

# W.H.O. is Sick?

## Basic Info

---

### Team:

- Paarth Lakhani : u0936913 : [paarth.lakhani@utah.edu](mailto:paarth.lakhani@utah.edu)
- Matt Willden : u0920335 : [willdenms@gmail.com](mailto:willdenms@gmail.com)
- Naoki Tominaga : u0876779 : [nrtominaga@gmail.com](mailto:nrtominaga@gmail.com)

### Github:

- [https://github.com/willdenms/WHO\\_is\\_sick](https://github.com/willdenms/WHO_is_sick)

## Background and Motivation

---

In looking at the world, it is interesting to note how diverse health problems can be. It seems like every region has a full variant of different diseases when comparing them to other areas; for instance, some areas will have a high frequency of cancer, while others have low cancer rates, yet other diseases are more rampant.

Health data, in general, is something that is well tracked by several organizations, but it seems like the mountain of data is very hard to discern. Further, it seems evident that there are more factors at play than just a countries region as to what contributes to mortality rates and disease burden.

The motivation of our visualization will be to first and foremost show how different countries are affected by various diseases including mortality rates, which will be broken down by country, age, and gender. Further, we would like to see how different socioeconomic data correlates to the mortality rates. The socioeconomic data could be GDP, health coverage, the standard of living, etc.

## Project Objectives

---

- What are the top countries affected by a certain disease?
- How are mortality rates affected by different socio-economic factors like: standard of living, GDP, and health coverage?
- What regions are more susceptible to specific disease?

- Optional) How does this data change over time?
- (Optional) What are the top diseases affecting the mortality rates of a selected country?

## Data

---

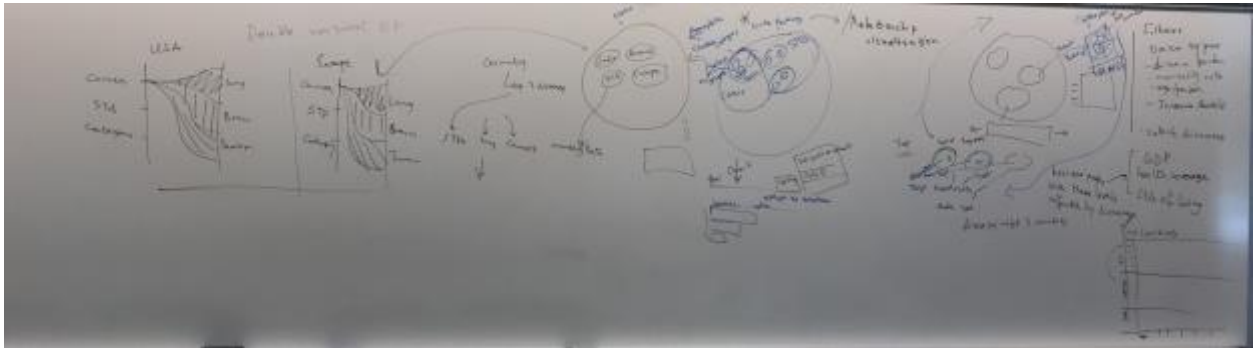
### Data Sources:

- Database reference: <http://ghdx.healthdata.org/data-sites-we-love>
  - A catalog of helpful data sites that relate to health
- W.H.O: <http://www.who.int/en/>
  - The World Health Organization, has a massive amount of data regarding mortality rates, disease burden and similar data sets.
- US Health data: <https://www.data.gov/health/>
  - This site seems to be focused on US health statistics so it might not prove to be helpful for world data, but it seems to be comprehensive for the US
- World Bank: <http://www.worldbank.org/en/who-we-are>
  - A large database that seems to collect a large set of socio-economic metrics.

### Data Processing:

**Data Cleanup:** We will be pulling datasets from at least a couple different sources, most likely the World Bank and the WHO; both of these data sets are incredibly large, and we will need to figure out how best to combine both of these datasets for our purposes. Another problem we've seen is that the WHO data only has data for every five years going back to 2000, while the World Bank has continuous yearly data going back over almost 20 years. We'll have to make these data sets match up for our purposes.

**Data Standardization:** In pulling from multiple datasets there will no doubt need to be some standardization in labeling and with dealing with empty dataset elements, such as where one data set

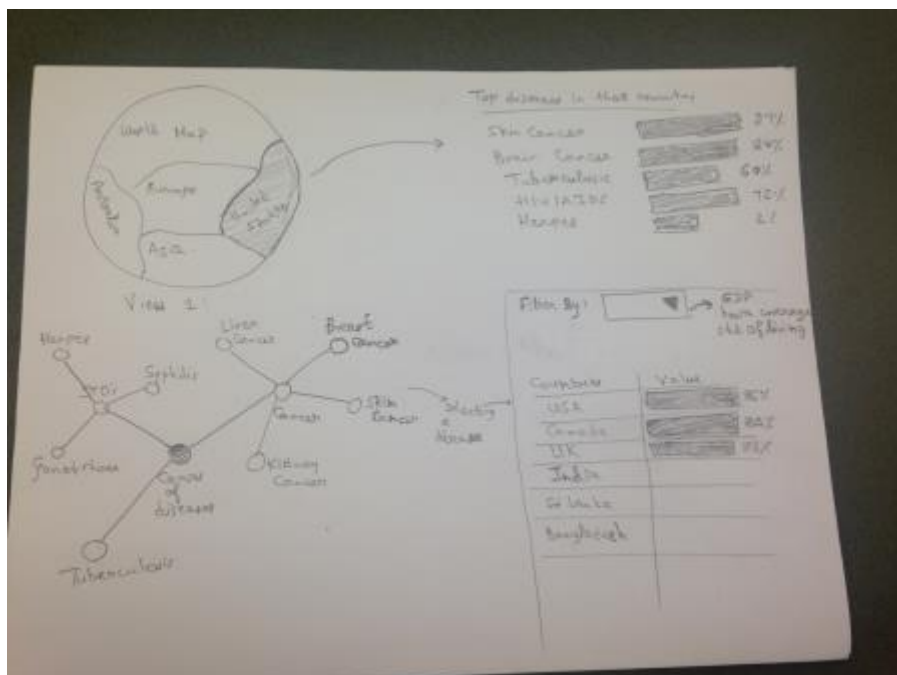


has values for a country or region, and the other one does not.

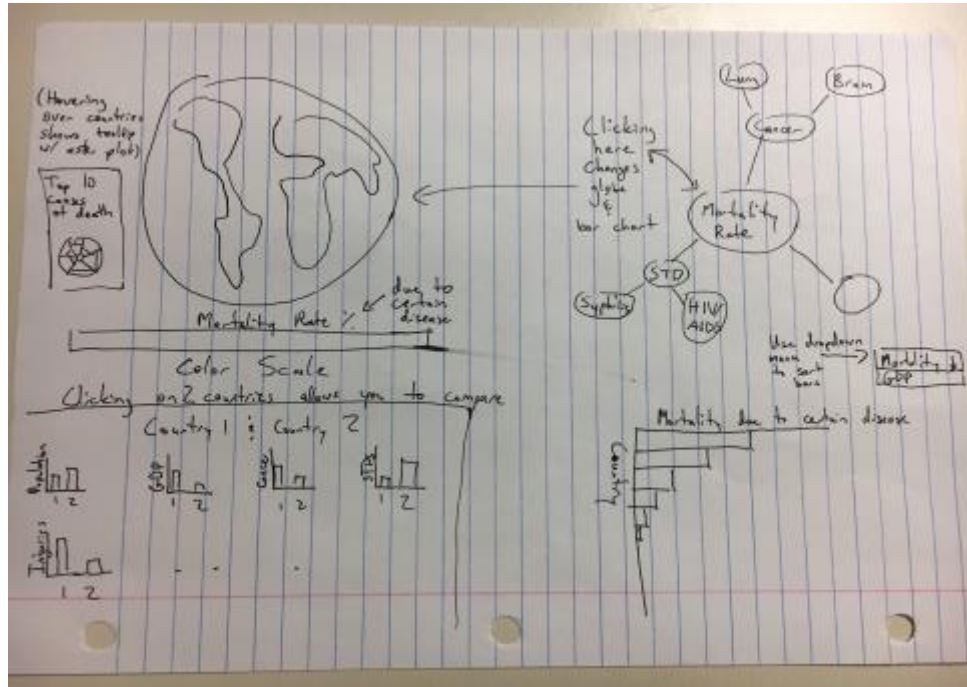
**Implementation:** We will be trying to get all of the data into 2 if not 1 csv file, then us D3 to read and manipulate the data.

## Visualization Design

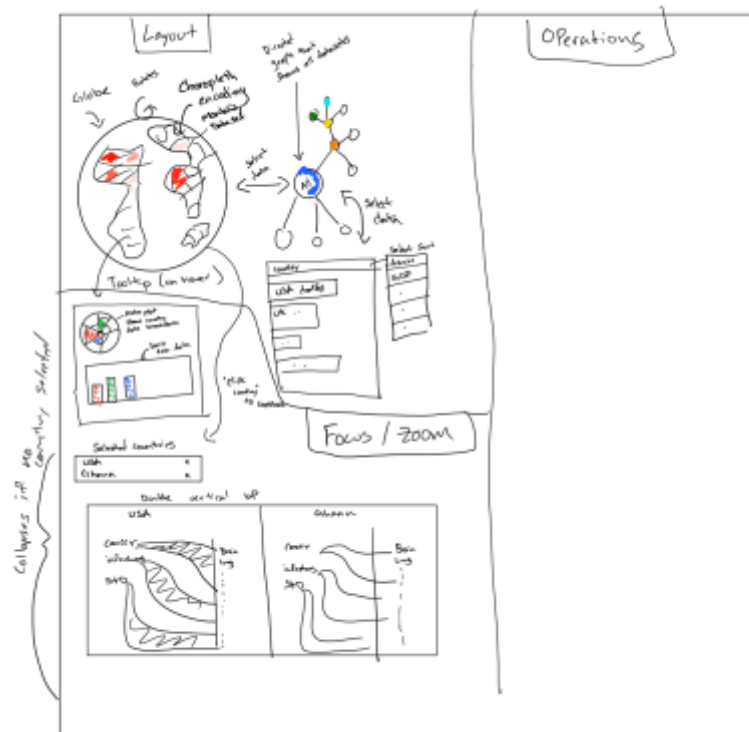
### Brainstorming



Design 1



Design 2



Design 3

## Final Design:

We created our three different designs independently from each other to try and get as creative and diverse designs as possible. What we came back with were three designs that had all of the same essential features. These main features included:

- **Choropleth globe**

- Let the user quickly see what countries are most affected by the affected data set

- **Relationship graph**

- Acts as a way to filter the selection, by a disease or a sub-disease
- Also lets the user see how the data is structured, meaning how diseases are related to each other

- **Bar Chart**

- This helps us overcome one of the most significant setbacks of a choropleth graph, which is seeing specific data. This will be a list of all countries, the length of the bar would be proportional to the data set selected from the relationship graph
- In addition to showing the data selection, the user can also sort the bar charts by different factors, such as gender, age, socio-economic data. This lets the user make meaningful correlations. For example, you could sort by highest - lowest GDP, and quickly see how GDP affects the current data set.

There were some more short features that we liked from different designs; because they helped explain the data even better. However, we thought that there might be more complicated to implement, and the visualization would still be useful without them:

- **Tool Tip**

- When you select a data set, you see a snapshot of how countries are affected by that dataset. However, a user may still be interested in seeing how other factors influence a country. Hovering over a country would produce a tooltip that would have an aster plot, showing the top 'x' diseases for that country. And a bar chart showing socio-economic data.

- **Bar Charts to Compare 2 Countries**

- This helps us to compare two countries affected by a particular disease and hence, draw out conclusions.

- **Compare Data by Year**

- Use brushing or a drop-down menu to view data summed up over different years or for a particular year that the user chooses.

## Must-Have Features:

- **Global Map:** We think that this is an excellent way to grab the user's attention. It is also critical to answering some of our questions about how specific regions are affected by different diseases.
- **Relationship Visualization:** This is a must-have feature because it allows us to filter data based on disease and is a good way to visualize how different diseases are categorized.
- **Bar Chart:** The bar chart will be a good visualization to have since it allows the user to see the data shown on the global map more clearly. Additionally, being able to sort based on socioeconomic factors will enable the user to view interesting and insightful correlations.

## Optional Features:

- **Tooltip:** This helps the user access data more quickly without having to change datasets, it's also an easy way to make more use out of the choropleth globe.
- **Comparison Chart:** This helps answer another question, in how do different countries compare with different diseases & socio-economic factors.
- **Data Sets by Years:** This helps the user make better correlations, but increasing the data available to them.

## Project Schedule

---

- **Week 1**
  - Data aggregated and formatted for use - Matt Willden
  - Globe Prototype finished - Matt Willden
  - Relationship Visualization - Naoki Tominaga
  - Bar Chart - Paarth Lakhani

- **Week 2**

- Start linking data to one or two of our modules

- **Week 3**

- Data has been linking to all modules.
- Make the visualizations better using CSS techniques

- **Week 4**

- Implement interactivity between graphs
- All must-have features are finished
- Continue fixing bugs
- Start working on optional features.

- **Week 5**

- Fine tune all must-have features
- Optional features at least somewhat integrated or figured out
- Finish website