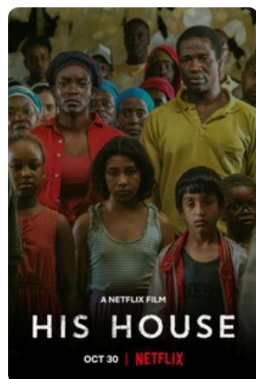# MTH208: Worksheet 7

Web scraping...

In this worksheet, we will continue scraping the web. First we load some packages

```
library(tidyverse)
library(rvest)
```

We will continue with our Movie database example, and extract more detailed information. Any given movie on top 100 movies to watch on Netflix right now, there is a website dedicated to that movie. For example, the movie `His House` has the following website:

Go to the URL: `https://www.rottentomatoes.com/m/his_house` to see what the page looks like.

1. Our goal in this worksheet is to extract the following information for all 100 movies:

   a. Tomatometer score

   b. Popcornmeter

   c. Number of Tomatometer Reviews

   d. Number of Popcornmeter ratings (if over 1000, save as just 1000)

Create a dataset of the final information and save the dataset in an `Rdata` file using command below (where `data` is the name of your object)

```
save(data, file = "TomatoList.Rdata")
```

2. For each movie, notice that there is a poster as well. I am interested in answering the question, does the poster have anything to do with the ratings of a movie. What can you do with the posters to address this question?