Naoki Tominaga u0876779
Paarth Lakhani u0936913
Matt Willden u0920335
https://willdenms.github.io/WHO_is_sick

Overview and Motivation

Mortality has been steadily falling during the 20th and 21st centuries. We wanted to see what the major causes of mortality are for each country and how socioeconomic factors affects mortality across the world. This kind of data is important to study because it gauges how well a country's health care system works as well as to focus resources to improve the health care systems.

Related Work

- HW3/HW4: The bar chart was inspired by these two assignments. Tools were used from these homeworks to appropriately display the bar chart.
- HW5: The main motivation for the relationship visualization/tree was derived from the playoff visualization from this homework. The table also gave us some inspiration for our bar chart.
- HW6: The choropleth from this assignment gave us a good idea for a design for our project. Although we did want a bit more detailed map for our visualization.

Questions

- What are the top countries affected by a certain disease?
- How are mortality rates affected by different GDP per capita?
- The diseases affecting a country's mortality rates?

Data

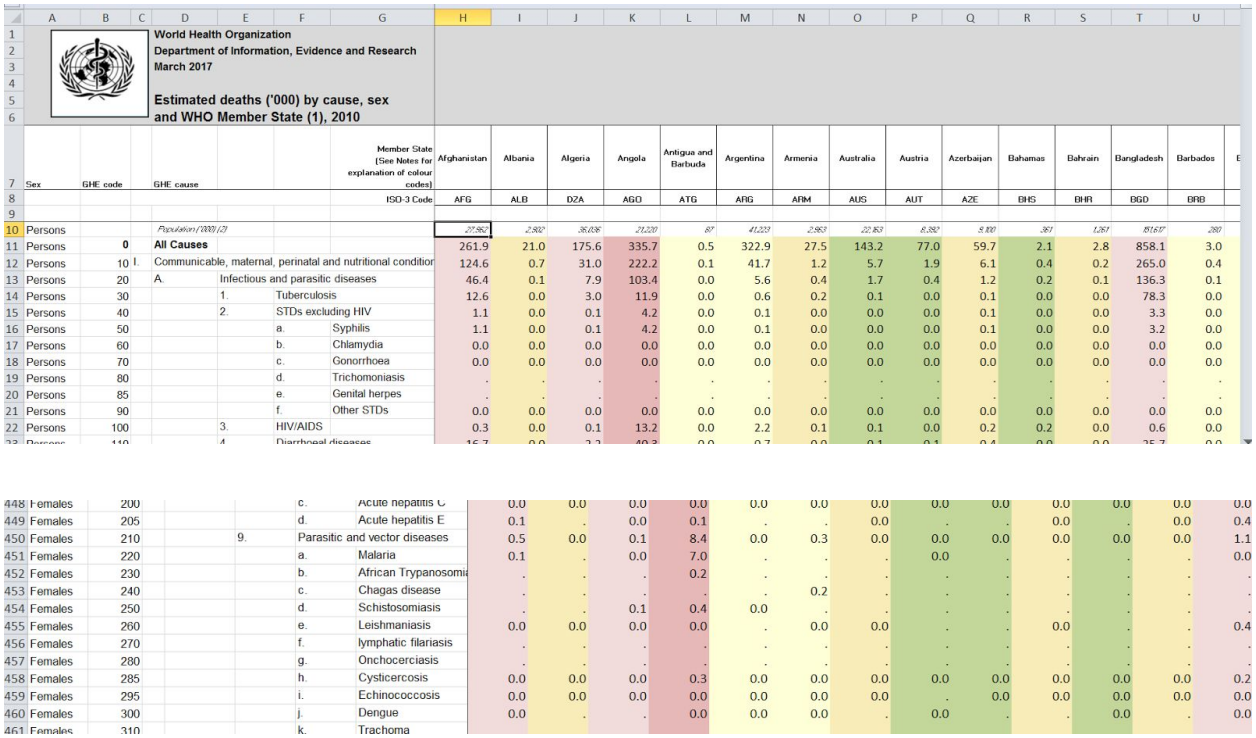
Most of our data was pulled and correlated from the WHO website. In addition to the WHO website, we also used the World Bank website to find population, GDP, and other socio-economic data. Below are the direct links:

- http://www.who.int/en/
- http://www.worldbank.org/en/who-we-are

It took quite a bit of work to get all of the data in the right order. To be able to have the modules we wanted we were required to write some special scripts to extract and combine the data we needed, as well and manually editing and transposing some csv tables.

To give an overview of how we achieved the ideal data structure, I'll list the phases that we went through.

Below is the initial dataset from the WHO website:

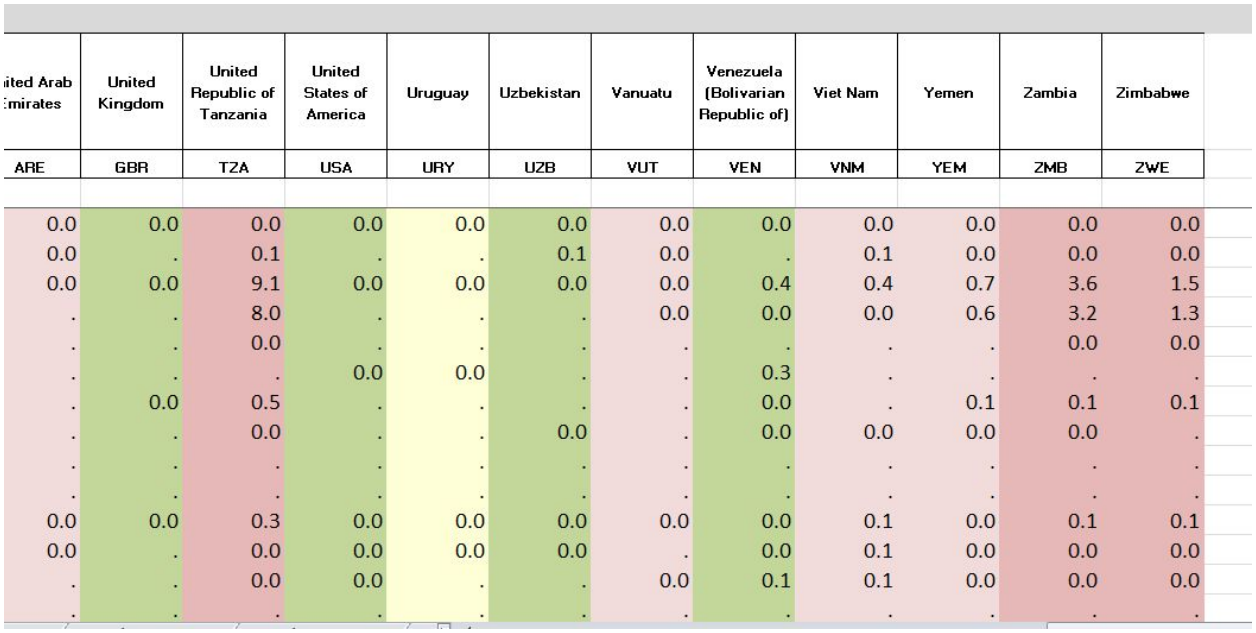


World Health Organization
Department of Information, Evidence and Research
March 2017

Estimated deaths ('000) by cause, sex and WHO Member State (1), 2010

Sex	GHE code	GHE cause	Member State (See Notes for explanation of colour codes)	Afghanistan	Albania	Algeria	Angola	Antigua and Barbuda	Argentina	Armenia	Australia	Austria	Azerbaijan	Bahamas	Bahrain	Bangladesh	Barbados
			ISO-3 Code	AFG	ALB	DZA	AGO	ATG	ARG	ARM	AUS	AUT	AZE	BHS	BHR	BGD	BRB
Persons		Population ('000) (2)		27,962	2,902	36,036	21,220	87	41,223	2,963	22,163	8,382	9,100	361	1,261	151,617	280
Persons	0	All Causes		261.9	21.0	175.6	335.7	0.5	322.9	27.5	143.2	77.0	59.7	2.1	2.8	858.1	3.0
Persons	10	I. Communicable, maternal, perinatal and nutritional conditior		124.6	0.7	31.0	222.2	0.1	41.7	1.2	5.7	1.9	6.1	0.4	0.2	265.0	0.4
Persons	20	A. Infectious and parasitic diseases		46.4	0.1	7.9	103.4	0.0	5.6	0.4	1.7	0.4	1.2	0.2	0.1	136.3	0.1
Persons	30	1. Tuberculosis		12.6	0.0	3.0	11.9	0.0	0.6	0.2	0.1	0.0	0.1	0.0	0.0	78.3	0.0
Persons	40	2. STDs excluding HIV		1.1	0.0	0.1	4.2	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	3.3	0.0
Persons	50	a. Syphilis		1.1	0.0	0.1	4.2	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	3.2	0.0
Persons	60	b. Chlamydia		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Persons	70	c. Gonorrhoea		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Persons	80	d. Trichomoniasis		.	.	.	.	.	.	.	.	.	.	.	.	.	.
Persons	85	e. Genital herpes		.	.	.	.	.	.	.	.	.	.	.	.	.	.
Persons	90	f. Other STDs		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Persons	100	3. HIV/AIDS		0.3	0.0	0.1	13.2	0.0	2.2	0.1	0.1	0.0	0.2	0.2	0.0	0.6	0.0
Persons	110	4. Diarrhoeal diseases		16.7	0.0	2.2	49.3	0.0	0.7	0.0	0.1	0.1	0.4	0.0	0.0	25.7	0.0

Females	200			c.	Acute hepatitis C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Females	205			d.	Acute hepatitis E	0.1	.	0.0	0.1	.	.	0.0	.	.	0.0	.	0.0	0.4
Females	210		9.	Parasitic and vector diseases		0.5	0.0	0.1	8.4	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	1.1
Females	220			a.	Malaria	0.1	.	0.0	7.0	.	.	.	0.0	.	.	.	.	0.0
Females	230			b.	African Trypanosomia	.	.	.	0.2	.	.	.	.	.	.	.	.	.
Females	240			c.	Chagas disease	.	.	.	.	.	0.2	.	.	.	.	.	.	.
Females	250			d.	Schistosomiasis	.	.	0.1	0.4	0.0	.	.	.	.	.	.	.	.
Females	260			e.	Leishmaniasis	0.0	0.0	0.0	0.0	.	0.0	0.0	.	.	0.0	.	.	0.4
Females	270			f.	lymphatic filariasis	.	.	.	.	.	.	.	.	.	.	.	.	.
Females	280			g.	Onchocerciasis	.	.	.	.	.	.	.	.	.	.	.	.	.
Females	285			h.	Cysticercosis	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
Females	295			i.	Echinococcosis	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.	0.0	0.0	0.0	0.0	0.0
Females	300			j.	Dengue	0.0	.	.	0.0	0.0	0.0	.	0.0	.	.	0.0	.	0.0
Females	310			k.	Trachoma													



ited Arab Emirates	United Kingdom	United Republic of Tanzania	United States of America	Uruguay	Uzbekistan	Vanuatu	Venezuela (Bolivarian Republic of)	Viet Nam	Yemen	Zambia	Zimbabwe
ARE	GBR	TZA	USA	URY	UZB	VUT	VEN	VNM	YEM	ZMB	ZWE
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	.	0.1	.	.	0.1	0.0	.	0.1	0.0	0.0	0.0
0.0	0.0	9.1	0.0	0.0	0.0	0.0	0.4	0.4	0.7	3.6	1.5
.	.	8.0	.	.	.	0.0	0.0	0.0	0.6	3.2	1.3
.	.	0.0	.	.	.	.	.	.	.	0.0	0.0
.	.	.	0.0	0.0	.	.	0.3	.	.	.	.
.	0.0	0.5	.	.	.	.	0.0	.	0.1	0.1	0.1
.	.	0.0	.	.	0.0	.	0.0	0.0	0.0	0.0	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.1
0.0	.	0.0	0.0	0.0	0.0	.	0.0	0.1	0.0	0.0	0.0
.	.	0.0	0.0	.	.	0.0	0.1	0.1	0.0	0.0	0.0
.	.	.	.	.	.	.	.	.	.	.	.

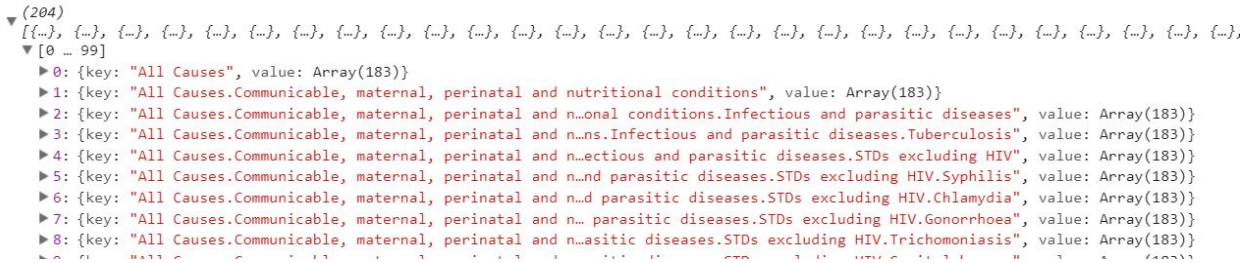
As you can see in the above screenshots, there are more than 460 rows and more than 100 columns. After spending considerable time cleaning up the data we produced the following CSV file which is much more cleaned.

	A	B	C	D	E	F	G	H
	Disease Name	AFG	ALB	DZA	AGO	ATG	ARG	ARM
2	All Causes	260	21.8	195.9	341.3	0.6	332.6	
3	All Causes.Communicable, maternal, perinatal and nutritional conditions	102.5	0.8	31.5	210.9	0.1	40.9	
4	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases	37.4	0.1	7.8	94.4	0	5.7	
5	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.Tu	13.3	0	3.1	11	0	0.7	
6	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.ST	0.9	0	0.1	4.9	0	0.1	
7	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.ST	0.9	0	0.1	4.9	0	0.1	
8	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.ST	0	0	0	0	.	0	
9	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.ST	0	0	0	0	.	0	
10	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.ST	.	.	.	.	.	.	.
11	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.ST	.	.	.	.	.	.	.
12	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.ST	0	0	0	0	.	0	
13	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.HI	0.3	0	0.1	12.4	0	2.3	
14	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.Di	12.1	0	2.1	33.2	0	0.6	
15	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.Ch	4.4	0	0.3	3.5	0	0.1	
16	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.Ch	1.4	0	0.2	0.9	.	0.1	
17	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.Ch	0.1	0	0	0	0	0	
18	All Causes.Communicable, maternal, perinatal and nutritional conditions.Infectious and parasitic diseases.Ch	2	0	0.1	0.4	.	0	

There is another CSV that lists the population and GDP of all the countries:

	A	B	C	D	E
1	Country N	Country C	Series Nar	2015 [YR2015]	
2	Afghanista	AFG	GDP (curre	1.92E+10	
3	Afghanista	AFG	GDP grow	1.519911	
4	Afghanista	AFG	GDP per ca	590.2695	
5	Afghanista	AFG	Surface ar	652860	
6	Albania	ALB	GDP (curre	1.15E+10	
7	Albania	ALB	GDP grow	2.56	
8	Albania	ALB	GDP per ca	3965.017	
9	Albania	ALB	Population	2889167	
10	Albania	ALB	Surface ar	28750	
11	Algeria	DZA	GDP (curre	1.67E+11	
12	Algeria	DZA	GDP grow	3.9	
13	Algeria	DZA	GDP per ca	4206.031	
14	Algeria	DZA	Population	39666519	
15	Algeria	DZA	Surface ar	2381740	
16	American	ASM	GDP (curre	..	
17	American	ASM	GDP grow	..	
18	American	ASM	GDP per ca	..	
19	American	ASM	Population	55538	
20	American	ASM	Surface ar	200	
21	Andorra	ADO	GDP (curre	..	
22	Andorra	ADO	GDP grow	..	
23	Andorra	ADO	GDP per ca	..	

So, in order to efficiently use two CSV files in one visualization, it is good to bind them together in some data structure to use in our visualization. Thus, we converted the two CSV files into an array of JSON objects that has the following structure:



It is an array of JSON objects where each JSON object has the disease name as the key and the value is an array of country objects where each country object has four fields: GDP, code, mortality, name.

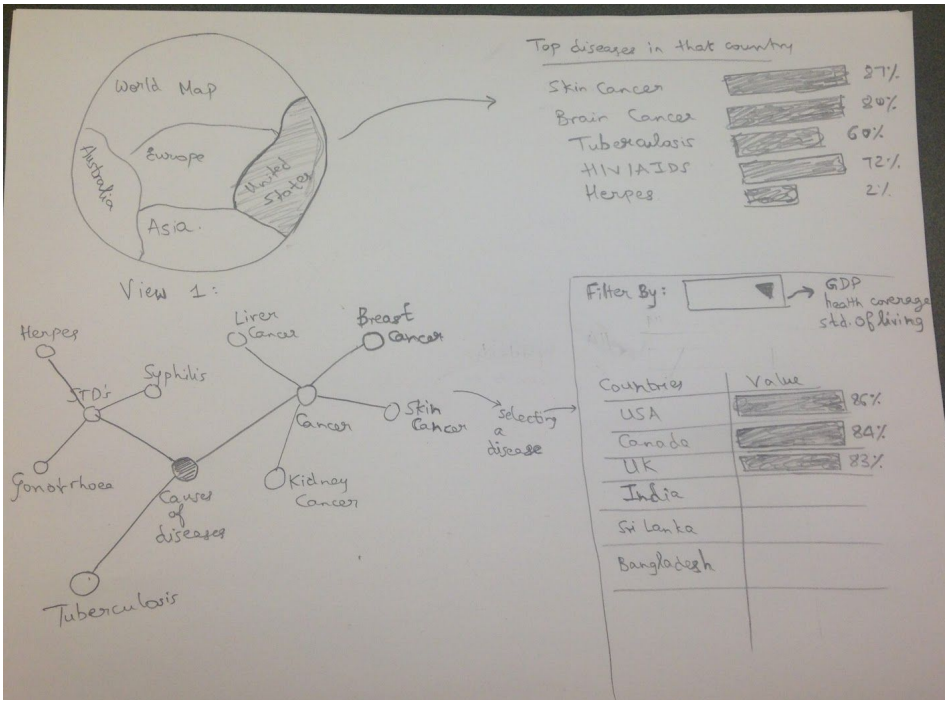
Thus, it was tricky to structure the data effectively.

Exploratory Data Analysis

We did not use any visualizations when we first discovered our data set. We only looked at the data through excel. It didn't occur to us that we should try to visualize the data. The sheer volume of the data set is all the insight we gained from it. In retrospect this was probably not the best way to determine a data set to use but our reasoning was that we didn't want to be hampered by a small data set.

Design Evolution

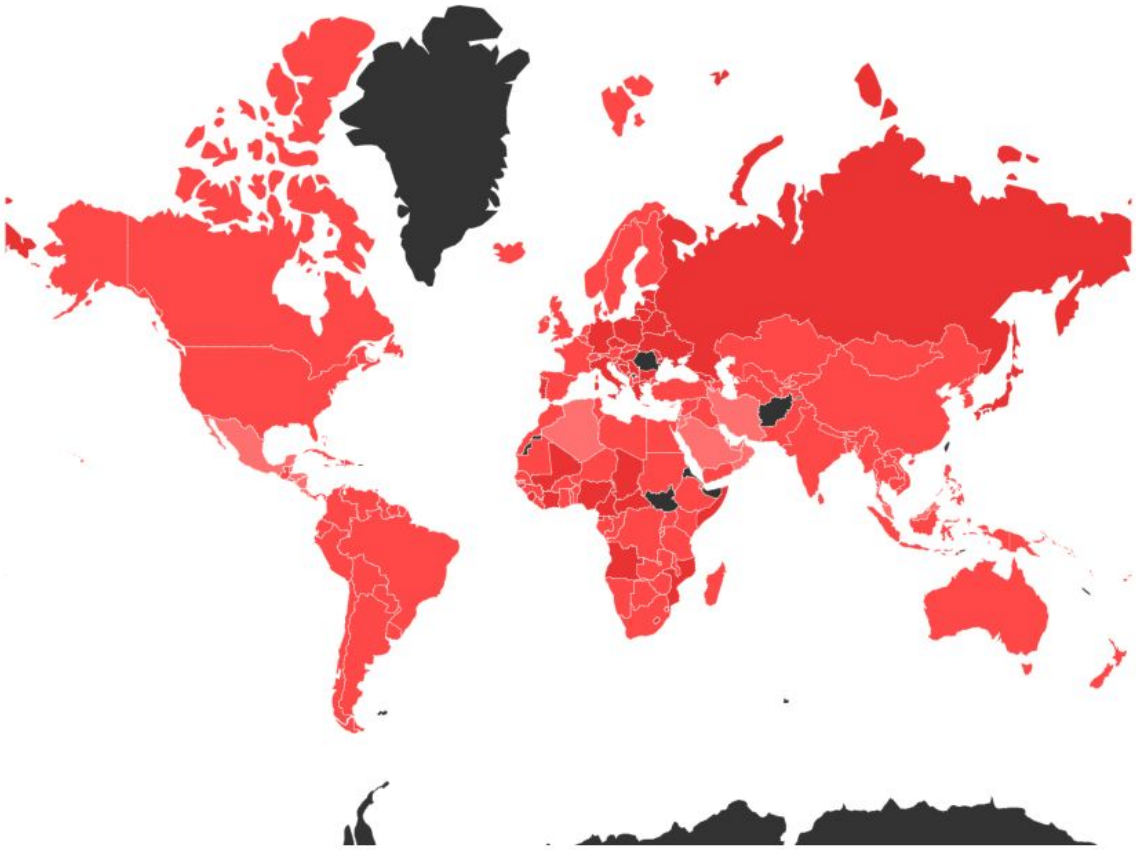
Our original design:



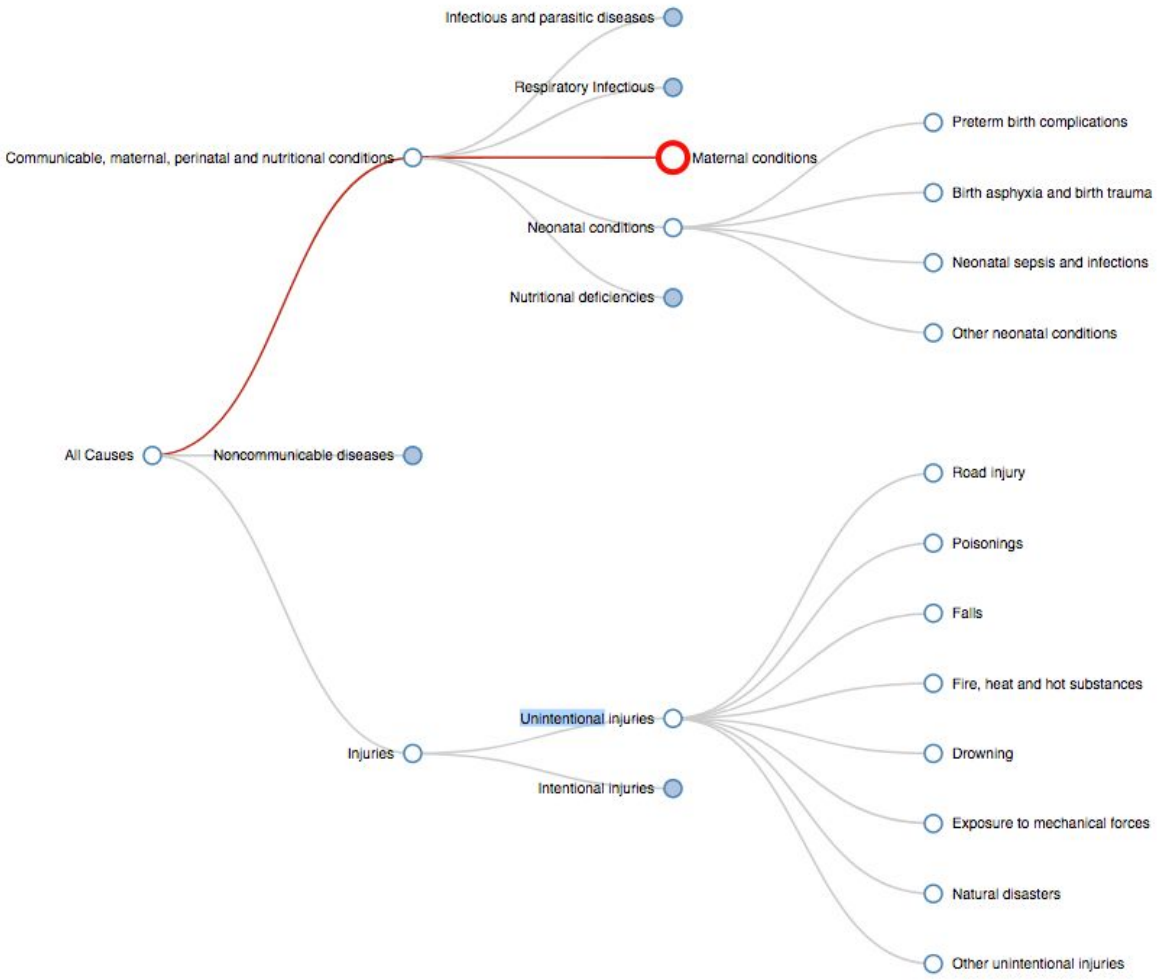
- **Choropleth globe**
    - Let the user quickly see what countries are most affected by the affected data set
- **Relationship graph**
    - Acts as a way to filter the selection, by a disease or a sub disease
    - Also lets the user see how the data is structured, meaning how diseases are related to each other

- **Bar Chart**
    - This helps us overcome one of the biggest 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 the GDP of all the countries to see how the current data set is affecting all the countries.

Our Final Design:



- **Choropleth map**:
    - We were unable to implement a globe. So we did a choropleth with a map projection instead. Hovering over the choropleth shows tooltips for the country you are hovering over. Inside the tooltip it shows GDP per capita, mortality due to the selected disease, population, etc.



- **Relationship Graph:**
    - The relationship graph originally was designed so the root of the tree was at the center. This caused a tree that was too cluttered and messy so it was changed to a tree that had a root on the left and expanded to the right.
    - Minor features were added to make usability easier. These included highlighting nodes when they are selected and path highlighting to show where the selected node is.

- **Bar Chart:**
    - The bar chart is a visualization that lists the mortality rates of the disease selected from the relationship graph. The bar chart lists all the countries with the highest to lowest mortality rates. The rows can be sorted according to the GDP that is displayed in the third column.

Implementation

It took far longer to for us to get our data into a manageable state for us to really get our views working. In addition to data mangling there was a significant setback with the choropleth. We originally wanted to have a rotating globe that the user would be able to rotate and see, this however proved impractical for a few reasons. The foremost being that we did not have enough time to port the version 3 D3 code into a version 4 environment. We fell back to a Mercator

projection, which we believe helps the user quickly see what areas have the highest mortality rate by the chosen disease.

Evaluation

We used 2015 dataset that lists the mortality rates of different diseases for all countries. It was hard to draw meaningful conclusions just by looking at the numbers in the data set. Soon after we put the data into visualizations, we drew out worthwhile conclusions. Some worthwhile conclusions are:

1. India, China, United States, Russia and Indonesia are 5 countries with highest mortality rates
2. In general, the GDP is inversely proportional to the mortality rates of the country
3. From the visualization, we observe that in 2015, non-communicable diseases resulted into ~17million deaths, Communicable, maternal, perinatal and nutritional conditions resulted into ~4.5million deaths and injuries resulted into ~2 million deaths. All the deaths are of the top 3 countries listed in the table.

There are couple of improvements:

1. Ability to sort the bar chart based on the mortality rates.
2. Ability to compare two countries and the causes affecting their mortality rates.
3. Add an aster view in the tooltip when user hovers on. This view would list top 10 diseases of that country.
4. Add the ability to see how mortality rates change over the years.