Codebook

Course project

Coursera Getting and Cleaning Data (in R)

Johns Hopkins University.

Antonio Rubiera

August 28, 2019

Description

In this codebook we describe the data contained in the file **selected_data_grouped.txt**. The original dataset this data is derived is from the dataset made public by Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra2 and Jorge L. Reyes-Ortiz as documented in their research paper "A Public Domain Dataset for Human Activity Recognition Using Smartphones."[1] The original data comes in two parts: a training dataset, labeled 'train' in the original dataset, containing 7,352 observations, and a test dataset, labeled 'test' in the original dataset, containing 2,947 observations.

There are two categories of data in the original dataset. The first category is composed of a sample of the raw inertial signal data for three variables. The second category is composed of a processed sample of 561 variables. In this codebook we describe a selection from the **second** category for 79 fields measuring mean values and standard deviations. The procedure we follow is as required for this assignment:

1. We merge the training and the test data sets to create one data set.
2. We extract 79 measurements on the mean and standard deviation for each measurement from the 561 measurements in the original dataset.
3. We use descriptive activity names to name the activities in the data.
4. We label the data set columns appropriately with descriptive variable names.
5. From the data set in step 4, we extract a second, independent tidy data set with the average (men) of each variable (the 79 columns) grouped by activity and subject. It is this extracted, tidy data set that is contained in the file **selected_data_grouped.txt**.

The github repository for this assignment is found at: https://github.com/rubiera/GitHubLink

[1] ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010

Data Description (Metadata)

The dataset contained in the file selected_data_grouped.txt has four columns for subjects and activities as group by variables for the mean and standard deviation of the acceleration (m/s^2) and gyroscope measurement (m) extracted from the original dataset

Subjects and Activities

Subject Number is in the range 1 to 30 and found in the data as the column SubjectNumber. Activity descriptions are in the column ActivityName and activity numbers are in the column ActivityNumber. Activity descriptions and their respective numbers are:

1. WALKING
2. WALKING_UPSTAIRS
3. WALKING_DOWNSTAIRS
4. SITTING
5. STANDING
6. LAYING

The data has a count for the number of observations N (in the range 36 to 95 Observations) for a given subject and a matching activity (subject 1 to 30 X activity 1 to 6, for a total of 180 possible groups). A snipped of the leftmost columns of the dataset (as extracted from txt into a csv file) is shown below.

	A	B	C	D	E	F	G
1		ActivityName	SubjectNumber	N	ActivityNumber	timeBodyAccerelometer-mean()-X	timeBodyAccerelometer-mean()-Y
2	1	WALKING	2	59	1	0.276619902	-0.01813606
3	2	WALKING	4	60	1	0.282200437	-0.019293435
4	3	WALKING	9	52	1	0.255528502	-0.016854968
5	4	WALKING	10	53	1	0.2773478	-0.015905978
6	5	WALKING	12	50	1	0.275332021	-0.021277246
7	6	WALKING	13	57	1	0.270651574	-0.019965793
8	7	WALKING	18	56	1	0.276552252	-0.012223713
9	8	WALKING	20	51	1	0.277413895	-0.012505082
10	9	WALKING	24	58	1	0.278561305	-0.015456722
11	10	WALKING	1	95	1	0.272628958	-0.013411706
12	11	WALKING	3	58	1	0.269043715	-0.022552792
13	12	WALKING	5	56	1	0.257347329	-0.022548869
14	13	WALKING	6	57	1	0.275101614	-0.021830524
15	14	WALKING	7	57	1	0.254956073	-0.029137351
16	15	WALKING	8	48	1	0.284614576	-0.015390657
17	16	WALKING	11	59	1	0.279504604	-0.017083205
18	17	WALKING	14	59	1	0.276323994	-0.020443872
19	18	WALKING	15	54	1	0.271288506	-0.020533644
20	19	WALKING	16	51	1	0.274075316	-0.015191057
21	20	WALKING	17	61	1	0.276884372	-0.01734567
22	21	WALKING	19	52	1	0.286058382	-0.014607829
23	22	WALKING	21	52	1	0.273987303	-0.014829676
24	23	WALKING	22	46	1	0.260226489	-0.017113433
25	24	WALKING	23	59	1	0.277088644	-0.022616648
26	25	WALKING	25	74	1	0.275826762	-0.017534954
27	26	WALKING	26	59	1	0.279680084	-0.015567649
28	27	WALKING	27	57	1	0.271317198	-0.019461182
29	28	WALKING	28	54	1	0.274562244	-0.020006204
30	29	WALKING	29	53	1	0.282277544	-0.017550681
31	30	WALKING	30	65	1	0.268480696	-0.020013703
32	31	WALKING_UPSTAIRS	2	48	2	0.279061071	-0.010394458
33	32	WALKING_UPSTAIRS	4	52	2	0.283685503	-0.015932698

Data Column Descriptions

The data is grouped by subject and activity, and shows the mean of each of the 79 measurement columns. There are therefore 180 categories of means, each containing 79 measurements. The data columns are described in the following figures.

	VariableName	Explanation	Unit of Measure
1	timeBodyAccerelometer-mean()-X	The time domain mean acceleration of the subject in the X dimension	Meters per Second Squared (m/s^2)
2	timeBodyAccerelometer-mean()-Y	The time domain mean acceleration of the subject in the Y dimension	Meters per Second Squared (m/s^2)
3	timeBodyAccerelometer-mean()-Z	The time domain mean acceleration of the subject in the Z dimension	Meters per Second Squared (m/s^2)
4	timeBodyAccerelometer-std()-X	The time domain standard deviation of the acceleration of the subject in the X dimension	Meters per Second Squared (m/s^2)
5	timeBodyAccerelometer-std()-Y	The time domain standard deviation of the acceleration of the subject in the Y dimension	Meters per Second Squared (m/s^2)
6	timeBodyAccerelometer-std()-Z	The time domain standard deviation of the acceleration of the subject in the Z dimension	Meters per Second Squared (m/s^2)
7	timeGravityAccerelometer-mean()-X	The time domain mean of the gravity measure of the accelerometer in the X dimension	Meters per Second Squared (m/s^2)
8	timeGravityAccerelometer-mean()-Y	The time domain mean of the gravity measure of the accelerometer in the Y dimension	Meters per Second Squared (m/s^2)
9	timeGravityAccerelometer-mean()-Z	The time domain mean of the gravity measure of the accelerometer in the Z dimension	Meters per Second Squared (m/s^2)
10	timeGravityAccerelometer-std()-X	The time domain standard deviation of the gravity measure of the accelerometer in the X dimension	Meters per Second Squared (m/s^2)
11	timeGravityAccerelometer-std()-Y	The time domain standard deviation of the gravity measure of the accelerometer in the Y dimension	Meters per Second Squared (m/s^2)
12	timeGravityAccerelometer-std()-Z	The time domain standard deviation of the gravity measure of the accelerometer in the Z dimension	Meters per Second Squared (m/s^2)

	VariableName	Explanation	Unit of Measure
13	timeBodyAccerelometerJerk-mean()-X	The time domain mean of the acceleration jerk of the subject in the X dimension	Meters per Second Squared (m/s^2)
14	timeBodyAccerelometerJerk-mean()-Y	The time domain mean of the acceleration jerk of the subject in the Y dimension	Meters per Second Squared (m/s^2)
15	timeBodyAccerelometerJerk-mean()-Z	The time domain mean of the acceleration jerk of the subject in the Z dimension	Meters per Second Squared (m/s^2)
16	timeBodyAccerelometerJerk-std()-X	The time domain standard deviation of the acceleration jerk of the subject in the X dimension	Meters per Second Squared (m/s^2)
17	timeBodyAccerelometerJerk-std()-Y	The time domain standard deviation of the acceleration jerk of the subject in the Y dimension	Meters per Second Squared (m/s^2)
18	timeBodyAccerelometerJerk-std()-Z	The time domain standard deviation of the acceleration jerk of the subject in the Z dimension	Meters per Second Squared (m/s^2)
19	timeBodyGyroscope-mean()-X	the time domain mean of the relative distance of the subject in the X dimension	Meters (m)
20	timeBodyGyroscope-mean()-Y	the time domain mean of the relative distance of the subject in the Y dimension	Meters (m)
21	timeBodyGyroscope-mean()-Z	the time domain mean of the relative distance of the subject in the Z dimension	Meters (m)
22	timeBodyGyroscope-std()-X	the time domain standard deviation of the mean of the relative distance of the subject in the X dimension	Meters (m)
23	timeBodyGyroscope-std()-Y	the time domain standard deviation of the mean of the relative distance of the subject in the Y dimension	Meters (m)
24	timeBodyGyroscope-std()-Z	the time domain standard deviation of the mean of the relative distance of the subject in the Z dimension	Meters (m)

	VariableName	Explanation	Unit of Measure
25	timeBodyGyroscopeJerk-mean()-X	the time domain mean of the relative jerk distance of the subject in the X dimension	Meters (m)
26	timeBodyGyroscopeJerk-mean()-Y	the time domain mean of the relative jerk distance of the subject in the Y dimension	Meters (m)
27	timeBodyGyroscopeJerk-mean()-Z	the time domain mean of the relative jerk distance of the subject in the Z dimension	Meters (m)
28	timeBodyGyroscopeJerk-std()-X	the time domain standard deviation of the mean of the relative jerk distance of the subject in the X dimension	Meters (m)
29	timeBodyGyroscopeJerk-std()-Y	the time domain standard deviation of the mean of the relative jerk distance of the subject in the Y dimension	Meters (m)
30	timeBodyGyroscopeJerk-std()-Z	the time domain standard deviation of the mean of the relative jerk distance of the subject in the Z dimension	Meters (m)
31	timeBodyAccerelometerMagnitude-mean()	The time domain mean of the magnitude of the acceleration of the subject	Meters per Second Squared (m/s^2)
32	timeBodyAccerelometerMagnitude-std()	The time domain standard deviation of the mean of the magnitude of the acceleration of the subject	Meters per Second Squared (m/s^2)
33	timeGravityAccerelometerMagnitude-mean()	The time domain mean of the magnitude of the gravity measure of the subject	Meters per Second Squared (m/s^2)
34	timeGravityAccerelometerMagnitude-std()	The time domain standard deviation of the mean of the magnitude of the gravity measure of the subject	Meters per Second Squared (m/s^2)
35	timeBodyAccerelometerJerkMagnitude-mean()	The time domain mean jerk acceleration magnitude of the subject	Meters per Second Squared (m/s^2)
36	timeBodyAccerelometerJerkMagnitude-std()	The time domain standard deviation of the mean jerk acceleration magnitude of the subject	Meters per Second Squared (m/s^2)

	VariableName	Explanation	Unit of Measure
37	timeBodyGyroscopeMagnitude-mean()	The time domain mean relative distance magnitude of the subject	Meters (m)
38	timeBodyGyroscopeMagnitude-std()	The time domain standard deviation of the mean relative distance magnitude of the subject	Meters (m)
39	timeBodyGyroscopeJerkMagnitude-mean()	The time domain mean jerk relative distance magnitude of the subject	Meters (m)
40	timeBodyGyroscopeJerkMagnitude-std()	The time domain standard deviation of the mean jerk relative distance magnitude of the subject	Meters (m)
41	freqBodyAccerelometer-mean()-X	The frequency domain mean acceleration of the subject in the X dimension	Meters per Second Squared (m/s^2)
42	freqBodyAccerelometer-mean()-Y	The frequency domain mean acceleration of the subject in the Y dimension	Meters per Second Squared (m/s^2)
43	freqBodyAccerelometer-mean()-Z	The frequency domain mean acceleration of the subject in the Z dimension	Meters per Second Squared (m/s^2)
44	freqBodyAccerelometer-std()-X	The frequency domain standard deviation of the acceleration of the subject in the X dimension	Meters per Second Squared (m/s^2)
45	freqBodyAccerelometer-std()-Y	The frequency domain standard deviation of the acceleration of the subject in the Y dimension	Meters per Second Squared (m/s^2)
46	freqBodyAccerelometer-std()-Z	The frequency domain standard deviation of the acceleration of the subject in the Z dimension	Meters per Second Squared (m/s^2)
47	freqBodyAccerelometer-meanFrequency()-X	The frequency domain mean frequency acceleration of the subject in the X dimension	Meters per Second Squared (m/s^2)
48	freqBodyAccerelometer-meanFrequency()-Y	The frequency domain mean frequency acceleration of the subject in the Y dimension	Meters per Second Squared (m/s^2)
49	freqBodyAccerelometer-meanFrequency()-Z	The frequency domain mean frequency acceleration of the subject in the Z dimension	Meters per Second Squared (m/s^2)

	VariableName	Explanation	Unit of Measure
50	freqBodyAccerelometerJerk-mean()-X	The frequency domain mean of the acceleration jerk of the subject in the X dimension	Meters per Second Squared (m/s^2)
51	freqBodyAccerelometerJerk-mean()-Y	The frequency domain mean of the acceleration jerk of the subject in the Y dimension	Meters per Second Squared (m/s^2)
52	freqBodyAccerelometerJerk-mean()-Z	The frequency domain mean of the acceleration jerk of the subject in the Z dimension	Meters per Second Squared (m/s^2)
53	freqBodyAccerelometerJerk-std()-X	The frequency domain standard deviation of the acceleration jerk of the subject in the X dimension	Meters per Second Squared (m/s^2)
54	freqBodyAccerelometerJerk-std()-Y	The frequency domain standard deviation of the acceleration jerk of the subject in the Y dimension	Meters per Second Squared (m/s^2)
55	freqBodyAccerelometerJerk-std()-Z	The frequency domain standard deviation of the acceleration jerk of the subject in the Z dimension	Meters per Second Squared (m/s^2)
56	freqBodyAccerelometerJerk-meanFrequency()-X	The frequency domain mean frequency jerk acceleration of the subject in the X dimension	Meters per Second Squared (m/s^2)
57	freqBodyAccerelometerJerk-meanFrequency()-Y	The frequency domain mean frequency jerk acceleration of the subject in the Y dimension	Meters per Second Squared (m/s^2)
58	freqBodyAccerelometerJerk-meanFrequency()-Z	The frequency domain mean frequency jerk acceleration of the subject in the Z dimension	Meters per Second Squared (m/s^2)
59	freqBodyGyroscope-mean()-X	the frequency domain mean of the relative distance of the subject in the X dimension	Meters (m)
60	freqBodyGyroscope-mean()-Y	the frequency domain mean of the relative distance of the subject in the Y dimension	Meters (m)
61	freqBodyGyroscope-mean()-Z	the frequency domain mean of the relative distance of the subject in the Z dimension	Meters (m)

	VariableName	Explanation	Unit of Measure
62	freqBodyGyroscope-std()-X	the frequency domain standard deviation of the mean of the relative distance of the subject in the X dimension	Meters (m)
63	freqBodyGyroscope-std()-Y	the frequency domain standard deviation of the mean of the relative distance of the subject in the Y dimension	Meters (m)
64	freqBodyGyroscope-std()-Z	the frequency domain standard deviation of the mean of the relative distance of the subject in the Z dimension	Meters (m)
65	freqBodyGyroscope-meanFrequency()-X	The frequency domain mean frequency relative distance of the subject in the X dimension	Meters (m)
66	freqBodyGyroscope-meanFrequency()-Y	The frequency domain mean frequency relative distance of the subject in the Y dimension	Meters (m)
67	freqBodyGyroscope-meanFrequency()-Z	The frequency domain mean frequency relative distance of the subject in the Z dimension	Meters (m)
68	freqBodyAccerelometerMagnitude-mean()	The frequency domain mean of the magnitude of the acceleration of the subject	Meters per Second Squared (m/s^2)
69	freqBodyAccerelometerMagnitude-std()	The frequency domain standard deviation of the mean of the magnitude of the acceleration of the subject	Meters per Second Squared (m/s^2)
70	freqBodyAccerelometerMagnitude-meanFrequency()	The frequency domain mean frequency magnitude of the acceleration of the subject	Meters per Second Squared (m/s^2)
71	freqBodyBodyAccerelometerJerkMagnitude-mean()	The frequency domain magnitude mean of the acceleration jerk of the subject	Meters per Second Squared (m/s^2)
72	freqBodyBodyAccerelometerJerkMagnitude-std()	The frequency domain magnitude standard deviation of the mean of the acceleration jerk of the subject	Meters per Second Squared (m/s^2)
73	freqBodyBodyAccerelometerJerkMagnitude-meanFrequency()	The frequency domain mean frequency magnitude of the acceleration jerk of the subject	Meters per Second Squared (m/s^2)

	VariableName	Explanation	Unit of Measure
74	freqBodyBodyGyroscopeMagnitude-mean()	The frequency domain mean magnitude of the relative distance of the subject	Meters (m)
75	freqBodyBodyGyroscopeMagnitude-std()	The frequency domain standard deviation of the mean magnitude of the relative distance of the subject	Meters (m)
76	freqBodyBodyGyroscopeMagnitude-meanFrequency()	The frequency domain mean frequency magnitude of the relative distance of the subject	Meters (m)
77	freqBodyBodyGyroscopeJerkMagnitude-mean()	The frequency domain mean magnitude of the relative distance jerk of the subject	Meters (m)
78	freqBodyBodyGyroscopeJerkMagnitude-std()	The frequency domain standard deviation of the mean magnitude of the relative distance jerk of the subject	Meters (m)
79	freqBodyBodyGyroscopeJerkMagnitude-meanFrequency()	The frequency domain mean frequency magnitude of the relative distance jerk of the subject	Meters (m)