Do Students' Motivations and Actions Differ on Campus than at Home? A case study on Khalifa University students

Begad Elsaid[1], Ahmed Saeed[2], Omran Al-Mofleh[3], and Hamza Riyad Hindi[4]

[1]Khalifa University, 100045069@ku.ac.ae
[2]Khalifa University, 100045067@ku.ac.ae
[3]Khalifa University, 100045079@ku.ac.ae
[4]Khalifa University, 100045100@ku.ac.ae

May 16, 2020

Abstract

This study aims to examine which predictors affect students' motivations and actions towards energy consumption. The data was collected by a survey conducted for Khalifa University students and includes 519 students from different campuses, genders and nationalities. The survey collects data about students' beliefs, social perceptions, motivations and actions towards saving energy. Descriptive Statistics, multiple linear regression models and subset selection processes were applied on the data collected to analyze and check the hypotheses. Our findings show that students tend to conserve energy more in home than on campus. It is also seen that motivations, beliefs and social perceptions impact the level of conservation. Results indicate that regardless of the different preferences, students are improving their energy saving behavior overtime.

Introduction of the Study

This study examines the relationships between conserving energy and many other controlling variables. Some of the controlling variables are personal (Gender, Nationality, Age. . .) and others are about the academic life (Education Level, Campus, Dorms. . .) as the survey was conducted on University students. The research also looks for the beliefs, social perceptions and motivations about saving energy. For the response variables, the study asks for the energy conservation levels at home and on campus.

Previous literature show that buildings' consumption of energy accounts for 20-40% of the total energy consumption [1]. This shows how important it is to increase the the conservative behavior towards saving energy in homes and campuses.

Data was collected by conducting a survey among Khalifa University students asking them questions to indicate the variables mentioned above. The survey collected 519 responses. In the coming sections, we first describe the survey and the data collected. Then, we state our hypotheses questions to proceed with the analysis. After that, we apply descriptive analysis by conducting graphs and examining basic statistics and indications. Moreover, we statistically examine the indications we observed. Last is the conclusion where we state our results from the research.

Survey Variables

The survey studied various control, independent, and response variables.

Control Variables included:

1. Age (20-29/30-39/40-49/50-59/60+).

2. Gender (A binary variable such that 1 = male or 0 = female).

3. Occupation (Student/Faculty/Post-doc/Staff/Other).

4. Education (School/High-school/Undergraduate/Graduate/PhD).

5. Location of residence (Dorms/ Home).

6. Nationality (Local / Non-Local).

7. The big 5 personality types: Openness, Agreeableness, Conscientiousness, Extraversion, Neuroticism (A Likert scale from 1 to 5 such that 1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, and 5 = strongly agree).

The independent variables measured are:

1. People's personal beliefs & norms on how conserving energy will:
   - Benefit Society
   - Protect the Natural Environment
   - Save money at home

2. Social perception of peoples' friends/colleagues/neighbors on conserving energy.

3. Social Network Characteristics:
   - What are the characteristics of the people you interact with? Same Age/Same Occupation/Same Nationality/Consider as friends
   - What are the locations of the interactions? Classroom/Labs/Dorms/Common Areas
   - How many people does the population interact with?
   - How often does the population interact with people?

4. Energy consumption patterns 6 and 12 months ago at home and on-campus.

Response Variables:

1. People's motivation to conserve energy at home and on-campus.

2. The extent of people's action of saving energy at home and on-campus.

Data Source & Exploration

The survey studied was conducted and data was provided by Dr. Elie Azar from the Industrial Engineering Department at Khalifa University (KU). Data was collected from 519 students from across all three campuses. Hot-deck imputation was used to fill out the missing data. The age and occupation variables were excluded from the model because almost all people were in the same age range and had the same occupation as students. The campus variable is also removed from the data because all people are KU students, and each campus has a very similar structure and system to another. Moreover, the responses from the questions were changed into a Likert scale from 1-5.

During the exploration of Cronbach-alpha values, beliefs on saving energy & social perception on saving energy variables seemed to be the only ones that can be grouped together with Cronbach alpha $\alpha = 0.7$ for both so we averaged the responses of the questions in both variables. The 5 personality variables all scored Cronbach alphas of $\alpha < 0.6$, thus they were removed from the analysis. We also conducted a pilot model on the energy consumption patterns at 6 and 12 months ago as response variables. We found that people ended up having the same values at different times and this may be due to people not remembering exactly what their energy consumption patterns are at these respective times. Therefore, we eliminate these variables and discuss more about them in the limitations and further research section.

Research Hypotheses

This report will document various statistical methods to respond to the below research questions:

1. Are there any differences in student's motivation and actions to conserve energy at home than on campus?

2. Which factors affect and what accounts for the differences in student's motivation and actions to conserve energy at home than on campus?

The answers will be in terms of diverse statistical information about the data's variables that will be discussed in the next parts.

Descriptive Statistics

To summarize the given data set, we will provide some descriptive statistics which will enable us to present the data in a more meaningful way, and allows a simpler interpretation of the data. We used a visualization tool called "Power Business Intelligence" to come up with the following charts.

In Figure 1, each pie chart represents the energy conservation at Home and Campus, respectively. These pie charts are divided depending on the responses recorded in the data and each color indicates a level of conservation from 1 to 5 (5 being very conservative). In general, the respondents gave answers to higher levels of energy conservation at home than at campus, and that can be seen by looking at the red and blue areas that are obviously bigger in the left chart (conservation at Home). Also, the average of conservation levels at home is significantly higher than the average at campus.

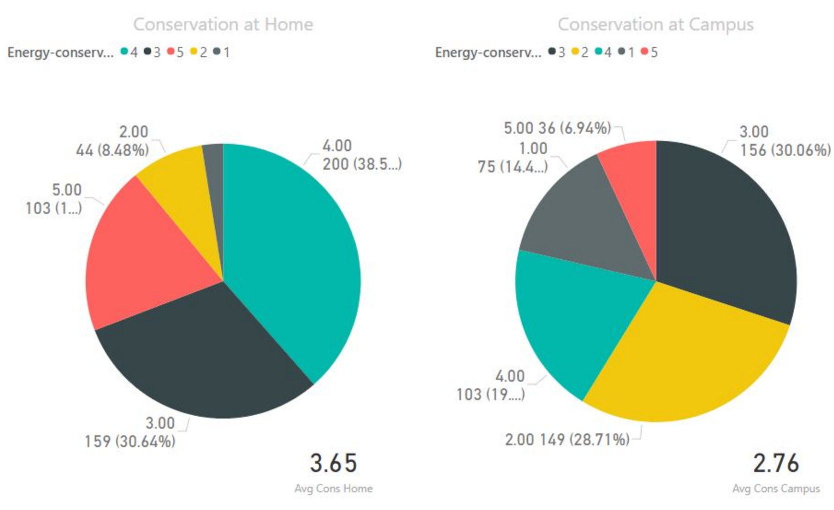


Figure 1: Energy conservation at home and at campus levels' responses

Next, given that the experiment was conducted on a total of 519 people, 235 males (45.3%) and 284 females (54.7%), we implemented the same charts for males in Figure 2 and females in Figure 3 below, to see if there exists any difference.

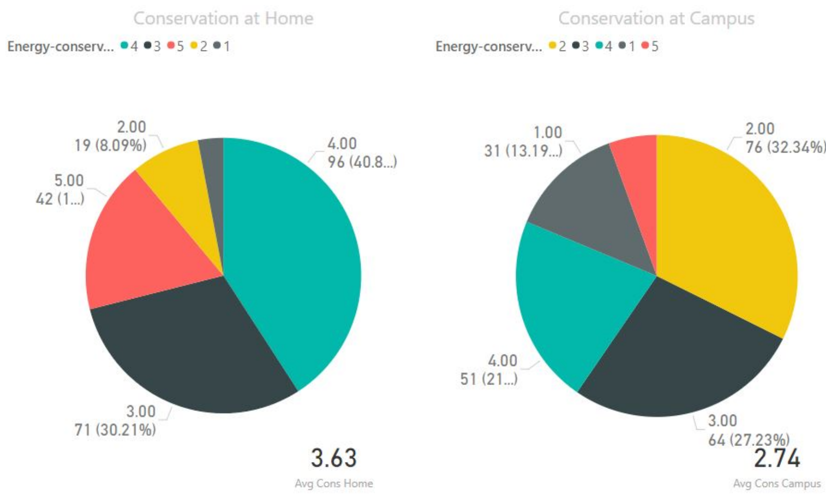


Figure 2: Energy conservation at home and at campus levels' male responses

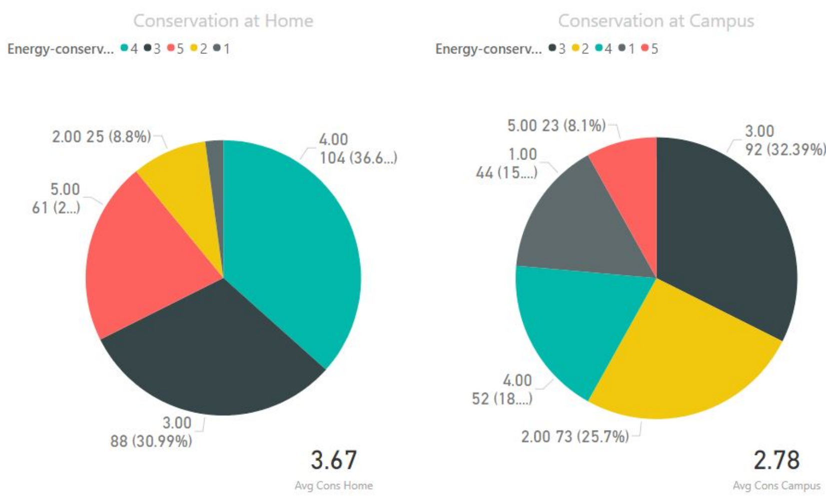


Figure 3: Energy conservation at home and at Campus levels' female responses

Looking at the previous charts, we noticed a slight difference in the average in both the males and females results. However, the average for the female respondents is a bit higher than both the males and the general averages.

Some other information that is interesting to check is seeing whether living in dorms or not would affect the level of energy conservation. The total data is divided into 295 respondents that do not live in dorms (56.5% of total) and 224 who live in dorms (43.16%). Figures 4 and 5 show the results for people who live in dorms and others who do not, respectively. The results show that there is no significant difference for the conservation at campus results, so we will deeply look at it in the quantitative statistics section to find out more about it. However, a significant change in the conservation at home charts is shown. The average increases when the students do not live in dorms, which means they would conserve less if they live in dorms. This result can indicate that students who live in dorms and go back to their homes would not care as much about saving energy. They may feel more comfortable and consume more.

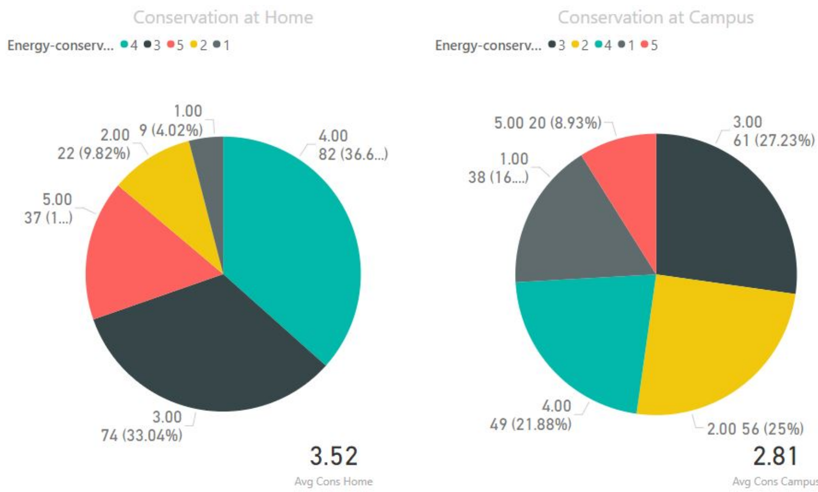


Figure 4: Energy conservation at home and at campus levels' responses of students living in dorms

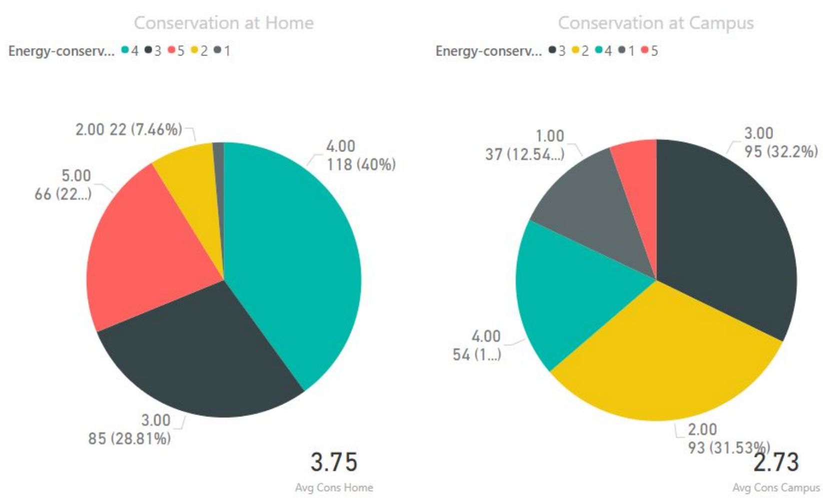


Figure 5: Energy conservation at home and at campus levels' responses of students not living in dorms

Then, we looked at the motivation data to see what are the levels of motivation the respondents get from home and campus. We can see in Figure 6 that motivation from home gives a higher average than from campus and this can be true because of the parents' contribution in saving energy at home or the inspiration that respondents get from their families are affecting their willingness to conserve energy.

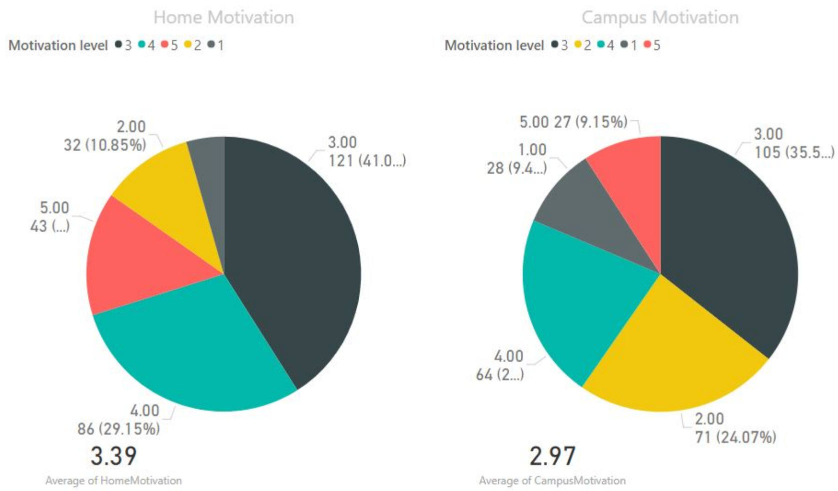


Figure 6: Home and Campus Motivation levels' responses and their averages

Next is the correlation between the motivation level that students get from home and the count of energy conservation at home responses. Basically, the graph, from motivation level 1 to 3, shows an increase in the responses regarding the levels on energy conservation which is common sense. When there is more motivation, students are more willing to conserve energy by consuming less. After the motivation level of 3, however, we can observe that the number of responses decreases and this is due to the high percentage of people who are motivated at level 3 compared to other levels of motivation. The red line, which is a level of 5 energy conservation, acts differently yet unsurprisingly. It increases whenever the motivation is increased, which means that students who conserve at the best level will perform better with higher motivation levels unlike the other 4 levels.

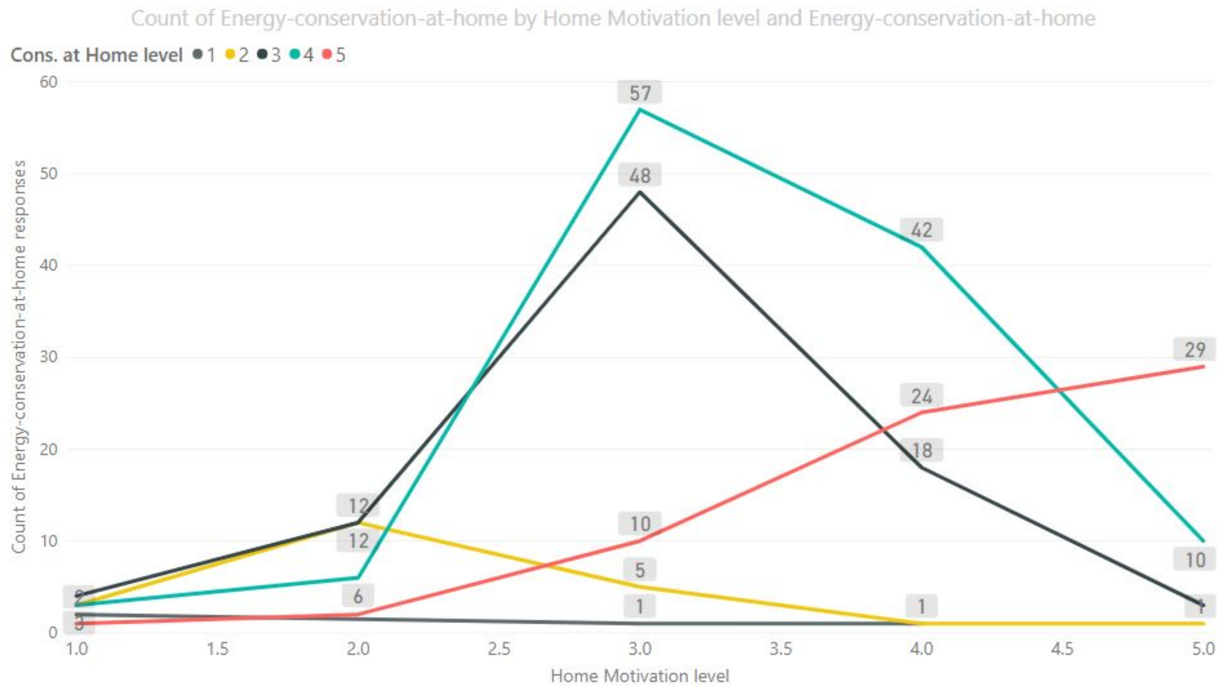
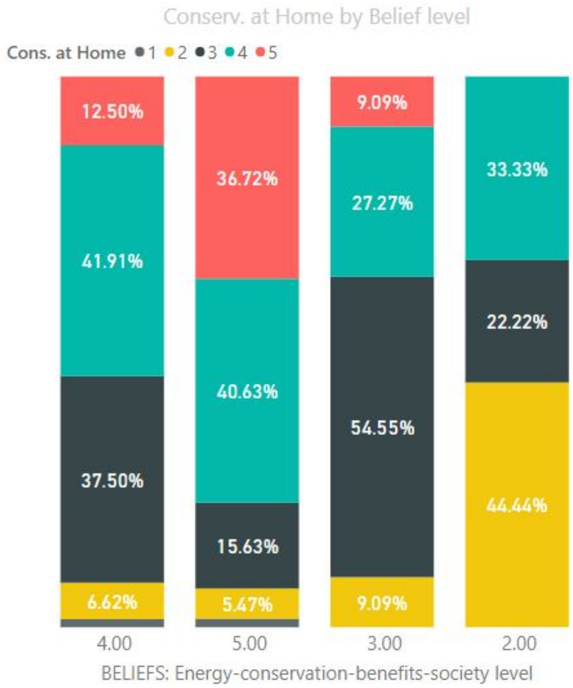


Figure 7: The relationship between the conservation of energy at home levels and the home motivation levels

Moreover, the respondents were asked to answer questions that indicate how much do they believe in certain statements and the first one is “Energy conservation benefits society”. We implemented two charts, one against conservation at home and the other against conservation at campus. Figure 8 shows how many responses of energy conservation we have for the first belief at each level. The figure displays the percentages of responses for each level of energy conservation at each belief level. For example, in the “conservation at Home” chart, we can see that 41.91% of the 4-level belief responses were recorded for the conservation of level 4. To conclude from the figure, we noticed that high levels of the mentioned belief have more responses with higher levels of energy conservation at home than at campus. This result supports our intention that people tend to conserve more at home than they do at campus. It also shows that beliefs are more likely to be applied within the family than in a public area, and this may be explained by the share of belief between family members and also could indicate that students have less control over energy consumption on campus than at home.



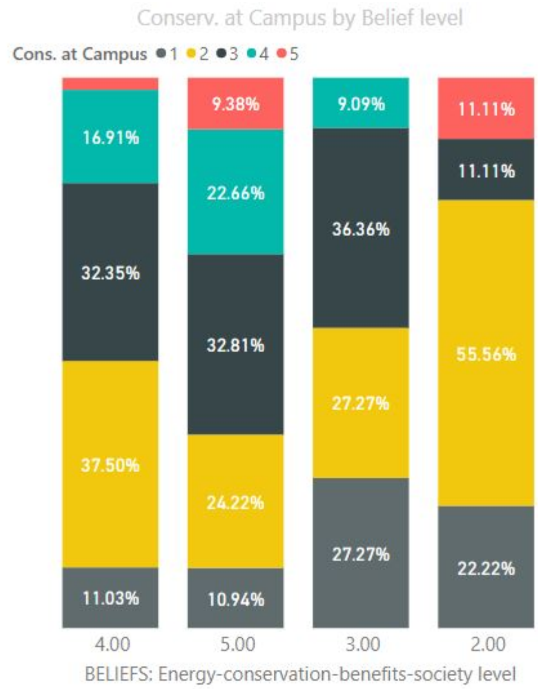


Figure 8: The count of each energy conservation level with respect to "Energy conservation benefits society" belief levels

Another belief that was tested is whether the conservation of energy protects the environment. Figure 9 shows the results by two graphs, one for conservation at home and the other for campus. we can observe quite the same result observed from Figure 8. For conservation at home, we can see that people who conserve at level 4 are the majority in all the belief levels (the green line is above all other lines). We can also observe that at level 5 belief, a big number of respondents do actually conserve at level 5. On the other hand, conservation on campus graph shows that most people at level 5 belief conserve at level 3 or 2. The interesting result is that the majority of people who answered 5 for this question are higher, which is different than in Figure 8 where the majority answered 4.

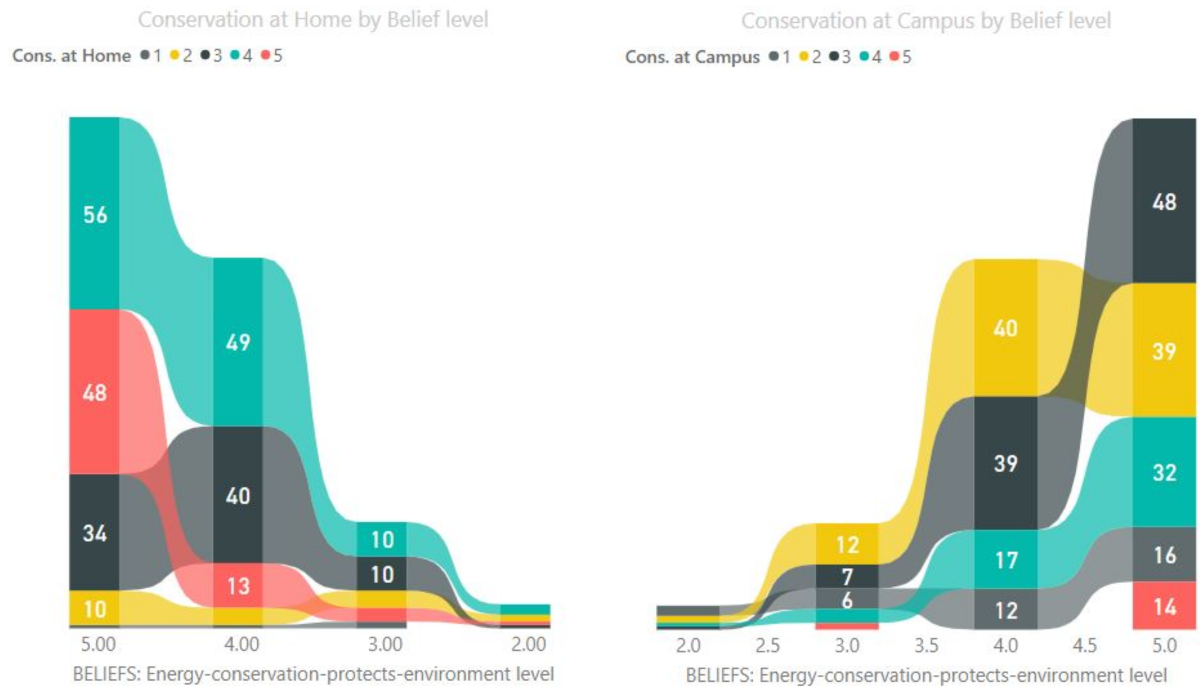


Figure 9: The count of each energy conservation level with respect to "Energy conservation protects the environment" belief levels

Moving to social perceptions results, the respondents were asked three questions about the level friends, colleagues, and neighbors save energy at, to see if that affects their motives to conserve energy. Figure 10 shows the results for energy conservation levels for friends against own conservation levels on campus. We can observe that nobody answered 5 meaning that no respondent has friends conserve at level 5. we also can see that the majority of people answered 2 or 3 for this question. we also can observe that majority of people who answered 2 and 3 are actually conserving at the same level on campus and this shows an indication that friends behavior affects respondents behavior on campus.

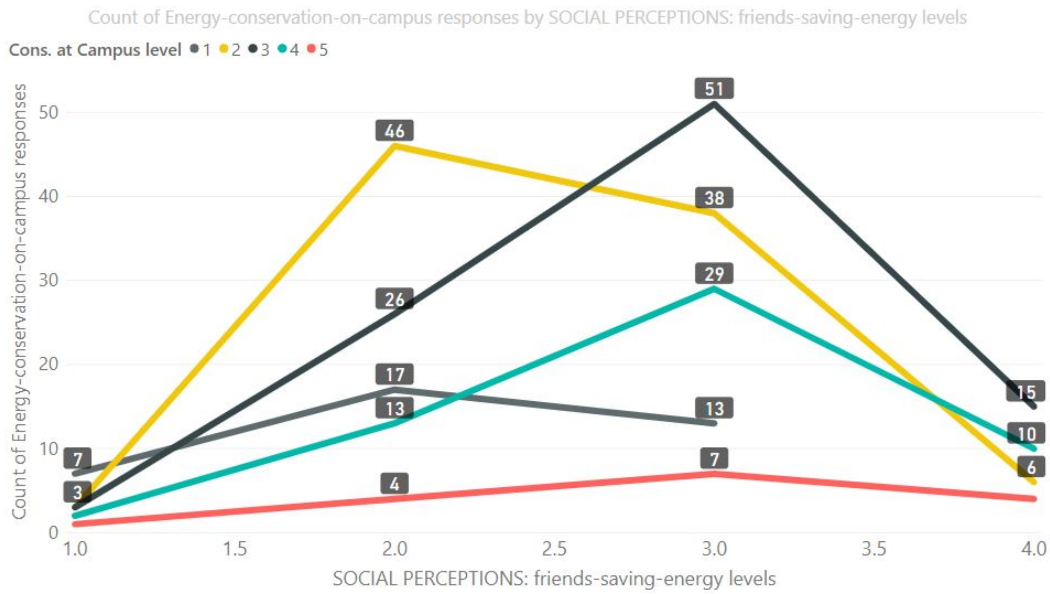


Figure 10: The count of each energy conservation level with respect to "Friends Saving Energy" Social Perception levels

Figure 11 shows the level of saving energy of neighbors against the respondents' actual energy consumption at home levels. We can notice that the majority of them answered 3 for the neighbors level of conservation while their actual conservation levels are mostly 4 and 5. This shows that people conserve at a level higher than their neighbors but we can not observe a link between the neighbors conservation levels and own conservation levels at home as all the answers to the neighbors question have the same pattern where most of the respondents actually conserve at level 4 or 5.

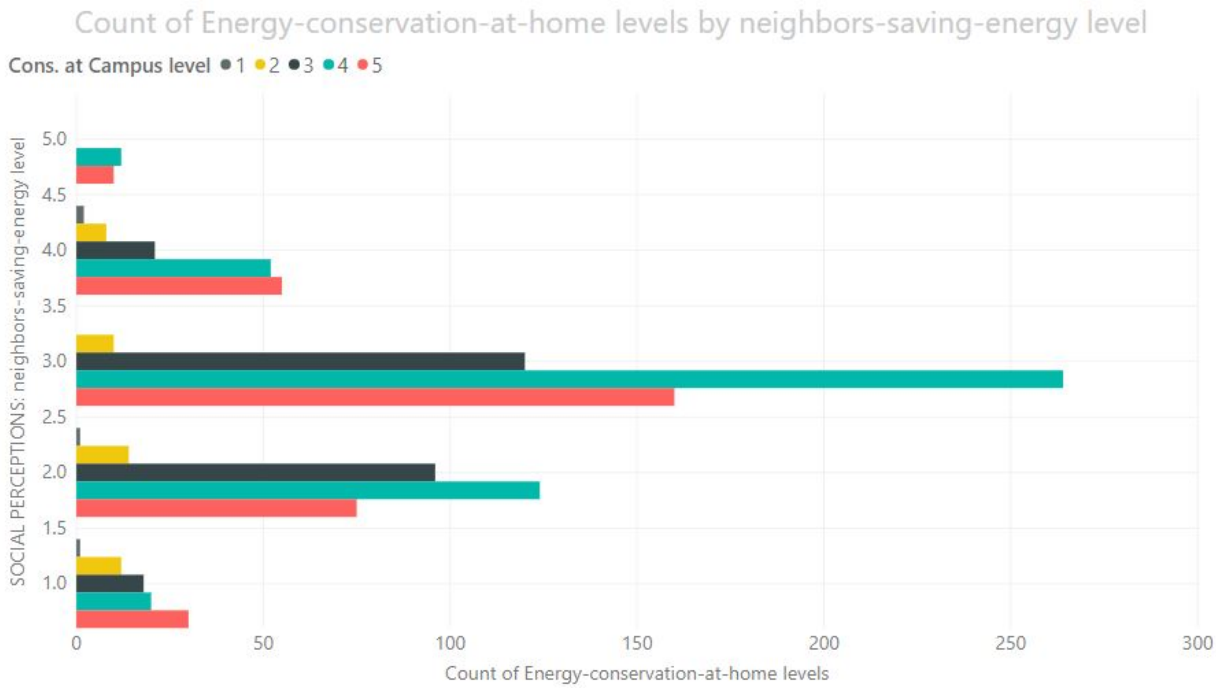


Figure 11: The count of each energy conservation at home level with respect to "Neighbors Saving Energy" Social Perception levels

For more interesting charts, please visit https://app.powerbi.com/groups/4262e68c-52c1-4afa-acb5-ac19d4141769/reports/014f4c31-2db5-4aa2-919f-485ed488cd19?ctid=08fe1c0a-19f5-4f24-a662-fdd5dd460025

Quantitative Analysis

To answer research question 1, we elected to use the Paired Two-Sample Wilcoxon test in part A. Furthermore, to find the factors that have a significant effect on student's motivations and energy consumption patterns, we investigated four Multiple Linear Regression models, Subset Selection and Ridge Regression. Finally, we further explored the connections between the variables studied using Principal Component Analysis (PCA) and Predictive Power Score (PPS).

Methodology

A. Paired Two-Sample Wilcoxon test

The first interesting research question explored in this section was whether student's motivation and actions to reduce energy consumption differ from home than on campus. Common techniques to answer this find such differences in population means are the analysis of variance (ANOVA) and the 2 sample t-test. However, by conducting the Shapiro-Wilk test, the assumptions of population linearity are not satisfied ($p_{Motivation} < 0.001$, $p_{Actions} < 0.001$) which is expected since the response variables are measured on a Likert scale. Therefore, a different test called the paired two-sample Wilcoxon test is used. The Wilcoxon test is a nonparametric test that finds differences in paired population means by ranking the difference in populations and calculating a test statistic W. Furthermore, the test does not require that the distribution of the mean of the population differences follow a normal distribution. Therefore, it is appropriate to use the test for the response variables studied.

B. MLR

To answer the second research question, we applied the multiple linear regression method, which combined the predictors to give information about the significance of the relationship between these and the four responses related to passion and extent of actions to save energy. We had tested different models:

Model 1: Predicts motivation at home

This model was used to predict the amount of the student's motivation in saving energy at homes (y_1). All of the independent and control variables mentioned above in the Survey Variables part, in addition to the response variables related to people's motivation to conserve energy at home and on-campus, were included as predictors in this model. The reason for making such particular responses as predictors is just to find the strength of cooperation between them. Some predictors had divided into more than one new predictor for sake of determining the difference between the effect of each of them on the responses, however, one of them should not appear when estimating the model as it is considered as a "reference group," i.e. a group of individuals who we use as a source to determine and form our thoughts, manners, and social customs [2]. For instance, the educational top degree of students had three choices: High-school, undergraduate, and graduate. The graduate students were chosen as a "reference group," so the rest of those choices became predictors in the process of estimating the model. This model came with β_i^1, the model's estimated coefficient of the i^{th}predictor, for each $i = 0, ..., 20$, and with ϵ_1, the model's error term. This model has the below linear equation:

$$y_1 = \beta_0^1 + \beta_1^1\,(Gender) + \beta_2^1\,(Education:\ High\ School) + \beta_3^1\,(Education:\ Undergraduate) + ...$$
$$+\beta_{18}^1\,(Campus\ motivation) + \beta_{19}^1\,(Average\ social\ perceptions) + \beta_{20}^1\,(Average\ beliefs) + \epsilon_1$$

Model 2: Predicts motivation on campus

We wanted in this model to estimate the amount of the motivation of students in saving campus's power (y_2). Each independent and control variable mentioned in the Survey Variables section, in addition to each response variable related to people's motivation to conserve energy at home and on-campus, was included as predictors in this model. This model came with β_i^2, which is the estimated coefficient of the i^{th}predictor, for each $i = 0, ..., 20$, and ϵ_2, which is the model's residual term. The equation represents this model is:

$$y_2 = \beta_0^2 + \beta_1^2\,(Gender) + \beta_2^2\,(Education:\ High\ School) + \beta_3^2\,(Education:\ Undergraduate) + ...$$
$$+\beta_{17}^2\,(Home\ motivation) + \beta_{18}^2\,(Control\ actions\ at\ campus) + \beta_{19}^2\,(Average\ social\ perceptions)$$
$$+\beta_{20}^2\,(Average\ beliefs) + \epsilon_2$$

Model 3: Predicts actions at home

In this model, we were measuring the extents of the student's actions in saving homes' energy, which is denoted as y_3. All control and dependent variables, in addition to the all response variables related to people's motivation and actions to conserve energy at home and on-campus, were included as predictors in this model. The model comes with β_i^3 , the third model's estimated coefficient of the i^{th}predictor, for each $i = 0, ..., 22$, and ϵ_3, the model's residual term. The relevant linear equation is:

$$y_3 = \beta_0^3 + \beta_1^3\,(Gender) + \beta_2^3\,(Education:\ High\ School) + \beta_3^3\,(Education:\ Undergraduate) + ...$$
$$+\beta_{17}^3\,(Home\ motivation) + \beta_{18}^3\,(Control\ actions\ at\ campus) + \beta_{19}^3\,(Campus\ motivation)$$
$$+\beta_{20}^3\,(Campus\ actions\ in\ current) + \beta_{21}^3\,(Average\ social\ perceptions) + \beta_{22}^3\,(Average\ beliefs) + \epsilon_3$$

Model 4: Predicts actions on campus

The goal of creating this linear model was to estimate the extent of students' actions in saving campus's power, denoted as y_4. All dependent, independent and control variables were engulfed as predictors in this model. The model includes β_i^4 , the fourth model's estimated coefficient of the predictor i, for each $i = 0, ..., 22$, and ϵ_4, the model's error term. This model has the following equation:

$$\begin{aligned} y_4 = \beta_0^4 + \beta_1^4\,(Gender) + \beta_2^4\,(Education:\ High\ School) + \beta_3^4\,(Education:\ Undergraduate) + ... \\ +\beta_{17}^4\,(Home\ motivation) + \beta_{18}^4\,(Home\ actions\ in\ current)\ + \beta_{19}^4\,(Control\ actions\ at\ campus) \\ +\beta_{20}^4\,(Campus\ motivation) + \beta_{21}^4\,(Average\ social\ perceptions) + \beta_{22}^4\,(Average\ beliefs) + \epsilon_4 \end{aligned}$$

The most significant predictors in every linear model will be arranged in descending order based on the importance of those variables, that is, "the absolute value of β" [3].

Multicollinearity in each linear model will be tested with the help of the variance inflation factor (VIF). High values of VIF (bigger than 5) indicate trouble and it should be fixed to reduce this case. Here is the VIF equation for each predictor i:

$$VIF_i = \frac{1}{1 - R_i^2},$$

where R_i^2 is the R-squared value of the i^{th} independent variable when regression over other variables. This formula could be adjusted also to find the overall VIF of the model using the R-squared value of this model, where the whole model's R^2 is defined as the proportion of a response variable' variance can be illustrated by the predictors in the model itself [4], [3] .

To select the set of variables to include in each model, we test both Subset Selection and Ridge Regression which are feature selection techniques that exclude irrelevant predictors from the ordinary least squares models discussed earlier [5].

C. Subset Selection

Subset Selection fits all possible multiple linear regression models from no predictors to all p predictors studied. Then, it filters through all the models and finds the best p models each having a different number of predictors chosen by having the highest R^2 value. Furthermore, Subset Selection compares all p models using certain statistics and chooses the best one. We elect to choose the C_p statistic to compare the p models which is defined as follows:

$$C_p = \frac{1}{n}\left(RSS + 2d\hat{\sigma}^2\right)$$

where RSS is the residual sum of squares, d is the number of predictors in the model and $\hat{\sigma}^2$ is an estimate of the variance of the residuals ϵ. The term $2d\hat{\sigma}^2$ accounts for the underestimation of the training error to the test error. In fact, it can be shown that C_p an unbiased estimator of the test MSE. Therefore, C_p is appropriate to compare the models and the models with the lowest C_p values are chosen as the best model. Note that the C_p statistic is proportional to the Akaike Information Criterion (AIC) and thus will have the same results.

D. Ridge Regression

Ridge Regression is a type of Shrinkage method that involves minimizing the RSS of a least squares model with all predictors with the addition of a regularization term. The regularization term $\lambda||\beta||_2$ consists of a tuning parameter $\lambda > 0$ and the l_2 norm of the least squares coefficient excluding the intercept coefficient. As $\lambda \to \infty$, all coefficients β shrink to zero from the minimization problem. The goal of ridge regression is to find a balance between shrinking the coefficients of the irrelevant predictors and keeping the significant predictors in the model. Moreover, a further advantage of ridge regression is that the regularization of the coefficients has the effect of reducing their variance. Different values of λ yield different coefficient estimates and the optimal value of λ can be found by performing ridge regression to a mesh of λ values and choosing the λ which gives the model with the lowest MSE. Note, that when $\lambda = 0$, the regularization term is zero, and thus the least squares model is fit.

E. PCA

Previously, we had identified Cronbach Alpha for groups of questions in the survey data to determine which questions are considered reliable and have a good correlation to response variables, and then we removed all questions that had a Cronbach Alpha of 0.6 or lower to make the data suitable for PCA. In this method, we explored more the second research inquiry by finding the optimal value of principal components (PCs) using a screeplot. Since there are some variables as binary and others as quantitative numbers, hence we had standardized predictors so that we can get a linear combination of these standardized variables and a reduced difference between variances of them. The variables extracted from PCA could be used to classify people based on their characteristics of motivation and actions to save power. For a better description, we used an orthogonal rotation of type "varimax" by maximizing the combination of sample variability of standardized loadings for every principal component over all PCs obtained.

F. Predictive Power Score

Predictive Power Score (PPS) is a new statistic proposed to find how well a variable predicts another variable. This method poses advantages over the person correlation as it uncovers non-linear effects and asymmetry as well as be able to analyze categorical variables. For example, when considering the relationship between students being in the dorms (Dorms) and having dorm interaction (Dorm interaction), Dorm interaction predicts Dorms very well ($PPS = 0.64$) while the converse relation is not as strong ($PPS = 0.25$). This is because if a person interacts a lot with people in the dorm then they are definitely living in the dorms. However, if a person lives in the dorms, they do not necessarily interact with people. A relationship like this example would not be uncovered by Pearson's Coefficient since it is symmetric. Furthermore, PPS is calculated by a cross-validated Decision Tree where PPS for continuous variables are found through a Decision Tree Regressor and calculating the Mean Absolute Error (MAE), and categorical variables can be found using a Decision Tree Classifier and calculating the weighted F1 statistic [6].

Results

A. Paired Two-Sample Wilcoxon test

The results from the Wilcoxon test show that student's motivations at home and on campus are different ($p_{Motivation} < 0.001$) where motivation to save energy at home ($\mu = 3.4$) is higher than their motivation to save energy on campus ($\mu = 2.98$). Moreover, students conserve energy more at home ($\mu = 3.65$) than on-campus ($\mu = 2.77$) with ($p_{Actions} < 0.001$). An interpretation of these results could be that students at home are self-aware of their use of electricity because either the family or themselves pay for it. However, on-campus, there are no repercussions to using excessive electric energy and thus their motivation to save is low there.

B. MLR

To investigate relations between predictors and dependent variables in the data, we had the different summaries (see Figures 12 to 15), and a brief description of VIFs of the models.

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              0.7887165  0.4703968   1.677  0.09423 .  
GenderMale              -0.0464740  0.0839096  -0.554  0.57992    
EducationHigh-school    -0.0308607  0.2793669  -0.110  0.91208    
EducationUndergraduate -0.0854467  0.2949575  -0.290  0.77217    
DormsYes                 0.1367596  0.1111109   1.231  0.21896    
NationalityUAE          -0.0103335  0.1333735  -0.077  0.93827    
InteractionFrequency     0.0881246  0.0535458   1.646  0.10044    
NumberOfInteractions    -0.0333632  0.0482144  -0.692  0.48927    
ClassroomInteraction     0.0952865  0.0526453   1.810  0.07090 .  
LabInteraction           0.0092701  0.0491553   0.189  0.85049    
DormsInteraction        -0.0401539  0.0359041  -1.118  0.26395    
CommonAreaInteraction    0.0007562  0.0419001   0.018  0.98561    
SameAge                  0.0709277  0.0461072   1.538  0.12460    
SameOccupation          -0.0487853  0.0440757  -1.107  0.26889    
SameNationality         -0.0169616  0.0389228  -0.436  0.66319    
AbsenceInteraction      -0.0831502  0.0523802  -1.587  0.11305    
FriendsInteraction       0.0258320  0.0481528   0.536  0.59188    
CampusActionControl      0.0538224  0.0434501   1.239  0.21603    
CampusMotivation         0.3111824  0.0408382   7.620 1.29e-13 ***
AvgSocialPercep          0.1181937  0.0651513   1.814  0.07026 .  
AvgBeliefs               0.2059737  0.0630910   3.265  0.00117 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8886 on 498 degrees of freedom
Multiple R-squared:  0.2506,	Adjusted R-squared:  0.2205 
F-statistic: 8.324 on 20 and 498 DF,  p-value: < 2.2e-16
```

Figure 12: MLR outcomes on Model 1

```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)           -0.772726   0.488619  -1.581   0.1144    
GenderMale             0.040517   0.087141   0.465   0.6422    
EducationHigh-school   0.010328   0.290102   0.036   0.9716    
EducationUndergraduate 0.174353   0.306215   0.569   0.5694    
DormsYes               0.076603   0.115504   0.663   0.5075    
NationalityUAE        -0.044799   0.138483  -0.323   0.7465    
InteractionFrequency  -0.043032   0.055720  -0.772   0.4403    
NumberOfInteractions   0.057891   0.050023   1.157   0.2477    
ClassroomInteraction  -0.032816   0.054827  -0.599   0.5498    
LabInteraction         0.008954   0.051044   0.175   0.8608    
DormsInteraction      -0.075720   0.037176  -2.037   0.0422 *  
CommonAreaInteraction  0.105124   0.043254   2.430   0.0154 *  
SameAge                0.024782   0.047979   0.517   0.6057    
SameOccupation        -0.002311   0.045825  -0.050   0.9598    
SameNationality        0.003244   0.040426   0.080   0.9361    
AbsenceInteraction     0.074056   0.054429   1.361   0.1743    
FriendsInteraction    -0.105661   0.049792  -2.122   0.0343 *  
HomeMotivation         0.335551   0.044036   7.620 1.29e-13 ***
CampusActionControl    0.273566   0.043494   6.290 6.97e-10 ***
AvgSocialPercep        0.203611   0.067261   3.027   0.0026 ** 
AvgBeliefs             0.309397   0.064744   4.779 2.32e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9227 on 498 degrees of freedom
Multiple R-squared:  0.3472,	Adjusted R-squared:  0.321 
F-statistic: 13.25 on 20 and 498 DF,  p-value: < 2.2e-16
```

Figure 13: MLR outcomes on Model 2

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)             0.864404   0.430161   2.009  0.04503 *  
GenderMale              0.086257   0.076304   1.130  0.25884    
EducationHigh-school    0.340880   0.254055   1.342  0.18029    
EducationUndergraduate  0.563604   0.268153   2.102  0.03607 *  
DormsYes               -0.274790   0.101166  -2.716  0.00683 ** 
NationalityUAE         -0.208261   0.121803  -1.710  0.08793 .  
InteractionFrequency    0.008232   0.048805   0.169  0.86613    
NumberOfInteractions   -0.102554   0.043917  -2.335  0.01993 *  
ClassroomInteraction    0.002004   0.048048   0.042  0.96674    
LabInteraction          0.008538   0.044686   0.191  0.84855    
DormsInteraction        0.025906   0.032677   0.793  0.42828    
CommonAreaInteraction   0.021002   0.038162   0.550  0.58234    
SameAge                -0.001043   0.042027  -0.025  0.98021    
SameOccupation          0.025140   0.040215   0.625  0.53216    
SameNationality         0.039880   0.035389   1.127  0.26033    
AbsenceInteraction      0.049602   0.047825   1.037  0.30017    
FriendsInteraction     -0.063161   0.043801  -1.442  0.14993    
HomeMotivation          0.373201   0.040778   9.152  < 2e-16 ***
CampusActionControl    -0.022134   0.041822  -0.529  0.59688    
CampusMotivation        0.060291   0.043090   1.399  0.16238    
CampusActionCurrently   0.091322   0.043874   2.081  0.03790 *  
AvgSocialPercep        -0.110590   0.060131  -1.839  0.06649 .  
AvgBeliefs              0.293537   0.058133   5.049 6.24e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8077 on 496 degrees of freedom
Multiple R-squared:  0.3403,	Adjusted R-squared:  0.311 
F-statistic: 11.63 on 22 and 496 DF,  p-value: < 2.2e-16
```

Figure 14: MLR outcomes on Model 3

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)            -0.8590442  0.4384101  -1.959  0.05062 .  
GenderMale              0.0179224  0.0778477   0.230  0.81801    
EducationHigh-school   -0.2048020  0.2591821  -0.790  0.42980    
EducationUndergraduate -0.1306595  0.2743927  -0.476  0.63416    
DormsYes                0.0666004  0.1038068   0.642  0.52144    
NationalityUAE         -0.2459835  0.1239883  -1.984  0.04782 *  
InteractionFrequency   -0.0086732  0.0497311  -0.174  0.86162    
NumberOfInteractions    0.0657441  0.0448984   1.464  0.14375    
ClassroomInteraction    0.0428562  0.0489221   0.876  0.38145    
LabInteraction          0.0119991  0.0455325   0.264  0.79225    
DormsInteraction        0.0008949  0.0333183   0.027  0.97858    
CommonAreaInteraction   0.0524417  0.0388269   1.351  0.17742    
SameAge                -0.0273570  0.0428071  -0.639  0.52307    
SameOccupation          0.0622229  0.0408986   1.521  0.12880    
SameNationality        -0.0129356  0.0361018  -0.358  0.72027    
AbsenceInteraction      0.0624434  0.0487046   1.282  0.20041    
FriendsInteraction     -0.0226465  0.0447136  -0.506  0.61275    
HomeMotivation          0.0082809  0.0449218   0.184  0.85382    
HomeActionCurrently     0.0948210  0.0455548   2.081  0.03790 *  
CampusActionControl     0.3089648  0.0403074   7.665 9.48e-14 ***
CampusMotivation        0.3973005  0.0402151   9.879  < 2e-16 ***
AvgSocialPercep         0.2194226  0.0606865   3.616  0.00033 ***
AvgBeliefs              0.0739171  0.0606490   1.219  0.22351    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.823 on 496 degrees of freedom
Multiple R-squared:  0.4958,	Adjusted R-squared:  0.4735 
F-statistic: 22.17 on 22 and 496 DF,  p-value: < 2.2e-16
```

Figure 15: 12 MLR outcomes on Model 4

The first two models show results of predicting motivation to save power. The overall first model in Figure 12 was significant ($p < 2.2 \times 10^{-16}$), which implies that all independent and control variables could significantly expect the motivation rate to conserve energy at home. The same model could explain only 25% of variance of the home motivation. However, the most significant predictors were the motivation of students to conserve energy at campuses ($\beta = 0.31,\ p = 1.29 \times 10^{-13}$) and then the average of students' beliefs ($\beta = 0.21,\ p = 0.001$). But, the rest of predictors had a very tiny effect on home motivation because their p-values were more than 0.05. Regarding the enthusiasm at KU campus to save energy, each predictor could have a significant linear relation with this response due to the fact that the second model contained them was significant ($p < 2.2 \times 10^{-16}$), and the campus motivation's variability covered was not more than 35%. The variables that most significantly predict energy motivation at KU campus were the motivation of students to conserve energy at home ($\beta = 0.34,\ p = 1.29 \times 10^{-13}$), the average of students' beliefs ($\beta = 0.31,\ p = 2.32 \times 10^{-6}$), the control of consuming campuses' powers ($\beta = 0.27,\ p = 6.97 \times 10^{-10}$), the average of social perceptions ($\beta = 0.2,\ p = 0.003$), the interaction with friends ($\beta = -0.106,\ p = 0.03$), the interactions at public areas on campus ($\beta = 0.105,\ p = 0.02$), and the interactions at dorms ($\beta = -0.08,\ p = 0.04$). The other predictors could not make, at all, a significant impact on campus motivation as their p-values exceed 0.05. As we see from first two models, the significance of relation between home motivation and campus motivation implies that what affects on the first affects also on the latter, however the predictors obtained in each model contradict this fact.

The last two models display the summaries of predicting recent actions to save energy. The whole third model was significant ($p < 2.2 \times 10^{-16}$), provided with a 34% of recent home action's variance explained by the model's predictors. Even though all predictors were statistically significant, the only predictors who took a huge part to this model were the students with an undergraduate degree ($\beta = 0.56,\ p = 0.04$), the motivation of students to conserve energy at home ($\beta = 0.37,\ p < 2 \times 10^{-16}$), the average of students beliefs ($\beta = 0.29,\ p = 6.24\ \times 10^{-7}$) , the students who have dorms ($\beta = -0.27,\ p = 0.01$), the aggregate number of student interactions ($\beta = -0.1,\ p = 0.02$), and the current actions to save energy at campuses ($\beta = 0.09,\ p = 0.04$). In the fourth model, all variables included expected the campus action significantly ($p < 2.2 \times 10^{-16}$), and the model itself had explained only 50% of variance in the current actions to save power in KU campuses. However, the variables importantly predict actions done at KU campus to save power were the motivation of students to conserve energy at campuses ($\beta = 0.4,\ p < 2 \times 10^{-16}$), the control on consuming campuses' powers ($\beta = 0.31,\ p = 9.48 \times 10^{-14}$), the UAE national students ($\beta = -0.25,\ p = 0.048$), and the mean social perceptions of students ($\beta = 0.22,\ p \approx 0$), and the current actions to save home's energy ($\beta = 0.09,\ p = 0.04$). But, the rest of variables took a tiny part in the model since their p-values were larger than 0.05. As seen from last two linear models, we noticed two important things: First, the significance of relationship between current home actions and current on-campus actions implies that what predicts the first predicts also the latter. However, the predictors obtained in every model do not agree with this fact. Second, the motivation rate at each area (home or campus) between student could have a massive impact on their amount of deeds to make the power sustainable as every increase in home and on-campus motivation implies an increase in actions to conserve energy at home and on-campus by 0.37 and 0.4, respectively.

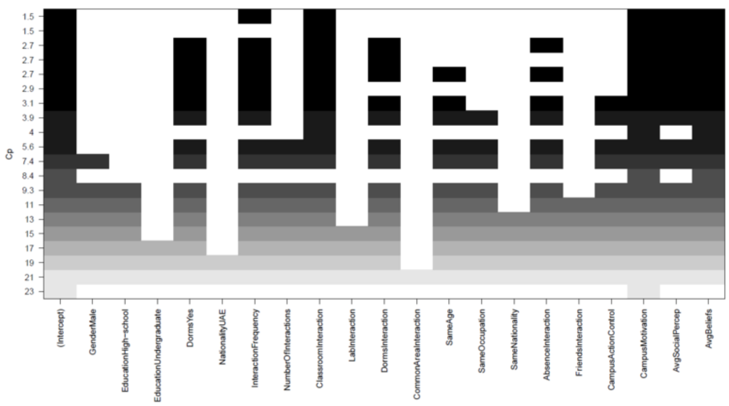
On the other side, the VIFs of predictors and the models were less than 5, thence there was no issue with multicollinearity in our multiple linear models.

C. Subset Selection

The four models on students' motivation and conserving energy actions at home and on campus were tested by backward subset selection. Results of models 1 to 4 are presented in Figures 16 to 20 respectively. The left panel of the figures shows the C_p values for all the regression models with different numbers of predictors. The model with the lowest C_p value (i.e. the one at the top of the graph) is chosen to be the best model. Furthermore, multiple linear regression is fit using the predictors in the best models (i.e. the predictors that are shaded in black) and the analysis is presented in the right panel. It can be observed that the R^2_{adj} values for the subset selected models are slightly higher than the multiple linear regression models in part B model. This indicates an improvement in the model's ability to predict the response variables and to eliminate irrelevant predictors.

When students' motivation to conserve energy is considered as a response variable (Model 1 and 2), students' motivation at home and on campus are dual predictors of each other. A similar observation can be seen for students' energy consumption actions at home and on campus (Model 3 and Model 4). This suggests that regardless of the location students are present in, students that are more motivated and take conservation actions in a specific place are likely to be motivated and take action in another place. However, a closer inspection of models 3 and 4 indicates that students' motivation at home was not a significant predictor of students' conservation patterns on campus and students' motivation on campus was not a significant predictor of students' conservation patterns on campus. An interpretation could be that students' energy consumption patterns at a specific location are strongly connected to their motivation to conserve energy at that specific location rather than their overall motivation to conserve energy. Furthermore, students' ability to impact energy conservation on campus (CampusActionControl) was a strong predictor of campus motivation signaling that the more control students have on saving energy on campus the more motivated they are to save energy.

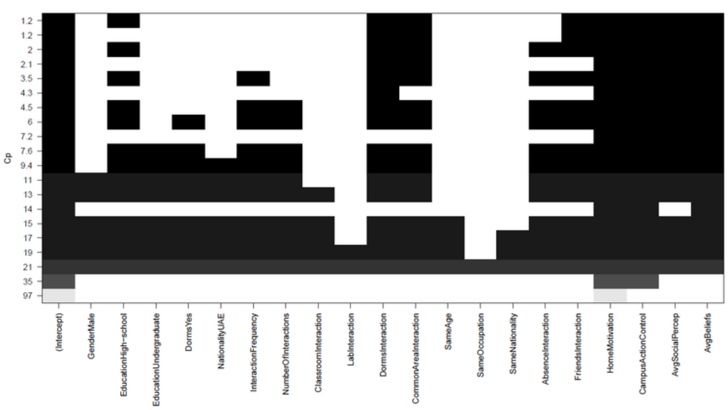
Other strong predictors that appear in almost all models are students' individual beliefs on saving energy and the social perception of saving energy of the people around them. Considering motivation as the response variable, both predictors seem significant ($p < 0.05$) were students that believe strongly in the benefits of saving and energy and have a positive support system are more likely to be motivated to save energy ($\beta_{Beliefs}, \beta_{SocialPercep} > 0$). However, exploring students' conservation actions as response variables, we see that student's beliefs are strongly connected to their home consumption ($p_{Beliefs} < 0.05,\ p_{SocialPercep} > 0.05$) while students' social perception affects their energy consumption actions on campus ($p_{SocialPercep} < 0.05$ and the Beliefs variable was excluded by subset selection). This presents an interesting result where an explanation could be that people act upon their motivations depending on the people around them and stick to their beliefs when alone. Moreover, interaction with people variables, such as Friends Interaction ($p_{Model2} > 0.1,\ \beta < 0$), Classroom Interactions ($p_{Model2} < 0.1,\ \beta > 0$), and Dorm Interaction ($p_{Model2} < 0.05,\ \beta < 0$), appear in some of the models yet do not appear significant to students' motivation and consumption patterns. This may be due to collinearity and that the real effects of the interaction variables are hidden in students' beliefs and social perception variables. Finally, there is some evidence that suggests that Nationals conserve energy less on campus($p_{Model4} < 0.1,\ \beta < 0$), students with an undergraduate degree consume less than students with a graduate degree at home ($p_{Model3} < 0.1,\ \beta > 0$) and students in the dorms conserve less at their homes ($p_{Model2} < 0.05,\ \beta < 0$).




```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           0.65565    0.31122   2.107 0.035626 *
InteractionFrequency  0.06019    0.04226   1.424 0.154984
ClassroomInteraction  0.08014    0.04355   1.840 0.066305 .
CampusMotivation      0.32548    0.03770   8.634  < 2e-16 ***
AvgSocialPercep       0.12054    0.06051   1.992 0.046897 *
AvgBeliefs            0.22130    0.06088   3.635 0.000306 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.885 on 513 degrees of freedom
Multiple R-squared:  0.2342,    Adjusted R-squared:  0.2267
F-statistic: 31.37 on 5 and 513 DF,  p-value: < 2.2e-16
```

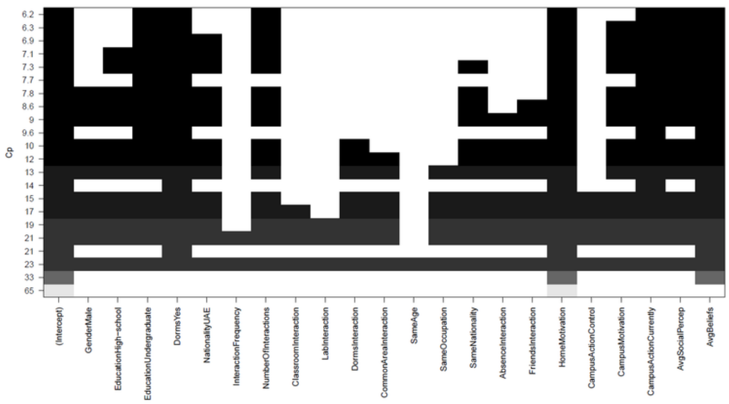
Figure 16: Subset Selection Results on Model 1 with Home Motivation as the response variable




```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             -0.21819    0.42209  -0.517  0.60543
EducationHigh-school    -0.30349    0.27107  -1.120  0.26341
EducationUndergraduate -0.16360    0.28973  -0.565  0.57256
FriendsInteraction      -0.07415    0.04559  -1.626  0.10451
CommonAreaInteraction    0.09490    0.03810   2.491  0.01306 *
DormsInteraction        -0.05829    0.02606  -2.237  0.02572 *
HomeMotivation           0.32854    0.04329   7.589 1.55e-13 ***
CampusActionControl      0.28800    0.04251   6.774 3.47e-11 ***
AvgSocialPercep          0.18606    0.06571   2.831  0.00482 **
AvgBeliefs               0.29005    0.06316   4.593 5.52e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9202 on 509 degrees of freedom
Multiple R-squared:  0.3357,    Adjusted R-squared:  0.3239
F-statistic: 28.58 on 9 and 509 DF,  p-value: < 2.2e-16
```

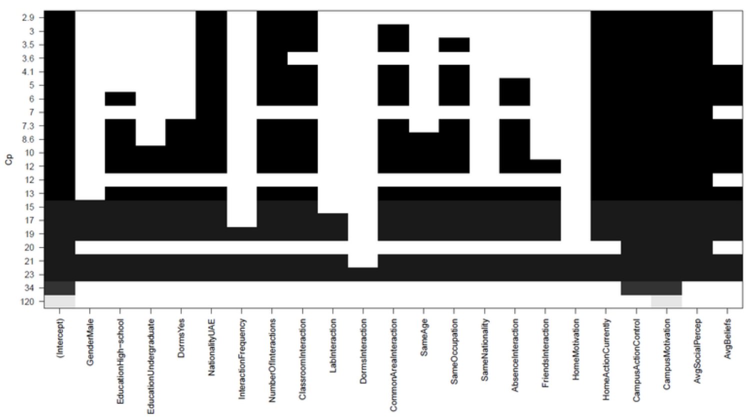
Figure 17: Subset Selection Results on Model 2 with Campus Motivation as the response variable




```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.07102    0.34510   3.103 0.002019 **
EducationHigh-school    0.22651    0.23911   0.947 0.343934
EducationUndergraduate  0.46311    0.25383   1.824 0.068665 .
DormsYes               -0.25919    0.07197  -3.601 0.000347 ***
NumberOfInteractions   -0.07114    0.03665  -1.941 0.052794 .
HomeMotivation          0.38090    0.03841   9.917  < 2e-16 ***
CampusActionCurrently   0.12617    0.03551   3.553 0.000416 ***
AvgSocialPercep        -0.11034    0.05668  -1.947 0.052114 .
AvgBeliefs              0.30047    0.05578   5.387  1.1e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8054 on 510 degrees of freedom
Multiple R-squared:  0.3255,	Adjusted R-squared:  0.3149
F-statistic: 30.76 on 8 and 510 DF,  p-value: < 2.2e-16
```

Figure 18: Subset Selection Results on Model 3 with Home Actions as response variable




```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -0.17148    0.27104  -0.633  0.52723
NationalityUAE        -0.28428    0.11225  -2.533  0.01162 *
NumberOfInteractions   0.07210    0.03992   1.806  0.07148 .
ClassroomInteraction   0.04159    0.04125   1.008  0.31382
HomeActionCurrently    0.11196    0.04222   2.652  0.00826 **
CampusMotivation       0.50625    0.03784  13.379  < 2e-16 ***
AvgSocialPercep        0.35824    0.05923   6.049 2.82e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8687 on 512 degrees of freedom
Multiple R-squared:  0.424,	Adjusted R-squared:  0.4172
F-statistic: 62.81 on 6 and 512 DF,  p-value: < 2.2e-16
```

Figure 19: Subset Selection Results on Model 4 with Campus Actions as response variable

D. Ridge Regression

When fitting the four multiple linear regression models with a ridge regression model, the results show that multiple linear regression outperformed ridge regression. We suspect the reasons behind this are:

1. Data splitting has not been incorporated into the analysis of MLR and Ridge Regression. The ridge regression may be doing worse on the data analyzed, but do better on other data.

2. Predictors analyzed are all measured on a Likert Scale. Therefore, the predictor coefficients values are low, and thus their l_2 norm would not have a significant impact in the regularization term of the minimization problem.

Nevertheless, we present the results of ridge regression on model 1. The optimal tuning parameter value was found to be $\lambda_{Optimal} = 0.316$. $R^2_{Ridge} = 0.240$ which is lower than the $R^2_{MLR} = 0.251$. The coefficients of the predictors are presented below.

```
        (Intercept)           GenderMale  EducationHigh-school EducationUndergraduate            DormsYes         NationalityUAE
        0.979914373          -0.043428902          0.028017033          0.006972957          0.071279748          0.002403547
InteractionFrequency  NumberOfInteractions  ClassroomInteraction       LabInteraction     DormsInteraction  CommonAreaInteraction
        0.064985452          -0.009000265          0.068734714          0.016114008         -0.026666870          0.014734259
            SameAge        SameOccupation      SameNationality    AbsenceInteraction   FriendsInteraction   CampusActionControl
        0.051976395          -0.029522339         -0.013764861         -0.048888204          0.008061828          0.065283372
   CampusMotivation       AvgSocialPercep           AvgBeliefs
        0.232932747           0.105752537          0.194609855
```

Figure 20: Ridge Regression coeffecients of Model 1

E. PCA

To further explore the factors of variation of different responses, we determined the optimal number of PCs by doing a screeplot as in Figure 21, with the optimal analysis of the model using varimax rotation as in Figure 22.

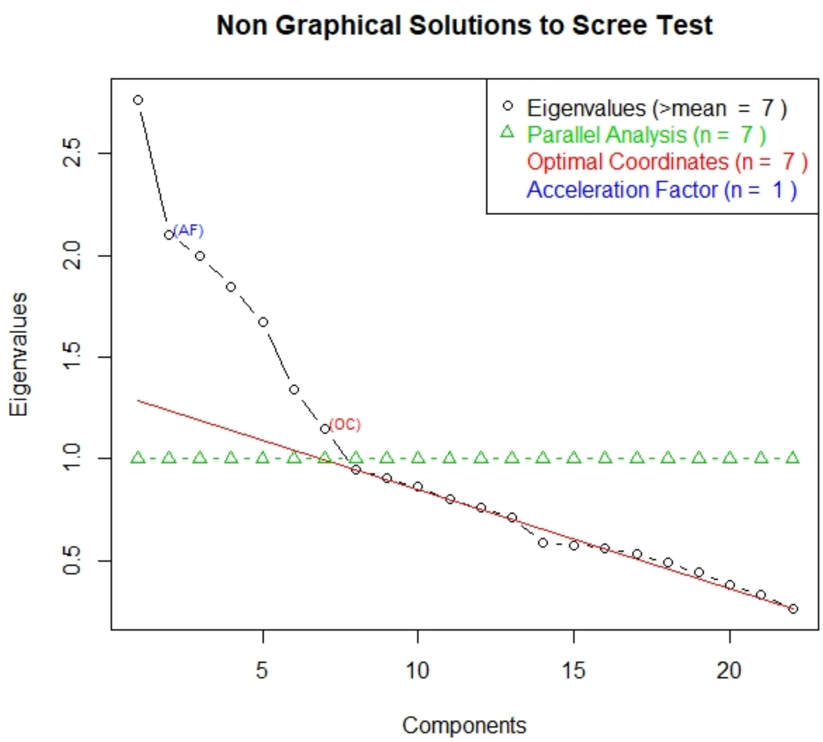


Figure 21: The optimal number of PCs

As seen in Figure 21, the optimal number (coordinates) of PCs needed is 7, which is as same as the number of eigenvalues more than one, thus we will stick with a 7 PCs model.

```
Standardized loadings (pattern matrix) based upon correlation matrix
                          RC2   RC1   RC3   RC4   RC6   RC5   RC7   h2   u2 com
Gender                   -0.12  0.09 -0.23 -0.27  0.08 -0.08  0.46 0.37 0.63 2.6
Education                -0.14  0.01  0.07  0.72  0.05 -0.03  0.07 0.55 0.45 1.1
Dorms                     0.03 -0.11  0.01  0.07  0.90  0.01 -0.08 0.84 0.16 1.1
Nationality               0.03 -0.06  0.04  0.74  0.15  0.02 -0.14 0.60 0.40 1.2
InteractionFrequency      0.07  0.59  0.09  0.13 -0.03 -0.23  0.43 0.63 0.37 2.4
NumberOfInteractions     -0.07 -0.35  0.14 -0.31 -0.02  0.06 -0.04 0.25 0.75 2.5
ClassroomInteraction     -0.02  0.79  0.10 -0.14 -0.05  0.16 -0.03 0.69 0.31 1.2
LabInteraction           -0.01  0.81  0.03 -0.10  0.07  0.06  0.02 0.68 0.32 1.1
DormsInteraction          0.07  0.14  0.00  0.15  0.88 -0.02  0.11 0.84 0.16 1.2
CommonAreaInteraction     0.19  0.44  0.08  0.24 -0.05 -0.27  0.41 0.54 0.46 3.8
SameAge                  -0.01 -0.01  0.01 -0.02  0.03  0.75  0.21 0.60 0.40 1.2
SameOccupation            0.06  0.08  0.02 -0.02 -0.04  0.76 -0.04 0.59 0.41 1.0
SameNationality           0.11 -0.09 -0.13  0.47  0.00  0.50  0.22 0.56 0.44 2.7
AbsenceInteraction        0.03  0.05  0.09 -0.04  0.03  0.12  0.67 0.48 0.52 1.1
FriendsInteraction        0.07  0.01  0.00  0.13 -0.04  0.27  0.68 0.56 0.44 1.4
CampusActionControl       0.60 -0.11  0.05 -0.12  0.15  0.02  0.00 0.41 0.59 1.3
BeliefSociety            -0.02  0.04  0.84 -0.10 -0.01 -0.01  0.07 0.72 0.28 1.0
BeliefEnvironment         0.02  0.02  0.84 -0.05  0.02 -0.03  0.03 0.71 0.29 1.0
BeliefMoney               0.13  0.07  0.66  0.23  0.02  0.00 -0.09 0.53 0.47 1.4
SocialPercepFriends       0.76  0.02 -0.03  0.05 -0.04  0.09 -0.01 0.60 0.40 1.0
SocialPercepColleagues    0.77  0.05 -0.01 -0.04 -0.02  0.03  0.07 0.61 0.39 1.0
SocialPercepNeighbours    0.70  0.14  0.13  0.08  0.02 -0.05  0.02 0.54 0.46 1.2

                       RC2  RC1  RC3  RC4  RC6  RC5  RC7
SS loadings           2.16 2.05 2.00 1.68 1.67 1.65 1.64
Proportion Var        0.10 0.09 0.09 0.08 0.08 0.08 0.07
Cumulative Var        0.10 0.19 0.28 0.36 0.43 0.51 0.58
Proportion Explained  0.17 0.16 0.16 0.13 0.13 0.13 0.13
Cumulative Proportion 0.17 0.33 0.48 0.61 0.74 0.87 1.00

Mean item complexity =  1.5
Test of the hypothesis that 7 components are sufficient.

The root mean square of the residuals (RMSR) is  0.07

Fit based upon off diagonal values = 0.79
```

Figure 22: The summary of rotating loadings in the PCA model

We had 7 principal components (PC1 is RC2, PC2 is RC1, PC3 is RC3, etc.), each PC has a certain proportion of variability in the survey data. By following the pattern of variance explained in the PCs, we noticed that it decreases through PCs. The first PC could explain 10% of variability in data and the last one could just have 7% of it. Overall, the 7 PCs could explain 58% of the variability in the study's data. The sum of squares loadings range was between 1.64 and 2.16 (bigger than 1), which implies the sufficiency of having each of seven principal components in the model.

The rotating correlation matrix between variables and PCs shows different loads of variables on each PC. In most cases, the variables with the highest positive or negative correlation with a certain PC has the higher load on that PC. A visual interpretation of this will be introduced in the loadings diagram in Figure 23.

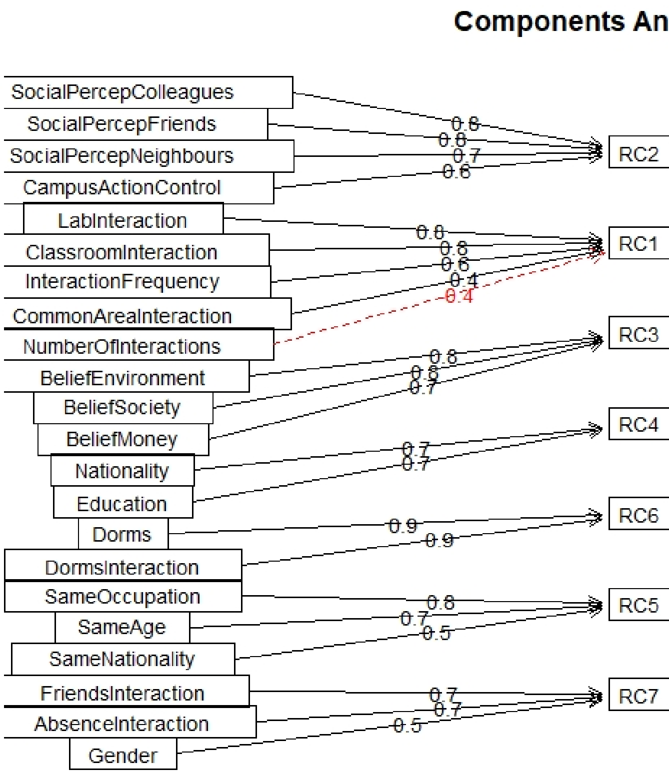


Figure 23: The highest loadings of each variable on the rotating version of PCA model

For the sake of explaining loads on the optimal PCA model we have, we assumed that all people who have a passion and do action to save power are initiators to conserve it. As long as we had 7 PCs, thus we divided the initiators into 7 types: Type A, type B, type C, type D, type E, type F, and type G.

The first type of proactive people to save power was represented by PC1, which was explaining 10% of data's variance and was highly loaded by the control on campus consumption, in addition to social perception about friends, colleagues, and neighbors conservation of energy. The type A of people believes that their friends, mates, and neighbors are doing their best to conserve energy. From this belief, they want to be like them and they control their consumption on energy at KU campuses by opening windows or closing computers when finishing their sessions.

The interactions in labs, common zones, and classrooms, then the frequency of contact, had the most positive loads on PC2, which explained 9% of data's variance. This PC represented the type B of initiators, and this group prefers to frequently interact with their colleagues at campus, especially in labs, classrooms, and common areas such as restaurants and public sitting zones. During conversations, they consider gathering information about energy consumption from their social environment at campus to form a good background about how to conserve the earth's power.

In the same PC, there was a significant negative load from students' number of interactions on it. The same type of proactive students was also characterized as a group who has a low daily number of interactions. This low amount of daily contacts on campus might be good for them as the chance of getting bad or wrong advises from non-expert people decreases. Consequently, they will feel more trustworthy about their own strategies to conserve power as long as they are based on experts' recommendations.

The beliefs about the impact of saving energy on money, society and environment had the top loads on PC3, which represented about 9% of data's variability. This PC illustrates that the type C of initiators strongly thinks that the benefits of conserving power are: Saving environment and keeping it free from pollution, lowering money spent for paying energy bills, and not causing any damage or noise to the surrounding humans and other living creatures. This belief gives these people a passion to do actions to conserve power.

The education and nationality had the top load on PC4, which had 8% of data's variability. The type D of initiators have a high education level as each improvement in level of education leads to more consciousness on energy and the best ways to use it.

Regarding nationality, we can say that it does not rely on whether the student is local or non-local, but it is all about manners and principals he or she arose on [7], [8]. So, the customs and socio-cultures may play a role in forming the behavior of consuming energy at home or on-campus. However, the type D considers only the good customs and ethics to save earth by reducing the unnecessary use of energy.

The dorms interactions, in addition to dorms, loaded significantly on PC5, that formed 8% of variance in this data. The same PC, which is called type E of proactive people, involved other students who are more comfortable with living in KU dorms and like to interact with their colleagues there.

The PC6, which explained 8% of variability in the survey data, was largely loaded by the interaction with people with same age, with same nationality, or with same job occupation. Obviously, the type F is likely in favor of contacting their colleagues from students who are in the same age range as this type, regardless of the place of interactions occurred, and the colleagues who like to interact with should be from the same nationality as them. In details, the local students contact their local ones and the non-local pupils like to interact with their non-local mates.

In both PCs 5 and 6, the proactive types mentioned consider gathering information about energy consumption from their colleagues and other people around them to have a knowledge about power conservation concept or may find friends who can convince them to actually save power to save the earth.

Finally, the PC7, with a 7% of variance explained, had a significant load from gender, friends interactions and absence interactions on this PC, and a significant negative load from education on the same PC. The last type of initiators in this PCA model (type G), which was represented by this PC, is in favor of having contacts with their friends and other people who communicate without the appearance of them, regardless of the place of contact. The same type also has a high similarity in rate of consuming energy between males and females.

The power consumption pattern does not often depend on gender. However, we assumed that men students are consuming power more than female students based on a report done by Räty and Carlsson-Kanyama [9], which concluded that "men used more energy than women." This result was applicable when comparing some European men with some European women in the rate of using different power sources, such as transportation and food, and this could be a rational reason for the effect of differences between genders on students' motivation and energy to conserve power (this effect in the study we have was proved in the descriptive statistics section). Type G basically does not have this hole between genders and they have approximately the same behaviors to conserve energy. This case may appear because all of the students under this kind are equally educated about the notions of energy in their different educational levels.

F. Predictive Power Score

Fig. 24 presents a heat map for the predictive power scores of all the predictors studied in this project. The y-axis represents the variables as response variables and the x-axis as predictors. By examining the row entries of the variables of interest, we can figure out the best predictors or features that explain those variables. Three interesting observations are made upon that examination:

1. **Motivation and Energy Consumption Action Variables:** The strongest predictors of students' motivation to conserve energy at home are students' energy consumption at home and their motivation to

save energy on campus ($PPS > 0.19$) while the strongest predictors of students' motivation to conserve energy on campus are students' ability to impact energy conservation on campus and their current energy consumption patterns on campus ($PPS > 0.2$). As for the strongest features that predict students' energy consumption at home and on campus, students' motivation to conserve energy at home and on campus, students' energy consumption patterns at home and on campus, and students' control over conserving energy on campus all had PPS values greater than 0.2. This solidifies the results from the multiple linear regression and subset selection models. Students' social perception and beliefs on conserving energy were also found to be predictors of their motivations and consumption averaging around a $PPS \approx 0.15$ for the variables. Furthermore, almost all other variables such as student's interactions and control variables have $PPS > 0$ signifying that the complexity of predicting such response variables.

2. **Belief and Social Perception Variables:** When looking at students' social perception variables with regards to friends and colleagues, we see that most students' interaction with people variables predicts them to some extent ($PPS > 0.1$). We do not see that with the neighbor social perception variable because there were no questions in the survey on students' interactions with neighbors. A similar result is found for students' belief variables with students' interactions ($PPS > 0.1$) which implies that students' beliefs are shaped by their interactions with friends and colleagues in different locations. However, students' belief that conserving energy saves the environment was an exception as most interaction variables had PPS values of 0. An explanation could be the clarity that saving energy saves the environment that students' interactions with people do not change their mindset.

3. **Longitudinal variables of energy consumption:** Student's consumption patterns at home and on campus six and twelve months prior were added to see whether removing them was a good idea. It can be observed that these variables are the strongest predictors of themselves ($PPS > 0.5$) and seem inconsistent in predicting students' motivation and current consumption patterns ($PPS < 0.1$). This could be due to students' to remembering exactly what their energy consumption was previously and thus had similar replies six and 12 months ago.

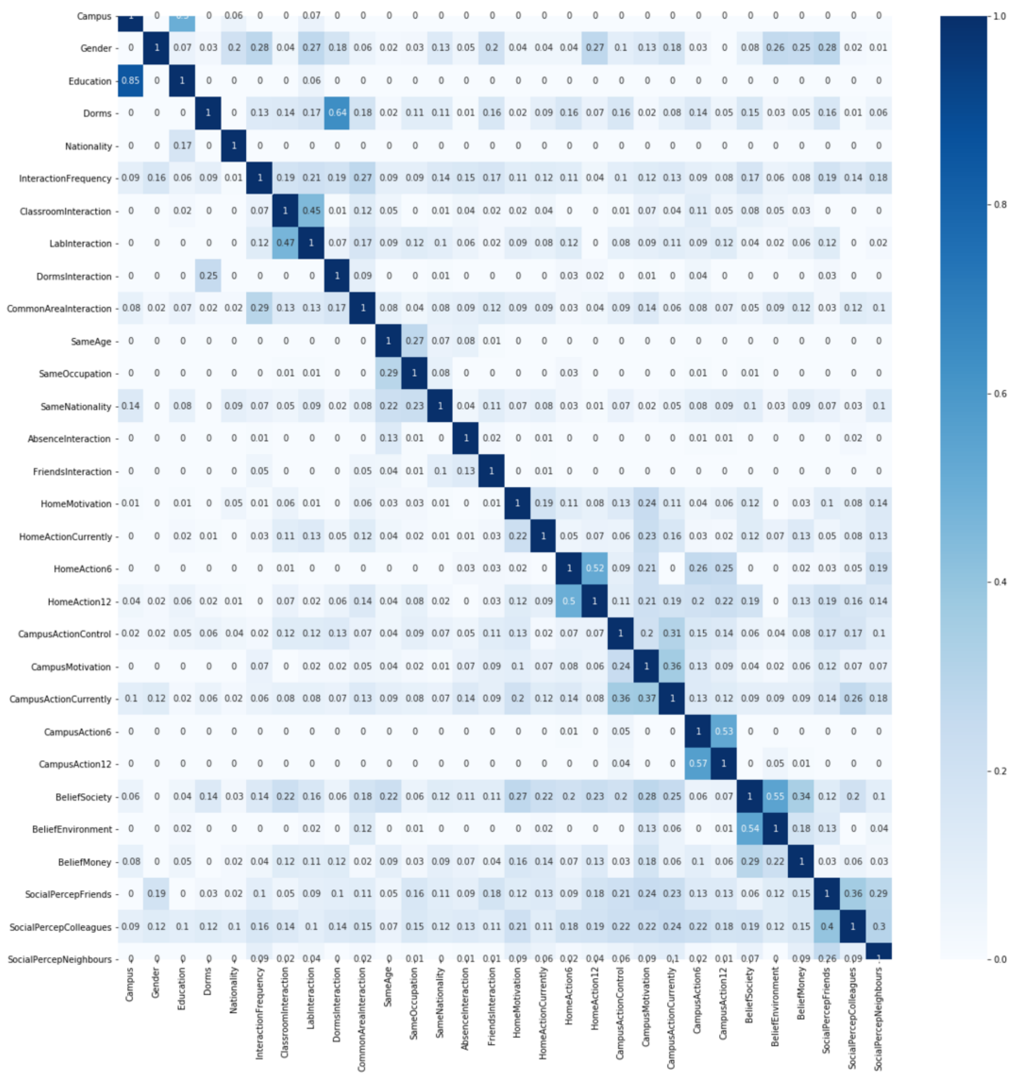


Figure 24: Predictive Power Score Heat Map of all variables studied

Limitations & Further Research

Some survey and analysis limitations present in this project include the following observations:

1. It is unknown how the survey was sampled and therefore we do not how well the survey represents the Khalifa University population. No information exists on the student's majors or undergraduate year of study. Furthermore, the age range in the survey was too wide and most of the student population are in the range of 20-29. The age ranges could be narrower to allow for analysis of age in energy consumption patterns.

2. Some variables of interest, such as personality traits, had too few questions effectively measure the real effects of the variables. There should be at least 4 or 5 well-written questions to uncover the effect of complex variables. This will allow Cronbach alpha analysis to solidify that the variables were indeed measured in the survey.

3. The survey did not indicate which year it was conducted, hence it is needed to be updated so we can know how accurate are its data and how relevant are these to this particular year. Longitudinal variables in surveys need to be measured at their respective times. For instance, for student's actions to conserve energy on campus and comparing it to 6 and 12 months ago, students may not remember exactly their energy consumption patterns at these times and thus provide inaccurate responses.

4. While VIF values for the multiple linear regression models analyzed were not concerning, we expect that there are collinearity issues with the variables. Therefore, while some variables may not appear directly significant to the response variables studied, they could be indirectly related to the response variable through unknown predictors not studied in this paper.

5. Data splitting was not incorporated into the analysis of this study due to time constraints. Therefore, different results could be observed when analyzing further future data on energy consumption.

Further research would look into refining the survey to uncover other relationships and combat the limitations in the analysis. Furthermore, R^2 values for the multiple linear regression models were significantly low, so researchers can explore alternative variables to explain students' motivations and actions to conserve energy. It would also be interesting to conduct a longitudinal study to see if the response variables change with time. Finally, other types of techniques not used in this project such as mediation analysis [10] may provide a clearer picture of students' energy consumption patterns.

Conclusion

This project aimed to understand whether there are differences in student's motivation to conserve energy and consumption patterns at home and on campus. The descriptive statistics conducted indicate that conservation levels at home are significantly higher than on campus. Averages of energy consumption show students in dorms conserve less. Results from quantitative analysis support that conclusion as it shows that indeed students were more likely to save energy and be more motivated at home than on campus. Decision makers can benefit from such result as they can focus on raising more awareness about energy conservation in public areas such as schools and universities.

Moving to the effect of variables such as personal specifications, beliefs and motivations on energy conservation. The indications did not show a significant difference between males' and females' behavior. Students with high motivation showed a good response as when the motivation is high, the actual conservation levels go up. Beliefs showed a significant impact on conservation levels at home. Social perceptions also showed an impact on students' behavior towards energy consumption. Through multiple linear regression models and subset selection, student's motivation, energy conservation beliefs, and the social perception around students were the strongest predictors of students' home and campus energy consumption. The multiple linear regression did not show precisely that the variables impact on home motivation should also impact campus motivation, and the same thing for current home actions and current campus motivations, due to the difference between predictors that mostly affect each response. The PCA approach divided people who do things to conserve energy to 8 categories. Regardless of differences in their preferences, they all try the best to reduce the massive consumption of energy nowadays.

Role of Each Member in this Report

Ahmed Saeed: Doing a quantitative analysis on the energy consumption survey data using PCA and MLR, and having a main contribution in reviewing the quantitative analysis paragraphs and the report design, and filling in the shortcomings (limitations) of the study and the appendices parts, in addition to a partial contribution in reviewing the rest of report, and writing the hypotheses, survey variables studied, and conclusions.

Hamza Riyad: (in collaboration with Omran) Descriptive statistics; creating graphs on PowerBI, commenting on the observations from the charts, assuming hypotheses that will be studied in the quantitative statistics. Writing the Introduction, descriptive analysis, conclusion, and abstract.

Begad: Worked exclusively on the Data Exploration & Imputation, Wilcoxon test, Subset selection, Ridge Regression, and Predictive Power score sections. Helped with the Multiple Linear Regression and had main contributions in the Limitations & Further Research, Variables studied, and Research Hypothesis sections and filling out the code in the Appendix.

R Code Appendix

```
library(readxl)
library(VIM)
library(psych)
library(tidyverse)
library(nFactors)
library(car)
library(psych)
library(leaps)

df <- read_excel("SurveyDataEdited.xlsx")
df <- hotdeck(as.data.frame(df))
Energy <- df[,1:44]

Energy[Energy == "Never"]      <- 1
Energy[Energy == "Rarely"]     <- 2
Energy[Energy == "Sometimes"]  <- 3
Energy[Energy == "Very often"] <- 4
Energy[Energy == "Always"]     <- 5

Energy[Energy == "Not at all"]  <- 1
Energy[Energy == "Slightly"]    <- 2
Energy[Energy == "Moderately"]  <- 3
Energy[Energy == "Very"]        <- 4
Energy[Energy == "Extremely"]   <- 5

Energy[Energy == "Strongly Disagree"]  <- 1
Energy[Energy == "Disagree"]           <- 2
Energy[Energy == "Undecided"]          <- 3
Energy[Energy == "Agree"]              <- 4
Energy[Energy == "Strongly Agree"]     <- 5

Energy[Energy == "Significantly decreased"]  <- 1
```

```
Energy[Energy == "Slightly decreased"]       <- 2
Energy[Energy == "Remained the same"]        <- 3
Energy[Energy == "Slightly increased"]       <- 4
Energy[Energy == "Significantly increased"]  <- 5

Energy[Energy == "Very Low"]  <- 1
Energy[Energy == "Low"]       <- 2
Energy[Energy == "Moderate"]  <- 3
Energy[Energy == "High"]      <- 4
Energy[Energy == "Very high"] <- 5

Energy[Energy == "Zero-to-four"]         <- 1
Energy[Energy == "Five-to-Nine"]         <- 2
Energy[Energy == "Ten-to-Ninteen"]       <- 3
Energy[Energy == "Twenty-to-Fortynine"]  <- 4
Energy[Energy == "Above 50"]             <- 5

Energy[,c(9:44)] <- data.matrix(Energy[,c(9:44)])

Socialpercep  <- Energy[,c(44,43,42)]
Beliefs       <- Energy[,c(39,40,41)]
Extraversion  <- Energy[,c(9,14)]
Agreeableness <- Energy[,c(10,15)]
Conscientious <- Energy[,c(11,16)]
Neuroticism   <- Energy[,c(12,17)]
Openness      <- Energy[,c(13,18)]
HomeAction    <- Energy[,c(31,32,33)]
CampusAction  <- Energy[,c(34,36,37,38)]

Energy <- Energy %>%
  mutate(AvgSocialPercep = (SocialPercepFriends+SocialPercepColleagues+SocialPercepNeighbours)/3) %>%
  mutate(AvgBeliefs = (BeliefSociety+BeliefEnvironment+BeliefMoney)/3)



## Cronbach Alpha
psych::alpha(Socialpercep)
psych::alpha(Beliefs)
psych::alpha(Extraversion, check.keys = TRUE)
psych::alpha(Agreeableness, check.keys = TRUE)
psych::alpha(Conscientious, check.keys = TRUE)
psych::alpha(Openness, check.keys = TRUE)
psych::alpha(Neuroticism, check.keys = TRUE)
psych::alpha(HomeAction)
psych::alpha(CampusAction)



## Part B: Multiple linear regression code
```

```
NewEnergy <- Energy[,-c(1,2,3,5,9:18,32:33,37:44)]

lmmodel1 <- lm(HomeMotivation~.-(CampusActionCurrently+HomeActionCurrently),data=NewEnergy)
summary(lmmodel1)
vif(lmmodel1)

sum((lmmodel1$residuals)^2)

lmmodel2 <- lm(CampusMotivation~.-(CampusActionCurrently+HomeActionCurrently),data=NewEnergy)
summary(lmmodel2)
vif(lmmodel2)

sum((lmmodel2$residuals)^2)

lmmodel3 <- lm(HomeActionCurrently~.,data=NewEnergy)
summary(lmmodel3)
vif(lmmodel3)

sum((lmmodel3$residuals)^2)

lmmodel4 <- lm(CampusActionCurrently~.,data=NewEnergy)
summary(lmmodel4)
vif(lmmodel4)

sum((lmmodel4$residuals)^2)



(sum((lmmodel1$residuals)^2)+sum((lmmodel2$residuals)^2)
+sum((lmmodel3$residuals)^2)+sum((lmmodel4$residuals)^2))/4


## Part C Regsubsets

# Model 1
regfit.m1 <- regsubsets(HomeMotivation~.-(CampusActionCurrently+HomeActionCurrently),
data=NewEnergy, nvmax = 25,method = "backward")

reg.summary <- summary(regfit.m1)
mincp <- which.min(reg.summary$cp)

plot(regfit.m1,scale="Cp")
coef(regfit.m1,mincp)

newlm1 <- lm(HomeMotivation~InteractionFrequency+ClassroomInteraction+
CampusMotivation+AvgSocialPercep+AvgBeliefs, data = NewEnergy)
summary(newlm1)

# Model 2
regfit.m2 <- regsubsets(CampusMotivation~.-(CampusActionCurrently+HomeActionCurrently),
```

```
data=NewEnergy, nvmax = 25,method = "backward")

reg.summary <- summary(regfit.m2)
mincp <- which.min(reg.summary$cp)

plot(regfit.m2,scale="Cp")
coef(regfit.m2,mincp)

newlm2 <- lm(CampusMotivation ~ Education + FriendsInteraction + CommonAreaInteraction +
DormsInteraction + HomeMotivation + CampusActionControl +AvgSocialPercep+AvgBeliefs,data = NewEnergy)
summary(newlm2)

# Model 3
regfit.m3 <- regsubsets(HomeActionCurrently~.,data=NewEnergy, nvmax = 25,method = "backward")

reg.summary <- summary(regfit.m3)
mincp <- which.min(reg.summary$cp)

plot(regfit.m3,scale="Cp")
coef(regfit.m3,mincp)

newlm3 <- lm(HomeActionCurrently ~ Education + Dorms + NumberOfInteractions +
CampusActionCurrently + AvgSocialPercep + AvgBeliefs,data = NewEnergy)
summary(newlm3)

# Model 4
regfit.m4 <- regsubsets(CampusActionCurrently~.,data=NewEnergy, nvmax = 25,method = "backward")

reg.summary <- summary(regfit.m3)
mincp <- which.min(reg.summary$cp)

plot(regfit.m4,scale="Cp")
coef(regfit.m4,mincp)

newlm4 <- lm(CampusActionCurrently ~ Nationality + NumberOfInteractions +
ClassroomInteraction + HomeActionCurrently + CampusMotivation + AvgSocialPercep,data = NewEnergy)
summary(newlm4)

# Part A Wilcoxon Test

# ANOVA did not work because data isn't normal (Likert Scale)

Motivation <- tibble(
  MotivationValues = c(Energy$HomeMotivation,Energy$CampusMotivation),
  Location = c(rep("Home",nrow(Energy)),rep("Campus",nrow(Energy)))
)

Action <- tibble(
  ActionValues = c(Energy$HomeActionCurrently,Energy$CampusActionCurrently),
  Location = c(rep("Home",nrow(Energy)),rep("Campus",nrow(Energy)))
)
```

```
fit <- aov(MotivationValues~Location,Motivation)
summary(fit)

# t-test trial
t.test(Energy$HomeMotivation,Energy$CampusMotivation, alternative = "two.sided", var.equal = TRUE)


# Are home & campus motivations any different?
leveneTest(MotivationValues~Location,Motivation)

shapiro.test(Motivation$MotivationValues)

wilcox.test(MotivationValues ~ Location, data = Motivation,
                   paired = TRUE,alternative = "less")

describeBy(Motivation$MotivationValues, Motivation$Location)

# Are home & campus actions different?

leveneTest(ActionValues~Location,Action)

shapiro.test(Action$ActionValues)

wilcox.test(ActionValues ~ Location, data = Action,paired = TRUE,alternative = "less")

describeBy(Action$ActionValues, Action$Location)




## Part D Ridge Regression

GoodEnergy <- NewEnergy[,-c(21,25:30)]
x = model.matrix(HomeMotivation~.,GoodEnergy)[,-1]
y = GoodEnergy$HomeMotivation

library(glmnet)
lambdas <- 10^seq(10, -2, by = -.1)
fit <- glmnet(x, y, alpha = 0, lambda = lambdas)
summary(fit)

cv_fit <- cv.glmnet(x, y, alpha = 0, lambda = lambdas)
opt_lambda <- cv_fit$lambda.min
y_predicted <- predict(fit, s = opt_lambda, newx = x)

# Sum of Squares Total and Error
sst <- sum((y - mean(y))^2)
sse <- sum((y_predicted - y)^2)

# R squared
```

```
rsq <- 1 - sse / sst
rsq

# Coeffecients
out = glmnet(x,y,alpha = 0)
predict(out,type="coefficients",s = opt_lambda)[1:21 ,]


## Part E: PCA code
library(clusterSim)
library(factoextra)
GoodEnergy <- read.csv("EnergyData.csv")
GoodEnergy <- GoodEnergy[,-c(1,2,3,4,6,23,24,28,29,36,37,38,39)]
GoodEnergy$Gender <- as.numeric(GoodEnergy$Gender)
GoodEnergy$Gender[GoodEnergy$Gender == 1] <- 0
GoodEnergy$Gender[GoodEnergy$Gender == 2] <- 1
GoodEnergy$Education <- as.numeric(GoodEnergy$Education)
GoodEnergy$Education[GoodEnergy$Education == 2] <- 0
GoodEnergy$Education[GoodEnergy$Education == 3] <- 2
GoodEnergy$Education[GoodEnergy$Education == 0] <- 3
GoodEnergy$Dorms <- as.numeric(GoodEnergy$Dorms)
GoodEnergy$Nationality <- as.numeric(GoodEnergy$Nationality)
GoodEnergy$NumberOfInteractions <- as.numeric(GoodEnergy$NumberOfInteractions)
#
GoodEnergy.homemot <- GoodEnergy$HomeMotivation
GoodEnergy$HomeMotivation <- NULL
GoodEnergy.campmot <- GoodEnergy$CampusMotivation
GoodEnergy$CampusMotivation <- NULL
GoodEnergy.homeactcur <- GoodEnergy$HomeActionCurrently
GoodEnergy$HomeActionCurrently <- NULL
GoodEnergy.campactcur <- GoodEnergy$CampusActionCurrently
GoodEnergy$CampusActionCurrently  <- NULL
#
(GoodEnergycor <- cor(GoodEnergy))
eigen(GoodEnergycor)
#finding the optimal number of PCs
plotnScree(nScree(x=eigen(GoodEnergycor)$values))
(GoodEnergy.pca <- principal(GoodEnergycor,nfactors=8,residuals = T,rotate = "varimax"))
fa.diagram(GoodEnergy.pca)

## Part F Predictive Power Score

import pandas as pd
import numpy as np
import ppscore as pps
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("SurveyDataEdited.csv")

df
df_matrix = np.around(pps.matrix(df),decimals=2)
```

```
fig, ax = plt.subplots(figsize=(20,20))

sns.heatmap(df_matrix, vmin=0, vmax=1, cmap="Blues", annot=True,ax=ax)
plt.savefig('energy_exploration.png')
```

References

[1]D. Bonino, F. Corno, and L. D. Russis, "Home energy consumption feedback: A user survey", *Energy and Buildings*, vol. 47, pp. 383–393, Apr. 2012, doi: 10.1016/j.enbuild.2011.12.017.

[2]A. Crossman, "How Sociological Reference Groups Shape Our Thoughts and Behavior". https://www.thoughtco.com/reference-group-3026518.

[3]G. K. Uyanık and N. Güler, "A Study on Multiple Linear Regression Analysis", *Procedia - Social and Behavioral Sciences*, vol. 106, pp. 234–240, Dec. 2013, doi: 10.1016/j.sbspro.2013.12.027.

[4]"R-Squared - Definition, Interpretation, and How to Calculate". https://corporatefinanceinstitute.com/resources/knowledge/other/r-squared/.

[5]G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning.* Springer New York, 2013.

[6]F. Wetschorek, "RIP correlation. Introducing the Predictive Power Score". https://towardsdatascience.com/rip-correlation-introducing-the-predictive-power-score-3d90808b9598.

[7]M. Bonera, E. Corvi, A. Codini, and R. Ma, "Does Nationality Matter in Eco-Behaviour?", *Sustainability*, vol. 9, no. 10, p. 1694, Sep. 2017, doi: 10.3390/su9101694.

[8]C. Simga-Mugan, B. A. Daly, D. Onkal, and L. Kavut, "The Influence of Nationality and Gender on Ethical Sensitivity: An Application of the Issue-Contingent Model", *Journal of Business Ethics*, vol. 57, no. 2, pp. 139–159, Mar. 2005, doi: 10.1007/s10551-004-4601-z.

[9]R. Räty and A. Carlsson-Kanyama, "Comparing energy use by gender, age and income in some European countries". https://www.compromisorse.com/upload/estudios/000/101/foir2800.pdf, 2009.

[10]F. Greb, J. Steffens, and W. Schlotz, "Modeling Music-Selection Behavior in Everyday Life: A Multilevel Statistical Learning Approach and Mediation Analysis of Experience Sampling Data", *Frontiers in Psychology*, vol. 10, Mar. 2019, doi: 10.3389/fpsyg.2019.00390.