



# Scratch

A Web Scraping Project

Chao Shi

05/09/2017

# Twitch? What? Why?

The screenshot shows a Twitch streamer's profile page for the channel **summit1g**. The page includes a banner image of a League of Legends game, a bio section, and a sidebar with social network integration.

**Summit1g Profile:**

- Videos: 354
- Clips
- Collections
- Events
- Followers: 1,979,289
- Following: 123
- Follow (button)
- Subscribe (button)

**Bio (Overlays):**

- It is a YouTube
- He is popular
- You can watch ads or sub to pay him
- Bob
- Streamer
- Watching now
- Total views: 140,083,375
- Or just donate
- 31,960
- ABOUT ME
- DONATE NOW WITH PayPal™
- PC SPECS

**Sidebar (Left):**

- Followed Channels
- Follow channels and easily access your favorites here. Here are a few you might like!
- DisguisedToastHS: Hearthstone
- DewyUnchained: Super Senso
- TwinGalaxiesLive: Smite
- SpeedGaming: The Legend of Zelda: A Link to ...
- Online Friends
- Twitch is better together. Add some Friends to see who's online and chat!
- Growing social network

**Chat (Right):**

- Top Cheer by Slixit: 1000
- xRazeOvation: DR WW
- jhrvy: that arrow tho
- adriez: WW MUNDO AND LADY TARIC
- Hopsin\_funkvolume: saved
- Jimboslicee: Dat ash arrow
- d0m3sto: BARON
- BOOKeeper: I feel so bad for players in bronze
- Komposite: DAT ASHE
- Leppo420: master lee!
- P3tro1um: 🌈
- LordMariusGrim: MAN???? LADY WAT IS TARIC AHHAAH
- xModernSkater: lmao!!!
- TexasLonghornz: Baron?
- xXHoTWiReXx: sumqueisha?
- Maplegogogo: TAKE FREE BARON
- Titan3736: Warwick
- dontcallmechuck24: Baron
- TriforceOfChozo1: @summit1g SUMMIT RAIDBOSS!!!
- asuuuhdudee: THAT WAS CRINGE LMFAOOOO
- burl3yb0y: ASHE SNIPE
- ForrestGump95: I WANT YOU BRO

**Bottom Buttons:**

- Send a message
- Chat



Get a \$300 trial credit today

## Amazon's \$970 million purchase of Twitch makes so much sense now: It's all about the cloud



Matt Weinberger



✉

Mar. 16, 2016, 3:53 PM

29,215



FACEBOOK



LINKEDIN



TWITTER



twitch



PRINT

- [Discover The Vast Unique Collections At World...  
\(Cost Plus World Market\)](#)
- [Earn a Master's Degree at DePaul University  
\(College of Computing and Digital Media\)](#)

- [This game will keep you up all night!  
\(Vikings: Free Online Game\)](#)
- [Surprising Way to Look Up Birth Records  
\(Ancestry\)](#)

Sponsored Links

Amazon raised a lot of eyebrows when it bought the very popular Twitch game-streaming service for \$970 million in August 2014.

Why it would spend so much cash for Twitch was a real head-scratcher — live game broadcasts on the internet isn't exactly what you would call core to Amazon's retail business.

It was only on March 16, walking around the Game Developers



Amazon founder and CEO Jeff Bezos. Mike Segar/Reuters

# Project Work Flow

Data  
Collection



**Scrapy**

Data Cleaning  
and EDA



**Python  
Jupyter Notebook**

Graph  
App



**R Studio  
Shiny**

# Data Collection: Scraping twtictools.com

<https://www.twitchtools.com/channels?q=&sort=followers&dir=DESC>

The screenshot shows a grid of channel cards. A purple arrow points from the 'Channel Info' tab in the top navigation bar down to the list of channels. The channels listed include Syndicate, Riot Games, summit1g, ESL\_CSGO, Nightblue3, imaqtpie, LIRIK, sodapoppin, TSM\_Bjergsen, CaptainSparklez, TSM\_Dyrus, JoshOG, GoldGlove, Gosu, Castro\_1021, boxbox, dreamhackcs, Trick2g, TimTheTatman, nl\_Kripp, Swiftor, GamesDoneQuick, SivHD, Doublelift, ijerichoii, shroud, TSM\_TheOddOne, Voyboy, Pewdiepie, and ELEAGUE TV. Each card displays the channel name, logo, and follower count.



<https://www.twitchtools.com/channel/sodapoppin>

The screenshot shows the detailed channel information for 'sodapoppin'. A purple arrow points from the 'Channel Stats' section down to the 'Teams Joined' section. The 'Channel Stats' section includes a summary of general information such as display name, account unique ID, channel followers, channel views, mature content status, twitch partner status, last game, account creation date, account update date, and twitch URL. The 'Teams Joined' section lists several teams: The Hammer Squad (52 members), Best Team NA (37 members), Polaris (23 members), Serious Gaming (14 members), Northern Gaming (18 members), and The Milk Men (5 members). Each team has a small logo and member count.

Loop over ~500 pages  
30 channels per page  
get channel names and url components

Loop over ~1500 urls  
get channel statistics and teams joined

Do the same for Team Info tab



# Key Data Columns Used and Derived

Columns from the scraped data used in this demonstration

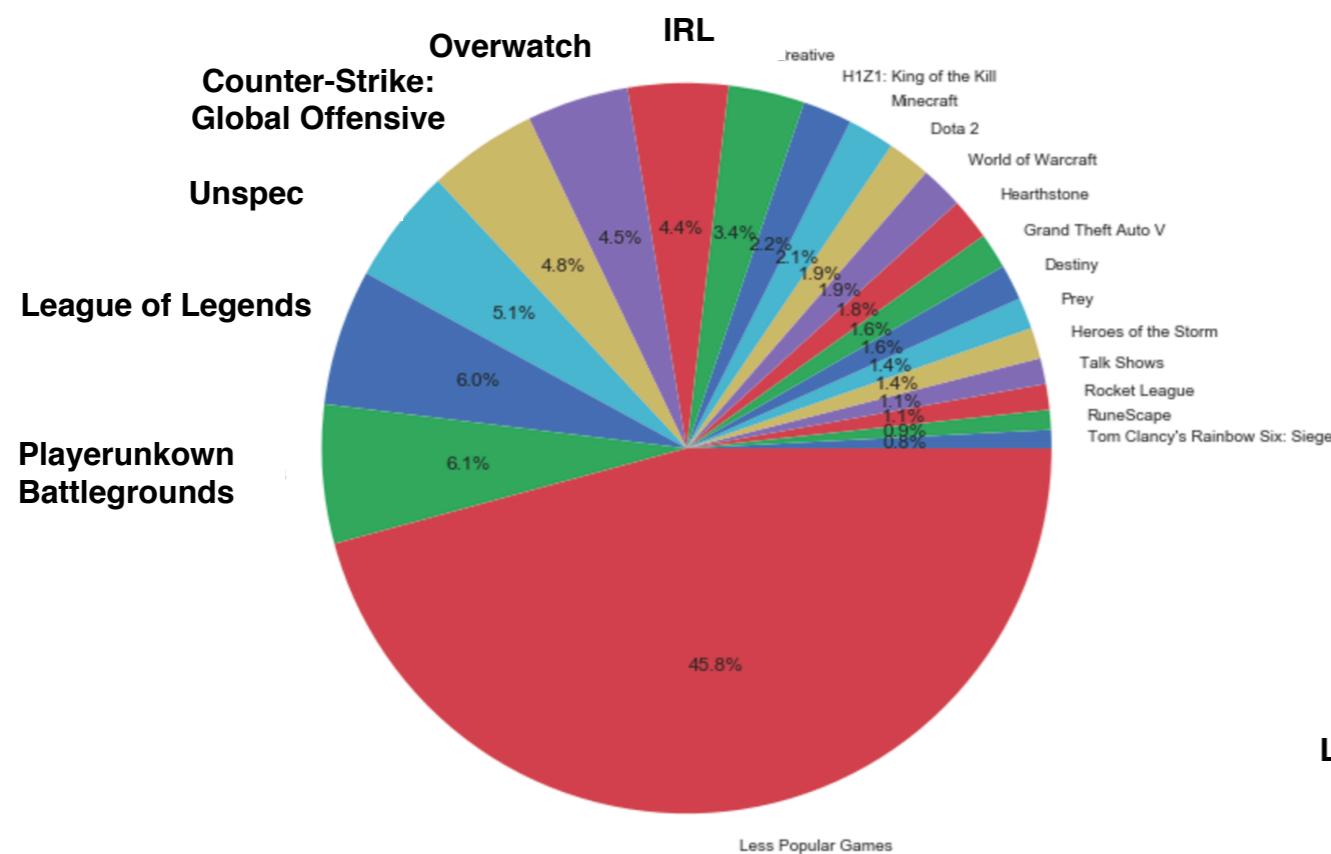
channel ID	teams joined	last game	account date info	followers	views		
display_name	engid	teams_joined	last_game	account_updated_date	account_created_date	channel_followers	channel_views
RouzVega	rouzvega		League of Legends	2017-05-07 14:34:00	2014-09-17 13:27:00	58290	882208
manuelferraratv	manuelferraratv		H1Z1: King of the Kill	2017-05-07 08:34:00	2016-08-10 19:53:00	58248	1420836
ESPORTNOW	esportnow	Polskie Strumyki;FantasyEX	Counter-Strike: Global Of	2017-05-07 14:35:00	2015-04-30 13:44:00	58188	8242040
Princess_PwnZ	princess_pwnz		PLAYERUNKNOWN'S B/	2017-05-07 17:06:00	2015-10-04 22:33:00	58171	837150
Tempo_Storm	tempo_storm	Tempo Storm	Hearthstone	2017-05-07 14:32:00	2015-01-11 20:35:00	58102	4029937
Frodan	frodan	Tempo Storm	Hearthstone	2017-05-07 16:02:00	2010-07-07 03:14:00	58050	1422397
scudpunk	scudpunk	rustNchill.com;DBH;The Ma	Less Popular Games	2017-05-07 14:30:00	2014-10-06 19:51:00	58006	1800907
jennyeatsbabies	jennyeatsbabies		League of Legends	2017-05-07 14:30:00	2012-05-16 00:32:00	57996	3302820
StressLevelZero	stresslevelzero		Creative	2017-05-07 08:31:00	2014-08-11 06:25:00	57995	748444
ScufJumpn	scufjumpn	Denial eSports	H1Z1: King of the Kill	2017-05-07 15:01:00	2014-06-13 15:04:00	57931	313584
ruhub_dotapit	ruhub_dotapit	RuHub	Dota 2	2017-05-07 14:31:00	2015-06-29 20:16:00	57880	7121322
minerva	minerva	CNB HyperX	League of Legends	2017-05-07 06:32:00	2013-09-27 14:22:00	57849	2068327
Mithrain	mithrain		PLAYERUNKNOWN'S B/	2017-05-07 14:06:00	2015-01-10 17:26:00	57814	2059887
tonkaaaP	tonkaaaap	Team Daddy	Less Popular Games	2017-05-07 14:35:00	2015-12-23 19:58:00	57786	4554136
Blazifyy	blazifyy		Counter-Strike: Global Of	2017-05-07 00:00:00	2014-05-04 22:02:00	57734	73535
Behzinga	behzinga	Sidemen	Overwatch	2017-05-07 12:30:00	2012-04-06 15:26:00	57594	82400
羅傑	roger9527	Team NVIDIA TW	Hearthstone	2017-05-07 14:34:00	2014-07-05 05:11:00	57549	4280036
CarcinogenSDA	carcinogensda	SpeedRunsLive;Team Lude	Less Popular Games	2017-05-07 13:33:00	2010-08-22 13:53:00	57423	4490197
兔頭	tim885885	★Jteam★	IRL	2017-05-07 06:31:00	2010-02-04 15:06:00	57376	5370114
Grimmax	grimmax	Game And Play TV	IRL	2017-05-07 14:35:00	2010-09-30 11:33:00	57330	2087945
Thorlar	thorlar	The Asgardians;Twitch Adm	Less Popular Games	2017-05-07 14:35:00	2012-08-26 16:02:00	57311	4867002
vikared	vikared		Dota 2	2017-05-07 15:02:00	2014-03-31 15:03:00	57261	1984984
Nathanias	nathanias	Digital Chaos	Less Popular Games	2017-05-07 12:31:00	2010-12-14 01:01:00	57240	6898224

Columns derived

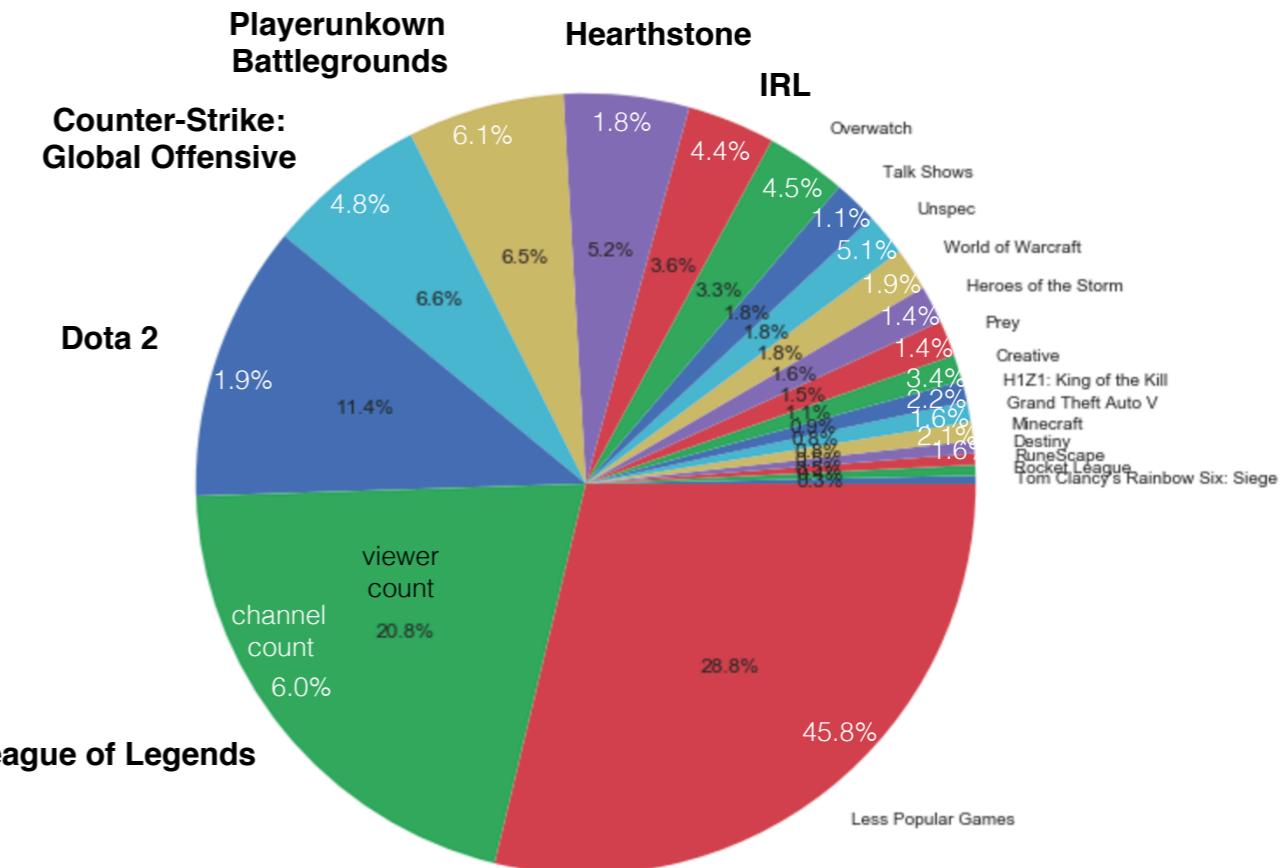
channel days	view per follower	view per day	follower per day	view_per_follower_per_day	nteams
963	15.134808715	916.10384216	60.5295950156	0.015716312269	0
269	24.3928718583	5281.91821561	216.535315985	0.0906798210344	0
738	141.645012717	11168.0758808	78.8455284553	0.191930911541	2
580	14.3911914872	1443.36206897	100.294827586	0.0248123991158	0
846	69.3596950191	4763.51891253	68.6784869976	0.0819854551053	1
2496	24.502962963	569.870592949	23.2572115385	0.00981689221273	1
943	31.0469089405	1909.76352068	61.512195122	0.0329235513685	4
1817	56.9490999379	1817.73252614	31.9185470556	0.0313423775112	0
1000	12.9053194241	748.444	57.995	0.0129053194241	0
1058	5.41306036492	296.393194707	54.7551984877	0.00511631414453	1
677	123.035970974	10518.9394387	85.4948301329	0.181737032459	1
1317	35.7538937579	1570.48367502	43.9248291572	0.0271479831115	1
847	35.6295533954	2431.97992916	68.2573789847	0.0420655884243	0
500	78.8103692936	9108.272	115.572	0.157620738587	1
1098	1.27368621609	66.9717668488	52.5810564663	0.00116000566129	0
1856	1.43070458728	44.3965517241	31.03125	0.0007708537647	1
1037	74.3720307912	4127.32497589	55.4956605593	0.0717184482075	1
2449	78.1950960417	1833.48182932	23.4475296039	0.0319293981387	6
2648	93.5951268823	2027.98867069	21.667673716	0.0353455917229	1
2411	36.4197627769	866.007880547	23.7785151389	0.0151056668506	1
1714	84.9226501021	2839.55775963	33.4369894982	0.0495464703046	2
1132	34.6655489775	1753.51943463	50.5839222615	0.0306232764819	0
2336	120.514046122	2953.00684932	24.5034246575	0.0515899170041	1

# Popular Content on Twitch

## Channel Count



## Viewer Count



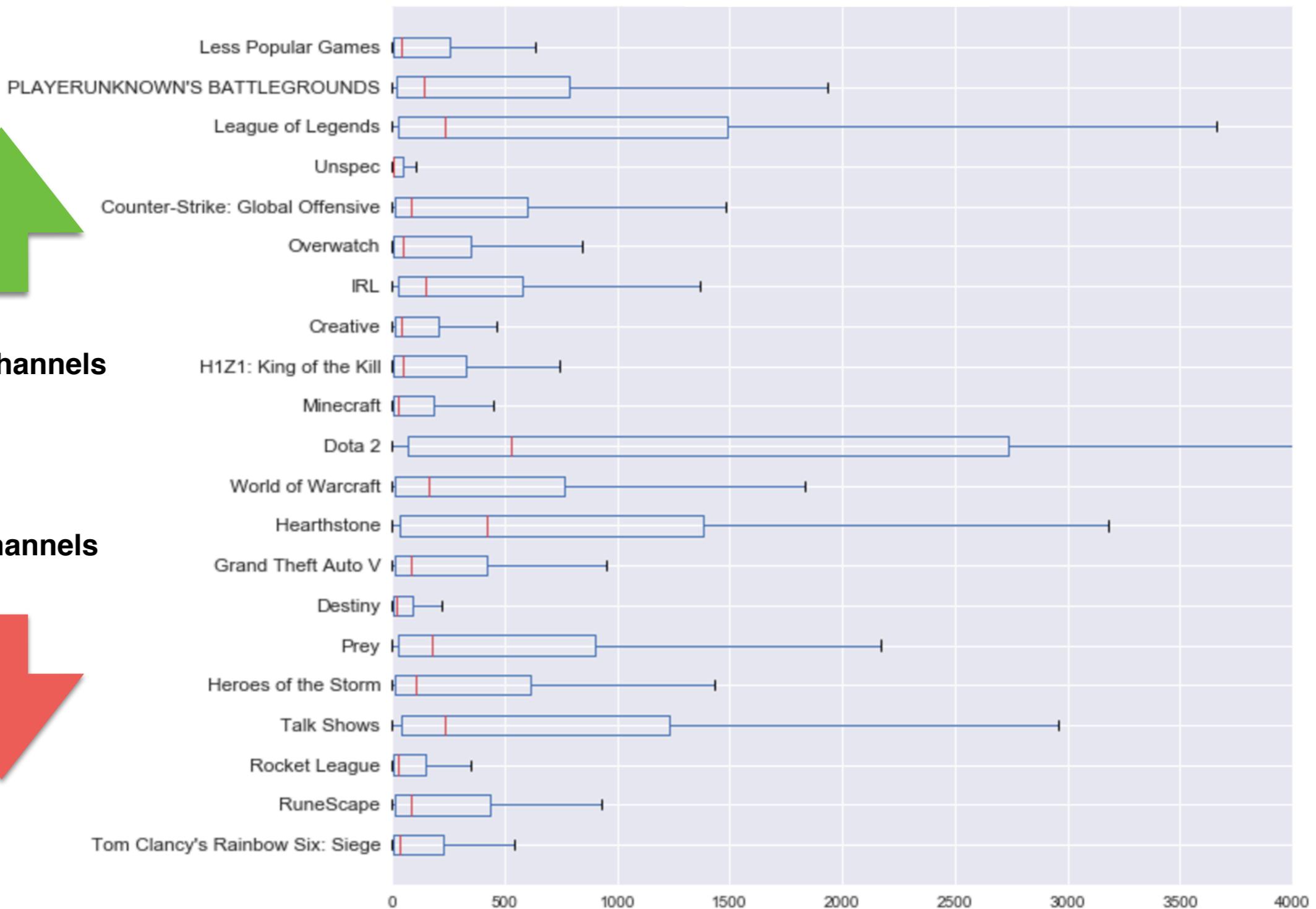
# Best Content for Attracting Viewers?



more channels

less channels

Viewers per Day



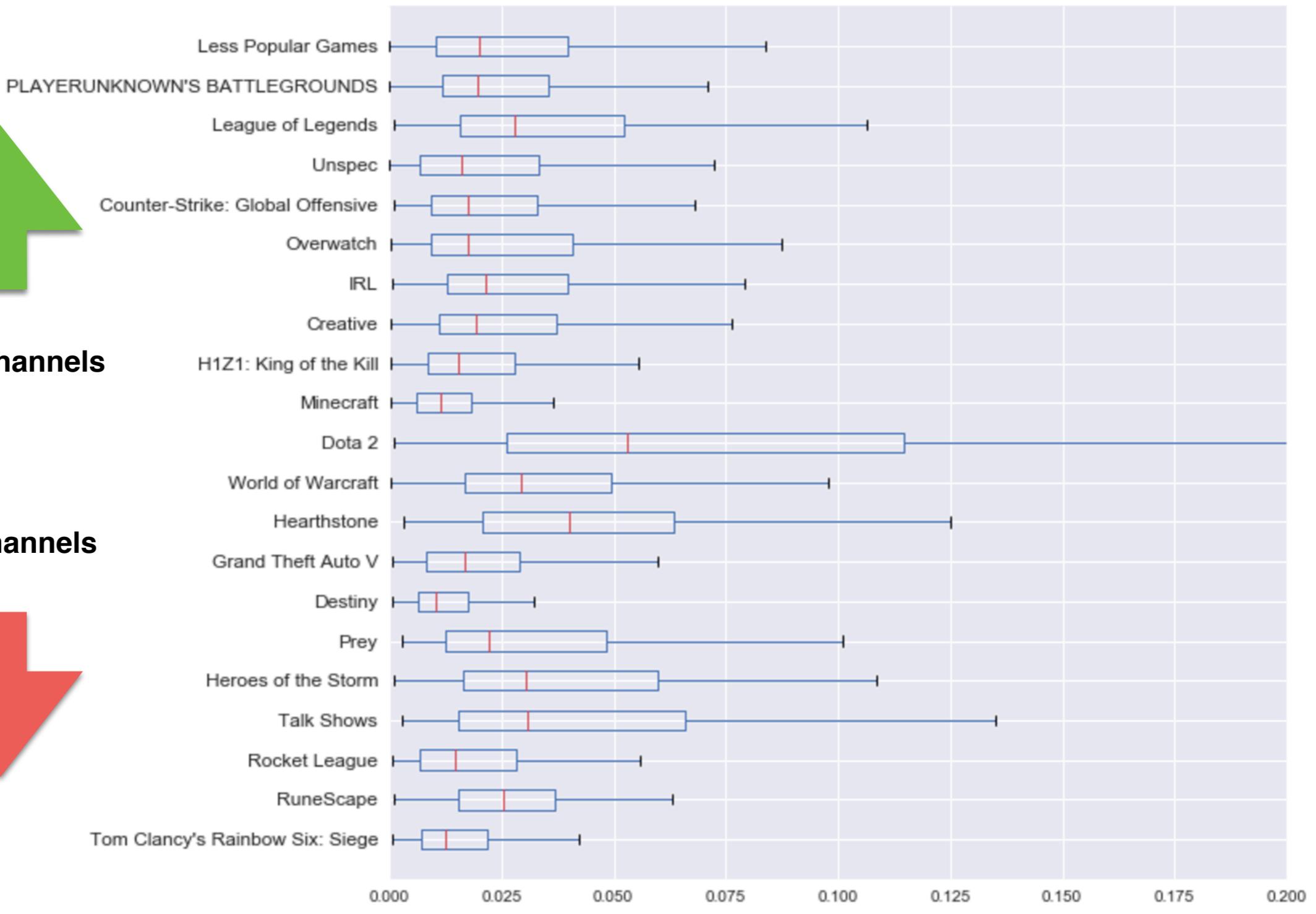
# Viewer “Loyalty” Proxy



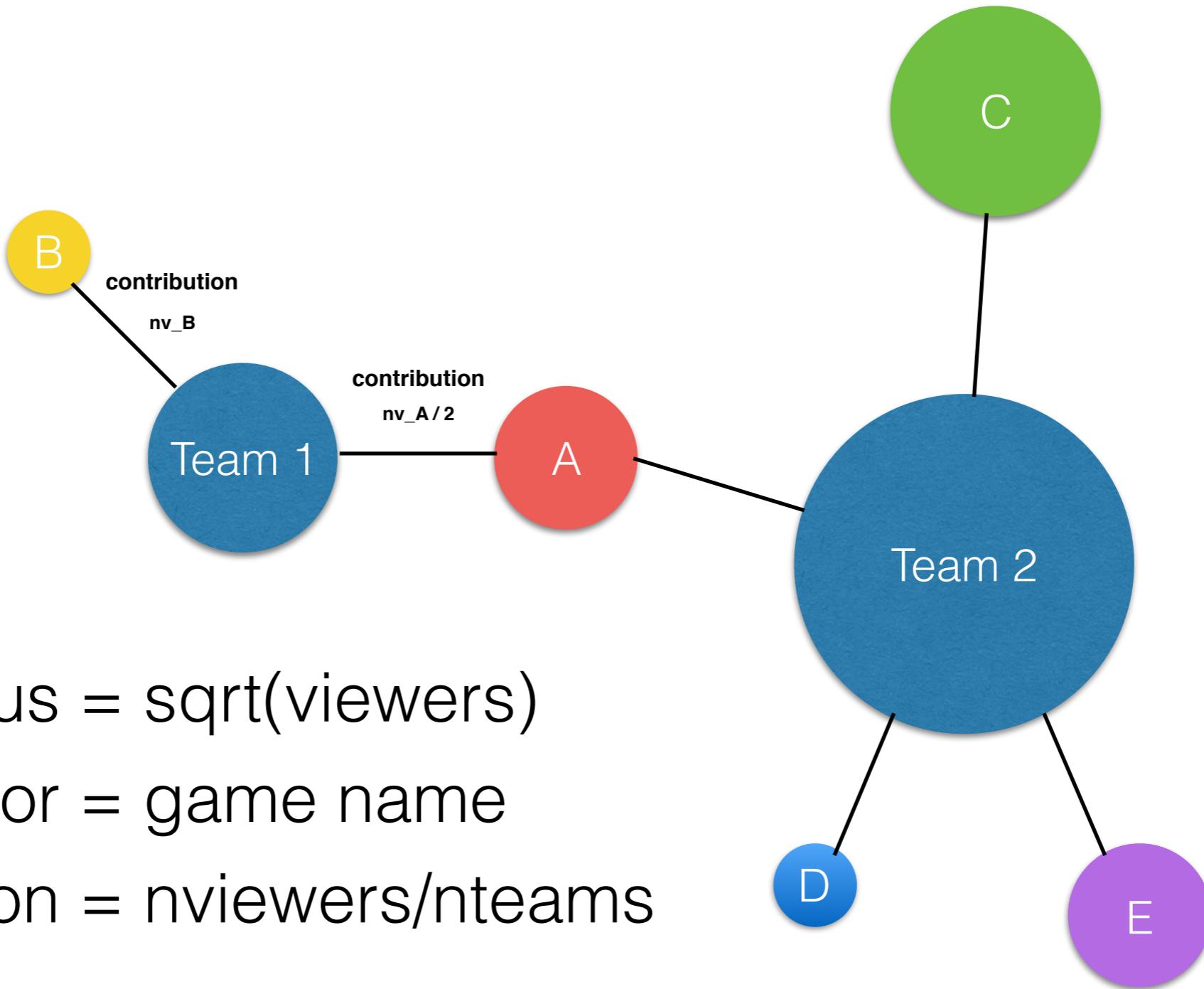
more channels

less channels

Views per Follower per Day



# Want to Sponsor a Channel or a Team?



[https://chaoshi.shinyapps.io/Twitch\\_Graph\\_R\\_Shiny/](https://chaoshi.shinyapps.io/Twitch_Graph_R_Shiny/)

# Recap

Popular Content

Suggested Streaming Content for Attracting Viewers

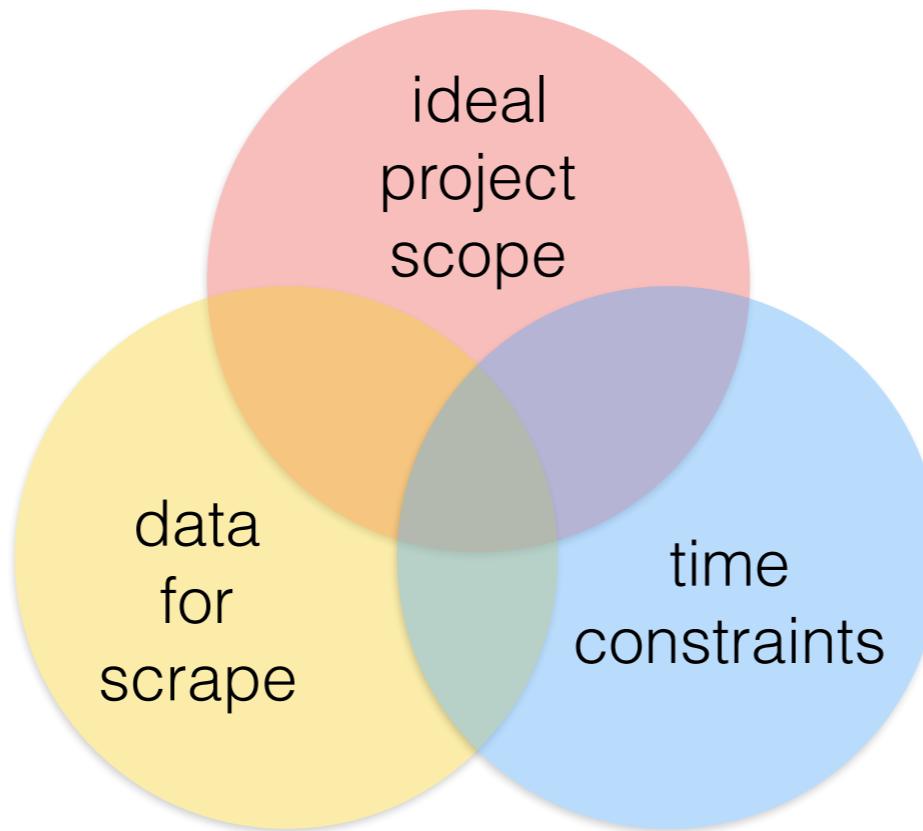
Viewer “Loyalty” Proxy

Channel-Team Interaction + Value Estimation Visualization

# Acknowledgement

- NYC Data Science Academy: Shu Yan, Zeyu Zhang, Luke Lin
- networkD3 developer Christopher Gandrud and others
- d3.js developer the amazing Mike Bostock
- Shiny influence from Joe Cheng

# Lessons Learned + Future



What kind of data is good for graph?  
Did I scrape the right network?