

# Visualizing the history of Nobel prize winners

---

Workplace education 2023 – 2024  
presenter: **Salma Rahmani**



# Project description

The Nobel Prize has been among the most prestigious international awards since 1901. Prizes are awarded annually in the fields of chemistry, literature, physics, physiology or medicine, economics and peace. In addition to honour, prestige and a significant amount of prize money, the award recipient also receives a gold medal bearing the image of Alfred Nobel (1833 - 1896), who founded the award. The Nobel Foundation has made available a dataset of all laureates from the awards' inception, 1901 to 2023. The dataset used in this project was taken from the Nobel Prize API.

Analyze the data of Nobel Prize winners in detail, using the information you have learned until you come to this data set (keep the analyzes in your codebook and you can talk about the important ones in the video you will shoot). Prepare a Power Point presentation containing your analysis and the visuals you created and answer the questions below. In addition to the questions, it is entirely up to your creativity to extract meaningful information from the data and support it with various visuals.





# Analysis of project

---

- The questions I answer in this project are:
- -Who won more Nobel Prizes, men or women?
- -What country has won more Nobel Prizes?
- -What is the gender of a typical Nobel Prize winner?
- -Who was the first woman to win a Nobel Prize?
- -Who have won more than one Nobel Prize?

---

# Project output:

---

The gender and birth country with the most Nobel laureates.

---

Analysis and visualization of the proportion of USA-born winners over decades.

---

Analysis and visualization of the proportion of female laureates per decade and category.

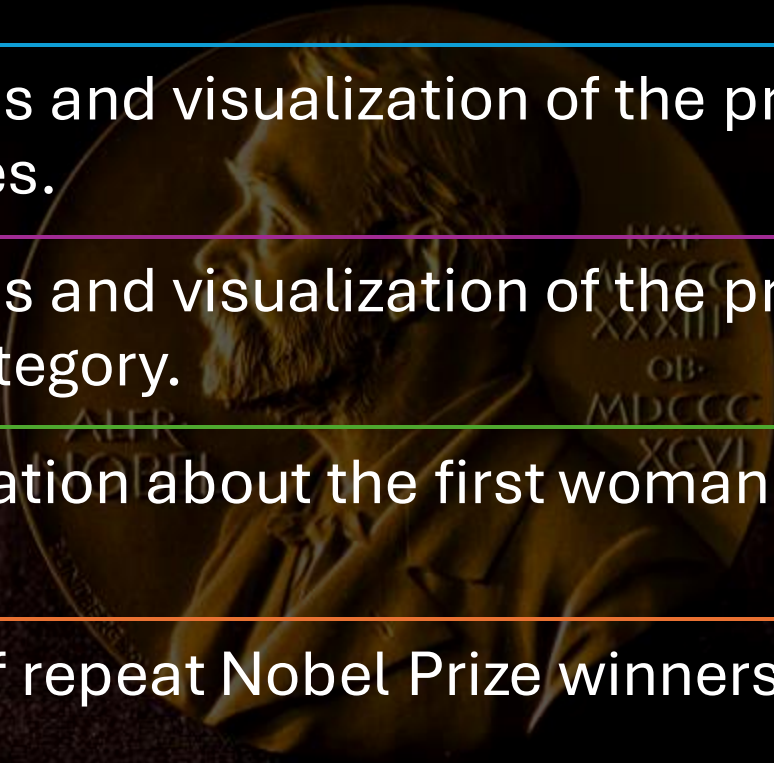
---

Information about the first woman to win a Nobel Prize.

---

A list of repeat Nobel Prize winners.

---





## Project steps:

---

- **Loading Libraries and Dataset**
- **Exploratory Data Analysis (EDA)**
- **USA-Born Winners Analysis**
- **Female Laureates Analysis**
- **First Woman Nobel Prize Winner**
- **Repeat Nobel Prize Winners**



# Question 1:

---

- 1. What is the most awarded gender and birth\_country? Save your answers as top\_gender and top\_country.
  - Answer:
  - This section of the code determines the predominant gender and country of birth among Nobel laureates by using the value\_counts() function to count occurrences. It then selects the gender and birth country with the highest count. The code demonstrates a streamlined analysis of the Nobel Prize dataset, extracting and presenting the
  - most common gender and birth country through print statements.
- 

```
⋮  # Start coding here!  
3  # Loading in required libraries  
4  import pandas as pd  
5  import seaborn as sns  
6  import numpy as np  
7
```

```
12 top_gender = nobel['sex'].value_counts().index[0]  
13 top_country = nobel['birth_country'].value_counts().index[0]  
14  
15 print("\n The gender with the most Nobel laureates is :", top_gender)  
16 print(" The most common birth country of Nobel laureates is :", top_country)  
17
```

## 1. OUTPUT:

- Answer:
- The most frequently awarded gender is **male** and the birth country is the **United States**.

The gender with the most Nobel laureates is : Male

The most common birth country of Nobel laureates is : United States of America



## Question 2:

- **Rearrange the data by decade (1900-1910-1920 ...etc.). Which year has the highest rate of US-born winners? Store this as an integer called max\_decade\_usa.**
  - Answer:
- This portion pinpoints the decade with the highest percentage of Nobel laureates born in the United States, storing this information in the variable max\_decade\_usa. The code facilitates the analysis of trends in the nationality of laureates by creating a Boolean column, organizing birth years into decades, and computing the proportion of laureates born in the United States for each decade.

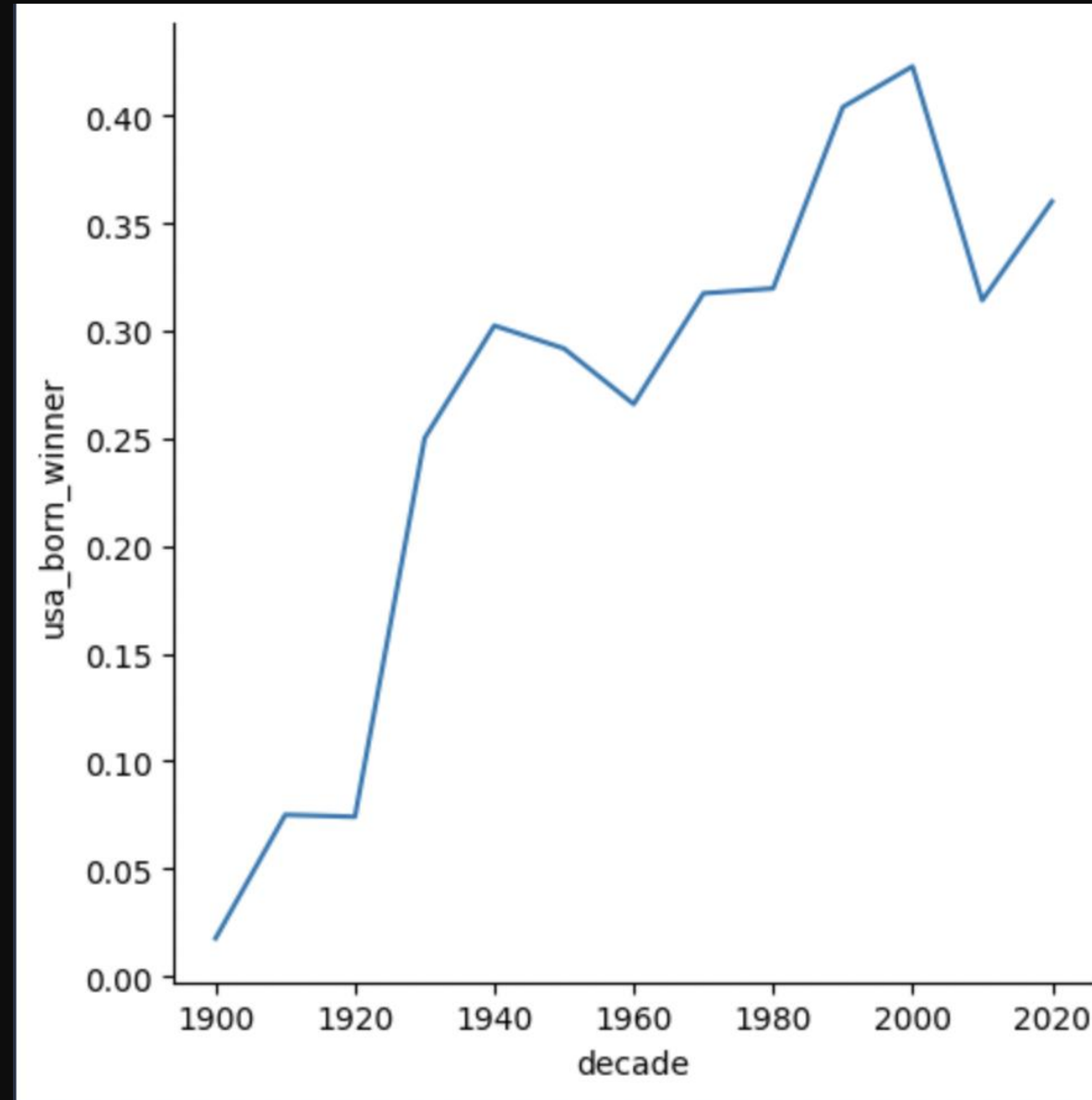
```
17
18 # Calculate the proportion of USA born winners per decade
19 nobel['usa_born_winner'] = nobel['birth_country'] == 'United States of America'
20 nobel['decade'] = (np.floor(nobel['year'] / 10) * 10).astype(int)
21 prop_usa_winners = nobel.groupby('decade', as_index=False)['usa_born_winner'].mean()
22
23 # Identify the decade with the highest proportion of US-born winners
24 max_decade_usa = prop_usa_winners[prop_usa_winners['usa_born_winner'] == prop_usa_winners['usa_born_winner'].max()]
25   ['decade'].values[0]
26
27 # Optional: Plotting USA born winners
28 ax1 = sns.relplot(x='decade', y='usa_born_winner', data=prop_usa_winners, kind="line")
```



## 2. OUTPUT:

---

- Answer:
  - This graph displays the percentage of Nobel Prizes awarded by decade, with the USA consistently showing improvement, particularly in the 2000s.
- 



# Question 3:

---

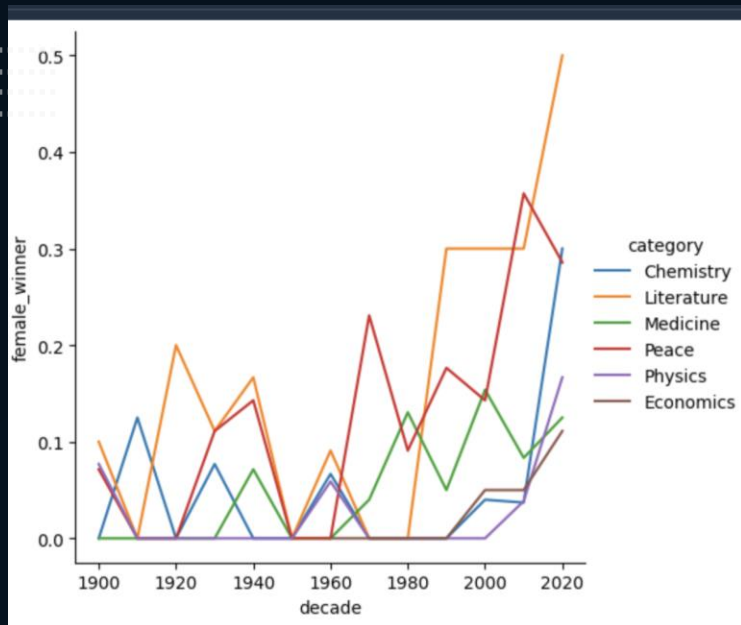
- Rearrange the data by decade (1900-1910-1920 ...etc.). Which decade and category pair had the highest proportion of female award winners? Store this as a dictionary called `max_female_dict`, where decade is key, and category is value.
  - Answer:
  - This lines of codes calculates and visualizes the proportion of female Nobel laureates per decade and category. It identifies the decade and category with the highest proportion of female laureates, storing the result. The information is organized into a dictionary. Additionally, an optional Seaborn line plot shows trends in female winners, using the y-axis to represent the percentage of female laureates.
- 

```
# Calculating the proportion of female laureates per decade
nobel['female_winner'] = nobel['sex'] == 'Female'
prop_female_winners = nobel.groupby(['decade', 'category'], as_index=False)['female_winner'].mean()

# Find the decade and category with the highest proportion of female laureates
max_female_decade_category = prop_female_winners[prop_female_winners['female_winner'] == prop_female_winners['female_winner'].max()]
[['decade', 'category']]

# Create a dictionary with the decade and category pair
max_female_dict = {max_female_decade_category['decade']: max_female_decade_category['category'].values[0]}

# Optional: Plotting female winners with % winners on the y-axis
ax2 = sns.relplot(x='decade', y='female_winner', hue='category', data=prop_female_winners, kind="line")
```



### 3. OUTPUT:

Answer:

- This chart makes it easy to spot differences in Nobel laureates, showing times when women have won and areas dominated by men. It's a simple way to see the variations in winning frequencies between genders.

## Question 4:

- Who is the first woman to receive a Nobel Prize and in what category did she receive this award? Save your answers as `first_woman_name` and `first_woman_category`.

- Answer:
- To provide the It begins by filtering the original dataset, 'nobel,' to create a new DataFrame, 'nobel\_women,' containing only female laureates. Subsequently, it narrows down the selection to the earliest year of female Nobel Prize recipients, creating a DataFrame, 'min\_row.' The code then extracts the name and category of the first woman laureate from 'min\_row' and prints this information in a formatted statement. Overall, this code systematically searches and retrieves details about the pioneering woman who was the first to achieve the prestigious honor of a Nobel Prize.

```
# Finding the first woman to win a Nobel Prize
nobel_women = nobel[nobel['female_winner']]
min_row = nobel_women[nobel_women['year'] == nobel_women['year'].min()]
first_woman_name = min_row['full_name'].values[0]
first_woman_category = min_row['category'].values[0]
print(f"\n The first woman to win a Nobel Prize was {first_woman_name}, in the category of {first_woman_category}.")
```

```

nobel['decade'] = np.floor(nobel['year'] / 10) * 10.astype(int)
prop_usa_winners = nobel.groupby('decade', as_index=False)['usa_born_winner'].mean()

# Identify the decade with the highest proportion of US-born winners
max_decade_usa = prop_usa_winners[prop_usa_winners['usa_born_winner'] == prop_usa_winners['usa_born_winner'].max()]['decade'].values[0]

# Optional: Plotting USA born winners
ax1 = sns.relplot(x='decade', y='usa_born_winner', data=prop_usa_winners, kind="line")

# Calculating the proportion of female laureates per decade
nobel['female_winner'] = nobel['sex'] == 'Female'
prop_female_winners = nobel.groupby(['decade', 'category'], as_index=False)['female_winner'].mean()

# Find the decade and category with the highest proportion of female laureates
max_female_decade_category = prop_female_winners[prop_female_winners['female_winner'] == prop_female_winners['female_winner'].max()]['decade', 'category']

# Create a dictionary with the decade and category pair
max_female_dict = {max_female_decade_category['decade'].values[0]: max_female_decade_category['category'].values[0]}

# Optional: Plotting female winners with % winners on the y-axis
ax2 = sns.relplot(x='decade', y='female_winner', hue='category', data=prop_female_winners, kind="line")

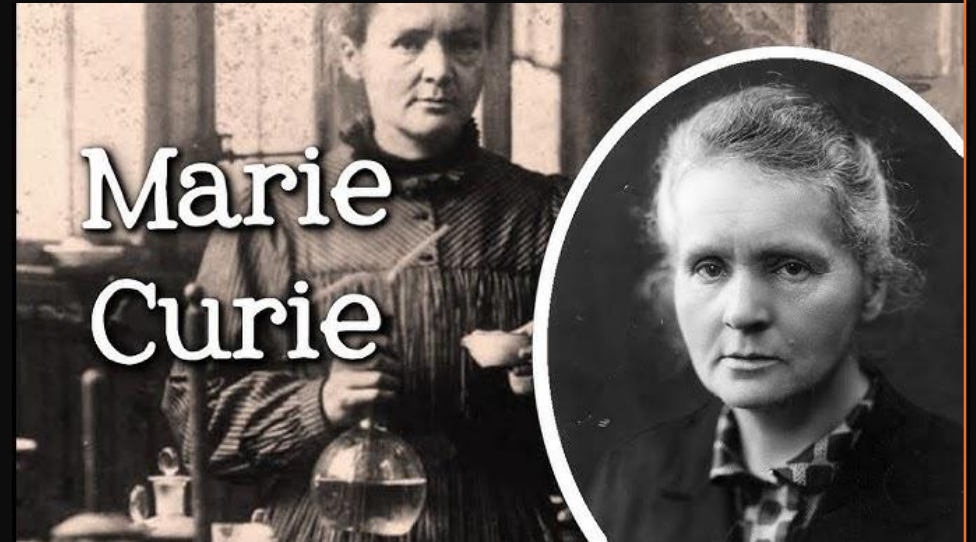
# Finding the first woman to win a Nobel Prize
nobel_women = nobel[nobel['female_winner']]
min_row = nobel_women[nobel_women['year'] == nobel_women['year'].min()]
first_woman_name = min_row['full_name'].values[0]
first_woman_category = min_row['category'].values[0]
print(f"\n The first woman to win a Nobel Prize was {first_woman_name}, in the category of {first_woman_category}.")

# Selecting the laureates that have received 2 or more prizes
counts = nobel['full_name'].value_counts()
repeats = counts[counts >= 2].index
repeat_list = list(repeats)

print("\n The repeat winners are :", repeat_list)
```

## 4. OUTPUT:

- Answer:
- The first woman to win a Nobel Prize was Marie Curie, in the category of Physics.



The first woman to win a Nobel Prize was Marie Curie, née Skłodowska, in the category of Physics.

## Question 5:

- Which individuals or organizations have won more than one Nobel Prize over the years? Store the full names in a list called `repeat_list`.
- Answer:
- This code snippet effectively identifies laureates who have received two or more Nobel Prizes. It achieves this by utilizing the `value_counts()` function to count the occurrences of each laureate's name in the dataset. Subsequently, it selects the laureates with counts equal to or exceeding 2 and prints the list of repeat winners. The result provides a concise overview of Nobel laureates who have been recognized multiple times.

```
# Selecting the laureates that have received 2 or more prizes
counts = nobel['full_name'].value_counts()
repeats = counts[counts >= 2].index
repeat_list = list(repeats)

print("\n The repeat winners are :", repeat_list)
```





## 5. OUTPUT:

---

- Answer:
- **Only six persons has won more than one Nobel Prize**

**NOBEL PRIZE  
WINNER  
LIST**



The repeat winners are : ['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)']

A cartoon illustration of a man in a tuxedo and bow tie, sitting on a globe. The globe features a portrait of Abraham Lincoln. The man is holding a red book. The background is a solid grey color.

# Thanks for your attention

---

You can access to the data and the code in my DataCamp profile.