This project started off straight forward. Simple API pulls, joining of data, and subsequent scatterplot visualizations and correlation analyses.

We started with the goal of gathering data at the lowest level of granularity that we could. Census data is incredibly flexible and allowed us to pull data down to the Census block level. The walk score was incredibly granular, so much so that adjacent buildings had different scores, but it did not scale up. Lastly, the FBI crime data did not have an API, but did have a website with data voluntarily submitted for 16 total cities and counties through Virginia. Unfortunately, that data was not any more granular than the city/county level. This left our group with crime data at the city/county level and census data that could be pulled at that level and the walk score at a very granular level. The challenge of this analysis was getting these data at a shared granularity so that we could move forward analysis.

We decided that the best solution was to select 100 points at random in each area that had FBI crime data to create area averages for our analysis, this way the common denominator in our data sets was the city/county level. We then ran the gleaned latitudes and longitudes through the walk score API. As you can see on the right-hand side of this slide, we used a swarmplot in the Seaborn library to visualize these scores for each of the city/county options to understand if the averages and spreads were congruent with what we thought these areas would be. Whereas it was not a perfect approach, heavily populated areas tended to have higher scores and rural areas tended to have lower scores, which made us feel confident moving forward with the analysis.

* Started off
* Goal was lowest granularity
  + Census – flexible
  + Walk score – granular
  + FBI – voluntary, 16, not granular
* Challenge – shared granularity
* Best solution – 100 random
* Gave averages – common denominator
* On the right – swarm plots
* Congruence with our thoughts
* Not perfect, but heavily pop = higher walk scores