Pulling data from Web of Science

A. Influential articles definition (`coreidd`)
    a. Identify & retrieve 23 influential records from WoS (call this generation 1)
        i. Query link (run on 11/25/2024): https://www.webofscience.com/wos/woscc/summary/7c87e64e-2a8c-44b2-86e0-57aab3b4244b-012b4e74ec/relevance/1
        ii. Export using Fast5000 as a txt file, copy contents into an Excel file, and save as gen1_download.csv
    b. To get info on these papers, run 01_gen1.py file to create input file for API call and output results of API call to csv
        i. This will first create files formatted properly for the API, then pull from the API; the process is similar for steps c-e
    c. To get info on the bibliographies of these papers, run 02_gen1_bibs.py to create input file for API call and output results of API call to csv
    d. Identify & retrieve articles that cite these highly cited articles (call this generation 2) using 03_gen2.py
    e. Identify & retrieve bibliographies of these citing articles (generation 2) using 04_gen2_bibs.py

B. Books definition (`books`)
    a. Identify books entries on WoS using below search, select all entries as of 04/13/2023 EXCEPT erroneous Intracellular Trafficking-titled entry
        i. Cited References Query: “Cited work = Infectious Diseases of Humans: Dynamics and Control OR Cited work = Modeling Infectious Diseases in Humans and Animals OR Cited work = Mathematical Tools for Understanding Infectious Disease Dynamics”
        ii. Export using Fast5000 as txt files, copy contents into an Excel file and save as gen1_download.csv
    b. Retrieve more information on these articles that cite the books (generation 1) by running 01_gen1.py
        i. This will first create files formatted properly for the API, then pull from the API; similar approach is used in step c
    c. Identify & retrieve bibliographies of these citing articles by running 02_gen1_bibs.py

Once these are done, move on to 05_join_datasets_clean.R, then to the files in process_data/