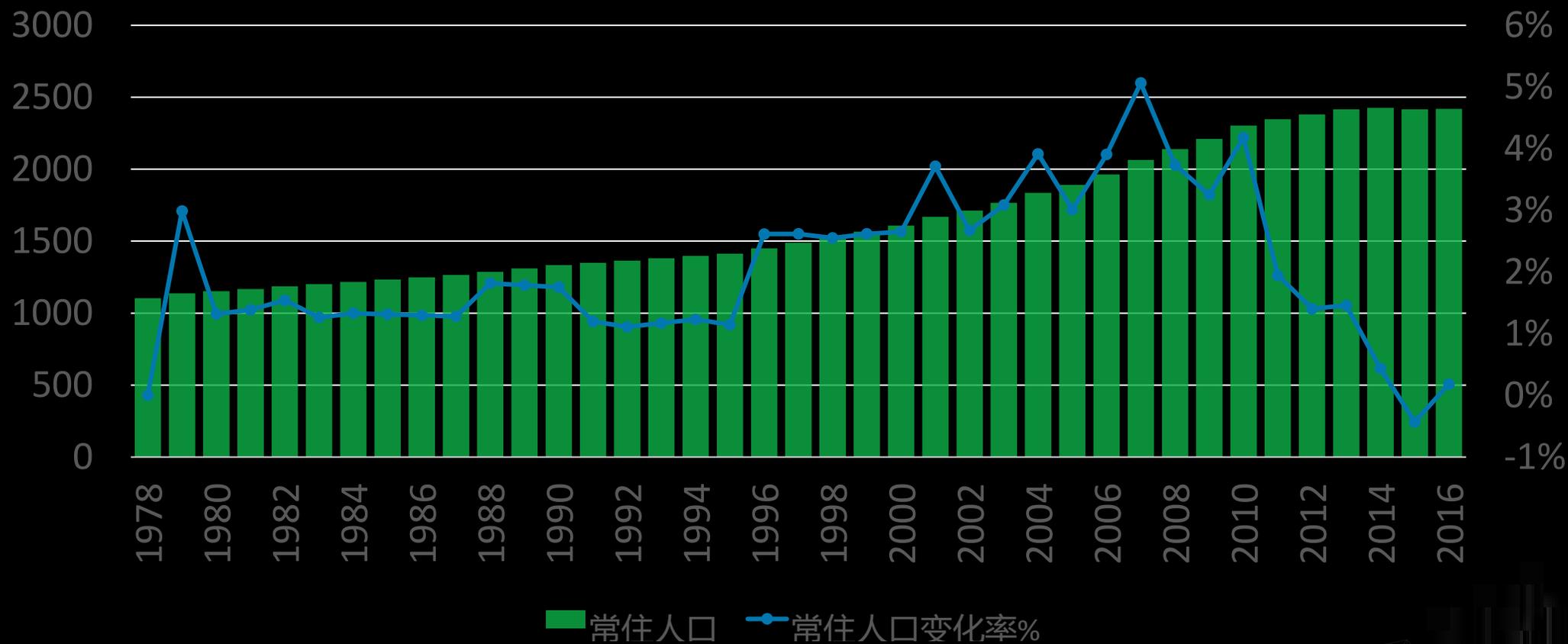


上海这座城市到有多少 多少人？

主讲人：汤舸



1978年-2016年上海市常住人口总量变化趋势



统计局 v.s. 公安局

常住人口

=

户籍人口

+

外来人口

*上海户籍，且常住在浦东区内

*非上海户籍，且在本地居住超过6个月以上，普查年具体操作按“离开户籍地半年以上”为准

≈ 户籍人口

实有人口

=

户籍人员

*户籍地在浦东

+

人在户不在

*上海户籍，且调查时发现人户分离

-

户在人不在

*户籍地在浦东，但调查时发现人户分离

+

实有来沪人员

*非上海户籍，原则上来沪三天就被计入

+

境外人员

*在浦东区居住超过六个月的外籍人员

统计局 v.s. 公安局

普查数据为基本，
年度据基本资料估算年度平均值

全市 每10年
人口普查 + 1%抽样调查

全市 年度
1.5% 人口变动抽样调查

各区 半年度
区县数据评估报告

各部门 持续
集合大数据判断趋势

全市 年末
汇总区县委办，综合统筹

实时更新，月度报告，年底考核

各居村委 月度
人户分离情况

综合协管队员 月度
来沪人员情况

居村委干部 和 综合协管队员
信息交流

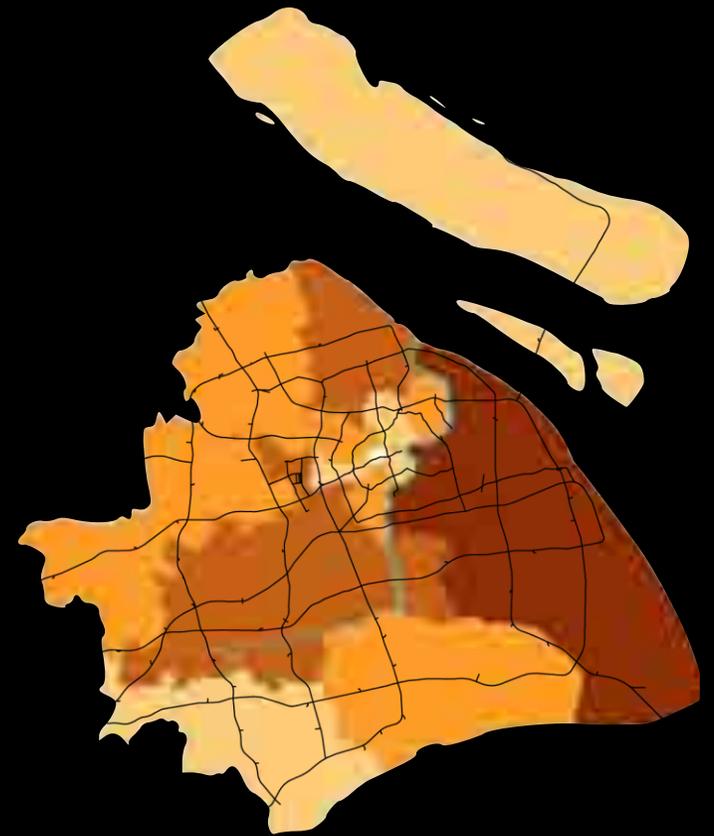
全市 每年10月
市人口办对各区人口登记率
和注销率情况进行抽查考核



2000年



2010年



2015年





上海每日实时人口规模推算

实时人口推算结果：

上海非春节期间实时人口区间值：

2700-3000万

平均值：**2870万**

(**300-500万短期逗留人口**)

不占用居住、教育等资源

可能占用交通、商业资源

上海春节期间实际人口最低值：

1500万



上海总人口

大尺度

- ❖ 集成：将多种“好而不同”的个体结果集成，从而能获得更接近真实的结果。
- ❖ 基于此的假设：在确保集成个体质量的条件下，集成越多个体的集成结果越能接近实际结果

如何定义“好”： 如何定义“不同”：

- 相关性系数
- 均值检验
- 聚类
- 来源不同？
- 类型不同？

分配比例

街镇人口

集成：

- 比例均值
- 比例中位数
- 比例加权平均

运营商数据：

- 联通
- 移动
- 电信

数据源检验：

- 异常数据源筛选：
- 相关性系数太低
 - 均值检验显著不同值

官方数据：

- 年鉴
- 六普数据

验证：
计算街镇人口与公安人口差值

通过

未通过

栅格人口 (X*X)

集成：

- 用地人口均值
- 加权平均

信令比例分配：

- 运营商数据

用地比例分配：

- 建筑容量、用地类型

验证

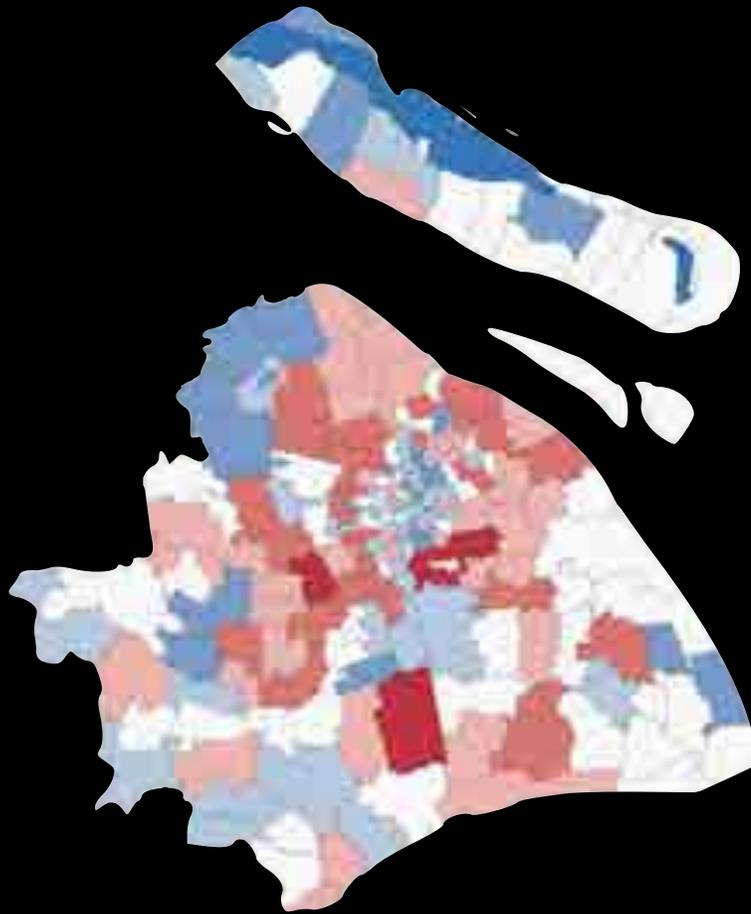
未通过

通过

验证步骤：以小区围栏为单元，按小区户数计算户均人口，若户均人口在1-6人范围内，则认定正常，否则为不正常，计算不正常小区的占比。

中尺度

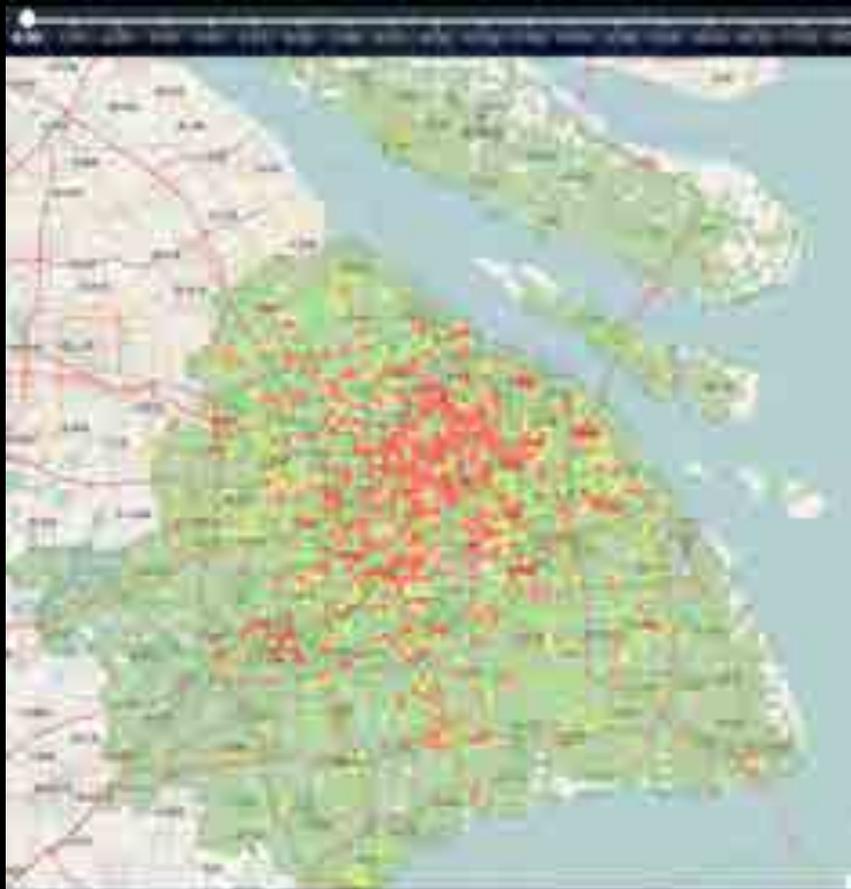
小尺度



官方统计数据

数据基底层





数据修正层1

数据基底层

腾讯定位大数据

小时/500米



数据修正层2

数据修正层1

数据基层层

Talkingdata定位大数据

月度/10米



房地产交易中心新房成交量

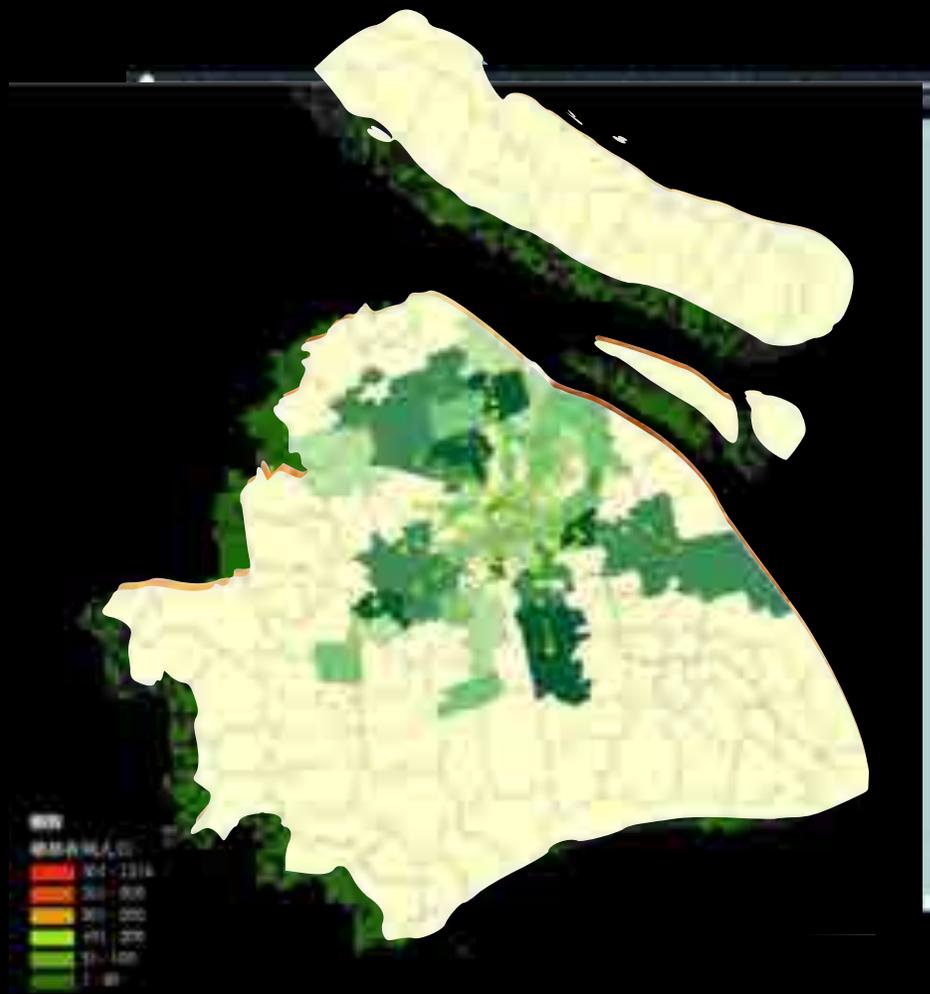
天/GPS

数据验证层1

数据修正层2

数据修正层1

数据基层层



轨道交通早高峰进站人数

天/GPS

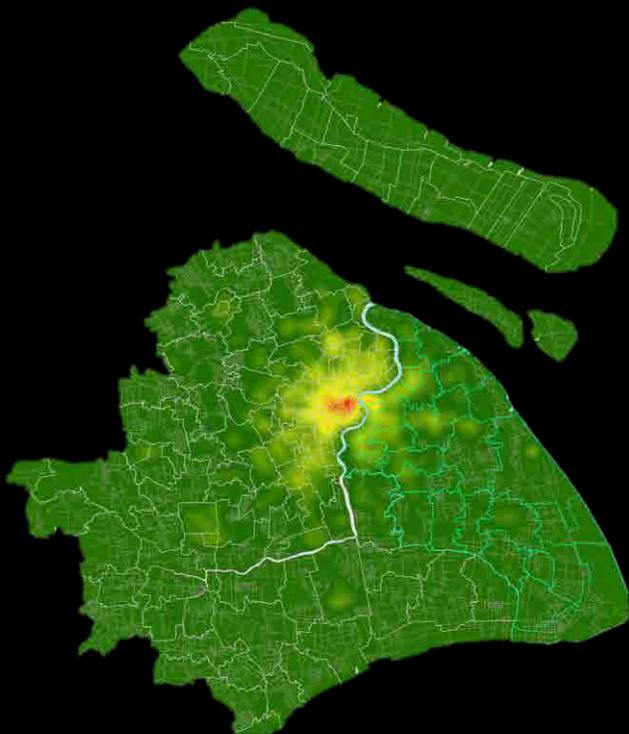
数据验证层2

数据验证层1

数据修正层2

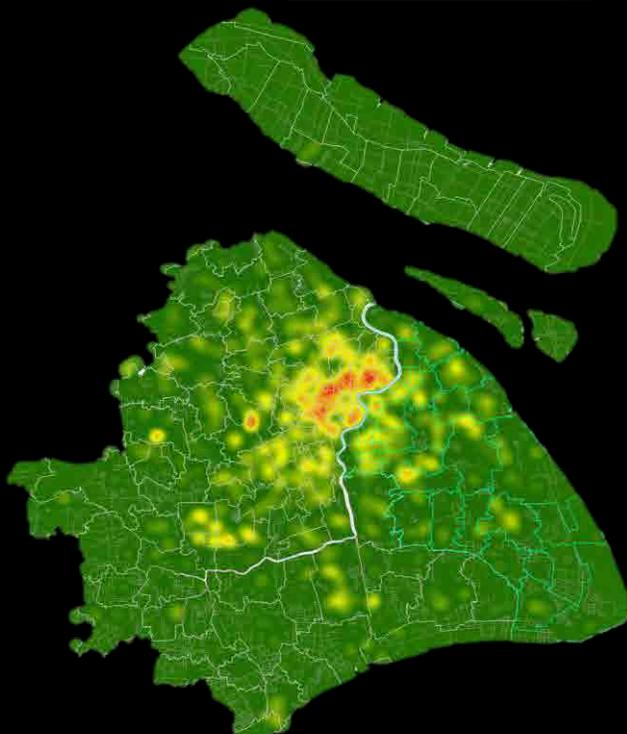
数据修正层1

数据基层层



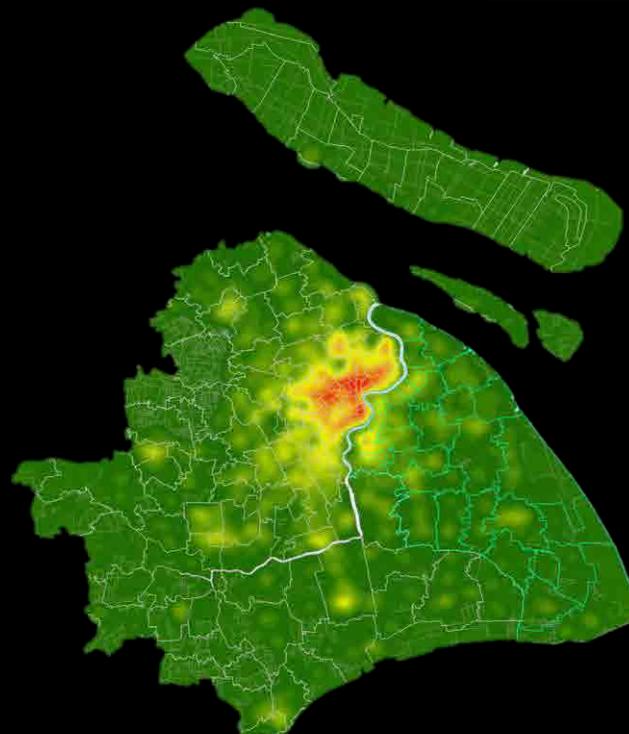
Talkingdata手机位置数据

获取时间：2016下半年 样本量：1000万+



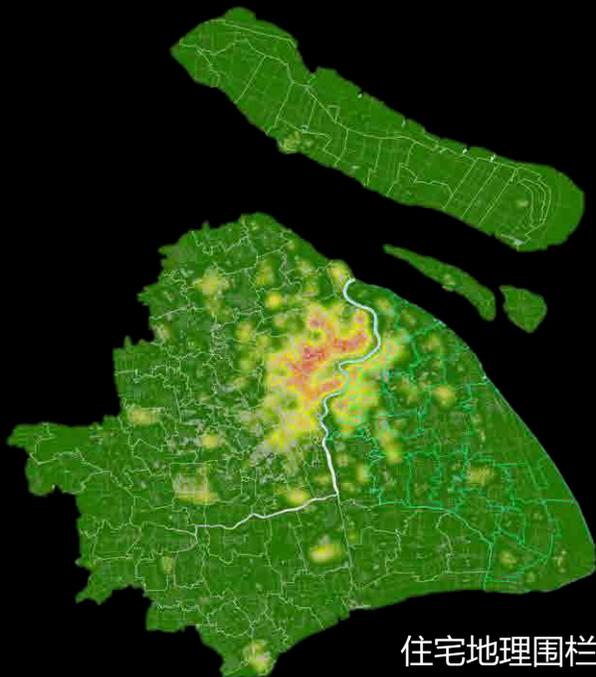
腾讯宜出行数据

获取时间：2017年3月 样本量：200万+



运营商信令数据

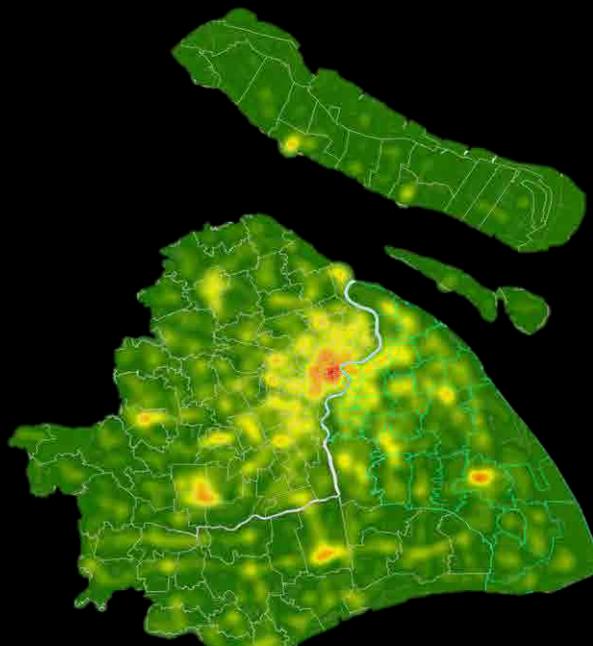
获取时间：2016下半年 样本量：1000万+



住宅地理围栏

在线地图点位数据

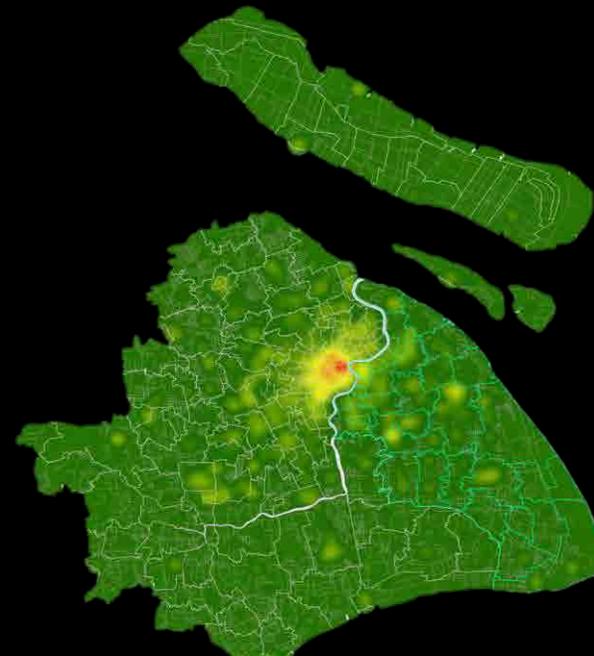
获取时间：2016下半年 样本量：400,000+



高德公交站点位置

房屋套数数据

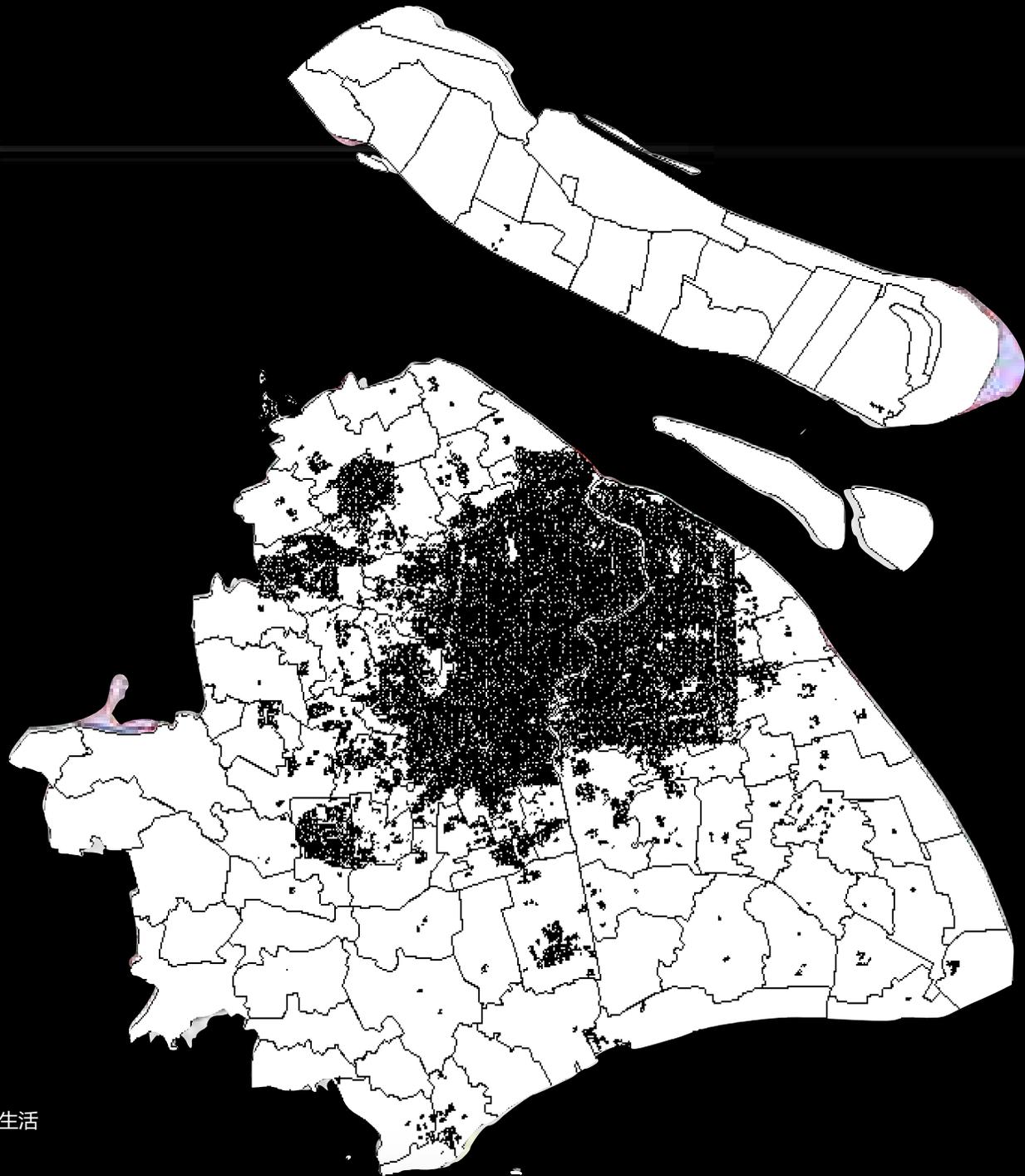
获取时间：2017年 样本量：7,600,000+

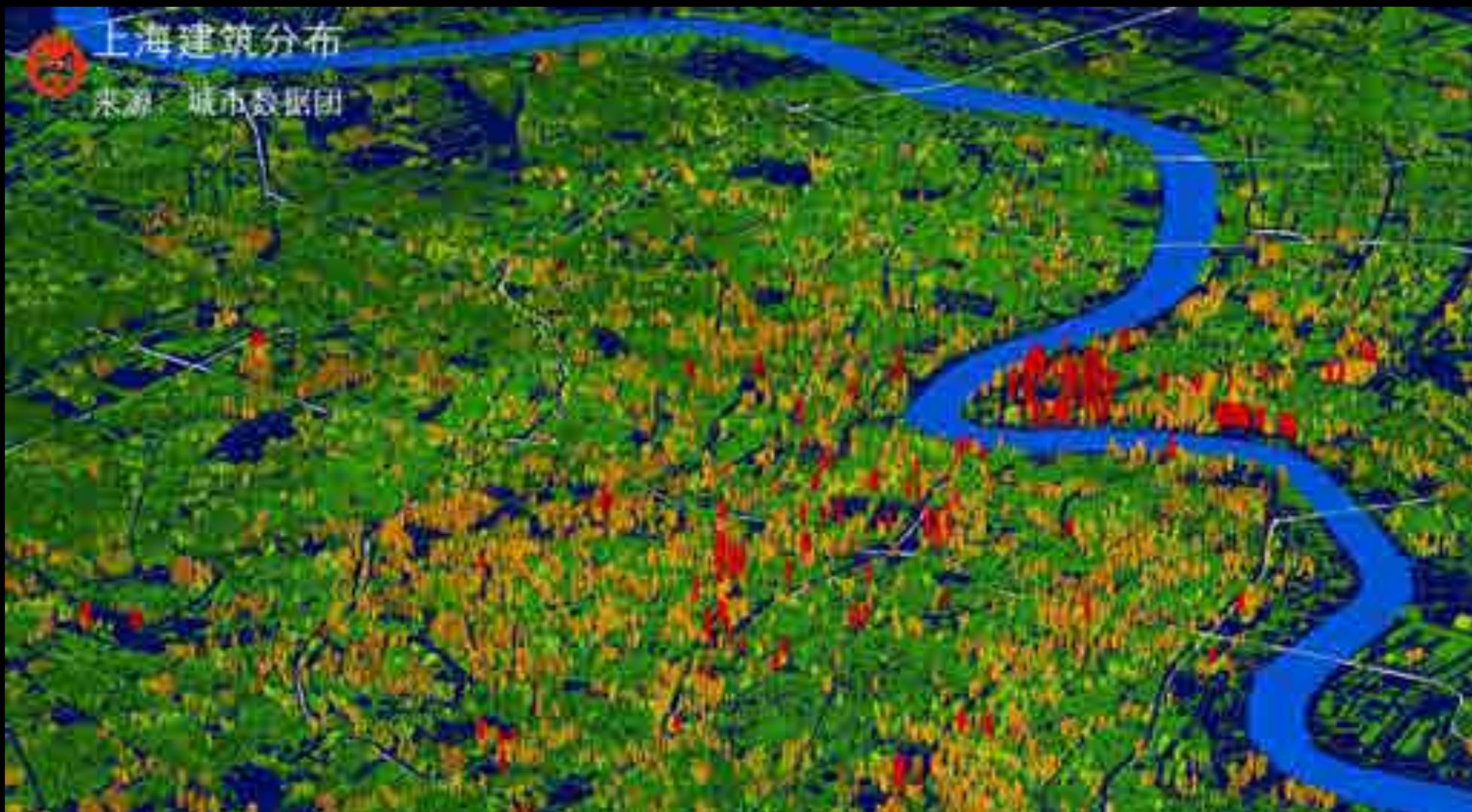


高德酒店位置

其他数据

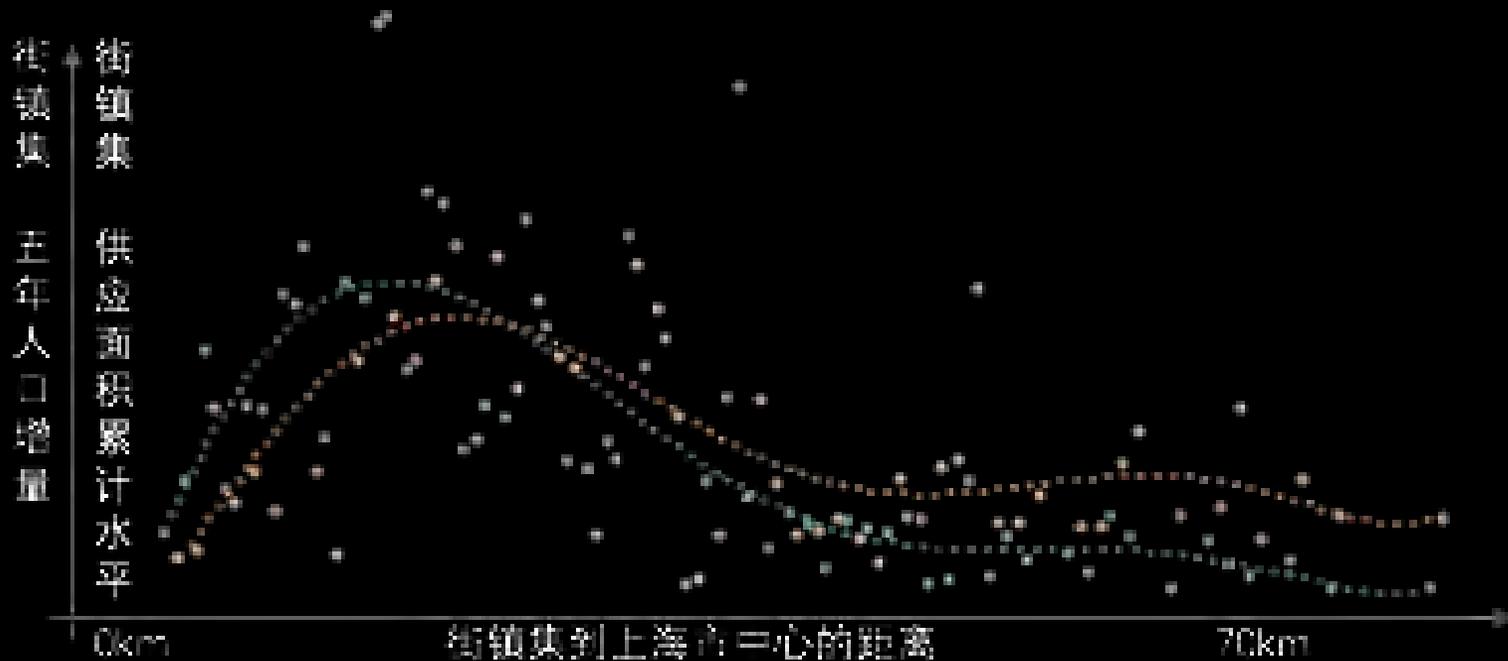
轨交进出站、生活用水、生活垃圾、在校人数等







人口增量与到市中心距离的关系 同时期新增一手房供应量与到市中心距离的关系



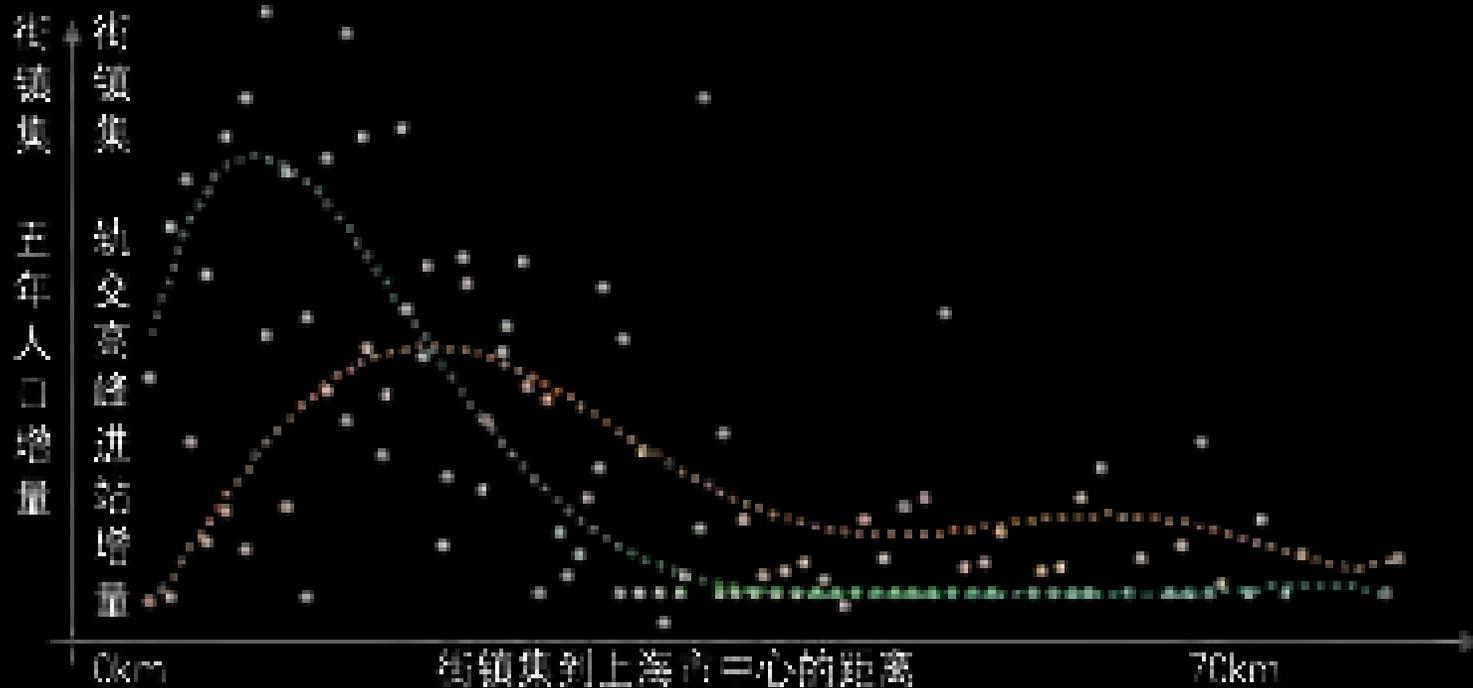
人口增量与一手房供应量回归结果：

街镇人口增量增加1万人；
一手房供应量增加9.56万平米；
相关性显著，显著性水平：1%

*注：街镇集：测量各街镇到市中心的直线距离，并按照1公里单位区间把街镇数据组合成街镇集



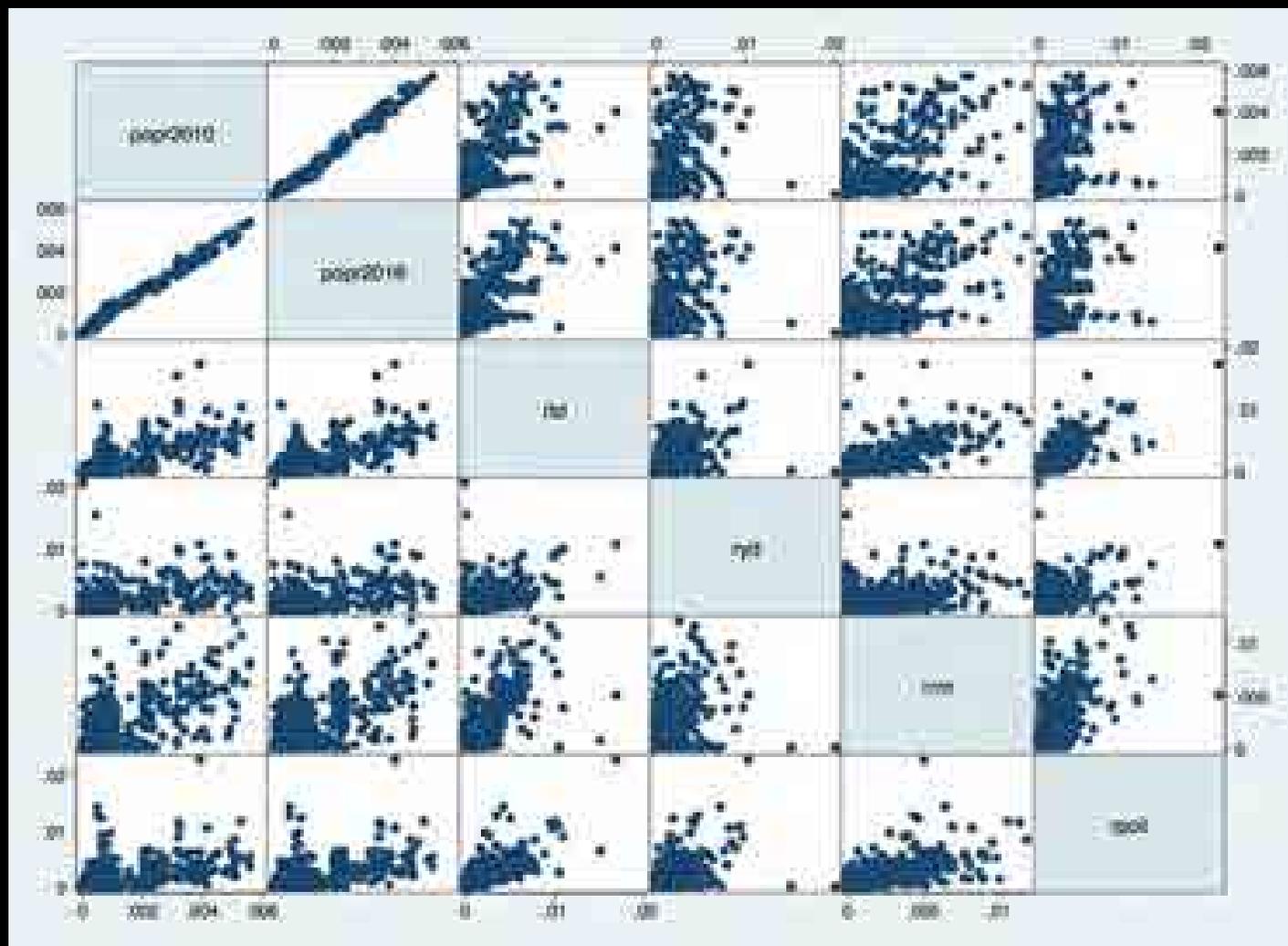
人口增量与到市中心距离的关系 轨交站点高峰时期进站人流增量与到市中心距离的关系



人口增量与一手房供应量回归结果：

街镇人口增量增加1万人
轨交高峰期进站人流增量提高781人
相关性显著，显著性水平：1%

*注：街镇集：测量各街镇到市中心的直线距离，并按照1公里单位区间把街镇数据组合成街镇集





街镇

栅格

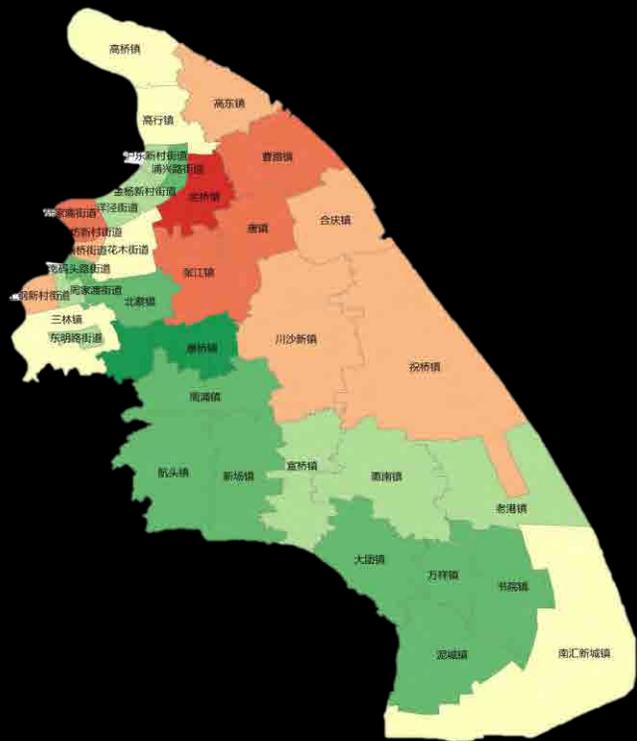
围栏



为什么要研究人口数据？

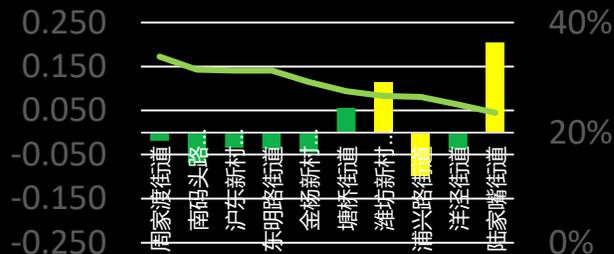


偏差系数计算：考虑城市发展层级差异的人口超常规活跃指数

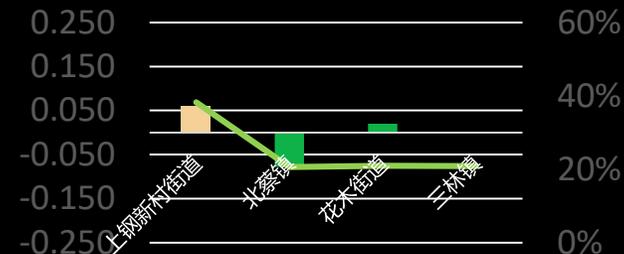


浦东新区各街镇：实有人口与修正常住人口的偏差可视化

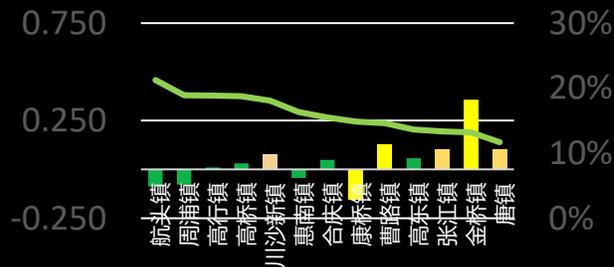
第一圈层人口偏差分析



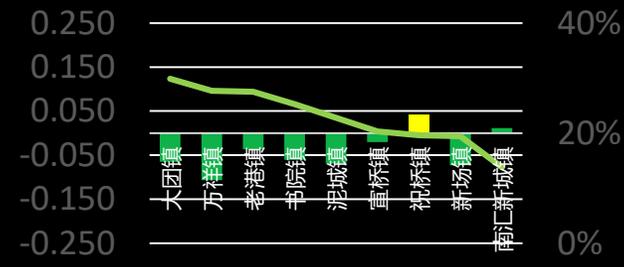
第二圈层人口偏差分析



第三圈层人口偏差分析



第四圈层人口偏差分析



图表解释：圈层人口偏差分析左侧纵轴代表人口偏差系数， $\text{偏差系数} = (\text{修正常住人口} - \text{官方实有人口}) / \text{修正常住人口} - 1$

右侧纵轴代表实有人口老龄化（大于60岁）水平 %

标签筛选

张江镇

劳动村

显示范围

步行15min

步行1h

劳动村综合评估

内容	达标率	缺失设施
教育	100%	
卫生	100%	
养老	100%	
文化	100%	
体育	100%	

年龄结构



教育 数量 类型 数量 万人均

<input type="checkbox"/>	幼儿园	幼儿园	4	18.56
<input checked="" type="checkbox"/>	小学	小学	2	332
<input type="checkbox"/>	中学	中学	1	1.67
<input type="checkbox"/>	中等职业学校			

名称 类型 级别 地址 建筑面积(m²) 占地面积(m²) 办学类型 在校学生数 班数 教职工

上海市浦东新区花木小学	小学	公办	浦东新区花木小学	13888	34900	公立	300	9	13
上海市浦东新区花木小学	小学	公办	浦东新区花木小学	11971.8	20500	公立	300	24	47

春节前夕人口流入区VS一线城市人口流出目的地





一座城市究竟有多少人——
这个问题重要吗？



数据 ——> 认知



THANKS