

数据可视化报告

成员：

时浩铭 1911463

林坤 1911433

周民宇 1911529

一、作品简介

作品主题与问题场景

自 2013 年“大气十条”实施以来，全国各地狠抓大气污染治理，我国大气污染防治成效显著。近年来，全国多处空气质量监测站收集到大量具有高维、时序特点的空气质量数据；使用可视化与可视分析技术，对识别主要大气污染源、分析污染成因、分析污染时空分布模式、比较各地大气污染差异，识别特定事件等方面进行分析，对解读该类数据、提供知识见解具有重要意义。

本作品主要使用 2018 年中国高分辨率大气污染再分析开放数据集，使用污染物特征值计算，设计并实现对 2018 年间全国大气污染状况的展示性可视化系统；同时重点对全国各地大气污染成因与类型、主要污染源、大气污染时空分布模式、时空演变态势展开可视分析，并提供对 2018 年内任意地区不同或相同时间点的对比分析视图，比较其大气污染物差异；为大气污染防治工作的分析、指挥和决策提供有效手段和决策依据。

目标用户与应用价值

本作品目标用户包括关注大气污染情况的大众，分析及评估大气污染成效并制定防治措施的部门，以及大气污染数据分析的相关研究人员与学者。

基于大气污染数据的可视化使大众能直观与直接地感受数据所展示出的某地区某时地污染情况，同时直观地为用户展示各地大气污染的情况，结合各地大气污染物差异的比较结果，对其相关措施的有效性与改进建议提供辅助支持；对于相关研究人员与学者，通过数据处理思路与可视化视图与交互的设计，提供大气污染数据分析的新思路。

二、数据介绍

二、数据介绍

2.1 数据来源与格式

结合设计主题与应用场景，本作品主要使用 2018 年中国高分辨率大气污染再分析开放数据集，以及由阿里云 DataV 提供的来自高德开放平台 2020 年 4 月更新的地理数据（<http://datav.aliyun.com/tools/atlas>）。通过 DataV Atlas 地图选择器，本作品获取了中国、全国 34 个省份地理数据文件作为系统中地图地理数据的输入。

提供的大气污染数据集由中国科学院大气物理研究所等单位发布，数据包括日均值数据，存储为.csv 格式，包括我国六项常规污染物、五个常用气象要素以及所在网格点的经纬度值。

2.2 数据预处理

结合可视分析需求，本作品对数据进行了不同方面的预处理。最终用于可视分析系统输入的数据均被转换为.csv 格式，具体的数据预处理过程如下。

2.2.1 逆地理编码与数据清洗

首先使用高德地图开放 API 对数据集中的经纬度做逆地理编码，获得每对经纬度坐标对应的行政区划信息到二级地级行政区。在此过程中，删除了发现的 38 条不在高德地图查询范围内的坐标数据；对直辖市地区统一使用直辖市的名称。获取到行政区划信息后，将该信息添加到源数据集的最后作为新的列，得到经过逆地理编码与数据清洗后的数据，该数据以.csv 格式存储。

对应代码在script文件夹下可以查看。

2.2.2 数据统计

对清洗后的数据，按照行政区划，分别统计某省份（省级行政区）和某城市（地级行政区）内所有经纬度点数据的平均值，得到按行政区划统计的日均值数据；在此数据的基础上，对各省份、城市的污染物变量及气象变量按日、月做统计，得到各行政区划不同时间粒度的污染物变量与气象变量均值。

2.2.3 IAQI AQI污染等级计算

空气质量指数（AQI, Air Quality Index），是根据空气中的各种成分占比，将监测的空气浓度简化成为单一的概念性指数值形式。它将空气污染程度和空气质量状况分级表示，适合于表示城市的短期空气质量状况和变化趋势。空气质量分指数（IAQI, Individual Air Quality Index）是指根据单项污染物计算得出的空气质量指数。

根据 2012 年 2 月 29 日国务院发布新修订的《环境空气质量标准》（GB 3095-2012）中对环境空气质量标准分级、主要污染物项目和这些污染物在各个级别下的浓度限值，本作品对各项污染物的 IAQI 与 AQI 进行了计算，并判断得出最终的污染等级。

(1) 污染物浓度标准

本作品中按照以下空气质量分指数及对应的污染物项目浓度限值标准计算各污染物空气质量分指数。

空气质量 分指数 (IAQI)	SO ₂ 24h 平均 (μg/m ³)	SO ₂ 1h 平均 (μg/m ³)	NO ₂ 24h 平均 (μg/m ³)	NO ₂ 1h 平均 (μg/m ³)	PM10 24h 平均 (μg/m ³)	CO 24h 平均 (mg/m ³)	CO 1h 平均 (mg/m ³)	O ₃ 8h 滑动平均 (μg/m ³)	PM2.5 24h 平均 (μg/m ³)
0	0	0	0	0	0	0	0	0	0
50	50	150	40	100	50	2	5	100	35
100	150	500	80	200	150	4	10	160	75
150	475	650	180	700	250	14	35	215	115
200	800	800	280	1200	350	24	60	265	150
300	1600	1600	565	2340	420	36	90	800	250
400	2100	2100	750	2090	500	48	120	1000	350
500	2620	2620	940	3840	600	60	150	1200	500

其中，SO₂ 1h 浓度高于 800μg/m³ 的，不再进行其空气质量分指数的计算，按 24h 平均浓度计算分指数报告；当 O₃ 8h 浓度高于 800μg/m³ 的，不再进行其空气质量分指数的计算，按1h 平均浓度计算分指数报告。

(2) IAQI 与 AQI 计算公式

根据相关标准，本作品中使用如下公式计算 IAQI 与 AQI：

$$IAQI_P = \frac{IAQI_{Hi} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}}(C_p - BP_{Lo}) + IAQI_{Lo}$$

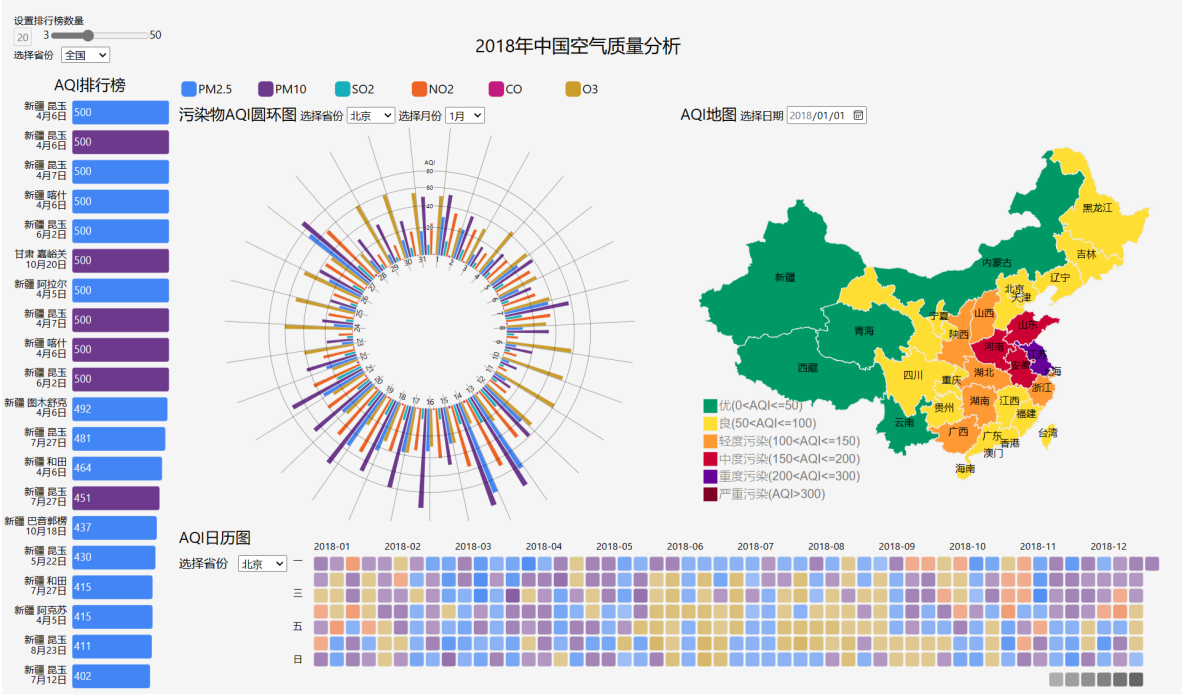
其中，BP_{Hi}代表某特定污染物所在的最小区间的上限。BP_{Lo}代表某特定污染物所在的最小区间的下限，BP_{Hi} - BP_{Lo}代表该区间的上下限差值，为一个固定值。IAQI_{Hi}代表某特定污染物所在的最小区间对应的空气质量分指数(IAQI)的区间的上限，IAQI_{Lo}代表某特定污染物所在的最小区间对应的空气质量分指数(IAQI)的区间的下限，IAQI_{Hi} - IAQI_{Lo}代表该区间的上下限差值，为一个固定值。C_p代表实际测得的污染物浓度，C_p - BP_{Lo}代表污染物实际浓度与浓度指标的地位值之差，它不是一个固定值。整个乘积再加上各级别的起点指数即可算出该污染物的空气质量分指数。每种指标按照对应时间区间取值，在相同时间区间下的所有污染物的 IAQI 计算结果的最大值作为这个时间区间的 AQI 数值。

三、可视化与交互设计

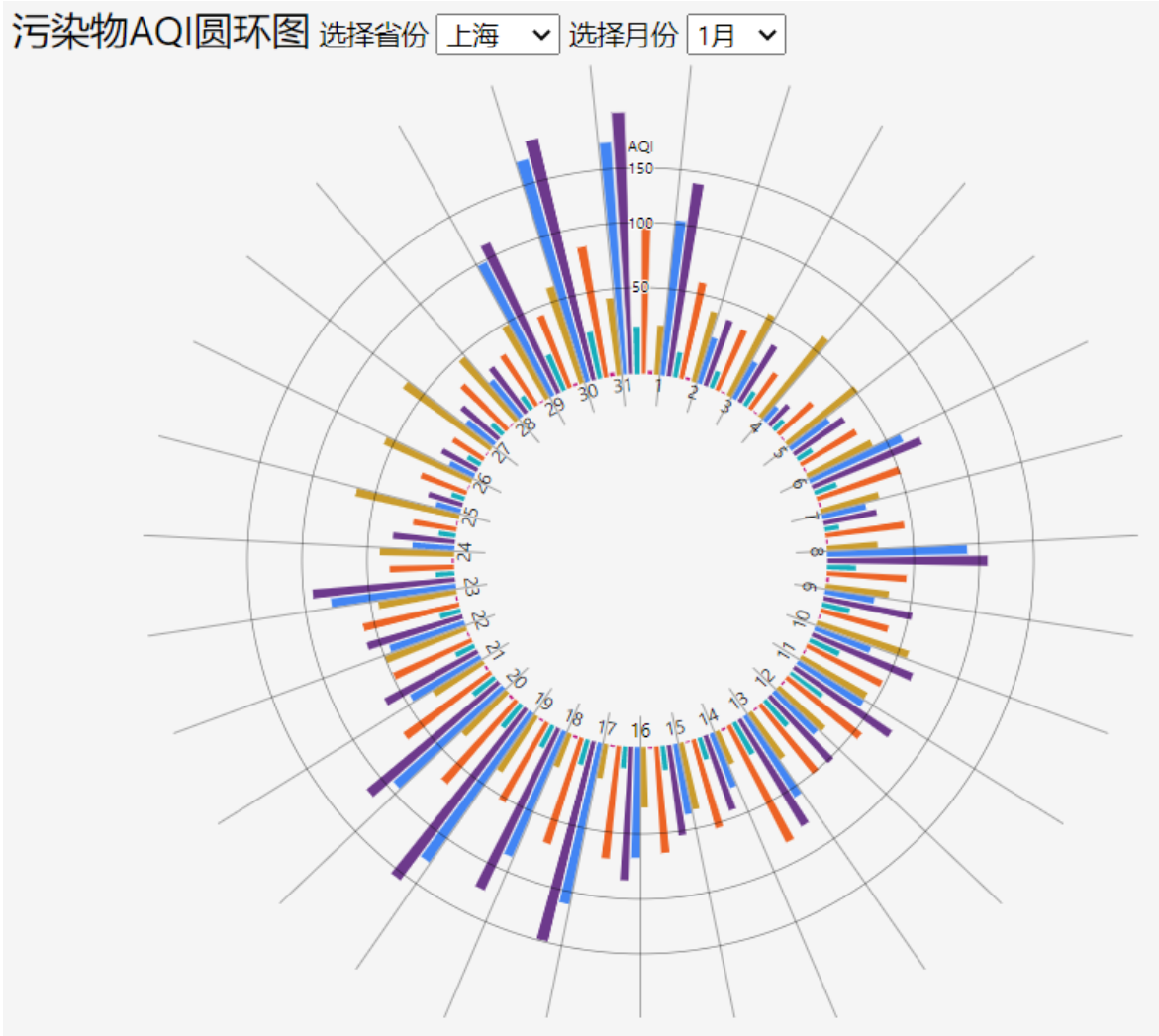
系统总体介绍

本作品系统以一个页面呈现，主要以日均值数据作为可视分析系统的输入，通过对于徐要可视化的污染物种类进行选择，可以切换显示不同类型污染物可视化效果。作品中设计实现了多种视图，辅以合理的视觉编码，同时支持灵活直观的交互设计，可以显示对应日期，不同地区等污染物的AQI数值，并支持按月份对污染物AQI数值进行可视分析，支持用户开展对大气污染时空分布模式、污染成因分析、污染类型分析、污染差异对比分析等分析任务。

整体系统显示图：



污染物AQI圆环直方图



(1) 视图设计

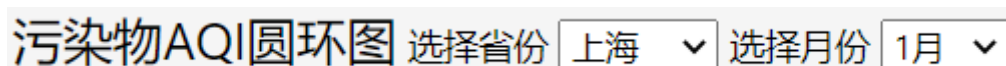
AQI圆环图用于展示某一个省份在某月份中的六项污染物对应AQI的变化情况。

本视图采用径向布局，展示每个月内污染物AQI数据的数值与变化情况。

环上被平均分为月份天数格，每格分别显示对应六种污染物AQI数值，同时直观通过柱状图展现每天内不同污染物的AQI值，并且能够通过柱状图高低看出当日何种污染物AQI值最高，以确定当日AQI数值。

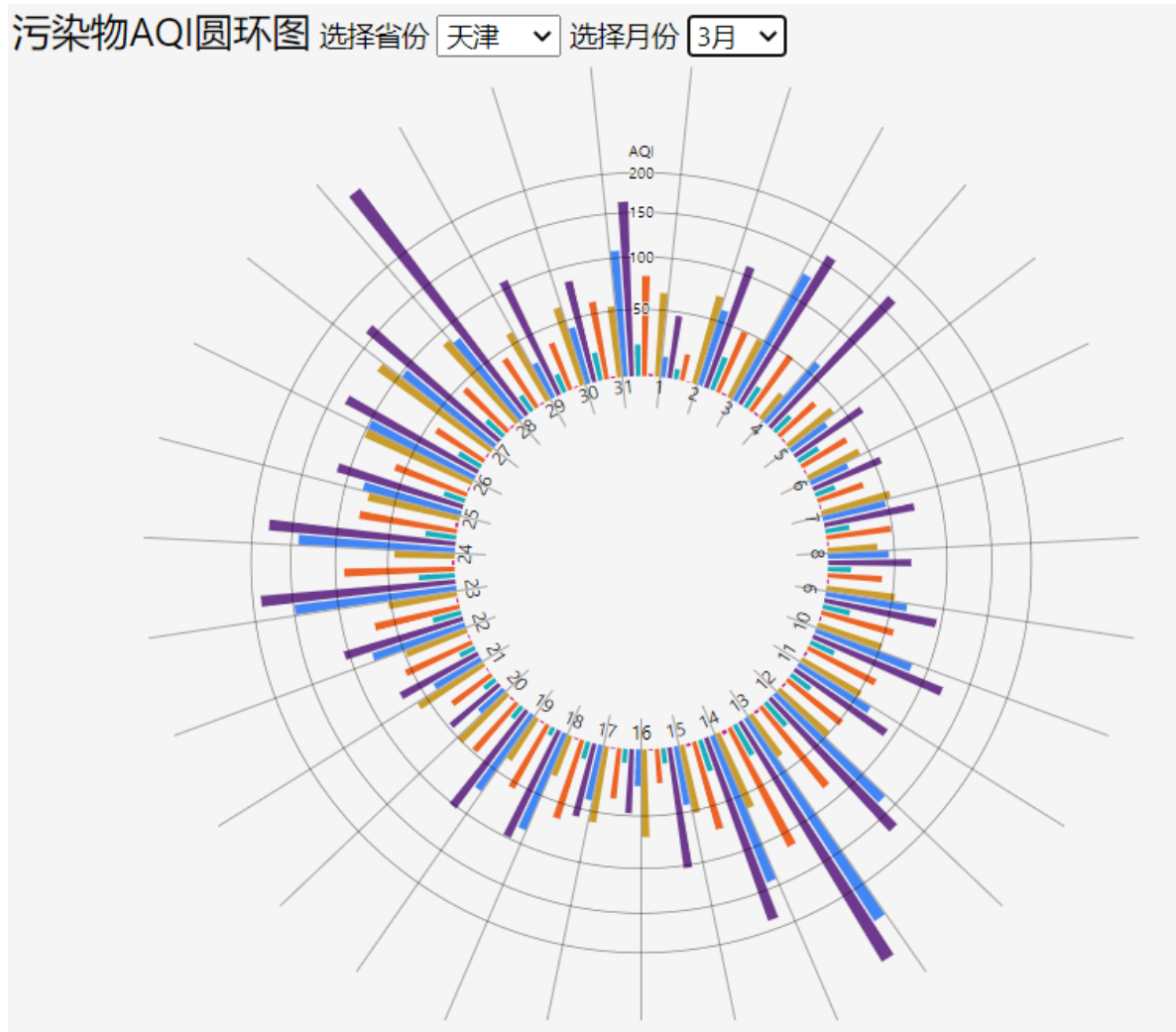
视图每格主要分为6根柱子，分别采用与整体污染物视觉编码相同的颜色。从日期开始数值到日期结束数值里依次展示 $PM_{2.5}$, PM_{10} , SO_2 , NO_2 , CO , O_3 六种污染物数值。同时对应纵轴提供对AQI数值的参考。

(2) 交互设计



用户通过切换省份以及月份可以观察不同数据，更新显示其中的多个视图。

月份省份切换显示：

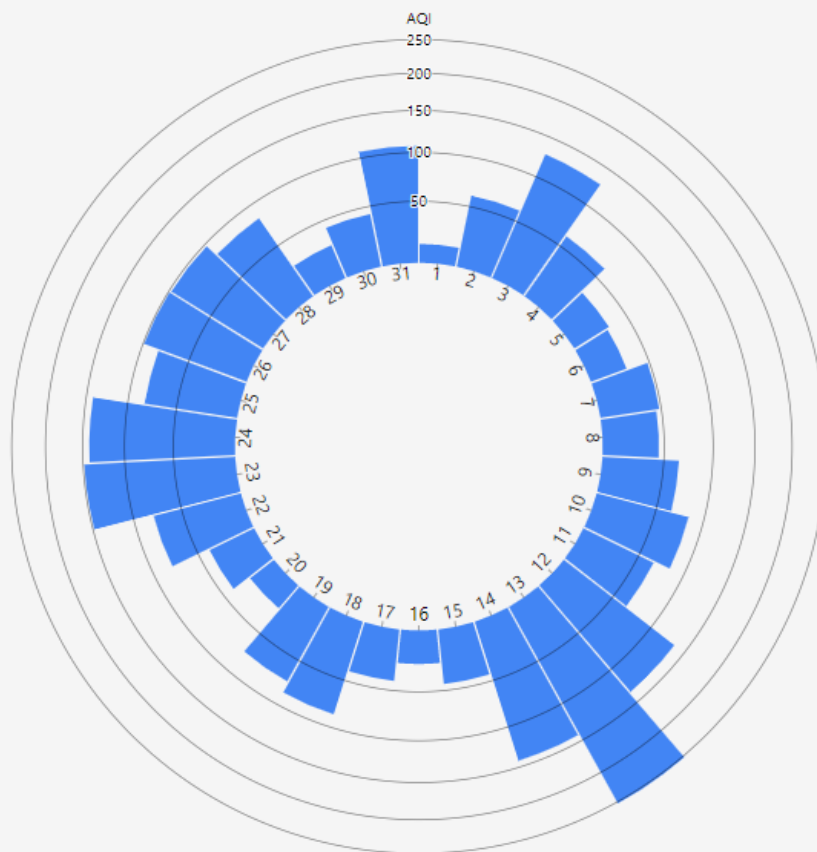


同时鼠标移动至不同的柱子会对应高亮，并且可以显示对应的AQI值。

(3)与污染物进行联动

选择PM2.5作为污染物时：

污染物AQI圆环图 选择省份 天津 选择月份 3月



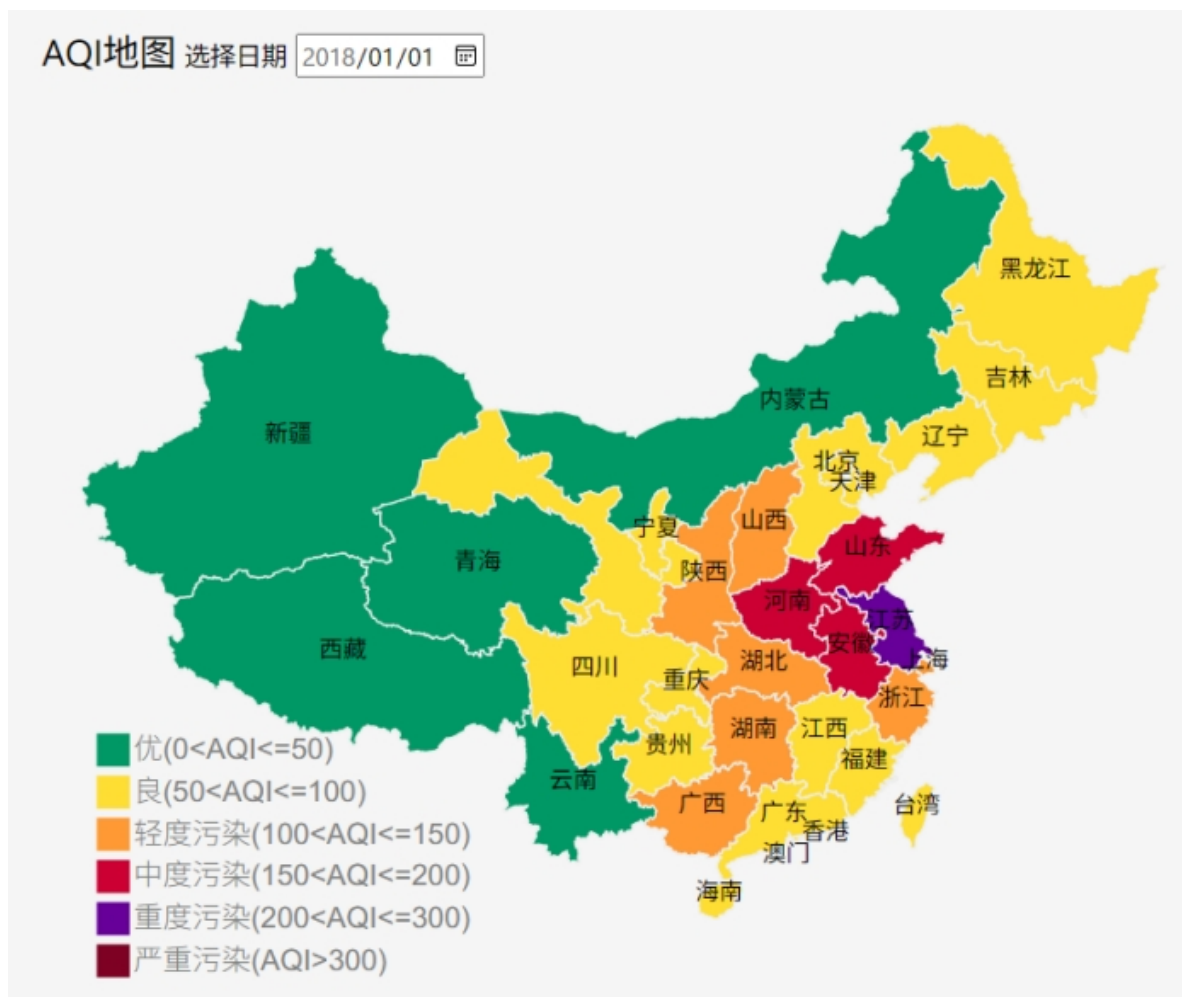
对应选择别的污染物也可以达到此效果。同时鼠标放置到对应柱子也可以显示AQI值。

(4) 代码实现

主要实现方法是利用d3提供的arc()函数设置内半径以及外半径，根据污染物排列顺序以及天数计算出污染物开始角度与结束角度，并利用返回值对path的d属性进行填充，对应地再补充x坐标轴以及y坐标轴即可。

污染物AQI地图

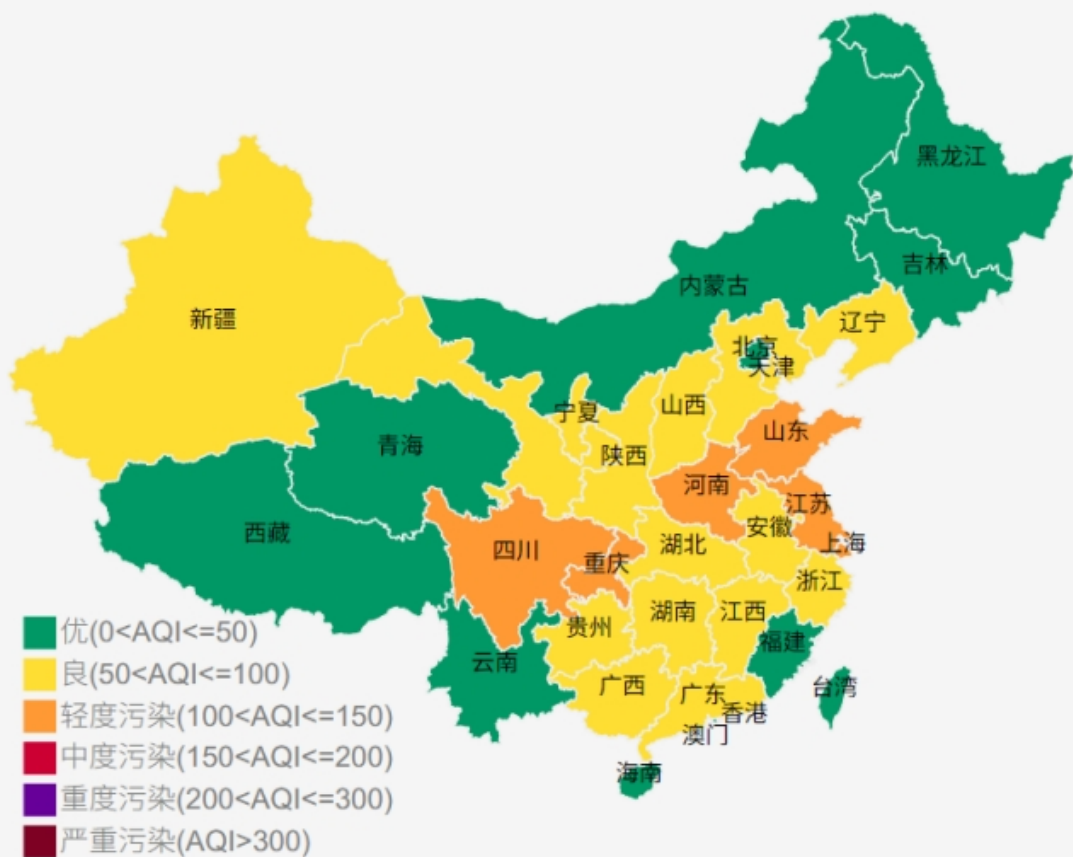
可视化与交互设计



AQI地图展示了2018年内某一特定日期的全国各省份的空气质量情况，根据AQI分级的颜色进行视觉映射。鼠标移动到某一省份上时将出现提示，内容为省份名称和AQI数值，同时省份的边界会加粗高亮。地图的左下角附有图例。通过上方的日期选择器可以更改日期，从而了解不同日期的空气质量情况。

(1)更改日期效果

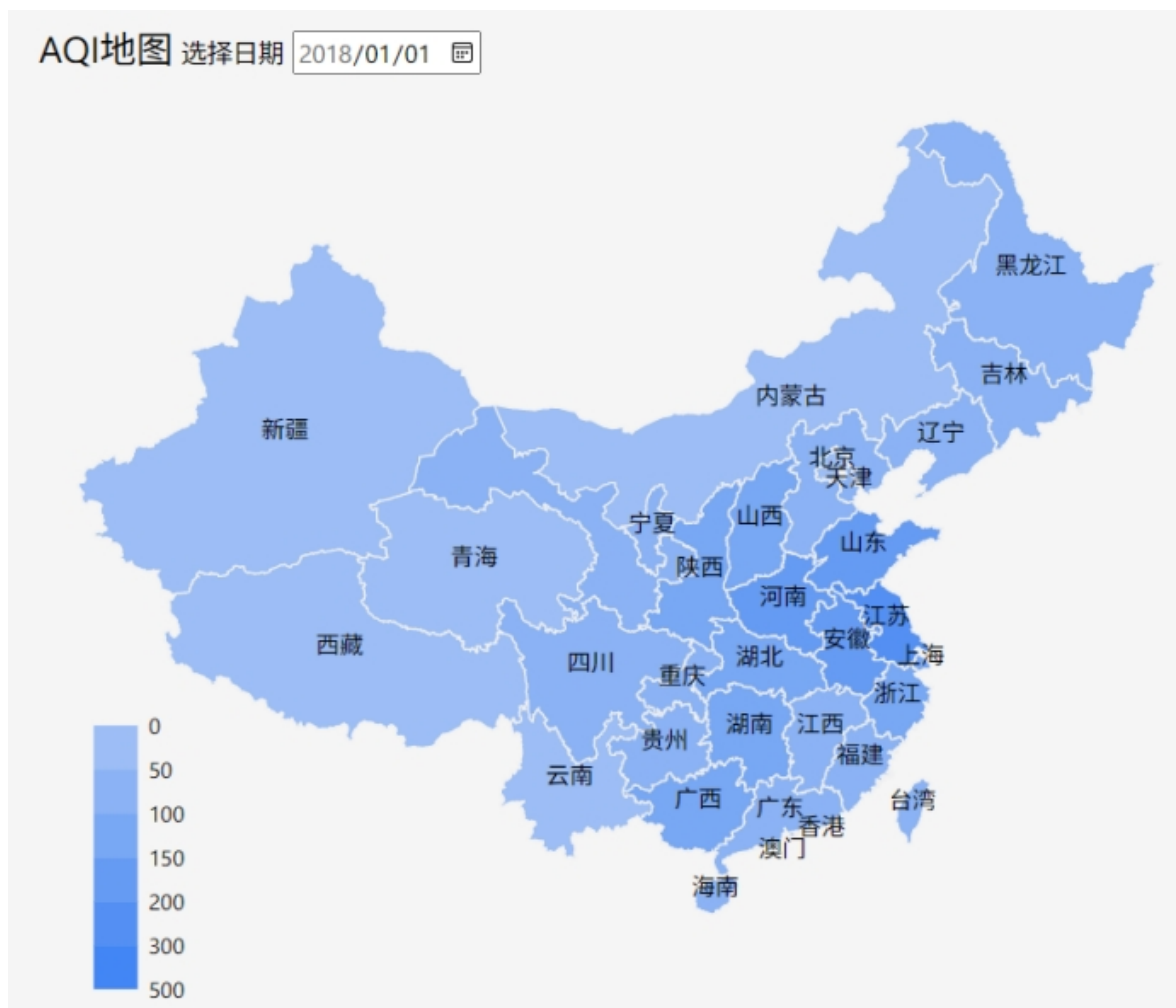
AQI地图 选择日期 2018/12/18



(2)与污染物选择的联动

当选择某一种污染物时，地图将同步切换为该种污染物的地理分布展示，同时左下角图例也切换为该种污染物的浓度换算的空气质量分指数。此时的视觉映射为颜色浓度，污染物浓度越高，颜色浓度也越高。

污染物选择PM2.5时效果：



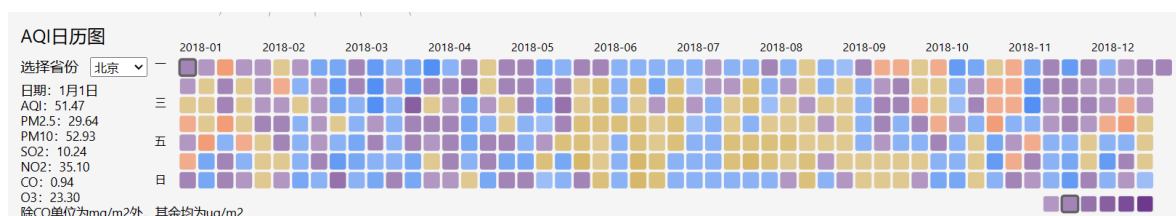
(3)代码实现

AQI地图使用墨卡托投影方式，关键函数为D3的`geoMercator()`和`geoPath()`函数，其中前者用于设置使用墨卡托投影方式，后者用于通过geojson数据生成地理边界的路径以作为`path`对象的属性。

从分省份空气质量数据中按照选定的污染物类型（未选择则为AQI数值）提取出对应的数据，并按照污染物类型和数值选择对应的颜色编码，生成SVG的`path`对象，并使用`stroke`属性设置描边。随后将省份的名称按照地理数据中的中心坐标生成`text`对象，其中为了防止遮挡使用`dx`和`dy`属性微调了若干省份名称的坐标位置。最后用`rect`对象和`text`对象绘制图例。

污染物AQI日历热力图

可视化与交互设计

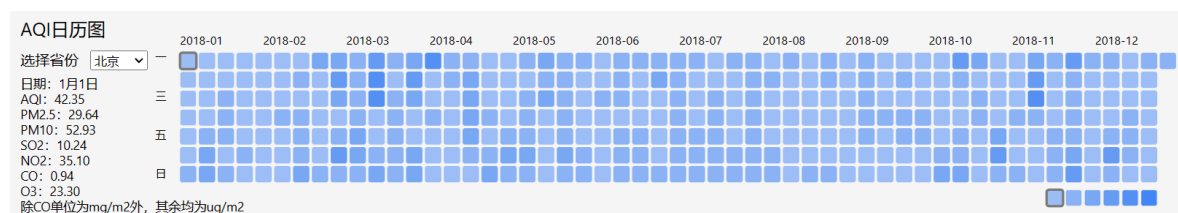


污染物AQI日历热力图展示了2018年某省份的空气质量情况，根据AQI分级以及主要污染物进行透明度和颜色的映射。鼠标移动到某一方格时左侧将显示出当天的具体信息，包括日期、AQI及各污染物浓度等，同时右下角图例中相同颜色的方块也会突出显示。通过上方的省份选择器可以更改省份，从而了解不同省份的空气质量情况。

(1)与污染物选择的联动

当选择某一种污染物时，日历图将同步切换为该种污染物的展示，同时左侧显示的AQI也为该污染物浓度计算出的AQI。此时的视觉映射为颜色浓度，污染物浓度越高，颜色浓度也越高。

污染物选择PM2.5时效果:



(2)代码实现

实现方法较为简单，首先绘制365个小方块，并给每个方块绑定每一天的数据，根据全局的污染物设置来选择颜色以及AQI。绑定 `mouseover` 与 `mouseout` 事件以实现 `hover` 效果，包括边框的设置，左侧信息的显示，以及图例中对应边框的显示。当全局污染物设置后，会修改颜色以及对应的AQI值。

AQI柱状图

可视化与交互设计

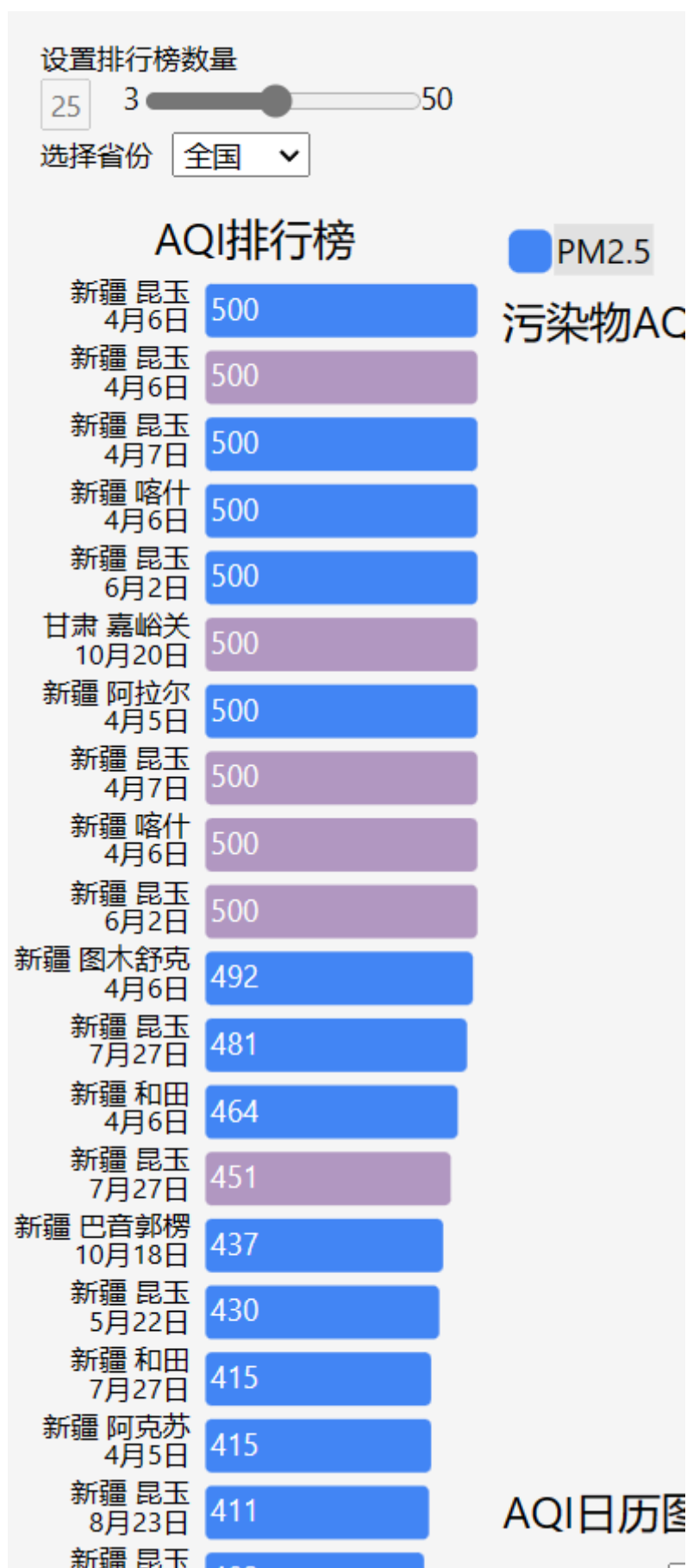


AQI柱状图展示了2018年按天计算的AQI排行情况，显示的信息包括AQI值、主要污染物及浓度、地区以及日期。其中主要污染物通过颜色来表示，浓度则通过 tooltip 实现。上方设置有排行数量以及省份的选择。当排行数量较大时，会将日期修改至 tooltip 中显示。

(1)与污染物选择的联动

当选择某一种污染物时，其余颜色将会设置透明度0.5，以凸显该污染物占比。

污染物选择PM2.5时效果:



(2)代码实现

实现方法也较为简单，首先需要根据图的大小计算每一个柱的高度，接着读取对应的AQI数据以及主要污染物来计算柱的长度和颜色，并将AQI值显示在柱上。当上方数量或省份切换发生改变时，会同步切换数据以显示对应省份和对应数量的数据。

讨论与总结

本作品使用2018年中国高分辨率大气污染再分析开放数据集，围绕分析大气污染时空分布模式、时空演变态势、识别主要污染源、分析污染成因、评估大气环境状况、发现特定事件等分析任务，对数据进行多方面处理，并最终设计实现了全国大气污染可视分析系统。

首先对数据进行逆地理编码、数据清洗、数据统计、IAQI与AQI等级计算等预处理操作。

使用AQI地图、污染物AQI圆环直方图、AQI排行榜、污染指标与AQI平行坐标等视图提供对污染时空分布模式与时空演变态势的分析；并结合灵活丰富的交互，用户利用该页面其他视图做更细粒度的兴趣点分析。并且增加对全局污染物设置分析的创新的设计与实现。

该可视化系统作品可以是相对于污染物的时空数据分析，同时根据污染物AQI数值的季节以及月份特性可以对应分析对应月份哪些污染物含量较高，哪些地区污染较严重，并根据此分析采取相关措施防治空气污染。