Pulling data from Web of Science

A. Highly cited articles definition
    a. Identify & retrieve 34 highly cited records from WoS (call this generation 1)
        i. Query link (run on 08/21/2023): https://www.webofscience.com/wos/woscc/summary/0fdf2693-bfa3-4651-a491-c4b1fc5e5460-9efab318/relevance/1
        ii. Export using Fast5000 as a txt file, copy contents into an Excel file, and save as gen1_download.csv
        iii. Optional: If you want info on these papers, run 01_gen1.py file to create input file for API call and output results of API call to csv
        iv. Optional: If you want 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
    b. Identify & retrieve articles that cite these highly cited articles (call this generation 2) using 03_gen2.py
        i. This will first create files formatted properly for the API, then pull from the API
    c. Identify & retrieve bibliographies of these citing articles (generation 2) using 04_gen2_bibs.py
        i. This will first create files formatted properly for the API, then pull from the API

B. Books definition
    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
    c. Identify & retrieve bibliographies of these citing articles by running 02_gen1_bibs.py
        i. This will first create files formatted properly for the API, then pull from the API

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