**CODEBOOK for Cleaning and Getting Data Project**

**Description of Project Datasets and Data Processing**

|  |  |  |  |
| --- | --- | --- | --- |
| Datasets | Description of datasets | Code | Remarks |
| **Train datasets** | Original datasets from raw text files for subjects in the train group |  |  |
| 1. Subject\_train | Contains subject identifiers for subjects in the train-group | Its range is from 1 to 30 | Read into R as an object called subject\_train |
| 1. X\_train | Contain a 561- feature vector measured as time and frequency from the accelerometer and gyroscope 3-axial raw signals for subjects in the train-group |  | Read into R as an object called train\_set and given column names from the features.txt file |
| 1. y\_train | Activity\_class for train-group | 1=Walking  2=Walking\_upstairs  3=Walking downstairs  4=Sitting  5=Standing  6=Laying | Read into R as an object called activity\_train and linked with activity labels (as described under code) |
| **Test datasets** | Original datasets from raw text files for test group |  |  |
| 1. Subject\_test | Contain subject identifiers for test group | Its range is from 1 to 30 | Read into R as an object called subject\_test |
| 1. X\_test | Contain a 561- feature vector measured as time and frequency from the accelerometer and gyroscope 3-axial raw signals for the subjects in the test group |  | Read into R as an object called test\_set and given column names from the features.txt file |
| 1. y\_test | Contain activity\_class for test group | 1=Walking  2=Walking\_upstairs  3=Walking downstairs  4=Sitting  5=Standing  6=Laying | Read into R as an object called activity\_test and linked with activity labels (as described under code) |
| **Derived datasets** | Datasets derived from original raw datasets above |  |  |
| train\_data | Derived as a single dataset for the train group by combining subject\_train, train\_set and activity\_train datasets |  |  |
| test\_data | Derived as a single dataset for the test group by combining subject\_test, test\_set and activity\_test datasets |  |  |
| merged\_data | Dataset combining both the test\_data and train\_data so that information from both test and train groups are now together in one dataset. |  |  |
| merged\_data\_means\_std | Dataset containing only the mean and standard deviation measurements |  | Extracted from merged\_data |
| mean\_activity\_data | Second independent tidy data obtained by averaging each variable in merged\_data\_means\_std for each activity and subject. |  |  |