Session 8 Recitation

1) Imagine that your PI has told you that they need you to do some data handling for them. They want you to take the genome of the West Nile virus (https://www.ncbi.nlm.nih.gov/nuccore/LC318700.1) and extract a particular gene region of interest (10000-10401 inclusive). Since a previous grad student wrote the code extractSeqRegion.py (on D2L), your PI tells you to just use that script and get the gene region by the end of the day. Try running the script on the genome file. What do you notice? How would you test and fix it?

2) Imagine that you fixed the script in #1 and your PI was so happy that they asked you to give your script to the summer intern so they can help you get sequence regions for other viruses. You've observed the summer intern being sloppy and not checking their work. What would you do to make your script more error-proof (don't actually code it; just think of edge cases)?

3) Go to the Jupyter website shown in the first video and look at some of the example Jupyter notebooks. Find something that looks relevant to your research. Click on either "Try it in your browser" or "Install the Notebook" on the homepage if you want to try running these notebooks.