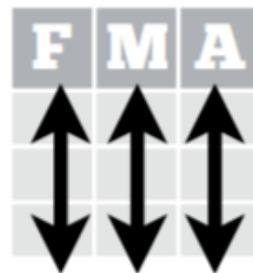








In a tidy
data set:

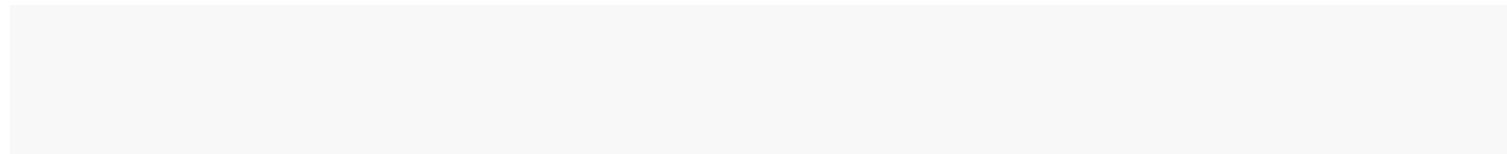


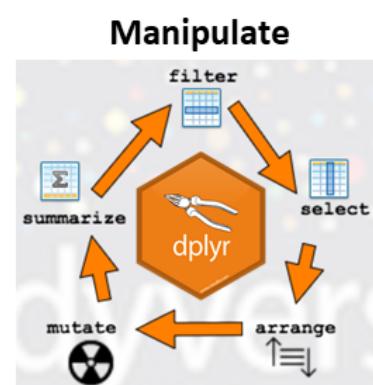
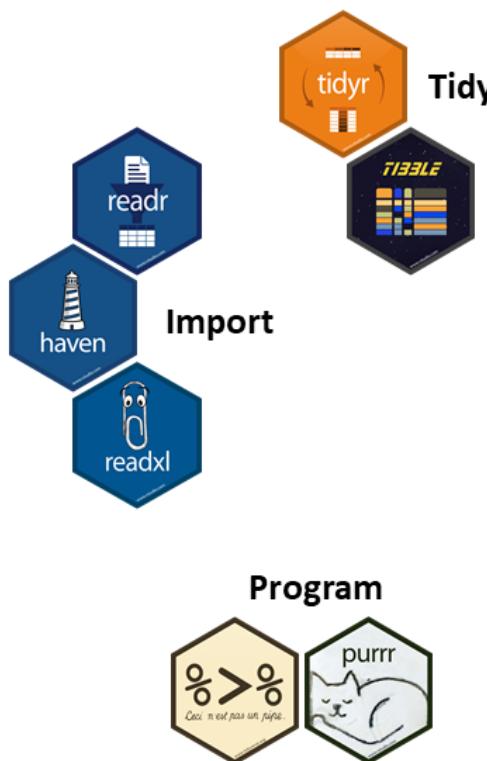
Each **variable** is saved
in its own **column**

&



Each **observation** is
saved in its own **row**





Transform & Clean

Visualize



Model



Dataset

TMDB 5000 Movie Dataset

Metadata on ~5,000 movies from TMDb

The Movie Database (TMDb) • updated a year ago (Version 2)

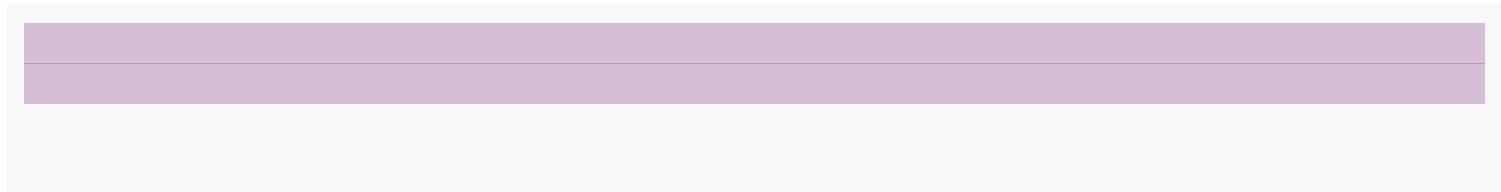
1292 voters

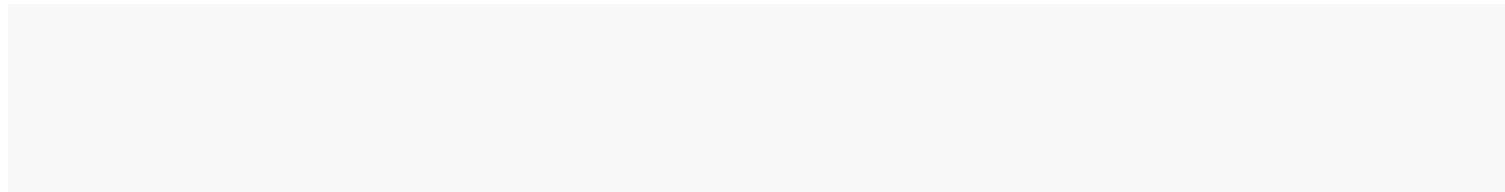
share

Data Overview Kernels (1,417) Discussion (50) Activity Download (9 MB) New Kernel

Data (9 MB) API [kaggle datasets download -d tmdb/tmdb-movie-meta...](#) ? [Download All](#) [X](#)

Data Sources	About this file	Columns
<ul style="list-style-type: none">tmdb_5000_credits.csv 4803 x 4tmdb_5000_movies.csv 4803 x 20	<p>Help us describe this file</p> <p>Edit</p>	<p>A movie_id</p> <p>A title</p> <p>A cast Mark Wahlberg</p> <p>A crew</p> <p>Edit</p>

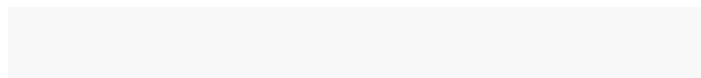








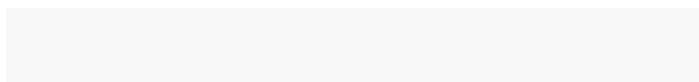
- old way : **verb(subject, complements)**

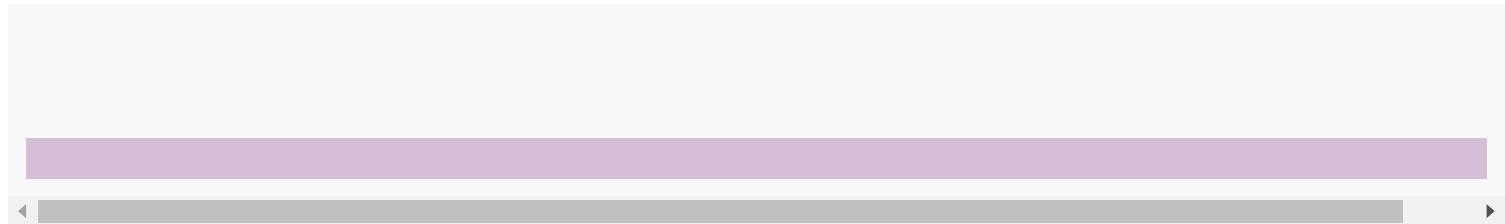


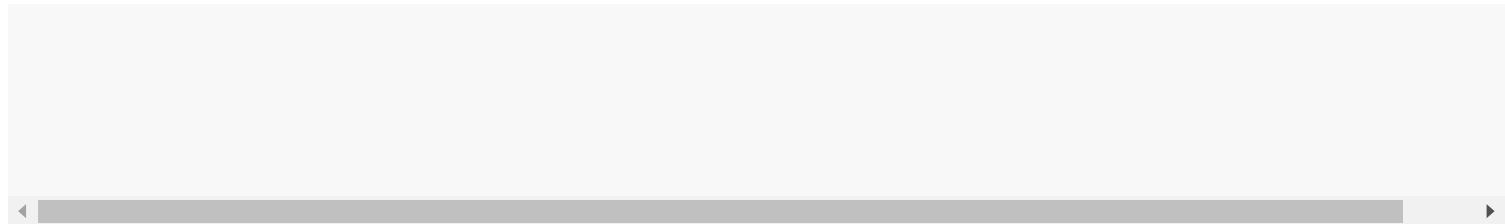


- old way : **verb(subject, complements)**

- pipe way : **subject %>% verb(complements)**



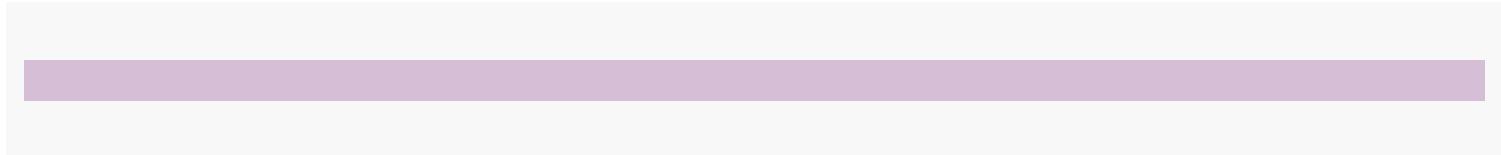




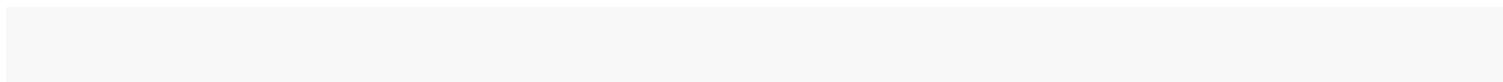


- => Rename **id** to **movie_id** + join both









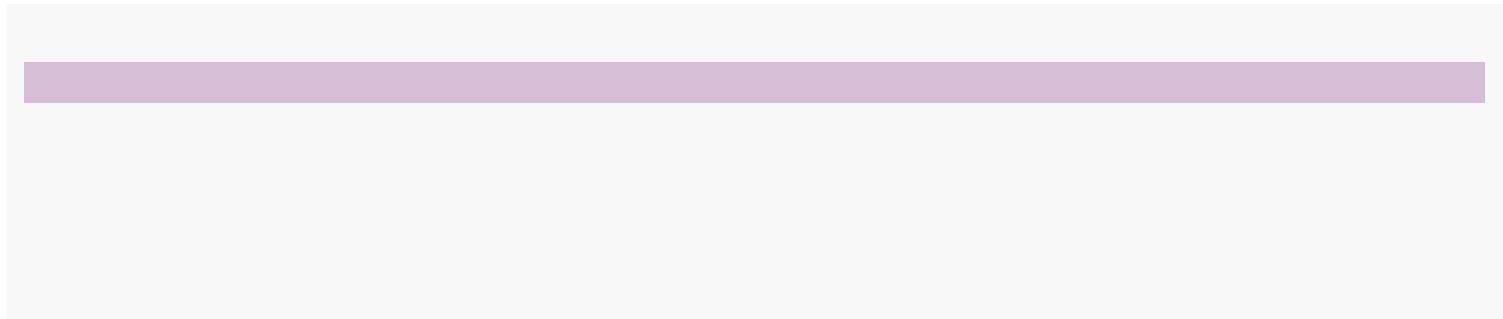


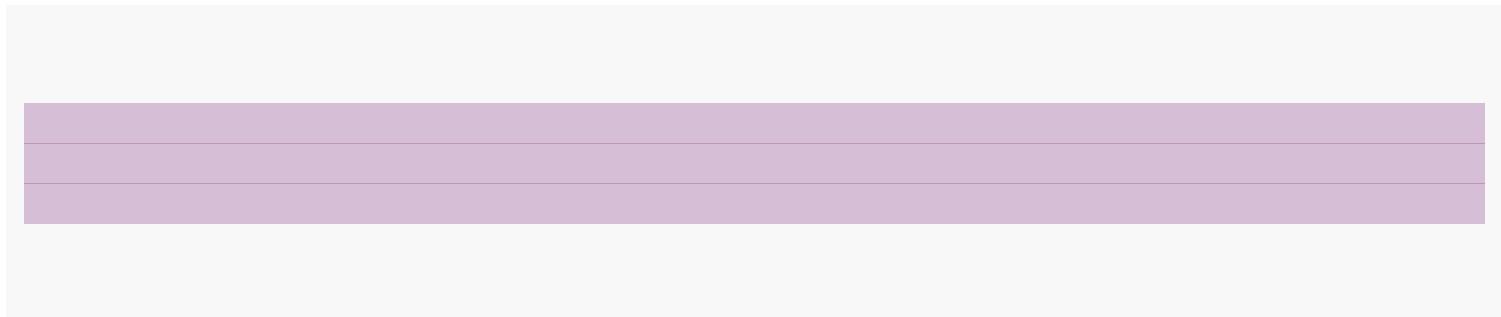
budget	genres	homepage
2.37e+08	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]	http://www.avatarmovie.com/

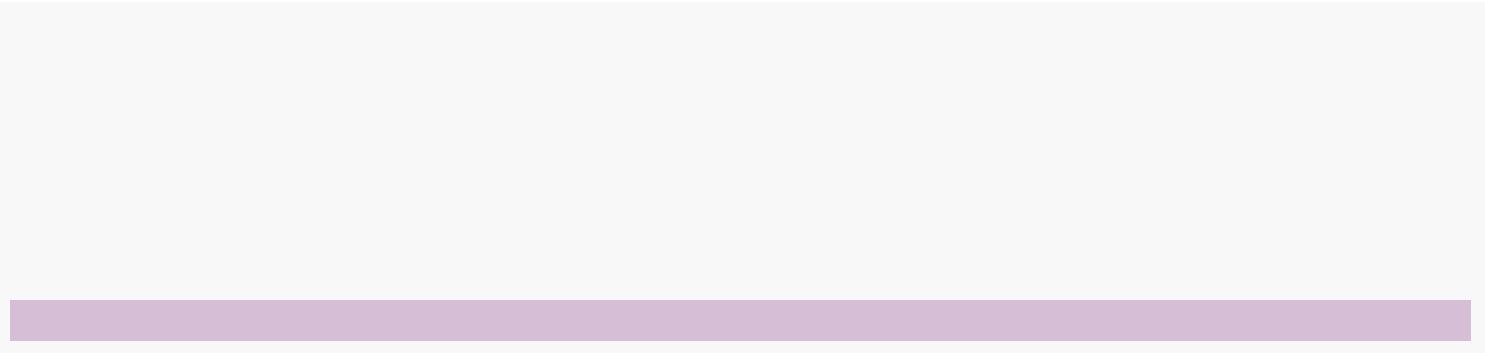




title	budget	vote_average	production_companies
Avatar	2.37e+08	7.2	[{"name": "Ingenious Film Partners", "id": 289}, {"name": "Twentieth Century Fox Film Corporation", "id": 306}, {"name": "Dune Entertainment", "id": 444}, {"name": "Lightstorm Entertainment", "id": 574}]
Pirates of the Caribbean: At World's End	3.00e+08	6.9	[{"name": "Walt Disney Pictures", "id": 2}, {"name": "Jerry Bruckheimer Films", "id": 130}, {"name": "Second Mate Productions", "id": 19936}]
Spectre	2.45e+08	6.3	[{"name": "Columbia Pictures", "id": 5}, {"name": "Danjaq", "id": 10761}, {"name": "B24", "id": 69434}]



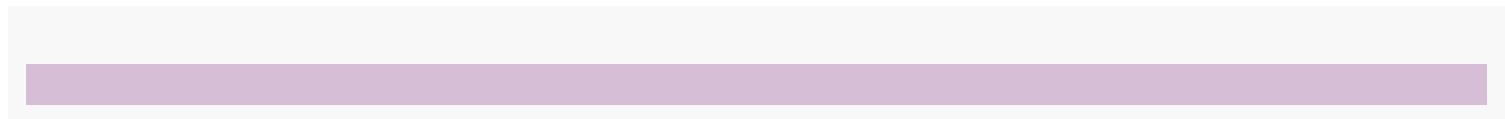




title	budget	vote_average	Production
Avatar	2.37e+08	7.2	Other
Pirates of the Caribbean: At World's End	3.00e+08	6.9	Disney
Spectre	2.45e+08	6.3	Other
The Dark Knight Rises	2.50e+08	7.6	DC
John Carter	2.60e+08	6.1	Disney
Spider-Man 3	2.58e+08	5.9	Marvel
Tangled	2.60e+08	7.4	Disney

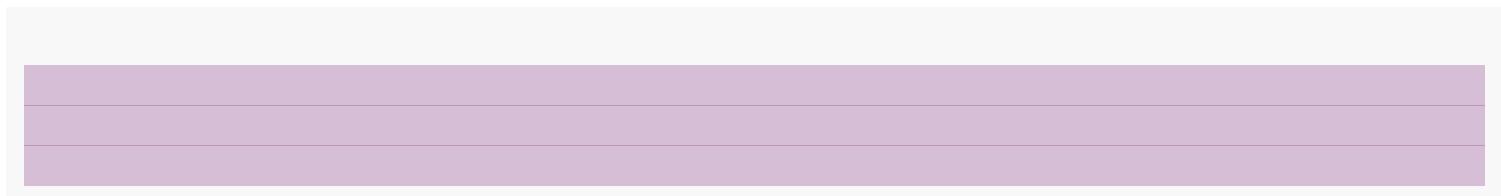
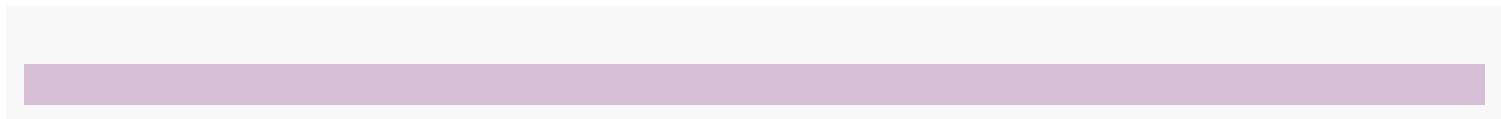


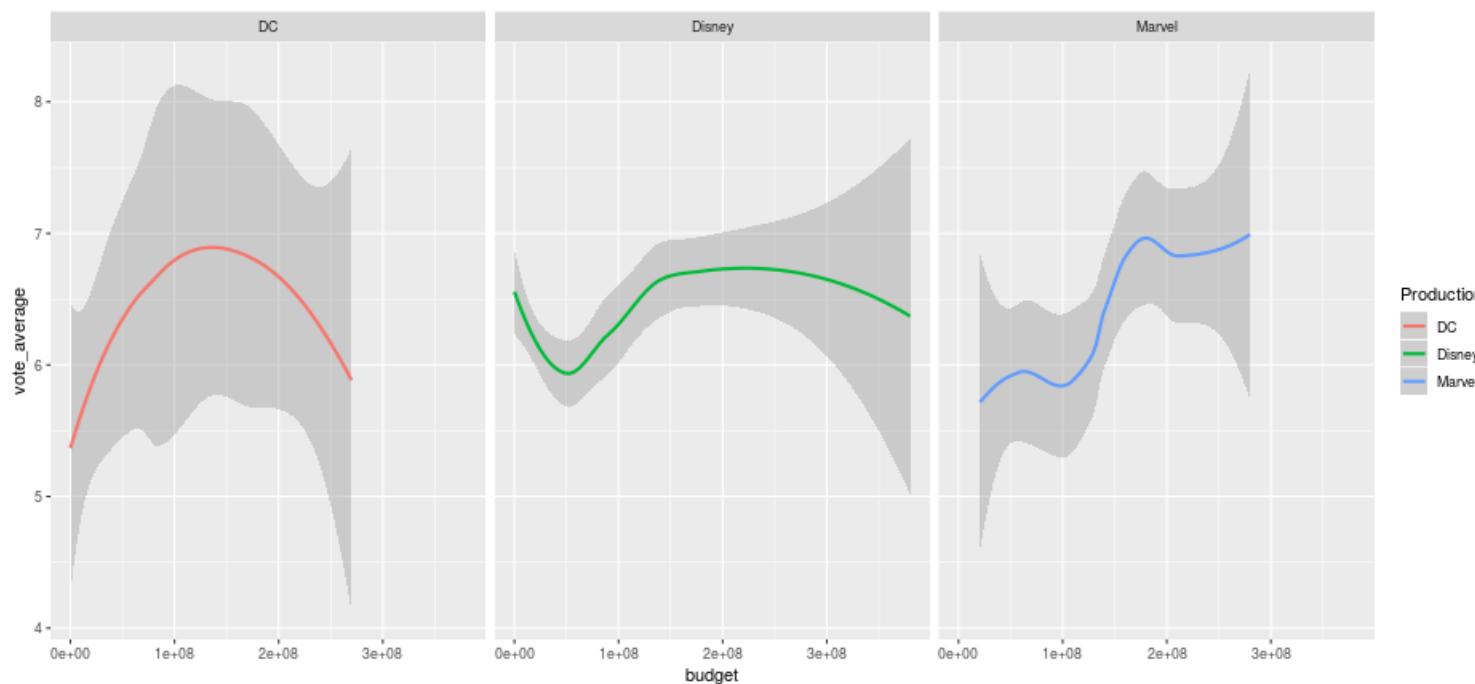
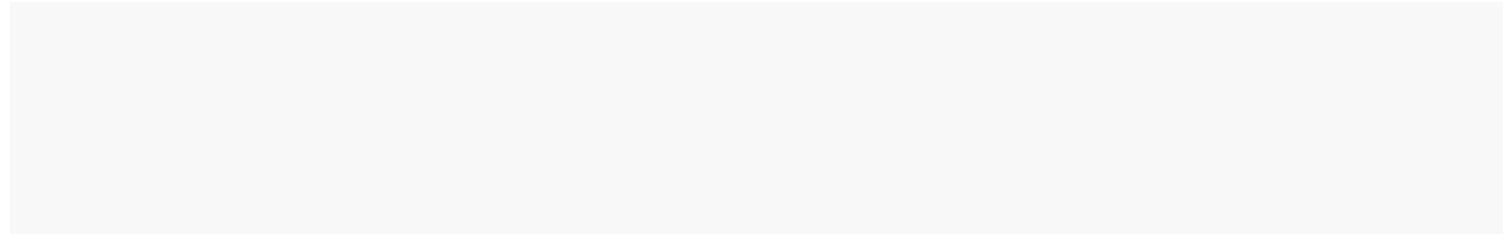
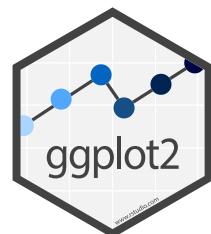
Keep Disney, Marvel and DC movies

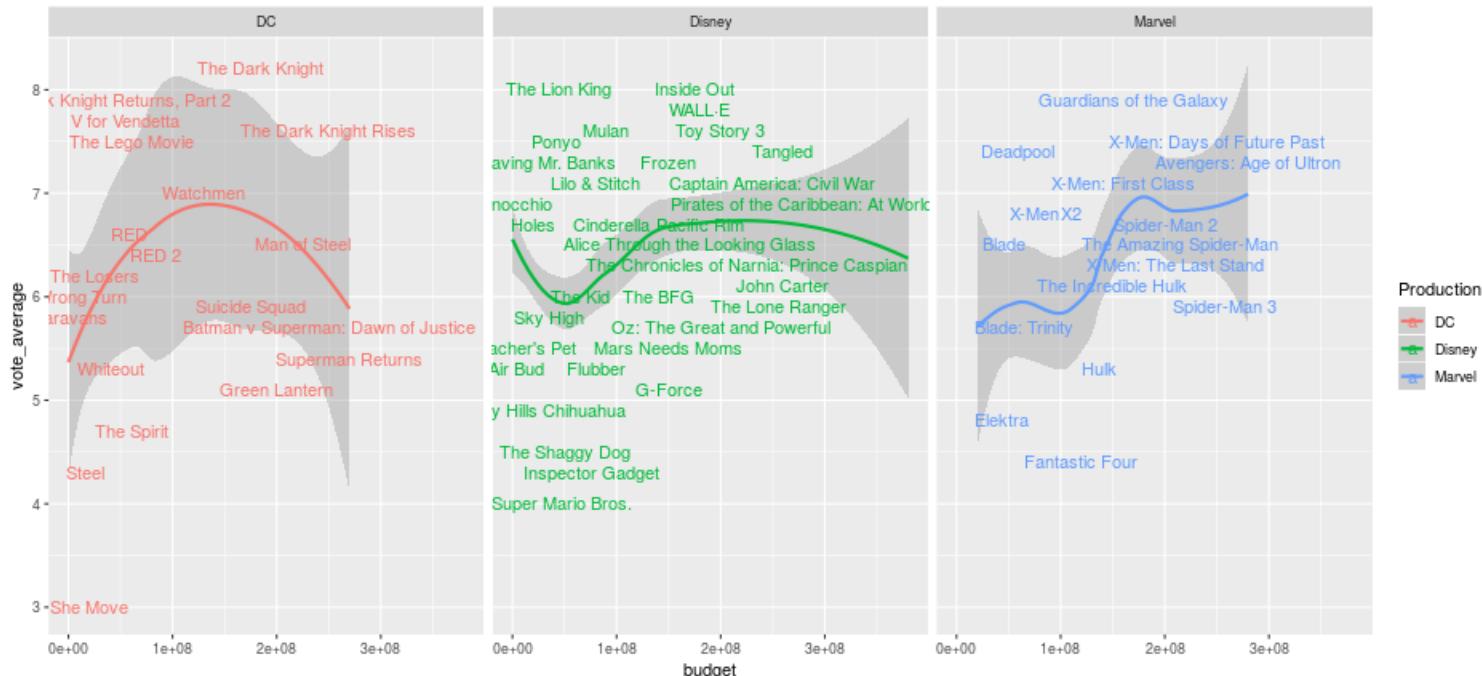
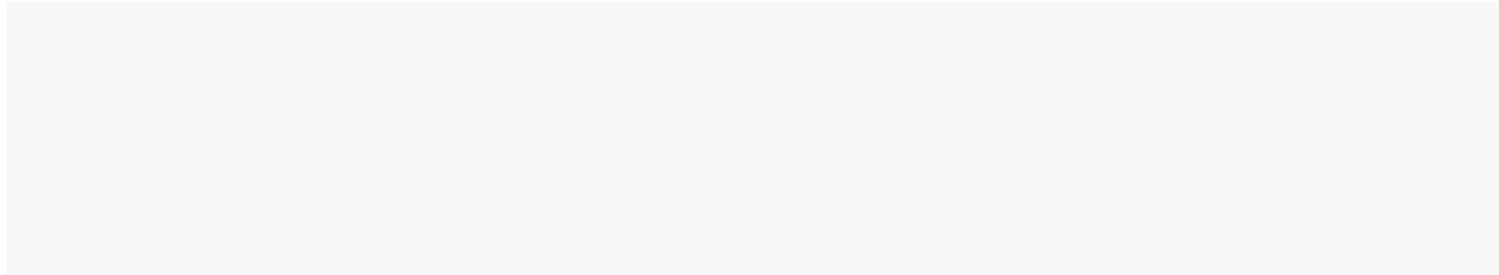
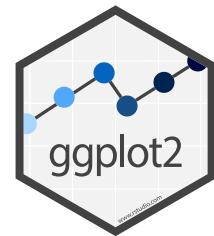




Keep Disney, Marvel and DC movies







- Cheat sheets



Thanks Dude

Slides created via the R package **xaringan** with the **R-Ladies theme**