



CAMPUS  
DE EXCELENCIA  
INTERNACIONAL

## DESIGN OF A NEW INTERACTIVE DATA ANALYSIS TOOL

### ABSTRACT:

This report explains the different steps of the design and implementation of a new Visual Analytics tool, it will be structured according to the four design abstraction levels. This tool is a Shiny Application developed in R language which aims to make comparisons between university masters in France regarding to the professional insertion statistics.

### SKILLS USED:

- Programing in R
- Development of a Shiny App
- Data Visualization
- Data Analysis

DEVELOPERS: Henri DESQUESSES  
Thomas JASSEM

REPORT AUTHORS: Henri DESQUESSES  
Thomas JASSEM

DATE: 21/01/2018



## Table of Contents

I.	Problem characterization in the application domain.....	3
II.	Data and Task Abstractions .....	3
1.	Data .....	3
2.	Task abstraction .....	4
III.	Interaction and visual encoding .....	5
IV.	Algorithm Implementation .....	8

## Table of illustrations:

Figure 1: Interactive map .....	5
Figure 2: Ranking .....	6
Figure 3: Grade attribution view .....	7
Figure 4: University information view.....	8



## I. Problem characterization in the application domain

The goal of this application is to have an easy access to several useful statistics about masters available in France regarding student professional insertion.

Thus target users of our application will not necessarily be professional data analysts but more lambda citizens who wants to get global or precise information on the offer. We will focus on the main target who can be interested in it: **students**.

For students the use will be the following, a student just finished high-school, he/she wants to choose the university in which he/she will do his/her studies but does not know anything about those are available and how to choose among them. The only things he/she knows is the location in which he/she wants his/her university to be, because of friends who told him/her that this or this city is great or because he wants to live at the cost or any other reason. Then the only thing that he/she is interested in is the professional insertion behind the diploma because studies are not for fun but to have a good condition and he has no real idea of what to do.

He/she can also know exactly in which university he/she wants to go and which master to do.

In any case his/her questions must be the following:

- Q1: What are the universities in this region/city?
- Q2: Which academy/university/master is the best?
- Q3: Are there a lot of masters in?
- Q4: Is the university fame with companies?
- Q5: What is the becoming of the students?
- Q6: Do they earn a lot of money?
- Q7: Are there a lot of students to meet people and how many girls (for males only)?

The problem is that there are a lot of websites which make available all the information about universities about professional insertion and a lot of other statistics including rankings, but their use is never really intuitive and friendly which can discourage the less motivated one. The goal of this application is to give to anyone the opportunity to visualize all the information he/she needs in a really simple way.

## II. Data and Task Abstractions

### 1. Data

Data on which the application is based on a dataset which comes from the French open data website. The dataset contains 4275 located records of plenty of different variables which can be gathered in several categories:



Information on the university:

- Etablissement (Establishment)
- academie (academy)
- domaine (field)
- discipline (course)
- de\_diplomes\_boursiers (percent of scholar)

Information on the survey:

- nombre\_de\_reponses (number of answers)
- taux\_de\_reponse (answer rate)
- nombre de femmes (number of women)

Information on the type of job:

- situation (situation)
- taux\_dinsertion (insertion rate)
- emplois\_cadre\_ou\_professions\_intermediaires (manager jobs)
- emplois\_cadre (manager jobs)
- emplois\_stables (stable jobs)
- emplois\_a\_temps\_plein (full-time jobs)
- emplois\_exterieurs\_a\_la\_region\_de\_luniversite (job external to the region)
- taux\_de\_chomage\_regional (regional unemployment rate)

Information on the incomes:

- salaire\_net\_median\_des\_emplois\_a\_temps\_plein (average income)
- salaire\_brut\_annuel\_estime    salaire\_net\_mensuel\_median\_regional    (average regional income)
- salaire\_net\_mensuel\_regional\_1er\_quartile
- salaire\_net\_mensuel\_regional\_3eme\_quartile

But as we said before, the goal is to have a simple interface and an easy access to the information so we decided to keep only the features which seemed to us essential for the target case of use. So we can focus the user attention on the most important.

## 2. Task abstraction

In order to allow the user to get quickly the data wanted the following tasks will be performed by the application:

- T1: Visualization of rankings as a function of several criteria
- T2: Geographic search
- T3: Assignment of a grade to each university
- T4: Aggregation by academy to have a global view
- T5: Accurate search of university and master

So our application is consuming data presenting information to a user and also discovering new knowledge by generation of our own grade classification.

Users will do lookups by searching directly the university in which they are interested in to see its information or looking at the different universities in a specific area. They will also to browse and explore taking a look to the different rankings and the classification of universities. Regarding to the queries, users will be able to either identify or summarize data with some aggregation and filtering functionalities.

### III. Interaction and visual encoding

In order to perform the tasks described previously and answer all the questions users can ask we have decided to use 4 different views: an interactive map, a multiple ranking view, a grade attribution view and finally a university information view.

#### Interactive map:

It is a geographic arrangement for the data under consideration where the universities are gathered in geographic clusters which deploy when they are zoomed. On hover on a marker you access to university name and on click information about the university is displayed in a pop-up.

It is possible to change the university scale to the academy scale which is more global by selecting it.

Associated tasks and questions: T2, T4, Q1, Q2, Q3, Q4, Q6, Q7

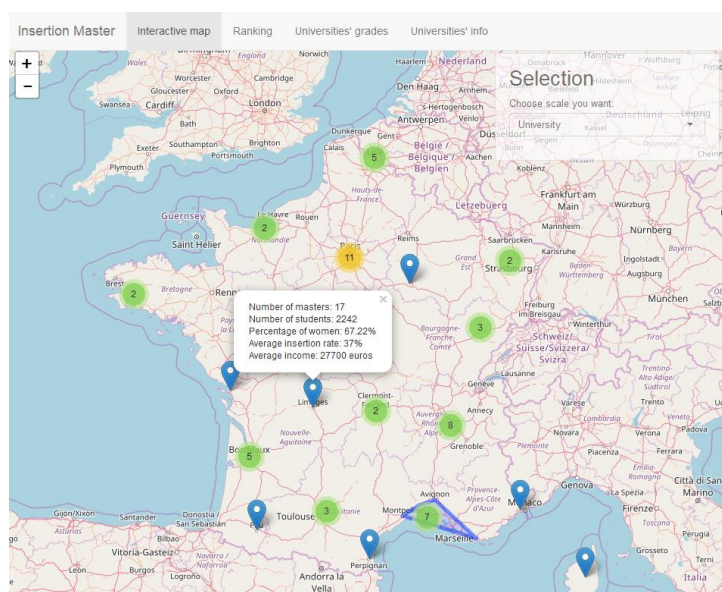



Figure 1: Interactive map

## Multiple ranking View:

This view is designed to give university rankings as a function of three selectable criteria: average income, insertion rate and proportion of quick managers. The ranking is rendered graphically with a bar plot which in addition to give the rank, gives the gap between each universities and makes trends visible. For each of three criteria you access to the values on the plot by hover. For the income one the average income of the region is given as a landmark and permit to compare to the region in addition to compare to the other universities.

It is possible to interact with the plot in several ways which are proposed on hover on the plot:

- Zoom in or out
- Shift the axis
- Visualize the values (both region and university values for income by choosing  )
- Download the plot as a PNG

Again it is possible to change the university scale to the academy scale by selecting it.

Associated tasks and questions: T1, T4, Q2, Q4, Q5, Q6

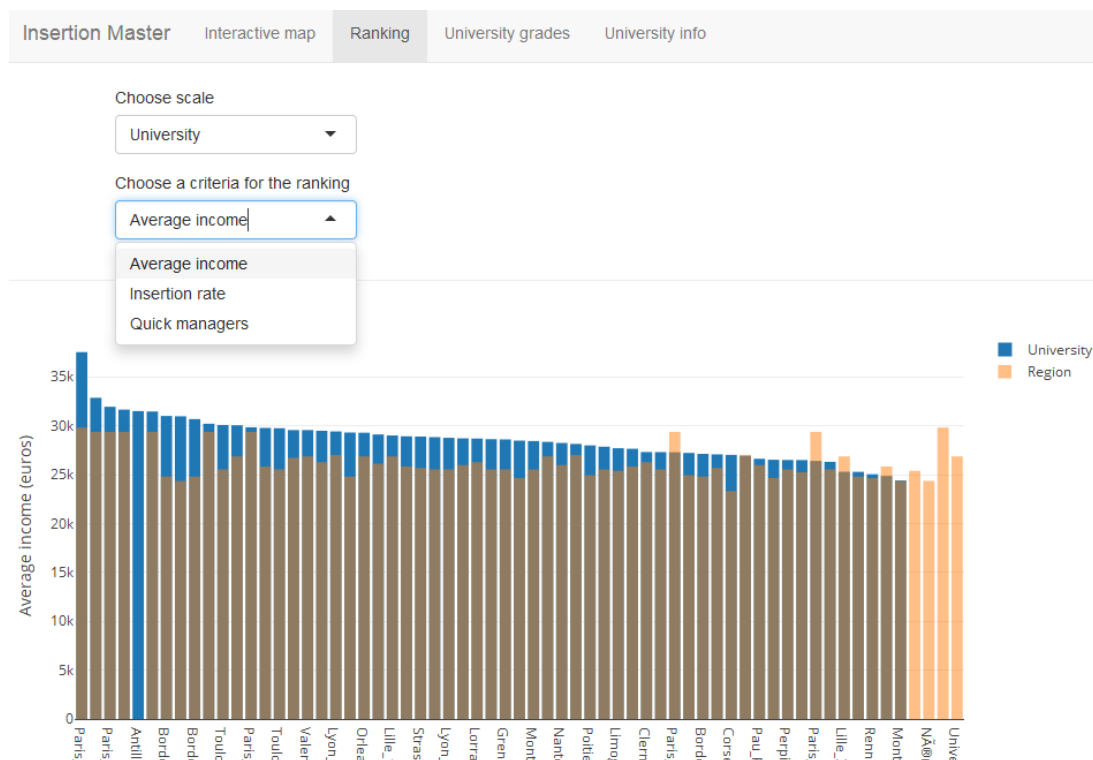


Figure 2: Ranking

## Grade attribution view:

This view permits to see on a basic scatter plot the results of our university clustering by two features (average income and insertion rate). The fact that it is impossible for the user to choose the features with which he/she wants to do the clustering is voluntary, because other choices produce irrelevant classifications. Again the values are displayed on hover on the points.

Associated tasks and questions: T3, Q2, Q4, Q6

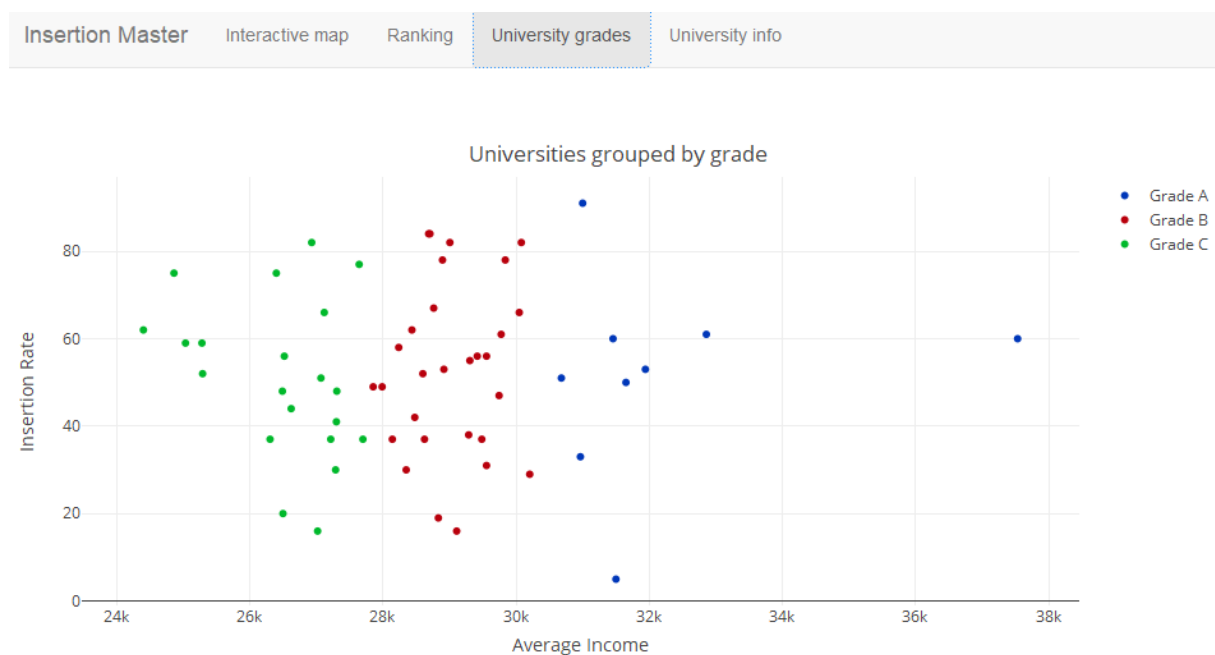


Figure 3: Grade attribution view

## University information view:

In this more traditional table-based view it is possible to select a university to visualize all the useful information related to it. When you have chosen a university it is possible to see an overview with the following statistics about it: grade, number of students (population), number of women, insertion rate, average income, percentage of scholar and proportion of quick managers.

Below the list of the masters available in the university are displayed and it is possible to rank them as a function of any of the features by clicking on the header. It is also possible to search for a specific master with its title.

Associated tasks and questions: T1, T3, T5, Q3, Q4, Q5, Q6, Q7

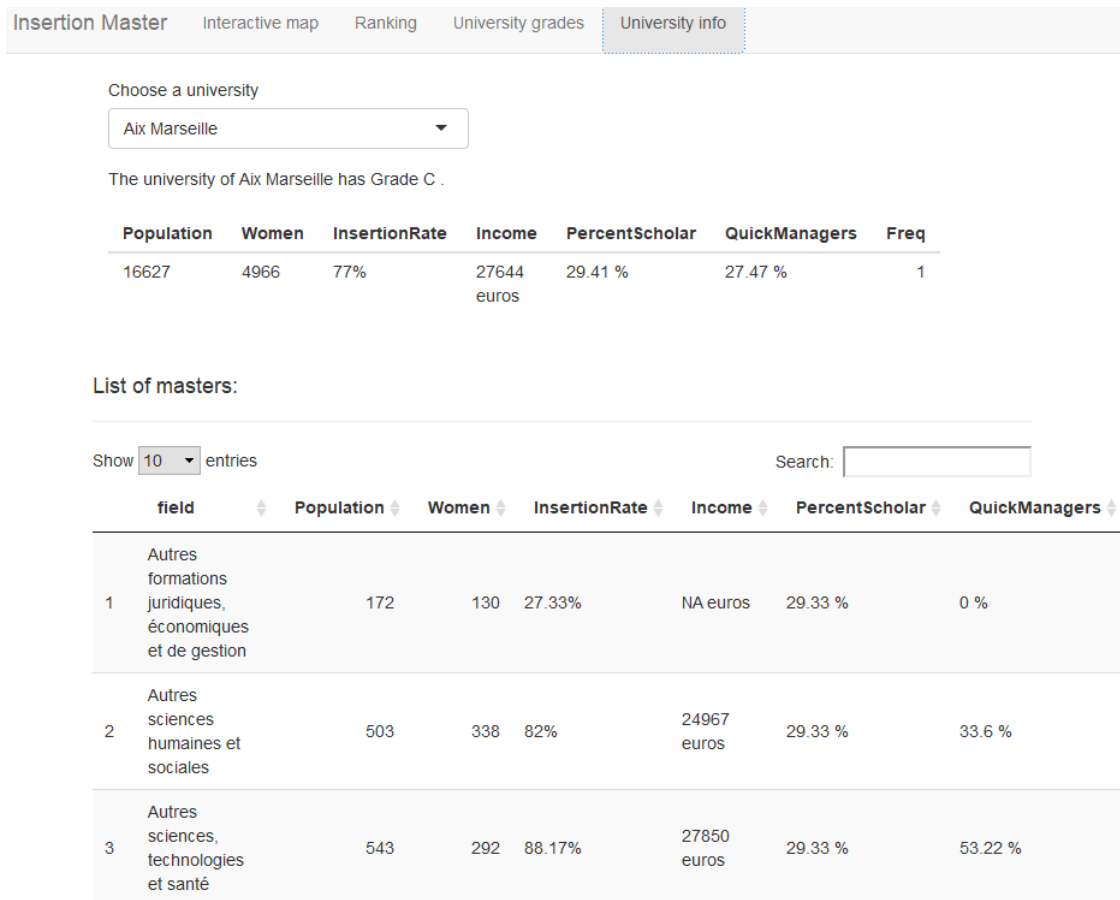


Figure 4: University information view

## IV. Algorithm Implementation

Most as possible, queries are made in global.R file in order to simplify the code in server.R file in which there is the processing of all views. Useful functions are coded in helpers.R file.

### Interactive map:

The map is from Leaflet library which offers a really natural and refine design, adapted to a simple view to just plot our markers. The name of university or academy (depending on the choice in the selectInput) is displayed on hover on a marker and on click on it a popup shows the following information:

- Number of masters
- Number of students
- Percentage of women





- Average insertion rate
- Average income

The markers are geographically clustered again to show a simpler and more refined view.

### Multiple ranking view:

After having tested several libraries to plot our rankings, we have chosen to use Plotly which offers a friendly customization and integrated useful interactive options which makes each plot easier to understand.

Two selectInputs permit to choose scale (university or academy) and criteria (average income, insertion rate or proportion of quick managers) of the ranking.

### Grade attribution view:

Grades are attributed to each university by clustering which is done with k-mean in global.R. There are 3 clusters to which we attribute a grade (A, B or C). The seed is fixed because k-mean function has a random part and the clustering was different each new time we ran the application. This was a problem because cluster names make sense.

Then clusters are shown in a basic scatter plot done again with Plotly.

### University information view:

After having chosen the university in a selectInput, it is possible to see all the information about it. In a first simple table there is a recap about the general statistics about the university (grade, number of students, number of women, insertion rate, income, percentage of scholar and proportion of quick managers). Detailed statistics about each master are visible below in a datatable from DT package which provides filtering, pagination, sorting, and many other features in the tables.