

BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ YÖNETİM BİLİŞİM SİSTEMLERİ BÖLÜMÜ VERİ MADENCİLİĞİ DERSİ DÖNEM PROJESİ

RUHİ ÇENET “LUT ÇÖLÜ” YOUTUBE VİDEOSU VERİ ANALİZİ

MERT YERÜK

İÇİNDEKİLER

İçindekiler

[Youtube Veri Çekme 3](#_Toc124192921)

[Excel Verilerini Kaydetme 5](#_Toc124192922)

[Veri Analizi İçin Paketler 6](#_Toc124192923)

[En Çok Kullanılan Kelimelerin Grafiği 8](#_Toc124192924)

[KELİME BULUTU 9](#_Toc124192925)

[Pozitif ve Negatif Kelime Analizi 10](#_Toc124192926)

[Duygu Analizi 11](#_Toc124192927)

[En çok kullanılan kelimeler 12](#_Toc124192928)

[Bar Chart Grafiği 13](#_Toc124192929)

[Pie Chart Grafiği 15](#_Toc124192930)

[Box Plot Grafiği 16](#_Toc124192931)

[Emojilerin Kullanımını Gösteren Grafik 17](#_Toc124192932)

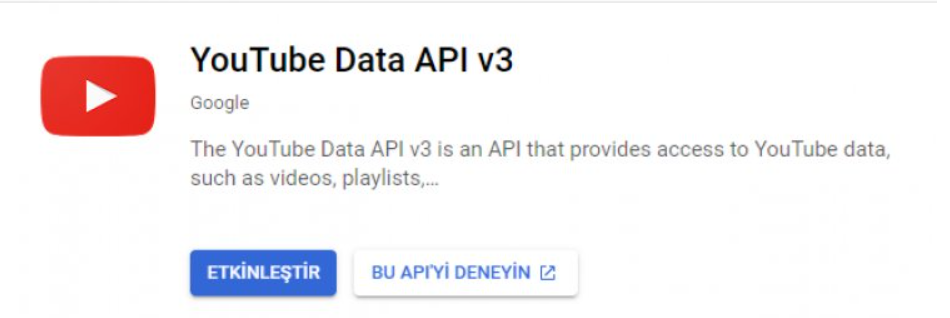
[Emoji Tipleri Grafiği 19](#_Toc124192933)

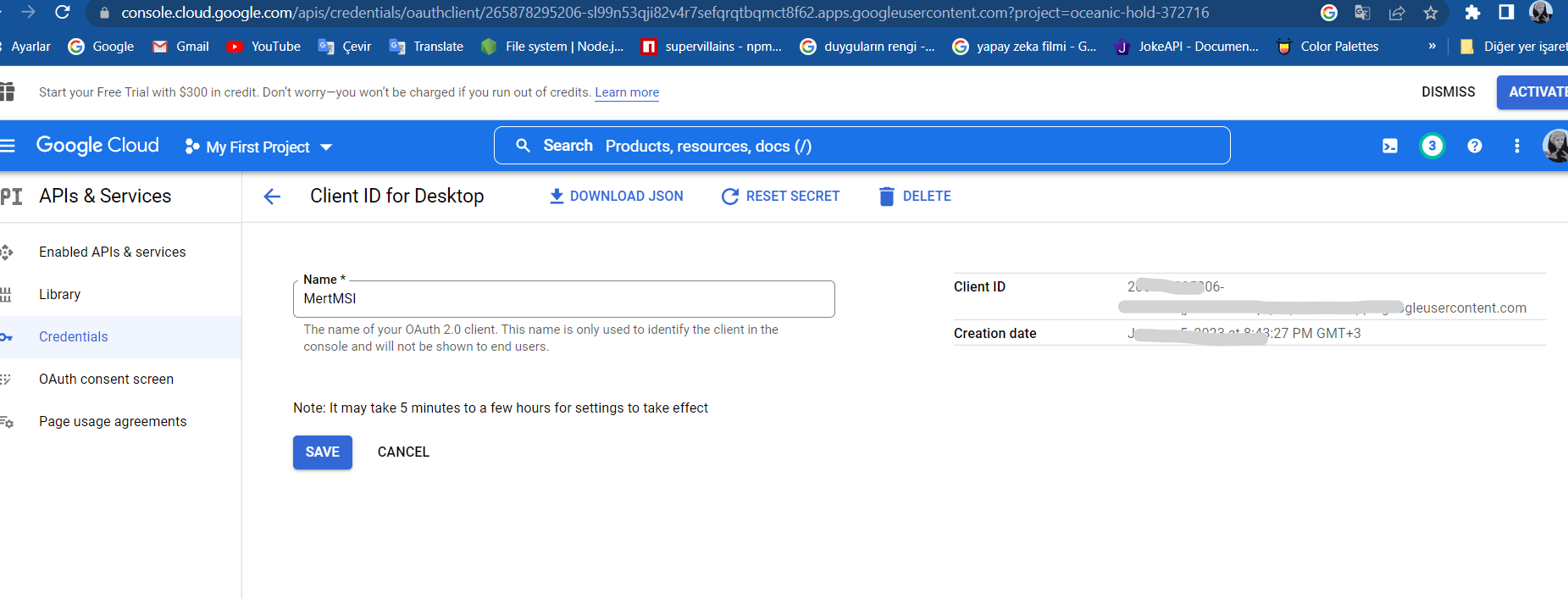
[Emoji Line Chart 20](#_Toc124192934)

[KAYNAKÇA 22](#_Toc124192935)

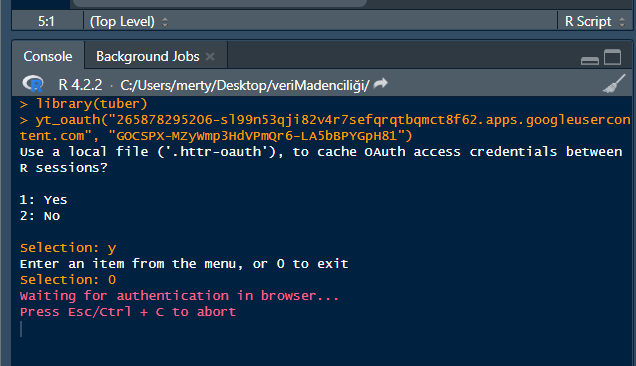
# Youtube Veri Çekme

Ruhi Çenet’in “LUT ÇÖLÜ” video yorumlarını analiz edebilmek için öncelikle Google tarafından bir hesap açıp verileri çekebilmem gerekiyordu.

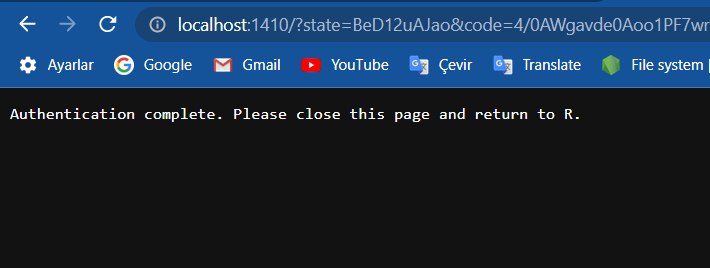




Youtube ofisiyle ilgili video ile bağlantı kurabilmek için gerekli paket olan “tuber’i” yükledim. Daha sonra Google tarafından özel olarak verilen Client ID ve Client secret kodunu yt\_oauth komutuna ekledim. Google sunucusuna bağlanmak için konsole ekranında evet diyerek bağlantı sağladım.ve daha sonra google sunucusuna bağlantı yapmış oldum.

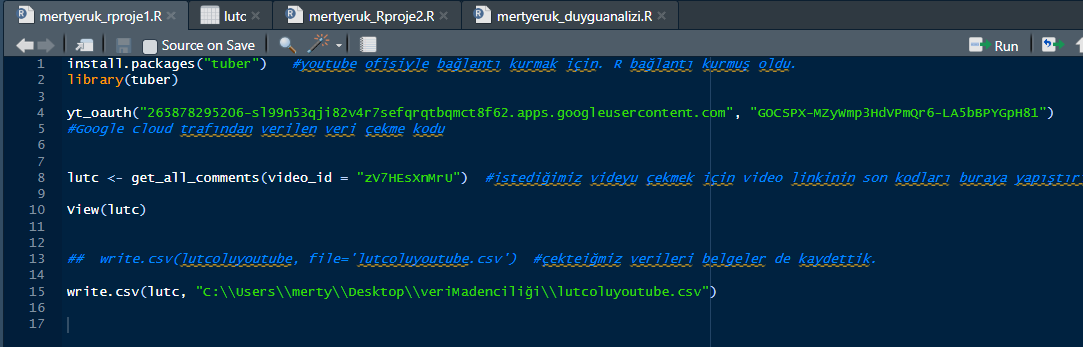


Google sunucusuna başarılı bir şekilde bağlantı sağladım.



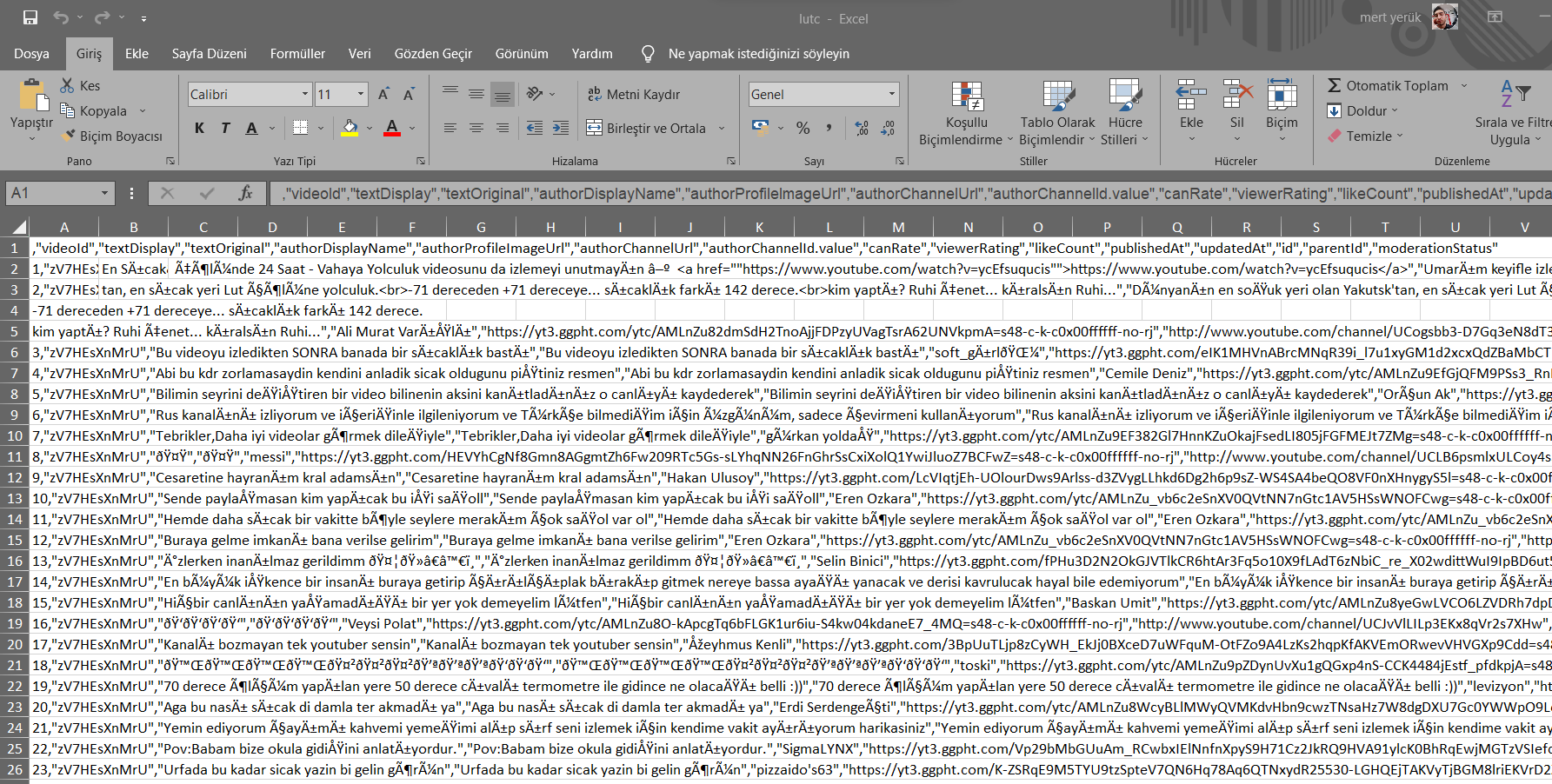
Lutc <- get\_all\_comments(video\_id = “9” ) kısmına istediğim videonun son kısmını ekleyerek yorumların veri olarak indirilmesini sağladım. (video linki = <https://youtu.be/zV7HEsXnMrU>)

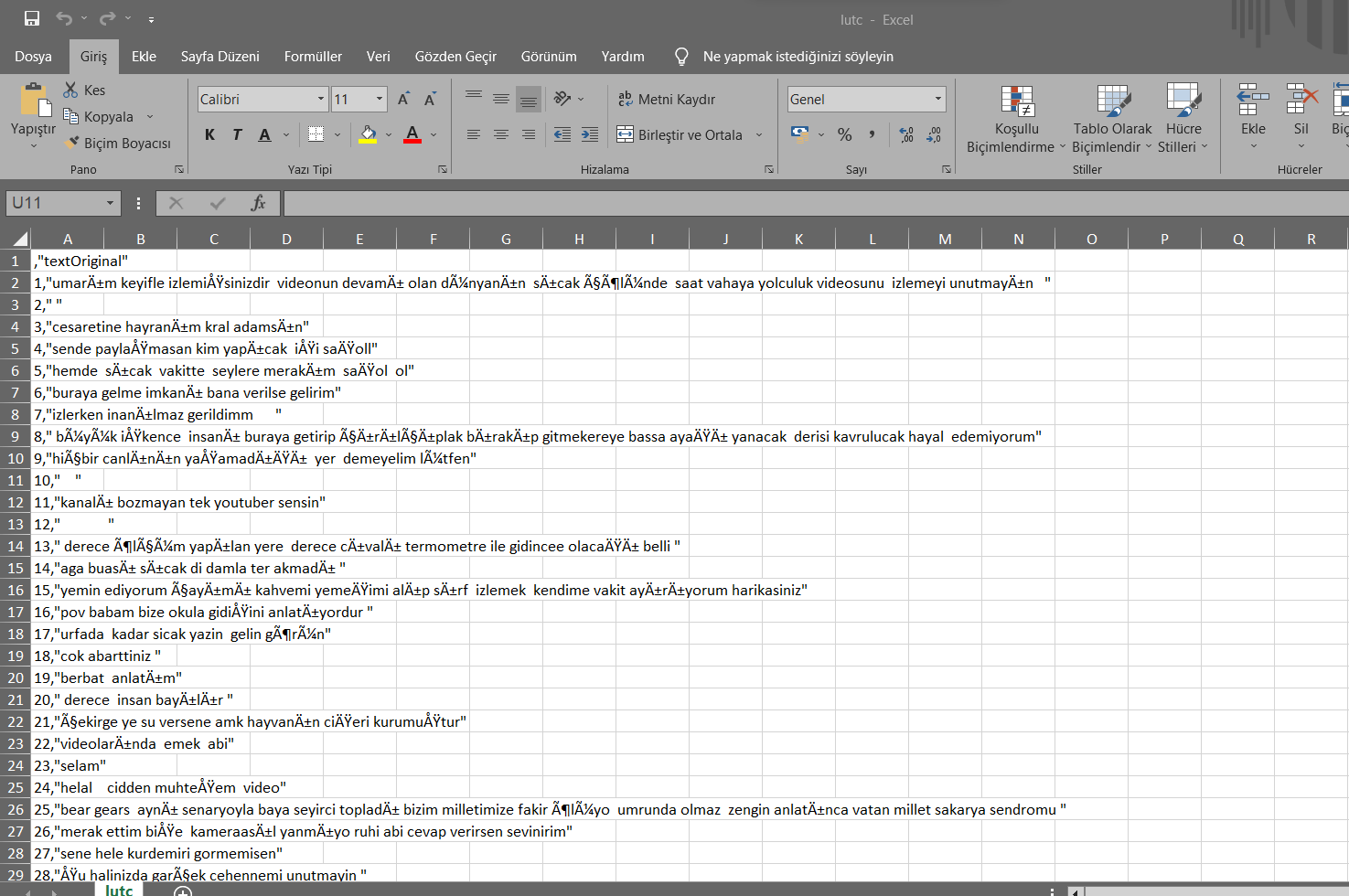
write.csv(lutc, "C:\\Users\\merty\\Desktop\\veriMadenciliği\\lutcoluyoutube.csv") formatı ile verilerimin nereye kaydedileceğini belirttim.

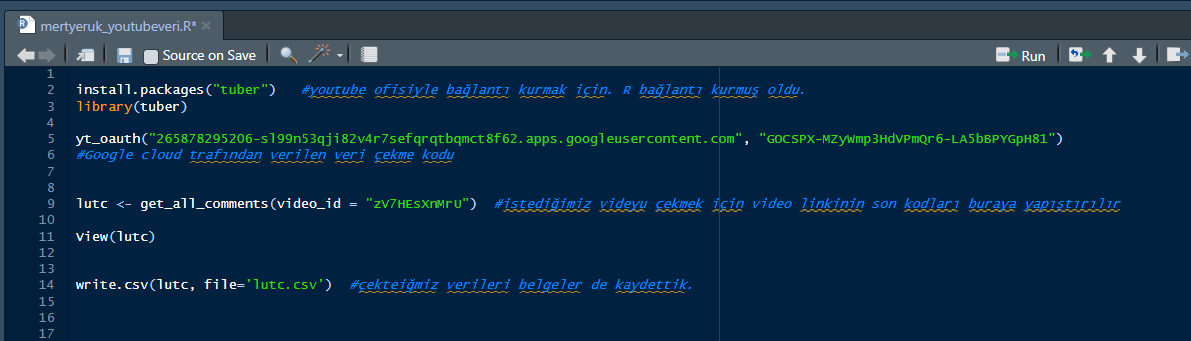


# Excel Verilerini Kaydetme

Excel olarak belgelere kaydedilen verileri öncelikle okunur bir hale getirdim temizleme işlemi yaparak.

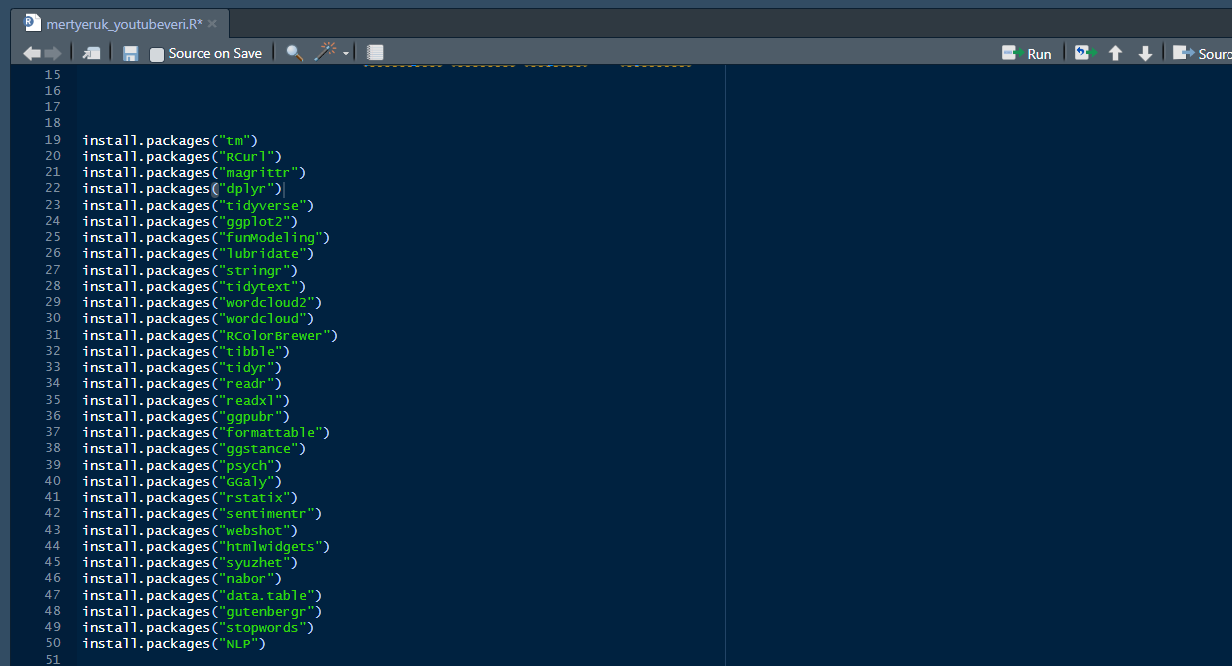


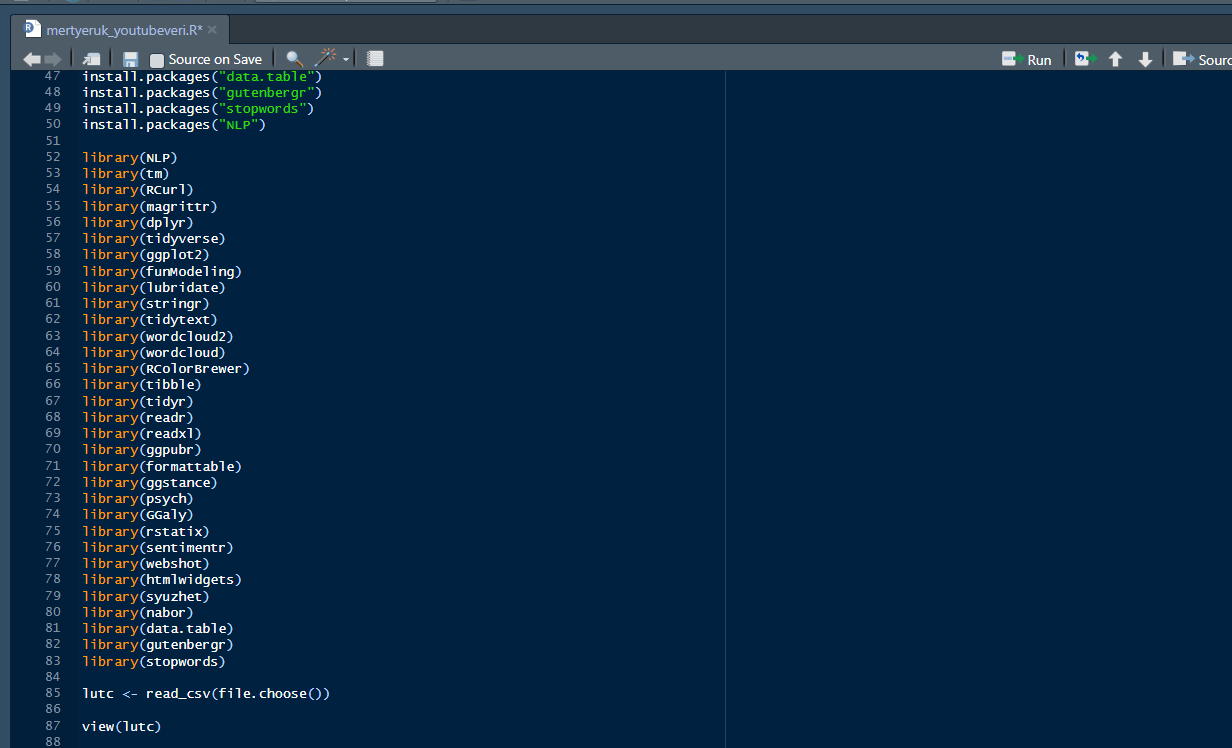




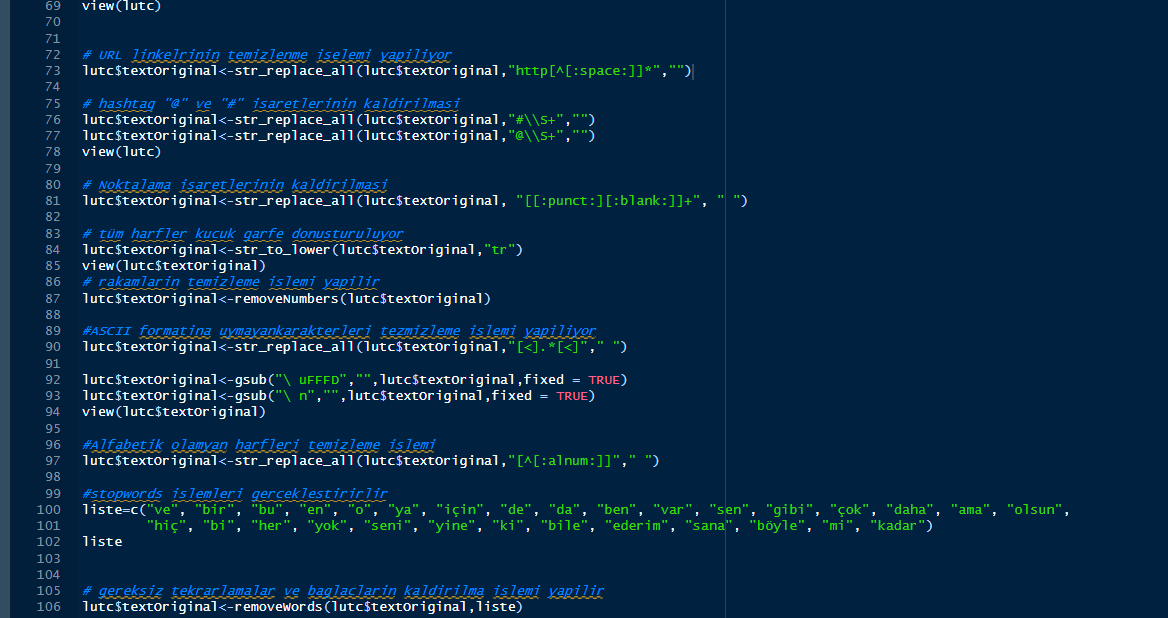
Verileri excel’e aktardıktan sonra veri temizleme işlemi yaptıktan sonra kelime analizine kelime bulutuna başlamam için gerekli paketleri yüklemem gerekiyordu onları sırasıyla yükleyip aktif ettim.

# Veri Analizi İçin Paketler

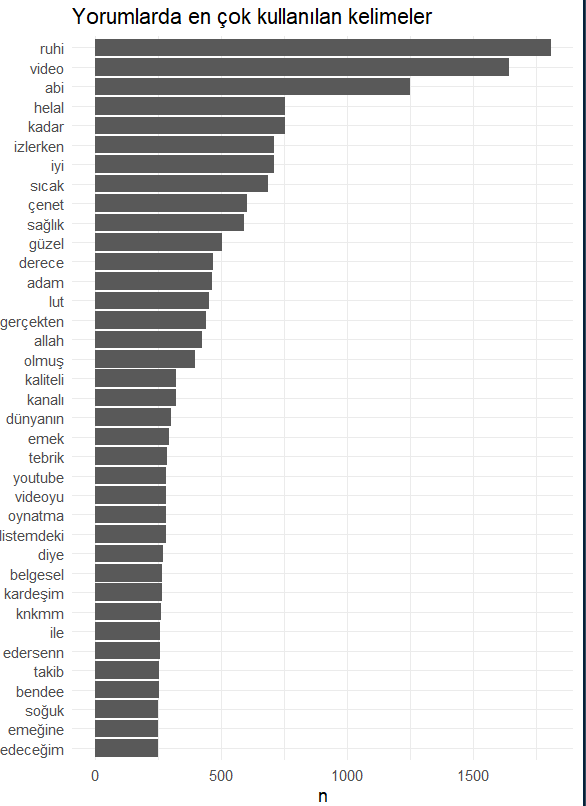




Gerekli paketler de yüklendikten sonra aşağıdaki komutlar ile verilerimin linklerini, hastag gibi işaretleri, noktalama işaretlerini kaldırıp temizleme işlemi yaptım. Tabi çektiğim verilerde veri analizini bozabilecek kelimeler bulunuyordu. Bağlaçlar gibi. Bunları temizledim. Çektiğim veri setinde bana lazım olan kısımlar sütunda gösterilen textOriginal kısmıydı sadece bunları alarak yorum analizine yapmaya başladım. Yorum analizini yapabilmem için bunları kelime haline getirerek temiz\_t olarak adlandırdım. Kullanıdığım kodlar ise;

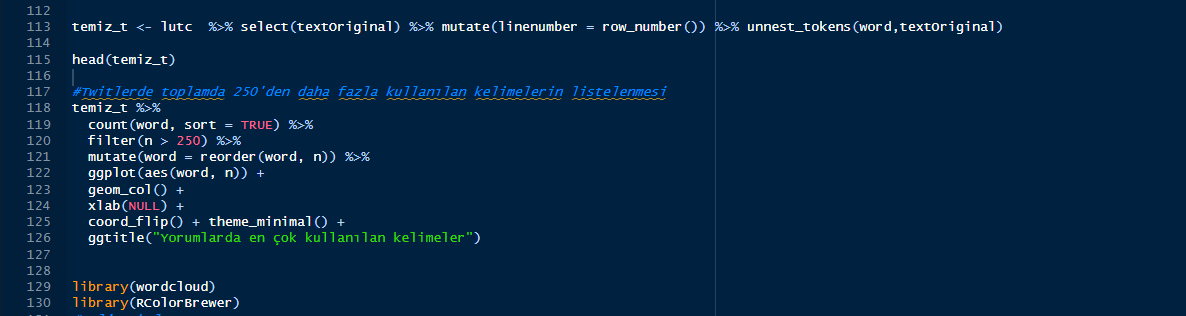


# En Çok Kullanılan Kelimelerin Grafiği



En çok kullanılan kelimelerin grafiğini getirebilmek için aşağıdaki kodları kullandım. Öncelikle verilerimi temiz bir kod haline getirdim temiz\_t olarak ve daha sonra bu kodu kullandım. Bu kod, "temiz\_t" adlı veri kümem içindeki kelimeleri sayıp ve sayısı 250'den büyük olan kelimeleri listeledi. Daha sonra, bu kelimelerin sayısına göre sıralayıp bir bar grafiği oluşturdu. Grafikte, kelimeler y-ekseninde, kelime sayıları ise x-ekseninde gösterilir.

Bu kelimeler doğrultusunda video sahibi Ruhi hakkında yorumlarda çok konuşulduğu belirtilmiş oluyor. Ruhi’ye abi, Çenet olarak hitap edildiği sonucunu çıkarıyoruz. Video’nun ise oldukça kaliteli olduğunu, iyi olduğunu ve çok fazla tebrik, teşekkür aldığını görmüş oluyoruz.



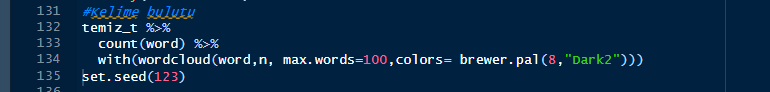
# KELİME BULUTU

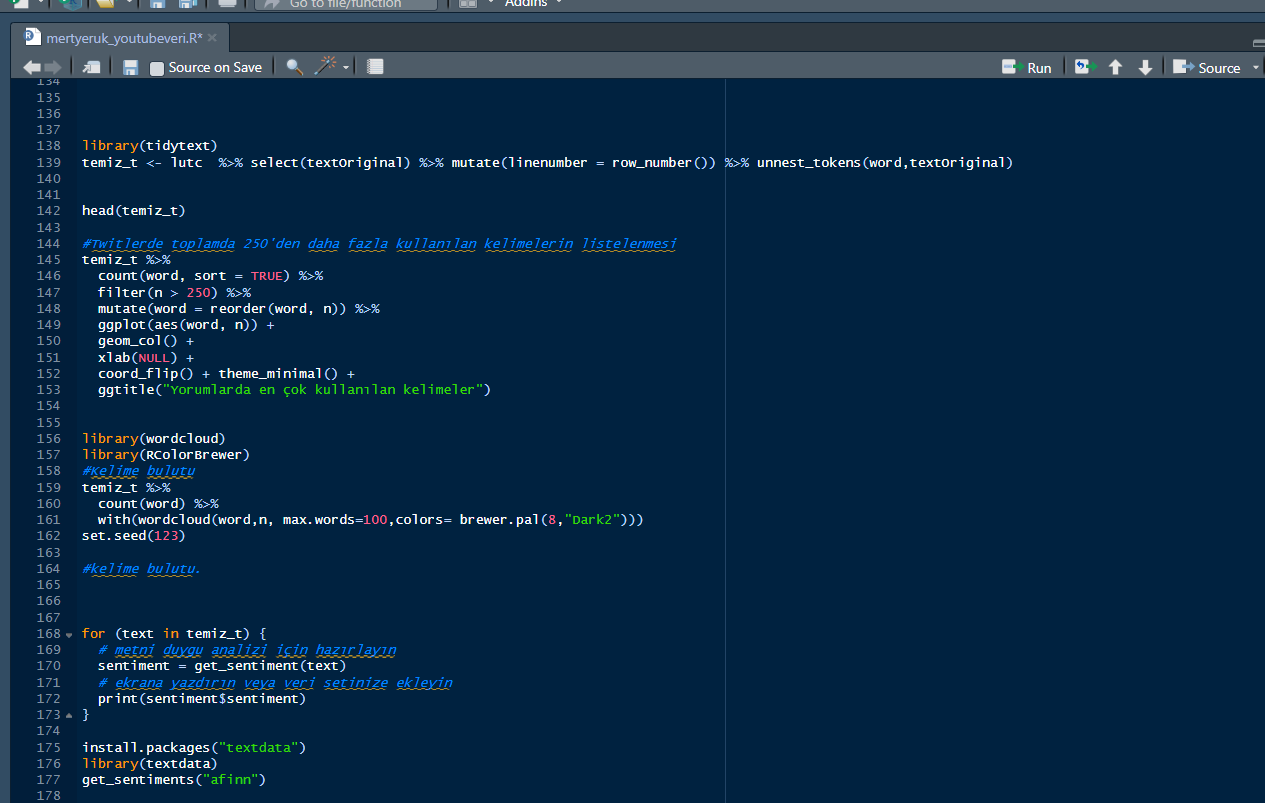
Kelime bulutu, veri kümemdeki kelimelerin büyüklüklerine göre kelime sayılarını gösterir. Çıkan bu kelimelere göre ise gerekli analizi pozitif mi negatif mi anlam içeriyor olarak yorumlar ve analizi yaparım.

Kelime bulutuna baktığımda ise en önce olan kelimelere göre şu analizi yapabilirim. Video sahibinin çok fazla sevildiğini video’nun da oldukça çok fazla beğenildiğini söyleyebilirim. Kelime bulutuna göre Lut Çölü’nde çekilen video’nun çekim sırasında oldukça sıcak bir günde yapıldığı belirtiliyor. Bu video’nun çekimi için çok fazla emek harcandığını bu yüzden de takip sayısında bir artış olduğunu gözlemliyoruz.



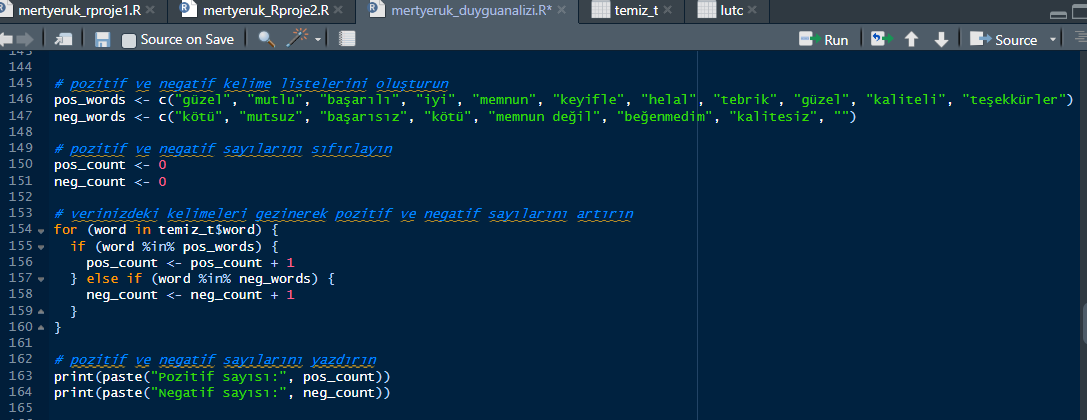
Kelime bulutunu yapabilmek için kullandığım gerekli kodlar aşağıdadır.



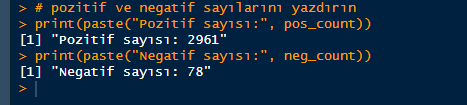


## Pozitif ve Negatif Kelime Analizi

En çok kullanılan kelimelere göre kelime analizi ile olumlu/olumsuz kelimelerin sayısını analiz etme. Bunun için kullandığım kodlar aşağıdadır.

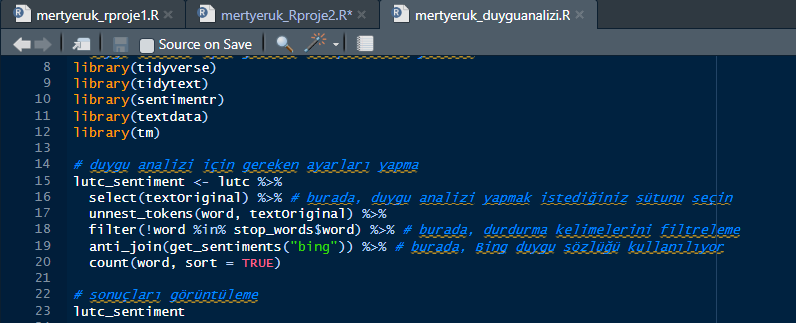


Bu işlemi yaparken kelime bulutunda, kelime grafiklerinde en çok kullanılan kelimelere göre pozitif kelimeleri ve negatif kelimeleri yazıp yorumların genel olarak olumlu mu olumsuz mu olduğunu tespit etmeye çalıştım. Analizimin sonucunda pozitif kelimelerin sayısı 2961 iken, negatif sayıların sayısı ise 78’di. Bu veriler ışığında video’nun oldukça fazlasıyla beğenildiğini söyleyebilirim.



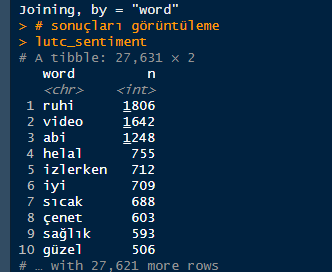
# Duygu Analizi

Duygu analizini yapabilmem için çok fazla kaynağa baktım. Bunların sonucunda en sonunda bir kaç yöntem bulabildim. Öncelikle gerekli kütüphanelerin yüklenmesi gerekiyordu. Bunlar ise library(tidyverse), library(tidytext), library(sentimentr), library(textdata), library(tm) kütüphaneleriydi. Bunları kelime bulutu için yüklemiştim tekrar gerek duymadım. Duygu analizine başlarken öncelikle gerekli ayarları yapmam gerekiyordu. Bunun için lutc\_sentiment adında yeni bir veri oluşturdum.

Duygu analizini yapabilmem için öncelikle lutc veri setimi lutc\_sentiment veri seti adında tekrardan oluşturmam gerekiyordu. Bunun için gerekli kod ayarlarını yaptım.

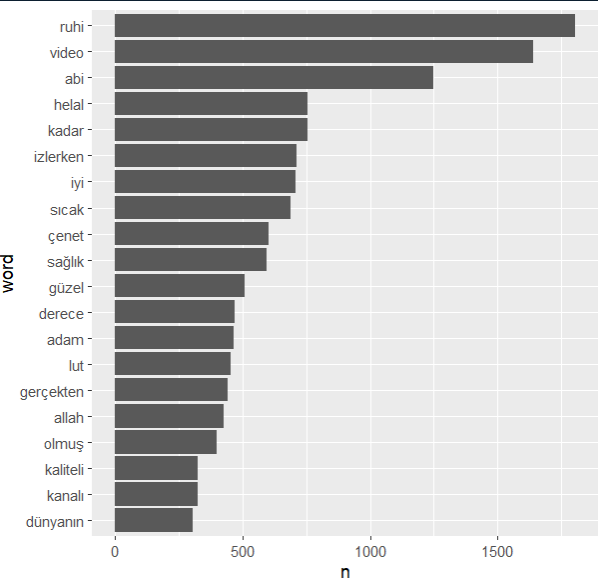
## En çok kullanılan kelimeler

En çok kullanılan kelimelere göre video’nun çok beğenildiği sonucuna ulaşıyoruz.

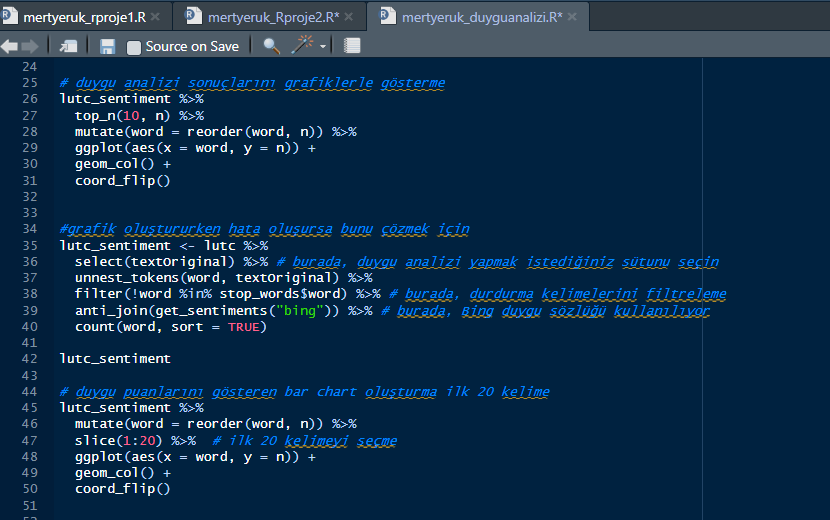


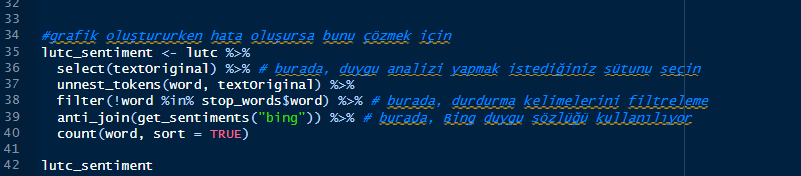
En çok kullanılan ilk 10 kelimeyi aşağıdaki formülle grafiğini getirdim. Bu grafik ise bizi kelimelerin frekans dağılımlarını 0 ile 1500 arasında göstermiş oldu. Grafiğe bakıldığı zaman en çok kullanılan kelimenin Ruhi olduğunu ve videonun “iyi”, “helal” gibi kelime yorumlarıyla çok beğenildiğini anlamış oluyoruz.

## Bar Chart Grafiği



Bar chart grafiğini oluşturmak için kullandığım kodlar.

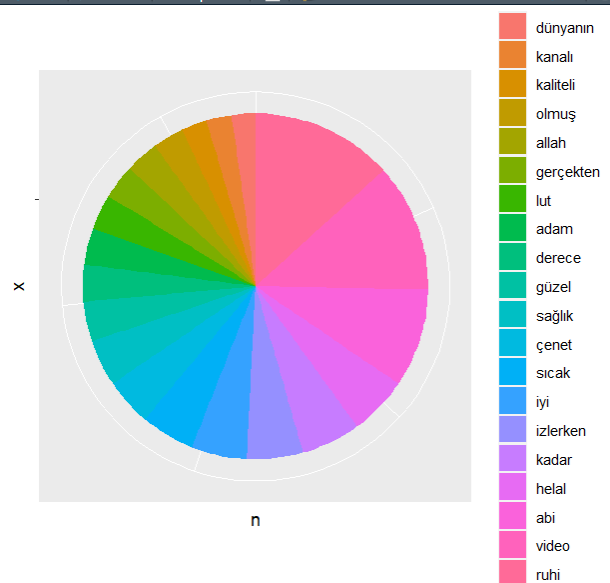


Grafikleri oluşturmaya devam edecektim ancak hatalar oluşabiliyordu bunları engellemek için aşağıdaki kod satırını aktif ettim. Bunun sayesinde kodlarımda oluşan hataların önüne geçilmiş oldu. 

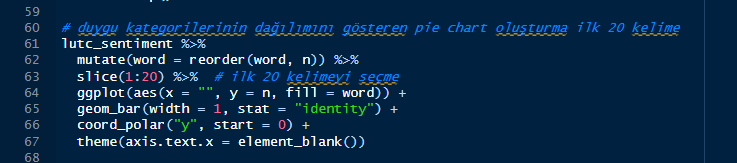
## Pie Chart Grafiği

Duygu kategorilerine göre pie chart grafiği ile kelimelerin renklerine göre dağılımını gösteren grafik.

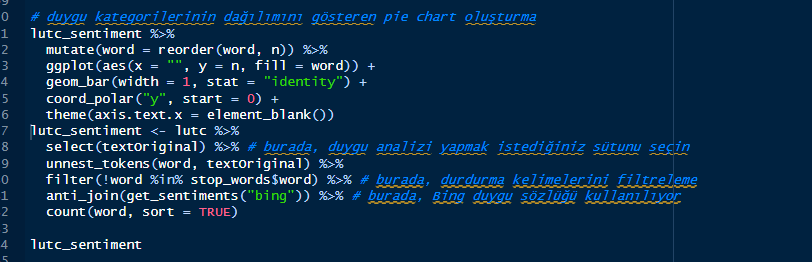
Bu grafiğe göre; Ruhi kelimesi en çok kullanıldığı için koyu pembe renkte, lut adam gibi kelimeler az kullanıldığı için ise yeşil renklerde gösterilmiştir.



Bunu yapmak için ise aşağıdaki kod satırlarını kullandım.

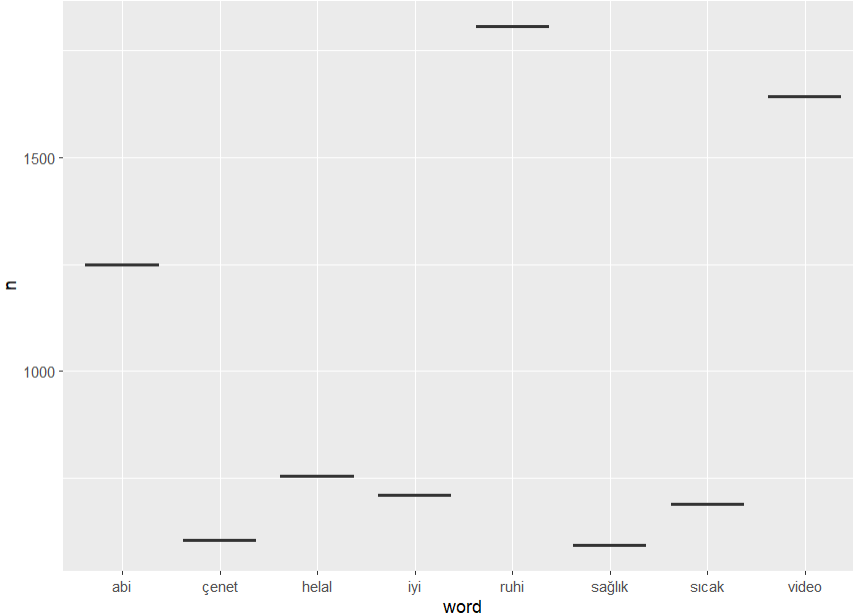


Ben Pie Chart’I yaparken ilk 20 kelimeyi kullandım. Çünkü bütün kelimeleri seçtiğim zaman kodda hata oluşuyordu. Bütün kelimelerin seçili olduğu kod satırı ise aşağıdaki resimdedir.

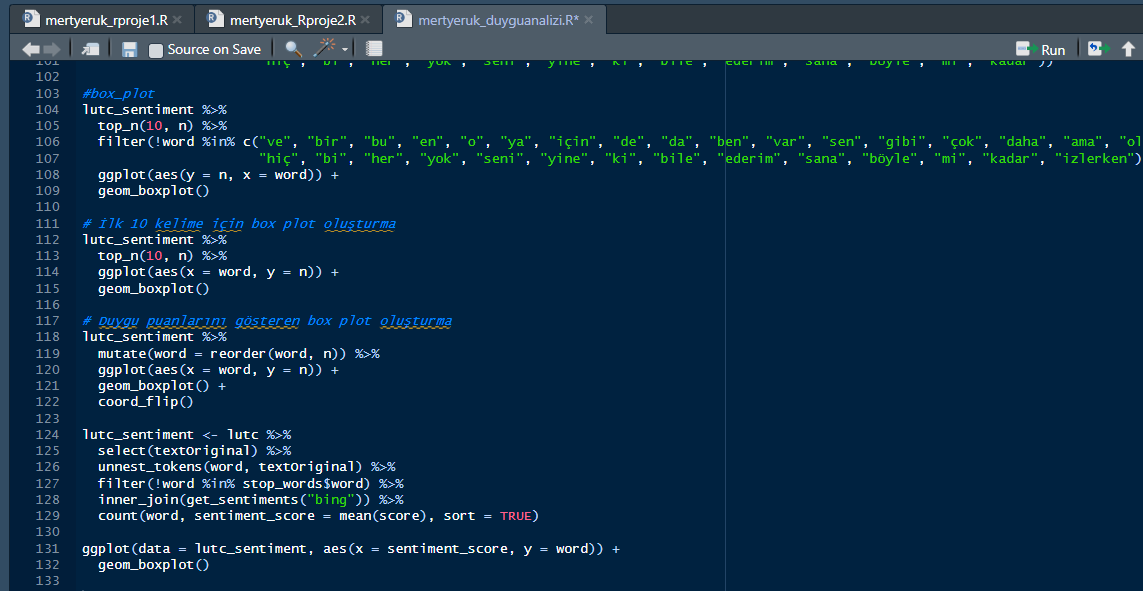


## Box Plot Grafiği

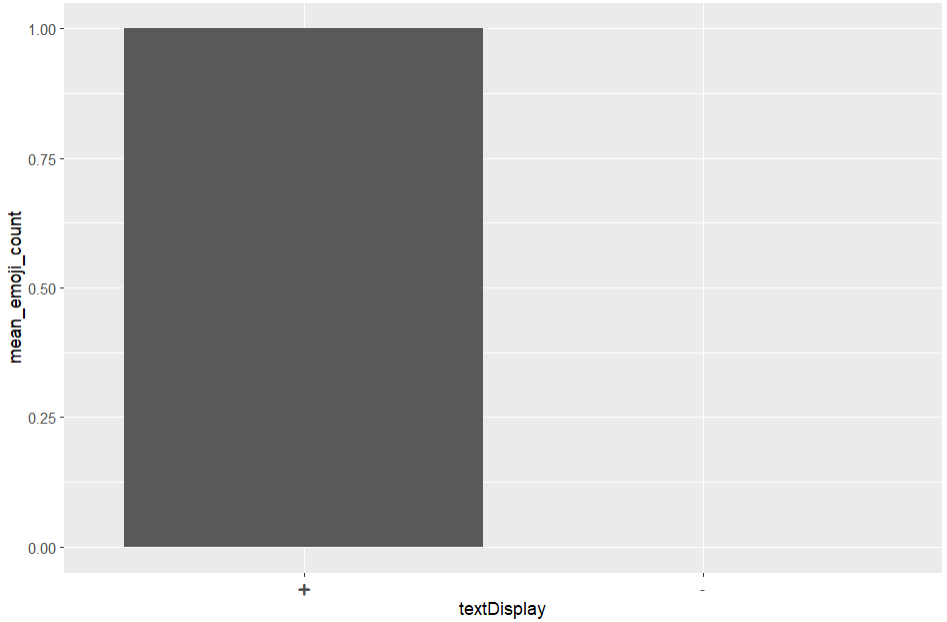
Duygu punlarını gösteren box plot grafiğini kullandım. Bunun için ilk 10 kelimeyi seçtim ve ilk 10 kelimeye göre frekans dağılımları ortaya çıktı. Bu grafiğe göre en çok kullanılan kelime “ruhi” 1500 frekansının üzerindedir.



Bunu yaparken kullandığım kod satırları ise aşağıdadır. Tabi bunu yaparken tekrar kelimeleri filtrelemem gerekiyordu. Öncelikle filtreleme işlemi yaptım. Daha sonra ise bu kodlar ile grafiği getirdim.

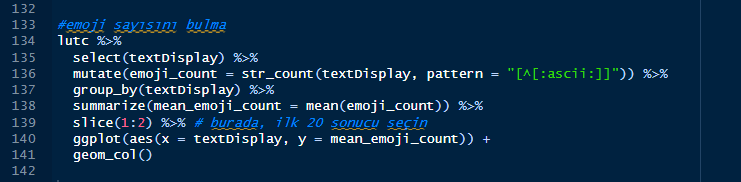


## Emojilerin Kullanımını Gösteren Grafik



Bu grafik ile her bir verinin içinde emoji karakterlerinin ortalama sayısını göstermektedir. Örneğin, eğer "textDisplay" sütunundaki bir veride hiç emoji yoksa, "emoji\_count" değişkenine 0 atanır ve bu verinin "mean\_emoji\_count" değeri de 0 olur. Eğer "textDisplay" sütunundaki bir veride bir tane emoji varsa, "emoji\_count" değişkenine 1 atanır ve bu verinin "mean\_emoji\_count" değeri de 1 olur. Bu grafikte, "mean\_emoji\_count" değişkeni 0 ile 1 arasında + - li ifadeler oluşturmaktadır. Bu ifadeler, "textDisplay" sütunundaki verilerin emoji karakterlerinin ortalama sayısını göstermektedir. Eğer "mean\_emoji\_count" değeri pozitif (+) ise, "textDisplay" sütunundaki verilerin emoji karakterlerinin ortalama sayısı da pozitiftir ve eğer "mean\_emoji\_count" değeri negatif (-) ise, "textDisplay" sütunundaki verilerin emoji karakterlerinin ortalama sayısı da negatiftir. Bu da (+) işareti verdiği için verilerde emoji ifadesinin olduğunu ortaya çıkarır.

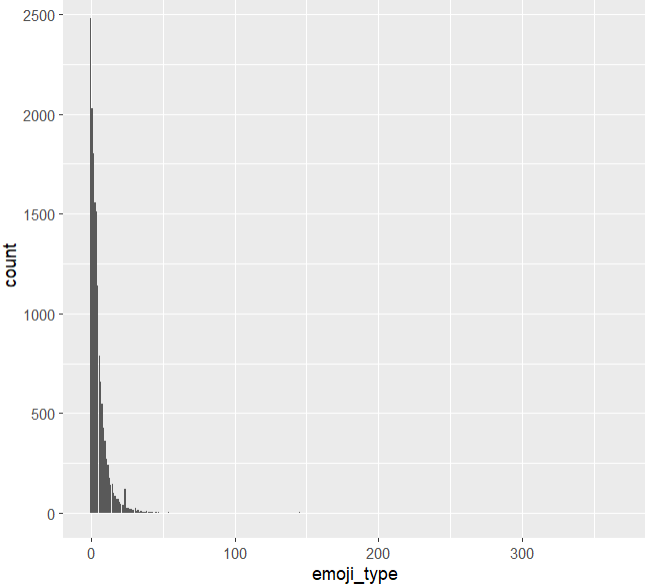
Bu grafiği yapabilmek için aşağıdaki kodları kullandım.



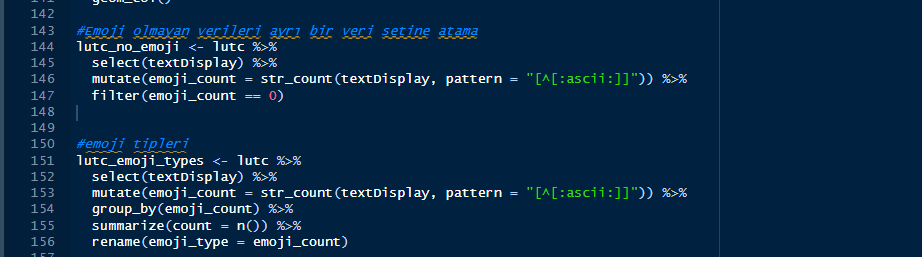
Daha sonra textDisplay de bulunan verilerin emojilerini lutc\_emoji\_types isimli bir veri seti oluşturarak emojilerin analizini yaptım.

Bu kod, "lutc" veri setindeki verileri seçer ve her bir verinin içinde kaç tane emoji olduğunu sayar. Daha sonra, emoji sayısına göre verileri gruplar ve her grup için kaç tane veri olduğunu sayar. Böylece, farklı tipte emojilerin hangi tipte olduğunu ve her tipte kaç tane olduğunu bulabilirsiniz. "lutc\_emoji\_types" isimli veri setinde, farklı tipte emojilerin hangi tipte olduğu ve her tipte kaç tane olduğu bilgisi bulunur. Bu veri setinde de istediğiniz analizlerini yaptım.

## Emoji Tipleri Grafiği

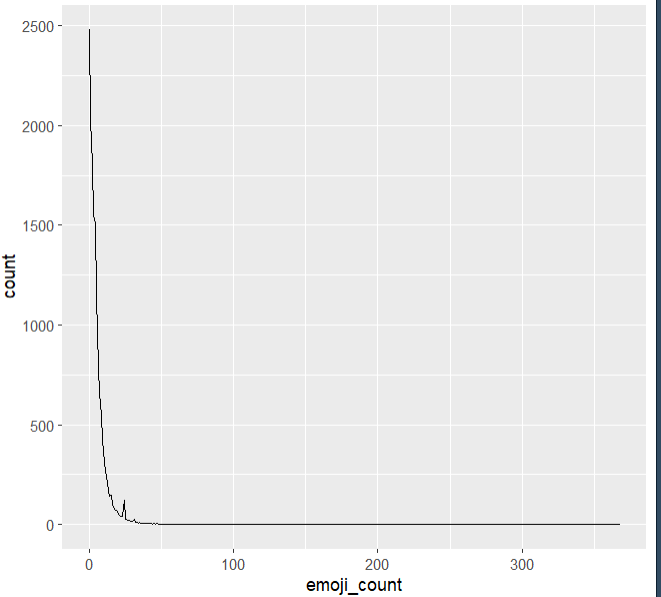


Bunu yaparken öncelikle emoji olmayan verileri ayırdım. Daha sonra bu işleme geçtim. Bunun için kullandığım kodlar ise aşağıdadır.

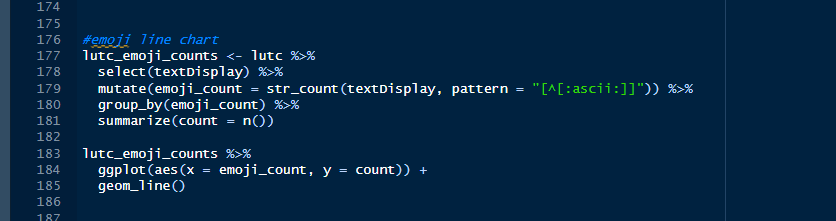


## Emoji Line Chart

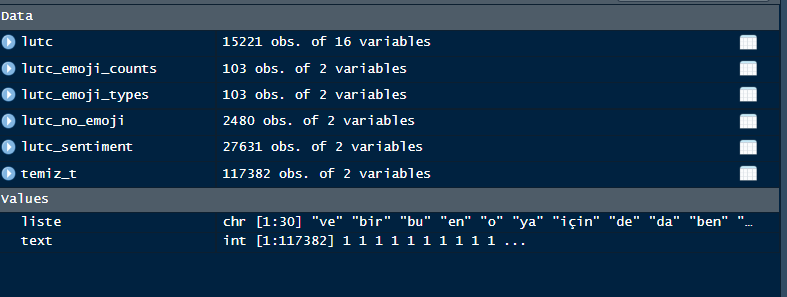
Line chart, verileri birleştirerek değişimlerini gösteren bir grafik türüdür. Bu grafikte, verilerin değişimini sürekli olarak gösterir ve veriler arasındaki ilişkiyi açıkça gösterir. Bu grafikte ise emoji sayılarına göre bir çizgi grafiği oluşturdum. Bu line chart'ta, x-ekseni emoji sayılarını, y-ekseni ise verilerin sayısını gösterir. Sadece emojileri kullandığım için emoji olmayan verilerin sayısı 0 olarak belirtildi. Bu şekilde, emoji sayısının hangi değerlerde daha fazla olduğunu ve hangi değerlerde daha az olduğunu gördük.



Emoji Line Chart Grafiği için kullanıdığım kod satırı ise aşağıdadır.



Veri setimde 15221 gözlem değeri 16 tane değişkenim vardır. Lutc veri setimi analiz edebilmek için öncelikle yorumları kelime kelime ayırmam gerekiyordu. Bu yüzden temiz\_t adlı kelimelerin (toplam 117382 kelime) olduğu bir veri seti oluşturdum. Oluşturduğum bu veri seti ile kelime anlizlerini, kelime bulutunu oluşturdum. Lutc\_emoji\_types ile sadece emoji içeren verileri yeniden oluşturup emoji analizi yaptım.



# KAYNAKÇA

<https://rpubs.com/wangyuy/youtube_analysis>

<https://www.w3schools.com/r/r_graph_line.asp>

<https://chat.openai.com/>

<https://bookdown.org/burak2358/SARP-TR/veri-setleri.html>

<https://www.r10.net/php/2453095-php-ile-youtube-dan-veri-cekme.html>

<https://www.veribilimiokulu.com/r-ile-metin-madeciligi-bolum-2/>

<https://medium.com/@mertyeruk34/r-programlama-dersleri-2-bd7c875350cd>

<https://www.youtube.com/watch?v=bTzpobQDlPc>

<https://www.youtube.com/watch?v=niHYSeSEpvQ>