R Notebook

Code ▾

Connect to Database and Obtain Data

Hide

```
# Create a database connection
con = dbConnect(MySQL(), user='deepAnalytics', password='Sqltask1234!', dbname='dataanalytics2018', host='data-analytics-2018.cbrosir2cswx.us-east-1.rds.amazonaws.com')
```

Hide

```
#summary of connection
summary(con)
```

```
<MySQLConnection:0,0>
  User:    deepAnalytics
  Host:    data-analytics-2018.cbrosir2cswx.us-east-1.rds.amazonaws.com
  Dbname: dataanalytics2018
  Connection type: data-analytics-2018.cbrosir2cswx.us-east-1.rds.amazonaws.com via TCP/IP

Results:
```

Hide

```
dbGetInfo(con)
```

```
$host
[1] "data-analytics-2018.cbrosir2cswx.us-east-1.rds.amazonaws.com"

$user
[1] "deepAnalytics"

$dbname
[1] "dataanalytics2018"

$conType
[1] "data-analytics-2018.cbrosir2cswx.us-east-1.rds.amazonaws.com via TCP/IP"

$serverVersion
[1] "5.6.10"

$protocolVersion
[1] 10

$threadId
[1] 113664

$rsId
list()
```

Hide

```
# List the tables contained in the database.
my_tables <- dbListTables(con)
my_tables
```

```
[1] "iris"    "yr_2006" "yr_2007" "yr_2008" "yr_2009" "yr_2010"
```

Hide

```
# there are 6 tables: "iris"    "yr_2006" "yr_2007" "yr_2008" "yr_2009" "yr_2010"
```

Hide

```
# Lists attributes contained in a table
list_db_fields_custom_function<- function (x) {dbListFields(con,x)}
lapply(my_tables,list_db_fields_custom_function)
```

```
[[1]]
[1] "id"              "SepalLengthCm" "SepalWidthCm"  "PetalLengthCm" "PetalWidthCm"
[6] "Species"

[[2]]
 [1] "id"                  "Date"                 "Time"
 [4] "Global_active_power" "Global_reactive_power" "Global_intensity"
 [7] "Voltage"             "Sub_metering_1"       "Sub_metering_2"
[10] "Sub_metering_3"

[[3]]
 [1] "id"                  "Date"                 "Time"
 [4] "Global_active_power" "Global_reactive_power" "Global_intensity"
 [7] "Voltage"             "Sub_metering_1"       "Sub_metering_2"
[10] "Sub_metering_3"

[[4]]
 [1] "id"                  "Date"                 "Time"
 [4] "Global_active_power" "Global_reactive_power" "Global_intensity"
 [7] "Voltage"             "Sub_metering_1"       "Sub_metering_2"
[10] "Sub_metering_3"

[[5]]
 [1] "id"                  "Date"                 "Time"
 [4] "Global_active_power" "Global_reactive_power" "Global_intensity"
 [7] "Voltage"             "Sub_metering_1"       "Sub_metering_2"
[10] "Sub_metering_3"

[[6]]
 [1] "id"                  "Date"                 "Time"
 [4] "Global_active_power" "Global_reactive_power" "Global_intensity"
 [7] "Voltage"             "Sub_metering_1"       "Sub_metering_2"
[10] "Sub_metering_3"
```

Hide

```
# tables for the years 2006 -2010 have the same attributes. Column names are the same.
```

Hide

```
# We are only using Date, Time and Submeters for our analysis.

yr_2006SELECT <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub
_metering_3 FROM yr_2006")
yr_2007SELECT <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub
_metering_3 FROM yr_2007")
yr_2008SELECT <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub
_metering_3 FROM yr_2008")
yr_2009SELECT <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub
_metering_3 FROM yr_2009")
yr_2010SELECT <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub
_metering_3 FROM yr_2010")
```

Explore and prepare data

Note: MySQL tables are read into R as data.frames, but without coercing character or logical data into factors. Similarly while exporting data.frames, factors are exported as character vectors. Integer columns are usually imported as R integer vectors, except for cases such as BIGINT or UNSIGNED INTEGER which are coerced to R's double precision vectors to avoid truncation (currently R's integers are signed 32-bit quantities). Time variables are imported/exported as character data, so you need to convert these to your favorite date/time representation.

Investigate Data

Hide

```
# Function to explore tables. Prints out structure, summary, head and tail of data for e
very table.
investigateDF <- function(df) {list(str(df), summary(df),head(df),tail(df))}
```

Hide

```
# Investigates tables from 2006 to 2010
investigateDF(yr_2006SELECT)
```

```
'data.frame':   21992 obs. of  5 variables:
 $ Date         : chr  "2006-12-16" "2006-12-16" "2006-12-16" "2006-12-16" ...
 $ Time         : chr  "17:24:00" "17:25:00" "17:26:00" "17:27:00" ...
 $ Sub_metering_1: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_2: num  1 1 2 1 1 2 1 1 1 2 ...
 $ Sub_metering_3: num  17 16 17 17 17 17 17 17 17 16 ...
[[1]]
NULL

[[2]]
     Date               Time           Sub_metering_1     Sub_metering_2     Sub_metering_3
 Length:21992       Length:21992       Min.   : 0.000     Min.   : 0.000     Min.   : 0.00
 Class :character   Class :character   1st Qu.: 0.000     1st Qu.: 0.000     1st Qu.: 0.00
 Mode  :character   Mode  :character   Median : 0.000     Median : 0.000     Median : 0.00
                                       Mean   : 1.249     Mean   : 2.215     Mean   : 7.41
                                       3rd Qu.: 0.000     3rd Qu.: 1.000     3rd Qu.:17.00
                                       Max.   :77.000     Max.   :74.000     Max.   :20.00

[[3]]
```

Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3
<chr>	<chr>	<dbl>	<dbl>	<dbl>
1 2006-12-16	17:24:00	0	1	17
2 2006-12-16	17:25:00	0	1	16
3 2006-12-16	17:26:00	0	2	17

	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
4	2006-12-16	17:27:00	0	1	17
5	2006-12-16	17:28:00	0	1	17
6	2006-12-16	17:29:00	0	2	17

6 rows

[[4]]

	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
21987	2006-12-31	23:54:00	0	0	0
21988	2006-12-31	23:55:00	0	0	0
21989	2006-12-31	23:56:00	0	0	0
21990	2006-12-31	23:57:00	0	0	0
21991	2006-12-31	23:58:00	0	0	0
21992	2006-12-31	23:59:00	0	0	0

6 rows

Hide

```
investigateDF(yr_2007SELECT)
```

```
'data.frame':   521669 obs. of  5 variables:
 $ Date          : chr  "2007-01-01" "2007-01-01" "2007-01-01" "2007-01-01" ...
 $ Time          : chr  "00:00:00" "00:01:00" "00:02:00" "00:03:00" ...
 $ Sub_metering_1: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_2: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_3: num  0 0 0 0 0 0 0 0 0 0 ...
[[1]]
NULL

[[2]]
     Date                Time            Sub_metering_1    Sub_metering_2    Sub_metering_3
 Length:521669      Length:521669      Min.   : 0.000    Min.   : 0.000    Min.   : 0.000
 Class :character   Class :character   1st Qu.: 0.000    1st Qu.: 0.000    1st Qu.: 0.000
 Mode  :character   Mode  :character   Median : 0.000    Median : 0.000    Median : 0.000
                                       Mean   : 1.232    Mean   : 1.638    Mean   : 5.795
                                       3rd Qu.: 0.000    3rd Qu.: 1.000    3rd Qu.:17.000
                                       Max.   :78.000    Max.   :78.000    Max.   :20.000

[[3]]
```

	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl>
1	2007-01-01	00:00:00	0	0	0
2	2007-01-01	00:01:00	0	0	0
3	2007-01-01	00:02:00	0	0	0
4	2007-01-01	00:03:00	0	0	0
5	2007-01-01	00:04:00	0	0	0
6	2007-01-01	00:05:00	0	0	0

6 rows

```
[[4]]
```

	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl>
521664	2007-12-31	23:54:00	0	0	18
521665	2007-12-31	23:55:00	0	0	18
521666	2007-12-31	23:56:00	0	0	18
521667	2007-12-31	23:57:00	0	0	18
521668	2007-12-31	23:58:00	0	0	18
521669	2007-12-31	23:59:00	0	0	18

6 rows

Hide

```
investigateDF(yr_2008SELECT)
```

```
'data.frame':   526905 obs. of  5 variables:
 $ Date         : chr  "2008-01-01" "2008-01-01" "2008-01-01" "2008-01-01" ...
 $ Time         : chr  "00:00:00" "00:01:00" "00:02:00" "00:03:00" ...
 $ Sub_metering_1: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_2: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_3: num  18 18 18 18 18 17 18 18 18 18 ...
[[1]]
NULL


[[2]]
     Date               Time           Sub_metering_1   Sub_metering_2   Sub_metering_3
 Length:526905      Length:526905      Min.   : 0.00    Min.   : 0.000   Min.   : 0.000
 Class :character   Class :character   1st Qu.: 0.00    1st Qu.: 0.000   1st Qu.: 0.000
 Mode  :character   Mode  :character   Median : 0.00    Median : 0.000   Median : 1.000
                                       Mean   : 1.11    Mean   : 1.256   Mean   : 6.034
                                       3rd Qu.: 0.00    3rd Qu.: 1.000   3rd Qu.:17.000
                                       Max.   :80.00    Max.   :76.000   Max.   :31.000


[[3]]
```

Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl>
1 2008-01-01	00:00:00	0	0	18
2 2008-01-01	00:01:00	0	0	18
3 2008-01-01	00:02:00	0	0	18
4 2008-01-01	00:03:00	0	0	18
5 2008-01-01	00:04:00	0	0	18
6 2008-01-01	00:05:00	0	0	17

6 rows

```
[[4]]
```

	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl>
526900	2008-12-31	23:54:00	0	0	0
526901	2008-12-31	23:55:00	0	0	0
526902	2008-12-31	23:56:00	0	0	0
526903	2008-12-31	23:57:00	0	0	0
526904	2008-12-31	23:58:00	0	0	0
526905	2008-12-31	23:59:00	0	0	0

6 rows

```
investigateDF(yr_2009SELECT)
```

```
'data.frame':   521320 obs. of  5 variables:
 $ Date           : chr  "2009-01-01" "2009-01-01" "2009-01-01" "2009-01-01" ...
 $ Time           : chr  "00:00:00" "00:01:00" "00:02:00" "00:03:00" ...
 $ Sub_metering_1: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_2: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_3: num  0 0 0 0 0 0 0 0 0 0 ...
[[1]]
NULL

[[2]]
      Date              Time           Sub_metering_1    Sub_metering_2    Sub_metering_3
 Length:521320     Length:521320     Min.   : 0.000    Min.   : 0.000    Min.   : 0.000
 Class :character  Class :character  1st Qu.: 0.000    1st Qu.: 0.000    1st Qu.: 0.000
 Mode  :character  Mode  :character  Median : 0.000    Median : 0.000    Median : 1.000
                                     Mean   : 1.137    Mean   : 1.136    Mean   : 6.823
                                     3rd Qu.: 0.000    3rd Qu.: 1.000    3rd Qu.:18.000
                                     Max.   :82.000    Max.   :77.000    Max.   :31.000

[[3]]
```

Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3
<chr>	<chr>	<dbl>	<dbl>	<dbl>
1 2009-01-01	00:00:00	0	0	0
2 2009-01-01	00:01:00	0	0	0
3 2009-01-01	00:02:00	0	0	0
4 2009-01-01	00:03:00	0	0	0
5 2009-01-01	00:04:00	0	0	0
6 2009-01-01	00:05:00	0	0	0

6 rows

```
[[4]]
```

	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
521315	2009-12-31	23:54:00	0	0	18
521316	2009-12-31	23:55:00	0	0	18

	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
521317	2009-12-31	23:56:00	0	0	19
521318	2009-12-31	23:57:00	0	0	18
521319	2009-12-31	23:58:00	0	0	18
521320	2009-12-31	23:59:00	0	0	19
6 rows					

Hide

```
investigateDF(yr_2010SELECT)
```

```
'data.frame':	457394 obs. of  5 variables:
 $ Date          : chr  "2010-01-01" "2010-01-01" "2010-01-01" "2010-01-01" ...
 $ Time          : chr  "00:00:00" "00:01:00" "00:02:00" "00:03:00" ...
 $ Sub_metering_1: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_2: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_3: num  18 18 19 18 18 19 18 18 19 18 ...
[[1]]
NULL

[[2]]
     Date               Time            Sub_metering_1   Sub_metering_2    Sub_metering_3
 Length:457394     Length:457394     Min.   : 0.0000   Min.   : 0.000    Min.   : 0.000
 Class :character  Class :character  1st Qu.: 0.0000   1st Qu.: 0.000    1st Qu.: 1.000
 Mode  :character  Mode  :character  Median : 0.0000   Median : 0.000    Median : 1.000
                                     Mean   : 0.9875   Mean   : 1.102    Mean   : 7.244
                                     3rd Qu.: 0.0000   3rd Qu.: 1.000    3rd Qu.:18.000
                                     Max.   :88.0000   Max.   :80.000    Max.   :31.000

[[3]]
```

	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
1	2010-01-01	00:00:00	0	0	18
2	2010-01-01	00:01:00	0	0	18
3	2010-01-01	00:02:00	0	0	19
4	2010-01-01	00:03:00	0	0	18
5	2010-01-01	00:04:00	0	0	18
6	2010-01-01	00:05:00	0	0	19
6 rows					

```
[[4]]
```

	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl>
457389	2010-11-26	20:57:00	0	0	0
457390	2010-11-26	20:58:00	0	0	0
457391	2010-11-26	20:59:00	0	0	0
457392	2010-11-26	21:00:00	0	0	0
457393	2010-11-26	21:01:00	0	0	0
457394	2010-11-26	21:02:00	0	0	0

6 rows

```
NA
```

Hide

```
#Combine tables into one dataframe (using dplyr)
df2006_2010 <- bind_rows(yr_2006SELECT,yr_2007SELECT,yr_2008SELECT,yr_2009SELECT,yr_2010
SELECT)
investigateDF(df2006_2010)
```

```
'data.frame':	2049280 obs. of  5 variables:
 $ Date         : chr  "2006-12-16" "2006-12-16" "2006-12-16" "2006-12-16" ...
 $ Time         : chr  "17:24:00" "17:25:00" "17:26:00" "17:27:00" ...
 $ Sub_metering_1: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_2: num  1 1 2 1 1 2 1 1 1 2 ...
 $ Sub_metering_3: num  17 16 17 17 17 17 17 17 17 16 ...
[[1]]
NULL


[[2]]
     Date                Time           Sub_metering_1     Sub_metering_2
 Length:2049280      Length:2049280      Min.   : 0.000     Min.   : 0.000
 Class :character    Class :character    1st Qu.: 0.000     1st Qu.: 0.000
 Mode  :character    Mode  :character    Median : 0.000     Median : 0.000
                                         Mean   : 1.122     Mean   : 1.299
                                         3rd Qu.: 0.000     3rd Qu.: 1.000
                                         Max.   :88.000     Max.   :80.000

 Sub_metering_3
 Min.   : 0.000
 1st Qu.: 0.000
 Median : 1.000
 Mean   : 6.458
 3rd Qu.:17.000
 Max.   :31.000


[[3]]
```

| | Date | Time | Sub_metering_1 | Sub_metering_2 | Sub_metering_3 |
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
1	2006-12-16	17:24:00	0	1	17
2	2006-12-16	17:25:00	0	1	16
3	2006-12-16	17:26:00	0	2	17
4	2006-12-16	17:27:00	0	1	17
5	2006-12-16	17:28:00	0	1	17
6	2006-12-16	17:29:00	0	2	17

6 rows

```
[[4]]
```

| | Date | Time | Sub_metering_1 | Sub_metering_2 | Sub_metering_3 |
	<chr>	<chr>	<dbl>	<dbl>	<dbl>
2049275	2010-11-26	20:57:00	0	0	0
2049276	2010-11-26	20:58:00	0	0	0

	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl>
2049277	2010-11-26	20:59:00	0	0	0
2049278	2010-11-26	21:00:00	0	0	0
2049279	2010-11-26	21:01:00	0	0	0
2049280	2010-11-26	21:02:00	0	0	0

6 rows

NA

Create DateTime Objects

```
# Combine Date and Time attribute values in a new attribute column
df2006_2010 <-cbind(df2006_2010,paste(df2006_2010$Date,df2006_2010$Time), stringsAsFacto
rs=FALSE)
head(df2006_2010)
```

	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl> ▶
1	2006-12-16	17:24:00	0	1	17
2	2006-12-16	17:25:00	0	1	16
3	2006-12-16	17:26:00	0	2	17
4	2006-12-16	17:27:00	0	1	17
5	2006-12-16	17:28:00	0	1	17
6	2006-12-16	17:29:00	0	2	17

6 rows | 1-6 of 6 columns

```
# Give the new attribute in the 6th column a header name
colnames(df2006_2010)[6] <-"DateTime"
head(df2006_2010)
```

	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering_3 <dbl>	DateTime <chr>
1	2006-12-16	17:24:00	0	1	17	2006-12-16 17:24:00
2	2006-12-16	17:25:00	0	1	16	2006-12-16 17:25:00

	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering_3	DateTime
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<chr>
3	2006-12-16	17:26:00	0	2	17	2006-12-16 17:26:00
4	2006-12-16	17:27:00	0	1	17	2006-12-16 17:27:00
5	2006-12-16	17:28:00	0	1	17	2006-12-16 17:28:00
6	2006-12-16	17:29:00	0	2	17	2006-12-16 17:29:00

6 rows

Hide

```
# Move the DateTime attribute within the dataset
df2006_2010 <- df2006_2010[,c(ncol(df2006_2010), 1:(ncol(df2006_2010)-1))]
head(df2006_2010)
```

	DateTime	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<db
1	2006-12-16 17:24:00	2006-12-16	17:24:00	0	1	
2	2006-12-16 17:25:00	2006-12-16	17:25:00	0	1	
3	2006-12-16 17:26:00	2006-12-16	17:26:00	0	2	
4	2006-12-16 17:27:00	2006-12-16	17:27:00	0	1	
5	2006-12-16 17:28:00	2006-12-16	17:28:00	0	1	
6	2006-12-16 17:29:00	2006-12-16	17:29:00	0	2	

6 rows

Hide

```
# Convert DateTime from character to POSIXct
df2006_2010$DateTime <- as.POSIXct(df2006_2010$DateTime, "%Y/%m/%d %H:%M:%S")
```

```
unknown timezone '%Y/%m/%d %H:%M:%S'unknown timezone '%Y/%m/%d %H:%M:%S'unknown timezone
'%Y/%m/%d %H:%M:%S'unknown timezone '%Y/%m/%d %H:%M:%S'
```

Hide

```
attr(df2006_2010$DateTime, "tzone") <- "Europe/Paris"

#Verify
str(df2006_2010)
```

```
'data.frame':   2049280 obs. of  6 variables:
 $ DateTime     : POSIXct, format: "2006-12-16 18:24:00" "2006-12-16 18:25:00" ...
 $ Date         : chr  "2006-12-16" "2006-12-16" "2006-12-16" "2006-12-16" ...
 $ Time         : chr  "17:24:00" "17:25:00" "17:26:00" "17:27:00" ...
 $ Sub_metering_1: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sub_metering_2: num  1 1 2 1 1 2 1 1 1 2 ...
 $ Sub_metering_3: num  17 16 17 17 17 17 17 17 17 16 ...
```

Hide

```
head(df2006_2010)
```

	DateTime	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering
	<S3: POSIXct>	<chr>	<chr>	<dbl>	<dbl>	<db
1	2006-12-16 18:24:00	2006-12-16	17:24:00	0	1	
2	2006-12-16 18:25:00	2006-12-16	17:25:00	0	1	
3	2006-12-16 18:26:00	2006-12-16	17:26:00	0	2	
4	2006-12-16 18:27:00	2006-12-16	17:27:00	0	1	
5	2006-12-16 18:28:00	2006-12-16	17:28:00	0	1	
6	2006-12-16 18:29:00	2006-12-16	17:29:00	0	2	

6 rows

Hide

```
# Create "year, quarter, month, week, weekday, day, dateTZ(different than original date
  [chr string] with time zone applied], hour, and minute attributes
df2006_2010$year <- year(df2006_2010$DateTime)
df2006_2010$quarter <- quarter(df2006_2010$DateTime)
df2006_2010$month <- month(df2006_2010$DateTime)
df2006_2010$week <- week(df2006_2010$DateTime)
df2006_2010$weekday <- weekdays(df2006_2010$DateTime)
df2006_2010$day <- day(df2006_2010$DateTime)
df2006_2010$dateTZ <- date(df2006_2010$DateTime)
df2006_2010$hour <- hour(df2006_2010$DateTime)
df2006_2010$minute <- minute(df2006_2010$DateTime)
```

Hide

```
# verify new attributes
head(df2006_2010)
```

	DateTime	Date	Time	Sub_metering_1	Sub_metering_2	Sub_metering
	<S3: POSIXct>	<chr>	<chr>	<dbl>	<dbl>	<db
1	2006-12-16 18:24:00	2006-12-16	17:24:00	0	1	
2	2006-12-16 18:25:00	2006-12-16	17:25:00	0	1	

	DateTime <S3: POSIXct>	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_metering <db
3	2006-12-16 18:26:00	2006-12-16	17:26:00	0	2	
4	2006-12-16 18:27:00	2006-12-16	17:27:00	0	1	
5	2006-12-16 18:28:00	2006-12-16	17:28:00	0	1	
6	2006-12-16 18:29:00	2006-12-16	17:29:00	0	2	

6 rows | 1-8 of 15 columns

Hide

```
tail(df2006_2010)
```

	DateTime <S3: POSIXct>	Date <chr>	Time <chr>	Sub_metering_1 <dbl>	Sub_metering_2 <dbl>	Sub_n
2049275	2010-11-26 21:57:00	2010-11-26	20:57:00	0	0	
2049276	2010-11-26 21:58:00	2010-11-26	20:58:00	0	0	
2049277	2010-11-26 21:59:00	2010-11-26	20:59:00	0	0	
2049278	2010-11-26 22:00:00	2010-11-26	21:00:00	0	0	
2049279	2010-11-26 22:01:00	2010-11-26	21:01:00	0	0	
2049280	2010-11-26 22:02:00	2010-11-26	21:02:00	0	0	

6 rows | 1-8 of 15 columns

Any NAs?

Hide

```
sum(is.na(df2006_2010))
```

```
[1] 0
```

Hide

```
# no missing values
```

Data Documentation

Source: http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption#
(http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption#)

Abstract: 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.

Attribute Information:

sub_metering_1: energy sub-metering No. 1 (in). It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas powered).

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.

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.