Visualizing the Top 400 Universities

Salwa Aljehane[1], Reem Alshahrani[1], and Maha Thafar[1]
saljehan@kent.edu, ralshahr@kent.edu, mthafar@kent.edu
[1]Department of Computer Science, Kent State University, Kent, OH, USA

Abstract— *Visualization is a powerful tool to represent and analyze the data. Meanwhile, visualization became essential for decision making. In this paper, we provide a visualization of the top 400 universities based on the Time Higher Education Universities Rankings [1]. The ranking is based on five different criteria: teaching, research, international outlook, industry income and citation. The visualization system is hosted on a Web- based platform. It visualizes the geographical distribution of the top universities, the score distribution in all ranking criteria and a classification of all universities by area of study in each country. The main goal is to present data in a friendly way that helps to promote people's understanding and help them gain insight and make decisions. We also utilized different visualization tools to show how it could be used for educational purposes.*

Keywords: Top Universities, Visualization, D3, Higher Education

1. Introduction

Every year, more students around the world are seeking a higher education with different criteria. The Internet has opened the world for all the students worldwide and made it easier for them to find opportunities and scholarships to study abroad. Universities rankings is one way to look for the university that best fits the students' ambitions. Additionally, graduates who are seeking jobs in high-profile universities, professors, and researchers are all interested in the annual rankings. There are some magazines that have developed methodologies and criteria to analyze and rate the universities worldwide, for example, US news, Times Higher education, and QS. The majority of the rankings use classic visualization tools such as tables or pie charts. However, these classic tools do not give the users the ability to navigate the universities rankings easily. Of course, that directly affects the usability of the rankings. In addition, these tools are out-dated in the era of big data and information visualization.

Providing a useful and efficient platform for users is essential to increasing the usability and readability of the data. It is also crucial to avoid overwhelming users with the clutter of data. Achieving this is not easy, due to the large volume of information related to the rankings. However, the advancement in visualization tools has led to a better visual perception. Well-designed data graphics is usually the simplest and the most powerful way to help the users think visually.

There are two main goals of this paper. The first goal is to present data in a friendly way that helps to promote peoples' understanding and help them gain insight from it. Thus, using visualization to support interactive visual representations of data helps to amplify cognition. It has been said "A picture is worth a thousand words.".

Reasoning is an important way to make decisions and comparisons. Using visualization tools could help to enforce the data to be within the scope of eye span. Consequently, more precise and reasonable decisions could be taken. The current way of visualizing the top universities does not provide an efficient representation of the data which make it harder for the users to draw a conclusion.

The other goal of this paper is to illustrate how to utilize the available visualization tools for educational purposes. Google Developers provide a wide range of tools that can be utilized to analyze and visualize different types of data. An example of that is Google Charts [2], which allows the developers to present raw data as a histogram, line chart, timeline, etc., using JavaScript. To visualize the top universities, Google Bar Chart was used to give an overview of the rankings in a bar chart as shown in section III. Google Maps was used to show the geographical distribution of the top universities as shown in section III. Data-Driven Documents (D3) [3] library was utilized to develop other visualization elements with more detailed information about each university as shown in section IV.

The contributions of this paper are summarized as follows:

- Utilizing the available visualization tools to present the top universities to help data mining and knowledge discovery.
- Visualization of the top 400 universities, clustered based on the location and area of study.
- Comparing universities to visually facilitate decision-making for the students who are seeking higher education.

Time Higher Education magazine's ranking of universities was used as a resource for the dataset, which is one of the most influential and widely observed rankings. They use 13 different metrics to provide comprehensive comparisons and a balanced evaluation. These metrics are grouped into five groups: teaching, research, citations, industry income, and international outlook. This visualization of the top 400 universities would help the students who are thinking of studying overseas, students who are planning on getting enrolled in Global Exchange programs, innovative students

who need funding for their research, and all other students locally and globally. The goal is to visualize the dataset of the top universities so the students can gain knowledge and insights about the universities, in order to decide their future institution of higher education. The visualization system visualizes the top universities with the following aspects:

- The overall ranking.
- The geographical distribution of the universities world-wide.
- Classification of the top universities by subject in each country.
- Detailed information (ranking by criteria) for all universities in each country.

The subjects or area of study were classified based on the classification used in the Times Higher Education magazine, which includes 5 subjects: arts and humanities, clinical, pre-clinical and health, engineering and technology, life sciences, and social sciences. In addition to that, the visualization will show the top 100 universities under 50 years old. Visualizing those universities can give an insight about the future of higher education.

2. Design and Development

2.1 Data Collection

No dataset was available. Therefore, we had to collect raw data, analyze and reproduce the data to use it. Data were collected from Time Higher Education World University Rankings magazine. Noteworthy, Times Higher Education provides detailed data about the first 200 universities such as the score for each evaluation criteria. Therefore, some parts of the visualization system were limited to the top 200 universities due to the lack of information from the main resource. The data was organized in many formats as needed such as Comma-Separated Value (CSV) and nested JavaScript Object Notation (JSON) [4] to represent the hierarchy of the data. To represent the geographical distribution of the top universities, OpenRefine [5] were used to retrieve the geocoding for each university.

2.1.1 Framework

The visualization system was developed on a Web-based platform using HTML, JavaScript, JQuery, and Bootstrap library to make it scalable to any browsing environment such as desktop, laptop, tablet, or smartphone. The website was carefully designed to ensure simplicity and usability. Guidelines were provided to the user throughout all the web pages.

3. Utilizing Google Charts

Google Chart package was used to give a general overview of the universities' rankings. It is also packed with many features such as tooltips and graph legend to customize the charts.
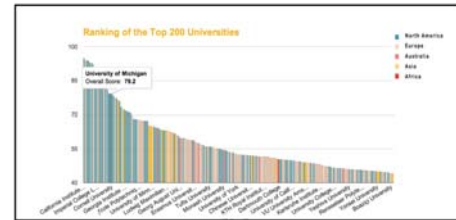

Fig. 1: Bar Chart

Fig. 2: Google Map with Map layout

3.1 Bar Chart

The bar chart was implemented to show the top universities due to its simplicity and most of the users are familiar with this kind of visualization. The column length represents the overall score of each university which reflects the rank. The colors were used to distinguish universities in different continents as shown in 1. It is also packed with many features such as tooltips and color encoding to customize the charts.

By moving the mouse over any column; the user will see a tooltip that shows the university's name and its overall score.

3.2 Google Map

Due to the need to represent a large number of universities, 400, it was necessary to tackle the scalability issue and represent these universities in a simple and intuitive way. Google Map API [4] was used to visualize the universities according to their geographical distribution. This way of visualization shows the location of each university so the user can get insights about the city size, terrain and location either coastal or mountainous area. Google Map provides different layouts which can be chosen by the user. The layouts include map, terrain, and satellite. These layouts are shown in 2 3 respectively.

It is also zoomable to facilitate the navigation through the map. Tiny markers were used to point to the countries that have universities in the top 400 rankings list. Additionally, each marker is linked with a tooltip that shows a general description of the selected university. Clicking on the university's name will open the university's website.
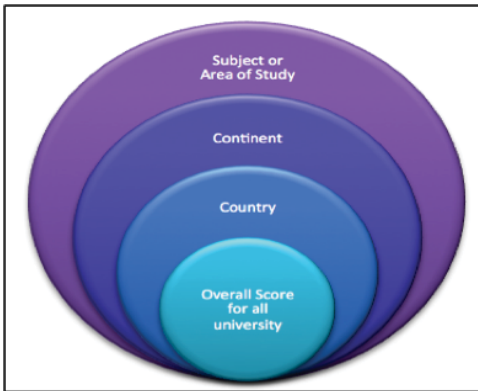
Fig. 3: Google Map with Satellite layout
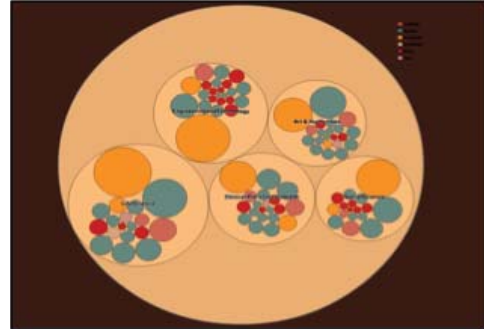

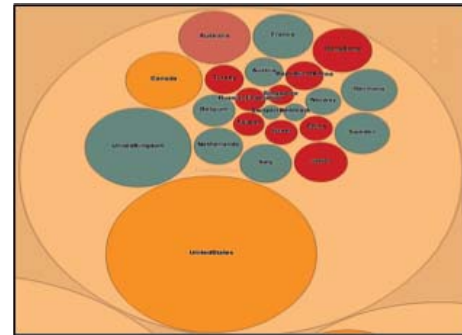Fig. 4: Data hierarchy

Fig. 5: Circle packing

Fig. 6: Circle packing after zooming in

4. Utilizing D3 Library

4.1 Circle Packing

This part of the visualization system was developed to represent the hierarchical data. The hierarchy of the data as shown in 4. It was used to visualize the top universities classified by the subject or area of study.

This circle packing answers the question: which subject is good at which country? so the student can decide on the countries that have a good reputation in his/her area of interest. The circle packing provides simple and effective interaction. By clicking on a subject, the circle will take the student into a lower level of the hierarchy to show the countries with top universities in that subject. Circle packing is useful to visualize the hierarchical data. The size of the circles represents the overall score and the color was utilized to represent the continent of each country. To make the circle packing more efficient, the zooming effect was added to make the circles zoomable so the student can navigate the hierarchy of the data easily. The final result of this part is shown in Figures 5 and 4 respectively.

4.2 World Map Visualization of The Universities With Heatmaps and Pie Charts

The main idea of this part is to simplify the searching for a university by utilizing different techniques. This visualization incorporates three different techniques, namely, geographical distribution on the world map, heatmaps and

pie charts. Figures 7 and 8 shows the final results of this part.

To make the navigation easier for the user, the map has markers that point to the countries with universities ranked as the top 200 universities. The student can interact with this visualization as follows:

Map Interaction:

The top 400 universities are distributed worldwide. The best way to visualize the universities in a simple and intuitive way is using the map. That makes it easier for the student to determine the location of the university and how far it is from his home country. By clicking on the country, a little tooltip will appear to give the option for the user to go to the detailed information about the universities in this country as shown in Fig. 7. The
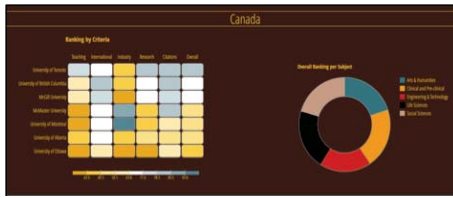

Fig. 7: World map visualization

Fig. 8: The detailed information for each country

link in the tooltip will take the student to a detailed information, which is represented by an interactive heatmap and pie chart as shown in Figure 7. To make the navigation through the map easier, the map is zoomable, so the user can zoom in and out.

Heatmap Interaction

The heatmap was implemented to represent the variation in the scores for each evaluation criteria. It is an effective technique to visualize the scores using the colors by normalizing the scores and mapping them into a specific color scale. That make it easier for the student to infer and compare the scores by the colors instead of reading the scores for each individual university and criterion. The order of the universities in the heatmap reflects the rank of the universities in the country as shown in Figure 8. By moving the mouse over the heatmap, the scores will appear to the student. Additionally, moving the mouse on the heatmap will show the scores that fall under the selected range.

Pie Chart Interaction

The pie chart is a simple and easy visualization tool to infer the information from. Additionally, most of the users are familiar with the pie chart. It shows the subjects that have the highest scores in each country which is calculated by the average overall score for all the universities in each subject. That reflects what this country is strong in and what are the offered subjects in the top universities in this country. Colors encoding were also used to distinguish different subjects. Figure 8 shows the pie chart beside the heatmap.

4.3 Bubble Chart

This visualization will show the top 100 universities under 50 years old. Visualizing those universities can give an insight about the future of the higher education. Specifically, it will show the prediction of the leading universities in the future. In addition, it might provide alternative choices for students who cannot pay the tuition for the high-profile universities such as Harvard and Oxford. When the user place the mouse over a circle it will show a tooltip that shows detailed information about the university such as university's
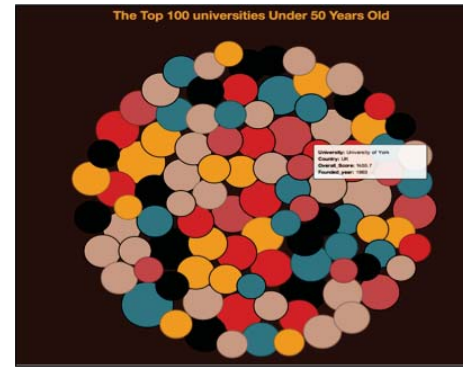

Fig. 9: 100 universities under 50 years bubble chart

name, country, overall score, and the founded year as shown in Figure 9. The bubble chart in the D3 library was used to implement this part. Each university is represented by a bubble. Six colors were used to encode the six continents based on the location of the university. The size of the circle reflects the overall score of the university.

5. Related Work

Information visualization techniques have been improved over the years. A few of these techniques have been utilized to visualize the top universities. Generally, tables are used by the magazines that announce the rankings, such as U.S. News and Times Higher Education. A team from Harvard University and Johannes Kepler University developed LineUp [6]. This tool visualizes the dataset in a tabular view with headers that show the histogram of the scores. It allows the user to explore, refine, and combine these attributes to see the effect on the rankings. However, the tabular view may lead to ambiguity due to the large volume of data. Additionally, it needs understanding of how to use the Lineup interface. Another effort had been made in [7] to visualize the top universities based on publications and citations. The authors represented the universities on a map to show the geographical distribution, which is easier to understand. However, the ranking was limited to certain criteria.

6. Implementation

The visualization system is hosted on a Web-based platform to ensure the accessibility and the availability for all the users. The entire visualization system is available on the following link: `http://www.cs.kent.edu/~ralshahr/Project/dashboard/`

7. Evaluation

A survey was developed to evaluate and test the usability of the visualization system by real users. It consisted of 10 questions using the Likert scale, ranging from strongly agree to disagree, and one open-ended question to allow students
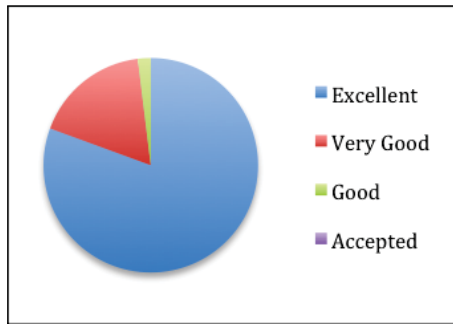
Fig. 10: Survey results

suggestions for any further improvement. This survey manipulates three major aspects:

- Navigation through the visualization system.
- How the system helps the student to gain insight about the universities.
- The interface and the design of the visualization system.

The survey was distributed to around 60 users and 52 responses were received. The survey was distributed to users from different countries and diverse backgrounds. The result of the survey is shown in 10.

8. Conclusion

In this paper, the Top 400 universities data was collected, analyzed, organized and visualized in different ways. Multiple visualizations were presented in order to help the students achieve the goal of this visualization system. Different techniques and tools were utilized to provide a meaningful view of the data. The system can be updated with the most recent rankings as needed.

References

[1] (2014) The times higher education world university rankings. Times Higher Education. 26 Red Lion Square, London. [Online]. Available: https://www.timeshighereducation.co.uk/world-university-rankings/2014-15/world-ranking

[2] Z. Ying, "Introducing google chart tools and google maps api in data visualization courses." *IEEE computer graphics and applications*, vol. 32, no. 6, pp. 6–9, 2011.

[3] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2301–2309, 2011.

[4] C. J. Ihrig, "Javascript object notation," in *Pro Node. js for Developers*. Springer, 2013, pp. 263–270.

[5] R. Verborgh and M. De Wilde, *Using OpenRefine*. Packt Publishing Ltd, 2013.

[6] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit, "Lineup: Visual analysis of multi-attribute rankings," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2277–2286, 2013.

[7] L. Bornmann, M. Stefaner, F. de Moya Anegón, and R. Mutz, "Ranking and mapping of universities and research-focused institutions worldwide based on highly-cited papers: A visualisation of results from multi-level models," *Online Information Review*, vol. 38, no. 1, pp. 43–58, 2014.