Air Pollution and Mortality in NYC

CUS 625 Project Presentation

Caifeng Ye

5-2-2016

Introduction

- Air pollution affected many cities adversely in developing country over recent years.

- How is the situation in developed countries? According to a recent American Lung Association (ALA) report, more than half of people in the U.S. still breathe air dirty enough to cause health problems and four out of ten people had unhealthy air quality in their communities. (1)

- As the most biggest and populous city in the United States, New York City draws my attention.

---

(1) America Lung Association."State of the Air 2015" 29 April 2015 Press Materials Web. 7 April. 2016

Harmful Air Pollution on Mankind

- Many studies show that exposure to air pollution has been associated with mortality.

- Six major air pollutants:
  1. Ozone(O3) ,
  2. Fine Particles(Particulate Matter that is 2.5 Micrometers in Diameter and Smaller, also called PM2.5)
  3. Particles of 10 micrometers or less (PM10),
  4. Carbon monoxide (CO)
  5. Sulfur dioxide(SO2),
  6. Nitrogen dioxide (NO2)

- U.S. Environmental Protection Agency (EPA) pointed that PM2.5 can aggravate heart and lung diseases.

- Ozone is linked with asthma attacks and respiratory diseases.

EPA Breakpoints for Six Pollutants for AQI

- The Air Quality Index (AQI) is the system used to warn the public when air pollution levels are dangerous.
- EPA Breakpoints for Six Pollutants for AQI (https://www3.epa.gov/airnow/aqi-technical-assistance-document-dec2013.pdf)

O_3 (ppb)	O_3 (ppb)	$PM_{2.5}$ (µg/m^3)	PM_{10} (µg/m^3)	CO (ppm)	SO_2 (ppb)	NO_2 (ppb)	AQI	AQI
C_{low} - C_{high} (avg)	C_{low} - C_{high} (avg)	C_{low}- C_{high} (avg)	C_{low} - C_{high} (avg)	C_{low} - C_{high} (avg)	C_{low} - C_{high} (avg)	C_{low} - C_{high} (avg)	I_{low} - I_{high}	Category
0-59 (8-hr)	-	0.0-12.0 (24-hr)	0-54 (24-hr)	0.0-4.4 (8-hr)	0-35 (1-hr)	0-53 (1-hr)	0-50	Good
60-75 (8-hr)	-	12.1-35.4 (24-hr)	55-154 (24-hr)	4.5-9.4 (8-hr)	36-75 (1-hr)	54-100 (1-hr)	51-100	Moderate
76-95 (8-hr)	125-164 (1-hr)	35.5-55.4 (24-hr)	155-254 (24-hr)	9.5-12.4 (8-hr)	76-185 (1-hr)	101-360 (1-hr)	101-150	Unhealthy for Sensitive Groups
96-115 (8-hr)	165-204 (1-hr)	55.5-150.4 (24-hr)	255-354 (24-hr)	12.5-15.4 (8-hr)	186-304 (1-hr)	361-649 (1-hr)	151-200	Unhealthy
116-374 (8-hr)	205-404 (1-hr)	150.5-250.4 (24-hr)	355-424 (24-hr)	15.5-30.4 (8-hr)	305-604 (24-hr)	650-1249 (1-hr)	201-300	Very Unhealthy
-	405-504 (1-hr)	250.5-350.4 (24-hr)	425-504 (24-hr)	30.5-40.4 (8-hr)	605-804 (24-hr)	1250-1649 (1-hr)	301-400	Hazardous
-	505-604 (1-hr)	350.5-500.4 (24-hr)	505-604 (24-hr)	40.5-50.4 (8-hr)	805-1004 (24-hr)	1650-2049 (1-hr)	401-500	Hazardous

Five Understanding

- Overall trends towards air emissions of New York City from 1995 to 2015. Specifically, year-round fine particular pollution (PM2.5) and Ozone(O3) pollution.

- Air pollution in different area in NYC. Where are the cleanest area and the most polluted area?

- Trends towards mortality from cardiovascular diseases (CVDs), chronic lower respiratory disease (CLRD) and pneumonia in NYC from 1997 to 2014.

- The association between mortality from cardiovascular diseases (CVDs), pneumonia and CLRD and the concentrations of air pollutants: ozone (O3) and PM2.5.

- Considering the living area, gender and age, who is at most risk?

Data Sources

- Pollutants Data

  Data on concentrations of PM2.5 (µg/m3) and O3 (ppm) in the air from 1995–2014 were obtained from annual summary files from EPA which contain data summarized on an annual basis. The database is open to the public through the EPA's website.(http://www.epa.gov/ttn/airs/air-saqs/detaildata/downloadaqsdata.htm)

- Mortality Data

  The causes of death were coded according to the International Classification of Diseases, Tenth Revision (ICD-10). The mortality data were classified as deaths due to total nonaccidental causes (ICD-10 codes A00–R99), cardiovascular disease (ICD-10 codes I00–I99), respiratory disease (ICD-10 codes J00–J98). All the annual mortality data from 1997 to 2014 were obtained from annual vital statistics from New York State Department of Health (NYSDOH). (https://www.health.ny.gov/statistics/vital_statistics/)

User Definition

- Users: Casual Users

- People who are interested in environment issues and have been concerned for the air pollution and related mortality. But these people do not have to be an expert in order to gain general understanding of the field.

- A casual user may be a resident who lives in NYC and wants to know air pollution condition in his/her living county. A casual user also can be a student who wants to do a research on comparing the air pollution situation over years.

- Requires : straightforward, easy to use, do self-explanatory

Story 1: Understanding 1 & 2

As is the case for casual users, the design adapts the story telling method. The visualization is designed into 3 stories. Each story tries to address certain understanding. Story one is about the trends of air pollutants (PM2.5 and Ozone) in different counties in NYC.

•Understanding 1 & 2:

- Overall trends towards air emissions of New York City from 1997 to 2014. Specifically, year-round fine particular pollution (PM2.5) and Ozone(O3) pollution.

- Air pollution in different area in NYC. Where are the cleanest area and the most polluted area?

Story 1: Understanding 1 & 2

- Background introduces general knowledge to the users in order to let users get quick grasp.

- The 1st graph is a site map, shows all the monitoring sites for PM2.5 and Ozone in NYC from 1995 to 2014. Users can use filters to see sites location for different pollutant and county.

- The 2nd & 3rd dashboards shows maps of PM2.5 and Ozone concentration across NYC from 1999 to 2014, I encode color with parameter measures to show changes of different county over years. Users can use filter to see the trend of a county.

- The 4th dashboard are time-series line plots which demonstrate PM2.5 and Ozone overall trends in NYC. The plots used reference lines to show under which point is good and above which point are supposed to be unhealthy.

- The 5th graph are gantt bars to show trends in different county over years. You may see that the second graph lacks the data of Kings.

Story 1: Understanding 1 & 2

•Key Interaction Features

As we know, comparing is one of the most effective ways to interact with data. The main interaction features used in this story is comparison. For example: by encoding colours with measures, the design uses simple heat-maps to exhibit the overall trends of PM2.5 and Ozone of different counties.

• What do you think the visuals say in support of the understanding or decision being made

We can clearly see from the story that the concentration of PM2.5 has been declined by 66.94% from 29.49 ug/m3 to 9.75 ug/m3 in past 20 years . While Ozone levels have not improved substantially. It was fluctuating and increasing recently.

As for the different county in NYC, we can see from the gannt bars that New York county has the highest average concentration of PM2.5 and Richmond has the highest average level of Ozone. In contrast, Richmond has the lowest PM2.5 level. Ozone level in Bronx is better than the other counties, but the trend showed the level is increasing yearly in Bronx.

Story 2: Understanding 3

Story 2 is about the trend of related death.

- Understanding 3: trends towards mortality from cardiovascular diseases (CVDs), chronic lower respiratory disease (CLRD) and pneumonia in NYC from 1997 to 2014.

Story 2: Understanding 3

- Background introduce data source and mortality classification to the users.

- The 1st graph are 2 trend lines, show total death number and mortality rate from 1997 to 2014 (Death from cardiovascular diseases (CVDs), chronic lower respiratory disease (CLRD) and pneumonia).

-The 2nd dashboard are gantt bars which show total mortality rate by county over years. We can find that Richmond has the highest overall mortality rate from the above 3 disease.

-The 3rd dashboard are also gantt bars which describe death rate by cause and county yearly. Even though the overall death rate shows decreasing, for each disease, it may not show the same trend. We only can conclude that the mortality rate of Cardiovascular(CVDs) shows overall declining . We also can detect that the mortality rate of CVDs has similar patterns in different county. So does the CLRD death rate.

- The 4th dashboard are bar charts of the mortality of each disease and was divided by age, gender groups. (In order to easy comparison, I highlight the male group with grey color to separate female and male.) The bar charts demonstrate that for each disease people aged over 64 deaths significantly increased. The other finding is that for people aged less than 64, male mortality is higher than female . While people aged over 65, more female deaths than male. for each disease, same pattern has been seen.

Story 2: Understanding 3

- Key Interaction Features

Also, comparing is the key interaction features in this story. For certain disease, by comparing the mortality rate changed over years, users can find similar patterns in different county.

Story 3: Understanding 4 & 5

Section 3 try to find whether there are some relationship between pollutant and death from certain disease. That is the association between mortality from cardiovascular diseases (CVDs), pneumonia and CLRD and the concentrations of air pollutants: ozone (O3) and PM2.5.

Also, considering the living area, gender and age, the design want to find which group of people are at most risk.

Story 3: Understanding 4 & 5

- The 1st graph is the overall trend of pollutant and mortality rate. As we see in story 1 and 2. PM2.5 level has been declined by 66.94% from 29.49 ug/m3 to 9.75 ug/m3 from 1995 to 2015 in NYC . Ozone levels have not improved substantially. The total number of death has decreased by 33.08% from 30432 to 20364 from 1997 to 2014, the mortality rate has been declined by 37.59% from 2029.6 (per 10000) to 1266.6 from 1997 to 2014.

- The 2nd dashboard shows trends of pollutant and mortality rate by county over years. By comparing time-series lines, we may see Richmond has the lowest average PM2.5 level and highest Ozone level where has the highest mortality rate. In contrast, New York county has the highest PM2.5 level but lowest mortality rate.

- The 3rd and 4th dashboards are scatter-plots try to find the relationship between pollutant (PM2.5 and Ozone) and mortality rate of certain disease.

Story 3 :Understanding 4 & 5

- The 3rd and 4th dashboards are scatter-plots which try to find the relationship between pollutant (PM2.5 and Ozone) and mortality rate of certain disease. For example, users may want to find the relationship between PM2.5 level and CVDs mortality. It is easy to see that they both rise or fall together in the same direction. We can also detect similar patterns in different county when analysis PM2.5 level and mortality from CLRD. But we could hardly find some meaningful relationship between Ozone level and mortality.

- The 5th graph are bar charts previously showed in story 2. It was put in story 3 again to combine with the result of correlation analysis. Thus, users may find which group of people are at most risk, say people's gender, age, county.

Discussion

- **Data**

  -PM2.5 and Ozone data

  The project used data summarized on an annual basis. It would be more reliable and useful to use daily and monthly data. Thus, we can find seasonal changes and patterns over years.

  -Mortality data

  Would prefer to use monthly data. When users do the correlation analysis by scatter-plot , more data will demonstrate the relationship more precisely. The shapes can be displayed more clearly.

Discussion

- **Further Work**

  - Optimizing aspect ratio and quantitative scales.

    In the design, the range of some quantitative values were not scaled properly. For example, when doing the correlation analysis, the scatter-plots does not show good view of patterns.

    To find meaningful patterns, by rescaling to spread the values across as much space as possible.

  - Comparing overall trend of Pollution and Mortality

    May add change rate to each time-serie line.