Stats Assignment 1

Group 48 - ABRMAR043

STA2005S Regression Assignment

Setup, Data loading and Objective:

Part One: Analysis

Section 1: Introduction

Problem & Unknown:

Analysis Summary:

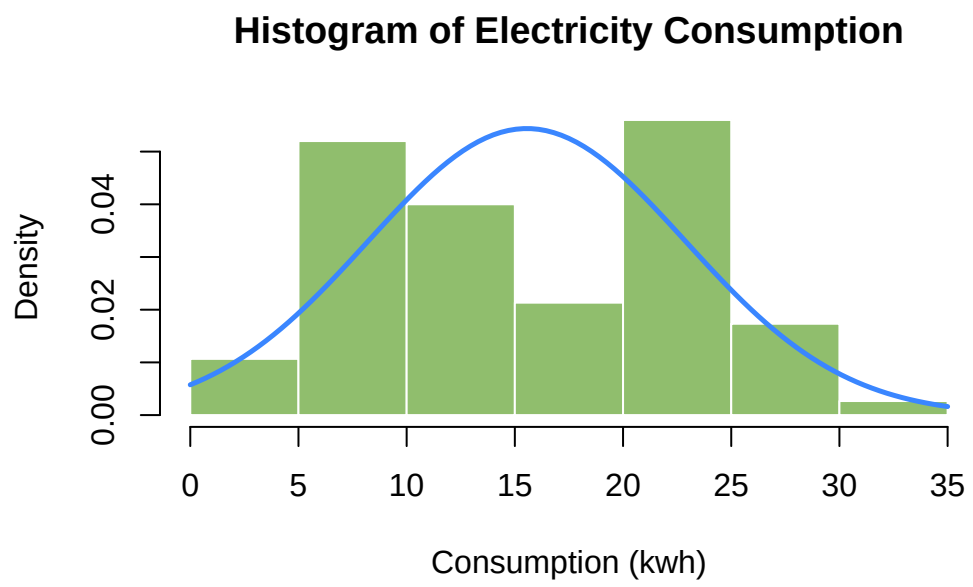
Nature of analysis:

Section 2: Data Exploration

Density Plots:

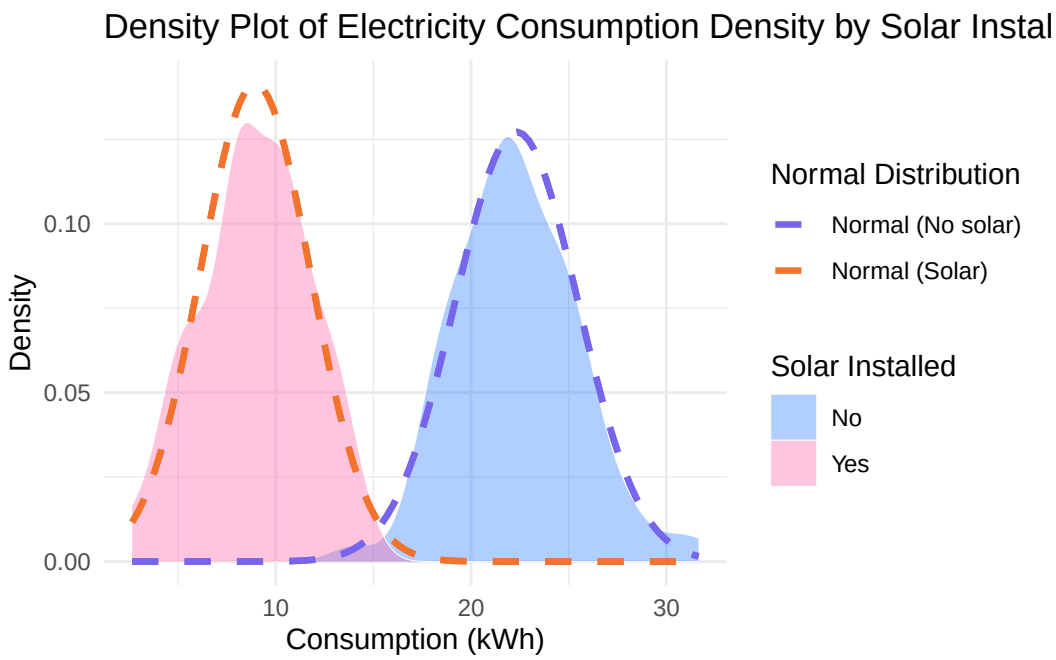
```
# Plot a histogram of consumption_kwh (freq = FALSE, if using base R) with an overlaid normal

mean_kwh <- mean(consumption_kwh, na.rm = TRUE)
sd_kwh <- sd(consumption_kwh, na.rm = TRUE)

hist(consumption_kwh,
     freq = FALSE,
     main = "Histogram of Electricity Consumption",
```

```
     xlab = "Consumption (kwh)",
     ylab = "Density",
     col = col_palette[7],
     border = "white"
     )

curve(dnorm(x, mean = mean_kwh, sd = sd_kwh),
      col = col_palette[11],
      lwd = 2.5,
      add = TRUE
      )
```




```
# Plot the density of consumption_kwh, stratified by solar_installed, with overlaid normal d

solar_yes <- data %>% filter(solar_installed == "Yes")
solar_no <- data %>% filter(solar_installed == "No")

mean_solar <- mean(solar_yes$consumption_kwh, na.rm = TRUE)
mean_no_solar <- mean(solar_no$consumption_kwh, na.rm = TRUE)
sd_solar <- sd(solar_yes$consumption_kwh, na.rm = TRUE)
sd_no_solar <- sd(solar_no$consumption_kwh, na.rm = TRUE)
```

```r
ggplot(data, aes(x = consumption_kwh, fill = solar_installed)) +
  geom_density(color = "white", linewidth = 0, alpha = 0.4) +

  stat_function(fun = dnorm,
                args = list(mean = mean_solar, sd = sd_solar),
                aes(linetype = "Normal (Solar)"),
                color = col_palette[4], linewidth = 1.15) +
  stat_function(fun = dnorm,
                args = list(mean = mean_no_solar, sd = sd_no_solar),
                aes(linetype = "Normal (No solar)"),
                color = col_palette[12], linewidth = 1.15) +

  scale_fill_manual(values= c("Yes"=col_palette[1], "No"=col_palette[11])) +
  scale_linetype_manual(values = c(2, 2), name = "Normal Distribution") +

  labs(title = "Density Plot of Electricity Consumption Density by Solar Installation",
       x = "Consumption (kWh)",
       y = "Density",
       fill = "Solar Installed"
       ) +
  theme_minimal()
```

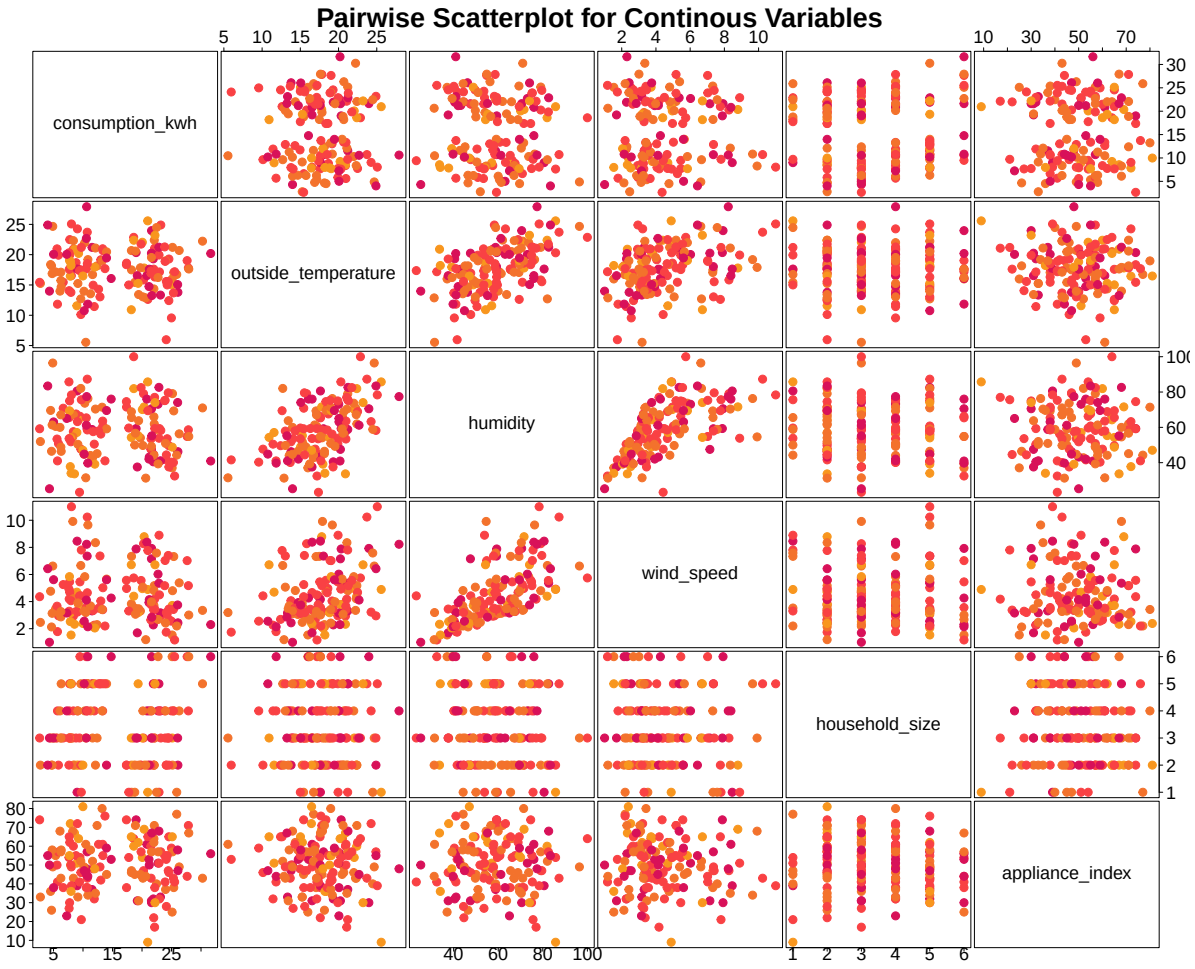


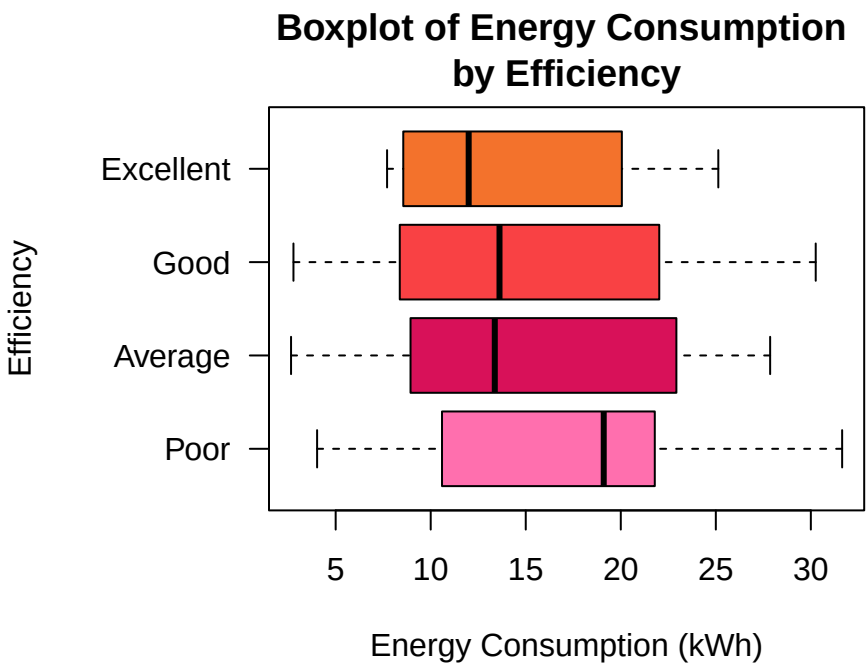
Pairwise Plots:

```
# Create pairwise scatterplots for all continuous variables, including consumption_kwh

cts_variables <- data %>%
  dplyr::select(consumption_kwh, outside_temperature, humidity,
                wind_speed, household_size, appliance_index)

day_numeric <- as.numeric(factor(data$day_of_week,
levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"), or

efficiency_numeric <- as.numeric(factor(data$energy_efficiency, levels = c("Poor", "Average"

solar_numeric <- as.numeric(factor(data$solar_installed, levels = c("Yes", "No"), ordered =
#_____________________________________________________________

# Graph
pairs(x = cts_variables, pch = 19,
      col = col_palette[efficiency_numeric + 1],
      main = " Pairwise Scatterplot for Continous Variables",
      gap = 0.5,
      las = 1,
      cex = 3,
      cex.labels = 3,
      cex.axis = 3,
      cex.main = 3
      )
```

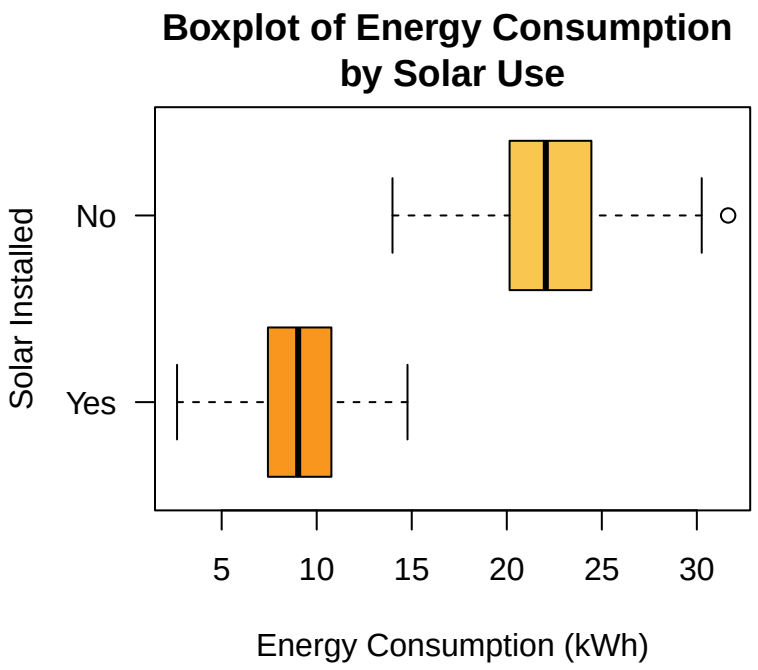
Categorical Variable Plots:

```
# Plot consumption_kwh against each categorical variable. Namely: Solar_installed, Energy_ef

par(mar = c(4, 10, 3, 2))
boxplot(consumption_kwh ~ efficiency_numeric, data = data,
        horizontal = TRUE,
        names = levels(energy_efficiency),
        xlab = "Energy Consumption (kWh)", ylab = "",
        main = "Boxplot of Energy Consumption \nby Efficiency",
        col = col_palette[1:4],
        las = 1
        )
mtext("Efficiency", side = 2, line = 6, cex = 1, font = 1)
```

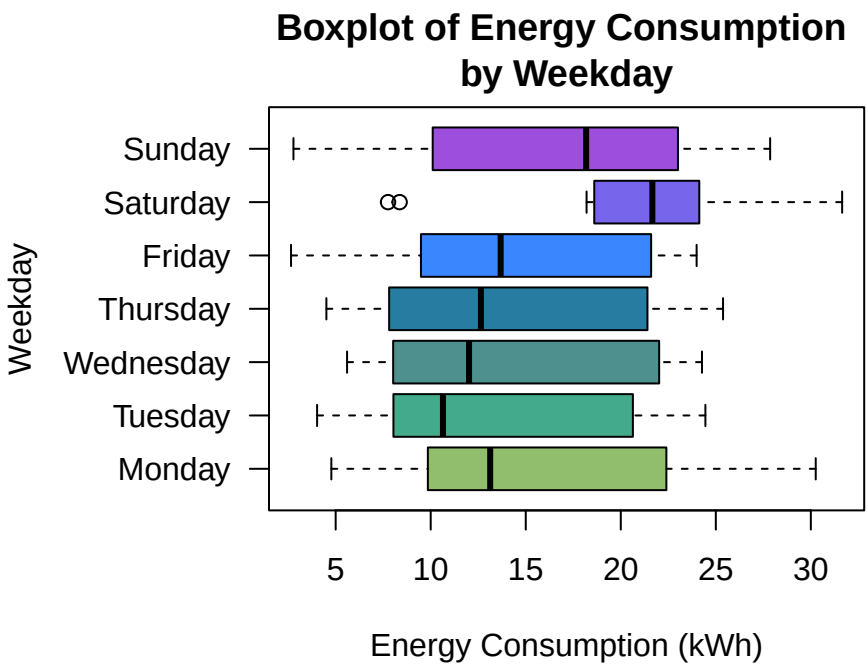

Pairwise Scatterplot for Continous Variables
consumption_kwh
outside_temperature
humidity
wind_speed
household_size
appliance_index



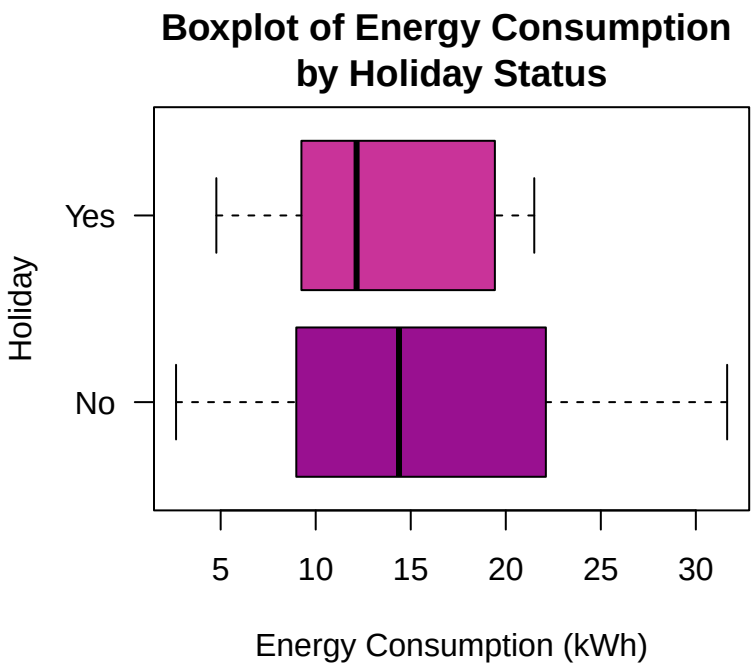

```r
par(mar = c(4, 10, 3, 2))
boxplot(consumption_kwh ~ solar_numeric, data = data,
        horizontal = TRUE,
        names = levels(solar_installed),
        xlab = "Energy Consumption (kWh)", ylab = "Solar Installed",
        main = "Boxplot of Energy Consumption \nby Solar Use",
        col = col_palette[5:6],
        las = 1
        )
```




```r
par(mar = c(4, 10, 3, 2))
boxplot(consumption_kwh ~ day_numeric, data = data,
        horizontal = TRUE,
        names = levels(day_of_week),
        xlab = "Energy Consumption (kWh)", ylab = "",
        main = "Boxplot of Energy Consumption \nby Weekday",
        col = col_palette[7:13],
        las = 1
        )
mtext("Weekday", side = 2, line = 6, cex = 1, font = 1)
```




```r
par(mar = c(4, 10, 3, 2))
boxplot(consumption_kwh ~ holiday, data = data,
        horizontal = TRUE,
        names = c("No", "Yes"),
        xlab = "Energy Consumption (kWh)", ylab = "Holiday",
        main = "Boxplot of Energy Consumption \nby Holiday Status",
        col = col_palette[14:15],
        las = 1
        )
```



Categorical Relationships:

```
\begin{table}[!h]
\begin{center}
\caption{Some important Table - put the caption at the top of a table} \vspace{.5cm}
\begin{tabular}{ |c|c| }
\hline
x & y \\
\hline
\hline
100 & 200 \\
150 & 259\\
120 & 249\\
111 & 212\\
139 & 221\\
342 & 435\\
\hline
\end{tabular}
\label{tab:somelabel} %this is how one adds a label for a table
\end{center}
\end{table}
```

Comments:

Section 3: Simple Linear Regression

Model Fitting:

Simultaneous Hypothesis Test:

Section 4: Multiple Linear Regression

Fit Model:

Hypothesis Testing:

Interpretation:

Section 5: Conclusion

Summary:

Recomendations:

Future Research:

Part Two: Simulation of Power

Scenario A: Baseline Normal Errors

Scenario B: Non-linear Truth/Exponential Relationship