COMP 333 — Lab Assignment 3

Motivation

The purpose of this assignment is for you to apply Python and pandas for data wrangling.

It builds on Lab 8 where you used OpenRefine to wrangle ecological data. Here you want to think about how to program those steps of data wrangling using Python and pandas.

You should test out your code using the ecological data as done in Lab 8.

Assignment

Create a Jupyter notebook using Python code and any of its libraries, but especially pandas, to wrangle the ecological data of Lab 8.

Think about each step of the wrangling, and if there is not a direct function in pandas (or a Python library) for the step, then structure your code by defining a Python function for the step.

You should write functions that perform the following steps:

- construct a text facet when given a column with text strings as entries;
- construct a numeric cluster when given a column with numeric entries, with the option of using a log scale;
- cluster a column of text entries (using an appropriate clustering method);
- for string processing by trimming all whitespace at the start and end of a string; and
- for string processing by replacing internal whitespace by a single space.

You may need to write other functions as well to structure your code.

Test your code by applying it to wrangle the ecological data.

Deliverable

Your deliverable is the completed ipynb notebook showing all code, computation, output, and plots.