



# U.S. Pollution Data

Zhixing Chang

Peng Yan

Xiaoyang Dou

Tianlun Zhao

# Description

This dataset deals with pollution in the U.S. Pollution in the U.S. has been well documented by the U.S. This data include a total of 28 fields, four major pollutants (Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide and Ozone) for every day from 2000 - 2016.



# Research Question

What is the main source that cause air pollution in different city?

How does the trend of air pollution during the day?

What are the main pollutants in each city?

which cities' pollution is getting better?

which cities' is getting worse?





## Prior Work

Nick Brooks searched the California pollution by U.S. pollution data in the U.S. since 2000. Megan Risdal analysis the U.S. pollution dataset to finished her professional research.

Jaeyoon Park was using the U.S. pollution data to make animation, basemap, plotly for Air Qulity Index.

Aleksey Bilogur provide several datasets website about air pollution.

- " (<https://www.kaggle.com/jihyeseo/seoulairreport>
- " <https://www.kaggle.com/epa/hazardous-air-pollutants>
- " <https://www.kaggle.com/shrutibhargava94/india-air-quality-data>
- " <https://www.kaggle.com/epa/air-quality>

and the U.S. pollution data to do his graduation research.



# Datasets:

## U.S. Pollution Data

**Pollution in the U.S. since 2000**

<https://www.kaggle.com/sogun3/uspollution>

We already download it in each member's computer.

# Proposed work

Data clean: Take out outlier to ensure the accuracy of the data

Miss value

Duplicate datas

Data preprocessing: Specific time for different data.

Data integration: Data visualization

Analyse the trend

Data reduction: Separate the whole database into small parts to speed up mining





# List of tool(s)

1. Python 3
2. Weka
3. Excel
4. Tableau (data visualization)



## Evaluate:

Base on our results we can determine if there is a correlation between different pollutants.

We are also going to make some trend graphs to see how the pollution increase or decrease and what happened to the certain city makes the pollution increase or decrease.

Using this result we could guide different city use different method reduce pollution.

