Assignment 2 - CS 458

Amber Horvath, Alannah Oleson, and Katherine Bajno

Abstract— On a large scale, many recent studies link high air pollution levels to the larger problem of climate change. On a somewhat smaller scale, a large amount of air pollution can not only affect the climate of a certain geographic region - it can also negatively affect the quality of life for those who live there. Over time, the rate at which the air quality of a region is increasing or decreasing can single the area out for inspection by climate scientists or give the inhabitants of the area an indicator of the overall healthiness of the environment there.

To address both issues in a way that the average person can understand, we created a visualization of air pollution levels in the western United States. We use a treemap format to represent the magnitude of 11 states' air pollution levels over the span of ten years. We also make use of a color gradient to indicate the rate at which polltion levels are increasing or decreasing. This is a novel representation because much of this kind of data is presented in tables that are rather hard to parse. Our visualization, on the other hand, gives the user a good overview of general trends in the region at a glance, while still allowing them to drill down into the details if they want to. This can benefit both average people (for instance, someone looking at which state in the region they want to move to) and those with specialized needs such as climate scientists.

◆

---

1 INTRODUCTION

1.1 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 Motivation

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 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

---

- *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*

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 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).

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 RELATED WORK

Climate change has been studied by many research groups across different domains. However, how groups choose to visualize the data changes depending upon the focus of the group and what questions they're trying to answer. The teams software capabilities also come in to play, with some groups researching air pollution having little to no background in programming or information visualization. Zell et al. (2010) found that, between the years 1955 and 2006, 26,253 different titles were published relating to air pollution from 124 different countries [Zell et al. 2010].

Judging by this influx of publications, air pollution is clearly an issue that has garned interest from people of all backgrounds and interest. It is also important to note that humans play a large role in air pollution, with NASA releasing satellite visualizations of our impact on air quality [NASA]. Air pollution is also extremely dangerous, with an estimated 2.4 million fatalities occurring each year due to unsafe living conditions caused by air pollution[WHO]. Different groups of researchers have attempted to create visualizations that both inform experts and the general public while also being used to glean meaningful information about different atmospheric trends.

In an effort to inform city planners on how to structure future traffic schemes, Zahran et al. ran a case study to test their implementation of a 3D city model with an overlay of the changing air pollution through different types of visualizations[Zahran et al.]. Through their different tests, they found that users best understood the metaphor of smog clouds covering the more polluted areas and less clouds over the more clear areas. Perhaps an easily understandable metaphor is a good way to inform the public on levels of air pollution - however, how to do this on a state by state basis and capture the temporal aspect may not be possible with the cloud metaphor.

Elbir created a system to estimate the ambient air pollution levels temporally and spatially with the goal of mapping emissions and air quality levels [Elbir 2004]. The target users are policy makers in order to inform and predict air quality changes within the region they are evaluating. Elbir uses darker markings on a map of the area to show concentration of harmful air pollutants. The advantage of Elbir's system is that, not only does it visualize current air pollution, it also predicts future pollution levels based upon the data currently available. This might be an avenue of future research for our visualization to help citizens of the states we are visualizing to better understand whether their state is at risk or to see if it is imporiving over time.

Wang et al. created a viewpoint-based method for rendering of visualizations that promotes speed and interactivity [Wang et al 2010]. Speed is important as, if the data is consistently being sampled, being able to render and update on the fly leads to the most accurate and relevant information being shared. This type of sampling is important for government officials to keep their patrons informed if a certain area is reaching dangerous levels of pollution. The scope of our project does not reach that level of criticality, but if we shifted our audience towards government officials, their system might be an avenue for implementation efforts.

The kind of work Wang et al. are doing is similar to the work of Li et al. Li et al. aimed to create many different visualizations to inform policy makers in Beijing what is causing the rapid increase in air pollution and to analyze the trends in order to combat the air pollution [Li et al 2016]. The authors attempted different methods of visualizations and found certain types of visualizations lead to different information being conveyed (e.g. scatter plots helped locate where data was missing). The authors also found that wind speed affected change in air pollution. Li et al. also leveraged some open source software in order to create their visualization. One such open source software is Openair, an R software package used to visual air quality [Openair]. These findings suggest our visualization could be improved through supporting different visualizations depending upon what question the user is trying to answer through our visualization. For future work, we might also consider leveraging some of these open source software options in order to keep our product flexible and receptive to incoming data.

Another different type of visualization researchers have tried is Bi et al.'s tree visualization [Bi et al.]. The tree is used to represent different polutant levels for each day throughout a month in a given area. This metaphor allows for a high level of accuracy, but does not provide a broader scale view or allow for easy comparison between different regions.

Beyond purely visualizations but in the vein of informing the public, Bohler et al. developed a system called APNEE (Air Pollution Network for Early warning and online information Exchange in Europe) as a ways of informing select citizens through mobile messages, online posts, and street panels when air pollution reaches certain critical levels [Bohler et al.]. Perhaps if we leveraged our software beyond merely a visualization and created an app, we could also distribute our information to patrons. For now, that is outside of the scope of our assignment, and wouldn't make a lot of sense unless we had more frequent updates on climate change (currently our system functions on a year-long comparison level).

4 BACKGROUND

After doing a review of previous literature, we found most previous reseach was aimed at informing policy makers and other officials of the state of their local air pollution rates. These visualizations, while similar to our work, differ slightly in that they are made by experts for experts. In our case, we are seeking to inform the greater public on a more general level of how their state is doing in comparison to other states and whether their state's air pollution is increasing or decreasing. Our goal is to provide an easy to understand visualization for non-experts to track their state's pollution levels.

5 METHODS

5.1 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

5.2 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.

5.3 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 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.
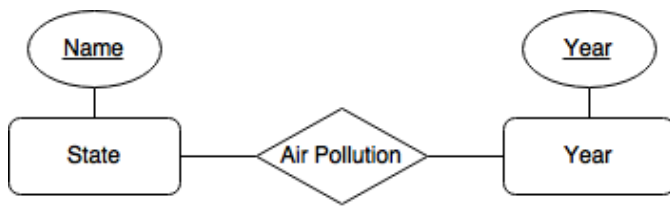
Fig. 1. An ER diagram showing the relationship between our 3 tables: the State table has a primary key "name" and the Year table has a primary key "year", both of which are used to query on the Air Pollution table

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.

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.
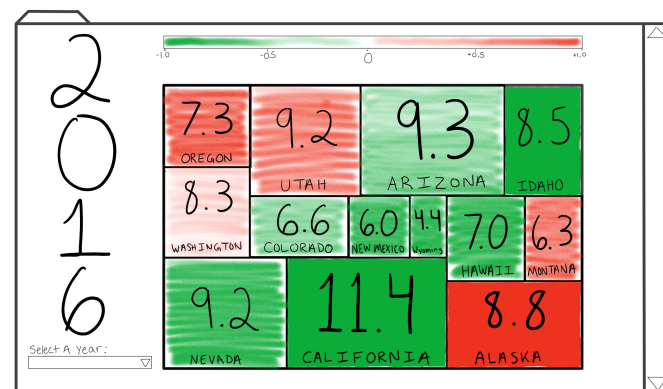

Fig. 4. A visualization using a tree map describing the levels of air pollution in the western region by block size.

5.4 Enhancement Over Existing Models

After reviewing the current models available, our model provides an enhancement by providing an easy-to-understand metaphor for non-experts to view how their state's air pollution is changing over time. We achieve this through size and color changing to represent whether the air pollution levels are rising or lowering over time. For example, a large orange** block means that a state's air pollution is rapidly increasing.

Previous models require an understanding of atmospheric sciences and are designed in mind for policy makers in the domain of city planning[Zahran et al.]. While users with more experience in the domain of atmospheric sciences may use our software, we hope to create a visualization that is accessible to users of all backgrounds.

6 IMPLEMENTATION

6.1 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.

6.2 Website

Alannah

6.3 Issues Encountered

Alannah

7 RESULTS

Alannah

7.1 Insights

** ??? I don't really know what kinds of insights we'd find but we will see I guess **

7.2 Data Set

** Yeah we already said this up in Methods so we could just reference that again I guess??? **

7.3 Dimensions Used

** Our visualization is both spatial (size of block) and temporal (can show change over time) so I guess talk about that/ why we chose that visualization/if it worked or not **

7.4 Performance

** no idea here since we're not running any sort of tests on it or doing any user studies lmao **

7.5 Supplementary Materials

** alannah said: link to source code - so I guess here we could link to our github or something **

7.6 Domain Expert

8 CONCLUSION AND FUTURE WORK

Amb

9 ACKNOWLEDGEMENTS

Amb