STAT1006

Regression and Nonparametric Inference

Semester 2, 2023

Final Project Report

Influence of Lifestyle Factors on Sleep

Atharva Dengle (20376888)

Bachelor of Advanced Science (Data Science) (Honours)

Declaration

The work presented in this report is my own work and all references are duly acknowledged.

This work has not been submitted, in whole or in part, in respect of any academic award at Curtin University or elsewhere.

Atharva Anand Dengle (Signature)

Atharva Anand Dengle

20/10/2023

Contents

1. Introduction to the Dataset

1.1 Overview

Sleeping is a vital activity that directly impacts our physical and mental health. As such, it is crucial to understand what contributing factors affect sleep to optimise them.

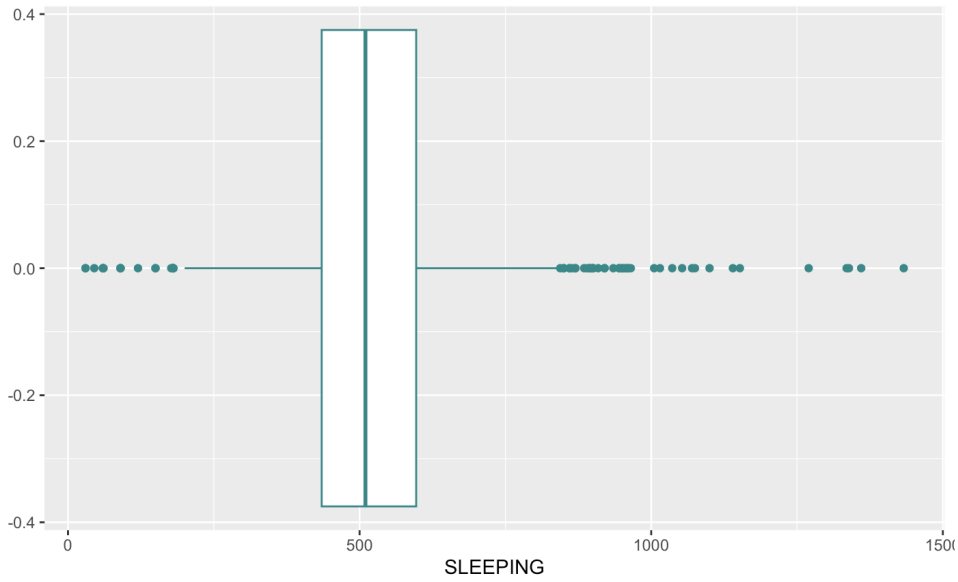
The "sleep" dataset is derived from the American Time Use Survey. It provides a list of 21 variables (activities) completed the previous day by survey participants, which are then compared to determine their influence on the response variable "SLEEPING". It contains three categorical variables, which include the highest education received, employment status and sex of the respondent, with the rest being numerical.

1.2 Exploratory Data Analysis (EDA)

Before beginning the regression analysis, viewing and understanding the dataset is important. As such, the variables and their plots require to be examined. The first six rows of data were viewed using the head function to see the presentation of the columns and rows (Appendix A). A further summary of the dataset was viewed (Appendix B) and displayed the minimum, median, mean, maximum and quantile values. Similarly, the 'structure' function lists all the column names, their data type and the first few observations (Appendix C).

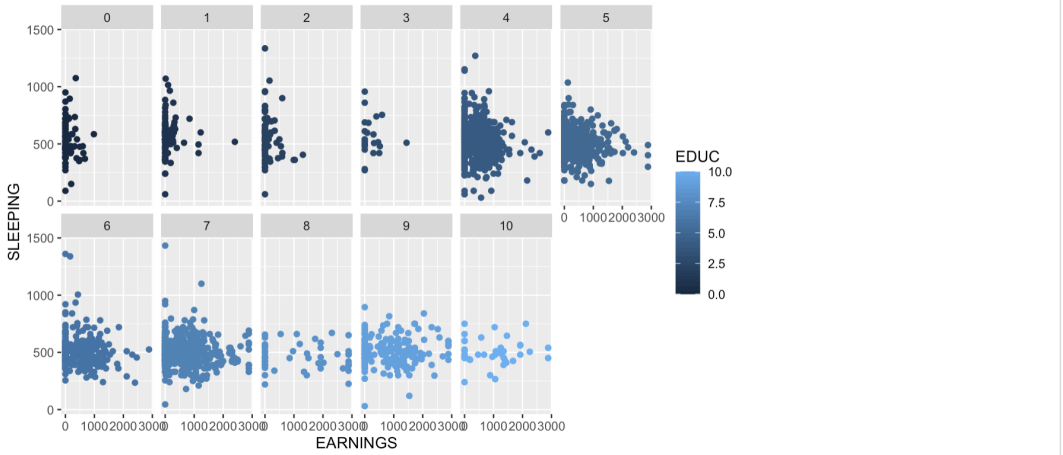
The response variable, SLEEPING, was also viewed to see its distribution, as shown by the boxplot below. The distribution is highly skewed towards the right, with its mean centered around 520 minutes. There are extreme outliers at both ends, with the ones above 1000 contributing the most to its skewness.

```{r}
library(ggplot2)
ggplot(sleep, aes(x=SLEEPING)) + geom_boxplot(color = "cyan4")
```



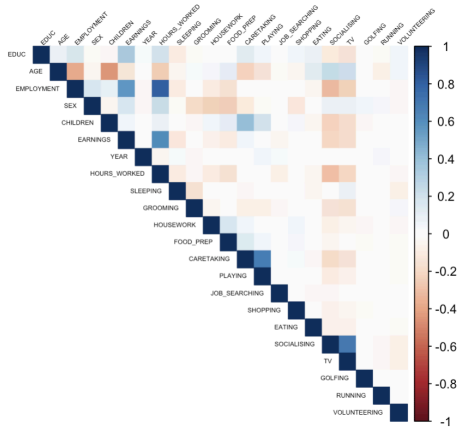
Subgraphs of different participant groups were also produced upon using the variables. It factored in the education level of the survey participants and compared their respective earning levels and sleep. As seen below, the lower levels of education (indicated by darker shades of blue) tend to have more varying levels of sleep with limited earnings. In graphs 5,6 and 7, the sleep is more centred around 500 minutes (approximately 8 hours), with a few outliers. A general pattern can be deduced: as the education level increases, the hours slept vary less, while the earnings increase.

```{r}
ggplot(data = sleep) + geom_point(mapping = aes(x = EARNINGS, y = SLEEPING, color = EDUC)) + facet_wrap(~EDUC,
    nrow = 2)
```



The EDA was resumed by checking the correlation of the explanatory variables to show which variables have a strong relationship with sleeping. The Pearson R correlation was the most appropriate for a linear model as it is used for continuous variables. The highest correlated variable would suggest that any changes to it would subsequently cause a change in sleep. Appendix D is a correlation chart that displays the Pearson correlation and the graph plots.

```{r}

suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(corrplot))

Corr <- cor(sleep)
corrplot(Corr, method = "color", type = "upper", tl.col = "black",
         tl.srt = 45, tl.cex = 0.4)
```



A better representation can be viewed (left) that uses R functions to create a mosaic-like plot displaying the strongly correlated variables as darker colours and the weaker correlations as lighter colours. It can be noticed that "GROOMING" is the darkest variable relating to sleeping hence, having the strongest correlation amongst the rest of the variables. The exact value of the pearson R correlation is viewed to be -0.1522 approximately as shown in Appendix E.

2. Best Variable Selection and Simple Linear Regression (SLR)

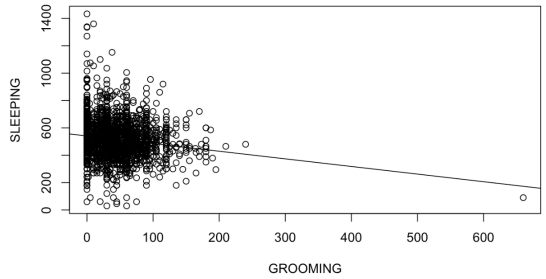
2.1 Variable Selection

To select a variable that would fit well with sleep, not only was the highest correlated variable selected, but the best subset method was also used for variable selection to verify. The best subset method calculates the optimal model based on the predictor variables that give the best RSS value. This procedure would be a model that minimises the residual sum of squares of the error. Upon analysing the summary of the results, "GROOMING" is the most highly correlated and best subset variable (Appendix F) that is selected for linear regression.

2.2 Simple Linear Regression

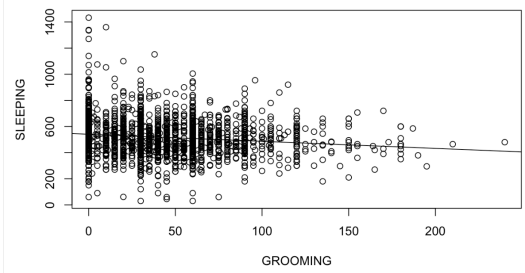
The linear regression was formed with "SLEEPING" being regressed against "GROOMING". The plot was graphed with the abline as shown below. The Pearson's correlation was also produced to be -0.1522111.

```r
plot(SLEEPING~GROOMING, data = sleep)
lm_slr <- lm(sleep$SLEEPING~sleep$GROOMING)
abline(lm_slr)
cor(sleep$SLEEPING, sleep$GROOMING)
```



However, as seen in the graph above, the majority of the data is concentrated to the left of the graph, with a singular outlier on the right. This outlier may affect the overall graph due to its singular but extreme influence, potentially skewing it. Hence, considering the highly dense data points, it was omitted to fit a better-representing model. After excluding the data point, the model's correlation decreased to -0.1393408, but the graph's shape and overall points were better spread out. The following graph was produced after the omission of the singular outlier.

```r
suppressPackageStartupMessages(library(dplyr))
sleep2 <- filter(sleep, sleep$GROOMING < 600)
plot(SLEEPING~GROOMING, sleep2)
lm_slr_2 <- lm(sleep2$SLEEPING~sleep2$GROOMING)
abline(lm_slr_2)
cor(sleep2$SLEEPING, sleep2$GROOMING)
```



Upon creating the linear model, an Analysis of Variance (ANOVA) table was also made, which displays a significant relationship between grooming behaviour and sleep patterns. The large F-value of 49.441 and small p-value of 2.628e-12 indicate that the changes in grooming explain a substantial amount of the changes in sleep as shown below. This indicates that GROOMING is a good predictor in comparison to an intercept-only model. However, the higher residual sum of squares (44881368) means there is a lot of unexplained variability that the model is unable to explain, which is explained by the previously low Pearson correlation value indicating a weak relationship.

```{r}
anova_table <- anova(lm_slr_2)
print(anova_table)
summary(lm_slr_2)
```

```
Analysis of Variance Table

Response: sleep2$SLEEPING
                 Df   Sum Sq Mean Sq F value    Pr(>F)
sleep2$GROOMING   1   888664  888664  49.441 2.628e-12 ***
Residuals      2497 44881368   17974
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Diagnostic plots along with the model's summary were also generated to see if the assumptions are met by the linear model. The summary shown (Appendix G) describes the linear model's fit. The coefficient of GROOMING decreases by -0.54105 for each 1 unit increase in SLEEPING. The p-values generated are both less than 0.05 and are very significant, but the low R-squared value of 0.01942 shows that changes in GROOMING very minimally affect the changes in SLEEPING.

The diagnostic plots (Appendix H) were created to see if the assumptions of normality, independence, constant variance and to see if there are any patterns or leverage points. The residuals vs fitted and scale-location graphs produce a generally straight red line with majority of the points clustered on the right of the graph, and spreads out through the middle with fewer points on the left side. There are a few notable observations on the extreme left that are outliers. There is a pattern of increasing points towards the right and hence the variance is not considered constant. The normal Q-Q plot predominantly has points lying on the line but trail off at the ends, suggesting skewness which was seen in the EDA, but normality is mostly met. The residuals vs leverage plot has points clustered to the left and spread through to the right, the red line is straight and there are a few leverage points but do not greatly alter the graph shape.

3. Multiple Linear Regression (MLR)

3.1 All variables

To begin the multiple linear regression, the whole model was initially selected to see the impact of all variables on sleep. This model would consider all variables affecting sleep.

The summary as seen below was generated.

```
Call:
lm(formula = SLEEPING ~ ., data = sleep)

Residuals:
    Min      1Q  Median      3Q     Max
-525.67  -80.70   -6.16   71.55  829.05

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.968e+03  1.787e+03  -1.101 0.270812
EDUC          -2.693e+00  1.333e+00  -2.020 0.043454 *
AGE           -5.530e-01  1.884e-01  -2.936 0.003359 **
EMPLOYMENT    -8.022e+00  5.202e+00  -1.542 0.123135
SEX           -1.208e+01  5.920e+00  -2.041 0.041403 *
CHILDREN      -1.662e+00  2.835e+00  -0.586 0.557773
EARNINGS      -1.155e-02  5.757e-03  -2.007 0.044892 *
YEAR           1.308e+00  8.901e-01   1.469 0.141855
HOURS_WORKED  -4.954e-01  2.183e-01  -2.269 0.023334 *
GROOMING      -6.334e-01  7.436e-02  -8.518  < 2e-16 ***
HOUSEWORK     -7.300e-02  3.443e-02  -2.120 0.034100 *
FOOD_PREP     -6.965e-02  5.225e-02  -1.333 0.182625
CARETAKING    -1.510e-01  4.854e-02  -3.110 0.001893 **
PLAYING        1.209e-01  8.835e-02   1.369 0.171190
JOB_SEARCHING -1.107e-01  1.117e-01  -0.991 0.321805
SHOPPING       4.375e-02  5.050e-02   0.866 0.386328
EATING        -1.067e-01  4.902e-02  -2.177 0.029609 *
SOCIALISING   -1.393e-01  1.928e-02  -7.222 6.77e-13 ***
TV             1.184e-01  2.275e-02   5.202 2.13e-07 ***
GOLFING       -1.341e-01  1.518e-01  -0.883 0.377149
RUNNING        4.565e-02  4.454e-01   0.102 0.918375
VOLUNTEERING  -1.707e-01  4.598e-02  -3.713 0.000209 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 130.5 on 2478 degrees of freedom
Multiple R-squared:  0.08126, Adjusted R-squared:  0.07347
F-statistic: 10.44 on 21 and 2478 DF,  p-value: < 2.2e-16
```

The summary indicates the model fit is explained by the R-squared value of 0.08126, indicating that 8.126% of the variability in sleeping is explained by the predictors. The lowest p-value variables were GROOMING, SOCIALISING and TV, which show that they had a strong impact on SLEEP. The model can be said to be significant as explained by the large F-value (10.44) and low p-value of (2.2e-16). The large spread of the residuals suggest the model doesn't accurately capture all the influences of lifestyle factors on sleep.

The diagnostic plots were then generated (Appendix I). The residuals vs fitted and scale-location graphs look similar with majority of the points on the right and a generally straight line, this indicates the model only struggled to fit some observations. The normal Q-Q plot lies mostly on the line but trails off extremely at the ends. The residuals vs leverage graph has majority of observations on the left with no observations exceeding the Cook's distance line. Overall, the model has seemingly constant variance as seen by the residuals vs fitted and scale-location graphs. The normality is not severely violated but there is skewness present in the model. The residual vs leverage graph suggests there aren't any leverage points and are within the acceptable range. Despite the diagnostic plots meeting most criteria, the summary indicates the model is weak in its ability to predict SLEEP using all the predictor variables.

3.2 Forward Selection

The forward selection method was used as another method of seeing which variables R would choose to fit the model. This model used 14 variables out of the 21 predictors to predict SLEEP.

The summary of the model was generated as seen below.

```{r}
mlr <- lm(SLEEPING ~ GROOMING + EARNINGS + EDUC + SOCIALISING + TV + HOURS_WORKED +
    VOLUNTEERING + CARETAKING + AGE + EATING + HOUSEWORK + SEX +
    PLAYING + EMPLOYMENT, data = sleep)
summary(mlr)
```

```
Call:
lm(formula = SLEEPING ~ GROOMING + EARNINGS + EDUC + SOCIALISING +
    TV + HOURS_WORKED + VOLUNTEERING + CARETAKING + AGE + EATING +
    HOUSEWORK + SEX + PLAYING + EMPLOYMENT, data = sleep)

Residuals:
    Min      1Q  Median      3Q     Max
-515.77  -80.98   -6.96   72.61  828.62

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  652.516243  13.799385  47.286  < 2e-16 ***
GROOMING      -0.630883   0.074091  -8.515  < 2e-16 ***
EARNINGS      -0.011087   0.005739  -1.932 0.053495 .
EDUC          -2.640461   1.324209  -1.994 0.046262 *
SOCIALISING   -0.137119   0.019151  -7.160 1.06e-12 ***
TV             0.117886   0.022728   5.187 2.31e-07 ***
HOURS_WORKED  -0.486508   0.216662  -2.245 0.024825 *
VOLUNTEERING  -0.171015   0.045849  -3.730 0.000196 ***
CARETAKING    -0.171355   0.045863  -3.736 0.000191 ***
AGE           -0.539190   0.173493  -3.108 0.001906 **
EATING        -0.106168   0.048939  -2.169 0.030146 *
HOUSEWORK     -0.076887   0.034266  -2.244 0.024931 *
SEX          -11.969308   5.766366  -2.076 0.038023 *
PLAYING        0.143159   0.087462   1.637 0.101797
EMPLOYMENT    -7.770404   5.172034  -1.502 0.133124
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 130.5 on 2485 degrees of freedom
Multiple R-squared:  0.07876,   Adjusted R-squared:  0.07357
F-statistic: 15.17 on 14 and 2485 DF,  p-value: < 2.2e-16
```

The summary indicates the model fit is explained by the R-squared value of 0.07876, indicating that 7.876% of the variability in sleeping is explained by the predictors. The lowest p-value variables were again GROOMING, SOCIALISING and TV, which show a strong impact on SLEEP. The model can be said to be significant as explained by the large F-value (15.17) and low p-value of (2.2e-16). The large spread of the residuals suggest the model doesn't accurately capture all the influences of lifestyle factors on sleep but the spread is slightly less compared to the all variable model.

The diagnostic plots for the forward model (Appendix J), has a resemblance to the plots of the entire model. In the residuals vs fitted graph, the red line followed a similar generally straight line with only a singular outlier compared to the two found in the entire model. The normal Q-Q plot seemed to be exactly similar to the entire model and showed majority of the points on the line, with extreme deviations at the ends. The scale-location graph was shown slightly better on the forward model due to the absence of a influential outlier, yet the red line was seen to curve near the cluster of points. However, the residuals vs leverage graph looked worse on the forward model with the line not as straight as the one in the all-variable model, but had significantly less outliers. The model seems to be similar to the all-variable model but has less outliers on the diagnostic plots, despite this, the low R-squared value indicates a weaker graph compared to all variables model.

3.3 Backwards Selection

The backwards selection method is another procedure used to see which variables are considered impactful on SLEEPING. This method chose 15 out of 21 predictor variables with the majority being the same as the forward selection but also chose YEAR as a variable affecting SLEEPING.

The summary is shown below.

```r
rmlr <- lm(SLEEPING ~ EDUC + AGE + EMPLOYMENT + SEX + EARNINGS + YEAR +
    HOURS_WORKED + GROOMING + HOUSEWORK + CARETAKING + PLAYING +
    EATING + SOCIALISING + TV + VOLUNTEERING, data = sleep)
summary(rmlr)
```

```
Call:
lm(formula = SLEEPING ~ EDUC + AGE + EMPLOYMENT + SEX + EARNINGS +
    YEAR + HOURS_WORKED + GROOMING + HOUSEWORK + CARETAKING +
    PLAYING + EATING + SOCIALISING + TV + VOLUNTEERING, data = sleep)

Residuals:
    Min      1Q  Median      3Q     Max
-521.26  -80.61   -6.23   72.46  833.50

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.864e+03  1.783e+03  -1.045 0.295911
EDUC         -2.699e+00  1.325e+00  -2.038 0.041666 *
AGE          -5.434e-01  1.735e-01  -3.132 0.001755 **
EMPLOYMENT   -7.795e+00  5.171e+00  -1.508 0.131805
SEX          -1.168e+01  5.769e+00  -2.025 0.042966 *
EARNINGS     -1.147e-02  5.745e-03  -1.997 0.045953 *
YEAR          1.254e+00  8.882e-01   1.411 0.158234
HOURS_WORKED -4.713e-01  2.169e-01  -2.173 0.029856 *
GROOMING     -6.269e-01  7.413e-02  -8.457  < 2e-16 ***
HOUSEWORK    -7.584e-02  3.427e-02  -2.213 0.026969 *
CARETAKING   -1.678e-01  4.593e-02  -3.653 0.000265 ***
PLAYING       1.330e-01  8.774e-02   1.515 0.129824
EATING       -1.048e-01  4.894e-02  -2.141 0.032346 *
SOCIALISING  -1.363e-01  1.916e-02  -7.113 1.47e-12 ***
TV            1.172e-01  2.273e-02   5.157 2.70e-07 ***
VOLUNTEERING -1.700e-01  4.584e-02  -3.708 0.000213 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 130.5 on 2484 degrees of freedom
Multiple R-squared:  0.0795,    Adjusted R-squared:  0.07394
F-statistic:  14.3 on 15 and 2484 DF,  p-value: < 2.2e-16
```

The summary indicates the model fit is explained by the R-squared value of 0.0795, indicating that 7.95% of the variability in sleeping is explained by the predictors. The lowest p-value variables were again GROOMING, SOCIALISING and TV, which show a strong impact on SLEEP. The model can be said to be significant as explained by the large F-value (14.3) and low p-value of (2.2e-16). The large spread of the residuals suggest the model doesn't accurately capture all the influences of lifestyle factors on sleep. The spread is the greatest in comparison to the all variable model and forward selection.

The diagnostic plots (Appendix K) are almost closely similar to the forward model. This is due to having selected the same variables but with the addition of YEAR. The residuals vs fitted graph follows the same structure and the red line is mostly straight indicating the assumption of independence is met. The normal Q-Q plot trails off at the ends, similar to both forward and all-variable models and can be said that the assumption of normality is mostly met. The scale-location and residuals vs leverage closely resemble the forward model. This means that the spread of residuals is approximately even and that no highly influential leverage points exist that cross the Cook's distance line. Overall, the graph performed slightly better than the forward model due to the higher r-squared value and the extra predictor variable.

4. Discussion

4.1 Conclusion

Upon considering the performance of the models and their ability to explain the changes in sleep, the lmall or all-variable model performed the best in comparison to the simple linear regression, backwards and forward selection models. The all-variable model despite showing the best R-squared value and a multitude of significant variables, had difficulty in explaining the changes in sleep based on the predictors. This could imply that even though 21 lifestyle predictor variables were used, there wasn't a robust correlation or linear relationship between a single or several variables, indicating that sleep is a complex variable to explain. The all-variable model also had the lower RSS value when compared with other models using the ANOVA table as seen in Appendix L, M and N. However, it can be concluded, that GROOMING, SOCIALISING and TV were significant variables when attempting to predict sleep patterns. These appeared in all models with extreme test values and low p-values signifying a statistically prominent effect on the response variable. Overall, the all-variable model was able to capture the strongest relative relationship with sleep using all 21 variables.

4 .2 Limitations

There are many limitations that may exist when trying to fit single and multiple linear regression using the sleep dataset. They include:

- Existence of multicollinearity - the relationship within the predictor variables become difficult to isolate and with multiple linear regression, there are several predictors that could potentially be interrelated with each other
- Assumption violations - the diagnostic plots show signs of potential violations even though there may not be an explicitly large fluctuation in the graphs
- Weak model performance - all models were unable to find a strong relationship between the predictors and SLEEPING which means a lot of the variability in sleep is unexplained
- External factors - the dataset doesn't account for lifestyle changes or health patterns and other factors that could affect sleep
- Variable selection - the strongest correlation between sleep is the variable GROOMING which was selected through the best subset selection method, this method undermines the possibility of overfitting the data which may result in the relationship being purely coincidental

Due to the complexity of the relationships between the 21 variables and SLEEPING, the simple linear and multiple linear regression models were unable to capture a significant correlation between variables, however, the model with the best fit was the all-variable model.

5. References

Blackboard. (2023). *Lectures 7-11 - Regression and Nonparametric Inference* Curtin.edu.au. https://lms.curtin.edu.au/webapps/blackboard/content/listContent.jsp?course_id=_127657_1&content_id=_11538516_1&mode=reset

Correlation and causation. (2023). Australian Bureau of Statistics; Australian Bureau of Statistics. https://www.abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/correlation-and-causation#:~:text=A%20correlation%20between%20variables%2C%20however,relationship%20between%20the%20two%20events.

Grolemund, G. (2013). *24 Model building | R for Data Science*. Had.co.nz. https://r4ds.had.co.nz/model-building.html

Grolemund, G. (2023a). *3 Data visualisation | R for Data Science*. Had.co.nz. https://r4ds.had.co.nz/data-visualisation.html

Grolemund, G. (2023b). *7 Exploratory Data Analysis | R for Data Science*. Had.co.nz. https://r4ds.had.co.nz/exploratory-data-analysis.html

in. (2016, October 5). *How do I determine the best predictor in a linear regression model?* Cross Validated. https://stats.stackexchange.com/questions/238501/how-do-i-determine-the-best-predictor-in-a-linear-regression-model#:~:text=Generally%20variable%20with%20highest%20correlation,change%20in%20R%2Dsquared%20value.

Multiple Linear Regression in R - Articles - STHDA. (2018, March 10). Sthda.com. https://www.sthda.com/english/articles/40-regression-analysis/168-multiple-linear-regression-in-r/

Normal QQ plot and general QQ plot—ArcMap | Documentation. (2021). Arcgis.com. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/normal-qq-plot-and-general-qq-plot.htm#:~:text=QQ%20plot%20example-,Examining%20data%20distributions%20using%20QQ%20plots,deviate%20from%20the%20reference%20line.

Residual Sum of Squares (RSS): What It Is, How to Calculate It. (2023). Investopedia. https://www.investopedia.com/terms/r/residual-sum-of-squares.asp#:~:text=The%20residual%20sum%20of%20squares%20(RSS)%20is%20a%20statistical%20technique,the%20residuals%2C%20or%20error%20term.

Stepwise Regression Essentials in R - Articles - STHDA. (2018, March 11). Sthda.com. https://www.sthda.com/english/articles/37-model-selection-essentials-in-r/154-stepwise-regression-essentials-in-r/

Understanding Diagnostic Plots for Linear Regression Analysis | UVA Library. (2023). Virginia.edu. https://library.virginia.edu/data/articles/diagnostic-plots

6. Appendices

Appendix A

```
```{r}
head(sleep)
```
```

Description: df [6 × 22]

	EDUC <int>	AGE <int>	EMPLOYMENT <int>	SEX <int>	CHILDREN <int>	EARNINGS <int>	YEAR <int>	HOURS_WORKED <int>	SLEEPING <int>
22445	4	59	0	1	0	0	2004	0	585
43594	9	31	1	0	1	0	2005	0	555
74835	4	49	2	0	0	229	2008	20	595
118646	5	41	2	1	1	615	2012	40	180
69054	1	55	2	1	0	2404	2008	55	518
42289	7	38	2	1	0	1085	2005	45	500

6 rows | 1–10 of 22 columns

Appendix B

```
```{r}
summary(sleep)
```
```

```
      EDUC            AGE          EMPLOYMENT         SEX           CHILDREN         EARNINGS           YEAR     
 Min.   : 0.000   Min.   :15.00   Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   :   0.0   Min.   :2003  
 1st Qu.: 4.000   1st Qu.:33.00   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:   0.0   1st Qu.:2004  
 Median : 5.000   Median :44.00   Median :2.000   Median :0.000   Median :0.0000   Median : 237.0   Median :2007  
 Mean   : 5.262   Mean   :46.14   Mean   :1.332   Mean   :0.448   Mean   :0.8704   Mean   : 469.3   Mean   :2007  
 3rd Qu.: 7.000   3rd Qu.:59.00   3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:2.0000   3rd Qu.: 769.0   3rd Qu.:2010  
 Max.   :10.000   Max.   :85.00   Max.   :2.000   Max.   :1.000   Max.   :9.0000   Max.   :2885.0   Max.   :2012  
  HOURS_WORKED       SLEEPING        GROOMING        HOUSEWORK        FOOD_PREP       CARETAKING        PLAYING      
 Min.   :  0.00   Min.   :  30.0   Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   :  0.000  
 1st Qu.:  0.00   1st Qu.: 435.0   1st Qu.: 15.00   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.000  
 Median : 32.00   Median : 510.0   Median : 32.00   Median :  0.00   Median : 10.00   Median :  0.00   Median :  0.000  
 Mean   : 24.67   Mean   : 518.6   Mean   : 40.72   Mean   : 40.69   Mean   : 34.74   Mean   : 32.34   Mean   :  9.144  
 3rd Qu.: 40.00   3rd Qu.: 597.0   3rd Qu.: 60.00   3rd Qu.: 55.00   3rd Qu.: 50.00   3rd Qu.: 15.00   3rd Qu.:  0.000  
 Max.   :100.00   Max.   :1433.0   Max.   :660.00   Max.   :720.00   Max.   :697.00   Max.   :870.00   Max.   :490.000  
 JOB_SEARCHING       SHOPPING         EATING        SOCIALISING          TV            GOLFING          RUNNING      
 Min.   :  0.000   Min.   :  0.00   Min.   :  0.00   Min.   :   0.0   Min.   :   0.0   Min.   :  0.000   Min.   :  0.000  
 1st Qu.:  0.000   1st Qu.:  0.00   1st Qu.: 30.00   1st Qu.: 120.0   1st Qu.:  30.0   1st Qu.:  0.000   "print"          
 Median :  0.000   Median :  0.00   Median : 60.00   Median : 245.0   Median : 120.0   Median :  0.000   Median :  0.000  
 Mean   :  2.275   Mean   : 26.89   Mean   : 69.17   Mean   : 282.2   Mean   : 159.7   Mean   :  1.102   Mean   :  0.506  
 3rd Qu.:  0.000   3rd Qu.: 30.00   3rd Qu.: 90.00   3rd Qu.: 400.0   3rd Qu.: 230.0   3rd Qu.:  0.000   3rd Qu.:  0.000  
 Max.   :420.000   Max.   :420.00   Max.   :595.00   Max.   :1200.0   Max.   :1010.0   Max.   :385.000   Max.   :150.000  
  VOLUNTEERING   
 Min.   :   0.00  
 1st Qu.:   0.00  
 Median :   0.00  
 Mean   :  11.46  
 3rd Qu.:   0.00  
 Max.   :1100.00  
```

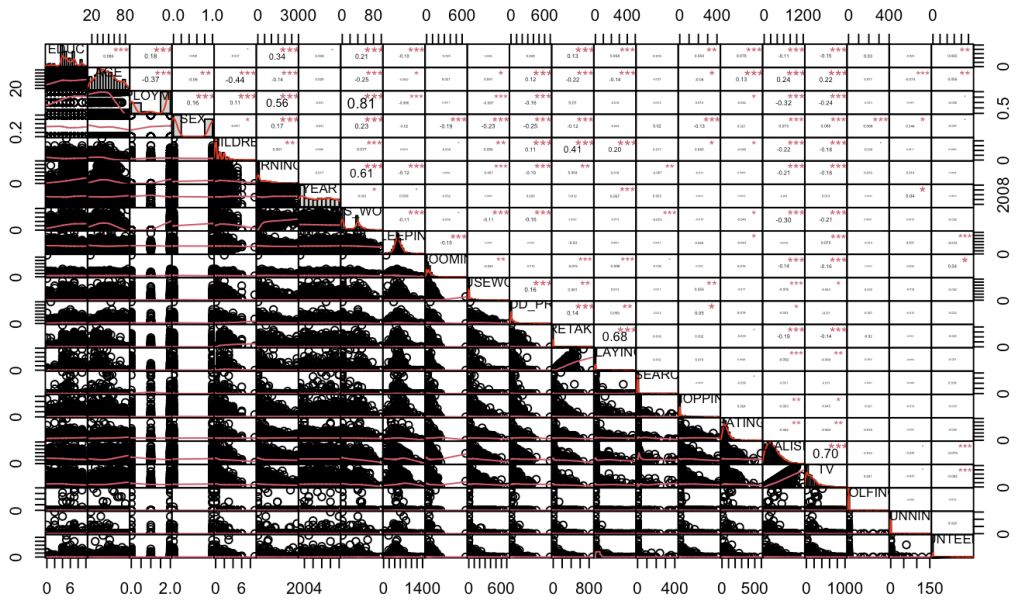
Appendix C

```{r}
str(sleep)
```

```
'data.frame':   2500 obs. of  22 variables:
 $ EDUC          : int  4 9 4 5 1 7 5 6 3 1 ...
 $ AGE           : int  59 31 49 41 55 38 20 38 39 16 ...
 $ EMPLOYMENT    : int  0 1 2 2 2 2 2 2 2 ...
 $ SEX           : int  1 0 0 1 1 1 1 0 0 1 ...
 $ CHILDREN      : int  0 1 0 1 0 0 2 2 0 1 ...
 $ EARNINGS      : int  0 0 229 615 2404 1085 245 924 400 84 ...
 $ YEAR          : int  2004 2005 2008 2012 2008 2005 2008 2012 2006 2007 ...
 $ HOURS_WORKED  : int  0 0 20 40 55 45 0 40 40 14 ...
 $ SLEEPING      : int  585 555 595 180 518 500 420 618 740 529 ...
 $ GROOMING      : int  0 30 90 90 30 30 50 85 40 45 ...
 $ HOUSEWORK     : int  0 40 10 0 60 0 0 15 30 5 ...
 $ FOOD_PREP     : int  0 35 45 0 0 0 25 40 30 5 ...
 $ CARETAKING    : int  0 300 0 55 0 0 0 0 0 0 ...
 $ PLAYING       : int  0 95 0 0 0 0 0 0 0 0 ...
 $ JOB_SEARCHING: int  0 0 0 0 0 0 0 0 0 0 ...
 $ SHOPPING      : int  0 0 60 0 35 0 0 70 0 10 ...
 $ EATING        : int  165 100 20 85 20 80 60 50 30 40 ...
 $ SOCIALISING   : int  618 90 370 435 125 145 0 313 45 168 ...
 $ TV            : int  250 0 180 0 125 145 0 120 0 120 ...
 $ GOLFING       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ RUNNING       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ VOLUNTEERING : int  0 100 0 0 0 0 0 0 0 0 ...
```

Appendix D

```{r}
suppressWarnings({
library(PerformanceAnalytics)
chart.Correlation(sleep, method = "pearson")
})
```



Appendix E

```{r}
cor(sleep, method = "pearson")
```

	EDUC	AGE	EMPLOYMENT	SEX	CHILDREN	EARNINGS	YEAR	HOURS_WORKED
EDUC	1.000000000	0.09503337	0.180885518	-0.02923703	-0.036680876	0.342565983	0.037923779	0.208154381
AGE	0.095033367	1.00000000	-0.373136512	-0.05956302	-0.440874810	-0.136941512	0.029190030	-0.253501835
EMPLOYMENT	0.180885518	-0.37313651	1.000000000	0.16009559	0.108504586	0.560321847	-0.030773743	0.808623863
SEX	-0.029237029	-0.05956302	0.160095595	1.00000000	-0.051452280	0.174478701	-0.030803276	0.230518711
CHILDREN	-0.036680876	-0.44087481	0.108504586	-0.05145228	1.000000000	0.061150074	0.008583451	0.076565633
EARNINGS	0.342565983	-0.13694151	0.560321847	0.17447870	0.061150074	1.000000000	0.017261553	0.614634704
YEAR	0.037923779	0.02919003	-0.030773743	-0.03080328	0.008583451	0.017261553	1.000000000	-0.042374296
HOURS_WORKED	0.208154381	-0.25350184	0.808623863	0.23051871	0.076565633	0.614634704	-0.042374296	1.000000000
SLEEPING	-0.103503447	-0.04199892	-0.095935252	-0.02044201	0.001640198	-0.123653359	0.039106440	-0.108648695
GROOMING	-0.025482902	0.03132163	0.001123322	-0.19131259	-0.037558051	-0.006355326	-0.031576657	-0.034778365
HOUSEWORK	0.008577703	0.04092799	-0.096774442	-0.23032920	0.055513904	-0.067231878	-0.005505898	-0.107520952
FOOD_PREP	0.029195932	0.11777162	-0.157525315	-0.24976490	0.114473631	-0.102796310	0.026390546	-0.158494049
CARETAKING	0.131603962	-0.22475812	0.010498890	-0.11839557	0.405076540	0.053634007	0.011829066	0.002733954
PLAYING	0.068218066	-0.14115060	-0.018336451	-0.02383117	0.201246193	0.018364558	0.067094350	-0.012654564
JOB_SEARCHING	0.018866944	-0.02079469	-0.015079043	0.02018863	0.017434185	-0.056689473	0.032591090	-0.073344647
SHOPPING	0.059362685	-0.04009661	0.014247518	-0.12507715	0.044632964	-0.010642686	-0.001239866	-0.015145667
EATING	0.078456713	0.12742264	-0.041980210	0.02081750	-0.048221092	0.001942604	-0.007631935	-0.045294376
SOCIALISING	-0.114878096	0.24191116	-0.318296916	0.07329848	-0.223977939	-0.209997351	-0.001014688	-0.298391953
TV	-0.148627195	0.22052783	-0.237088528	0.08540632	-0.177829774	-0.175169264	0.013217827	-0.206213501
GOLFING	0.029526401	0.01722212	-0.012877306	0.06798544	-0.037665945	0.018957793	-0.003266649	0.003068785
RUNNING	-0.022649123	-0.07441480	-0.006670806	0.04399531	0.001141422	0.019189299	0.040105954	0.001819230
VOLUNTEERING	0.064910459	0.05634709	-0.037650635	-0.03732520	0.006253553	-0.003374573	-0.007852049	-0.031689594

	SLEEPING	GROOMING	HOUSEWORK	FOOD_PREP	CARETAKING	PLAYING	JOB_SEARCHING	SHOPPING
EDUC	-0.103503447	-0.0254829019	0.008577703	0.029195932	0.131603962	0.068218066	0.018866944	0.059362685
AGE	-0.041998924	0.0313216331	0.040927995	0.117771619	-0.224758120	-0.141150602	-0.020794692	-0.040096614
EMPLOYMENT	-0.095935252	0.0011233215	-0.096774442	-0.157525315	0.010498890	-0.018336451	-0.015079043	0.014247518
SEX	-0.020442008	-0.1913125859	-0.230329202	-0.249764904	-0.118395573	-0.023831169	0.020188632	-0.125077154
CHILDREN	0.001640198	-0.0375580513	0.055513904	0.114473631	0.405076540	0.201246193	0.017434185	0.044632964
EARNINGS	-0.123653359	-0.0063553256	-0.067231878	-0.102796310	0.053634007	0.018364558	-0.056689473	-0.010642686
YEAR	0.039106440	-0.0315766574	-0.005505898	0.026390546	0.011829066	0.067094350	0.032591090	-0.001239866
HOURS_WORKED	-0.108648695	-0.0347783650	-0.107520952	-0.158494049	0.002733954	-0.012654564	-0.073344647	-0.015145667
SLEEPING	1.000000000	-0.1522110953	-0.004666137	-0.005263403	-0.030054135	0.008080965	0.001175768	0.025217397
GROOMING	-0.152211095	1.0000000000	-0.054688717	-0.016365354	-0.075516352	-0.087567516	-0.035897684	-0.005208028
HOUSEWORK	-0.004666137	-0.0546887166	1.000000000	0.161432160	0.060905038	0.012809091	-0.010857700	0.056483686
FOOD_PREP	-0.005263403	-0.0163653536	0.161432160	1.000000000	0.142135234	0.058628363	-0.011993909	0.049675645
CARETAKING	-0.030054135	-0.0755163520	0.060905038	0.142135234	1.000000000	0.675476776	0.018183770	0.031677143
PLAYING	0.008080965	-0.0875675158	0.012809091	0.058628363	0.675476776	1.000000000	0.011838762	0.019028929
JOB_SEARCHING	0.001175768	-0.0358976844	-0.010857700	-0.011993909	0.018183770	0.011838762	1.000000000	-0.007288088
SHOPPING	0.025217397	-0.0052080276	0.056483686	0.049675645	0.031677143	0.019028929	-0.007288088	1.000000000
EATING	-0.042827963	0.0154260879	-0.016559202	0.018546246	-0.033171824	0.002763771	-0.035969465	0.024181747
SOCIALISING	-0.016024881	-0.1422592331	-0.078339585	-0.042818155	-0.189715880	-0.091679940	-0.030541642	-0.062548997
TV	0.074649678	-0.1561714522	-0.042692396	-0.010061788	-0.141514067	-0.063535290	-0.013438218	-0.045334368
GOLFING	-0.018826030	-0.0004869388	-0.032625688	-0.027228982	-0.019556373	-0.014315513	-0.006132246	-0.021463756
RUNNING	0.007009270	0.0047701395	-0.017569198	-0.014841860	-0.002067097	-0.005263322	-0.008233395	-0.012668568
VOLUNTEERING	-0.071516631	0.0403316162	-0.031790872	-0.023474476	-0.024673994	-0.020668899	0.018374757	-0.018029349

	EATING	SOCIALISING	TV	GOLFING	RUNNING	VOLUNTEERING
EDUC	0.078456713	-0.114878096	-0.14862720	0.0295264008	-0.022649123	0.064910459
AGE	0.127422642	0.241911160	0.22052783	0.0172221191	-0.074414800	0.056347087
EMPLOYMENT	-0.041980210	-0.318296916	-0.23708853	-0.0128773063	-0.006670806	-0.037650635
SEX	0.020817497	0.073298482	0.08540632	0.0679854386	0.043995307	-0.037325200
CHILDREN	-0.048221092	-0.223977939	-0.17782977	-0.0376659450	0.001141422	0.006253553
EARNINGS	0.001942604	-0.209997351	-0.17516926	0.0189577933	0.019189299	-0.003374573
YEAR	-0.007631935	-0.001014688	0.01321783	-0.0032666487	0.040105954	-0.007852049
HOURS_WORKED	-0.045294376	-0.298391953	-0.20621350	0.0030687846	0.001819230	-0.031689594
SLEEPING	-0.042827963	-0.016024881	0.07464968	-0.0188260302	0.007009270	-0.071516631
GROOMING	0.015426088	-0.142259233	-0.15617145	-0.0004869388	0.004770140	0.040331616
HOUSEWORK	-0.016559202	-0.078339585	-0.04269240	-0.0326256878	-0.017569198	-0.031790872
FOOD_PREP	0.018546246	-0.042818155	-0.01006179	-0.0272289815	-0.014841860	-0.023474476
CARETAKING	-0.033171824	-0.189715880	-0.14151407	-0.0195563733	-0.002067097	-0.024673994
PLAYING	0.002763771	-0.091679940	-0.06353529	-0.0143155128	-0.005263322	-0.020668899
JOB_SEARCHING	-0.035969465	-0.030541642	-0.01343822	-0.0061322459	-0.008233395	0.018374757
SHOPPING	0.024181747	-0.062548997	-0.04533437	-0.0214637558	-0.012668568	-0.018029349
EATING	1.000000000	-0.062419145	-0.06426299	0.0194421442	-0.004354778	-0.028616880
SOCIALISING	-0.062419145	1.000000000	0.69663764	0.0118462060	-0.034259367	-0.078585248
TV	-0.064262992	0.696637636	1.00000000	0.0212566710	-0.036566691	-0.081971002
GOLFING	0.019442144	0.011846206	0.02125667	1.0000000000	-0.005458757	-0.012665953
RUNNING	-0.004354778	-0.034259367	-0.03656669	-0.0054587571	1.000000000	0.028726174
VOLUNTEERING	-0.028616880	-0.078585248	-0.08197100	-0.0126659532	0.028726174	1.000000000

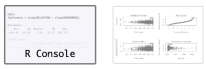
Appendix F

```{r}
suppressPackageStartupMessages(library(caret))
suppressPackageStartupMessages(library(leaps))
best_subsets <- regsubsets(SLEEPING~., data = sleep, nvmax = 22, nbest = 1)
summary(best_subsets)
```

```
Subset selection object
Call: regsubsets.formula(SLEEPING ~ ., data = sleep, nvmax = 22, nbest = 1)
21 Variables  (and intercept)
             Forced in Forced out
EDUC             FALSE      FALSE
AGE              FALSE      FALSE
EMPLOYMENT       FALSE      FALSE
SEX              FALSE      FALSE
CHILDREN         FALSE      FALSE
EARNINGS         FALSE      FALSE
YEAR             FALSE      FALSE
HOURS_WORKED     FALSE      FALSE
GROOMING         FALSE      FALSE
HOUSEWORK        FALSE      FALSE
FOOD_PREP        FALSE      FALSE
CARETAKING       FALSE      FALSE
PLAYING          FALSE      FALSE
JOB_SEARCHING    FALSE      FALSE
SHOPPING         FALSE      FALSE
EATING           FALSE      FALSE
SOCIALISING      FALSE      FALSE
TV               FALSE      FALSE
GOLFING          FALSE      FALSE
RUNNING          FALSE      FALSE
VOLUNTEERING     FALSE      FALSE
1 subsets of each size up to 21
Selection Algorithm: exhaustive
          EDUC AGE EMPLOYMENT SEX CHILDREN EARNINGS YEAR HOURS_WORKED GROOMING HOUSEWORK FOOD_PREP CARETAKING PLAYING
1  ( 1 )  " "  " " " "        " " " "      " "      " "  " "          "*"      " "       " "       " "       " "
2  ( 1 )  " "  " " " "        " " " "      "*"      " "  " "          "*"      " "       " "       " "       " "
3  ( 1 )  "*"  " " " "        " " " "      " "      " "  "*"          "*"      " "       " "       " "       " "
4  ( 1 )  " "  " " " "        " " " "      " "      " "  "*"          "*"      " "       " "       " "       " "
5  ( 1 )  "*"  " " " "        " " " "      " "      " "  "*"          "*"      " "       " "       " "       " "
6  ( 1 )  "*"  " " " "        " " " "      " "      " "  "*"          "*"      " "       " "       " "       " "
7  ( 1 )  " "  "*" " "        " " " "      " "      " "  "*"          "*"      " "       " "       "*"       " "
8  ( 1 )  " "  "*" " "        " " " "      "*"      " "  "*"          "*"      " "       " "       "*"       " "
9  ( 1 )  " "  "*" " "        " " " "      "*"      " "  "*"          "*"      " "       " "       "*"       " "
10 ( 1 )  "*"  "*" " "        " " " "      "*"      " "  "*"          "*"      " "       " "       "*"       " "
11 ( 1 )  "*"  "*" " "        " " " "      "*"      " "  "*"          "*"      "*"       " "       "*"       " "
12 ( 1 )  "*"  "*" " "        "*" " "      "*"      " "  "*"          "*"      "*"       " "       "*"       " "
13 ( 1 )  "*"  "*" " "        "*" " "      "*"      " "  "*"          "*"      "*"       " "       "*"       "*"
14 ( 1 )  "*"  "*" "*"        "*" " "      "*"      " "  "*"          "*"      "*"       " "       "*"       "*"
15 ( 1 )  "*"  "*" "*"        "*" " "      "*"      "*"  "*"          "*"      "*"       " "       "*"       "*"
16 ( 1 )  "*"  "*" "*"        "*" " "      "*"      "*"  "*"          "*"      "*"       "*"       "*"       "*"
17 ( 1 )  "*"  "*" "*"        "*" " "      "*"      "*"  "*"          "*"      "*"       "*"       "*"       "*"
18 ( 1 )  "*"  "*" "*"        "*" " "      "*"      "*"  "*"          "*"      "*"       "*"       "*"       "*"
19 ( 1 )  "*"  "*" "*"        "*" " "      "*"      "*"  "*"          "*"      "*"       "*"       "*"       "*"
20 ( 1 )  "*"  "*" "*"        "*" "*"      "*"      "*"  "*"          "*"      "*"       "*"       "*"       "*"
21 ( 1 )  "*"  "*" "*"        "*" "*"      "*"      "*"  "*"          "*"      "*"       "*"       "*"       "*"
          JOB_SEARCHING SHOPPING EATING SOCIALISING TV  GOLFING RUNNING VOLUNTEERING
1  ( 1 )  " "           " "      " "    " "         " " " "     " "     " "
2  ( 1 )  " "           " "      " "    " "         " " "*"     " "     " "
3  ( 1 )  " "           " "      " "    " "         " " " "     " "     " "
4  ( 1 )  " "           " "      " "    "*"         "*" " "     " "     " "
5  ( 1 )  " "           " "      " "    "*"         "*" " "     " "     " "
6  ( 1 )  " "           " "      " "    "*"         "*" " "     " "     "*"
7  ( 1 )  " "           " "      " "    "*"         "*" " "     " "     "*"
8  ( 1 )  " "           " "      " "    "*"         "*" " "     " "     "*"
9  ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
10 ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
11 ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
12 ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
13 ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
14 ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
15 ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
16 ( 1 )  " "           " "      "*"    "*"         "*" " "     " "     "*"
17 ( 1 )  "*"           " "      "*"    "*"         "*" " "     " "     "*"
18 ( 1 )  "*"           " "      "*"    "*"         "*" "*"     " "     "*"
19 ( 1 )  "*"           "*"      "*"    "*"         "*" "*"     " "     "*"
20 ( 1 )  "*"           "*"      "*"    "*"         "*" "*"     " "     "*"
21 ( 1 )  "*"           "*"      "*"    "*"         "*" "*"     "*"     "*"
```

Appendix G

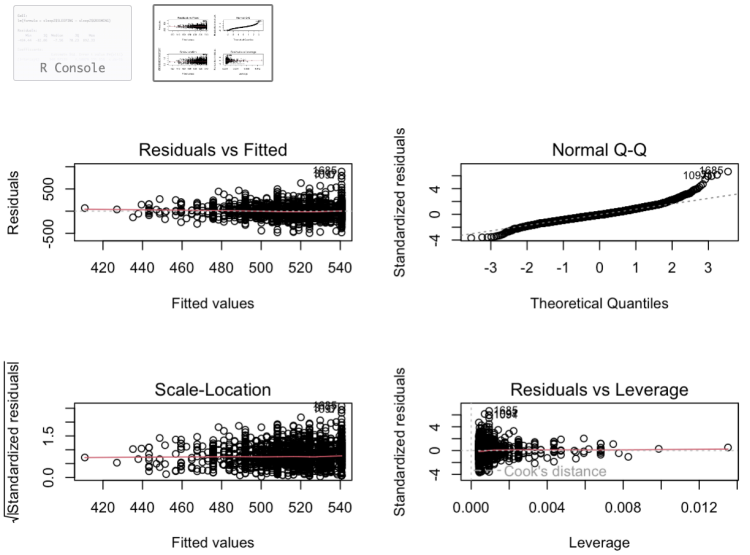
```{r}
summary(lm_slr_2)
par(mfrow = c(2, 2))
plot(lm_slr_2, which = 1)
plot(lm_slr_2, which = 2)
plot(lm_slr_2, which = 3)
plot(lm_slr_2, which = 5)
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))
```




```
Call:
lm(formula = sleep2$SLEEPING ~ sleep2$GROOMING)

Residuals:
    Min      1Q  Median      3Q     Max
-494.44  -82.06   -7.56   70.23  892.33

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     540.67130    4.10976 131.558  < 2e-16 ***
sleep2$GROOMING  -0.54105    0.07695  -7.031 2.63e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 134.1 on 2497 degrees of freedom
Multiple R-squared:  0.01942,   Adjusted R-squared:  0.01902
F-statistic: 49.44 on 1 and 2497 DF,  p-value: 2.628e-12
```

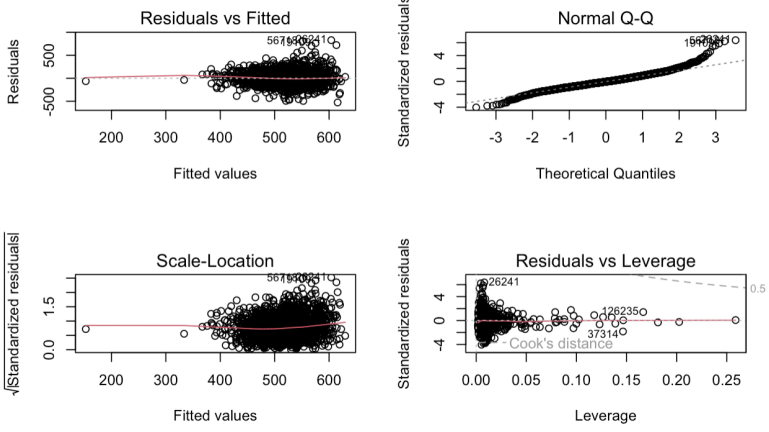
Appendix H

```r
summary(lm_slr_2)
par(mfrow = c(2, 2))
plot(lm_slr_2, which = 1)
plot(lm_slr_2, which = 2)
plot(lm_slr_2, which = 3)
plot(lm_slr_2, which = 5)
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))
```



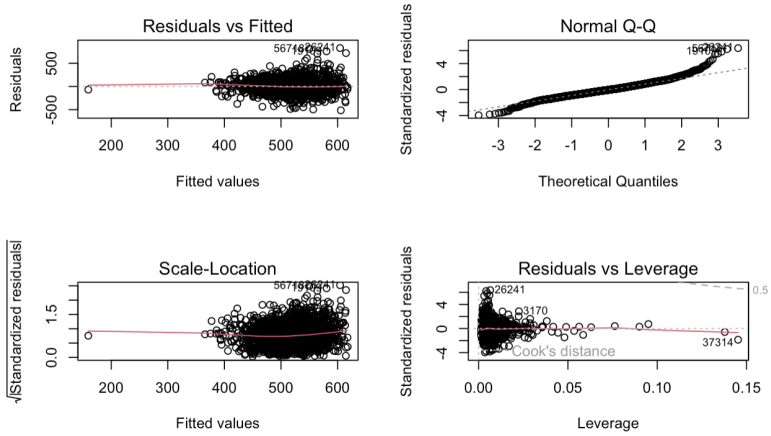
Appendix I

```r
par(mfrow = c(2, 2))
plot(lmall, which = 1)
plot(lmall, which = 2)
plot(lmall, which = 3)
plot(lmall, which = 5)
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))
```



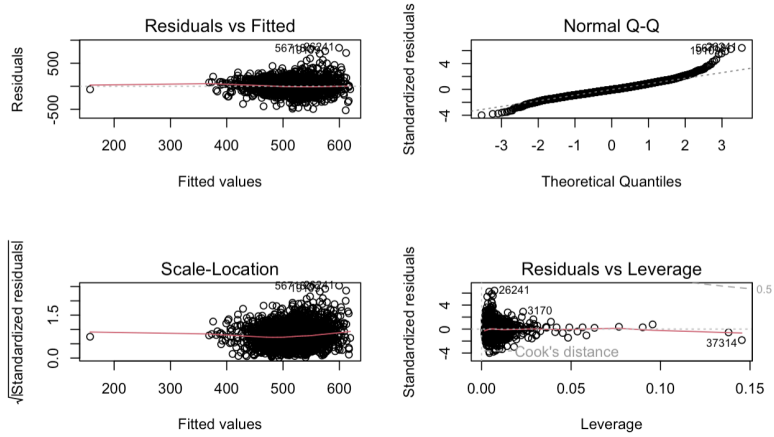
Appendix J

```r
par(mfrow = c(2, 2))
plot(mlr, which = 1)
plot(mlr, which = 2)
plot(mlr, which = 3)
plot(mlr, which = 5)
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))
```



Appendix K

```r
par(mfrow = c(2, 2))
plot(rmlr, which = 1)
plot(rmlr, which = 2)
plot(rmlr, which = 3)
plot(rmlr, which = 5)
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))
```



Appendix L

```
```{r}
anova(lmall, mlr)
```
```

```
Analysis of Variance Table

Model 1: SLEEPING ~ EDUC + AGE + EMPLOYMENT + SEX + CHILDREN + EARNINGS +
    YEAR + HOURS_WORKED + GROOMING + HOUSEWORK + FOOD_PREP +
    CARETAKING + PLAYING + JOB_SEARCHING + SHOPPING + EATING +
    SOCIALISING + TV + GOLFING + RUNNING + VOLUNTEERING
Model 2: SLEEPING ~ GROOMING + EARNINGS + EDUC + SOCIALISING + TV + HOURS_WORKED +
    VOLUNTEERING + CARETAKING + AGE + EATING + HOUSEWORK + SEX +
    PLAYING + EMPLOYMENT
  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1   2478 42219819
2   2485 42334545 -7   -114726 0.9619 0.4574
```

Appendix M

```
```{r}
anova(lmall, mlr_r)
```
```

```
Analysis of Variance Table

Model 1: SLEEPING ~ EDUC + AGE + EMPLOYMENT + SEX + CHILDREN + EARNINGS +
    YEAR + HOURS_WORKED + GROOMING + HOUSEWORK + FOOD_PREP +
    CARETAKING + PLAYING + JOB_SEARCHING + SHOPPING + EATING +
    SOCIALISING + TV + GOLFING + RUNNING + VOLUNTEERING
Model 2: SLEEPING ~ EDUC + AGE + EMPLOYMENT + SEX + EARNINGS + YEAR +
    HOURS_WORKED + GROOMING + HOUSEWORK + CARETAKING + PLAYING +
    EATING + SOCIALISING + TV + VOLUNTEERING
  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1   2478 42219819
2   2484 42300619 -6    -80800 0.7904 0.5773
```

Appendix N

```
```{r}
anova(lmall, lm_slr_1)
```
```

```
Analysis of Variance Table

Model 1: SLEEPING ~ EDUC + AGE + EMPLOYMENT + SEX + CHILDREN + EARNINGS +
    YEAR + HOURS_WORKED + GROOMING + HOUSEWORK + FOOD_PREP +
    CARETAKING + PLAYING + JOB_SEARCHING + SHOPPING + EATING +
    SOCIALISING + TV + GOLFING + RUNNING + VOLUNTEERING
Model 2: SLEEPING ~ GROOMING
  Res.Df      RSS  Df Sum of Sq      F    Pr(>F)
1   2478 42219819
2   2498 44889139 -20  -2669320 7.8335 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```