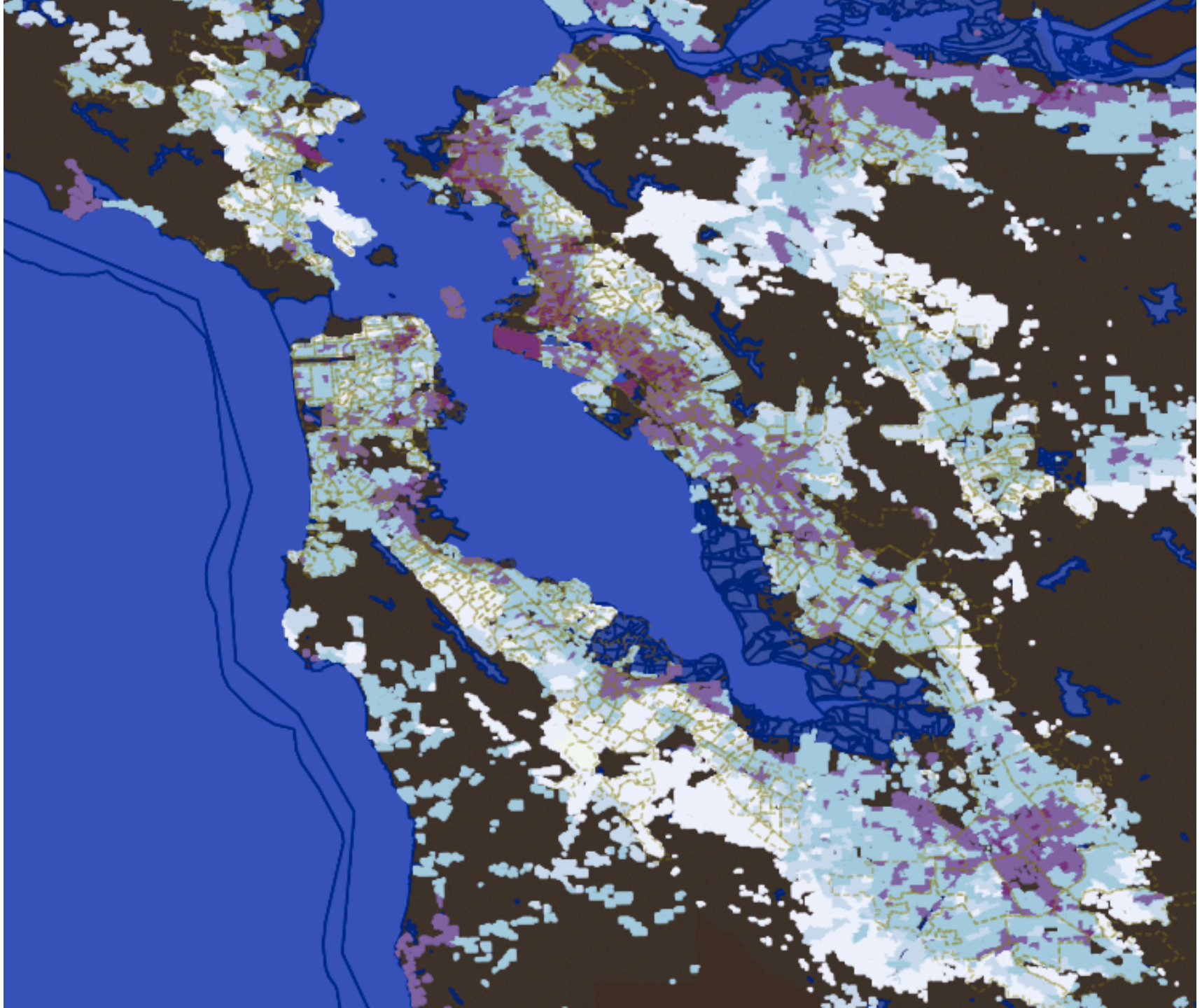
Exploring concepts from gentrification

Brian M. Hamlin
Sarah Quadri (CSU East Bay)
Kenneth Joseph (Carnegie Mellon Univ.)



150m Grid Base Data

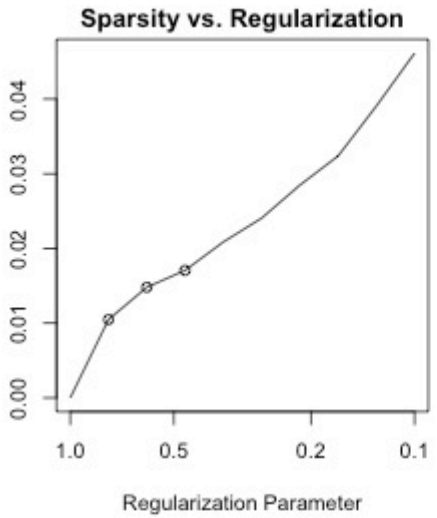
UrbanFootprint 1.0 Data Dictionary

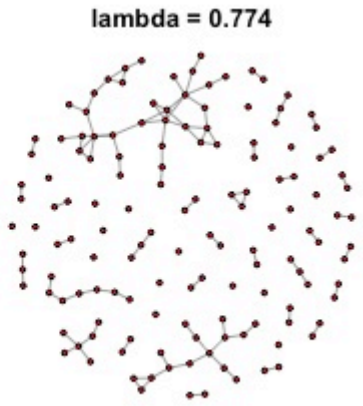
This UrbanFootprint-California (v1.0) Base/Existing Conditions Dataset
(Copyright (C) 2012 Calthorpe Associates) is made available under the Open Database License.
Any rights in individual contents of the database are licensed under the Database Contents License
whose text can be found at http://opendatacommons.org/licenses/dbcl/+

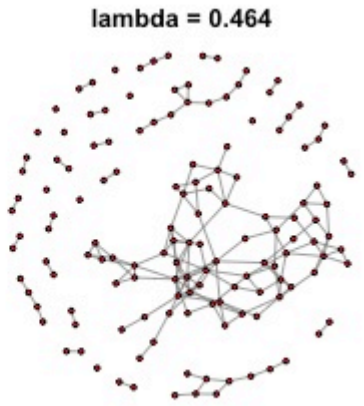
contact:
engineer: Brian M Hamlin, maplabs@light42.com
principal: Joe DiStefano joed@calthorpe.com

Variable Name	Definition
ID_Grid	Grid ID
County	County Name
Area Variables by Use and Landtype	
Parcel_SqFt	Parcel square feet
Acres_Grid	Grid cell acres (approx 5.55987)
Acres_Grid_Urban	Urban grid cell acres
Acres_Grid_GF	Greenfield grid field acres
Acres_Grid_Con	Constrained grid field acres *(Note: These acres can be divided into any land coverage type of conser farmland classes below, but it is anticipated that we will include habitat, carbon sequestration land*

Understanding Census variable correlations





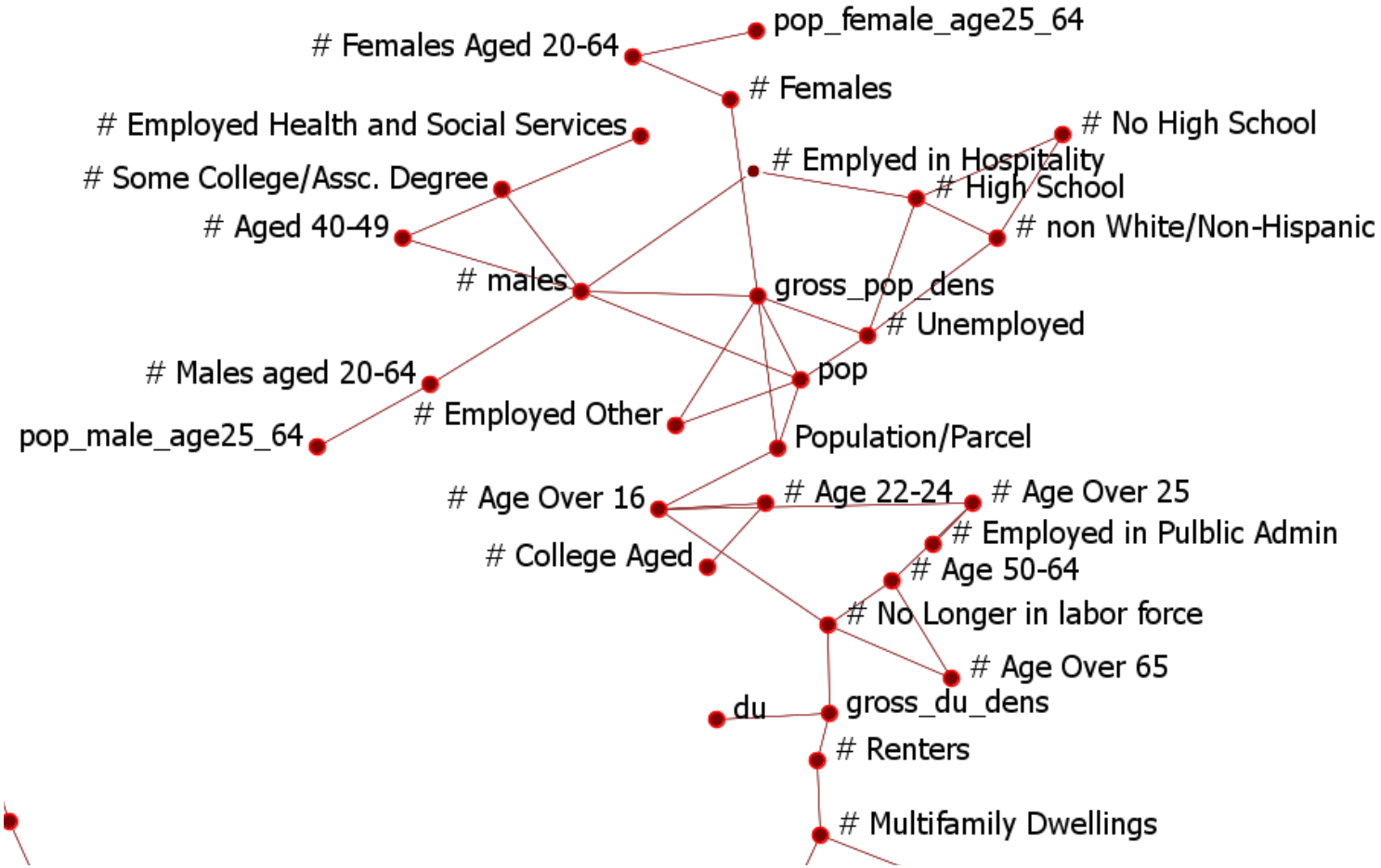


Use a nonparametric method developed for high-dimensional data where d~n or d > n

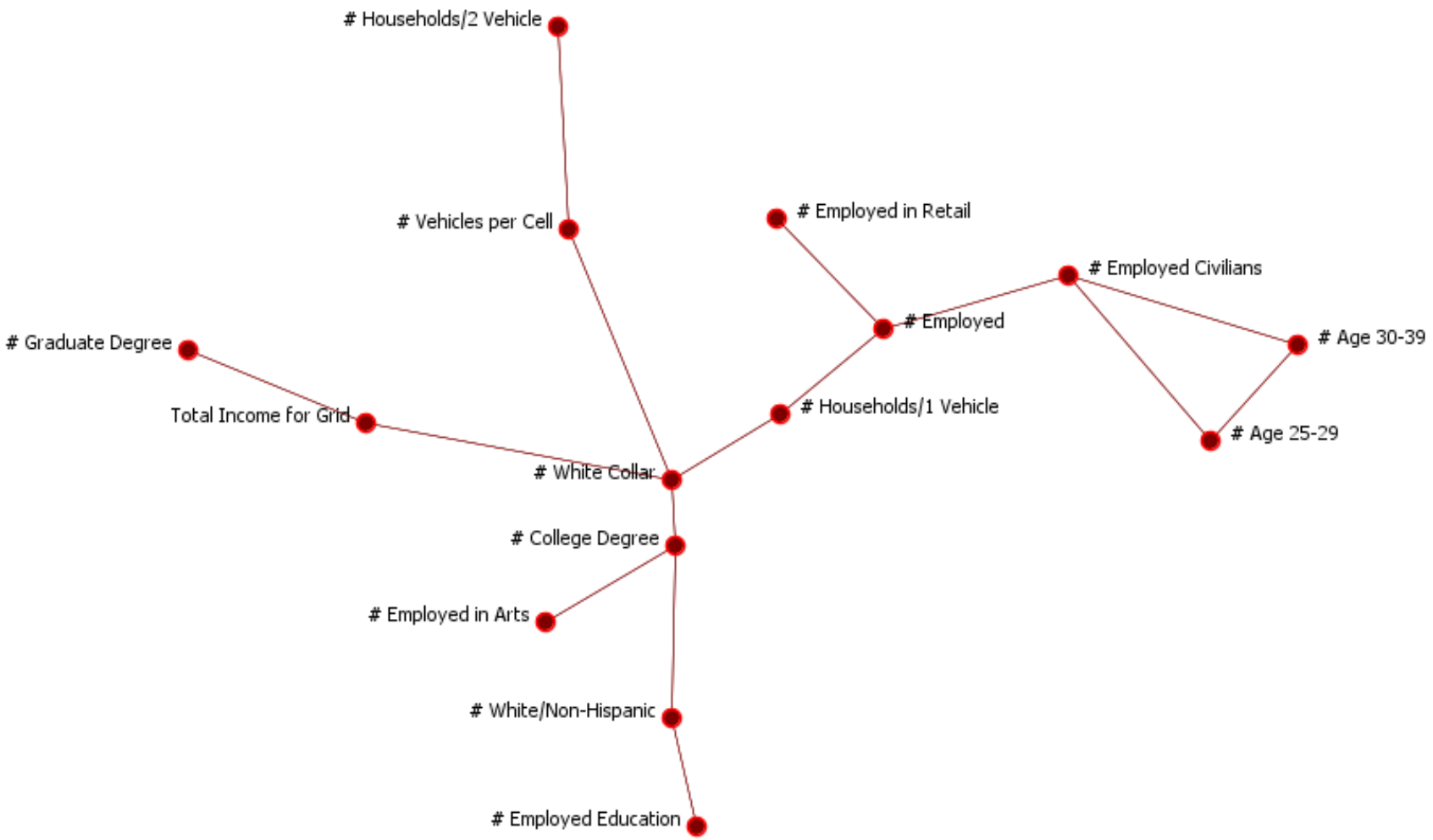
(560 neighborhoods, 200 census variables)

T. Zhao and H. Liu. The huge Package for High-dimensional Undirected Graph Estimation in R. Journal of Machine Learning Research, 2012

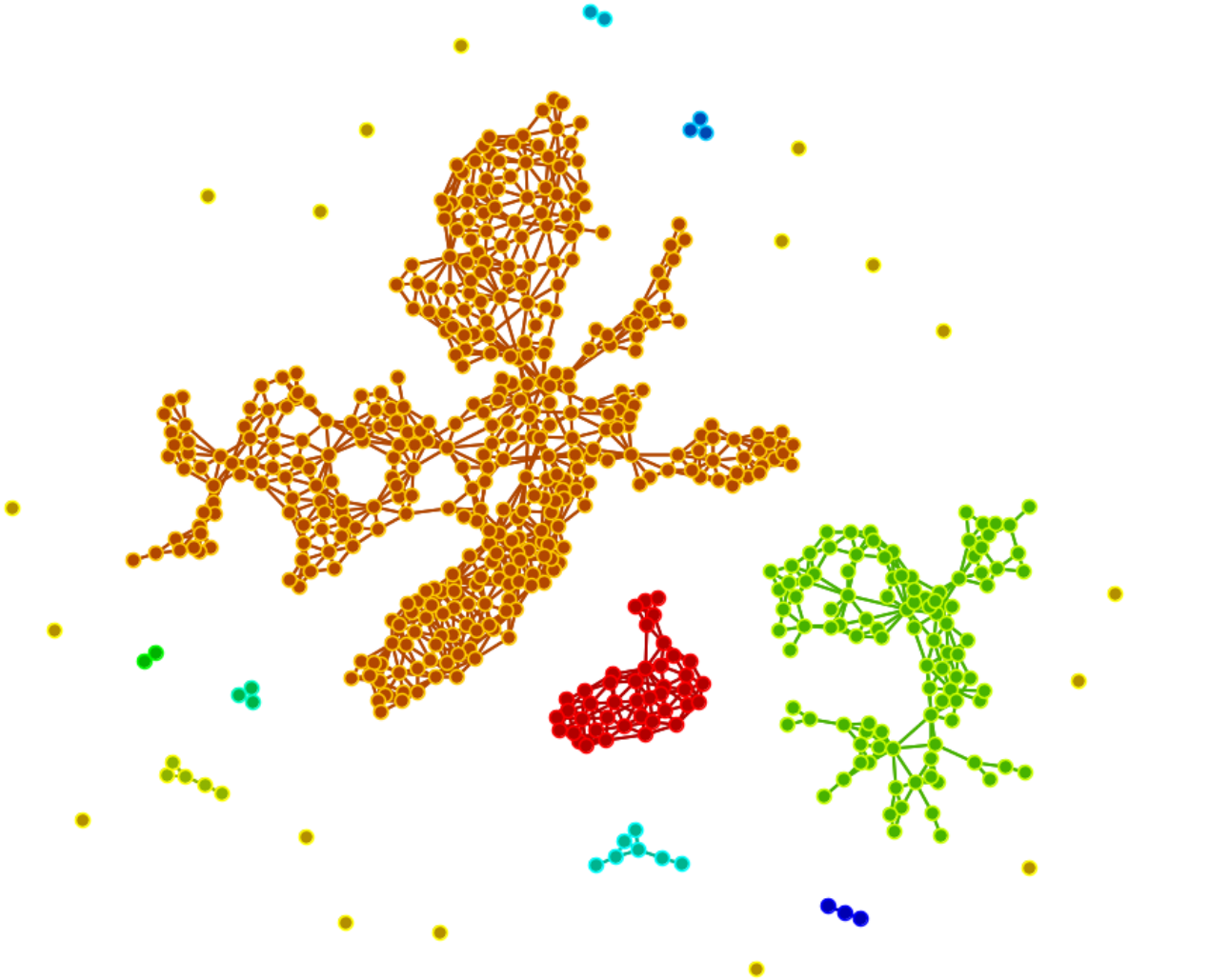
“Population variables”



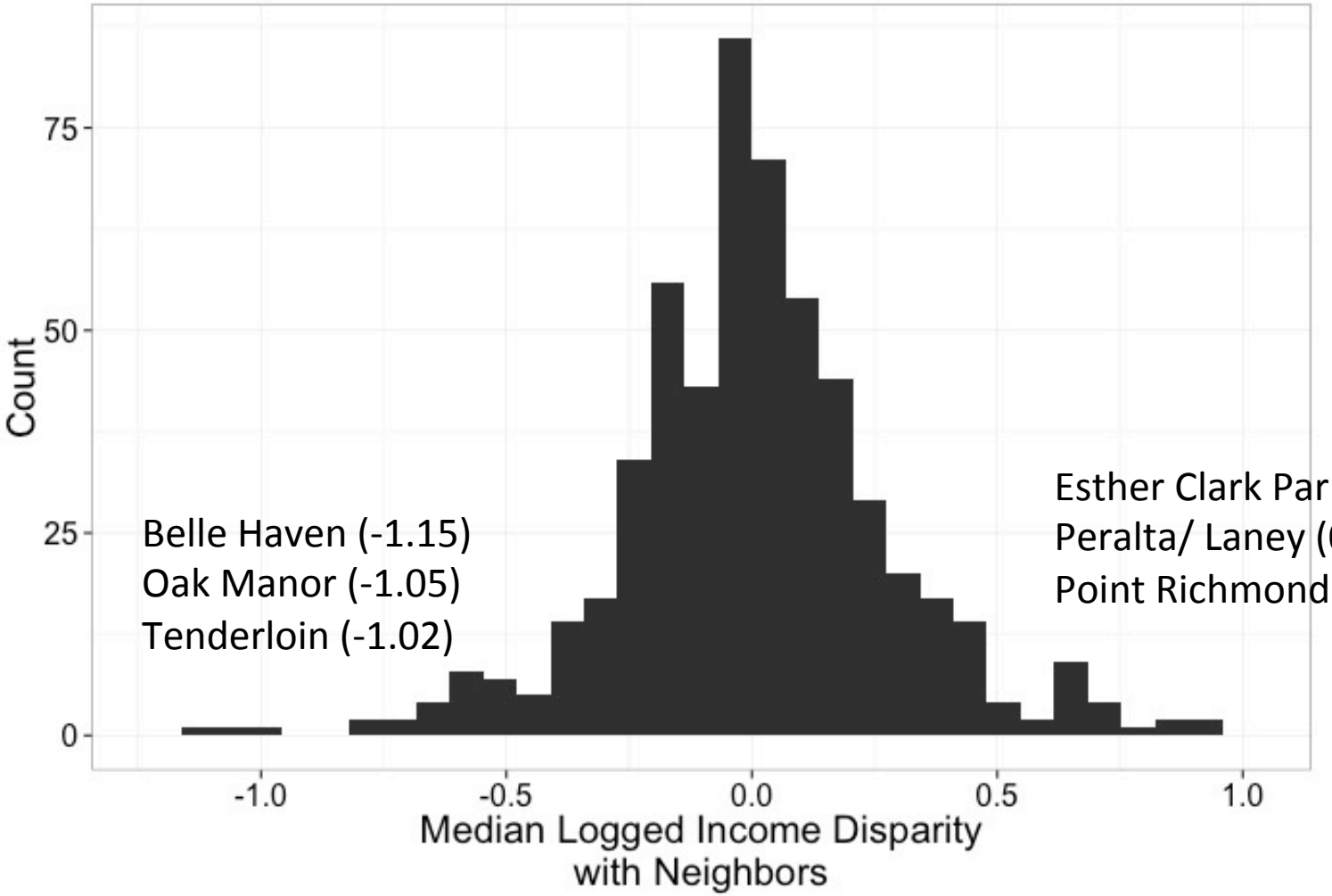
“Gentrification Variables”



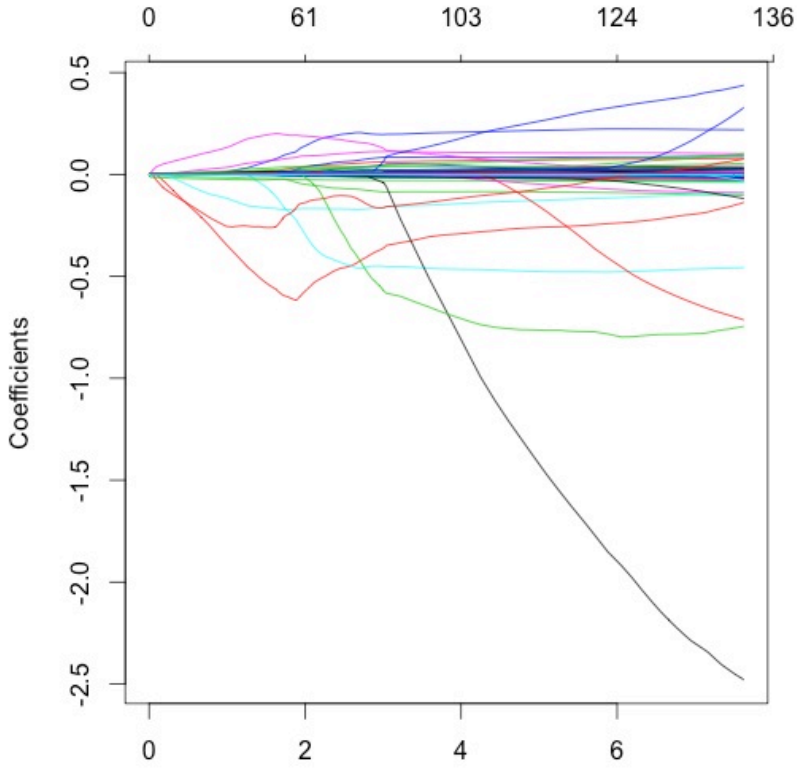
Understanding “Pushing out phemomena”

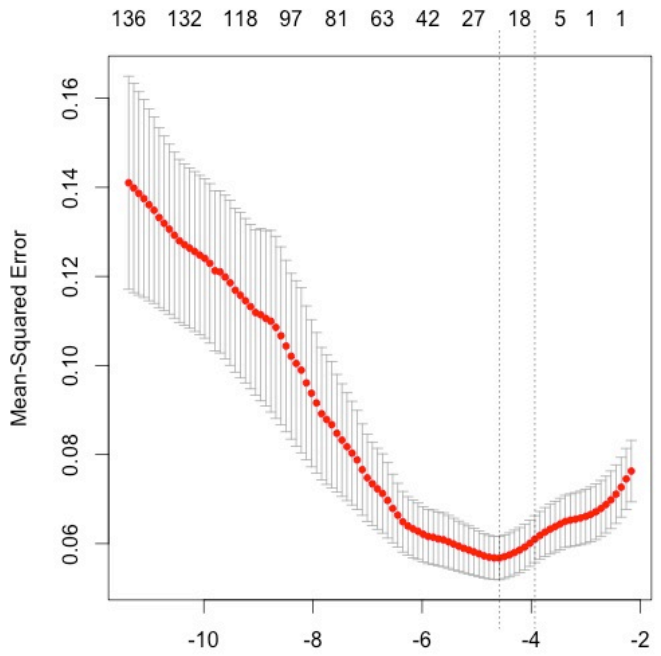


Nonparametric correlation graph discovery



Regression Results





Regression Results

du_occ_rate	-0.345
emp_ag	-0.255
pop_ind_extract	-0.117
pop_age20	-0.024
pop_ind_other	-0.015
hh_veh3p	0.007
pop_ind_public	0.008
hh_avg_veh	0.011
pop_ind_whole	0.011
hh_avg_size	0.013
emp_transware	0.016
pop_ind_util	0.035
pop_employed_af	0.132