

## HATIRLATMA

1. TEK DEĞİŞKENLİ GÖRSELLEŞTİRME
2. İKİ DEĞİŞKENLİ GÖRSELLEŞTİRME
3. GRAFİK BÖLME
4. ZAMAN SERİSİ GÖRSELLEŞTİRME
5. ÖRNEKLER
6. KAYNAKÇA

# R veri Görselleştirme Ders Notu

Kevser Bahadır

## HATIRLATMA

```
library(tibble)
library(dplyr)
```

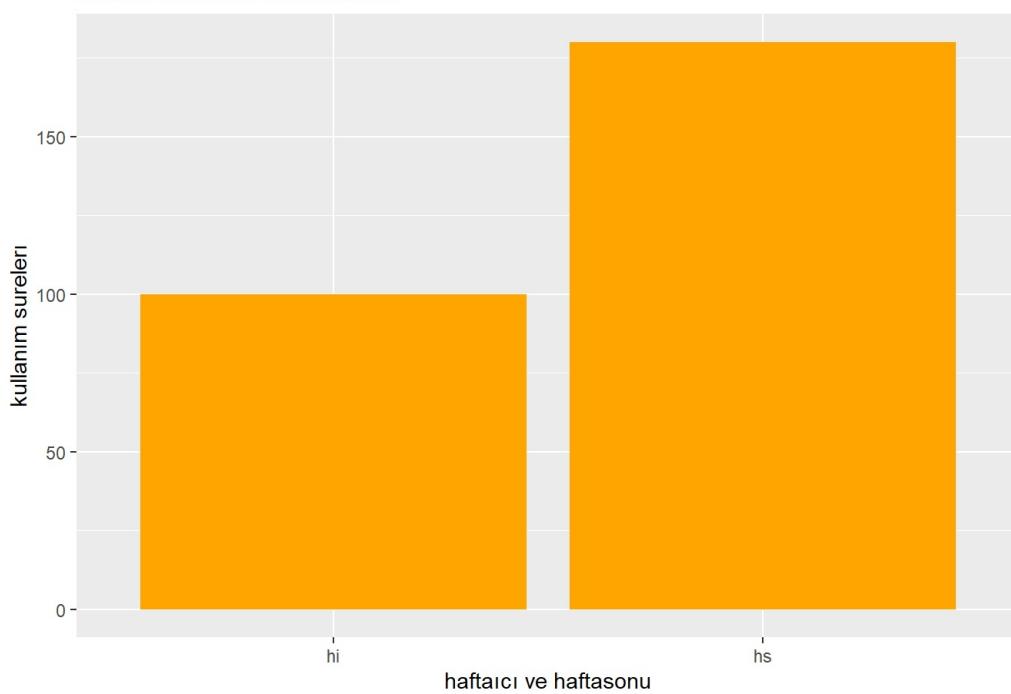
```
##
## Attaching package: 'dplyr'
```

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

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

```
library(ggplot2)
library(tidyr)

df <- tibble(
  kullanım = factor(c("hi" , "hs") , levels= c("hi", "hs")),
  sure= c(100,180)
)
ggplot(data = df ,aes(x=kullanım , y=sure , fill = kullanım))+
  geom_bar(colours ="black", fill = "orange" , stat = "identity") +
  guides(fill = FALSE) +
  xlab("haftaici ve haftasonu") + ylab("kullanım sureleri") +
  ggtitle("telefon kullanım istatistikleri")
```



## 1. TEK DEĞİŞKENLİ GÖRSELLEŞTİRME

### 1.1 KESİKLİ DEĞİŞKEN

Bu bir kategorik değişken anlamına gelir örneğin cinsiyet, semt, ilçe anne, baba, rütbe vs. bar plot ile görselleşecektir ise iki türlü yapılır. birincisi frekans diğeri mutlak değerine göre.

```
df <- tibble(
  kullanım = c(rep("hi" , 20), rep("hs" , 10))
)
```

*#Burada evet bir kategorik değişken var ama karşısında değeri ok gözlemleyeceğim şey sadece frekansı*

### 1.2 SÜREKLİ DEĞİŞKEN

#### AÇIKLAMA

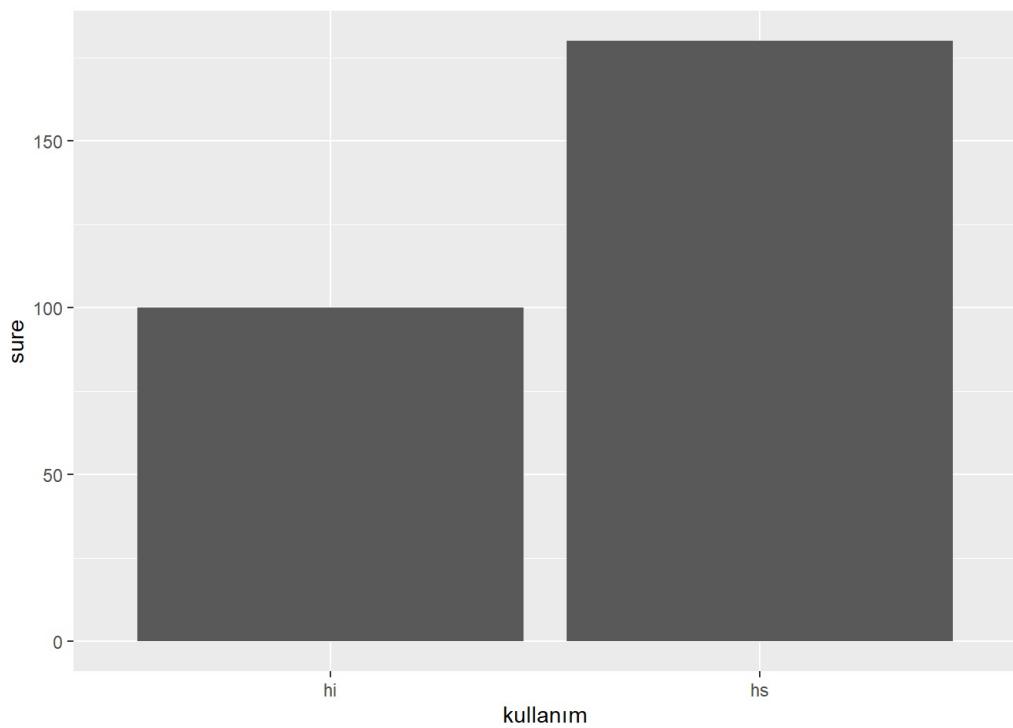
```
df <- tibble(
  kullanım = factor(c("hi", "hs"), levels= c("hi", "hs")),
  sure = c(100,180)
)
df
```

```
## # A tibble: 2 × 2
##   kullanım   sure
##   <fct>     <dbl>
## 1 hi          100
## 2 hs          180
```

## KESİKLİ DEĞİŞKEN

### SÜTUN GRAFİĞİ (*geom\_bar*)

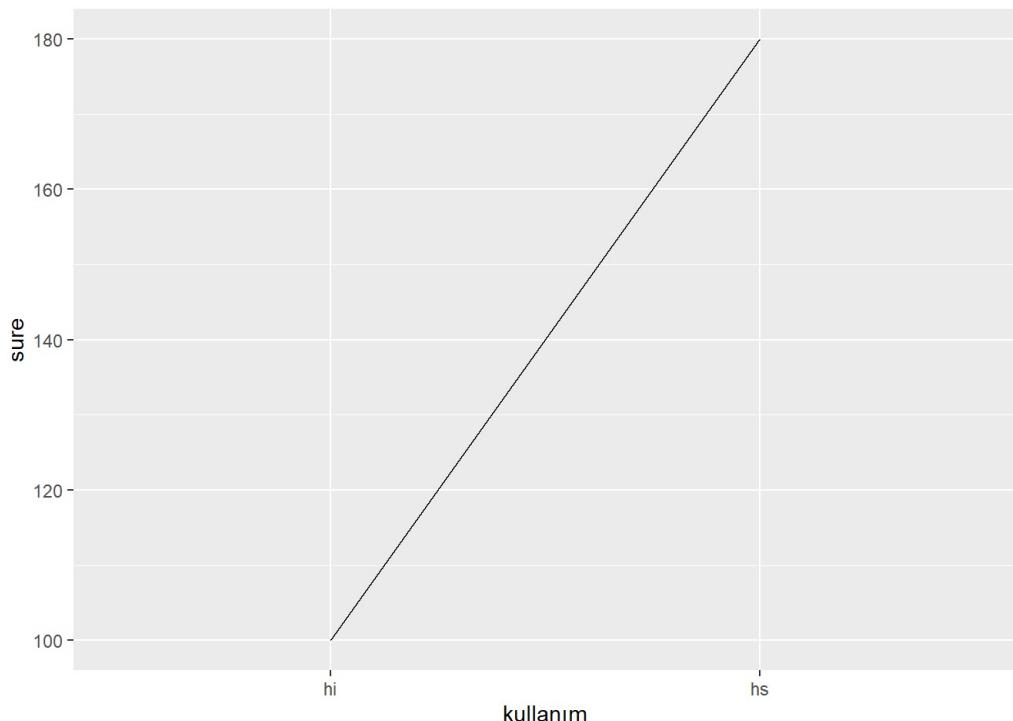
```
df <- tibble(
  kullanım = factor(c("hi" , "hs") , levels= c("hi", "hs")),
  sure= c(100,180)
)
ggplot(df, aes( x = kullanım , y = sure)) +
  geom_bar( stat = "identity")
```



## ÇİZGİ GRAFİĞİ (geom\_line)

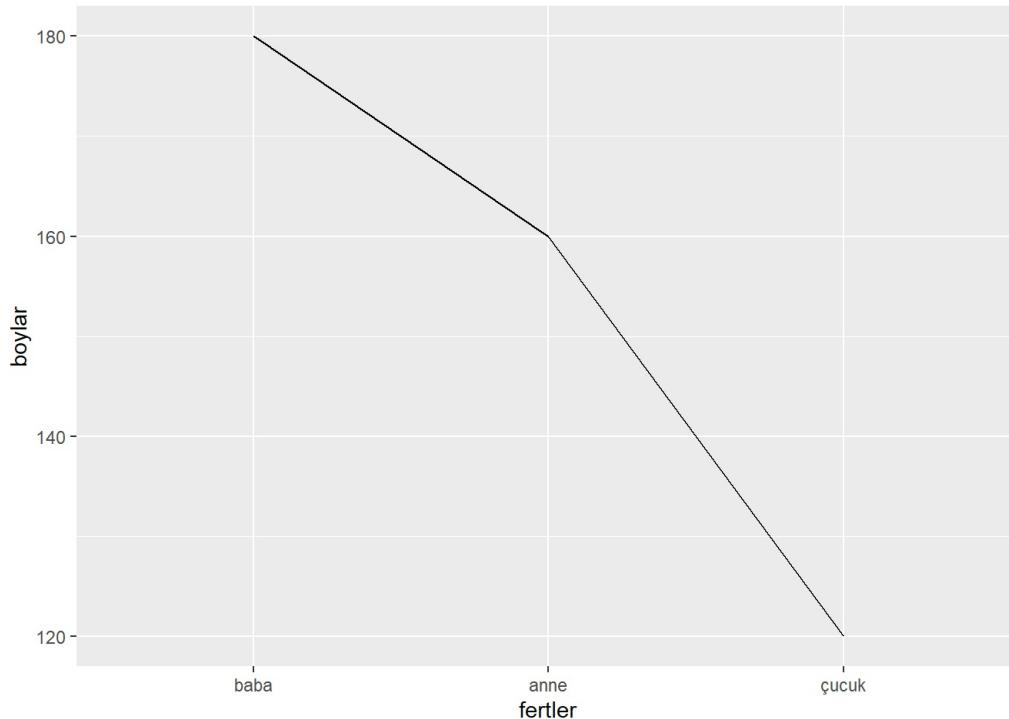
Kategorik değişkenlerin grafiğidir.

```
df <- tibble(
  kullanım = factor(c("hi" , "hs") , levels= c("hi", "hs")),
  sure= c(100,180)
)
ggplot(df, aes( x = kullanım,y= sure , group = 1)) +
  geom_line()
```



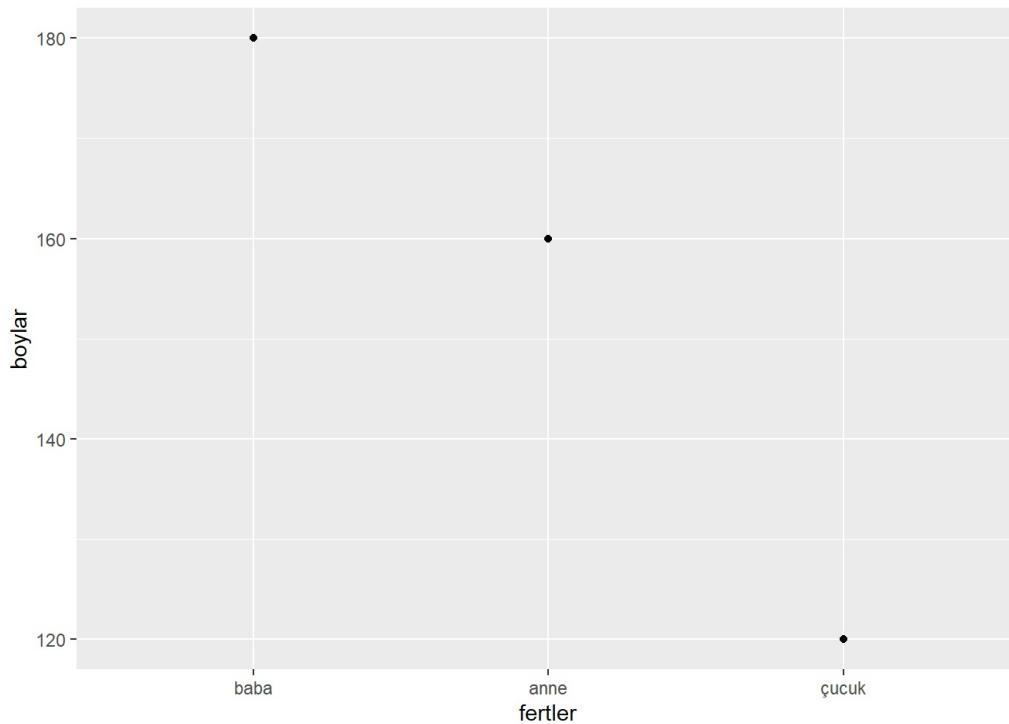
örneğin aile fertlerinin boyalarının çizgi grafiği

```
df <- tibble(
  fertler = factor(c("baba" , "anne" , "çucuk"), levels = c("baba" , "anne" , "çucuk")),
  boyalar= c(180,160,120)
)
ggplot(df, aes( x = fertler, y= boyalar , group = 1)) +
  geom_line()
```



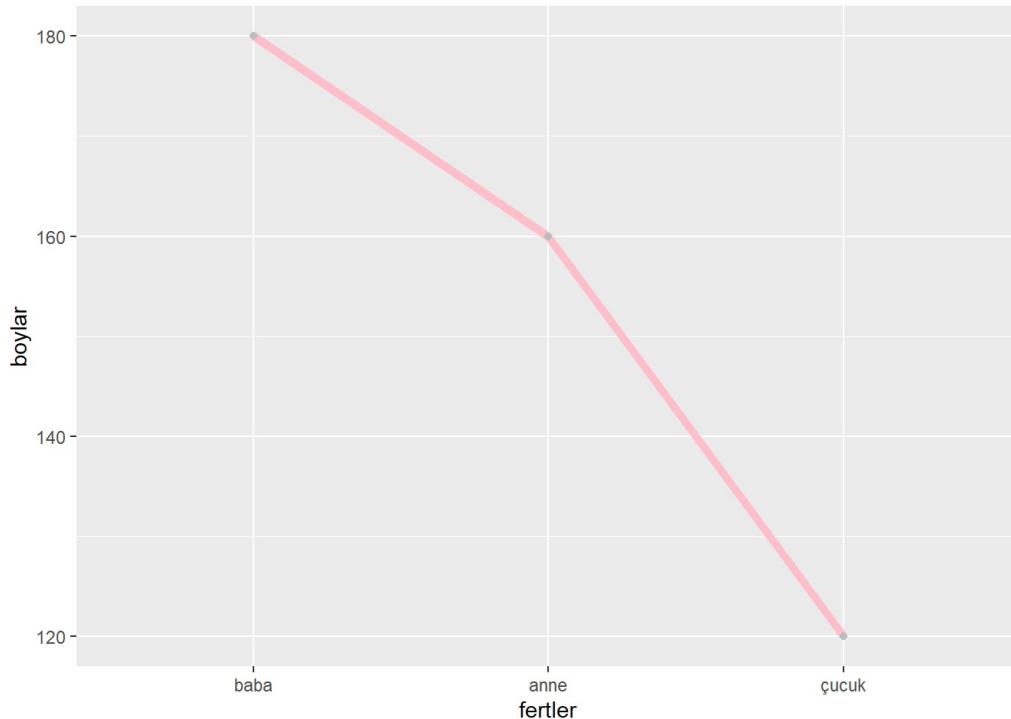
## NOKTA GRAFİĞİ (GEOM\_POINT)

```
df <- tibble(
  fertler = factor(c("baba" , "anne" , "çucuk"), levels = c("baba" , "anne" , "çucuk")),
  boyalar= c(180,160,120)
)
ggplot(df, aes( x = fertler, y= boyalar , group = 1)) +
  geom_point()
```



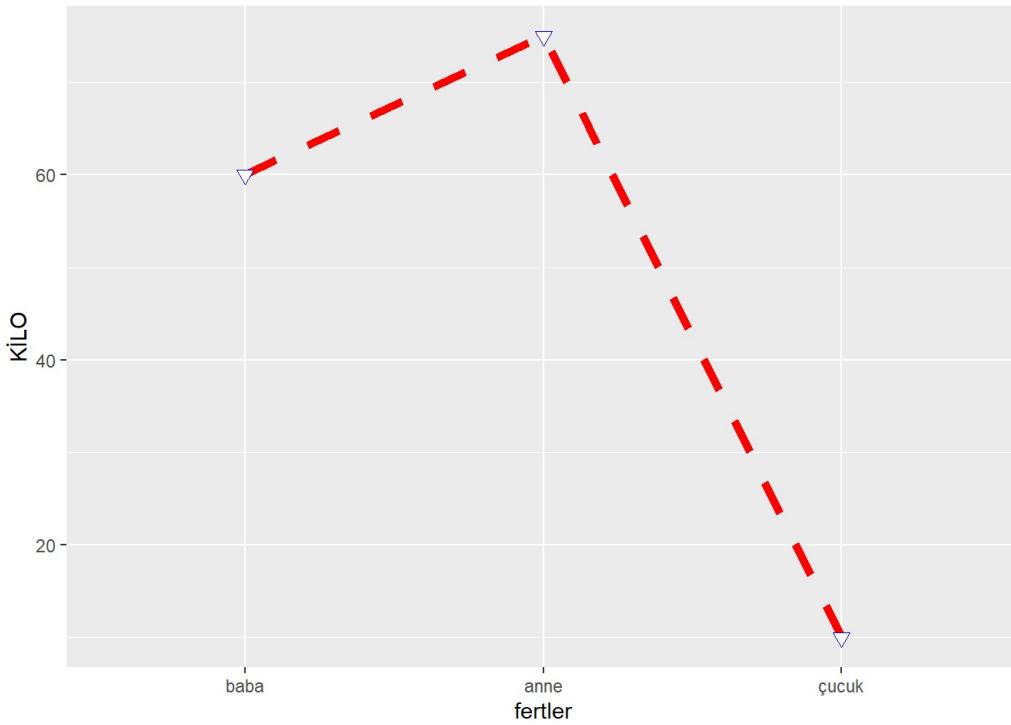
## ÇİZGİ GRAFİĞİ İLE NOKTA GRAFİĞİNİN BİRLEŞİMİ

```
df <- tibble(
  fertler = factor(c("baba" , "anne" , "çucuk"), levels = c("baba" , "anne" , "çucuk")),
  boyalar= c(180,160,120)
)
ggplot(df, aes( x = fertler, y= boyalar , group = 1)) +
  geom_line(colour= "pink" , size = 2.0) +
  geom_point( colour = "gray" , )
```



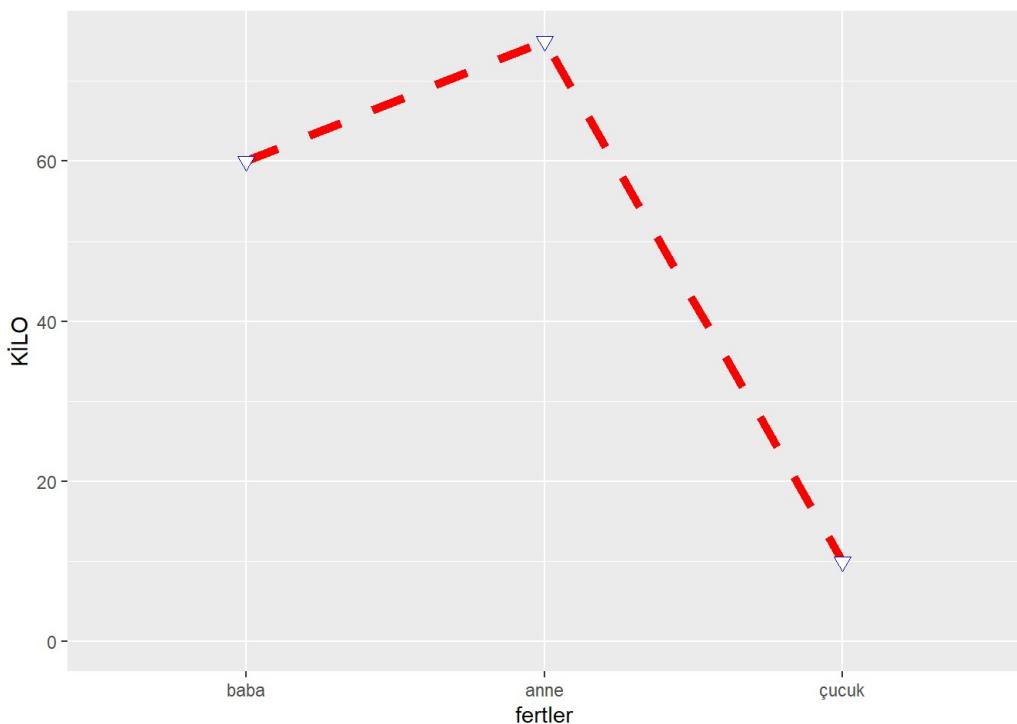
ÖZELLEŞTİREBİLİRİZ ÖRNEĞİN;

```
df <- tibble(
  fertler = factor(c("baba" , "anne" , "çucuk") , levels = c("baba" , "anne" , "çucuk")) ,
  KİLO = c(60,75,10)
)
ggplot(df, aes( x = fertler, y= KİLO , group = 1)) +
  geom_line( colour = "red" , linetype = "dashed" , size= 2.0) +
  geom_point(colour = "blue" , size = 2.5 , shape = 25 , fill = "white")
```



KİLO, yani y eksenini 0'dan başlatabarak grafiği daha anlamlı yapabiliriz. (expand\_limits)

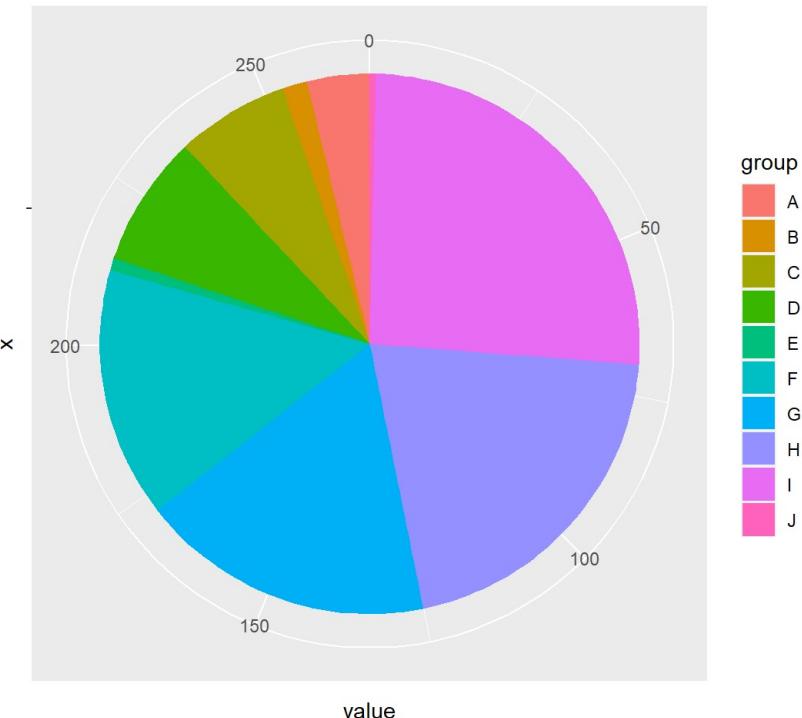
```
df <- tibble(
  fertler = factor(c("baba" , "anne" , "çucuk") , levels = c("baba" , "anne" , "çucuk")) ,
  KİLO = c(60,75,10)
)
ggplot(df, aes( x = fertler, y= KİLO , group = 1)) +
  geom_line( colour = "red" , linetype = "dashed" , size= 2.0) +
  geom_point(colour = "blue" , size = 2.5 , shape = 25 , fill = "white") +
  expand_limits( y = 0)
```



## PASTA GRAFİĞİ

```
data <- data.frame(
  group=LETTERS[1:10],
  value=c(10,4,18,21,2,40,47,55,69,1)
)

ggplot(data, aes(x="", y=value, fill=group)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)
```



## 1.2 SÜREKLİ DEĞİŞKEN

verimiz günlük telefon kullanımı. iki tane kategorik değişkenim var.

```
df <- tibble(
  cinsiyet = factor(rep(c("e" , "k") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
df
```

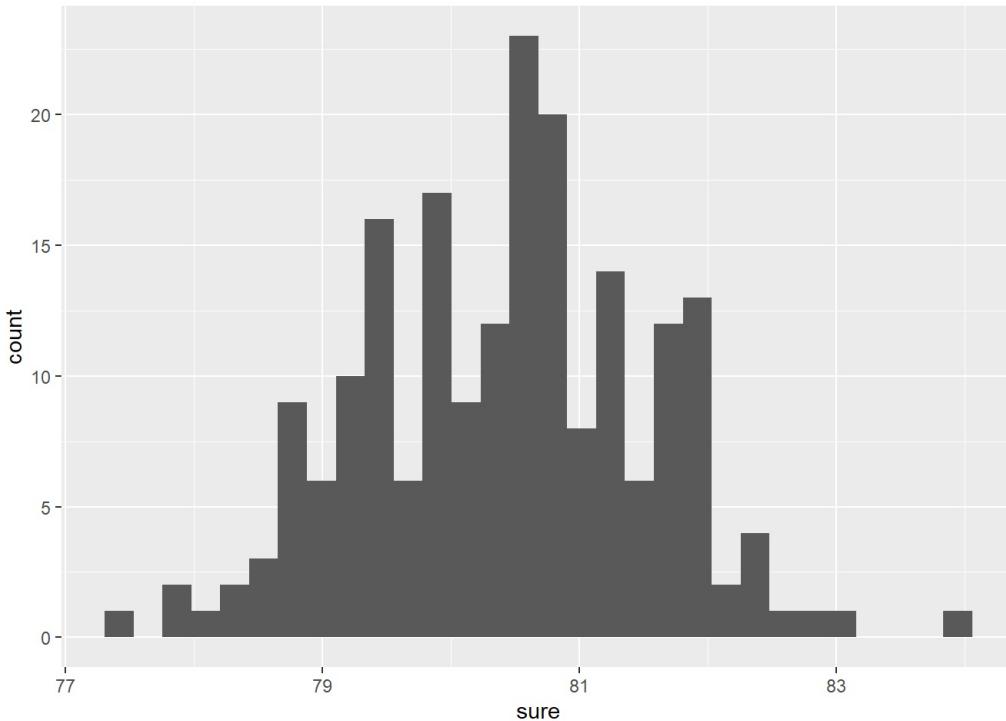
```
## # A tibble: 200 x 2
##   cinsiyet    sure
##   <fct>     <dbl>
## 1 e          78.7
## 2 e          79.2
## 3 e          79.0
## 4 e          81.4
## 5 e          79.7
## 6 e          78.5
## 7 e          79.5
## 8 e          78.3
## 9 e          79.0
## 10 e         78.4
## # ... with 190 more rows
```

ben bu veriyi nasıl görselleştirebilirim?

## HİSTOGRAM (*GEOM\_HISTOGRAM*)

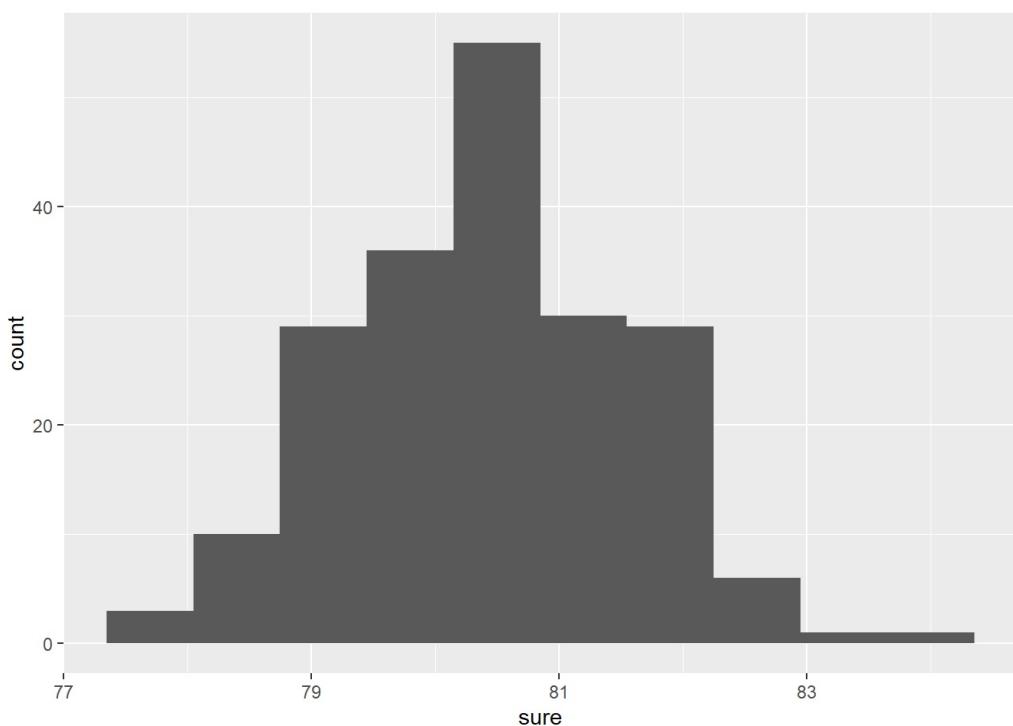
```
ggplot(df, aes(sure)) +
  geom_histogram()
```

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



daha okunabilir olması için;

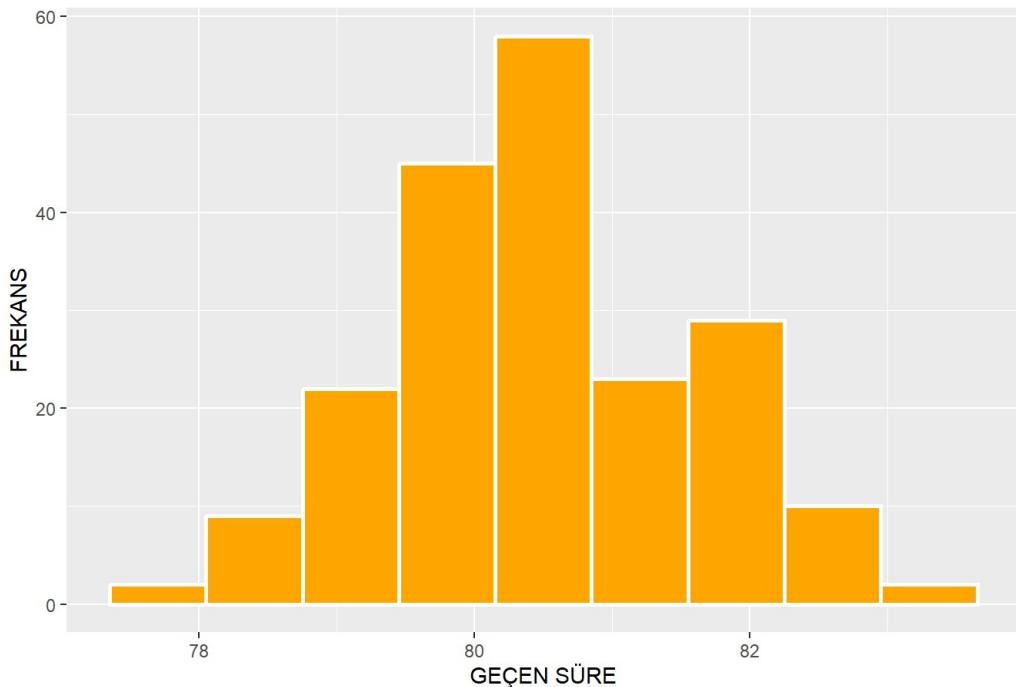
```
ggplot(df, aes(sure)) +
  geom_histogram(binwidth = 0.7)
```



Özelleştirelim;

```
df <- tibble(
  cinsiyet = factor(rep(c("e" , "k") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure)) +
  geom_histogram(binwidth = 0.7, colour = "white" , fill= "orange" , size=1) +
  xlab("GEÇEN SÜRE") + ylab("FREKANS") +
  ggtitle("TELEFONDA GEÇİRİLEN SÜRE VE KİŞİ SAYISI HİSTOGRAM GRAFİĞİ")
```

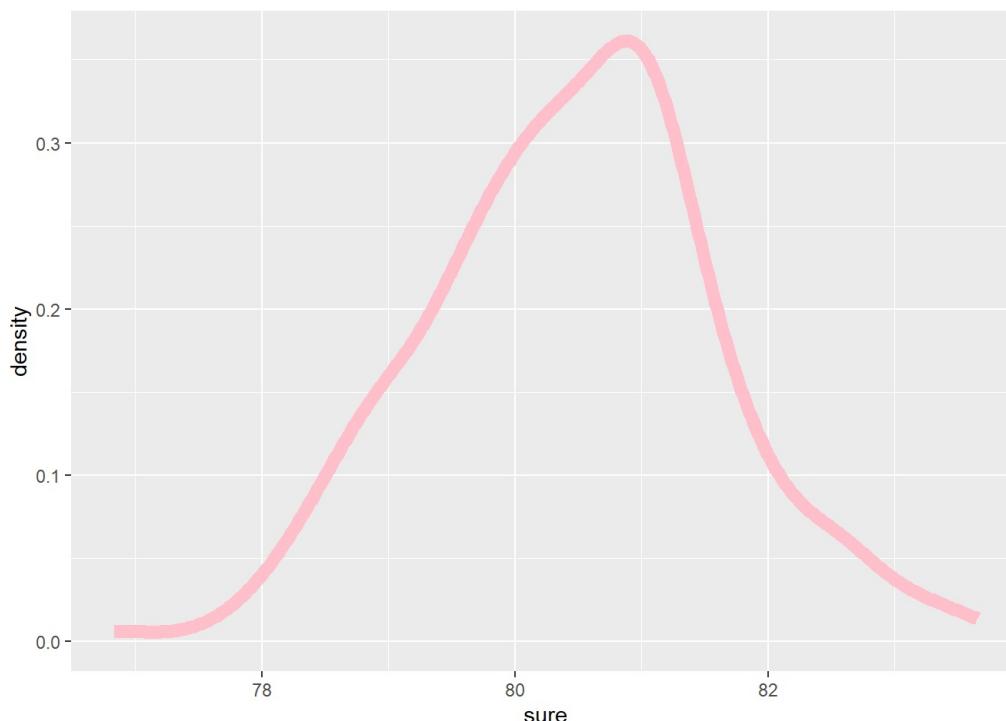
TELEFONDA GEÇİRİLEN SÜRE VE KİŞİ SAYISI HİSTOGRAM GRAFİĞİ



YORUM = Telefonda 78 dakika süre geçiren kişi sayısı 10 ile 0 arasındadır.

## YOĞUNLUK FONKSİYONU (*geom\_density*)

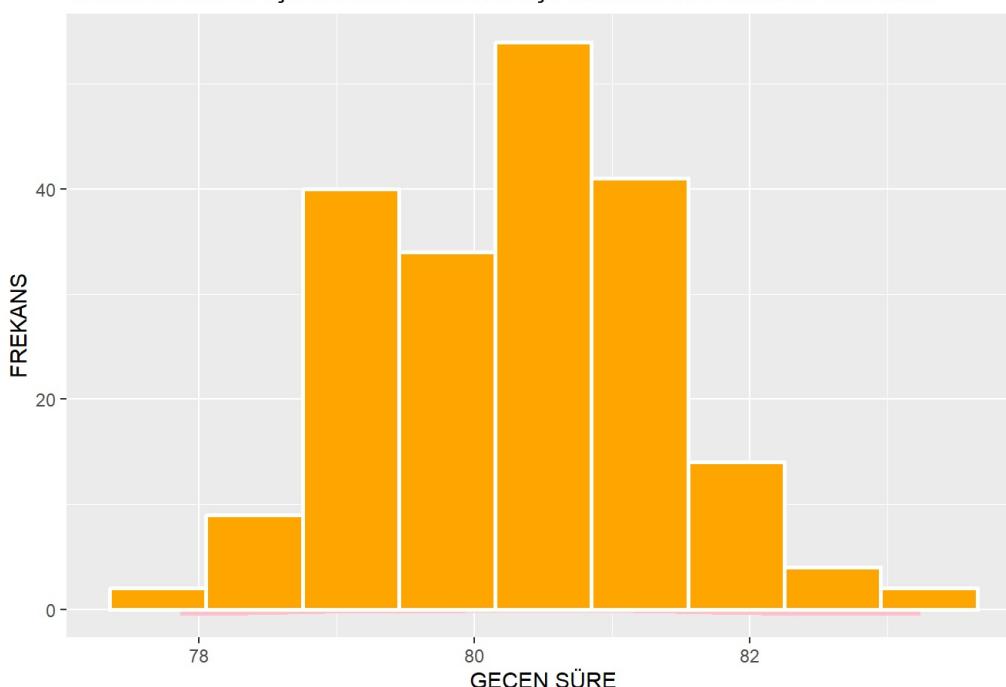
```
df <- tibble(
  cinsiyet = factor(rep(c("e" , "k") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure)) +
  geom_density(colour = "pink" , size= 3)
```



HİSTOGRAMIN ÜSTÜNE ÇİZDİRMEK İSTİYORUM AMA YEKSENİ UYUŞMAZLIĞI YAŞANIYOR BKNZ

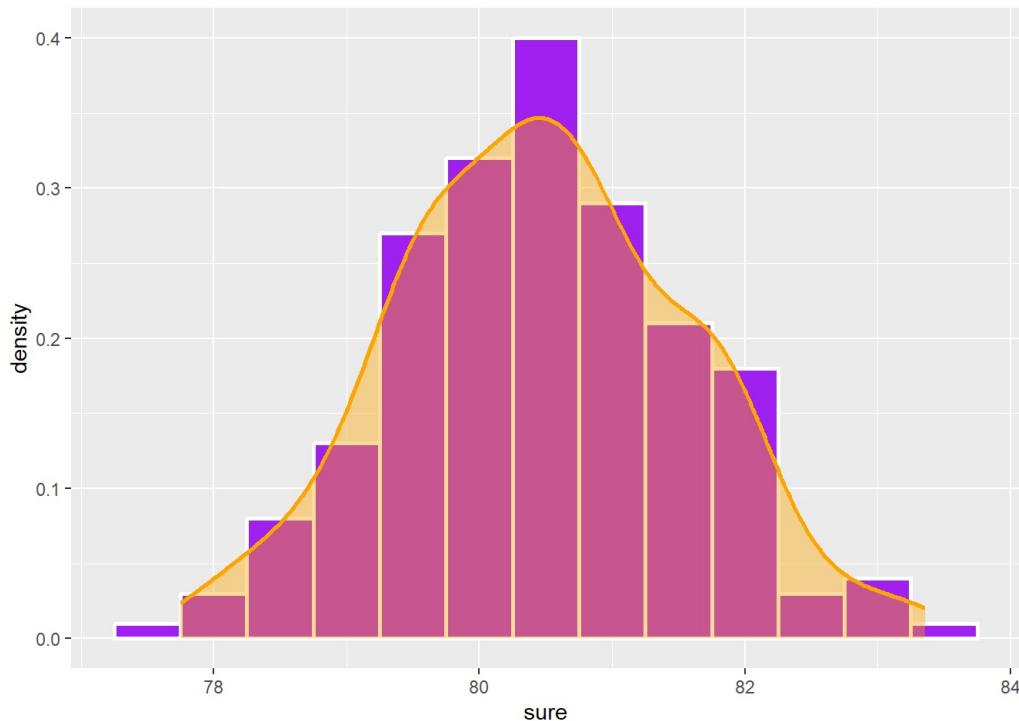
```
df <- tibble(
  cinsiyet = factor(rep(c("e" , "k") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure)) +
  geom_density(colour = "pink" , size= 3) +
  geom_histogram(binwidth = 0.7, colour = "white" , fill= "orange" , size=1) +
  xlab("GEÇEN SÜRE") + ylab("FREKANS") +
  ggtitle("TELEFONDA GEÇİRİLEN SÜRE VE KİŞİ SAYISI HİSTOGRAM GRAFİĞİ")
```

TELEFONDA GEÇİRİLEN SÜRE VE KİŞİ SAYISI HİSTOGRAM GRAFİĞİ



PEKİ NASIL DÜZELTİRİZ?

```
df <- tibble(
  cinsiyet = factor(rep(c("e" , "k") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure)) +
  geom_histogram(aes(y=..density..) , binwidth = 0.5, colour = "white" , fill= "purple" , size=1) +
  geom_density( alpha= 0.4, fill= "orange" , colour = "orange" , size= 1)
```



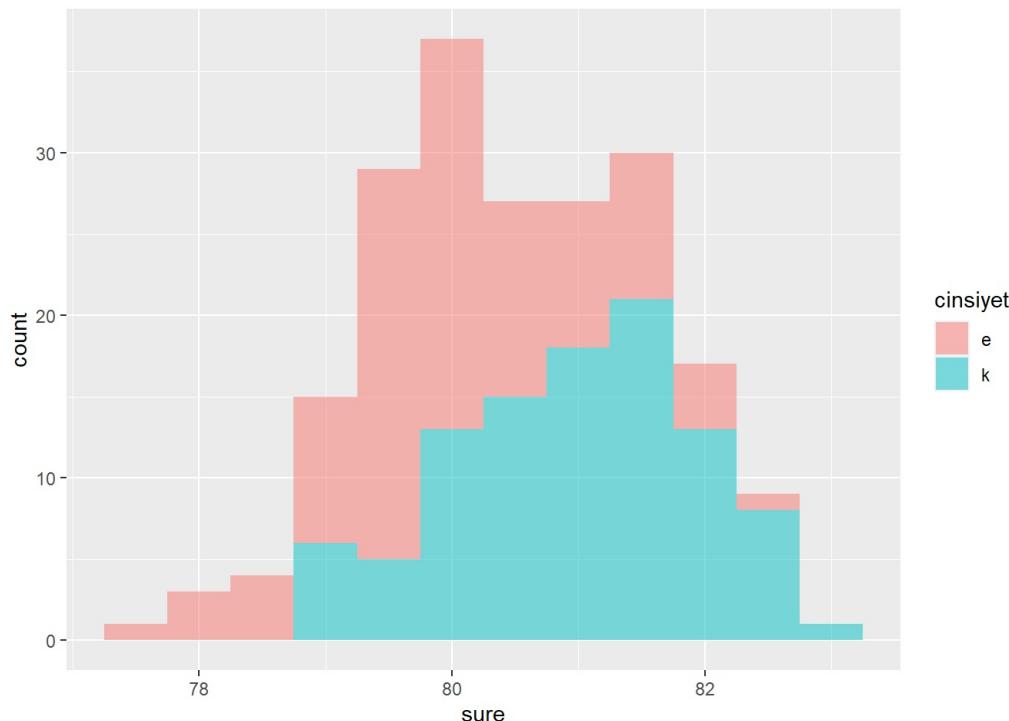
not= alpha transparan özellik

#### KIRILMALARA GÖRE HİSTOGRAM VE YOGUNLUK GRAFİĞİ

Elimizde bir veri seti var. Bu verimizde bir adet sürekli bir adet kategorik veri seti var. normalilik sağa veya sola çarpıklık direkt histogramı yazdırarak yorumlamak doğru değil. bu yorumu yapabilmek için kırılıp tekrar göstermemiz gereklidir.

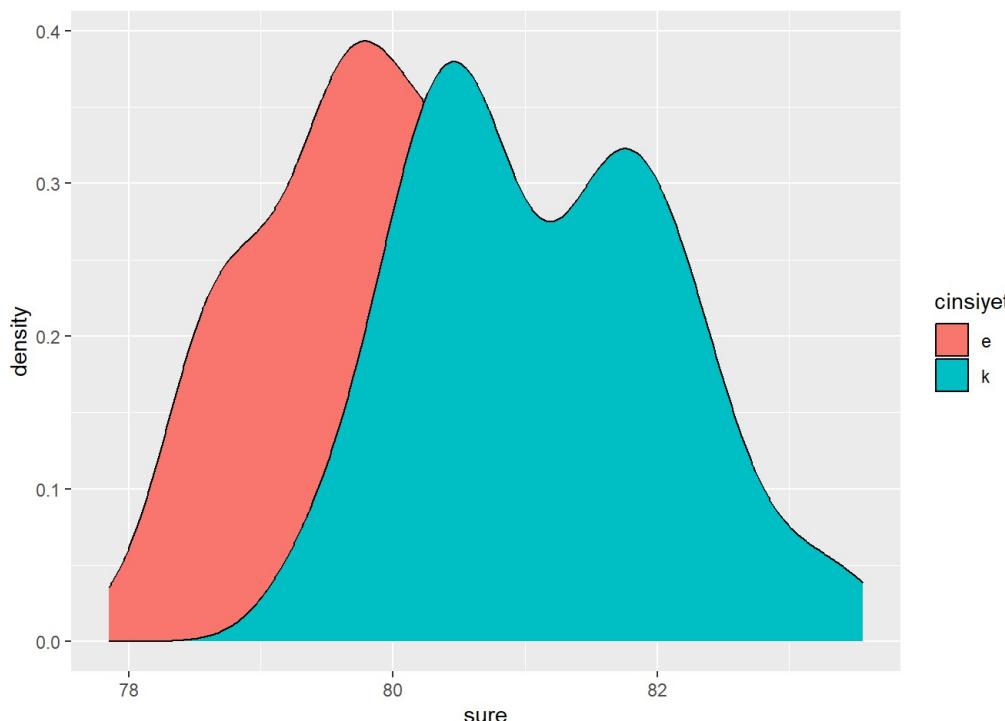
#### histogram

```
df <- tibble(
  cinsiyet = factor(rep(c("e" , "k") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure , fill = cinsiyet)) +
  geom_histogram(binwidth = 0.5, alpha= 0.5, possiton="identity")
```



#### yoğunluk

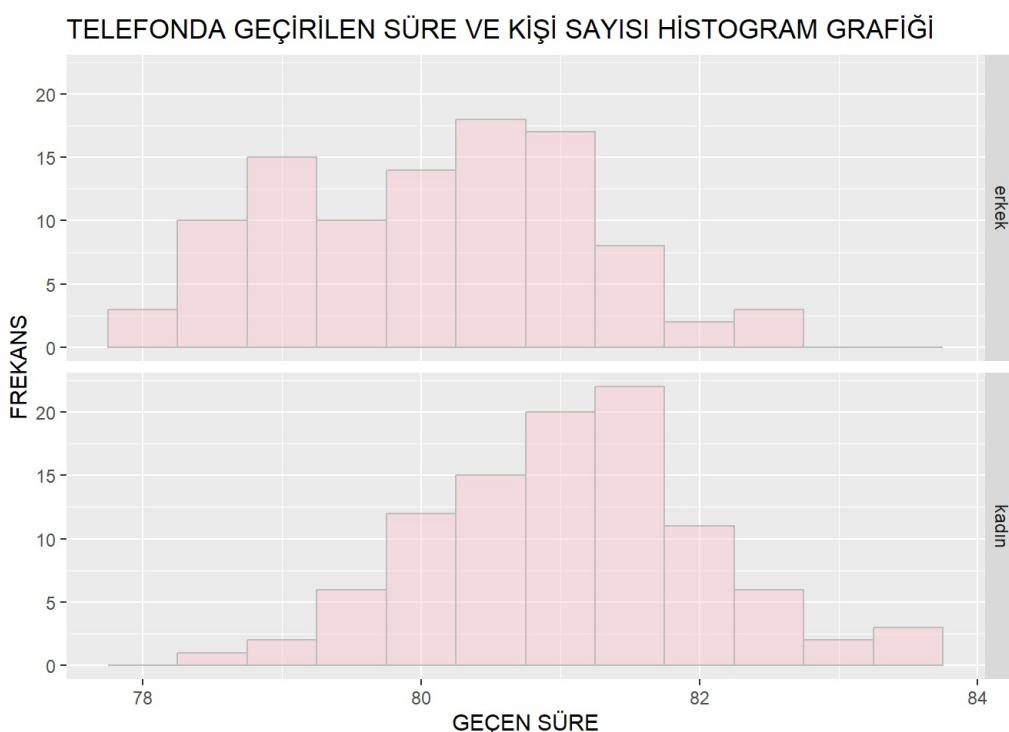
```
df <- tibble(
  cinsiyet = factor(rep(c("e" , "k") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure , fill = cinsiyet)) +
  geom_density()
```



#### KIRILMALARI İKİ AYRI GRAFİKLE GÖSTERMEK İÇİN (facet\_grid())

grafikleri bir arada (yukarıdaki örnek) ya da farklı iki ayrı grafik olarak gösterebiliriz.

```
df <- tibble(
  cinsiyet = factor(rep(c("erkek" , "kadın") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure , fill = cinsiyet)) +
  geom_histogram(binwidth = 0.5 , colour = "gray" , fill = "pink" , alpha= 0.4)+ 
  facet_grid(cinsiyet ~ .) +
  xlab("GEÇEN SÜRE") + ylab("FREKANS") +
  ggtitle("TELEFONDA GEÇİRİLEN SÜRE VE KİŞİ SAYISI HİSTOGRAM GRAFİĞİ")
```



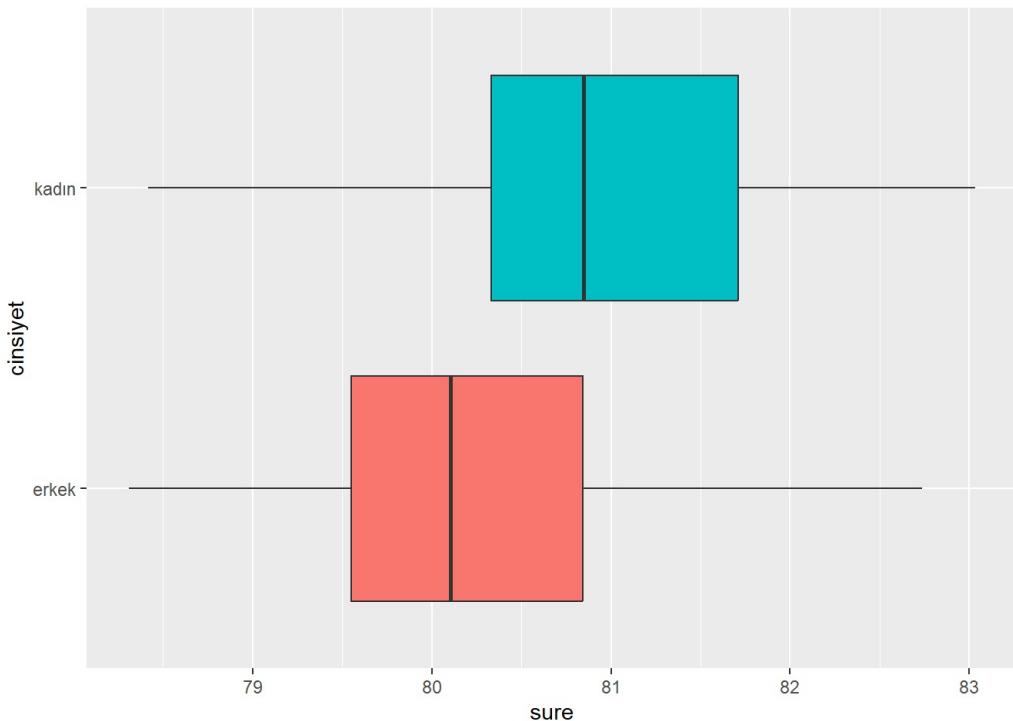
#### BOX / KUTU GRAFİĞİ (geom\_boxplot)

Verinin kümelenmesini, yiğilmasının dağılımlarını gösterir.

```

df <- tibble(
  cinsiyet = factor(rep(c("erkek" , "kadın") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes( x = cinsiyet , y= sure , fill =cinsiyet)) +
  geom_boxplot()+
  guides(fill= FALSE)+ 
  coord_flip()

```



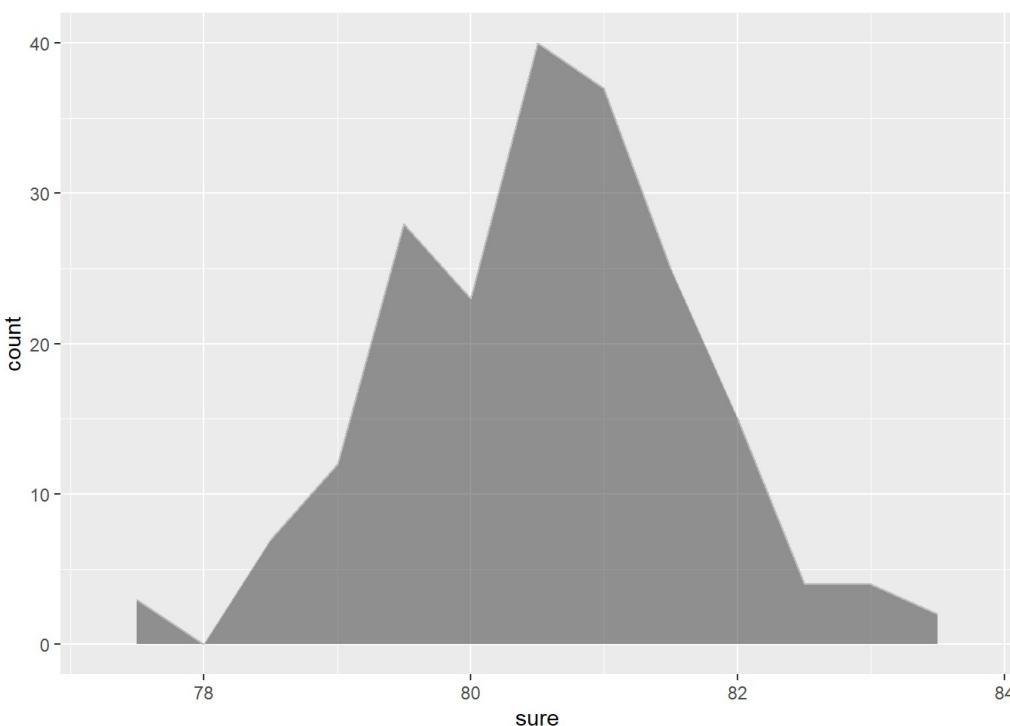
not; `coord_flip` = şekil yataydan dikey hale getirir. `guides(fill= FALSE)` ise grafiğin yanında gereksiz açıklamavardı onu siler.

## ALAN GRAFİĞİ (`geom_area`)

```

df <- tibble(
  cinsiyet = factor(rep(c("erkek" , "kadın") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure)) +
  geom_area(stat="bin" , binwidth= 0.5, colour = "gray" , alpha= .5)

```

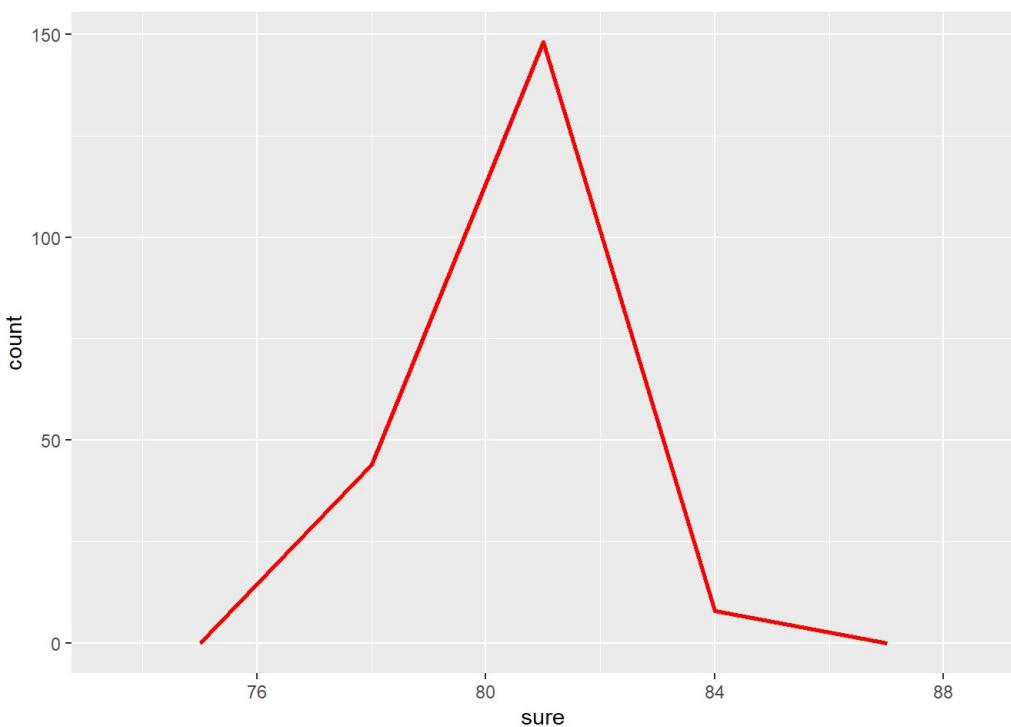


## FREKANS GRAFİĞİ (`geom_freqpoly`)

```

df <- tibble(
  cinsiyet = factor(rep(c("erkek" , "kadın") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure)) +
  geom_freqpoly(colour = "red" , size = 1 , bins = 3)

```

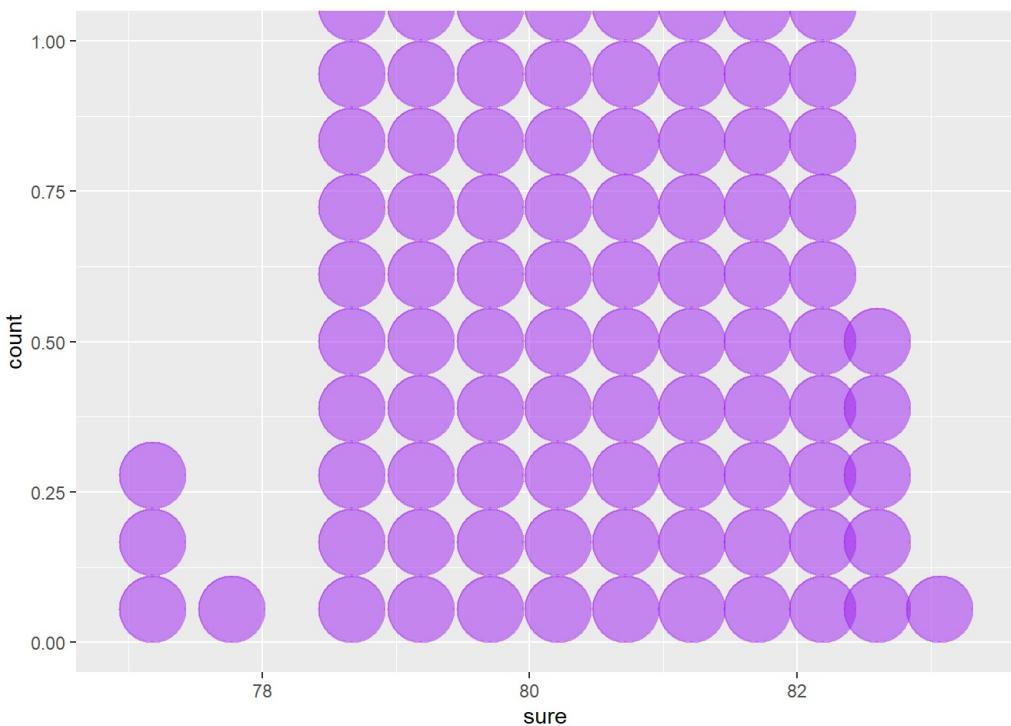


### DOT PLOT (*geom\_dotplot*)

```

df <- tibble(
  cinsiyet = factor(rep(c("erkek" , "kadın") , each = 100)),
  sure = c(rnorm(100, mean=80) , rnorm(100, mean = 81))
)
ggplot(df, aes(sure)) +
  geom_dotplot(binwidth = 0.5 , colour= "purple" , fill ="purple" , alpha= 0.5)

```



## 2. İKİ DEĞİŞKENLİ GÖRSELLEŞTİRME

### 2.1 SÜREKLİ X , SÜREKLİ Y

## SCATTER-PLOT(en çok kullanılan)

verilen iki sürekli değişkenin birbirleri ile ilişkisini inceler.

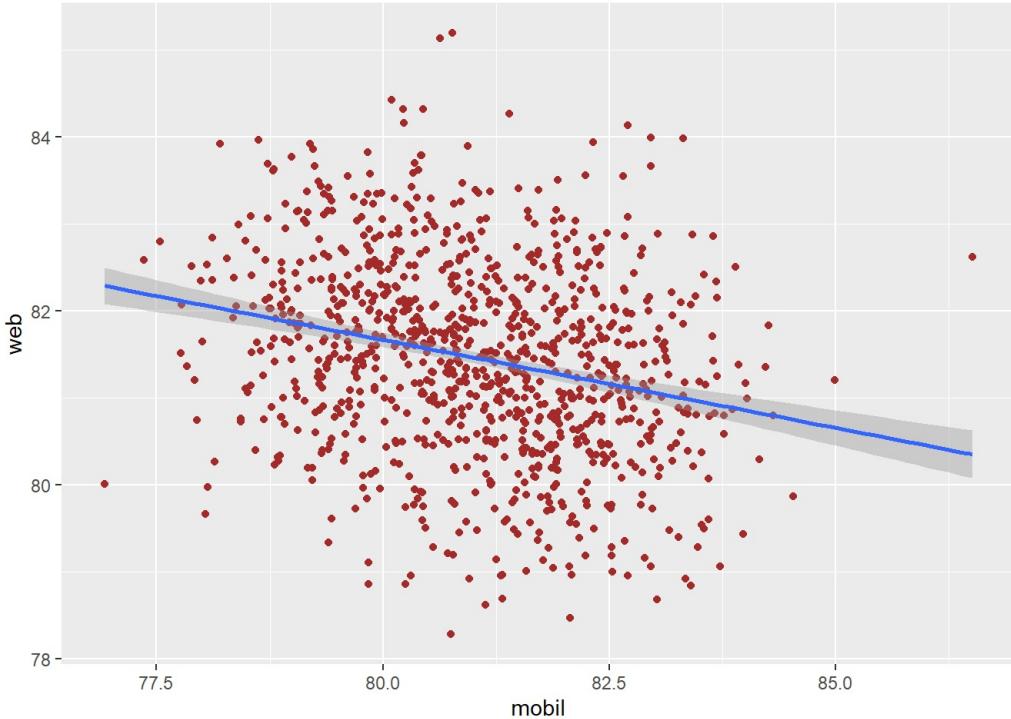
```
df <- tibble(
  cinsiyet = factor(rep(c("E", "K") , each=500)),
  mobil = c(rnorm(500, mean= 80), rnorm(500, mean=82)),
  web = c(rnorm(500, mean= 82) , rnorm(500, mean= 81))
)
df
```

```
## # A tibble: 1,000 × 3
##   cinsiyet mobil   web
##   <fct>     <dbl> <dbl>
## 1 E          81.5  82.0
## 2 E          80.8  81.7
## 3 E          79.9  83.6
## 4 E          80.8  81.5
## 5 E          80.9  82.2
## 6 E          80.1  82.9
## 7 E          79.0  82.4
## 8 E          81.2  82.4
## 9 E          79.7  83.0
## 10 E         79.0  82.0
## # ... with 990 more rows
```

soru = mobil ile web de geçirilen süre arasında bir korelasyon ilişkisi var mı?

```
ggplot(df, aes(mobil,web)) +
  geom_point(colour= "brown") +
  geom_smooth(method= lm )
```

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

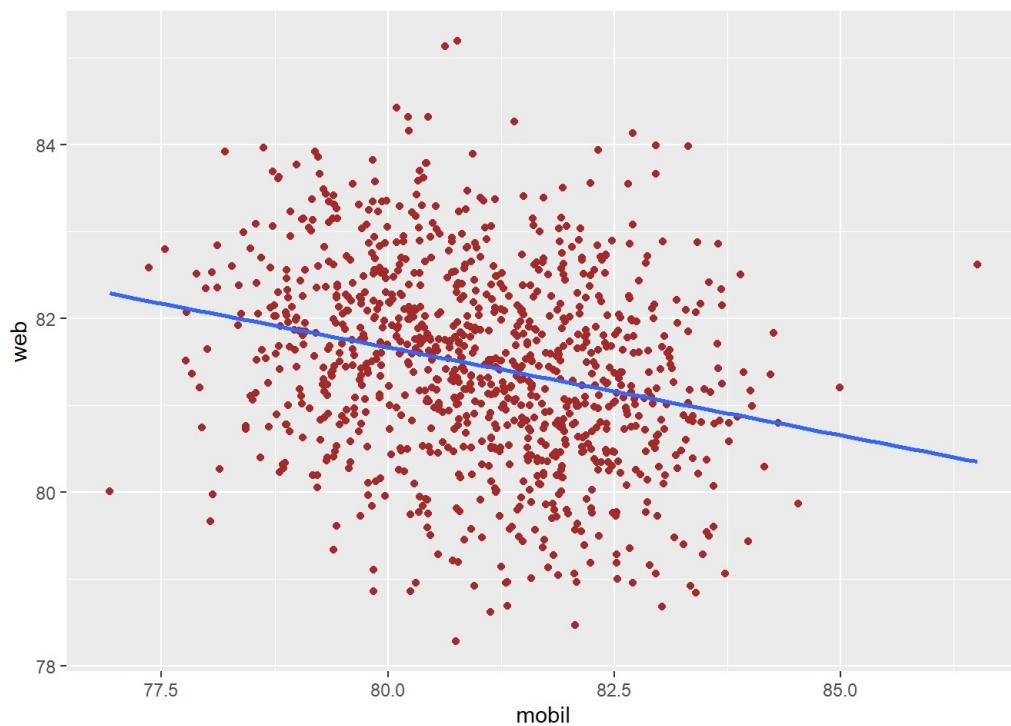


yorum= mobil kullanımı arttığında web kullanımı artmıyor.

geom\_smooth(method= lm) lm = lineer method

```
ggplot(df, aes(mobil,web)) +
  geom_point(colour= "brown") +
  geom_smooth(method= lm , se = FALSE)
```

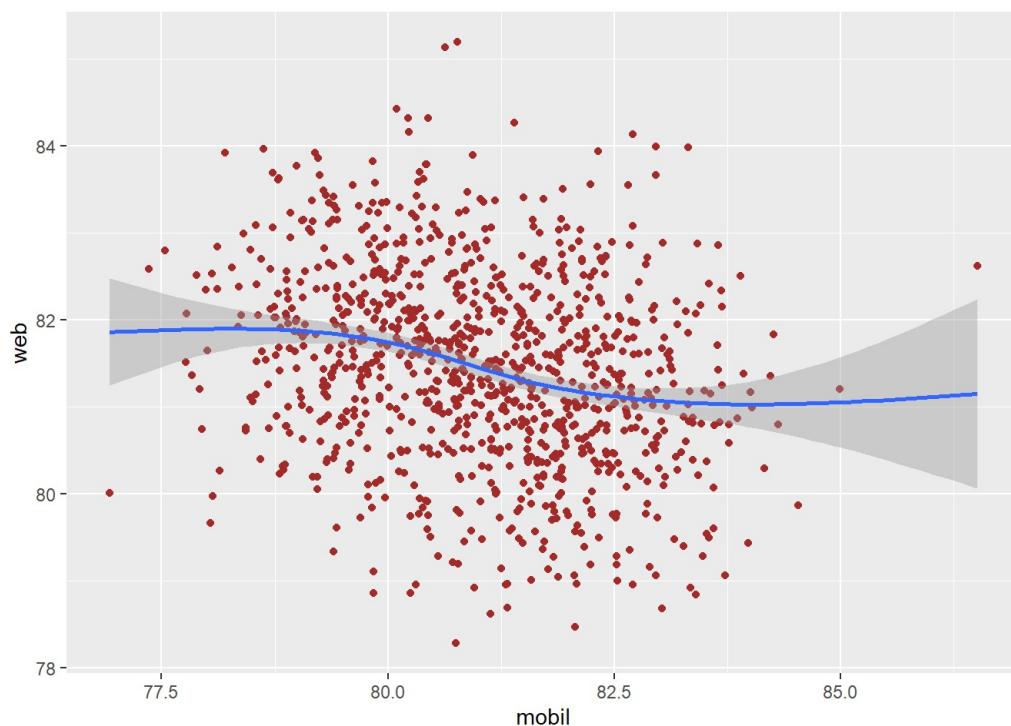
```
## `geom_smooth()` using formula = 'y ~ x'
```



se = FALSE = standard error demek. Çizginin etrafındaki yoğunluk gider.

```
ggplot(df, aes(mobil,web)) +
  geom_point(colour= "brown") +
  geom_smooth()
```

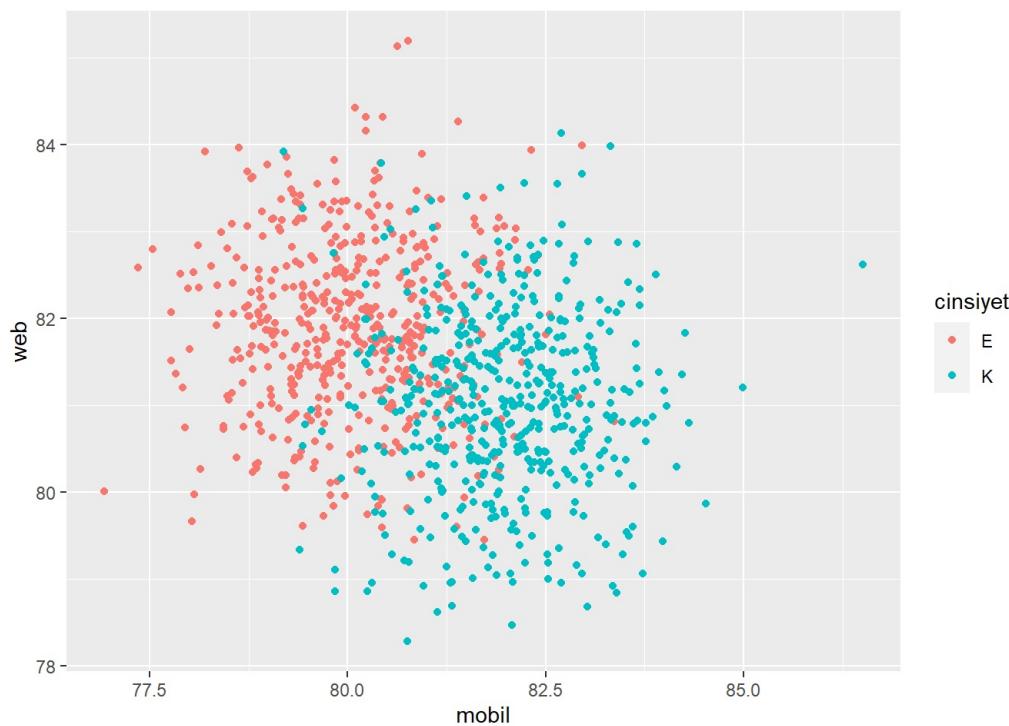
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



veriyi temsil kabileyeti daha kuvvetli bir veridir.

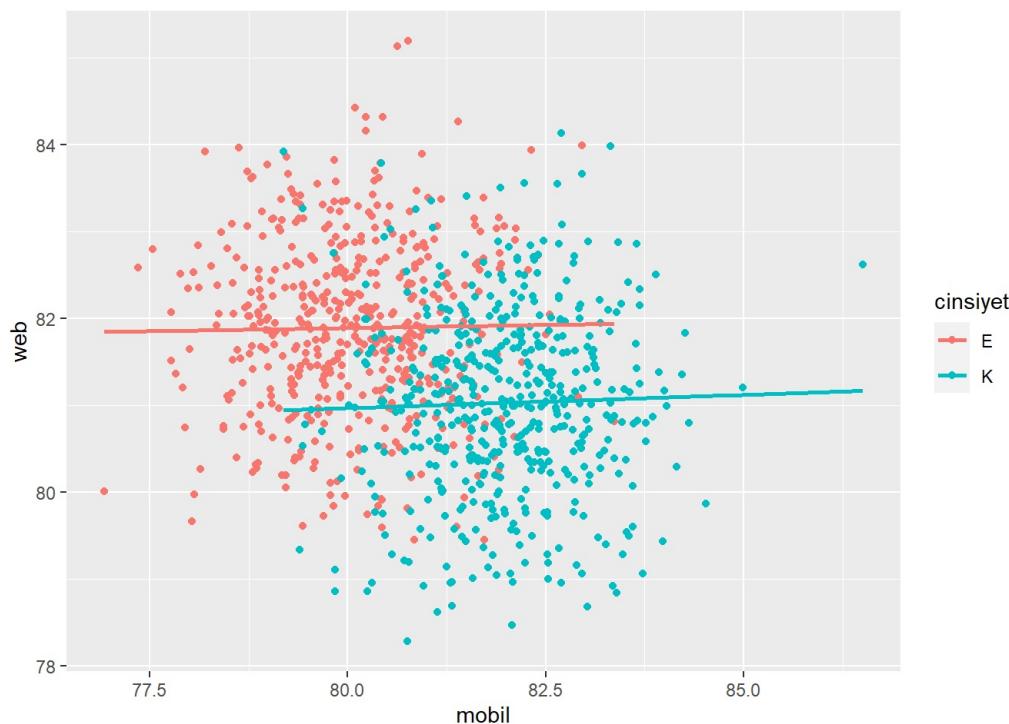
CİNSİYETE GÖRE BAKALIM;

```
ggplot(df, aes(mobil,web, color= cinsiyet)) +
  geom_point()
```



Renkleri ayırdı. hangisinin kadın olduğunu hangisinin erkek olduğunu görebiliyoruz.

```
ggplot(df, aes(mobil, web, color = cinsiyet)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE, formula = 'y ~ x')
```

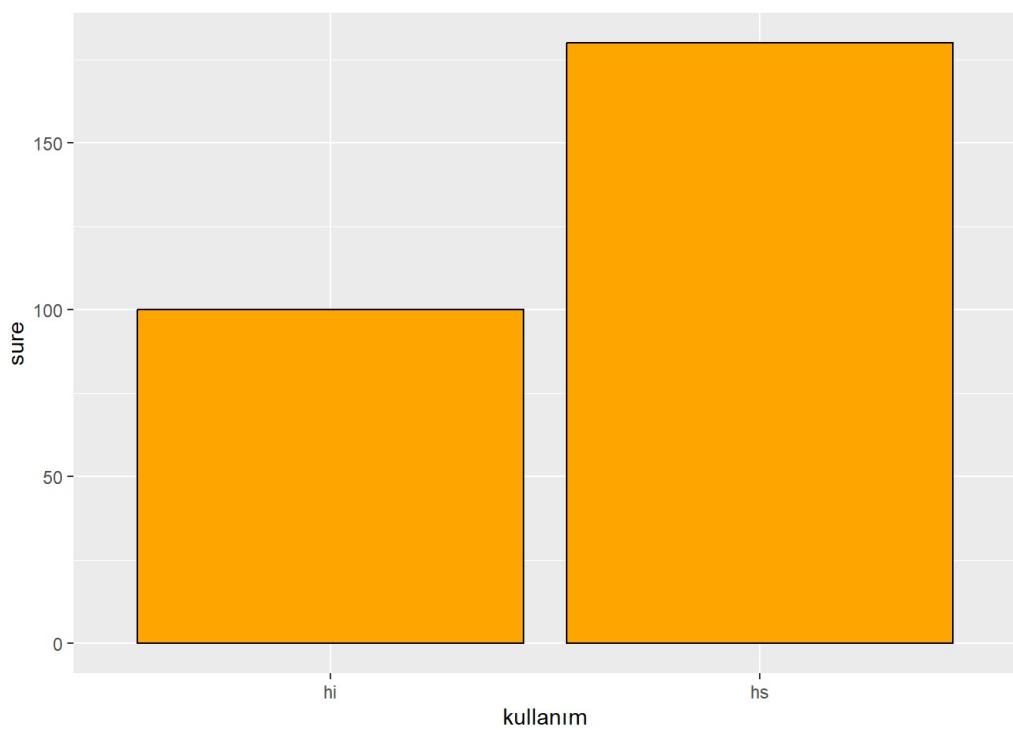


eğimleri değişmiş oldu.

## 2.2 Kesikli X, Sürekli Y

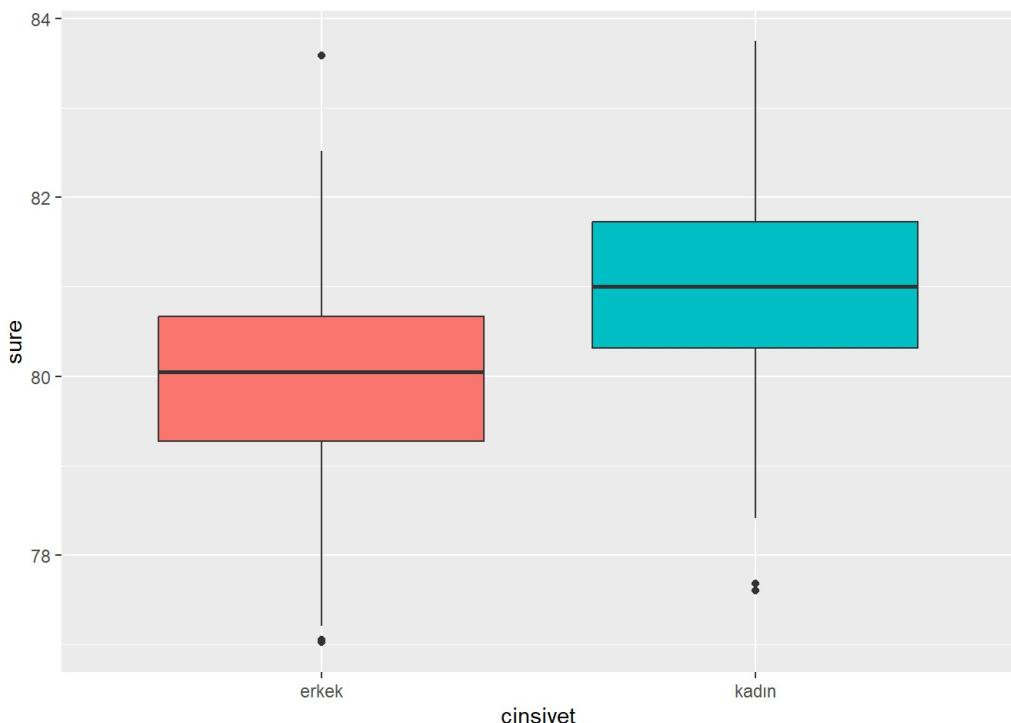
Sütun grafikleri (*geom\_bar*)

```
df <- tibble(
  kullanım= factor(c("hi" , "hs"), levels= c("hi", "hs")),
  sure =c(100,180)
)
ggplot(df, aes(kullanım , sure , fill =kullanım)) +
  geom_bar(colour = "black", fill= "orange", stat ="identity")
```



## BOX PLOT (*geom\_boxplot*)

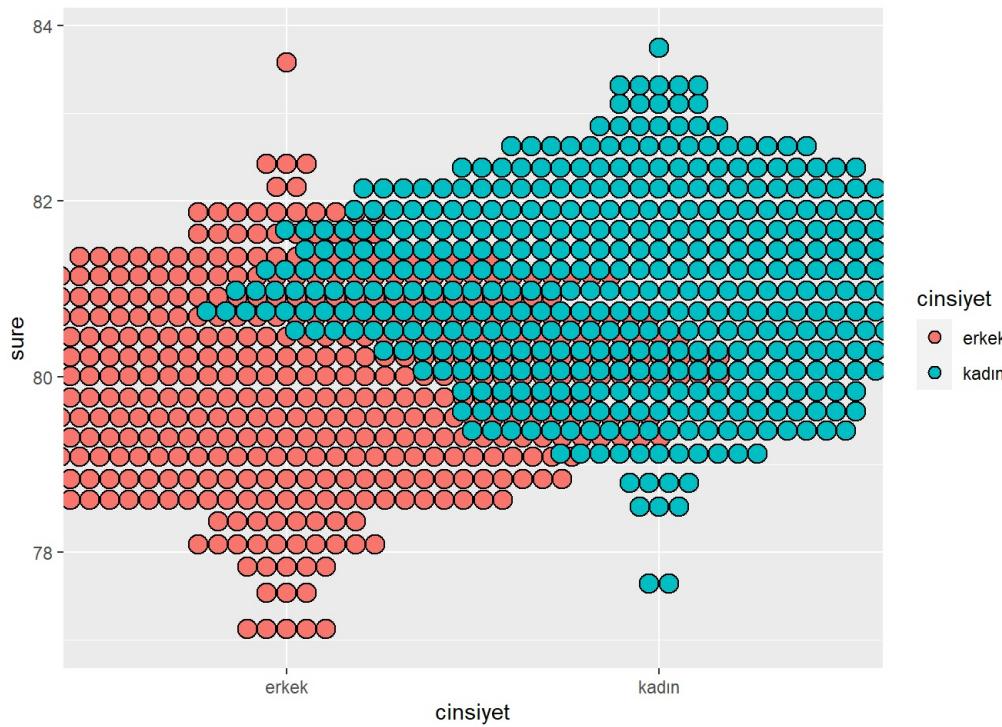
```
df <- tibble(
  cinsiyet = factor(rep(c("erkek" , "kadın") , each = 500)),
  sure = c(rnorm(500, mean=80) , rnorm(500, mean = 81))
)
ggplot(df, aes( x = cinsiyet , y= sure , fill =cinsiyet)) +
  geom_boxplot()+
  guides(fill= FALSE)
```



## ####DOT PLOT (*geom\_dotplot*)

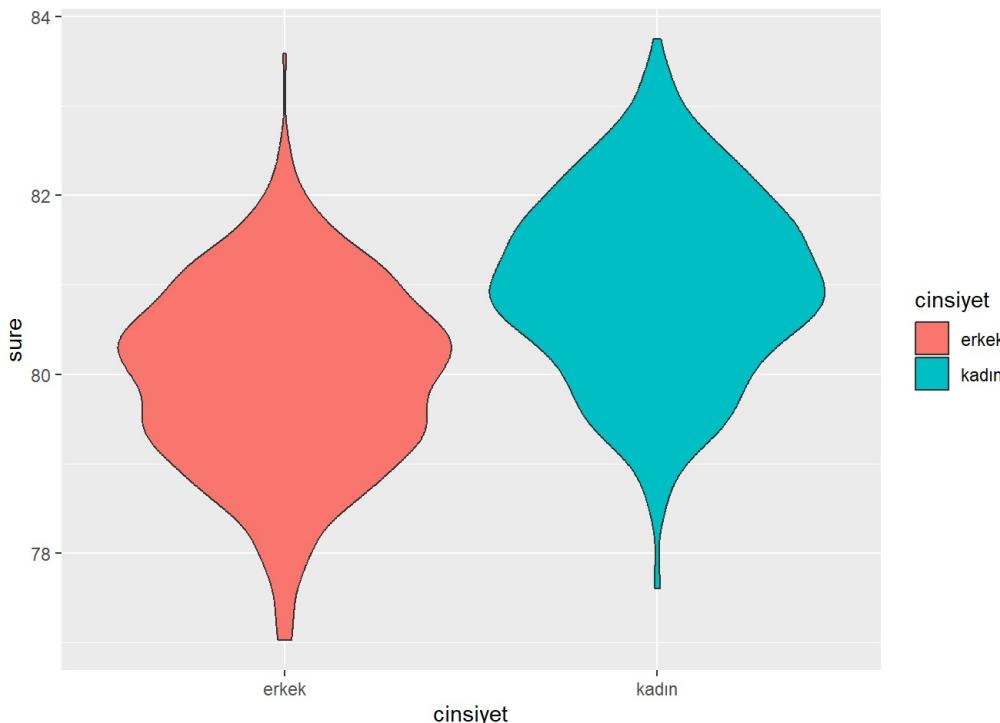
```
ggplot(df, aes( x = cinsiyet , y= sure , fill =cinsiyet)) +
  geom_dotplot(binaxis = "y" , stackdir = "center")
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with
## `binwidth`.
```



## VIOLİN (geom\_violin)\*

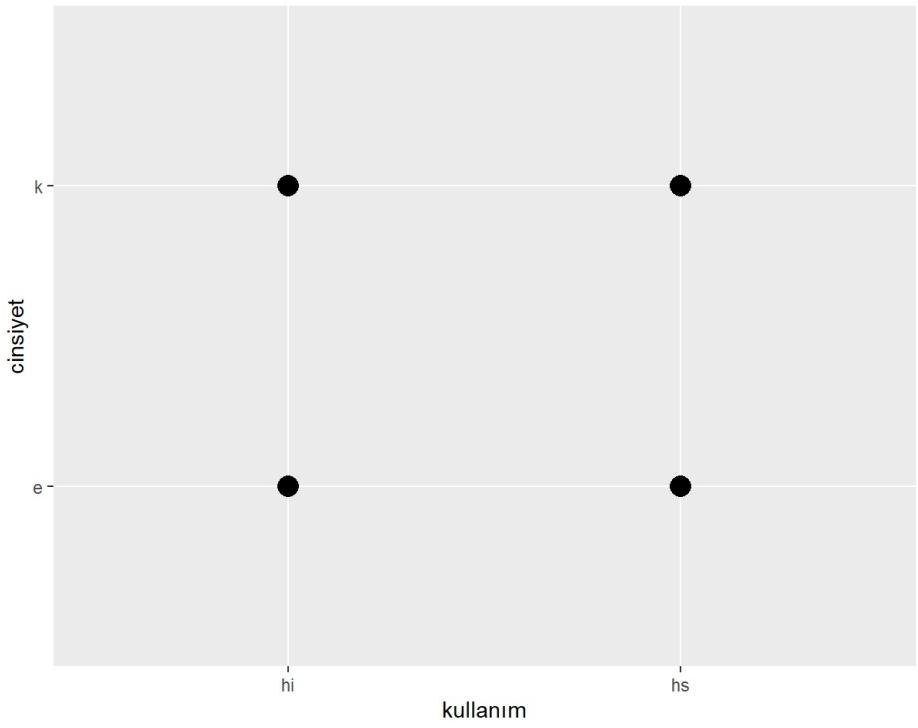
```
ggplot(df, aes( x = cinsiyet , y= sure , fill =cinsiyet)) +
  geom_violin(scale= "area")
```



## 2.3 KESİKLİ X, KESİKLİ Y

(ikisi de kategorik)

```
df<- tibble(
  kullanım = factor(c("hi","hs","hi", "hs"), levels=c("hi","hs")),
  sure=c(100,180,90,200),
  cinsiyet= factor(c("e","e","k","k"))
)
ggplot(df, aes(kullanım, cinsiyet)) +
  geom_count()
```

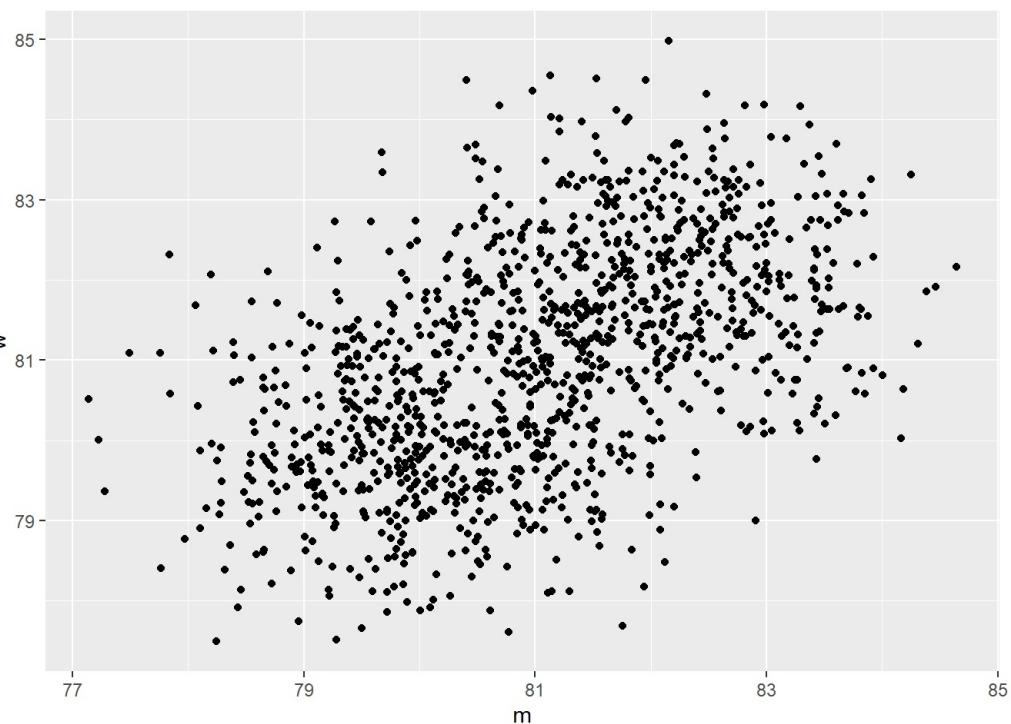


### 3. GRAFİK BÖLME

```
df<- tibble(
  c= factor(rep(c("e","k") , each=600)),
  m = c(rnorm(600, mean= 80), rnorm(600, mean=82)),
  w= c(rnorm(600, mean= 80), rnorm(600, mean=82)),
  me= factor(rep(c("a","b","c") , each= 400))
)
df
```

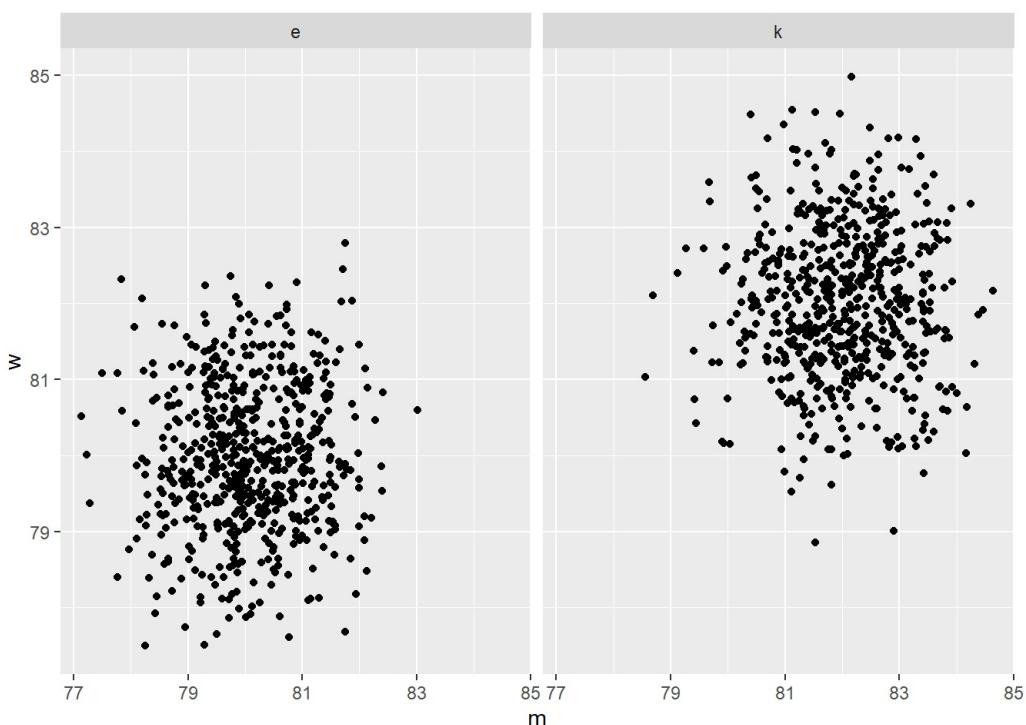
```
## # A tibble: 1,200 × 4
##   c      m     w me
##   <fct> <dbl> <dbl> <fct>
## 1 e     80.6  79.2 a
## 2 e     77.7  78.9 a
## 3 e     80.1  79.8 a
## 4 e     80.2  77.6 a
## 5 e     80.2  78.8 a
## 6 e     79.8  79.4 a
## 7 e     79.1  79.7 a
## 8 e     79.6  81.8 a
## 9 e     81.5  79.2 a
## 10 e    79.7  80.3 a
## # ... with 1,190 more rows
```

```
df<- tibble(
  c= factor(rep(c("e","k") , each=600)),
  m = c(rnorm(600, mean= 80), rnorm(600, mean=82)),
  w= c(rnorm(600, mean= 80), rnorm(600, mean=82)),
  me= factor(rep(c("a","b","c") , each= 400))
)
t <- ggplot(df, aes(m,w)) +
  geom_point()
t
```



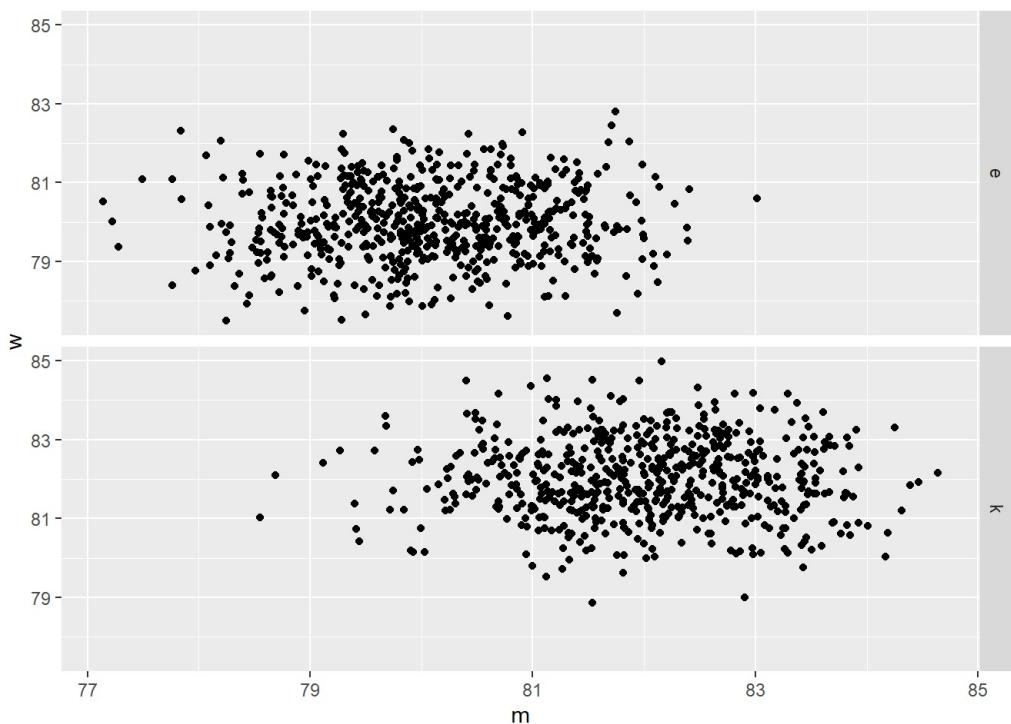
sütun bazında bölme

```
t + facet_grid(. ~ c)
```

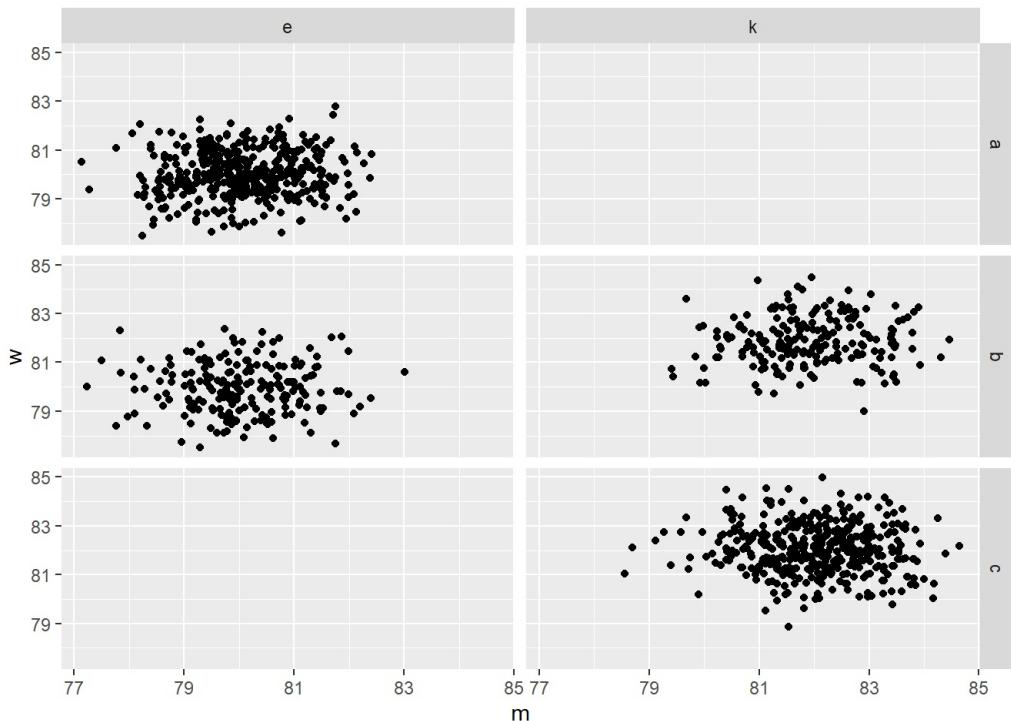


satırda bölmek için (yıllarda falan kullanıyoruz genelde)

```
t + facet_grid( c~ . )
```



```
t + facet_grid(me ~ c)
```

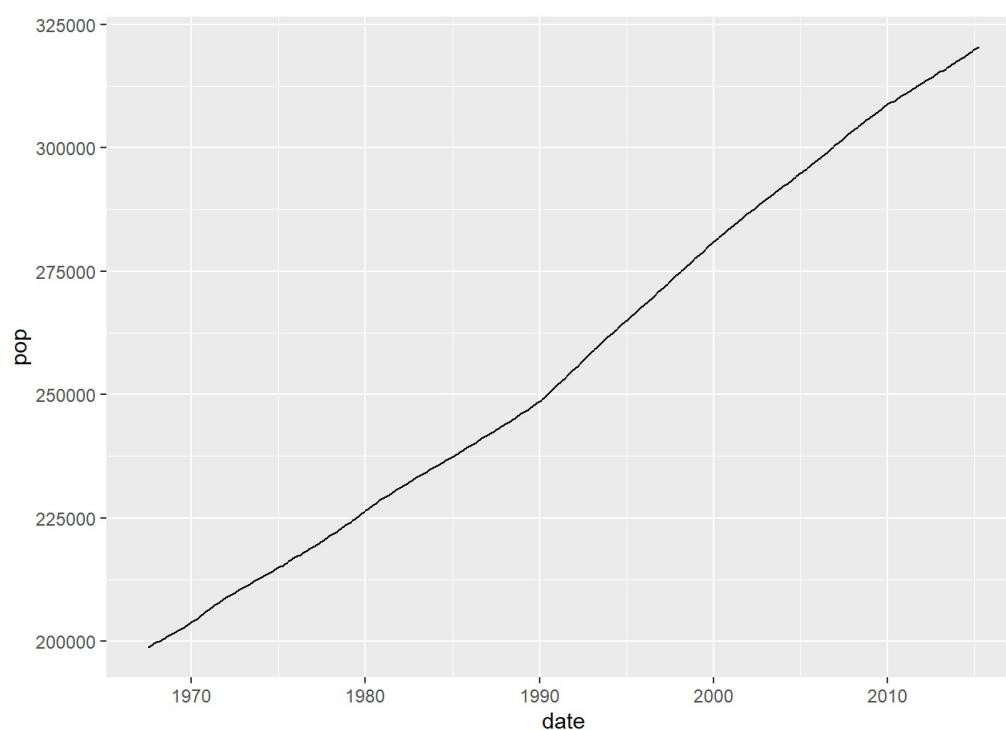


## 4. ZAMAN SERİSİ GÖRSELLEŞTİRME

```
head(economics)
```

```
## # A tibble: 6 × 6
##   date      pce    pop psavert uempmed unemploy
##   <date>  <dbl>  <dbl>    <dbl>    <dbl>    <dbl>
## 1 1967-07-01 507. 198712    12.6     4.5    2944
## 2 1967-08-01 510. 198911    12.6     4.7    2945
## 3 1967-09-01 516. 199113    11.9     4.6    2958
## 4 1967-10-01 512. 199311    12.9     4.9    3143
## 5 1967-11-01 517. 199498    12.8     4.7    3066
## 6 1967-12-01 525. 199657    11.8     4.8    3018
```

```
d <- economics
ggplot(d , aes(date , pop)) +
  geom_line()
```

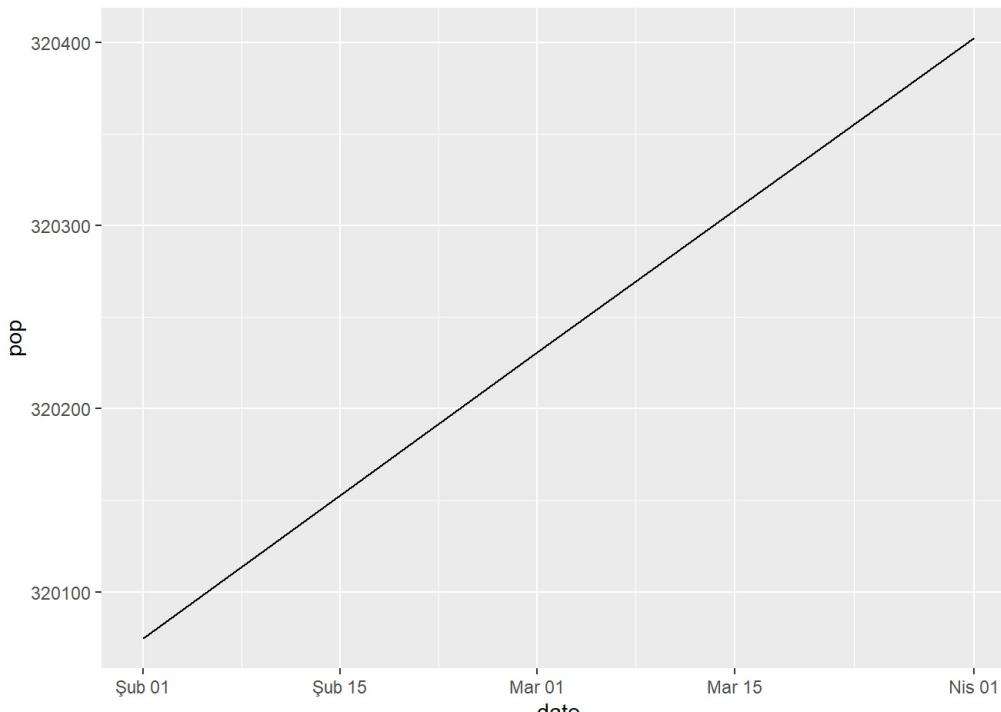


tarih değişkeninin tarih değişkeni olduğunu r a göstermemiz lazım. önce yıl ay gün şeklinde dir.

```
head(economics)
```

```
## # A tibble: 6 × 6
##   date      pce    pop psavert uempmed unemploy
##   <date>    <dbl>  <dbl>    <dbl>    <dbl>    <dbl>
## 1 1967-07-01 507. 198712    12.6     4.5    2944
## 2 1967-08-01 510. 198911    12.6     4.7    2945
## 3 1967-09-01 516. 199113    11.9     4.6    2958
## 4 1967-10-01 512. 199311    12.9     4.9    3143
## 5 1967-11-01 517. 199498    12.8     4.7    3066
## 6 1967-12-01 525. 199657    11.8     4.8    3018
```

```
d <- economics
s <- d %>% filter( date > as.Date("2015-1-25"))
ggplot(s, aes(date , pop)) +
  geom_line()
```

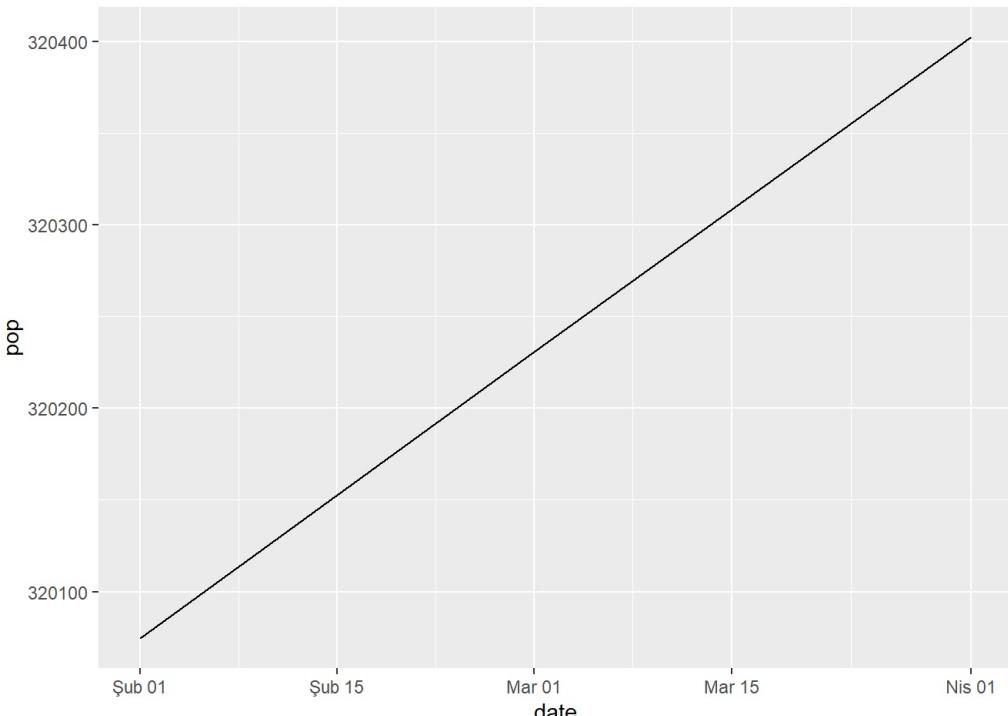


aynı veri setindeki değişikleri ggplot içinde de yapılabilir.

```
head(economics)
```

```
## # A tibble: 6 × 6
##   date      pce    pop psavert uempmed unemploy
##   <date>    <dbl>  <dbl>    <dbl>     <dbl>
## 1 1967-07-01 507. 198712    12.6     4.5     2944
## 2 1967-08-01 510. 198911    12.6     4.7     2945
## 3 1967-09-01 516. 199113    11.9     4.6     2958
## 4 1967-10-01 512. 199311    12.9     4.9     3143
## 5 1967-11-01 517. 199498    12.8     4.7     3066
## 6 1967-12-01 525. 199657    11.8     4.8     3018
```

```
d <- economics
s <- d %>% filter( date > as.Date("2015-1-25"))
ggplot(s, aes(date , pop)) +
  geom_line()
```



```
#şurda geom_line içinde bir şey yazıyordu bak
```

## ÇOKLU ZAMAN SERİSİ GÖRSELLEŞTİRME

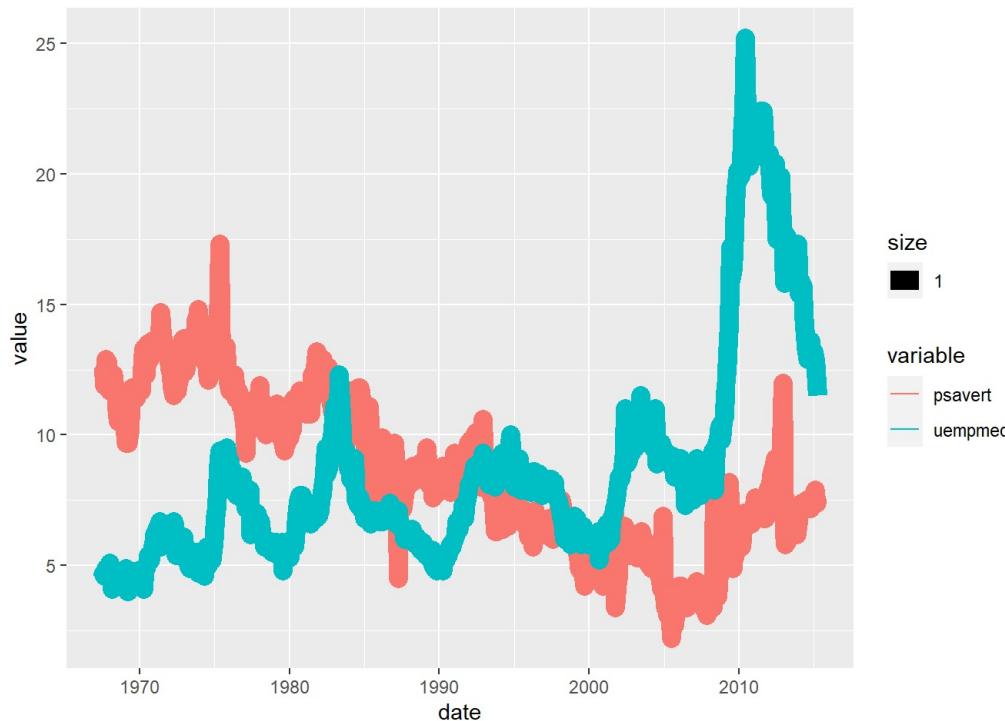
variable kategorik değişken iki tane değişkeni var. çizgi olarak gösterdik.

## ÇOKLU ZAMAN SERİSİNDE ÇİZGİ GRAFİĞİ

```
df <- economics %>%
  select(date, psavert, uempmed) %>%
  gather(key = "variable" , value = "value" , -date)
head(df, 3)
```

```
## # A tibble: 3 × 3
##   date      variable value
##   <date>    <chr>    <dbl>
## 1 1967-07-01 psavert    12.6
## 2 1967-08-01 psavert    12.6
## 3 1967-09-01 psavert    11.9
```

```
ggplot(df, aes(date, value)) +
  geom_line(aes(color = variable , size = 1))
```

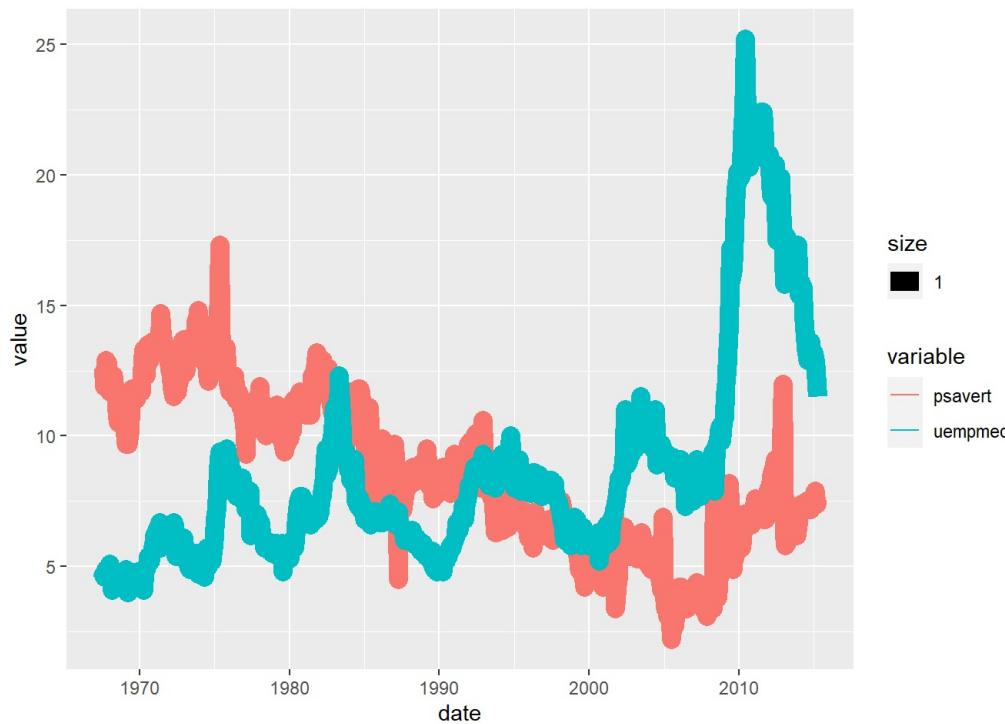


## ÇOKLU ZAMAN SERİSİNDE ALAN GRAFİĞİ

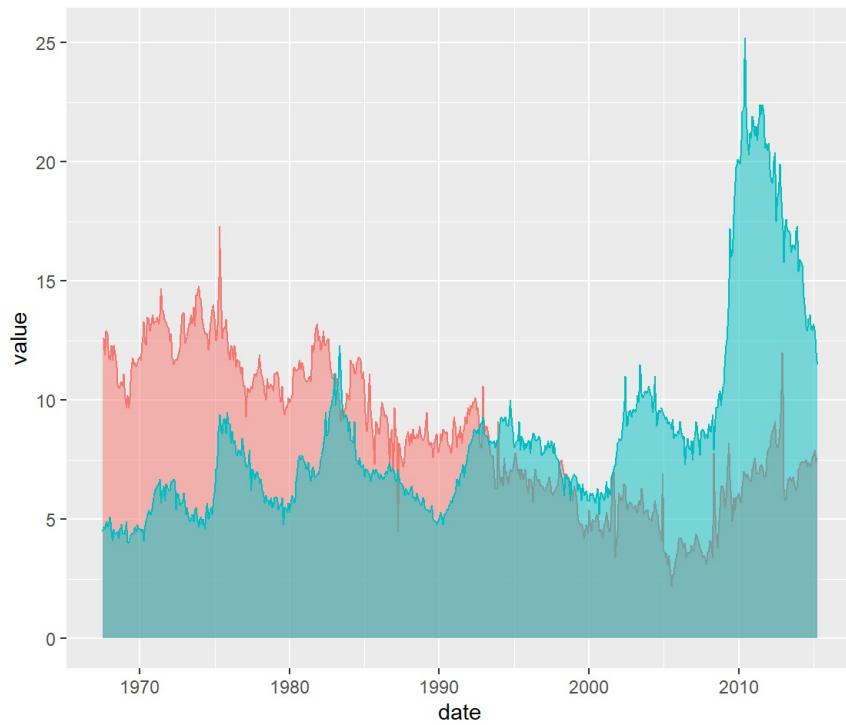
bitcoin ile ilgili zaman grafiği için uygun bir grafik olabilir.

```
library(tidyr)
df <- economics %>%
  select(date, psavert, uempmed) %>%
  gather(key = "variable" , value = "value" , -date)

ggplot(df, aes( date , value )) +
  geom_line(aes(color= variable , size = 1))
```



```
ggplot(df, aes(date,value)) +
  geom_area(aes(color = variable, fill=variable),
            alpha= 0.5, position = position_dodge(0.8)
           )
```



## 5. ÖRNEKLER

### MTCARS

Öncelikle mtcars'in içinde yer alandataların kısaltmaları ne anlama geliyor inceliyoruz.

([https://rstudio-pubs-static.s3.amazonaws.com/61800\\_faea93548c6b49cc91cd0c5ef5059894.html](https://rstudio-pubs-static.s3.amazonaws.com/61800_faea93548c6b49cc91cd0c5ef5059894.html)) ([https://rstudio-pubs-static.s3.amazonaws.com/61800\\_faea93548c6b49cc91cd0c5ef5059894.html](https://rstudio-pubs-static.s3.amazonaws.com/61800_faea93548c6b49cc91cd0c5ef5059894.html))

1	mpg	Miles/US Gallon	mpg is the determinant of fuel efficiency
2	cyl	Number of cylinders	Data includes vehicles with 4,6,8 cylinder engines.
3	disp	Displacement (cu.in.)	Displacement measures overall volume in the engine as a factor of cylinder circumference, depth and total number of cylinders. This metric gives a good proxy for the total amount of power the engine can generate.
4	hp	Gross horsepower	Gross horsepower measures the theoretical output of an engine's power output; notably, <i>gross</i> rating is of the engine in an isolated environment outside any specific vehicle. When installed in a car, exhaust systems, carburetor, alternator, power systems, etc all influence the power that actually gets to the drive train. Moreover, according to online sources, in the early 1970s, regulatory changes influenced how gross horsepower was measured. As this dataset is from the early-mid 1970s, it's unclear if hp metrics may be used as reliable comparators of engine power across models as it's uncertain how manufacturers are reporting.
5	drat	Rear axle ratio	The rear axle gear ratio indicates the number of turns of the drive shaft for every one rotation of the wheel axle. A vehicle with a high ratio would provide more torque and thus more towing capability, for example. Transmission configuration can often influence a manufacturer's gearing ratio.
6	wt	Weight (lb/1000)	The overall weight of the vehicle per 1000lbs (half US ton)
7	qsec	1/4 mile time	A performance measure, primarily of acceleration. Fastest time to travel 1/4 mile from standstill (in seconds).
8	vs	V/S	Binary variable signaling the engine cylinder configuration a V-shape (vs=0) or Straight Line (vs=1). V==0 and S==1. Configuration offers trade offs in power/torque, design usage in terms of space/size of engine and performance or center of gravity of vehicle. The geometry and placement of the engine, as influenced by its cylinder head, can have numerous knock-on influences on the vehicle beyond the technical engineering considerations of the cylinder angle.
9	am	Transmission Type	A binary variable signaling whether vehicle has automatic (am=0) or manual (am=1) transmission configuration.
10	gear	Number of forward gears	Number of gears in the transmission. Manual transmissions have either 4 or 5 forward gears; Automatic either 3 or 4
11	carb	Number of carburetors	The number of carburetor barrels. Engines with higher displacement typically have higher

Datamızı çağrıralım.

```
mtcars
```

```
##          mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant      18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360    14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D     24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230      22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280      19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C     17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE     16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL     17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC    15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial 14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128       32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic     30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla 33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona   21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger 15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin     15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28      13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird 19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9        27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2     26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa     30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L    15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino      19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora     15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E       21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

Bazı istatiksel özelliklerini inceleyelim.

```
str(mtcars)
```

```
## 'data.frame': 32 obs. of 11 variables:
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...
## $ disp: num 160 160 108 258 360 ...
## $ hp  : num 110 110 93 110 175 105 245 62 95 123 ...
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt  : num 2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num 16.5 17 18.6 19.4 17 ...
## $ vs  : num 0 0 1 1 0 1 0 1 1 1 ...
## $ am  : num 1 1 1 0 0 0 0 0 0 0 ...
## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...
## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...
```

```
summary(mtcars)
```

```

##      mpg          cyl         disp        hp
##  Min.   :10.40   Min.   :4.000   Min.   :71.1   Min.   :52.0
##  1st Qu.:15.43  1st Qu.:4.000  1st Qu.:120.8  1st Qu.:96.5
##  Median :19.20  Median :6.000  Median :196.3  Median :123.0
##  Mean   :20.09  Mean   :6.188  Mean   :230.7  Mean   :146.7
##  3rd Qu.:22.80  3rd Qu.:8.000  3rd Qu.:326.0  3rd Qu.:180.0
##  Max.   :33.90  Max.   :8.000  Max.   :472.0  Max.   :335.0
##      drat         wt         qsec        vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50  Min.   :0.0000
##  1st Qu.:3.080  1st Qu.:2.581  1st Qu.:16.89  1st Qu.:0.0000
##  Median :3.695  Median :3.325  Median :17.71  Median :0.0000
##  Mean   :3.597  Mean   :3.217  Mean   :17.85  Mean   :0.4375
##  3rd Qu.:3.920  3rd Qu.:3.610  3rd Qu.:18.90  3rd Qu.:1.0000
##  Max.   :4.930  Max.   :5.424  Max.   :22.90  Max.   :1.0000
##      am          gear        carb
##  Min.   :0.0000  Min.   :3.000  Min.   :1.000
##  1st Qu.:0.0000  1st Qu.:3.000  1st Qu.:2.000
##  Median :0.0000  Median :4.000  Median :2.000
##  Mean   :0.4062  Mean   :3.688  Mean   :2.812
##  3rd Qu.:1.0000  3rd Qu.:4.000  3rd Qu.:4.000
##  Max.   :1.0000  Max.   :5.000  Max.   :8.000

```

Mtcars'ın içinde yer alan verileri faktör haline getirelim.

```

mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)

```

faktör haline getirilmiş datamızdan herhangi sütununu çağırabiliyor muyuz diye test edelim.

```
summary(mtcars$mpg)
```

```

##      Min. 1st Qu. Median  Mean 3rd Qu. Max.
##  10.40  15.43  19.20  20.09  22.80  33.90

```

Datamızda kayıp veri var mı? incelemesi yapalım.

```
table(is.na(mtcars))
```

```

##
## FALSE
## 352

```

Mtcars datamız da yer alan sütunların isimlerini tekrar sorgulayalım. İsteğe bağlı sütun ekleme ya da çıkartma yapılabilir. Fakat ben bu örnekte gerek duymadım.

```
names(mtcars)
```

```

## [1] "mpg"   "cyl"   "disp"  "hp"    "drat"  "wt"    "qsec" "vs"    "am"    "gear"
## [11] "carb"

```

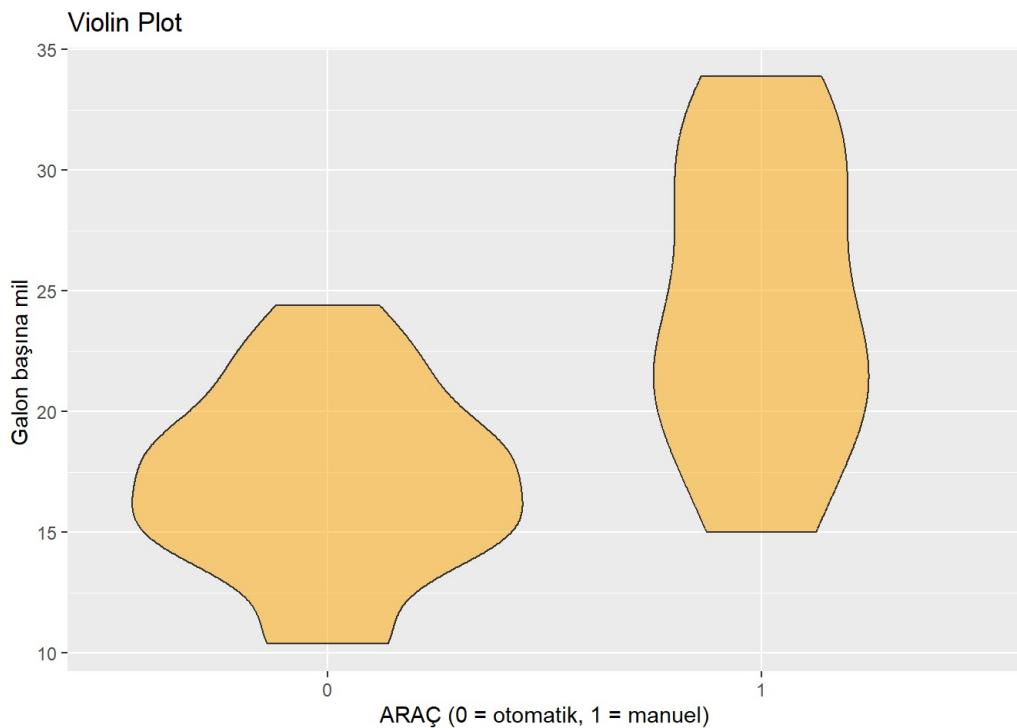
**soru : Acaba otomatik ve manuel araçların galon başına mil ne kadar?** Bu sorunumuzu violin plot yardımıyla inceleyelim.

```

df <- ggplot(data = mtcars, aes(x=am, y=mpg),
              draw_quantiles=TRUE)

df + geom_violin( fill = "orange" , alpha = 0.5) + labs(title="Violin Plot", x="ARAÇ (0 = otomatik, 1 = manuel)", y ="Galon başına mil ")

```



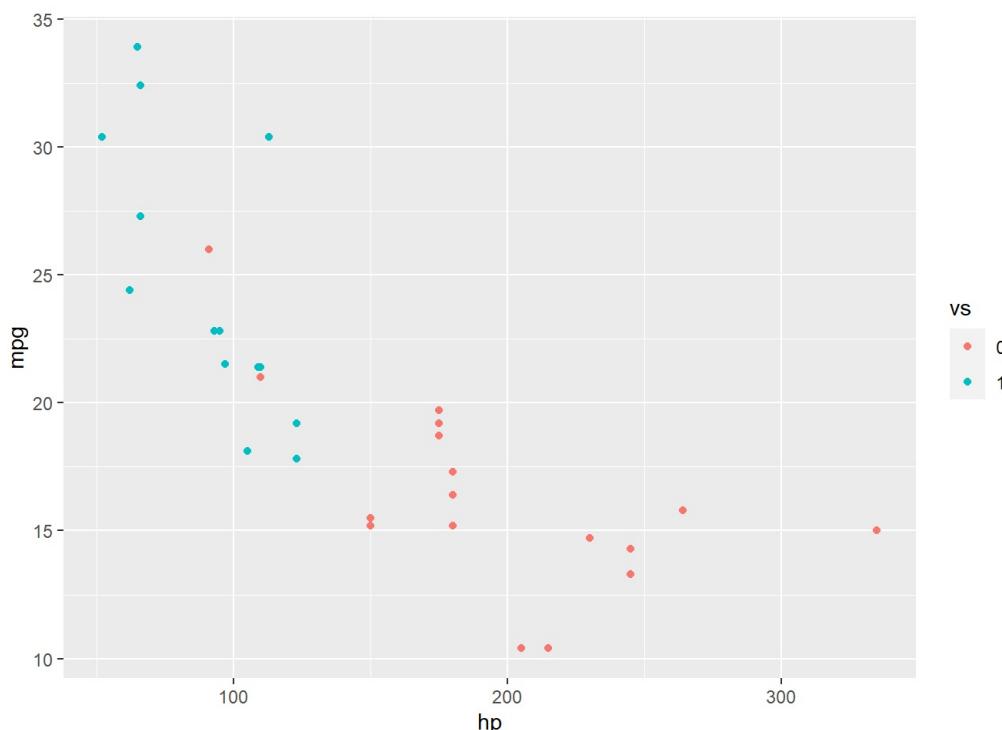
Grafik, otomatik vitesli araçlara göre manuel şanzımanlı araçlarda MPG için bir yükseliş sebep olmasını görüyoruz.

otomatik şanzımanlı araçlar için mpg'nin medyan çevresinde yoğun olduğu manuel için ilk yüzdelik (%25) dilim çevresinde yoğunluk olduğu görülmektedir.

#### soru2- motor ve beygir gücünün ilişkisi hakkında ne可以说呢?

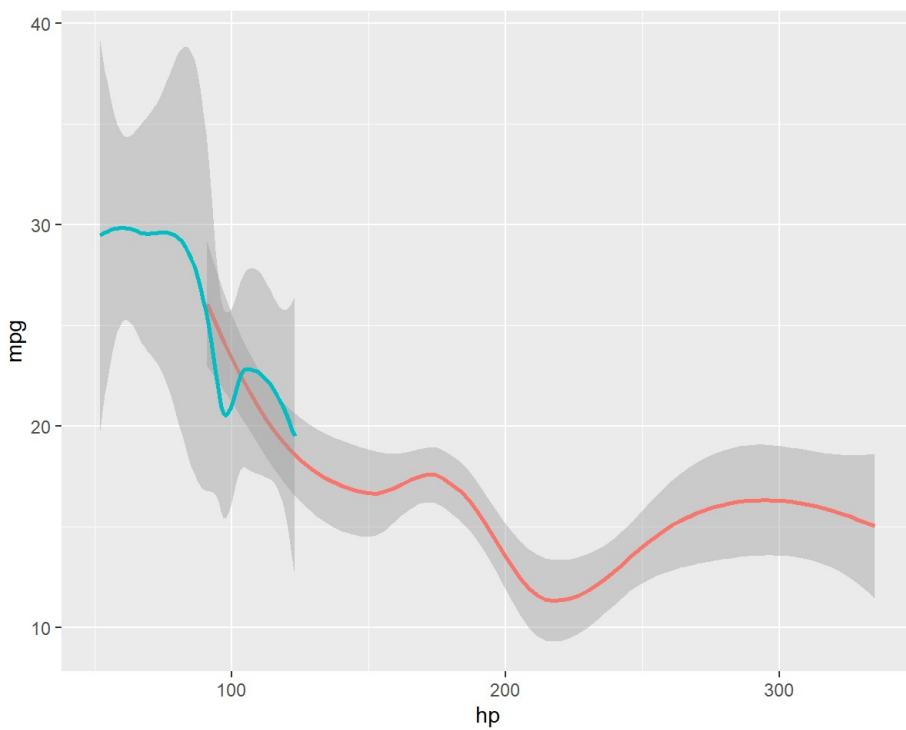
yakıt verimliliğin beygir gücüne etkisi;

```
ggplot(mtcars, aes(x = hp, y = mpg, color = vs)) +
  geom_point()
```



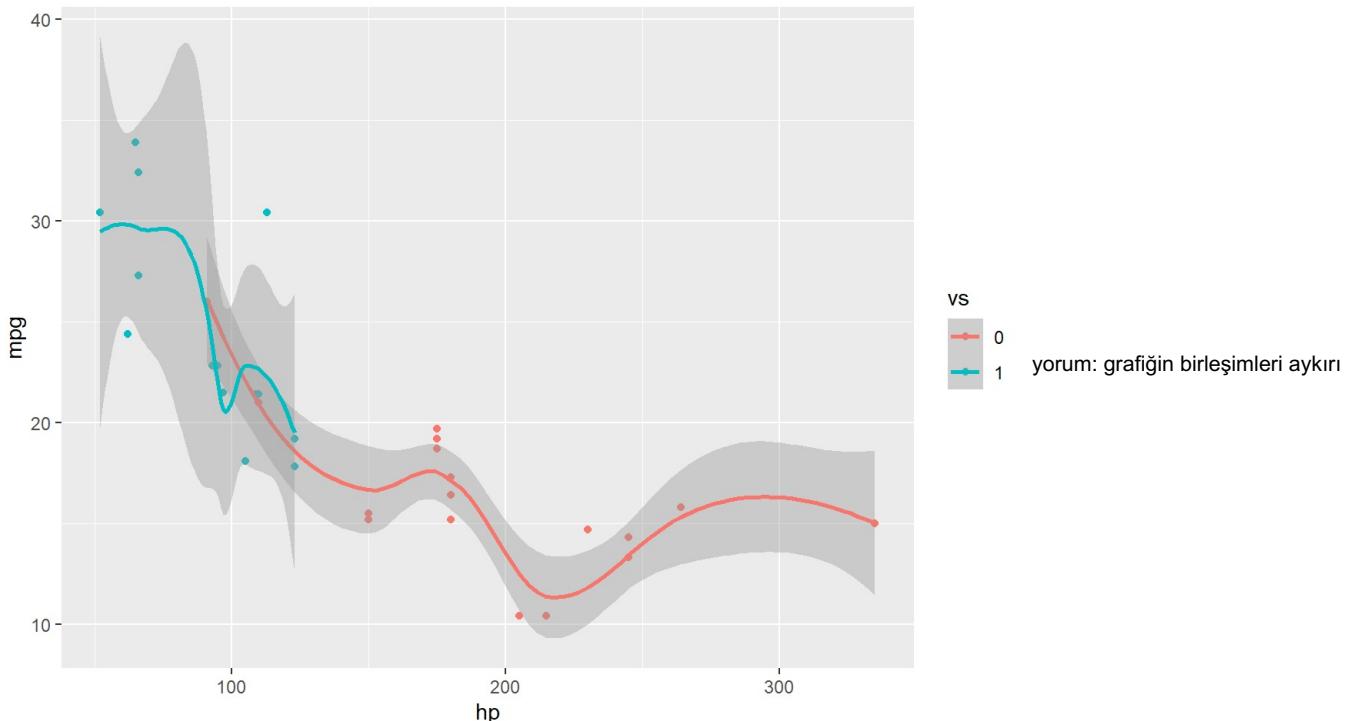
geom\_smooth() komutu verideki trendi çizgi ile özetlemeyi graftır.

```
ggplot(mtcars, aes(x = hp, y = mpg, color = vs)) +
  geom_smooth( method = "loess" , formula = 'y ~ x')
```



Şimdi bu iki grafiği birleştiriyoruz;

```
ggplot(mtcars, aes(x = hp, y = mpg, color = vs)) +
  geom_point() + geom_smooth(method = "loess", formula = 'y ~ x')
```



değerleri bulmamıza yardımcı oldu.

Staright motorları için, daha düşük beygir(hp) gücünün ortalama ve maksimum değer arasında daha yüksek mpg verdiği,

V-biçimli motorlarda ise, daha yüksek beygir gücüne(hp) rağmen mpg değerlerinin en küçük niceliksel aralığın altına düşüğünü aktarır.

## DIAMONDS

Datamızı çağrıyoruz.

```
diamonds
```

```

## # A tibble: 53,940 × 10
##   carat cut      color clarity depth table price     x     y     z
##   <dbl> <ord>    <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1 0.23 Ideal    E     SI2     61.5   55   326  3.95  3.98  2.43
## 2 0.21 Premium  E     SI1     59.8   61   326  3.89  3.84  2.31
## 3 0.23 Good    E     VS1     56.9   65   327  4.05  4.07  2.31
## 4 0.29 Premium I     VS2     62.4   58   334  4.2    4.23  2.63
## 5 0.31 Good    J     SI2     63.3   58   335  4.34  4.35  2.75
## 6 0.24 Very Good J     VVS2    62.8   57   336  3.94  3.96  2.48
## 7 0.24 Very Good I     VVS1    62.3   57   336  3.95  3.98  2.47
## 8 0.26 Very Good H     SI1     61.9   55   337  4.07  4.11  2.53
## 9 0.22 Fair    E     VS2     65.1   61   337  3.87  3.78  2.49
## 10 0.23 Very Good H    VS1     59.4   61   338  4     4.05  2.39
## # ... with 53,930 more rows

```

Verilerimizi incelemek için `summary()` komutunu çağırıyoruz.

```

d <- diamonds
summary(d)

```

```

## #> #> carat          cut       color      clarity      depth
## #> #> Min.   :0.2000  Fair      : 1610  D: 6775  SI1     :13065  Min.   :43.00
## #> #> 1st Qu.:0.4000  Good     : 4906  E: 9797  VS2     :12258  1st Qu.:61.00
## #> #> Median  :0.7000  Very Good:12082  F: 9542  SI2     : 9194  Median  :61.80
## #> #> Mean    :0.7979  Premium   :13791  G:11292  VS1     : 8171  Mean    :61.75
## #> #> 3rd Qu.:1.0400  Ideal     :21551  H: 8304  VVS2    : 5066  3rd Qu.:62.50
## #> #> Max.   :5.0100
## #> #>           I: 5422  VVS1    : 3655  Max.   :79.00
## #> #>           J: 2808  (Other) : 2531
## #> #> table          price      x           y
## #> #> Min.   :43.00  Min.   : 326  Min.   : 0.000  Min.   : 0.000
## #> #> 1st Qu.:56.00  1st Qu.: 950  1st Qu.: 4.710  1st Qu.: 4.720
## #> #> Median  :57.00  Median : 2401  Median : 5.700  Median : 5.710
## #> #> Mean    :57.46  Mean   : 3933  Mean   : 5.731  Mean   : 5.735
## #> #> 3rd Qu.:59.00  3rd Qu.: 5324  3rd Qu.: 6.540  3rd Qu.: 6.540
## #> #> Max.   :95.00  Max.   :18823  Max.   :10.740  Max.   :58.900
## #>
## #> #> z
## #> #> Min.   : 0.000
## #> #> 1st Qu.: 2.910
## #> #> Median  : 3.530
## #> #> Mean    : 3.539
## #> #> 3rd Qu.: 4.040
## #> #> Max.   :31.800
## #

```

verileri inceliyoruz. bakıyorum `carat` belki sürekli değişken olabilir.

`cut` değişkeni yani kaliteleri ile ilgili bir değişken; `fair` good, very good, premium ideal bu kategorik olabilir peki ordinal mı diye merak ediyoruz ve bunu araştırıyorum. `/str(d)/` ile yapılır.

```
str(d)
```

```

## #> #> #> tibble [53,940 × 10] (S3: tbl_df/tbl/data.frame)
## #> #> $ carat  : num [1:53940] 0.23 0.21 0.23 0.29 0.31 ...
## #> #> $ cut     : Ord.factor w/ 5 levels "Fair" <"Good" <...: 5 4 2 4 2 3 3 3 1 3 ...
## #> #> $ color   : Ord.factor w/ 7 levels "D" <"E" <"F" <"G" <...: 2 2 2 6 7 7 6 5 2 5 ...
## #> #> $ clarity : Ord.factor w/ 8 levels "I1" <"SI2" <"SI1" <...: 2 3 5 4 2 6 7 3 4 5 ...
## #> #> $ depth   : num [1:53940] 61.5 59.8 56.9 62.4 63.3 ...
## #> #> $ table   : num [1:53940] 55 61 65 58 58 57 55 61 61 ...
## #> #> $ price   : int [1:53940] 326 326 327 334 335 336 336 337 337 ...
## #> #> $ x       : num [1:53940] 3.95 3.89 4.05 4.2 4.34 3.94 3.95 4.07 3.87 4 ...
## #> #> $ y       : num [1:53940] 3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 ...
## #> #> $ z       : num [1:53940] 2.43 2.31 2.31 2.63 2.75 2.48 2.47 2.53 2.49 2.39 ...

```

`cut` olan değişkene bakıyorum ve evet diyorum, kategorik ve ordinal yani sıralı bir değişkendir.

`price` a karşılık bir `cut` (*kalite*) karşılaştırmak güzel olabilir. sınıf sayısı fazlalaşınca sıkıntısı yaratılabilir.

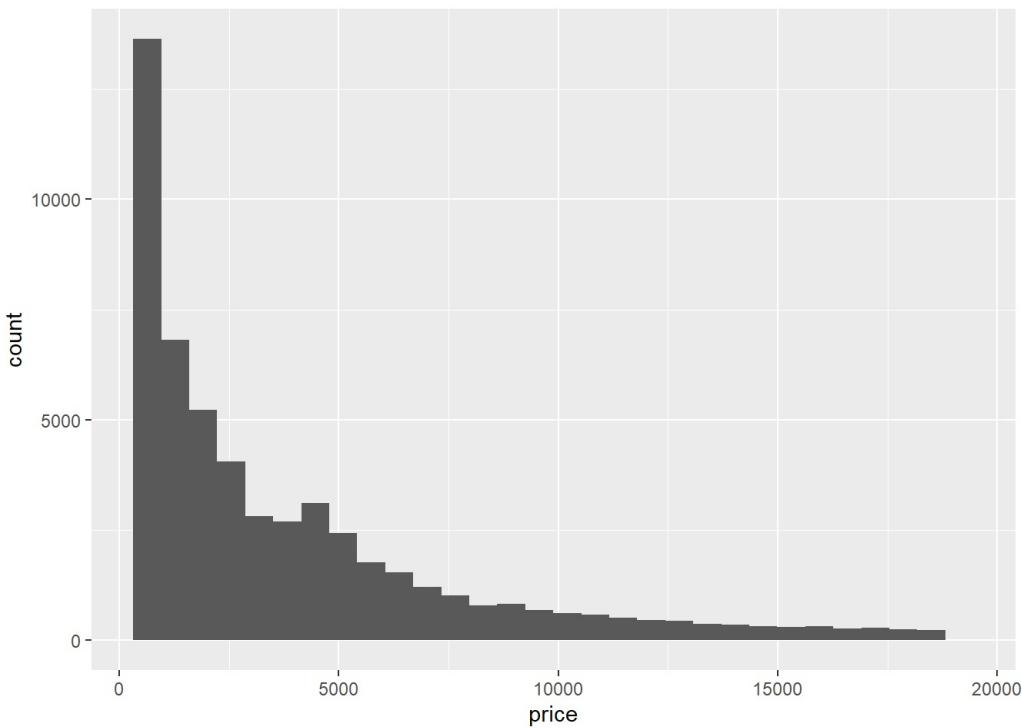
ilk olarak `price`'ı görselleştirelim.

**Soru: Ürünlerimizin fiyat performansı nedir?**

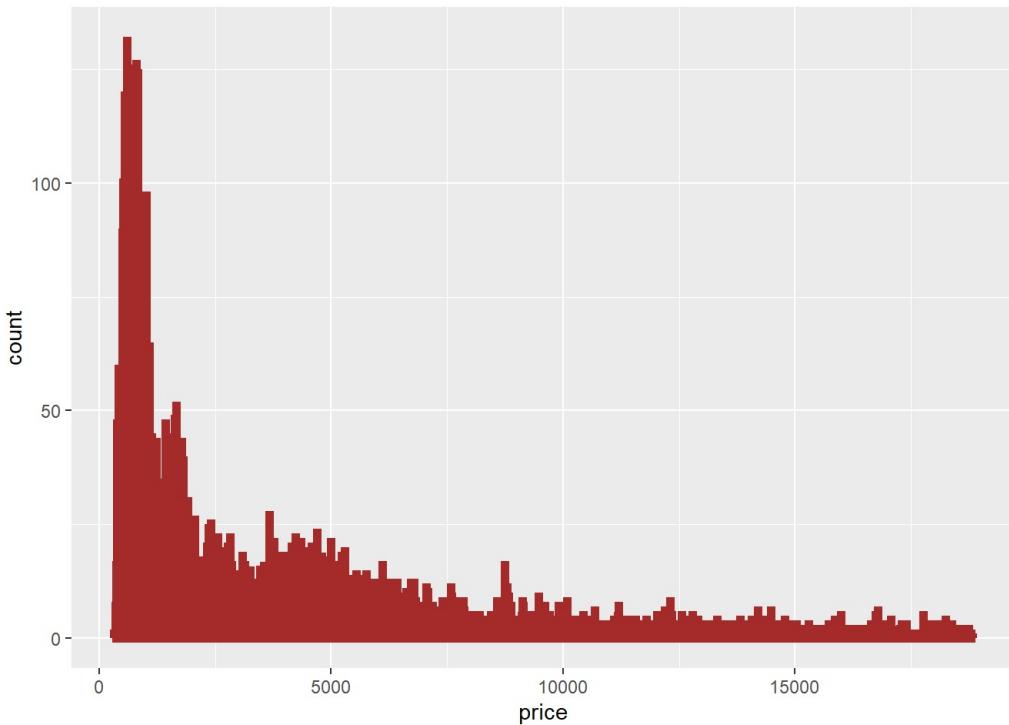
Histogram grafiği ile yorumlamaya çalışalım.

```
ggplot(d, aes(price)) +
  geom_histogram()
```

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



```
ggplot(d, aes(price)) +  
  geom_histogram(binwidth = 0.5, colour = "brown", size = 2 )
```



yorum : Satmış olduğumuz ürünlerin skalarası var. fiyat dağılımı bu şekilde. ürünlerimin büyük bir çoğunluğu 5000 tl'den daha az ürünleri kullanıyorlar.

Tüm veriyi temsil ediyor diye değerlendirmek yanlış olur. Bazı farklı yapılar olabilir. peki bu farklı yapıları nasıl analiz ederiz?

price'ı hangi kategorik değişkenler etkileyebilir? cut(kalite) olabilir. belki color (renk) olabilir.

price'ın bazı kategorik değişkenlerdeki kırılımını meerak ediyoruz. Görselleştirmeden önce ölçmek istiyorum. ( d%>% )

ilk olarak cut'ı sınıflandırdık. sonra cut'a göre sınıflandırılmış değişkenlerin price ortalama fiyatını verecek.

```
d %>% group_by(cut) %>%  
  summarise(n=n() , ort= mean(price))
```

```
## # A tibble: 5 × 3
##   cut      n    ort
##   <ord>  <int> <dbl>
## 1 Fair     1610  4359.
## 2 Good     4906  3929.
## 3 Very Good 12082  3982.
## 4 Premium   13791  4584.
## 5 Ideal    21551  3458.
```

şimdi ortalamalar gözle görülür bir fark yok birbirlerine çok yakın çıktı.Hani sanki bu kategorik çok etkilemiyor gibi geldi.

```
d %>% group_by(color) %>%
  summarise(n=n() , ort= mean(price))
```

```
## # A tibble: 7 × 3
##   color      n    ort
##   <ord>  <int> <dbl>
## 1 D       6775  3170.
## 2 E       9797  3077.
## 3 F       9542  3725.
## 4 G      11292  3999.
## 5 H      8304  4487.
## 6 I      5422  5092.
## 7 J      2808  5324.
```

Sınıflar arası geçişler yumuşak ama çok sınıf olduğu için D ile J arasında geçiş farkı büyük.Histogramın farklılığı için bu farkı aldım verdim kabul ettim.

```
d %>% group_by(clarity) %>%
  summarise(n=n() , ort= mean(price))
```

```
## # A tibble: 8 × 3
##   clarity      n    ort
##   <ord>  <int> <dbl>
## 1 I1       741  3924.
## 2 SI2      9194  5063.
## 3 SI1     13065  3996.
## 4 VS2     12258  3925.
## 5 VS1      8171  3839.
## 6 VVS2     5066  3284.
## 7 VVS1     3655  2523.
## 8 IF       1790  2865.
```

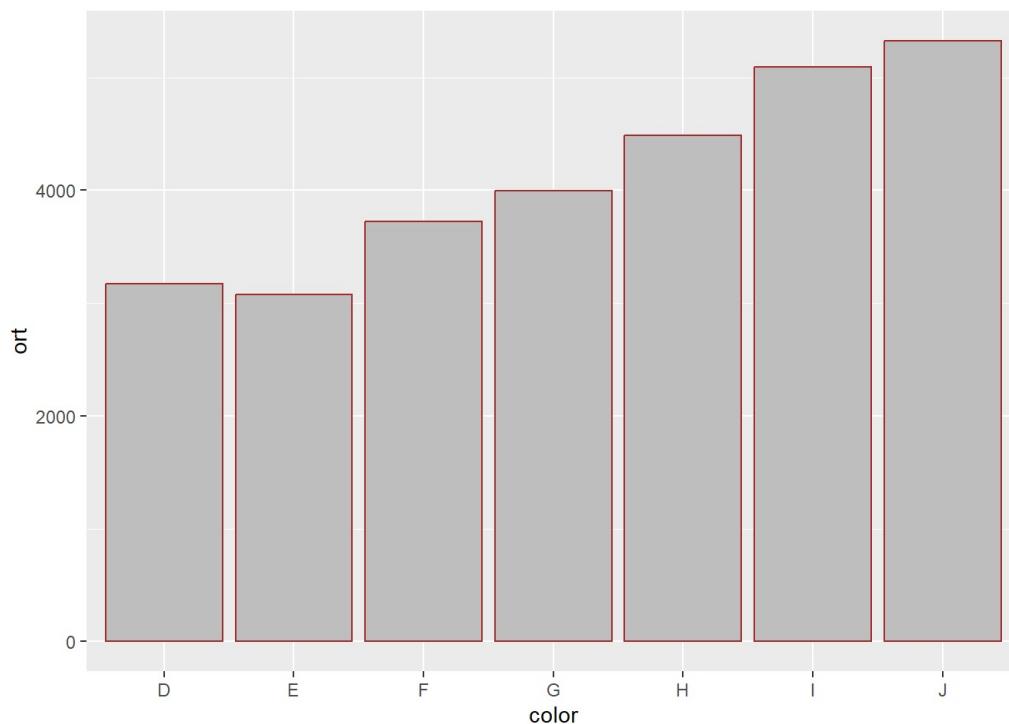
yne çok fark yok 8 sınıf'a nazaran. Colorın daha çok etilediğini düşünüyorum.

şimdi color bir kesikli değişken mi? Evet. ya bu kesikli değişkenin frekansı elimizde olacak ya da bu değere karşılık gelen mutluk değer net değer elimizde olmalı.

O zaman nasıl görselleştirebilirim? sütun, çizgi, nokta, pasta kullanabilirim.

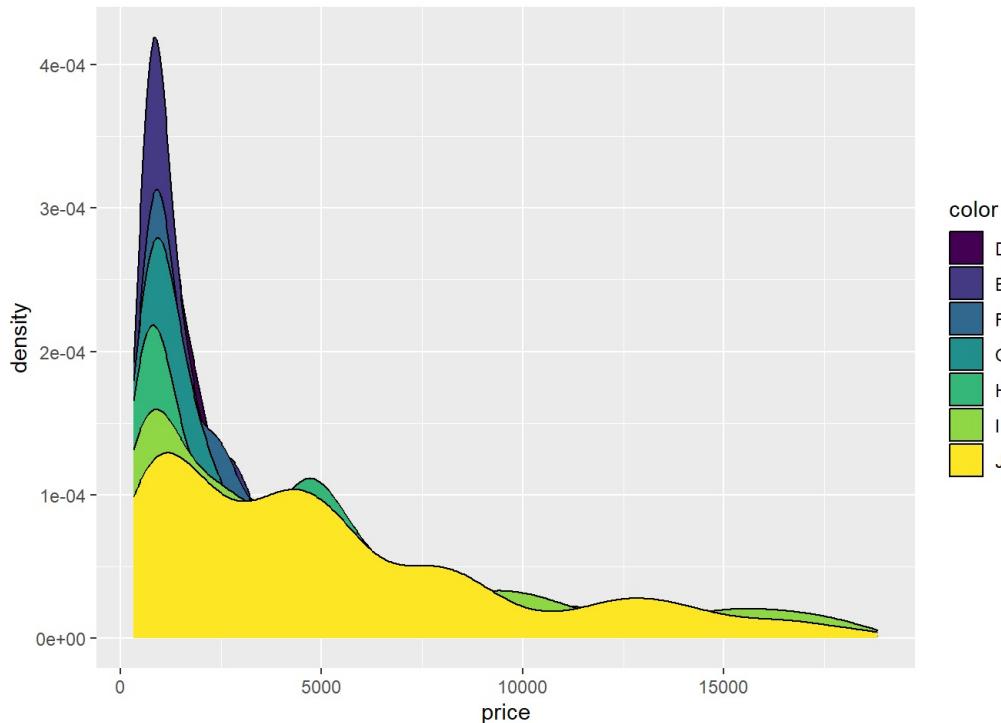
```
k<- d %>% group_by(color) %>%
  summarise(n=n() , ort= mean(price))

ggplot(k, aes(color, ort)) +
  geom_bar(stat = "identity", color ="brown", fill = "gray")
```



yoğunluk grafiği

```
ggplot(d, aes(price, fill =color)) +
  geom_density()
```



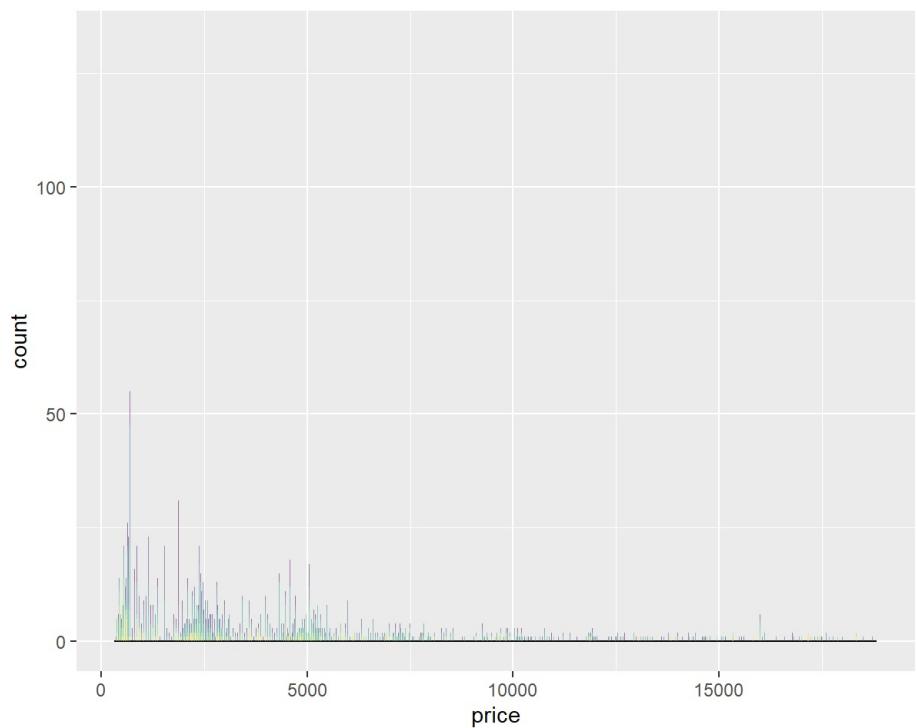
o ilk tablo ile karşılaştırdık. (brown olan) aslında histograma gitmeye bile gerek yok çünkü istediğimiz eristik.

İlk grafikte dedikki tüm veriyi temsil eder mi? priceların yoğunluğunu tüm verinin yoğunluğu olarak mı ifade etmeliyim? tamam yoğunlum bu ama bu yoğunluğu oluşturan etkenler ne gibi bir araştırmaya gitmiştık.

youmlamaya gittiğimizde D str(d) yaptığımızda en yüksek değerdi. ama bakıldığımda e onu geçmiş yoğunluğu, d yi baskılampiş.

```
ggplot(d, aes(price, fill =color)) +
  geom_density() +
  geom_histogram(binwidth = 0.5, alpha= 0.5, possiton="identity")
```

```
## Warning in geom_histogram(binwidth = 0.5, alpha = 0.5, possiton = "identity"):
## Ignoring unknown parameters: `possiton`
```



## 6. KAYNAKÇA

Ders notlarını oluşturanken yardım aldığım eğitim platformları aşağıdaki gibidir.

### UDEMY

Yetişkinlere ve öğrencilere yönelik, alanında profesyonel gönüllü eğitmenler tarafından oluşturulan kursları içeren eğitim teknolojisi, kitlesel çevrimiçi açık ders ve bir çevrimiçi öğrenme platformudur. [<https://tr.wikipedia.org/wiki/Udemy> (<https://tr.wikipedia.org/wiki/Udemy>)]

[udemy \(<https://www.udemy.com/course/r-programlama/>\)](https://www.udemy.com/course/r-programlama/)



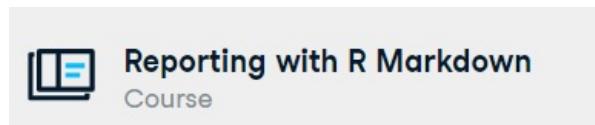
### Modern R Programlama Eğitimi

Vahit Keskin, Veri Bilimi Okulu

### DataCamp

Datacamp, veri bilimleri ile ilgilenen ve bu konuda kendini geliştirmek isteyen herkes için kolaylık sağlayan bir eğitim sitesidir. [<https://www.abprojeyonetimi.com/datacamp-nedir-ne-ise-yarar/> (<https://www.abprojeyonetimi.com/datacamp-nedir-ne-ise-yarar/>)]

[datacamp \(<https://app.datacamp.com/learn/courses/reporting-with-rmarkdown>\)](https://app.datacamp.com/learn/courses/reporting-with-rmarkdown)



### Mimar Sinan Güzel Sanatlar Üniversitesi

İleri düzey R programlama dersi esnasında aklımda kalan bilgilerin ders notumu hazırlamamda etkisi büyktür.

# Mimar Sinan Güzel Sanatlar Üniversitesi

tarafından Arthipo | 20 Şubat 2022 | Sanat Tarihi

