**7/1/21 Research Meeting with Prof. Antonelli**

* Particulate matter PM­­2.5
  + Estimates of air pollution everywhere
  + Data across all US
  + EPA has air pollution monitors spread out
    - Sort of limited in the number of them that there are
  + What we *want* to know is the *surface of pollution*
    - Satellites that orbit earth
    - Send down a beam of light
    - Based on how much light disperses, you can use that to predict amount of air pollution
  + Get that exposure data, learn some things about it
  + Aim: how the particles vary across the US
    - Understand how exposure profiles differ across the US
    - Distinct groups of profiles of pollution; want to be able to identify these things
    - Need to understand spatial variability over time in this air pollution
* Second project: looking at aggregate zip-code-level outcomes
  + How many deaths in particular zip codes? Is this associated with pollution profiles for that region?
  + Belief: people don’t stay in their zip code
    - Work in other zip codes, people move around
    - Hypothesis: by looking at just zip code level exposure, we might underestimate the impact of pollution on health
  + Construct two notions of air pollution exposure:
    - 1) pollution at your zip code
    - 2) some kind of weighted average of nearby zip codes
      * How do you weight them? Distance?
      * Commuter data from census info

**8/16/21**

* “Exposure prediction”
* “aerosol optical depth” (what satellite data measures)
* Step 1: Getting the data (publicly available at <http://fizz.phys.dal.ca/~atmos/martin/?page_id=140>)
  + May be massive files; can start with a specific region
  + See how big it is, try to visualize it and get some insights
* Long-term:
  + As we make progress, he needs to get me access to the UF cluster of “remote computing”
    - Cloud computing for some of the more computationally heavy computing that I will need to do
* Download some small part of the dataset, try and working with it in R
  + Start with PM, pick a year, throw it into R, try to visualize it
  + Plot a heat-map on top of the continental US with PM in some year
  + If that’s easy (and computationally easy), try and do it for all of the exposures (individual components of PM)
  + If that’s easy too, start looking at how the data varies over time (i.e. seasonality)
* Long-term end goal
  + This research aims to figure out: what would happen if we changed the air pollution mixture
    - Mixture meaning: not just how much pollution, but what it is made up of
    - You can have same pollution level but the components of the pollution differ; certain mixtures might be better/worse for health than others
  + Try to figure out
    - Pollution has been ever-declining since the Clean Air act in the late 20th century to lower pollution levels
    - They set these standards for long-term pollution levels
    - PM2.5 is measured in microns per cubic meter; acceptable level is 12 microns per cubic meter
    - Research is finding there are still health problems below this ‘acceptable’ benchmark of 12
      * There may not even be a true level that is ‘acceptable’
    - Huge public health impact: premature death, increase in respiratory related hospitalizations
    - It is getting harder and harder to lower air pollution; maybe we can just change what makes up the pollution instead of focusing on lower levels
      * Not all air pollution mixtures are possible
      * Try to find realistic “shifts” in the pollution make-up for a given region, to lessen the public health impact
      * Interventions that are realistic and possible, and evaluate the extent to which they would lead to improvements in health
  + I am working on discovering what these interventions are: variety of ways to deal with it
    - Start by just getting to know the data and visualizing it
    - Getting on the cloud computing resource to host the data and run the computations on more powerful computers
    - Eventually get to know which interventions are possible
      * Look at how things change over time
      * Could also look at specific interventions; shut off power plants, installations that change their emissions, doing something with automobiles, etc.
      * Clustering regions of the country to determine areas where a certain intervention may be most effective