

HOMEWORK 4
STAT 4410/8416 Section 001
FALL 2014
Due: October 30, 2014 by midnight

1. **Working with date-time data:** The object `myDate` contains the date and time of now as below. Based on this object answer the following questions.

```
library(lubridate)
myDate <- now()
myDate

## [1] "2014-10-22 02:58:51 CDT"
```

- (a) Write your codes so that it displays the week day of `myDate`.
- (b) Add one month with `myDate` and display the resulting date time. Explain why the time zone has changed even though you did not ask for time zone change.
- (c) This homework is due on October 30, 2014 by 11.59PM. Compute and display how many minutes are remaining for submission of this homework?
2. **Exploring XML data:** In this problem we will read the xml data. For this we will obtain a xml data called olive oils from the link <http://www.ggobi.org/book/data/olive.xml>. Please follow the directions in each step and provide your codes and output.

- (a) Parse the xml data from the above link and store in a object called `olive`. Obtain the root of the xml file and display its name.
- (b) Examine the actual file by going to the link above and identify the path of categorical variables in the xml tree. Use that path to obtain the categorical variable names. Please keep the names, not nick names and store them in `cvNames`. Display `cvNames`.
- (c) Now examine the file by going to the link and identify the path of real variables in the xml tree. Use that path to obtain the real variable names. Please keep the names, not nick names and store them in `rvNames`. Display `rvNames`.
- (d) Notice the path for the data in xml file. Use that path to obtain the data and store the data in a data frame called `oliveDat`. Change the column names as you have obtained the column names. Display some data.
- (e) Generate a plot of your choice to display any feature of `oliveDat` data. Notice that the column names are different fatty acids. The values are % of fatty acids found in the Italian olive oils coming from different regions and areas.
- (f) Explain what these two lines of codes are doing.

```
r <- xmlRoot(olive)
xmlSApply(r[[1]][[2]], xmlGetAttr, "name")
```

3. **Boston hubway data:** This question will explore Boston hubway data. Please carefully answer each question below including your codes and results.

- (a) Obtain the compressed data, `bicycle-rents.csv.zip`, from blackboard and display few data rows.
- (b) Based on the rent date column, create two new columns `weekDay` and `hourDay` which represent week day name and hour of the day respectively. Store the data in `myDat` and display few records of the data. Hint: For weekday use function `wday()`.

- (c) Summarize myDat by weekDay based on the number of rents for each weekDay and store the data in weekDat. Display some data.
 - (d) Create a suitable plot of the data you stored in weekDay so that it displays number of bike rents for each week day.
 - (e) Now we want to investigate what happens in each day. Summarize myDat again but this time by weekDay and hourDay and obtain the number of rents. Store the data in hourDat and Display some data.
 - (f) The dataframe hourDat is now ready for plotting. Generate line plots showing number of bike rents vs hour of the day and colored by weekDay.
4. **Creating HTML Page:** In this problem we would like to create a basic HTML page. Please follow the each step below and finally submit your HTML file on the blackboard.
- (a) Open a notepad or any plain text editor. Write down some basic HTML codes as shown in Lecture 15, slide 6 and modify according to the following questions. Save the file as hw4.html and upload on the blackboard as a separate file. You don't need to answer these questions here in the .Rnw file.
 - (b) Write "What is data science?" in the first header tag, `<h1></h1>`
 - (c) Hw1 solution contains the answer of what is data science. The answer has three paragraphs. Write the three paragraphs of text about data science in three different paragraph tags `<p></p>`. You can copy the text from hw1 solution.
 - (d) Write "What we learnt from hw1" in second heading under tag `<h2></h2>`
 - (e) Copy all the points we learnt in hw1 solution. List all the points under ordered list tag ``. Notice that each item of the list should be inside list item tag ``.
 - (f) Now we want to make the text beautiful. For this we would write some CSS codes in between `<head></head>` tag under `<style></style>`. For this please refer to lecture 15 slide 8. First change the fonts of the body tag to Helvetica Neue.
 - (g) For the paragraph that contains the definition of data science, give an attribute `id='dfn'` and in CSS change the color of 'dfn' to white and background-color to steelblue.
 - (h) Write CSS so that color of h1,h2 becomes steelblue.
5. **Bonus for undergraduate (3 points) mandatory for graduate students:** The following link contains the complete texts of Romeo and Juliet written by Shakespeare. Read the complete text and generate a plot similar to Romeo and Juliet case study in lecture 13 (last plot).
- http://shakespeare.mit.edu/romeo_juliet/full.html
6. **Bonus (2 points) question for all :** GSS data challenge information can be found in the following link. Download the data sets and explore. Provide some plots and numerical summary that creates some interest about this data.

<http://www.amstat.org/sections/sgovt/dataac15.htm>