

University Ranking Project

COLUCELLI MICHELE, DI LISA MARCO AND MALERBA MARTINA
DIAG - Sapienza Università di Roma

August 12, 2020



Figure 1: Dean Mode with Sapienza University

Abstract

Education is such a big deal to people nowadays and millions of students apply to Universities each year. Whatever is their reason for doing it, be it to find the job of their dreams or broaden their knowledge, tracking down the Universities that satisfy their needs is not an easy task at all. With this application we want to help those Students make a choice, by offering many attributes to compare from three different famous university rankings. This application is also for those Deans that are looking for ways to improve their Universities, from quality of education to international outlook.

I. INTRODUCTION

The project that we're presenting in this paper comes into two modes: Student mode and Dean mode; the aim is to allow a Student to easily choose the best university for them and for a Dean to improve their University, with the aid of different parameters.

II. DATASET

The dataset was obtained by joining three different university ranking datasets : the *Times Higher Education World University Ranking*, the *Center for World University Ranking* and the *Shanghai Ranking*. By joining the three datasets, we obtained a variety of different attributes that can improve the experience of a Student or a Dean.

First, we did the inner join of the dataset by taking into consideration only the universities that were in all three datasets; then, for the complete dataset, we did a final inner join to add the geographic coordinate for each instance. Finally we added the data from the *Numbeo database*, which is the world's largest cost of living database and also a crowd-sourced global database of quality of life informations including housing indicators, perceived crime rates, and quality of healthcare, among many other statistics. At the end we did the merge of all the datasets in a single dataset.

From the *Times Higher Education World University Ranking* we used the following attributes:

- the rank: the final rank regarding the Times Higher Education World University Rankings;
- the overall: the global strength of an institution;
- the teaching : it examines the perceived prestige of institutions in teaching. The responses were statistically representative of the global academy's geographical and subject mix;
- the research: it looks at a university's reputation for research excellence among its peers, based on the responses to their annual Academic Reputation Survey;
- the citations: an influence indicator looks at universities' role in spreading new knowledge and ideas;
- the industry income: a university's ability to help industry with innovations, inventions and consultancy has become a core mission of the contemporary global academy;
- the international outlook: the ability of a university to attract undergraduates, post-graduates and faculty from all over the planet is key to its success on the world stage.

From the *Center for World University Ranking*:

- cwur rank: ranks each university among all other universities in the world;

- national rank: ranks each university on a national level;
- quality of education: represents the number of alumni who have won major international awards, prizes, medals, etc.;
- alumni employment: measure of number of alumni who have held CEO positions at companies on Forbes Global 2000 list;
- quality of faculty: measure of number of academics who have won major international awards, prizes, and medals;
- research performance: measure of research output (total # of papers), high-quality publications (# of papers in top journals), influence (# of papers in highly influential journals), and citations (# of highly cited papers);
- quality of publications, influence, citations, research output: values used to measure the research performance in years before 2019;
- score: weighted sum of other numeric metrics (details available in CWUR methodology);

From the *Shanghai Ranking*:

- shg world rank: ranks each university among all other universities in the world;
- shg national rank: ranks each university on a national level;
- shg score: the highest scoring university is assigned a score of 100, and other universities are calculated as a percentage of the top score;
- shg score on alumni: a percentage of the total number of the alumni of a university winning Nobel Prizes and Field Medals. Alumni are defined as those who obtain bachelor's, master's or doctoral degrees from the university. Different weights are set according to the periods of obtaining degrees. The weight is 100 for alumni obtaining degrees after 2011, and then it decreases for each year as you go back in time;
- shg score on HiCi: a percentage of the number of Highly Cited Researchers that gives a better idea of the quality of faculty;

- shg score on pub: the total number of papers indexed in Science Citation Index-Expanded and Social Science Citation Index.

Finally from the *Numbeo*:

- Cost of Living Index : describes how much is the living costs on average;
- Rent Index:describes how much an apartment rental costs on average;
- Cost of Living Plus Rent Index: is the average of the first two described above;
- Groceries Index:food cost on average;
- Restaurant Price Index:restaurant cost on average;
- Local Purchasing Power Index:the average purchasing power in that country;

These indices are relative to New York City (NYC). Which means that for New York City, each index should be 100(%). If another city has, for example, rent index of 120, it means that on an average in that city rents are 20% more expensive than in New York City. If a city has rent index of 70, that means on average rent in that city is 30% less expensive than in New York City.

i. Dataset Preprocessing

Since we wanted to reach two different goals, we decided to compute two different PCAs, one for the Student mode and one for the Dean mode, on the same dataset acting on different attributes. So the pre-processing is executed two times on the final dataset adding four more attributes. To compute the principal components first we standardized the features by removing the mean and scaling to unit variance using sklearn library in python. Those principal components have been computed once one, so they don't change overtime. They were then used in the corresponding scatter plot for each user mode

For the Student mode we did the pca on cwur wordl rank, sha rank,the rank,latitude,longitude,Cost of living plus rent index,Grocery Index,the teaching,the research,the citations,the international outlook.

For the Dean we did instead the pca on cwur quality of education, cwur quality of faculty, cwur influence , the research ,the citations,the international outlook and the income.

III. VISUALIZATION AND GOALS

We will now describe the different views that were used for the two user modes. They were both done using D3.js , which is a Javascript library for Visual Analytics purposes.

IV. STUDENT

Education is such a big deal to people nowadays and millions of Students apply to universities each year. Tracking down the universities that satisfy their needs is not an easy task at all, so with this application we want to help those Students make a choice, by offering many attributes to compare from three different famous university rankings.

i. Geographical Map

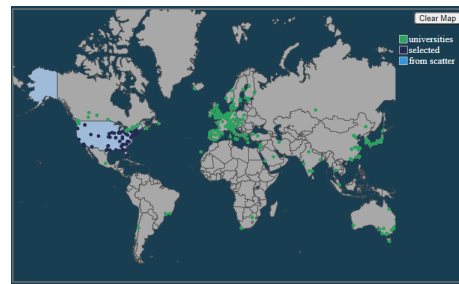


Figure 2: Selection of the USA country

The geolocation of the universities is represented using a geographic map where each university is represented by their *latitude* and *longitude*. This view helps the Student to choose in a easy way the university, knowing the country. Each university can change color when they generate events from the scatter plot and when they choose the countries they're interested in. The Student has also the possibility to pan and zoom on the map, which is very

useful when there is a cluster of universities that would make them hard to distinguish.

ii. Parallel graph

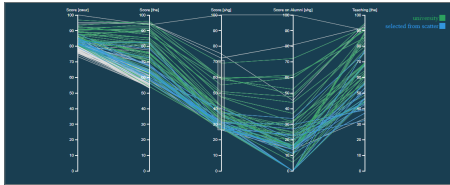
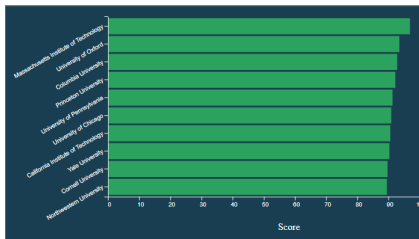


Figure 3: the parallel with brush on the score of shanghai and with scatter interaction

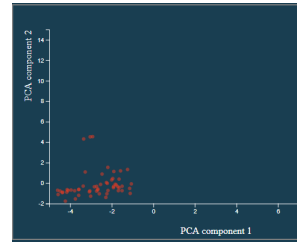
The parallel graph allows to compare the features of several individual observations on a set of numeric values. Each vertical bar represents a variable and often has its own scale. The values at the end are plotted as series of lines connected across each axis. The attributes that were used are the *Score cwur*, *Score the*, *Score shg*, *Score on alumni* and *Teaching*.

iii. Bar chart



The bar chart shows the relationship between a numeric and a categorical variable. Each entity of the categorical variable represented as a bar. The size of the bar represents its numerical value. In the Student mode, the numeric variable is the *cwur* score and the categorical variables are the universities; In the Dean mode, the numeric variable is the average of the scores selected in the checkbox.

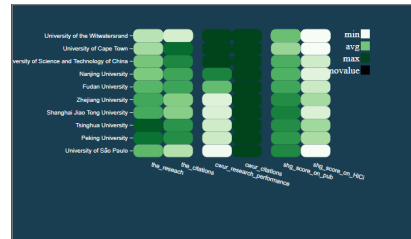
iv. Scatter Plot



The scatter plot takes the PCA that was tailored for the Student as input.

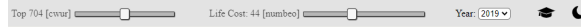
When there are clusters of universities, it can be hard to tell them apart, so we included a brush function that allows to select a subset of them, and a zoom function that zooms in on said subset. To zoom out and reset the scatter plot, the Student can simply double click on it. Tooltips were also included in order to let the Student know what universities are depicted in the graph.

v. Heat map



We figured that the Student might want to compare the three rankings on how well a university does on *Research Performance* and *Citations*. The universities selected by the Student on the parallel graph appear on the heat map: a scale of greens is used to represent the 'intensity' of each attribute; the darker the color is, the higher the value of the attribute is. Some attributes may appear in black, which is not a color that belongs to the scale, but it simply depicts that there is no available data for those particular attributes.

vi. Header



The header for this part permits to the Student to filter the data selecting the range of top cwur , the numbeo range life cost to consider and finally the year .Finally Is possible by clicking on the moon to change the lightmode and by clicking on the hat is possible moving to the Dean's mode.

V. DEAN

This application is also for those Deans that are looking for ways to convince more Students to apply to their universities, offering them a tool to easily measure quality of education, international outlook, quality of teaching etc. and how they can improve them in order to place higher in rankings.

i. Legend

The Legend permit to choose which university show in the radar chart and in the connected by clicking the circle of each item.



Figure 4: Sapienza University of Rome, University of Turin and Bologna University in 2019

ii. Radar Chart

A radar chart is a two-dimensional chart type designed to plot one or more series of values over multiple quantitative variables. Each variable has its own axis, all axes are joined in the

center of the figure. We decided to apply it on the *citations*, *research*, *income* and *international outlook* in order to give to the Dean the possibility to see immediately five attributes that help the growth of their university. The not opaque instances correspond to the ones chosen in the legend.

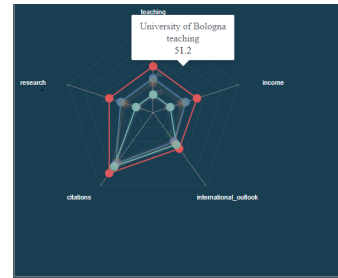


Figure 5: Sapienza University of Rome, University of Turin and Bologna University in 2019

iii. Lollipop Chart

The idea behind the lollipop chart is to allow the Dean to further compare their university with other universities of the same country by using three different attributes: *Quality of Education*, *Quality of Faculty* and *Influence*. Since the lollipop plot shows the relationship between a numeric and a categorical variable, we placed the universities on the x axis and the values of the attributes on the y axis.

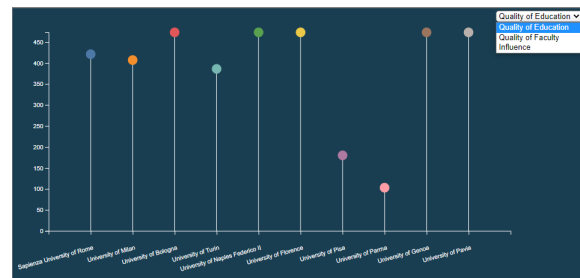


Figure 6: Italian Universities for the 2019 academic year

iv. Connected Scatter Plot

The scatterplot permit to the Dean to visualize for the attributes Teaching, Research and Citations how each university of that country and

for the range computed involved between 2012 and 2009. The not opaque instances correspond to the ones chosen in the legend.

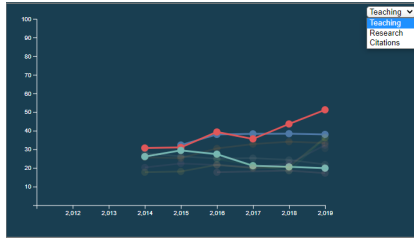
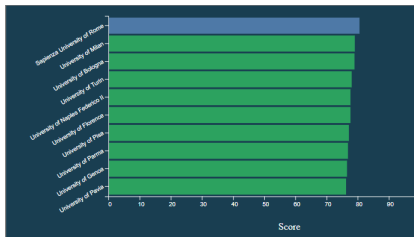


Figure 7: *Sapienza University of Rome, University of Turin and Bologna University evolution in Teaching*

v. Scatter Plot

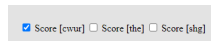
It works just like the scatter plot in Student mode, but this one takes as input the PCA for Dean instead.

vi. Bar chart



This bar chart works just like the bar chart in the student mode but in this case the selected university is shown in the color assigned in the legend.

vii. Header



The header for this part permits to the Dean to select on which score doing analytics. As in the other mode it is possible by clicking on the moon to change the lightmode and by clicking on the hat is possible moving to the Dean's mode.

VI. INTERACTIONS AND ANALYTICS

This project has many possible interactions permitting the two users to perform a better analysis of the dataset created from us for a better achievement of their goals.

- **Parallel coordinates:** it allows brush selection on each axis performing multiple filters on the data modifying the content of the other views;
- **Geographical map:** it permits the selection of the countries and the two actions explained above in the respective section: pan and zoom;
- **Scatter plot:** it interacts with the parallel graph; the Student can make a selection on the parallel graph and then the selected universities will be depicted in the scatter plot. The Student can then brush and zoom in to take a better look at the results.
- **Header:** in the Student mode, the application allows the student to filter universities by year, life cost and cwur score; in the Dean mode, the application allows to filter again the universities by year but it also allows to filter the data on the average of the scores selected in the checkboxes.
- **Legend:** it permits to better distinguish any universities from the others by assigning a color to it, just by clicking on the circle.

VII. STUDENT'S EXPERIENCE

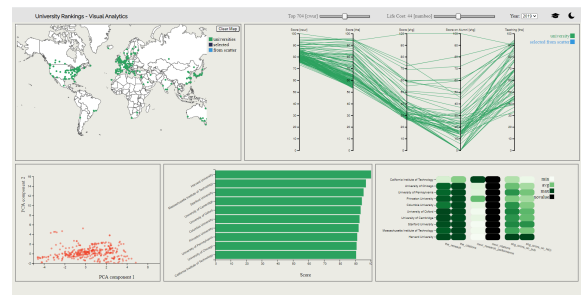


Figure 8: *Student mode in lightmode*

The student's experience starts from the sliders of type range located in the header i.e Top n cwur and Life Cost. From the Top n cwur slider the student can choose the set of universities on which they will perform the analysis. We gave this possibility because the dataset is too big, so the student can start the analysis from a small group of universities and then extend the research on more universities if they didn't find what they were looking for. For this slider we chose the cwur ranking because in our opinion it gives more importance to the student: indeed, because within its seven macroareas, "quality of education" and "alumni employment" make up the fifty percent of what determines the final position in the ranking. Indeed, from cwur documentation:

- Quality of Education, is measured by the number of a university's alumni who have won major academic distinctions relative to the university's size (25%)
- Alumni Employment, is measured by the number of a university's alumni who have held top executive positions at the world's largest companies relative to the university's size (25%)

From the Life Cost slider the student can choose the set of countries that are within that range of life cost values. The student can also use a drop down menu to choose the year, to further filter the dataset.

The parallel graph is the heart of the student mode: when the student interacts with the parallel coordinates chart all the other charts change accordingly. Parallel chart uses the following attributes: *score cwur*, *score the*, *score shg*, *score on alumni* and *the teaching*. We explained earlier in the *Dataset* section what these attributes mean.

Then the student can pick from the map the countries they are most interested in and from the scatterplot, thanks to the PCA components, they will be able to see how the universities differ from each other on the attributes used in the Student's PCA. Finally, the heat map allows the student to compare the universities on the three different rankings on *Research Perfor-*

mance and *Citations*. The heat map could also be used as a way to check which of the three rankings is the most reliable one: if two out of three rankings "behave the same way", then maybe the Student should be careful with the ranking that behaves differently.

VIII. DEAN'S EXPERIENCE

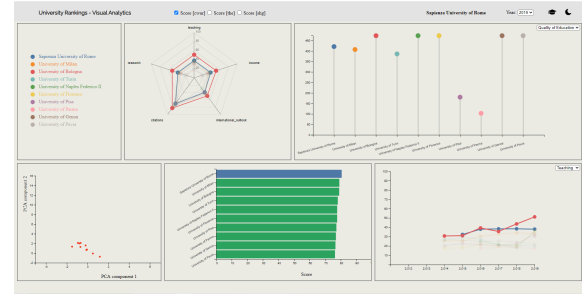


Figure 9: Dean mode in lightmode

The Dean's experience starts choosing the university from the dropdown menu. Once the Dean chooses the university, the application will show all the universities of that specific country, since laws, investments, way of teaching etc. differ in each country. In this mode will be shown the first ten universities or the universities that are the closest to the Dean's University and each university was assigned a different color to make the analysis easier. The ten universities shown are selected based on the scores selected in the checkboxes in the header. These colors are used in the lollipop, in the connected and in the radar chart too. On the radar chart we paid attention to the following attributes:

- the teaching : it examines how the university is perceived from the outside from the geographical point of view and also from the single subjects.
- international outlook: the ability of the university to attract new students, which is seen as a sign of global success.
- income: how the university can help the economy with inventions and innovations.

-
- the citations: this indicator focuses on how the university can spread knowledge and new ideas.
 - the research: it estimates the reputation of the university concerning the research, as obtained from the annual Academic Reputation Survey.

[6] <https://pandas.pydata.org/>

[7] <https://scikit-learn.org/stable/>

The student has the possibility to show the universities that prefers by clicking on the legend. In this way the color of the university will change from opaque to the corresponding color assigned in the legend.

In the lollipop chart the Dean can compare the universities listed in the legend on three different attributes :

- Quality of Education measures the average of alumni that won important awards and medals that are related to the size of the university. With alumni we refer to the students that got bachelor's degree, master's degree and Phd.
- Quality of faculty: measures the average of alumni that won important awards and medals.
- Influence: measure the number of papers that were published on influential magazines.

In the connected scatterplot the Dean can see the evolution through the years of the universities listed in the legend, on three different attributes : *Research*, *Teaching* and *Citations*. The barchart instead shows the average of the scores selected in the checkboxes.

REFERENCES

[1] <http://www.shanghairanking.com/>

[2] <https://cwur.org/>

[3] <https://www.timeshighereducation.com/world-university-rankings>

[4] <https://www.numbeo.com/cost-of-living/>

[5] <https://d3js.org/>