

Project Proposal

Air condition of major cities in California

Names: Spencer Fronberg, Mattia Grespan, Qian Zuo

October 28, 2018

1 Basic Info

- The project title: **"Air Conditions of major cities in California"**
- Mattia Grespan, mattia.grespan@utah.edu, u1131814
- Spencer Fronberg, u0766439@utah.edu, u0766439
- Qian Zuo, u1142882@utah.edu, u1142882
- Project repository link: <https://github.com/spence521/dataviscourse-pr-Air-Condition.git>

2 Background and Motivation

Many environmental factors can affect our health, but the one that have the most impact to us is air pollution. We breathe air to live and what we breathe has a direct impact on our health. Environmental Temperature and humidity are also important factors that influence human's living standard. Let us talk about these three factors' effect towards human as below.

Low quality air can put us at a higher risk for asthma and other respiratory diseases. Air pollutants are mostly carcinogens and living in a polluted area can put people at risk of Cancer. Coughing and wheezing are common symptoms observed on city folks.

So how the temperature influence us? According a recent research, death rates become progressively higher when outdoor air temperature rises above or falls below 20-25°C. This is largely attributable to the direct effects of exposure to heat and cold on the human body in general, and on the circulatory system in particular. Besides, for cold-related mortality, the analysis yielded evidence of an important indirect effect involving increased incidence of influence of influenza and other respiratory infections [1].

When it comes to the humidity of the air, people in different part of the United States have rather different feelings. Breathing dry air is a potential health hazard which can

cause such respiratory ailments as asthma, bronchitis, sinusitis, and nosebleeds, or general dehydration since body fluids are depleted during respiration. Skin moisture evaporation can cause skin irritations and eye itching [2].

In the past research, people focused on only one issue of air pollution, temperature and humidity, however, these three issues influence human's living standards a lot. We should do something to combine these issues together. So in this project, we will collect and show data of these three aspects for major cities in California. The reason we choose California for one part is the three factors varies a lot among cities in California, for the other part is California is a wonderful working place for software engineers and we will graduate soon, this is a good chance for us to give the students who want to work there some suggestions. After all, fresh air, convenient temperature, and suitable humidity are important for human's health, let's do something meaningful.

3 Project Objectives

We want to create a very friendly and informative visualization about the air conditions (temperature, humidity and pollution) in eight major cities in California. This visualization should help users to have a general idea about the air conditions in these cities in the simplest and interactive way possible. The cities are: Sacramento, San Francisco, Los Angeles, San Diego, San Jose, Bakerfield, Fresno and Redding.

4 Data

We are downloading our data from the PurpleAir website (<https://www.purpleair.com/sensorlist>). Our data includes three different levels of pollution, but we are only focused on the PM2.5 particles level measured as CF ATM ug/m³. PM2.5 is a microscopic particle 2.5 microns in width and almost 30 times smaller than the diameter of a human hair. When levels are high, PM2.5 particles form a haze in the sky, making their way into peoples respiratory tracts and reaching the lungs [3].

Our data also includes humidity and temperature. Each entry in the data sets is recorded by sensors every one to two minutes. There are many different data sets that are available to us. We will focus on the vicinity of downtown (including it) areas for the major cities in California. There are about 20 to 30 sensors for each city. Our data is the format of csv files.

5 Data Processing

For each data set, we plan to process it to get an average pollution, humidity, and temperature for each sensor in increments of either one hour or half hour. We might even reduce it to ten minute increments. We then plan to take few of the averaged data sets per city and average them so we get accurate data to represent each city. We will then have the data

that represents each city with their averages in increments of half hour to an hour (we have not decided yet). We plan to do this for eight or nine cities in California to see trends in each city.

6 Visualization Design

The decisions we propose as the final design for this visualization are the following:

1. Three buttons to select the condition of the air that the visualization will work with.
2. A brush bar to select any sequence of days. The visualization will use these information and generate the average of the values during that period of time.
3. Use the shape of the California state as an spatial region. This could be seen as a mark, but is also a channel to convey to give the user a better general sense of the location and distance between the cities the visualization is showing.
4. Circles to indicate each city (marks). This circles will be colored according to the value of the condition the visualization is currently showing. There is a color saturation bar with a specific hue gradient for each condition. Even though the saturation of color can be effective in showing the intensity of the selected air condition, we decided that the circles will also include the number of units that the color is encoding. Note that this last step is redundant but we think is important in order to the lecture of the data even easier to the user.
5. Once the user interactively clicks (or hover) on a city, a plot of the time in function of the current selected air condition will show up. This plot will use time by hours. In order to compare the selected air condition on different cities the user can click more than one city and the graph for both cities will show up on top of the other.
6. Display more information of the city: population, number of autos, traffic, etc., as well as the final description summarizing the data for the current selected condition according to the data during the selected period of time.

The following Figure 1 and Figure 2 show digital sketches of our final visualization design. The original sketches (on paper) were included in the appendix of this proposal.

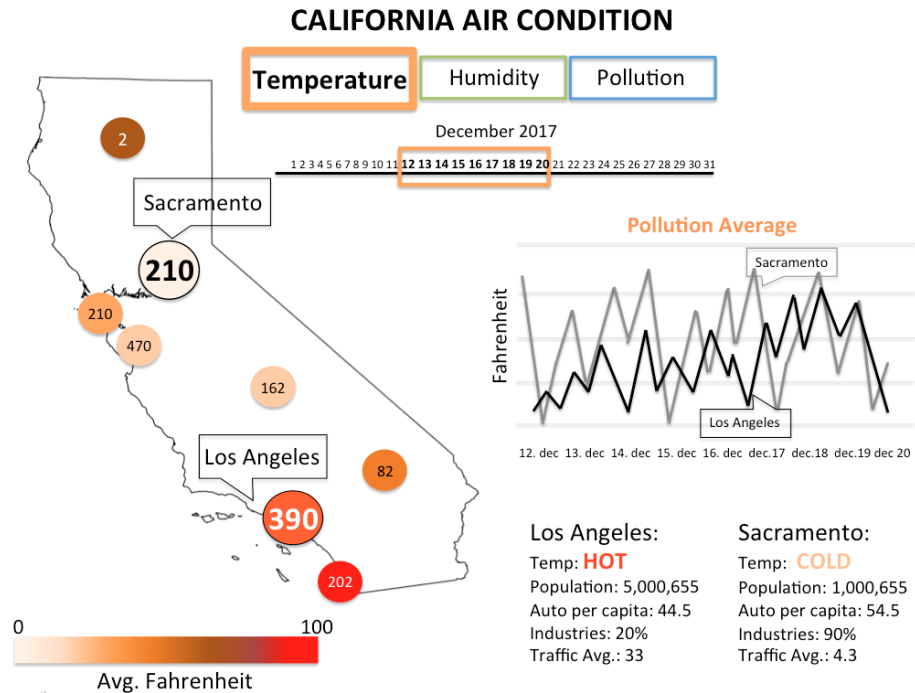


Figure 1: Final design (Pollution)

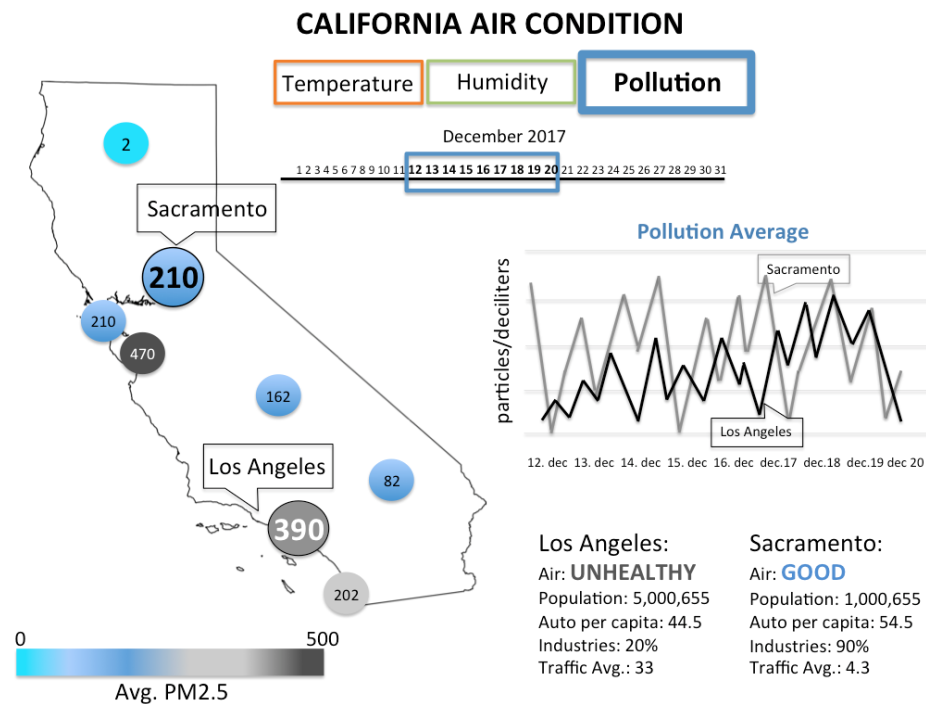


Figure 2: Final design (Temperature)

Furthermore, Figure 3 shows the digital sketch for one of the first alternative prototype designs we created.

California Air Condition

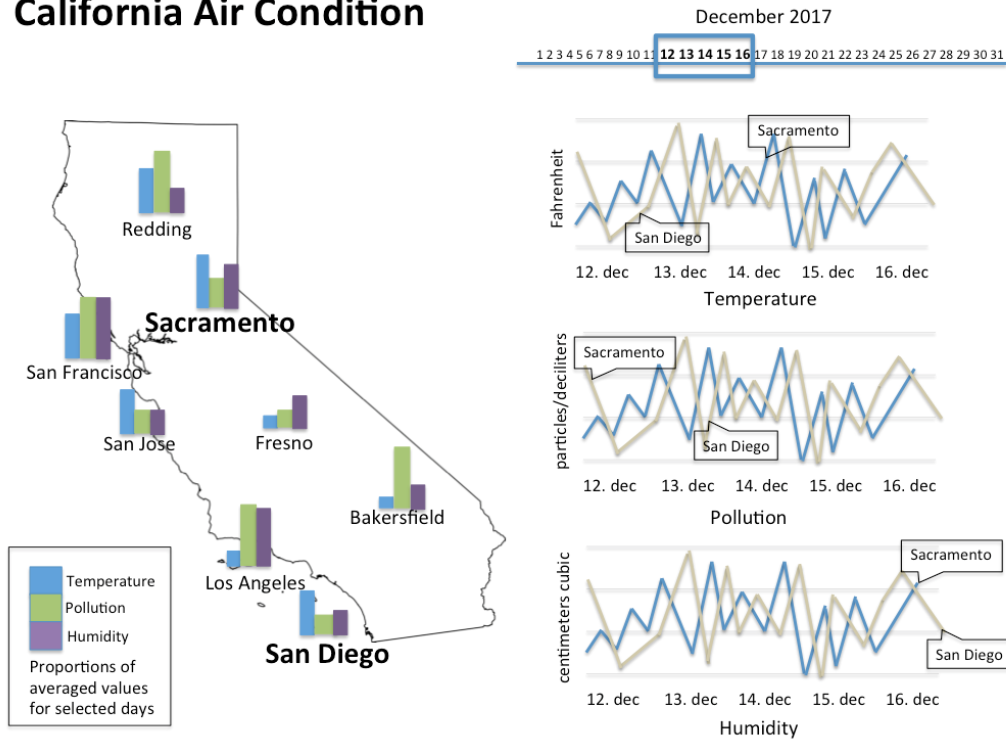


Figure 3: Prototype design

Some of the problems we found on this first approach were:

- Occlusion in case of adding more cities to the visualization. Furthermore, the little bar charts and the name of the cities would not be readable.
- Different units. Even though the intention of the bar chart is to show simple proportions of each of the values of the air conditions (pollution in blue, humidity in green and temperature in purple), the units of each of them are different. This is not a very good way to encode this values at once. We tried to encode this three values in different ways but none of them convinced us (See sketches in appendix).
- In general one of the main problem we faced was to encode the three variables Temperature, Humidity and Pollution on the same map. We tried several options. Some of our different initial approaches to solve this can be traced in the sketches of the appendix.

7 Must-Have Features

Every feature described on the visualization design should be included in our final visualization.

8 Optional Features

- Give suggestions to people: when someone has some specific requirements for the air conditions of the city he/she wants to live, we will provide some range for these three factors. That's to say, we will give three drop-down boxes for choices, search the goal cities according to the requirements and return the ideal cities for him/her.

9 Project Schedule

- Week1: Scrape, clean up and extract data.
- Week2: Implement the platform and build the frame of it.
- Week3-4: Implement the Must-have features.
- Week5: Try to implement the Optional features.

References

- [1] Anton E. Kunst Casper W. N. Looman Johan P. Mackenbach. Outdoor Air Temperature and Mortality in the Netherlands: A Time-Series Analysis. American Journal of Epidemiology, Volume 137, Issue 3, 1 February 1993, Pages 331341, <https://doi.org/10.1093/oxfordjournals.aje.a116680>
- [2] Infoplease. Effects of Dry Air on the Body. <https://www.infoplease.com/science-health/weather/effects-dry-air-body>
- [3] The World Bank. IBRD. Understanding Air Pollution and the Way It Is Measured. <http://www.worldbank.org/en/news/feature/2015/07/14/understanding-air-pollution-and-the-way-it-is-measured>

APPENDIX

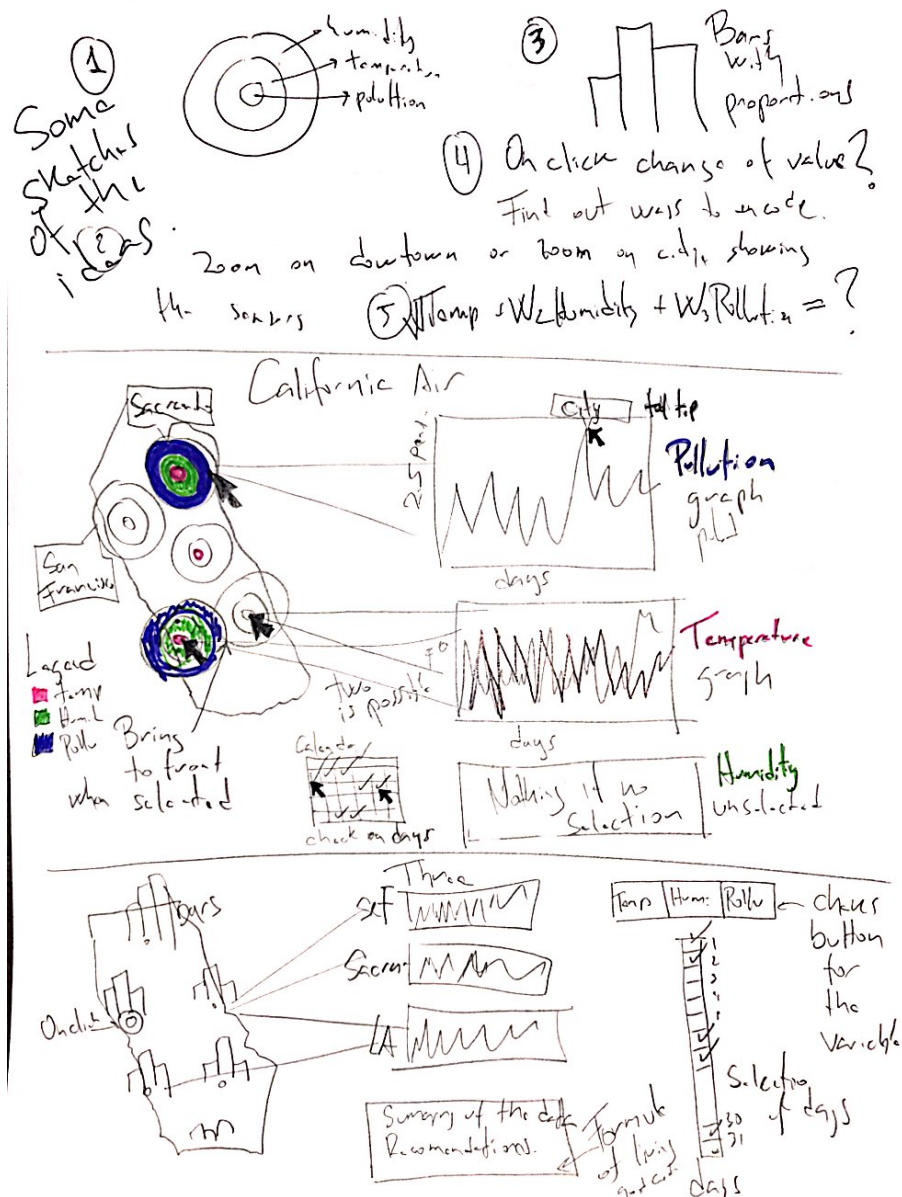


Figure 4: Prototype designs

