

CodeBook

Laurie Cuffney

7/22/2019

Codebook WearableTechDataAnalysis

Address Variables

- testXAddress: address of test variable data file
- testYAddress: address of test activity label file
- testSubAddress: address of test subject Id label file
- trainXAddress: address of training variable data file
- trainYAddress: address of training activity label file
- trainSubAddress: address of training subject Id label file
- featuresAddress: address of features file

Data

- testX: test data for 561 measured variables
- testY: activity labels for test data
- testSub: subject labels for test data
- testData: mean and standard deviation test data with activity and subject IDs
- trainX: training data for 561 measured variables
- trainY: activity labels for training data
- trainSub: subject labels for training data
- trainData: mean and standard deviation training data with activity and subject IDs
- fullData: merged test and training data
- skinnyData: skinny data set over subject Id and activity
- averageData: averages of 66 mean and standard deviation variables for each subject and activity level
- finaldata: average data arranged by activity level and subject

Functions

- numRemove removes leading numbers
- renameAct: replace activity numbers labels with a readable label

Variables

- features: variable names for 561 measured variables.
- feat: feature labels with no numbers and only descriptive label
- activities: activities names ordered by their ID number from original codebook
- testYreadable: test activity labels in a readable format
- trainYreadable: training activity labels in a readable format
- locationMeanStd: indices of mean and standard deviation variables
- testXMeanStd: contains only test mean and standard deviation variables
- trainXMeanStd: contains only training mean and standard deviation variables

Transformation

- Data headers are renamed into readable formats using provided files in data set UCI HAR Dataset.

- Data is reduced down to only mean and standard deviation variables using indices found with `grep`.
- Training and Test data are merged together and then melted into a skinny data set using the `subject` Ids and activity labels over the 66 mean and standard deviation variables.
- Averages for each subject and activity level are calculated by using `dcast` on the skinny data set.
- The obtained averages are arranged into a final tidy data set using the `arrange` function from the `dplyr` package.