Project Notes/Log

Gift and Kunal

Research Questions

* Is it safe for gardening?
* Is infrastructure or point source contributing more? If so, how? Why? What can we do about it?
  + Where do we put our energy?
* Not too interested in differences between contaminants.
* How interested are we in comparing mining and urban
* We are trying to make a generalizable conclusion, studying individual community dynamics is not as useful

Priority action Items

* Continue testing pli models (KP)
* How to handle missing data (KP)
* Cf models (Gift)
* Play around with multiple metals in a model, starting with two or three (KP and GC)
* Figure out how to represent, understand, communicate complex interactions (GC)
* Add pH and EC into the models
* Figure out how to look at multiple rwhi questions simultaneously – maybe just compare p-values
* Look at dimensional reduction methods – PCA to figure out which metals to focus on? (KP)

Regular action items

* MFA with hds info?
* Add pH and EC into all analysis

Double check the following

* Al summary stats, some values might actually be lead stats
  + Why was period:season taken out of the model? It was a better fit, but then period was visualized later?
* Do the control samples fall into PH sampling scheme? Do we exclude control samples that were collected after our project ended?

List of assumptions, data notes

Control Modeling options

* Put all the metals together?
  + Would the different metal scales fuck up the model interactions?
    - Metal:season and metal:community and metal:period and metal:period:season?
* Model PLI -> model cfactor comparing metals, with signif variables from pli -> individual concentration models with highest contributors -> HDS models
  + 1) PLI – narrow variables down, generalized pollution analysis
  + 2) Cfactor – assess contribution only
  + 3) Concentration – highest contributor analyte trends
  + 4) Concentration + HDS - highest contributor analyte trends
    - Do we need pli models for hds?
    - Can we add hds questions together? How do we handle missing hds data?
* IF we can analyze multiple HDS ?s in one PLI model, repeat above analysis
  + Model PLI -> model cfactor comparing metals, with signif variables from pli -> individual concentration models with highest contributors -> HDS models
    - 1) PLI + HDS – narrow variables down, generalized pollution analysis
    - 2) Cfactor + HDS – assess contribution only
    - 3) Concentration + HDS – highest contributor analyte trends
* How do we graph complex interactions?

HDS modeling

* How do we rank different hds questions?
  + Can we compare p-values to each other?
* Do we look at multiple metals at once for one hds question at a time?
* Add distance to point source to hds models? Which is more important?
* How do we get creative with HDS? Can we create an index?
  + For ex every home before 1960 is a minus 1, first flush is a plus 1, etc etc
  + Normalize for number of responses
  + But some questions are really hard – 1s and 0s and -1s could be dependent on analyte
  + Only variables that people realistically have control over?
  + Conservative, risk averse approach
    - Wherever there is a NA or missing value, we assume they are not following the best practice

Pull back a little bit

* Take top 5 average cfactor analytes by community, by season
  + Ranked differently, but perhaps we see trends? Cu in all the mining communities for example
  + And then go to infrastructure
* We look at PLI split by community or land use

Questions

* How to handle missing categorical data from home description survey? In order to attach to contaminant data.
  + Do we create an index of decision making? How? How to normalize? Assume NAs are 0s? Bootstrapping and imputation?
  + How do we assess home age and roof type then?
* Advice on the best way to pick metals based on contamination?
* How should we approach collinearity in the modeling?
* Do we model each community individually? Combine analytes?
* Do we need to use cfactor or pli for individual analyte analysis? Or concentrations?
* Do we need enrichment factors to help us understand pollution?