

R Worksheet 1: Data Management

Please document your code for answering the following questions in an R script and check that your code compiles.

For all of the following exercises, use the dataset mgus2.xlsx

Exercise 1: importing and checking data

- Import the file mgus2.xlsx from your working directory into R.
- Generate a list of the variables used in the data you imported.
- Familiarize yourself with the variable names and find out what they stand for, using the R Documentation page <https://www.rdocumentation.org/packages/survival/versions/3.1-8/topics/mgus2>.
- Make sure that R interprets all categorical variables and all continuous variables accordingly, as this does not happen automatically in all cases.

Exercise 2: modifying data

- Generate a sub dataset that only contains female patients.
- Save the sub dataset with only female patients as a new file (.xlsx) in your working directory.
- For analysing your data, you will not need the ID. Delete the variable from the dataset.
- Add a new categorical variable from the variable "fuptime" called "fu_cat" that distinguishes between (1) less than 5 years, (2) 5 – 10 years, (3) 10 – 15 years, (4) over 15 years. Make sure your new variable is a factor.
- Add a new categorical variable with your data from the variable "age" called "age_qt" that groups patients into four categories so each category holds 25% of the data (Tip: use quartiles). Make sure your new variable is a factor.