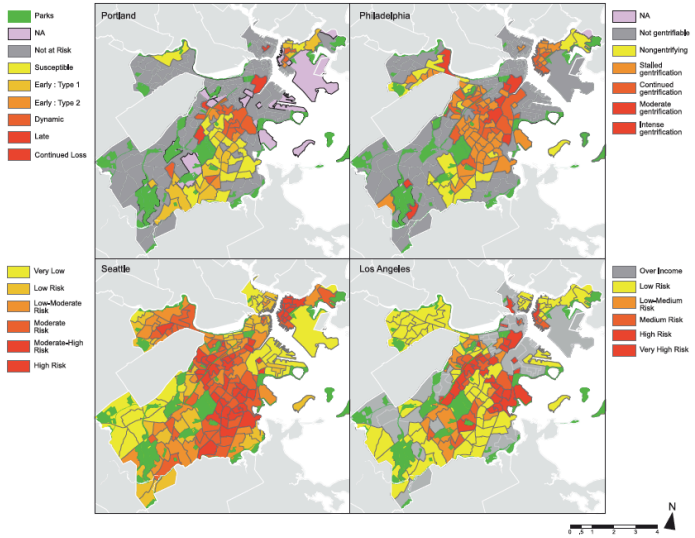
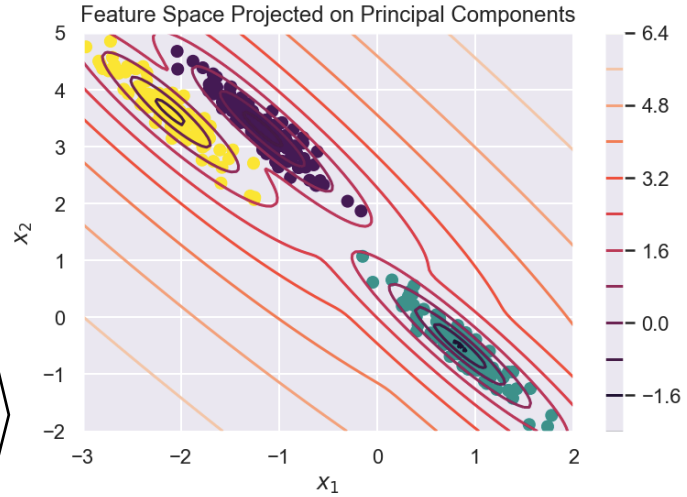
Applying Clustering to Identify Gentrification in New York Metro Area

Phenomenon so complex that applying different methodologies to the same data often yields strikingly different conlcusions …



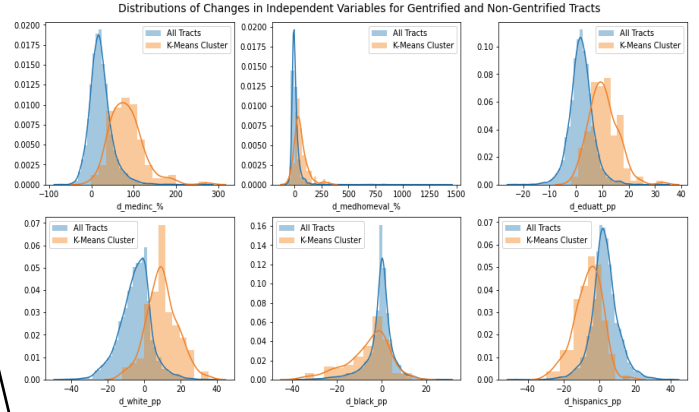
- The map above illustrates lack of consensus among reseachers about what constitutes gentrification best and demonstrates four most cited methodologies applied to the same area by Preis et al.
- Most of the methods apply arbitary cut-offs to different independent variables without regard for interdependence between the variables, which is often equivalent to discarding the lowest-varying features
- This is further exacerbated by one or two variables (median household income and median home value in our case) completely dominating the variation of features across census tracts

… which can be solved by applying theory-agnostic clustering algorithm to identify statistical regularities within data without imposing any assumptions driven by pre-conceptions …



- Instead of making explicit assumptions about the nature of the phenomenon, we suggest applying clustering algorithm that separates unlabelled data into groups which can be interpreted based on their statistical properties
- The advantage of this approach is that it takes into account non-linear relationship between variables and defines clusters based on the measure of distance in the feature space, which allows to separate tracts that materially differ even along just one dimension, into different groups
- To avoid being stuck in local minima, the algorithm runs many times and only the best results are presented

… leading to easily interpretable and robust labelling of census tracts undergoing gentrification that also generalizes well out of sample



- By reserving three more variables as external, we were also able to check if census tracts identified as undergoing gentrification differ along other dimensions from non-gentrified tracts
- We found that our algorithm is better able to identify tracts that have seen steeper decline in the number of violent crimes and exhibit higher rates of property price growth than in surrounding tracts that haven't been identified as gentrifying
- The number of eviction rates in those tracts, however, is lower than for non-gentrifying tracts, which raises additional questions about the cause-effect relationship of gentrification
- In our view, the results are very encouraging and motivate further study of potential application of clustering algorithms to social phenomena