# Exercises 1 Exploring data and Variables

1. Please load the data *WoON2015small.dta* in Stata
2. keep the following variables in memory, dropping the remaining variables: value of houses (WOZwaarde), number of children in the household ( AantKind), gender and age of the respondent interviewed (gslop, LftOP), number of rooms (Kamer5), housing tenure (huko) and municipalities ( g4\_5). command [keep WOZwaarde AantKind gslop LftOP Kamer5 huko g4\_5]
3. Could you identify which variables are continue variables and which variables are factor (categorical)?

Give a brief description: ………..

1. Consider now factor variables only. What can you say about the distribution of these factor variables? what do these tell about houses and households?

what stands out: ……..

1. Consider continue variables. How could you characterize the distribution of continue variables?

Give a brief description: ……..

1. Could you calculate mean housing values for rental and owner-occupied sectors?

The average price of a rental house…………… and an owner-occupied house ……….….

1. In which municipality is the share of rental housing market the lowest and In which municipality the highest?

…………………………………………..

1. Could you generate another age variable (to be named agecat3) including three categories: 1) young, 2) medium-aged and 3) senior? Which group is overrepresented in the data?

………………………

1. Could you generate a dummy variable for young people?

OLS regression:

1. Could you run a OLS regression model to explain housing value (WOZwaarde) from age (LftOP ). What is the effect of age? ………….
2. Please run another OLS model including another variable, number of kids in the household (AantKind). So, this model should include two explanatory variables: age and household members. What is the effect of these variables? …………………
3. Please estimate another OLS model with an additional variable, housing tenure (huko). So, this model should include 3 variables: LftOP, AantKind and huko. What has changed in the effects of the variables with respect to the estimates in j) and k)? do you have an explanation for these changes? …………………