Assignment #7: Due November 1, 2017, 1pm

INSTRUCTIONS: Please upload **ONLY** the following file to Canvas: your R program (saved as LASTNAME_FIRSTNAME_HW7). Make sure to use the setwd() function to create a path as was done in class and use that path to locate your data sets. Please make sure to comment your code.

You are to repeat Assignment #2 using R.

There is a file *AdoptionData.csv* which contains data from Tuberculosis Screening for International Adoptees. In addition, there is a file *Coding Sheet.doc* which contains the labels (Question Description) and coding (Field Type, Field Format, Field Range) for the data set. You are to create an **analysis data set** (R data frame called *analysis*) by doing the following:

- Reduce the data set down to contain the following variables: uniqueid, COUNTRY, LIVING, ADOPTAGE, AGETST1, WTTST1, HTTST1, INDTST1, RESTST1, CXRDONE, GENDER
- Create **labels** (like formats in SAS) for all of the variables in the reduced data set
- Convert all values of 99 and 88 appropriately in the reduced data set, making sure that you have an analysis data set when completed. (HINT: Think about how these should be coded)
- Create the necessary variables to do the calculations below

NOTE: You may use several steps to create the data set, but the final data set should be called *analysis*.

Please calculate the following:

(1) GENDER and CXRDONE frequency distributions
(2) Mean BMI (using WTTST1 and HTTST1)

Again, I do not want any output – please provide the code to do the above.

NOTE: An analysis data set is a data set that has been **fully** cleaned. All data has been checked, all issues and anomalies have been resolved and the user should not have to do anything further in order to modify the data.