

Lab 2: Screen-Scraping

Srishti Saha (ss1078)

04 February, 2020

Question 1: Navigate to the wikipedia page on highest grossing films. One there, utilize the inspect function of your web browser in conjunction with Rvest to copy the table of highest grossing films. Print out the head() of this table

```
films_wiki<-read_html("https://en.wikipedia.org/wiki/List_of_highest-grossing_films")
films_table_section<-html_node(films_wiki, xpath='//*[@id="mw-content-text"]/div/table[1]')
#head(films_table_section)
films_table<-html_table(films_table_section)

knitr::kable(
  head(films_table),
  format = 'latex',
  caption="Top 5 highest grossing films",
  booktabs = T
)%>%kable_styling(c("striped","condensed"),full_width = F, latex_options = "hold_position",font_size = 7)
```

Table 1: Top 5 highest grossing films

Rank	Peak	Title	Worldwide gross	Year	Reference(s)
1	1	Avengers: Endgame	\$2,797,800,564	2019	[# 1][# 2]
2	1	Avatar	\$2,789,679,794	2009	[# 3][# 4]
3	1	Titanic	\$2,187,463,944	1997	[# 5][# 6]
4	3	Star Wars: The Force Awakens	\$2,068,223,624	2015	[# 7][# 8]
5	4	Avengers: Infinity War	\$2,048,359,754	2018	[# 9][# 10]
6	3	Jurassic World	\$1,671,713,208	2015	[# 11][# 12]

Question 3: Unfortunately data we want isn't always as nicely structured as a wikipedia table. Navigate to a this blog post about famous movie quotes. Using the Selector Gadget for you browser, pull the first 10 quotes from the list, print the results.

```
quotes_page<-read_html("https://blog.hubspot.com/sales/famous-movie-quotes")
quotes<-html_nodes(quotes_page, css="p:nth-child(13) , p:nth-child(12) ,
  p:nth-child(11) , p:nth-child(10) , p:nth-child(9) , p:nth-child(8) , p:nth-child(7) ,
  p:nth-child(6) , p:nth-child(5) , p:nth-child(4) , p:nth-child(3) .epsilon")
```

The above quite will store the quotes in an object.

```
## {xml_nodeset (10)}
## [1] <p>1. "May the Force be with you." -<em>Star Wars, 1977</em></p>
## [2] <p>2. "There's no place like home." -<em>The Wizard of Oz, 1939</em> ...
## [3] <p>3. "I'm the king of the world!" -<em>Titanic, 1997</em></p>
## [4] <p>4. "Carpe diem. Seize the day, boys. Make your lives extraordina ...
## [5] <p>5. "Elementary, my dear Watson." -<em>The Adventures of Sherlock ...
## [6] <p>6. "It's alive! It's alive!" -<em>Frankenstein, 1931</em></p>
## [7] <p>7. "My mama always said life was like a box of chocolates. You n ...
## [8] <p>8. "I'll be back." -<em>The Terminator, 1984</em></p>
## [9] <p>9. "You're gonna need a bigger boat." -<em>Jaws, 1975</em></p>
## [10] <p>10. "Here's looking at you, kid." -<em>Casablanca,1942</em></p>
```