

```
library(tidyverse)
```

```
library(rvest)
```

```
url <- "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

```
print(url)
```

```
[1] "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

```
imdb <- read_html(url)
```

```
imdb
```

```
{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" width="1" alt="Facebook logo" data-bbox="125 645 145 655" />
```

```
titles <- imdb %>%
  html_nodes("h3.list-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 <- imdb %>%
  html_nodes("div.ratings-imdb-rating") %>%
  html_text2()
```

rating

'9.3' · '9.2' · '9.0' · '9.0' · '9.0' · '9.0' · '9.0' · '9.0' · '8.9' · '8.8' · '8.8' · '8.8' · '8.8' · '8.8' · '8.8' · '8.7' · '8.7' · '8.7' ·  
 '8.7' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.6' · '8.5' · '8.5' · '8.5' ·  
 '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5' · '8.5'

```
vote <- imdb %>%
  html_nodes("p.sort-num_votes-visible") %>%
  html_text2()
```

vote

'Votes: 2,707,491 | Gross: \$28.34M | Top 250: #1' · 'Votes: 1,880,247 | Gross: \$134.97M | Top 250: #2' ·  
 'Votes: 2,680,785 | Gross: \$534.86M | Top 250: #3' · 'Votes: 1,368,180 | Gross: \$96.90M | Top 250: #6' ·  
 'Votes: 1,863,692 | Gross: \$377.85M | Top 250: #7' · 'Votes: 1,283,834 | Gross: \$57.30M | Top 250: #4' ·  
 'Votes: 799,821 | Gross: \$4.36M | Top 250: #5' · 'Votes: 2,078,622 | Gross: \$107.93M | Top 250: #8' ·  
 'Votes: 1,893,174 | Gross: \$315.54M | Top 250: #9' · 'Votes: 2,378,789 | Gross: \$292.58M | Top 250: #14' ·  
 'Votes: 2,151,780 | Gross: \$37.03M | Top 250: #12' · 'Votes: 2,103,636 | Gross: \$330.25M | Top 250: #11' ·  
 'Votes: 1,682,822 | Gross: \$342.55M | Top 250: #13' · 'Votes: 768,640 | Gross: \$6.10M | Top 250: #10' ·  
 'Votes: 1,174,350 | Gross: \$46.84M | Top 250: #17' · 'Votes: 1,931,800 | Gross: \$171.48M | Top 250: #16' ·  
 'Votes: 1,016,362 | Gross: \$112.00M | Top 250: #18' · 'Votes: 1,304,362 | Gross: \$290.48M | Top 250: #15' ·  
 'Votes: 1,865,964 | Gross: \$188.02M | Top 250: #25' · 'Votes: 1,447,535 | Gross: \$130.74M | Top 250: #22' ·  
 'Votes: 1,671,792 | Gross: \$100.13M | Top 250: #19' · 'Votes: 1,316,145 | Gross: \$136.80M | Top 250: #27' ·  
 'Votes: 1,405,142 | Gross: \$216.54M | Top 250: #23' · 'Votes: 1,376,753 | Gross: \$322.74M | Top 250: #28' ·  
 'Votes: 1,109,823 | Gross: \$204.84M | Top 250: #29' · 'Votes: 774,815 | Gross: \$10.06M | Top 250: #31' ·  
 'Votes: 763,108 | Gross: \$7.56M | Top 250: #24' · 'Votes: 702,696 | Gross: \$57.60M | Top 250: #26' ·  
 'Votes: 467,290 | Top 250: #21' · 'Votes: 349,163 | Gross: \$0.27M | Top 250: #20' ·  
 'Votes: 59,195 | Top 250: #44' · 'Votes: 881,646 | Gross: \$13.09M | Top 250: #42' ·  
 'Votes: 826,225 | Gross: \$53.37M | Top 250: #34' · 'Votes: 1,219,749 | Gross: \$210.61M | Top 250: #30' ·  
 'Votes: 1,515,559 | Gross: \$187.71M | Top 250: #37' · 'Votes: 1,338,305 | Gross: \$132.38M | Top 250: #39' ·  
 'Votes: 1,346,914 | Gross: \$53.09M | Top 250: #41' · 'Votes: 674,580 | Gross: \$83.47M | Top 250: #53' ·  
 'Votes: 1,173,805 | Gross: \$19.50M | Top 250: #35' · 'Votes: 892,000 | Gross: \$78.90M | Top 250: #51' ·  
 'Votes: 1,094,128 | Gross: \$23.34M | Top 250: #40' · 'Votes: 1,070,060 | Gross: \$422.78M | Top 250: #36' ·  
 'Votes: 1,131,955 | Gross: \$6.72M | Top 250: #38' · 'Votes: 842,593 | Gross: \$32.57M | Top 250: #32' ·  
 'Votes: 869,162 | Gross: \$13.18M | Top 250: #46' · 'Votes: 333,471 | Gross: \$5.32M | Top 250: #48' ·  
 'Votes: 577,115 | Gross: \$1.02M | Top 250: #43' · 'Votes: 678,736 | Gross: \$32.00M | Top 250: #33' ·  
 'Votes: 282,014 | Top 250: #45' · 'Votes: 264,631 | Gross: \$11.99M | Top 250: #50'

```
#build a data set
df <- data.frame(
  title = titles,
  rating = rating,
  vote = vote
)
head(df)
```

A data.frame: 6 x 3

	title	rating	vote
	<chr>	<chr>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,707,491   Gross: \$28.34M   Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,880,247   Gross: \$134.97M   Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,680,785   Gross: \$534.86M   Top 250: #3
4	4. Schindler's List (1993)	9.0	Votes: 1,368,180   Gross: \$96.90M   Top 250: #6
5	5. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,863,692   Gross: \$377.85M   Top 250: #7
6	6. The Godfather Part II (1974)	9.0	Votes: 1,283,834   Gross: \$57.30M   Top 250: #4