First I created a copy of the provided CSV files opened them as an Excel file and deleted the first 3 rows and the empty row that didn't contain any relevant information, formatted the time column to be the same for every row as hh:mm:ss and then downloaded as a CSV file.
Save them in the folder Date.

First start with data manipulation, as I extract the df for every year, and choose the pollutants that I need to plot later, date and time . as I keep date and time separate.
Then use rename() to change the name of the pollutants to make them easily readable.
Then modify the data frame and select only particular data with select() that I need to work with - date, time PM10, NO, NO2.
Then create the data frame named under the required date with filter()
I checked every data frame for missing data filled the gaps where possible, and kept the gaps where there were too many.
I chose not to fill in all of the missing data, but to visualize with them, because the missing information is at the beginning of the day and is 12 rows, which would give irrelevant information, I chose to fill in a single missing data with approx.
Convert time from chr to hh:mm:ss and date from chr to Date type.
Then start to plot the data to see if everything goes through.
I choose to plot the data with lines and in some plots with points because is a better view of timeline data, and clearly can see the spikes if there are any.
In r script extract_data I load all libraries that will need to work out my extraction, and average the months for 2020 to produce plots.
I use R markdown to create an HTML to show my visualisation and write my observations for the final plots as a final report.

References

Hadley Wickham, Romain François, Lionel Henry, Kirill Müller, Davis Vaughan. (n.d.). *A Grammar of Data Manipulation*. Retrieved 26 March 2025, from https://dplyr.tidyverse.org/

Tidyverse. (2019.). Tidyverse packages. https://www.tidyverse.org/packages/

Ehsan Karim. (2020, October 14). *Lab 6 (part 1a) Missingness Types and Patterns*. YouTube. https://www.youtube.com/watch?v=KnBkYkHfGG0

Statistics Ninja. (2021, January 4). *Exploring missing values with the mice package*. YouTube. https://www.youtube.com/watch?v=fsqAi9ddHU0

Adam Petrie. (2018, August 31). *Unit2 - Analyzing and Handling Missing Values in R*. YouTube. https://www.youtube.com/watch?v=ptynHDzVLxo

Handling missing values in R (2021, April 24). R-Bloggers. https://www.r-bloggers.com/2021/04/handling-missing-values-in-r/#google_vignette

na.approx function - RDocumentation. (2023). Rdocumentation.org. https://www.rdocumentation.org/packages/zoo/versions/1.8-12/topics/na.approx

Tidyverse. (n.d.). HMS: Harmonized multiple-imputation software. https://hms.tidyverse.org/

Yan, Q. (2019, November 28). R Visualization Tips: Using the scales package. R Visualization Tips. https://bookdown.org/Maxine/ggplot2-maps/posts/2019-11-27-using-scales-package-to-modify-ggplot2-scale/

u/jwilliman. (2022, October 25). How to calculate monthly averages from daily data? [Online forum post]. Reddit. https://www.reddit.com/r/RStudio/comments/yd6bxp/how_to_calculate_monthly_averages_from_daily_data/?rdt=36308

Stack Overflow User. (2021, August 24). Finding monthly average of columns using group by function in R [Online forum post]. Stack Overflow. https://stackoverflow.com/questions/68910425/finding-monthly-average-of-columns-using-group-by-function-in-r

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). *The New S language*. Wadsworth & Brooks/Cole. https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/mean

mean function - RDocumentation. (n.d.). Www.rdocumentation.org. https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/mean

Plotly Technologies Inc. (n.d.). *Line and scatter plots in R* [Documentation]. Plotly. https://plotly.com/r/line-and-scatter/

Wickham, H., François, R., Henry, L., & Müller, K. (n.d.). *group_by: Group by one or more variables* [Function documentation]. Tidyverse. https://dplyr.tidyverse.org/reference/group_by.html

Wickham, H., François, R., Henry, L., & Müller, K. (n.d.). summarise(): Summarise multiple values into a single value. *dplyr*. Retrieved from https://dplyr.tidyverse.org/reference/summarise.html

Codecademy. (n.d.). R Aggregates Cheat Sheet. *Codecademy*. https://www.codecademy.com/learn/learn-r/modules/learn-r-aggregates/cheatsheet

Sanderson, S. P., II. . (2024, January 23). How to Extract Month from Date in R (with Examples). *R-bloggers*. https://www.r-bloggers.com/2024/01/how-to-extract-month-from-date-in-r-with-examples/

Khati, O. (2021, May 20). How to get month name as output from date object. *Stack Overflow*. https://stackoverflow.com/questions/67624454/how-to-get-month-name-as-output-from-date-object

Plotly Technologies Inc. (n.d.). ggplotly: Getting Started. *Plotly*. https://plotly.com/ggplot2/getting-started/

Plotly Technologies Inc. (n.d.). Facet Wrap. *Plotly*. https://plotly.com/ggplot2/Facetting/facet_wrap/

Plotly Technologies Inc. (n.d.). Facet Wrap. *Plotly*.

https://plotly.com/ggplot2/Facetting/facet_wrap/