STAT 158 Homework 4: Plots and loops Due Friday March 29th

Your solutions should contain all the code you ran and any relevant R output (**including plots**), copy / pasted from R Studio. Solutions should be numbered so that it is clear which code and output go with which question. The pdf you turn in should thus show a record of you doing all the things the homework questions ask you to do.

1. Using the iris dataset in R, make a scatterplot with petal width on the x-axis and petal length on the y-axis. Distinguish between species by making the species have different colors and symbol types. Give your plot an appropriate title and appropriate axis labels.

2. Generate random values from three different normal distributions, each with a different mean and standard deviation. Pick whatever values you like, and whatever sample sizes you like. Then make a boxplot that should have the following features:

    a. Three boxes are plotted side by side, one for each distribution (note: boxplot(a,b,c) will plot a, b, and c next to each other).
    b. Each box is colored a different shade of grey
    c. Each box has a meaningful label below it. You can use the "names" plotting argument to give each boxplot a name that identifies its mean and standard deviation (e.g. "mean=0, sd=2").
    d. The numbers on the y axis and labels on the x axis are displayed in a smaller size than R's default size.
    e. There is not a full square box drawn around the plot; only the horizontal and vertical axes are drawn with lines.

3. In the notes there is a Central Limit Theorem example in which a sampling distribution of means is created using a for loop, and then this distribution is plotted. This distribution should look approximately like a normal distribution. However, not all statistics have normal sampling distributions. For this problem, you'll create a sampling distribution of standard deviations rather than means.

    a. Using a for loop, draw 10,000 samples of size n=30 from a uniform distribution with min=100 and max=300. For each sample, compute the standard deviation. Make a histogram of the distribution of standard deviations. Give your plot appropriate labels.
    b. Repeat the procedure in a., but this time draw samples of n=50.
    c. Repeat the procedure in a, but this time draw from a uniform distribution with min=100 and max=500.
    d. Redo the plots from a, b, and c, this time making the x and y axes for each plot have the same scale, so that they are easy to compare to one another. What effect did changing the sample size have on the sampling distribution of the standard deviation? What effect did changing the max value for the uniform distribution that the samples were drawn from have on the sampling distribution of the standard deviation?

4. Create a vector whose values are the logarithms of a random sample of size 30 from a standard normal distribution (which has mean = 0 and sd = 1) Use the ifelse function to replace all the missing values with 999. Use cbind to show the original normally distributed values, the log transformed values, and the log transformed values with the missing ones replaced all next to each other.