

Veri Analizi Projesi

H. Simay Özgöl & Ataman Önel Ük

2024-05-17

UYKU BOZUKLUK TÜRLEİNİN NEDENİNE İLİŞKİN VERİ ANALİZİ

VERİ SETİNİN TANIMI

Toplamda 374 gözlem ve 13 değişkenden oluşan veri setinde kayıp gözlem bulunmamaktadır. Aşağıda veri setinin linki verilmiş ve değişkenler açıklanmıştır. Veri setinden hareketle; kişilerin sahip olduğu stres seviyesi, uyku kalitesi veya uyku sürelerinin uyku bozukluğunda ne kadar etkili olduğu ele alınacaktır.

- <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>

GÖZLEM NUMARASI: Kişilere ait gözlem numaralarını içermektedir. CİNSİYET: Kişilerin cinsiyetini içerir. YAŞ: Kişilere ait yaşları veren değişkendir. MESLEK: Gözlemlere ait meslekleri vermektedir. UYKU SÜRESİ: Saat cinsinden uyku süresini vermektedir. UYKU KALİTESİ: 1 ile 10 arasında ölçülmüş uyku kalitesini vermektedir. FİZİKSEL AKTİVİTE: Fiziksel aktivite seviyesini “dakika/gün” cinsinden veren değişkendir. STRES SEVİYESİ: 1 ile 10 arasında ölçülmüş stres seviyesini vermektedir. VKİ: Kişiyeye ait vücut kitle indeksini içermektedir. KAN BASINCI: Kişinin kan basıncı ölçümünü “sistolik/diastolik” cinsten veren değişkendir. KALP ATIŞ HIZI: Kişinin dinlenme halindeki dakika başına kalp atış hızını vermektedir. GÜNLÜK ADIM: Günlük atılan adım sayısını vermektedir. UYKU BOZUKLUĞU: Uyku bozukluğu durumunu 3 ayrı kategoride inceleyen değişkendir.

Bağımlı değişken olarak seçilen uyku bozukluğuna ait değışkende bulunan 3 ayrı kategori ve açıklaması aşağıda verilmektedir.

1- Insomnia (Uykusuzluk): Uykuya dalmayı ve uykuda kalmayı zorlaştıran ya da çok erken uyanmaya ve yeniden uykuya dönememeye neden olan yaygın bir uyku bozukluğudur.

2- Sleep Apnea (Uyku Apnesi) : Uyku sırasında üst solunum yolunun tıkanması ile solunumun tekrar tekrar durup başladığı ciddi bir uyku bozukluğudur. Bu, hava yollarının tıkanması (obstrüktif uyku apnesi) ya da beynin solunumu doğru şekilde kontrol etmemesi (merkezi apne) nedeniyle olur.

3- None: Herhangi bir uyku bozukluğu yaşamayan gözlemler için bu kategori uygun görülmüştür.

VKİ değişkeninde kategoriler arasında hem “Normal Weight” hem de “Normal” bulunduğundan düzenlenmiş olup “Normal Weight” kategorisi kaldırılmış “Normal” kategorisi içerisine dahil edilmiştir.

VERİ SETİNİN R ORTAMINA TANITILMASI

```
library(readr)
Uyku_Bozuklugu <-
read.table("C:/Users/w10/OneDrive/Desktop/Sleep_health_and_lifestyle_dataset.
csv", header= TRUE, sep=",")
View(Uyku_Bozuklugu)
```

Veri seti yukarıda yazılan kod yardımıyla R Markdown'a tanımlanmıştır.

```
str(Uyku_Bozuklugu)

## '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" ...

Uyku_Bozuklugu$Sleep.Disorder <- factor(Uyku_Bozuklugu$Sleep.Disorder,
levels=c("None", "Sleep Apnea", "Insomnia"))
Uyku_Bozuklugu$Gender <- factor(Uyku_Bozuklugu$Gender,
levels=c("Female", "Male"))
Uyku_Bozuklugu$BMI.Category <- factor(Uyku_Bozuklugu$BMI.Category,
levels=c("Normal", "Obese", "Overweight", "Normal Weight"))
Uyku_Bozuklugu$Quality.of.Sleep <- factor(Uyku_Bozuklugu$Quality.of.Sleep,
levels = 1:10, ordered=TRUE)
Uyku_Bozuklugu$Stress.Level<- factor(Uyku_Bozuklugu$Stress.Level, levels =
1:10, ordered=TRUE)

levels(Uyku_Bozuklugu$BMI.Category)[levels(Uyku_Bozuklugu$BMI.Category) ==
"Normal Weight"] <- "Normal"
levels(Uyku_Bozuklugu$BMI.Category)

## [1] "Normal" "Obese" "Overweight"
```

“Normal” ve “Normal Weight” adlı iki kategori yukarıda yazılan kod yardımıyla tek bir grup haline getirilmiştir.

```
str(Uyku_Bozuklugu)
```

```
## 'data.frame':   374 obs. of  13 variables:
## $ Person.ID      : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Gender         : Factor w/ 2 levels "Female","Male": 2 2 2 2 2
2 2 2 2 2 ...
## $ 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 : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 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      : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 6
8 8 8 8 8 7 6 6 6 ...
## $ BMI.Category      : Factor w/ 3 levels "Normal","Obese",...: 3 1 1
2 2 2 2 1 1 1 ...
## $ 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    : Factor w/ 3 levels "None","Sleep Apnea",...: 1
1 1 2 2 3 3 1 1 1 ...
```

Gerekli düzenlemeler sayesinde veri setindeki değişkenlerde bulunan kategoriler, yapılacak olan analizlerde hata alınmaması için R'a özellikle belirtilmiştir.

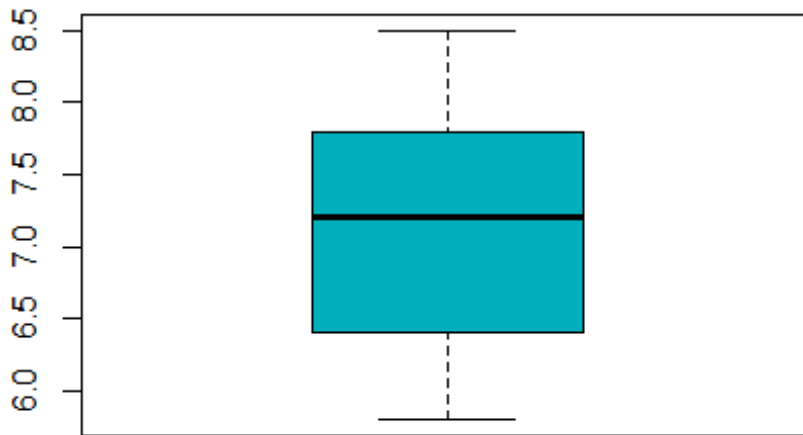
```
quantile(Uyku_Bozuklugu$Sleep.Duration)
```

```
##    0%   25%   50%   75%  100%
##  5.8   6.4   7.2   7.8   8.5
```

```
median(Uyku_Bozuklugu$Sleep.Duration)
```

```
## [1] 7.2
```

```
boxplot(Uyku_Bozuklugu$Sleep.Duration, col = "#00AFBB")
```



Eğitim ve test veri setleri oluşturulmadan önce orjinal veri seti üzerinden uyku süresi değişkeninin kartil ve medyan hesapları yapılmış, sonrasında box-plot'u elde edilmiştir.

1. kartil değeri 6.4, 3. kartil değeri ise 7.8 gelmiş olup, bireylerin ortalama uyku süresi 7.2 saat olarak bulunmuştur. Bu grafikte herhangi bir uç değere rastlanmamıştır.

```
library(ggpubr)

## Zorunlu paket yükleniyor: ggplot2

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

freq_sd <- Uyku_Bozuklugu %>%
  count(Gender, Sleep.Disorder) %>%
  rename(count = n)
ggbarplot(freq_sd, x = "Sleep.Disorder", y = "count",
  fill = "Gender", palette = c("Male" = "lightblue", "Female" =
"lightpink"),
```

```

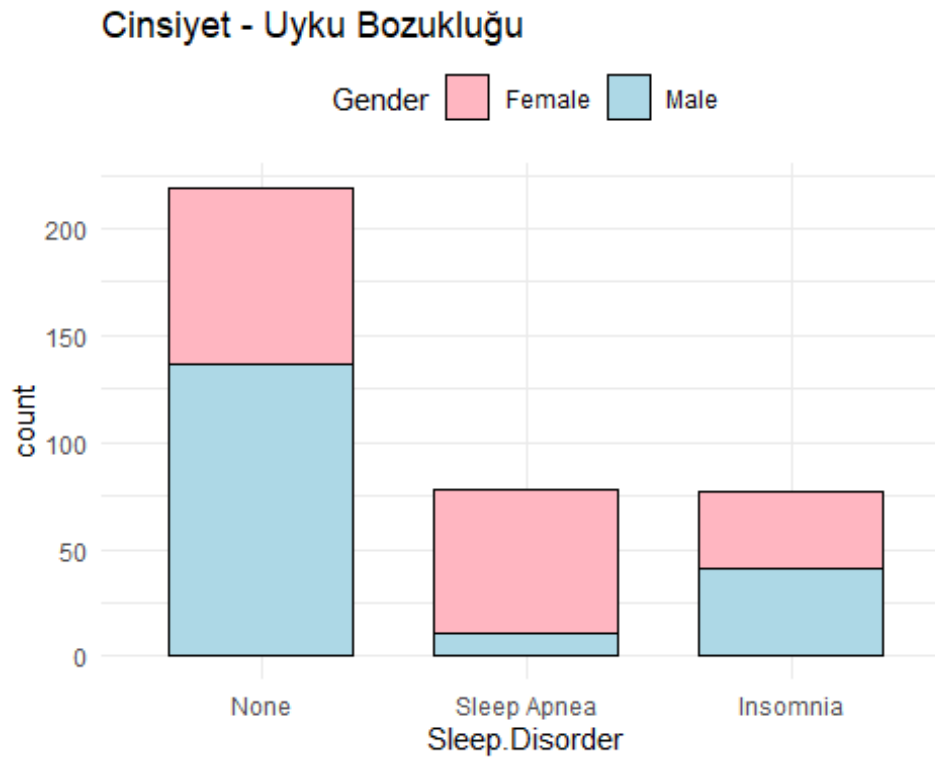
title = "Cinsiyet - Uyku Bozukluğu",
legend.title = "Gender") +
theme_minimal() +
theme(legend.position = "top")

```

```

## Warning: No shared levels found between `names(values)` of the manual
scale and the
## data's colour values.

```



Uyku bozukluğu vakaları için cinsiyet dağılımlarına bakılmış olup, aşağıdaki yorumlar elde edilmiştir: -Uykusuzluk (Insomnia) rahatsızlığı için cinsiyetlerin hemen hemen eşit dağıldığı söylenebilir. -Herhangi bir rahatsızlığı olmayan bireylerin bulunduğu grupta, erkeklerin sayısının kadınlara göre biraz daha fazla olduğu görülmektedir. -Uyku apnesi (Sleep Apnea) grubunda ise kadınların sayısının daha fazla olduğu görülmektedir.

EĞİTİM VE TEST VERİ SETLERİNİN ELDE EDİLMESİ

```

set.seed(400)
trainIndex <- sample(1:nrow(Uyku_Bozuklugu), size =
round(0.7*nrow(Uyku_Bozuklugu)), replace=FALSE)
train<- Uyku_Bozuklugu[trainIndex,]
test<- Uyku_Bozuklugu[-trainIndex,]

```

Yazılan kod yardımıyla eğitim ve test veri setleri sırasıyla %70 ile %30 oranlarında rastgele olacak şekilde örneklemeler oluşturulmuştur. Oluşturulan bu veri setlerini excel ortamında görüntülemek için ise "write.xlsx" fonksiyonu ile aşağıdaki adımlar takip edilmiştir.

```

library(openxlsx)
write.xlsx(train, 'train_data.xlsx')
write.xlsx(test, 'test_data.xlsx')

library(dplyr)
glimpse(train)

## Rows: 262
## Columns: 13
## $ Person.ID          <int> 101, 279, 343, 220, 342, 156, 14, 13, 142,
327...
## $ Gender             <fct> Female, Female, Female, Male, Female,
Male, Ma...
## $ Age                <int> 36, 50, 56, 43, 56, 39, 29, 29, 38, 53,
31, 35...
## $ Occupation         <chr> "Teacher", "Nurse", "Doctor",
"Salesperson", "...
## $ Sleep.Duration     <dbl> 7.2, 6.1, 8.2, 6.5, 8.2, 7.2, 6.0, 6.1,
7.1, 8...
## $ Quality.of.Sleep   <ord> 8, 6, 9, 6, 9, 8, 6, 6, 8, 9, 7, 8, 7, 8,
9, 6...
## $ Physical.Activity.Level <int> 60, 90, 90, 45, 90, 60, 30, 30, 60, 30,
75, 60...
## $ Stress.Level       <ord> 4, 8, 3, 7, 3, 5, 8, 8, 5, 3, 6, 4, 4, 4,
3, 8...
## $ BMI.Category       <fct> Normal, Overweight, Normal, Overweight,
Normal...
## $ Blood.Pressure     <chr> "115/75", "140/95", "118/75", "130/85",
"118/7...
## $ Heart.Rate         <int> 68, 75, 65, 72, 65, 68, 70, 70, 68, 65,
70, 65...
## $ Daily.Steps        <int> 7000, 10000, 10000, 6000, 10000, 8000,
8000, 8...
## $ Sleep.Disorder     <fct> None, Insomnia, None, Sleep Apnea, None,
None,...

summary(train)

##      Person.ID      Gender      Age      Occupation
Sleep.Duration
## Min.   : 2.0   Female:129   Min.    :28.00   Length:262      Min.
:5.80
## 1st Qu.:102.5   Male  :133   1st Qu.:36.00   Class :character 1st
Qu.:6.40
## Median :200.5                Median :43.00   Mode  :character Median
:7.20
## Mean   :196.7                Mean   :42.94                Mean
:7.17
## 3rd Qu.:294.8                3rd Qu.:50.00                3rd
Qu.:7.80
## Max.   :374.0                Max.    :59.00                Max.

```

```

:8.50
##
## Quality.of.Sleep Physical.Activity.Level Stress.Level BMI.Category
## 6 :76 Min. :30.00 3 :58 Normal :152
## 8 :70 1st Qu.:45.00 8 :52 Obese : 3
## 9 :59 Median :60.00 5 :50 Overweight:107
## 7 :50 Mean :59.02 4 :40
## 5 : 4 3rd Qu.:75.00 7 :33
## 4 : 3 Max. :90.00 6 :29
## (Other): 0 (Other): 0
## Blood.Pressure Heart.Rate Daily.Steps Sleep.Disorder
## Length:262 Min. :65.00 Min. : 3000 None :154
## Class :character 1st Qu.:68.00 1st Qu.: 5275 Sleep Apnea: 58
## Mode :character Median :70.00 Median : 7000 Insomnia : 50
## Mean :69.92 Mean : 6851
## 3rd Qu.:72.00 3rd Qu.: 8000
## Max. :86.00 Max. :10000
##

```

Burada “str()” fonksiyonu kullanmak yerine “train” veri setini daha detaylı incelemek adına “glimpse()” fonksiyonu tercih edilmiştir. Bu fonksiyon veri setini satır yerine sütun halinde göstermekte olup özellikle geniş veri setleriyle çalışırken daha okunaklı sonuçlar vermektedir. Ardından “summary()” fonksiyonu ile istatistiksel özetler elde edilmiştir.

```

train$Sleep.Disorder <- factor(train$Sleep.Disorder, levels=c("None", "Sleep
Apnea", "Insomnia"))
train$Gender <- factor(train$Gender, levels=c("Female", "Male"))
train$BMI.Category <- factor(train$BMI.Category,
levels=c("Normal", "Obese", "Overweight", "Normal Weight"))
train$Quality.of.Sleep <- factor(train$Quality.of.Sleep, levels = 1:10,
ordered=TRUE)
train$Stress.Level<- factor(train$Stress.Level, levels = 1:10, ordered=TRUE)
levels(train$BMI.Category)[levels(train$BMI.Category) == "Normal Weight"] <-
"Normal"
levels(train$BMI.Category)
## [1] "Normal" "Obese" "Overweight"

```

İnceleme sonucunda R’da tanımlanmış olan veri setinin hatalı şekilde okunduğu görülmüş ve gerekli görülen değişkenlerin türleri düzenlenmiştir. Ek olarak asıl veri setinde 4 kategoriden oluşan, vücut kitle indeksi bilgisini veren değişken “Normal”, “Obez” ve “Çok Kilolu” olacak şekilde 3 kategoriye ayrılmıştır. Düzenlenmiş veri setini kontrol etmek için yeniden “glimpse()” ve “summary()” fonksiyonlarından yararlanılmıştır.

```

library(dplyr)
glimpse(train)

## Rows: 262
## Columns: 13
## $ Person.ID <int> 101, 279, 343, 220, 342, 156, 14, 13, 142,
327...

```

```
## $ Gender      <fct> Female, Female, Female, Male, Female,
Male, Ma...
## $ Age         <int> 36, 50, 56, 43, 56, 39, 29, 29, 38, 53,
31, 35...
## $ Occupation  <chr> "Teacher", "Nurse", "Doctor",
"Salesperson", "...
## $ Sleep.Duration <dbl> 7.2, 6.1, 8.2, 6.5, 8.2, 7.2, 6.0, 6.1,
7.1, 8...
## $ Quality.of.Sleep <ord> 8, 6, 9, 6, 9, 8, 6, 6, 8, 9, 7, 8, 7, 8,
9, 6...
## $ Physical.Activity.Level <int> 60, 90, 90, 45, 90, 60, 30, 30, 60, 30,
75, 60...
## $ Stress.Level <ord> 4, 8, 3, 7, 3, 5, 8, 8, 5, 3, 6, 4, 4, 4,
3, 8...
## $ BMI.Category <fct> Normal, Overweight, Normal, Overweight,
Normal...
## $ Blood.Pressure <chr> "115/75", "140/95", "118/75", "130/85",
"118/7...
## $ Heart.Rate  <int> 68, 75, 65, 72, 65, 68, 70, 70, 68, 65,
70, 65...
## $ Daily.Steps <int> 7000, 10000, 10000, 6000, 10000, 8000,
8000, 8...
## $ Sleep.Disorder <fct> None, Insomnia, None, Sleep Apnea, None,
None,...
```

`summary(train)`

```
##   Person.ID      Gender      Age      Occupation
Sleep.Duration
##   Min.   : 2.0   Female:129   Min.    :28.00   Length:262      Min.
:5.80
##   1st Qu.:102.5   Male  :133   1st Qu.:36.00   Class :character  1st
Qu.:6.40
##   Median :200.5                Median :43.00   Mode  :character  Median
:7.20
##   Mean   :196.7                Mean   :42.94                Mean
:7.17
##   3rd Qu.:294.8                3rd Qu.:50.00                3rd
Qu.:7.80
##   Max.    :374.0                Max.    :59.00                Max.
:8.50
##
##   Quality.of.Sleep Physical.Activity.Level Stress.Level  BMI.Category
##   6      :76      Min.    :30.00      3      :58      Normal    :152
##   8      :70      1st Qu.:45.00      8      :52      Obese      : 3
##   9      :59      Median :60.00      5      :50      Overweight:107
##   7      :50      Mean   :59.02      4      :40
##   5      : 4      3rd Qu.:75.00      7      :33
##   4      : 3      Max.    :90.00      6      :29
##   (Other): 0      (Other): 0
```



```
## Blood.Pressure      Heart.Rate      Daily.Steps      Sleep.Disorder
## Length:262          Min.      :65.00    Min.      : 3000    None       :154
## Class :character     1st Qu.:68.00    1st Qu.: 5275    Sleep Apnea: 58
## Mode  :character     Median :70.00    Median : 7000    Insomnia   : 50
##                      Mean      :69.92    Mean      : 6851
##                      3rd Qu.:72.00    3rd Qu.: 8000
##                      Max.      :86.00    Max.      :10000
##
```

Değişkenler için istatistiksel hesaplamaların doğru bir şekilde yapıldığı sonucuna varılmıştır.

EĞİTİM VERİ SETİNİ GÖRSELLEŞTİRME

```
library(funModeling)
```

```
## Zorunlu paket yükleniyor: Hmisc
```

```
##
```

```
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
```

```
##
```

```
##      src, summarize
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      format.pval, units
```

```
## funModeling v.1.9.5 :)
```

```
## Examples and tutorials at livebook.datascienceheroes.com
```

```
## / Now in Spanish: librovivodecienciadedatos.ai
```

```
profiling_num(train)
```

```
##          variable      mean      std_dev variation_coef      p_01
## 1      Person.ID 196.748092 109.6070546    0.55709336    7.660
## 2           Age  42.942748   8.9026539    0.20731449   28.610
## 3 Sleep.Duration   7.170229   0.8260783    0.11520947    5.961
## 4 Physical.Activity.Level 59.019084 21.5410954    0.36498525   30.000
## 5      Heart.Rate  69.916031   3.7895565    0.05420154   65.000
## 6    Daily.Steps 6850.763359 1592.2644028    0.23242146 3883.000
##      p_05  p_25  p_50  p_75  p_95  p_99      skewness kurtosis
iqr
## 1  24.05 102.5 200.5 294.75 360.95 371.39 -0.07451139 1.773856
192.25
## 2  30.00  36.0  43.0  50.00  59.00  59.00  0.18500656 1.976059
14.00
## 3   6.00   6.4   7.2   7.80   8.40   8.50 -0.01886109 1.617917
1.40
## 4  30.00  45.0  60.0  75.00  90.00  90.00  0.06383172 1.651603
30.00
## 5  65.00  68.0  70.0  72.00  75.00  81.78  0.98473852 4.834624
4.00
```

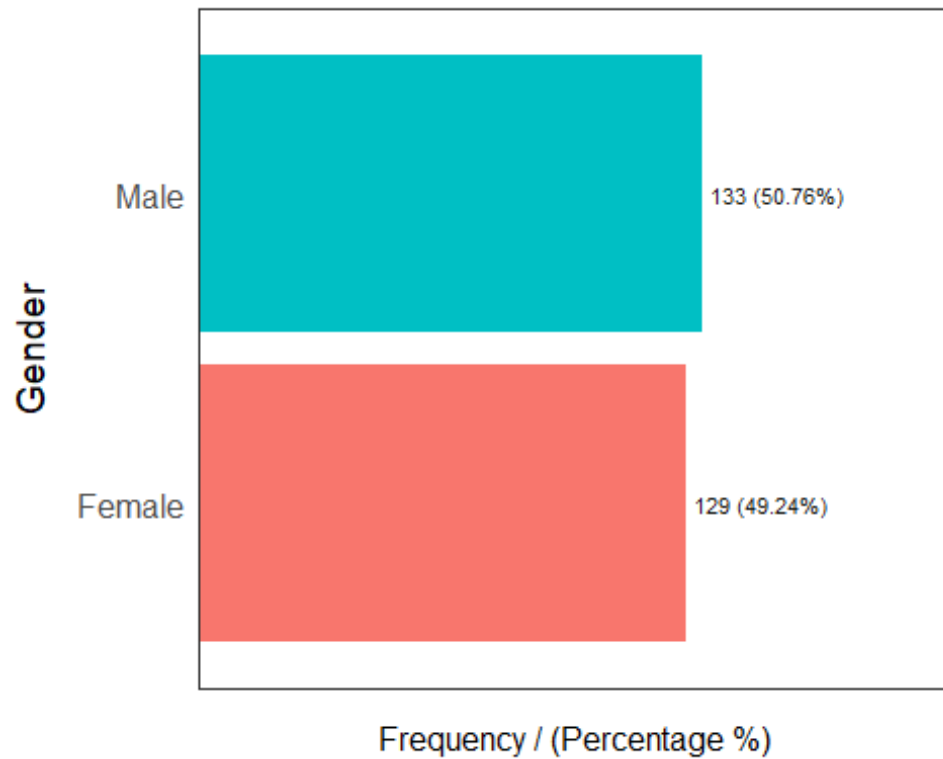
```
## 6 5000.00 5275.0 7000.0 8000.00 10000.00 10000.00 0.30169114 2.421918
2725.00
##           range_98      range_80
## 1 [7.66, 371.39] [44.1, 342.9]
## 2 [28.61, 59] [31, 56]
## 3 [5.961, 8.5] [6.01, 8.29]
## 4 [30, 90] [30, 90]
## 5 [65, 81.78] [65, 75]
## 6 [3883, 10000] [5000, 8000]
```

```
plot_num(train)
```

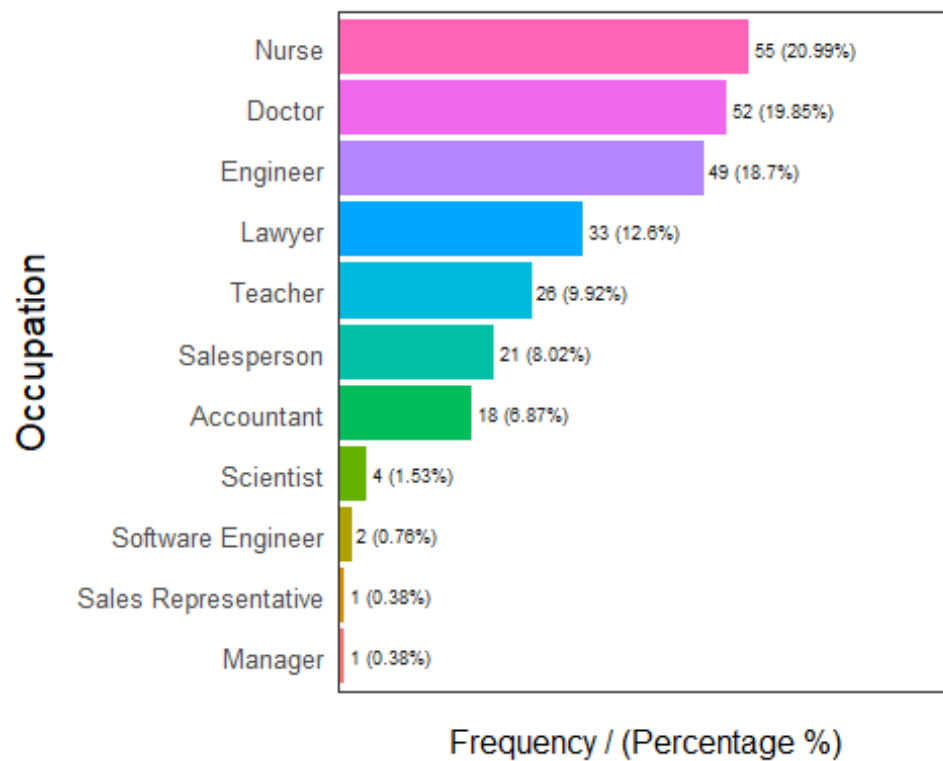


```
freq(train)
```

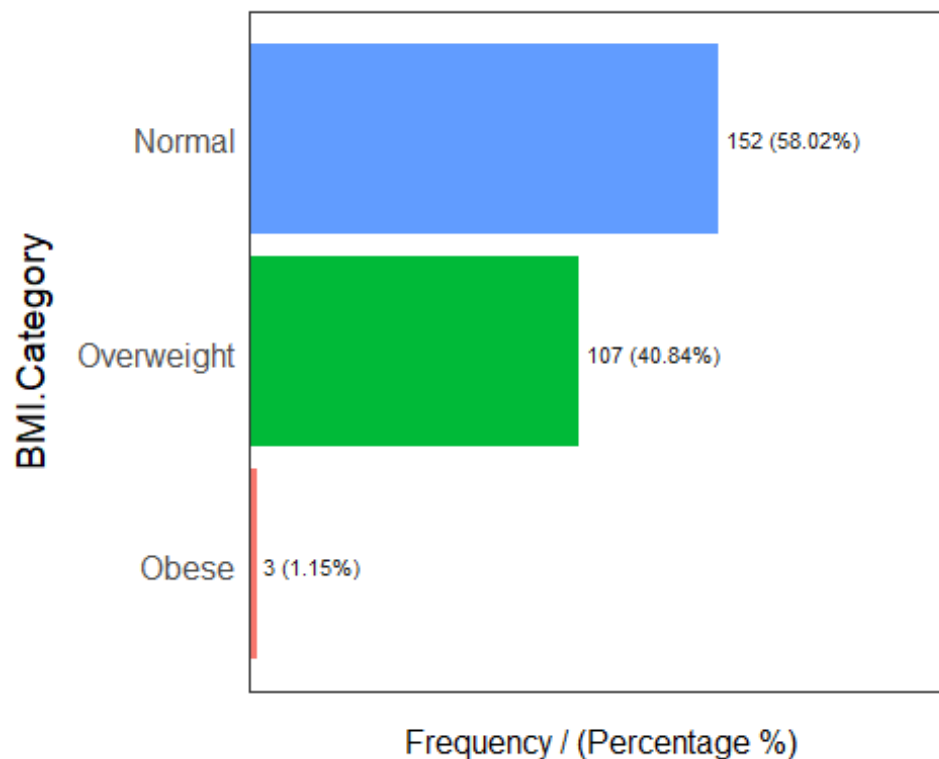
```
## Warning: The `scale` argument of `guides()` cannot be `FALSE`. Use
"none" instead as
## of ggplot2 3.3.4.
## i The deprecated feature was likely used in the funModeling package.
## Please report the issue at
<https://github.com/pablo14/funModeling/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



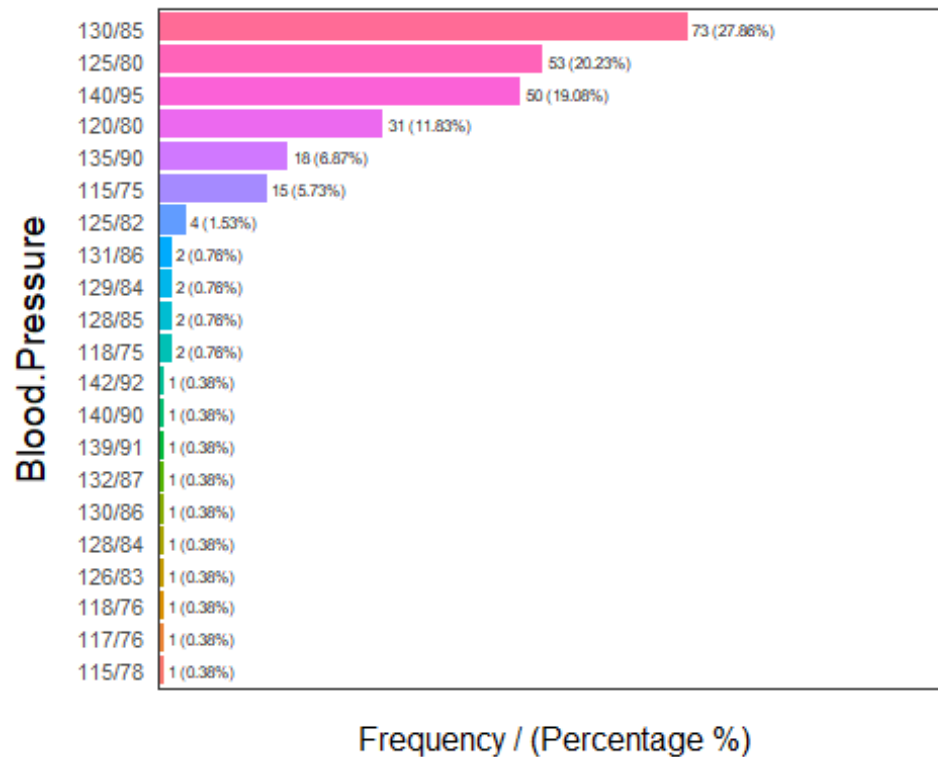
##	Gender	frequency	percentage	cumulative_perc
## 1	Male	133	50.76	50.76
## 2	Female	129	49.24	100.00



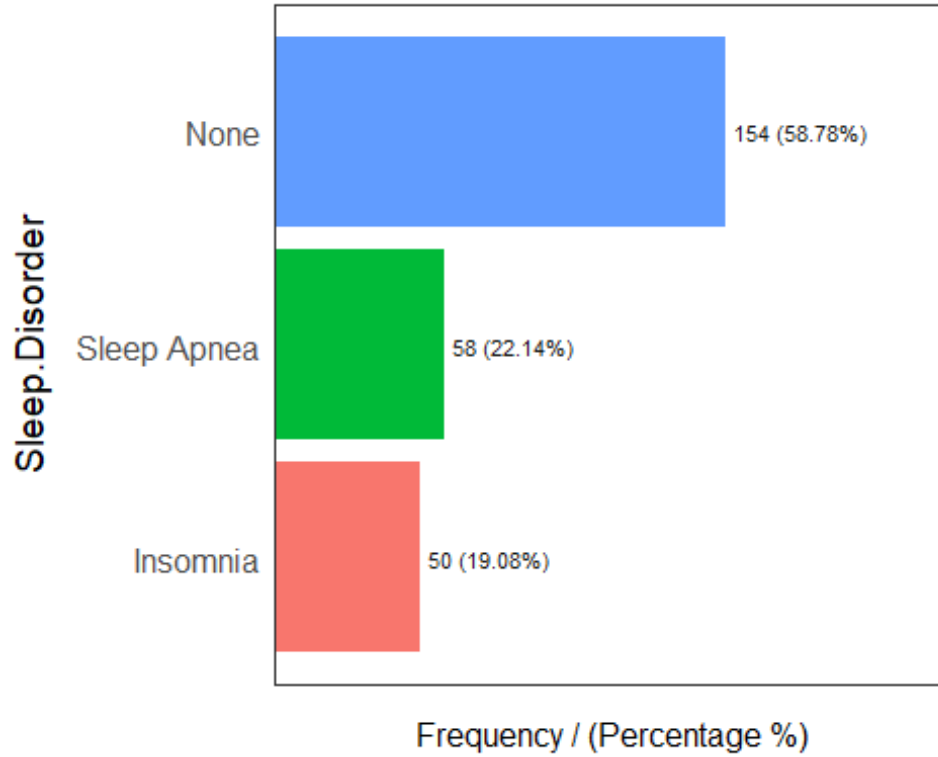
##	Occupation	frequency	percentage	cumulative_perc
## 1	Nurse	55	20.99	20.99
## 2	Doctor	52	19.85	40.84
## 3	Engineer	49	18.70	59.54
## 4	Lawyer	33	12.60	72.14
## 5	Teacher	26	9.92	82.06
## 6	Salesperson	21	8.02	90.08
## 7	Accountant	18	6.87	96.95
## 8	Scientist	4	1.53	98.48
## 9	Software Engineer	2	0.76	99.24
## 10	Manager	1	0.38	99.62
## 11	Sales Representative	1	0.38	100.00



##	BMI.Category	frequency	percentage	cumulative_perc
## 1	Normal	152	58.02	58.02
## 2	Overweight	107	40.84	98.86
## 3	Obese	3	1.15	100.00



##	Blood.Pressure	frequency	percentage	cumulative_perc
## 1	130/85	73	27.86	27.86
## 2	125/80	53	20.23	48.09
## 3	140/95	50	19.08	67.17
## 4	120/80	31	11.83	79.00
## 5	135/90	18	6.87	85.87
## 6	115/75	15	5.73	91.60
## 7	125/82	4	1.53	93.13
## 8	118/75	2	0.76	93.89
## 9	128/85	2	0.76	94.65
## 10	129/84	2	0.76	95.41
## 11	131/86	2	0.76	96.17
## 12	115/78	1	0.38	96.55
## 13	117/76	1	0.38	96.93
## 14	118/76	1	0.38	97.31
## 15	126/83	1	0.38	97.69
## 16	128/84	1	0.38	98.07
## 17	130/86	1	0.38	98.45
## 18	132/87	1	0.38	98.83
## 19	139/91	1	0.38	99.21
## 20	140/90	1	0.38	99.59
## 21	142/92	1	0.38	100.00



```
## Sleep.Disorder frequency percentage cumulative_perc
## 1 None 154 58.78 58.78
## 2 Sleep Apnea 58 22.14 80.92
## 3 Insomnia 50 19.08 100.00
```

```
## [1] "Variables processed: Gender, Occupation, BMI.Category,
Blood.Pressure, Sleep.Disorder"
```

Elde edilen grafikler sonucunda cinsiyet gruplarının dağılımının birbirine hemen hemen eşit olduğu görülmektedir.

Histogram grafiklerine bakıldığında ise herhangi bir değişkenin normal dağılmadığı kanısına varılabilir. Örneğin “Kalp Atış Hızı” bilgisini veren değişkenin genel olarak gözlemlenen konumları 65-75 aralığında olup sağa çarpık bir dağılıma sahip olduğu söylenebilir. “Fiziksel Aktivite Seviyesi” değişkeninin ise kendi arasında kümeleşip, 5 ana gruba ayrıldığı söylenebilir.

Uyku Bozukluğu değişkeninde bulunan kategoriler için ise “Uyku Apnesi” ile “Uykusuzluk” rahatsızlıklarının dağılımlarının birbirine yakın olduğu, herhangi bir uyku rahatsızlığına sahip olmayanların dağılımının ise çok daha fazla olduğu söylenebilir.

Meslek grupları için en çok frekansa sahip olan mesleklerin sırasıyla hemşirelik, doktorluk ve mühendislik olduğu görülmektedir. Aynı zamanda veri setinin bu değişken adına çok farklı kategorilere sahip olduğu söylenebilir.

```
library(psych)
```

```
##
## Attaching package: 'psych'

## The following object is masked from 'package:Hmisc':
##
##      describe

## The following objects are masked from 'package:ggplot2':
##
##      %+%, alpha

describeBy(train,train$Sleep.Disorder)

##
## Descriptive statistics by group
## group: None
##
```

	vars	n	mean	sd	median	trimmed	mad
min							
## Person.ID	1	154	154.88	105.54	133.0	149.22	110.45
2							
## Gender	2	154	1.64	0.48	2.0	1.68	0.00
1							
## Age	3	154	39.66	8.31	38.0	39.06	7.41
28							
## Occupation	4	154	3.29	2.11	3.0	2.91	1.48
1							
## Sleep.Duration	5	154	7.39	0.77	7.6	7.43	0.59
6							
## Quality.of.Sleep	6	154	7.64	1.02	8.0	7.68	1.48
6							
## Physical.Activity.Level	7	154	57.27	22.01	60.0	56.60	28.17
30							
## Stress.Level	8	154	5.13	1.65	5.0	5.04	1.48
3							
## BMI.Category	9	154	1.16	0.54	1.0	1.00	0.00
1							
## Blood.Pressure	10	154	8.18	4.19	7.0	8.21	1.48
1							
## Heart.Rate	11	154	68.97	2.70	70.0	68.88	2.97
65							
## Daily.Steps	12	154	6822.08	1457.38	7000.0	6825.81	1482.60
4200							
## Sleep.Disorder	13	154	1.00	0.00	1.0	1.00	0.00
1							
##		max	range	skew	kurtosis	se	
## Person.ID		360.0	358.0	0.52	-0.98	8.50	
## Gender		2.0	1.0	-0.59	-1.66	0.04	
## Age		59.0	31.0	0.66	-0.74	0.67	
## Occupation		11.0	10.0	2.19	5.17	0.17	
## Sleep.Duration		8.5	2.5	-0.49	-0.73	0.06	
## Quality.of.Sleep		9.0	3.0	-0.35	-1.01	0.08	

```

## Physical.Activity.Level    90.0    60.0  0.03    -1.34    1.77
## Stress.Level               8.0     5.0  0.42    -0.81    0.13
## BMI.Category               3.0     2.0  3.12     7.78    0.04
## Blood.Pressure            20.0    19.0  0.47     0.15    0.34
## Heart.Rate                77.0    12.0  0.24    -0.05    0.22
## Daily.Steps              10000.0 5800.0 -0.11    -1.21   117.44
## Sleep.Disorder             1.0     0.0   NaN      NaN     0.00
## -----
## group: Sleep Apnea
##               vars  n    mean      sd median trimmed      mad
min
## Person.ID           1 58   280.72  102.01   295.5   297.31   85.25
4.0
## Gender              2 58    1.12   0.33    1.0    1.04    0.00
1.0
## Age                 3 58    50.41   8.90   50.0   51.65   10.38
28.0
## Occupation          4 58    6.19   1.57    6.0    6.06    0.00
2.0
## Sleep.Duration      5 58    7.07   0.98    6.8    7.08    1.41
5.8
## Quality.of.Sleep    6 58    7.26   1.70    6.0    7.38    2.97
4.0
## Physical.Activity.Level 7 58   73.78  18.65   75.0   76.46   22.24
30.0
## Stress.Level        8 58    5.59   2.36    7.0    5.60    1.48
3.0
## BMI.Category        9 58    2.79   0.59    3.0    2.96    0.00
1.0
## Blood.Pressure     10 58   18.62   2.97   20.0   19.23    0.00
6.0
## Heart.Rate         11 58   72.57   4.69   73.0   72.00    7.41
68.0
## Daily.Steps        12 58  7603.45 2028.31 7000.0 7737.50 2816.94
3000.0
## Sleep.Disorder     13 58    2.00   0.00    2.0    2.00    0.00
2.0
##               max range  skew kurtosis      se
## Person.ID      374.0 370.0 -1.40    0.99   13.39
## Gender          2.0   1.0  2.27    3.20    0.04
## Age            59.0  31.0 -1.05    0.22    1.17
## Occupation     11.0   9.0  0.86    4.01    0.21
## Sleep.Duration  8.2   2.4  0.03   -1.91    0.13
## Quality.of.Sleep 9.0   5.0 -0.17   -1.53    0.22
## Physical.Activity.Level 90.0 60.0 -1.20    0.25    2.45
## Stress.Level    8.0   5.0 -0.09   -1.93    0.31
## BMI.Category    3.0   2.0 -2.53    4.73    0.08
## Blood.Pressure  20.0  14.0 -2.17    4.41    0.39
## Heart.Rate     86.0  18.0  0.73   -0.04    0.62
## Daily.Steps    10000.0 7000.0 -0.21   -0.98   266.33

```



```
## Sleep.Disorder      2.0      0.0      NaN      NaN      0.00
## -----
## group: Insomnia
##               vars  n      mean      sd median trimmed  mad  min
## Person.ID      1 50  228.30  47.29  234.5  229.38 38.55  68
## Gender         2 50   1.54   0.50   2.0    1.55  0.00   1
## Age           3 50  44.38   3.54  44.0   44.12  1.48  33
## Occupation     4 50   7.76   3.34   8.0    8.18  4.45   1
## Sleep.Duration  5 50   6.60   0.41   6.5    6.52  0.15   6
## Quality.of.Sleep 6 50   6.62   0.73   7.0    6.55  1.48   5
## Physical.Activity.Level 7 50  47.30  12.13  45.0   45.00  0.00  30
## Stress.Level    8 50   5.82   1.49   7.0    5.88  0.00   3
## BMI.Category    9 50   2.78   0.62   3.0    2.98  0.00   1
## Blood.Pressure 10 50  14.40   2.71  13.0   14.40  0.00   7
## Heart.Rate     11 50  69.76   4.10  72.0   69.53  1.48  65
## Daily.Steps    12 50 6066.00 886.09 6000.0 6000.00  0.00 3500
## Sleep.Disorder 13 50   3.00   0.00   3.0    3.00  0.00   3
##               max range skew kurtosis  se
## Person.ID     316.0 248.0 -0.74   1.79  6.69
## Gender         2.0   1.0 -0.16  -2.01  0.07
## Age           53.0  20.0  0.17   2.42  0.50
## Occupation     11.0  10.0 -0.80  -0.63  0.47
## Sleep.Duration  8.3   2.3  2.37   6.06  0.06
## Quality.of.Sleep 9.0   4.0  0.69   0.87  0.10
## Physical.Activity.Level 90.0 60.0 2.50   6.60  1.72
## Stress.Level    8.0   5.0 -0.35  -1.65  0.21
## BMI.Category     3.0   2.0 -2.42   4.05  0.09
## Blood.Pressure  21.0  14.0 -0.14   0.63  0.38
## Heart.Rate     83.0  18.0  0.42   0.27  0.58
## Daily.Steps   10000.0 6500.0 1.73   7.75 125.31
## Sleep.Disorder  3.0   0.0   NaN    NaN  0.00
```

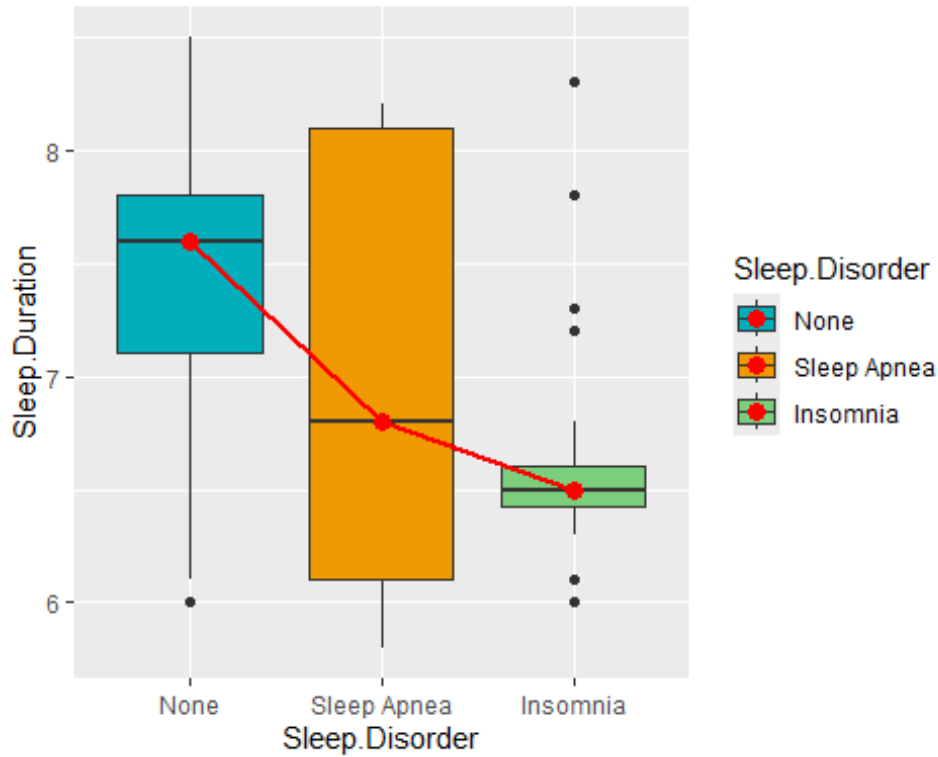
Yukarıda yazılı olan “describeBy()” fonksiyonu ile kategorik değişkenimiz olan “Uyku Bozukluğu” diğer değişkenler için 3 ayrı grup adına ayrıntılı olarak incelenmiştir.

Örneğin herhangi bir rahatsızlığa sahip olmayan bireylerin yaşlarının ortalama 39, uyku apnesine sahip bireylerin 50 ve uykusuzluk sıkıntısı çeken bireylerin ise 44 yaşında olduğu görülmektedir.

UYKU SÜRELERİ İLE RAHATSIZLIK TÜRLERİ İÇİN BOX-PLOT GRAFİĞİ

```
library(ggplot2)
ggplot(train, aes(x= Sleep.Disorder, y= Sleep.Duration, fill=Sleep.Disorder))+
  geom_boxplot() +
  stat_summary(fun = median, geom="point", color= "red", size = 3) +
  stat_summary(fun = median, geom="line", aes(group=1), color= "red", size =
1) +
  scale_fill_manual(values = c("None" = "#00AFBB", "Insomnia" = "palegreen3",
"Sleep Apnea" = "orange2"))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```



İki değişken için hazırlanmış olan box-plot'lara bakıldığında varyansların homojen olmadığını, gruplara göre farklılık gösterdiğini söylemek mümkündür. Herhangi bir uyku rahatsızlığına sahip olmayan bireyleri içeren grubun ortalama uyku süresinin diğer gruplara göre daha fazla olduğu ve gruplar arasında anlamlı bir farklılık olduğu görülmektedir.

Bir diğer göze çarpan bilgi ise “Uykusuzluk” rahatsızlığını belirten grupta çok sayıda aykırı değer olduğu göze çarpmaktadır. Aynı zamanda bu grupta değişimin en az ve “Uyku Apnesi” grubunda ise değişimin en fazla olduğu söylenebilir.

NOKTA ÖLÇÜLERİ

```
n <- nrow(train)  
sorted_train <- train[order(train$Sleep.Duration),]  
  
# Çift ise;  
a <- (n/2)  
b <- (n/2)+1  
(sorted_train$Sleep.Duration[a] + sorted_train$Sleep.Duration[b])/2  
  
## [1] 7.2
```

```
median(train$Sleep.Duration)
```

```
## [1] 7.2
```

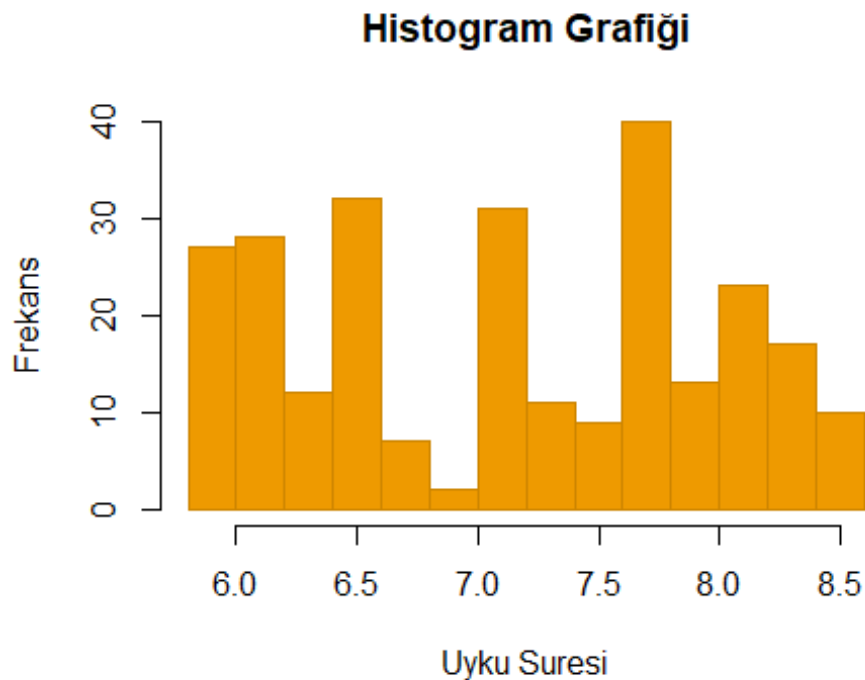
```
mean(train$Sleep.Duration)
```

```
## [1] 7.170229
```

Bu bölümde veri setinde bulunan gözlemlerin sahip olduğu uyku sürelerine bağlı olarak küçükten büyüğe sıralanmış; elde edilen analizler neticesinde uyku sürelerinin ortalaması 7.17 ile ortancası 7.2 saat olarak bulunup, histogram ve yoğunluk grafikleri elde edilmiştir.

YOĞUNLUK VE HISTOGRAM GRAFİĞİ

```
hist(train$Sleep.Duration,  
      xlab = "Uyku Suresi" ,  
      ylab = "Frekans",  
      main = "Histogram Grafiği",  
      col = "orange2",  
      border = "orange3")
```



```
# Histogram grafiği için çarpıklık, basıklık ve değişim katsayısı hesapları;  
library(e1071)
```

```
##
```

```
## Attaching package: 'e1071'
```

```

## The following object is masked from 'package:Hmisc':
##
##      impute

skewness(train$Sleep.Duration) #çarpıklık

## [1] -0.01875321

kurtosis(train$Sleep.Duration) #basıklık

## [1] -1.39441

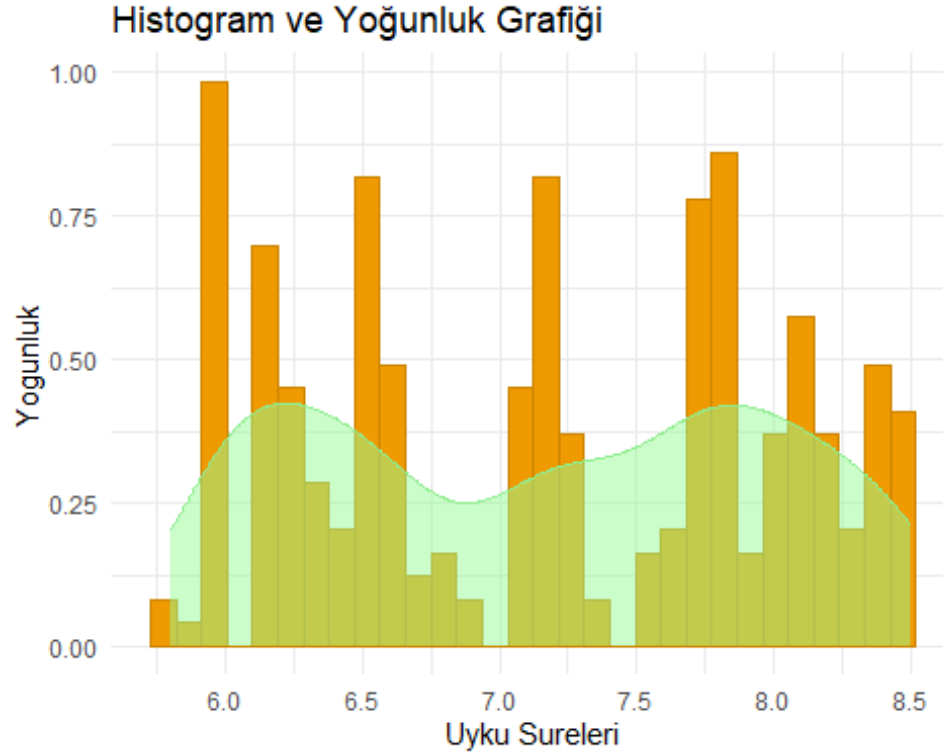
DK_Sleep.Duration <- (sd(train$Sleep.Duration) /
mean(train$Sleep.Duration))*100
DK_Sleep.Duration

## [1] 11.52095

# Yoğunluk ile birlikte histogram grafiği;
library(ggplot2)
ggplot(train, aes(x = Sleep.Duration)) +
  geom_histogram(aes(y = ..density..), fill = "orange2", color = "orange3",
bins = 30) +
  geom_density(alpha = 0.5, fill = "palegreen", color= "palegreen2") +
  labs(title = "Histogram ve Yoğunluk Grafiği", x = "Uyku Sureleri", y =
"Yogunluk") +
  theme_minimal()

## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2
3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```



Histogram grafiğinin daha doğru analiz edilebilmesi amacı ile bir alt satırda bulunan çarpıklık, basıklık ve değişim katsayısı hesaplamaları yapılmıştır. Kısaca özetlemek gerekirse;

- Çarpıklığın “-0.018” gelmesi ile veri setinin sağa çarpık bir dağılım gösterdiğini fakat değerin sıfıra oldukça yakın olması nedeniyle dağılımın simetrik kabul edilebileceğini,
- Basıklığın “-1.40” gelmesi değişkenin dağılımının düzleştirilmiş, veri kümesindeki değerlerin normal dağılıma göre daha düşük bir zirveye sahip olduğunu ve genelde daha yaygın olarak dağıldığını,
- Değişim katsayısının “11.52” gelmesi ise ele alınan değişkenin ortalamasına göre %11.52 oranında değişkenlik gösterdiğini, ortalamasına kıyasla standart sapmasının (aynı zamanda varyansının) daha yüksek olduğunu söylemektedir.

STRES SEVİYESİNDE UYKU SÜRESİNİN YAYILIMINI İNCELEME

```
std_dk <- function(x)
  {c(std <-sd(x), dk <-(sd(x)/mean(x))*100)}
tapply(train$Sleep.Duration, train$Stress.Level, std_dk)

## $`1`
## NULL
##
## $`2`
## NULL
##
## $`3`
```

```
## [1] 0.2015968 2.4471772
##
## $`4`
## [1] 0.4511808 6.4408392
##
## $`5`
## [1] 0.2962349 3.9497984
##
## $`6`
## [1] 0.5240755 7.0264404
##
## $`7`
## [1] 0.1590693 2.4483609
##
## $`8`
## [1] 0.09375283 1.54815721
##
## $`9`
## NULL
##
## $`10`
## NULL
```

1 ile 10 arasında puanlanmış olan stres seviyelerine göre uyku sürelerinin yayılımı incelendiğinde, veri setinde stres seviyelerinin 3 ile 8 puan aralığında olduğunu; 1,2,9 ve 10 seviyeleri için yeterli bir bilgiye sahip olunmadığı gözlemlenmiştir. Örneğin stres seviyesi 4 olan bireylerde uyku süresinin standart sapması “0.45” ve değişim katsayısı “6.44” gelirken; stres seviyesi 8 olan bireylerde uyku süresinin standart sapması “0.093” ve değişim katsayısı “1.54” gelmiştir. Bu durum, stres seviyesi 8 seviyesinde olan grup içinde değişimin diğer seviyelere göre de en az olduğunu göstermektedir.

Genel olarak stres seviyesinin artması ile uyku süresinin azalacağı düşünülmektedir. Fakat veri setinden hareketle yapılmış olan hesaplamalar sonucunda bu konuda net bir bilgiye ulaşılamamış, gruplara göre standart sapma ve değişim katsayısı oranları farklılık göstermiştir.

```
# Geometrik ortalama hesabı;
library("psych")
geometric.mean(train$Sleep.Duration)

## [1] 7.122402

# Kesilmiş ortalama ve kalan gözlemlerin hesabı;
p <- 0.1
mean(train$Sleep.Duration, trim = p)

## [1] 7.162381

n <- nrow(train)
kgs <- n-(as.integer(2*p*n))
kgs
```

```
## [1] 210
```

Kırılmış ortalama, verinin dağılımının ne kadar sağlam olduğunu anlamamıza yardımcı olmakta, daha az etkilenmiş bir merkezi eğilim ölçüsü sağlamaktadır.

İlk olarak yapılmış olan işlemde, uyku süresinin geometrik ortalamasına bakılmış ve veri setindeki ortalama uyku süresi 7.12 saat olarak bulunmuştur. Ardından uyku sürelerinin kesilmiş ortalamaları hesaplanmış ve sonuç 7.16 gelmiştir. Bu durumda uç değerlerin ortalamayı önemli ölçüde değiştirmediğini söylemek mümkündür.

İkinci olarak ise veri setinin kırılması sonucunda kalan gözlem sayısı hesaplanmıştır. Train veri setinde normal şartlarda 262 gözlem varken, işlem sonrasında 210 gözlem üzerinden uyku sürelerinin ortalama hesabının yapılmış olduğu görülmektedir.

VKI İLE UYKU RAHATSIZLIĞI İÇİN ÇAPRAZ TABLO İNCELEMESİ

Çapraz tablolar kategorik veriler arasındaki ilişkileri incelemek için tasarlanmıştır;

```
tablo <- table(train$Sleep.Disorder,train$BMI.Category)
prop.table(tablo,2)
```

```
##
##           Normal      Obese Overweight
##  None          0.93421053 0.00000000 0.11214953
## Sleep Apnea    0.03289474 0.66666667 0.47663551
##  Insomnia      0.03289474 0.33333333 0.41121495
```

```
round(100*prop.table(tablo,2), 2)
```

```
##
##           Normal Obese Overweight
##  None          93.42  0.00      11.21
## Sleep Apnea    3.29 66.67      47.66
##  Insomnia      3.29 33.33      41.12
```

```
addmargins(round(prop.table(tablo,2), 2),1)
```

```
##
##           Normal Obese Overweight
##  None          0.93  0.00      0.11
## Sleep Apnea    0.03  0.67      0.48
##  Insomnia      0.03  0.33      0.41
##  Sum           0.99  1.00      1.00
```

Vücut kitle indeksi ile uyku rahatsızlığı arasındaki ilişkiler çapraz tablo yardımıyla incelenmiştir. Örneğin, “Normal” kategorisindeki bireylerin % 93.42’si ve “Aşırı Kilolu” (Overweight) kategorisinde bulunan bireylerin % 11.21’i uyku bozukluğu yaşamamaktadır. “Aşırı Kilolu” kategorisinde geriye kalan gözlemlerin %47.66’sı uyku apnesi, %41.12’si uykusuzluk sorunu yaşamaktadır. Obez” (Obese) kategorisindeki bireylerin ise % 66.7’si uyku apnesi ve % 33.3’ü uykusuzluk sorunu yaşamakta olup, bu grupta herhangi bir rahatsızlığa sahip olmayan gözleme rastlanmamıştır.

Bu sonuçlar; bireylerin kilosunun normal seviyelerde olması durumunda uyku bozukluğu riskini azaltabileceğini, obezitenin uyku apnesi ile güçlü bir ilişkisi olabileceğini ve aşırı kilonun hem uyku apnesi hem de uykusuzluk riskini taşıdığını göstermektedir.

```
library(DescTools)
```

```
##  
## Attaching package: 'DescTools'  
  
## The following objects are masked from 'package:psych':  
##  
##     AUC, ICC, SD  
  
## The following objects are masked from 'package:Hmisc':  
##  
##     %nin%, Label, Mean, Quantile
```

```
Assocs(tablo)[1:3,1]
```

## Contingency Coeff.	Cramer V	Kendall Tau-b
## 0.6381234	0.5860520	0.7540547

Yapılmış olan inceleme sonucunda, uyku rahatsızlıkları ve vücut kitle indeksi kategorileri arasında önemli bir ilişki olduğu söylenebilmektedir. Özellikle Kendall's Tau-b değeri, sıralı verilerde 0.75 ile güçlü bir ilişki olduğunu belirtir.

```
library("gplots")
```

```
## Registered S3 method overwritten by 'gplots':  
##   method      from  
##   reorder.factor DescTools  
  
##  
## Attaching package: 'gplots'  
  
## The following object is masked from 'package:DescTools':  
##  
##     reorder.factor  
  
## The following object is masked from 'package:stats':  
##  
##     lowess
```

```
balloonplot(t(tablo), main = "VKI & Uyku Rahatsızlıkları", xlab = "VKI",  
            ylab = "Rahatsızlık",  
            label = TRUE, show.margins = FALSE)
```


VKI & Uyku Rahatsızlıkları

Rahatsızlık	VKI		
	Normal	Obese	Overweigh
None	142		12
Sleep Apnea	5	2	51
Insomnia	5	1	44

Son olarak elde edilmiş olan çapraz tablo grafiğinde kategoriler arası ilişkiler daha net anlaşılmaktadır. En yüksek frekansa sahip olan grup, 142 ile normal kilo ve rahatsızlığı olmayan bireyleri kapsamaktadır. Veri seti içerisinde obez olup uykusuzluk yaşayan sadece 1 gözleme rastlanılmıştır.

BALON GRAFİĞİ

```
tablo2 <- xtabs(~ Sleep.Disorder+BMI.Category+Gender+Occupation, data=train)
tablo2
```

```
## , , Gender = Female, Occupation = Accountant
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           13     0         0
## Sleep Apnea       0     0         0
##   Insomnia        0     0         4
##
## , , Gender = Male, Occupation = Accountant
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           1     0         0
## Sleep Apnea       0     0         0
##   Insomnia        0     0         0
##
## , , Gender = Female, Occupation = Doctor
##
```

```

##          BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           2     0         0
##   Sleep Apnea     0     0         0
##   Insomnia        0     0         0
##
## , , Gender = Male, Occupation = Doctor
##
##          BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           46     0         0
##   Sleep Apnea     1     1         0
##   Insomnia        1     1         0
##
## , , Gender = Female, Occupation = Engineer
##
##          BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           26     0         0
##   Sleep Apnea     0     0         0
##   Insomnia        1     0         0
##
## , , Gender = Male, Occupation = Engineer
##
##          BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           18     0         0
##   Sleep Apnea     1     0         0
##   Insomnia        2     0         1
##
## , , Gender = Female, Occupation = Lawyer
##
##          BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           0     0         0
##   Sleep Apnea     0     0         0
##   Insomnia        0     0         0
##
## , , Gender = Male, Occupation = Lawyer
##
##          BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           29     0         2
##   Sleep Apnea     1     0         0
##   Insomnia        1     0         0
##
## , , Gender = Female, Occupation = Manager
##
##          BMI.Category
## Sleep.Disorder Normal Obese Overweight

```

```

##      None      0      0      1
##      Sleep Apnea      0      0      0
##      Insomnia      0      0      0
##
## , , Gender = Male, Occupation = Manager
##
##              BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      0      0
##      Insomnia      0      0      0
##
## , , Gender = Female, Occupation = Nurse
##
##              BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      2      0      5
##      Sleep Apnea      2      0      45
##      Insomnia      0      0      1
##
## , , Gender = Male, Occupation = Nurse
##
##              BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      0      0
##      Insomnia      0      0      0
##
## , , Gender = Female, Occupation = Sales Representative
##
##              BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      0      0
##      Insomnia      0      0      0
##
## , , Gender = Male, Occupation = Sales Representative
##
##              BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      1      0
##      Insomnia      0      0      0
##
## , , Gender = Female, Occupation = Salesperson
##
##              BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      0      0

```

```

##      Insomnia      0      0      0
##
## , , Gender = Male, Occupation = Salesperson
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      0      1
##      Insomnia      0      0      20
##
## , , Gender = Female, Occupation = Scientist
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      2
##      Sleep Apnea      0      0      2
##      Insomnia      0      0      0
##
## , , Gender = Male, Occupation = Scientist
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      0      0
##      Insomnia      0      0      0
##
## , , Gender = Female, Occupation = Software Engineer
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      0      0      0
##      Sleep Apnea      0      0      0
##      Insomnia      0      0      0
##
## , , Gender = Male, Occupation = Software Engineer
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      2      0      0
##      Sleep Apnea      0      0      0
##      Insomnia      0      0      0
##
## , , Gender = Female, Occupation = Teacher
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##      None      3      0      1
##      Sleep Apnea      0      0      2
##      Insomnia      0      0      17
##

```

```
## , , Gender = Male, Occupation = Teacher
##
##           BMI.Category
## Sleep.Disorder Normal Obese Overweight
##   None           0     0         1
##   Sleep Apnea    0     0         1
##   Insomnia       0     0         1

tablo2_df <- as.data.frame(ftable(tablo2))
tablo2_df
```

	Sleep.Disorder	BMI.Category	Gender	Occupation	Freq
## 1	None	Normal	Female	Accountant	13
## 2	Sleep Apnea	Normal	Female	Accountant	0
## 3	Insomnia	Normal	Female	Accountant	0
## 4	None	Obese	Female	Accountant	0
## 5	Sleep Apnea	Obese	Female	Accountant	0
## 6	Insomnia	Obese	Female	Accountant	0
## 7	None	Overweight	Female	Accountant	0
## 8	Sleep Apnea	Overweight	Female	Accountant	0
## 9	Insomnia	Overweight	Female	Accountant	4
## 10	None	Normal	Male	Accountant	1
## 11	Sleep Apnea	Normal	Male	Accountant	0
## 12	Insomnia	Normal	Male	Accountant	0
## 13	None	Obese	Male	Accountant	0
## 14	Sleep Apnea	Obese	Male	Accountant	0
## 15	Insomnia	Obese	Male	Accountant	0
## 16	None	Overweight	Male	Accountant	0
## 17	Sleep Apnea	Overweight	Male	Accountant	0
## 18	Insomnia	Overweight	Male	Accountant	0
## 19	None	Normal	Female	Doctor	2
## 20	Sleep Apnea	Normal	Female	Doctor	0
## 21	Insomnia	Normal	Female	Doctor	0
## 22	None	Obese	Female	Doctor	0
## 23	Sleep Apnea	Obese	Female	Doctor	0
## 24	Insomnia	Obese	Female	Doctor	0
## 25	None	Overweight	Female	Doctor	0
## 26	Sleep Apnea	Overweight	Female	Doctor	0
## 27	Insomnia	Overweight	Female	Doctor	0
## 28	None	Normal	Male	Doctor	46
## 29	Sleep Apnea	Normal	Male	Doctor	1
## 30	Insomnia	Normal	Male	Doctor	1
## 31	None	Obese	Male	Doctor	0
## 32	Sleep Apnea	Obese	Male	Doctor	1
## 33	Insomnia	Obese	Male	Doctor	1
## 34	None	Overweight	Male	Doctor	0
## 35	Sleep Apnea	Overweight	Male	Doctor	0
## 36	Insomnia	Overweight	Male	Doctor	0
## 37	None	Normal	Female	Engineer	26
## 38	Sleep Apnea	Normal	Female	Engineer	0

## 39	Insomnia	Normal	Female	Engineer	1
## 40	None	Obese	Female	Engineer	0
## 41	Sleep Apnea	Obese	Female	Engineer	0
## 42	Insomnia	Obese	Female	Engineer	0
## 43	None	Overweight	Female	Engineer	0
## 44	Sleep Apnea	Overweight	Female	Engineer	0
## 45	Insomnia	Overweight	Female	Engineer	0
## 46	None	Normal	Male	Engineer	18
## 47	Sleep Apnea	Normal	Male	Engineer	1
## 48	Insomnia	Normal	Male	Engineer	2
## 49	None	Obese	Male	Engineer	0
## 50	Sleep Apnea	Obese	Male	Engineer	0
## 51	Insomnia	Obese	Male	Engineer	0
## 52	None	Overweight	Male	Engineer	0
## 53	Sleep Apnea	Overweight	Male	Engineer	0
## 54	Insomnia	Overweight	Male	Engineer	1
## 55	None	Normal	Female	Lawyer	0
## 56	Sleep Apnea	Normal	Female	Lawyer	0
## 57	Insomnia	Normal	Female	Lawyer	0
## 58	None	Obese	Female	Lawyer	0
## 59	Sleep Apnea	Obese	Female	Lawyer	0
## 60	Insomnia	Obese	Female	Lawyer	0
## 61	None	Overweight	Female	Lawyer	0
## 62	Sleep Apnea	Overweight	Female	Lawyer	0
## 63	Insomnia	Overweight	Female	Lawyer	0
## 64	None	Normal	Male	Lawyer	29
## 65	Sleep Apnea	Normal	Male	Lawyer	1
## 66	Insomnia	Normal	Male	Lawyer	1
## 67	None	Obese	Male	Lawyer	0
## 68	Sleep Apnea	Obese	Male	Lawyer	0
## 69	Insomnia	Obese	Male	Lawyer	0
## 70	None	Overweight	Male	Lawyer	2
## 71	Sleep Apnea	Overweight	Male	Lawyer	0
## 72	Insomnia	Overweight	Male	Lawyer	0
## 73	None	Normal	Female	Manager	0
## 74	Sleep Apnea	Normal	Female	Manager	0
## 75	Insomnia	Normal	Female	Manager	0
## 76	None	Obese	Female	Manager	0
## 77	Sleep Apnea	Obese	Female	Manager	0
## 78	Insomnia	Obese	Female	Manager	0
## 79	None	Overweight	Female	Manager	1
## 80	Sleep Apnea	Overweight	Female	Manager	0
## 81	Insomnia	Overweight	Female	Manager	0
## 82	None	Normal	Male	Manager	0
## 83	Sleep Apnea	Normal	Male	Manager	0
## 84	Insomnia	Normal	Male	Manager	0
## 85	None	Obese	Male	Manager	0
## 86	Sleep Apnea	Obese	Male	Manager	0
## 87	Insomnia	Obese	Male	Manager	0
## 88	None	Overweight	Male	Manager	0

## 89	Sleep Apnea	Overweight	Male		Manager	0
## 90	Insomnia	Overweight	Male		Manager	0
## 91	None	Normal	Female		Nurse	2
## 92	Sleep Apnea	Normal	Female		Nurse	2
## 93	Insomnia	Normal	Female		Nurse	0
## 94	None	Obese	Female		Nurse	0
## 95	Sleep Apnea	Obese	Female		Nurse	0
## 96	Insomnia	Obese	Female		Nurse	0
## 97	None	Overweight	Female		Nurse	5
## 98	Sleep Apnea	Overweight	Female		Nurse	45
## 99	Insomnia	Overweight	Female		Nurse	1
## 100	None	Normal	Male		Nurse	0
## 101	Sleep Apnea	Normal	Male		Nurse	0
## 102	Insomnia	Normal	Male		Nurse	0
## 103	None	Obese	Male		Nurse	0
## 104	Sleep Apnea	Obese	Male		Nurse	0
## 105	Insomnia	Obese	Male		Nurse	0
## 106	None	Overweight	Male		Nurse	0
## 107	Sleep Apnea	Overweight	Male		Nurse	0
## 108	Insomnia	Overweight	Male		Nurse	0
## 109	None	Normal	Female	Sales	Representative	0
## 110	Sleep Apnea	Normal	Female	Sales	Representative	0
## 111	Insomnia	Normal	Female	Sales	Representative	0
## 112	None	Obese	Female	Sales	Representative	0
## 113	Sleep Apnea	Obese	Female	Sales	Representative	0
## 114	Insomnia	Obese	Female	Sales	Representative	0
## 115	None	Overweight	Female	Sales	Representative	0
## 116	Sleep Apnea	Overweight	Female	Sales	Representative	0
## 117	Insomnia	Overweight	Female	Sales	Representative	0
## 118	None	Normal	Male	Sales	Representative	0
## 119	Sleep Apnea	Normal	Male	Sales	Representative	0
## 120	Insomnia	Normal	Male	Sales	Representative	0
## 121	None	Obese	Male	Sales	Representative	0
## 122	Sleep Apnea	Obese	Male	Sales	Representative	1
## 123	Insomnia	Obese	Male	Sales	Representative	0
## 124	None	Overweight	Male	Sales	Representative	0
## 125	Sleep Apnea	Overweight	Male	Sales	Representative	0
## 126	Insomnia	Overweight	Male	Sales	Representative	0
## 127	None	Normal	Female		Salesperson	0
## 128	Sleep Apnea	Normal	Female		Salesperson	0
## 129	Insomnia	Normal	Female		Salesperson	0
## 130	None	Obese	Female		Salesperson	0
## 131	Sleep Apnea	Obese	Female		Salesperson	0
## 132	Insomnia	Obese	Female		Salesperson	0
## 133	None	Overweight	Female		Salesperson	0
## 134	Sleep Apnea	Overweight	Female		Salesperson	0
## 135	Insomnia	Overweight	Female		Salesperson	0
## 136	None	Normal	Male		Salesperson	0
## 137	Sleep Apnea	Normal	Male		Salesperson	0
## 138	Insomnia	Normal	Male		Salesperson	0

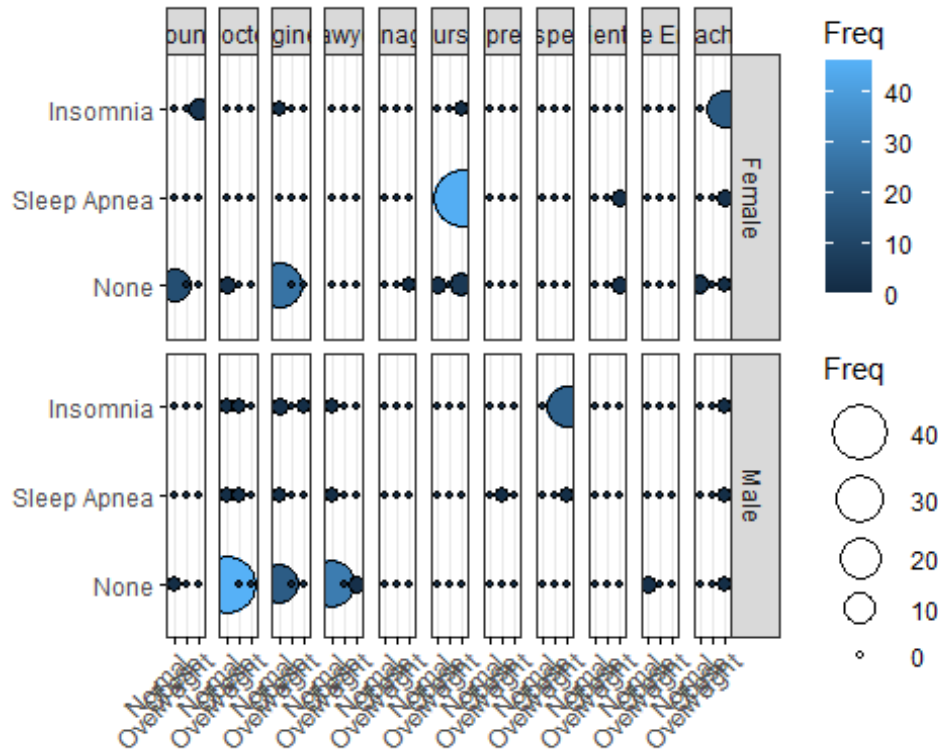
## 139	None	Obese	Male	Salesperson	0
## 140	Sleep Apnea	Obese	Male	Salesperson	0
## 141	Insomnia	Obese	Male	Salesperson	0
## 142	None	Overweight	Male	Salesperson	0
## 143	Sleep Apnea	Overweight	Male	Salesperson	1
## 144	Insomnia	Overweight	Male	Salesperson	20
## 145	None	Normal	Female	Scientist	0
## 146	Sleep Apnea	Normal	Female	Scientist	0
## 147	Insomnia	Normal	Female	Scientist	0
## 148	None	Obese	Female	Scientist	0
## 149	Sleep Apnea	Obese	Female	Scientist	0
## 150	Insomnia	Obese	Female	Scientist	0
## 151	None	Overweight	Female	Scientist	2
## 152	Sleep Apnea	Overweight	Female	Scientist	2
## 153	Insomnia	Overweight	Female	Scientist	0
## 154	None	Normal	Male	Scientist	0
## 155	Sleep Apnea	Normal	Male	Scientist	0
## 156	Insomnia	Normal	Male	Scientist	0
## 157	None	Obese	Male	Scientist	0
## 158	Sleep Apnea	Obese	Male	Scientist	0
## 159	Insomnia	Obese	Male	Scientist	0
## 160	None	Overweight	Male	Scientist	0
## 161	Sleep Apnea	Overweight	Male	Scientist	0
## 162	Insomnia	Overweight	Male	Scientist	0
## 163	None	Normal	Female	Software Engineer	0
## 164	Sleep Apnea	Normal	Female	Software Engineer	0
## 165	Insomnia	Normal	Female	Software Engineer	0
## 166	None	Obese	Female	Software Engineer	0
## 167	Sleep Apnea	Obese	Female	Software Engineer	0
## 168	Insomnia	Obese	Female	Software Engineer	0
## 169	None	Overweight	Female	Software Engineer	0
## 170	Sleep Apnea	Overweight	Female	Software Engineer	0
## 171	Insomnia	Overweight	Female	Software Engineer	0
## 172	None	Normal	Male	Software Engineer	2
## 173	Sleep Apnea	Normal	Male	Software Engineer	0
## 174	Insomnia	Normal	Male	Software Engineer	0
## 175	None	Obese	Male	Software Engineer	0
## 176	Sleep Apnea	Obese	Male	Software Engineer	0
## 177	Insomnia	Obese	Male	Software Engineer	0
## 178	None	Overweight	Male	Software Engineer	0
## 179	Sleep Apnea	Overweight	Male	Software Engineer	0
## 180	Insomnia	Overweight	Male	Software Engineer	0
## 181	None	Normal	Female	Teacher	3
## 182	Sleep Apnea	Normal	Female	Teacher	0
## 183	Insomnia	Normal	Female	Teacher	0
## 184	None	Obese	Female	Teacher	0
## 185	Sleep Apnea	Obese	Female	Teacher	0
## 186	Insomnia	Obese	Female	Teacher	0
## 187	None	Overweight	Female	Teacher	1
## 188	Sleep Apnea	Overweight	Female	Teacher	2

## 189	Insomnia	Overweight	Female	Teacher	17
## 190	None	Normal	Male	Teacher	0
## 191	Sleep Apnea	Normal	Male	Teacher	0
## 192	Insomnia	Normal	Male	Teacher	0
## 193	None	Obese	Male	Teacher	0
## 194	Sleep Apnea	Obese	Male	Teacher	0
## 195	Insomnia	Obese	Male	Teacher	0
## 196	None	Overweight	Male	Teacher	1
## 197	Sleep Apnea	Overweight	Male	Teacher	1
## 198	Insomnia	Overweight	Male	Teacher	1

Burada yazılan kodlar ile “Uyku Rahatsızlığı”, “VKI”, “Cinsiyet” ve “Meslek” kategorileri ile farklı bir veri seti oluşturulmuştur.

Oluşturulan veri seti sayesinde birbiriyle ilişkilendirilmiş her bir grup için frekans değerleri gözlemlenmiştir. Örneğin veri seti içerisinde uyku bozukluğu yaşamayan, kilosu normal, cinsiyeti erkek ve doktor olan 46 birey; uyku bozukluğu yaşamayan, kilosu normal, cinsiyeti kadın ve mühendis olan 26 birey; uyku apnesi bulunan, aşırı kilolu, cinsiyeti kadın ve hemşire olan 45 birey; uykusuzluk yaşayan, aşırı kilolu, cinsiyeti erkek ve satış personeli olan 20 birey bulunmaktadır.

```
library(ggpubr)
ggballoonplot(
  tablo2_df, x = "BMI.Category", y = "Sleep.Disorder",
  size = "Freq", fill = "Freq",
  facet.by = c("Gender", "Occupation"),
  ggtheme = theme_bw())
```



Elde edilen balon grafiğinde her bir balon, veri çerçevesindeki belirli bir kombinasyonu göstermektedir. Balonların boyutu frekansı temsil ederken, renkler de frekansın bir göstergesi olarak kabul edilmektedir.

Kadın ve erkek olarak iki gruba ayrılmış grafik incelendiğinde; - Kadınlar için en yüksek frekansın aşırı kilolu, uyku apnesine sahip ve mesleği hemşire olan bireylerde - Erkekler için en yüksek frekansın normal kilolu, uyku rahatsızlığı olmayan ve mesleği doktor olan bireylerde olduğu söylenebilir.

```
chisquare <- table(train$BMI.Category, train$Sleep.Disorder)
exp_tablo <- chisq.test(chisquare)$expected

## Warning in chisq.test(chisquare): Chi-squared approximation may be
incorrect

rowcs <- function(i, obs, exp) {
  sum(((obs[i,] - exp[i,])^2)/exp[i,])
}

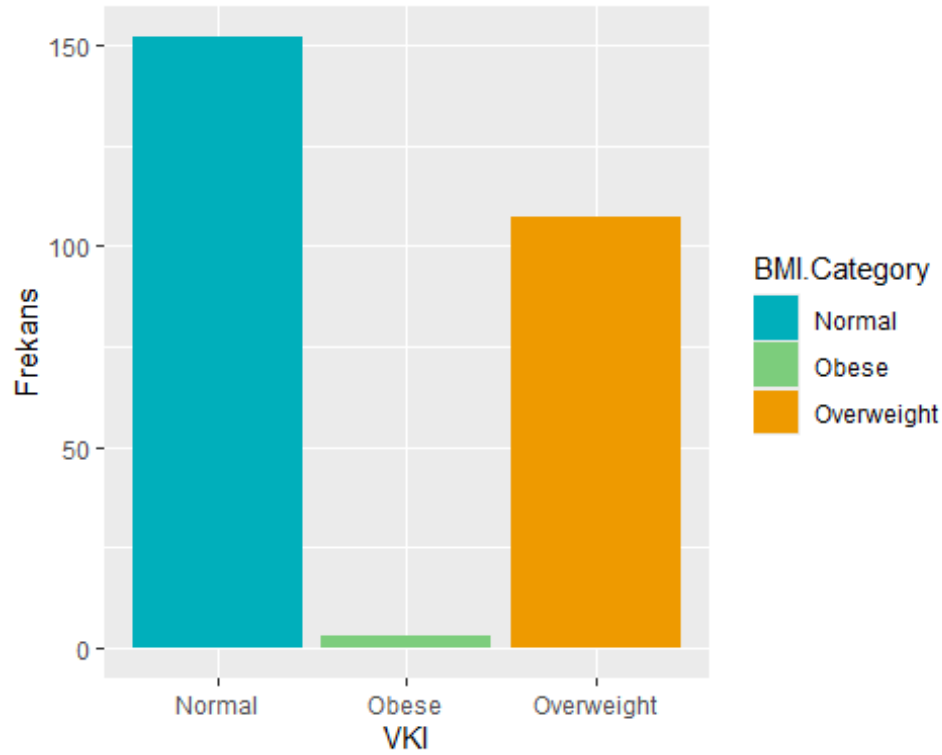
chi_tablo <- as.matrix(lapply(seq_len(nrow(chisquare)), rowcs, obs =
chisquare, exp = exp_tablo))
rownames(chi_tablo) <- rownames(chisquare)
chi_tablo

##           [,1]
## Normal      75.29551
## Obese        4.769655
## Overweight  99.90628
```

Yukarıda yazılan kod yardımı ile VKI ve Uyku Rahatsızlığı için ki-kare testi uygulanmıştır. Test sonucunda VKI kategorileri uyku bozukluklarının dağılımında önemli farklılıklar olduğunu göstermektedir. Özellikle “Normal” (sonuç= 75.29) ve “Aşırı Kilolu” (sonuç= 99.9) kategorileri, uyku bozuklukları açısından beklenen ve gözlemlenen değerler arasında büyük farklılıklar göstermektedir. Bu durum, bu kategorilerdeki bireylerin uyku bozuklukları konusunda daha fazla incelemeye tabi tutulmasını gerektirebilir.

BAR-PLOT GRAFİĞİ

```
library(ggplot2)
ggplot(train, aes(BMI.Category, fill= BMI.Category)) +
  scale_fill_manual(values = c("Normal" = "#00AFBB", "Obese" = "palegreen3",
"Overweight" = "orange2")) +
  xlab("VKI") + ylab("Frekans") +
  geom_bar()
```



```
library(inspectdf)
library(dplyr)
train %>% inspect_types()

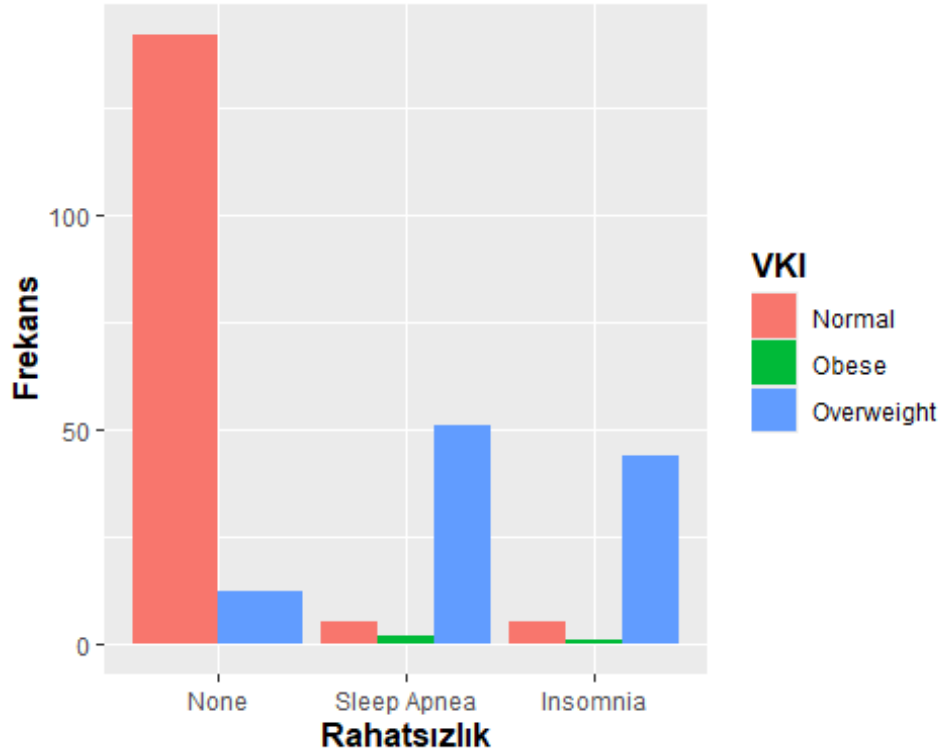
## # A tibble: 5 × 4
##   type          cnt  pcnt col_name
##   <chr>        <int> <dbl> <named list>
## 1 integer           5 38.5  <chr [5]>
## 2 factor           3 23.1  <chr [3]>
## 3 character        2 15.4  <chr [2]>
## 4 ordered factor   2 15.4  <chr [2]>
## 5 numeric          1  7.69 <chr [1]>

tra_cat <- train %>% inspect_cat()
tra_cat$levels$BMI.Category

## # A tibble: 3 × 3
##   value      prop  cnt
##   <chr>    <dbl> <int>
## 1 Normal  0.580    152
## 2 Overweight 0.408    107
## 3 Obese    0.0115     3
```

VKI için genel bir hesaplama yapıldığı takdirde veri setinde bulunan gözlemlerin %58.0'inin normal, %40.8'nin aşırı kilolu ve %1.15'nin obez kategorisi içerisinde olduğu görülmektedir. Bar-plot grafiği ile bu sonuçları görsel olarak incelemek mümkündür.

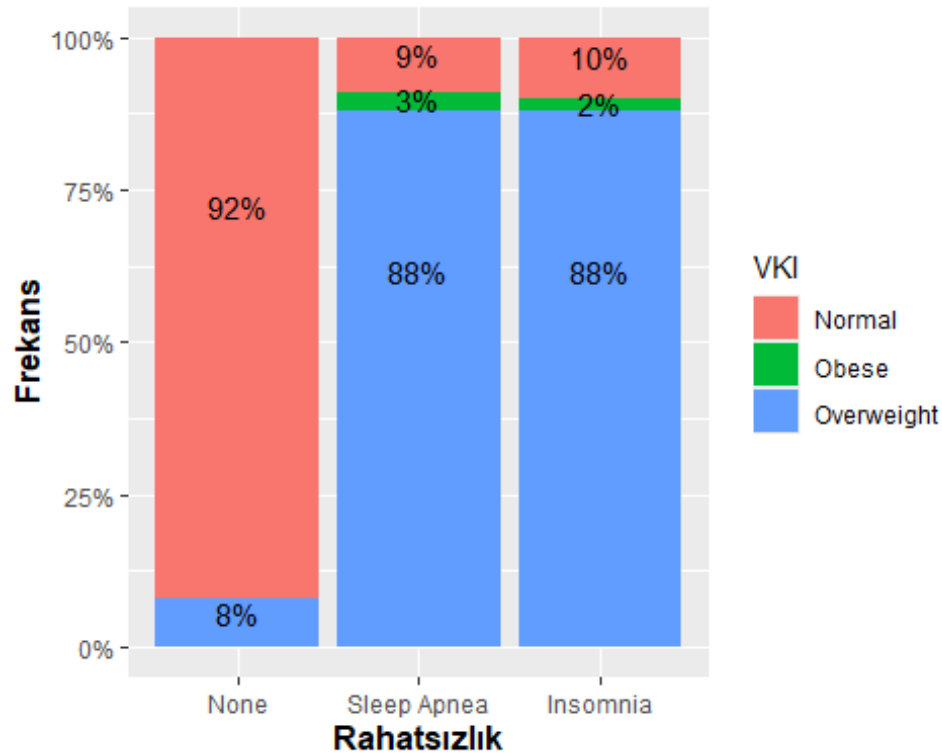
```
library(ggplot2)
ggplot(train, aes(Sleep.Disorder, fill= BMI.Category))+
  geom_bar(position=position_dodge())+
  xlab("Rahatsızlık")+
  ylab("Frekans")+
  scale_fill_discrete(name = "VKI")+
  theme(axis.title.x = element_text(color="black", face="bold", size=12),
        axis.title.y = element_text(color="black", face="bold", size=12),
        legend.title = element_text(colour="black", face="bold", size=12))
```



İki kategorik değişkeni ele alarak oluşturulan bar-plot grafiğinde uyku rahatsızlıkları ile vücut kitle indeksleri incelenmiştir. Herhangi bir rahatsızlığı olmayan bireylerin çoğunlukla normal kiloda olduğu görülmekte olup, bu kategoride obez bireylere rastlanmamaktadır. Aşırı kilolu bireyler için uyku apnesi veya uykusuzluk rahatsızlığına sahip olma durumu birbirine yakın gelmekte, rahatsızlığa sahip olmayan bireylerin bu kategoride daha az sıklıkta dağıldığı görülmektedir.

```
# Yüzdelik şekilde ifade etmek için;
library(dplyr)
tra_pct <- train %>%
  group_by(Sleep.Disorder, BMI.Category) %>%
  summarise(count = n(), .groups = "drop_last") %>%
  mutate(pct = round(count / sum(count), 2))
ggplot(tra_pct, aes(Sleep.Disorder, pct, fill = BMI.Category)) +
  geom_bar(stat='identity') +
  xlab("Rahatsızlık")+
  ylab("Frekans")+
  scale_y_continuous(labels = scales::percent)
```

```
scale_fill_discrete(name = "VKI") +
theme(axis.title.x = element_text(color="black", face="bold", size=12),
      axis.title.y = element_text(color="black", face="bold", size=12)) +
geom_text(aes(label=scales::percent(pct)), position = position_stack(vjust
= .7)) +
scale_y_continuous(labels = scales::percent)
```



Bar-plot grafiğinde görülen sütunları yüzdelik şekilde ifade edebilmek için yukarıdaki işlemler yapılmıştır. İncelendiği takdirde elde edilen değerlerin, yeni ve bir önceki grafik ile doğru şekilde örtüştüğü görülecektir.

PASTA GRAFİĞİ

```
library(plotly)
```

```
##
## Attaching package: 'plotly'

## The following object is masked from 'package:Hmisc':
##
##   subplot

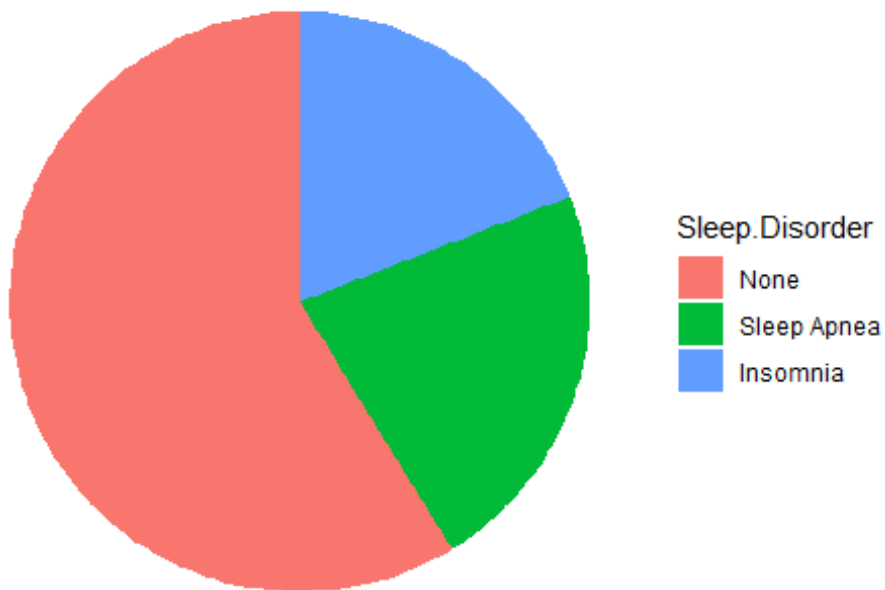
## The following object is masked from 'package:ggplot2':
##
##   last_plot

## The following object is masked from 'package:stats':
##
##   filter
```

```
## The following object is masked from 'package:graphics':
##
##      layout

cross<-as.data.frame(prop.table(table(train$Sleep.Disorder)))
colnames(cross) <- c("Sleep.Disorder", "Freq")
ggplot(cross, aes(x = "", y = Freq, fill = Sleep.Disorder)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  labs(title = "Uyku Rahatsızlıkları Dağılımı") +
  theme_void() +
  theme(legend.position = "right")
```

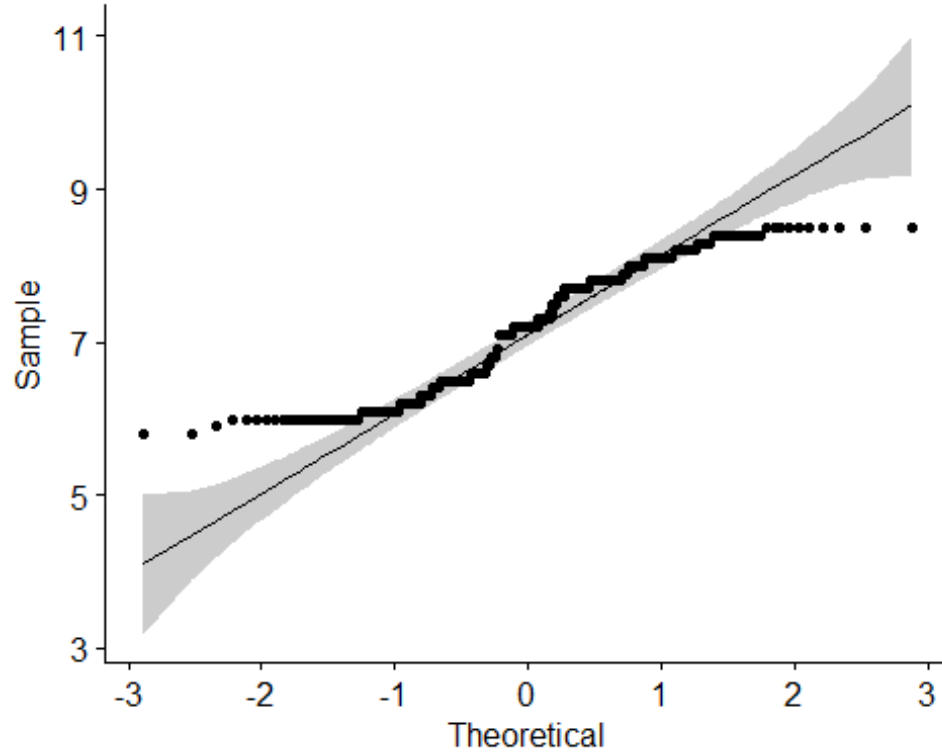
Uyku Rahatsızlıkları Dağılımı



VKI'ne göre yapılan pasta grafiğinde en büyük yüzdeliği % 58.7 ile “None” grubu oluşturmaktadır. Daha sonrasında % 22.1 ile “Sleep Apnea” ve %19.0 ile “Insomnia” grubu oluşturur.

Q-Q PLOT GRAFİĞİ

```
library(ggpubr)
ggqqplot(train$Sleep.Duration)
```

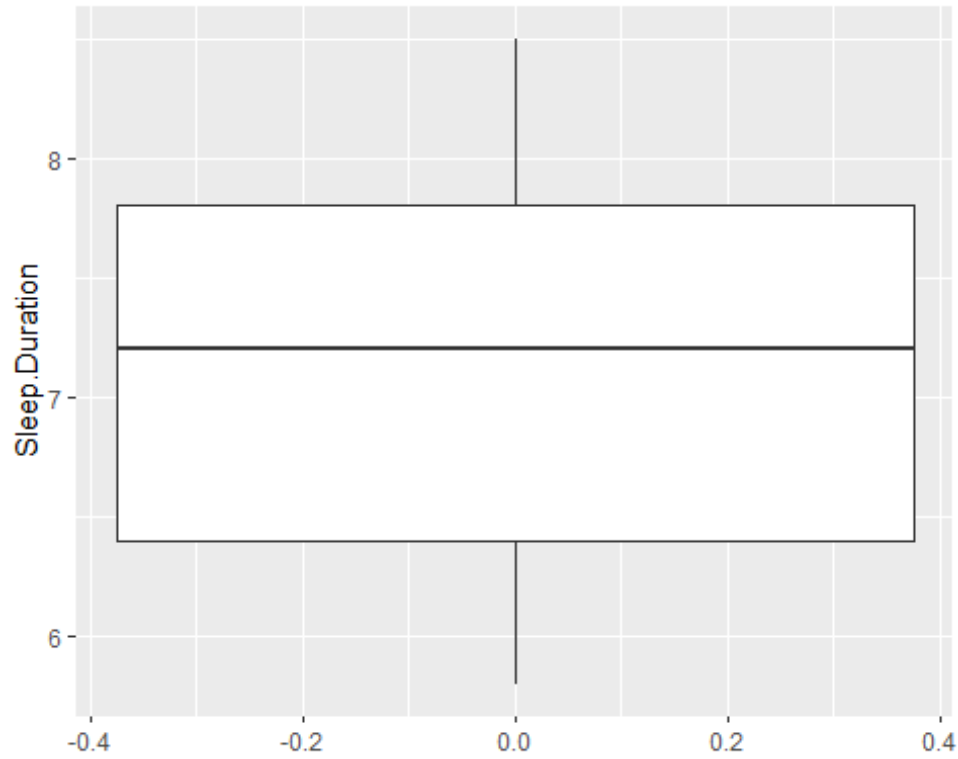


Q-Q plot grafiğinde veri noktalarının teorik çizgi boyunca hizalanma derecesi, verinin dağılımı hakkında önemli bilgiler sağlamaktadır. Uyku süreleri için çizilmiş olan Q-Q plot'ta dağılımın konkav şekilde meydana geldiğini, dolayısıyla sağa çarpık olduğunu söylemek mümkündür. Uç noktalarda belirgin sapmalar görüldüğünden, verilerin kuyruklarının teorik dağılımın kuyruklarından farklı olduğu sonucuna varılmaktadır.

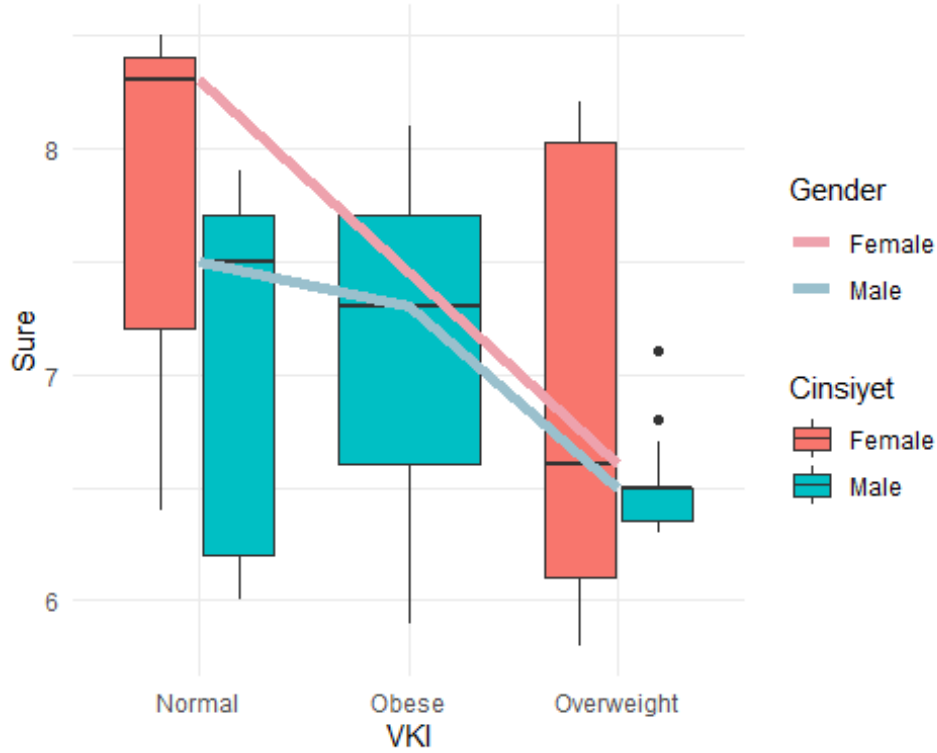
(BURAYA Bİ TEKRAR BAK !!!!!!!!)

VKI DÜZEYİNDE UYKU SÜRELERİNİN DETAYLI İNCELENMESİ

```
ggplot(train, aes(y=Sleep.Duration))+  
  geom_boxplot()
```



```
ggplot(train, aes(x=BMI.Category, y=Sleep.Duration, fill= Gender))+  
  geom_boxplot()+  
  labs(x="VKI", y = "Sure")+  
  scale_fill_discrete(name = "Cinsiyet")+  
  stat_summary(aes(group = Gender, color = Gender),  
               fun = median, geom = "line", size = 2) +  
  scale_color_manual(values = c("Male" = "lightblue3", "Female" =  
"lightpink2")) +  
  theme_minimal()
```

Bu kod ile elde edilmiş olan ilk grafikte dağılımın oldukça simetrik olduğu, aşırı uç değerlerin bulunmadığı ve uyku sürelerinin genel olarak homojen olduğu söylenebilmektedir. Aynı zamanda ortalama uyku süresinin yaklaşık 7 saat olduğu görülmektedir.

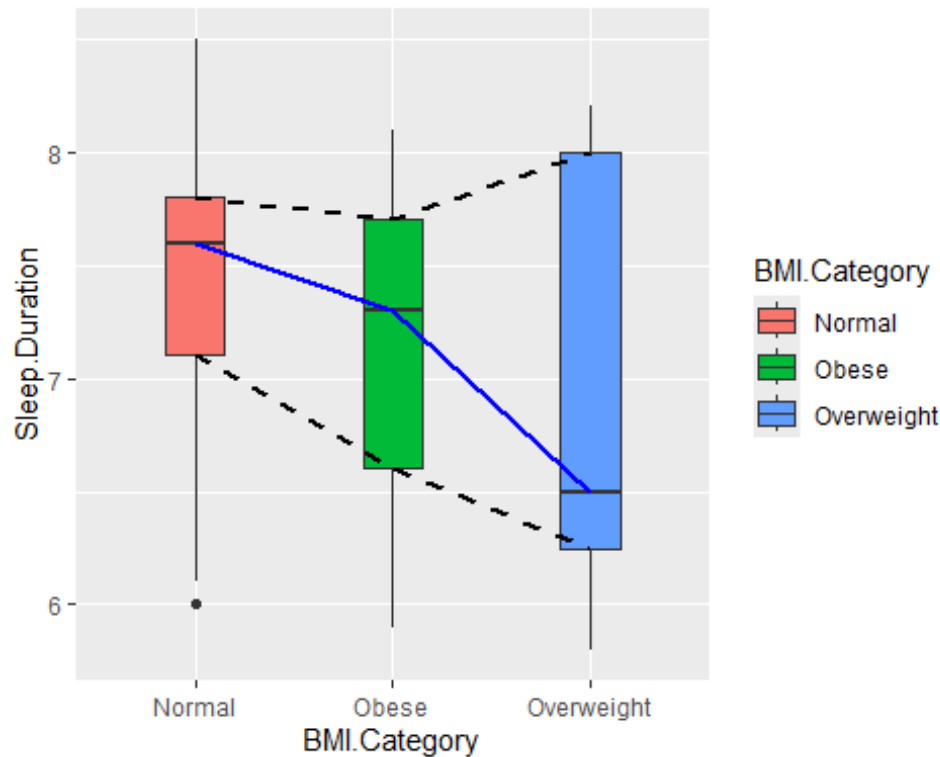
İkinci box-plot grafiğinde ise uyku süreleri VKI'lerine göre kategorize edilmiş ve cinsiyete göre gruplandırılmıştır. Bakıldığında “Obez” kategorisinde herhangi bir kadın bireyin olmadığı, gruplar arası varyansın farklılık gösterdiği ve homojen olmadığı, kilonun arttıkça her iki cinsiyet için de ortalama uyku süresinin azaldığı görülmektedir.

Ek olarak grafikte görüldüğü üzere; iki cinsiyet için çizilen eğrilerin paralel olduğu, dolayısıyla etkileşimin olmadığı söylenebilir.

```
ggplot(train, aes(x=BMI.Category, y=Sleep.Duration, fill=BMI.Category)) +
  geom_boxplot(width=0.3) +
  stat_summary(fun.y = median, geom="line", group= 1, color= "blue", size =
1) +
  stat_summary(geom = 'line', linetype = "dashed", group=1, size = 1, fun.y =
quantile, fun.args = list(probs = 0.25), position = position_dodge(0.75)) +
  stat_summary(geom = 'line', group= 1, linetype = "dashed", size = 1, fun.y =
quantile, fun.args = list(probs = 0.75), position = position_dodge(0.75))

## Warning: The `fun.y` argument of `stat_summary()` is deprecated as of
ggplot2 3.3.0.
## i Please use the `fun` argument instead.
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Bu grafikte ise cinsiyet gruplarından bağımsız olarak sadece VKI'leri için süre dağılımları kodda daha detaya girilerek incelenmiştir. Burada farklı olarak birinci ve üçüncü kartiller de kesikli çizgiler ile belirtilmek istenmiştir. Aşırı kilolu bireylerin bulunduğu grupta değişimin en çok, normal kilolu bireylerin bulunduğu grupta ise değişimin en az olduğu görülmektedir. Benzer şekilde varyanslar gruplara göre farklılık göstermekte ve VKI düzeyi arttıkça ortalama uyku süresi azalmaktadır.

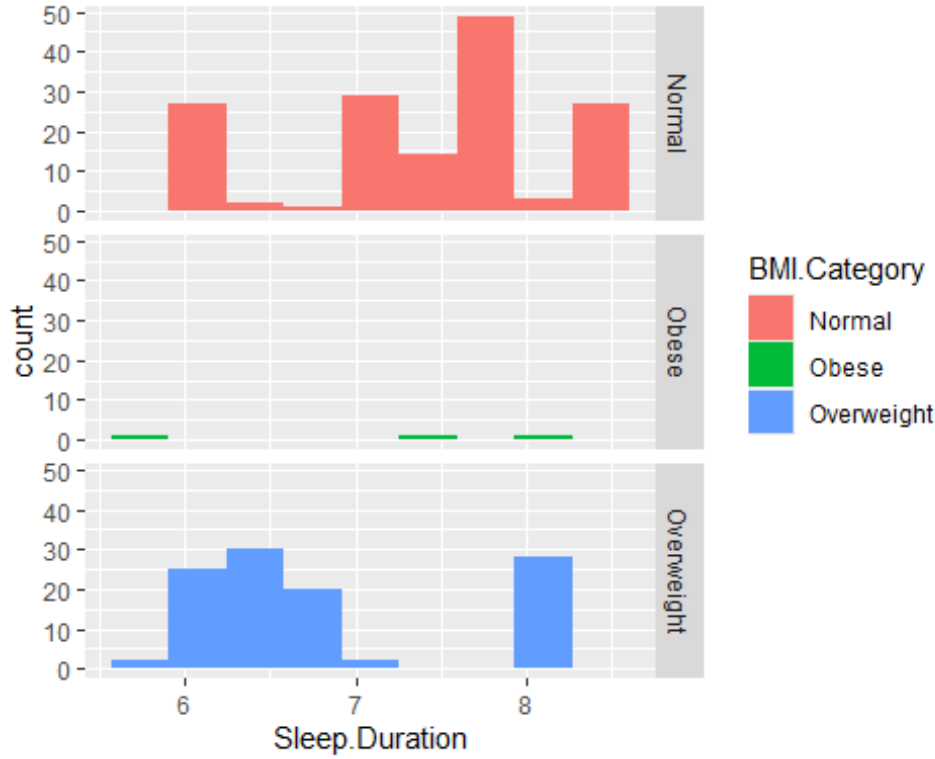
```
k <- ceiling((log(2*nrow(train)))+1)
genislik <- max(train$Sleep.Duration)-min(train$Sleep.Duration)
binw <- genislik/k
genislik

## [1] 2.7

binw

## [1] 0.3375

# Değişken düzeylerinde kırılımı görmek için;
library(ggplot2)
ggplot(train,aes(Sleep.Duration,fill=BMI.Category))+
  geom_histogram(binwidth=binw)+
  facet_grid(BMI.Category~.)
```



Başlangıçta yapılan işlemler neticesinde veri seti için genişlik “2.7” ve binw değeri “0.33” bulunmuştur. Ardından VKI kategorisine bağlı olarak uyku sürelerinin ne kadar fark yaratabileceği incelenmiştir. “Obez” kategorisinde sadece 3 gözlem olup birbiri ile çok farklı değerlere sahip olmaları nedeniyle yorum yapmak hatalı olacaktır. Ancak diğer iki kategori olan “Normal” ve “Aşırı Kilolu” kategorileri için yorum yapılacaktır; - Normal kilolu bireyler için uyku sürelerine ilişkin histogram grafiğinin dağılımının sola çarpık olduğunu, dağılımın geniş ve birden fazla tepe noktasına sahipliği, daha fazla çeşitlilik olduğunu göstermektedir. - Aşırı kilolu bireyler için uyku süresinin genel olarak 6 ile 7 saat arasında dağıldığı görülmektedir. Ancak 8 saat ve üzeri uyuyan ayrı bir grubun bu kategoride farklı bir küme oluşturduğunu söylemek de mümkündür.

UYKU KALİTESİ İLE UYKU SÜRELERİNİN İNCELENMESİ

Saçılım grafiği;

library(tidyverse)

```
## — Attaching core tidyverse packages — tidyverse
2.0.0 —
```

```
## ✓ forcats 1.0.0      ✓ stringr 1.5.1
```

```
## ✓ lubridate 1.9.3    ✓ tibble 3.2.1
```

```
## ✓ purrr 1.0.2       ✓ tidyr 1.3.1
```

```
## — Conflicts —
```

```
tidyverse_conflicts() —
```

```
## ✗ psych::%+__() masks ggplot2::%+__()
```

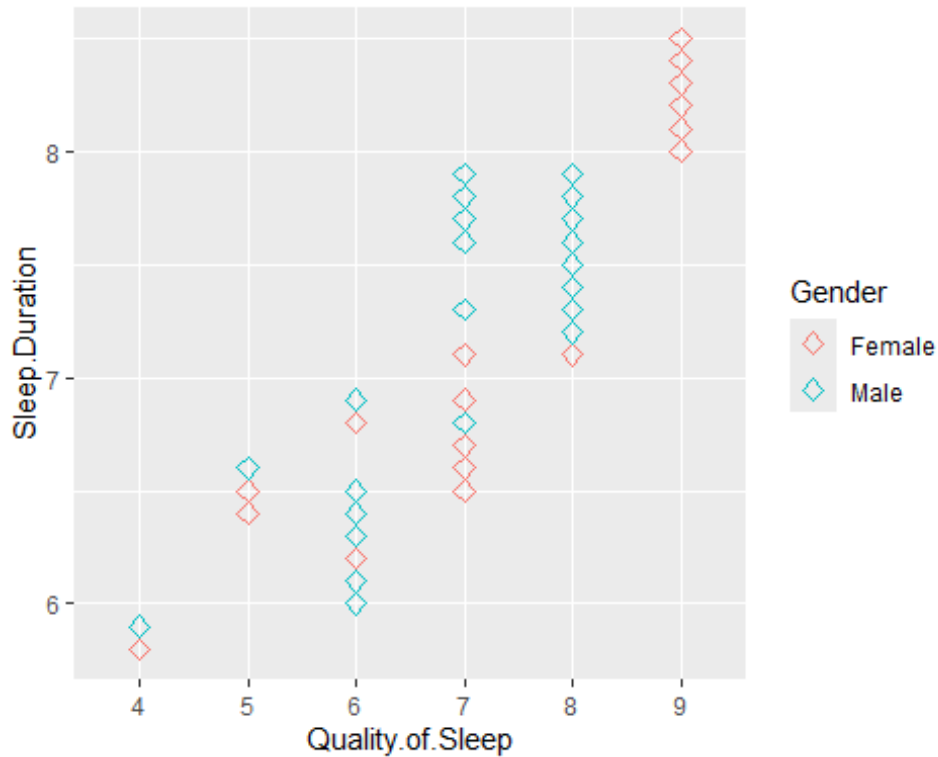
```
## ✗ psych::alpha() masks ggplot2::alpha()
```

```
## ✗ plotly::filter() masks dplyr::filter(), stats::filter()
```

```
## X dplyr::lag()      masks stats::lag()
## X Hmisc::src()      masks dplyr::src()
## X Hmisc::summarize() masks dplyr::summarize()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

ggplot(train, aes(Quality.of.Sleep, Sleep.Duration, color= Gender))+
  geom_point(size=2, shape= 5, stroke=1)+
  geom_smooth(method = "glm", col="black", se = FALSE)

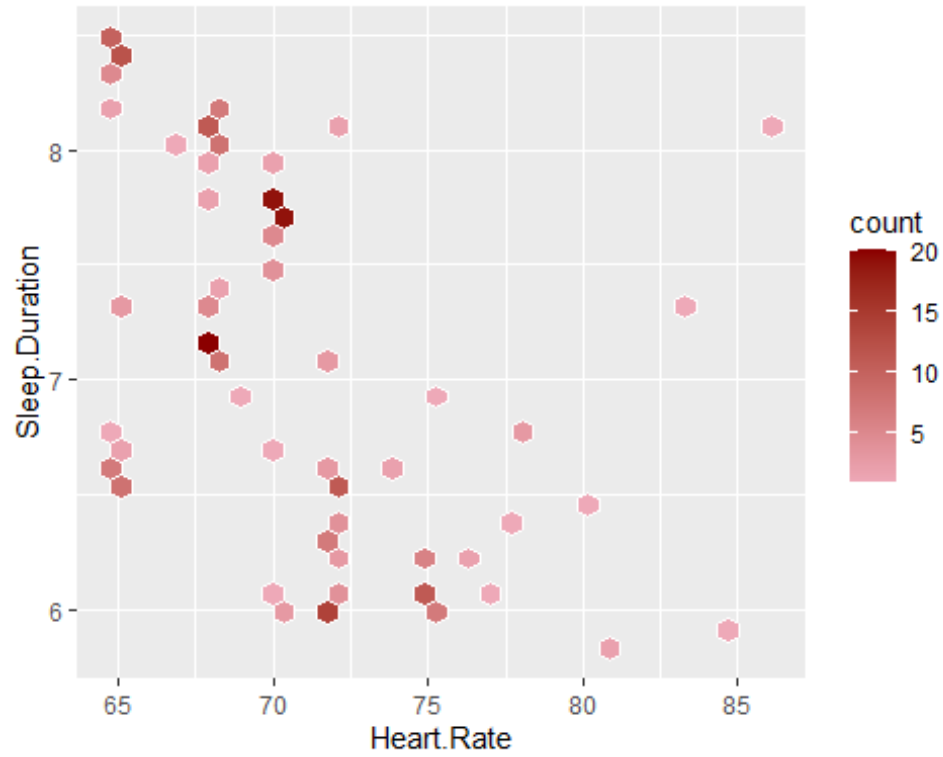
## `geom_smooth()` using formula = 'y ~ x'
```



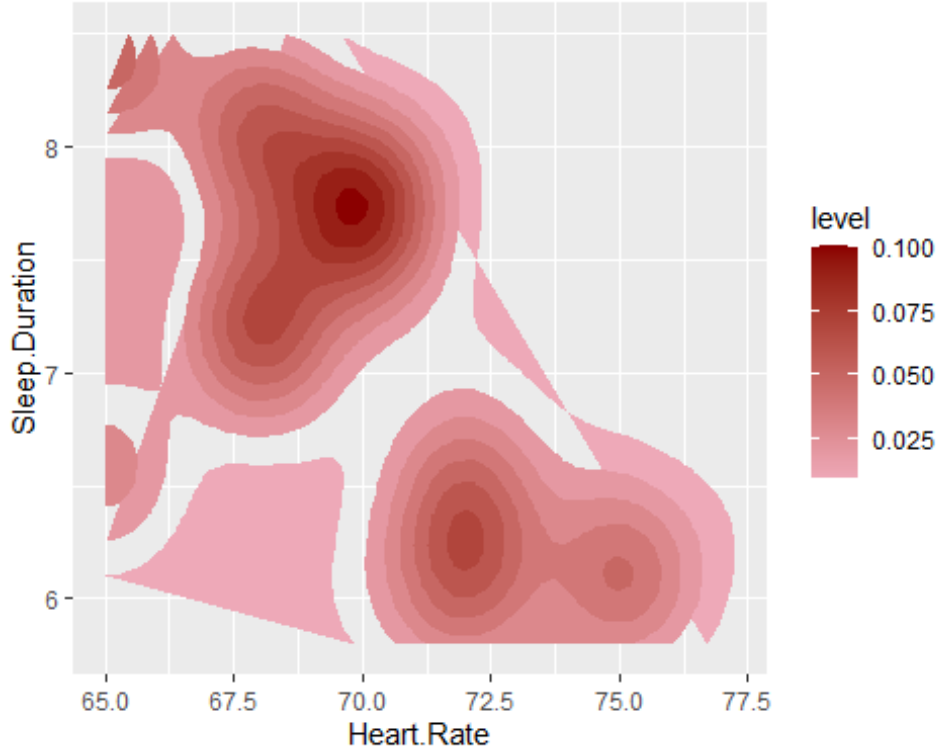
Son satırda bulunan “method = glm” fonksiyonu genelleştirilmiş doğrusal modelleri temsil etmek amacı ile tercih edilmiştir.

Yukarıdaki kod yardımı ile oluşturulan saçılım grafiği cinsiyet bazında iki gruba ayrılmıştır. Grafik incelendiğinde uyku kalitesi ve süresi en yüksek olan bireylerin cinsiyetinin hep kadın olduğu gözlemlenmiştir. Veri setinde uyku süresinin ortalama 7 saat olduğu bilinmesiyle birlikte, kalitenin arttığı noktada sürenin de genel anlamda arttığı söylenebilmektedir.

```
#Altıgen çizimi;
library(hexbin)
ggplot(train, aes(x=Heart.Rate, y=Sleep.Duration))+
  geom_hex(bins=30, color = "white")+
  scale_fill_gradient(low="pink2", high="red4")
```



```
#Kontur çizimi;
library(ggplot2)
ggplot(train, aes(x=Heart.Rate, y=Sleep.Duration) ) +
  stat_density_2d(aes(fill= ..level..), geom = "polygon")+
  scale_fill_gradient(low = "pink2", high = "red4")
```



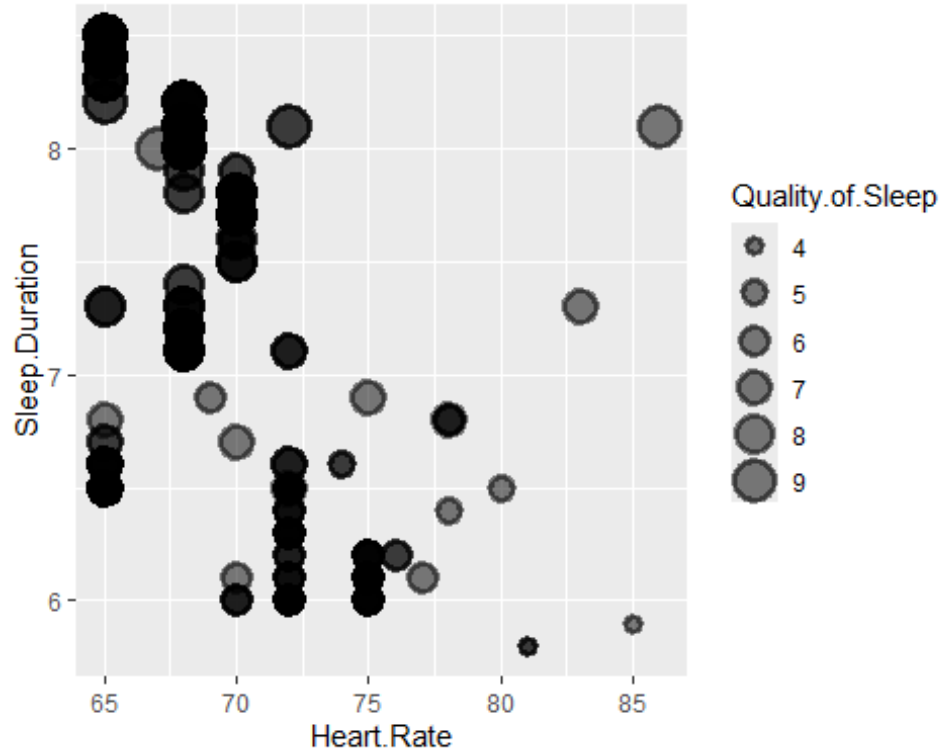
Kalp atış hızı için ortalama 60-100 bpm normal kabul edilmektedir. Ancak bu aralığın üst sınırına yaklaşan değerler, stres, kafein alımı, sağlık durumları (örneğin tiroid problemleri) veya düşük fiziksel kondisyon gibi çeşitli faktörlerden etkilenebilir. Yapılan bu testte kalp atış hızları 65 ile 85 arasında gelmiş olup, yorumlaması da ona göre yapılmıştır.

Kodlar sayesinde elde edilmiş olan grafikler uyku sürelerinin kalp atış oranları ile ilişkisi incelenmiştir. İlk oluşturulan grafikte aykırı değerleri de görmek mümkün iken, ikinci olarak oluşturulan grafikte bu değerlerden kurtularak daha genel ve doğru bir dağılım elde edilmiştir.

Kalp atış hızı 65-75 aralığında olan bireylerin genel olarak uyku sürelerinin daha yüksek olduğu görülmektedir. Fakat özellikle kalp atış hızı 70 iken uyku süresinin de ortalama 8 saat olduğu söylenebilir. Hız 80 ve üzeri seviyelere çıktığı takdirde uyku sürelerinde azalma olacağı beklenmektedir.

Bu iki değişkenin arasındaki ilişki uyku kalitesi ile incelenmek istenirse;

```
# Kabarcık grafiği;  
library(ggplot2)  
ggplot(train, aes(Heart.Rate, Sleep.Duration, size=Quality.of.Sleep)) +  
  geom_point(alpha=0.5, stroke=2)
```



Kabarcık grafiğinde yeniden uyku sürelerinin kalp atış oranları ile ilişkisi incelenmiş olup, farklı olarak uyku kalitesi düzeyinde gruplandırma yapılmıştır.

Grafikten hareketle, kalp atış hızı özellikle 65-75 aralığında olan bireylerin genel olarak uyku kalitelerinin yüksek ve uyku sürelerinin de normal olduğunu söylemek mümkündür. Aykırı değerler dışında, kalp atış hızı arttıkça diğer iki değişkenin de benzer oranda azaldığı söylenebilir.

SAÇILIM MATRİSLERİ

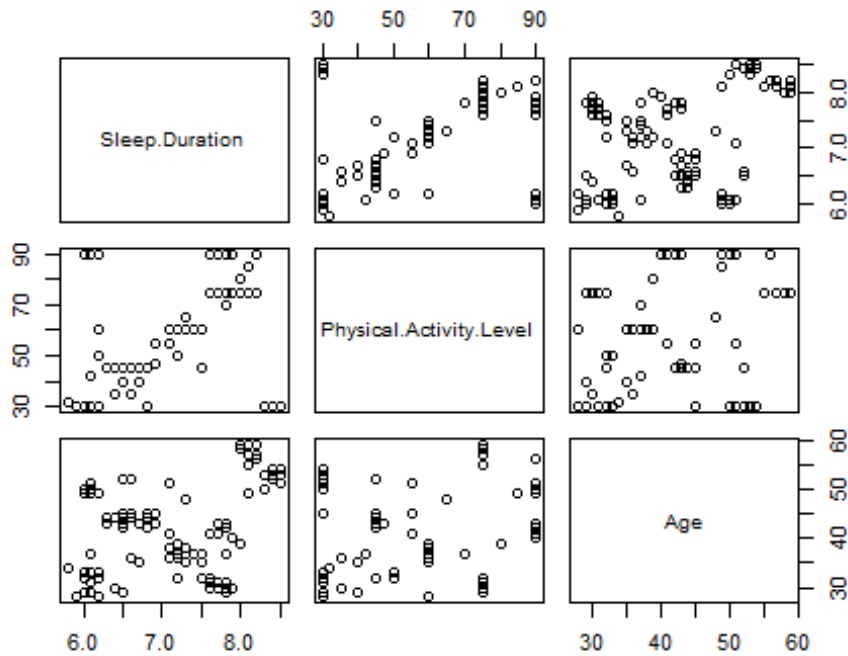
Bu bölümde değişkenlerin korelasyon ilişkileri incelenecektir.

```
korelasyon <- train[,c(5,7,3)]
```

```
cor(korelasyon)
```

```
##              Sleep.Duration Physical.Activity.Level      Age
## Sleep.Duration          1.0000000          0.2017374 0.4180600
## Physical.Activity.Level  0.2017374          1.0000000 0.1774586
## Age                     0.4180600          0.1774586 1.0000000
```

```
plot(korelasyon)
```



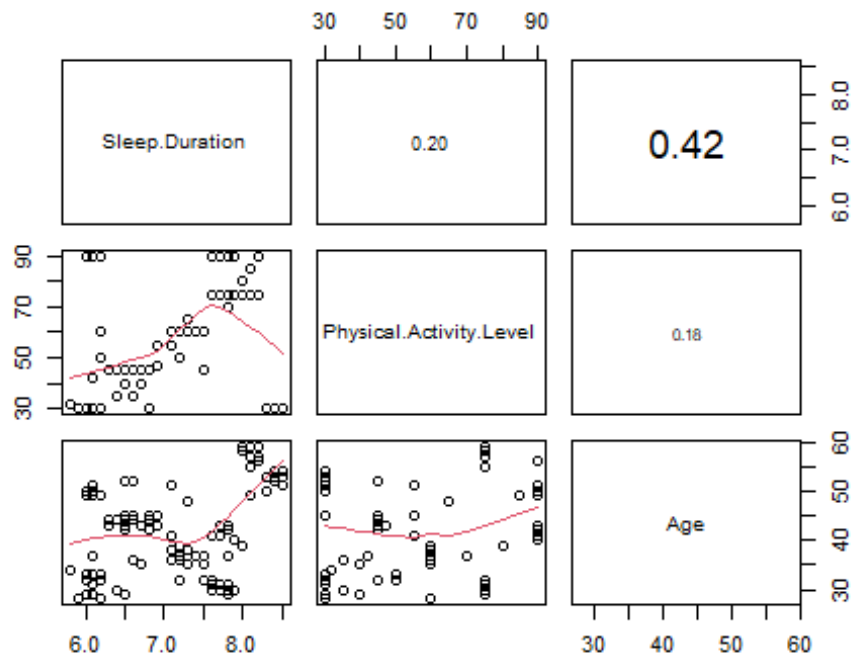
Aydın Hoca'nın yöntemi ile saçılım grafiği oluşturulmak istenirse;

```
panel.cor <- function(x,y,digits=2,prefix="",cex.cor)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr=c(0,1,0,1))
  r=(cor(x,y))
  txt <- format(c(r,0.123456789),digits=digits)[1]
  txt <- paste(prefix, txt, sep="")
  if(missing(cex.cor)) cex <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex=cex*abs(r))
}

panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col="#F08080", ...)
}

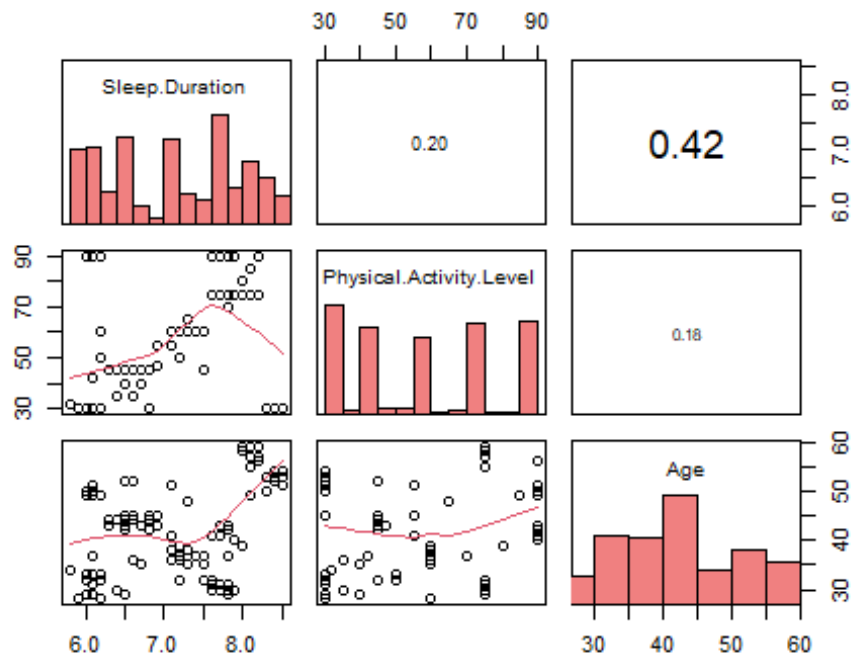
pairs(korelasyon, lower.panel=panel.smooth, upper.panel=panel.cor)
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
```

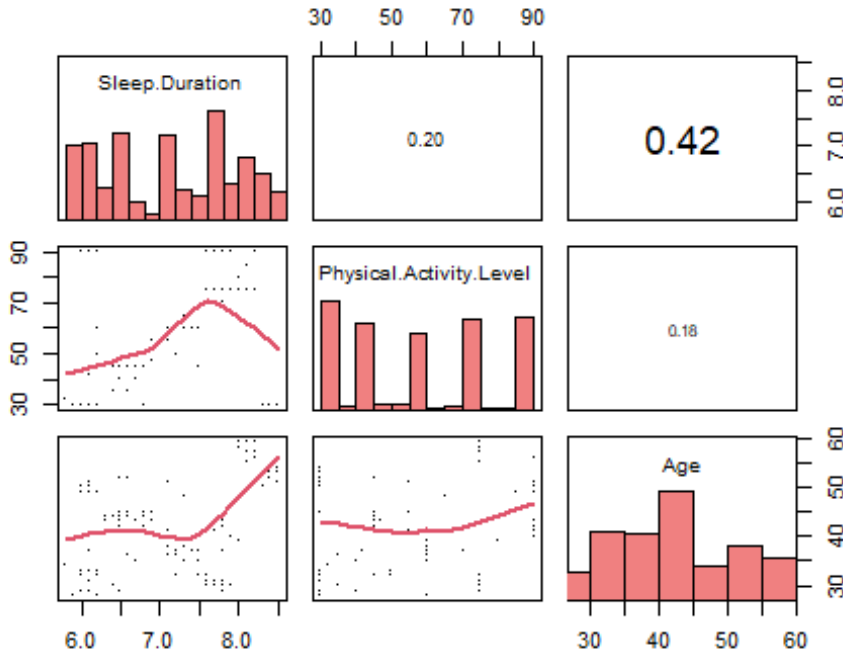
```
pairs(korelasyon, diag.panel=panel.hist, lower.panel=panel.smooth,  
upper.panel=panel.cor)
```

[illegible]



```
pairs(korelasyon, diag.panel=panel.hist, lower.panel=function(x,y)
panel.smooth(x, y, pch=".", lwd=2), upper.panel=panel.cor)
```

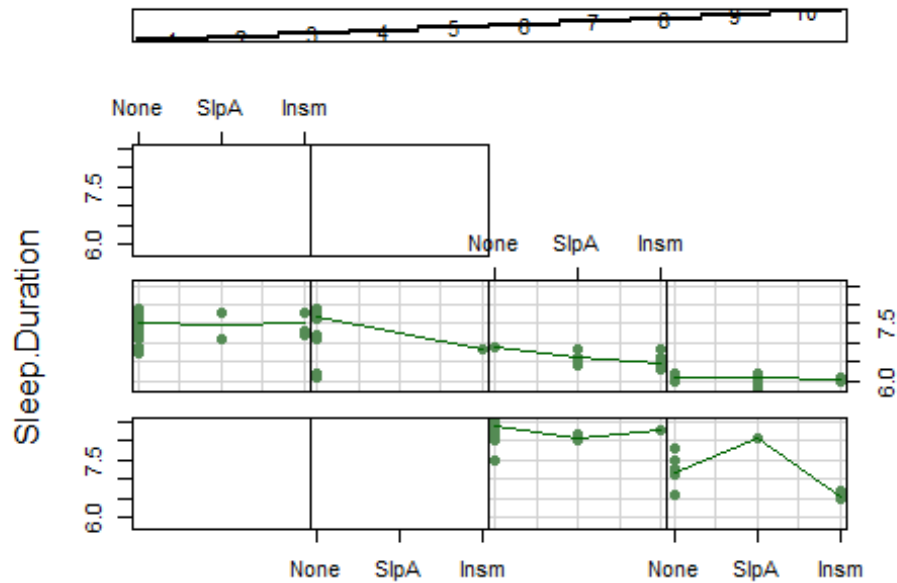
[illegible]



Uyku süresi, fiziksel aktivite ve yaş değişkenleri için saçılım grafiği, korelasyon katsayıları ve yoğunluk grafikleri elde edilmiş ve aşağıda yorumlaması yapılmıştır. -Üç değişken için de korelasyon ilişkisinin pozitif ve * işaretinden hareketle anlamlı olduğu söylenebilir. Uyku süresinin fiziksel aktivite için korelasyon katsayısı 0.20, yaş değişkeni için 0.42 gelirken; fiziksel aktivite ile yaş arasındaki korelasyon katsayısı 0.18 ile en az değere sahiptir. - Yoğunluk grafikleri incelendiğinde, uyku süresi için oluşturulan grafiğin çift tepe noktasına sahip olduğundan iki ayrı uyku süresi grubunun olabileceğini, fiziksel aktivite için oluşturulan yoğunluk grafiğinin nispeten düz olması verilerin geniş bir aralığa yayılmış olabileceğini ve yaş için oluşturulan yoğunluk grafiği, yaşların yaklaşık 30-40 aralığında yoğunlaştığını göstermektedir. -Saçılım grafikleri incelendiğinde, uyku süresi ve fiziksel aktivite seviyesi arasında hafif pozitif bir ilişkinin olması, daha fazla fiziksel aktivitenin daha uzun uyku süreleri ile ilişkilendirilebileceğini; uyku süresi ve yaş arasında orta düzeyde pozitif bir ilişki olması, yaşın da ilerlemesi ile uyku süresinin artabileceğini ve fiziksel aktivite seviyesi ile yaş arasında hafif pozitif bir ilişkinin olması, yaş ilerledikçe fiziksel aktivite seviyesinin biraz artabileceğini göstermektedir.

```
k_matris <- function(x, y, pch=20, col=1, ...) {
  points(x=x, y=y, col=col, pch=pch, type="p", ...)
  lines(lowess(x=x, y=y), col="darkgreen")
}
coplot(Sleep.Duration~Sleep.Disorder|Stress.Level, col="palegreen4",
data=train,panel=k_matris, pch=20, cex=1.5)
```

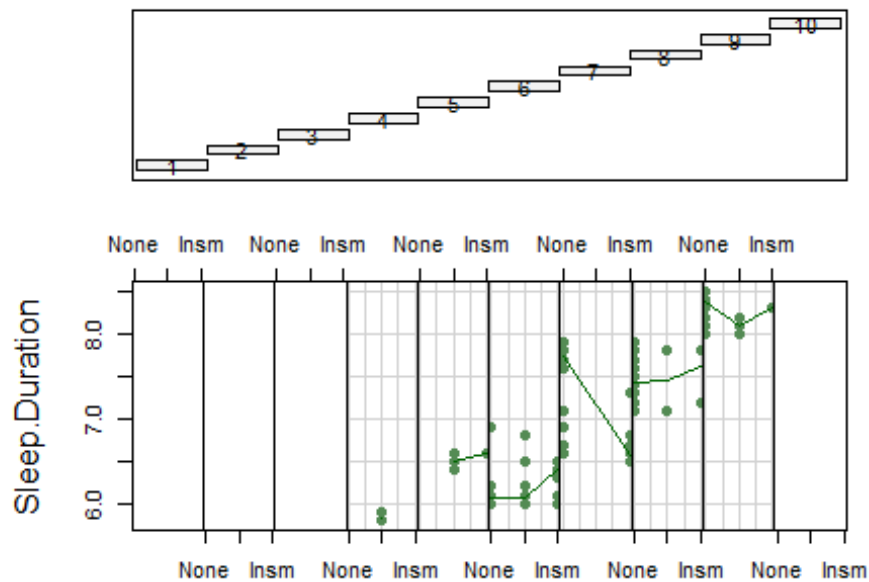
Given : Stress.Level



Sleep.Disorder

```
coplot(Sleep.Duration~Sleep.Disorder|Quality.of.Sleep, col="palegreen4",
data=train, panel=k_matris, rows=1, pch=20, cex=1.5)
```

Given : Quality.of.Sleep



Sleep.Disorder

Yukarıda yazılan kod yardımı ile uyku süreleri ve uyku bozukluklarının, stres seviyesi ve uyku kalitesine göre koşullu matrisleri elde edilmiştir.

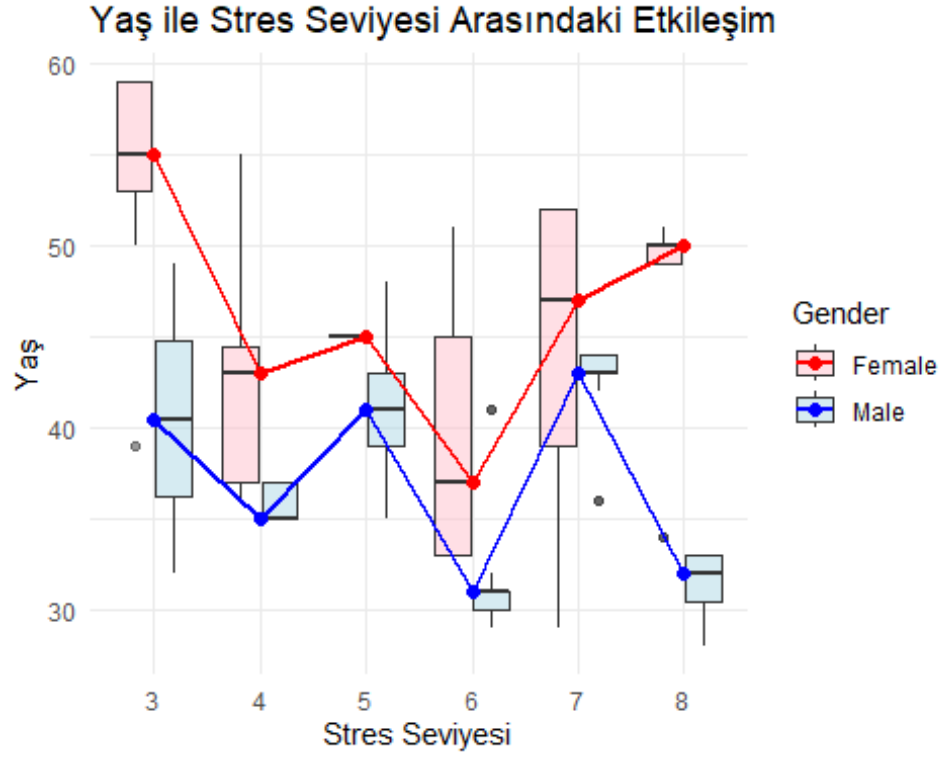
Grafik incelendiğinde daha yüksek uyku kalitesine sahip olan bireylerin genellikle daha uzun süre uyuduğu; "Uykusuzluk" (Insomnia) rahatsızlığı çeken bireylerin uyku sürelerinin, bu rahatsızlığa sahip olmayan bireylere göre genellikle daha kısa veya düzensiz olduğu ve bu bireylerin için uyku sürelerinde daha fazla dalgalanma yaşandığı söylenebilmektedir.

YAŞ İLE STRES SEVİYESİ İÇİN ETKİLEŞİM GRAFİĞİ

```
library(dplyr)
etkilesim_yst <- train%>%
  group_by(Stress.Level, Gender)%>%
  summarise(Median=median(Age,na.rm = TRUE), .groups = 'drop')
etkilesim_yst

## # A tibble: 12 × 3
##   Stress.Level Gender Median
##   <ord>         <fct>   <dbl>
## 1 3           Female    55
## 2 3           Male     40.5
## 3 4           Female    43
## 4 4           Male     35
## 5 5           Female    45
## 6 5           Male     41
## 7 6           Female    37
## 8 6           Male     31
## 9 7           Female    47
## 10 7          Male     43
## 11 8          Female    50
## 12 8          Male     32

ggplot(train, aes(x = as.factor(Stress.Level), y = Age, fill = Gender)) +
  geom_boxplot(alpha = 0.5) +
  geom_line(data = etkilesim_yst, aes(x = as.factor(Stress.Level), y =
Median, color = Gender, group = Gender), size = 1) +
  geom_point(data = etkilesim_yst, aes(x = as.factor(Stress.Level), y =
Median, color = Gender), size = 2) +
  labs(title = "Yaş ile Stres Seviyesi Arasındaki Etkileşim",
    x = "Stres Seviyesi",
    y = "Yaş") +
  theme_minimal() +
  scale_fill_manual(values = c("Female" = "pink", "Male" = "lightblue")) +
  scale_color_manual(values = c("Female" = "red", "Male" = "blue"))
```



Yaş ile stres seviyesi arasında etkileşimin olup olmadığı merak edilmiş ve kod yardımı ile grafik oluşturulmuştur.

Her bir stres seviyesi için ayrı kategoriler neticesinde kadın ve erkek olmak üzere iki grubun stres yaşlarına bakıldığında -Stres seviyesinin 3 olduğu grup için kadınlarda ortalama yaşı 55, erkeklerde yaklaşık 40 civarlarında olduğu -Stres seviyesinin 4 olduğu grup için kadınlarda ortalama yaşı yaklaşık 45, erkeklerde ortalama 35 civarlarında olduğu -Stres seviyesinin 5 olduğu grup için kadınlarda ortalama yaşı ortalama 45, erkeklerde yaklaşık 41 civarlarında olduğu -Stres seviyesinin 6 olduğu grup için kadınlarda ortalama yaşı yaklaşık 37, erkeklerde yaklaşık 32 civarlarında olduğu -Stres seviyesinin 7 olduğu grup için kadınlarda ortalama yaşı yaklaşık 47, erkeklerde yaklaşık 44 civarlarında olduğu -Stres seviyesinin 8 olduğu grup için kadınlarda ortalama yaşı 50, erkeklerde yaklaşık 33 civarlarında olduğu görülmektedir.

Etkileşim eğrilerinin özellikle stres seviyelerinin 3,4,5,6,7 olduğu durumlarda her iki cinsiyet için benzer etkiye sahip olduğunu söylemek mümkündür. Stres seviyesinin 8 olduğu durumda ise yaş medyanında büyük değişiklikler olduğu, stres seviyesinin gruplar üzerinde farklı etkiler yarattığı gözlemlenmiştir.

DÖNÜŞÜM UYGULANMASI

Log dönüşümü:

```
library(ggplot2)
```

```
sleep.duration_log <- log(train$Sleep.Duration)
```

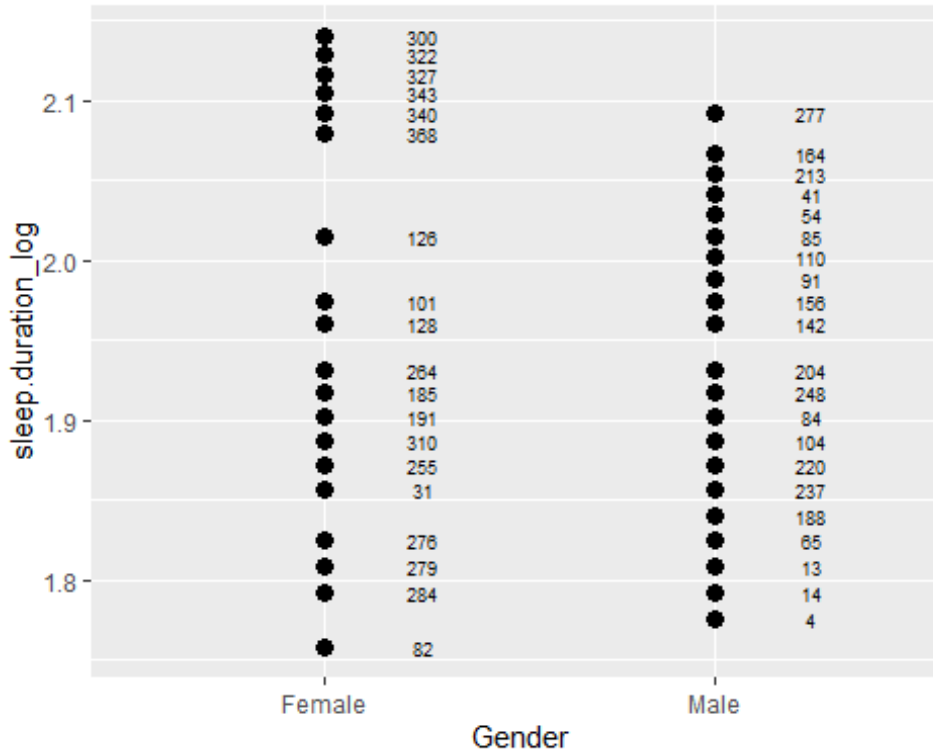
```
sleep.duration_log
```

[1] 1.974081 1.808289 2.104134 1.871802 2.104134 1.974081 1.791759
1.808289
[9] 1.960095 2.116256 2.041220 1.987874 1.871802 1.974081 2.104134
1.791759
[17] 1.757858 1.791759 1.808289 1.824549 1.871802 2.116256 1.871802
2.028148
[25] 1.916923 2.054124 2.091864 2.028148 2.001480 1.871802 1.987874
1.791759
[33] 1.871802 1.840550 2.014903 1.987874 2.054124 1.987874 1.871802
2.079442
[41] 2.140066 1.824549 1.808289 2.041220 1.902108 1.871802 1.974081
2.066863
[49] 1.808289 1.871802 2.104134 1.824549 1.871802 1.824549 2.128232
2.128232
[57] 2.128232 1.791759 1.824549 1.974081 2.128232 2.091864 1.774952
2.140066
[65] 2.128232 1.871802 2.091864 1.856298 1.960095 1.808289 1.791759
1.808289
[73] 1.960095 1.987874 1.960095 1.791759 1.974081 2.041220 2.041220
2.091864
[81] 2.128232 1.974081 2.128232 2.104134 1.824549 1.902108 2.140066
1.791759
[89] 1.824549 1.887070 1.974081 2.140066 2.091864 2.066863 1.808289
1.974081
[97] 2.054124 2.054124 1.960095 1.987874 1.987874 2.054124 1.887070
1.974081
[105] 2.140066 2.054124 2.054124 2.140066 2.116256 2.140066 2.041220
1.791759
[113] 2.054124 1.960095 2.054124 2.079442 1.856298 1.791759 1.791759
1.871802
[121] 1.871802 1.808289 2.079442 1.916923 2.054124 2.128232 1.987874
1.871802
[129] 2.116256 1.887070 1.974081 2.054124 2.079442 1.974081 2.091864
1.824549
[137] 2.091864 1.840550 2.041220 1.791759 1.960095 1.791759 2.041220
1.791759
[145] 1.871802 1.808289 1.887070 1.974081 2.014903 2.041220 2.054124
2.104134
[153] 1.808289 2.041220 2.091864 2.014903 2.041220 2.079442 2.041220
1.840550
[161] 1.840550 1.856298 1.887070 1.871802 2.091864 1.887070 2.054124
1.791759
[169] 2.066863 2.028148 1.856298 2.028148 1.960095 1.791759 2.079442
1.974081
[177] 2.041220 1.887070 2.128232 2.091864 1.871802 1.791759 2.041220
1.808289
[185] 1.960095 1.916923 1.791759 2.054124 2.116256 2.128232 2.104134
2.079442
[193] 1.974081 2.104134 2.054124 2.091864 2.014903 2.054124 2.140066
2.054124

```
## [201] 1.960095 2.091864 2.041220 1.808289 1.931521 2.140066 1.871802
1.808289
## [209] 1.840550 2.079442 2.041220 1.960095 1.808289 1.791759 1.887070
2.028148
## [217] 2.054124 1.791759 2.079442 1.974081 1.887070 1.931521 1.791759
2.128232
## [225] 2.054124 2.054124 1.840550 1.824549 1.824549 1.987874 1.902108
2.091864
## [233] 1.757858 1.871802 2.041220 1.791759 1.791759 2.140066 1.974081
1.824549
## [241] 2.041220 1.808289 2.091864 2.041220 1.808289 1.916923 1.974081
2.066863
## [249] 2.104134 1.840550 2.001480 1.887070 1.887070 2.041220 1.974081
1.871802
## [257] 1.791759 1.974081 1.887070 1.856298 2.128232 2.054124

ggplot(train, aes(Gender, sleep.duration_log, label=rownames(train)))+
  geom_point(size=3)+
  geom_text(label=rownames(train), nudge_x=0.25, check_overlap=TRUE, size=2.5)+
  geom_smooth(method="auto", col="palegreen2", se=FALSE)

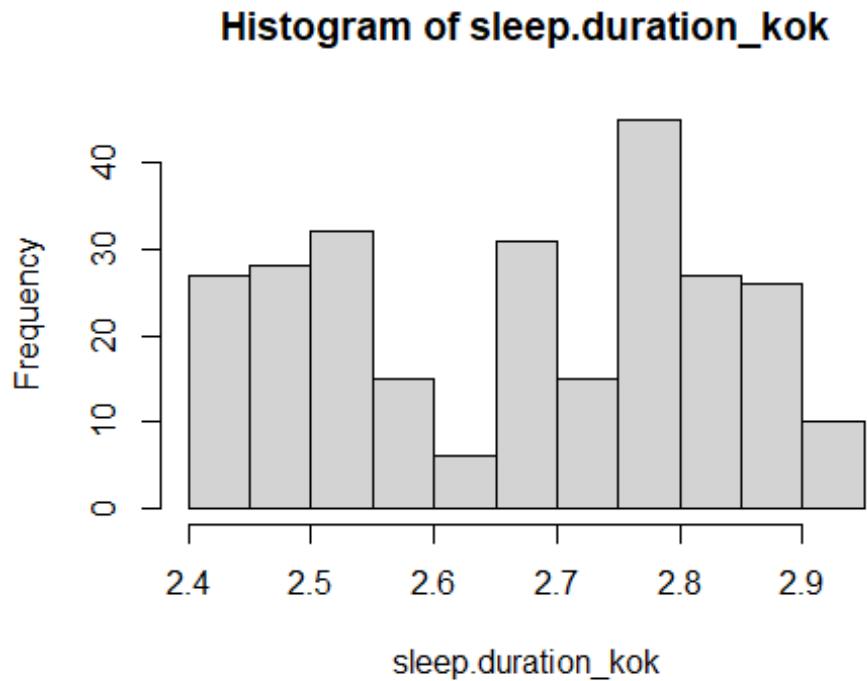
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



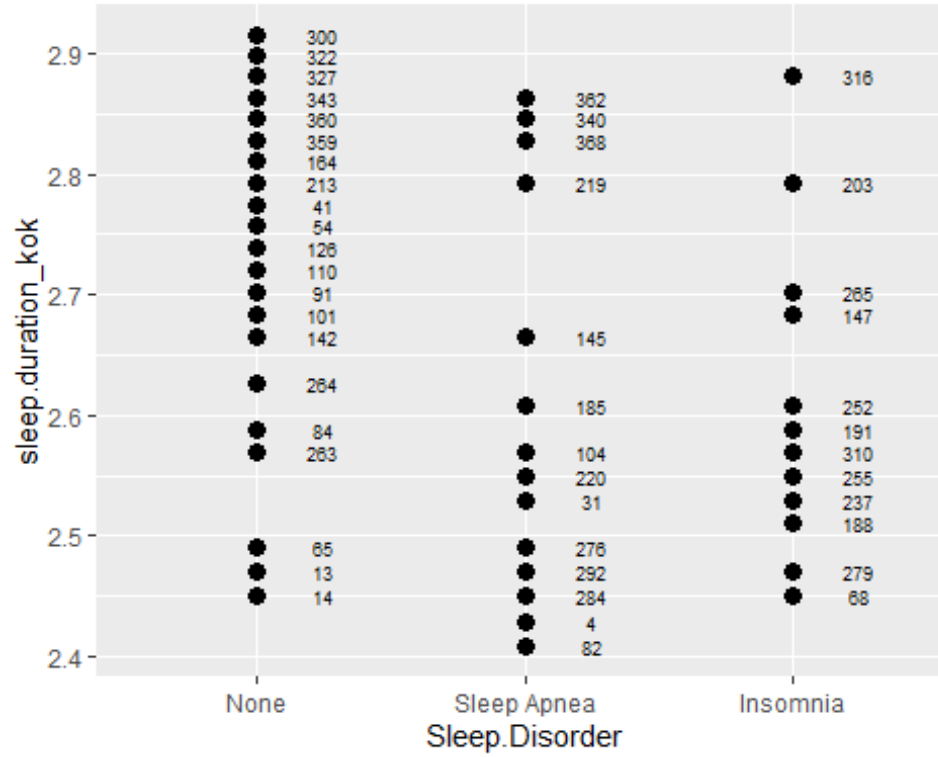
Yukarıdaki kod yardımıyla uyku sürelerinin logaritmik dönüşümü yapılmıştır. Daha sonra dönüşüm uygulanan uyku süreleri ile cinsiyet bazında bir grafik oluşturulmuştur. Genel olarak her iki cinsiyet grubu arasında belirli uyku süresi logaritmik değerlerinde benzer yoğunluklar gözlemlenebilirken, bazı değerlerde (özellikle uç değerlerde) farklılıklar

gözlemlenmektedir. Bu görselleştirme sayesinde cinsiyetler arası uyku süresi farklılıklarını daha net bir şekilde anlamak mümkündür.

```
# Kök dönüşümü:  
sleep.duration_kok <-sqrt(train$Sleep.Duration)  
hist(sleep.duration_kok)
```



```
library(ggplot2)  
ggplot(train, aes(Sleep.Disorder, sleep.duration_kok))+  
  geom_point(size=3)+  
  geom_text(label=rownames(train), nudge_x=0.25, check_overlap=TRUE, size=2.5)+  
  geom_smooth(method = "auto", se = FALSE)  
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Kök dönüşümü uygulanmış uyku süresi için uyku bozuklukları ile oluşturulan grafik incelendiğinde her üç grup arasında belirgin bir fark olmamakla birlikte, “Uyku Apnesi” grubunun uyku süresi dağılımı daha dar bir aralıkta yoğunlaştığı görülmektedir. Rahatsızlığı olmayan bireylerin olduğu grup ile ve “Uykusuzluk” grubunda uyku süresi dağılımı daha geniş ve bireyler arasında uyku süresi değerlerinin oldukça çeşitlilik gösterdiği söylenebilir. Her grupta bireysel farklılıkların oldukça belirgin olduğunu söylemek mümkündür.

#polynomial merkezileştirme:

```
ort_sleep.duration <- mean(train$Sleep.Duration)
ort_sleep.duration
```

```
## [1] 7.170229
```

```
sapma <- (train$Sleep.Duration-ort_sleep.duration)
sapma
```

```
## [1] 0.02977099 -1.07022901 1.02977099 -0.67022901 1.02977099
0.02977099
```

```
## [7] -1.17022901 -1.07022901 -0.07022901 1.12977099 0.52977099
0.12977099
```

```
## [13] -0.67022901 0.02977099 1.02977099 -1.17022901 -1.37022901 -
1.17022901
```

```
## [19] -1.07022901 -0.97022901 -0.67022901 1.12977099 -0.67022901
0.42977099
```

```
## [25] -0.37022901 0.62977099 0.92977099 0.42977099 0.22977099 -
0.67022901
```

```
## [31] 0.12977099 -1.17022901 -0.67022901 -0.87022901 0.32977099
```

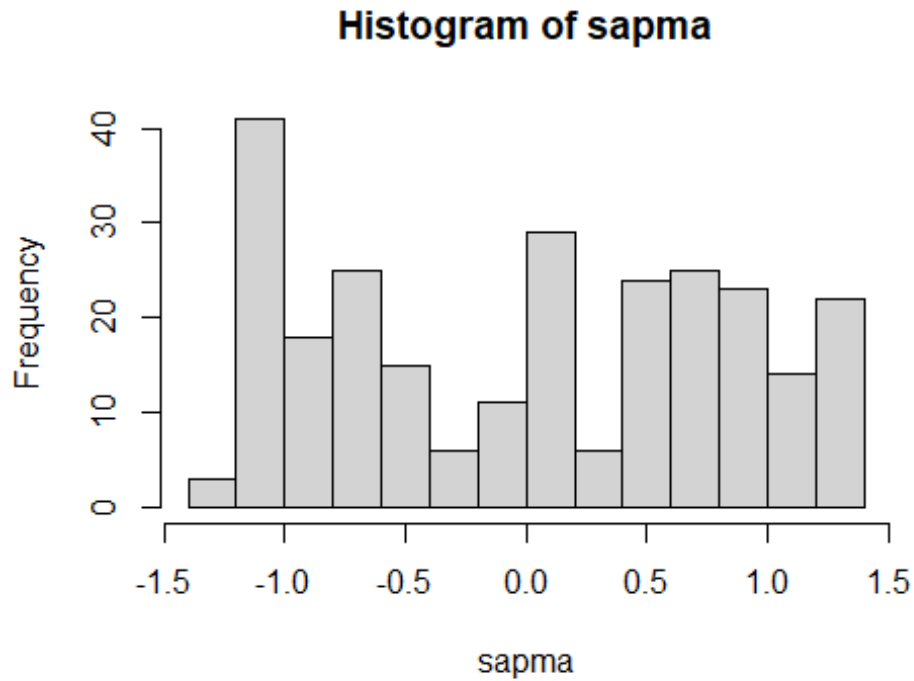
0.12977099
[37] 0.62977099 0.12977099 -0.67022901 0.82977099 1.32977099 -
0.97022901
[43] -1.07022901 0.52977099 -0.47022901 -0.67022901 0.02977099
0.72977099
[49] -1.07022901 -0.67022901 1.02977099 -0.97022901 -0.67022901 -
0.97022901
[55] 1.22977099 1.22977099 1.22977099 -1.17022901 -0.97022901
0.02977099
[61] 1.22977099 0.92977099 -1.27022901 1.32977099 1.22977099 -
0.67022901
[67] 0.92977099 -0.77022901 -0.07022901 -1.07022901 -1.17022901 -
1.07022901
[73] -0.07022901 0.12977099 -0.07022901 -1.17022901 0.02977099
0.52977099
[79] 0.52977099 0.92977099 1.22977099 0.02977099 1.22977099
1.02977099
[85] -0.97022901 -0.47022901 1.32977099 -1.17022901 -0.97022901 -
0.57022901
[91] 0.02977099 1.32977099 0.92977099 0.72977099 -1.07022901
0.02977099
[97] 0.62977099 0.62977099 -0.07022901 0.12977099 0.12977099
0.62977099
[103] -0.57022901 0.02977099 1.32977099 0.62977099 0.62977099
1.32977099
[109] 1.12977099 1.32977099 0.52977099 -1.17022901 0.62977099 -
0.07022901
[115] 0.62977099 0.82977099 -0.77022901 -1.17022901 -1.17022901 -
0.67022901
[121] -0.67022901 -1.07022901 0.82977099 -0.37022901 0.62977099
1.22977099
[127] 0.12977099 -0.67022901 1.12977099 -0.57022901 0.02977099
0.62977099
[133] 0.82977099 0.02977099 0.92977099 -0.97022901 0.92977099 -
0.87022901
[139] 0.52977099 -1.17022901 -0.07022901 -1.17022901 0.52977099 -
1.17022901
[145] -0.67022901 -1.07022901 -0.57022901 0.02977099 0.32977099
0.52977099
[151] 0.62977099 1.02977099 -1.07022901 0.52977099 0.92977099
0.32977099
[157] 0.52977099 0.82977099 0.52977099 -0.87022901 -0.87022901 -
0.77022901
[163] -0.57022901 -0.67022901 0.92977099 -0.57022901 0.62977099 -
1.17022901
[169] 0.72977099 0.42977099 -0.77022901 0.42977099 -0.07022901 -
1.17022901
[175] 0.82977099 0.02977099 0.52977099 -0.57022901 1.22977099
0.92977099
[181] -0.67022901 -1.17022901 0.52977099 -1.07022901 -0.07022901 -

```

0.37022901
## [187] -1.17022901  0.62977099  1.12977099  1.22977099  1.02977099
0.82977099
## [193]  0.02977099  1.02977099  0.62977099  0.92977099  0.32977099
0.62977099
## [199]  1.32977099  0.62977099 -0.07022901  0.92977099  0.52977099 -
1.07022901
## [205] -0.27022901  1.32977099 -0.67022901 -1.07022901 -0.87022901
0.82977099
## [211]  0.52977099 -0.07022901 -1.07022901 -1.17022901 -0.57022901
0.42977099
## [217]  0.62977099 -1.17022901  0.82977099  0.02977099 -0.57022901 -
0.27022901
## [223] -1.17022901  1.22977099  0.62977099  0.62977099 -0.87022901 -
0.97022901
## [229] -0.97022901  0.12977099 -0.47022901  0.92977099 -1.37022901 -
0.67022901
## [235]  0.52977099 -1.17022901 -1.17022901  1.32977099  0.02977099 -
0.97022901
## [241]  0.52977099 -1.07022901  0.92977099  0.52977099 -1.07022901 -
0.37022901
## [247]  0.02977099  0.72977099  1.02977099 -0.87022901  0.22977099 -
0.57022901
## [253] -0.57022901  0.52977099  0.02977099 -0.67022901 -1.17022901
0.02977099
## [259] -0.57022901 -0.77022901  1.22977099  0.62977099

```

```
hist(sapma)
```



Ortalama uyku süresinin hesabı yapılmış olup 7.17 saat olarak bulunmuştur. Daha sonra ise genel olarak uyku sürelerinin ortalamadan ne kadar saptığı incelenmiştir. Negatif değerler ortalamanın ne kadar altında kaldığını gösterirken, pozitif değerler ortalamanın üzerinde kalan değerleri göstermektedir.

Hesaplama sonrası incelenmiş olan histogram grafiğinde 40 ve üzeri bireyin ortalama uyku süresinden daha az bir süre zarfında uyuduğu ve buradan hareketle genel olarak dağılımın sağa çarpık olduğu gözlemlenmiştir.

TUKEY TESTİ YARDIMI İLE REGRESYON MODELİ KURMA

```
library(rcompanion)
```

```
##
## Attaching package: 'rcompanion'

## The following object is masked from 'package:psych':
##
##   phi

Tukey_Sleep.Duration <- transformTukey(train$Sleep.Duration,plotit=FALSE)

##
##      lambda      W Shapiro.p.value
## 459    1.45 0.924      2.556e-10
##
## if (lambda > 0){TRANS = x ^ lambda}
```

```

## if (lambda == 0){TRANS = log(x)}
## if (lambda < 0){TRANS = -1 * x ^ lambda}

Tukey_Heart.Rate <- transformTukey(train$Heart.Rate, plotit=FALSE)

##
##      lambda      W Shapiro.p.value
## 252 -3.725 0.9325      1.432e-09
##
## if (lambda > 0){TRANS = x ^ lambda}
## if (lambda == 0){TRANS = log(x)}
## if (lambda < 0){TRANS = -1 * x ^ lambda}

Tukey_Pysical.Activity.Level <- transformTukey(train$Daily.Steps,
plotit=FALSE)

##
##      lambda      W Shapiro.p.value
## 409  0.2 0.9269      4.543e-10
##
## if (lambda > 0){TRANS = x ^ lambda}
## if (lambda == 0){TRANS = log(x)}
## if (lambda < 0){TRANS = -1 * x ^ lambda}

Tukey_Daily.Steps <- transformTukey(train$Daily.Steps, plotit=FALSE)

##
##      lambda      W Shapiro.p.value
## 409  0.2 0.9269      4.543e-10
##
## if (lambda > 0){TRANS = x ^ lambda}
## if (lambda == 0){TRANS = log(x)}
## if (lambda < 0){TRANS = -1 * x ^ lambda}

# Uyku Süresi için Box-Cox dönüşümü
library(MASS)

## Warning: package 'MASS' was built under R version 4.4.1

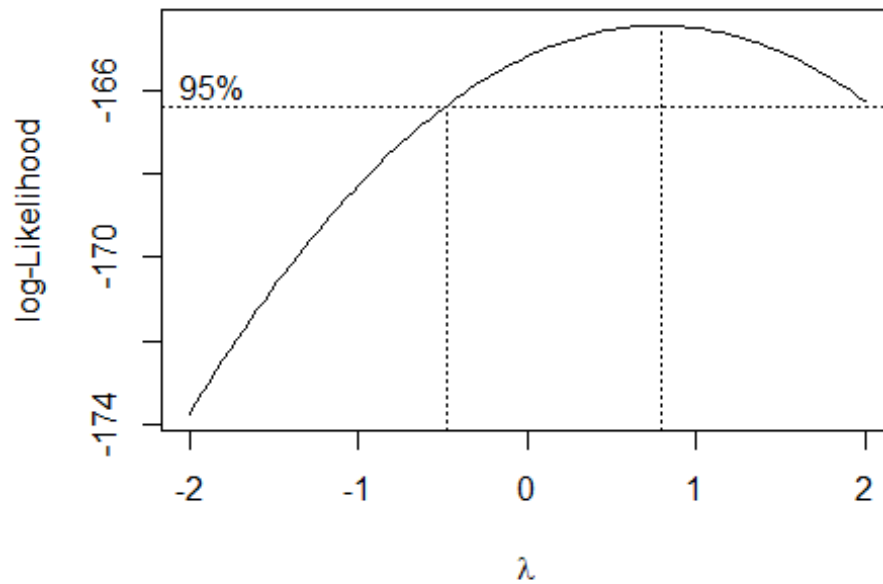
##
## Attaching package: 'MASS'

## The following object is masked from 'package:plotly':
##
##      select

## The following object is masked from 'package:dplyr':
##
##      select

Sleep.Duration_box <- boxcox(train$Sleep.Duration~ 1,
lambda = seq(-2,2,0.1))

```



```
Sleep.Duration_cox <- data.frame(Sleep.Duration_box$x, Sleep.Duration_box$y)
Sleep.Duration_cox <- Sleep.Duration_cox[order(-
Sleep.Duration_cox$Sleep.Duration_box.y),]
Sleep.Duration_cox[1,]

##      Sleep.Duration_box.x Sleep.Duration_box.y
## 70      0.7878788      -164.4648

lambda <- Sleep.Duration_cox[1, "Sleep.Duration_box.x"]
lambda

## [1] 0.7878788

lambda_sleep <- 0.7878788
train$Sleep.Duration_transformed <- train$Sleep.Duration ^ lambda_sleep

# Kalp Atış Hızı için Box-Cox dönüşümü
lambda_heart <- -3.725
train$Heart.Rate_transformed <- -1 * train$Heart.Rate ^ lambda_heart

# Regresyon Modeli Kurma:
model <- lm(Sleep.Duration_transformed ~ Heart.Rate_transformed + Age +
Gender + Sleep.Disorder + Stress.Level, data = train)

summary(model)

##
## Call:
```

```
## lm(formula = Sleep.Duration_transformed ~ Heart.Rate_transformed +
##     Age + Gender + Sleep.Disorder + Stress.Level, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.61026 -0.07815  0.00322  0.08158  0.39193
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.192e+00  1.286e-01  32.586 < 2e-16 ***
## Heart.Rate_transformed -1.178e+05  6.595e+05  -0.179  0.8583
## Age            1.100e-02  2.200e-03   5.002 1.07e-06 ***
## GenderMale      1.835e-01  3.268e-02   5.615 5.19e-08 ***
## Sleep.DisorderSleep Apnea -6.861e-02  3.168e-02  -2.166  0.0313 *
## Sleep.DisorderInsomnia  -2.732e-01  3.330e-02  -8.207 1.20e-14 ***
## Stress.Level.L    -7.153e-01  4.216e-02 -16.966 < 2e-16 ***
## Stress.Level.Q    -1.309e-01  2.778e-02  -4.713 4.05e-06 ***
## Stress.Level.C    -2.611e-01  2.500e-02 -10.444 < 2e-16 ***
## Stress.Level^4     1.659e-01  3.056e-02   5.429 1.34e-07 ***
## Stress.Level^5     7.137e-02  2.983e-02   2.393  0.0175 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1359 on 251 degrees of freedom
## Multiple R-squared:  0.9035, Adjusted R-squared:  0.8996
## F-statistic: 234.9 on 10 and 251 DF,  p-value: < 2.2e-16
```

Tukey testi uygulanmış olan nicel değişkenler incelenmiştir: -Uyku süresi değişkeni için Shapiro-Wilk testi sonucu, normal dağılım varsayımını p-değeri çok küçük olduğundan reddetmektedir. Lambda değeri 1.45 iken bu değişken, sağa çarpık bir dağılım göstermektedir. -Kalp atış hızı değişkeni için Shapiro-Wilk testi sonucu, normal dağılım varsayımını p-değeri yine çok küçük olduğundan reddetmektedir. Lambda değeri -3.725 olduğundan, bu değişken soluk çarpık bir dağılım göstermektedir. -Fiziksel aktivite ve günlük atılan adım değişkenleri için bakılacak olursa benzer şekilde p değeri çok küçük geldiğinden Shapiro-Wilk testi normal dağılım varsayımını reddetmektedir. İki değişken için de Lambda değeri 0.2 gelmiş ve dağılımların sağa çarpık olduğu görülmüştür.

Kurulan regresyon modeli ile aşağıdaki sonuçlar elde edilmiştir: -Kalp atış hızı değişkeninin, uyku süresinin Box-Cox dönüşümü üzerinde istatistiksel olarak anlamlı bir etkisi yok iken ($p > 0.05$); yaş, cinsiyet, uyku bozukluğu ve stres seviyesi değişkenleri uyku süresi üzerinde anlamlı etkilere sahiptir. -Modelin yüksek bir R-kare değeri vardır. Bir diğer deyişle, bağımsız değişkenler uyku süresinin Box-Cox dönüşümünün büyük bir kısmını açıklamaktadır. -Stres seviyesinin farklı derecelerden bileşenleri de modelde önemli rol oynamaktadır. -Bu model, uyku süresini etkileyen faktörlerin analiz edilmesi ve anlamlı etkilerin belirlenmesi açısından oldukça değerlidir.

```
asil <- train[,c(3,7,11)]
library(PerformanceAnalytics)
```



```

## Zorunlu paket yükleniyor: xts
## Zorunlu paket yükleniyor: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

##
## ##### Warning from 'xts' package
## #####
## #
## #
## # The dplyr lag() function breaks how base R's lag() function is supposed
to #
## # work, which breaks lag(my_xts). Calls to lag(my_xts) that you type or
#
## # source() into this session won't work correctly.
#
## #
#
## # Use stats::lag() to make sure you're not using dplyr::lag(), or you can
add #
## # conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop
#
## # dplyr from breaking base R's lag() function.
#
## #
#
## # Code in packages is not affected. It's protected by R's namespace
mechanism #
## # Set `options(xts.warn_dplyr_breaks_lag = FALSE)` to suppress this
warning. #
## #
#
##
#####
##

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##   first, last

##
## Attaching package: 'PerformanceAnalytics'

```

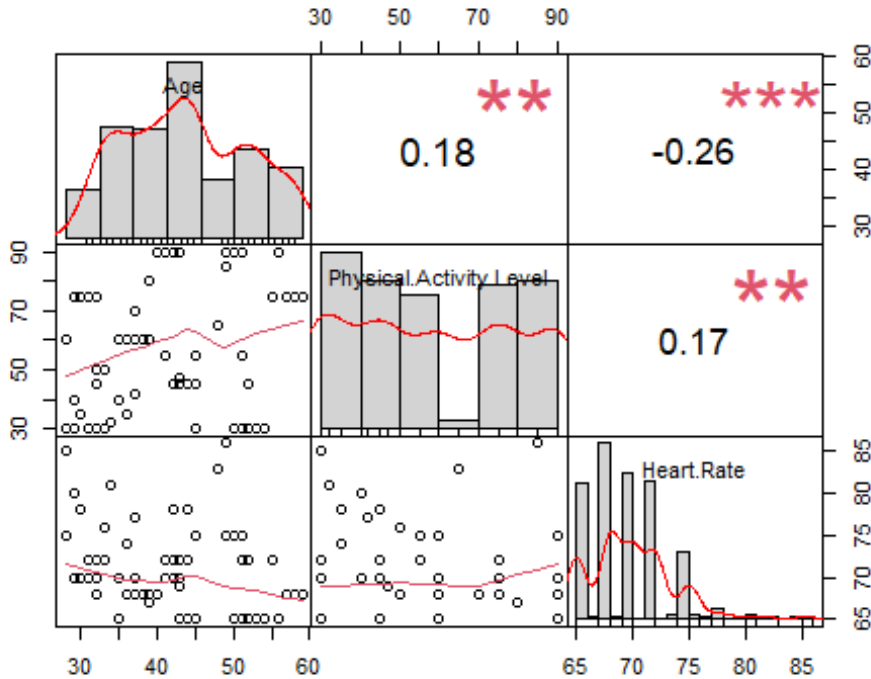
```
## The following object is masked from 'package:gplots':
##
##     textplot

## The following objects are masked from 'package:e1071':
##
##     kurtosis, skewness

## The following object is masked from 'package:graphics':
##
##     legend

chart.Correlation(asil, histogram=TRUE, pch=15)

## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
```



Elde edilen grafikte; yaş, fiziksel aktivite ve kalp atış hızı için korelasyon ilişkileri, saçılım ve histogram grafikleri incelenmiştir. -Yaş ile fiziksel aktivite arasında 0.18 korelasyon katsayısı ile pozitif bir ilişki, yaş ile kalp atış hızı arasında -0.26 korelasyon katsayısı ile negatif yönlü bir ilişki, fiziksel aktivite ile kalp atış hızı arasında ise 0.17 korelasyon katsayısı ile pozitif yönlü bir ilişki olduğu görülmektedir. Ek olarak değerler üzerinde çıkmış olan * işaretlerinden hareketle, değişkenler arasındaki ilişkinin anlamlı olduğu sonucuna varılmaktadır. -Histogram grafikleri incelendiğinde yaşların dağılımının neredeyse normal, fiziksel aktivite dağılımının çok tepeli, kalp atış hızının ise sağa çarpık olduğu söylenebilir. -Saçılım grafikleri incelendiğinde yaş ile fiziksel aktivitenin pozitif, yaş

ile kalp atış hızının negatif, fiziksel aktivite ile kalp atış hızının ise pozitif yönlü eğilim gösterdiği görülmektedir.

KATEGORİLER ARASI MOZAİK ÇİZİMİ

```
mozaik <- xtabs(~Stress.Level+Quality.of.Sleep+Sleep.Disorder, data=train)
ftable(mozaik)
```

```
##                               Sleep.Disorder None Sleep Apnea Insomnia
## Stress.Level Quality.of.Sleep
## 1                1                0                0                0
##                2                0                0                0
##                3                0                0                0
##                4                0                0                0
##                5                0                0                0
##                6                0                0                0
##                7                0                0                0
##                8                0                0                0
##                9                0                0                0
##               10                0                0                0
## 2                1                0                0                0
##                2                0                0                0
##                3                0                0                0
##                4                0                0                0
##                5                0                0                0
##                6                0                0                0
##                7                0                0                0
##                8                0                0                0
##                9                0                0                0
##               10                0                0                0
## 3                1                0                0                0
##                2                0                0                0
##                3                0                0                0
##                4                0                0                0
##                5                0                0                0
##                6                0                0                0
##                7                0                0                0
##                8                1                0                0
##                9               32               24                1
##               10                0                0                0
## 4                1                0                0                0
##                2                0                0                0
##                3                0                0                0
##                4                0                0                0
##                5                0                0                0
##                6                0                0                0
##                7                1                0               16
##                8               21                0                0
##                9                0                2                0
##               10                0                0                0
## 5                1                0                0                0
```

##	2	0	0	0
##	3	0	0	0
##	4	0	0	0
##	5	0	0	0
##	6	0	0	0
##	7	2	0	1
##	8	42	2	3
##	9	0	0	0
##	10	0	0	0
## 6	1	0	0	0
##	2	0	0	0
##	3	0	0	0
##	4	0	0	0
##	5	0	0	0
##	6	3	0	0
##	7	24	0	1
##	8	1	0	0
##	9	0	0	0
##	10	0	0	0
## 7	1	0	0	0
##	2	0	0	0
##	3	0	0	0
##	4	0	0	0
##	5	0	3	1
##	6	1	3	20
##	7	0	0	5
##	8	0	0	0
##	9	0	0	0
##	10	0	0	0
## 8	1	0	0	0
##	2	0	0	0
##	3	0	0	0
##	4	0	3	0
##	5	0	0	0
##	6	26	21	2
##	7	0	0	0
##	8	0	0	0
##	9	0	0	0
##	10	0	0	0
## 9	1	0	0	0
##	2	0	0	0
##	3	0	0	0
##	4	0	0	0
##	5	0	0	0
##	6	0	0	0
##	7	0	0	0
##	8	0	0	0
##	9	0	0	0
##	10	0	0	0
## 10	1	0	0	0

```
##           2           0           0           0
##           3           0           0           0
##           4           0           0           0
##           5           0           0           0
##           6           0           0           0
##           7           0           0           0
##           8           0           0           0
##           9           0           0           0
##          10           0           0           0
```

Stres seviyesi ile uyku bozukluğuna ilişkin grafik:

```
library(ggplot2)
library(ggmosaic)
ggplot(train) +
  geom_mosaic(aes(x = product(Stress.Level), fill=Sleep.Disorder)) +
  facet_grid(Sleep.Disorder~.)
```

```
## Warning: The `scale_name` argument of `continuous_scale()` is deprecated
as of ggplot2
```

```
## 3.5.0.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: The `trans` argument of `continuous_scale()` is deprecated as of
ggplot2 3.5.0.
```

```
## i Please use the `transform` argument instead.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: `unite_()` was deprecated in tidyr 1.2.0.
```

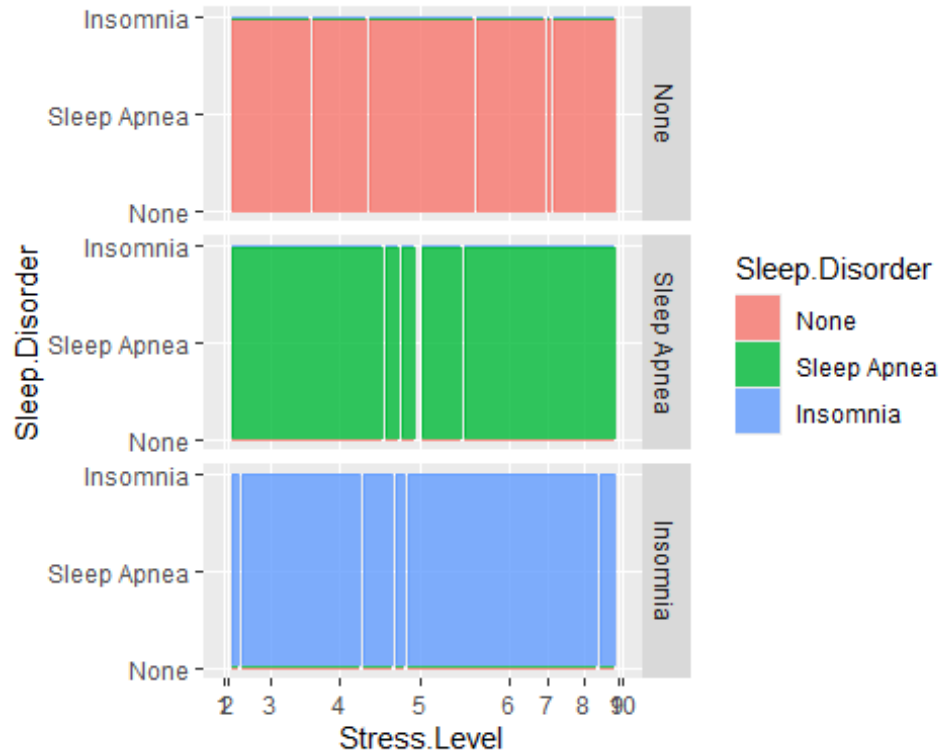
```
## i Please use `unite()` instead.
```

```
## i The deprecated feature was likely used in the ggmosaic package.
```

```
## Please report the issue at <https://github.com/haleyjeppson/ggmosaic>.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Bu tablo stres seviyeleri, uyku kalitesi ve uyku bozuklukları arasındaki olası ilişkileri anlamak için kullanılmıştır. -Stres seviyesi 1 ve 2 için uyku kalitesi ne olursa olsun hiçbir uyku bozukluğu vakaları gözlenmemiştir. -Stres seviyesi 3 ve üzerinde uyku kalitesi arttıkça, bazı uyku bozukluğu vakaları gözlemlenmiştir. -Stres seviyesi arttıkça uyku bozukluğu vakalarının (özellikle Insomnia) arttığı görülmektedir.

Buradan hareketle, özellikle stres seviyesi yüksek olan bireylerde, belirli uyku bozukluklarının yaygın olduğunu söylemek mümkündür.

TEST VERİ SETİNİ GÖRSELLEŞTİRME

```
library(readxl)
test <- read_excel("C:/Users/w10/OneDrive/Desktop/H.Simay Özgül_Ataman Önel
Ük_verianalizi/test_data.xlsx")

test$Sleep.Disorder <- factor(test$Sleep.Disorder, levels=c("None", "Sleep
Apnea", "Insomnia"))
test$Gender <- factor(test$Gender, levels=c("Female", "Male"))
test$BMI.Category <- factor(test$BMI.Category,
levels=c("Normal", "Obese", "Overweight", "Normal Weight"))
test$Quality.of.Sleep <- factor(test$Quality.of.Sleep, levels = 1:10,
ordered=TRUE)
test$Stress.Level <- factor(test$Stress.Level, levels = 1:10, ordered=TRUE)
levels(test$BMI.Category)[levels(test$BMI.Category) == "Normal Weight"] <-
"Normal"
levels(test$BMI.Category)
```

```
## [1] "Normal"      "Obese"      "Overweight"

str(test)

## tibble [112 × 13] (S3: tbl_df/tbl/data.frame)
## $ Person.ID      : num [1:112] 1 5 6 7 8 9 11 15 19 23 ...
## $ Gender         : Factor w/ 2 levels "Female","Male": 2 2 2 2 2
2 2 2 1 2 ...
## $ Age           : num [1:112] 27 28 28 29 29 29 29 29 30 ...
## $ Occupation    : chr [1:112] "Software Engineer" "Sales
Representative" "Software Engineer" "Teacher" ...
## $ Sleep.Duration : num [1:112] 6.1 5.9 5.9 6.3 7.8 7.8 6.1 6 6.5
7.7 ...
## $ Quality.of.Sleep : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 6
4 4 6 7 7 6 6 5 7 ...
## $ Physical.Activity.Level: num [1:112] 42 30 30 40 75 75 30 30 40 75 ...
## $ Stress.Level      : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<...: 6
8 8 7 6 6 8 8 7 6 ...
## $ BMI.Category     : Factor w/ 3 levels "Normal","Obese",...: 3 2 2
2 1 1 1 1 1 1 ...
## $ Blood.Pressure   : chr [1:112] "126/83" "140/90" "140/90"
"140/90" ...
## $ Heart.Rate       : num [1:112] 77 85 85 82 70 70 70 70 80 70 ...
## $ Daily.Steps      : num [1:112] 4200 3000 3000 3500 8000 8000 8000
8000 4000 8000 ...
## $ Sleep.Disorder   : Factor w/ 3 levels "None","Sleep Apnea",...: 1
2 3 3 1 1 1 1 3 1 ...
```

Test veri seti train veri seti gibi kodlar yardımı ile elde edilip, gerekli düzenlemeler yapılmıştır.

REGRESYON MODELİNİN KURULMASI

```
test$Sleep.Duration_log <- log(test$Sleep.Duration)
test$Age_log <- log(test$Age)
test$Physical.Activity.Level_log <- log(test$Physical.Activity.Level)
test$Daily.Steps_log <- log(test$Daily.Steps)

test_model <- lm(Sleep.Duration_log ~
Age_log+Stress.Level+Gender+Physical.Activity.Level_log, data=test)
summary(test_model)

##
## Call:
## lm(formula = Sleep.Duration_log ~ Age_log + Stress.Level + Gender +
##     Physical.Activity.Level_log, data = test)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.188130 -0.015981  0.002032  0.016627  0.084878
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.949467   0.102949  18.936 < 2e-16 ***
## Age_log       -0.062103   0.028783  -2.158  0.0333 *
## Stress.Level.L -0.224847   0.010266 -21.902 < 2e-16 ***
## Stress.Level.Q -0.010969   0.012014  -0.913  0.3634
## Stress.Level.C -0.076576   0.010038  -7.629 1.23e-11 ***
## Stress.Level^4  0.073003   0.009648   7.567 1.67e-11 ***
## Stress.Level^5  0.002448   0.011357   0.216  0.8298
## GenderMale     0.010064   0.010942   0.920  0.3598
## Physical.Activity.Level_log 0.056637   0.012727   4.450 2.18e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03839 on 103 degrees of freedom
## Multiple R-squared:  0.8679, Adjusted R-squared:  0.8576
## F-statistic: 84.57 on 8 and 103 DF,  p-value: < 2.2e-16
```

Kurulan regresyon modelinde yaş, stres seviyesi, cinsiyet ve fiziksel aktivite değişkenlerinin logaritmik dönüşüm uygulanmış olan uyku sürelerindeki etkisi incelenmiştir. Elde edilen sonuçlar doğrultusunda aşağıdaki yorumları yapmak mümkündür. -Artıklar, modelin tahmin ettiği değerler ile gerçek değerler arasındaki farkları gösterir. Bu dağılım oldukça dar ve simetrik görünmekte olup, modelin iyi bir uyum sağladığını söylemektedir. -R kare değeri 0.86 geldiğinden modelin genel uyumunun oldukça yüksek olduğu görülmektedir. Bağımsız değişkenlerin büyük kısmının uyku süresi üzerinde anlamlı etkilere sahip olduğu gözlemlenmiştir.

TEST VERİ SETİ ÜZERİNDEN TAHMİN YAPILMASI

```
tahmin <- predict(test_model, test)
tahmin
```

```
##          1          2          3          4          5          6          7          8
## 1.996419 1.790282 1.790282 1.877707 2.024820 2.024820 1.788102 1.788102
##          9         10         11         12         13         14         15         16
## 1.867642 2.022715 2.022715 2.022715 1.857974 1.981985 1.783961 2.020678
##         17         18         19         20         21         22         23         24
## 2.020678 2.020678 2.020678 2.130464 2.018707 2.018707 1.781989 1.780078
##         25         26         27         28         29         30         31         32
## 1.780078 1.982542 1.961810 1.971874 1.971874 1.971874 2.005507 1.960060
##         33         34         35         36         37         38         39         40
## 1.960060 1.960060 1.960060 1.960060 1.960060 1.960060 1.958359 2.002056
##         41         42         43         44         45         46         47         48
## 1.958359 1.958359 1.958359 1.958359 1.958359 1.956702 2.000399 2.000399
##         49         50         51         52         53         54         55         56
## 1.956702 2.000399 2.000399 1.956702 1.990335 1.859308 1.973392 2.140701
##         57         58         59         60         61         62         63         64
## 1.998786 1.977218 1.977218 2.018645 1.999408 2.018645 2.017148 2.017148
##         65         66         67         68         69         70         71         72
## 2.017148 1.932732 1.859915 1.859915 1.859915 1.971728 2.015687 2.015687
##         73         74         75         76         77         78         79         80
## 1.858487 1.858487 1.931304 1.858487 1.931304 1.931304 1.858487 1.858487
```



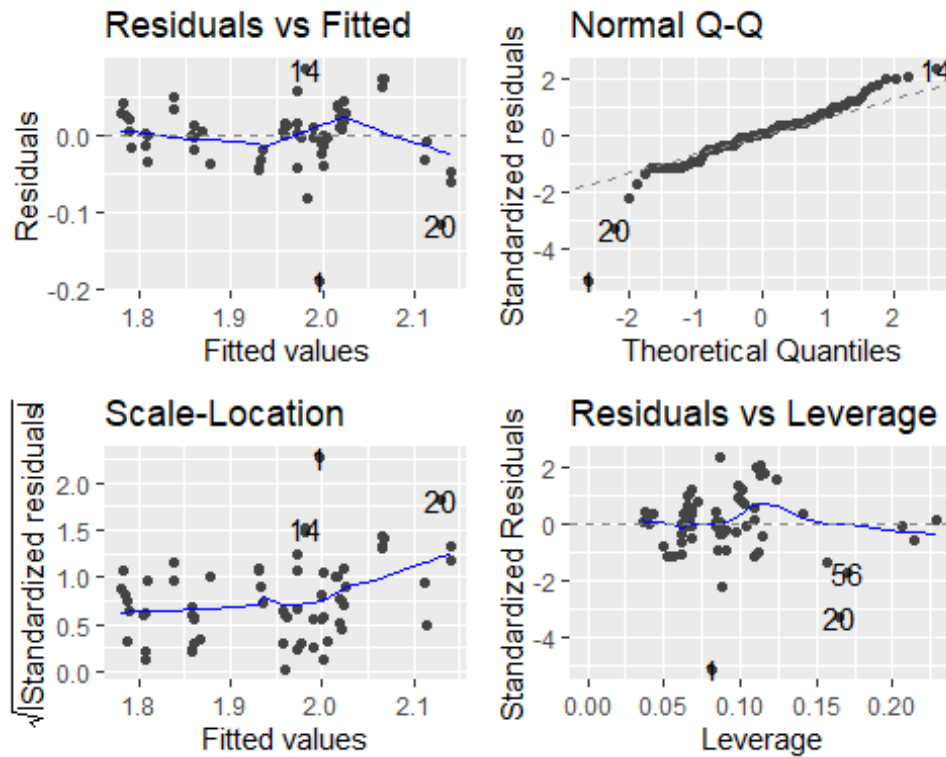
```
##      81      82      83      84      85      86      87      88
## 1.858487 1.858487 1.858487 1.935573 1.929909 1.929909 1.929909 1.929909
##      89      90      91      92      93      94      95      96
## 1.808966 1.990425 1.807686 2.140024 1.806431 1.806431 1.806431 1.806431
##      97      98      99     100     101     102     103     104
## 1.806431 1.806431 1.806431 2.068490 2.068490 1.805201 1.838048 1.838048
##     105     106     107     108     109     110     111     112
## 2.066101 2.066101 2.064940 2.113479 2.113479 2.112398 2.112398 2.112398
```

Tahmin edilen logaritmik uyku süreleri genellikle 1.78 ile 2.14 arasında değişmektedir. En düşük tahmin 1.781989 ve en yüksek tahmin 2.140701 olarak görülmektedir.

Tahminlerin dar bir aralıkta olması, modelin iyi bir performans gösterdiğine işaret edebilir ancak daha kesin değerlendirme için residual analizi ve diğer performans metriklerine bakılmalıdır. Aşağıda da bunun için kod yazılmış ve grafikler elde edilmiştir;

```
defaultSummary <- function(data)
{summary_data <- summary(data)
  return(summary_data)}

library(ggfortify)
autoplot(test_model)
```



Sırası ile grafikler yorumlanacak olursa;

Residuals vs Fitted grafiği için artıkların çoğunun sıfır etrafında toplanmış olduğu görünüyor. Ancak, grafikte belirli bir eğilim veya desenin görünmesi, modelde bazı sistematik hataların olduğunu gösterebilmektedir.

Normal Q-Q Plot grafiği için noktaların genel olarak düz bir çizgi boyunca sıralanmış olduğu ancak uçlarda sapmaların olduğu görülmektedir. Özellikle düşük ve yüksek uçlarda bazı noktaların çizgiden saptmakta olması, normal dağılımdan sapmaların olduğunu gösterebilmektedir.

Scale Location grafiği için bir miktar eğilim var gibi görünmekte, özellikle yüksek uyum değerlerinde artan bir varyansın olduğu söylenebilmektedir.

Residuals vs Leverage grafiği standartlaştırılmış artıkların kaldıraç değerlerine karşı dağılımını gösterir. Yüksek kaldıraç ve büyük artık değerleri olan noktalar, model üzerinde büyük bir etkiye sahip olabilir ve potansiyel etkili gözlemler olarak tanımlanabilir. Bu grafikte bu bilgiler göz önünde bulundurulduğunda 20, 14 ve 56 gibi bazı noktalar dikkat çekmektedir.

Özetle bu sonuçlar, modelin bazı varsayımlarının ihlal edildiğini ve modelin iyileştirilmesi gerekebileceğini göstermektedir.

```
test_model2 <-lm(Sleep.Duration_log~
Age_log+Physical.Activity.Level_log+Heart.Rate, data=test)
summary(test_model2)

##
## Call:
## lm(formula = Sleep.Duration_log ~ Age_log + Physical.Activity.Level_log +
##     Heart.Rate, data = test)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.169913 -0.068822 -0.005195  0.056067  0.250805
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.325208   0.230008  10.109 < 2e-16 ***
## Age_log        -0.027023   0.044884  -0.602  0.548389
## Physical.Activity.Level_log  0.087583   0.025086   3.491  0.000697 ***
## Heart.Rate     -0.008931   0.001787  -4.998  2.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08939 on 108 degrees of freedom
## Multiple R-squared:  0.2489, Adjusted R-squared:  0.228
## F-statistic: 11.93 on 3 and 108 DF,  p-value: 8.32e-07

test_model3 <-lm(Sleep.Duration_log~
Sleep.Disorder+Heart.Rate+Physical.Activity.Level_log, data=test)
summary(test_model3)
```

```
##
## Call:
## lm(formula = Sleep.Duration_log ~ Sleep.Disorder + Heart.Rate +
##     Physical.Activity.Level_log, data = test)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18311 -0.03923 -0.01637  0.04541  0.23638
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.163545    0.176103   12.286 < 2e-16 ***
## Sleep.DisorderSleep Apnea  -0.040756    0.025686   -1.587 0.115523
## Sleep.DisorderInsomnia  -0.067111    0.021064   -3.186 0.001890 **
## Heart.Rate          -0.006779    0.001881   -3.604 0.000478 ***
## Physical.Activity.Level_log  0.071056    0.027142    2.618 0.010128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08565 on 107 degrees of freedom
## Multiple R-squared:  0.3168, Adjusted R-squared:  0.2913
## F-statistic: 12.41 on 4 and 107 DF,  p-value: 2.518e-08

liste <- list(test_model, test_model2, test_model3)

PRESS <- function(linmodel)
{pr <- residuals(linmodel)/(1 - lm.influence(linmodel)$hat)
  sum(pr^2)}
for (model in liste) {
  press_value <- PRESS(model)
  print(paste("PRESS:", round(press_value, 3)))
}

## [1] "PRESS: 0.186"
## [1] "PRESS: 0.972"
## [1] "PRESS: 0.898"
```

Yapılan hesaplamalar sonunda elde edilen “PRESS” değerleri incelenmiştir. “PRESS” değeri, her bir gözlem için tahmin hatalarının karesinin toplamını ifade etmektedir. Daha düşük “PRESS” değeri, modelin daha iyi bir tahmin performansına sahip olduğunu göstermekte, gözlemleri daha iyi tahmin ettiğini ve daha az hata yaptığını belirtmektedir.

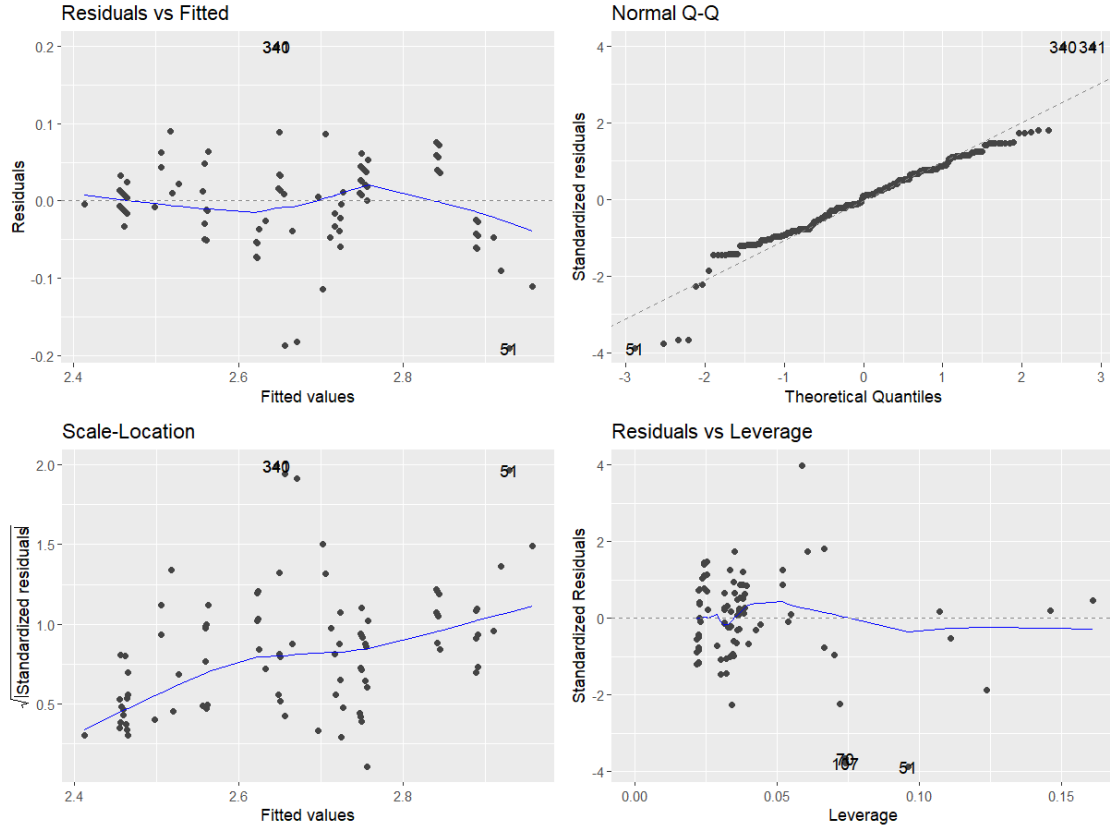
Birinci modelin “PRESS” değeri 0.186 ile en düşük değerdir. Dolayısıyla, birinci modelin diğer modellere göre gözlemleri daha iyi tahmin ettiğini söylemek mümkündür.

ÖNEMLİ NOT

Aşağıdaki kod yardımı ile model için bir grafik oluşturulmuştur ancak R markdown ortamında “knit” seçeneğine basıldığında hata alındığından ve sonuç düzeltilemediğinden

kod, alıntı şeklinde rapora eklenmiş ve çıkan grafik görsel şeklinde konulmuştur. Grafiğin yorumları ise aşağıda yer almaktadır.

```
<testAgelog < -log(testAge) testPhysical.Activity.Levellog <
-log(testPhysical.Activity.Level) library(ggfortify) autoplot(test_model)
par(mfrow=c(2,2)) plot(test_model) new_data <- train[-c(14,20),] model_newdata <-
lm(sleep.duration_kok ~ Stress.Level + Age + Physical.Activity.Level + Gender, data =
new_data) summary(model_newdata) autoplot(model_newdata)
defaultSummary(data.frame(obs=trainSleep.Duration, pred =
predict(model_newdata, train)))defaultSummary(data.frame(obs =
testSleep.Duration, pred=predict(model_newdata, test))>
```



grafikler

Yazılan kod yardımıyla 14 ve 20. satırda bulunan artık gözlemler eğitim veri setinden çıkarılıp yeni bir veri seti elde edilmiştir. Sonrasında yeniden birinci model üzerinden grafikler oluşturulmuştur.

Residuals vs Fitted grafiği için artıkların varyansının değişken olduğu görünmektedir. Çizgi eğimli ve doğrusal olmayan bir ilişki göstermektedir. Dolayısıyla, modelde doğrusal olmayan bir bileşenin olabileceği kanısına varılmaktadır.

İşlem sonrası bir önceki Q-Q Plot grafiğine göre yeni grafikte dağılımın daha düzgün olduğu söylenebilir. Ancak burada da bazı noktalarda çizgiden sapmalar vardır.

Scale Location Plot grafiği için elde edilen mavi eğim çizgisinin sabit ve yatay olması beklenir. Ancak burada çizgi hafif bir eğim göstermekte, bu da varyansın değişken olduğunu söylemektedir.

Residuals vs Leverage grafiği için Leverage değerleri yüksek olan noktalar, modelin tahminlerini önemli ölçüde etkileyebilecek gözlemlerdir. Burada bazı yüksek leverage değerlerine sahip noktalar göze çarpmaktadır.

CART YÖNTEMİ İLE TEST VE EĞİTİM VERİ SETİ MODELLERİNİ GÖRSELLEŞTİRME

```
# CART - Nicel değişkenler için (train veri seti ile);
```

```
library(rpart)
```

```
library(rpart.plot)
```

```
cart<-
```

```
rpart(sleep.duration_kok~Age+Stress.Level+Gender+Physical.Activity.Level,  
data=train)
```

```
cart$variable.importance
```

```
##          Stress.Level Physical.Activity.Level          Age  
##          5.2120505          1.8921633          1.8511408  
##          Gender  
##          0.5219144
```

```
cart
```

```
## n= 262
```

```
##
```

```
## node), split, n, deviance, yval
```

```
##      * denotes terminal node
```

```
##
```

```
## 1) root 262 6.25346600 2.673268
```

```
## 2) Stress.Level=7,8,9,10 85 0.20562760 2.494923
```

```
## 4) Stress.Level=8,9,10 52 0.01855474 2.460775 *
```

```
## 5) Stress.Level=1,2,3,4,5,6,7 33 0.03088778 2.548732 *
```

```
## 3) Stress.Level=1,2,3,4,5,6 177 2.04591300 2.758914
```

```
## 6) Age< 48.5 118 0.88587640 2.704068
```

```
## 12) Physical.Activity.Level< 57.5 27 0.08968956 2.577038 *
```

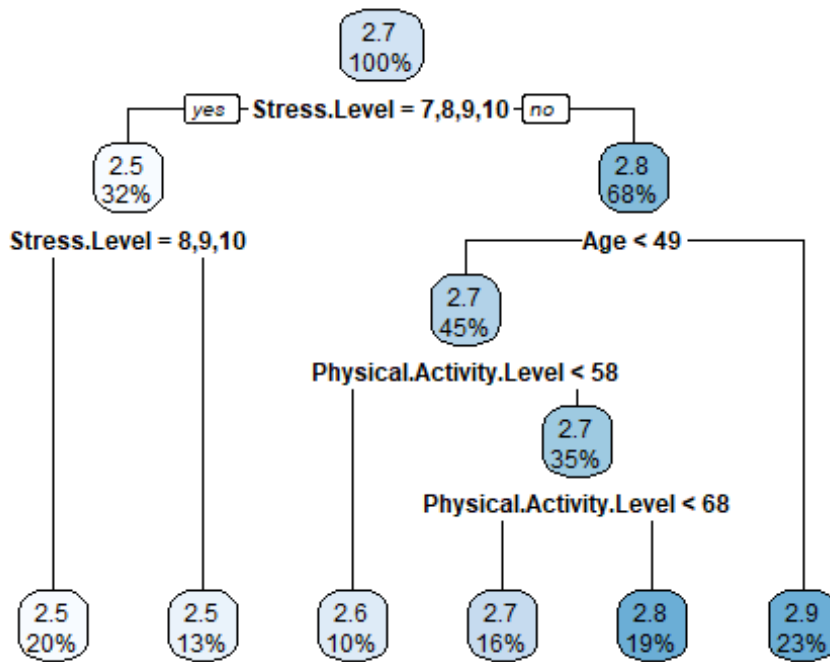
```
## 13) Physical.Activity.Level>=57.5 91 0.23122330 2.741758
```

```
## 26) Physical.Activity.Level< 67.5 41 0.01620731 2.689563 *
```

```
## 27) Physical.Activity.Level>=67.5 50 0.01172676 2.784558 *
```

```
## 7) Age>=48.5 59 0.09518774 2.868605 *
```

```
rpart.plot(cart)
```



CART - Kategorik değişkenler için (test veri seti ile);

```
library(rpart)
```

```
library(rpart.plot)
```

```
cart<-rpart(Sleep.Duration_log~Stress.Level+Gender+Quality.of.Sleep+Age_log,
data=test)
```

```
cart$variable.importance
```

```
## Quality.of.Sleep      Stress.Level      Age_log      Gender
##      0.929479874      0.893730219      0.396858935      0.009618902
```

```
cart
```

```
## n= 112
```

```
##
```

```
## node), split, n, deviance, yval
```

```
##      * denotes terminal node
```

```
##
```

```
## 1) root 112 1.148913000 1.946894
```

```
## 2) Quality.of.Sleep=1,2,3,4,5,6 34 0.033682150 1.823104
```

```
## 4) Stress.Level=8,9,10 18 0.003300778 1.797208 *
```

```
## 5) Stress.Level=1,2,3,4,5,6,7 16 0.004730961 1.852237 *
```

```
## 3) Quality.of.Sleep=7,8,9,10 78 0.367100600 2.000854
```

```
## 6) Quality.of.Sleep=1,2,3,4,5,6,7,8 66 0.197199400 1.981395
```

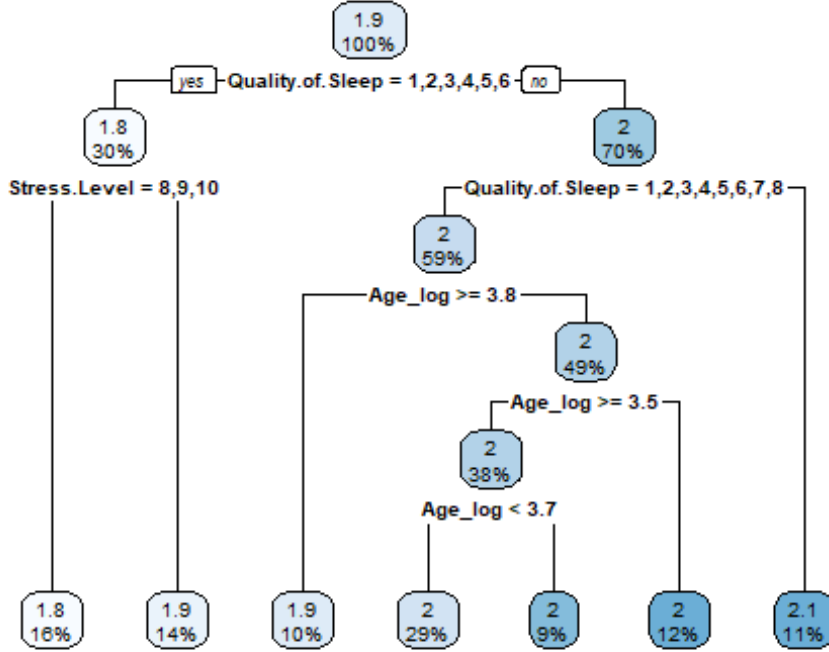
```
## 12) Age_log>=3.772695 11 0.010075440 1.897560 *
```

```
## 13) Age_log< 3.772695 55 0.094349090 1.998162
```

```
## 26) Age_log>=3.510542 42 0.053717690 1.983638
```

```
##          52) Age_log< 3.701226 32 0.011048760 1.970867 *
##          53) Age_log>=3.701226 10 0.020749110 2.024505 *
##          27) Age_log< 3.510542 13 0.003145547 2.045087 *
##          7) Quality.of.Sleep=9,10 12 0.007455788 2.107880 *
```

`rpart.plot(cart)`



Her iki grafik de CART yöntemini kullanarak regresyon ağaçlarını görselleştirmiştir. Bu ağaçlar, veri setlerindeki değişkenlerin nasıl bölündüğünü ve her bölünme noktasında hangi değişkenlerin etkili olduğunu görsel olarak sunmaktadır. Ağaçların derinliği arttıkça modelin daha fazla detayı öğrenmeye çalıştığını ve değişkenler arasındaki ilişkileri nasıl kullandığını görmek mümkündür.

Ağaçlar belirli faktörlere göre gruplandırmalar yaparak ortalama değerleri göstermektedir. İlk karar ağacında; stres seviyesi, uyku bozukluğu ve fiziksel aktivite gibi değişkenler dikkate alınmıştır. İkinci karar ağacında ise uyku kalitesi, stres seviyesi ve logaritmik olarak dönüştürülmüş yaş dikkate alınmıştır. Karar ağaçları, belirli grupların ortalama değerlerini ve bu grupların yüzdelik dağılımlarını göstermektedir. Renkler ve yüzdelikler, bu grupların ortalama değerlerini ve büyüklüklerini daha iyi anlamamıza yardımcı olmaktadır.

SONUÇ

Yapılan incelemeler sonunda analize başlamadan önce beklenildiği gibi uyku bozukluklarının; bireylerin stres seviyesi, uyku kalitesi, fiziksel aktivite seviyesi, kalp atış hızı ve hatta yaptıkları meslekler ile ilgisi olduğu kanısına varılmıştır. Sonuçlar farklı

grafikler ile görselleştirilmiş, regresyon modelleri yardımı ile matematiksel şekilde ifade edilmiştir.