QS World University Rankings 2025

Visual Analytics 2025/2026

Dariga Shokayeva (2242443), Yelizaveta Tskhe (2224260)

Introduction

Global university rankings are widely used by prospective students and researchers to compare higher education institutions. Among these, the QS World University Rankings comprises multiple indicators, including academic reputation, employer reputation, citation impact, internationalization, and sustainability, which are combined into a single overall score and a final rank. While rankings provide a convenient summary, they often overlook the underlying multidimensional characteristics of universities, making it difficult to understand why institutions differ or cluster in certain positions.

From a visual analytics perspective, university ranking data represent a high-dimensional dataset in which each institution is described by multiple indicators. Traditional tabular representations are insufficient for exploratory analysis, as they do not support the identification of patterns or similarities between institutions beyond their rank. Users may be interested in comparing groups of universities with similar profiles, identifying outliers, or understanding how different indicators contribute to overall positioning.

The goal of this project is to design and implement an interactive visual analytics system that supports exploratory analysis of the QS World University Rankings 2025 dataset. The system integrates dimensionality reduction techniques with coordinated visualizations to enable users to explore the multidimensional structure of the data. By combining a two-dimensional projection obtained via Principal Component Analysis (PCA) with interactive profile-based visualizations, the system allows users to explore beyond ranking comparison and instead focus on characteristics and performance patterns.

The intended users of the system include prospective graduate students comparing universities based on multiple criteria, as well as university administrators interested in benchmarking institutions. The system is designed to support tasks such as identifying clusters of similar universities, comparing selected institutions across performance indicators, and deriving insights about trends in higher education performance.

Dataset

The dataset used in this project is based on the QS World University Rankings 2025 from a Kaggle repository. The dataset contains information about 1,503 universities worldwide and includes both ranking information and a set of performance indicators used by QS to compute the overall score.

Each university is described by a combination of attributes and indicators reflecting different dimensions of academic performance. After preprocessing and cleaning, the dataset includes the following attributes:

- Academic Reputation
- Employer Reputation
- Faculty Student Ratio
- Citations per Faculty
- International Faculty
- International Students
- International Research Network
- Employment Outcomes
- Sustainability
- QS Overall Score
- 2025 Rank

To support dimensionality reduction and exploratory analysis, ordinal ranking variables (2024 and 2025 ranks) were excluded from the PCA computation, as they represent derived summaries of the same indicators and would introduce redundancy and bias.

The resulting dataset has the Angelini-Santucci index (AS) of about 13,000, which falls within the required range for the assignment.

Task Abstraction

The system supports a set of exploratory and comparative tasks aligned with the intended users' goals:

- T1 – Explore similarity and clustering
- T2 - Compare individual institutions
- T3 – Analyze selected groups
- T4 – Identify distinguishing features
- T5 – Quantify similarity and deviation

These tasks emphasize exploratory and comparative analysis rather than reproducing the published ranking order.

Analytical Goals

The primary goal of the system is to support profile-based reasoning about universities as multidimensional items. By combining dimensionality reduction and interactive selection, the system enables users to uncover patterns and relationships that are not visible in traditional ranking tables.

Specifically, the system allows users to:

- Explore the global structure of university performance using a PCA-based projection
- Compare individual universities against both global averages and nearest neighbors in feature space
- Understand which indicators most strongly characterize or differentiate selected groups
- Quantify similarity using distance-based metrics rather than relying only on rank

Related Work

Visual analytics has been applied to higher-education and ranking datasets, specifically to help users interpret complex performance indicators across universities. Aljehane et al. (2015) developed a visualization platform for the Times Higher Education Top 400 University rankings that integrates maps, bar charts, and score-based comparisons across teaching, research, citations, international outlook, and industry income [1]. Their system proves the need for interactive interfaces to navigate multi-attribute university rankings. However, they do not use dimensionality reduction triggered by user selection. Muzamhindo et al. (2017) investigate the use of Principal Component Analysis (PCA) on global university features to derive ranking dimensions [2]. Their findings show that PCA components capture dominant variance patterns in ranking metrics, such as global reputation and research influence, but their work remains non-interactive and without visual exploration or coordinated views.

Research on multi-attribute ranking visualization provides a technical foundation relevant to our system design. LineUp by Gratzl et al. (2013) is a system for visual analysis of ranked multidimensional items, which allows users to manipulate weights and observe how rankings change accordingly [3]. While it offers strong support for ranking transparency and attribute contribution explanation, its mechanism does not rely on dimensionality reduction. The RankBooster by Puri et al. (2020) is a visual analytics tool, specifically for university rankings, that focuses on ranking prediction and comparison between universities [4]. It includes multiple coordinated views and higher level analytics, but it concentrates more on predictive ranking changes than data exploration.

Dimensionality reduction methods, such as PCA and t-SNE, are widely used in visual analytics to reveal the high-dimensional structure. The system that is used in the education sector is SPEET (Dominguez et al., 2018) [5]. It applied PCA and t-SNE-based visualizations to analyze student performance. It demonstrates that dimensionality reduction scatterplots indeed uncover latent similarity structures that are otherwise difficult to detect through traditional tabular or bar chart visualizations.

Our project combines these three research directions (ranking visualization, multi-attribute analysis, and dimensionality reduction techniques). Unlike LineUp or RankBooster, which emphasize ranking sensitivity and prediction, our system focuses on structural analysis of the QS 2025 University Rankings through PCA. Unlike previous ranking visualizations, we use analytics triggered by user interactions, such as nearest neighbor computation in the feature space and dynamically computed summaries for selected subsets. Moreover, previous work usually uses dimensionality reduction as a static embedding, whereas our system integrates PCA into the analysis flow as a central mechanism for exploration, coordinated with a profile view that allows deeper interpretation. The resulting tool offers an analytically meaningful environment for understanding similarity patterns and performance profiles across global universities.

Visual Analytics Cycle

The design of the system follows the standard Visual Analytics Cycle, which combines automated analysis techniques with interactive visual representations and user-driven exploration.

The pipeline is:

1. **Data Transformation**

Raw QS ranking data is cleaned, converted to numeric form, and standardized.

2. **Automated Analysis**

PCA is applied to reduce the dimensionality of the dataset from nine performance indicators to two principal components, preserving the variance structure.

3. **Visual Mapping**

The PCA results are visualized in a two-dimensional scatterplot, providing an overview of similarities and clusters among universities. Additional coordinated views are used to show detailed profiles.

4. **Interactive Exploration and Analytics**

Users interactively select individual universities or groups on the PCA plot. These interactions trigger computations, including aggregation, distance calculation, and feature comparison, which are reflected in the coordinated views.

This enables users to explore patterns and derive insights by combining visual perception with analytical computation.

Dimensionality Reduction

The system uses Principal Component Analysis (PCA) as a dimensionality reduction technique. PCA is applied to the nine performance indicators describing each university. The first two principal components are used as the basis for the main overview visualization. These components capture the dominant variance in the dataset and provide a meaningful low-dimensional embedding in which institutions with similar performance profiles are positioned close to each other. The PCA is treated as a central element of the analysis flow. Users begin their exploration on the PCA map, identify regions of interest, and then zoom into detailed views that explain the observed spatial relationships.

Visual Encodings and Coordinated Views

The system is composed of three coordinated visual components:

- PCA Scatterplot

The primary view is a two-dimensional scatterplot representing the PCA projection of universities. Each point corresponds to a university, positioned according to its values on the two principal components. Color encodes the geographical location of the institution, supporting comparison across regions, while point size reflects academic reputation to provide an additional visual cue. This view provides an overview of the global structure of the dataset, revealing clusters of similar universities and outliers with distinct profiles.

- Radar Chart (Profile View)

The radar chart provides a profile view of selected universities. For group selections, the chart displays individual institution profiles alongside the mean profile of the selection and the global mean, enabling comparison across all performance indicators.

- Bar Chart

A horizontal bar chart explicitly highlights analytical differences. For group selections, the bar chart displays the features that most strongly differentiate the selected group from the global mean. For single university selections, it shows the absolute indicator values of the selected institution.

Interaction Techniques and Analytics

The system supports a set of interaction techniques that trigger analytical computations:

- Direct selection

Users can select universities using box or lasso selection on the PCA scatterplot to define custom groups, or click on individual points to explore single institutions.

- Selection-driven aggregation

For group selections, the system computes aggregate statistics, including mean indicator values, average QS overall score, and average distance from the global mean.

- Distance-based analysis

Euclidean distances in feature space are computed to quantify similarity between institutions. For single university selections, the system identifies the nearest neighbors based on profile similarity.

- Feature differentiation analysis

The system computes differences between the selection mean and the global mean for each indicator.

- Reset interaction

A dedicated button allows users to clear the current selection and restart exploration.

These interactions ensure that analytics are explicitly triggered by visual interaction, satisfying the requirements of the assignment and reinforcing the tight coupling between visualization and computation.

Discovered Insights

Using the visual analytics system, several insights were derived through interactive exploration of the QS World University Rankings 2025 dataset. The combination of dimensionality reduction, coordinated views, and interaction-triggered analytics enabled the identification of global patterns, group characteristics, and institution profiles that are not apparent from ranking tables alone.

- Global Structure of University Performance

Exploration of the PCA scatterplot reveals a clear global structure in the dataset. Universities are distributed along the first principal component in a manner that corresponds to overall academic strength, with institutions characterized by high academic reputation, strong citation impact, and high employment outcomes clustering on one side of the plot. Universities with lower overall performance across these indicators tend to cluster on the opposite side. This observation confirms that PCA embedding captures meaningful variance related to core performance dimensions rather than arbitrary structure.

- Group Comparison and Differentiation

By selecting groups of universities using lasso or box selection on the PCA plot, the system enables direct comparison between selected groups and other universities. For example, selecting a cluster of highly ranked universities reveals a consistent profile characterized by above-average academic reputation, strong employer reputation, and high citation impact. However, indicators such as international students and international faculty show more deviations from the global mean, suggesting that global prestige is more strongly associated with research and reputation metrics rather than internalization alone. In the bar chart, it is noticeable in many group selections, academic reputation and citations per faculty are the most distinctive features, while sustainability and international students contribute less strongly to group differentiation.

- Single University Profile Analysis

When a single university is selected, the system compares the institution's profile against the global mean. The radar chart clearly represents strengths and weaknesses across individual indicators, enabling rapid qualitative assessment. In several cases, universities with similar overall scores exhibit different profiles. For example, one university may outperform the global average in research impact while underperforming in internalization indicators, whereas another institution may exhibit the opposite pattern. This highlights the limitations of relying only on ranking positions and demonstrates the values of profile-based comparisons.

- Similarity and Neighbor Analysis

For a single university selection, the system computes Euclidean distances in feature space to identify the most similar institutions. The neighbor analysis reveals that universities with similar profiles are not always geographically close or similarly ranked. In some cases, the university's nearest neighbors in feature space belong to different regions but share strengths across multiple indicators. This proves that performance similarity is driven more by institutional characteristics than by geographic location or ordinal rank position. Moreover, distance-based similarity analysis provides a quantitative measure of how atypical a university's profile is relative to the other universities. Universities with large distances from the global mean often correspond to highly specialized elite institutions with extreme values on specific indicators.

Overall, the discovered insights demonstrate that ordinal rankings alone are insufficient to fully characterize university performance. The system reveals hidden structure, trade-offs, and similarities that are not visible in traditional tables.

Conclusion

In this project, we designed and implemented an interactive visual analytics system for exploring the QS World University Rankings 2025 dataset. The system combines dimensionality reduction, coordinated multiple views, and interaction-triggered analytics to support exploratory analysis of multidimensional university performance data.

The system demonstrates how visual analytics can increase transparency in ranking data and facilitate more nuanced comparisons than ordinal rank positions alone. Furthermore, it illustrates the value of combining automated analysis and interactive visualization for the exploration of complex ranking datasets and provides a foundation for future extensions.

References

[1] Puri, A., Ku, B. K., Wang, Y., & Qu, H. (2020). Rankbooster: Visual analysis of ranking predictions. *arXiv preprint arXiv:2004.06435*.

[2] Muzamhindo, S., Kong, Y., & Famba, T. (2017). Principal component analysis as a ranking tool-A case of world universities. *International Journal of Advanced Research*, *5*(6), 2114-2135.

[3] Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., & Streit, M. (2013). Lineup: Visual analysis of multi-attribute rankings. *IEEE transactions on visualization and computer graphics*, *19*(12), 2277-2286.

[4] Puri, A., Ku, B. K., Wang, Y., & Qu, H. (2020). Rankbooster: Visual analysis of ranking predictions. *arXiv preprint arXiv:2004.06435*.

[5] Domınguez, M., Vilanova, R., Prada, M. A., Vicario, J., Barbu, M., Varanda, M. J., ... & Paganoni, A. (2018). SPEET: Visual data analysis of engineering students performance from academic data.