

WHO Cares about Healthcare?

Trisha Dwivedi (td2704), Pooja Mukund (pm3148), Ragyie Rawal (rr3423), Safiya Sirota (sns2189), Eric Wang (ew2717)

Motivation

The Global Health Expenditure Database (GHED) is an open-access data source which contains health care spending data from 2000 to 2019 for almost 190 countries. The GHED is utilized by the World Health Organization (WHO) to monitor availability and distribution of global health resources. Through our analysis of this dataset, we hope to understand and present healthcare funding patterns in various countries and across different income levels and WHO regions. We want to show who is bearing the brunt of the costs and how spending distributions differ throughout the world. Several countries rely on the WHO for healthcare, treatment, and research funding, so recognizing underfunded areas is paramount to ensure care for everyone globally. This project is overall motivated by the health disparities throughout the world. We also hope to compare spending in specific countries across different areas, including: primary health care, preventative care, curative care, infectious diseases, and noncommunicable diseases. We want to analyze the availability of resources for health and the extent to which they are used efficiently and equitably.

Related Work

When we were brainstorming ideas for this project, we realized we all have a common interest in examining global healthcare systems and understanding global health disparities. We found the WHO Health Financing website, which is where we discovered the importance of the GHED and decided to conduct our own analysis of this dataset.

The United Nations (UN) has a document which details country classifications, and this inspired us to focus some of our analysis on examining health care spending over time in developed and developing countries.

WHO Health Financing: <https://www.who.int/data/gho/data/themes/topics/health-financing>

UN Country Classification:

https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf

Initial Questions

Initially, we wanted to learn more about the following questions:

1. How does spending compare per capita between countries with high and low socioeconomic status?
2. Are countries putting enough resources into their most prevalent diseases?
3. How much do different countries spend on curative vs. preventive care?
4. How much do out of pocket costs, government funding, and private insurance companies contribute?

We managed to evaluate most of the questions, except for the second one. After some evaluation of the dataset, we realized there was not enough data available to extensively evaluate expenditure on specific diseases in countries outside of the WHO Africa region.

While looking through the international data, we realized we also wanted to focus in on the United States. Specifically, we wanted to investigate patterns of government spending over time and in different states. We were able to use our original WHO dataset to find information on yearly spending since 2000, but we did not have state-level data. Therefore we sought out a second, more granular dataset to compare expenditure between U.S. states.

Data

Data Sources

Our primary data source is the Global Health Expenditure Database:

<https://apps.who.int/nha/database>

Our secondary data source is the State and Local Finance Database, which we used for evaluating patterns in health care spending in specific US states:

<https://state-local-finance-data.taxpolicycenter.org/pages.cfm>

Data Cleaning

We did not need to do any scraping since we downloaded the dataset directly from the “Data Explorer” tab on the GHED website. After discussing with our TA, we decided to download the whole available dataset containing all variables instead of selecting for specific variables. This was because we wanted to have the whole dataset available in our repository for reproducibility purposes, and so we could have access to all variables in case we wanted to expand our exploratory analysis.

The data we downloaded was already very clean, though there were columns with many “NA” values representing missing data. At the beginning of our analyses, we imported the data and used ``janitor::clean_names()``. After that, we wrangled the dataset appropriately for each separate analysis.

Exploratory Analysis

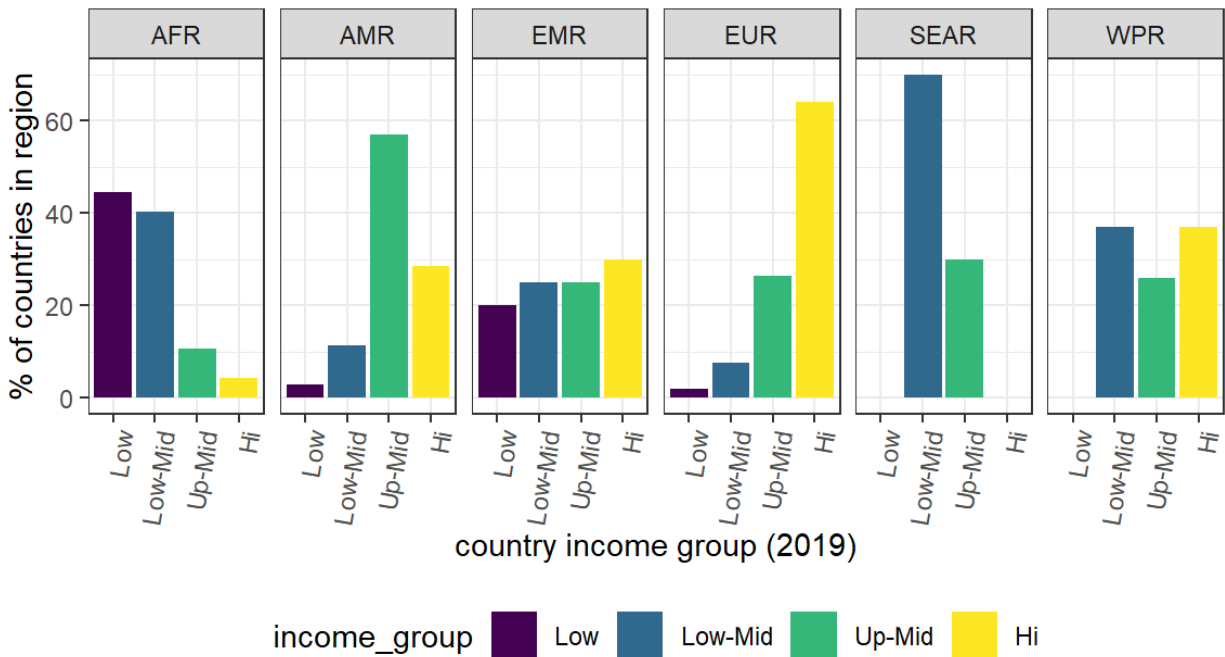
We started our exploratory analysis by defining which income levels, WHO regions, and specific countries to compare and analyze.

To conduct income-level analysis, we used the income levels which were pre-defined in the GHED: Low, Lower-Middle, Upper-Middle, and High. To conduct region analysis, we utilized the WHO regions designated in the GHED dataset. These regions are: AFR (African Region), AMR (Region of the Americas), EMR (Eastern Mediterranean Region), EUR (European Region), SEAR (South-East Asian Region), and WPR (Western Pacific Region).

We considered multiple strategies when selecting which countries to compare and analyze. Ultimately, we decided to use the UN Country Classification document as a reference source (https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf) to identify developed and developing countries in order to compare government spending in developed countries with high incomes and government spending in developing countries with low incomes. For developed countries, we decided to analyze the G-7 countries since this group is made up of the world's largest developed economies. For developing countries, we picked one country from each defined region to try to get a global overview of developing economies.

Global Analysis

Distribution of income levels of countries in each WHO region



The bar plots above show the distribution of income levels of countries (categorized in 2019 since this is the most recent available data) in each WHO region. The figure depicts how the highest proportion of low income is in the AFR region, low to middle income is in the SEAR region, upper to middle income is in the AMR region, and high income is in the EUR region.

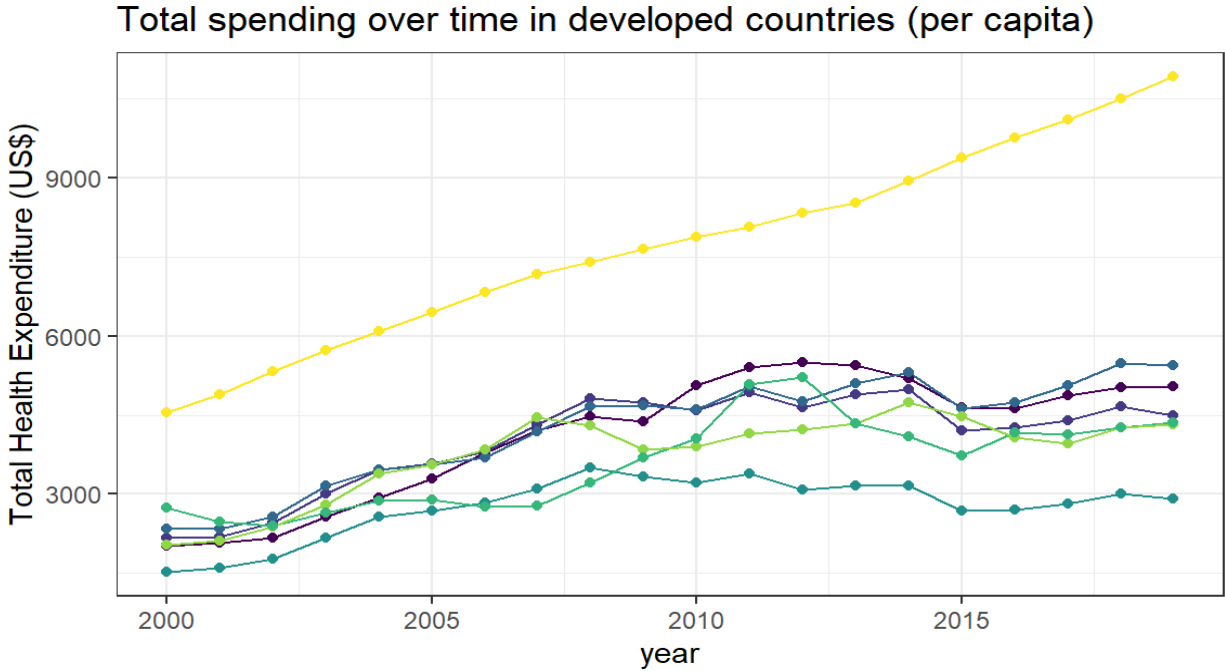
income_group	Mean Current Health Expenditure as % of GDP	Mean Current Health Expenditure per Capita (US\$)
Low	6.01	39.25
Low-Mid	5.08	126.69
Up-Mid	6.90	482.74
Hi	7.69	2937.29

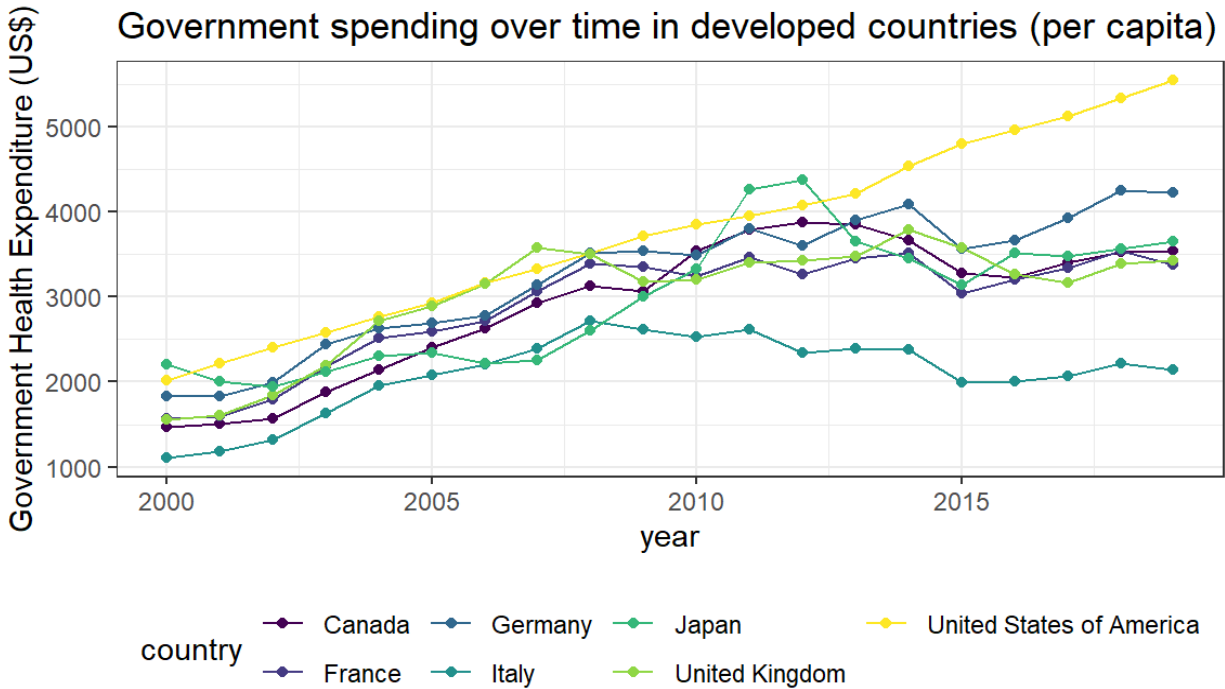
The table above shows that the mean current health expenditure (CHE) per capita in US\$ increases going from low income to high income groups. The highest CHE as % of gross domestic product (GDP) belongs to countries in the highest income group at 7.67%; however, we see that this percentage for all income groups ranges from roughly 5-8%, so the amount per capita a country is able to dedicate for health expenditure depends and varies based on country GDP and income level.

region_who	Mean Current Health Expenditure as % of GDP	Mean Current Health Expenditure per Capita (US\$)
AFR	5.33	128.38
SEAR	4.42	186.29
EMR	5.52	634.23
AMR	7.20	1081.32
WPR	7.32	1107.48
EUR	7.63	2461.09

As shown by the table above, the AFR region has the lowest mean CHE per capita (US\$) while the EUR region has the highest. While the expenditure as % of GDP ranges from 4-8%, we note that the EUR region has countries that, on average, spend the highest percentage of their GDP on health at 7.6%. So even though the EUR region has the highest proportion of high income countries, these countries still dedicate the highest proportion of their GDP to healthcare, followed by the WPR and AMR regions.

The major developed economies we examined are the G-7 countries, consisting of: Canada, Japan, France, Germany, Italy, United Kingdom, and the United States of America.

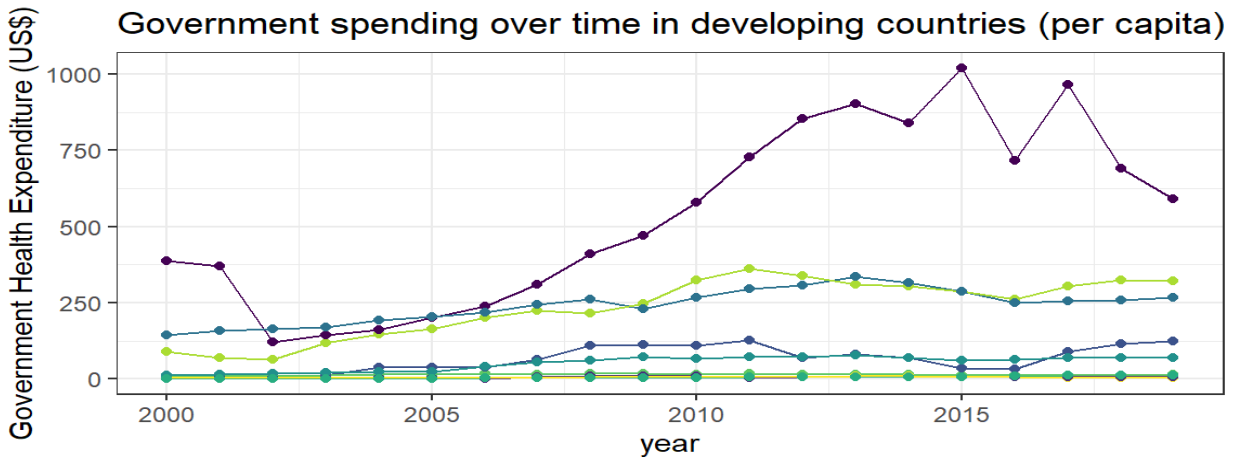
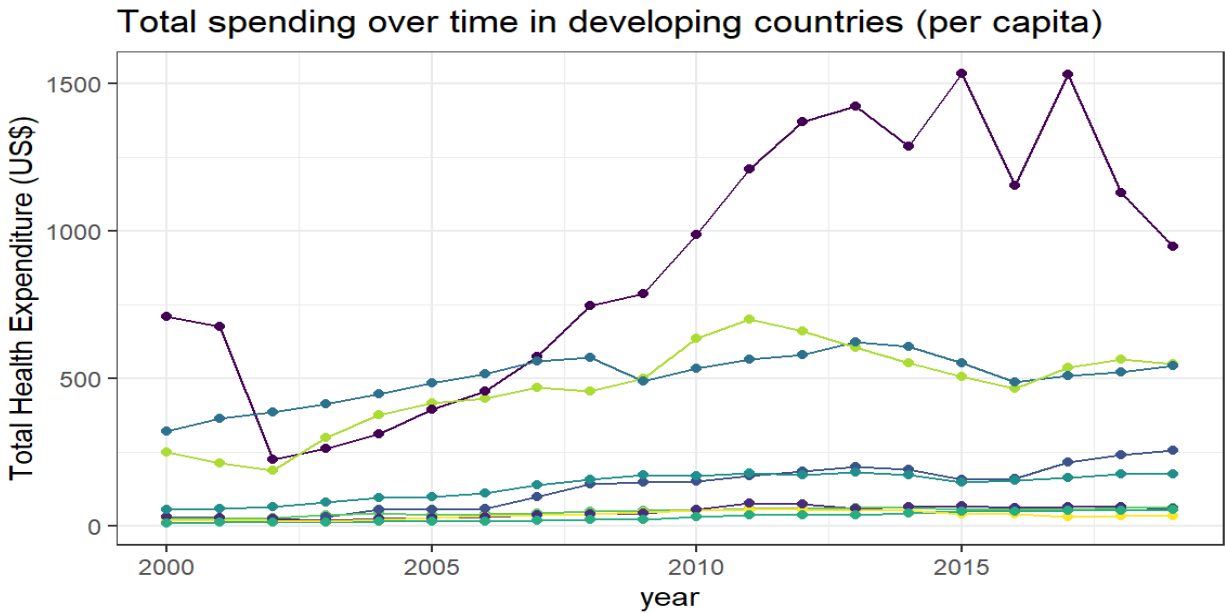




As we can see from the above line plots, the United States has the highest total and government spending per capita over time and grows at a steeper rate year over year than the other countries. All countries have higher health expenditure in 2019 compared to 2000. It's interesting to note that Japan had a spike in government health expenditure around 2011 and 2012, exceeding the US - upon looking into this, we note that the deadly earthquake and tsunami that hit Japan in 2011 coincides with this increase. With thousands dead and injured and many resources scarce, the government stepped in to provide relief to its citizens and help with recovery and health concerns after the natural disaster struck.

Note that total spending per capita ranges from \$1500 to \$11000 and government spending per capita ranges from \$1000 to \$5500.

The developing economies we dive into are: Morocco, Uganda, South Africa, Senegal, Nepal, Iraq, Haiti, Mexico, and Argentina.



country ● Argentina ● Haiti ● Mexico ● Morocco ● Senegal ● South Africa ● Uganda

As we can see from the above line plots, among these developing economies, Argentina has the highest total and government spending per capita over time and grows at a steeper rate year over year. According to an Economist article (https://solidfoundations.economist.com/reports/Amgen_Argentina_English_Final.pdf), Argentina's government has been devoting more resources to healthcare over the years to address healthcare inequalities. Some of the African countries, such as Senegal, South Africa, and Uganda, don't seem to have increased health spending much, if at all, over the last two decades.

Note that total spending per capita ranges from close to \$0 to \$1500 and government spending per capita ranges from close to \$0 to \$1000. Both these ranges are much less than that of the G-7

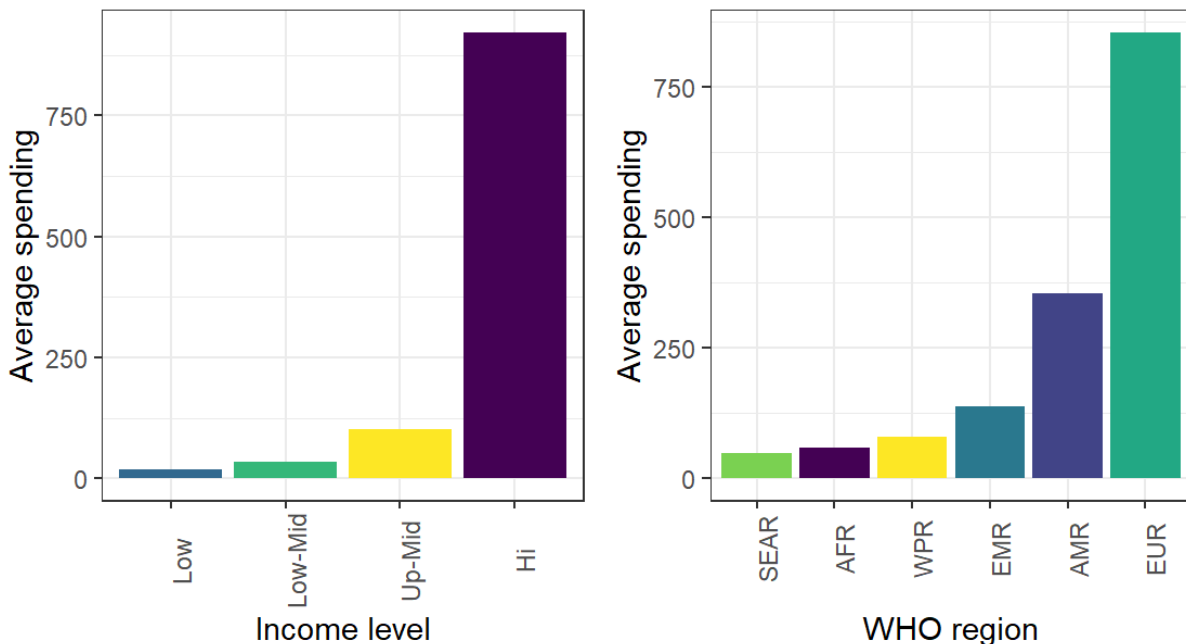
developed economies mentioned above: total spending per capita from \$1500 to \$11000 and government spending per capita from \$1000 to \$5500.

Health Expenditure by Category

We wanted to continue our global analysis by examining health care expenditure by category across the defined income levels and WHO regions. We focused on expenditure in 2019 since this was the most recent available data. We wanted to examine primary health care, preventative care vs. curative care, and infectious diseases vs. noncommunicable diseases. We focused on evaluating expenditure in million constant US\$, expenditure as a percent of GDP, and expenditure as a percent of CHE. We used million constant US\$ as the unit of expenditure in this analysis to help standardize comparison across the countries in each region, and because there was more data available for this particular unit.

We started by examining the average primary health care spending across income levels and WHO regions.

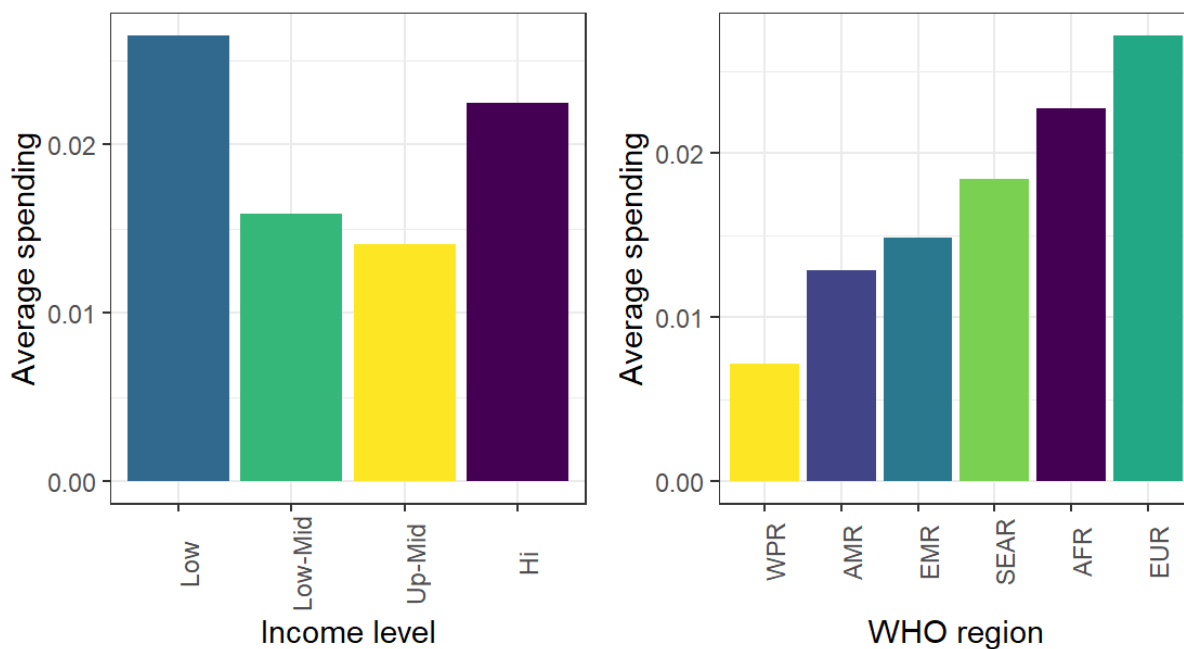
Primary health care spending in million constant USD\$



The bar plots above show the average primary health care (PHC) spending in million current US\$ in 2019 across income levels and WHO regions.

High income countries spent the most on average on PHC, while low income countries spent the least. The European region spent the most on average on PHC, while the South-East Asian region spent the least on average on PHC.

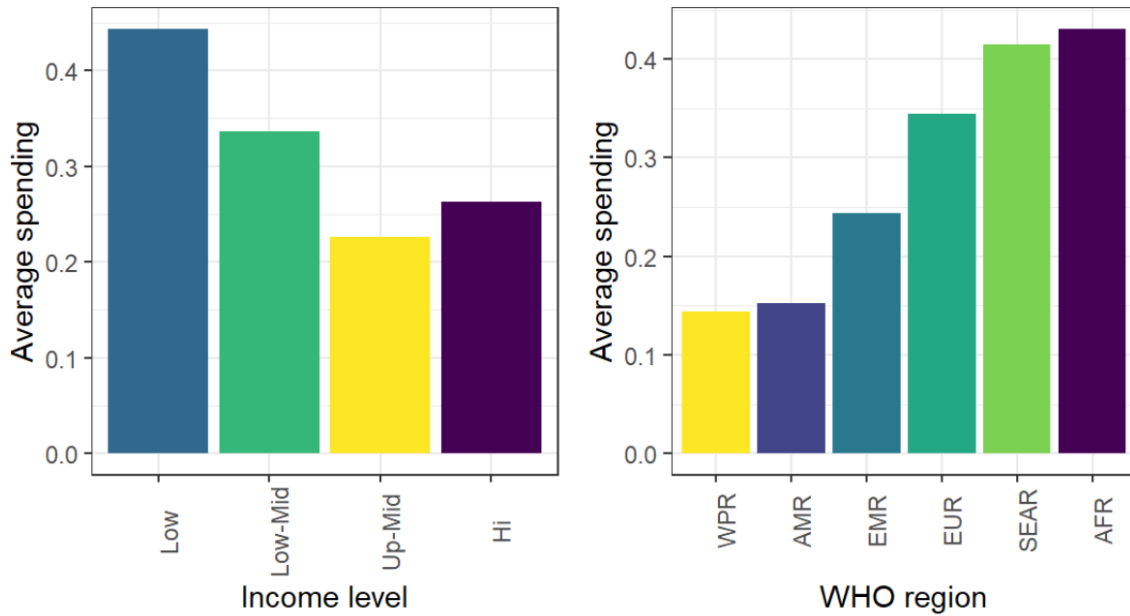
Primary health care spending per capita GDP



The bar plots above show the average primary health care (PHC) spending as a percent of GDP in 2019 across income levels and WHO regions.

Low income countries spent the most on average on PHC as a proportion of GDP, while up-mid income countries spent the least. It is interesting that low income countries actually spend more as compared to high income countries. The European region spent the most on average on PHC as a proportion of GDP, while the Western Pacific region spent the least.

Primary health care spending per capita current health expenditure

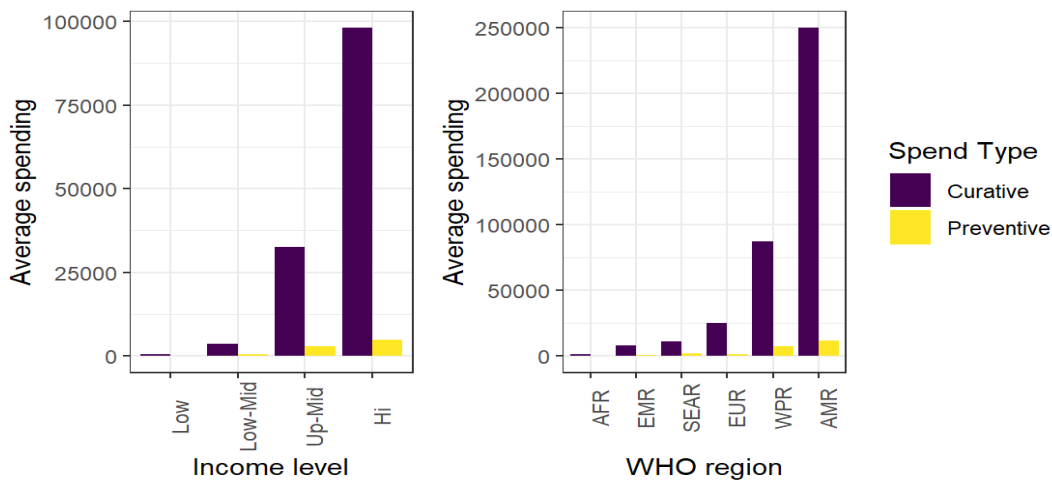


The bar plots above show the average primary health care (PHC) spending as a percent of current health expenditure in 2019 across income levels and WHO regions.

Low income countries spent the most on average on PHC as a proportion of current health expenditure, while upper-middle income countries spent the least. Low income countries actually spent significantly more in comparison to high income countries. Interestingly, African region countries spent the most on average on PHC as a proportion of current health expenditure, while Region of the Americas and Western Pacific region spent the least.

We continued our analysis by comparing average preventative spending and curative spending.

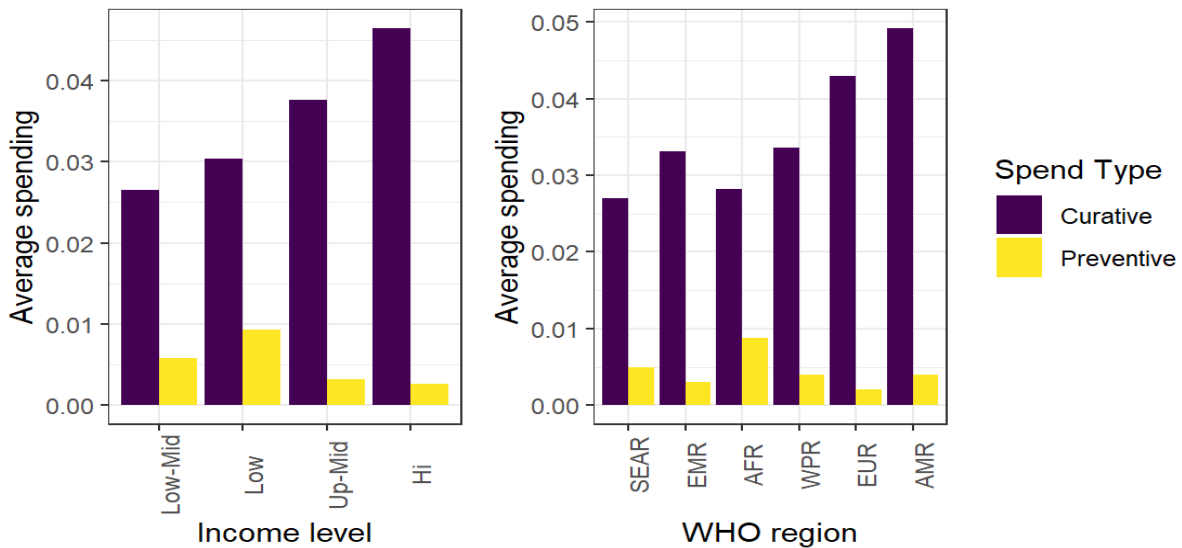
Preventive vs. curative spending in million constant USD



The bar plots above show the average preventive vs. curative spending in million current US\$ in 2019 for income levels and WHO regions.

Across all income levels and WHO regions, countries on average spent much more on curative care than on preventative care. This reflects that global health care practices seem to prioritize curative care, and focus less on preventative care for patients. On average, high income countries spent the most on both curative and preventative care, while low income countries spent the least. The region of the Americas spent the most on both curative and preventative care, while the African region spent the least.

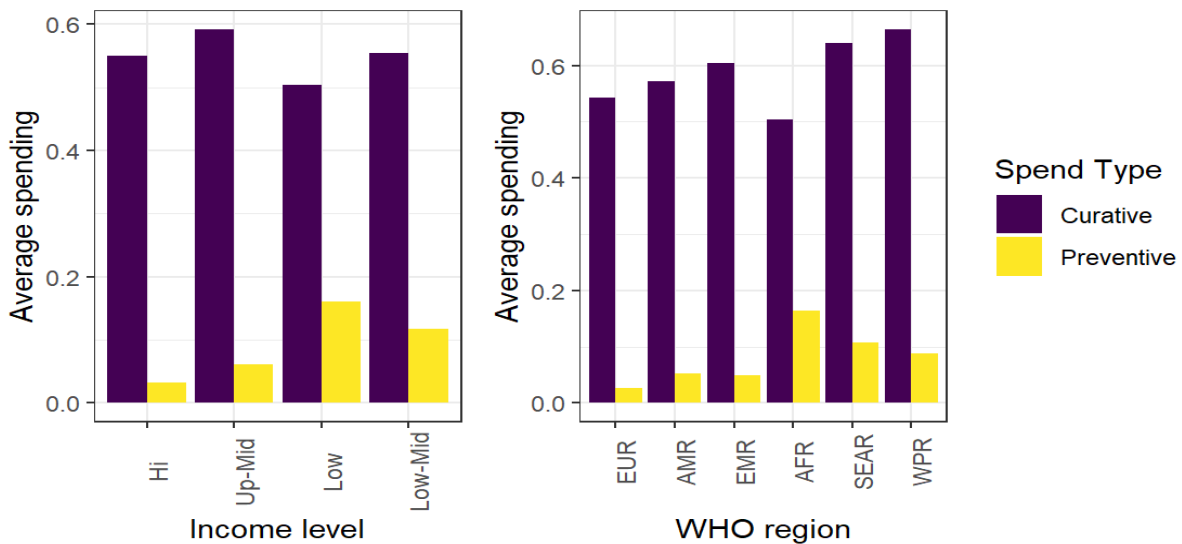
Preventive vs. curative spending per capita GDP



The bar plots above show the average preventive vs. curative spending as a percent of GDP in 2019 for income levels and WHO regions.

On average, countries across all income groups and WHO regions spent much more on curative care than on preventative care as a percent of GDP. High income countries spent the most on average on curative care. However, low income countries actually spent the most on preventative care as a percent of GDP. The region of the Americas spent the most on average on curative care. It is interesting that the African region actually spent the most on preventative care as a percent of GDP.

Preventive vs. curative spending per capita current health expenditure



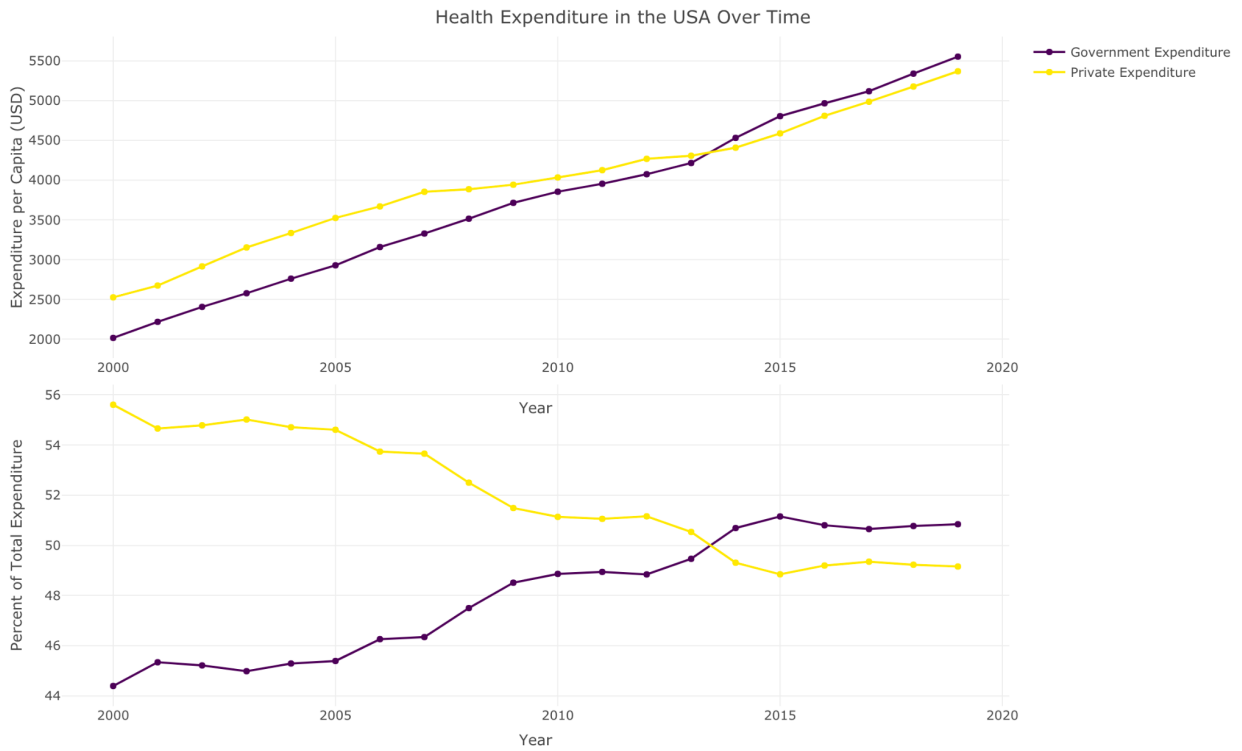
The bar plots above show the average preventive vs. curative spending as a percent of current health expenditure in 2019 for income levels and WHO regions..

On average, countries in all income levels and WHO regions spent significantly more on curative care than preventative care as a percent of current health expenditure. Upper-middle income countries actually spent the most on curative care, while low income countries spent the least. High income and lower-middle income countries spent relatively similar amounts on curative care as a percent of current health expenditure. Again, low income countries spent the most on average on preventative care as a percent of current health expenditure. The Western Pacific region actually spent the most on curative care, while the European region only spent the second highest on curative care. The African region spent the least on curative care, though countries in this region actually spent the most on average on preventative care as a percent of current health expenditure.

We hoped to continue our analysis by comparing infectious and non-communicable disease spending between countries of different regions and income-level, similarly to how we compared preventive vs. curative spending above. However, we found that data on these categories of disease, as well as how much spending goes towards more specific categories (e.g HIV, reproductive health, injuries etc.) was sparse for countries outside of the African region. Thus we did not feel we could conduct an in-depth analysis for this.

Zoom In: USA

We decided to further our analysis by specifically zooming in on the United States (US). We were interested in health care expenditure patterns in the US because we all live here and have experienced the health care system first hand. In addition, based on the line graphs displaying the total spending over time in developed countries per capita, the US appears to spend the most and grows at a steeper rate each year. This is a trend which we wanted to further analyze.

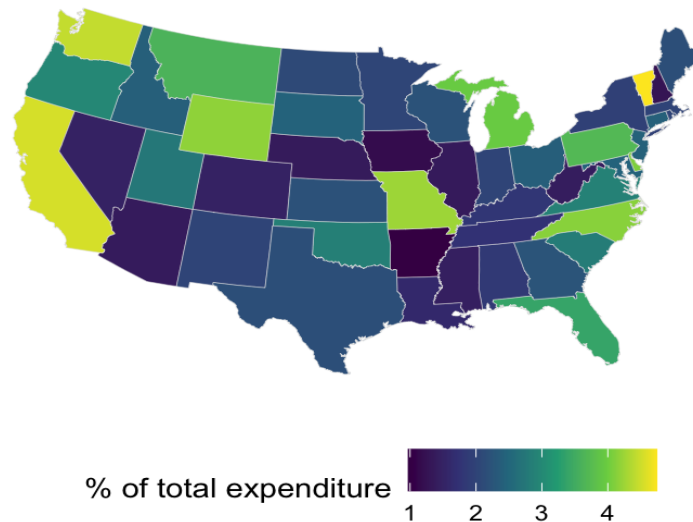


These line plots above visualize health spending in the U.S. from 2000-2019 in two different ways: in per capita US\$ and in % of total healthcare expenditure. The graph on the top highlights how the cost of health per person has increased each year for the last two decades. The graph on the bottom shows how government spending on health gradually overtook private spending around 2013. Both graphs show that the amount and percent of government expenditure is increasing in the U.S. In the early 2000s, private expenditure was a majority of the total US health expenditure, but this is continuing to decrease as the US puts a greater emphasis on government expenditure. This is likely a result of recent elected officials focusing on creating more government-sponsored healthcare programs.

Another interesting trend we noticed from these line plots is a shift towards more government expenditure per capita US\$ and in % of total healthcare expenditure starting around 2013. In 2000, there was significantly more private expenditure, but over time private expenditure decreased and there was a switch to there being more government expenditure around 2013. This could be due to the introduction of the Affordable Care Act and its implications. Based on this report (<https://www.ncbi.nlm.nih.gov/books/NBK241401/>), the President Obama introduced the Affordable Care Act in 2010, and open enrollment into the program began in 2013. This could explain why there was a shift towards more government expenditure around 2013.

We also wanted to examine patterns in health care expenditure in the US in specific states. We did this by creating a map displaying the % of total expenditure towards health in 2019 for each state.

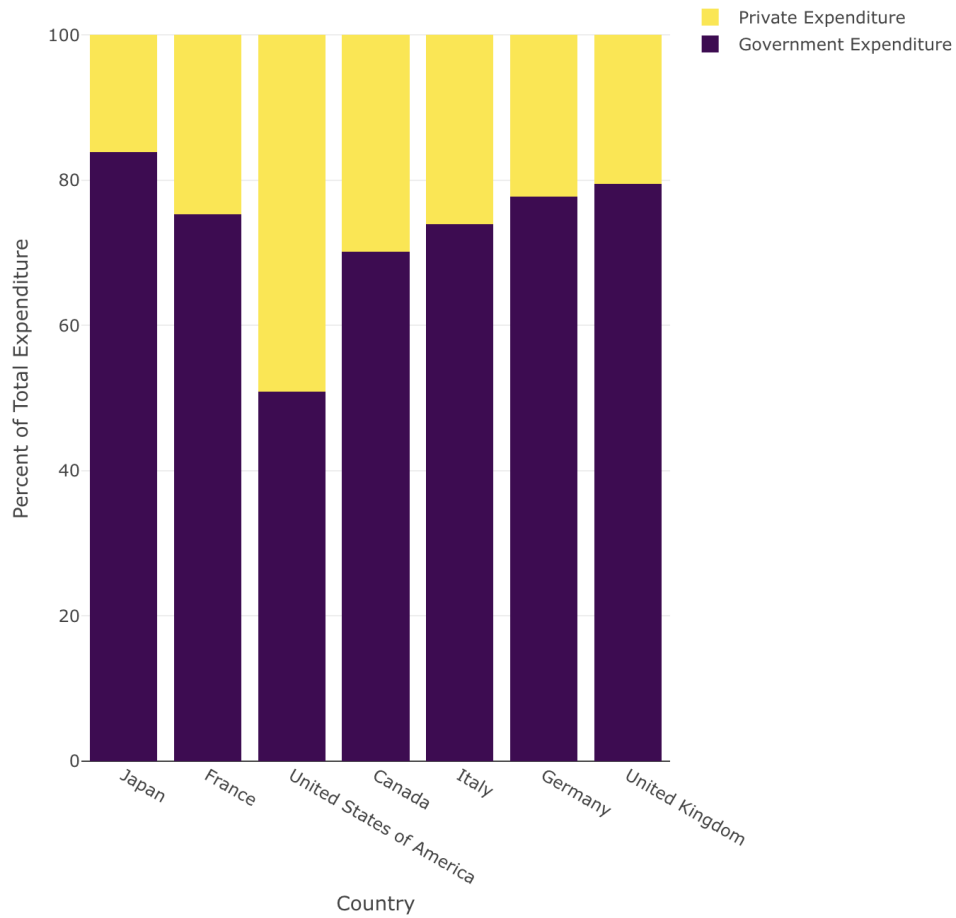
% of Total Expenditure Put Towards Health in 2019



We can see here from this map that the states that put the most of their state and local spending towards health care are Vermont, California, and Washington. The states that contribute the least of their expenditure towards health care are Arkansas, Iowa, and New Hampshire. It's interesting to see how spending varies so widely, even within regions. Take the drastic difference between New Hampshire and Vermont as an example. They are neighbors, but their health expenditure varies significantly.

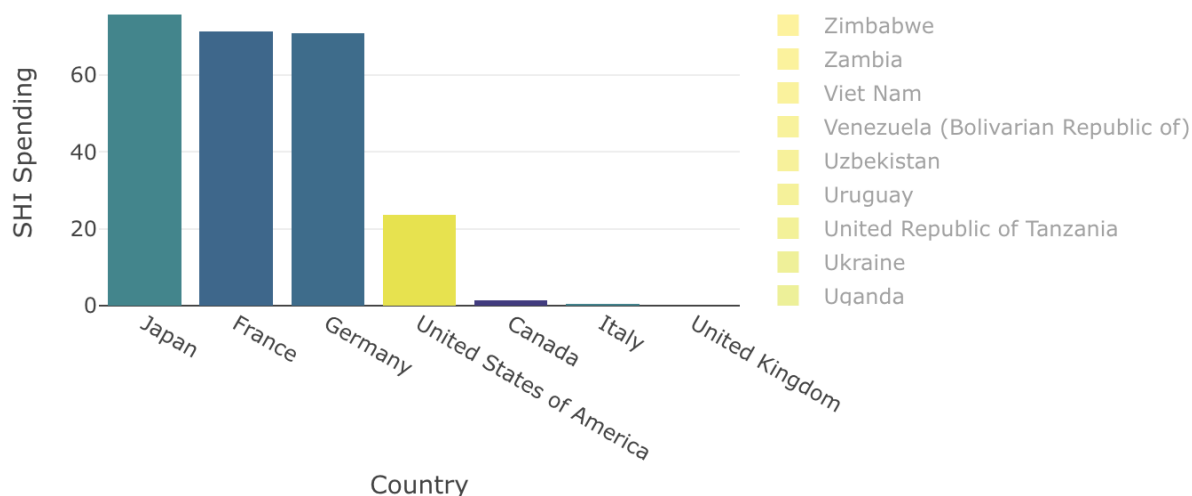
Dashboard: Like a G7

Previous studies have shown that the US spends more on healthcare compared to other developing nations, but has worse healthcare outcomes (http://blogs.bu.edu/ellisrp/files/2013/04/2013_EllisChenLuscombe_Comparisons_EHE_20130425b.pdf). Many other developed countries also have universal healthcare coverage or government programs for citizens, but the US relies much more heavily on private insurance programs. We wanted to explore this topic by comparing the US to other developed countries in terms of different types of health expenditure. To do this, we created a dashboard displaying types of expenditure in developed countries.



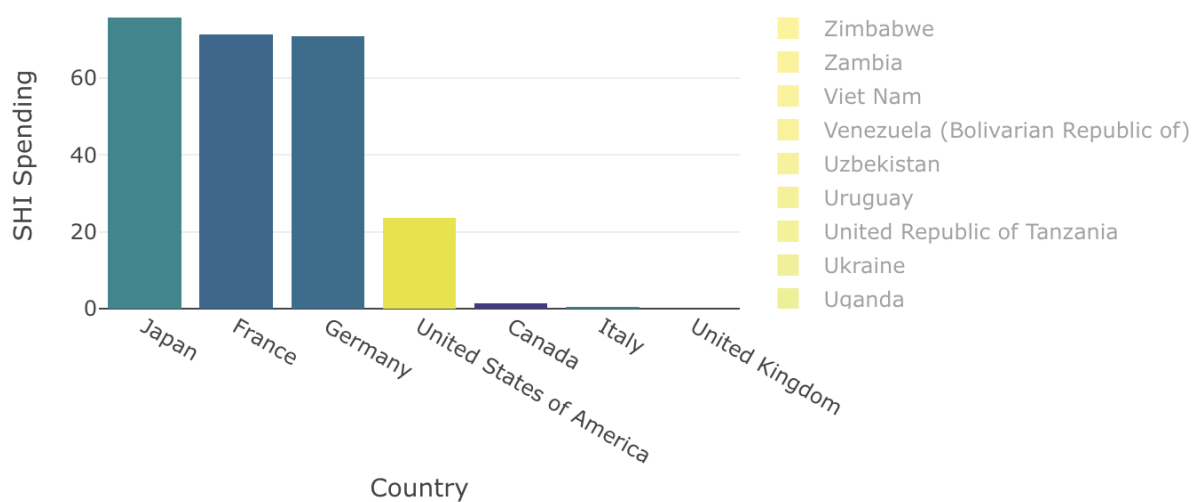
As shown in the stacked barplot above, we first looked at the percent of private expenditure to government expenditure in developed countries. Not surprisingly, the US had the lowest percent of government expenditure with 50.84% of expenditure coming from the government and 49.15% of expenditure coming from private organizations. All other developing countries' expenditure reflected a majority coming from the government. Japan had the highest percentage of government expenditure at 83.86% followed by the United Kingdom (UK) at 79.47%.

Social Health Insurance as % of Current Health Expenditure



As shown in the barplot above, we then decided to look at social health insurance, and found that the UK and Italy had the lowest percent of social health insurance. Whereas Japan, France, and Germany had the highest in that order. Germany, France, and Japan have wider-known social health insurance systems, but the UK started with social health insurance but over time moved away from this system (<https://www.healthaffairs.org/doi/10.1377/hlthaff.2019.00874>). The US has started to put more emphasis on social insurance programs with the development of Medicare and Medicaid in the 1960s, but still heavily relies on other forms of private and government health schemes.

Social Health Insurance as % of Current Health Expenditure



As shown in the barplot above, we looked at Out-of-pocket (OOP) expenditure as a percentage of current health expenditure. We found that the Republic of Korea had the highest percentage of OOP expenditure, while France had the lowest percentage of OOP expenditure. France and Germany have had historically low OOP expenditure which matches what we see in the graph. Interestingly, the US has the second to last OOP expenditure, which is not what we expected to see. It would be interesting to dive deeper in a later analysis to understand why the US reports lower OOP costs. The Republic of Korea, Italy, and the UK had the highest OOP costs. On the dashboard on our website, we also added the option to look at all countries with available data for the last two plots.

Additional Analysis

After looking at global and regional trends, we were curious about the possibility of predicting the income group given a country in a certain WHO region, and amounts of private and government expenditure and external funding.

We wanted to apply linear models as learned in class to this problem. But after trying an initial linear model with income group as the response variable, we realized that since income group is a categorical variable, it didn't make much sense using 'lm' to predict a 4-level categorical variable (1: low, 2: low-mid, 3: up-mid, 4: high). Thus, we looked to other potential classification models. We settled on using a Random Forest model. Random Forest is an algorithm that generates multiple decision trees that branch into different variables to create paths for classifying an input vector as a specific output using feature values and thresholds; the classification model classifies an input based on what class is assigned by the majority of the decision trees created.

We used 2,340 as a train set, and 1,000 as a test set. We use the train set to train the Random Forest model and test the prediction accuracy using the test set. The confusion matrix for the model created is shown below. We note that class errors for each of the income groups range from 13-22%, meaning a 78-87% accuracy for each of the classes.

```
Confusion matrix:
      1   2   3   4 class.error
1 263  55   8   0  0.1932515
2  56 491  63   1  0.1963993
3   6  76 464  48  0.2188552
4   0   3  75 489  0.1375661
```

We then ran our test dataset through the trained Random Forest model from above. The following is the confusion matrix for the predictions of our 1,000 size test set.

	pred_1			
	1	2	3	4
1	129	21	1	0
2	20	225	32	0
3	5	29	237	20
4	0	0	30	251

We get an accuracy of 84.2% based on the above results. So 84.2% of the 1000 test set data rows were classified to the correct income group based on WHO region, and amounts of private and government expenditure and external funding, showing this Random Forest model performs well using these predictors.

General Difficulties

We struggled at the beginning to create the website. We did not have 'github.io' in the original repository we created, so we created a new repository with an updated repository name which solved the problem. We also had some trouble with committing to the repository, which required some trial and error to solve. Lastly, we struggled with determining which variables to use to compare different categories of interest, since the categories we originally picked had a lot of NA values and no data for certain countries and regions.

Discussion

Key Findings

We noticed some interesting trends through this analysis. On a global level, the EUR region has the highest proportion of high income countries; these countries dedicate the highest proportion of their GDP to healthcare, followed by the WPR and AMR regions. Among the highly developed economies of the G-7 countries, the United States has the highest total and government spending per capita over time and grows at a steeper rate year over year than the other countries. Japan had a spike in government health expenditure around 2011 and 2012, exceeding the US - upon looking into this, we note that the deadly earthquake and tsunami that hit Japan in 2011 coincides with this increase.

In addition, countries with developing economies spent close to \$0 to \$1500 per capita on total healthcare spending while G-7 developed economies spent between \$1500 to \$11000 per capita. In spite of this, taking a look at categories on a detailed level, we noticed that low income countries spent the most on average on primary healthcare as a proportion of GDP, while upper-middle income countries spent the least.

Across all income levels and WHO regions, countries on average spent much more on curative care than on preventative care. This reflects that global health care practices seem to prioritize curative care, and focus less on preventative care for patients. The AMR region spends the most on average on curative care, and it's interesting that the AFR region actually spent the most on preventative care as a percent of GDP.

Finally, based on WHO region and amounts of private and government expenditure and external funding, we found that a Random Forest classification model performs well using these variables to predict income level groups, with an 84.2% accuracy.

Further Analysis

We noticed that the AFR region had lots of country-level data for most of the countries in its region (more than any other region) when it comes to classifying spending based on disease type, such as noncommunicable diseases, infectious diseases, and more. For future work, we'd take a closer look at the specific categories of disease spending for countries in the AFR region, and search for news articles and outbreak history to help correlate spending to historical and current health issues that countries in the AFR region have experienced.

We could also search for a new dataset on just US state-level data to see if we note any stark differences in health spending and spending categories beyond what we already looked at as percent of total expenditure that is devoted to health. It would be interesting to see if there are state-level or region-level differences within the United States.