EN.540.635 "Software Carpentry"

Lab 7 - Data Organization

Invariably as scientists and engineers we will need to collect data. Whether the data is collected ourselves, via instrumentation, or by collaborators, it is rare to have said data perfectly organized in an easy to utilize fashion. The focus of this lab is to practice basic python tools to process messy data. In today's lab you will encounter a new file type: *pickle*. Pickle is a module that allows you to convert data types into a byte stream and write them to a file. It also provides functions that do the inverse, and thus you can read your pickled data. We have provided such a pickle file which you will be working on today. Your task is to do the following:

1. Download and read the pickle file.

2. Assess the data types within the pickle file. In this case, we will tell you that you have a list of data; however, the data itself is mixed up. You will have some entries correlating to pokemon, some correlating to elements in the periodic table, and the final entries correlating to the top 1000 books.

3. Parse out the list into three individual lists for the various data types.

4. Sort the books from the top to 999.

5. Sort the Periodic Table by element number.

6. Sort the pokemon by their attack.

7. Generate and plot a histogram of the top 1000 books by author name i.e plot the frequency of author occurrences in the top 1000 against the author name. Who is most represented in this list?

8. Plot the elements weight on the the x-axis, and the corresponding van der Waals radii on the y-axis. Is there a trend?

9. Plot the Pokemon attack on the x-axis and defense on the y-axis. Is there a trend? Does this indicate an "ideal" pokemon?

In regards to the pokemon data, you will notice several numbers exist. They correlate to the following data in the corresponding order:

Total, Health Points, Attack, Defense, Special Attack, Special Defense, Speed