

1. Cleaning the GSS 2017: "CAnD3 RRWM 1.do"
 - a. Import the GSS.csv 2017 data set
 - b. Create a dummy variable for physical health of the respondent (srh_110)
 - The categories should be good or poor health
 - Recode the missing values (7,8,9) as "."
 - For easier interpretation label the categories appropriately
 - c. Create a categorical variable for age (agec)
 - Drop respondents aged below 20 and those older than 65. Sample should be respondents aged 20 to 65.
 - Recode missing values as "." Note: According to the codebook, there should be no missing values.
 - Create a new variable named 'age group' and create the following categories:
 - 20-29, 30-39, 40-49, 50-59, 60+
 - For easier interpretation, label the categories
 - d. Create a dummy variable for marital status (marstat).
 - Recode the missing values (97, 98) as "."
 - Create a new binary variable for marital status and create the following categories
 - Single
 - Married / Living Common Law
 - For easier interpretation label the categories
 - e. Create categorical variables for main activity of the spouse (map_110)
 - Recode missing values (96,97,98,99) as "."
 - Create the following categories
 - Working, Unemployed / Job Seeking, Studying, Homemaking/Caregiving, Not in the labour force, Other
 - For easier interpretation label the categories
 - f. Create a dummy variable for spouse work status (map_130)
 - Recode missing values (6,7,8,9) as "."
 - Create the following categories:
 - Paid job or self-employed in last 12 mos
 - Not employed in last 12 mos
 - For easier interpretation label the categories
 - g. Create a categorical variable for spouse's work schedule (map_155)
 - Recode missing (96,97,98,99) as "."
 - Create new categories and label them as follows:
 - Day shift/Standard, Evening/Night shift, Rotating/Irregular shift, Split shift or other
 - h. Create a dummy variable for respondent's work status (worklyr)
 - Recode missing values as "." Note there should not be any missing values according to the codebook.
 - Create and label the following categories

- Worked last yr
 - Did not work last yr
 - i. Create a dummy variable for sex (sex)
 - Recode missing values as “.” Note there should not be any missing values according to the codebook.
 - Create / label the following categories
 - Female
 - Male
 - j. Create a categorical variable for respondent’s level of education (ehg3_01b)
 - Recode the missing values (97,98,99) as “.”
 - Create and label the following categories
 - Less than a high school degree
 - High school degree or equivalent
 - College or trade certificate
 - University degree or equivalent certificate
 - k. Create a categorical variable for family income (famincg2)
 - Recode the missing values as “.” Note there should not be any missing values according to the codebook.
 - Create and label the following categories
 - Less than \$25k
 - \$25k-\$49,999k
 - \$50k-74,999k
 - \$75k-99,999k
 - \$100k +
 - l. Create a categorical variable for average hour worked per week (uhw_16gr)
 - Recode missing values (7,8,9) as “.”
 - Note: valid skips (6) here are our unemployed category - worklyr shows us that those who said they did not work in the past year are unemployed (7087) - this same number is present in our uhv_16gr variable where valid skips are 7087 respondents = they are our unemployed category. labelling them as unemployed rather than missing
 - Create and label the following categories
 - 0 hrs
 - Less than 30 hrs
 - 30-40 hrs
 - 40.1-50 hrs
 - "50.1+ hrs
 - Unemployed
2. Create the following tables:
- a. Using the summarize function summarise all variables that were cleaned up in stage 1 (summary descriptives table)
 - b. Create a cross-tabulation of physical health by spouse’s main activity (row percentages)

- c. Create a cross-tabulation of physical health by spouse's employment status (row percentages)
 - d. Create a cross-tabulation of physical health by a spouse's work schedule (row percentages)
- 3. Create the following regression on the binary of self-reported physical health using logistic regression
 - a. Physical health & spouse's main activity (make sure to control for all variables that were cleaned up, except for spouse employment status in last 12 mos, spouse work schedule)
 - b. Physical health and spouse employment status in last 12 mos (make sure to control for all variables that were cleaned up, except for spouse's main activity, spouse work schedule)
 - c. Physical health and spouse work schedule (make sure to control for all variables that were cleaned up, except for spouse's main activity, spouse employment status in last 12 mos)