

全国城市空气质量数据可视化

杨宇菲 计算机系 2020215224

一、可视化任务

近年来,随着我国城市化、工业化进程不断加快,能源消耗迅速增加,大气污染日益严重,城市空气质量成为人们关注的焦点问题。

2012年3月国家发布的新空气质量评价标准中,综合二氧化硫、二氧化氮、PM10、PM2.5、一氧化碳和臭氧共6种污染物的浓度监测数据,用AQI(Air Quality Index, 空气质量指数)来定量评价空气的污染程度及其对健康的影响。在中国环境监测总站网站上,每小时更新全国各地的AQI和各项污染物监测数据。

针对全国城市空气质量的数据可视化,目前已存在的可视化平台大多仅对各区实时空气质量进行地图热力图绘制及排序,或对未来数天内各时段的空气质量进行预测和可视化,而对历史数据在时间维度上的可视化探索并不常见。因此,此提案希望在实现区域空气质量可视化的基础上,对每个城市的数年间的历史空气质量数据进行多角度分析和可视化,以此对各城市空气质量的长期发展变化产生更清晰和直观的认知,也可作为城市发展和环境治理成效的侧写。

二、数据相关

1. 数据来源

全国空气质量数据来自中国环境监测总站的全国城市空气质量实时发布平台。已有个人网页 <https://quotsoft.net/air/> 提供从2014年5月13日至今的历史数据打包下载(每周日更新)以及单日数据的下载(第二日02:00更新前一日数据),共包括191个中国主要城市,数据每小时更新。

2. 数据内容和格式

每日数据分别存为一个csv文件,名称为china_cities_日期,例如:2021年3月20日的全国空气质量数据文件名为china_cities_20200320.csv。

csv文件的第一行为列名,以下行为数据,每列的列名和对应数据的描述和格式如下:

- 1) date: 日期,格式同文件名
- 2) hour: 时段,从0到23的整数
- 3) type: 数据类型

type	数据类型	单位
AQI	AQI 实时值	N/A
PM2.5	PM2.5 实时浓度	(微克/立方米)
PM2.5_24h	PM2.5 的 24 小时滑动均值	(微克/立方米)
PM10	PM10 实时浓度	(微克/立方米)
PM10_24h	PM10 的 24 小时滑动均值	(微克/立方米)
SO2	SO2 实时浓度	(微克/立方米)
SO2_24h	SO2 的 24 小时滑动均值	(微克/立方米)
NO2	NO2 实时浓度	(微克/立方米)
NO2_24h	NO2 的 24 小时滑动均值	(微克/立方米)
O3	O3 实时浓度	(微克/立方米)
O3_24h	O3 的 24 小时最大值	(微克/立方米)
O3_8h	O3 的 8 小时滑动均值	(微克/立方米)
O3_8h_24h	O3 的 8 小时滑动均值的 24 小时最大值	(微克/立方米)
CO	CO 实时浓度	(毫克/立方米)
CO_24h	CO 24 小时滑动均值	(毫克/立方米)

4) 各个城市的名称，以下每行为某时段对应城市的相应种类污染物数据

date,	hour,	type,	北京,	天津,	石家庄,	唐山,	秦皇岛,	...
20210320,	0,	AQI,	73,	81,	68,	75,	88	
20210320,	0,	PM2.5,	53,	60,	49,	55,	65	
20210320,	0,	PM2.5_24h,	72,	75,	56,	78,	38	
20210320,	0,	PM10,	16,	59,	46,	47,	56	
20210320,	0,	PM10_24h,	88,	106,	123,	128,	65	
...								

3. 数据处理

原数据为按小时计的实时空气质量数据，需要在此基础上计算单日、月度、季度和年度的平均值/最大值/最小值，以提供多分辨率、多尺度的历史空气质量数据可视化。并在此基础上进行其他数据分析如排名、比例统计等，进一步探索我国空气质量的地域和时间特性。

三、设计方案

可视化方案可归纳为以下几个部分。

1. 基于地图的多分辨率可交互城市空气质量可视化：以全国地图为表示平面绘制散点图，使用图例颜色的深浅表示空气质量指标的大小。用户对地图进行放大，当放大到一定阈值后，从显示主要城市变为显示区域内全部城市。用户可切换可视化空气质量指标为 AQI/PM2.5/PM10/SO2/NO2/O3/CO，并可在不同时间层面（单日/月度/季度/年度）选择不同时间的数据进行可视化展示。将鼠标放置在代表某城市的散点上时，出现浮窗，以雷达图的方式展现此城市在所选时段六项污染物指标的值。
2. 多分辨率可交互城市空气质量历史数据可视化：在 1 的地图上点击代表某城

市的散点，进入此城市的历史数据可视化页面。在全部时间层次，以散点图（小时指标-时间）的形式可视化全时长的历史空气质量数据，也可切换为箱型图（年度指标-年份）形式的年度空气质量分析可视化。在年度层次，可切换散点图（小时指标-时间）与月度箱线图（月度指标-月份）。在月度层次，可切换折线图（小时指标-时间）与日箱线图（单日指标-日期）。在单日层次，展示小时指标随时间变化的折线图。同时，在一侧用饼图或堆叠图的方式展示所选时间范围内各指标范围所占的天数比例（全部时长/年度/月度层次）或时段比例（单日层次）。对于细分时间段的选择，当鼠标悬停在图表中相应时段区域内时进行高亮，点击即可聚焦此时段，进入对偶尺度模式。同上，用户可切换可视化空气质量指标为 AQI/PM2.5/PM10/SO2/NO2/O3/CO。

3. 动态可视化：

- 1) 统计每季度 AQI 最好和最差的 20 座城市，对它们进行柱状图排名的动态可视化。
- 2) 对全国各地区（华中/华北/华东/华南/西北/东北/西南）城市的 AQI 进行柱状图排名的动态可视化。
- 3) 统计各季度 AQI 最好和最差的 20 座城市的省份或地区，对它们的地域分布进行饼状图的动态可视化。
4. （可选内容）结合国家统计局提供的各省市森林资源和造林面积等年度数据，绘制散点图，探索空气质量与绿化覆盖率等的相关性。

四、工作量预估

- 数据预处理（20%）
- 核心可视化代码实现：城市空气质量的区域可视化（20%）、城市空气质量历史数据可视化（20%）、动态可视化（10%）
- 交互功能实现（20%）
- 网页设计与视觉效果优化（10%）

预计共需 10 周完成。

五、个人联系方式

姓名：杨宇菲

学号：2020215224

学院：计算机系

邮箱：yangyufe20@mails.tsinghua.edu.cn

手机：13691080309

组队情况：暂无