

hw6

This homework is not for a grade. It's for cool points. (It's worth 15 cool points, FYI.)

IMDB

For the moment, let's pretend it's 2014, since that's when this dataset was generated.

I, Mr. Moneybags M. TvPerson, have recently acquired copyrights to *all* television shows. There are many of them, including old ones that no longer run. I want to make some decisions about which shows I should invest in. I managed to get this dataset from the IMDB database, which describes, for every episode of every show listed, the popularity ratings of the show.

The columns are:

- **name**: The name of the show, suffixed with the year it first aired. ("Reboots" have the same name, but I consider them different shows, so I just added the first year of every show to the name.)
- **first_year**: the year the show first aired
- **episode_name**: the name of the episode reviewed
- **season**: which season of the show the episode aired in.
- **episode_number**: the episode number within the season
- **star_1**: how many 1-star ratings this episode got on IMDB
- **star_2**: how many 2-star ratings the episode got

...

```
imdb <- read.table("imdb_tv_shows.txt",
  header = TRUE,
  sep = "\t",
  stringsAsFactors = FALSE,
  quote = "", # no characters are used to "quote" entries in columns
  comment.char = "") # no characters are used for comments
```

```
print(head(imdb))
```

```
##           name first_year episode_name season
## 1 $#! My Dad Says 2010      2010      Code Ed      1
## 2 $#! My Dad Says 2010      2010      Corn Star      1
## 3 $#! My Dad Says 2010      2010      Dog Ed Pursuit      1
## 4 $#! My Dad Says 2010      2010      Easy, Writer      1
## 5 $#! My Dad Says 2010      2010      Ed Goes to Court      1
## 6 $#! My Dad Says 2010      2010 Family Dinner for Schmucks      1
## episode_number star_1 star_2 star_3 star_4 star_5 star_6 star_7 star_8
## 1           4      2      2      0      2      8      8      13      13
## 2          14      2      0      2      2      2      7      11      11
## 3           7      2      0      0      2      2      7      7      11
## 4           6      3      0      3      3      3      9      15      9
## 5          15      2      0      0      0      2      7      16      7
## 6          11      2      0      2      2      2      7      11      7
## star_9 star_10
## 1      2      8
## 2      2      7
## 3      7      7
## 4      3      9
```

```
## 5      2      7
## 6      2     11
```

Help me impress my friends and make lots of money

Your goal, should you choose to accept it, is to help me use the data to determine which shows were successful (don't ask me what makes a show successful; I'm just the boss) and which were losers. Some of my friends are data scientists, so there should probably be some statistics involved to impress them. On the other hand, most of my friends like pretty pictures, so some super-sweet data visualizations are a definite plus.