Assignment 1
Submission deadline: Friday Sep 9th 2016 by 8 PM
Submission format: upload document in Canvas

1. **(20 points)** Using R, compute the percentage of your life that you have spent at Virginia Tech. Report R code and output.

   Hint: Here is the mathematical formula:

$$\text{Pct} = \frac{(\textit{current year} - \textit{admission year}) \times 12 + (\textit{current month} - \textit{admission month})}{(\textit{current year} - \textit{birth year}) \times 12 + (\textit{current month} - \textit{birth month})} \times 100$$

   For example, if you were born in March 1995 and started at Virginia Tech in August 2014, then in August 2016 the above formula gives you

$$\text{Pct} = \frac{(2016 - 2014) \times 12 + (8 - 8)}{(2016 - 1995) \times 12 + (8 - 5)} \times 100 = 9.41$$

2. **(20 points)** Using R, compute the first 1000 Fibonacci numbers starting with 1, and plot the ratios $\frac{x_{n+1}}{x_n}$ for n=1,2,...,999. Report R code and R plot.

   Hint: The first and second Fibonacci numbers are $x_1 = 1, x_2 = 1,$ and successive numbers can be generated by the formula $x_{n+1} = x_n + x_{n-1}$, for example $x_3 = x_2 + x_1 = 1 + 1 = 2, \quad x_4 = x_3 + x_2 = 2 + 1 = 3,$ and the sequence goes as 1,1,2,3,5,8,....

3. **(20 points)** From UC Irvine's machine learning depository, consider the Iris data set at http://archive.ics.uci.edu/ml/datasets/Iris.

(a) Using your own words, describe the dataset in 2-3 sentences.
(b) Import the data in R, and convert the data set into a data frame. Define column names for the data frame using the "Attribute Information" from the above webpage.
(c) For each "class", calculate the mean and standard deviation of sepal length, sepal width, petal length, and petal width. Write this summary information into an R matrix called "summary". Report this summary matrix.
(d) Save the iris data frame and the summary matrix in an R workspace called "iris.RData". Report R code and answers.

4. **(20 points)** Install the R package "babynames". Load the babynames data and answer the following questions. Report R code and answers.

(a) Describe the dataset in two sentences. How many rows and columns does the dataset have?
(b) How many unique names are there in the dataset? Why is this number different from the number of rows in (a)?

(c) What were the most popular male names for the years 1900, 1925, 1950, 1975, 2000? What were the most popular female names for the years 2010, 2011, 2012, 2013, 2014?

5. **(20 points)** What are the 10 most popular male baby names across years? What are the 10 most popular female baby names across years?

Assignment instructions:

1. **Honor code:** The Virginia Tech honor pledge for assignments is as follows:
   "I have neither given nor received unauthorized assistance on this assignment."

   The pledge is to be written out on all graded assignments at the university and signed by the student. Type up your name to sign.

2. Submit your assignment as a document (word, pdf or similar) to Canvas, clearly marked with student's name and assignment number, eg. Sengupta_Srijan_HW1.pdf. Your submission should include R code and answers to problems.

3. Late assignments will not be accepted. Check Canvas regularly for assignments and submission dates.

4. You are free to discuss assignment problems with your classmates, but submitted work (answers and codes) **must** be your own work. Students are not allowed to copy computer codes or answers from each other, and must write their own codes and answers.