D2: Dispersion and Correlation

This week's assignment (D2) is also very straightforward.

We would like you to write code that

1) Builds on D1. So the data from D1 should already be loaded. We'll reuse it here.

2) Computes the standard deviation of the ratings of each movie, across all participants (the output here should be a variable that contains 400 standard deviations)

3) Computes the mean absolute deviation of the ratings of each movie, across all participants (the output here should be a variable that contains 400 mean absolute deviations)

4) Computes the mean and median of the standard deviations resulting from 2)

5) Computes the mean and median of the mean absolute deviations resulting from 3)

6) Computes the pairwise Pearson correlation between all ratings of movies. (the output here should be 400x400 correlation matrix – it's ok to correlate ratings of a given movie with itself, for the sake of simplicity. It will also allow you to check if this worked, as the diagonal should contain 1s).

7) Computes the mean and median correlation resulting from 6)

Hint: Most of these can easily be done with nested loops. But there are many ways to do this assignment (with different tradeoffs).

Caution: The most common way to get this assignment wrong is to
   a) Being off by one (row or column). This is particularly common in Python, as Python indexes from 0. This is known as the "fencepost problem" or a "off-by-one-error". Double-check everything and read up on it, if you are unfamiliar with this issue.
   b) Being unable to handle missing values. Most entries in the dataset are empty, as most people have not seen most of these movies. That's fine. Just make sure to handle it properly (don't impute zeroes and make sure not to include the nans in the calculations)

If you feel adventurous and want to work ahead, make – separate – histograms of the resulting distributions of the results from 2), 3) and 6). We'll use them later.