

# CS458/519 Information Visualization Assignment 1

Emmitt Johnson, Cody Malick, and Melissa Sheridan



---

## 1 INTRODUCTION

Twitch.tv is a popular platform for gamers to stream their game play for others to watch and comment. The chat system on Twitch.tv is a great way for viewers to interact with the player, and have allowed for some streamers to gain immense popularity get better from feedback and interact with others. While Twitch.tv is pretty robust for what it is, it does not show statistics that would be useful for the viewers and streamers. There are plenty of statistics that would be interesting, but we will be focusing on the chat statistics as this is the way users interact with the streamer directly. These chats statistic would be useful for the streamer to see what kind of messages they receive, how messages are sent, when the chat was the most active, and who was the most active on the chat and what they sent. This information is huge and would not be easy to parse or understand if it was in a list or a pure database. Visualizing this data will make it more accessible to the users. This information is also based on time, and having a way to see changes over time, or popularity over time, easier to understand. The potential user groups would be people who have an account on Twitch.tv and actively use the platform. Our general approach is to record data from twenty different Twitch streams, and choose the ten most popular (with the most data) for data processing. With the data,

we can then take summary data using Python such as messages per minute, total messages, most popular emoji's, etc. Once we have the summary information, we can use a graphing library such as D3Js to model and visualize our data.

## 2 VISUALIZATION TASKS.

- Which channel had the most active chat?
- What time frame had the most messages sent?
- Who sent the most messages?
- Number of times a message was sent within a time frame?

## 3 DATA SOURCES

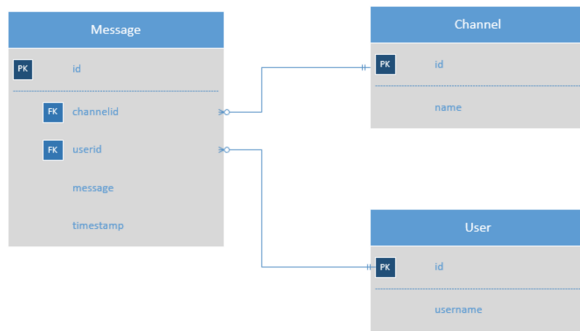
A set of twenty popular streaming channels on Twitch.tv. Twitch chat data across those channels. We will pull one day worth of data from these twenty channels.

## 4 DATA ORGANIZATION IN TERMS OF ER DIAGRAMS

We will be needed a way to format this massive amount of data from Twitch.tv. We decided that it would be best if we were to separate users, channels, and messages. Users and channels will be fairly straight forward with a name and primary key; messages will has a foreign key reference to both a channel and user as well as a time stamp of when the message was sent. This should give us a lot of control over how we complete our visualization tasks.

- 
- Emmitt Johnson.
  - Cody Malick.
  - Melissa Sheridan.

*Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x.*  
*For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.*  
*Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxx/*



## 5 IMPLEMENTATION OF THE ER DIAGRAMS

```

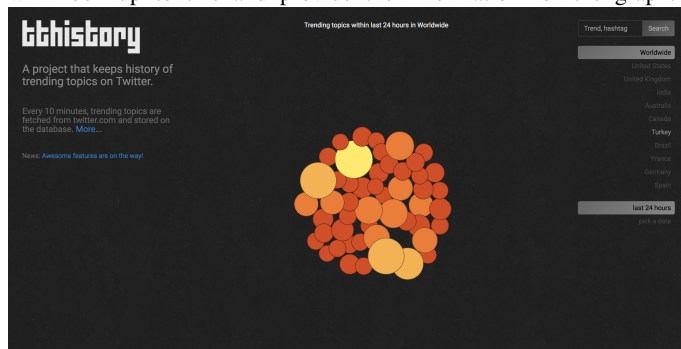
MariaDB [go-twitch-bot]> describe user;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id    | int(11) | NO | PRI | NULL | auto_increment |
| username | text | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
2 rows in set (0.00 sec)

MariaDB [go-twitch-bot]> describe channel;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id    | int(11) | NO | PRI | NULL | auto_increment |
| name  | text | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
2 rows in set (0.00 sec)

MariaDB [go-twitch-bot]> describe message;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id    | int(11) | NO | PRI | NULL | auto_increment |
| channelid | int(11) | YES | MUL | NULL | |
| userid | int(11) | YES | MUL | NULL | |
| message | text | YES | | NULL | |
| timestamp | int(11) | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)
  
```

## 6 A DESIGN OF THE VISUALIZATION INTERFACE

This interface is on D3js. This has the ability to display live information or static information from different categories. In this case it is trending hashtags from Twitter, and it can be sorted by world wide information, you can search for a specific hashtag, or narrowed down to a specific country. In our case we will be able to sort based on channel, message, time, and user. This is an interactive visualization that can be hosted on a website or local server. Our databases will hook up to this and provide the information for the graph.



## ACKNOWLEDGMENTS

The authors wish to thank Twitch.tv.