

Reading in Data

1. Exercise

- Create a matrix of 1:676 with dimensions 26 x 26
- Add row names and column names of `a-z` and `a-z`
- Find the six numbers that relate to these six positions in the matrix:
(`'i','f'`) (`'i','t'`) (`'i','s'`) (`'t','o'`) (`'b','e'`) (`'s','o'`)
- Write this matrix to a comma-separated values (.csv) file. Keep the column names; drop the row names.

2. Exercise

- Read in the file `df_unknown_miss.csv`. Can you figure out what might be the missing value? (HINT: use a scatter plot)
- Re-read the data telling R what the missing value is.
- Are there perhaps any data entry mistakes in one of the variables? (HINT: use a scatter plot)
- Replace the data entry with the mean of the other data points in the variable. (HINT: use `which()`, with a logical expression as a parameter, to find out which value to replace)
- What is the mean and the standard deviation of each variable?

3. Exercise

- Load in data: `testsemicolon.txt`, `Header = TRUE` Separated by `';`
- Plot a graph of weight against price. (USE: `plot(x,y)`)
- Create a histogram of weights. (USE: `hist()`)

4. Exercise

The file 'ExamResults.xls' contains Geography and History exam results for 53 pupils. The results for the two subjects are on two separate worksheets in the file.

- Read in the file to R using the RODBC program.
- Combine the data from the two worksheets into a single data frame showing the name, age and combined exam results of each pupil.