* This is a personal log of the books I have read spanning roughly 19 years (2001-2020). The majority of the books don’t have the year they were read in, only those from 2011 to 2020 are dated.
* The first stage of the data cleaning involved joining the two data sets I had for the books, one of them in a Word Document and the second one in Excel.
* To achieve this, I manually joined the two files in Excel. The data in the Word document had both the title and the author’s name on the same line, separated by a dash. I used the tools in Excel to separate these into two columns.
* Once I had both data sets on the same file. I loaded the csv into Jupyter Notebook to continue working with the data and filling out the gaps.
* In preparing the data for the Goodreads API, I had to go back and modify some of the author entries, which contained author name and series name or genre in the same column. The API also missed authors whose names were misspelled in the data set, so I went in and fixed the typos.
* Using an API from Goodreads, I was able to add genre and number of pages for most books.
* Having this additional information, I loaded the data into an SQLlite database. See images for ERD diagram.
* The next step was to create the Flask API to pull data from the database.
* Once the data was loaded, I proceeded to create my visualizations using D3
  + Books per year
  + Books divided by author sex
  + Map of books divided by author nationality.
* I loaded these into a dashboard deployed on using github.