

Mini Project -my anime list web scraping

```
library(tidyverse)
library(rvest)# scrap data
library(dplyr)
library(readr)
```

```
url = "https://myanimelist.net/topanime.php"
```

```
#read html
Top_Anime <- read_html(url)
```

```
#title
title <- Top_Anime %>%
  html_nodes("h3.hoverinfo_trigger") %>%
  html_text2()
```

```
title[1:10]
```

'Fullmetal Alchemist: Brotherhood' · 'Bleach: Sennen Kessen-hen' · 'Kaguya-sama wa Kokurasetai: Ultra Romantic' · 'Steins;Gate' · 'Gintama' · 'Shingeki no Kyojin Season 3 Part 2' · 'Gintama\' · 'Gintama: The Final' · 'Gintama\': Enchousen' · 'Hunter x Hunter (2011)'

```
#rating
rating <- Top_Anime %>%
```

```
html_nodes("div.js-top-ranking-score-col") %>%  
html_text2()
```

rating

'9.12' · '9.10' · '9.09' · '9.08' · '9.07' · '9.06' · '9.05' · '9.05' · '9.04' · '9.04' · '9.03' · '9.02' · '8.99' · '8.95' · '8.95' · '8.94' ·
'8.94' · '8.92' · '8.92' · '8.91' · '8.89' · '8.88' · '8.86' · '8.86' · '8.85' · '8.83' · '8.82' · '8.82' · '8.81' · '8.81' · '8.80' · '8.79' ·
'8.79' · '8.78' · '8.77' · '8.77' · '8.77' · '8.75' · '8.75' · '8.74' · '8.74' · '8.73' · '8.73' · '8.72' · '8.72' · '8.71' · '8.71' · '8.71' ·
'8.71' · '8.70'

```
#info  
info <- Top_Anime %>%  
  html_nodes("div.information") %>%  
  html_text2()
```

info

'TV (64 eps)\nApr 2009 - Jul 2010\n3,021,625 members' · 'TV (? eps)\nOct 2022 -\n283,906 members' ·
 'TV (13 eps)\nApr 2022 - Jun 2022\n690,177 members' · 'TV (24 eps)\nApr 2011 - Sep 2011\n2,326,522 members' ·
 'TV (51 eps)\nApr 2015 - Mar 2016\n563,920 members' · 'TV (10 eps)\nApr 2019 - Jul 2019\n1,958,286 members' ·
 'TV (51 eps)\nApr 2011 - Mar 2012\n497,023 members' · 'Movie (1 eps)\nJan 2021 - Jan 2021\n116,914 members' ·
 'TV (13 eps)\nOct 2012 - Mar 2013\n294,996 members' · 'TV (148 eps)\nOct 2011 - Sep 2014\n2,504,049 members' ·
 'OVA (110 eps)\nJan 1988 - Mar 1997\n291,941 members' · 'TV (13 eps)\nApr 2021 - Jun 2021\n385,655 members' ·
 'TV (12 eps)\nJan 2017 - Mar 2017\n282,034 members' · 'TV (22 eps)\nOct 2017 - Mar 2018\n366,402 members' ·
 'Movie (1 eps)\nSep 2016 - Sep 2016\n2,066,845 members' ·
 'TV (24 eps)\nOct 2008 - Mar 2009\n1,111,233 members' ·
 'TV (201 eps)\nApr 2006 - Mar 2010\n976,211 members' · 'Movie (1 eps)\nJul 2013 - Jul 2013\n219,058 members' ·
 'Movie (1 eps)\nSep 2020 - Sep 2020\n488,248 members' ·
 'TV (25 eps)\nApr 2008 - Sep 2008\n1,628,142 members' · 'TV (7 eps)\nAug 2017 - Aug 2017\n363,347 members' ·
 'TV (14 eps)\nJul 2018 - Oct 2018\n174,001 members' · 'TV (12 eps)\nOct 2022 -\n816,577 members' ·
 'Movie (1 eps)\nAug 2016 - Aug 2016\n2,461,751 members' ·
 'TV (74 eps)\nApr 2004 - Sep 2005\n915,058 members' · 'TV (11 eps)\nDec 2021 - Feb 2022\n1,038,533 members' ·
 'TV (26 eps)\nApr 2020 - Oct 2021\n80,793 members' · 'TV (16 eps)\nDec 2020 - Mar 2021\n1,704,063 members' ·
 'TV (13 eps)\nJan 2019 - Apr 2019\n1,326,168 members' · 'TV (12 eps)\nJan 2018 - Mar 2018\n188,101 members' ·
 'Movie (1 eps)\nJan 2017 - Jan 2017\n420,536 members' ·
 'TV (10 eps)\nOct 2016 - Dec 2016\n1,083,676 members' · 'TV (12 eps)\nJan 2022 - Apr 2022\n979,869 members' ·
 'Movie (1 eps)\nJul 2001 - Jul 2001\n1,677,657 members' · 'TV (26 eps)\nJul 2013 - Dec 2013\n627,620 members' ·
 'ONA (11 eps)\nApr 2021 - Jul 2021\n248,449 members' · 'TV (26 eps)\nApr 2022 - Oct 2022\n43,740 members' ·
 'TV (26 eps)\nApr 1998 - Apr 1999\n1,685,678 members' · 'TV (75 eps)\nOct 2000 - Mar 2002\n506,741 members' ·
 'TV (10 eps)\nOct 2014 - Dec 2014\n244,990 members' · 'TV (12 eps)\nJan 2017 - Mar 2017\n174,777 members' ·
 'TV (12 eps)\nApr 2022 - Jun 2022\n1,103,693 members' · 'TV (24 eps)\nJul 2019 - Dec 2019\n1,200,648 members' ·
 'TV (12 eps)\nJul 2022 - Sep 2022\n308,586 members' · 'TV (12 eps)\nOct 2021 - Dec 2021\n675,124 members' ·
 'TV (13 eps)\nApr 2021 - Jun 2021\n347,259 members' · 'OVA (4 eps)\nFeb 1999 - Sep 1999\n262,606 members' ·
 'TV (12 eps)\nOct 2021 - Mar 2022\n364,950 members' · 'TV (47 eps)\nOct 1980 - Aug 1981\n44,524 members' ·
 'TV (25 eps)\nOct 2006 - Jul 2007\n2,063,204 members'

```
#build a dataset
df <- data.frame(
  titles = title,
  ratings = rating,
  infos = info
```

```
)  
head(df)
```

A data.frame: 6 × 3

	titles	ratings	infos
	<chr>	<chr>	<chr>
1	Fullmetal Alchemist: Brotherhood	9.12	TV (64 eps) Apr 2009 - Jul 2010 3,021,625 members
2	Bleach: Sennen Kessen-hen	9.10	TV (? eps) Oct 2022 - 283,906 members
3	Kaguya-sama wa Kokurasetai: Ultra Romantic	9.09	TV (13 eps) Apr 2022 - Jun 2022 690,177 members
4	Steins;Gate	9.08	TV (24 eps) Apr 2011 - Sep 2011 2,326,522 members
5	Gintama°	9.07	TV (51 eps) Apr 2015 - Mar 2016 563,920 members
6	Shingeki no Kyojin Season 3 Part 2	9.06	TV (10 eps) Apr 2019 - Jul 2019 1,958,286 members