**Research Meetings with Prof. Antonelli**

**7/1/21**

* 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

**9/1/21**

* Scales sometimes too wide / too narrow
* Hard code file names
* Before next meeting
  + Try to go to terminal
  + Figure out how to sign into HiperGator
  + ssh [nwibert@hpg.rc.ufl.edu](mailto:nwibert@hpg.rc.ufl.edu)
  + password: gatorlink password
  + cd /blue/jantonelli/nwibert
  + just make sure that you can sign on, exit
* for computations, keep values in matrices
* save data yearly
  + same way I have it now, but make a list for each one
  + for example, a list called “PM\_2000” has three columns: longitude, latitude, and PM, all for the year 2000
  + eventually we will get monthly data
  + in remote server, make directory structure like “PM/2000/annual”, “PM/2000/January”
* start R environment: module load R
* q()
* use scp to get files into the remote server
* either into my own folder, or to the shared folder
* checked and verified!!

**USING ftp IN LINUX**

* monthly files are stored in an ftp server (file transfer protocol)
* <ftp://stetson.phys.dal.ca/Aaron/V4NA02/>
* If you paste this into file explorer, you can look at the files manually
* How do we get them all into HiPerGator? Manually downloading everything and moving it all is a bit tedious
* Using ftp command, can access this same directory within linux terminal
* ftp stetson.phys.dal.ca
  + when asked for username, type ‘anonymous’
  + don’t enter a password
  + once in the server, run command ‘pass’ to enter passive mode
  + cd into /Aaron/V4NA02
    - separate folders for every component (BC, SO4, etc.)
    - each of those folders has data for every month of every year (a lot)
    - once in a component directory, use command ‘mget \*.nc’ to download everything (you will have to confirm one at a time)
    - they will be downloaded to the directory you were in when you launched ftp
      * so, create directory for component and navigate into it before launching ftp

**9/29/21**

* how pollution mixtures change temporally
* over months, years
  + are the clusters/centers changing over time
* plot:
  + for a given period of time, you have pollution clusters + their mean values
  + want to see what those clusters represent in terms of pollution mixture
  + heatmap that shows this
    - row for each component, 8th row for total pollution
    - (add up 7 components too, compare “total level of pollution” across areas)
    - Column for each cluster
    - Use cluster centers to get values
    - Legend is continuous colors
* Library(maps); library(sp)
  + Test = map(database=”usa”)
  + Point.in.polygon(point.x, point.y, pol.x=test$x, pol.y=test$y)
  + Try and get the point in polygon working
  + Get rid of those excess points outside of the USA
* Pick 2 years that are far apart (2000, 2015)
* Plot a single component in both years
* Grant
  + Main goal: try to understand how different exposures within PM2.5 affect health
  + Figure out the “shift” to reduce public health impact
  + What components are most harmful / would have the greatest impact if levels reduced
  + What components would be the most viable/practical/realistic to shift
  + What kinds of industries / pollution generators should be targeted