1. Should all variables be used in our model?(seller, data and postcode, Address)? We should list enough reasons.
2. Is there any variables that are highly correlated?
3. Variables suburb and CouncilArea have too many levels, like 34 levels for CouncilArea, should we keep this variable in the process of building model? I guess that we should drop the two variables because it is too difficult to interpret each level.
4. Detect relationship between independent variables and dependent variable(price). Use plots, scatterplots and add labels, edit axis.
5. After model building, use plot () function to check model assumption.