

+ MI  ...

 $\{...\}$ 

+   ...

```
{html_document}
<html xmlns:og="http://ogp.me/ns#" xmlns:fb="http://www.facebook.com/2008/fbml">
[1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-8 ...
[2] <body id="styleguide-v2" class="fixed">\n                <img height="1" widt ...
```

```
# movie title
titles <- imdb %>%
  html_nodes("h3.lister-item-header") %>%
  html_text2()
```

```
titles[1:10]
```

'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. The Dark Knight (2008)' · '4. Schindler's List (1993)' · '5. The Lord of the Rings: The Return of the King (2003)' · '6. The Godfather Part II (1974)' · '7. 12 Angry Men (1957)' · '8. Pulp Fiction (1994)' · '9. The Lord of the Rings: The Fellowship of the Ring (2001)' · '10. Inception (2010)'

```
# rating
ratings <- imdb %>%
  html_nodes("div.ratings-imdb-rating") %>%
  html_text2() %>%
  as.numeric()
```

```
# number of votes
num_votes <- imdb %>%
  html_nodes("p.sort-num_votes-visible") %>%
  html_text2()
```

```
# build a dataset
df <- data.frame(
  title = titles,
  rating = ratings,
  num_vote = num_votes
)
```

```
head(df)
```

	title	rating	num_vote
	<chr>	<dbl>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,702,956   Gross: \$28.34M   Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,876,645   Gross: \$134.97M   Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,676,781   Gross: \$534.86M   Top 250: #3
4	4. Schindler's List (1993)	9.0	Votes: 1,366,252   Gross: \$96.90M   Top 250: #6
5	5. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,861,122   Gross: \$377.85M   Top 250: #7
6	6. The Godfather Part II (1974)	9.0	Votes: 1,281,988   Gross: \$57.30M   Top 250: #4