

数据可视化报告——北京市空气治理

田润泽-2023201906

乙巳年腊月十五

摘要

本文通过对北京市空气质量数据的分析，探讨了北京市空气质量的现状，以及北京市近年来空气治理的情况。通过对数据的处理和可视化，我们发现了一些有趣的现象，例如北京市空气质量的季节性变化，以及不同污染物之间的相关性。我们还通过对数据的分析，得到了空气质量的分布与变化，以帮助人们合理安排出行与出游。本报告的一切数据、代码与可视化成果均已同步至个人 github 平台：<https://github.com/Welldefine/Data-Visualization>

1 研究背景

近年来，北京市作为中国的首都，其空气质量问题备受社会各界关注。由于地理位置、快速的城市化进程以及能源消费结构的影响，北京长期以来面临着空气污染的严峻挑战，尤其是 PM2.5 和 PM10 等颗粒物污染问题。这不仅对市民的健康造成了威胁，也对城市的可持续发展和国际形象产生了重要影响。

为应对这一问题，北京市政府及相关部门采取了一系列治理措施，包括调整能源结构、优化交通管理、推进工业减排和增加绿化覆盖率等。在国家政策的支持下，北京还积极参与了区域联防联控，与周边城市协同治理大气污染。这些努力已取得显著成效，例如 2013 年至今，北京的年均 PM2.5 浓度逐步下降，空气质量优良天数显著增加。

然而，尽管空气质量治理取得了一定进展，北京的空气污染问题仍然存在一些复杂性和不确定性。例如，冬季采暖季的污染反弹、区域传输效应和气象条件对空气质量的影响等，仍是当前研究和治理的难点。

本研究以北京市空气质量的治理与现状为主题，通过数据可视化技术，分析北京市空气质量的特点，近年来北京市空气质量的改善趋势、治理措施的成效。通过直观的可视化展示，不仅可以帮助公众更好地理解空气质量治理的重要性，也为政府决策和未来研究提供参考依据。

2 研究目的

本研究旨在通过数据可视化技术，全面分析北京市近年来空气质量的治理成果与现状。

- 展示变化趋势通过对北京市近十年来空气质量数据（如 PM2.5、PM10、SO2、NO2 等指标）的分析，直观呈现空气质量的总体变化趋势，探究不同时间节点的治理成效与环境改善情况。
- 展示空气质量特征通过对北京市空气质量数据的分析，展示不同季节、不同区域的空气质量特征，探究空气质量的时空分布规律。

3 数据说明与数据来源

数据来源：北京市环境保护检测中心 (2013.12.06–2024.11.16) - 王晓磊

3.1 数据描述

本研究所使用的数据集包括 2014 年到 2024 年北京市空气质量数据，总计包含 11 个文件夹，每个文件夹对应一年内的空气质量记录。具体数据内容如下：

- 文件结构：每个文件夹包含逐日空气质量数据（1 月 1 日至 12 月 31 日）。
- 主要文件：
 - `beijing_all_20xxxxxx.csv`: 包括 PM2.5、PM10、AQI 等主要空气质量指标。
 - `beijing_extra_20xxxxxx.csv`: 包括 SO2、NO2、CO、O3 等附加空气质量指标。
- 数据格式：CSV 文件。第一行为列名，包括日期、小时、数据类型以及各监测点名。以下每行为某时刻的一种类型数据。

3.2 数据类型

数据集中所包含的空气质量指标及其单位如表 1 所示。

表 1: 数据类型及单位

Type	数据类型	单位
AQI	AQI 实时值	N/A
PM2.5	PM2.5 1 小时均值	$\mu\text{g}/\text{m}^3$
PM2.5_24h	PM2.5 24 小时均值	$\mu\text{g}/\text{m}^3$
PM10	PM10 1 小时均值	$\mu\text{g}/\text{m}^3$
PM10_24h	PM10 24 小时均值	$\mu\text{g}/\text{m}^3$
SO2	SO2 1 小时均值	$\mu\text{g}/\text{m}^3$
SO2_24h	SO2 24 小时均值	$\mu\text{g}/\text{m}^3$
NO2	NO2 1 小时均值	$\mu\text{g}/\text{m}^3$
NO2_24h	NO2 24 小时均值	$\mu\text{g}/\text{m}^3$
O3	O3 1 小时均值	$\mu\text{g}/\text{m}^3$
O3_24h	O3 24 小时最大值	$\mu\text{g}/\text{m}^3$
O3_8h	O3 8 小时滑动均值	$\mu\text{g}/\text{m}^3$
O3_8h_24h	O3 8 小时滑动均值的 24 小时最大值	$\mu\text{g}/\text{m}^3$
CO	CO 1 小时均值	mg/m^3
CO_24h	CO 24 小时均值	mg/m^3

3.3 数据样例

以下是数据文件的样例：

date, hour, type,	东四, 天坛, 宫园, 万寿西宫, ...
20131205, 1, PM2.5,	93, 93, 63, 79, ...
20131205, 1, PM2.5_24h,	93, 108, 99, 123, ...
20131205, 1, PM10,	103, 124, 81, 107, ...
20131205, 1, PM10_24h,	97, 130, 122, 141, ...
20131205, 1, AQI,	123, 141, 130, 161, ...

4 数据清洗

4.1 清洗结果

output_data 中包含了数据的初步清洗和分区整理后得到两个子文件夹：

- **first_clean:** 该文件夹将同一年的 _all 数据和 _extra 数据进行了合并，生成了初始数据集，为后续分析奠定了基础。

- **by_district:** 该文件夹下包含以下七个子文件夹, 对应 AQI, CO, NO2, O3, PM2.5, PM10, SO2 七类空气质量指标的数据。这些数据按时间维度（逐年、逐月、逐小时）进行了整理，并完成了以下清洗工作：
 - 检测并处理缺失值。
 - 识别并修正异常值。
 - 按照北京市行政区划（如东城区、西城区等）对监测站点数据进行了分类汇总。

4.2 数据格式

清洗后的数据以 CSV 格式存储，文件命名格式为 Beijing_AirQuality_YYYYXX_days.csv，其中 YYYY 表示年份，XX 表示空气质量指标（如 AQI、PM2.5 等）。具体数据格式说明如下：

- 第一行为列名，包括日期和各行政区名称（如东城区、西城区等）。
- 以下每行为某日期某指标的观测数值。

4.3 数据样例

以下是清洗后数据文件的样例 (Beijing_AirQuality_2014AQI_days.csv):

date,	东城区 ,	西城区 ,	海淀区 ,	...
20140101,	88.198,	90.625,	105.33,	...
20140102,	119.354,	112.277,	128.236,	...
20140103,	123.979,	123.264,	146.125,	...
20140104,	105.813,	102.916,	104.569,	...
20140105,	172.614,	102.917,	104.569,	...

5 数据可视化

5.1 空气治理成效可视化

5.1.1 年均 AQI 逐年变化情况

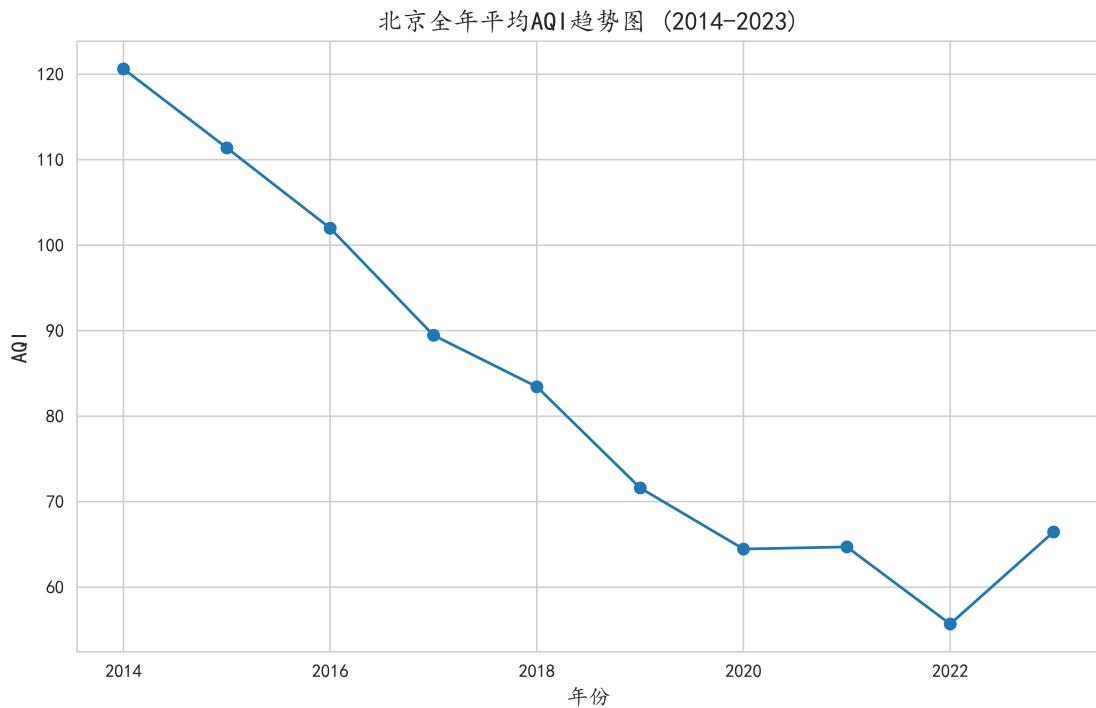


图 1: 北京市年均 AQI 逐年变化情况

图 1 展示了 2014 年至 2023 年北京市全年平均空气质量指数 (AQI) 的变化趋势。从图中可以看出，北京市的空气质量在过去十年中呈现出明显的改善趋势。

在 2014 年，北京市的全年平均 AQI 为 120 左右，显示出较高的空气污染水平。然而，随着时间的推移，AQI 值逐年下降，到 2018 年降至约 85。这一趋势表明北京市在空气质量改善方面取得了显著进展。尽管在 2023 年 AQI 值出现了轻微的波动，但总体上仍保持在较低水平，显示出空气质量的持续改善。2021 年，AQI 值下降至约 55，达到十年来的最低点。

5.1.2 主要污染物的逐年变化情况

图 2 详细描绘了 2015 年至 2023 年北京市六种主要空气污染物浓度相对于前一年变化的百分比。这些污染物包括一氧化碳 (CO)、二氧化氮 (NO₂)、臭氧 (O₃)、可吸入颗粒物 (PM10)、细颗粒物 (PM2.5) 和二氧化硫 (SO₂)。图中的数据并非直接表示污染物的浓度，而是展示了每种污染物浓度相对于前一年的变化百分比，这有助于更直观地理解污染物浓度的增减趋势。

从图中可以观察到，臭氧 (O₃) 的变化百分比相对稳定，二氧化硫 (SO₂) 的变化百分比

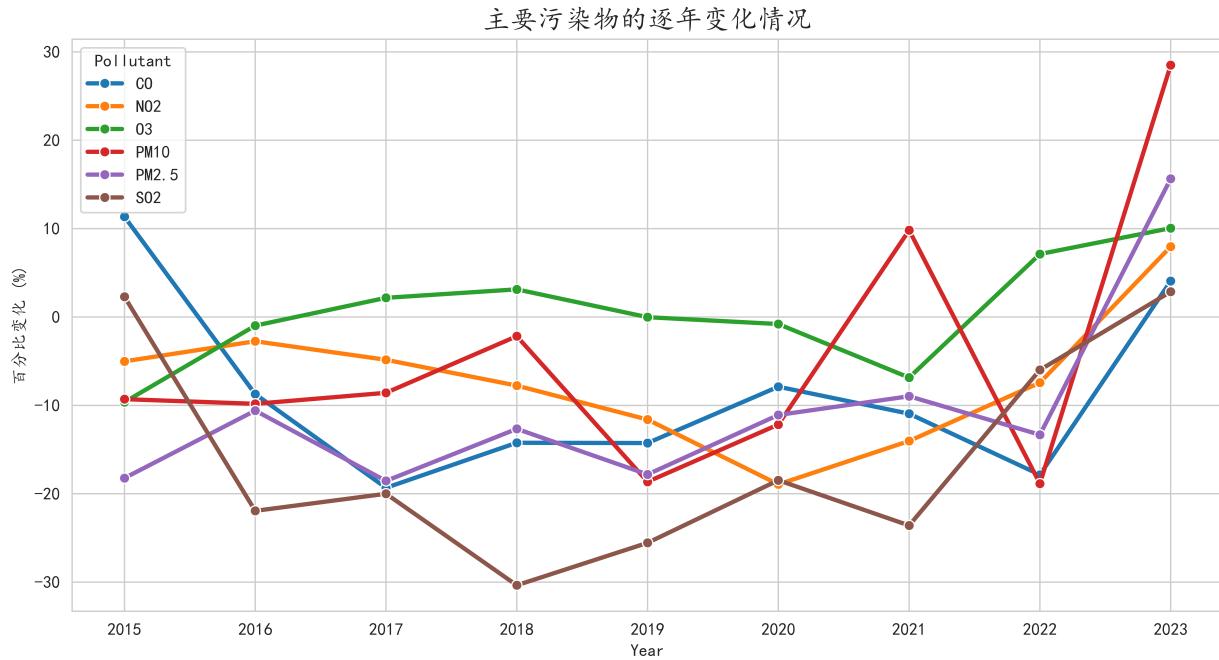


图 2: 北京市主要污染物浓度逐年变化情况

几乎一直保持为负值，表明其浓度持续降低，而其他污染物也在大多年份表现出负的百分变化，说明了其浓度几乎一直在降低，反映了北京市空气质量的持续改善。

5.1.3 近年来全年空气质量情况

图 3 展示了 2016 年、2018 年、2020 年和 2023 年北京市全年空气质量情况的日均 AQI 热力图。图中使用了从浅绿色到深蓝色的渐变色来表示不同的 AQI 值。浅绿色代表较低的 AQI 值，而深蓝色则代表较高的 AQI 值。横轴表示月份，从一月到十二月；纵轴表示一周中的每一天，从星期一到星期日。并且图上展示了展示了四个不同年份（2016 年、2018 年、2020 年和 2023 年）的空气质量数据，便于观察和比较不同年份之间的空气质量变化。

从图中可以看出，北京市的空气质量在不同季节存在明显差异。通常，冬季（特别是 12 月和 1 月）的空气质量较差，这可能与供暖季节的燃煤排放有关。相比之下，夏季（特别是 6 月、7 月和 8 月）的空气质量较好，这可能与较高的温度和较强的大气扩散条件有关。

通过比较不同年份的数据，可以观察到北京市空气质量的改善趋势。例如，2023 年的热力图显示出更多的浅绿色区域，表明这一年的空气质量整体上有所改善。2016 年的图中深蓝色区域较多，表明那一年的空气质量相对较差。随着时间的推移，深蓝色区域逐渐减少，显示出空气质量的持续改善。

图中还显示出一周内空气质量的变化。通常，周末的空气质量相对较好，这可能与工作日的交通排放和工业活动有关。

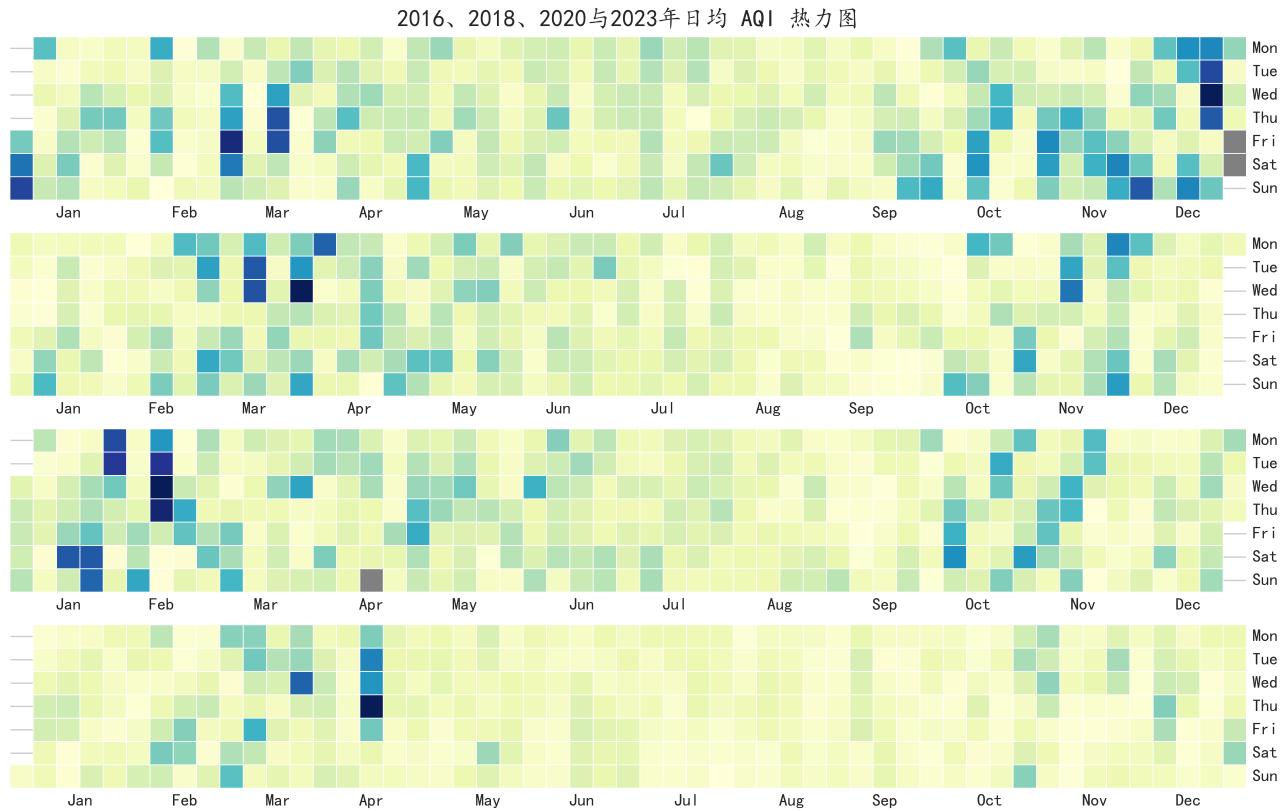


图 3: 近年来北京市全年空气质量情况

5.2 北京市空气质量的时空特征

5.2.1 AQI 的分布曲线

图 4 展示了北京市空气质量指数 (AQI) 的分布曲线。该图是一个直方图，横轴表示 AQI 值，纵轴表示频数，即每个 AQI 区间内出现的次数。

图中的分布曲线呈现出右偏的形态，即大部分数据集中在较低的 AQI 值区域，随着 AQI 值的增加，频数迅速下降。这种分布形态表明北京市的空气质量大多数时间处于较好的状态，高 AQI 值的情况较少。

5.2.2 北京市空气质量的区域特征

图 5 是一个箱线图，展示了 2014 年至 2023 年北京市不同区域的空气质量指数 (AQI) 分布情况。图中展示了北京市多个区域的 AQI 分布情况，包括东城区、西城区、朝阳区、海淀区、丰台区、石景山区、房山区、大兴区、通州区、顺义区、昌平区、门头沟区、平谷区、怀柔区、密云区、延庆区以及全市平均 (mean_aqi)。每个区域的 AQI 分布通过一个箱线图表示，可以直观地比较不同区域之间的空气质量差异。

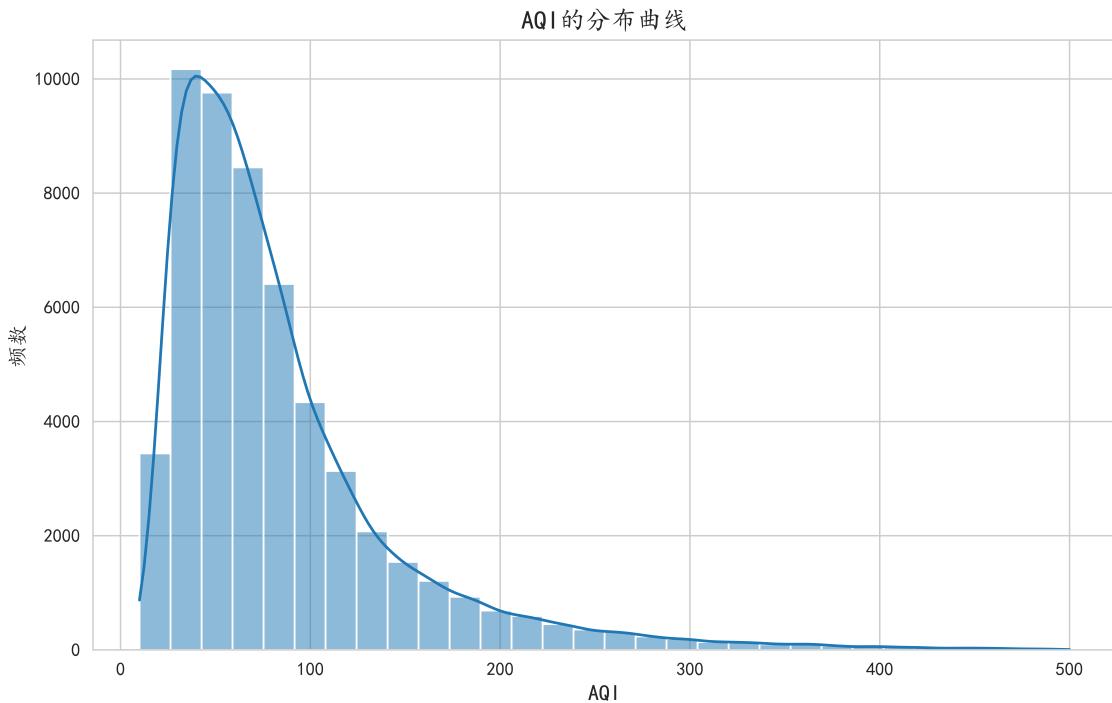


图 4: 北京市 AQI 分布曲线

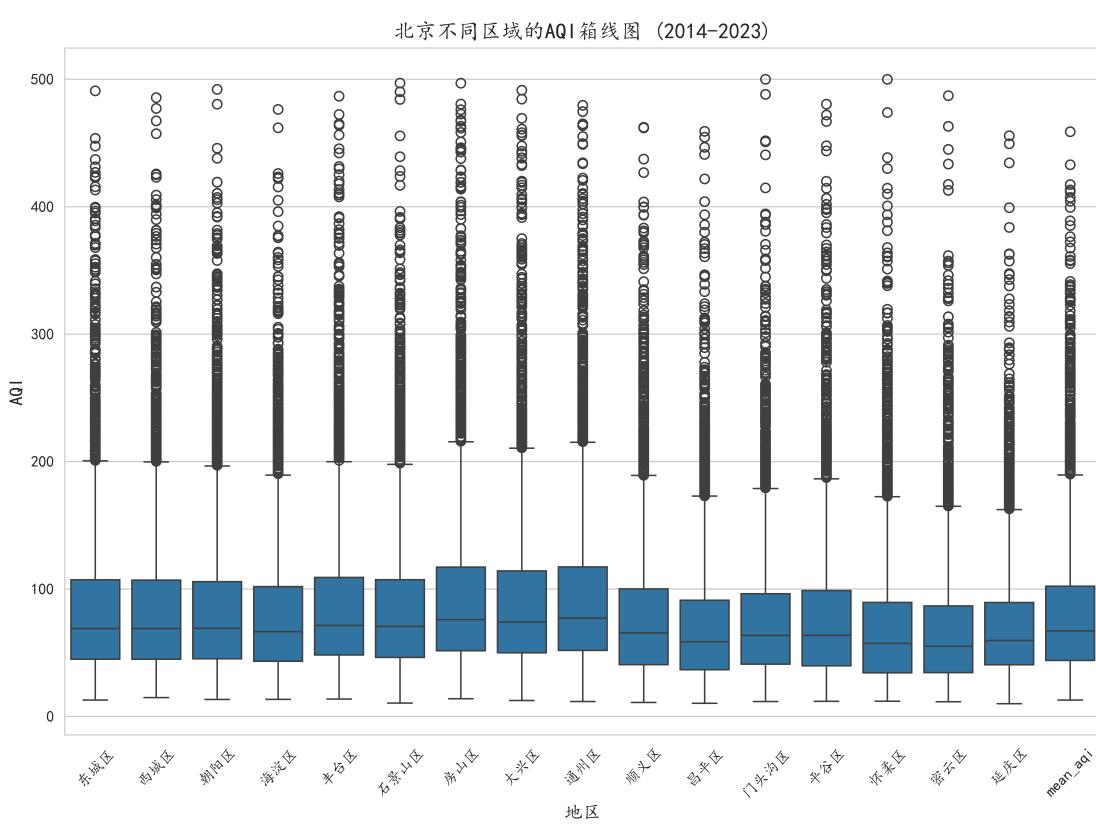


图 5: 北京市不同区域 AQI 箱线图

图 6 是一个柱状图，展示了 2014 年至 2023 年北京市各区域强污染的天数。图中每个柱子代表一个区域，柱子的高度表示该区域在这段时间内 AQI 大于 300 的天数，即强污染天数。图中还标出了平均值为 51.75 天的参考线，用蓝色虚线表示。

从图中可以看出，不同区域的强污染天数存在显著差异。一些区域的柱状较高，表明这些区域在 2014 年至 2023 年间经历了更多的强污染天数。例如，大兴区、通州区和房山区的柱状明显高于其他区域，表明这两个区域的强污染天数较多。推测可能与当地的产业结构相关。

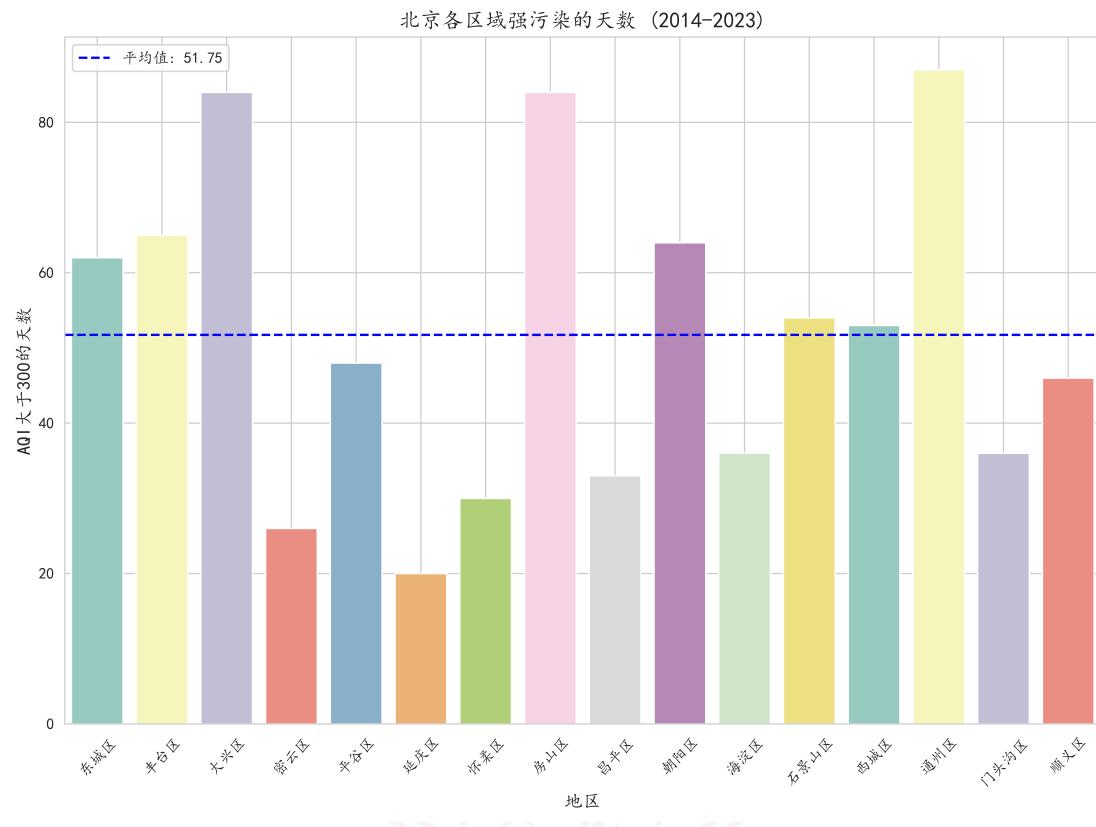


图 6: 北京市各区域强污染的天数

5.2.3 北京市空气质量的时间特征

图 7 展示了一年四季中不同季节空气质量指数 (AQI) 在一天内的变化情况。图中用不同颜色的线条分别表示了春季 (Spring)、夏季 (Summer)、秋季 (Autumn) 和冬季 (Winter) 的 AQI 变化趋势。

从图中可以看出：春季和冬季的 AQI 值在一天中的变化趋势相似，均在早晨和晚上较高，而在下午时段较低。夏季的 AQI 值在一天中的变化较为平稳，但在下午时段略有上升。秋季的 AQI 值在一天中的变化趋势与春季和冬季相似，但在下午时段的下降幅度较小。

所有季节的 AQI 值在早晨时段（大约 5 点到 8 点）达到较高水平，这可能与交通高峰期的排放有关。在下午时段（大约 14 点到 16 点），除了夏季外，其他季节的 AQI 值都有所下降，这

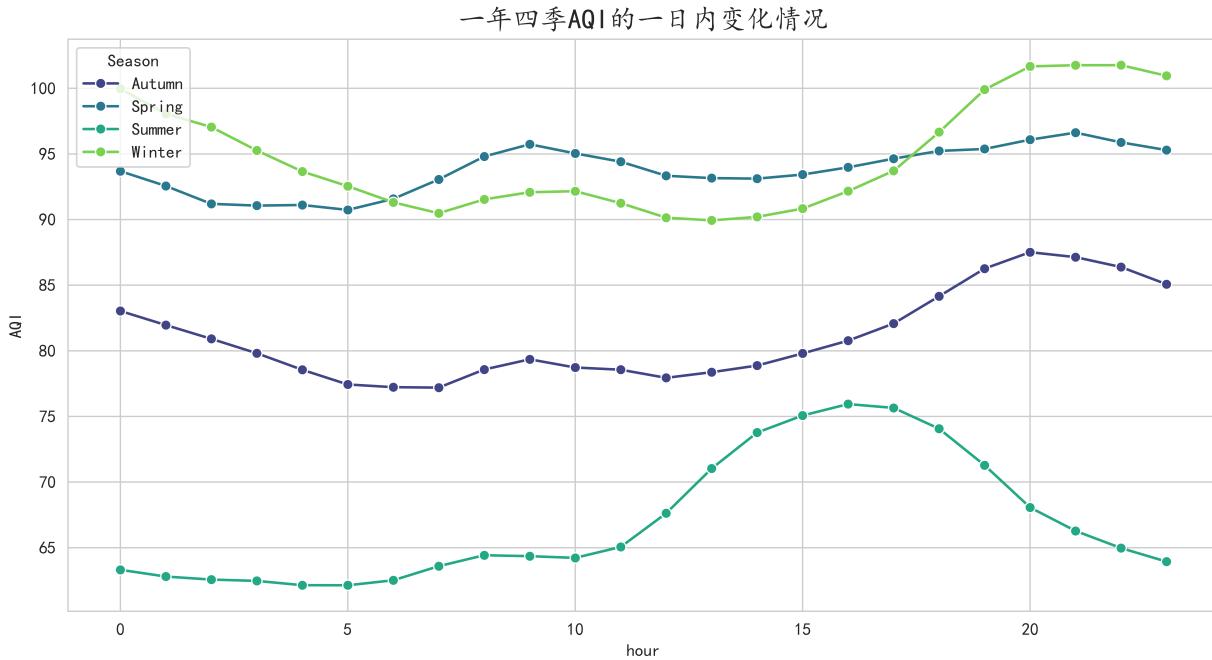


图 7: AQI 一日内变化

可能与日间的大气扩散条件较好有关。晚上时段（大约 20 点以后），AQI 值再次上升，可能与夜间供暖、烹饪等活动有关。

5.2.4 北京市空气质量的时空相关性

图 8 是一个时空聚类热力图，展示了北京市不同区域在不同月份的空气质量情况。热力图通过颜色的深浅来表示空气质量指数（AQI）的高低，颜色越深表示 AQI 值越高，即空气质量越差。

从图中可以看出，不同区域和不同月份的空气质量存在显著差异。而且根据聚类结果，我们可以大致得到和之前类似的结论：夏季的空气质量相对较好，冬季的空气质量相对较差；中心城区的空气质量相对较好，郊区的空气质量相对较差。而且我们还能够看出，具体每一个城区，在哪些月份的空气质量相对较差，这对于未来的空气质量治理有一定的参考意义。

5.2.5 AQI 与主要污染物的相关关系

图 9 是一个散点图矩阵，展示了空气质量指数（AQI）与不同主要污染物之间的相关关系。这种图表形式有助于观察多个变量之间的两两关系，包括它们之间的相关性和分布特征。

对角线上的每个图是一个变量的直方图，展示了该变量的分布情况。例如，AQI 的直方图显示了 AQI 值的分布，其他污染物（如 CO、NO₂、O₃ 等）的直方图也类似。非对角线上的每个图是两个变量之间的散点图，展示了这两个变量之间的关系。例如，AQI 与 CO 之间的散点图显示了这两个变量之间的关系。

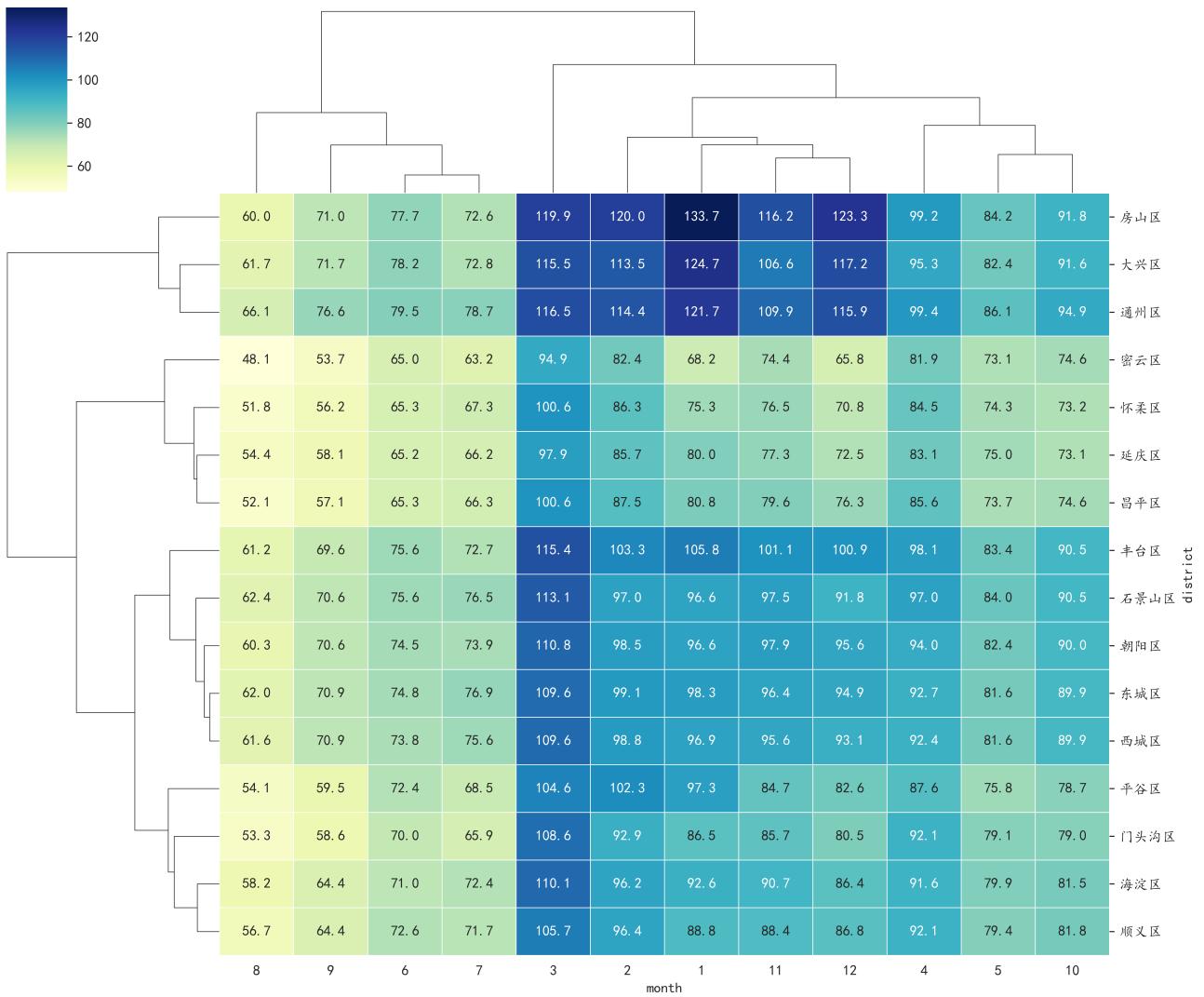


图 8: 北京市空气质量的时空聚类热力图

通过观察散点图的分布，可以初步判断两个变量之间是否存在相关性。如果散点图中的点大致沿着一条直线分布，那么这两个变量之间可能存在线性相关性。例如，AQI 与 PM2.5 之间的散点图显示出明显的正相关趋势，那么可以推断 AQI 的增加可能与 PM2.5 浓度的增加有关。

借助该图，我们可以快速识别变量之间的关系，包括相关性、分布特征和异常值。从而进一步理解空气质量的成因、评估污染物的影响以及制定空气质量管理策略。

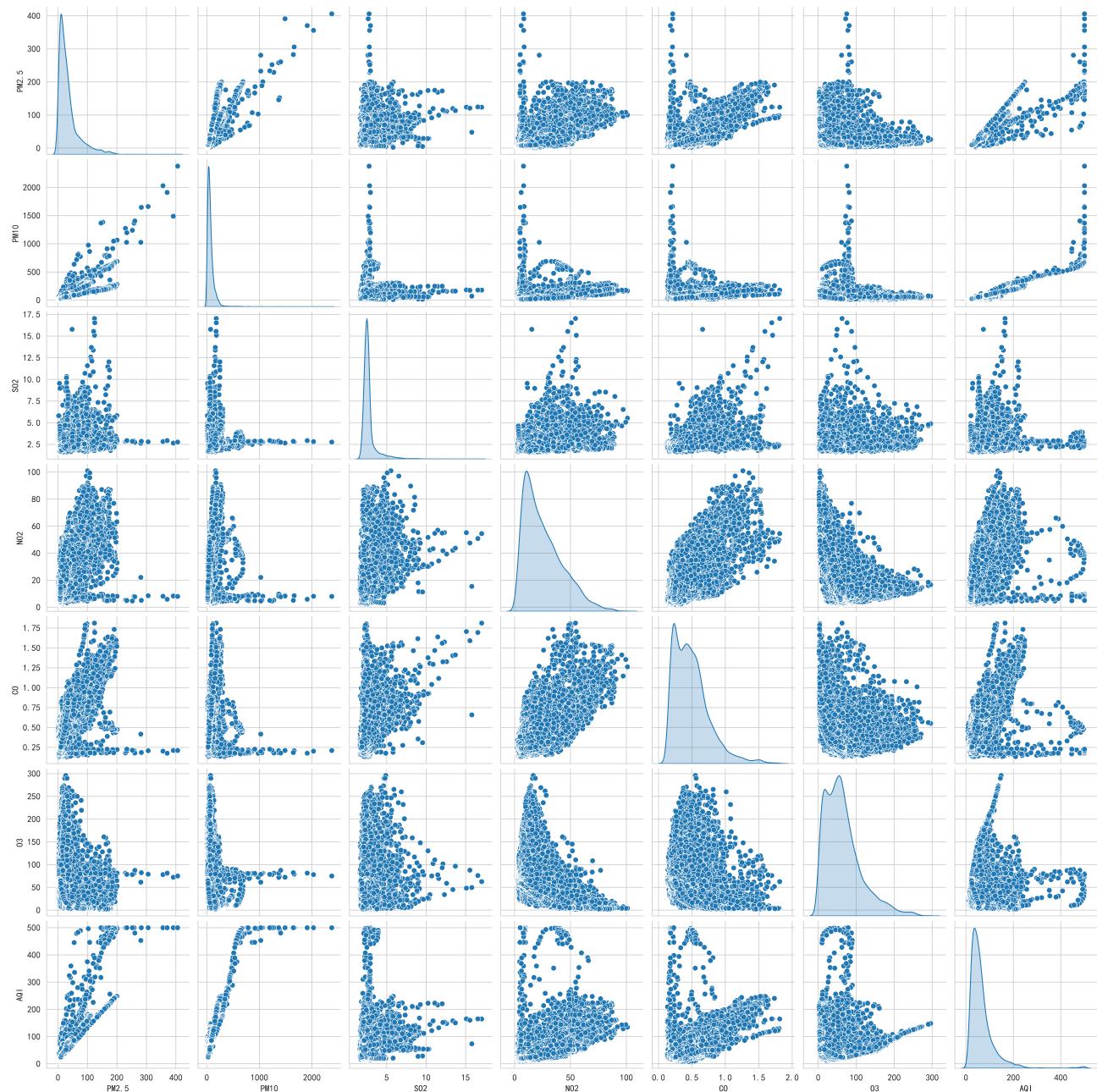


图 9: AQI 与主要污染物的相关关系

6 结论

本研究通过对北京市近年来空气质量数据的分析，探讨了北京市空气质量的现状和治理成效。通过数据的处理和可视化，我们发现了一些有趣的现象，例如北京市空气质量的季节性变化，不同区域空气质量的差异，以及不同污染物之间的相关性。

具体来说，我们得出了以下结论：

- 北京市的空气质量在过去十年中呈现出明显的改善趋势，AQI 值逐年下降。
- 北京市的主要污染物浓度在过去十年中也呈现出下降趋势，尤其是 SO₂ 的浓度持续降低。
- 北京市的空气质量在不同季节和不同区域存在显著差异，冬季和郊区的空气质量相对较差。
- 北京市的空气质量在一天内也存在明显变化，早晨和晚上的 AQI 值较高，下午较低。
- AQI 与主要污染物之间存在一定的相关性，例如 AQI 与 PM_{2.5} 之间存在明显的正相关趋势。

通过对数据的分析，我们可以更好地理解空气质量的时空特征，为未来的空气质量治理提供参考依据。本报告的一切数据、代码与可视化成果均已同步至个人 github 平台：<https://github.com/Welldefine/Data-Visualization>

7 致谢

感谢北京市环境保护检测中心和王晓磊提供的数据，为本研究提供了重要的数据支持。感谢北京市政府及相关部门对空气质量治理工作的支持和努力。感谢数据可视化技术为我们提供了一种直观、有效的分析工具。