CT5102: Programing for Data Analytics 2018/19

Assignment 4: Manipulating Data Frames (10 marks)

Given the following vectors:

```
set.seed(1000)
ids    <- rep(as.character(1001:1005,2))
module <- c(rep("CT101",5),rep("CT102",5))
result <- c(rnorm(n = 5,mean = 70,sd = 5),
            rnorm(n = 5,mean = 50,sd = 8))
```

Create this data frame:

```
> dataf
    ids module   result
1  1001  CT101       NA
2  1002  CT101 63.97072
3  1003  CT101 70.20563
4  1004  CT101 73.19694
5  1005  CT101 66.06723
6  1001  CT102 46.91609
7  1002  CT102 46.19306
8  1003  CT102 55.75801
9  1004  CT102 49.85196
10 1005  CT102 39.01506
```

Next, write an aggregation function for a data frame with the following function definition.

```
my_aggregate <- function(df, group_id, data_id, f, ...){}
      # df is the data frame
      # group_id stores the column name (grouping variable)
      # data_id is the data column name (data to be aggregated)
      # f is an R function to be applied to the data
```

The following validity checks should be performed:

- **df** must be a valid data frame
- **group_id** must be a valid categorical/string column name belonging to the data frame
- **res_id** must be a valid numeric column name belonging to the data frame
- **f** must be a valid function

Sample output is as follows:

```
> my_aggregate(dataf, "module", "result", mean)
   CT101    CT102
      NA 47.54683

> my_aggregate(dataf, "module", "result", mean, na.rm=T)
   CT101    CT102
68.36013 47.54683

> my_aggregate(dataf, "ids", "result", mean)
    1001     1002     1003     1004     1005
      NA 55.08189 62.98182 61.52445 52.54114

> my_aggregate(dataf, "ids", "result", mean,na.rm=T)
    1001     1002     1003     1004     1005
46.91609 55.08189 62.98182 61.52445 52.54114
```

Examples of error checking include (hint the function **class()** can be useful to detect data frame objects as they are S3 classes).

```
> my_aggregate(1:10, "module", "result", mean, na.rm=T)
Error in my_aggregate(1:10, "module", "result", mean, na.rm = T) :
  First parameter is not a data frame object


> my_aggregate(dataf, "modul", "result", mean)
Error in my_aggregate(dataf, "modul", "result", mean) :
  Error modul is not a valid column


> my_aggregate(dataf, "module", "ids", mean)
Error in my_aggregate(dataf, "module", "ids", mean) :
  Error ids is not a numeric column

> my_aggregate(dataf, "module", "result", 10)
Error in my_aggregate(dataf, "module", "result", 10) :
  Error 10 is not a function
```