INTRODUCTION

➢ Problem

You Tube has become a great streaming platform for people to share videos belonging to a wide spectrum of genres. It provides an opportunity to talented people to share their content. Since it is a free platform, a person can showcase his/ her talent by uploading videos and can see the impact that it makes. Any person can become a content creator and has the liberty to publish videos belonging to any genre, whether it's entertainment, food, education, travel, et. al. It helps one find the right audience, be it in any part of the world. Also, it has become one of the quickest media to gain popularity and earn money. This is also one of the motivations for many content creators on You Tube. In order to achieve this, understanding the parameters that are responsible for making any video trending on You Tube is of paramount importance. This can be achieved by analyzing the underlying pattern of various videos of different categories that get uploaded every day and become trending. This information will help a content creator of a particular genre to refine his/ her video and upload it using the important factors which will increase the chances of his/ her video becoming trending. Hence, this calls for a need to have good visualizations that can unambiguously depict the parameters to make a video of a particular category trending/ viral and reach the intended audience in the targeted countries. This will help the person to become one of the top performers on You Tube and earn many benefits.

➢ User

Ms. Kiara is a data analyst working in a renowned data analytics firm. As part of her work, she has to present data regarding factors that have high probability of making a music video trending on You Tube. She should analyze all the important parameters that have been responsible for making music videos trending/ viral in the past in countries namely USA, Canada, Germany and India which cover mainly all continents. The data should provide information regarding the most popular music channels which broadcasted videos that later became top trending. It should also represent the best time to upload videos to obtain maximum views based on the upload times of past top trending videos in different countries. Moreover, it should depict the most frequently used tags for music videos which will allow maximum views based on the tags used for past trending videos. All this information will be used by a new music content creator who wants to publish his/her video on You Tube and is interested in making it trending/ viral in various countries in a short time. He/ She can use this information to pick the right channel, best upload time and key tags to target its intended audience in different countries to achieve maximum views.

➢ Scenario

Since Music strikes a chord with majority of people, it is one of the most important categories for video upload on You Tube. Ms. Kiara has been assigned the task of designing visualizations for a musician which can help him gain instant popularity. This is possible if his videos achieve maximum views across different countries. This calls for analyzing the pattern in the past trending videos to understand the most important parameters that can make a video trending. The data analyzed by Ms. Kiara should be presented to the musician in a form which is easy to understand in one glance and can help him formulate a strategy to make his video trending. The visualization should be devoid of any technical jargons and should be visually appealing. It should provide the musician with key information for any video upload such as right keywords to use in video tags, best channel to launch the video and the best time to upload a video to target audience across all continents majorly in countries USA, Canada, Germany and India. This may not be a sure shot formula for the musician to become popular but such analysis will increase his chances of becoming one of the top performers on You Tube. If the data analyst is successful in making the right predictions for the musician, he can associate himself with the analytics firm after gaining popularity which will be beneficial for the company also. He can also ask for

sentiment analysis of his first video to further increase the probability of his later uploads to become trending.

➢ Datasets

Trending You Tube video statistics: This dataset contains data regarding trending You Tube videos of various genres across different countries. Information related to different factors like number of views, likes, dislikes, tags, et. al which make a video trending is also included. Although the dataset contains information for various countries, we have considered data only for USA, Canada, Germany and India in the form of csv files for this analysis as that covers all continents. Each of the csv files contains 16 variables which represent information about videos like their title, description, tags, thumbnail, channel on which they are launched, number of views, likes, dislikes and comments they received and date and time of upload. Along with these, 4 JSON files are also considered which contain information for the fields namely: kind, etag, id, snippet, channelId, title and assignable. They represent video category information for each of the countries.

Link to Dataset (Total Number of Datasets is 8 (csv and JSON files):

Canada Dataset:

https://www.kaggle.com/datasnaek/youtube-new/downloads/CAvideos.csv/114

https://www.kaggle.com/datasnaek/youtube-new/downloads/CA_category_id.json/114

Germany Dataset:

https://www.kaggle.com/datasnaek/youtube-new/downloads/DEvideos.csv/114

https://www.kaggle.com/datasnaek/youtube-new/downloads/DE_category_id.json/114

India Dataset:

https://www.kaggle.com/datasnaek/youtube-new/downloads/INvideos.csv/114

https://www.kaggle.com/datasnaek/youtube-new/downloads/IN_category_id.json/114

USA Dataset:

https://www.kaggle.com/datasnaek/youtube-new/downloads/USvideos.csv/114

https://www.kaggle.com/datasnaek/youtube-new/downloads/US_category_id.json/114

PRE-PROCESSING

➢ Cleaning

Following cleaning operations have been performed on the datasets:

- The missing values (NAs) have been found only in the field ‘description’ which has been replaced with space.
- The date format in the field ‘trending_date’ was not appropriate. It has been modified into the format YYYY-DD-MM by replacing ‘.’ With ‘-‘for further analysis.

➢ Wrangling

Following wrangling operations have been performed on the dataset:

- The data frames for each of the countries namely USA, Canada, Germany and India have been modified by adding a new column named 'country' to present the country information so that these data frames can be merged into one data frame.
- For each of the countries, category id and category information representing different genres has been extracted from JSON file and mapped to the data frame containing data from csv files. A new column named 'category' has been added to data frames of each country to represent genres.
- All data frames representing different countries have been merged together to create a final data frame to be used for further analysis as the fields in data frames are same representing same information.
- The merge data frame is further filtered to create a new data frame only for the category: Music as our analysis is for a music content creator targeting different countries.
- New fields namely 'publish_date' and 'publish_hour' have been calculated for the final data frame by performing operations on date time type of field in final data frame namely 'publish_time'.

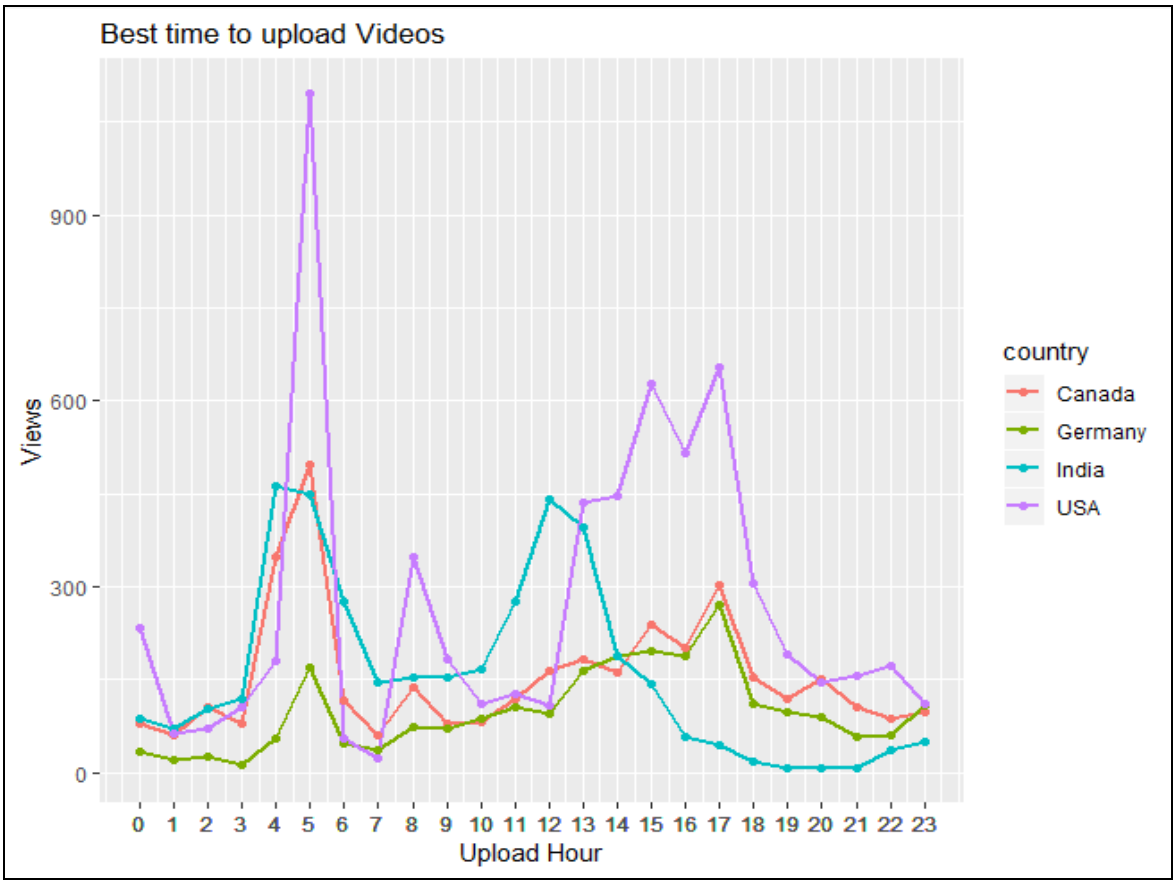
VISUALIZATIONS

➢ Visualization 1

The above visualization depicts 5 Most Popular Music Channels in different Countries by Views. It is very important for a musician to associate himself with the right music channel which had high success rate in the past. In USA, PTXOfficial, Charlie Puth, ibighit, SMTOWN and Billboard music channels had videos with highest views in order. In Canada, Speed Records, T-Series, T-Series Apna Punjab, WORLDSTARHIPHOP and Humble Music music channels had videos with highest views in order. In Germany, 385ideal, MrRevillz, Deutch Rap, Kontra K and netd muzik music channels had videos with

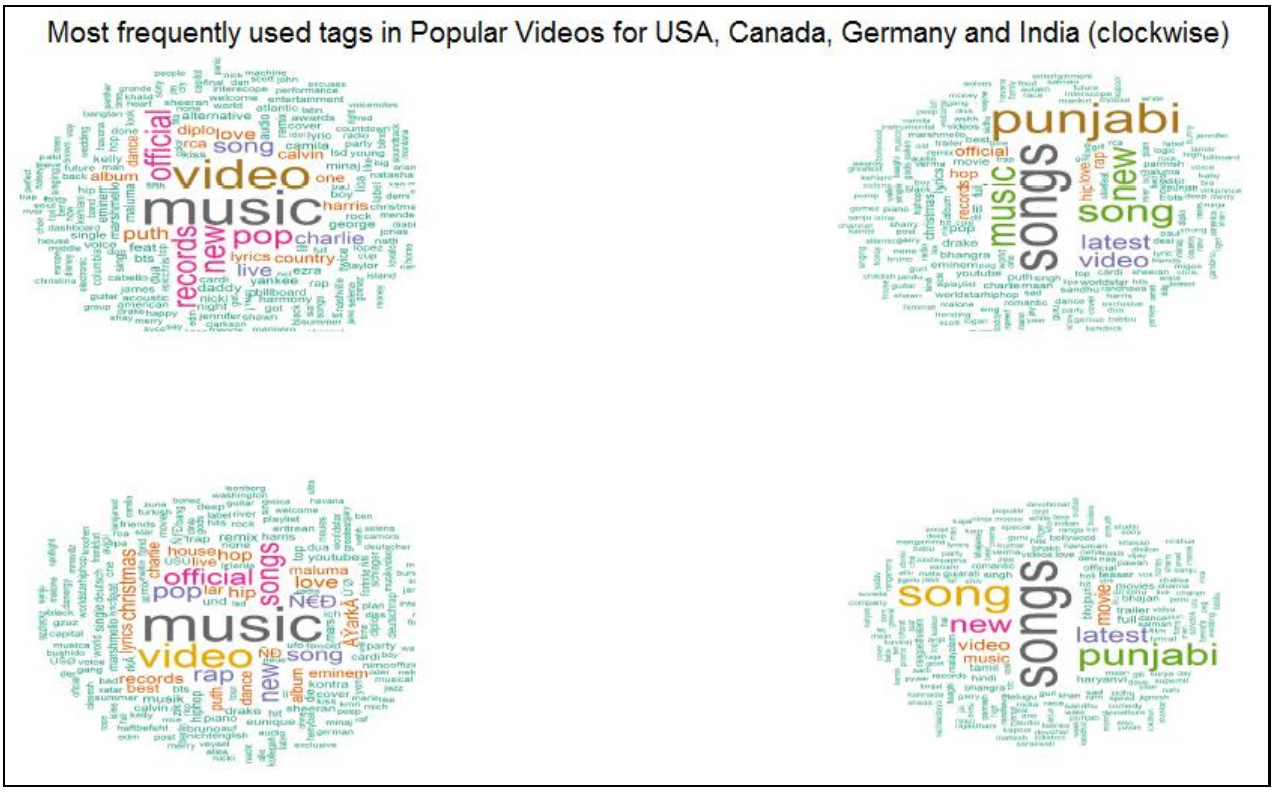
highest views in order. In India, T-series, Speed Records, Aditya Music, T-Series Apna Punjab and Zee Music Company music channels had videos with highest views in order. The top trending videos in the past were associated with these channels in different countries. This implies that if the music content creator uploads his videos on any of these channels, which are feasible to him, his video is highly likely to receive huge number of views based on the past records. This will increase the probability of his videos to become trending/ viral across continents.

➢ Visualization 2



The above visualization depicts information regarding the Best Hour of any day to upload music videos to target different countries based on maximum views received at different hours. Based on the visualization, the music content creator can infer that if his target audience is USA, then the best hours of the day to upload video is 05, 08, 13, 15 and 17 as peaks in Views were received at these hours with maximum at 05. For Canada, the best hours of the day to upload video is 05, 08, 12, 13, 15 and 17 as peaks in Views were received at these hours with maximum at 05. For Germany, the best hours of the day to upload video is 05, 08, 11, 13, 15 and 17 as peaks in Views were received at these hours with maximum at 17. For India, the best hours of the day to upload video is 04, 05, 12 and 13 as peaks in Views were received at these hours with maximum at 04. This information helps the musician to decide the best time to upload his videos based on past records.

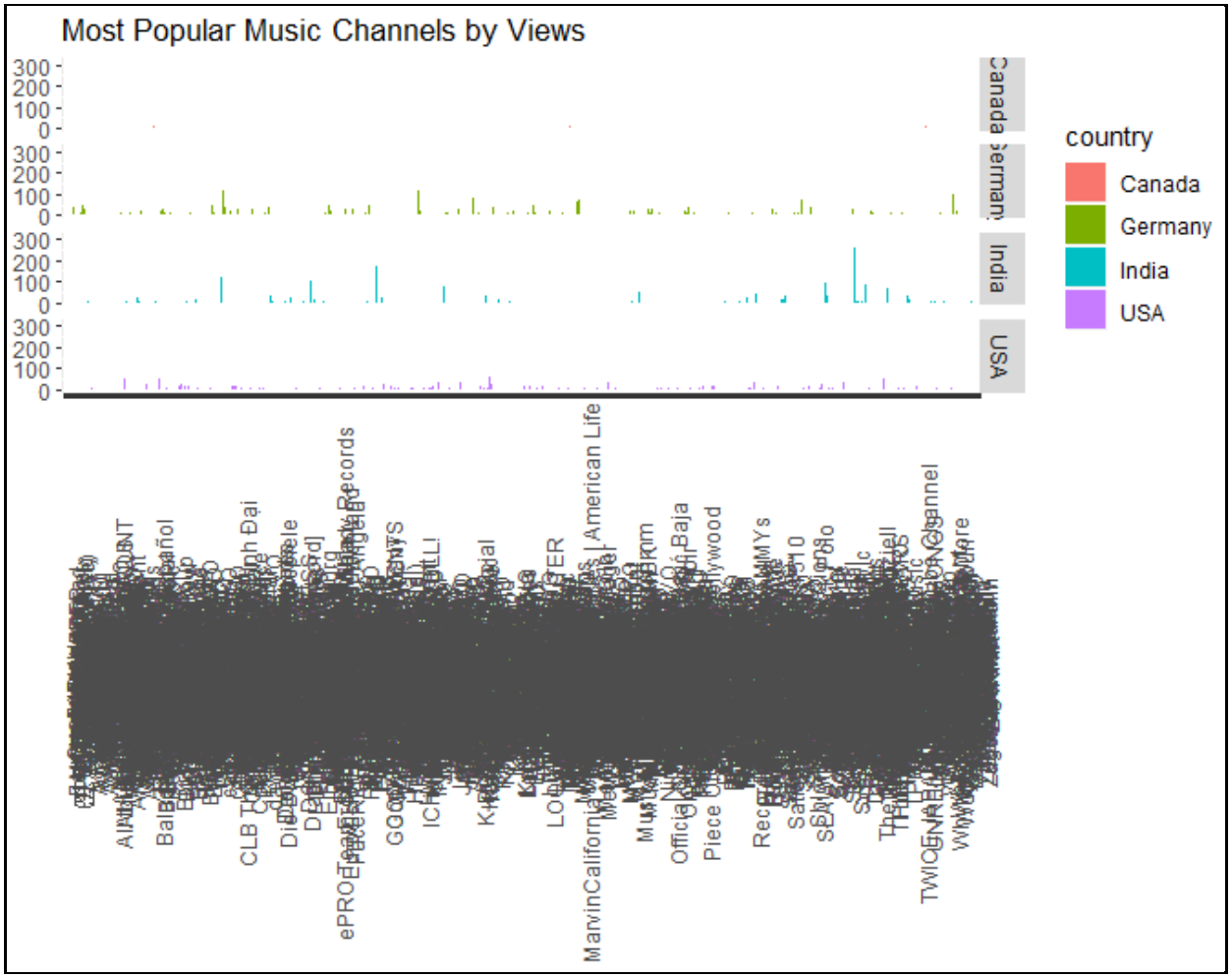
➢ Visualization 3



The above visualization depicts information regarding the most frequently used tags in popular videos for different countries. The tags used while uploading a video is one of the important parameters which decide the number of views that will be obtained by a video. The tags should be strong enough to match the keywords of the search phrases used by people while looking for a video. This is an important determinant of the number of views for a video which is based on the match between search phrase and tag. From the records of past trending videos, it can be observed that for USA, the key tags are music, video, pop, official, new, records, puth, charli, love, calvin, album, lyrics, live, country on and harris. For Canada, key tags are songs, punjabi, music, song, new, latest, video, official, hop, love and rap. For Germany, key tags are music, video, official, songs, rap, new, love, hop, house, eminem, lyrics, Christmas, best, records, puth and liar are key tags. For India, songs, song, new, punjabi, video, music, movie and latest are key tags. If the musician uses any of these tags for target countries, it will increase his chances of video to come at the top while a user searches the videos. Hence, it will increase his chances of getting maximum views.

PREVIOUS ITERATIONS

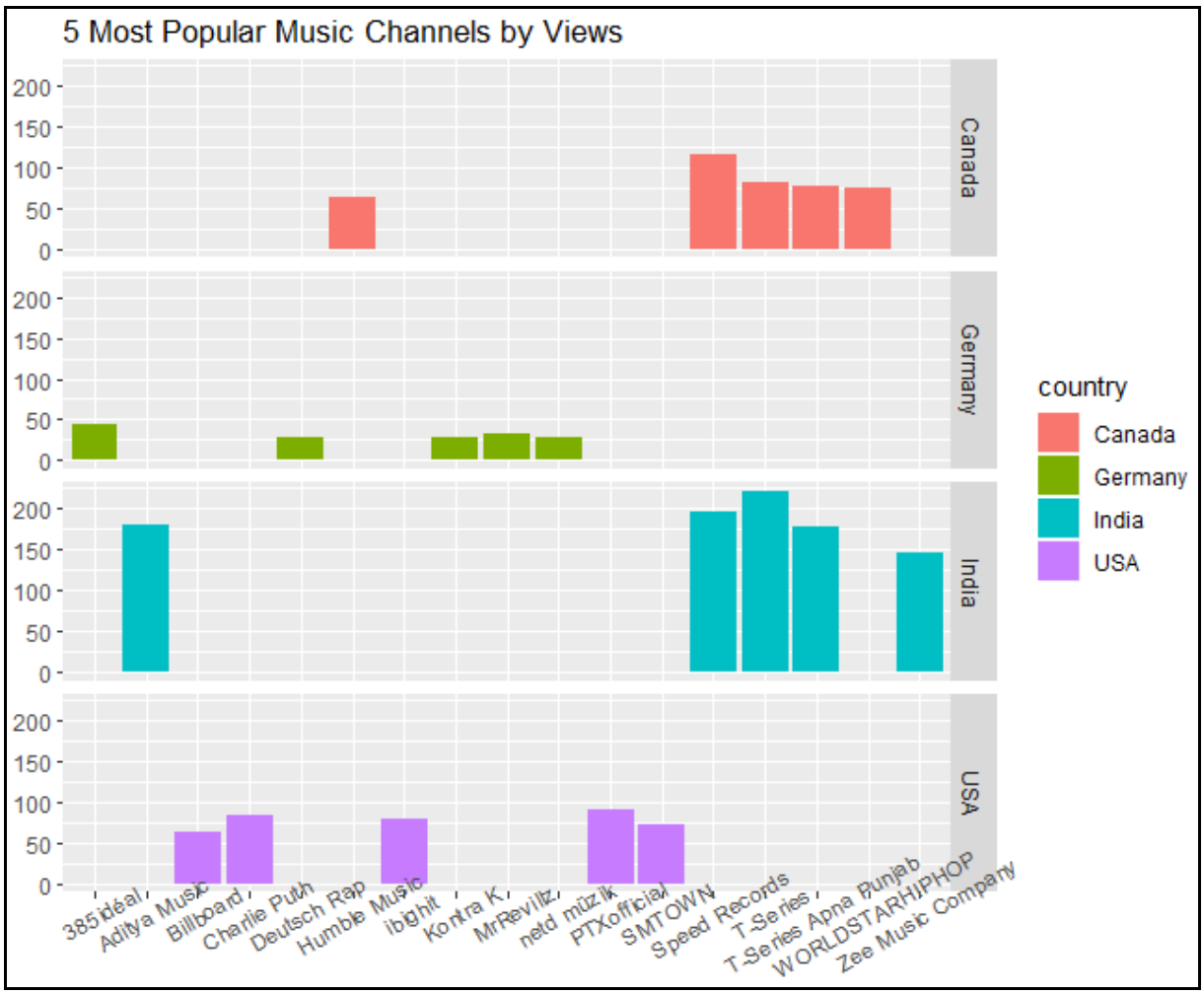
➢ Question 1 Iteration 1



In this iteration, following problem was encountered:

- Since all music channels were considered to determine popularity, the x-axis became over-crowded and hence, was difficult to interpret. As we don't need information about all the channels, we reduce the count to top 5 music channel in final visualization.

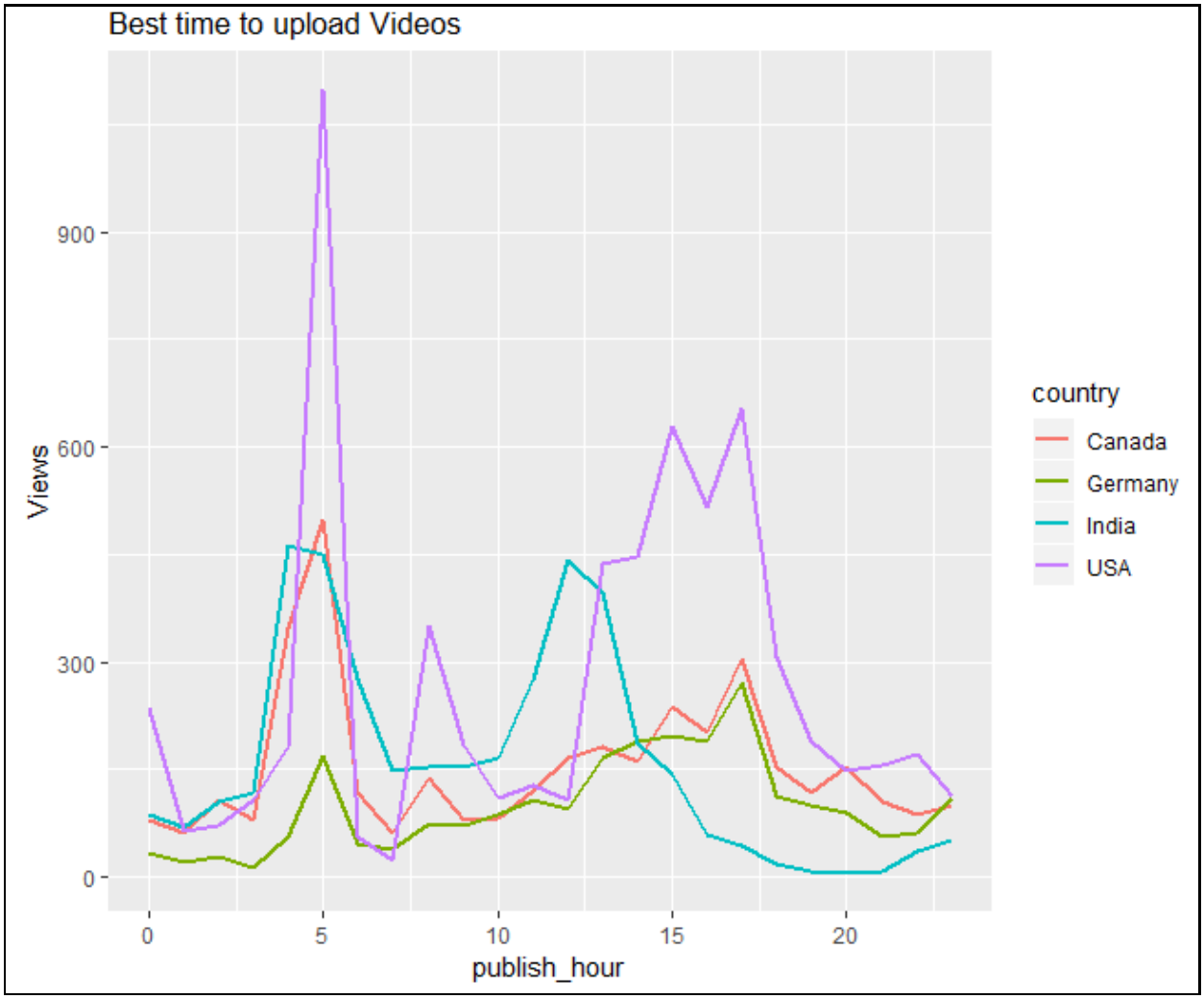
➢ Question 1 Iteration 2



In this iteration, following problem was encountered:

- After reducing the number of music channels to top 5 based on number of views from previous iteration, we found that the top 5 music channels for different countries are different. Hence, there are empty bars for various countries. If the empty bars would have been represented here, it might have given wrong information that a particular music channel which is top performing in another country is also present in a different country with lesser views. It might be possible that a particular channel is not at all present in different countries, hence it is misleading to represent names of top 5 music channels for all countries in this manner. Only the top channels for different countries should be represented individually which is represented in final visualization.

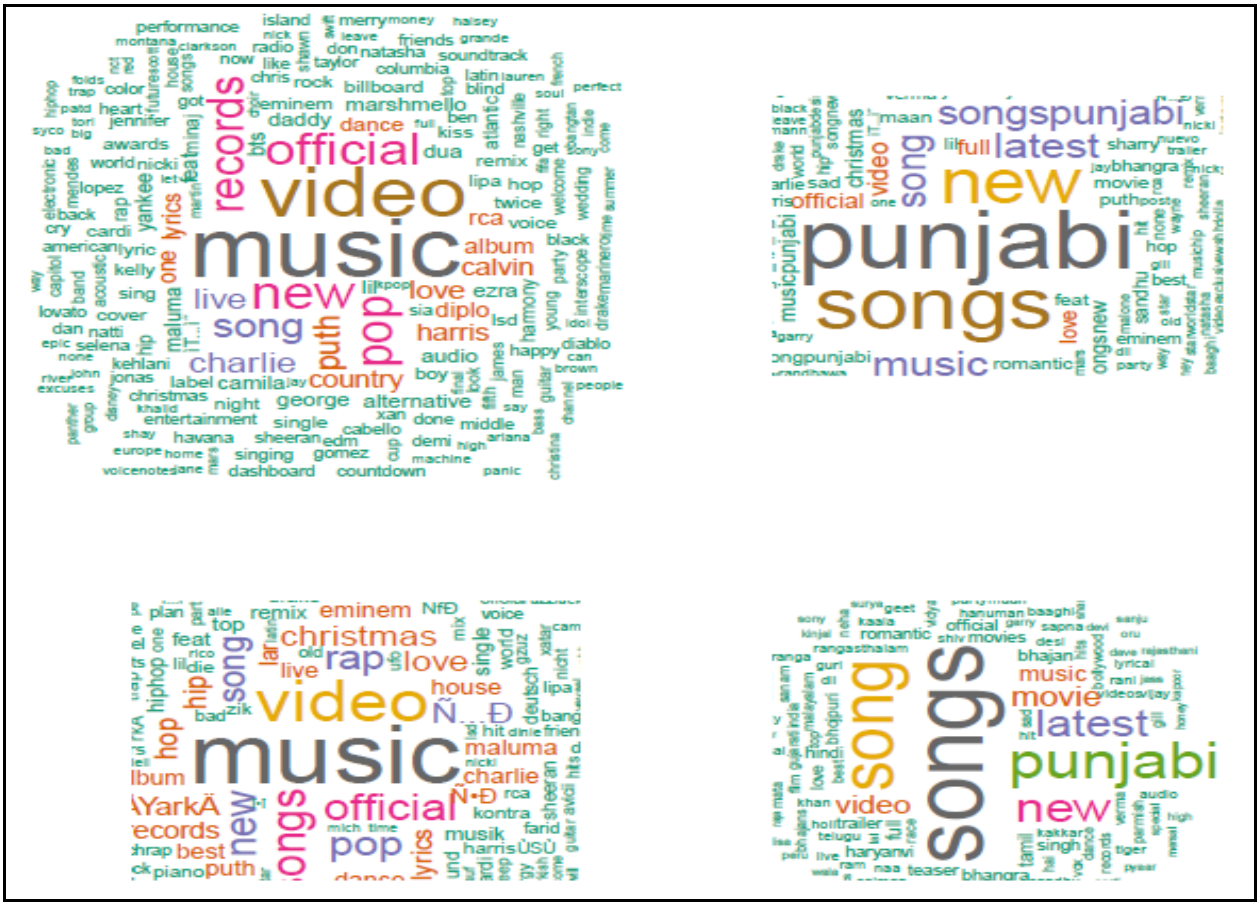
➢ Question 2 Iteration 1



In this iteration, following problems were encountered:

- The visualization did not represent all the values from 0 to 23 to represent various hour values on the x-axis. Which is an important parameter for this analysis.
- Points have not been included on the line graph to represent peak values.
- X-axis label requires refining for correct interpretation which is done in the final visualization.

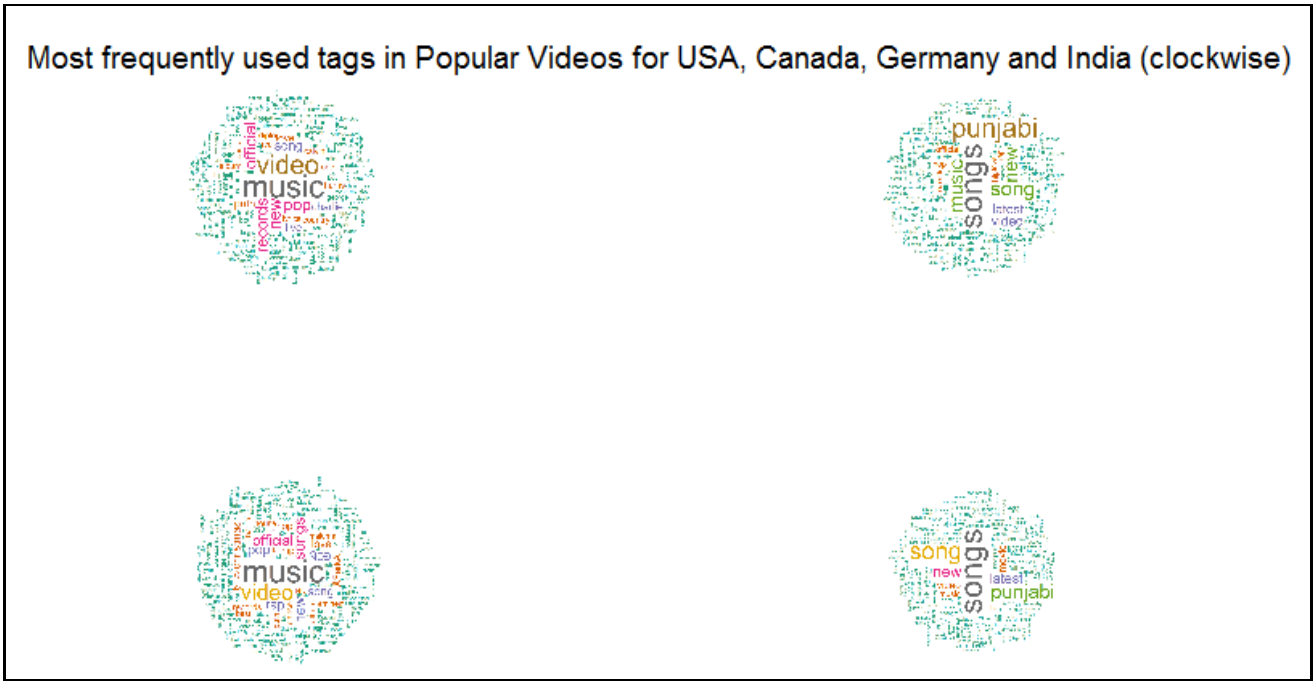
➢ Question 3 Iteration 1



In this iteration, following problems were encountered:

- There is no way to identify which word cloud belongs to which country as no description has been included while plotting multiple graphs together.
- The shape of all the word clouds as different and has not been completely represented as the white spaces overlap the word cloud. This reduces the overall information to be depicted and is incorrect representation
- Although all the word clouds are circular, but such multiple plotting led to non-uniformity in the graph which makes it visually less appealing. All these cases have been dealt with in the final graph.

Question 3 Iteration 2



In this iteration, following problem was encountered:

- The word cloud represented for different countries were very small in size compared to the width of the plot window available and hence, required change in both width and height so that the key tags were clearly visible and could be easily interpreted. Since it is created by combining 4 different png files into one plot, it requires resizing. This graph is not representative of a clear visualization and has been dealt with within limits for clear visibility in final visualization.

APPENDICES-CODE

```
# 1. load packages

# Load package to read csv files

library('readr')

# Load package for date and time manipulations

library('lubridate')

# Load package to perform string operations

library('stringr')

# Load the package required to read JSON files.

library("rjson")
```

```
# Load package to perform data manipulation
library('dplyr')
# Load package for creating plots
library('ggplot2')
# Load library to perform sql function on dataframe
library('sqldf')
install.packages('gsubfn')
library('gsubfn')
library('proto')
library('RSQLite')
# Load package to plot multiple graphs
library('gridExtra')
library('wordcloud')
library('RColorBrewer')
# load packages for text mining
library('tm')
library('NLP')
library('xtable')

# 2. Read Youtube trending videos data for 4 countries
CA_data<-read_csv('CAvideos.csv')
US_data<-read_csv('USvideos.csv')
DE_data<-read_csv('DEvideos.csv')
IN_data<-read_csv('INvideos.csv')

# 3. Data wrangling

# add new column to dataframes
CA_data$country<-'Canada'
US_data$country<-'USA'
DE_data$country<-'Germany'
```

```
IN_data$country<-'India'

# Map category information in JSON file to csv file for differnt Countries
category<-fromJSON(file = "US_category_id.json")
print(category)
df<-as.data.frame(NULL)
for(i in 1:32){
  df[i,1]<-category$items[[i]]$id
  df[i,2]<-category$items[[i]]$snippet$title
}
colnames(df)<-c('category_id','category')

US_data$category_id<-as.character(US_data$category_id)
US_data<-left_join(US_data,df,by='category_id')
apply(is.na(US_data), 2, which)

CA_data$category_id<-as.character(CA_data$category_id)
CA_data<-left_join(CA_data,df,by='category_id')
apply(is.na(CA_data), 2, which)

DE_data$category_id<-as.character(DE_data$category_id)
DE_data<-left_join(DE_data,df,by='category_id')
apply(is.na(DE_data), 2, which)

IN_data$category_id<-as.character(IN_data$category_id)
IN_data<-left_join(IN_data,df,by='category_id')
apply(is.na(IN_data), 2, which)

# combine all files in 1 dataframe
all_country_data<-rbind(CA_data,US_data,DE_data,IN_data)

# extract data only for Music category
```

```
music_data<-filter(all_country_data,all_country_data$category=='Music')

# convert date and time in correct formats
music_data$trending_date<-str_replace_all(music_data$trending_date,'\\.','-')
music_data$publish_date<-as.Date(music_data$publish_time,format='%Y-%m-%d')
music_data$publish_time<-strftime(music_data$publish_time,format='%H:%M:%S')
music_data$publish_hour<-hms(music_data$publish_time)$hour

# 4.Data cleaning

# find missing values in data frame
table(is.na(music_data))
apply(is.na(music_data), 2, which)

# replace missing values in description
music_data$description<-str_replace_na(music_data$description)
music_data$description<-str_replace_all(music_data$description,'NA','')

# Visualizations

# Visualization 1: 5 Most Popular Music Channels in different Countries by Views

# Iteration 1

all_channels<-sqldf("select channel_title, count(views) as views, country from music_data  group by channel_title")
all_channels <- all_channels[with(all_channels, order(-views)), ]

ggplot(data=all_channels,aes(x=channel_title,y=views))+
  labs(title='Most Popular Music Channels by Views')+
  geom_bar(stat = 'identity',aes(fill=country))+
  facet_grid(country~.)+
```

```
  theme(axis.title.y = element_blank(),axis.title.x = element_blank(),axis.text.x = element_text(angle = 90))


# Iteration 2


US_channels<-sqldf("select channel_title, count(views) as views, country from music_data where country='USA' group by channel_title")

US_channels <- US_channels[with(US_channels, order(-views)), ]


CA_channels<-sqldf("select channel_title, count(views) as views, country from music_data where country='Canada' group by channel_title")

CA_channels <- CA_channels[with(CA_channels, order(-views)), ]


DE_channels<-sqldf("select channel_title, count(views) as views, country from music_data where country='Germany' group by channel_title")

DE_channels <- DE_channels[with(DE_channels, order(-views)), ]


IN_channels<-sqldf("select channel_title, count(views) as views, country from music_data where country='India' group by channel_title")

IN_channels <- IN_channels[with(IN_channels, order(-views)), ]


top5_channels<-rbind(US_channels[1:5,],CA_channels[1:5,],DE_channels[1:5,],IN_channels[1:5,])


ggplot(data=top5_channels,aes(x=channel_title,y=views))+

  labs(title='5 Most Popular Music Channels by Views')+

  geom_bar(stat = 'identity',aes(fill=country))+

  facet_grid(country~.)+

  theme(axis.title.y = element_blank(),axis.title.x = element_blank(),axis.text.x = element_text(angle = 30))


# final visualization


z =filter(top5_channels,top5_channels$country=='USA')
```

```
p1<-ggplot(data =z )
p1<-p1+
  labs(title='5 Most Popular Music Channels in USA by Views')+
  geom_bar(stat = 'identity',fill='firebrick',width=.5,aes(x=channel_title,y=views))+
  theme(axis.title.y = element_blank(),axis.title.x = element_blank(),axis.text.x = element_text(angle =
30))+
  scale_x_discrete(limits= z$channel_title)
p1


z =filter(top5_channels,top5_channels$country=='Canada')
p2<-ggplot(data =z )
p2<-p2+
  labs(title='5 Most Popular Music Channels in Canada by Views')+
  geom_bar(stat = 'identity',fill='steelblue',width=.5,aes(x=channel_title,y=views))+
  theme(axis.title.y = element_blank(),axis.title.x = element_blank(),axis.text.x = element_text(angle =
30))+
  scale_x_discrete(limits= z$channel_title)
p2


z =filter(top5_channels,top5_channels$country=='Germany')
p3<-ggplot(data =z)
p3<-p3+
  labs(title='5 Most Popular Music Channels in Germany by Views')+
  geom_bar(stat = 'identity',fill='darkgreen',width=.5,aes(x=channel_title,y=views))+
  theme(axis.title.y = element_blank(),axis.title.x = element_blank(),axis.text.x = element_text(angle =
30))+
  scale_x_discrete(limits= z$channel_title)
p3


z =filter(top5_channels,top5_channels$country=='India')
p4<-ggplot(data =z )
p4<-p4+
```

```
  labs(title='5 Most Popular Music Channels in India by Views')+
  geom_bar(stat = 'identity',fill='tomato1',width=.5,aes(x=channel_title,y=views))+
  theme(axis.title.y = element_blank(),axis.title.x = element_blank(),axis.text.x = element_text(angle = 30))+
  scale_x_discrete(limits= z$channel_title)
p4


grid.arrange(p1,p2,p3,p4,nrow=2,ncol=2)


# Visualization 2: Best time to upload Videos in different Countries


publish_data<-sqldf("select publish_hour,count(views)as views, country from music_data group by publish_hour,country order by country")


# Iteration 1
# without x-axis labels


ggplot(data=publish_data,aes(x=publish_hour,y=views))+
  labs(y='Views',title='Best time to upload Videos')+
  geom_line(aes(colour=country),size=1)


# final visualization
ggplot(data=publish_data,aes(x=publish_hour,y=views))+
  labs(y='Views',title='Best time to upload Videos')+
  geom_line(aes(colour=country),size=1)+geom_point(aes(colour=country))+
  scale_x_continuous("Upload Hour", labels = as.character(publish_data$publish_hour), breaks = publish_data$publish_hour)


# Visualization 3: Most frequently used tags in Popular Videos across different Countries


# Word Cloud for USA
```

```
# Load the data as a corpus
data1<-subset(music_data,music_data$country=='USA')$tags
docs1 <- Corpus(VectorSource(data1))

# Text Transformation
# Replacing “/”, “@” and “|” with space
toSpace1 <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs1 <- tm_map(docs1, toSpace1, "/")
docs1 <- tm_map(docs1, toSpace1, "@")
docs1 <- tm_map(docs1, toSpace1, "\\|")

# Cleaning text

# Convert the text to lower case
docs1 <- tm_map(docs1, content_transformer(tolower))
# Remove numbers
docs1 <- tm_map(docs1, removeNumbers)
# Remove english common stopwords
docs1 <- tm_map(docs1, removeWords, stopwords("english"))
# Remove punctuations
docs1 <- tm_map(docs1, removePunctuation)
# Eliminate extra white spaces
docs1 <- tm_map(docs1, stripWhitespace)

inspect(docs1)

# Build a term-document matrix

dtm1 <- TermDocumentMatrix(docs1)
m1 <- as.matrix(dtm1)
v1 <- sort(rowSums(m1),decreasing=TRUE)
```

```
d1 <- data.frame(word = names(v1),freq=v1)

# Generate word cloud

set.seed(1234)
wordcloud(words = d1$word, freq = d1$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

 png("USACloud.png", width=12, height=8, units="in", res=300)
 wordcloud(words = d1$word, freq = d1$freq, min.freq = 1,
           max.words=200, random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))
 dev.off()

# Word Cloud for Canada

# Load the data as a corpus
data2<-subset(music_data,music_data$country=='Canada')$tags
docs2 <- Corpus(VectorSource(data2))

# Text Transformation
# Replacing “/”, “@” and “|” with space
toSpace2 <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs2 <- tm_map(docs2, toSpace2, "/")
docs2 <- tm_map(docs2, toSpace2, "@")
docs2 <- tm_map(docs2, toSpace2, "\\|")

# Cleaning text

# Convert the text to lower case
```

```
docs2 <- tm_map(docs2, content_transformer(tolower))
# Remove numbers
docs2 <- tm_map(docs2, removeNumbers)
# Remove english common stopwords
docs2 <- tm_map(docs2, removeWords, stopwords("english"))
# Remove punctuations
docs2 <- tm_map(docs2, removePunctuation)
# Eliminate extra white spaces
docs2 <- tm_map(docs2, stripWhitespace)

# Build a term-document matrix

dtm2 <- TermDocumentMatrix(docs2)
m2 <- as.matrix(dtm2)
v2 <- sort(rowSums(m2),decreasing=TRUE)
d2 <- data.frame(word = names(v2),freq=v2)

# Generate word cloud

set.seed(1234)
wordcloud(words = d2$word, freq = d2$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

 png("CanadaCloud.png", width=12, height=8, units="in", res=300)
 wordcloud(words = d2$word, freq = d2$freq, min.freq = 1,
           max.words=200, random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))
 dev.off()

# Word Cloud for Germany
```

```
# Load the data as a corpus
data3<-subset(music_data,music_data$country=='Germany')$tags
docs3 <- Corpus(VectorSource(data3))

# Text Transformation
# Replacing “/”, “@” and “|” with space
toSpace3 <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs3 <- tm_map(docs3, toSpace3, "/")
docs3 <- tm_map(docs3, toSpace3, "@")
docs3 <- tm_map(docs3, toSpace3, "\\|")

# Cleaning text

# Convert the text to lower case
docs3 <- tm_map(docs3, content_transformer(tolower))
# Remove numbers
docs3 <- tm_map(docs3, removeNumbers)
# Remove english common stopwords
docs3 <- tm_map(docs3, removeWords, stopwords("english"))
# Remove punctuations
docs3 <- tm_map(docs3, removePunctuation)
# Eliminate extra white spaces
docs3 <- tm_map(docs3, stripWhitespace)

# Build a term-document matrix

dtm3 <- TermDocumentMatrix(docs3)
m3 <- as.matrix(dtm3)
v3 <- sort(rowSums(m3),decreasing=TRUE)
d3 <- data.frame(word = names(v3),freq=v3)
```

```
# Generate word cloud
set.seed(1234)
wordcloud(words = d3$word, freq = d3$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

 png("GermanyCloud.png", width=12, height=8, units="in", res=300)
 wordcloud(words = d3$word, freq = d3$freq, min.freq = 1,
           max.words=200, random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))
 dev.off()

# Word Cloud for India

# Load the data as a corpus
data4<-subset(music_data,music_data$country=='India')$tags
docs4 <- Corpus(VectorSource(data4))

# Text Transformation
# Replacing “/”, “@” and “|” with space
toSpace4 <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs4 <- tm_map(docs4, toSpace4, "/")
docs4 <- tm_map(docs4, toSpace4, "@")
docs4 <- tm_map(docs4, toSpace4, "\\|")

# Cleaning text

# Convert the text to lower case
docs4 <- tm_map(docs4, content_transformer(tolower))
# Remove numbers
docs4 <- tm_map(docs4, removeNumbers)
# Remove english common stopwords
```

```
docs4 <- tm_map(docs4, removeWords, stopwords("english"))
# Remove punctuations
docs4 <- tm_map(docs4, removePunctuation)
# Eliminate extra white spaces
docs4 <- tm_map(docs4, stripWhitespace)

# Build a term-document matrix

dtm4 <- TermDocumentMatrix(docs4)
m4 <- as.matrix(dtm4)
v4 <- sort(rowSums(m4),decreasing=TRUE)
d4 <- data.frame(word = names(v4),freq=v4)

# Generate word cloud
set.seed(1234)
wordcloud(words = d4$word, freq = d4$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

 png("IndiaCloud.png", width=12, height=8, units="in", res=300)
 wordcloud(words = d4$word, freq = d4$freq, min.freq = 1,
           max.words=200, random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))
 dev.off()

# Iteration 1

par(mfrow=c(2,2))
wordcloud(words = d1$word, freq = d1$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
wordcloud(words = d2$word, freq = d2$freq, min.freq = 1,
```

```
        max.words=200, random.order=FALSE, rot.per=0.35,
        colors=brewer.pal(8, "Dark2"))
wordcloud(words = d3$word, freq = d3$freq, min.freq = 1,
        max.words=200, random.order=FALSE, rot.per=0.35,
        colors=brewer.pal(8, "Dark2"))
wordcloud(words = d4$word, freq = d4$freq, min.freq = 1,
        max.words=200, random.order=FALSE, rot.per=0.35,
        colors=brewer.pal(8, "Dark2"))
dev.off()



# Iteration 2
rl <- lapply(list("USACloud.png","CanadaCloud.png","GermanyCloud.png","IndiaCloud.png"), png::readPNG)
gl <- lapply(rl, grid::rasterGrob)
plot.new()
do.call(gridExtra::grid.arrange, gl)
title(paste0("Most frequently used tags in Popular Videos for USA, Canada, Germany and India (clockwise)"), font.main= 1)


# Final visualization


img1 <- readPNG( "USACloud.png")
img2 <- readPNG( "CanadaCloud.png")
img3 <- readPNG( "GermanyCloud.png")
img4 <- readPNG( "IndiaCloud.png")
g1 <- rasterGrob(img1, interpolate=TRUE, width = unit(11,"in"), height=unit(5,"in"))
g2 <- rasterGrob(img2, interpolate=TRUE,width = unit(11,"in"), height=unit(5,"in"))
g3 <- rasterGrob(img3, interpolate=TRUE, width = unit(11,"in"), height=unit(5,"in"))
g4 <- rasterGrob(img4, interpolate=TRUE,width = unit(11,"in"), height=unit(5,"in"))
plot.new()
grid.arrange(g1, g2,g3,g4, nrow=2,widths=c(7,7))
```

```
title(paste0("Most frequently used tags in Popular Videos for USA, Canada, Germany and India (clockwise)"), font.main= 1)

dev.off()
```