Air pollutant emission

Data

1. Data source

There many online database about air pollution, built by environmental protection bureau and public affair organizations. Since I am most interested in air pollutant emissions in European countries, I obtained the air emission data by accounts from 37 countries from here, including some countries outside Europa. This dataset includes emission data from 2011 to 2018. Emissions from different sources, such as industry and land use and so on, and the amount of each air pollutants are also contained in the dataset. The information of all listed counties, like country area and population, are produced with the package countryinfo.

2. Data cleaning

First, data are downloaded from the website and uploaded to the IBM cloud. While reading in data, repeated and meaningless columns are dropped. I acquire the total emission of all air pollutants for each year and for each country to discuss the emission trend in all countries and to categorize these countries. Missing values in certain years are replace by mean values from other years. From this step, countries are clustered based on their average emission over years.

Second, country emission data and their geographical data are combined to visualize the clusters of countries on a map.

Third, emission data and countryinfo data are merged to get average emission per million people (“density_pop”) and per square kilometer (“density_area”). By sort the data and apply linear regression, relations between emisison and land/population can be revealed.

By comparing emission among countries and among years, which country are emitting more air pollutants can be intuitively observed. And whether emission trend during these years in single countries is changing can also be judged. The use of the linear regression model will give us hints on how to slow down climate change with contributions from our own.

In the end, the Foursquare location data will be used to help me out if I would like to present my analysis in schools as an appeal to the young to make contribution to protect our environment and climate, for example in Cologne. It would be helpful as well for me to how to present my analysis to different levels of pupils or high school students.