



**QCon** 全球软件开发大会  
INTERNATIONAL SOFTWARE  
DEVELOPMENT CONFERENCE

BEIJING 2018

# 《2018，你不知道的前端黑科技》

演讲者 / TakingData 马骥

# 目录：

- TalkingData简介
- 绘制百万行的数据表格
- web端数据科学实践
  - 自动化配色
  - 异常数据检测
  - 基于密度的离群点检测
- web端并行计算架构设计

# TalkingData介绍



## #1 覆盖体量

60亿  
累计设备  
Device Coverage

2.5亿  
日活设备  
Daily Active Devices

6.5亿  
月活设备  
Monthly Active Devices



## #2 数据吞吐

14T  
每天新增数据  
Daily Ingested Data

34亿  
每天交互会话  
Daily Sessions

370亿  
每天处理事件  
Daily Events



## #3 线下采集

80个  
城市  
Cities

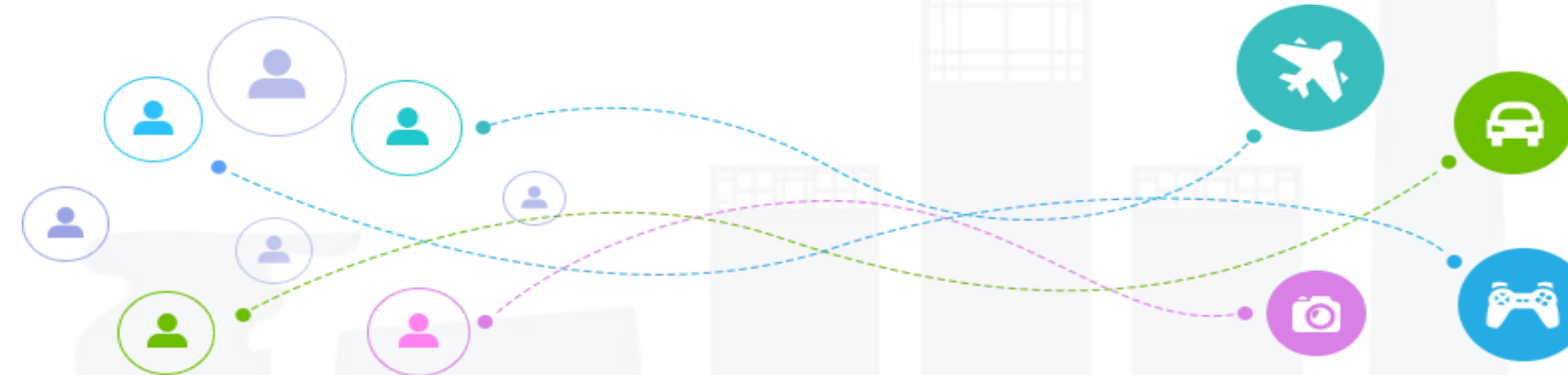
3,000个  
商城  
Shopping Malls

400万  
WIFI指纹  
WIFI Fingerprints

4,200万  
位置信息  
Points Of Interest



## #4 标签体系



12大类, 800+个标签



## #5 合作伙伴



人口统计学



金融



地产



零售



医疗



政府



航空



酒店



汽车



广告

Partner Ecosystem

# 绘制百万行的数据表格



高效、卓越的整套前端解决方案

Name	Age	Province	City	Action
John Brown	18	America	New York	View Edit
Jim Green	24	America	Washington, D.C.	View Edit
Joe Black	30	Australian	Sydney	View Edit

Name	Age	Address
John Brown	18	New York No. 1 Lake Park
Jim Green	25	London No. 1 Lake Park
Joe Black	30	Sydney No. 1 Lake Park
Jon Snow	26	Ottawa No. 2 Lake Park

展示
  唤醒
  登录
  点击
  激活
  7日留存
  30日留存
  次日留存
  日活跃

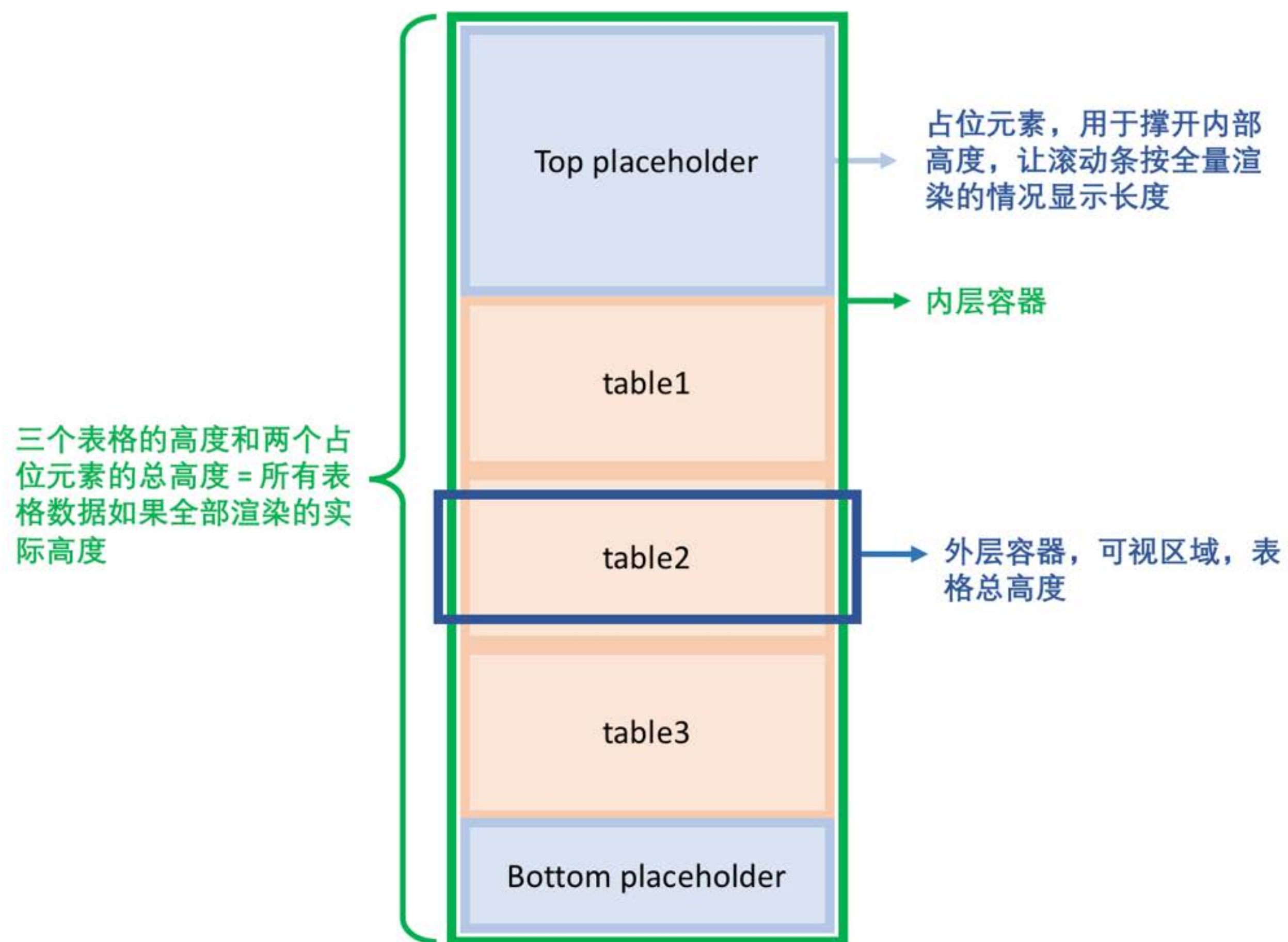
Name	Show	Weak	Activ
☆ Name1	1895	2356	5081
☆ Name2	4077	2797	6960
☆ Name3	5461	9518	2980
☆ Name4	4720	8570	9800
☆ Name5	8403	3981	5068

Name	Status	Portrayal	People	Sampling Time	Updated Time
Business56	Fail	5	3	Almost7days	2018-04-07
Business9	Fail	5	3	Almost4days	2018-04-07
Business40	Fail	5portrayals	3	Almost7days	2018-04-07
Business65	Fail	City	3	Almost5days	2018-04-07
Business47	Fail	People	3	Almost5days	2018-04-07
Business29	Fail	Cost	3	Almost5days	2018-04-07
Business72	Success	Life	3	Almost1days	2018-04-07
Business52	Fail	Entertainment	3	Almost7days	2018-04-07
Business19	Working	5	3	Almost1days	2018-04-07
Business71	Success	5	3	Almost7days	2018-04-07

# 绘制百万行的数据表格 - 有图有真相

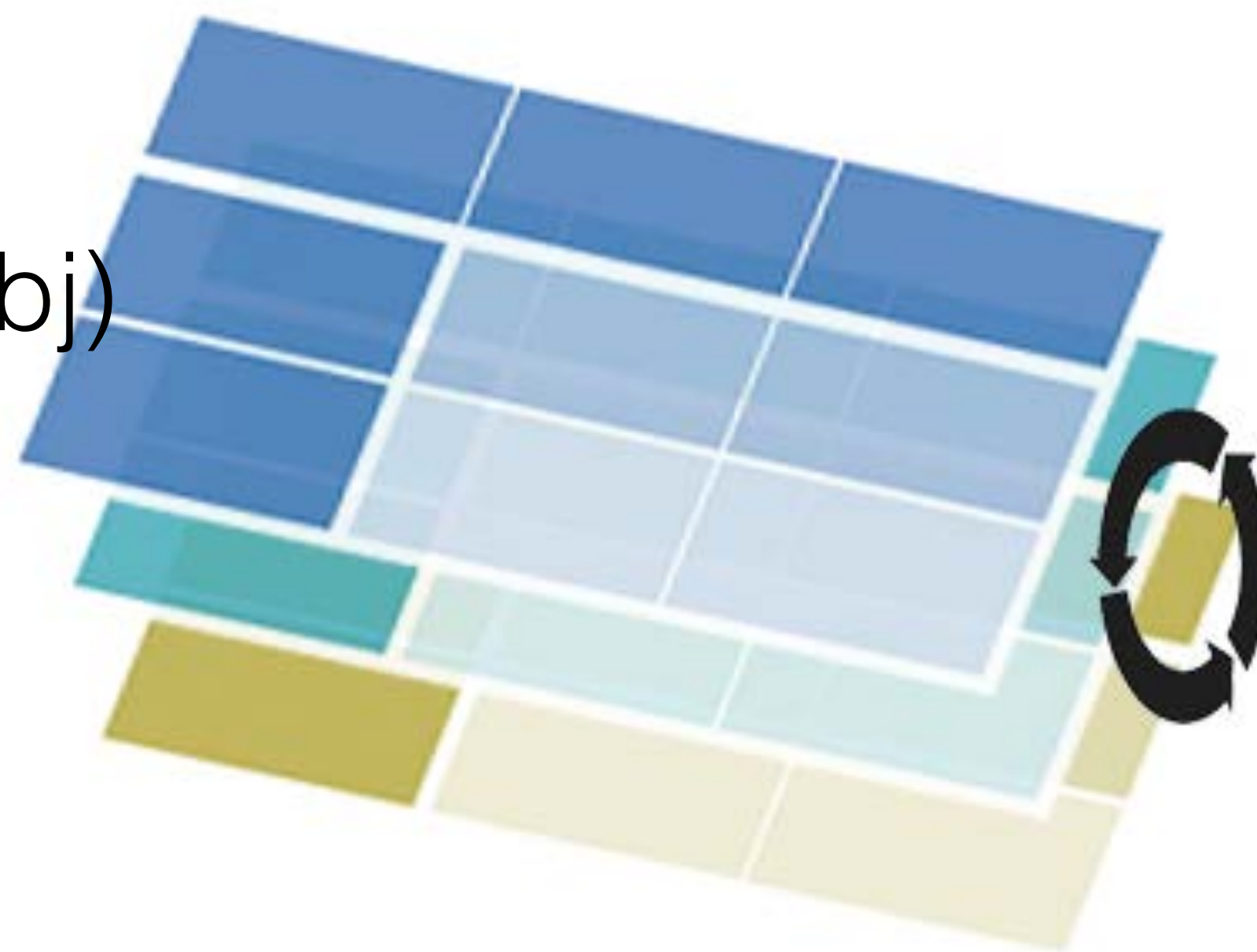


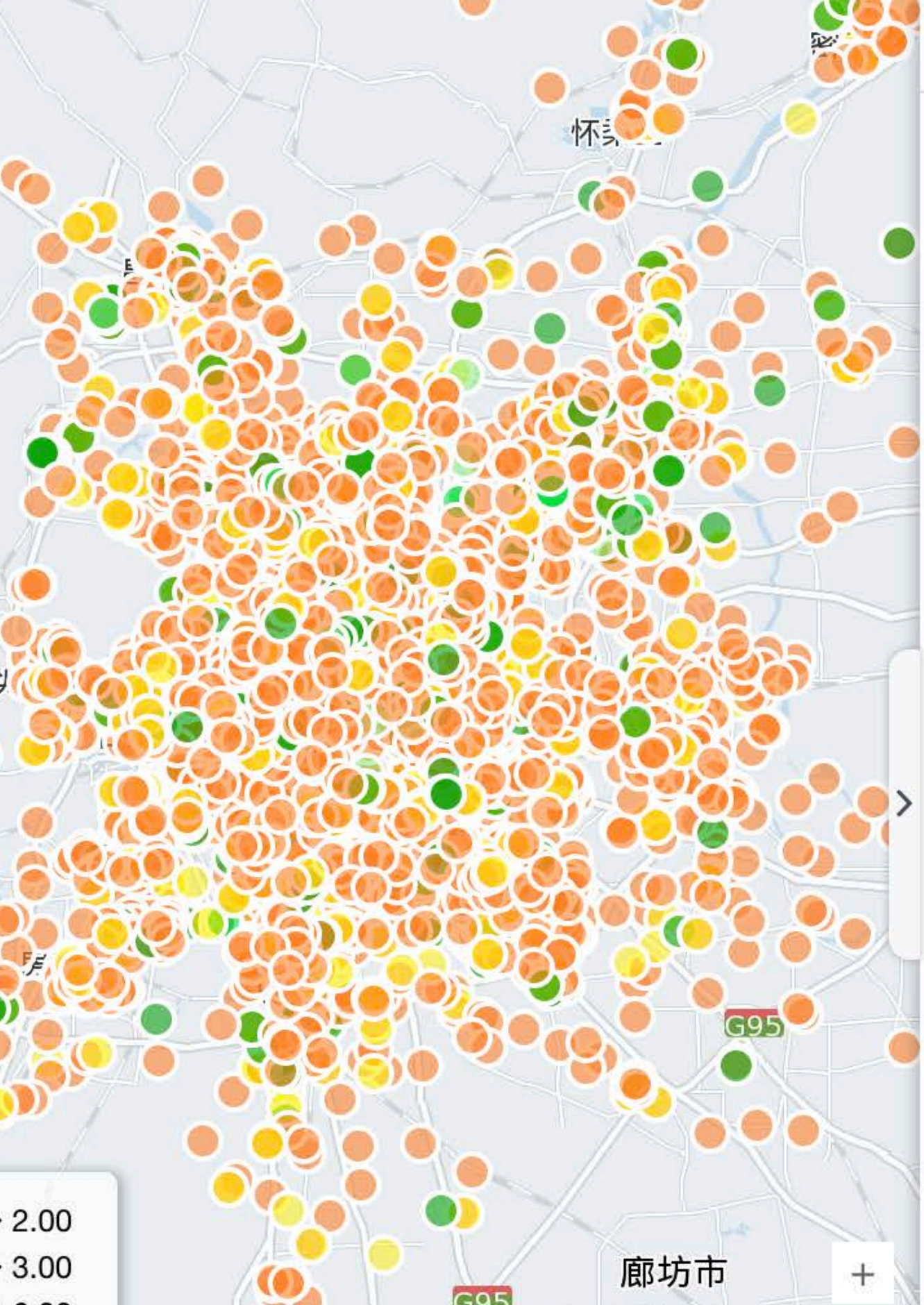
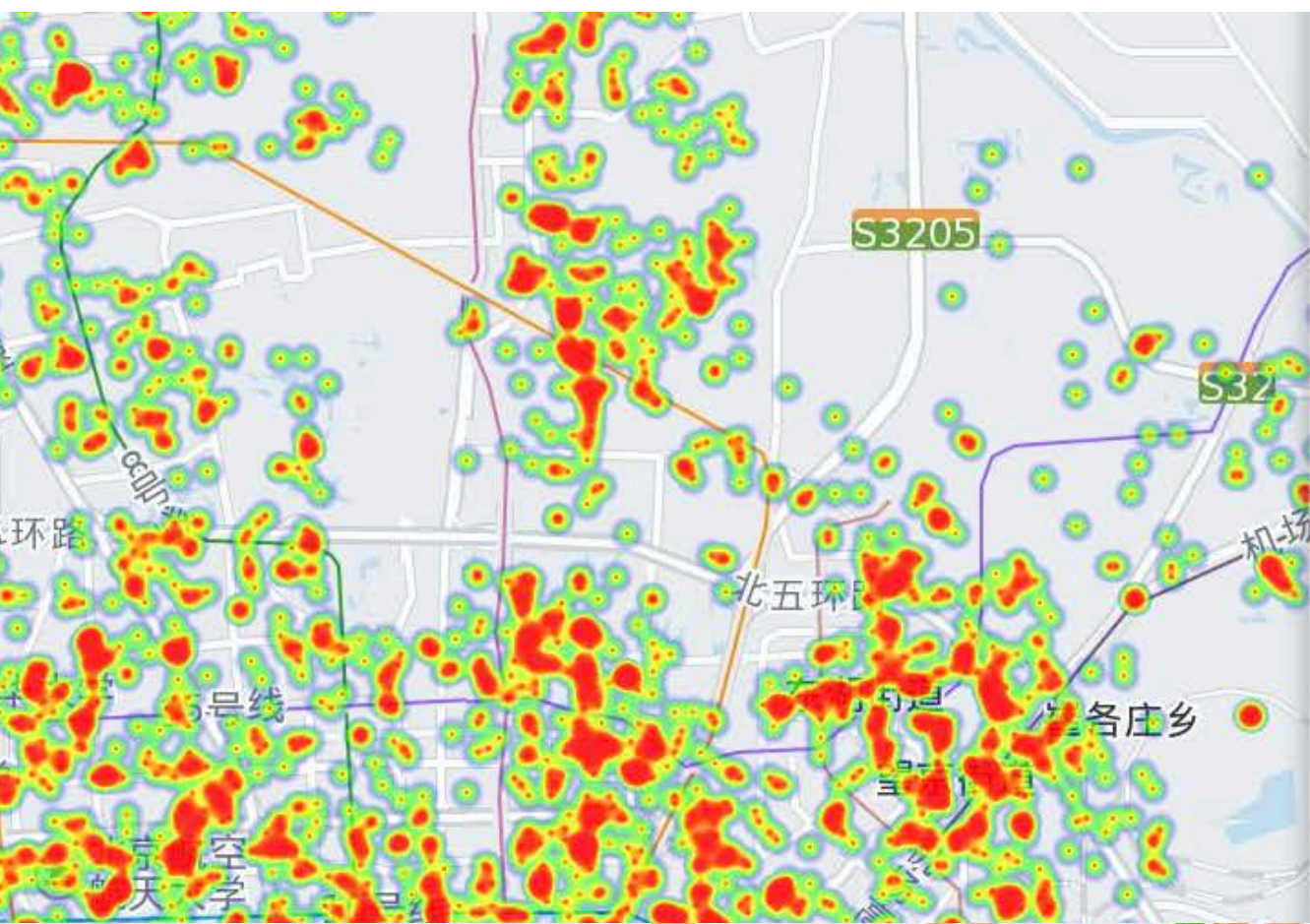
# 绘制百万行的数据表格



# 绘制百万行的数据表格

- getter和setter惹的祸
- 扩展对象Object.preventExtensions(obj)
- 密封对象Object.seal(obj)
- 冻结对象Object.freeze(obj)





视觉样式 筛选器

图例样式 地图样式

透明度

透明度

渲染半径

渲染半径

视觉样式 筛选器

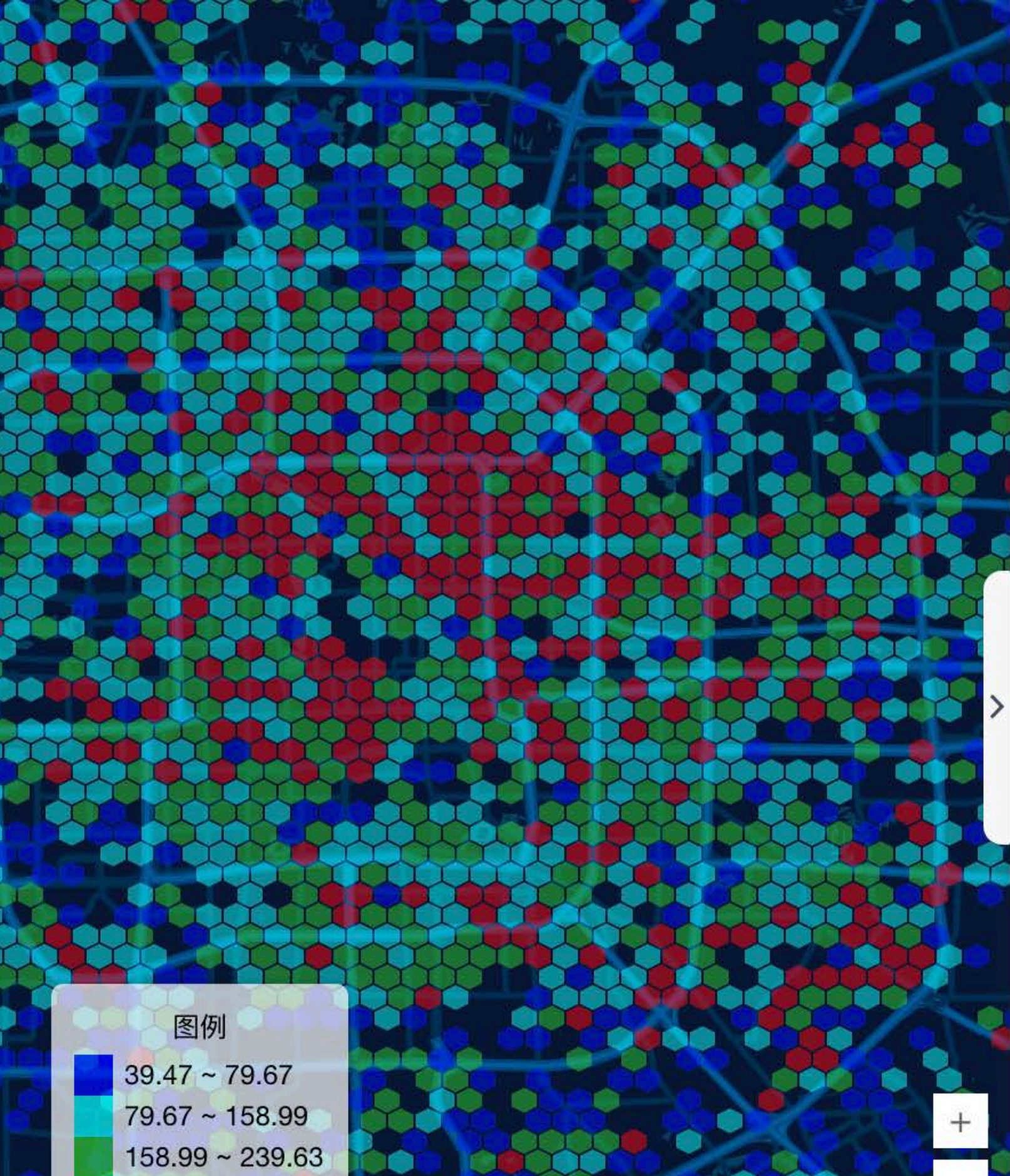
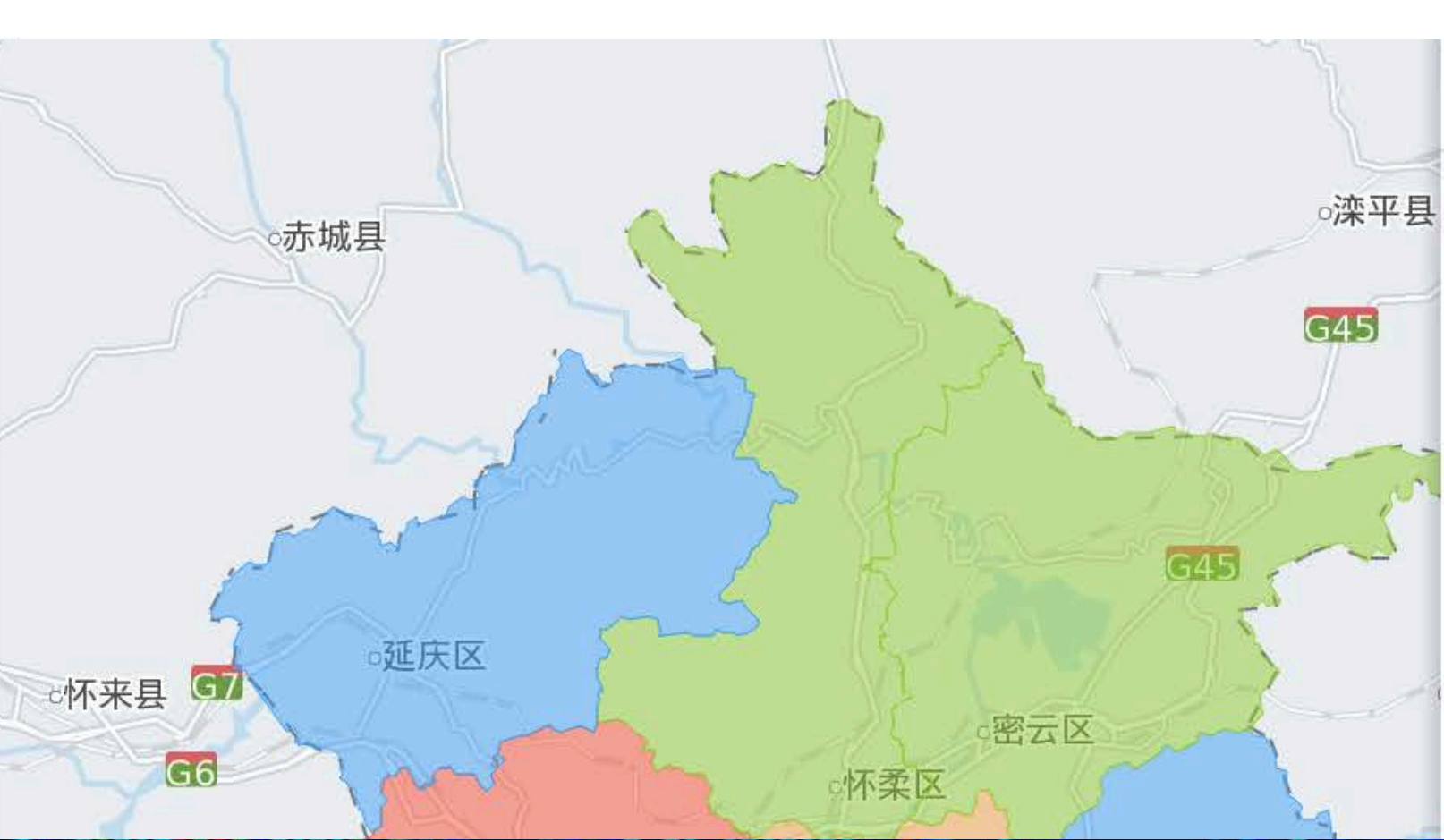
图例样式 地图样式

调色板

5

透明度

视觉焦点 ?



视觉样式 筛选器

图例样式 地图样式

调色板

5

视觉样式 筛选器

图例样式 地图样式

调色板

5

透明度  0.5

网格形状

正方形  六边形

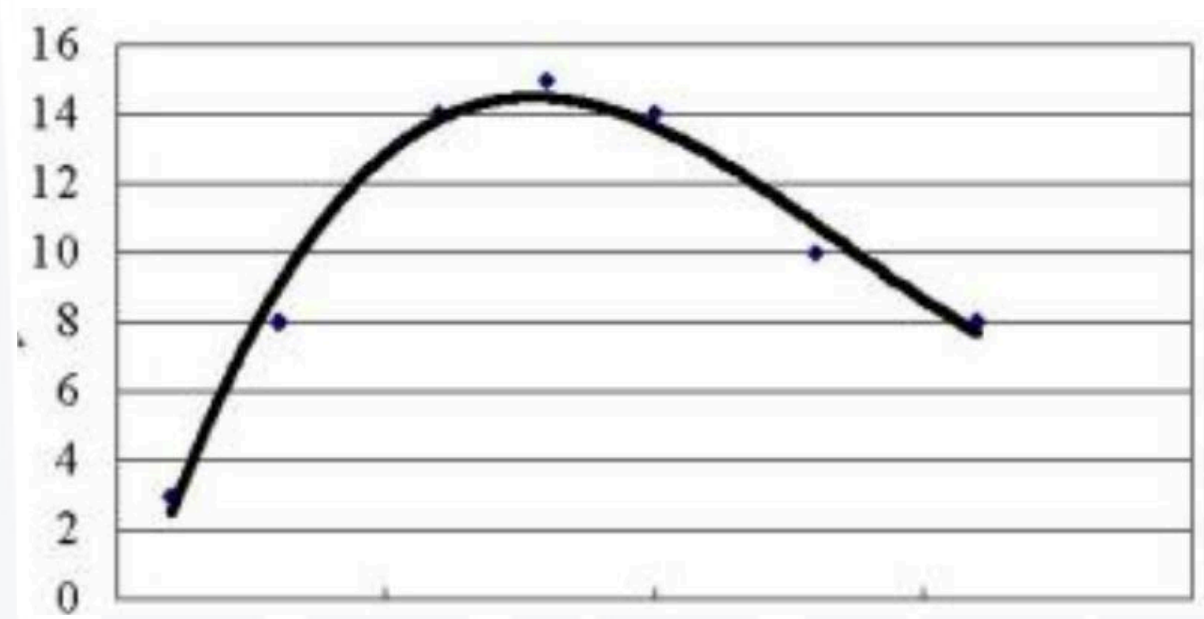
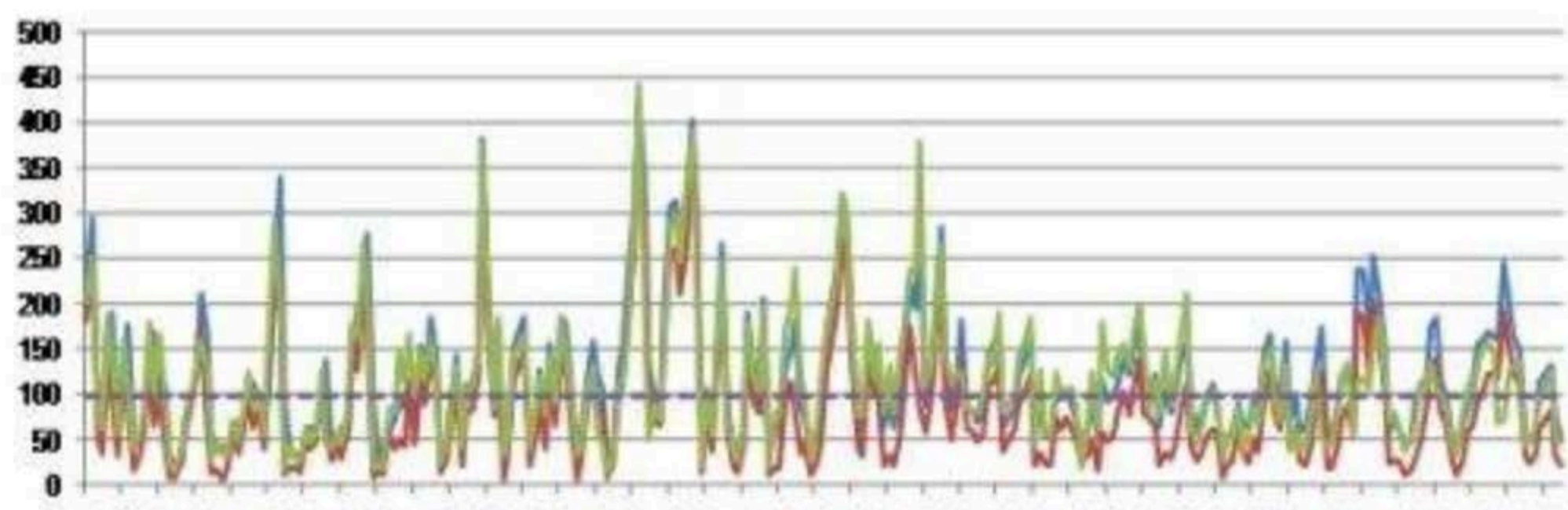
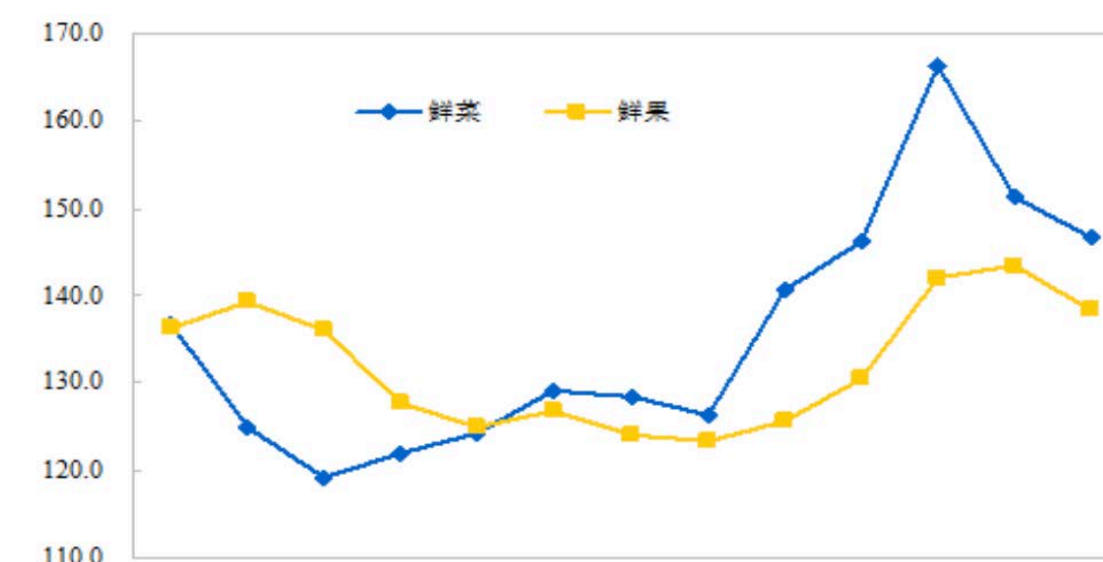
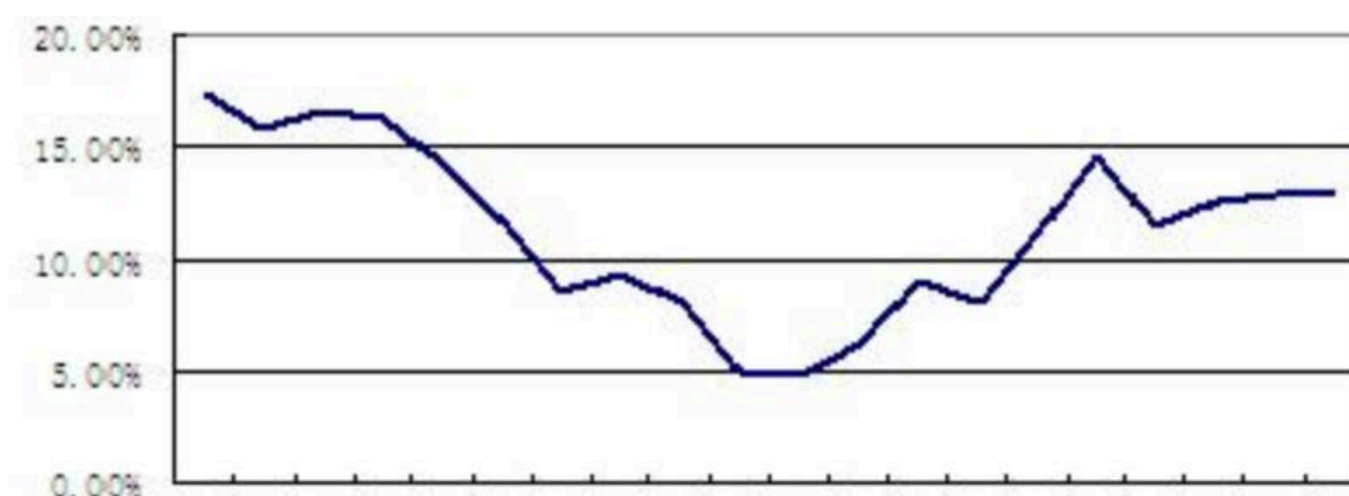
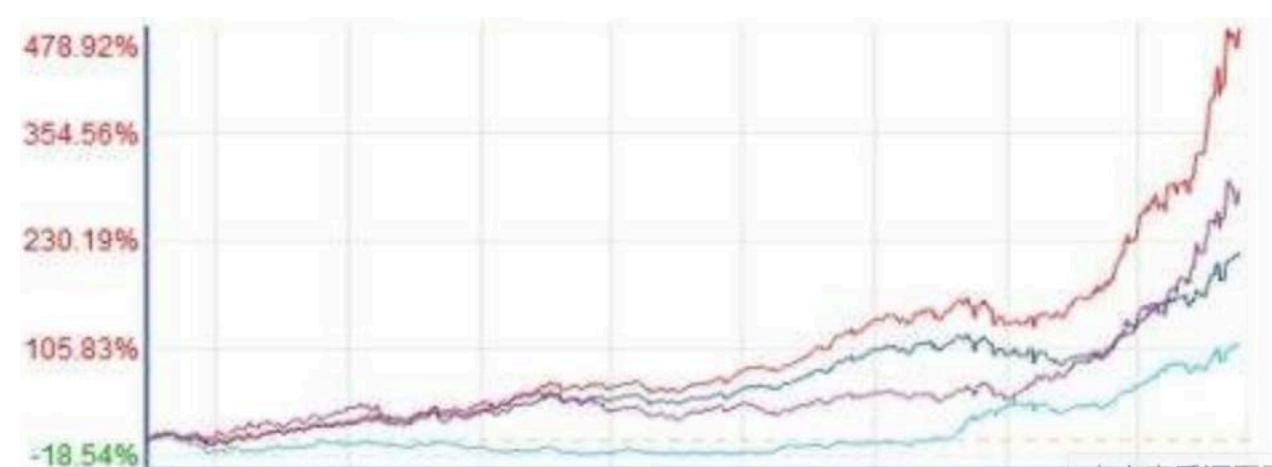
渲染半径

渲染半径  12



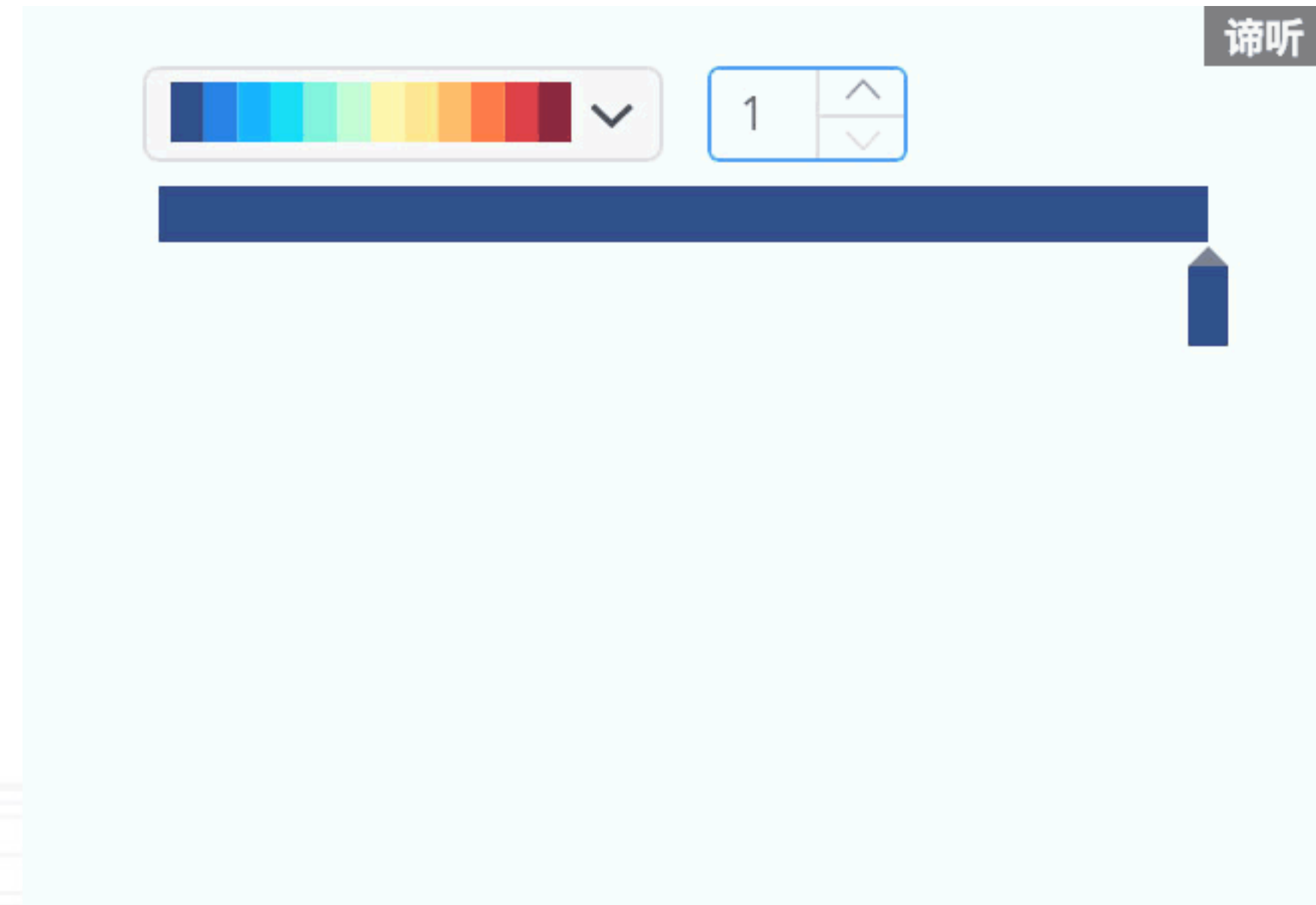


# web端数据科学实践-自动化配色



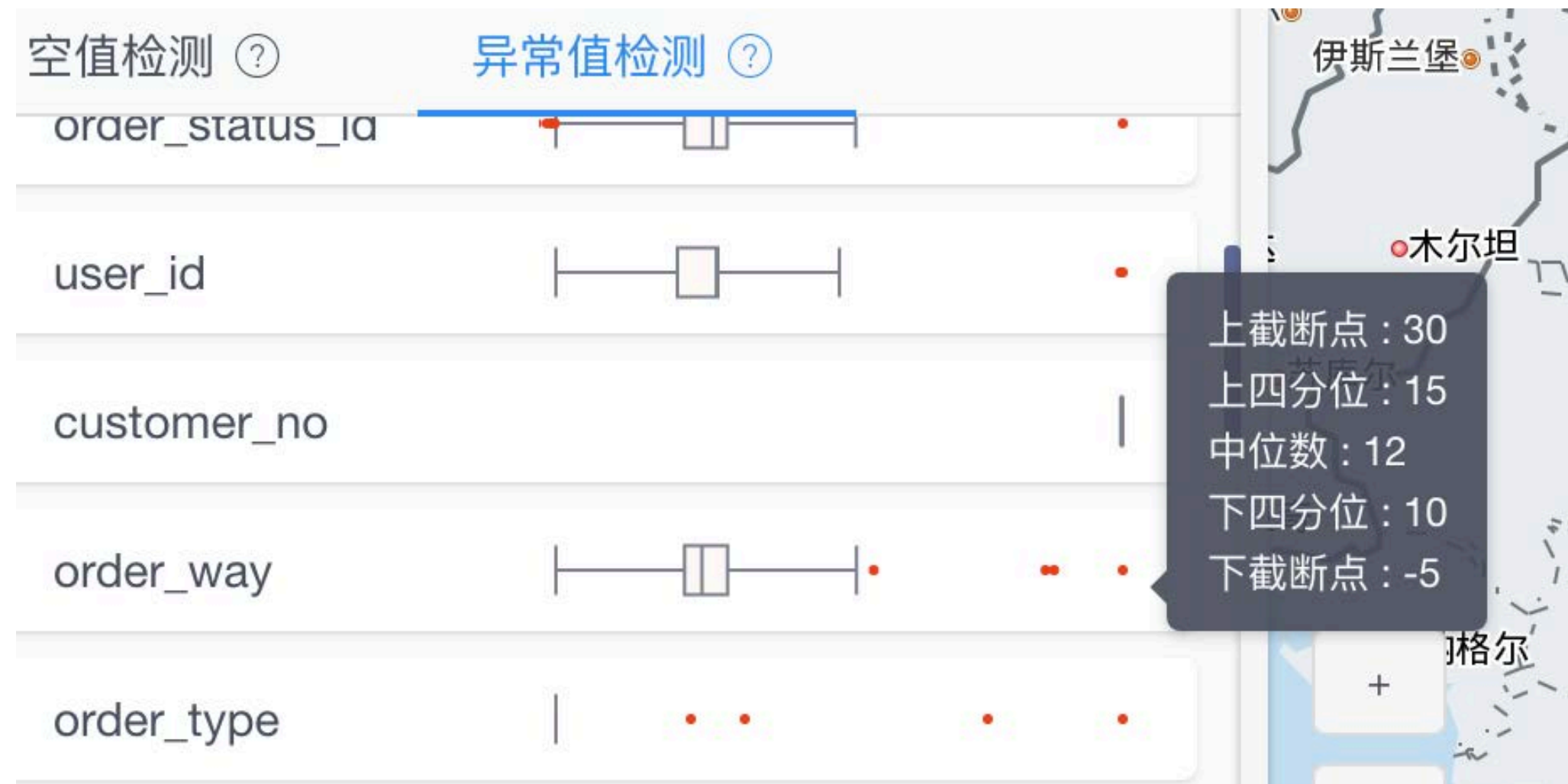
# web端数据科学实践-自动化配色

- 数据形态强相关（分布和趋势）
- 四分位识别
- 色谱面积等分点模型



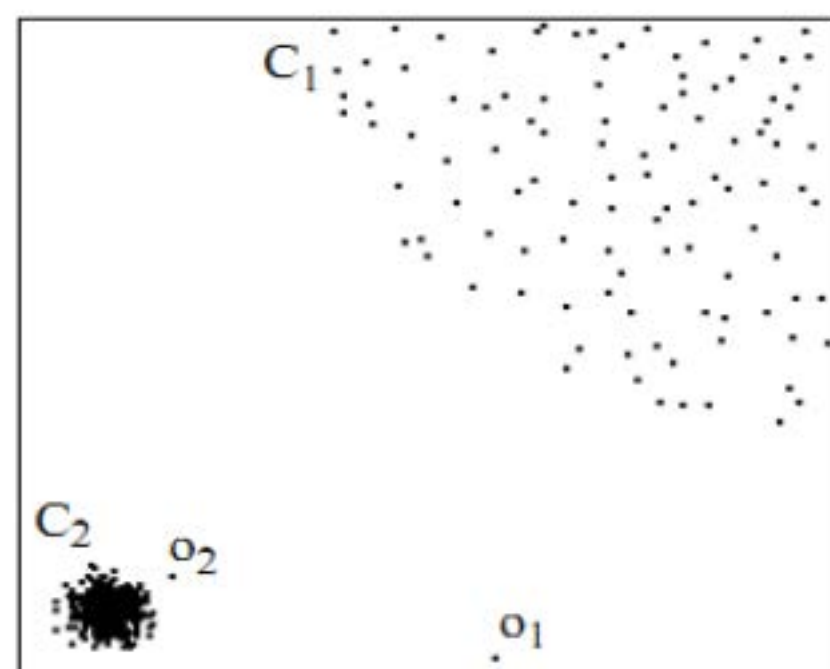
# web端数据科学实践 - 异常数据检测

- 空值、类型检测
- 数值类型箱线图检测



# web端数据科学实践 - 异常数据检测

- 位置数据基于密度的离群点检测

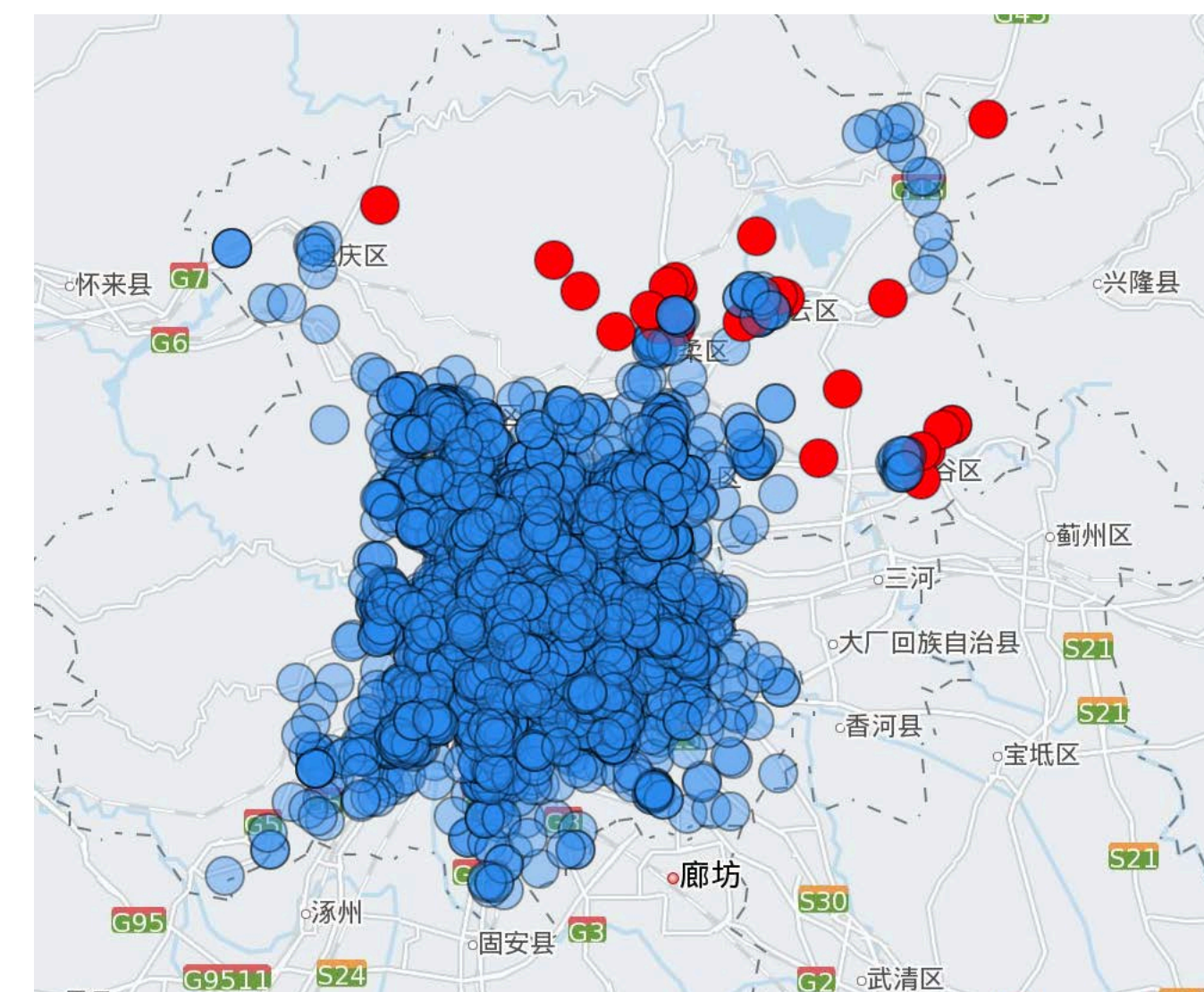


- 用视觉直观的感受一下，对于C1集合的点，整体间距，密度，分散情况较为均匀一致，可以认为是同一簇；对于C2集合的点，同样可认为是一簇。o1、o2点相对孤立，可以认为是异常点或离散点。

# web端数据科学实践 - 异常数据检测

- 位置数据基于密度的离群点检测LOF

LOF算法，主要是通过比较每个点 $p$ 和其邻域点的密度来判断该点是否为异常点，如果点 $p$ 的密度越低，越可能被认定是异常点。至于密度，是通过点之间的距离来计算的，点之间距离越远，密度越低，距离越近，密度越高，完全符合我们的理解。



# 大数据下数据可视化的挑战 - 小结

- 业务逻辑上移
- 数据科学算法模型的支持
- 大数据量浏览器渲染
- 大数据量计算性能



# 地图可视化传统的架构图

Client



Server



# 地图可视化架构图

- 前后端协同
- 提升Client端计算能力
- 降低Server服务压力
- 成本优势

Client

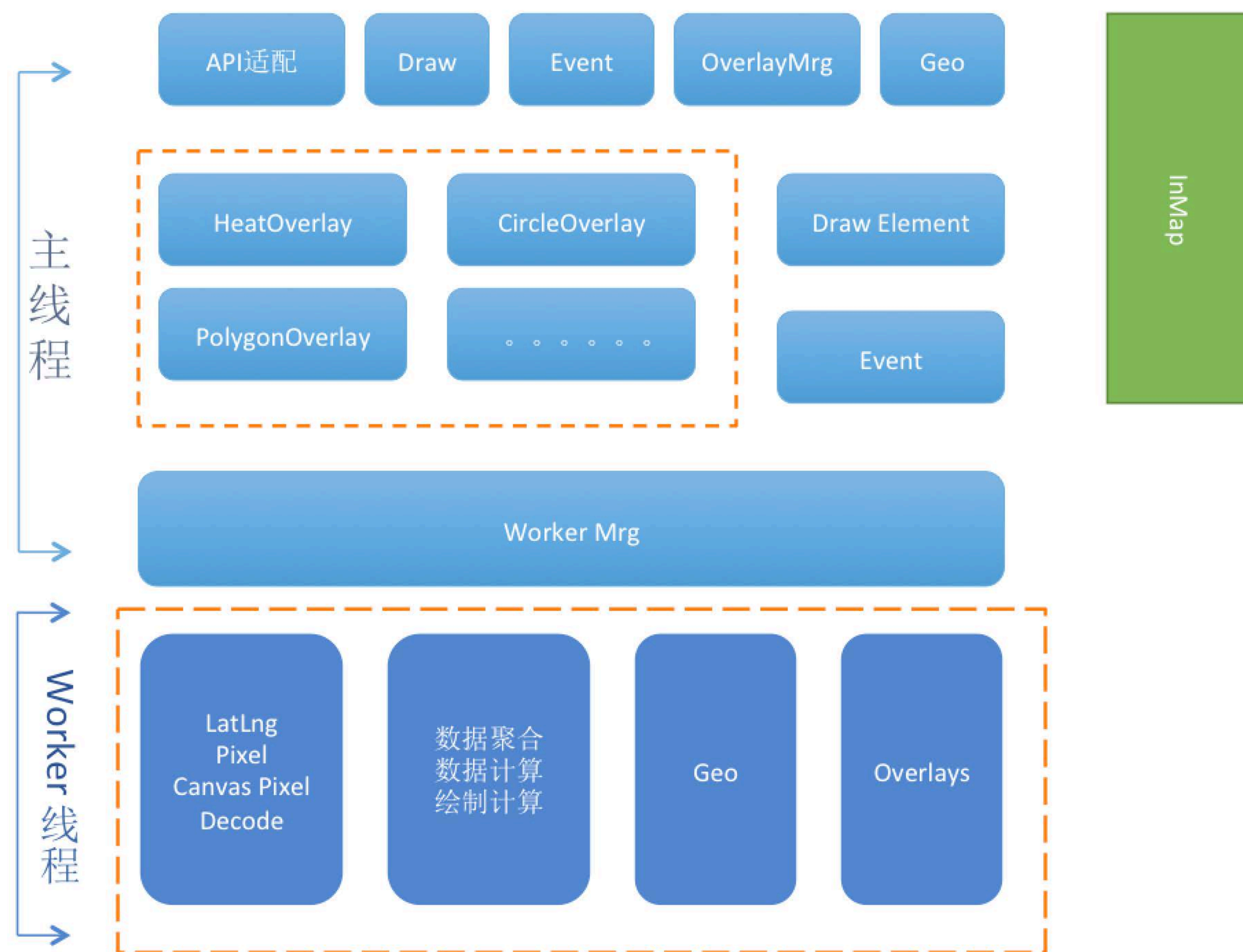


Server

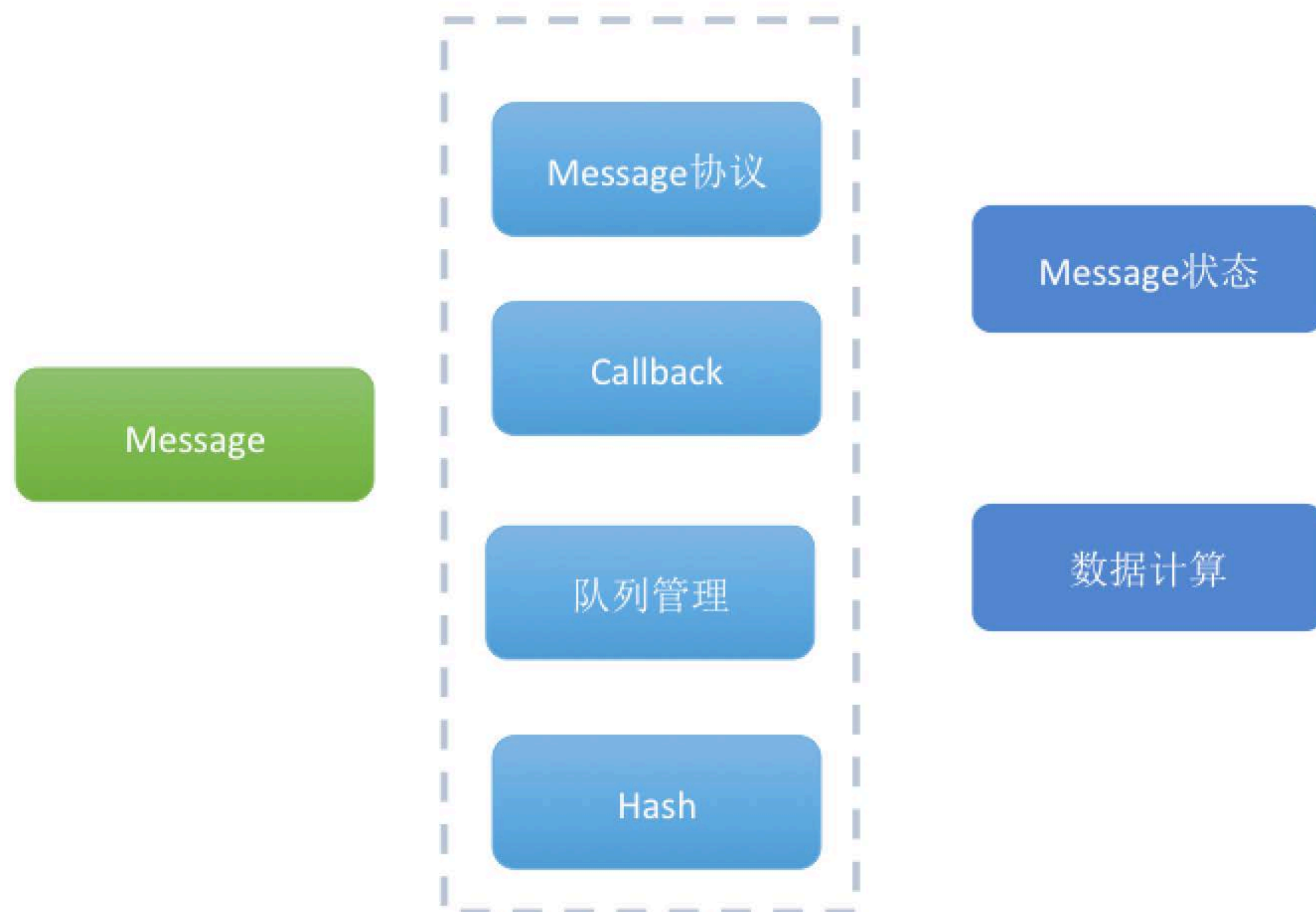




# 地图可视化架构图



# 地图可视化架构图





关注QCon微信公众号，  
获得更多干货！

# Thanks!



主办方 **Geekbang** > **InfoQ**  
极客邦科技