Harry Liu
August Ning
Nathaniel Brooke
Gouttham Chandrasekar
Gabriela Rodriguez-Florido

COMPSCI 316 Milestone #1

Description:

Our goal is to provide a user friendly way for users to get aggregate information about various conditions. We will merge health data from several sources such as the IHME, CDC, and Enigma. The queries are informational, so you would be filtering down based on location, demographic, or other relevant data point information. For example, you might be curious about the number of male adults above age 65 who were affected by malaria in low income countries. One area where data could be updated would be null values - where we simply lack the information, but that information could be gathered. Additional updates could be insertion of new years and disease metrics.

Plan to Populate Data:

We have found relevant data from the following sources:

i. **Global Health Data Exchange** - a global catalog of global health and demographic data compiled by the The Institute for Health Metrics and Evaluation (IHME), an independent global health research center at the University of Washington.
ii. **CDC Health Data** (healthdata.gov) - the U.S. government collection of data
iii. **Enigma** - a public database collection from Enigma

These sources provide us with the health data information we need, but we will need to look elsewhere (depending on the dataset) to acquire population, GDP, and average income information for each place. The plan is to get these from FRED (economic resource database for the US), or for world data from the World Bank.

Sample Data:

See included Excel spreadsheet and SQL code.

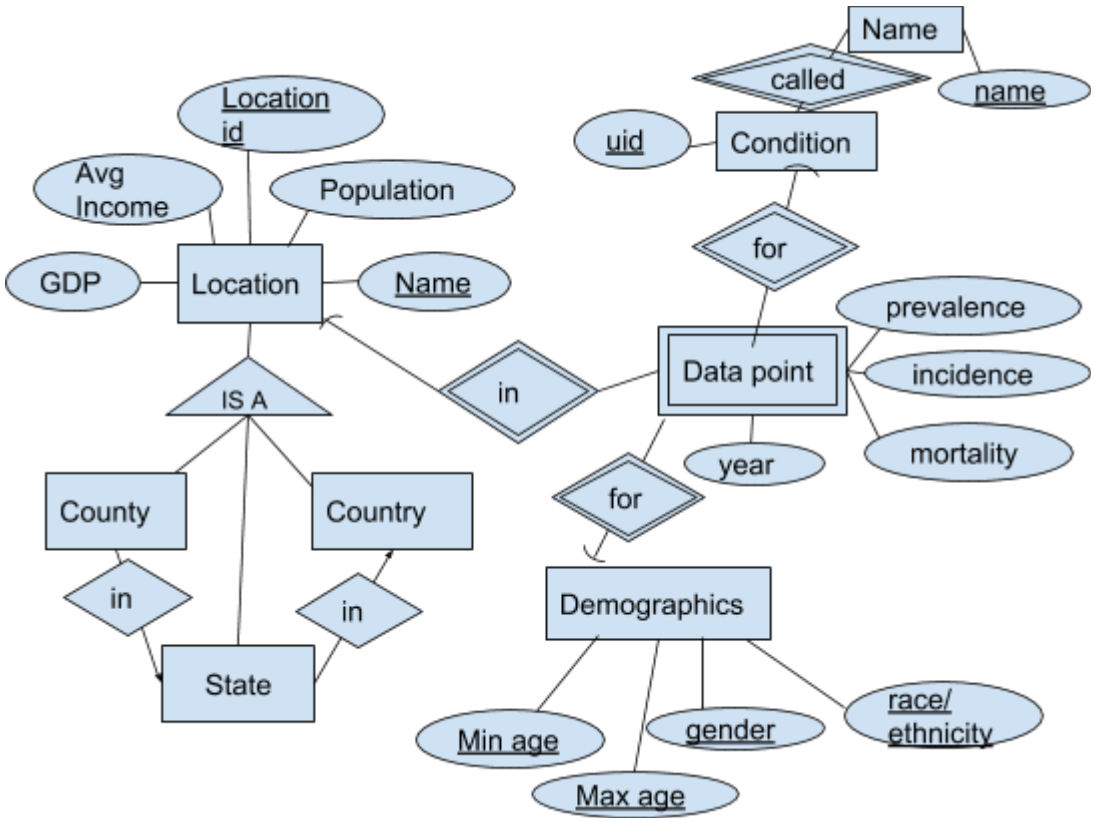
Assumptions About the Data:

- We are assuming that there will be a given age range for the age data, so there are variables named minimum and maximum age for the age in the demographics entity set.
- We are not assuming that condition names are a unique identifier, so we have given each condition a unique id (uid) to account for conditions that may have multiple names or common names.
- The same as above is true for the country, state, and county names.

Harry Liu
August Ning
Nathaniel Brooke
Gouttham Chandrasekar
Gabriela Rodriguez-Florido

- We are assuming that every disease potentially has a prevalence, incidence, and mortality; we may have these fields as NULL for conditions where this data is currently unavailable to us, but this does not mean that the data will not exist at some point.

E/R Diagram:

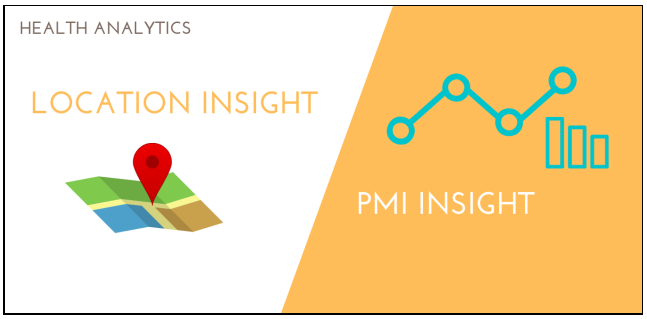


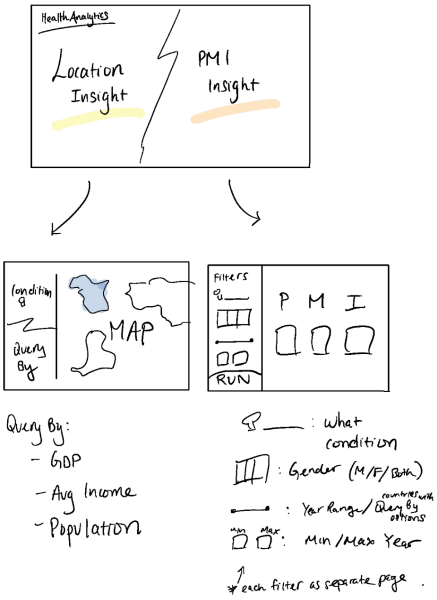
Tables:

DataPoint(condition id, location id, prevalence, incidence, mortality, year, min age, max age, gender, race/ethnicity)
Condition(condition id)
ConditionName(name, condition_id)
Location(uid, name, population, avg_income, gdp, type)
In(uid, enclosing_id)

Harry Liu
August Ning
Nathaniel Brooke
Gouttham Chandrasekar
Gabriela Rodriguez-Florido

Description of Web Interface (and link):





In our first iteration of the web application, the user will have two ways to interact with our health database. There will be an option for location insights and PMI (prevalence, mortality, and Incidence) insights. If the user clicks on the location insights option on the home landing page, they will see a map and filter options. After specifying a condition, the map will highlight regions with the highest mortality related to that condition. Additionally, the user can choose the Query By options. In that case, the user can adjust the region filters and the top 10 countries with the highest mortality related to that condition will be highlighted. For the second option, PMI insights, the user will be able to adjust filters such as gender, region, year, and min/max age to obtain PMI (prevalence, mortality, and incidence) values for the specified restrictions.