Assignment 1 - CS 458

Amber Horvath, Alannah Oleson, and Katherine Bajno

1 Introduction

1.1 What's the problem?

Within the last decade, numerous works have surfaced that suggest climate change has detrimental effects on many aspects of the environment. One indicator of climate change in a geographic region is air quality, which is measured in parts per million of particulate matter. Generally, any particulate less than 2.5 microns in diameter meets the standards for dangerous particulates. An area that has a high concentration of dangerous particulates - a high number on the air quality index, which corresponds to bad air quality - can be both a symptom of or a catalyst for climate change. Identifying areas in which the air quality is markedly bad or decreasing over time can provide a way to focus climate change studies and environmental science efforts.

1.2 Why does this problem need a visualization?

Though there is much data available on the topic of air quality, very little of it is not simply presented in a table or list. Of the visualizations that do exist, many are simply colored maps that make darker or more saturated colors correspond to worse air qualities. Thus, it can be hard to see just which areas present a problem over time (signified by either a continual or sudden, severe decrease in air quality). A visualization that could clearly show both the magnitude of the air quality index and give an indication of how the quality was increasing or decreasing over time would be very useful to researchers in the field.

1.3 Who are our potential users?

Our potential users include researchers and climate change/air quality scientists who study geographic regions in the western US. (We will focus our visualization on this region in order to make the scope appropriate for this project.) This visualization will ideally give scientists a quick overview of air quality trends in an area, which could indicate that the region requires more study or analysis in future work.

In addition, we should not forget that the general population might benefit from a good visualization of this data as well. For instance, perhaps a person who has asthma might view the data as part of a decision on whether or not to relocate to a certain state. A citizen activist might also be interested in this data in order to raise awareness in their region about the dangers of poor air quality. Multiple cases such as these exist and might be well served by this visualization.

1.4 What is our general approach?

We will pull data from the American Health Rankings site by the United Health Foundation. We will use this to aggregate data from the air quality measures of 13 western region states over the past 10 years. We then intend to make a series of tree maps that visualize two dimensions of the data: the magnitude of the air pollution levels (represented by the size of the block; larger = worse) and the rate of change in air quality levels (represented by the color of the block; green = decreasing/getting better, red = increasing/getting worse, more saturated = changing faster).

- *Amber Horvath, computer science student at Oregon State University: E-mail: horvatha@oregonstate.edu.*
- *Alannah Oleson, computer science student at Oregon State University. E-mail: olesona@oregonstate.edu.*
- *Katherine Bajno, computer science student at Oregon State University. E-mail: kbajno@oregonstate.edu*

2 Visualization Tasks

- Our visualization aims to address the following questions:
  - How fast is the air quality increasing and decreasing in each state?
  - What states on the west coast are most at risk of bad air quality?
  - What trends in air quality can we identify in air quality on the west coast over the past 10 years?
  - Which states can be identified as danger zones for further research (air pollution that is quickly increasing)?

3 Data Sources

We will be pulling data from the United Health Foundations website that catalogues data about each state by year. We will be studying 11 different states in the western US. We chose the western United States as the data set was relevant to our team's interest and to keep the project within a manageable scope.

- Washington: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/WA
- Oregon: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/OR
- California: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/CA
- Alaska: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/AK
- Hawaii: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/HI
- Montana: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/MT
- Wyoming: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/WY
- Colorado: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/CO
- New Mexico: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/NM
- Idaho: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/ID
- Utah: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/UT
- Arizona: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/AZ
- Nevada: http://www.americashealthrankings.org/explore/2015-annual-report/measure/air/state/NV

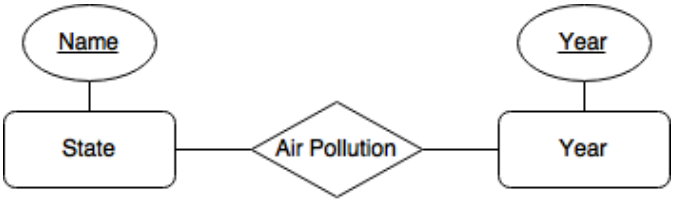

Fig. 1. An ER diagram showing the relationship between our 3 tables: the State table has a primary key ”name” and the Year table has a primary key ”year”, both of which are used to query on the Air Pollution table

Sort by key: None

	Name
	Alaska
	Arizona
	California
	Colorado
	Hawaii
	Idaho
	Montana
	Nevada
	New Mexico
	Oregon
	Utah
	Washington
	Wyoming

Fig. 2. A visualization of the contents of the ”state” table, with the ”name” column serving as the primary key and key to index the ”air pollution” table.

4 Data Organization

We set up our data in the form of 3 tables, 1 for the states, one for the years, and one for the air pollution. The entity-relationship diagram for these tables can be seen in Figure 1. The state name and year are used as keys to index the values stored within the air pollution table. This data was retrieved from the United Health Foundation’s report on air pollution across the United States.

5 Implementation of Data Organization

We created a relational database with 3 tables to store our data. The ”state” table and ”year” database are used to index values stored within the ”air pollution” database, as seen in the ER diagram in Figure 1. We implemented the database on the ONID Database Server provided by Oregon State University. The contents of the ”state” diagram can be seen in Figure 2 and the ”air pollution” configuration can be found in Figure 3.

6 Design of the Interface

We represent our data using an interactive tree map on a webpage. On the left hand side, users can select what year they want to view from

	Field	Type	Collation	Attributes	Null	Default	Extra	Action
	Air_Pollution	varchar(3)	latin1_swedish_ci		No			
	name	varchar(20)	latin1_swedish_ci		No			
	year	varchar(4)	latin1_swedish_ci		No			

Check All / Uncheck All With selected:

Fig. 3. The ”air pollution” table’s columns include ”air pollution”, ”name”, and ”year” with ”air pollution” containing all the values of the air pollution for each state and year, and ”name” and ”year” serving as keys.

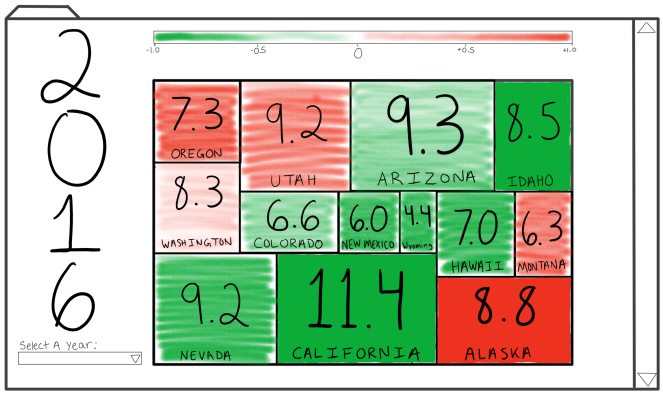

Fig. 4. A visualization using a tree map describing the levels of air pollution in the western region by block size.

2007 to 2016, giving them the ability to see the changes in air pollution over the years. The size of the boxes on the tree map represent how much or how little air pollution there is in a state. The larger the box, the more air pollution exists. The color of the boxes represent the change in air pollution from the previous year to the one selected. The range is from -1 to +1, with -1 representing the highest positive change and the +1 representing the highest negative change in air pollution. The visualization is represented in Figure 4 .