Visualizing Associations Between Health, Economic, and Environmental Data by State

Andrew Pitney

Synopsis

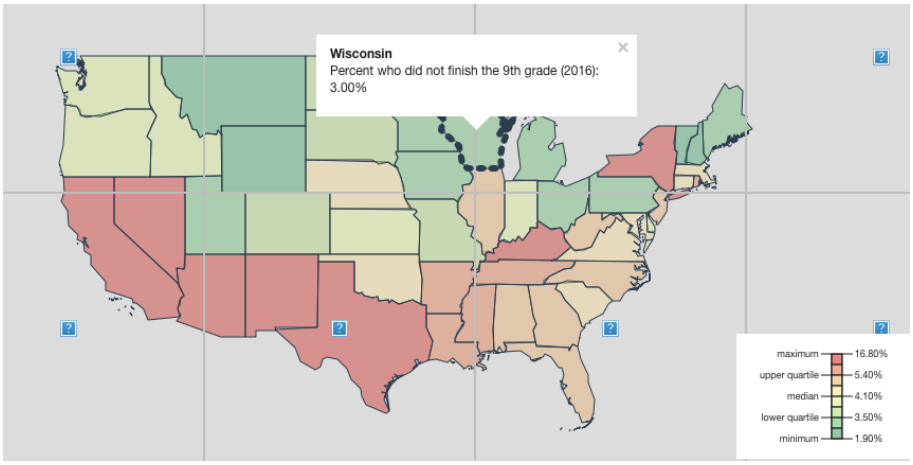
Every day institutions collect mountains of data. In 2013, the world had collected 4.4 zettabytes of information. By 2020, that number is set to rise to 44 zettabytes (or about 44 *trillion* gigabytes) (www.northeastern.edu). But how can the public make meaningful sense this information? How can an everyday citizen look at correlations between intimidating diseases like cancer and environmental issues like pollution? This project sets out to make state health, economic, and environmental data more accessible, visual, and relatable to anyone with internet access. On the front end, I imagine this project as website. Behind the scenes, there will be a small database of compiled research data. In the database, values will be speared by state, health, economic, and environmental data. With simple selections, a user will be able to view data in 3 different ways. The first will be a list view of attributes for a selected state. The second will be a heat map for positive and negative attributes for all states. A user may select a second attribute in the map view, where the map will show positive or negative correlations as a heat map. Lastly, users may look at a graph view of 2 selected attributes. states will be displayed on this graph based on x and y-axis. This will give insight whether correlation exists between things like drinking rate and physical inactivity. Future upgrades to the site may include county-based data for the entire United States, in addition to expanded attributes.

Functionality

There will be three main uses of the platform by a user:

1) **By-State View**

   A user selects a "by-state" view. This selection will give a simple list-rundown of a single selected state's economic, health, and environmental data. This use case will require a number of SELECT-FROM-WHERE statements that gather data from each tuple with the associated state, as well as aggregate functions to determine averages for regions like the Midwest, Northeast, West,..etc.. The following is an example of what the web page would display with "Percent that did not finish 9th grade" selected:



Obtainted from https://www.opendatanetwork.com/entity/0400000US55/Wisconsin/education.graduation_rates.percent_less_than_9th_grade?ref=entity-question&year=2016

2) **Map View**

   A user selects a "map" view and either one or two attributes. This view will likely use GROUP BY queries to determine value thresholds for colors of each state on the map.

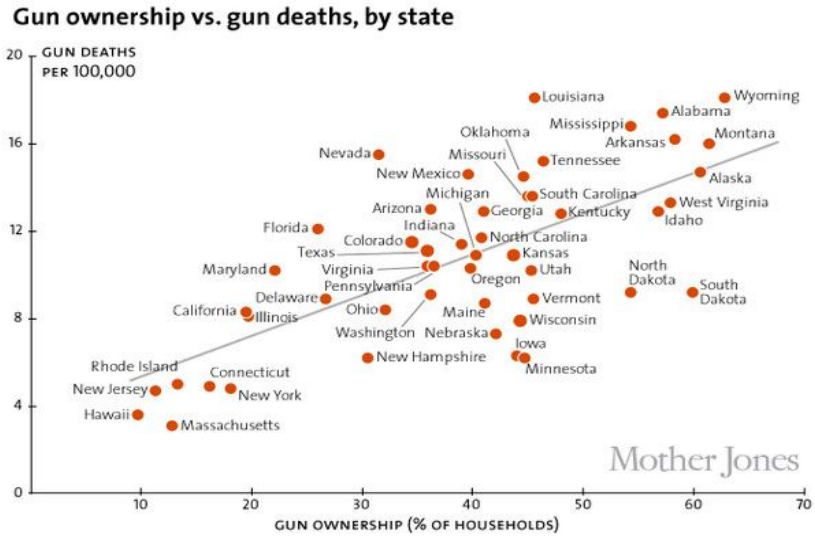
   a. If one attribute is selected, states with extreme rates of the attribute will show up as more red and states with less extreme rates of the attributes will show up as more green. For example, states with lower median incomes will be more red than states with higher median income, and a separate search with states that have higher rates of a certain type cancer will be more red compared to states with lower rates of a cancer types.

   b. If two attributes are selected, the system will look at positive and negative correlations between positive or negatively associated attributes. Scenarios for positive and negative correlations are as follows:

      *High State Cancer Rate * Low Median Income = Red*
      *Low State Cancer Rate * Low Median Income = Red-Green*
      *Low State Cancer Rate * High Median Income = Green*

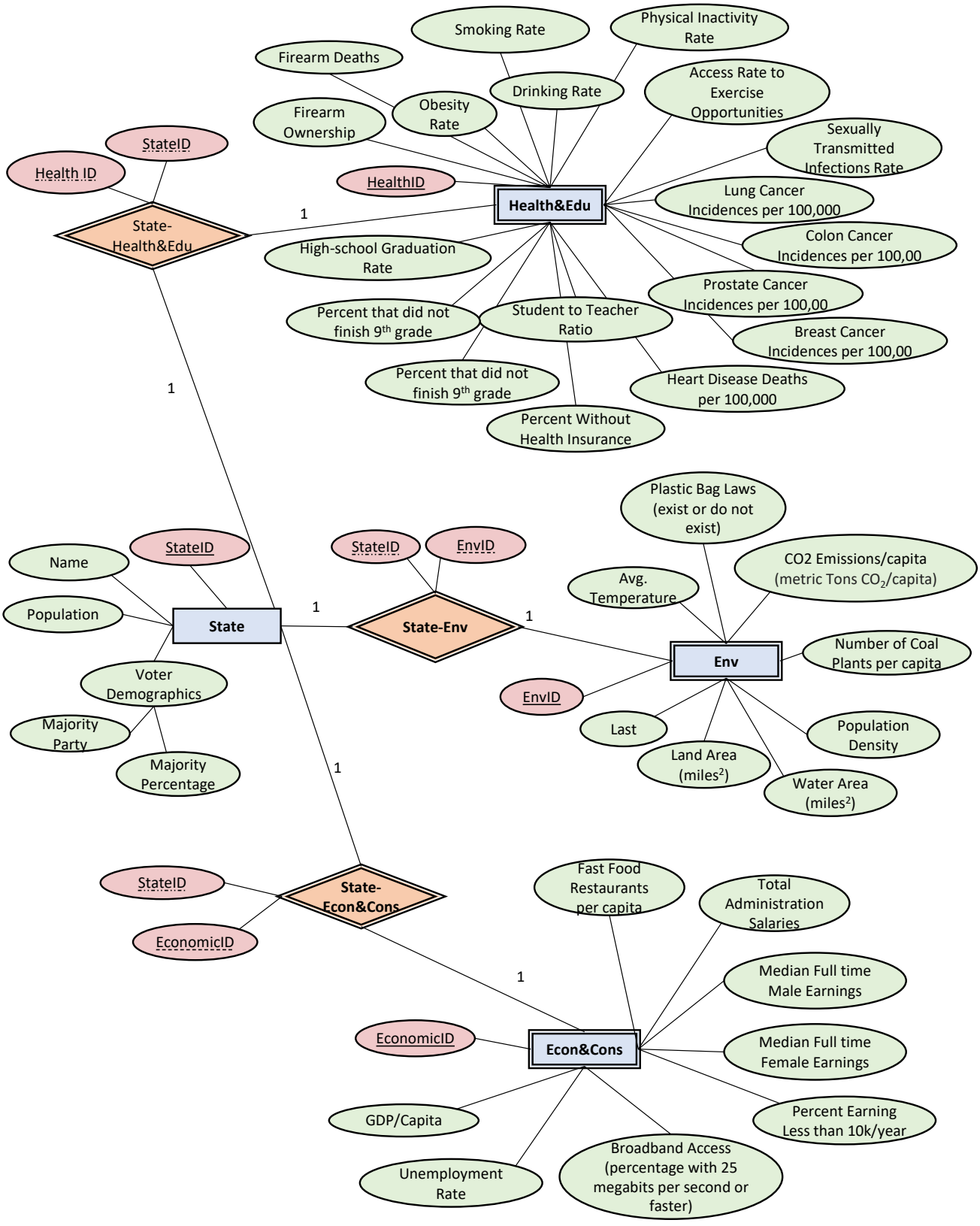
3) Graph View

A user selects a "graph" view and two attributes- one for the x-axis, and one for the y-axis. From here, a graph will be constructed based on the range of values that exist for each attribute, with each point being a state's position in each axis. This view will likely use ORDER BY queries to display and connect data together, as wel as MIN and MAX aggregate functions to determine the scaling of the graph. The following is an example of what an output graph might look like. This graph relates gun ownership with gun death rate:



Obtained from: https://www.vox.com/policy-and-politics/2017/10/2/16399418/us-gun-violence-statistics-maps-charts

Database


Smoking Rate
Physical Inactivity Rate
Firearm Deaths
Drinking Rate
Access Rate to Exercise Opportunities
Firearm Ownership
Obesity Rate
Sexually Transmitted Infections Rate
StateID
Health ID
HealthID
Health&Edu
Lung Cancer Incidences per 100,000
State-Health&Edu
1
Colon Cancer Incidences per 100,00
High-school Graduation Rate
Prostate Cancer Incidences per 100,00
Percent that did not finish 9th grade
Student to Teacher Ratio
Breast Cancer Incidences per 100,00
Percent that did not finish 9th grade
Heart Disease Deaths per 100,000
Percent Without Health Insurance
1
Plastic Bag Laws (exist or do not exist)
StateID
StateID
EnvID
Name
CO2 Emissions/capita (metric Tons CO2/capita)
Avg. Temperature
Population
State
1
State-Env
1
Env
Number of Coal Plants per capita
Voter Demographics
EnvID
Last
Population Density
Majority Party
Land Area (miles2)
Majority Percentage
Water Area (miles2)
1
StateID
State-Econ&Cons
Fast Food Restaurants per capita
Total Administration Salaries
EconomicID
1
Median Full time Male Earnings
EconomicID
Econ&Cons
Median Full time Female Earnings
GDP/Capita
Percent Earning Less than 10k/year
Unemployment Rate
Broadband Access (percentage with 25 megabits per second or faster)

State Entity
The State table stores basic demographic data for each state, and serves as a major access point for database information. This table will be used by the website to show trends among all states and attributes found on Health&edu, Env, and Env&Cons tables.

Health&Edu Entity
Stores information relevant to the public health and education of each state. Values range from cancer rates, graduation rates, and activity rates.

Env Entity
This table stores information about the environment of a selected state. Values include geographical information, as well as sources of pollution like coal plants and if plastic bag laws exist.

Econ&Cons
This economy and consumer table stores information relevant to the economic and consumer health of each state. Values range from the number of fast food restaurants per capita, as well as poverty rates, and rates of internet access

Relational Tables
Relational tables like State-Econ&Cons, State-Env, State-Health&Edu serve to connect states with their respective health information, as well as multi relational joins. Multi-relational joins will rate multiple attributes of a particular state. Relationships and connections between attributes are visualized when joined attributes for each state are put on a graph or map.

Stakeholders

Users of this database could range from those curious about quality of life in their state might be. Other users may be the public health sector, looking at how their state compares to others, or even analyzing vague correlations between different health, economic data, and environmental data. Researchers might use this site as a stepping stone toward looking at why some states are doing well, and why other states aren't.

Technical Requirements

I envision this platform as an HTML/CSS/JavaScript based web application. The site will used SQL-powered searches, based on user input, of coupled data tables. Tables will be filled with data posted on websites like www.opendatanetwork.com or www.data.gov. A user will be able to visualize relationships between data in a by state view, map view, or by a graph view.