# R: In-Class Activity 6

For this activity, you'll create a .Rmd file and output a PDF file. There is no need to upload these files. This is a chance for you to practice with your instructor (and other students) present to obtain help.

Notes: - Although we aren't turning this in, you should use good programming practices (see R File Submission Guidelines).
- Provide a brief narrative using markdown throughout the program explaining the upcoming code's purpose.
- In your code chunks, use spacing and indentation to improve readability.

The goal of this activity is to practice summarizing data using contingency tables and other common numeric summaries.

## Dataset:

The dataset for this activity is available at https://www4.stat.ncsu.edu/~online/datasets/bankData.txt and has information about direct marketing campaigns of a Portuguese banking institution.

The variable names and descriptions are given at the site: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

## Programming questions

Write a brief narrative and code to answer the questions below.

1. Create code to import the bankData.txt file using functions from the tidyverse and save it as an R object. Note: The delimiter is a semicolon and the column names are included in the raw data file.

2. Create a new R object that takes the dataset from above and does the following:

   a) Renames the y variable to be more meaningful
   b) Subsets the observations to only include rows where age is less than 75 and marital is not "unknown."

From this point forward, use your newly created dataset.

3. Create a two-way contingency table as seen below. Interpret the number in the top left.

   ```
   ##
   ##              no   yes
   ##   no      29939  3722
   ##   unknown   879   105
   ##   yes      5533   661
   ```

4. Create a three-way contingency table as seen below. Interpret the number in the top left of the first table.

   ```
   ## , ,  = divorced
   ##
   ##
   ##              no   yes
   ##   no       3382   354
   ##   unknown   112     7
   ##   yes       596    63
   ##
   ## , ,  = married
   ##
   ##
   ##              no   yes
   ##   no      18405  2026
   ##   unknown   526    59
   ```

```
##   yes        3385    362
##
## , ,  = single
##
##
##              no    yes
##   no       8152   1342
##   unknown   241     39
##   yes      1552    236
```

5. Replicate the summary statistics created below.

```
##       age             duration      emp.var.rate
##  Min.   :17.00   Min.   :   0   Min.   :-34.000
##  1st Qu.:32.00   1st Qu.: 102   1st Qu.:-18.000
##  Median :38.00   Median : 179   Median : 11.000
##  Mean   :39.76   Mean   : 258   Mean   :  1.086
##  3rd Qu.:47.00   3rd Qu.: 319   3rd Qu.: 14.000
##  Max.   :74.00   Max.   :4918   Max.   : 14.000
```

6. Replicate the finding of the minimum, 1st quartile, sample mean, median, 3rd quartile, and maximum for the duration variable. This was done for every combination of the marital status and loan variables (i.e. for all subgroups created by these two variables).

```
## # A tibble: 9 x 8
## # Groups:   marital [3]
##   marital  loan     minimum    Q1  mean median    Q3 maximum
##   <chr>    <chr>      <dbl> <dbl> <dbl>  <dbl> <dbl>   <dbl>
## 1 divorced no             0 102    254.   180   315     3253
## 2 divorced unknown        8  94.5  210.   164   266.    1120
## 3 divorced yes            7  98.5  258.   176   320     2139
## 4 married  no             0 102    257.   179   318     4199
## 5 married  unknown        8 102    247.   167   308     2926
## 6 married  yes            3  99.5  258.   177   310     3322
## 7 single   no             1 104    262.   184   328.    4918
## 8 single   unknown        7 104    274.   178   317     1580
## 9 single   yes            3  98    259.   174.  311.    3076
```

Knit this to a PDF and you're done! Great work!