Climate Change and its Impact on Health

Ameya Shanbhag, avs431

Titash Mandal, tm2761

Vikram Sunil Bajaj, vsb259

TABLE OF CONTENTS

Introduction and Motivation

Weather is the most critical for human in many aspects of life. The Earth's climate has changed throughout history. Just in the last 650,000 years there have been seven cycles of glacial advance and retreat, with the abrupt end of the last ice age about 7,000 years ago marking the beginning of the modern climate era — and of human civilization. Most of these climate changes are attributed to very small variations in Earth's orbit that change the amount of solar energy our planet receives. In the past 100 years, the global average temperature has risen by about 0.74 degrees Celsius (34 Fahrenheit). The study and knowledge of how weather Temperature evolves over time in some location or country in the world can be beneficial for several purposes.

Our project wanted to explore the results of climate data after processing, Collecting and storing of huge amounts of weather data. The data collection is done by the Meteorological department. The Meteorological departments use different types of sensors such as temperature, humidity etc. to get the data. The sensors volume and velocity of data in each of the sensor make the data processing time consuming and complex. We aim at analyzing the climate change and its impact on the environment as well as human health by using the available **Big data technologies**.

We have exploited the opportunities to **mine large climate datasets, temperature datasets and a global health and death dataset with an emphasis on the visualization of the mined data.** We applied Big Data technologies to analyze and draw correlations among the radical features causing climate change. Machine Learning (ML) is all about predicting future data based on patterns in existing data. Weather systems travel large distances on a time scale of hours and days, so recent weather observations from around the country can be used to predict the future weather of one specific site. The last step of our project was to use the **Machine Learning Algorithms** to **predict** tomorrow's **temperature**, by feeding the test set to the model.

Datasets used for the Project:

1. Pollution levels in different cities in the United States:

This dataset deals with pollution in the U.S. and has been well documented by the U.S Environmental Pollution Agency. The dataset displays four major pollutants (Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide and Ozone) for every day from 2000 - 2016.

Feature Engineering:

The original dataset had a total of 28 fields namely:

1. **State Code:** The code allocated by US EPA to each state
2. **County Code:** The code of counties in a specific state allocated by US EPA
3. **Site Num:** The site number in a specific county allocated by US EPA
4. **Address:** Address of the monitoring site
5. **State**: State of monitoring site
6. **County**: County of monitoring site
7. **City:** City of the monitoring site
8. **Date Local:** Date of monitoring

The four pollutants (NO2, O3, SO2 and O3) each has 5 specific columns. For instance, for NO2:

- NO2 Units: The units measured for NO2
- NO2 Mean: The arithmetic means of concentration of NO2 within a given day
- NO2 AQI: The calculated air quality index of NO2 within a given day
- NO2 1st Max Value: The maximum value obtained for NO2 concentration in a given day
- NO2 1st Max Hour: The hour when the maximum NO2 concentration was recorded in a given day

Out of these components, we decided to select the most informative attributes which are relevant to our data mining and analysis procedure.

The features used for data mining are :

NO2 Mean, NO2 AQI, State, City, Date Local

2. GLOBAL TEMPERATURE ANALYSIS

The Global Temperatures by City dataset has monthly average temperatures from 1743-2013, for each city. Since our other datasets had records from 2000-2016, we only considered temperature records for the corresponding years and computed their yearly average temperatures. We also considered only records corresponding to the same cities in the other datasets.

	Attribute Name	Description
1.	dt	1743-2013 (monthly)
2.	Average Temperature	Average Temperature in C
3.	Average Temperature Uncertainty	Average uncertainty in measurement of temperature
4.	City	City Names
5.	Country	Country Names
6.	Latitude	Latitude values
7.	Longitude	Longitude values

Fig-1.1 Figure showing the description of the temperature dataset.

3. NCHS Leading Causes of Deaths: United States

This dataset is based on the information from all resident death certificates filed in the 50 states and the District of Columbia using demographic and medical characteristics. Age-adjusted death rates (per 100,000 population) are based on the 2000 U.S. standard population. Populations used for computing death rates after 2010 are post estimates based on the 2010 census, estimated as of July 1, 2010. Rates for census years are based on populations enumerated in the corresponding censuses. Rates for non-census years before 2010 are revised using updated intercostal population estimates and may differ from rates previously published.

Causes of death are classified by the **International Classification of Diseases**, Tenth Revision (ICD–10) are ranked according to the number of deaths assigned to rank-able causes. Cause of death statistics are based on the underlying cause of death.

The dataset had the following attributes:

	Name	Description
1.	Year	1999-2015
2.	113 Cause Name	Disease Name
3.	Cause name	Disease Category
4.	State	State name
5.	Deaths	Total number of deaths
6.	Age-adjusted Death Rate	To make fairer comparisons between groups with different age distributions

Based on the different causes of death we decided to filter out deaths caused by certain factors like Accidents, Assaults, murder as we realized that they do not have any direct relationship with the pollution levels in the atmosphere. The features we filtered out from our dataset is as follows:

```
'All Causes',
'Assault (homicide) (*U01-*U02,X85-Y09,Y87.1)',
'Assault (homicide) by discharge of firearms (*U01.4,X93-X95)',
'Assault (homicide) by other and unspecified means and their sequelae (*U01.0-*U01.3,*U01.5-*U01.9,*U02,X85-X92,X
'Legal intervention (Y35,Y89.0)',
'Events of undetermined intent (Y10-Y34,Y87.2,Y89.9)',
'Discharge of firearms, undetermined intent (Y22-Y24)',
'Other and unspecified events of undetermined intent and their sequelae (Y10-Y21,Y25-Y34,Y87.2,Y89.9)',
'Operations of war and their sequelae (Y36,Y89.1)',
'Complications of medical and surgical care (Y40-Y84,Y88)',
'Other land transport accidents (V01,V05-V06,V09.1,V09.3-V09.9, V10-V11,V15-V18,V19.3,V19.8-V19.9,V80.0-V80.2,V8
'Water, air and space, and other and unspecified transport accidents and theirsequelae (V90-V99,Y85)',
'Nontransport accidents (W00-X59,Y86)',
'Falls (W00-W19)',
'Accidental discharge of firearms (W32-W34)',
'Accidental drowning and submersion (W65-W74)',
'Accidental exposure to smoke, fire and flames (X00-X09)',
'Accidental poisoning and exposure to noxious substances (X40-X49)',
'Other and unspecified nontransport accidents and their sequelae (W20-W31,W35-W64,W75-W99,X10-X39,X50-X59,Y86)',
'Intentional self-harm (suicide) (*U03,X60-X84,Y87.0)',
'Intentional self-harm (suicide) by discharge of firearms (X72-X74)',
'Intentional self-harm (suicide) by other and unspecified means and their sequelae (*U03,X60-X71,X75-X84,Y87.0)',
'Accidents (unintentional injuries) (V01-X59,Y85-Y86)',
'Transport accidents (V01-V99,Y85)',
'Motor vehicle accidents (V02-V04,V09.0,V09.2,V12-V14,V19.0-V19.2,V19.4-V19.6,V20-V79,V80.3-V80.5,V81.0-V81.1,V82
'All other diseases (Residual) (D65-E07,E15-E34,E65-F99,G04-G14,G23-G25,G31-H93,K00-K22,K29-K31,K50-K66,K71-K72,K
```

Fig- 1.2 These are the list of death causes which we filtered from our dataset as they do not bear any direct relationship with the air pollutants present in the atmosphere. Below is the cleaned table after cleaning in SPARK.

```
+----------+--------------------+-----+------------+
|Max_Deaths|              Causes|Years|       State|
+----------+--------------------+-----+------------+
|   71930.0|Diseases of heart...| 1999|  California|
|   58987.0|Diseases of heart...| 1999|    New York|
|   53067.0|Malignant neoplas...| 1999|  California|
|   51434.0|Diseases of heart...| 1999|     Florida|
|   43418.0|Diseases of heart...| 1999|       Texas|
|   41707.0|Diseases of heart...| 1999|Pennsylvania|
|   38478.0|Malignant neoplas...| 1999|     Florida|
|   37609.0|Malignant neoplas...| 1999|    New York|
|   33387.0|Diseases of heart...| 1999|    Illinois|
|   33192.0|Diseases of heart...| 1999|        Ohio|
+----------+--------------------+-----+------------+
only showing top 10 rows
```

DATA MODELING AND ANALYSIS

AQI- AIR QUALITY INDEX

The AQI is an index for reporting THE daily air quality. It tells us how clean or polluted the air is, and what associated health effects might be a concern for you. The AQI focuses on health effects that people may experience within a few hours or days after breathing polluted air. EPA calculates the AQI for five major air pollutants regulated by the Clean Air Act: **ground-level ozone, particle pollution (also known as particulate matter), carbon monoxide, sulfur dioxide, and nitrogen dioxide**. For each of these pollutants, EPA has established national air quality standards to protect public health. Ground-level ozone and airborne particles are the two pollutants that pose the greatest threat to human health in this country.

Given the air quality index of various gases in our dataset, we realized there were multiple values for AQI each day for each city for years ranging 2000-2016. Thus, our main aim was to first calculate the average daily AQI for each city, then calculate the average monthly AQI levels for each city and filter out the cities with AQI levels lower than 80 because an AQI level above 80 is considered as dangerous by the Environment pollution agency. We used pyspark to analyze our dataset and the result we obtained was as follows.

State	City	Max_NO2_AQI	Max_SO2_AQI	MAX_O3_AQI	CO_AQI	Yearly
Missouri	St. Louis	80.0	103.0	140.0	52.0	2000
California	Not in a city	84.0	21.0	154.0	35.0	2000
Pennsylvania	Reading	57.0	135.0	172.0	22.0	2000
Virginia	Alexandria	74.0	109.0	150.0	25.0	2000
California	Long Beach	108.0	65.0	114.0	63.0	2000
Michigan	Detroit	93.0	134.0	114.0	51.0	2000
California	Davenport	33.0	33.0	47.0	9.0	2000
California	Goleta	54.0	13.0	74.0	18.0	2000
California	Bethel Island	41.0	26.0	124.0	17.0	2000
Pennsylvania	Lancaster	55.0	79.0	177.0	19.0	2000
Kansas	Kansas City	74.0	101.0	132.0	57.0	2000
Illinois	Cicero	84.0	79.0	87.0	38.0	2000
Illinois	Calumet City (PU ...	58.0	136.0	101.0	24.0	2000
California	San Pablo	64.0	39.0	74.0	22.0	2000
Missouri	St. Ann	84.0	91.0	142.0	22.0	2000
California	Vandenberg Air Fo...	31.0	6.0	80.0	6.0	2000
California	San Francisco	72.0	27.0	36.0	36.0	2000
Pennsylvania	Pittsburgh	74.0	117.0	169.0	18.0	2000
Pennsylvania	Norristown	64.0	44.0	207.0	16.0	2000
Indiana	Indianapolis (Rem...	58.0	89.0	142.0	33.0	2000
District Of Columbia	Washington	109.0	104.0	164.0	58.0	2000
California	Concord	72.0	29.0	147.0	31.0	2000
Pennsylvania	Greensburg	49.0	110.0	145.0	17.0	2000
Illinois	Chicago	97.0	102.0	42.0	26.0	2000
New York	Holtsville	67.0	86.0	203.0	32.0	2000
California	Rubidoux	94.0	115.0	192.0	48.0	2000
California	Victorville	94.0	36.0	179.0	18.0	2000
Texas	Houston	101.0	172.0	203.0	42.0	2000
Virginia	Seven Corners	87.0	89.0	161.0	32.0	2000
Pennsylvania	Freemansburg	44.0	59.0	201.0	26.0	2000

Fig-1.3 Table showing the Maximum yearly AQI levels of different gases grouped by year , city and state. This table is not filtered based on the AQI values > 80.

As pollution is growing everyday due to increased wastes produced from factories, cars, household products, ill-practices, we decided to identify which cities have been constantly featured as having a higher pollution level as compared to other cities. We ranked every city based on their pollution level for years ranging from 2000-2016 and displayed the top polluted cities. These cities have constantly shown up in many consecutive years as having an AQI Level for the gases above 80.

State	City	Count
California	Calexico	16
California	Los Angeles	15
California	Long Beach	14
New York	New York	13
California	San Diego	11
Colorado	Welby	11
District Of Columbia	Washington	11
California	Burbank	11
Texas	El Paso	10
Pennsylvania	Philadelphia	10

Fig-1.4 Table showing the top polluted cities and their count. The count represents the number of years for which they have constantly had the AQI levels of gases above 80. It is interesting to notice that New York also features on this list.

Data mining using spark also helped us visualize two cities which have shown a very high level pollution of 3 gases in a certain year. As shown in the table below Houston and Kansas city have 3 gases in above a normal pollution level which comes as a shocking observation.

State	City	Max_NO2_AQI	MAX_SO2_AQI	O3_AQI	NO2_AQI	Yearly
Texas	Houston	101	86.0	150.0	101	2000
Kansas	Kansas City	103	97.0	200.0	103	2005

Fig- 1.5 Figure showing two cities with 3 gases exceeding the save environmental AQI level.

DECISION TREE MODELING USING ML LIB

A decision tree is a map of the possible outcomes of a series of related choices. It allows an individual or organization to weigh possible actions against one another based on their costs, probabilities, and benefits. They can be used either to drive informal discussion or to map out an algorithm that predicts the best choice mathematically.

Decision trees where the target variable can take continuous are called **regression trees**. **We used the Decision tree for a regression analysis to predict the Average Temperature of different from based on their levels of pollution in air.**

Our Target variable was: Temperature

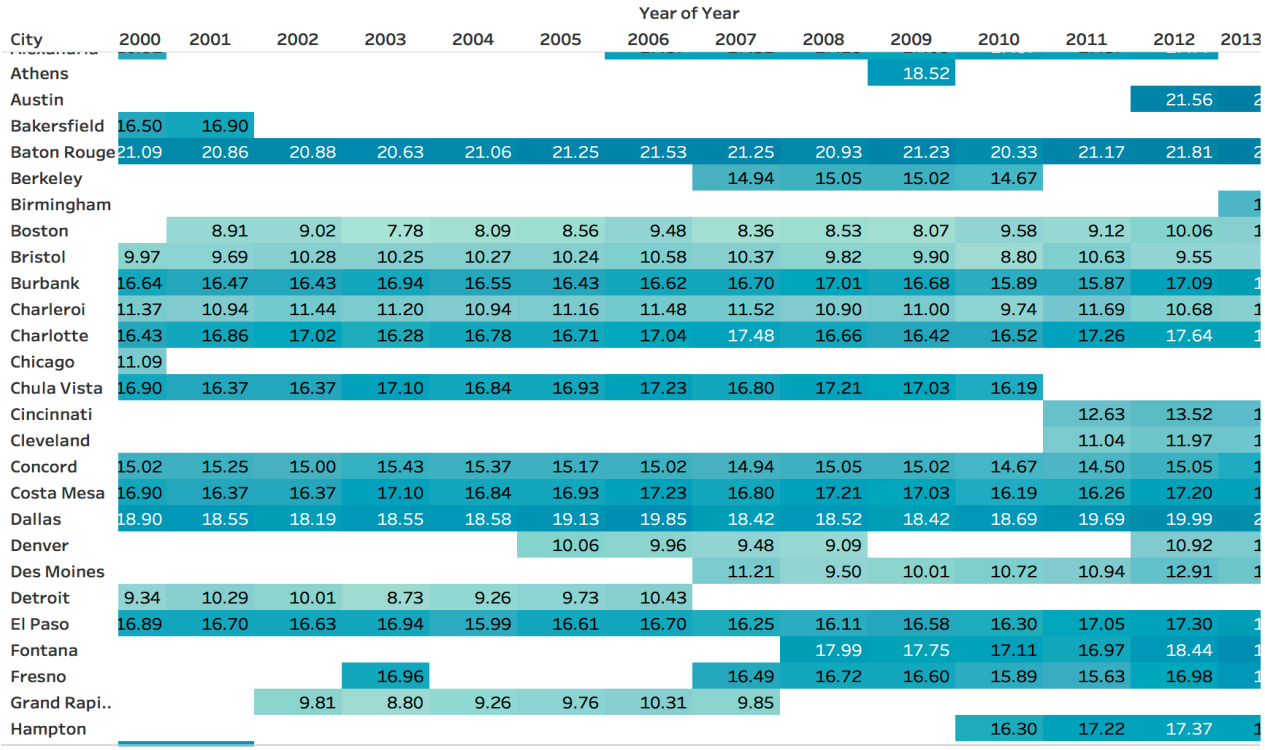
Our feature vectors were: NO2, SO2, CO, O3 AQI levels for all years ranging 2000-2013.

```
Test Mean Squared Error = 16.65643623052192
Learned regression tree model:
DecisionTreeModel regressor of depth 5 with 59 nodes
  If (feature 1 <= 30.0)
   If (feature 1 <= 10.0)
    If (feature 3 <= 100.0)
     If (feature 0 <= 38.0)
      Predict: 21.556083333333333
     Else (feature 0 > 38.0)
      If (feature 1 <= 7.0)
       Predict: 10.955869047619046
      Else (feature 1 > 7.0)
       Predict: 14.403327380952382
    Else (feature 3 > 100.0)
     If (feature 1 <= 9.0)
      If (feature 3 <= 112.0)
       Predict: 15.372856481481483
      Else (feature 3 > 112.0)
```

Fig-1.6 This figure shows the decision tree modeling with only a mean squared error of 16.65%

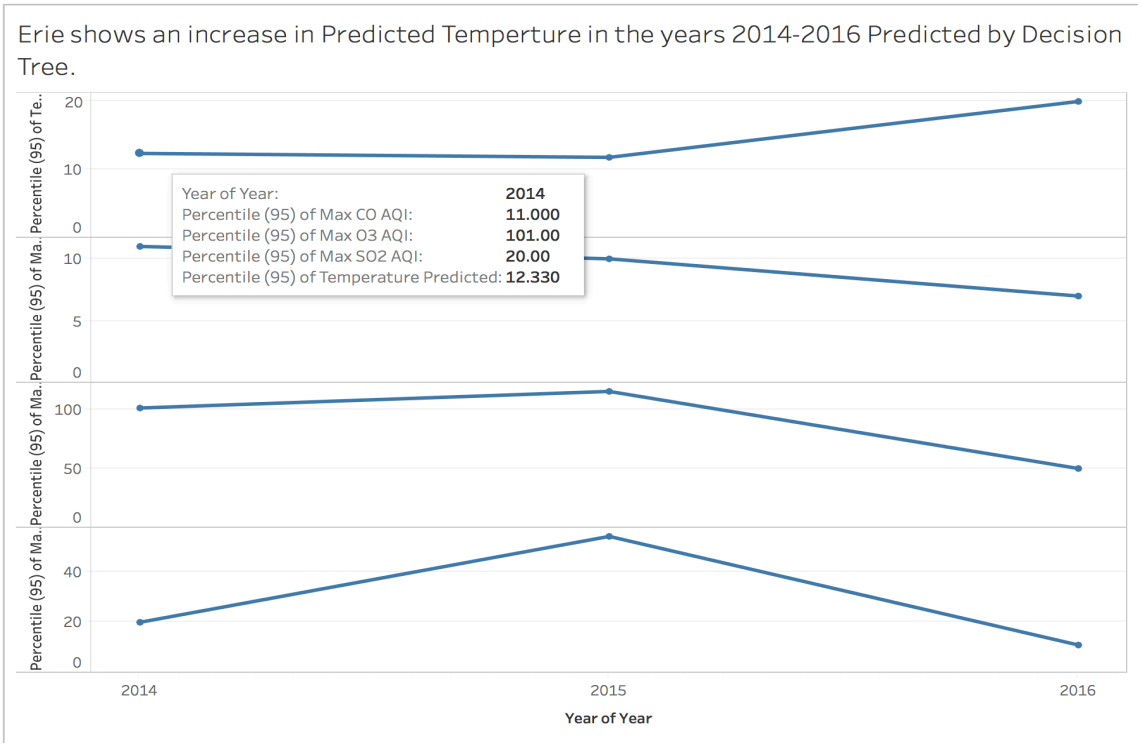
	Year	City	Max_NO2_AQI	Max_SO2_AQI	Max_CO_AQI	Max_O3_AQI	Temperature_Predicted
0	2014-01-01	Albuquerque	43.0	92.0	15.0	87.0	13.949487
1	2014-01-01	Arden-Arcade	41.0	7.0	19.0	122.0	15.797569
2	2014-01-02	Austin	34.0	10.0	5.0	97.0	19.923520
3	2014-01-01	Baton Rouge	55.0	54.0	47.0	129.0	11.556602
4	2014-01-01	Beltsville	40.0	20.0	10.0	101.0	26.339146
5	2014-01-01	Bethel Island	31.0	14.0	8.0	101.0	19.923520
6	2014-01-01	Birmingham	82.0	97.0	11.0	90.0	11.067763
7	2014-05-02	Blaine	37.0	11.0	8.0	93.0	19.923520
8	2014-01-01	Boston	60.0	40.0	13.0	54.0	13.949487
9	2014-01-01	Bristol	50.0	20.0	16.0	84.0	18.436983
10	2014-01-01	Burbank	71.0	6.0	34.0	129.0	18.675667
11	2014-01-01	Calexico	93.0	16.0	43.0	151.0	17.036500
12	2014-01-01	Camden	62.0	14.0	17.0	115.0	18.436983

Fig-1.7 The table of predicted temperature values of different cities.



City	Year of Year													
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Athens										18.52				
Austin													21.56	2
Bakersfield	16.50	16.90												
Baton Rouge	21.09	20.86	20.88	20.63	21.06	21.25	21.53	21.25	20.93	21.23	20.33	21.17	21.81	2
Berkeley								14.94	15.05	15.02	14.67			
Birmingham														1
Boston		8.91	9.02	7.78	8.09	8.56	9.48	8.36	8.53	8.07	9.58	9.12	10.06	1
Bristol	9.97	9.69	10.28	10.25	10.27	10.24	10.58	10.37	9.82	9.90	8.80	10.63	9.55	
Burbank	16.64	16.47	16.43	16.94	16.55	16.43	16.62	16.70	17.01	16.68	15.89	15.87	17.09	1
Charleroi	11.37	10.94	11.44	11.20	10.94	11.16	11.48	11.52	10.90	11.00	9.74	11.69	10.68	1
Charlotte	16.43	16.86	17.02	16.28	16.78	16.71	17.04	17.48	16.66	16.42	16.52	17.26	17.64	1
Chicago	11.09													
Chula Vista	16.90	16.37	16.37	17.10	16.84	16.93	17.23	16.80	17.21	17.03	16.19			
Cincinnati												12.63	13.52	1
Cleveland												11.04	11.97	1
Concord	15.02	15.25	15.00	15.43	15.37	15.17	15.02	14.94	15.05	15.02	14.67	14.50	15.05	1
Costa Mesa	16.90	16.37	16.37	17.10	16.84	16.93	17.23	16.80	17.21	17.03	16.19	16.26	17.20	1
Dallas	18.90	18.55	18.19	18.55	18.58	19.13	19.85	18.42	18.52	18.42	18.69	19.69	19.99	2
Denver						10.06	9.96	9.48	9.09				10.92	1
Des Moines								11.21	9.50	10.01	10.72	10.94	12.91	1
Detroit	9.34	10.29	10.01	8.73	9.26	9.73	10.43							
El Paso	16.89	16.70	16.63	16.94	15.99	16.61	16.70	16.25	16.11	16.58	16.30	17.05	17.30	1
Fontana									17.99	17.75	17.11	16.97	18.44	1
Fresno				16.96				16.49	16.72	16.60	15.89	15.63	16.98	1
Grand Rapi..			9.81	8.80	9.26	9.76	10.31	9.85						
Hampton											16.30	17.22	17.37	1

Fig – 1.8 This chart shows the city wise temperatures predicted by the Decision Tree for years 2000-2013. There is an average stemperature increase by 2 degrees as seen.



Above is the Temperature vs pollution graph of an outlier Erie, which shows temperature increase despite the pollution levels decreasing.

Visualizations

To plot the table of the top polluted cities based on their AQI levels on a graph using SparkR, we first pivoted the tabled based on the cities, with their yearly AQI values as rows.

City Yearly	Burbank	Calexico	El Paso	Long Beach	Los Angeles	New York	Philadelphia	San Diego	Washington	Welby
2000	113.0	118.0	99.0	108.0	111.0	101.0	95.0	104.0	109.0	109.0
2001	129.0	108.0	101.0	105.0	112.0	102.0	87.0	110.0	88.0	102.0
2002	131.0	108.0	111.0	106.0	112.0	102.0	85.0	106.0	101.0	NaN
2003	108.0	111.0	100.0	107.0	113.0	105.0	100.0	110.0	101.0	108.0
2004	105.0	102.0	81.0	105.0	112.0	109.0	101.0	106.0	104.0	103.0
2005	88.0	107.0	82.0	108.0	106.0	91.0	85.0	103.0	104.0	96.0
2006	101.0	101.0	103.0	101.0	103.0	91.0	102.0	97.0	104.0	89.0
2007	86.0	102.0	89.0	102.0	102.0	101.0	87.0	101.0	81.0	103.0
2008	102.0	110.0	88.0	106.0	105.0	82.0	NaN	105.0	101.0	84.0
2009	87.0	101.0	NaN	103.0	104.0	95.0	NaN	90.0	82.0	83.0
2010	81.0	NaN	NaN	104.0	88.0	102.0	106.0	82.0	85.0	NaN
2011	NaN	106.0	NaN	102.0	103.0	NaN	86.0	NaN	NaN	NaN
2012	NaN	90.0	NaN	97.0	NaN	NaN	NaN	NaN	NaN	NaN
2013	NaN	111.0	NaN	NaN	89.0	NaN	NaN	NaN	NaN	87.0
2014	NaN	93.0	NaN	107.0	86.0	NaN	NaN	NaN	NaN	102.0
2015	NaN	82.0	93.0	NaN	86.0	97.0	NaN	NaN	NaN	NaN
2016	NaN	83.0	NaN	NaN	NaN	88.0	NaN	NaN	NaN	NaN

We had to first remove the NaN numbers after pivoting, We replaced the missing values with average of the Pollution levels for that city for that years. Then plotted the graph.

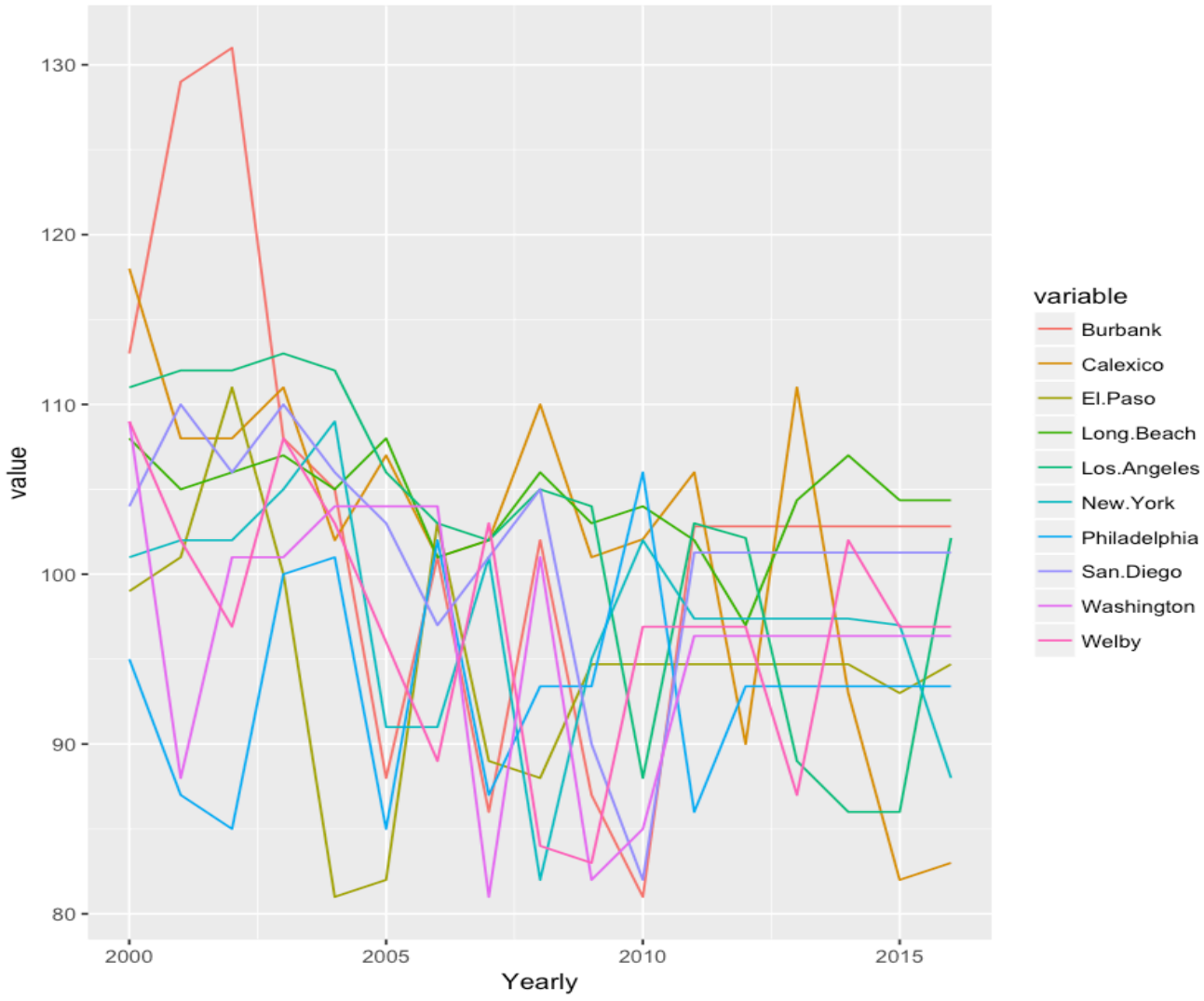
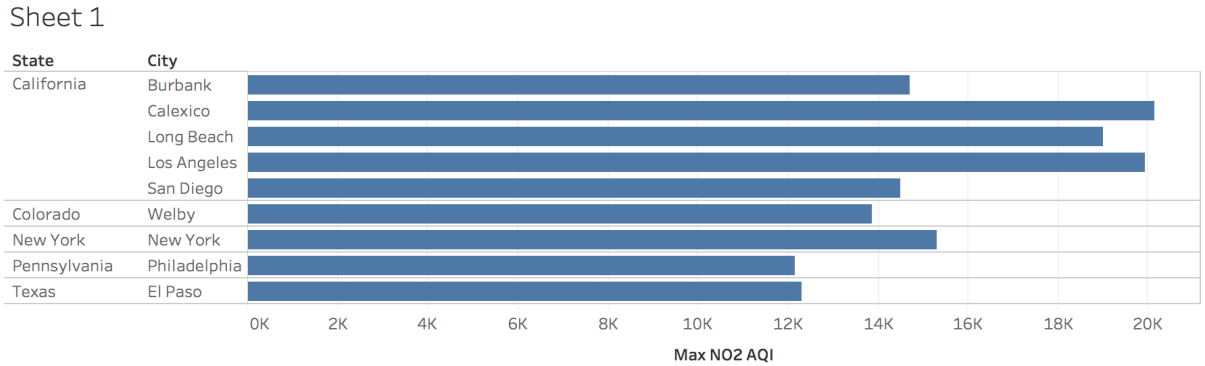


Fig- 1.9 Graph showing the AQI levels of different cities for years ranging 2000-2015.



TEMPERATURE-POLLUTION ANALYSIS

We joined the temperature data with the pollution data and extracted the maximum yearly NO2, SO2, CO and O3 values per city along with their average annual temperatures. We combined the deaths dataset and the temperature dataset of the top 10 polluted cities and visualized them on tableau.

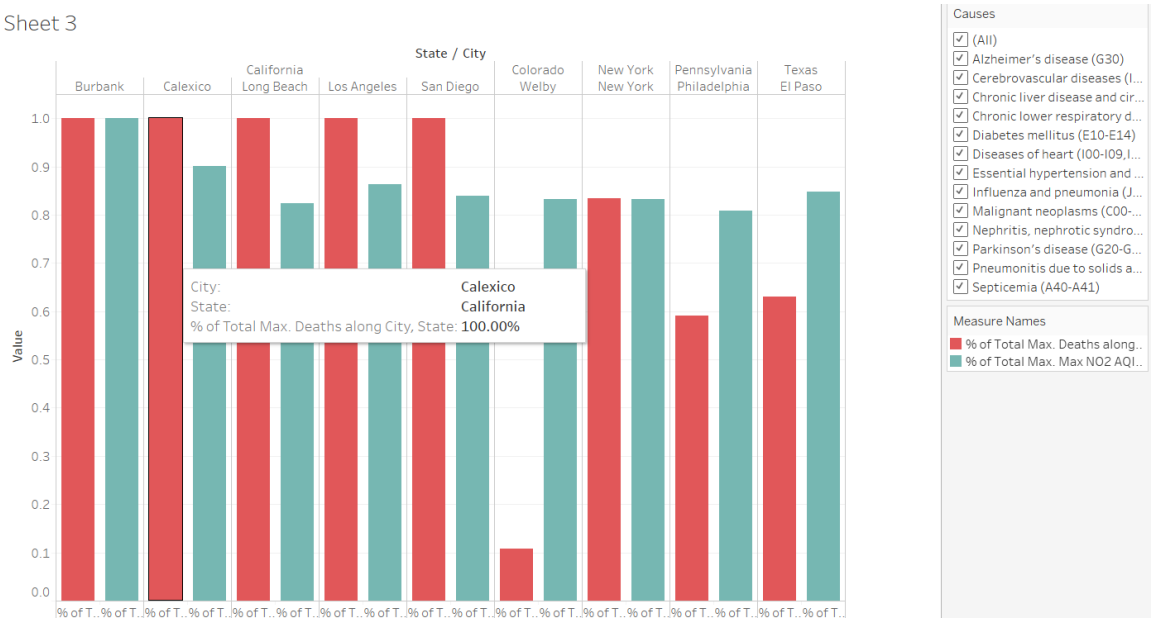
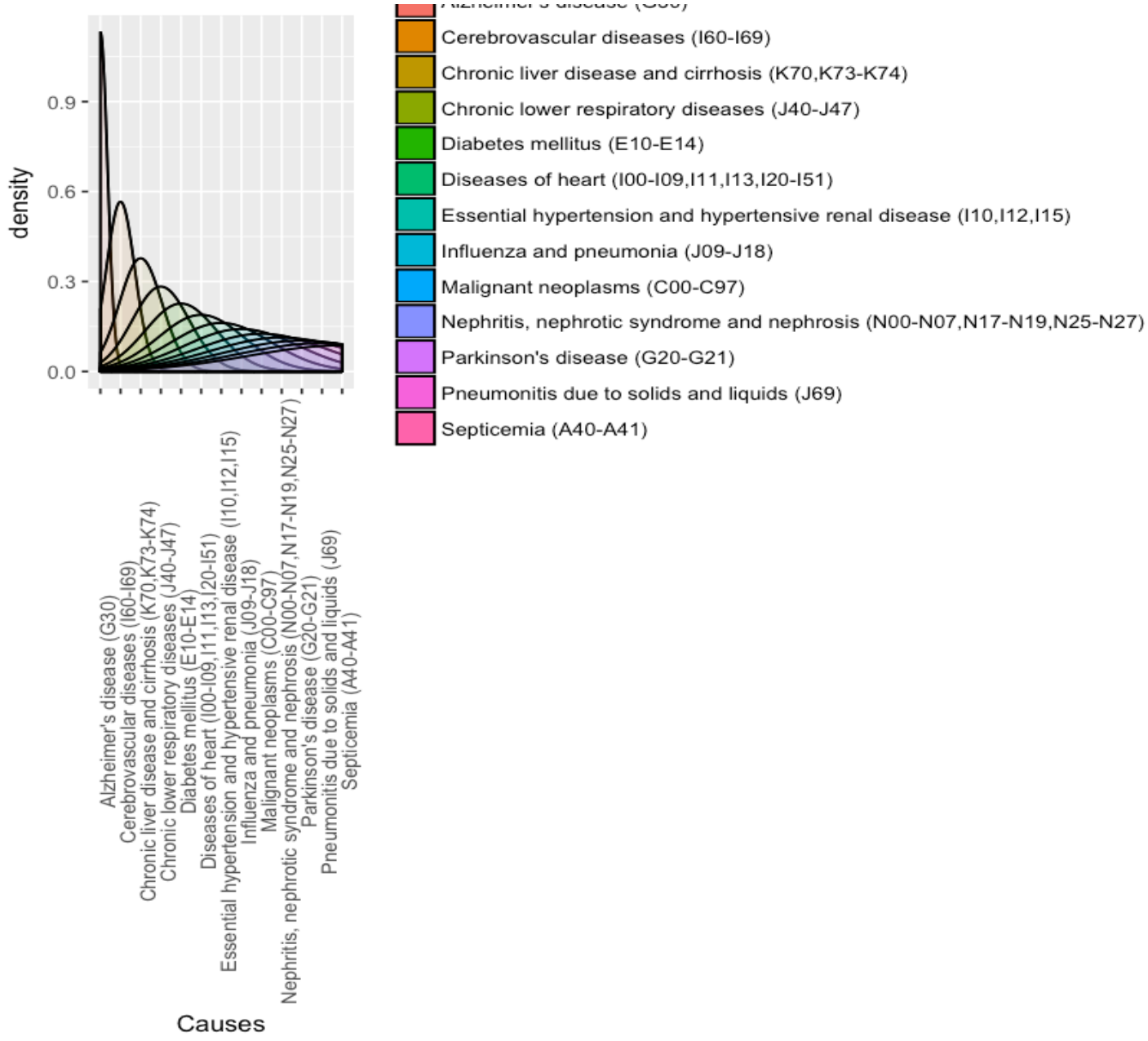


Fig-2.0 Deaths and NO2 AQI levels of top 10 polluted cities.
This is an interactive graph and when a particular disease is selected, the corresponding pollution levels can also be mapped.


density
0.9
0.6
0.3
0.0
Alzheimer's disease (G30)
Cerebrovascular diseases (I60-I69)
Chronic liver disease and cirrhosis (K70,K73-K74)
Chronic lower respiratory diseases (J40-J47)
Diabetes mellitus (E10-E14)
Diseases of heart (I00-I09,I11,I13,I20-I51)
Essential hypertension and hypertensive renal disease (I10,I12,I15)
Influenza and pneumonia (J09-J18)
Malignant neoplasms (C00-C97)
Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)
Parkinson's disease (G20-G21)
Pneumonitis due to solids and liquids (J69)
Septicemia (A40-A41)
Causes
Cerebrovascular diseases (I60-I69)
Chronic liver disease and cirrhosis (K70,K73-K74)
Chronic lower respiratory diseases (J40-J47)
Diabetes mellitus (E10-E14)
Diseases of heart (I00-I09,I11,I13,I20-I51)
Essential hypertension and hypertensive renal disease (I10,I12,I15)
Influenza and pneumonia (J09-J18)
Malignant neoplasms (C00-C97)
Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)
Parkinson's disease (G20-G21)
Pneumonitis due to solids and liquids (J69)
Septicemia (A40-A41)

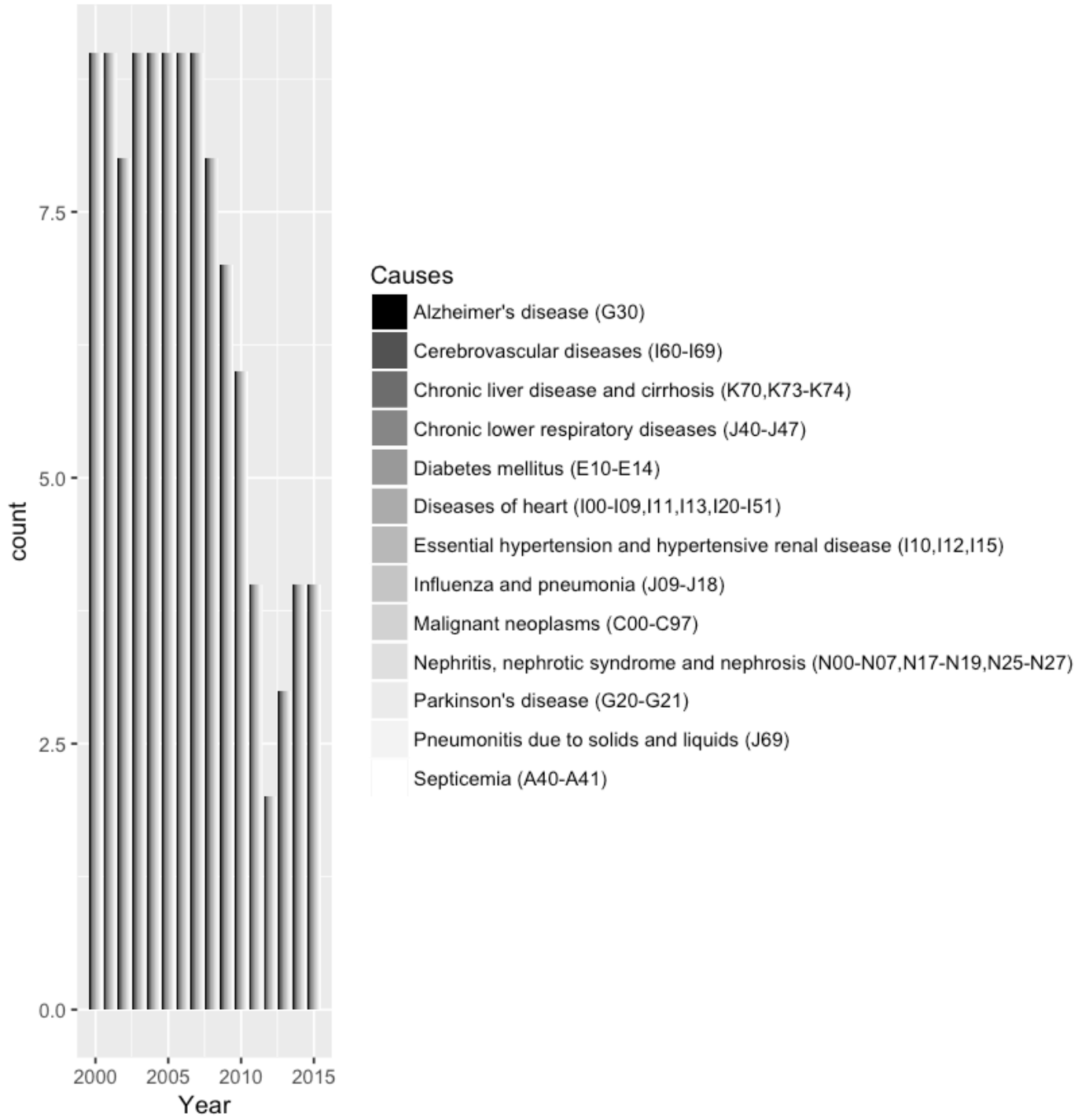


Fig-2.1 Figure showing the year wise death rates based on different diseases.

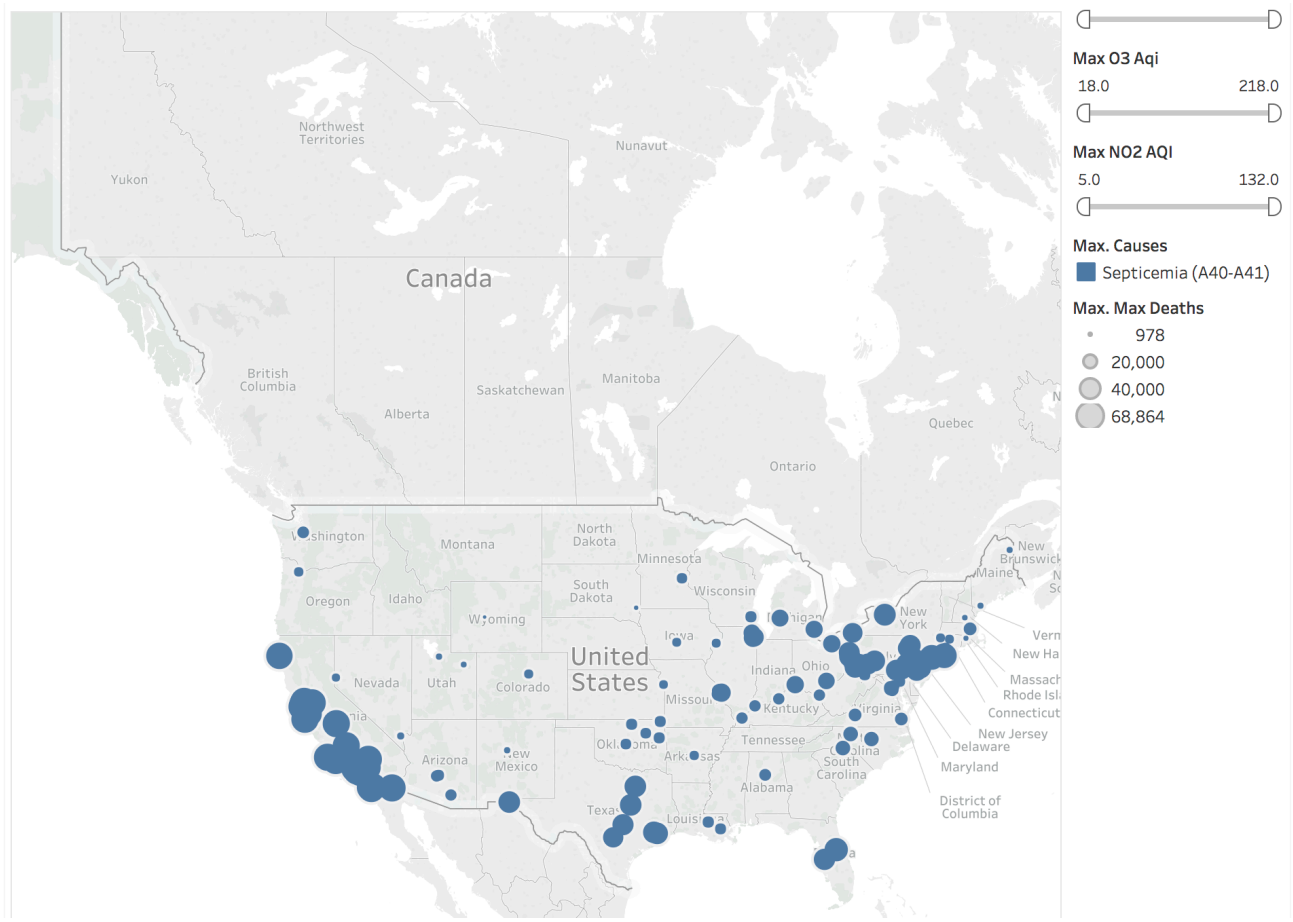
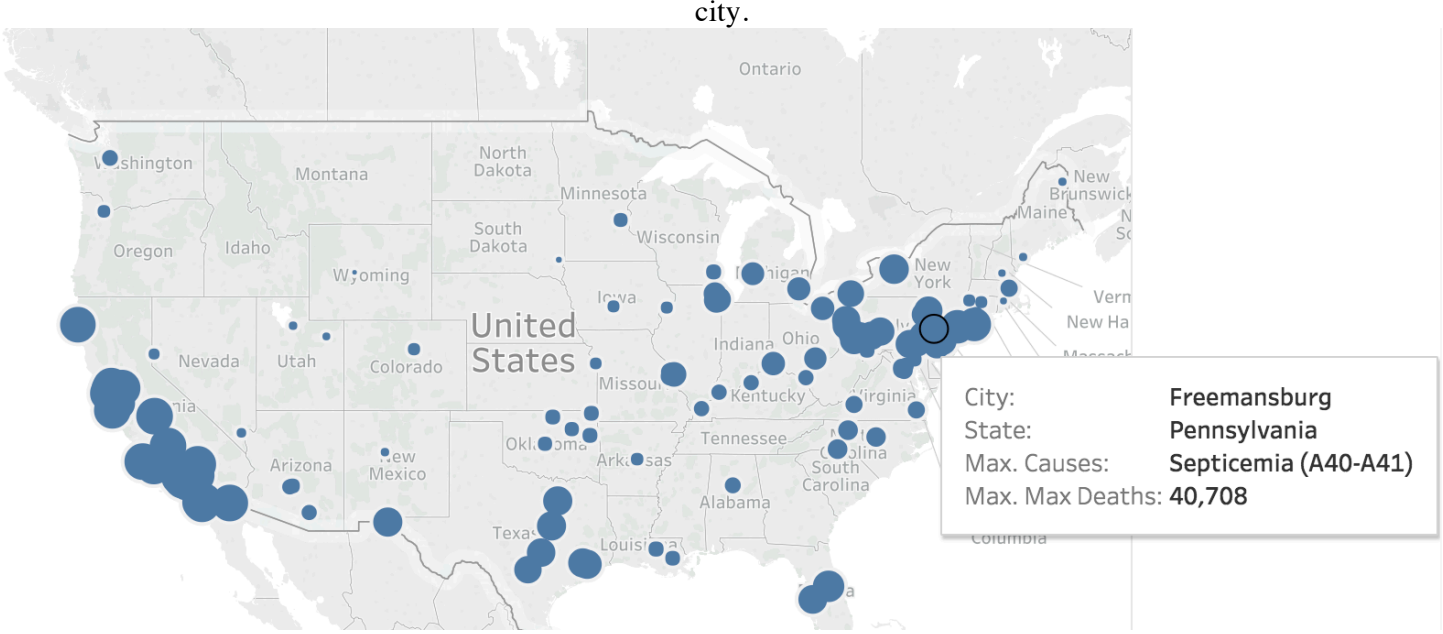


Fig- 2.2 This is an interactive graph showing the different cities in our dataset.
Hovering over the graph, one can see the Pollution levels, also the deaths caused due to a particular disease in that city.



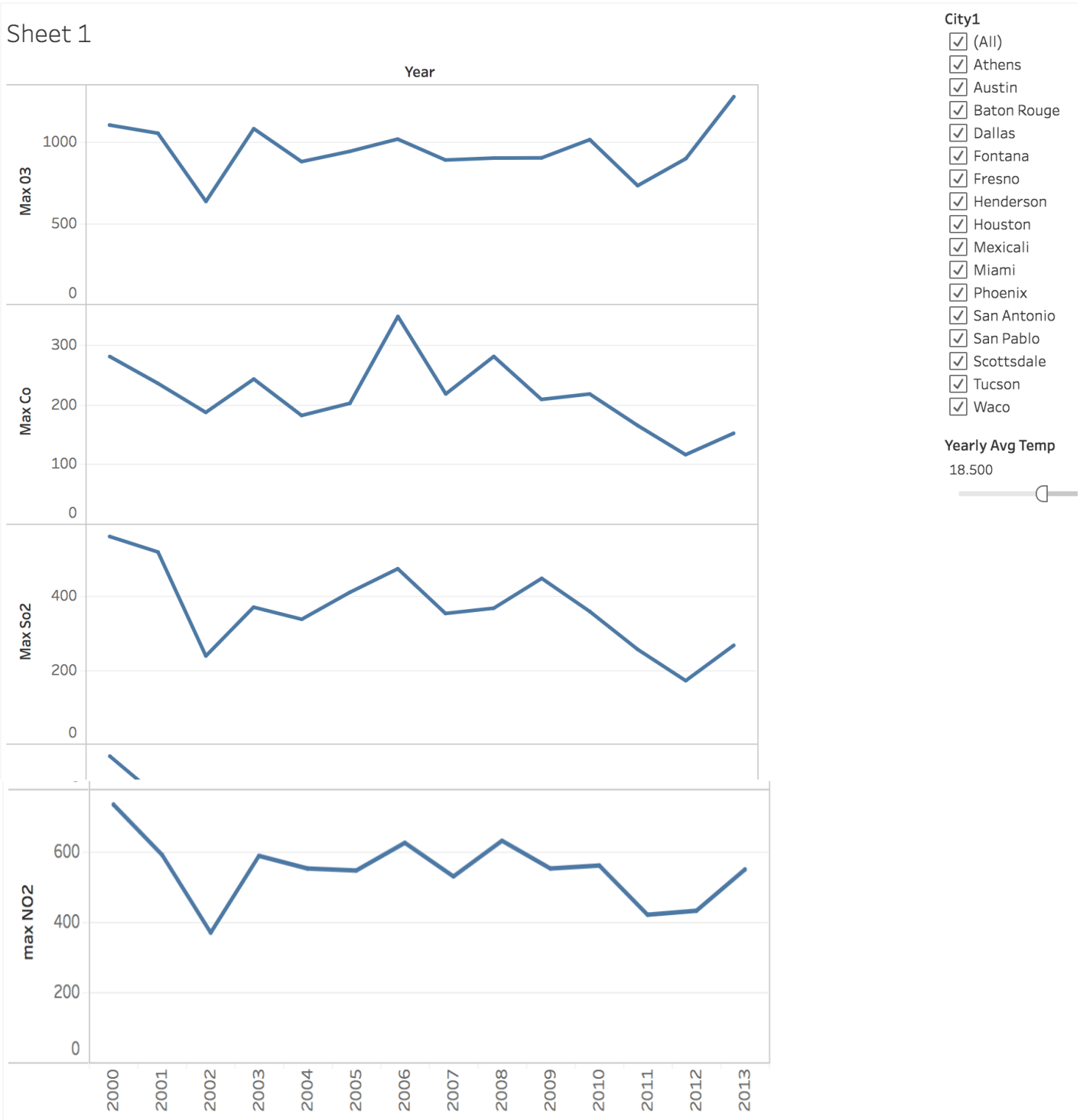


Fig- 2.3 This graph shows the pollution levels changing in past 13 years. Users can select a city, they wish to observe and check the levels of pollution progression.

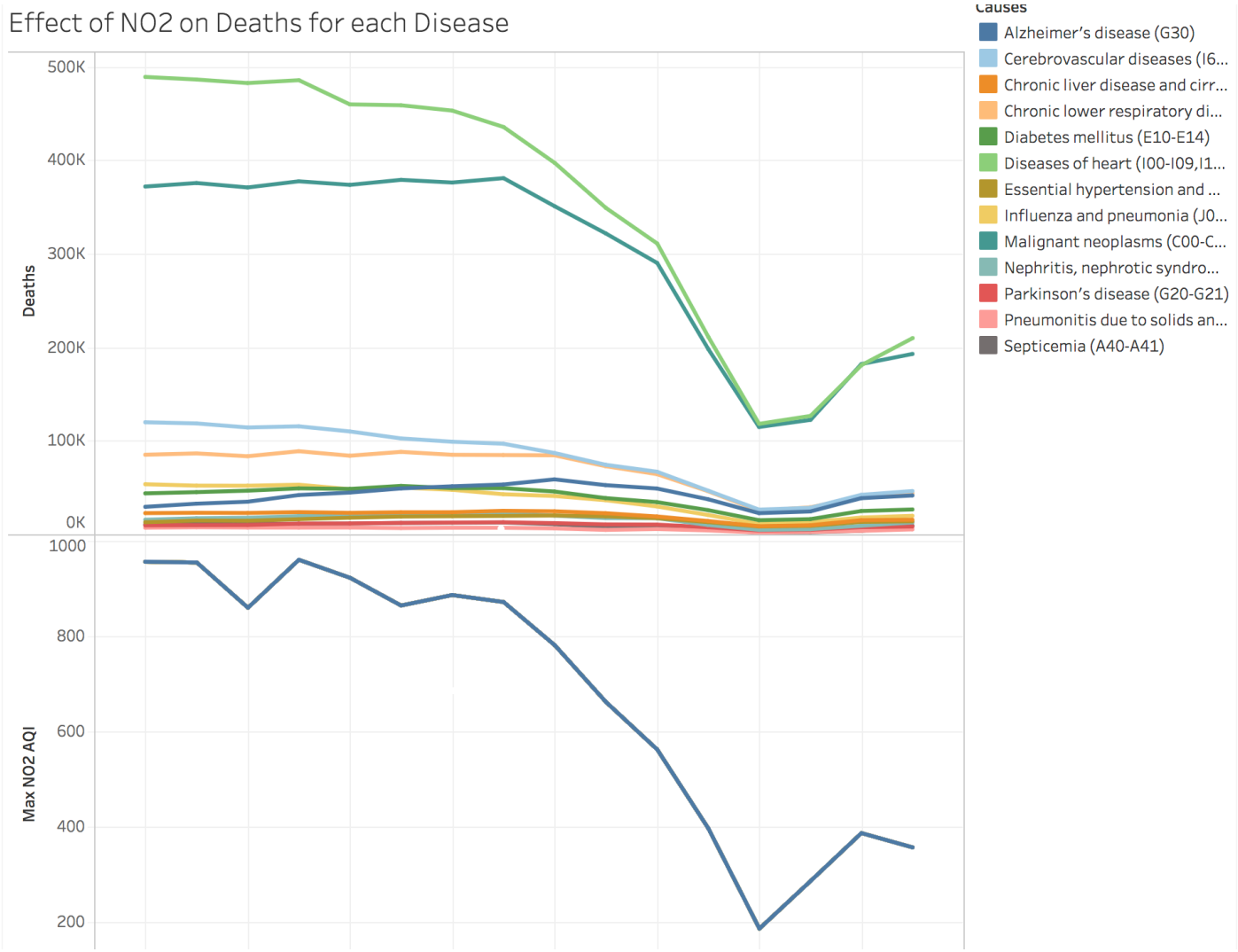


Figure showing the correlation between the deaths due to a particular disease and the pollution levels of gases.

TWITTER SENTIMENT ANALYSIS

Disruptive climate events are often reflected in social media. Monitoring changes in these sentiments during and after the disasters can reveal awareness of people with respect to climate change and mental health. We have used **ensemble algorithm with multiple models like Logistic Regression, BernoulliNB Classifier, Naïve Bayes, Linear SVC and synthesized their results into a single score in order to compare and improve the accuracy of predictive analytics.** Based on the keyword **"Climate Change"** we have calculated the sentiment values of each tweet that contains the keyword and plotted the real-time sentiment analysis of the tweet.

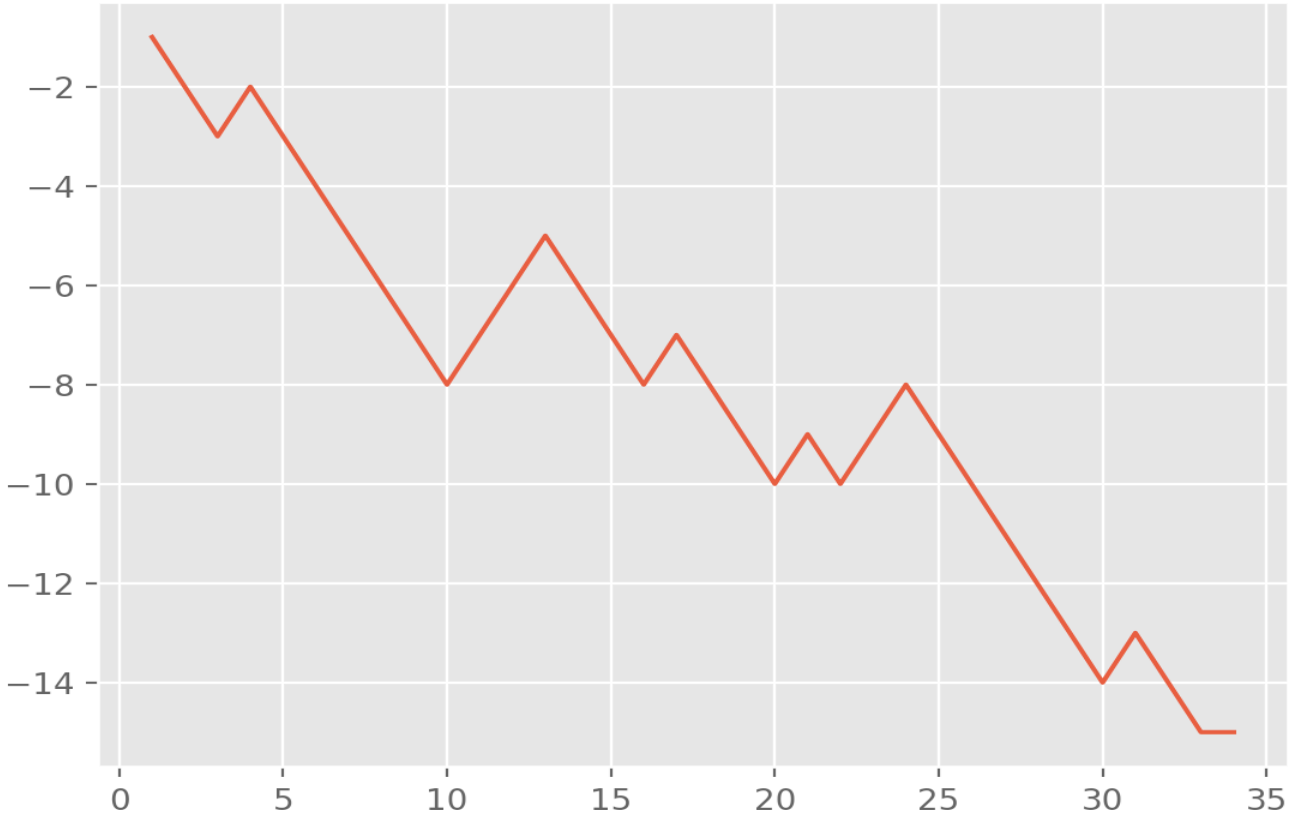


Fig- 2.2 Figure showing the graph of the positive vs negative tweets.

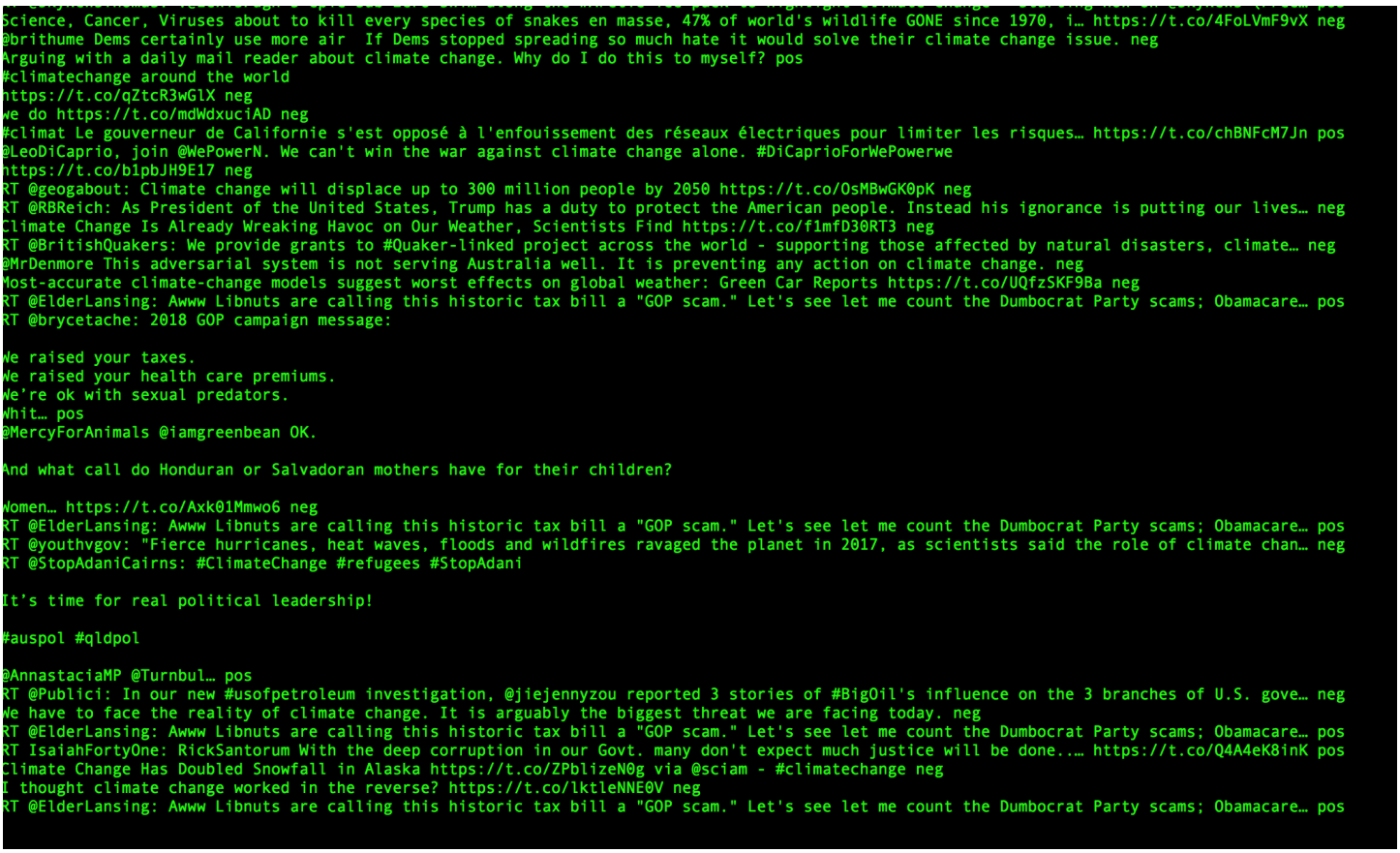


Fig-2.3 Live Tweets coming in real time on the terminal

CONCLUSION

- We learnt how to integrate Jupyter notebook with Spark, PySpark and R on Windows. Data Cleaning and identifying the most informative attributes was an important part to start our project.

- In a nutshell, we analyzed the impact of increasing pollution levels on temperature change and the subsequent impact on death counts for certain diseases.

- There was a strong correlation between the pollution levels, temperature increase and death counts for certain cities and diseases, especially for Cardiovascular diseases, however, there were a few outliers for which temperature and deaths decrease even though pollution levels increased.

- We also identified that the average increase in temperature over the last 4 years was about 2C, indicating the need for imminent action to curb the rising temperatures.

- Additionally, our Twitter sentiment analysis revealed that there is an increasing awareness about climate change among Twitter users.

- Our website: http://www.vikrambajaj.me/bigdata