Proposal

I have always been fascinated by the power of statistical and mathematical models. It is truly amazing that everything — from the expected sales of a store to the future of the earth we are living on — can be described by mathematical models. I feel extremely blessed that I am one of the few people that not only can appreciate the beauty of it but also be able to have a career based on it. Speaking of the future of the earth, more precisely, the future of human beings, we are actually at of critical point of it as we can still prevent the collapse of humans if we start to take some serious actions now. We, as statisticians, can also contribute to saving humans by developing good statistical models to describe and predict ecological issues.

Ozone, as a natural substance, plays an important role in environmental sustainability. It can be used as a water disinfectant and natural pesticide in agriculture. However, ground-level ozone can have a harmful effect on humans since it is mostly formed of "fog". Therefore, for this project, I decided to study the distribution of ground-level ozone in the last ten years (2011-2020) in the United States. The aim of this project is to build a multilevel model with ground-level ozone score being the response and the time and date ozone scores were measured corresponding temperature, and measurement locations being predictors to predict the future distribution of ground-level ozone in the United States.

The data used for analysis are ground-level ozone at more than 85 rural sites throughout the United States collected by United Environmental Protection Agency (EPA) from 2010 to 2020. All the datasets were downloaded at <https://java.epa.gov/castnet/clearsession.do>.

**Project Timeline:**

EDA: Nov 4th - Nov 8th

Data processing: Nov 9th - Nov 12th

Modeling and Validation: Nov 14th - Nov 20th

Write up: Nov 21th - Nov 30th