Lab Exercise 10

1)

A)

Download all the reviewed human proteins from UniProt in XML format. For each protein, calculate the frequency of aspartic acid (i.e. the frequency of D in the amino-acid sequence) and the frequency of residues with "turn" secondary-structure (the latter is found in UniProt's data as features). Include in your analysis only proteins with a positive "turn frequency".

Plot the aspartic acid frequency vs. turn frequency for the human proteins, and include linear regression in your graph. Are the two frequencies correlated (according to both Pearson's and Spearman's correlations)? Is the correlation significant? Is it strong?

B)

Do transmembrane proteins (i.e. with a "Transmembrane" keyword in UniProt) have a different proportion of hydrophobic amino acids (say *I*, *L*, *V* and *M*)? Is the association significant? Is it strong?