# 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 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

********** FILL IN WITH PREVIOUS WORK IN THE FIELD OF AIR POLLUTION RESEARCH AND VISUALIZATIONS ***********

- [Zahran et al 2013]: did case study in area with increasing air pollution to test their 3d city model that incorporates different visualizations to represent the air pollution in the city. They tried 3D point shapes, 3D planar surfaces, and 3D volumetric clouds. The clouds worked the best due to the metaphor of them appearing like storm clouds. Benefits of visualizations include teaching the public about raising levels of pollution and possibly influencing traffic schemes to take into account changing pollution levels. Implications for our work: perhaps clouds are useful for visualition however do they accurately convey more or less pollution versus just pollution versus no pollution?

- [Elbir 2004] Created system based on CALPUFF dispersion model - estimates ambient air pollution levels temporally and spatially. Allows for mapping of emissions and air quality levels. Implication for our work: their display is more akin to an atmospheric visualization over an area as opposed to comparing distinct regions within an area so we're tackling different problems.

• *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*

- [Wang et al 2010] Created viewpoint-based method during rendering, promotes speed and interactivity. Ran experimental tests, believe it can benfit environment monitoring. Can load large data sets and display them quickly. Implications for our work: We're implementing a 2D interface, so perhaps if we moved to a 3D space we could leverage this work, but it's outside of the scope of our implementation efforts.

- [Zell et al 2010] meta study about the topic of air pollution visualization research. Basically stated that there's been a lot of publications recently about air pollution research because it's so critically important to current world affairs. Implication for our work: air pollution is important so our work is important.

- [Li et al 2016] Researchers in China wanted to analyze air pollution trends happening in Beijing. They used open source tools (PivotTable.js and D3.js) to create many different visualizations to visualize spatio-temporal visualizations. They used many different visualizations such as line graphs, area graphs, tables, and so on. Implication for our work: perhaps we can leverage the open source tools they used as both are hosted on Github. Perhaps utilizing some different visualizations along with our square plot will help lead scientists to come to different conclusions based upon our work.

## 4 Background

********** FILL IN WITH QUESTIONS WE'RE TRYING TO ANSWER *************

## 5 Methods

### 5.1 Data Sources

**note: should probably just move all these links to citations and just summarize here **double note: also need to add eastern region states

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

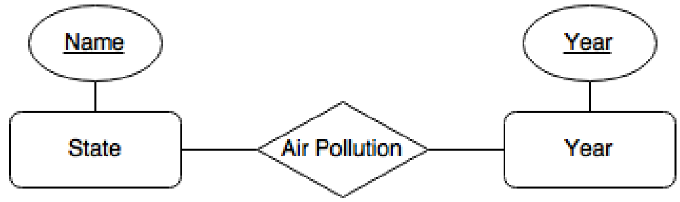


Fig. 1. An ER diagram showing the relationship between our 3 tables: the State table has a primary key "name" and the Year table has a primary key "year", both of which are used to query on the Air Pollution table

Sort by key: None

| ←T→ | | | 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.

### 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 **??**. 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 **??** .

| | 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.

### 5.4 Enhancement Over Existing Models

** Fill in with why our design is better/more helpful than other already existing models - should tie back to Related Works ** ** Idea - whoever writes this section should also be in charge of related work **

## 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 **??**. 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 **??** and the ”air pollution” configuration can be found in Figure **??**.

### 6.2 Website

** fill in with how we created website - might need Kat’s help here **

### 6.3 Issues Encountered

** fill in with any issues encountered probably while creating visualization/website - again might need Kat’s help here **

## 7 RESULTS

** intro to what we found **

### 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 **

## 8 CONCLUSION AND FUTURE WORK

** some longform paragraph here about our model, why it’s awesome, etc. how we would want to change it in the future **

## 9 ACKNOWLEDGEMENTS

** TO AMBER FROM AMBER: i think the original tex doc he supplied had some special thing for acknowledgements so copy paste** ** Here we just acknowledge the website we got data from (?) and Eugene and the TAs???? **