# C4 - Data Analytics

'Smart Home' Project

Objective: To find evidence to help support marketing claims of sub-meters providing owners with 'useful' power usage analytics. Perform an analytical deep dive of sub-metering generated data and producing high quality visualizations taht support a positive narrative around the findings.

Data: 47 months of energy usage data
##C4T1 ### IOT Analytics and their onboarding process:

(1) Conduct research to become informed on the client's business
(2) Identify any analytic skill/knowledge gaps foreseen for the project and plug those gaps with self-learning.
(3) Perform an initial exploration of the data
(4) Hold a project kick-off meeting with the client to close the deal

Produce a PPT for the home developer's management team.

Include:
1. How we will conduct the analysis
2. What they're likely to gain
3. Keep in mind this is business people not technical

**Produce an initial report for IOT Analytics' clients. PPT including initial insights of business relevance.**

**Research the Domain**

- What role do sub-meters play in the power industry?
- What kinds of power usage analytics are currently offered?
- What can be learned from the analytics?
- What are the benefits to consumers?

Also analyze the data before making the EDA;

- How is power measured?
- In what kind of units?
- What household areas are each of the sub-meters measureing?
- Is there additional power and power realted information that would benefit analytics in the future if added to the data set?
- Are there any changes to the sub-metering data collection structure that would help future analytics?

**Take aways from Predictive Analytics Myths** Summary

Although Predictive Analytics is a powerful optimization technique, it is not always the best solution. Even though we know that nothing in life is 100% guaranteed, we often overlook this fact of life in the way we use predictive analytics for our daily business use. In reality the projects that drive measurable business results encompass more than just good models. They typically incorporate an effective process such as Aryng's BADIR™: 5 steps from "data to decisions"™ framework. Key takeaways from this paper are:

1. Process (BADIR) is key to good business result
2. Right Talent + Good Tools = Great Models 
3. Models needs to be maintained
4. Models are not perfect
5. Use simpler techniques till PA can be justified

### Dplyr Summary

- *Piping* - chaining 2+ functions together using the pipe operator or %>%. Use pipe betwen two functions you want to combine, executed in the order they're written. The result of the first funciton is transfered into the second.
- inside dplyr "<-" doesnt exist, we use "=" instead
- no quotes used around variable names (some exeptions)
- do not use $ to refer to variables within data frame, the pipe auto does it.

Basic Struture

**new.df <- old.df %>% function1(.) %>% function2(.)**

this creates a new data frame from an old one, where the old is dumped into function1 and then the results are dumped into function2.

Using the pipe operator i dont need to input the data frame as the first argument.

Some helpful functions:

- filter(df, Prey.species != "Unknown" & nchar(Prey.species) > 0)
- mutate(Prey.species = tolower(Prey.species))
- group_by(Lion.ID, Prey.species) This will show the number of species that each lion killed.
- summarise( nb_killed = sum(State..kill..1..or.non.kill..0..)) This aggregates the results of the previews one
- arrange(desc(nb_kille)) arranges in decending order
- select(Lion = Lion.ID, renames the columns at the end Prey = Prey.species, nb_killed)

### Data Collection

```
#only need once
#install.packages("RMariaDB")
#library(RMariaDB)
```

```
require(pacman)
```

```
## Loading required package: pacman
```

```
pacman:: p_load(pacman, dplyr, GGally, ggplot2, ggrepel, patchwork, gifski, ggforce, ggthemes, maps, sf
```

```
## Installing package into 'C:/Users/chelo/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## Warning: package 'TinyTeX' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```
## Warning: Perhaps you meant 'tinytex' ?
```

```
## Warning: unable to access index for repository http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contril
##   cannot open URL 'http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib/4.1/PACKAGES'
```

```
## Warning in p_install(package, character.only = TRUE, ...):
```

```
## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'TinyTeX'
```

```
## Installing package into 'C:/Users/chelo/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## Warning: package 'Latex' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```
## Warning: unable to access index for repository http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contril
##   cannot open URL 'http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib/4.1/PACKAGES'
```

```
## Warning in p_install(package, character.only = TRUE, ...):
```

```
## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'Latex'
```

```
## Warning in pacman::p_load(pacman, dplyr, GGally, ggplot2, ggrepel, patchwork, : Failed to install/loa
## TinyTeX, Latex
```

```
con = dbConnect(MariaDB(), user='deepAnalytics', password='Sqltask1234!',
                dbname='dataanalytics2018', host='data-analytics-2018.cbrosir2cswx.us-east-1.rds.amazona

dbListTables(con)
```

### Creating a Database connection

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

Using the dbListsFields function you can learn the attributes associated with a particular table (yr_2006).

```
## Lists attributes contained in a table
dbListFields(con, 'yr_2006')
```

```
##  [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"
```

```
yr_2006 <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub_metering_3 FROM yr_2
yr_2007 <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub_metering_3 FROM yr_2
yr_2008 <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub_metering_3 FROM yr_2
yr_2009 <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub_metering_3 FROM yr_2
yr_2010 <- dbGetQuery(con, "SELECT Date, Time, Sub_metering_1, Sub_metering_2, Sub_metering_3 FROM yr_2
```

### Pulling the necessary data into R

```
head(yr_2006)
```

### Investigation of Data

```
##         Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
tail(yr_2006)
```

```
##             Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
#--------------
head(yr_2007)
```

```
##         Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
tail(yr_2007)
```

```
##              Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
#--------------
head(yr_2008)
```

```
##         Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
tail(yr_2008)
```

```
##              Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
#--------------
head(yr_2009)
```

```
##         Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
tail(yr_2009)
```

```
##              Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 521315 2009-12-31 23:54:00              0              0             18
## 521316 2009-12-31 23:55:00              0              0             18
## 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
```

```
#--------------
head(yr_2010)
```

```
##         Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
tail(yr_2010)
```

```
##              Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

This tells me that 2006 and 2010 we dont have complete data for all the days. 2006 has 16-Dec to 31-Dec; 2010 has 1-Jan to 26-Nov.

```
alldata <- bind_rows(yr_2007,yr_2008, yr_2009)

head(alldata)
```

**Combining tables into one dataFrame using dplyr**

```
##         Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 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
```

```
tail(alldata)
```

```
##               Date     Time Sub_metering_1 Sub_metering_2 Sub_metering_3
## 1569889 2009-12-31 23:54:00              0              0             18
## 1569890 2009-12-31 23:55:00              0              0             18
## 1569891 2009-12-31 23:56:00              0              0             19
## 1569892 2009-12-31 23:57:00              0              0             18
## 1569893 2009-12-31 23:58:00              0              0             18
## 1569894 2009-12-31 23:59:00              0              0             19
```

Data is in the correct order! :)

```
alldata <- cbind(alldata, paste(alldata$Date, alldata$Time),
                 stringsAsFactors=FALSE)

### Giving the new attribute a name

colnames(alldata)[6] <- "DateTime"

alldata <- alldata[,c(ncol(alldata), 1:(ncol(alldata)-1))]
head(alldata)
```

**Combine Date and Time**

```
##              DateTime       Date     Time Sub_metering_1 Sub_metering_2
## 1 2007-01-01 00:00:00 2007-01-01 00:00:00              0              0
## 2 2007-01-01 00:01:00 2007-01-01 00:01:00              0              0
## 3 2007-01-01 00:02:00 2007-01-01 00:02:00              0              0
## 4 2007-01-01 00:03:00 2007-01-01 00:03:00              0              0
## 5 2007-01-01 00:04:00 2007-01-01 00:04:00              0              0
## 6 2007-01-01 00:05:00 2007-01-01 00:05:00              0              0
##   Sub_metering_3
## 1              0
## 2              0
## 3              0
## 4              0
## 5              0
## 6              0
```

```
alldata$DateTime <- as.POSIXct(alldata$DateTime, tz="Europe/Paris",  "%Y-%m-%d %H:%M:%S")  #converts fr
#attr(alldata$DateTime, "tzone") <- "Europe/Paris"  #adds time zone to avoid problems
str(alldata)
```

### Changing data to POSIXct

```
## 'data.frame':    1569894 obs. of  6 variables:
##  $ DateTime      : POSIXct, format: "2007-01-01 00:00:00" "2007-01-01 00:01:00" ...
##  $ 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 ...
```

```
head(alldata)
```

```
##              DateTime       Date     Time Sub_metering_1 Sub_metering_2
## 1 2007-01-01 00:00:00 2007-01-01 00:00:00              0              0
## 2 2007-01-01 00:01:00 2007-01-01 00:01:00              0              0
## 3 2007-01-01 00:02:00 2007-01-01 00:02:00              0              0
## 4 2007-01-01 00:03:00 2007-01-01 00:03:00              0              0
## 5 2007-01-01 00:04:00 2007-01-01 00:04:00              0              0
## 6 2007-01-01 00:05:00 2007-01-01 00:05:00              0              0
##   Sub_metering_3
## 1              0
## 2              0
## 3              0
## 4              0
## 5              0
## 6              0
```

```
alldata$year <- year(alldata$DateTime)
alldata$quarter <- quarter(alldata$DateTime)
alldata$month <- month(alldata$DateTime)
alldata$week <- week(alldata$DateTime)
alldata$weekdays <- weekdays(alldata$DateTime)
alldata$day <- day(alldata$DateTime)
alldata$hour <- hour(alldata$DateTime)
alldata$minute <- minute(alldata$DateTime)
head(alldata)
```

### Lubridate to extract data

```
##              DateTime       Date     Time Sub_metering_1 Sub_metering_2
## 1 2007-01-01 00:00:00 2007-01-01 00:00:00              0              0
## 2 2007-01-01 00:01:00 2007-01-01 00:01:00              0              0
## 3 2007-01-01 00:02:00 2007-01-01 00:02:00              0              0
## 4 2007-01-01 00:03:00 2007-01-01 00:03:00              0              0
```

```
## 5 2007-01-01 00:04:00 2007-01-01 00:04:00              0              0
## 6 2007-01-01 00:05:00 2007-01-01 00:05:00              0              0
##   Sub_metering_3 year quarter month week weekdays day hour minute
## 1              0 2007       1     1    1   Monday   1    0      0
## 2              0 2007       1     1    1   Monday   1    0      1
## 3              0 2007       1     1    1   Monday   1    0      2
## 4              0 2007       1     1    1   Monday   1    0      3
## 5              0 2007       1     1    1   Monday   1    0      4
## 6              0 2007       1     1    1   Monday   1    0      5
```

**Data Documentation**

Each sub-meter is measured in watt-hour and is the energy consumed in:

Energy sub-metering 1 = kitchen (dishwasher, oven, microwave)
Energy sub-metering 2 = laundry room (washing machine, tumble drier, refrigerator, light)
Energy sub-metering 3 = electric water heater and Air Conditioner.

- can you find what is using the most power? the least?
- anything to learn from the min and max?
- Research other basic statistics

```
alldata %>% select(Sub_metering_1, Sub_metering_2, Sub_metering_3, year) %>% filter(year == 2007) %>% su
```

```
##  Sub_metering_1   Sub_metering_2   Sub_metering_3        year
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   :2007
##  1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.:2007
##  Median : 0.000   Median : 0.000   Median : 0.000   Median :2007
##  Mean   : 1.232   Mean   : 1.638   Mean   : 5.795   Mean   :2007
##  3rd Qu.: 0.000   3rd Qu.: 1.000   3rd Qu.:17.000   3rd Qu.:2007
##  Max.   :78.000   Max.   :78.000   Max.   :20.000   Max.   :2007
```

```
alldata %>% select(Sub_metering_1, Sub_metering_2, Sub_metering_3, year) %>% filter(year == 2008) %>% su
```

```
##  Sub_metering_1  Sub_metering_2   Sub_metering_3        year
##  Min.   : 0.00   Min.   : 0.000   Min.   : 0.000   Min.   :2008
##  1st Qu.: 0.00   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.:2008
##  Median : 0.00   Median : 0.000   Median : 1.000   Median :2008
##  Mean   : 1.11   Mean   : 1.256   Mean   : 6.033   Mean   :2008
##  3rd Qu.: 0.00   3rd Qu.: 1.000   3rd Qu.:17.000   3rd Qu.:2008
##  Max.   :80.00   Max.   :76.000   Max.   :31.000   Max.   :2008
```

```
alldata %>% select(Sub_metering_1, Sub_metering_2, Sub_metering_3, year) %>% filter(year == 2009) %>% su
```

```
##  Sub_metering_1   Sub_metering_2   Sub_metering_3        year
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   :2009
##  1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.:2009
##  Median : 0.000   Median : 0.000   Median : 1.000   Median :2009
##  Mean   : 1.137   Mean   : 1.136   Mean   : 6.823   Mean   :2009
##  3rd Qu.: 0.000   3rd Qu.: 1.000   3rd Qu.:18.000   3rd Qu.:2009
##  Max.   :82.000   Max.   :77.000   Max.   :31.000   Max.   :2009
```

```
group_by(alldata, year) %>%
  summarise(Energy_Meter_1 = sum(Sub_metering_1),
            Energy_Meter_2 = sum(Sub_metering_2),
            Energy_Meter_3 = sum(Sub_metering_3))
```

```
## # A tibble: 4 x 4
##    year Energy_Meter_1 Energy_Meter_2 Energy_Meter_3
##   <dbl>          <dbl>          <dbl>          <dbl>
## 1  2007         642548         854350        3022840
## 2  2008         584631         661905        3178353
## 3  2009         592657         592145        3556816
## 4    NA            153             10            834
```

Why are there so many NA values? Lets check:

```
alldata[!complete.cases(alldata),]
```

```
##         DateTime       Date     Time Sub_metering_1 Sub_metering_2
## 119637      <NA> 2007-03-25 02:00:00               0               0
## 119638      <NA> 2007-03-25 02:01:00               0               0
## 119639      <NA> 2007-03-25 02:02:00               0               1
## 119640      <NA> 2007-03-25 02:03:00               0               0
## 119641      <NA> 2007-03-25 02:04:00               0               0
## 119642      <NA> 2007-03-25 02:05:00               0               0
## 119643      <NA> 2007-03-25 02:06:00               0               0
## 119644      <NA> 2007-03-25 02:07:00               0               0
## 119645      <NA> 2007-03-25 02:08:00               0               0
## 119646      <NA> 2007-03-25 02:09:00               0               0
## 119647      <NA> 2007-03-25 02:10:00               0               1
## 119648      <NA> 2007-03-25 02:11:00               0               0
## 119649      <NA> 2007-03-25 02:12:00               0               0
## 119650      <NA> 2007-03-25 02:13:00               0               0
## 119651      <NA> 2007-03-25 02:14:00               0               0
## 119652      <NA> 2007-03-25 02:15:00               0               0
## 119653      <NA> 2007-03-25 02:16:00               0               1
## 119654      <NA> 2007-03-25 02:17:00               0               0
## 119655      <NA> 2007-03-25 02:18:00               0               0
## 119656      <NA> 2007-03-25 02:19:00               0               0
## 119657      <NA> 2007-03-25 02:20:00               0               0
## 119658      <NA> 2007-03-25 02:21:00               0               0
## 119659      <NA> 2007-03-25 02:22:00               0               1
## 119660      <NA> 2007-03-25 02:23:00               0               0
## 119661      <NA> 2007-03-25 02:24:00               0               0
## 119662      <NA> 2007-03-25 02:25:00               0               0
## 119663      <NA> 2007-03-25 02:26:00               0               0
## 119664      <NA> 2007-03-25 02:27:00               0               0
## 119665      <NA> 2007-03-25 02:28:00               0               1
## 119666      <NA> 2007-03-25 02:29:00               0               0
## 119667      <NA> 2007-03-25 02:30:00               0               0
## 119668      <NA> 2007-03-25 02:31:00               0               0
## 119669      <NA> 2007-03-25 02:32:00               0               0
## 119670      <NA> 2007-03-25 02:33:00               0               1
```

```
## 119671      <NA> 2007-03-25 02:34:00              0              0
## 119672      <NA> 2007-03-25 02:35:00              0              0
## 119673      <NA> 2007-03-25 02:36:00              0              0
## 119674      <NA> 2007-03-25 02:37:00              0              0
## 119675      <NA> 2007-03-25 02:38:00              0              0
## 119676      <NA> 2007-03-25 02:39:00              0              1
## 119677      <NA> 2007-03-25 02:40:00              0              0
## 119678      <NA> 2007-03-25 02:41:00              0              0
## 119679      <NA> 2007-03-25 02:42:00              0              0
## 119680      <NA> 2007-03-25 02:43:00              0              0
## 119681      <NA> 2007-03-25 02:44:00              0              0
## 119682      <NA> 2007-03-25 02:45:00              0              1
## 119683      <NA> 2007-03-25 02:46:00              0              0
## 119684      <NA> 2007-03-25 02:47:00              0              0
## 119685      <NA> 2007-03-25 02:48:00              0              0
## 119686      <NA> 2007-03-25 02:49:00              0              0
## 119687      <NA> 2007-03-25 02:50:00              0              0
## 119688      <NA> 2007-03-25 02:51:00              0              1
## 119689      <NA> 2007-03-25 02:52:00              0              0
## 119690      <NA> 2007-03-25 02:53:00              0              0
## 119691      <NA> 2007-03-25 02:54:00              0              0
## 119692      <NA> 2007-03-25 02:55:00              0              0
## 119693      <NA> 2007-03-25 02:56:00              0              0
## 119694      <NA> 2007-03-25 02:57:00              0              1
## 119695      <NA> 2007-03-25 02:58:00              0              0
## 119696      <NA> 2007-03-25 02:59:00              0              0
## 649945      <NA> 2008-03-30 02:00:00             37              0
## 649946      <NA> 2008-03-30 02:01:00             38              0
## 649947      <NA> 2008-03-30 02:02:00             10              0
## 649948      <NA> 2008-03-30 02:03:00              1              0
## 649949      <NA> 2008-03-30 02:04:00              1              0
## 649950      <NA> 2008-03-30 02:05:00              1              0
## 649951      <NA> 2008-03-30 02:06:00              1              0
## 649952      <NA> 2008-03-30 02:07:00              2              0
## 649953      <NA> 2008-03-30 02:08:00              1              0
## 649954      <NA> 2008-03-30 02:09:00              1              0
## 649955      <NA> 2008-03-30 02:10:00              1              0
## 649956      <NA> 2008-03-30 02:11:00              1              0
## 649957      <NA> 2008-03-30 02:12:00              2              0
## 649958      <NA> 2008-03-30 02:13:00              1              0
## 649959      <NA> 2008-03-30 02:14:00              1              0
## 649960      <NA> 2008-03-30 02:15:00              1              0
## 649961      <NA> 2008-03-30 02:16:00              1              0
## 649962      <NA> 2008-03-30 02:17:00              2              0
## 649963      <NA> 2008-03-30 02:18:00              1              0
## 649964      <NA> 2008-03-30 02:19:00              1              0
## 649965      <NA> 2008-03-30 02:20:00              1              0
## 649966      <NA> 2008-03-30 02:21:00              2              0
## 649967      <NA> 2008-03-30 02:22:00              1              0
## 649968      <NA> 2008-03-30 02:23:00              1              0
## 649969      <NA> 2008-03-30 02:24:00              1              0
## 649970      <NA> 2008-03-30 02:25:00              1              0
## 649971      <NA> 2008-03-30 02:26:00              2              0
## 649972      <NA> 2008-03-30 02:27:00              1              0
```

```
## 649973      <NA> 2008-03-30 02:28:00             1             0
## 649974      <NA> 2008-03-30 02:29:00             1             0
## 649975      <NA> 2008-03-30 02:30:00             1             0
## 649976      <NA> 2008-03-30 02:31:00             2             0
## 649977      <NA> 2008-03-30 02:32:00             1             0
## 649978      <NA> 2008-03-30 02:33:00             1             0
## 649979      <NA> 2008-03-30 02:34:00             1             0
## 649980      <NA> 2008-03-30 02:35:00             1             0
## 649981      <NA> 2008-03-30 02:36:00             2             0
## 649982      <NA> 2008-03-30 02:37:00             1             0
## 649983      <NA> 2008-03-30 02:38:00             1             0
## 649984      <NA> 2008-03-30 02:39:00             1             0
## 649985      <NA> 2008-03-30 02:40:00             2             0
## 649986      <NA> 2008-03-30 02:41:00             1             0
## 649987      <NA> 2008-03-30 02:42:00             1             0
## 649988      <NA> 2008-03-30 02:43:00             1             0
## 649989      <NA> 2008-03-30 02:44:00             2             0
## 649990      <NA> 2008-03-30 02:45:00             1             0
## 649991      <NA> 2008-03-30 02:46:00             1             0
## 649992      <NA> 2008-03-30 02:47:00             1             0
## 649993      <NA> 2008-03-30 02:48:00             1             0
## 649994      <NA> 2008-03-30 02:49:00             2             0
## 649995      <NA> 2008-03-30 02:50:00             1             0
## 649996      <NA> 2008-03-30 02:51:00             1             0
## 649997      <NA> 2008-03-30 02:52:00             1             0
## 649998      <NA> 2008-03-30 02:53:00             2             0
## 649999      <NA> 2008-03-30 02:54:00             0             0
## 650000      <NA> 2008-03-30 02:55:00             1             0
## 650001      <NA> 2008-03-30 02:56:00             1             0
## 650002      <NA> 2008-03-30 02:57:00             1             0
## 650003      <NA> 2008-03-30 02:58:00             2             0
## 650004      <NA> 2008-03-30 02:59:00             1             0
## 1173908     <NA> 2009-03-29 02:00:00             0             0
## 1173909     <NA> 2009-03-29 02:01:00             0             0
## 1173910     <NA> 2009-03-29 02:02:00             0             0
## 1173911     <NA> 2009-03-29 02:03:00             0             0
## 1173912     <NA> 2009-03-29 02:04:00             0             0
## 1173913     <NA> 2009-03-29 02:05:00             0             0
## 1173914     <NA> 2009-03-29 02:06:00             0             0
## 1173915     <NA> 2009-03-29 02:07:00             0             0
## 1173916     <NA> 2009-03-29 02:08:00             0             0
## 1173917     <NA> 2009-03-29 02:09:00             0             0
## 1173918     <NA> 2009-03-29 02:10:00             0             0
## 1173919     <NA> 2009-03-29 02:11:00             0             0
## 1173920     <NA> 2009-03-29 02:12:00             0             0
## 1173921     <NA> 2009-03-29 02:13:00             0             0
## 1173922     <NA> 2009-03-29 02:14:00             0             0
## 1173923     <NA> 2009-03-29 02:15:00             0             0
## 1173924     <NA> 2009-03-29 02:16:00             0             0
## 1173925     <NA> 2009-03-29 02:17:00             0             0
## 1173926     <NA> 2009-03-29 02:18:00             0             0
## 1173927     <NA> 2009-03-29 02:19:00             0             0
## 1173928     <NA> 2009-03-29 02:20:00             0             0
## 1173929     <NA> 2009-03-29 02:21:00             0             0
```

```
## 1173930     <NA> 2009-03-29 02:22:00             0             0
## 1173931     <NA> 2009-03-29 02:23:00             0             0
## 1173932     <NA> 2009-03-29 02:24:00             0             0
## 1173933     <NA> 2009-03-29 02:25:00             0             0
## 1173934     <NA> 2009-03-29 02:26:00             0             0
## 1173935     <NA> 2009-03-29 02:27:00             0             0
## 1173936     <NA> 2009-03-29 02:28:00             0             0
## 1173937     <NA> 2009-03-29 02:29:00             0             0
## 1173938     <NA> 2009-03-29 02:30:00             0             0
## 1173939     <NA> 2009-03-29 02:31:00             0             0
## 1173940     <NA> 2009-03-29 02:32:00             0             0
## 1173941     <NA> 2009-03-29 02:33:00             0             0
## 1173942     <NA> 2009-03-29 02:34:00             0             0
## 1173943     <NA> 2009-03-29 02:35:00             0             0
## 1173944     <NA> 2009-03-29 02:36:00             0             0
## 1173945     <NA> 2009-03-29 02:37:00             0             0
## 1173946     <NA> 2009-03-29 02:38:00             0             0
## 1173947     <NA> 2009-03-29 02:39:00             0             0
## 1173948     <NA> 2009-03-29 02:40:00             0             0
## 1173949     <NA> 2009-03-29 02:41:00             0             0
## 1173950     <NA> 2009-03-29 02:42:00             0             0
## 1173951     <NA> 2009-03-29 02:43:00             0             0
## 1173952     <NA> 2009-03-29 02:44:00             0             0
## 1173953     <NA> 2009-03-29 02:45:00             0             0
## 1173954     <NA> 2009-03-29 02:46:00             0             0
## 1173955     <NA> 2009-03-29 02:47:00             0             0
## 1173956     <NA> 2009-03-29 02:48:00             0             0
## 1173957     <NA> 2009-03-29 02:49:00             0             0
## 1173958     <NA> 2009-03-29 02:50:00             0             0
## 1173959     <NA> 2009-03-29 02:51:00             0             0
## 1173960     <NA> 2009-03-29 02:52:00             0             0
## 1173961     <NA> 2009-03-29 02:53:00             0             0
## 1173962     <NA> 2009-03-29 02:54:00             0             0
## 1173963     <NA> 2009-03-29 02:55:00             0             0
## 1173964     <NA> 2009-03-29 02:56:00             0             0
## 1173965     <NA> 2009-03-29 02:57:00             0             0
## 1173966     <NA> 2009-03-29 02:58:00             0             0
## 1173967     <NA> 2009-03-29 02:59:00             0             0
##         Sub_metering_3 year quarter month week weekdays day hour minute
## 119637               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119638               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119639               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119640               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119641               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119642               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119643               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119644               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119645               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119646               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119647               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119648               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119649               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119650               0   NA      NA    NA   NA     <NA>  NA   NA     NA
## 119651               0   NA      NA    NA   NA     <NA>  NA   NA     NA
```

```
## 119652     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119653     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119654     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119655     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119656     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119657     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119658     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119659     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119660     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119661     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119662     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119663     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119664     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119665     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119666     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119667     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119668     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119669     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119670     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119671     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119672     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119673     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119674     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119675     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119676     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119677     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119678     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119679     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119680     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119681     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119682     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119683     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119684     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119685     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119686     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119687     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119688     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119689     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119690     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119691     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119692     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119693     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119694     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119695     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 119696     0   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649945    17   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649946    17   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649947    18   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649948    17   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649949    18   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649950    18   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649951    17   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649952    18   NA   NA   NA   NA   <NA>   NA   NA   NA
## 649953    18   NA   NA   NA   NA   <NA>   NA   NA   NA
```

```
## 649954              17   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649955              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649956              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649957              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649958              17   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649959              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649960              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649961              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649962              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649963              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649964              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649965              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649966              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649967              17   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649968              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649969              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649970              17   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649971              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649972              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649973              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649974              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649975              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649976              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649977              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649978              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649979              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649980              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649981              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649982              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649983              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649984              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649985              17   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649986              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649987              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649988              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649989              17   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649990              18   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649991              16   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649992               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649993               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649994               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649995               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649996               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649997               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649998               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 649999               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 650000               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 650001               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 650002               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 650003               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 650004               0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173908              0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173909              0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173910              0   NA      NA    NA    NA     <NA>  NA    NA     NA
```

```
## 1173911     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173912     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173913     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173914     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173915     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173916     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173917     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173918     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173919     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173920     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173921     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173922     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173923     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173924     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173925     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173926     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173927     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173928     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173929     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173930     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173931     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173932     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173933     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173934     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173935     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173936     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173937     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173938     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173939     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173940     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173941     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173942     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173943     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173944     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173945     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173946     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173947     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173948     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173949     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173950     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173951     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173952     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173953     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173954     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173955     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173956     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173957     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173958     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173959     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173960     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173961     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173962     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173963     0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173964     0   NA      NA    NA    NA     <NA>  NA    NA     NA
```

```
## 1173965              0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173966              0   NA      NA    NA    NA     <NA>  NA    NA     NA
## 1173967              0   NA      NA    NA    NA     <NA>  NA    NA     NA
```

By checking the times and dates of the NA values, we can see that this is due to a time change. Probably in the last week of March there is a change in time so the time zone of Europe/Paris doesn't recognize this as a valid time/date, therefore, it stores the value as NA until its not an issue. This happens for all three years.

**Summary of values**

- Can you find what is using the most power? the least?

**Sub_metering_3** is what shows the most power consumption of all with about **3 Million Watt-hours per year**, this corresponds to the electric water heater and Air conditioner. It also appears to be the one with the lowest Maximum energy consumption per minute with about 30 watt-hours, by comparison the other two sub meters are more than double.

**Sub_metering_1** is the one with the least power consumption with around **600,000 watt-hours per year** and a maximum of about 77 watt-hours per minute. This corresponds to the kitchen (dishwasher, oven, microwave)

- Anything to learn from the min and max?

The kitchen and laundry room appliances might consume the most power per minute but because they aren't used as often, the overall consumption cost depends mainly on the electric water heater and air conditioning power management.

**High-level Recommendations**

Thermostat set point
Outdoor temps
dedicated submeters for heater and cooling

This energy information coming to and from your home through your smart meter can be run through a home energy management System (EMS), which will allow you to view it in an easy-to-understand format on your computer or hand-held device. A home EMS allows you to track your energy use in detail to better save energy. For instance, you can see the energy impact of various appliances and electronic products simply by monitoring your EMS while switching the devices on and off.

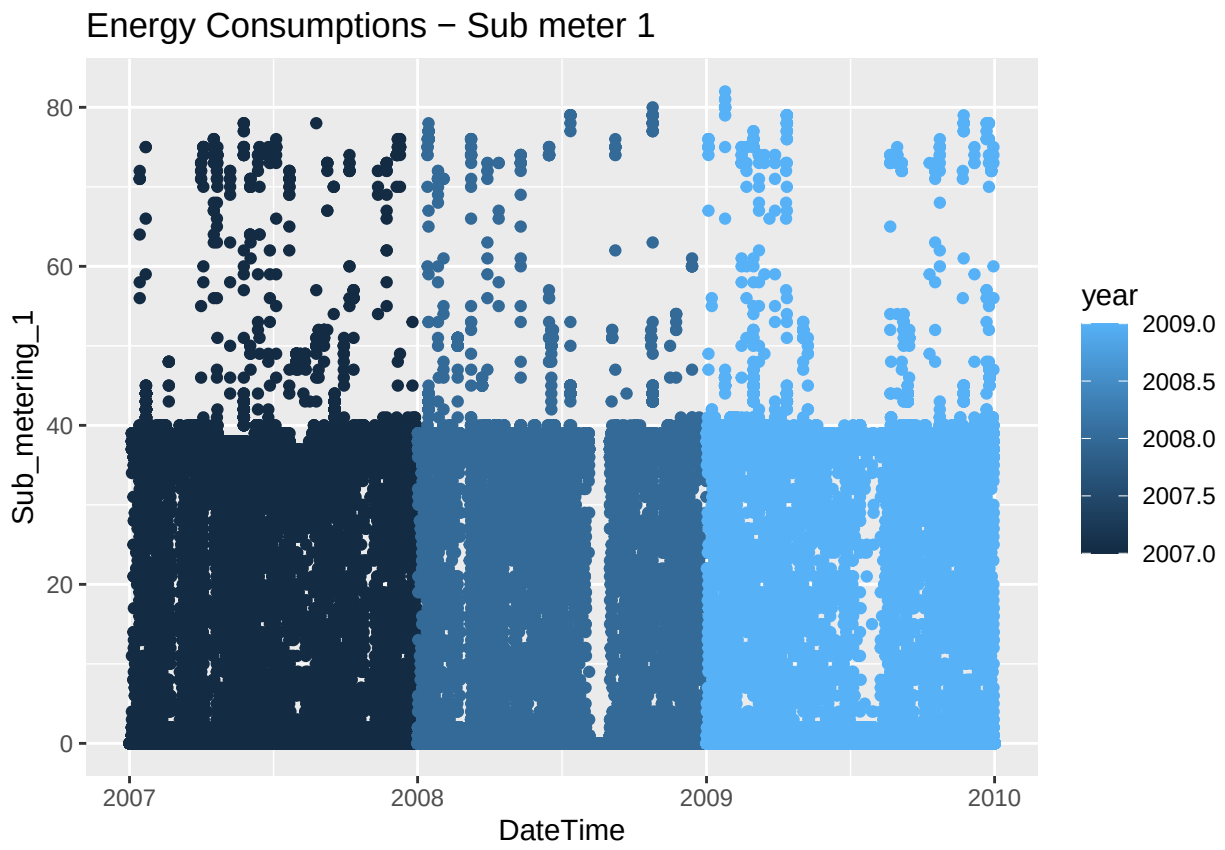
An EMS also allows you to monitor real-time information and price signals from your utility and create settings to automatically use power when prices are lowest. You can also choose settings that allow specific appliances and equipment to turn off automatically when a large demand threatens to cause an outage—avoiding peak demand rates, helping to balance the energy load in your area, and preventing blackouts. Your utility may provide financial incentives for doing so.

## C4T2

**Extra - Data Visualizations (part of task 2)**

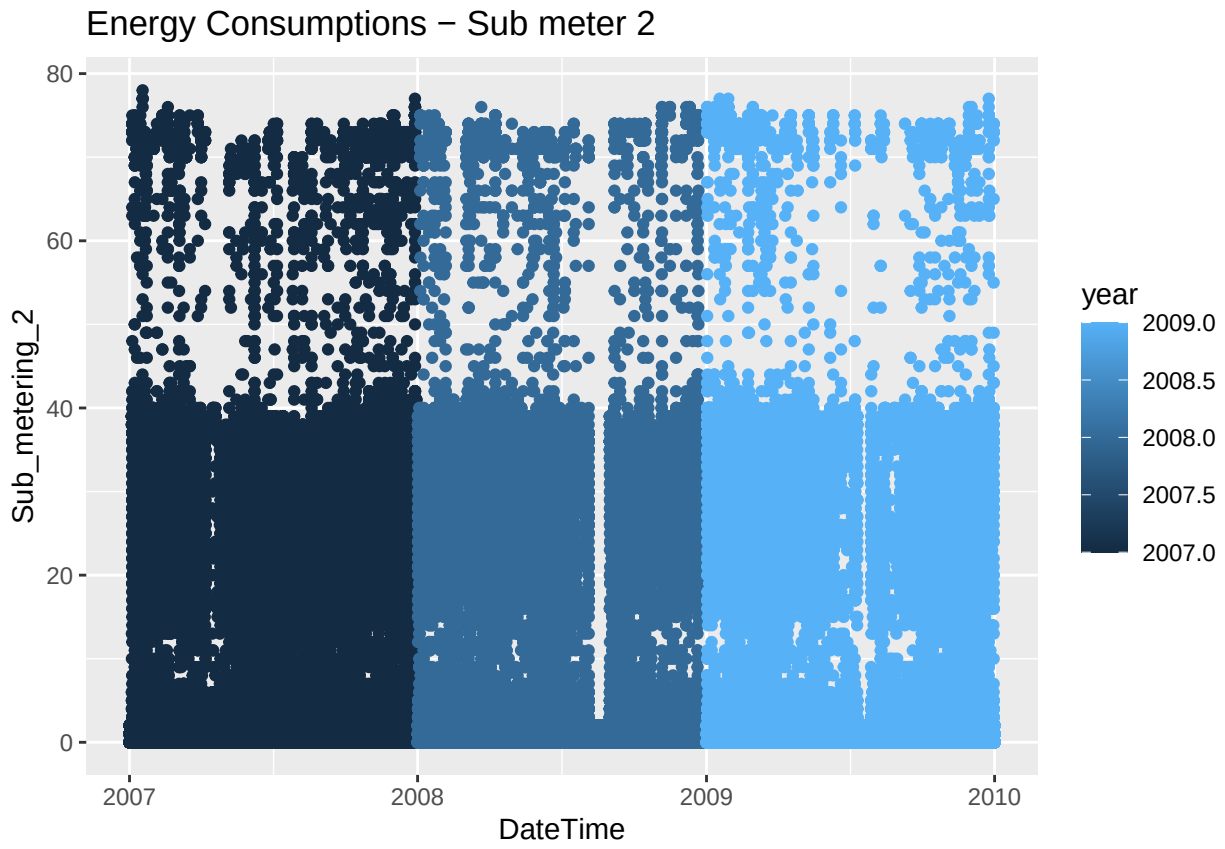
```
ggplot(alldata) +
  geom_point(aes(DateTime, Sub_metering_1, colour=year))+
  labs(title = "Energy Consumptions - Sub meter 1")
```

```
## Warning: Removed 180 rows containing missing values (geom_point).
```



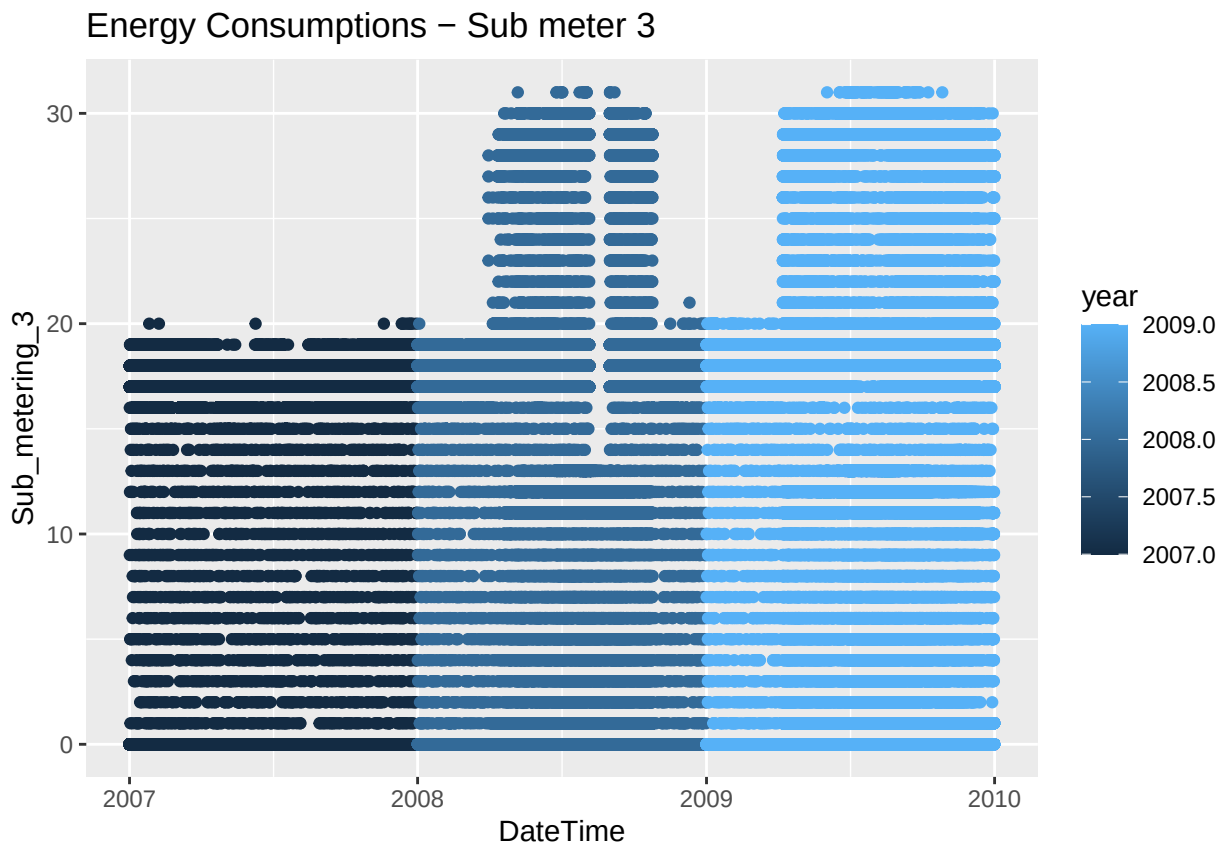

```
ggplot(alldata) +
  geom_point(aes(DateTime, Sub_metering_2, colour=year))+
  labs(title = "Energy Consumptions - Sub meter 2")
```

```
## Warning: Removed 180 rows containing missing values (geom_point).
```

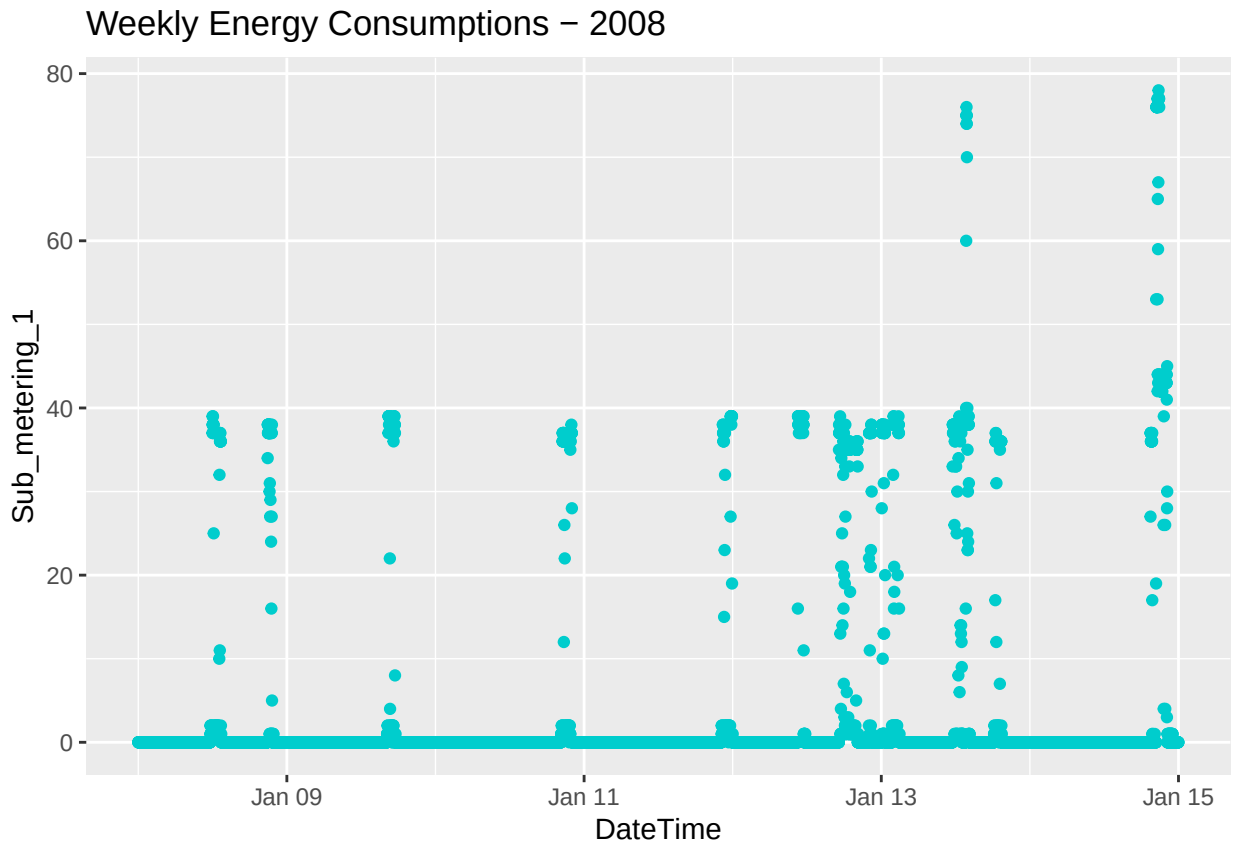



```
ggplot(alldata) +
  geom_point(aes(DateTime, Sub_metering_3, colour=year))+
  labs(title = "Energy Consumptions - Sub meter 3")
```

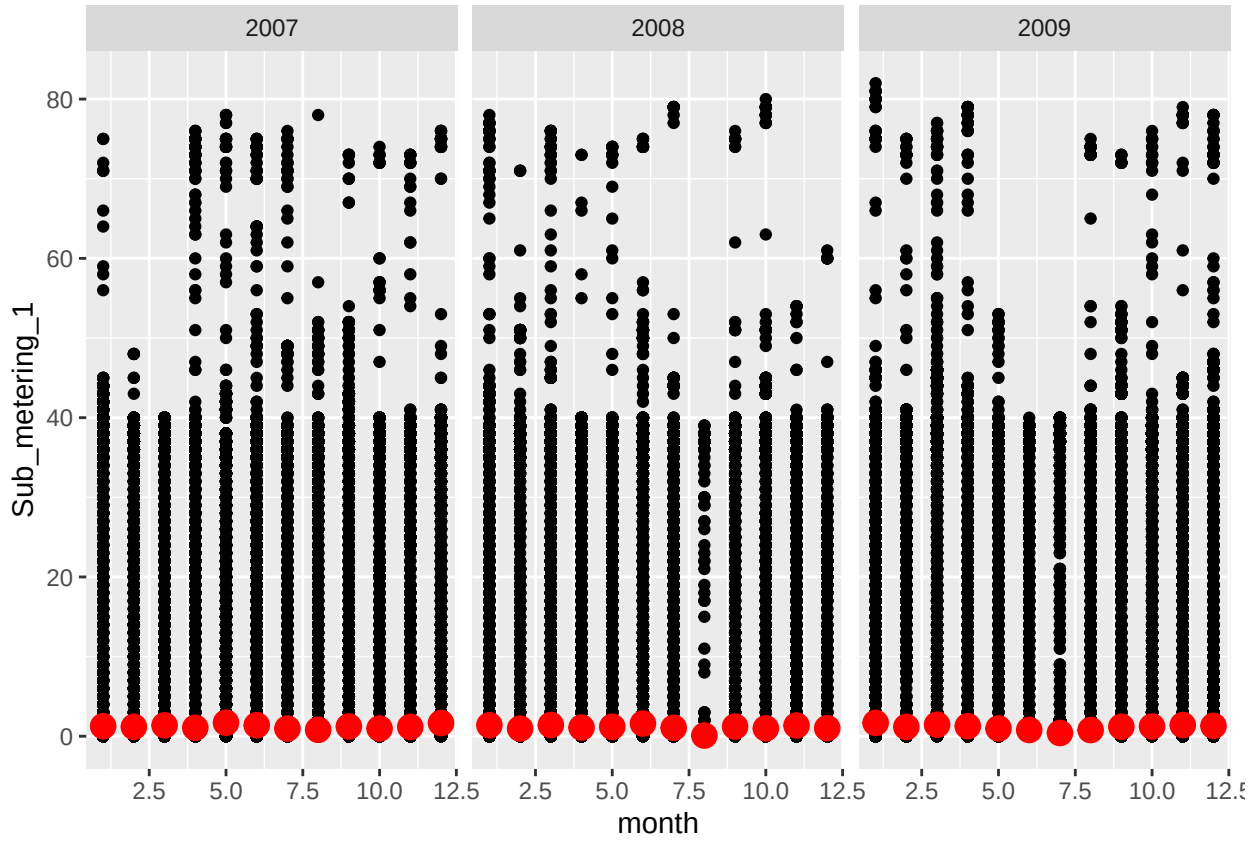
```
## Warning: Removed 180 rows containing missing values (geom_point).
```



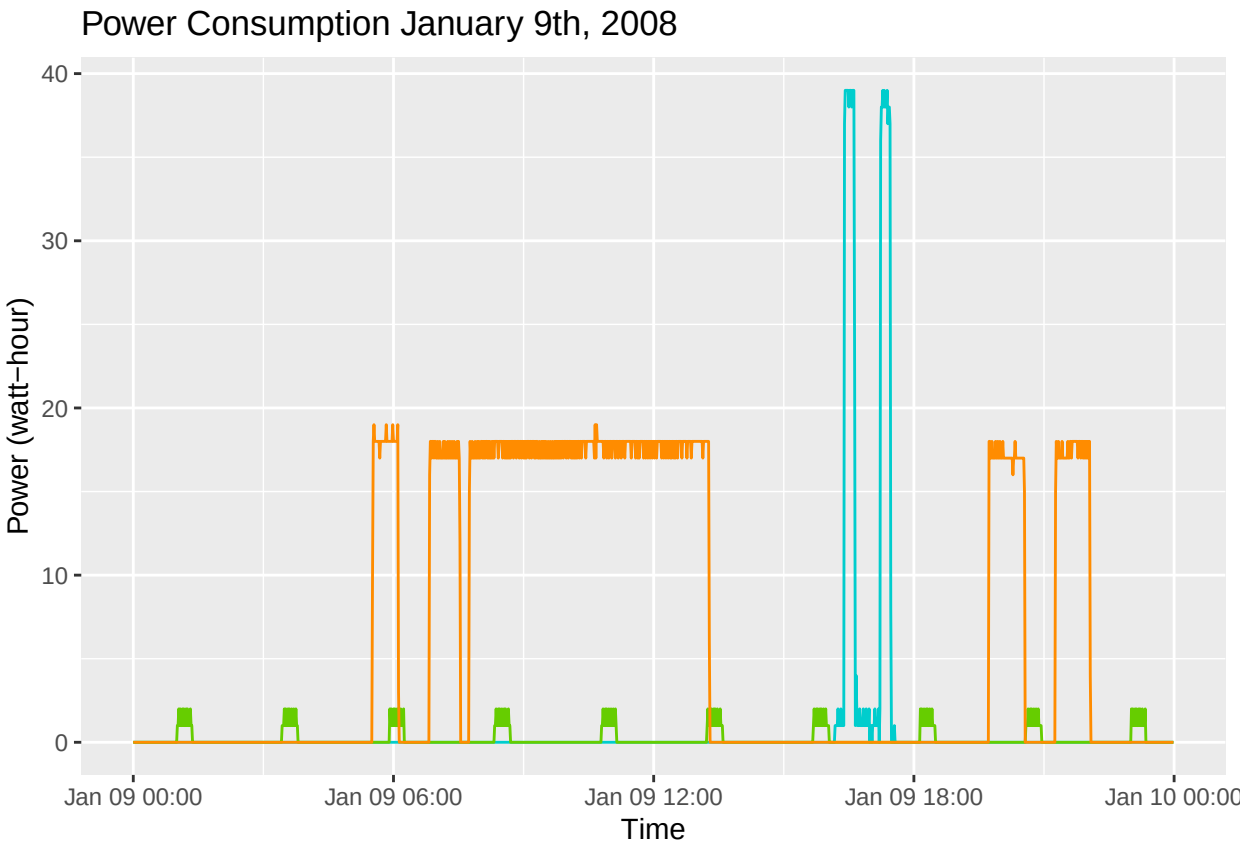

```
alldata_weekly <- filter(alldata, year==2008 & week==2)

ggplot(alldata_weekly) +
  geom_point(aes(DateTime, Sub_metering_1), colour = 'cyan3')+
  labs(title = "Weekly Energy Consumptions - 2008")
```



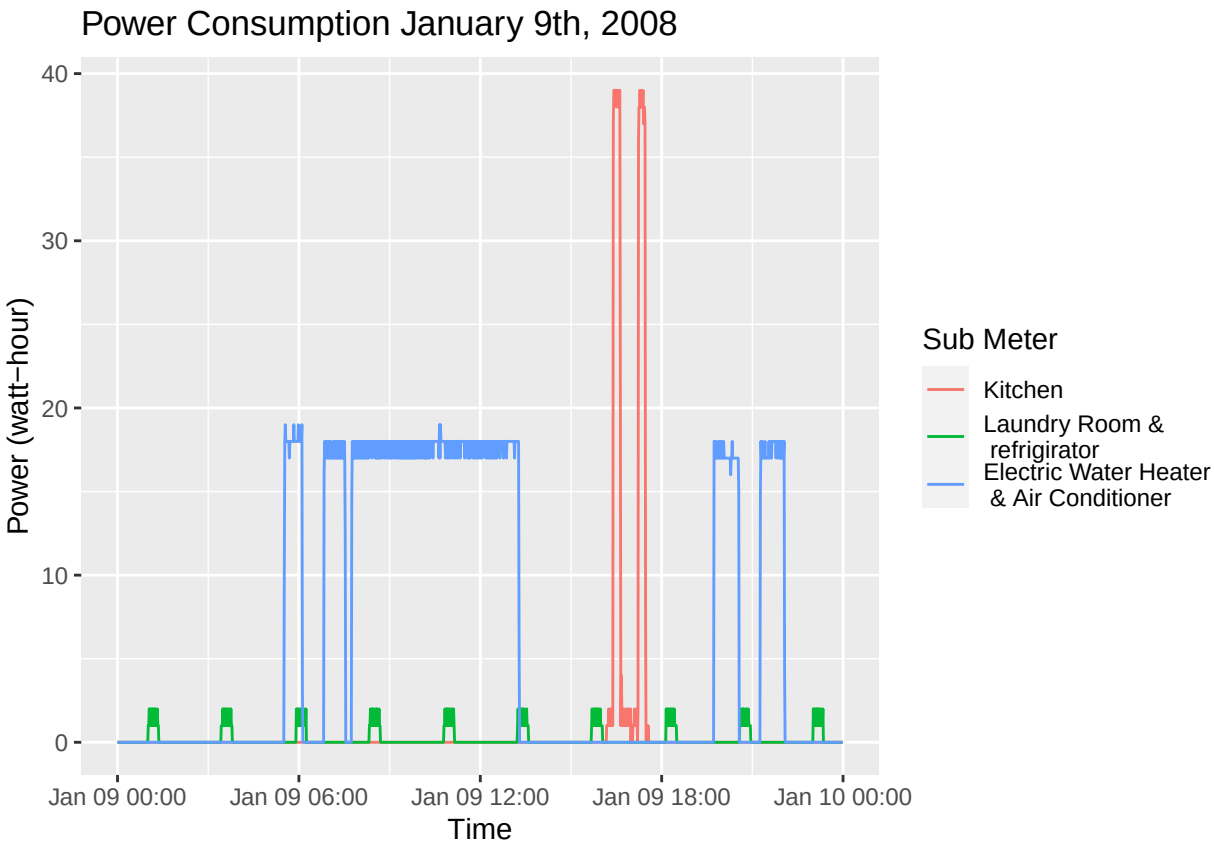

```
alldata1 <- filter(alldata, is.na(alldata)==FALSE)

ggplot(alldata1) +
  geom_point(aes(month, Sub_metering_1, )) +
  stat_summary(aes(month, Sub_metering_1), fun = mean, geom = 'point', colour = 'red', size = 4) +
  facet_grid(~ year) # two variablies
```



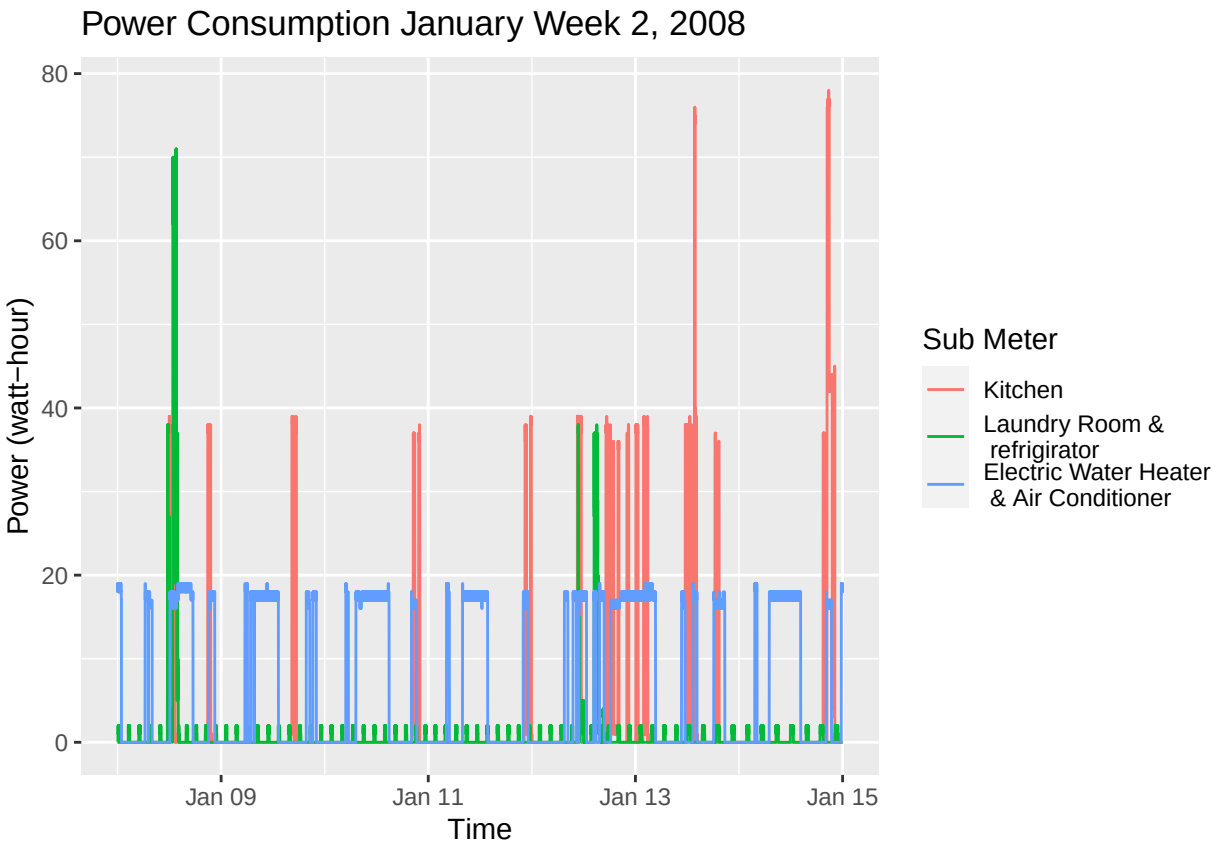

```
alldata_daily <- filter(alldata, year==2008 & month==1 & day==9)

ggplot(alldata_daily) +
  geom_line(aes(DateTime, Sub_metering_1), colour = "cyan3") +
  geom_line(aes(DateTime, Sub_metering_2), colour = "chartreuse3") +
  geom_line(aes(DateTime, Sub_metering_3), colour = "darkorange") +
  labs(title = "Power Consumption January 9th, 2008",
       x = "Time",
       y = "Power (watt-hour)")
```



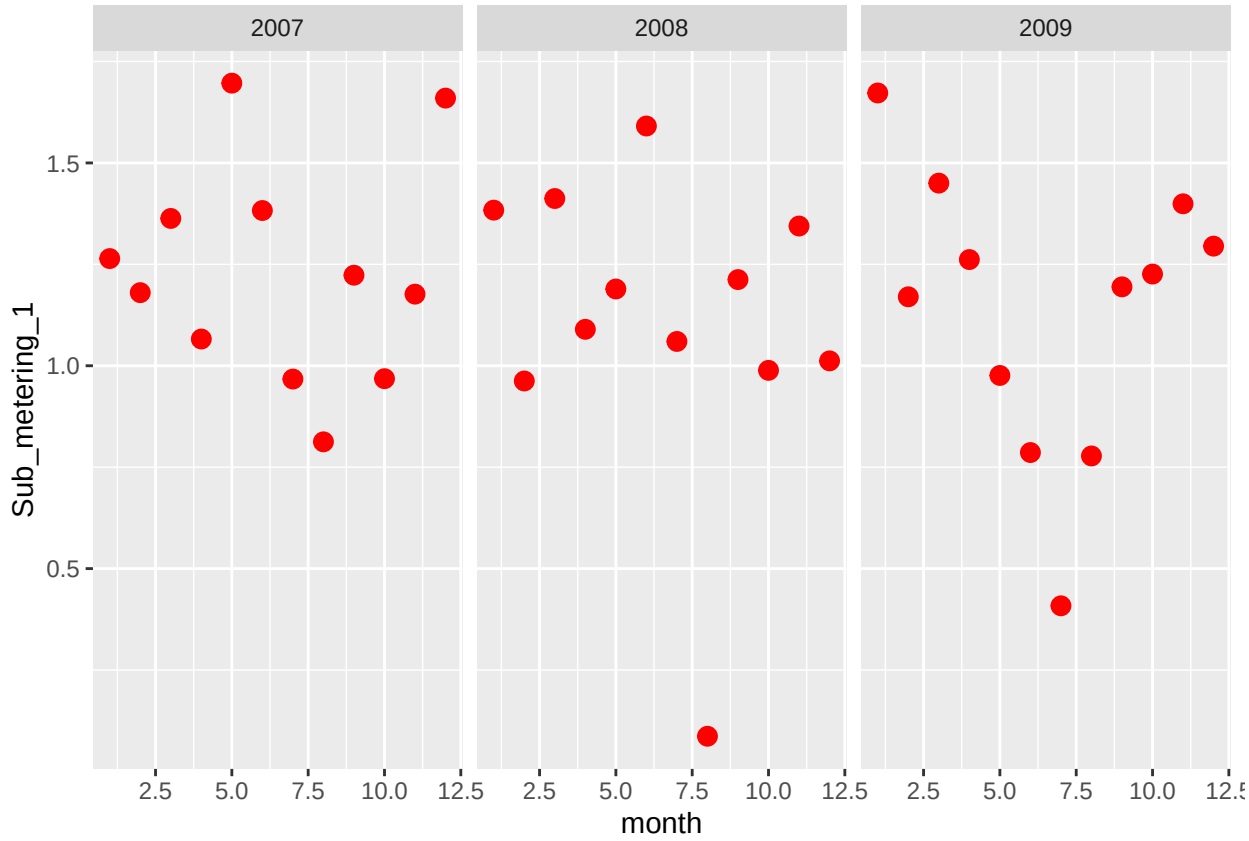

```
#This ^ had no legend and it looked lengthy to put one so i tried a different method:
#Now I used gather to group everything in a way ggplot can understand and plot it.

dfdaily <- alldata_daily %>% gather(key = Meter, value = Value, Sub_metering_1:Sub_metering_3)

ggplot(dfdaily) +
  geom_line(aes(DateTime, Value, colour=Meter)) +
  scale_color_discrete(name = "Sub Meter", labels = c("Kitchen", "Laundry Room &\n refrigirator", "Elec
  labs(title = "Power Consumption January 9th, 2008",
       x = "Time",
       y = "Power (watt-hour)")
```



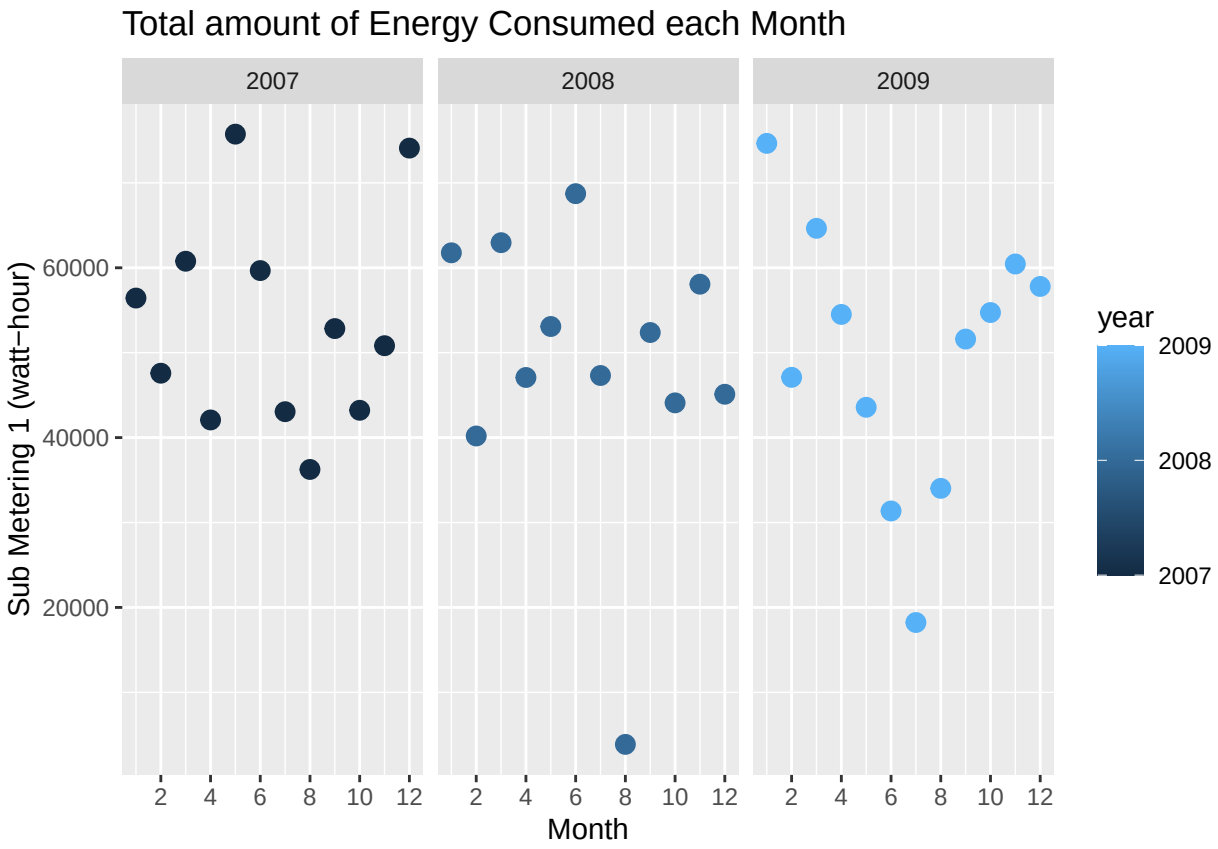

```
alldata_weekly <- filter(alldata, year==2008 & month==1 & week==2)

dfweek <- alldata_weekly %>% gather(key = Meter, value = Value, Sub_metering_1:Sub_metering_3)

ggplot(dfweek) +
  geom_line(aes(DateTime, Value, colour=Meter)) +
  scale_color_discrete(name = "Sub Meter", labels = c("Kitchen", "Laundry Room &\n refrigirator", "Elec
  labs(title = "Power Consumption January Week 2, 2008",
       x = "Time",
       y = "Power (watt-hour)")
```



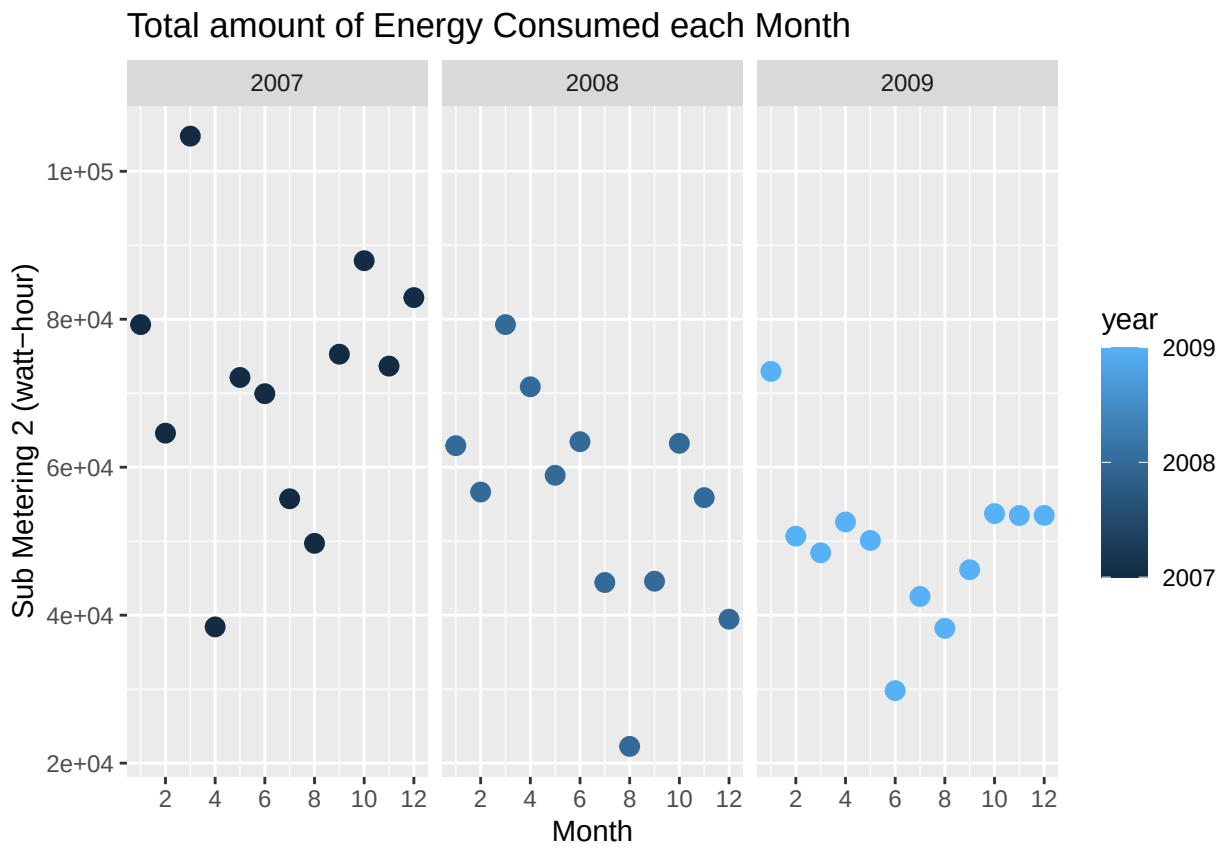

```
ggplot(alldata1) +
  #geom_point(aes(month, Sub_metering_1, )) +
  stat_summary(aes(month, Sub_metering_1), fun = mean, geom = 'point', colour = 'red', size = 3) +
  facet_grid(~ year) # two variablies
```



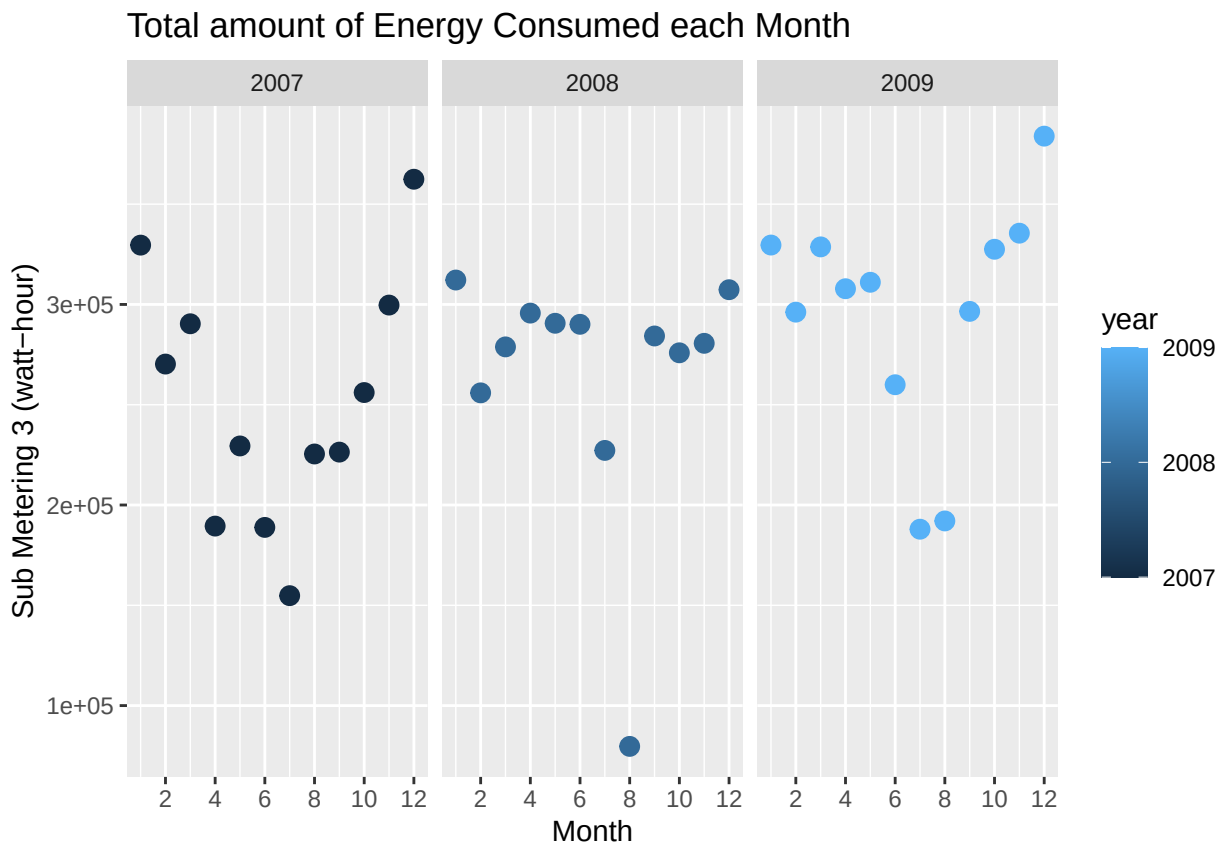

```
ggplot(alldata1) +
  #geom_point(aes(month, Sub_metering_1, )) +
  stat_summary(aes(month, Sub_metering_1, colour = year), fun = sum, geom = 'point', size = 3) +
  facet_grid(~year) +
  scale_color_continuous(breaks = c(2007, 2008, 2009))+
  scale_x_continuous(breaks=c(2,4,6,8,10,12))+
  labs(title = "Total amount of Energy Consumed each Month",
       x = "Month",
       y = "Sub Metering 1 (watt-hour)")
```



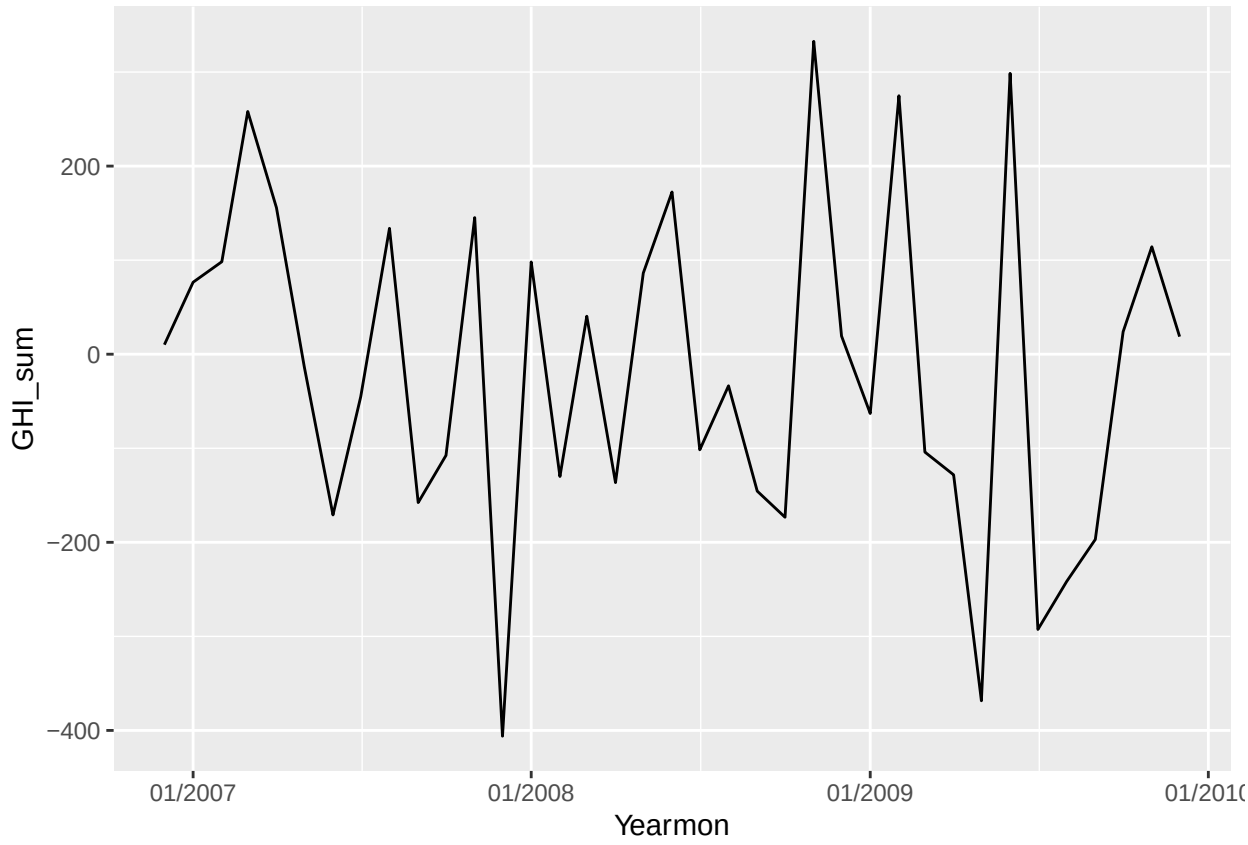

```
ggplot(alldata1) +
  #geom_point(aes(month, Sub_metering_1, )) +
  stat_summary(aes(month, Sub_metering_2, colour = year), fun = sum, geom = 'point', size = 3) +
  facet_grid(~year) +
  scale_color_continuous(breaks = c(2007, 2008, 2009))+
  scale_x_continuous(breaks=c(2,4,6,8,10,12))+
  labs(title = "Total amount of Energy Consumed each Month",
       x = "Month",
       y = "Sub Metering 2 (watt-hour)")
```




```
ggplot(alldata1) +
  #geom_point(aes(month, Sub_metering_1, )) +
  stat_summary(aes(month, Sub_metering_3, colour = year), fun = sum, geom = 'point', size = 3) +
  facet_grid(~year) +
  scale_color_continuous(breaks = c(2007, 2008, 2009))+
  scale_x_continuous(breaks=c(2,4,6,8,10,12))+
  labs(title = "Total amount of Energy Consumed each Month",
       x = "Month",
       y = "Sub Metering 3 (watt-hour)")
```



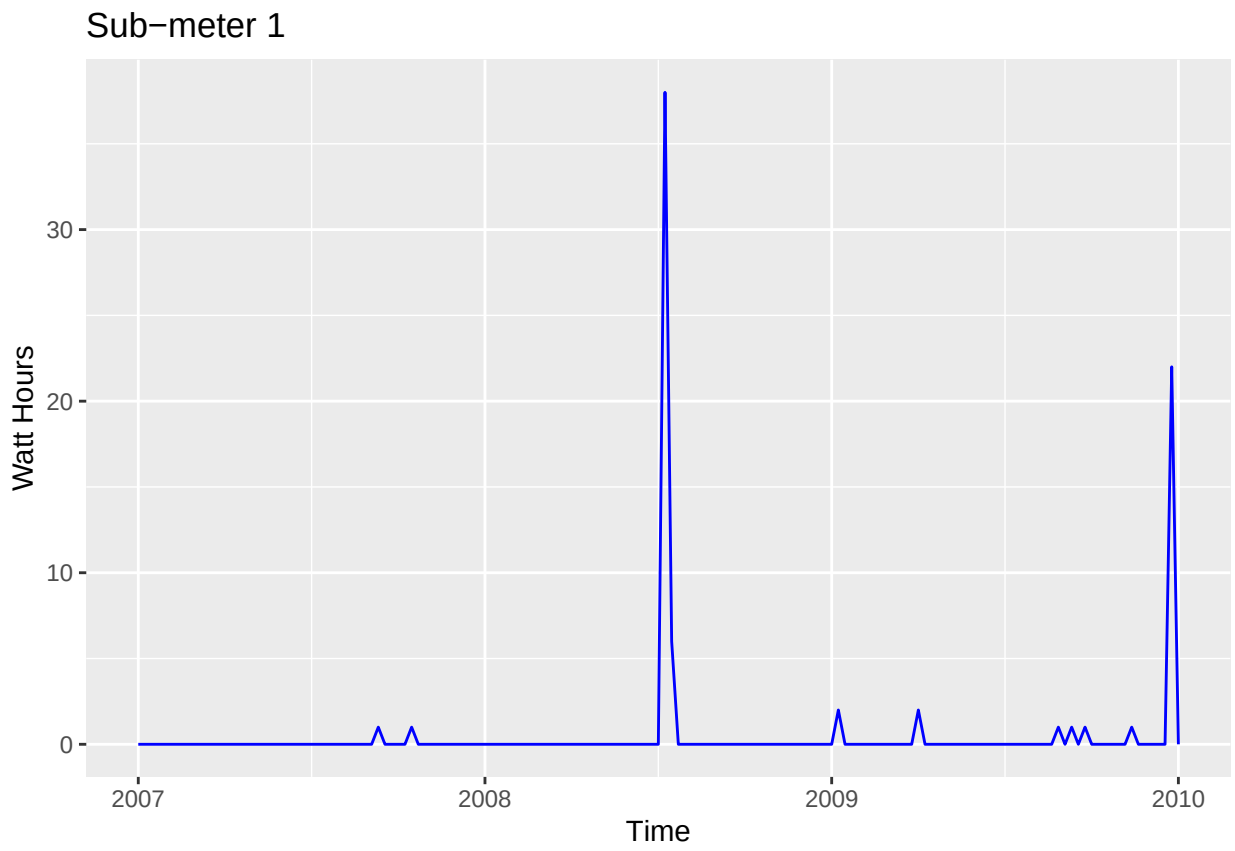

```
dummy <- data.frame(Time = alldata1$DateTime,
                    GHI = rnorm(length(alldata1$DateTime)))

df <- mutate(dummy,
             Yearmon = as.yearmon(alldata1$DateTime)) %>%
  group_by(Yearmon) %>%
  summarise(GHI_sum = sum(GHI)) %>%
  ungroup() %>%
  mutate(Yearmon = as.Date(Yearmon))

ggplot(df, aes(Yearmon, GHI_sum)) +
  geom_line()+
  scale_x_date(labels= date_format("%m/%Y"))
```



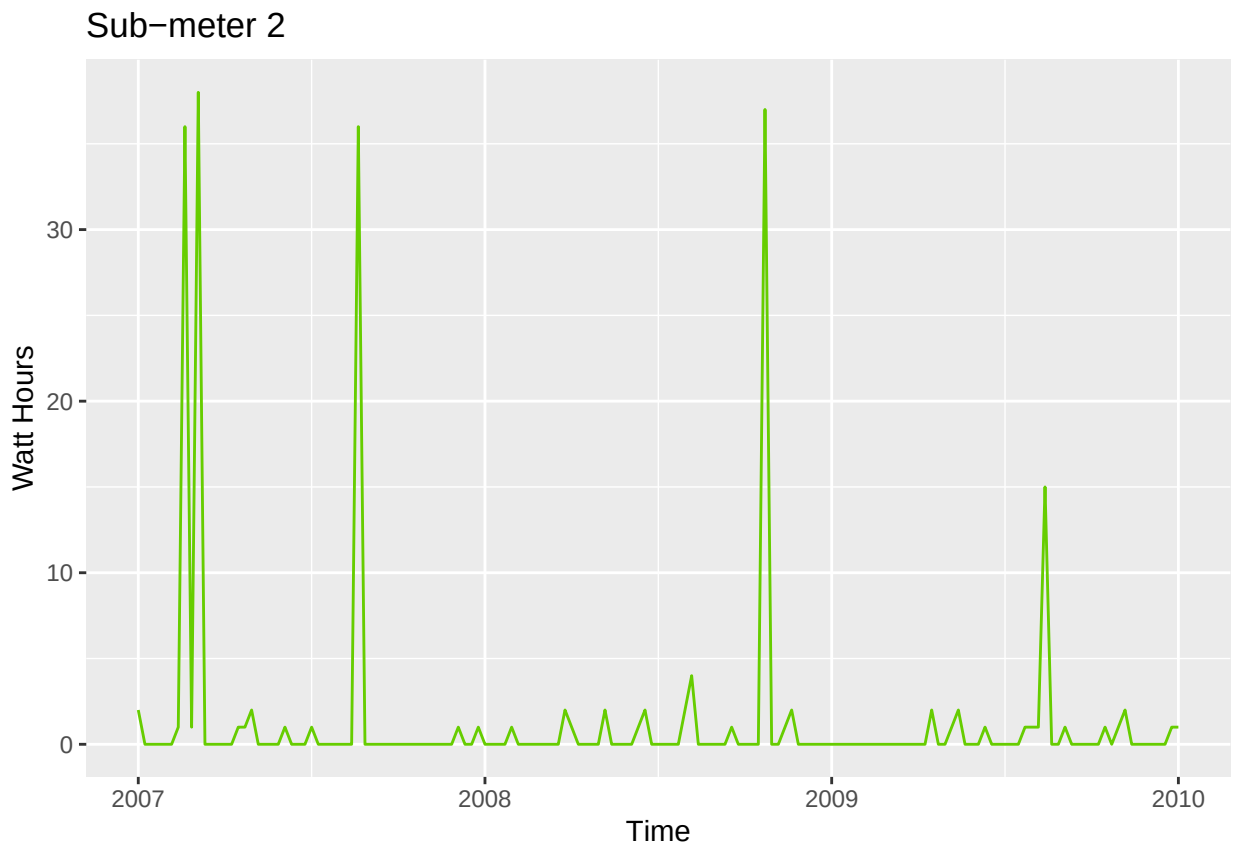

```
#ggplot(alldata1) +
#  geom_point(aes(month, Sub_metering_1, group=month)) +
#  facet_grid(~ year)
```

**Time Series Analysis**

```
House070809weekly <- filter(alldata, weekdays == 'Monday' & hour==20 & minute==1)

#create time series object

tsSM1_070809weekly <- ts(House070809weekly$Sub_metering_1, frequency=52, start=c(2007,1))
tsSM2_070809weekly <- ts(House070809weekly$Sub_metering_2, frequency=52, start=c(2007,1))
tsSM3_070809weekly <- ts(House070809weekly$Sub_metering_3, frequency=52, start=c(2007,1))

#Plot sub-meter 3 with autoplot

#autoplot(tsSM3_070809weekly)

#plot.ts(tsSM3_070809weekly)

autoplot(tsSM1_070809weekly,colour = 'blue', xlab = "Time", ylab = "Watt Hours", main = "Sub-meter 1")
```




```
autoplot(tsSM2_070809weekly, colour = 'chartreuse3', xlab = "Time", ylab = "Watt Hours", main = "Sub-me
```




```
autoplot(tsSM3_070809weekly, colour = 'red', xlab = "Time", ylab = "Watt Hours", main = "Sub-meter 3")
```


```
?autoplot
```

```
## starting httpd help server ... done
```

**Forcast**

```
#Apply time series linear regression to the sub-meter 3 ts object and use summary to obtain R2 and RMSE

fitSM1 <- tslm(tsSM1_070809weekly ~ trend + season)
#summary(fitSM1)

fitSM2 <- tslm(tsSM2_070809weekly ~ trend + season)
#summary(fitSM2)

fitSM3 <- tslm(tsSM3_070809weekly ~ trend + season)
#summary(fitSM3)

accuracy(fitSM1)
```

```
##                         ME     RMSE       MAE MPE MAPE      MASE      ACF1
## Training set -8.099957e-17 2.889331 0.7636169 NaN  Inf 0.6681648 0.1102239
```

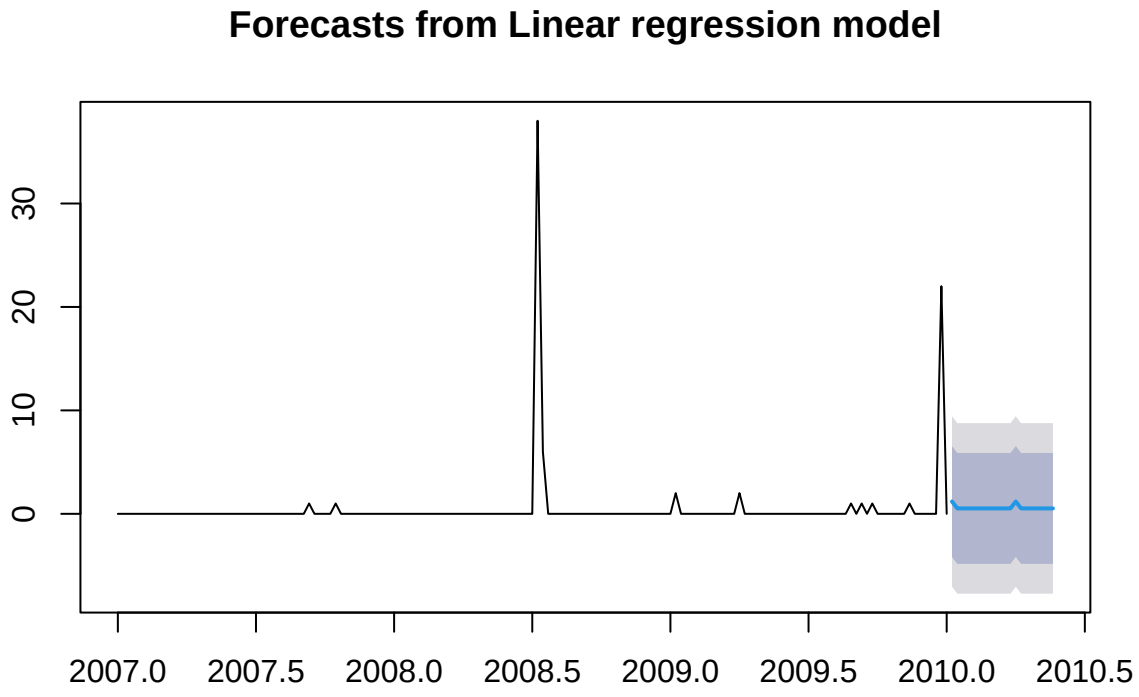
```
accuracy(fitSM2)
```

```
##                         ME     RMSE      MAE MPE MAPE      MASE        ACF1
## Training set 5.159974e-18 4.867134 2.029347 NaN  Inf 0.8291106 -0.05728754
```

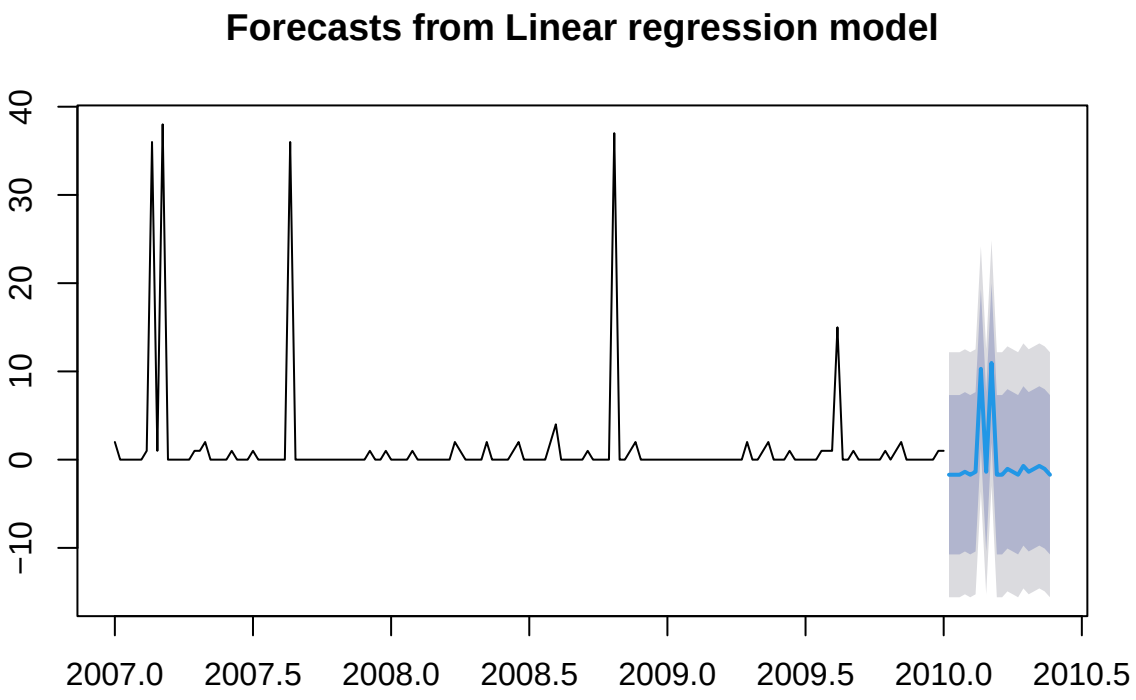
```
accuracy(fitSM3)
```

```
##                          ME     RMSE      MAE MPE MAPE      MASE      ACF1
## Training set -1.176131e-16 7.362187 6.073596 NaN  Inf 0.6295435 0.1033361
```

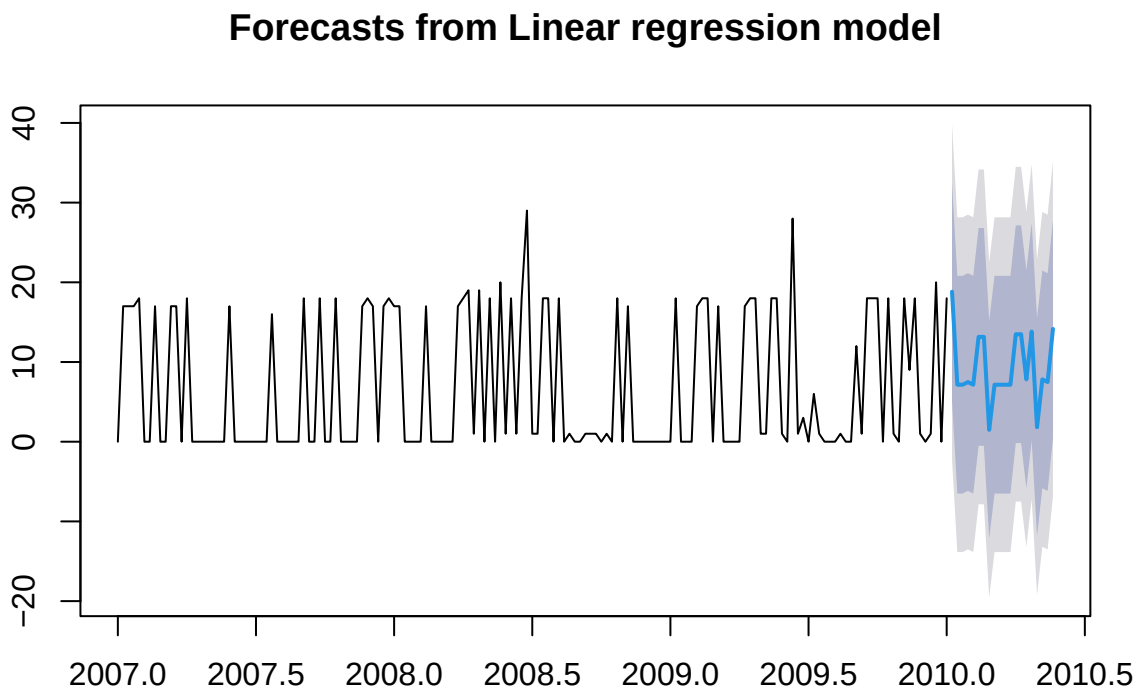
```
# Create the forecast for sub-meter 3, focasting ahead 20 time periods

forecastfitSM1 <- forecast(fitSM1, h=20)
forecastfitSM2 <- forecast(fitSM2, h=20)
forecastfitSM3 <- forecast(fitSM3, h=20)

# Plot the forecast for sub-meter 3

plot(forecastfitSM1)
```



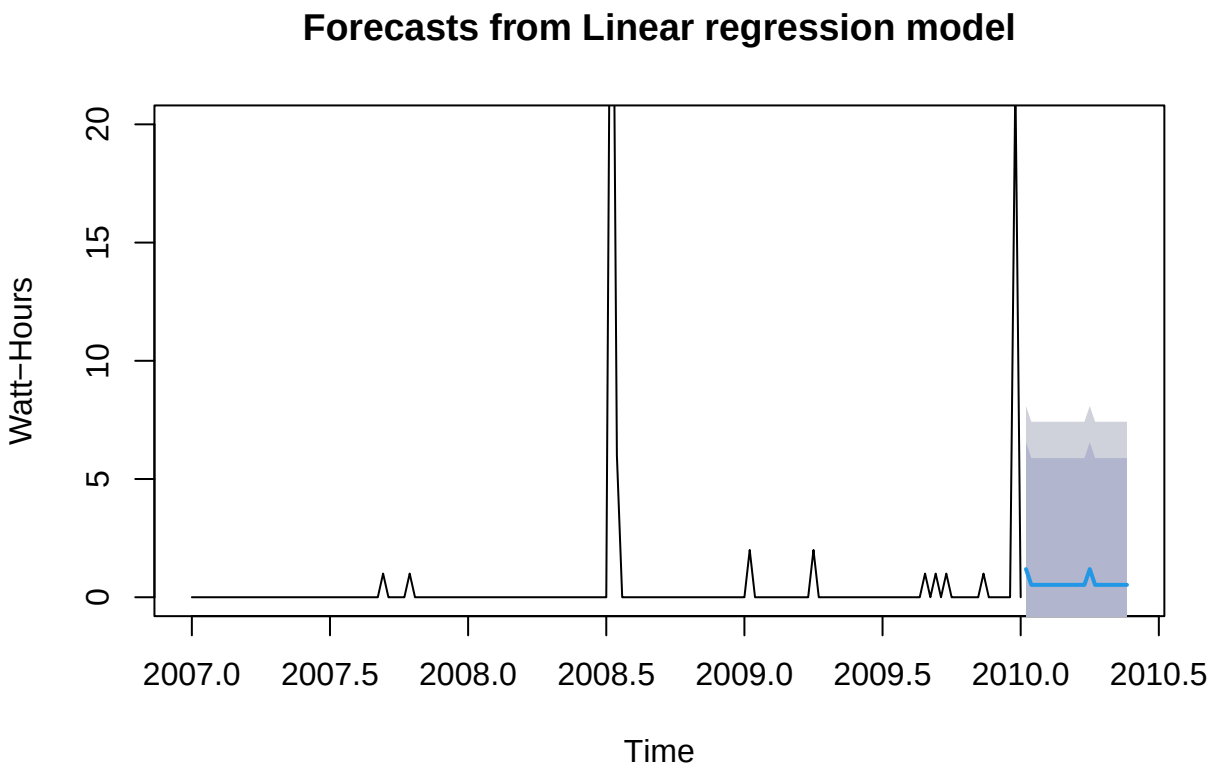

```
plot(forecastfitSM2)
```



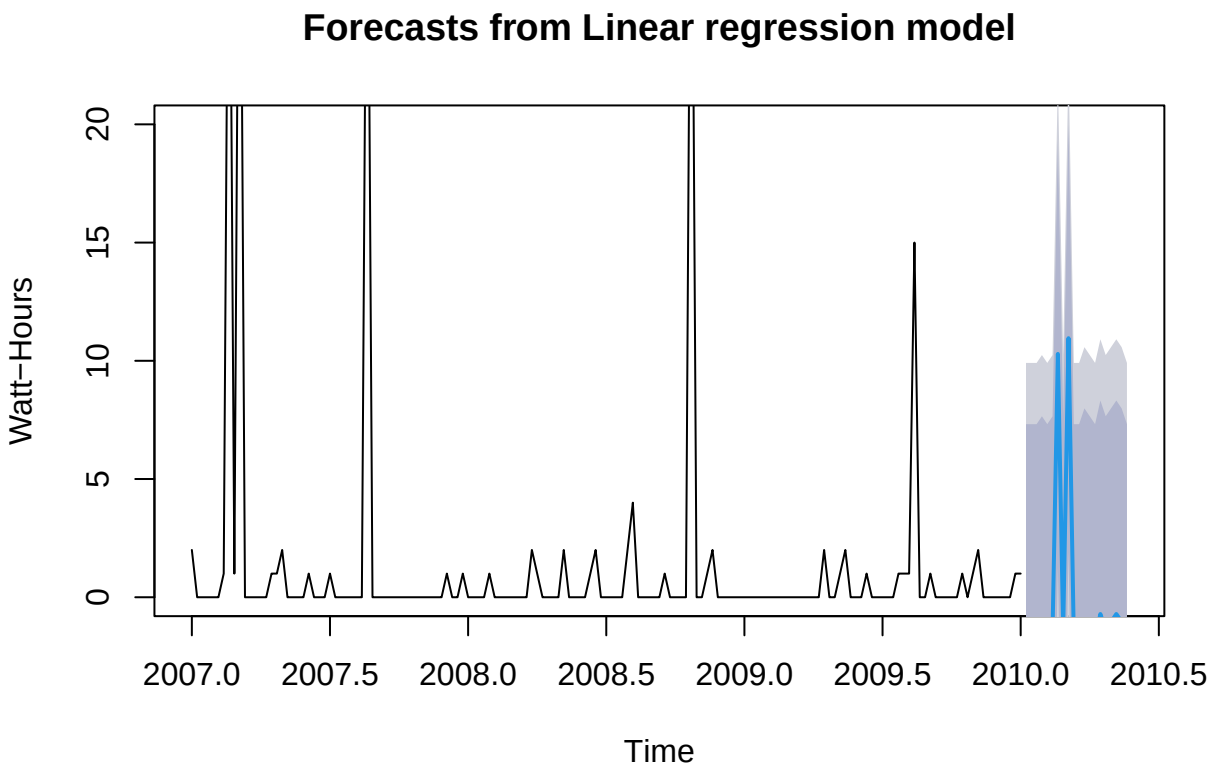

```
plot(forecastfitSM3)
```



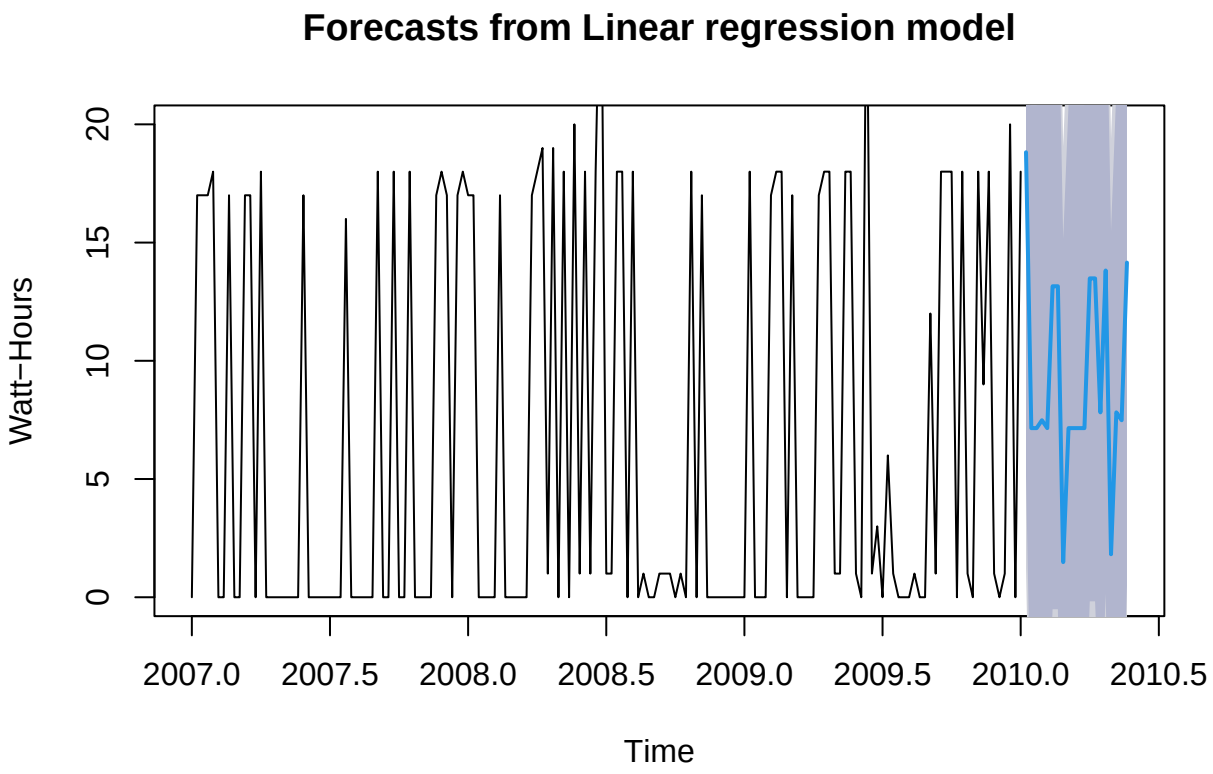

```
## Create sub-meter 3 forecast with confidence levels 80 and 90

forecastfitSM1c <- forecast(fitSM1, h=20, level=c(80,90))
forecastfitSM2c <- forecast(fitSM2, h=20, level=c(80,90))
forecastfitSM3c <- forecast(fitSM3, h=20, level=c(80,90))

## Plot sub-meter 3 forecast, limit y and add labels

plot(forecastfitSM1c, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time")
```




```
plot(forecastfitSM2c, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time")
```

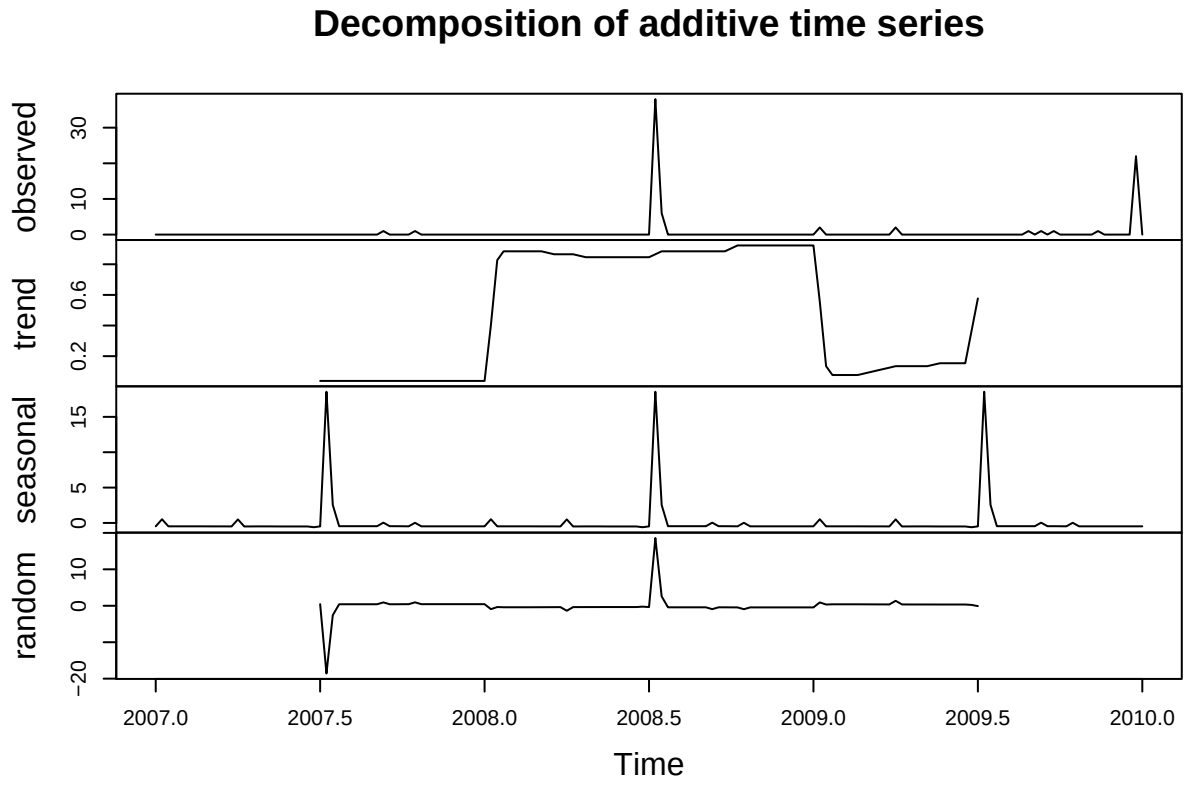



```
plot(forecastfitSM3c, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time")
```

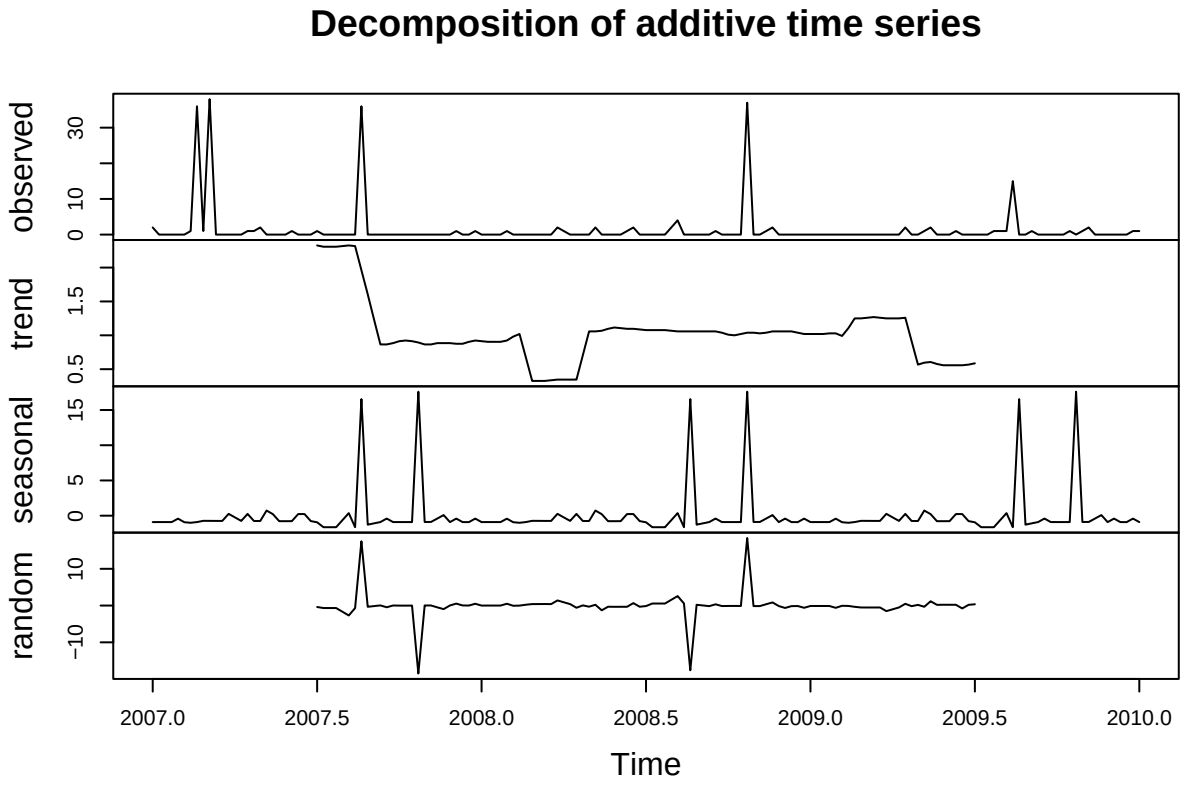



### Adjusting for Seasonality

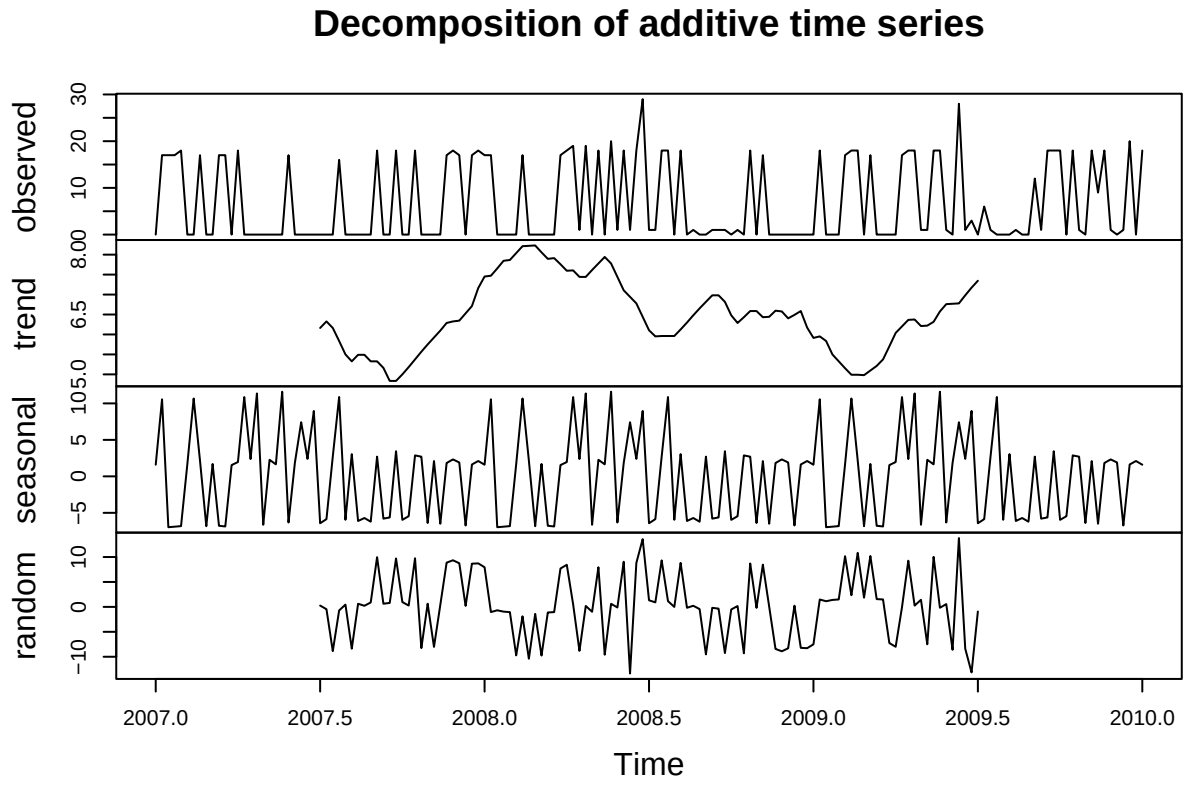
```
## Decompose Sub-meter 3 into trend, seasonal and remainder

components070809SM1weekly <- decompose(tsSM1_070809weekly)
components070809SM2weekly <- decompose(tsSM2_070809weekly)
components070809SM3weekly <- decompose(tsSM3_070809weekly)

## Plot decomposed sub-meter 3

plot(components070809SM1weekly)
```




```
plot(components070809SM2weekly)
```




```
plot(components070809SM3weekly)
```




```
## Check summary statistics for decomposed sub-meter 3

summary(components070809SM1weekly)
```

```
##          Length Class  Mode
## x        157    ts     numeric
## seasonal 157    ts     numeric
## trend    157    ts     numeric
## random   157    ts     numeric
## figure    52    -none- numeric
## type       1    -none- character
```

```
summary(components070809SM2weekly)
```

```
##          Length Class  Mode
## x        157    ts     numeric
## seasonal 157    ts     numeric
## trend    157    ts     numeric
## random   157    ts     numeric
## figure    52    -none- numeric
## type       1    -none- character
```

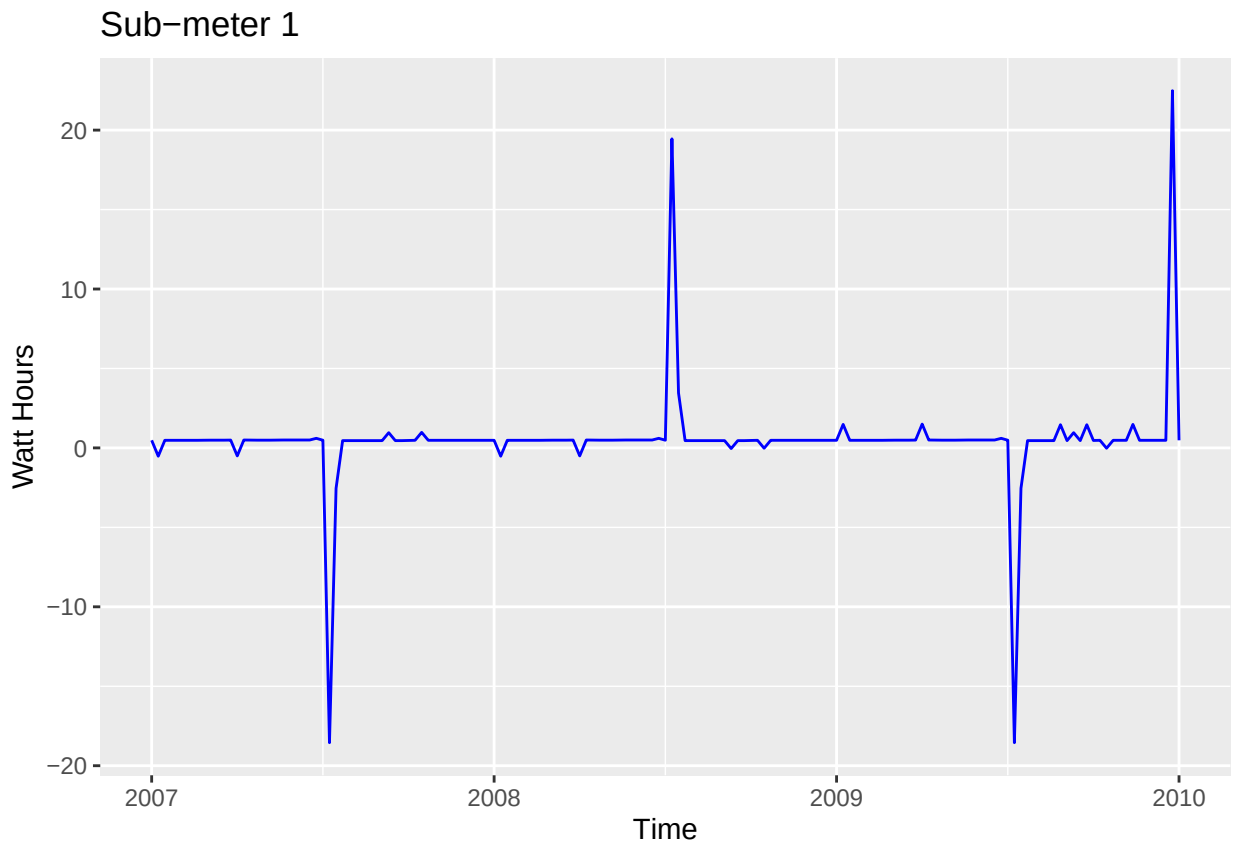
```
summary(components070809SM3weekly)
```

```
##          Length Class  Mode
## x        157    ts     numeric
## seasonal 157    ts     numeric
## trend    157    ts     numeric
## random   157    ts     numeric
## figure    52    -none- numeric
## type       1    -none- character
```

Does sub-meter 3 show a trend in power usage?
Would this information be important to a homeowner trying to understand their power consumption?
Does sub-meter 3 show seasonal effects on power usage?
What may or may not cause this?

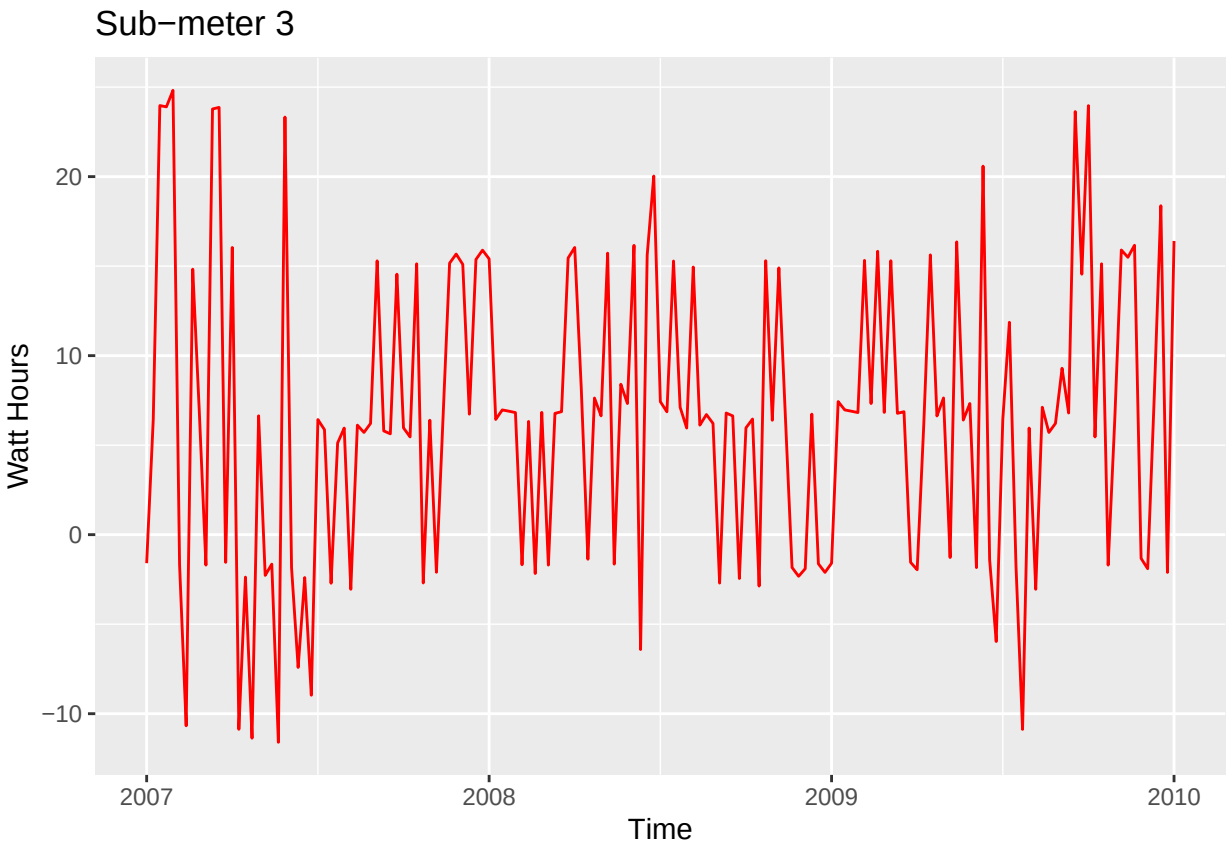
### Holt-Winters Forecasting

```
## Seasonal adjusting sub-meter 3 by subtracting the seasonal component & plot
tsSM1_070809Adjusted <- tsSM1_070809weekly - components070809SM1weekly$seasonal
tsSM2_070809Adjusted <- tsSM2_070809weekly - components070809SM2weekly$seasonal
tsSM3_070809Adjusted <- tsSM3_070809weekly - components070809SM3weekly$seasonal


autoplot(tsSM1_070809Adjusted, colour = 'blue', xlab = "Time", ylab = "Watt Hours", main = "Sub-meter 1
```

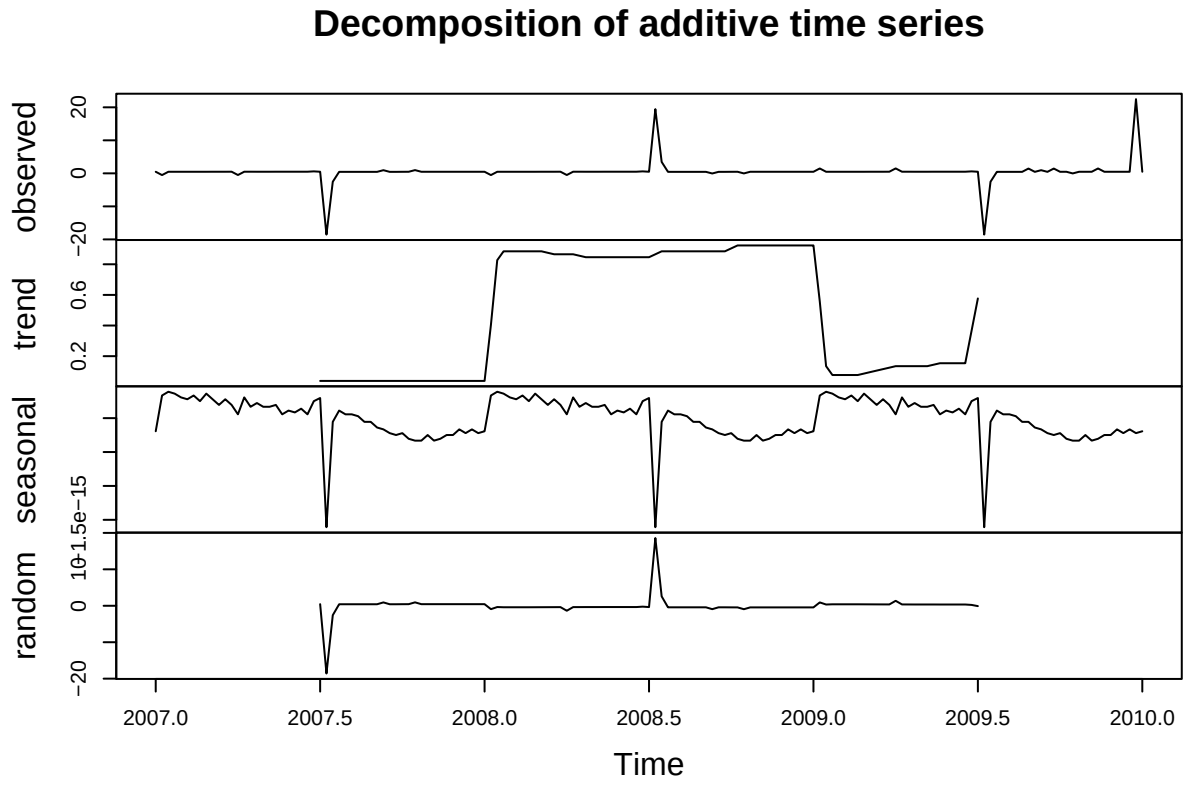


```
autoplot(tsSM2_070809Adjusted, colour = 'chartreuse3', xlab = "Time", ylab = "Watt Hours", main = "Sub-m
```

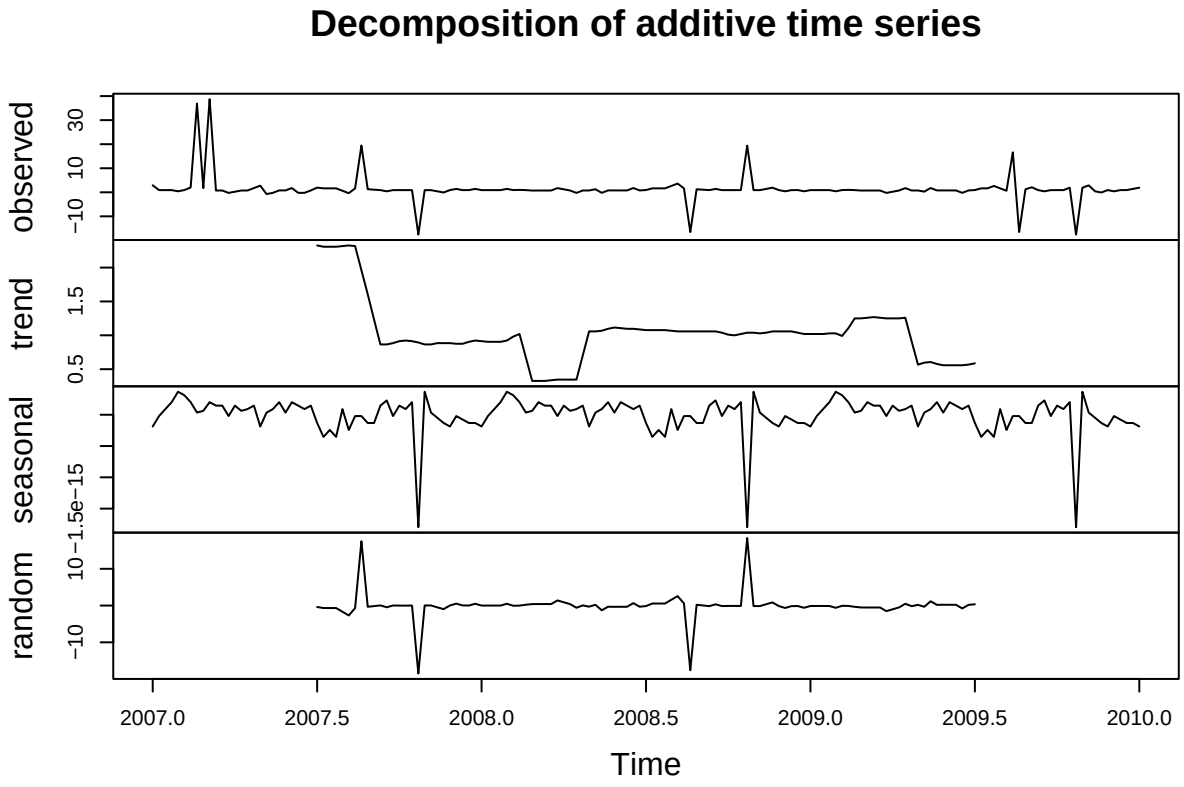

```
autoplot(tsSM3_070809Adjusted, colour = 'red', xlab = "Time", ylab = "Watt Hours", main = "Sub-meter 3"
```



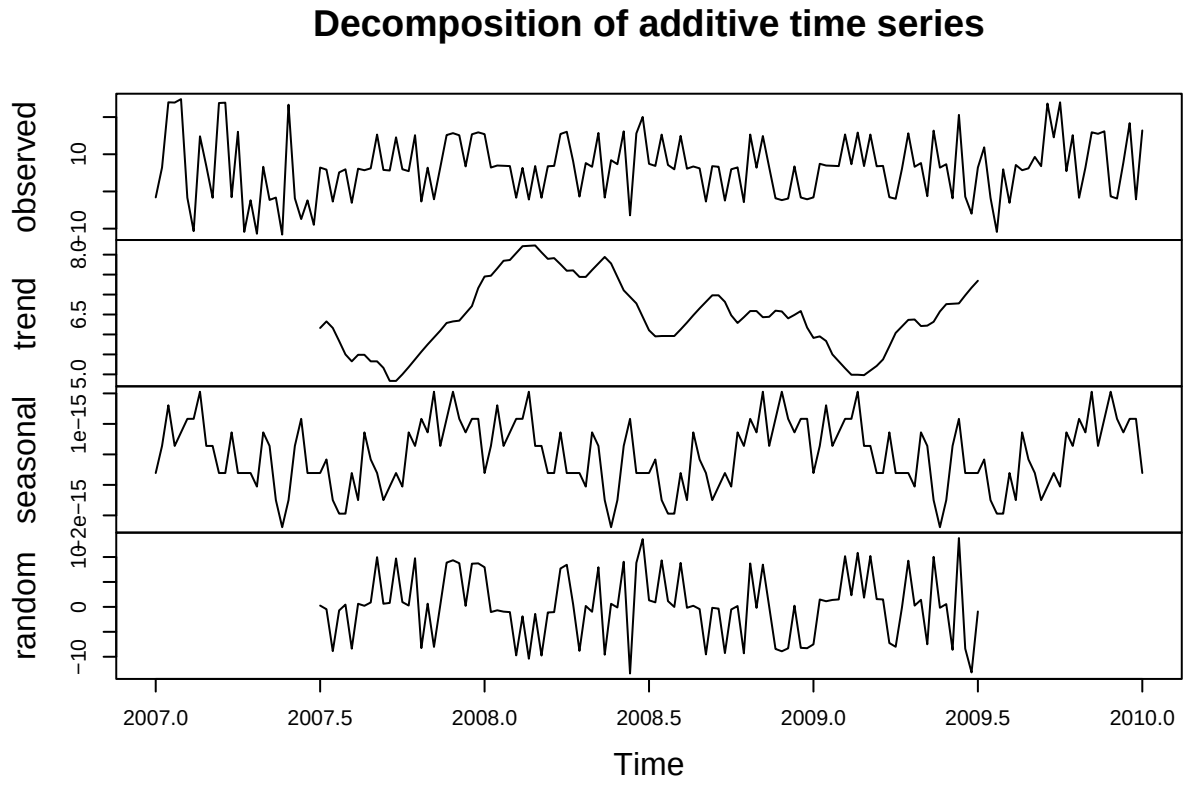

```
## Test Seasonal Adjustment by running Decompose again. Note the very, very small scale for Seasonal
plot(decompose(tsSM1_070809Adjusted))
```



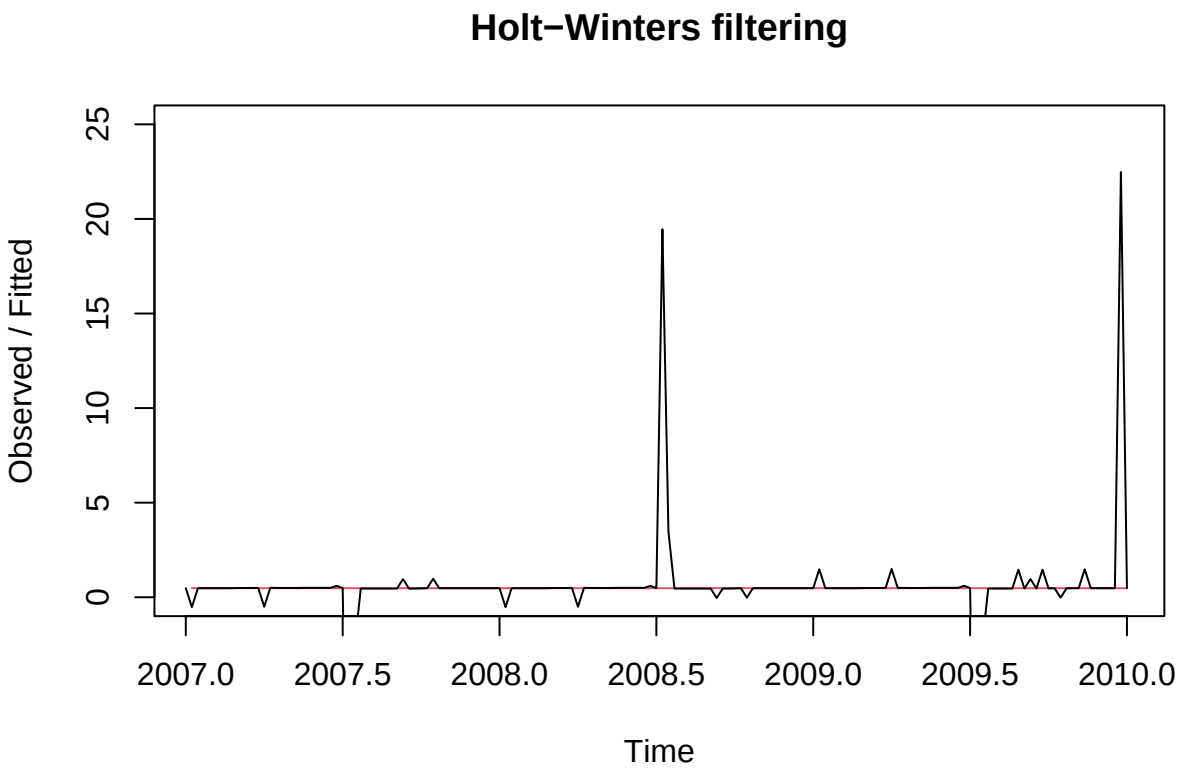

```
plot(decompose(tsSM2_070809Adjusted))
```



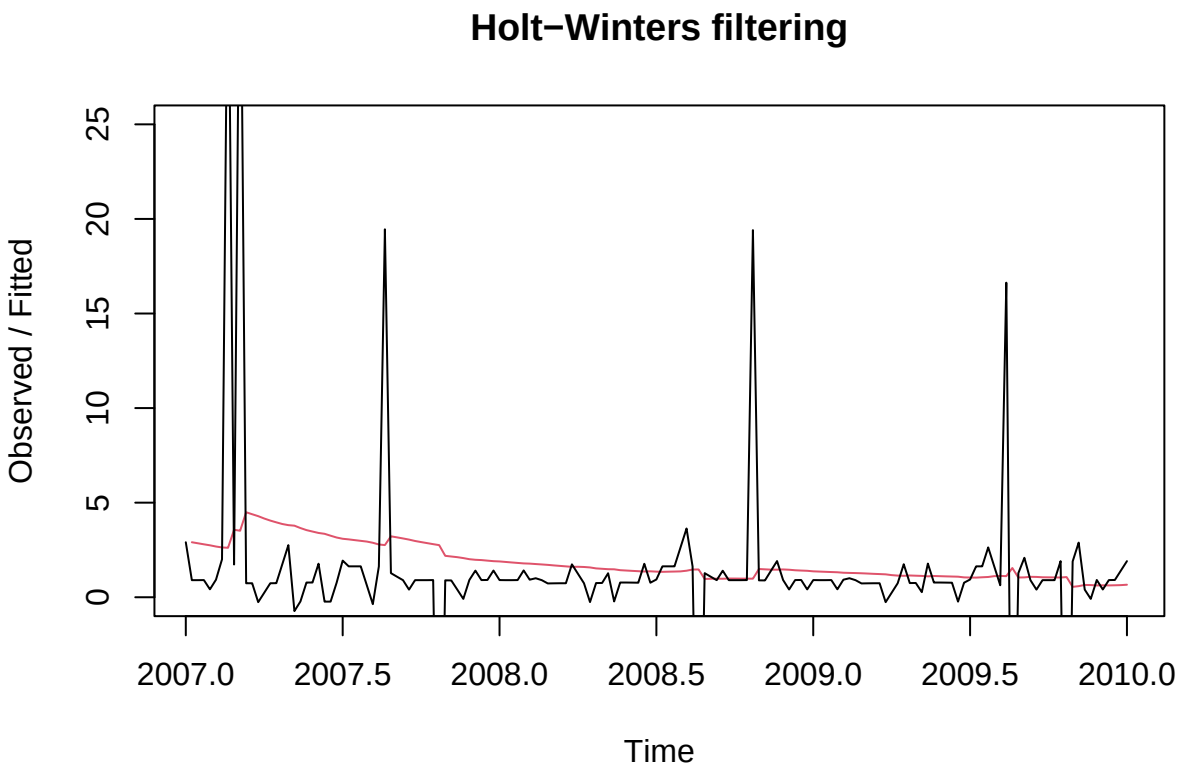

```
plot(decompose(tsSM3_070809Adjusted))
```



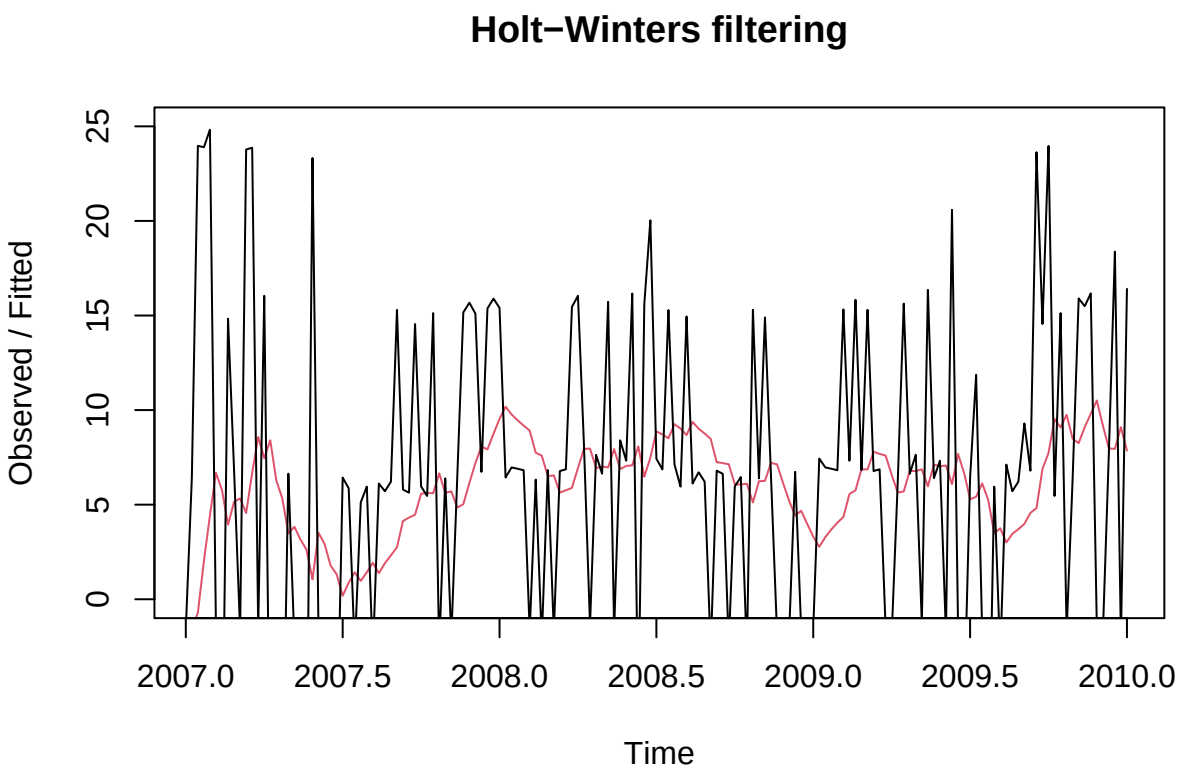

```
## Holt Winters Exponential Smoothing & Plot
tsSM1_HW070809 <- HoltWinters(tsSM1_070809Adjusted, beta=FALSE, gamma=FALSE)
tsSM2_HW070809 <- HoltWinters(tsSM2_070809Adjusted, beta=FALSE, gamma=FALSE)
tsSM3_HW070809 <- HoltWinters(tsSM3_070809Adjusted, beta=FALSE, gamma=FALSE)


plot(tsSM1_HW070809, ylim = c(0, 25))
```



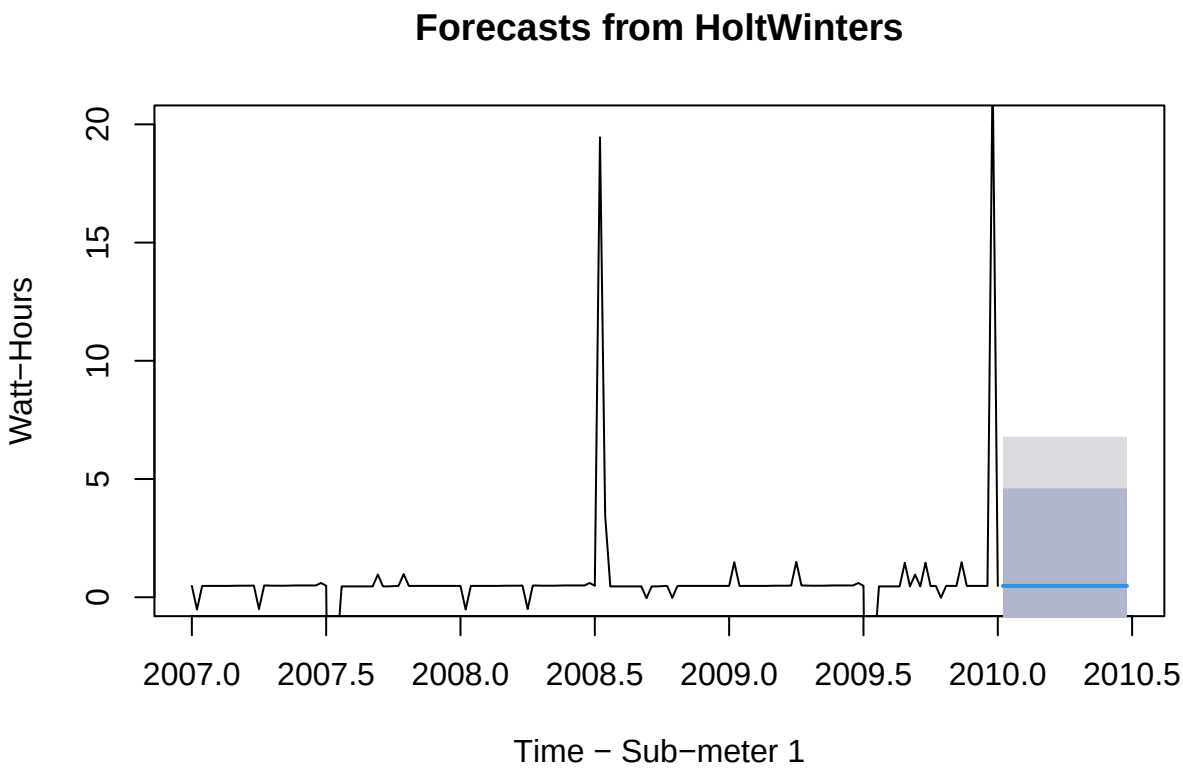

```
plot(tsSM2_HW070809, ylim = c(0, 25))
```



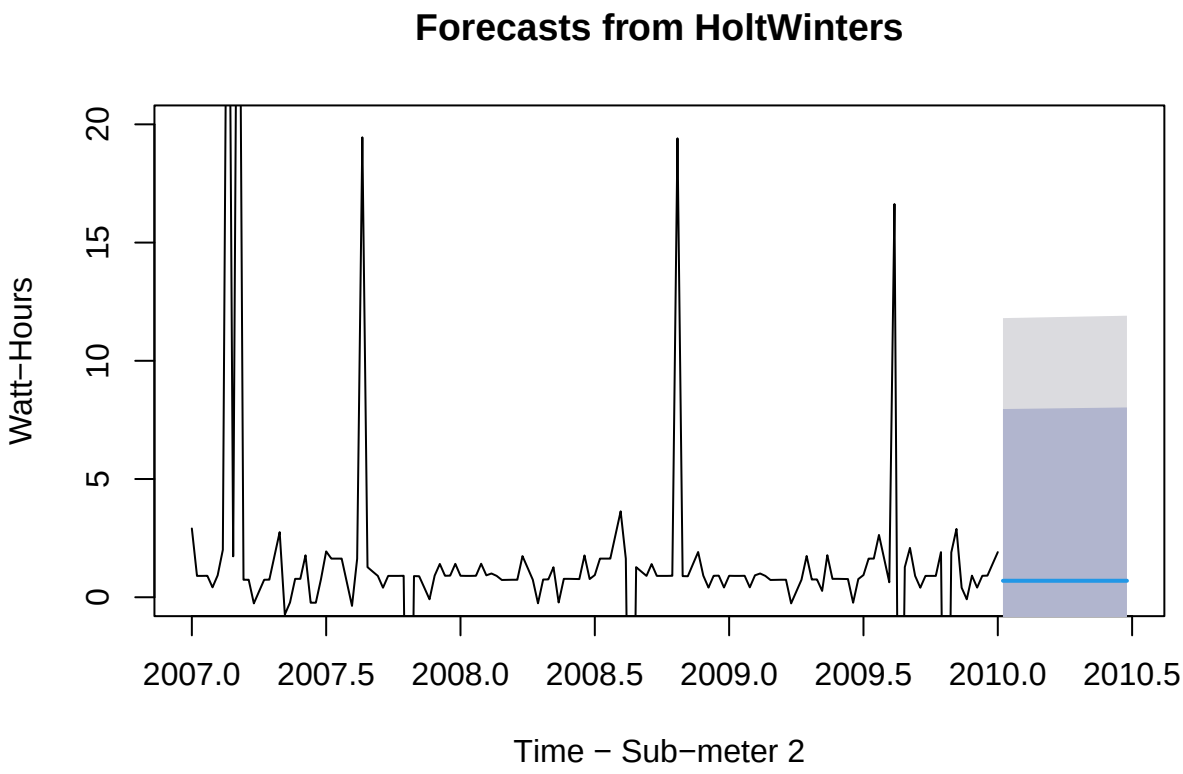

```
plot(tsSM3_HW070809, ylim = c(0, 25))
```



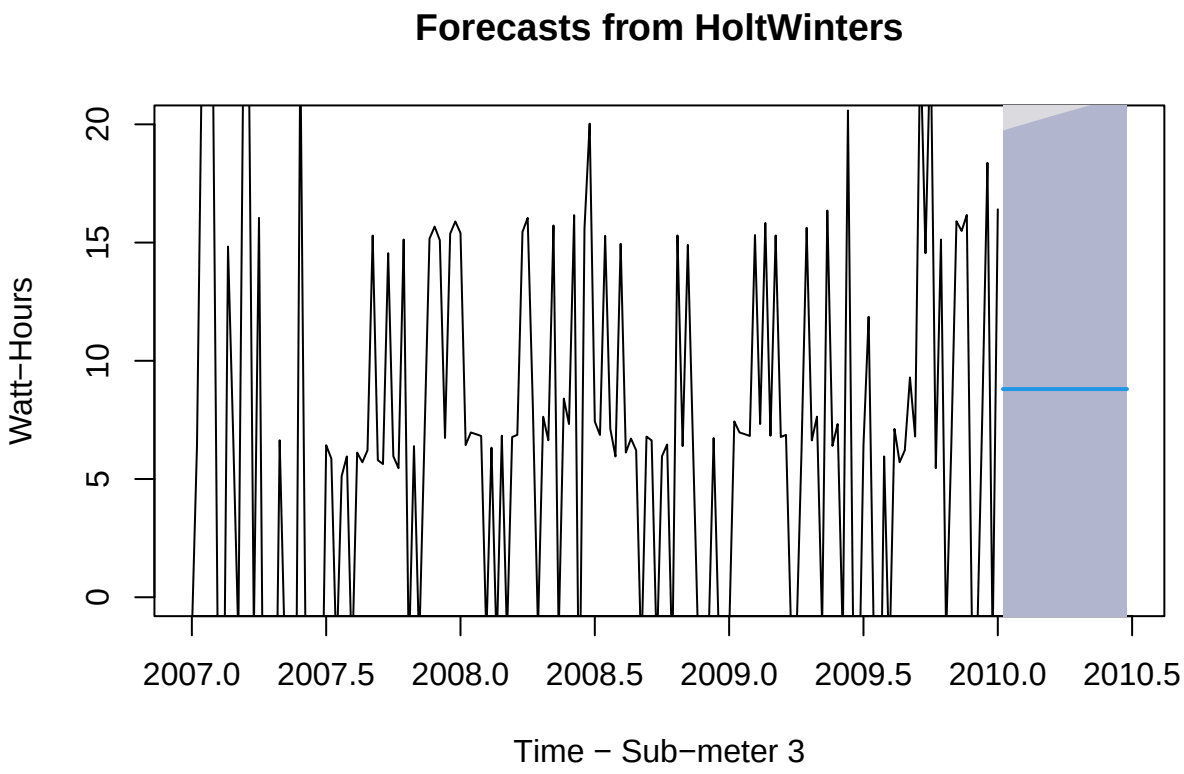

```
## HoltWinters forecast & plot
tsSM1_HW070809for <- forecast(tsSM1_HW070809, h=25)
tsSM2_HW070809for <- forecast(tsSM2_HW070809, h=25)
tsSM3_HW070809for <- forecast(tsSM3_HW070809, h=25)


plot(tsSM1_HW070809for, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time - Sub-meter 1")
```



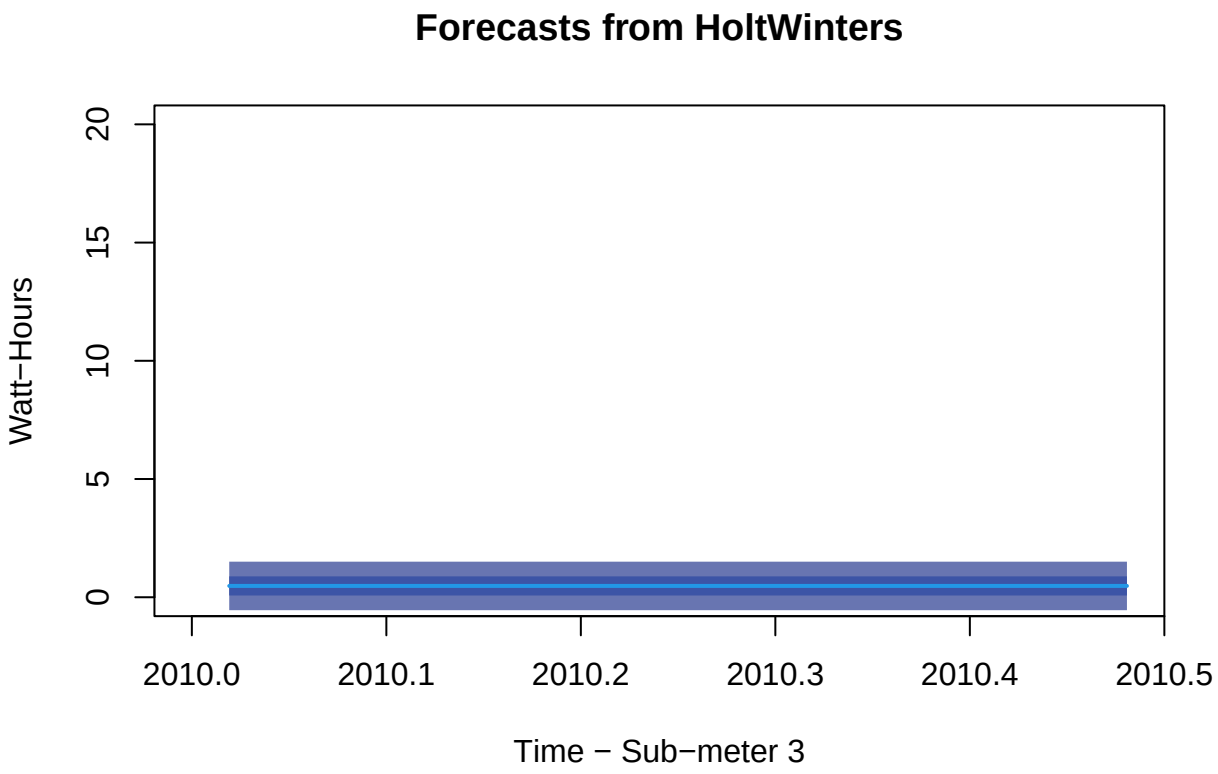

```
plot(tsSM2_HW070809for, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time - Sub-meter 2")
```



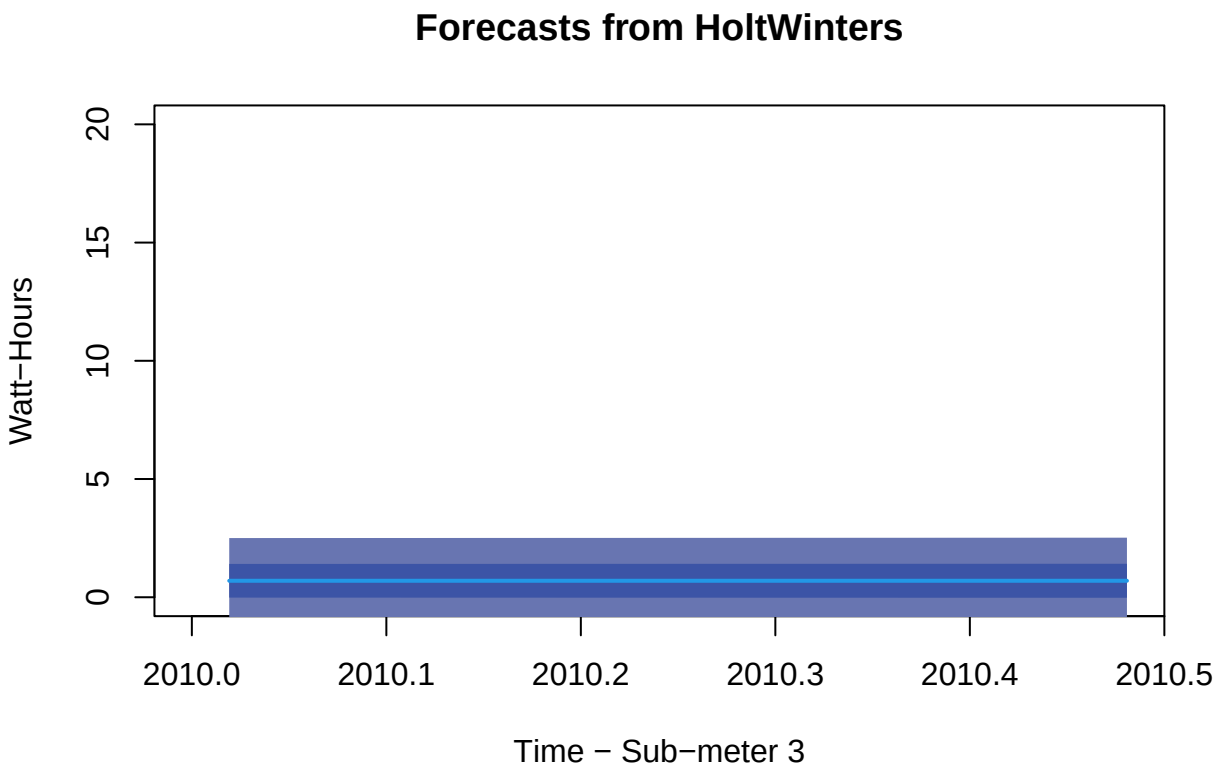

```
plot(tsSM3_HW070809for, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time - Sub-meter 3")
```



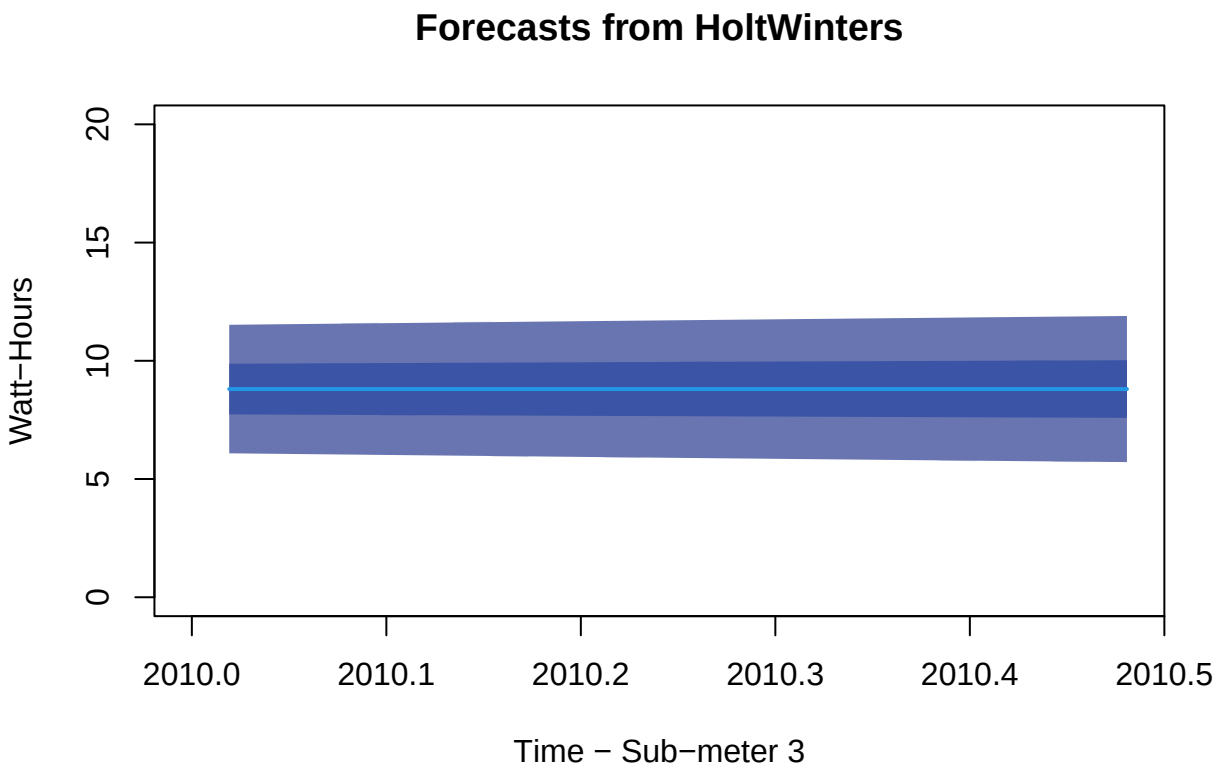

```
## Forecast HoltWinters with diminished confidence levels
tsSM1_HW070809forC <- forecast(tsSM1_HW070809, h=25, level=c(10,25))
tsSM2_HW070809forC <- forecast(tsSM2_HW070809, h=25, level=c(10,25))
tsSM3_HW070809forC <- forecast(tsSM3_HW070809, h=25, level=c(10,25))
## Plot only the forecasted area
plot(tsSM1_HW070809forC, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time - Sub-meter 3", start(2010))
```




```
plot(tsSM2_HW070809forC, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time - Sub-meter 3", start(2010))
```




```
plot(tsSM3_HW070809forC, ylim = c(0, 20), ylab= "Watt-Hours", xlab="Time - Sub-meter 3", start(2010))
```




**Report**

PPT Report to management.

1. Visualizations
2. Time Series Visualizations
3. Linear Regression Forecast Visualizations
4. Decomposition Visualizations + Chart comparing the summary statistics for the seasonal, trend, and remainder components
5. Holt-Winters Forecasting
6. Any correlations/predictions from the data. What other type of data could be used for the analysis?
7. Summary Statement addressing the goal of this project
8. Five Business recommendations you can suggest
9. Lessons learned