

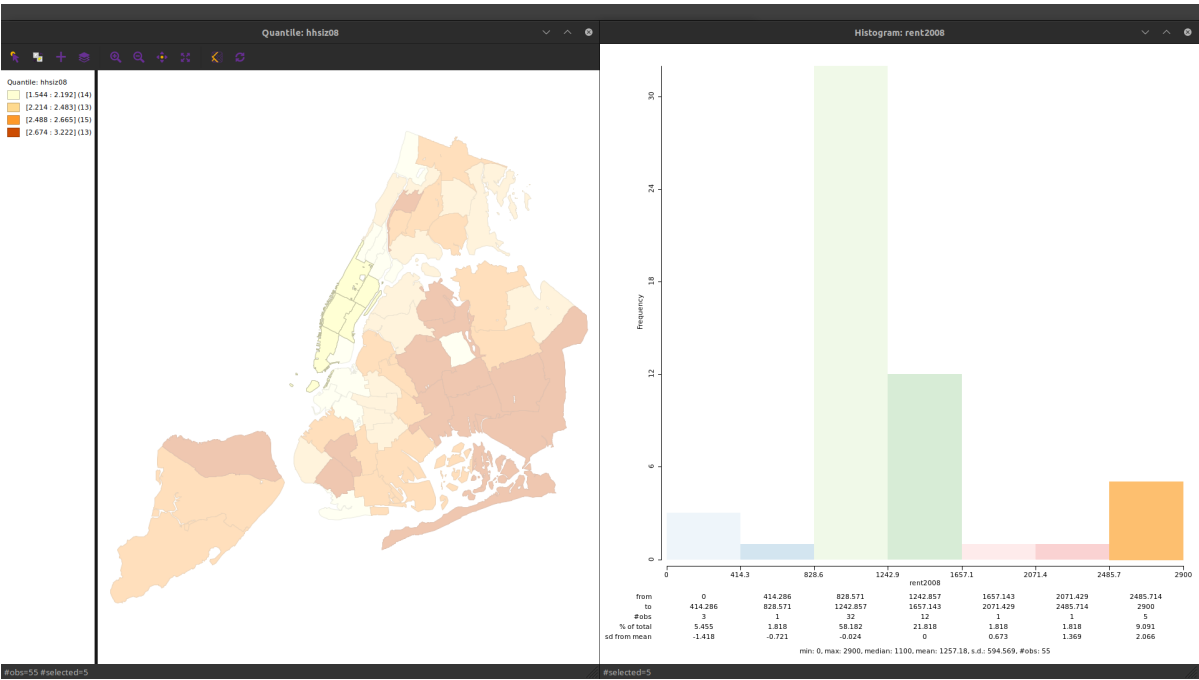
Lab 1, Spatial Data Analysis

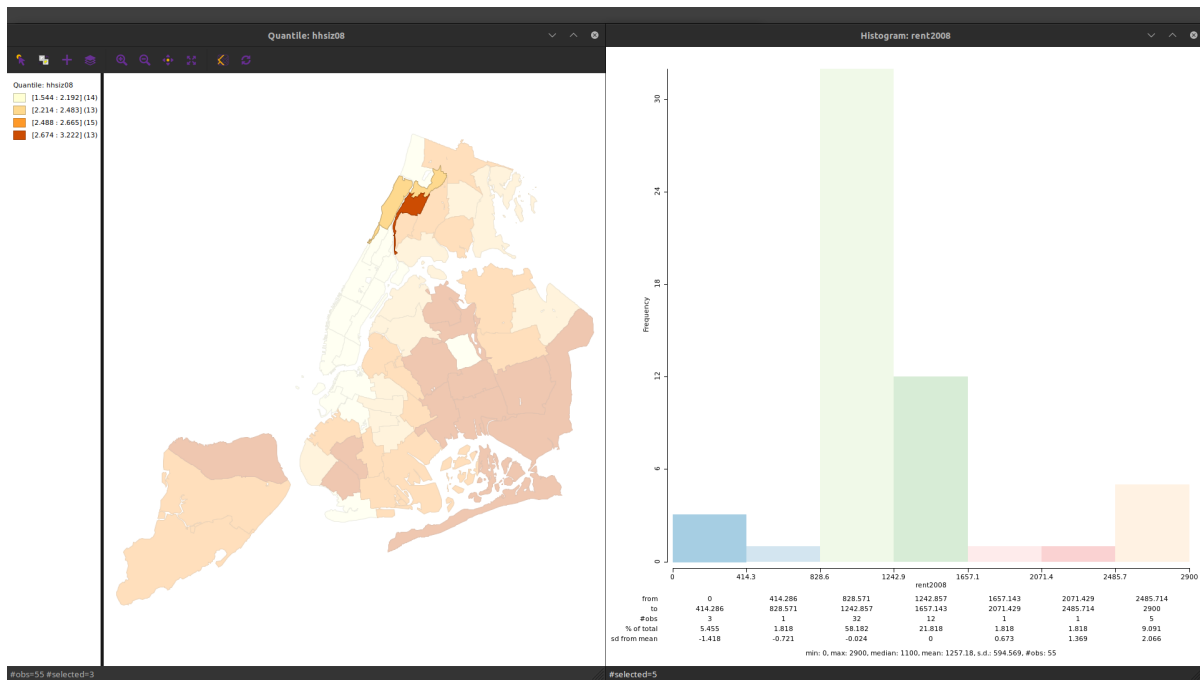
Ian Grant

February 15, 2024

1 Boxplots

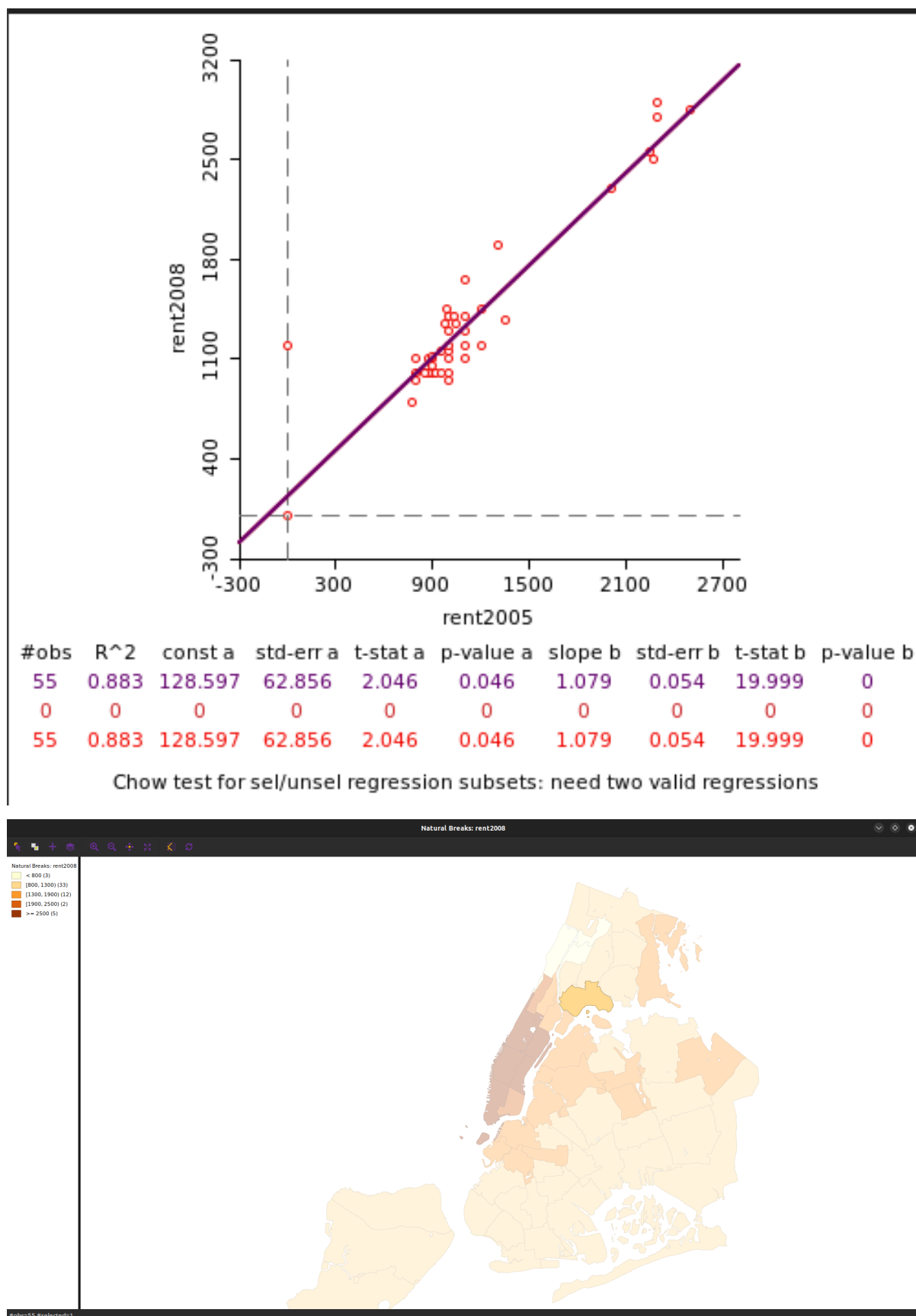
The boxplots in the screenshots shown below show the distribution of rent in 2008. The maps show the distribution of household size. In the first image, selecting lower Manhattan reveals that this group of sub-boroughs has low household size (highlighted on the map) and high rent (highlighted in the boxplot, on the right side). In the second image, selecting two sub-boroughs in northern Manhattan and the Bronx shows that while both have very low rent (highlighted in the boxplot), they diverge sharply in household size.





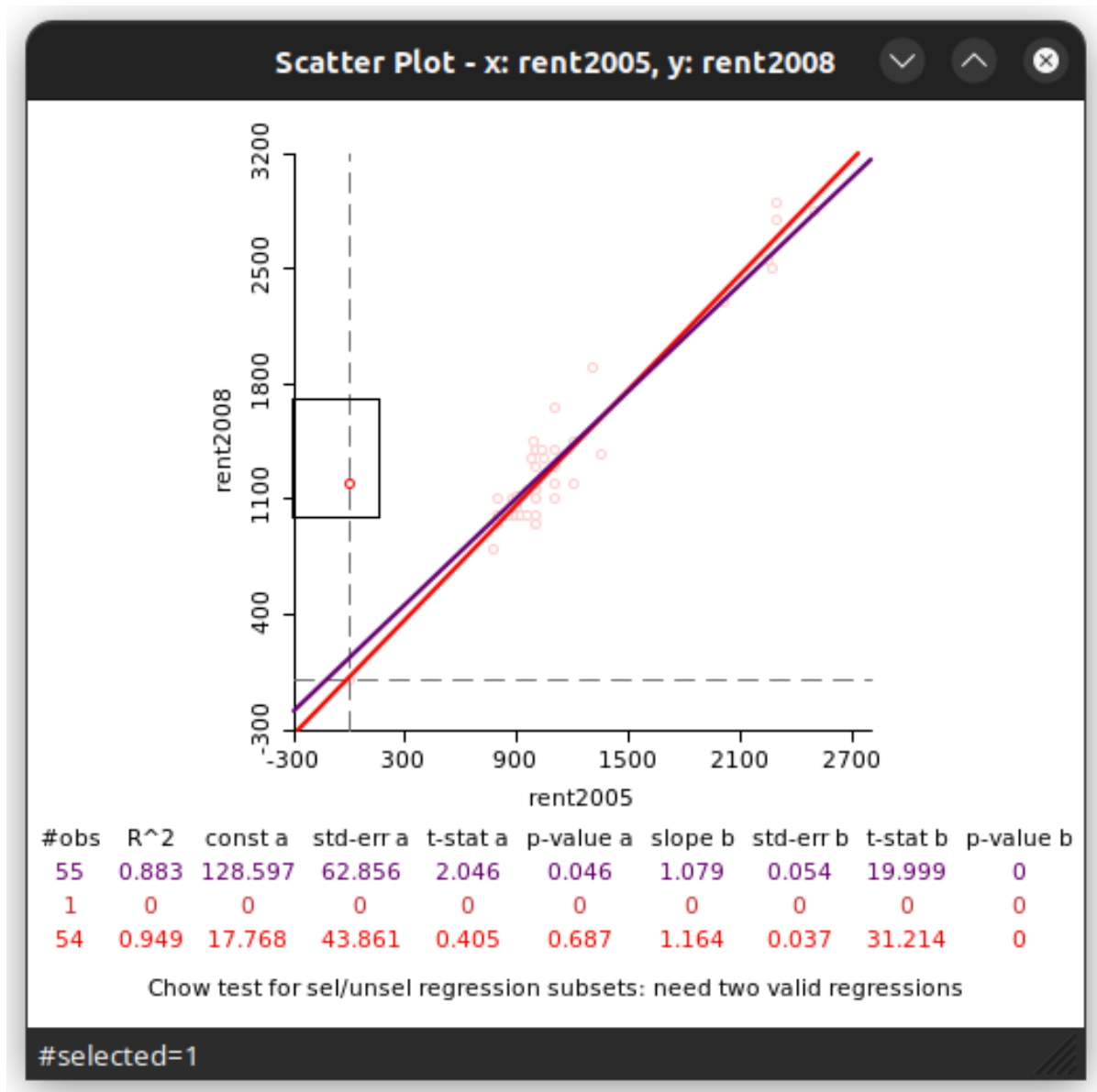
2 Scatterplots

We can using a scatterplot and the accompanying regression to characterize how rent changed from 2005 to 2008. The scatterplot shown below shows that rent increases were relatively uniform with the exception of a sub-borough in the South Bronx which is an outlier, highlighted in the map below.

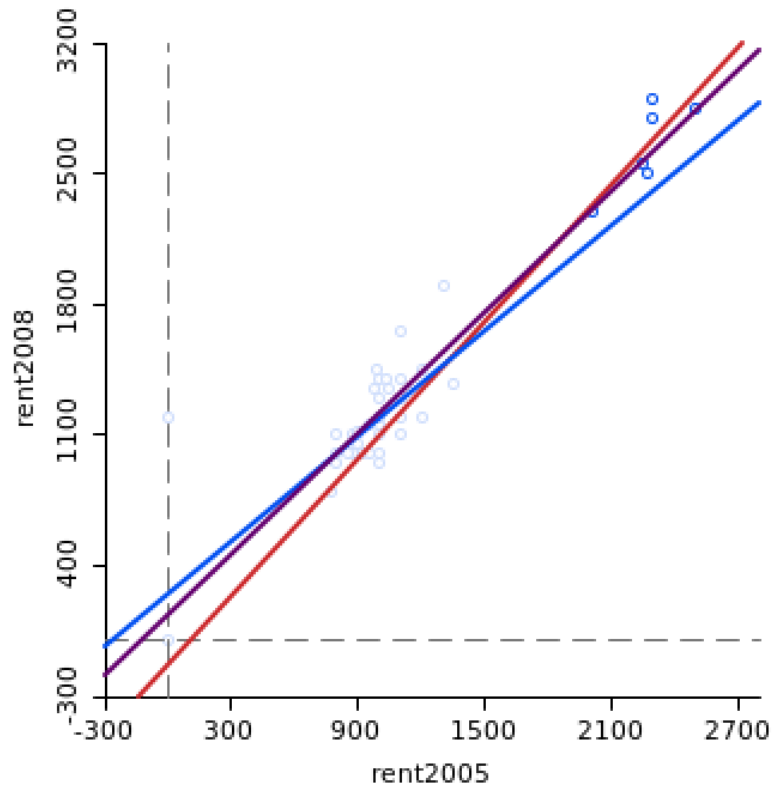


Because this outlier is on the extreme low end of the 2005 rent distribution, we can expect it to have an outsize impact on the regression fit. Indeed, if we exclude the outlier from the regression, the R squared

increases from .883 to .949 (below). Also, the intercept decreases from \$128.60 to \$17.768 and its p-value becomes non-significant. An intercept indistinguishable from zero makes sense if we expect rent increases to occur as a factor of existing rent, not as a fixed increase regardless of current rent. Since the slope of the new regression is 1.164 and the intercept is near zero, we can conclude that the sub-boroughs' median rent increased by an average of about 16.5% over the three-year period from 2005 to 2008.



Finally, we can ask whether some geographical areas experienced different patterns of rent increase. Let's look at lower Manhattan, which has the city's highest rents. Selecting just these sub-boroughs, shown at the top right of the scatterplot below, produces slightly different slopes. But the Chow test for a difference in slopes produces a non-significant p-value of .252; we cannot conclude that Manhattan's rent-increase trend is different from the rest of the city's.

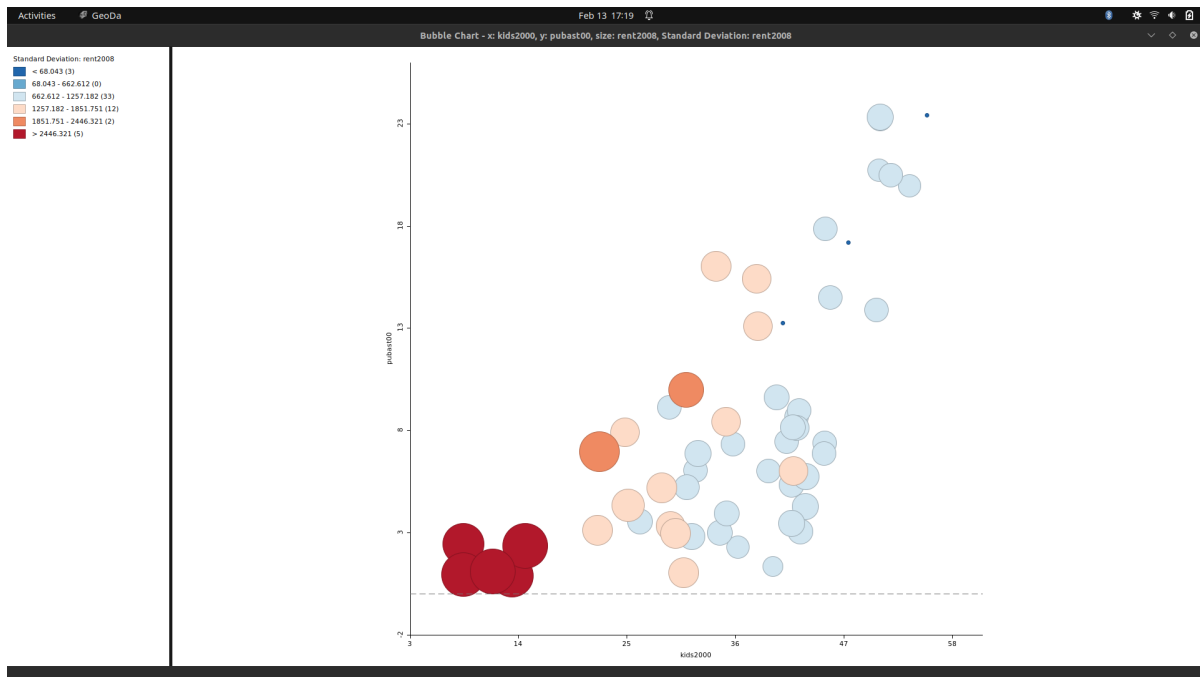


#obs	R ²	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
55	0.883	128.597	62.856	2.046	0.046	1.079	0.054	19.999	0
6	0.657	-133.031	1007.064	-0.132	0.901	1.225	0.442	2.769	0.050
49	0.648	245.361	95.104	2.580	0.013	0.939	0.101	9.309	0.000

Chow test for sel/unsel regression subsets: distrib=F(2,51), ratio=1.4161, p-val=0.2520

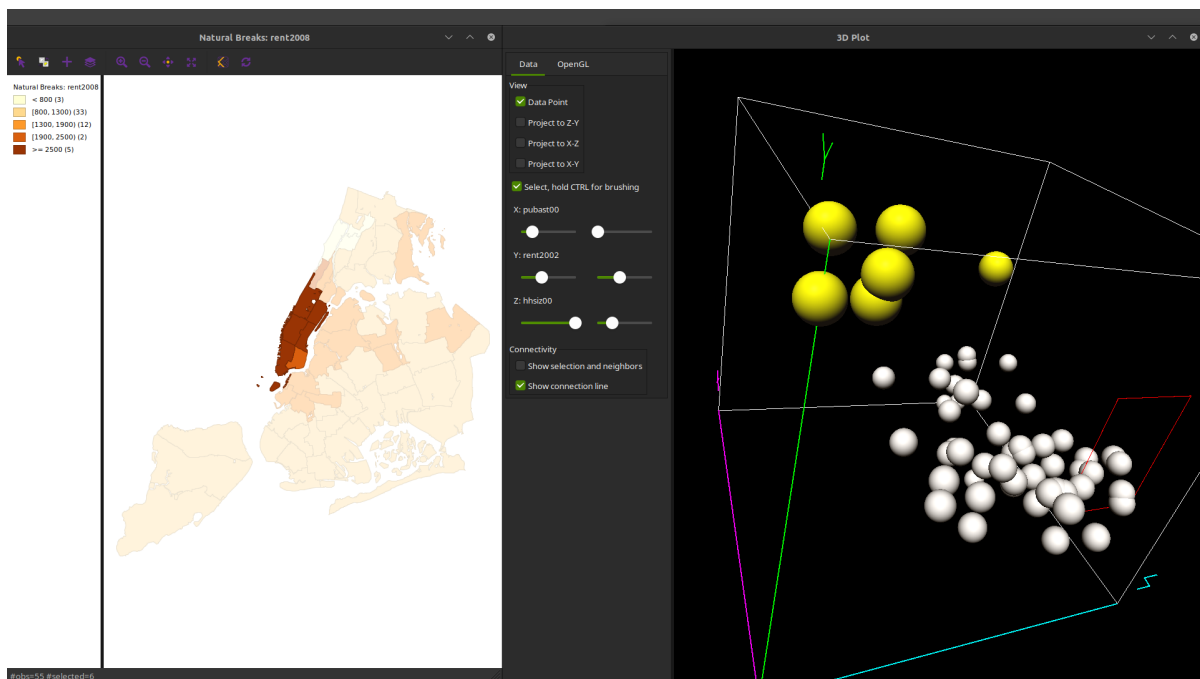
3 Bubble plot

This same group of rich sub-boroughs in lower Manhattan stands out on a bubble plot of percent public assistance and percent children. The sub-boroughs are shown on the lower left of the plot, with the lowest percentage of children and the lowest percent on public assistance. Meanwhile, the size and shape of the bubble shows that rent tracks closely with these two variables, with the highest-rent sub-boroughs (lower Manhattan) also having the lowest values for children and public assistance.



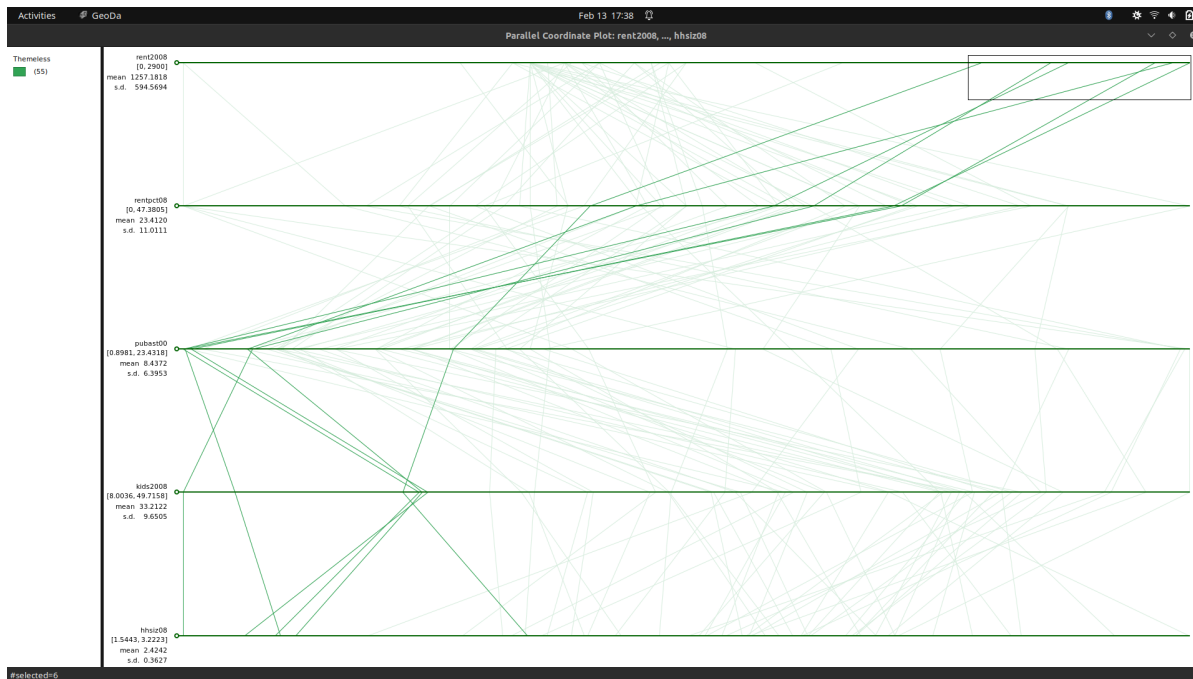
4 3D scatter plot

These outlying lower and midtown Manhattan sub-boroughs are even clearer in a 3D scatterplot. In the screenshot below, the sub-boroughs are shown in dark red on the map and in yellow in the 3D scatterplot. Their placement in the 3D plot indicates that they have lower percent public assistance, high rent, and low household size.

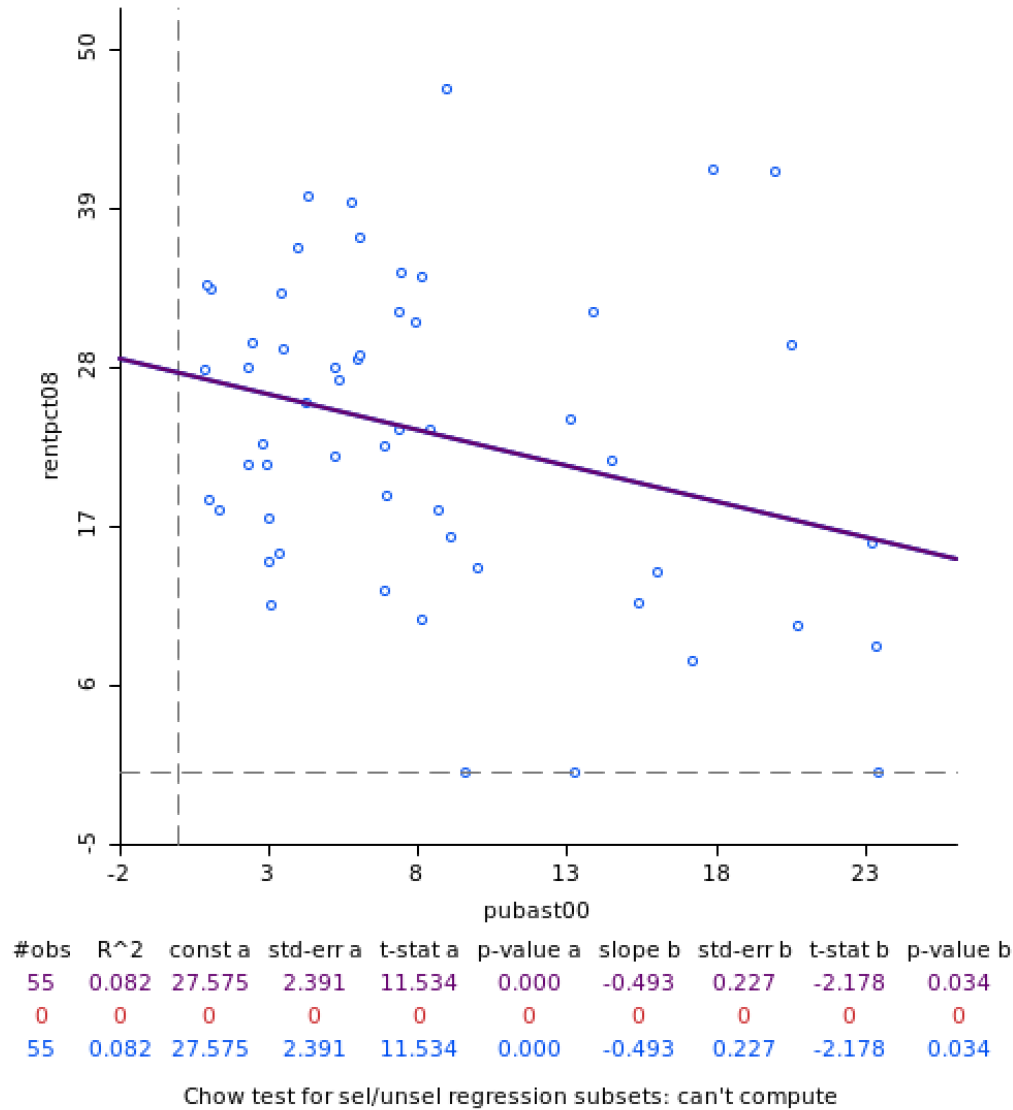


5 Parallel coordinates plot

A similar picture emerges with the parallel coordinate plot. Across rent, percent of housing stock that is rental units, percent of population on public assistance, percent children, and household size, the rich boroughs of lower and midtown Manhattan follow a similar trajectory, with high rent, medium percent housing stock, and low values for the other variables.



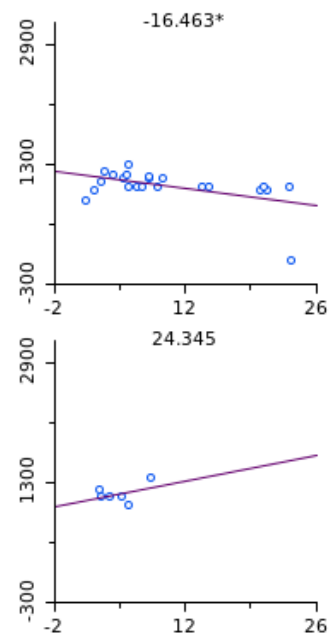
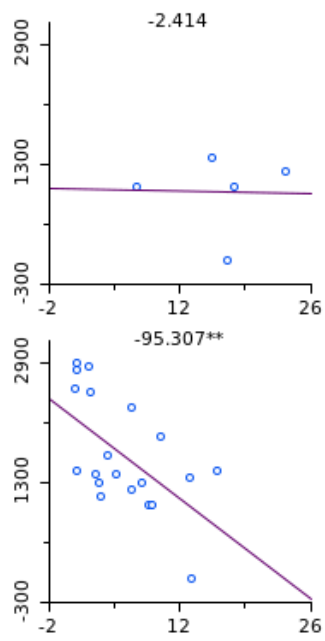
6 Conditional scatterplot



Overall, rent in 2008 shows a significant but weak relationship with public assistance in 2000 (the most recent available data in the dataset; scatterplot shown above). However, there is a lot of variation around the trend line and the R squared value of this fit is .082. A conditional scatterplot may help to clarify the relationship.

Below, the same scatterplot is reproduced with children in 2008 and household size in 2008 as conditions. Now we can see that there is a strong negative relationship between 2008 rent and 2000 public assistance for sub-boroughs with fewer children and smaller households (bottom left). By selecting the points in this sub-plot, we can see that they are mostly relatively central neighborhoods of the city in Manhattan and western Brooklyn and Queens, though there are some Bronx and eastern Queens sub-boroughs included as well.

Quantile vert cat var: kids2008, dep. var: rent2008
33.594



2.486
Quantile horiz cat var: hhsiz08, ind. var: pubast00

