**Unearthing Patterns in the Laureate Age at the Time of the Award**

How old are the Nobel laureates at the time when they win the prize? Does this vary by category? Also, how has the age of the laureates changed over time?

Challenge 1

Calculate the age of the laureate in the year of the ceremony and add this as a column called winning\_age to the df\_data DataFrame. Hint: you can use [this](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.dt.html) to help you.

Challenge 2

Who were the oldest and the youngest winners?

* What are the names of the youngest and oldest Nobel laureate?
* What did they win the prize for?
* What is the average age of a winner?
* 75% of laureates are younger than what age when they receive the prize?
* Use Seaborn to [create histogram](https://seaborn.pydata.org/generated/seaborn.histplot.html) to visualise the distribution of laureate age at the time of winning. Experiment with the number of bins to see how the visualisation changes.

Challenge 3

* Calculate the descriptive statistics for the age at the time of the award.
* Then visualise the distribution in the form of a histogram using [Seaborn's .histplot() function](https://seaborn.pydata.org/generated/seaborn.histplot.html" \t "_blank).
* Experiment with the bin size. Try 10, 20, 30, and 50.

Challenge 4

Are Nobel laureates being nominated later in life than before? Have the ages of laureates at the time of the award increased or decreased over time?

* Use Seaborn to [create a .regplot](https://seaborn.pydata.org/generated/seaborn.regplot.html?highlight=regplot#seaborn.regplot) with a trendline.
* Set the lowess parameter to True to show a moving average of the linear fit.
* According to the best fit line, how old were Nobel laureates in the years 1900-1940 when they were awarded the prize?
* According to the best fit line, what age would it predict for a Nobel laureate in 2020?

Challenge 5

How does the age of laureates vary by category?

* Use Seaborn's [.boxplot()](https://seaborn.pydata.org/generated/seaborn.boxplot.html?highlight=boxplot#seaborn.boxplot) to show how the mean, quartiles, max, and minimum values vary across categories. Which category has the longest "whiskers"?
* In which prize category are the average winners the oldest?
* In which prize category are the average winners the youngest?
* You can also use plotly to create the box plot if you like.

Challenge 6

* Now use Seaborn's [.lmplot()](https://seaborn.pydata.org/generated/seaborn.lmplot.html?highlight=lmplot#seaborn.lmplot) and the row parameter to create 6 separate charts for each prize category. Again set lowess to True.
* What are the winning age trends in each category?
* Which category has the age trending up and which category has the age trending down?
* Is this .lmplot() telling a different story from the .boxplot()?
* Create a third chart with Seaborn. This time use .lmplot() to put all 6 categories on the same chart using the hue parameter.

.

.

..

...

..

.

.

**Solution 1: Calculate the Age at the Time of Award**

First, we need to extract the year as a number from the birth\_date column:

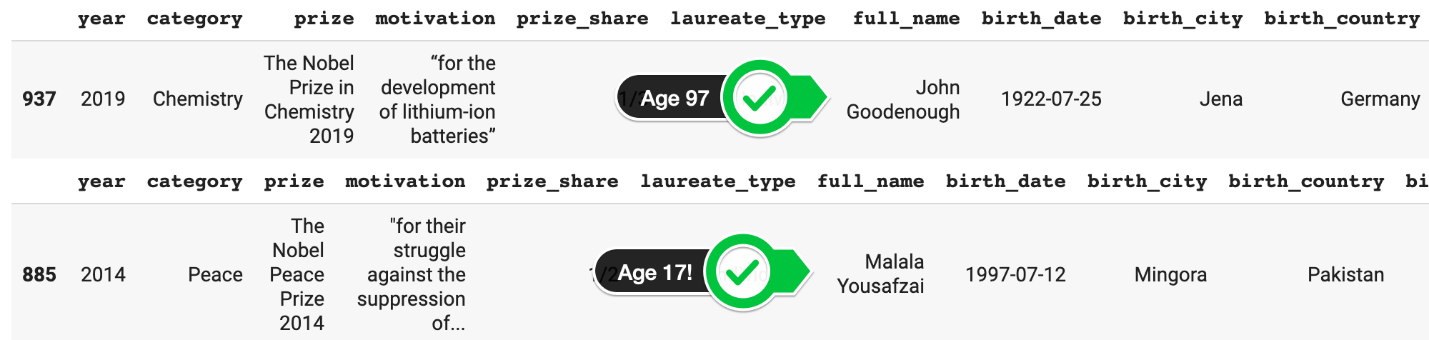
1. birth\_years = df\_data.birth\_date.dt.year

Now we can work out the age at the time of the award:

1. df\_data['winning\_age'] = df\_data.year - birth\_years

**Solution 2: Oldest and Youngest Winners**

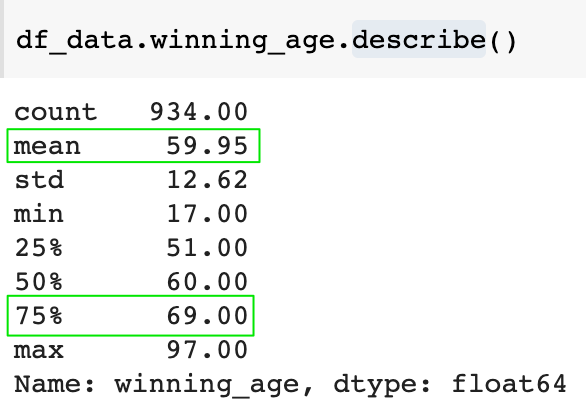
1. display(df\_data.nlargest(n=1, columns='winning\_age'))
2. display(df\_data.nsmallest(n=1, columns='winning\_age'))



John Goodenough was 97 years old when he got the Nobel prize!!! Holy moly. Interestingly John was born to American parents while they were in Germany. This is one example where our analysis of countries counts an extra "German" prize even though he is an American citizen. Too bad we don't have a nationality column in our dataset! Nonetheless, this goes to show it is never too late to win a Nobel prize. I'm keeping my fingers crossed for you!

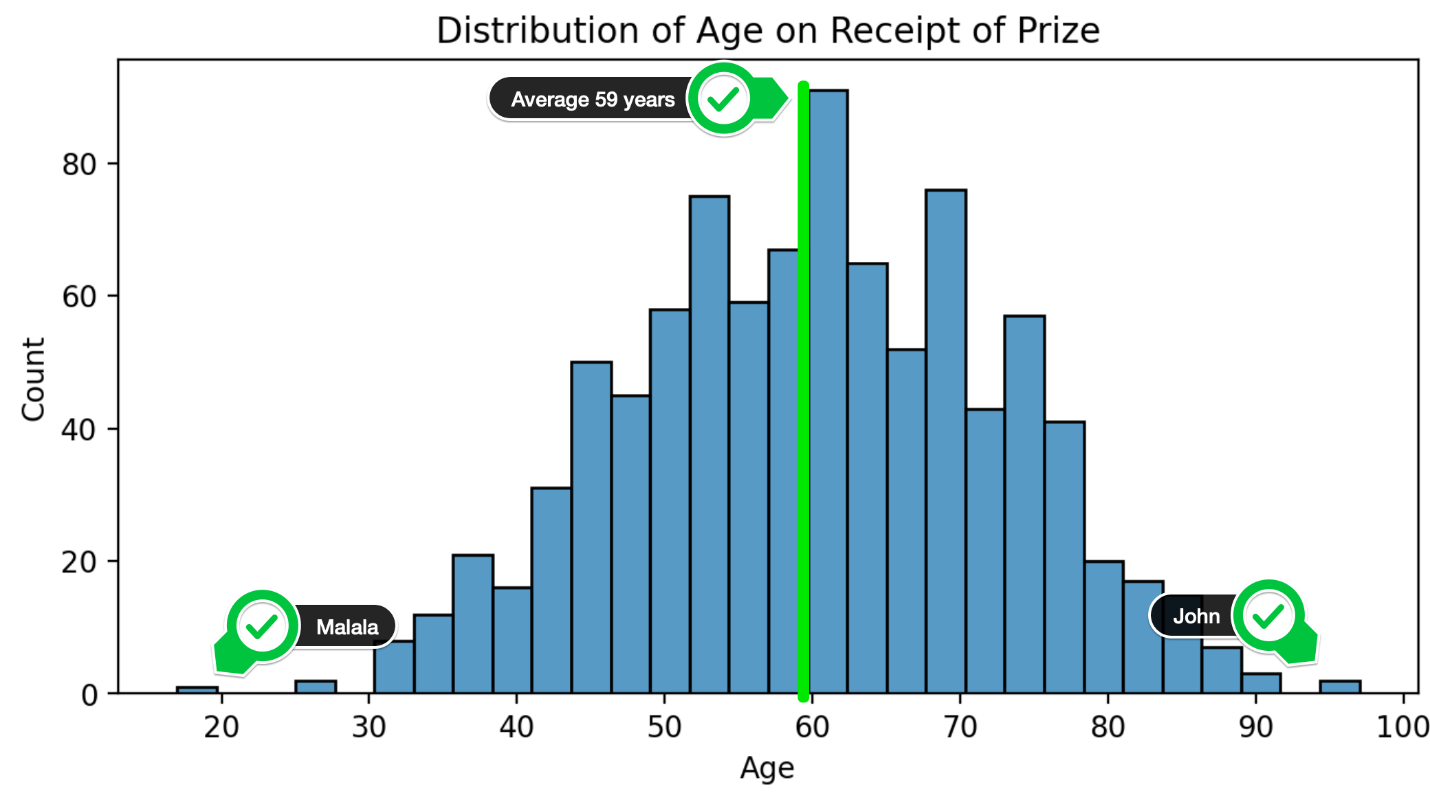
**Solution 3: Descriptive Statistics and Histogram**

Using .describe() is a fantastic way to get a feeling for how the numbers are distributed in a particular column. However, actually visualising them on a histogram to see their distribution is highly recommended too since it allows us to see if we have a bell-shaped curve or something else.



Here's what the histogram looks like:

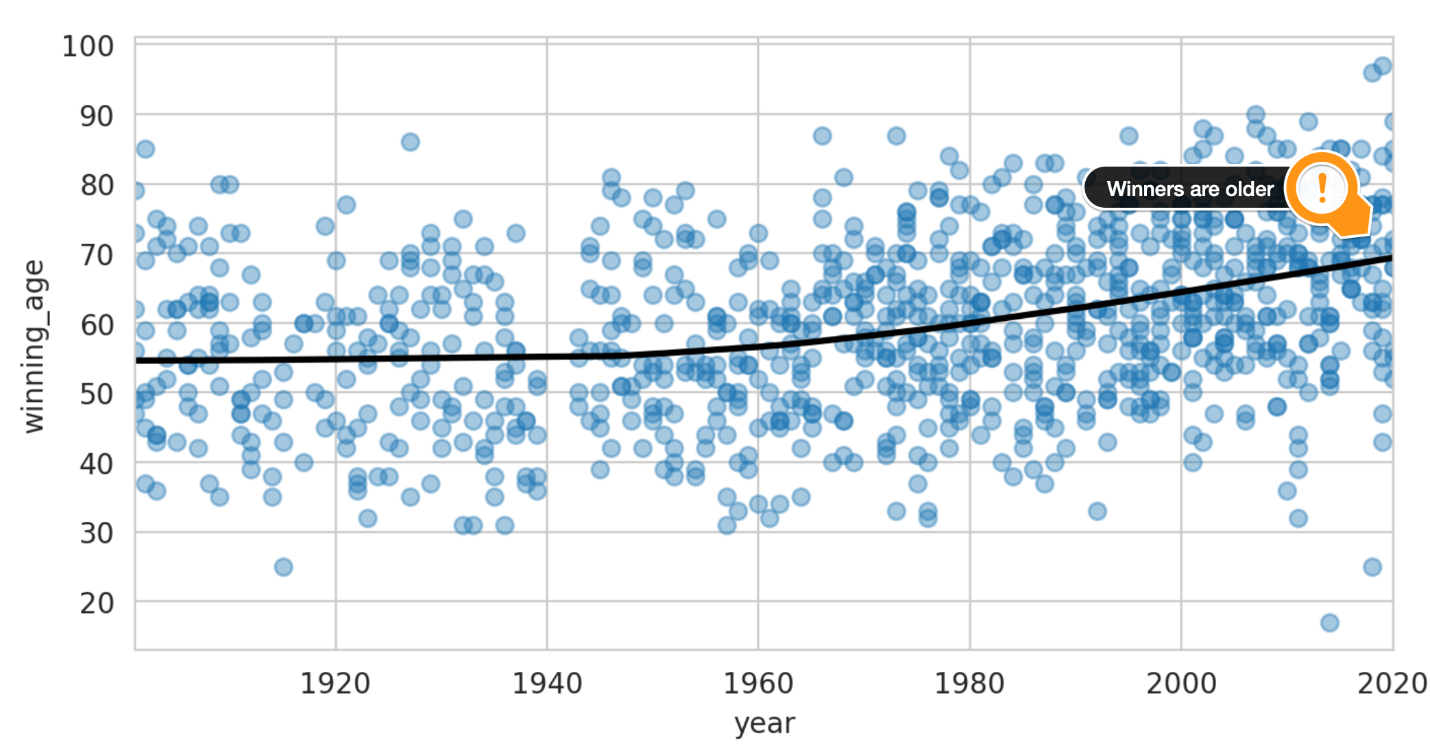
1. plt.figure(figsize=(8, 4), dpi=200)
2. sns.histplot(data=df\_data,
3. x=df\_data.winning\_age,
4. bins=30)
5. plt.xlabel('Age')
6. plt.title('Distribution of Age on Receipt of Prize')
7. plt.show()



**Solution 4: Winning Age Over Time (All Categories)**

The histogram above shows us the distribution across the entire dataset, over the entire time period. But perhaps the age has changed over time.

1. plt.figure(figsize=(8,4), dpi=200)
2. with sns.axes\_style("whitegrid"):
3. sns.regplot(data=df\_data,
4. x='year',
5. y='winning\_age',
6. lowess=True,
7. scatter\_kws = {'alpha': 0.4},
8. line\_kws={'color': 'black'})
10. plt.show()



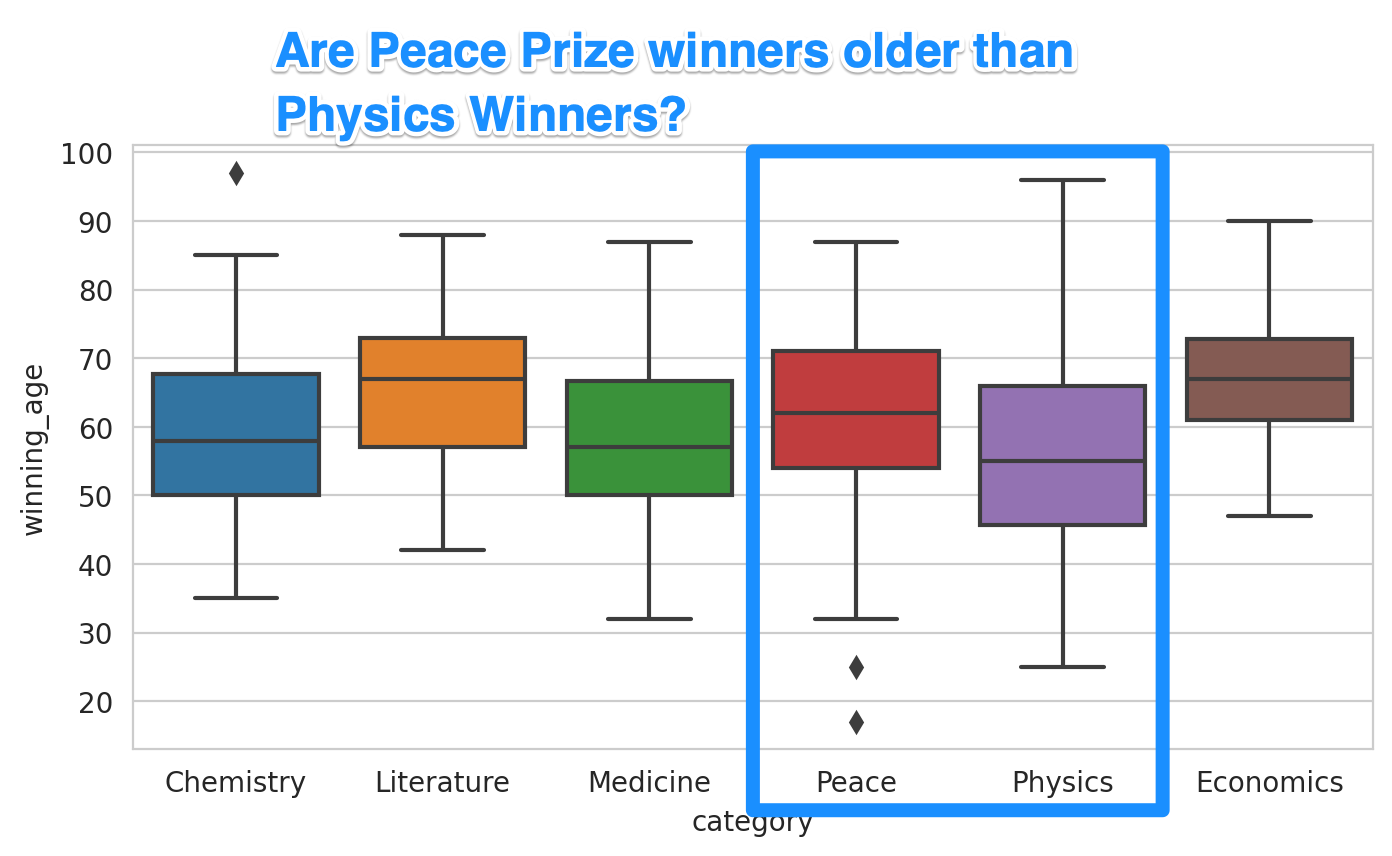
Using the lowess parameter allows us to plot a local linear regression. This means the best fit line is still linear, but it's more like a moving average which gives us a non-linear shape across the entire series. This is super neat because it clearly shows how the Nobel laureates are getting their award later and later in life. From 1900 to around 1950, the laureates were around 55 years old, but these days they are closer to 70 years old when they get their award! The other thing that we see in the chart is that in the last 10 years the spread has increased. We've had more very young and very old winners. In 1950s/60s winners were between 30 and 80 years old. Lately, that range has widened.

**Solution 5: Age Differences between Categories**

Seaborn allows us to create the above chart by category. But first, let's look at a box plot by category.

1. plt.figure(figsize=(8,4), dpi=200)
2. with sns.axes\_style("whitegrid"):
3. sns.boxplot(data=df\_data,
4. x='category',
5. y='winning\_age')
7. plt.show()

The box plot shows us the mean, the quartiles, the maximum and the minimum values. It raises an interesting question: "Are peace prize winners really older than physics laureates?".

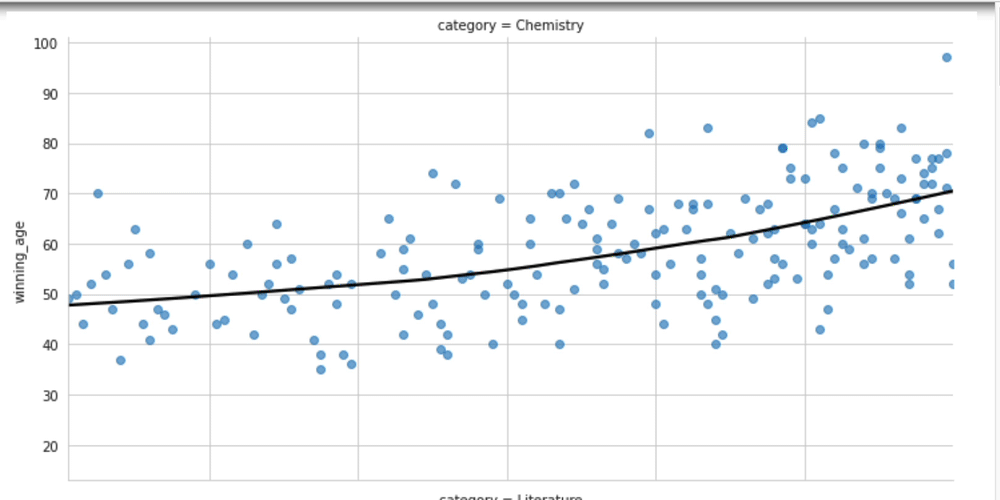


**Solution 6: Laureate Age over Time by Category**

To get a more complete picture, we should look at how the age of winners has changed over time. The box plot above looked at the dataset as a whole.

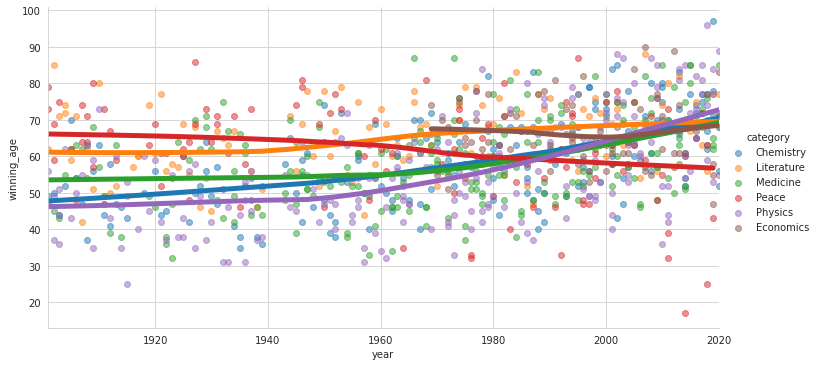
1. with sns.axes\_style('whitegrid'):
2. sns.lmplot(data=df\_data,
3. x='year',
4. y='winning\_age',
5. row = 'category',
6. lowess=True,
7. aspect=2,
8. scatter\_kws = {'alpha': 0.6},
9. line\_kws = {'color': 'black'},)
11. plt.show()

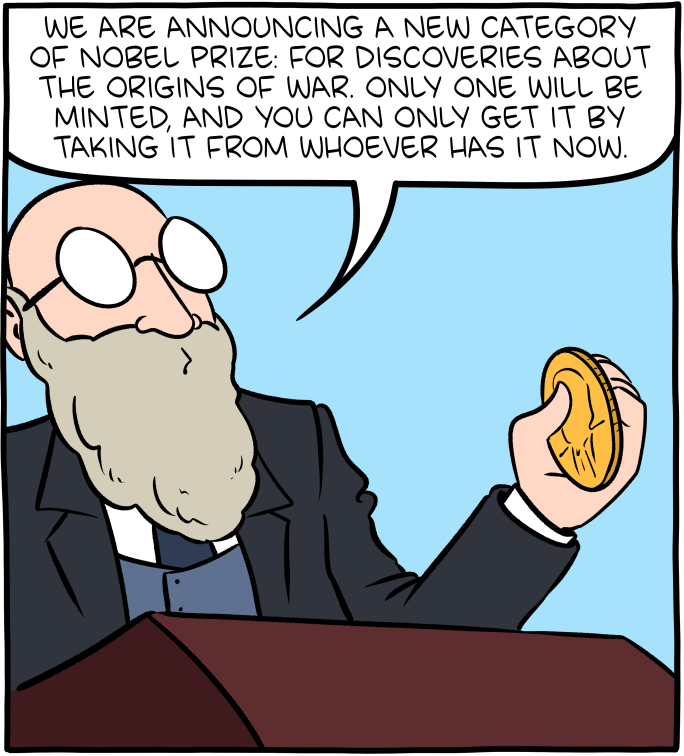
We see that winners in physics, chemistry, and medicine have gotten older over time. The ageing trend is strongest for physics. The average age used to be below 50, but now it's over 70. Economics, the newest category, is much more stable in comparison. The peace prize shows the opposite trend where winners are getting younger! As such, our scatter plots showing the best fit lines over time and our box plot of the entire dataset can tell very different stories!



To combine all these charts into the same chart, we can use the hue parameter

1. with sns.axes\_style("whitegrid"):
2. sns.lmplot(data=df\_data,
3. x='year',
4. y='winning\_age',
5. hue='category',
6. lowess=True,
7. aspect=2,
8. scatter\_kws={'alpha': 0.5},
9. line\_kws={'linewidth': 5})
11. plt.show()





Source: [smbc-comics.com](https://www.smbc-comics.com/comic/prize)