

PS5

Becky Crouse

March 7, 2023

Question 3

For this exercise I scraped a table of historical income tax bracket information from Wikipedia. This table provides the tax rate applicable to the lowest and highest tax brackets since inception of the income tax in the U.S.

I think the information in this table is very interesting and while my studies typically focus on business taxes, I do think it possible that it could be useful to me in the future if I look at individual tax issues or the intersection of business and individual taxation.

I was having difficulty using the SelectorGadget browser extension, so I consulted ChatGPT. After continuing to run into issues, I eventually found the correct CSS selector information by simply using the inspect feature in Google Chrome. Finally, I executed the R script from Osker and got an error regarding an SSL certificate. A quick Google search brought me to the solution on the httr github page which I incorporated into my code.

Question 4

I have an interest in scraping data from firm annual filings available from the SEC EDGAR database. Typically I would obtain this data from Compustat, but information that is only found in 10-k footnote disclosures isn't always complete in Compustat.

I ran a program in SAS to find example companies with incomplete UTB (uncertain tax benefit) footnotes. I chose a company with information from 2008 missing and got its CIK number.

My first attempt to scrape this was using the rvest package for the homework Question 3. I wasn't able to get it to work right from the SEC .htm 10-k filing page so I decided to try using the free API offered by the SEC instead.

I pulled the JSON data from the SEC API and formatted the UTB footnote information into a table. Unfortunately the year I was looking for was not included. The particular API that I'm able to get working only includes XBRL data and XBRL tags weren't required widely until 2009. Thus, I realized the missing data issue from Compustat is possibly related to the absence of XBRL tags in the footnotes, meaning I'll need to keep exploring ways to efficiently obtain the missing years' data that I cannot find through the API.