**Getting and Cleaning Data: Course Project**

**Introduction**

This repository contains my work for the course project for the Coursera course "Getting and Cleaning data", part of the Data Science specialization. What follows first are my notes on the original data.

**About the raw data**

The features (561 of them) are unlabelled and can be found in the x\_test.txt and x\_train.txt file. The activity labels are in the y\_test.txt and y\_train.txt file. The test subjects are in the subject\_test.txt and subject\_train.txt file. Labels features are in features\_test.txt file. Names of activities are in activity\_labels.txt file.

**About the script and the tidy dataset**

I created a script called run\_analysis.R which will merge the test and training sets together and give the final tidy dataset as output. Prerequisites for the code:

1. Data must be downloaded and save in a folder called Data.

After merging testing and training, labels are added and only columns that have to do with mean and standard deviation are kept.

Lastly, the script will create a tidy data set containing the means of all the columns per test subject and per activity. This tidy dataset will be written to a tab-delimited file called tidy.txt, which can also be found in this repository.

**About the Code Book**

The CodeBook.md file explains the transformations performed and the resulting data and variables.