APPM2720 Spring 2017 Quiz 2

You should work on this quiz on your own, without get help from others. You are however, encouraged to use the web and any other reference materials and resources to complete these questions. All the numbered questions count equally for 10 points.

All data files are included in the APPM2720 Quiz2 directory you should also use the lectures and R code from the class as back ground and examples.

Please submit your work to me by email nychka@ucar.edu by midnight on Monday February 22, 2016.

You can submit your completed quiz using either:

- using the markdown format in Rstudio along with the knitr button to create an html file that you email to me (recommended!).
- add your text answers as comments in an R script, use the notebook button and convert your R script to an html file.

If you have some errors in your code and your script can not be converted. You can either comment out the code so it still can be graded but will not be run.

In R markdown you can include a block of code that is not run by omitting the `{r}` in beginning of the code block.

The text file **MobyDick.txt** is a text version of this classic work of American literature and has been downloaded from Project Gutenberg.

(1) The following R code can be used to read in this file:

```
everything<- scan("MobyDick.txt", what="a", quote="%")
everything2<- gsub('[();,:\"]',' ', everything )
everything2<- gsub('--',' ', everything2, fixed=TRUE )
everything2<- gsub('[.?!]',' XXX ', everything2 )  # STEP A
everything2<- scan( text= everything2, sep=" ",
                          what="raw", quote='@@')  #STEP B
MobyDick<- everything2[ everything2!=""]           #STEP C
```

Explain carefully what are the purposes of STEPS A,B and C in the R code above. How many words are in the R dataset `MobyDick` ?

(2) **How many chapters are in this book? Make a boxplot of word length by chapters. What is the longest word used?**

(NOTE: The book begins with some introduction before the chapters start and the chapters end with the final section being an "Epilogue". Consider the `rep` function demonstrated in class to create the indicator for the chapters.)

(3) Moby Dick is a white sperm whale that Captain Ahab is driven to hunt and kill. **Explain what the following R code does:**

''' tapply(MobyDick=="white", chapterInd, sum) ``` **Now use this code and also one for the string Ahab to identify some chapters that might focus specifically on Ahab and the whale (as opposed to only one of these topics.)**

The file **CUNorlinQuad.jpg** is an aerial photo of Norlin Quad.

(4) **Read this image into R. What are the dimensions of this image and the total number of pixels?**

(5) **Convert this color image to a grey scale image**

(6) **Use a simple thresholding scheme to find the sidewalks on campus in this image. Which image works better the color one or the grey scale?**

(7) If I is a data set with values in the range [0,1] the R code `Inew <- ( round( I*10))/10` will create a new set of values that are rounded to 10 equally spaced values in the [0,1] range. Use this technique on the *CUNorlin* image and allow for **15** different levels in each channel. **For the new image how different *possible* colors can be represented? Convert the new image to raster format and using the table command report how many different colors are actually in the image. How many colors appear in more than 100 pixels?**

Pick a company that you are interested in and go to Yahoo Finance Type in the company name in the search window and in the company page choose the item *Historical Prices* in the menu on the left side.

Download prices for the past two years in csv format. (Change the dates -> Click on apply -> click on download When I did it the data file was called table.csv)

(8) **Read these into R and make a plot of the daily closing and opening prices over time. Comment on any patterns.**

Note: You can convert the dates in the first column to a better form by using the **as.Date** function. For example if the **stockData** is your data frame then `dateObject <- as.Date(stockData$Date)` will do it.

(9) Investigate how well a simple forecast strategy works for predicting the prices: Find the mean price of the past 10 closing values and use this to predict the next day close. Make a boxplot of the forecast errors (Actual prices - forecast), comment on the distribution, and also investigate whether the errors depend on the day of the week.

Hints:

You may have to reverse the time order of your data -- check the order!

You can extract the day of the week from your date object above using for example
`dayOfWeek<- weekdays(dateObject)`