**Charley Ferrari**

**IS 602 Semester Project Specifications**

**Data Dictionary:** [**http://www.nyc.gov/html/dcp/pdf/bytes/pluto\_datadictionary.pdf**](http://www.nyc.gov/html/dcp/pdf/bytes/pluto_datadictionary.pdf)

**Room to Grow: An Analysis of New York City’s PLUTO zoning database**

The PLUTO database is a database of every Tax Lot in New York City. The full database for all five boroughs contains 859,560 lots, with 82 variables. There are five main types of variables:

**Geographic:** Different Geographic definitions, including Census tract, zip code, etc.

**Building Info:** Information about the building, including type, number of units, year built, etc.

**Lot Info:** Information about the lot itself, such as its size.

**Zoning Info:** Zoning codes of the lot. Lots can have multiple zoning codes, so this is listed in a few ways. These roughly relate to the type of use: Residential, Commercial, etc.

**FAR Info:** FAR is the Floor-to-Area ratio of the lot. In order to calculate, you take the floor area of all floors and divide it by the lot area. For a building taking up an entire lot, a 6 floor building would have an FAR of 6. This is divided into different types, commercial, residential, etc.

Each lot has a unique ID composed of three variables: Tax Lot, Tax Block, and Borough. Each Tax Block in a Borough is unique and each Tax Lot in a Tax Block is unique. This combination of variables will be my smallest unit of analysis, and corresponds to a single unique line item in this dataset.

I’m mainly interested in residential data, so I will filter my data to only focus on R zoning districts. Depending on how the data is affected, I may analyze buildings that are only residential, or include mixed use buildings that contain some residences. Mixed use buildings will probably tend to be multi-family units, and not including them might skew my data.

Once I have filtered my data, my analysis will center on the FAR variables. Two variables in particular, Built Floor Area Ratio and Maximum Allowable Residential FAR, will allow me to calculate how much space is allowed to be developed in the lot. This is the variable I’m interested in comparing across these lots.

I plan on comparing this calculated non-built FAR to the other variables in the data to see if any patterns emerge. Geographically, I will probably focus on Census Tracts and Census Blocks. These are very fine grained units of geography (blocks roughly correspond to city blocks) and nest cleanly into each other. The same cannot be said for the other various districts (zip codes, city council districts, and school districts for example probably won’t align.)

I believe the geographic patterns will be the most interesting, but an example of a non-geographic hypothesis I’d like to test out is the effect of building type on the unbuilt FAR. Is a neighborhood of individually owned single family homes less likely to densify than a neighborhood of multifamily buildings for example? It’s unclear as of now if I can find out whether a multi-family building is rental, co-op, or condo (it doesn’t appear I can) but this would be another factor I’d like to test out.

I’m unclear at the moment of how I would integrate it, but I have two additional data sources I might want to add to this analysis. Zillow’s price data is the most promising, as they calculate a neighborhood level Zillow Home Value Index (ZHVI). I would have to research how Zillow defines their neighborhoods. If they do roll them up to Census tract or block, I would be able to add a neighborhood column to my data and join the two sources to find out what effect home prices have on unbuilt FAR (one would imagine higher prices would encourage more development.) I would also love to include a distance to subway metric, but am not sure if I could get exactly what I want from the google maps API.

**Visualization**

A few of the non-geographic variables would lend themselves to traditional data visualizations. A bar graph for example would be able to show types of homes and their average unbuilt FARs.

I would like to have a geographic component to this, but at the moment an unclear on how exactly I will accomplish this. At its crudest (but still effective,) I could graph the x and y coordinates as a scatter plot. I would probably have to bin the data to get a heat map, but something like this should give a rough idea of geographic clusters of unbuilt FAR. Hopefully, this should visually correspond to a map of New York City.

I imagine google maps would not be powerful enough for this sort of analysis. D3.js is another option, although it may prove to be too complex for a project at this level.

**Discussion**

This is the sort of project I can see growing in scope beyond this class. I’m interested in how cities evolve and grow, and this data could provide a very interesting view on how zoning restrictions may affect city growth. It has been argued that excessive zoning can actually cause suburban sprawl. There is an underlying demand for real estate, and if the urban areas of a city aren’t allowed to grow to accommodate this demand, it will spill into the suburban fringe. Looking at this aspect of the zoning data can show if this is really happening.

I also think that we’re in an interesting time when it comes to urban transportation. In the past few years we are beginning to get more decentralized modes of transport. Uber and bike share are two examples of transportation networks that are completely flexible. I believe that urban growth in the future will happen more organically, with these transportation networks growing along with the city. This sort of analysis is the beginnings of trying to quantify how this growth can occur.

If a patch of unbuilt FAR is found within biking distance of a subway station for example, a large developer could build denser buildings in conjunction with a small bike share network centered around the subway station. In this way, the area of urban land can be increased. Instead of assuming these far flung areas will be accessible by car, dense development could be supported. If a city is trying to promote sustainable development, this would be a great way to go about doing it.