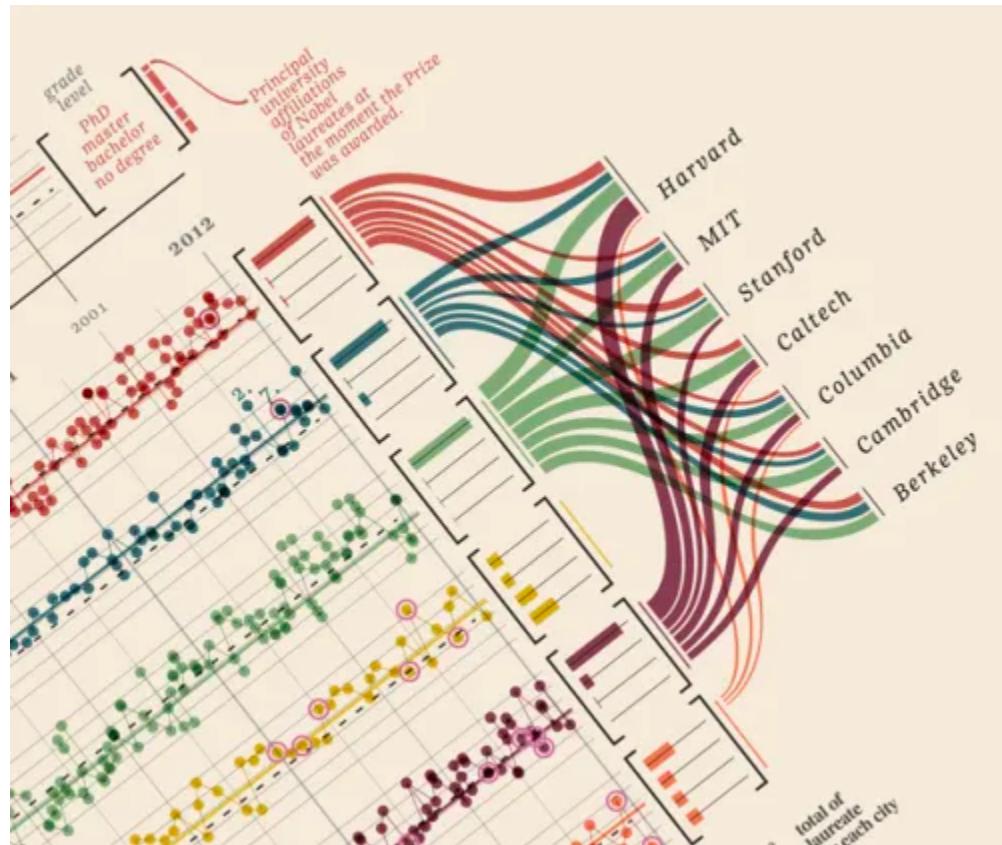


I590 - DATA VISUALIZATION

VISUALIZING NOBEL PRIZE WINNERS

Final Project Paper



[Source: <https://www.themarginalian.org/>]

By Vishwas Shivakumar, Saurav Subash Prasad, Karan Gupta

Abstract:

In this proposal, we have highlighted our motivation to work on the topic of the Nobel Prize. We have raised some questions that we intend to answer through effective visualizations by the end of the course. We have used a Kaggle dataset which has all the required columns, but we will have to add some recent data using the Nobel prize API. We have also done some research on the existing work and have provided our critique on some of the existing visualizations.

Introduction:

The Nobel Prizes are a set of annual international awards given for landmark achievements or discoveries made during the preceding year. They are widely regarded as the most prestigious honors in the fields of literature, medicine, physics, chemistry, peace, and economics. A separate committee awards each prize, but the processes are similar. About 3000 people, usually academics, have the right to nominate candidates. From these, the Nobel committees choose around 300 potential recipients and prepare a report reflecting the advice of experts. The prize awarding institutions then select the laureates by a majority vote. There is a maximum of three laureates per award, except for the peace prize which can be awarded to institutions. [1]

Motivation:

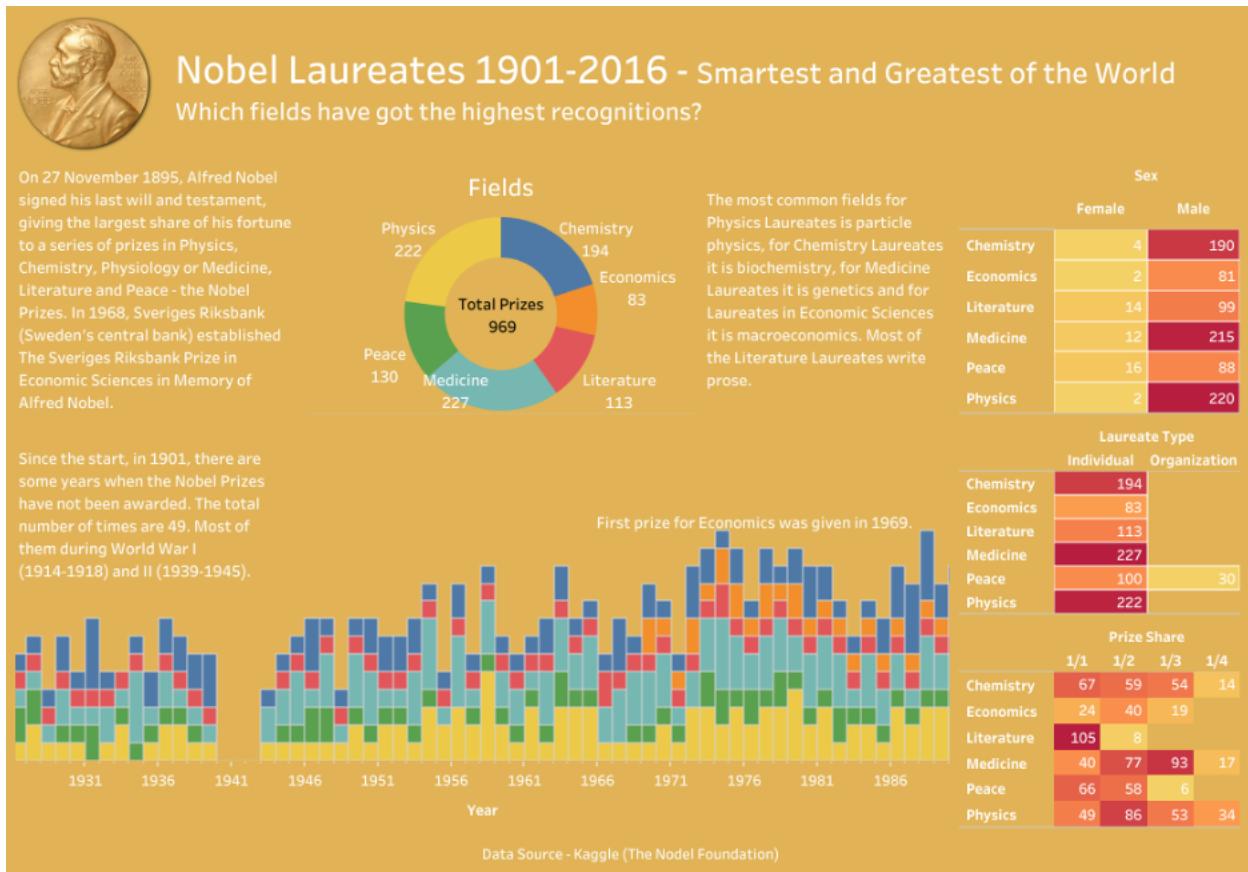
Recently, Philip H. Dybvig, an Indiana University Alumnus was awarded the Nobel Prize for Economics for his research on banks and the financial crisis. Over the course of Indiana University's history, ten individuals with ties to the university have been awarded a Nobel Prize. [2] Currently, Harvard University tops the list with producing the greatest number of Nobel Laureates (161), either as faculties or as graduates. [3] We were astonished to learn that the top

seven universities with the most Nobel Laureates are in the United States of America. [4] The USA has 398 Nobel Laureates, followed by the United Kingdom which has 137, almost one-third to that of the USA. [5] Nobel Prizes started in 1901 and there are a total of 989 Prizes awarded to date. [6] That means that more than 40% of the Nobel Prizes have been awarded to Laureates from the USA.

This motivated us to dig deeper into understanding the distribution of countries of these Nobel Laureates and how this has changed over the years. We are also curious to know whether a country where a person is born affects their chances of the Nobel Prize. These chances could be affected by the university they attend and the education they get. A person born in an underdeveloped or developing country might have a less chance to get this award. But what if a person born in these countries migrates to the developed countries and settle there? This motivated us to find the data for all the Nobel Laureates till date and the countries they were born. We are trying to do some interesting visualizations that explain how countries are doing in each category of the Nobel Prize awards and how that has changed over time.

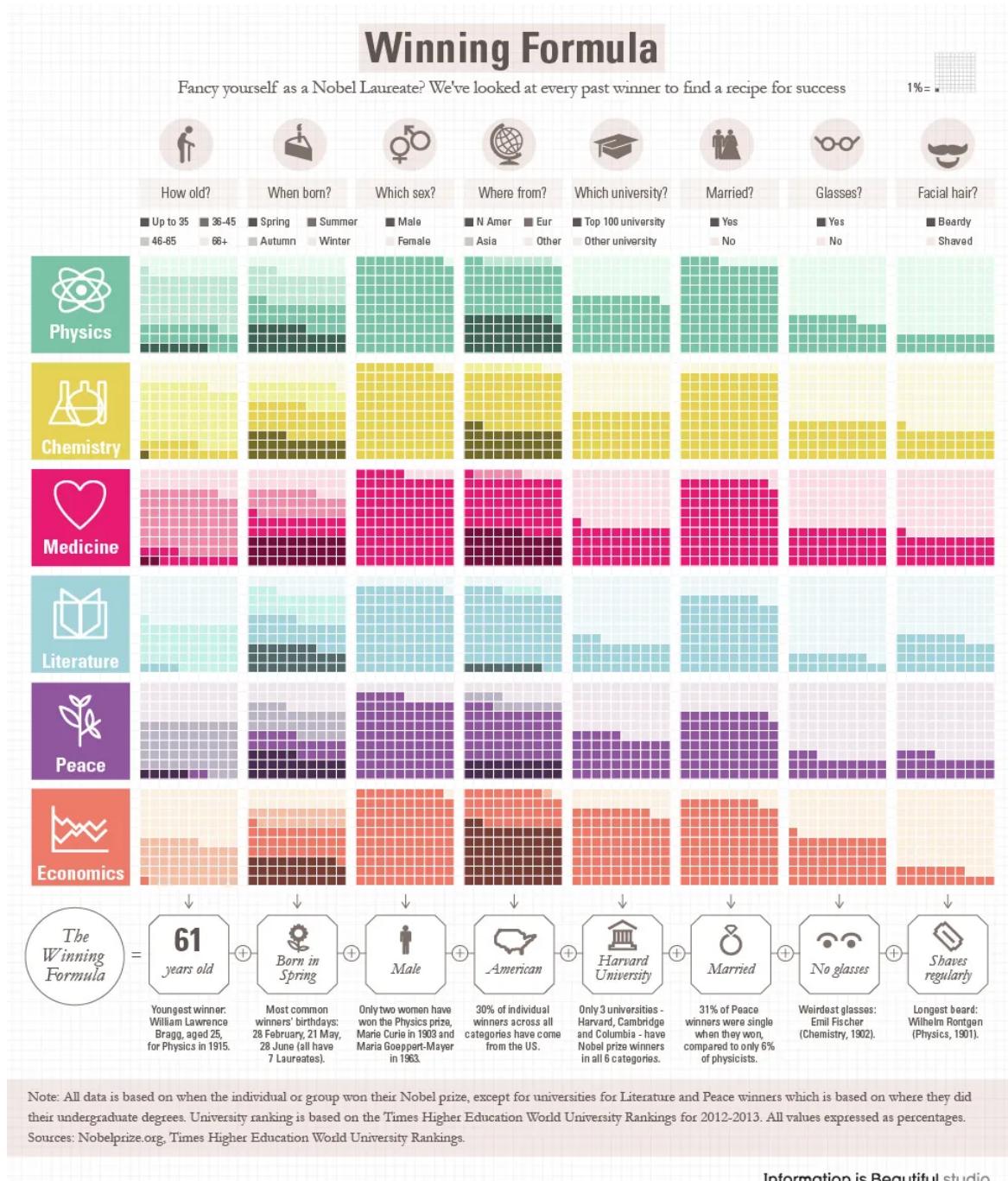
A part of our motivation was also that there is a constant controversy over how universities count the Nobel Prizes being awarded to them. A Nobel Prize awardee is usually an alumnus of multiple universities, and all the universities try to claim the prize, but it is only awarded to the university with which the Nobel Prize winner is currently associated. This motivated us to answer a series of questions which are listed in the ‘objectives’ section later in the report.

Research on Existing Work:

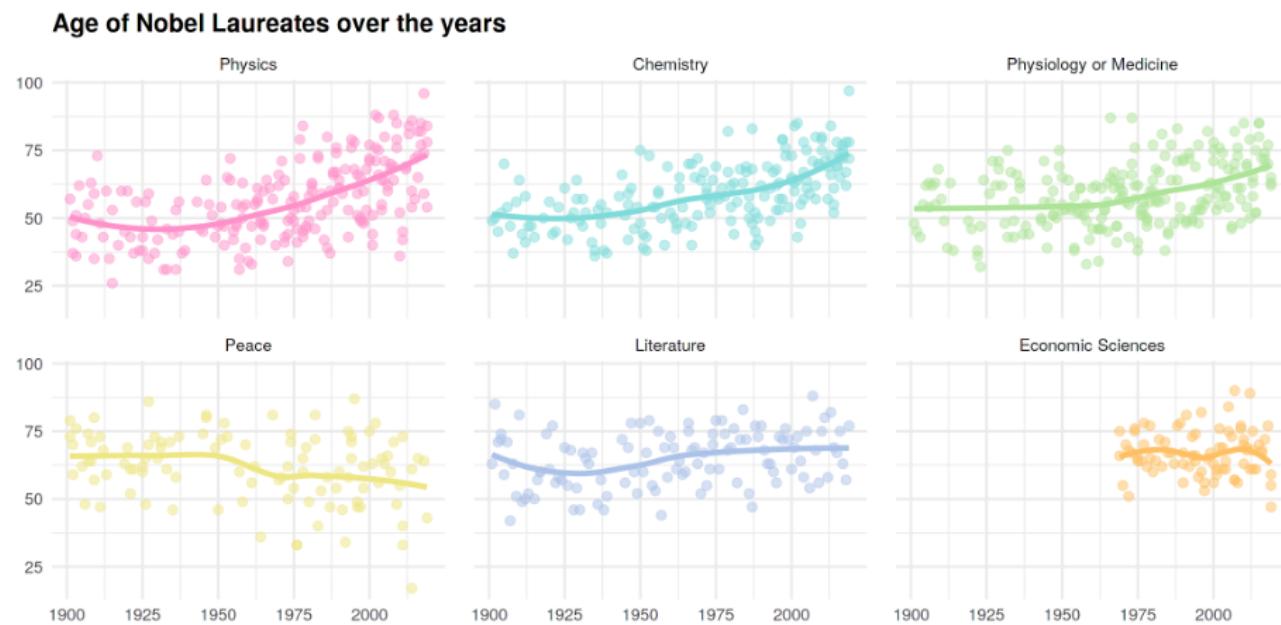


This is a Tableau Dashboard [10] giving us insights on which fields have got the highest recognition. This visualization includes the laureates' gender, as well as the Laureate type and prize share. To indicate the intensity of numbers, it uses colors that transition from yellow to red. Gender discrimination is clearly visible among the winners. Other than the field of Peace, which contains organizations as winners, the Laureate type is usually Individual. It also provides some background information on the Nobel Laureates. The author uses a histogram to plot the number of prizes in each field, which has been color coded. The color selection could have been better because it is not grayscale protected as all the colors are of the same saturation value or nearly so. Since red has

been used extensively, the visualization is not colorblind friendly. The values on the histogram are not labeled, making it difficult for the viewer to determine the exact number of prize winners.



This Visualization [11] is fascinating as it attempts to deduce a winning formula for Nobel Laureates. It includes a plethora of details divided among the six Nobel Prize categories. It includes the following information: age, birth season, gender, birth continent, university, marital status, glasses, and facial hair. The first element is based on the average age of Nobel Laureates, which is 61. This may or may not be related to the likelihood of winning a prize [12]. The most common Birth month is Spring, which is a classic example of showing correlation isn't causation. The month you were born in has absolutely no bearing on your chances of winning the Nobel Prize. Gender is the next component in the formula, and there is a lot of evidence that shows gender discrimination is prevalent in the workplace [13]. The visualization also demonstrates the dominance of the United States among the laureates. There has been a lot of research done to show why the United States dominates the Nobel Prize. [14] [15]. The last three factors are a person's marital status, whether or not they wear glasses, and whether or not they have facial hair. All three are cases where there is a correlation but no causation. Most Nobel laureates are married because they begin their research at a much later age, and there is sufficient research showing the correlation between age and marriage, indicating that there is no direct correlation between marital status and winning the Nobel Prize. [16] [17]



[18] This visualization implements a scatter plot to depict the ages of Nobel laureates in relation to their fields, with different colors used for each. There is also a line graph that shows the trend of ages. This visualization is very effective in conveying the message that Economic Sciences as a Field began much later than the other fields, but the exact year is difficult to deduce from the graph. The colors used could have been better because they are not grayscale protected and would look the same on a grayscale. Also, the colors are too light on a white background, making it difficult to read. The viewer may have difficulty determining the exact age of the Nobel Laureates based on the visualization. 'Alpha' was used to make the dots of the scatter plot more transparent, providing better insights; however, the 'Alpha' value could have been higher, making the plot more visible against the white background.

Objectives / Contribution:

Our motivation to see which university has the right to claim the Nobel Prize brings us to the next question, which is which universities presently hold the most Nobel Prizes? If we answer this question, we would be interested to see the same among countries. We know that the US has the highest number of Nobel Prizes. Has it always been the same? How has that number changed over time? Which nations have historically ranked the highest?

Studying the existing work on visualizing the Nobel Prize winners also inspired us to answer a few more questions, such as:

- What is the breakdown of Nobel Prizes by category? How has that number varied over time?
- What is the gender and country distribution of the Nobel Prizes till date?
- What was the age and gender distribution for every category?

Data & Methods:

Data Sources:

We are utilizing a Kaggle dataset [7]. The data was gathered via the SPARQL Endpoint of the Nobel Prize API [8]. This dataset profiles every Nobel Prize recipient since the award's introduction in 1901 through 2016. The event takes place every year on December 10th, and winners are revealed at the start of October. The Nobel Prize was established in 1895 by Swedish inventor Alfred Nobel, and the first award was made in 1901. It recognizes luminaries in the domains of literature, peace, biology, chemistry, physics, and economics. The really great thing about this dataset is it is adequate for our project it contains all the necessary data such as, age, sex, city of birth, motivation, and a few others. It also contains a clearly defined data dictionary, data types for each column, and a short, well-documented data lineage.

	df.sample(5)																		
:>	index	id	firstname	surname	born	died	bornCountry	bornCountryCode	bornCity	diedCountry	...	diedCity	gender	year	category	overallMotivation	share	mot	
	198	198	202	Otto	Hahn	03/08/1879	7/28/1968	Germany	DE	Frankfurt-on-the-Main	West Germany (now Germany)	G_ttingen	male	1944.0	chemistry	NaN	1.0	"discusses the first heavy metals"	
	612	612	372	Francis Harry Compton	Crick	6/8/1916	7/28/2004	United Kingdom	GB	Northampton	USA	San Diego, CA	male	1962.0	medicine	NaN	3.0	"discusses the first communication methods"	
	561	561	361	Werner Forssmann		8/29/1904	6/1/1979	Germany	DE	Berlin	West Germany (now Germany)	Schopfheim	male	1956.0	medicine	NaN	3.0	"discusses the first communication methods"	
	819	819	169	Otto Wallach		03/27/1847	2/26/1931	Germany (now Russia)	RU	Koenigsberg (now Kaliningrad)	Germany	G_ttingen	male	1910.0	chemistry	NaN	1.0	"receives services"	

This screenshot shows how our data looks , this shows all the nobel prize winners and all the details related to them

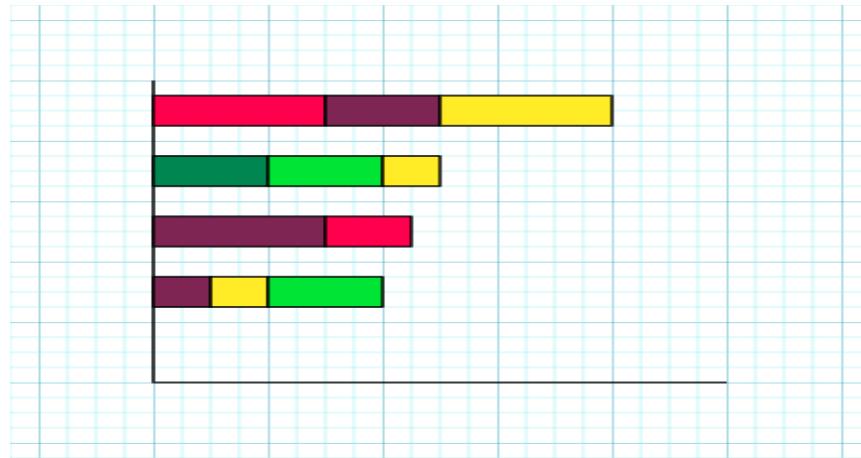
We lack the data after 2016 that can be obtained using the Nobel Prize API version 2 [8], which makes it simple to obtain the data in csv or JASON format. We will also use the information from the "Nobel Prize Facts" [9] website to create some original and fascinating visualizations on

"Family Nobel Prize laureates" and Nobel laureates who were imprisoned at the time of their award.

Ideas & Sketches along with Visualization Method Selection:

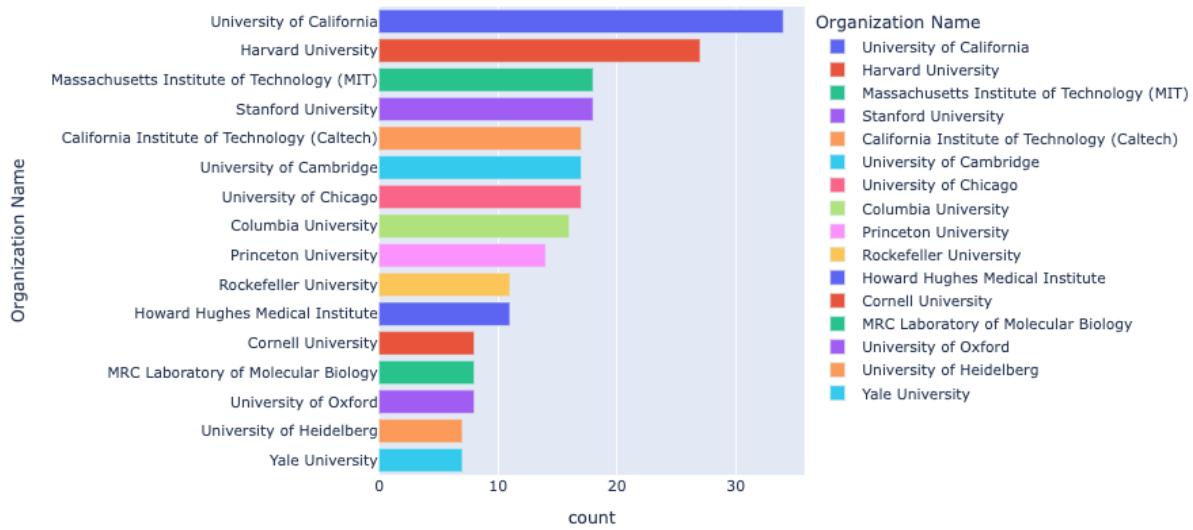
Visualization: Which university has the most Nobel awards?

Plotting a bar graph with the count of prizes (across all categories) on the x-axis and the schools on the y-axis was our initial plan to visualize the university with the most Nobel laureates. We picked a horizontal bar graph since we wanted to display the data for the top 15 universities. But we soon understood that the problem with this would be that we couldn't see how many prizes the college had received in each category. In order to solve this problem, we will create an interactive horizontal stacked bar graph that displays the number of awards received in each category. Below is a basic sketch of the same.



Method Selection:

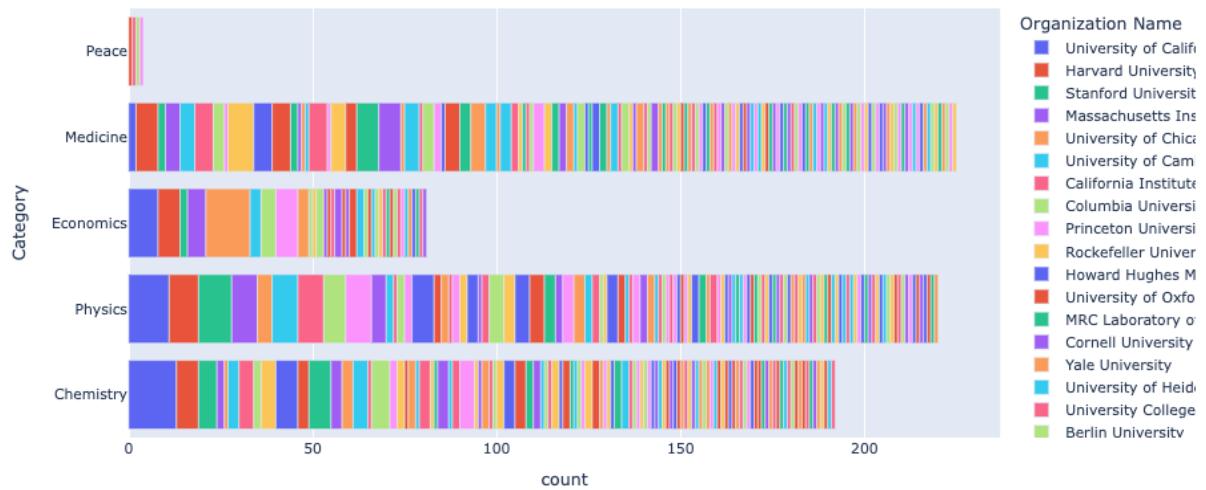
We decided to display the Universities with the most noble medals using a horizontal stacked bar graph, as was described in our visualization design for this section. This was caused by a multitude of factors; if we had merely plotted the bar graph, we would have known only how many awards each university had received.



Initial plot with no info on categories

We built a horizontal stacked plot with the categories on the y axis, the number of awards on the x axis, and each stack in the bar representing a university in order to include the category data in the charts. This was hardly the plot we were searching for, as we quickly understood since we received no information to help us with our analysis.

We choose the plot below to address these problems. It is a horizontally stacked bar plot with the number of prizes and the universities in the y-axis, each stack representing a separate category. Additionally, this is interactive, allowing us to see the specifics when we hover over the bars.



Visualization: US domination in Nobel Awards over the decades?

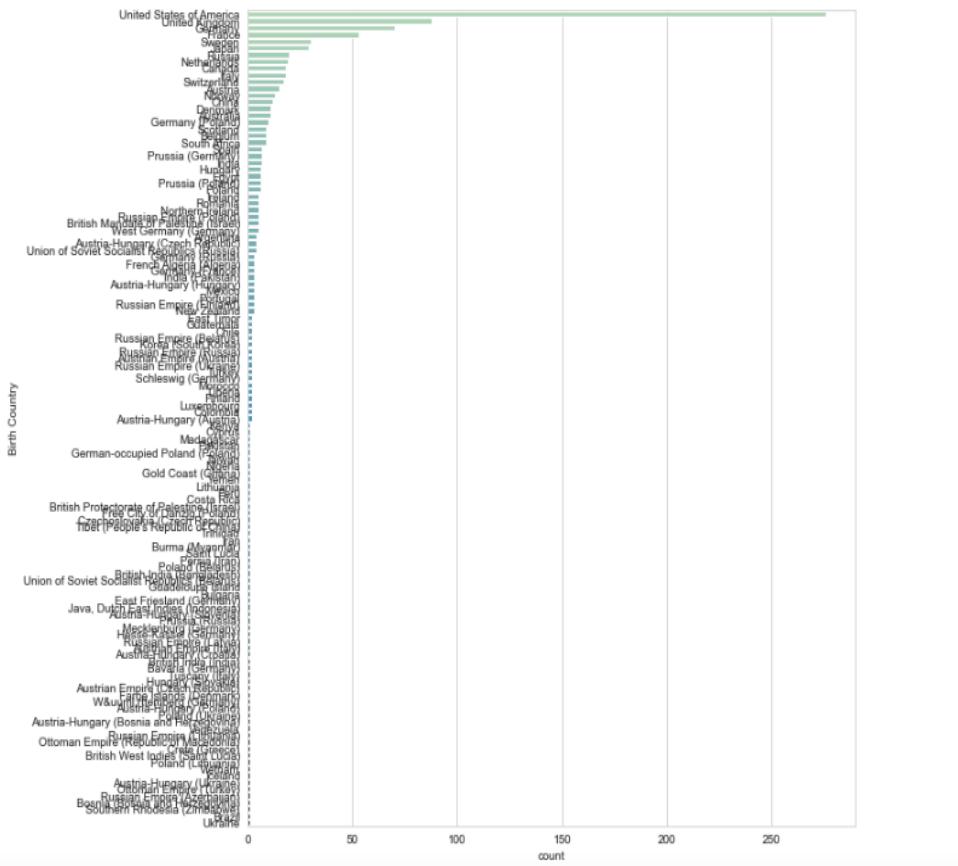
We decided that a choropleth map would be the best approach to illustrate the US' dominance of the Nobel Prizes because a bar graph (countries versus number of prizes) would make it difficult to view the data for a specific country. When we hover over a country of interest on an interactive choropleth map, all the facts for that nation are displayed.

Here, we'll utilize a colormap to display the number of awards each nation has received across the globe. The user will be able to select a decade and view the distribution of the awards by country for that decade by using a slider we'll provide underneath the map.



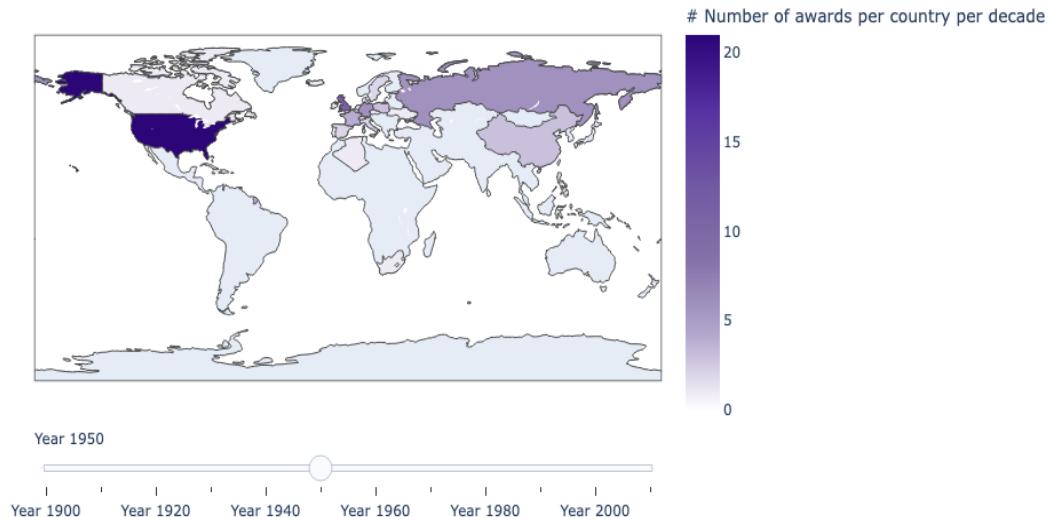
Method Selection:

Our initial idea was to represent this with a bar graph. The issue with this, though, is that there will be a bar for every nation, and since the majority of nations only receive one or two awards, it will be challenging to examine a single nation and see how it develops over time. The first effort at plotting it as a bar graph is shown below.



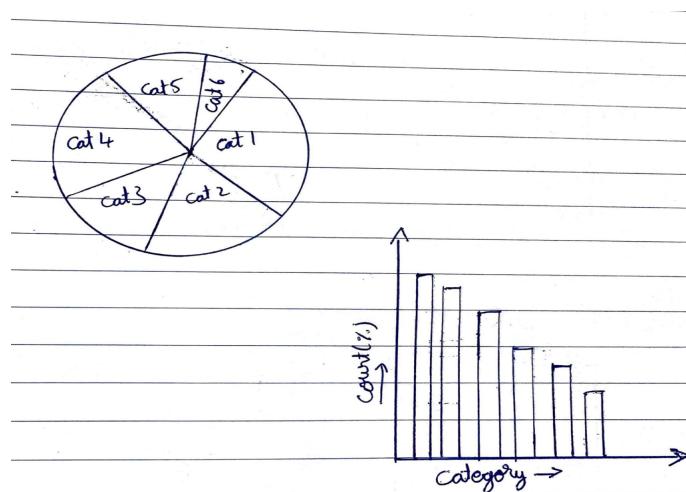
We utilized a choropleth map of the world with a slider to select the decade in order to see the trend through time and to get the answer to our question. Additionally, we made a deliberate choice to leave the scale flexible so that it might adjust based on our choice, showing that the number of awards they give away has grown over the years.

Number of awards per country per decade

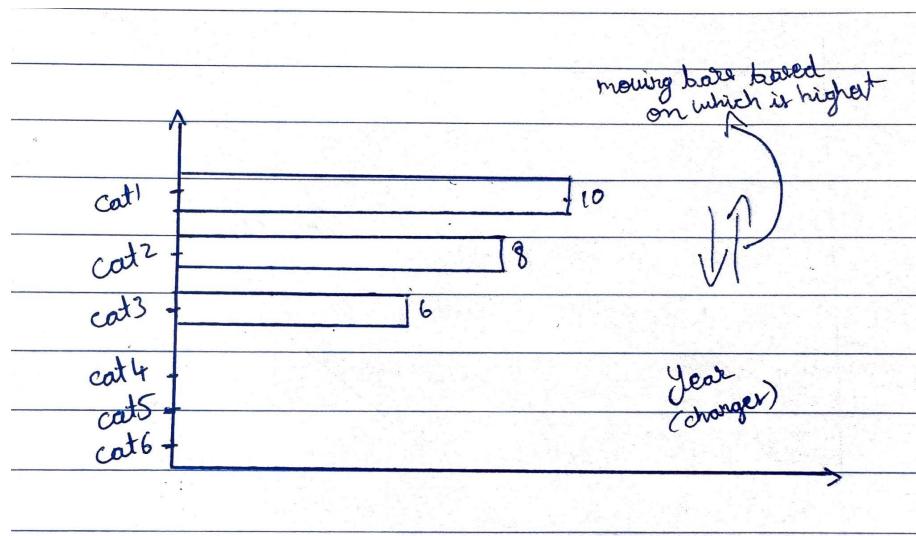


Visualization: Breakdown of Nobel prize by category.

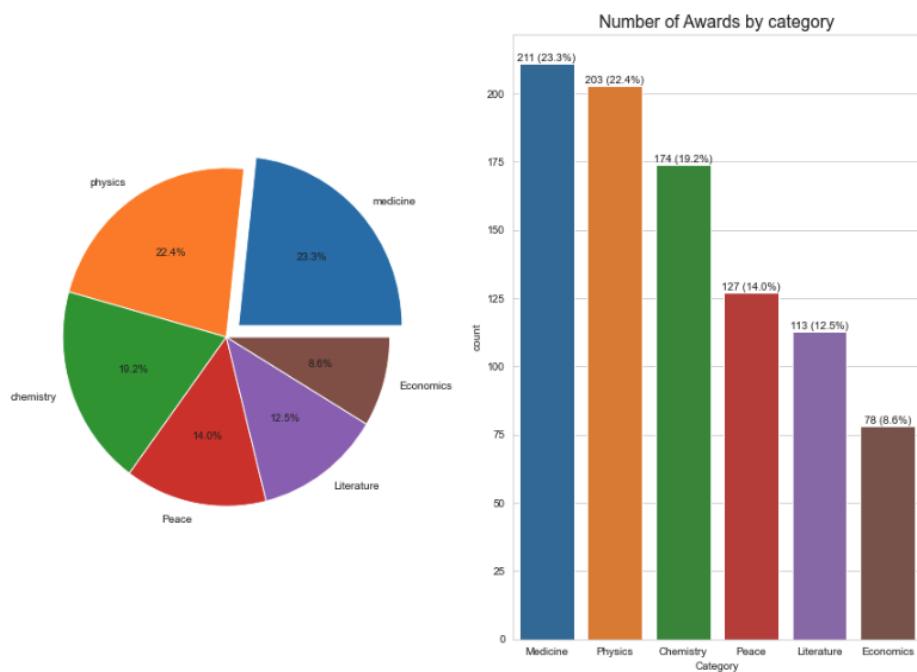
In order to facilitate understanding, we will first create a pie chart showing the proportion of total awards that each category has received. This chart will be supported by a bar graph showing the total number of awards that each category has received together with the percentage value next to it. This could match the illustration below.



We reasoned that these plots by themselves wouldn't provide us with the complete picture, so we also plotted a moving horizontal bar graph to display the statistics of cumulative award distribution among categories over time. Here is a quick drawing of an idea for it.

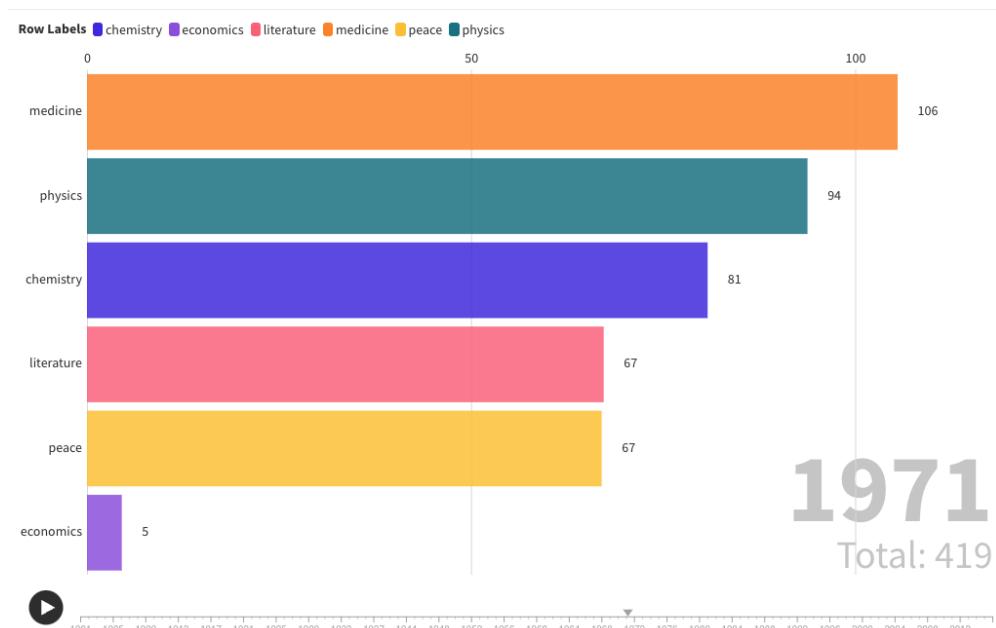


Method Selection:



We constructed a pie chart according to our original concept, showing the percentage distribution of awards given to each category over the years. The absolute values and percentage shares were also shown on a bar graph.

The moving bar graph plot of cumulative data throughout the years is shown below. We choose this moving bar graph because it displays the total number of prizes given over time, which enables us to derive inferences and findings. To do this, the data were categorized by category and year and transformed to Tidy format, which made plotting the data simpler.

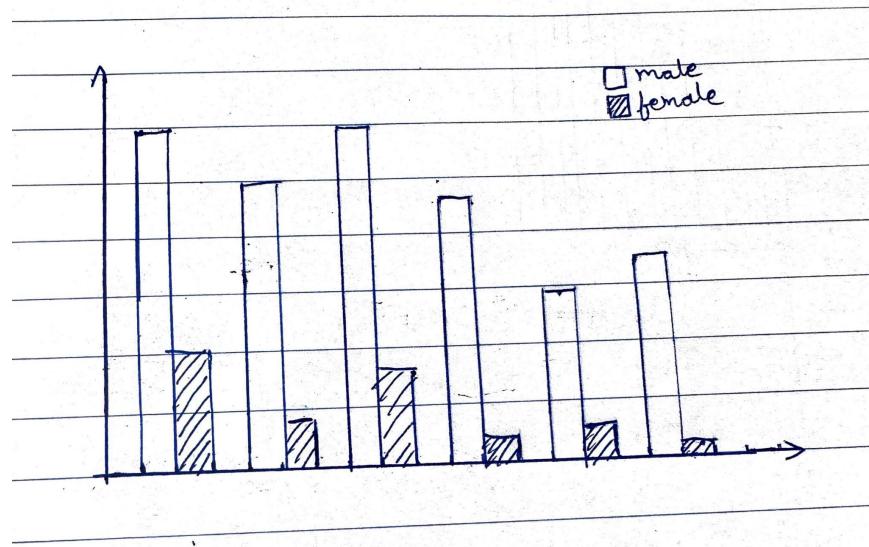


Visualization: Gender and country distribution of Nobel prizes till date

We aimed to highlight the fact that many nations have yet to produce a female Nobel laureate. We used a choropleth map once more, quite similar to the one in the part before, but this time we'll be exhibiting a map with cumulative data rather than data spanning decades. We will produce two choropleth maps, the first of which will display data for all Nobel laureates and the

second of which will only display data for female Nobel laureates. In order to compare the distribution of male and female Nobel laureates among nations, this is done.

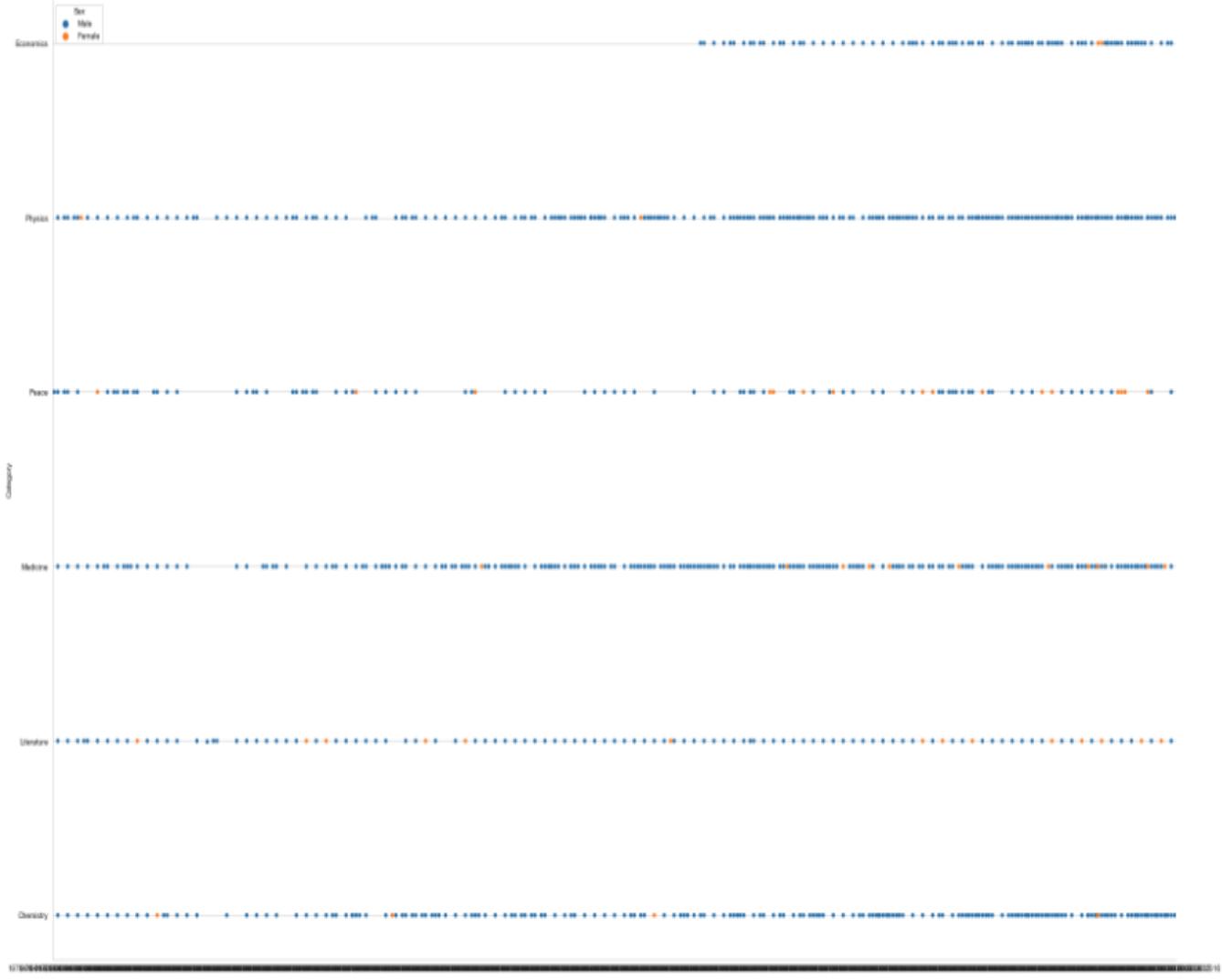
We will draw a multiple-bar bar chart with each bar in a group representing a gender and each group being a category to show the distribution of prizes by gender and by category. The y-axis will display the percentage for each bar as well as the number of awards for each bar. something like the illustration below. To show the distribution of awards in each category by percentages we will plot pie charts for each category.



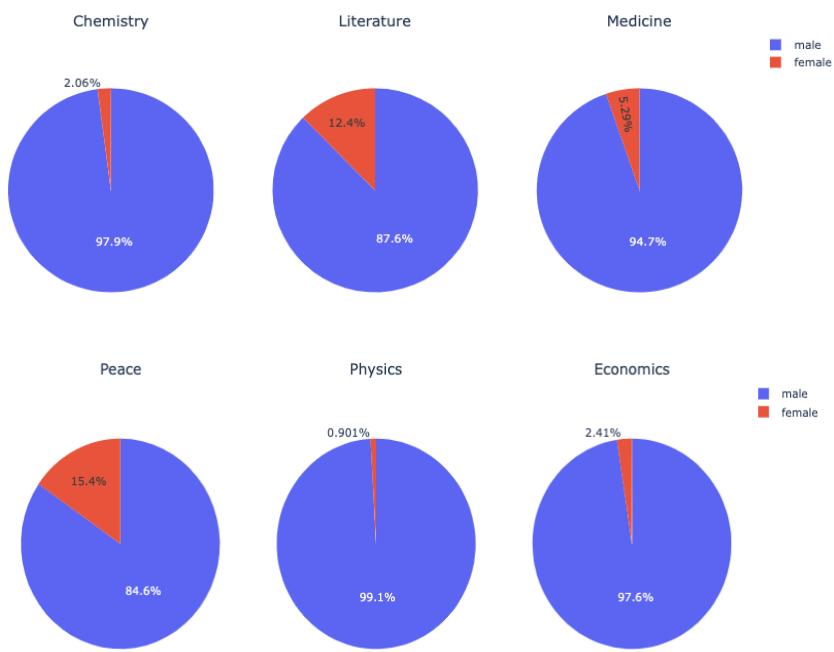
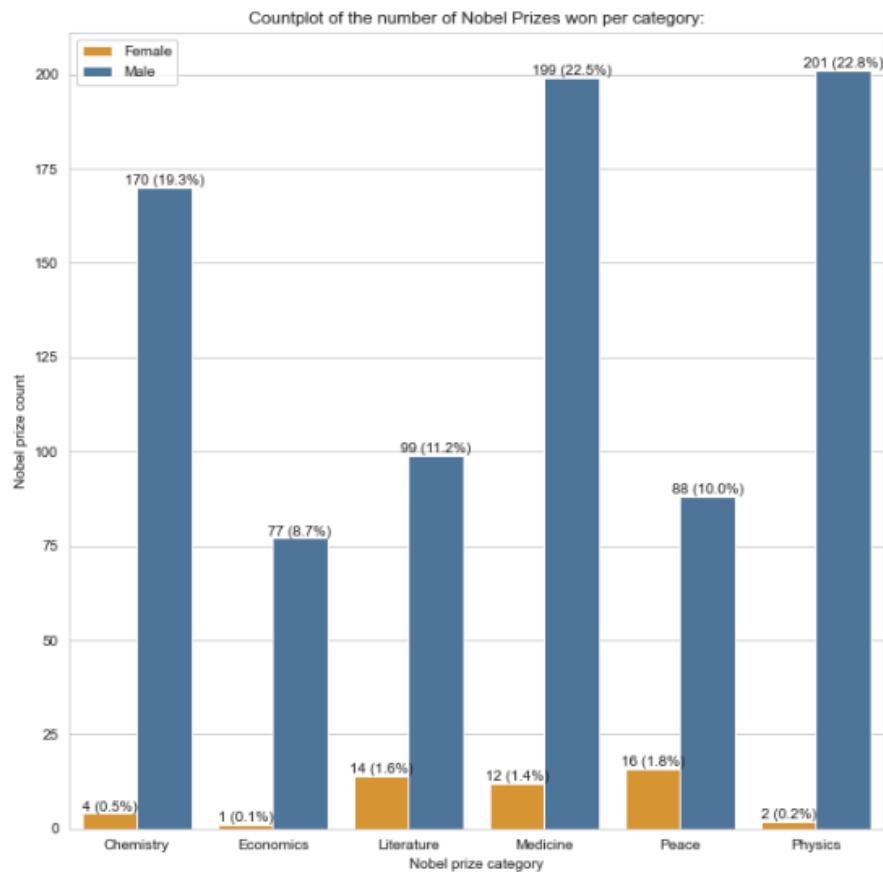
Method Selection:

To show the gender disparity in terms of number of awards given out we tried using a 1D swarm plot without any jitter, but it was difficult to make anything out of it, so we chose a stacked bar graph to show the distribution of awards among male and females each year.

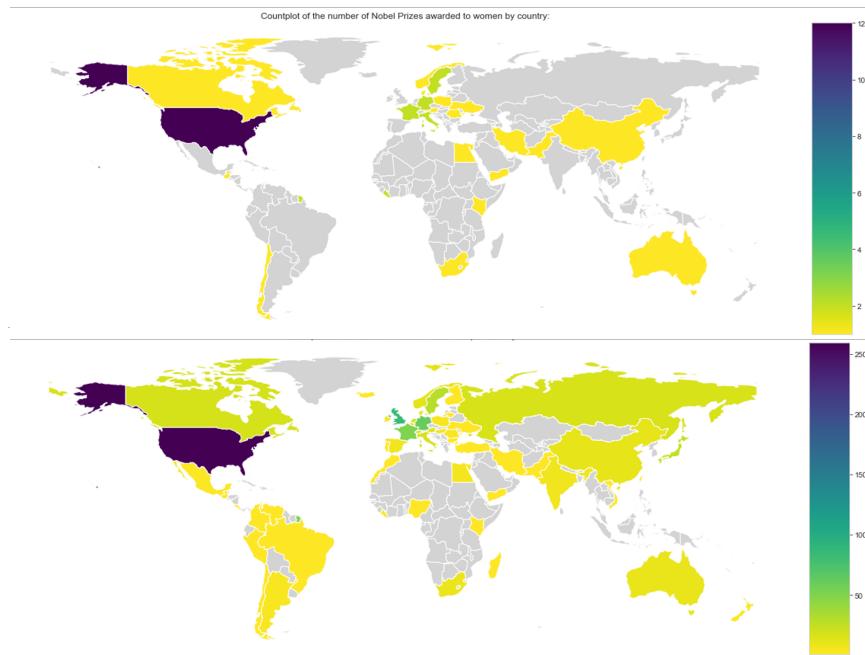
Gender distribution of prize winners by year and category



We decided it would be better to present the data on a multibar bar graph with each group representing a category because the previous plot made no sense. We will learn valuable information from this plot in conjunction with the pie charts showing the percentages that each gender has in each category.

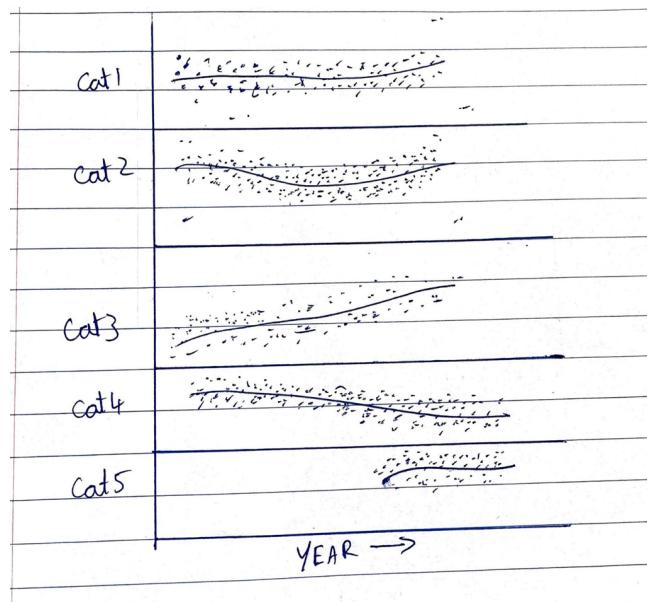


As per our plan to see the difference in the number of countries females have won from, we plotted a choropleth map, which is interactive and can provide us with the details of each country. We have chosen the world map as it is the easiest way to visualize the gender distribution over the world. We selected "Viridis r" as our colormap since it was the colormap that was consistently perceived across all our testing. We first tried using "Viridis" along with other colormaps, such as "Magma," but it was suppressing a lot of the details, so we moved on to "Viridis r."

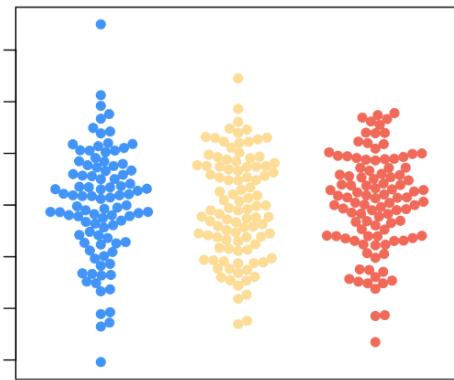


Visualization: Age and gender distribution for each category.

We were interested in the age vs. years plot's trend. We utilized a scatter plot for this, but it didn't provide us with much information, so we intended to use a scatter plot with a trend line to gain more from it. Something similar to the sketch below.

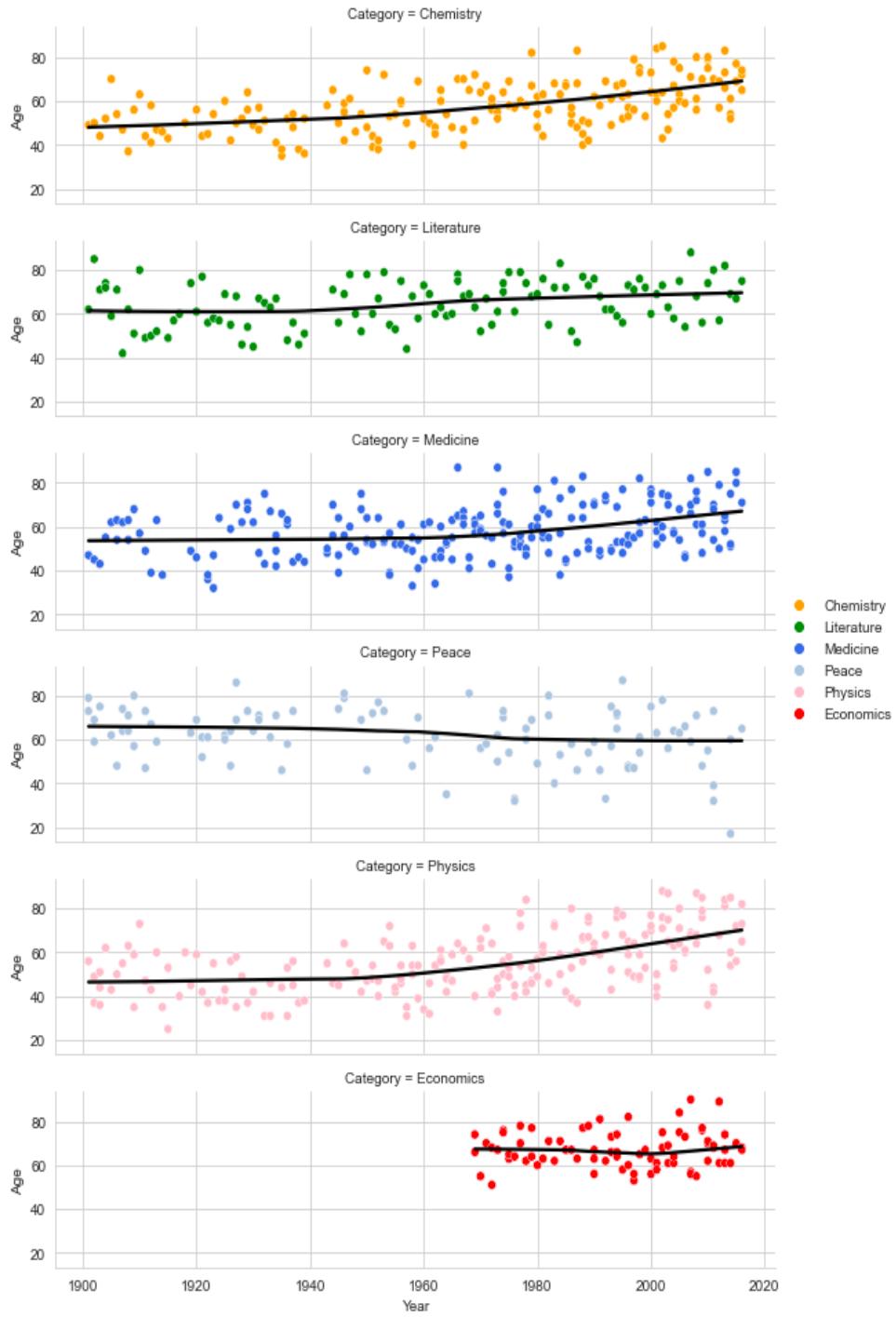


To further see the trend of age vs category separated by gender we will be plotting a beeswarm plot which will be easy to read and catch any outliers if any and gain insights from it.

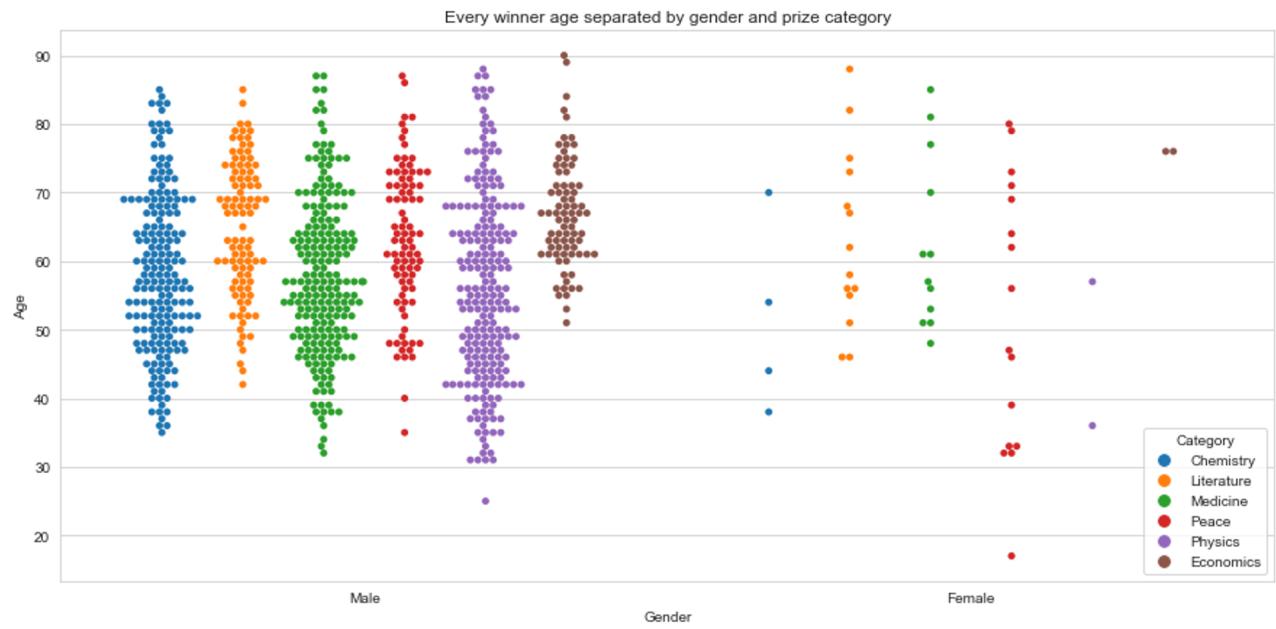


Method Selection:

To see the trend of the age at which Nobel laureates get an award has changed over the years, we plotted multiple scatter plots with a trend line. We used a scatter plot to clearly see the outliers and understand the reason behind those outliers.



In order to understand what the distribution of age is in each category and compare it between genders, we thought that beeswarm plot was the best choice. Since we are not doing any statistics with our data, box plots didn't make any sense.



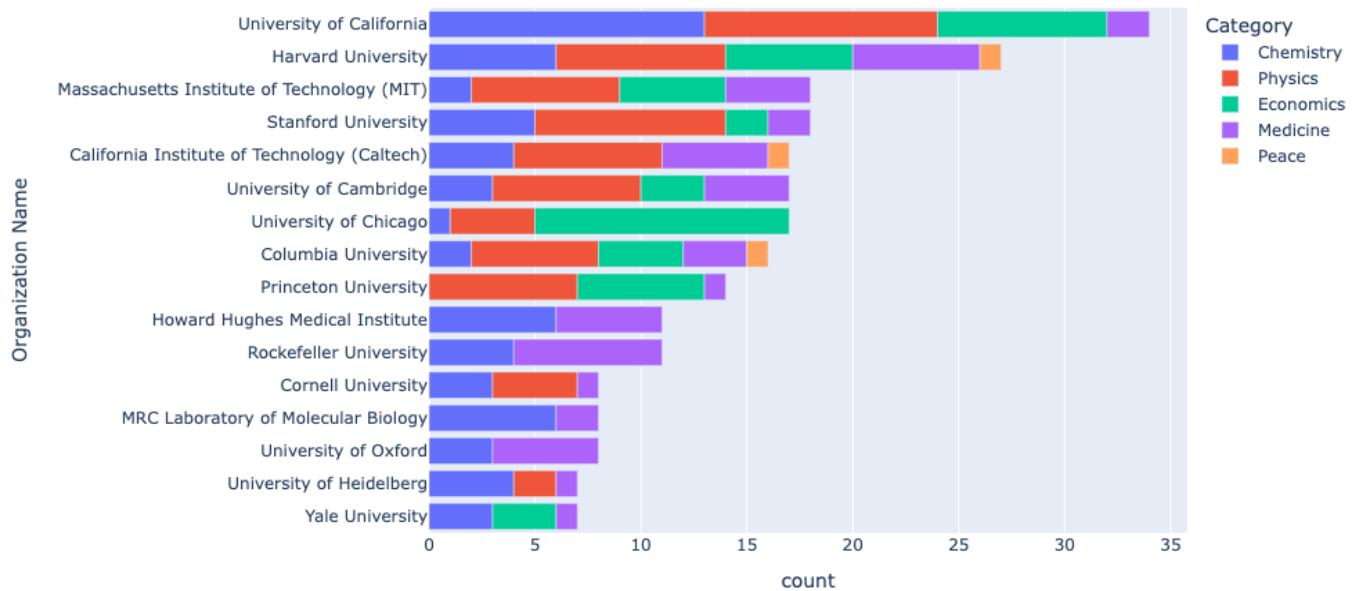
Results:

Visualization: Which university has the most Nobel awards?

The majority of the top universities on the list are from the US, as can be seen from the image below. This is explained by the fact that the US spends a lot of money on research. In the section that follows, it will be further addressed why the US has been in the lead.

We can also see that the primary shareholders in each university are the faculties of medicine, physics, and chemistry. This demonstrates that the majority of research funding and scholarships are awarded to scientific fields, and the outcomes are obvious. The other intriguing finding is that those who have achieved success in the field of literature are not affiliated with any universities. This is because most literary works are the creation of a single person. Additionally,

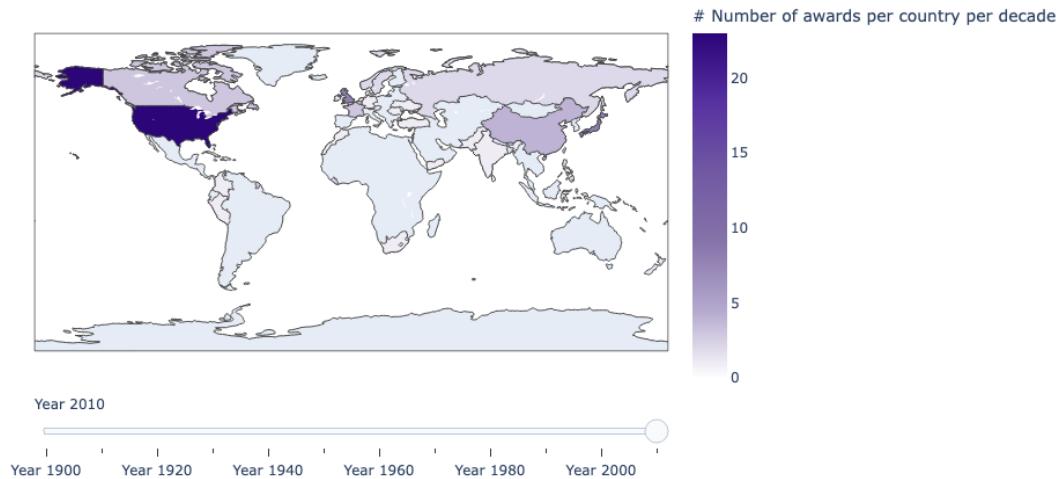
with the exception of a few times, none of the recipients of the peace prize are connected to academia.



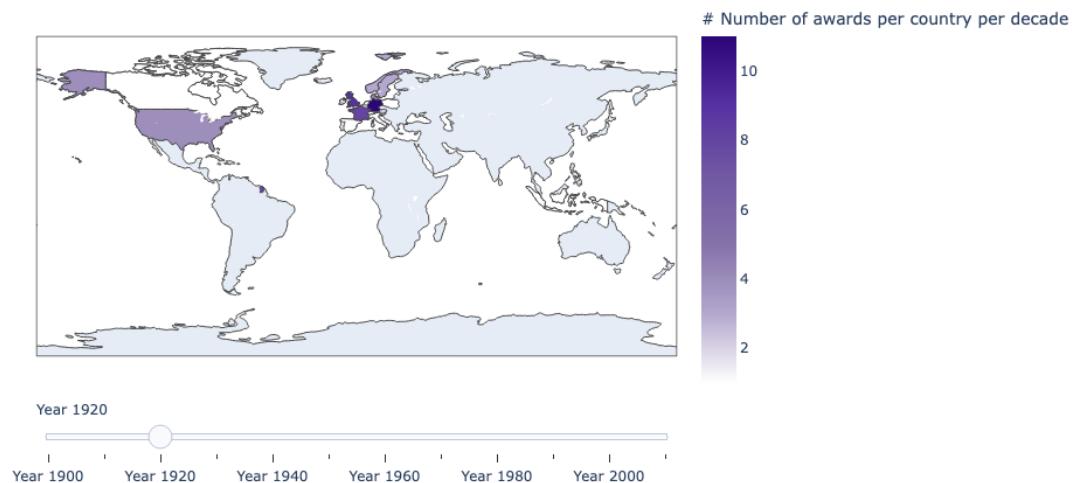
Visualization: US domination in Nobel Awards over the decades?

Our slider-equipped choropleth map makes it very evident that the US has dominated the Nobel Prize since the 1930s before which the UK and Europe had most Nobel awards. This is a result of the substantial research funding that was committed after the War to promote scientific advancements. In regions like Europe and Japan, there used to be enormous groups headed by extremely senior professors, and a younger person wouldn't take over until that professor retired, by which time they might no longer have their best ideas, which explains why the US absolutely dominates other nations.

Number of awards per country per decade



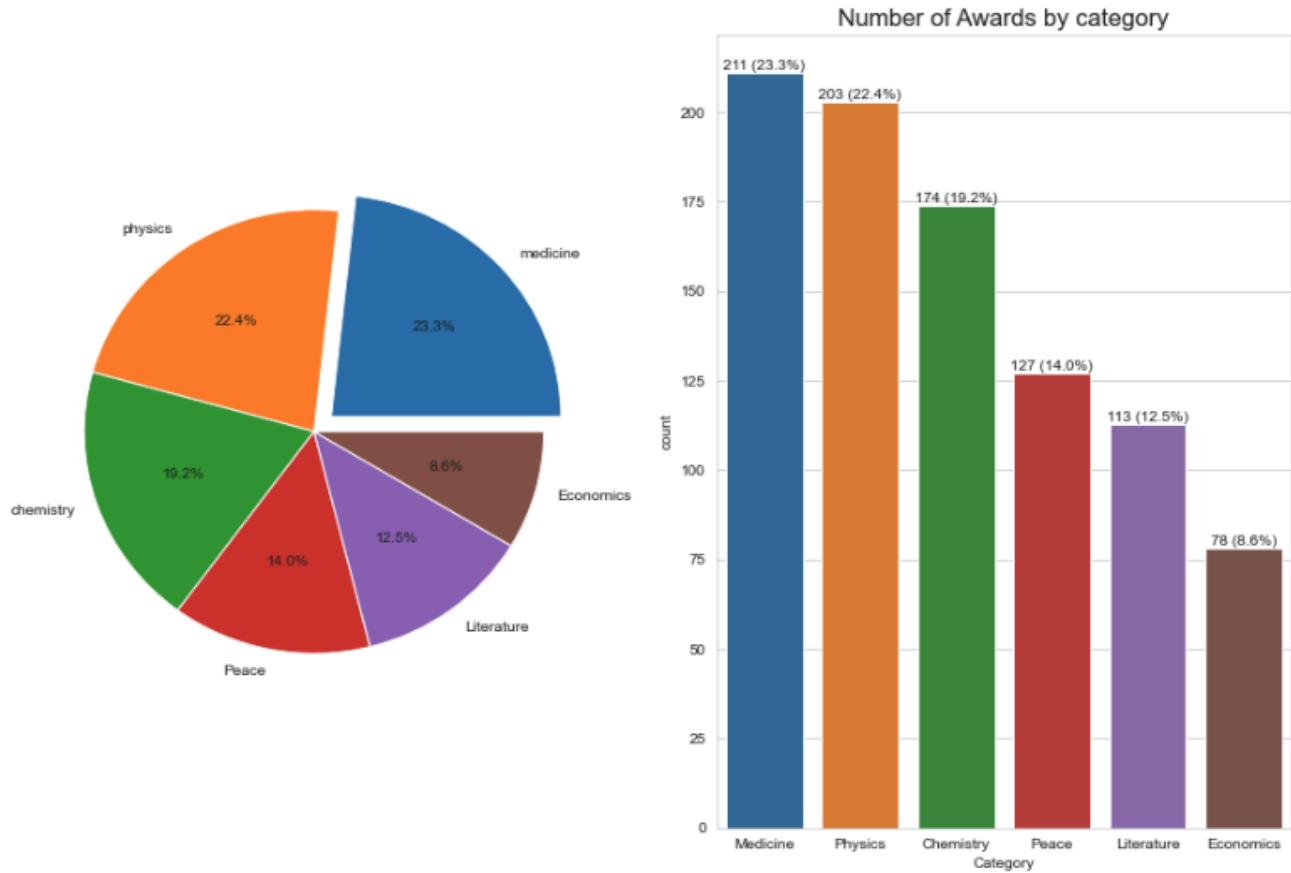
Number of awards per country per decade



It is also important to note that many immigrants from other nations conduct their study here and win a Nobel Prize since the US has been open to immigration. But the escalating xenophobic and nationalist sentiments that are reducing interest in traveling to the United States. This is especially true for Chinese students, who were scrutinized under the presidency of former President Donald Trump due to fears about espionage.

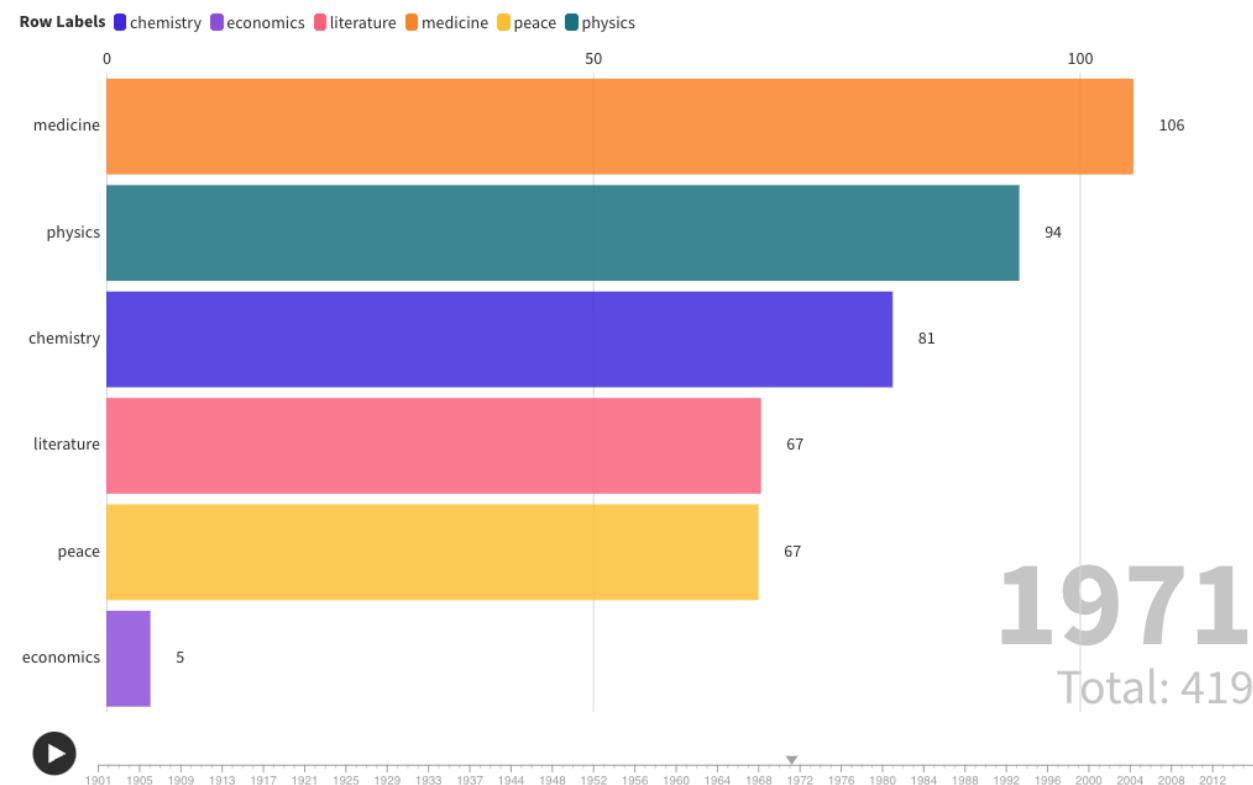
Visualization: Breakdown of Nobel prize by category

In accordance with our plan, we made a pie chart showing the cumulative data and the percentage shares of each group. The intriguing finding from this was that the scientific fields have received the most awards, again reflecting the amount of money dedicated to their study. To support our arguments that scientific fields receive greater money, we also plotted the bar graph with absolute values for the amount of awards each category received.

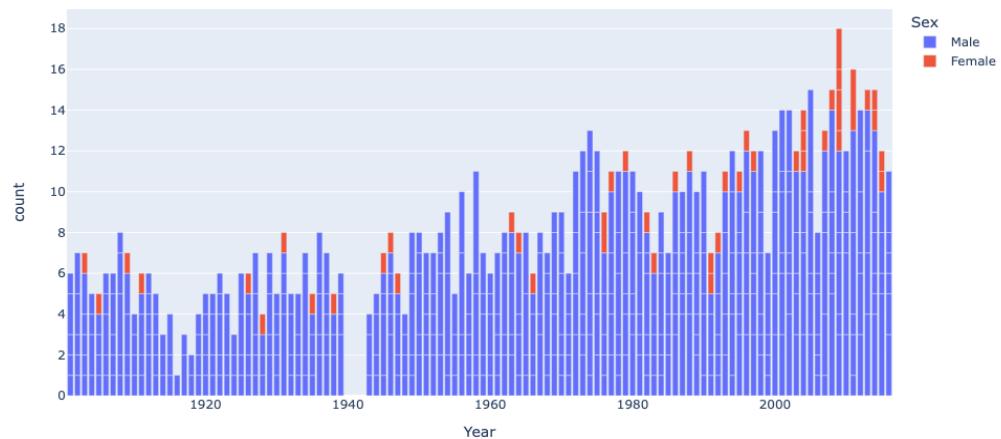


We also created a moving bar graph which updates the graph each year showing us the trend of each award. By looking at this we noticed a few things, there were no awards given in the years between 1939-1943. Economics was not part of the Nobel prizes till 1968, it was established

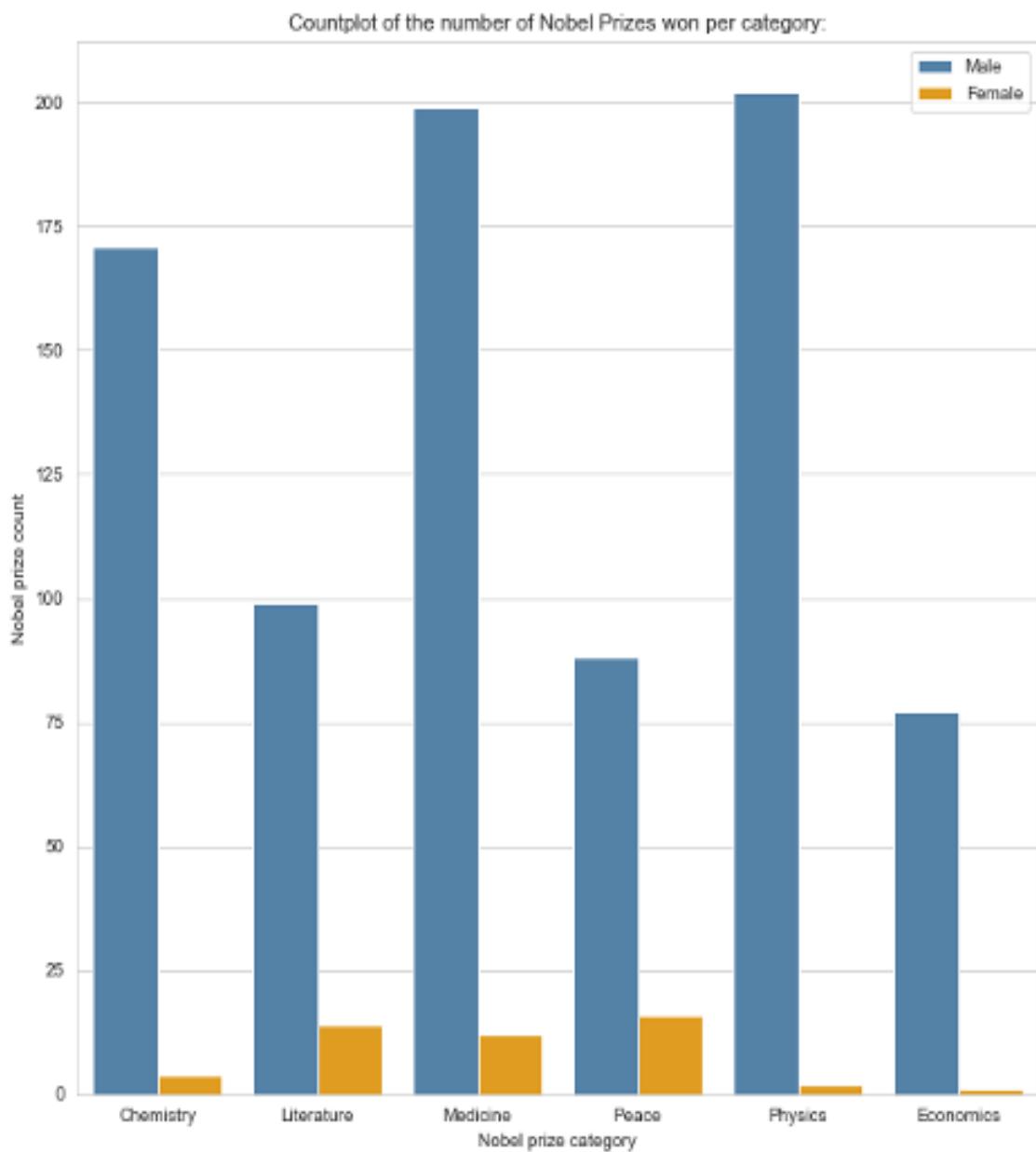
in 1968 and the first award was given out in 1969. IU Political Science Distinguished Professor Elinor Ostrom (1933-2012) was the first woman to be awarded the Nobel Prize for Economic Sciences in 2009.

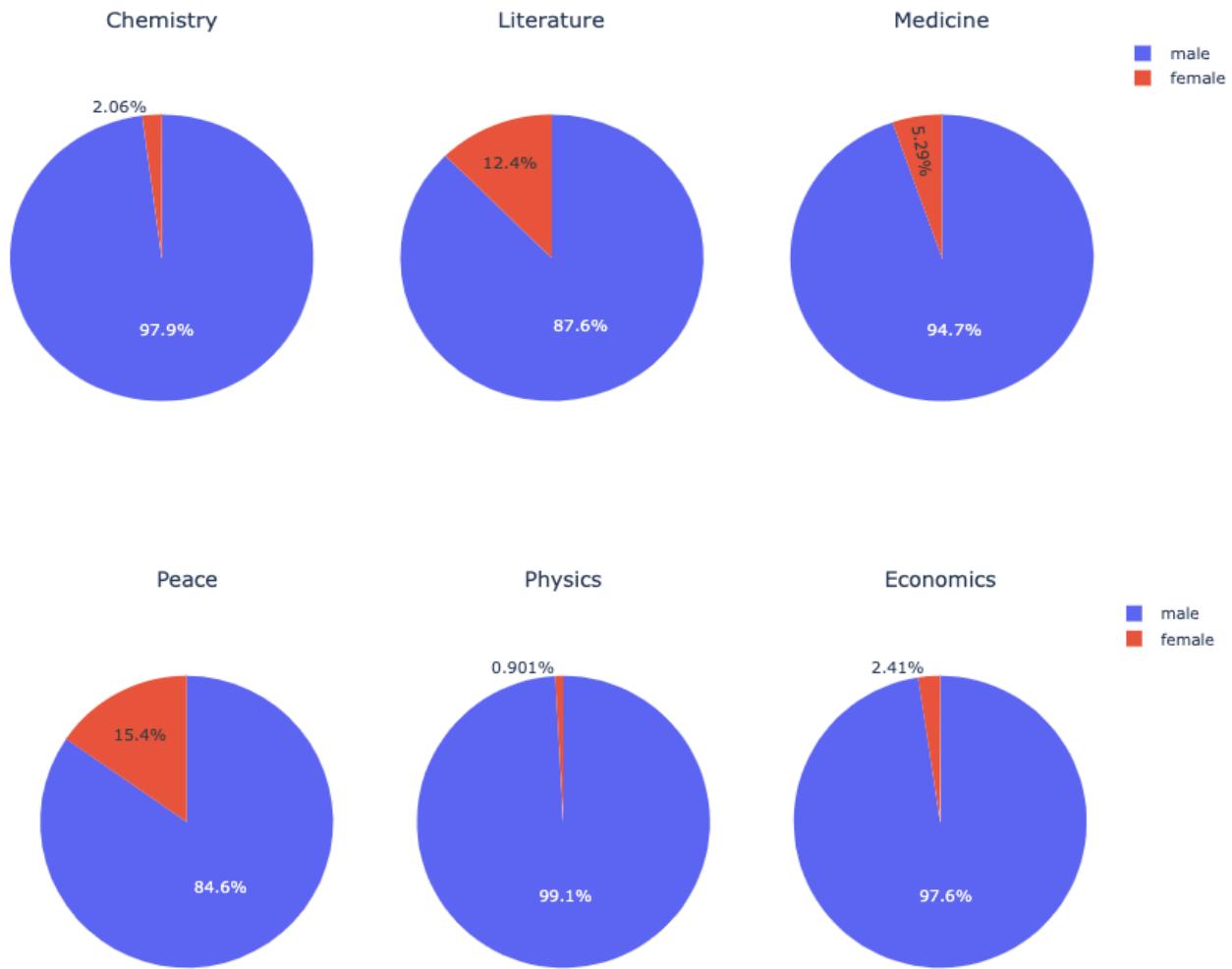


Visualization: Gender and country distribution of Nobel prizes till date



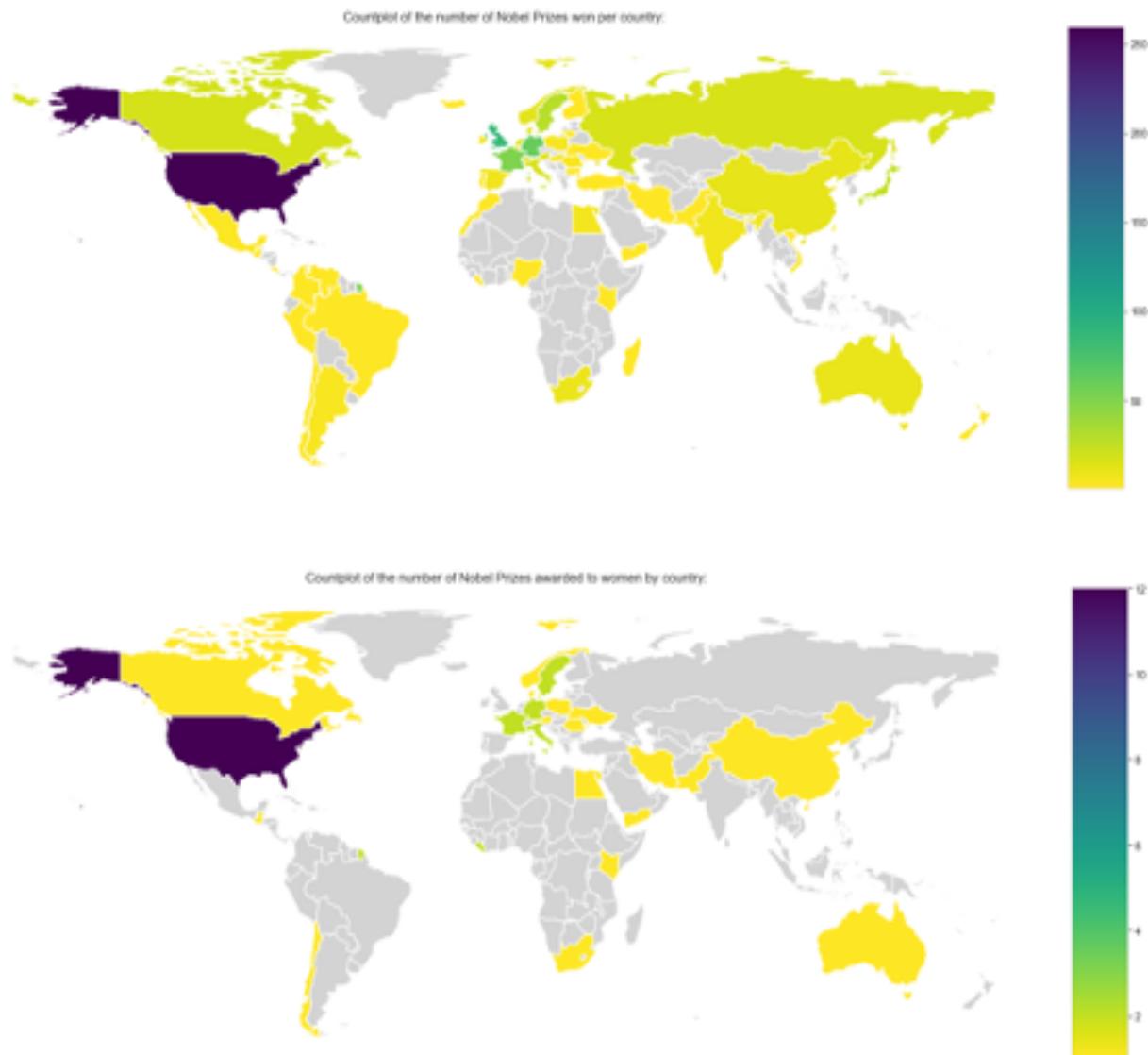
There is a gender gap in the number of Nobel laureates. The figure above makes it abundantly evident that there are significantly fewer female winners than there are male winners. The maximum number of honors given to female laureates in one year—nine—was given to them in 2009, making it the best year for female contenders. The years 1940 to 1942 are missing from the chart since no awards were granted during this time due to World War II. Another insight derived from the plot is that very few awards were given during World War I. In 2001, there were 15 awards granted, which was the most ever.





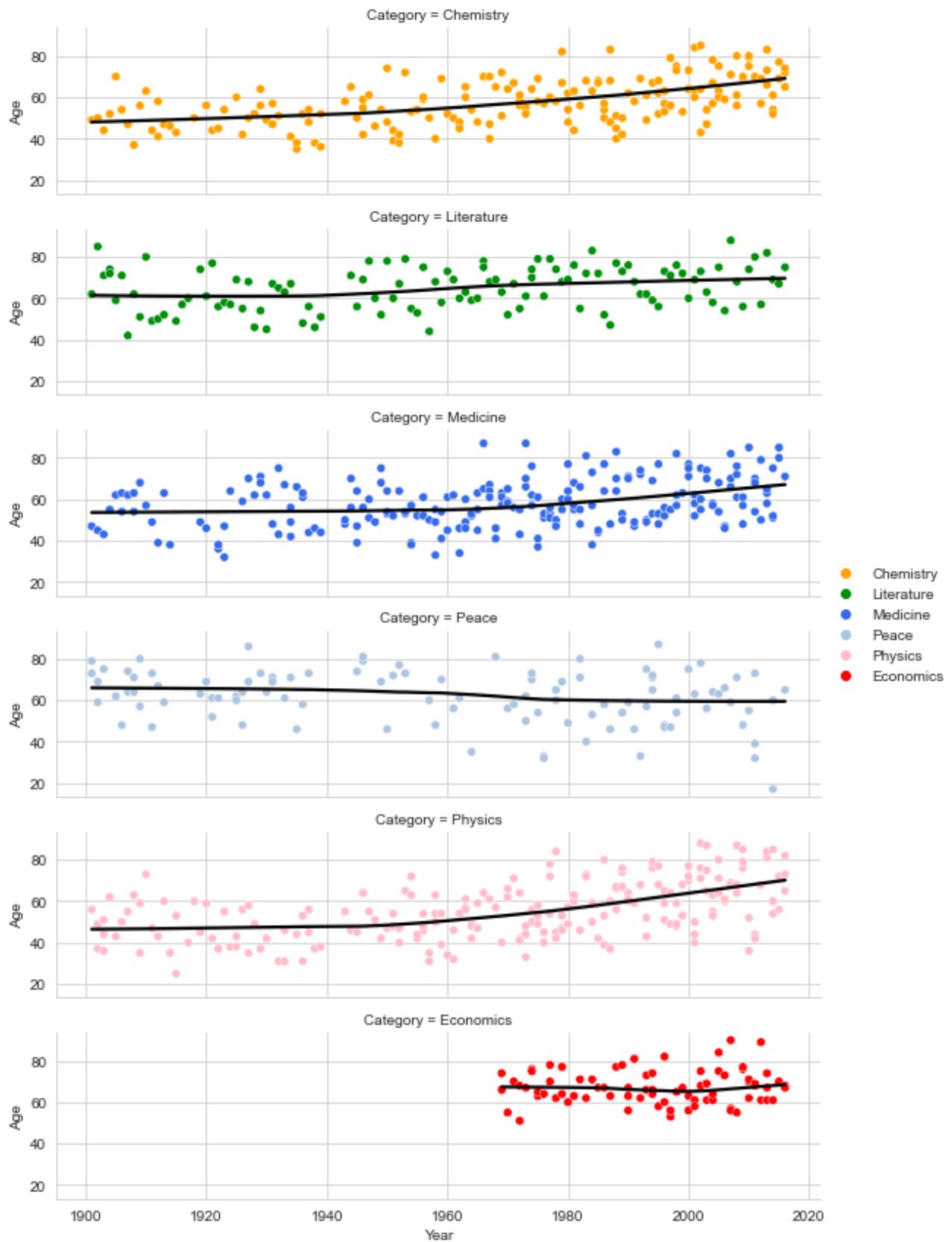
These Plots display the awards' percentage distribution in each of their gender-specific color-coded categories. Even while the gender gap is evident in every category, the most unbalanced situation is in Physics, where only 0.9% of the total awards were granted to women. Despite the fact that there have been two female laureates in both physics and economics, which is the same amount overall, the ratio in economics is higher at 2.4%. While working for Indiana University, Elinor Ostrom was awarded one of these Nobel Prizes in 2009 for her outstanding

contributions to the field of economics. She was the first woman to receive the Nobel Prize in Economic Sciences. Female Nobel laureates have been the most successful in the Peace Category, with a percentage of 15.4%.



The dominance of the United States is again evident here, with 12 prizes won by the country compared to 2 or 1 awards won by the other nations. Significant countries like India, United Kingdom, Japan, and Russia have not yet produced a single Female Laureate.

Visualization: Age and gender distribution for each category



The 2016 Nobel laureates for physics, medicine, and chemistry: all men, at least 65 years old and mostly over 72. However, the average laureate back in the first half of the 20th century was "only" 56. The average age of the physics laureates, who are now typically men in their late 60s, was formerly 47. Since the 1950s and up until the present, there has been a noticeable tendency of Nobel laureates earning awards later in life in all of the traditional disciplines. Every category has an increasing trend except Peace. The age of laureates in the Peace category is decreasing with every passing year. This appears to be the trend given the large number of applicants for the awards and the difficulty of winning in comparison to earlier periods. The most pronounced age trend is shown in the subject of physics, where the average age of laureates is rising. This can be attributed to the early 20th century scientific revolution, which was triggered by the developing discipline of quantum mechanics. Mr. Källstrand (Prominent Researcher) asserts that throughout the first decade of the 20th century, "physics was a fast-expanding field; many physicists were young, and they were creating discoveries swiftly." This was made clear to the Nobel committee. "The committee understood this. They had a keen interest in the subject and were eager to acknowledge achievements "he explained. This explains the trend seen in the category of Physics



It looks like Physics has a bulk of the youngest laureates. Chemistry and Medicine follow next. Malala Yousuf, who is the youngest recipient of an award in the area of peace at the age of 16, is the cause of the outlier in the peace category, which is significantly younger than the average age of Nobel Prize recipients. The field of economics appears to have the oldest winners with points above 90. Female Nobel laureates in literature are older than their male contemporaries, whereas female laureates in chemistry are younger.

Conclusion:

By working on this project, we gained some practical insights on making creative and interactive visualizations keeping in mind the psychological and other important factors. It also allowed us to gain a deeper understanding on the subject of Nobel Prizes by answering the questions that we raised earlier in the report. Some of the most interesting insights gathered by these visualizations were the dominance of certain countries in getting the most Nobel Prizes, the need for immigration support in the US, and helping women in technology to effectively reduce the gender inequality in Nobel Prizes.

In the future, we would like to incorporate data with latitude and longitude values so that we may visualize the actual cities where Nobel laureates live. We wish to incorporate locations on our globe map that will enhance our visualization. We would also like to incorporate data on committee members and their nationalities to determine whether there is a correlation between the committee members' nationalities and the Nobel laureate's nationality. We would like to create an API that can obtain real-time data or data that refreshes once a year, automatically replotted all our visuals and keeping them up to date.

References:

1. simpleshow foundation. (2014b, December 10). *The Nobel Prize – explained in a nutshell*. YouTube. Retrieved October 14, 2022, from <https://www.youtube.com/watch?v=94YQ3el6LMc>
2. *Nobel Prize*. (n.d.). University Honors & Awards. Retrieved October 14, 2022, from <https://honorsandawards.iu.edu/awards/research-creative/nobel-prize.html>
3. Harvard University. (2022, March 9). *Nobel Laureates*. Retrieved October 14, 2022, from <https://www.harvard.edu/about/history/nobel-laureates/>
4. Amazing Rankings. (2019, October 16). *Top Universities that Produced the most Science Nobel Prize Winners 1901-2019*. YouTube. Retrieved October 14, 2022, from <https://www.youtube.com/watch?v=k5XJ--XyPec>
5. BRAINIAC. (2021, November 14). *Countries with most number of NOBEL prize winners 2021*. YouTube. Retrieved October 14, 2022, from <https://www.youtube.com/watch?v=kZwjjuE2qu8>
6. *All Nobel Prizes*. (n.d.). NobelPrize.org. Retrieved October 14, 2022, from <https://www.nobelprize.org/prizes/lists/all-nobel-prizes/>
7. *Nobel Prize Winners*. (2022, October 8). Kaggle. Retrieved October 14, 2022, from <https://www.kaggle.com/datasets/thedevastator/a-complete-history-of-nobel-prize-winners>

8. *Nobel Prizes 2022*. (2021, July 6). NobelPrize.org. Retrieved October 14, 2022, from
<https://www.nobelprize.org/about/developer-zone-2/>
9. *Nobel Prize facts*. (n.d.). NobelPrize.org. Retrieved October 14, 2022, from
<https://www.nobelprize.org/prizes/facts/nobel-prize-facts/>
10. .jn, umesh. (n.d.).
https://public.tableau.com/app/profile/umeshjn/viz/Nobel_15/NobelPrizes Tableau.
Retrieved October 14, 2022, from
https://public.tableau.com/app/profile/umeshjn/viz/Nobel_15/NobelPrizes
11. Locke, S. (2014, October 14). *The age, sex, education, and facial hair of every Nobel Prize winner, in one chart*. Vox. Retrieved October 14, 2022, from
<https://www.vox.com/xpress/2014/10/14/6966291/nobel-prize-winners-youngest-chart-graphic-age-demographics-facial-hair>
12. *Age and the Nobel prize revisited*. (1993, November 1). SpringerLink. Retrieved October 14, 2022, from
https://link.springer.com/article/10.1007/BF02026517?error=cookies_not_supported&code=df1c4714-7116-4eab-8b9e-4c47795fbef2
13. Robinson, B., PhD. (2021, February 15). *Gender Discrimination Is Still Alive And Well In The Workplace In 2021*. Forbes. Retrieved October 14, 2022, from
<https://www.forbes.com/sites/bryanrobinson/2021/02/15/gender-discrimination-is-still-alive-and-well-in-the-workplace-in-2021/?sh=2ad5e5f27f1c>

14. Staff, S. X. (2021, October 12). *Why the United States dominates the Nobels*. Retrieved October 14, 2022, from <https://phys.org/news/2021-10-states-dominates-nobels.html>
15. *Why Do So Many Americans Win the Nobel Prize?* (2017, October 6). Inside Science. Retrieved October 14, 2022, from <https://www.insidescience.org/news/why-do-so-many-americans-win-nobel-prize>
16. Bjørk, R. (2019). *The age at which Noble Prize research is conducted*. Welcome to DTU Research Database. Retrieved October 14, 2022, from <https://orbit.dtu.dk/en/publications/the-age-at-which-noble-prize-research-is-conducted>
17. *What Is the Average Age of Marriage in the U.S.?* (2022, April 22). Brides. Retrieved October 14, 2022, from <https://www.brides.com/what-is-the-average-age-of-marriage-in-the-u-s-4685727>
18. *Exploring historic Nobel prize data.* (2020, August 2). Kaggle. Retrieved October 14, 2022, from <https://www.kaggle.com/code/imdevskp/exploring-historic-nobel-prize-data>