Assignment 4
Exercises: Data extraction Wednesday Feb. 12, 2014
Due date: Fri. Feb 21, 2014

During the course of the class, problems will become increasingly word-based with decreasing help on how to use R to accomplish your objectives. The hope is that over time, you will become comfortable transforming word probllems (that do not include any information on R) into solutions derived using R. You must eail the script that generates your results to Dan Smith together with your report. Your report contains text that explains what you did and charts. You can do screen dumps of the charts, or ask Dan to help you create jpeg or png files of your charts.

As usual, you must do all your work in a directory with a name related to this assignment.

In this assignment, you will solve a few data extraction problems and create some plots. These skills are crucial in all future problems.

Question 1: 20 pts
Read the following datasets built into R and provide their class: "Titanic", "ToothGrowth", "UKgas", "infert", "VADeath". Provide a brief description of each dataset. Convert each dataset to a dataframe.

Question 2: 30 pts
(a) Read in the file "Maze_UniversityOfIllinois.csv" (found on the course website), and create a dataframe of all women who participated in the experiment.
(b) Using the same data, extract from the file, a data.frame which contains data for all the men who participated in the experiment, and who completed the maze on the 5th trial in less than 10 seconds. What was the mean and standard deviation of the times? What was the average error rate?
(c) Using the same data, transform the age column into a factor, and create two plots: the left plot shows the timings of the mean time to completion of the 10th run through the maze for women, and the right plot shows the timings of the mean time to completion of the 10th run through the maze for men. In other words, for each age, compute the mean time to compute the 10th run through the maze, and plot this mean time as a function of the age.) (Hint: use the `plot` command in the form `plot(vector  factor)` to create the graph. The vector is the time vector (for men), and the factor would be the age. The result will be a box plot. Please use Google to find out how to interpret a box plot and explain this in your lab report. You will need a separate plot for men and women.)

Question 3: 20pts
(a) Read the dataset "WorldPhones" from within R. What type of object is returned? Convert this object to a data.frame. Plot the number of phones over the 7 continents as a function of the year. Put all the plots onto the same graph. Add labels and a title.

(b) Create a separate chart that shows the total number of phones in the world between 1951 and 1961. Create appropriate labels and titles.

Question 4: 30pts
(a) Read the data "quakes" from within R. What does this data represent? Please provide a description in your own words after reading the help data on it.
(b) How many different magnitudes taken by the earthquakes (you are not allowed to count by hand.)
(c) Plot the longitude versus latitude (geographic location) of the earthquakes of magnitude less than 5, in red. Plot the longitude versus latitude of the earthquakes of magnitude greater or equal than 5 in blue. Put both plots on the same graph.
(d) Is the mean depth of earthquakes of magnitude less than 5.0 significantly different than the mean depth of earthquakes of magnitude greater or equal than 5.0? (Of course, explain how you reach your conclusion.) This is a problem of Null Hypothesis.