Questions:

* Meetings? Frequency? Lab meetings?
* Method for sharing papers/literature review notes?
* Coding languages, etc.?
* Method for sharing code? (GitHub? Travis?)

Major questions

* Pollutant (all criteria?)
* Sites
* Turn on/off
* Monitoring frequency (continuous? seasonal? etc.)
* Control for monitoring requirements?
* Control for modeled concentrations? (DUN DUN DUN) Actual measured concentrations?
* American community survey data

SES, correlated with industrialization, correlated with monitoring

How to tease apart industrialization with monitoring (per capita, monitoring less intensive where they live?) . . . but if they’re far from any source of pollution.

Correlation, monitoring station toward where the

Forget where monitors are, ask where emitters are, where people are, etc. . . . if we found that . . . relative to some

Emitters versus monitors, how do we calculate spatial correlation. Spatial correlation between emitters and monitors. . . . If we found that economic status was correlated with emitters more strongly than monitors are correlated with emitters.

Check on criteria pollutant emissions.

Total number of people, total number of people in current SES, if population density were uniform, spatially, we know that people don’t live near oil refinery, they live there out of economic necessity, notwithstanding. If we imagine that SES is correlated with emission . . . socially, what’s the right answer. More monitoring where there’s more emissions. Density of monitors that’s adequate. If we established a threshold, more people, not enough monitors, compared to what you think you need.

If we talk about actual air quality, problems that make that difficult. We do know where the monitors are, we kind of know where the people are, we know what the emissions are. Some idea of local winds, etc. If have emitter in the middle of a basin, trapping things, versus

Values recorded at the monitors . . . does the density of monitors tend to be lower where values are higher? . . . if we think that lower SES is going to be associated . . .

Two hypotheses:

* Industry is successful at keeping monitors out of places where monitors would find problems.
* Monitoring is lower where people with low SES live.

How to tease apart deliberately polluting where race/SES live versus trying to avoid getting caught polluting.

Look at public process on siting or removing a monitor

Methodology from underlying data, SES goes along with two different kinds of emitters

* Agricultural (industrial farms)
* Industrial (sites)

How to quantify

Tranche of SES, look at the average distance . . . average inverse distance between a human head and an emitter, compare that to what happens for different SES tranches.

Population not equal across census tracts, census block groups, etc., or headcount of people, of the entire Black population, what fraction lives within half a kilometer of an emitter

Look at how people are quantifying statistical significance . . .

Inverse weighted distance, inverse weighted distance to all emitters

First piece, association between where people live, doesn’t take into account travel

Elementary schools . . . for public schools in CA, list of headcounts, breakdown by race

Meal subsidies and proximity to polluters

Look at elementary school populations, etc. . . . geographic coordinates school.

Kids are spending their time in school . . . I like that, that’s where they are. There a big chunk of their year. Clearly what happens to them matters.

how close are the schools to emitters, how close are they to monitors, take all of Black children, look at the average distance between where they are, per capita average, inverse weighted distance to all the emitters, then white kids. Then on average white kids are further from emitters. Permutation test significance . . . 1 million, and 20 million people, null, random distribution, do that over and over. Nice nonparametric test. Monitors in there, distance to monitor, question then is . . . differential rate in how close they are to monitors, and then how close to emitters, Black kids are closer to emitters, by an amount that can’t be reasonably ascribed to chance.

130,000 rows of data, 8 or 10 rows, 25 rows per school. file format tab-delimited, straightforward. Aggregate the different ethnic groups and genders. So detailed and great.

Think about whether to limit to elementary schools versus high schools.

cde.ca.gov/ds/sd/sd/fsenr.asp

Nonparametric tests. Permutation tests. Python package for permutation tests.

github or Berkeley’s github instance . . . works primarily in Python, JavaScript-y.

R libraries. Python. They corrected random integer generation.

Stratified permutation tests in Python. Stratify by county. Within each school district, do Black kids tend to be closer than White kids, then compare in a different. To the extent to which you imagine that you can move kids around a school district but not across, could stratify on district.

Stationary sources CA PM2.5 emissions reporting.

TO DO:

* Download monitoring sites
* Download monitor data
* Download and format school data
* Geospatial analysis