other statistics

Code

What is a normal rate of power usage per day? per week? When should the homeowner be alerted that power is exceeding a set amount?

To do this, we will look at:

– Submeter 1: Kitchen There are going to be lots of days where there is no kitchen power use. That's logically ok. Therefore, we will look at a table of Kitchen use that's not zero, per day. We will find the average, as well as the maximum, and even a threshold. The homeowner could be alerted when the amount exceeds a certain limit.

Hide

```
# filter 0 out

submetersByDay_noZeros <- submetersByDay %>%
  filter(Kitchen > 0, LaundryRoom >0 , WaterHeater_AC >0)
summary(submetersByDay_noZeros)
```

```
      date                Kitchen         LaundryRoom      WaterHeater_AC  TotalEnergy_pe
rDay
 Min.   :2006-12-17   Min.   :    1.0   Min.   :   43.0   Min.   :  212   Min.   :  769
 1st Qu.:2007-12-13   1st Qu.:  979.2   1st Qu.:  455.5   1st Qu.: 7748   1st Qu.:10544
 Median :2008-12-02   Median : 1386.5   Median :  959.5   Median : 9739   Median :13352
 Mean   :2008-11-27   Mean   : 1931.2   Mean   : 2051.3   Mean   : 9905   Mean   :13888
 3rd Qu.:2009-11-10   3rd Qu.: 2489.2   3rd Qu.: 3031.8   3rd Qu.:12037   3rd Qu.:16739
 Max.   :2010-11-26   Max.   :11859.0   Max.   :12126.0   Max.   :23815   Max.   :31521
   day_index
 Min.   :   2.0
 1st Qu.: 362.2
 Median : 717.0
 Mean   : 711.7
 3rd Qu.:1059.8
 Max.   :1433.0
```

Hide

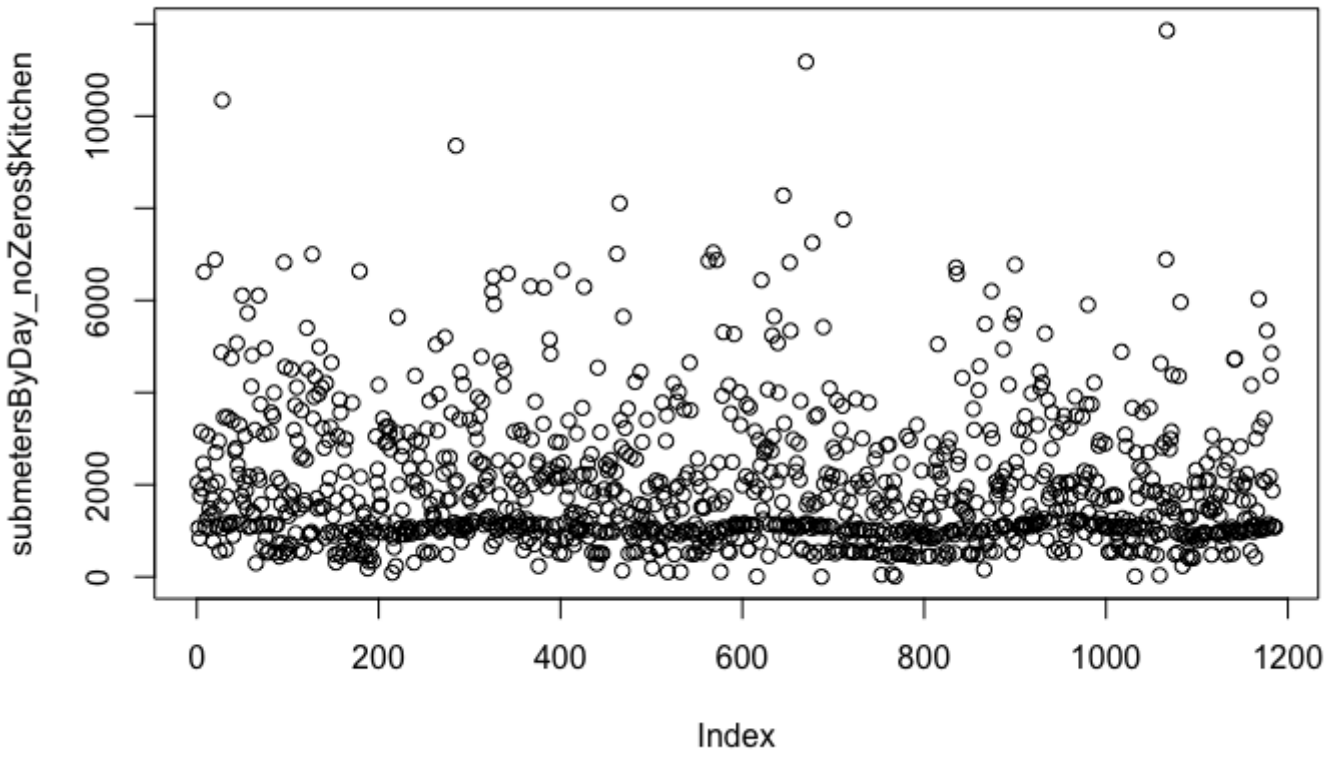
```
# Look at how many values in the data set are above the 3rd quarter threshold
sum(submetersByDay$Kitchen > 2489.2)
```

```
[1] 298
```

298 is way too many. We would be notifying the homeowner almost everyday for 10 months.

Hide

```
plot(submetersByDay_noZeros$Kitchen)
```



Based on the visual above, let's notify them of usage above 8000. That wuld indicate something irregular.The incidents where the power consumption is higher than 8000 could even indicate days where they perhaps forgot to turn off the oven.

Hide

```
submetersByDay %>%  filter ( Kitchen > 8000)
```

date <date>	Kitchen <dbl>	LaundryRoom <dbl>	WaterHeater_AC <dbl>	TotalEnergy_perDay <dbl>	day_index <int>
2007-01-21	10343	5550	11599	27492	37
2007-12-01	9356	338	16781	26475	350
2008-06-18	8109	3032	11369	22510	550
2009-01-31	8277	6606	14241	29124	777
2009-03-01	11178	2713	9350	23241	806
2010-06-06	11859	600	10405	22864	1266

6 rows

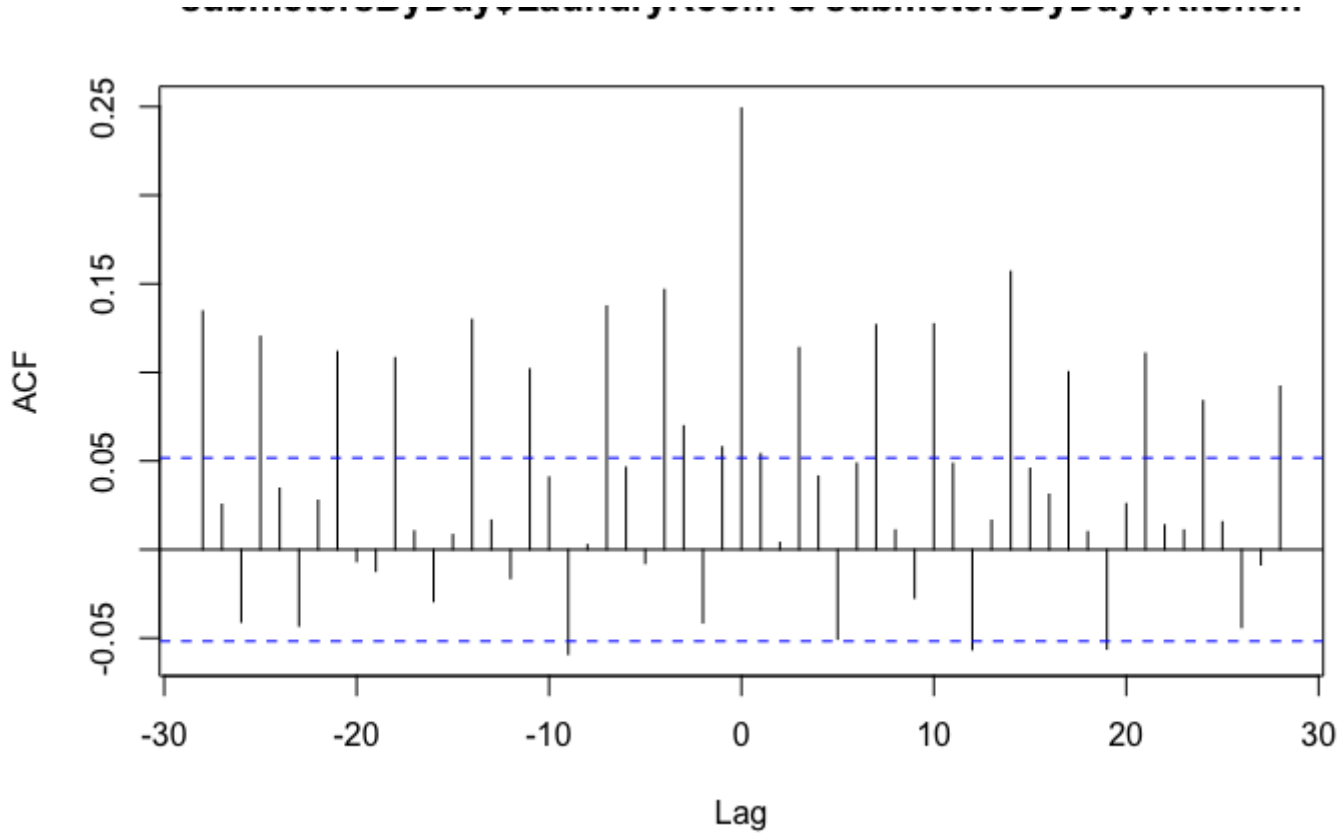
only two days in which kitchen energy usage exceeds 8000 W-H, does WaterHeater_AC exceeds it's 3rd quarter threshold of 12037 W-H. This means that highest kitchen usage is not necessarily associated with high water heater (dishwasher) or AC(heating) usage. Is there a better way to find this correlation?

We will use the crosscorrelation function

Correlation

Hide

```
ccf (submetersByDay$LaundryRoom, submetersByDay$Kitchen)
```



I will need to come back to this. I don't have much time… unfortunately.

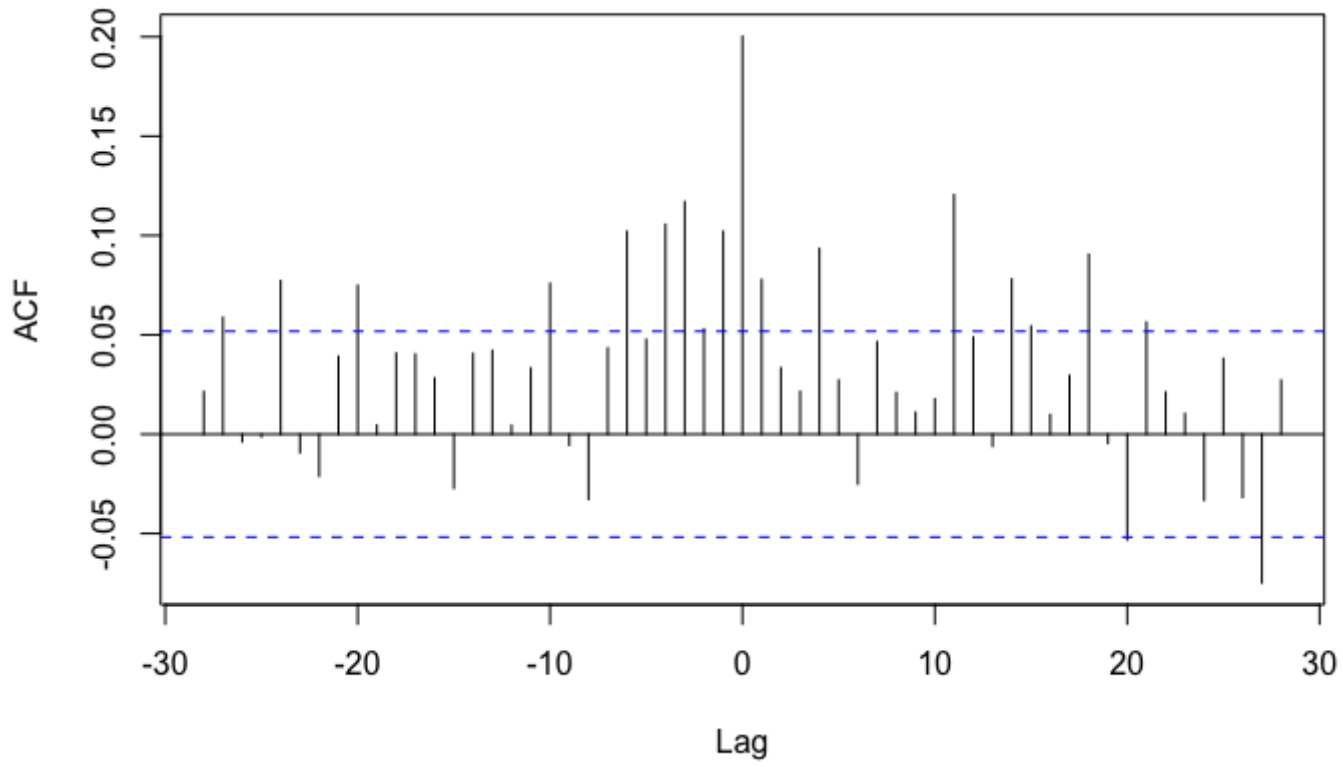
For future reading:

https://homepage.univie.ac.at/robert.kunst/prognos4.pdf
(https://homepage.univie.ac.at/robert.kunst/prognos4.pdf)

https://online.stat.psu.edu/stat510/lesson/8/8.2 (https://online.stat.psu.edu/stat510/lesson/8/8.2)

Hide

```
ccf (submetersByDay$LaundryRoom, submetersByDay$WaterHeater_AC)
```



Hide

```
ccf (submetersByDay$WaterHeater_AC,submetersByDay$LaundryRoom)
```