BUAN 6357 (Johnston)

Homework 3

Code Due: 9 February 2019 (6PM)

Part B Due: 10 February 2019 (11:59PM)

Points available: 100.

This assignment is about application of a specified penalty function (within-group error sum of squares, *a.k.a.* WSS) in a clustering process which does not directly provide support for that calculation. Perform the cluster analysis using hclust() and squared euclidean distance to track with kmeans(). Use the dataset "Boston" from MASS for this assignment. (Notice the capitalization.) The number of clusters considered is to range from 1 to 25, inclusive. For this assignment you will need the packages "MASS", "tidyverse" and "broom", in that order. You should not use any additional packages. You should use only the "require()" or "library()" statement in your code. Any use of the install.packages() function in submitted code will result in a score of 0 for that submission.

The first commands of your code MUST be:

setwd("c:/data/BUAN6357/HW_3"); source("prep.txt", echo=T)

and the last command of your code MUST be:

source("validate.txt", echo=T)

Be careful with the quote characters as they must ALL be the same at the beginning and end of a string. (Use the single or double quote character from the key next to "Enter".) Inclusion of these lines is required BEFORE your code will be tested.

Submit the code to eLearning as an ASCII file which can be copied directly into R.

You may submit this assignment as many times as needed until you get full credit.

Deliverables (all names lower case) :

1. hc hclust() result object
2. hc.twss vector of total WSS values by number of clusters

Part B of HW 3 will direct you to explore the deliverables and answer questions about them. You may submit answers to HW 3 part B as many times as you wish but only the score for the last submitted code will be retained.

Note: you will also need the cutree() function to truncate the hclust() result at the various numbers of clusters.