140.886: Intro to R

Homework 3

Assigned Date: 1/10/2013

Due Date: 1/11/2013

Name:

Please submit:

1. Annotated R analysis script (emailed to me)
2. Answers written/typed on this sheet (turned in tomorrow morning)

For (a), the script should be source-able (ie have no errors) and divided into sections (corresponding to questions) using comments. The answer to the question should be the last line(s) of code evaluated in that section. For example:

#####

# q1

[code goes here]

[have answer print here]

Instructions:

1. Use the mortality dataset from Lab 2 for the following questions:

**http://biostat.jhsph.edu/~ajaffe/files/indicatordeadkids35.csv**

Questions

1. Create a integer ‘year’ variable using the column names, and print the head() and class() of this new vector
2. Transpose the dataset such that countries are the columns and years are the rows. Append the ‘year’ vector from above. Show [1:5,1:5] of this transposed data frame.
3. Read in the ‘Data’ sheet in the xlsx file: [**http://biostat.jhsph.edu/~ajaffe/files/country\_pop.csv**](http://biostat.jhsph.edu/~ajaffe/files/country_pop.xls)
4. Determine the population of each country in our dataset, and then sort the columns of our dataset by that population, from smallest to largest. Show [1:5,1:5] of this sorted data frame.
5. Plotting: we talked about lattice::levelplot() but R has two similar plots called image() and heatmap(). Create the following levelplots using lattice::levelplot() and image():
	1. Years 1840-1900 for countries that start with the letter ‘A’
	2. Years 1950-1975 for the 10 largest countries and the 10 smallest countries
6. Plotting: create “spaghetti”/line plots (one line per country) using lattice for Years 1975-2010 for countries that have between 7 and 10 characters (including spaces) in their names.