

youtube

Attulaya

2024-02-01



- **YouTube** is an American video-sharing website headquartered in San Bruno, California. The service was created by three former PayPal employees—Chad Hurley, Steve Chen, and Jawed Karim—in February 2005. Google bought the site in November 2006 for US\$1.65 billion; YouTube now operates as one of Google's subsidiaries.

Loading libraries

```
set.seed(123)
```

```
# Data manipulation
```

```
library(data.table)
```

```
library(dplyr)
```

```
library(DT)
```

```
# Time manipulation
```

```
library(lubridate)
```

```
# Visualization
```

```
library(ggplot2)

library(plotrix)

library(corrplot)

library(ggdendro)

library(ggrepel)

# WordCloud

library(wordcloud)

# Text manipulation

library(tidytext)

library(stringr)

library(tm)

library(sentimentr)

library(wordcloud)

library(RSentiment)
```

Reading and preparing data

```
gb <- tail(fread("GBvideos.csv",encoding = "UTF-8"),20000)

gb[, "Location":= "GB"]

fr <- tail(fread("FRvideos.csv",encoding = "UTF-8"),20000)

fr[, "Location":= "FR"]

ca <- tail(fread("CAvideos.csv",encoding = "UTF-8"),20000)

ca[, "Location":= "CA"]

us <- tail(fread("USvideos.csv",encoding = "UTF-8"),20000)

us[, "Location":= "US"]

de <- tail(fread("DEvideos.csv",encoding = "UTF-8"),20000)
```

```

de[, "Location":= "DE"]

videos <- as.data.table(rbind(gb, fr, ca, us, de))

videos$trending_date <- ydm(videos$trending_date)

videos$publish_time <- ymd(substr(videos$publish_time, start = 1, stop = 10))

videos$dif_days <- videos$trending_date - videos$publish_time

```

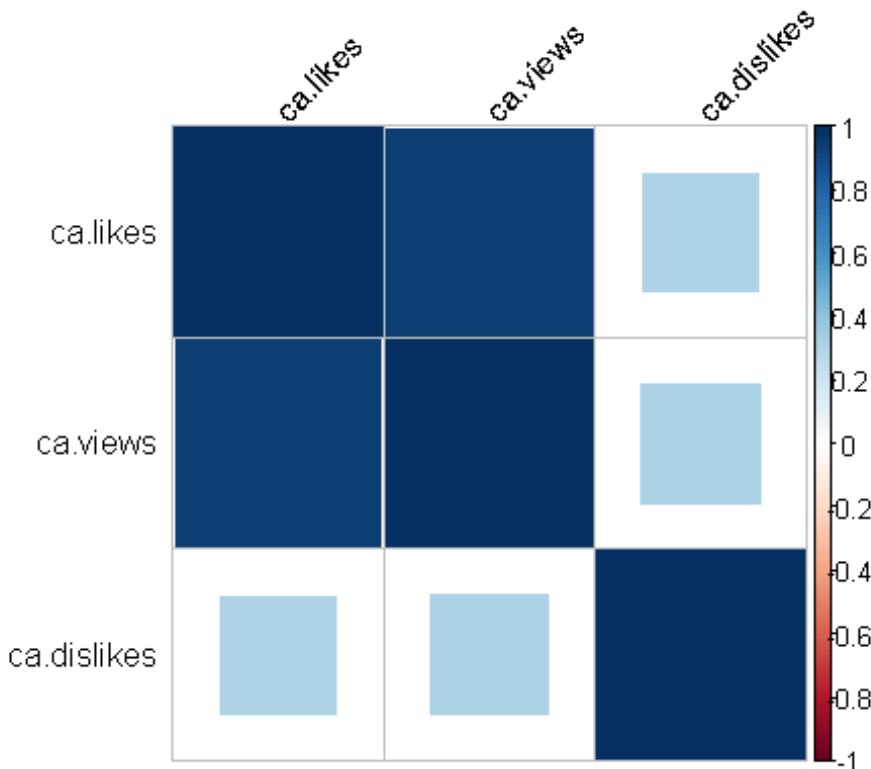
- We can see that between views and likes we have a high correlation, I thought that we will have a similar correlation between views and dislikes, but is almost half of the like correlation.
- lets check the correlation

```

cacorr <- data.frame(ca$likes, ca$views, ca$dislikes)
cacorr <- cor(cacorr)

corrplot(cacorr, method = "square", shade.col = NA, tl.col = "black", tl.srt = 45)

```



```

ca_head <- head(ca)
ca_head

```

```

##      video_id trending_date
## 1: n1WpP7iowlc 17.14.11
## 2: 0dBIkQ4Mz1M 17.14.11
## 3: 5qpjK5DgCt4 17.14.11
## 4: d380meD0W0M 17.14.11
## 5: 2Vv-BfVoq4g 17.14.11
## 6: 0yIWz1XEeyc 17.14.11
##
## title
## 1: Eminem - Walk On
Water (Audio) ft. Beyoncé
## 2: PLUSH -
Bad Unboxing Fan Mail
## 3: Racist Superman | Rudy Mancuso,
King Bach & Lele Pons
## 4: I
Dare You: GOING BALD!?
## 5: Ed Sheeran - Perfect
(Official Music Video)
## 6: Jake Paul Says Alissa Violet CHEATED with LOGAN PAUL! #DramaAlert Team
10 vs Martinez Twins!
##   channel_title category_id      publish_time
## 1:   EminemVEVO          10 2017-11-10 17:00:03
## 2:   iDubbzbzTV          23 2017-11-13 17:00:00
## 3:   Rudy Mancuso         23 2017-11-12 19:05:24
## 4:   nigahiga            24 2017-11-12 18:01:41
## 5:   Ed Sheeran           10 2017-11-09 11:04:14
## 6:   DramaAlert           25 2017-11-13 07:37:51
##
## tags
## 1: Eminem"|"Walk"|"On"|"Water"|"Aftermath/Shady/Interscope"|"Rap
## 2: plush"|"bad unboxing"|"unboxing"|"fan
mail"|"idubbzbztv"|"idubbzbztv2"|"things"|"best"|"packages"|"plushies"|"chonten
t chop
## 3: racist
superman"|"rudy"|"mancuso"|"king"|"bach"|"racist"|"superman"|"love"|"rudy
mancuso poo bear black white official music video"|"iphone x by
pineapple"|"lelepons"|"hannahstocking"|"rudymancuso"|"inanna"|"anwar"|"sarkis
"|"shots"|"shotsstudios"|"alesso"|"anitta"|"brazil"|"Getting My Driver's
License | Lele Pons
## 4:
ryan"|"higa"|"higatv"|"nigahiga"|"i dare you"|"idy"|"rhpc"|"dares"|"no
truth"|"comments"|"comedy"|"funny"|"stupid"|"fail
## 5:
edsheeran"|"ed
sheeran"|"acoustic"|"live"|"cover"|"official"|"remix"|"official
video"|"lyrics"|"session
## 6:

```

```

#DramaAlert" | "Drama" | "Alert" | "DramaAlert" | "keemstar" | "youtube news" | "jake
paul" | "team 10" | "alissa violet" | "cheated" | "logan paul" | "logan paul alissa
violet" | "jake paul alissa violet" | "Martinez Twins" | "left team 10" | "faze
banks" | "erika costell
##           views   likes dislikes comment_count
## 1: 17158579  787425    43420      125882
## 2: 1014651   127794     1688       13030
## 3: 3191434   146035     5339       8181
## 4: 2095828   132239     1989       17518
## 5: 33523622  1634130    21082      85067
## 6: 1309699   103755     4613       12143
##                                     thumbnail_link comments_disabled
## 1: https://i.ytimg.com/vi/n1WpP7iowLc/default.jpg          FALSE
## 2: https://i.ytimg.com/vi/0dBIkQ4Mz1M/default.jpg          FALSE
## 3: https://i.ytimg.com/vi/5qpjK5DgCt4/default.jpg          FALSE
## 4: https://i.ytimg.com/vi/d380meD0W0M/default.jpg          FALSE
## 5: https://i.ytimg.com/vi/2Vv-BfVoq4g/default.jpg          FALSE
## 6: https://i.ytimg.com/vi/0yIWz1XEeyc/default.jpg          FALSE
##           ratings_disabled video_error_or_removed
## 1:             FALSE          FALSE
## 2:             FALSE          FALSE
## 3:             FALSE          FALSE
## 4:             FALSE          FALSE
## 5:             FALSE          FALSE
## 6:             FALSE          FALSE
##
description
## 1:
Eminem's new track Walk on Water ft. Beyoncé is available everywhere:
http://shady.sr/WOWEminem \nPlaylist Best of Eminem:
https://goo.gl/AquNpo\nSubscribe for more: https://goo.gl/DxCrDV\n\nFor
more visit:
\nhttp://eminem.com\nhttp://facebook.com/eminem\nhttp://twitter.com/eminem
\nhttp://instagram.com/eminem\nhttp://eminem.tumblr.com\nhttp://shadyrecords.com\nhttp://facebook.com/shadyrecords\nhttp://twitter.com/shadyrecords
\nhttp://instagram.com/shadyrecords\nhttp://trustshady.tumblr.com\n\nMusic
video by Eminem performing Walk On Water. (C) 2017 Aftermath
Records\nhttp://vevo.ly/gA7xKt
## 2:
Still got a lot of packages. Probably will last for another year. On a side
note, more 2nd channel vids soon. editing with premiere from now on, gon' be
a tedious transition, but i think it's for the best. \n\n_\n\nSUBSCRIBE
► http://www.youtube.com/subscription\_center?add\_user=iDubbbztv\n\nMain
Channel ► https://www.youtube.com/user/iDubbbzTV\nSecond Channel ►
https://www.youtube.com/channel/UC-tsNNJ3yIW98MtPH6PWFAQ\nGaming Channel ►
https://www.youtube.com/channel/UCVhfFXNY0z3-mbrTh10YRXA\n\nWebsite ►
http://www.idubbbz.com\n\nInstagram ►
https://instagram.com/idubbbz\nTwitter ►
https://twitter.com/Idubbbz\nFacebook ►
http://www.facebook.com/IDubbbz\nTwitch ► http://www.twitch.tv/idubbbz\n_

```

```

## 3:
WATCH MY PREVIOUS VIDEO ► \\n\\nSUBSCRIBE ►
https://www.youtube.com/channel/UC5jkXpfnBh1Djqh0ir5FsIQ?sub\_confirmation=1\\n\\nTHANKS FOR WATCHING! LIKE & SUBSCRIBE FOR MORE VIDEOS!\\n-----\\n-----\\nFIND ME ON: \\nInstagram |
http://instagram.com/rudymancuso\\nTwitter |
http://twitter.com/rudymancuso\\nFacebook |
http://facebook.com/rudymancuso\\n\\nCAST: \\nRudy Mancuso |
http://youtube.com/c/rudymancuso\\nLele Pons |
http://youtube.com/c/lelepons\\nKing Bach |
https://youtube.com/user/BachelorsPadTv\\n\\nVideo Effects: \\nCaleb Natale |
https://instagram.com/calebnatale\\n\\nPA:\\nPaulina Gregory\\n\\n\\nShots Studios Channels:\\nAlesso | https://youtube.com/c/alesso\\nAnitta |
https://youtube.com/c/anitta\\nAnwar Jibawi |
https://youtube.com/c/anwar\\nAwkward Puppets |
https://youtube.com/c/awkwardpuppets\\nHannah Stocking |
https://youtube.com/c/hannahstocking\\nInanna Sarkis |
https://youtube.com/c/inanna\\nLele Pons |
https://youtube.com/c/lelepons\\nMaejor | https://youtube.com/c/maejor\\nMike Tyson | https://youtube.com/c/miketyson \\nRudy Mancuso |
https://youtube.com/c/rudymancuso\\nShots Studios |
https://youtube.com/c/shots\\n\\n#Rudy\\n#RudyMancuso
## 4:
I know it's been a while since we did this show, but we're back with what might be the best episode yet!\\nLeave your dares in the comment section!
\\n\\nOrder my book how to write good \\nhttp://higatv.com/ryan-higas-how-to-write-good-pre-order-links\\n\\nJust Launched New Official Store\\nhttps://www.gianthugs.com/collections/ryan\\n\\nHigaTV Channel\\nhttps://www.youtube.com/higatv\\n\\nTwitter\\nhttps://www.twitter.com/theralryanhiga\\n\\nFacebook\\nhttps://www.facebook.com/higatv\\n\\nWebsite\\nhttps://www.higatv.com\\n\\nInstagram\\nhttps://www.instagram.com/notryanhiga\\n\\nSend us mail or whatever you want here!\\nPO Box 232355\\nLas Vegas, NV 89105
## 5: 🎵 : https://ad.gt/yt-perfect\\n💰 : https://atlanti.cr/yt-album\\nSubscribe to Ed's channel:
http://bit.ly/SubscribeToEdSheeran\\n\\nFollow Ed on...\\nFacebook:
http://www.facebook.com/EdSheeranMusic\\nTwitter:
http://twitter.com/edsheeran\\nInstagram:
http://instagram.com/teddysphotos\\nOfficial Website:
http://edsheeran.com\\n\\nDirector: Jason Koenig\\nProducer: Honna Kimmerer\\nStarring: Ed Sheeran & Zoey Deutch \\nDirector of Photography: Johnny Valencia\\nProduction Company: Anonymous Content\\nExec Producer: Nina Soriano\\nProduction Manager: Doug Hoff\\nCommissioner: Dan Curwin\\nProduction Designer: John Lavin\\nLead Casting: Amy Hubbard \\n\\nWritten by: Jason Koenig, Ed Sheeran, Andrew Kolvet, Jenny Koenig, Murray Cummings\\n\\nEdited by: Jason Koenig & Johnny Valencia\\nVFX: Ian Hubert\\n\\nCast: Bo Valencia, Dennis Ranalta, Arthur Pauli\\n \\nSki Cinematography: Corey Koniniec\\nSpecialty Camera op: Ryan Haug\\n1st AC: Ryan Brown\\n \\n1st Assistant Director: Ole Zapatka\\nArt Director: Klaus Hartl\\nSnow fx: Lucien Stephenson\\n\\nGaffer: Thomas Berz\\nStylist:
```

```

Claudia Lajda\\nHair & Makeup: Christel Thoresen\\nAustrian Casting: Ursula
Kiplinger\\n \\nAdditional VFX: Zoic\\n\\nSpecial Thanks to: The Hintertux
Glacier, Austria;\\nThe Tenne, and Hotel Neuhintertux
## 6:
► Follow for News! - https://twitter.com/KEEMSTAR\\n\\n► Also follow
#DramaAlert on:\\n★ Instagram: https://instagram.com/DramaAlert\\n★ Twitter:
https://twitter.com/DramaAlert\\n★ Facebook:
https://facebook.com/DramaAlert\\n\\n► Follow for livestreams! -
https://twitch.tv/KEEMSTAR\\n\\n► KEEM Merch\\nhttp://keem.shirtz.cool\\n\\n►
USE CODE (KEEM)\\nhttps://gfuel.com/pages/keemstar\\n\\nDollar in the Woods!
(OUT NOW)\\n► iTunes\\nhttps://itunes.apple.com/us/album/dollar-in-the-woods-single/id1295414119\\n\\n► Spotify
\\nhttps://open.spotify.com/track/3uUHoKWqPbJ5qoREGbguC9?si=v4CgSBBR\\n\\n►
YouTube (Music Video)\\nhttps://youtu.be/n38Qxi7TVWo\\n\\nAdpocalypse! (My
New Game)\\n► Apple (iOS)\\nhttps://itunes.apple.com/us/app/the-adpocalypse/id1263621591\\n\\n►
Android\\nhttps://play.google.com/store/apps/details?id=com.projectorgames.howtogetahead
##     Location
## 1:      CA
## 2:      CA
## 3:      CA
## 4:      CA
## 5:      CA
## 6:      CA

```

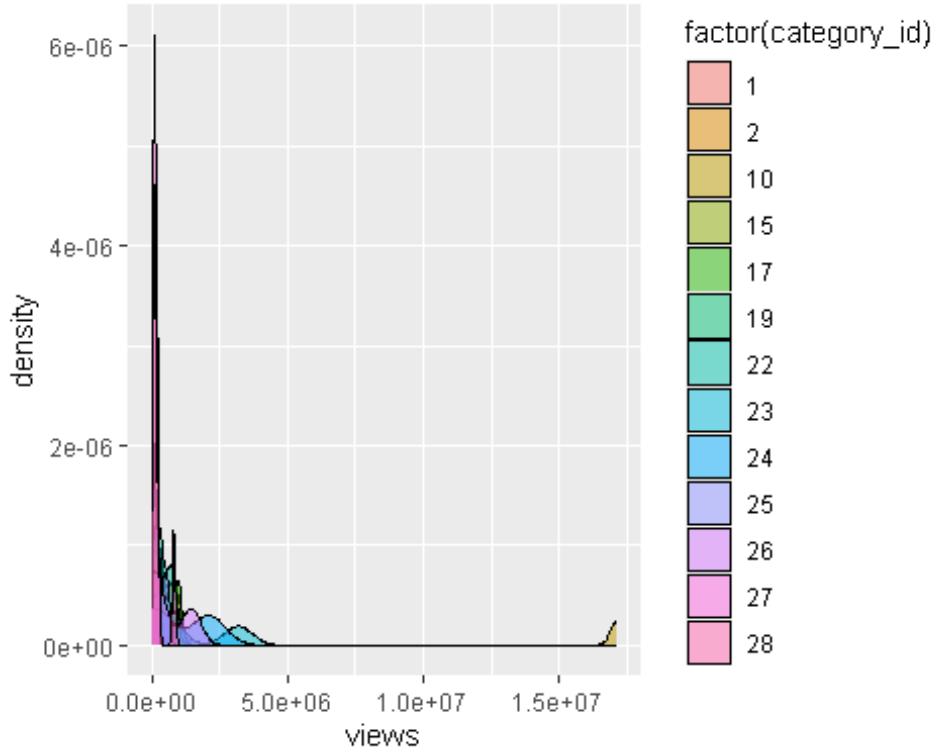
density graph for USA

```

library(ggplot2)

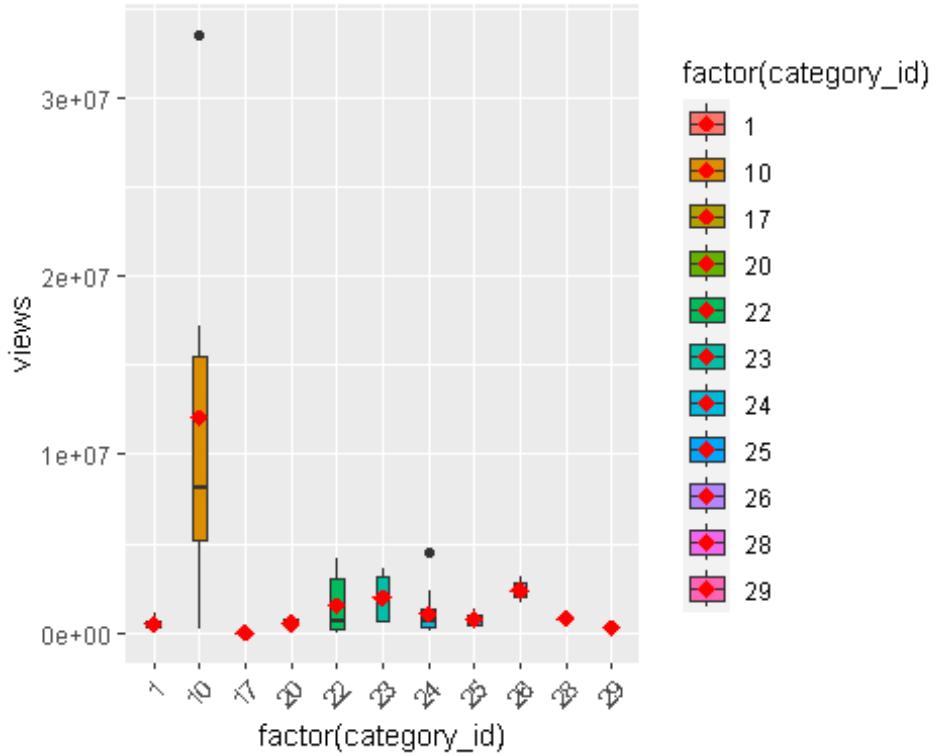
ggplot(us, aes(x = views, fill = factor(category_id))) +
  geom_density(alpha = 0.5)

```



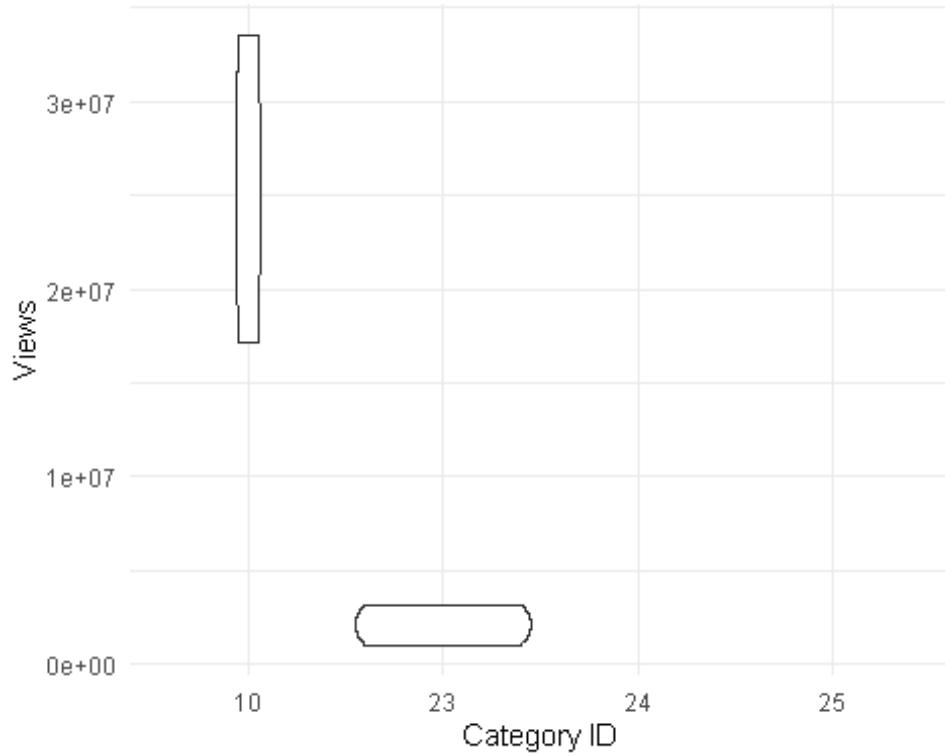
Box Plot for Canada

```
ggplot(head(ca,50), aes(x = factor(category_id), y = views, group = factor(category_id), fill=factor(category_id))) +
  geom_boxplot(width = 0.3) + # Adjust the width as needed
  stat_summary(fun = "mean", geom = "point", shape = 18, size = 3, color = "red") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Adjust x-axis
Label rotation if needed
```



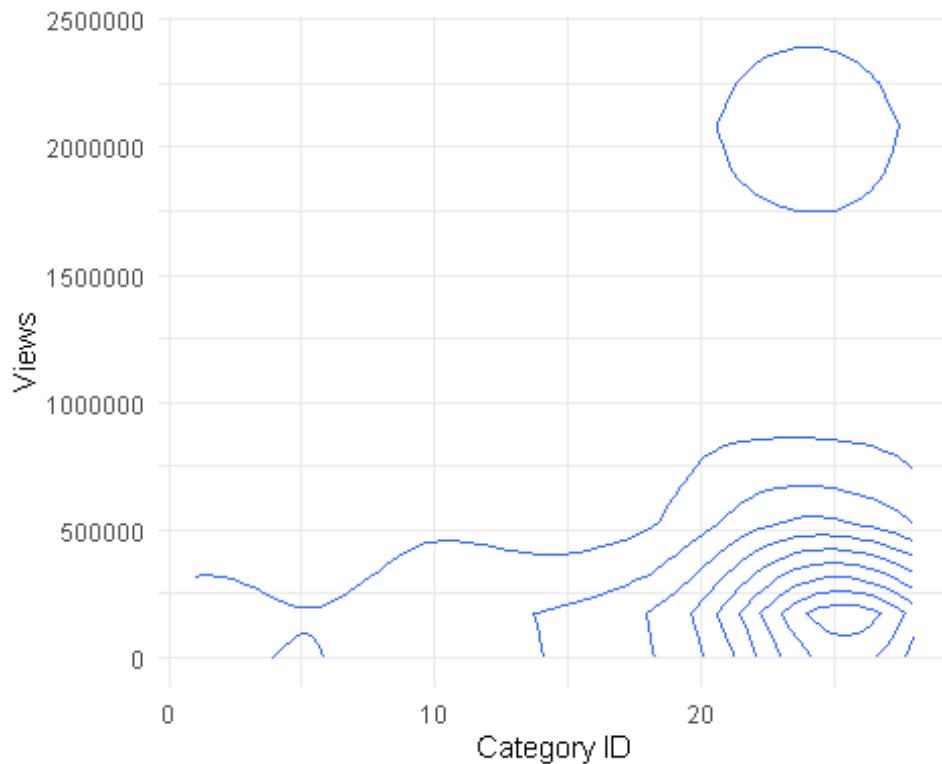
smoothed Violin Graph

```
ggplot(head(ca), aes(x = factor(category_id), y = views)) +  
  geom_violin(adjust=2) +  
  labs(x = "Category ID", y = "Views") +  
  theme_minimal()
```



density graph for US

```
ggplot(us, aes(x = category_id, y = views)) +  
  geom_density_2d() +  
  labs(x = "Category ID", y = "Views") +  
  theme_minimal()
```

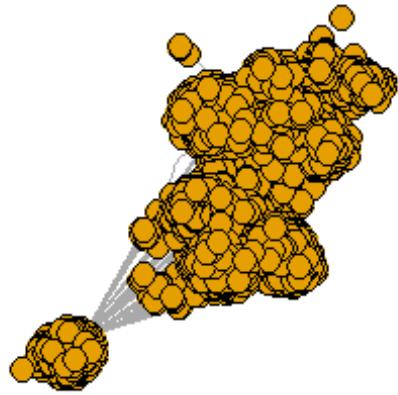


```
library(igraph)

set.seed(42) # for reproducibility

# Create a graph
graph <- graph.data.frame(videos[, c("category_id", "views", "likes")],
                           directed = FALSE)

# Plot the network graph
plot(graph,
      layout = layout_with_fr(graph), # Use Fruchterman-Reingold Layout
      vertex.label = NA, # Display category_id as node Labels
      edge.label = E(graph)$views) # Display views as edge labels
```

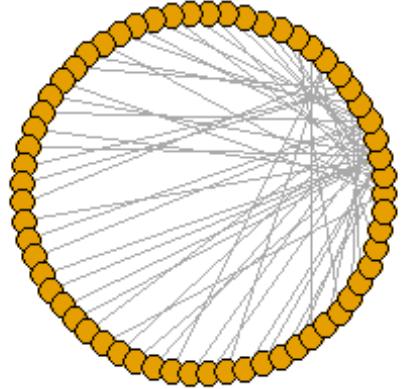


network graph for head(ca,50)

```
set.seed(42) # for reproducibility

# Create a graph
graph <- graph.data.frame(head(ca,50)[, c("category_id", "views", "likes")],
                           directed = FALSE)

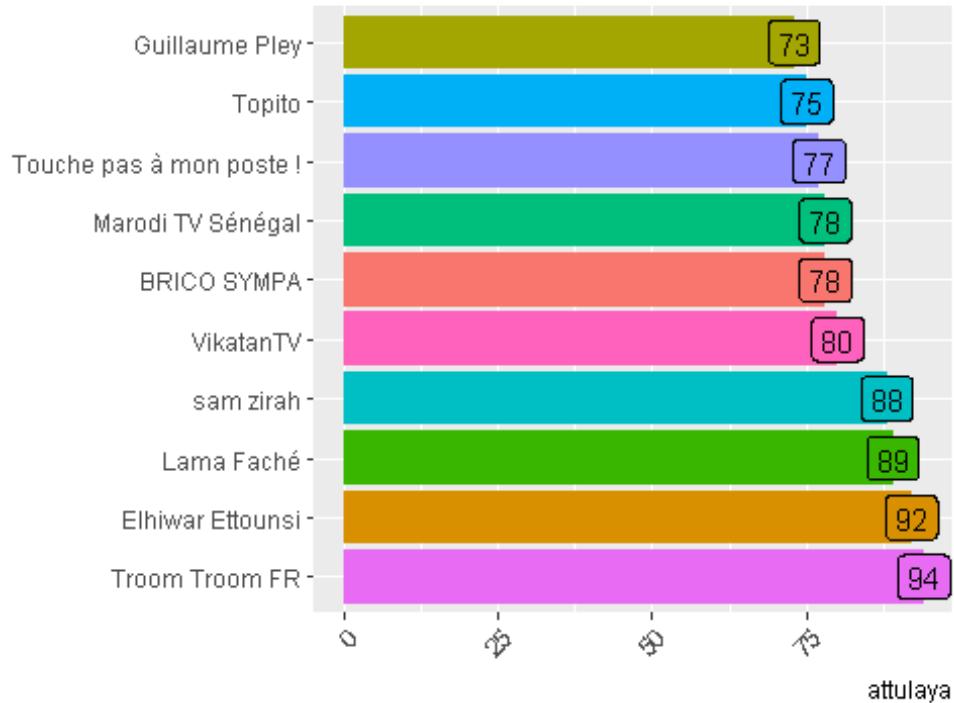
# Plot the network graph
plot(graph,
      layout = layout.circle(graph), # Use Fruchterman-Reingold Layout
      vertex.label = NA, # Display category_id as node Labels
      edge.label = E(graph)$views) # Display views as edge Labels
```



Top trending Channels in all countries

```
ggplot(videos[, .N, by=channel_title][order(-N)][1:10], aes(reorder(channel_title, -N), N, fill=channel_title))+geom_bar(stat="identity")+geom_label(aes(label=N))+guides(fill="none") + theme(axis.text.x = element_text(angle = 45, hjust = 1)) + labs(caption="attulaya", title=" Top trending channel titles in all countries") + xlab(NULL)+ylab(NULL)+coord_flip()
```

Top trending channel titles in all countries



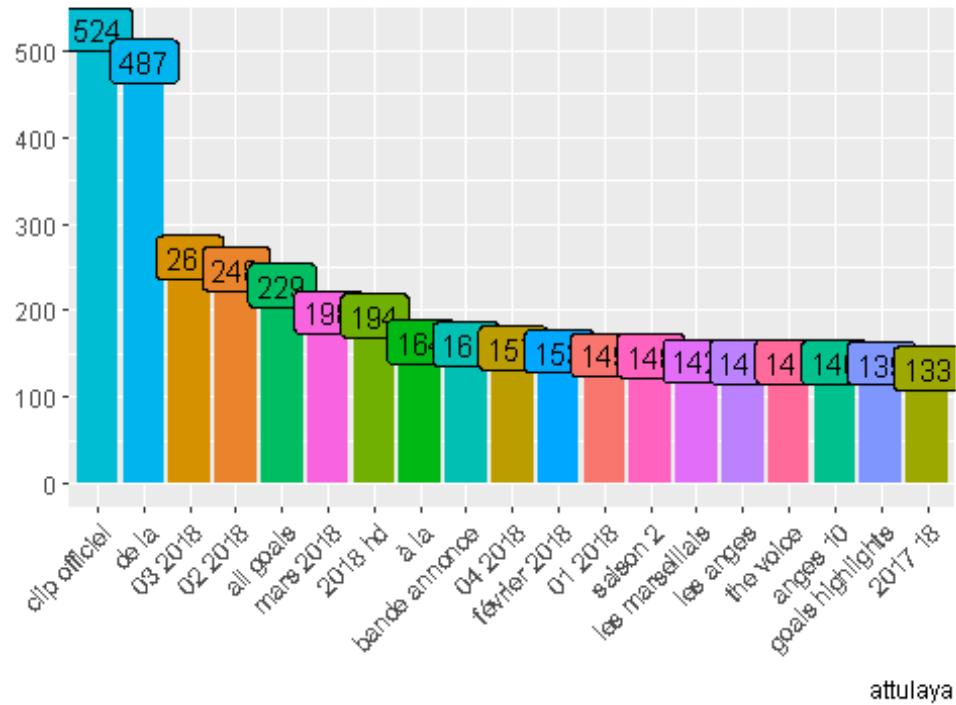
Title Bigrams

```
biga <- unnest_tokens(videos,bigram, title, token = "ngrams", n = 2)

biga <- as.data.table(biga)

ggplot(biga[,.N,by=bigram][order(-N)][1:19],aes(reorder(bigram,-N),N,fill=bigram))+geom_bar(stat="identity")+geom_label(aes(label=N))+guides(fill="none") +theme(axis.text.x = element_text(angle = 45,hjust = 1))+ labs(caption="attulaya",title="Top Title bigrams") +xlab(NULL)+ylab(NULL)
```

Top Title bigrams

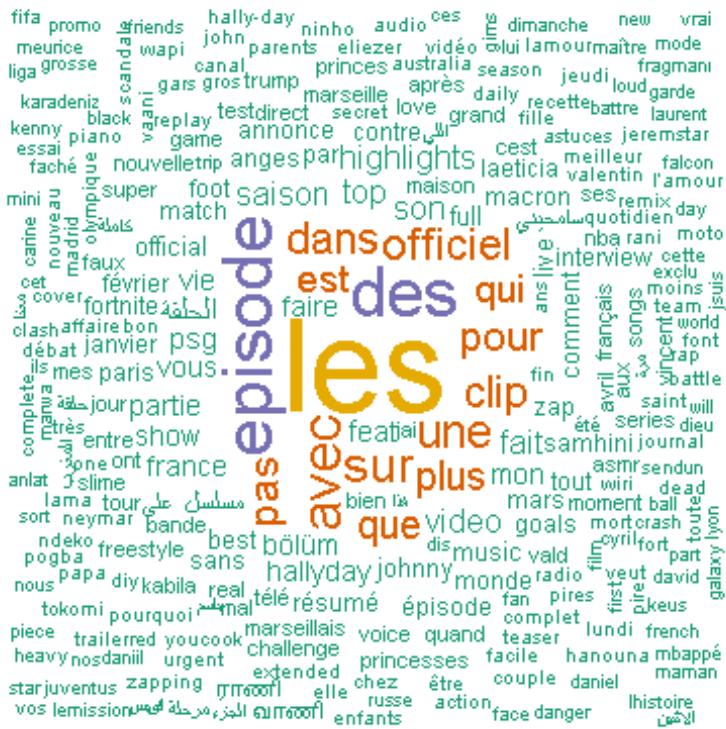


attulaya

- There are mainly bigrams related to *music*.

Title wordcloud

```
wordcloud(sent_video$text,sent_video$freq,  
min.freq=5,colors=brewer.pal(6,"Dark2"),random.order = F)
```



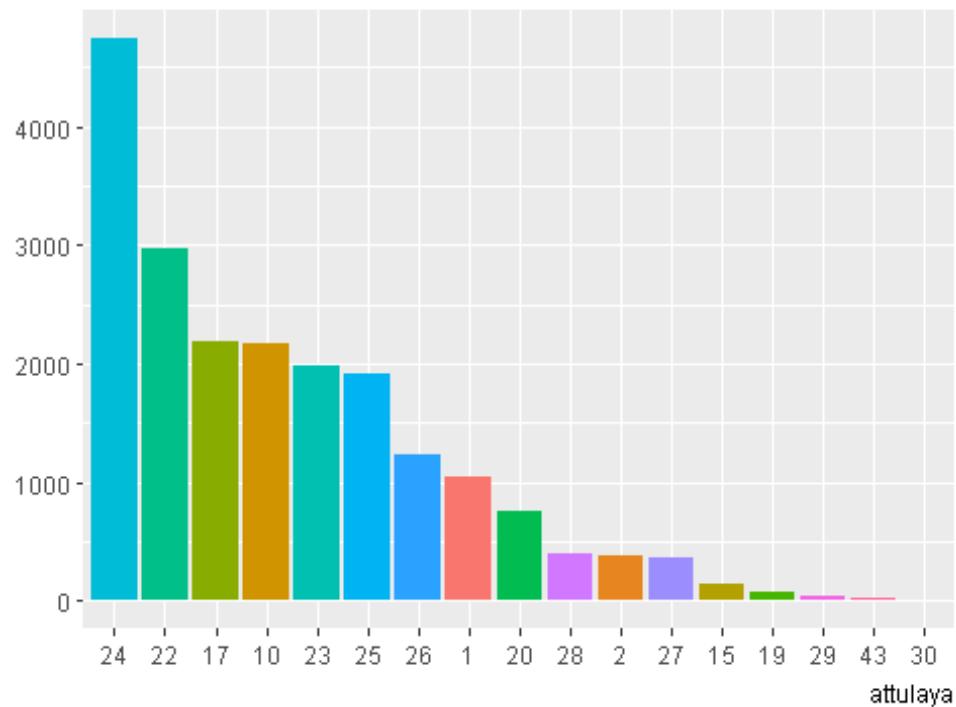
- We can see that a lot of the trending videos are music videos.

Top Category ID

```

ggplot(videos[,.N,by=category_id][order(-N)],aes(reorder(category_id,-N),N,fill=as.factor(category_id)))+geom_bar(stat="identity")+guides(fill="none")+labs(caption="attulaya",title=" Top Category ID")+
  xlab(NULL)+ylab(NULL)
  
```

Top Category ID

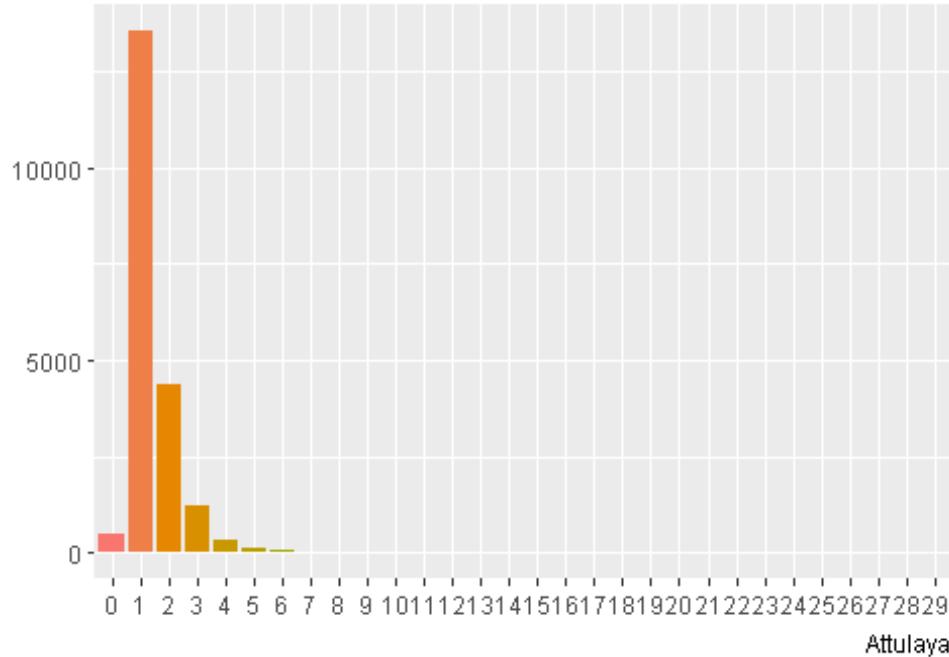


How much time passes between published and trending?

```
ggplot(videos[dif_days<30],aes(as.factor(dif_days),fill=as.factor(dif_days)))
+geom_bar()+guides(fill="none")+
  labs(caption="Attulaya",title=" Time between
published and trending",subtitle="In days")+
  xlab(NULL)+ylab(NULL)
```

Time between published and trending

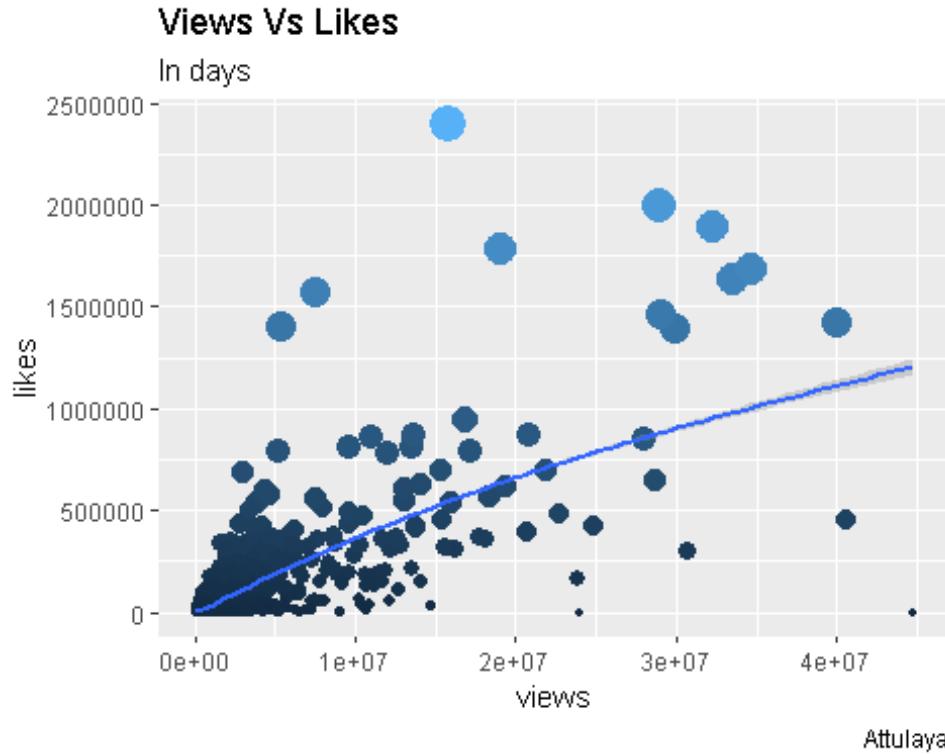
In days



- It seems that the videos never trend in the same day it is published.
- [none] is displayed if there are no tags, after none we can see tags as *new*, *iphone*, *episode* and tags related to *music*.

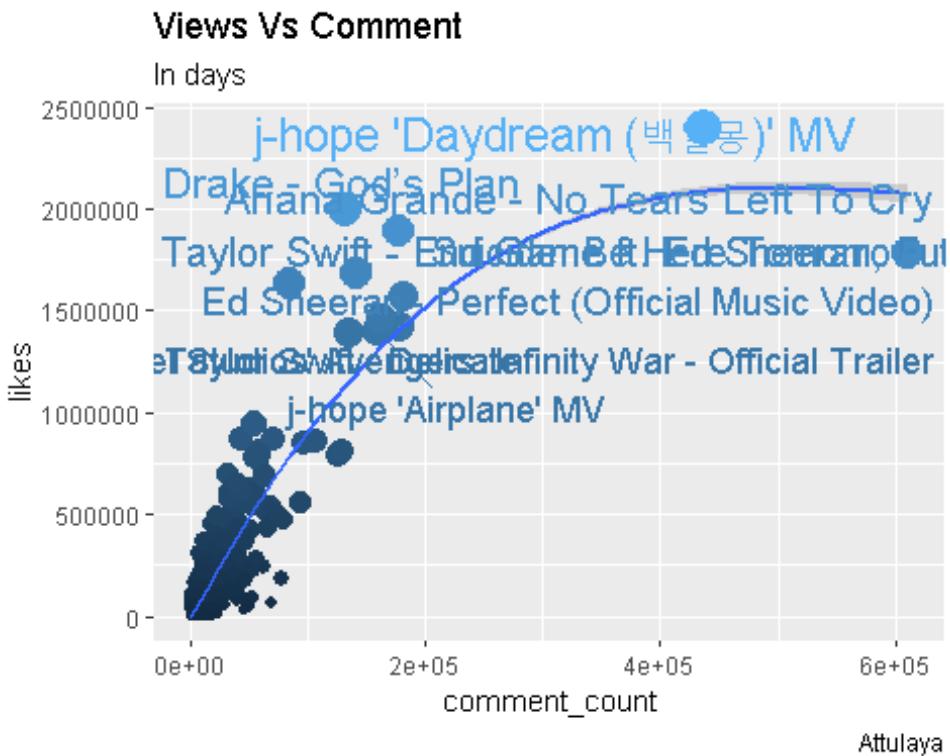
Views Vs Likes

```
ggplot(videos[.,("views"=max(views),"likes"=max(likes)),by=title],aes(views,likes,colour=likes,size=likes))+geom_jitter()+geom_smooth()+guides(fill="none")+labs(caption="Attulaya",title="Views Vs Likes",subtitle="In days") +theme(legend.position = "none") +geom_text_repel(data=subset(videos[.,("views"=max(views),"likes"=max(likes)),by=title], views > 5e+07), aes(views,likes,label=title),check_overlap=T)
```



Likes Vs Comment

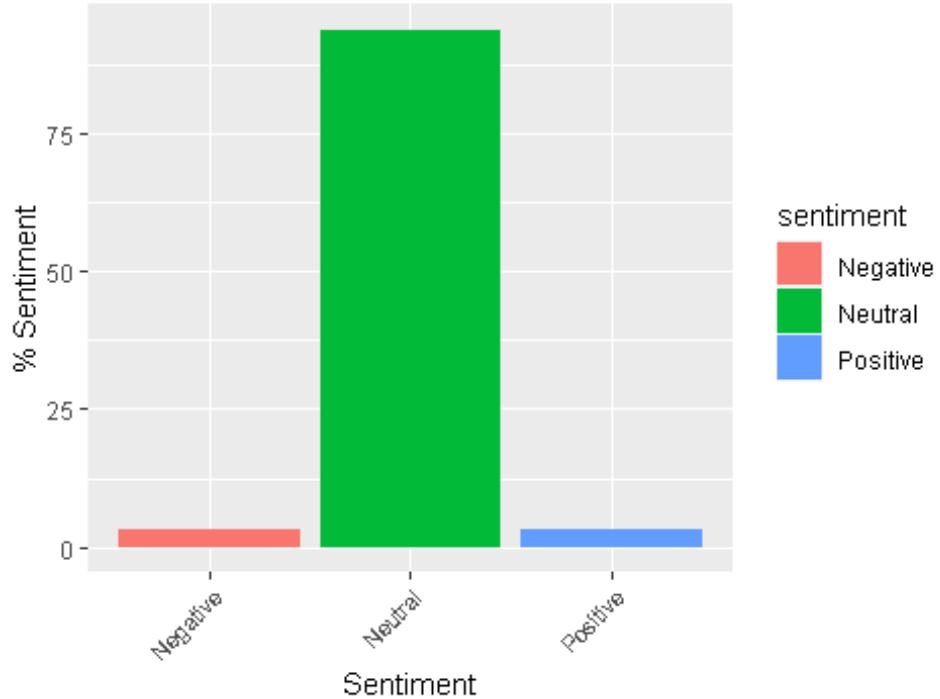
```
ggplot(videos[.,("comment_count"=max(comment_count),"likes"=max(likes)),by=title],aes(comment_count,likes,colour=likes,size=likes))+geom_jitter()+geom_smooth()+guides(fill="none")+labs(caption="Attulaya",title="Views Vs Comment",subtitle="In days")+
  theme(legend.position =
"none")+geom_text_repel(data=subset(videos[.,("comment_count"=max(comment_count),"likes"=max(likes)),by=title], likes > 1e+06),
  aes(comment_count,likes,label=title),check_overlap=T)
```



Sentiment Analysis Description field (Sample)

```
ggplot(sentiments1,aes(x = sentiment,y = Percentage ,fill=sentiment ))+
  geom_bar(stat = "identity") +
  ggtitle("Description Sentiments (Sample)")+xlab("Sentiment")+ylab("%
  Sentiment")+
  theme(axis.text.x = element_text(angle = 45, size=8,hjust = 1))
```

Description Sentiments (Sample)



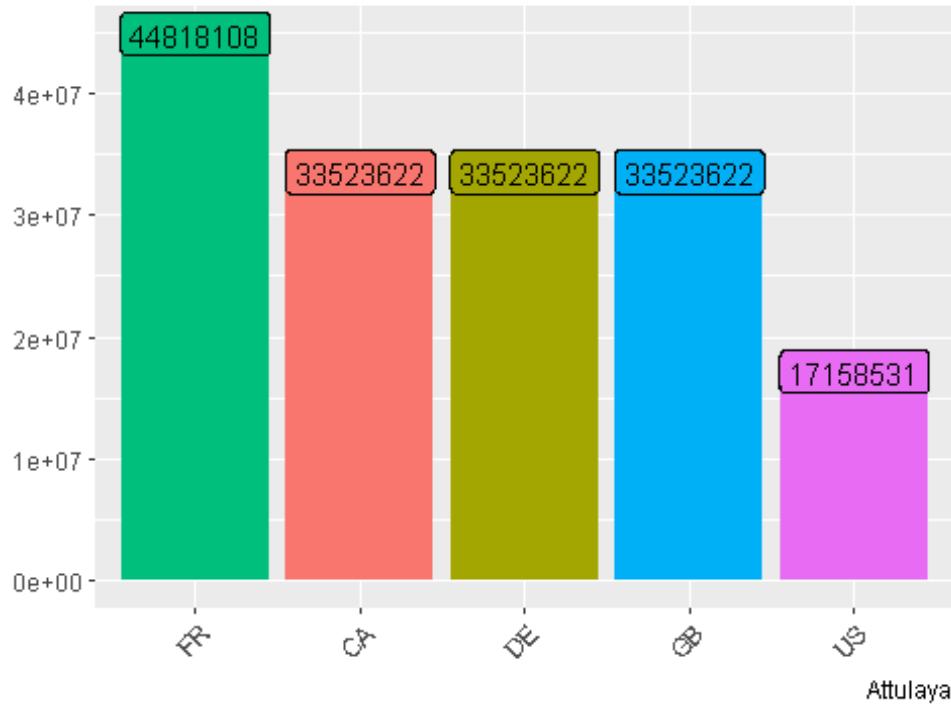
- Here we can see that the sentiments in the description field are basically neutral.
- We can see that the video description is clearly more positive than negative

Top Countries in Absolute numbers

In total number of views

```
ggplot(videos[,"Total_VIEWS"=max(videos)],by=Location],aes(reorder(Location,  
-  
Total_VIEWS),Total_VIEWS,fill=Location))+geom_bar(stat="identity")+geom_label(aes(label=Total_VIEWS))+guides(fill="none")+theme(axis.text.x =  
element_text(angle = 45,hjust = 1))+ labs(caption="Attulaya",title=" Total  
Views by Countries")+xlab(NULL)+ylab(NULL)
```

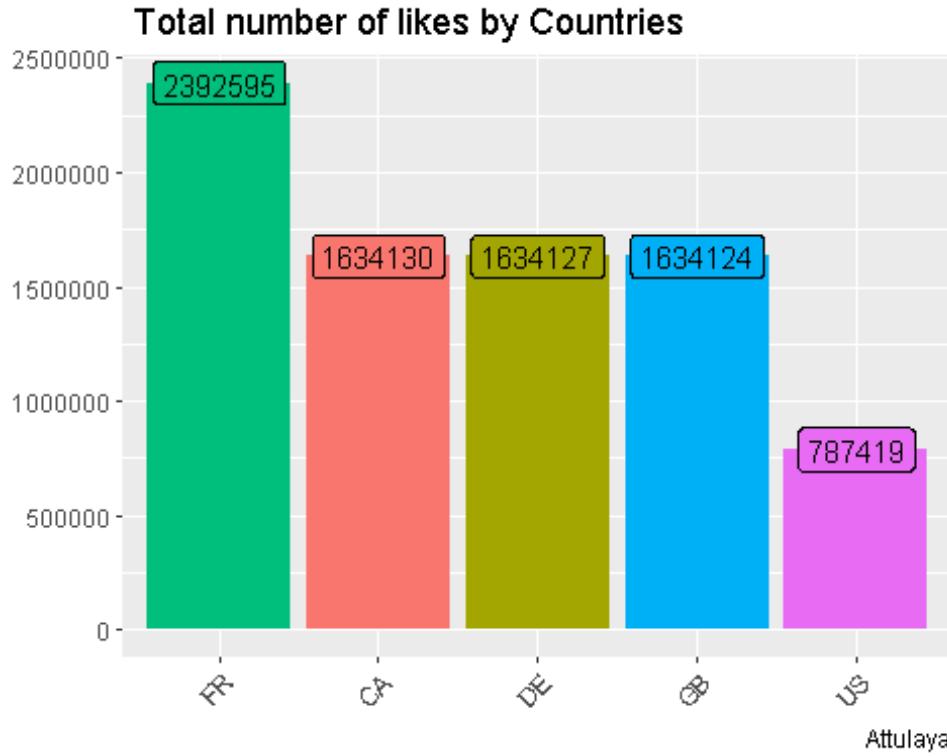
Total Views by Countries



- GB is the Country with most viewed videos in the trending field with significative difference with the other countries, almost doubled the second country.

In total number of likes

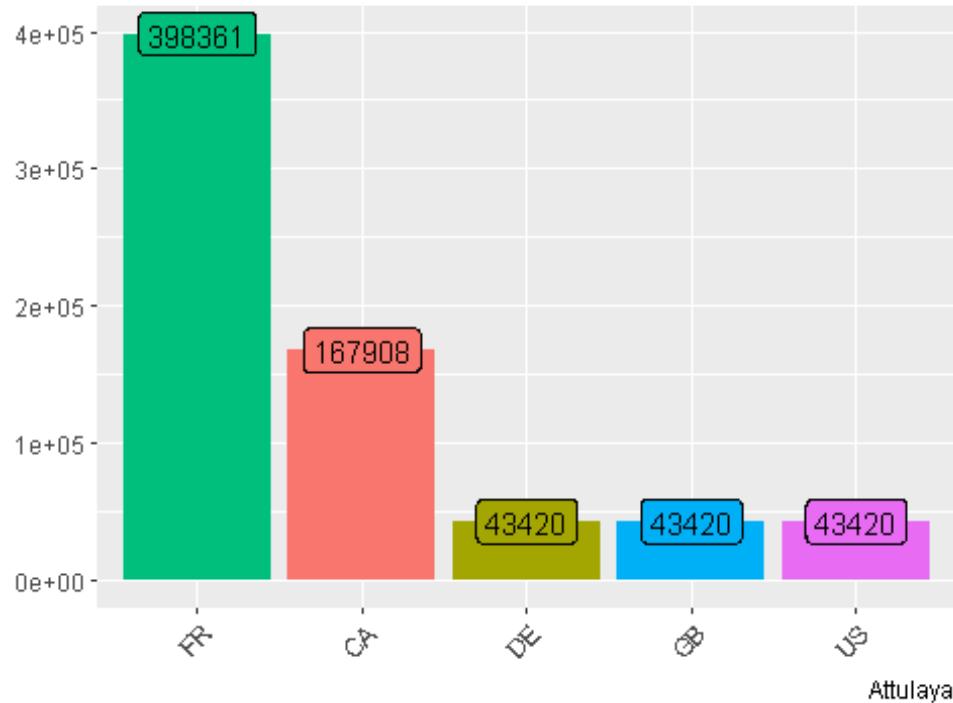
```
ggplot(videos[,"Total_Likes"=max(likes)],by=Location],aes(reorder(Location,
-
Total_Likes),Total_Likes,fill=Location))+geom_bar(stat="identity")+
geom_label(aes(label=Total_Likes))+guides(fill="none")+
theme(axis.text.x = element_text(angle = 45,hjust = 1))+
  labs(caption="Attulaya",title=" Total
number of likes by Countries")+xlab(NULL)+ylab(NULL)
```



In total number of dislikes

```
ggplot(videos[,"Total_Dislikes"=max(dislikes)],by=Location],aes(reorder(Location,-Total_Dislikes),Total_Dislikes,fill=Location))+geom_bar(stat="identity")+geom_label(aes(label=Total_Dislikes))+guides(fill="none") + theme(axis.text.x = element_text(angle = 45,hjust = 1))+ labs(caption="Attulaya",title=" Total Dislikes by Countries")+xlab(NULL)+ylab(NULL)
```

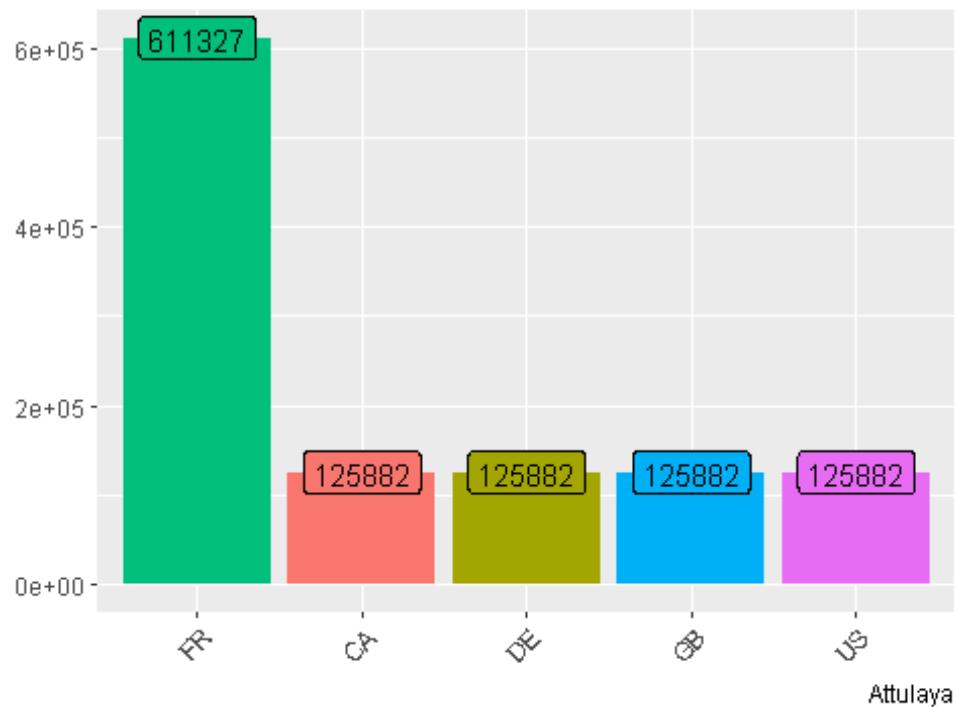
Total Dislikes by Countries



In total number of comments

```
ggplot(videos[,"Total_Comments"=max(comment_count)],by=Location],aes(reorder(Location,-Total_Comments),Total_Comments,fill=Location))+geom_bar(stat="identity")+geom_label(aes(label=Total_Comments))+guides(fill="none") + theme(axis.text.x = element_text(angle = 45,hjust = 1))+ labs(caption="Attulaya",title=" Total Comments by Countries")+xlab(NULL)+ylab(NULL)
```

Total Comments by Countries



Title length in words

```
videos[, "Word_len":= str_length(title)]  
  
ggplot(videos[, .N, keyby=Word_len], aes(Word_len, N, fill=N)) + geom_bar(stat =  
  "identity") + guides(fill="none") + labs(caption="Donyoe", title="Title length in  
words") + xlab(NULL) + ylab(NULL)
```

Title length in words

