Energy Usage Analysis

An Analysis of Solar Generation Effectiveness

Student Number: 2710017

15 November 2024

Introduction (20 marks)

Objectives

This report looks at the use of solar energy during September 2024 in a domestic property in the southern highlands of Scotland. The property previously used oil for heating and electricity for lighting plus cooking but this was changed in order to reduce running costs and the carbon-footprint of the property. A solar array, battery and ground source heat pump were installed in order to attempt to rely solely on solar generated electricity and supplemented with import from the electricity grid only to compensate for short-falls.

The main aims of this data analysis is to answer the following questions:

1. How effectively does the solar generation cover energy consumption?
2. When specifically does the solar generation fall short and require grid import?
3. Would additional battery storage capacity cover these shortfalls?
4. Does the installation meet the expectations at the time of purchase?

Summary of The Data

The data was collated from three original sources and combined into a single .txt file, then imported as an R dataframe.

```r
# Import all data from the tab-separated data file which is held in the data sub-folder
file_path <- './Data/Energy_September_2024.txt'
energy_df <- read.delim((file_path))
# Convert the string date to a valid date format
energy_df$Date <- as.Date(energy_df$Date, '%d/%m/%Y')
```

The data analysed comprises four parts, all daily data, 30 observations, for each day in September 2024:

- Weather: Temperature and solar irradiance readings
- Energy Use: Electricity consumption
- Energy Source: The source of electricity: solar, battery or import from the grid
- *Occupied*: The approximate number of hours the house is occupied each day

All data and supporting files can be found online at Github[1].

[1] https://github.com/StuartG24/Home-Solar-Usage-Analysis

Weather

Weather data is sourced from the Balquhidder Weather Station[2] and consists of:

- *Temp* - the mean daily temperature in °C and is derived from 6 readings taken at 4 hourly intervals over a 24 hour period
- *Irrdnce* - irradiance, a measure of the solar energy experienced over a specified area, units are kW/m^2 or W/m^2 and this is used to calculate the theoretical power generated from an array of solar panels[3]

Energy Use & Source

The distribution of power for the house is managed by a Tesla Powerwall and Controller and an iPhone app is used to monitor this, see Figure 1. All electricity data was downloaded via this app.

Electricity used and where it is sourced from:

- *Home_Total* - total energy used by the house, in kWh or Wh
- *From_Solar* - solar power generated by an array of 36 solar panels
- *From_PWall* - battery storage
- *From_Grid* - the national power grid

Electricity generated by the solar panels and where it is used (the controller intelligently makes the routing decisions):

- *Solar_Total* - total energy generated by the solar panels
- *To_Home* - consumption by the house
- *To_PWall* - for battery storage
- *To_Grid* - export to the national power grid

TO DO: *Tidy up table columns display .... look at Pandoc options etc to change layout for table and whole document? https://pandoc.org/MANUAL.html#synopsis*

Once the sources of data have been collated and loaded the dataset consists of 30 observations and 12 columns. The first 6 rows are shown below:

```r
# Display the first 6 rows of the data
#head(energy_df)
kable(head(energy_df), caption = 'First 6 Rows of the source dataset')
```

Table 1: First 6 Rows of the source dataset

Date	Home_Total	From_Pwall	From_Solar	From_Grid	Solar_Total	To_Home	To_Pwall	To_Grid	Temp	Irrdnce	Occupied
2024-09-01	19048	11652	2180	5217	2360	2180	172	9	12.9	73.18	24
2024-09-02	11304	5224	1171	4909	1232	1171	50	11	13.4	67.08	6
2024-09-03	13867	7062	3690	3115	6372	3690	2664	18	9.9	269.36	0

[2] https://www.blscc.org/weather
[3] Wikipedia: https://en.wikipedia.org/wiki/Solar_irradiance

Date	Home_Total	From_Pwall	From_Solar	From_Grid	Solar_Total	To_Home	To_Pwall	To_Grid	Temp	Irrdnce	Occupied
2024-09-04	16241	9315	4380	2546	6768	4380	2378	10	9.5	344.50	0
2024-09-05	17960	9127	5744	3090	11036	5744	5204	88	14.5	336.36	12
2024-09-06	16015	10998	4617	400	14726	4617	9152	957	16.6	297.88	24

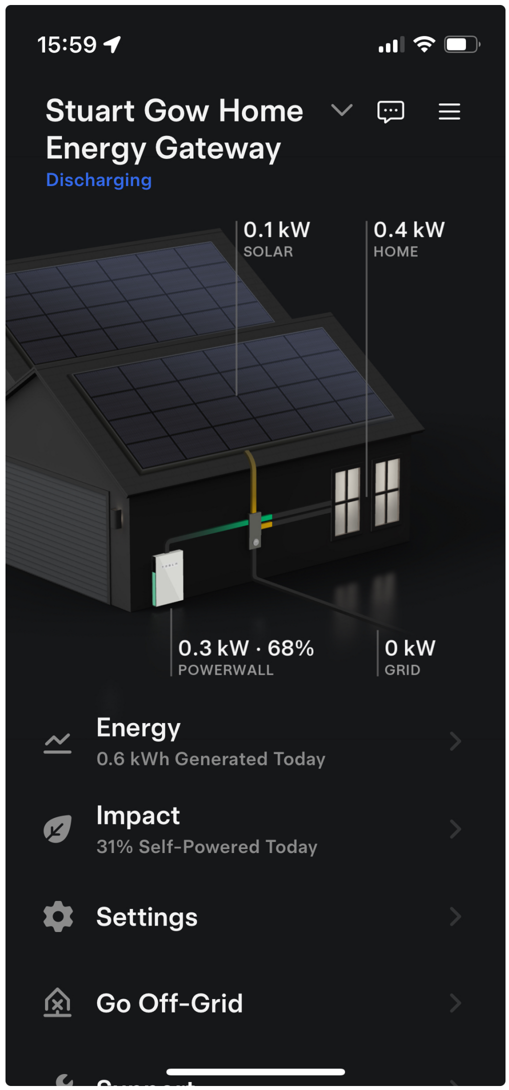


Figure 1: Tesla Powerwall

Methods and Results - (40 marks)

The data was analysed in several groups and themes, in summary:

- House Consumption & Solar Energy Sufficiency

- Solar Energy Generation Effectivness
- xxx

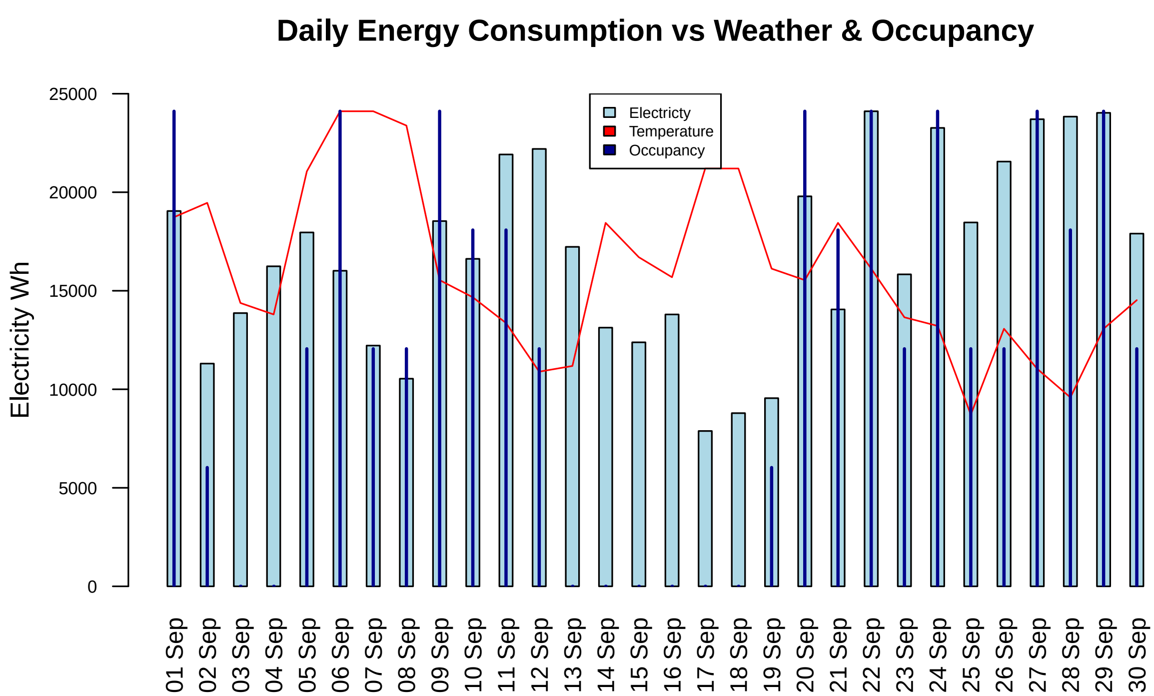
WIP - Analysis, expected conclusions:

1) Consumption is covered by solar .. or not? how much? how much grid still needed
2) Consumption is linked to temperature and house occupancy
3) Solar generation is linked to irradiance .. but how much?
4) Any other links such as temperature? .. probably not usage or occupancy though?
5) ?? Cannot account for £ cost and different costs at times of day .. battery importing then for example
6) ?? Battery timing in and out complicates the analysis?
7) Increased battery will smooth out across days? forecast storage/impact .. but can't see the intra-day detail to better analyse
8) Solar generation meets the forecasts at purchase .. need original data/estimates!?

House Consumption & Solar Energy Effectiveness

Energy Demand

The energy consumption of the house can be compared to the weather and its occupancy and the figure below shows this (NB: the temperature and occupancy values have been scaled to only show the relationships and so no values are displayed). At first glance, there does not appear to be a strong link between the energy demand and the outside temperature but potentially there is with occupancy.



The first bar plot below shows the total energy consumed per day by the house and a breakdown of where this energy is sourced from. It appears that the energy generated by the solar panels only meets a small proportion of the total daily consumption, on average 26%.

```
average_solar_percent <- mean(c(energy_df$From_Solar / energy_df$Home_Total)) * 100
print(paste("Straight percentage:", average_solar_percent))
```

```
## [1] "Straight percentage: 26.0395444324628"
```

However, this is misleading as generated solar energy is often first stored in the battery for later use or exported to the gird if the battery becomes full. The second bar plot shows this more clearly. With solar energy distributed to the home, the battery or exported to the grid on average: 62%, 30% and 8% respectively.

```
solar_home_percent <- mean(c(energy_df$To_Home / energy_df$Solar_Total)) * 100
solar_powerwall_percent <- mean(c(energy_df$To_Pwall / energy_df$Solar_Total)) * 100
solar_grid_percent <- mean(c(energy_df$To_Grid / energy_df$Solar_Total)) * 100
print(paste("To Home:", solar_home_percent, "To Powerwall:", solar_powerwall_percent,
            "To Grid:", solar_grid_percent))
```

```
## [1] "To Home: 62.0971977996547 To Powerwall: 29.9891405679381 To Grid: 7.91530834160956"
```

A good measure of how well the solar generation meets the needs of the house's energy consumption is to simply look at how much usage is catered for without any import from the grid, and this is 80% on average.

```
adjusted_solar_percent <- mean(c((energy_df$Home_Total - energy_df$From_Grid) /
                                    energy_df$Home_Total)) * 100
print(paste("Without Grid Imported:", adjusted_solar_percent))
```

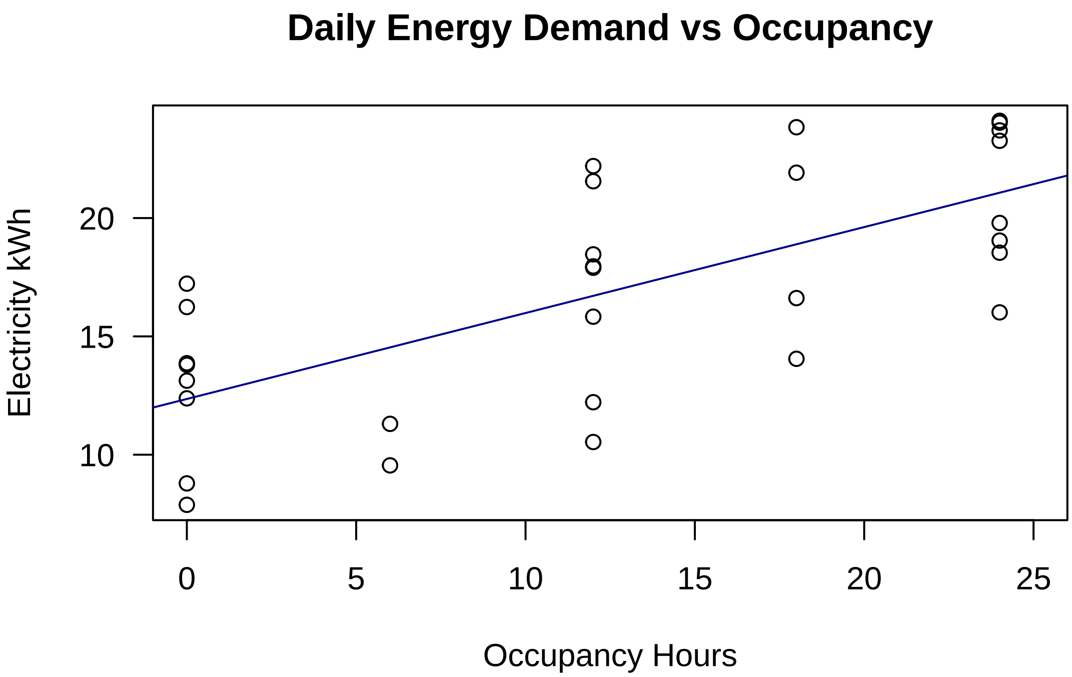
```
## [1] "Without Grid Imported: 79.9652308577059"
```

It is expected that the energy demand from the house should be related to the occupancy and also the temperature. So three linear regressions were carried out. The null hypothesis for these are that energy demand is not linked to occupancy or to temperature.

First looking at the relationship between energy demand and occupancy.

```
# Plots and linear regression for energy and occupancy
regression_model <- lm(energy_df$Home_Total/1000 ~ energy_df$Occupied)
regression_summary <- summary((regression_model))
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta <- regression_summary$coefficients["energy_df$Occupied", "Estimate"]
p_value <- regression_summary$coefficients["energy_df$Occupied", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared

plot(energy_df$Home_Total/1000 ~ energy_df$Occupied, main="Daily Energy Demand vs Occupancy",
     xlab="Occupancy Hours", ylab="Electricity kWh", las=1, xlim=c(0,25))
abline(regression_model, col="darkblue")
```



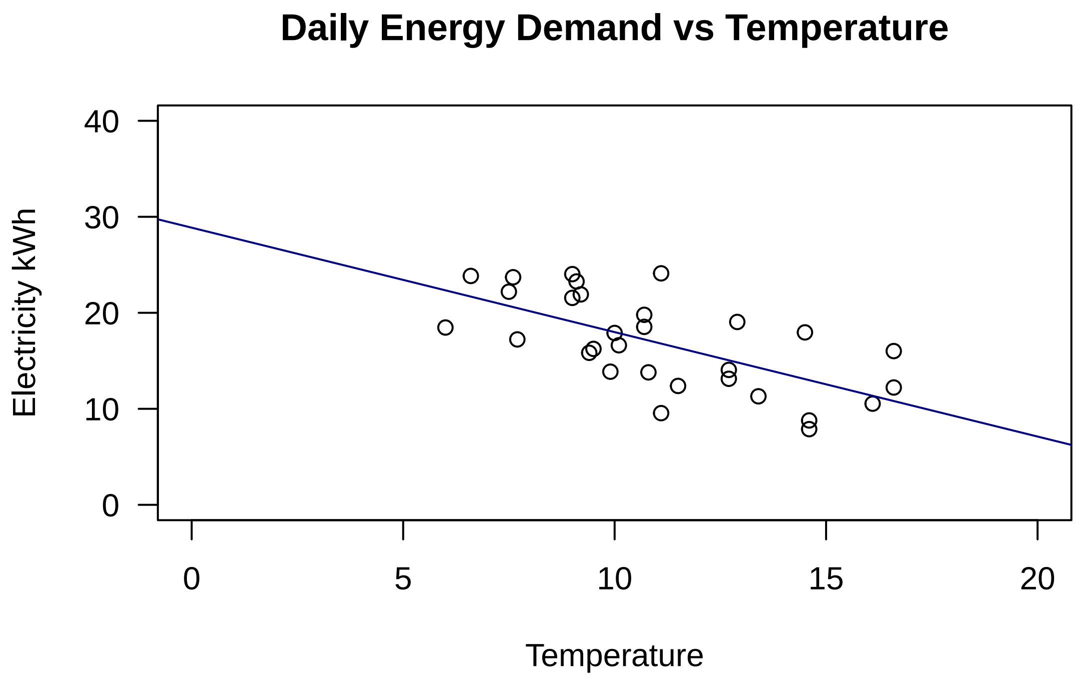

```
print(sprintf("Alpha: %.3f, Beta: %.3f", alpha, beta))
```

```
## [1] "Alpha: 12.356, Beta: 0.363"
```

```
print(sprintf("p-value: %.4f, Adj R-Squared: %.3f", p_value, adj_r_squared))
```

```
## [1] "p-value: 0.0000, Adj R-Squared: 0.461"
```

Then looking at the relationship between energy demand and temperature (NB: same R Code so not printed to save space).




```
## [1] "Alpha: 28.868, Beta: -1.088"
```

```
## [1] "p-value: 0.0001, Adj R-Squared: 0.397"
```

The regression analyses above do indicate a link between energy demand and occupancy and also inversely with temperature, however the adjusted r-squared values are relatively low and so the correlations are not strong. So a a multi-linear regression was completed to look at the relationship between energy demand and occupancy plus temperature (NB: A three dimensional scater plot was not attempted). The p-value for this is near zero and below the 5% critical value, it is statistically significant and we can reject the null hypothesis. From this we can infer that energy demand is linked to occupancy and temperature. In addition the high adjusted r-squared value of 0.8 suggests that there is a strong correlation between energy demand and with the combined occupancy and temperature.

!! TO DO: For conclusions section ... temperature not massive due to good insulation and also not extremely cold etc ... but intercept does indicate as it gets colder then increases

```
## [1] "Alpha: 23.448, Beta Occupancy: 0.328, Beta Temp: -0.965"
```

```
## [1] "p-value Occp: 0.0000, p-value Temp: 0.0000, Adj R-Squared: 0.789"
```

Solar Energy Generation Effectiveness

The amount of energy generated by solar panels is a function of their size and the level of sunshine received, measured by irradiance. There is also a loss factor that reflects several things including the panel's efficiency and the invertor's efficiency. Additionally, irradiance observations are taken from local weather stations which may not experience the same shading from the sun as that experienced at the site of the solar panels.

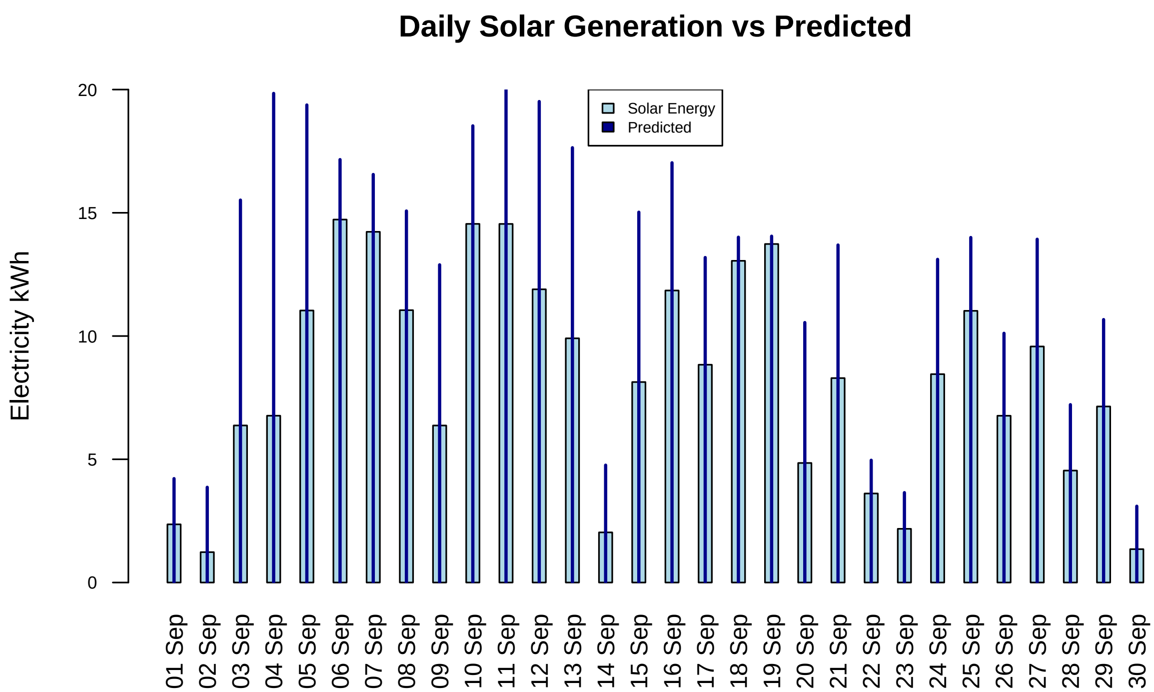
$$Power(kWh) = Area(m^2) * Irradiance(kWh/m^2) * LossFactor$$

At installation, the annual generation power for this solar array was calculated as 7,920 kWh pa using an area of 72 m^2, irradiance 137.5 kWh/m^2 pa and a loss factor of 0.8. Using the September observations of irradiance and generated solar energy, the effectiveness of the installation can be compared to the estimates (sales promises) made originally. It is very likely that the sales estimates were optimistic.

It is expected that the generated solar energy is related to the irradiance and a quick scatter plot and linear regression confirmed this, with a p-value < 0.005 and a high adjusted r-squared 0.71.

T-Test to Evaluate Effectiveness

A prediction of the daily solar energy was calculated using the formulae above with the actual irradiance observations. The actual and predicted solar energy daily totals are compared in the plot below and the means for the month are 8.35 kWh and 12.8 kWh respectively.



The null hypothesis is that the installation performs as well as the sales promises; specifically that the daily energy production in September μ_{act} is the same as that expected μ_{exp}.

$$H_0 : \mu_{act} = \mu_{exp}$$

the alternative hypothesis is that the installation does not perform as predicted

$$H_1 : \mu_{act} < \mu_{exp}$$

A T-Test was performed to evaluate this and calculate the p-value which is the probability, if H_0 is true, of obtaining the observation, or an observation more extreme. In this case more extreme is that the actual mean observed is less than the predicted mean because we expect the installation does not to perform as well as the sales estimates. A one-tailed test was used.

```r
# Evaluate a T-Test
t_test_result <- t.test(energy_df$Solar_Total/1000, solar_predicted, alternative = "less")
print(sprintf("p-value: %.4f", t_test_result$p.value))
```

```
## [1] "p-value: 0.0004"
```

The p-value is < 0.005 which is statistically significant and the null hypothesis can be rejected at the 5% level. From this we can infer that the installation is not generating solar energy as effectively as predicted.

xxxxxxxxxxxxxxxxx

Predict annual generation using non September

?? monthly irradiance data: https://re.jrc.ec.europa.eu/pvg_tools/en/#MR

Conclusions - 20 marks

##?? Discussion

TO DO: Structure into two sections

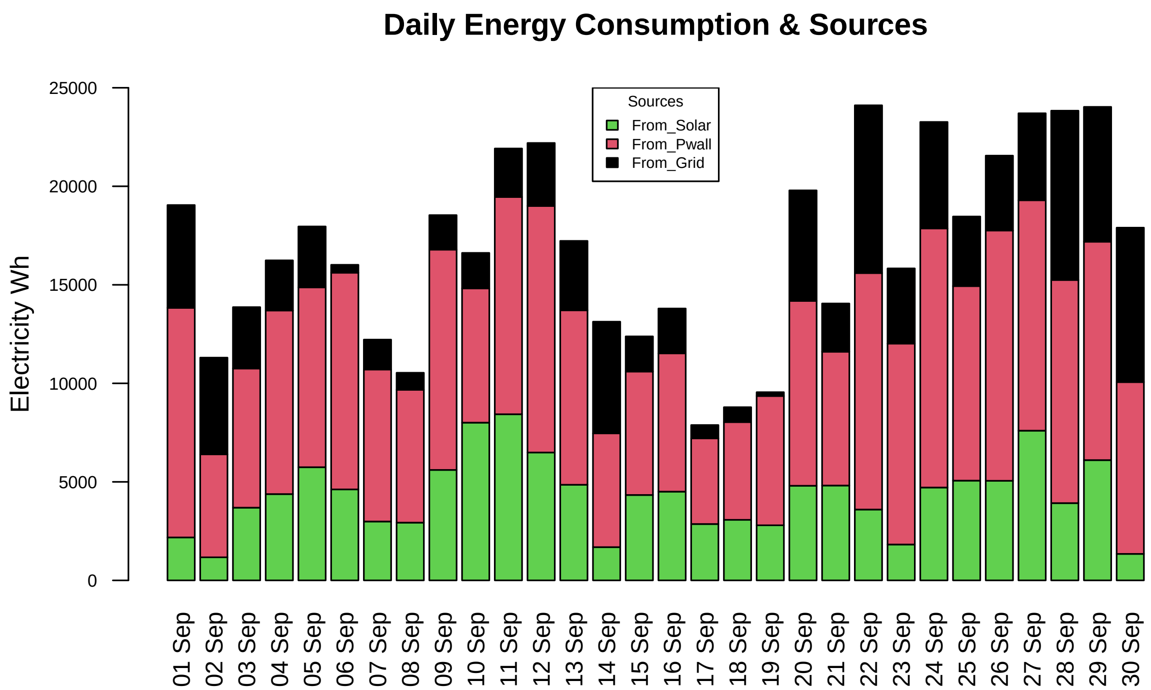
##?? Conclusion

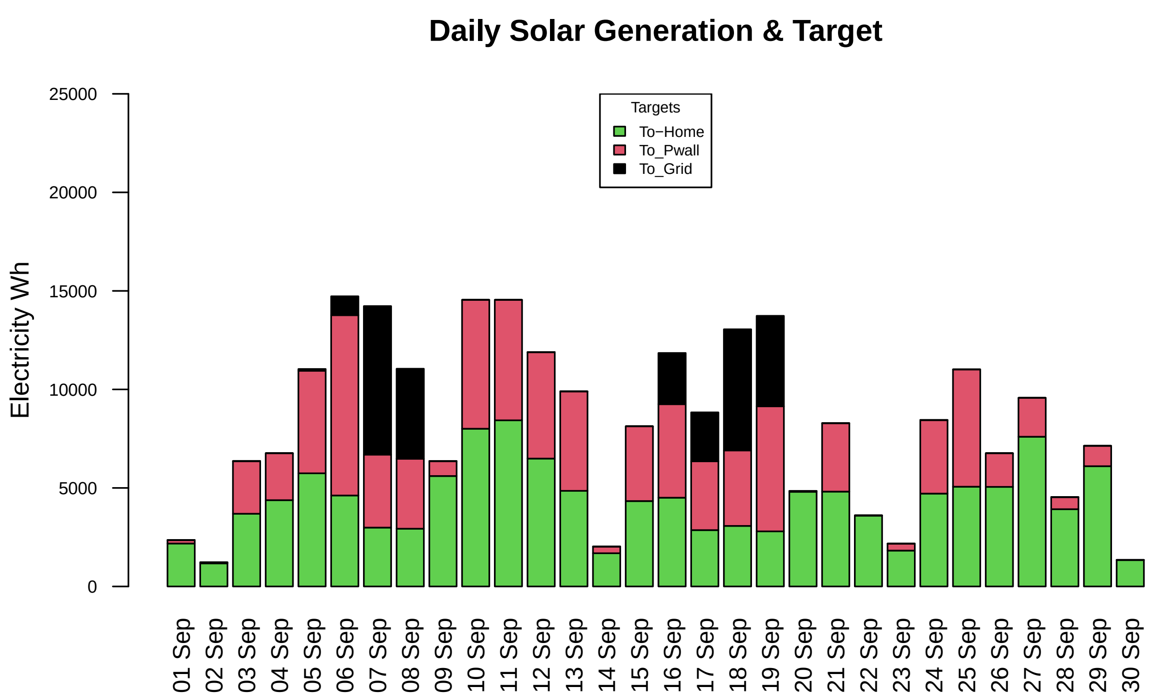
Test citations (Crawley, 2014) and as Fraix-Burnet (2016)

Spiegel and Schiller (2012)

Appendices

?? Energy Consumption ....





References

Crawley, M.J. (2014) *Statistics: An introduction using R*. 2nd Edition. John Wiley & Sons.

Fraix-Burnet, D. (2016) ‘Introduction to R’, *Statistics for Astrophysics: Clustering and Classification*, Volume 77(2016), pp. 3–12. Available at: https://doi.org/10.1051/eas/1677002.

Spiegel, M.R. and Schiller, J. (2012) *Schaum’s outline probability and statistics*. 4th edn. McGraw Hill.