STAT 613 - Homework 7

Jacob Henkels

3/21/2021

## Warning: package 'rvest' was built under R version 4.0.4

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## -- Attaching packages --------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v stringr 1.4.0  
## v tidyr 1.1.2 v forcats 0.5.0  
## v readr 1.3.1

## Warning: package 'forcats' was built under R version 4.0.3

## -- Conflicts ------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x readr::guess\_encoding() masks rvest::guess\_encoding()  
## x dplyr::lag() masks stats::lag()

## Warning: package 'repurrrsive' was built under R version 4.0.4

# 1.

Link <- "https://www.imdb.com/list/ls091520106/"  
page = read\_html(Link)  
Title = page%>% html\_nodes(".lister-item-header a")%>%  
html\_text()  
Title

## [1] "The Shawshank Redemption"   
## [2] "The Godfather"   
## [3] "The Godfather: Part II"   
## [4] "The Dark Knight"   
## [5] "12 Angry Men"   
## [6] "Schindler's List"   
## [7] "The Lord of the Rings: The Return of the King"   
## [8] "Pulp Fiction"   
## [9] "The Good, the Bad and the Ugly"   
## [10] "Fight Club"   
## [11] "Joker"   
## [12] "The Lord of the Rings: The Fellowship of the Ring"   
## [13] "Forrest Gump"   
## [14] "Inception"   
## [15] "Star Wars: Episode V - The Empire Strikes Back"   
## [16] "The Lord of the Rings: The Two Towers"   
## [17] "The Matrix"   
## [18] "One Flew Over the Cuckoo's Nest"   
## [19] "Goodfellas"   
## [20] "Seven Samurai"   
## [21] "Se7en"   
## [22] "City of God"   
## [23] "Life Is Beautiful"   
## [24] "The Silence of the Lambs"   
## [25] "Star Wars: Episode IV - A New Hope"   
## [26] "It's a Wonderful Life"   
## [27] "Saving Private Ryan"   
## [28] "Spirited Away"   
## [29] "The Green Mile"   
## [30] "Léon: The Professional"   
## [31] "Hara-Kiri"   
## [32] "Interstellar"   
## [33] "The Usual Suspects"   
## [34] "The Lion King"   
## [35] "American History X"   
## [36] "Back to the Future"   
## [37] "The Pianist"   
## [38] "Modern Times"   
## [39] "Terminator 2: Judgment Day"   
## [40] "The Intouchables"   
## [41] "Psycho"   
## [42] "Gladiator"   
## [43] "City Lights"   
## [44] "The Departed"   
## [45] "Whiplash"   
## [46] "Once Upon a Time in the West"   
## [47] "The Prestige"   
## [48] "Avengers: Endgame"   
## [49] "Casablanca"   
## [50] "Grave of the Fireflies"   
## [51] "Rear Window"   
## [52] "Cinema Paradiso"   
## [53] "Alien"   
## [54] "Raiders of the Lost Ark"   
## [55] "Memento"   
## [56] "Apocalypse Now"   
## [57] "The Great Dictator"   
## [58] "The Lives of Others"   
## [59] "Avengers: Infinity War"   
## [60] "Django Unchained"   
## [61] "Spider-Man: Into the Spider-Verse"   
## [62] "The Shining"   
## [63] "Paths of Glory"   
## [64] "WALL·E"   
## [65] "Sunset Blvd."   
## [66] "Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb"  
## [67] "Princess Mononoke"   
## [68] "Oldboy"   
## [69] "Witness for the Prosecution"   
## [70] "The Dark Knight Rises"   
## [71] "Once Upon a Time in America"   
## [72] "Parasite"   
## [73] "Aliens"   
## [74] "American Beauty"   
## [75] "Coco"   
## [76] "Your Name."   
## [77] "Braveheart"   
## [78] "Das Boot"   
## [79] "3 Idiots"   
## [80] "Taare Zameen Par"   
## [81] "Star Wars: Episode VI - Return of the Jedi"   
## [82] "Toy Story"   
## [83] "Reservoir Dogs"   
## [84] "Amadeus"   
## [85] "Dangal"   
## [86] "Good Will Hunting"   
## [87] "Inglourious Basterds"   
## [88] "M"   
## [89] "Requiem for a Dream"   
## [90] "2001: A Space Odyssey"   
## [91] "Vertigo"   
## [92] "Eternal Sunshine of the Spotless Mind"   
## [93] "Citizen Kane"   
## [94] "Full Metal Jacket"   
## [95] "The Hunt"   
## [96] "North by Northwest"   
## [97] "A Clockwork Orange"   
## [98] "Snatch"   
## [99] "Amélie"   
## [100] "The Kid"

Ratings = page%>% html\_nodes(".ipl-rating-star.small .ipl-rating-star\_\_rating")%>%  
html\_text()  
Ratings

## [1] "9.3" "9.2" "9" "9" "9" "8.9" "8.9" "8.9" "8.8" "8.8" "8.4" "8.8"  
## [13] "8.8" "8.8" "8.7" "8.7" "8.7" "8.7" "8.7" "8.6" "8.6" "8.6" "8.6" "8.6"  
## [25] "8.6" "8.6" "8.6" "8.6" "8.6" "8.5" "8.6" "8.6" "8.5" "8.5" "8.5" "8.5"  
## [37] "8.5" "8.5" "8.5" "8.5" "8.5" "8.5" "8.5" "8.5" "8.5" "8.5" "8.5" "8.4"  
## [49] "8.5" "8.5" "8.4" "8.5" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4"  
## [61] "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.4" "8.6"  
## [73] "8.3" "8.3" "8.4" "8.4" "8.3" "8.3" "8.4" "8.4" "8.3" "8.3" "8.3" "8.3"  
## [85] "8.4" "8.3" "8.3" "8.3" "8.3" "8.3" "8.3" "8.3" "8.3" "8.3" "8.3" "8.3"  
## [97] "8.3" "8.3" "8.3" "8.3"

Runtime = page%>% html\_nodes(".runtime")%>%  
html\_text()  
Runtime

## [1] "142 min" "175 min" "202 min" "152 min" "96 min" "195 min" "201 min"  
## [8] "154 min" "178 min" "139 min" "122 min" "178 min" "142 min" "148 min"  
## [15] "124 min" "179 min" "136 min" "133 min" "146 min" "207 min" "127 min"  
## [22] "130 min" "116 min" "118 min" "121 min" "130 min" "169 min" "125 min"  
## [29] "189 min" "110 min" "133 min" "169 min" "106 min" "88 min" "119 min"  
## [36] "116 min" "150 min" "87 min" "137 min" "112 min" "109 min" "155 min"  
## [43] "87 min" "151 min" "106 min" "165 min" "130 min" "181 min" "102 min"  
## [50] "89 min" "112 min" "155 min" "117 min" "115 min" "113 min" "147 min"  
## [57] "125 min" "137 min" "149 min" "165 min" "117 min" "146 min" "88 min"   
## [64] "98 min" "110 min" "95 min" "134 min" "120 min" "116 min" "164 min"  
## [71] "229 min" "132 min" "137 min" "122 min" "105 min" "106 min" "178 min"  
## [78] "149 min" "170 min" "165 min" "131 min" "81 min" "99 min" "160 min"  
## [85] "161 min" "126 min" "153 min" "99 min" "102 min" "149 min" "128 min"  
## [92] "108 min" "119 min" "116 min" "115 min" "136 min" "136 min" "102 min"  
## [99] "122 min" "68 min"

movieDF = data.frame(Title, Ratings, Runtime)  
movieDF

## Title  
## 1 The Shawshank Redemption  
## 2 The Godfather  
## 3 The Godfather: Part II  
## 4 The Dark Knight  
## 5 12 Angry Men  
## 6 Schindler's List  
## 7 The Lord of the Rings: The Return of the King  
## 8 Pulp Fiction  
## 9 The Good, the Bad and the Ugly  
## 10 Fight Club  
## 11 Joker  
## 12 The Lord of the Rings: The Fellowship of the Ring  
## 13 Forrest Gump  
## 14 Inception  
## 15 Star Wars: Episode V - The Empire Strikes Back  
## 16 The Lord of the Rings: The Two Towers  
## 17 The Matrix  
## 18 One Flew Over the Cuckoo's Nest  
## 19 Goodfellas  
## 20 Seven Samurai  
## 21 Se7en  
## 22 City of God  
## 23 Life Is Beautiful  
## 24 The Silence of the Lambs  
## 25 Star Wars: Episode IV - A New Hope  
## 26 It's a Wonderful Life  
## 27 Saving Private Ryan  
## 28 Spirited Away  
## 29 The Green Mile  
## 30 Léon: The Professional  
## 31 Hara-Kiri  
## 32 Interstellar  
## 33 The Usual Suspects  
## 34 The Lion King  
## 35 American History X  
## 36 Back to the Future  
## 37 The Pianist  
## 38 Modern Times  
## 39 Terminator 2: Judgment Day  
## 40 The Intouchables  
## 41 Psycho  
## 42 Gladiator  
## 43 City Lights  
## 44 The Departed  
## 45 Whiplash  
## 46 Once Upon a Time in the West  
## 47 The Prestige  
## 48 Avengers: Endgame  
## 49 Casablanca  
## 50 Grave of the Fireflies  
## 51 Rear Window  
## 52 Cinema Paradiso  
## 53 Alien  
## 54 Raiders of the Lost Ark  
## 55 Memento  
## 56 Apocalypse Now  
## 57 The Great Dictator  
## 58 The Lives of Others  
## 59 Avengers: Infinity War  
## 60 Django Unchained  
## 61 Spider-Man: Into the Spider-Verse  
## 62 The Shining  
## 63 Paths of Glory  
## 64 WALL·E  
## 65 Sunset Blvd.  
## 66 Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb  
## 67 Princess Mononoke  
## 68 Oldboy  
## 69 Witness for the Prosecution  
## 70 The Dark Knight Rises  
## 71 Once Upon a Time in America  
## 72 Parasite  
## 73 Aliens  
## 74 American Beauty  
## 75 Coco  
## 76 Your Name.  
## 77 Braveheart  
## 78 Das Boot  
## 79 3 Idiots  
## 80 Taare Zameen Par  
## 81 Star Wars: Episode VI - Return of the Jedi  
## 82 Toy Story  
## 83 Reservoir Dogs  
## 84 Amadeus  
## 85 Dangal  
## 86 Good Will Hunting  
## 87 Inglourious Basterds  
## 88 M  
## 89 Requiem for a Dream  
## 90 2001: A Space Odyssey  
## 91 Vertigo  
## 92 Eternal Sunshine of the Spotless Mind  
## 93 Citizen Kane  
## 94 Full Metal Jacket  
## 95 The Hunt  
## 96 North by Northwest  
## 97 A Clockwork Orange  
## 98 Snatch  
## 99 Amélie  
## 100 The Kid  
## Ratings Runtime  
## 1 9.3 142 min  
## 2 9.2 175 min  
## 3 9 202 min  
## 4 9 152 min  
## 5 9 96 min  
## 6 8.9 195 min  
## 7 8.9 201 min  
## 8 8.9 154 min  
## 9 8.8 178 min  
## 10 8.8 139 min  
## 11 8.4 122 min  
## 12 8.8 178 min  
## 13 8.8 142 min  
## 14 8.8 148 min  
## 15 8.7 124 min  
## 16 8.7 179 min  
## 17 8.7 136 min  
## 18 8.7 133 min  
## 19 8.7 146 min  
## 20 8.6 207 min  
## 21 8.6 127 min  
## 22 8.6 130 min  
## 23 8.6 116 min  
## 24 8.6 118 min  
## 25 8.6 121 min  
## 26 8.6 130 min  
## 27 8.6 169 min  
## 28 8.6 125 min  
## 29 8.6 189 min  
## 30 8.5 110 min  
## 31 8.6 133 min  
## 32 8.6 169 min  
## 33 8.5 106 min  
## 34 8.5 88 min  
## 35 8.5 119 min  
## 36 8.5 116 min  
## 37 8.5 150 min  
## 38 8.5 87 min  
## 39 8.5 137 min  
## 40 8.5 112 min  
## 41 8.5 109 min  
## 42 8.5 155 min  
## 43 8.5 87 min  
## 44 8.5 151 min  
## 45 8.5 106 min  
## 46 8.5 165 min  
## 47 8.5 130 min  
## 48 8.4 181 min  
## 49 8.5 102 min  
## 50 8.5 89 min  
## 51 8.4 112 min  
## 52 8.5 155 min  
## 53 8.4 117 min  
## 54 8.4 115 min  
## 55 8.4 113 min  
## 56 8.4 147 min  
## 57 8.4 125 min  
## 58 8.4 137 min  
## 59 8.4 149 min  
## 60 8.4 165 min  
## 61 8.4 117 min  
## 62 8.4 146 min  
## 63 8.4 88 min  
## 64 8.4 98 min  
## 65 8.4 110 min  
## 66 8.4 95 min  
## 67 8.4 134 min  
## 68 8.4 120 min  
## 69 8.4 116 min  
## 70 8.4 164 min  
## 71 8.4 229 min  
## 72 8.6 132 min  
## 73 8.3 137 min  
## 74 8.3 122 min  
## 75 8.4 105 min  
## 76 8.4 106 min  
## 77 8.3 178 min  
## 78 8.3 149 min  
## 79 8.4 170 min  
## 80 8.4 165 min  
## 81 8.3 131 min  
## 82 8.3 81 min  
## 83 8.3 99 min  
## 84 8.3 160 min  
## 85 8.4 161 min  
## 86 8.3 126 min  
## 87 8.3 153 min  
## 88 8.3 99 min  
## 89 8.3 102 min  
## 90 8.3 149 min  
## 91 8.3 128 min  
## 92 8.3 108 min  
## 93 8.3 119 min  
## 94 8.3 116 min  
## 95 8.3 115 min  
## 96 8.3 136 min  
## 97 8.3 136 min  
## 98 8.3 102 min  
## 99 8.3 122 min  
## 100 8.3 68 min

# 2.

### A. Explain or describe nested Lists. (Three or four sentences)

A nested list essentially boils down to a list within a list. For instance, a list may represent a group of people, and includes nested lists with information like their favorite movies. These can occur when the data includes matrices of information.

### B. Open and explore data for Git Hub repos by running the following code gh\_repo. Comment on the structure and arrangement of the data (Two or three sentences)

gh\_repos  
#Hid results since it was very long

By running ‘gh\_repo’, we get a long dataset with multi-level lists (nested). The list contains github users and nested lists for the repositories and info on them.

### C. In two or three sentences, explain what happens to the data when you run the code;

repos <- tibble(repo = gh\_repos)  
repos

## # A tibble: 6 x 1  
## repo   
## <list>   
## 1 <list [30]>  
## 2 <list [30]>  
## 3 <list [30]>  
## 4 <list [26]>  
## 5 <list [30]>  
## 6 <list [30]>

Using the code above produces a tibble of the 6 Github users. We can also see that for each Github user, there are 30 lists nested inside of that list.

### D. Describe the data representation when the listserver tool is applied. And go into detail explaining why we have a nested List configuration. What do the numbers 30 and 68 represent ? (6 or 7 sentences)

listviewer::jsonedit(gh\_repos)  
#Hid results so that I could knit to .docx

By using the listserver tool, we are able to get lists & nested lists. It begins with an array of 6, each representing a Github user. There are 30 nested lists that represent a Github user’s repositories. Within those repositories, there appears to be 68 fields of info on each repository, including things like login data. They also include information like urls, descriptions, and when they were created. Overall, with this code, we are able to dive into gh\_repos and observe the nested lists inside, including 6 users, 30 repositories, and 68 rows of info about each repository.

### E. Use and show R unesting code to produce the tibble below.

tibble(repo = gh\_repos) %>%  
 unnest\_auto(repo) %>%  
 unnest\_auto(repo)

## Using `unnest\_longer(repo)`; no element has names

## Using `unnest\_wider(repo)`; elements have 68 names in common

## # A tibble: 176 x 67  
## id name full\_name owner private html\_url description fork url   
## <int> <chr> <chr> <lis> <lgl> <chr> <chr> <lgl> <chr>  
## 1 6.12e7 after gaborcsa~ <nam~ FALSE https:/~ Run Code i~ FALSE http~  
## 2 4.05e7 argu~ gaborcsa~ <nam~ FALSE https:/~ Declarativ~ FALSE http~  
## 3 3.64e7 ask gaborcsa~ <nam~ FALSE https:/~ Friendly C~ FALSE http~  
## 4 3.49e7 base~ gaborcsa~ <nam~ FALSE https:/~ Do we get ~ FALSE http~  
## 5 6.16e7 cite~ gaborcsa~ <nam~ FALSE https:/~ Test R pac~ TRUE http~  
## 6 3.39e7 clis~ gaborcsa~ <nam~ FALSE https:/~ Unicode sy~ FALSE http~  
## 7 3.72e7 cmak~ gaborcsa~ <nam~ FALSE https:/~ port of cm~ TRUE http~  
## 8 6.80e7 cmark gaborcsa~ <nam~ FALSE https:/~ CommonMark~ TRUE http~  
## 9 6.32e7 cond~ gaborcsa~ <nam~ FALSE https:/~ <NA> TRUE http~  
## 10 2.43e7 cray~ gaborcsa~ <nam~ FALSE https:/~ R package ~ FALSE http~  
## # ... with 166 more rows, and 58 more variables: forks\_url <chr>,  
## # keys\_url <chr>, collaborators\_url <chr>, teams\_url <chr>, hooks\_url <chr>,  
## # issue\_events\_url <chr>, events\_url <chr>, assignees\_url <chr>,  
## # branches\_url <chr>, tags\_url <chr>, blobs\_url <chr>, git\_tags\_url <chr>,  
## # git\_refs\_url <chr>, trees\_url <chr>, statuses\_url <chr>,  
## # languages\_url <chr>, stargazers\_url <chr>, contributors\_url <chr>,  
## # subscribers\_url <chr>, subscription\_url <chr>, commits\_url <chr>,  
## # git\_commits\_url <chr>, comments\_url <chr>, issue\_comment\_url <chr>,  
## # contents\_url <chr>, compare\_url <chr>, merges\_url <chr>, archive\_url <chr>,  
## # downloads\_url <chr>, issues\_url <chr>, pulls\_url <chr>,  
## # milestones\_url <chr>, notifications\_url <chr>, labels\_url <chr>,  
## # releases\_url <chr>, deployments\_url <chr>, created\_at <chr>,  
## # updated\_at <chr>, pushed\_at <chr>, git\_url <chr>, ssh\_url <chr>,  
## # clone\_url <chr>, svn\_url <chr>, size <int>, stargazers\_count <int>,  
## # watchers\_count <int>, language <chr>, has\_issues <lgl>,  
## # has\_downloads <lgl>, has\_wiki <lgl>, has\_pages <lgl>, forks\_count <int>,  
## # open\_issues\_count <int>, forks <int>, open\_issues <int>, watchers <int>,  
## # default\_branch <chr>, homepage <chr>