Case Study 2

Due Friday, November 10th by 11:59 PM

Problem description

Does pollution kill people? Data in one early study designed to explore this issue came from 60 cities in the United States from 1959-1961.(Data from McDonald, G.C. and Ayers, J.A. (1978). Some Applications of the "Chernoff Faces": A Technique for Graphically Representing Multivariate Data in Wang, P.C.C. (ed.) Graphical Representation of Multivariate Data, Academic Press.). Total age age-adjusted mortality from all causes, in deaths per 100,000 residents, is the response variable. The explanatory variables include relative pollution potential of oxides of nitrogen and sulfur dioxide. Relative pollution potential is the produce of the tons emitted per day per square kilometer and a factor correcting for SMA dimension and exposure. Other socioeconomic indicators are included in the data as well, included the percentage of the population that is nonwhite, median number of school years completed, and mean annual precipitation. **Is there evidence that mortality is associated with either of the pollution variables, after the effects of climate and socioeconomic variables are accounted for?** For this case study, you must include both a model that only includes the measures of pollution, and one that incorporates a larger set of socioeconomic and climate variables (it is up to you to determine which you want to include). Analyze the data and write a report of the findings, including any important limitations of the study.[1]

As a reminder, the case study should be 1-2 pages and only include details to answer the question at hand. Details such as condition checks, decisions to transform the dependent variable, and other pieces of information should be included in the appendix at the end of the report. All tables must be nicely formatted. The code below will load the data for the case study.

```
# Loading the Statistical Sleuth data library
library(Sleuth2)
library(tidyverse)
# Reading in and saving the data
mort <- Sleuth2::ex1123
```

[1]This exercise was borrowed from *The statistical sleuth: a course in methods of data analysis.*