**Kimberly Grauer- Final Project 2 write up – Housing and Vacancy Survey**

**Project Design Writeup**

* **Requirements:**
  + Well-articulated problem statement with "specific aim" and hypothesis, based on your lightning talk

The problem with housing prices and affordability in NYC is ubiquitous. Rent regulated housing is one way to approach the crisis in housing affordability. The problem is, the data is not well organized and difficult to understand. I want to look at the HVS from 2014 and see if any of the variables can be used as predictors for whether someone will or will not receive a rent subsidy.

* + An outline of any potential methods and models

I will create a dummy variable for receive and not receive rent stabilization. I will then select a handful of covariates and test, train the data to see if certain demographics can predict if someone will receive rent stailized housing. I will then ask, which demographics are the most likely to receive subsidies, and which types of subsidies are the most received.

* + Detailed explanation of extant data available (ie: build a data dictionary or link to pre-built data dictionaries)

Data is available but in a poor format. It is a text file and needs to be organized and sorted and a data dictionary needs to be created.

* + Describe any outstanding questions, assumptions, risks, caveats

Some major outstanding questions are whether anything will even come from this analysis. The questions I am asking is a major political issue and if it was possible to find anything insightful from publicly available data, I am sure someone would have already done so! Some caveats are the data is kind of old and I have not figured out a way to incorporate a time element to my analysis, which may be more interesting.

* + Demonstrate domain knowledge, including specific features or relevant benchmarks from similar projects

A similar project looked into how the absolute number of rent stabilized units has changed over time and in what parts of NYC have the largest changes occurred, down to the block and lot number.

The Furman Center frequently runs analysis on housing numbers, but there has been no public attempt to look at demographics or predictors.

* + Define your goals and criteria, in order to explain what success looks like

I would like to better understand the housing market in NYC and learn how to clean a highly unstructured database fluently. More specifically and relating to the project, I would like to see if there are underlying data trends that are good predictors of rent stabilization rates.