Chapter 5: Variables & Manipulation

Exercises

*Brian Fogarty*

*15 May 2018*

**EXERCISE I**

Using an abbreviated version of the 2015 UK Millennium Cohort survey dataset (mcs.dta), perform the following exercises. Note: the survey was carried out in 2015 to 14 year-old pupil in the UK. The dataset consists of nearly 12,000 observations and 52 variables. You need to use the haven package to read-in the data.

Perform recoding and labeling on the variables below. Note: for all the variables, you need to first convert them to factors using the as.factor () function and then to use the function recode

1. mths - rename as maths and label the values as 1=’1. Strongly Disagree’;2=’2. Disagree’;3=’3. Agree’;4=’4. Strongly Agree’. This variable includes pupils’ responses to whether they were good at mathematics.

2. scien - rename as science and label the values as 1=’1. Strongly Disagree’;2=’2. Disagree’;3=’3. Agree’;4=’4. Strongly Agree’. This variable includes pupils’ responses to

whether they were good at science.

3. sex - rename as gender and label the values as 0=’0. Female’;1=’1. Male’.

4. best - rename as bestsch and label the values as 1=’1. Never’; 2=’2. Sometimes’;3=’3. Most

Times’;4=’4. Always’. This variable includes pupils’ responses to how often they do their best at school.

**EXERCISE II**

Provide the level of measurement for the variables you recoded in Exercise I (in other words what kind of variable do you have)

**Exercise from GBCS**

**Aurelien Boucher**

**EXERCISE III**

In the Great Britain Class Survey, import the dataset gfk\_cleaed\_eul

For this exercise you may need the package labelled and use the function dplyr::recode

1. Identify the character variable to transform into factor variable. How many are they.
2. Recode the variable educ, whichclass, mrock, mmetal according to the information in the excel file 7616\_gbcs\_codebook\_eul
3. For mrock and mmetal creat a new variable whith 3 categories of answers as follow “dislike or dislike a lot”, “neither like or dislike”, “Like or like a lot”
4. Create a variable generation, with the “Boomers” (born between 1946 and 1964), the GenX (born between 1965 and 1984), “Millenium” (born between 1985 and 1996), the “born in 1945 or before”, and GenZ for those born after 1996. Use the cut function

**EXERCISE IV**

1. Import the excel file gfk\_excel\_version into R (from the menu file->import data set->from Excel)
2. Inspect the variables hhincome? And describe what are the different levels of those variables
3. Convert the level refuse to NA, using the function Na\_if
4. Reorganise hhincome from the lowest to the highest income category using the function factor
5. Verify that you still have 23 levels after recoding using the function nlevel