A first look at the effect of tech on animation movies

Research question: How has the prevalence and quality of animation movies developed as animation techniques have 'improved'?

The technology, tools, and special effects available to people making animation movies has been developing rapidly for many years. How has technological availability affected the number of animation movies being made? Have average ratings increased or decreased? In what years were the top 50 rated animated movies available to the database released?

In [40]:

```
#Importing NumPy, Pandas, Matplotlib
import numpy as np
import pandas as pd
import functools
import matplotlib.pyplot as plt
import matplotlib.axes as ax
```

Importing various databases

In [41]:

```
#Importing movie database
movies = pd.read_csv('/Users/anton/Desktop/movielens/movies.csv', sep=',')
movies.head()
```

Out[41]:

	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy

In [42]:

```
tags = pd.read_csv('/Users/anton/Desktop/movielens/tags.csv', sep=',')
tags.head()
```

Out[42]:

	userId	movieId	tag	timestamp
0	18	4141	Mark Waters	1240597180
1	65	208	dark hero	1368150078
2	65	353	dark hero	1368150079
3	65	521	noir thriller	1368149983
4	65	592	dark hero	1368150078

In [43]:

```
ratings = pd.read_csv('/Users/anton/Desktop/movielens/ratings.csv', sep=',', parse_dates=
['timestamp'])
ratings.head()
```

Out[43]:

	userId	movieId	rating	timestamp
0	1	2	3.5	1112486027
1	1	29	3.5	1112484676
2	1	32	3.5	1112484819
3	1	47	3.5	1112484727
4	1	50	3.5	1112484580

Creating a seperate column with the year

In [44]:

```
#count number of total movies
movies['year'] = movies['title'].str.extract('.*\((.*)\).*', expand=True)
movies.head()
```

Out[44]:

	movieId	title	genres	year
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	1995
1	2	Jumanji (1995)	Adventure|Children|Fantasy	1995
2	3	Grumpier Old Men (1995)	Comedy|Romance	1995
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance	1995
4	5	Father of the Bride Part II (1995)	Comedy	1995

In [45]:

```
movie_count = movies[['year','movieId']].groupby('year', as_index = False).count()
movie_count = movie_count[movie_count['year'] != '1975-1979']
movie_count = movie_count[movie_count['year'] != '1983)']
movie_count = movie_count[movie_count['year'] != '2007-']
movie_count = movie_count[movie_count['year'] != '2009-']
movie_count = movie_count[movie_count['year'] != 'Bicicleta, cullera, poma']
movie_count = movie_count[movie_count['year'] != 'Das Millionenspiel']
movie_count['number_of_movies'] = movie_count['movieId']
del movie_count['movieId']
movie_count.head()
```

Out[45]:

	year	number_of_movies
0	1891	1
1	1893	1
2	1894	2
3	1895	2
4	1896	2

Filtering for animation movies

In [46]:

```
#Filtering for animation moves - save the rows in which 'genres' contains 'animation'

animation_movies = movies[movies['genres'].str.contains('Animation')]
#animation_movies = animation_movies.set_index('movieId')
animation_movies.head()
```

Out[46]:

	movieId	title	genres	year
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	1995
12	13	Balto (1995)	Adventure|Animation|Children	1995
47	48	Pocahontas (1995)	Animation|Children|Drama|Musical|Romance	1995
236	239	Goofy Movie, A (1995)	Animation|Children|Comedy|Romance	1995
241	244	Gumby: The Movie (1995)	Animation|Children	1995

Cleaning up the data, albeit not as efficiently as possible

In [47]:

```
animation_movie_count = animation_movies[['year','movieId']].groupby('year', as_index = False).count()
animation_movie_count = animation_movie_count[animation_movie_count['year'] != '1975-1979']
animation_movie_count = animation_movie_count[animation_movie_count['year'] != '1983)']
animation_movie_count = animation_movie_count[animation_movie_count['year'] != '2007-']
animation_movie_count = animation_movie_count[animation_movie_count['year'] != '2009-']
animation_movie_count = animation_movie_count[animation_movie_count['year'] != 'Bicicleta, cullera, poma']
animation_movie_count = animation_movie_count[animation_movie_count['year'] != 'Das Millionenspiel']
animation_movie_count['number_of_animation_movies'] = animation_movie_count['movieId']
del animation_movie_count['movieId']
animation_movie_count.head()
```

Out[47]:

	year	number_of_animation_movies
0	1912	1
1	1914	1
2	1926	1
3	1928	1
4	1930	1

Counting the number of ratings corresponding to each movieId

In [48]:

```
rating_count = ratings[['movieId','rating']].groupby('movieId', as_index = False).count()
rating_count.head()
```

Out[48]:

	movieId	rating
0	1	49695
1	2	22243
2	3	12735
3	4	2756
4	5	12161

... and filtering out movies that have had less than 10 reviews. 10 was chosen arbitrarily.

In [49]:

```
at_least_ten_ratings = rating_count[rating_count['rating'] >= 10]
at_least_ten_ratings.head()
```

Out[49]:

	movieId	rating
0	1	49695
1	2	22243
2	3	12735
3	4	2756
4	5	12161

Finding the average rating for each movie, then deleting the column 'userId'

In [50]:

```
#Find average movie rating in ratings for each movie in animation_movies

average_rating = ratings.groupby('movieId', as_index = False).mean()
average_rating.head()
```

Out[50]:

	movieId	userId	rating
0	1	69282.396821	3.921240
1	2	69169.928202	3.211977
2	3	69072.079388	3.151040
3	4	69652.913280	2.861393
4	5	69113.475454	3.064592

In [51]:

```
del average_rating['userId']
average_rating.head()
```

Out[51]:

	movieId	rating
0	1	3.921240
1	2	3.211977
2	3	3.151040
3	4	2.861393
4	5	3.064592

Unused attempts, kept as references

In [52]:

```
#fulldf = pd.merge(animation_movies, average_rating, how = 'inner')
#fulldf.head()
```

In [53]:

```
#fulldf['rating'] = pd.to_numeric(fulldf['rating'])
#fulldf.head()
```

Merging the dataframes containing only animation movies, the average rating of each movie, and movies with at least 10 reviews/ratings, then making some edits.

In [54]:

```
from functools import reduce
used_dfs = [animation_movies, average_rating, at_least_ten_ratings]
fulldf = functools.reduce(lambda left, right: pd.merge(left, right, on='movieId'), used_
dfs)
fulldf.head()
```

Out[54]:

	movieId	title	genres	year	rating_x	rating_y
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	1995	3.921240	49695
1	13	Balto (1995)	Adventure|Animation|Children	1995	3.272416	1461
2	48	Pocahontas (1995)	Animation|Children|Drama|Musical|Romance	1995	2.946114	13046
3	239	Goofy Movie, A (1995)	Animation|Children|Comedy|Romance	1995	3.054751	3178
4	244	Gumby: The Movie (1995)	Animation|Children	1995	2.380702	285

In [55]:

```
#fulldf['year'] = fulldf['title'].str.slice(start=-5, stop=-1)
#fulldf.head()

#fulldf['year'] = fulldf['title'].str.extract('.*\((.*)\).*', expand=True)
fulldf['year'] = pd.to_numeric(fulldf['year'], downcast = 'float')
fulldf['avg_rating'] = pd.to_numeric(fulldf['rating_x'], downcast = 'float')
fulldf['number_of_reviews'] = pd.to_numeric(fulldf['rating_y'], downcast = 'float')
del fulldf['rating_x']
del fulldf['rating_y']

fulldf.head()
```

Out[55]:

	movieId	title	genres	year	avg_rating	number_of_reviews
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	1995.0	3.921240	49695.0
1	13	Balto (1995)	Adventure|Animation|Children	1995.0	3.272416	1461.0
2	48	Pocahontas (1995)	Animation|Children|Drama|Musical|Romance	1995.0	2.946114	13046.0
3	239	Goofy Movie, A (1995)	Animation|Children|Comedy|Romance	1995.0	3.054751	3178.0
4	244	Gumby: The Movie (1995)	Animation|Children	1995.0	2.380702	285.0

Finding the average rating per year (of animation movies)

In [56]:

```
average_rating_per_year = fulldf.groupby('year', as_index = False).mean()
average_rating_per_year.head()
```

Out[56]:

	year	movieId	avg_rating	number_of_reviews
0	1912.0	71472.0	3.892857	14.0
1	1926.0	40033.0	3.644737	38.0
2	1928.0	2102.0	3.443932	1302.0
3	1937.0	594.0	3.610492	17766.0
4	1939.0	2899.0	3.336780	677.0

Filtering out bad ratings, decided it wasn't needed.

In [57]:

```
#filter out bad ratings

#fulldf_good = fulldf[fulldf['avg_rating'] >= 3.5]
#fulldf_good.head()

#fulldf.drop(fulldf_index, inplace=True)
#fulldf_highratings.sort_values(by='avg_rating', ascending = False)
```

PRESENTATION OF RESULTS -

Table 1

Top 50 rated animation movies of all (up to 2015, with at least 10 ratings)

In [58]:

```
table_top_50 = fulldf.sort_values(by='avg_rating', ascending = False)[:50]
del table_top_50['movieId']
del table_top_50['genres']
del table_top_50['number_of_reviews']
table_top_50.head(50)
```

Out[58]:

	title	year	avg_rating
145	Spirited Away (Sen to Chihiro no kamikakushi) ...	2001.0	4.203810
34	Wallace & Gromit: The Wrong Trousers (1993)	1993.0	4.181068
22	Wallace & Gromit: A Close Shave (1995)	1995.0	4.167315
154	My Neighbor Totoro (Tonari no Totoro) (1988)	1988.0	4.149481
93	Creature Comforts (1989)	1989.0	4.120697
20	Wallace & Gromit: The Best of Aardman Animatio...	1996.0	4.109473
85	Princess Mononoke (Mononoke-hime) (1997)	1997.0	4.096299
171	Nausicaä of the Valley of the Wind (Kaze no ta...	1984.0	4.092082
149	Grave of the Fireflies (Hotaru no haka) (1988)	1988.0	4.089744
36	Grand Day Out with Wallace and Gromit, A (1989)	1989.0	4.066765
241	Howl's Moving Castle (Hauru no ugoku shiro) (2...	2004.0	4.066078
160	Laputa: Castle in the Sky (Tenkû no shiro Rapy...	1986.0	4.061917
607	Cowboy Bebop (1998)	1998.0	4.042373
326	WALL·E (2008)	2008.0	4.038929
366	Up (2009)	2009.0	4.038266
422	Toy Story 3 (2010)	2010.0	4.012973
340	FLCL (2000)	2000.0	4.006628
411	How to Train Your Dragon (2010)	2010.0	4.000420
384	Mary and Max (2009)	2009.0	3.995127
517	Presto (2008)	2008.0	3.982143
536	For the Birds (2000)	2000.0	3.976526
311	Persepolis (2007)	2007.0	3.971987
252	Man Who Planted Trees, The (Homme qui plantait...	1987.0	3.971939
217	Whisper of the Heart (Mimi wo sumaseba) (1995)	1995.0	3.966157
213	Kiki's Delivery Service (Majo no takkyûbin) (1...	1989.0	3.949879

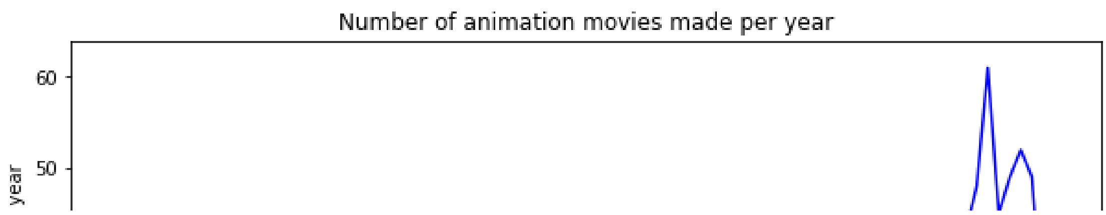
	title	year	avg_rating
21	Ghost in the Shell (Kôkaku kidôtai) (1995)	1995.0	3.944641
337	Waltz with Bashir (Vals im Bashir) (2008)	2008.0	3.928430
318	Girl Who Leapt Through Time, The (Toki o kaker...	2006.0	3.924745
253	Vincent (1982)	1982.0	3.923695
0	Toy Story (1995)	1995.0	3.921240
216	Porco Rosso (Crimson Pig) (Kurenai no buta) (1...	1992.0	3.918792
197	Incredibles, The (2004)	2004.0	3.908572
388	Balance (1989)	1989.0	3.902174
520	One Man Band (2005)	2005.0	3.895238
480	Lifted (2006)	2006.0	3.893701
380	Cameraman's Revenge, The (Mest kinematografich...	1912.0	3.892857
162	Finding Nemo (2003)	2003.0	3.886228
295	Ratatouille (2007)	2007.0	3.883866
557	Batman: The Dark Knight Returns, Part 2 (2013)	2013.0	3.882038
119	More (1998)	1998.0	3.881216
222	Neon Genesis Evangelion: The End of Evangelion...	1997.0	3.880711
126	Monsters, Inc. (2001)	2001.0	3.879444
551	Paperman (2012)	2012.0	3.878049
608	La Luna (2011)	2011.0	3.877778
159	Cowboy Bebop: The Movie (Cowboy Bebop: Tengoku...	2001.0	3.872483
458	BURN-E (2008)	2008.0	3.870927
167	Triplets of Belleville, The (Les triplettes de...	2003.0	3.868469
185	Lupin III: The Castle Of Cagliostro (Rupan san...	1979.0	3.867045
566	Wolf Children (Okami kodomo no ame to yuki) (2...	2012.0	3.866667
390	Fantastic Mr. Fox (2009)	2009.0	3.862851

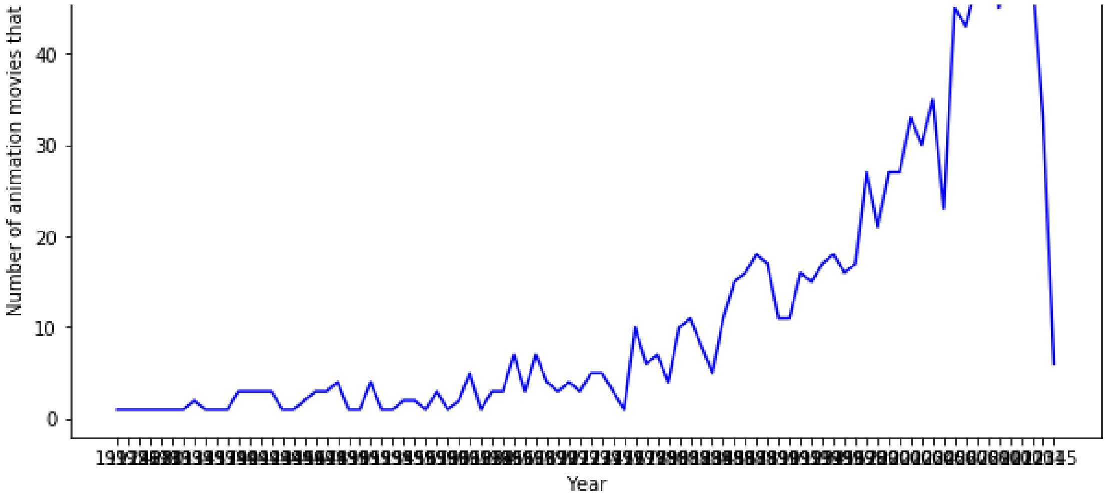
Graph 1

Number of animation movies total

In [59]:

```
x_number = animation_movie_count['year']
y_number = animation_movie_count['number_of_animation_movies']

plt.figure(figsize=(10, 6))

plt.plot(x_number, y_number, color='blue')

plt.xlabel('Year')
plt.ylabel('Number of animation movies that year')

plt.title('Number of animation movies made per year')

#plt.locator_params(axis='x', nticks=10)

plt.show()
```



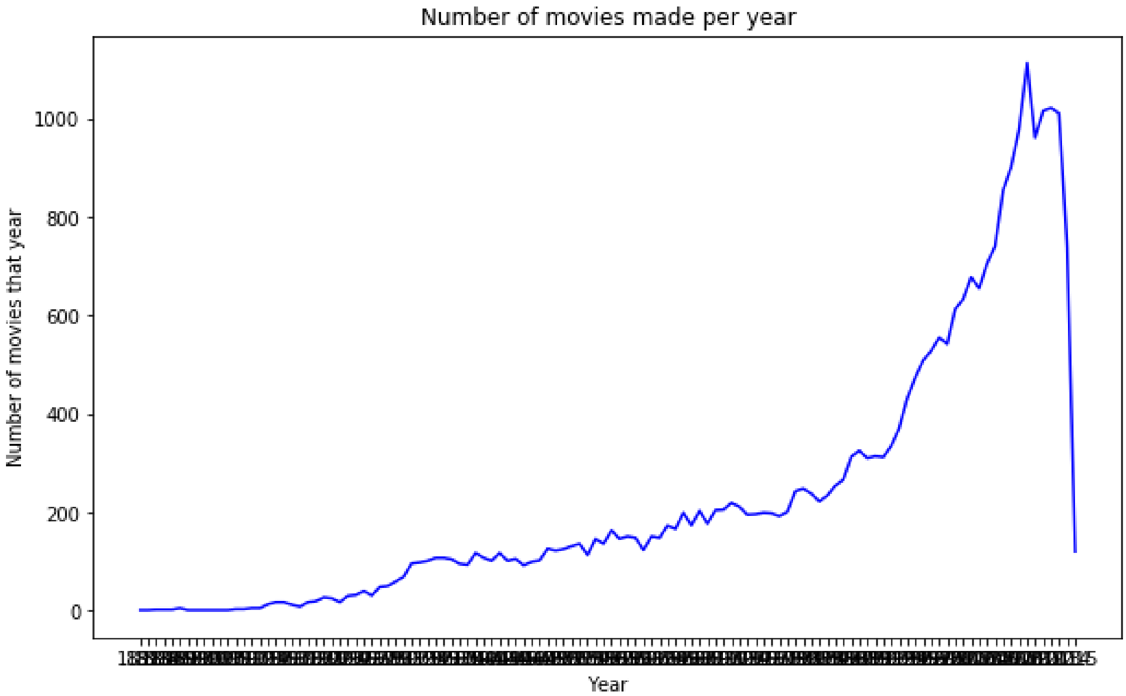


Graph 2

Number of movies total

In [60]:

```
x_number2 = movie_count['year'].values
y_number2 = movie_count['number_of_movies'].values

plt.figure(figsize=(10, 6))

plt.plot(x_number2, y_number2, color='blue')

plt.xlabel('Year')
plt.ylabel('Number of movies that year')

plt.title('Number of movies made per year')
plt.show()
```



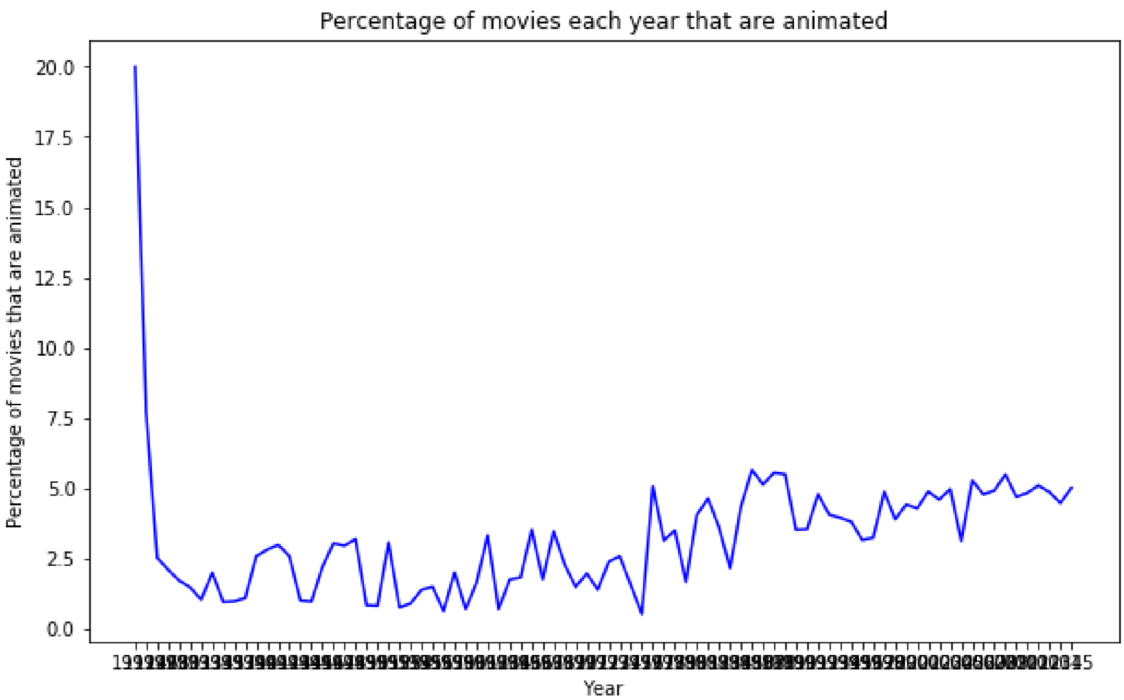
Graph 3

Number of movies that are animated movies

In [61]:

```
year_calc = pd.merge(movie_count, animation_movie_count, how = 'inner')
```

```
x_number3 = year_calc['year'].values
y_number3 = (year_calc['number_of_animation_movies'].values / year_calc['number_of_movies
'].values) * 100

plt.figure(figsize=(10, 6))

plt.plot(x_number3, y_number3, color='blue')

plt.xlabel('Year')
plt.ylabel('Percentage of movies that are animated')

plt.title('Percentage of movies each year that are animated')
plt.show()
```

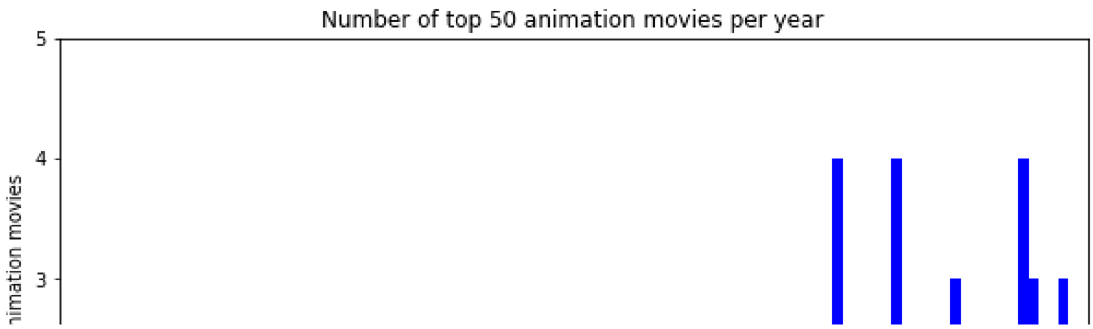


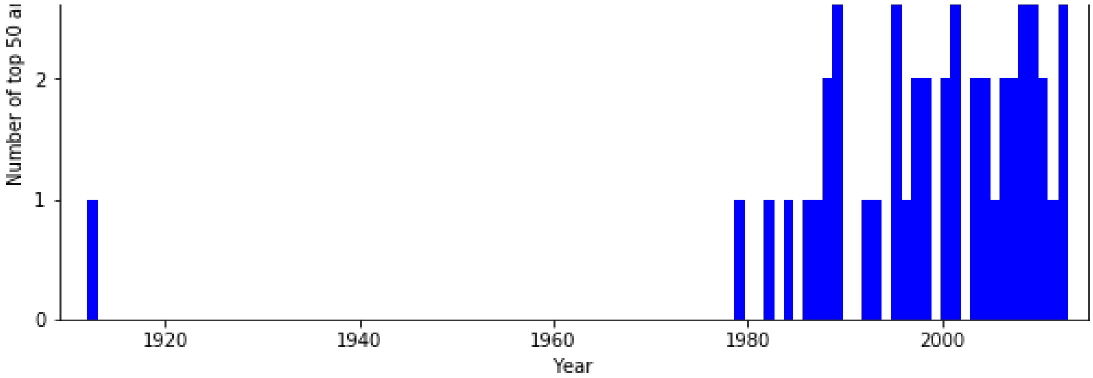
Graphs 4a and 4b

Histogram of years with number of top 50 movies

In [62]:

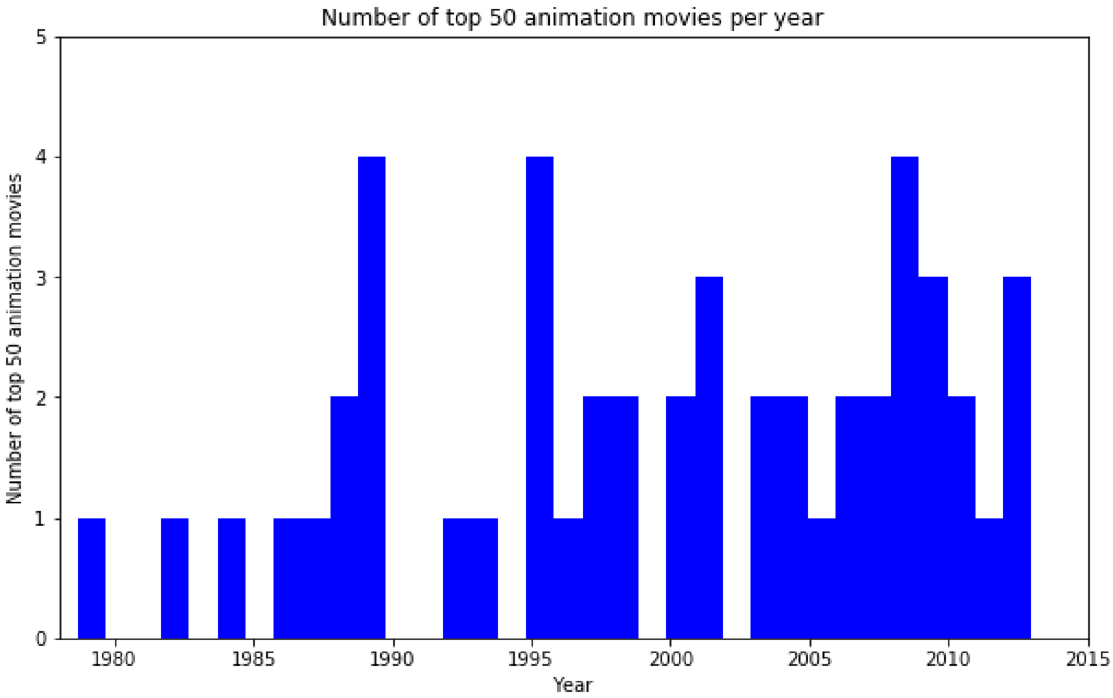
```
m_number = table_top_50['year'].values

plt.figure(figsize=(10, 6))

plt.hist(m_number, 100, density=False, facecolor='blue')

plt.xlabel('Year')
plt.ylabel('Number of top 50 animation movies')
plt.title('Number of top 50 animation movies per year')

plt.axis([1909, 2015, 0, 5])

plt.show()
```





In [63]:

```
m_number = table_top_50['year'].values

plt.figure(figsize=(10, 6))

plt.hist(m_number, 100, density=False, facecolor='blue')

plt.xlabel('Year')
plt.ylabel('Number of top 50 animation movies')
plt.title('Number of top 50 animation movies per year')

plt.axis([1978, 2015, 0, 5])

plt.show()
```



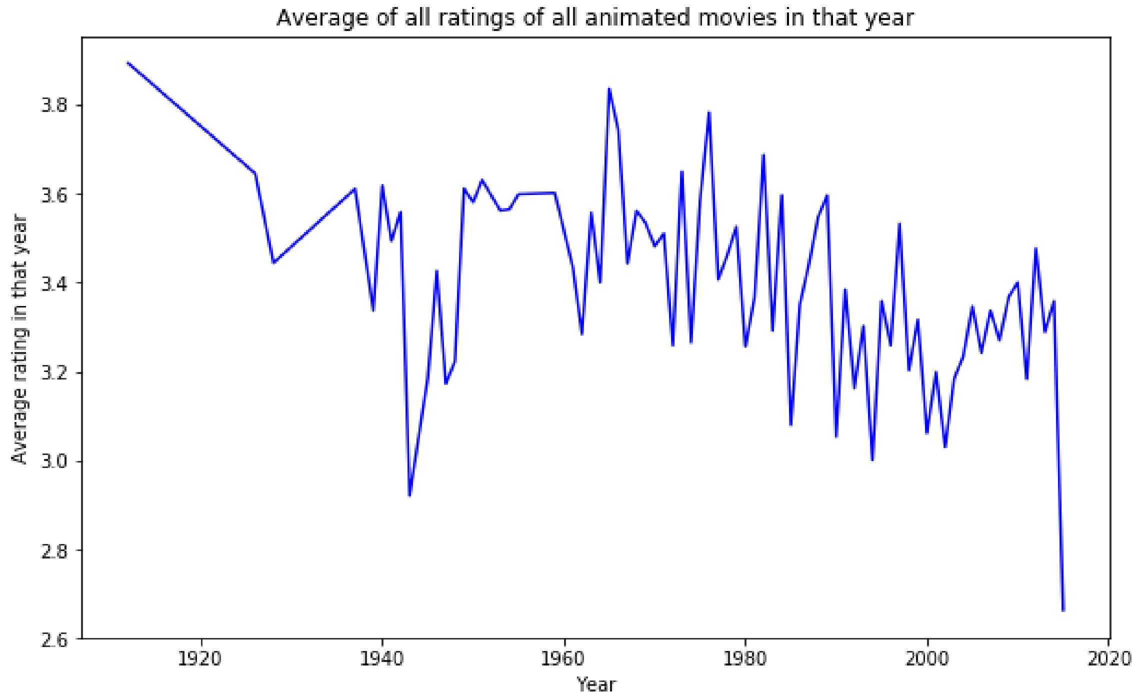
Graph 5

Average rating of animation movies per year (all ratings)

In [64]:

```
x_avg = average_rating_per_year['year'].values
y_avg = average_rating_per_year['avg_rating'].values

plt.figure(figsize=(10, 6))

plt.plot(x_avg, y_avg, color='blue')

plt.xlabel('Year')
plt.ylabel('Average rating in that year')
```

```
plt.title('Average of all ratings of all animated movies in that year')
plt.show()
```



3 Preliminary Conclusions:

1. **As time has gone on and for a number of reasons, the number of movies made per year has increased (graph 2). The number of animated movies made per year has followed a similar pattern (graph 1). The number of animation movies as a percentage of the total number of movies has slowly but surely increased (graph 3).**
2. **As animation movies have become more prevalent, so have the number of highly rated animation movies. However, the frequency of top 50 animation movies (judging by the criteria that they receive at least 10 reviews) has been fairly consitent since around 1980, as shown in the histogram (graph 4). All but one of the top 50 animation movies have been released since 1979.**
3. **The average of all ratings of animated movies (graph 5) shows that an increase in the number of animation movies has not lead to an increase in the overall (average) quality of said movies.**

It would seem that technology has enabled animation movies to be made more easily. Most of the best animation movies have been released in the last 40 years, indicating that advancing technology has played it's part. Within this time period, there is no clear correlation between the dramatic advances in animation movie technology and the overall quality of animation movies OR the frequency of excellent animation movies.

In [ ]:

Does quantity mean quality for animation movies?

Anton Baker
from Berlin, Germany

Dataset Used

The IMDB Movie Dataset was used for this Mini Project.

Motivation

As a fan of ‘The Simpsons’, I am very fond of the first couple of seasons. Apart from being subjectively better written, I very much prefer the more cartoon-style animations of the old episodes to the computer-style animations of today(see next slide). I find some modern animations and CGI effects in movies to be a little too over-the-top and soulless.

This got me thinking: the rapid evolution of computers and the technological possibilities they offer to people making animation movies must have had an effect on the number of animation movies made. It would be interesting to see whether or on the perceived quality of these movies has also increased or decreased.

Old vs. New Animation Graphics

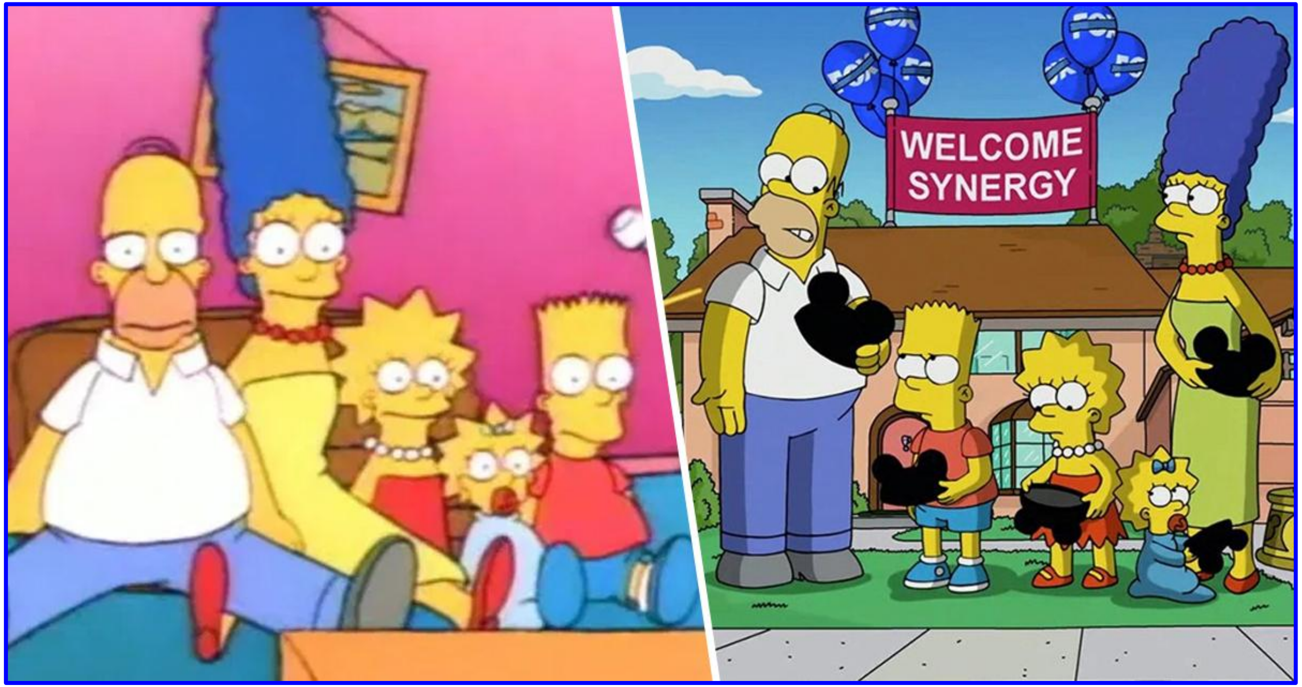


Image Source: 20th Television / Disney / Twitter via https://www.unilad.co.uk/film-and-tv/the-simpsons-is-30-years-old-today/

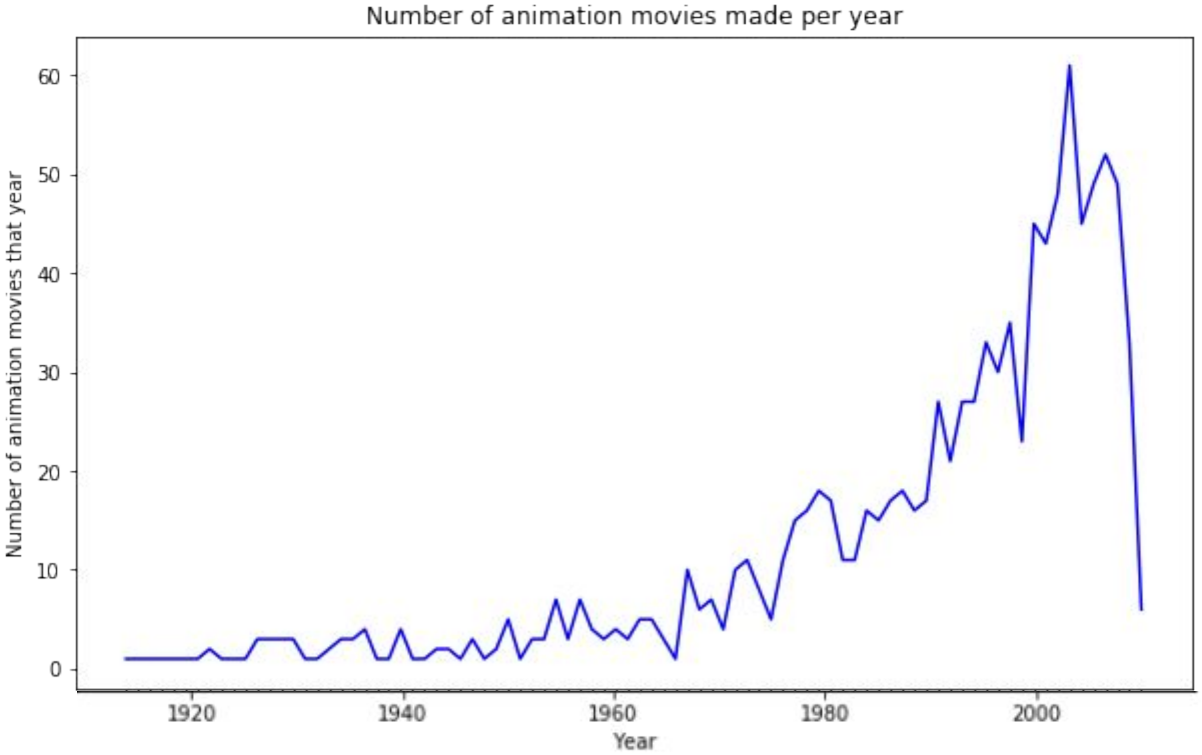
Research Question(s)

How has the prevalence and quality of animation movies developed as animation techniques have evolved as judged by average ratings on the Movie Dataset provided in the Python for Data Science course?

Noticeable Increase in Animation Movies

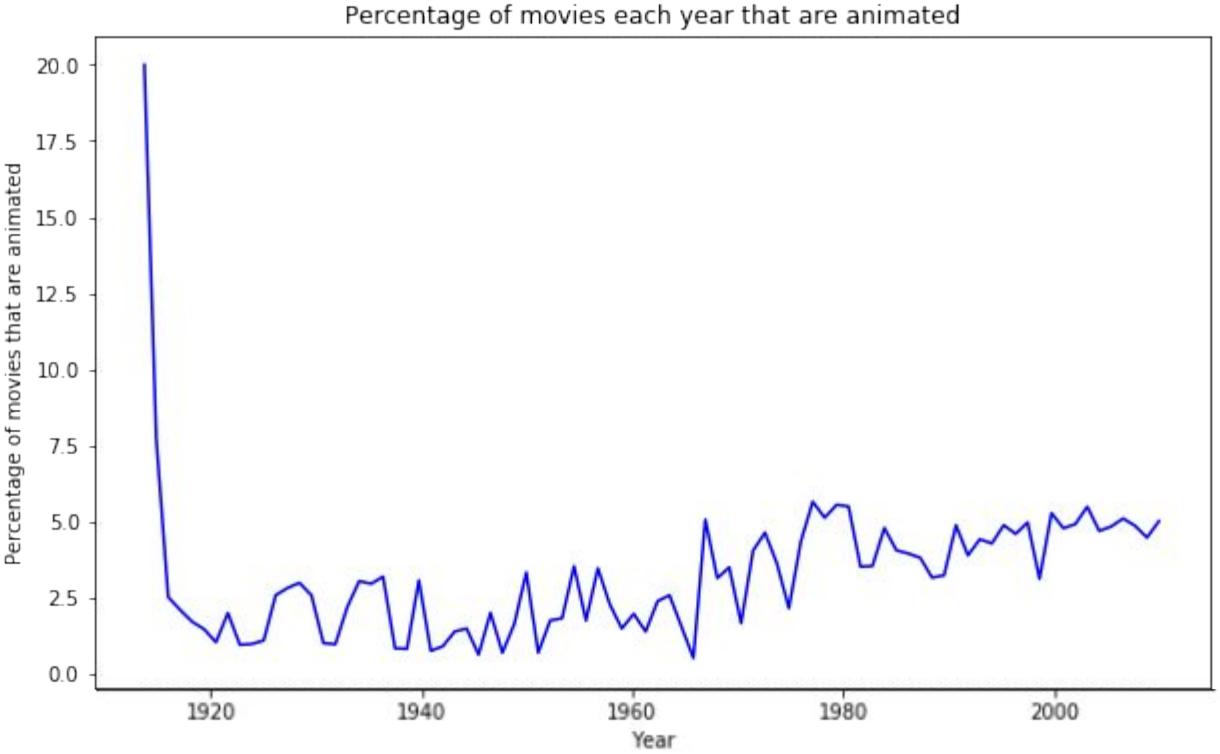
More and more animation movies have been made each year, with the number starting to increase more rapidly in the 1970s.

The data for the year 2015 does not indicate a sudden downturn, merely that the data from this database ends somewhere in the middle of 2015.



Slowly Growing Share of Animation Movies

As a percentage of total movies, animation movies have slowly but surely become more and more popular, with roughly 1-in-20 movies being an animation movie by the year 2000.



Ranking the Top Animation Movies

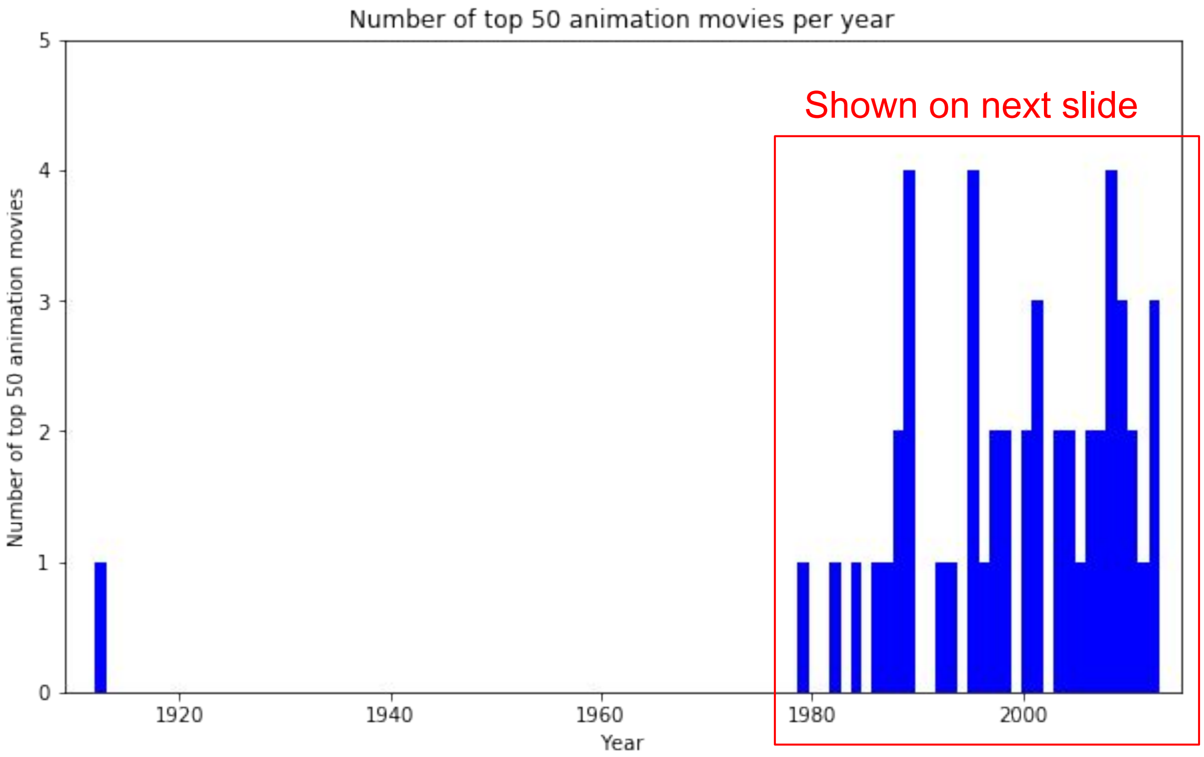
Seen on the right is a list of the top rated animation movies (up until 2015). The criteria was that the movie must have received at least 10 ratings, the average of which would determine the score (out of 5).

Of the top 15 movies seen on this slide, all were made between 1984 and 2009. The histogram on the next slide shows how many of the top 50 animation movies were released each year.

	title	year	avg_rating
145	Spirited Away (Sen to Chihiro no kamikakushi) ...	2001.0	4.203810
34	Wallace & Gromit: The Wrong Trousers (1993)	1993.0	4.181068
22	Wallace & Gromit: A Close Shave (1995)	1995.0	4.167315
154	My Neighbor Totoro (Tonari no Totoro) (1988)	1988.0	4.149481
93	Creature Comforts (1989)	1989.0	4.120697
20	Wallace & Gromit: The Best of Aardman Animatio...	1996.0	4.109473
85	Princess Mononoke (Mononoke-hime) (1997)	1997.0	4.096299
171	Nausicaä of the Valley of the Wind (Kaze no ta...	1984.0	4.092082
149	Grave of the Fireflies (Hotaru no haka) (1988)	1988.0	4.089744
36	Grand Day Out with Wallace and Gromit, A (1989)	1989.0	4.066765
241	Howl's Moving Castle (Hauru no ugoku shiro) (2...	2004.0	4.066078
160	Laputa: Castle in the Sky (Tenkû no shiro Rapy...	1986.0	4.061917
607	Cowboy Bebop (1998)	1998.0	4.042373
326	WALL·E (2008)	2008.0	4.038929
366	Up (2009)	2009.0	4.038266

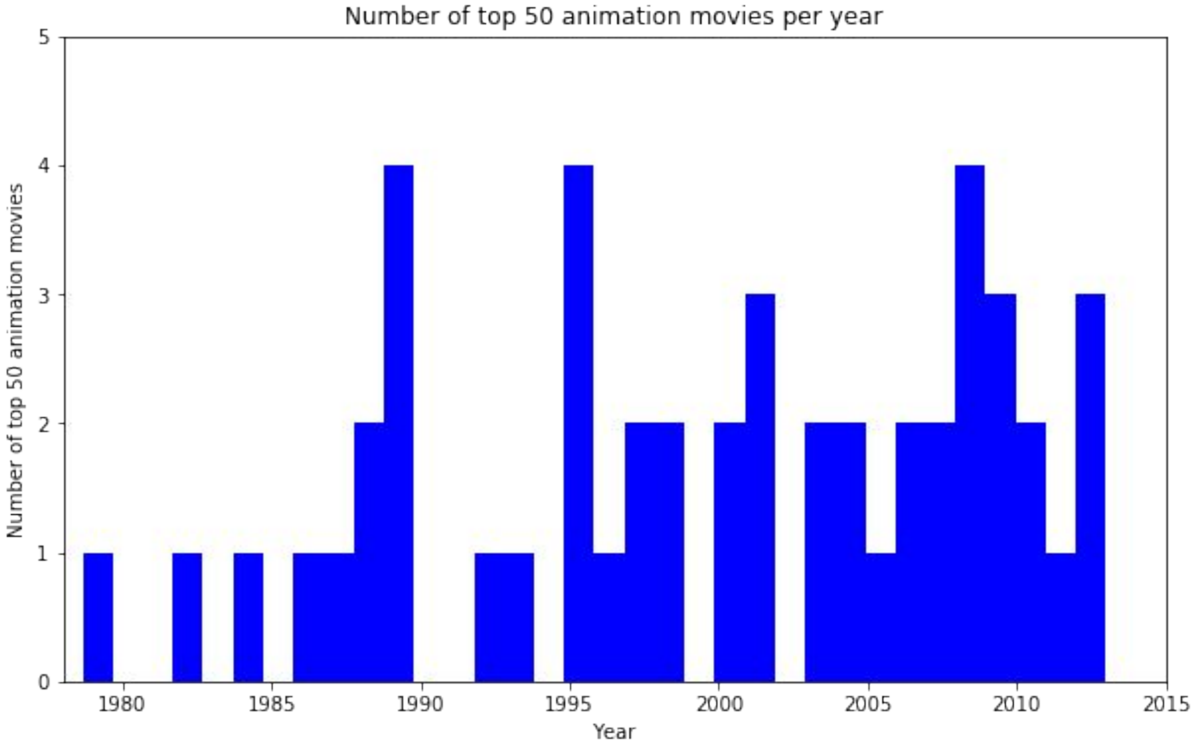
Most of the Top 50 since 1979

With the exception of **The Cameraman's Revenge** from 1912, all other 49 of the top 50 animation movies have come out since 1979. It would seem that certain technological advances have seen animation movies become a late bloomer among cinematic genres.



Consistent at the Top since 1979

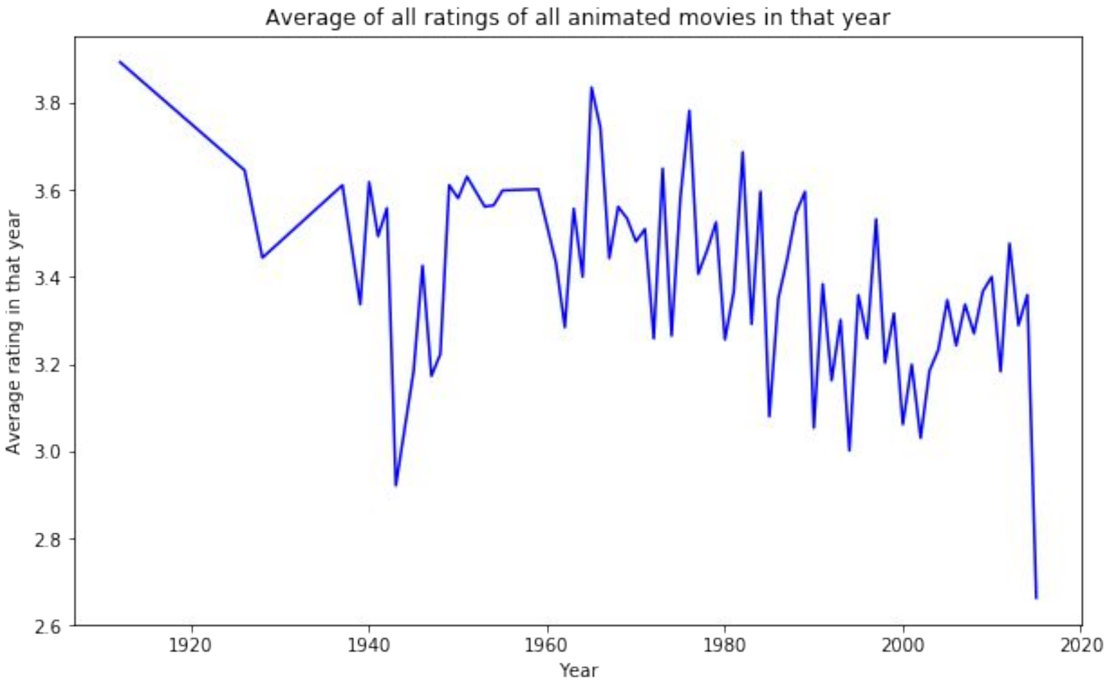
Among the 49 remaining movies, 26/49 have been released from 2000-2015 and 23/49 from 1979-1999. The distribution of excellent animation movies has been fairly consistent, even though there have been many advances in animation technology. It suggests that good animation technology ≠ good animation movie.



Declining Average Quality

By looking at the average rating of all animation movies (that received at least 10 ratings), we can see that the average perceived quality of animation movies has overall been declining over the past 50 years, although there has been a small increase since 2000.

Perhaps animation movies have become less novel.



Discussion

- ❖ Almost all of the top 50 animation movies have been released since 1979, suggesting that certain technological advances were necessary in giving filmmakers the tools they needed to make good animation movies.
- ❖ The number of animation movies has increased as a percentage of total movies made, the frequency of top animation movies has remained fairly consistent.
- ❖ Due to the fact that a lot more animation movies have been made, we have received more excellent animation movies recently than we used to.
- ❖ The overall average ratings of animation movies has fluctuated between 3.0/5.0 and 3.8/5.0.

Final Conclusion

By looking at the IMDB ratings of Simpsons episodes (the show has been running since 1989), we see that 49 out of the top 50 rated Simpsons episodes were released before 2000 - the exception being *Trilogy of Error* at 43rd place from the year 2001.

Although special effects and neat animation tricks can be fun, they can not obscure or counter what is otherwise a badly written or directed movie.

Plot, story, and characters are just as important in animation movies as they are in other genres.

Limitations

Movie ratings are objective, which is why only movies with at least 10 ratings were considered.

The link between technology and animation movies is only assumed based on general knowledge of how IT has developed over the past decades. There is no directly provable causality, merely an explainable correlation.

Acknowledgements

I have not received feedback from friends/ colleagues on this mini-project.

I have done some background research by looking at episode ratings of Simpsons episodes on IMDB (link below).

https://www.imdb.com/search/title/?series=tt0096697&view=simple&count=250&sort=user_rating,desc&ref_=tt_eps_rhs_sm

References

I have completed all the work on my own, excluding the countless pages on stackoverflow.com, pandas.pydata.org et al. visited. The picture source is listed directly below the picture, as is the IMDB link with the Simpsons episodes.

Thank you for your interest and attention :)