

# CATalytics: A Visual Analytics Tool for Exploring and Analysing University Rankings

*Albert Bingei, Cornelia Tisandinia Larasati, Timothy Tan Swee Guang*

**ABSTRACT** - World university rankings, which are essential for universities in attracting talent and producing new knowledge, have long been under limelight for their measures that proclaim to represent academic quality. No two university rankings are ever homogenous and the effect of the various performance indicators and demographics of universities are often left unscrutinised. Addressing this need for analysis, we have designed and developed CATalytics, a dynamic and interactive visual analytics dashboard to allow key stakeholders, whether students, academics or university management, to gain valuable insights on how various performance indicators affect the ranking of universities, and how these change over the years. CATalytics thus utilises a combination of time series graphs, a global map, star plots (glyphs), slope graphs and ternary plots to aid users in exploring and analysing what makes a good university.

**INDEX TERMS** - Data visualisation, University rankings, Time series, Star plot, Glyph, Mercator projection, Slope graph



## 1. INTRODUCTION

World university rankings have long been under scrutiny for their measures to represent academic quality. The Times Higher Education (THE) World University Ranking is one of these many university rankings available which rank universities across the world according to a set of performance indicators. These rankings are important for universities in attracting talent and producing new knowledge and affect various stakeholders in different ways [1].

Universities which rank highly often see rising student applications while those declining in ranking suffer a decline. This is because these rankings are an important factor in decision-making for students applying or planning to apply for higher education. According to a survey of more than 14,000 international students in Australia and the UK, 77% of students listed both university and subject rankings as “very important” when deciding a place to study [2].

Academics are also not immune to a university’s rank. A ‘halo’ effect exists on academics as their perceptions of a university’s quality are shaped by university rankings [3]. This creates a competitive environment where the most talented and decorated professors apply to top universities so as to be associated with the prestige of the university, without having actual knowledge of the program quality [4]. This leads to a cyclical self-fulfilling prophecy, where the top universities will improve in academic rigour, while lower-ranked universities do not attract new and attractive talent.

Given the importance of these rankings on a university’s academic rigor and subsequent performance, this project hopes to provide an interactive outlet for stakeholders to explore what makes a good university beyond its own ranking.

## 2. MOTIVATION AND OBJECTIVES

Most ranking systems online merely list out the universities and their rankings relative to each other, without delving into the breakdown of performance indicators that may affect the rankings. This makes it difficult for anyone to truly visualize the information being displayed, such as how various performance indicators affect the universities’ rankings and how universities have moved within the ranks through the years.

CATalytics (Colleges And Tertiaries Analytics) aims to provide users with an integration of functions that are usually provided as individual modules. Specifically, CATalytics seeks to provide the following analysis requirements:

1. To present the trends of university rankings across years and locations
2. To analyse the relative effect of performance indicators on university ranking
3. To visualise the relative effect of demographics on university ranking

## 3. DATA TRANSFORMATION

There are various insights that can be developed from data. However, most data that is found cannot be immediately used as they are usually in the form that was convenient for the uploader/creator. Hence, it is necessary to transform data, more commonly known as data cleaning, to use them effectively and make them more relevant for our current purposes.

The data was transformed mainly in these ways:

1. Recompute total scores based on weightage of various performance indicators

2. Rank every university based on recomputed total scores
3. Normalise ratios

In the original data, each performance indicator of each university was given a score between 0-100 and a total score based on these scores were given in another column. However, many of the total scores had missing values as they were not pre-calculated by THE. With reference to THE's methodology [5], our group assigned the respective weight to each performance indicator and thus, calculated the total scores for all the universities. This enabled us to fill in the missing total scores.

As some of the total scores of universities were not calculated, especially those that were not in the top 200 rankings of each year, many universities were just assigned a ranking within a range of 100 rankings. For example, the university ranked directly after #200 would be given a rank of 201-300 in the "Rank" column. This prevented our group from sorting through the universities in a meaningful ranking. Hence, by filling in the missing total scores, our group was able to rank the universities from 1 to 800, with the #1 university having the highest total score for that particular year. From this set, only the top 400 of each year were retained as 2012 only had 400 universities on the rankings. This was done to standardise the number of universities per year that can be analysed.

The ratios for each university, namely Student-Staff, International Student and Female-Male ratios, were not standardised across a single base. Some of these ratios were also missing as the values of essential statistics were also missing. For example, a missing value in the "number of students" column would also prevent Student-Staff and International Student ratios from being calculated. Thus, our group first filled in the missing statistics and calculated the raw ratios. Then, to tackle the problem of a lack of standardisation, the ratios were normalised to bring all of 3 ratios into proportion with one another.

Further data cleaning also included:

1. Renaming and reformatting the data headers
2. Ensuring university names with diacritics were replaced with readable Roman script
3. Filling in missing statistics of individual universities through extensive literature reviews
4. Inserting latitude and longitude coordinates for every university

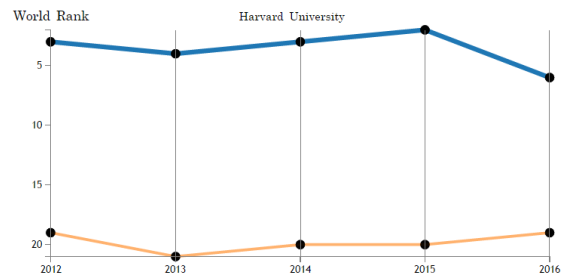
#### 4. VISUALISATION DESIGN

To visualise the data, our group chose to use 3 main elements - time series, glyph, and ternary plot - and 2

supporting elements - global map and slope graph. Each of the visualisations will allow users to explore the data in a different way.

##### a. Time-series Graph

A time series is a series of data points graphed in time order taken at successive equally spaced points in time. According to the feature of the dataset, a time series line graph will be the best way to display an overview of how the rankings of each university change over the years.



This visualisation will be used as a base filter for users to select which universities and which year they would like to explore further. This selection will then be reflected on the global map, glyph and ternary plots for comparison purposes and further insights.

##### b. Global Map: Mercator Projection

A zoomable map will provide users with an overview of the distribution of ranked universities in the world. While the map is based off the Mercator projection (a cylindrical map projection presented by the Flemish geographer and cartographer Gerardus Mercator) which may distort certain sizes of countries, this projection is well-suited as an interactive world map that can be zoomed in and out seamlessly [6].

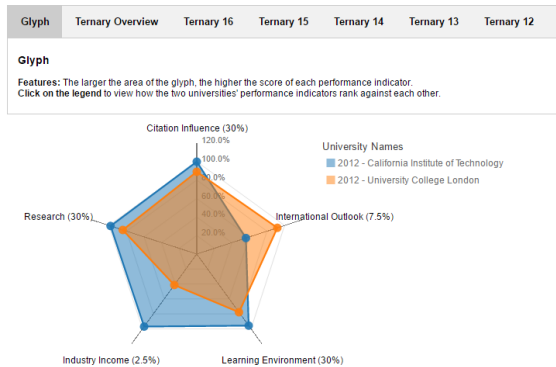
Without any selection, the map will show all the data points ("pins") of all universities by default. Upon selection of a universities the pins of the selected universities will remain on the map while the unselected universities will not appear. As shown below, a tooltip with the university's name will appear if users hover over the pin on the map.



For more information on the university, users can click the pins on the map to be brought to a Google search of the university.

### c. Star Plot (Glyph)

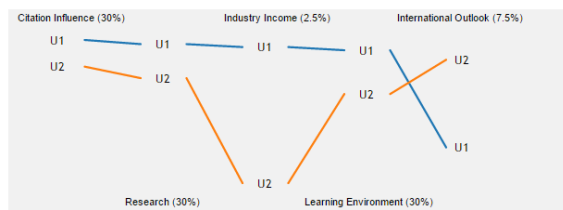
A glyph (radar chart) can be used to display multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point [7]. This visualisation is also useful in identifying any outliers within the performance indicators [8]. In our dataset, our group has identified 5 main performance indicators which contribute to the total score and thus affect the university's rankings.



Glyphs also allow layering of data for comparison among different universities in the same year. The larger the plotted glyph area, the higher the scores of the different performance indicators as well as the ranking of the selected university. Hence, at a glance, users can observe which universities are ranked higher than the other. Glyphs also allow us to pinpoint which performance indicators of a specific university are more highly ranked.

### d. Slope Graph

A slope graph displays how each performance indicator of a certain university changes relative to the rest of universities each year, as well as its rate of change.



The benefit of this visualisation is that there is virtually no data ink (ie. no excess visualisation that does not represent the data) [9]. This allows it to be viewed within a tooltip. Slope graphs can also display multiple alternative rankings on the same set of items, in this case, across a set of performance indicators.

### e. Ternary Plot

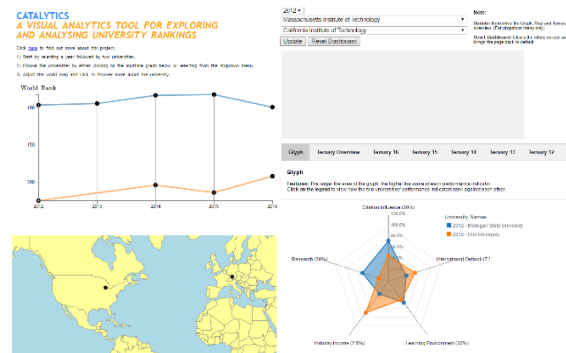
Ternary plots allow users to observe the correlation of 3 variables - Student-Staff, International Student-Student

and Female-Male ratios, with each of the 3 apexes representing one ratio [10]. This allows users to observe the movement in trend across years of specific universities through the movement of data points within a ternary chart. From this chart, users can also observe the significance of one ratio over another, and how this affects the ranking of the specific university.

## 5. USER INTERFACE DESIGN

The intention of CATalytics is to let users explore what makes a well-ranked university. Thus, our dashboard aims to facilitate the familiarisation and reasoning of existing data through a range of user-driven experiences. The design of CATalytics closely follows Schneiderman's data visualisation general guidelines [11]: over, zoom and filter, details on demand.

The application will display the main dashboard.

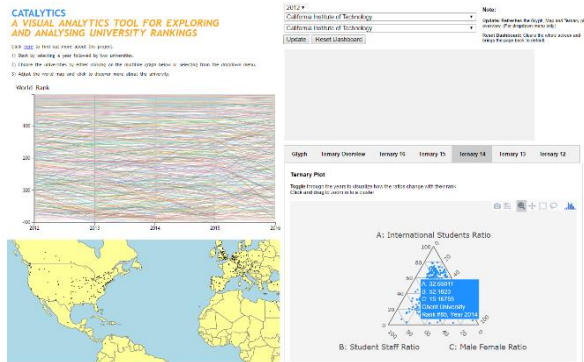


All visualisations will appear on the same computer screen, and achieves the 4 main characteristics that define an effective information dashboard design, as outlined by Few [12]. All universities across all years will be displayed by default on the time series graph at the top of the page and the global map to the left of the page. Upon selection of 2 specific universities within a year, a glyph will appear, displaying the 2 universities, to the right of the global map. The global map will also update to show just the 2 universities selected.

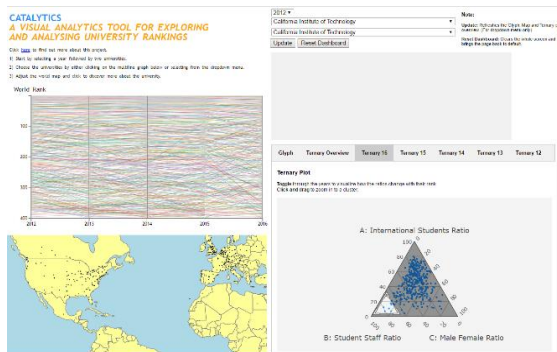
These elements are linked through the coordinated linked view (CLV), which is a common data visualisation technique that allows users to manipulate the visual presentation and coordinating the interactions between these elements [13]. This is crucial to synchronise the presentation of the data, especially since the user will need to interact with the elements, which leads to an update in the space-time-attribute parameters for an overview-to-detail display.

Clicking on the legend of the glyph will allow a slope graph to appear at the top right of the page. This supports users in providing details on demand. The slope graph will also reflect the ranking of each performance indicator of a university (multivariate data) in relation to the other university selected.

Selecting the “Ternary Plot” tab will replace the glyph visualisation with a ternary plot visualisation within the same screen.



Users are able to select a single year to be displayed on the ternary plot. Hovering over a data point also creates a tooltip with information about the demographic ratios. Users can also zoom in onto the ternary plot using a lasso function (shown below) to study a cluster of universities. This scaling function allows users to zoom in on crowded regions of the plot for greater insight.



The combination of these visuals will give the user greater control over the data and allow users to explore the data through different visualisations to aid them in drawing conclusions of what produces a well-ranked university.

## 6. WEB-BASED APPLICATION

### a. System Architecture

This project is presented as a web-based application - designed and developed through Javascript. This was especially after taking into consideration the latest changes in the mobile computing and rich client applications. Through this method, users will be able to access the application on any device with their web browsers. The D3.js and Plot.ly libraries are used to design the application.

### b. Data Visualisation

CATALYTICS was developed with a strong emphasis on interactivity and usability. In line with the best practices for dashboard design, this project set out to produce a

dashboard that provided an exploratory point of view through a historical time horizon. The dashboard should also be able to provide a common interface for interacting with and analysing the key data.

Hence, all the data visualisation elements were created using Data-Driven Documents (D3.js) which is a JavaScript library producing dynamic, interactive data visualizations in web browsers which users are able to manipulate [14]. It makes use of the widely implemented Scalable Vector Array (SVG), HTML5, and CSS standards to produce elements which can be added, removed or edited according to the contents of the dataset [15]. D3.js also handles large data efficiently and the elements encoded update automatically based on changes in the data, providing a visualisation that is visually pleasing and understandable.

### c. Web Design

Following the best practices for dashboard design, each component of the dashboard is in a modular format, developed through the CSS functionalities of d3.js. Creating a dashboard independent of any third-party app support allows the application to be supported on many common browsers such as Google Chrome, Firefox, Internet Explorer, Opera, and Safari browsers.

This also allows us to implement best practices for dashboard design, such as by ensuring that the boundaries of a single page are not exceeded and that the visual elements are arranged in an organised format. These attributes will improve user-friendliness of our dashboard.

## 7. CASE STUDY

To give an example of how the application is used, we will undertake a case study to explore the application.

Users will start immediately at the main dashboard. All universities across all years will be displayed by default on the time series graph and the global map.

To start off, there are 2 ways to select universities for comparison. Firstly, users can select 2 specific universities with a year using the drop down menu and the “update” button. Alternatively, users can also select a year first before clicking on 2 universities in the multiline time-series plot. This allows users greater flexibility in choosing which university to compare.

Upon selection of 2 specific universities within a year and clicking on the “Update” button, 3 elements will change - the time series graph will reflect only these 2 universities, the global map will update to show just the 2 universities selected and a glyph will appear, displaying the 2 universities on the glyph. Users can hover over the data points in the time series graph to discover the rank

of the particular university in a particular year. On the glyph, the higher the score of the respective performance indicator, the closer the data point is to the outside axis. Clicking on the legend of the glyph will allow a slope graph to appear. The slope graph will reflect the relative ranking of each performance indicator of a particular university to the other university selected.

Selecting the “Ternary Plot” tab will replace the glyph visualisation with a ternary plot visualisation within the same screen. Users are able to select a single year to be displayed on the ternary plot, through the other tabs, labelled by year. Users can also zoom in onto the ternary plot using a lasso function to study a cluster of universities that need further analysis. Double-clicking the ternary plot will reset the ternary plot.

Users can reset the dashboard by clicking the “Reset Dashboard” button at the top.

## 8. CONCLUSION

CATalytics enables users to explore and analyse the factors that could affect a university’s rankings to gain valuable insights for their field - whether they are potential students, academics or university management. Through a comparison between two competing universities, or an overview of the trend of demographic ratios, users will hopefully be able to make important decisions in their respective backgrounds, whether it is to decide on a future university to enroll into or to improve the curriculum and academic support. Best practices from other universities can also be derived from this application.

CATalytics has a lot of potential to be enhanced further to provide a more rigorous analysis of factors that affect university rankings, especially since only THE World University Rankings was used for this application. Alternative rankings from other ranking companies, such as QS World University Rankings, Center for World University Rankings (CWUR) and Shanghai Ranking, have other attributes that contribute to a university’s total score as well as utilise different scoring systems. These various factors can be compared and analysed in a single application to gain insight into how multiple heterogeneous attributes affect the rankings. Furthermore, this application could be enhanced by adding an independent algorithm to combine all these attributes and providing our own comprehensive and independent ranking system.

The education landscape is constantly changing, as various teaching pedagogies and university focuses develop through the years. Having an application which can comprehensively and consistently track these

changes in the focus of various universities annually would definitely be the ultimate goal.

## 9. REFERENCES

- [1] Sadlak, J. (2010). Ranking in Higher Education: Its Place and Impact. Retrieved November 19, 2016, from <http://www.educationarena.com/pdf/sample/sample-essay-sadlak.pdf>
- [2] University rankings most important factor for international students. (2015). Retrieved November 07, 2016, from <https://www.timeshighereducation.com/world-university-rankings/news/university-rankings-most-important-factor-for-international-students>
- [3] Marginson, S. (2006). Australian universities in the global context. *Campus Review*, 16(11), 8-9.
- [4] Clarke, M. (2002). 'Some guidelines for academic quality rankings', *Higher Education in Europe* 27(4), 443-459.
- [5] World University Rankings 2015-2016 methodology. (2016). Retrieved November 16, 2016, from <https://www.timeshighereducation.com/news/ranking-methodology-2016>
- [6] Grafarend, E. (1995). The Optimal Universal Transverse Mercator Projection. *International Association of Geodesy Symposia Geodetic Theory Today*, 51-51. doi:10.1007/978-3-642-79824-5\_13
- [7] Borgo, R., Kehrer, J., Chung, D. H., Maguire, E., Laramée, R. S., Hauser, H., . . . Chen, M. (2013). Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. *EUROGRAPHICS*. Retrieved from <http://www.iu.uib.no/vis/publications/publication/2013/pdfs/Borgo13GlyphBased.pdf>
- [8] Radar Chart. (n.d.). Retrieved November 19, 2016, from [http://datavizcatalogue.com/methods/radar\\_chart.html](http://datavizcatalogue.com/methods/radar_chart.html)
- [9] Edward Tufte's "Slopegraphs". (n.d.). Retrieved November 19, 2016, from <http://charliepark.org/slopegraphs/>
- [10] Reading a Ternary Diagram And Creating Rock Names. (n.d.). Retrieved November 19, 2016, from <https://www.peeranswer.com/question/56bbc9ec938137ad032f2996>
- [11] Shneiderman, B. (2005) "The eyes have it: A task by data type taxonomy for information visualization" *IEEE Conference on Visual Languages (VL96)*, pp. 336-343.
- [12] Few, S. (2006). *Information dashboard design: The effective visual communication of data*. Beijing: O'Reilly.
- [13] Robert, J.C. "State of the Art: Coordinated & Multiple Views in Exploratory Visualization", *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*.
- [14] Bostock, M. (2012) *Data-Driven Document*. Available at [d3js.org](http://d3js.org). Accessed 31 Oct. 2012.
- [15] Dewar, M (2012) *Getting Started with D3*. O'Reilly Media Inc. USA.
- [16] Mark Otto, Jacob Thornton, and Bootstrap contributors. (n.d.). *Bootstrap* · The world's most popular mobile-first and responsive front-end framework. Retrieved November 18, 2016, from <http://getbootstrap.com/>

## 10. AUTHORS' BIOGRAPHY

**Albert Bingei** is currently a final year student pursuing his Bachelor of Business Management (BBM) in the Lee Kong Chian School of Business at Singapore Management University, with a second major in Analytics.

**Cornelia Tisandinia Larasati** is currently a final year student pursuing her Bachelor of Accountancy (BAcc) in the School of Accountancy at Singapore Management University, with a second major in Analytics.

**Timothy Tan Swee Guang** is currently a final year student pursuing his Bachelor of Social Science (BSocSc) in the School of Social Sciences at Singapore Management University, with a second major in Analytics.