PROJECT 1 QUESTION 2

Introduction

There are different movie genres available in IMDb website for viewers to choose their preferences. Among the genres, musical movies and animation movies are selected to find the possible relationship referred to the selected info of movies. Singings are usually involved in musical movies while animation movies are mostly children friendly.

Data Collection and Preparation

Due to 50 movies listed in one webpage, there are 150 most popular movies from the first three pages of the site for each genre. The information gained from the website are movie title, released year, duration (in minutes), rating (from 1 to 10), number of votes (number of users cast in rating) and the name(s) of director(s).

The URL links of webpages are used in the functions to scrap the required data from the website and the path needed in the functions for the extraction of components is produced using SelectorGadget. Then, the function do.call() is used to execute all the values of functions. Further works of extraction, removal and structure change are performed to get the exact values needed for a data frame. Basic data frames are created each for both genres of movies with all the directors in one row. Another data frames are with directors in separate rows individually with other variables kept the same and duplicated.

The name of data, values and functions are shown in Figure 1. The names are indicated by "m" or "musical" for their relations to musical movies information whereas musical movies information by "a" or "animation". Figure 2 shows the first six rows of basic data frames. The movie titles generated in the data which differ from the titles shown in website may be due to translation factor. Figure 3 shows the data of first ten rows with directors split into rows.Data with missing values in any variables are listed as illustrated in Figure 4 to check with the website.

The missing values of rating and votes only happen for the movies released during 2023, 2024 and 2025 which are not available to rate yet. Those with duration values are the movies confirmed to be released within these two months. Movies with directors without release year are in post-production stage with release date unscheduled. Meanwhile, Moana is the only movie in the list with no values for all variables. It is just a project in development announced as live action of Moana (2016). The data are maintained as the missing values occur due to

unconfirmed release date, production in progress or votes for rating not available. However, the NA will be omitted in the data analysis if necessary.



Figure 1: Collection of objects, variables and functions

>	head(musical)						
	title year duration r	ating	votesNo			director	
1	Peter Pan & Wendy 2023 106	4.2	13915			David Lowery	
2	The Little Mermaid 2023 135	NA	NA			Rob Marshall	
- 3	Wish 2023 NA	NA	NA	Chris Bu	ıck, Fawr	Neerasunthorn	
4	Wonka 2023 NA	NA	NA			Paul King	
-	Wicked: Part One 2024 NA	NA	NA			Jon M. Chu	
	Joker: Folie à Deux 2024 NA	NA	NA			Todd Phillips	
>	head(animation)						
	title	year	duration	rating	votesNo		director
1	The Super Mario Bros. Movie	2023	92	7.3	89144	Aaron Horvath,	Michael Jelenic, Pierre Leduc, Fabien Polack
2	Suzume no Tojimari	2022	122	7.8	14419		Makoto Shinkai
3		2023	NA	NA.	NA		Chris Buck, Fawn Veerasunthorn
4	Puss in Boots: The Last Wish	2022	102	7.9	124163		Joel Crawford, Januel Mercado
-	Spider-Man: Across the Spider-Verse		140	NA	NA		m Dos Santos, Kemp Powers, Justin K. Thompson
6	Spider-Man: Into the Spider-Verse	2018	117	8.4	558374	E	Bob Persichetti, Peter Ramsey, Rodney Rothman

Figure 2: Basic data frames

```
> head(directors_m,10)
                 title year duration rating votesNo
                                                                  director
     Peter Pan & Wendy 2023
                                 106
                                         4.2
                                               13915
                                                             David Lowery
                                                             Rob Marshall
    The Little Mermaid 2023
                                  135
                                          NA
                                                  NA
3
                  Wish 2023
                                                               Chris Buck
                                  NA
                                          NA
                                                  NA
                  Wish 2023
                                  NA
                                          NA
                                                  NA
                                                      Fawn Veerasunthorn
5
                 Wonka 2023
                                  NA
                                          NA
                                                                Paul King
                                                  NA
     Wicked: Part One 2024
6
                                  NA
                                          NA
                                                  NA
                                                                Jon M. Chu
                                                            Todd Phillips
  Joker: Folie à Deux 2024
                                  NA
                                          NA
                                                  NA
8
                Grease 1978
                                 110
                                         7.2
                                              286036
                                                           Randal Kleiser
                Sing 2 2021
                                         7.4
                                               74555
                                  110
                                                           Garth Jennings
                Sing 2 2021
10
                                               74555 Christophe Lourdelet
                                  110
 head(directors_a,10)
                                  title year duration rating votesNo
                                                                                director
                                                         7.3
7.3
1
           The Super Mario Bros. Movie 2023
                                                   92
                                                               89144
                                                                           Aaron Horvath
2
           The Super Mario Bros. Movie 2023
                                                   92
                                                                89144
                                                                         Michael Jelenic
                                                         7.3
7.3
3
           The Super Mario Bros. Movie 2023
                                                   92
                                                               89144
                                                                           Pierre Leduc
4
           The Super Mario Bros. Movie 2023
                                                   92
                                                               89144
                                                                           Fabien Polack
                                                         7.8
                                                               14419
                    Suzume no Tojimari 2022
                                                  122
                                                                         Makoto Shinkai
                                  Wish 2023
                                                   NA
                                                                  NA
                                                                             Chris Buck
                                                          NA
                                   Wish 2023
                                                   NA
                                                          NA
                                                                   NA Fawn Veerasunthorn
          Puss in Boots: The Last Wish 2022
                                                  102
                                                         7.9
                                                              124163
                                                                           Joel Crawford
                                                                          Januel Mercado
          Puss in Boots: The Last Wish 2022
                                                  102
                                                         7.9
                                                              124163
                                                  140
                                                                   NA Joaquim Dos Santos
10 Spider-Man: Across the Spider-Verse 2023
```

Figure 3: Director data frames

```
> musical[!complete.cases(musical),]
         title year duration rating votesNo
The Little Mermaid 2023 135 NA NA
                                                                                                          director
                                                                                                    Rob Marshall
                                                                            Chris Buck, Fawn Veerasunthorn
                           Wish 2023
                          Wonka 2023
            Wicked: Part One 2024
                                                 NA
                                                                                                       Jon M.
                                                                                                   Todd Phillips
Marc Webb
       Joker: Folie à Deux 2024
                    Snow White 2024
     The Color Purple 2023
Mufasa: The Lion King 2024
51
58
97
                                                                                                  Blitz Bazawule
                                                 NA
                                                           NΑ
                                                                     NA Samantha Jayne, Arturo Perez Jr.
         Mean Girls Musical
                                                 NA
                                                           NA
       Miraculous: Le Film 2023
Moana NA
99
                                                                                                       Jeremy Zag
120
                                                 NA
                    Aladdin 2 2025
134
                                                                                                     Guy Ritchie
> animation[!complete.cases(animation),]
                    title year duration rating votesNo
Wish 2023 NA NA NA
Spider-Man: Across the Spider-Verse 2023 140 NA NA
                                                                                                                                       Chris Buck, Fawn Veerasunthorn
                                                                                                       NA Joaquim Dos Santos, Kemp Powers, Justin K. Thompson
NA Jeff Rowe, Kyler Spears
NA Peter Sohn
15
34
         Teenage Mutant Ninja Turtles: Mutant Mayhem
Elemental
37
                                        Sonic the Hedgehog 3 2024
                                                                                                                                                                Jeff Fowler
    Chicken Run: Dawn of the Nugget NA
Strays 2023
The Lord of the Rings: The War of the Rohirrim 2024
Mufasa: The Lion King 2024
                                                                                                                                                                     Sam Fell
                                                                                                                                                             Josh Greenbaum
61
                                                                                            NA
                                                                                                                                                             Kenji Kamiyama
                                                                                                                                                   Barry Jenkins
Walt Dohrn, Tim Heitz
chell, Stephanie Stine
                                                                                            NA
                                       Trolls Band Together 2023
Kung Fu Panda 4 2024
98
100
                                                                                                                                       Mike Mitchell,
                             Ruby Gillman, Teenage Kraken 2023
Garfield 2024
                                                                                                                                             Kirk DeMicco, Faryn Pearl
Mark Dindal
116
118
                                                      Migration 2023
                                                                                                                                          Benjamin Renner, Guylo Homs
135
                              Resident Evil: Death Island 2023
                                                                                                                                                           Eiichirô Hasumi
```

Figure 4: Movies with missing values

Data Analysis and Comparison

The summary() function is applied for basic data frames to obtain the minimum and maximum values of numerical variables. By setting a maximum level, the function also gives the top director lists in descending order. The highest number of productions are 6 directed movies for both genre lists. The filter() function is used to check the movie details of the highest and lowest number of votes collected. The output is shown in Figure 5 and Figure 6 shows some comparisons made between the two genres.

```
rating
Min. :2.800
1st Qu:6.600
Median :7.050
Mean :6.857
3rd Qu:7.400
Max. :8.500
                                                                                                                     votesNo
                                                                                                                                                 directo
                                 Min. :1937
1st Qu.:1981
                                                         Min. : 74.00
1st Qu.: 92.75
                                                                                                                Min. : 149
1st Qu.: 38478
Median : 98837
                                                                                                                                              Length: 150
  Class :character
                                                                                                                                              Class :character
                                                         Median :106.00
Mean :109.24
3rd Qu.:123.25
  Mode :character
                                 Median :2005
Mean :1998
                                                                                                                Median: 98837
Mean: 138766
                                                                                                                                              Mode
                                                                                                                                                       :character
                                 Mean :1998
3rd Qu.:2019
                                                                                                                Mean :
3rd Qu.:
                                 Max.
                                             :2025
                                                                      :181.00
                                                                                                                            :1081243
                                            : 2
 > summary(animation)
    title
                                 year
Min.
                                                                                     rating
Min. :4.500
1st Qu.:6.800
Median :7.300
Mean :7.255
                                                                                                                                                 director
                                                              duration
                                                                                                                     votesNo
 Length:150
Class :character
Mode :character
                                 Min. :1953
1st Qu.:2004
Median :2014
                                                         Min. : 69.00
1st Qu.: 90.00
Median : 97.00
Mean : 98.31
                                                                                                                Min. : 3470
1st Qu.: 117252
Median : 199842
                                                                                                                                              Length:150
Class :character
                                                                                                                                                       :character
                                 Mean
                                             :2011
                                                                                                                Mean
                                                                                                                               279061
                                 3rd Qu.:2021
Max. :2024
NA's :1
                                                          3rd Qu.:105.00
Max. :140.00
NA's :13
                                                                                      3rd Qu.:7.800
Max. :8.600
NA's :15
                                                                                                                3rd Qu.: 359047
> # find out the director(s) with highest number of movie productions
> # it will be in decreasing order and just check on the top 6 director(s)
> summary(as.factor(directors_m$director),maxsum = 6)
Hamilton Luske Wilfred Jackson Clyde Geronimi John Musker Ron Cle
                                                                                                                                                 (Other)
 6 6 5
> summary(as.factor(directors_a$director),maxsum = 6)
                           Ron Clements Hayao Miyazaki
6 5
     John Musker
                                                                          John Lasseter
5.7 3470 Juan Jesús García Galocha
8.4 1138970 Andrew Stanton
```

Figure 5: Output of summary () and filter ()

Genre of movies	Musical	Animation
Latest year in the list	2025	2024
Oldest year in the list	1937	1953
Longest time (min)	181	140
Shortest time (min)	74	69
Highest rating (/10)	8.5	8.6
Lowest rating (/10)	2.8	4.5
Most voted movie (Titles/ Year/ No. of votes)	The Lion King/ 1994/ 1081243	WALL: E/ 2008/ 1138970
Least voted movie (Titles/ Year/ No. of votes)	Carmen/ 2022/ 149	Mummies/ 2023/ 3470
Most involved director(s)	Hamilton Luske, Wilfred Jackson	John Musker, Ron Clements

Figure 6: Comparison of musical movies and animation movies

Musical movies older than animation movies are still in the popular list and new ones are scheduled one year further than the animation genre. The rating range of musical movies (5.7) is also wider than animation movies (4.1) at the same time. From the table of comparison, the duration of musical movies (74 minutes to 181 minutes) is usually longer than that of animation movies (69 minutes to 140 minutes). The reason of least number of votes for both genres of movies may due to the short-accumulated time as they are both assumed to be the latest movies released. At the same time, the most voted movies are both very well known until today.

The two basic datasets are combined to create other two new dataset: one with movies that categorised under both genres and one with all movies involved. From 150 top musical movies and 150 top animation movies, there are altogether 265 movies with 35 of them (shown in Figure 7) are musical and animation movies.

Figure 7: Musical and animation genre list of movies

The most involved directors are acquired from both new datasets with split rows for the variable "director". John Musker and Ron Clements are found the most in musical animation movies. They are usually working together for a movie production. Hamilton Luske and Wilfred Jackson are observed to be close co-workers in musical movies except in Sleeping Beauty (1959) and Snow White and the Seven Dwarfs (1937). Peter Pan (1953) is their only musical animation production. Clyde Geronimi worked together with Hamilton Luske besides Fantasia (1940). Toy Story 2 (1999) is the only cooperation between John Lasseter and Lee Unkrich. Both of them and Hayao Miyazaki are the main directors in their own animation production.

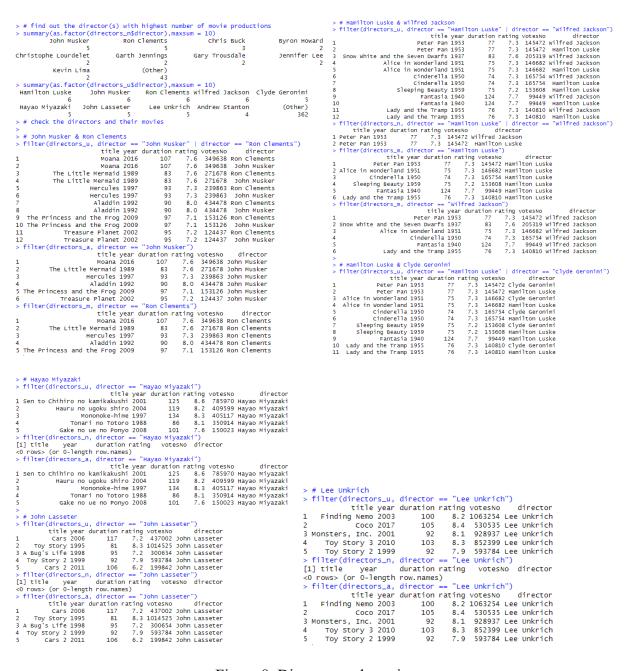
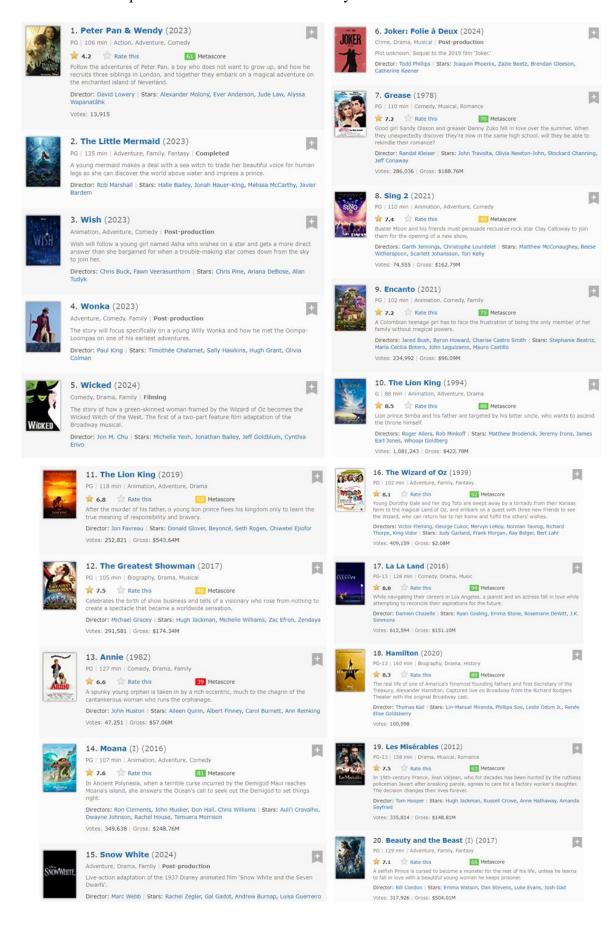


Figure 8: Directors and movies

Conclusion

The directors who work with musical movies tend to be related in animation genre and usually co-workers are involved which make the genres of movie musical and animation. Animation-focused directors are usually not involved in musical elements. Genres and directors make better movies.

APPENDIX A: Top 20 Musical movies as of 06 May 2023



APPENDIX B: Top 20 Animation movies as of 06 May 2023

