

# notebook

July 27, 2022

## 0.1 1. The most Nobel of Prizes

The Nobel Prize is perhaps the world's most well known scientific award. Except for the honor, prestige and substantial prize money the recipient also gets a gold medal showing Alfred Nobel (1833 - 1896) who established the prize. Every year it's given to scientists and scholars in the categories chemistry, literature, physics, physiology or medicine, economics, and peace. The first Nobel Prize was handed out in 1901, and at that time the Prize was very Eurocentric and male-focused, but nowadays it's not biased in any way whatsoever. Surely. Right?

Well, we're going to find out! The Nobel Foundation has made a dataset available of all prize winners from the start of the prize, in 1901, to 2016. Let's load it in and take a look.

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

# Reading in the Nobel Prize data
nobel = pd.read_csv('datasets/nobel.csv')

# Taking a look at the first several winners
nobel.head(n = 6)
```

```
[ ]:  year      category      prize \
0  1901    Chemistry  The Nobel Prize in Chemistry 1901
1  1901  Literature  The Nobel Prize in Literature 1901
2  1901    Medicine  The Nobel Prize in Physiology or Medicine 1901
3  1901      Peace  The Nobel Peace Prize 1901
4  1901      Peace  The Nobel Peace Prize 1901
5  1901    Physics  The Nobel Prize in Physics 1901

      motivation prize_share laureate_id \
0  "in recognition of the extraordinary services ..."      1/1      160
1  "in special recognition of his poetic composi..."      1/1      569
2  "for his work on serum therapy, especially its..."      1/1      293
3                                NaN      1/2      462
4                                NaN      1/2      463
5  "in recognition of the extraordinary services ..."      1/1       1
```

	laureate_type	full_name	birth_date	birth_city \
0	Individual	Jacobus Henricus van 't Hoff	1852-08-30	Rotterdam
1	Individual	Sully Prudhomme	1839-03-16	Paris
2	Individual	Emil Adolf von Behring	1854-03-15	Hansdorf (Lawice)
3	Individual	Jean Henry Dunant	1828-05-08	Geneva
4	Individual	Frédéric Passy	1822-05-20	Paris
5	Individual	Wilhelm Conrad Röntgen	1845-03-27	Lennepe (Remscheid)

	birth_country	sex	organization_name	organization_city \
0	Netherlands	Male	Berlin University	Berlin
1	France	Male	NaN	NaN
2	Prussia (Poland)	Male	Marburg University	Marburg
3	Switzerland	Male	NaN	NaN
4	France	Male	NaN	NaN
5	Prussia (Germany)	Male	Munich University	Munich

	organization_country	death_date	death_city	death_country
0	Germany	1911-03-01	Berlin	Germany
1	NaN	1907-09-07	Châtenay	France
2	Germany	1917-03-31	Marburg	Germany
3	NaN	1910-10-30	Heiden	Switzerland
4	NaN	1912-06-12	Paris	France
5	Germany	1923-02-10	Munich	Germany

## 0.2 2. So, who gets the Nobel Prize?

Just looking at the first couple of prize winners, or Nobel laureates as they are also called, we already see a celebrity: Wilhelm Conrad Röntgen, the guy who discovered X-rays. And actually, we see that all of the winners in 1901 were guys that came from Europe. But that was back in 1901, looking at all winners in the dataset, from 1901 to 2016, which sex and which country is the most commonly represented?

(For country, we will use the birth\_country of the winner, as the organization\_country is NaN for all shared Nobel Prizes.)

```
[ ]: # Display the number of (possibly shared) Nobel Prizes handed
# out between 1901 and 2016
display(len(nobel))

# Display the number of prizes won by male and female recipients.
display(nobel['sex'].value_counts())

# Display the number of prizes won by the top 10 nationalities.
nobel['birth_country'].value_counts().head(10)
```

911

Male	836
Female	49

Name: sex, dtype: int64

```
[ ]: United States of America    259
      United Kingdom             85
      Germany                   61
      France                    51
      Sweden                    29
      Japan                     24
      Netherlands               18
      Canada                    18
      Italy                     17
      Russia                    17
      Name: birth_country, dtype: int64
```

### 0.3 3. USA dominance

Not so surprising perhaps: the most common Nobel laureate between 1901 and 2016 was a man born in the United States of America. But in 1901 all the winners were European. When did the USA start to dominate the Nobel Prize charts?

```
[ ]: # Calculating the proportion of USA born winners per decade
nobel['usa_born_winner'] = (nobel['birth_country'] == 'United States of_
    ↪America')
nobel['decade'] = (np.floor(nobel['year'] / 10) * 10).astype(np.int64)
prop_usa_winners = nobel.groupby('decade', as_index = False)['usa_born_winner'].
    ↪mean()

# Display the proportions of USA born winners per decade
display(prop_usa_winners)
```

	decade	usa_born_winner
0	1900	0.017544
1	1910	0.075000
2	1920	0.074074
3	1930	0.250000
4	1940	0.302326
5	1950	0.291667
6	1960	0.265823
7	1970	0.317308
8	1980	0.319588
9	1990	0.403846
10	2000	0.422764
11	2010	0.292683

### 0.4 4. USA dominance, visualized

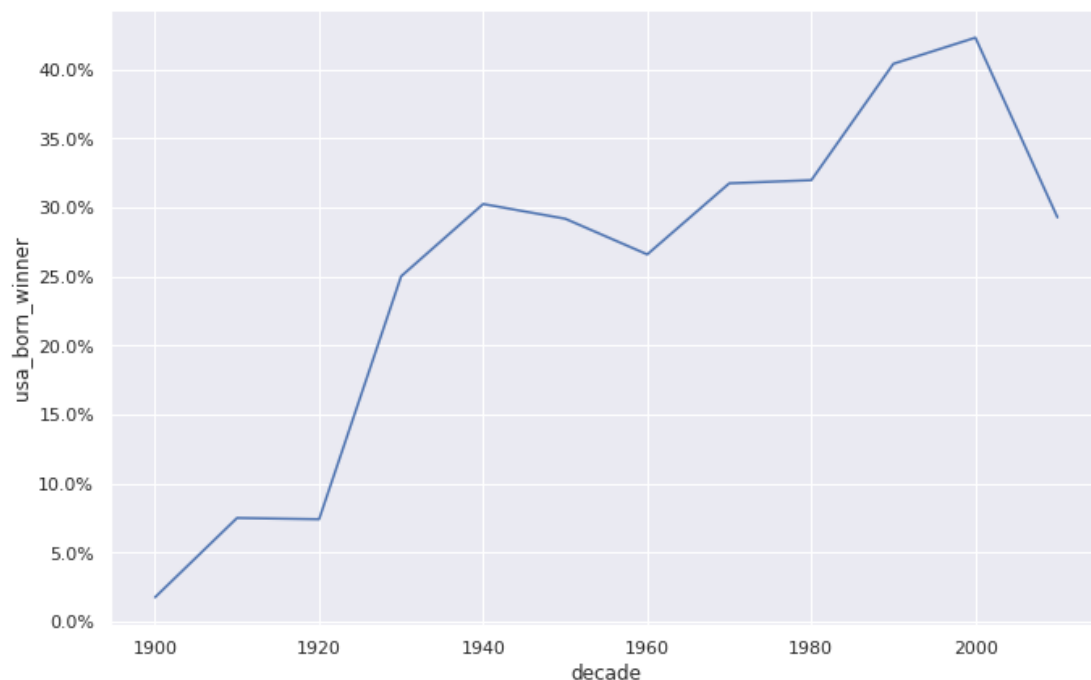
A table is OK, but to see when the USA started to dominate the Nobel charts we need a plot!

```
[ ]: # Setting the plotting theme
sns.set()

# and setting the size of all plots.
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = [11, 7]

# Plotting USA born winners
ax = sns.lineplot(data = prop_usa_winners, x = 'decade', y = 'usa_born_winner')

# Adding %-formatting to the y-axis
from matplotlib.ticker import PercentFormatter
ax.yaxis.set_major_formatter(PercentFormatter(1.0))
```



## 0.5 5. What is the gender of a typical Nobel Prize winner?

So the USA became the dominating winner of the Nobel Prize first in the 1930s and had kept the leading position ever since. But one group that was in the lead from the start, and never seems to let go, are men. Maybe it shouldn't come as a shock that there is some imbalance between how many male and female prize winners there are, but how significant is this imbalance? And is it better or worse within specific prize categories like physics, medicine, literature, etc.?

```
[ ]: # Calculating the proportion of female laureates per decade
nobel['female_winner'] = nobel['sex'] == 'Female'
```

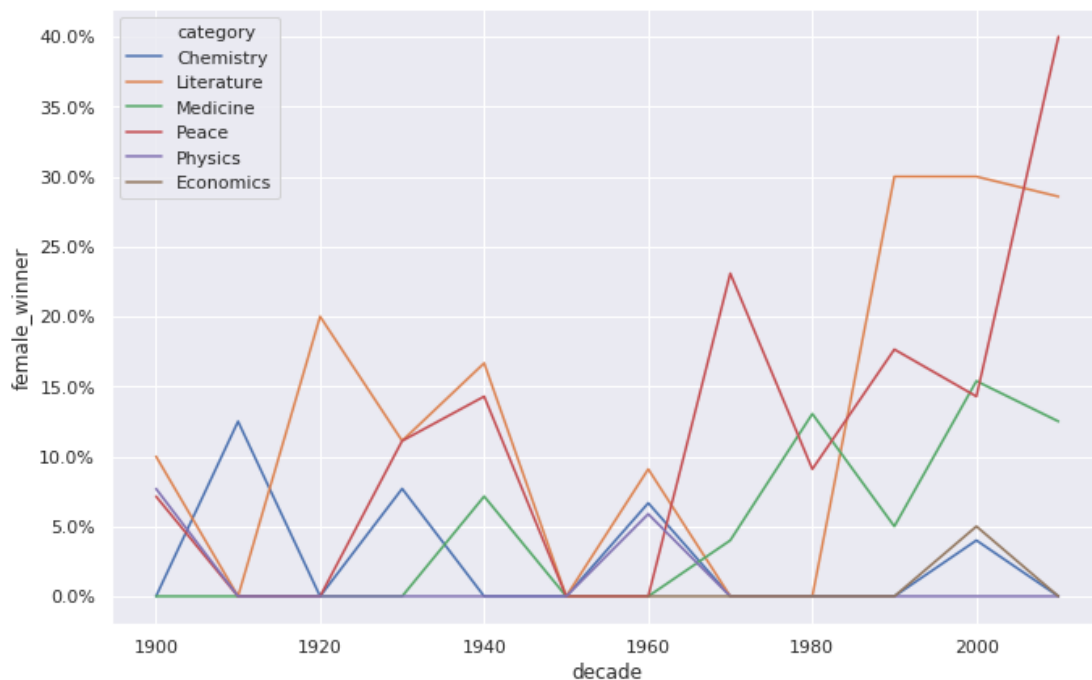
```

prop_female_winners = nobel.groupby(['decade', 'category'], as_index =_
↳False)['female_winner'].mean()

# Plotting USA born winners with % winners on the y-axis
ax = sns.lineplot(data = prop_female_winners, x = 'decade', y =_
↳'female_winner', hue = 'category')

# Adding %-formatting to the y-axis
from matplotlib.ticker import PercentFormatter
ax.yaxis.set_major_formatter(PercentFormatter(1.0))

```



## 0.6 6. The first woman to win the Nobel Prize

The plot above is a bit messy as the lines are overplotting. But it does show some interesting trends and patterns. Overall the imbalance is pretty large with physics, economics, and chemistry having the largest imbalance. Medicine has a somewhat positive trend, and since the 1990s the literature prize is also now more balanced. The big outlier is the peace prize during the 2010s, but keep in mind that this just covers the years 2010 to 2016.

Given this imbalance, who was the first woman to receive a Nobel Prize? And in what category?

```

[ ]: # Picking out the first woman to win a Nobel Prize
nobel[nobel['sex'] == 'Female'].nsmallest(1, 'year', keep = 'first')

```

```
[ ]:      year category                                prize \
19  1903  Physics  The Nobel Prize in Physics 1903

                                motivation prize_share \
19  "in recognition of the extraordinary services ...      1/4

    laureate_id laureate_type                                full_name birth_date \
19              6      Individual  Marie Curie, née Sklodowska 1867-11-07

    birth_city      ...      sex organization_name organization_city \
19      Warsaw      ...      Female      NaN      NaN

    organization_country death_date death_city death_country usa_born_winner \
19              NaN  1934-07-04  Sallanches      France      False

    decade female_winner
19      1900              True

[1 rows x 21 columns]
```

## 0.7 7. Repeat laureates

For most scientists/writers/activists a Nobel Prize would be the crowning achievement of a long career. But for some people, one is just not enough, and few have gotten it more than once. Who are these lucky few? (Having won no Nobel Prize myself, I'll assume it's just about luck.)

```
[ ]: # Selecting the laureates that have received 2 or more prizes.
nobel.groupby('full_name').filter(lambda x: len(x) >= 2)
```

```
[ ]:      year  category                                prize \
19   1903   Physics  The Nobel Prize in Physics 1903
62   1911  Chemistry  The Nobel Prize in Chemistry 1911
89   1917    Peace    The Nobel Peace Prize 1917
215  1944    Peace    The Nobel Peace Prize 1944
278  1954  Chemistry  The Nobel Prize in Chemistry 1954
283  1954    Peace    The Nobel Peace Prize 1954
298  1956   Physics  The Nobel Prize in Physics 1956
306  1958  Chemistry  The Nobel Prize in Chemistry 1958
340  1962    Peace    The Nobel Peace Prize 1962
348  1963    Peace    The Nobel Peace Prize 1963
424  1972   Physics  The Nobel Prize in Physics 1972
505  1980  Chemistry  The Nobel Prize in Chemistry 1980
523  1981    Peace    The Nobel Peace Prize 1981

                                motivation prize_share \
19  "in recognition of the extraordinary services ...      1/4
62  "in recognition of her services to the advance...      1/1
```

89		NaN	1/1
215		NaN	1/1
278	"for his research into the nature of the chemi...		1/1
283		NaN	1/1
298	"for their researches on semiconductors and th...		1/3
306	"for his work on the structure of proteins, es...		1/1
340		NaN	1/1
348		NaN	1/2
424	"for their jointly developed theory of superco...		1/3
505	"for their contributions concerning the determ...		1/4
523		NaN	1/1

	laureate_id	laureate_type	\
19	6	Individual	
62	6	Individual	
89	482	Organization	
215	482	Organization	
278	217	Individual	
283	515	Organization	
298	66	Individual	
306	222	Individual	
340	217	Individual	
348	482	Organization	
424	66	Individual	
505	222	Individual	
523	515	Organization	

	full_name	birth_date	\
19	Marie Curie, née Sklodowska	1867-11-07	
62	Marie Curie, née Sklodowska	1867-11-07	
89	Comité international de la Croix Rouge (Intern...	NaN	
215	Comité international de la Croix Rouge (Intern...	NaN	
278	Linus Carl Pauling	1901-02-28	
283	Office of the United Nations High Commissioner...	NaN	
298	John Bardeen	1908-05-23	
306	Frederick Sanger	1918-08-13	
340	Linus Carl Pauling	1901-02-28	
348	Comité international de la Croix Rouge (Intern...	NaN	
424	John Bardeen	1908-05-23	
505	Frederick Sanger	1918-08-13	
523	Office of the United Nations High Commissioner...	NaN	

	birth_city	...	sex	\
19	Warsaw	...	Female	
62	Warsaw	...	Female	
89	NaN	...	NaN	
215	NaN	...	NaN	

278	Portland, OR	...	Male
283	NaN	...	NaN
298	Madison, WI	...	Male
306	Rendcombe	...	Male
340	Portland, OR	...	Male
348	NaN	...	NaN
424	Madison, WI	...	Male
505	Rendcombe	...	Male
523	NaN	...	NaN

	organization_name	organization_city	\
19	NaN	NaN	
62	Sorbonne University	Paris	
89	NaN	NaN	
215	NaN	NaN	
278	California Institute of Technology (Caltech)	Pasadena, CA	
283	NaN	NaN	
298	University of Illinois	Urbana, IL	
306	University of Cambridge	Cambridge	
340	California Institute of Technology (Caltech)	Pasadena, CA	
348	NaN	NaN	
424	University of Illinois	Urbana, IL	
505	MRC Laboratory of Molecular Biology	Cambridge	
523	NaN	NaN	

	organization_country	death_date	death_city	\
19	NaN	1934-07-04	Sallanches	
62	France	1934-07-04	Sallanches	
89	NaN	NaN	NaN	
215	NaN	NaN	NaN	
278	United States of America	1994-08-19	Big Sur, CA	
283	NaN	NaN	NaN	
298	United States of America	1991-01-30	Boston, MA	
306	United Kingdom	2013-11-19	Cambridge	
340	United States of America	1994-08-19	Big Sur, CA	
348	NaN	NaN	NaN	
424	United States of America	1991-01-30	Boston, MA	
505	United Kingdom	2013-11-19	Cambridge	
523	NaN	NaN	NaN	

	death_country	usa_born_winner	decade	female_winner
19	France	False	1900	True
62	France	False	1910	True
89	NaN	False	1910	False
215	NaN	False	1940	False
278	United States of America	True	1950	False
283	NaN	False	1950	False



298	United States of America	True	1950	False
306	United Kingdom	False	1950	False
340	United States of America	True	1960	False
348	NaN	False	1960	False
424	United States of America	True	1970	False
505	United Kingdom	False	1980	False
523	NaN	False	1980	False

[13 rows x 21 columns]

## 0.8 8. How old are you when you get the prize?

The list of repeat winners contains some illustrious names! We again meet Marie Curie, who got the prize in physics for discovering radiation and in chemistry for isolating radium and polonium. John Bardeen got it twice in physics for transistors and superconductivity, Frederick Sanger got it twice in chemistry, and Linus Carl Pauling got it first in chemistry and later in peace for his work in promoting nuclear disarmament. We also learn that organizations also get the prize as both the Red Cross and the UNHCR have gotten it twice.

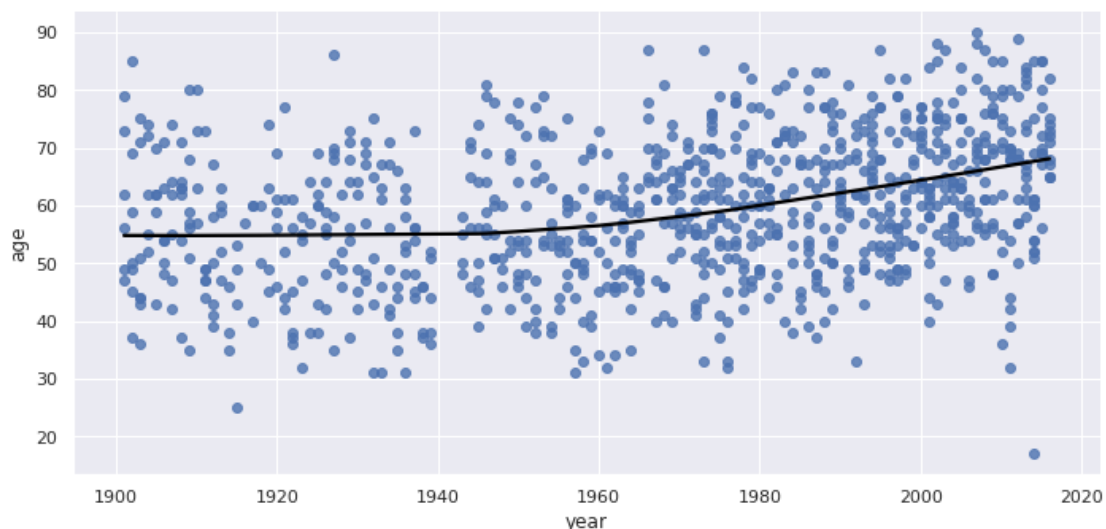
But how old are you generally when you get the prize?

```
[ ]: # Converting birth_date from String to datetime
nobel['birth_date'] = pd.to_datetime(nobel['birth_date'])

# Calculating the age of Nobel Prize winners
nobel['age'] = nobel['year'] - nobel['birth_date'].dt.year

# Plotting the age of Nobel Prize winners
sns.lmplot(x='year', y='age', data=nobel, lowess=True,
           aspect=2, line_kws={'color' : 'black'})
```

```
[ ]: <seaborn.axisgrid.FacetGrid at 0x7fd0911539e8>
```



## 0.9 9. Age differences between prize categories

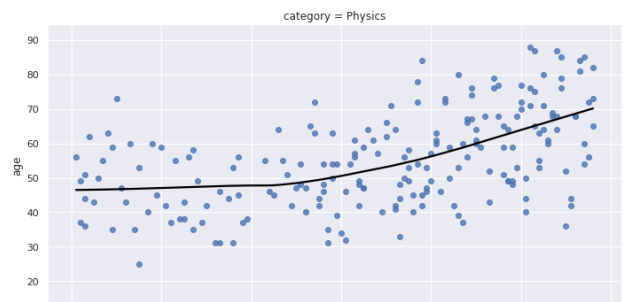
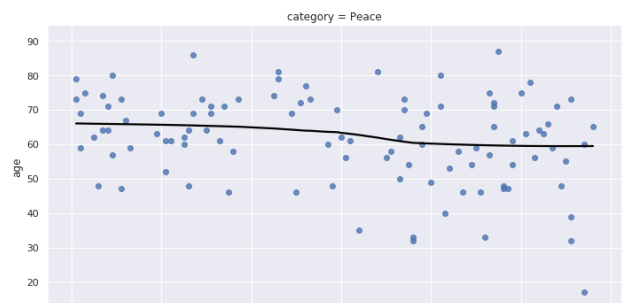
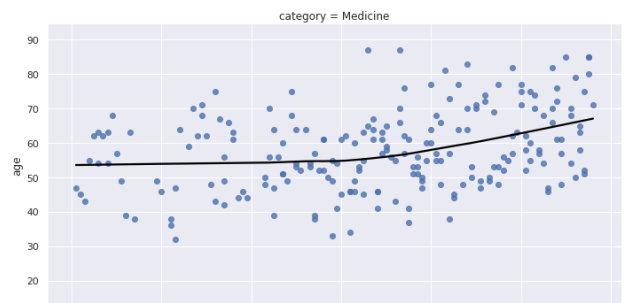
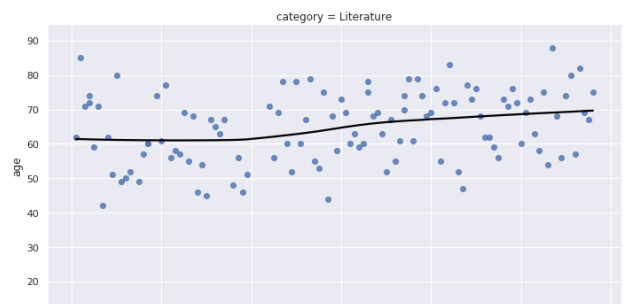
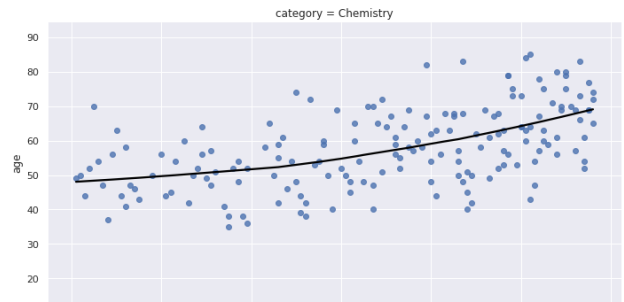
The plot above shows us a lot! We see that people use to be around 55 when they received the prize, but nowadays the average is closer to 65. But there is a large spread in the laureates' ages, and while most are 50+, some are very young.

We also see that the density of points is much high nowadays than in the early 1900s – nowadays many more of the prizes are shared, and so there are many more winners. We also see that there was a disruption in awarded prizes around the Second World War (1939 - 1945).

Let's look at age trends within different prize categories.

```
[ ]: # Same plot as above, but separate plots for each type of Nobel Prize
sns.lmplot(x='year', y='age', data=nobel, lowess=True,
           aspect=2, line_kws={'color' : 'black'}, row = 'category')
```

```
[ ]: <seaborn.axisgrid.FacetGrid at 0x7fd09184a320>
```



## 0.10 10. Oldest and youngest winners

More plots with lots of exciting stuff going on! We see that both winners of the chemistry, medicine, and physics prize have gotten older over time. The trend is strongest for physics: the average age used to be below 50, and now it's almost 70. Literature and economics are more stable. We also see that economics is a newer category. But peace shows an opposite trend where winners are getting younger!

In the peace category we also a winner around 2010 that seems exceptionally young. This begs the questions, who are the oldest and youngest people ever to have won a Nobel Prize?

```
[ ]: # The oldest winner of a Nobel Prize as of 2016
display(nobel.nlargest(1, 'age'))

# The youngest winner of a Nobel Prize as of 2016
nobel.nsmallest(1, 'age')
```

```
      year  category                                     prize \
793  2007  Economics  The Sveriges Riksbank Prize in Economic Scienc...

      motivation prize_share \
793  "for having laid the foundations of mechanism ...      1/3

      laureate_id laureate_type      full_name birth_date birth_city ... \
793           820    Individual  Leonid Hurwicz 1917-08-21    Moscow ...

      organization_name organization_city      organization_country \
793  University of Minnesota  Minneapolis, MN  United States of America

      death_date      death_city      death_country usa_born_winner \
793  2008-06-24  Minneapolis, MN  United States of America      False

      decade  female_winner  age
793    2000           False  90.0

[1 rows x 22 columns]
```

```
[ ]:      year category                                     prize \
885  2014    Peace  The Nobel Peace Prize 2014

      motivation prize_share \
885  "for their struggle against the suppression of...      1/2

      laureate_id laureate_type      full_name birth_date birth_city ... \
885           914    Individual  Malala Yousafzai 1997-07-12    Mingora ...
```

	organization_name	organization_city	organization_country	death_date	\
885	NaN	NaN	NaN	NaN	

	death_city	death_country	usa_born_winner	decade	female_winner	age
885	NaN	NaN	False	2010	True	17.0

[1 rows x 22 columns]

## 0.11 11. You get a prize!

Hey! You get a prize for making it to the very end of this notebook! It might not be a Nobel Prize, but I made it myself in paint so it should count for something. But don't despair, Leonid Hurwicz was 90 years old when he got his prize, so it might not be too late for you. Who knows.

Before you leave, what was again the name of the youngest winner ever who in 2014 got the prize for "[her] struggle against the suppression of children and young people and for the right of all children to education"?

```
[ ]: # The name of the youngest winner of the Nobel Prize as of 2016
youngest_winner = 'Malala Yousafzai'
```