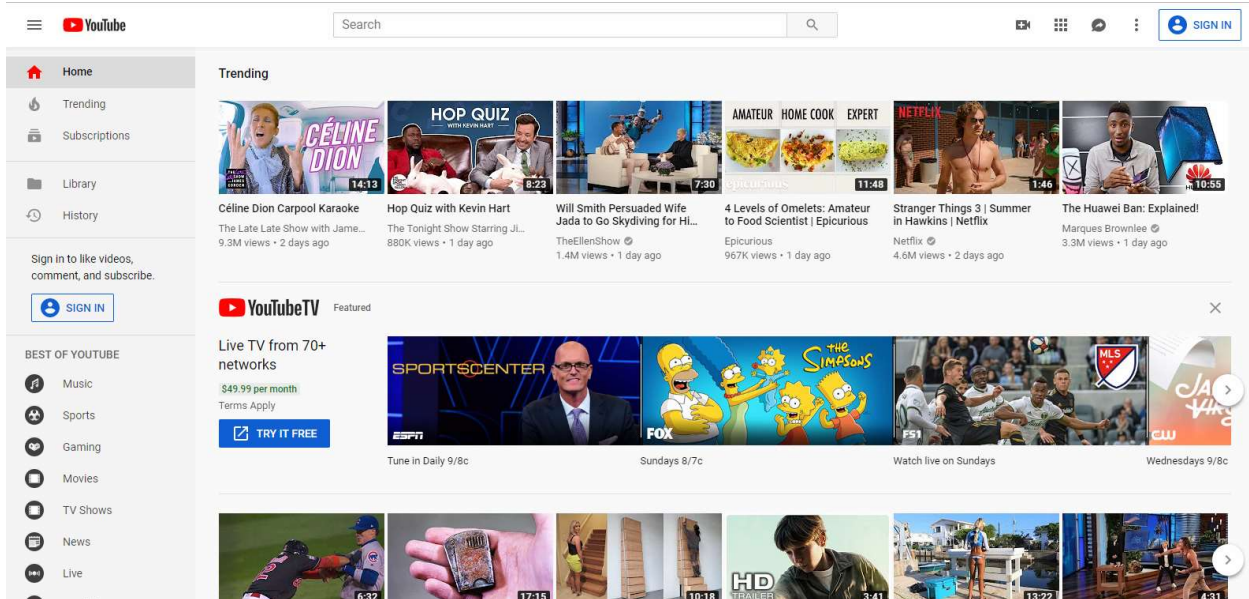


# ENTERING THE DARK SIDE OF YOUTUBE

When you visit youtube.com, the videos you see suggested are the ones Google has decided you will want to click on. These suggestions are boosted to the top because other viewers have made them popular. But according to unofficial sources, there are more than 5 billion videos on YouTube. That's 5 billion videos that you and I will never see. I want to figure out the best way to get into the deepest parts, at the very bottom of that pile.

After logging myself out so that my viewing history and subscriptions are cleared, it looks something like this:



A random word generator will help me conduct random queries. I can use it like this, as shown in the pictures:



jobless police drink distribute kitten disease equal resolute



I'd like to know what effect of the number of random words has on how many views the videos I get from the search have.

If I run a query and get this right away, then I will count the result as NA and try another search string:

 FILTER

Did you mean: **jobless** police drink distribute kitten disease *equally* resolute

Otherwise, I'll sample the last result –the **very** last result.

This is a test run.



Guinea | Wikipedia audio article

wikipedia tts • 31 views • 6 months ago

This is an audio version of the Wikipedia Article: Guinea. Listening is a more natural way of learning, when compared to ...

CC

No more results

What else can I do to increase the unpopularity of the results? I can add in a two-character alphanumeric string from random.org to some of my queries. I will treat this action as a second factor.

Answering some questions:

Why did you log yourself out?

I wanted to view YouTube from a fresh perspective. When sending unusual queries the results won't be so different, but I want the bottom result to be closer to what anyone who goes on YouTube would expect. It may also be better not to let YouTube think I'm interested in Wikipedia TTS videos. Side note: Wikipedia TTS is a really interesting channel. I think the creator might use a bot.

Why the random generators?

Choosing my own searches could add some bias to the kind of results I see. I compared three word generators and chose the one that seemed to have the broadest dictionary and uniform results. It's also a supplement to creativity.

Isn't scrolling until there are no more results a waste of time?

It might be. Hopefully they won't usually take too long to get to. Shortly after starting this I found that holding "page down" is an effective way to continuously request more from the server.

Why are you doing this?

I think that knowing what helps to find obscure content will be useful for people like me who like randomness. It's also nice to bring up something that would not normally get viewed.

And what about Naomi?

Oh, just stay tuned. [https://en.wikipedia.org/wiki/Love\\_of\\_Chair](https://en.wikipedia.org/wiki/Love_of_Chair)

It's time to collect some data.

LOADING

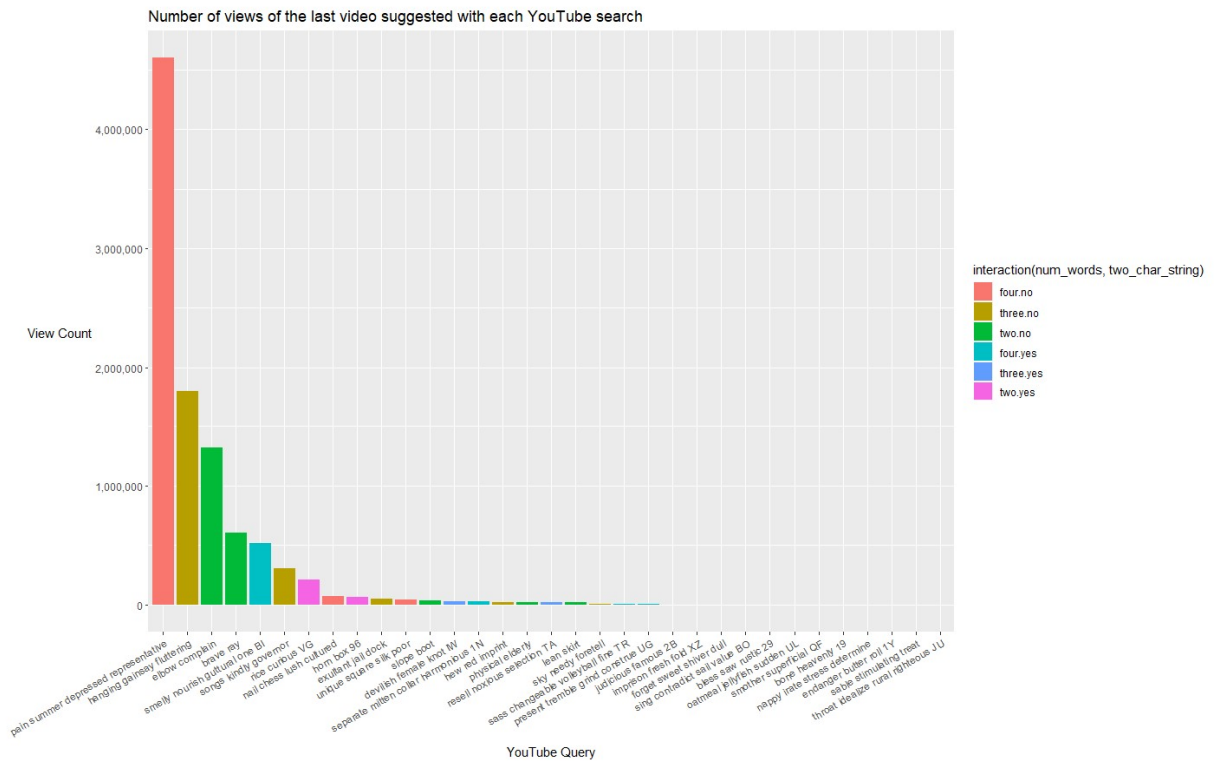
Results:

num_words	two_char_string	search	result_title	result_views	comment
two	no	elbow complain	Slow Loris Tickle (Sonya The Tickle Lover)	1,320,078	
two	no	slope boot	How to Cut a Hole in a Metal Roof for a Vent Stack - Part 1	33,387	The bottom video was flagged because of potentially offensive Nazi content and it wouldn't show the number of views. I used the second last instead.
two	no	lean skirt	jacqueline fernandez hot oops moments on stage	15,754	I'm starting to worry about this experiment
two	no	physical elderly	The Widows Might - Full Film	21,901	I didn't want to click on the bottom result. I wanted to be impartial, but I'm going to take the second last.
two	no	brave ray	Jewelry Commercial - Diamonds by Raymond Lee	606,561	
two	yes	judicious famous 2B	04/27/17 Planning Commission Meeting	755	chmod?
two	yes	rice curious VG	Auburn Coach Wife Kristi Malzahn Agrees with Match & eHarmony: Men are Jerks	208647	
two	yes	bone heavenly 19	Biblical cosmology   Wikipedia audio article	13	
two	yes	horn box 96	Why I will never buy a Ford - Transit problems and nightmares	60,918	
two	yes	smother superficial QF	Pillow Smothering	18	I took the second last result with this one too.
three	no	sky needy foretell	[我的怙我的某] - 第 26 集 / My Sweet Home	7,648	
three	no	hew red imprint	A Conversation with Rick Riordan	21,914	
three	no	songs kindly governor	Yousuf Zulekha Bangla Episode-91 (ইউসুফ জুলেখা, পর্ব- ৯১) drama series Dubbing SATV BD  Part 91	303,243	
three	no	hanging gainsay fluttering	Mili (HD) - Amitabh Bachchan   Jaya Bachchan - Superhit Hindi Movie - (With Eng Subtitles)	1,797,354	
three	no	sable stimulating treat	NA	NA	No matching results

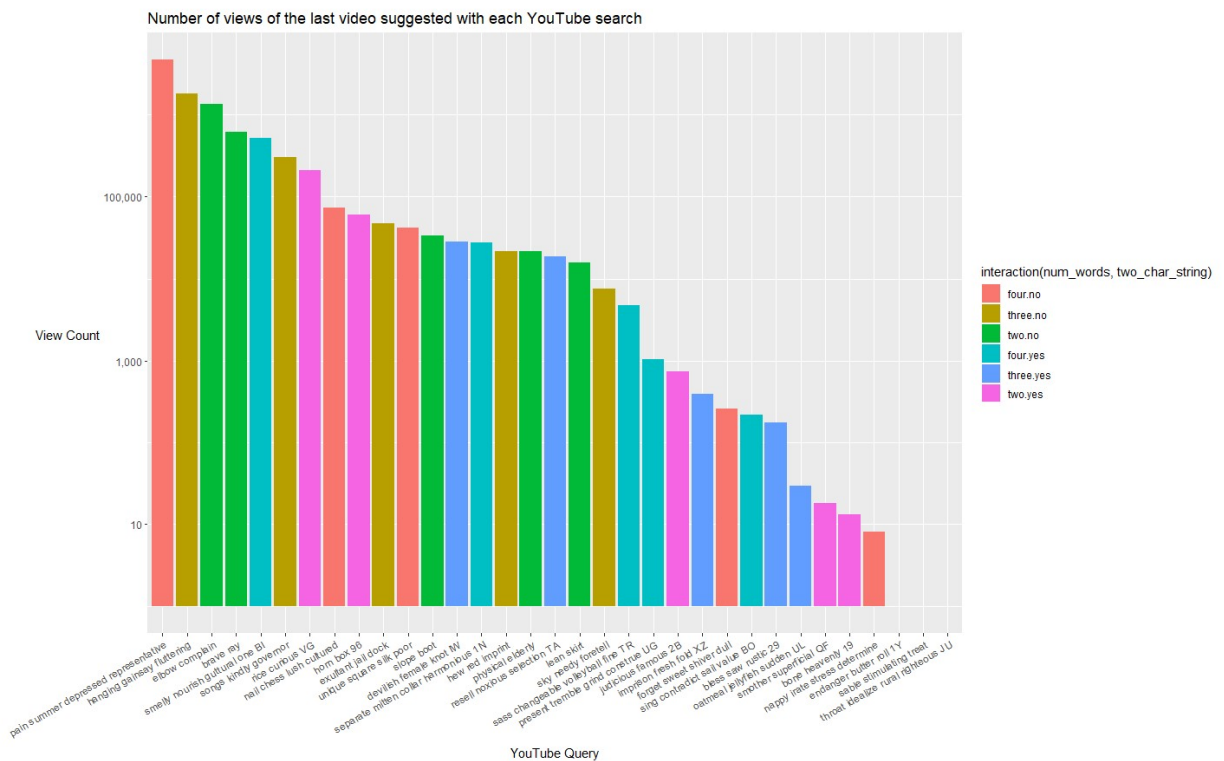
three	no	exultant jail dock	1 guy makes 25 guys run from him during a brawl in Cook County Supermax Prison	47,617	
three	yes	imprison fresh fold XZ	Kínai sátor összehajtás Fura módra :)	390	
three	yes	oatmeal jellyfish sudden UL	Rolled Wheat and FLOOR Rolled Aots and barley	29	
three	yes	endanger butter roll 1Y	NA	NA	No results
three	yes	bless saw rustic 29	Lembranças.	172	
three	yes	resell noxious selection TA	THRIFT HAUL 26 ITEMS - Goodwill & Salvation Army eBay Resale	18,749	
three	yes	devilish female knot IW	NYSTV - The Seven Archangels in the Book of Enoch - 7 Eyes and Spirits of God - Multi Language	28,248	
four	no	nappy irate stress determine	Customer stress....	8	
four	no	unique square silk poor	FanBox Video Earnings	42,293	I skipped two results this time
four	no	forget sweet shiver dull	Free Printable "Prenuptial Agreement" Forms	255	Ok, then
four	no	nail chess lush cultured	Aprenda flor de coração e xadrez	72,744	
four	no	pain summer depressed representative	ZEITGEIST: THE MOVIE   2007 (HD)	4,599,219	
four	yes	sass changeable volleyball fine TR	SASS Tutorial in hindi (Crash Course)	4,763	
four	yes	throat idealize rural righteous JU	NA	NA	No results
four	yes	present tremble grind construe UG	The Woodlanders by Thomas Hardy   Full Audiobook with subtitles   Part 1 of 2	1,056	
four	yes	sing contradict sail value BO	View on Steel Authority of India Ltd, and NMDC Ltd : StockAxis	216	
four	yes	smelly nourish guttural one BI	How to Get Rid of Vaginal Odor, Home remedies for Vaginal Odor, How to get rid of feminine odor	514,729	
four	yes	separate mitten collar harmonious 1N	910 The Man Who Married a Toad , Multi-subtitles	27,398	

I introduced a little bit of bias by avoiding pornographic content, but it is probably not too detrimental to this study. I think that better results might have been found if I made it a rule to always take the third from the bottom instead of the very bottom, because it seems like the videos at the very bottom are extra special in some way or another. I would like to know more about what the algorithm does, but I don't. I've realized that many of these videos that are unpopular should be unpopular, but there are a few that have less attention than they deserve. Like the moon, the dark side of YouTube may not be what you expect or were hoping for.

I will now analyze the results.



View count is a natural ratio that resembles exponential growth, so I will use the logs of the view counts for comparison and analysis.



A two-way ANOVA yields these results:

```
              Df Sum Sq Mean Sq F value Pr(>F)
num_words      2   0.75    0.38   0.030 0.9701
two_char_string 1  67.30   67.30   5.454 0.0282 *
num_words:two_char_string 2  30.86   15.43   1.251 0.3043
Residuals     24 296.13   12.34
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3 observations deleted due to missingness
```

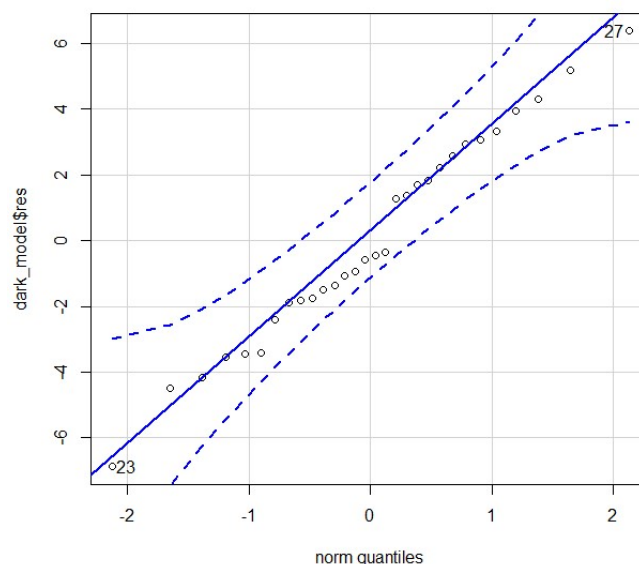
The only factor identified as significant here is whether the random two-character string was added to the search.

I can check the validity of this model by looking at these details:

First from this numerical study, I notice that smallest and largest standard deviations are too far apart, so I will accept the assumption of constant variance.

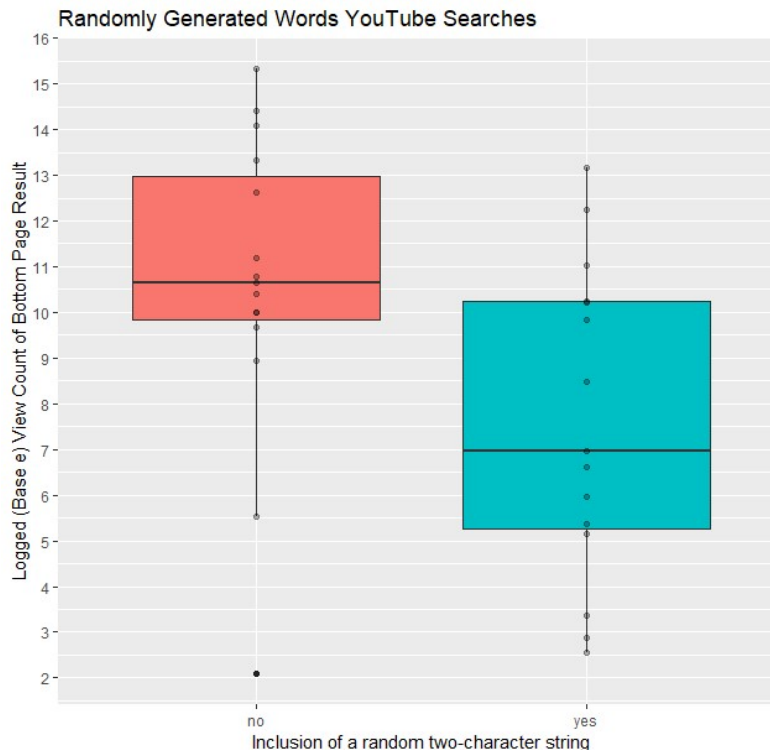
```
> favstats(log(result_views) ~ two_char_string + num_words, data = views)
two_char_string.num_words   min      Q1   median      Q3    max    mean      sd n missing
1          no.four 2.079442 5.541264 10.652377 11.194702 15.34140  8.961836 5.186764 5      0
2          yes.four 5.375278 6.962243  8.468633 10.218225 13.15140  8.835155 3.006481 5      1
3          no.three 8.942199 9.994881 10.770945 12.622290 14.40183 11.346428 2.173790 5      1
4          yes.three 3.367296 5.147494  5.966147  9.838896 10.24878  6.913722 3.011388 5      1
5           no.two  9.664850 9.994288 10.415922 13.315561 14.09320 11.496764 2.051285 5      0
6           yes.two  2.564949 2.890372  6.626718 11.017284 12.24840  7.069544 4.482135 5      0
```

This QQ-Plot shows normally distributed errors. I will assume that the errors are also normally distributed.



The model is found to be valid. The popularities of the video results is different depending on whether the random two-character string is included. These are the same statistics from above, isolated to just the two-character string groups.

```
> favstats(log(result_views) ~ two_char_string, data = views)
two_char_string  min      Q1    median      Q3     max      mean      sd  n missing
1             no 2.079442 9.829569 10.652377 12.96893 15.3414 10.601676 3.418095 15      1
2             yes 2.564949 5.261386  6.962243 10.23350 13.1514  7.606141 3.424460 15      2
```



The average number of views for a video picked from searches containing the two-character string is 2,010 and the average is 40,202 without it. Thus, if you are looking for highly random and unpopular results include a randomly generated alphanumeric string. We know that two-character strings work well, but perhaps other lengths like three or one work even better. That is a worthy investigation requiring more data collection. For now, I could try this for fun:



Yay!