

# Wiki Table Resubmit

*Paine (HE Peiwei, 54471952)*

*2017.2.10*

## Research Topic: Wiki Table Scraping

First, I scrap the wiki page of The Fast and the Furious, and get a table of the detailed info of the serie films.

```
library(rvest)
```

```
## Loading required package: xml2
```

```
url = "https://en.wikipedia.org/wiki/The_Fast_and_the_Furious"
```

```
film = read_html(url)
```

```
table=film%>%html_node("table.wikitable")%>%html_table(header = NA, trim = TRUE, fill=TRUE)
```

Then, I extra links of all the films.

```
links = film%>%html_node("table.wikitable")%>% html_nodes("a") %>% html_attr("href")
```

```
links = paste("https://en.wikipedia.org", links, sep = "")
```

```
table$link = links
```

Then, I create 2 functions to get the Starring and Director of each film, according to the details in the infomation boxes.

```
get_Starring = function(page) {  
  details = page %>% html_node("table.infobox") %>% html_table  
  colnames(details) <- c("name", "value")  
  details$value[details$name == "Starring"]  
}
```

```
get_Director = function(page) {  
  details = page %>% html_node("table.infobox") %>% html_table  
  colnames(details) <- c("name", "value")  
  details$value[details$name == "Directed by"]  
}
```

Finally, I apply the funtions and add 2 new columns to the table.

```
table$Starring = ""  
table$Director = ""  
for (i in seq_along(links)) {  
  message(i)  
  page = read_html(links[i])  
  table$Starring[i] = get_Starring(page)  
  table$Director[i] = get_Director(page)  
}
```

```
## 1
```

```
## 2
```

```
## 3
```

```
## 4
```

```
## 5
## 6
## 7
## 8
## 9
## 10
table
```

##	Timeline order	Title
## 1	1	The Fast and the Furious
## 2	2	The Turbo Charged Prelude for 2 Fast 2 Furious
## 3	3	2 Fast 2 Furious
## 4	4	Los Bandoleros
## 5	5	Fast & Furious
## 6	6	Fast Five
## 7	7	Fast & Furious 6
## 8	8	The Fast and the Furious: Tokyo Drift
## 9	9	Furious 7
## 10	10	The Fate of the Furious

##	Release date
## 1	2001-06-22 !June 22, 2001
## 2	2003-06-03 !June 3, 2003
## 3	2003-06-06 !June 6, 2003
## 4	2009-07-28 !July 28, 2009
## 5	2009-04-03 !April 3, 2009
## 6	2011-04-29 !April 29, 2011
## 7	2013-05-24 !May 24, 2013
## 8	2006-06-16 !June 16, 2006
## 9	2015-04-03 !April 3, 2015
## 10	2017-04-14 !April 14, 2017

##	link
## 1	<a href="https://en.wikipedia.org/wiki/The_Fast_and_the_Furious_(2001_film)">https://en.wikipedia.org/wiki/The_Fast_and_the_Furious_(2001_film)</a>
## 2	<a href="https://en.wikipedia.org/wiki/The_Fast_and_the_Furious#Short_films">https://en.wikipedia.org/wiki/The_Fast_and_the_Furious#Short_films</a>
## 3	<a href="https://en.wikipedia.org/wiki/2_Fast_2_Furious">https://en.wikipedia.org/wiki/2_Fast_2_Furious</a>
## 4	<a href="https://en.wikipedia.org/wiki/Los_Bandoleros_(film)">https://en.wikipedia.org/wiki/Los_Bandoleros_(film)</a>
## 5	<a href="https://en.wikipedia.org/wiki/Fast_%26_Furious_(2009_film)">https://en.wikipedia.org/wiki/Fast_%26_Furious_(2009_film)</a>
## 6	<a href="https://en.wikipedia.org/wiki/Fast_Five">https://en.wikipedia.org/wiki/Fast_Five</a>
## 7	<a href="https://en.wikipedia.org/wiki/Fast_%26_Furious_6">https://en.wikipedia.org/wiki/Fast_%26_Furious_6</a>
## 8	<a href="https://en.wikipedia.org/wiki/The_Fast_and_the_Furious:_Tokyo_Drift">https://en.wikipedia.org/wiki/The_Fast_and_the_Furious:_Tokyo_Drift</a>
## 9	<a href="https://en.wikipedia.org/wiki/Furious_7">https://en.wikipedia.org/wiki/Furious_7</a>
## 10	<a href="https://en.wikipedia.org/wiki/The_Fate_of_the_Furious">https://en.wikipedia.org/wiki/The_Fate_of_the_Furious</a>

```
##
## 1
## 2 Vin Diesel (1, 3-8)\nPaul Walker (1-2, 4-7)\nDwayne Johnson (5-8)\nMichelle Rodriguez (1, 4, 6-8)
## 3
## 4
## 5
## 6
## 7
## 8
## 9
```

## 10	
##	Director
## 1	Rob Cohen
## 2	Rob Cohen (1)\nJohn Singleton (2)\nJustin Lin (3-6)\nJames Wan (7)\nF. Gary Gray (8)
## 3	John Singleton
## 4	Vin Diesel
## 5	Justin Lin
## 6	Justin Lin
## 7	Justin Lin
## 8	Justin Lin
## 9	James Wan
## 10	F. Gary Gray