Stat 123 Homework Assignment 2
Due Friday February 18th by 9:00pm

Using R Markdown, please complete the following assignment. If an answer does not require any R code, you can type the answer to the question outside of a chunk. Make sure that your assignment is well labelled so that it is clear where each question's answer begins. Your assignment should be submitted as a pdf (whether you knit directly to PDF, or knit to HTML or Word and then convert the file to a pdf).

1. The built-in Titanic data set is a 4-dimensional array that contains the following information:

   - Dimension 1: Class of the passenger (1 = 1st, 2 = 2nd, 3 = 3rd, 4 = Crew member)
   - Dimension 2: Sex of the passenger (1 = male, 2 = female)
   - Dimension 3: Age of the passenger (1 = child, 2 = adult)
   - Dimension 4: Survival of the passenger (1 = died, 2 = survived)

   If you wanted to determine, for example, how many male, adult, crew members survived, you could type in Titanic[4,1,2,2] to get this value. If you wanted to create a table with how many 1st class passengers (of all genders and ages) died, you could type Titanic[1, , ,1].

   (a) Create (and print out) a table which contains the adult passengers (of all classes and genders) who survived.

   (b) Create (and print out) a vector called *survived* which contains the adult passengers who survived. Hint: You may need to use rowSums() on your answer from part (a).

   (c) Create a barplot displaying the *survived* vector. Make sure to include a main title and to label your x-axis. Also, make sure that each bar is a different colour.

   (d) What does the bar graph imply about the survival of adult passengers based on class?

   (e) Create (and print out) a vector called *died* which contains the adult passengers who did not survive.

   (f) Create (and print out) a vector called *percent.Survived* which contains the percentage of adult passengers who survived in each class.

   (g) Create a pie chart that displays the *percent.Survived* data. Be sure to include a main title for your pie chart.

   (h) What does the pie chart imply about the survival of the adult passengers based on class? Does this imply something different than the bar graph did? If yes, why?

2. The following question deals with the data set *NHLData.csv* which you will need to download from the assignment page.

   (a) Create (**but do not print**) a vector called *points* containing the number of points for each player (the variable P in the data set).

   (b) Create a histogram displaying the distribution of this variable. Be sure to have both a main title and a title on your x-axis. Also, make sure that the scale on the x axis goes to 30 and the scale on the y axis goes to 400.

   (c) Describe the shape of the distribution (symmetric, left-skewed, right-skewed).

   (d) What is an appropriate measure of the center of the distribution (mean or median), why?

   (e) Compute the appropriate center value and the corresponding measures of variability.

3. (You may wish to do this question by hand) Consider the following sample of points from the NHL data set:

```
19
20  sample(nhl$P,20)
21
22 ```
```

```
 [1]  3  1  8  5  3  1  2  0  5  2  0  1  3  3  2  1 11  0
[19]  6  1
```

   (a) Create a stemplot of the distribution of the sample.

   (b) Does the distribution resemble the one seen in question 2? Explain why there might be some differences.