Group Name : Insta

Group Member : Yunsung Jung, Yuna Lee, Sungwon Lee

<Which factors affect Property Price in Seoul?>

▪ How many unique observations to you have?

We have one unique observation, Jung-gu. Property Price of Jung-gu is much higher than Property Price of the others.

▪ What information/features/characteristics do you have for each observation?

First of all, we analyze multiple regression of Property Price. The multiple regression model is significantly meaningful (P-Value<.005, R-Squared = 0.6216). In this model, Transportation and Welfare Facilities have significant effects. We also analyze simple regression of each factor, such as Transportation, Education, Welfare Facilities, Crime Accidents, and Income. Transportation, Education, Crime Accidents, and Income have positive relationship. However, Welfare Facilities has only negative relationship.

▪ What are the min/max/mean/median/sd values for each of these features? What is the distribution of the core features (show a histogram)?

	Min	Max	Mean	Median	SE Value
Property Price	1876000	6409000	3144914	2792000	192834.6
Transportation	0	8	2.12	2	0.3431
Education	0	6	2	2	0.3055
Welfare Facilities	141	770	405.88	392	29.8052
Crime Accidents	3124	8851	5226.96	5289	273.6773
Income	241.8	479.8	320	307.8	10.7568

*Property Price – Won

*Transportation – the number of transfer stations

*Education – the number of high schools which have high education-achievement (2013)

*Welfare Facilities – the number of facilities, such as Parks, Facilities for Handicapped, Nursery Facilities (2015)

*Crime Accidents – the number of major crimes (2014)

*Income – monthly income (unit. 10000 Won) (2008)

▪ Are there obvious trends in the data (over time, across subgroups, etc.), and are the differences statistically significant?

We would like to search difference of Property Price between Gangnam and Gangbuk. First of all, we exercise T-Test with two groups. However, there is no significant difference (P-Value : 0.5186). So, we erase an outlier data of Gangbuk group. After that, we do T-Test again and get a significant difference at a 90% significance level (P-Value : 0.077).

▪ What are the other salient aspects of the data (e.g. geospatial factors, text content, etc.)?

We search correlation of each factor. Transportation only has a positive relationship (P-Value < .05). Welfare Facilities has a negative relationship (P-Value<.05). The other factors, Education, Crime Accidents, and Income have a high P-Value, so they do not reject null hypothesis.

▪ Provide a bullet-list of the next 5-10 tasks you will perform in analyzing your dataset.

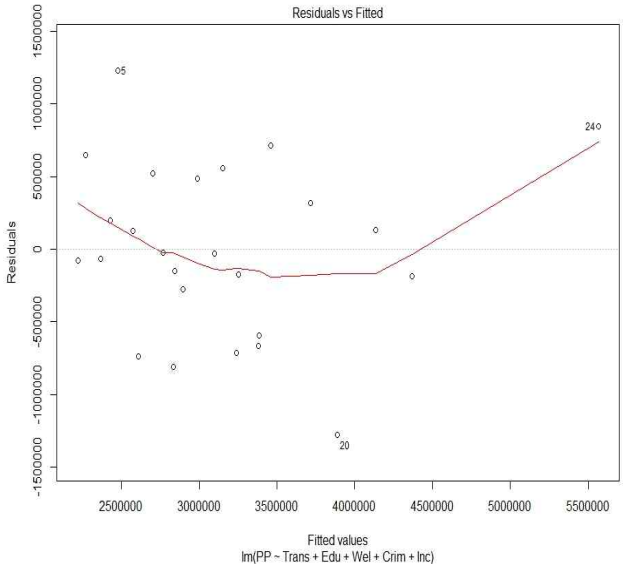
- Why does Welfare Facilities have a negative relationship?

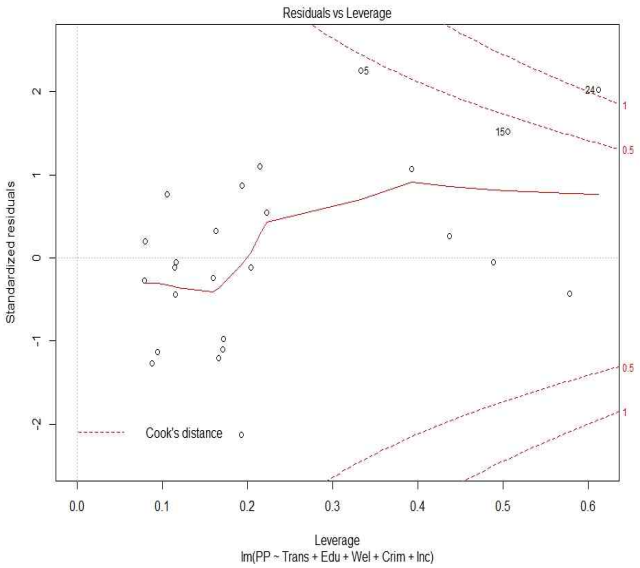
- In our multiple regression model, R-Squared is only 0.6216. We have some questions about other factors which affect Property Price.

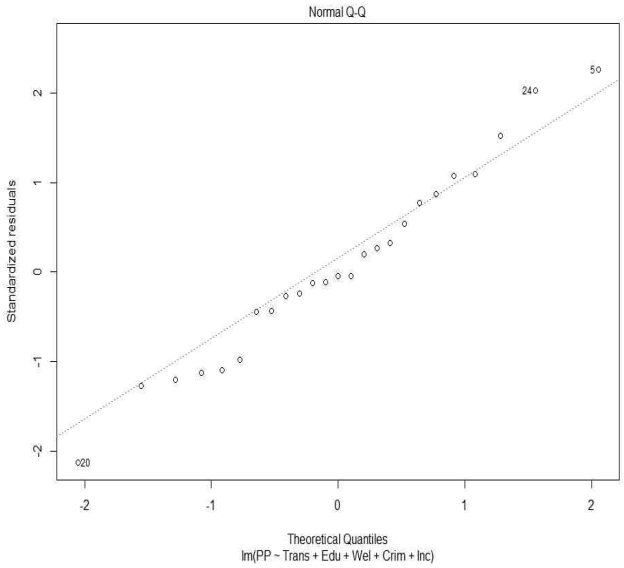
- Why does not Education affect Property Price significantly?

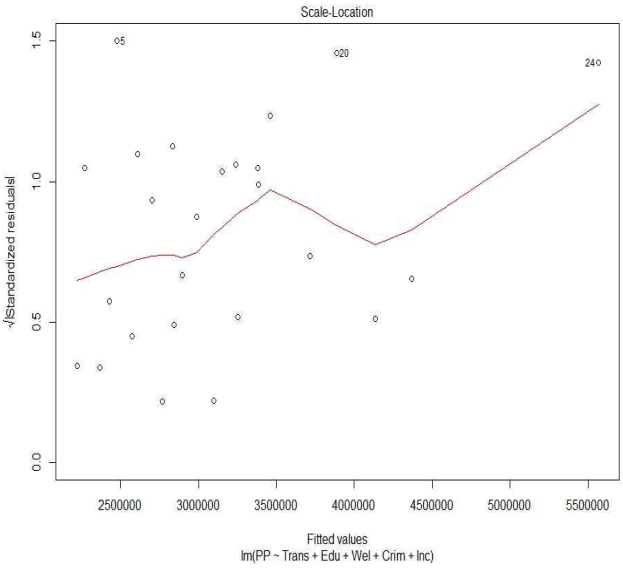
- If we erase an outlier data, how would the result be changed?

● Plots of multiple regression

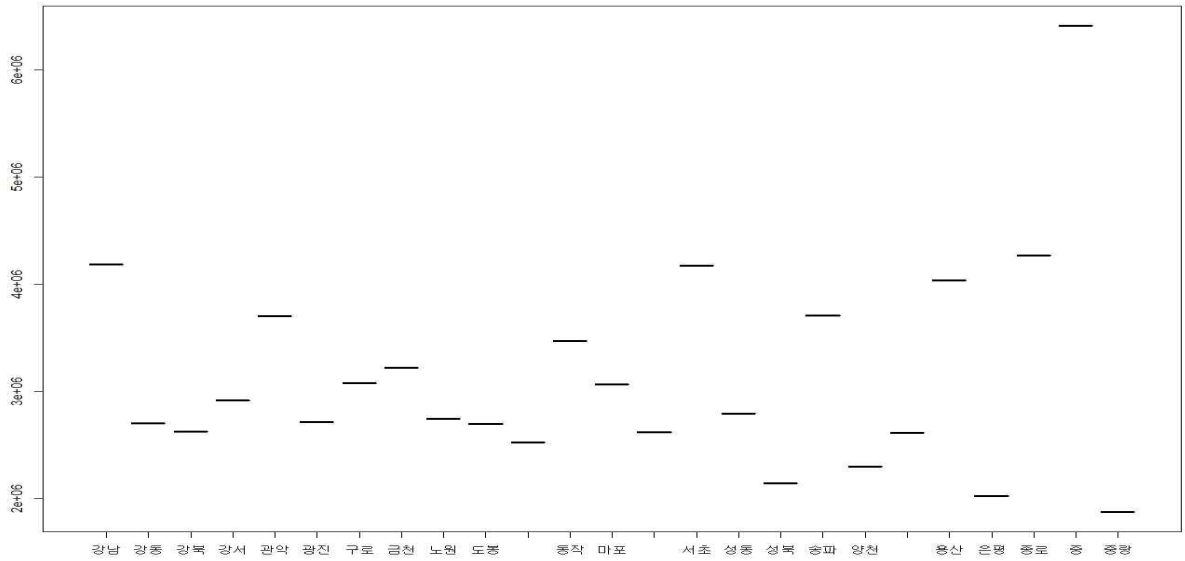


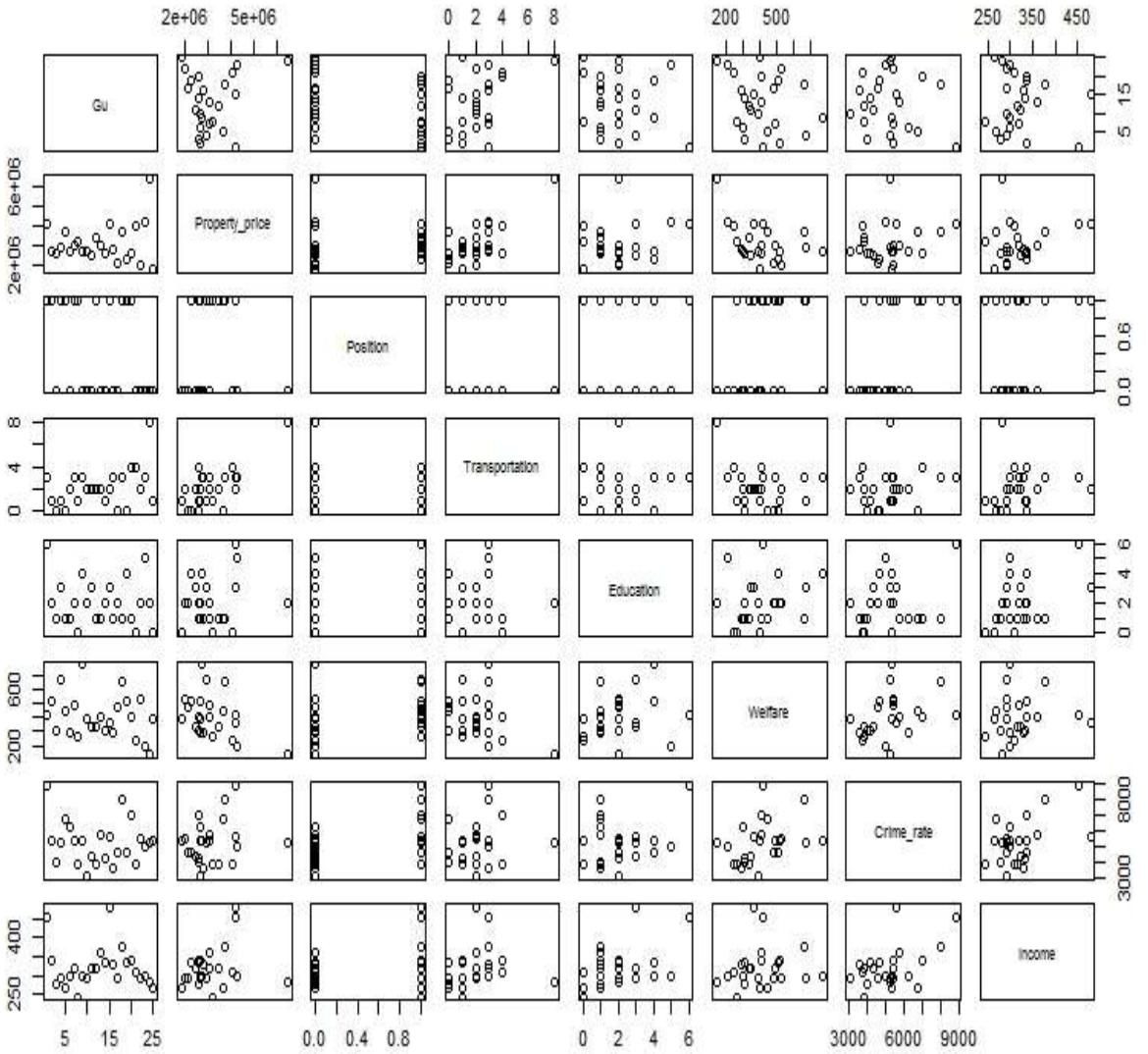






● Plots of property price by boroughs and factors





* Plots of simple regression