STAT 306
Group Project Proposal: Beijing PM2.5 data

Presented to:
Professor Bruce Dunham

Date Submitted: November 4, 2022

Group B1:

Zejun Su - 72603681
Pontus Sjostrand - 90888793
Yulu Duan - 23366081
Yifan Wu - 99677320

- **The source of the data**

The data set contains PM2.5 data from the US embassy in Beijing and meteorological data from Beijing International airport. The data time period is from Jan 1st, 2010 to Dec 31st, 2014. Source of data is from Songxi, Chen, csx@gsm.pku.edu.cn, Guanghua School of Management, Center for Statistical Science, Peking University.

- **A brief description of the variables measured**

In the Beijing PM2.5 dataset, we use pm2.5 which is the PM2.5 concentration measured in micrograms (one-millionth of a gram) per cubic meter of air (ug/m^3) as the response variable. TEMP, the temperature measured in Celsius (℃). PRES, the pressure measured in hectopascal (*hPa*). Iws, the cumulated wind speed measured in meters per second (m/s). These three variables mentioned above are our explanatory variables. All these explanatory and response variables were measured between Jan 1st, 2010 to Dec 31st, 2014 in US Embassy in Beijing and Beijing International airport.

- **Motivation behind the analysis of the data**

It is of interest to predict air quality based on variables which are easy to measure. If we find a relationship between measured air quality (in this case PM 2.5), temperature, air pressure and wind speed, we can predict air quality based on metrics that are easy and cheap to measure. The need for an expensive air quality monitor would become less important and recommendations to people regarding air quality could be made even without a specific air quality measurement device if a clear relationship between the explanatory variables and the response variable can be observed.

- **Responsibilities undertaken by each team member for final report**

Yulu Duan	R coding and data analysis
Pontus Sjostrand	Main responsibility for the conclusion part of the report. Translates the results from R output into words.
Zejun Su	R coding and data analysis
Yifan Wu	Introduction and general structure of the report.