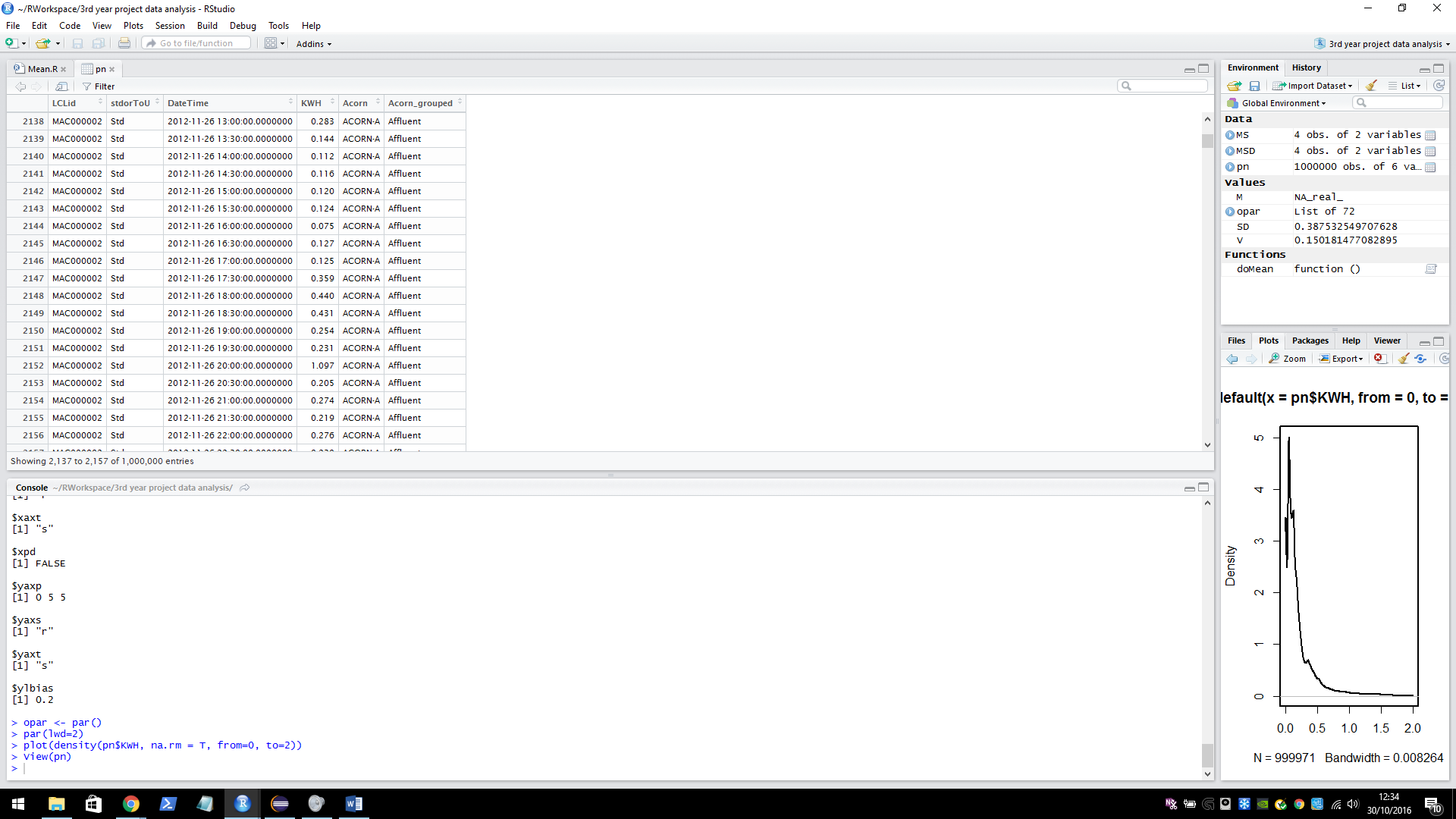
**Data analysis**

**Low-Carbon-London-Dataset**

The data set includes KWH per half hour readings for a number of households. The data set is grouped into Affluent, Comfortable, Adversity and ACORN-U depending on the customer status.

The first section from this containing 1 million entries was used in the analysis. The complete dataset contains 167 million entries. The sample below shows 8 rows of the data set with the column names.

**LCLid**: the unique house identifier.

**stdorToU**: Tariff.

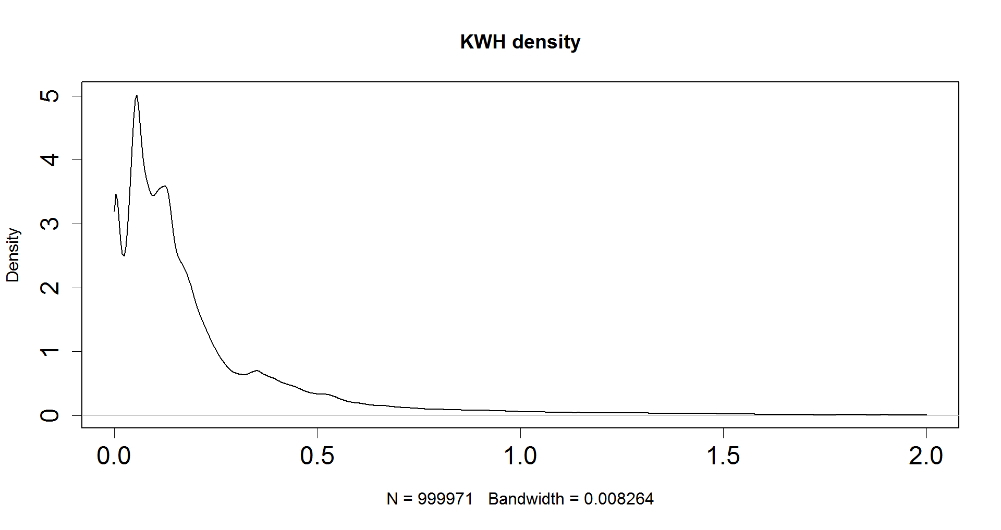
**KWH**: energy used KWH per half hour.  
**Acorn/Acorn grouped**: the grouping of the customer.

The data set can be found here <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>. Details on the ACORN groups can be found here <http://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf>.

**First analysis of data**

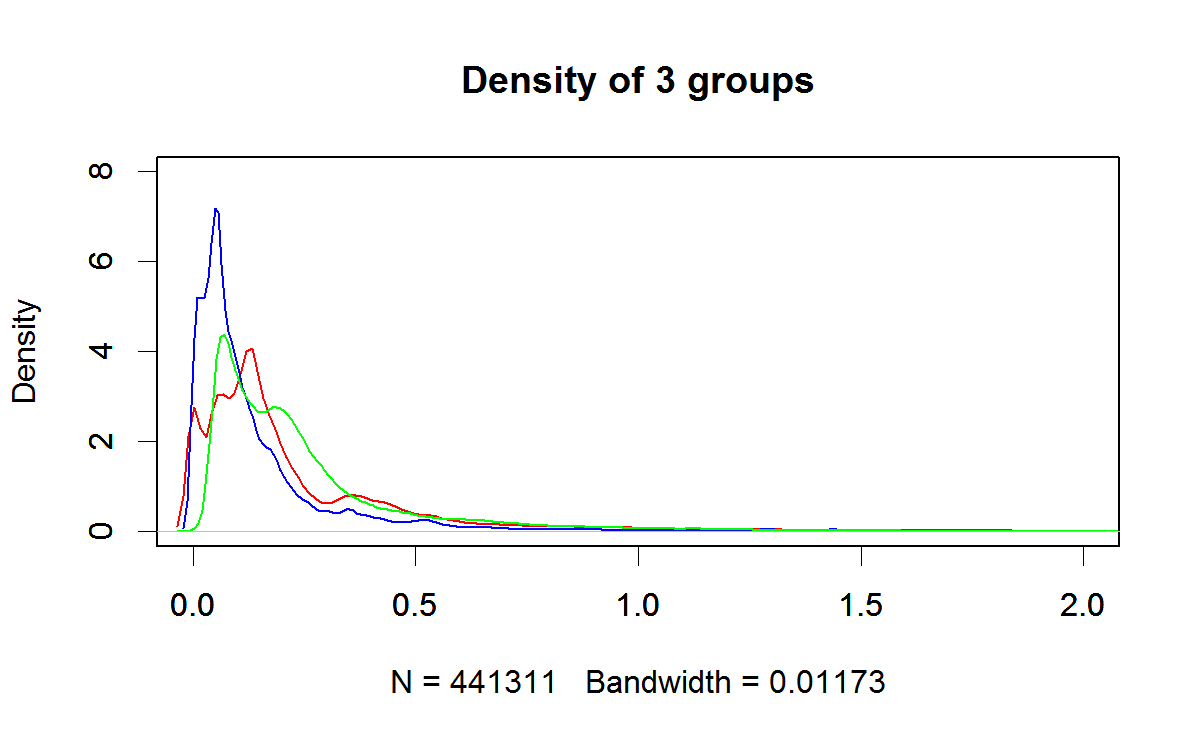
All graphs below were generated using R from the first 1 million entries.

Density Plot for all KWH in the data set.

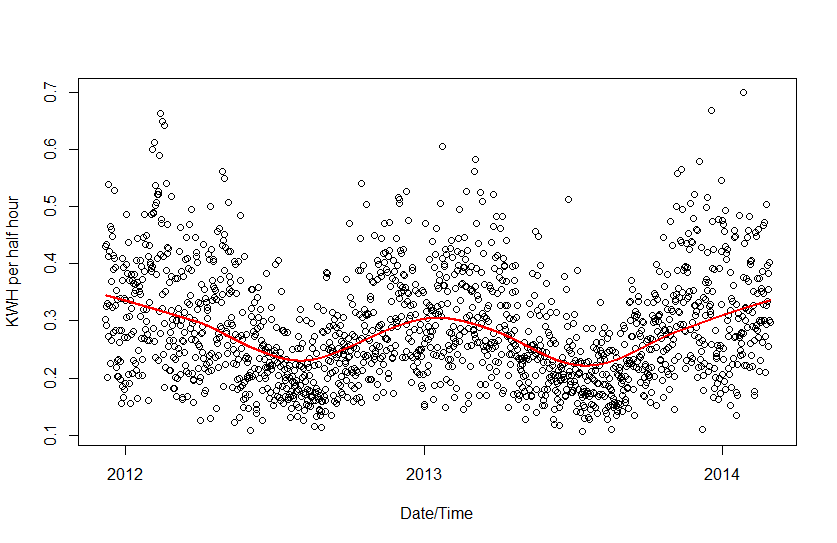


­­

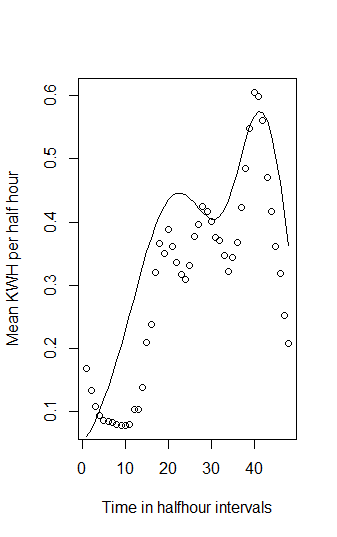
Density plot (KWH) for the three customer groups (Affluent (red), Comfortable (green), Adversity (blue)).



Plot below shows seasonality in the data. We can see that the average KWH usage is higher in the winter than in the summer. (Data was taken as the mean KWH per half hour at half day intervals)



Attempt at fitting two normal curves to the mean KWH per half hour of the ACORN-U group.



**Building a data sampling model**