7PAM2000 Applied Data Science 1

Assignment 1: Visualizations

Sharon Zacharia

Student ID: 22074407

MSc Data Science (Sandwich)

Repository Link: https://github.com/SHARONZACHARIA/ADS-Visualization-Assgn-1.git

INTRODUCTION

In this Visualization Assignment the '*Nobel Prize from 1901 till 2020'* Data set will be analyzed, and three different types of visualization techniques will be applied.

The visualization techniques that are used in the data set were obtained from the public data set repository in Kaggle are Line plot, Pie chart and Bar chart. The Pyplot library in matplotlib is used for drawing the plots.

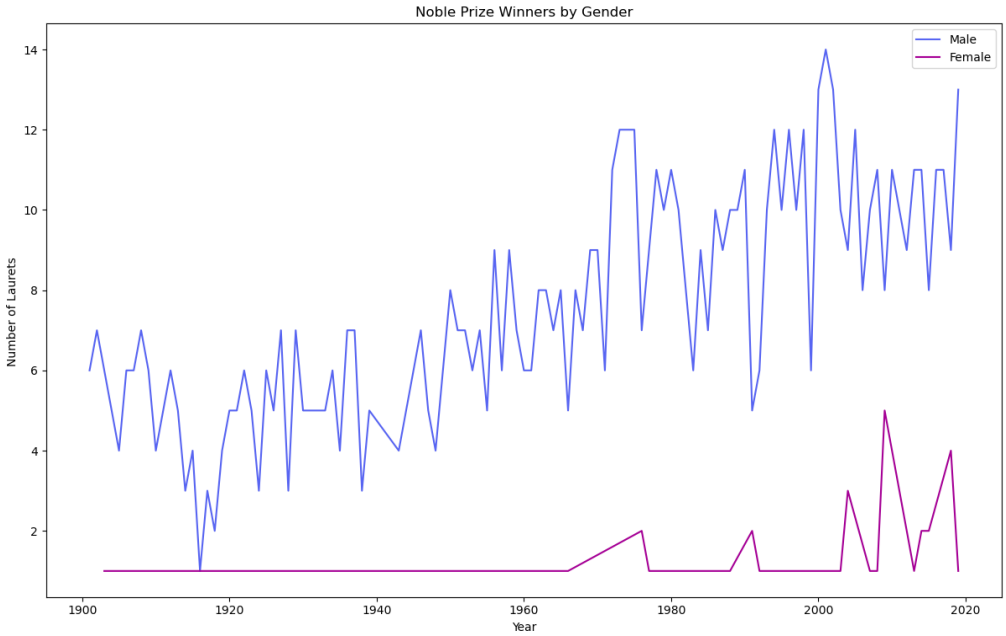
DATASET

The dataset was downloaded from the open-source dataset website "Kaggle." After doing adequate research a dataset that shows the list of Nobel prize laureates which includes the name of the laureate, born country code, gender, year, category and the age in which Nobel prize was obtained from Kaggle.

The source of the dataset is:
https://www.kaggle.com/datasets/bahramjannesarr/nobel-prize-from-1901-till-2020

Visualization 1: Line Plot (Nobel Prize Winners by Gender)

The plotted line graph shows the number of Nobel laureates from 1901 to 2020. The graph shows the variation in the number of male and female Nobel laureates over the period.



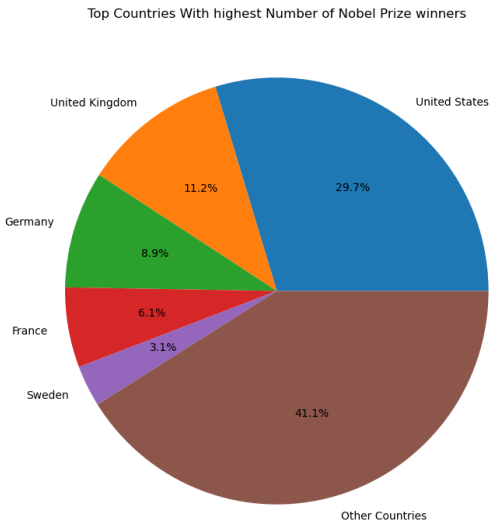
Year is taken along the x axis and count of laureates along y axis. From the graph we can infer that there were relatively few numbers of female Nobel prize laureates in the initial years and most of the laureates were men.

The graphics highlights the slight increase in number of female laureates after the year 2000.

The number of male laureates keeps on increasing from the beginning and the graph clearly shows the drastic gender gap in the fields where Nobel prize is awarded.

Visualization 2: Pie Chart (Top Countries with Highest Number of Winners)

The second Visualization technique used is Pie chart. Pie chart shows the relative size of the data using slices since it was used to show the number of laureates in each country. The plotted pie chart shows the countries with highest number of Nobel prize winners over the time. Each slice shows the size of the total number of winners from each country.

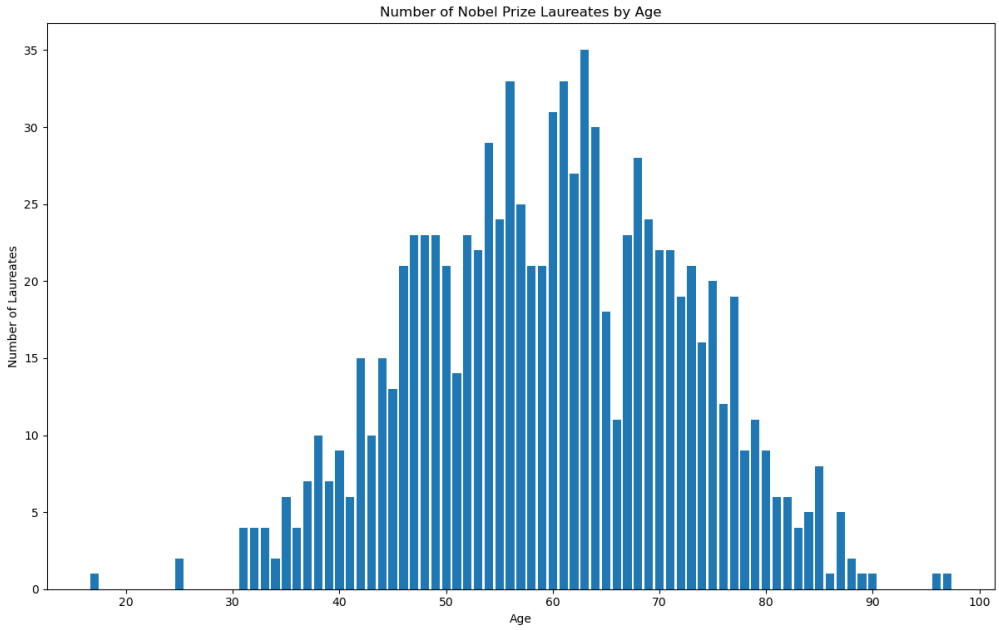


From the chart we can infer that the United States has the greatest number of Nobel prize winners with 29.7 % of the total share followed by United Kingdom with 11.2 %.

Germany, France and Sweden share the number of winners by 8.9%, 6.1 % and 3.1 % respectively and rest of the countries has a total share of 41.1%.

Visualization 3: Bar Chart (Number of Nobel Prize Laureates by Age)

The last visualization technique applied is the Bar chart which shows values for different category as bars, each category points to a single bar, and the length of each bar corresponds to the bar's value.



The plotted Bar graph shows total number of laureates in each age group.

Age data is marked across the x axis and Number of Laureates along the y axis.

From the Bar chart we can find out that most of the winners falls between the ages of 60- 65 and there are only few laureates below the age 30 and over 90 years old.

CONCLUSION

Three types of visualization techniques were applied on the selected data set and inferences were drawn from line graph, pie chart and bar chart. Pandas was used to read csv file and plots were drawn using matplotlib. Conclusions were drawn regarding the differences in the number of male and female laureates, Countries of Nobel laureates and number of laureates in each age group.