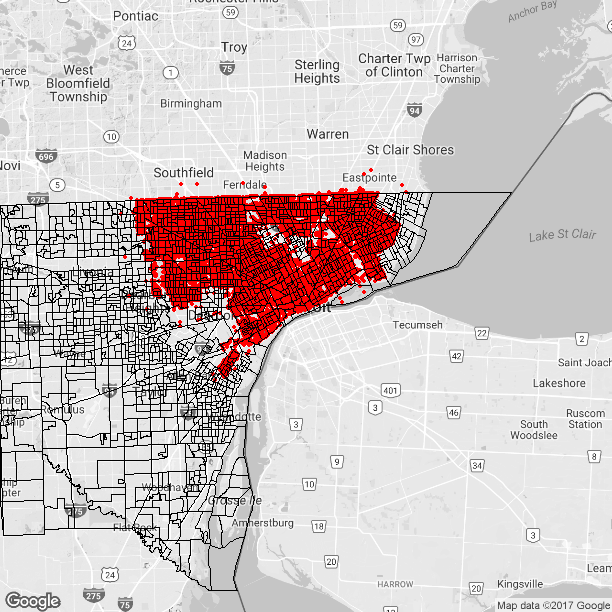
Claire Kelling

10/31 Progress Update

**1) Exploratory Data Analysis (EDA)-** I have gathered and cleaned most of the crime data in R. There are quite a few NA values for some variables of interest.

* There are 601,841 observations
* **Variables of possible interest** include Call Priority, Call Code, Call Description (longer version of Category, such as traffic stop), Category, Call Time (and date), Officer Initiated (yes/no), many measures for response times, longitude and latitude
* 5.9% of the variables have missing values for latitude longitude (I just removed these)
* The data is extremely dense- I could subset the data for a certain time period (overlapping with time period for ACS variables?) or a certain priority level



2) **ACS-**  I have pretty much completed my collection of ACS variables, for the variables that we have discussed. This turned out to be a bit of a challenge. I have collected the following variables- feel free to tell me if I should collect more! You can see the list of the available variables/tables from the ACS [here](https://censusreporter.org/topics/).

* Median income (as well as other income variables that aren’t as clear)
* Total population
* Employment
* Gender
* Age

3)**Spatial Modeling-**As we discussed in our first meeting, the plan is to start by creating a basic model and then see if there are any spatial trends in the residuals. I'm not quite sure how to do this and I'm hoping for some input. I'm more confident in doing so when there is some count data as a response on each point. For our data, we have response time for each coordinate, that we could model. However, I am thinking that this might be a point process model, which I am learning this week actually in my Spatial Statistics class. Unless you think I should model response time, which I have started to look at, I think I will hold off on the spatial modeling until after I am done with point process lectures in class.

4) **Categorical Data Analysis Class Project-** I am hoping to apply what I am learning for this project to my Categorical Data Analysis (STAT 544) class project. The professor, Dr. Le Bao, is supportive of this idea. I could use some ideas here- I am thinking that in order to deal with “count” data, I will need to aggregate so that it is areal data? Alternatively, I could wait until I have learned how to do a Log Gaussian Cox Point Process, which I learned today in class is somewhat similar to a Poisson GLMM (?). **I could also think that I could a classification problem in terms of the call priority?**

5) **Github**- I will use Github for my code, unless you see any reason not to do so. You can see my project folder [here](https://github.com/ckelling/bdss_igert_project).

That’s all for now! Please let me know if you have any questions, and I hope to hear your input on next steps.