India Air Quality Index - Data Analysis

Rishi Dey Chowdhury

7/9/2021

INTRODUCTION

In this Data Analysis Project, I am going to work with **Air Quality Index (AQI) Data of India**. I will be using several Statistical Tools to Analyze the Data which includes Exploratory Data Analysis, Techniques and methodologies used for Inference and Modelling.

MOTIVE

- To understand the relationship that holds between different parameters which we measure as a part of AQI Data of India.
- To find out if COVID-19 Nation-wise Lockdowns, Social Distancing and Closure of Industries, Factories and Suspension of movement via private vehicles and public transport had a significant impact on India's AQI.

UNDERSTANDING THE DATA

Let's take a look at the Data

- This Dataset contains 263890 rows and 10 columns.
- The Year ranges from 2014 to 2021 (till June), with observations recorded on each of the 30 /31 days of the month for 12 months for the last 3 years.
- The Data is generated from the 22 cities from various Stations located near that Cities. The Cities include:

- The parameters which we measure at the different Stations are given under the Specie Column and it includes -

Table 1: First few rows of the Air Quality Index Data

Year	Month	Day	City	Specie	count	min	max	median	variance
2014	12	29	Delhi	pm25	24	296.0	460.0	394.0	27226.40
2014	12	29	Hyderabad	pm25	13	159.0	162.0	161.0	8.59
2014	12	29	Delhi	pm10	82	79.0	999.0	218.0	634717.00
2014	12	29	Delhi	o3	79	0.1	87.4	3.2	2324.38
2014	12	29	Delhi	so2	91	0.3	21.2	4.2	231.83
2014	12	29	Delhi	pm25	83	139.0	747.0	307.0	215149.00

Table 2: Last few rows of the Air Quality Index Data

Year	Month	Day	City	Specie	count	min	max	median	variance
2021	6	24	Kolkata	o3	48	2.9	105.7	8.4	4611.99
2021	6	24	Kolkata	pm25	48	45.0	104.0	63.0	1398.61
2021	6	24	Kolkata	pressure	56	996.9	1007.5	999.3	67.94
2021	6	24	Kolkata	wind-speed	56	0.1	4.2	1.1	10.87
2021	6	24	Kolkata	dew	37	28.0	28.0	28.0	0.00
2021	6	24	Kolkata	co	48	1.0	5.2	2.3	16.41

Table 3: City Stations

State	City	Number of Stations
Andhra_Pradesh	Visakhapatnam	1
Arunachal_Pradesh	Visakhapatnam	1
Bihar	Patna	6
Chandigarh	Chandigarh	1
Delhi	Delhi	40
Kerala	Thiruvananthapuram	2
Kerala	Thrissur	1
MadhyaPradesh	Bhopal	1
Maharashtra	Mumbai	21
Maharashtra	Nagpur	1
Maharashtra	Nashik	1
Meghalaya	Shillong	1
Rajasthan	Jaipur	3
Tamil_Nadu	Chennai	8
Telangana	Hyderabad	6
Uttar_Pradesh	Lucknow	6
Uttar_Pradesh	Muzaffarnagar	1
West_Bengal	Kolkata	7

Table 4: Specie Description

Parameters	Description	Units
pm25	Particle pollution/particulate matter(particles less than or equal to 2.5 micrometers in diameter)	microgr
pm10	Particle pollution/particulate matter(particles less than or equal to 10 micrometers in diameter)	microgr
o3	Ground-level ozone	microgr
so2	Sulphur dioxide	microgr
no2	Nitrogen dioxide	microgr
co	Carbon Monoxide	miligra
temperature	Temperature	Celcius
pressure	Air Pressure	Torr
wind-gust	Wind Gust/Force	kmph
humidity	Relative Humidity	No Uni
wind-speed	Wind Speed	kmph
dew	Dew Point	Celcius
precipitation	Precipitation	milimet

Table 5: Significance of the AQI Values

AQI Values	Level of Health Concern
0-50	Good
51-100	Moderate
101-150	Unhealthy for sensitive group
151-200	Unhealthy
201-300	Very Unhealthy
301-500	Hazardous

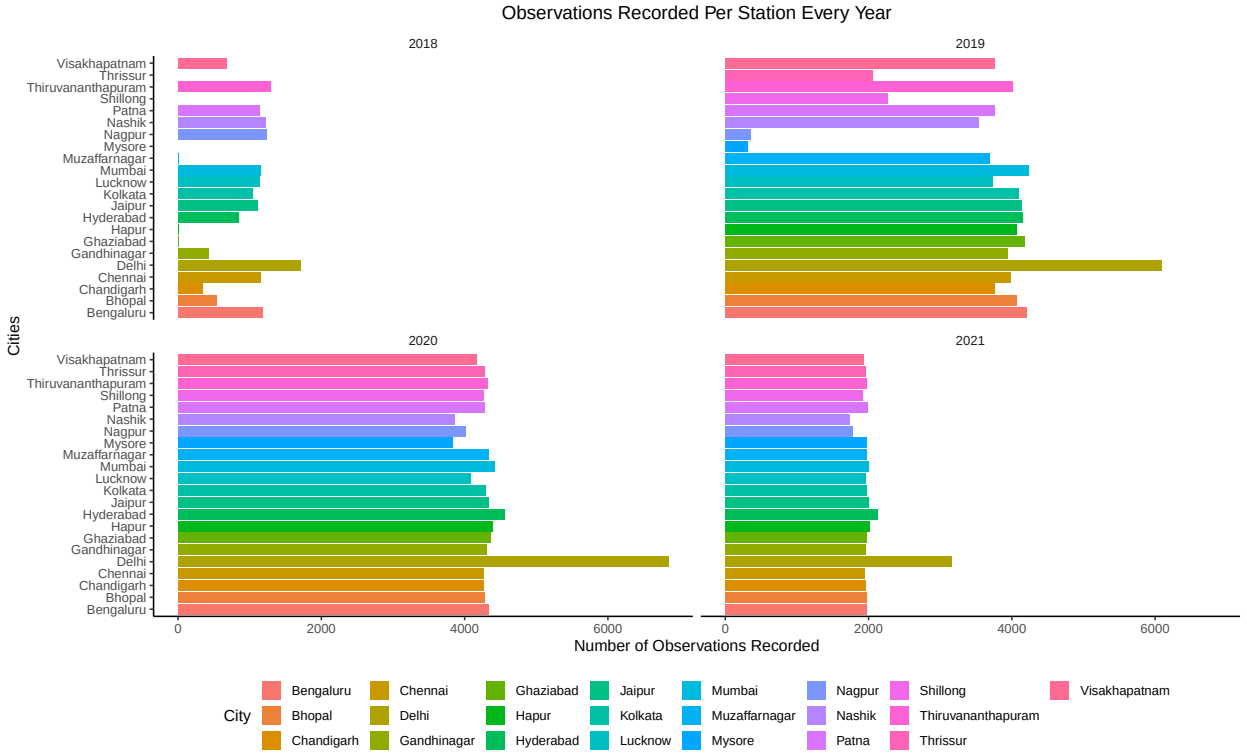
AQI is a comparable and communicable way of measuring the parameters in the Air. It is calculated when atleast 3 of the top 6 parameter's data is available of which one must be pm10 or pm25. It is the max of the parameters recorded given they satisfy the above condition.

- It also helps in identifying faulty standards and inadequate monitoring programmes.
- AQI helps in analysing the change in air quality (improvement or degradation).
- Comparing air quality conditions at different locations/cities.
- It can be easily interpreted by anyone, without knowing about background details.

In further Analysis we will refer *pm25, pm10, o3, so2, no2 and co2 as pollutants* and the *remaining weather parameters as non-pollutants.*

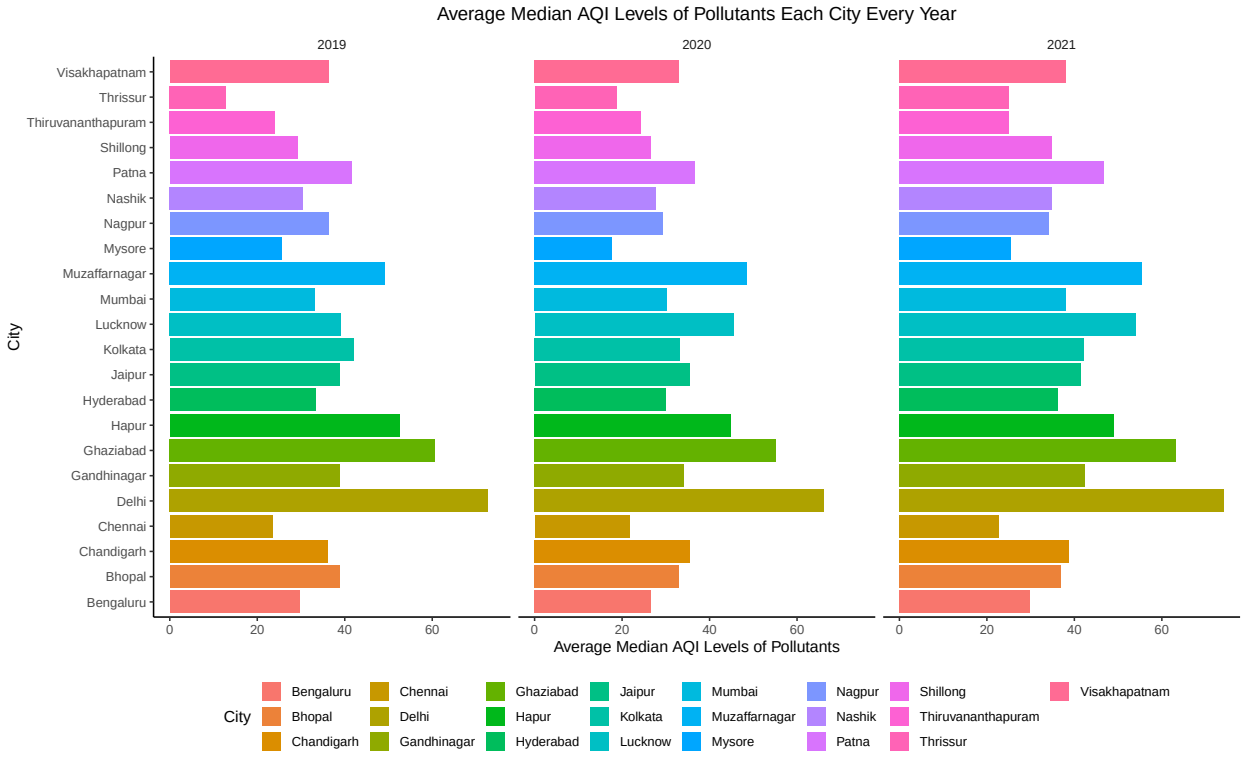
TOP CITY STATIONS

Here we look at *which city records how many observations per year.*



It seems **Delhi is the most monitored cities among the others**. We see all the other stations have almost equal number of observations per year. Our data seems to have lot of missing values for the year 2018 and the years before that; Hence we will only consider the data of the year 2019, 2020 and 2021.

Speculating on the reason *why Delhi is so heavily monitored* we look at *how the Average Median AQI Levels of pollutants at each city every year*.

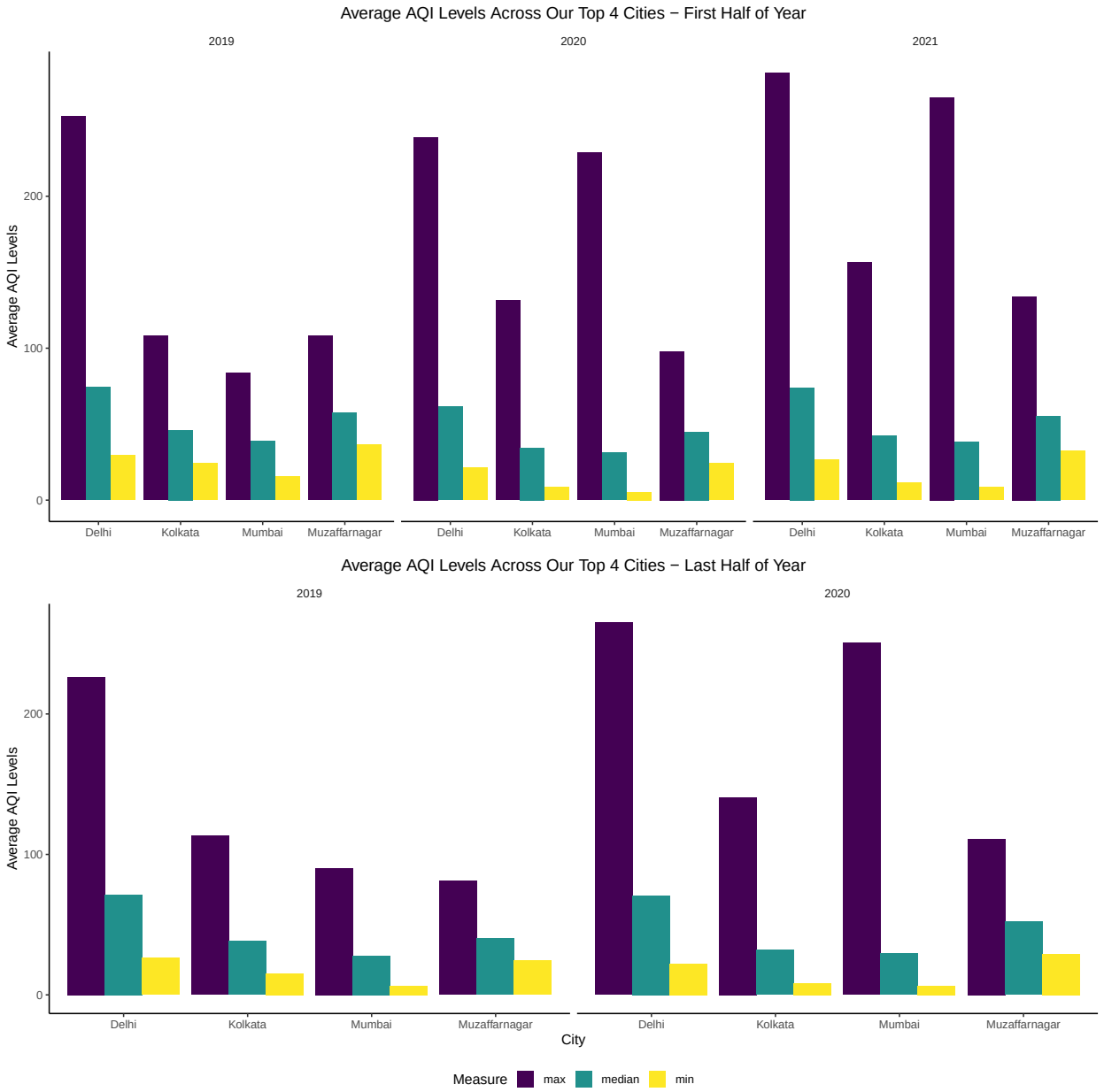


Indeed it is clearly visible that **Delhi's AQI Levels of Pollutants is higher** than all other cities. This makes Delhi our city of main focus in all the further analysis. Every city owing to it's location at different parts of the country, different development status, different population, different local weather conditions has different measured values of AQI. So, it makes sense to look at them separately. So, whenever we will look at City-wise Analysis, we will look at these Cities as our **Top 4 Cities - Delhi, Muzaffarnagar, Kolkata, Mumbai.**

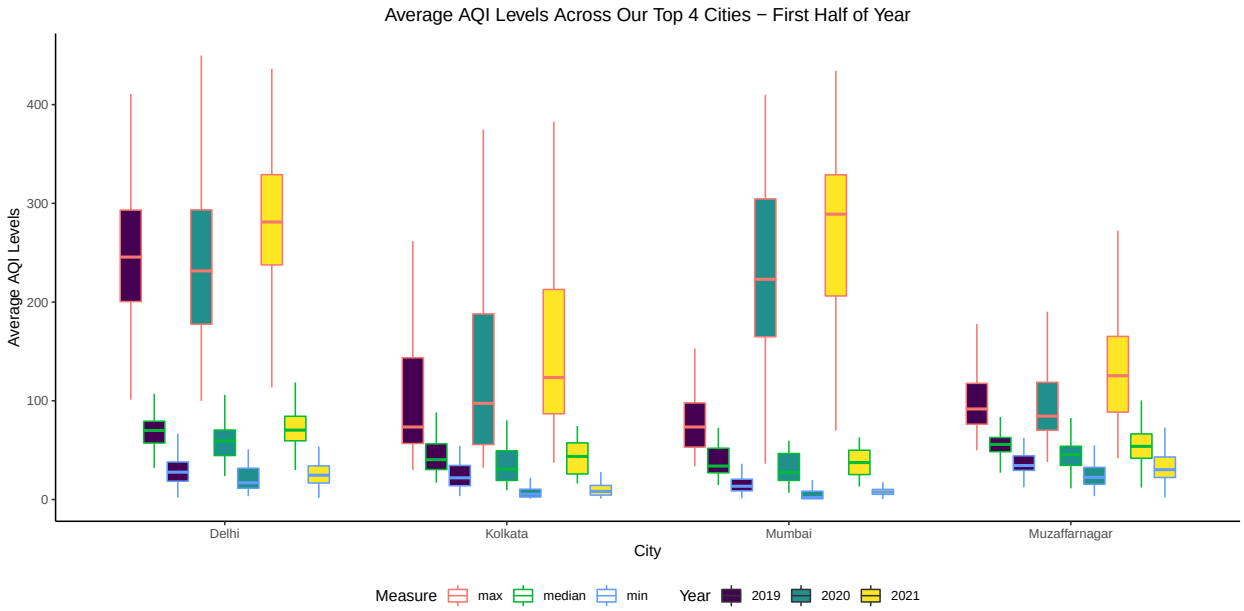
VISUAL OVERVIEW

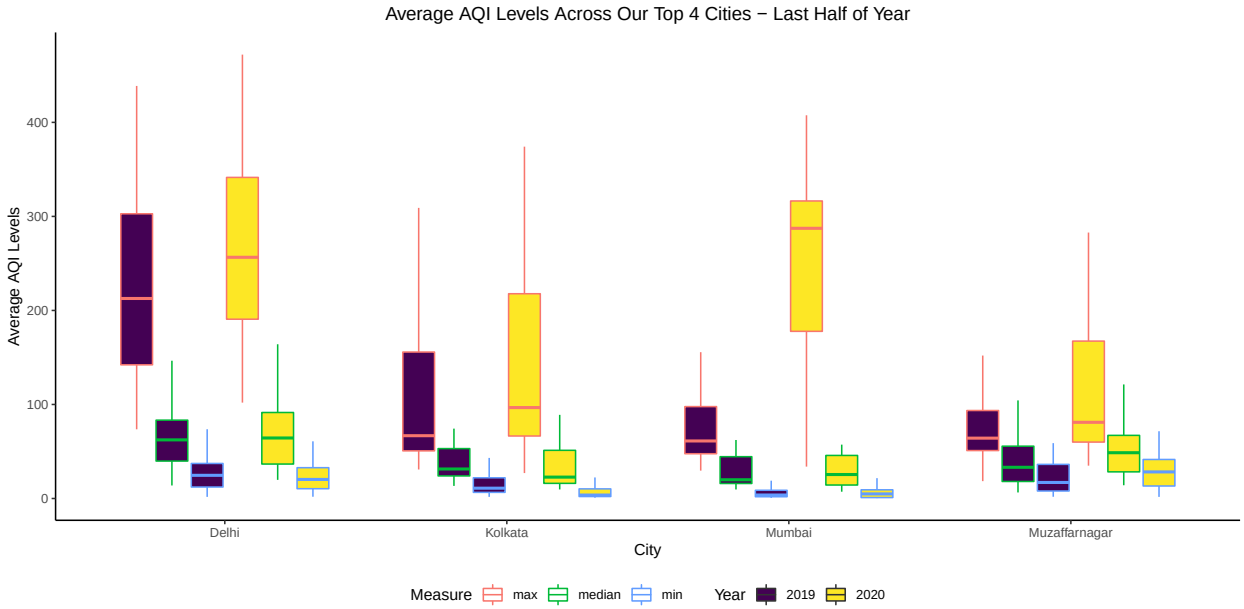
Here we will take a visual tour of the Data through Exploratory Data Analysis.

- Let's see the AQI Levels in the Top 4 across all 3 measures - Max, Median and Min splitting the year into two halves.



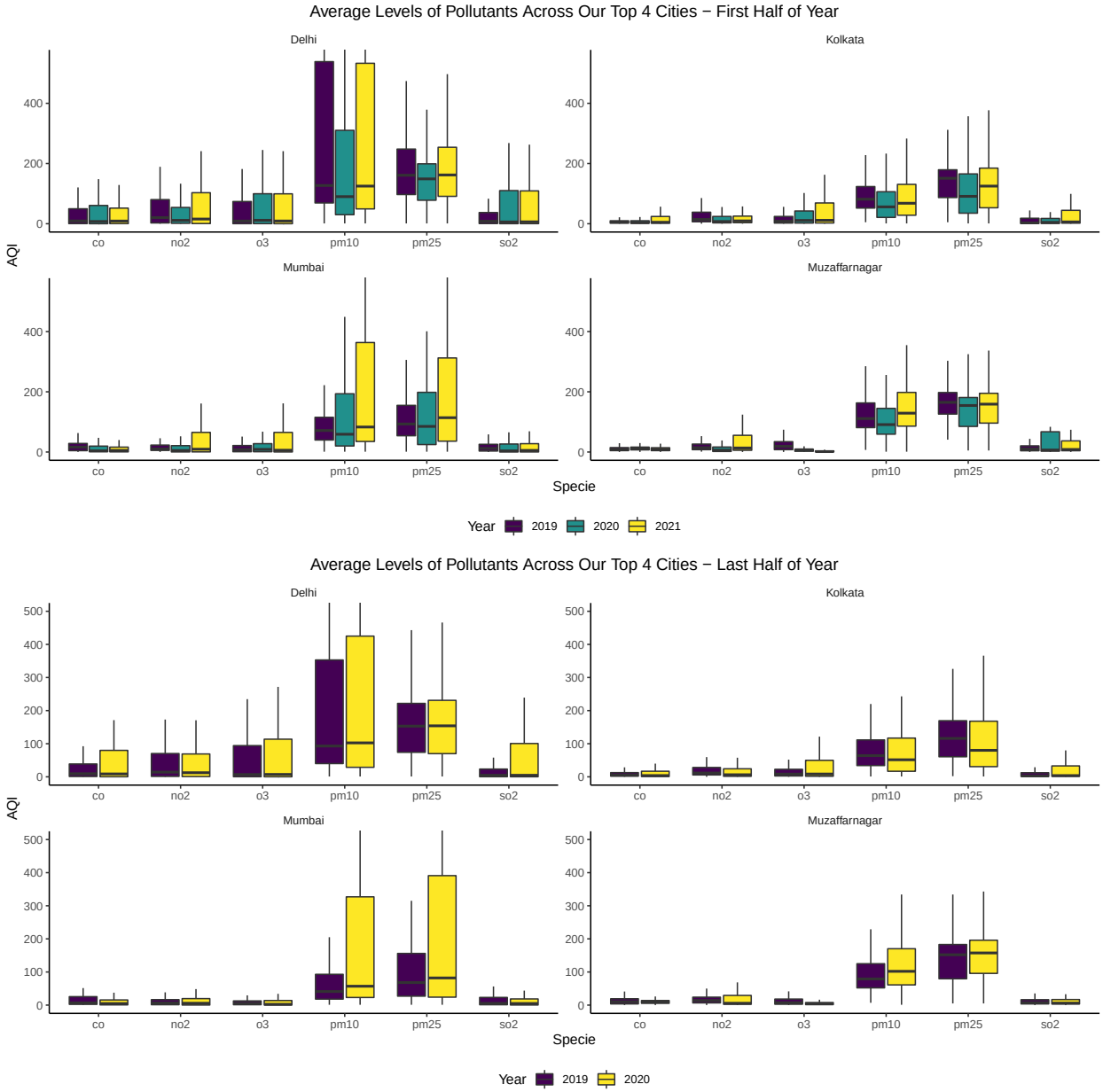
The first bit of observation is the **Median & Min AQI Levels of all 4 cities have dropped in 2020 and 2021 is similar to 2019**. But the **Max AQI Levels have increased in 2020 and even more in 2021** even though we had so many restrictions in 2020!





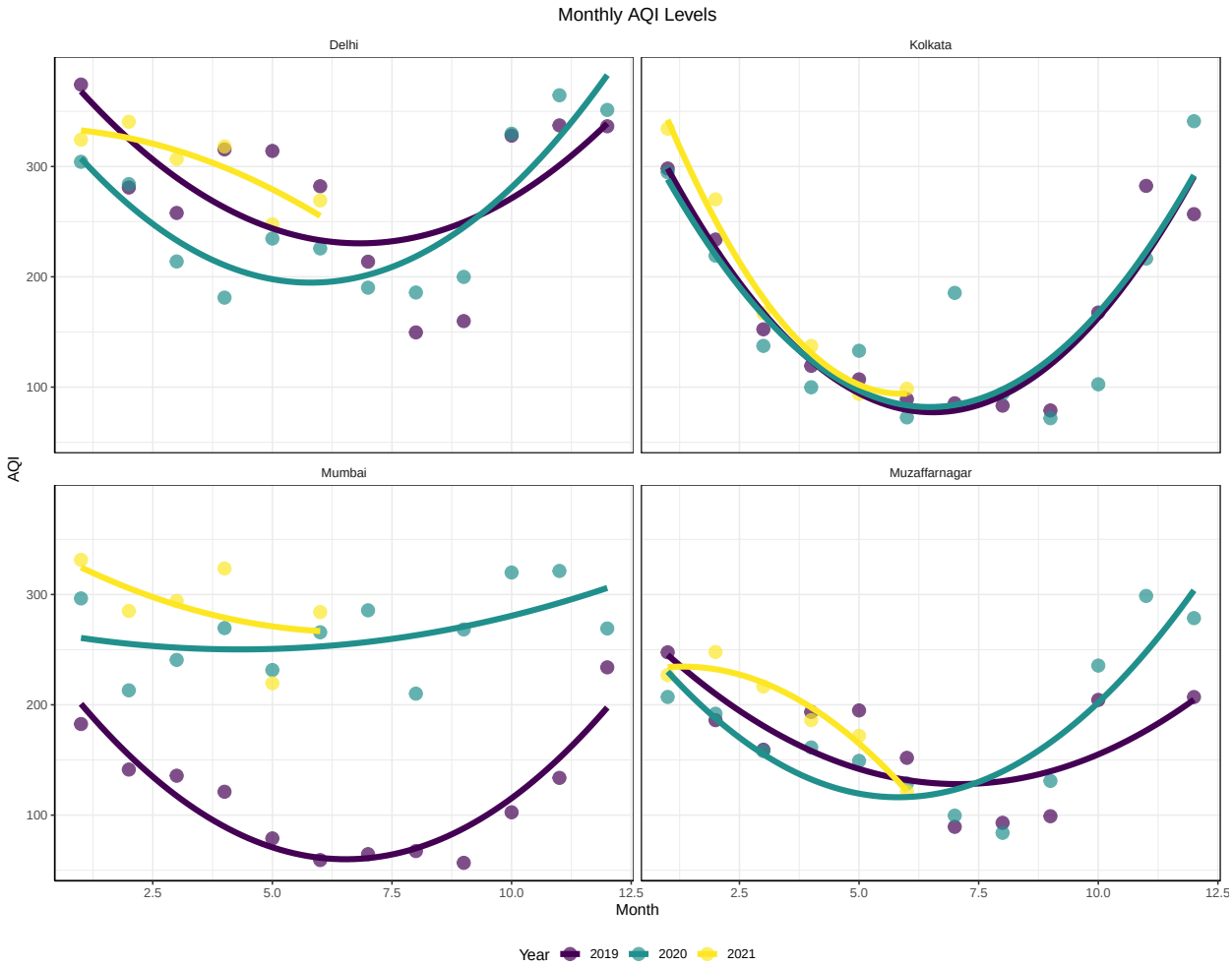
In the first half of the year 2020 Min and Median AQI Levels have dropped whereas in the second half of the year it either increased or remained same as that of 2019. 2021's observed values are pretty much the same as that of 2019 if not more. Max AQI Levels kept rising across the years irrespective of the part of the years. **The spread(Variance) of the resp. measures look the same across the years**.

- Let's take a look at the pollutant-wise Levels in our Top 4 Cities.



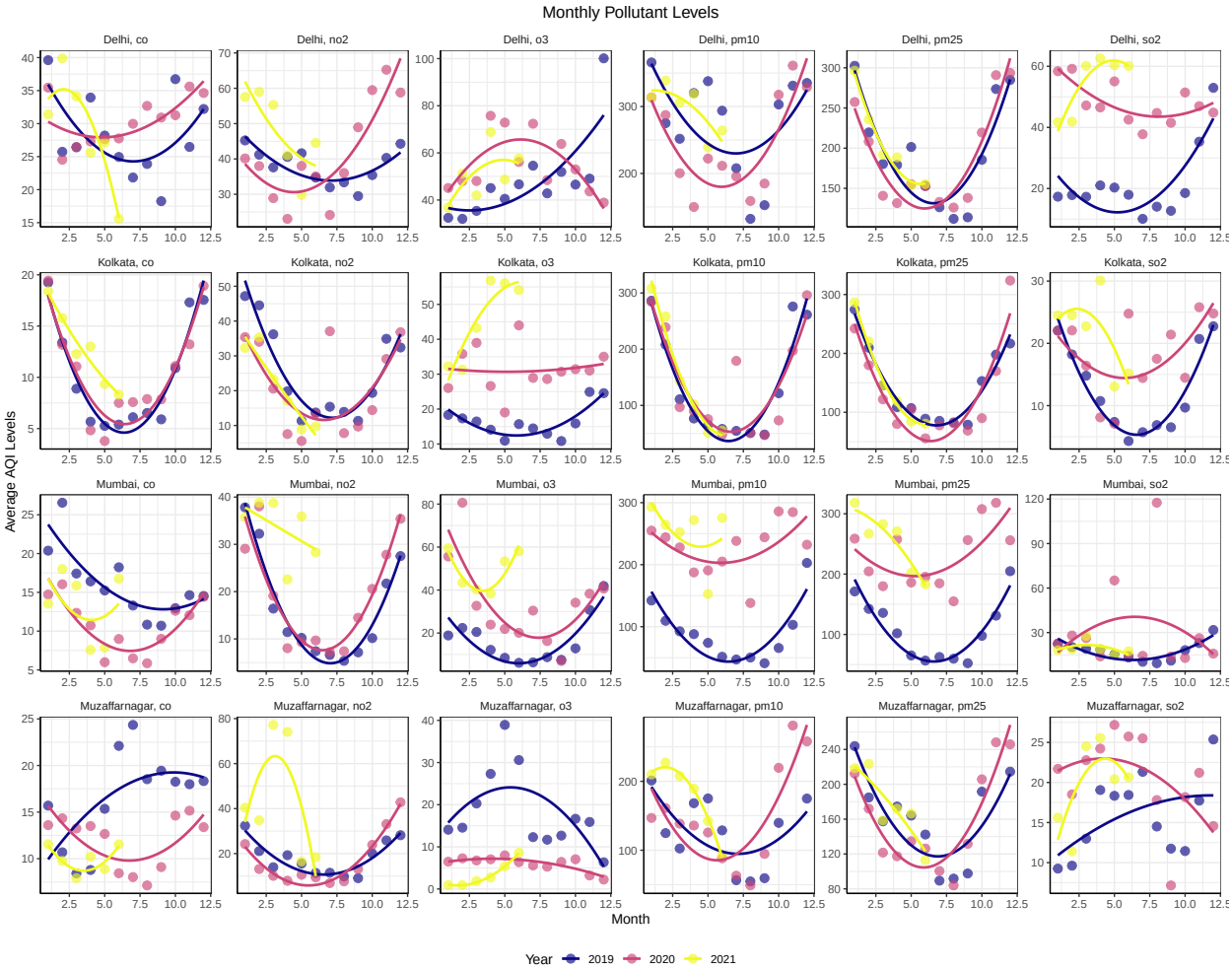
Here we see an interesting effect of Lockdowns. **All the pollutants have decreased considerably in the First half of 2020 whereas in the Second half it has bumped up again to match the Levels of 2019 or even more! While the Levels of 2021 & 2019 are very similar**.

- Pollutant Gases and Particulate Matter, whose levels despite being affected by Human Intervention and Industrial Activities, are nothing but Natural Gases which follow nature's process. Here we will take a look at *how the AQI Levels change with seasons throughout the year.*



Indeed it looks like the **AQI Levels have dropped in the first half of 2020 compared to 2019, with the exception of Mumbai. But it has increased to levels more than 2019 in the later half. The AQI Levels of 2021 is more that both the previous years**. I have fit here a Quadratic Model which seems to model the situation well though we shouldn't focus on the model fit to 2021 as we only have the data pertaining to first 6 months.

Now, we will see the above observed AQI Levels in terms of Levels of each Pollutants.



Except for Ground-Level Ozone and Sulphur Dioxide most of the Pollutant Levels have decreased during the first half of 2020 which saw an increase in the second half when compared to 2019. The surprising thing is that even **particulate matter increased to much higher levels in Mumbai** even though it saw a massive lockdown and transportation as well as industrial work were suspended.

- The Pollutant Levels are affected to a great deal by the weather conditions of a place. Here we try to look at *how the Pollutant Levels are related to the weather parameters.*