**Analysing trends in Noble Prize winners using R**

**Summer Internship 2020**

Submitted by:

**Bhartendu Dubey (9917103102)**

Under the supervision of:

**Ms. Ambalika Sarkar**



**Department of CSE/IT**

**Jaypee Institute of Information Technology University, Noida**

**June 2020**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
|  |  | **Page No.** |
| **1.** | **INTRODUCTION**  **1.1 General Description**  **1.2 Objective and scope of the project** | 3  3  3 |
| **2.** | **REQUIREMENT ANALYSIS**  **2.1 Hardware Requirements**  **2.2 Software Requirements** | 4  4  4 |
| **3.** | **DETAILED ANALYSIS & IMPLEMENTATION** | 5 |
| **4.** | **CONCLUSION** | 21 |

1. **INTRODUCTON**

The Nobel Prize is a set of annual international awards bestowed in several categories by [Swedish](https://en.wikipedia.org/wiki/Sweden) and [Norwegian](https://en.wikipedia.org/wiki/Norway) institutions in recognition of academic, cultural, or scientific advances around the world. The [will](https://en.wikipedia.org/wiki/Will_and_testament) of the [Swedish](https://en.wikipedia.org/wiki/Swedes) chemist, engineer and industrialist [Alfred Nobel](https://en.wikipedia.org/wiki/Alfred_Nobel) established the five Nobel prizes in 1895. The prizes in [Chemistry](https://en.wikipedia.org/wiki/Nobel_Prize_in_Chemistry), [Literature](https://en.wikipedia.org/wiki/Nobel_Prize_in_Literature), [Peace](https://en.wikipedia.org/wiki/Nobel_Peace_Prize), [Physics](https://en.wikipedia.org/wiki/Nobel_Prize_in_Physics), and [Physiology or Medicine](https://en.wikipedia.org/wiki/Nobel_Prize_in_Physiology_or_Medicine) were first awarded in 1901. The prizes are widely regarded as the most prestigious awards available in their respective fields.

* 1. **General Description**

The Nobel Foundation, a private institution established in 1900, has ultimate responsibility for fulfilling the intentions in Alfred Nobel’s will. The main mission of the Nobel Foundation is to manage Alfred Nobel’s fortune in a manner that ensures a secure financial standing for the Nobel Prize over the long term and that the prize-awarding institutions are guaranteed independence in their work of selecting recipients.

The Nobel Foundation also strives to safeguard the prize-awarding institutions’ common interests and to represent the Nobel organisation as a whole. In the past two decades a number of outreach activities have been developed with the aim of inspiring and disseminating knowledge about the Nobel Prize.

* 1. **Objective and scope of the project**

The prime objective of this project work is to analyse the rich history of over 100 years of the Nobel Prize winners. This Project will give useful insights using various data visualisations of R language. The Nobel laureate data was acquired from the Nobel Prize API which includes a record for every individual or organization that was awarded the Nobel Prize since 1901.

1. **REQUIREMENT ANALYSIS**
   1. **Hardware Requirements:**

Processor Intel® Core™ i5-8250U CPU@ 1.6GHz

RAM 8.00 GB

GPU Nvidia GeForce MX150

GPU Size 4.00 GB

System type 64-bit operating system, x-64 based processor

* 1. **Software requirements:**

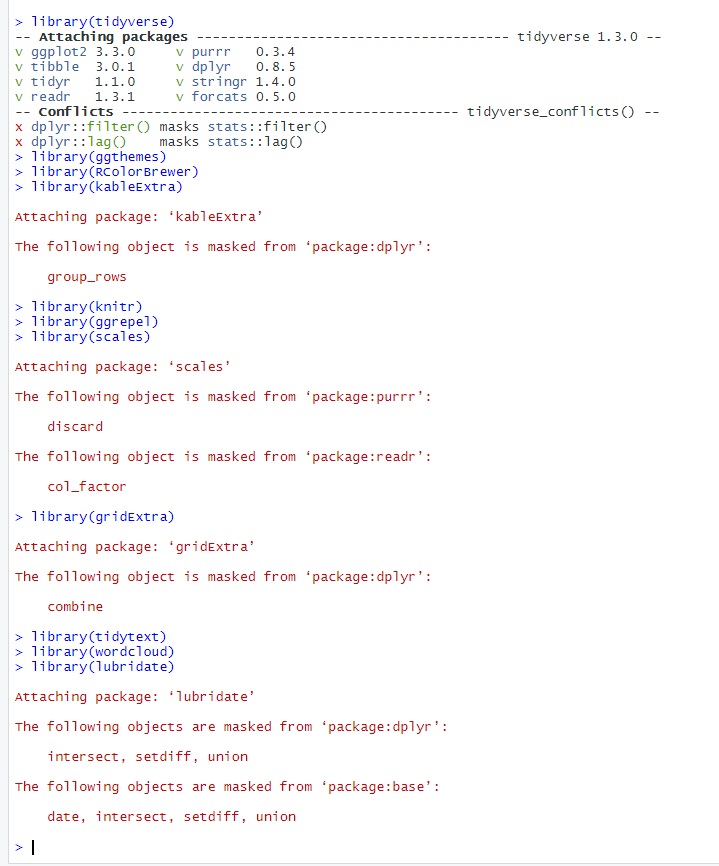
R lang.

RStudio

1. **DETAILED ANALYSIS & IMPLEMENTATION**

Following Libraries were used for working with the Nobel Data.

* **tidyverse**: It collects some of the most versatile R packages (ggplot2, dplyr, tidyr, readr, purrr, and tibble) which work in harmony to clean, process, model, and visualize data.
* **ggthemes**: It provides Extra Themes, Scales and Geoms for 'ggplot2'.
* **RColorBrewer**: It is an R package that contains a ready-to-use color palettes for creating beautiful graphics for different plots.
* **kableExtra**: It helps in building common complex tables and manipulate table styles.
* **knitr**: It is an engine for dynamic report generation with R. It is a package in the statistical programming language R that enables integration of R code into LaTeX, LyX, HTML, Markdown, AsciiDoc, and reStructuredText documents.
* **ggrepel**: It is an add-in for the ggplot2 package and it makes the text label show up nicely.
* **scales**: This package is to customise to control the appearance of axis and legend labels.
* **gridExtra**: It Provides a number of user-level functions to work with "grid" graphics, notably to arrange multiple grid-based plots on a page, and draw tables.
* **tidytext**: This package provide functionality to tokenize elements by commonly used units of text like these and convert to a one-term-per-row format.
* **wordcloud**: The word cloud generator package (wordcloud) are available in R to analyze texts and to quickly visualize the keywords as a word cloud.
* **lubridate**: It is an R package that makes it easier to work with dates and times.

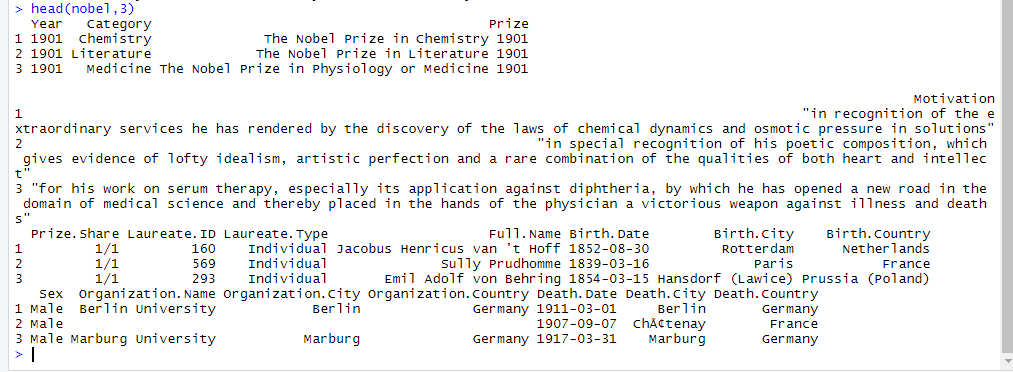


*Fig.1. Required Libraries*

* **DataSet Properties:**



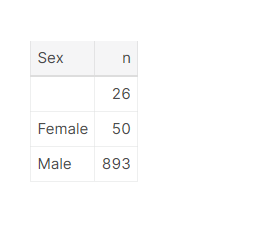
*Fig.2. Structure of Dataset*



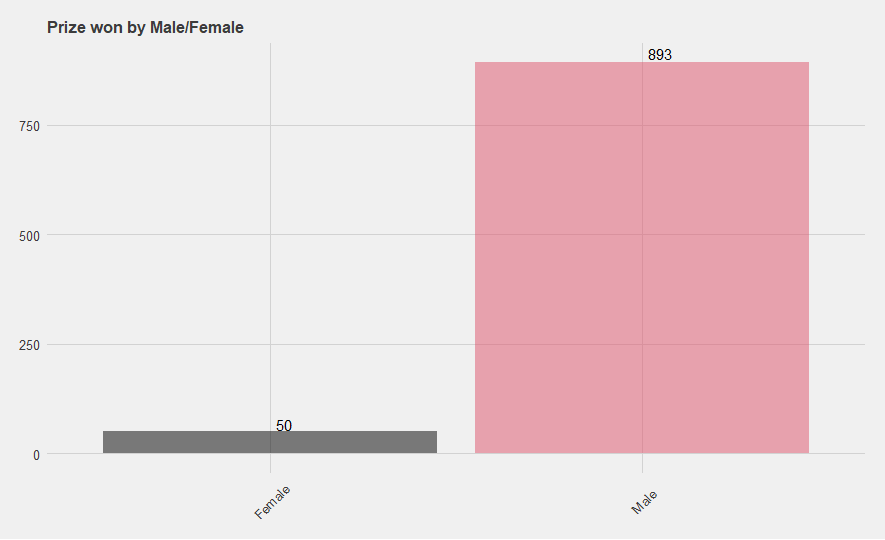
*Fig.3.The Noble Dataset*

* **Exploratory Data Analysis [EDA]:**
* **#Gender**

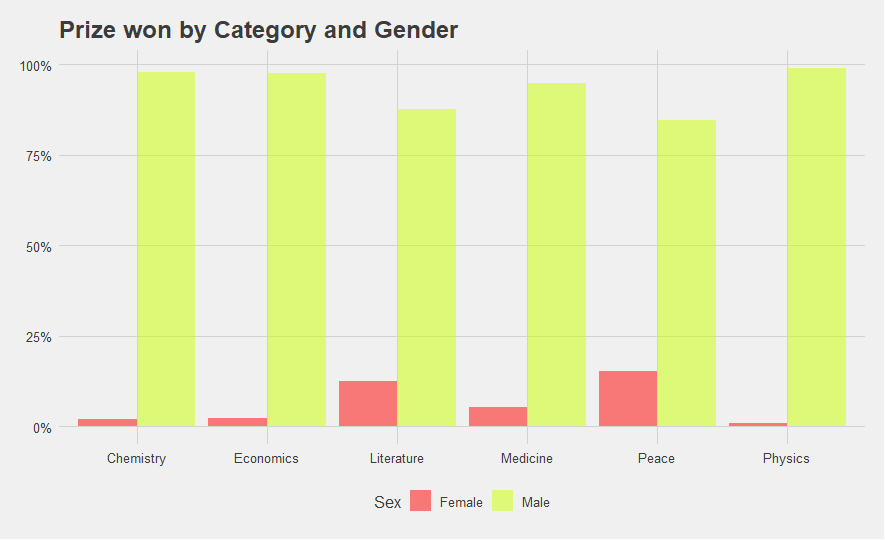
Finding gender ratio of Nobel prize winners.



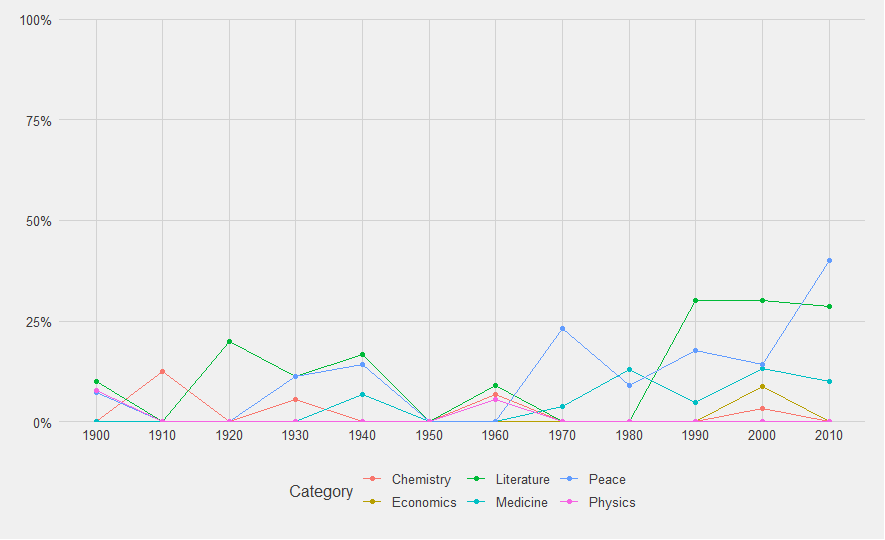
*Fig.4. Gender Ratio*



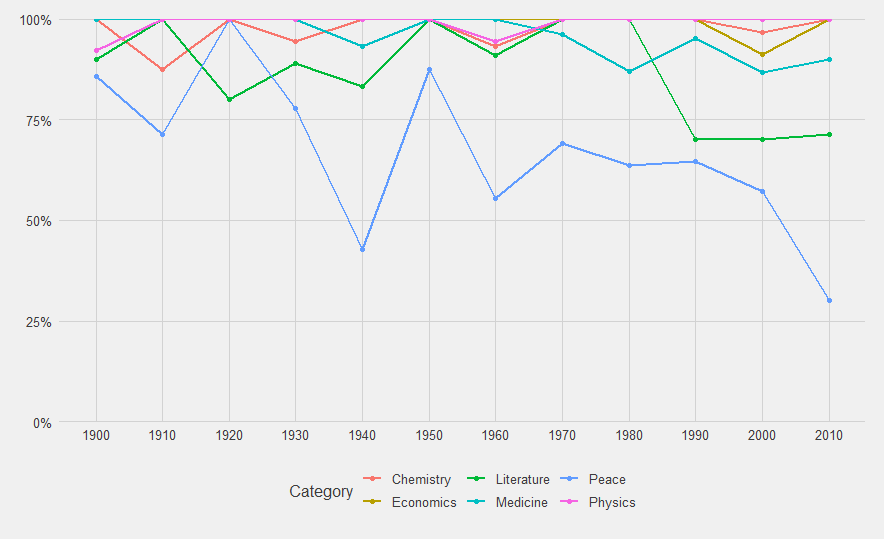
*Fig.5. Gender Ratio Plot*



*Fig.6. Category vs Gender plot*

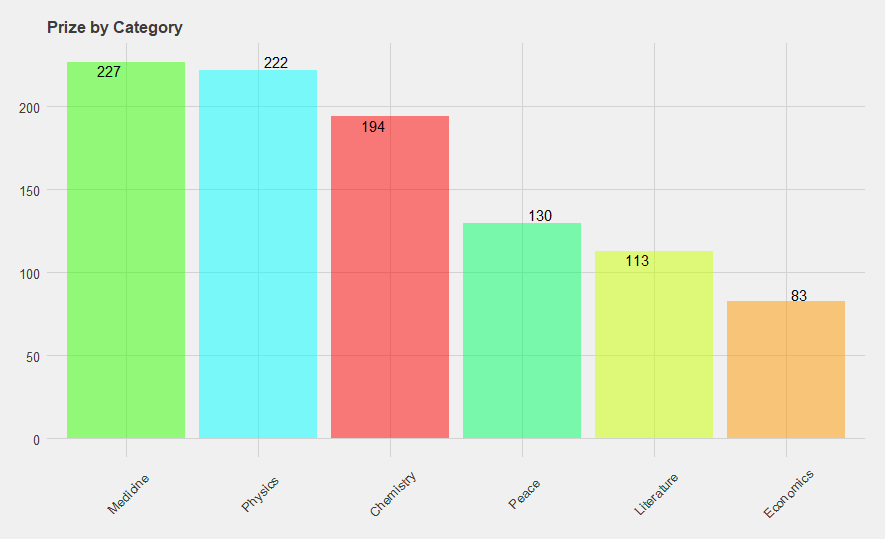


## *Fig.7. Female Laureates Proportion per decade*



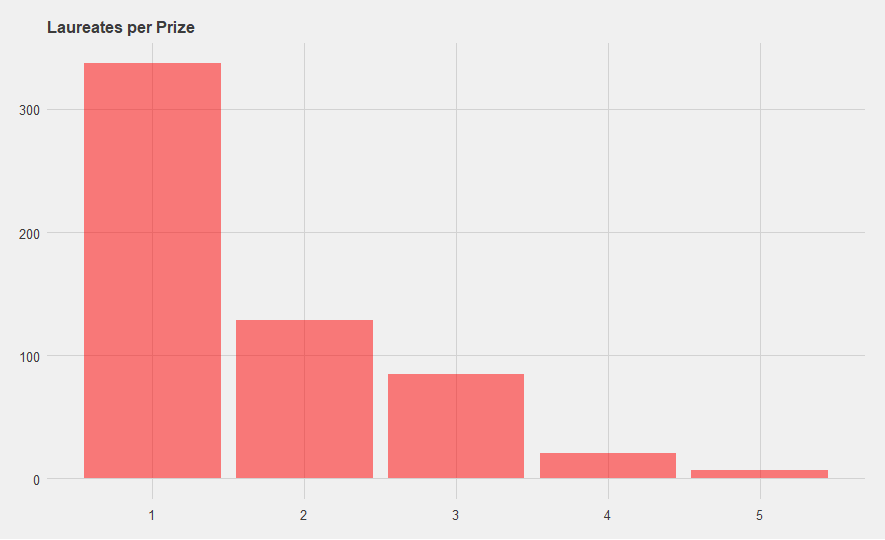
## *Fig.8. Male Laureates Proportion per decade*

* **#Category:**



*Fig.9. Prizes by category*

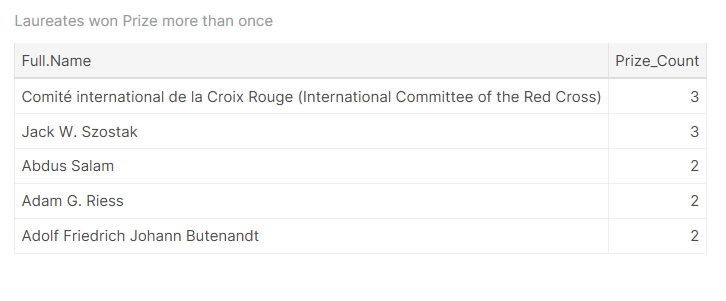
* **#Sharing Prizes:**



*Fig.10. Laureates count Per Prize*

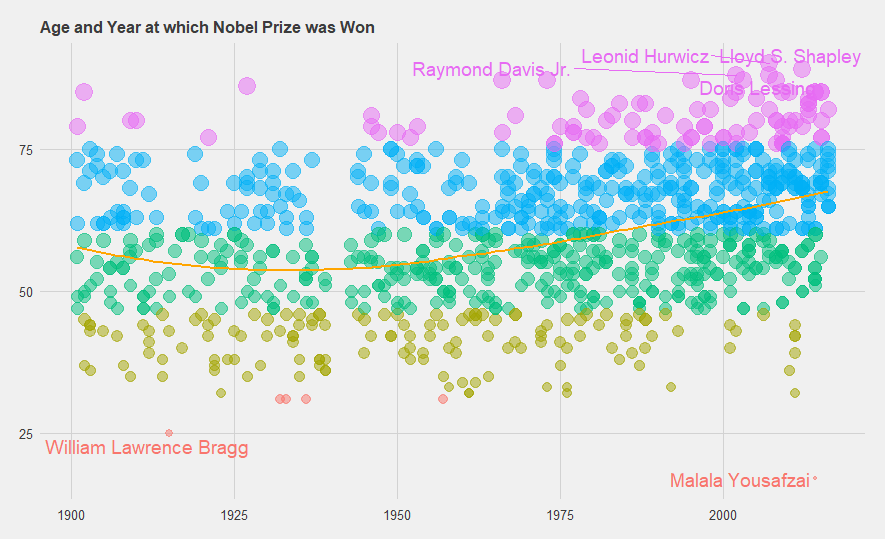
* **#Multiple Nobel Laureates:**

Laureates/organizations may receive Nobel prize more than once, for their work in same or different field/topic.

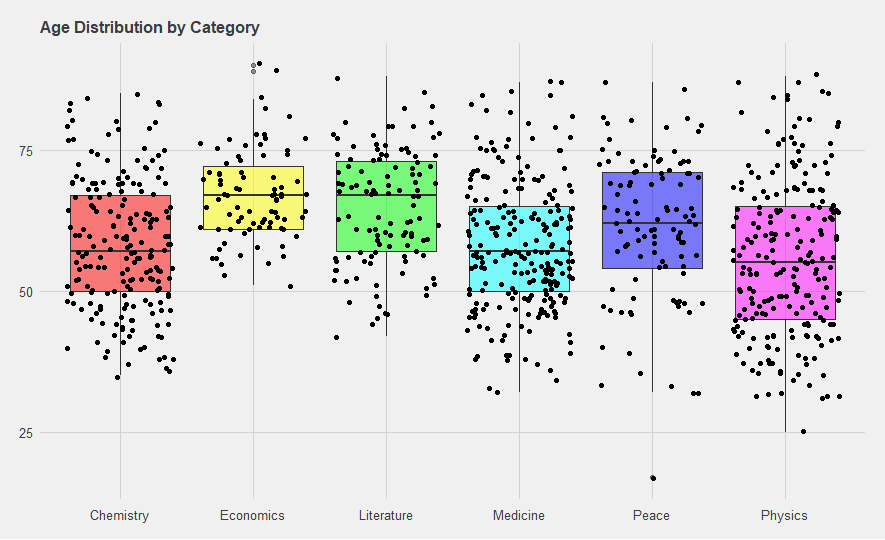


*Fig.11. Laureates with multiple Nobel Prizes*

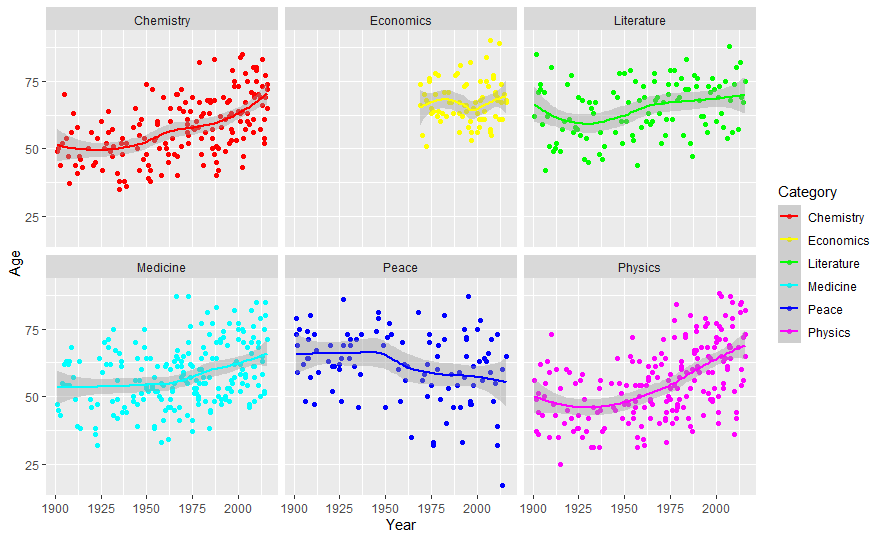
* **#Age:**



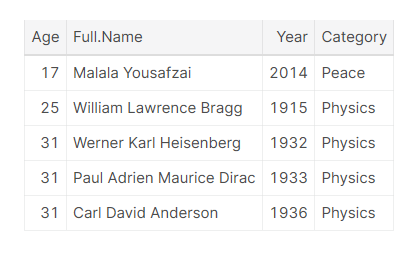
*Fig.12. Age vs Year of Nobel Prize Win*



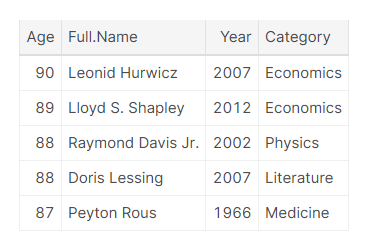
*Fig.13. Age Distribution by Category*



## *Fig.14. Age Trend for receiving Nobel price per Category*



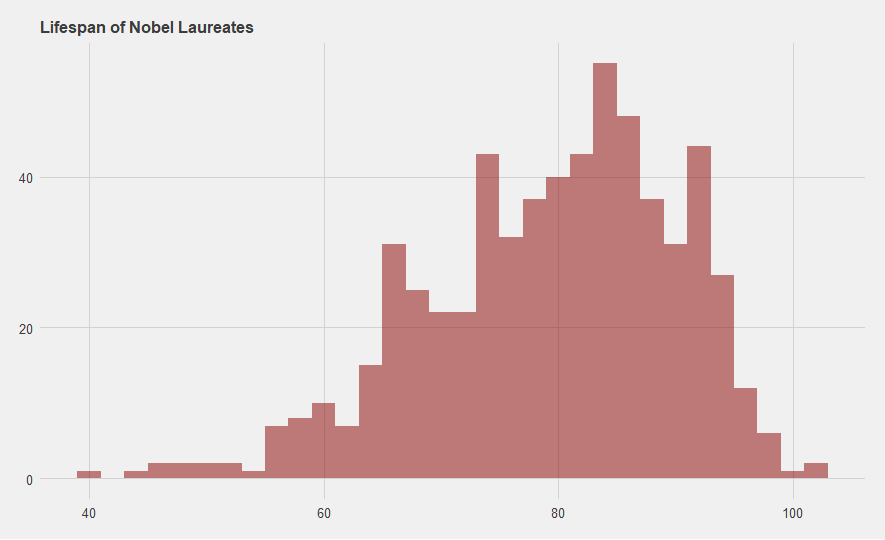
*Fig.15. Youngest Nobel laureates*



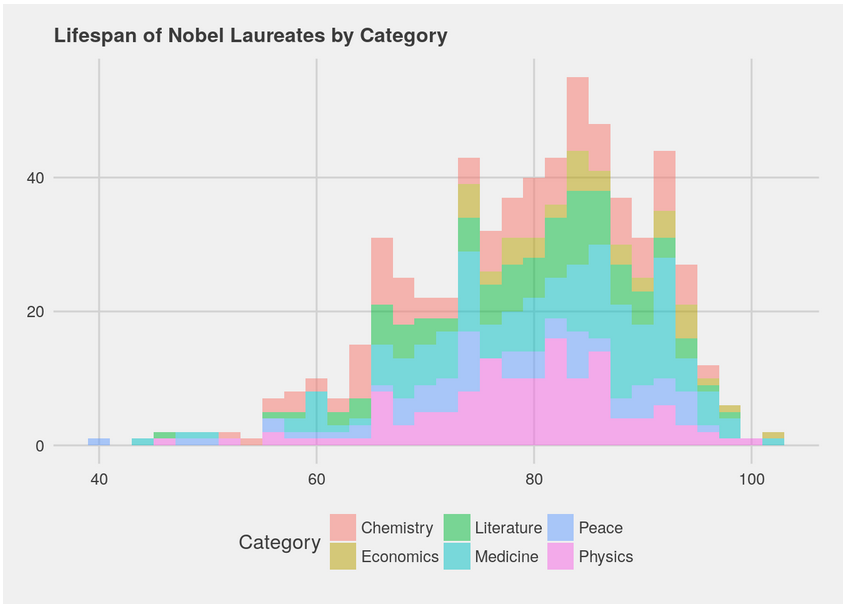
*Fig.16. Oldest Nobel laureates*



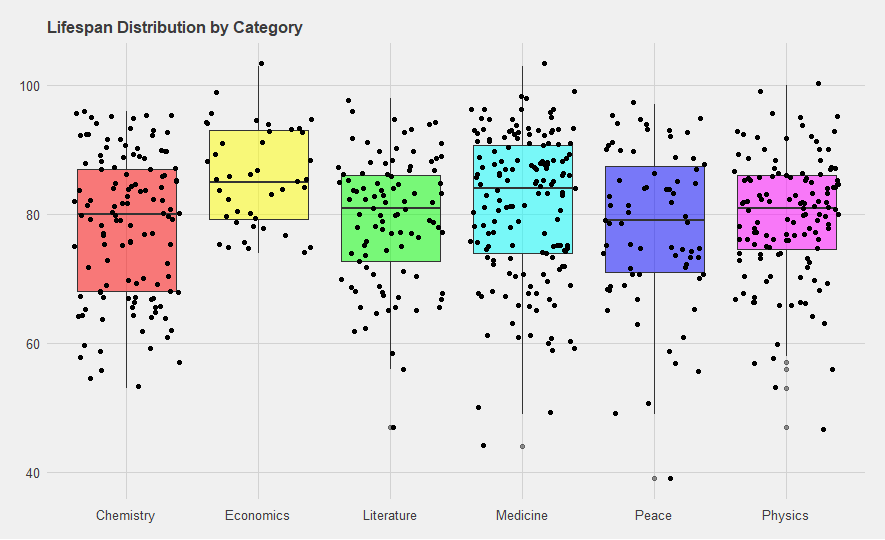
*Fig.17. First Female Laureate in each category*

* **#Life Span:**

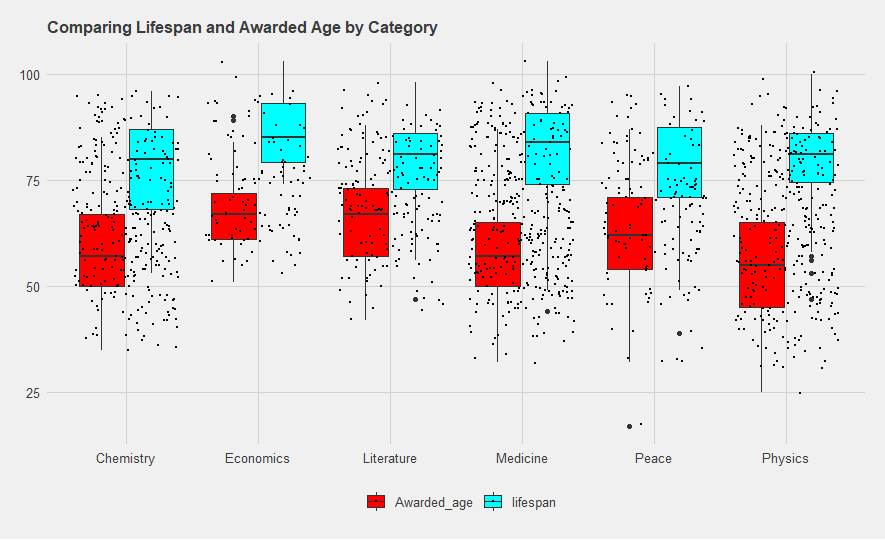
*Fig.18. Life Span of Nobel Laureates*



*Fig.19. Life Span By category*

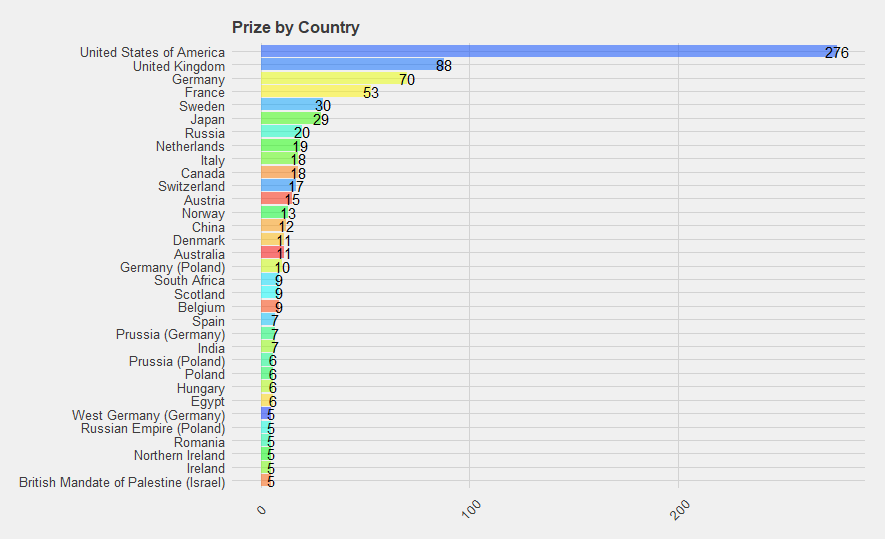


*Fig.20. LifeSpan Distribution by Category*

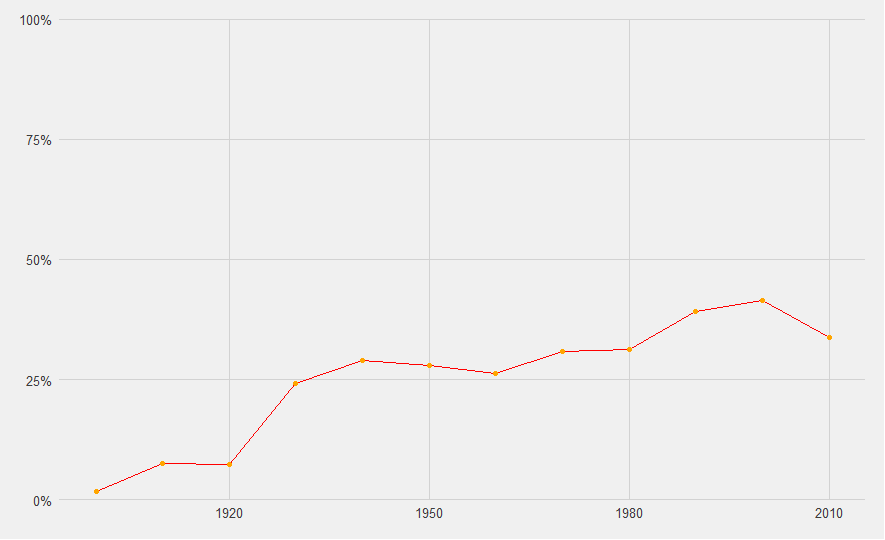


*Fig.21. LifeSpan Vs Awarded age*

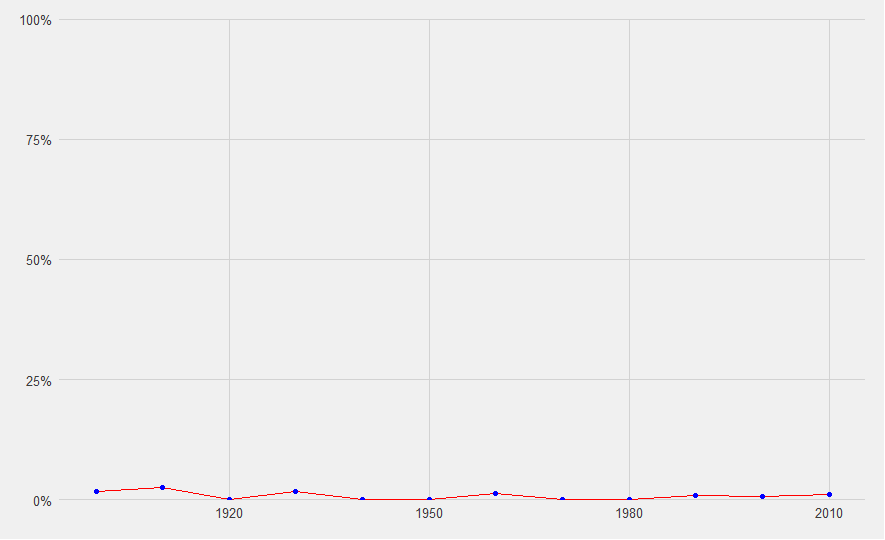
* **#Country:**



*Fig.22. Prizes by country*



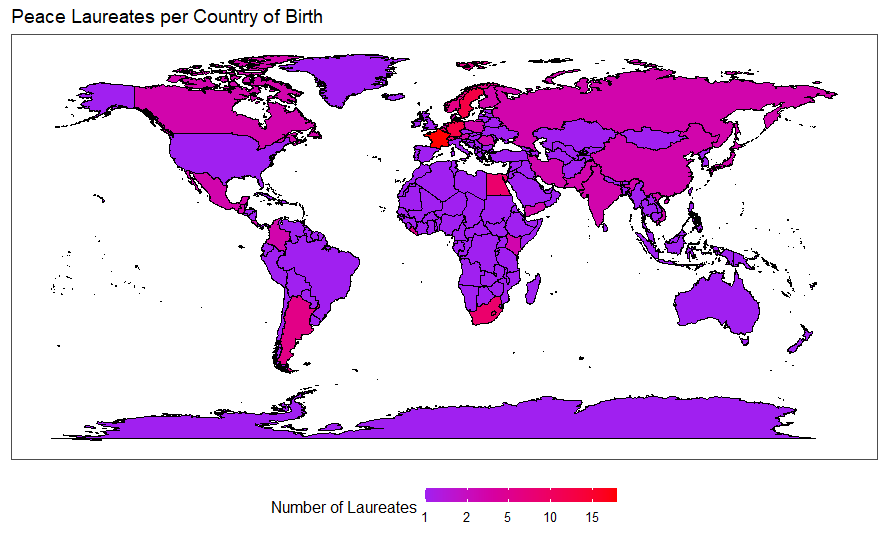
*Fig.23. USA-Year wise proportion*



*Fig.24. India-Year wise proportion*

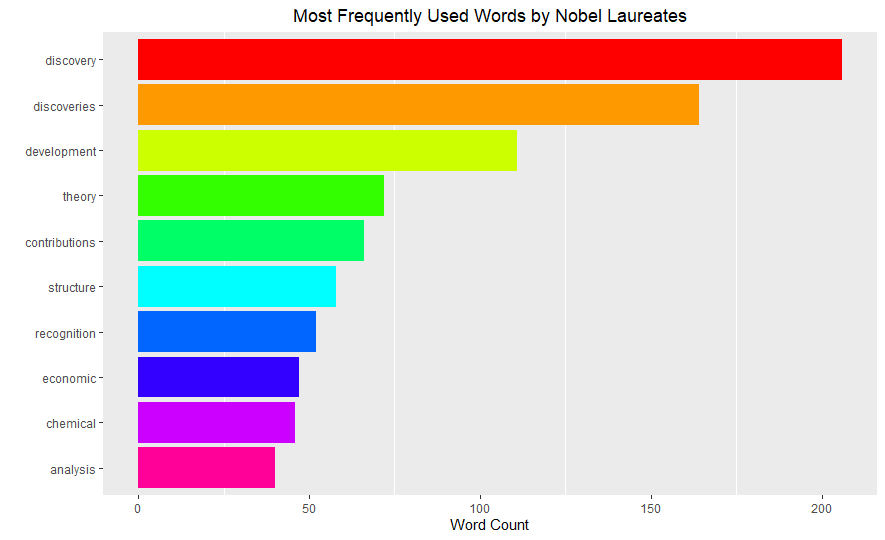


*Fig.25. Peace Laureates by Birth Country*

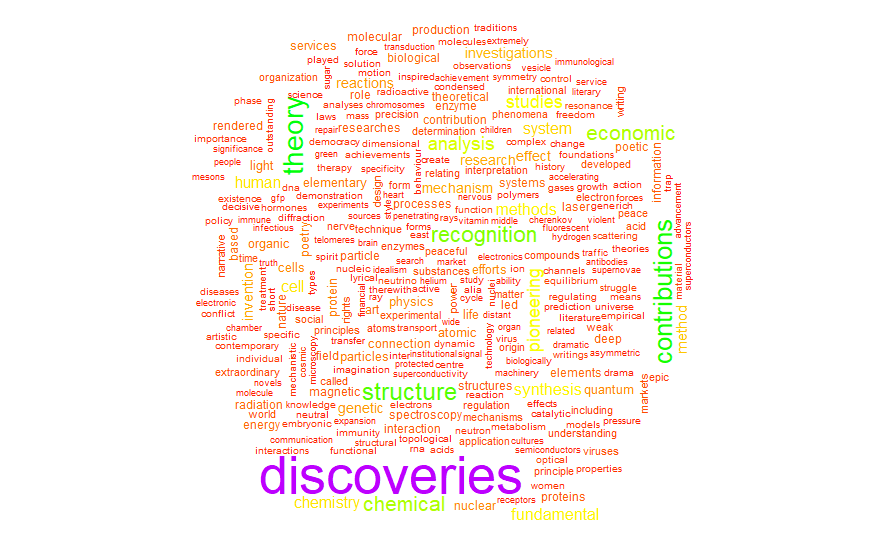


*Fig.26. Peace Laureates country map*

* **#Finding Prize Motivation:**



*Fig.27. Most frequently used words by Nobel laureates*



*Fig.28. Word Cloud of motivation key-terms*

1. **CONCLUSION**

The data was analysed and visualised to gain useful insights for a detailed study of over a century old rich history of The Nobel Prize winners. The insights such as country winning the most prizes in each category and most frequently written words in the prize motivation were analysed based on age, gender, and nationality of past Nobel laureates. This data can be used further to predict the next Nobel laureates based on the certain factors such as age, gender, nationality of nominees.