Downloading “Electric power consumption” dataset file for Course Project 1, Exploratory Data Analysis, Coursera

The dataset for this project was made available as a zip file on the course website and a link to this dataset was posted as “Electric power consumption” in the project instructions webpage. The url was obtained by right clicking and copying the link address.

A “DataArchive” directory was created in the working directory, and the copied link address, with quotes, was assigned to fileUrl. The “download.file” function was used to download the zip file from the fileUrl to the “DataArchive” directory as “household_power_consumption.zip”, and the date of download was recorded.

(Note that the R-WIN_INTERNET2 must be set to TRUE (i.e. “setInternet2(use = TRUE”) in order for the https download to work properly with the knitr package in R Markdown files.)

```
if(!file.exists("DataArchive")) {
    dir.create("DataArchive")
}

setInternet2(use = TRUE)
fileUrl <-
"https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip"

download.file(fileUrl, destfile =
"./DataArchive/household_power_consumption.zip")

list.files("./DataArchive")
```

```
## [1] "household_power_consumption.zip"
```

```
dateDownloaded.household_power_consumption.zip <- date()
dateDownloaded.household_power_consumption.zip
```

```
## [1] "Sun Jun 08 10:15:37 2014"
```

From the Project Instructions:

Introduction

This assignment uses data from the UC Irvine Machine Learning Repository, a popular repository for machine learning datasets. In particular, we will be using the “Individual household electric power consumption Data Set” which I have made available on the course web site:

Dataset: Electric power consumption [20Mb]

Description: Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available.

The following descriptions of the 9 variables in the dataset are taken from the UCI web site:

1. Date: Date in format dd/mm/yyyy
2. Time: time in format hh:mm:ss
3. Global_active_power: household global minute-averaged active power (in kilowatt)
4. Global_reactive_power: household global minute-averaged reactive power (in kilowatt)
5. Voltage: minute-averaged voltage (in volt)
6. Global_intensity: household global minute-averaged current intensity (in ampere)
7. Sub_metering_1: energy sub-metering No. 1 (in watt-hour of active energy). It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas powered).
8. Sub_metering_2: energy sub-metering No. 2 (in watt-hour of active energy). It corresponds to the laundry room,

containing a washing-machine, a tumble-drier, a refrigerator and a light.
9. Sub_metering_3: energy sub-metering No. 3 (in watt-hour of active energy). It corresponds to an electric water-heater and an air-conditioner.

The dataset has 2,075,259 rows and 9 columns.

Note that in this dataset missing values are coded as “?”.

In addition, from the UCI web site, semi-colon attribute separators are used in the dataset.