

CodeBook

Rafael Otega

12 de junio de 2016

Process

1. I have imported the test and train data sets.
2. I have included the names of the columns to the "x" set.
3. I have use the information in "activity labels.txt" to search and replace the numbers in set "y" for the name of the activity.
4. The variables have been filtered so that only those which are the mean or the standard deviation of something have been selected.
5. I have merged "x", "y" and the id of the subject. This have originated the sets "train" and "test".
6. I have merged "test" and "train" to create "data".
7. I have performed the requiered calculus of the mean by groupen over the id and the activity variables.

Variables

Two new variables have been added:

activity, that describes with words, using the information provided in the file "activity labels.txt", the name of the activity the subject was performing when the measurement was taken.

id, this means the number given to the subject that undertook the activity.

The variables have been filtered so that only the mean and the standard deviation have been selected. As it might een seen, the name of this variables (columns) now contain the word “mean” at the end, to represent the fact that the mean has been calculated. This mean have been done over the same values of the id and activity. In other words, data have been grouped by the id and the name of the activity prior to calculate the mean.

Units

The “id” column has no unit. It is just an identifier. The column“activity” has not unit either. It is the name of the activity the subject was undertaken during the measurement. The rest of the variables are given in multiples of “g” and are normalized to be between -1 and 1.