DS8 Midterm Proposal - Andrew Richard

Project Title: Exploring the relationship between out of pocket medical expenses and prevalence of diseases and disorders.

Executive Summary

The United States is largely viewed as a leader in healthcare practices, but also has some of the highest expenses for healthcare in the world. This app will explore the relationship between various health metrics and the out of pocket expenditure by country over time. Data from the World Bank, WHO, Global Health Data Exchange, OECD, and potentially others will be used for reference in this app.

Motivation

Health and healthcare data are some of the most prevalent but underutilized data out there. There is a wealth of information that is collected that can be used for analysis. As there is more coverage on the U.S.'s healthcare costs, it might be expected that there would be some correlation between higher cost and higher prevalence of some disorders / diseases.

Data Question

What is the relationship between the amount a person must spend out of pocket on medical expenses and the prevalence of diseases like obesity, heart disease, or others of the like? Is the relationship statistically significant?

Minimum Viable Product (MVP)

The MVP will include:

- interactive visualizations showing trends in health data over time in relation to out of pocket medical expenses for various countries.
- interactive tables displaying the data
- multiple filters to select data desired

Schedule (through 2/15/2025)

1. *Get the Data (finish date)*
2. *Clean & Explore the Data (finish date)*
3. *Create Presentation and Shiny App (finish date)*
4. *Internal Demos (2/11/2025)*
5. *Midcourse Project Presentations (2/15/2025)*

Data Sources

[1] "Global Health Data Exchange | GHDx." Accessed January 24, 2025. https://ghdx.healthdata.org/.

[2] "OECD Data Explorer • Healthcare Coverage." Accessed January 24, 2025. https://data-explorer.oecd.org/vis?fs[0]=Topic%2C1%7CHealth%23HEA%23%7CHealth%20expenditure%20and%20financing%23HEA_EXP%23&pg=0&fc=Topic&bp=true&snb=5&vw=ov&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_HEALTH_PROT%40DF_HEALTH_PROT&df[ag]=OECD.ELS.HD&df[vs]=1.0&dq=.A...&pd=1990%2C&to[TIME_PERIOD]=false&ly[cl]=TIME_PERIOD&ly[rs]=INSURANCE_TYPE&ly[rw]=REF_AREA%2CCOMBINED_UNIT_MEASURE.

[3] DataBank | The World Bank.Available at https://databank.worldbank.org/home.

[4] Health Nutrition and Population Statistics | DataBank.Available at https://databank.worldbank.org/source/health-nutrition-and-population-statistics#.

Known Issues and Challenges

establishing relationships between the datasets may prove to be a challenge, since the data come from different sources and will be formatted differently. There will also be some navigating times and ensuring data line up.

Making visualizations interactive may also prove to be a challenge.