Mapping Retention Rates in Informatics Studies across European Countries

Daniel Schnurpfeil[a]

[a]student of the master degree program at Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Plzeň, Czech Republic

Abstract

Informatics studies are essential for technological progress in Europe, but student retention rates may limit their impact. This study develops interactive visualization to investigate the factors influencing retention rates in different European countries. We introduce interactive visualizations built on Vega[1] framework to analyze retention data collected from national education statistics, surveys and institutional reports. This approach provides users with a dynamic platform for exploring complex datasets, ultimately providing a comprehensive picture of trends and differences in retention within European computing education.

Keywords: student retention rates, interactive visualization, vega, map, charts, economic visualization

1. Introduction

This study investigates the factors influencing student retention rates within European informatics programs. This study develops interactive visualizations to analyze retention data across diverse European countries. By using national educational statistics, surveys, and institutional reports,[2] we aim to create a comprehensive picture of informatics education retention. These interactive visualizations, built upon Vega[1] framework, provide a dynamic platform for users to explore complex datasets. This facilitates deeper understanding of retention trends within the European informatics education landscape.

The study also examines how retention rates differ between men and women. It examines the impact of the COVID-19 pandemic on retention rates and hypothesizes different impacts due to the disruption of the pandemic. In addition, it examines whether clusters of countries share trends in student retention and seeks to uncover systemic influences. These findings serve as a basis for strategies to improve student retention in European computer science programs.

2. Related Work

Although several studies have looked at visualizing educational data, especially in European computer science education, the task of visualizing graduation rates is different from tracking student retention. The Informatics Education Higher Education (IEHE) data portal [2] offers a valuable resource for examining trends in European informatics education. This portal provides data on the number of students enrolled, degrees awarded, and gender distribution in different countries. It uses maps, tables and graphs and allows users to break down these trends by country, programme level (bachelor, master, doctorate) and year.

3. Methods

This section details the data processing, visualisation tools and analytical techniques used to investigate the factors influencing student retention rates in European computer science programs.

3.1. Data Acquisition and Preprocessing

The study utilizes data on student enrollment and retention rates for bachelor's programs in informatics across various European countries. The data source consists of an Excel spreadsheet containing multiple columns:

- Institution Type: Categorizes institutions as research universities (RU), universities of applied sciences (UAS), or combined (RU+UAS).

- country: Specifies the country for each data point.
- year: Represents the academic year of the data.
- types of gender: Includes separate entries for total, female, and male students.

A Python script facilitates the conversion of this data from its initial Excel format into a structured JSON format. This conversion into JSON offers several advantages:

- Structured format: The JSON schema enforces a consistent data structure, simplifying analysis and visualization.
- Single file: By consolidating data into a single JSON file, it becomes readily accessible for visualization tools like Vega.
- Separation of concerns: The schema separates data organization from specific visualizations, promoting flexibility and code re-usability.

3.2. Retention Formulas

The provided data models are computed by following given formulas:

- first model - Retention Rate:

$$R = \frac{S[X+1]}{S[X] - D[x] + F[x+1]}$$

- second model - Year-on-year Retention Rate:

$$R = \frac{S[X+1] - F[x+1]}{S[X] - D[x]}$$

where R is retention rate, S is amount of students in certain year X, D is amount of graduated students for certain year X and F are first year students.

3.3. Visualization Tools and Techniques

The dashboard incorporates user-friendly filtering options to enable focused exploration across various dimensions:

- Year Slider: A slider allows users to dynamically select a specific academic year or range of years for analysis.
- Gender drop-down menu: Users can select a specific gender (total, male, female or female-male comparison) to filter the visualizations and examine retention trends for that group.
- Institution type drop-down menu: A drop-down menu allows users to filter data based on institution type (RU, UAS, or RU+UAS), enabling comparisons between different types of institutions.

These filter options dynamically update all visualizations within the dashboard, ensuring a cohesive exploration experience. Interactive Map The dashboard features an interactive map that provides a geographical context for student enrollment data. Users can:

- Zoom and Scroll: Users can navigate the map to explore retention trends across different European regions.
- Hover Interactions: Hovering over a specific country reveals additional details such as overall retention rates, gender breakdowns, and institution type distribution.
- Shift + Click Selection: shift + clicking on a country adds or removes it from a group displayed in the dynamic charts, allowing users to focus on specific regions for comparison.

The map utilizes a color-coded scheme to represent retention rates with scaled values for better visual differentiation:

- Gray: Represents countries with no available data for the selected filters.
- Blue: Denotes countries with retention rates within a pre-defined normal range.
- Red: Highlights countries with retention rates deviating significantly from the norm, potentially indicating outliers that warrant further investigation.

This color scheme facilitates rapid identification of potential areas of concern or interest.

3.4. Dynamic Bar Chart

The dashboard incorporates dynamic bar charts to visualize retention rates across years for a user-selected group of countries. This group selection is linked to the interactive map through shift + clicking. The charts display data for a maximum of seven countries within the chosen group. For groups exceeding seven countries, labels will be displayed for the top three, while the remaining data

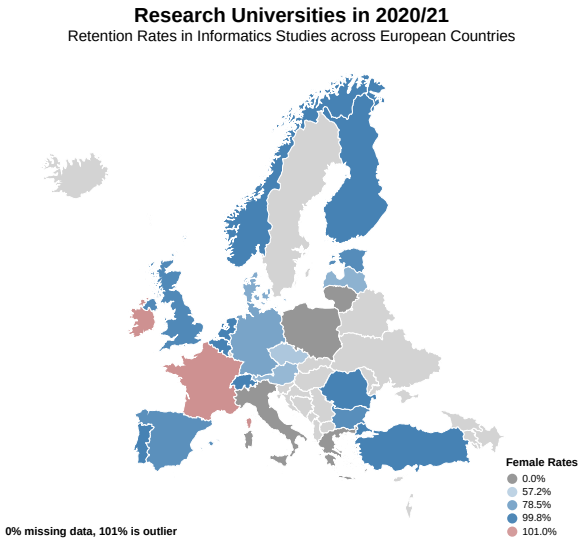

Figure 1: Map Example

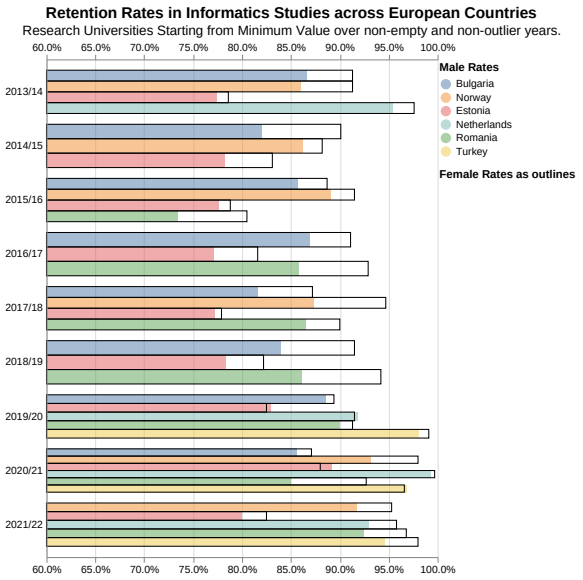

Figure 2: Countries with better female rates in second retention model

is still included but not explicitly labeled. This approach balances information density with readability. Outliers and empty data points (countries with no data for the selected filters) are excluded from the bar charts to ensure clarity and focus on the core trends.

3.5. Dynamic Line Charts

The dashboard also includes a line chart that provides a high-level overview of data completeness across different countries and years. This allows users to quickly identify any missing data points that might require further investigation or data collection efforts.

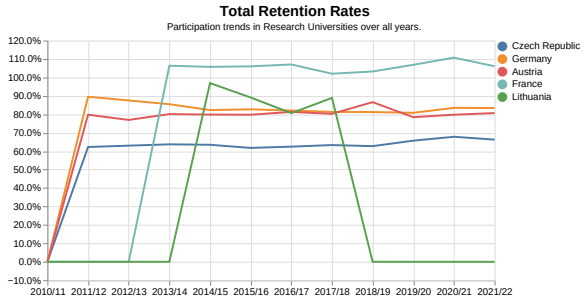

Figure 3: Line Chart Example

4. Results

The dashboard revealed key trends in the retention of computer science students in Europe. Based on the analysis from the first retention model, we've observed a small decrease in retention rates across several European countries (Figure 4) in academic year 2020/2021 when the COVID-19 pandemic has begun.

Although the overall analysis did not reveal any significant differences in retention rates between genders in Europe (based on Retention Model 2), examining specific countries and program types could yield more nuanced results. (Figures 2 and 5).

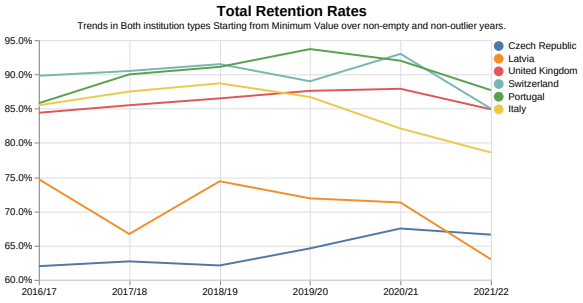

Figure 4: COVID-19 affection in first retention model

The provided data have shown no significant clusters. However, the lower rates for Czech Republic are seen in all available data (Figure 6). The source

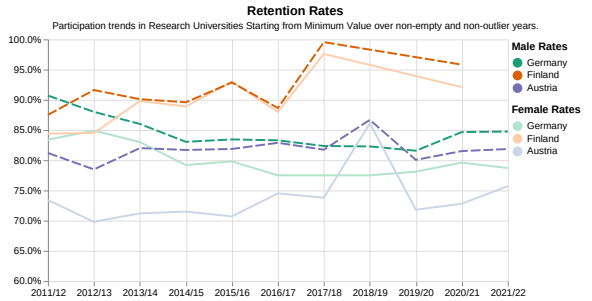

Figure 5: Countries with better male rates in second retention model

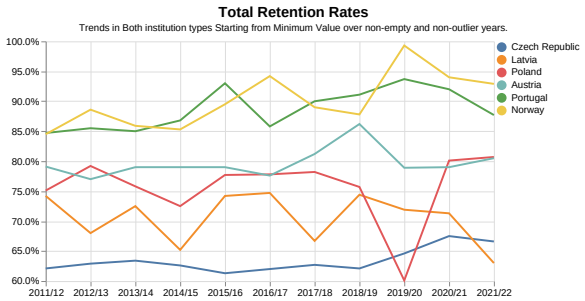

Figure 6: Clusters in second retention model

code can be found at GitHub[1] repository.

5. Conclusion

In conclusion, this study and interactive visualization development provided valuable insights into the complex landscape of student retention in European informatics programs.

It's important to note that while the interactive dashboard relied on the Vega visualization framework, data transformation for the analysis was performed mainly in pure JavaScript (JS). This approach offered greater flexibility and control over the data manipulation process compared to using Vega's built-in data transformation capabilities regardless of the higher amount of time spent by developing this visualisation.

5.1. Limitations & Future Work

While the current dashboard provides valuable insights, future iterations could benefit from expanding its functionality. Including additional year filters would allow users to explore historical trends and identify long-term patterns in student retention. Additionally, implementing responsive design principles would optimize the dashboard for smaller screens and mobile devices, enhancing accessibility and user experience. This would be particularly valuable for users who want to zoom in on the interactive map for a closer look at retention trends across different regions.

Acknowledgement

I would like to express my thanks to my supervisor, Doc. Ing. Přemysl Brada, MSc., Ph.D., for his guidance throughout the project and Informatics Europe for providing the valuable data that formed the foundation of this research.

[1] https://github.com/danschnurp/kiv-vi

References

[1] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, J. Heer, Vega: A visualization grammar, 2016. URL: http://vega.github.io/.

[2] E. Di Nitto, I. García-Varea, M. Jazayeri, D. A. Tamburri, S. Tikhonenko, Informatics higher education in europe: A data portal and case study, Commun. ACM 66 (2023) 61–67. URL: https://doi.org/10.1145/3591107. doi:10.1145/3591107.