PROJECT: VISUALIZING THE HISTORY OF NOBEL PRIZE WINNERS

datalab

The Nobel Prize has been among the most prestigious international awards since 1901. Each year, awards are bestowed in chemistry, literature, physics, physiology or medicine, economics, and peace. In addition to the honor, prestige, and substantial prize money, the recipient also gets a gold medal with an image of Alfred Nobel (1833 - 1896), who established the prize.

The Nobel Foundation has made a dataset available of all prize winners from the outset of the awards from 1901 to 2023. The dataset used in this project is from the Nobel Prize API and is available in the `nobel.csv` file in the `data` folder.

In this project, you'll get a chance to explore and answer several questions related to this prizewinning data. And we encourage you then to explore further questions that you're interested in!

```
# Loading in required libraries
import pandas as pd
import seaborn as sns
import numpy as np

# Start coding here!
```

```
nobel=pd.read_csv('data/nobel.csv')
nobel.head()
```

	...	c...	prize	motivation	pri...	lau...
0	1901	Chemistry	The Nobel Prize in Chemistry 1901	"in recognition of the extraordinary services ...	1/1	16
1	1901	Literature	The Nobel Prize in Literature 1901	"in special recognition of his poetic composit...	1/1	56
2	1901	Medicine	The Nobel Prize in Physiology or Medicine 19...	"for his work on serum therapy, especially its...	1/1	29
3	1901	Peace	The Nobel Peace Prize 1901	null	1/2	46
4	1901	Peace	The Nobel Peace Prize 1901	null	1/2	46

Rows: 5 Expand

```
top_gender=nobel['sex'].value_counts().index[0]
top_gender
```

```
'Male'
```

```
top_country=nobel['birth_country'].value_counts().index[0]
top_country
```

```
'United States of America'
```

```
nobel['decade']=(np.floor(nobel['year']/10)*10).astype(int)
us_winners=nobel[nobel['birth_country']=='United States of America']

us_winners
```

		c...	prize	motivation	pri...	lau...
35	1906	Peace	The Nobel Peace Prize 1906	null	1/1	
72	1912	Peace	The Nobel Peace Prize 1912	null	1/1	
79	1914	Chemistry	The Nobel Prize in Chemistry 1914	"in recognition of his accurate determination...	1/1	
95	1919	Peace	The Nobel Peace Prize 1919	null	1/1	
117	1923	Physics	The Nobel Prize in Physics 1923	"for his work on the elementary charge of ele...	1/1	
124	1925	Peace	The Nobel Peace Prize 1925	null	1/2	
138	1927	Physics	The Nobel Prize in Physics 1927	"for his discovery of the effect named after h...	1/2	
149	1929	Peace	The Nobel Peace Prize 1929	null	1/1	
152	1930	Literature	The Nobel Prize in Literature 1930	"for his vigorous and graphic art of descripti...	1/1	
160	1931	Peace	The Nobel Peace Prize 1931	null	1/2	
161	1931	Peace	The Nobel Peace Prize 1931	null	1/2	
162	1932	Chemistry	The Nobel Prize in Chemistry 1932	"for his discoveries and investigations in surf...	1/1	
168	1933	Medicine	The Nobel Prize in Physiology or Medicine 19...	"for his discoveries concerning the role playe...	1/1	
172	1934	Chemistry	The Nobel Prize in Chemistry 1934	"for his discovery of heavy hydrogen"	1/1	
174	1934	Medicine	The Nobel Prize in Physiology or Medicine 19...	"for their discoveries concerning liver therap...	1/3	
175	1934	Medicine	The Nobel Prize in Physiology or Medicine 19...	"for their discoveries concerning liver therap...	1/3	

Rows: 291 Expand

```
us_winners['decade'].value_counts().index
```

```
Int64Index([2000, 1990, 2010, 1970, 1980, 1950, 1960, 2020, 1930, 1940, 1920,
            1910, 1900],
           dtype='int64')
```

```
max_decade_usa2=(us_winners['decade'].value_counts()/nobel['decade'].value_counts()).max()
max_decade_usa2
```

```
0.42276422764227645
```

```
us_decade=us_winners.groupby('decade', as_index=False)['decade'].value_counts()
us_decade
```

index	decade	count
0	1900	
1	1910	
2	1920	
3	1930	
4	1940	
5	1950	
6	1960	
7	1970	
8	1980	
9	1990	
10	2000	
11	2010	
12	2020	

Rows: 13 Expand

```
max_decade_usa1=(us_winners.groupby('decade', as_index=False)['decade'].value_counts()/nobel.groupby('decade', as_index=False)
['decade'].value_counts()) #.max()
max_decade_usa1['decade']=us_decade['decade']
max_decade_usa=max_decade_usa1[max_decade_usa1['count']==max_decade_usa1['count'].max()]['decade'].values[0]
max_decade_usa
```

```
2000
```

```
Male=nobel.groupby('decade')['sex'].value_counts().values[0]
Male
```

```
53
```

```
nobel['female']=nobel['sex']=='Female'
nobel['female']
```

0	False
1	False
2	False
3	False
4	False
5	False
6	False
7	False
8	False
9	False
10	False
11	False
12	False
13	False
14	False
15	False

Rows: 1,000 Expand

```
grp_decat=nobel.groupby(['decade', 'category'])
total_count=grp_decat.count()
fem_count=grp_decat['sex'].apply(lambda x: (x=='Female').sum())
total_count
```

	c…			m…	pri…	lau…	laurea…	f.	b…	b…
1900	Chemistry	9	9	9	9	9	9	9	9	9
1900	Literature	10	10	10	10	10	10	10	10	10
1900	Medicine	11	11	11	11	11	11	11	11	11
1900	Peace	14	14	0	14	14	14	14	13	13
1900	Physics	13	13	13	13	13	13	13	13	13
1910	Chemistry	8	8	8	8	8	8	8	8	8
1910	Literature	9	9	9	9	9	9	9	9	9
1910	Medicine	6	6	6	6	6	6	6	6	6
1910	Peace	7	7	0	7	7	7	7	5	5
1910	Physics	10	10	10	10	10	10	10	10	10
1920	Chemistry	10	10	10	10	10	10	10	10	10
1920	Literature	10	10	10	10	10	10	10	10	10
1920	Medicine	11	11	11	11	11	11	11	11	11
1920	Peace	11	11	0	11	11	11	11	11	11
1920	Physics	12	12	12	12	12	12	12	12	12
1930	Chemistry	13	13	13	13	13	13	13	13	13

Rows: 72 Expand

```
# Updated code
# prop = fem_count['Female'] / total_count['Female']
fem_count
```

	c…	
1900	Chemistry	0
1900	Literature	1
1900	Medicine	0
1900	Peace	1
1900	Physics	1
1910	Chemistry	1
1910	Literature	0
1910	Medicine	0
1910	Peace	0
1910	Physics	0
1920	Chemistry	0
1920	Literature	2
1920	Medicine	0
1920	Peace	0
1920	Physics	0
1930	Chemistry	1

Rows: 72 Expand

```
grouped=nobel.groupby(['decade', 'category'], as_index=False)['female'].mean()
max_prop=grouped[grouped['female']==grouped['female'].max()]
max_prop
```

		c…	
68	2020	Literature	0.5

Rows: 1 Expand

```
max_female_dict = {max_prop['decade'].iloc[0]: max_prop['category'].iloc[0]}
max_female_dict
```

```
{2020: 'Literature'}
```

```
first_woman=[nobel[nobel['sex']=='Female'].values[0]]
first_woman=pd.DataFrame(first_woman, columns=nobel.columns)
first_woman_name=first_woman['full_name'][0]
first_woman_name
first_woman_category=first_woman['category'][0]
first_woman_category
```

```
'Physics'
```

```
nobel.columns
```

```
Index(['year', 'category', 'prize', 'motivation', 'prize_share', 'laureate_id',
       'laureate_type', 'full_name', 'birth_date', 'birth_city',
       'birth_country', 'sex', 'organization_name', 'organization_city',
       'organization_country', 'death_date', 'death_city', 'death_country',
       'decade', 'female'],
      dtype='object')
```

```
repeat_list = []
repeat_names = nobel['full_name'].value_counts()
repeat_names = repeat_names[repeat_names > 1].index.tolist()
repeat_list.extend(repeat_names)
repeat_list
```

```
['Comité international de la Croix Rouge (International Committee of the Red Cross)',
 'Linus Carl Pauling',
 'John Bardeen',
 'Frederick Sanger',
 'Marie Curie, née Sklodowska',
 'Office of the United Nations High Commissioner for Refugees (UNHCR)']
```