**R Proficiency Test**

1. **(5 pts.)** Write a function fun1 that receives as input a number x. If x is a multiple of 2, print the word "eeny" to the console. If x is a multiple of 3, print the word "meeny." If x is a multiple of 5, print the word "miny." Otherwise, print the word "moe."
2. **(5 pts.)** Write a function fun2 that receives as input a vector v of numbers and returns the number of even numbers in this vector. For full credit, use the apply functions.
3. **(10 pts.)** Write a function fun3 that receives as input a list l and returns the number of words (character strings) in this list starting with the letter "R" (or "r"). Note that the list may contain elements of different data types. For full credit, use the apply functions.
4. **(10 pts.)** Write a function fun4 that receives as input four numbers w, x, y, and z. Your function should create a matrix mat1 with x rows and y columns such that every element in mat1 is equal to w, and a matrix mat2 with y rows and x columns such that every element in mat2 is equal to z. Your function should then combine (by rows) mat1 with the transpose of mat2 and return the resulting matrix.
5. **(10 pts.)** Suppose you have a data frame diamonds that looks as follows (only the first three rows are shown):

carat cut color clarity depth table price x y z

0.23 Ideal E SI2 61.5 55 326 3.95 3.98 2.43

0.21 Premium E SI1 59.8 61 326 3.89 3.84 2.31

0.23 Good E VS1 56.9 65 327 4.05 4.07 2.31

...

Write a function fun\_diamonds that receives as input data frame diamonds and returns the average price of diamonds with "Ideal" cut.

1. **(5 pts.)** Suppose you have a data frame temperatures that looks as follows (only the first three rows are shown):

temp

75

80

55

...

* 1. **(2.5 pts.)** Write a function temp\_hist that receives as input data frame temperatures and plots a histogram of the data. Your function should specify the number of breaks (5 breaks), the title ("Histogram of temperatures") and the axis label ("Temperatures").

* 1. **(2.5 pts.)** Write a function temp\_boxplot that receives as input data frame temperatures and creates a boxplot of the data. Your function should specify the title ("Boxplot of temperatures") and the axis label ("Temperatures").

1. **(10 pts.)** Suppose you have three data frames dat1, dat2, and dat3 that look as follows (only the first three rows of one of the data frames are shown):

x y

10 33

42 24

57 15

...

Write a function fun\_plot that receives as input data frames dat1, dat2, and dat3, and plots them on the same plot. Data points from dat1 should be plotted as red circles. Data points from dat2 should be plotted as black diamonds. Data points from dat3 should be plotted as blue crosses. Your function should also add a title ("Plot of three datasets"), axis labels ("x" and "y"), and a legend to the plot.

1. **(45 pts.)** Given the following files in csv format containing student grades for a class (with headers):

**student\_data\_01.csv:**

firstname,lastname,gender,id,age,a1,a2,a3,a4,a5,a6,a7,a8,a9,a10

Wallace,Barton,M,236290,14,90,89,87,89,86,89,96,95,95,89

Ben,Massey,M,682582,16,83,85,84,82,83,84,84,83,83,83

Angelica,Johnson,F,864083,13,83,77,85,83,89,79,91,79,91,84

Bobbie,Stanley,M,347045,13,87,89,86,89,88,87,95,87,89,88

Rudolph,Simmons,M,106369,13,76,94,76,99,72,76,80,74,87,77

Joseph,Simpson,M,979705,15,86,91,97,89,95,82,82,78,87,76

Crystal,Armstrong,F,815414,16,86,84,86,76,81,87,85,91,81,81

Cheryl,Silva,F,382579,15,91,91,86,99,89,90,96,88,95,40

Joe,Patrick,M,542978,16,88,100,80,84,97,78,83,89,81,91

Kristen,Lamb,F,925099,16,85,80,84,90,91,88,85,84,81,84

**student\_data\_02.csv:**

lastname,firstname,gender,id,age,a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11

Russell,Doris,F,861836,14,89,96,97,78,94,97,86,92,90,99,90

Wong,Whitney,F,184294,14,74,75,76,71,68,83,67,77,72,75,76

Mendez,Bernadette,F,581916,13,72,68,73,75,84,72,68,77,65,71,70

Colon,Ronald,M,599992,16,73,69,69,68,67,70,70,72,74,72,71

Pratt,Albert,M,431907,14,82,87,94,85,100,91,90,92,91,93,92

Johnson,Angelica,F,864083,13,83,77,85,83,89,79,91,79,91,84,78

Massey,Ben,M,682582,16,83,85,84,82,83,84,84,83,83,83,77

Webb,Nettie,F,882618,16,78,80,77,77,80,79,76,78,78,75,76

Crawford,Gertrude,F,425515,14,87,73,82,88,79,79,84,82,78,84,80

Gonzalez,Casey,M,732776,13,78,89,99,95,89,85,94,94,98,97,95

* 1. **(15 pts.)** Write a function read\_data that receives as input the name of the two files ("student\_data\_01.csv" and "student\_data\_02.csv"). This function should read the two files, merge their contents into a single data frame df\_merged, write data frame df\_merged to a file in csv format ("student\_data\_merged.csv") and then return data frame df\_merged.

Note that any student who is missing a grade for an assignment should be given a value of NA for that assignment. If a student is in both datasets, ensure that their grades for each assignment match; if a grade for an assignment does not match, or if the student is in both datasets but is missing a grade for an assignment in one of them, then assign a value of NA for that assignment. Also, ensure that the data types in the final data frame are correct (i.e., grades, ID, and age should be of type numeric).

* 1. **(10 pts.)** Write a function compute\_mean\_median that receives as input the data frame from problem 8.1 (df\_merged). This function should add two columns to the data frame containing the average and the median grade for each student, respectively, and then return the modified data frame (df\_merged\_new). Ensure that NA values are being handled appropriately (i.e., not included in the calculation). For full credit, use the apply functions.
  2. **(5 pts.)** Write a function compute\_mean\_gender that receives as input the data frame from problem 8.2 (df\_merged\_new). This function should calculate the mean of the average grades for male students and the mean of the average grades for female students. Print both means to the console (e.g., "Male students have a mean grade of ...").
  3. **(10 pts.)** Write a function compute\_letter\_grades that receives as input the data frame from problem 8.2 (df\_merged\_new). This function should return a new data frame df\_letters, which, instead of numeric grades, contains letter grades in a 10 point scale (100-90=A, 89-80=B, etc.).
  4. **(5 pts.)** Write a function count\_letter\_grades that receives as input the data frame from problem 8.4 (df\_letters), finds the total counts for each letter grade (count only the letter grades corresponding to the average grades computed in problem 8.2), and prints the results to the console.