

# ARE YOU A YOUTUBE ENTHUSIAST ?

An analysis of personal Youtube watch history of  
two college students

---

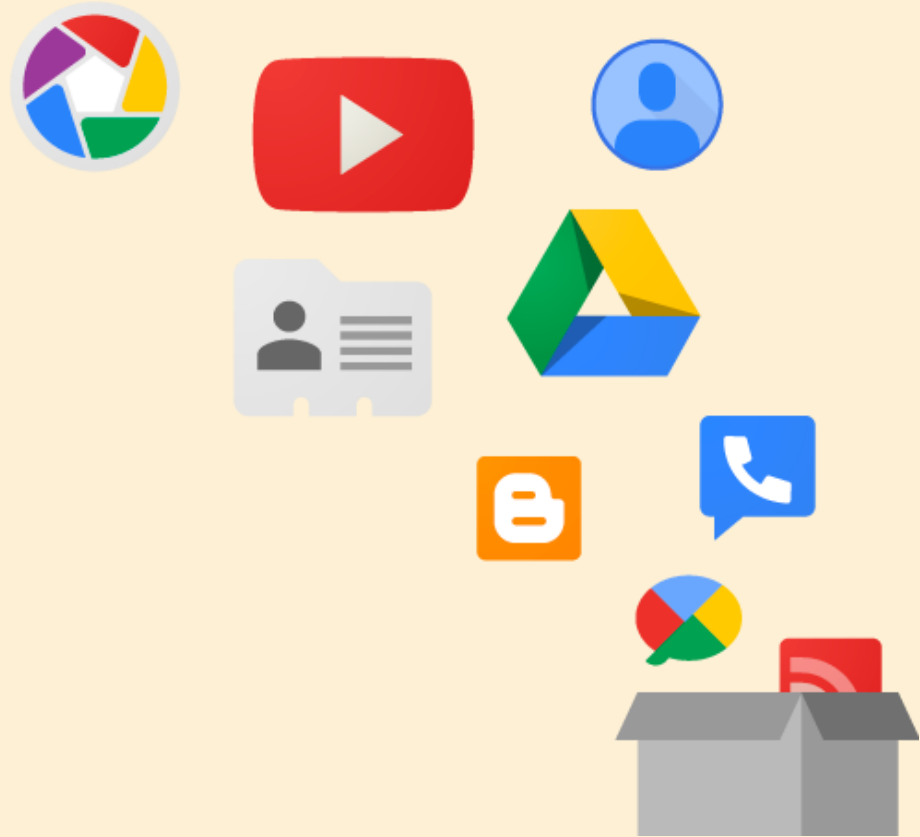
A project by Tam Nguyen & Ngoc Ha Vu



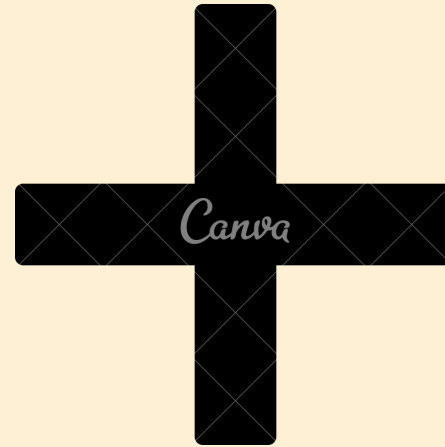
# Agenda

- 1 Introduction
- 2 Data Extraction & Cleaning
- 3 An overview of our Youtube data
- 4 An analysis of personal preference
- 5 Reference

# Introduction

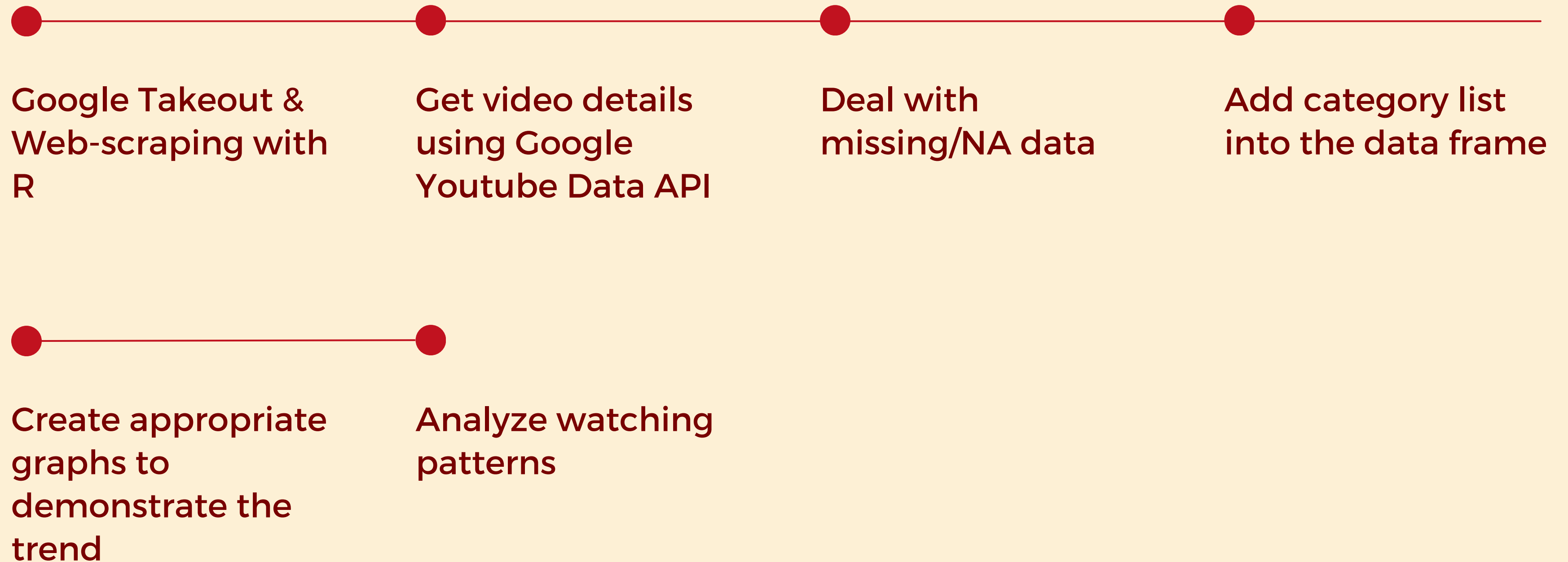


Google Takeout

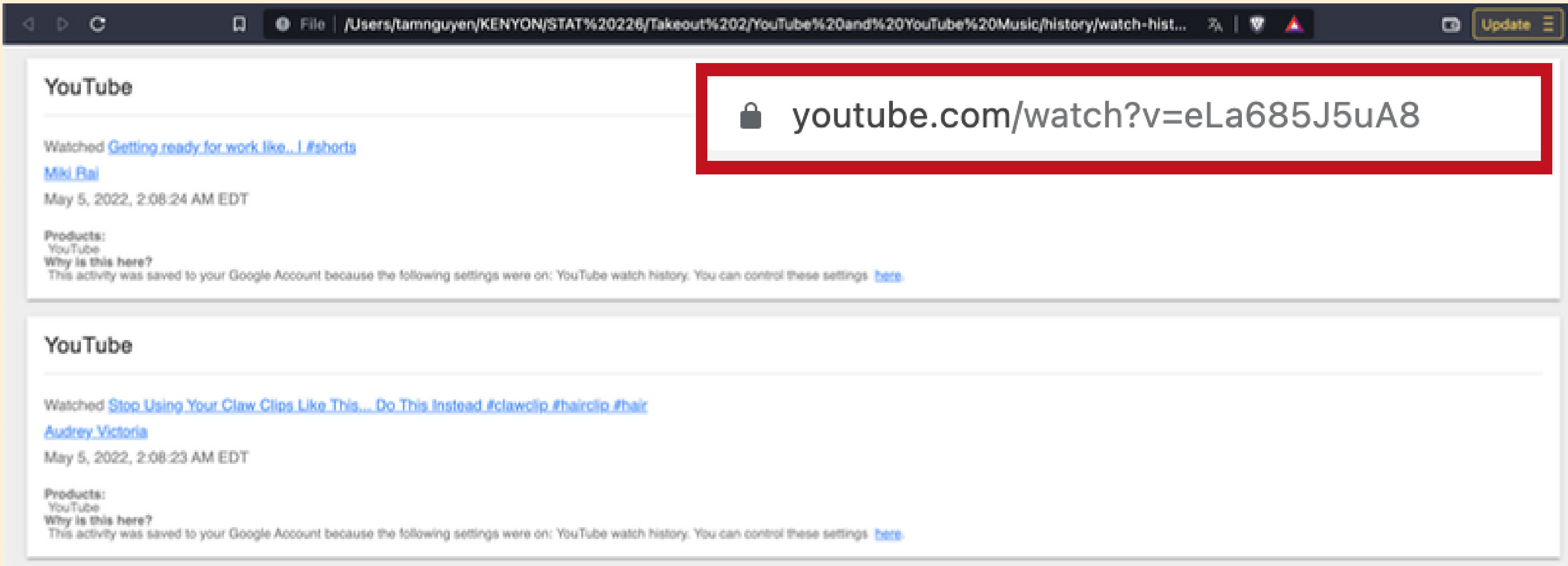


Youtube Data API

# Data Extraction & Cleaning Process



# Data extraction highlights



# stringr package

[illegible]

# Data extraction highlights

```
> watched_video_id
[1] "SGLBQR-ftVI" "NLBTbCfR-Fg" "UCXao7aTDQM" "GrdgyD4wFGU" "nIL_gX9W5qQ" "i0oSBvWBeP8"
[7] "T0sHaz4H9MQ" "Ern8Z4I1mIA" "wW3x1RjGsDE" "-L_o2M_FyaM" "GjSi40xJ0RY" "i_hdxA_SSyo"
[13] "redFrGBZoJY" "LspZCNP6aE" "n6Pnzi6r9NU" "ixdSsW5n2rI" "Kma3NpC3JKQ" "wXsBaFCfXb0"
[19] "ddaEtF0sFeM" "ntEoGvhoVac" "fgpwTxIv76Q" "AE3eD-XV4vg" "AmvA-XJF0j8" "ilB9h1pfjc8"
[25] "6DufHnGH690" "GgQF08dL5XQ" "xB2qsCnqAXA" "neCmEbI2VWg" "GQ4F9k4USfA" "KKc_RMln5UY"
[31] "hTGcMk_QXEg" "vVhKA9Av6vA" "1ntkDbxTnXA" "gq5rusq-u2s" "AsAWJMt4kCw" "9MtjSD0gtlk"
[37] "QeyRd8rhPJ0" "--44CW7yid4" "JuK93FeTmWU" "VWCtdepXYLQ" "oF6rBWct4G0" "qKTZNRo21WY"
[43] "l1IQZlEf8Tg" "n9NM7PZs8RM" "hTGcMk_QXEg" "0tWS7x0Ebn0" "wW3x1RjGsDE" "Kma3NpC3JKQ"
[49] "neCmEbI2VWg" "KKc_RMln5UY" "vVhKA9Av6vA" "1ntkDbxTnXA" "pI00Fx9x9zY" "kNw7nruHmgY"
[55] "nuW2Cu5Q8No" "lIZUmUbRSJM" "Yfe6ZTaSBEG" "GQ4F9k4USfA" "vVhKA9Av6vA" "1ntkDbxTnXA"
[61] "iiDw68JgUng" "WaoPg2Qhd0I" "9L8J9VBEXxI" "EZhVL4iiovI" "m7swtn7m7CI" "Uq1bQJc0l_A"
[67] "wHI77Ea_h4k" "Kma3NpC3JKQ" "cL4uhaQ58Rk" "2P8RU_dHyi4" "AmvA-XJF0j8" "GQ4F9k4USfA"
[73] "GgQF08dL5XQ" "xB2qsCnqAXA" "9MtjSD0gtlk" "neCmEbI2VWg" "KKc_RMln5UY" "vVhKA9Av6vA"
```

# Data extraction highlights

## httr package

```
youtubeAPIKey <- "XXXXX"
connectionURL <- "https://www.googleapis.com/youtube/v3/videos"
get_id <- list()
for (i in 1:nrow(watched_video_df)) {
  videoID <- watched_video_df[i,1]
  get_id[[i]] <- GET(connectionURL,
                    query = list(
                      key = youtubeAPIKey,
                      id = videoID,
                      fields = "items(id, snippet(channelId, title,
                                                                    categoryId))",
                      part = "snippet")
  ) %>%
  content("parsed") %>%
  pluck(1,1,2,3)
}

watched_video_df <- watched_video_df %>%
  left_join(category_list, by=c("category_id" = "id"))
```

# Data extraction highlights

get_id	list [5409]	List of length 5409
[[1]]	character [1]	'10'
[[2]]	character [1]	'22'
[[3]]	character [1]	'10'
[[4]]	character [1]	'10'
[[5]]	character [1]	'10'
[[6]]	character [1]	'10'
[[7]]	character [1]	'10'
[[8]]	character [1]	'10'
[[9]]	character [1]	'10'
[[10]]	character [1]	'10'
[[11]]	character [1]	'10'
[[12]]	character [1]	'10'
[[13]]	character [1]	'10'
[[14]]	character [1]	'10'
[[15]]	character [1]	'10'
[[16]]	character [1]	'10'
[[17]]	character [1]	'10'
[[18]]	character [1]	'10'



# Data extraction highlights

```
# Importing category list
category_list <- Youtube_Category_ID...Sheet1 %>%
  mutate(id = as.character(id))

# Mutate into the watched_video_df
watched_video_df <- watched_video_df %>%
  mutate(vid_category = as.character(get_id))
watched_video_df

# Left_join() => Watched
watched_video_df <- watched_video_df %>%
  left_join(category_list, by = c("vid_category" = "id"))
glimpse(watched_video_df)
```

id	scrapedTitle	scrapedTime	time	vid_category	category_name
5QIBQR-rvI	DẤU Cho Mai Về - Sau (Official Music Video) / buitruonglinh	Sep 28, 2022, 12:58:56 AM EDT	2022-09-28 00:58:56	10	Music
NLBtsCR-Fg	ẢNH KHÔNG EM / V&O, (Original)	Sep 28, 2022, 12:54:49 AM EDT	2022-09-28 00:54:49	22	People & Blogs
UCXao7aTDQM	HÀ Anh Tuấn - Thương TÀI LÃ LÃ - LÃ LÃ ĐÔI CẢ-FA EM (Official MV)	Sep 28, 2022, 12:48:38 AM EDT	2022-09-28 00:48:38	10	Music
Gr8gyD4wFGU	CẢ ƠI Ơ - ẢNH KHẢ NG? - 14 Casper &mp; Ben (Official)	Sep 28, 2022, 12:43:36 AM EDT	2022-09-28 00:43:36	10	Music
nl_gX9W5qQ	V&O - MÀ NÀ NÀ - CẢ-FA EM / OFFICIAL MV	Sep 28, 2022, 12:39:17 AM EDT	2022-09-28 00:39:17	10	Music

# An overview of our Youtube Adventure in numbers



# A summary of our data

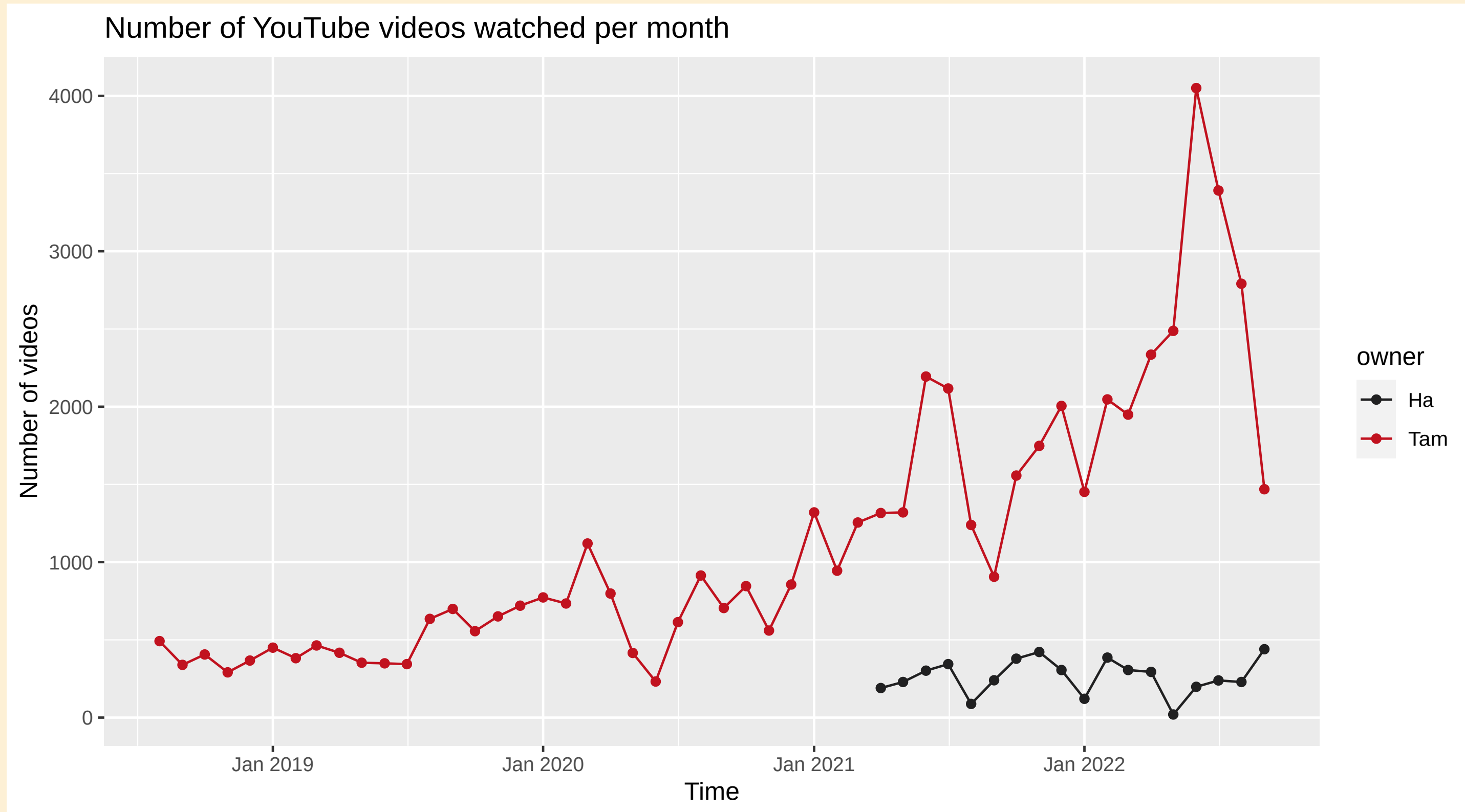
## Ha's

- April 13, 2021 - Sept. 28, 2022
- Total: **5409** videos
- Time spent: 1 year, 5 months, 14 days, and 18 hours

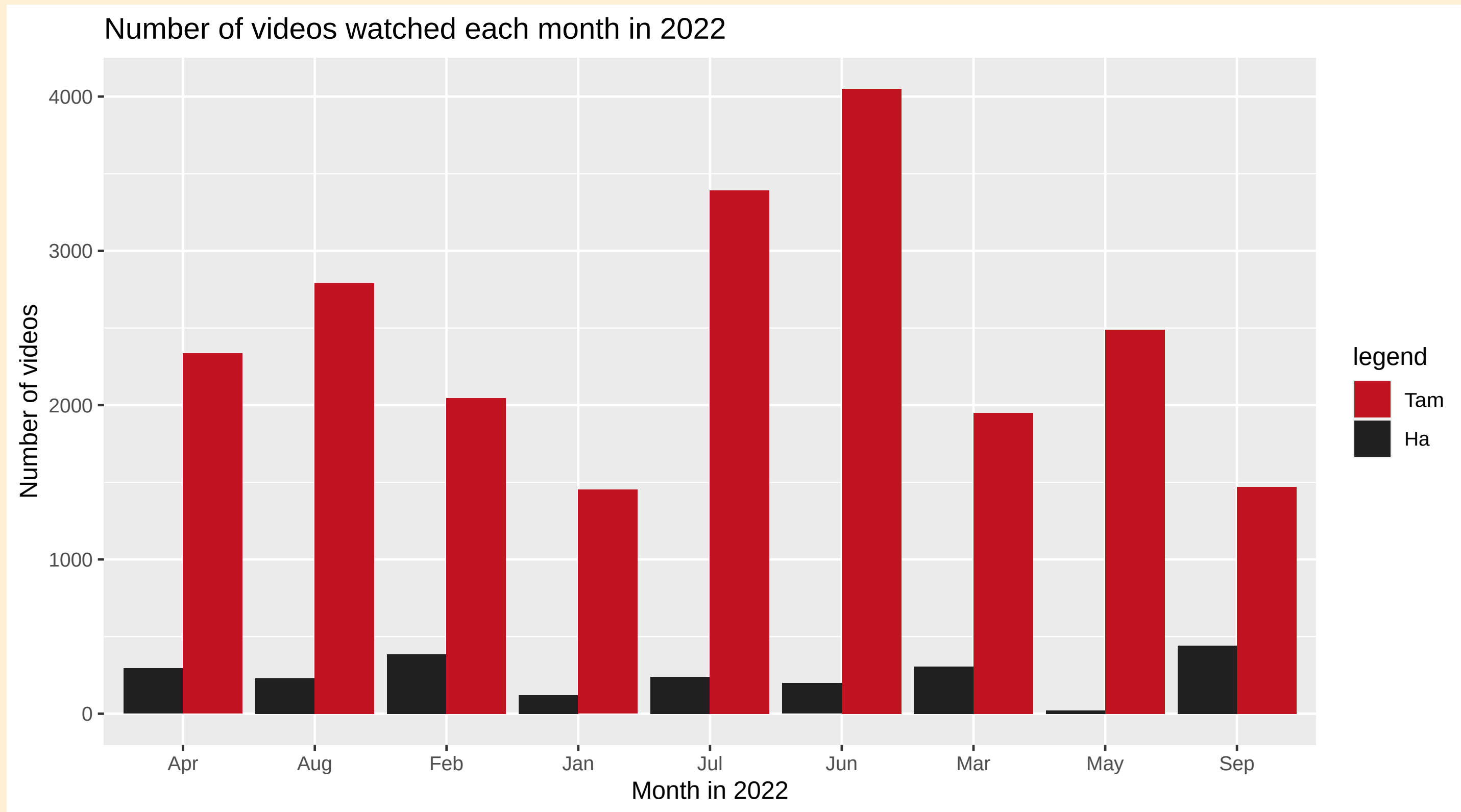
## Tam's

- Aug. 04, 2018 - Sept. 28, 2022
- Total: **56400** videos
- Time spent: 4 years, 1 month, 23 days, 12 hours and 31 minutes

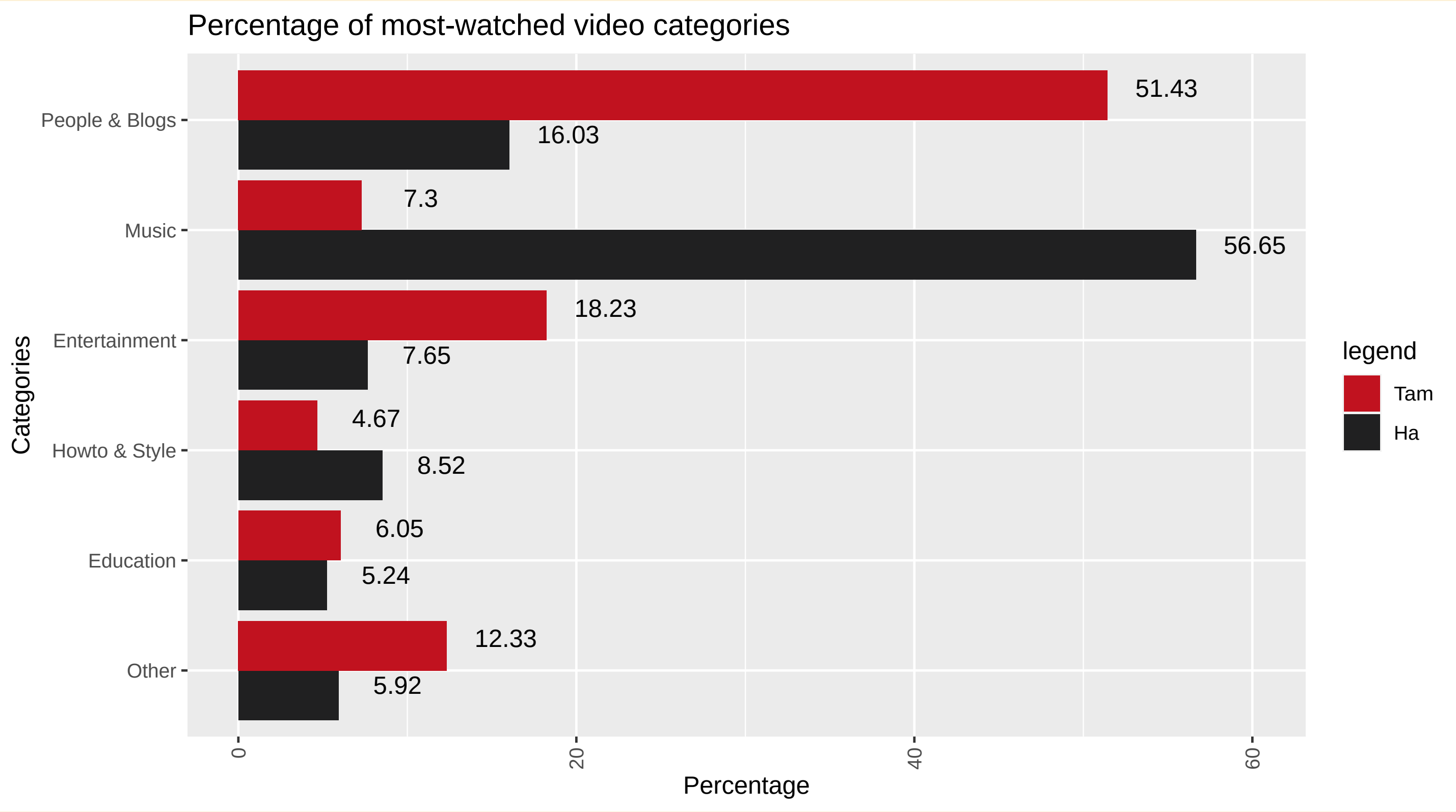
# The trend throughout the years...



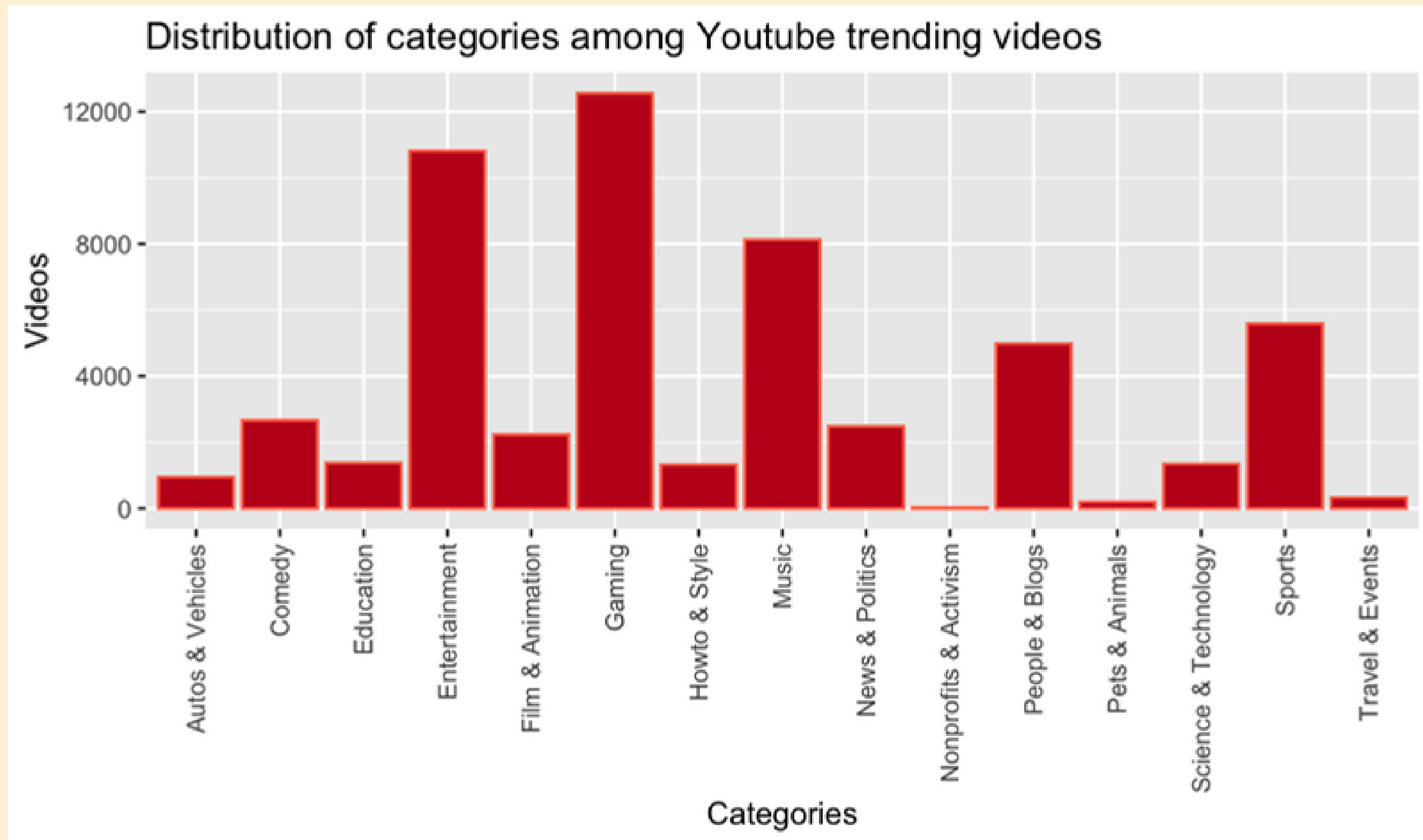
# Let's zoom in 2022's data!



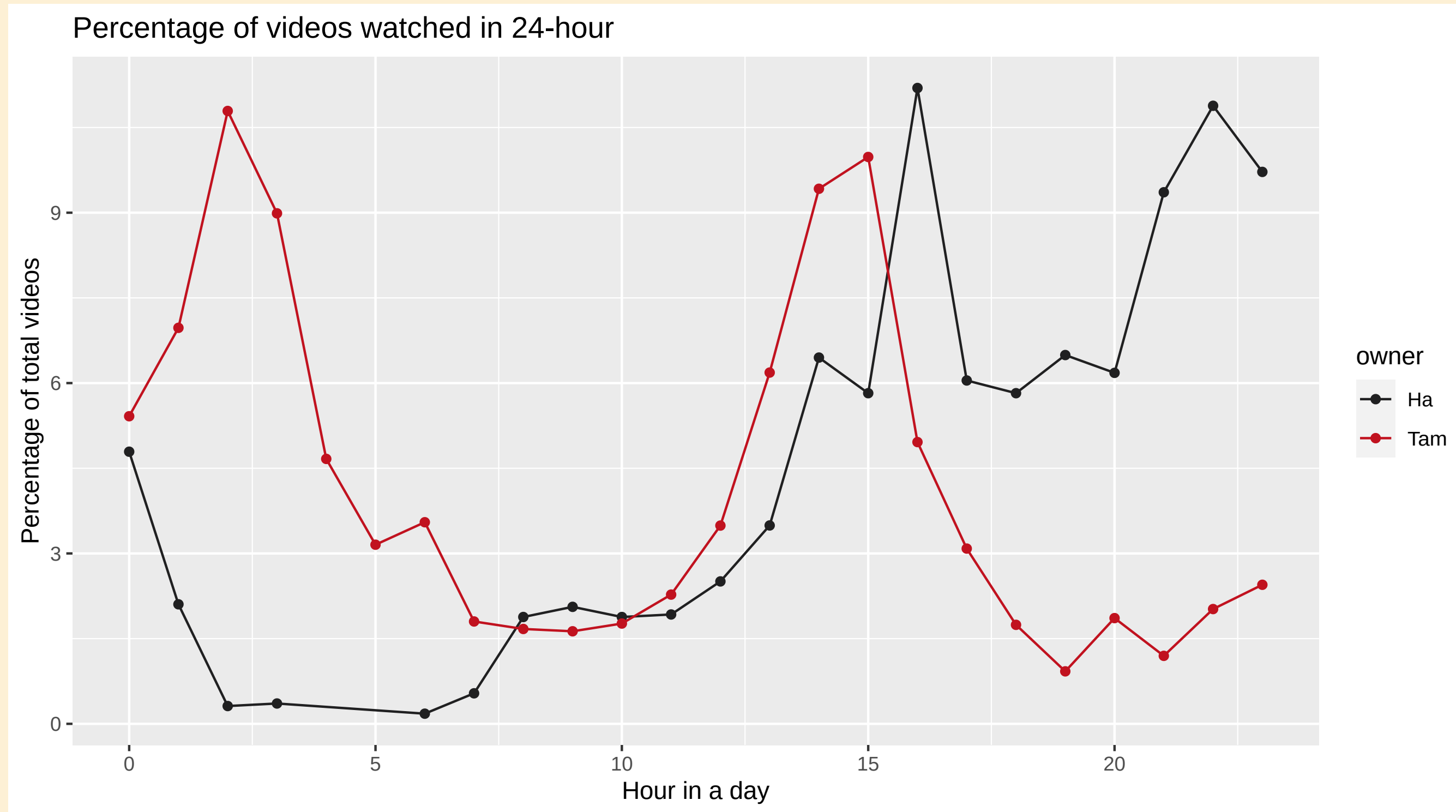
# Let's look at what we watch the most!



# But the most trending categories are:



# In a day, we watched Youtube the most at...





# References

- How to get Youtube video data:  
<https://tinyurl.com/mryxhnfr>
- List of YouTube categories with category IDs:  
<https://gist.github.com/dgp/1b24bf2961521bd75d6c>
- Kaggle:  
[https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset?](https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset?select=US_youtube_trending_data.csv)  
[select=US\\_youtube\\_trending\\_data.csv](https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset?select=US_youtube_trending_data.csv)

