* Step0 = figure out the URL that you will for scraping data
  + In the MU ISA Course Bulletin, this is somewhat trivial because we have that information (at the beginning or from a quick google search)
  + When we scrape multiple pages, we will need to figure out how the URL changes and how we can programmatically account for it (e.g., <https://www.imdb.com/search/title/?companies=co0144901&start=51&ref_=adv_nxt>)
* Step1 = read the backend of the HTML to R (or your software of choice)
  + Conceptually: view-source:https://www.imdb.com/search/title/?companies=co0144901&start=51&ref\_=adv\_nxt
  + In R, we will use rvest package and a function called read\_html()
* Step 2: We subset/extract only the relevant “things” we want from a webpage (e.g., for ISA courses, you might want to keep only the course name and course description)
  + In R, we will use either html\_element() or html\_elements()
  + Element is if you wanted a single result and the plural if you wanted multiple results
* Step3: We will clean the results from step2 (we will remove the HTML tags)
  + Text: 🡪 html\_text2()
  + Table: 🡪 html\_table()
* Step4: we will sometimes combine the results from steps 2 and 3 (typically put it in a dataframe or a tibble)