You can really learn about sexuality through a MOOC!

Vladimir E. Cuevas, Juan Carlos Oyuela y Andres Segura Tinoco

December 4th, 2018

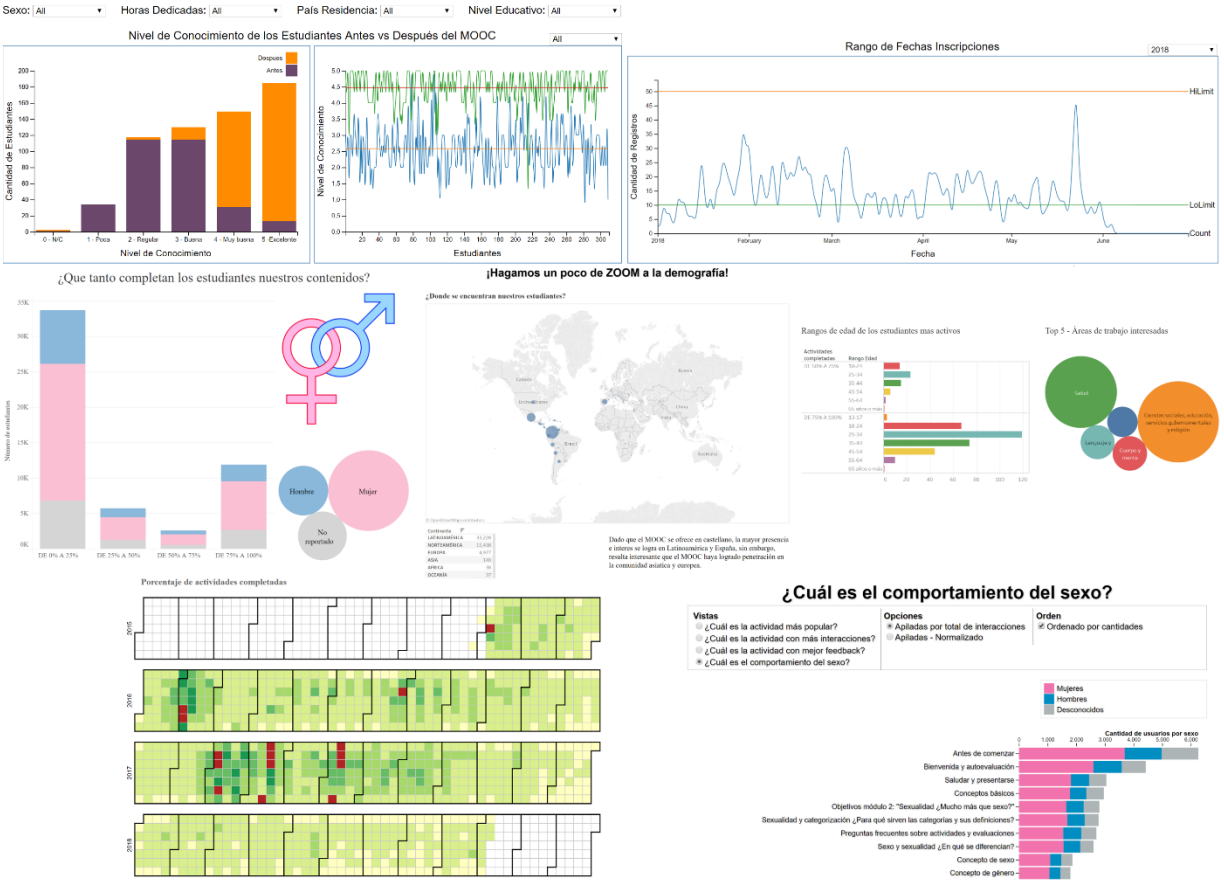

Fig. 1. Visualization idioms - You can really learn about sexuality through a MOOC!

Abstract — The research group formed by teachers Angela Maria Rojas and Elvia Esther Vargas, who belong to the Psychology Department of the Universidad de los Andes, wishes to achieve an understanding and use of the data collected during the last 3 years and which refer to the MOOC "Sexuality ... much more than sex ", provided on the Coursera platform, in order to verify the effectiveness of the online course. This article illustrates the way in which the problem was approached from a perspective of visual analytics information, generating great value to the research group. For the development of this project, the possibility offered by Coursera of accessing the download of its databases and the guide to understand them was taken advantage of, in such a way that it was possible to solve a set of specific tasks raised by the users, through interactive visualizations built in D3 framework and the Tableau system. Thanks to these visualizations, it was possible to understand and discover the way in which students achieve a significant improvement in the capacities to approach topics related to sexuality, see the most interesting activities of the course are and additionally understand the student population from a point of demographic view. This exercise allowed to have a starting point, based on descriptive analytics, which offers a fundamental tool for the research team in the decision making on the MOOC and, additionally, to guide its promotion towards specific populations, turning it into a means with the which could be analyzed and explore the other 24 MOOCs offered by the Universidad de los Andes, since its data structure is identical.

Index Terms — DataViz, MOOC, Sexuality, Coursera, Uniandes, D3, Tableau.

1. Introduction

The Universidad de los Andes, within its process of innovation and exploration of the educational scheme, has developed a set of virtual courses (MOOCs) that aim to extend teaching far beyond its physical boundaries. The Psychology Department has created the MOOC "Sexuality ... much more than sex", which aims to improve the students experience to talk about sexuality with girls, boys, students, partners or colleagues. This course is implemented within the COURSERA platform, from which it is intended to perform an analysis and construction of data visualization, which strengthen the positioning of the course and the possibility of improving its contents.

2. Problem approach

Currently the course in reference is the most successful within the set of MOOCs offered by the Universidad de los Andes, however, some members of the community feel skepticism about this virtual scheme, especially because it is a sensitive issue, so that the belief that you cannot transmit these key concepts effectively, through a virtual platform has been evidenced.

The Psychology department, and, the professors who developed this idea want to show the community that the course is successful, that people get knowledge and value when they finish the lessons and, above all, that this is an efficient scheme of knowledge transfer.

Therefore, our solution aims to attack this problem from the data visualization viewpoint, since currently the queries and graphics offered by Coursera, are too basic and fail to take advantage of the potential of the students' behavior embedded within the available information. More details about the solution can be found in the What, Why and How chapters.

Users of the solution:

- Teaching group (course managers)
- CONECTA-TE Innovation Center
- Representative of the Ministry of Education

3. State-of-the-art

Currently, for people who are in an average age range it is normal to take online courses because of the boom they have had since, in 2007, David Wiley of the Utah State University, started exploring this market. After this, many online learning platforms have been raised with great success, this document presents Coursera, founded in April 2012, as a case of success and study since its founders, Andrew Ng and Daphne Koller (current teachers from Stanford University) never thought that this would cause so much impact and that it would climb so quickly [1].

The main objective, according to John Bravman (vice president of undergraduate education), wanted to make more space in the curriculum to generate a significant commitment between Faculty and students. Although it seemed that the student-university relationship was colder in this way, what was expected was to be able to devote more time to more meaningful activities between the students and the teacher, instead of repeating the same classes repeatedly.

Once they formed Coursera, Koller and Ng quickly approached other universities to become partners in the content production.

Coursera in numbers

As of February 2017, the total number of members is 149 in 29 countries. Coursera works mainly with universities and colleges, but also with governments. The university's partners include the University of São Paulo in Brazil, the University of London in the United Kingdom, the Indian School of Business in India, the University of Yonsei in Korea, and institutions such as Yale and the University of Pennsylvania. It should be noted that the University of Los Andes, where this work is carried out, has courses within the platform [2].

As of 2017 Coursera offers complete masters. They began with the Master of Innovation and Entrepreneurship (OMIE) of HEC Paris and the Master of Accounting (iMSA) of the University of Illinois but continued to offer the Master of Science in Computer Science in Data and the Master of Business Administration (iMBA) both from the University of Illinois.

MOOC: Sexualidad...mucho más que sexo

This document focuses specifically on the Massive Open Online Course (MOOC) course of the University of Los Andes, called "Sexuality ... much more than sex" dictated by:

- Elvia Vargas Trujillo, Associate Professor of the Psychology Department and Phd from the Autonomous University of Madrid in Behavioral Sciences Methodology.
- Ángela María Rojas Martínez, Consultant in Sexual Education and PhD in Psychology at the Universidad de los Andes Colombia.
- Marta Carolina Ibarra A, PhD student at the Universidad de los Andes.
- Carlos Hermosa Bosano, Clinical Psychologist at the San Francisco University of Quito (Ecuador) and master's in psychology at the Universidad de los Andes (Colombia).

The course lasts four weeks with a dedication of four to seven hours per week. It is dictated in Spanish and as of the date of completion of this document it has a score of 4.8 stars with a vote of 1860 people. The Universidad de los Andes currently has 29 courses in Coursera.

With respect to the honors granted in the course, this MOOC always grants them since the course is designed so that they always finish it, since for those responsible for the course it is not important to look at whether the student is good or bad, but to understand if it was achieved improve the perception of the student's competences in order to talk about sexuality issues.

Coursera's Introductory Human Physiology Course: Factors that Characterize Successful Completion of a MOOC - Deborah Engle, Chris Mankoff, and Jennifer Carbrey

This study examines students in the MOOC of Coursera, introductory human physiology. Of the 33,378 students who accessed the course, around 15,000 responded to the pre-course survey items on their age, educational background, English proficiency, and plans to participate in the course. The students who completed the pre-course survey in groups were classified according to the number of exams completed and the level of achievement of the corresponding course. Chi-squared goodness-of-fit tests were used to analyze the distribution of student responses in the pre-course survey and the associated achievement level [3].

The results of the analyzes showed the following ingishts: of the students who responded to the pre-course survey and passed with distinction, a higher percentage self-identified as fluent in English, while a smaller percentage self-identified as beginners. Students with graduate degrees were more likely to pass the course or pass with distinction than students with only some college experience or a bachelor's degree. Students who completed some or all of the exams were more likely to self-report to complete all course activities than students who did not take any exam. A greater proportion of students who passed the course or passed with distinction published two or more times in the discussion forum of the course. Understanding the MOOC students and the characteristics that lead them to their success allows the modification of the courses to increase the student performance and can also inform the teaching in the traditional classroom.

4. First results

From the gathering of information obtained in the first meetings, where the data sets and the general needs of the client were collected, other work sessions were carried out where the main tasks and the secondary tasks of the project were thoroughly reviewed and prioritized, Ensuring uniqueness in the understanding between the business expert and the visualization expert. As a result of this detailed review, the scope of the construction of 3 main tasks and 7 secondary tasks were defined, based on their importance and impact on the client's research process.

The sessions of understanding have been framed within the framework seen in the course of Visual Analytics of the Universidad de los Andes, which has facilitated the work of confirming the data

sources (What), prioritization, selection and debugging of the activities (Why), and tuning the proposed idioms (How).

5. WHAT

The main dataset used for visualization is of the temporary and static type and contains the relevant information of the students who register daily in the MOOC. The attributes of the dataset are the following:

- Student ID: ordinal and sequential
- Grade Note: quantitative and sequential (from 0.0 to 1.0)
- Country Origin: categorical
- Country Residence: categorical
- Gender: categorical
- Education Level: categorical
- Age Range: categorical
- Previous Completed Courses: categorical
- Date Enrollment: ordinal and sequential (only dates greater than 2010)
- Start Date: quantitative and sequential (only dates greater than 2010)
- Last Update: quantitative and sequential (only dates greater than 2010)

For the main task number 3, another table-type dataset with static availability is used, which contains the activities (items) that the students are doing during the course. The following attributes are used only:

- Student ID: ordinal and sequential
- Item Name: categorical
- Item Sequence: quantitative and sequential
- State: categorical (Started and Completed).

Another table-type dataset with static availability is also used, which contains the information of the course users who carry out activities. The following attributes are used only:

- Uniandes_user_id: ordinal y sequential
- Reported_or_inferred_gender: categorical (woman, man y non-disclosed)

Similarly, another table-type dataset with static availability is used, which contains the information of the feedbacks of each activity provided by the user. The following attributes are used:

- Course_item_id: ordinal y sequential
- Feedback_rating: categorical ("Thumbs up" y "Thumbs down")

In a complementary way we use the initial and final surveys that students take to add information to the demographic or professional profile. The table-type dataset consists of:

- Student ID: ordinal y sequential (key with other datasets)
- Update date: ordinal and sequential
- Question: Categorical
- Answer: Depending on the question, the possible answers can be: Categorical or Sequential Ordinals.

Finally, to fulfill the main tasks of the visualization, the following attributes will be derived:

- Grade Level: categorical, derived from Grade Note
- Activity Level: categorical, derived from the number of Items that a student sees during the course
- Daily Inscriptions: quantitative and sequential

The result of the surveys are data that support the main task.

6. WHY

The tasks that are listed below are related to the analysis and understanding of the data collected, so that the organizers can improve the image and the characteristics of the course and, in turn, maintain the active participation of the students.

All the tasks were reviewed with the organizers of the course, to refine them and prioritize them, according to the impact of these. Therefore, 3 main and 7 secondaries were defined, all relevant to the business.

Main Tasks

TP1: Identify the characteristics that indicate that the MOOC is effective and that it really improves the competences and abilities to speak of "Sexuality" in the participants, even though it does not have a teacher in person **(Identify - Features).**

TP2: Discover the distribution of the demographic variables of the students who complete the course, to identify which communities show greater interest in completing and finishing the course **(Discover - Distributions).**

TP3: Identify the most popular and least popular activities available in the course, comparing the number of times an activity was completed against the number of times an activity was not completed **(Part-to-whole - Relationship)**.

Secondary Tasks

1. Determine if there is an activity or item in which students withdraw from the course on a frequent basis, in such a way that it is possible to identify a key moment of desertion **(Locate - Outliers)**.
2. Explore the distribution and counting of students globally, considering that the course is dictated online **(Explore - Distribution)**.
3. To profile the users of the course, and to identify the most active populations **(Compare-Features)**.
4. Find the periods or ranges of dates with the highest enrollment of students, in order to strengthen and focus on certain populations the marketing campaigns of the course, in such a way that it is possible to increase the annual participation rate **(Locate - Outliers)**.
5. Identify the most popular and least popular activities (items) available in the course, through its Like / Dislike **(Find / Search - Trends)** and **(Lookup - Values)** type rating.
6. Identify the most popular and least popular activities available in the course, through the number of interactions each item has, where each activity that is initiated has a status of "initiated" and each activity that is completed is completed by the user changes status to "completed". The activity can be initiated and completed many times by the same user **(Find / Search - Trends)** and **(Lookup - Values)**.
7. Identify the most popular and least popular activities (items) available in the course, counterstand it against gender or sex **(Find / Search - Trends)** and **(Lookup - Values)**.

7. HOW

In the process of defining the visualization, considering the selected tasks and the data, the HOW proposal was made, which consists of the idioms described below:

TP1's Idiom, composed of 2 graphics (a bar chart on the left and a multi series line chart) juxtaposed:

- Marks: for the bar chart, vertical lines and for the ms line chart, points linked by lines.
- Channels: length on the Y axis to express quantity (of students in the bar chart and expectation for the ms line chart) and color hue to differentiate the categories of the variables (curves).
- How-Encode: for the bar chart arrange express on the Y axis, and on the X axis, separate, order and align for the possible answers of the surveys. For the ms line chart, arrange express for both axes.
- How-Face: juxtapose of the 2 charts (bar chart and multi series line chart).
- How-Reduce: the data can be reduced from a set of common filters, by gender, country of residence, hours dedicated and educational level.
- Coordinate views: multiform, since they use the same data, but with different encoding.
- Description: this idiom is composed of 2 juxtaposed graphs, which help identify characteristics or patterns that show the relevance of this MOOC.

TP2's idiom, Vertical Stacked Bar Chart, Horizontal BarChart, Bubble chart and Map view:

- Brands: vertical lines and Areas (Circles).
- Channels: length and 2D Area to express the number of students. Color (Hue) to separate the categories in the demographic variables Sex, Work area, Age range. The Y axis is ordered expressly. Area to express the volume of students by sex. Area for number of students and spatial location on the map by country of residence of the student
- How-Encode: grouping of the populations by quartiles in percentage ranges for the completeness of activities. Separate, order & align.
- How-Manipulate: change
- How-Reduce: aggregate
- Description: the number of students in the Y axis is presented, grouped by quartiles and complemented by the volume in each demographic variable.

TP3's idiom, a Stacked bar chart and Normalized stacked bar chart:

- Data: Table of multiple dimensions, 2 categorical attributes, 1 quantitative attribute.
- Data derived for Normalized stacked bar chart: 1 quantitative attribute, which is the normalized version of the original attribute.
- Marks: lines stacked horizontally
  - Glyph: object composed of sub-bars of different colors on top of each one. Only for the Normalized stacked bar chart, each of the bars is stretched to the maximum possible length showing percentages instead of absolute counts. For the stacked bar chart, absolute values are shown.
- Channels:
  - Length to express quantitative values
  - Hue, color: to express categories
  - Spatial region: one per brand
  - Aligned: The lowest category or component of the stacked bar (for both, stacked bar chart and normalized stacked bar chart).
  - Not aligned: other components of the bar or sub-bars.
  - The normalized version of the original attribute is used for the normalized stacked bar chart.
  - There are two ways of ordering, one is by order in which the activities are presented in Coursera and the second is by quantities of the quantitative attribute.
- Main task: (Part-to-whole relationship) and for the standardized version (part-to-whole judgments). The sum of the quantitative attributes by category must coincide with the normalized data, such as percentages, where the parties must add up to 100%. You can also use a single bar in a standard stacked bar chart to show this property with the most accurate length judgments channel.
- How-Manipulate - Change the order, to show more popular activities that were never finished, against interactions that the user makes about the activities, gender or gender of the user against the activities and feedback of the users about the activities, ordering all the aggregations two ways:
  - For quantities in descending order or in order that the activities are presented in the Coursera MOOC.
- How-Manipulate: (Select elements (activities and categories) with hover), to show totals (example, total by category or total per activity) and data that does not fit on the screen (example, module path, lesson and the lesson to the activity), through a composite infotip.
- How-Encode: (Separate, order and align with 2 keys matrix), for two (2) dataset keys projected onto stacked bar charts and normalized stacked bar chart.

With respect to idioms for secondary tasks:

- For secondary task 1, a vertical bar chart is proposed, which accounts for (summarize) which the last activity was carried out by each student, in order to identify if there is an activity where a significant group of students withdraw from the course. The marks will be vertical lines, and the channels will be length to express the number of students who arrived at said activity. The Y axis will be ordered expressly, and the X axis will use the separator, order and align to position the bars.
- For secondary tasks of 2 and 3, the use of Stacked Bar Chart Horizontal (Normalized), like that proposed for main task 3, is proposed as idioms.
- For secondary task 4, it is proposed to use an idiom consisting of a Calendar View and a Line Chart. Calendar View will summarize the trends of all years, while in the Line Chart, information will be displayed only for 1 specific year. In addition, the Line Chart, will show lines of the dash type that cut the Y axis, to define the tolerance zones from which the outliers are identified. The data of both graphs will be ordered expressly on the X axis. They will also be ordered and aligned on the Y axis, with respect to the number of entries per date. The line chart will be updated, based on an interaction of the type select over a combo box.

8. INSIGHTS

After an analysis on all points related to users and tasks, the state of the art, data exploration, analysis of relationships between customer needs, business questions, data, the possibilities offered by visual analysis and mainly the created visualization, the following list of insights was generated by task.

The most important insights of main task 1 and its secondary tasks are listed below:

- **I1**: Based on the surveys conducted, the effectiveness of the MOOC is observed in the students, since before studying it they have knowledge and expectations on the subject between regular and good (2.5 / 5.0 on average),

but after studying it, improves your knowledge between very good or excellent (4.5 / 5.0 on average).

- **I2**: Students who devoted more hours of study to MOOC (5 or more) stated that their knowledge about sexuality improved in a greater percentage, than those who devoted few hours to the course (4 or less), specifically, almost 10% of more knowledge about the subject.
- **I3**: The historical tendency of the inscriptions is very positive, since until 2016, for no day had more than 50 registrations, however, as of 2016, in mid-February and July there are peaks with more than 100 inscriptions However, the only worrying thing is that by 2018 the number of daily registrations has decreased with respect to 2016 and 2017.

The most important insights of main task 2 and its secondary tasks are listed below:

- **I4**: Although in principle it was thought that the balance of men and women participating in the course was similar, and even that the male population could be higher, we have detected that women are the ones who are most interested in MOOC contents, and not only for browsing, are the ones that complete most of the activities from the course, surpassing men 2.5 times in participation.
- **I5**: The population most interested in the contents of the course are health professionals and social sciences, who are between 18 and 44 years old, with the area of language and humanities as the protagonist with more than 50% of the population.
- **I6**: We have found a particular behavior with the volume of enrollments during the 3 years that the online course has taken place, where mainly the volume of registrations has been decreasing, however, there are some points in time where the registrations are going up, as is the case of June 10, 2017, the date immediately following a journal publication, evidencing the effectiveness of these advertising campaigns.

The most important insights of main task 3 and its secondary tasks are listed below:

- **I7**: There is a trend of between 41% for the most popular activity and 56% for the less popular activity of not completing the activity.
- **I8**: There is a correlation between the number of users by sex or gender, with respect to the use of these by activity. Women occupy the first place with a total of 50,248 users, those who did not report sex are in second place with 18,388 users, men are those who least see the course with a total of 18,223 users. Previously it was believed that most users were men.
- **I9**: 99.1% of the feedback from the course is positive.
- **I10**: Of the 86,859 users that appear in the database, most are women with 54% of use for the least popular activity and 69% for the most popular.

9. Used Technologies

For the development of the project, the following technologies were used:

- Sublime Text 3 was used as a development IDE.
- HTML and CSS, to design the website.
- JavaScript and the d3.js framework to create the graphics and the respective interaction with them and control the html itself.
- Tableau to make the graphics of the main task 2.
- GitHub to store the code of the visualization and the data used.

10. Conclusión

The effectiveness of the MOOC was verified in a significant way, demonstrating through the visualization of the data, the positive impact it generates on its students and the overall performance that the course has achieved, reaching mainly the young female population. It is concluded that yes individuals can learn effectively about a subject as sensitive as sexuality through a distance course, and that the success of its contents is unquestionable thanks to the fact that they have been gestated from a highly professional perspective, evidencing that most relevant and best-qualified activities are those related to conceptual issues.

11. Acknowledgments

We want to thank cordially, Ph.D. Ángela María Rojas Martínez and Ph.D. Elvia Vargas Trujillo, both professors of the psychology department of the Universidad de los Andes, for their knowledge and support throughout the project.
https://psicologia.uniandes.edu.co/index.php/profesores/elvia-vargas-trujillo
https://www.coursera.org/instructor/arojas

12. References

[1] C. Severance. Teaching the World: Daphne Koller and Coursera, University of Michigan. August 2012. https://ieeexplore.ieee.org/document/6272250.
[2] Wikipedia. Coursera. Citado en octubre del 2018. https://es.wikipedia.org/wiki/Coursera
[3] D. Engle, C. Mankoff, J. Carbrey. Coursera's introductory human physiology course: Factors that characterize successful completion of a MOOC. April 2015. http://www.irrodl.org/index.php/irrodl/article/view/2010.

13. Glossary

[1] MOOC: Massive Online Open Courses (or massive and open online courses). The massive courses are constituted as the evolution of open education on the internet.
[2] COURSERA: Virtual platform that offers elements to be able to offer, distribute and manage MOOC's.
[3] INSIGHT: Ability to achieve a deep understanding of something that is not obvious, in our context, through the visualized data.