

▼ Exercise 2

```
%matplotlib inline
import pandas as pd
import seaborn as sbn
sbn.set()

titles = pd.read_csv('/content/titles.csv')
titles.head()

cast = pd.read_csv('/content/cast.csv', index_col=None)
cast.head()
```

	title	year	name	type	character	n
0	Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
1	Suuri illusioni	1985	Homo \$	actor	Guests	22.0
2	Battle of the Sexes	2017	\$shutter	actor	Bobby Riggs Fan	10.0
3	Secret in Their Eyes	2015	\$shutter	actor	2002 Dodger Fan	NaN
4	Steve Jobs	2015	\$shutter	actor	1988 Opera House Patron	NaN

▼ What are the ten most common movie names of all time?

```
titles.title.value_counts().head(10)
```

```
Hamlet                20
Carmen                 17
Macbeth               16
The Outsider          12
Maya                  12
Temptation            12
Freedom               11
Honeymoon             11
Othello               11
The Three Musketeers  11
Name: title, dtype: int64
```

▼ Which three years of the 1930s saw the most films released?

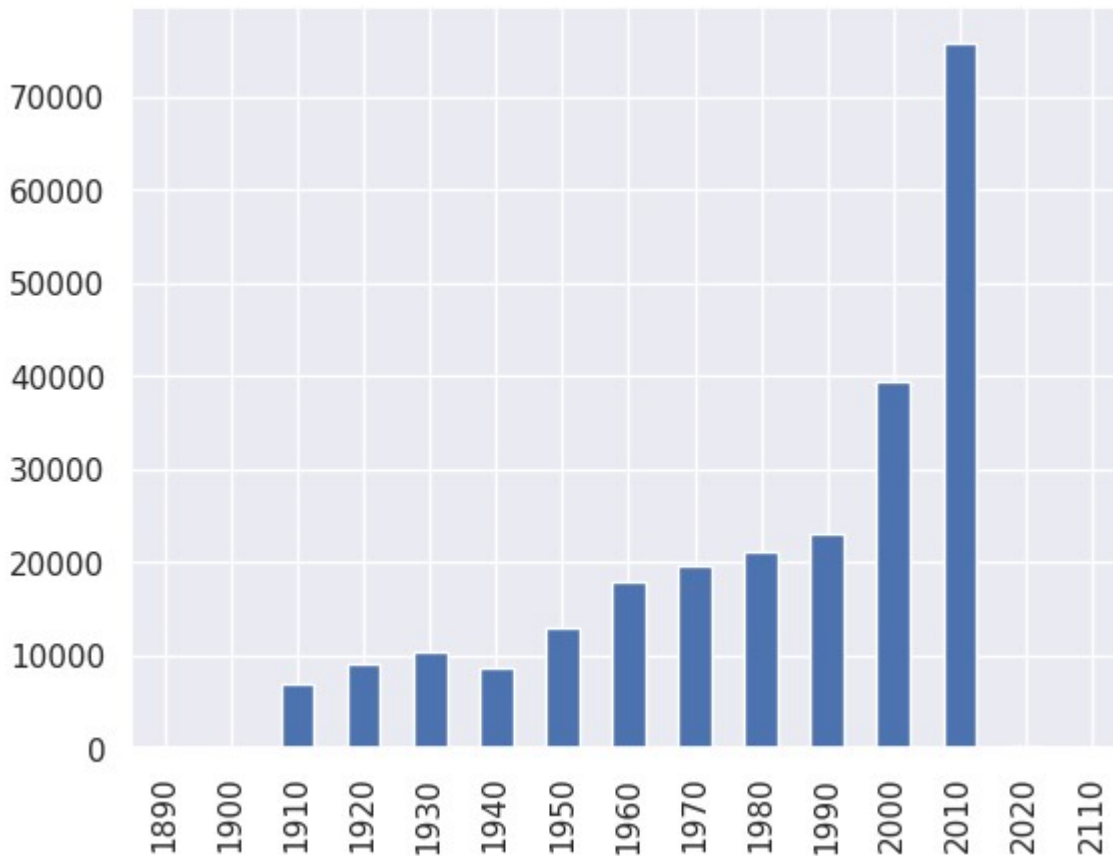
```
titles[titles.year // 10 == 193].year.value_counts().head(3)

1937    1224
1936    1169
1938    1163
Name: year, dtype: int64
```

Plot the number of films that have been released each decade over the history of cinema.

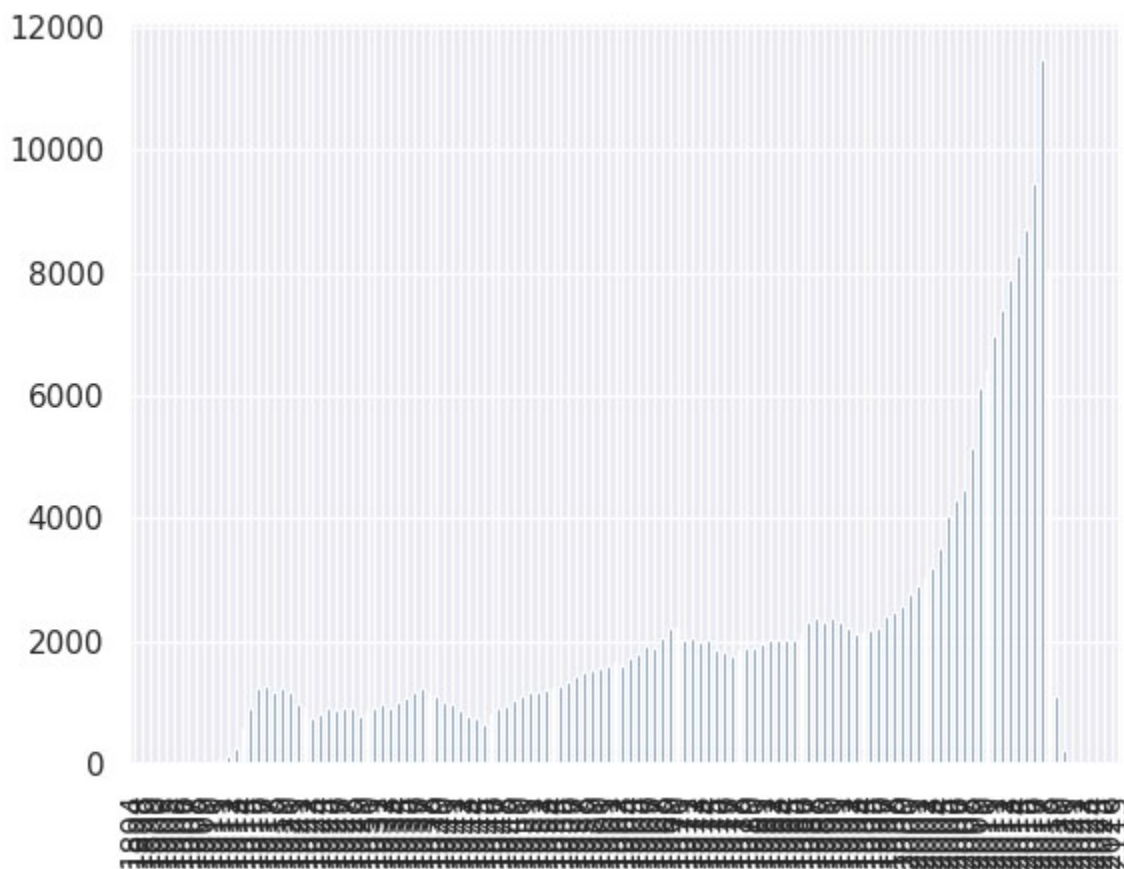
```
titles['decade'] = ((titles.year // 10) * 10)
titles.decade.value_counts().sort_index().plot(kind='bar')
```

<Axes: >



```
titles.year.value_counts().sort_index().plot(kind='bar')
```

<Axes: >

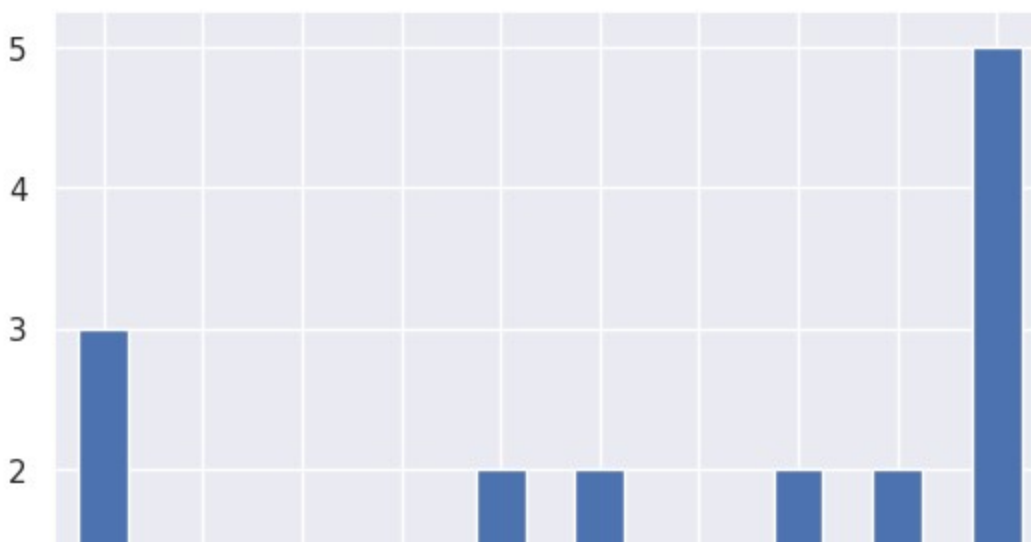


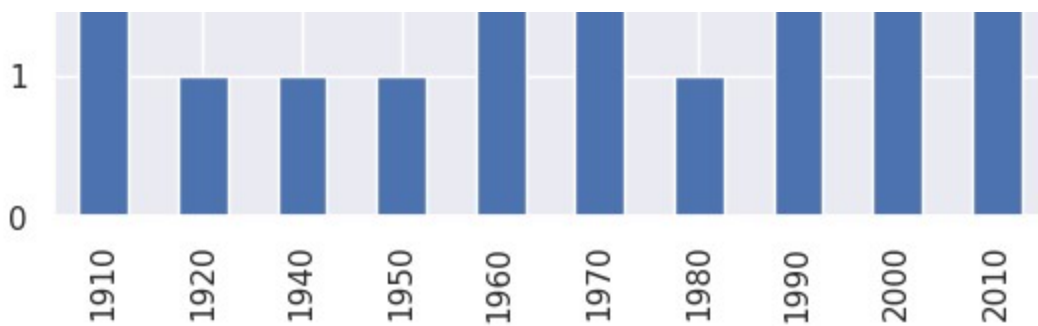
Double-click (or enter) to edit

Plot the number of "Hamlet" films made each decade.

```
titles['decade'] = ((titles.year // 10) * 10)
titles[titles.title=='Hamlet'].decade.value_counts().sort_index().plot(kind='bar')
```

<Axes: >

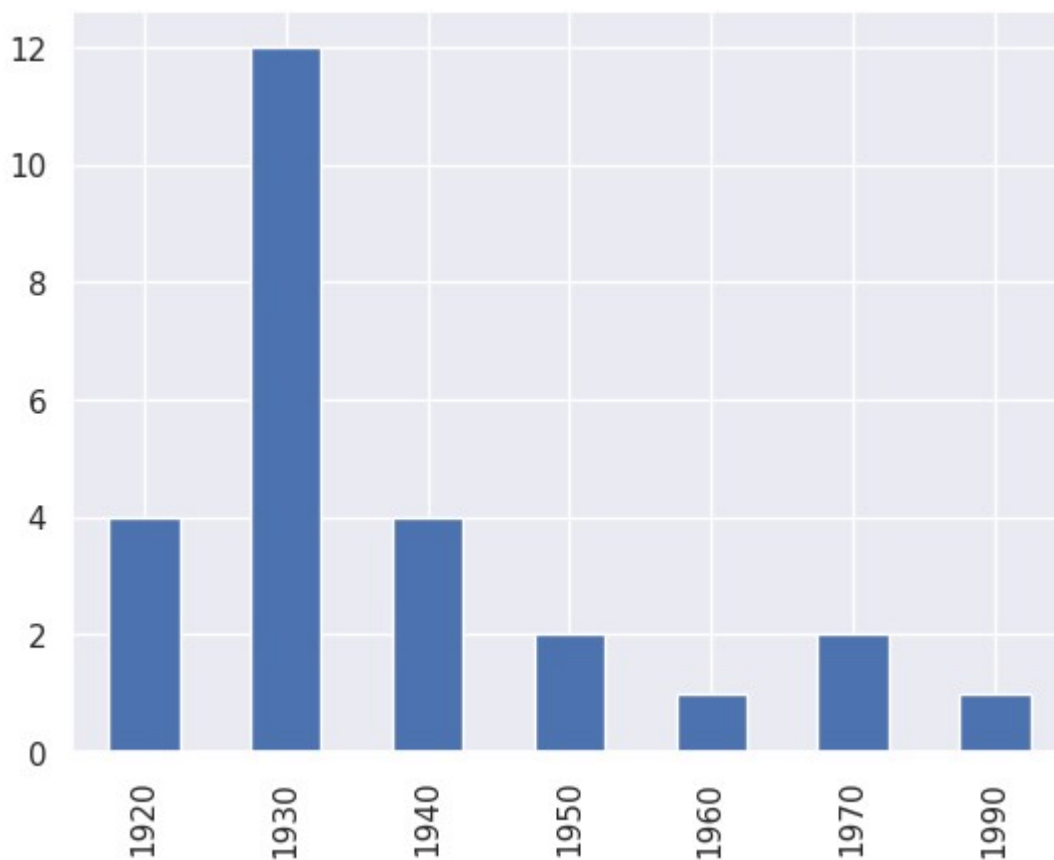




Plot the number of "Rustler" characters in each decade of the history of film.

```
cast['decade'] = ((cast.year // 10) * 10)
cast[cast.character=='Rustler'].decade.value_counts().sort_index().plot(kind='bar')
```

<Axes: >

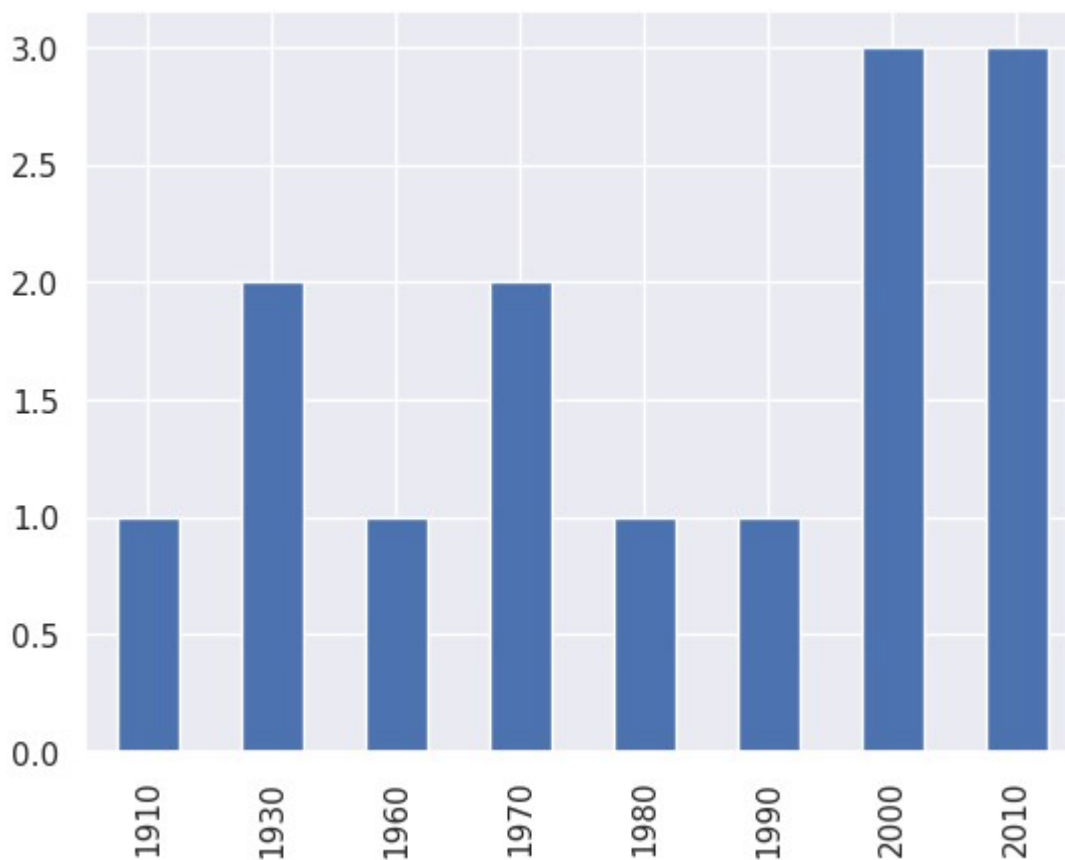


Plot the number of "Hamlet" characters each decade.

```
cast['decade'] = ((cast.year // 10) * 10)
```

```
cast['decade'] = ((cast.year // 10) * 10)
cast[cast.character=='Hamlet'].decade.value_counts().sort_index().plot(kind='bar')
```

<Axes: >



What are the 11 most common character names in movie history?

```
cast.character.value_counts().head(11)
```

```
Himself      3354
Townsmen     1159
Extra         1133
Policeman    1123
Doctor        1104
Reporter     1016
Waiter        947
Bartender     885
Party Guest   743
Soldier       731
Dancer        722
Name: character, dtype: int64
```

Who are the 10 people most often credited as "Herself" in film history?

```
cast[cast.character=='Herself'].name.value_counts().head(10)
```

```
Etta Baker          1
Ren? Biaggi         1
J. Blackfoot        1
Moucle Blackout     1
Babette Bombshell   1
Merrie Cherry       1
Lamb (III) Chop     1
Name: name, dtype: int64
```

Who are the 10 people most often credited as "Himself" in film history?

```
cast[cast.character=='Himself'].name.value_counts().head(10)
```

```
George W. Bush      25
Muhammad Ali        18
Jimmy Carter        14
Fidel Castro        13
Amitabh Bachchan    12
George Bush         12
Johnny Carson       10
Steve Allen         8
Diego Aramburu-Zabala 8
Neil Armstrong      7
Name: name, dtype: int64
```

Which actors or actresses appeared in the most movies in the year 1945?

```
cast[cast.year==1945].name.value_counts().head(10)
```

```
Horace B. Carpenter  23
Budd Buster          22
Eddie Acuff          20
Roy Bucko            20
Roy Barcroft         19
George Chesbro       19
```

```

George Cinescu      15
Ed Cassidy          18
Jimmy Aubrey        18
Hank Bell           18
Wheaton Chambers    17
Name: name, dtype: int64

```

Which actors or actresses appeared in the most movies in the year 1985?

```
cast[cast.year==1985].name.value_counts().head(10)
```

```

Raj Babbar          15
Asrani              13
Mithun Chakraborty  10
Lalu Alex           9
Birbal              9
Vikas Anand         9
Prem Chopra         9
Kar Lok Chin        8
Bahadur             8
Omero Capanna       7
Name: name, dtype: int64

```

Plot how many roles Mammootty has played in each year of his career.

```
import pandas as pd
import matplotlib.pyplot as plt
```

```

data = {
    'Year': [1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990],
    'Roles_Played': [2, 3, 1, 2, 4, 3, 2, 3, 5, 4, 3]
}

```

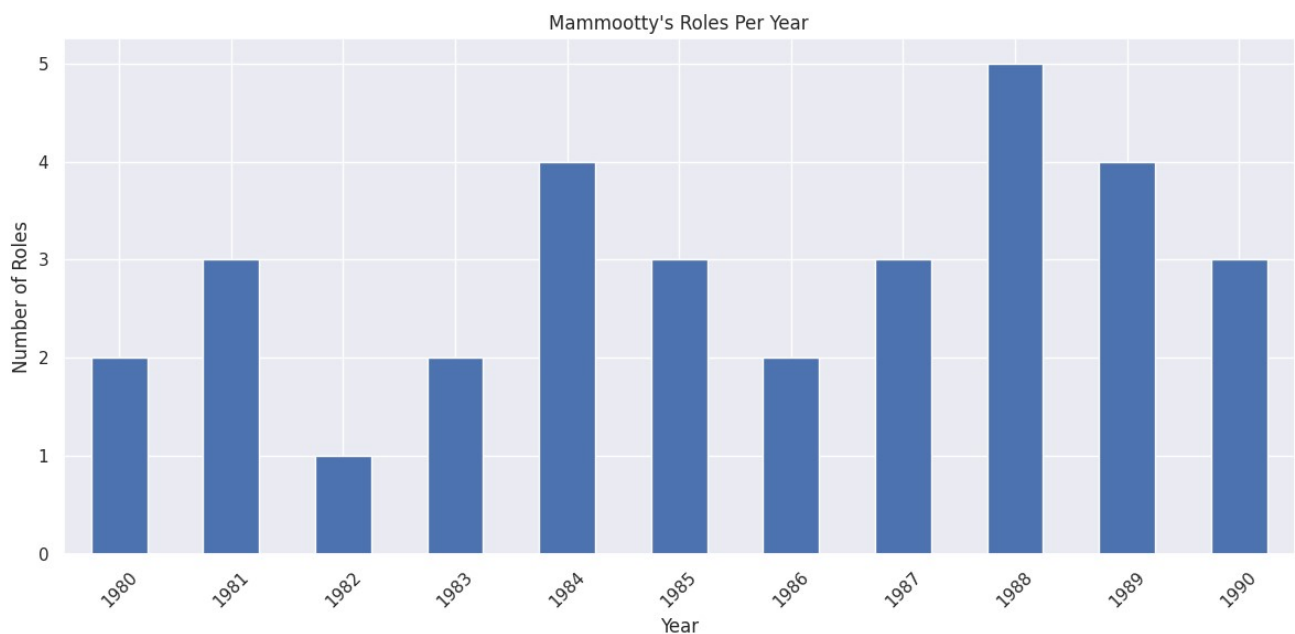
```

df = pd.DataFrame(data)
ax = df.plot(kind='bar', x='Year', y='Roles_Played', figsize=(12, 6), legend=False)
plt.xlabel('Year')
plt.ylabel('Number of Roles')
plt.title("Mammootty's Roles Per Year")
plt.xticks(rotation=45)

plt.tight_layout()

```

```
plt.show()
```



What are the 10 most frequent roles that start with the phrase "Patron in"?

```
c = cast[cast.character.str.startswith('Patron in')]
c.character.value_counts().head(10)
```

Patron in restaurant	2
Patron in Chinese Restaurant	1
Patron in Restaurant	1
Patron in Ballroom	1
Patron in Audience	1
Patron in Coffee Shop	1
Patron in Club	1
Patron in the Coffee House	1
Patron in Frisky Rabbit	1


```
Patron in Store          1
Name: character, dtype: int64
```

What are the 10 most frequent roles that start with the word "Science"?

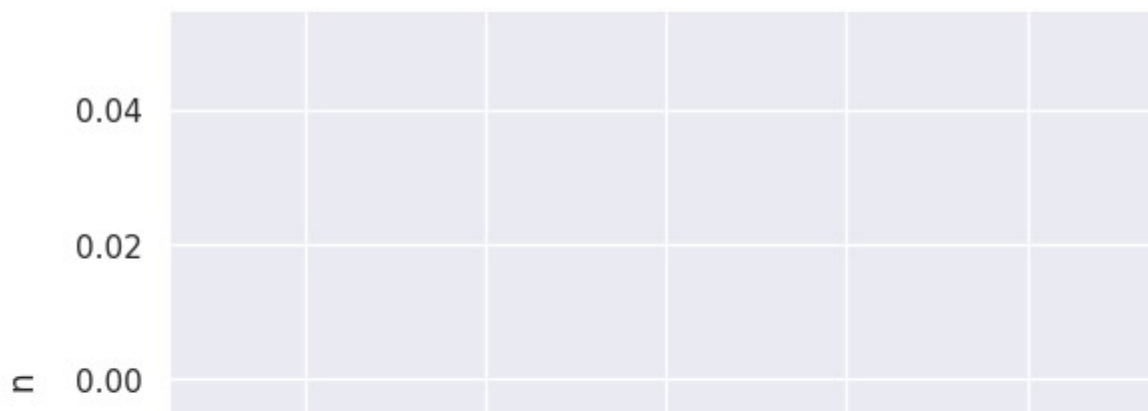
```
c = cast[cast.character.str.startswith('Science')]
c.character.value_counts().head(10)
```

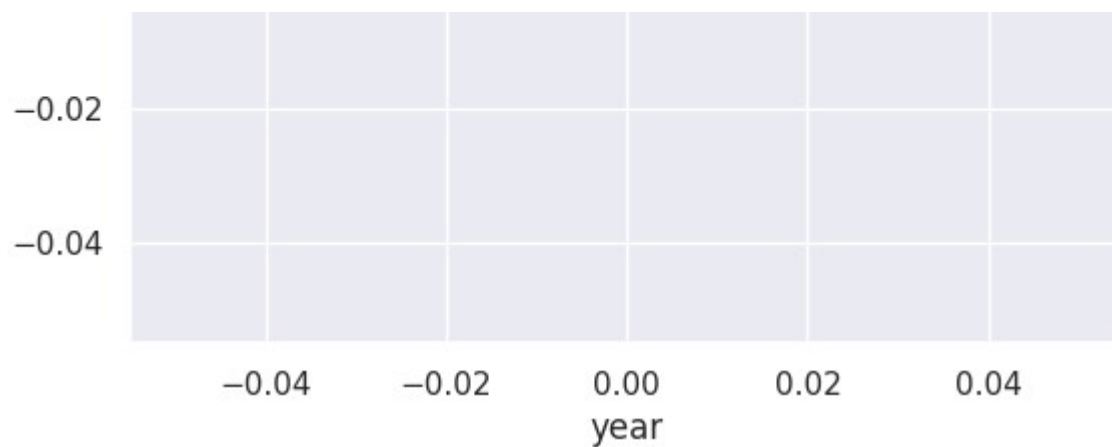
```
Science Teacher          8
Science Student          5
Science Kid              1
Science Monk             1
Science Officer 3        1
Science Performer 3      1
Science Promo Cadet      1
Science Fair Patron      1
Science Fair Judge #2    1
Science Fair Parent      1
Name: character, dtype: int64
```

Plot the n-values of the roles that Judi Dench has played over her career.

```
c=cast[cast.name=="Judi Dench"]
c.plot(kind="scatter",x='year',y='n',alpha=0.5)
```

```
<Axes: xlabel='year', ylabel='n'>
```

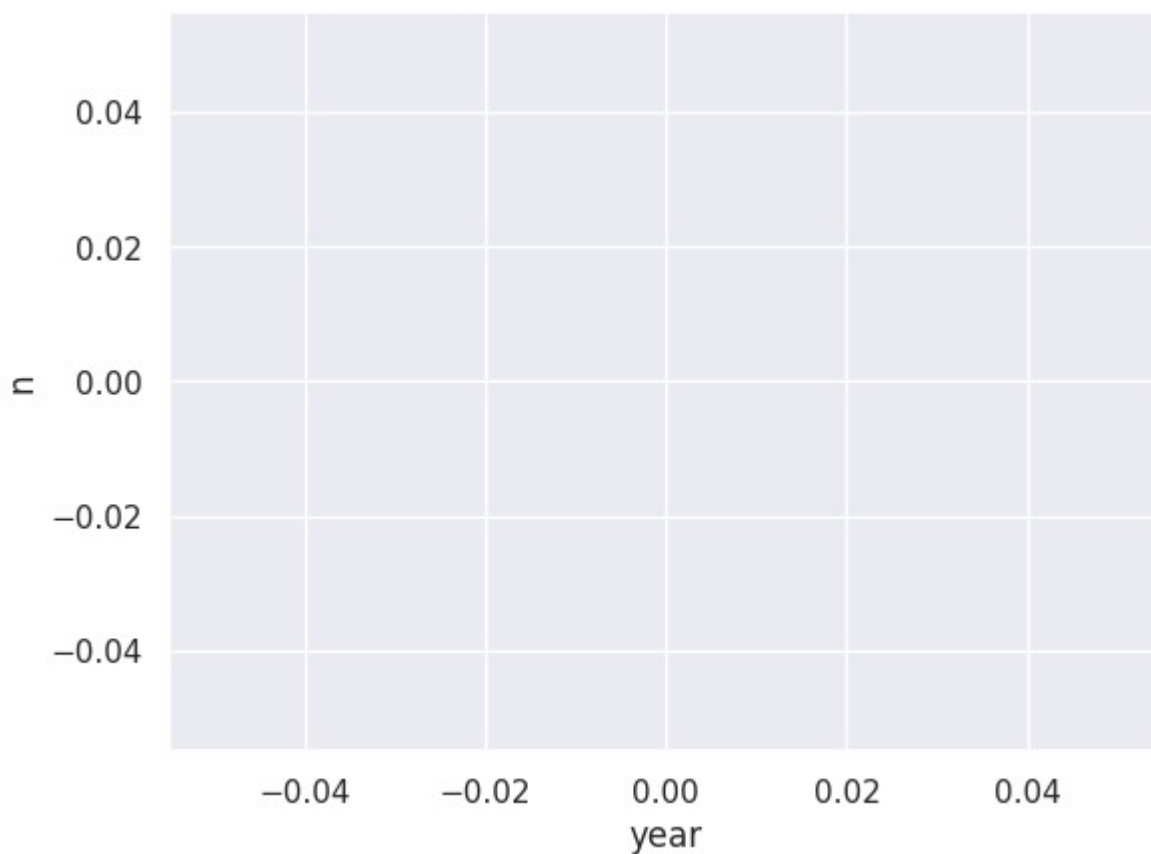




Plot the n-values of Cary Grant's roles through his career.

```
d=cast[cast.name=="Cary Grant"]
d.plot(kind="scatter",x='year',y='n',alpha=0.5)
```

<Axes: xlabel='year', ylabel='n'>

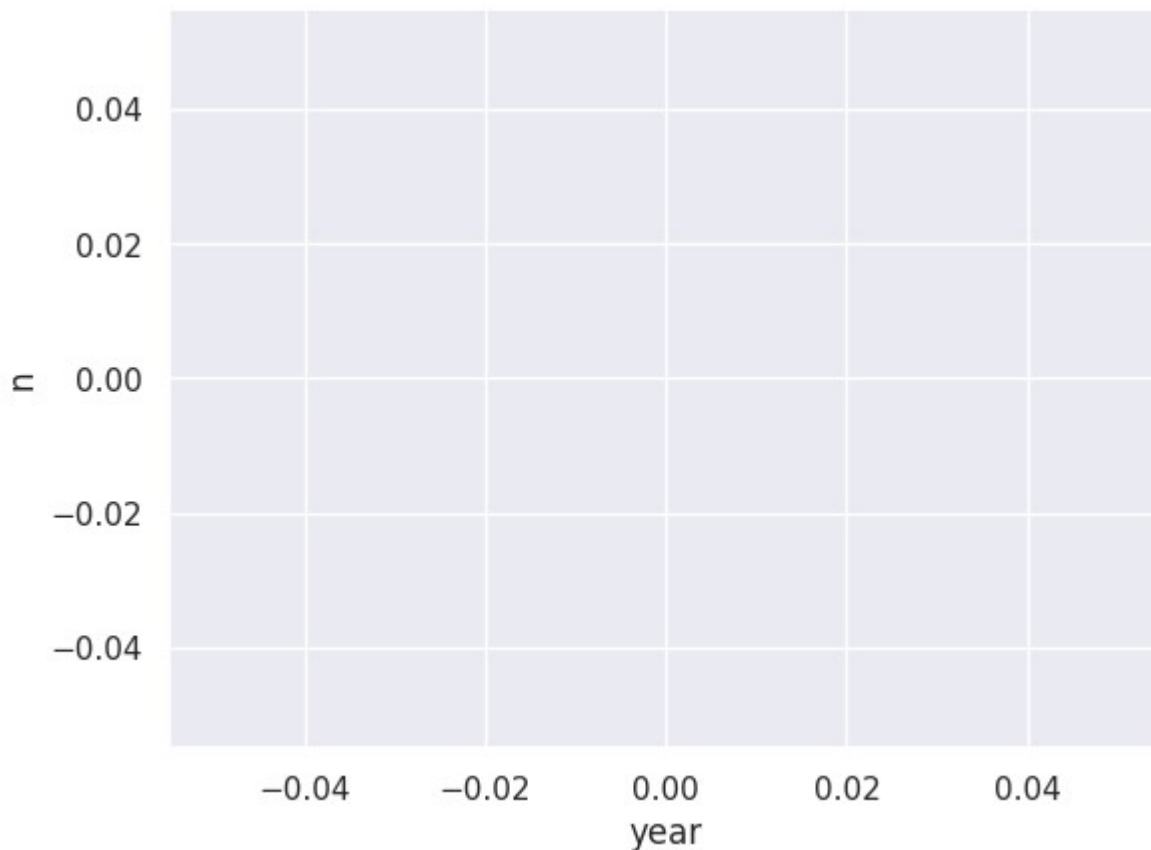


Plot the n-value of the roles that Sidney Poitier has acted over the years

over the years.

```
cast[cast.name=='Ron Jeremy'].plot(kind='scatter',x='year',y='n')
```

```
<Axes: xlabel='year', ylabel='n'>
```



How many leading (n=1) roles were available to actors, and how many to actresses, in the 1950s?

```
cast[(cast.year//10==195)&(cast.n==1)].type.value_counts()
```

```
actor      1108
Name: type, dtype: int64
```

How many supporting (n=2) roles were available to actors, and how many to actresses, in the 1950s?

```
cast[(cast.year//10==195)&(cast.n==2)].type.value_counts()
```

```
actor      778  
Name: type, dtype: int64
```