```r
setwd("~/Documents/Classes/Stat423")
df <- read.csv("sleep_cycle_productivity.csv")

df$Gender <- as.factor(df$Gender)

rating_vars <- c("Sleep.Quality", "Productivity.Score", "Mood.Score", "Stress.Level")
for (var in rating_vars) {
  new_var <- paste0(var, "Cat")
  df[[new_var]] <- cut(df[[var]], breaks = c(0, 3, 7, 10),
                       labels = c("Low", "Medium", "High"), right = TRUE)
  df[[new_var]] <- as.factor(df[[new_var]])
}

summary(df)
```

```
##      Date              Person_ID         Age           Gender
##  Length:5000        Min.   :1000   Min.   :18.00   Female:1675
##  Class :character   1st Qu.:3258   1st Qu.:28.00   Male  :1718
##  Mode  :character   Median :5603   Median :39.00   Other :1607
##                     Mean   :5527   Mean   :38.59
##                     3rd Qu.:7750   3rd Qu.:49.00
##                     Max.   :9998   Max.   :59.00
##  Sleep.Start.Time Sleep.End.Time Total.Sleep.Hours Sleep.Quality
##  Min.   :20.00    Min.   :0.56   Min.   :4.500     Min.   : 1.000
##  1st Qu.:21.02    1st Qu.:3.66   1st Qu.:5.690     1st Qu.: 3.000
##  Median :22.02    Median :4.97   Median :6.960     Median : 5.000
##  Mean   :22.01    Mean   :4.98   Mean   :6.975     Mean   : 5.521
##  3rd Qu.:23.00    3rd Qu.:6.31   3rd Qu.:8.210     3rd Qu.: 8.000
##  Max.   :23.98    Max.   :9.42   Max.   :9.500     Max.   :10.000
##  Exercise..mins.day. Caffeine.Intake..mg. Screen.Time.Before.Bed..mins.
##  Min.   : 0.00       Min.   :  0.0        Min.   :  0.00
##  1st Qu.:22.00       1st Qu.: 73.0        1st Qu.: 46.00
##  Median :44.00       Median :144.0        Median : 92.00
##  Mean   :43.96       Mean   :146.7        Mean   : 91.42
##  3rd Qu.:66.00       3rd Qu.:220.0        3rd Qu.:136.00
##  Max.   :89.00       Max.   :299.0        Max.   :179.00
##  Work.Hours..hrs.day. Productivity.Score   Mood.Score      Stress.Level
##  Min.   : 4.000       Min.   : 1.000     Min.   : 1.000   Min.   : 1.000
##  1st Qu.: 6.033       1st Qu.: 3.000     1st Qu.: 3.000   1st Qu.: 3.000
##  Median : 7.998       Median : 6.000     Median : 5.000   Median : 6.000
##  Mean   : 7.988       Mean   : 5.644     Mean   : 5.371   Mean   : 5.548
##  3rd Qu.: 9.905       3rd Qu.: 8.000     3rd Qu.: 8.000   3rd Qu.: 8.000
##  Max.   :11.999       Max.   :10.000     Max.   :10.000   Max.   :10.000
##  Sleep.QualityCat Productivity.ScoreCat Mood.ScoreCat Stress.LevelCat
##  Low   :1473      Low   :1428           Low   :1561   Low   :1489
##  Medium:2035      Medium:1964           Medium:2047   Medium:1966
##  High  :1492      High  :1608           High  :1392   High  :1545
##
##
##
```

```r
head(df)
```

```
##         Date Person_ID Age Gender Sleep.Start.Time Sleep.End.Time
```

```
## 1 2024-04-12      1860  32  Other            23.33           4.61
## 2 2024-11-04      1769  41 Female            21.02           2.43
## 3 2024-08-31      2528  20   Male            22.10           3.45
## 4 2024-02-22      8041  37  Other            23.10           6.65
## 5 2024-02-23      4843  46  Other            21.42           4.17
## 6 2024-07-08      7439  38   Male            21.77           6.41
##   Total.Sleep.Hours Sleep.Quality Exercise..mins.day. Caffeine.Intake..mg.
## 1              5.28             3                  86                   87
## 2              5.41             5                  32                   21
## 3              5.35             7                  17                   88
## 4              7.55             8                  46                   34
## 5              6.75            10                  61                  269
## 6              8.64            10                  88                  251
##   Screen.Time.Before.Bed..mins. Work.Hours..hrs.day. Productivity.Score
## 1                           116            8.808920                   8
## 2                            88            6.329833                  10
## 3                            59            8.506306                  10
## 4                            80            6.070240                   8
## 5                            94           11.374994                   8
## 6                           123            6.207993                   1
##   Mood.Score Stress.Level Sleep.QualityCat Productivity.ScoreCat Mood.ScoreCat
## 1          3            6              Low                  High           Low
## 2          3            7           Medium                  High           Low
## 3          9           10           Medium                  High          High
## 4          4            2             High                  High        Medium
## 5          7            9             High                  High        Medium
## 6          9            7             High                   Low          High
##   Stress.LevelCat
## 1          Medium
## 2          Medium
## 3            High
## 4             Low
## 5            High
## 6          Medium
```

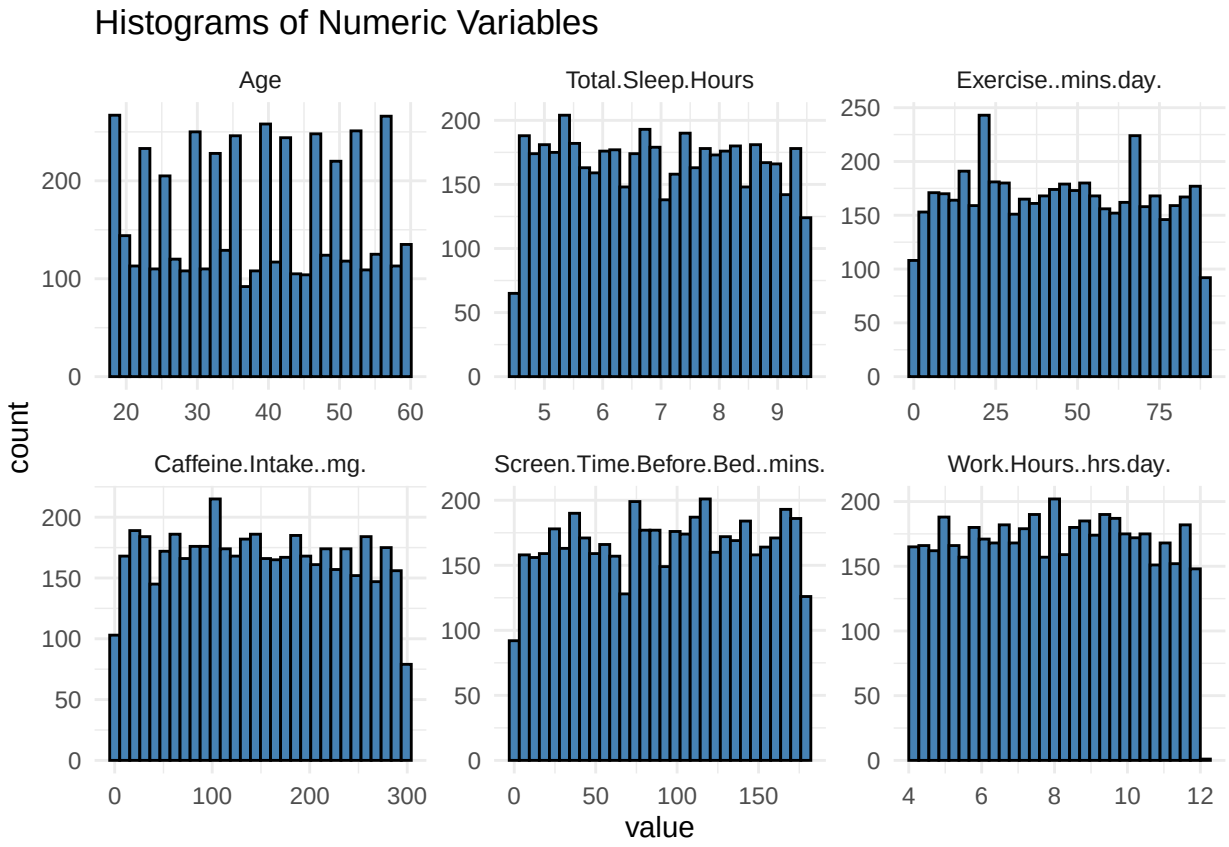
```r
numeric_vars <- c("Age", "Total.Sleep.Hours", "Exercise..mins.day.",
                  "Caffeine.Intake..mg.", "Screen.Time.Before.Bed..mins.", "Work.Hours..hrs.day.")

df_numeric <- df[, numeric_vars]
summary(df_numeric)
```

```
##       Age        Total.Sleep.Hours Exercise..mins.day. Caffeine.Intake..mg.
##  Min.   :18.00   Min.   :4.500     Min.   : 0.00       Min.   :  0.0
##  1st Qu.:28.00   1st Qu.:5.690     1st Qu.:22.00       1st Qu.: 73.0
##  Median :39.00   Median :6.960     Median :44.00       Median :144.0
##  Mean   :38.59   Mean   :6.975     Mean   :43.96       Mean   :146.7
##  3rd Qu.:49.00   3rd Qu.:8.210     3rd Qu.:66.00       3rd Qu.:220.0
##  Max.   :59.00   Max.   :9.500     Max.   :89.00       Max.   :299.0
##  Screen.Time.Before.Bed..mins. Work.Hours..hrs.day.
##  Min.   :  0.00                Min.   : 4.000
##  1st Qu.: 46.00                1st Qu.: 6.033
##  Median : 92.00                Median : 7.998
##  Mean   : 91.42                Mean   : 7.988
##  3rd Qu.:136.00                3rd Qu.: 9.905
##  Max.   :179.00                Max.   :11.999
```

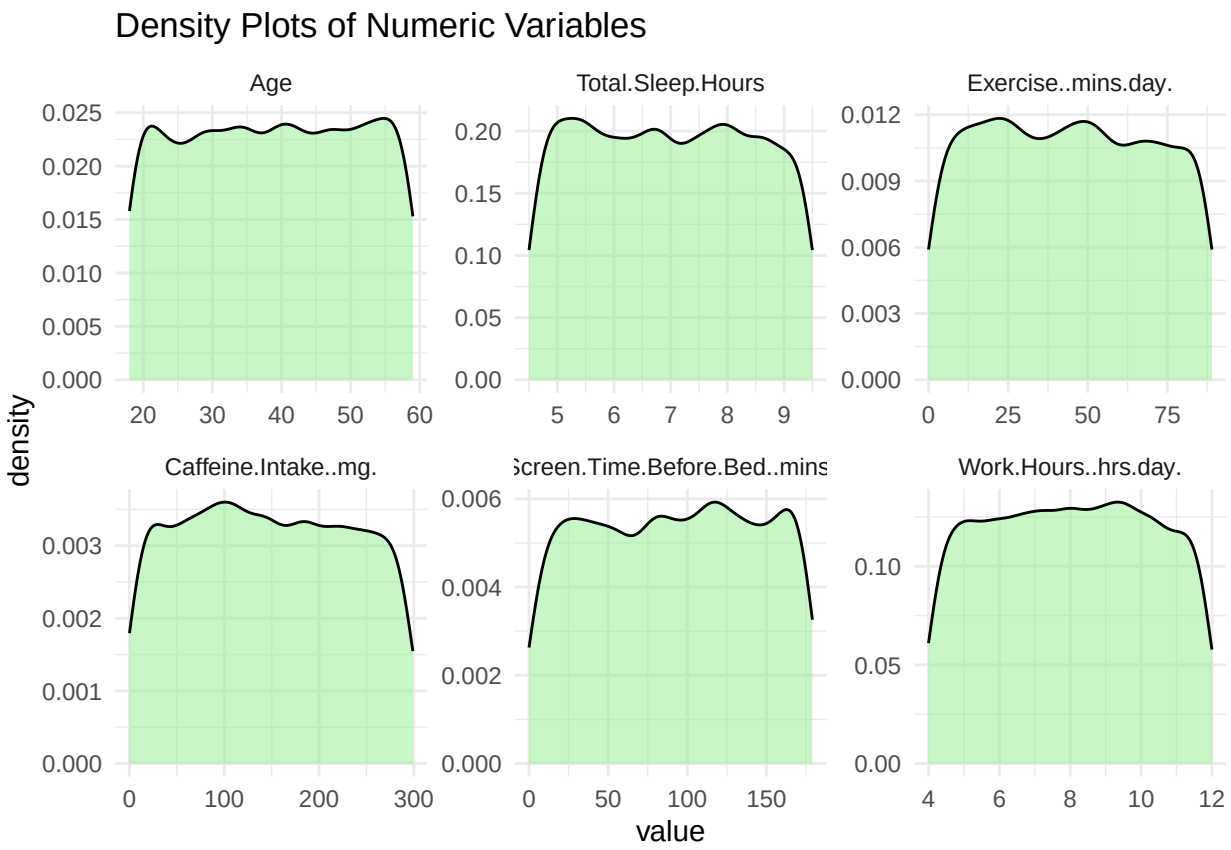
```
library(ggplot2)
library(reshape2)
df_numeric_long <- melt(df_numeric)
```

```
## No id variables; using all as measure variables
```

```
ggplot(df_numeric_long, aes(x = value)) +
  geom_histogram(bins = 30, fill = "steelblue", color = "black") +
  facet_wrap(~ variable, scales = "free") +
  theme_minimal() +
  ggtitle("Histograms of Numeric Variables")
```

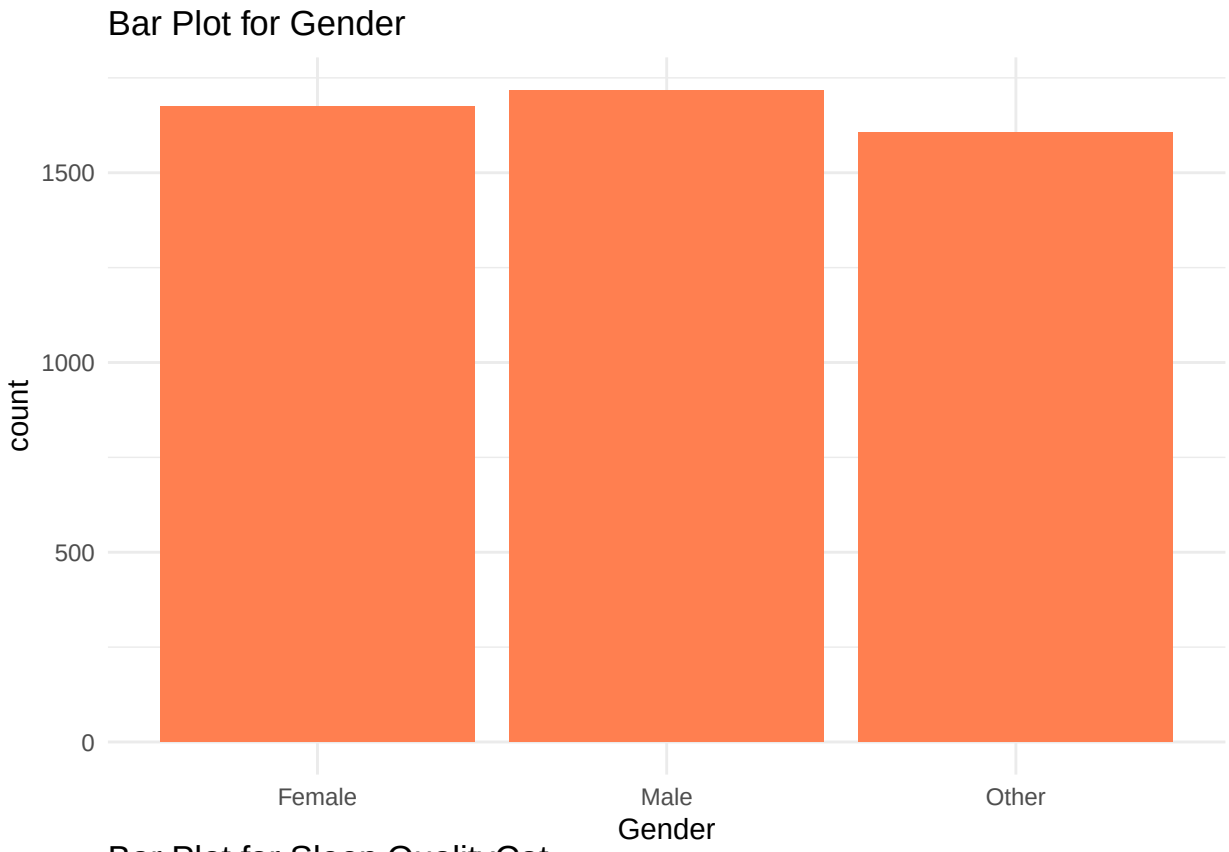


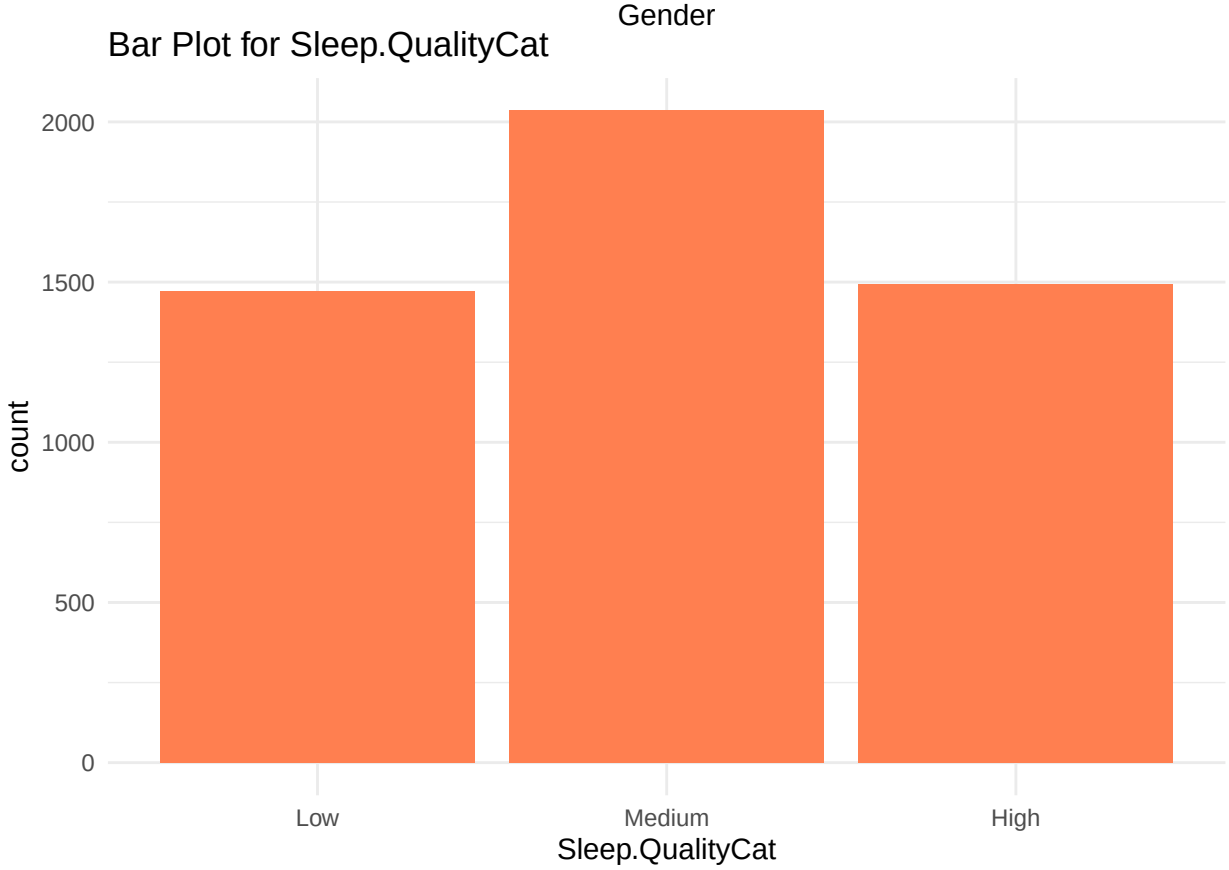

```
ggplot(df_numeric_long, aes(x = value)) +
  geom_density(fill = "lightgreen", alpha = 0.5) +
  facet_wrap(~ variable, scales = "free") +
  theme_minimal() +
  ggtitle("Density Plots of Numeric Variables")
```

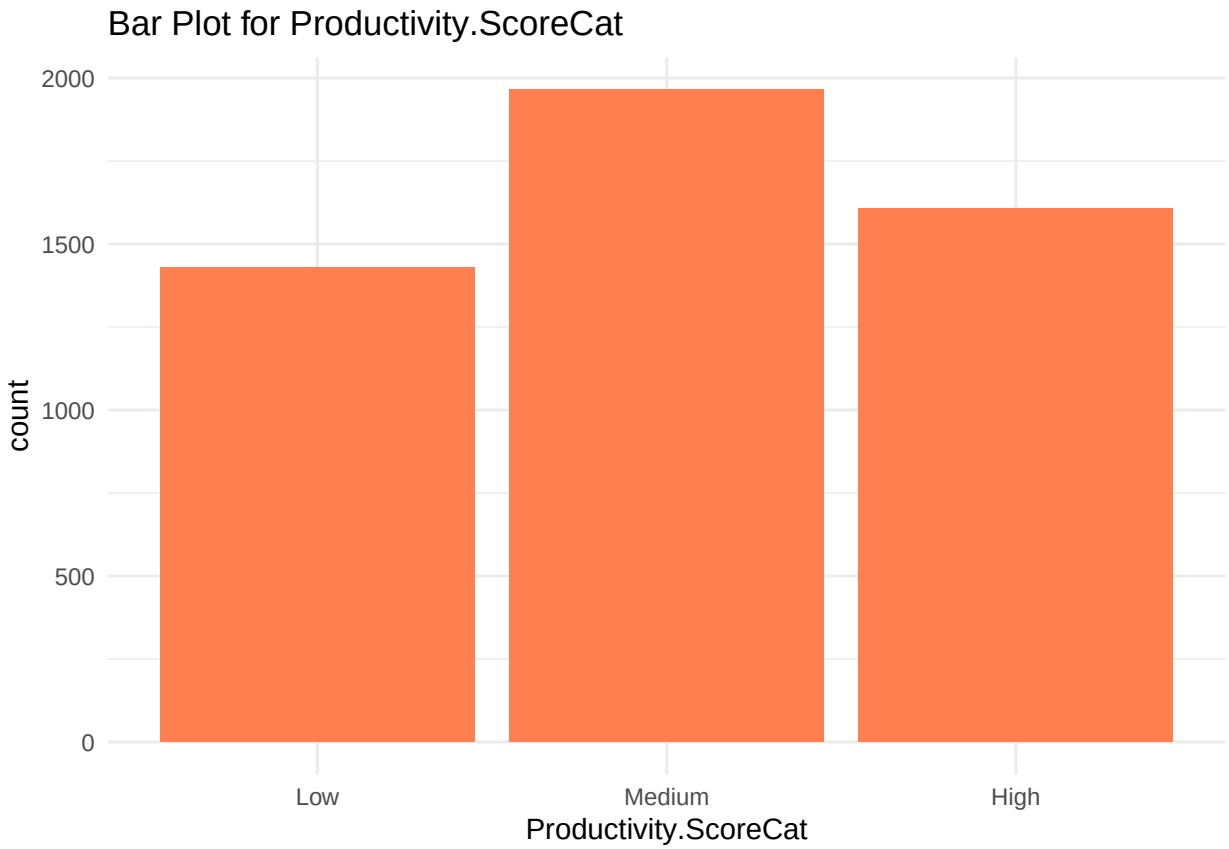


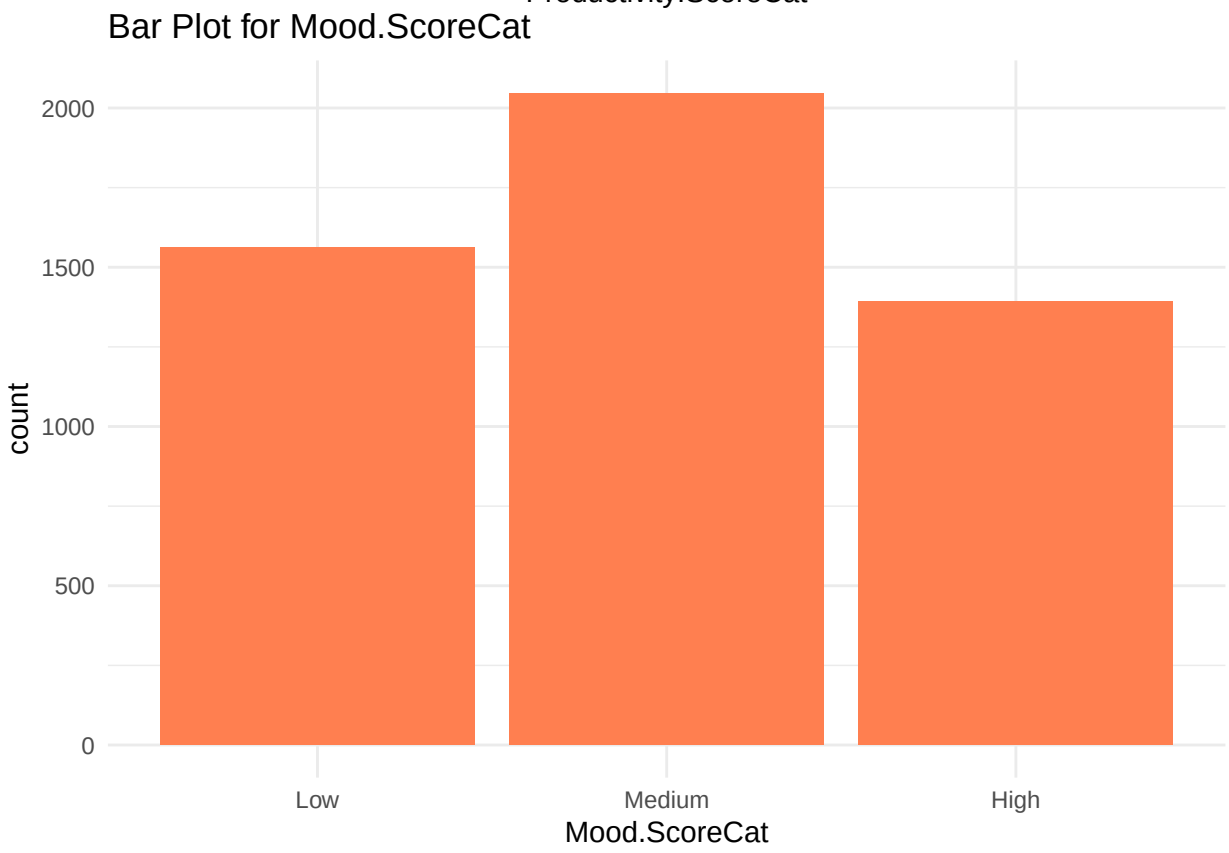

```
categorical_vars <- c("Gender", "Sleep.QualityCat", "Productivity.ScoreCat", "Mood.ScoreCat", "Stress.L

par(mfrow = c(3,2))

for (var in categorical_vars) {
  print(ggplot(df, aes_string(x = var)) +
  geom_bar(fill = "coral") +
  theme_minimal() +
  ggtitle(paste("Bar Plot for", var)))
}
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```


Bar Plot for Gender
1500
1000
500
0
count
Female
Male
Other
Gender


Bar Plot for Sleep.QualityCat
2000
1500
1000
500
0
count
Low
Medium
High
Sleep.QualityCat

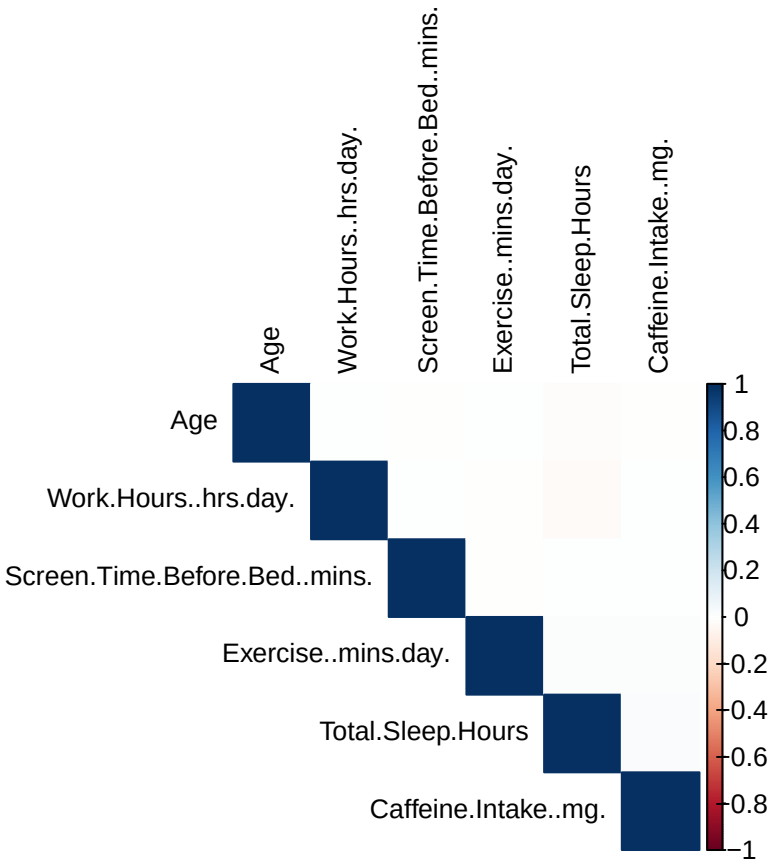

Bar Plot for Productivity.ScoreCat
2000
1500
1000
500
0
count
Low
Medium
High
Productivity.ScoreCat


Bar Plot for Mood.ScoreCat
2000
1500
1000
500
0
count
Low
Medium
High
Mood.ScoreCat


```
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```
numeric_vars <- c("Age", "Total.Sleep.Hours", "Exercise..mins.day.",
                  "Caffeine.Intake..mg.", "Screen.Time.Before.Bed..mins.", "Work.Hours..hrs.day.",
                  "Sleep.Quality")

#df_numeric <- df[, numeric_vars]

cor_matrix <- cor(df_numeric, use = "complete.obs")
corrplot(cor_matrix, method = "color", type = "upper", order = "hclust",
         tl.col = "black", tl.cex = 0.8)
```




```
setwd("~/Documents/Classes/Stat423")
df <- read.csv("sleep_cycle_productivity.csv")
df$Gender <- as.factor(df$Gender)

rating_vars <- c("Sleep.Quality", "Productivity.Score", "Mood.Score", "Stress.Level")
for (var in rating_vars) {
  new_var <- paste0(var, "Cat")
  df[[new_var]] <- cut(df[[var]],
                       breaks = c(0, 3, 7, 10),
                       labels = c("Low", "Medium", "High"),
                       right = TRUE)
  df[[new_var]] <- as.factor(df[[new_var]])
}

df_model <- df[, !(names(df) %in% c("Date", "Person_ID", "Sleep.Start.Time", "Sleep.End.Time",
                                    "Sleep.Quality", "Productivity.Score", "Mood.Score", "Stress.Level"))]

library(bestNormalize)

bn_sleep <- bestNormalize(df_model$Total.Sleep.Hours)
```

```
## Warning: `progress_estimated()` was deprecated in dplyr 1.0.0.
## i The deprecated feature was likely used in the bestNormalize package.
##   Please report the issue to the authors.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
print(bn_sleep$chosen_transform)
```

```
## orderNorm Transformation with 5000 nonmissing obs and ties
##  - 501 unique values
##  - Original quantiles:
##   0%  25%  50%  75% 100%
## 4.50 5.69 6.96 8.21 9.50
```

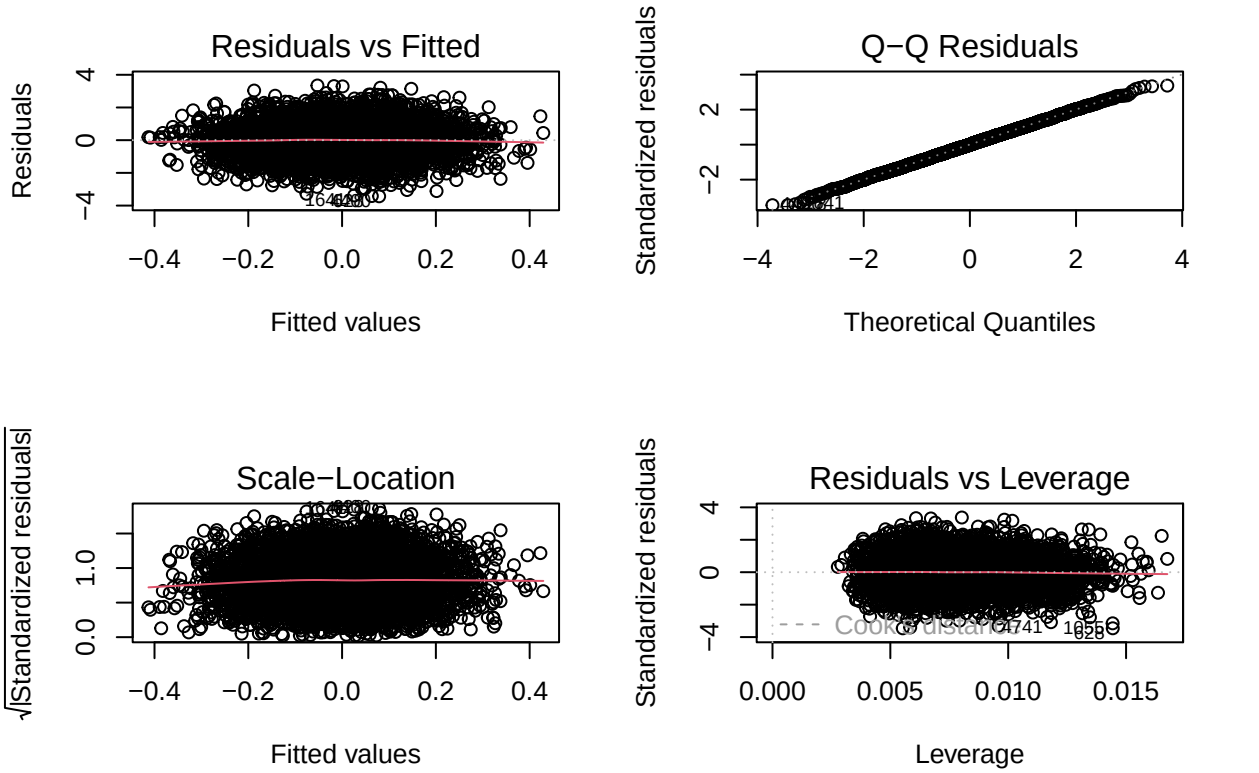
```r
df_model$bn_TotalSleep <- predict(bn_sleep)

model_sleep_bn <- lm(bn_TotalSleep ~ (Age + Gender + Exercise..mins.day. +
                                      Caffeine.Intake..mg. + Screen.Time.Before.Bed..mins. +
                                      Work.Hours..hrs.day. + Sleep.QualityCat + Productivity.ScoreCat +
                                      Mood.ScoreCat + Stress.LevelCat)^2, data = df_model)

model_sleep_reduced <- step(model_sleep_bn, direction = "both", trace = 0)
summary(model_sleep_reduced)
```

```
##
## Call:
## lm(formula = bn_TotalSleep ~ Age + Gender + Exercise..mins.day. +
##     Caffeine.Intake..mg. + Screen.Time.Before.Bed..mins. + Work.Hours..hrs.day. +
##     Sleep.QualityCat + Productivity.ScoreCat + Mood.ScoreCat +
##     Stress.LevelCat + Age:Sleep.QualityCat + Gender:Caffeine.Intake..mg. +
##     Exercise..mins.day.:Caffeine.Intake..mg. + Caffeine.Intake..mg.:Sleep.QualityCat +
##     Caffeine.Intake..mg.:Mood.ScoreCat + Caffeine.Intake..mg.:Stress.LevelCat +
##     Screen.Time.Before.Bed..mins.:Sleep.QualityCat + Work.Hours..hrs.day.:Mood.ScoreCat +
##     Sleep.QualityCat:Mood.ScoreCat + Productivity.ScoreCat:Mood.ScoreCat,
##     data = df_model)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4319 -0.6622 -0.0060  0.6754  3.3431
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                     4.403e-01  1.712e-01
## Age                                            -4.020e-04  2.103e-03
## GenderMale                                     -9.963e-02  6.829e-02
## GenderOther                                    -1.426e-01  6.888e-02
## Exercise..mins.day.                            -1.079e-03  1.091e-03
## Caffeine.Intake..mg.                           -1.760e-03  5.950e-04
## Screen.Time.Before.Bed..mins.                  -2.250e-04  4.962e-04
## Work.Hours..hrs.day.                           -2.702e-02  1.104e-02
## Sleep.QualityCatMedium                         -1.575e-01  1.471e-01
## Sleep.QualityCatHigh                            1.217e-01  1.577e-01
## Productivity.ScoreCatMedium                    -3.444e-02  6.284e-02
## Productivity.ScoreCatHigh                       1.338e-01  6.528e-02
## Mood.ScoreCatMedium                            -2.153e-01  1.559e-01
## Mood.ScoreCatHigh                              -2.056e-01  1.700e-01
## Stress.LevelCatMedium                          -1.313e-01  6.789e-02
## Stress.LevelCatHigh                             1.601e-02  7.266e-02
## Age:Sleep.QualityCatMedium                      2.024e-03  2.781e-03
## Age:Sleep.QualityCatHigh                       -4.596e-03  2.965e-03
```

```
## GenderMale:Caffeine.Intake..mg.                          6.811e-04  3.988e-04
## GenderOther:Caffeine.Intake..mg.                         8.302e-04  4.050e-04
## Exercise..mins.day.:Caffeine.Intake..mg.                 1.181e-05  6.474e-06
## Caffeine.Intake..mg.:Sleep.QualityCatMedium              1.293e-05  3.982e-04
## Caffeine.Intake..mg.:Sleep.QualityCatHigh                8.575e-04  4.256e-04
## Caffeine.Intake..mg.:Mood.ScoreCatMedium                 7.809e-04  3.950e-04
## Caffeine.Intake..mg.:Mood.ScoreCatHigh                   1.717e-04  4.277e-04
## Caffeine.Intake..mg.:Stress.LevelCatMedium               7.264e-04  4.004e-04
## Caffeine.Intake..mg.:Stress.LevelCatHigh                -8.950e-05  4.264e-04
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatMedium     8.663e-04  6.532e-04
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatHigh      -4.542e-04  7.051e-04
## Work.Hours..hrs.day.:Mood.ScoreCatMedium                 2.950e-02  1.474e-02
## Work.Hours..hrs.day.:Mood.ScoreCatHigh                   1.984e-02  1.614e-02
## Sleep.QualityCatMedium:Mood.ScoreCatMedium              -8.371e-02  8.115e-02
## Sleep.QualityCatHigh:Mood.ScoreCatMedium                -7.089e-02  8.744e-02
## Sleep.QualityCatMedium:Mood.ScoreCatHigh                 2.204e-01  8.883e-02
## Sleep.QualityCatHigh:Mood.ScoreCatHigh                   3.893e-02  9.534e-02
## Productivity.ScoreCatMedium:Mood.ScoreCatMedium          4.910e-02  8.262e-02
## Productivity.ScoreCatHigh:Mood.ScoreCatMedium           -1.320e-01  8.641e-02
## Productivity.ScoreCatMedium:Mood.ScoreCatHigh           -7.061e-02  9.131e-02
## Productivity.ScoreCatHigh:Mood.ScoreCatHigh             -2.176e-01  9.492e-02
##                                                         t value Pr(>|t|)
## (Intercept)                                               2.572  0.01013 *
## Age                                                      -0.191  0.84839
## GenderMale                                               -1.459  0.14466
## GenderOther                                              -2.070  0.03850 *
## Exercise..mins.day.                                      -0.989  0.32278
## Caffeine.Intake..mg.                                     -2.958  0.00312 **
## Screen.Time.Before.Bed..mins.                            -0.453  0.65027
## Work.Hours..hrs.day.                                     -2.447  0.01442 *
## Sleep.QualityCatMedium                                   -1.070  0.28453
## Sleep.QualityCatHigh                                      0.772  0.44033
## Productivity.ScoreCatMedium                              -0.548  0.58366
## Productivity.ScoreCatHigh                                 2.049  0.04048 *
## Mood.ScoreCatMedium                                      -1.380  0.16753
## Mood.ScoreCatHigh                                        -1.209  0.22655
## Stress.LevelCatMedium                                    -1.934  0.05318 .
## Stress.LevelCatHigh                                       0.220  0.82567
## Age:Sleep.QualityCatMedium                                0.728  0.46673
## Age:Sleep.QualityCatHigh                                 -1.550  0.12120
## GenderMale:Caffeine.Intake..mg.                           1.708  0.08769 .
## GenderOther:Caffeine.Intake..mg.                          2.050  0.04046 *
## Exercise..mins.day.:Caffeine.Intake..mg.                  1.825  0.06808 .
## Caffeine.Intake..mg.:Sleep.QualityCatMedium               0.032  0.97411
## Caffeine.Intake..mg.:Sleep.QualityCatHigh                 2.015  0.04397 *
## Caffeine.Intake..mg.:Mood.ScoreCatMedium                  1.977  0.04810 *
## Caffeine.Intake..mg.:Mood.ScoreCatHigh                    0.401  0.68811
## Caffeine.Intake..mg.:Stress.LevelCatMedium                1.814  0.06969 .
## Caffeine.Intake..mg.:Stress.LevelCatHigh                 -0.210  0.83376
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatMedium      1.326  0.18480
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatHigh       -0.644  0.51947
## Work.Hours..hrs.day.:Mood.ScoreCatMedium                  2.002  0.04538 *
## Work.Hours..hrs.day.:Mood.ScoreCatHigh                    1.230  0.21891
## Sleep.QualityCatMedium:Mood.ScoreCatMedium               -1.032  0.30229
```

```
## Sleep.QualityCatHigh:Mood.ScoreCatMedium               -0.811  0.41756
## Sleep.QualityCatMedium:Mood.ScoreCatHigh                2.481  0.01312 *
## Sleep.QualityCatHigh:Mood.ScoreCatHigh                  0.408  0.68300
## Productivity.ScoreCatMedium:Mood.ScoreCatMedium         0.594  0.55240
## Productivity.ScoreCatHigh:Mood.ScoreCatMedium          -1.528  0.12661
## Productivity.ScoreCatMedium:Mood.ScoreCatHigh          -0.773  0.43938
## Productivity.ScoreCatHigh:Mood.ScoreCatHigh            -2.292  0.02193 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9952 on 4961 degrees of freedom
## Multiple R-squared:  0.01554,    Adjusted R-squared:  0.007997
## F-statistic: 2.061 on 38 and 4961 DF,  p-value: 0.0001421
```

```
par(mfrow = c(2,2))
plot(model_sleep_reduced)
```




```
bn_work <- bestNormalize(df_model$Work.Hours..hrs.day.)
```

```
print(bn_work$chosen_transform)
```

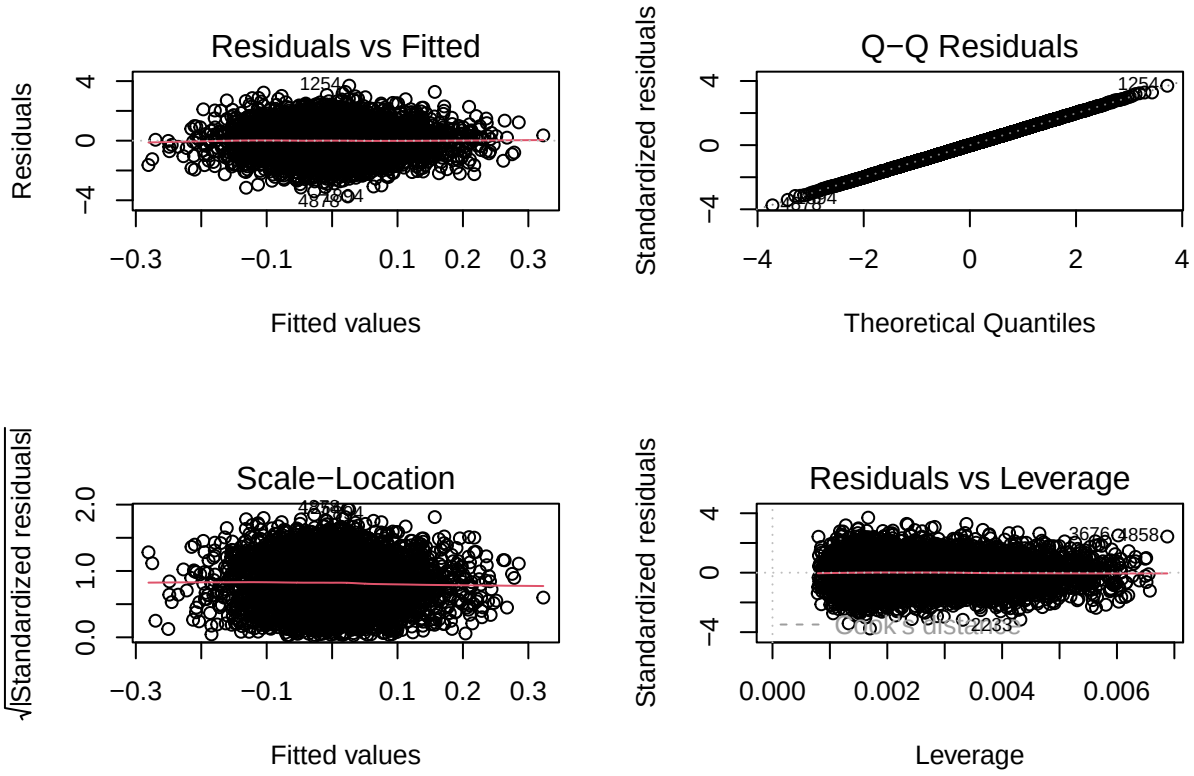
```
## orderNorm Transformation with 5000 nonmissing obs and no ties
##  - Original quantiles:
##     0%    25%    50%    75%   100%
##  4.000  6.033  7.998  9.905 11.999
```

```
df_model$bn_WorkHours <- predict(bn_work)

model_work_bn <- lm(bn_WorkHours ~ (Age + Gender + Exercise..mins.day. +
                                    Caffeine.Intake..mg. + Screen.Time.Before.Bed..mins. +
                                    Total.Sleep.Hours + Sleep.QualityCat + Productivity.ScoreCat +
```

```
                                    Mood.ScoreCat + Stress.LevelCat)^2,
                  data = df_model)

model_work_reduced <- step(model_work_bn, direction = "both", trace = 0)
summary(model_work_reduced)
```

```
##
## Call:
## lm(formula = bn_WorkHours ~ Age + Exercise..mins.day. + Total.Sleep.Hours +
##     Sleep.QualityCat + Mood.ScoreCat + Age:Exercise..mins.day. +
##     Age:Sleep.QualityCat + Age:Mood.ScoreCat, data = df_model)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7430 -0.6781 -0.0013  0.6624  3.6928
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -3.087e-01  1.517e-01  -2.035  0.04193 *
## Age                         1.102e-02  3.369e-03   3.271  0.00108 **
## Exercise..mins.day.         5.207e-03  1.813e-03   2.872  0.00410 **
## Total.Sleep.Hours          -1.655e-02  9.727e-03  -1.702  0.08887 .
## Sleep.QualityCatMedium      2.083e-01  1.128e-01   1.846  0.06494 .
## Sleep.QualityCatHigh        1.713e-01  1.200e-01   1.428  0.15348
## Mood.ScoreCatMedium        -7.182e-02  1.104e-01  -0.651  0.51531
## Mood.ScoreCatHigh           2.111e-01  1.209e-01   1.747  0.08077 .
## Age:Exercise..mins.day.    -1.344e-04  4.479e-05  -3.002  0.00270 **
## Age:Sleep.QualityCatMedium -6.141e-03  2.780e-03  -2.209  0.02720 *
## Age:Sleep.QualityCatHigh   -5.308e-03  2.964e-03  -1.791  0.07339 .
## Age:Mood.ScoreCatMedium     2.439e-03  2.730e-03   0.893  0.37171
## Age:Mood.ScoreCatHigh      -4.444e-03  2.981e-03  -1.491  0.13600
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9985 on 4987 degrees of freedom
## Multiple R-squared:  0.005231,   Adjusted R-squared:  0.002837
## F-statistic: 2.185 on 12 and 4987 DF,  p-value: 0.01012
```

```
par(mfrow = c(2,2))
plot(model_work_reduced)
```


Residuals vs Fitted
Residuals
Fitted values
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage