## HU Extension Assignment 01 E63 Big Data Analytics

## 

### Handed out: 01/30/2016 Due by 11:59PM on Friday, 02/05/2016

**Download and install the latest version of R and R Studio.**

**Problem 1.** Create a vector V with 8 elements (7,2,1,0,3,-1,-3,4).

* Transform that vector into a rectangular matrix A of dimensions 4X2 (4- rows, 2-columns).
* Create a matrix transpose to the above matrix A. Call that matrix AT.
* Calculate matrix products: A\*AT and AT\*A. Present the results. What are the dimensions of those two product matrices.
* Square matrixes sometimes have an inverse matrix. Try calculating inverse matrices (or matrixes, if you prefer) of above matrices (matrixes) A\*AT and AT\*A.
* Extend the above vector V with the ninth number of value -2. Do it elegantly by concatenating two vectors (☺).
* Transform that extended vector into a 3X3 matrix B.
* Calculate the inverse matrix of matrix B. Call it Binv. Demonstrate that the product of B and Binv is the same as the product of Binv and B and is equal to what?
* Determine the eigenvectors of matrixes B.
* Construct a new matrix C which is made by using each eigenvector of matrix B as a column. Calculate the product of matrix C and matrix B and the product of matrix B and C. Is there any significance to the elements of the product matrixes.
* Transform matrix B into a matrix with names columns and named rows.
* Transformed that fully “named” matrix into a data.frame.
* Ask the object you just created what is its class().

**Problem 2.** Consider file 2006Data.csv upload to the class site in Assignment 01 folder. File represents actual measurement of power consumption in a country somewhere in a California. Import data contained in that file into a data frame. You are expected to Google and find a function that will let you perform that import. Create a scatter plot of power consumption vs. temperature and power consumption vs. hour of the day. Subsequently create a boxplot with power on the vertical axis and hour of the day on the horizontal axis. The objective is to present the distribution (variation) of power consumption for every hour of the day.

Problem 3. Separate temperature scale in a reasonable number of intervals: 50 or 100. Calculate average power consumption, minimum power consumption and maximum power consumptions for every interval. Present those three sets of values on a single scatter graph (perhaps in different colours). Calculate covariance matrix between daily temperature and each of those power indicators. Give interpretations of the results.

**SUBMISSION INSTRUCTIONS:**

Your main submission should be an MS Word document containing your code, results produced by that code and brief textual descriptions of what you did and why. Typically, you just copy your code and results from the R console and past them into the Word document. Start with this text of homework assignment as the template. Please add any other files that you might have used or generated.

Please read detailed submission and grading instructions on class site.