Is gentrification a race phenomenon or a wealth phenomenon?

Citadel Datathon Oct 2020 – Team 9

Highlights

- Gentrification is difficult to define. We decided to define gentrification narrowly as an increase in a tract's median household price quantile.

- The features with the greatest explanatory power were *not* purely wealth-related. The Caucasian median income and the change in the proportion of graduate degree-holders were more important.

- Gentrification is a multifaceted problem and is not purely a wealth phenomenon.

What is gentrification?

Most definitions of gentrification posit that the influx of affluence causes home prices to increase, displacing the original poorer inhabitants of a neighbourhood. This definition is problematic because it encompasses several nodes in a causal chain:
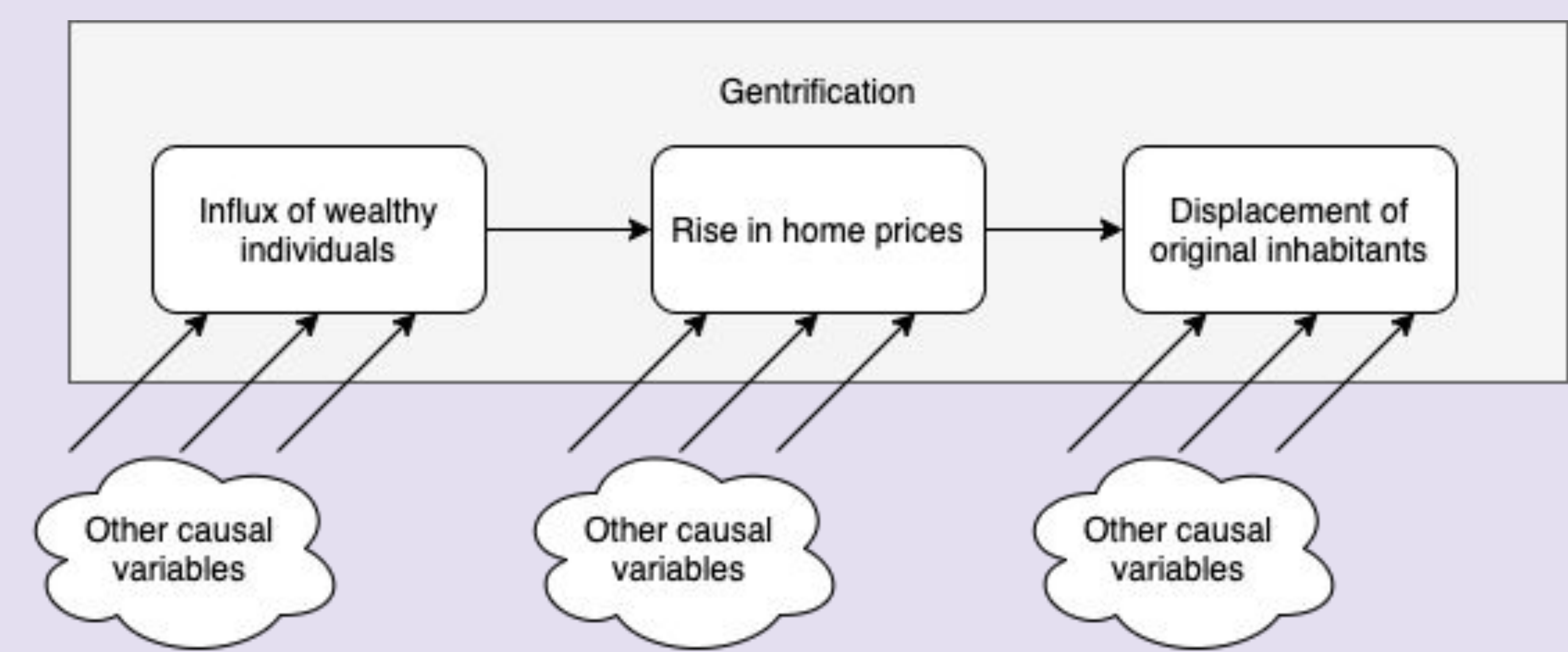
We decided to measure gentrification by the change in the quantile of the median house price. This saves us from having to include arbitrary parameters in the definition (which constitute additional degrees of freedom).

Data

We used tract data from the US Census Bureau, in particular: racial demographics, education levels, median household income (by race). We computed per-year quantiles, then used the change in these quantiles over a 10y period as features in our model.


Modelling

Our objective with modelling was explainability over predictivity. To that end, we began with a simple OLS model fit using both quantile changes and the initial quantile values of different demographic quantities as features. Select betas are shown below. The variables with the largest positive beta are the fraction of the tract population that have graduate degrees, and the Caucasian median income. This initially suggests that the wealth aspect of gentrification is not dominant.

Variable	Beta
Hispanic percentage of pop.	-0.26
Black percentage of pop.	-0.19
Graduate degree-holders	0.05
Caucasian median income	0.04

To gain a deeper understanding into the pertinent features and understand nonlinear interactions between features, we fit a Random Forest regressor to the dataset. After fitting the model, we computed Shapley values, allowing us to gain a more nuanced picture of how each feature contributes to the change in median house prices. The feature importances corroborated the results of the OLS model.
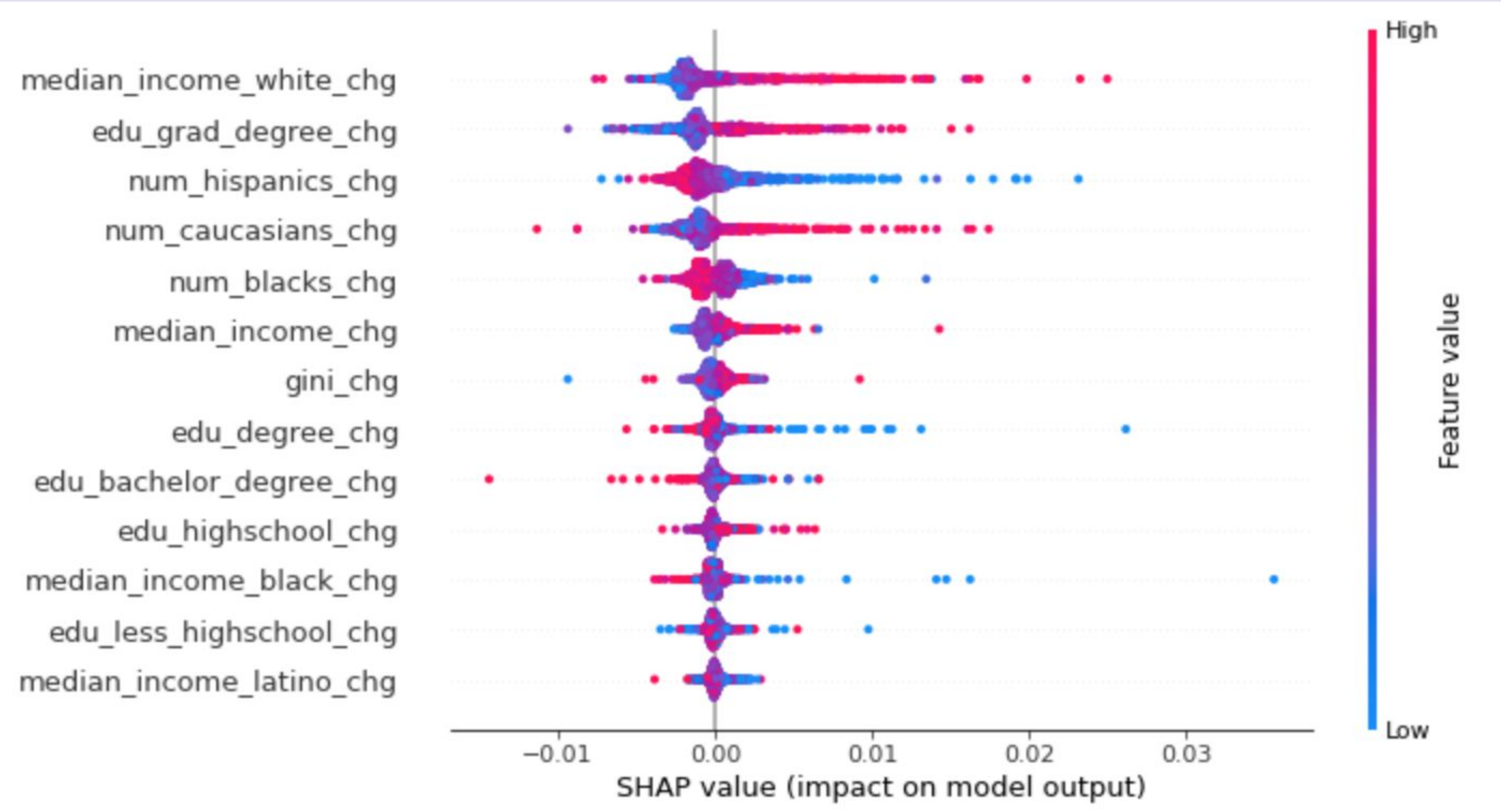
Results

Based on both the linear regression model and the Random Forest regressor, one of the most important features in determining gentrification, as measured by the change in the quantile of the median home values in a tract, is the change in the (quantile of the) median Caucasian income.
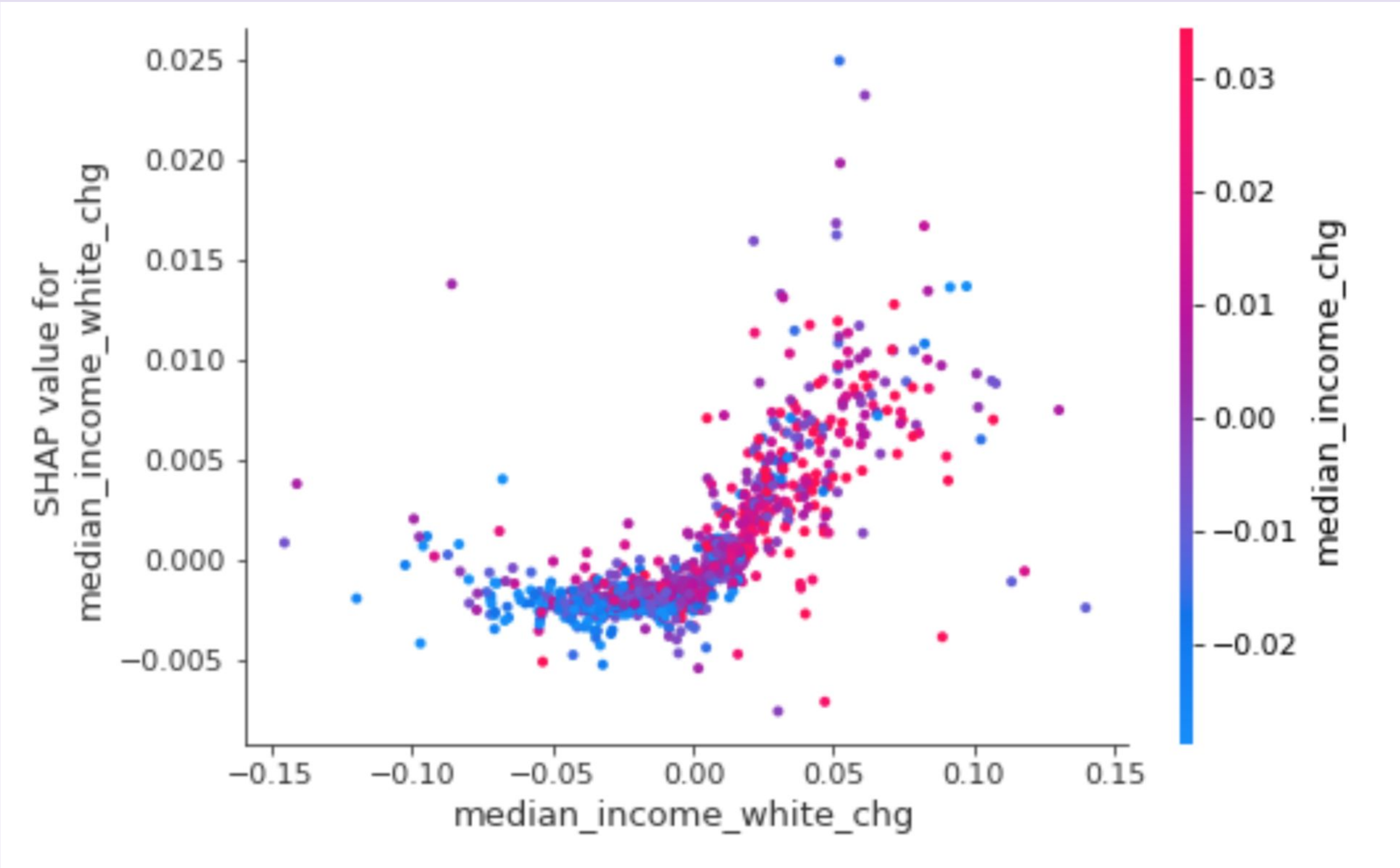
A rise in the level of affluence is *not* the key cause of gentrification. Race and education are at least as important.

This suggests that existing initiatives such as the Mayor of London's affordable housing scheme, may not be sufficient to prevent the displacement of individuals from neighbourhoods