Analysis of Different Patterns in Nobel Prize Laureates

Group 15

Introduction

Our group aims to create a report examining the evolution and patterns of Nobel Prize awards over more than a century. The Nobel Prize dataset (1901-2016) [1] (via Kaggle) contributes the data we will use to accomplish our goal of understanding how recognition of intellectual and humanitarian achievement has evolved globally over time.

Our group's analysis initially started broad but began to focus in and begin addressing narrower questions that were obtainable and that we believe will interest both academic audiences and the general public in an intuitive and easily understandable format. After going through some of the critiques and recommendations. Based on the recommendations, we placed the choropleth map into Shiny to allow users to zoom in and look at countries that are smaller and underrepresented in a map setting, and added more introductory information as well as a data table allowing users to compare the data with visualizations as shown. We also changed the way we gather the visualizations to allow for a higher resolution view. After the critiques and reworks, we wanted users to look at the visualizations and the information available to them to answer a variety of questions as below:

- Are there specific demographics and information related to Nobel Prize laureates that tell a unique story?
- How are the Nobel recipients distributed throughout the world? Which countries have the most recipients, and does this relate to any potential historical factors at play?
- How has the share of recipients changed over time? Are certain awards represented more? Do we see an increase in the number of multiple-recipient awards?

Literature Review

We focused our search across two domains: temporal pattern visualization and geographic and network visualization. Then, we made observations about recurring themes.

Temporal Pattern Visualization

- The excerpt from our Stat 436 course notes [3] on Time series graphics breaks down how to interface with Tidyverse and R and actually model temporal data.
- *Chaining and the temporal dynamics of scientists' publishing behaviour* [4] is a paper that models a similar problem. In this case, the authors stratify the papers produced by a single Turing award winner over the course of his career. However, the strategies can be adapted and generalized for our use case.
    - Useful: See the figures for this paper here.

- *Visualization of Time-Oriented Data 2e* [5] by Aigner et al. is a comprehensive resource on representing temporal data.

Geographic and Network Visualization

- *Mapping the Global South: Equal-Area Projections for Choropletheth Maps* [5] serves as an introduction to the considerations researchers and scientists use when choosing projections and ideas for choropleth maps which we plan on using as a way to guide our choice in projections.
- *Genius Cliques: Mapping Out The Nobel Network* [6] is a paper that provides a set of socio-geographic representations of the Nobel laureates. In particular, it groups laureates based on how related their stories/focus were.
- *The Nobel "Pride" Phenomenon: An analysis of Nobel Prize discoveries and their recognition* [7] is a paper which creates visualizations to show the following: Nobel Prize discoveries are highly concentrated in just five countries. 30% of all Nobel laureates were immigrants; most moved before their Nobel discoveries. 77 institutions with Nobel discoveries never were recognized with a Nobel Prize. Only three countries outperform the US in terms of discoveries per capita.

Visualizations

The Shiny dashboard is available online for viewing at nobel.salm.dev or http://159.203.146.138:3839/ – if both domains do not resolve, please contact salm@cs.wisc.edu, and I will promptly restore it.

Critical Report:

We began with a series of static visualizations present in earlier Milestone reports. Here, we have replaced them with a dynamic dashboard using R, Shiny, and a variety of R packages. The source code is available on GitHub.

For each visualization, we create a reactive object that holds the summary data frame of the statistics we are displaying for that given visualization. These reactive objects automatically update whenever the filter selected by the user changes. The benefit of using reactivity is the automatic and efficient updating of data in response to user input.

Our data [1] comprises 911 data entries with 18 columns. Numeric and character

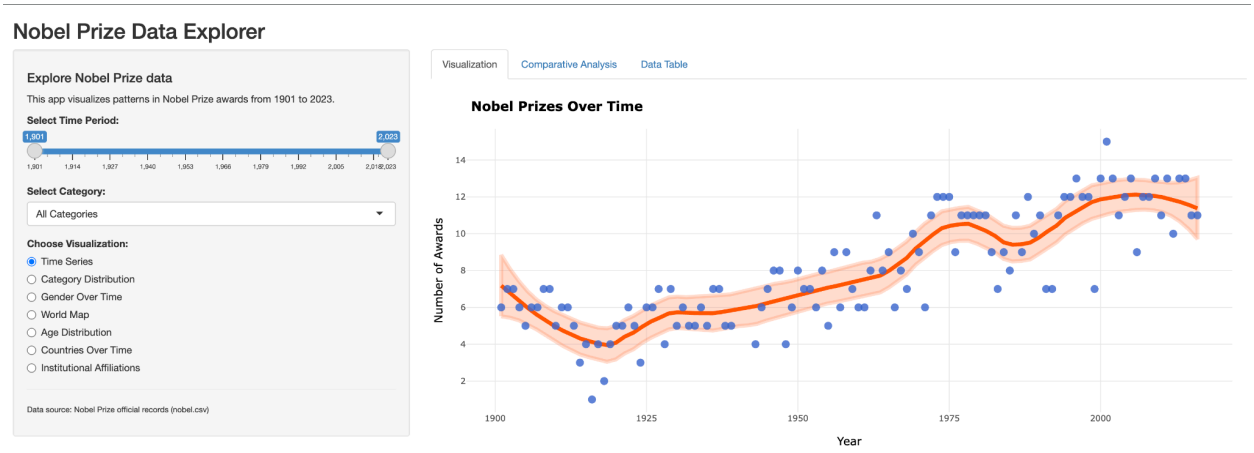
There exist numeric columns like "year" (of prize acquisition) and "laureate id". The remainder of the columns are character strings, and include information like "birth country", "organization name" (their institution), and "death city" (where they died).

We preprocessed the data by editing the "birth date" and "death date" columns with lubridate to become of type date-time, with the rest of the columns as text/id being used to describe some sort of information related to the laureate. When digging deeper into the data, most of the

columns have some sort of missing information as well as untidy information that had to be manually updated via Excel and R before further analysis.

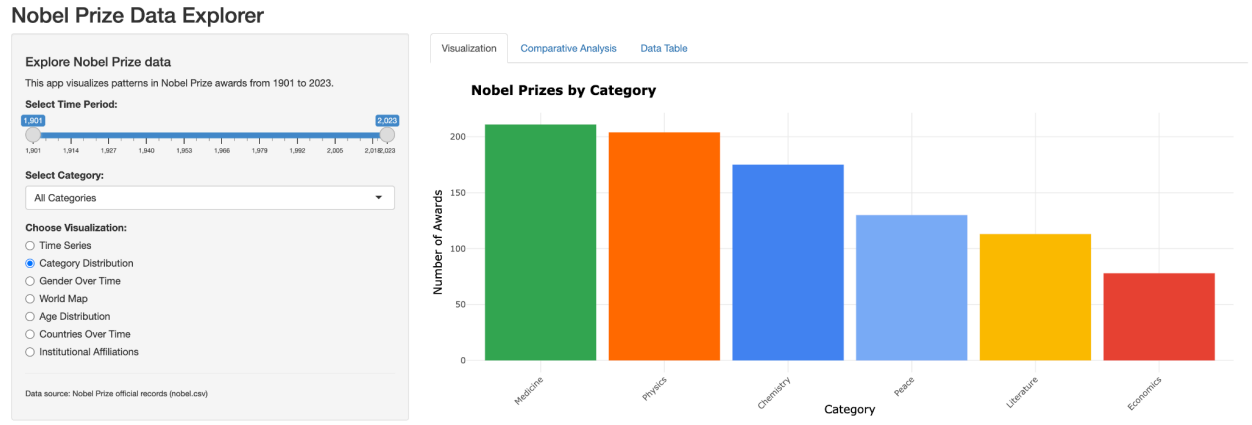
Exploratory Interface:

Our hope the addition of these visualizations will aid in telling the story we've set out to tell – so let's see what we can discover:



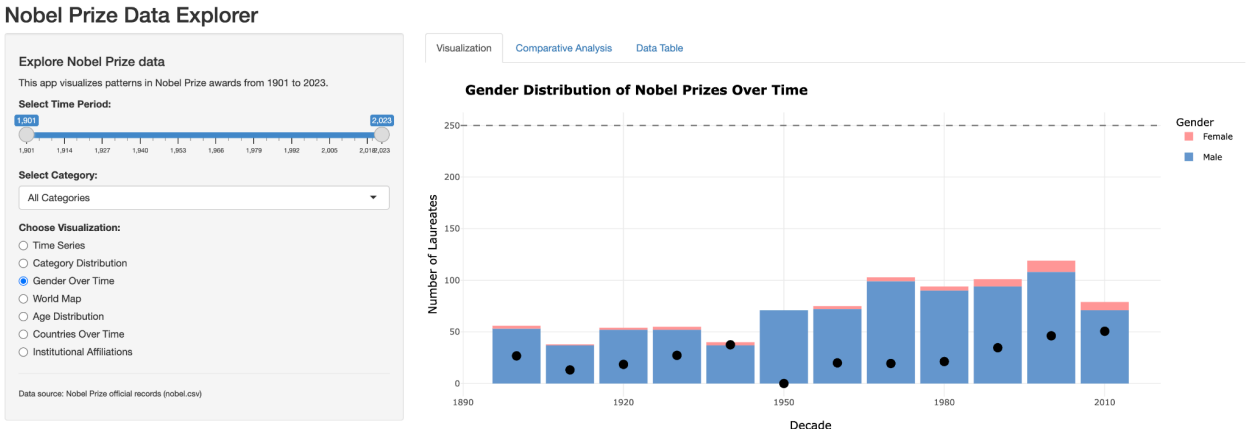
Screenshot 1: Time Series with Trend Line.

This visualization (above) displays the number of Nobel Prizes awarded annually from 1901 to 2023, with a LOESS smoothing trend line overlaid. The blue dots represent individual years, while the orange trend line reveals long-term patterns in award frequency. This time series shows periods of decreased awards during World Wars, followed by a significant increase post-1950s. The visualization addresses the question of how award patterns have changed over time, showing a general upward trend in the number of awards given annually, suggesting an increase in multiple-recipient awards in modern times.



Screenshot 2: Category Distribution Bar Chart.

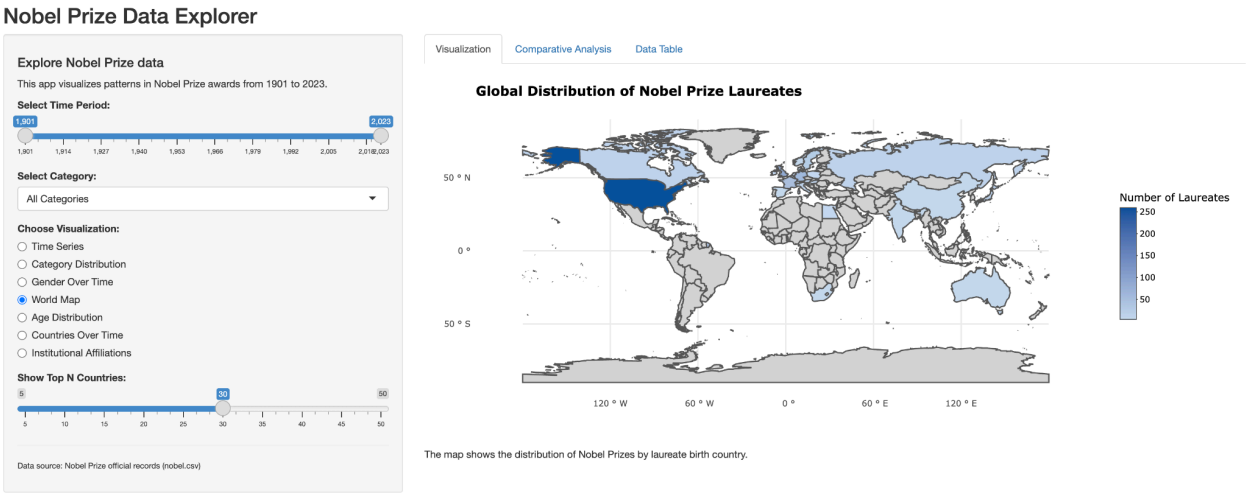
This bar chart (above) displays the total count of Nobel Prizes by category (Physics, Chemistry, Medicine, Literature, Peace, and Economics). The color-coded bars provide a clear comparison of category representation, showing which fields have received more recognition. Physics, Chemistry, and Medicine lead in several awards, while Economics (added only in 1969) has the fewest. This visualization directly addresses the question of whether certain awards are represented more throughout Nobel Prize history.



Screenshot 3: Gender Distribution Over Time.

This stacked bar chart (above) with line overlay shows the gender distribution of Nobel laureates by decade. The blue and pink segments represent male and female laureates, respectively, with a black line indicating the percentage of female representation. The visualization starkly illustrates the significant gender disparity in Nobel Prizes, with minimal female representation until recent decades. This addresses the demographic question by revealing historical gender biases in award recipients and showing a gradual improvement in female representation, particularly after 1970.

Screenshot 4: The view (above) is another way of looking at the information presented in Screenshot 3. It calculates and displays an “Equality Index”. That is, it calculates and displays $\frac{\left|\frac{Female\ Prizes}{Total\ Prizes} - 0.5\right|}{0.5}$. The dashed line indicates the 50/50 female/male mark. The story and conclusions match.

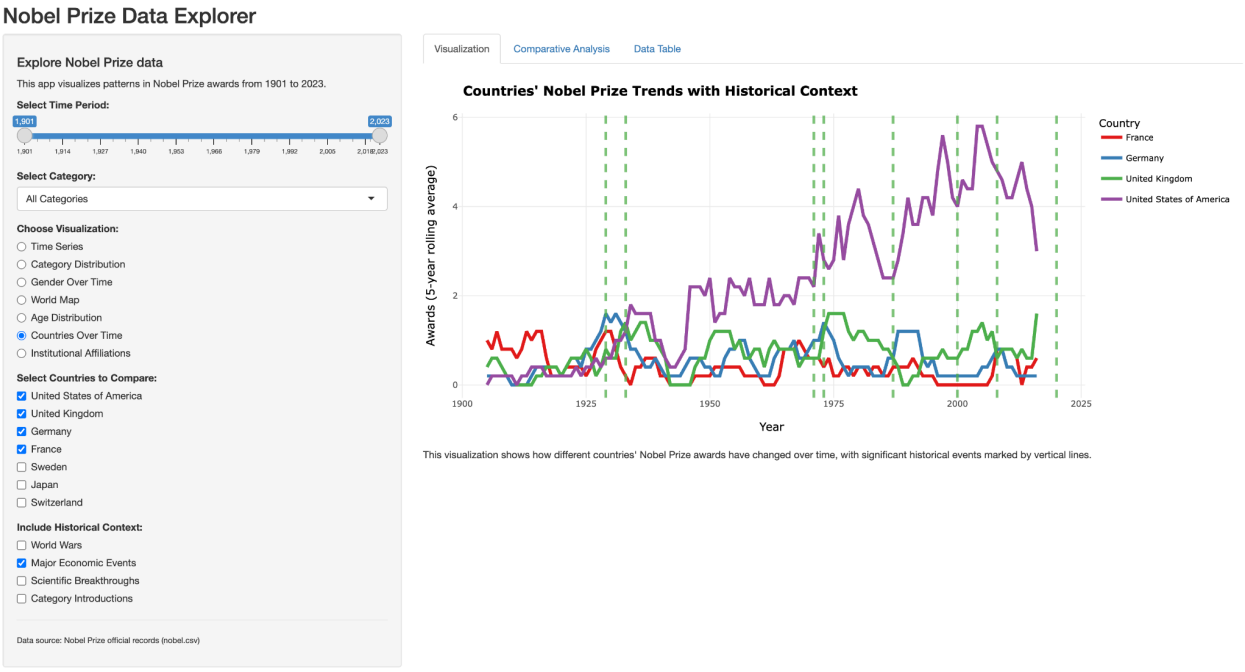


Screenshot 5: World Map (Choropleth).

This choropleth map [5] (above) displays the global distribution of Nobel Prize laureates by birth country. The color gradient (light to dark blue) indicates the number of laureates from each country, with the United States and Western European nations showing the highest concentration. This visualization directly answers the question about the geographic distribution of laureates, highlighting the Western-centric nature of the awards and allowing exploration of potential correlations with geopolitical influence, educational investment, and research infrastructure.

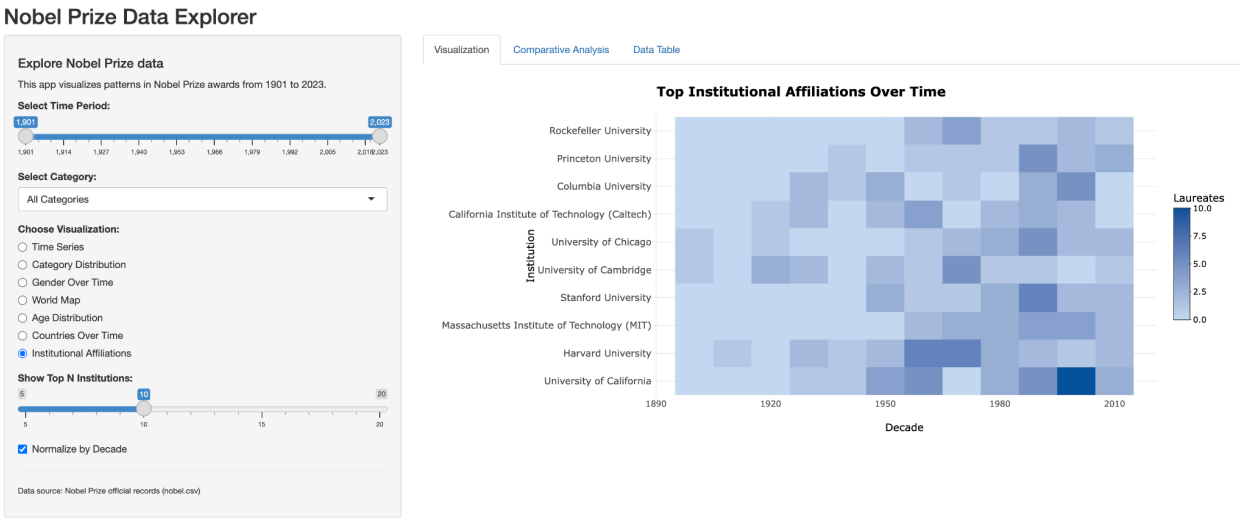
Screenshot 6: Age Distribution By Category

This box plot visualization (above) shows the age distribution of laureates at the time of receiving their award, segmented by prize category. The boxes indicate the interquartile range, with horizontal lines marking the median age. Notable patterns include younger laureates in Physics and Chemistry compared to Literature and Peace. This visualization contributes to the demographic analysis by revealing how age patterns differ across disciplines, potentially reflecting the nature of breakthrough work in different fields.



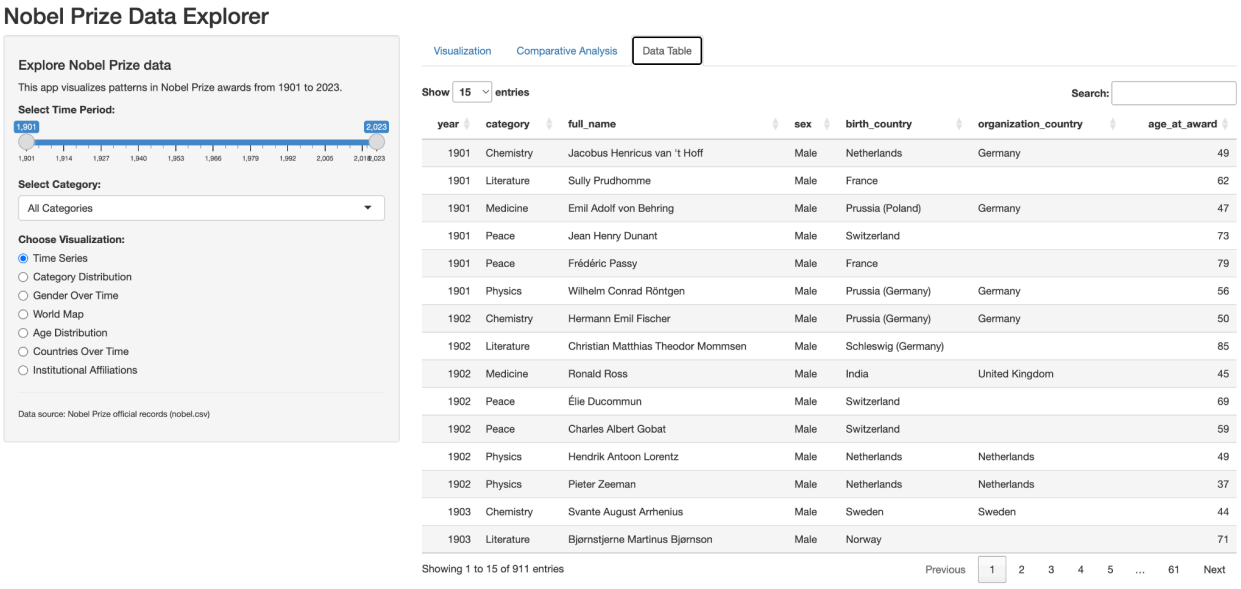
Screenshot 7: Countries' Nobel Prize Trends with Historical Context.

This line graph (above) shows the 5-year rolling average of Nobel Prizes awarded to different countries over time, with vertical dashed lines marking significant historical events. The visualization reveals how geopolitical events (wars, economic changes) correlate with shifts in national scientific output and recognition. The United States shows dramatic increases post-WWII, while other countries show fluctuations corresponding to historical circumstances. This directly addresses the question about historical factors influencing the geographic distribution of laureates.



Screenshot 8: Top Institutional Affiliations Over Time.

This heatmap (above) displays the distribution of Nobel laureates across top research institutions by decade. Darker blue cells indicate higher concentrations of laureates. The visualization reveals institutional centers of excellence like Rockefeller University, Harvard, MIT, and Cambridge. This contributes to understanding both the demographic and geographic questions by showing how certain institutions have dominated Nobel recognition and how institutional leadership has shifted over time.



Screenshot 9: Data Table.

The data table (above) allows granular interaction with the data itself.

All together, these visualizations provide a comprehensive analysis of Nobel Prize patterns across time, geography, and demographics, effectively addressing all three research questions while revealing both expected trends and subtle factors that have influenced the distribution of these prestigious awards.

Conclusion

After considering the visualizations and reviewing the relevant information. We see that throughout time, Nobel prize laureates are prone to a variety of expected factors that are at play as well as more subtle reasoning when looking deeper into the different visualizations.

When looking at the distribution of awards, we see several uneven distributions that are at play. When looking at the demographic information of laureates, older men dominate the distribution, potentially showing sexism due to historical factors limiting women from participating in academia. Additionally, we see uneven distribution of awards when looking at the different countries. To be expected, wealthier and more developed countries had more laureates due to their excess investment and resources that are available, but there are more subtle factors at play when looking at geographic distributions. An example of this comes from Russia. Things like the Cold War, leading to the Nobel Prize laureates coming out of Russia, took a rapid decline following the end of World War 2 and the beginning of the Cold War.

Although the Nobel Awards have gone a long way since 1900, with more group awards being given. Understanding and looking for ways to resolve issues relating to uneven distribution, bias, and external factors helps further push the legitimacy of the award and serves as a way to get more eyes and people interested in academia, which in turn furthers the development of the world.