Prediction for 2018 U.S. News College Ranking

Yi Le

The problem definition and motivation

The air pollution in China has become extreme recent years, with PM2.5 as an important index to represent air quality. The PM2.5 changes every day in each city, which can be regarded as a Markov Chain. The aim of this project is to predict the future trend of PM2.5 in five main cities in China and reveal the latent variables that influence air pollution.

Methods

R is the only language that will be implemented. Several packages will be involved including mcmc and matplotlib.

The link to the data

Data source:
http://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities

This hourly data set contains the PM2.5 data in Beijing, Shanghai, Guangzhou, Chengdu and Shenyang. Meanwhile, meteorological data for each city are also included.

Team Member

I formed a group individually, so all work will be done by myself, including data collecting and cleansing, data analysis, visualization and writing final report.