GettingAndCleaningData final project

Christopher Krolak

2/24/2021

High level process flow:

1. download zip file from URL
2. extract zip file to temp folder
3. build the test data set:
    1. read "X_test.txt" into data frame
    2. use "features.txt" for column names
    3. decode activity index from "Y_test.txt" with data from "activity_labels.txt", prepend this as the first data column.
4. build the training data set:
    1. read "X_train.txt" into data frame
    2. use "features.txt" for column names
    3. decode activity index from "Y_train.txt" with data from "activity_labels.txt", prepend this as the first data column.
5. Merge the test and traiing data sets
6. subset the merged data to included 2 attribute columns (subject, activity) and then only data columns with "mean" or "std" in the column name. This is the mergedData set.
7. create a new tidyData set with column means for all numberic columns. This is the "tidyData" set.

Automated Script: 'run_analysis.R'

All of this data reduction is automated in the script "run_analysis.R". To execute, do the following: 1. import script: source("run_analysis.R") 2. run script: main()

Resulting global objects:

mergedData = merged and reduced data set of 10299 observations (rows) and 81 variables (columns)
tidyData = summarized data with 40 observations (rows) and 81 variables. Each row is a group mean by subject and activity.

Resulting data file:

tidyData.txt is the summarized data file, it will be in your working directory.