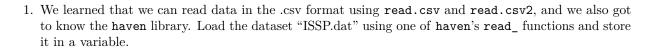
Using the tidyverse

Data Manipulation



- 2. The ISSP dataset consists of many single observations from many countries. We are only interested in observations from Germany. The variable V3 signifies the country, the value 276.1 stands for West Germany, the value 276.2 for East Germany. Using dplyr's filter(), select only the rows containing observations from Germany. Store the resulting table in the variable "Germany_raw".
- 3. Now, we want to do some subsequent data manipulation.
 - a. First, the dataset still consists of unnecessarily many variables. We only want to look at the variables "sex", "age", and "DE_RINC" (Real Income). Use dplyr's select() to select only the variables in question.
 - b. Looking at the result, it becomes apparent that the "DE_RINC" variable is stored as a string. Since we want to work with numbers, we need to convert the variable to a numeric format. Use mutate() to redefine the "DE_RINC" variable by calling as.numeric() on it.
 - c. We have also noted that we cannot really infer anything from our sex variable, since it only consists of 1s and 2s. Use mutate() and factor() to convert sex to a factor with the levels / labels 1="m", 2="w".
 - d. Lastly, we want to create a dichotomous variable "highInc" that indicates whether the person in question has a monthly real income of 5000 or more. Use mutate() in combination with if_else (or if you like: case_when()) to create this variable.
 - e. Store the resulting table in a variable called "Germany"

^{4.} Have a look at the share of men and women who have a "high income". Of all men, what is the percentage of men with a high income? And of all women, what is the percentage of women with a high income? Use group_by() and summarise() to calculate the mean of the "highInc" variable within each group. Finally, arrange the resulting table by your percentage variable (ascending).