

# Hot days, high concentration NO2

Mariken van Strien, 1854399

## Introduction

**Hypothesis: The proportion of days with high a high concentration of No2 is higher when it is hot.**

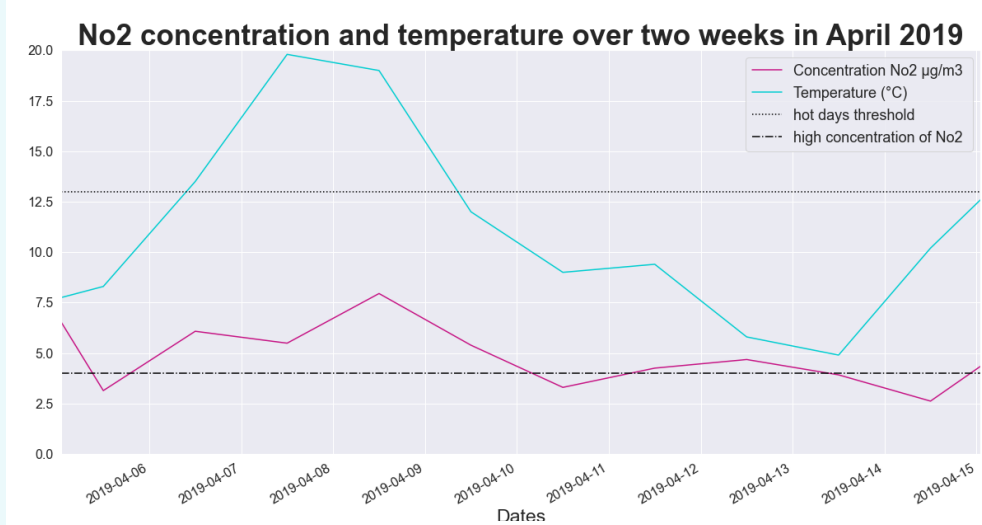
NO2 is released from a number of places, including industrial activities, power plants, and automobile exhaust. The amount of NO2 in metropolitan areas can change based on the weather, traffic patterns, and other elements (Tzortziou et al., 2022).

According to one theory, hot weather increases the proportion of days with high NO2 concentrations. This theory is predicated on the idea that rising temperatures cause changes in air conditions that can trap pollutants close to the ground as well as an increase in emissions from machinery and other sources. If this theory is correct, it may have significant repercussions for public health (Grigorieva, E., & Lukyanets, A., 2021). Because of this I researched if there is a relation between the concentration of NO2 and the temperature. This resulted in the hypothesis stated above.

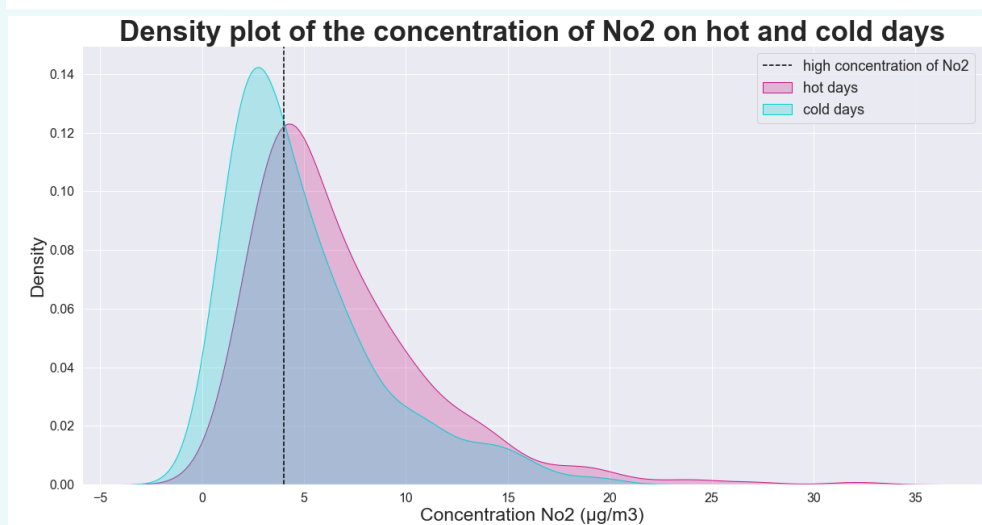
## Methods

To determine if this hypothesis is true, first the data that was given was subtracted and a dataframe was created to make the data usable. This data was then 'cleared' from missing data using interpolation. After this a line graph was created to visualize the concentration of NO2 vs temperature and to see if there could be a possible correlation between these two. Next a kernel density plot of the temperature in the data set was created and a threshold value was determined. This threshold helped dividing the data set in to two sets. Using these sets another density plot was created, visualizing the distribution of the NO2 concentration in both these data sets, to be able to compare them. And look at the concentration on hot days versus cold days. After this the p-value was determined and a z-test was done on the data.

## Results



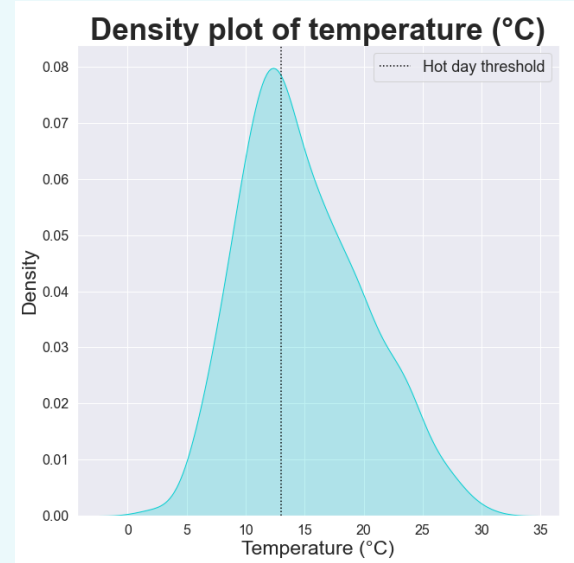
The graph above visualizes the NO2 concentration, and the temperature over two weeks in April 2019, the two horizontal lines show the dividing of "hot" and "cold" days and "high" and "low" concentration. We can see that when the temperature is considered high (above the line), the proportion of days above the concentration line is higher than below the line.



## Discussion

The initial steps of creating a dataframe and clearing the data from missing values using interpolation were important to ensure that the data was usable for further analysis. The line graph showing the concentration of NO2 vs temperature helped identify a potential correlation between these variables. The density plots that were created helped with defining how the data should be divided and this dividing helped with analyzing if the hypothesis was correct: if the temperature actually influences the proportion of days with a high concentration NO2. The accuracy of the analysis was also determined by the p-value and the z-test. Overall, the analysis seems to have followed a sound methodology to test the hypothesis and provide evidence for a relationship between temperature and NO2 concentration. However it is important to realise that the data used was measured at a specific time and location, which can influence accuracy. Next to that it is also good to keep in mind that other factors, weather factors or other, could also be of influence.

The graph on the right visualizes the distribution of temperatures in the data set. With this plot the threshold value was determined. The vertical line visualizes this value. The value has been chosen as it is close to the peak, which means it is the most common value of the data. And it is a good value between a low and high temperature.



The graph on the left visualizes the distribution of NO2 concentration in the data set one of the graphs shows the data on hot days and one of them shows the data on cold days. The vertical line shows the chosen value that divides the sets in a high and a low concentration NO2. If you look at this graph you can see that on hot days (the pink graph) most of the concentration values are considered high, the majority of the concentration values are on the right side of the line. And on cold days the majority of the concentration values are considered low.

The confidence interval value has been computed, this is: 0.5395, 0.6179, this means that there is a 95% chance that a bigger amount of days with high NO2 concentrations is related to the weather being hot. Next to this the p value also has been computed: 1.9321466302096588e-08) and this indicates strong evidence against the null hypothesis.

## Conclusion

Based on the analysis, it can be concluded that there is evidence to support the hypothesis that there is a higher proportion of days with high NO2 concentration when it is hot. The analysis followed a clear methodology to provide evidence for this relationship, this includes clearing of unnecessary/unclear data, creating visualizations and doing statistical tests.

Tzortziou, M., Kwong, C. F., Goldberg, D. E., Schiferl, L. D., Commane, R., Abuhassan, N., Szykman, J., & Valin, L. (2022, February 22). Declines and peaks in NO<sub>2</sub> pollution during the multiple waves of the COVID-19 pandemic in the New York metropolitan area. Atmospheric Chemistry and Physics; Copernicus Publications. <https://doi.org/10.5194/acp-22-2399-2022>  
Grigorieva, E., & Lukyanets, A. (2021, June 19). Combined Effect of Hot Weather and Outdoor Air Pollution on Respiratory Health: Literature Review. Atmosphere; MDPI. <https://doi.org/10.3390/atmos12060790>