# Contents

```
weatherAUS <- read.csv(file = "C:/Users/halil ibrahim kaya/Documents/GitHub/R-Project/weaterAUS/dataset/
```

```
# dplyr paketini aktif hale getirelim
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

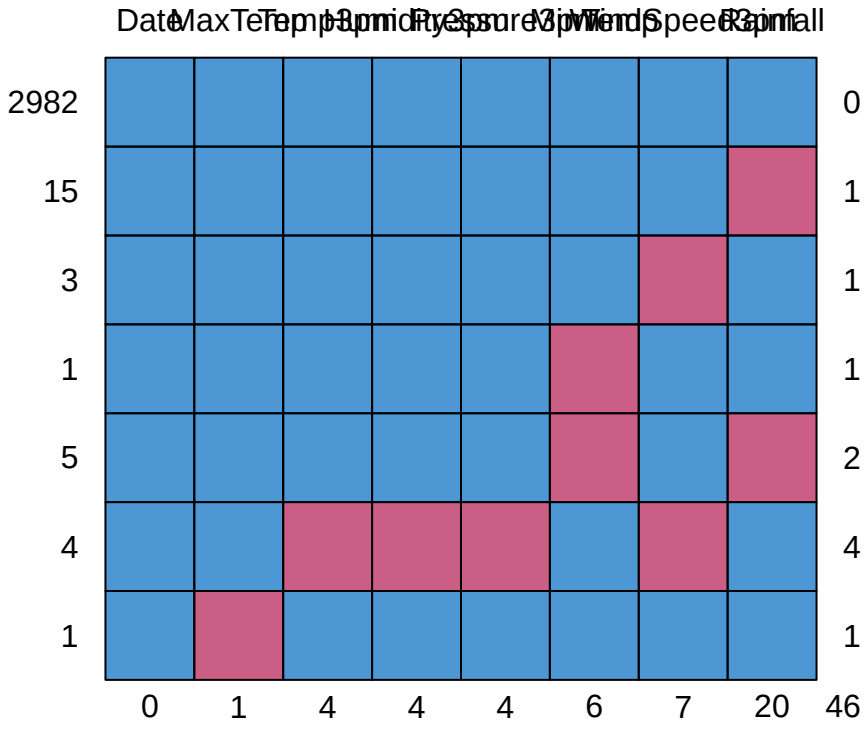
```
# Sadece tek bir konum icin analiz yapalim
AlburyWeather <- weatherAUS %>% filter(Location=="Albury")
# Analiz ve tahmin icin gerekli verilerimizi seciyoruz
AlburyWeather <- AlburyWeather %>% select(Temp3pm,MinTemp,MaxTemp,Rainfall,WindSpeed3pm,Humidity3pm,Pres
head(AlburyWeather)
```

```
##   Temp3pm MinTemp MaxTemp Rainfall WindSpeed3pm Humidity3pm Pressure3pm
## 1    21.8    13.4    22.9      0.6           24          22      1007.1
## 2    24.3     7.4    25.1      0.0           22          25      1007.8
## 3    23.2    12.9    25.7      0.0           26          30      1008.7
## 4    26.5     9.2    28.0      0.0            9          16      1012.8
## 5    29.7    17.5    32.3      1.0           20          33      1006.0
## 6    28.9    14.6    29.7      0.2           24          23      1005.4
##         Date
## 1 2008-12-01
## 2 2008-12-02
## 3 2008-12-03
## 4 2008-12-04
## 5 2008-12-05
## 6 2008-12-06
```

```
# Eksik veri kontrolu yapalim
library(mice)
```

```
##
## Attaching package: 'mice'

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

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

```
md.pattern(AlburyWeather)
```




```
##      Date MaxTemp Temp3pm Humidity3pm Pressure3pm MinTemp WindSpeed3pm Rainfall
## 2982    1       1       1           1           1       1            1        1
## 15      1       1       1           1           1       1            1        0
## 3       1       1       1           1           1       1            0        1
## 1       1       1       1           1           1       0            1        1
## 5       1       1       1           1           1       0            1        0
## 4       1       1       0           0           0       1            0        1
## 1       1       0       1           1           1       1            1        1
##         0       1       4           4           4       6            7       20
##
## 2982  0
## 15    1
## 3     1
## 1     1
## 5     2
## 4     4
## 1     1
##      46
```

```
# Toplam 46 degerim NA geldi. 3011 veri icerisinde az oldugundan dolayi bunlari atabiliriz
```

```r
head(which(is.na(AlburyWeather)))
```

```
## [1] 1045 1046 1047 1048 3294 3451
```

```r
# NA olarak atanmis verilerimin Index numaralarina goz attim
# Iki sekilde bu gozlemlerden kurtulabiliriz


#Birinci Yontem
index <- which(is.na(AlburyWeather))
head(AlburyWeather[-index,])
```

```
##   Temp3pm MinTemp MaxTemp Rainfall WindSpeed3pm Humidity3pm Pressure3pm
## 1    21.8    13.4    22.9      0.6           24          22      1007.1
## 2    24.3     7.4    25.1      0.0           22          25      1007.8
## 3    23.2    12.9    25.7      0.0           26          30      1008.7
## 4    26.5     9.2    28.0      0.0            9          16      1012.8
## 5    29.7    17.5    32.3      1.0           20          33      1006.0
## 6    28.9    14.6    29.7      0.2           24          23      1005.4
##         Date
## 1 2008-12-01
## 2 2008-12-02
## 3 2008-12-03
## 4 2008-12-04
## 5 2008-12-05
## 6 2008-12-06
```

```r
# Ikinci Yontem
head(na.omit(AlburyWeather))
```

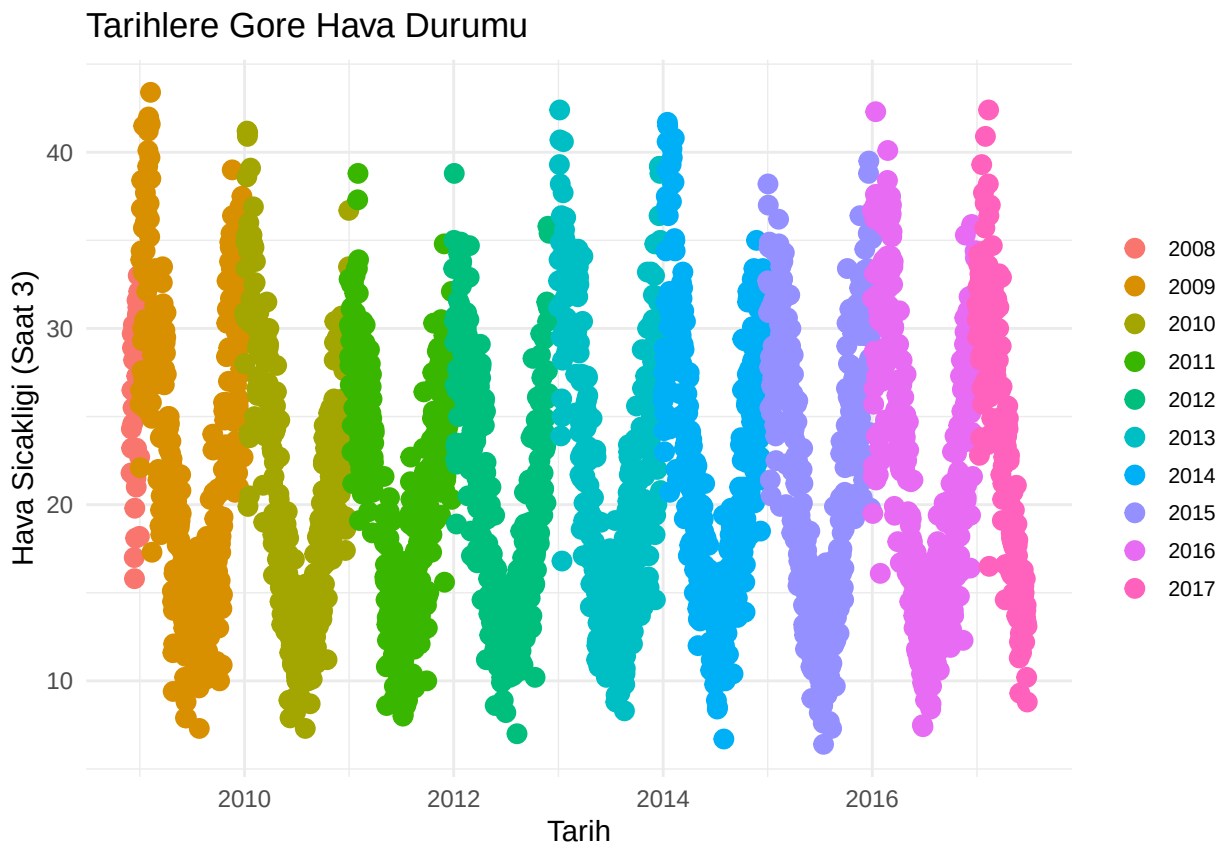
```
##   Temp3pm MinTemp MaxTemp Rainfall WindSpeed3pm Humidity3pm Pressure3pm
## 1    21.8    13.4    22.9      0.6           24          22      1007.1
## 2    24.3     7.4    25.1      0.0           22          25      1007.8
## 3    23.2    12.9    25.7      0.0           26          30      1008.7
## 4    26.5     9.2    28.0      0.0            9          16      1012.8
## 5    29.7    17.5    32.3      1.0           20          33      1006.0
## 6    28.9    14.6    29.7      0.2           24          23      1005.4
##         Date
## 1 2008-12-01
## 2 2008-12-02
## 3 2008-12-03
## 4 2008-12-04
## 5 2008-12-05
## 6 2008-12-06
```

```r
# Bu yontemi kullanarak NA iceren tum satiri atiyoruz. Bu yontemi kullanalim
AlburyWeather <- na.omit(AlburyWeather)
head(AlburyWeather)
```

```
##   Temp3pm MinTemp MaxTemp Rainfall WindSpeed3pm Humidity3pm Pressure3pm
## 1    21.8    13.4    22.9      0.6           24          22      1007.1
```

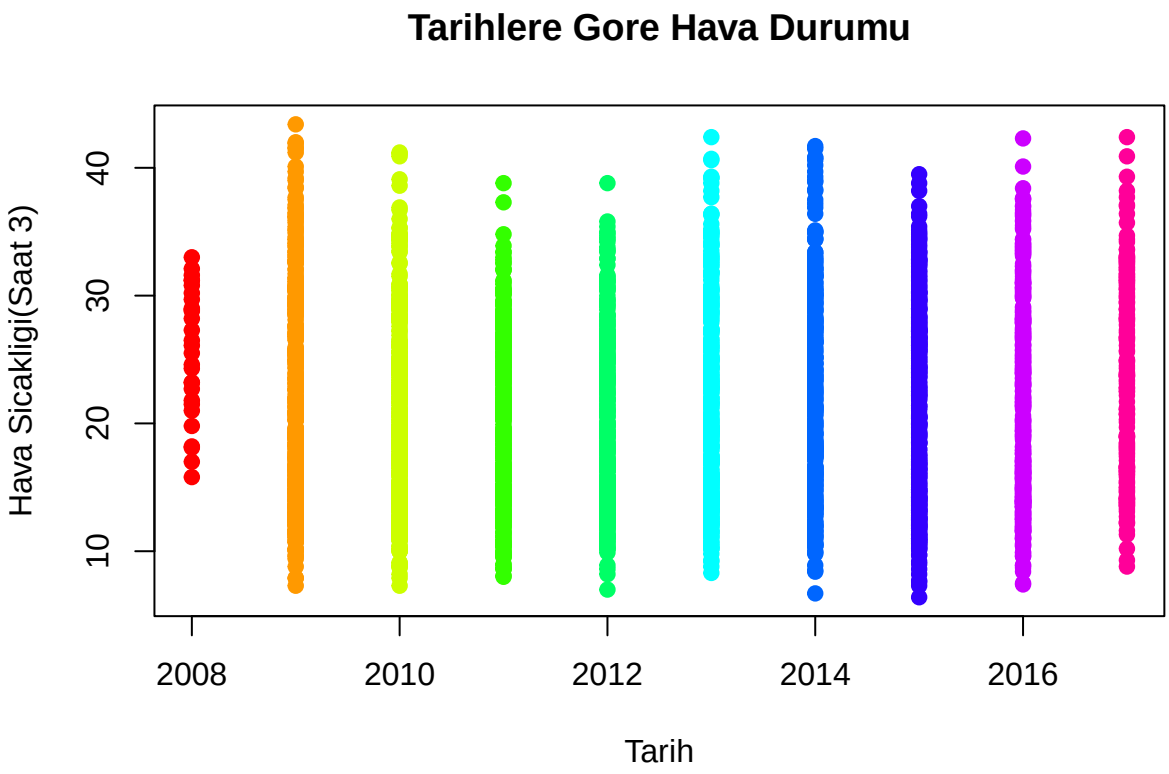
```
## 2    24.3     7.4    25.1      0.0           22          25      1007.8
## 3    23.2    12.9    25.7      0.0           26          30      1008.7
## 4    26.5     9.2    28.0      0.0            9          16      1012.8
## 5    29.7    17.5    32.3      1.0           20          33      1006.0
## 6    28.9    14.6    29.7      0.2           24          23      1005.4
##         Date
## 1 2008-12-01
## 2 2008-12-02
## 3 2008-12-03
## 4 2008-12-04
## 5 2008-12-05
## 6 2008-12-06
```

```r
AlburyWeather$Year <- format(as.Date(AlburyWeather$Date, format="%Y-%m-%d"), "%Y")
# Veri setimizdeki tarih bolumune ek olarak Yil bolumu olusturduk
years <- unique(AlburyWeather$Year)
# Her bir yili tekil olarak ayarladik
colors <- rainbow(length(years))
# Her tekil yila renk atamasi yaptik
AlburyWeather$Color <- colors[as.factor(AlburyWeather$Year)]
# faktorlere gore renk atamasi yaptik
AlburyWeather$Date <- as.Date(AlburyWeather$Date, format="%Y-%m-%d")
# tarih degiskenimi date'e donusturme islemini gerceklestirdim

havadurumu <- ggplot(AlburyWeather, aes(x = Date, y = Temp3pm, color = Year)) +
  geom_point(size = 3) +
  labs(
    title = "Tarihlere Gore Hava Durumu",
    x = "Tarih",
    y = "Hava Sicakligi (Saat 3)"
  ) + theme_minimal() +
  theme(
    legend.title = element_blank(),
    legend.text = element_text(size = 8),
    legend.key.size = unit(0.5, "cm")
  )
print(havadurumu)
```




```
# Grafigimizi olusturduk.
# yillara gore renklendirilmis sekilde sicaklik degerlerimizi bu sekilde inceliyoruz
```

```
plot(y=AlburyWeather$Temp3pm,x=AlburyWeather$Year, main =  "Tarihlere Gore Hava Durumu",
      ylab = "Hava Sicakligi(Saat 3)",xlab = "Tarih",col = AlburyWeather$Color, pch = 19)
```

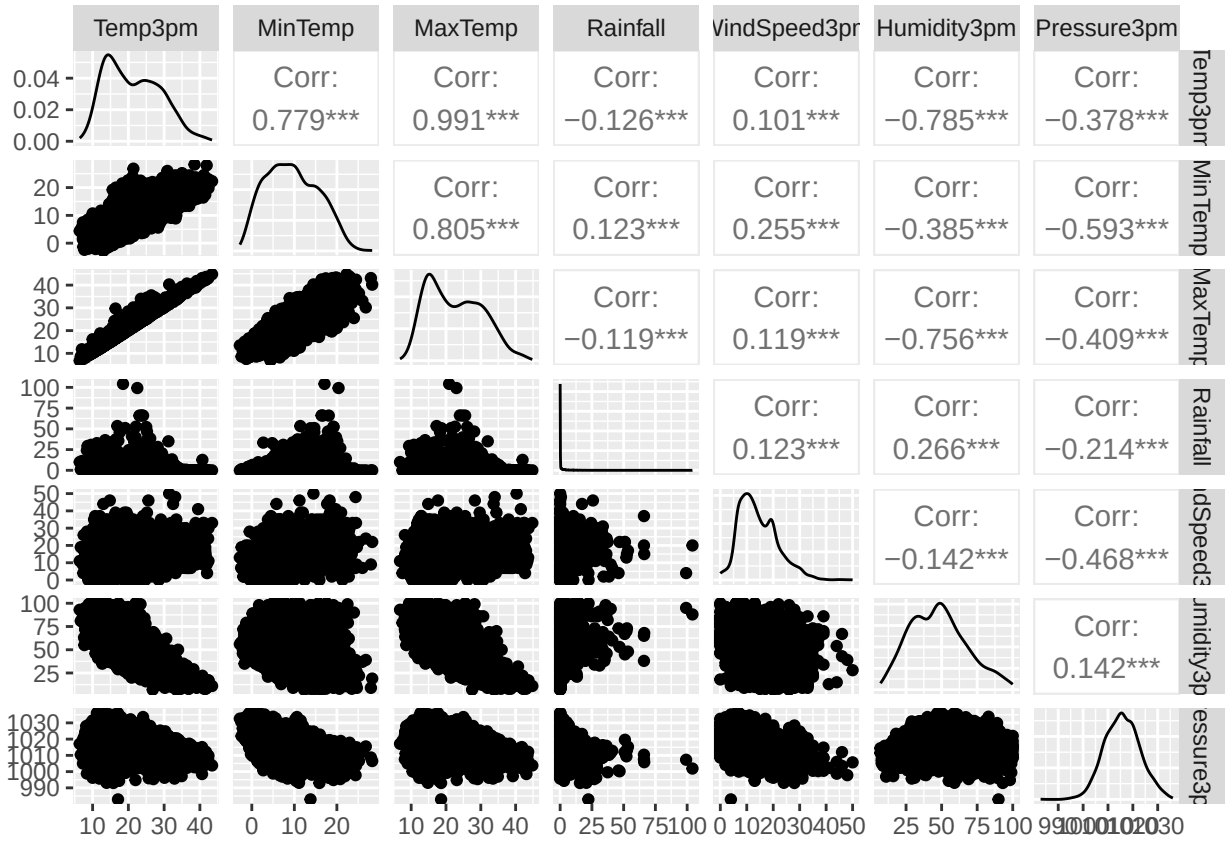



```
# verilerimizin yillara gore sicaklik aralşklarini incelemek icin Year degiskenimizle bir plot cizdirdi
```

```
library(GGally)
```

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

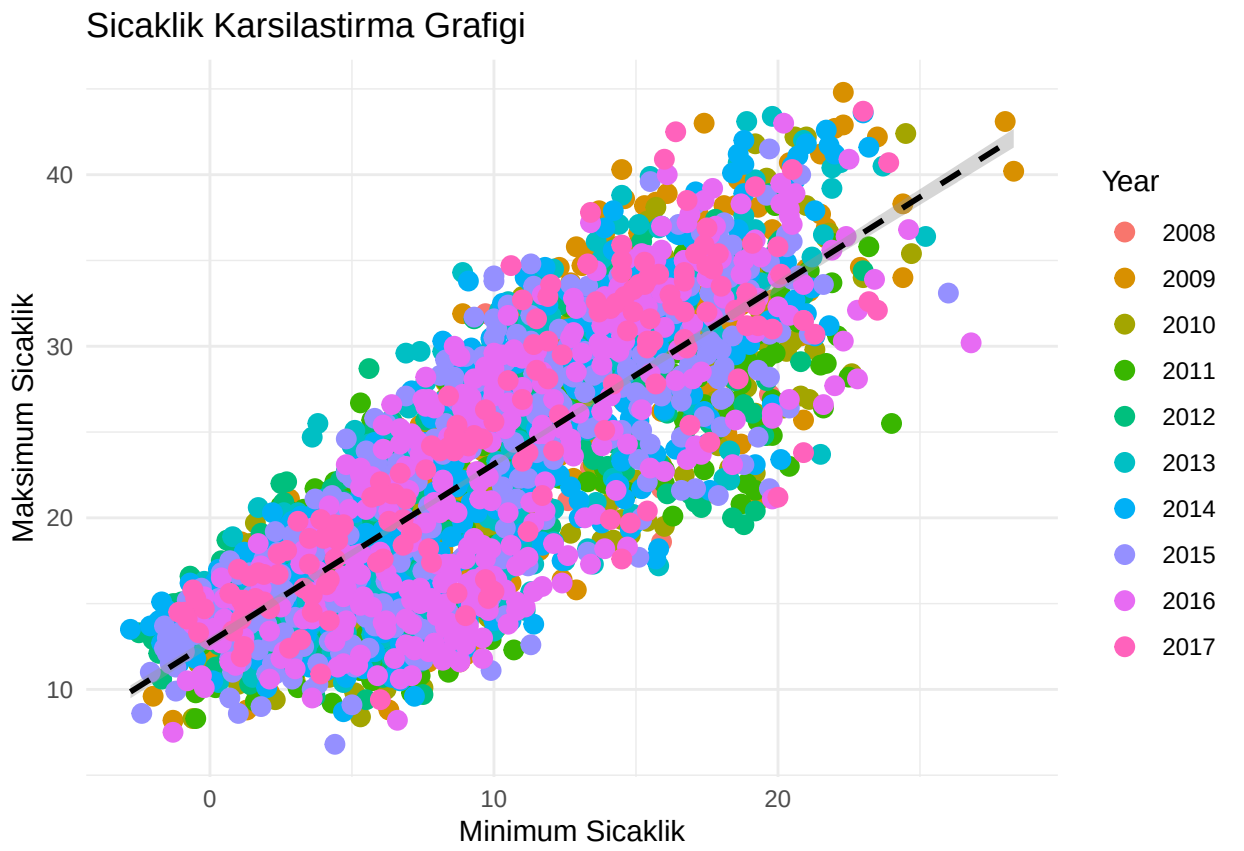
```
# Albury sehrine ait verileri aciklayici olarak gorsellestiren bir fonksiyonumuz:
ggpairs(AlburyWeather,columns = 1:7)
```




```
# Corr degerlerimiz aralarindaki baglanti boyutunu ve yonunu belirliyor.
```

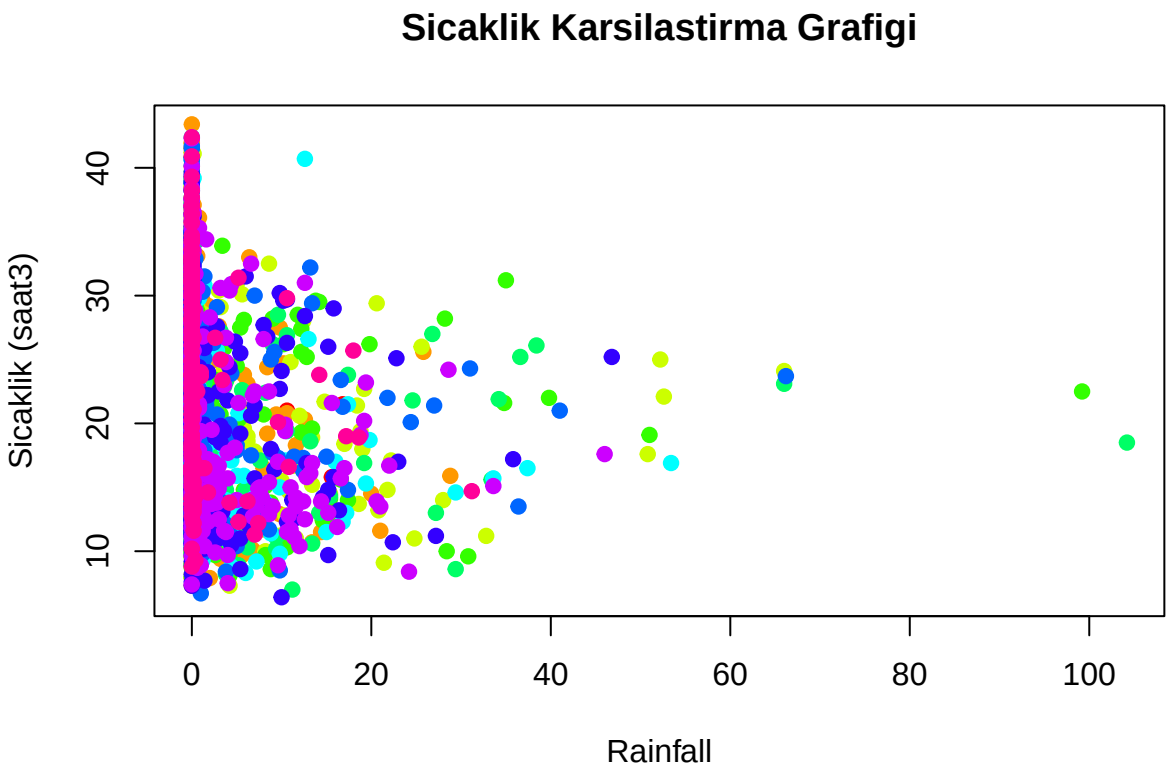
```
MinMaxTemp <- ggplot(AlburyWeather, aes(x = MinTemp, y = MaxTemp, color = Year)) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", color = "black", linetype = "dashed") +
  labs(
    x = "Minimum Sicaklik",
    y = "Maksimum Sicaklik",
    title = "Sicaklik Karsilastirma Grafigi"
  ) +
  theme_minimal()
print(MinMaxTemp)
```

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



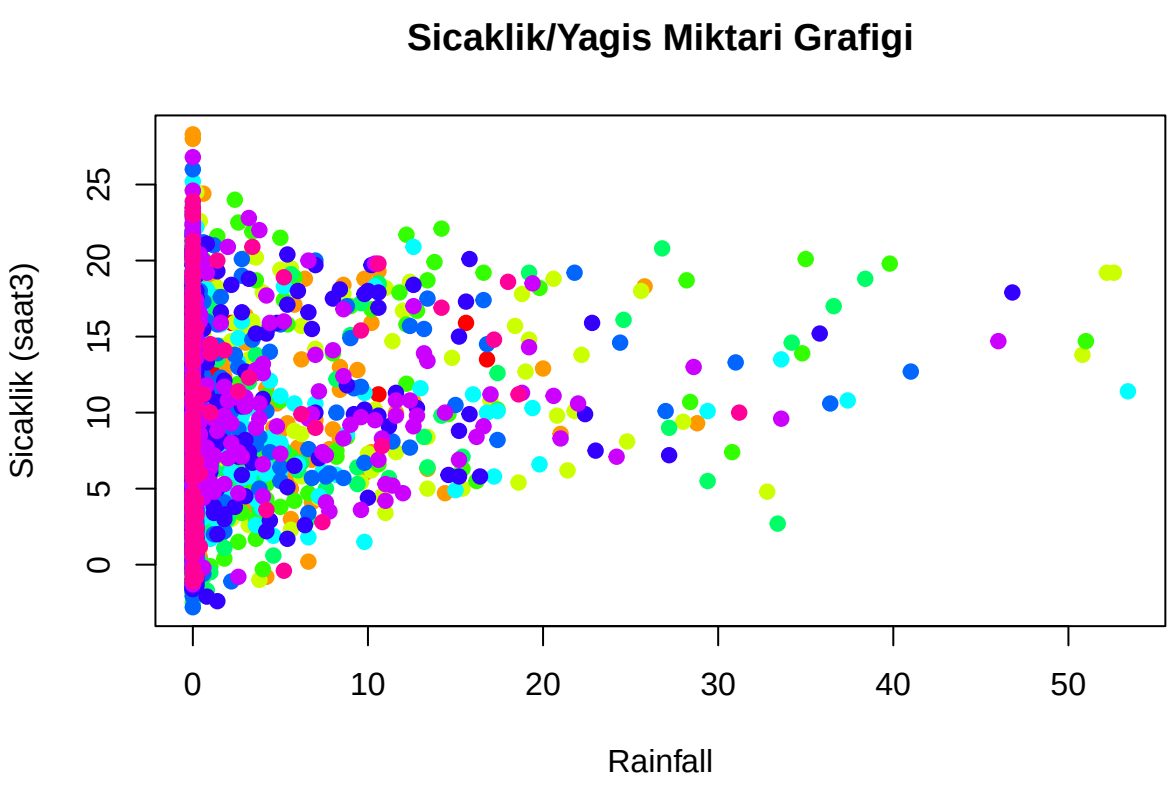

```
# Minimum ve Maksimum degerlerinin birbiriyle iliskili sekilde nasil degistigini gosteren grafigimiz
# lineer bir goruntu goruyroz. korelasyonlarimiz da fazla ve arti derecede cikmasini bekliyoruz
# ayni zamanda linear model cizgisini de cizdirip goruntuluyoruz
```

```
plot(AlburyWeather$Temp3pm~AlburyWeather$Rainfall,col = AlburyWeather$Color, pch = 19,
     xlab = "Rainfall", ylab = "Sicaklik (saat3)", main ="Sicaklik Karsilastirma Grafigi")
```




```
# verilerimizde bulunan ayriki degerler sebebiyle grafigimiz kotu bir sekilde cikti bu yuzden bir alt s
# aykiri degerlerden kurtulup o sekilde tekrardan cizelim
```

```
library(tidyverse)
rainVStemp3 <- AlburyWeather %>% select(Rainfall,MinTemp,Year,Color) %>% filter(Rainfall<60)
plot(rainVStemp3$MinTemp~rainVStemp3$Rainfall,col = AlburyWeather$Color, pch = 19,
     xlab = "Rainfall", ylab = "Sicaklik (saat3)", main ="Sicaklik/Yagis Miktari Grafigi")
```




```
# aralarindaki iliskiyi simdi daha guzel bir sekilde gorebiliyoruz
```

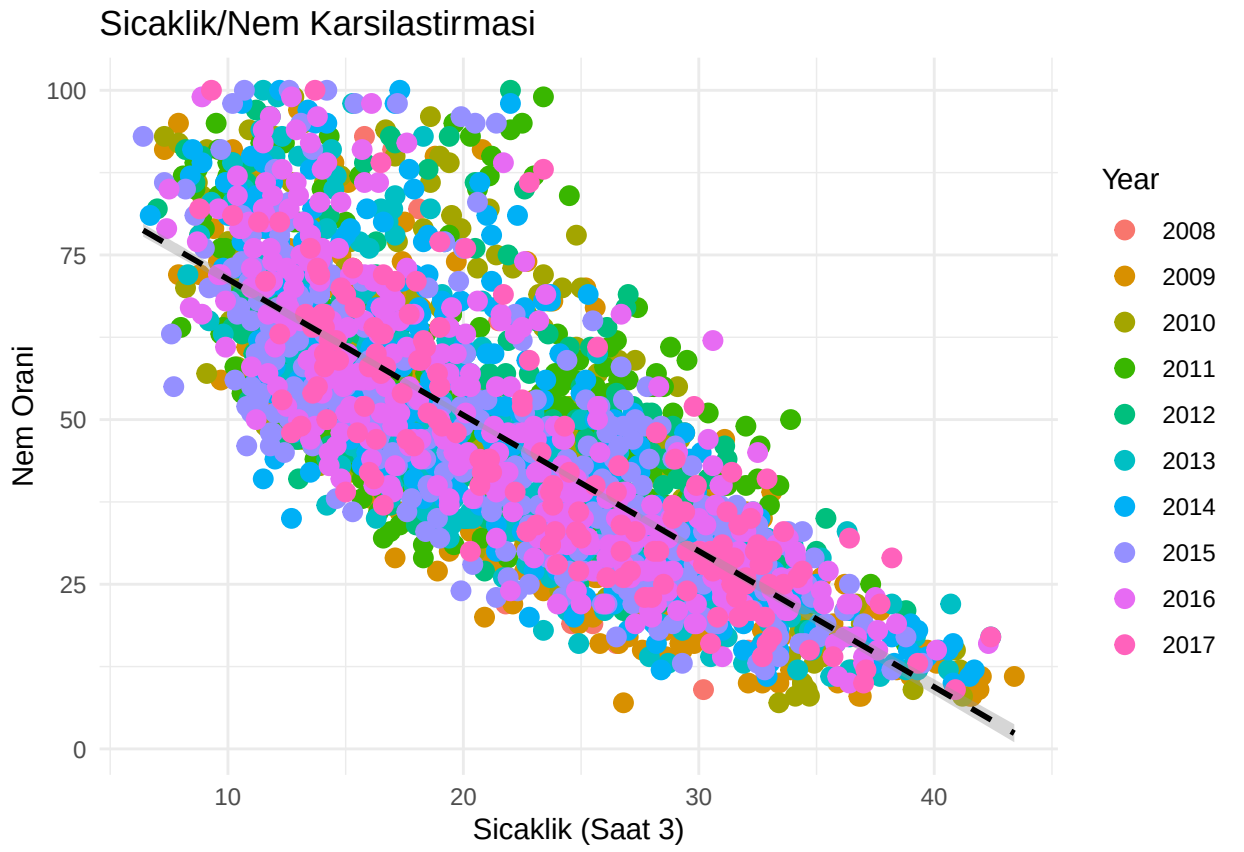
```
library(gifski)
library(ggplot2)
library(gganimate)
p <- ggplot(AlburyWeather, aes(x = as.Date(Date), y = MaxTemp, color = Year)) +
  geom_line() +
  theme_minimal() +
  labs(title = "Yillara Göre Maksimum Sicaklik", x = "Tarih", y = "Maksimum Sicaklik")
animated_plot <- p + transition_time(as.numeric(Year)) +
  labs(title = 'Yıl: {frame_time}')

# burada yillara gore degisen sicakliklarin gif'ini cizdirdik ancak yorumlayicida cikmadi.
# buraya ait gif goruntusunu proje icerisindeki klasorde bulabilirsiniz
```

```
AlburyWeather$Year <- format(as.Date(AlburyWeather$Date), "%Y")
AlburyWeather$Month <- format(as.Date(AlburyWeather$Date), "%m")
AlburyWeather$YearMonth <- format(as.Date(AlburyWeather$Date), "%Y-%m")
# aylara gore olan degisimin de grafiklerini hareketli hazirlamak icin bu sekilde ayirma islemleri gerc
pAy <- ggplot(AlburyWeather, aes(x = Date, y = MaxTemp, color = YearMonth)) +
  geom_line() +
  theme_minimal() +
  labs(title = "Aylara Göre Maksimum Sıcaklık", x = "Tarih", y = "Maksimum Sıcaklık",
       theme(
    legend.position = "right",
```

```
    legend.text = element_text(size = 8),
    legend.title = element_text(size = 10),
    legend.key.size = unit(0.5, "cm")))

animated_plotAy <- pAy + transition_time(as.numeric(factor(YearMonth)))

# bu sekilde aylara gore artis animasyonunu da kaydettim. yine bunu da proje dosyalarinda bulabilirsini
```

```
tempHumi <- ggplot(AlburyWeather, aes(x = Temp3pm, y = Humidity3pm, color = Year)) +
  geom_point(size = 3) +  # Noktaları gösterir
  geom_smooth(method = "lm", color = "black", linetype = "dashed") +  # Regresyon doğrusunu ekler
  labs(
    x = "Sicaklik (Saat 3)",
    y = "Nem Orani",
    title = "Sicaklik/Nem Karsilastirmasi"
  ) +
  theme_minimal()  # Temayı sadeleştirir

print(tempHumi)
```

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



```r
# saat 3 oldugunda sicaklik ve nem orani ne sekilde degisiklik gosterir bunu grafigini cizip bu sekilde
# ayni zamanda dogrusal iliskiyi ifaden eden linear dogrumuzu da ciziyoruz
```

```r
library(tidyverse)
# her bir ayın saat 3 oldugundaki sicaklik degerlerinin dagilisini kontrol edelim
AlburyWeather %>% group_by(YearMonth) %>% summarise(shapiro.test(Temp3pm)$p.value)
```

```
## # A tibble: 100 x 2
##    YearMonth `shapiro.test(Temp3pm)$p.value`
##    <chr>                                <dbl>
##  1 2008-12                            0.204
##  2 2009-01                            0.412
##  3 2009-02                            0.219
##  4 2009-03                            0.00719
##  5 2009-04                            0.211
##  6 2009-05                            0.277
##  7 2009-06                            0.0666
##  8 2009-07                            0.185
##  9 2009-08                            0.458
## 10 2009-09                            0.759
## # i 90 more rows
```

```r
# hepsini tek tek inceledigimde normallik gosterdigini gorduk. ilk varsayim testimizi basariyla tamamla
```

```r
t.test(AlburyWeather$Temp3pm[AlburyWeather$Year==2009],mu = 20,alternative = "two.sided",conf.level = 0
```

```
##
##  One Sample t-test
##
## data:  AlburyWeather$Temp3pm[AlburyWeather$Year == 2009]
## t = 4.6427, df = 361, p-value = 4.822e-06
## alternative hypothesis: true mean is not equal to 20
## 95 percent confidence interval:
##  21.16287 22.87193
## sample estimates:
## mean of x
##   22.0174
```

```r
# H0: 2009 yilindaki hava sicakliklarinin ortalama degeri 20'dir
# Ha: 2009 yilindaki hava sicakliklarinin ortalama degeri 20 degildir
# yillara gore sicaklik degerlerinin ortalamasini karsilastirmak icin onden bir alistirma yaptik
# verilerimiz normallik gosterdigi icin t.test kullandik ancak iki degiskenli istatistik analizi icin
# varyans homojenliklerini de kontrol etmemiz gerekecek.
# p degerigimiz 0.05den kucuk geldigi icin %95 dogruluk oraninda 2009 yilina ait saat 3 sicaklik degeri
# 20'ye esit degildir diyoruz.
```

```r
bartlett.test(Temp3pm~Year, data = AlburyWeather[AlburyWeather$Year==c(2009,2010),])
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Temp3pm by Year
## Bartlett's K-squared = 1.9068, df = 1, p-value = 0.1673
```

```r
# H0: Varyans homojenligi gosteriyor
# Ha: Varyans homojenligi gostermiyor
# 2009 ve 2010 yillarindaki saat 3 sicaklik degerlerinin ortalama degerlerini karsilastirabilmek icin y
# p degerim 0.16 cikti. %95 dogruluk oraninda H0 kabul edildi
# verilerimiz varyans homojenligi gosteriyor
```

```r
yil9 <- AlburyWeather[AlburyWeather$Year==2009,]
yil10 <- AlburyWeather[AlburyWeather$Year==2010,]

t.test(x = yil9$Temp3pm ,y = yil10$Temp3pm, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  yil9$Temp3pm and yil10$Temp3pm
## t = 2.5454, df = 702.41, p-value = 0.01113
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.3407389 2.6393127
## sample estimates:
## mean of x mean of y
##  22.01740  20.52738
```

```r
# H0: 2009 ve 2010 yillari arasindaki saat 3e ait sicaklik degerlerimin ortalamalari aynidir
# Ha: 2009 ve 2010 yillari arasindaki saat 3e ait sicaklik degerlerimin ortalamalari farklidir
# %95 dogruluk oranin p degerim 0.05den kucuk ciktigindan dolayi H0 hipotezimi reddettim. ikisinin de o
```

```r
bartlett.test(Temp3pm~Year, data = AlburyWeather[AlburyWeather$Year==c(2011,2012),])
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Temp3pm by Year
## Bartlett's K-squared = 0.057944, df = 1, p-value = 0.8098
```

```r
# H0: Varyans homojenligi gosteriyor
# Ha: Varyans homojenligi gostermiyor
# 2011 ve 2012 yillarindaki saat 3 sicaklik degerlerinin ortalama degerlerini karsilastirabilmek icin y
# p degerim 0.80 cikti. %95 dogruluk oraninda H0 kabul edildi
# verilerimiz varyans homojenligi gosteriyor
```

```r
yil11 <- AlburyWeather[AlburyWeather$Year==2011,]
yil12 <- AlburyWeather[AlburyWeather$Year==2012,]

t.test(x = yil11$Temp3pm, y = yil12$Temp3pm, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  yil11$Temp3pm and yil12$Temp3pm
```

```
## t = 1.4241, df = 654.91, p-value = 0.1549
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.282016  1.770855
## sample estimates:
## mean of x mean of y
##  20.67685  19.93243
```

```
# H0: 2011 ve 2012 yillari arasindaki saat 3e ait sicaklik degerlerimin ortalamalari aynidir
# Ha: 2011 ve 2012 yillari arasindaki saat 3e ait sicaklik degerlerimin ortalamalari farklidir
# 2011 ile 2012 yillarina ait saat 3 sicaklik degerlerimin ortalamalari %95 dogruluk oraniyla ayni kabu
# p degerim 0.15. H0 kabul edildi
```

```
# elimizdeki 2008 veri sayisi az oldugundan anova testi oncesi onu cikaralim
library(tidyverse)
alb <- AlburyWeather %>% filter(Year!=2008)
anova_result <- aov(Temp3pm ~ Year, data = alb)
summary(anova_result)
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## Year           8   2318  289.69   5.149 2.13e-06 ***
## Residuals   2944 165634   56.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(anova_result)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Temp3pm ~ Year, data = alb)
##
## $Year
##                   diff         lwr        upr     p adj
## 2010-2009 -1.49002579 -3.23922605  0.2591745 0.1689337
## 2011-2009 -1.34055146 -3.12116924  0.4400663 0.3201349
## 2012-2009 -2.08497088 -3.85284541 -0.3170964 0.0078533
## 2013-2009 -0.94006193 -2.71071620  0.8305924 0.7779840
## 2014-2009  0.07335299 -1.66329377  1.8099997 1.0000000
## 2015-2009 -0.50891016 -2.23594829  1.2181280 0.9922073
## 2016-2009 -0.68707545 -2.41293837  1.0387875 0.9486263
## 2017-2009  1.62247764 -0.55104640  3.7960017 0.3319280
## 2011-2010  0.14947433 -1.64922866  1.9481773 0.9999994
## 2012-2010 -0.59494509 -2.38103386  1.1911437 0.9826140
## 2013-2010  0.54996387 -1.23887635  2.3388041 0.9896657
## 2014-2010  1.56337878 -0.19180631  3.3185639 0.1263712
## 2015-2010  0.98111563 -0.76456288  2.7267941 0.7186795
## 2016-2010  0.80295035 -0.94156552  2.5474662 0.8869246
## 2017-2010  3.11250343  0.92413886  5.3008680 0.0003576
## 2012-2011 -0.74441942 -2.56128794  1.0724491 0.9394842
## 2013-2011  0.40048954 -1.41908389  2.2200630 0.9989917
## 2014-2011  1.41390445 -0.37259292  3.2004018 0.2538971
```

```
## 2015-2011  0.83164130 -0.94551699  2.6087996 0.8770357
## 2016-2011  0.65347602 -1.12254024  2.4294923 0.9677315
## 2017-2011  2.96302910  0.74947142  5.1765868 0.0011132
## 2013-2012  1.14490896 -0.66219595  2.9520139 0.5670079
## 2014-2012  2.15832387  0.38452751  3.9321202 0.0050883
## 2015-2012  1.57606072 -0.18832933  3.3404508 0.1239049
## 2016-2012  1.39789544 -0.36534431  3.1611352 0.2517277
## 2017-2012  3.70744852  1.50412869  5.9107684 0.0000067
## 2014-2013  1.01341491 -0.76315193  2.7899818 0.7017404
## 2015-2013  0.43115176 -1.33602352  2.1983270 0.9978873
## 2016-2013  0.25298648 -1.51304031  2.0190133 0.9999597
## 2017-2013  2.56253956  0.35698872  4.7680904 0.0095689
## 2015-2014 -0.58226315 -2.31536265  1.1508363 0.9816399
## 2016-2014 -0.76042843 -2.49235684  0.9715000 0.9115842
## 2017-2014  1.54912465 -0.62921873  3.7274680 0.4009143
## 2016-2015 -0.17816528 -1.90045874  1.5441282 0.9999968
## 2017-2015  2.13138780 -0.03930302  4.3020786 0.0589477
## 2017-2016  2.30955308  0.13979716  4.4793090 0.0269034
```

```
# gtum gruplarin birbirleriyle tek tek karsilastirilmasi icin olusturdugumuz anova testi ve test sonras
# yillar arasi sicaklik farkini ve ona bagli p degeri bize aciklayici ifadeler sunuyor.
```

```
head(AlburyWeather)
```

```
##   Temp3pm MinTemp MaxTemp Rainfall WindSpeed3pm Humidity3pm Pressure3pm
## 1    21.8    13.4    22.9      0.6           24          22      1007.1
## 2    24.3     7.4    25.1      0.0           22          25      1007.8
## 3    23.2    12.9    25.7      0.0           26          30      1008.7
## 4    26.5     9.2    28.0      0.0            9          16      1012.8
## 5    29.7    17.5    32.3      1.0           20          33      1006.0
## 6    28.9    14.6    29.7      0.2           24          23      1005.4
##         Date Year   Color Month YearMonth
## 1 2008-12-01 2008 #FF0000    12   2008-12
## 2 2008-12-02 2008 #FF0000    12   2008-12
## 3 2008-12-03 2008 #FF0000    12   2008-12
## 4 2008-12-04 2008 #FF0000    12   2008-12
## 5 2008-12-05 2008 #FF0000    12   2008-12
## 6 2008-12-06 2008 #FF0000    12   2008-12
```

```
# makina ogrenmesi modeli olusturmak icin hani degiskenleri kullanmam gerektigi konusunda ufak goz gezd
library(tidyverse)
dataModel <- AlburyWeather %>% select(Temp3pm,MinTemp,MaxTemp,Rainfall,WindSpeed3pm,Humidity3pm)
# bu verileri kullanmaya karar verdim. bu verileri kullanmama karar vermemde yukarida grafigini cizdigi
head(dataModel)
```

```
##   Temp3pm MinTemp MaxTemp Rainfall WindSpeed3pm Humidity3pm
## 1    21.8    13.4    22.9      0.6           24          22
## 2    24.3     7.4    25.1      0.0           22          25
## 3    23.2    12.9    25.7      0.0           26          30
## 4    26.5     9.2    28.0      0.0            9          16
## 5    29.7    17.5    32.3      1.0           20          33
## 6    28.9    14.6    29.7      0.2           24          23
```

```r
# modelimiz icin trainSet ve testSet olusturalim. bunlar icin oncelikle rastgele indexler secmemiz gere

set.seed(125)
ind <- sample(1:nrow(dataModel), size = nrow(dataModel)*0.75)
# verilerimin %75'ini modelimi egitmek icin %25'ini modelimi test etmek icin kullanacagim
trainSet <- dataModel[ind,]
testSet <- dataModel[-ind,]
# train ve test verisetlerimizi olusturduk
# oncelikle linear model kullanacagiz daha sonra ayni modeli MARS modeliyle olusturup modellerimizi kar
# Saat 3 sicaklik degerimiz uzerine bir model olusturalim
# verilerim dogrusal dagildiklari icin linear model kullanacagim

# Linear Model
modelLM <- lm(Temp3pm ~ . , data = trainSet)
summary(modelLM)
```

```
##
## Call:
## lm(formula = Temp3pm ~ ., data = trainSet)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8564 -0.3730  0.1109  0.5747  2.2735
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.791714   0.224150  16.916  < 2e-16 ***
## MinTemp       0.018894   0.007623   2.479   0.0133 *
## MaxTemp       0.870145   0.007917 109.912  < 2e-16 ***
## Rainfall      0.013301   0.003349   3.971 7.37e-05 ***
## WindSpeed3pm -0.029370   0.003043  -9.653  < 2e-16 ***
## Humidity3pm  -0.040121   0.001958 -20.490  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9471 on 2230 degrees of freedom
## Multiple R-squared:  0.9843, Adjusted R-squared:  0.9842
## F-statistic: 2.789e+04 on 5 and 2230 DF,  p-value: < 2.2e-16
```

```r
# modelimizin genel sonuclari bu sekilde cikti
# genel olarak iyi bir sonuc gibi gozukuyor.
```

```r
## tahmin islemini gerceklestirelim
tahminLM <- predict(modelLM,testSet)
head(tahminLM)
```

```
##        2        6       18       19       20       26
## 24.12298 28.28585 21.93940 25.03321 28.11651 30.93878
```

```r
# tahmin degerlerime ufak goz attik
```

```r
# modelimizi degerlendirelim
library(caret)
```

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

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
R2(tahminLM,testSet$Temp3pm)
```

```
## [1] 0.9892858
```

```r
# R2 degerimiz 0.98 geldi cok iyi bir puan yakalamisiz bu da modelimizin dogrulugu konusunda guven veri
# diger testlerimize bakalim
```

```r
RMSE(tahminLM,testSet$Temp3pm)
```

```
## [1] 0.7814156
```

```r
# ortalama hatalarin kareler toplami 0.78 geldi
# sicaklik verilerimi goz onune bulundurdugumda hatali sonuclarimin karelerinin ortalamasi 0.78 gelmis
# bu da modelimin iyi bir is cikarttigini gosteriyor
```

```r
MAE(tahminLM,testSet$Temp3pm)
```

```
## [1] 0.6020751
```

```r
# hatalarimin toplaminin ortalamasi 0.60 cikti
# yine bu da tahminlerimdeki hatalarin ortalamasinin 1 derece dahi olmadigini gosteriyor
# modelimiz guzel sonuclar vermis.

# ayni modeli MARS yontemiyle de olusturup testlerimizi gerceklestirelim
```

```r
library(earth)
```

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

```
## Zorunlu paket yükleniyor: plotmo
```

```
## Zorunlu paket yükleniyor: plotrix
```

```
modelMARS <- earth(Temp3pm ~ . , data = trainSet, penalty = 2, degree=2, nk=200,
            nfold=10, keepxy=T)
summary(modelMARS)
```

```
## Call: earth(formula=Temp3pm~., data=trainSet, keepxy=T, degree=2, nfold=10,
##             penalty=2, nk=200)
##
##                   coefficients
## (Intercept)         28.1099176
## h(32.4-MaxTemp)     -0.8769149
## h(MaxTemp-32.4)      0.9585053
## h(97-Humidity3pm)    0.0359004
## h(Humidity3pm-97)    0.4628308
##
## Selected 5 of 5 terms, and 2 of 5 predictors
## Termination condition: RSq changed by less than 0.001 at 5 terms
## Importance: MaxTemp, Humidity3pm, MinTemp-unused, Rainfall-unused, ...
## Number of terms at each degree of interaction: 1 4 (additive model)
## GCV 0.9251104  RSS 2051.928  GRSq 0.9837384  RSq 0.9838546  CVRSq 0.9832815
##
## Note: the cross-validation sd's below are standard deviations across folds
##
## Cross validation:   nterms 5.80 sd 1.03    nvars 2.40 sd 0.52
##
##      CVRSq    sd     MaxErr   sd
##      0.983 0.004      -9.82 1.87
```

```
# genel sonuclarimizda yine Rsq yani R2 degerimiz iyi gozukuyor
# ancak model egitimlerinde sonuc iyi olsa da tahminlerde iyi cikmayabilir bu da modelimizde overfittin
# tahminler uzerinden yorumlarimiza bakalim
```

```
tahminMARS <- predict(modelMARS,testSet)
head(tahminMARS)
```

```
##       Temp3pm
## [1,] 24.29327
## [2,] 28.39888
## [3,] 21.76198
## [4,] 24.69582
## [5,] 27.86861
## [6,] 31.16215
```

```
# tahmin degerlerime ufak goz atalim
```

```
# modelimizi degerlendirelim
library(caret)
R2(tahminMARS,testSet$Temp3pm)
```

```
##        Temp3pm
## [1,] 0.9889846
```

```
RMSE(tahminMARS,testSet$Temp3pm)
```

```
## [1] 0.7919226
```

```
MAE(tahminMARS,testSet$Temp3pm)
```

```
## [1] 0.5958188
```

```
# modelimize ait sonuclar bu sekildedir. MARS yontemi linear veya non-linear bir cok model icin kullani
# parametrelerinin bulundugu bir modeldir. kullanimi gittikce yayginlastigindan dolayi burdaki linear v
```

bu projem buraya kadardi sonraki projelerimde gorusmek uzere