Project suggestion 1 (15P):
In this project you will be given an output file of a common tool (BLAST) used in biology. BLAST is used to identify sequences in a given dataset that have a certain similarity to a given input query sequence. The identity tells you how similar the input sequence is to the hits found in the database and the e-value gives you the probability of that a hit was identified by chance.

1) Compile two lists of BLAST hits from the provided BLAST hit output file that meet the following criteria (a) an identity of 35% or higher (list1) or (b) an e-value of at least 10^{-7} or better (list2). Store the lists in new files. Determine how many hits are identical, how many are unique to criterion (a) and how many are unique to criterion (b) (**3P**). Print the answers to the screen. Does (a) or (b) retrieve more candidates (**2P**)? Print your answer to the screen. **Total 5P**

2) Create a heatmap illustrating the shared BLAST hits from criteria (1a) and (1b) with the IDs displayed on the left. The first column of the heatmap represents the identity and the second column represents the e-value (**2P**). Arrange the rows of the heatmap in descending order based on identity (**2P**) and provide legends for identity and e-value color codes to the right of the heatmap (**2P**). **Total 6P**

3) Retrieve the sequences of the shared best hits identified from criteria (1a) and (1b) from the provided sequence file. Print these sequences into a new FASTA file ensuring they adhere to the typical FASTA format **(4P)**. **Total 4P**