

Problem Set 4

Samuel Hunt

February 2025

1 Data Scraping Ideas

The focus of the research project I am engaged with this semester is on estimating the effects of tariffs on a number of economic and non-economic variables using econometric methods. Acquiring imports/exports data for countries worldwide would be helpful for this project, and so I am interested in scraping data from the UN Comtrade database. This website produces tables of imports/exports data for over 200 countries and over many years, with specific data offered for product sector type.

Additionally, I will need some health data for my project to measure possible effects on health outcomes. Therefore, I am also interested in scraping the Center for Disease Control's National Center for Health Statistics Database. This website also contains a large database with U.S. statistics over a number of variables and years.

For a final random data source, I am a big NBA fan, and if I was going to practice data scraping for fun I think a good place to practice would be getting statistics off of the nba.com website, which contains a number of historical statistics for players and teams.

2 sparklyr questions

Unfortunately, I was never able to load the sparklyr package onto OSCER, this is because the install command resulted in a "nonzero exit status" error. Adding the piece of code to my bash profile did not resolve this, and I have heard that it is an issue caused by OSCER having an out of date command line interface package that creates lots of issues upon attempting to install packages. If a fix exists for this issue I would absolutely love to go back and complete this section.