Homework 1

Homework can be submitted via the github link which will create a repository for you with basic template of files you can edit to solve the homework. See the the table for homework submission links which will help you create a github repository in the class team. https://piazza.com/ucr/fall2020/gen220/resources

1. Write a script called `download_count.sh` which does the following.
   - Download the data file https://ftp.ncbi.nlm.nih.gov/pub/UniVec/UniVec_Core from NCBI
   - Print out the count of the number of FASTA format sequences in this file - see Wikipedia FASTA format - each record starts with a `>`
2. Write a script called `summary_exons.sh` which summarizes the total length of exons in the file data/rice_random_exons.bed. These data are in the BED file format. The columns are "Chromosome", "Start position", "Stop position". The length of a feature (or exon in this case) is computed by doing the computation: STOP - START
   - read in the file
   - use a loop structure to read each line
   - add up the length of each exon by summing this into a variable
   - Print out the total length of exon features at the end.
   - You do not need to save this for each chromosome, just print out the total length for this example - however if this is too easy for you, go ahead and make a more sophisticated report which presents, per chromosome, the total length of exons as well as the total number of exons, and the average length of exons.
3. Write a script called `strand_gene_count.sh` to calculate the number of genes that are on the positive (+) and negative (-) strand in the file.

- https://fungidb.org/common/downloads/release-48/ScerevisiaeS288c/gff/data/FungiDB-48_ScerevisiaeS288c.gff