STAT463 Project: Sleep Health and Lifestyle

Lisa Lu 31088272

```
# Import libraries
library(dplyr)
library(ggplot2)
library(stringr)
library(gridExtra)
```

Data Exploration

```
# Read in dataset
dataset <- read.table("Sleep_health_and_lifestyle_dataset.csv",
                      header = TRUE, sep = ',', na.strings = "na")

# Print the first few rows of data frame
head(dataset)
```

```
##   Person.ID Gender Age          Occupation Sleep.Duration Quality.of.Sleep
## 1         1   Male  27    Software Engineer            6.1                6
## 2         2   Male  28              Doctor            6.2                6
## 3         3   Male  28              Doctor            6.2                6
## 4         4   Male  28 Sales Representative            5.9                4
## 5         5   Male  28 Sales Representative            5.9                4
## 6         6   Male  28    Software Engineer            5.9                4
##   Physical.Activity.Level Stress.Level BMI.Category Blood.Pressure Heart.Rate
## 1                      42            6   Overweight         126/83         77
## 2                      60            8       Normal         125/80         75
## 3                      60            8       Normal         125/80         75
## 4                      30            8        Obese         140/90         85
## 5                      30            8        Obese         140/90         85
## 6                      30            8        Obese         140/90         85
##   Daily.Steps Sleep.Disorder
## 1        4200           None
## 2       10000           None
## 3       10000           None
## 4        3000    Sleep Apnea
## 5        3000    Sleep Apnea
## 6        3000       Insomnia
```

```
# Explore the structure of the dataset
str(dataset)
```

```
## 'data.frame':    374 obs. of  13 variables:
```

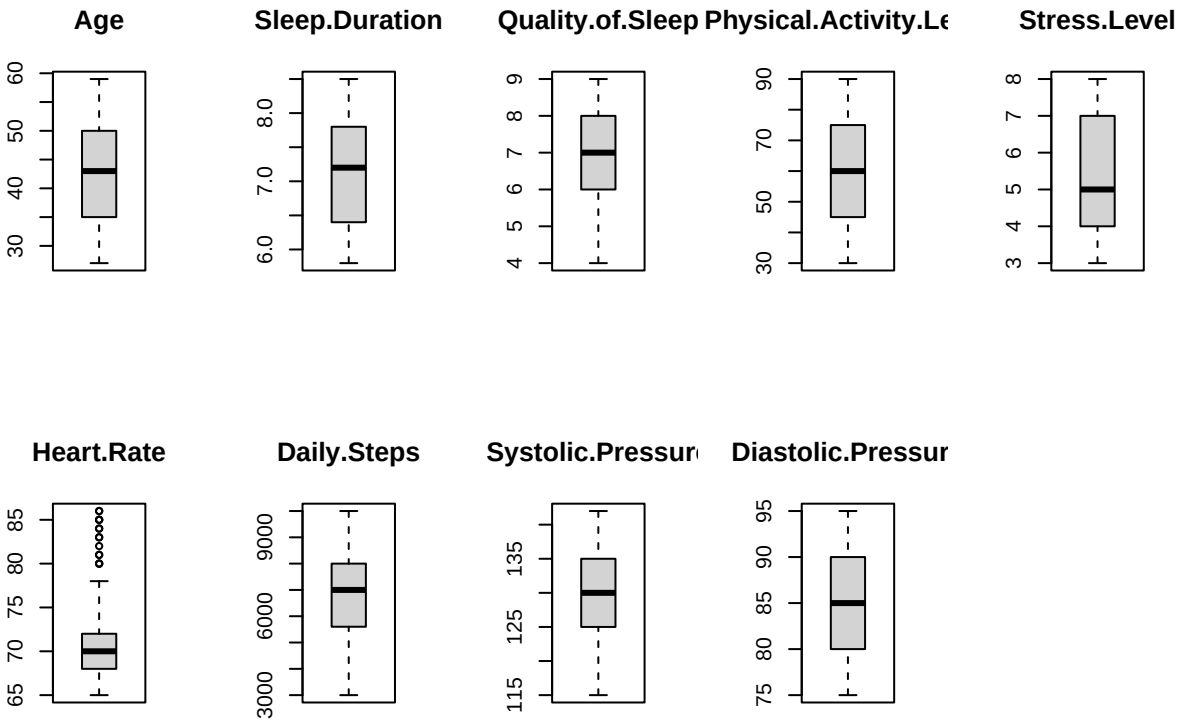
```
##  $ Person.ID              : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Gender                 : chr  "Male" "Male" "Male" "Male" ...
##  $ Age                    : int  27 28 28 28 28 28 29 29 29 29 ...
##  $ Occupation             : chr  "Software Engineer" "Doctor" "Doctor" "Sales Representative" ...
##  $ Sleep.Duration         : num  6.1 6.2 6.2 5.9 5.9 5.9 6.3 7.8 7.8 7.8 ...
##  $ Quality.of.Sleep       : int  6 6 6 4 4 4 6 7 7 7 ...
##  $ Physical.Activity.Level: int  42 60 60 30 30 30 40 75 75 75 ...
##  $ Stress.Level           : int  6 8 8 8 8 8 7 6 6 6 ...
##  $ BMI.Category           : chr  "Overweight" "Normal" "Normal" "Obese" ...
##  $ Blood.Pressure         : chr  "126/83" "125/80" "125/80" "140/90" ...
##  $ Heart.Rate             : int  77 75 75 85 85 85 82 70 70 70 ...
##  $ Daily.Steps            : int  4200 10000 10000 3000 3000 3000 3500 8000 8000 8000 ...
##  $ Sleep.Disorder         : chr  "None" "None" "None" "Sleep Apnea" ...
```

```
# Get a descriptive statistics
summary(dataset)
```

```
##    Person.ID         Gender               Age         Occupation
##  Min.   :  1.00   Length:374         Min.   :27.00   Length:374
##  1st Qu.: 94.25   Class :character   1st Qu.:35.25   Class :character
##  Median :187.50   Mode  :character   Median :43.00   Mode  :character
##  Mean   :187.50                      Mean   :42.18
##  3rd Qu.:280.75                      3rd Qu.:50.00
##  Max.   :374.00                      Max.   :59.00
##  Sleep.Duration  Quality.of.Sleep Physical.Activity.Level  Stress.Level
##  Min.   :5.800   Min.   :4.000    Min.   :30.00           Min.   :3.000
##  1st Qu.:6.400   1st Qu.:6.000    1st Qu.:45.00           1st Qu.:4.000
##  Median :7.200   Median :7.000    Median :60.00           Median :5.000
##  Mean   :7.132   Mean   :7.313    Mean   :59.17           Mean   :5.385
##  3rd Qu.:7.800   3rd Qu.:8.000    3rd Qu.:75.00           3rd Qu.:7.000
##  Max.   :8.500   Max.   :9.000    Max.   :90.00           Max.   :8.000
##  BMI.Category       Blood.Pressure       Heart.Rate     Daily.Steps
##  Length:374         Length:374         Min.   :65.00   Min.   : 3000
##  Class :character   Class :character   1st Qu.:68.00   1st Qu.: 5600
##  Mode  :character   Mode  :character   Median :70.00   Median : 7000
##                                        Mean   :70.17   Mean   : 6817
##                                        3rd Qu.:72.00   3rd Qu.: 8000
##                                        Max.   :86.00   Max.   :10000
##  Sleep.Disorder
##  Length:374
##  Class :character
##  Mode  :character
##
##
##
```

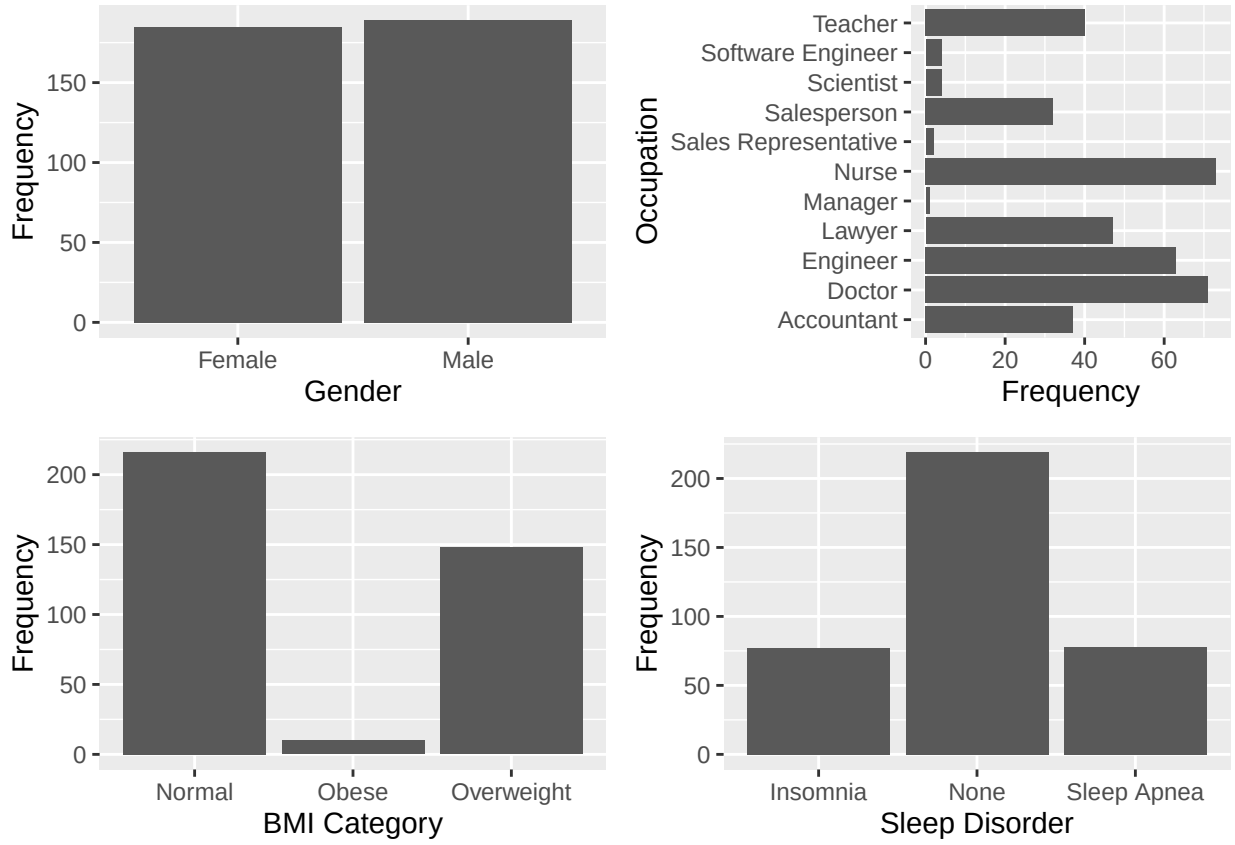
```
# Data preprocessing
# Split Blood Pressure column into systolic and diastolic pressure as numeric data
dataset[c('Systolic.Pressure', 'Diastolic.Pressure')] <- as.numeric(str_split_fixed(dataset$Blood.Pressu

# Combine "Normal" and "Normal Weight" values in BMI.Category
dataset$BMI.Category[dataset$BMI.Category == "Normal Weight"] <- "Normal"
```

```r
# Convert stress level as factors
# categorical.data$Stress.Level <- as.factor(dataset$Stress.Level)


# Separate categorical and numeric data
categorical.data <- dataset[,c("Gender","Occupation", "BMI.Category", "Sleep.Disorder" )]
numeric.data <- dataset[,c(3,5,6,7,8,11,12,14,15)]

# Plot the data -- Boxplot for numeric data and Histogram for categorical data
# Boxplots

par(mfrow=c(2,5))
for (i in 1:length(numeric.data)) {
        boxplot(numeric.data[,i], main=names(numeric.data[i]), type="l")

    }
```




```r
# Frequency chart
gender <- ggplot(data = dataset, aes(x = Gender)) +
geom_bar() +
labs(y = "Frequency", x = "Gender")

occupation <- ggplot(data = dataset, aes(y = Occupation)) +
geom_bar() +
labs(y = "Occupation", x = "Frequency")
```

```
bmi <- ggplot(data = dataset, aes(x = BMI.Category)) +
geom_bar() +
labs(y = "Frequency", x = "BMI Category")

sleep_disorder <- ggplot(data = dataset, aes(x = Sleep.Disorder)) +
geom_bar() +
labs(y = "Frequency", x = "Sleep Disorder")

grid.arrange(gender,occupation, bmi, sleep_disorder, ncol = 2, nrow = 2)
```




```
systolic <- ggplot(data = dataset, aes(x =Systolic.Pressure)) +
geom_bar() +
labs(y = "Frequency", x = "Systolic Pressure")

diastolic <- ggplot(data = dataset, aes(x = Diastolic.Pressure)) +
geom_bar() +
labs(y = "Frequency", x = "Diastolic Pressure")
```

Explore the factors affecting quality of sleep

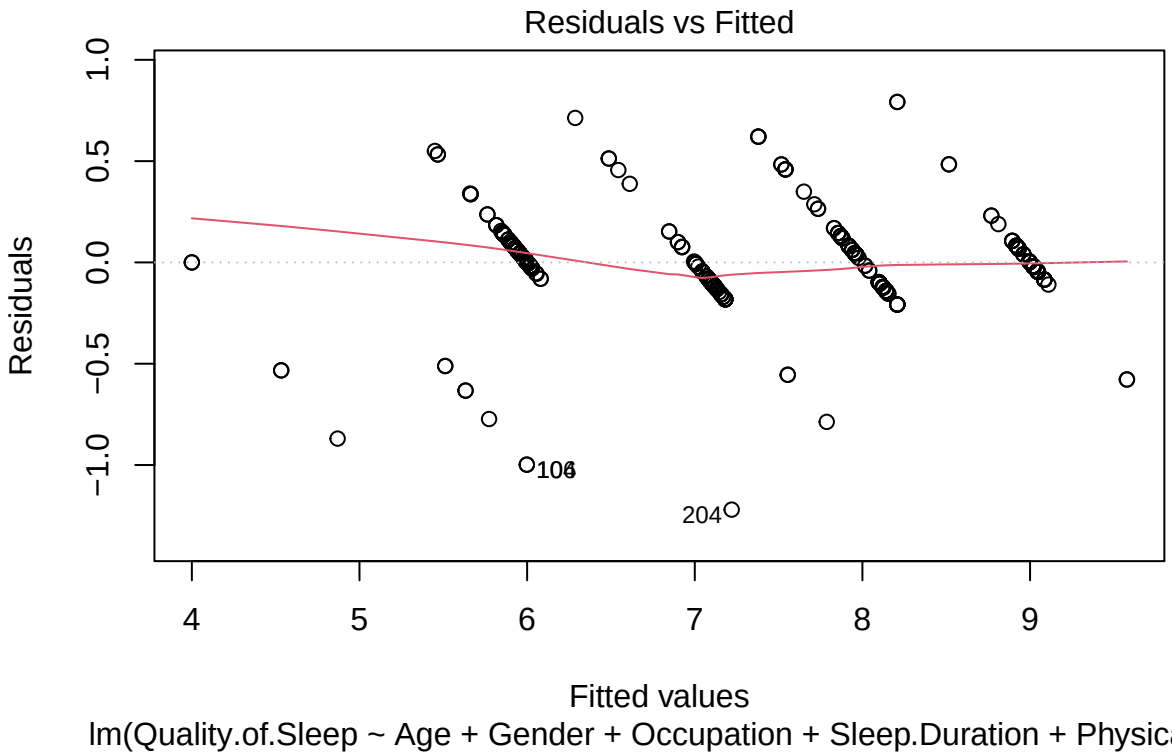
```
# Use linear regression model
sleep <- dataset[, -c(1, 10)]
```

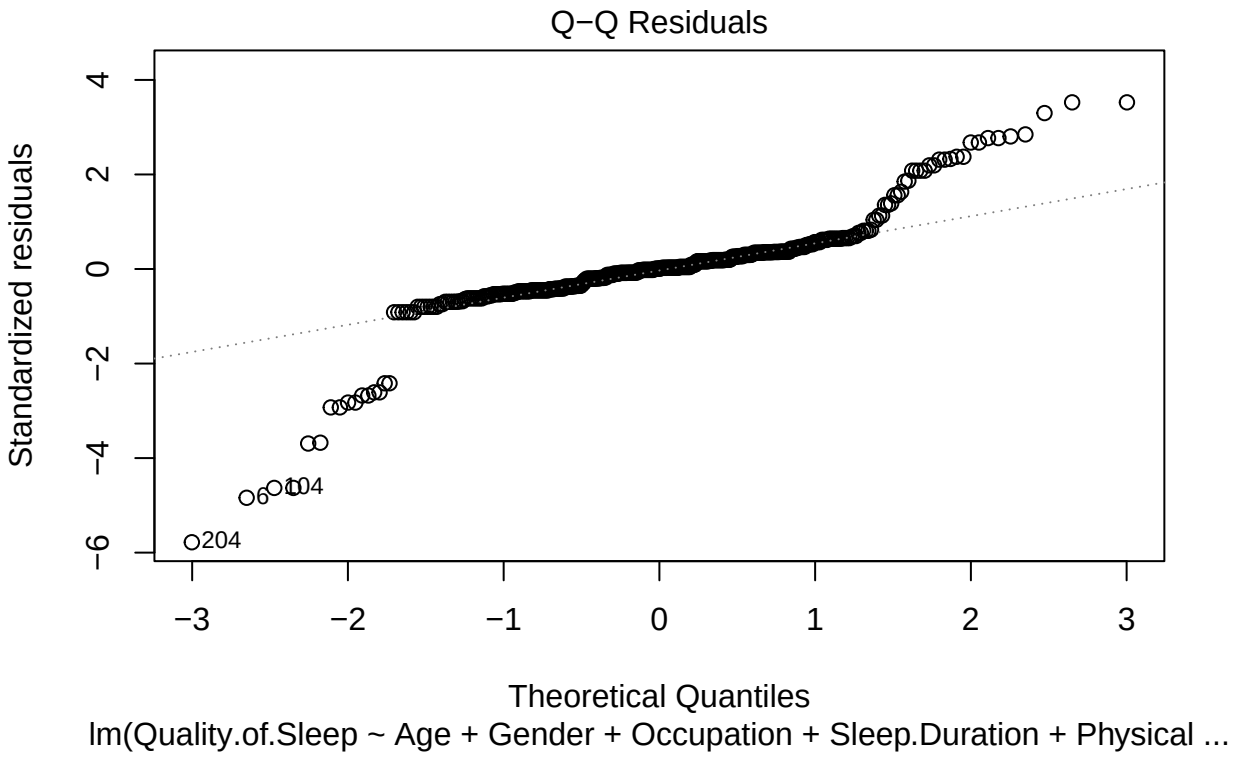
```r
# Get the full multiple linear regression model
lr_full <- lm(Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration + Physical.Activity.Level +

summary(lr_full)
```

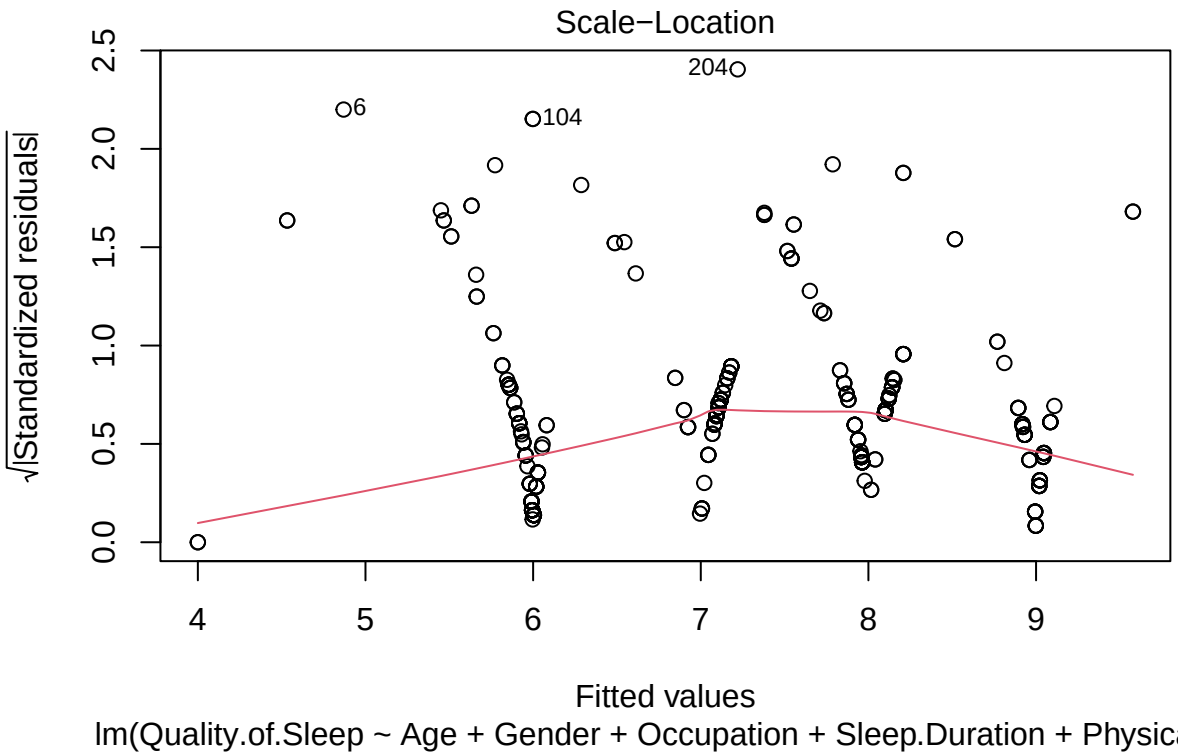
```
##
## Call:
## lm(formula = Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Physical.Activity.Level + Stress.Level + Heart.Rate + Daily.Steps +
##     Systolic.Pressure + Diastolic.Pressure + BMI.Category, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22024 -0.09512  0.00160  0.08038  0.79203
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     8.746e+00  1.003e+00   8.717  < 2e-16 ***
## Age                             6.250e-02  5.215e-03  11.985  < 2e-16 ***
## GenderMale                      6.391e-01  7.248e-02   8.818  < 2e-16 ***
## OccupationDoctor               -4.883e-01  8.497e-02  -5.747 1.97e-08 ***
## OccupationEngineer             -6.404e-01  8.180e-02  -7.829 5.82e-14 ***
## OccupationLawyer               -3.326e-01  9.580e-02  -3.472 0.000581 ***
## OccupationManager              -2.897e-01  2.430e-01  -1.192 0.234016
## OccupationNurse                -1.086e-01  9.296e-02  -1.168 0.243669
## OccupationSales Representative -1.329e+00  2.117e-01  -6.280 9.98e-10 ***
## OccupationSalesperson          -9.861e-01  9.315e-02 -10.586  < 2e-16 ***
## OccupationScientist            -4.390e-01  1.531e-01  -2.867 0.004395 **
## OccupationSoftware Engineer    -4.599e-01  1.396e-01  -3.294 0.001090 **
## OccupationTeacher              -4.996e-01  7.781e-02  -6.421 4.39e-10 ***
## Sleep.Duration                  2.421e-01  4.879e-02   4.962 1.09e-06 ***
## Physical.Activity.Level        -1.166e-03  1.539e-03  -0.758 0.449200
## Stress.Level                   -3.802e-01  2.631e-02 -14.450  < 2e-16 ***
## Heart.Rate                     -2.352e-02  9.411e-03  -2.499 0.012902 *
## Daily.Steps                     4.319e-05  2.137e-05   2.021 0.044076 *
## Systolic.Pressure              -3.892e-03  1.620e-02  -0.240 0.810249
## Diastolic.Pressure             -1.818e-02  2.141e-02  -0.849 0.396526
## BMI.CategoryObese              -1.067e-01  1.843e-01  -0.579 0.563090
## BMI.CategoryOverweight         -4.705e-01  9.063e-02  -5.192 3.53e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2312 on 352 degrees of freedom
## Multiple R-squared:  0.9648, Adjusted R-squared:  0.9627
## F-statistic: 459.2 on 21 and 352 DF,  p-value: < 2.2e-16
```

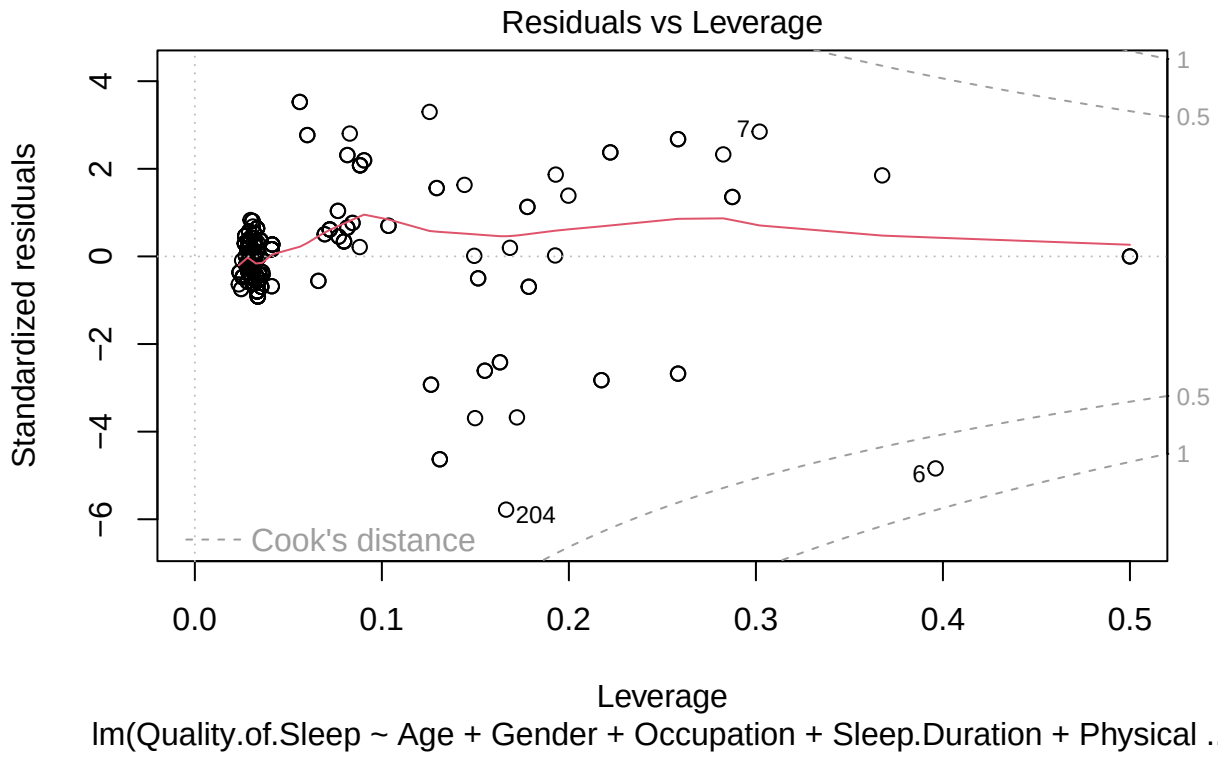
```r
plot(lr_full)
```

```
## Warning: not plotting observations with leverage one:
##   264
```


Residuals vs Fitted
Residuals
1.0
0.5
0.0
−0.5
−1.0
4
5
6
7
8
9
106
204
Fitted values
lm(Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration + Physical ...


Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration + Physical ...
6
104
204


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
6
104
204
4
5
6
7
8
9
Fitted values
lm(Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration + Physical ...




```
lr_backward <- step(lr_full, direction = "backward")
```

```
## Start:  AIC=-1074.03
## Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Physical.Activity.Level + Stress.Level + Heart.Rate + Daily.Steps +
##     Systolic.Pressure + Diastolic.Pressure + BMI.Category
##
##                          Df Sum of Sq    RSS      AIC
## - Systolic.Pressure       1    0.0031 18.822 -1075.96
## - Physical.Activity.Level 1    0.0307 18.850 -1075.42
## - Diastolic.Pressure      1    0.0385 18.858 -1075.26
## <none>                              18.819 -1074.03
## - Daily.Steps             1    0.2183 19.038 -1071.71
## - Heart.Rate              1    0.3339 19.153 -1069.45
## - Sleep.Duration          1    1.3164 20.136 -1050.74
## - BMI.Category            2    1.5538 20.373 -1048.35
## - Gender                  1    4.1575 22.977 -1001.37
## - Age                     1    7.6799 26.499  -948.03
## - Occupation             10   10.8539 29.673  -923.72
## - Stress.Level            1   11.1631 29.982  -901.84
##
## Step:  AIC=-1075.96
## Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Physical.Activity.Level + Stress.Level + Heart.Rate + Daily.Steps +
##     Diastolic.Pressure + BMI.Category
```

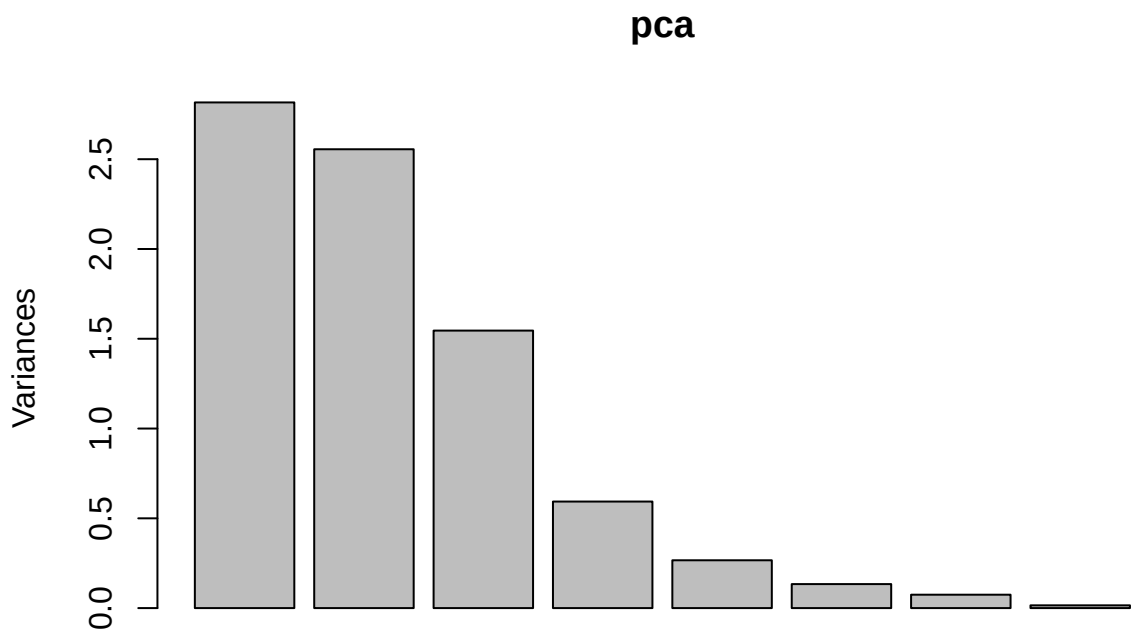
```
##
##                            Df Sum of Sq    RSS      AIC
## - Physical.Activity.Level  1    0.0323 18.855 -1077.32
## <none>                                  18.822 -1075.96
## - Heart.Rate               1    0.3321 19.154 -1071.42
## - Daily.Steps              1    0.3370 19.159 -1071.33
## - Diastolic.Pressure       1    0.5862 19.409 -1066.49
## - Sleep.Duration           1    1.6133 20.436 -1047.21
## - BMI.Category             2    1.9700 20.792 -1042.74
## - Gender                   1    4.3215 23.144 -1000.66
## - Occupation              10   11.6330 30.455  -915.99
## - Stress.Level             1   11.5548 30.377  -898.95
## - Age                      1   12.0491 30.872  -892.91
##
## Step:  AIC=-1077.32
## Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Stress.Level + Heart.Rate + Daily.Steps + Diastolic.Pressure +
##     BMI.Category
##
##                       Df Sum of Sq    RSS      AIC
## <none>                             18.855 -1077.32
## - Heart.Rate           1    0.4823 19.337 -1069.88
## - Daily.Steps          1    0.5654 19.420 -1068.27
## - Diastolic.Pressure   1    0.6597 19.514 -1066.46
## - Sleep.Duration       1    1.6251 20.480 -1048.40
## - BMI.Category         2    1.9533 20.808 -1044.45
## - Gender               1    4.6338 23.489  -997.14
## - Occupation          10   11.9244 30.779  -914.03
## - Stress.Level         1   11.8205 30.675  -897.30
## - Age                  1   12.0602 30.915  -894.39
```

```
summary(lr_backward)
```

```
##
## Call:
## lm(formula = Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Stress.Level + Heart.Rate + Daily.Steps + Diastolic.Pressure +
##     BMI.Category, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20581 -0.09929  0.00322  0.07850  0.78272
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     8.970e+00  7.962e-01  11.266  < 2e-16 ***
## Age                             6.176e-02  4.104e-03  15.048  < 2e-16 ***
## GenderMale                      6.160e-01  6.604e-02   9.327  < 2e-16 ***
## OccupationDoctor               -4.672e-01  8.027e-02  -5.821 1.31e-08 ***
## OccupationEngineer             -6.266e-01  7.629e-02  -8.212 4.11e-15 ***
## OccupationLawyer               -3.152e-01  8.690e-02  -3.627 0.000329 ***
## OccupationManager              -2.906e-01  2.422e-01  -1.200 0.230934
## OccupationNurse                -1.005e-01  9.222e-02  -1.090 0.276379
## OccupationSales Representative -1.304e+00  2.050e-01  -6.362 6.16e-10 ***
```

```
## OccupationSalesperson          -9.679e-01  8.869e-02 -10.914  < 2e-16 ***
## OccupationScientist            -4.328e-01  1.495e-01  -2.895 0.004025 **
## OccupationSoftware Engineer    -4.296e-01  1.333e-01  -3.222 0.001389 **
## OccupationTeacher              -4.989e-01  7.534e-02  -6.622 1.32e-10 ***
## Sleep.Duration                  2.362e-01  4.277e-02   5.524 6.44e-08 ***
## Stress.Level                   -3.771e-01  2.531e-02 -14.897  < 2e-16 ***
## Heart.Rate                     -2.618e-02  8.701e-03  -3.009 0.002807 **
## Daily.Steps                     3.425e-05  1.051e-05   3.258 0.001231 **
## Diastolic.Pressure             -2.403e-02  6.829e-03  -3.519 0.000489 ***
## BMI.CategoryObese              -1.026e-01  1.837e-01  -0.559 0.576586
## BMI.CategoryOverweight         -4.576e-01  8.073e-02  -5.668 2.99e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2308 on 354 degrees of freedom
## Multiple R-squared:  0.9647, Adjusted R-squared:  0.9628
## F-statistic: 509.4 on 19 and 354 DF,  p-value: < 2.2e-16
```

```r
 # Use PCA analysis
# Standardise the Quality of sleep and stress level
Quality.of.Sleep <- data.frame(scale(dataset$Quality.of.Sleep))
Stress.Level <- data.frame(scale(dataset$Stress.Level))
Physical.Level <-data.frame(scale(dataset$Physical.Activity.Level))

pca_dataset <- select(dataset, Age, Physical.Activity.Level, Stress.Level, Sleep.Duration, Heart.Rate, D
pca <- prcomp(scale(pca_dataset))
screeplot(pca)
```




```
summary(pca)
```

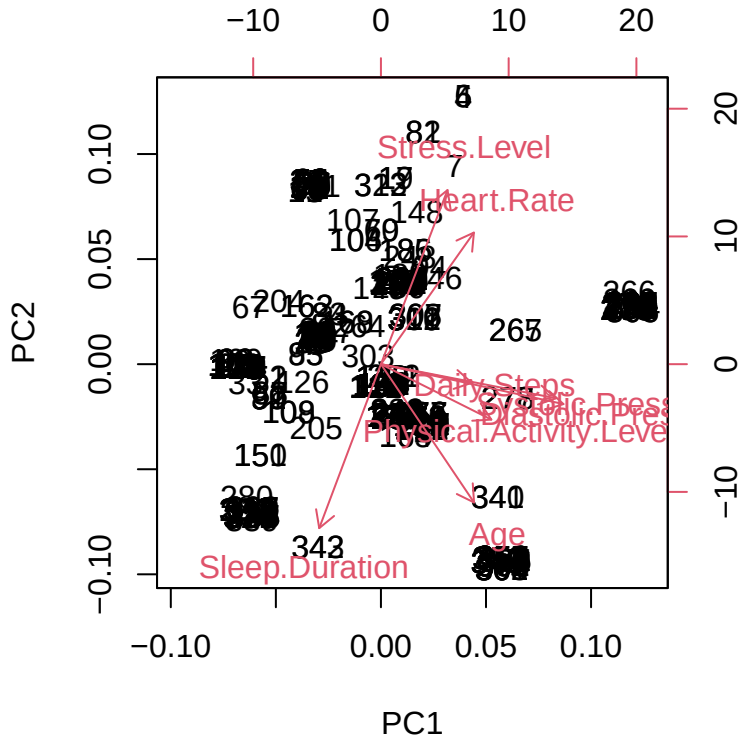
```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     1.678 1.5986 1.2432 0.77024 0.51598 0.36558 0.27302
## Proportion of Variance 0.352 0.3194 0.1932 0.07416 0.03328 0.01671 0.00932
## Cumulative Proportion  0.352 0.6715 0.8646 0.93879 0.97207 0.98878 0.99810
##                           PC8
## Standard deviation     0.1234
## Proportion of Variance 0.0019
## Cumulative Proportion  1.0000
```

```
pca$rotation
```

```
##                                 PC1         PC2         PC3         PC4
## Age                       0.2808238 -0.43769719  0.27248573 -0.14363711
## Physical.Activity.Level   0.3329906 -0.17543724 -0.57607747  0.30983532
## Stress.Level              0.2008117  0.55107947 -0.07629239 -0.15796884
## Sleep.Duration           -0.1860260 -0.51905048 -0.10913196  0.47568105
## Heart.Rate                0.2802932  0.41650382  0.05688401  0.70772865
## Daily.Steps               0.2740358 -0.05901064 -0.66000771 -0.35711415
## Systolic.Pressure         0.5279852 -0.10791817  0.31048115 -0.02246459
## Diastolic.Pressure        0.5503599 -0.12624159  0.20226580 -0.05682280
##                                 PC5         PC6         PC7          PC8
```

```
## Age                          0.76427190  0.1901990  0.11543150 -0.032671389
## Physical.Activity.Level -0.11634837  0.2780977  0.58274938  0.001931354
## Stress.Level                 0.32129927 -0.5909898  0.40619632 -0.087828120
## Sleep.Duration               0.07549208 -0.6669310 -0.07729110 -0.036451708
## Heart.Rate                   0.26540810  0.1661846 -0.38121319  0.021503167
## Daily.Steps                  0.14844879 -0.1019411 -0.55903047  0.115487413
## Systolic.Pressure       -0.29342565 -0.2038101  0.02105469  0.696133317
## Diastolic.Pressure      -0.33888365 -0.1134555 -0.13351966 -0.701059927
```

```
biplot(pca)
```




```
step.model <- step(lr_full, direction = "backward")
```

```
## Start:  AIC=-1074.03
## Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Physical.Activity.Level + Stress.Level + Heart.Rate + Daily.Steps +
##     Systolic.Pressure + Diastolic.Pressure + BMI.Category
##
##                           Df Sum of Sq    RSS      AIC
## - Systolic.Pressure        1    0.0031 18.822 -1075.96
## - Physical.Activity.Level  1    0.0307 18.850 -1075.42
## - Diastolic.Pressure       1    0.0385 18.858 -1075.26
## <none>                                 18.819 -1074.03
## - Daily.Steps              1    0.2183 19.038 -1071.71
## - Heart.Rate               1    0.3339 19.153 -1069.45
```

```
## - Sleep.Duration           1    1.3164 20.136 -1050.74
## - BMI.Category             2    1.5538 20.373 -1048.35
## - Gender                   1    4.1575 22.977 -1001.37
## - Age                      1    7.6799 26.499  -948.03
## - Occupation              10   10.8539 29.673  -923.72
## - Stress.Level             1   11.1631 29.982  -901.84
##
## Step:  AIC=-1075.96
## Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Physical.Activity.Level + Stress.Level + Heart.Rate + Daily.Steps +
##     Diastolic.Pressure + BMI.Category
##
##                           Df Sum of Sq    RSS      AIC
## - Physical.Activity.Level  1    0.0323 18.855 -1077.32
## <none>                                 18.822 -1075.96
## - Heart.Rate               1    0.3321 19.154 -1071.42
## - Daily.Steps              1    0.3370 19.159 -1071.33
## - Diastolic.Pressure       1    0.5862 19.409 -1066.49
## - Sleep.Duration           1    1.6133 20.436 -1047.21
## - BMI.Category             2    1.9700 20.792 -1042.74
## - Gender                   1    4.3215 23.144 -1000.66
## - Occupation              10   11.6330 30.455  -915.99
## - Stress.Level             1   11.5548 30.377  -898.95
## - Age                      1   12.0491 30.872  -892.91
##
## Step:  AIC=-1077.32
## Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Stress.Level + Heart.Rate + Daily.Steps + Diastolic.Pressure +
##     BMI.Category
##
##                       Df Sum of Sq    RSS      AIC
## <none>                             18.855 -1077.32
## - Heart.Rate           1    0.4823 19.337 -1069.88
## - Daily.Steps          1    0.5654 19.420 -1068.27
## - Diastolic.Pressure   1    0.6597 19.514 -1066.46
## - Sleep.Duration       1    1.6251 20.480 -1048.40
## - BMI.Category         2    1.9533 20.808 -1044.45
## - Gender               1    4.6338 23.489  -997.14
## - Occupation          10   11.9244 30.779  -914.03
## - Stress.Level         1   11.8205 30.675  -897.30
## - Age                  1   12.0602 30.915  -894.39
```

```
summary(step.model)
```

```
##
## Call:
## lm(formula = Quality.of.Sleep ~ Age + Gender + Occupation + Sleep.Duration +
##     Stress.Level + Heart.Rate + Daily.Steps + Diastolic.Pressure +
##     BMI.Category, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20581 -0.09929  0.00322  0.07850  0.78272
##
```

```
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     8.970e+00  7.962e-01  11.266  < 2e-16 ***
## Age                             6.176e-02  4.104e-03  15.048  < 2e-16 ***
## GenderMale                      6.160e-01  6.604e-02   9.327  < 2e-16 ***
## OccupationDoctor               -4.672e-01  8.027e-02  -5.821 1.31e-08 ***
## OccupationEngineer             -6.266e-01  7.629e-02  -8.212 4.11e-15 ***
## OccupationLawyer               -3.152e-01  8.690e-02  -3.627 0.000329 ***
## OccupationManager              -2.906e-01  2.422e-01  -1.200 0.230934
## OccupationNurse                -1.005e-01  9.222e-02  -1.090 0.276379
## OccupationSales Representative -1.304e+00  2.050e-01  -6.362 6.16e-10 ***
## OccupationSalesperson          -9.679e-01  8.869e-02 -10.914  < 2e-16 ***
## OccupationScientist            -4.328e-01  1.495e-01  -2.895 0.004025 **
## OccupationSoftware Engineer    -4.296e-01  1.333e-01  -3.222 0.001389 **
## OccupationTeacher              -4.989e-01  7.534e-02  -6.622 1.32e-10 ***
## Sleep.Duration                  2.362e-01  4.277e-02   5.524 6.44e-08 ***
## Stress.Level                   -3.771e-01  2.531e-02 -14.897  < 2e-16 ***
## Heart.Rate                     -2.618e-02  8.701e-03  -3.009 0.002807 **
## Daily.Steps                     3.425e-05  1.051e-05   3.258 0.001231 **
## Diastolic.Pressure             -2.403e-02  6.829e-03  -3.519 0.000489 ***
## BMI.CategoryObese              -1.026e-01  1.837e-01  -0.559 0.576586
## BMI.CategoryOverweight         -4.576e-01  8.073e-02  -5.668 2.99e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2308 on 354 degrees of freedom
## Multiple R-squared:  0.9647, Adjusted R-squared:  0.9628
## F-statistic: 509.4 on 19 and 354 DF,  p-value: < 2.2e-16
```

Explore the causes of sleep disorder

```
# Multinomial Logistic Regression: predict the probablities of categorically dependent variable
library(nnet)
mlr <- multinom(Sleep.Disorder ~ Gender + Occupation + Stress.Level + Physical.Activity.Level, data = sl
```

```
## # weights:  45 (28 variable)
## initial  value 410.880996
## iter  10 value 187.520509
## iter  20 value 165.987919
## iter  30 value 165.007256
## iter  40 value 164.985977
## iter  50 value 164.985254
## final  value 164.985250
## converged
```

```
summary(mlr)
```

```
## Call:
## multinom(formula = Sleep.Disorder ~ Gender + Occupation + Stress.Level +
##     Physical.Activity.Level, data = sleep)
```

```
##
## Coefficients:
##             (Intercept) GenderMale OccupationDoctor OccupationEngineer
## None           4.174937  0.9987048         2.512166        -0.02797914
## Sleep Apnea  -15.661379  0.9581789        16.336806        12.78346560
##             OccupationLawyer OccupationManager OccupationNurse
## None               0.6853457        13.8571159       0.2371525
## Sleep Apnea       14.4654205        -0.9224896      18.4769314
##             OccupationSales Representative OccupationSalesperson
## None                             -7.138586              -3.47016
## Sleep Apnea                      36.052875              12.89285
##             OccupationScientist OccupationSoftware Engineer OccupationTeacher
## None                  22.71830                  -0.2097068         -2.875763
## Sleep Apnea           39.87144                 -10.0204086         13.754666
##             Stress.Level Physical.Activity.Level
## None          -0.7117069              0.01342945
## Sleep Apnea   -0.5291236              0.04769968
##
## Std. Errors:
##             (Intercept) GenderMale OccupationDoctor OccupationEngineer
## None           1.614257  0.8720345         1.133950          0.8875872
## Sleep Apnea    1.529513  1.4111518         1.033262          1.1554892
##             OccupationLawyer OccupationManager OccupationNurse
## None                1.096940      3.389199e-07        1.010095
## Sleep Apnea         1.101842      8.552055e-14        1.242848
##             OccupationSales Representative OccupationSalesperson
## None                         3.939857e-11              1.169389
## Sleep Apnea                  5.012748e-08              1.157615
##             OccupationScientist OccupationSoftware Engineer OccupationTeacher
## None                  0.6847528               1.607970e+00         0.6806111
## Sleep Apnea           0.6847528               2.508357e-11         0.7881444
##             Stress.Level Physical.Activity.Level
## None           0.2141819              0.01538044
## Sleep Apnea    0.2340595              0.01769048
##
## Residual Deviance: 329.9705
## AIC: 385.9705
```

```
exp(summary(mlr)$coefficients)
```

```
##              (Intercept) GenderMale OccupationDoctor OccupationEngineer
## None        6.503573e+01   2.714763     1.233161e+01       9.724086e-01
## Sleep Apnea 1.578882e-07   2.606945     1.244471e+07       3.562776e+05
##             OccupationLawyer OccupationManager OccupationNurse
## None            1.984458e+00      1.042483e+06    1.267634e+00
## Sleep Apnea     1.915368e+06      3.975281e-01    1.057863e+08
##             OccupationSales Representative OccupationSalesperson
## None                         7.938735e-04          3.111206e-02
## Sleep Apnea                  4.545323e+15          3.974610e+05
##             OccupationScientist OccupationSoftware Engineer OccupationTeacher
## None              7.352486e+09               8.108219e-01      5.637313e-02
## Sleep Apnea       2.069893e+17               4.448277e-05      9.409694e+05
##             Stress.Level Physical.Activity.Level
## None           0.4908057               1.013520
```

```
## Sleep Apnea    0.5891211                1.048856
```

```
step(mlr, direction = 'backward')
```

```
## Start:  AIC=385.97
## Sleep.Disorder ~ Gender + Occupation + Stress.Level + Physical.Activity.Level
##
## trying - Gender
## # weights:  42 (26 variable)
## initial  value 410.880996
## iter  10 value 177.332652
## iter  20 value 166.100777
## iter  30 value 165.664981
## iter  40 value 165.659175
## final  value 165.659127
## converged
## trying - Occupation
## # weights:  15 (8 variable)
## initial  value 410.880996
## iter  10 value 275.973053
## final  value 275.307174
## converged
## trying - Stress.Level
## # weights:  42 (26 variable)
## initial  value 410.880996
## iter  10 value 181.771178
## iter  20 value 172.152231
## iter  30 value 171.327531
## iter  40 value 171.306792
## final  value 171.306663
## converged
## trying - Physical.Activity.Level
## # weights:  42 (26 variable)
## initial  value 410.880996
## iter  10 value 184.396149
## iter  20 value 170.255837
## iter  30 value 169.188662
## iter  40 value 169.165176
## final  value 169.164844
## converged
##                            Df      AIC
## - Gender                   26 383.3183
## <none>                     28 385.9705
## - Physical.Activity.Level 26 390.3297
## - Stress.Level             26 394.6133
## - Occupation                8 566.6143
## # weights:  42 (26 variable)
## initial  value 410.880996
## iter  10 value 177.332652
## iter  20 value 166.100777
## iter  30 value 165.664981
## iter  40 value 165.659175
## final  value 165.659127
## converged
```

```
##
## Step:  AIC=383.32
## Sleep.Disorder ~ Occupation + Stress.Level + Physical.Activity.Level
##
## trying - Occupation
## # weights:  12 (6 variable)
## initial  value 410.880996
## iter  10 value 317.374616
## final  value 317.373969
## converged
## trying - Stress.Level
## # weights:  39 (24 variable)
## initial  value 410.880996
## iter  10 value 175.993980
## iter  20 value 171.714846
## iter  30 value 171.639418
## iter  40 value 171.638355
## iter  40 value 171.638354
## iter  40 value 171.638354
## final  value 171.638354
## converged
## trying - Physical.Activity.Level
## # weights:  39 (24 variable)
## initial  value 410.880996
## iter  10 value 176.721991
## iter  20 value 170.811786
## iter  30 value 170.466124
## iter  40 value 170.465132
## final  value 170.465129
## converged
##                            Df      AIC
## <none>                     26 383.3183
## - Physical.Activity.Level 24 388.9303
## - Stress.Level            24 391.2767
## - Occupation               6 646.7479

## Call:
## multinom(formula = Sleep.Disorder ~ Occupation + Stress.Level +
##     Physical.Activity.Level, data = sleep)
##
## Coefficients:
##             (Intercept) OccupationDoctor OccupationEngineer OccupationLawyer
## None            3.15448          3.28623          0.5967847         1.464939
## Sleep Apnea   -17.44269         18.14536         14.4517381        16.288232
##             OccupationManager OccupationNurse OccupationSales Representative
## None                12.239080     -0.09764979                      -7.428463
## Sleep Apnea         -1.655021     19.19850370                      37.022331
##             OccupationSalesperson OccupationScientist
## None                    -2.663533            18.94831
## Sleep Apnea             14.688950            37.11949
##             OccupationSoftware Engineer OccupationTeacher Stress.Level
## None                          0.6758843         -2.650177   -0.5865970
## Sleep Apnea                  -6.1401651         14.952246   -0.4238748
##             Physical.Activity.Level
```

```
## None                     0.02091099
## Sleep Apnea              0.05229846
##
## Residual Deviance: 331.3183
## AIC: 383.3183
```

Generalised Linear Mixed Model

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
m1 <- lmer(Quality.of.Sleep ~ Sleep.Duration + Stress.Level + (1|Occupation), data = sleep)
summary(m1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Quality.of.Sleep ~ Sleep.Duration + Stress.Level + (1 | Occupation)
##    Data: sleep
##
## REML criterion at convergence: 257
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7982 -0.5191  0.1304  0.5253  2.4210
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Occupation (Intercept) 0.3545   0.5954
##  Residual               0.1005   0.3170
## Number of obs: 374, groups:  Occupation, 11
##
## Fixed effects:
##                Estimate Std. Error t value
## (Intercept)     6.43184    0.51339  12.528
## Sleep.Duration  0.40225    0.05211   7.719
## Stress.Level   -0.42721    0.02358 -18.120
##
## Correlation of Fixed Effects:
##             (Intr) Slp.Dr
## Sleep.Durtn -0.924
## Stress.Levl -0.868  0.869
```

```
m2 <- lmer(Quality.of.Sleep ~ Sleep.Duration + (1|Occupation), data=sleep)
summary(m2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Quality.of.Sleep ~ Sleep.Duration + (1 | Occupation)
##    Data: sleep
##
```

```
## REML criterion at convergence: 485.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.2502 -0.4130  0.0418  0.6357  2.2046
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Occupation (Intercept) 0.3603   0.6003
##  Residual               0.1913   0.4374
## Number of obs: 374, groups:  Occupation, 11
##
## Fixed effects:
##                Estimate Std. Error t value
## (Intercept)    -1.65672    0.30774  -5.384
## Sleep.Duration  1.22701    0.03544  34.622
##
## Correlation of Fixed Effects:
##             (Intr)
## Sleep.Durtn -0.790
```