Context of the problem.

Air pollution is a concern that has been addressed by people in the last few years, yet the concepts for air quality, the measurement techniques and the consequences of breathing the different molecules on the air has been less addressed than I would have liked personally. Most of the people simply want to know if it is “dangerous” to go outside, but sadly the situation can be a little more complex than that.

Investigating a little about the Air Quality Index (AQI) which is the measure used to know the levels of toxicity in the air, is not the easiest to calculate because the amount of data needed, that’s why most analysis seems to be using a “simplified” version of the concept which actually leaves out multiple molecules that are of concern. This situation is specially important in Mexico because of all the cities, it looks that not everyone of them is equipped enough to calculate a real AQI; Even for Guadalajara, which has the most information that I could find, there is not enough data to actually calculate the real AQI.

Taking all this into account I decided to create a prototype to analyze all the information at our disposal and gives use simple recommendations depending on the molecule and the harm it could have by inhaling it in high concentrations. The idea is that you don’t need to be an expert to be able to know how to take care of yourself.

Solutions

I decided to make a program which could analyze the data of multiple years and molecules at the same time using concurrency, this with the objective of having the process done fast, as it would be parsing multiple csv’s of more than 8 thousand lines each one. After the data has been analyzed, in the case of the prototype a simple average of each molecule through the year on the multiple zones that the city is divided by the INECC, a CSV is written with the results. The last CSV is imported by a prolog program which evaluates by zone and year, and finally returns a “Tip” for dealing with every molecule.

Results

With the available data I could see that the pollution situation, at least in Guadalajara and at least in the year it could be tested, was not as severe as I had thought, at the very worst we could find ourselves on a moderate level of concentration of dangerous molecules. It is important to note that this is by using complete years, If we take a look at it by seasons we can see that every season is different and ones can have more concentration of certain molecule.

Setup Instructions: