Cynthia Wong
Week 5 Assignment

1. Re-run all code above to ensure it works.

All codes work with the exception of the first section involving Staten Island due to datasheet not being included in the resources.

2. Replicate the omitted variable bias example inducing negative correlation,
   - rather than positive correlation. (Pay attention to Z1.)
   - Given the materials presented to date, how do you interpret your finding?

Given the material presented, I would say that there is no way to interpret my findings confidently. By leaving out relevant independent variables, we create biased results that may overstate or understate the data.

If I must go through the process, I would drop a X variable which would result in a larger coefficient value, but a negative correlation as the correlation between the variable with the omitted variable.

3. Zillow's Zestimate is a hedonic price algorithm used to predict US house prices.
   - Other than the features used above,
   - what other features would you consider relevant in determining US house prices?

The features used in the example included: house size, land size, age, and location (if house is in Todt Hill).

Other features that may be relevant when determining housing prices can include nominal inventory at time of sale [supply], proximity to transportation (by miles or minutes), vacancy rates at time of sale [within the submarket of the home], and transaction volume in the submarket in the given area. If we were also looking to target a specific demographic, I would also look into proximity to schools, bars, and other amenities.

4. In most circumstances, we seek to draw inferences about the effects of some type of policy or action.
   - Considering the ideas discussed in class regarding causality:
     Do opportunity zones actually result in beneficial development, and how could we evaluate potential benefits using data?

Ideally opportunity zones will result in economic development for low income census tracts a whole. While there is no clear way to determine if opportunity zones do indeed result in beneficial development. We can make an attempt to.

Causation is an indication that one event will have a direct impact on another while a correlation is an indication that values vary together. From our discussion, correlation is not causation. Thus, meaning that even if two values are correlated, one cannot infer that it can cause the other. A real-world example of such can be an example in which ice cream sales increased during summer as well as sales of sunglasses. Logically the increase in either are not caused by one another, but they will appear as if they are correlated on a scatter plot.

We can evaluate potential benefits of opportunity zone by measuring

- the pricing of the property before and after in the vicinity of the opportunity zone.

We should expect to see a pricing increase for all properties and not just the ones that receive a tax break.
We can also potentially track:

- if median wages are increasing in the particular census tract
- increases in employment numbers
- changes in demographic break out over time