W209 Data Visualization Final Project Proposal

Fall 2017 Section 3

Team Members

Varadarajan Srinivasan
Amy Smessaert
Matthew Nelson

Project Concept

We have formulated a few different ideas for Air Quality Index visualizations (depending on available time). Ideally, we will complete them all as a series of visualizations on the same webpage.

Purpose of Visualization

Our main goal of the visualization will be to be to take a data-rich, often-neglected subject and make it quick and intuitive to understand. We have captured some of our going-in hypotheses/rationale for each visual but we also expect it to generate more follow-up questions as we delve into the data. To accomplish this goal, we plan to create 4 different categories of visuals, with an emphasis on interaction to enhance the viewer's engagement while answering important questions. The visualizations can also be seen as a catalyst for debunking (or proving) commonly stated AQI “facts” such as AQI is better after it rains.

Category A: Visuals that focus on identifying cities/counties that are more impacted than others
Hypothesis/Questions: Are there counties that have differential air quality over others. How has air quality changed over time? Are counties with poor air quality index directly proportional to population or do we see any outliers?

1. A single viz showing air quality index by county
    a. Averaged by year
    b. Choropleth map Visualization
    c. Interactive sliding time-scale
    d. Interaction to show a vertical line passing through Viz #3 as time scale moved here

Hypothesis/Questions: Are there cities that have differential air quality over others. How does that compare to other cities and rest of the country? Does this change by day of week (week day vs. weekend vs. holidays etc.)

2. A single viz showing the change in AQI by city
    a. Averaged by day
    b. Line Chart Visualization
    c. User-selection for City
    d. Range bars for Upper quartile / Lower Quartile of other cities in the same county/ whole country?
    e. The ability to select multiple cities and compare?

Hypothesis/Questions: Deep dive into cities with poorest AQIs. Is this a new trend or has this been true over a period of time. Is there an inflexion point that may coincide with policy changes/regulation etc. Is this heavily skewed towards more populated cities?

Streamgraph to show the top 20 worst Cities and their AQI change over time

    f. May need to normalize to show on a percentage basis?
    g. Hover and Highlight interactions.
    h. Interaction to highlight the same city on Viz #1 and sync time scale on X Axis
    i. Click on a city to bring it up in Viz #2

Category B: Visuals that show second order factors such as topography and its influence on air quality

Hypothesis/Questions: How much of this influenced by geography/topography of the region. Do topographical variances alleviate/enhance the effects of AQI such as prolonged smog or lesser rain etc.

3. Visualize topography as related to AQI
    a. Draw topographical map
    b. Layer on county AQI data
    c. Visually show relationship between topography and AQI levels
        i. E.g. Do valleys trap smog?

Category C: Visuals that focus on environmental impacts caused by air quality index

Hypothesis/Questions: How does AQI impact other natural and man-made phenomena such as precipitation and smog. Do cities with higher AQI also see a greater incidence of smog. How is precipitation impacted by smog for those cities

4. Histograms of daily smog by city to visualize patterns/predictors of AQI changes
    a. Example: https://flowingdata.com/2017/07/12/peak-times-for-leisure-and-sports/
    b. Vertical axis = City
    c. Horizontal axis = Day
    d. Group cities by proximity to each other
    e. Visually display typical smog movement
        i. E.g. If AQI is high in Wyoming today, will AQI rise in Colorado tomorrow?
5. Visualization of precipitation to AQI levels

a. Line graphs
    i. One line for AQI
    ii. One line for precipitation
b. I.e. Is it true that precipitation makes AQI decrease?
c. Dual vertical axes for AQI and precipitation
d. Horizontal axis for time/day

Category D: Visuals that focus on potential health risk assessment, and policy decision making
Hypothesis/Questions: Do cities with vehicular traffic caused pollution show a differential composition of AQI compared to pollutants generated by industrial smoke. Why is this important? Potential policy decisions may vary based on pollutant composition and relative health risks of nearby population.

7. Relative concentration of various pollutants in metropolitan cities vs. industrial suburbs
    a. Stacked bar chart with saturation/contrast representing less risky pollutants to more risk pollutants
    b. Has this changed over a period of time. Click on the city/county from stacked bar chart a) to zoom in city/county stats over time

Hypothesis/Questions: Do cities/states with poorer AQI show a greater incidence of respiratory/lung diseases? Why? Cause has to create an effect. Is there tangible/measurable human impact?

8. Plot cities/state maps with incidences and deaths by lung and bronchos cancer

Datasets

The EPA keeps a very clean Air Quality Index Dataset on their website. This data shows daily averages and is separated at the both city and county level.

https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI

Weather data: OpenWeather API

Topographical data: http://www.webgis.com/index.html
Lung Cancer data: https://www.cdc.gov/cancer/lung/statistics/state.htm

Challenges

With such large datasets being loaded into our page, it may be difficult to keep user latency down.