Covid and Air Quality

Big Data Project Group 10:

Vin Liu   Lillian Zhang   Hang Zhang

https://github.com/vin-lz/cs-gy-6513-big-data-project

INTRODUCTION

◎ In this project, we utilized Air Quality Index, and six Air Quality Parameters to measure air quality.

◎ Analyzed the data and map them to major events during the pandemics, such as major lockdowns in the cities around the world.

◎ Explored the change of air quality correlated with the lockdowns.

SOME PRIOR KNOWLEDGE

◎ AQI stands for Air Quality Index, from 0 to 500, the greater the value, the higher level of air pollution and more of health concern.

◎ The AQI is divided into six categories, with color coded. Each category corresponds to a different level of health concern.

◎ Air Quality Parameters (PM2.5, PM10, Ozone, CO, NO_2, and SO_2) are used interchangeably with Six Common Pollutants.
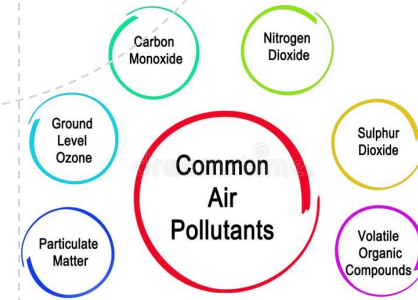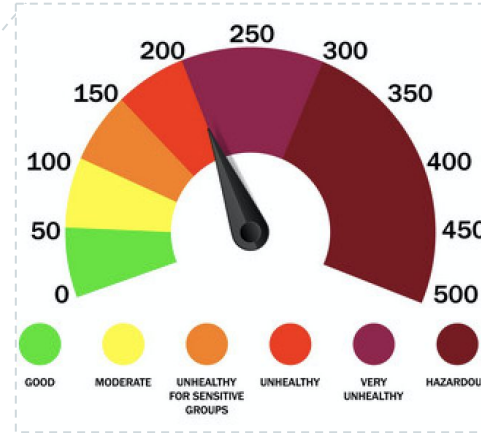
QUESTIONS TO BE ANSWERED

1. Is there a significant change in the air quality?
2. Did air quality get better or worse, and the reasons behind?
3. Are the changes in a similar pattern?
4. Which areas exhibited a significant reduction of certain pollutants during the lockdown?
5. Other affecting factors?

Gathering data

Processing data

Analysis data
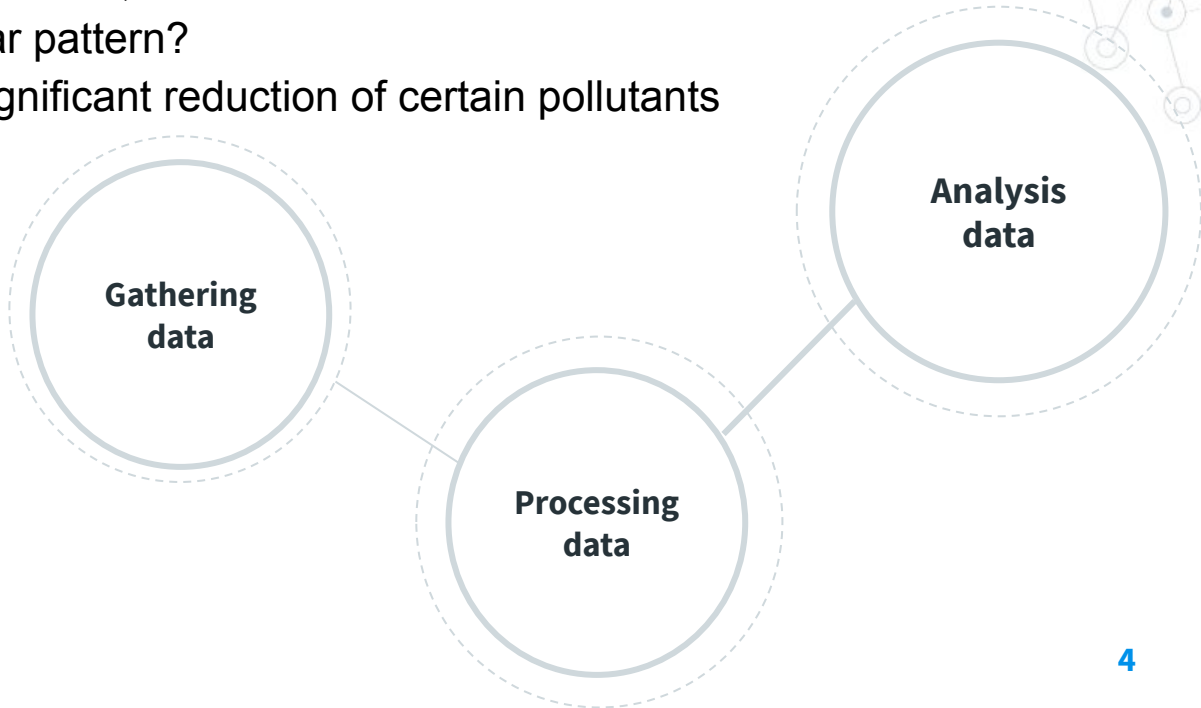
DATASETS

1. World Air Quality Open Data Platform (WAQI)
2. United State Environmental Protection Agency (EPA) Outdoor Air Quality Data
3. Oxford COVID-19 Government Response Tracker (Oxford)
4. China Air Quality Historical Data
5. Air Quality Improvement under COVID Visualized
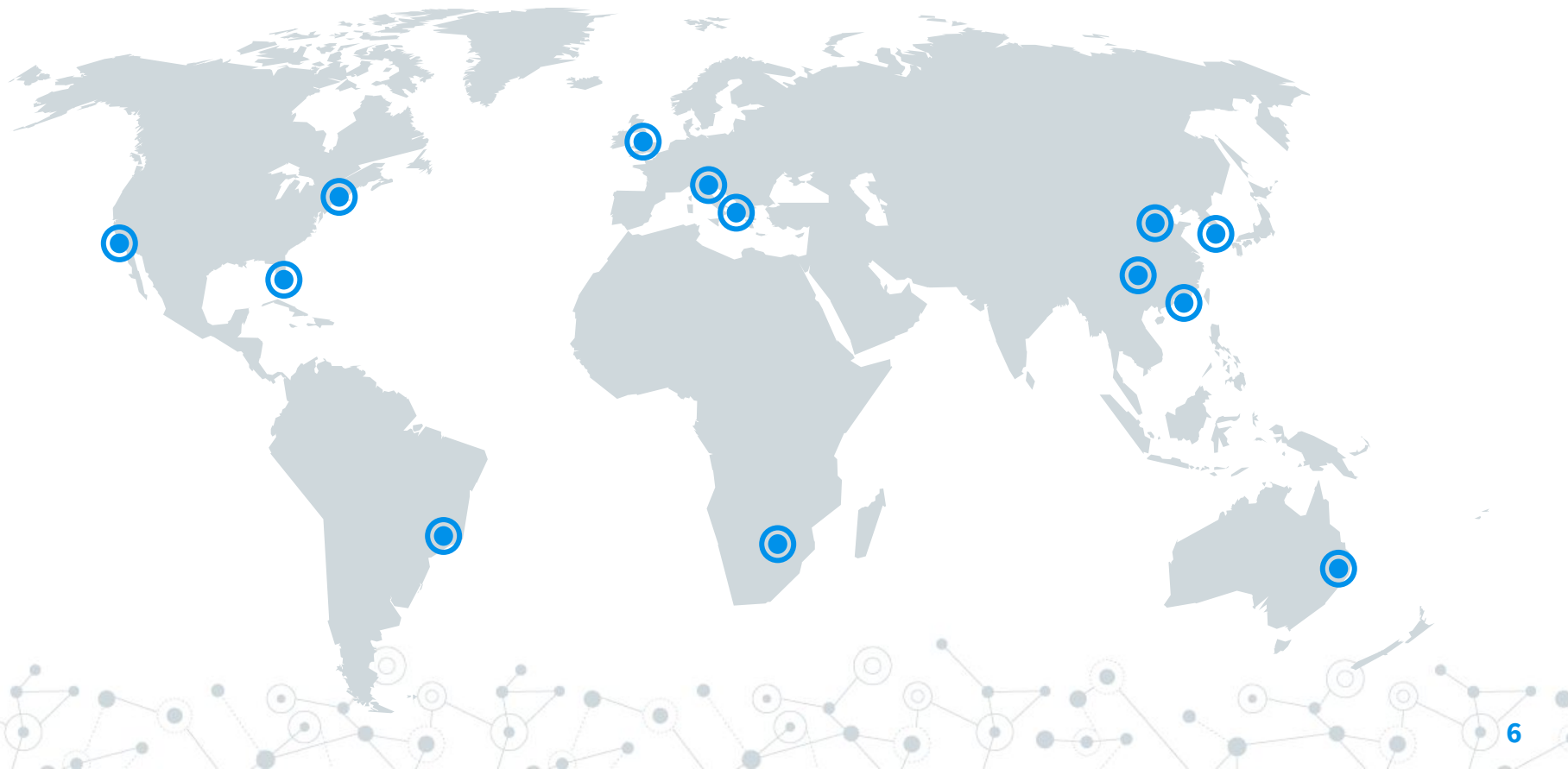6. US/Local COVID Policy reference

Available from:

https://aqicn.org/data-platform/COVID19/.
https://www.epa.gov/outdoor-air-quality-data/download-daily-data.
https://quotsoft.net/air/
https://github.com/ncydexter/Air-Quality-Improvement-under-COVID-Visualized.
https://github.com/OxCGRT/COVID-policy-tracker/blob/master
https://raw.githubusercontent.com/OxCGRT/USA-COVID-policy/master

Air quality data around the world

METHODS

HEATMAP

TREEMAP

BARPLOT

Data Analysis and Visualization

Analysis in three groups
1. Cities with clear lockdown dates: NYC, Wuhan
2. Cities with multiple-lockdown: LA, London
3. Outlier: Miami
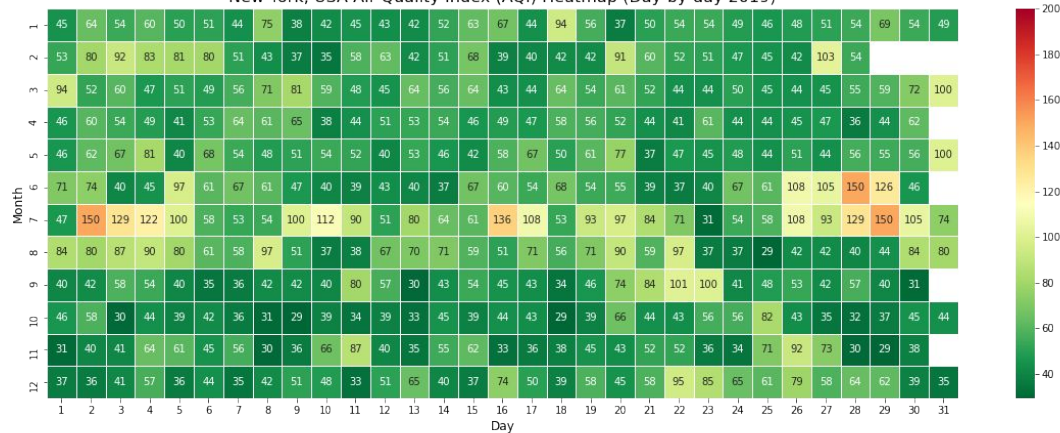
Let's zoom into New York City

Heatmap   **Bar chart**   **Tree map**
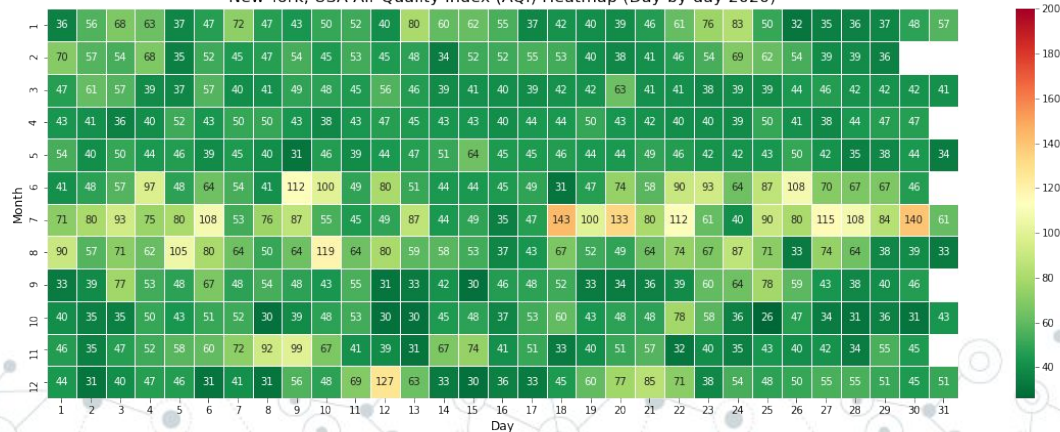
Heatmap

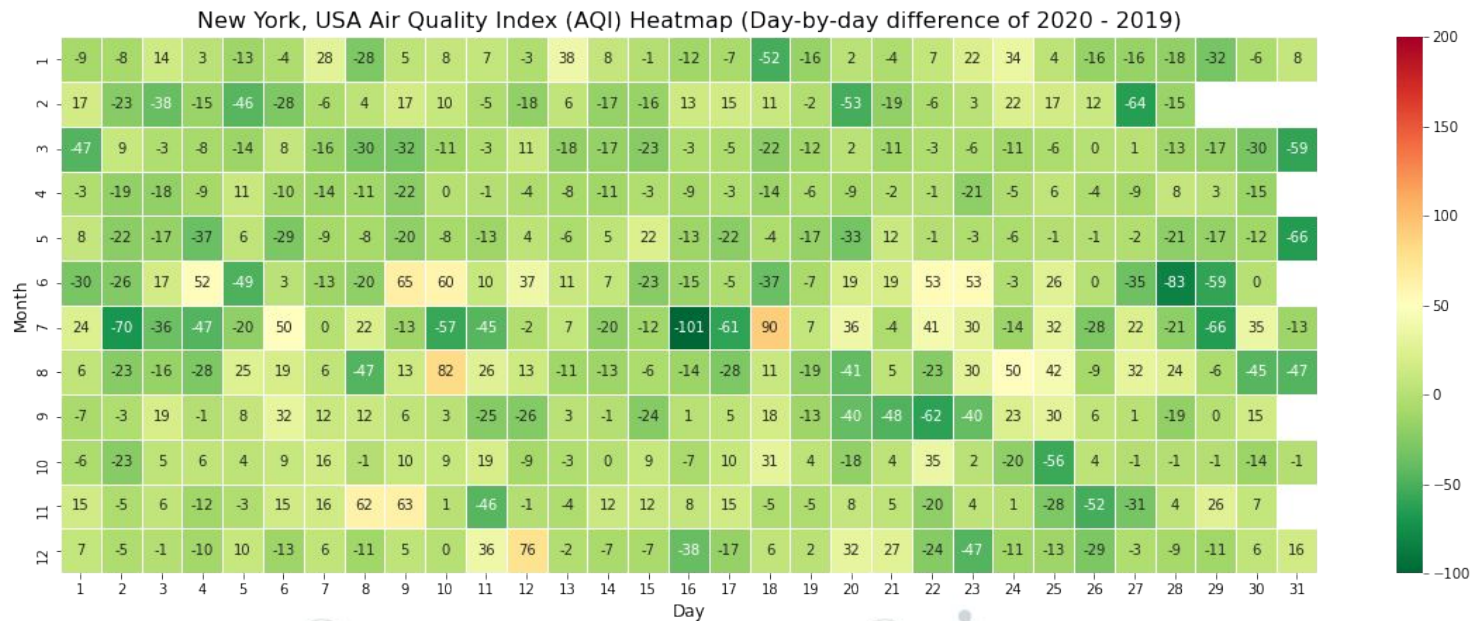New York, USA Air Quality Index (AQI) Heatmap (Day-by-day 2019)

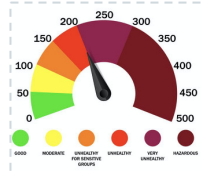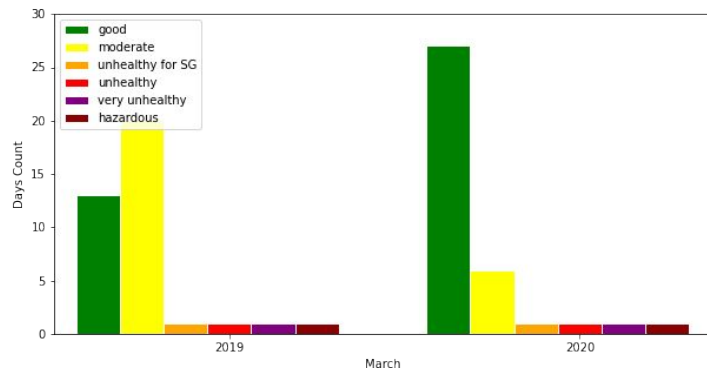New York, USA Air Quality Index (AQI) Heatmap (Day-by-day 2020)

Heatmap (difference)

AQI of 2020 minus AQI of 2019

New York, USA Air Quality Index (AQI) Heatmap (Day-by-day difference of 2020 - 2019)
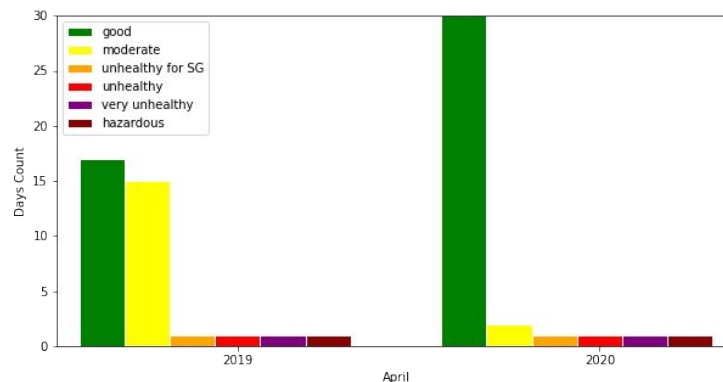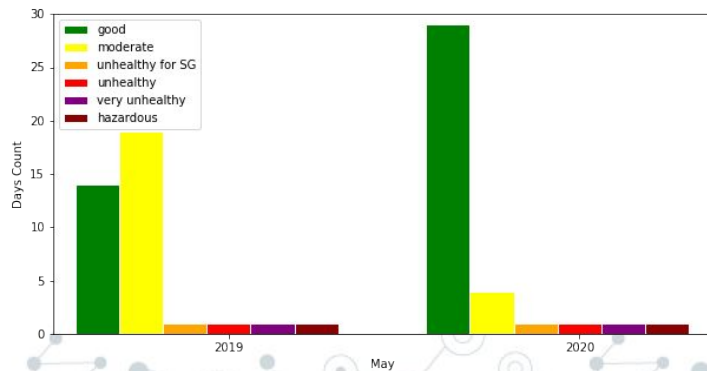
Bar charts

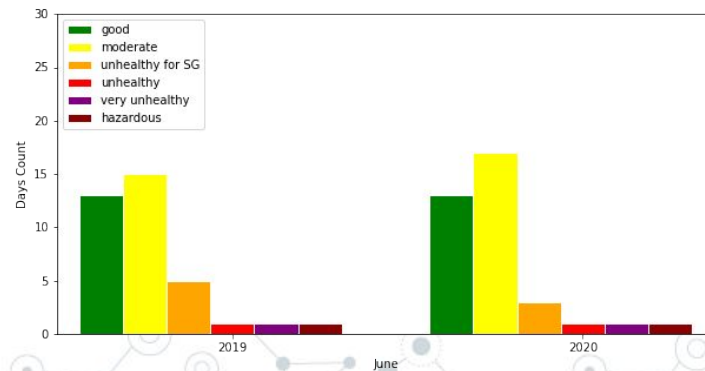New York, USA Day Counts of Level of Health Concern in March 2019 vs. 2020

New York, USA Day Counts of Level of Health Concern in April 2019 vs. 2020

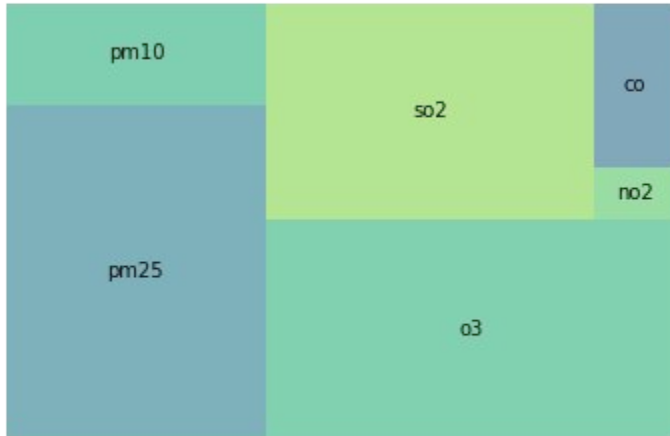New York, USA Day Counts of Level of Health Concern in May 2019 vs. 2020

New York, USA Day Counts of Level of Health Concern in June 2019 vs. 2020
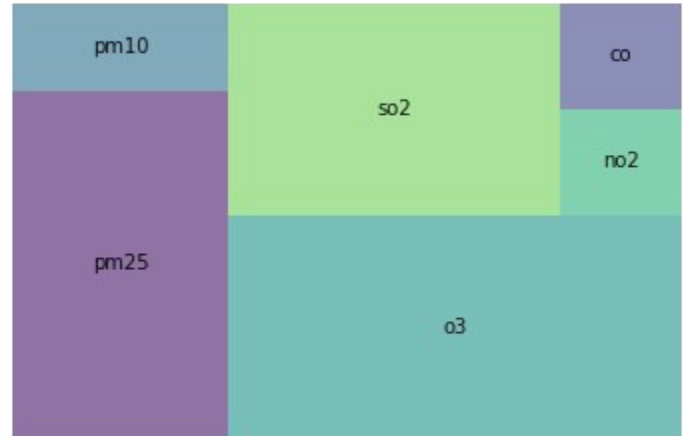
11

Tree map

New York first time, Average Pollutant Proportion in 2019

New York first time, Average Pollutant Proportion in 2020

AQI change in a glance

City	AQI change(%)
Wuhan	22.79
Beijing	7.62
Seoul	10.28
Hong Kong	18.92
New York City	17.15
Miami	-22.52

FINDINGS

1. We can see significant changes in the AQI in many major cities around the world.
2. Policy has a strong impact on the AQI. Cities with strict lockdown measurements had seen better air quality over the lockdown period.
3. Anomalies in our observation can be explained by some special events. For example, wildfires in California and population influx in Florida.
4. Not all pollutants are created equal. In general, the amount of all pollutants decreased. But, when we look into individual pollutant, we find that CO decrease more and Ozone decrease less proportionally.

CHALLENGES AND LIMITATIONS

1. The amount of high-quality data is limited online.
2. WAQI data contains 380 cities. The dataset is too big for OpenRefine. Loading data and processing it took a very long time.
3. Some cities do not have clear policy on lockdowns. It is hard to find a clear cutoff date for lockdown or reopen. Also, the lockdown measurements are different across countries. Some are strictly enforced, some are voluntary stay-at-home.
4. Oxford policy dataset uses a special numerical scheme to encode the policy content. It needs a codebook to understand the data.

Thank You!