Assignment 11: Correlation in Coal Quality Data

45 Points scaled to 20 Points

Introduction

In this assignment you will explore the correlation between the concentration of different elements or substances in coal samples. These data are form the United States Geological Survey Coal Quality (COALQUAL) database (https://ncrdspublic.er.usgs.gov/coalqual/). The data were prepared and provided by Geology PhD student Rachel Yesenchak.

Objectives

- *Prepare data for analyses*
- *Assess correlation using correlation coefficients*

Deliverables

- *Jupyter Notebook (Python) or R Markdown file (R) with all code and graphs embedded. Files can be rendered to HTML webpages if your instructor requires this. Questions or prompts should be stated and answered within Markdown cells.*

Background Questions

Question 1. Explain the difference between a linear correlation and a monotonic correlation. (5 Points)

Question 2. Explain the difference between the Pearson and Spearman correlation coefficients? (5 Points)

Questions and Tasks

This assignment can be conducted using either Python or R, whichever you prefer or whichever you instructor requires. Generate code to perform the following tasks and answer the associated questions.

Task 1. The data are in the wrong shape for the analysis. The rows should be the columns and the columns should be the rows. Reshape the data so that the columns, which represent different coal samples, are converted to the rows, and the rows, which represent the different chemical measurements, are the columns. (5 Points)

Task 2. Extract out just the measurements associated with Ash ("Ash"), Silicon ("Si"), Aluminum ("Al"), Calcium ("Ca"), Magnesium ("Mg"), Sodium ("Na"), Potassium ("K"), Mercury ("Hg"), Phosphorous ("P"), and Lead ("Pb"). (5 Points)

Task 3. Calculate all Pearson correlation coefficients for all pairs of chemical measurements. Present the results in a table. (5 Points)

Task 4. Calculate all Spearman correlation coefficients for all pairs of chemical measurements. Present the results in a table. (5 Points)

Task 5. Create three scatterplots to compare two of the measurements. For each of the scatterplots, write a short paragraph describing the relationship between the variables. In the paragraph, also include a discussion of the Pearson and Spearman correlation coefficients for the pair. (15 Points)