EDA LA1

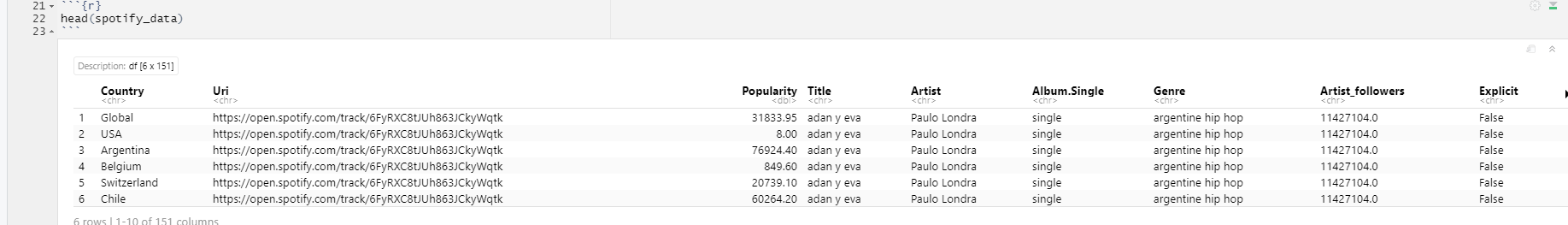
Sachin Kumar and Shivam Kumar

1.Read the csv data file

Spotify data=read.csv (file=file. Choose (), header = TRUE)

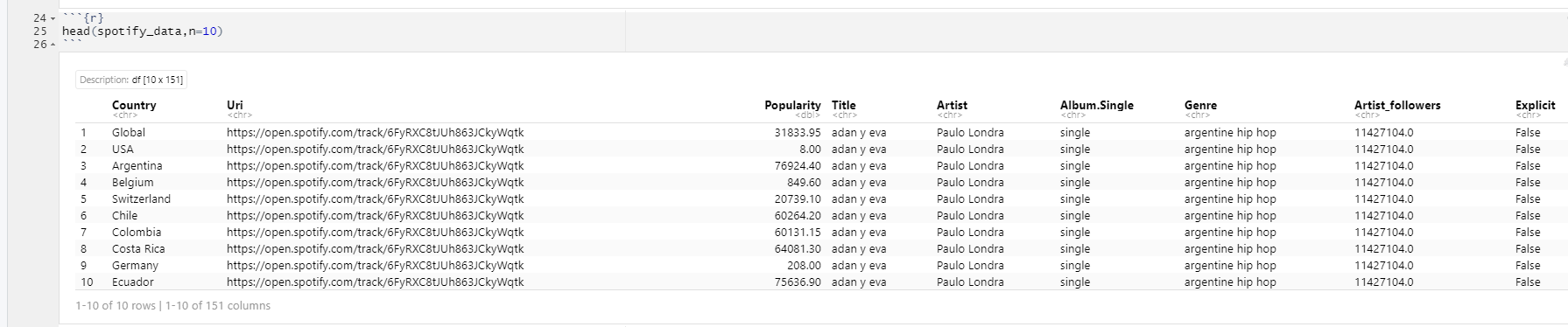
2.Display head of data

head (Spotify data)



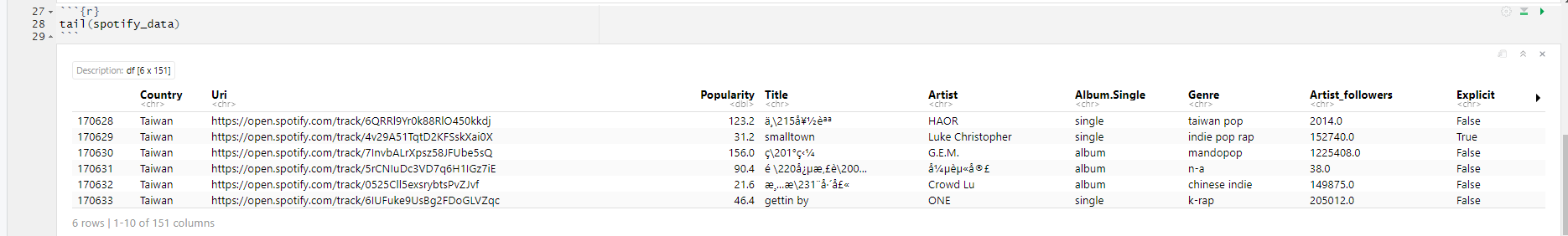
3.Display first n rows specified.

head (spotify\_data, n=10)



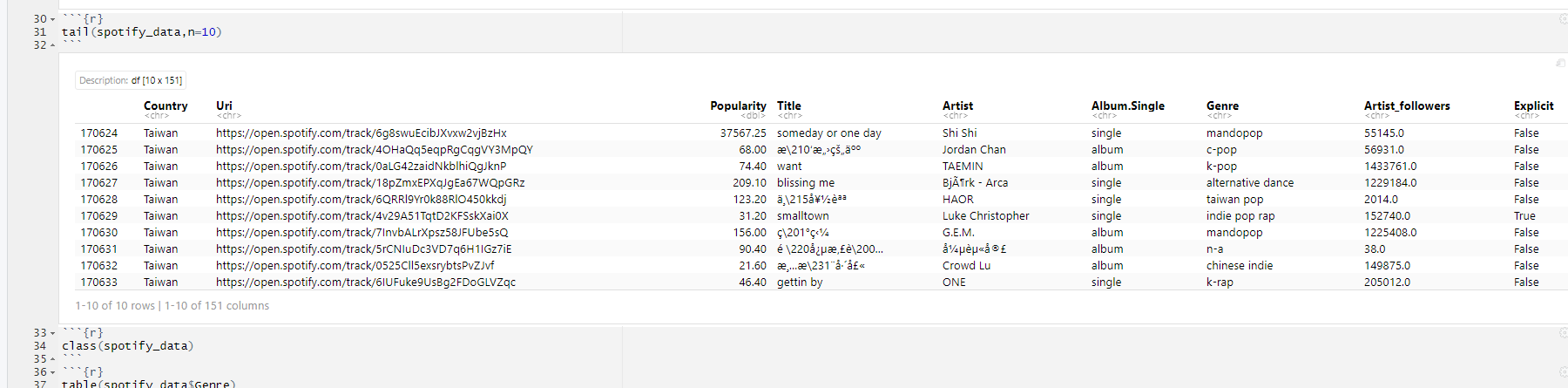
4.Display tail of data.

tail(spotify\_data)

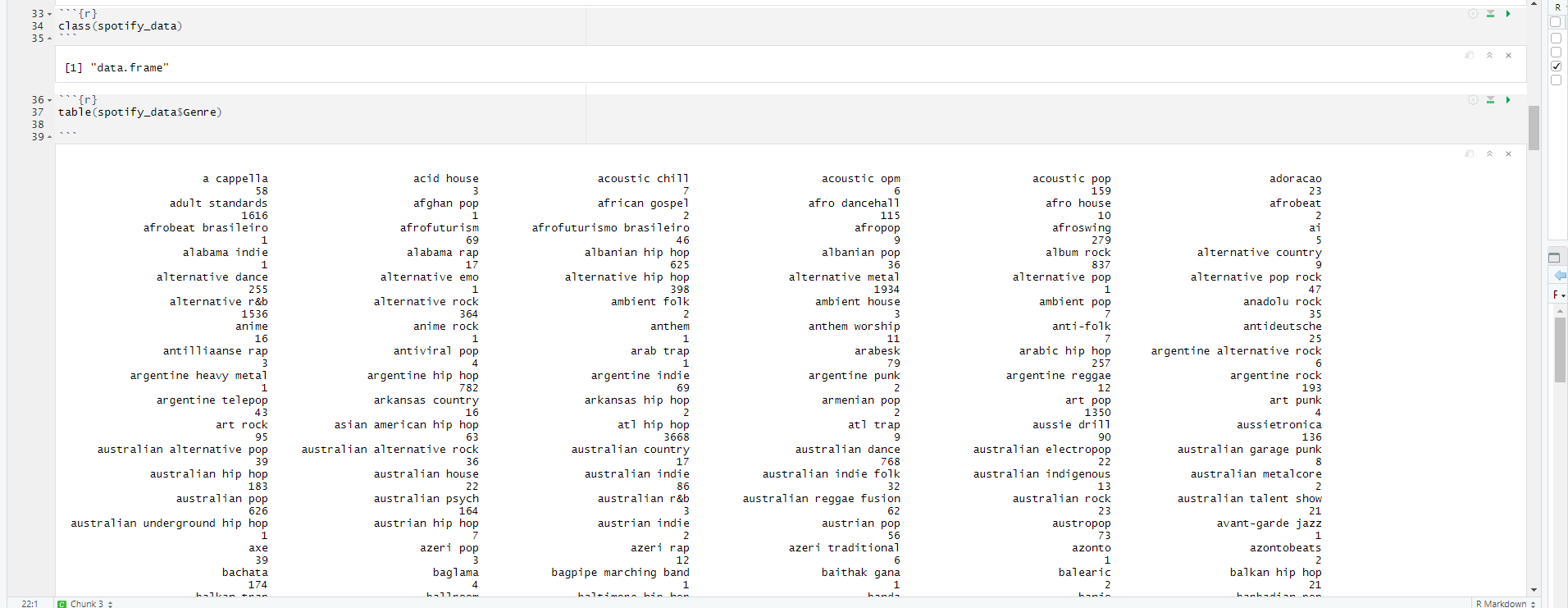


5.Display n rows specified from bottom

tail (spotify\_data, n=10)

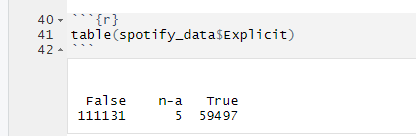


6.Determining Type of data class(spotify\_data)



7.Table command table(spotify\_data$Genre)

table(spotify\_data$Explicit)



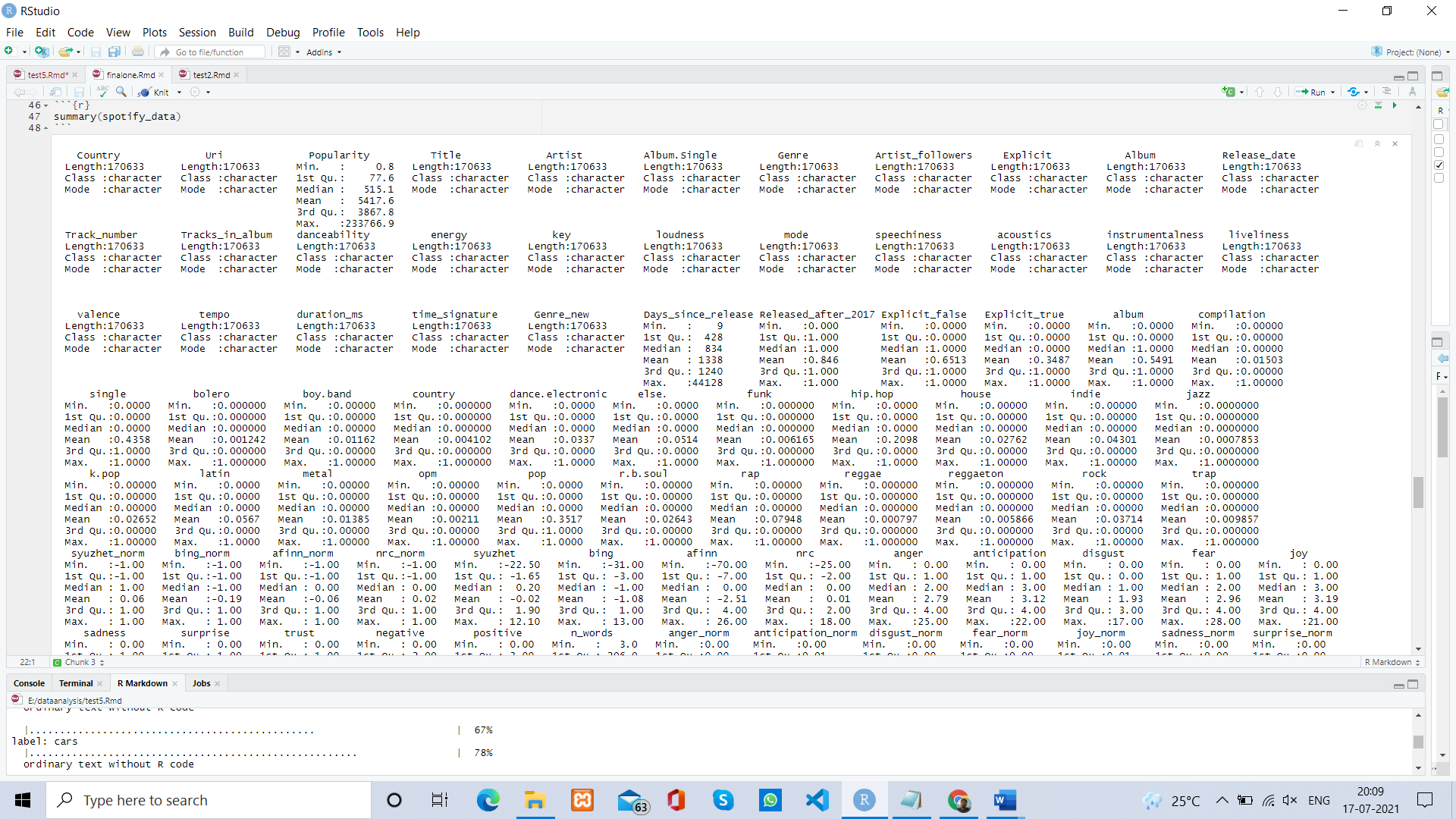
8.Determine the structure of data

str(spotify\_data)



9.Summarising the data

summary(spotify\_data)



10.Displaying Dimension of the data

dim(spotify\_data)

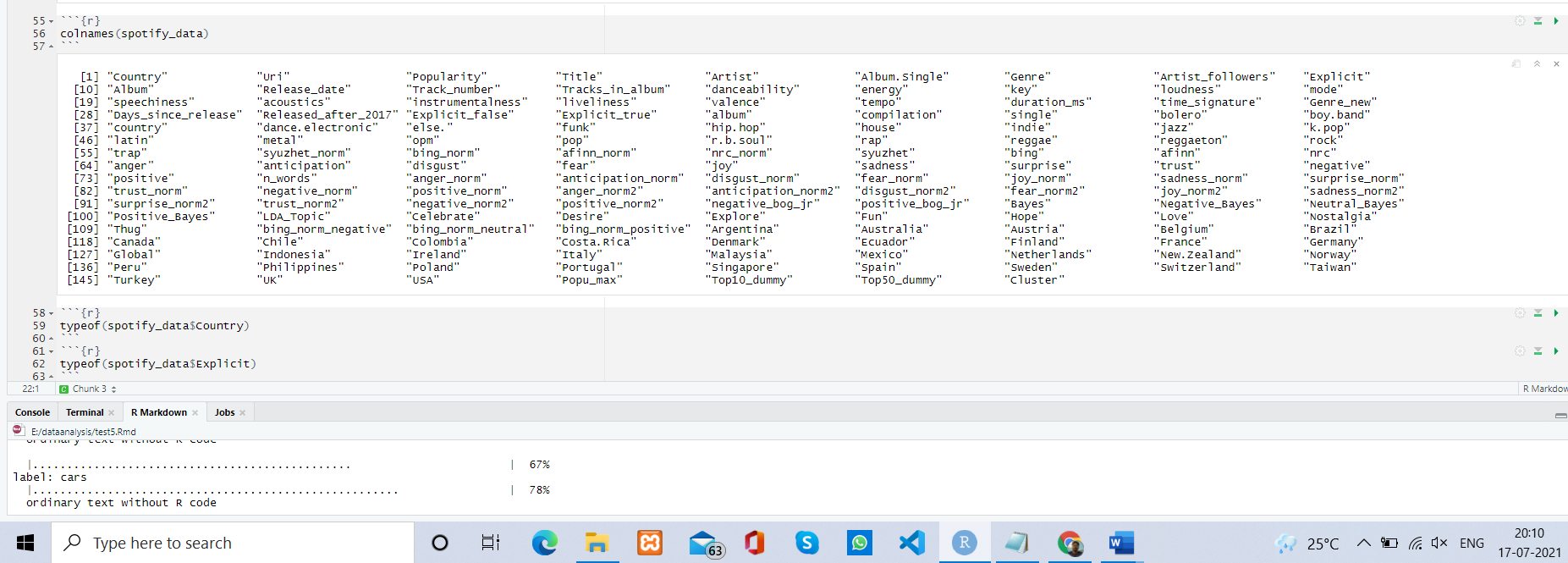
11.Displaying length of Genre column

length(spotify\_data$Genre)



12.Displaying column names of data

colnames(spotify\_data)



13.Displaying structure of some columns in the data

class(spotify\_data$Name) typeof(spotify\_data$Name)

14.Displaying type of some data structure in the data

typeof(spotify\_data$Explicit)



15.List of variables present in Spotify data ls(spotify\_data)



16.Some pattern matching operations on variable of spotify\_data

ls(spotify\_data,pattern="^Sp")

ls(spotify\_data,pattern="^[AD]")

ls(spotify\_data,pattern="t.l")

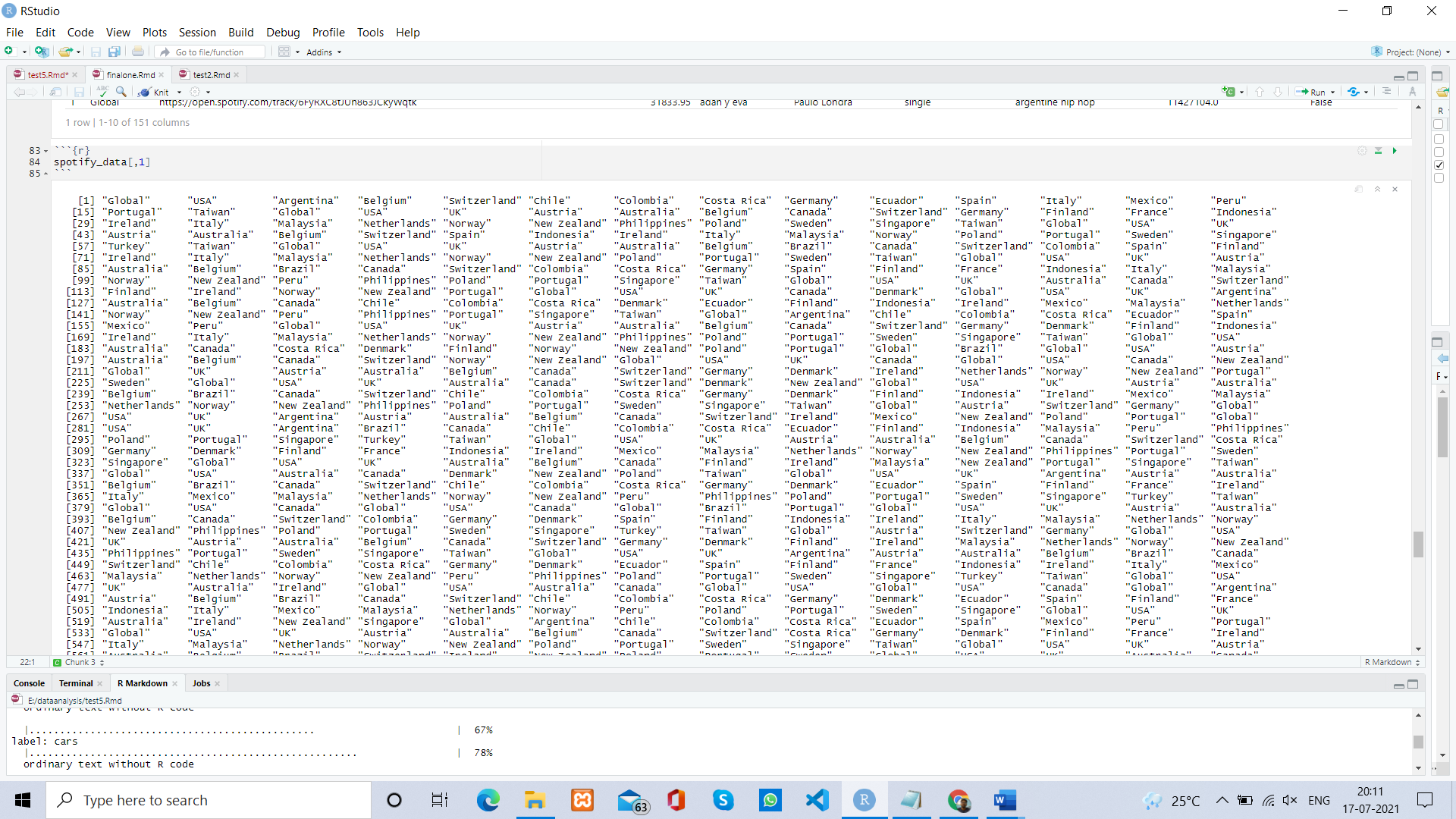
ls(spotify\_data,pattern="ce$")



17.Display 1st row and all columns of data frame spotify\_data[1,]

18.Display all rows and 1st column of data frame

spotify\_data[,1]



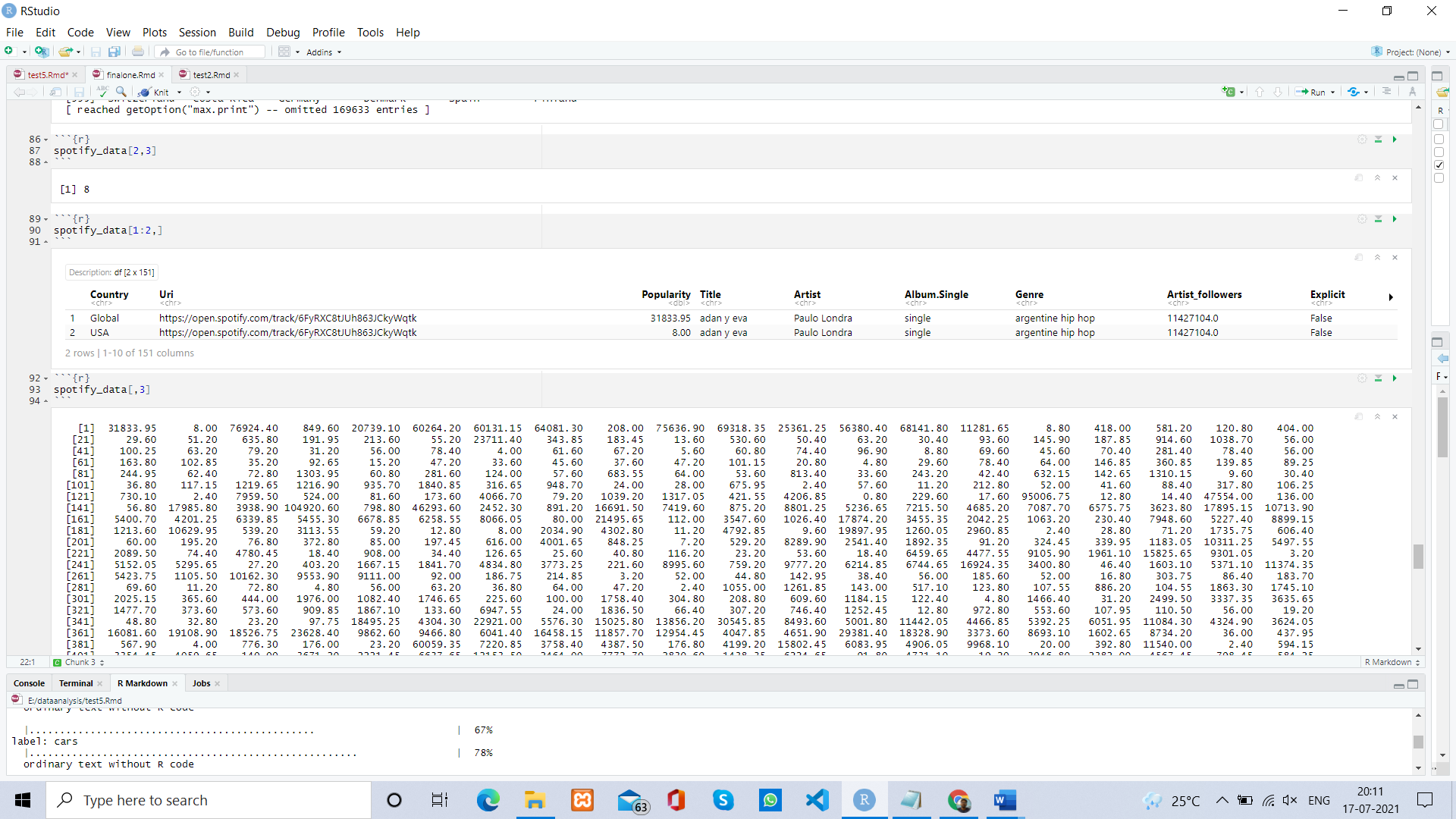
19.Display data in 2nd row and 3rd column of the data frame

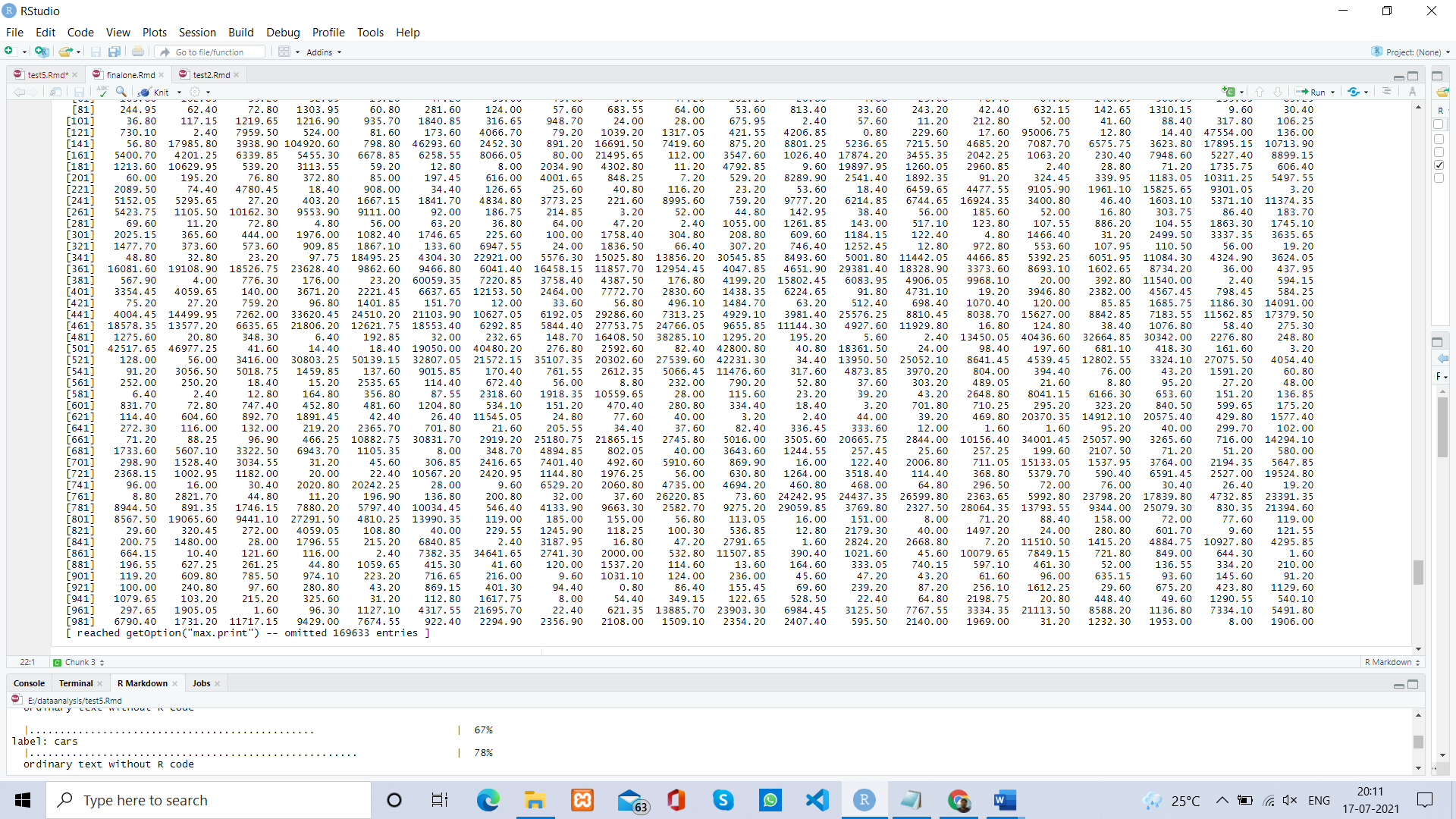
spotify\_data[2,3]

20.Display 1st and 2nd row and all columns spotify\_data[1:2,]

21.Display all rows and first 3 columns

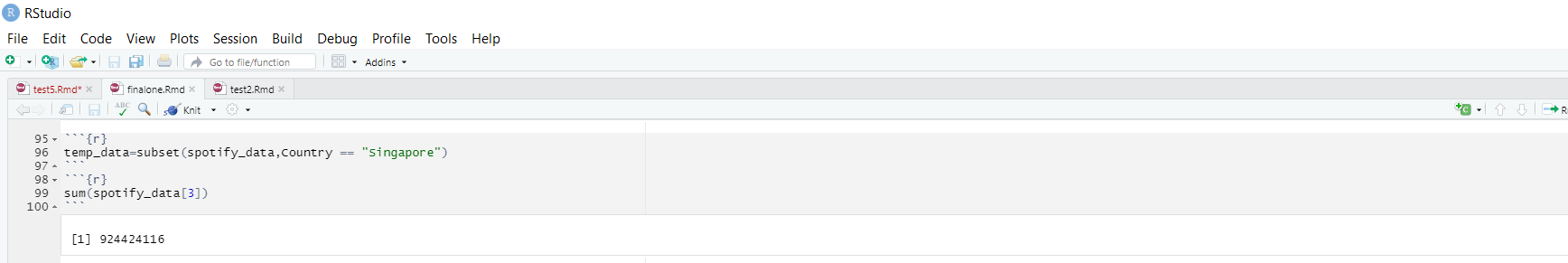
spotify\_data[,3]





22.Selecting data where spotify name is Singapore with subset operator

temp\_data=subset(spotify\_data,Country=="Singapore")



23.Renaming a column in data frame temp\_spotify=spotify\_datanames(temp\_data)[names(temp\_spotify)=="Total"]<-"Total\_Number") temp\_spotify[1,]

24.Adding a new column to dataframe temp\_spotify[["New\_col"]]<-rep(c(1,2,3,4,5),209) temp\_spotify[1:10,]

25.Display Sum of Popularity column sum(spotify\_data[3])

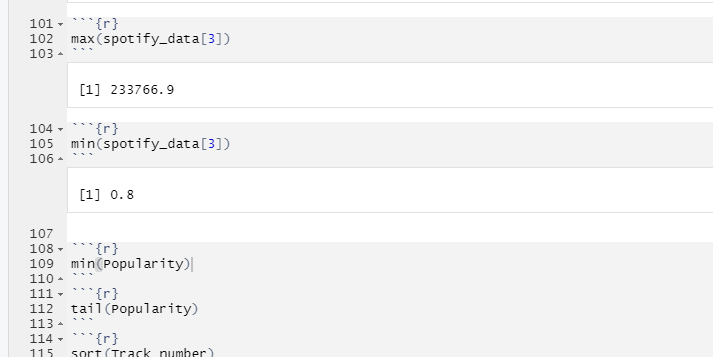


26.Display the maximum value of the Popularity column

max(spotify\_data[3])

27.Display the minimum value of the Popularity column

min(spotify\_data[3])



28.Attaching spotify data

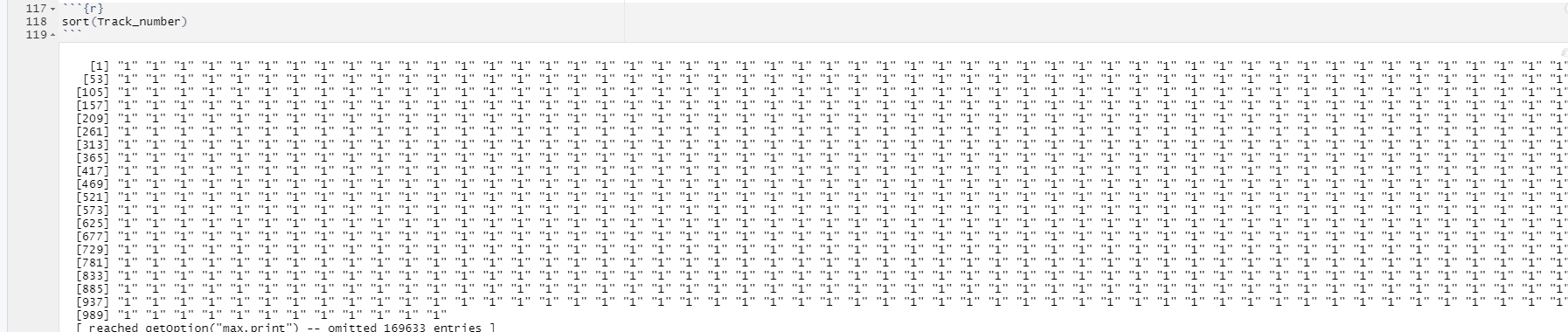
attach(spotify\_data)

29.Now we can use variables inside spotify data min(Popularity)

tail (Popularity)

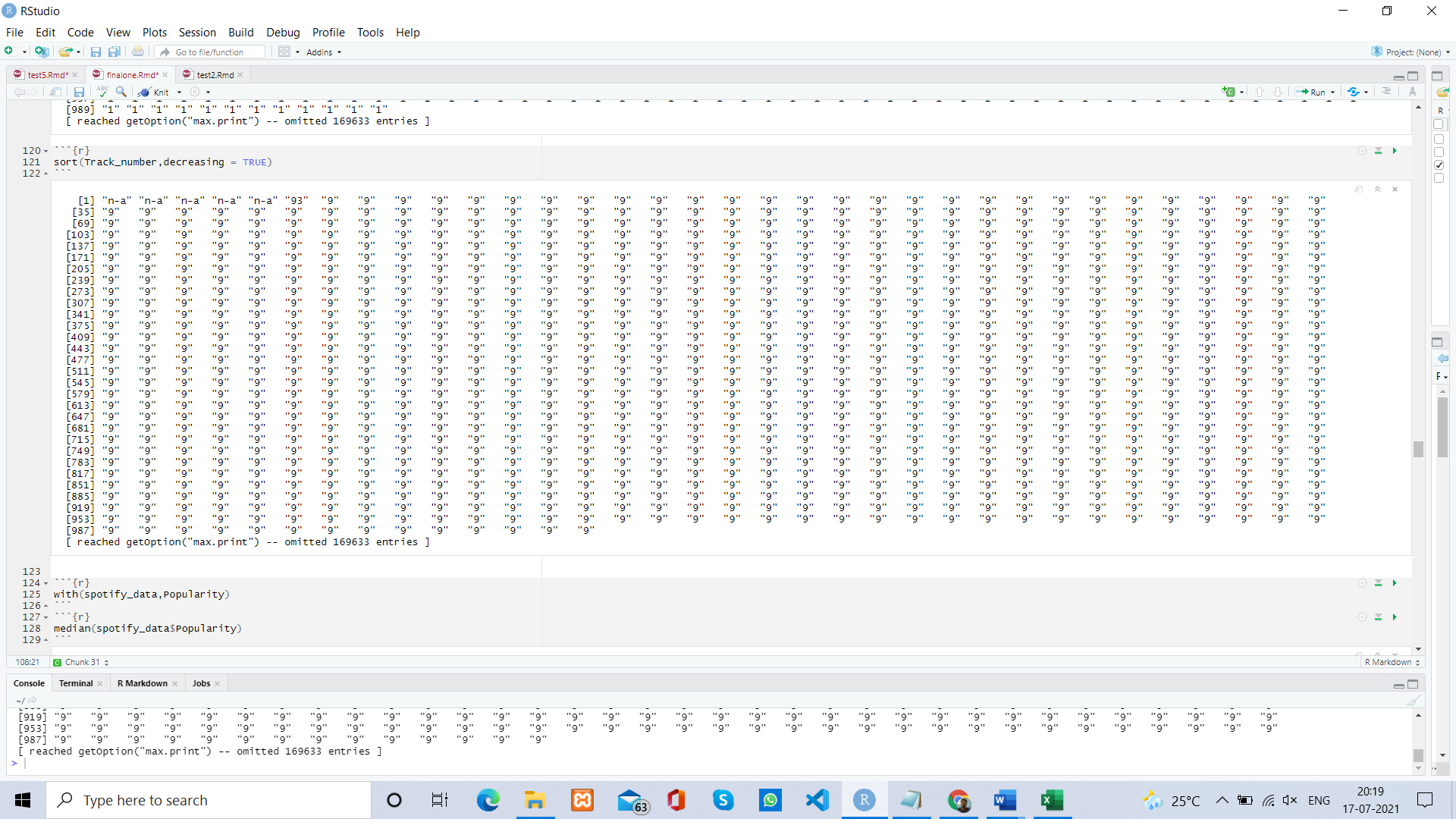
30.Sorting Track\_number variable in ascending order

sort(Track\_number)



31.Sorting Track number variable in descending order

sort(Track\_number,decreasing = TRUE)

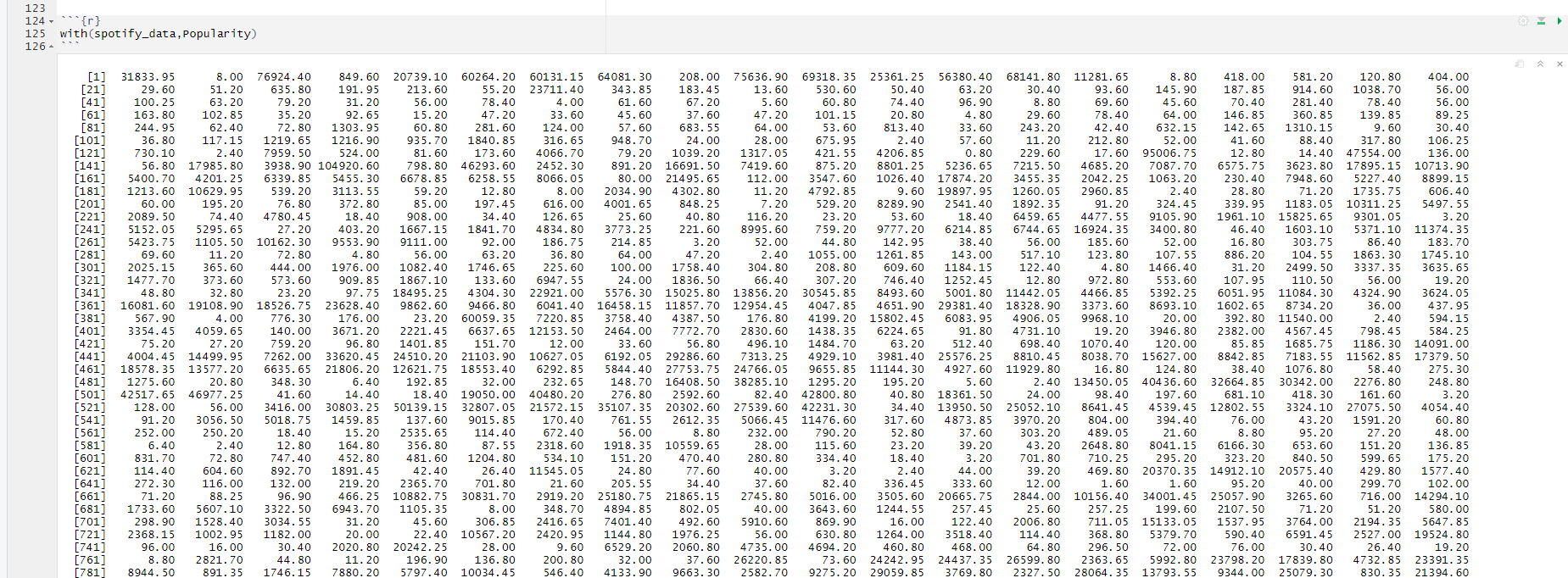


32.Detaching spotify data

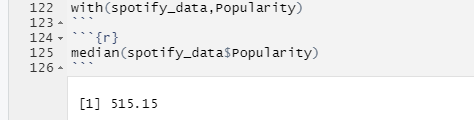
detach(spotify\_data)

33.Using with operator to use variables inside data

with(spotify\_data,Popularity)



34.Finding median of data median(spotify\_data$ Popularity)



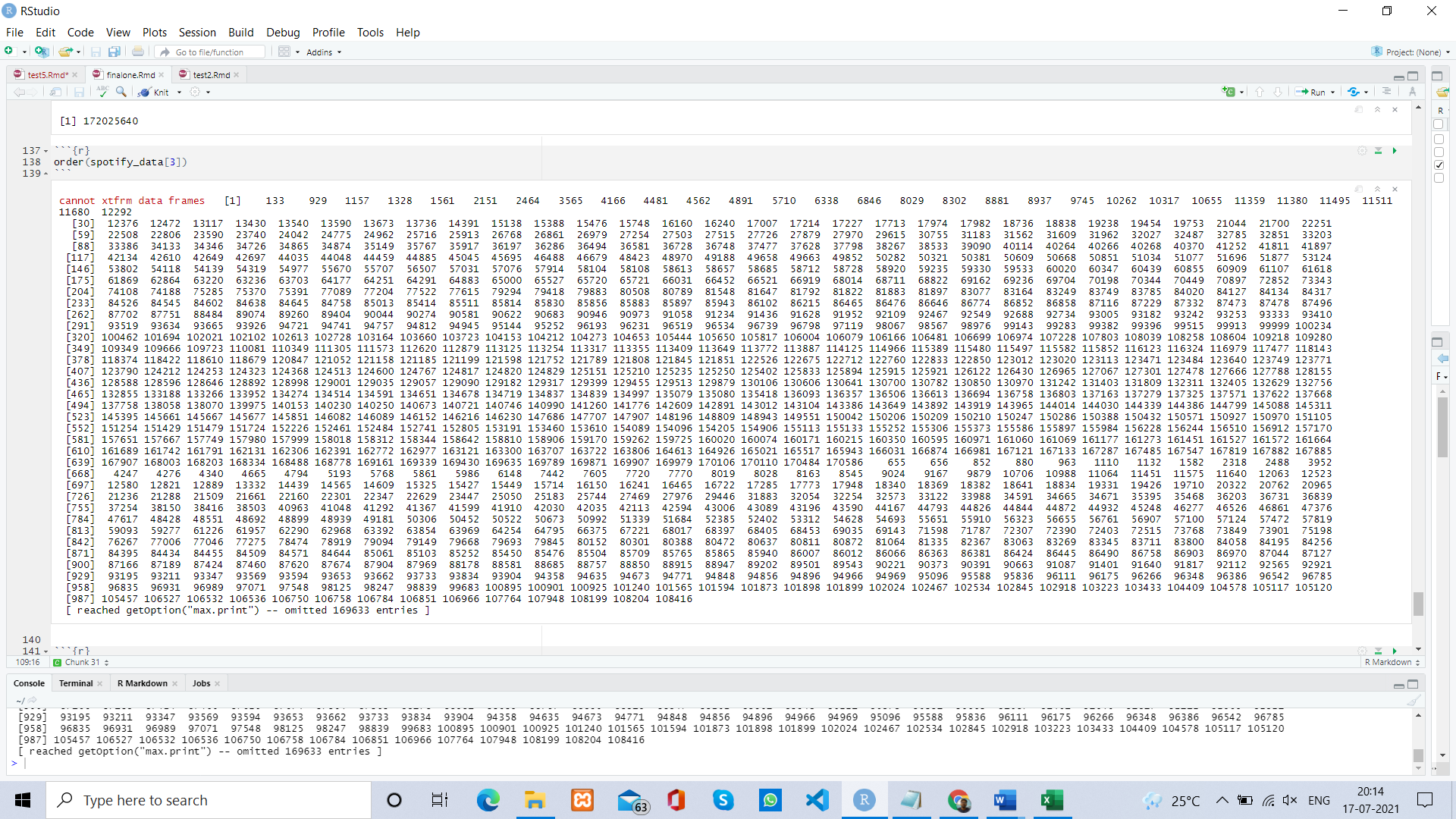
35.Finding mean of data mean(spotify\_data$Popularity)

36.Finding standard deviation of data sd(spotify\_data$ Popularity)

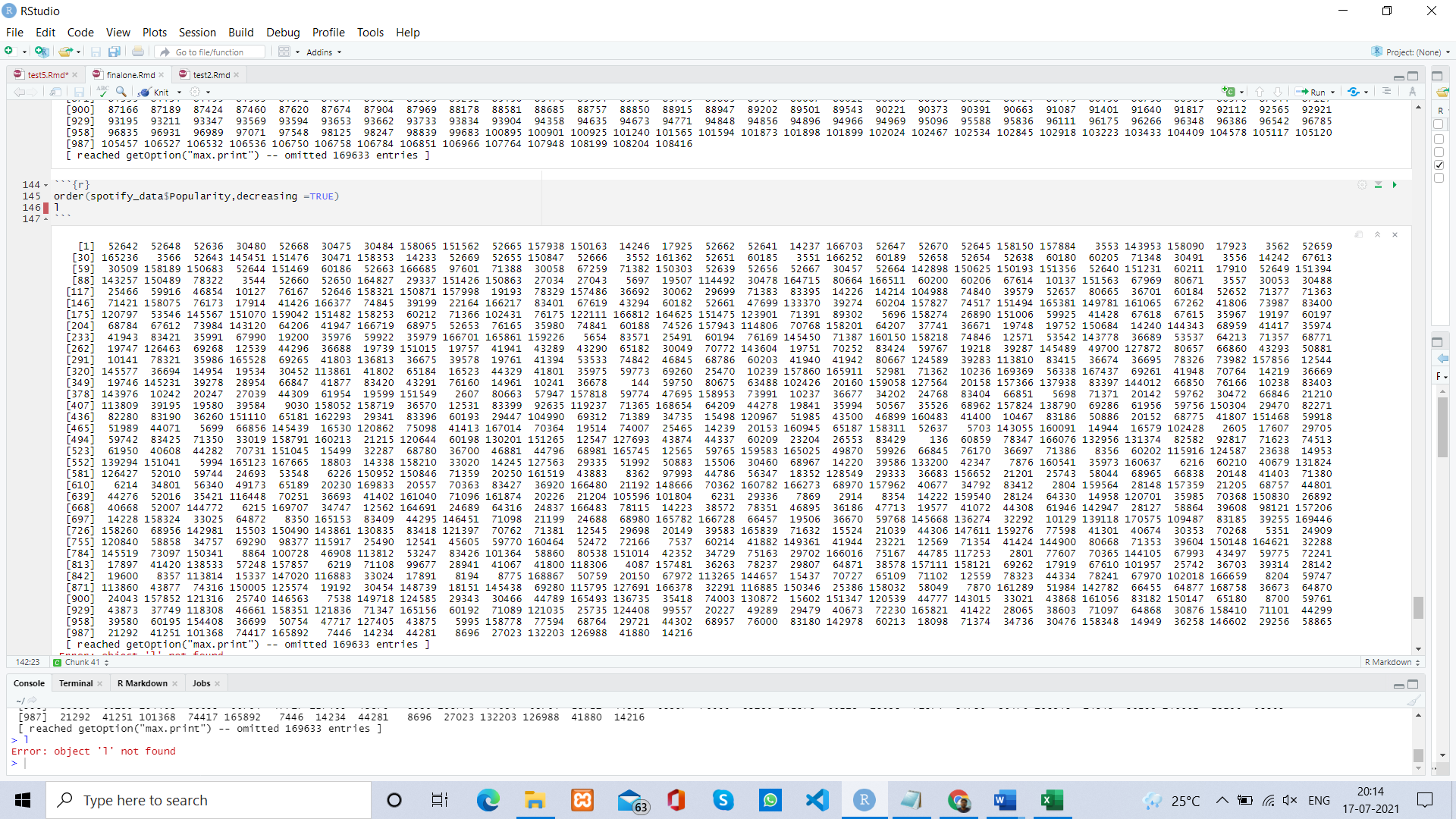
37.Finding variance of data var(spotify\_data$ Popularity)



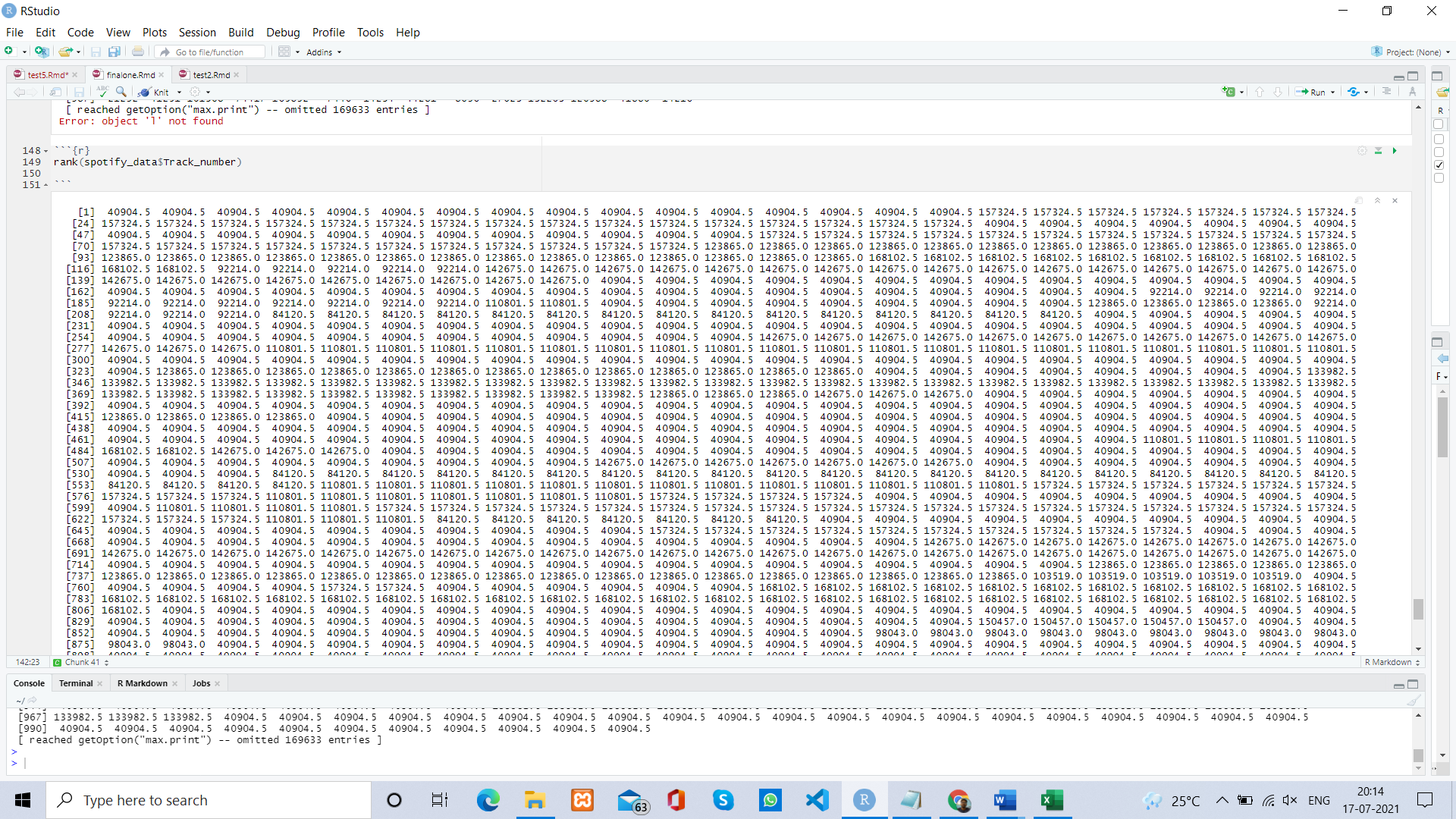
38.Order the Attack column in ascending order order(spotify\_data[3])



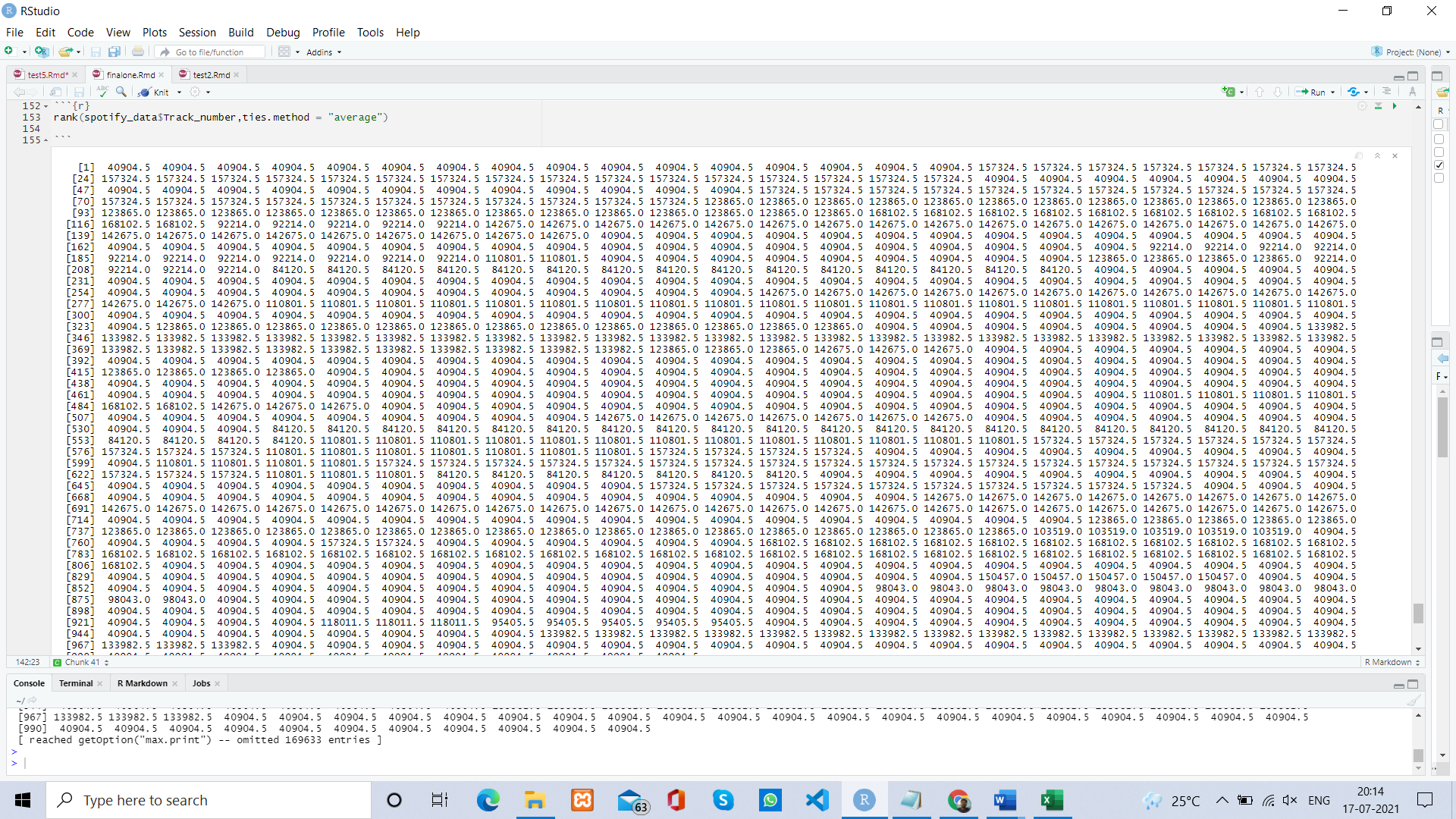
39.Order the Popularity column in descending order order(spotify\_data$Poplarity,decreasing =TRUE)



40.Rank of Speed column rank(spotify\_data$Track\_number)



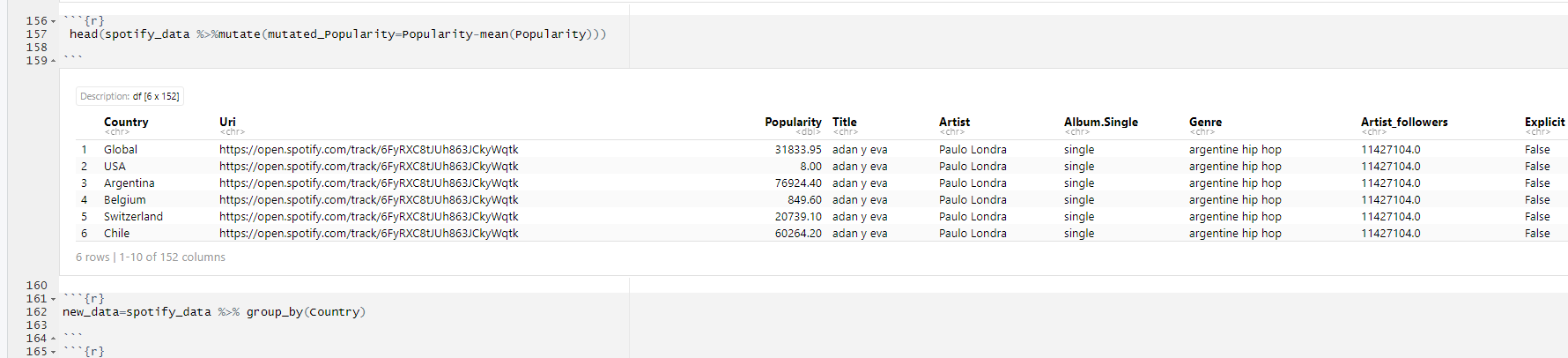
41.Rank of speed column with average as tie breaker rank(spotify\_data$ Track\_number,ties.method = "average")



DPLYR operations

42.Usage of mutate function library(dplyr) Attaching package: 'dplyr'

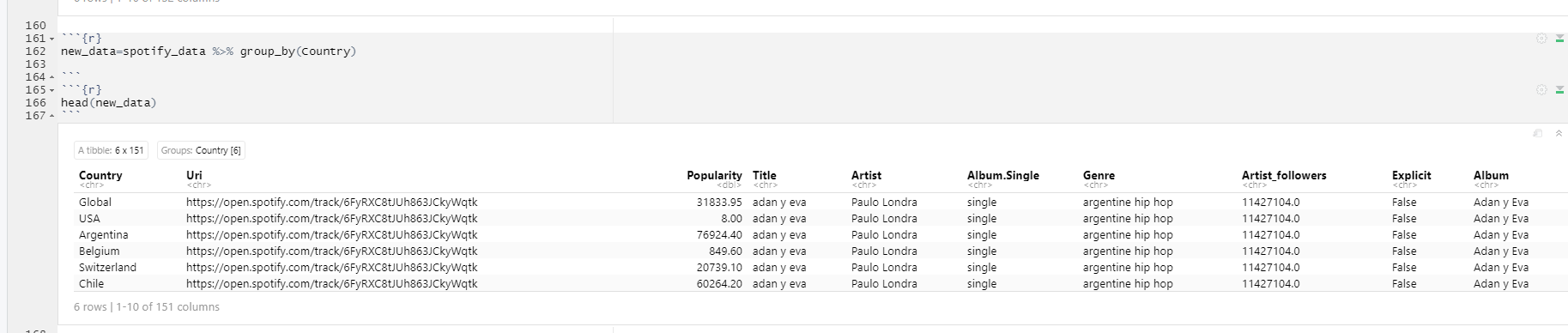
head(spotify\_data %>%mutate(mutated\_Popularity=Popularity-mean(Popularity)))



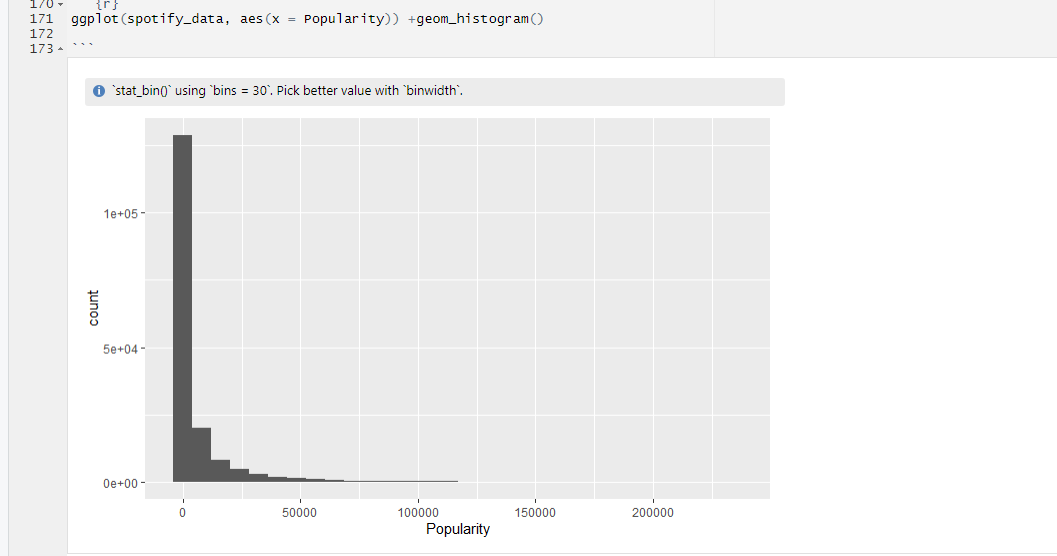
43.Adding extra column with user created vector vec = rep(c(1,2,3,4,5),209) head(spotify\_data %>% mutate(newcol = vec))

44.Group by function new\_data=spotify\_data %>% group\_by (Country)

head(new\_data)



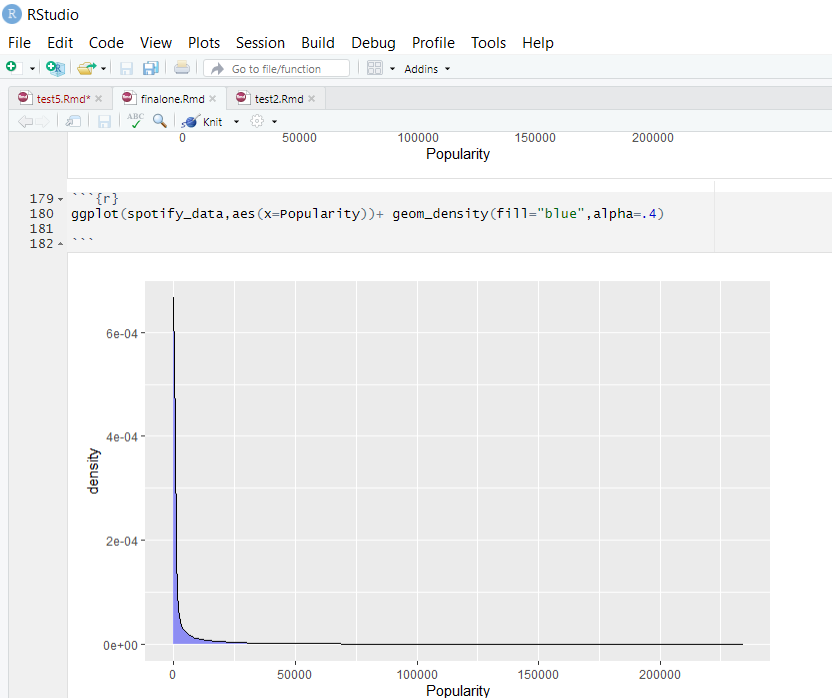
45.Histogram ggplot(spotify\_data, aes(x = Popularity)) +geom\_histogram()



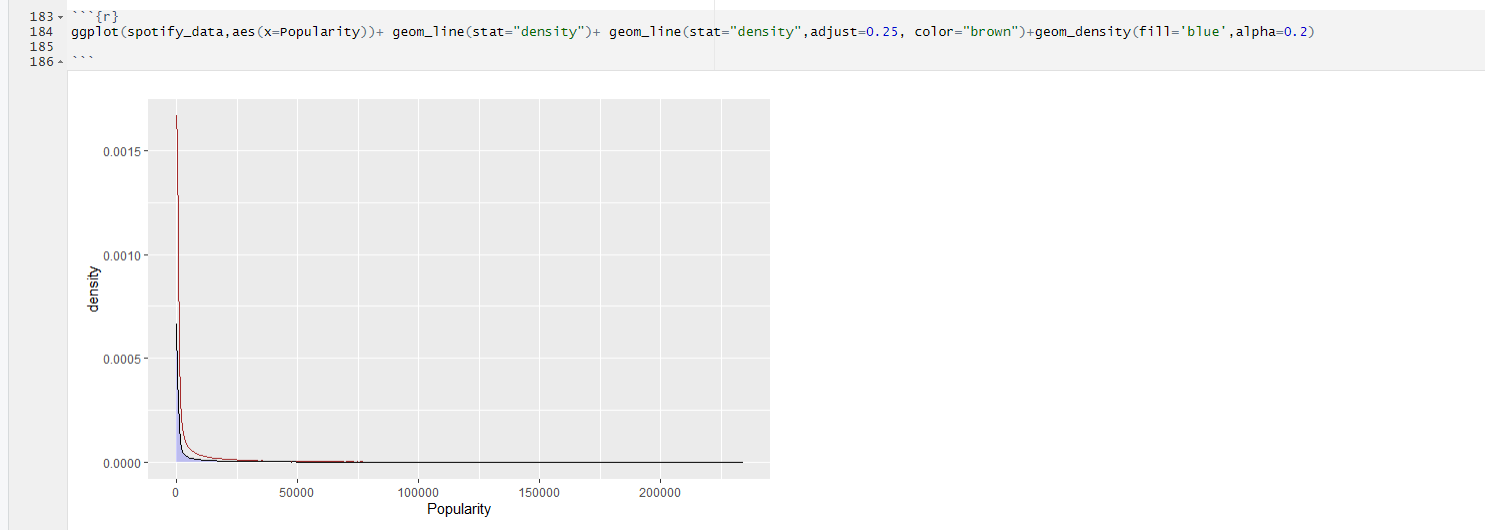
46.Histogram of Popularity column and its density ggplot(spotify\_data,aes(x=Popularity))+ geom\_histogram(fill="cornsilk",color="blue", size=0.2)+geom\_density(color="black")



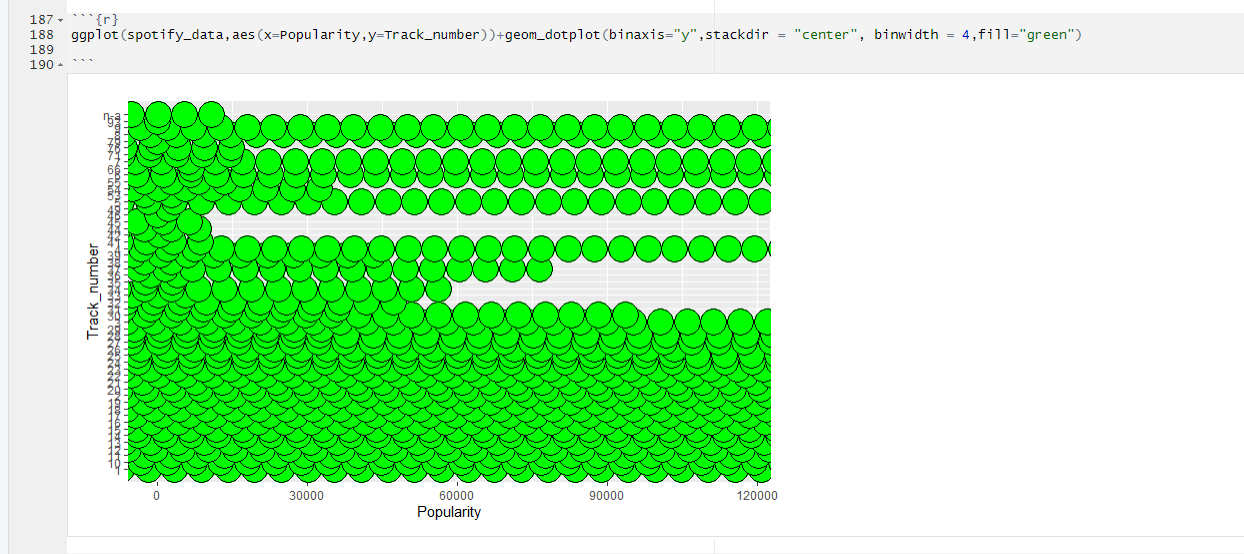
47.Line graph of Popularity column and its density ggplot(spotify\_data,aes(x=Popularity))+ geom\_density(fill="blue",alpha=.4)



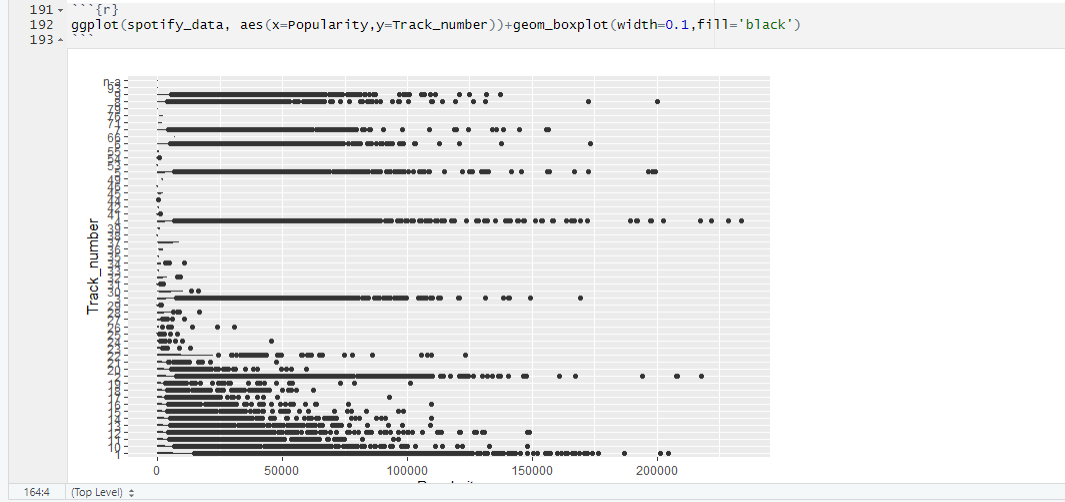
48.Line graph of Popularity column taking two alpha values ggplot(spotify\_data,aes(x=Popularity))+ geom\_line(stat="density")+ geom\_line(stat="density",adjust=0.25, color="brown")+geom\_density(fill='blue',alpha=0.2)



49.Dot Plot library(ggplot2) ggplot(spotify\_data,aes(x=Popularity,y=Track\_number))+geom\_dotplot(binaxis="y",stackdir = "center", binwidth = 4,fill="green")



50.Box Plot ggplot(spotify\_data, aes(x=Popularity,y=Track\_number))+geom\_boxplot(width=0.1,fill='black')



51.Violin Plot Attack and HP ggplot(spotify\_data,aes(x=Popularity,y=Track\_number))+geom\_violin()

