

STAT463 Project: Sleep Health and Lifestyle

Lisa Lu 31088272

```
library(dplyr)
library(ggplot2)
library(stringr)
```

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

```
# Split Blood Pressure column into systolic and diastolic pressure
```

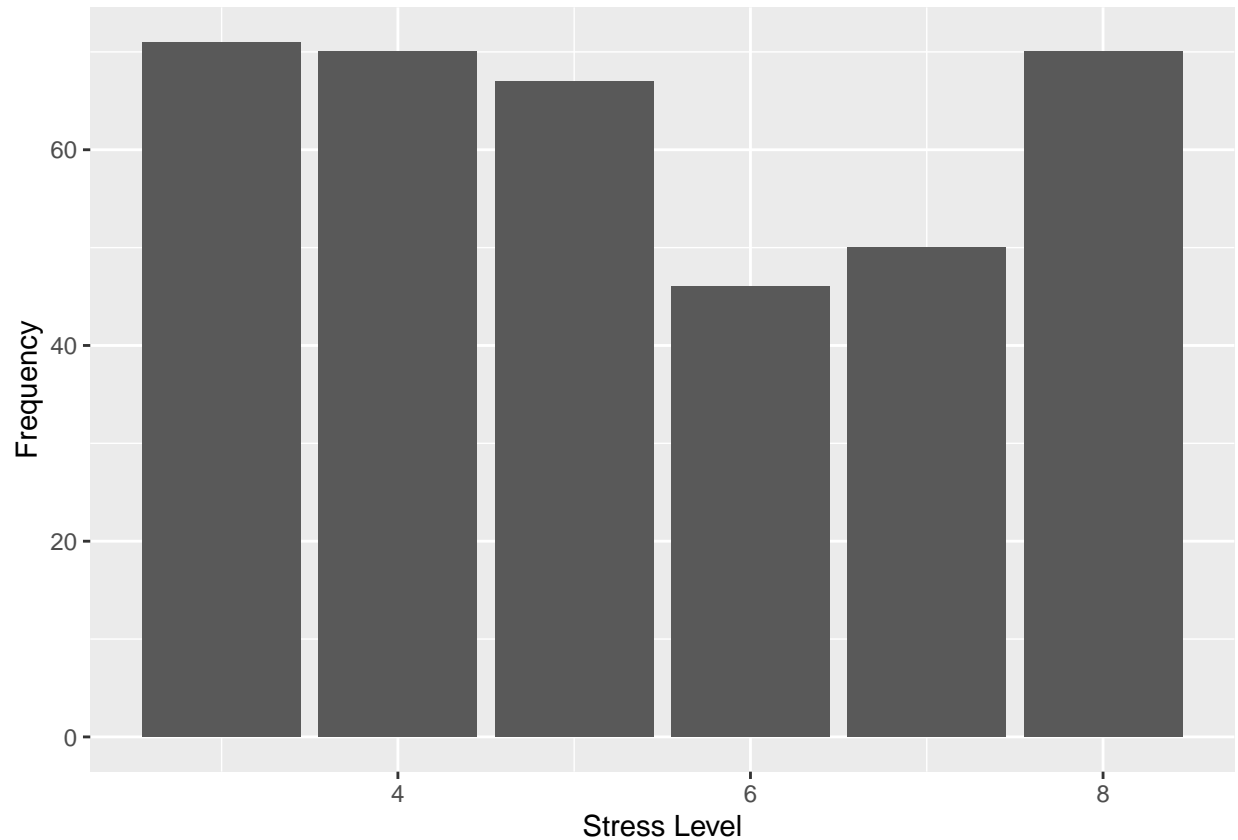
```
dataset[c('Systolic.Pressure', 'Diastolic.Pressure')] <- as.numeric(str_split_fixed(dataset$Blood.Pressure, 10, 2))
```

```
sleep <- dataset[, -10]
```

```
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   Systolic.Pressure  Diastolic.Pressure
## Length:374      Min.   :115.0    Min.   :75.00
## Class :character 1st Qu.:125.0    1st Qu.:80.00
## Mode  :character Median :130.0     Median :85.00
##                      Mean   :128.6     Mean   :84.65
##                      3rd Qu.:135.0     3rd Qu.:90.00
##                      Max.   :142.0     Max.   :95.00
```

```
# Frequency chart
dataset %>%
  ggplot(aes(x = Stress.Level)) +
  geom_bar() +
  labs(y = "Frequency", x = "Stress Level")
```



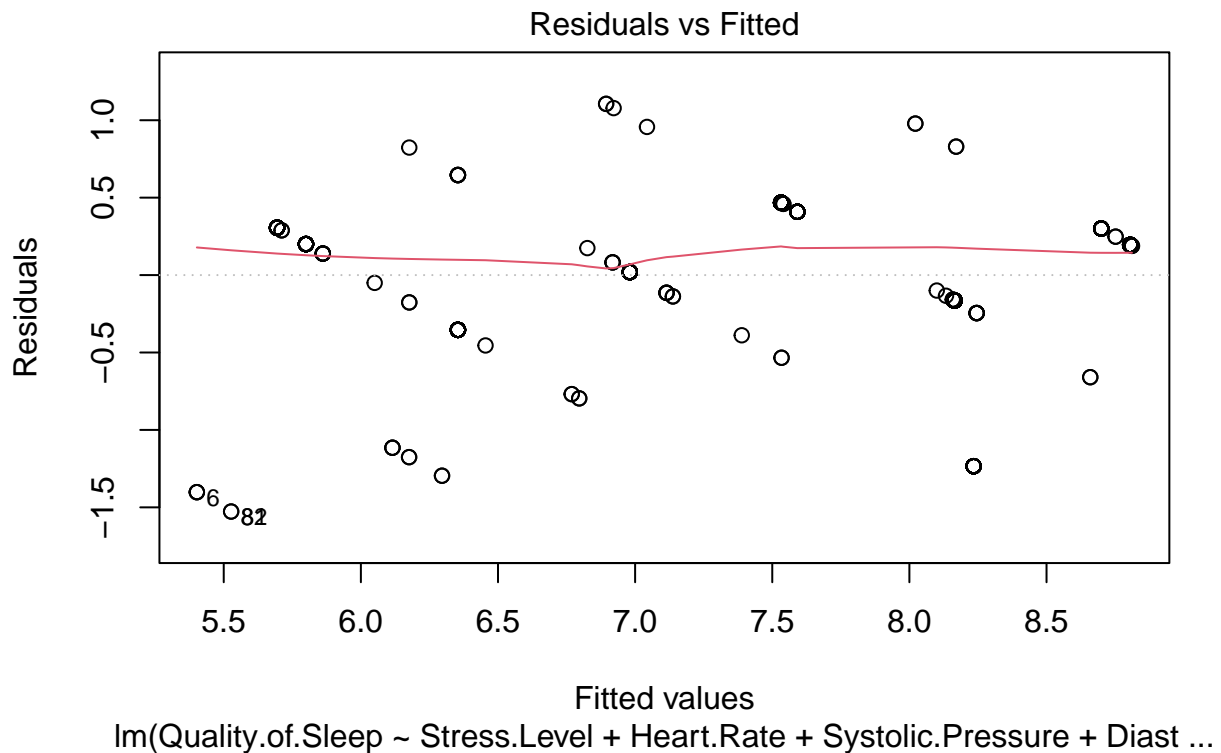
Explore the factors affecting quality of sleep

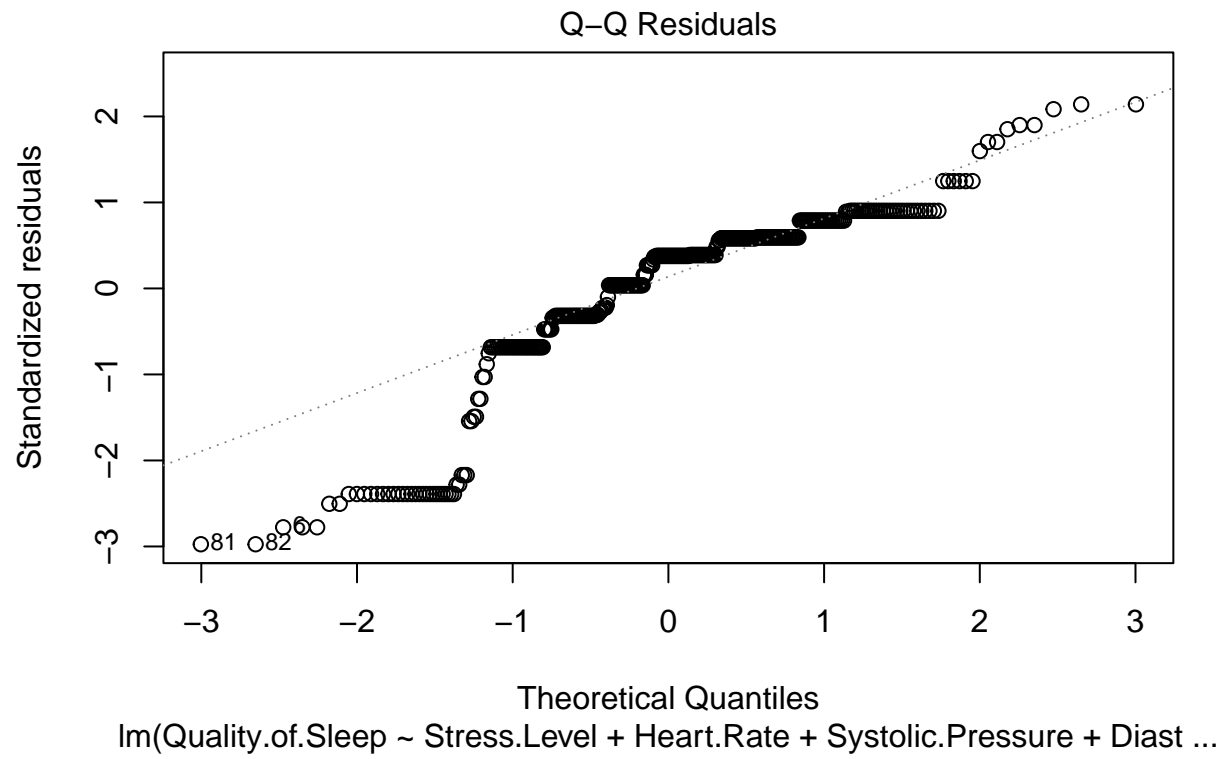
```
# Use linear regression model
lr <- lm(Quality.of.Sleep ~ Stress.Level + Heart.Rate + Systolic.Pressure + Diastolic.Pressure, data = sleep)
summary(lr)
```

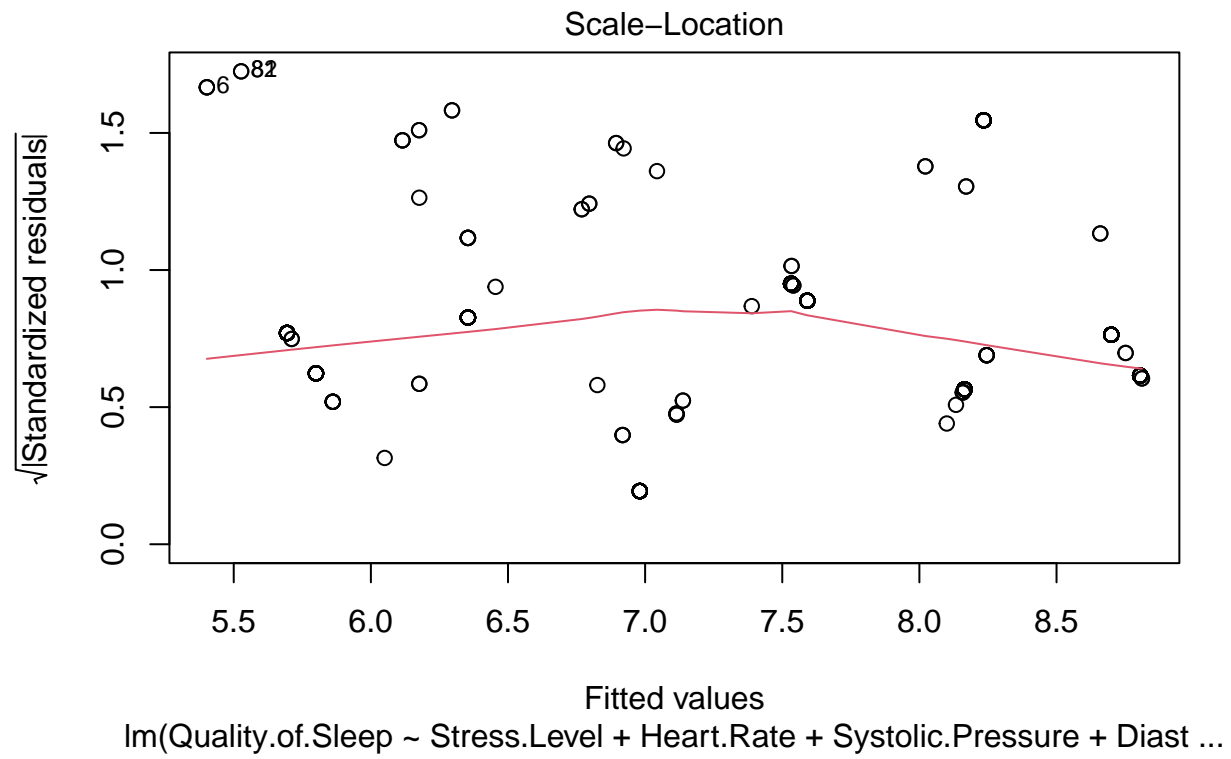
```
##
## Call:
## lm(formula = Quality.of.Sleep ~ Stress.Level + Heart.Rate + Systolic.Pressure +
##     Diastolic.Pressure, data = sleep)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5270 -0.1646  0.1950  0.3056  1.1057
##
## Coefficients:
```

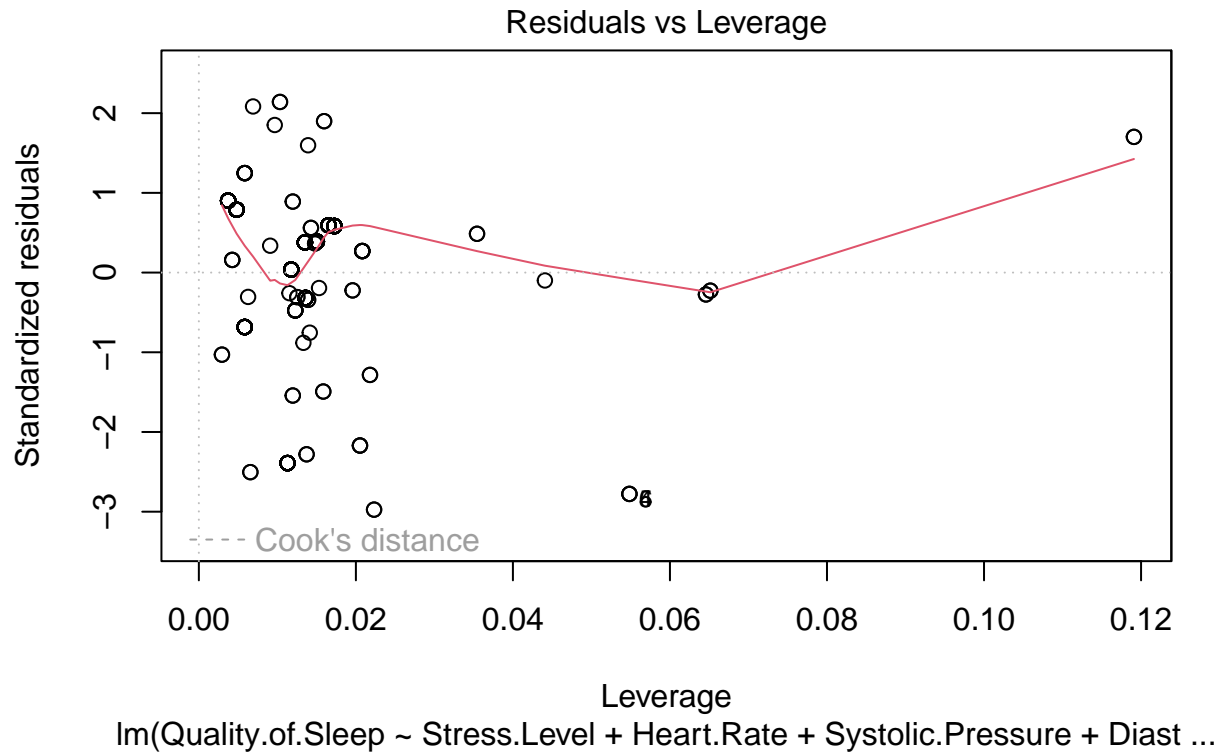
```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    12.5162267   0.6688836  18.712 < 2e-16 ***
## Stress.Level   -0.5596411   0.0205943 -27.175 < 2e-16 ***
## Heart.Rate     -0.0295663   0.0092121  -3.209 0.00145 **
## Systolic.Pressure -0.0004741 0.0151603  -0.031 0.97507
## Diastolic.Pressure -0.0006404 0.0189115  -0.034 0.97300
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5192 on 369 degrees of freedom
## Multiple R-squared:  0.8138, Adjusted R-squared:  0.8118
## F-statistic: 403.3 on 4 and 369 DF,  p-value: < 2.2e-16
```

```
plot(lr)
```



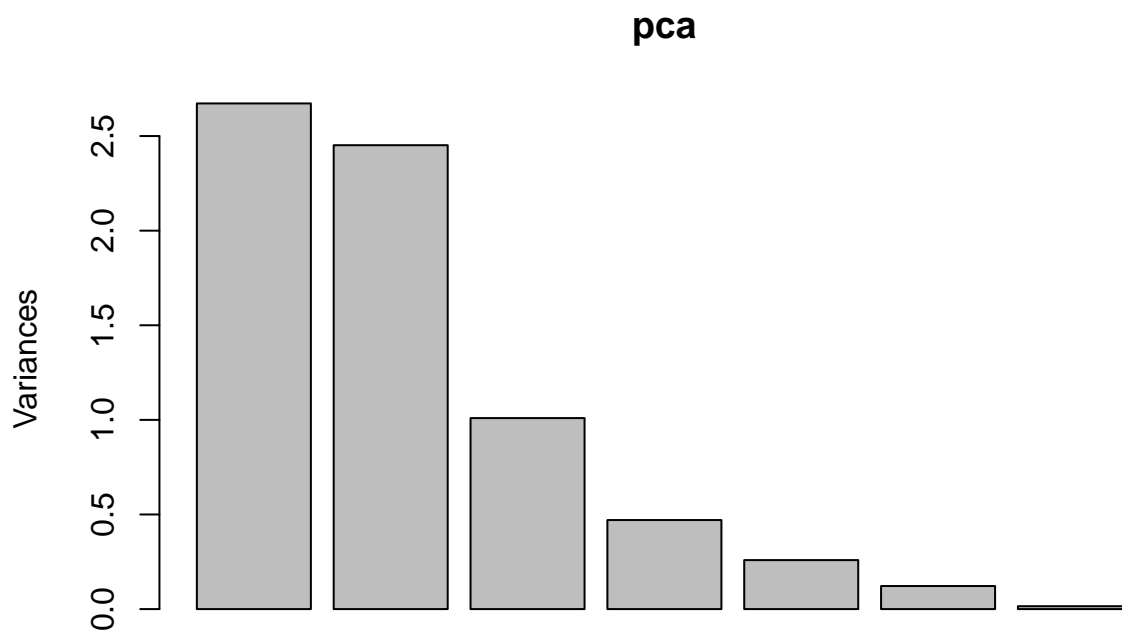






```
# Use PCA analysis
# Standardise the Quality of sleep and stress level
Quality.of.Sleep <- data.frame(scale(sleep$Quality.of.Sleep))
Stress.Level <- data.frame(scale(sleep$Stress.Level))

pca_dataset <- select(sleep, Age, Stress.Level, Sleep.Duration, Heart.Rate, Daily.Steps, Systolic.Pressure)
pca <- prcomp(scale(pca_dataset))
screplot(pca)
```



```
summary(pca)
```

```
## Importance of components:
```

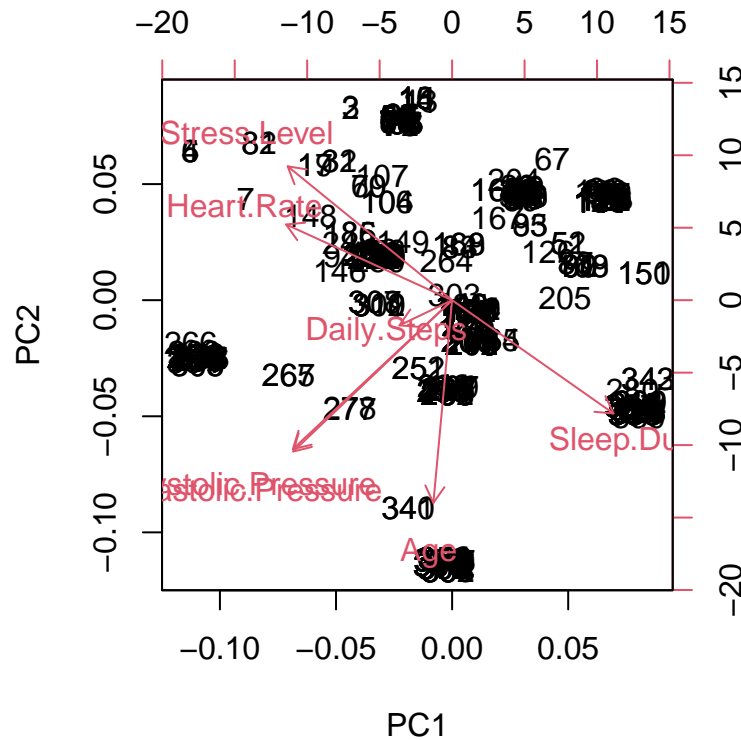
##	PC1	PC2	PC3	PC4	PC5	PC6	PC7
## Standard deviation	1.6347	1.5657	1.0047	0.68599	0.50887	0.34915	0.12343
## Proportion of Variance	0.3818	0.3502	0.1442	0.06723	0.03699	0.01741	0.00218
## Cumulative Proportion	0.3818	0.7320	0.8762	0.94342	0.98041	0.99782	1.00000

```
pca$rotation
```

##	PC1	PC2	PC3	PC4	PC5
## Age	-0.05110437	-0.57728121	0.064471833	-0.10305251	0.80041333
## Stress.Level	-0.44861708	0.38128629	-0.105540137	-0.09632353	0.31770580
## Sleep.Duration	0.44030710	-0.32273014	-0.039563998	0.63186886	-0.06229899
## Heart.Rate	-0.45287362	0.21562391	0.237130370	0.74394202	0.16806079
## Daily.Steps	-0.14365254	-0.07172991	-0.953260327	0.15490561	0.02264232
## Systolic.Pressure	-0.43430485	-0.42409286	0.134808931	-0.05473117	-0.30270358
## Diastolic.Pressure	-0.43348524	-0.43095107	-0.005462661	-0.02006787	-0.36624281
##	PC6	PC7			
## Age	-0.08722894	0.03298360			
## Stress.Level	0.72400112	0.08849828			
## Sleep.Duration	0.54404336	0.03587480			
## Heart.Rate	-0.33164139	-0.02237938			
## Daily.Steps	-0.16521063	-0.11709201			

```
## Systolic.Pressure 0.18453279 -0.69618517
## Diastolic.Pressure 0.02995660 0.70064825
```

```
biplot(pca)
```



Explore the causes of sleep disorder

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

```
## # weights: 48 (30 variable)
## initial value 410.880996
## iter 10 value 196.700033
## iter 20 value 152.981920
## iter 30 value 152.134735
## iter 40 value 151.957185
## iter 50 value 151.945410
## final value 151.945234
## converged
```



```
summary(mlr)
```

```
## Call:
## multinom(formula = Sleep.Disorder ~ Age + Gender + Occupation +
##   Stress.Level + Physical.Activity.Level, data = sleep)
##
## Coefficients:
##           (Intercept)           Age GenderMale OccupationDoctor
## None           12.44314 -0.19430468 -0.9609361           2.455351
## Sleep Apnea    -14.06096 -0.04390343  0.2236164           15.785526
##           OccupationEngineer OccupationLawyer OccupationManager
## None           1.868821           1.866562           18.128481
## Sleep Apnea     13.578401           14.886638           -1.159636
##           OccupationNurse OccupationSales Representative
## None           1.021504                   -8.092882
## Sleep Apnea     18.189903                   34.512717
##           OccupationSalesperson OccupationScientist
## None           -1.315795                   15.46199
## Sleep Apnea     12.581912                   32.80801
##           OccupationSoftware Engineer OccupationTeacher Stress.Level
## None           -0.5859557                   -2.902574 -0.7193111
## Sleep Apnea     -9.0215559                   13.594178 -0.2285089
##           Physical.Activity.Level
## None           0.01439774
## Sleep Apnea     0.03108405
##
## Std. Errors:
##           (Intercept)           Age GenderMale OccupationDoctor
## None           3.054673 0.05347296  1.068607           1.2422732
## Sleep Apnea     3.283065 0.06498043  1.334686           0.9618483
##           OccupationEngineer OccupationLawyer OccupationManager
## None           1.133166           1.329285           2.104373e-08
## Sleep Apnea     1.221164           1.161461           2.418316e-15
##           OccupationNurse OccupationSales Representative
## None           1.230607                   4.803752e-11
## Sleep Apnea     1.374842                   4.336635e-08
##           OccupationSalesperson OccupationScientist
## None           1.429002                   0.7033655
## Sleep Apnea     1.417162                   0.7033655
##           OccupationSoftware Engineer OccupationTeacher Stress.Level
## None           1.647423e+00                   0.8616426  0.2220350
## Sleep Apnea     8.416435e-11                   0.7150064  0.2203134
##           Physical.Activity.Level
## None           0.01614930
## Sleep Apnea     0.02005558
##
## Residual Deviance: 303.8905
## AIC: 363.8905
```

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

```
##           (Intercept)           Age GenderMale OccupationDoctor
```

```

## None          2.535046e+05 0.8234070 0.3825346 1.165052e+01
## Sleep Apnea   7.823530e-07 0.9570464 1.2505912 7.170786e+06
##              OccupationEngineer OccupationLawyer OccupationManager
## None          6.480653e+00 6.466026e+00 7.466197e+07
## Sleep Apnea    7.889047e+05 2.918669e+06 3.136004e-01
##              OccupationNurse OccupationSales Representative
## None          2.777369e+00 3.057073e-04
## Sleep Apnea    7.939163e+07 9.742776e+14
##              OccupationSalesperson OccupationScientist
## None          2.682611e-01 5.188699e+06
## Sleep Apnea    2.912427e+05 1.771487e+14
##              OccupationSoftware Engineer OccupationTeacher Stress.Level
## None          0.5565736664 5.488176e-02 0.4870877
## Sleep Apnea    0.0001207781 8.014498e+05 0.7957192
##              Physical.Activity.Level
## None          1.014502
## Sleep Apnea    1.031572

```

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

```

## Start: AIC=363.89
## Sleep.Disorder ~ Age + Gender + Occupation + Stress.Level + Physical.Activity.Level
##
## trying - Age
## # 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
## trying - Gender
## # weights: 45 (28 variable)
## initial value 410.880996
## iter 10 value 189.916107
## iter 20 value 153.158851
## iter 30 value 152.730924
## iter 40 value 152.606094
## iter 50 value 152.604760
## final value 152.604753
## converged
## trying - Occupation
## # weights: 18 (10 variable)
## initial value 410.880996
## iter 10 value 248.945828
## iter 20 value 238.663986
## iter 20 value 238.663986
## iter 20 value 238.663986
## final value 238.663986
## converged
## trying - Stress.Level
## # weights: 45 (28 variable)

```

```

## initial value 410.880996
## iter 10 value 178.169405
## iter 20 value 159.713073
## iter 30 value 159.078407
## iter 40 value 159.044076
## iter 50 value 159.043249
## final value 159.043246
## converged
## trying - Physical.Activity.Level
## # weights: 45 (28 variable)
## initial value 410.880996
## iter 10 value 171.174411
## iter 20 value 153.943455
## iter 30 value 153.320015
## iter 40 value 153.190516
## iter 50 value 153.186923
## final value 153.186884
## converged
##
##           Df      AIC
## - Gender      28 361.2095
## - Physical.Activity.Level 28 362.3738
## <none>        30 363.8905
## - Stress.Level 28 374.0865
## - Age          28 385.9705
## - Occupation   10 497.3280
## # weights: 45 (28 variable)
## initial value 410.880996
## iter 10 value 189.916107
## iter 20 value 153.158851
## iter 30 value 152.730924
## iter 40 value 152.606094
## iter 50 value 152.604760
## final value 152.604753
## converged
##
## Step: AIC=361.21
## Sleep.Disorder ~ Age + Occupation + Stress.Level + Physical.Activity.Level
##
## trying - Age
## # 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 255.793079
## final value 250.542216
## converged
## trying - Stress.Level

```

```

## # weights: 42 (26 variable)
## initial value 410.880996
## iter 10 value 189.687514
## iter 20 value 164.880744
## iter 30 value 163.978377
## iter 40 value 163.960457
## final value 163.960006
## converged
## trying - Physical.Activity.Level
## # weights: 42 (26 variable)
## initial value 410.880996
## iter 10 value 174.113130
## iter 20 value 154.352748
## iter 30 value 153.931286
## iter 40 value 153.881833
## iter 50 value 153.881081
## iter 50 value 153.881080
## iter 50 value 153.881080
## final value 153.881080
## converged
##
##           Df      AIC
## - Physical.Activity.Level 26 359.7622
## <none>                    28 361.2095
## - Stress.Level            26 379.9200
## - Age                     26 383.3183
## - Occupation              8 517.0844
## # weights: 42 (26 variable)
## initial value 410.880996
## iter 10 value 174.113130
## iter 20 value 154.352748
## iter 30 value 153.931286
## iter 40 value 153.881833
## iter 50 value 153.881081
## iter 50 value 153.881080
## iter 50 value 153.881080
## final value 153.881080
## converged
##
## Step: AIC=359.76
## Sleep.Disorder ~ Age + Occupation + Stress.Level
##
## trying - Age
## # 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
## trying - Occupation
## # weights: 12 (6 variable)
## initial value 410.880996
## iter 10 value 270.787373

```

```

## iter 20 value 270.733386
## iter 20 value 270.733383
## iter 20 value 270.733383
## final value 270.733383
## converged
## trying - Stress.Level
## # weights: 39 (24 variable)
## initial value 410.880996
## iter 10 value 178.569112
## iter 20 value 166.985617
## iter 30 value 166.468862
## iter 40 value 166.449375
## final value 166.449284
## converged
##           Df      AIC
## <none>      26 359.7622
## - Stress.Level 24 380.8986
## - Age          24 388.9303
## - Occupation   6 553.4668

## Call:
## multinom(formula = Sleep.Disorder ~ Age + Occupation + Stress.Level,
##           data = sleep)
##
## Coefficients:
##           (Intercept)           Age OccupationDoctor OccupationEngineer
## None              13.15049 -0.18302686           1.621257           1.236146
## Sleep Apnea      -13.38628 -0.01803231           15.933803           13.592256
##           OccupationLawyer OccupationManager OccupationNurse
## None              1.231801           20.64043           1.496128
## Sleep Apnea       15.710380           -1.53705           18.778673
##           OccupationSales Representative OccupationSalesperson
## None              -10.92876           -2.250945
## Sleep Apnea       36.84379           12.443432
##           OccupationScientist OccupationSoftware Engineer OccupationTeacher
## None              23.88502           -1.425999           -3.163346
## Sleep Apnea       40.92281           -8.507530           13.295103
##           Stress.Level
## None              -0.8015531
## Sleep Apnea      -0.2341703
##
## Residual Deviance: 307.7622
## AIC: 359.7622

```

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 ~ Stress.Level + (1|Occupation), data=sleep)
summary(m2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Quality.of.Sleep ~ Stress.Level + (1 | Occupation)
## Data: sleep
##
## REML criterion at convergence: 308.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6672 -0.2194  0.2041  0.2810  2.7420
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## Occupation (Intercept) 0.4568  0.6758
## Residual              0.1160  0.3406
## Number of obs: 374, groups: Occupation, 11
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  10.09206    0.22035  45.80
```

```
## Stress.Level -0.58522    0.01252 -46.74
##
## Correlation of Fixed Effects:
##      (Intr)
## Stress.Levl -0.326
```