Initial Analysis Documentation

Jon McCall
CP255: Urban Informatics and Visualization
Professor Sam Maurer

My work on my final python project this semester has been, as I imagine most research and modeling projects are, productive but frustratingly slow, full of data cleaning, and punctuated by lots of pausing to troubleshoot. The initial plan for the semester was composed of four parts:

1. To estimate the availability of housing affordable to households in different income brackets over time in the greater Bay Area.
2. To develop a model that incorporates household incomes as well as geographic location to estimate GHG emissions from VMT and household energy usage by
    a. Repeating the methodology of one of two precedent studies I identified in my literature review.
    b. Or by incorporating the outputs of the models developed in one of those studies without needing to recreate them myself.
3. To input the estimates from step one into the model in step two, thus estimating the changes in GHG emissions in the greater Bay Area that have resulted from shifts in the availability of housing affordable to low- and middle-income households.
4. To produce visualizations of the outputs of these models, for example bar graphs showing city and regional GHG emissions over time and interactive maps that would allow the viewer to see the amount of GHG emissions in different cities and counties year by year.

I had hoped that by this point in the semester I would have completed steps 1 – 3 and at least started step 4 in the above outline. This would have allowed me to submit some visualizations of greenhouse gas emissions estimates, even if they were very preliminary. However, thus far I have only been able to complete a draft version of step 1, and that was completed only this past week.

In addition to simply being overly optimistic about how much I could get done, this slowdown has been due to the challenges that I faced in cleaning and incorporating county-level data on area median incomes (so that I can, for example, answer the question “what is the maximum rent that an extremely low-income family of 4 in San Mateo County can afford?”). This challenge, as well as the iterative process of repeating my analyses for each year from 2012 – 2018, took significantly more time than I had anticipated. However, I am confident that my methodology makes sense and will allow me to have faith in the input calculations that I use eventually.

One lingering question I have before I move onto steps 2 – 4 of the outline above is how to deal with housing units that are owned outright, either because mortgages have been paid off or because the owners never had a mortgage on them. This accounts for about 7% of the households surveyed in the 2018 ACS PUMS data. These housing units are currently counted in my analysis as affordable to extremely low-income households, despite the fact that they are likely owned by people who do not fall into this category.