

PROJECT: VISUALIZING THE HISTORY OF NOBEL PRIZE WINNERS



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
import matplotlib.pyplot as plt

#importing nobel.csv into python
nobel = pd.read_csv("data/nobel.csv")
nobel.head()

#What is the most commonly awarded gender and birth country?

top_gender = nobel.value_counts("sex", ascending=False).index[0]
plt.figure(1)
g = sns.countplot(data=nobel, x="sex")
plt.title("Count of Nobel Prize winners by gender")
g.bar_label(g.containers[0])
print("top_gender: " + str(top_gender))

top_country = nobel.value_counts("organization_country", ascending=False).index[0]
plt.figure(2)
c = sns.countplot(data=nobel, x="organization_country")
plt.xticks(rotation=90)
plt.title("Count of Nobel Prize winners by country")
c.bar_label(c.containers[0])
print("top_country: " + str(top_country))

#note: counting organization_country instead of birth_country or death_country due
to the award being given to the laureate(s) based on the country where they
submitted their work

#Which decade had the highest ratio of US-born Nobel Prize winners to total winners
in all categories?

nobel["USA_winners"] = nobel["birth_country"] == "United States of America"
nobel["decade"] = (np.floor(nobel["year"]/10)*10).astype(int)
US_born_ratio = nobel.groupby("decade", as_index=False)["USA_winners"].mean()
max_decade_usa = US_born_ratio.loc[US_born_ratio["USA_winners"].idxmax(), "decade"]
plt.figure(3)
sns.relplot(x="decade", y="USA_winners", kind="line", data=US_born_ratio, ci=None)
plt.title("Ratio of US born Nobel Prize winners to total Nobel Prize winners by
decade")
print("max_decade_usa: " + str(max_decade_usa))

#Which decade and Nobel Prize category combination had the highest proportion of
female laureates?
```

```
nobel["female_winner"] = nobel["sex"] == "Female"
female_win_ratio = nobel.groupby(["decade", "category"], as_index=False)
["female_winner"].mean()
max_female_win_ratio =
female_win_ratio.loc[female_win_ratio["female_winner"].idxmax()]
highest_prop_female_dict = {max_female_win_ratio["decade"]:
max_female_win_ratio["category"]}
plt.figure(4)
sns.relplot(x="decade", y="female_winner", hue="category", kind="line",
data=female_win_ratio, ci=None)
plt.title("Female Nobel Prize winners by decade and category")
print("highest_prop_female_winners: " + str(highest_prop_female_dict))
```

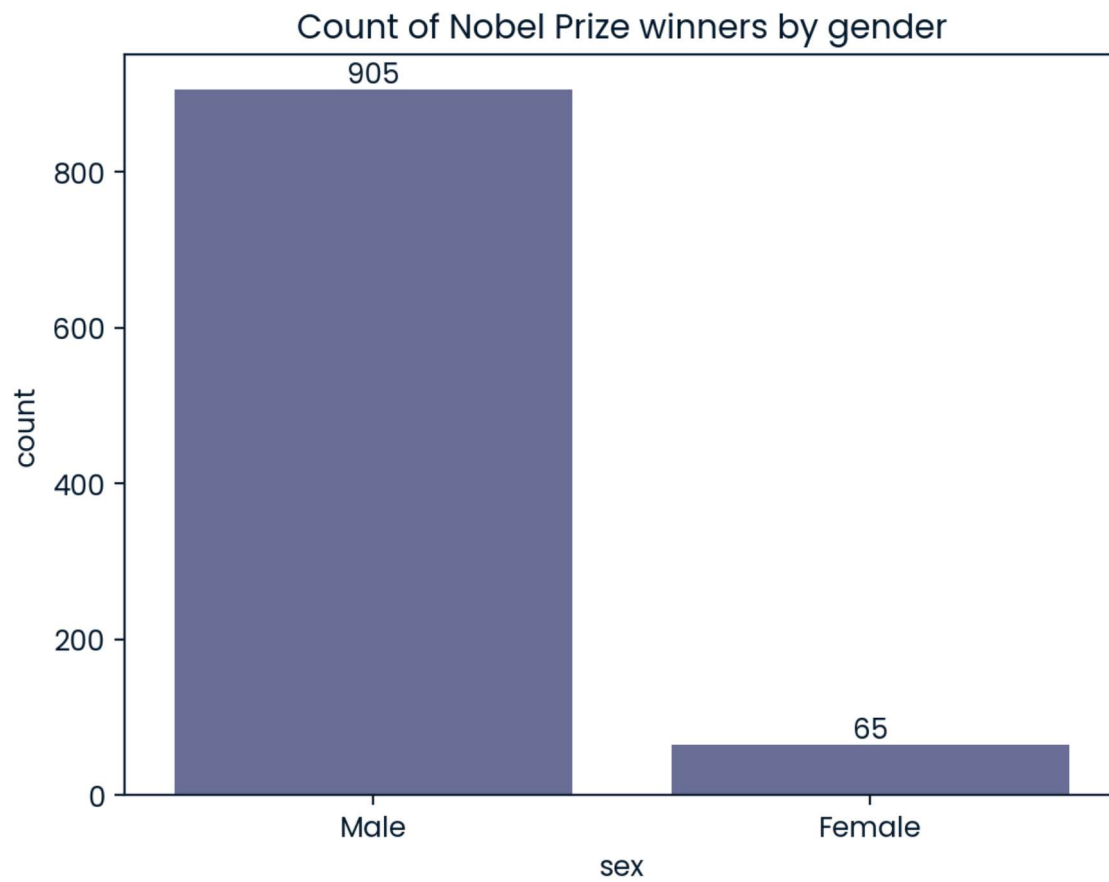
#Who was the first woman to receive a Nobel Prize, and in what category?

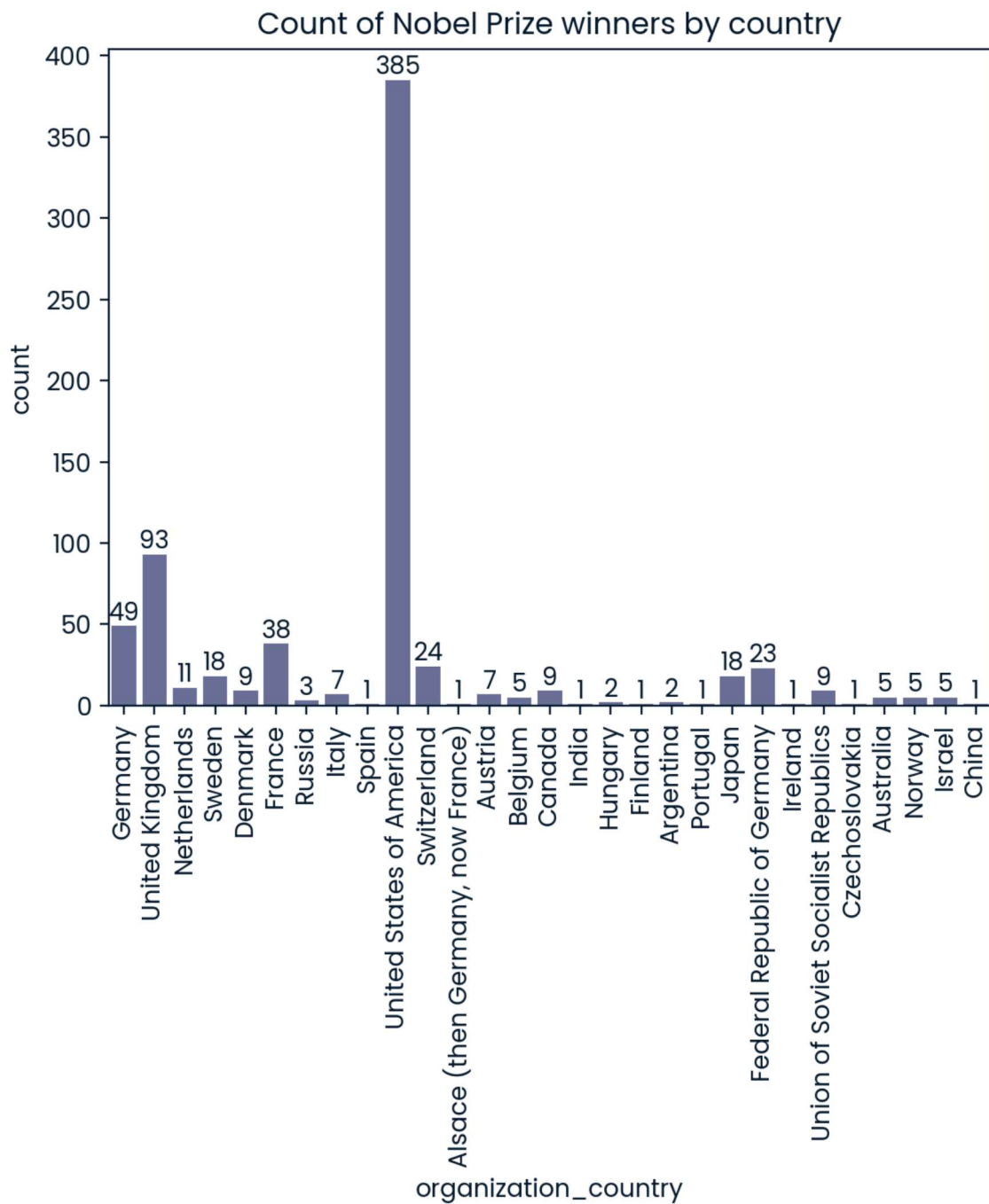
```
female_win_df = nobel[nobel["female_winner"] == True]
first_woman = female_win_df.loc[female_win_df["year"].idxmin()]
first_woman_name = first_woman["full_name"]
first_woman_category = first_woman["category"]
print("first_woman_name_and_category" + ": " + first_woman_name + "; " +
first_woman_category)
```

#Which individuals or organizations have won more than one Nobel Prize throughout the years?

```
name_counts = pd.Series(nobel["full_name"].value_counts())
name_awards = name_counts[name_counts >= 2].index
repeat_list = list(name_awards)
print("repeat_winners: ", repeat_list)
```

```
top_gender: Male
top_country: United States of America
max_decade_usa: 2000
highest_prop_female_winners: {2020: 'Literature'}
first_woman_name_and_category: Marie Curie, née Skłodowska; Physics
repeat_winners: ['Comité international de la Croix Rouge (International Committee
of the Red Cross)', 'Linus Carl Pauling', 'John Bardeen', 'Frederick Sanger', 'Marie
Curie, née Skłodowska', 'Office of the United Nations High Commissioner for Refugees
(UNHCR)']
```





<Figure size 640x480 with 0 Axes>

Ratio of US born Nobel Prize winners to total Nobel Prize winners by decade

