keşifçi veri analizi ve görselleştirme

ilke

2022-08-05

veri setinin 6 boyutlu gösterimi 1.boyut: değişkenler 2.'': değişkenlerin dağılımı(histogram) 3.boyut : yoğunluk 4.boyut: ilişki 5.boyut: korelasyona ait p-value 6.boyut:scatter plot

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✔ ggplot2 3.3.6     ✔ purrr   0.3.4
## ✔ tibble  3.1.8     ✔ dplyr   1.0.9
## ✔ tidyr   1.2.0     ✔ stringr 1.4.0
## ✔ readr   2.1.2     ✔ forcats 0.5.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
df <- mpg

df$class <- factor(df$class)
glimpse(df)
```

```
## Rows: 234
## Columns: 11
## $ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi", "…
## $ model        <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro", "…
## $ displ        <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.…
## $ year         <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200…
## $ cyl          <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 8, 8, …
## $ trans        <chr> "auto(l5)", "manual(m5)", "manual(m6)", "auto(av)", "auto…
## $ drv          <chr> "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4", "4…
## $ cty          <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1…
## $ hwy          <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2…
## $ fl           <chr> "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p…
## $ class        <fct> compact, compact, compact, compact, compact, compact, com…
```

```
df$disp_yeni <- ifelse(df$displ > 2, "bir", "sifir")  #displ değerleri ikiden büyükse 1 değilse 0
df$disp_yeni <- factor(df$disp_yeni)
df$disp_iki <- cut(df$displ, breaks = c(0,2, max(df$displ)))
```

ilk bakış

```
colnames(mpg)
```

```
##  [1] "manufacturer" "model"        "displ"        "year"         "cyl"
##  [6] "trans"        "drv"          "cty"          "hwy"          "fl"
## [11] "class"
```

```
rownames(mpg)
```

```
##   [1] "1"   "2"   "3"   "4"   "5"   "6"   "7"   "8"   "9"   "10"  "11"  "12"
##  [13] "13"  "14"  "15"  "16"  "17"  "18"  "19"  "20"  "21"  "22"  "23"  "24"
##  [25] "25"  "26"  "27"  "28"  "29"  "30"  "31"  "32"  "33"  "34"  "35"  "36"
##  [37] "37"  "38"  "39"  "40"  "41"  "42"  "43"  "44"  "45"  "46"  "47"  "48"
##  [49] "49"  "50"  "51"  "52"  "53"  "54"  "55"  "56"  "57"  "58"  "59"  "60"
##  [61] "61"  "62"  "63"  "64"  "65"  "66"  "67"  "68"  "69"  "70"  "71"  "72"
##  [73] "73"  "74"  "75"  "76"  "77"  "78"  "79"  "80"  "81"  "82"  "83"  "84"
##  [85] "85"  "86"  "87"  "88"  "89"  "90"  "91"  "92"  "93"  "94"  "95"  "96"
##  [97] "97"  "98"  "99"  "100" "101" "102" "103" "104" "105" "106" "107" "108"
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"
## [145] "145" "146" "147" "148" "149" "150" "151" "152" "153" "154" "155" "156"
## [157] "157" "158" "159" "160" "161" "162" "163" "164" "165" "166" "167" "168"
## [169] "169" "170" "171" "172" "173" "174" "175" "176" "177" "178" "179" "180"
## [181] "181" "182" "183" "184" "185" "186" "187" "188" "189" "190" "191" "192"
## [193] "193" "194" "195" "196" "197" "198" "199" "200" "201" "202" "203" "204"
## [205] "205" "206" "207" "208" "209" "210" "211" "212" "213" "214" "215" "216"
## [217] "217" "218" "219" "220" "221" "222" "223" "224" "225" "226" "227" "228"
## [229] "229" "230" "231" "232" "233" "234"
```

```
nrow(mpg)
```

```
## [1] 234
```

```
ncol(mpg)
```

```
## [1] 11
```

```
head(mpg)
```

```
## # A tibble: 6 × 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa…
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa…
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa…
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa…
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa…
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa…
```

```
str(mpg)
```

```
## tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
##  $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
##  $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
##  $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
##  $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
##  $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
##  $ drv         : chr [1:234] "f" "f" "f" "f" ...
##  $ cty         : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
##  $ fl          : chr [1:234] "p" "p" "p" "p" ...
##  $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
```

```
glimpse(mpg)  #veri seti genel yapı hakkında bilgi
```

```
## Rows: 234
## Columns: 11
## $ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi", "…
## $ model        <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro", "…
## $ displ        <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.…
## $ year         <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200…
## $ cyl          <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 8, 8, …
## $ trans        <chr> "auto(l5)", "manual(m5)", "manual(m6)", "auto(av)", "auto…
## $ drv          <chr> "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4", "4…
## $ cty          <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1…
## $ hwy          <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2…
## $ fl           <chr> "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p…
## $ class        <chr> "compact", "compact", "compact", "compact", "compact", "c…
```

```
psych::headTail(mpg)  #ilk 4 son 4 gözlem
```

```
##   manufacturer  model displ year cyl      trans  drv cty hwy   fl   class
## 1         audi     a4   1.8 1999   4   auto(l5)    f  18  29    p compact
## 2         audi     a4   1.8 1999   4 manual(m5)    f  21  29    p compact
## 3         audi     a4     2 2008   4 manual(m6)    f  20  31    p compact
## 4         audi     a4     2 2008   4   auto(av)    f  21  30    p compact
## 5         <NA>   <NA>   ...  ... ...       <NA> <NA> ... ... <NA>    <NA>
## 6   volkswagen passat     2 2008   4 manual(m6)    f  21  29    p midsize
## 7   volkswagen passat   2.8 1999   6   auto(l5)    f  16  26    p midsize
## 8   volkswagen passat   2.8 1999   6 manual(m5)    f  18  26    p midsize
## 9   volkswagen passat   3.6 2008   6   auto(s6)    f  17  26    p midsize
```

```
summary(df)
```

```
##  manufacturer          model               displ            year
##  Length:234         Length:234         Min.   :1.600   Min.   :1999
##  Class :character   Class :character   1st Qu.:2.400   1st Qu.:1999
##  Mode  :character   Mode  :character   Median :3.300   Median :2004
##                                        Mean   :3.472   Mean   :2004
##                                        3rd Qu.:4.600   3rd Qu.:2008
##                                        Max.   :7.000   Max.   :2008
##
##       cyl           trans               drv                 cty
##  Min.   :4.000   Length:234         Length:234         Min.   : 9.00
##  1st Qu.:4.000   Class :character   Class :character   1st Qu.:14.00
##  Median :6.000   Mode  :character   Mode  :character   Median :17.00
##  Mean   :5.889                                         Mean   :16.86
##  3rd Qu.:8.000                                         3rd Qu.:19.00
##  Max.   :8.000                                         Max.   :35.00
##
##       hwy            fl                 class     disp_yeni     disp_iki
##  Min.   :12.00   Length:234         2seater  : 5   bir  :191   (0,2]: 43
##  1st Qu.:18.00   Class :character   compact  :47   sifir: 43   (2,7]:191
##  Median :24.00   Mode  :character   midsize  :41
##  Mean   :23.44                      minivan  :11
##  3rd Qu.:27.00                      pickup   :33
##  Max.   :44.00                      subcompact:35
##                                     suv      :62
```

```
fivenum(df$displ)
```

```
## [1] 1.6 2.4 3.3 4.6 7.0
```

```
levels(df$class)
```

```
## [1] "2seater"    "compact"    "midsize"    "minivan"    "pickup"
## [6] "subcompact" "suv"
```

hızlı detaylı genel bakış

```
library(funModeling)
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

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

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

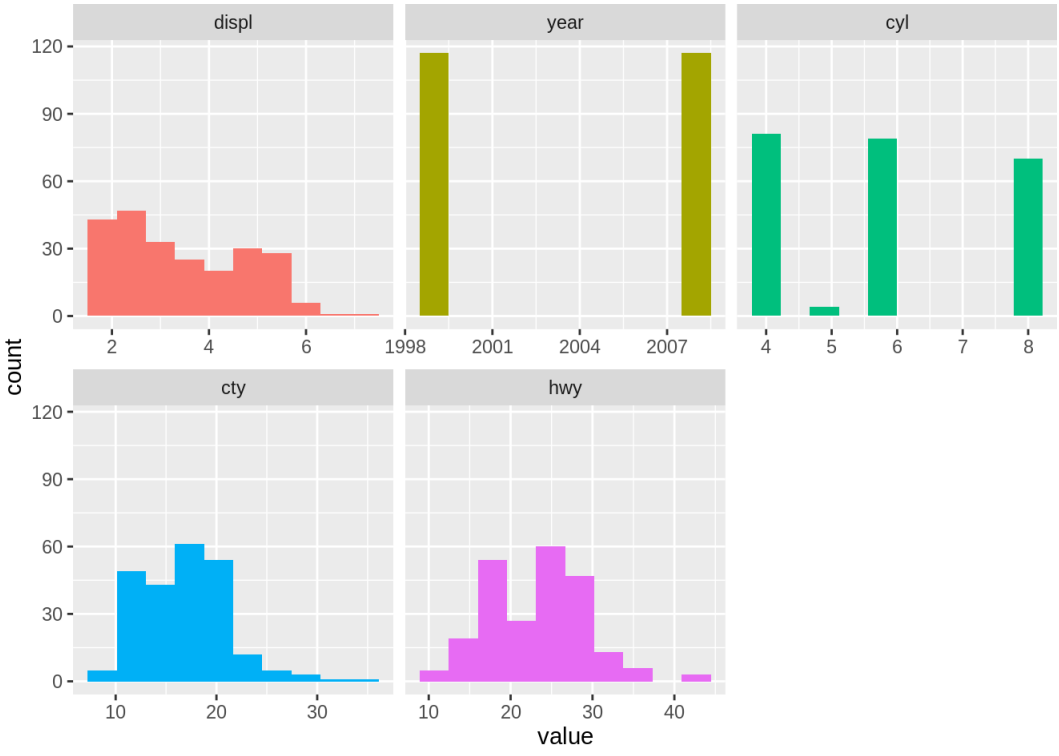
```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

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

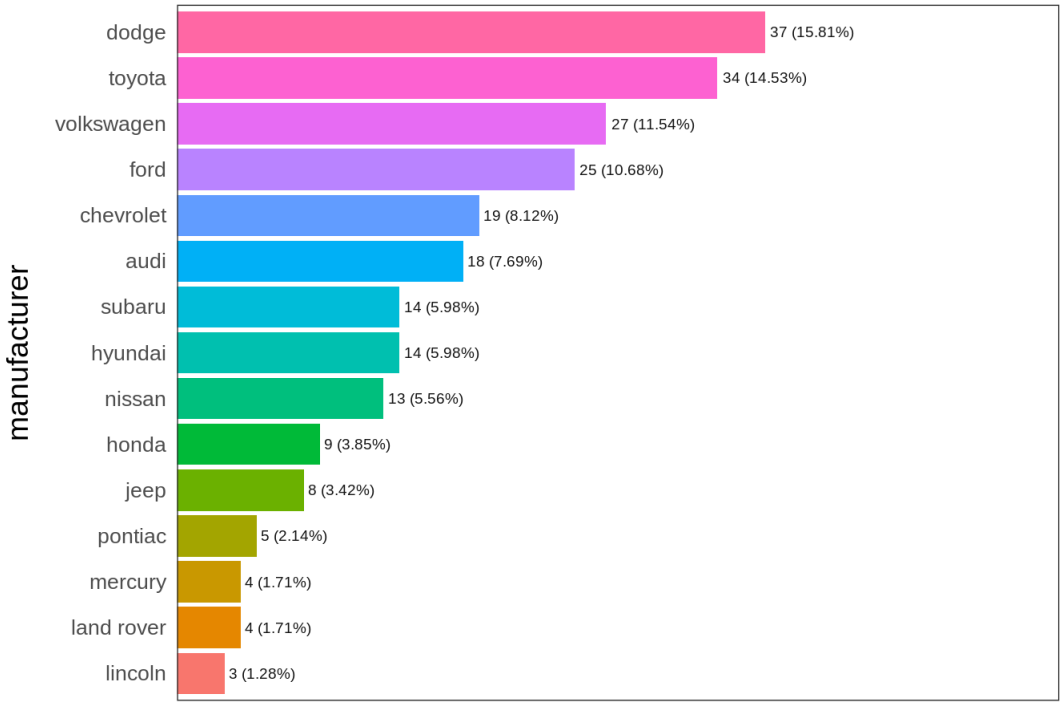
```
profiling_num(df)  #sürekli değişkenlerin özet bazı istatistiklerini verir
```

```
##   variable        mean  std_dev variation_coef   p_01   p_05   p_25   p_50
## 1    displ    3.471795 1.291959    0.372130002    1.6    1.8    2.4    3.3
## 2     year 2003.500000 4.509646    0.002250884 1999.0 1999.0 1999.0 2003.5
## 3      cyl    5.888889 1.611534    0.273656799    4.0    4.0    4.0    6.0
## 4      cty   16.858974 4.255946    0.252443926    9.0   11.0   14.0   17.0
## 5      hwy   23.440171 5.954643    0.254035837   12.0   15.0   18.0   24.0
##     p_75   p_95    p_99  skewness kurtosis iqr               range_98
## 1    4.6    5.7    6.20 0.4414630 2.107412 2.2             [1.6, 6.2]
## 2 2008.0 2008.0 2008.00 0.0000000 1.000000 9.0          [1999, 2008]
## 3    8.0    8.0    8.00 0.1130695 1.549122 4.0               [4, 8]
## 4   19.0   24.0   28.67 0.7914453 4.468651 5.0             [9, 28.67]
## 5   27.0   32.0   39.68 0.3668650 3.163929 9.0 [12, 39.6799999999999]
##       range_80
## 1     [2, 5.4]
## 2 [1999, 2008]
## 3       [4, 8]
## 4     [11, 21]
## 5   [16.3, 30]
```

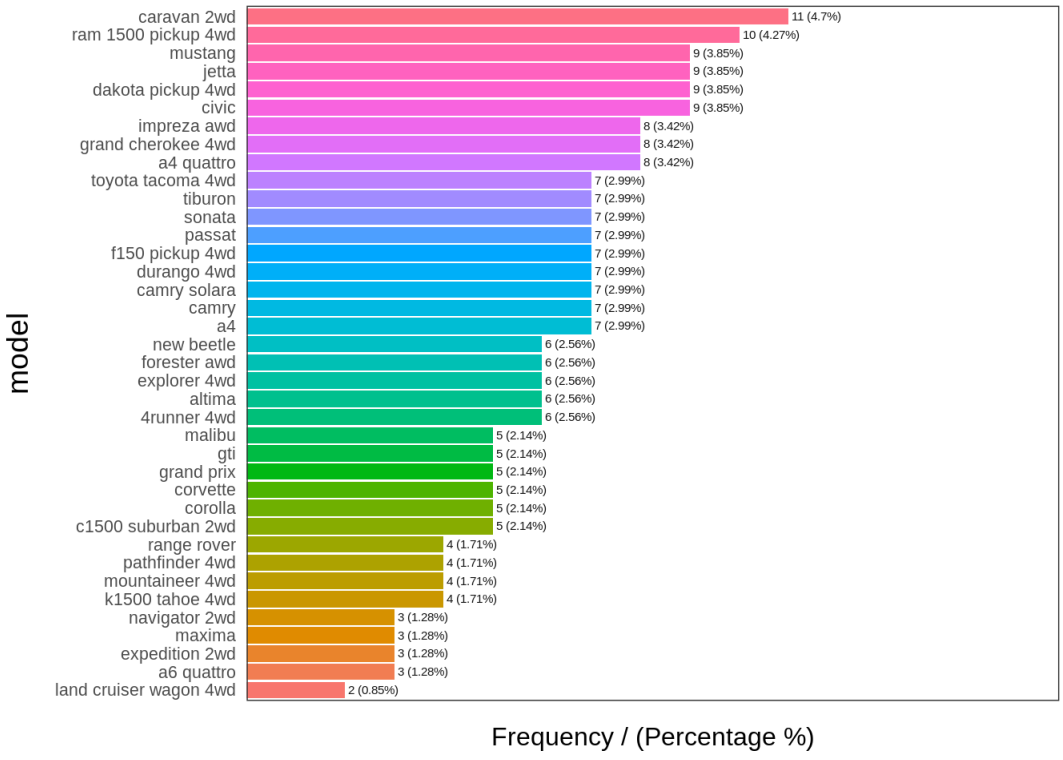
```
plot_num(df)  #genel histogram
```



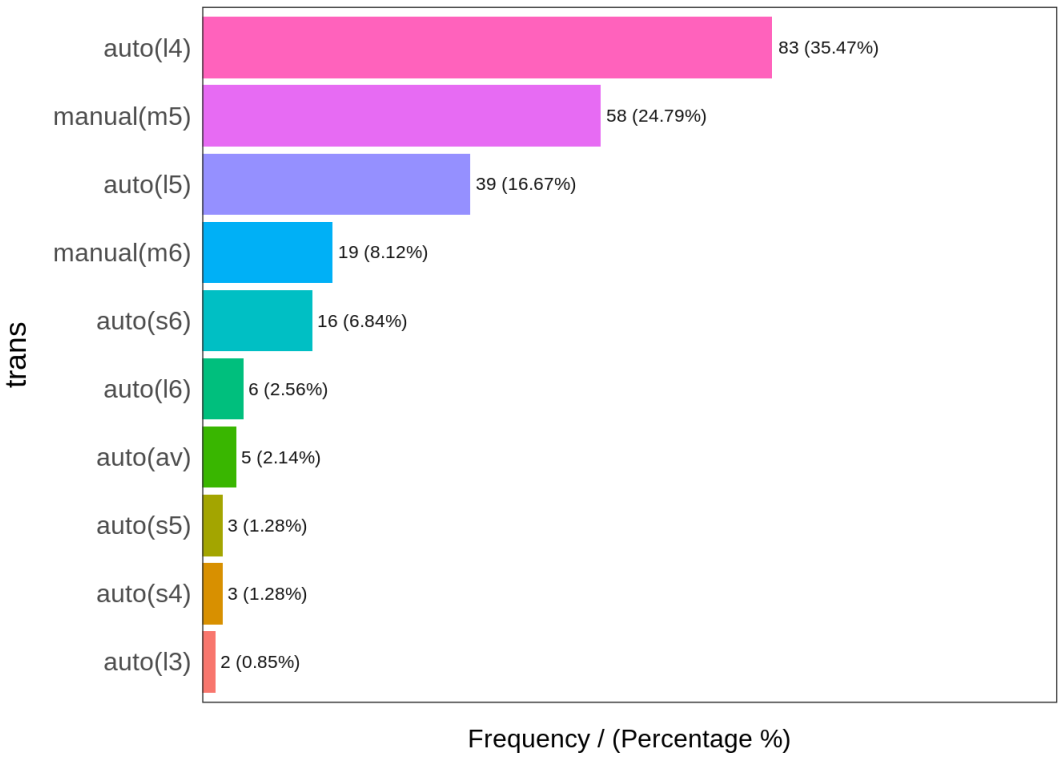
```
freq(df)      #detaylı data frame,grafikler verir
```



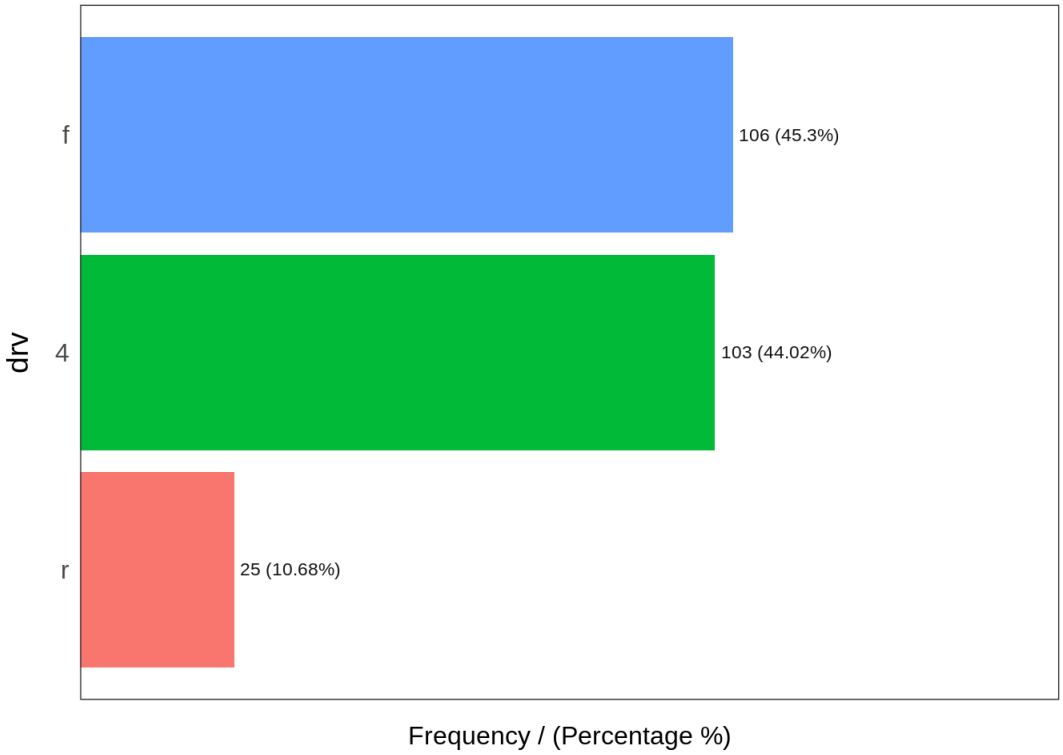

```
##    manufacturer frequency percentage cumulative_perc
## 1         dodge        37      15.81           15.81
## 2        toyota        34      14.53           30.34
## 3    volkswagen        27      11.54           41.88
## 4          ford        25      10.68           52.56
## 5     chevrolet        19       8.12           60.68
## 6          audi        18       7.69           68.37
## 7       hyundai        14       5.98           74.35
## 8        subaru        14       5.98           80.33
## 9        nissan        13       5.56           85.89
## 10        honda         9       3.85           89.74
## 11         jeep         8       3.42           93.16
## 12      pontiac         5       2.14           95.30
## 13   land rover         4       1.71           97.01
## 14      mercury         4       1.71           98.72
## 15      lincoln         3       1.28          100.00
```



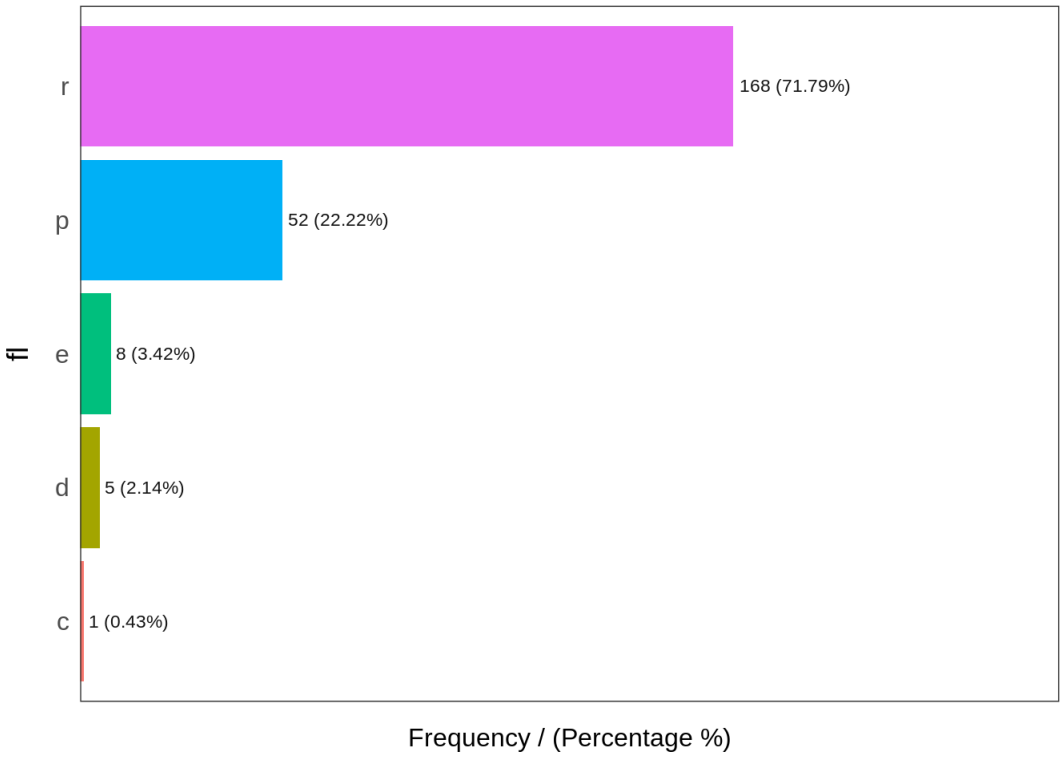
```
##                  model frequency percentage cumulative_perc
## 1              caravan 2wd        11       4.70           4.70
## 2    ram 1500 pickup 4wd        10       4.27           8.97
## 3                  civic         9       3.85          12.82
## 4      dakota pickup 4wd         9       3.85          16.67
## 5                  jetta         9       3.85          20.52
## 6                mustang         9       3.85          24.37
## 7              a4 quattro        8       3.42          27.79
## 8     grand cherokee 4wd         8       3.42          31.21
## 9            impreza awd         8       3.42          34.63
## 10                   a4         7       2.99          37.62
## 11                camry         7       2.99          40.61
## 12           camry solara        7       2.99          43.60
## 13           durango 4wd         7       2.99          46.59
## 14        f150 pickup 4wd        7       2.99          49.58
## 15                passat         7       2.99          52.57
## 16                sonata         7       2.99          55.56
## 17               tiburon         7       2.99          58.55
## 18     toyota tacoma 4wd         7       2.99          61.54
## 19           4runner 4wd         6       2.56          64.10
## 20                altima         6       2.56          66.66
## 21          explorer 4wd         6       2.56          69.22
## 22          forester awd         6       2.56          71.78
## 23            new beetle         6       2.56          74.34
## 24    c1500 suburban 2wd         5       2.14          76.48
## 25                corolla        5       2.14          78.62
## 26              corvette         5       2.14          80.76
## 27            grand prix         5       2.14          82.90
## 28                   gti         5       2.14          85.04
## 29                malibu         5       2.14          87.18
## 30       k1500 tahoe 4wd         4       1.71          88.89
## 31       mountaineer 4wd         4       1.71          90.60
## 32        pathfinder 4wd         4       1.71          92.31
## 33          range rover          4       1.71          94.02
## 34            a6 quattro         3       1.28          95.30
## 35       expedition 2wd          3       1.28          96.58
## 36               maxima          3       1.28          97.86
## 37         navigator 2wd         3       1.28          99.14
## 38 land cruiser wagon 4wd        2       0.85         100.00
```



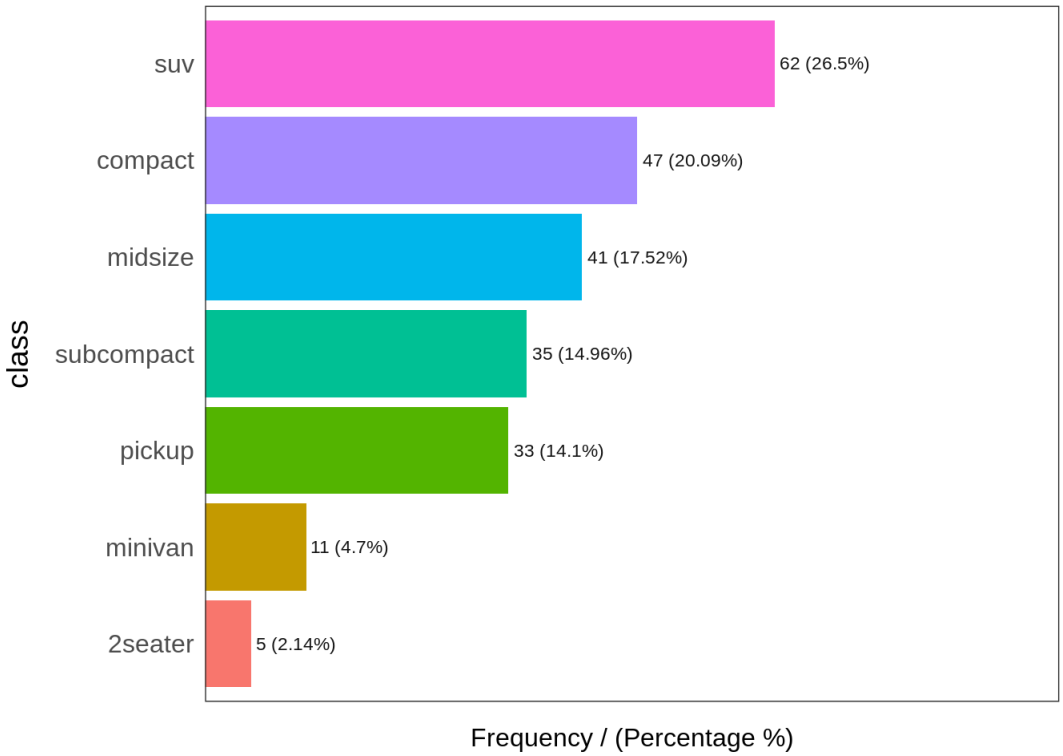
```
##         trans frequency percentage cumulative_perc
## 1    auto(l4)        83      35.47           35.47
## 2  manual(m5)        58      24.79           60.26
## 3    auto(l5)        39      16.67           76.93
## 4  manual(m6)        19       8.12           85.05
## 5    auto(s6)        16       6.84           91.89
## 6    auto(l6)         6       2.56           94.45
## 7    auto(av)         5       2.14           96.59
## 8    auto(s4)         3       1.28           97.87
## 9    auto(s5)         3       1.28           99.15
## 10   auto(l3)         2       0.85          100.00
```



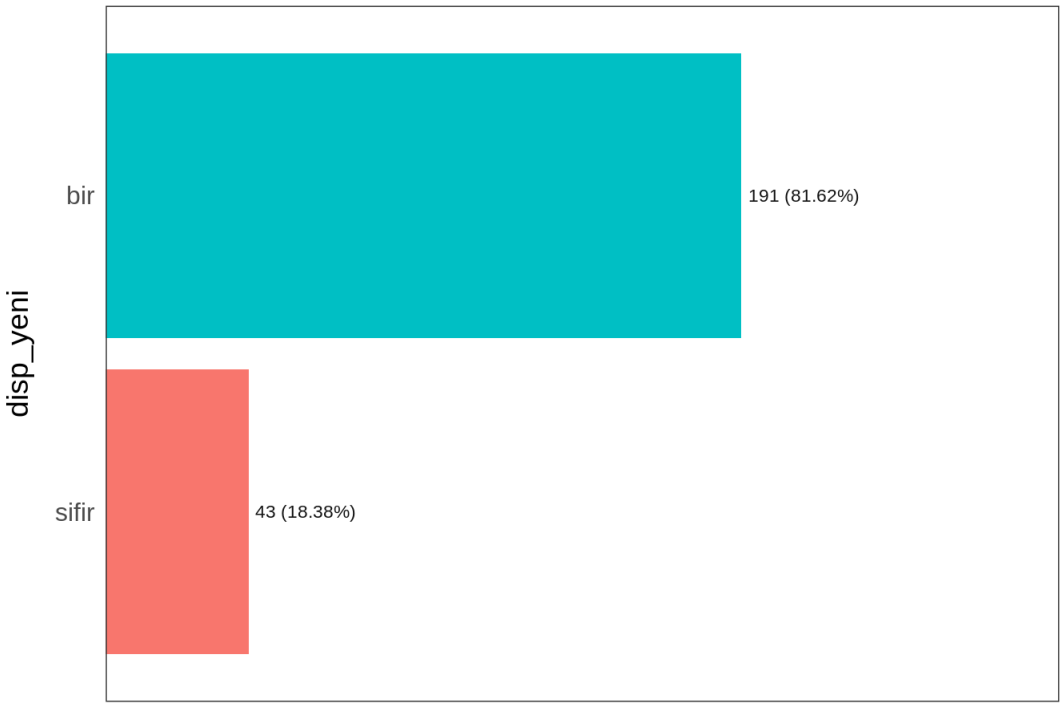

```
##   drv frequency percentage cumulative_perc
## 1   f       106      45.30           45.30
## 2   4       103      44.02           89.32
## 3   r        25      10.68          100.00
```



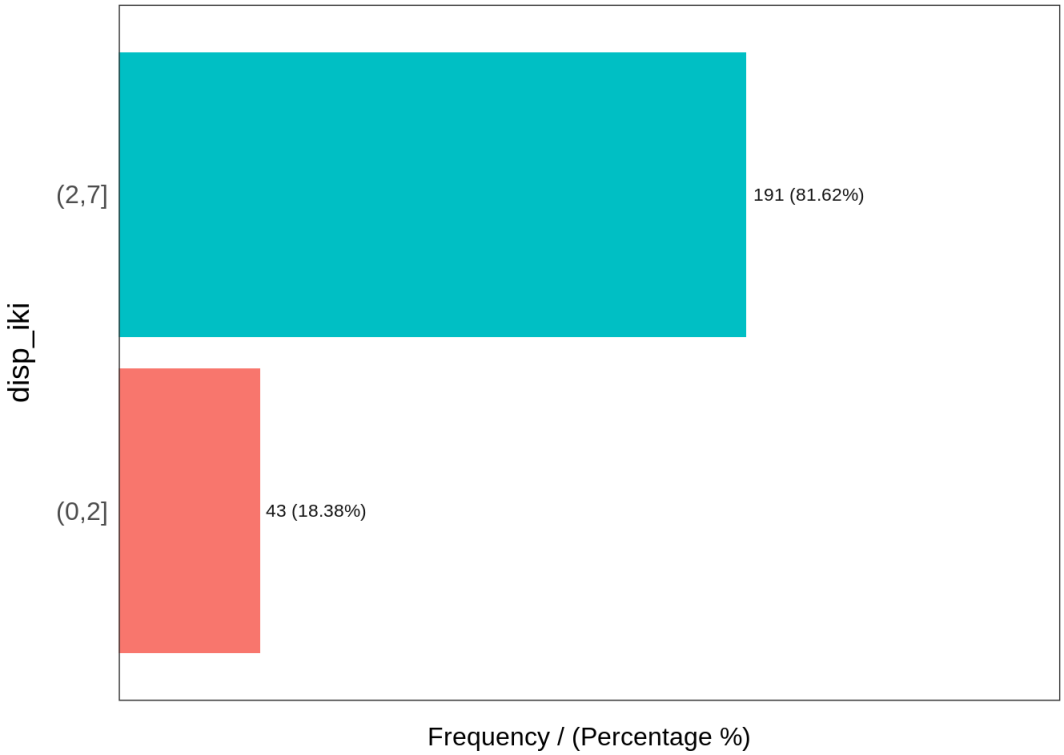
```
##   fl frequency percentage cumulative_perc
## 1  r       168      71.79          71.79
## 2  p        52      22.22          94.01
## 3  e         8       3.42          97.43
## 4  d         5       2.14          99.57
## 5  c         1       0.43         100.00
```




```
##        class frequency percentage cumulative_perc
## 1        suv        62      26.50          26.50
## 2    compact        47      20.09          46.59
## 3    midsize        41      17.52          64.11
## 4 subcompact        35      14.96          79.07
## 5     pickup        33      14.10          93.17
## 6    minivan        11       4.70          97.87
## 7    2seater         5       2.14         100.00
```




```
##   disp_yeni frequency percentage cumulative_perc
## 1       bir       191      81.62          81.62
## 2     sifir        43      18.38         100.00
```




```
##   disp_iki frequency percentage cumulative_perc
## 1    (2,7]       191      81.62          81.62
## 2    (0,2]        43      18.38         100.00
```

```
## [1] "Variables processed: manufacturer, model, trans, drv, fl, class, disp_yeni, disp_iki"
```

Detaylı

```
library(psych)
```

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

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

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

```
describe(mpg$displ)  #displ bazı istatistikleri
```

```
##    vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 234 3.47 1.29    3.3    3.39 1.33 1.6   7   5.4 0.44    -0.91 0.08
```

```
library(Hmisc)
Hmisc::describe(mpg$displ)  #en küçük en yüksek değerleri verir. aykırı gözlem tespiti için kullanılabilir
```

```
## mpg$displ
##       n  missing distinct     Info     Mean      Gmd      .05      .10
##     234        0       35    0.997    3.472    1.471      1.8      2.0
##     .25      .50      .75      .90      .95
##     2.4      3.3      4.6      5.4      5.7
##
## lowest : 1.6 1.8 1.9 2.0 2.2, highest: 6.0 6.1 6.2 6.5 7.0
```

```
library(pastecs)
```

```
##
## Attaching package: 'pastecs'
```

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

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
stat.desc(mpg)  #sürekli değişkenlerin tümünün min,median,range vs vs detaylı data.frame olarak verir
```

```
##          manufacturer model      displ         year          cyl trans drv
## nbr.val            NA    NA 234.000000 2.340000e+02  234.0000000    NA  NA
## nbr.null           NA    NA   0.000000 0.000000e+00    0.0000000    NA  NA
## nbr.na             NA    NA   0.000000 0.000000e+00    0.0000000    NA  NA
## min                NA    NA   1.600000 1.999000e+03    4.0000000    NA  NA
## max                NA    NA   7.000000 2.008000e+03    8.0000000    NA  NA
## range              NA    NA   5.400000 9.000000e+00    4.0000000    NA  NA
## sum                NA    NA 812.400000 4.688190e+05 1378.0000000    NA  NA
## median             NA    NA   3.300000 2.003500e+03    6.0000000    NA  NA
## mean               NA    NA   3.471795 2.003500e+03    5.8888889    NA  NA
## SE.mean            NA    NA   0.084458 2.948048e-01    0.1053493    NA  NA
## CI.mean            NA    NA   0.166399 5.808237e-01    0.2075589    NA  NA
## var                NA    NA   1.669158 2.033691e+01    2.5970434    NA  NA
## std.dev            NA    NA   1.291959 4.509646e+00    1.6115345    NA  NA
## coef.var           NA    NA   0.372130 2.250884e-03    0.2736568    NA  NA
##                   cty          hwy fl class
## nbr.val   234.0000000  234.0000000 NA    NA
## nbr.null    0.0000000    0.0000000 NA    NA
## nbr.na      0.0000000    0.0000000 NA    NA
## min         9.0000000   12.0000000 NA    NA
## max        35.0000000   44.0000000 NA    NA
## range      26.0000000   32.0000000 NA    NA
## sum      3945.0000000 5485.0000000 NA    NA
## median     17.0000000   24.0000000 NA    NA
## mean       16.8589744   23.4401709 NA    NA
## SE.mean     0.2782199    0.3892672 NA    NA
## CI.mean     0.5481481    0.7669333 NA    NA
## var        18.1130736   35.4577785 NA    NA
## std.dev     4.2559457    5.9546434 NA    NA
## coef.var    0.2524439    0.2540358 NA    NA
```

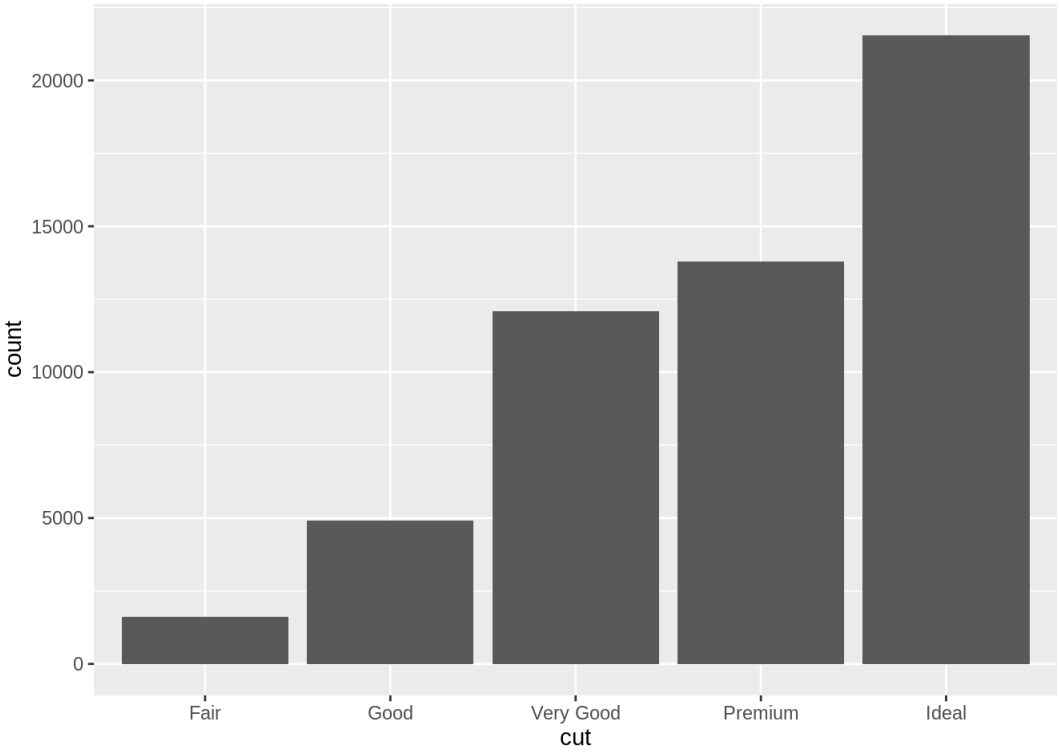
```
library(d3Tree)
d3tree(list(root = df2tree(rootname ='Titanic',
                           struct = as.data.frame(Titanic)),
                           layout = 'collapse'))                  #anlamlandırmayı kolaylaştırır
```

bar plot (sütun grafiği)

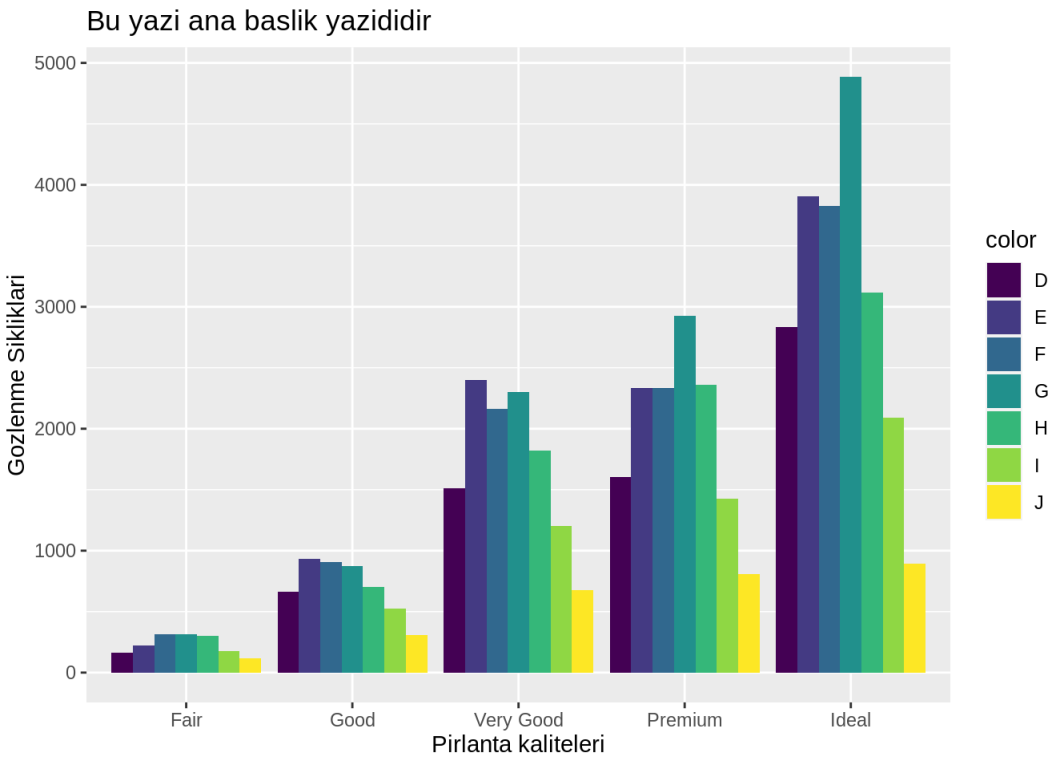
```
#örnek1
df <- diamonds
glimpse(df)
```

```
## Rows: 53,940
## Columns: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0.…
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ver…
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I,…
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS1, …
## $ depth   <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4, 64…
## $ table   <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62, 58…
## $ price   <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340, 34…
## $ x       <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00, 4.…
## $ y       <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05, 4.…
## $ z       <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39, 2.…
```

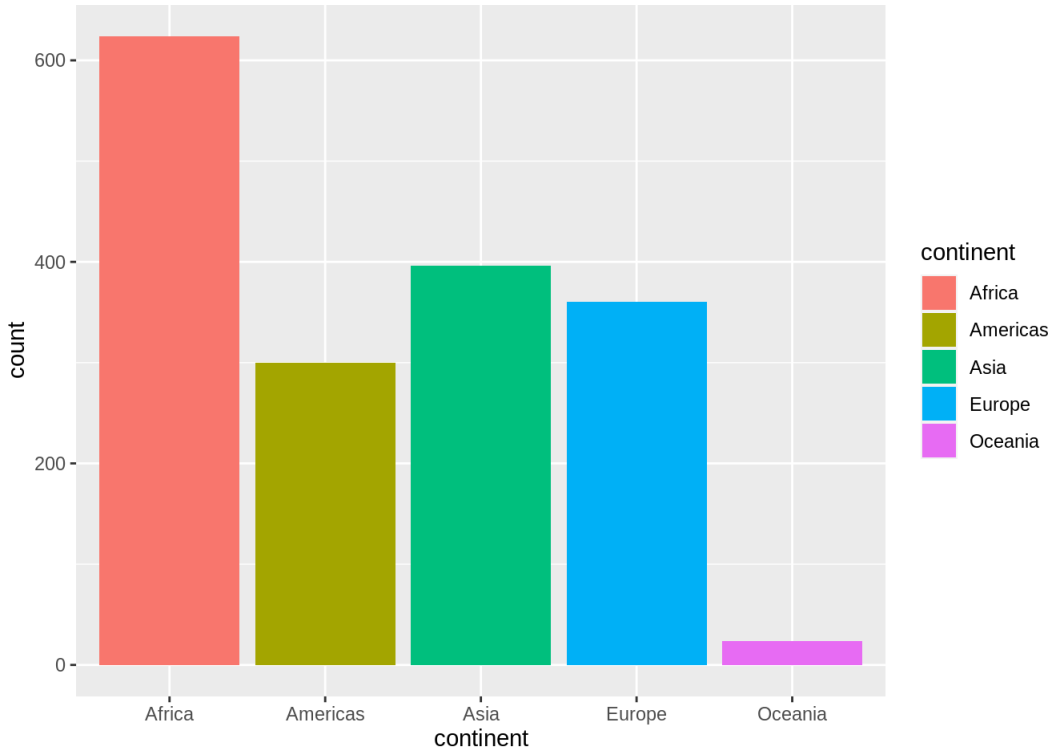
```
ggplot(df, aes(cut)) +
  geom_bar()
```




```
ggplot(df, aes(cut, fill = color)) +
  geom_bar(position = position_dodge()) +
  ggtitle("Bu yazi ana baslik yazididir") +
  xlab("Pirlanta kaliteleri") +
  ylab("Gozlenme Sikliklari")
```




```
#örnek2

library(gapminder)
data("gapminder")


ggplot(gapminder, aes(x=continent,fill=continent)) +
  geom_bar()
```

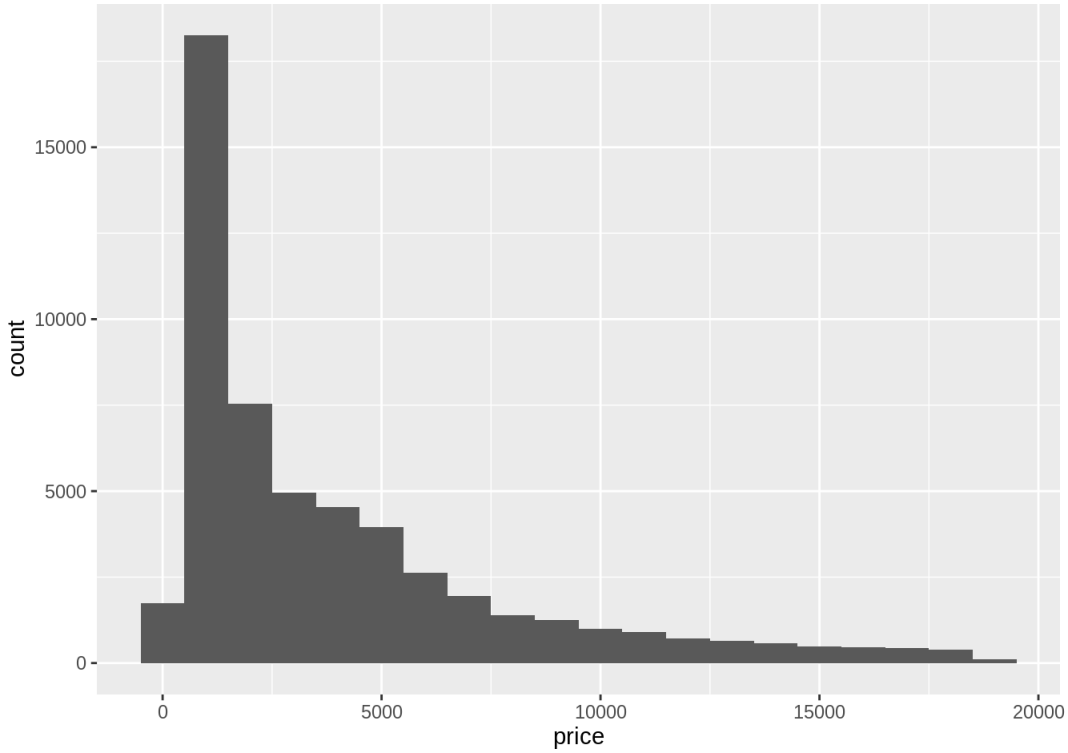


histogram

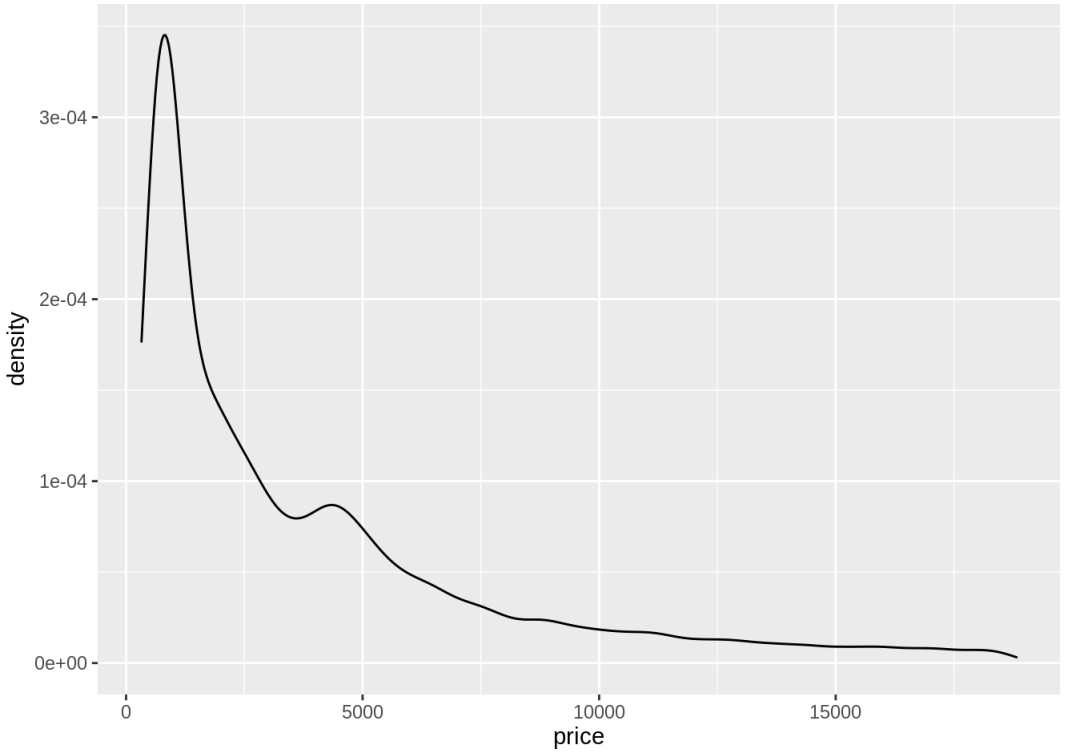
```
glimpse(df)
```

```
## Rows: 53,940
## Columns: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0.…
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ver…
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I,…
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS1, …
## $ depth   <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4, 64…
## $ table   <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62, 58…
## $ price   <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340, 34…
## $ x       <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00, 4.…
## $ y       <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05, 4.…
## $ z       <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39, 2.…
```

```
ggplot(df, aes(price)) +                #histogram
  geom_histogram(binwidth = 1000)
```

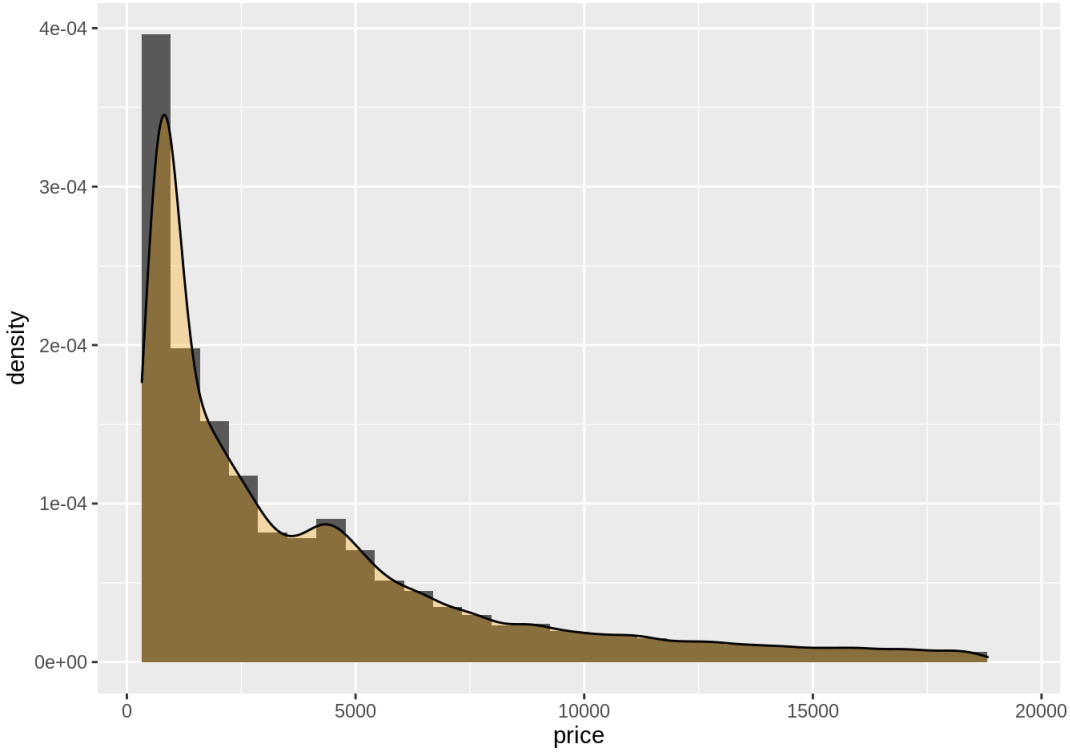



```
ggplot(df, aes(price)) +        #yoğunluk
  geom_density()
```



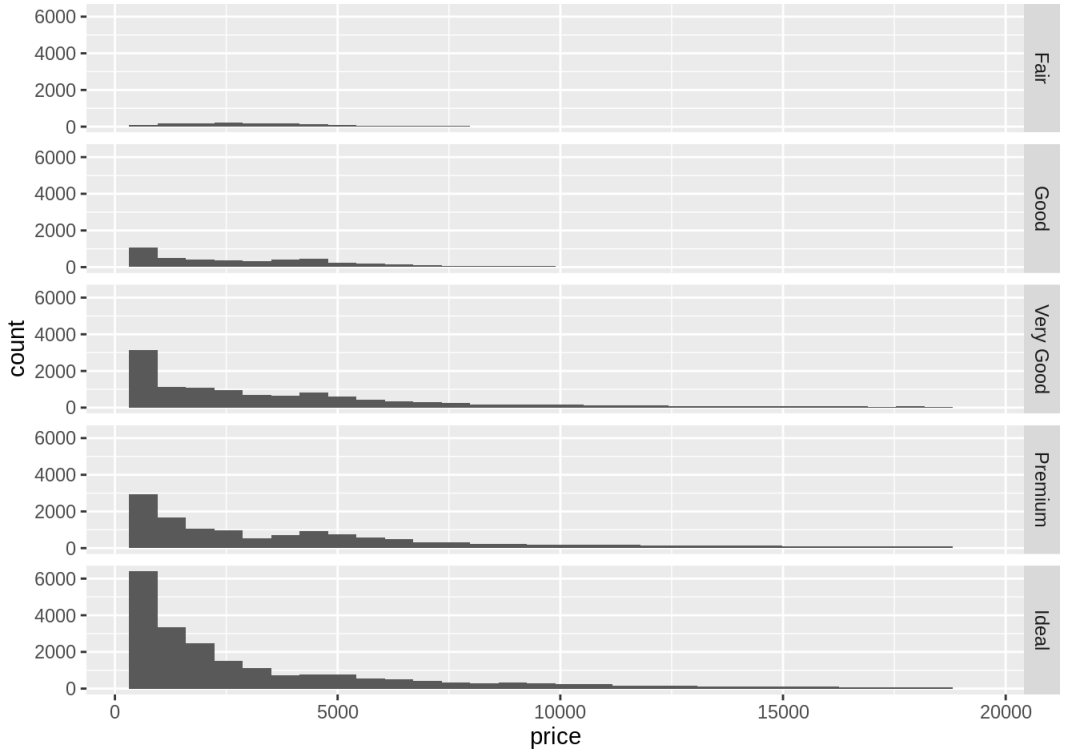
```
ggplot(df, aes(price)) +
  geom_histogram(aes(y = ..density..)) +          #beraber
  geom_density(alpha = .3, fill = "orange")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(df, aes(price)) +
  geom_histogram() +                #ayrı ayrı verir
  facet_grid(cut ~. )
```

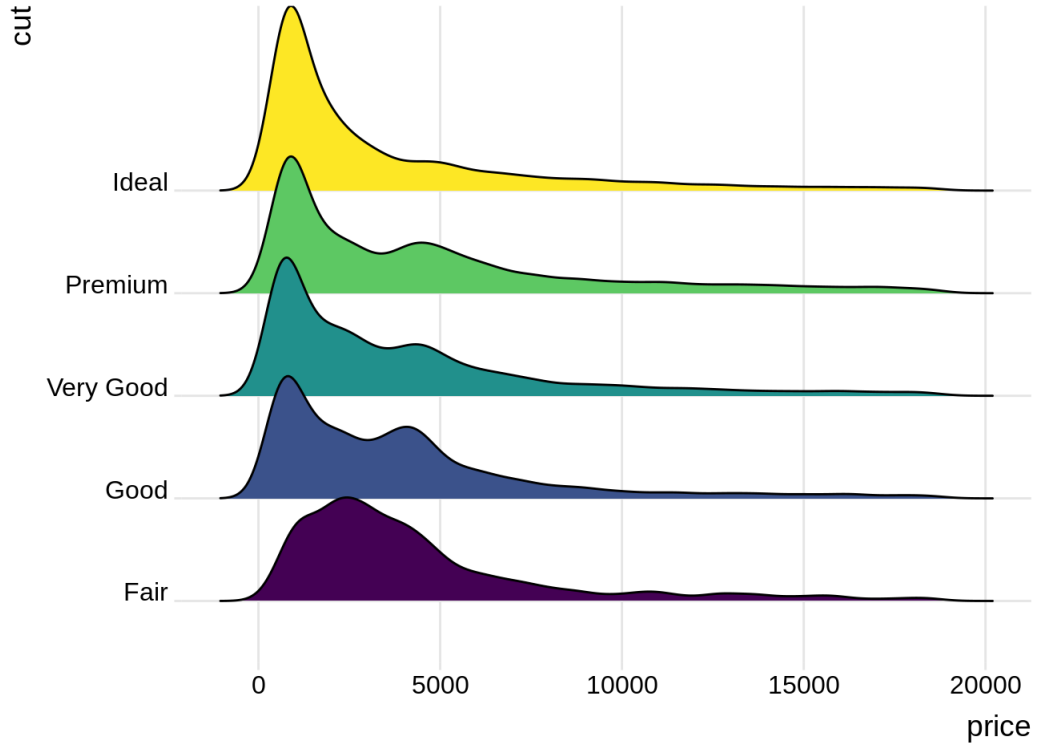
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



readline

```
library(ggridges)

ggplot(df, aes(x = price, y = cut, fill = cut)) +
  geom_density_ridges() +
  theme_ridges() +
  theme(legend.position = "none")
```

```
## Picking joint bandwidth of 458
```



interaktik histogram

```
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
```

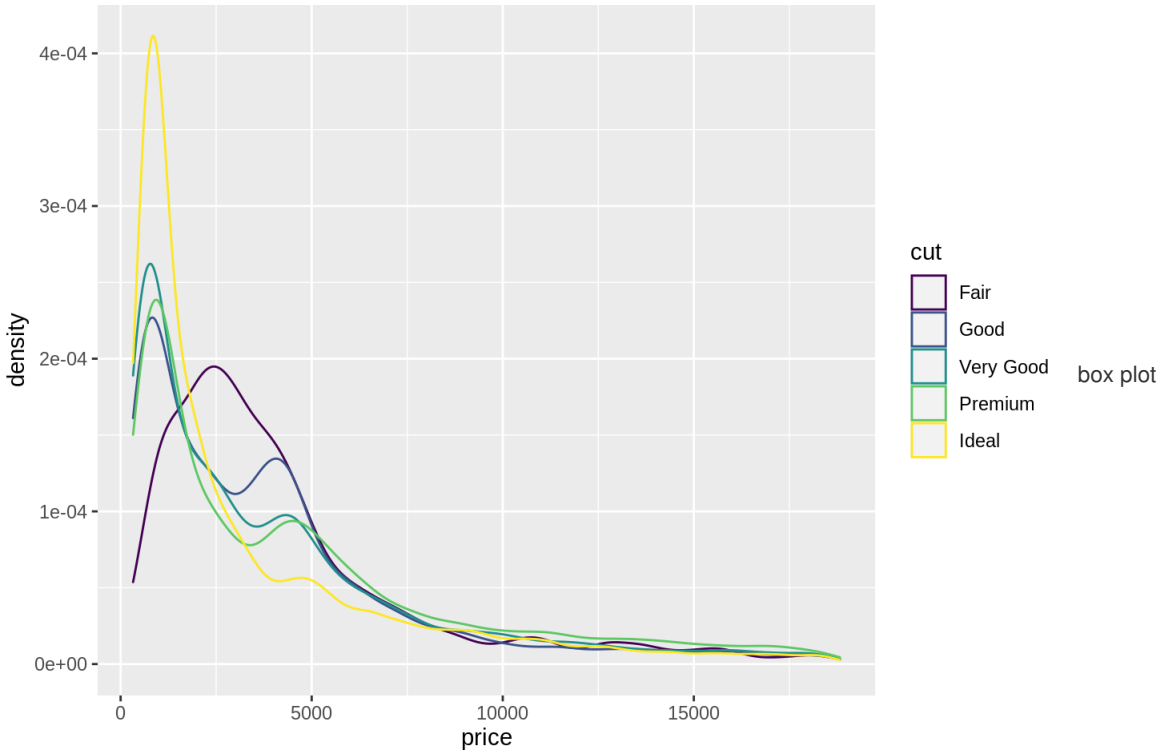
```
g <- plot_ly(x = rnorm(500), type = "histogram")
g
```

```
plot_ly(x = rnorm(500), opacity = 0.6, type = "histogram") %>%
 add_trace(x = rnorm(500) + 1) %>%
 layout(barmode =  "overlay")
```

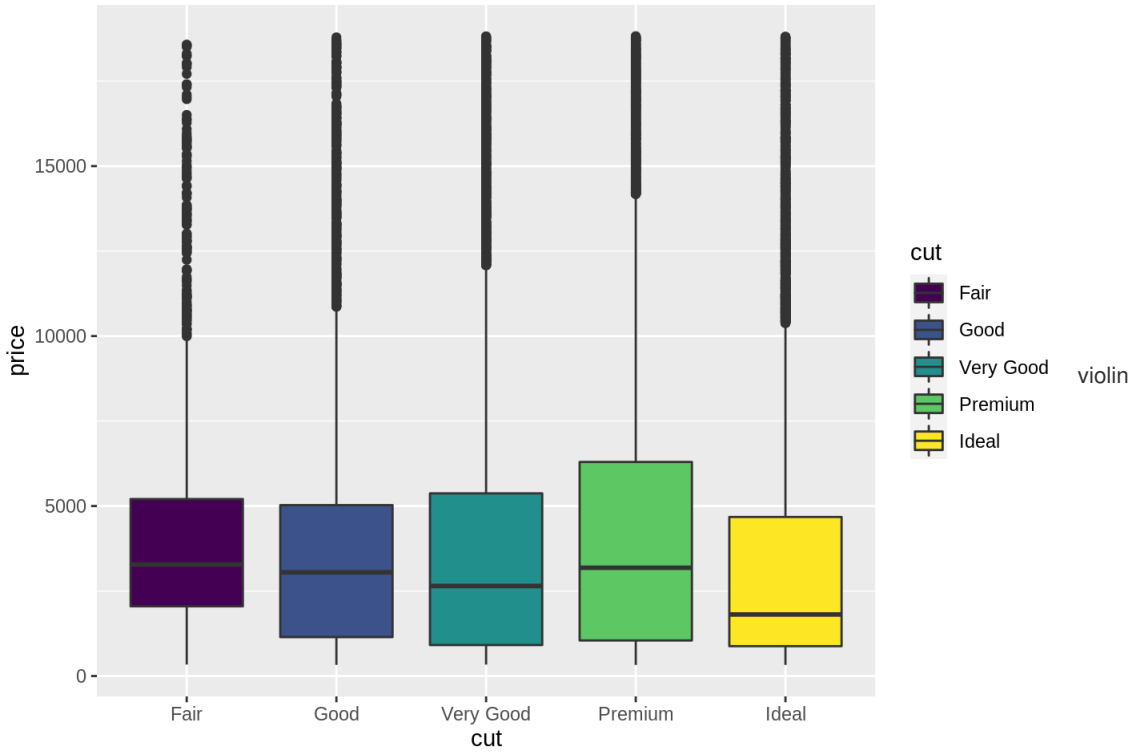
çoklu frekans

```
ggplot(df, aes(price, y = ..density..)) +
  geom_density(aes(colour = cut), binwidth = 500)
```

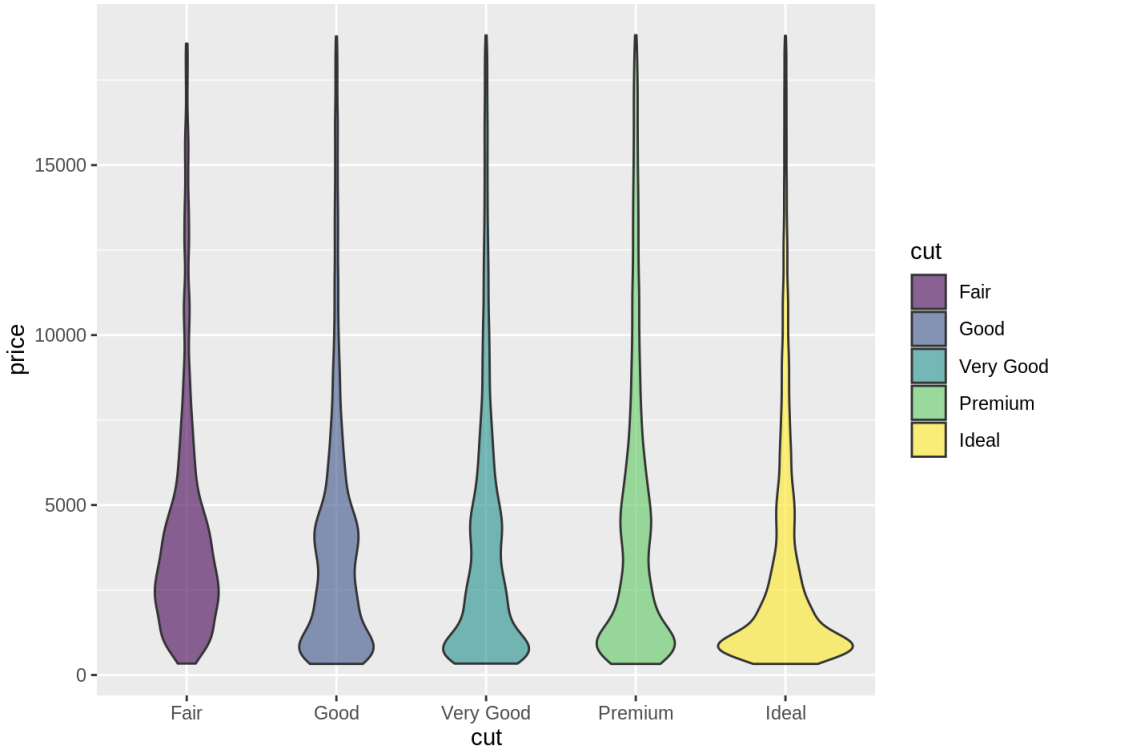
```
## Warning: Ignoring unknown parameters: binwidth
```



```
ggplot(df, aes(x = cut, y = price, fill = cut)) +
    geom_boxplot()
```




```
ggplot(df, aes(x = cut, y = price, fill = cut)) +
    geom_violin(alpha = 0.6)
```



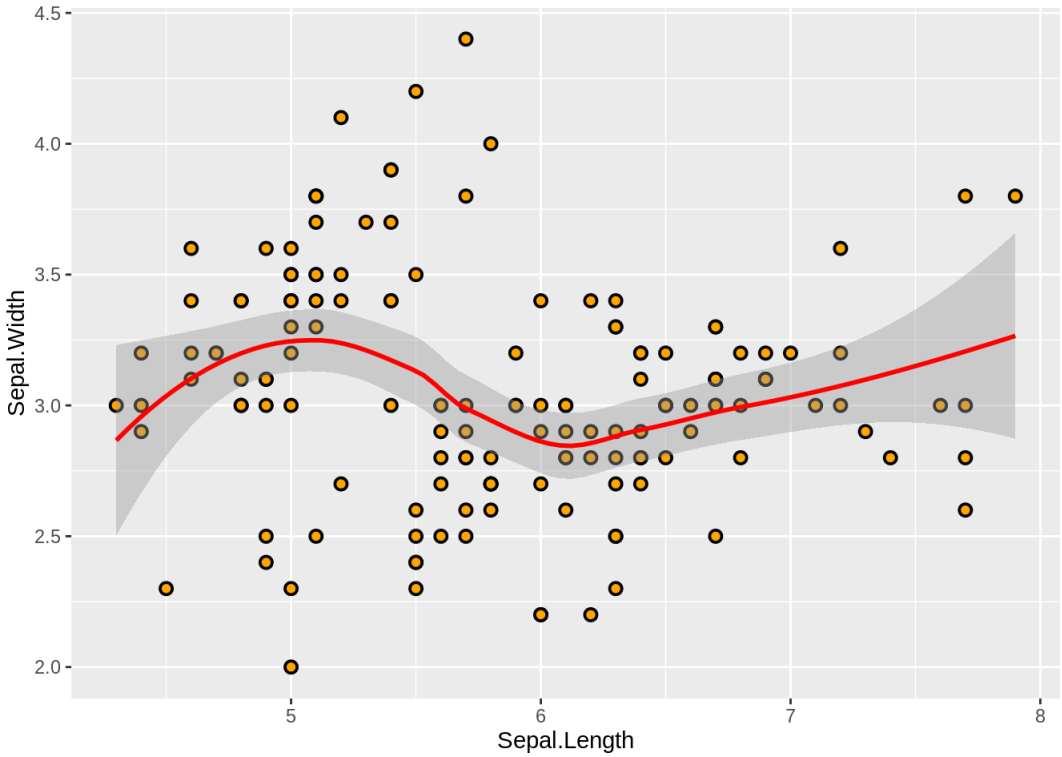
korelasyonların incelenmesi

```
glimpse(iris)
```

```
## Rows: 150
## Columns: 5
## $ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0, 4.4, 4.9, 5.4, 4.…
## $ Sepal.Width  <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4, 2.9, 3.1, 3.7, 3.…
## $ Petal.Length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5, 1.5, 1.…
## $ Petal.Width  <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.2, 0.…
## $ Species      <fct> setosa, setosa, setosa, setosa, setosa, setosa, setosa, s…
```

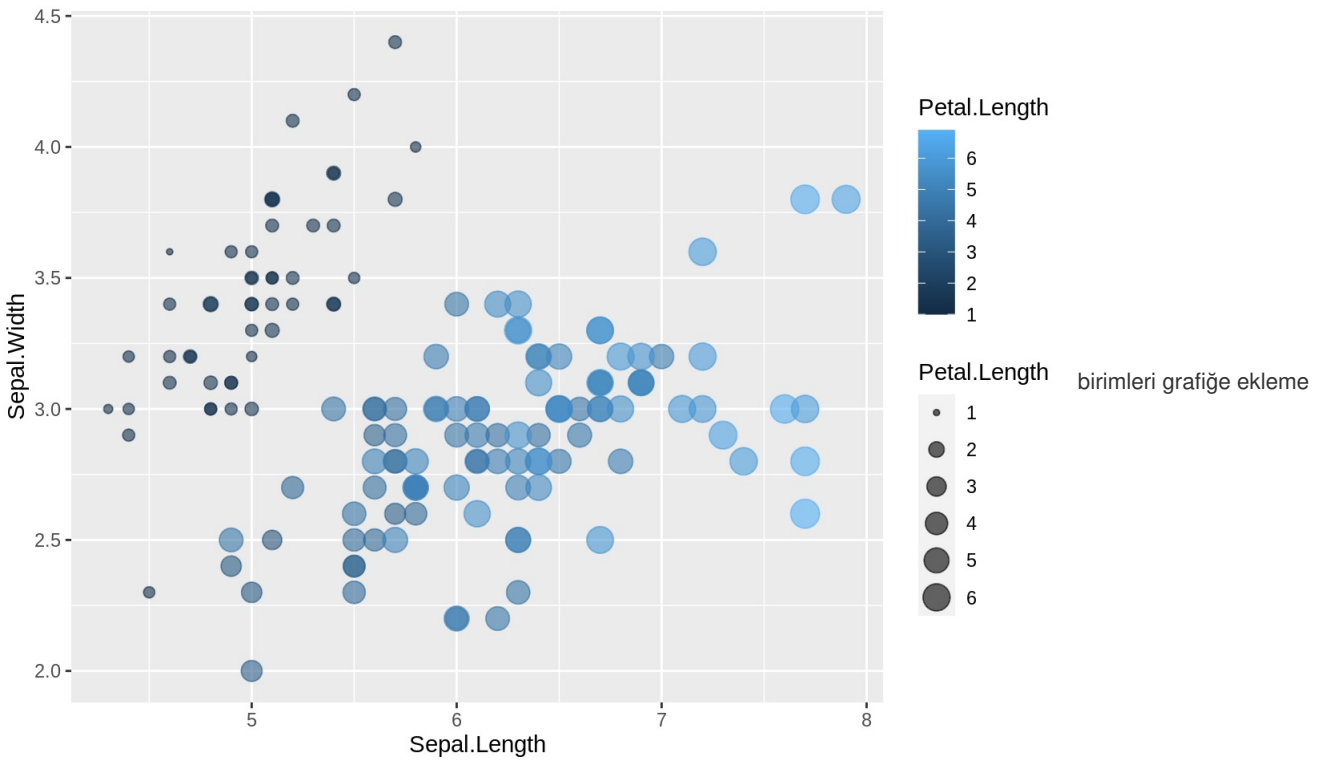
```
df1 <- iris

ggplot(df1, aes(Sepal.Length , Sepal.Width)) +
  geom_point(size = 2,
             shape = 21,
             stroke = 1,
             color = "black",
             fill = "orange") +
  geom_smooth(color = "red")
```

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




```
ggplot(df1, aes(Sepal.Length, Sepal.Width, color = Species, shape = Species )) +
  geom_point(size = 6, alpha = 0.6)
```

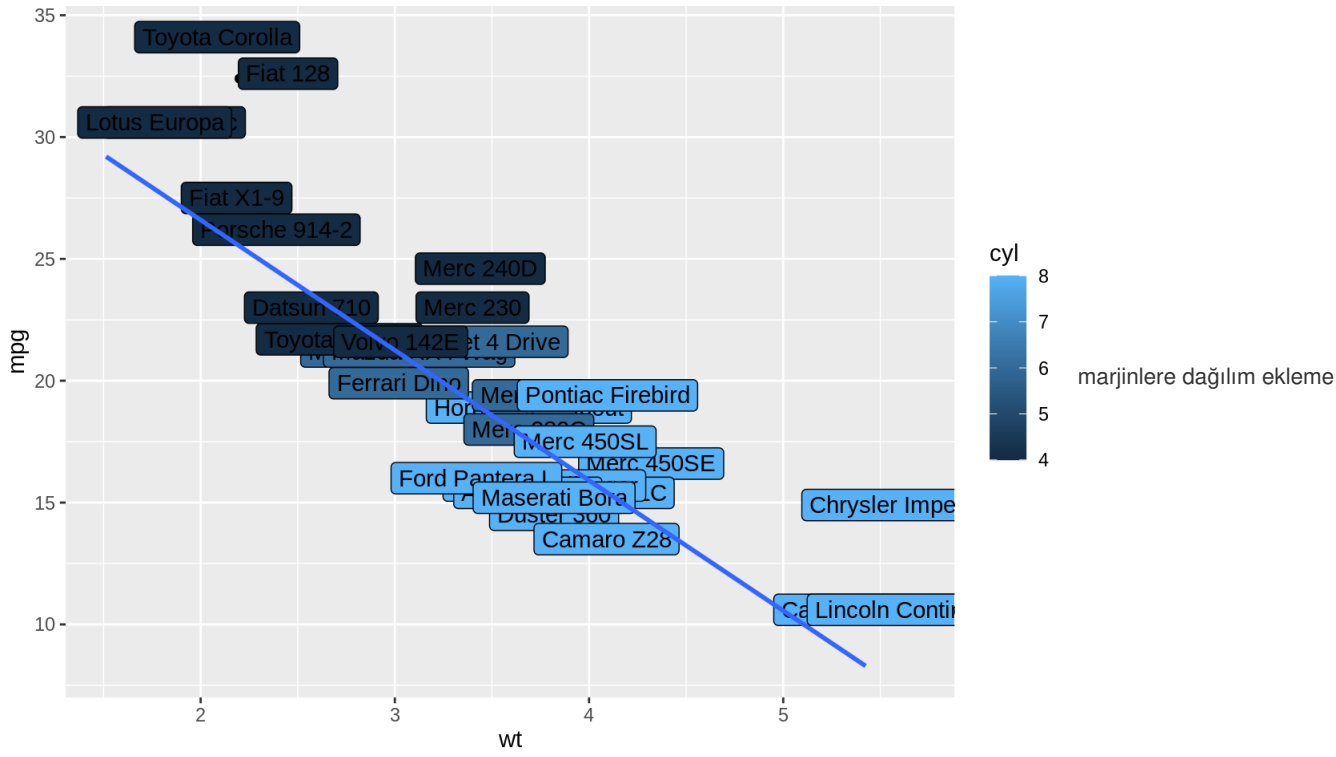

```
ggplot(df1, aes(Sepal.Length, Sepal.Width, color = Petal.Length, size = Petal.Length )) +
  geom_point(alpha = 0.6)
```



birimleri grafiğe ekleme

```
df <- mtcars

ggplot(df, aes(x = wt, y = mpg, fill = cyl)) +
  geom_point() +
  geom_label(label = rownames(df),
             nudge_x = 0.25,
             nudge_y = 0.2) +
  geom_smooth(method = lm, se = FALSE)
```

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



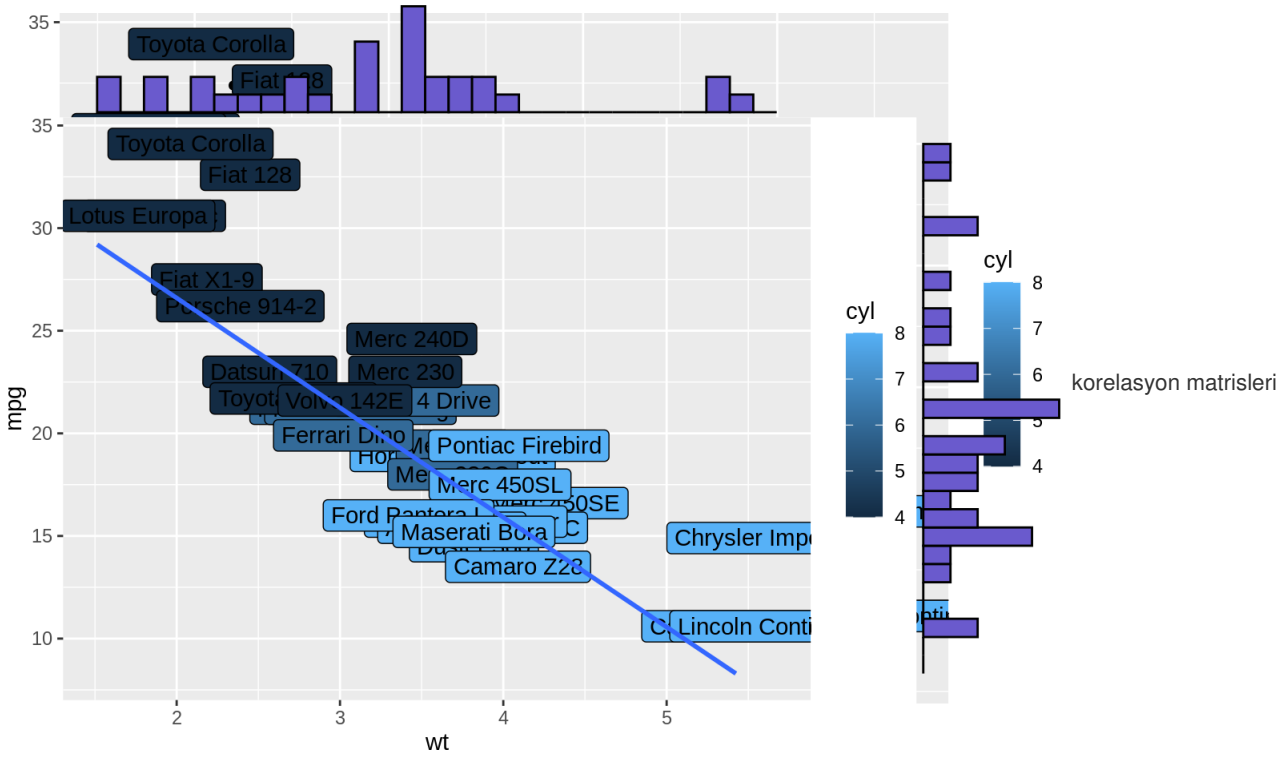
marjinlere dağılım ekleme

```
library(ggExtra)

g <- ggplot(df, aes(x = wt, y = mpg, fill = cyl)) +
  geom_point() +
  geom_label(label = rownames(df),
             nudge_x = 0.25,
             nudge_y = 0.2) +
  geom_smooth(method = lm, se = FALSE)
g
```

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

```
ggMarginal(g, type = "histogram", fill = "slateblue")
```

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




```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
##
## Attaching package: 'GGally'
```

```
## The following object is masked from 'package:funModeling':
##
##     range01
```

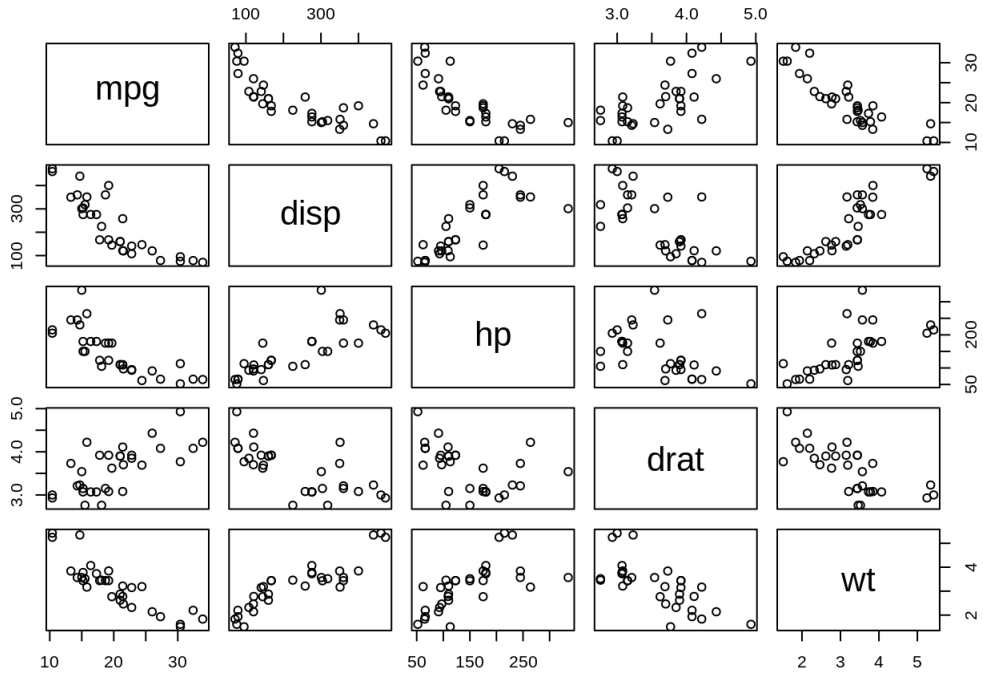
```
df <- mtcars[ , c(1,3:6)]

cor(df)
```

```
##             mpg       disp         hp       drat         wt
## mpg   1.0000000 -0.8475514 -0.7761684  0.6811719 -0.8676594
## disp -0.8475514  1.0000000  0.7909486 -0.7102139  0.8879799
## hp   -0.7761684  0.7909486  1.0000000 -0.4487591  0.6587479
## drat  0.6811719 -0.7102139 -0.4487591  1.0000000 -0.7124406
## wt   -0.8676594  0.8879799  0.6587479 -0.7124406  1.0000000
```

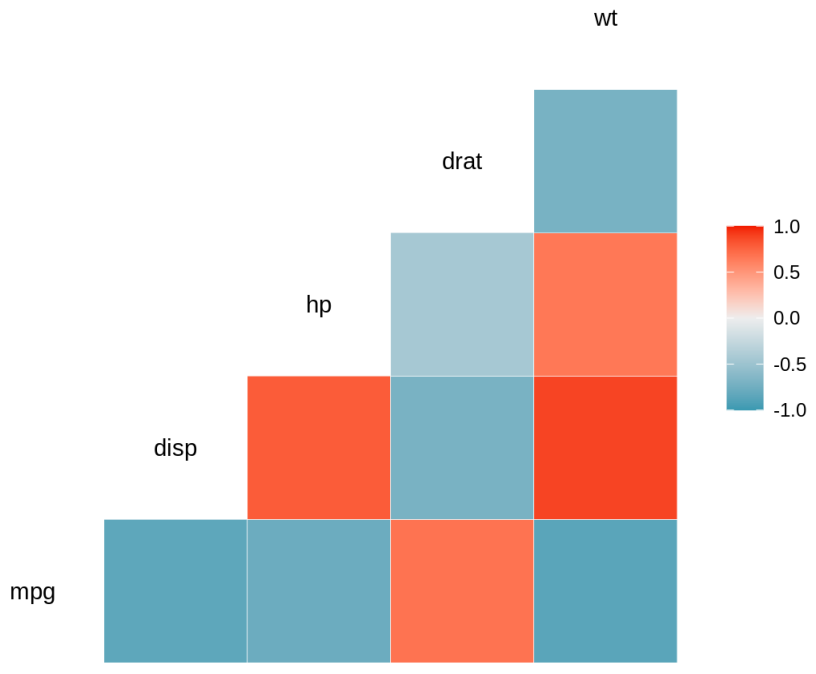
```
cor.test(df$mpg, df$wt)
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$mpg and df$wt
## t = -9.559, df = 30, p-value = 1.294e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9338264 -0.7440872
## sample estimates:
##       cor
## -0.8676594
```

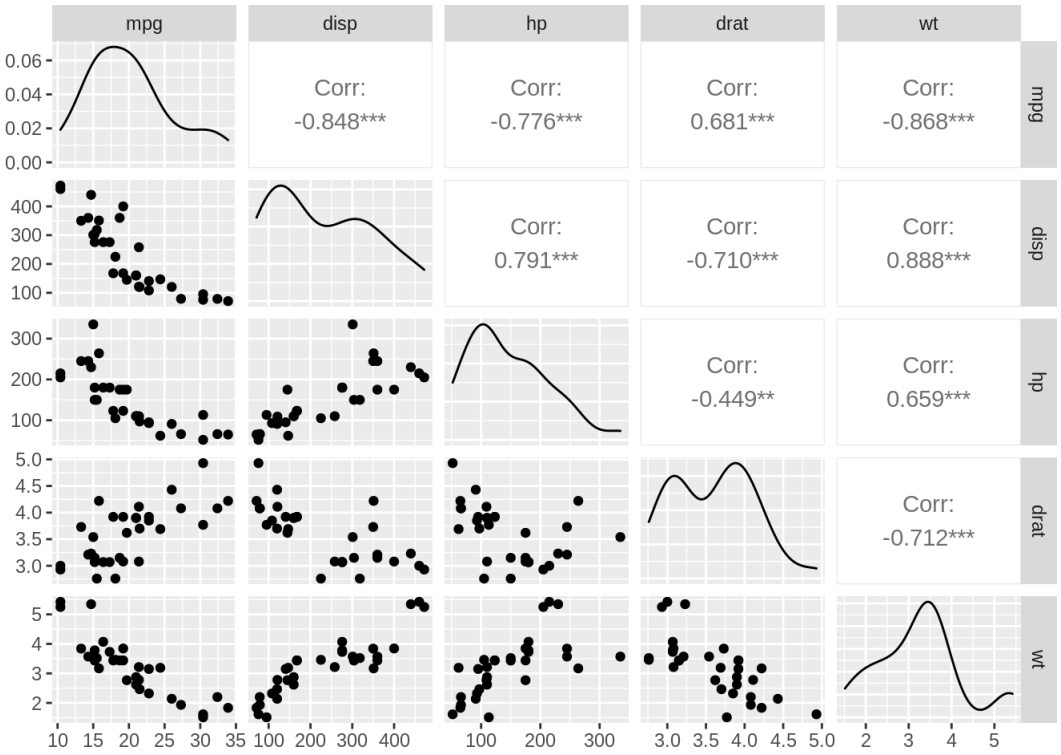
```
plot(df)        #6li
```




```
ggcorr(df)
```




```
ggpairs(df)
```



zaman serisi görselleştirme

```
df <- economics
ggplot(df, aes(date, psavert)) +
  geom_line() +
  stat_smooth()
```

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