

# Problem Set 5

Lilly Kirby-Rivera

February 27, 2024

## Q3

I used the `rvest` package in R to scrape a simple table of pharmaceutical company sales on the pharmaceutical industry Wikipedia. It told me that Pfizer is the top company in pharmaceutical sales and is the only one to surpass a billion dollars in revenue. Additionally, the United States seems to have the majority of top grossing companies compared to other countries. I thought this would be interesting to look at since I have been trying to find a way to gauge the influence of pharmaceutical manufacturers. I did not use any online tutorials, just class notes.

## Q4

For this part, I used the NIH RePORTER API to collect data on funding for SSRI (selective serotonin reuptake inhibitor) related research projects. I decided to use this API since it has been recently improved and I was already loosely familiar with it from a previous research project. Additionally, I have been investigating the firm behavior and pricing of certain SSRI manufacturers, so I figured this search could contribute some insight. I used the data to generate a plot of average project funding by year, but I did not account for inflation, so the data is only in nominal USD for each calendar year. As expected, the nominal amount rises by quite a bit over the time period, but it is more likely consistent in real USD.

Conveniently enough, there is an existing R package called `"repoRter.nih"`, which I used to interact with the API. I also installed the recommended packages for setting up the workspace, as described in the package's documentation. Aside from that, I mainly used `ggplot` to get a visual representation of the data.