



# Visual Mapping of Nobel Prizes

S. Aparicio<sup>1</sup> , R. Lousada<sup>1</sup> , P. Pires<sup>1</sup>, D. Gonçalves<sup>1</sup>  & A. Ferreira<sup>1</sup>

<sup>1</sup>INESC-ID, Instituto Superior Técnico, University of Lisbon

## Abstract

*Nobel Prizes tendencies can be an essential indicator of country development. The main objective of this work is to enable the visualization of the evolution of the chemistry and physics Nobel Prizes according to spatial, temporal, and relationship dimensions. To validate the conceptual proposal, we prototyped a visual dashboard that considers elements such as geography, age, and top-winning universities, throughout the years. While still work in progress, the proposed visualization allows the observation of several factors regarding Nobel Prize winners, namely some imbalances in gender, geographic distribution, and universities. We believe that further development will provide more valuable insights.*

## CCS Concepts

• **Human-centered computing** → *Graphical user interfaces; Interaction design process and methods; Visualization theory, concepts and paradigms; Visualization toolkits;*

## 1. Introduction

The development of a country is related to the concentration of talents, research investment, and Nobel Prizes [KKOP16]. Researchers observed that Nobel Prize attribution follows various patterns or biases [KFZP15], namely in Chemistry and Physics [NP18]. However, existing studies do not take advantage of data visualization techniques to provide more accessible insights into Nobel Prize attribution, such as the prominence of specific universities, countries, and even the relation between Nobel laureates and their supervisors. To that end, we developed the artifact on Figure 1 that enables the visualization of Chemistry and Physics Nobel Prizes since 1901 to 2018. We followed a design science research methodology [HC10] to undertake this objective. After creating a conceptual model for a Nobel Prizes dashboard, we developed an interac-

tive prototype that allows the exploration of the dataset we created by scraping Nobel Prize Official website and Wikipedia.

## 2. Conceptual Model

Inspired by Lupi's "Visual History of Nobel Prizes and Notable Laureates" published at "La Lettura" [Lup12], the Sunday cultural supplement of the Italian newspaper "Corriere della Sera", we devised an interactive dashboard to analyze the influence, find correlations and compare Nobel Prizes based on factors such as institution, advisory, nationality, living city, age, and longevity. On this initial prototype, we focused only on Physics and Chemistry, aiming at a straightforward way for the user to explore Nobel Prize awardees over time, together with a range of relevant related information.

## 3. Dataset

After collecting this data, we processed and cleaned the dataset using Pentaho Data Integration (<https://pentaho.com>) tool. For instance, to minimize loading time, we computed some derived features from the data, such as the age of the winner, number of winners per university, number of winners per country, university ranking, number of wins with a specific prize share, average winner age for each category, and bidirectional advisory relationships.

The resulting dataset consisted of a set of JSON files with a total size under 1 MByte, perfectly suitable for an online interactive tool. Even considering that we are only reporting on two Nobel Prize categories, we believe this scenario scales well for a broader study.

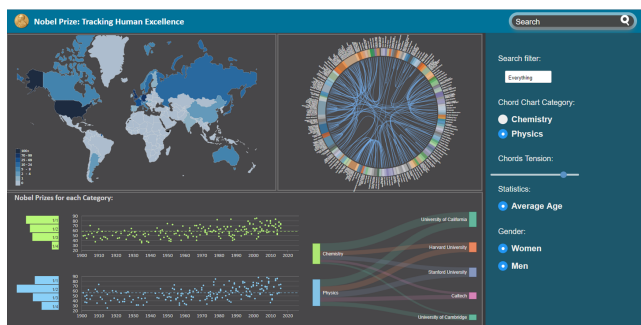


Figure 1: Overall Layout

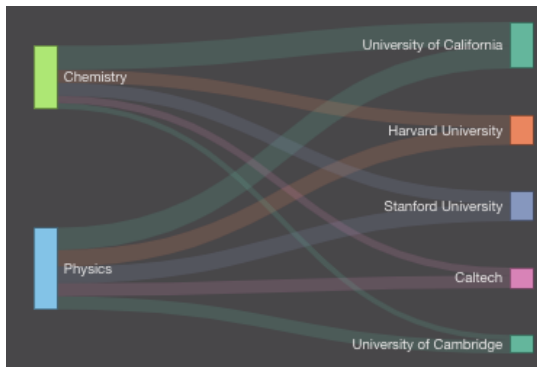


Figure 2: Sankey Diagram with top 5 universities

### 3.1. Visualization

We based the prototype for a Nobel Prize Winners Visual Mapping tool on a dashboard composed of two sections. The visualization section comprises seven linked visual idioms, while complemented by configuration panel. Each idiom answers individual questions, but a holistic view renders a deeper and more elaborate analysis. The background is a neutral black to prevent interference with other colors and the color scheme as a whole follows the readability guidelines proposed by Few [Few12]. Considering the conceptual model proposal, we chose to include the following linked idioms:

- Two Bar Charts to represent the distribution of Prize Shares on each category awards.
- Two Scatter Plots (3) to represent the relationship between time, and age of the winner, for each category.
- Sankey Diagram [RHF05, DV09, TML85, RO04], presented in Figure 2, which connects the multiple categories with the most awarded universities, and shows edges with thicknesses proportional to the number of wins. Colors for the category are those used in the scatter plots, and do not overlap with university colors (in future versions another encoding may be used for those to avoid confusion).
- Choropleth Map [Dup27, Gue33, CC09, AA99] displaying the geographical distribution of wins in the world.
- Chord Chart applying Hierarchical Edge Bundling [Hol06] for a selected category, which displays the relationships between winners and their students and advisors. In figure 4 the outer circle shows different colors for each organization. While selecting a specific winner in this chart, their relationships with their advisor (in green) and students (in red) are highlighted.

The connection between the seven idioms allow users to analyze in a simple dashboard. Some patterns are related to the laureates geographic position, nationality, institution, relationship, age or prize share. It is also possible to evaluate time, space and influence in specific idioms. To properly extract information from the complete dashboard, when a specific mark of the idiom is selected, other visualizations will highlight information regarding the user's selection. For example, when we study Nobel Prizes over time, it is possible to see that the age of the laureates tends to increase. To analyze the specific outliers, by hovering them, we can see connections regarding their country of origin, relations, or institution.

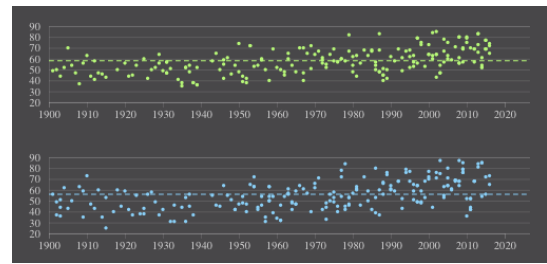


Figure 3: Scatter Plots representing awardees' age over time

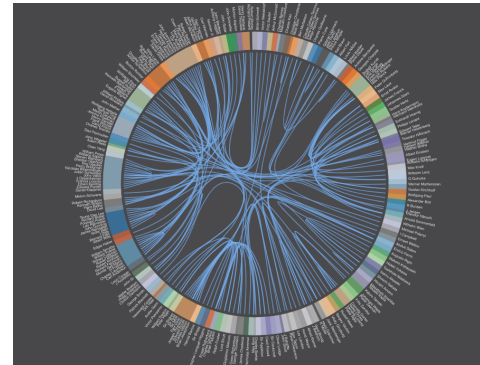


Figure 4: Advisors and Winners Relations' Chord Chart

The visualization also allows the user to zoom the Choropleth Map, brush the Scatter Plot, and rotate the Chord Chart.

### 4. Conclusions and Future Work

The developed artifact efficiently highlights how nationality, age, and advisories impact the probability to win the Nobel Prize. The scatterplot also allows the user to better analyze the influence of specific organizations and countries through time.

This work in progress may implement in the future, the rest of the Nobel prize categories such as Medicine and Physiology, Economic Sciences, Literature, and Peace. It should also be possible to configure the number of universities displayed in the Sankey Diagram, switch for an alternative chart if the current one is not helpful for the analysis, or even allow the full configuration of the dashboard. At the moment, this dashboard is only available for computer layout. Therefore, additional future work could generate a responsive design, adaptable to multiple platforms, as well as implement the search bar on the top-right corner. Finally, we plan to perform evaluations with real users to validate both the usability and efficiency of the Nobel Prize Winners Visual Mapping tool.

### Acknowledgements

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2019 and project VisBig - PTDC/CCI-CIF/28939/2017.

## References

- [AA99] ANDRIENKO G. L., ANDRIENKO N. V.: Interactive maps for visual data exploration. *International Journal of Geographical Information Science* 13, 4 (1999), 355–374. 2
- [CC09] CROMLEY R. G., CROMLEY E. K.: Choropleth map legend design for visualizing community health disparities. *International Journal of Health Geographics* 8, 1 (2009), 52. 2
- [Dup27] DUPIN C.: *Carte figurative de l'instruction populaire de la France*. Bachelier, 1827. 2
- [DV09] DIETMAIR A., VERL A.: A generic energy consumption model for decision making and energy efficiency optimisation in manufacturing. *International Journal of Sustainable Engineering* 2, 2 (2009), 123–133. 2
- [Few12] FEW S.: Visual business intelligence – black or white: What color works best for the background of a screen?, 2012. URL: <https://www.perceptualedge.com/blog/?p=1445>. 2
- [Gue33] GUERRY A.-M.: *Essai sur la statistique morale de la France*. Clearwater, 1833. 2
- [HC10] HEVNER A., CHATTERJEE S.: *Design research in information systems: theory and practice*, vol. 22. Springer Science & Business Media, 2010. 1
- [Hol06] HOLTEN D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 741–748. 2
- [KFZP15] KIM M. C., FENG Y., ZHU Y., PING Q.: Quantitative exploration into the diffusion process of creative ideas in economics: Nobel prize laureates. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–4. 1
- [KKOP16] KERR S., KERR W., OZDEN C., PARSONS C.: Global talent flows. *Journal of Economic Perspectives* 30, 4 (9 2016), 83–106. doi: 10.1257/jep.30.4.83. 1
- [Lup12] LUPI G.: A visual history of nobel prizes and notable laureates, 1901-2012, 2012. URL: <http://giorgialupi.com/lalettura>. 1
- [NP18] NILESH B., PRANAV B.: Statistical analysis of nobel prizes in physics: From its inception till date. *Journal of Physical Studies* 22, 3 (2018). 1
- [RHF05] RIEHMANN P., HANFLER M., FROEHLICH B.: Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. (2005), IEEE, pp. 233–240. 2
- [RO04] RASER J. M., O'SHEA E. K.: Control of stochasticity in eukaryotic gene expression. *science* 304, 5678 (2004), 1811–1814. 2
- [TML85] TOUCHETTE P. E., MACDONALD R. F., LANGER S. N.: A scatter plot for identifying stimulus control of problem behavior. *Journal of Applied Behavior Analysis* 18, 4 (1985), 343–351. 2