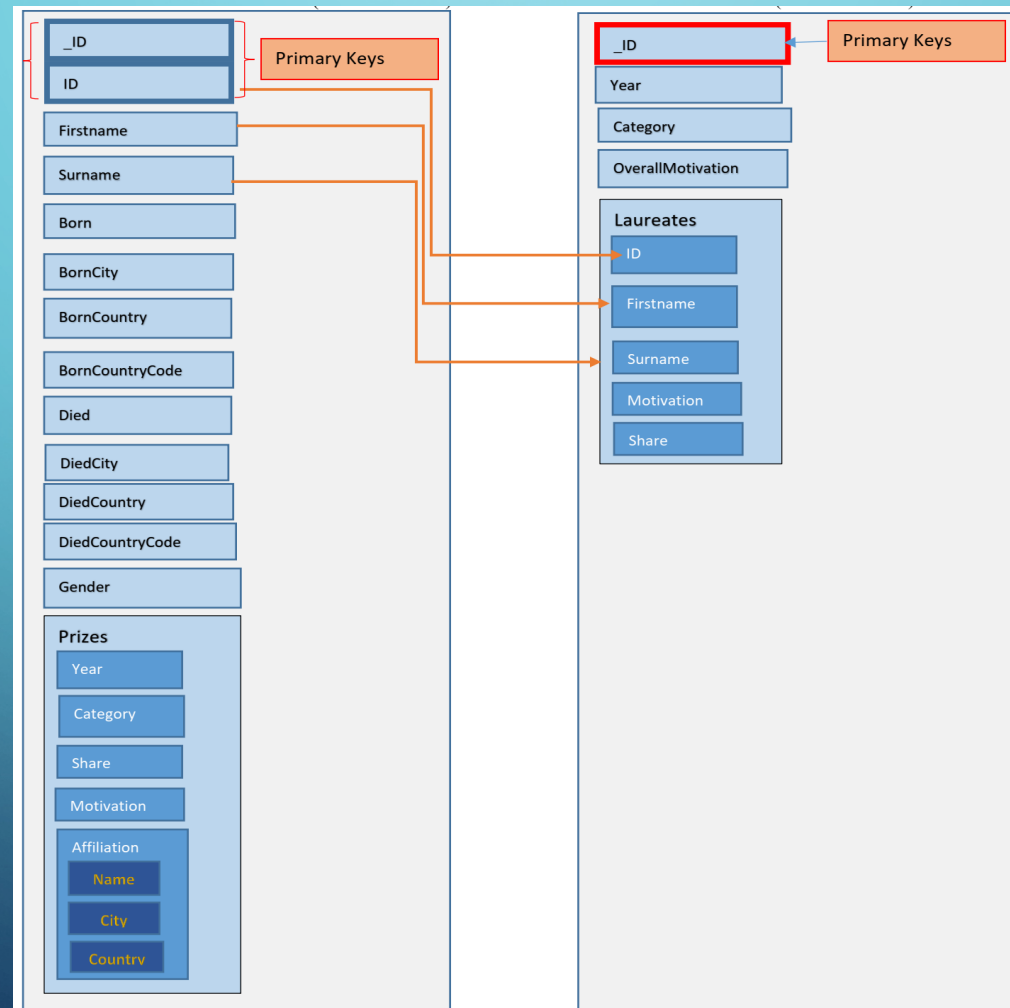# *Exploring the Nobel prize data from 1901 to 2018*

OVERVIEW: The data contains information about who has been awarded the Nobel Prize, when, in what prize category and the motivation, as well as basic information about the Nobel Laureates such as birth data and the affiliation at the time of the award.

I. Database model:

"Nobel" database
contains 2 collections:
Laureates and Prizes

"Laureates" collection (934 documents)

"Prizes" collection (590 documents)

_ID
ID

Primary Keys

Firstname
Surname
Born
BornCity
BornCountry
BornCountryCode
Died
DiedCity
DiedCountry
DiedCountryCode
Gender

Prizes
- Year
- Category
- Share
- Motivation
- Affiliation
  - Name
  - City
  - Country

_ID
Year
Category
OverallMotivation

Primary Keys

Laureates
- ID
- Firstname
- Surname
- Motivation
- Share

II. Data Analysis Intuition:

Categorical Data:

Prizes Collection:

```
_id: ObjectId('63d92672b9986af2fac4d33c')
year: "2018"
category: "physics"
overallMotivation: ""for groundbreaking inventions in the field of laser physics""
laureates: Array
  ▼ 0: Object
      id: "960"
      firstname: "Arthur"
      surname: "Ashkin"
      motivation: ""for the optical tweezers and their application to biological systems""
      share: "2"
```

Laureates Collection:

```
_id: ObjectId('63d9265cb9986af2fac4cf95')
id: "2"
firstname: "Hendrik Antoon"
surname: "Lorentz"
born: "1853-07-18"
died: "1928-02-04"
bornCountry: "the Netherlands"
bornCountryCode: "NL"
bornCity: "Arnhem"
diedCountry: "the Netherlands"
diedCountryCode: "NL"
gender: "male"
prizes: Array
  ▼ 0: Object
      year: "1902"
      category: "physics"
      share: "2"
      motivation: ""in recognition of the extraordinary service they rendered by their re_"
    ▼ affiliations: Array
      ▼ 0: Object
          name: "Leiden University"
          city: "Leiden"
          country: "the Netherlands"
```

Quantitative Data: Only the number of shares of each prize is quantitative data in both collections

Prizes Collection:

```
_id: ObjectId('63d92672b9986af2fac4d33c')
year: "2018"
category: "physics"
overallMotivation: ""for groundbreaking inventions in the field of laser physics""
laureates: Array
  0: Object
    id: "960"
    firstname: "Arthur"
    surname: "Ashkin"
    motivation: ""for the optical tweezers and their application to biological systems""
    share: "2"
```

Laureates Collection:

```
_id: ObjectId('63d9265cb9986af2fac4cf95')
id: "2"
firstname: "Hendrik Antoon"
surname: "Lorentz"
born: "1853-07-18"
died: "1928-02-04"
bornCountry: "the Netherlands"
bornCountryCode: "NL"
bornCity: "Arnhem"
diedCountry: "the Netherlands"
diedCountryCode: "NL"
gender: "male"
prizes: Array
  0: Object
    year: "1902"
    category: "physics"
    share: "2"
    motivation: ""in recognition of the extraordinary service they rendered by their re…"
    affiliations: Array
```

MongoDB

III. Exploration of data:

1. "Prizes" :

Category: There are 6 types of Categories prizes and it is shared quite evenly among them, except economics area.

Row Labels	Count of category
chemistry	110
economics	50
Literature	110
Medicine	109
Peace	99
Physics	112
Grand Total	**590**

Year: The period of this data is from 1901 to 2018.

2. "Laureates" :

Born: The data have laureates born from 1817 to 1997, in which some document does not have date of birth.

BornCity: There are 614 different born cities in total.

BornCountry: There are 120 different born countries in total, including some countries have changed names during the reported period.

Died: The list includes laureates died during the period of 1903-2018.

DiedCity: There are 296 different died cities in total.

Gender: There are 3 types of genders, including "male" & "female" for individual laureates & "org" for organisations laureates

Prizes: there are 858 prizes in total affiliated for the list of laureates

IV. Data Visualization and Conclusion:

1. Distribution of prizes
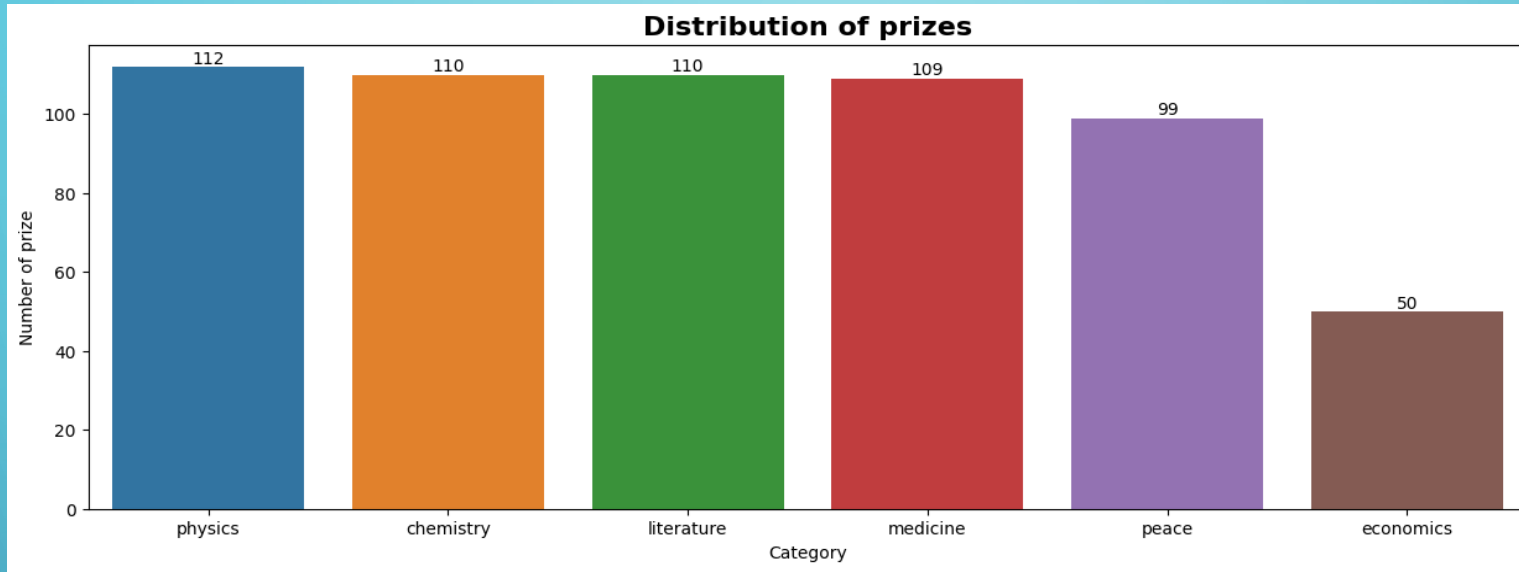
Economics is the field with the lowest number of prizes, which has half as many awards as other fields.
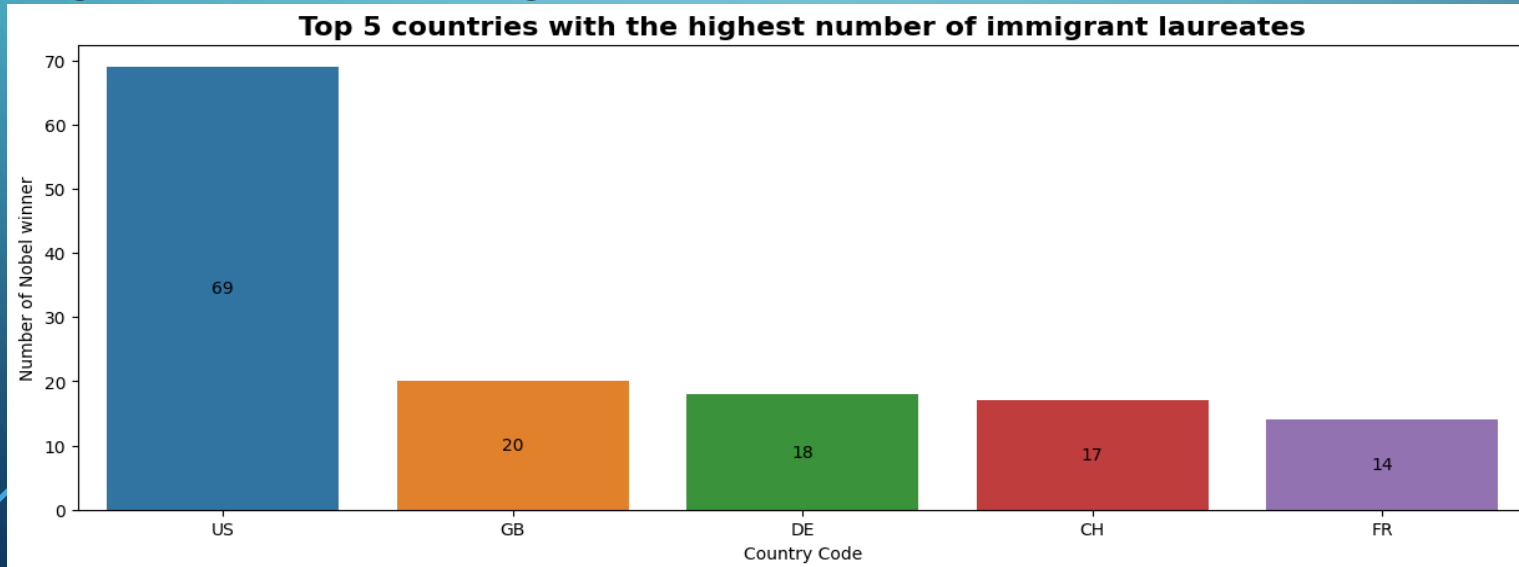
2. Most Nobel Winner Countries
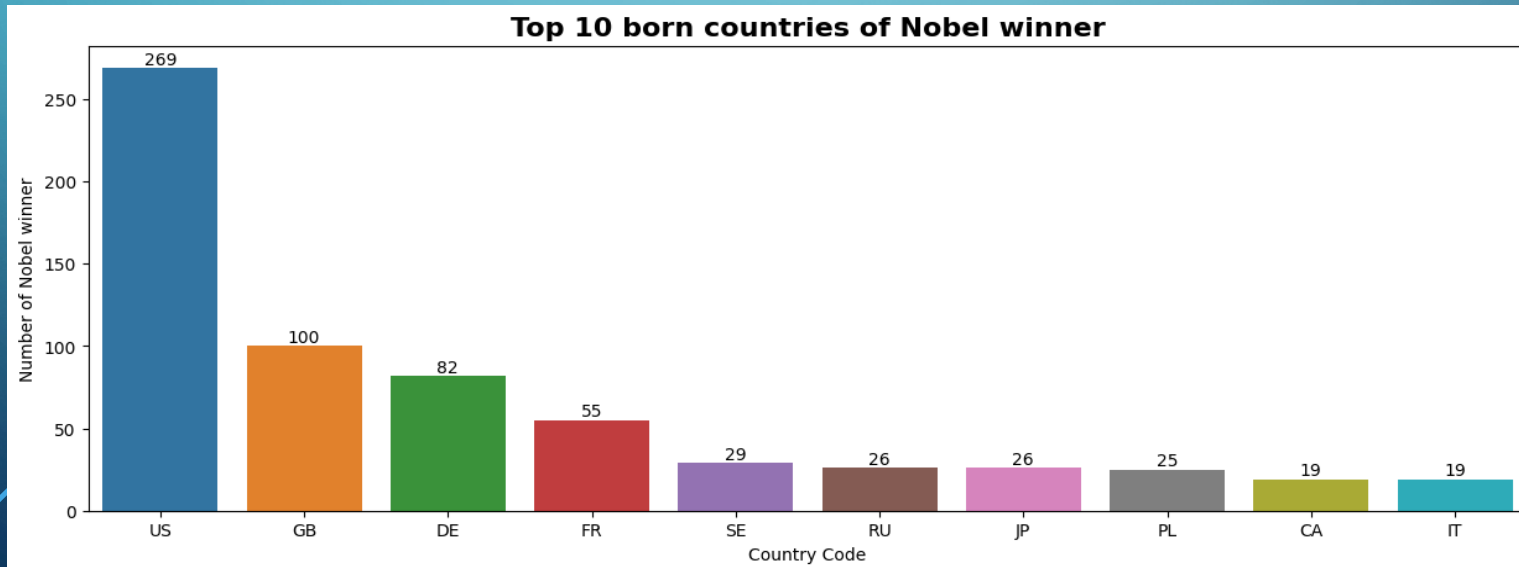
USA, Great Britain, Deutschland (Germany) had the most laureates in these years with remarkable distance

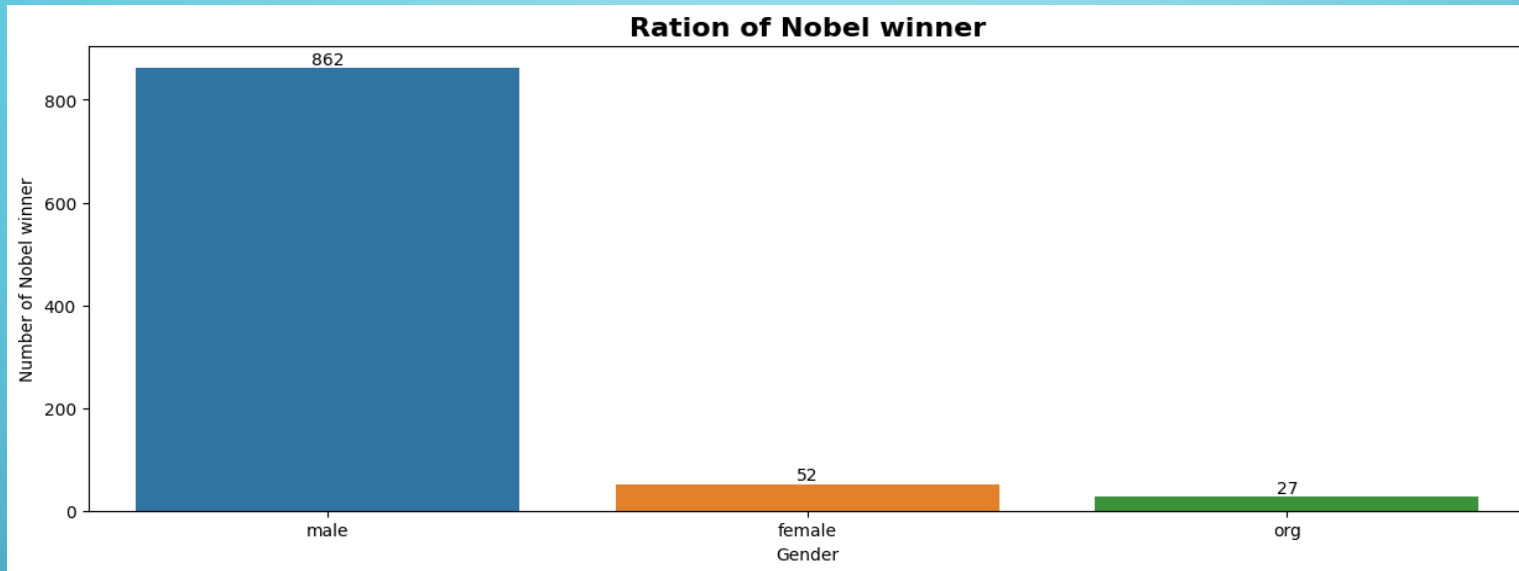# 3. Born countries with the highest number of laureates
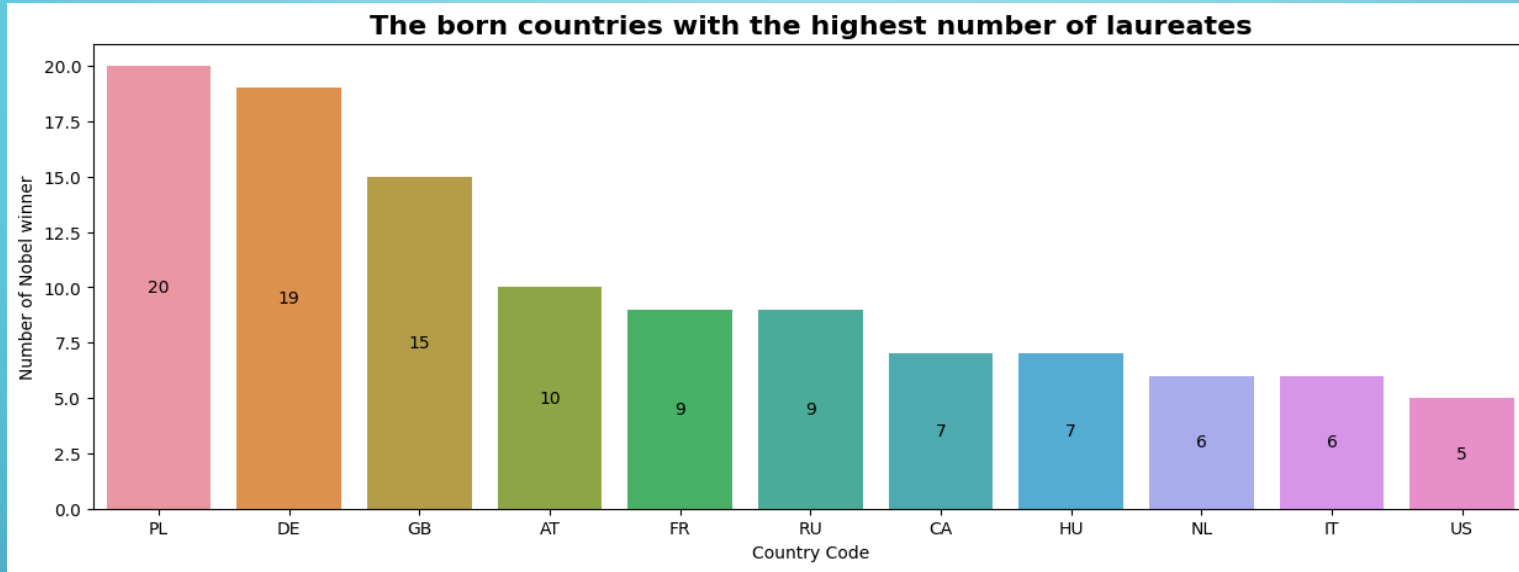
Poland, Deutschland (Germany), and Great Britain had the most laureates that won Nobel prizes.

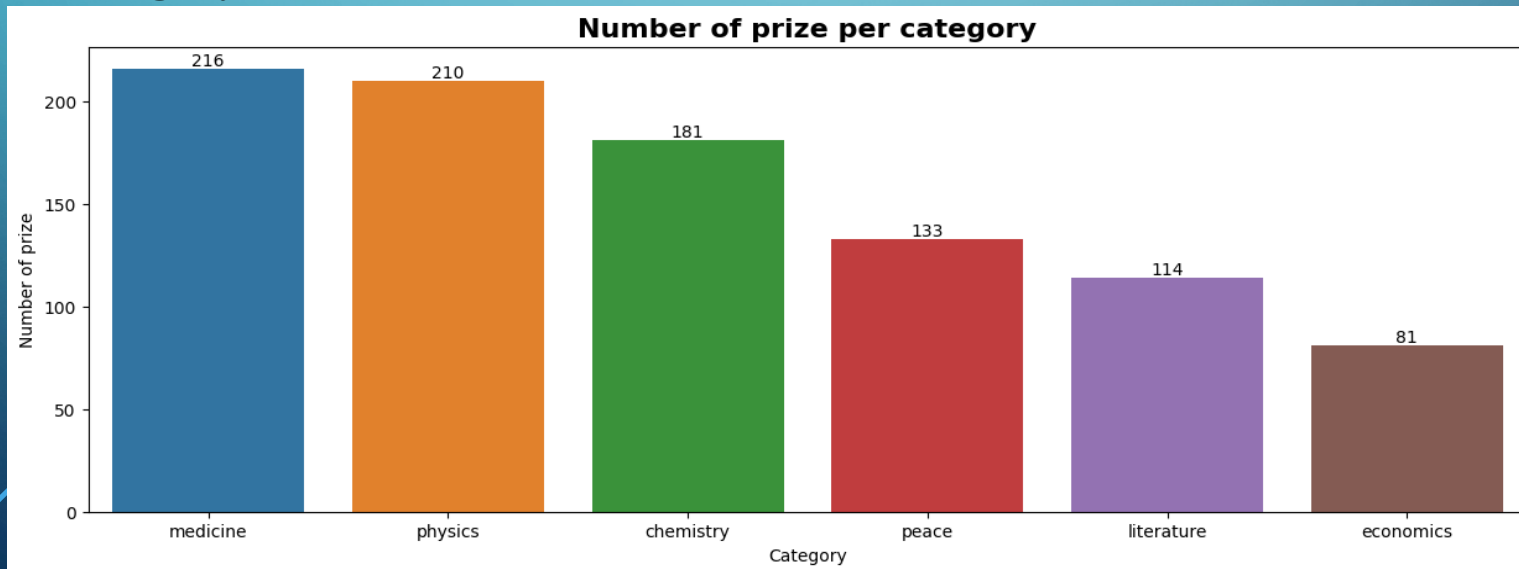# 4. Countries with the highest number of immigrant laureates

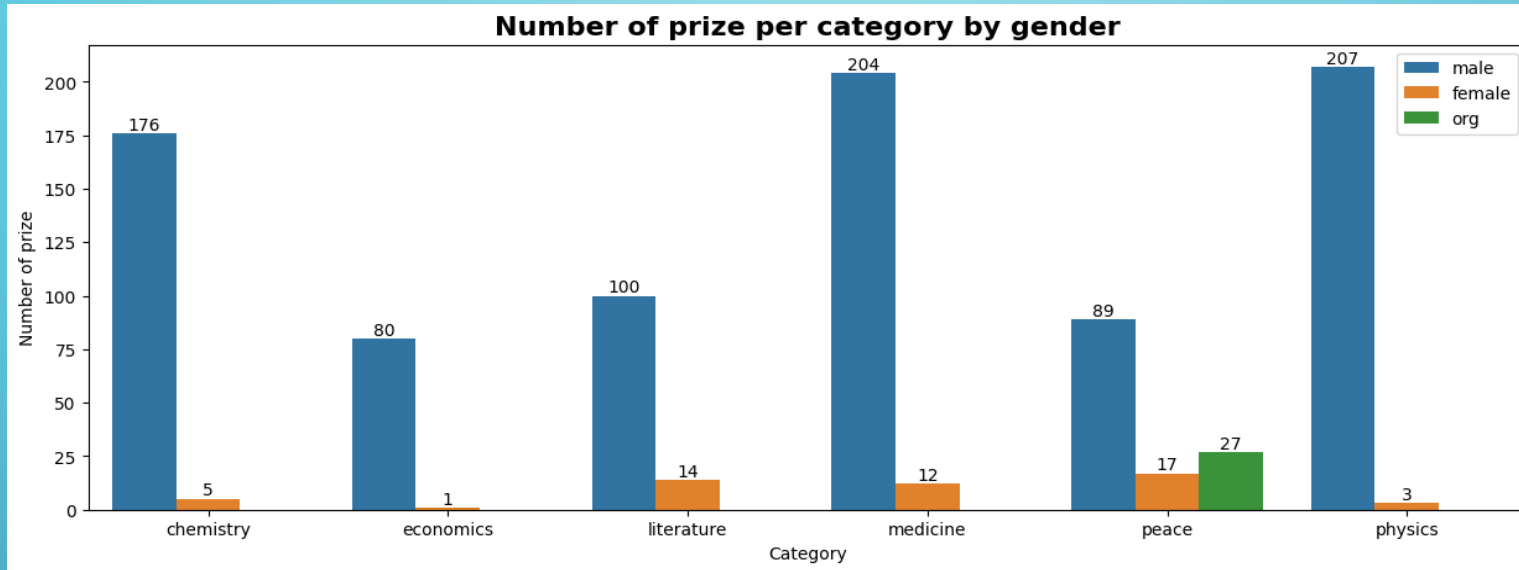The USA had the most immigrant laureates

5. The ratio of Winners

More than 850 Men won prizes, about 50 Women did and less than 30 for Organizations

# 6. Number of prize per category

Medicine, Physics and Chemistry are the fields with the highest number of prizes
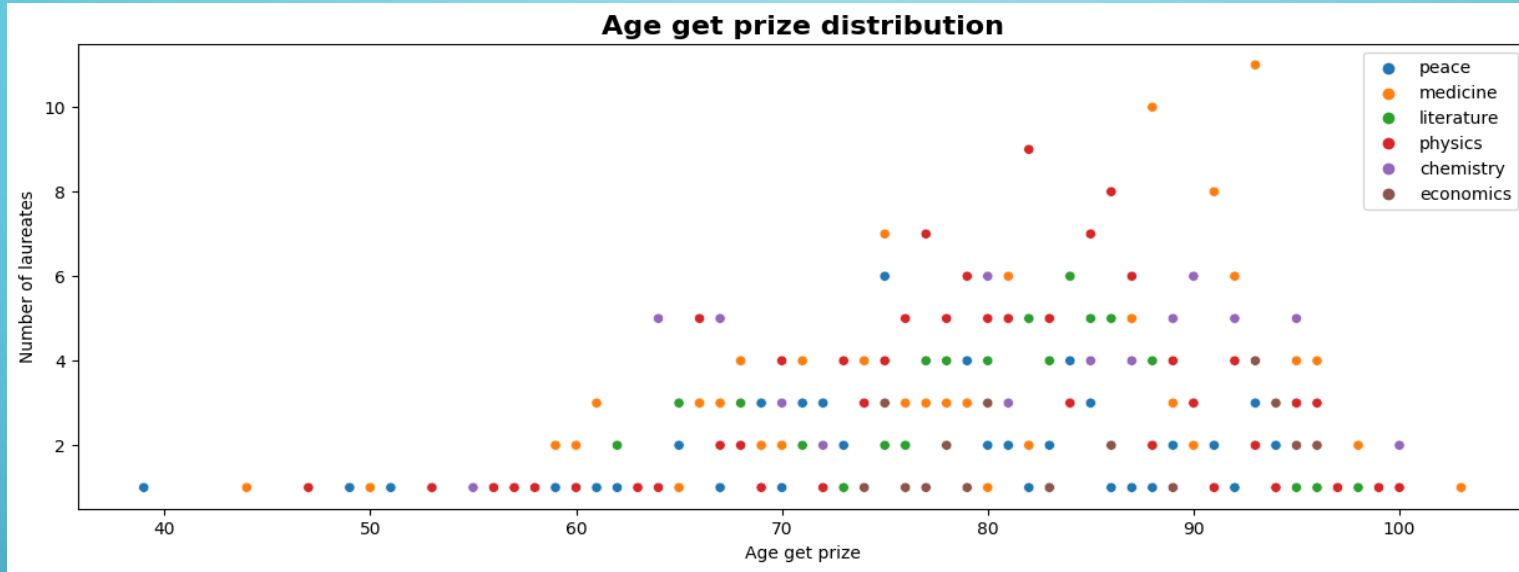
7. Category By Gender

Women have won the most prizes in Medicine, Literature and Peace while Physics, Medicine and Chemistry are the most popular fields for men. Organizations have solely received honors in the sphere of peace.

8. Top Universities that Nobel winners studied
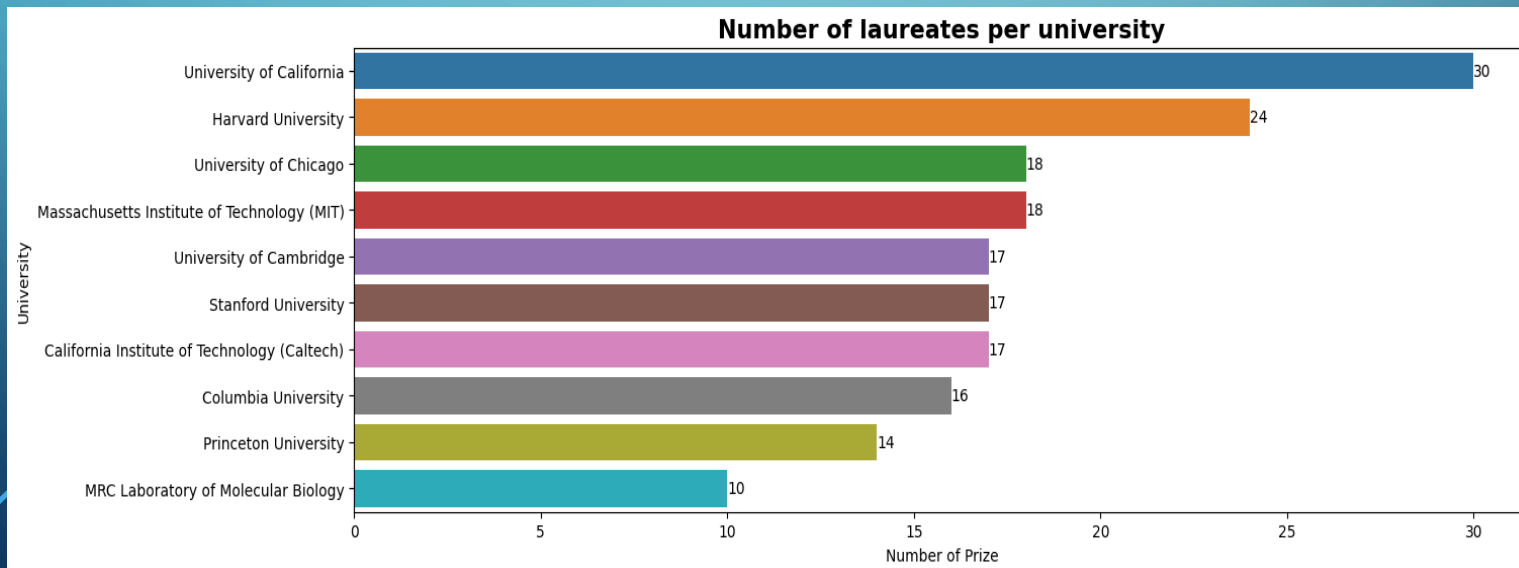
University of California, Harvard University, University of Chicago and Massachusetts Institute of Technology (MIT) are top 3 Universities that Nobel winners studied there

9. Distribution of Age Winners get prize

The ages between 75 and 95 years old are the highest number of laureates, particularly the peak at 85 ages.

10. Age of winners separated by gender and prize category
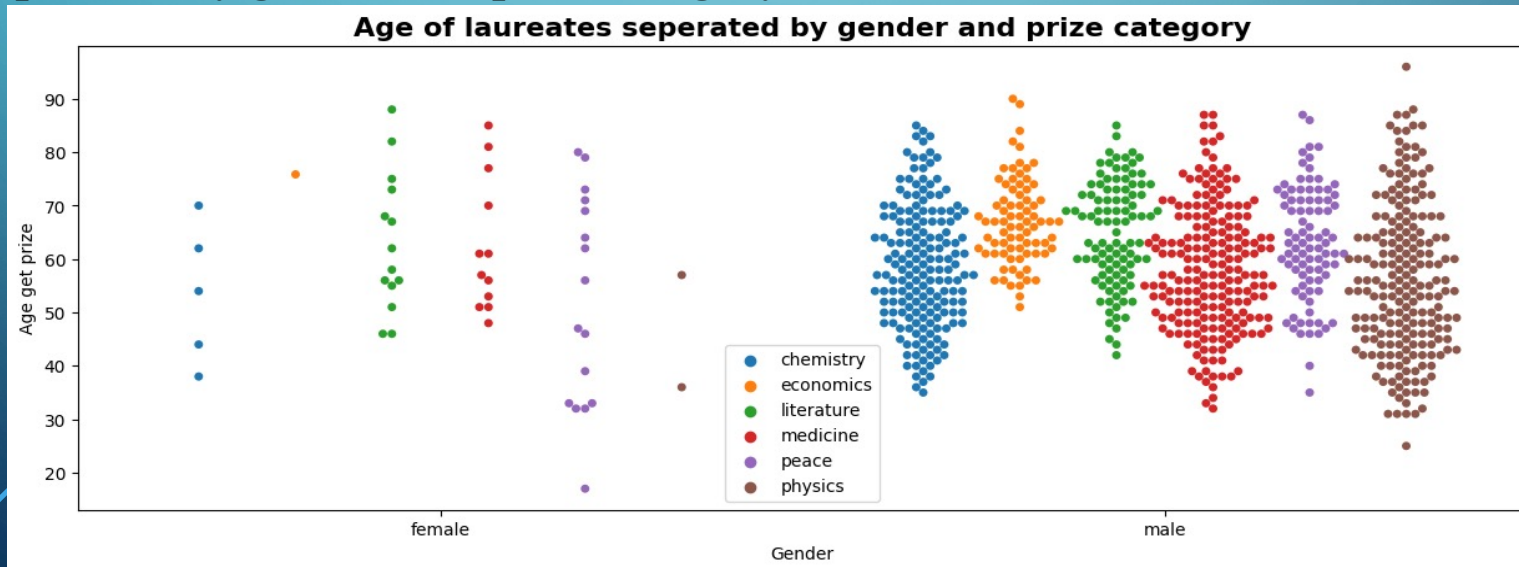
The youngest female Nobel laureate is 17 years old in the field of peace, while the Nobel prize in physics was awarded by the oldest 96-year-old man.

At the end of each section, I make a few comments about what I have studied through data analysis according to each criterion of the section.

I hope that this mini project will bring high efficiency to apply to practical situations in businesses.

Wishing everyone a ton of health, happiness, and success.

Sincerely thank you.

THE END!