# Introduction

We have been appointed as a consultant in a large building management company. The purpose of this report is to provide some insights about building electricity usage. In particular, we are interested in the daily load shapes and the corresponding consumption patterns as this knowledge has the potential to improve energy reduction recommendations. We will start by focusing our attention to one particular building, for which we have smart meter data. For this building, we will find electricity load patterns, and hypothesize about the reasons behind these load patterns. Next, we will study a set of different building, for which we have data on the itemized uses of electricity. For these buildings, we will try to understand further on the dimensionality of the electricity itemized uses, i.e. what are the possible relations between these itemized uses and which combinations of these usages account for 90 - 95% of variance in electricity loads. We will also explore what are the hypotheses that can possibly explain those principal uses, and investigate that, whether a normalization for data with same units is still necessary, or not.

# Dataset

We have been provided with the following two datasets:

– One year’s electricity consumption time series data from some building’s smart meter. The measurement time interval is 15 minutes, and the unit is kWh.

– Data set with statistics on over 5000 commercial buildings. The data could be found in elec\_end\_use.csv, and a description of the columns could be found in elec\_end\_use\_descrip.txt.

First, we will load these data files into our analysis.

Now, the one year’s electricity consumption data provided for the smart meter in our building is a one-dimensional time series data. We need to convert this into a matrix form, with each column representing a daily load so that we can explore daily load shapes and patterns. Now, to do that, we know that our values correspond to the electricity consumption every 15 minutes. So, in order to re-arrange the data so that we get one day per row, we need 4x24 (4 15 min intervals in an hour and 24 hours per day) number of columns and 35040/(4x24) number of columns as we have a total of 35040 entries in the smart meter data and each row will take up 4x24 data points.

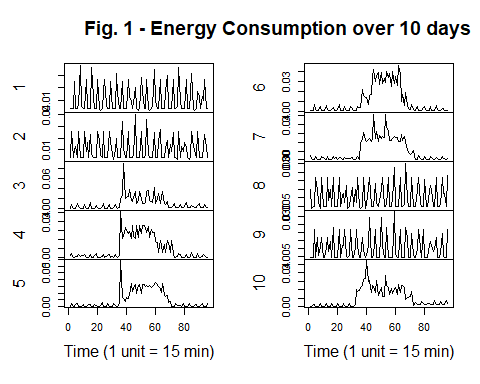
As we can now, see, our smart meter building data has 96 rows (each row representing 15 min) and 365 columns, each representing 1 day recording of electricity consumption. Now, let us look at the data provided in elec\_end\_use. Here, we can see that we have been provided data on 26 variables for over 5000 buildings. The key information provided is for building Square footage (SQFT8) and a categorization of it, year of construction category (YRCONC8), Principal building activity (PBA8), Electricity used (ELUSED8), Natural Gas Used (NGUSED8), and data providing the amount of electricity used (in thousand BTUs) for various usages such as heating, cooling use (thous Btu), ventilation, water heating, lighting, ,cooking, refrigeration, office equipment, computer use and miscellaneous.

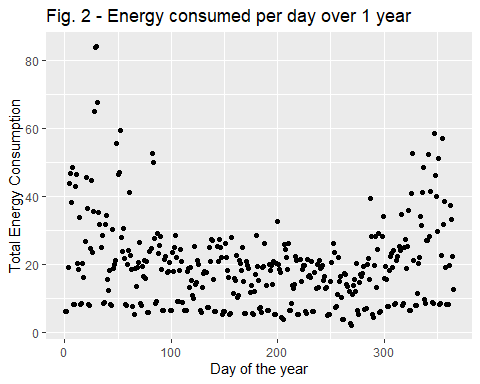
The last data adjustment that we will do is to add a row at the end of our smart meter data that shows the day of the year.

# Analysis

We will begin our analysis by plotting the daily consumption of electricity on different days in our “smart” building (the building with smart meter):

Plotting the electricity usage on day 1 of year (without normalization) As we can see, there seems to be a specific period of some days where the majority of activity happens in the building. In others, there seems to be no electricity usage spike at all. Plotting the consumption pattern across multiple days of the year should give us some insight into the energy usage patterns based on day of the week:



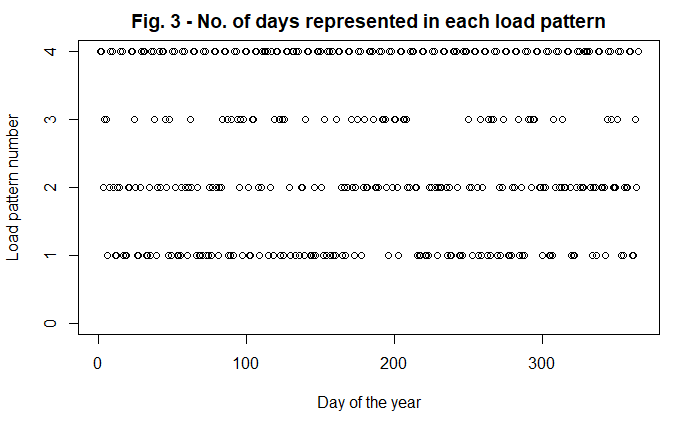
From these plots, we see that there is some cyclicity to the energy use patterns. the days 1,2,8,9,15,16… have very similar patterns with very low energy consumption. On the other hand, the days 3, 4, 5, 6 & 7 have characteristic “humps” in electricity usages. So, we can postulate that the days 1 & 2 are weekends, where no activity is done in the building, while the rest of the days represent weekdays where we have various activities in the building through 9am to 5PM. Let us also explore the data using a visualization of the total energy consumption of each day over a full year: 

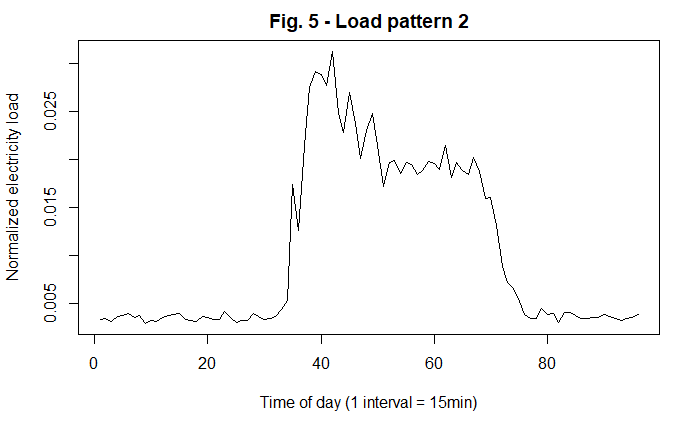
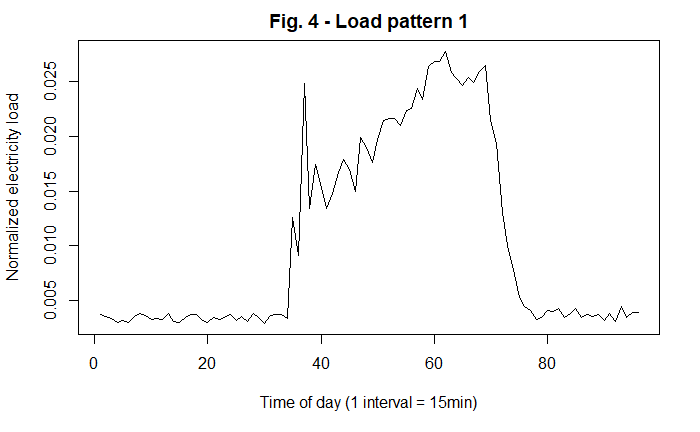
We see that the energy consumption is much higher during the wintertime as compared to the rest of the year, which tells us that this building is located in an area where heating loads are more significant than cooling loads.

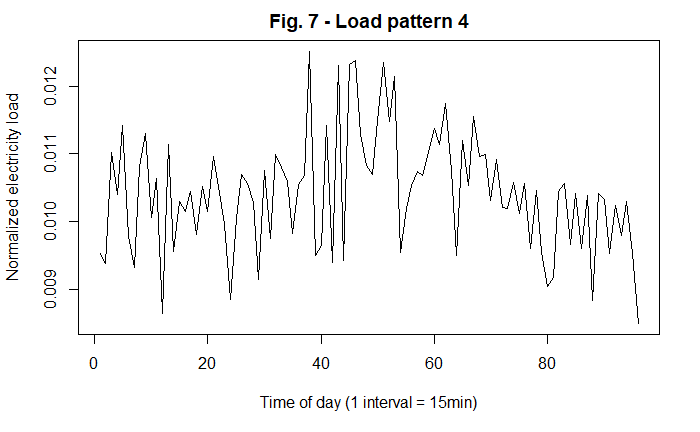
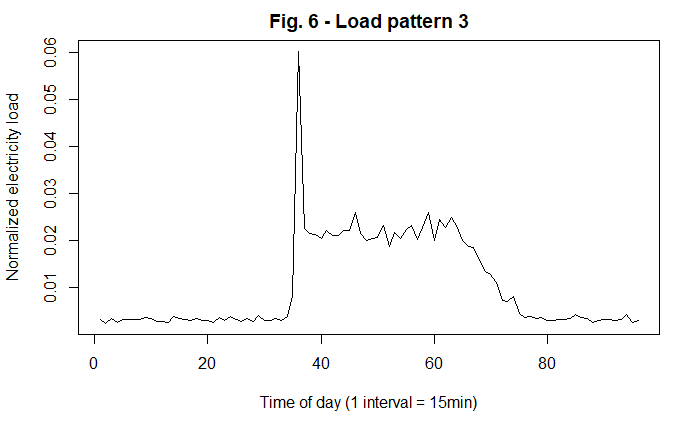
As we can see from our individual day plots (Figure 1), there is a certain pattern to the electricity usage shapes on different days with each day having a slightly different shape. We will now try to arrive at a limited number of load patterns that can explain the electricity usage of most of the days. This will help us understand the nature of electricity demand on any given day in the future and help make an informed decision about energy generation or possible integration of PVs, battery systems or EV charging stations in our building. However, it is hard for us to arrive at a few representative patterns using just our eyes. We arrive at these limited representative patterns through clustering, starting with an iteration of 7 possible clusters (Assuming that each day would have its own unique pattern):

However, we see that if we make 7 groups, we end up with a representative pattern that actually represents very small number of cases (9). Additionally, we see that three of the clusters have very similar load shapes. Additionally, we see that almost all days fall into 4 categories.

So, we iterate again with 4 clusters and we see the results: Iterating with 4 cases:



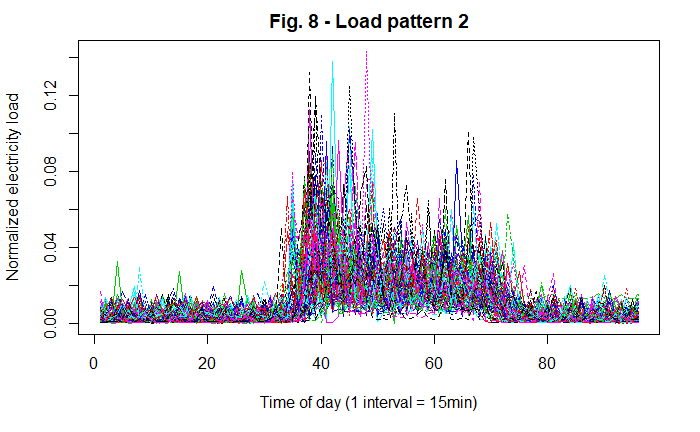


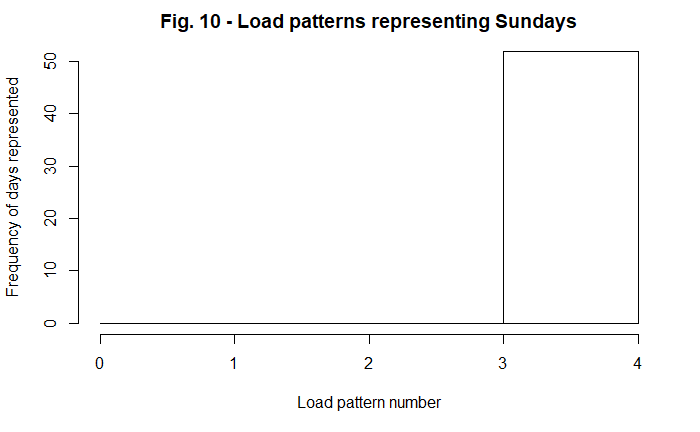
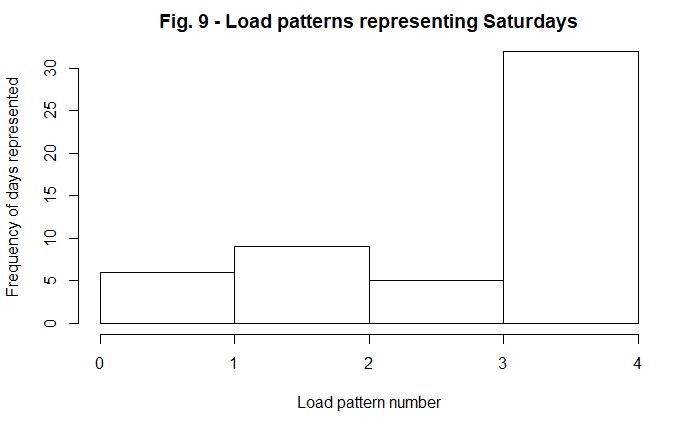


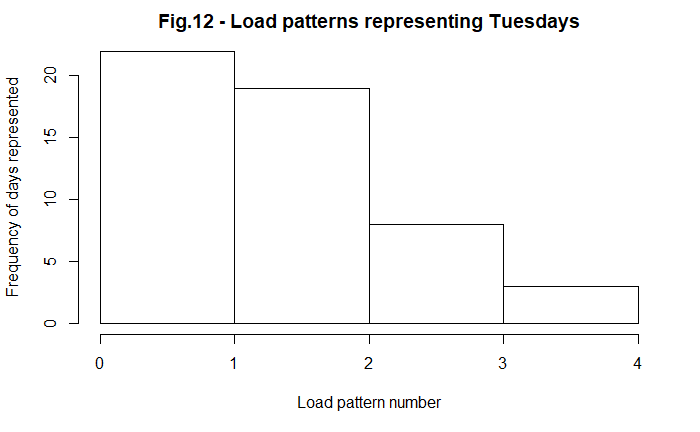
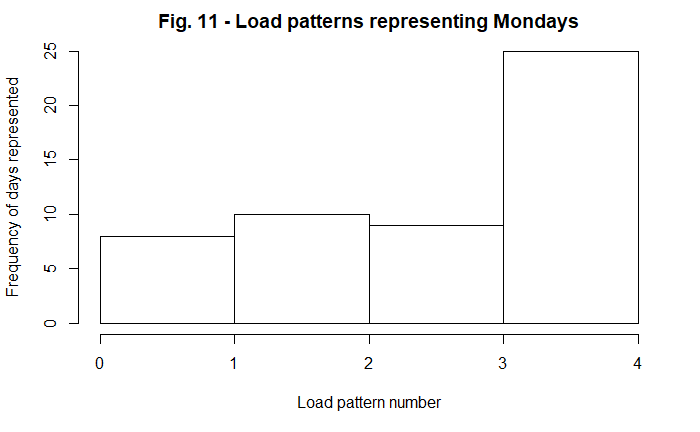
Here, we see a fairly even distribution where each load patern represents a sizeable number of days.

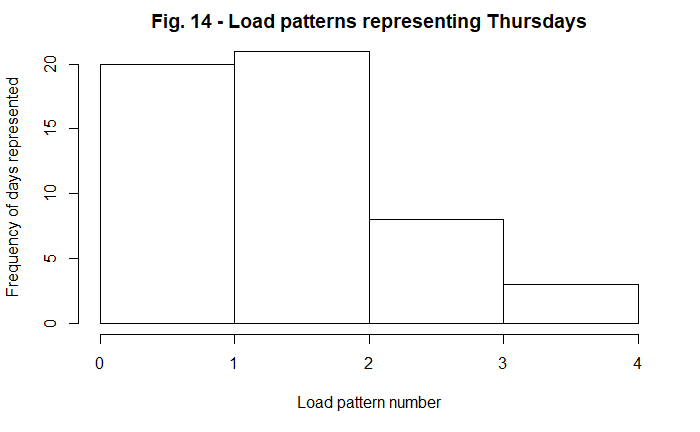
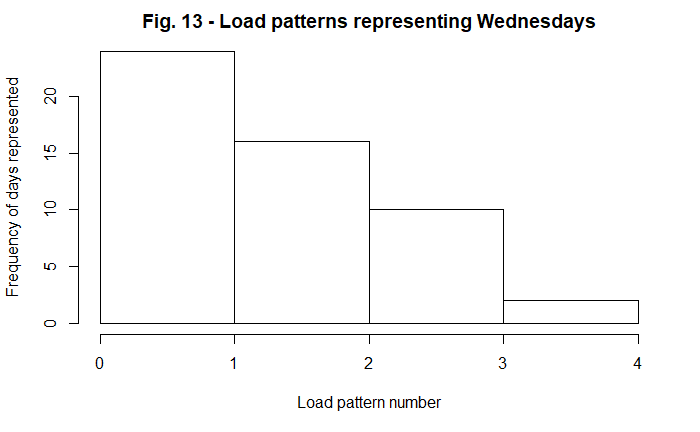
Now, we will check how well these days are represented in each of these clustered patterns:

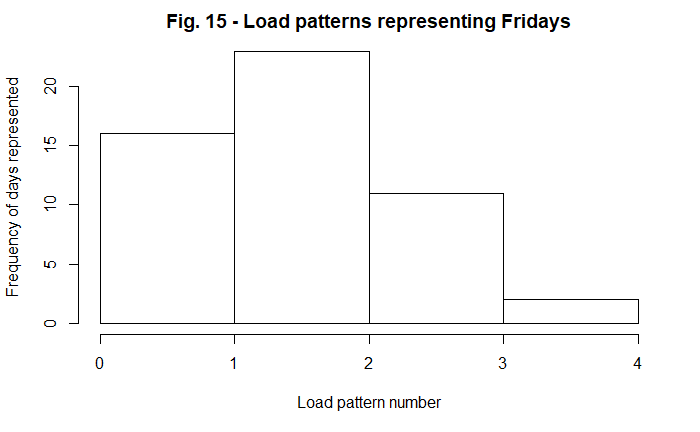
Let us see which days of the week are included in each of our representative load figures:







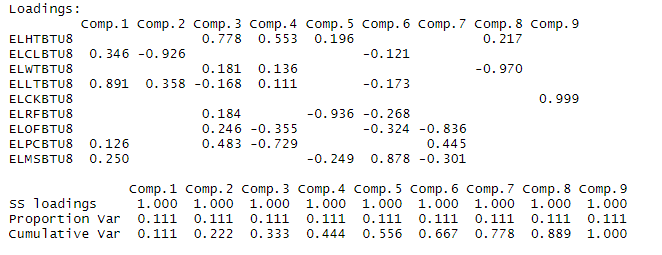




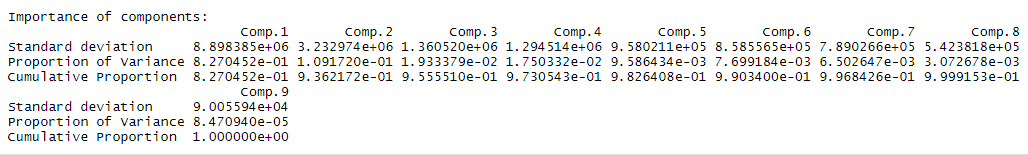
So, we can see that all Saturdays and Sundays fall into category 4. Mondays fall predominantly in category 4 with some Mondays falling in 1,2 & 3. Tuesdays and Wednesdays fall in category 1 ,2 & 3 in decreasing order. Thursdays and Fridays fall into categories 2,1 and 3 in descending order. This would indicate that this is an office building or a building which sees most of its use over weekdays with a weekly cyclical nature. The outliers could be public holidays, when the occupancy would be close to zero.

Now, we focus our attention to the second dataset. We are also very interested in the electricity consumption (in kBTU) for different end uses in a building, and in assessing the dependencies among these end uses. Therefore, we want to only focus on the data available ono heating, cooling, water heating, lighting, cooking, refrigeration, office equipment, computer, and misc. use. Therefore, we will extract the data on these end uses form our buildings dataset with the aim of finding how many loading vectors are required to explain the majority of variance in building electricity end usage, for example, 90-95% of variance.

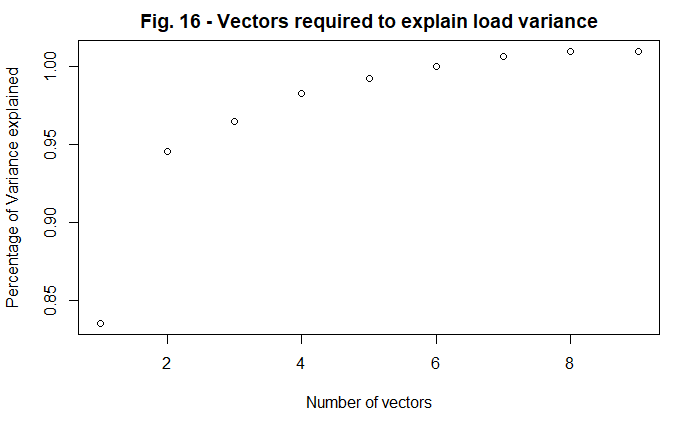
We will start by performing some exploratory data analysis on this data by looking at the plots of total electricity usage of a building and the individual usages. We see that only Cooling use, lighting use, PC use and miscellaneous usage have a clear linear trend with respect to the building total electricity usage. This shows that these usages will certainly have a significant impact on the total electricity usage of a building. Then, we try to find the principal components that effect the building electricity usage. the results are shown in the following table:



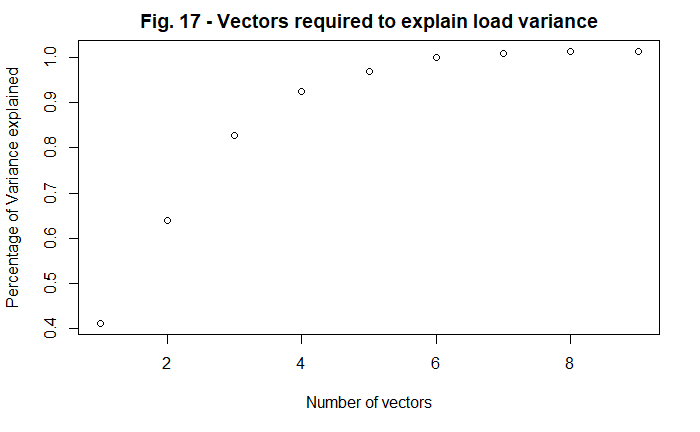
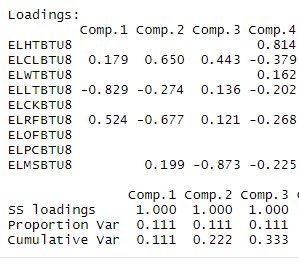
The below table shows the summary of our principal component analysis:



Based on the above, we see that 93.6% of the variance can be explained by Cooling loads, Lighting loads, PC usage and miscellaneous loads. But to explain 95% of the load, everything but the cooking load is required. Plotting this variance, we see the inflection point that showcases the decreasing impact of additional loading vectors on the variance in building electricity loads:



Now, let us repeat this analysis using normalized data.



Once again, we see that to explain 95% variance in the electricity consumption, every end-use except for cooking use is required. However, we can clearly see that the most important ones are Lighting, Cooling and refrigeration, which account for 81% of the variance.

When we were looking at this data without normalization, we saw that Cooling loads, Lighting loads, PC usage and miscellaneous loads could explain 93% of the variance while with normalization, refrigeration becomes more important and PC loads and miscellaneous reduce their impact. This is because each of the variables have a wide variety of ranges for the different buildings. Therefore, the various end-uses need to be normalized with respect to to the total building energy consumption in order to ascertain their importance in determining the total building electricity load.

# Conclusion:

Based on our analysis of the two datasets, we found four (4) different load patterns that can explain most of the daily load distribution for the year for the building with smart meter. Using these four patterns, we can better understand the energy use patterns of the building. Using this knowledge, we can now also set up the buildings Building Management System or integrate a remote / cloud-based optimization device which can track the usage of the various devices within the building and track any behavior that is outside of modelled quantities.

For the data of the 5000 buildings, we found that close to 80% of the variance in building electricity loads can be explained by Lighting, Cooling and Refrigeration. This is probably because as we have seen earlier, these buildings are in a hotter climate region and that these are buildings with high occupancy (explaining the need for lighting as well). We could use this data to perform further analysis and explore which of these end users are dominant for a building and start our energy optimization based on that. Additionally, we could also focus our attention on these end users to achieve maximum change with our efforts to make these buildings more energy efficient.