Top YouTube channel

Olena

19 02 2021

Title: “Top YouTube channel”

Synopsis

The goal of this study was to find out what is the most popular YouTube channel in some countries. This study involves exploring the dataset downloaded from https://www.kaggle.com/datasnaek/youtube-new. This dataset is a daily record of the top trending YouTube videos. Data is included for several countries (USA, Great Britain, Germany, Canada, France, Russia, Mexico, South Korea, Japan and India). Data includes the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count. The estimation of most popular YouTube channel was assessed taking into consideration numbers of views per each channel as well as numbers of likes per each channel. The analysis shows the 10 most popular YouTube channels by views and by likes.

Data processing

Load needed libraries and file

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
CA <- read.csv(file="CAvideos.csv")
DE <- read.csv(file="DEvideos.csv")
FR <- read.csv(file="FRvideos.csv")
GB <- read.csv(file="GBvideos.csv")
IN <- read.csv(file="INvideos.csv")
JP <- read.csv(file="JPvideos.csv")
KR <- read.csv(file="KRvideos.csv")
MX <- read.csv(file="MXvideos.csv")
RU <- read.csv(file="RUvideos.csv")
US <- read.csv(file="USvideos.csv")
```

Subset necessary data

```
CA1 <- CA[, c("channel_title", "views", "likes")]
DE1 <- DE[, c("channel_title", "views", "likes")]
FR1 <- FR[, c("channel_title", "views", "likes")]
GB1 <- GB[, c("channel_title", "views", "likes")]
IN1 <- IN[, c("channel_title", "views", "likes")]
JP1 <- JP[, c("channel_title", "views", "likes")]
KR1 <- KR[, c("channel_title", "views", "likes")]
MX1 <- MX[, c("channel_title", "views", "likes")]
RU1 <- RU[, c("channel_title", "views", "likes")]
US1 <- US[, c("channel_title", "views", "likes")]
```

Joining the dataframes

```
CHAN <- rbind(CA1, DE1, FR1, GB1, IN1, JP1, KR1, MX1, RU1, US1)
```

Look at data

```
str(CHAN)
```

```
## 'data.frame':    375942 obs. of  3 variables:
##  $ channel_title: Factor w/ 37824 levels "- ж¬ўиїЋи®ўй\230… -жµ\231ж±џеЌ«и§†гЂђеГ”и·‘еђ§гЂ‘е®\230ж–\271йў‘йЃ“",..: 1258 1
874 3548 3018 1221 1173 4549 702 2478 3698 ...
##  $ views        : int  17158579 1014651 3191434 2095828 33523622 1309699 2987945 748374 4477587 505161 ...
##  $ likes        : int  787425 127794 146035 132239 1634130 103755 187464 57534 292837 4135 ...
```

Converting variable “views” and “likes” to numeric

```
CHAN$views <- as.numeric(as.character(CHAN$views))
CHAN$likes <- as.numeric(as.character(CHAN$likes))
```

Missing values identification

```
sum(is.na(CHAN)) # There are no NA's.
```

```
## [1] 0
```

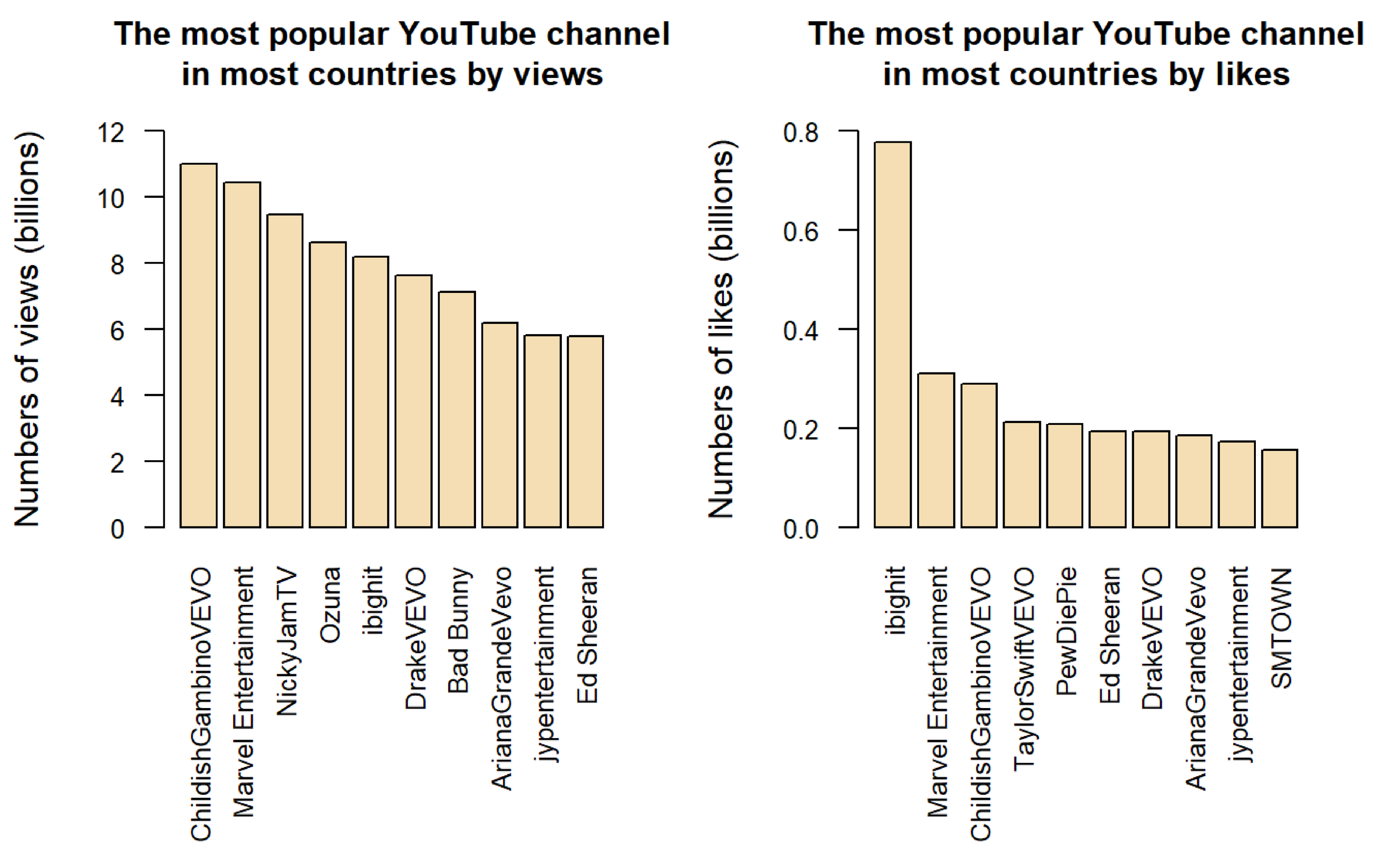
Results

Answering the question “What is the most popular YouTube channel in some countries?”

```
viewCHAN <- aggregate(views~channel_title, CHAN, sum)
likeCHAN <- aggregate(likes~channel_title, CHAN, sum)

top10_view <- arrange(viewCHAN, desc(viewCHAN$views))[1:10,]
top10_likes <- arrange(likeCHAN, desc(likeCHAN$likes))[1:10,]

par(mfrow = c(1, 2), mar=c(11,4,4,2))
barplot(top10_view$views/(10^9), las=2, names.arg = top10_view$channel_title, cex.main=1,  ylim=c(0,12), cex.axis=0.8, cex.n
ames=0.8, main = c(paste("The most popular YouTube channel"),paste("in most countries by views")), ylab = "Numbers of views
(billions)", col="wheat")
barplot(top10_likes$likes/(10^9), las=2, names.arg = top10_likes$channel_title, cex.main=1,  ylim=c(0,0.8), cex.axis=0.8, ce
x.names=0.8, main = c(paste("The most popular YouTube channel"), paste("in most countries by likes")), ylab = "Numbers of li
kes (billions)", col="wheat")
```



So, the 10 most popular YouTube channel by views are: ChildishGambinoVEVO, Marvel Entertainment, NickyJamTV, Ozuna, ibighit, DrakeVEVO, Bad Bunny, ArianaGrandeVevo, jypentertainment, Ed Sheeran.

While the 10 most popular YouTube channel by likes are: ibighit, Marvel Entertainment, ChildishGambinoVEVO, TaylorSwiftVEVO, PewDiePie, Ed Sheeran, DrakeVEVO, ArianaGrandeVevo, jypentertainment, SMTOWN.