Twitch_Stats

Oscar Monroy

4/29/2022

Bringing the Data onto R

```
library(dplyr)
library(ggplot2)
library(stringr)
t <- read.csv("Twitch_game_data.csv") # Converts csv into data frame

sum(duplicated(t)) # Checks to see if there is any duplicated observations
```

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

```
# Now we must add a variable that has the actual name (or rather abbreviation)
# of the months by using the numerical month variable.
t <- t %>%
  mutate(month = month.abb[Month]) # Adds month variable with month abb.

date <- str_c(t$month, " ", t$Year) # Creates new varaible with format "Month Year"

tw <- cbind(t, "date" = date) # Binds date variable to "t" to create "tw"

tw$date <- factor(tw$date, levels = unique(tw$date)) # Changes date variable to factor

head(tw, 10) # Small preview of data frame
```

```
##    Rank                            Game Month Year Hours_watched
## 1     1                League of Legends     1 2016      94377226
## 2     2 Counter-Strike: Global Offensive     1 2016      47832863
## 3     3                          Dota 2     1 2016      45185893
## 4     4                      Hearthstone     1 2016      39936159
## 5     5      Call of Duty: Black Ops III     1 2016      16153057
## 6     6                        Minecraft     1 2016      10231056
## 7     7                World of Warcraft     1 2016       8771452
## 8     8                Z1: Battle Royale     1 2016       7894571
## 9     9            Talk Shows & Podcasts     1 2016       7688369
## 10   10                          FIFA 16     1 2016       6988475
##    Hours_Streamed Peak_viewers Peak_channels Streamers Avg_viewers Avg_channels
## 1   1362044 hours       530270          2903    129172      127021         1833
## 2    830105 hours       372654          2197    120849       64378         1117
```

```
## 3    433397 hours       315083          1100     44074       60815          583
## 4    235903 hours       131357           517     36170       53749          317
## 5   1151578 hours        71639          3620    214054       21740         1549
## 6    490002 hours        64432          1538     88820       13769          659
## 7    342978 hours        46130          1180     33375       11805          461
## 8    205569 hours        41588           460     21396       10625          276
## 9     53235 hours        84051           148     10779       10347           71
## 10   203646 hours       145728           756     46462        9405          274
##    Avg_viewer_ratio month     date
## 1             69.29   Jan Jan 2016
## 2             57.62   Jan Jan 2016
## 3            104.26   Jan Jan 2016
## 4            169.29   Jan Jan 2016
## 5             14.03   Jan Jan 2016
## 6             20.88   Jan Jan 2016
## 7             25.57   Jan Jan 2016
## 8             38.40   Jan Jan 2016
## 9            144.42   Jan Jan 2016
## 10            34.32   Jan Jan 2016
```

Finding the most popular Twitch games of all time

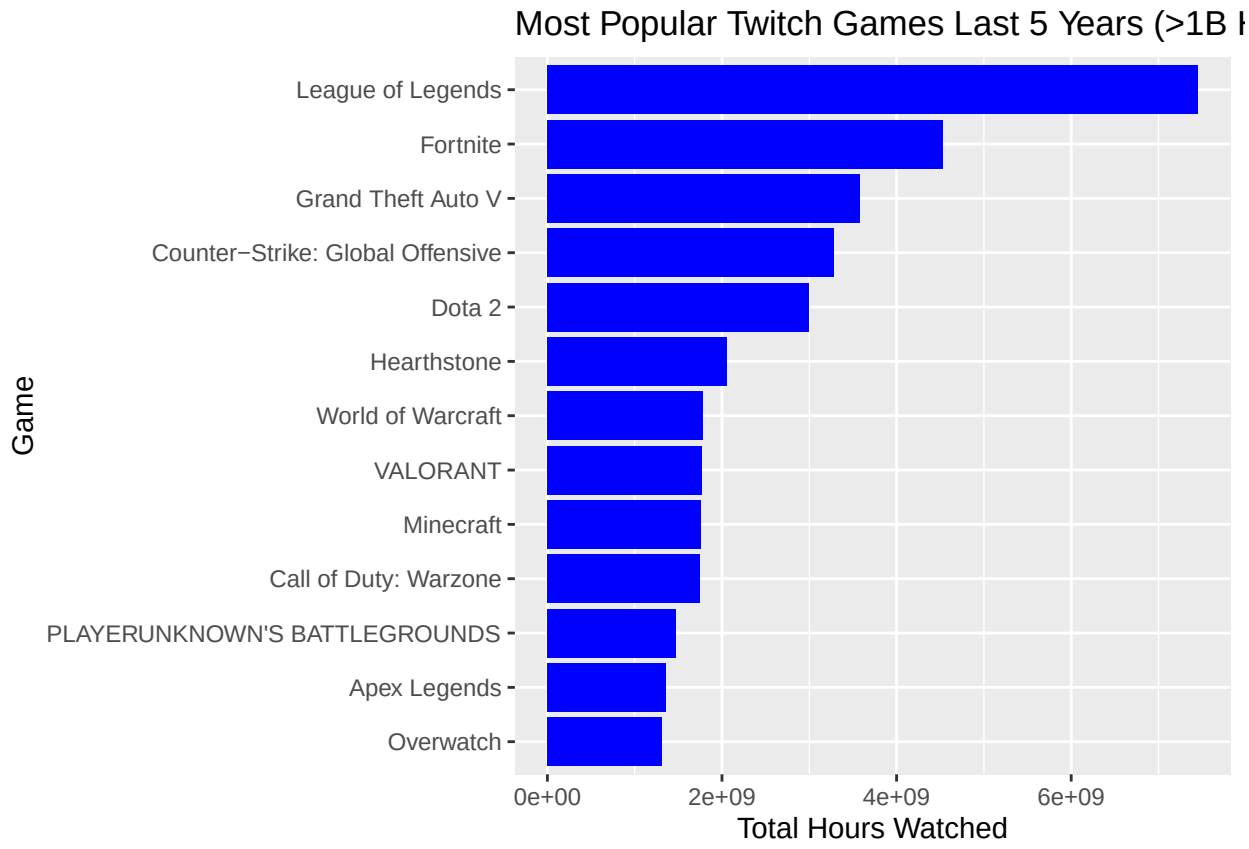
To find the most popular games on this streaming service, we'll need to use the Hours_watched as that is the most clear indicator of audience viewership. We'll also exclude non-game streams like the "Just Chatting" category of streaming and limit the top games to those with over 1 billion hours of views.

```
top <- tw %>%
  select(Game, Hours_watched) %>%
  group_by(Game) %>%
  filter(Game != "Just Chatting") %>%
  summarise(
    Total_Hours_Watched = sum(Hours_watched)
  ) # Gives a new data frame w/ the sum of the Hours_watched variable per game

tw_top <- top %>% arrange(desc(Total_Hours_Watched)) # Descending order

# Now we get games with over 1 billion hours of watch time.
top_games <- tw_top[which(tw_top$Total_Hours_Watched >= 1000000000), ]

top_games
```

```
## # A tibble: 13 x 2
##    Game                          Total_Hours_Watched
##    <fct>                               <dbl>
##  1 League of Legends                7450882313
##  2 Fortnite                         4527130088
##  3 Grand Theft Auto V               3583159500
```

```
##  4 Counter-Strike: Global Offensive          3287101217
##  5 Dota 2                                   2997538494
##  6 Hearthstone                              2057442484
##  7 World of Warcraft                        1784204317
##  8 VALORANT                                 1765325241
##  9 Minecraft                                1760655752
## 10 Call of Duty: Warzone                    1745395328
## 11 PLAYERUNKNOWN'S BATTLEGROUNDS            1475296705
## 12 Apex Legends                             1357189236
## 13 Overwatch                                1306750592
```

```r
ggplot(top_games, aes(x = reorder(Game, Total_Hours_Watched), y = Total_Hours_Watched)) +
  geom_bar(stat = "identity", fill = "blue") +
  coord_flip() +
  ggtitle("Most Popular Twitch Games Last 5 Years (>1B Hrs)") +
  xlab("Game") +
  ylab("Total Hours Watched")
```



It goes without saying that League of Legends is the king of Twitch games. Personally, I'm not exactly sure why the game is insanely popular. I find it very boring to watch and it's a slog just to see people spend 40 minutes trying to level up just to get killed by 5 super common meta characters. But if I were to wager a guess, I'd say it's because Riot Games (creator of League of Legends) started a very strong marketing campaign years ago which included putting popular players in ads and even going as far as making an animated TV show based on the lore of the game.

And now all that marketing pays off with League of Legends being one of the most played, most watched game of all time with currently over 7 billion hours of watch time within 5 years on Twitch alone (and that's only with the current data we have; Twitch began operations in 2011 and League of Legends released in 2009).

As for the rest, there'll be some other games that even the average everyday Joe will recognize. Minecraft and Grand Theft Auto V are two of the highest selling games of all time and are very popular streaming games for the sheer variety of gameplay possible within those game worlds. Fortnite was an overnight success story when Epic Games added a battle royale mode to their simple survival game and overtook PLAYERUNKOWN'S BATTLEGROUNDS, also known as PUBG, as the most popular battle royale game. Now Fortnite is one of the most referenced games in mainstream media. DOTA 2, League of Legend's smaller competitor that is overshadowed by said game, still remains fairly strong in the Twitch charts despite League of Legends. Counter-Strike: Global Offensive, another game made by the same developers as DOTA 2, continues to make waves every once in a while and this is just a remake of a mod from 1999. Hearthstone, World of Warcraft, and Overwatch all make appearances as the most popular streaming games due to the popularity of Blizzard Entertainment games in general. Call of Duty: Warzone simply combined two things people already loved: Call of Duty and battle royale, and now it's one of the more consitently popular games on Twitch. Apex Legends is yet another twist on the popular battle royale genre which features unique, recognizable characters and incredibly fast paced action. And last but not least, Valorant comes in at #8 on this list as it's one of the games that Riot Games recently released. Coming with a pedigree from being made by Riot Games and gameplay similar to Counter-Strike, it made a huge splash on release and continues to be incredibly popular.

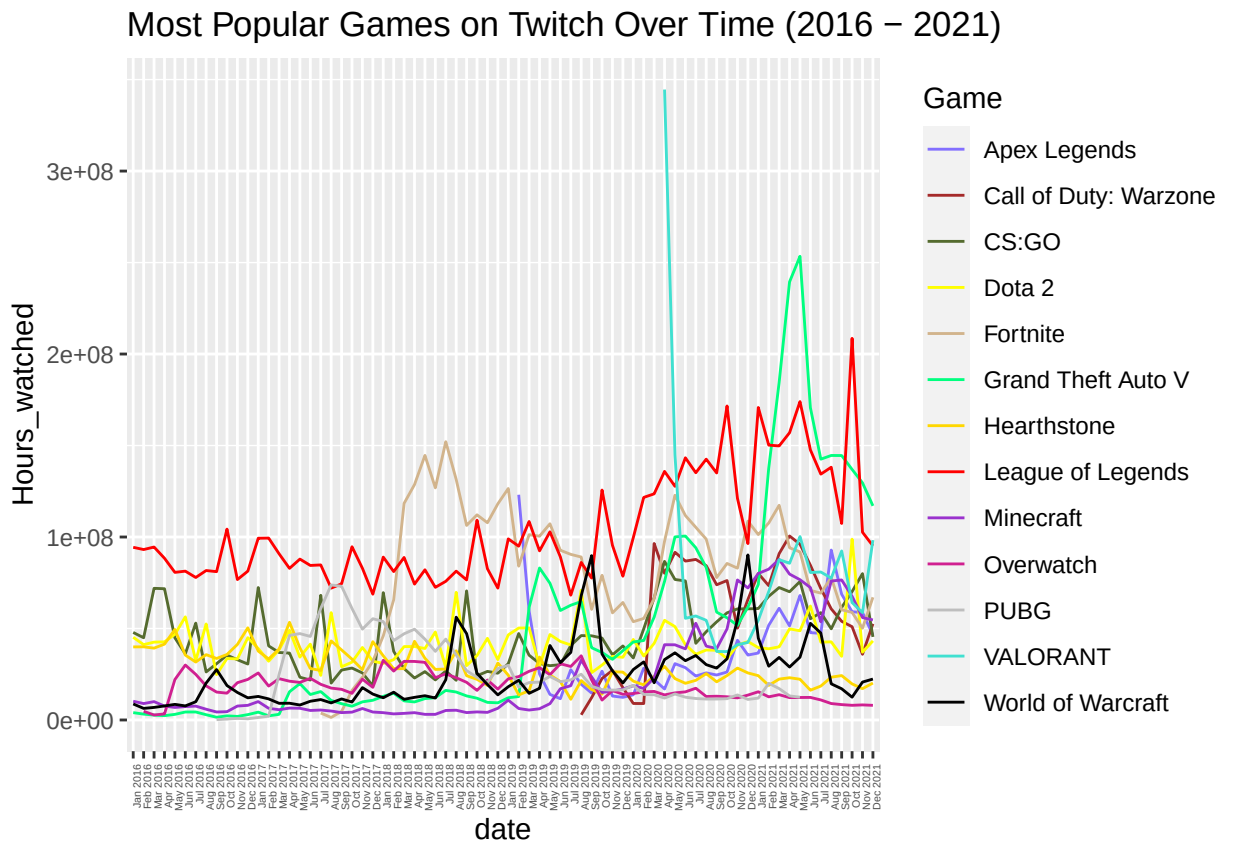
Twitch's Most Popular Games Over Time

Here we'll build a graph to show the evolution of these 13 super popular

Twitch games from 2016 to 2021.

```
# Here I grab 13 handpicked colors for best contrast
cl <- colors()[c(595, 32, 85, 652, 620, 611, 142, 552, 98, 641, 152, 635, 24)]

# We'll get only the data for the top 13 games.
t2 <- tw %>%
  filter(Game %in% as.character(top_games$Game))

# Here I'll just be changing the names of two games to
# their more recognizable abbreviations so they don't
# clog up the graph with their long names.
rn <- gsub(pattern = "PLAYERUNKNOWN'S BATTLEGROUNDS",
           replacement = "PUBG", t2$Game)
rn2 <- gsub(pattern = "Counter-Strike: Global Offensive",
            replacement = "CS:GO", rn)
t2$Game <- rn2

ggplot(t2, aes(x = date, y = Hours_watched, group = Game)) +
  geom_line(aes(color = Game)) +
  scale_colour_manual(values = cl) +
```

```
  ggtitle("Most Popular Games on Twitch Over Time (2016 - 2021)") +
  theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust=1, size = 4))
```



This graph shouldn't surprise anyone. All these games continue to be consistently popular on Twitch. Just look at League of Legends (in red) continue to reign supreme on Twitch. However, let's talk about the elephant in the room. Just look at April 2020. The Valorant beta released in April 2020 and automatically took the world by storm. It was Riot Games' first foray into a completely new genre that they haven't touched: FPS. Riot Games had only been known as the League of Legends developers until Valorant entered the scene with great anticipation. Even though the astronomical success didn't last too long, it continues to be one of the most watched games on the platform and played by tons of people. Grand Theft Auto V also had a massive surge in April-May 2021, though I can't pinpoint the reason for this happening after some research.

Finding More Specific Stats

Here, we'll be looking at specific stats like what game had the most views ever

in a month and which game had the most streamers in a month.

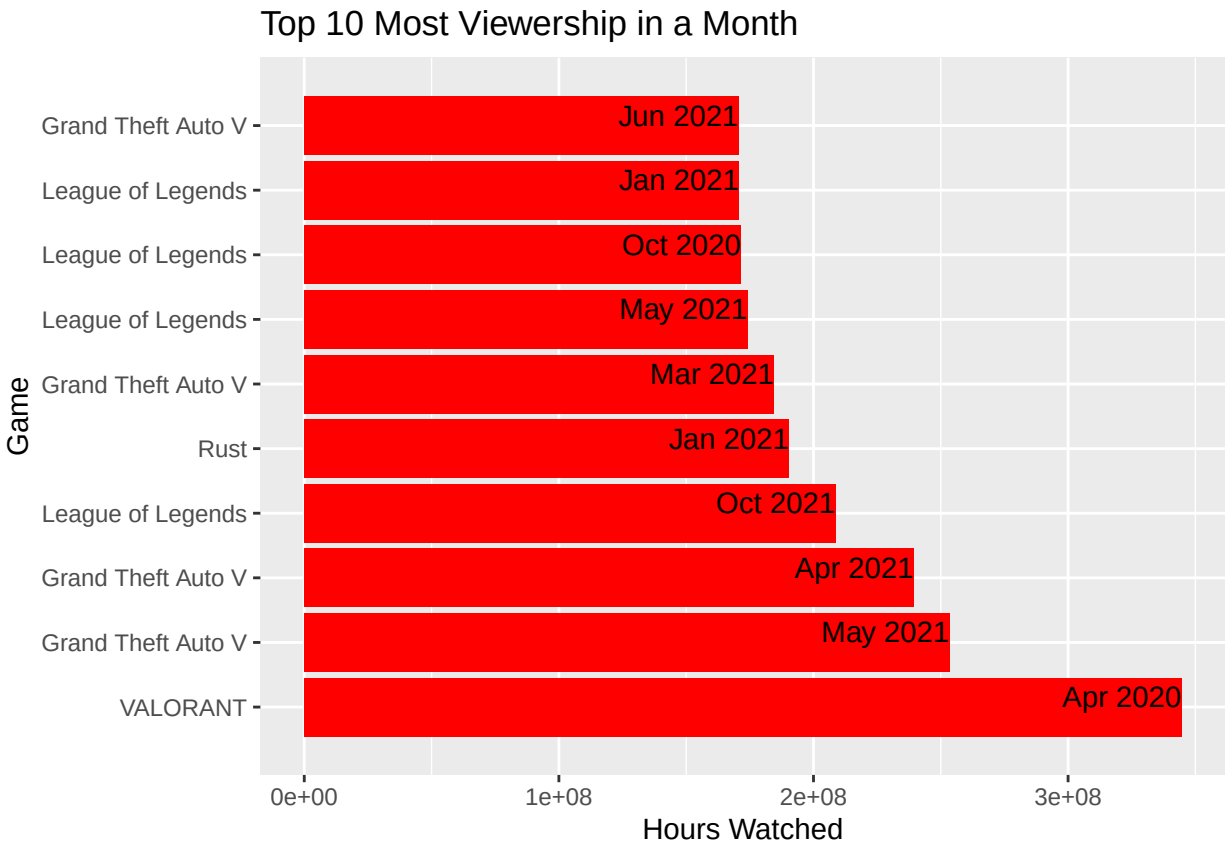
```
# So the game with the most views in a single goes to...
tw[which(tw$Hours_watched == max(tw$Hours_watched)), ]
```

```
##       Rank     Game Month Year Hours_watched Hours_Streamed Peak_viewers
## 10201    1 VALORANT     4 2020     344551979  4588347 hours      1728977
##       Peak_channels Streamers Avg_viewers Avg_channels Avg_viewer_ratio month
## 10201         15710    319709      479209         6381            75.09   Apr
##           date
## 10201 Apr 2020
```

```
# Obviously Valorant in April 2020. Again, just see the graph.
# It had a whopping 344 million hours of watch time, breaking records.

# I suppose this means that Valorant has the most streamers in a single
# month too, but heck, let's see.
tw[which(tw$Streamers == max(tw$Streamers)), ]
```

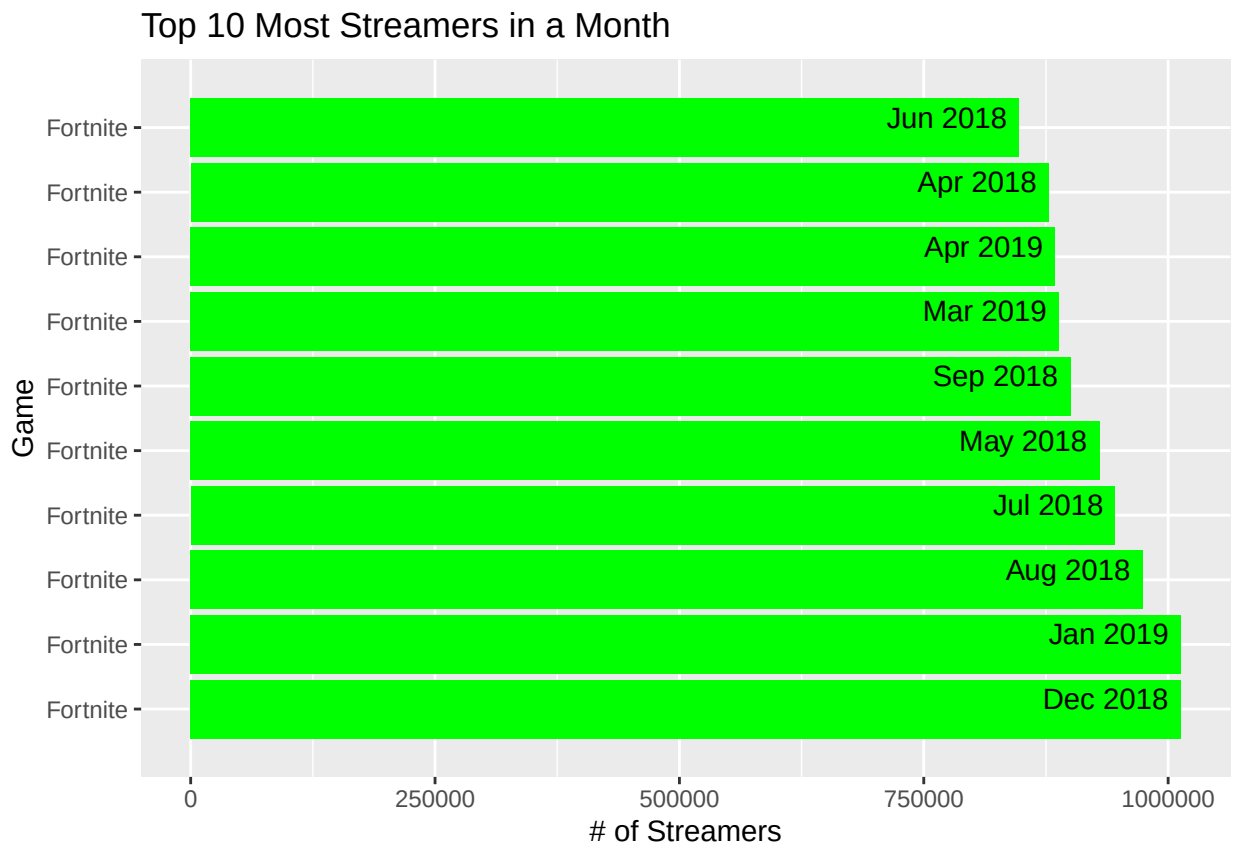
```
##      Rank     Game Month Year Hours_watched Hours_Streamed Peak_viewers
## 7001    1 Fortnite    12 2018     118143183  9027163 hours       418403
##      Peak_channels Streamers Avg_viewers Avg_channels Avg_viewer_ratio month
## 7001         22408   1013029      159008        12149            13.09   Dec
##          date
## 7001 Dec 2018
```

```
# I am honestly surprised, but now that I think about it, it makes sense.
# Despite the record breaking watch time Valorant had, it was still a
# limited access beta that mostly popular streamers had access to. Fortnite
# in 2018 was an unstoppable juggernaut of a game that drew in tons of
# streamers due to to the battle royale aspect and all the pop cultural
# references placed into the game (and outside it; see Marvel x Fortnite crossovers).

# Instead of just picking one game per category, let's grab the top 10.
hw_t10 <- head(tw %>%
                 filter(Game != "Just Chatting") %>%
                 arrange(desc(Hours_watched)), 10)

ggplot(hw_t10, aes(x = 1:nrow(hw_t10), y = Hours_watched)) +
  geom_bar(stat = "identity", fill = "red") +
  coord_flip() +
  ggtitle("Top 10 Most Viewership in a Month") +
  xlab("Game") +
  ylab("Hours Watched") +
  scale_x_discrete(labels = hw_t10$Game, breaks = 1:nrow(hw_t10),
                   limits = 1:nrow(hw_t10), name = "Game") +
  geom_text(aes(label = date), vjust = 0, hjust = 1)
```



No surprises here. GTA:V, League of Legends, and Valorant make up the majority of the most viewed games in a month. However there is a game sticking out like a sore thumb and it's not even in the top 13 most watched games list: Rust. It is an online multiplayer survival game where you play in a sandbox of sorts and build forts and craft weapons/clothing. It has also been released all the way back in 2013. So why the sudden surge of success in Jan. 2021? The Rust developers began a massive rollout of updates for this year beginning in January, and so tons of streamers jumped on the bandwagon along with viewers curious about the update.

```
# Now we take a look at the top 10 games with the most
# streamers in a single month.
s_t10 <- head(tw %>%
                filter(Game != "Just Chatting") %>%
                arrange(desc(Streamers)), 10)

ggplot(s_t10, aes(x = 1:nrow(s_t10), y = Streamers)) +
  geom_bar(stat = "identity", fill = "green") +
  coord_flip() +
  ggtitle("Top 10 Most Streamers in a Month") +
  xlab("Game") +
  ylab("# of Streamers") +
  scale_x_discrete(labels = s_t10$Game, breaks = 1:nrow(s_t10),
                   limits = 1:nrow(s_t10), name = "Game") +
  geom_text(aes(label = date), vjust = 0, hjust = 1.1)
```



Huh, well damnn...

I did mention that Fortnite was an unstoppable juggernaut on Twitch in 2018 (and 2019), but I didn't think it would be quite like this. So people looking at this might be asking "What's with the disparity in number of streamers in a month and total hours watched in a month?" And while I don't officially know the answer to that question, I do know that Ninja (a streamer at one point called the most subscribed-to streamer) and several other popular streamers had a ton of success with streaming Fortnite and more than likely caused an influx of new streamers hoping to replicate those success stories by also streaming Fortnite. Even then, it still seems that League of Legends continues to pull in the most views despite Fortnite drawing streamers in like moths to a light.

Maybe League of Legends and Grand Theft Auto V are the sercrets to
drawing in the most people to a stream if a person is new to streaming.
Maybe not as most of the views in those games are from top streamers.
I wish I had a new, unique streamers variable for this data.

Perhaps we should check the most popular games for streamers

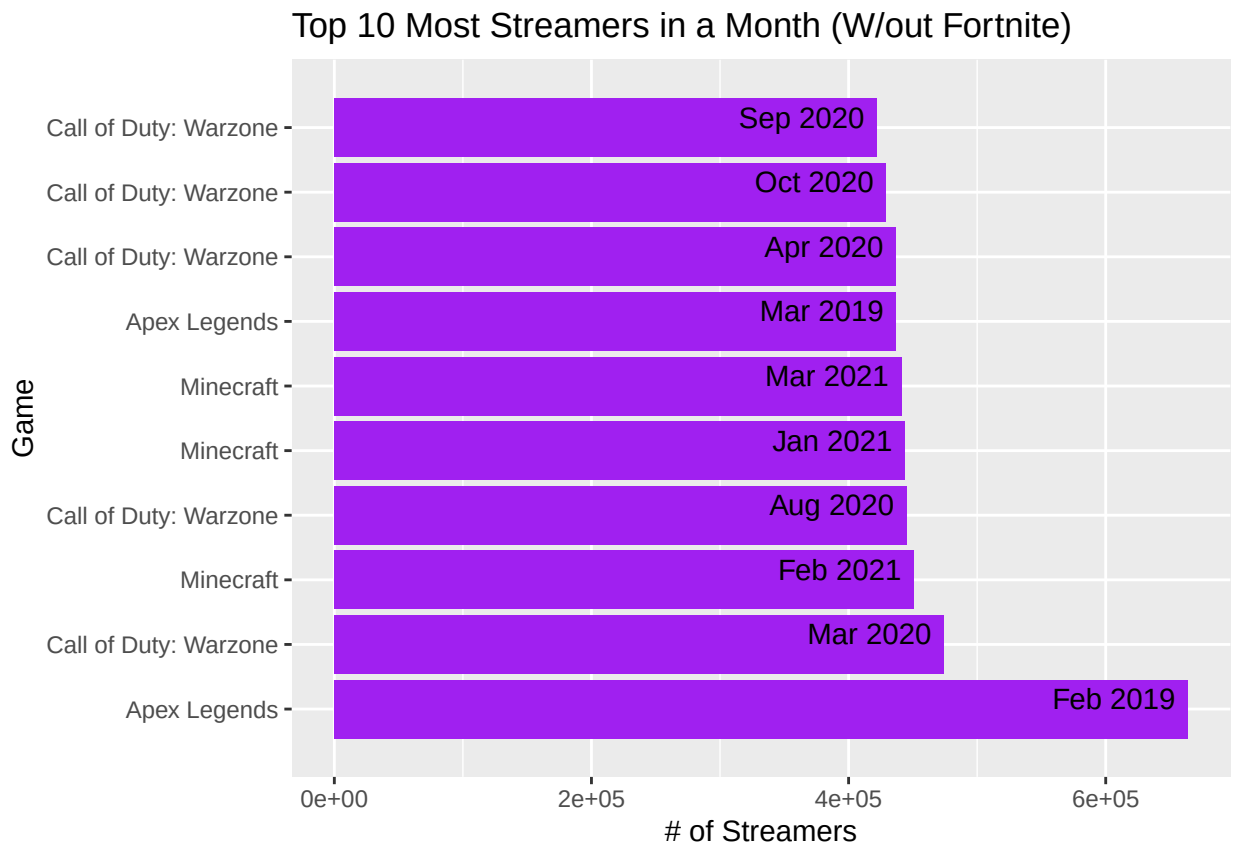
that isn't named Fortnite.

```
s <- tw %>%
    filter(Game != "Just Chatting") %>%
```

```
    arrange(desc(Streamers))

s10 <- head(s[which(s$Game != "Fortnite"), ], 10)
s10[, c(2, 14, 9)]
```

```
##                     Game     date Streamers
## 23          Apex Legends Feb 2019    663781
## 39 Call of Duty: Warzone Mar 2020    474044
## 42             Minecraft Feb 2021    450593
## 43 Call of Duty: Warzone Aug 2020    445223
## 44             Minecraft Jan 2021    443494
## 45             Minecraft Mar 2021    441025
## 47          Apex Legends Mar 2019    437017
## 48 Call of Duty: Warzone Apr 2020    436487
## 50 Call of Duty: Warzone Oct 2020    428944
## 52 Call of Duty: Warzone Sep 2020    422050
```

```
ggplot(s10, aes(x = 1:nrow(s10), y = Streamers)) +
  geom_bar(stat = "identity", fill = "purple") +
  coord_flip() +
  ggtitle("Top 10 Most Streamers in a Month (W/out Fortnite)") +
  xlab("Game") +
  ylab("# of Streamers") +
  scale_x_discrete(labels = s10$Game, breaks = 1:nrow(s10),
                   limits = 1:nrow(s10), name = "Game") +
  geom_text(aes(label = date), vjust = 0, hjust = 1.1)
```



Fortnite's insane popularity with streamers is even more absurd when you look at the index all the way to the left after we filter out Fortnite. The first non-Fortnite game on this list starts at #28 and then it skips to #39 for second place.

Aside from that, we see 3 new games in this list: Apex Legends, CoD: Warzone, and Minecraft. These games are popular with streamers for how easy it is to stream and also due to new streamers trying to replicate the success of top streamers within those games. They also contain tons of action to keep both streamers and viewers invested.