Group A3D: Domain and Data Selection

Group members: Andrea Bruttomesso 2120933
Alessandro Corrò
Davide Seghetto 2122548
Andrea Stocco 2108885

Links to the datasets: Nobel Laureates
Research Papers
Journal Ranking
Government budget allocations for R&D

Domain of Interest and Main Challenges

For our project we have chosen the domain of scientific research. Specifically, we aim to analyze potential correlations among Nobel Prize winners, their publications, and the research funding invested by various countries. This domain was selected because it allows us to reveal potential historical and geographical patterns in scientific research. For example, we want to examine whether, in certain years, the number of published papers in a specific field of science has increased or decreased, and identify in which states this has occurred. We will also analyze whether countries that invest more in Research and Development produce a higher number of scientific papers or Nobel Prizes, additionally, it is interesting to investigate whether Nobel Prize winners collaborate with each other or if their works inspire others through citations. Finally, we want to answer questions like: is a Nobel Prize topic something very studied in a year, or is it a winners's stroke of genius? Which is the venue that published more Nobel Prize author's papers? Which organizations funded the most important scientific discoveries?

Datasets

Nobel Laureates

The dataset contains information about the Nobel prize awarded from 1901 to 2016. The main data are:

- the year
- the category
- the winner
- the birthdate of the winner
- the birthplace of the winner
- the sex of the winner
- the winner organization
- the state of the organization

Research papers

The dataset contains information about research papers published between 1937 and 2017. For istance, we use this dataset in order to check whether a specific topic, winner of the prize, was something very studied during the year or it was totally a new discovery. They main data are:

- the year
- the title
- the authors
- the venue (journal or conference)
- the number of citations

Journal ranking

The dataset contains information about academic and research journals in which research articles relating to a particular academic discipline are published. We use this dataset in order to have more information about the venue of the research articles published by the winners of the Nobel. The main data are:

- overall rank
- the title
- the country
- the hirsh index
- the total number of references

Government budget allocations for R&D

This dataset contains the budget allocated by the major states from 1981 to 2023. We use this dataset to analyze, for example, how much more the countries that win the most Nobel Prizes spend on research compared to others.