# Data Analyses

How does location correlate to pollution?

Our bar graphs show the Air Quality Index (AQI) for the most populated cities in Europe, Asia, Oceania, and the Americas. One would assume that the AQI would be consistent across the cities in each country, however what was found is that while some continents are consistent (such as North America), others such as Oceania and Asia have extreme peaks in the Gold Coast and Ahmedabad, respectively. Asia has the highest, and therefor poorest, AQI which could be easily passed off as a higher population density compared to ‘Western’ cites but would be interesting to research further to find other contributing factors.

How do the extremities of pollution differ based on location?

Our pie charts show us the most common pollutants of each region.

North and South Americas both have high amounts of Fine particle Matter (PM2.5), which may be linked to respiratory health and illnesses, however South America’s average Nitrogen Dioxide (NO2) is higher. A common cause for NO2 is burning of fuel, and further research into this may provide some direct origins.

NO2 is also the largest contributing factor in Oceania, and as it is located in the southern hemisphere with South America, perhaps there is a link between the two that would be worth exploring in the future.

Asia joins North America with the most common pollutant being Fine Particle Matter.

How will location affect the characteristics of pollutions?

We find that there is a minor correlation between regions and their air pollutions, but there is still so much research we could do to find more concrete links between them. As it stands, we find that Fine Particle Matter is the largest contributor to pollution across most regions, followed by Coarse Particle Matter and Nitrogen Dioxide. However, the minor correlations between specific pollutants, and AQI’s of regions cannot confirm our hypothesis.

Challenges and Limitations

One of the challenges in collecting the live data is the incomplete data sets. Not all cities had information on all pollutant, which resulted in too many zeros, and the entire row scrapped for the purpose of finding the Mean. This then created a smaller sample size for the region.

Another challenge was the time to process each filter due to the amount of data that was collected.

The biggest limitation we had was not collecting historical data. That was due to the cost involved in purchasing said data. Ideally, we would be able to collect this historic data and how that compares to the live data, and if there are any significant changes that events such as the COVID pandemic had between the data sets.