For your 2nd homework exercise (due Thu 29 Feb @ 12:45), please:

- Open a new issue on Github and give it a brief title like "Homework_2".
- Using VScode, develop a BASH script that performs the following operations (executables to be used are in parantheses):
    - Downloads unmasked release 58 version of the *E. coli* MG1655 genome sequence (*Chromosome.fa.gz) and GFF3 annotation (*Chromosome.gff3.gz) files from Ensembl (curl): https://bacteria.ensembl.org/Escherichia_coli_str_k_12_substr_mg1655_gca_000005845/Info/Index
    - Unzips the genomic fasta and GFF files (gunzip)
    - Loads the BEDOPS, BEDtools, SAMtools, and UCSC modules (BEDOPS/2.4.41-foss-2021b, BEDTools/2.30.0-GCC-11.2.0, SAMtools/1.14-GCC-11.2.0, and ucsc/443)
    - Converts the GFF file to BED format (convert2bed)
    - Filters the BED file to create a new BED file with only CDS regions (grep)
    - Creates a "genome" index file for BEDtools (samtools faidx)
    - Uses the CDS region BED file to create a complementary set of BED intervals for non-CDS regions (bedtools complement)
    - Generates two files of fasta sequences for all CDS and all intergenic (non-CDS) regions respectively (bedtools getfasta)
    - Computes the GC content of CDS and intergenic (non-CDS) regions (faCount -summary). Note: GC content (https://en.wikipedia.org/wiki/GC-content) is not the same as CpG composition (https://en.wikipedia.org/wiki/CpG_site).
- Be sure that you script includes a SLURM header in this script that sends stderr/stdout to log files & code your script so that output files are sent to a “scratch” directory (in /work/gene8940/yourusername).
- Commit this script into your local git repository, push the script to github (using VScode or GitHub Desktop), pull the script down on the teaching cluster (using `git pull` after you `cd` inside your class repository) and submit the script to the teaching cluster scheduler (using `sbatch`). When the script has executed successfully, post the following information in the Homework 2 issue:

1) The GC content of CDS and non-CDS intergenic regions;
2) A URL to the location of the script on GitHub;
3) The git revision of the script used for this analysis.

Good luck and please post a comment in Slack or reach out to me/Oliver if you have any questions.