Energy Usage Analysis

An Analysis of Solar Generation Effectiveness

Student Number: 2710017

25 November 2024

Introduction

Scope & Objectives

About six years ago, I decided to remove my reliance on oil and to reduce my use of electricity imported from the National Grid. My aim was to try to reduce my energy costs and also to make a small contribution towards reducing the use of fossil fuels. So I obtained installation costs and received estimates of how much energy I could generate by using solar panels. The existing setup of oil powered central heating was replaced by the installation of a solar panel array, battery storage and ground source heat pump. I hoped to be able to only import from the grid when solar generation and storage failed to meet the demands of the house.

The main objective of this anaysis is to see if the installation has met the predictions made during the sales process and to what extent I no longer have to rely on imported electricity, or, in other words, how effective the overall installation is. I used a sample of data for the month of September 2024 as a hopefully representative month (ie not the height of summer with peak solar generation and not winter with low solar generation and peak consumption).

In the final report, R code and outputs have been shown to demonstrate working, however to reduce the size of the document this is suppressed if repeating previous workings. All coding details are included in the submitted .rmd file.

Summary of The Data

The data used in this analysis was collated from three sources and combined into a single .txt file, then imported into an R dataframe.

```
# 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, with 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].

Weather

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

- *Temp* - the mean daily temperature in °C. 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. The imported electricity data was in Wh but is usually reported in kWh.

Electricity used and where it is sourced from:

- *Home_Total* - total energy used by the house
- *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

R Dataset

The sources of data were collated and loaded into an R dataset consisting of 30 observations and 12 columns, the first 4 rows of which are shown below:

```
# Display the first 4 rows of the data
head(energy_df,4)
```

```
##         Date Home_Total From_Pwall From_Solar From_Grid Solar_Total To_Home
## 1 2024-09-01      19048      11652       2180      5217        2360    2180
## 2 2024-09-02      11304       5224       1171      4909        1232    1171
## 3 2024-09-03      13867       7062       3690      3115        6372    3690
## 4 2024-09-04      16241       9315       4380      2546        6768    4380
##   To_Pwall To_Grid Temp Irrdnce Occupied
## 1      172       9 12.9   73.18       24
## 2       50      11 13.4   67.08        6
## 3     2664      18  9.9  269.36        0
## 4     2378      10  9.5  344.50        0
```

[1] https://github.com/StuartG24/Home-Solar-Usage-Analysis
[2] https://www.blscc.org/weather
[3] Wikipedia: https://en.wikipedia.org/wiki/Solar_irradiance

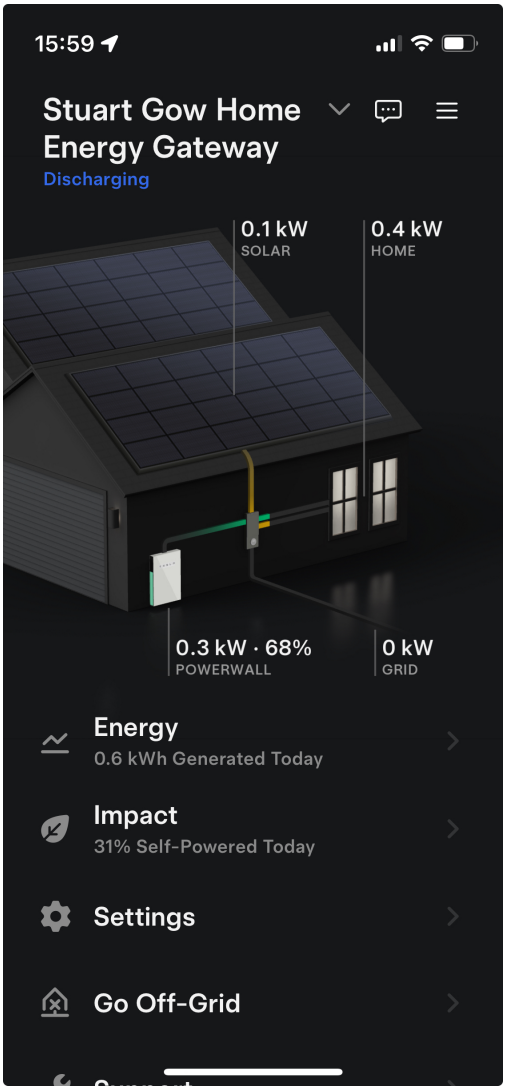


Figure 1: Tesla Powerwall

Methods and Results

The data was analysed in three themes and each of these are described in the following sections. In summary:

- Energy Consumption - What drives energy consumption?
- Solar Generation Effectiveness - How well does the installation meet sales promises?
- Solar Energy Sufficiency - How well does solar generation meet the energy demand?

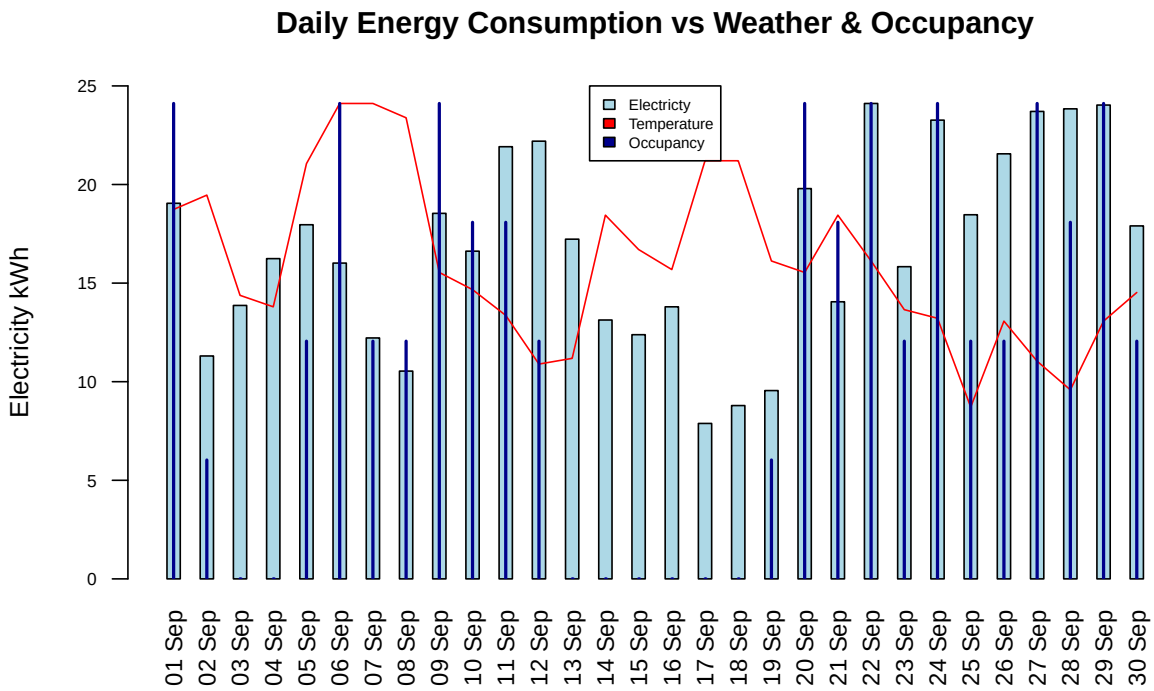
Energy Consumption

Sanity Check

The energy consumption (electricity in kWh) of the house was compared to the weather and to occupancy and the figure below summarises this (nb: temperature and occupancy have been scaled to only show the relative size and so no values are displayed).

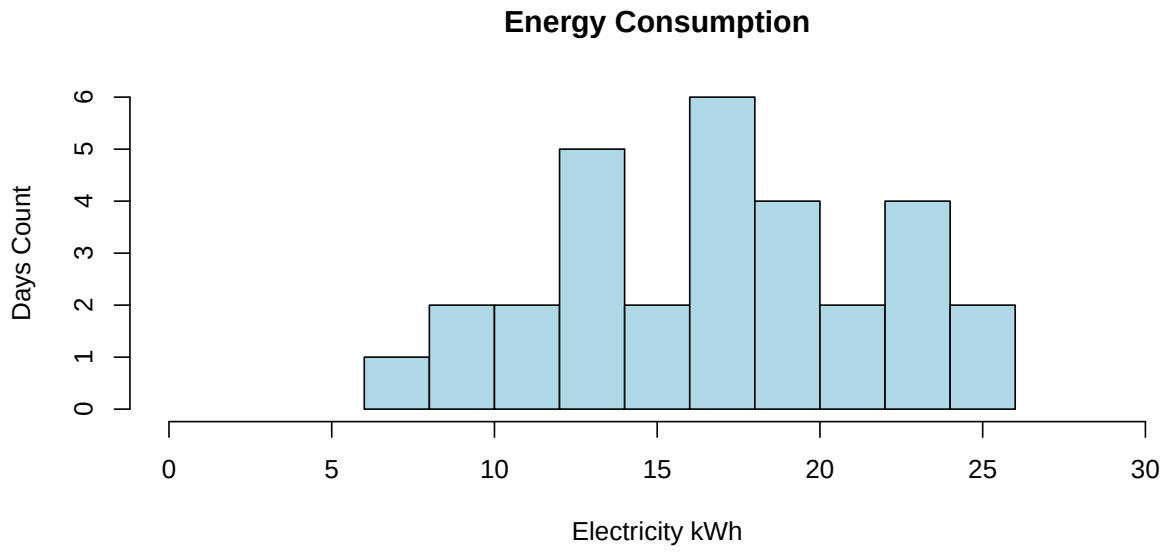
```
# Base barplot of energy consumption
barplot_result <- barplot(energy_df$Home_Total/1000, names.arg = format(energy_df$Date, "%d %b"),
        cex.name=0.9, las=2, cex.axis=0.7, main="Daily Energy Consumption vs Weather & Occupancy",
        ylab="Electricity kWh", ylim=c(0,25), col = 'lightblue', space = 1.5)

# Add scaled lines for temperature and occupancy
scaled_temp <- energy_df$Temp * max(energy_df$Home_Total/1000) / max(energy_df$Temp)
lines(barplot_result, scaled_temp, type = 'l', col = "red")
scaled_occ <- energy_df$Occupied * max(energy_df$Home_Total/1000) / max(energy_df$Occupied)
lines(barplot_result, scaled_occ, type = 'h', col = "darkblue", lwd = 2)
legend('top', legend=c("Electricty", "Temperature", "Occupancy"),
       fill=c("lightblue", "red", "darkblue"), cex = 0.6)
```

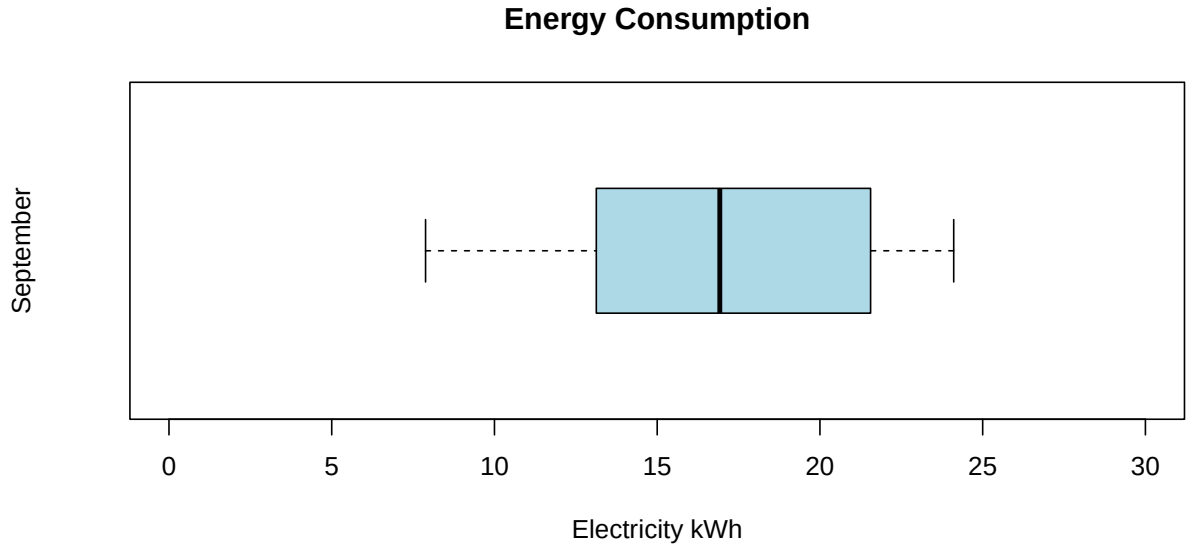


A quick examination of the energy demand was completed to visualise the distribution and identify any outliers.

```
# The distribution of energy consumption in September
hist(energy_df$Home_Total/1000, main="Energy Consumption", xlab= "Electricity kWh", ylab="Days Count",
         col = 'lightblue', xlim=c(0,30))
```




```
# Box plot of energy consumption in September
boxplot(energy_df$Home_Total/1000, main="Energy Consumption", xlab= "Electricity kWh", ylab="September"
         col = 'lightblue', ylim=c(0,30), horizontal = TRUE)
```

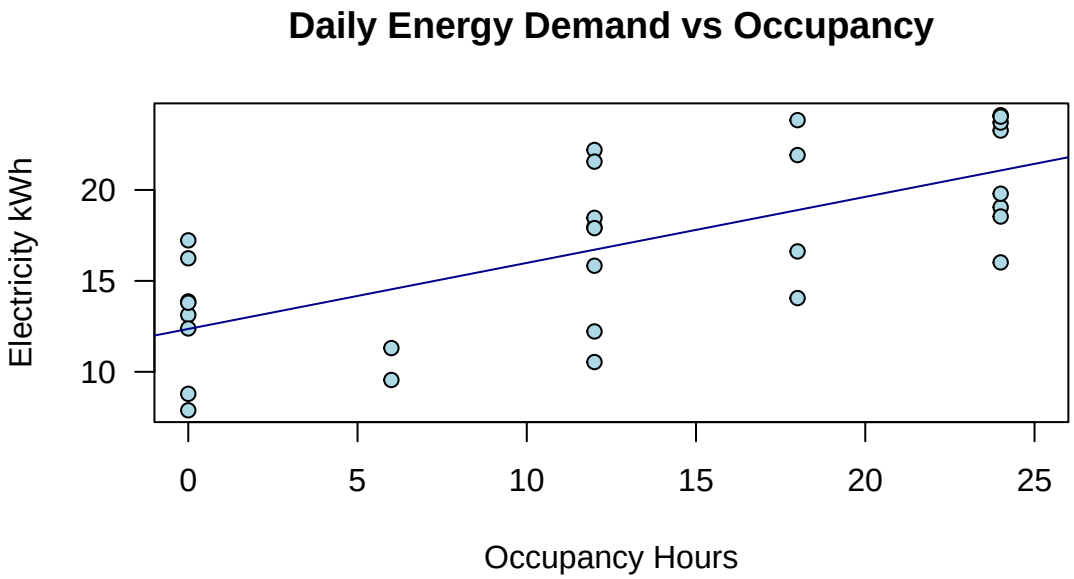


Relationships Between Demand and Other Factors

It was expected that the energy demand from the house would be related to the occupancy and the temperature, however visually there does not appear to be a strong link with temperature but potentially there is a link with occupancy. So three linear regressions were carried out to better identify any relationships.

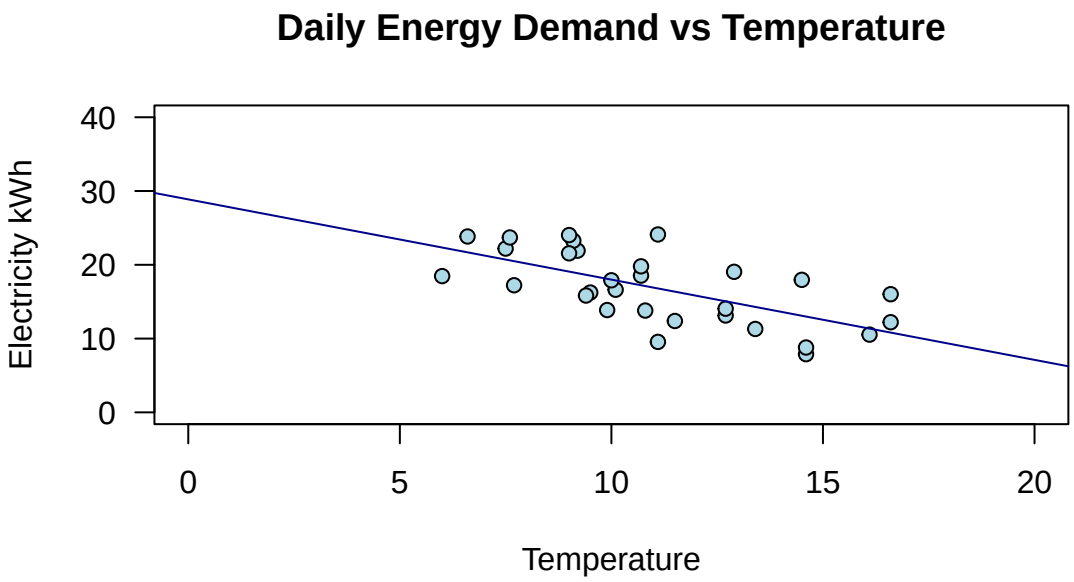
First looking at the relationship between energy consumption and occupancy.

```
# Calculate 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
# Scatter plot with fitted regression line
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),
     pch=21, bg="lightblue")
abline(regression_model, col="darkblue")
```



The scatter plot shows the calculated regression line as energy = 12.36 (α) + 0.36 (β) * occupancy. This fit had a p-value of 2.24e-05 and adjusted R-squared of 0.46. For a linear regression the null hypothesis is that β is zero, ie $H_0 : \beta = 0$. Here the p-value is less than the 5% critical value and so we can reject H_0 and infer that energy and occupancy are correlated, however the relative low value of adjusted R-squared suggests that the correlation is not very strong.

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



The regression line determined was energy = 28.87 (α) + -1.09 (β) * temperature. This fit had a p-value of 0.00011 and adjusted R-squared of 0.4. Here the p-value is less than the 5% critical value and so we can reject H_0 and infer that energy and temperature are inversely correlated, however the relative low value of adjusted R-squared suggests that the correlation is not very strong.

Then a multi-linear regression was performed looking at the relationship between energy demand and temperature plus occupancy (NB: A three dimensional scatter plot was not attempted).

```
# Multiple linear regression for energy and occupancy + temperature
regression_model <- lm(energy_df$Home_Total/1000 ~ energy_df$Temp + energy_df$Occupied)
regression_summary <- summary((regression_model))
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta0 <- regression_summary$coefficients["energy_df$Occupied", "Estimate"]
p_value0 <- regression_summary$coefficients["energy_df$Occupied", "Pr(>|t|)"]
beta1 <- regression_summary$coefficients["energy_df$Temp", "Estimate"]
p_value1 <- regression_summary$coefficients["energy_df$Temp", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared
#print(sprintf("Alpha: %.3f, Beta Occupancy: %.3f, Beta Temp: %.3f", alpha, beta0, beta1))
#print(sprintf("p-value Occp: %.4f, p-value Temp: %.4f, Adj R-Squared: %.3f", p_value0, p_value1, adj_r
```

The regression line determined was energy = 23.45 (α) + 0.33 (β) * occupancy + -0.97 (β) * temperature This fit had a p-value of 7.59e-08 for occupancy and 3.5e-07 for temperature and adjusted R-squared of 0.79. Here both p-values are less than the 5% critical value and so we can reject H_0 and infer that energy and occupancy plus temperature (inversely) are correlated, the high value of adjusted R-squared suggests that the correlation is quite strong.

Solar Generation Effectiveness

Background

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 efficiency of the

solar panels and the invertor. 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.

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

```
# The installation assumptions
solar_panels_count <- 36
solar_panels_area <- solar_panels_count * 2
solar_panels_max <- solar_panels_count * 275
solar_estimated_pa <- 7920
loss_factor <- solar_estimated_pa / solar_panels_max
irradiance_assumed <- solar_panels_max / solar_panels_area
```

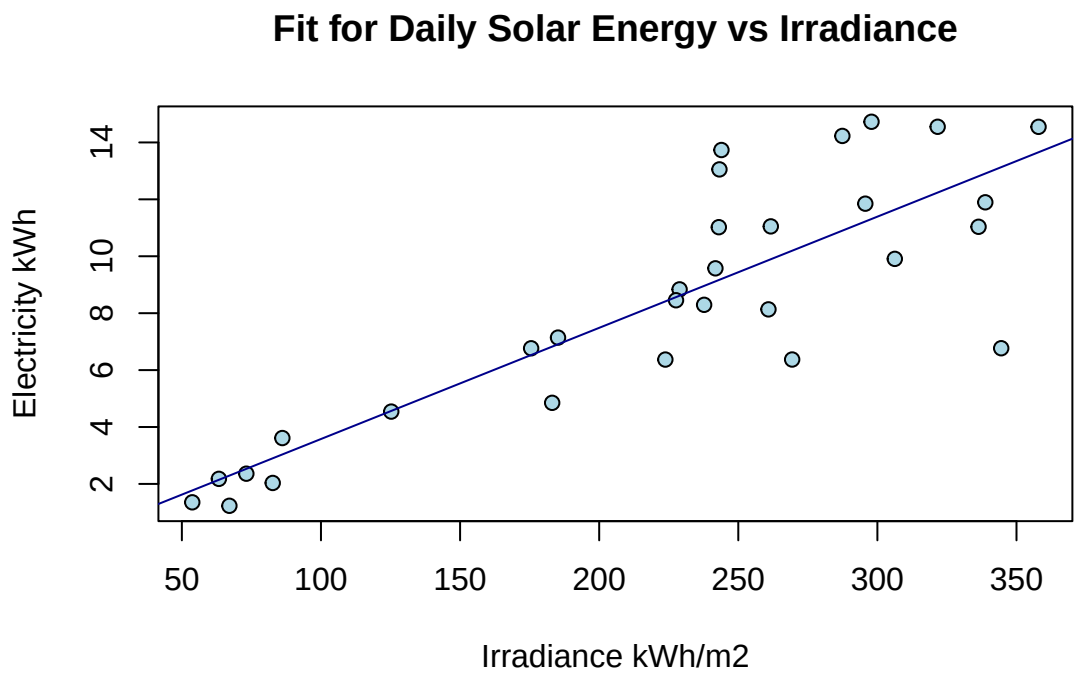
At installation, the annual generation power for this solar array was calculated as 7,920 kWh pa using an area of 72 m^2, irradiance 137.5 kWh/m^2 pa and a loss factor of 0.8. Using the September observations of irradiance and generated solar energy, the effectiveness of the installation can be compared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

Link Between Energy Generation & Irradiance

It is expected that solar generation is strongly linked to irradiance and this was checked using a scatter plot and linear regression.

```
# For solar generation, how well is this predicted by irradiance?
regression_model <- lm(energy_df$Solar_Total/1000 ~ energy_df$Irrdnce)
regression_summary <- summary(regression_model)

plot(energy_df$Solar_Total/1000 ~ energy_df$Irrdnce,
     main="Fit for Daily Solar Energy vs Irradiance", xlab="Irradiance kWh/m2", ylab="Electricity kWh",
     pch=21, bg="lightblue")
abline(regression_model, col="darkblue")
```

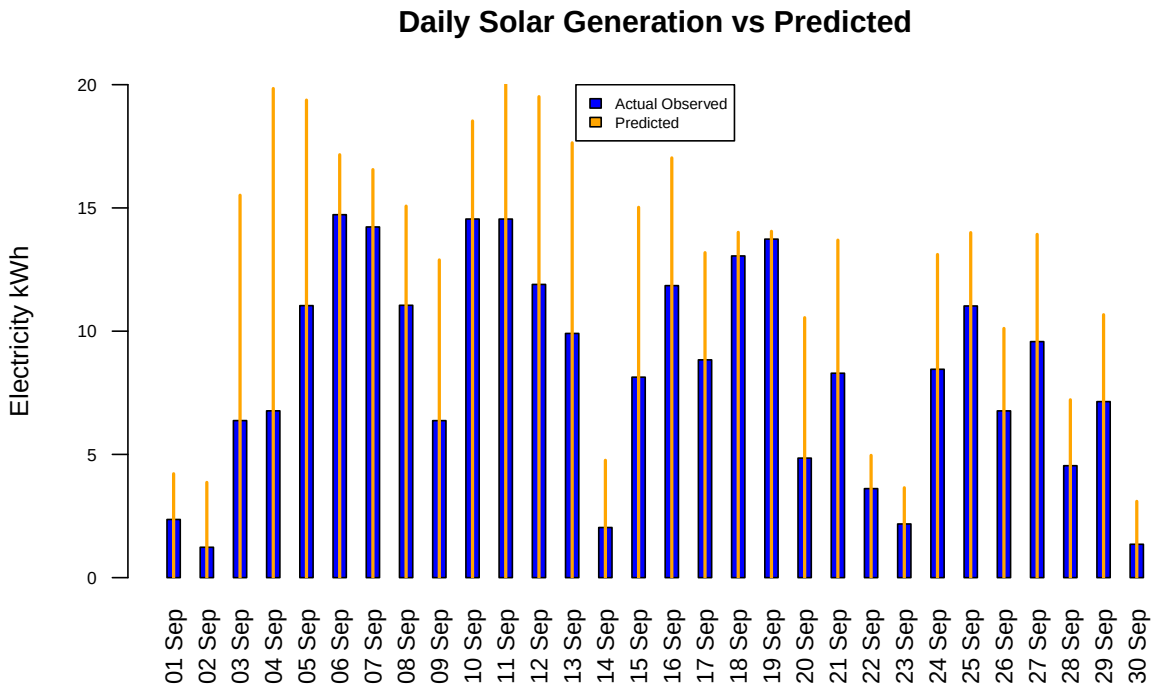


```
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta0 <- regression_summary$coefficients["energy_df$Irrdnce", "Estimate"]
p_value1 <- regression_summary$coefficients["energy_df$Irrdnce", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared
#print(sprintf("Alpha: %.3f, Beta Irradiation: %.3f", alpha, beta0))
#print(sprintf("p-value Irradiation: %.4f, Adj R-Squared: %.3f", p_value0, adj_r_squared))
```

The scatter plot and linear regression confirmed the relationship, with a p-value < 0.005 and a high adjusted r-squared 0.71.

Evaluate Effectiveness

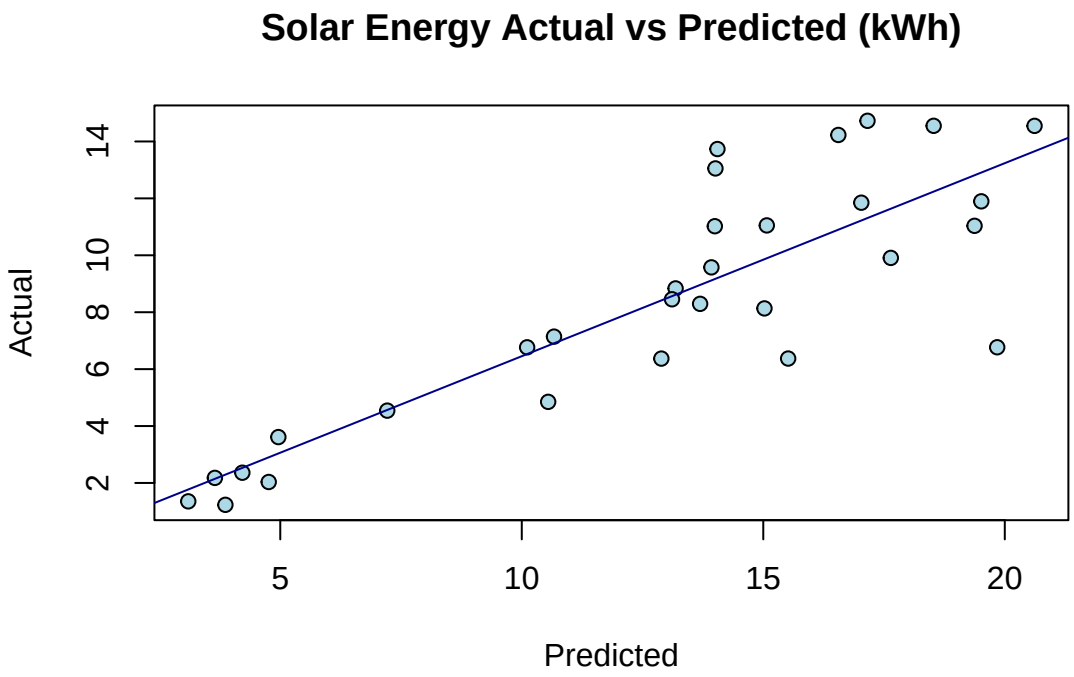
The sales prediction for the daily solar energy was calculated using the formulae above and with the irradiance observations during September. The means for the month are 8.35 kWh and 12.8 kWh respectively. The daily totals are compared in the plot below.



The figure clearly shows that the predicted generation is higher than the actual, although the differences do vary. Further analysis was completed to better understand this. A scatter plot below shows the relationship between the predicted and generated energy.

```
# Scatter plot and regression for solar generation & solar predicted
regression_model <- lm(energy_df$Solar_Total/1000 ~ solar_predicted)
regression_summary <- summary(regression_model)

plot(energy_df$Solar_Total/1000 ~ solar_predicted,
     main="Solar Energy Actual vs Predicted (kWh)", xlab="Predicted", ylab="Actual",
     pch=21, bg="lightblue")
abline(regression_model, col="darkblue")
```




```
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta0 <- regression_summary$coefficients["solar_predicted", "Estimate"]
p_value1 <- regression_summary$coefficients["solar_predicted", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared
#print(sprintf("Alpha: %.3f, Beta Irradiation: %.3f", alpha, beta0))
#print(sprintf("p-value Irradiation: %.4f, Adj R-Squared: %.3f", p_value0, adj_r_squared))
```

As expected the actual generated solar energy vs predicted do appear to have a strong relationship, with a p-value < 0.005 and a high adjusted r-squared 0.71.

Evaluate Effectiveness - T-Test

To try to better understand the relationship between the actual and predicted energy, a T-Test was completed. The null hypothesis is that the installation performs as well as the sales promises; specifically that the average daily energy production in September μ_{act} is the same as that promised (expected) μ_{exp}.

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

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

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

A T-Test was performed to evaluate H_0 and to 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 expected mean because we suspect that the installation does not perform as well as the sales promises; so a one-tailed paired test was used. Log values were used to try to make both samples closer to a normal distribution.

```
# Evaluate the T-Test
#t_test_result <- t.test(energy_df$Solar_Total/1000, solar_predicted, alternative = "less", paired = TR
t_test_result <- t.test(log10(energy_df$Solar_Total/1000), log10(solar_predicted), alternative = "less"
```

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 originally promised.

Evaluate Effectiveness - Binomial Probability

```
# How often is the generated energy close to the predicted value?
# Number of days generated is within 75% of the predicted value
nearPredicted <- energy_df$Solar_Total/1000 >= (0.75 * solar_predicted)
#table((nearPredicted))
p_value <- sum(dbinom(0:6,30,0.5))
```

An alternative way to look at the effectiveness of solar generation is to consider how often the actual solar generated power is near to the energy predicted during the sales process; within 75% seems a reasonable level. In September, this occurred on 6 days out of 30. A null hypothesis is that the 75% level can be met half of the time; so $H_0 : p(success) = 0.5$. The p-value is the probability, if the null hypothesis is true, of the observation or a more extreme observation. The calculated p-value is 0.001 which is below the critical value of 5% and therefore we can reject the null hypothesis and infer that the predicted energy is not met most of the time.

xxxxxxxxx

```
# How often are the actual values less then the predicted?
below_predicted <- energy_df$Solar_Total/1000 < solar_predicted
occurrences <- sum(below_predicted)
```

Further analysis was performed, to look at the frequency of actual daily generation being less than the predicted generation ($energy_{act} < energy_{exp}$). In September this happened on 30 days out of 30. The null hypothesis is that we should always achieve the predicted generation, so

$$H_0 : p(energy_{act} < energy_{exp}) = 0$$

```
# Determine the p-value for act < predicted on the given number of occurrences
# In a 30 day month and H0 saying the probability is zero
p_value <- dbinom(occurrences, 30, 0)
```

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 originally promised.

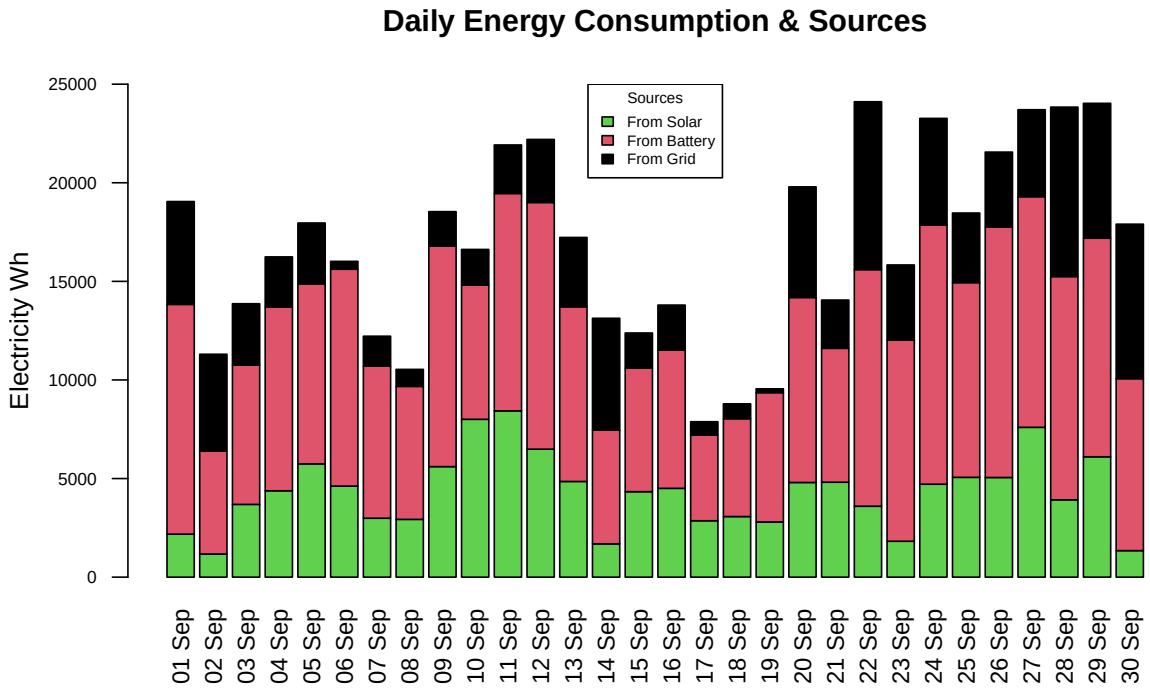
xxxxx ## Solar Energy Sufficiency

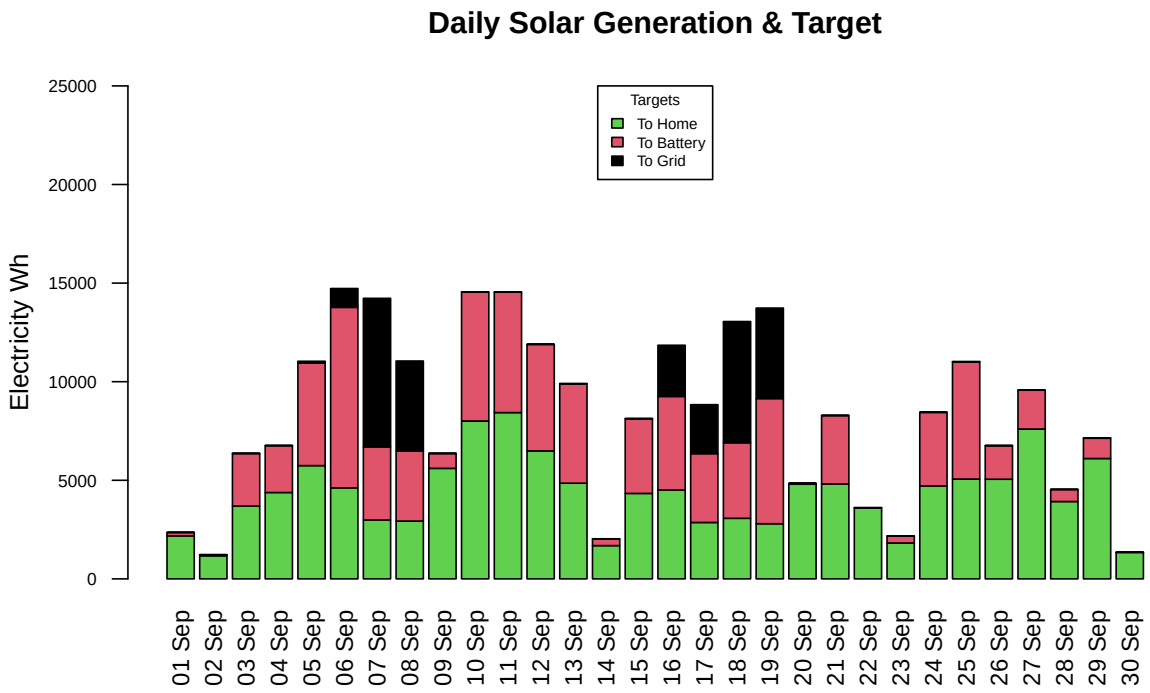
Visual Examination

The energy demands of the house were examined to see how well solar generated electricity matched this demand and how much any shortfalls needed to covered by imported electricity. The first stacked bar plot below shows the daily total energy consumed and how much is provided by: solar generation; the battery; or imported electricity. The second plot shows the daily total solar energy generated and where this is distributed to.

```r
# Stacked bar plot of daily energy consumption and the three sources of energy
# Transpose to matrix to use for stacked bar plot
usage <- t(energy_df)[c("From_Solar", "From_Pwall", "From_Grid"),]

barplot(usage, names.arg = format(energy_df$Date, "%d %b"),
        cex.name=0.9, las=2, cex.axis=0.7,
        main="Daily Energy Consumption & Sources", ylab="Electricity Wh",
        col=3:1, ylim=c(0,25000))
legend('top', legend=c("From Solar", "From Battery", "From Grid"), title="Sources",
       fill = 3:1, cex = 0.6)
```






```
# Calculate how the energy is consumed and distributed
average_solar_percent <- mean(c(energy_df$From_Solar / energy_df$Home_Total)) * 100
#print(paste("Straight percentage:", average_solar_percent))

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:", sola
# Calculate the approximate use of solar
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))
```

The first bar plot appears to clearly show that the generated solar energy is not sufficient to meet the total consumption needs, on average it only meets 26% of the daily demand. However, this is misleading as generated solar energy is often first stored in the battery for later use (nb it is sometime also to 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%, 8% respectively.

An alternative approximation of how well the solar generation meets the needs of the house's energy consumption is to look at how much usage is catered for without any import from the grid which suggests on average approximately 80% is ultimately provided by solar energy.

Successfully Meeting Demand - Fisher's Exact Test

In a previous section, looking at energy consumption, it appeared that occupancy was the bigger predictor of daily energy use. So it is interesting to see the relationship between successfully meeting demand (say, solar provides $> 75\%$ of energy) and significant occupancy (say, occupied $> 50\%$ of the day). I suspect that the generated solar energy does not meet the demands of a fully occupied house.

For September, the true/false occurrence of meeting demand and of being occupied were determined and a contingency table prepared.

```
# Prepare a contingency table looking at meeting demand vs occupation

# Determine true / false for each combination
demand_met <- ((energy_df$Home_Total - energy_df$From_Grid) / energy_df$Home_Total) > 0.75
occupied <- (energy_df$Occupied / 24) > 0.5
demandT_occpT <- sum(demand_met & occupied)
demandT_occpF <- sum(demand_met & !occupied)
demandF_occpT <- sum(!demand_met & occupied)
demandF_occpF <- sum(!demand_met & !occupied)

# Create matrix and display
contingency_table <- matrix(c(demandT_occpT,demandT_occpF,demandF_occpT,demandF_occpF), 2, 2)
dimnames(contingency_table) <- list(Demand_Met=c(TRUE,FALSE), Occupied=c(TRUE,FALSE))
print(contingency_table)
```

```
##           Occupied
## Demand_Met TRUE FALSE
##      TRUE     7     5
##      FALSE   15     3
```

This shows that demand is met in an occupied house on only 7 days out of 30, about 23% of September; however, we need to determine if this is statistically significant. The null hypothesis to consider is

$$H_0 : p(demandmet|occupied) = p(demandmet|\overline{occupied})$$

.

The contingency table is used to perform a Fisher's exact test to determine the p-value which is the probability, if the null hypothesis is true, of obtaining the observation, or an observation more extreme. In this case more extreme is that demand is met less frequently, so a one-sided test is used.

```
# Perform Fisher's Exact Test - a one-sided test
result = fisher.test(contingency_table, alternative = "less")
p_value_f = result$p.value
```

The results of a one-sided Fisher's exact test gave a p-value of 0.137, this is above the critical 5% value and so we cannot reject the null hypothesis; there is no evidence to suggest that the demand is not met when occupied. This does seem surprising given the visualisation.

Conclusions

- eg what conclusions were unexpected or surprising
- what have you learned from the data
- eg how has statistics helped you understand the data
- (conclusions are supported in the main body of the text)

Energy Consumption

- Analysis showed that energy consumption is strongly linked to a combination of temperature and occupancy
- However the link with temperature only was surprisngly weak
  - The temperature in September was only in the range 6 to 17 with a mean of 11 and this is not very extreme.
  - Good insulation with the above may have reduced the impact, colder weather will have a bigger impact, the calcuated regression line suggesst this with an aenergy demand of nearl 30 kWh at freezing point
  - Occupancy also not strong but similarly not big temperature demands, also good insulation and heating left at a constant level so just utilities such as washing and cooking and lighting, tv etc which is probably lower than the heating?

Solar Generation Effectiveness

- Analysis showed that the solar power generated is consistently less than that predicted during the month of September ... but not why? ... is it ineficciency or is it overforecast of irradiance?
- ?? probably throughout the year but only have monthly
- ?? but is this caused by inefficiency in the generation equipment or is it caused because the irradiance expereinced is less than that predicted? don't have annual irradiance and generation figures to determine this

Solar Energy Sufficiency

- ?? looks like majority of demands are met
- ?? but a surprising statistical result

?????? further work

?? Increased battery will smooth out across days? forecast storage/impact .. but can't see the intra-day detail to better analyse ?? Predict full year / annual generation using non September ?? monthly irradiance data: https://re.jrc.ec.europa.eu/pvg_tools/en/#MR ?? box plots

##?? Discussion

TO DO: Structure into two sections

##?? Conclusion

Test citations (Crawley, 2014) and as Spiegel and Schiller (2012)

Appendices

?? Energy Consumption ....

References

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

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