*Links to original UCI HAR data and study explanation:*

* Data: <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>
* Explanation: <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

The information below describes the data frame objects created from the UCI HAR raw data referenced above:

**Raw training data objects (70% of volunteers)**

*train\_subjectIDs*

* Numbers 1 – 30 indicating the study volunteer associated with each observation
* 7,352 observations in 1 column
* Data frame

*train\_activityIDs*

* Numbers 1 – 6 corresponding with activity labels (walking, etc.)
* 7,352 observations in 1 column
* Data frame

*train\_data*

* Measurements and calculations for each volunteer, activity and feature
* 7,352 observations in 561 columns
* Data frame

**Raw test data objects (30% of volunteers)**

*test\_subjectIDs*

* Numbers 1 – 30 indicating the study volunteer associated with each observation
* 2,947 observations in 1 column
* Data frame

*test\_activityIDs*

* Numbers 1 – 6 corresponding with activity labels (walking, etc.)
* 2,947 observations in 1 column
* Data frame

*test\_data*

* Measurements and calculations for each volunteer, activity and feature
* 2,947 observations in 561 columns
* Data frame

**Other raw data objects**

*activity\_labels*

* 6 activities the participants performed - walking, walking\_upstairs, Walking\_downstairs, sitting, standing, laying
* 6 observations in two columns
  + Column 1 = numbers 1 – 6
  + Column 2 = activity descriptions, i.e. “LAYING”, “SITTING”
* Data frame

*features*

* The descriptions of the metrics recorded and calculated in the study
* 561 observations in two columns
  + Column 1 = numbers 1 – 561
  + Column 2 = metric descriptions, i.e. "tBodyAcc-mean()-X"
* Data frame

**R objects created as part of creating the tidy data set**

*training*

* This data frame is a combination (created using the cbind function) of the following data frame objects described above:
  + *train\_subjectIDs*
  + *train\_activityIDs*
  + *train\_data*
* Column 1 is labeled ‘volunteerid’ and contains volunteer ID numbers
* Column 2 is labeled ‘activityid’ and contains the ID number associated with each activity performed by the volunteers
* Column 3 is labeled ‘activity’ and contains the activity description (i.e. STANDING, WALKING, etc) associated with each activity ID number
* Columns 4 – 564 contain training data related to each of the measurements and calculations in the ‘features’ data frame. The columns are labeled with each feature name.
* 7,352 observations of 564 variables

*test*

* This data frame is a combination (created using the cbind function) of the following data frame objects described above:
  + *test\_subjectIDs*
  + *test\_activityIDs*
  + *test\_data*
* Column 1 is labeled ‘volunteerid’ and contains volunteer ID numbers
* Column 2 is labeled ‘activityid’ and contains the ID number associated with each activity performed by the volunteers
* Column 3 is labeled ‘activity’ and contains the activity description (i.e. STANDING, WALKING, etc) associated with each activity ID number
* Columns 4 – 564 contain test data related to each of the measurements and calculations in the ‘features’ data frame. The columns are labeled with each feature name.
* 2,947 observations of 564 variables

*train\_test*

* This data frame is a combination of the *training* and *test* data frames described above (combined using rbind)
* 10,299 observations of 564 variables

*train\_test\_noDups*

* *train\_test* contains duplicate column names that are not needed for the analysis and that cause problems with later dplyr functions
* I removed the unnecessary duplicate columns to create the *train\_test\_noDups* data frame. It keeps the mean() and std() columns needed for the analysis.
* 10,299 observations of 480 variables

The data frame objects below subset the desired columns from *train\_test\_noDups*:

* *train\_test\_vol*
  + Subsets only the ‘volunteerid’ column
  + 10,299 observations of 1 variable
* *train\_test\_act*
  + Subsets only the ‘activity’ column
  + 10,299 observations of 1 variable
* *train\_test\_mean*
  + Subsets all columns containing a mean() calculation.
  + I explicitly choose to include both mean() and meanFreq() measurements in an effort to include all potentially useful data.
  + I exclude angle measurements that appear to be based on underlying mean data because they do not appear to meet the criteria of being mean measurements themselves.
  + 10,299 observations of 46 variables
* *train\_test\_stdev*
  + Subsets all columns containing a std() calculation (standard deviation).
  + 10,299 observations of 33 variables

*train\_test\_clean*

* Combines the following data frames using cbind:
  + *train\_test\_vol*
  + *train\_test\_act*
  + *train\_test\_mean*
  + *train\_test\_std*
* *train\_test\_clean* is the tidy data set requested by the assignment deliverable #4.
  + It is a combination of the training and test data sets
  + It includes only measurements of mean() and std()
  + It contains a ‘volunteerid’ column which contains the ID numbers associated with each study participant
  + It contains an ‘activity’ column containing descriptive activity names that correspond with each observation
  + All columns are labeled with descriptive variable names
* 10,299 observations of 81 variables

*train\_test\_summary*

* *train\_test\_summary* is the tidy data set requested by the assignment deliverable #5.
  + It uses the group\_by() and summarise\_each() functions to group *train\_test\_clean* by volunteer and by activity and then calculates the mean of each measurement for each grouping.
* 180 observations of 81 variables