Homework 1

All homework must by typed. Include all relevant R code your submission. Typed responses to questions should be mixed inline with R code. For this assignment, copying and pasting R code into a word document or .txt file is fine. Upload the completed homework assignment into the course webpage dropbox.

1. Lab: Note there is **no** lab to review for week 1.
2. Reading (accessed via springerlink on the UW library website)
   (a) Data Manipulation with R
      i. Chapter 1
      ii. Chapter 2.1-2.3
   (b) A Beginner's Guide to R
      i. Chapter 1.4, 1.8, 1.9
      ii. Chapter 2.1
      iii. Chapter 3.1,3.2
3. Describe some kinds of data that are of interest to you, either professionally, personally or (if nothing comes to mind) something that is publicly available. Describe briefly some variables of interest, i.e. identify them as categorical or quantitative, what the ranges and possible values of the data area (in statistics we call this the "support" of the variables), and the relationships there may or may not be between them. Are there some that are more obviously *dependent* or *response* variables, and others that might be considered *explanatory*? This exercise is a very preliminary opportunity to start thinking statistically, and to brainstorm for an eventual final project.
4. Create a vector of all the *even* numbers from 2 to 100 using the constraints given below:
   (a) using the `seq()` function.
   (b) using the ':' operator.
   (c) using a combination of only the `cumsum()` and `rep()` functions. We haven't talked about these functions in class, so first learn about them using `?cumsum` and `?rep`.
5. In statistics we perform a lot of sums, which R's native vector-friendly behavior makes quite easy. Generate some vector of numbers X (for example the one in the problem above). Note that for a vector X, the ith element of X is denoted X_i (where $i = \{1, 2, 3...n\}$). Compute the following quantities. Create whatever intermediate variables you might need. The 'sum' and 'length' functions may be useful here.
   (a) $\sum_{i=1}^{n} X_i$
   (b) $SS = \sum_{i=1}^{n} X_i^2$
   (c) $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
   (d) $\sum_{i=1}^{n} (X_i - \overline{X})^2$
   (e) $SS - n\overline{X}^2$

Optional: *Note that two of the quantities above should be equal. Identify those two quantities and prove using algebra that this will always be the case.*

6. Copy and paste the following code into `R`, which will generate a list of names, sex, and grades for students in a class.

```
Names <- c("Alana", "Bettie", "Consuela", "Dona", "Elaine", "Frances",
           "Gerri", "Helene", "Ichabod", "Jin", "Kenyatta", "Larry",
           "Mikhailo","Nick","Odin")
Sex <- c(rep("F",8), rep("M",7))
Grades <- round(runif(15,50,100))
```

   Note that each time the `Grades` object is created, they are different, so make sure you only create it once!

   (a) Use a logical expression to create a logical vector indicating which elements of Grades are greater than or equal to 90%. Assign that logical vector to an object named `gt90`.

   (b) Using the logical vector you created above and the bracket syntax discussed in lecture, print only the names of the students whose grades were greater than or equal to 90 (assuming that elements of Names and Grades correspond to each other, in order).

   (c) Extract the names of the students (if any) with grades under 60%. Unlike in part b, do this *without* first assigning a logical vector to a named object.

   (d) Extract the grades of male students and the female students and create new objects called `Grades.M` and `Grades.F`.

   (e) Use the `sum()` and `length()` functions to obtain average grades of the male and female students. Which have higher grades?

7. Basic R coding

   (a) Look at the help file for the `rm()` function. The first argument listed as "..." accepts an arbitrary number of object names to be removed. I.e, to remove variables `x` and `y` simultaneously we could type `rm(x,y)`. Alternatively, this function also allows us to remove variables given a character vector corresponding to object names. What argument of `rm` would we have to use to get `rm` to remove variables given a character vector of names? How would we easily remove all variables in the global environment? (Hint: which function returns a character vector of objects in the global environment?)

   (b) Explain why `c(0,2) + c(3,4,5,6)` returns `3 6 5 8`. What's the term used in R to describe this phenomenon?

   (c) What is the main difference between a vector and a list?

   (d) What two functions did we discuss in lecture which return characteristics of an object?

   (e) Where in a help file might you find a description of the type of object a function returns?