PhotoVoltaic (PV) panel performance at TN13 1SX

analysis of 'produced_kWh' 10 year records

emmanuel mavromatakis

04/03/2021

objective

the objective of this procedure is confirm the continuous "expected performance" of the PV panels. the data used are the monthly kWh readings taken from the date of installation of the panels to the end of 2020. the monthly output is compared with the "expected kwh production" (exp_kWh). the exp_kWh was provided by the installation contractor based on published data up to 2011.

summary

analysis of the recorded kWh suggests that the efficiency of the panels has not been reduced.

recommendations

1. identify parameters affecting performance of the panels and investigate correlations
2. obtain daily records at specified intervals, recording:
   - date,
   - time,
   - degree of cloud,
   - temperature
   - kWh reading
3. survey published "expected performance" of PV panels at SE England

analysis and results

required library

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------
tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

read_csv solar panel readings and store it in tibble pnl

```
file <- "solar_pv_monthly_reading_orchardsend_aug2011_to_dec2020_rev2.csv"

pnl <- read_csv(file)

## Parsed with column specification:
## cols(
##   year = col_double(),
##   month = col_character(),
##   season = col_character(),
##   reading_kwh = col_double(),
##   day_length_hour = col_double(),
##   e_grid_expt_kwh = col_double(),
##   sunshine_hrs_per_day = col_double()
## )
```

read the expected kWh for SE England. the expected kWh per month was provided at the time of installation (aug.2011)

```
exp_kwh <- read_csv("solar_pv_expected_southeast.csv")

## Parsed with column specification:
## cols(
##   month = col_character(),
##   exp_kwh = col_double(),
##   exp_temp = col_double()
## )
```

store the months in calender order and use it set the levels of the month vector

```
clevel <- c("jan", "feb", "mar", "apr", "may", "jun",
            "jul", "aug", "sep", "oct", "nov", "dec")
```

change the class of the pnl$month vector to factor with levels the clevel defined earlier

```
pnl <- pnl %>% mutate(month = factor(month, levels = clevel))
```

select the parameters “year”, “month” and “reading_kwh” to create data for plotting

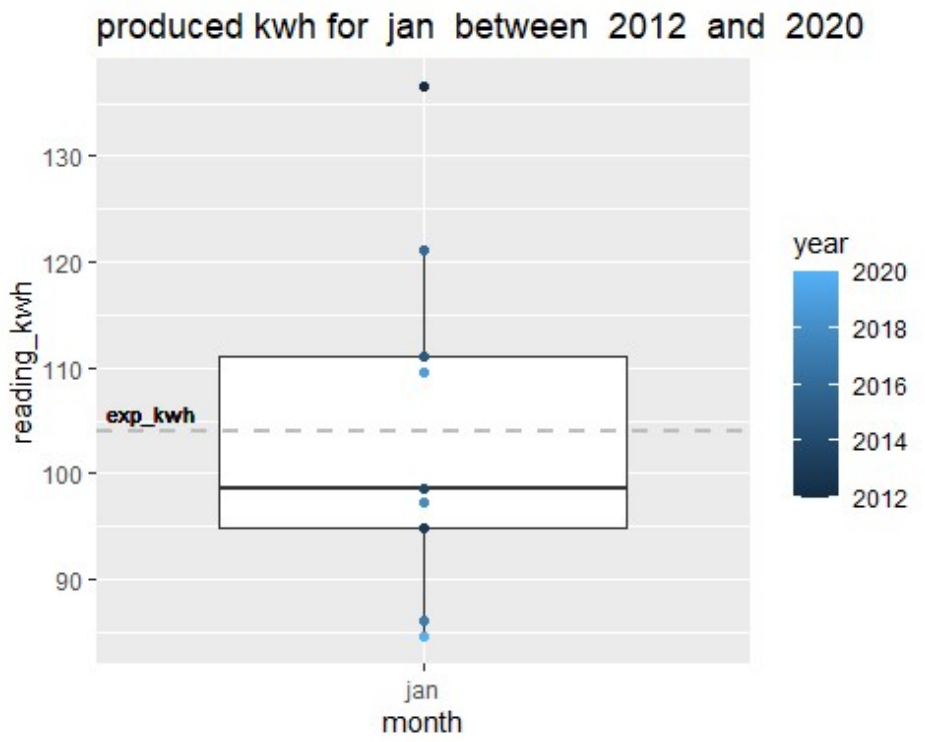
```
pnl <- pnl %>% select(year, month, reading_kwh)
```

for each month over the period between aug_2011 and dec_2020 calculate a. the kWh produced
b. the deviation (+/-) from exp_kWh

```
delta_mean <- data.frame(month = clevel, delta = 1:12)

for (i in 1:12){
  xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>% mutate(year = factor(year)) %>%
         mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
                group = seq(1, nrow(df), 1))
  df_stat <- df %>% summarise(avg_kwh = mean(reading_kwh),
                              sd_kwh = sd(reading_kwh))
  delta_mean$delta[i] <- df_stat$avg_kwh - exp_kwh$exp_kwh[i]
}
```

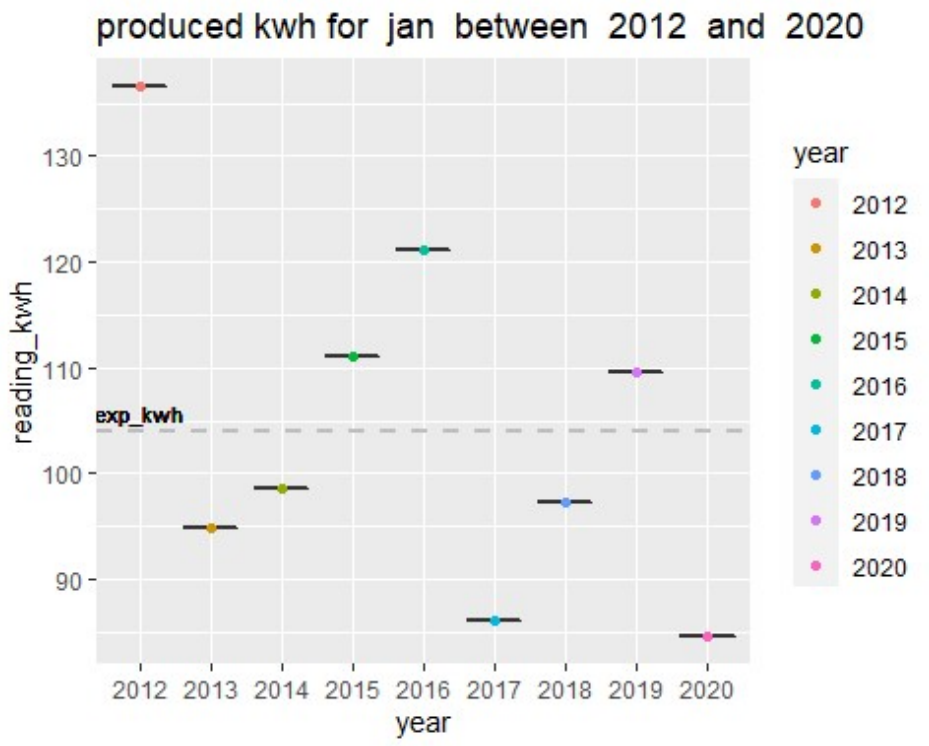
results for January

```
i <- 1
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "grey", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the above figure shows that the majority of the readings are below the expected kWh and there is an outlier for 2012

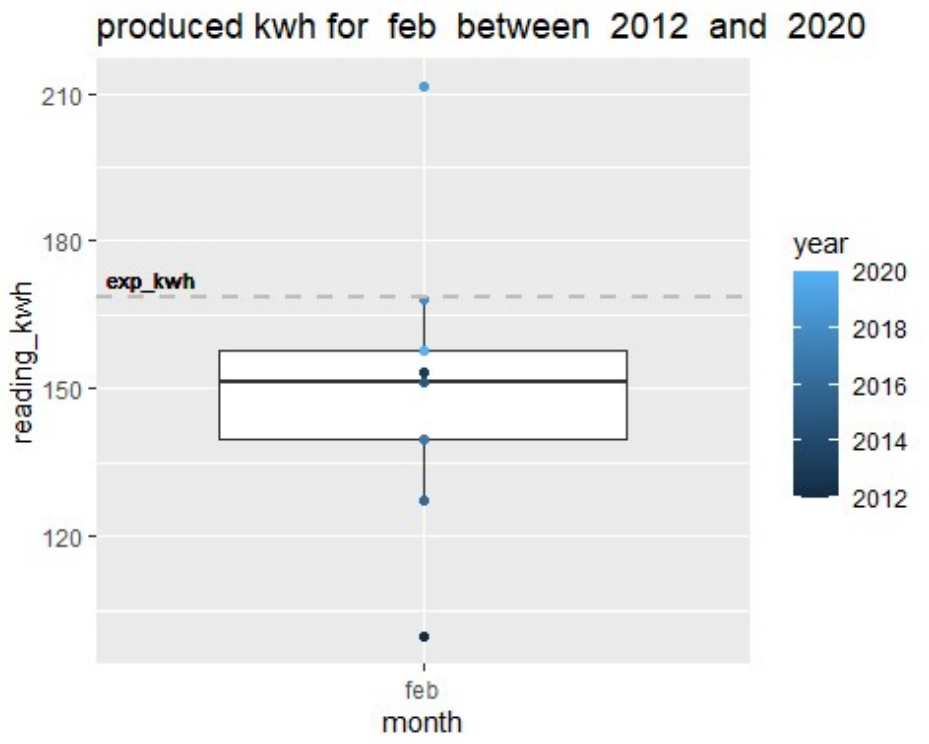
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "grey", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



this plot confirms that the outlier corresponds to 2012

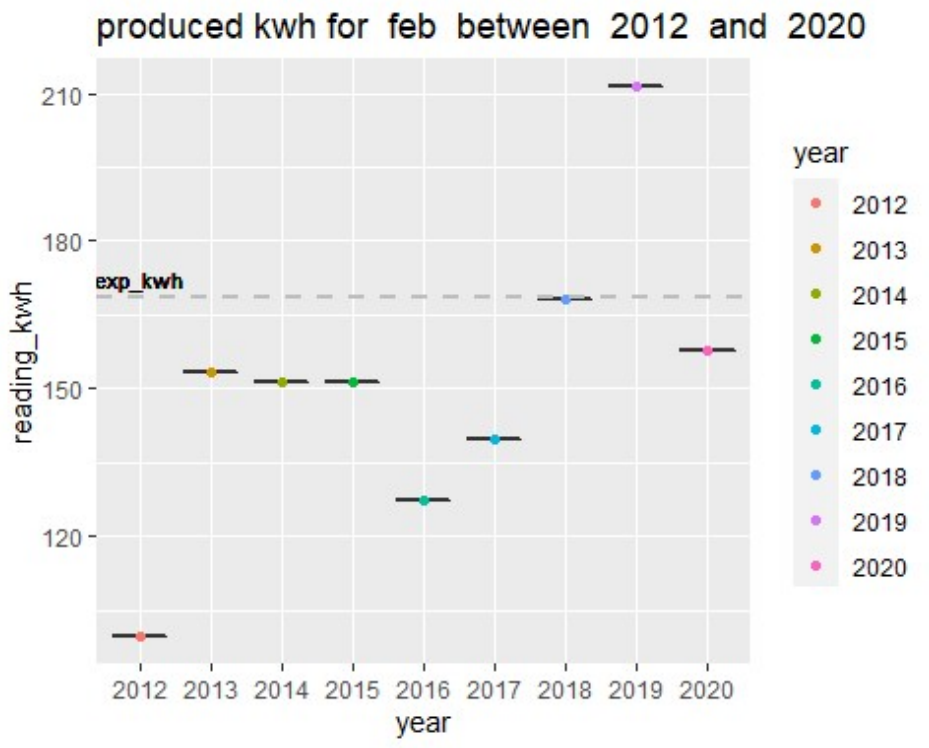
results for February

```
i <- 2
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the above figure shows that the readings for February are below the expected kWh and there are 2 outliers

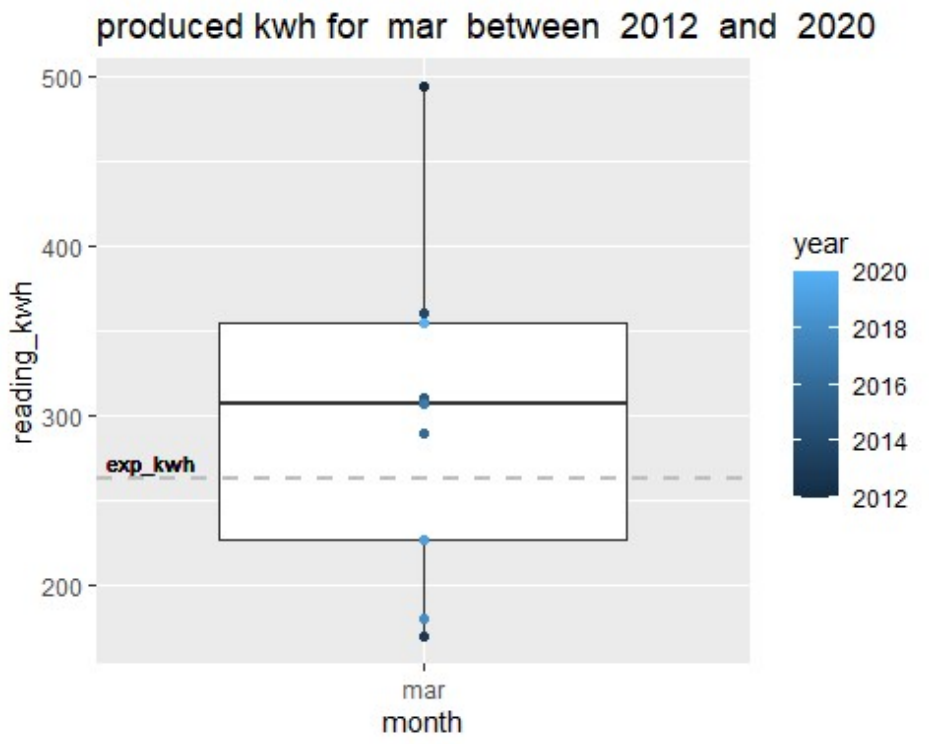
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



the above figure shows that the outliers correspond to 2012 and 2019 and year 2018 is the only one with produced kWh close to the expected value for February

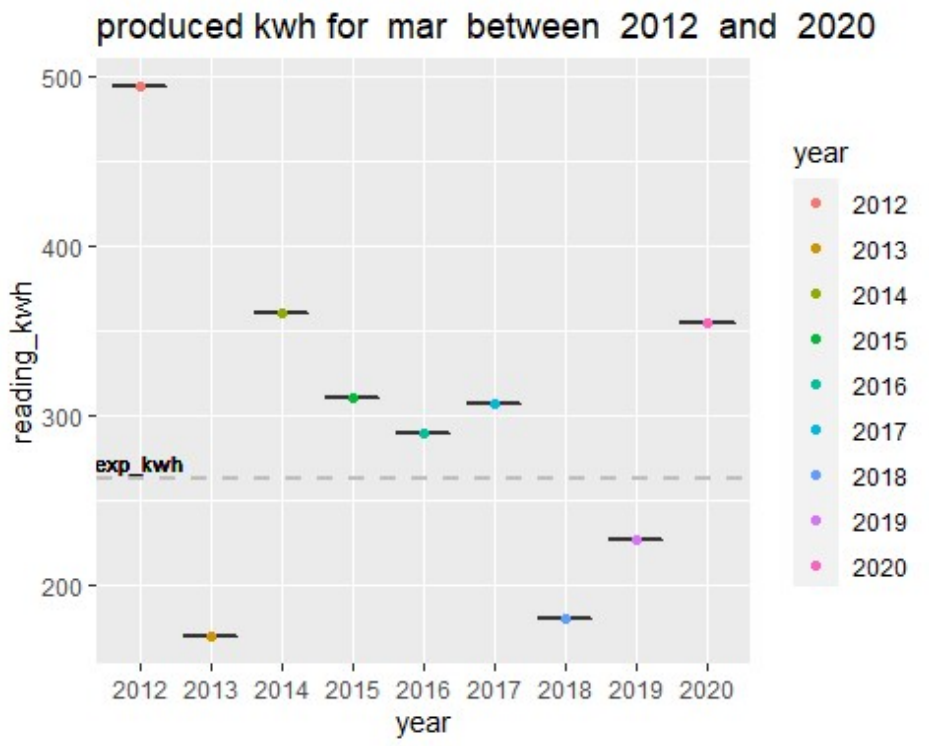
results for March

```
i <- 3
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "grey", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the produced kWh for March exceeded the expected kWh with the majority of the readings above the exp_kWh

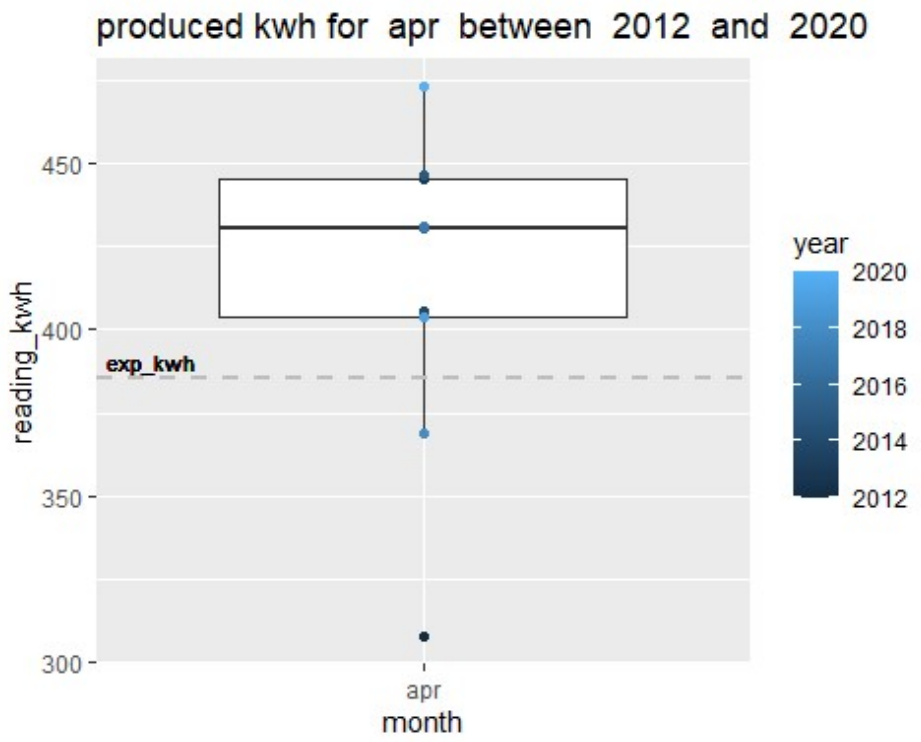
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



the above figure shows the variation of the produced kWh over the study period

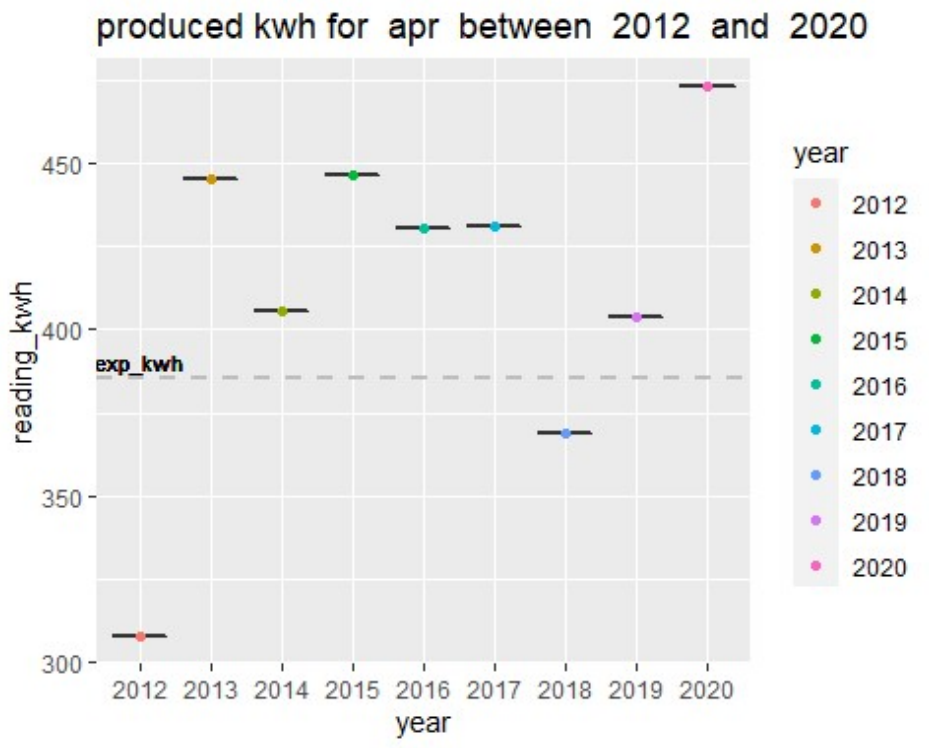
results for April

```
i <- 4
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the above figure shows that the majority of the readings for April are well above expected kWh with a negative outlier

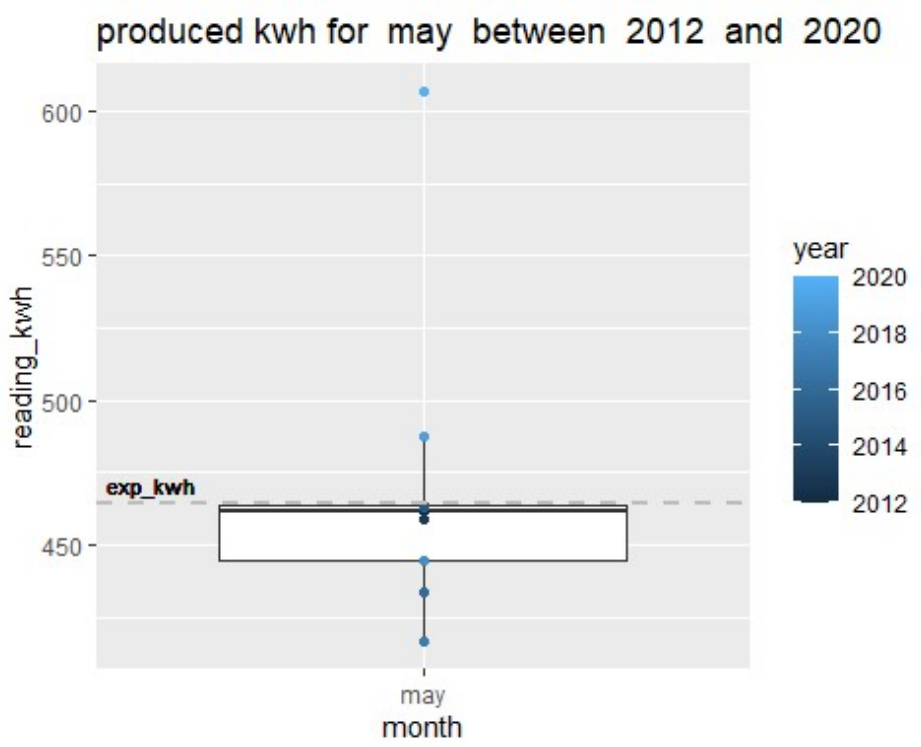
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



the above figure shows that the outlier for April corresponds to 2012. the Met Office records show that April 2012 was the wettest April on record in the UK since records began in 1910!

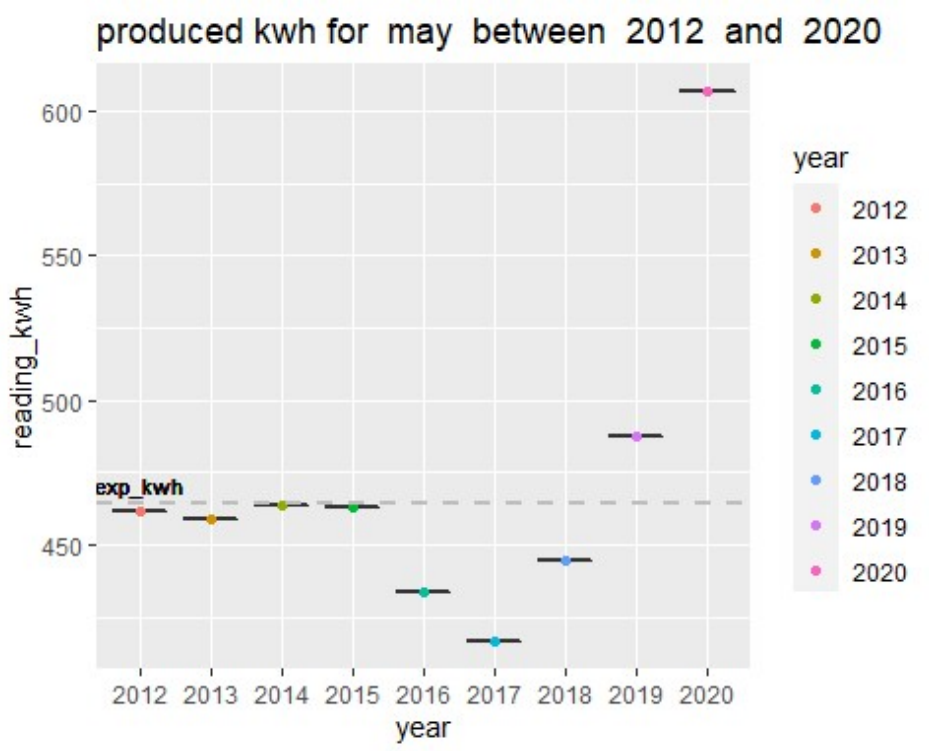
results for May

```
i <- 5
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the above figure shows that the readings for May are below the expected kWh and there is 1 positive outlier

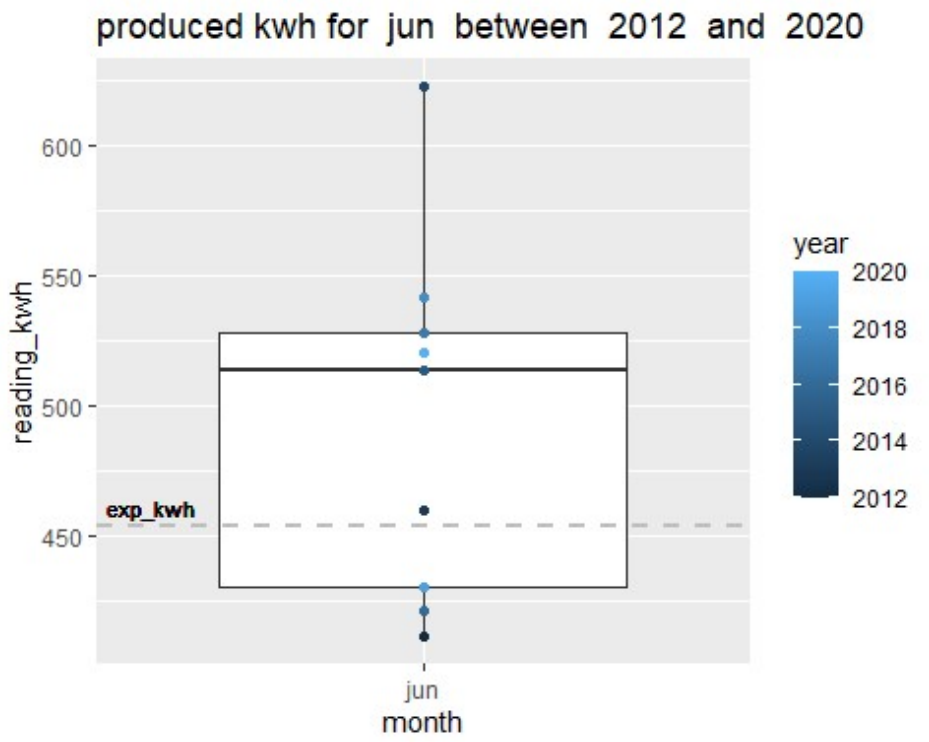
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



the above figure shows that the positive outlier corresponds to 2020 which was a sunny and warm month (2 months within the 1st lockdown)

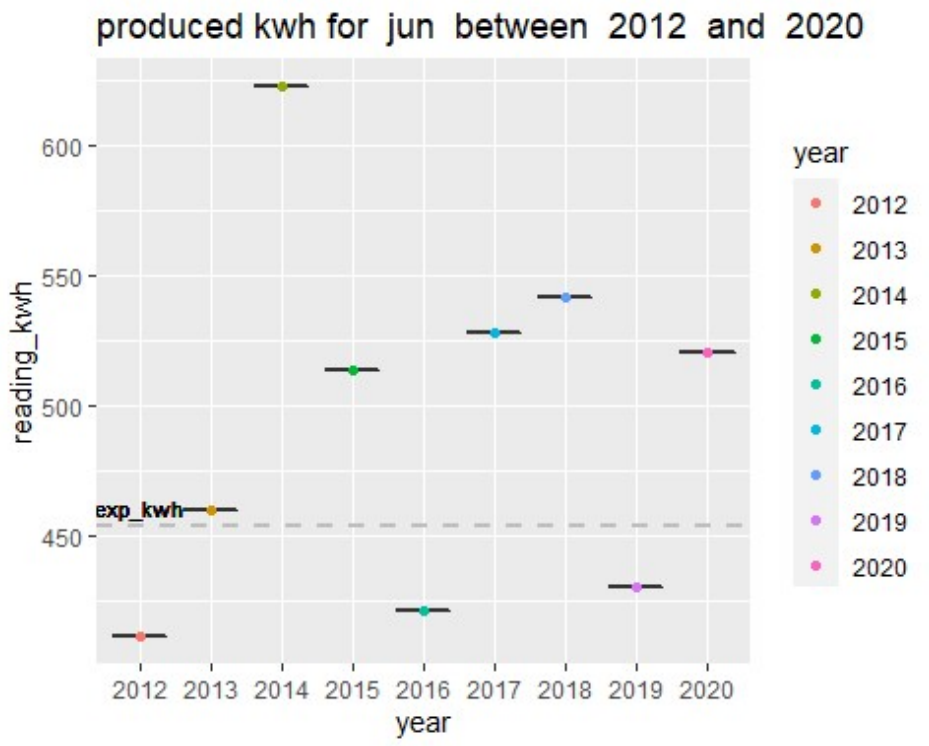
results for June

```
i <- 6
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the above figure shows that most of the readings for June are above the expected kWh

```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



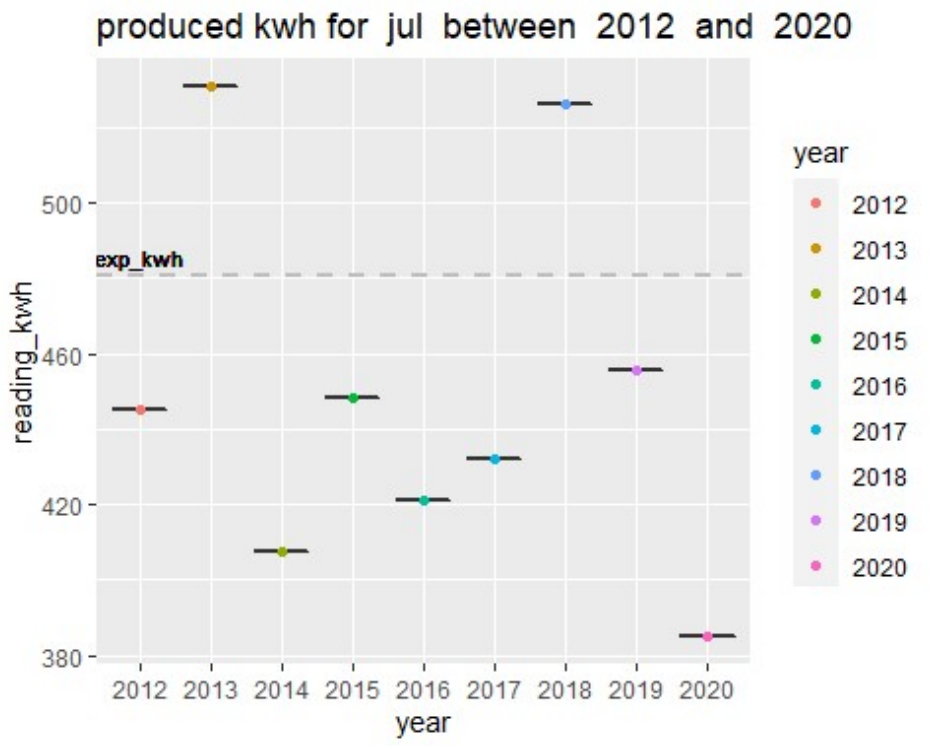
the low value for 2019 and high for 2020 are in line with the values one would expect based on weather conditions only as the Met Office records show

results for July

```
i <- 7
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```

the above figure shows that the July production was well below the expected kWh and there are 2 positive outliers

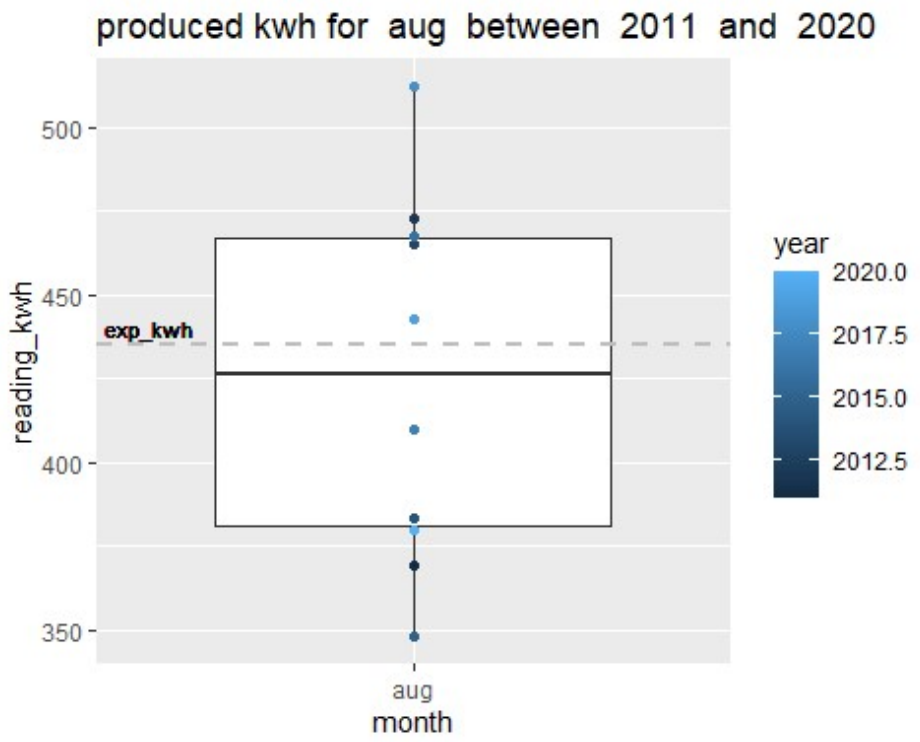
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



the above figure shows that the outliers correspond to 2013 and 2020

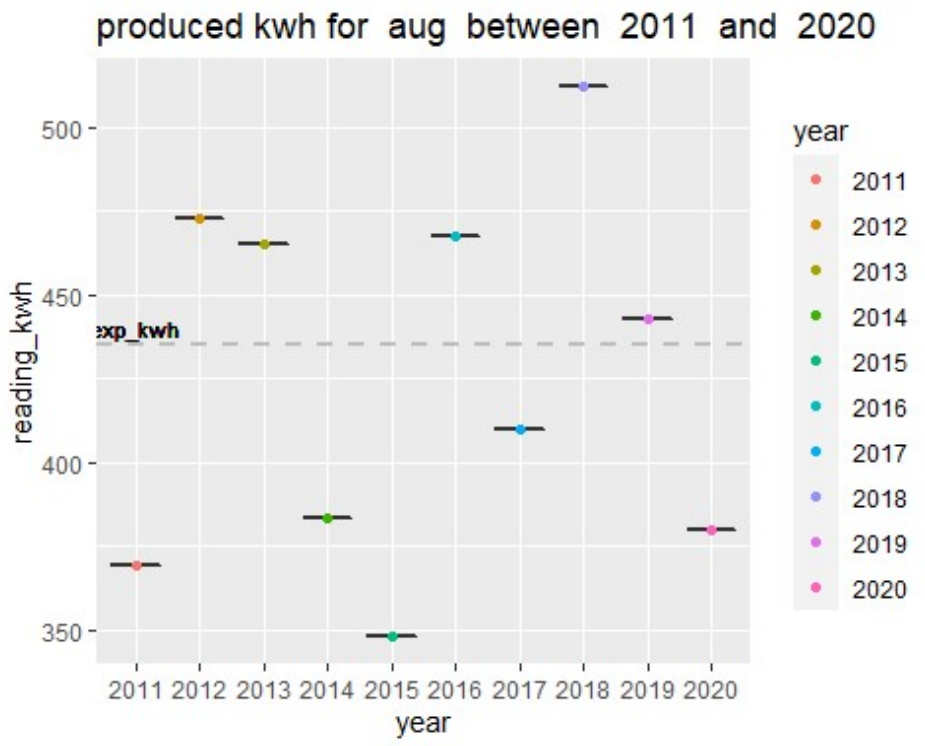
results for August

```
i <- 8
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the productions for August are practically symmetrical around the expected value

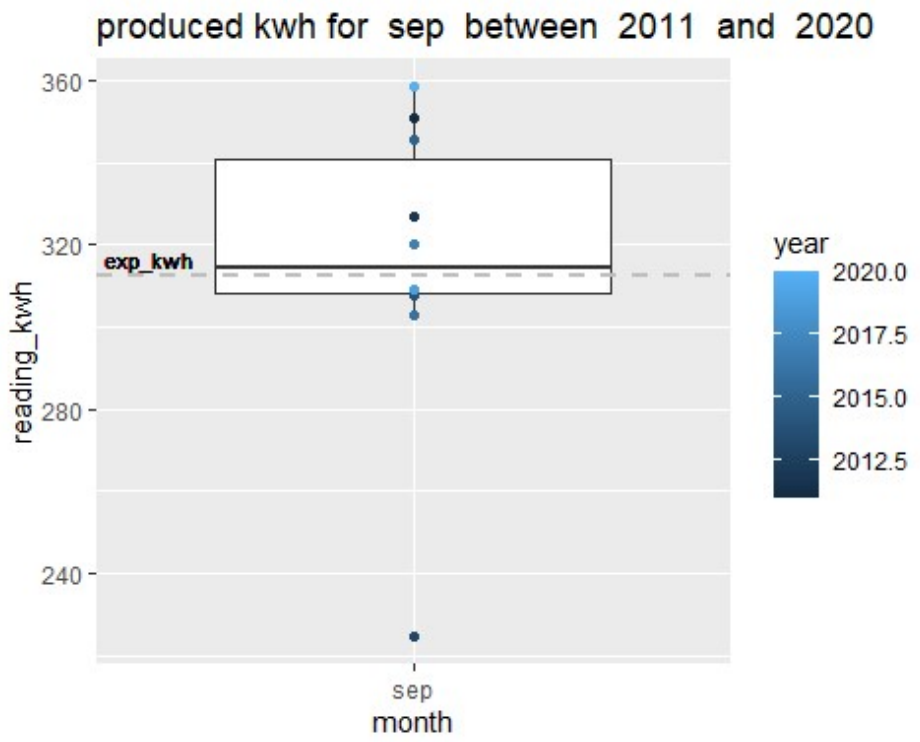
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
   ggtitle(paste0("produced kwh for  ", xm,
                  "  between  ", min(df$year),
                  "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



distribution of produced kWh for August over the study period

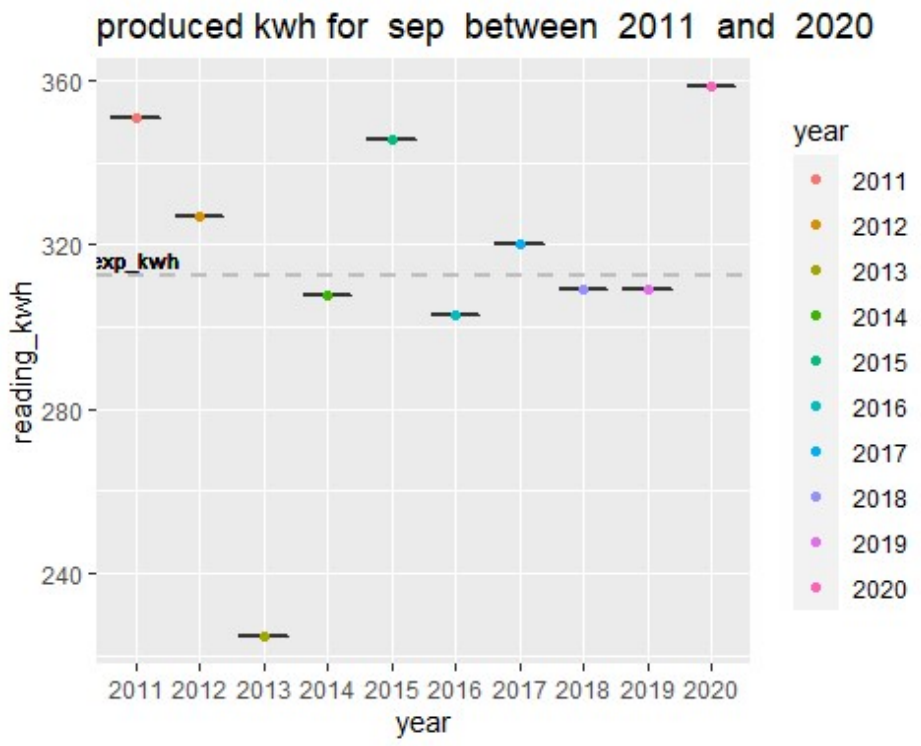
results for September

```
i <- 9
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



September readings produced a negative outlier

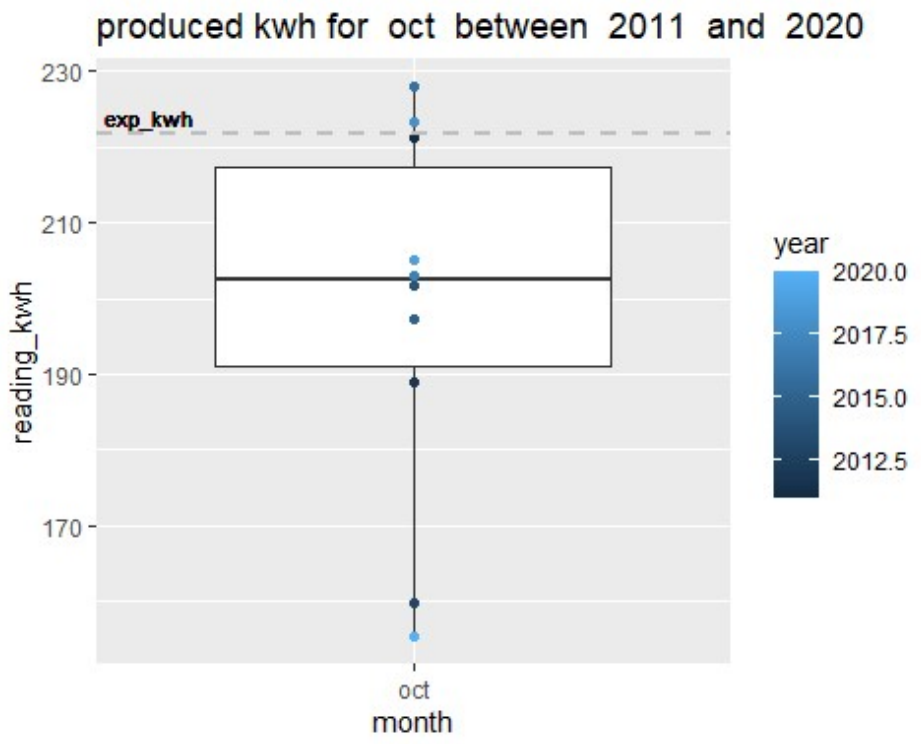
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



the negative outlier corresponds to 2013. the majority of the kWh produced is near the expected kWh

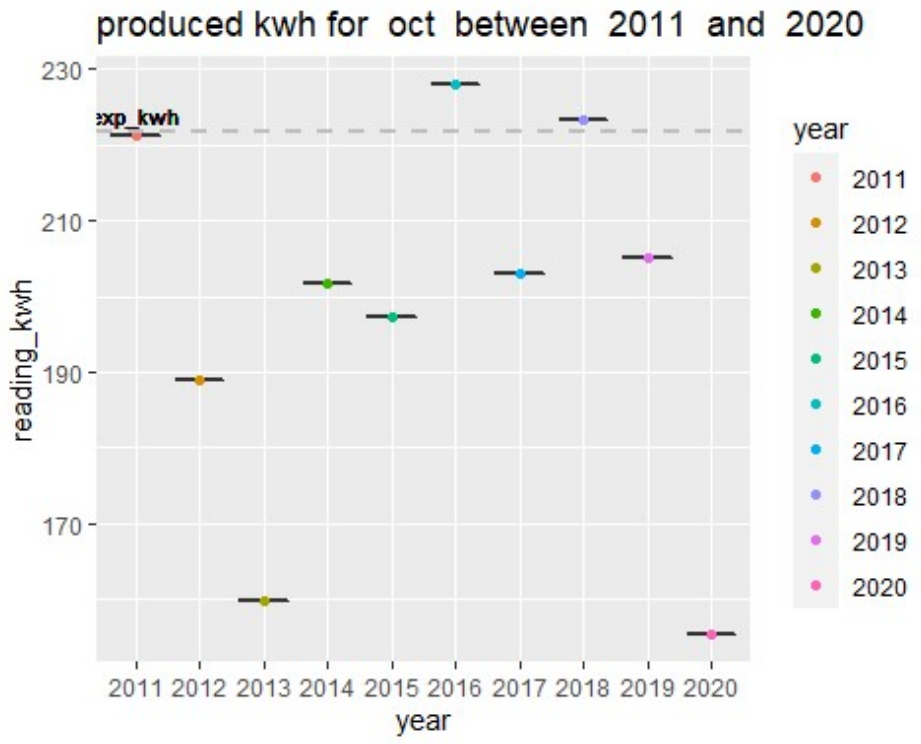
results for October

```
i <- 10
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



large deviations from the expected kWh for October with most of the values corresponding to lower kWh

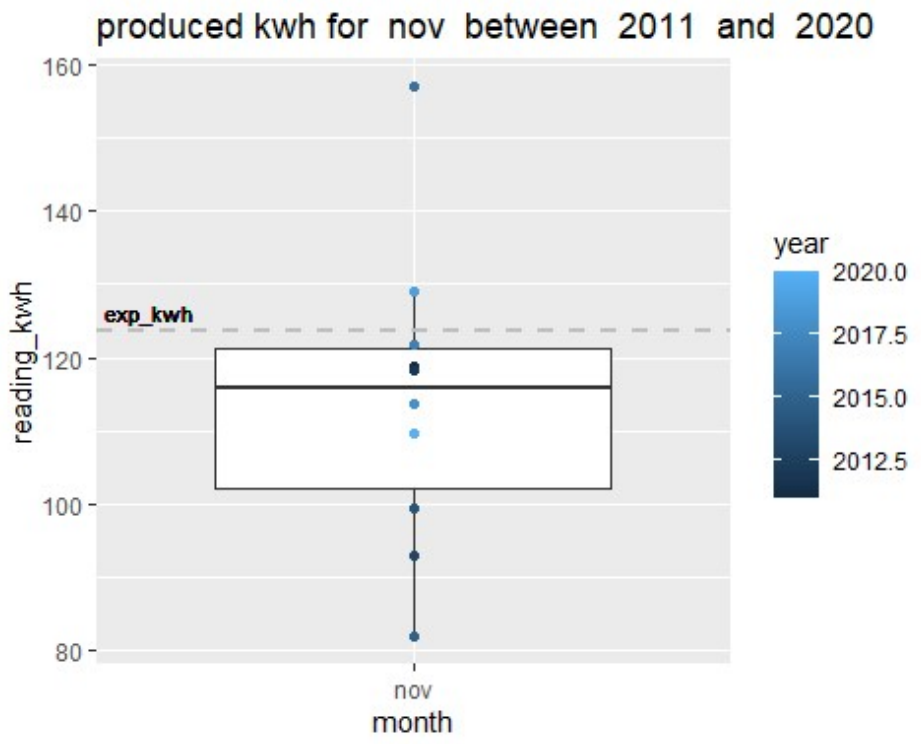
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



2020 has the lowest kWh over the study period. Met office records show that October 2020 was a very wet month

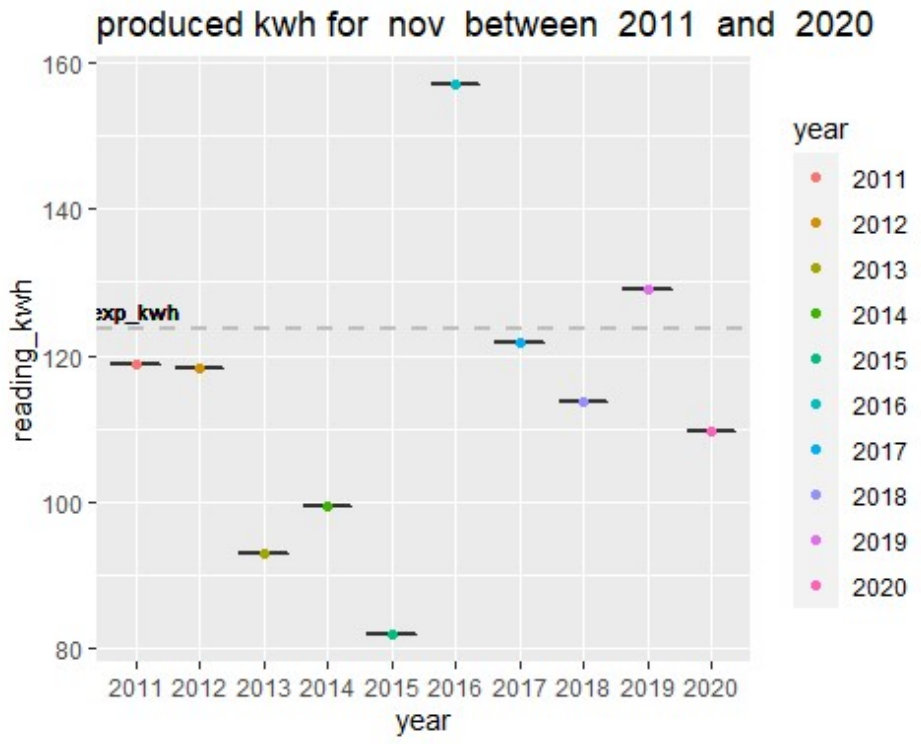
results for November

```
i <- 11
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



similar picture to October for November with one positive oulier

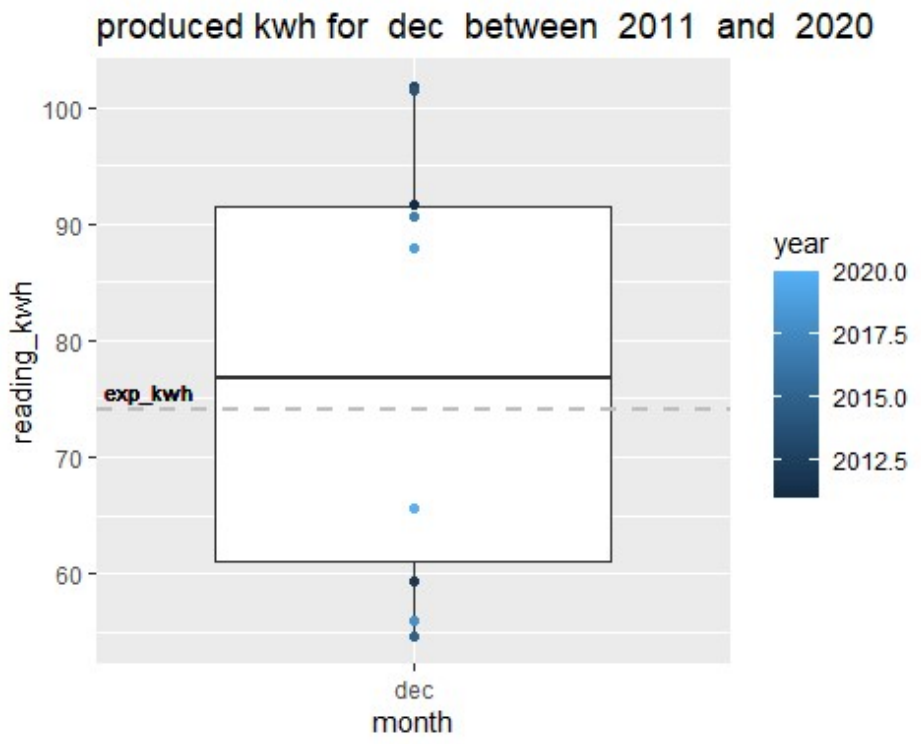
```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "gray", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



the above figure shows that November 2015 and November 2016 have mirror image kWh readings. The corresponding sunshine hours for SE England are 33 and 80 which explains the large difference in kWh

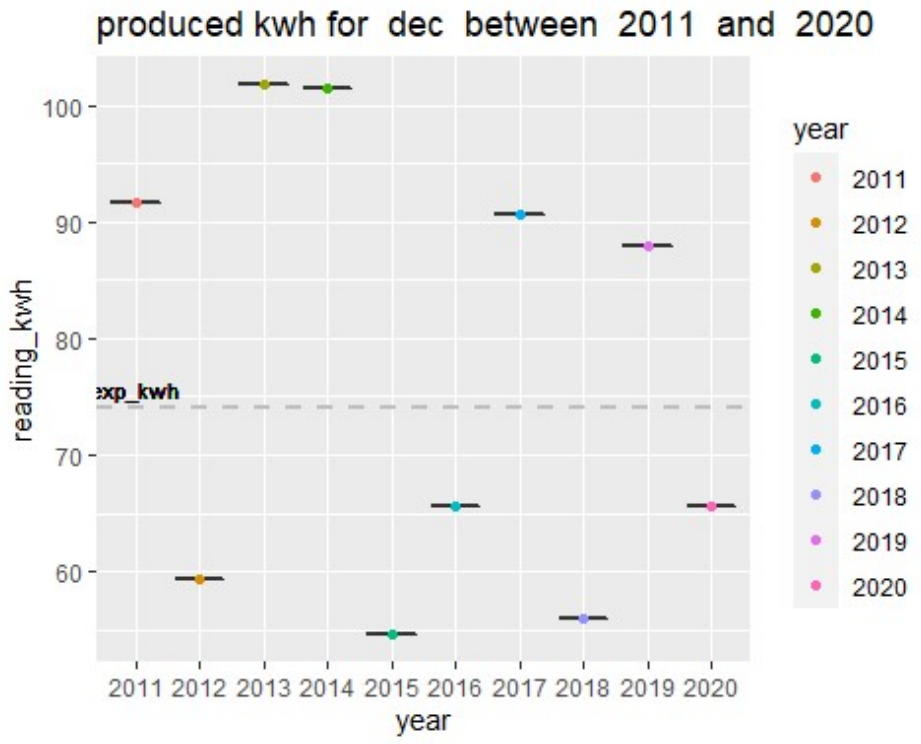
results for December

```
i <- 12
xm <- clevel[i]
  df <- filter(pnl, month == xm)
#
  df %>%
     ggplot(aes(x = month, y = reading_kwh)) +
     geom_boxplot() +
     geom_point(aes(color = year)) +
     geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
                color = "gray", size = 1) +
     geom_text(aes(x = 0.5, y = exp_kwh$exp_kwh[i],
                   label = "exp_kwh") , vjust = -0.5, cex = 3) +
     ggtitle(paste0("produced kwh for  ", xm,
                    "  between  ", min(df$year),
                    "  and  ", max(df$year)))
```



the above figure shows that the kWh production during December is evenly balanced around the expected kWh.

```
df %>% mutate(year = factor(year)) %>%
  mutate(exp_kwh = rep(exp_kwh$exp_kwh[i],nrow(df)),
         group = seq(1, nrow(df), 1)) %>%
  ggplot(aes(x = year, y = reading_kwh)) +
  geom_boxplot()+
  geom_point(aes(color = year)) +
  geom_hline(yintercept = exp_kwh$exp_kwh[i], linetype = 2,
             color = "grey", size = 1) +
  ggtitle(paste0("produced kwh for  ", xm,
                 "  between  ", min(df$year),
                 "  and  ", max(df$year))) +
  geom_text(aes(x = 1.0, y = exp_kwh,
                label = "exp_kwh") , vjust = -0.5, cex = 3)
```



there were positive and negative relative to expected kWh are evenly distributed for December.

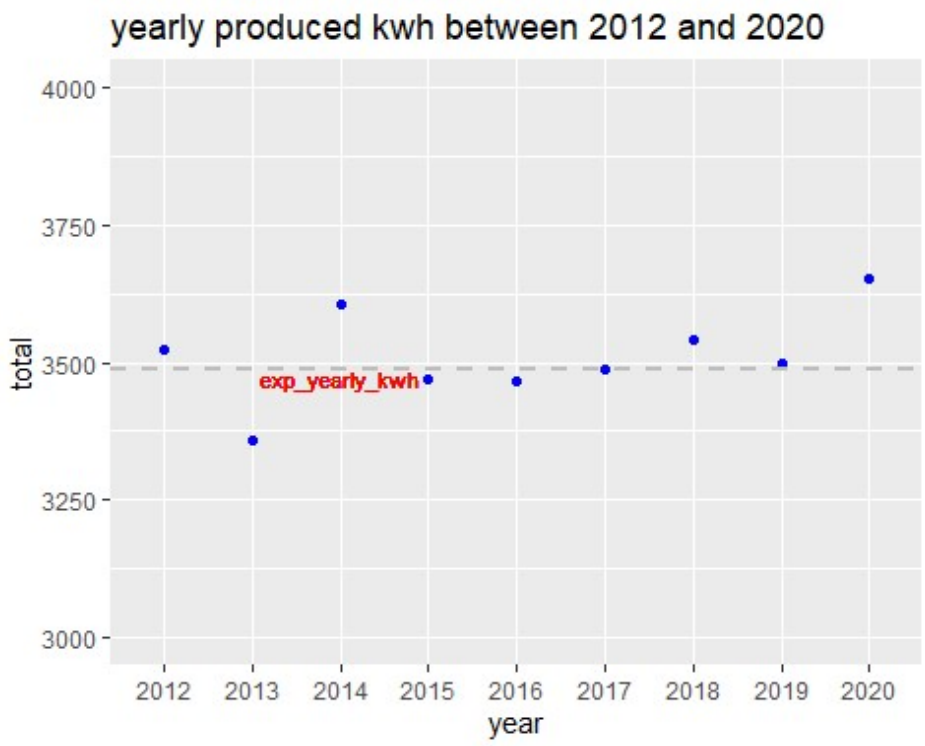
results for yearly produced kWh

```
  pnl <- pnl %>% mutate(year = factor(year))
  df_total <- pnl %>% group_by(year) %>%
     summarise(total = sum(reading_kwh)) # store the totals per year

## `summarise()` ungrouping output (override with `.groups` argument)

  total_exp <- sum(exp_kwh$exp_kwh) # calculate the total expected
```

generate plot of yearly production

```
  p <- df_total %>%                          # store plot in p
    filter(as.numeric(year) > 1) %>%
    ggplot(aes(x = year, y = total)) +
    ylim(3000, 4000)

  p + geom_point(color = "blue") +     # plot the expected production
    geom_hline(yintercept = total_exp,
               linetype = 2,
               color = "grey",
               size = 1) +
    ggtitle(paste0("yearly produced kwh between 2012 and 2020")) +
    geom_text(aes(x = 3, y = total_exp,
```

```
            label = "exp_yearly_kwh") ,
            vjust = 1, color= "red", cex = 3)
```



the graph of the yearly produced kWh shows that the per year produced kWh does not deviate much from the yearly expected kWh.

conclusion

the small deviation between “exp_kWh” and recorded kWh suggests that the efficiency of the panels has not been reduced.

recommendations

a. research into the reasons for the identified outliers and suggest potential relationships (correltations) among the parameters affecting performance of the panels

b. increase the number of observations by obtaining daily records at specified intervals, recording: date, time, degree of cloud, temperature and kWh reading.

c. perform literature survey for “expected performance” of PV panels at SE England