

## Lab 4

Assigned Monday Feb. 8, 2015

Due before lab Feb. 15, 2015

Return all scripts, plots, and results. Everything should be included in a single .zip file returned to the TA. The scripts should be in separate files. Name your working directory appropriately. Include your last name in the name of the working directory. For example: "lab04\_slice.zip".

The objective of this lab is to familiarize yourself more with the creation of graphs using the plot() and hist() functions. To this end, we will use a mixture of contrived data and data.frames that will be obtained from the data() command. As usual, any concepts and functions not encountered before should be investigated using help(), Google, and then the TA.

All functions should have "reasonable" names.

Scoring: each question has equal weighing.

### Task 1

You have created a survey of 170 people, asking for their age and weight. After some analysis, you have determined that their ages lie between 35 and 57, and their weights between 150 and 190 lbs. Since you do not have the actual data, you will construct a "fake" dataset using random numbers. Thus, we will assume that the ages are samples from a Normal distribution with mean 45 and standard deviation of 8, while the weights are taken from a normal distribution with mean 170 and standard deviation 15. Create a plot of weight (ordinate, vertical axis) versus age (abscissa, horizontal axis). Create appropriate labels for the two axes. Provide your plot with a title. Create a histogram of the weight data labelled and titled appropriately.

NOTE: We ignore the fact that this would probably not be a good model for weight. In the US, mean male body weight is closer to 195lbs and for females, 165. Using a mean of 170 might suggest a mixed sample of males and females, but then randomly sampling from a single, normal distribution would likely be inappropriate. We ignore this, but this is the type of thing you should question when someone presents you with data.

### Task 2

Now we create a second survey, this time with 130 participants asking just for their age. The second time around, we find that the ages lie between 35 and 57, and follow a normal distribution with mean 40 and standard deviation of 10. Again create a fake dataset for this second age group. Starting with the dataset created for Task 1, create a different plot, with two curves (one red and one blue) corresponding to each dataset. On the horizontal axis, I'd like the ID number corresponding to the person questioned. This ID ranges from 1 to 170 for the first dataset, and from 1 to 130 for the second dataset. Create a plot of two histograms stacked atop each other showing the distribution of the ages and identically scaled.

HINT: use the plot() command for the first line plot, and the lines() command for the second plot to make sure that it overlays the first instead of replacing it.)

### Task 3

Create a function with a single argument (n) that creates a vector of n random numbers and plots their square against their original values. Superimpose on this plot another scattergram that consists of the cube of these numbers against their values. Both plots should be created within the same function and

in different colors. Make plots for  $n = 10$ ,  $n = 100$ , and  $n = 1000$ . Make sure all data points are plotted.

For the next question, I will provide a function and a plot, and you are to create the R code that will generate this plot.

#### **Task 4**

Given the function

```
read.data = function() {  
  # returns a data.frame  
  data(USArrests)  
  arr = USArrests  
  df = arr[ , c("Murder", "Assault", "Rape", "UrbanPop")]  
  return(df)  
}
```

Create a function `my.plot()` that duplicates the plot in Figure 1 (The figure is in a file called `lab4_task04_fig.pdf` and can be found on the class website, along with this assignment.) Call the function to generate the plot. Comparing the plots for Assault, Rape and UrbanPop, discuss the differences and draw conclusions regarding the influence of population size on arrests of different types. Include your plot with your report.