Has_Netflix_Become_Oversaturated_Over_Time_

December 9, 2023

[ ]:
```
from datascience import *
import numpy as np
%matplotlib inline
```

0.0.1 Has Netflix Become Oversaturated Over Time?

Recently a perception of the streaming giant as emerged, the content library of Netflix is so vast and overwhelming that it has impacted the usability and enjoyment of the service. But is there any truth to this perception, or have we just gotten used to the streaming landscape so that the vast libraries are no longer astonishing, but a mere annoyance?

Dataset provided by https://data.world/back2vizbasics

[ ]:
```
netflix_imdb = Table.read_table( "Netflix TV Shows and Movies.csv" )
netflix_imdb.show( 3 )
```

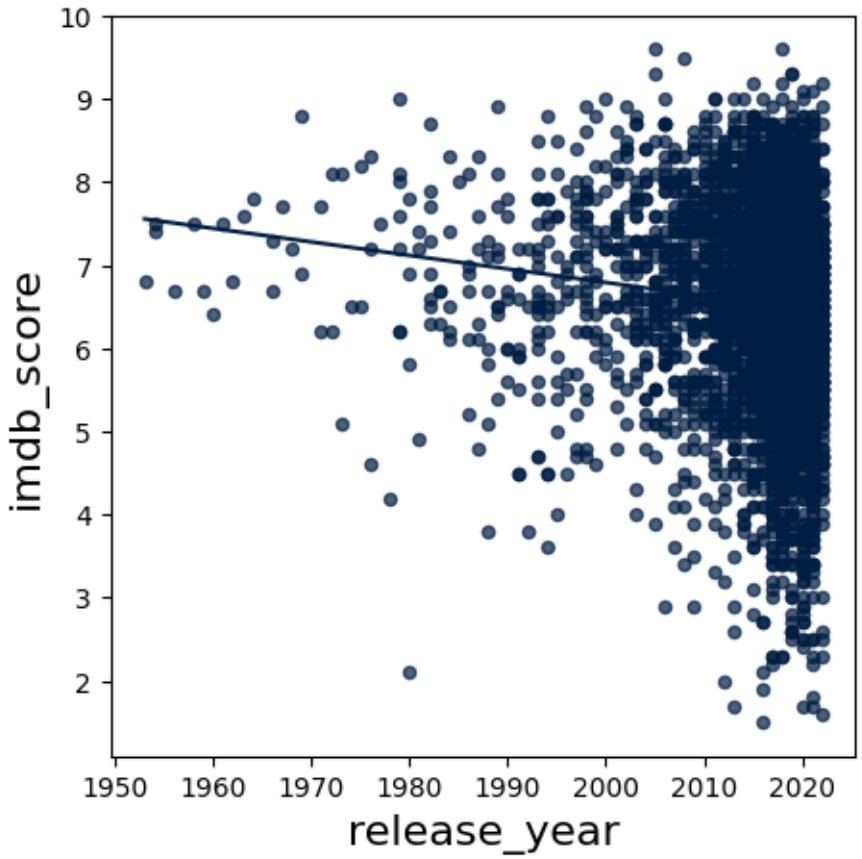
```
<IPython.core.display.HTML object>
```

From this dataset, we are shown items from Netflix's library, along with their associated IMDb Scores. Admittedly a flawed system for ranking the quality of programs, but a standard utilized by many consumers of media.

(TV Shows are ranked using the average of all episodes within the series)

#Let's begin by seeing the trend in quality overtime.

[ ]:
```
imdb_score_and_year = netflix_imdb.select(['imdb_score', 'release_year'])
imdb_score_and_year.scatter('release_year', fit_line = True )
```

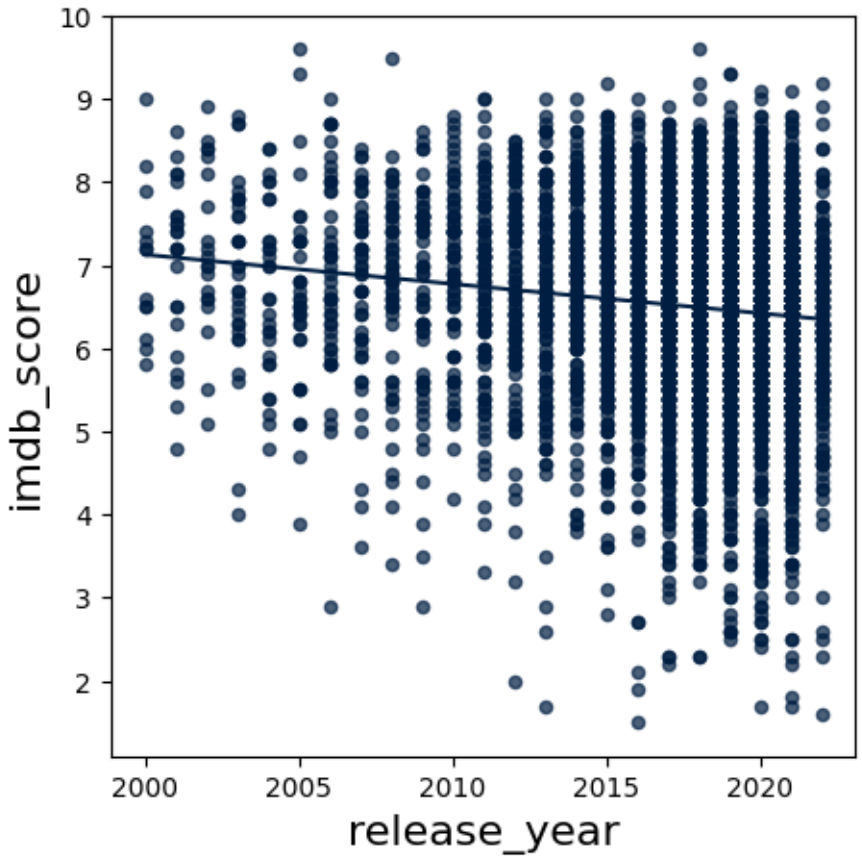


As we can already see, there has been an incredible increase of items since ~2005. However, there is so little data surrounding items from the 1999 and before. It may be beneficial to remove these from our dataset. Especially since Netflix would have minimal need to host old programs if they:

1. Weren't highly rated
2. Didn't have a large potential viewerbase

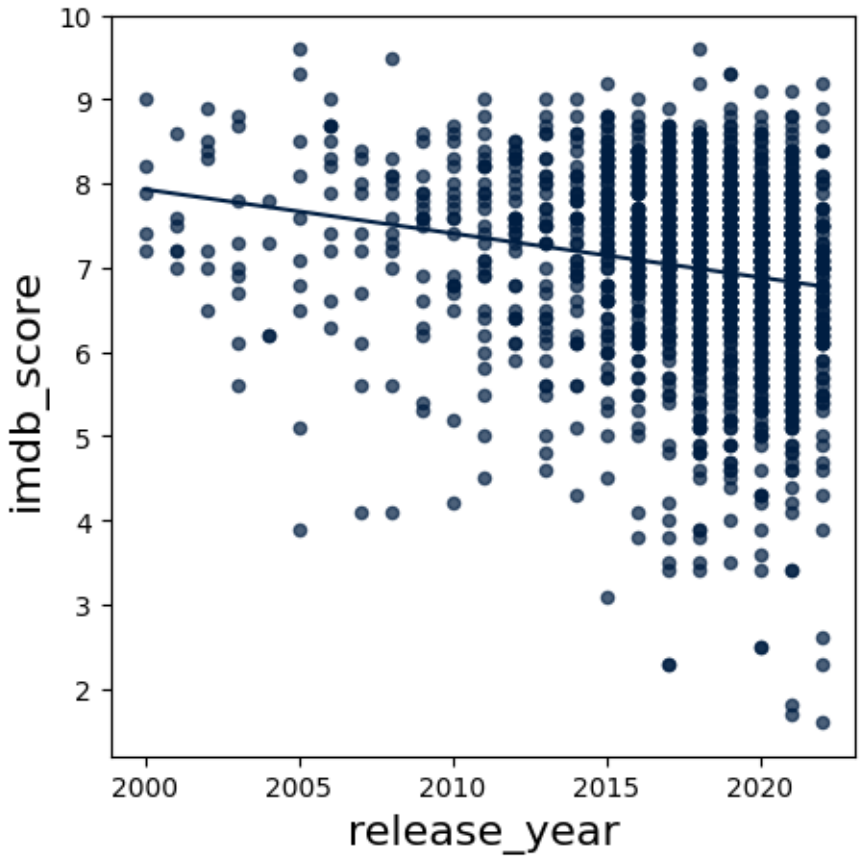
So let's limit our search.

[ ]:
```
after_2000 = netflix_imdb.where( netflix_imdb['release_year'] > 1999 )
after_2000_plot = after_2000.select( ['imdb_score', 'release_year'] )
after_2000_plot.scatter('release_year', fit_line = True )
```

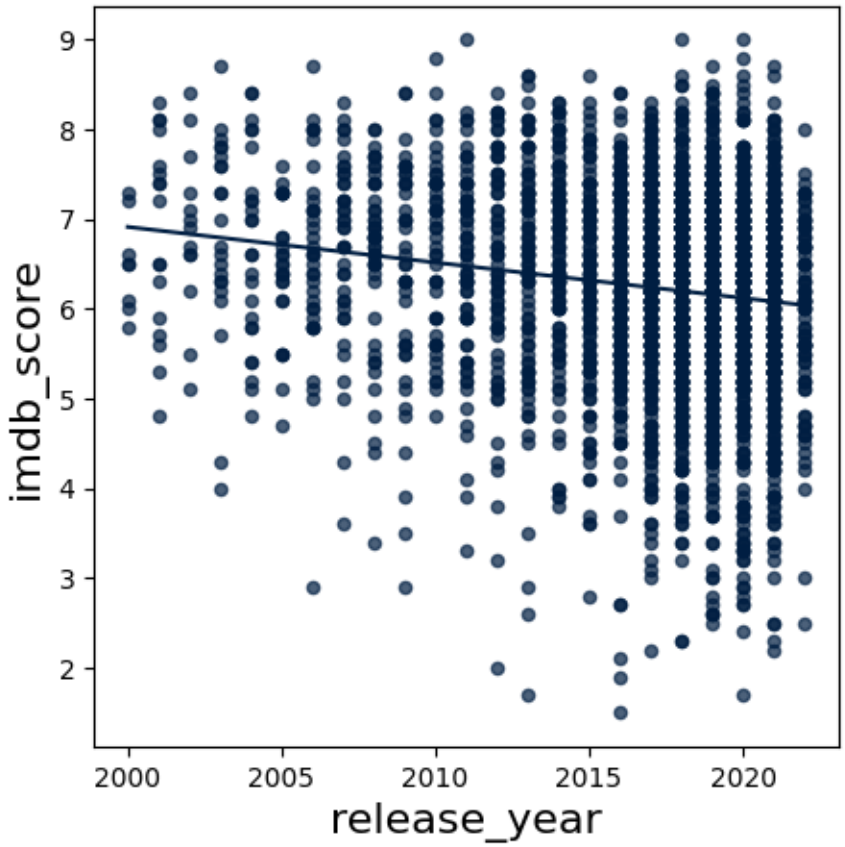


Much better, now we can see that there has been a very slight dip in quality over the years, especially with shows ranking under a two start showing up around 2012-2013 (This is an abysmally bad score). But how about movies and tv shows separated?

[ ]:
```
tv_shows = after_2000.where( after_2000['type'], 'SHOW' )
tv_shows_plot = tv_shows.select( ['imdb_score', 'release_year'] )
tv_shows_plot.scatter('release_year', fit_line = True )
```



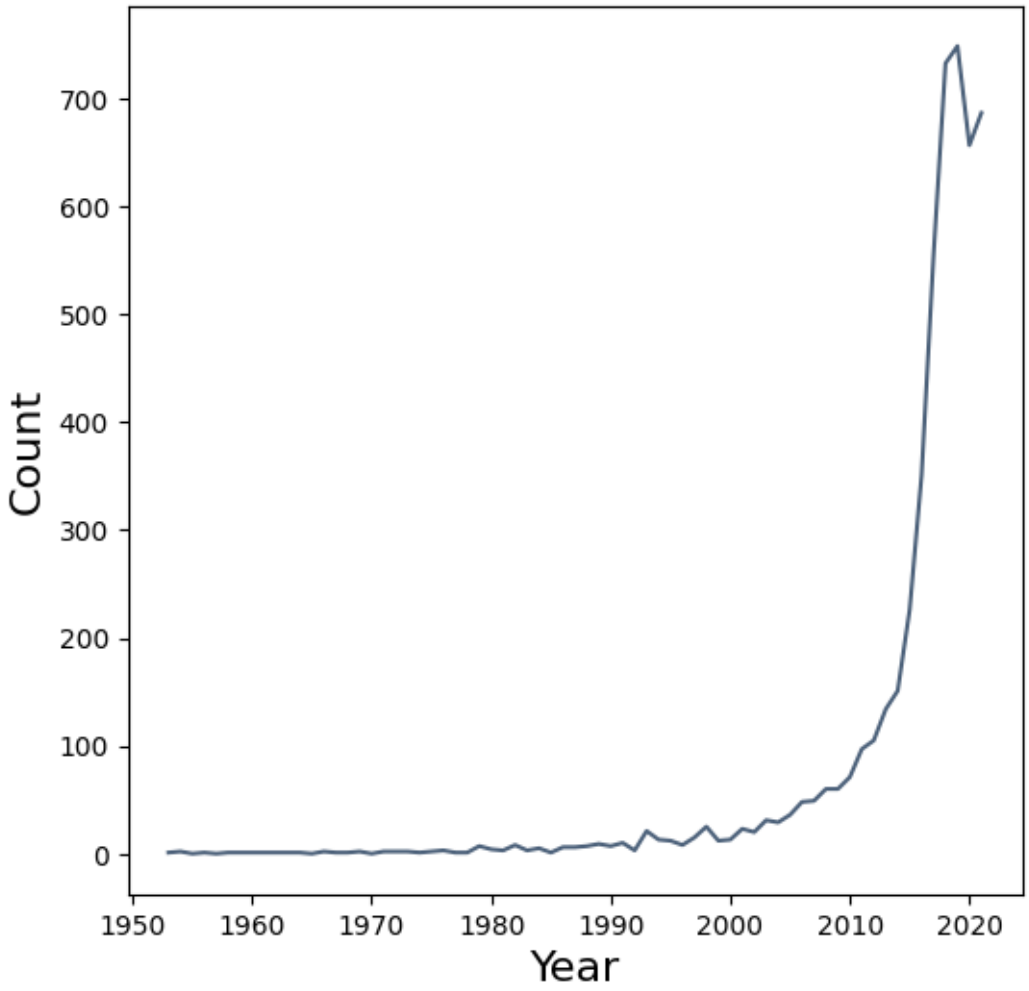
[ ]:
```
movies = after_2000.where( after_2000['type'], 'MOVIE' )
movies_plot = movies.select( ['imdb_score', 'release_year'] )
movies_plot.scatter( 'release_year', fit_line = True )
```

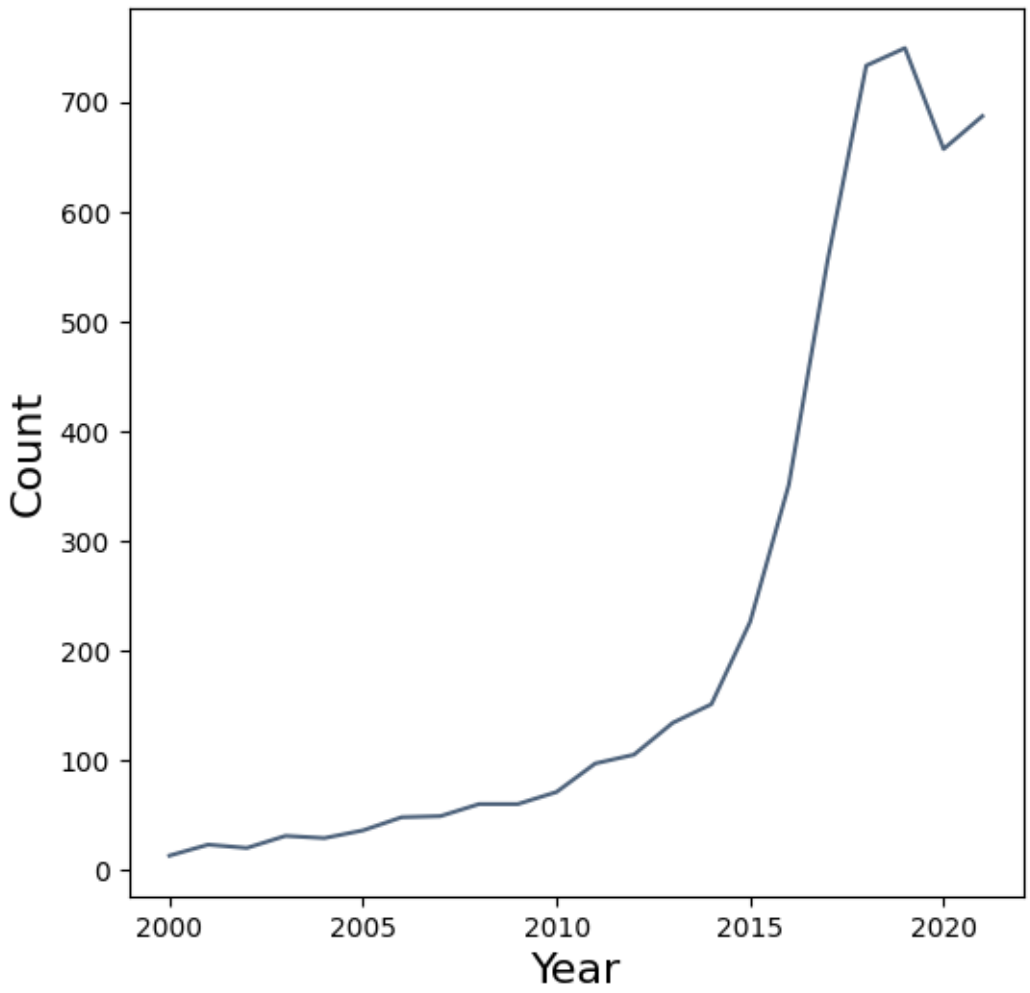


TV shows tend to be better rated overall, however there has been a steeper drop in quality than the Movies.

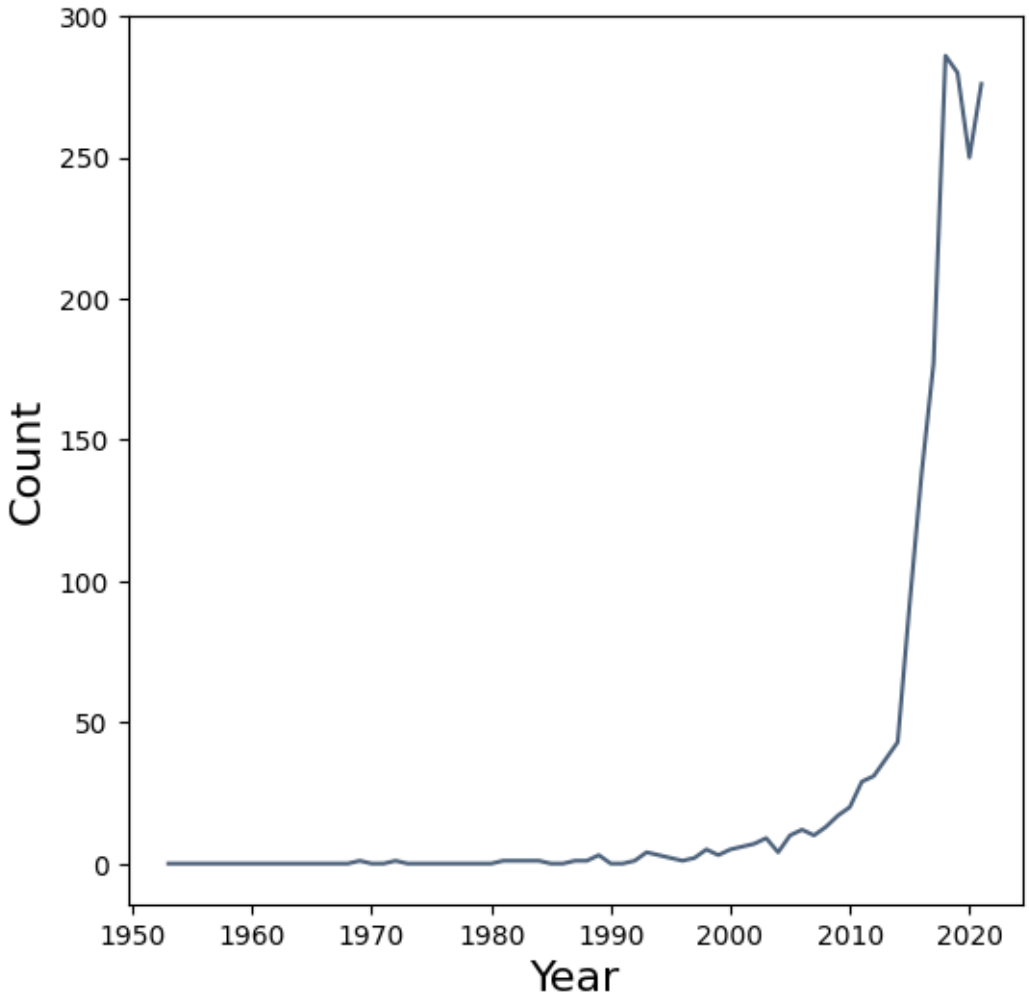
However let's see how the amount of content they added to the site changed overtime.

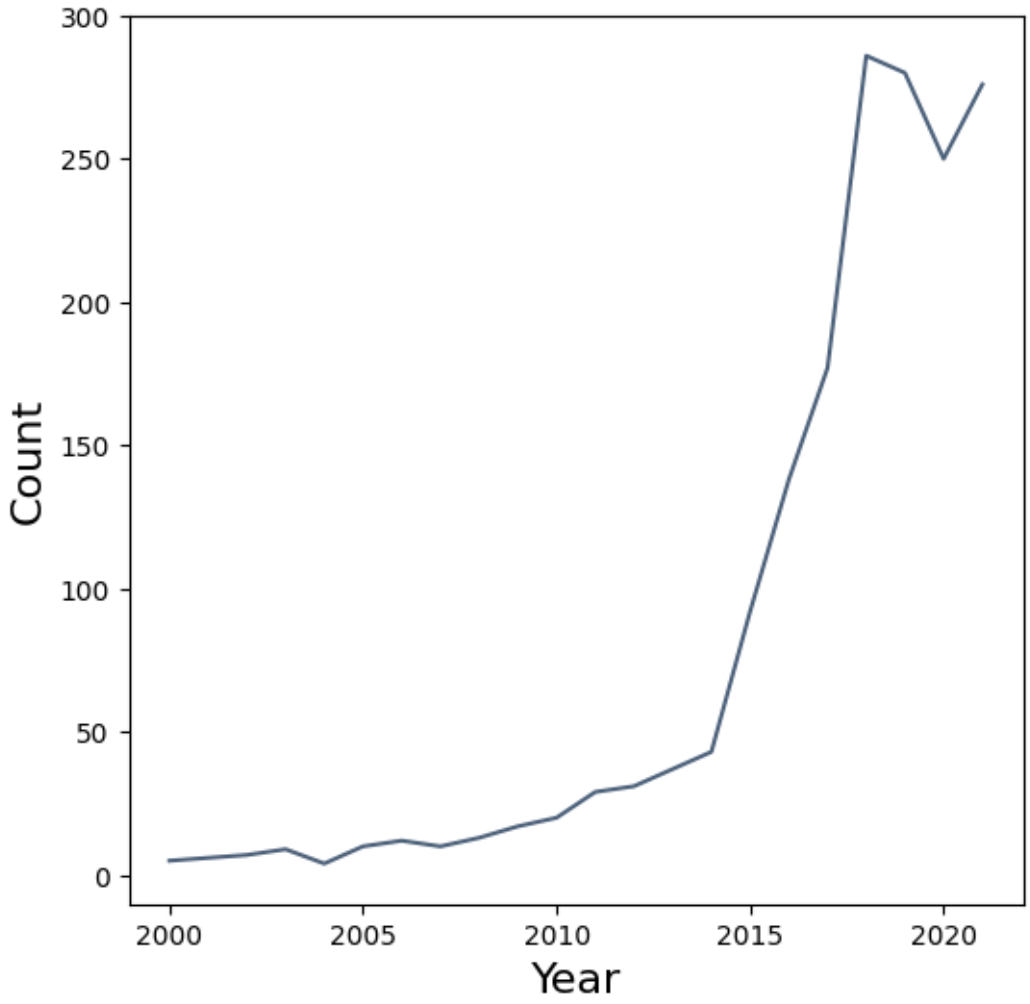
[ ]:
```
movies_per_year = Table.read_table( 'Movies_per_Year.csv')
movies_per_year.sort('Year').plot('Year')
movies_per_year_before_2000 = movies_per_year.where( movies_per_year['Year'] >=␣
↪2000 )
movies_per_year_before_2000.sort('Year').plot('Year')
```


Count
Year
0
100
200
300
400
500
600
700
1950
1960
1970
1980
1990
2000
2010
2020



[ ]:
```
shows_per_year = Table.read_table( 'Shows_per_Year.csv')
shows_per_year.sort('Year').plot('Year')
shows_per_year_before_2000 = shows_per_year.where( shows_per_year['Year'] >=␣
↪2000 )
shows_per_year_before_2000.sort('Year').plot('Year')
```


300
250
200
150
100
50
0
Count
1950
1960
1970
1980
1990
2000
2010
2020
Year



The vast, vast majority of items on netflix were from the year 2000 and beyond, with even more being added in around 2010. This trend continues to the time data is unavailable for this set (2022). This means that every year, Netflix's library is exploding with hundreds of items, creating that feeling of overwhelming indecision.

In conclusion there is truth to that experience, with Netflix's library slowly getting more and more flooded with items that are on average, not as good as what is in the rest of their library. This phenomenon has a name, 'shovelware', the process of filling the system with low-mid quality media to increase consumer pull, which the data shows Netflix may be suffering from.