Code Book

NOTE: Author of run\_analysis is 71 year-old former COBOL programmer so longer data names used plus capital letters used

run\_analysis script used to do followng:

1) Merge training & test sets to create 1 data set

2) extracts only measurements on mean & standard deviation

for each measurement

NOTE: descriptive activity names used for activities in merged

data set

NOTE: appropriately label merged data set with descriptive

variable names

3) from merged data set, creates 2nd independent tidy data set

with the average of each variable for each activity & each subject

Data

This project is an exercise in acquiring and cleaning data

Processing

Step 1: load dplyr as package need for several steps

Step 2: set up steps for getting file GCDFILE to start working with

A) set GCDFILE as zip file in authors area

B) check to see if GCDFILE does not exist until download made

C) set GCDURL to http address of data to download

D) download file

E) unzip GCDFILE

Step 3: Assign names to each of the 8 data frames (column heading

also added)

1) UHDFEATS for “features.txt"

2) UHDACTS for “activity\_labels.txt"

3) UHDTEST\_S for “subject\_test.txt"

4) UHDTEST\_X for “test/x\_test.txt"

5) UHDTEST\_Y for “test/y\_test.txt"

6) UHDTRAIN\_S for “subject\_train.txt"

7) UHDTRAIN\_X for “train/x\_train.txt"

8) UHDTRAIN\_Y for “train/y\_train.txt"

NOTE:

FEATNAMES created as list of UHDFEATS FUNCTIONS

Step 4: Combine data by rows for x, y & subject

A) CTTX row bind of UHDTEST\_X & UHDTRAIN\_X

B) CTTY row bind of UHDTEST\_Y & UHDTRAIN\_Y

C) CTTS row bind of UHDTEST\_S & UHDTRAIN\_S

Step 5: Merge by columns combined data sets of x, y & subject

MSYX column bind of CTTS, CTTX & CTTY

Step 6: Using the merge data set, create data set with ‘mean’ &

‘standard deviation’

ONLYNEED created from MSYX with selecting SUBJECT & CODE with

‘mean’ & standard deviation

Step 7: give descriptive names to the activites in just created data set

NOTE: full names of activites made

Step 8: create a final data set from prior created data set that is

grouped by subject & activity & summarized by ‘mean’ of all

activites

LASTNEED created from ONLYNEED group\_by(SUBJECT, ACTIVITY)

summarized

Step 9: final data set written out to table to ‘finalData.txt’