

Exercise 1. Data manipulation

Sasha D. Hafner

27 February, 2024

Overview

In this exercise you will carry out some basic data processing steps. The measurement data are from an experiment on the performance of two air cleaners on removal of particular volatile organic compounds (VOC). Measurement data were kindly provided by Kasper Kristensen.

1. Reading in data

Read in the data in the file `air_cleaners.csv`. Take a look at the top of the resulting data frame. Interpretation of most columns is pretty straightforward. Concentration (`concentration`) is in parts per billion on a gas volume basis (ppbv).

2. Checking data

Generate a summary of the data. What is the range in concentrations? Are there any missing values? How many air cleaners (`aircleaner`) are there?

3. Subsetting

Take a subset of your data frame that has results for only one air cleaner. Print a summary to your console to check it. Check the change in number of rows. Can you limit this subset to a single compound and flow direction (`flow_dir`)?

4. Merging

The file `mol_mass.csv` has molar masses of these compounds. Read in those data and merge them with the air cleaner data. Check for dropped rows.

5. Adding columns

Calculate concentration in g m^{-3} . You can use `mol_mass * pressure / gas_constant / temperature`. Numeric values do not matter so much here but you can use `gas_constant = 8.2057E-5` for the universal gas constant in $\text{atm m}^3\text{k mol}^{-1} \text{K}^{-1}$. Add mass flow rate of each compound based on air flow rates of $130 \text{ m}^3 \text{hr}^{-1}$ for PureAir2000 and $100 \text{ m}^3 \text{hr}^{-1}$ for NoSmell4.2. Think about whether you need a grouped operation for this task.

6. Dates and times

Check that `timestamp` has been correctly interpreted as a date/time object, and convert it if necessary. Add a new column with elapsed time since 10 March 2022 13:40 in minutes.

7. Grouped operations

Calculate the mean and standard deviation of concentration by air cleaner model, flow direction, and compound. Save the results in a new data frame.

8. Export

Export the summary you made in the previous step as a csv file.

9. Integration

Integrate mass flow data by air cleaner model, flow direction, and compound to calculate total mass flow.

10. Reshaping

Reshape the data frame so you can calculate removal in g by model and compound.