**plotly Bar & Donut Charts: Analyse Prize Categories & Women Winning Prizes**

Challenge 1: Come up with 3 Questions

A big part of data science is coming up with questions that you'd like to explore. This is the most difficult aspect to teach in a tutorial because it's completely open-ended and requires some creativity. Often times you will be asking questions of the data, that it actually cannot answer - and that's ok. That's all part of the process of discovery.

Pause here for a moment and think about the kind of data you saw in the columns. Write down at least 3 questions that you'd like to explore as part of this analysis. For example, your question might go like: "What percentage of the Nobel laureates were women?" or "How many prizes were given out in each category". **Practice coming up with a few of your own questions**.

In the upcoming lessons, you might find that we will write the code to answer some of your questions. And if not, your questions make for a great exercise to take this analysis even further.

The challenges below are all based on questions we're going to ask the data:

Challenge 2

Create a [donut chart using plotly](https://plotly.com/python/pie-charts/) which shows how many prizes went to men compared to how many prizes went to women. What percentage of all the prizes went to women?

Challenge 3

* What are the names of the first 3 female Nobel laureates?
* What did the win the prize for?
* What do you see in their birth\_country? Were they part of an organisation?

Challenge 4

Did some people get a Nobel Prize more than once? If so, who were they?

Challenge 5

* In how many categories are prizes awarded?
* Create a plotly bar chart with the number of prizes awarded by category.
* Use the color scale called Aggrnyl to colour the chart, but don't show a color axis.
* Which category has the most number of prizes awarded?
* Which category has the fewest number of prizes awarded?

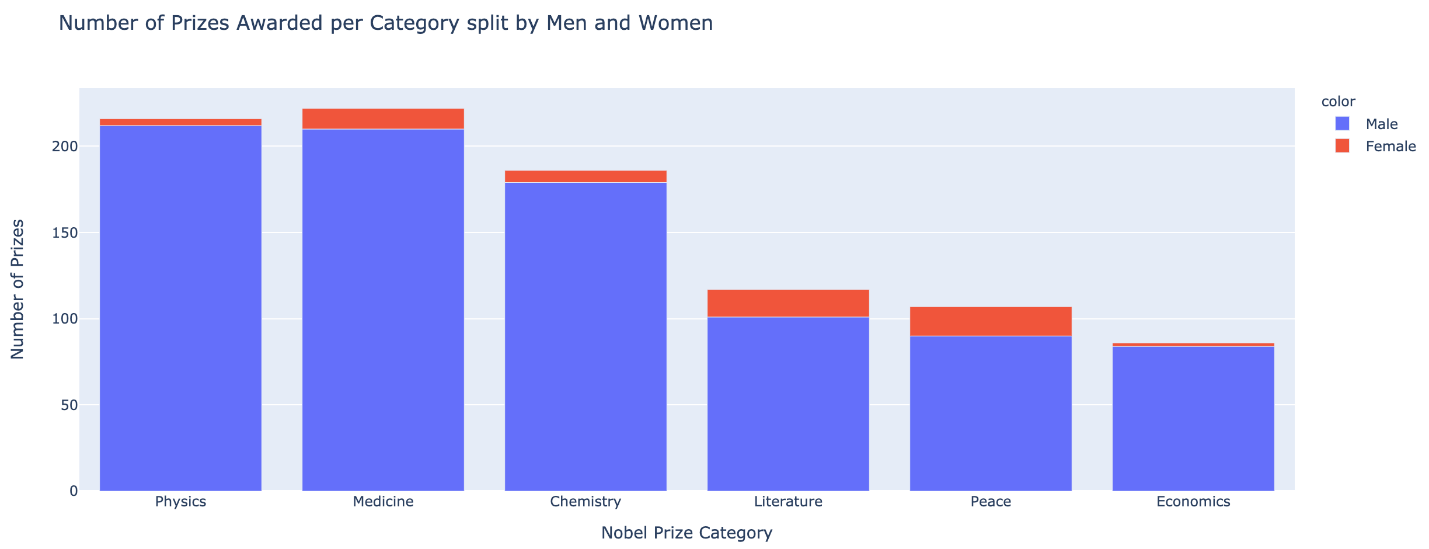
Challenge 6

* When was the first prize in the field of Economics awarded?
* Who did the prize go to?

Challenge 7

Create a [plotly bar chart](https://plotly.com/python/bar-charts/" \t "_blank) that shows the split between men and women by category.

* Hover over the bar chart. How many prizes went to women in Literature compared to Physics?



.

.

..

...

..

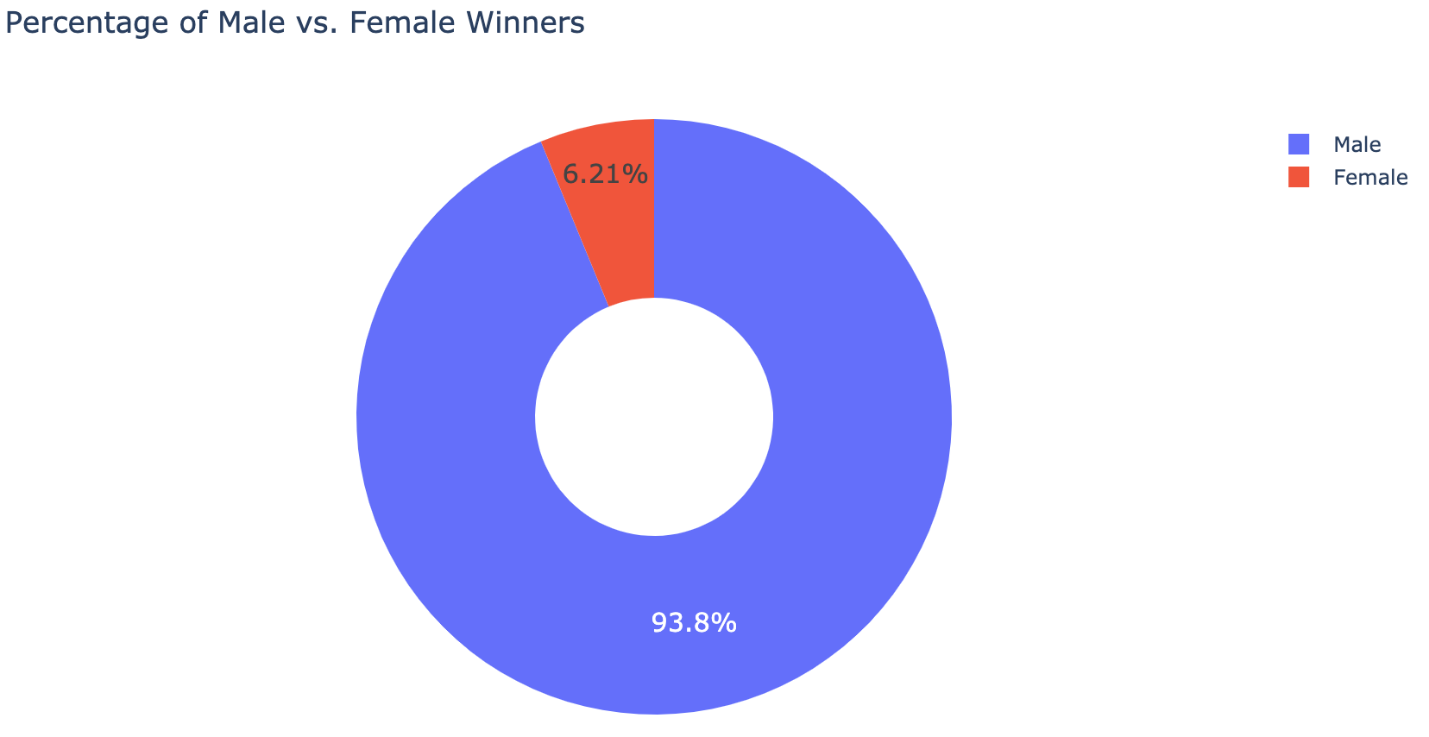
.

.

**Solution 2: Creating a Donut Chart with Plotly**

To create the chart we use the our .value\_counts() method together with plotly's .pie() function. We see that out of all the Nobel laureates since 1901, only about 6.2% were women.

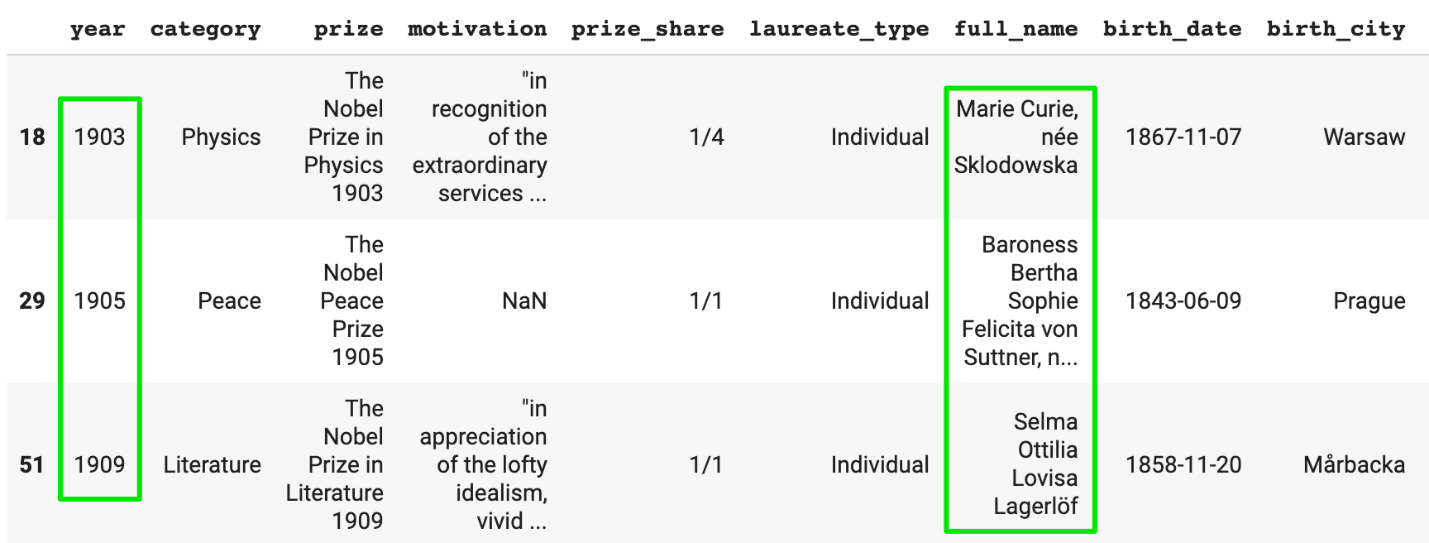
1. biology = df\_data.sex.value\_counts()
2. fig = px.pie(labels=biology.index,
3. values=biology.values,
4. title="Percentage of Male vs. Female Winners",
5. names=biology.index,
6. hole=0.4,)
8. fig.update\_traces(textposition='inside', textfont\_size=15, textinfo='percent')
10. fig.show()



**Solution 3: The first 3 women to win**

Even without looking at the data, you might have already guessed one of the famous names: Marie Curie.

1. df\_data[df\_data.sex == 'Female'].sort\_values('year', ascending=True)[:3]



**Solution 4: The Repeat Winners**

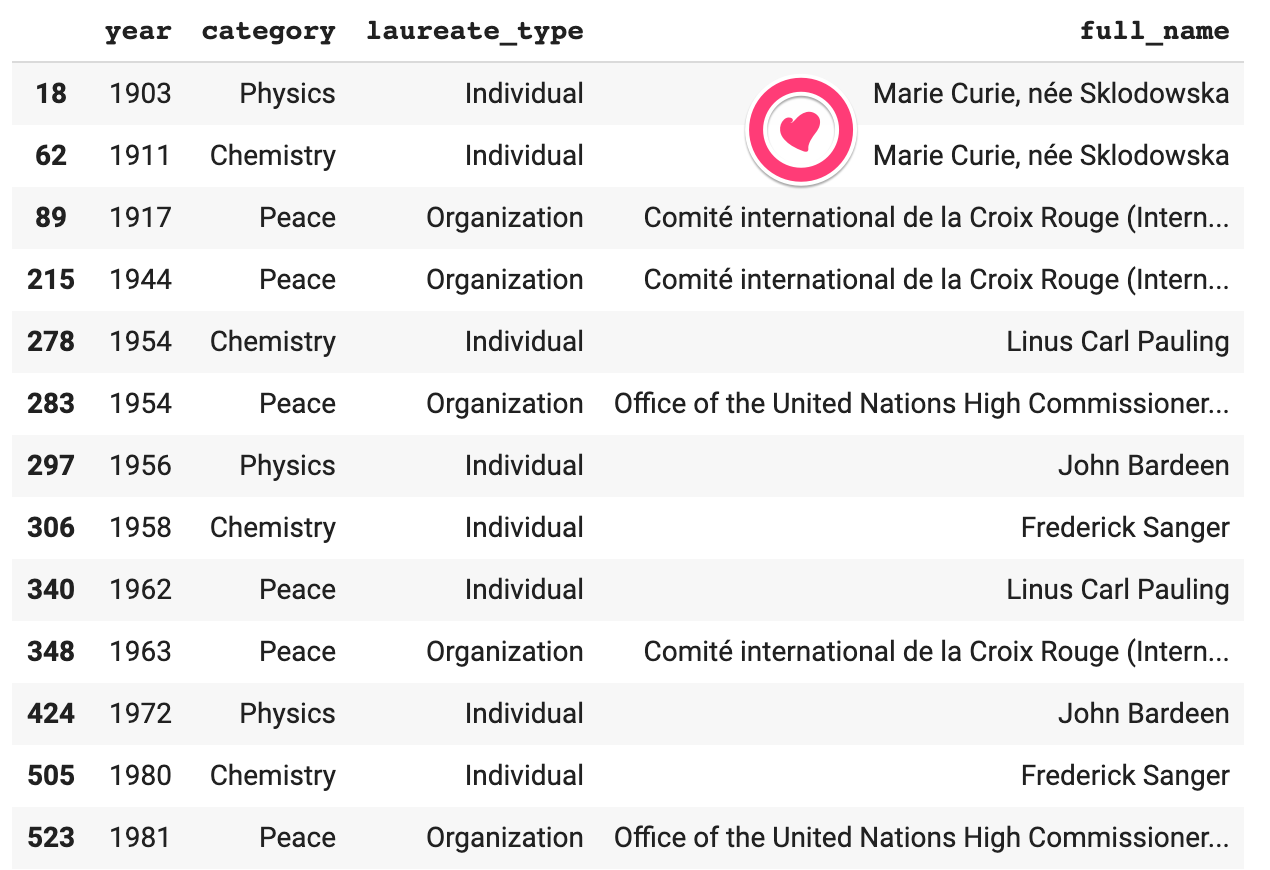
Winning a Nobel prize is quite an achievement. However, some folks have actually won the prize multiple times. To find them, we can use many different approaches. One approach is to look for duplicates in the full\_name column:

1. is\_winner = df\_data.duplicated(subset=['full\_name'], keep=False)
2. multiple\_winners = df\_data[is\_winner]
3. print(f'There are {multiple\_winners.full\_name.nunique()}' \
4. ' winners who were awarded the prize more than once.')

There are 6 winners who were awarded the prize more than once.

1. col\_subset = ['year', 'category', 'laureate\_type', 'full\_name']
2. multiple\_winners[col\_subset]

Only 4 of the repeat laureates were individuals.



We see that Marie Curie actually got the Nobel prize twice - once in physics and once in chemistry. Linus Carl Pauling got it first in chemistry and later for peace given his work in promoting nuclear disarmament. Also, the International Red Cross was awarded the Peace prize a total of 3 times. The first two times were both during the devastating World Wars.

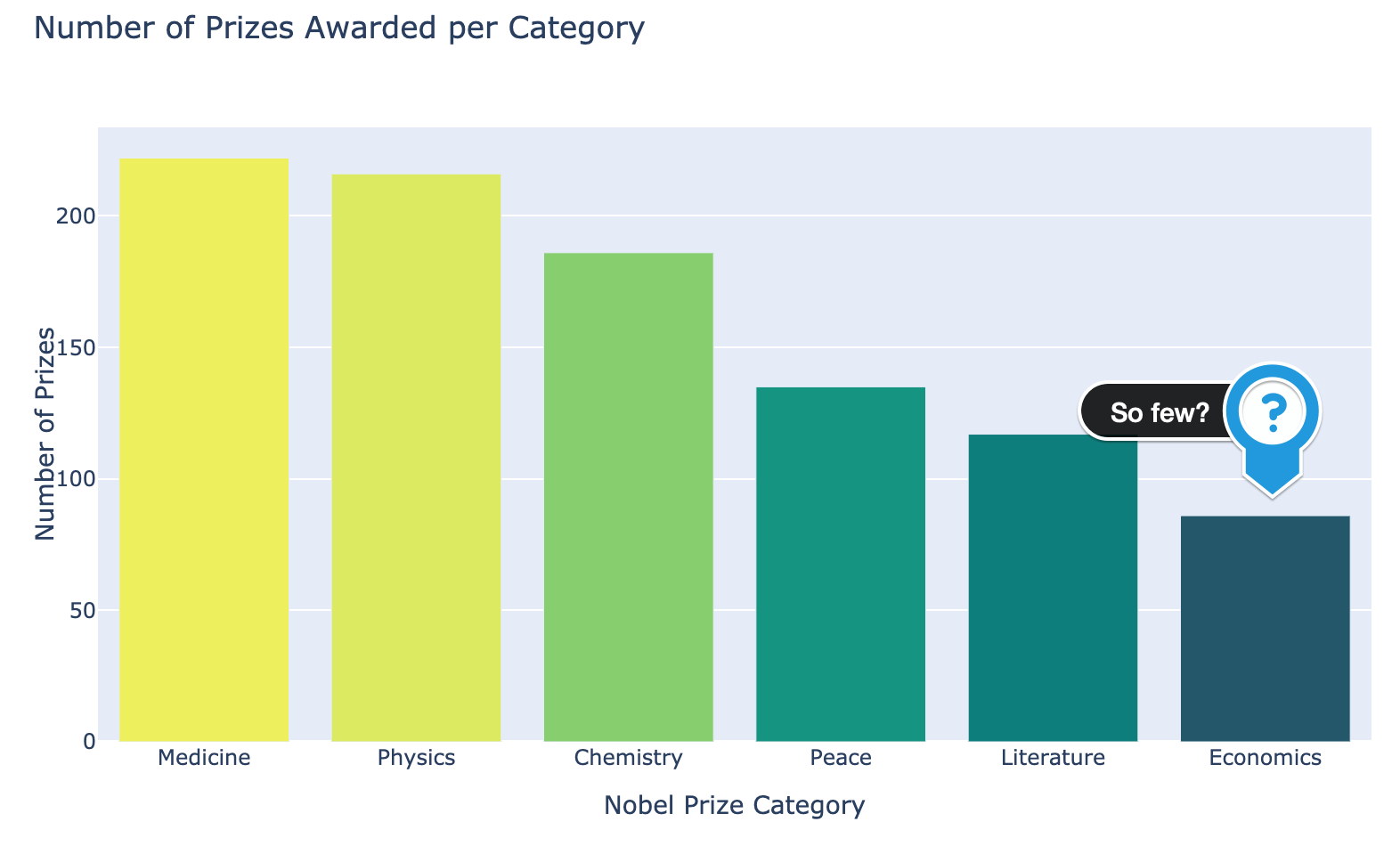
**Solution 5: Number of Prizes per Category**

To find the number of unique categories in a column we can use:

1. df\_data.category.nunique()

To generate the vertical plotly bar chart, we again use .value\_counts():

1. prizes\_per\_category = df\_data.category.value\_counts()
2. v\_bar = px.bar(
3. x = prizes\_per\_category.index,
4. y = prizes\_per\_category.values,
5. color = prizes\_per\_category.values,
6. color\_continuous\_scale='Aggrnyl',
7. title='Number of Prizes Awarded per Category')
9. v\_bar.update\_layout(xaxis\_title='Nobel Prize Category',
10. coloraxis\_showscale=False,
11. yaxis\_title='Number of Prizes')
12. v\_bar.show()

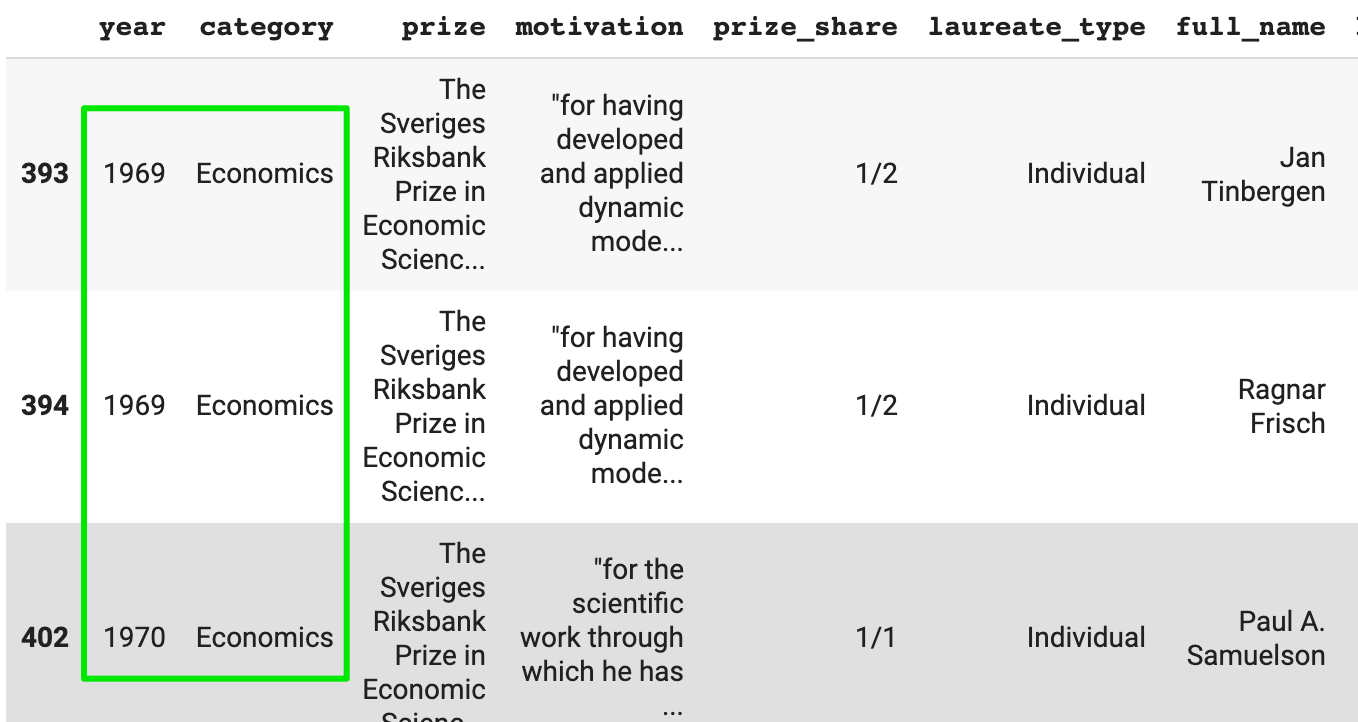


**Solution 6: The Economics Prize**

The chart above begs the question: "Why are there so few prizes in the field of economics?". Looking at the first couple of winners in the economics category, we have our answer:

1. df\_data[df\_data.category == 'Economics'].sort\_values('year')[:3]

The economics prize is much newer. It was first awarded in 1969, compared to 1901 for physics.

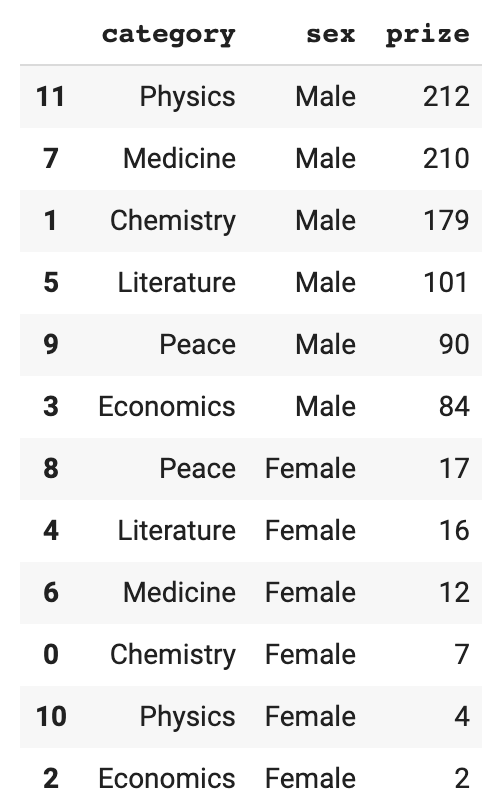


**Solution 7: Male and Female Winners by Category**

We already saw that overall, only 6.2% of Nobel prize winners were female. Does this vary by category?

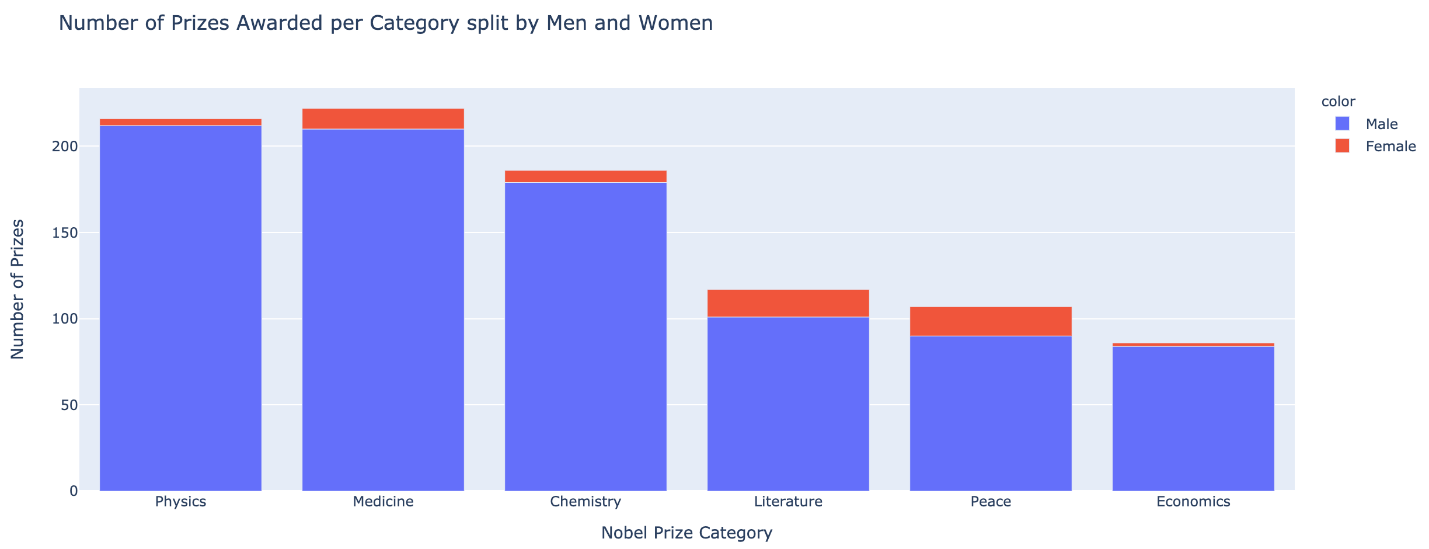
1. cat\_men\_women = df\_data.groupby(['category', 'sex'],
2. as\_index=False).agg({'prize': pd.Series.count})
3. cat\_men\_women.sort\_values('prize', ascending=False, inplace=True)

We can combine .groupby() and .agg() with the .count() function. This way we can count the number of men and women by prize category.



We can then use .color the parameter in the .bar() function to mark the number of men and women on the chart:

1. v\_bar\_split = px.bar(x = cat\_men\_women.category,
2. y = cat\_men\_women.prize,
3. color = cat\_men\_women.sex,
4. title='Number of Prizes Awarded per Category split by Men and Women')
6. v\_bar\_split.update\_layout(xaxis\_title='Nobel Prize Category',
7. yaxis\_title='Number of Prizes')
8. v\_bar\_split.show()



We see that overall the imbalance is pretty large with physics, economics, and chemistry. Women are somewhat more represented in categories of Medicine, Literature and Peace. Splitting bar charts like this is an incredibly powerful way to show a more granular picture.