D3: Simple linear regression

We would like you to write code that

1) Builds on D1 and D2. So the data from D1 should already be loaded. We'll reuse it here.

2) Finds the ratings of users who have rated both Star Wars I and Star Wars II. We will only consider these ratings (where the pair of ratings from both movies is jointly present) going forward.

3) Builds a simple regression model – predicting the ratings of Star Wars I from the ratings of Star Wars II (only). We will keep this as simple as possible (not using multiple regression or avoid overfitting – we'll consider that in a later assignment, do not forget to include the intercept term, however)

4) Returns the betas and residuals of this model.

5) Finds the ratings of users who have rated both Star Wars I and Titanic. Make sure you are not off by one (!)

6) Builds a simple regression model – predicting the ratings of Titanic from the ratings of Star Wars I (only).

7) Returns the betas and residuals of this model.

Hint: Compute the RMSE (Root Mean Squared error) as follows:
a) Calculate the deviation (difference) between two numbers, e.g. prediction vs. actual rating.
b) Square this deviation to get rid of the sign
c) Do this for all such deviations (e.g. for all users)
d) Sum up all squared deviations
e) Divide by the number of deviations to get the mean
f) Take the square root to undo the squaring in step b) and get back to the original units