

Analysis of *Nobel Prize laureates* dataset

January 2023

Dataset insight

Data used in this project comes from *public.opendatasoft.com* website and it includes Nobel Prizes from the very 1901 to last year - 2022.

At the beginning, let's take a look at the dataset by checking the column names.

```
## [1] "Firstname"          "Surname"            "Born"
## [4] "Died"               "Born.country"       "Born.country.code"
## [7] "Born.city"          "Died.country"       "Died.country.code"
## [10] "Died.city"          "Gender"             "Year"
## [13] "Age"                "Category"           "Overall.motivation"
## [16] "Motivation"         "Organization.name"   "Organization.city"
## [19] "Organization.country"
```

There are plenty of columns with both - numerical and text - values. In this case study I focused mostly on succeeding variables:

- Category - category the Prize was awarded at,
- Gender - gender of the laureate,
- Born - year when the laureate was born,
- Year - year when Nobel Prize was awarded,
- Age - calculated as a difference between *Year* and *Born* values,
- Born.country - place of birth of the laureate,
- Organization - place of work of the laureate,
- Motivation - argumentation for awarding the laureate.

Below you can find data analysis and some visualizations grouped in five sections:

- General statistics,
- Gender,
- Age,
- Background,
- Prize motivation.

Please enjoy!

O.

General statistics

The Nobel Prize laureates have been selected since

```
## [1] 1901
```

and there are 6 categories:

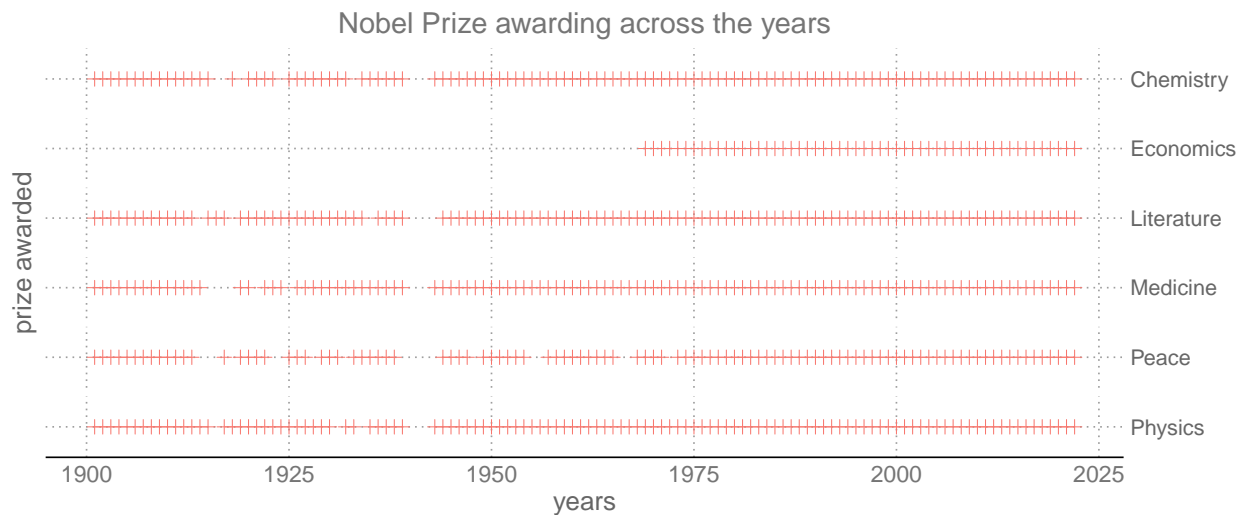
```
## [1] "Chemistry" "Economics" "Literature" "Medicine" "Peace"  
## [6] "Physics"
```

In the table below years when the Prizes started to be awarded are presented:

Category	First Year to be awarded
Chemistry	1901
Economics	1969
Literature	1901
Medicine	1901
Peace	1901
Physics	1901

Not for all of the categories the Prizes were awarded from the very beginning. It started with 5 of current 6 categories. The youngest category - Economics - was added almost 70 years later.

Let's take a look at the categories the Nobel Prizes were presented at through the years:



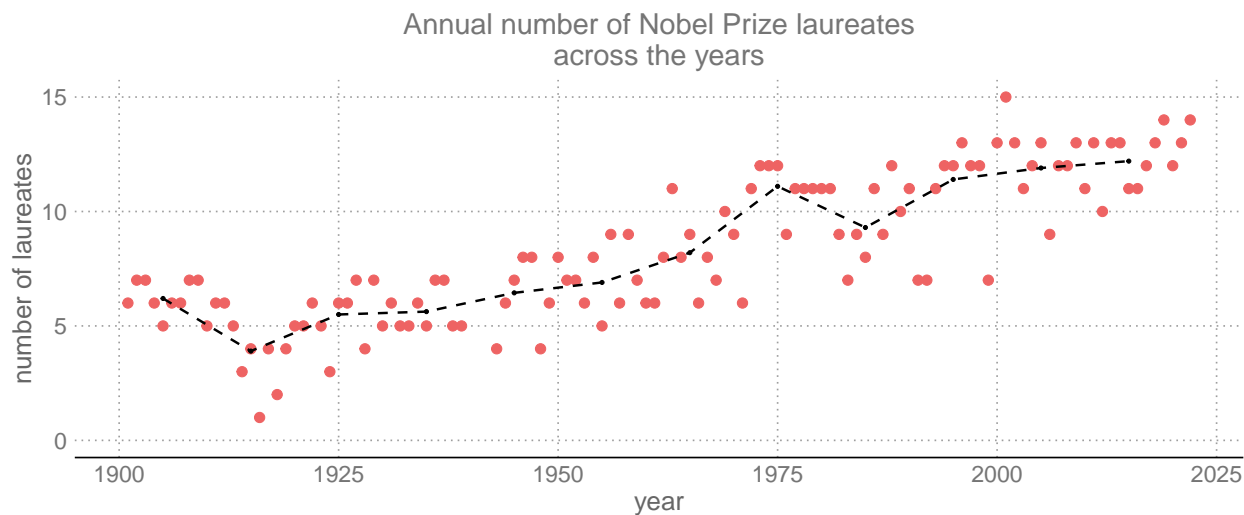
There is a noticable break of **3** years in Prize awarding from 1940 to 1942.

Since **1973** Prizes are always presented in all six categories.

Years when Prizes were awarded in least number of categories:

Year	Number of categories	Category
1916	1	Literature
1918	2	Chemistry, Physics

Moreover, number of laureates has been changing throughout the years as well:



There is a observable increase of number of annually awarded laureates.

Years with most laureates: **2001** (15), **2019** (14) and **2022** (14).

Years with least laureates: **1916** (1), **1918** (2), **1914** (3) and **1924** (3).

Now, let's take a closer look at yearly statistics of number of winners in each category:

Category	Average	Median	Maximum
Chemistry	1.7	1	3
Economics	1.7	2	3
Literature	1.0	1	2
Medicine	2.0	2	3
Peace	1.4	1	3
Physics	1.9	2	3

In each category total numbers of laureates are:

Category	Total number of laureates
Economics	92
Literature	119
Peace	140
Chemistry	191
Physics	222
Medicine	225

There are laureates who were awarded more than once:

Full Name	Prize year	Category	Age
International Committee of the Red Cross	1917	Peace	NA
International Committee of the Red Cross	1944	Peace	NA
International Committee of the Red Cross	1963	Peace	NA
Barry Sharpless	2001	Chemistry	60
Barry Sharpless	2022	Chemistry	81
Frederick Sanger	1958	Chemistry	40
Frederick Sanger	1980	Chemistry	62
John Bardeen	1956	Physics	48
John Bardeen	1972	Physics	64
Linus Pauling	1954	Chemistry	53
Linus Pauling	1962	Peace	61
Marie Curie	1903	Physics	36
Marie Curie	1911	Chemistry	44
Office of the United Nations High Commissioner for Refugees	1954	Peace	4
Office of the United Nations High Commissioner for Refugees	1981	Peace	31

Fun fact - most popular names among the Nobelists are:

Firstname	Number of laureates
John	12
Paul	10
Albert	8
Peter	8
Robert	8

However, there are not 12 but 11 different Johns, because John Bardeen is a double Nobel laureate. Considering the popularity of John name, let's view the countries all Johns were born in:

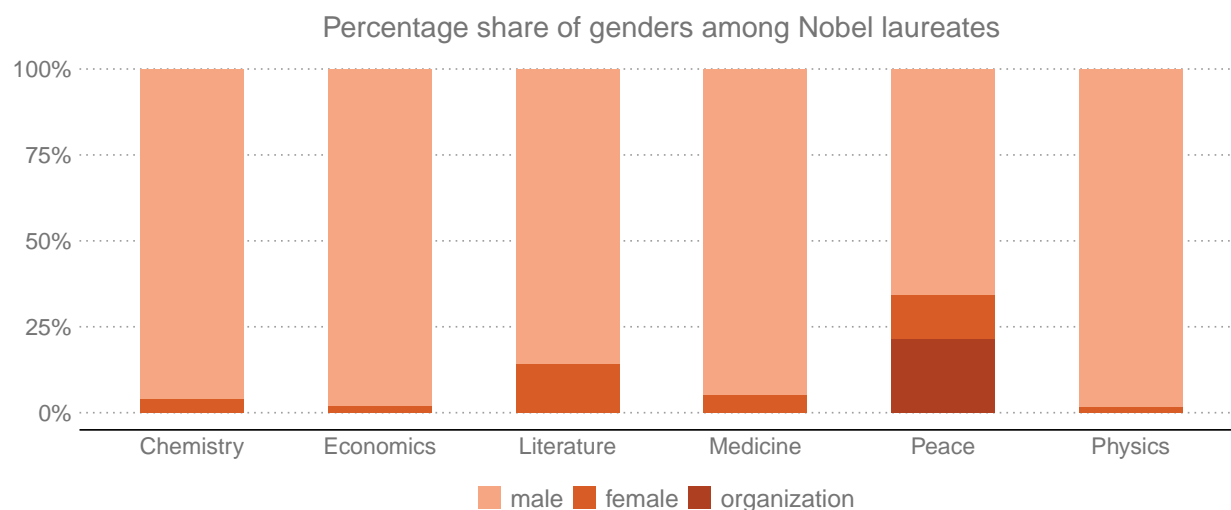
Full Name	Born.country
John Macleod	Scotland
John Galsworthy	United Kingdom
John Boyd Orr	Scotland
John Cockcroft	United Kingdom
John Bardeen	USA
John Steinbeck	USA
John Cornforth	Australia
John Pople	United Kingdom
John Hume	Northern Ireland
John O'Keefe	USA
John Goodenough	Germany

Gender

At the beginning, let's look how many women and men there are among all laureates:

Gender	Number of winners
female	61
male	898
organization	30

Women Nobelists represent **6.4%** of all winners. And within all categories:



First year when women were prized in all categories:

Firstname	Surname	Year	Category
Marie	Curie	1911	Chemistry
Elinor	Ostrom	2009	Economics
Selma	Lagerlof	1909	Literature
Gerty	Cori	1947	Medicine
Bertha	von Suttner	1905	Peace
Marie	Curie	1903	Physics

Marie Curie is the youngest woman Nobelist in two categories and at all.

In below table you can see years, when most women were awarded with comparison to all laureates:

Year	Awarded women	All
2009	5	13
2020	4	12
2018	4	13
2011	3	13
2004	3	12

There are multiple years when women were not awarded at all. For **119** years of Nobel Prizes, at least one woman was awarded **42** times. It gives **77** times with only men being awarded (**65.7%**).

Age

Now, let's see some statistics for the Nobelists' age.

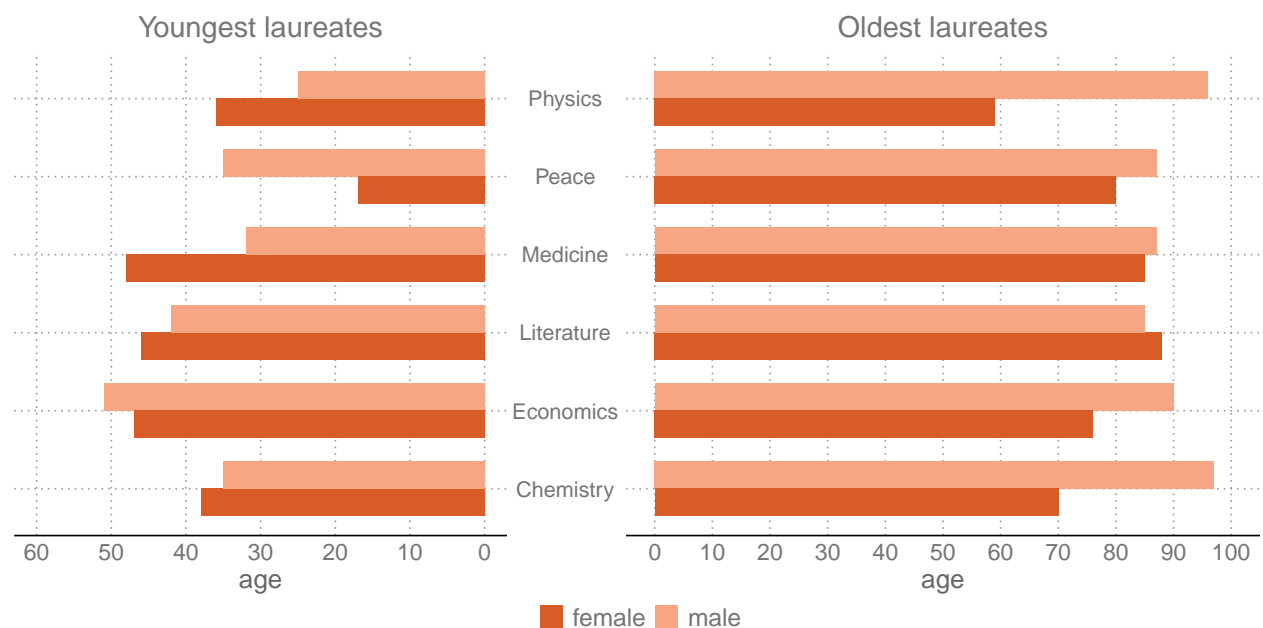
Oldest laureates:

Firstname	Surname	Year	Age	Category
John	Goodenough	2019	97	Chemistry
Arthur	Ashkin	2018	96	Physics
Leonid	Hurwicz	2007	90	Economics
Klaus	Hasselmann	2021	90	Physics
Syukuro	Manabe	2021	90	Physics
Lloyd S.	Shapley	2012	89	Economics
Roger	Penrose	2020	89	Physics

Youngest laureates:

Firstname	Surname	Year	Age	Category
Malala	Yousafzai	2014	17	Peace
Lawrence	Bragg	1915	25	Physics
Werner	Heisenberg	1932	31	Physics
Paul A.M.	Dirac	1933	31	Physics
Carl D.	Anderson	1936	31	Physics
Tsung-Dao	Lee	1957	31	Physics

Below chart presents the comparison between youngest and oldest men and women Nobelists in all categories:

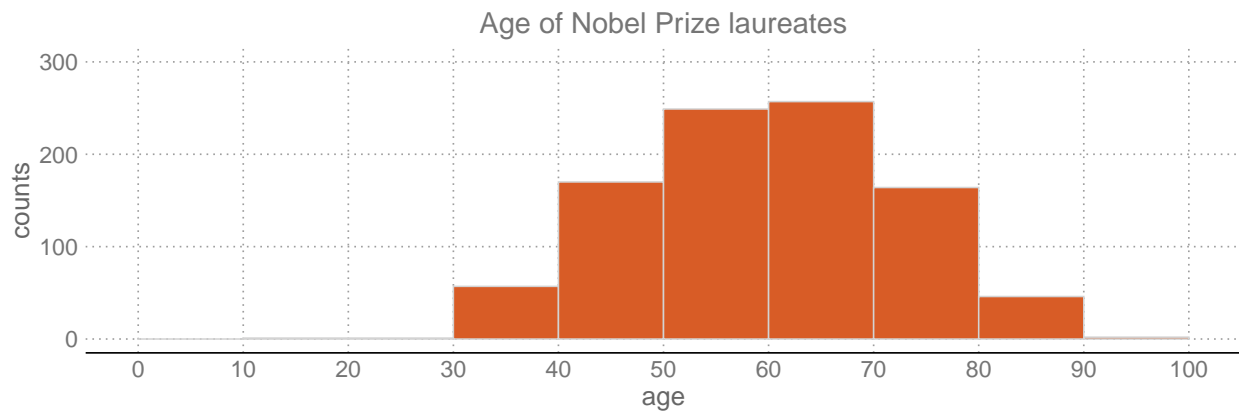


In below table, you can see statistics for laureates' **age** in each category:

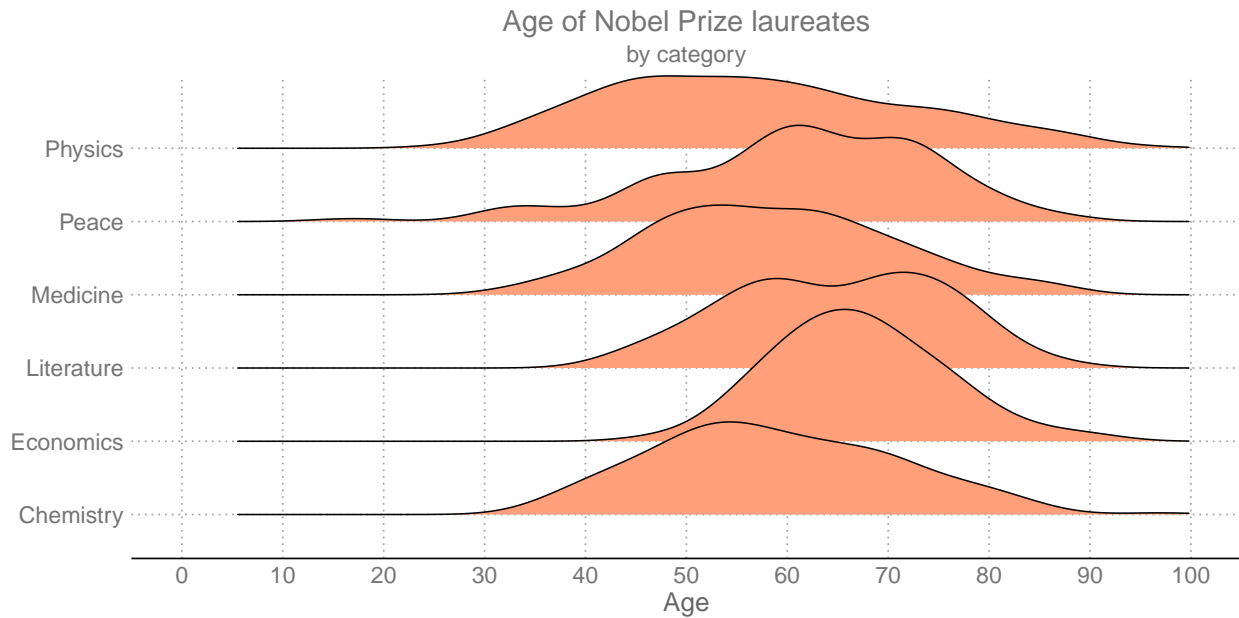
Category	Mean	Median	Stand dev
Chemistry	58.9	58	11.9
Economics	67.0	67	8.3
Literature	64.9	67	10.3
Medicine	58.6	57	11.7
Peace	60.4	62	14.2
Physics	57.1	56	14.8

Average age of Nobel Prize winner is **56**.

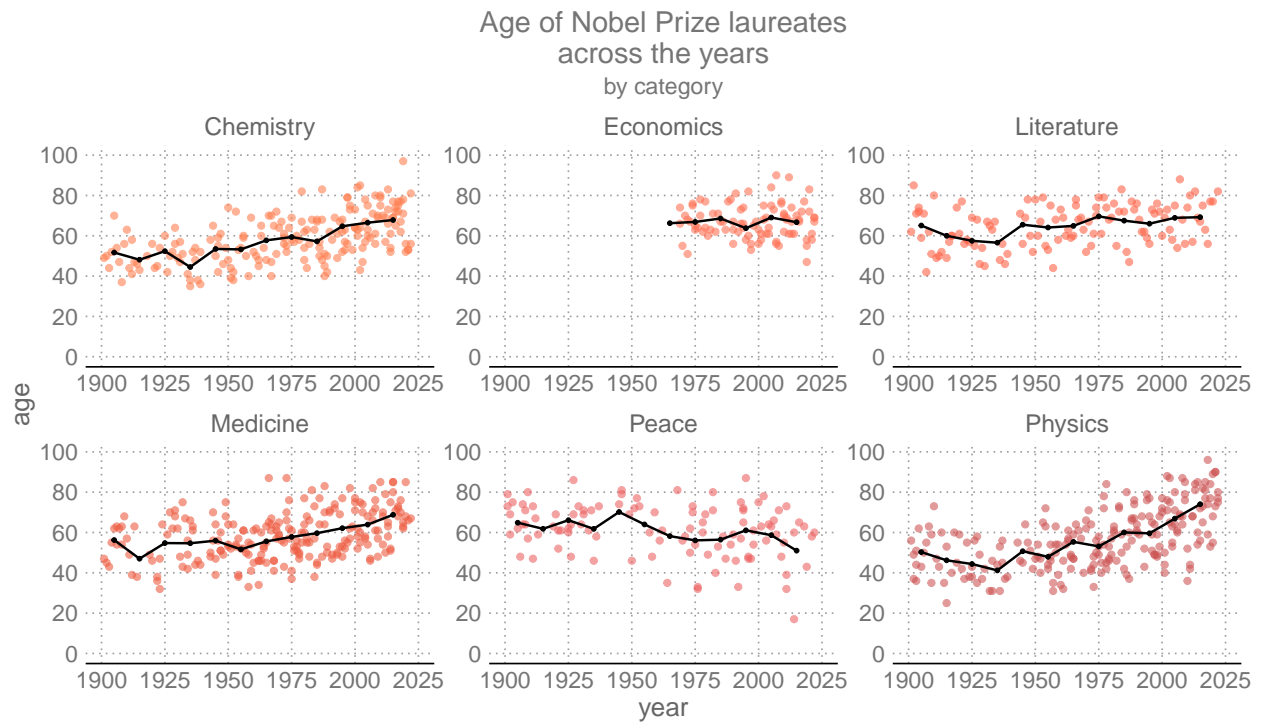
Histogram of laureates' age:



And after dividing into categories:



We can also see how the age of laureates in each category has been changing through the years:



There is a visible **increase** for **Chemistry**, **Physics** and **Medicine** and **decrease** for **Peace**.

Background

The dataset gives an opportunity to see where the Nobelists come from, by country and by research unit.

Most laureates were born in:

(note: born does not equal nationality)

Born.country	Number of laureates
USA	289
United Kingdom	90
Germany	67
France	56
	31
Sweden	30
Japan	28
Canada	21
Switzerland	19
the Netherlands	19

People born in United States of America make up a significant percentage of all laureates. Let's take a look at years when most of them were prized:

Year	Number of USA-born laureates	All
1972	8	11
1990	8	11
2001	8	15
1946	7	8
1993	7	11
2004	7	12

Although, there is no year with only American laureates awarded.

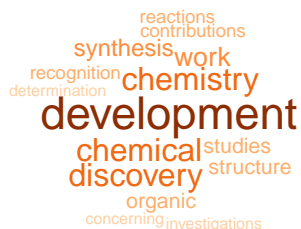
We can also find out, what universities most laureates are connected with:

Organization.name	Scientific unit
	262
University of California	36
Harvard University	27
Stanford University	22
Massachusetts Institute of Technology (MIT)	21
University of Chicago	19

Prize motivation

The last section in this report is the Prize motivation. For each category I carried out analysis of the motivation text by counting the frequency of appearance of words. The results are presented below in form of **word clouds**:

Chemistry



Economics



Literature



Medicine



Peace



Physics



And last but not least - most popular two-word combinations in the motivations (**bigrams**):

