# Overview

In this weeks post, we examine a perennial question: that of how our incomes determine our housing situation in the local housing markets.

# the Data

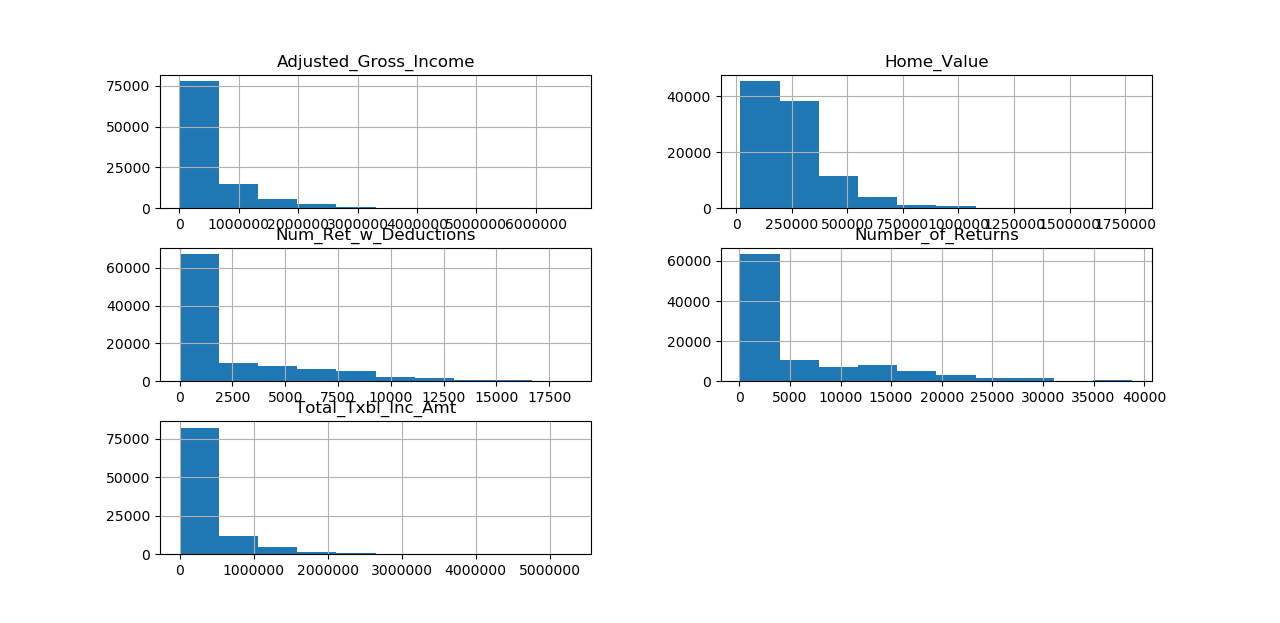
To answer this question, we used housing data from Zillow dating back to 1996, posted monthly. The Zillow Housing Index creates smoothed estimates of housing values over a period of time, and records these values monthly.

We then linked IRS tax data, filed yearly, using the zip codes for these areas. The IRS data pulled were from 2011 onward. The final dataset linked between 2011-2017, and covered the DC – Maryland – Virginia regions. In the words of a Tribe Called Quest: “DC – Maryland – Virginia, we got the vibes”.

# Results: forecast

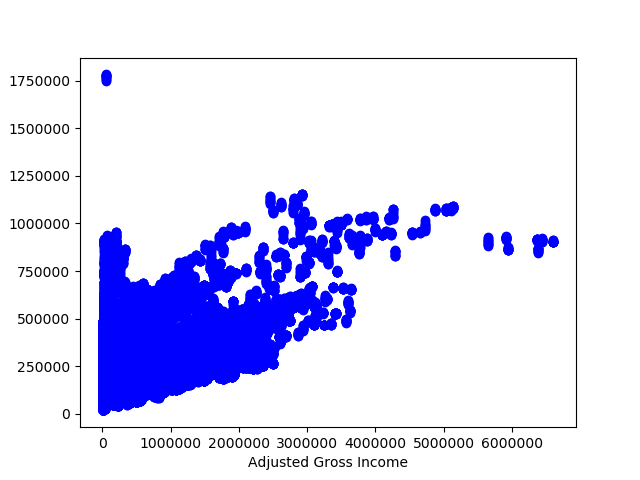
One would imagine that of our predictors used, income would be the biggest predictor of housing values. However, surprisingly, the strongest predictor of housing values in a given zip code (our unit of analysis), was the number of returns with deductions, and the number of returns with taxable income, with the latter being negatively correlated.

# Some graphs: EDA



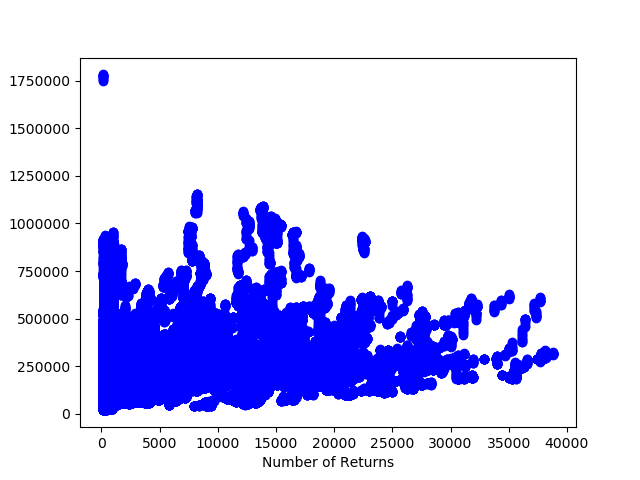
Our preliminary graphs show a strong right skew, with only home values showing a slightly less skewed distribution than the other metrics – AGI, Taxable Income Amount, Number of Returns, Number of Returns with Deductions.

# home values vs. income



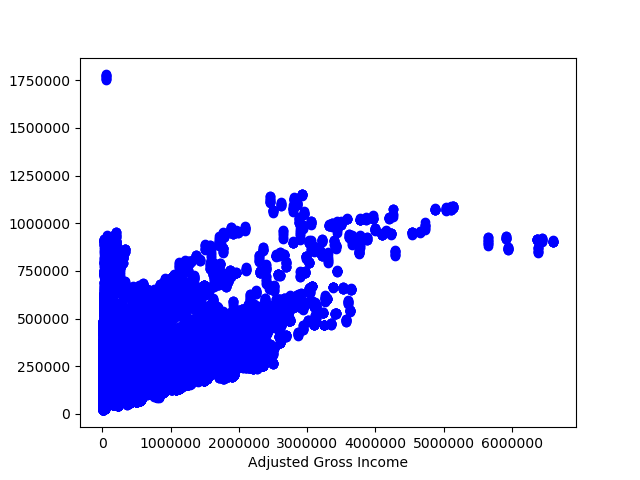
Inferences about the effect of adjusted gross income and of the number of taxpayers in a zip code can be similarly made from the provided scatter plots. Needless to say, the relations are overall positive. With a slight tapering off of values towards the upper ends, signaling a stronger relation between housing values and income at the lower end of the spectrum, and little effect of incomes on housing values at the upper end.

# home values vs. number of tax returns



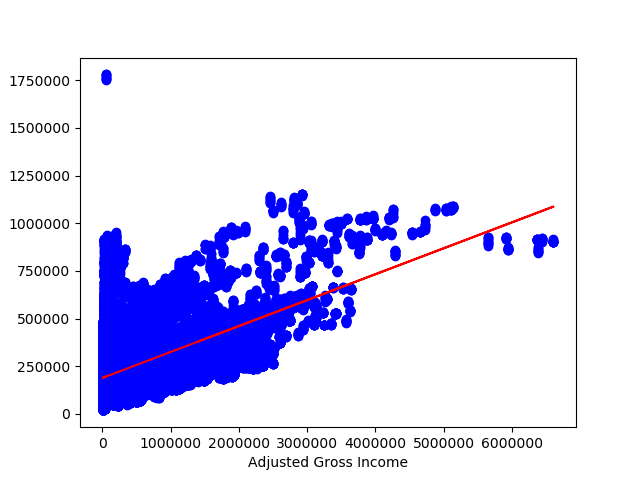
The relationship between home values and the number of tax returns filed in a neighborhood or a zip code is relatively flatter. There is an outlier or two, with an extremely high housing value in lower income neighborhoods, but it is not indicative of the overall trend.

# housing vs. income, training dataset



In this model, we used the scikit learn package, and Zillow’s monthly data to train our model. The above is the training dataset from a simple test-train split, and shows the same direct relationship between housing and income.

# housing vs. income, training data, line fitted



A simple model fit indicates the less than proportional relationship of income to housing prices in a neighborhood or a zip code. In other words, for a more than proportional increase in incomes in a neighborhood, the housing prices adjust less than proportionally. This is likely due to the lag it takes to establish a neighborhood.

# Regression model: housing values in a neighborhood, or IRS tax information by Zipcode



In what is probably the most impressive and surprising finding of our model fit, the Number of Returns with Deductions, and the Number of Returns with Taxable Income (last two coefficients shown above) have a more outsized effect than any of our predictors. Moreover, the number of Returns with Taxable Income correlate negatively to the housing values of a neighborhood. Total number of Returns, and Total Taxable Income Amount, while correlated, have a lesser effect than the other two variables mentioned above.

# Interpretation:

What does this say. That high tax deductions in a given neighborhood correlate strongly with the housing values in that neighborhood? Or that the number of returns with any taxable income strongly correlates negatively with the housing values of that neighborhood?

Maybe the overall total number of tax returns includes a number of duds, and empty tax returns filed for income. Whereas returns with any greater than 0 income shown, are ones with a taxable income.

In that case is it telling that people with a lower taxable income are living in the higher property vale neighborhoods? And that they receive higher tax deductions than other neighborhoods?

I think it is.

And I think it is a fascinating incentivization scheme, of getting people to become homeowners and property owners in more stable neighborhoods, or to realize their potential contribution to the US job and real estate market through appropriate property ownership.