

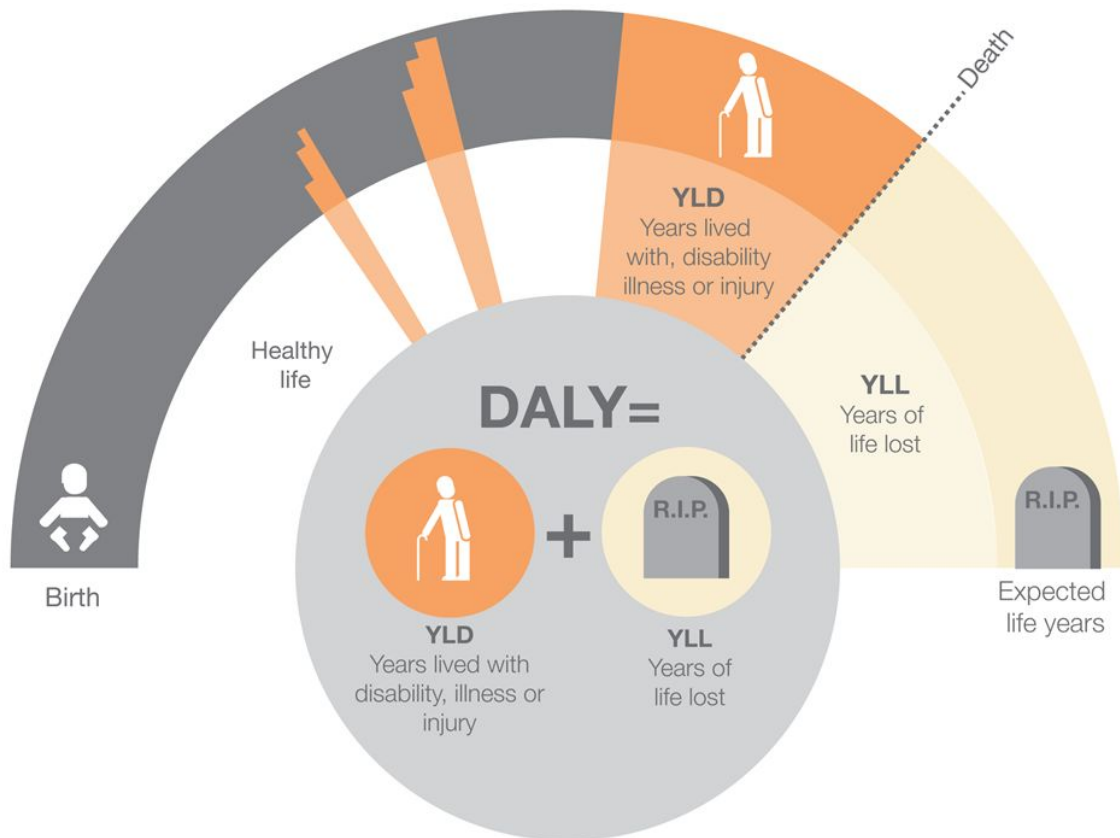
Disability-Adjusted Life Years (DALY)

From Around the World

Zhejing Luo, Jonathon O'Reilly (Group 3)

Project site: <http://oreillyjw.github.io/cs171finalproject/index.html>

Screencast: <https://www.youtube.com/watch?v=H6qm4MrxwDo>



[1]

Motivation & Goal

To measure the burden of disease, the World Health Organization (WHO) calculates a metric called disability-adjusted life year (DALY). DALY is expressed as the sum of the number of the years lived with disability (YLD) and the number of years lost (YLL). DALY is a highly valuable metric for understanding public health status, it represents "...the gap between current health status and an ideal health situation"[1].

It needs be noted that the word “disability” here does not only refer to the physical or mental handicap like blindness that people commonly think, but a much wider range of causes that negatively impacts the people’s quality of life, including sickness and injury.

Our project aims to learn how diseases and disabilities have impacted the world by creating effective DALY data visualization. We attempt to answer these questions with our visualization:

- What causes are included in DALY? What are the categories of the causes?
- Which regions have the highest/lowest DALY?
- What are some of the top DALY causes?
- How does various DALY causes impact people during different life stages?

Data

The DALY estimates of year 2000 and 2012 are published on the WHO official Global Health Observatory (GHO) data repository (<http://apps.who.int/gho/data/node.main.DALY?lang=en>). DALY data are available in these dimensions: cause, age, sex, region, country, income group.

We extracted four levels of DALY causes from the raw data in a hierarchy. The DALY are divided into 3 main cause categories (Level I), they are communicable diseases & other group I causes, non-communicable diseases, and injury. Each main cause category includes multiple subcategories (Level II). For example, malignant neoplasms and neurological conditions are both subcategories under the non-communicable diseases main category. Some of the sub-categories are divided into finer categories or specific causes, like road injury is a specific cause under the unintentional injury subcategory, which belongs to the injury main category. For simplification, we are only using the top three levels in the visualizations.

Causes listed in a hierarchy

1	I.	Communicable, maternal, perinatal and nutritional conditions	959,921,355
2	A.	Infectious and parasitic diseases	452,171,844
3		1. Tuberculosis	42,239,954
4		2. STDs excluding HIV	10,767,089
5		a. Syphilis	7,904,033
6		b. Chlamydia	1,416,433
7		c. Gonorrhoea	541,987
8		d. Trichomoniasis	167,265
9		e. Other STDs	737,371
10		3. HIV/AIDS	95,226,413
11		4. Diarrhoeal diseases	118,788,857
12		5. Childhood-cluster diseases	29,335,172
13		a. Whooping cough	8,108,644
14		b. Diphtheria	220,017
15		c. Measles	14,777,542
16		d. Tetanus	6,228,969
17		6. Meningitis	30,864,495
18		7. Encephalitis	6,485,725
19		8. Acute hepatitis B	6,242,734

There are two types of DALY datasets. Type 1 is available on the country level but only has DALY data for main cause categories (Level I) including Communicable Diseases & Other Group I, Non-communicable Diseases, and Injuries. Type 2 has DALY data for causes of all main and sub-categories (Level I, II, III, IV), however it is only available on region/income group level which was aggregated from the country level data.

To work with this situation, we created two groups of visualizations and put them on separate pages, each only using with one type of the dataset. Please see details in the sections below.

Type I dataset example

Country	Year	All Causes	Communicable & other Group I	Noncommunicable diseases	Injuries
Afghanistan	2012	68,970	23,259	35,454	10,256
Afghanistan	2000	86,566	37,561	37,894	11,111
Albania	2012	29,903	3,297	23,098	3,508
Albania	2000	36,919	5,068	27,014	4,837
Algeria	2012	34,790	5,963	25,650	3,177
Algeria	2000	38,868	8,253	26,537	4,078
Angola	2012	93,709	52,615	32,236	8,858
Angola	2000	115,496	76,145	30,594	8,757
Argentina	2012	26,808	3,706	20,093	3,009
Argentina	2000	29,479	4,348	21,592	3,540
Armenia	2012	35,083	3,719	28,090	3,274

Type 2 dataset example

		2012	2012	2012	2012	2012	2012
		Both sexes	Female	Female	Female	Female	Female
Region	Causes	All ages (total)	0-27 days	1-59 months	5-14 years	15-29 years	30-49 years
High-income	All Causes	388,667,723	2,140,761	2,531,291	3,636,272	4,137,743	677,133
High-income	Communicable	26,468,870	1,470,351	965,222	498,702	1,055,234	2,132,445
High-income	Infectious and	11,675,618	35,122	386,704	137,918	439,019	1,260,271
High-income	Tuberculosis	1,538,846	246	1,665	1,691	59,206	141,836
High-income	STDs excluding	438,813	2,188	1,145	4,531	100,363	125,197
High-income	Syphilis	28,537	2,188	1,145	316	606	1,905
High-income	Chlamydia	162,601	-	-	1,042	26,031	22,398
High-income	Gonorrhoea	87,647	-	-	677	16,187	20,678
High-income	Trichomoniasis	27,628	-	-	566	15,727	9,955
High-income	Other STDs	132,400	-	-	1,930	41,813	70,261
High-income	HIV/AIDS	3,861,568	207	22,934	5,748	135,950	666,413
High-income	Diarrhoeal dise	1,767,199	3,100	220,970	46,619	58,823	118,444
High-income	Childhood-clus	58,793	2,266	20,260	1,061	300	398
High-income	Whooping cou	40,338	1,148	15,165	140	43	115
High-income	Diphtheria	1,140	2	185	179	28	1
High-income	Measles	12,451	-	4,900	691	190	162
High-income	Tetanus	4,864	1,117	10	51	39	120
High-income	Meningitis	394,541	2,289	31,893	11,889	17,981	27,434

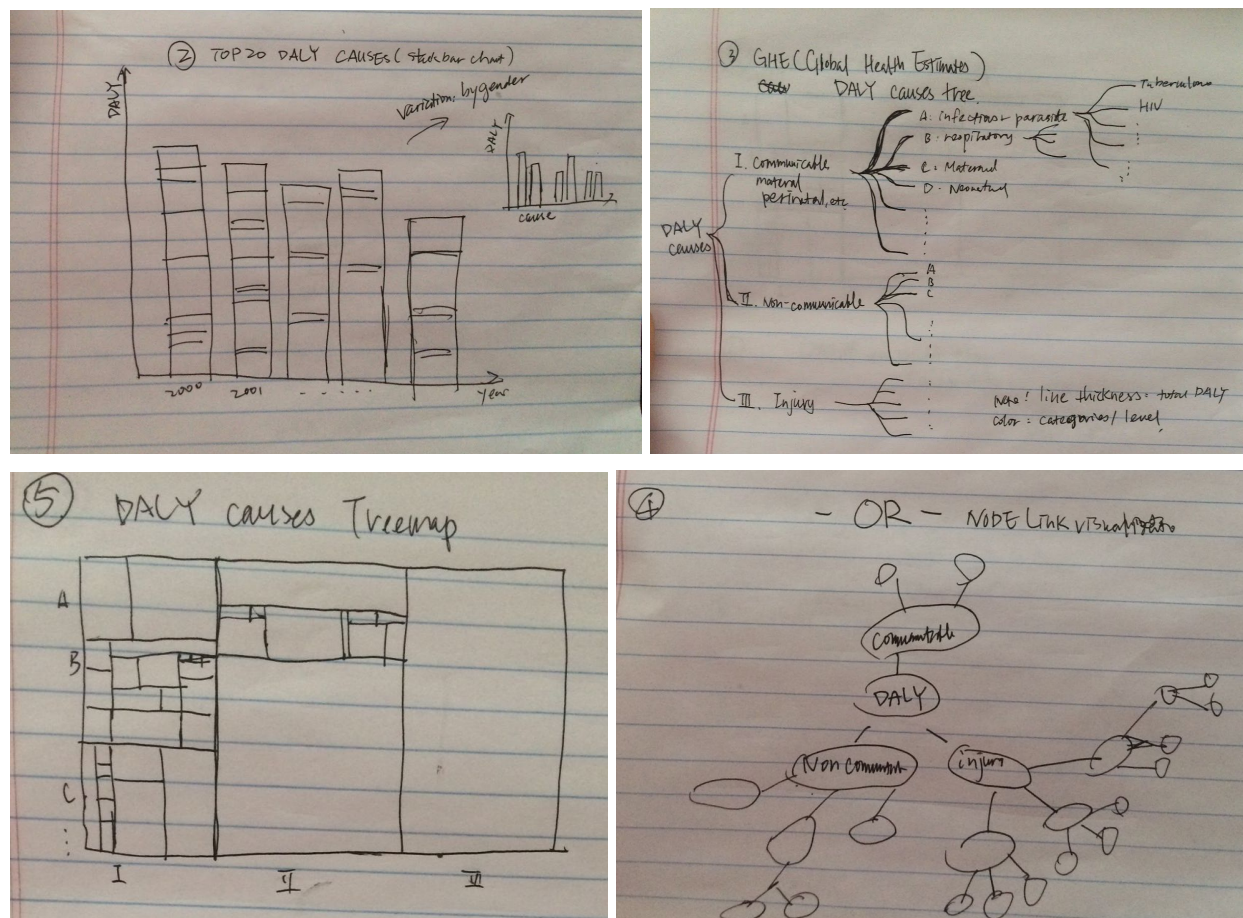
Outside of DALY data, we also used the Life Expectancy data

(<http://apps.who.int/gho/data/node.main.687?lang=en>) and Global Health Expenditure data

(<http://www.who.int/health-accounts/ghed/en/>) from WHO in our visualizations.

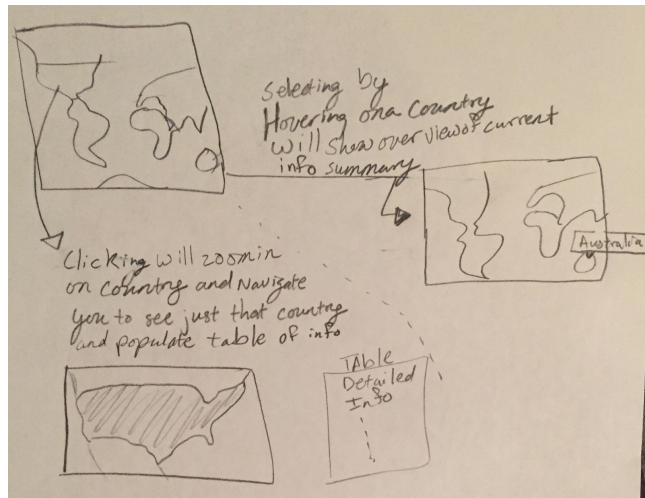
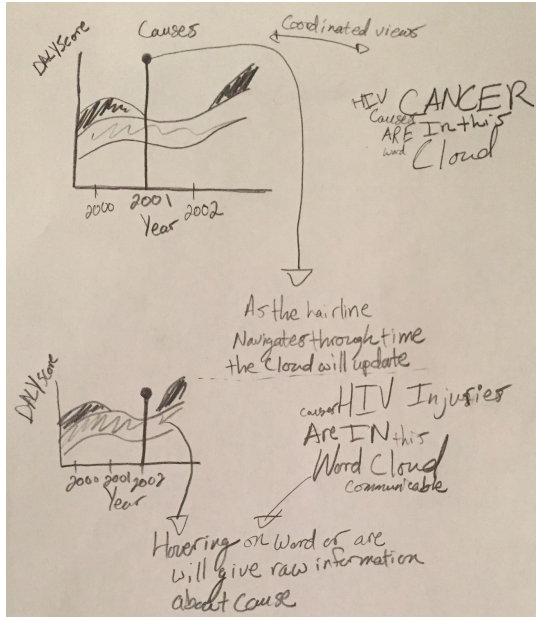
Design Evolution

Our design has evolved a lot in the iterations. When we first started this project, we proposed relatively simple and common visualizations like bar charts and cluster trees:

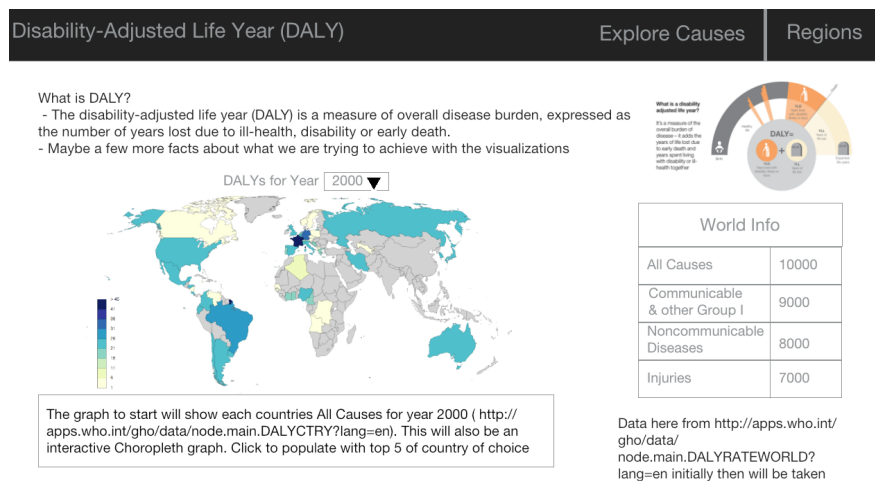


We were assigned text visualization exploration during the “Visualization Zoo” class and the lab session. The visualizations we had seen during the exploration inspired us to consider more possibilities. We compared our answers and started to look into less common and more complicated designs.

Interactivity were also given more consideration at this point. Apart from sketches for individual visualizations, we also created storyboards. Examples include:



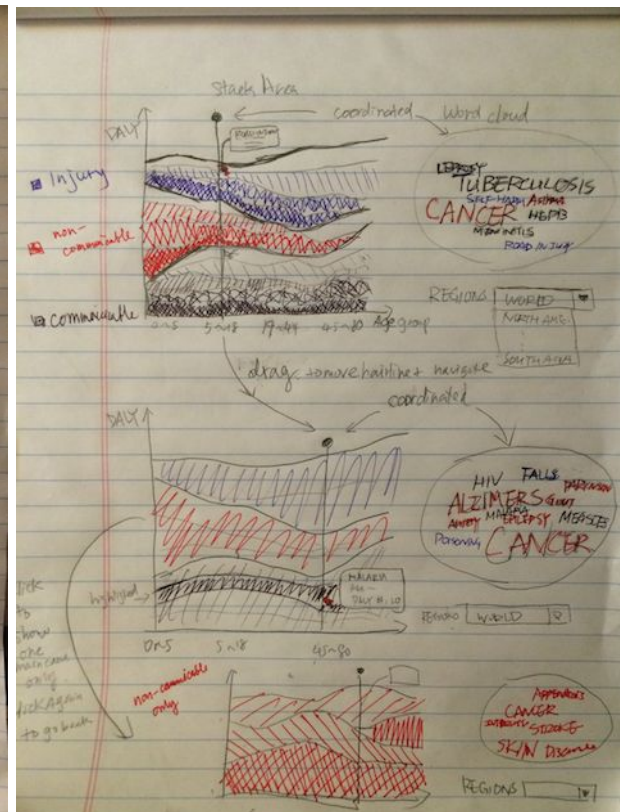
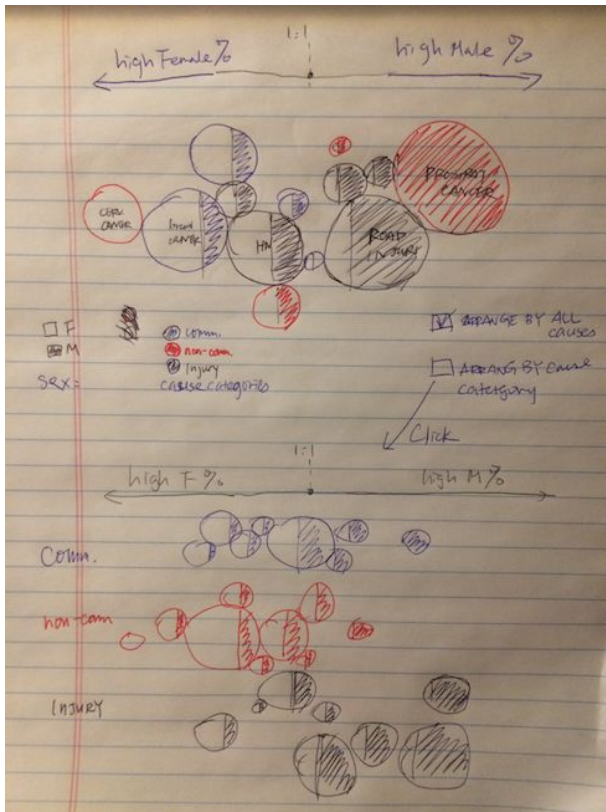
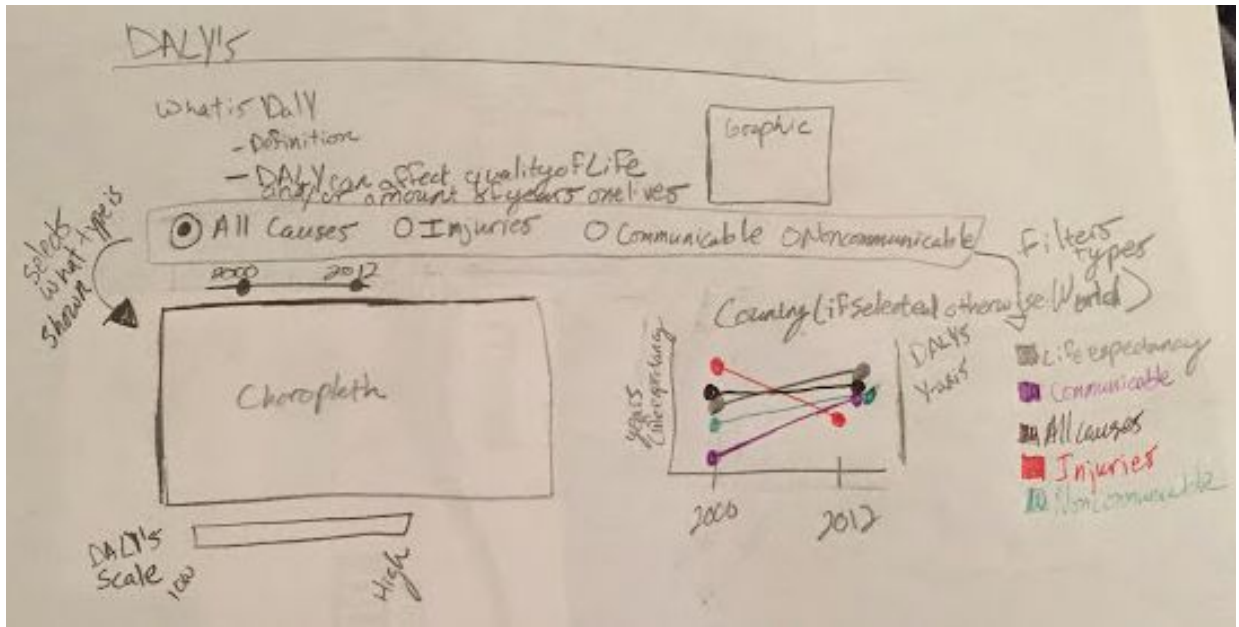
We also created some concept project website layouts:

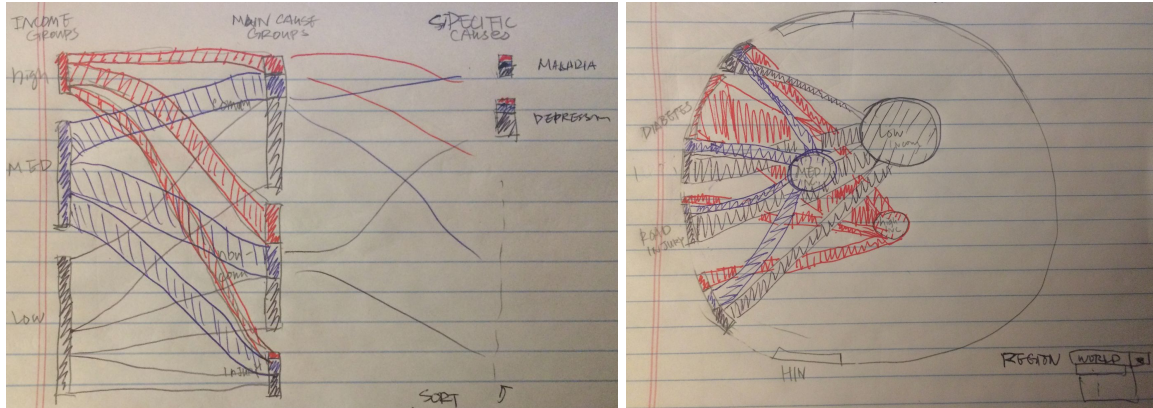


We came up with the below detailed feature list of the visualizations we were considering to implement.

- Choropleth Graph
 - Initially this would be used to show the overall outlook of the world and all countries as a whole
 - Clicking on a country will zoom into the country you want to look at, in addition to loading information about that country in a table
 - Clicking a second time on the country will zoom the user out or we will provide zoom in and out buttons
 - In addition to this, hovering over a country will provide a summary amount of information
- Cluster Tree
 - Select and Navigate a given node to which a user selects
 - This will expand and show more details of a given cluster tree section
 - Hovering on a given node will give more detail about a specific year
- Stacked Area chart
 - A hairline will give the ability to go through time which will then update both the Word Cloud and Bubble Chart
 - Hover on the area to get more details of a cause
 - Clicking on a given cause will only show that cause
 - Will need to determine exactly how the Word Cloud changes for this
 - But Bubbles would automatically change to represent the selected area
- Line Graph (Optional)
 - This will have a corresponding hairline to the Stacked Area Chart
 - Hover on a point to get more information about the Expenditure
- Word Cloud
 - This will display all causes the largest being the cause that has the largest DALY
 - Clicking on a word will change the Bubble Chart
 - Ability to invert the causes so the smallest DALY will now be the largest (optional)
 - Hover on the word to get more details of a cause
- Bubble Chart
 - This will show how the top effect by default from the Word Cloud is distributed for all ages in a given region and then split with gender
 - If you select a word in the word cloud this will automatically change to that word
 - Take the average of both sexes(total of both sexes), double it then for each age range or bubble show how proportionately men and women are affected.

After the evaluation session with other groups and the TF, we decided to come up with more changes and think twice about the innovation design based on the feedback. Examples include:



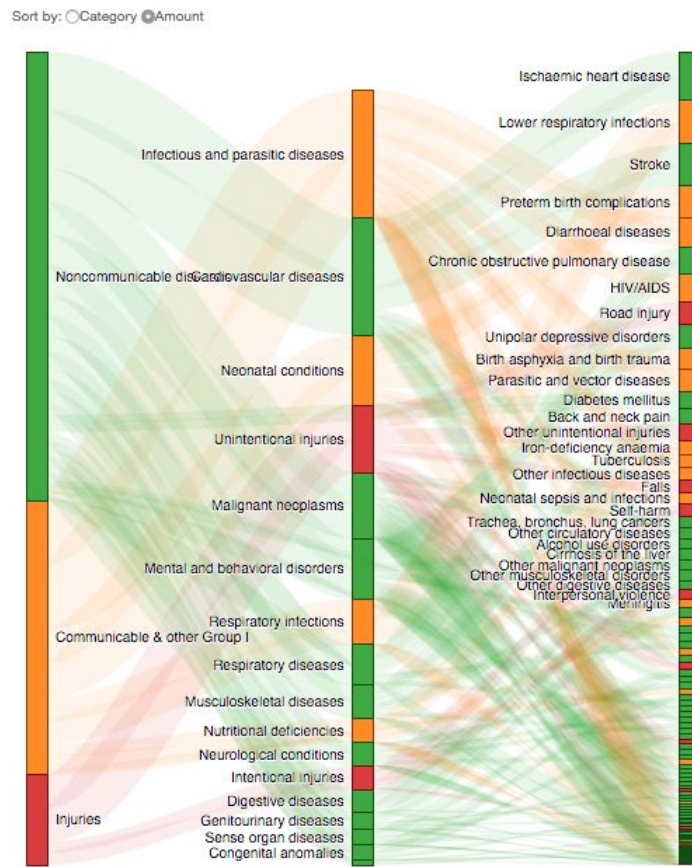


We realized that we do not have enough time to bring all our visions to life when we started implementing prototype I. We built two visualizations first:

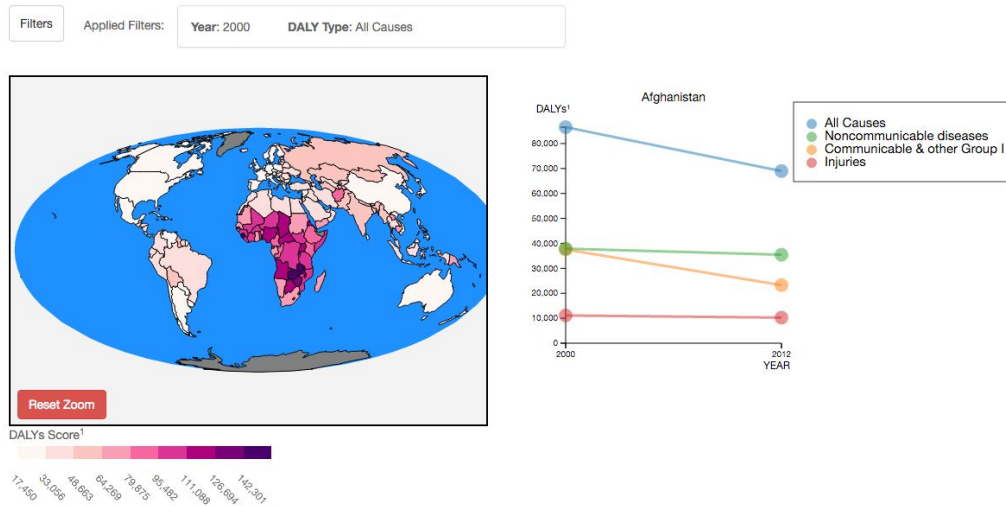
- Choropleth & multiple axis line chart for the country level DALY data
- Sankey chart for the more specific DALY causes and DALY cause hierarchy.

We finished the rough draft of these two visualizations during the Prototype v1 week.

Sankey Chart for specific DALY causes (v1)



Choropleth & Line Chart for country level DALY (v1)



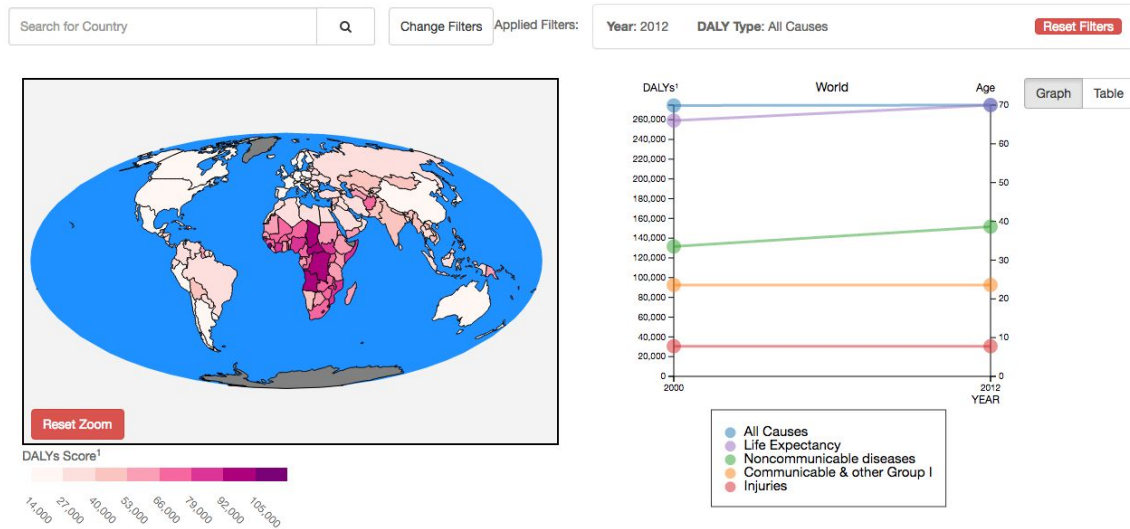
During the Prototype V2 & final touches week, we improved the above two visualizations, and added two more visualizations.

- Scatterplot Matrix chart on the country page
- Stacked area chart on the region/income groups page

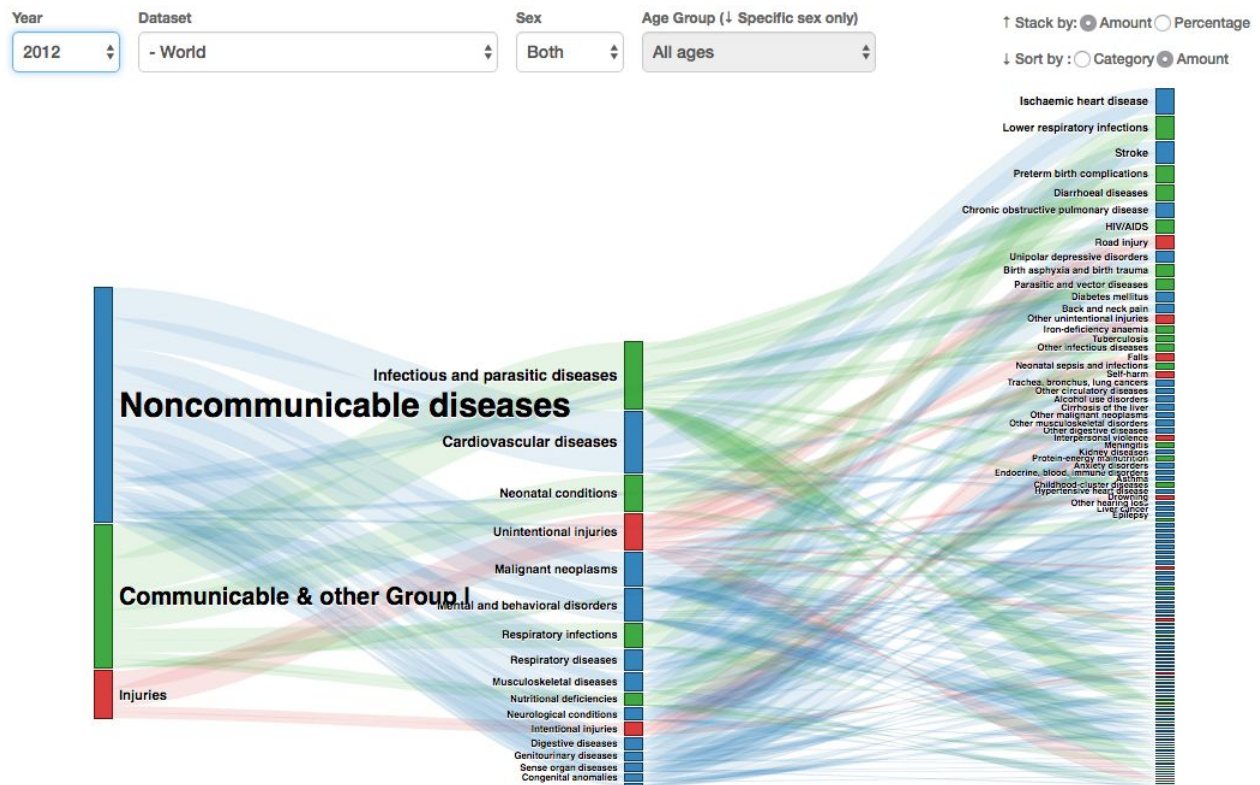
We introduced an additional Y-axis for the life expectancy in the line chart next to the choropleth, and moved the components around to make the design more visually appealing. For the Sankey chart, we matched the node/link colors with the stacked area chart on the same page, and used the font size together with node size to encode the DALY amount of each cause.

All four visualizations were made to be interactive, allowing users to use filters and select the part of the data they are interested in, and see more details for specific data points.

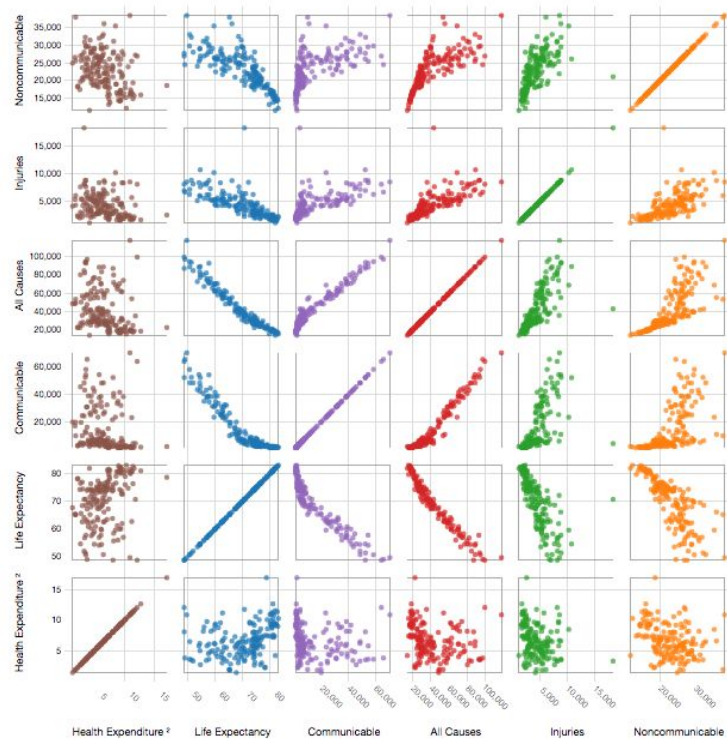
Choropleth & Line Chart for country level DALY (Final)



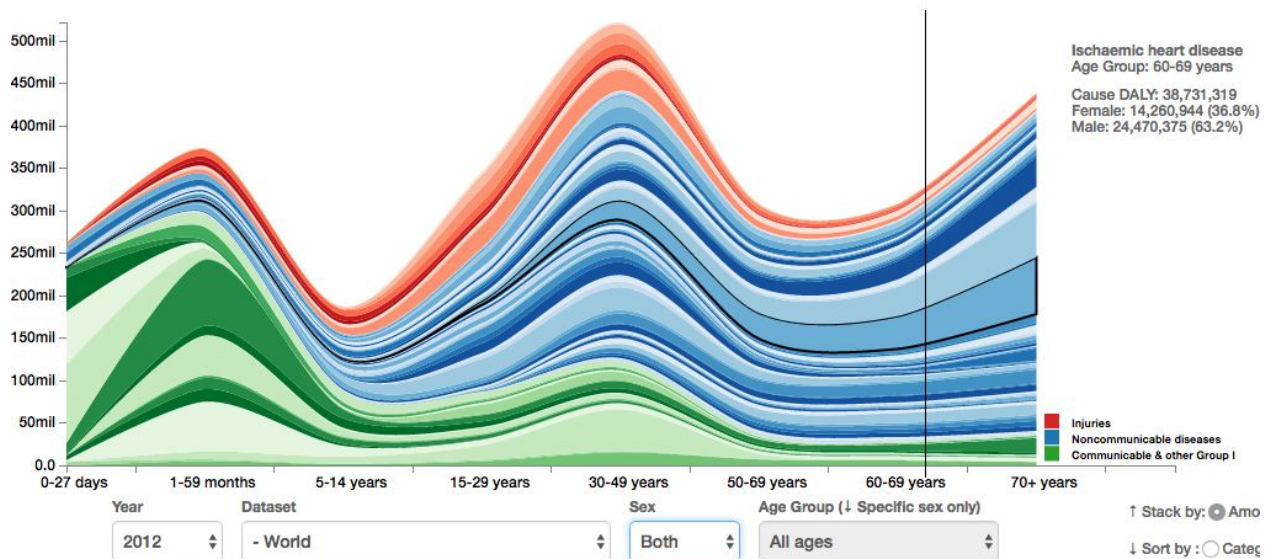
Sankey Chart for specific DALY causes (Final)



Scatterplot Matrix Chart for country level DALY (Final)



Stacked Area chart by specific causes on region/income level (Final)



Project Site Features

- **Country page**
 - Filters/Sorting
 - Years (2000/2012)

- Only show one year data at a time in scatterplot matrix and choropleth
 - Info Types:
 - All Cause (DALY)
 - Communicable & Other Group I (DALY)
 - Injuries (DALY)
 - Noncommunicable Diseases (DALY)
 - Life Expectancy
 - Remove Country/Countries
 - Add a Country to remove from Choropleth and Scatterplot matrix
 - Graphs
 - Choropleth
 - Click on country to zoom in
 - Click on same country selected, or reset zoom to zoom out
 - Line Graph
 - Shows data depending on what filter is selected
 - Toggle to show table of raw data for given country.
 - Scatter Plot Matrix
 - There is a switch
 - Selecting an area and brushing which will brush all graphs in matrix
 - Selecting a point in any graph will bring all of its data to the line graph, zoom into the country's location in the choropleth, and will highlight the country in the graphs of the scatterplot matrix.
 - Other
 - Search for a Country and then once you find an available country, choropleth will zoom on the country you selected, Line Graph will change to the country you select, and scatter plot will highlight the country searched on
- **Region/Age Group page**
 - Filters/Sorting
 - Years (2000/2012)
 - Only show one year data at a time in the stacked area and sankey
 - Datasets
 - Region
 - World
 - East Asia and Pacific
 - Europe and Central Asia
 - Latin America and Caribbean
 - Middle East and North Africa

- South Asia
 - Sub-Saharan Africa
 - Income Group
 - High income
 - Upper middle income
 - Lower middle income
 - Low income
 - Sex - Both (Can only show all Ages groups) / Male / Female
 - Age Group:
 - All ages (show for all sexes options)
 - Various ranges from 0 months to 70+ years for specific sexes
 - Stacked Chart Mode:
 - Stack by Amount and by Percentage
 - Sort Sankey:
 - Sort by Category or by Amount
- Graphs
 - Stacked Area Graph
 - When hovering over each stacked area, more information of the DALY cause are shown to the right
 - Sankey Graph
 - When hovering over a DALY cause node

Project files structure

- CSS
 - bootstrap.min.css (Downloaded)
 - colorbrewer.css (Downloaded)
 - fastselect.min.css (Downloaded)
 - font-awesome.min.css (Downloaded)
 - jquery-ui.css (Downloaded)
 - style.css
 - Created by Lo and Jonathon
- data
 - countrydalys.csv (Data Used for Choropleth/Line Graph and Scatter Plot Matrix)
 - healthdatabycountry.csv (Data Used for Choropleth/Line Graph and Scatter Plot Matrix)
 - lifeexpectancy.csv (Data Used for Choropleth/Line Graph and Scatter Plot Matrix)
 - region_top_level_data.csv (Data Used for Choropleth/Line Graph and Scatter Plot Matrix)
 - sankey_lookup.csv (Data Used for Sankey and Stacked Area Graph)
 - world-110m.json (Data Used for Choropleth)
 - world-country-names.tsv (Data Used for Choropleth)

- DALY_2000_2012 (Data Used for Sankey and Stacked Area Graph)
 - Age-standardized by country.xml
 - World Bank Income Groups High Income.xml
 - World Bank income groups Low income.xml
 - World Bank income groups Lower middle income.xml
 - World Bank income groups Upper middle income.xml
 - World Bank regions East Asia and Pacific.xml
 - World Bank regions Europe and Central Asia.xml
 - World Bank regions High income.xml
 - World Bank regions Latin America and Caribbean.xml
 - World Bank regions Middle East and North Africa.xml
 - World Bank regions South Asia.xml
 - World Bank regions Sub-Saharan Africa.xml
 - World by Cause.xml
- fonts
- img
- js
 - choropleth.js (Created by Jonathon)
 - Created to display the world map when the country tab is selected.
 - d3 (Downloaded d3 libraries)
 - d3.geo.projection.v0.min.js
 - d3.js
 - d3.legend-color.js
 - d3.legend.js (modified to avoid conflict with legend-color)
 - d3.min.js
 - d3.sankey.js
 - d3.tip.js
 - linegraphdalys.js (Created by Jonathon)
 - main.js (Created by Jonathon)
 - otherThirdParty (Downloaded libraries)
 - bootstrap.min.js
 - Used for layout bootstrap
 - colorbrewer.js
 - Used for colors in the choropleth graph.
 - fastselect.js
 - Used for the multi-select filter for countries to remove.
 - highlight.min.js
 - jquery-ui.js
 - Used to change pages
 - jquery.min.js
 - Basic jquery selector package
 - queue.min.js
 - Used to load the all files necessary for processing

- topojson.js
- sankey.js (Created by Lo/ Updated by Jonathon)
- scatterplotmatrix.js (Created by Jonathon)
- stack.js (Created by Lo/ Updated by Jonathon)

Progress Log

Week: Mar 14 - Mar 20

- Started Research Data
 - Starting to come up with the questions we would like to ask
 - Started answering the Who, What, When, Where, Why questions
- Setup Code Repository
- Setup Shared Drive for Misc. Project Information
- Research Visualization to use for Project

Week: Mar 21 - Mar 28

- Both Signed team agreement
- Wrote Goals & Tasks essay
- Wrote Description of Data document
- Drew visualization sketches
- Made elevator pitch slides
- Created screencast video & uploaded to youtube

Week: Mar 29 - April 4

- Finalized questions to answer:
- Finalized features of visualizations (See Feature List documentation)
- Website Layout/Storytelling
- Drew Visualization sketches version 2.0
- Project Timeline

Week: April 5 - April 11

- Received feedback from Group 2 on our visualization
- Gave feedback to Group 4 on their visualizations
- Worked through feedback from Group 2
- Setup basic website structure and tested homepage.
- Write an entry in your process book about the feedback you received in the expert evaluation and how it influences your project re-design.
 - The feedback that we received in our expert evaluation demonstrated to us that we needed to make our visualizations more clear to achieve our goal, which is to show how DALY's have impacted the world. We realized that we need to limit and be more precise in what DALY's have impacted. In order to do this we have

decided to update the choropleth graph on the front page to include not just a summary table of the data, but instead added another graph that includes life expectancy. Not only did adding the graph give the same information, we are now able to also show change over years and Life expectancy with double Y axes. In addition to making the front page convey our objective immediately, we also updated the secondary level pages to reflect our ideas better as well making it more clear that this section is dealing with regional information as defined by World Bank. We are going to try and create a Bubble and Radial Sankey diagram, which will replace the Cluster Tree, this could be an innovative solution to convey the message and bring the data together and minimize the number of pages that user would navigate to. Making some of the functionality simpler as noted by the feedback will make it easier for the user to use. Finally, making the bubble graph easier to read and helps better communicate our message.

- Also finalized who will be doing which visualizations and have the visualizations details almost completely worked out.
- We have decided to stick with the years 2000 and 2012 due to issues with the data sets being incomplete for other years.

Week: April 12 - April 19

- Received Feedback from TF:
 1. educate the user about the implication of the term DALY
 2. present a question or questions
 3. let the users discover and answer the questions by using your visualization
 4. Considering using: health expenditure data, scatterplot matrix
- Started building the visualizations and the site
 - Choropleth with line chart for DALY on country-level
 - Sankey chart for DALY cause groups

Week: April 20 - April 26

- Improved the choropleth and sankey visualizations from last week
- Added two more visualizations: scatterplot matrix chart on the Country page and Stacked area chart on the regions / income groups page
- Set up the project site on github

Week: April 27 - May 2

- Visualization: minor fixes on the details
- Added DALY introduction to the project site
- Recorded presentation screencast
- Completed process book
- Completed the project poster

References

[1] The burden of disease and what it means in England. Retrieved from Public Health England <https://publichealthmatters.blog.gov.uk/2015/09/15/the-burden-of-disease-and-what-it-means-in-england/>

[2] Metrics: Disability-Adjusted Life Year (DALY). (n.d.). Retrieved from http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/