

```

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
##
##
## 
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

library(ggplot2)
library(reshape2)

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

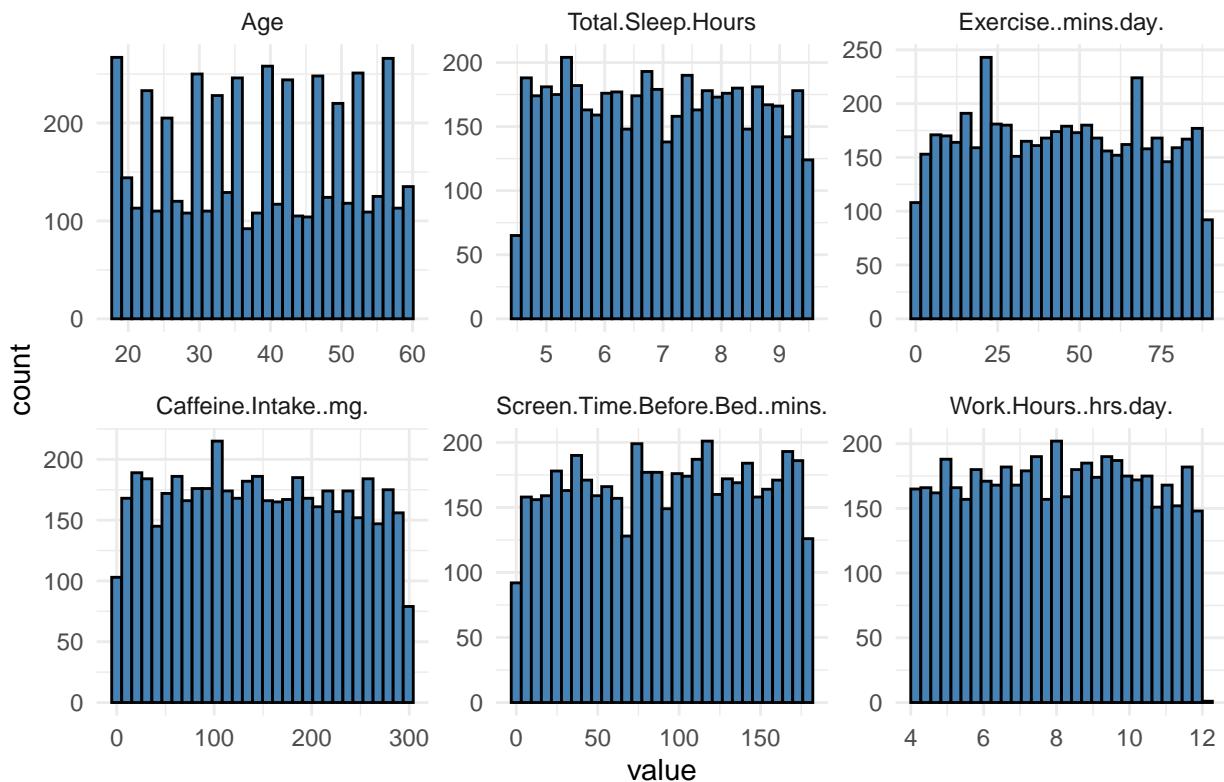
df_numeric <- df[, numeric_vars]

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")

```

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")

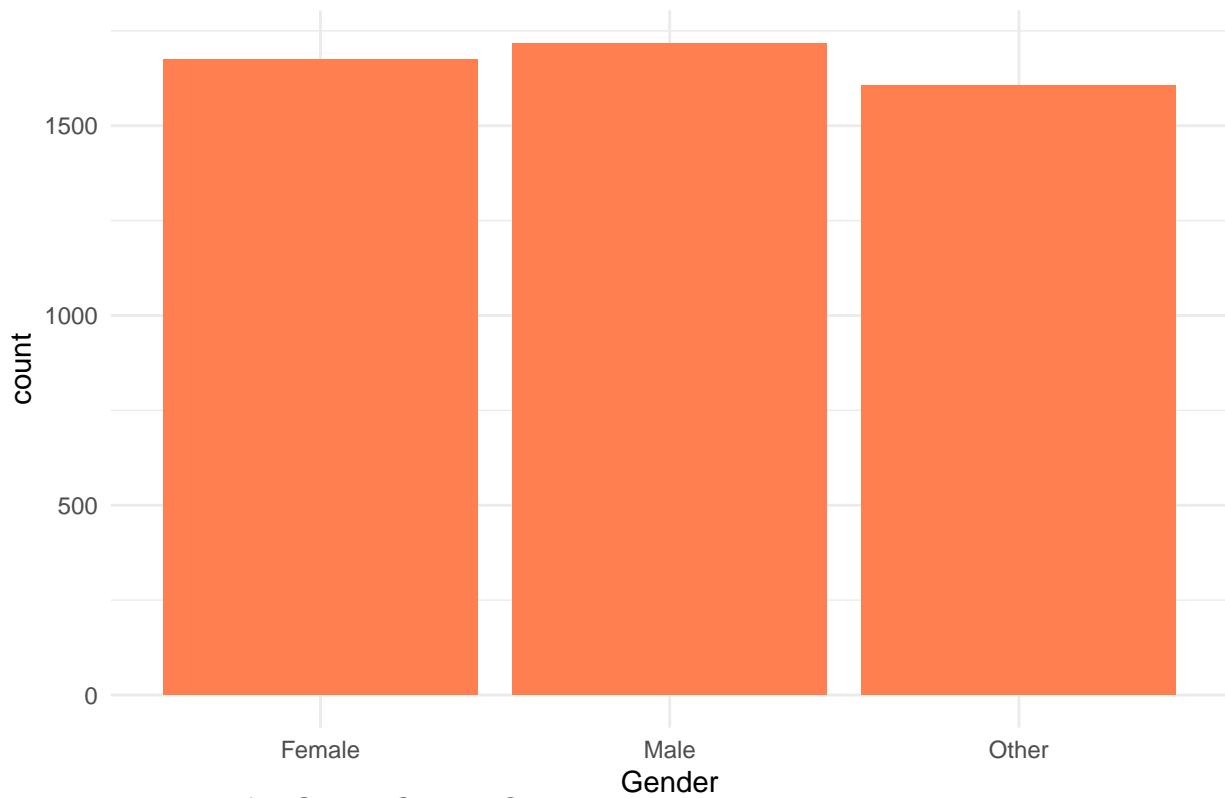
categorical_vars <- c("Gender", "Sleep.QualityCat", "Productivity.ScoreCat", "Mood.ScoreCat", "Stress.LifeScoreCat")

par(mfrow = c(2,4))

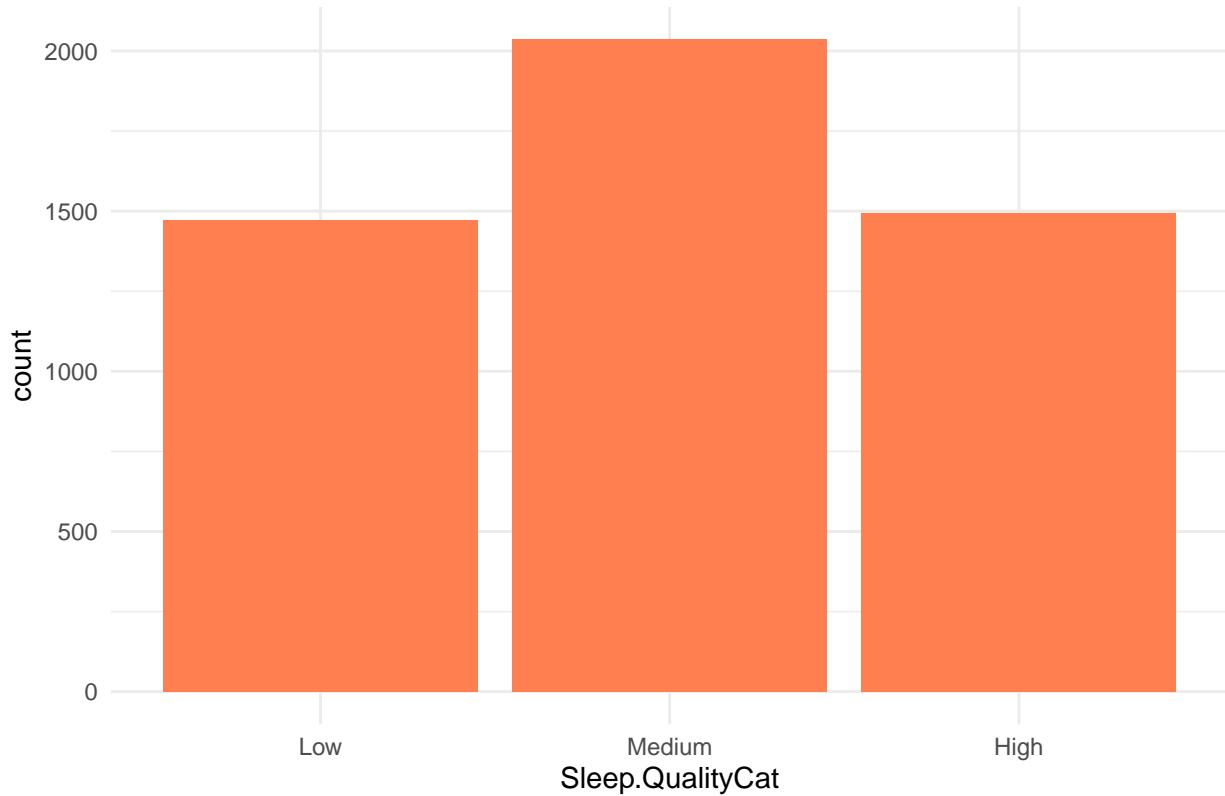
for (var in categorical_vars) {
  print(ggplot(df, aes_string(x = var)) +
    geom_bar(fill = "coral") +
    theme_minimal() +
    ggtitle(paste("Bar Plot for", var)))
}

## 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.
```

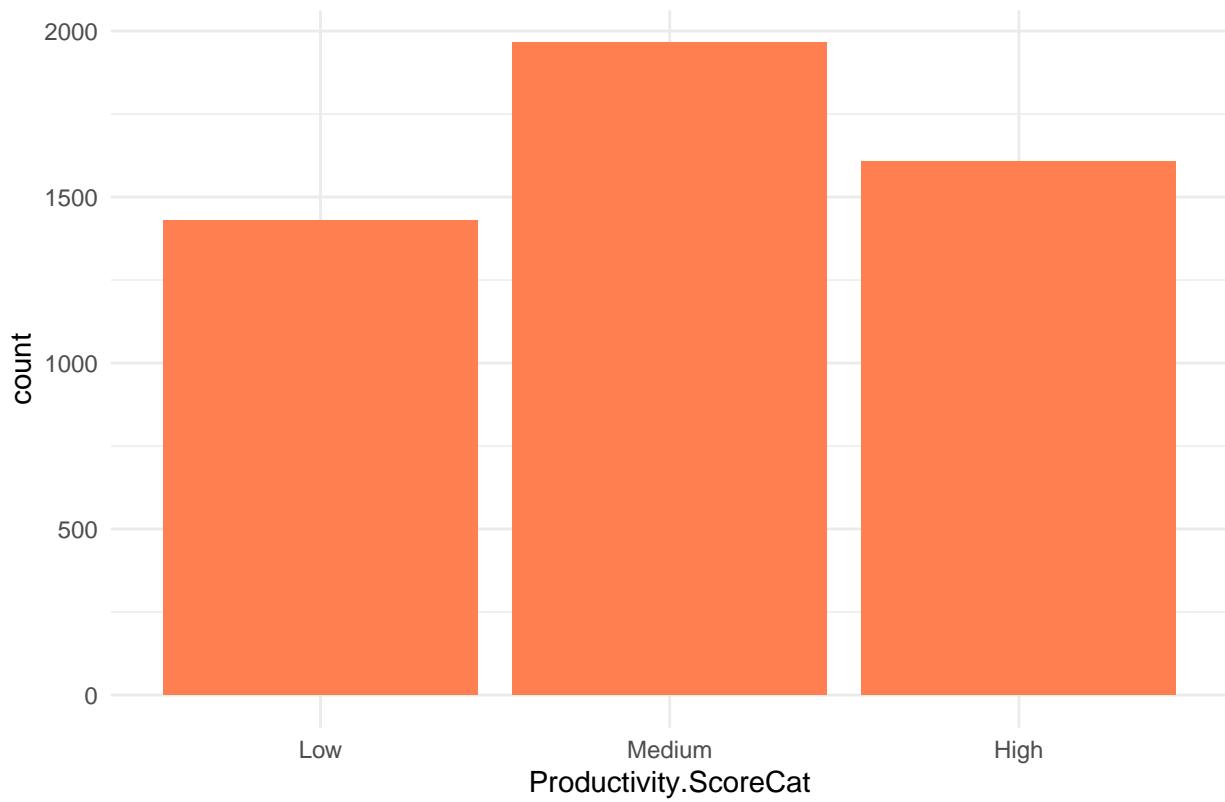
Bar Plot for Gender



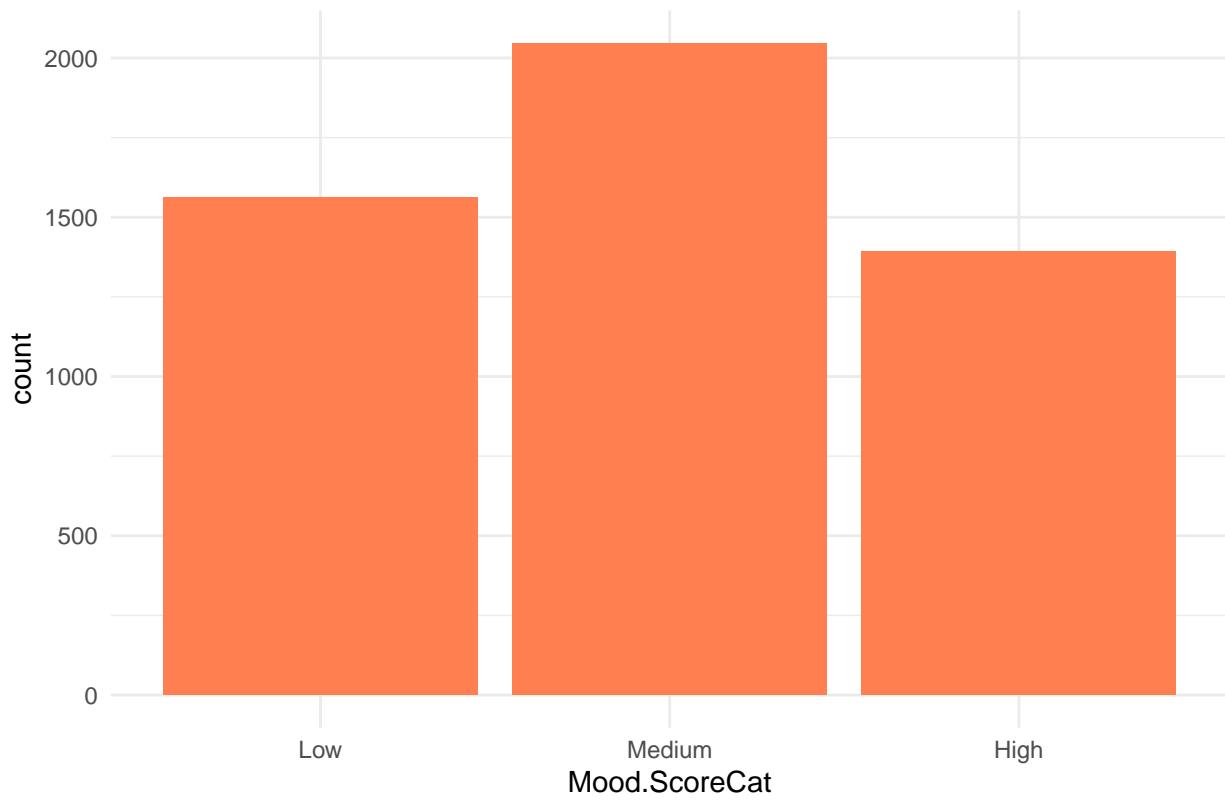
Bar Plot for Sleep.QualityCat



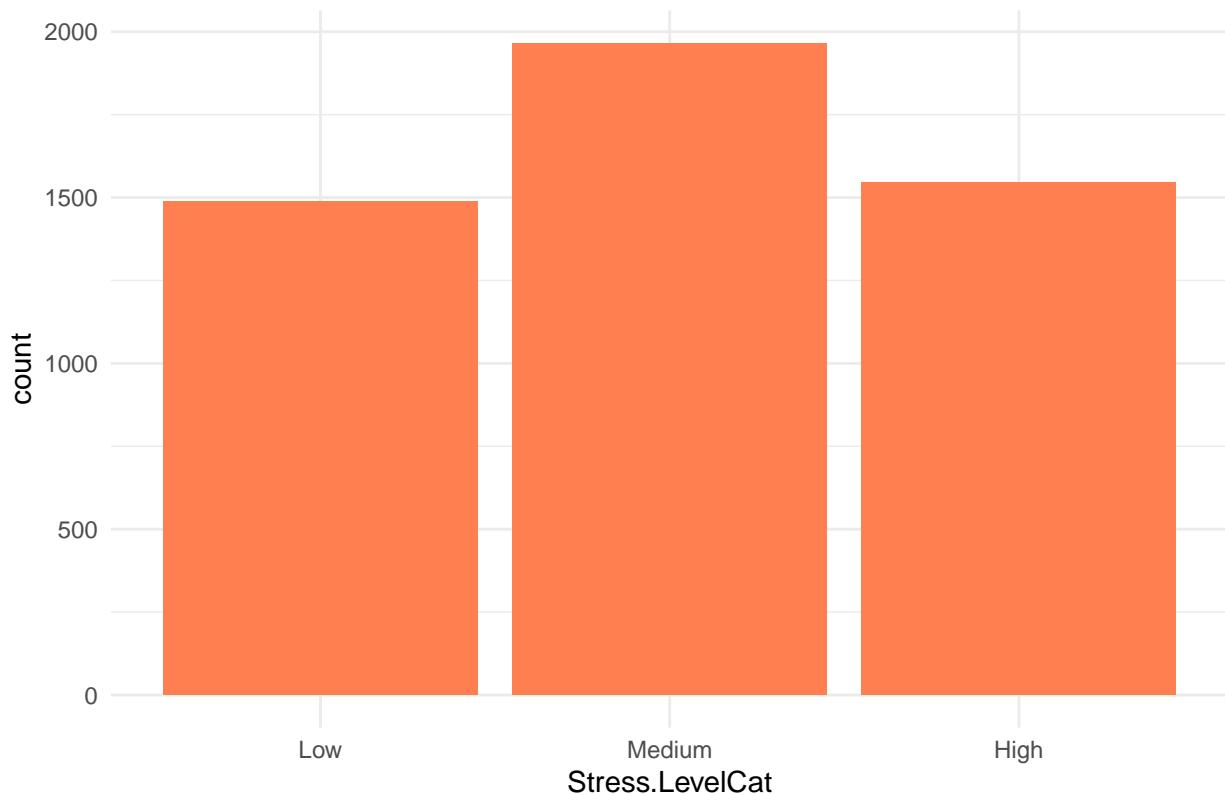
Bar Plot for Productivity.ScoreCat



Bar Plot for Mood.ScoreCat



Bar Plot for Stress.LevelCat

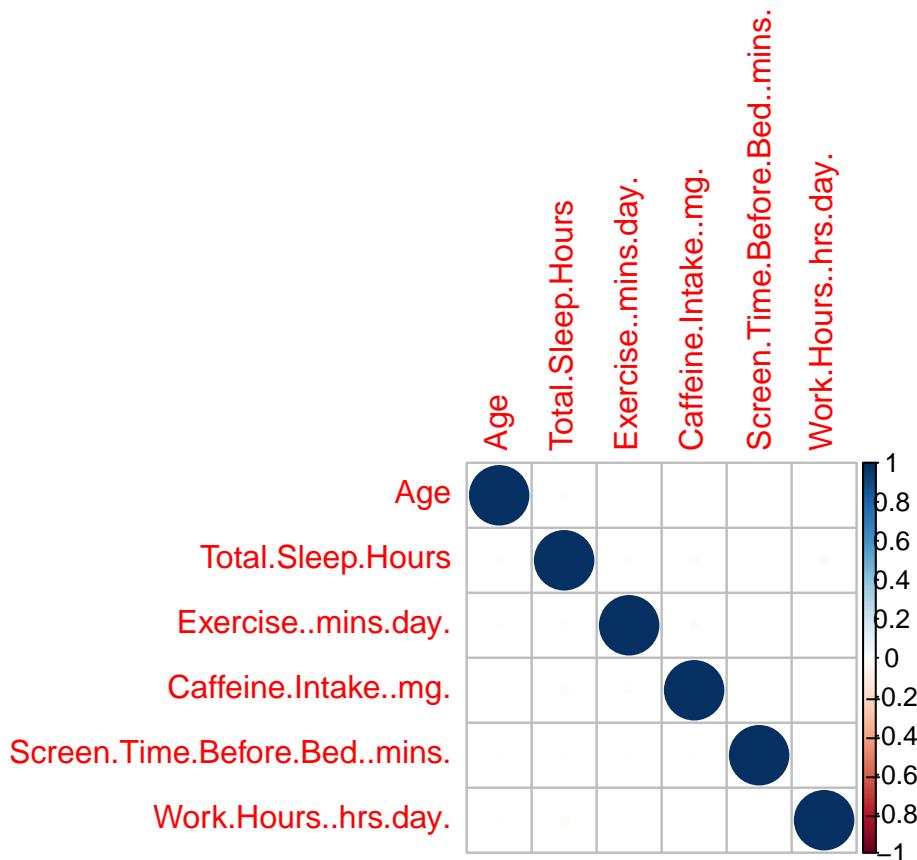


```
library(corrplot)

## 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.")

df_numeric <- df[, numeric_vars]

cor_matrix <- cor(df_numeric, use = "complete.obs")
corrplot(cor_matrix, method = "circle")
```



```

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.End.Time",
                                     "Sleep.Quality", "Productivity.Score", "Mood.Score", "Stress.Level"))]

df_model$Exercise..hours.day. <- df_model$Exercise..mins.day./60

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.

#print(bn_sleep$chosen_transform)
#paste("Transformation used: ", bn_sleep$chosen_transform)
df_model$bn_TotalSleep <- predict(bn_sleep)

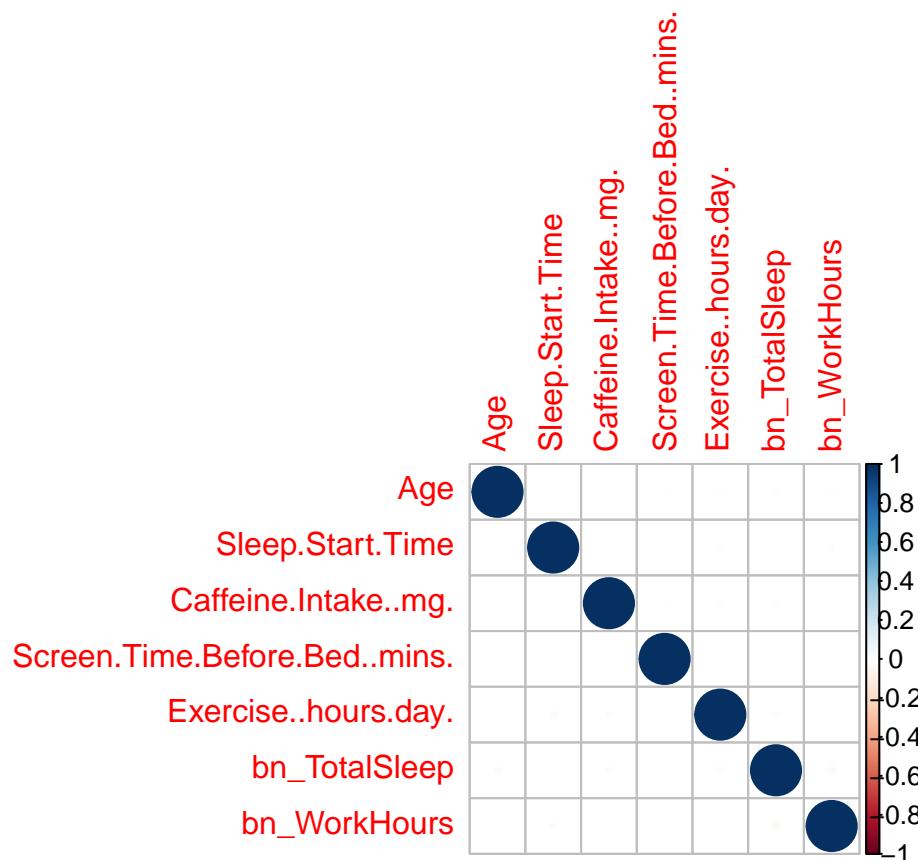
bn_work <- bestNormalize(df_model$Work.Hours..hrs.day.)

#paste("Transformation used: ", bn_work$chosen_transform)
df_model$bn_WorkHours <- predict(bn_work)

df_model_corr_plot <- df_model[, !names(df_model) %in% c("Total.Sleep.Hours", "Work.Hours..hrs.day.", "bn_TotalSleep", "bn_WorkHours")]

cor_matrix <- cor(df_model_corr_plot[, sapply(df_model_corr_plot, is.numeric)], use = "complete.obs")
corrplot(cor_matrix, method = "circle")

```



```

model_sleep_full <- lm(bn_TotalSleep ~ (Age + Gender + Exercise..hours.day. + Sleep.Start.Time +
                                         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_full, direction = "backward", trace = 0)
summary(model_sleep_reduced)

##
## Call:
## lm(formula = bn_TotalSleep ~ Age + Gender + Exercise..hours.day. +
## 
```

```

## Sleep.Start.Time + 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..hours.day.:Sleep.Start.Time +
## Exercise..hours.day.:Caffeine.Intake..mg. + Sleep.Start.Time:Work.Hours..hrs.day. +
## Sleep.Start.Time:Productivity.ScoreCat + Sleep.Start.Time:Mood.ScoreCat +
## 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
## -3.4710 -0.6678 -0.0094  0.6750  3.3694
##
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                3.767e-01  1.237e+00
## Age                     -5.288e-04  2.102e-03
## GenderMale               -1.003e-01  6.823e-02
## GenderOther               -1.384e-01  6.886e-02
## Exercise..hours.day.      1.358e+00  6.253e-01
## Sleep.Start.Time          3.441e-03  5.583e-02
## Caffeine.Intake..mg.      -1.783e-03  5.948e-04
## Screen.Time.Before.Bed..mins. -2.519e-04  4.960e-04
## Work.Hours..hrs.day.      -2.345e-01  1.185e-01
## Sleep.QualityCatMedium   -1.559e-01  1.470e-01
## Sleep.QualityCatHigh     1.109e-01  1.576e-01
## Productivity.ScoreCatMedium -7.118e-01  6.615e-01
## Productivity.ScoreCatHigh 8.758e-01  6.929e-01
## Mood.ScoreCatMedium       1.012e+00  6.533e-01
## Mood.ScoreCatHigh         -4.108e-02  7.104e-01
## Stress.LevelCatMedium    -1.323e-01  6.787e-02
## Stress.LevelCatHigh       1.967e-02  7.265e-02
## Age:Sleep.QualityCatMedium 1.926e-03  2.780e-03
## Age:Sleep.QualityCatHigh -4.467e-03  2.963e-03
## GenderMale:Caffeine.Intake..mg. 6.804e-04  3.985e-04
## GenderOther:Caffeine.Intake..mg. 8.196e-04  4.048e-04
## Exercise..hours.day.:Sleep.Start.Time -6.462e-02  2.828e-02
## Exercise..hours.day.:Caffeine.Intake..mg. 7.129e-04  3.881e-04
## Sleep.Start.Time:Work.Hours..hrs.day. 9.401e-03  5.360e-03
## Sleep.Start.Time:Productivity.ScoreCatMedium 3.060e-02  2.990e-02
## Sleep.Start.Time:Productivity.ScoreCatHigh -3.372e-02  3.137e-02
## Sleep.Start.Time:Mood.ScoreCatMedium -5.568e-02  2.894e-02
## Sleep.Start.Time:Mood.ScoreCatHigh -7.342e-03  3.170e-02
## Caffeine.Intake..mg.:Sleep.QualityCatMedium 1.352e-05  3.980e-04
## Caffeine.Intake..mg.:Sleep.QualityCatHigh 8.769e-04  4.253e-04
## Caffeine.Intake..mg.:Mood.ScoreCatMedium 7.914e-04  3.948e-04
## Caffeine.Intake..mg.:Mood.ScoreCatHigh 1.901e-04  4.274e-04
## Caffeine.Intake..mg.:Stress.LevelCatMedium 7.478e-04  4.001e-04
## Caffeine.Intake..mg.:Stress.LevelCatHigh -1.051e-04  4.263e-04
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatMedium 9.024e-04  6.527e-04
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatHigh -4.432e-04  7.053e-04
## Work.Hours..hrs.day.:Mood.ScoreCatMedium 2.968e-02  1.474e-02

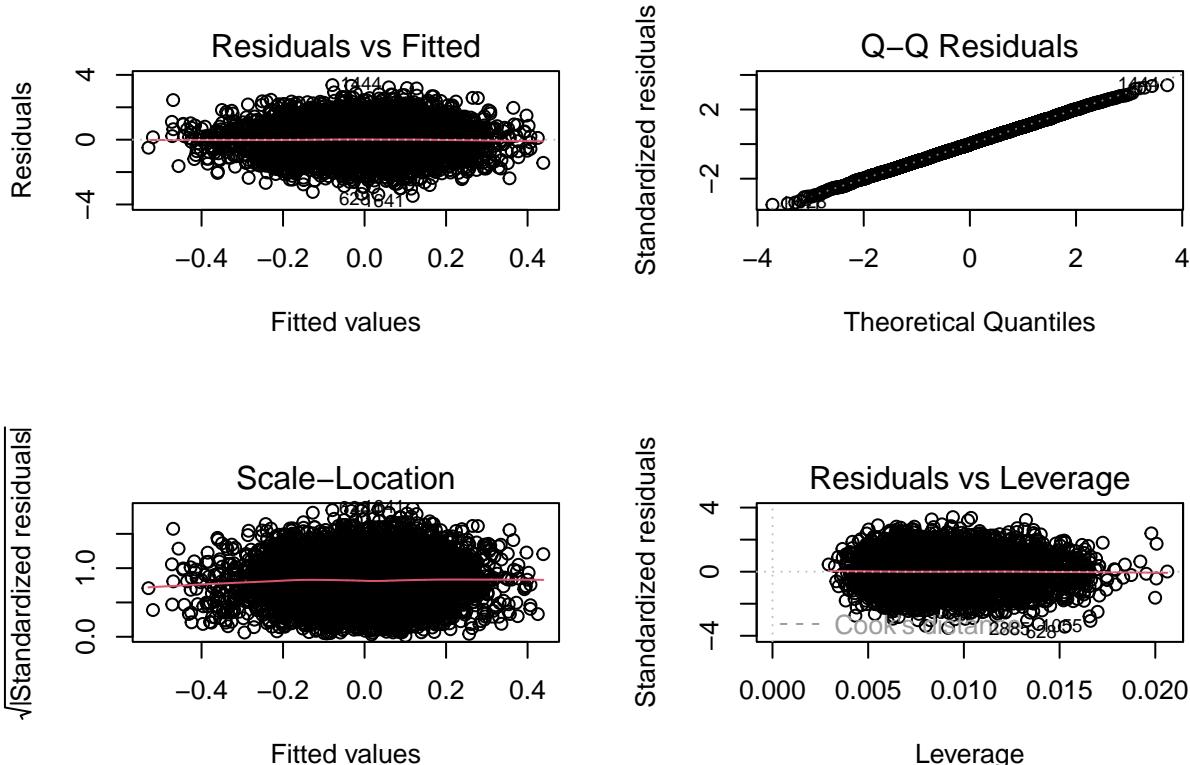
```

## Work.Hours..hrs.day.:Mood.ScoreCatHigh	1.960e-02	1.614e-02
## Sleep.QualityCatMedium:Mood.ScoreCatMedium	-8.769e-02	8.108e-02
## Sleep.QualityCatHigh:Mood.ScoreCatMedium	-6.845e-02	8.736e-02
## Sleep.QualityCatMedium:Mood.ScoreCatHigh	2.153e-01	8.876e-02
## Sleep.QualityCatHigh:Mood.ScoreCatHigh	3.858e-02	9.527e-02
## Productivity.ScoreCatMedium:Mood.ScoreCatMedium	5.110e-02	8.260e-02
## Productivity.ScoreCatHigh:Mood.ScoreCatMedium	-1.348e-01	8.634e-02
## Productivity.ScoreCatMedium:Mood.ScoreCatHigh	-7.018e-02	9.132e-02
## Productivity.ScoreCatHigh:Mood.ScoreCatHigh	-2.157e-01	9.489e-02
##		
## (Intercept)	0.304	0.76077
## Age	-0.252	0.80134
## GenderMale	-1.470	0.14164
## GenderOther	-2.010	0.04452 *
## Exercise..hours.day.	2.171	0.02998 *
## Sleep.Start.Time	0.062	0.95085
## Caffeine.Intake..mg.	-2.997	0.00274 **
## Screen.Time.Before.Bed..mins.	-0.508	0.61154
## Work.Hours..hrs.day.	-1.979	0.04787 *
## Sleep.QualityCatMedium	-1.060	0.28904
## Sleep.QualityCatHigh	0.704	0.48161
## Productivity.ScoreCatMedium	-1.076	0.28194
## Productivity.ScoreCatHigh	1.264	0.20633
## Mood.ScoreCatMedium	1.550	0.12125
## Mood.ScoreCatHigh	-0.058	0.95390
## Stress.LevelCatMedium	-1.950	0.05126 .
## Stress.LevelCatHigh	0.271	0.78662
## Age:Sleep.QualityCatMedium	0.693	0.48852
## Age:Sleep.QualityCatHigh	-1.507	0.13175
## GenderMale:Caffeine.Intake..mg.	1.707	0.08781 .
## GenderOther:Caffeine.Intake..mg.	2.025	0.04295 *
## Exercise..hours.day.:Sleep.Start.Time	-2.285	0.02236 *
## Exercise..hours.day.:Caffeine.Intake..mg.	1.837	0.06627 .
## Sleep.Start.Time:Work.Hours..hrs.day.	1.754	0.07949 .
## Sleep.Start.Time:Productivity.ScoreCatMedium	1.023	0.30614
## Sleep.Start.Time:Productivity.ScoreCatHigh	-1.075	0.28237
## Sleep.Start.Time:Mood.ScoreCatMedium	-1.924	0.05442 .
## Sleep.Start.Time:Mood.ScoreCatHigh	-0.232	0.81685
## Caffeine.Intake..mg.:Sleep.QualityCatMedium	0.034	0.97289
## Caffeine.Intake..mg.:Sleep.QualityCatHigh	2.062	0.03927 *
## Caffeine.Intake..mg.:Mood.ScoreCatMedium	2.005	0.04506 *
## Caffeine.Intake..mg.:Mood.ScoreCatHigh	0.445	0.65654
## Caffeine.Intake..mg.:Stress.LevelCatMedium	1.869	0.06166 .
## Caffeine.Intake..mg.:Stress.LevelCatHigh	-0.246	0.80531
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatMedium	1.383	0.16687
## Screen.Time.Before.Bed..mins.:Sleep.QualityCatHigh	-0.628	0.52977
## Work.Hours..hrs.day.:Mood.ScoreCatMedium	2.013	0.04415 *
## Work.Hours..hrs.day.:Mood.ScoreCatHigh	1.214	0.22463
## Sleep.QualityCatMedium:Mood.ScoreCatMedium	-1.082	0.27947
## Sleep.QualityCatHigh:Mood.ScoreCatMedium	-0.784	0.43334
## Sleep.QualityCatMedium:Mood.ScoreCatHigh	2.426	0.01532 *
## Sleep.QualityCatHigh:Mood.ScoreCatHigh	0.405	0.68557
## Productivity.ScoreCatMedium:Mood.ScoreCatMedium	0.619	0.53618
## Productivity.ScoreCatHigh:Mood.ScoreCatMedium	-1.561	0.11848

```

## Productivity.ScoreCatMedium:Mood.ScoreCatHigh      -0.769  0.44218
## Productivity.ScoreCatHigh:Mood.ScoreCatHigh      -2.273  0.02307 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9941 on 4954 degrees of freedom
## Multiple R-squared:  0.01911,   Adjusted R-squared:  0.0102
## F-statistic: 2.145 on 45 and 4954 DF,  p-value: 1.469e-05
par(mfrow = c(2,2))
plot(model_sleep_reduced)

```



```

model_work_full <- lm(bn_WorkHours ~ (Age + Gender + Exercise..hours.day. + Sleep.Start.Time +
                                         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_full, direction = "backward", trace = 0)
summary(model_work_reduced)

```

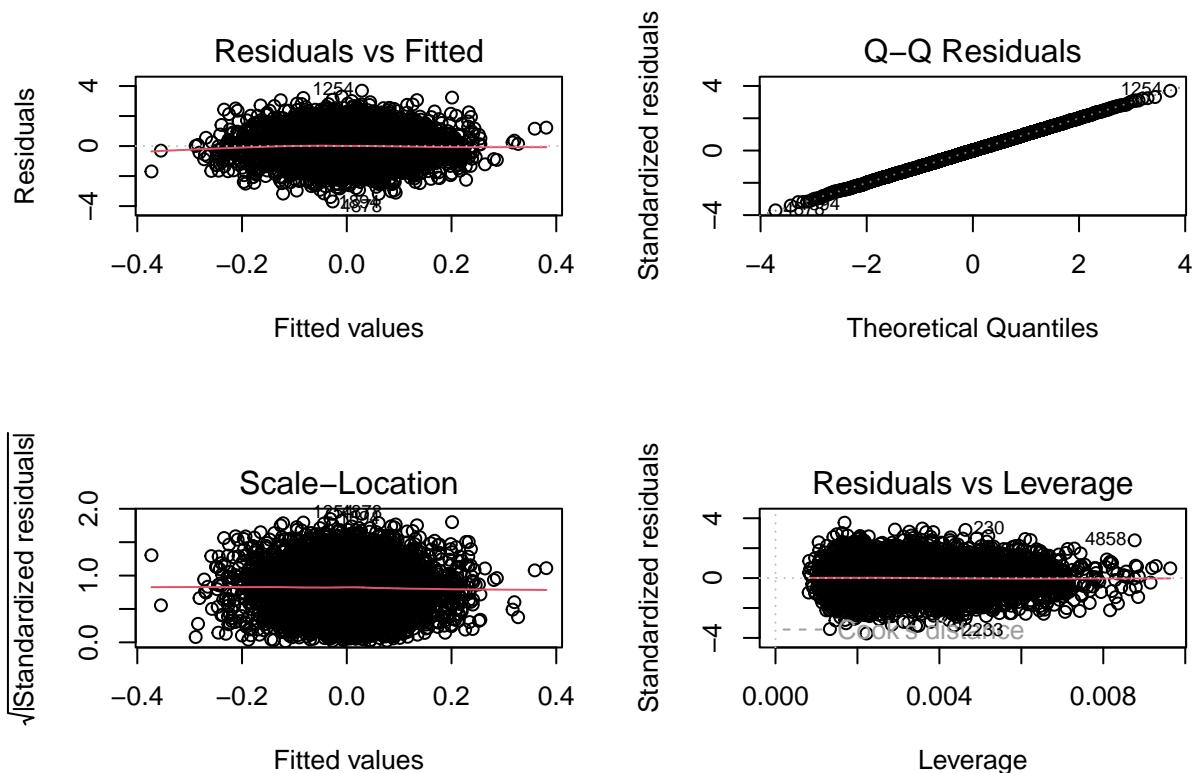
```

##
## Call:
## lm(formula = bn_WorkHours ~ Age + Exercise..hours.day. + Sleep.Start.Time +
##     Total.Sleep.Hours + Sleep.QualityCat + Mood.ScoreCat + Age:Exercise..hours.day. +
##     Age:Sleep.QualityCat + Age:Mood.ScoreCat + Sleep.Start.Time:Total.Sleep.Hours +
##     Sleep.Start.Time:Mood.ScoreCat, data = df_model)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -1.000000 -0.250000  0.000000  0.250000  1.000000
## 
```

```

## -3.6920 -0.6788 -0.0006  0.6669  3.6901
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.434681  1.386584  1.035  0.30086
## Age                      0.010990  0.003368  3.263  0.00111 **
## Exercise..hours.day.      0.309049  0.108895  2.838  0.00456 **
## Sleep.Start.Time          -0.079081  0.062677 -1.262  0.20711
## Total.Sleep.Hours         -0.400419  0.185634 -2.157  0.03105 *
## Sleep.QualityCatMedium   0.207793  0.112783  1.842  0.06547 .
## Sleep.QualityCatHigh     0.168340  0.120012  1.403  0.16077
## Mood.ScoreCatMedium       1.417800  0.648012  2.188  0.02872 *
## Mood.ScoreCatHigh         0.743275  0.709638  1.047  0.29497
## Age:Exercise..hours.day. -0.007912  0.002690 -2.942  0.00328 **
## Age:Sleep.QualityCatMedium -0.006142  0.002779 -2.210  0.02714 *
## Age:Sleep.QualityCatHigh -0.005234  0.002964 -1.766  0.07746 .
## Age:Mood.ScoreCatMedium  0.002460  0.002729  0.902  0.36736
## Age:Mood.ScoreCatHigh    -0.004548  0.002979 -1.527  0.12692
## Sleep.Start.Time:Total.Sleep.Hours 0.017409  0.008421  2.067  0.03875 *
## Sleep.Start.Time:Mood.ScoreCatMedium -0.067599  0.028971 -2.333  0.01967 *
## Sleep.Start.Time:Mood.ScoreCatHigh   -0.023934  0.031732 -0.754  0.45074
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9979 on 4983 degrees of freedom
## Multiple R-squared:  0.00728,   Adjusted R-squared:  0.004093
## F-statistic: 2.284 on 16 and 4983 DF,  p-value: 0.002495
par(mfrow = c(2,2))
plot(model_work_reduced)

```



```

remove_terms <- function(model, terms, threshold = 0.05) {
  current_model <- model
  for (term in terms) {
    current_terms <- attr(terms(current_model), "term.labels")
    if (!(term %in% current_terms)) {
      next
    }
    updated_model <- update(current_model, as.formula(paste(". ~ . -", term)))

    anova_result <- anova(updated_model, current_model)
    p_val <- anova_result[["Pr(>F)"]][2]

    if (is.na(p_val)) {
      next
    }

    # Since we are removing terms, we want a high p-value, which suggests removing the feature does not
    # Significantly effect the model.
    if (p_val > threshold) {
      cat("Removing", term, "\n")
      current_model <- updated_model

      # Here if a base feature is removed, interactions containing that feature, will be removed
      if (!grepl(":", term)) {
        current_terms <- attr(terms(current_model), "term.labels")
        interactions_to_remove <- current_terms[grepl(term, current_terms) & grepl(":", current_terms)]
        if (length(interactions_to_remove) > 0) {
          removal_formula <- paste(interactions_to_remove, collapse = " - ")
          current_model <- update(current_model, as.formula(paste(". ~ . -", removal_formula)))
        }
      }
    }
  }
  return(current_model)
}

terms_to_remove <- c("Sleep.Start.Time",
                     "Mood.ScoreCatHigh",
                     "Screen.Time.Before.Bed..mins.",
                     "Age:Sleep.QualityCatMedium",
                     "Sleep.QualityCat:Mood.ScoreCat",
                     "Productivity.ScoreCat:Mood.ScoreCat",
                     "Sleep.QualityCat",
                     "Stress.LevelCat",
                     "Gender",
                     "Age",
                     "Exercise..hours.day.",
                     "Screen.Time.Before.Bed..mins.",
                     "Work.Hours..hrs.day.",
                     "Stress.LevelCat",
                     "Work.Hours..hrs.day.:Mood.ScoreCat",
                     "Productivity.ScoreCat",
                     "Caffeine.Intake..mg.:Stress.LevelCat",

```

```

    "Caffeine.Intake..mg.:Mood.ScoreCat",
    "Stress.LevelCat",
    "Work.Hours..hrs.day.",
    "Caffeine.Intake..mg.",
    "Mood.ScoreCat",
    "Stress.LevelCat")

model_sleep_final <- remove_terms(model_sleep_reduced, terms_to_remove, threshold = 0.05)

## Removing Sleep.Start.Time
## Removing Productivity.ScoreCat:Mood.ScoreCat
## Removing Sleep.QualityCat
## Removing Gender
## Removing Age
## Removing Exercise..hours.day.
## Removing Screen.Time.Before.Bed..mins.
## Removing Work.Hours..hrs.day.:Mood.ScoreCat
## Removing Productivity.ScoreCat
## Removing Caffeine.Intake..mg.:Mood.ScoreCat
## Removing Work.Hours..hrs.day.

summary(model_sleep_final)

##
## Call:
## lm(formula = bn_TotalSleep ~ Caffeine.Intake..mg. + Mood.ScoreCat +
##     Stress.LevelCat + Caffeine.Intake..mg.:Stress.LevelCat, data = df_model)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4227 -0.6716 -0.0049  0.6689  3.3572
##
## Coefficients:
##                               Estimate Std. Error t value
## (Intercept)                1.021e-02  5.596e-02  0.182
## Caffeine.Intake..mg.        -1.287e-04 3.053e-04 -0.422
## Mood.ScoreCatMedium         5.762e-02 3.356e-02  1.717
## Mood.ScoreCatHigh          -2.275e-02 3.681e-02 -0.618
## Stress.LevelCatMedium      -1.406e-01 6.790e-02 -2.071
## Stress.LevelCatHigh         1.915e-02 7.258e-02  0.264
## Caffeine.Intake..mg.:Stress.LevelCatMedium 7.850e-04 4.005e-04  1.960
## Caffeine.Intake..mg.:Stress.LevelCatHigh   -9.016e-05 4.261e-04 -0.212
## Pr(>|t|)
## (Intercept)                  0.8552
## Caffeine.Intake..mg.          0.6734
## Mood.ScoreCatMedium           0.0861 .
## Mood.ScoreCatHigh              0.5365
## Stress.LevelCatMedium          0.0385 *
## Stress.LevelCatHigh             0.7920
## Caffeine.Intake..mg.:Stress.LevelCatMedium 0.0500 .
## Caffeine.Intake..mg.:Stress.LevelCatHigh    0.8325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```

## Residual standard error: 0.9985 on 4992 degrees of freedom
## Multiple R-squared:  0.002835,   Adjusted R-squared:  0.001436
## F-statistic: 2.027 on 7 and 4992 DF,  p-value: 0.04811

terms_to_remove <- c("Sleep.Start.Time",
                      "Mood.ScoreCatHigh",
                      "Screen.Time.Before.Bed..mins.",
                      "Age:Sleep.QualityCatMedium",
                      "Mood.ScoreCat",
                      "Age:Mood.ScoreCat",
                      "Sleep.QualityCat:Mood.ScoreCat",
                      "Mood.ScoreCat",
                      "Sleep.QualityCat",
                      "Total.Sleep.Hours")

model_work_final <- remove_terms(model_work_reduced, terms_to_remove, threshold = 0.05)

## Removing Sleep.Start.Time
## Removing Age:Mood.ScoreCat
## Removing Mood.ScoreCat
## Removing Sleep.QualityCat
## Removing Total.Sleep.Hours

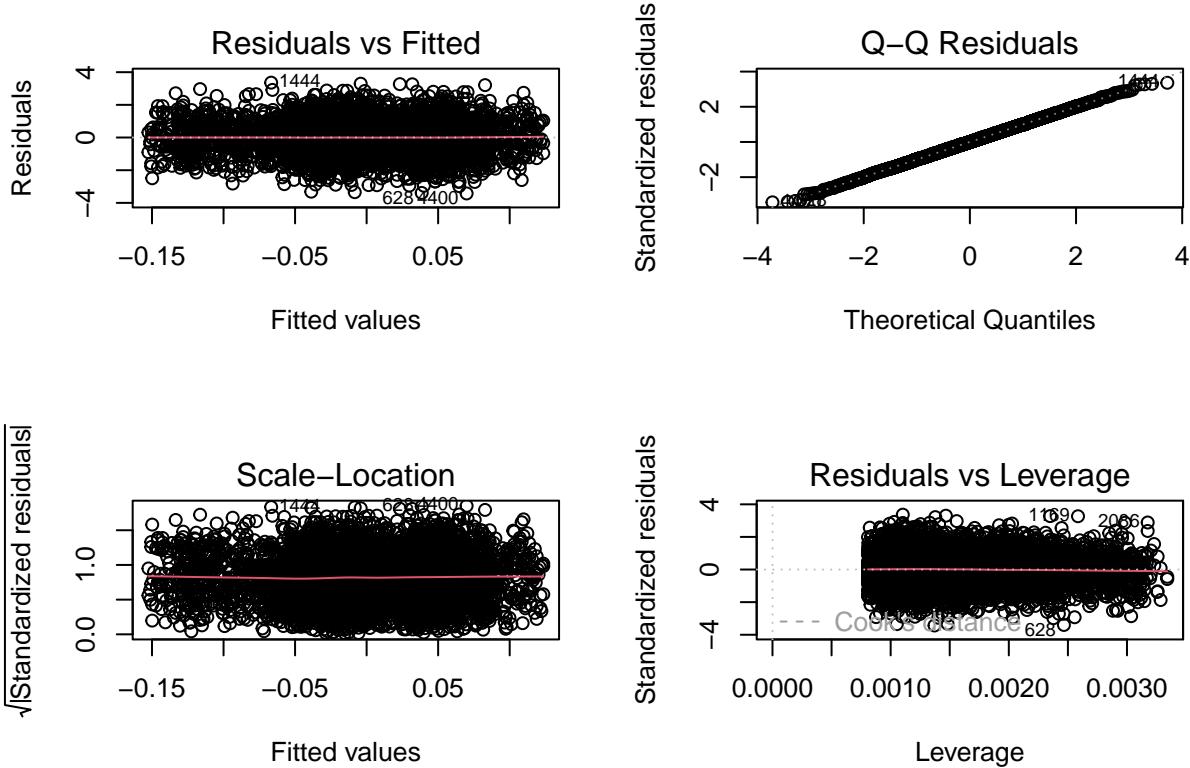
summary(model_work_final)

##
## Call:
## lm(formula = bn_WorkHours ~ Age + Exercise..hours.day. + Age:Exercise..hours.day.,
##      data = df_model)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -3.7293 -0.6788 -0.0008  0.6740  3.7088 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.260300  0.092568 -2.812  0.00494 ***
## Age          0.006784  0.002293  2.959  0.00310 ***
## Exercise..hours.day. 0.312016  0.108826  2.867  0.00416 ***
## Age:Exercise..hours.day. -0.008124  0.002688 -3.022  0.00252 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9993 on 4996 degrees of freedom
## Multiple R-squared:  0.001919,   Adjusted R-squared:  0.00132 
## F-statistic: 3.202 on 3 and 4996 DF,  p-value: 0.02231

library(car)

## Loading required package: carData
par(mfrow = c(2,2))
plot(model_sleep_final)

```



```

print(shapiro.test(residuals(model_sleep_final)))

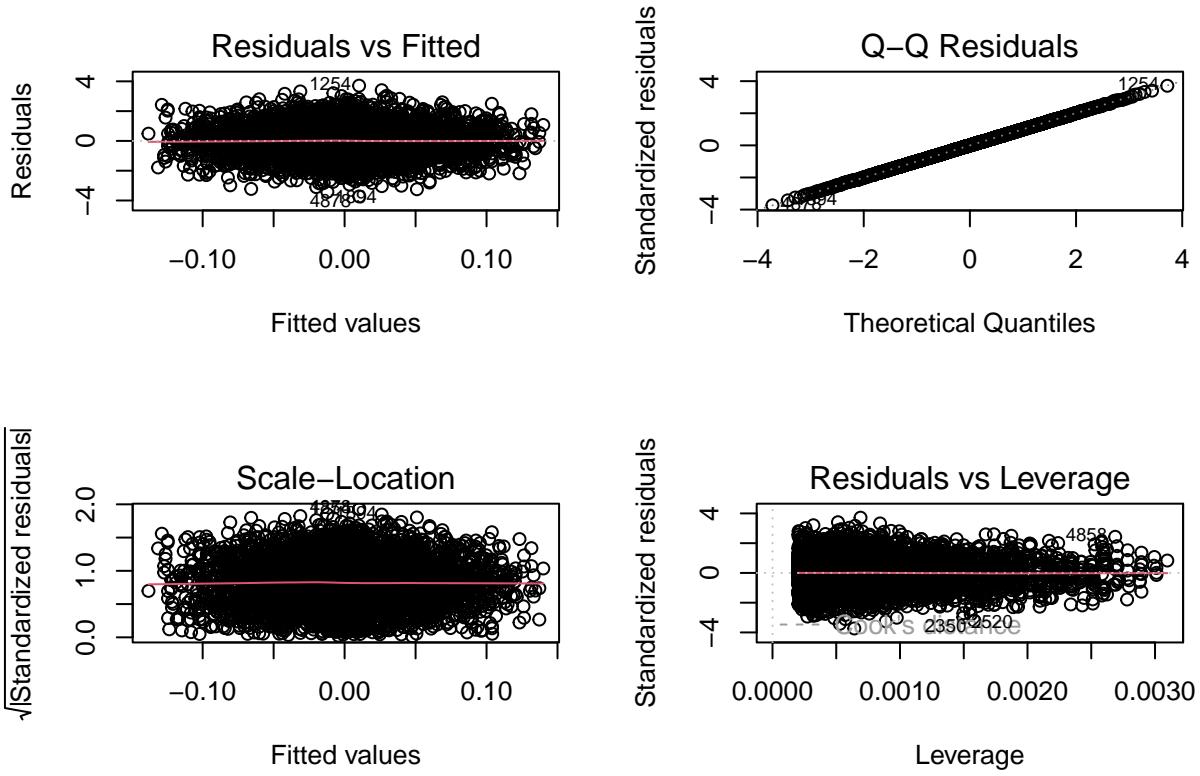
##
## Shapiro-Wilk normality test
##
## data: residuals(model_sleep_final)
## W = 0.99983, p-value = 0.9806
print(ncvTest(model_sleep_final))

##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.1385748, Df = 1, p = 0.7097
print(vif(model_sleep_final))

##
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif

##
##                                     GVIF Df GVIF^(1/(2*Df))
## Caffeine.Intake..mg.           3.442715  1    1.855456
## Mood.ScoreCat                  1.000935  2    1.000234
## Stress.LevelCat                15.550300  2    1.985796
## Caffeine.Intake..mg.:Stress.LevelCat 25.068956  2    2.237608
par(mfrow = c(2,2))
plot(model_work_final)

```



```

print(shapiro.test(residuals(model_work_final)))

##
## Shapiro-Wilk normality test
##
## data: residuals(model_work_final)
## W = 0.99996, p-value = 1
print(ncvTest(model_work_final))

##
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 2.086715, Df = 1, p = 0.14859
print(vif(model_work_final))

##
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
##
##          Age      Exercise..hours.day. Age:Exercise..hours.day.
##          3.983189     10.960762        14.047608

```