RProgramlama.Devamsızlık

2023-02-03

R Markdown

```
##install packages tidyverse
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggpubr)
library(rstatix)
```

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

```
library(car)
```

```
## Zorunlu paket yükleniyor: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(broom)
library(dplyr)
library(tidyr)
library(haven)
```

```
##Veri çağırılır
library(readxl)
devamsızlık <- read_excel("devamsızlık.xlsx")
View(devamsızlık)
```

```
library(readxl)
data <- read_excel("devamsızlık.xlsx")
data<-as.data.frame(data)
```

```
class(data)
```

```
## [1] "data.frame"
```

```
data <- data.frame
```

```
##Değişken isimleri
names(devamsızlık)
```

```
##  [1] "ID"                          "Reason_for_absence"
##  [3] "Month_of_absence"            "Day_of_the_week"
##  [5] "Seasons"                     "Transportation_expense"
##  [7] "Distance_from_Residence_to_Work" "Service_time"
##  [9] "Age"                         "Work_load_Average_day"
## [11] "Hit_target"                  "Disciplinary_failure"
## [13] "Education"                   "Son"
## [15] "Social_drinker"              "Social_smoker"
## [17] "Pet"                         "Weight"
## [19] "Height"                      "Body_mass_index"
## [21] "Absenteeism_time_in_hours"
```

```
head(devamsızlık)
```

```
## # A tibble: 6 x 21
##       ID Reason_f~1 Month~2 Day_o~3 Seasons Trans~4 Dista~5 Servi~6   Age Work_~7
##    <dbl>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <dbl>   <dbl>
## 1     11         26       7       3       1     289      36      13    33  239554
## 2     36          0       7       3       1     118      13      18    50  239554
## 3      3         23       7       4       1     179      51      18    38  239554
## 4      7          7       7       5       1     279       5      14    39  239554
## 5     11         23       7       5       1     289      36      13    33  239554
## 6      3         23       7       6       1     179      51      18    38  239554
## # ... with 11 more variables: Hit_target <dbl>, Disciplinary_failure <dbl>,
## #   Education <dbl>, Son <dbl>, Social_drinker <dbl>, Social_smoker <dbl>,
## #   Pet <dbl>, Weight <dbl>, Height <dbl>, Body_mass_index <dbl>,
## #   Absenteeism_time_in_hours <dbl>, and abbreviated variable names
## #   1: Reason_for_absence, 2: Month_of_absence, 3: Day_of_the_week,
## #   4: Transportation_expense, 5: Distance_from_Residence_to_Work,
## #   6: Service_time, 7: Work_load_Average_day
```

```
#Eksik gözlemler incelenmistir.
colSums(is.na(devamsızlık))
```

```
##                            ID              Reason_for_absence
##                             0                               0
##               Month_of_absence                 Day_of_the_week
##                             0                               0
##                        Seasons          Transportation_expense
##                             0                               0
## Distance_from_Residence_to_Work                    Service_time
##                             0                               0
##                            Age           Work_load_Average_day
##                             0                               0
##                     Hit_target            Disciplinary_failure
##                             0                               0
##                     Education                             Son
##                             0                               0
##                 Social_drinker                   Social_smoker
##                             0                               0
##                           Pet                          Weight
##                             0                               0
##                        Height                 Body_mass_index
##                             0                               0
##      Absenteeism_time_in_hours
##                             0
```

```
numeric <- c(devamsızlık$ID)
numeric <- c(devamsızlık$Reason_for_absence)
numeric <- c(devamsızlık$Month_of_absence)
numeric <- c(devamsızlık$Day_of_the_week)
numeric <- c(devamsızlık$Seasons)
numeric <- c(devamsızlık$Transportation_expense)
numeric <- c(devamsızlık$Distance_from_Residence_to_Work)
numeric <- c(devamsızlık$Service_time)
numeric <- c(devamsızlık$Age)
numeric <- c(devamsızlık$Work_load_Average_day)
numeric <- c(devamsızlık$Hit_target)
numeric <- c(devamsızlık$Disciplinary_failure)
numeric <- c(devamsızlık$Education)
numeric <- c(devamsızlık$Son)
numeric <- c(devamsızlık$Social_drinker)
numeric <- c(devamsızlık$Social_smoker)
numeric <- c(devamsızlık$Pet)
numeric <- c(devamsızlık$Weight)
numeric <- c(devamsızlık$Height)
numeric <- c(devamsızlık$Body_mass_index)
numeric <- c(devamsızlık$Absenteeism_time_in_hours)
```

```
#Verideki kategorik değişkenler faktör olarak tanımlanır.
devamsızlık$Seasons <- factor(devamsızlık$Seasons, levels=c("1","2","3","4"),labels=c("yaz","sonbahar","kış","ilk
bahar"))
devamsızlık$Day_of_the_week <- factor(devamsızlık$Day_of_the_week, levels=c("2","3","4","5","6"),labels=c("pazart
esi","salı","carsamba","persembe","cuma"))
devamsızlık$Disciplinary_failure <- factor(devamsızlık$Disciplinary_failure, levels=c("0","1"),labels=c("var","yo
k"))
devamsızlık$Education <- factor(devamsızlık$Education, levels=c("1","2","3","4"),labels=c("lise","mezun","yüksek
lisans","doktora"))
```

```
library(dplyr)
glimpse(devamsızlık)
```

```
## Rows: 740
## Columns: 21
## $ ID                              <dbl> 11, 36, 3, 7, 11, 3, 10, 20, 14, 1, 20~
## $ Reason_for_absence              <dbl> 26, 0, 23, 7, 23, 23, 22, 23, 19, 22, ~
## $ Month_of_absence                <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,~
## $ Day_of_the_week                 <fct> salı, salı, carsamba, persembe, persem~
## $ Seasons                         <fct> yaz, yaz, yaz, yaz, yaz, yaz, yaz, yaz~
## $ Transportation_expense          <dbl> 289, 118, 179, 279, 289, 179, 361, 260~
## $ Distance_from_Residence_to_Work <dbl> 36, 13, 51, 5, 36, 51, 52, 50, 12, 11,~
## $ Service_time                    <dbl> 13, 18, 18, 14, 13, 18, 3, 11, 14, 14,~
## $ Age                             <dbl> 33, 50, 38, 39, 33, 38, 28, 36, 34, 37~
## $ Work_load_Average_day           <dbl> 239554, 239554, 239554, 239554, 239554~
## $ Hit_target                      <dbl> 97, 97, 97, 97, 97, 97, 97, 97, 97, 97~
## $ Disciplinary_failure            <fct> var, yok, var, var, var, var, var, var~
## $ Education                       <fct> lise, lise, lise, lise, lise, lise, li~
## $ Son                             <dbl> 2, 1, 0, 2, 2, 0, 1, 4, 2, 1, 4, 4, 4,~
## $ Social_drinker                  <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,~
## $ Social_smoker                   <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Pet                             <dbl> 1, 0, 0, 0, 1, 0, 4, 0, 0, 1, 0, 0, 0,~
## $ Weight                          <dbl> 90, 98, 89, 68, 90, 89, 80, 65, 95, 88~
## $ Height                          <dbl> 172, 178, 170, 168, 172, 170, 172, 168~
## $ Body_mass_index                 <dbl> 30, 31, 31, 24, 30, 31, 27, 23, 25, 29~
## $ Absenteeism_time_in_hours       <dbl> 4, 0, 2, 4, 2, 2, 8, 4, 40, 8, 8, 8, 8~
```

```
summary(devamsızlık)
```

```
##        ID          Reason_for_absence Month_of_absence  Day_of_the_week
## Min.   : 1.00   Min.   : 0.00      Min.   : 0.000   pazartesi:161
## 1st Qu.: 9.00   1st Qu.:13.00      1st Qu.: 3.000   salı     :154
## Median :18.00   Median :23.00      Median : 6.000   carsamba :156
## Mean   :18.02   Mean   :19.22      Mean   : 6.324   persembe :125
## 3rd Qu.:28.00   3rd Qu.:26.00      3rd Qu.: 9.000   cuma     :144
## Max.   :36.00   Max.   :28.00      Max.   :12.000
##     Seasons      Transportation_expense Distance_from_Residence_to_Work
## yaz     :170   Min.   :118.0          Min.   : 5.00
## sonbahar:192   1st Qu.:179.0          1st Qu.:16.00
## kıs     :183   Median :225.0          Median :26.00
## ilkbahar:195   Mean   :221.3          Mean   :29.63
##               3rd Qu.:260.0          3rd Qu.:50.00
##               Max.   :388.0          Max.   :52.00
##   Service_time        Age          Work_load_Average_day   Hit_target
## Min.   : 1.00   Min.   :27.00   Min.   :205917        Min.   : 81.00
## 1st Qu.: 9.00   1st Qu.:31.00   1st Qu.:244387        1st Qu.: 93.00
## Median :13.00   Median :37.00   Median :264249        Median : 95.00
## Mean   :12.55   Mean   :36.45   Mean   :271490        Mean   : 94.59
## 3rd Qu.:16.00   3rd Qu.:40.00   3rd Qu.:294217        3rd Qu.: 97.00
## Max.   :29.00   Max.   :58.00   Max.   :378884        Max.   :100.00
## Disciplinary_failure        Education         Son          Social_drinker
## var:700              lise        :611   Min.   :0.000   Min.   :0.0000
## yok: 40              mezun       : 46   1st Qu.:0.000   1st Qu.:0.0000
##                      yüksek lisans: 79  Median :1.000   Median :1.0000
##                      doktora      :  4  Mean   :1.019   Mean   :0.5676
##                                         3rd Qu.:2.000   3rd Qu.:1.0000
##                                         Max.   :4.000   Max.   :1.0000
##  Social_smoker         Pet            Weight           Height
## Min.   :0.00000   Min.   :0.0000   Min.   : 56.00   Min.   :163.0
## 1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.: 69.00   1st Qu.:169.0
## Median :0.00000   Median :0.0000   Median : 83.00   Median :170.0
## Mean   :0.07297   Mean   :0.7459   Mean   : 79.04   Mean   :172.1
## 3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.: 89.00   3rd Qu.:172.0
## Max.   :1.00000   Max.   :8.0000   Max.   :108.00   Max.   :196.0
##  Body_mass_index Absenteeism_time_in_hours
## Min.   :19.00   Min.   :  0.000
## 1st Qu.:24.00   1st Qu.:  2.000
## Median :25.00   Median :  3.000
## Mean   :26.68   Mean   :  6.924
## 3rd Qu.:31.00   3rd Qu.:  8.000
## Max.   :38.00   Max.   :120.000
```

```r
library(funModeling)
```

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

```
## Zorunlu paket yükleniyor: lattice
```

```
## Zorunlu paket yükleniyor: survival
```

```
## Zorunlu paket yükleniyor: Formula
```

```
##
## 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.1.9.4 :)
## Examples and tutorials at livebook.datascienceheroes.com
##  / Now in Spanish: librovivodecienciadedatos.ai
```

```r
profiling_num(devamsızlık)
```

```
##                              variable          mean       std_dev variation_coef
## 1                                  ID  1.801757e+01  1.102125e+01     0.61169452
## 2                   Reason_for_absence  1.921622e+01  8.433406e+00     0.43886922
## 3                    Month_of_absence  6.324324e+00  3.436287e+00     0.54334451
## 4               Transportation_expense  2.213297e+02  6.695222e+01     0.30249991
## 5  Distance_from_Residence_to_Work  2.963108e+01  1.483679e+01     0.50071708
## 6                        Service_time  1.255405e+01  4.384873e+00     0.34927947
## 7                                 Age  3.645000e+01  6.478772e+00     0.17774410
## 8                Work_load_Average_day  2.714902e+05  3.905812e+04     0.14386564
## 9                          Hit_target  9.458784e+01  3.779313e+00     0.03995559
## 10                                Son  1.018919e+00  1.098489e+00     1.07809267
## 11                      Social_drinker  5.675676e-01  4.957487e-01     0.87346194
## 12                      Social_smoker  7.297297e-02  2.602681e-01     3.56663624
## 13                                Pet  7.459459e-01  1.318258e+00     1.76723032
## 14                             Weight  7.903514e+01  1.288321e+01     0.16300612
## 15                             Height  1.721149e+02  6.034995e+00     0.03506376
## 16                     Body_mass_index  2.667703e+01  4.285452e+00     0.16064205
## 17           Absenteeism_time_in_hours  6.924324e+00  1.333100e+01     1.92524172
##       p_01    p_05    p_25    p_50    p_75    p_95      p_99   skewness   kurtosis
## 1        1       3       9      18      28      34     36.00  0.016572227   1.748527
## 2        0       0      13      23      26      28     28.00 -0.913455957   2.733731
## 3        1       1       3       6       9      12     12.00  0.069227850   1.745400
## 4      118     118     179     225     260     361    378.00  0.395385099   2.675758
## 5       10      10      16      26      50      51     52.00  0.311449828   1.738729
## 6        3       4       9      13      16      18     18.00 -0.004709991   3.670407
## 7       28      28      31      37      40      50     56.05  0.696288348   3.420605
## 8   205917  222196  244387  264249  294217  343253 378884.00  0.959506612   3.605922
## 9       81      88      93      95      97      99    100.00 -1.259149221   5.394631
## 10       0       0       0       1       2       4      4.00  1.084261825   3.735182
## 11       0       0       0       1       1       1      1.00 -0.272772363   1.074405
## 12       0       0       0       0       0       1      1.00  3.283659682  11.782421
## 13       0       0       0       0       1       4      6.83  2.730166944  12.601483
## 14      56      56      69      83      89      98    106.00  0.016966890   2.084139
## 15     165     167     169     170     172     182    196.00  2.560855289  10.259791
## 16      19      19      24      25      31      32     38.00  0.304426971   2.679648
## 17       0       0       2       3       8      24     80.00  5.709125258  41.507696
##      iqr                range_98                  range_80
## 1     19                 [1, 36]                   [3, 34]
## 2     13                 [0, 28] [6.90000000000001, 28]
## 3      6                 [1, 12]                   [2, 11]
## 4     81              [118, 378]               [118, 291]
## 5     34                [10, 52]                  [11, 51]
## 6      7                 [3, 18] [8.90000000000001, 18]
## 7      9  [28, 56.0500000000001]                  [28, 47]
## 8  49830          [205917, 378884]         [230290, 326452]
## 9      4                [81, 100]                  [91, 99]
## 10     2                  [0, 4]                    [0, 2]
## 11     1                  [0, 1]                    [0, 1]
## 12     0                  [0, 1]                    [0, 0]
## 13     1  [0, 6.83000000000004]                    [0, 2]
## 14    20                [56, 106]                  [65, 95]
## 15     3              [165, 196]               [168, 178]
## 16     7                [19, 38]                  [22, 31]
## 17     6                 [0, 80]                    [1, 8]
```
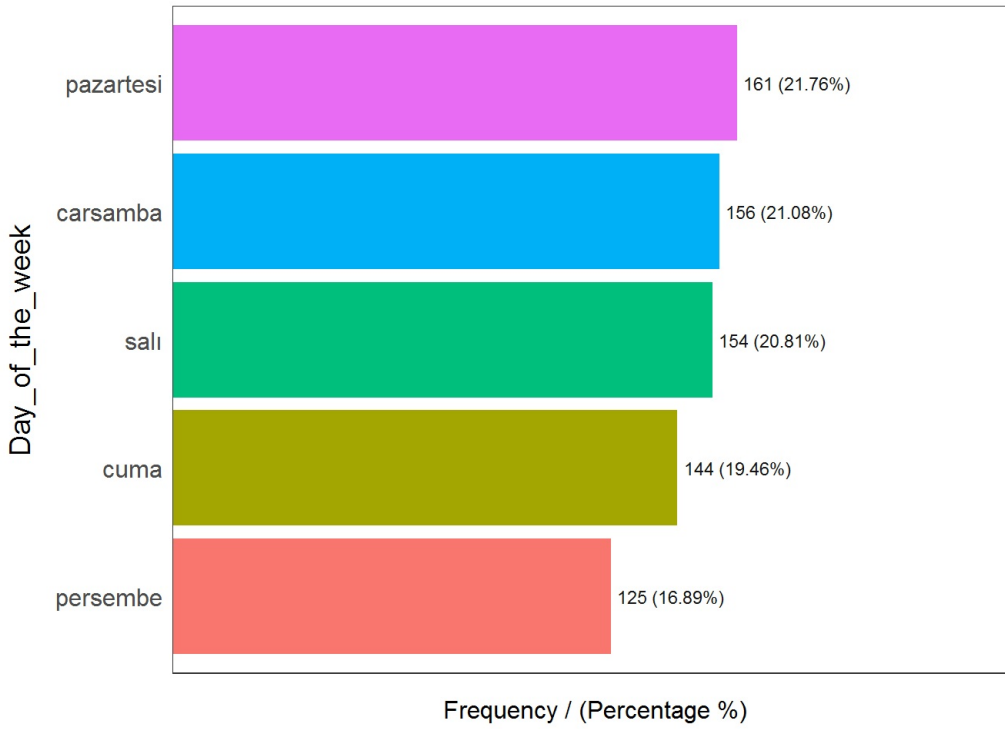
```
plot_num(devamsızlık)
```

```
## 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 <]8;;https://github.com/pablo14/funModeling/issueshttps://github.com/pablo14/funM
odeling/issues]8;;>.
```
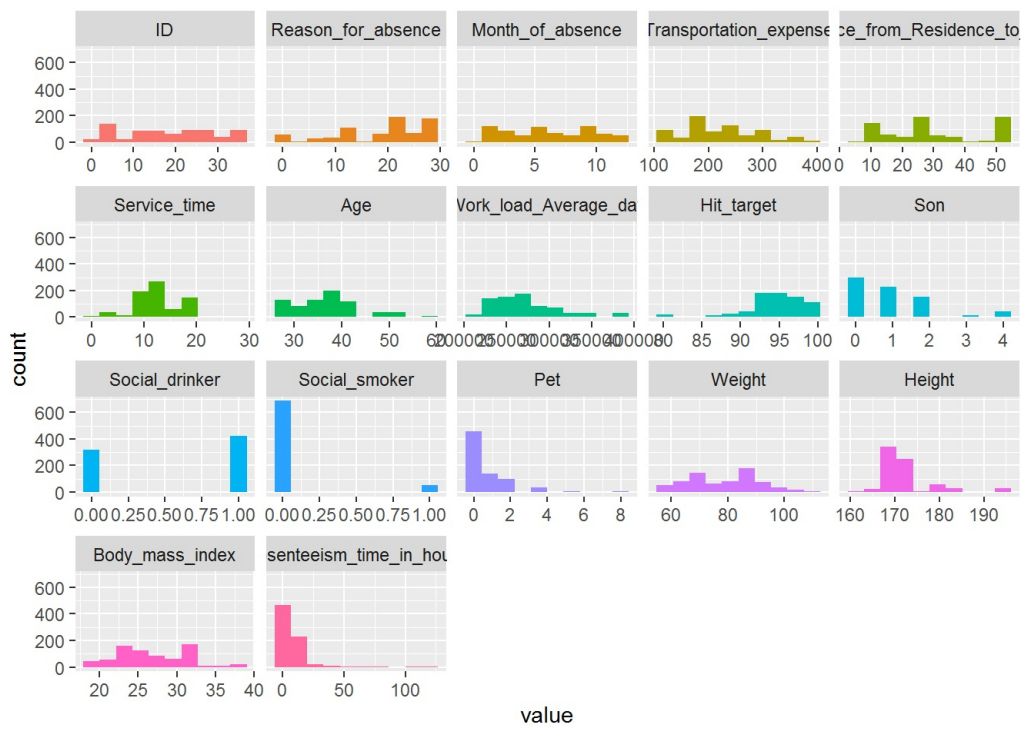
```
freq(devamsızlık)
```



```
##    Day_of_the_week frequency percentage cumulative_perc
## 1        pazartesi       161      21.76           21.76
## 2         carsamba       156      21.08           42.84
## 3             salı       154      20.81           63.65
## 4             cuma       144      19.46           83.11
## 5         persembe       125      16.89          100.00
```
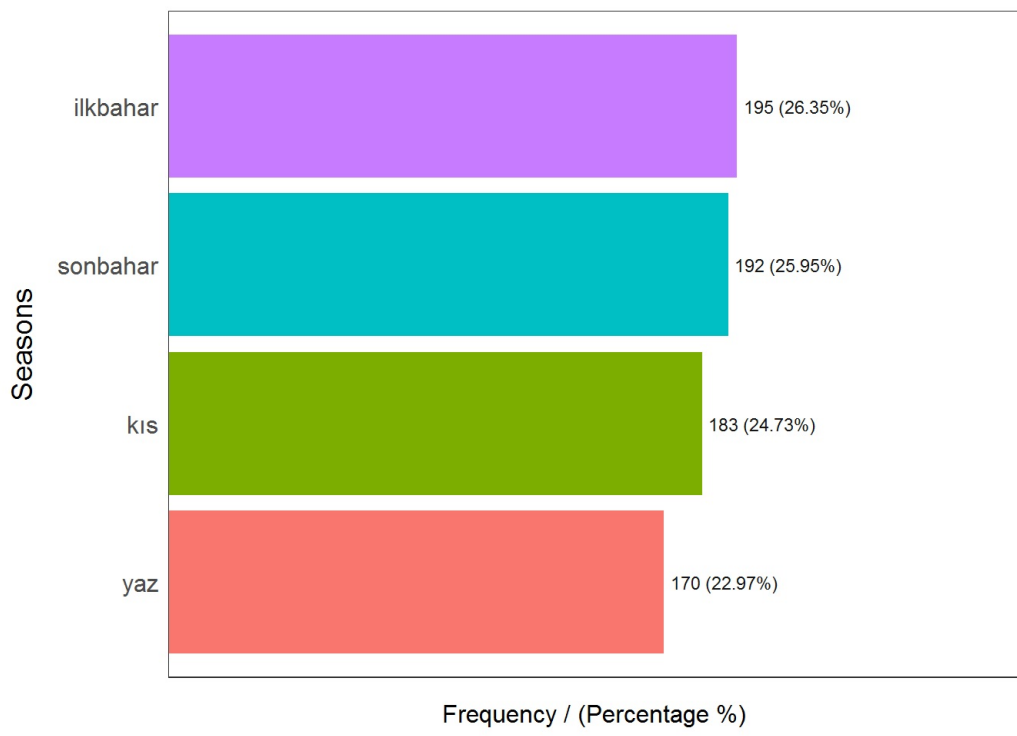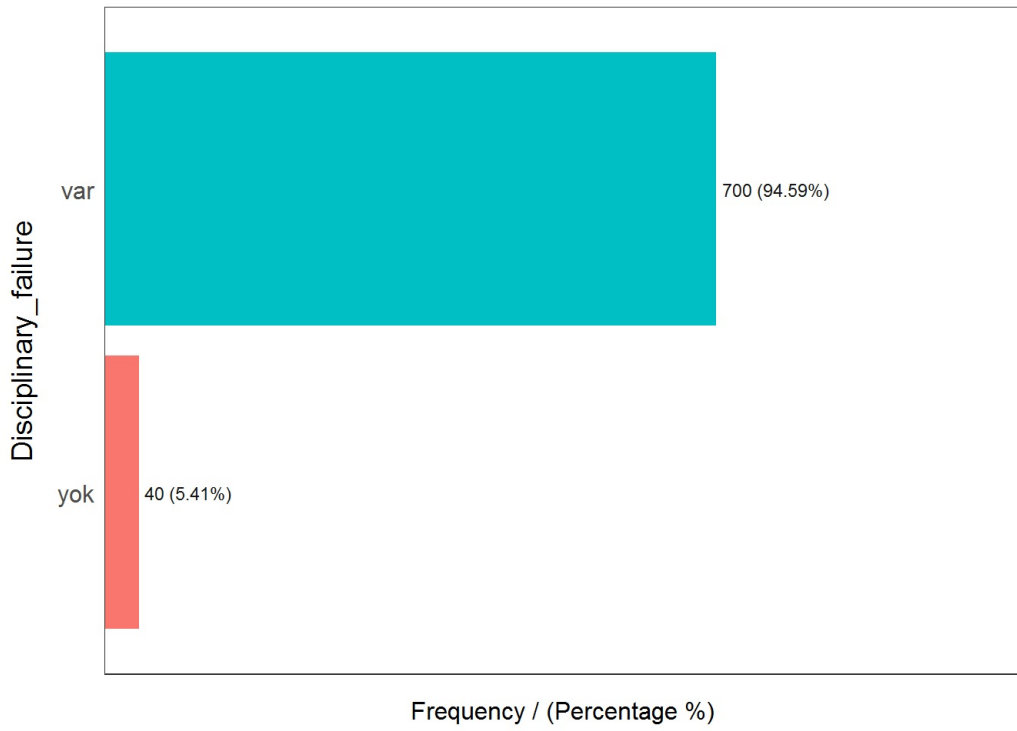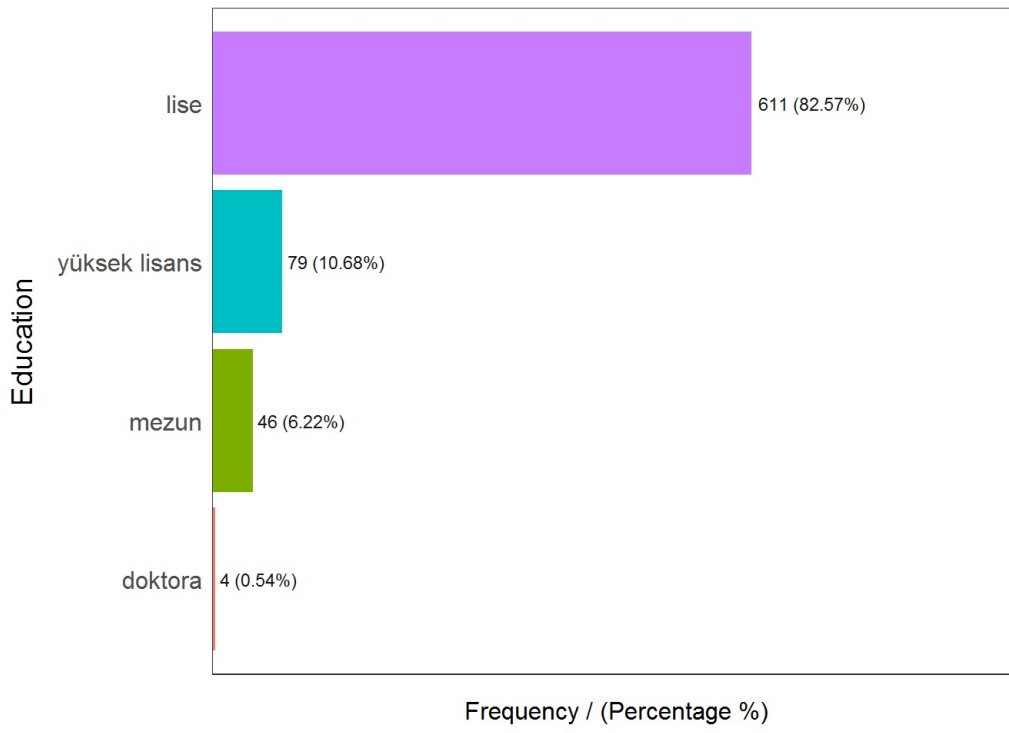
```
##     Seasons  frequency  percentage  cumulative_perc
## 1  ilkbahar       195       26.35            26.35
## 2  sonbahar       192       25.95            52.30
## 3       kıs       183       24.73            77.03
## 4       yaz       170       22.97           100.00
```



```
##   Disciplinary_failure  frequency  percentage  cumulative_perc
## 1                  var        700       94.59            94.59
## 2                  yok         40        5.41           100.00
```

```
##         Education frequency percentage cumulative_perc
## 1           lise       611      82.57           82.57
## 2 yüksek lisans        79      10.68           93.25
## 3          mezun        46       6.22           99.47
## 4        doktora         4       0.54          100.00
```

```
## [1] "Variables processed: Day_of_the_week, Seasons, Disciplinary_failure, Education"
```

```
geom_histogram()
```

```
## geom_bar: na.rm = FALSE, orientation = NA
## stat_bin: binwidth = NULL, bins = NULL, na.rm = FALSE, orientation = NA, pad = FALSE
## position_stack
```

```
ggplot(data=devamsızlık, aes(x=Seasons,y=Age))+
  geom_boxplot()
```
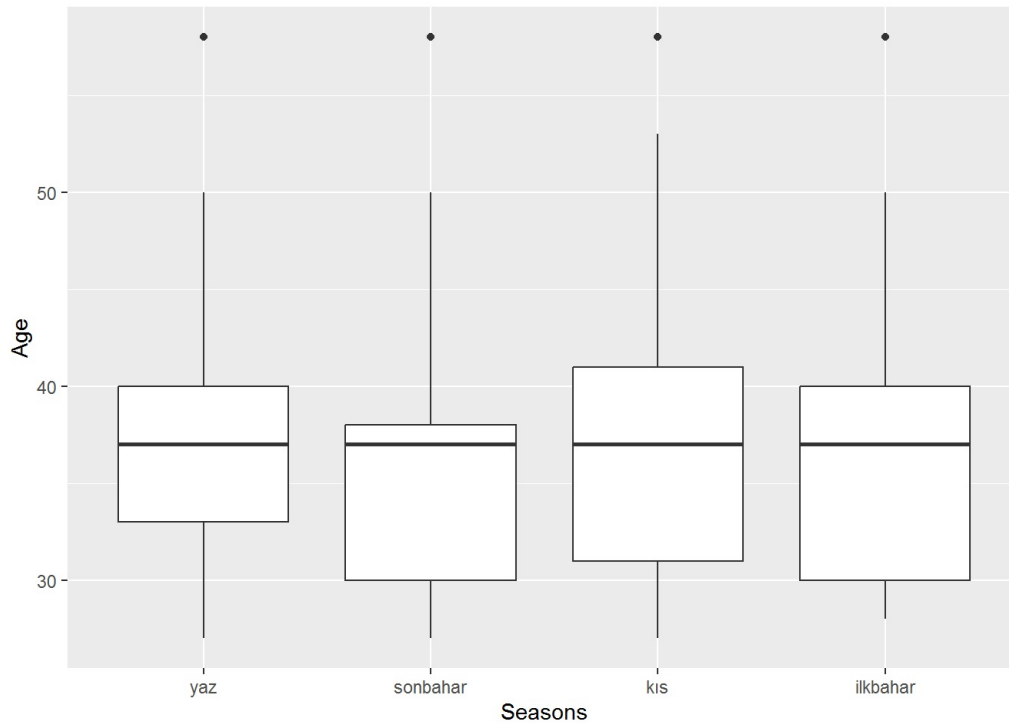
```
ggplot(data=devamsızlık, aes(x=Seasons,y=Age))+
  geom_boxplot()+
  scale_fill_manual(values = c("seagreen","red","skyblue","orange"))+
  labs(x="Seansons",y="Age")+
  theme_bw()
```
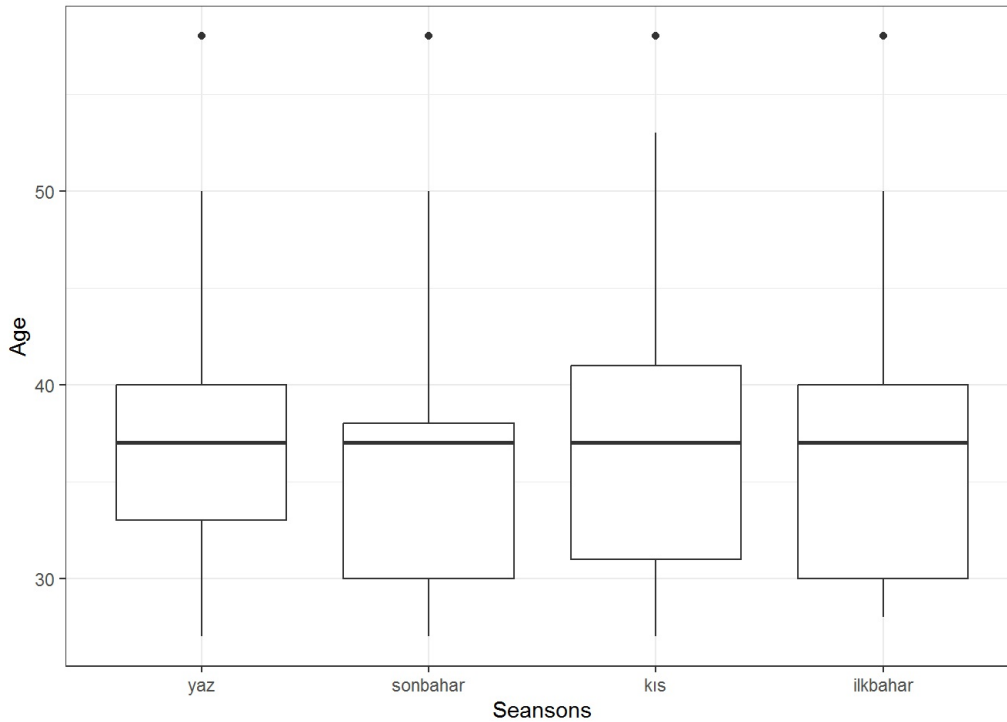


```
#install.packages("gapminder")
library(gapminder)
head(gapminder)
```

```
## # A tibble: 6 x 6
##   country     continent  year lifeExp      pop gdpPercap
##   <fct>       <fct>     <int>  <dbl>    <int>     <dbl>
## 1 Afghanistan Asia       1952   28.8  8425333      779.
## 2 Afghanistan Asia       1957   30.3  9240934      821.
## 3 Afghanistan Asia       1962   32.0 10267083      853.
## 4 Afghanistan Asia       1967   34.0 11537966      836.
## 5 Afghanistan Asia       1972   36.1 13079460      740.
## 6 Afghanistan Asia       1977   38.4 14880372      786.
```

```
ggplot(devamsızlık,aes(x=Age))+
  geom_histogram(binwidth = 5000)
```
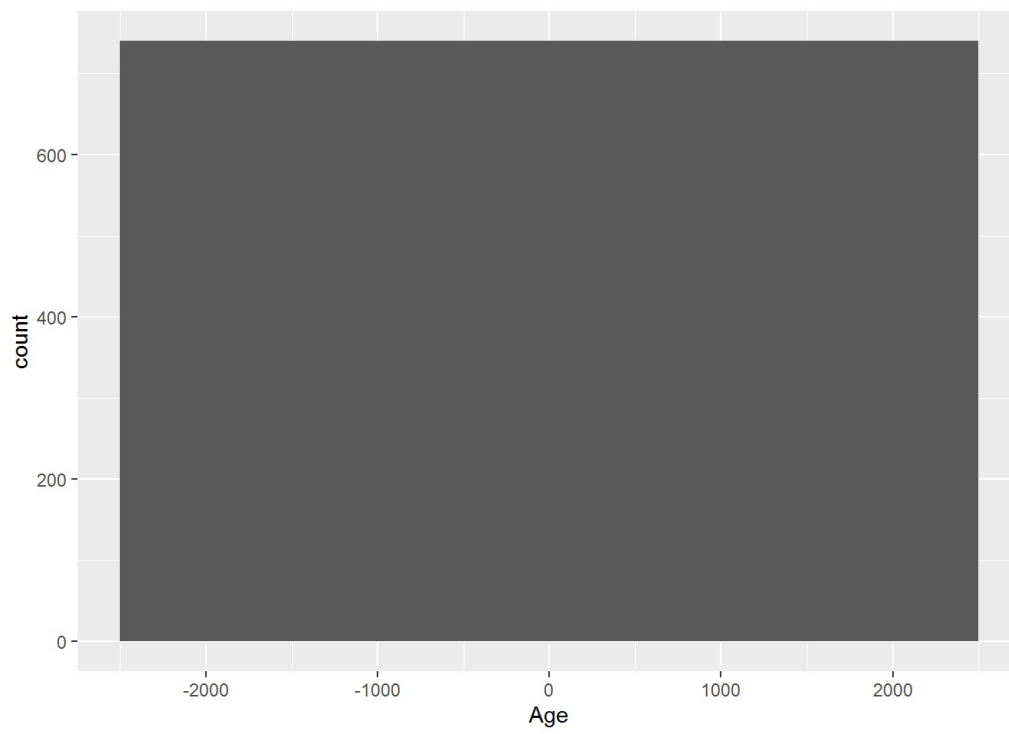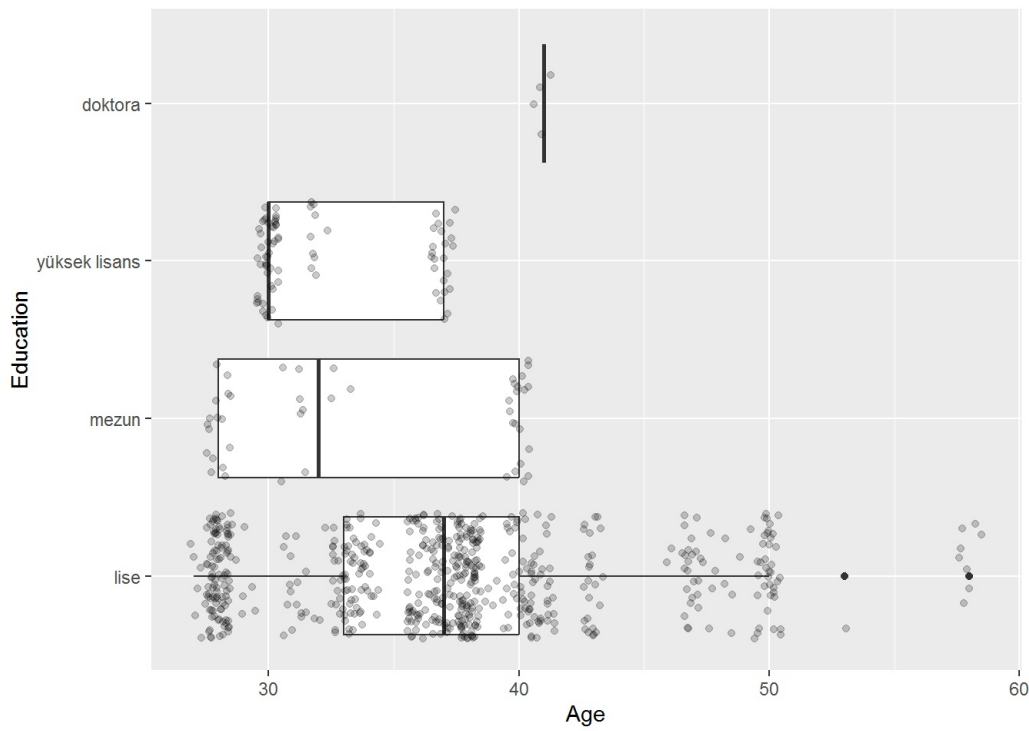
```
ggplot(devamsızlık,aes(x=Age))+
  geom_density(color="Blue")
```
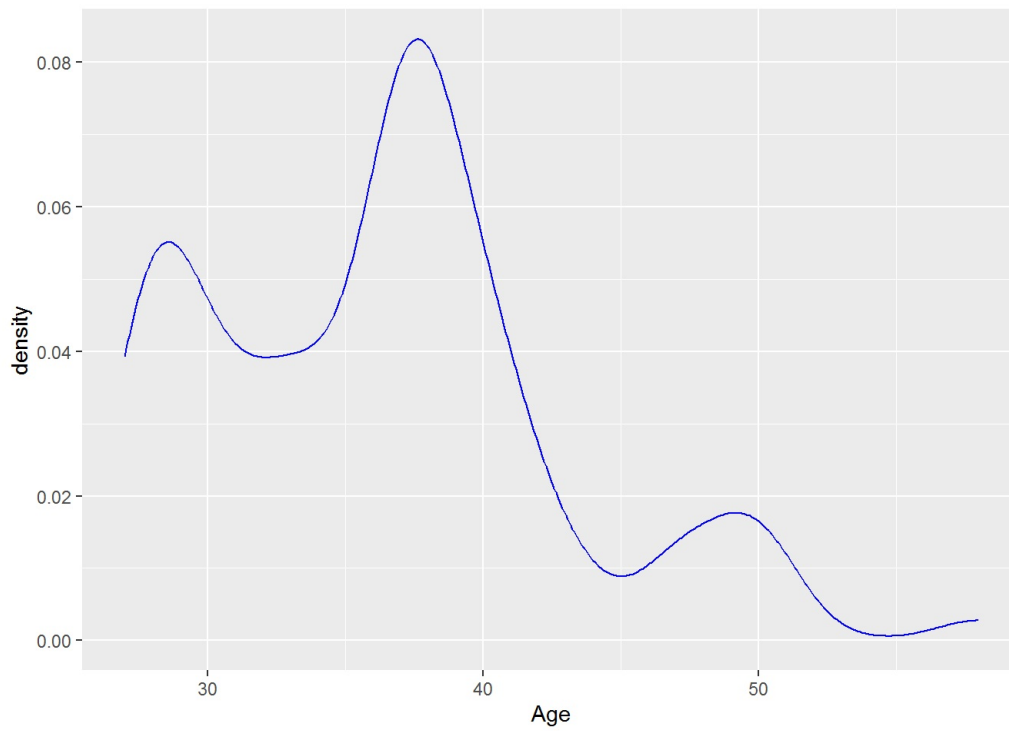


```
ggplot(devamsızlık,aes(x=Age, y=Education))+
  geom_boxplot()+
  geom_point()
```
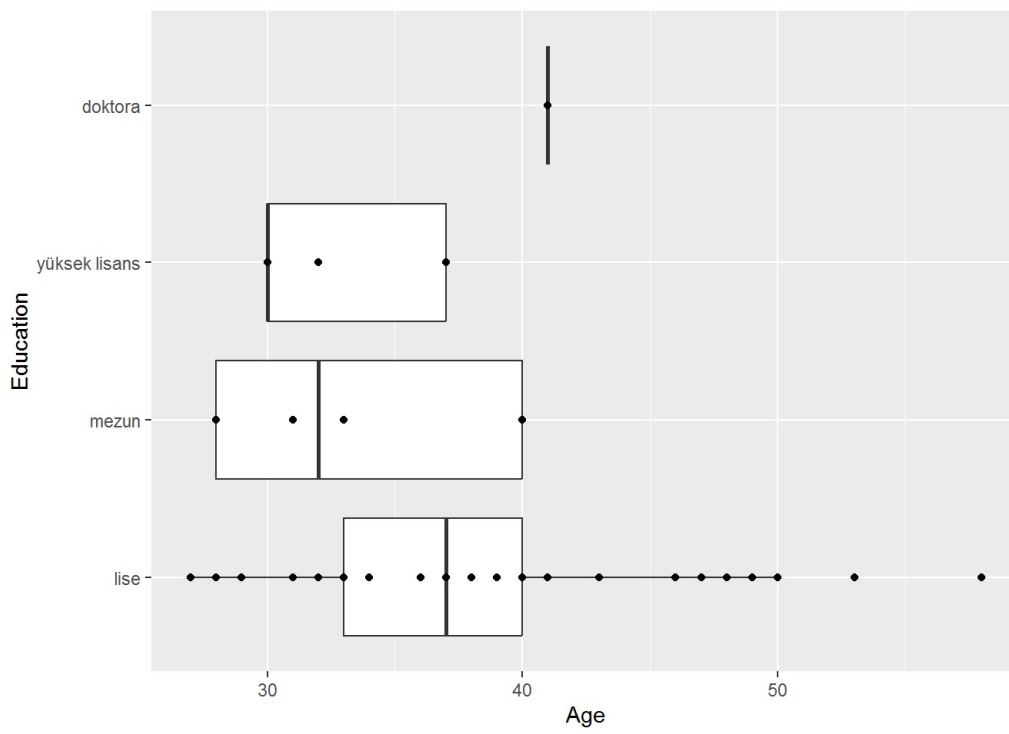
```
ggplot(devamsızlık,aes(x=Age, y=Education))+
  geom_boxplot()+
  geom_jitter(width=0.5,alpha=0.2)
```



```
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
```
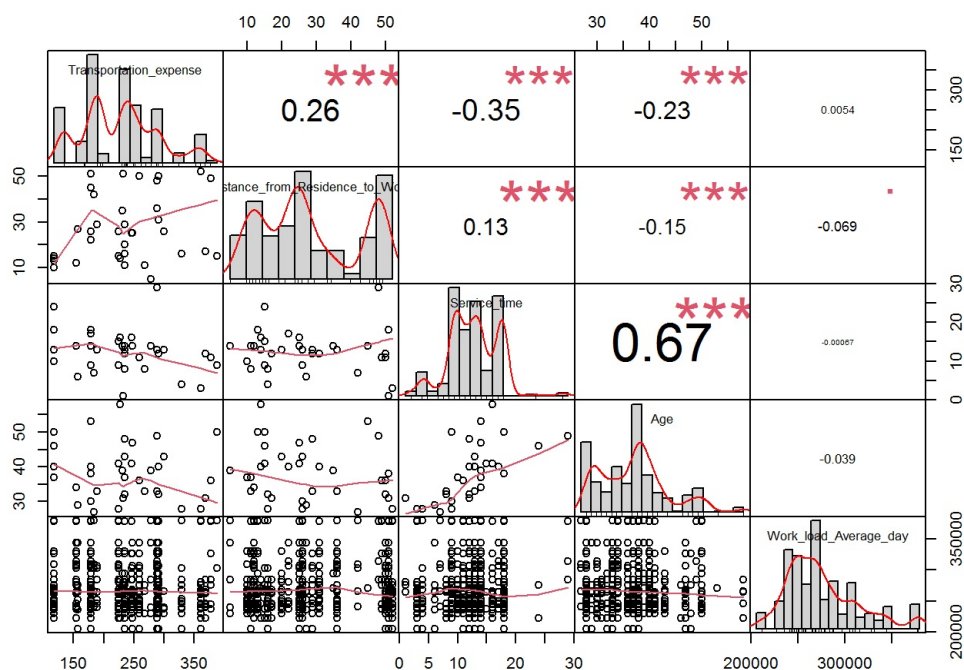
```
##
## Attaching package: 'xts'
```

```
chart.Correlation(devamsızlık[,6:10], histogram=TRUE, pch=19)
```

```
## 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

## 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

## 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

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



```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
corrplot(cor(devamsızlık[,6:10]))
```