

Assignment 6
Data extraction Monday Feb. 22, 2016
Due: Monday, February 29, 2016

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 problems (that do not include any information on R) into solutions derived using R. You must mail the scripts that generate your results to the TA together with your report. Your report contains text that explains what you did and charts.

Your scripts should be neat using comments and white space. (easy to read by somebody other than you). Separate different sections from one another. You should have one script per question to make grading and reading easier. Please run each of your scripts to make sure it executes correctly. (The TA tests each of your scripts.)

As usual, you must do all your work in a directory with a name related to this assignment. This assignment involves reading data, data extraction, and plotting. These skills are crucial for most problems.

Question 1: 40 pts

In this problem, we will analyze data from the "Maze UniversityOfIllinois.csv" and calculate various statistics of different subsets of the data, compared to the full dataset.

- (a) Read in the "Maze UniversityOfIllinois.csv" (found on the course website in the "data" directory) and create two data frames, one for all the men, and one for all the women.
- (b) Compare the mean time run by men in the last trial, against the mean run by men (in the last trial) who completed the maze in less than 10 seconds in the 5th trial. Repeat by replacing mean time by the error rate. (You are not allowed to use the summary() function.)
- (c) Do the same for the women.
- (d) Using the same data, transform the age column into a factor, and create two plots: the left plot shows the time to completion of the 10th run through the maze for women, and the right plot shows the time to completion of the 10th run through the maze for men. The two plots should be in the same figure and should be done using the technique described in "Additional Information" below.

(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 class notes or 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 2: 30pts

- (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. Be careful of your x and y limits (recall, these are determined by the first plot you draw.)

(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 3: 30pts

(a) Read the data "quakes" from within R. What does this data represent? Please provide a description in your own words.

(b) How many different magnitudes taken by the earthquakes (you are not allowed to count by hand.) (Hint: use the function unique .)

(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. Be careful of axis scaling.

Additional Information

1. How to transform a vector into a factor: use the function `as.factor()`.

For example, given the vector `weights = c(162; 151; 175; 162; 151)`, it is converted into a vector of factors by executing `weight.factors = as.factor(weights)`. Given a dataframe named `df`, `df[,1]` and `df$heights` are vectors (assuming there is a column named "heights").

2. To create two plots side by side, execute

```
par(mfrow=c(1,2))
```

```
plot(...)
```

```
plot(...)
```

The first plot command will create a plot on the left. The second plot command will create a plot on the right. `mfrow=c(2,1)` means that the plots will be on one row and two columns. Similarly,

```
par(mfrow=c(2,1))
```

```
plot(...)
```

```
plot(...)
```

will create two plots, one above the other.

3. To convert a matrix to a dataframe, use the function `as.data.frame`. So if you have a data matrix named `mx`, `as.data.frame(mx)` will convert `df` into a data frame. It works for tables and matrices but not for time series.