# Yue Wu

Reading, Understanding and Predicting Real Estate Price
## Xuhui District, Shanghai, China

1st plot – Diachronic analysis of housing price

Noticing that the housing market is never static but consistently in the situation of change, one of the variables to measure its character is the variable *Change* which indicates the change of price (compared with last month). According to the scatter plot (distributed with the variable of Price) in the following, the price change is comparatively dispersive, ranging from 0 to 1, which does not perform an apparent pattern, neither show a strong relationship with Price itself. In this way, the correlation of price change, using cross reference of the housing price with other variables across different years as a methodology will be investigated in the following.
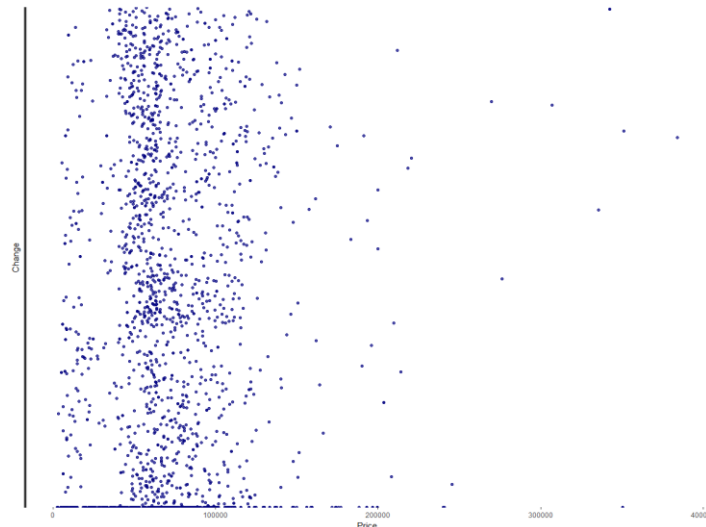
Fig 01-a: The scatter plot of price change housing price mean aggregated by different zones
The y-axis ranges from 0 to 1

After merging 3 different datasets altogether using the same name of real estate as an index, the diachronic prices by three datasets are listed within single data frame (more than 670 rows with same name index are sorted and merged). Comparing the line plot of housing price across different years in *Fig 01-b (a sample of 255 rows is selected and represented)*, the pattern of *Price* from 05/2015 to 04/2016 is comparatively clear, mainly showing an increasing tendency regardless of a small amount of extremum.
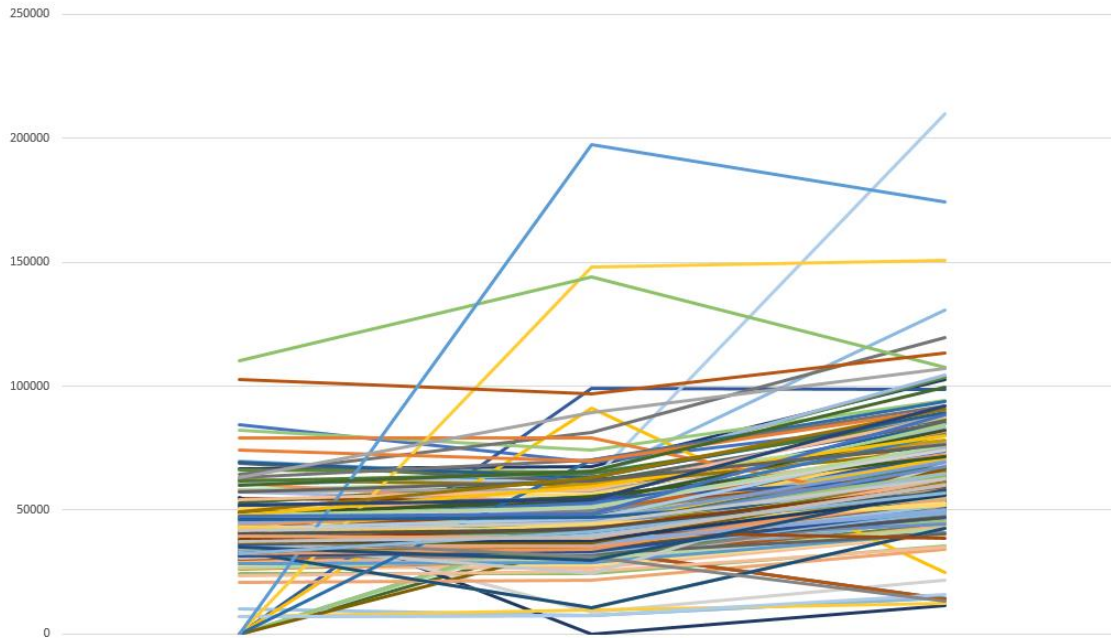
Fig 01-b: The boxplot of housing price across different years from 2015 to 2016 (this graph is done by Excel)

Running a multi-linear regression model, it indicates that there are certain but not strong correlation indicators that shows the relationship between variable and housing Price. Among the different variables, the change of housing price is mostly related to the original housing price itself, demonstrating that the house that has a higher starting price tends to have a propensity of getting a larger price change (increase). On the other hand, it reflects the tendency displayed in the real world that all the price of houses with different variables show a strong increasing tendency.

Variable	Estimate	Std.	Error	T value	Pr(>|t|)
Price	0.00000229044258	0.00000048322678	4.73989	0.0000030	***
Metro	0.14722278452926	0.05054153985413	2.91291	0.0037781	**
School	-0.0790560047503	0.03036264264384	-2.6037	0.0095589	**
Environment	8587.1	3008.4	2.854	0.00439	**
Green_Ratio	-0.3717807376132	0.19825904943950	-1.8752	0.0614773	.
ZoneChangqiao	0.24137315657099	0.11786048292500	2.04796	0.0412055	*
ZoneHuaihaixi	-0.2372351723513	0.11724619522105	-2.0234	0.0436857	*

Chart 01-a: The correlation chart of variables in terms of price change showing the most correlated variables within the multi-linear regression model