Jun 25th, 9:00 AM

Making posters to understand statistics: Towards a didactical approach in communication design

Michele Mauri
Politecnico di Milano, Italy

Simone Vantini
Politecnico di Milano, Italy

Beatrice Gobbo
Politecnico di Milano, Italy

Tommaso Elli
Politecnico di Milano, Italy

Elena Aversa
Politecnico di Milano, Italy

See next page for additional authors

Citation

Mauri, M., Vantini, S., Gobbo, B., Elli, T., Aversa, E., Benedetti, A., de los Ángeles Briones Rojas, M., and Colombo, G. (2022) Making posters to understand statistics: Towards a didactical approach in communication design, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), *DRS2022: Bilbao*, 25 June - 3 July, Bilbao, Spain. https://doi.org/10.21606/drs.2022.597

Authors

Michele Mauri, Simone Vantini, Beatrice Gobbo, Tommaso Elli, Elena Aversa, Andrea Benedetti, María de los Ángeles Briones Rojas, and Gabriele Colombo

Making posters to understand statistics: towards a didactical approach in communication design

Michele Mauri[a], Simone Vantini[b], Beatrice Gobbo[a*], Tommaso Elli[a], Elena Aversa[a], Andrea Benedettti[a], Maria De Los Ángeles Briones Rojas[a], Gabriele Colombo[a]

[a] DensityDesign Lab, Politecnico di Milano

[b] Politecnico di Milano

*Corresponding author e-mail: beatrice.gobbo@polimi.it

https://doi.org/10.21606/drs.2022.597

Abstract: The paper describes a didactical approach aimed at introducing statistics to communication design students at the master level. The approach is aimed at helping them in developing a critical attitude towards data manipulation and information visualization, acknowledging a lack of education on such areas despite their growing relevance in the communication design field. In previous experiences, we observed how theoretical lessons of statistics were inefficient because perceived as distant from the communication design practice. We therefore adopted a "thinking-through-doing" approach: instead of asking students to study statistical methods, we asked them to design a poster explaining them. In the paper we present the didactical experience discussing the outcomes. The approach brought students to better understand statistical methods and the implications of the decision taken in setting the analysis. In conclusion, we argue that it succeeds in making students more aware of the intersections between design and statistics.

Keywords: communication design; statistics; visual explanations

1. Introduction

The main need from which this paper stems is how to provide statistical learning to communication design students. Many questions could arise: why should we teach statistics to communication design students, for what purposes, and why we found it so difficult?

The overall goal of the presented didactical experience is twofold: on one hand, teach them various statistical methods that can be directly useful in their design activities; on the other hand, make them aware of the processes behind the production of data through statistical means and their influence on communication design.

The presented didactical experience is part of a larger course focused on information visualization at Politecnico di Milano, at the second year of the Master's Degree in Communication

Design. The course, called “Final Synthesis Design Studio”, is structured in 180 hours of lesson and about 200 hours of personal work with 50-60 students attending the course. Students have a solid background in visual and graphic design, but no previous experience in working with data and in information design. For simplicity, in this paper we will refer to communication design students simply as “students” or “design students”.

The course is structured in three different modules: information design and data visualization, statistics, and semiotics: this interdisciplinary approach introduces not just the needed design knowledge but also relevant concepts from the surrounding disciplines. In its seventeen editions, various didactical strategies were defined to equip students with the needed knowledge to deal with information visualization (Mauri 2020; Mauri et al. 2019; Valsecchi et al. 2010). Given the overall scale of the course, students work in the field of information design in three phases of work supervised by the design teachers, while semiotics and statistics are provided by dedicated teachers coming from respective fields.

In our experience, students are not used to working with data: at the beginning of the course, in the self-evaluation assessment, most of the students declare that they have never worked with data[1]. Data is seen as something distant from their domain, and it seems irrelevant to understand how data is produced since it’s handled by someone else, and they should focus just on its visual representation. The misconception of seeing data as raw material, as something “objective”, is deeply rooted in popular culture. In the academic literature instead, data is seen as the result of a series of actions:

> While many analysts may accept data at face value, and treat them as if they are neutral, objective, and pre-analytic in nature, data are in fact framed technically, economically, ethically, temporally, spatially and philosophically. (Kitchin 2014:28)

The risk of focusing on the visual representation without having a critical view of data production could reinforce such perception in public opinion (Correll 2019), or worse, bring to “weaponised design” (Diehm 2018): misuses of the work done by designers who are unaware of, or indifferent to, the political power of the devices, interfaces and data structures that mediate access to information.

For these reasons, a module in statistics was introduced, to highlight to students the similarities and dissimilarities between data production and the design process, and of enabling them to critically reflect on how their design is influenced by and influences data.

Concepts of statistics were introduced from the ninth edition of the course and become more substantial from the thirtieth edition. In this edition the module was framed as a theoretical one, meant to support the studio activities but without a direct connection. During

[1] On a sample of 36 responses, 19 students (52,8%) displayed a superficial knowledge of data analysis software, 7 (19,4%) never experienced the use of such software, 8 (22,2%) assessed to be familiar with them, while only 2 (5,6%) assessed to be proficient users. Similarly, they were given the space to express their expectations in the course: most of them claimed their lack of background in statistics and data analysis as well as their intention in learning the processes behind data visualization.

the course, most of the students looked scarcely interested, and the impression was that they struggled to understand the relevance of the module in the course. Additionally, this didactical approach reinforced in some students the feeling that statistical knowledge was useless both for their practice in general and their coursework. This feeling was further confirmed in the evaluations submitted by the students and provided by the institution each year at the end of the course. However, by discussing with former students now working in the information design field, they recognised the relevance of such a module. In the last editions of the course, as first delivery students were asked to design a visualization of data compiled from reputable sources on various issues. As members of the course's faculty (both design teachers, the statistics one, and assistants) we decided to reframe this delivery to experiment with a novel approach to teaching statistics in this context, preparing a hands-on exercise for the students.

Looking into literature, while it is possible to find examples of bringing design knowledge to computer scientists (Kerren, Stasko, and Dykes 2008; Lo, Ming, and Qu 2019) and in the usage of visualization for teaching statistics[2], however statistical training for communication design students emerges to be an understudied field. This paper presents a didactical approach to overcome this gap, followed by the outcomes and a first evaluation.

2. Methodology

2.1 Think through doing

A possible solution has been indirectly suggested by the literature related to Controversy Mapping, an approach to the study of society through the analysis of technological controversies. In presenting it, Latour (as cited in Venturini 2010) said:

> I have stopped, in the engineering school where I teach, to give a social science class: I only ask the young engineers to follow for one year, in real time, a scientific or technical controversy… They learn more science – meaning research – and it just happens that, without even noticing it, they learn also more law, economics, sociology, ethics, psychology, science policy and so on, since all those features are associated with the piece of science they have chosen to follow.

We adopted the same approach for introducing statistics to the students: instead of asking them to study it from the general concept to algorithmic, methodological, and technical specifications, we asked them to design a poster explaining via illustration how a very specific statistical method or algorithm works. Our working hypothesis was that the need to clearly represent the data analysis procedure would have pushed the students to both acquire the necessary statistical knowledge and to better understand the relationship with

[2] An example is the project "Seeing theory" available at https://seeing-theory.brown.edu/. The website uses visualizations and more generally visual strategies to explain statistical concepts such as variances, probability distributions, and Bayesian inferences.

communication design. To represent a statistical method, they should study it, get used to the related statistical concepts, and test it with small case studies.

The approach is not new in design education: *making* to understand the logics and processes behind the design of something is deeply rooted in the field. The concepts of “thinking through making” (Swanson 2020) and “learning by doing” (Özkar 2007) are proven didactical approaches in which the students are faced with a real-life design problem to solve. By dealing with the realization of the project, design students are forced to reflect on all the technical details, issues and choices that must be made. The difference in our case is that instead of using a project to directly reflect on topics related to design, we use it to promote reflection on the represented domain – in this case, statistics.

As a result, the statistics teacher overturned the lessons’ structure. The standard approach of teaching statistics to master students in the STEM[3] fields and economics is the one we can define as “downstream” or deductive: starting from the macro problem to be analysed, the approach first presents the available methods, then lists the possible variation for each one and provides technical details for specific implementations, and finally show the application of these methods to real problems. This approach is based on the idea that students, after graduation, should be able - in front of a specific applicative problem - to critically select the most suited data analysis methods among the learnt ones.

This is not, of course, the teaching aim of the presented module, whose goal is instead to make students aware of the existence and nature of what statistical learning is and make them able to fruitfully interact after graduation with experts in the field. For this reason, in the presented didactical experience, we swapped instead to an “upstream” or inductive approach: starting from a very specific implementation of a properly selected method, students had to approach it and understand its logics, moving from technical details toward more general concepts. An accurate selection of the methods proposed to the students was crucial for naturally and progressively enabling this drift from a specific method toward general concepts. We will come back to this issue in Section 2.3.

2.2 Visual explanations as think-through-doing mediators

To define the studio work proposed to the students, we also leveraged the concept of “research-led teaching” (Healey 2005), meaning the usage of current research topics to inform didactical experiences. More specifically, drawing on our current research, we asked students to design *visual explanations* of selected statistical methods.

With the term *visual explanations*, employed in the Information Design field, we refer to those visual strategies designed to explain a complex phenomenon that is not straightway

---

[3] STEM is an acronym used to group together the following academic disciplines: science, technology, engineering, and mathematics

graspable (Tufte et al. 1998): from mechanisms to living creatures' bodies. Such a term is directly related to the growing field of Visual Explanations of Artificial Intelligence (XAI) algorithms, which focuses on developing visual communication tactics to target audiences that are non-expert users. Although it has recently undergone a considerable expansion, the territory of Explainable Artificial Intelligence is still overseen by computer scientists designing visual explanations for expert users, from algorithms developers to domain experts like surgeons or policymakers (Correll 2019; Kolkman 2020).

We thought that this topic suits well the assignment and we proposed it to students: since part of the statistical methods have an algorithmic nature, we can test how XAI approaches can be employed to explain such methods.

We asked each group of students to deliver a 50*70 cm printed poster depicting the visual explanation of a chosen statistical method among the proposed ones. The faculty members wilfully set spatial and media constraint to reduce technical complexities and push students to focus on the contents.

2.3 Proposed topics

As mentioned above, we believe that a proper and thoughtful selection of the methods proposed to the students is needed to guarantee the success of this upstream approach. We defined two constraints in the selection. First, each method must be easily accessible and comprehensible by students without a solid mathematical education and mindset. Second, each method has also to stimulate discussion and thinking about general aspects of statistical learning far beyond the specific needs of the method.

The first constraint drove us in selecting algorithmic methods rather than model-based methods. Statistical learning collects a vast variety of approaches ranging from the "algorithmic" ones coming from the tradition of machine learning and computer science (which are based on the idea of learning by imitation and conveying the acquired knowledge in the form of an algorithm) to the "model-based" ones coming from the tradition of classical statistics and probability (which are based on the idea of identifying general rules and conveying the acquired knowledge in the form of a mathematical model). We opted for the former type of methods since: they have a much lower entry barrier for design students; they do not reinforce but rather weaken a possible math-phobic aptitude of design students toward quantitative disciplines, and they could reduce the risk of disorientation being the notion of coding and process familiar to them.

The second constraint instead drove us to select methods that (even though possibly developed for a particular problem) are general enough to be used to solve many different applicative issues, in various domain fields. This choice aims to stimulate students to abstract from a specific data analysis to the family of problems that the method can address.

Below the list of identified methods:

- **Bootstrap Estimation** is a computationally intensive method to quantify the uncertainty of statistical estimates in scenarios in which a theoretical study of this uncertainty is out of reach. By resampling data with repetition from the original source, this method generates artificial data sets meant to mimic the alternative ones that could be observed.
- **Classification Trees** are sequences of binary decision rules aiming at assigning data points with unknown class membership to the most likely class starting from their attributes. They are based on a training set made of data points whose relative class is known and they are driven by the idea of identifying sequential splitting rules (based on attribute values) creating sub-groups more and more homogeneous with respect to the class memberships. They are popular for their easy-readable open-box structure.
- **Control Charts** are anomaly detection tools that are used for the real-time monitoring of random stationary processes evolving over time. They are based on the identification of a minimal and maximal value of some attributes describing the status of the current process. These limits are built by looking at the distribution of these attributes during a training phase in which the process was known not to be affected by anomalies.
- **Hierarchical Clustering** refers to a family of algorithms aiming at identifying, within a multivariate data set, subgroups of homogeneous data points known as clusters. The algorithm is based on the idea of progressively merging into larger clusters (according to their relative distance) clusters generated at previous steps of the algorithm starting from an initial scenario in which each data point is a one-unit cluster.
- **Sentiment analysis.** The classical and more popular algorithm for performing a sentiment analysis of a corpus of texts is the Hopkin King Algorithm. Learning from some texts whose sentiment has been hand-coded by a human, it estimates the proportion of texts belonging to each sentiment category in the corpus by comparing the frequencies of words in the corpus with the ones observed in the sentiment-specific hand-coded sets.
- **Neural Networks** are a popular family of predictive algorithms able to accurately predict the value of a new unknown observation from past known observations by mimicking the human brain learning. Even though they are known to have very good predictive performances, they are sometimes criticized because of their black-box structure.
- **Page Rank Algorithm** is a famous algorithm used by search engines to rank web pages. It is a probabilistic algorithm based on the idea of a random web walker surfing the web by randomly clicking links from one page to other pages. The rank is determined by the number of visits to each web page if the walker could surf for an infinite amount of time.

- **Simplicial Depth Measure** is geometry-based method aiming at ranking multivariate data points from the most central ones the most peripherical ones relying on the relative positions of data points in the multivariate space. As an example, in a two-dimensional case the depth of a data point is computed as the numbers of triangles (with vertices in correspondence data points) able to enclose the data point.

Being all selected methods very popular in the data science community, detailed descriptions (for both experts and non-experts) can be easily found on the web together with public codes in different programming languages.

3. Didactical activities

In the entirety of the course, composed of three main modules, didactical activities were led by five teachers and five assistants. In the described didactical experience, presented as a course module, two teachers (one with a background in statistics, one in design) coordinated the activities and lectures, helped by four assistants. We will refer to “faculty” following the latter configuration of teachers and assistants. The module had a duration of four weeks, encompassing 49 hours of lessons and a similar number of students’ work. Students were divided into eight groups of 6/7 people, and each one chose a topic among the proposed ones (see section 2.3).

The eight methods were briefly illustrated during the second lesson to the entire class (about 10 minutes dedicated to each method). Half of the time was spent illustrating the tasks that the method could solve, emphasizing inputs and outputs, and giving an intuition about the main steps of the method; the rest of the time was invested in showing some popular applicative examples in which that method has been successfully used to solve a real problem. Finally, some time for Q&A was left after the presentation of each method. By purpose, mathematical formulas were neither shown nor mentioned during the presentation of the method and no bibliographic references were given to the students, encouraging them to identify proper sources and discuss them with the teachers.

For one month, students tackled the proposed statistical algorithms from initial explorations of the existing literature to the design of the final poster. A series of reviews was structured to assess intermediate results and to partially evaluate the correctness of the artifacts in progress. Reviews also functioned as moments for design feedback focusing on readability, visual hierarchy, and overall content presentation.

3.1 Didactical aims and evaluation criteria

As stated in the introduction, didactical aims are twofold: focusing on transferring correct knowledge about statistical methods and forming a germinal awareness of manipulating data and algorithms as communication design professionals. These two aims are reflected in the evaluation criteria to assess students’ work:

1. *Comprehension of the statistical method and its explanation (evaluated by the statistics teacher):* the correctness of the process they described — This criterion evaluates students' understanding of the algorithm and the technical language's appropriateness. This evaluation criterion answers the questions: *"Do the students correctly understand the process they describe? Are they able to explain it clearly and consistently?"*
2. *Visual translation of the process (evaluated by the design teacher)*: This criterion assesses students' ability to construct a communication design product that is cohesive, consistent, and functional to the explanation they are providing. Hierarchies, accuracy, illustrations, schemas, and visualizations are considered. This evaluation criterion answers the questions: *"Do students convey their understanding of the algorithm with visual clarity and intention? Is the layout of the poster easy to parse and decode by non-experts?*"

The structure of the evaluation criteria is, again, bifold: both disciplines of statistics and design converge in evaluating complementary aspects of the result of the teaching activity. By separating the two aspects of the evaluation, students have a clearer understanding on how they interacted with each other.

3.2 Structure of the module (and materials provided)

The module was articulated in a series of lectures and reviews that covered a variety of topics, both from statistics and the design discipline.

Lesson	Topic
Lesson 1	**Introduction to statistics and data science**. Main concepts of statistics are introduced to frame the exercise and to give basic tools to perform data analysis.
Lesson 2	**Presentation of algorithms.** Algorithms that students will explain in their posters are introduced with a short lecture that focuses on the overall aim and characteristics of each method.
Lessons 3	**Introduction to information visualization and visual explanations**. Students are provided with a background on information visualization and introduced to the history of visual explanations of complex artificial machines and how they can be considered a communication artefact.
Lessons 4,5,6	**Group reviews.**
Lesson 7	**Delivery, evaluation, and peer assessment.**

Table 1: The module was organized in a total of seven lessons: four lessons were devoted to reviews, two focused on statistics and one focused on visual explanations from a design point of view.

3.3 Delivery

The delivery of the poster was organized as an exhibition where faculty members had an appropriate amount of time to read and analyse the poster. This part of the delivery was crucial in assessing if the content of the poster was clear even without an appropriate introduction, as some members of the faculty did not follow the entire design process. After the initial moment of reading, students presented the design process. In this instance, the faculty had also the opportunity to ask questions.

4. Course results discussion and evaluation

After being introduced to the statistical methods identified and described in paragraph 2.3., students were asked to express a preference in dealing with them as contents of their posters. Considering the order of such preferences, we assigned the algorithms as follows:

- Group 1 – Simplicial Depth Measure (Figure 1.1)
- Group 2 – Neural Networks (Figure 1.2)
- Group 3 – Sentiment Analysis (Figure 1.3)
- Group 4 – Page Rank Algorithm (Figure 1.4)
- Group 5 – Hierarchical Clustering (Figure 2.1)
- Group 6 – Control Charts (Figure 2.2)
- Group 7 – Bootstrap Estimation (Figure 2.3)
- Group 8 – Classification Trees (Figure 2.4)

4.1 Faculty evaluation of the module

Acknowledging a lack of formal training in statistical learning among the students, we had not excluded the possibility of a significant failure of this teaching experiment. The outcome was instead over our best expectations: five groups quickly tackled the challenge in an optimal way and with excellent results, and two groups found their way with extra guidance help from the teachers. Only one group had major difficulties. We observed a strong correlation between the depth of the comprehension of the statistical method (as evaluated by the statistics teacher) and the quality of its visual translation.

The perceived average interest of the students with respect to this part of the course was much larger than the one shown in the previous editions and characterized by a higher level of interaction with teachers both during the lecture and by email. The statistics teacher acknowledged that the teacher-student interaction was higher characterized by questions and answers with a complexity often comparable with the one typically observed in classical statistics courses for STEM and economics students (even though mostly related to the specific algorithms) proving as false the misconception (diffused among teachers and students themselves) that design students do not have the background skill needed for an in-depth understanding of statical methods. This was also proven by the different aptitude of students while actively interacting with teachers being generally more self-confident and less apologetic.

A last but very relevant unexpected outcome of this upstream approach was a strong stimulus towards interdisciplinary interactions. While we were not expecting design students to read scientific papers (indeed, only one group used this strategy), we thought that students would have used web tutorials. Even though web tutorials (both in the form of web pages and videos) were an essential source of knowledge, two groups asked for help and suggestions from friends (most of them being master students in engineering at Politecnico di Milano).

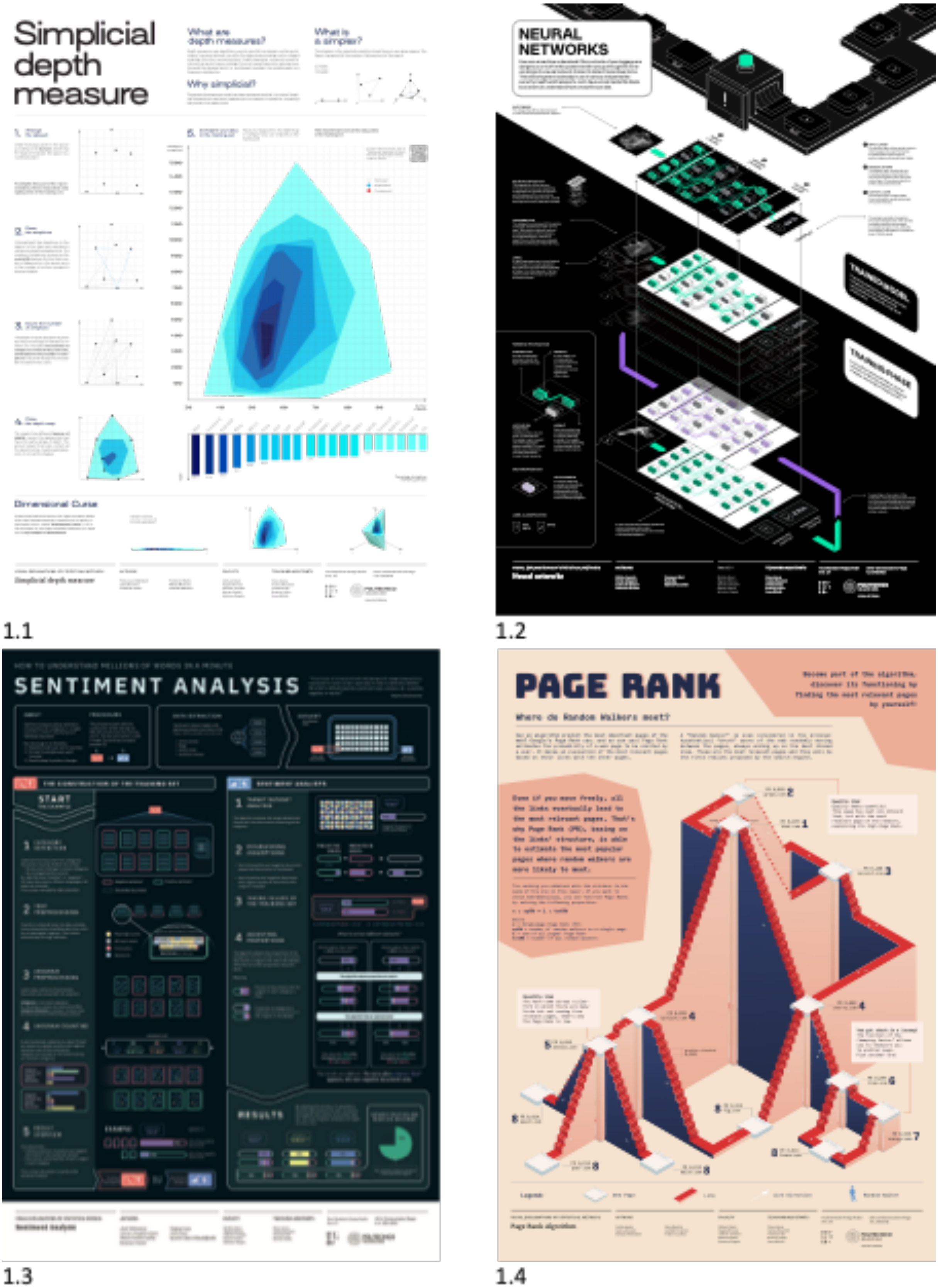
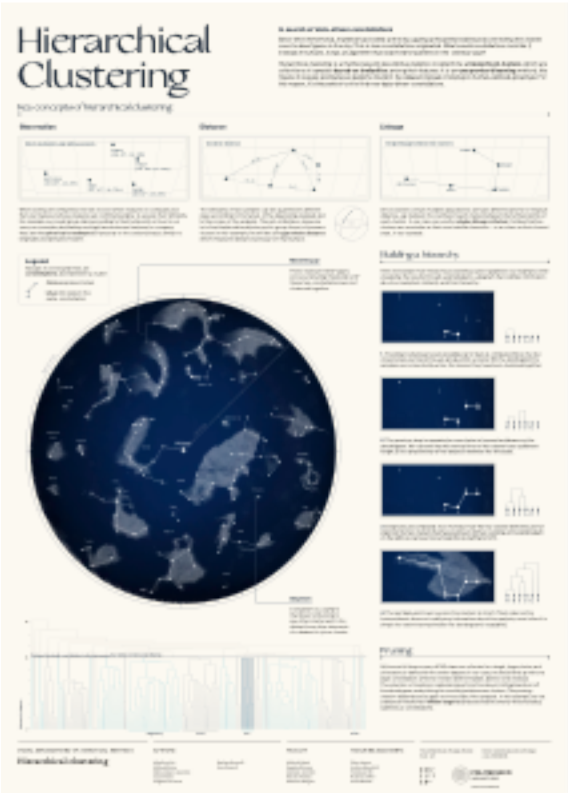
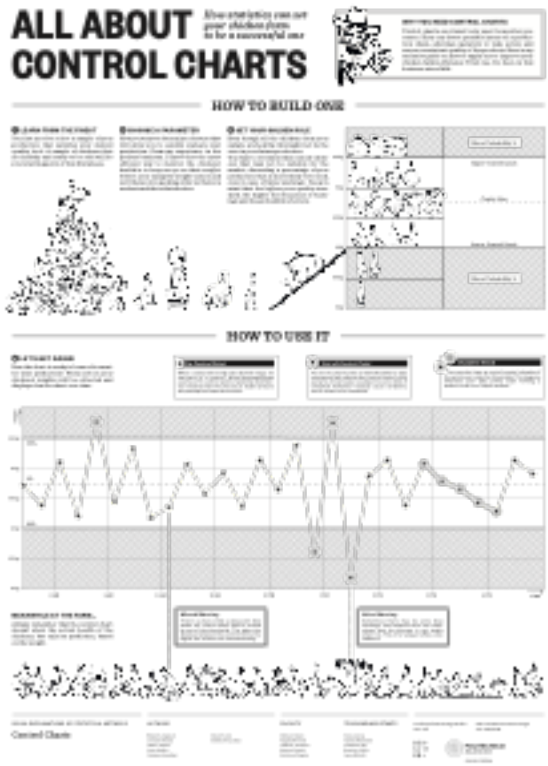


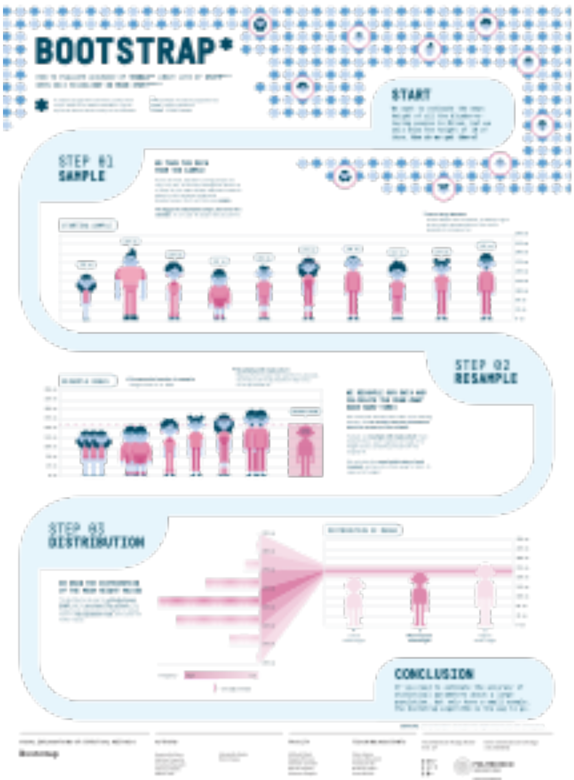
Figure 1 Final posters discussed during the delivery day. The order is sequential and follows the groups' number (groups 1-4). Authors list in the acknowledgements. Higher resolution here.



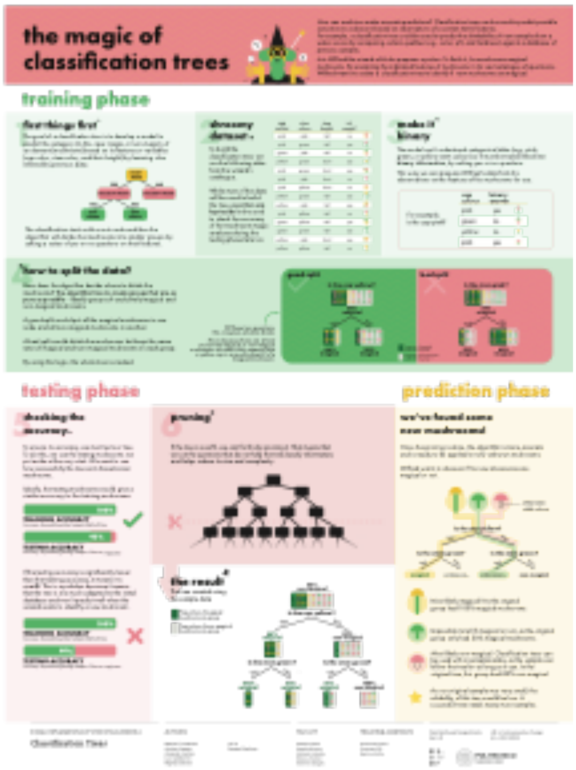
2.1



2.2



2.3



2.4

Figure 2 Final posters discussed during the delivery day. The order is sequential and follows the groups' number (groups 5-8). Authors list in the acknowledgements. Higher resolution here.

4.2 Students’ evaluation of the module

We used semi-structured interviews to collect insights about the students’ perception of the module and to evaluate its solidity. Interviews indeed offer the opportunity to better follow the singularities of students’ design processes through the employment of follow-up questions. Furthermore, it allowed us to understand the emergence of students’ reflections.

Participants were recruited from the classroom through an open call involving one member per group. Interviews lasted approximately 25 minutes, they were held in a meeting room equipped with a widescreen suitable for projecting posters, and an ambient microphone for recording the conversations. The recordings were transcribed and then analysed using thematic analysis (Braun and Clarke 2006).

Following the thematic analysis framework, results were organized around three focus points. First, the design process followed by students, to understand how they coped with studying statistics. Second, the eventual usage of examples or metaphors, given the recognized relevance of explanation by example in the field (Ehsan et al. 2021; Miller 2019). Third, the students’ considerations about the impact and role of the exercise in their education.

Each participant is indicated with the letter “P” and a sequential number corresponding to the working group to anonymously report results. Among a total number of eight interviews, six were conducted in Italian, while the remaining two in English.

Figure 3 Three participants dealing with reordering activity. Starting from the left, a participant positions actions as a sequence. In the middle, a participant positions actions crafting a "Gantt" visualization to point out the simultaneously of some actions. On the right, a participant is positioning actions in the space according to the way in which they divided into subgroups of work.

Learning statistics by doing. As first activity, students were asked to describe the design process by reordering four paper-cards corresponding to different actions: reference search, algorithm study, sketches, and visual explanation. This activity was helpful to understand how students collected information on their statistical method, and how they merged it with the design process. Students composed the process in different ways: (a) as a sequence, (b) as a Gantt to show simultaneous actions, (c) or as groups. The activity suggests that there is not a singular and shared process and that every group acted differently. (Figure 3).

Seven out of eight participants stated that the process began with an in-depth study of the analysed statistical method, either as a stand-alone activity or concurrently with others. In four cases [P1, P3, P4, P8] students decided to devote individual moments to the study of the method, and then discuss with the rest of the group doubts and discrepancies. Moreover, three students [P1, P3, P6] acknowledged that weekly reviews with the faculty, as well as conversations with friends into STEM [P1, P6] were useful to assess their understanding of the analysed method. Moreover, the understanding of the statistical method has often derived from the manipulation and experimentation with data and tools. Hence, while one group acted by translating the contents of scientific papers into a form suitable for a general audience [P3], the other students used online simulators [P4, P5, P7] and successfully learned the basics of *R*[4] [P1, P5, P7], despite its learning curve is often recognized as steep. (See Table 2)

	Discussion with experts		**Simulation**				**External sources**			**Other**
	Peer studying stats	Revisions with professors	Online simulators	Real-time simulation	Discoursive simulation	Simulation with code	Scientific articles	Web articles	Youtube videos	Background knowledge
P1										
P2										
P3										
P4										
P5										
P6										
P7										
P8										

Table 2: Table summarizing how participants coped with statistics. Grey cells correspond to the strategies adopted by each group.

Contextualizing the explanation. In seven out of eight cases students based their works on top of examples of the application of the assigned algorithm. Students used examples as a strategy to help the audience to understand statistical methods. Reasoning by examples helped students to share learned notions within the group and with the faculty members [P3, P8]. In two cases [P2, P5], one of which is reported as "*a safe solution*" [P2], they drew examples from the scientific literature. In the others, students relied on examples derived from their experiences or real-world scenarios [P1, P4, P6, P7]; only in one case, the example was derived from students' knowledge [P8]. While in some cases students carried the same example from the first to the last review [P7], others iteratively considered different options [P5]. In one case they reported a belated inspiration after watching a documentary [P6].

[4] R is a widely adopted free software for statistical computing. Know more at: https://www.r-project.org/

Educational take-aways. Interviewees were prompted to share general considerations about the academic and didactic contribution of the module. Specifically, we report information concerning how the exercise made students reflect on the algorithms' role in a society increasingly relying on them. Three participants appreciated how the exercise pushed them to think about the need for demystifying algorithms in society [P2], to produce an artifact aimed at opening the so-called black boxes [P7] and tackle a complex process by dissecting and representing it [P4].

Participants also reflected about the role of humans within the algorithmic process. P1, P2 and P5 stated that the exercise made them ponder on the relationship between data input and output, and on the huge influence that training dataset has on the results. P1 and P3 weighed on the important role that humans play when making choices and interpreting the results. Generally, students did not mention the risks of embedding human biases in the generation of samples or in the supervision of algorithms' learning processes. The faculty deem this aspect important since it leads to a deeper reflection on the idea that technologies are not neutral nor objective.

In general, the interviewed students recognized the importance of the module not only within the course but also within the entire communication design curriculum supplied by the University. Indeed, the module helped them develop sensitivity towards data and be prepared in dealing with future challenges [P6]. In addition, they also stated that the module completes knowledge gathered through other courses of the curriculum investigating the relationship between media, algorithms, and society from sociological perspectives. [P1, P4, P5].

Moreover, it allows the faculty to reflect on the interaction of the two evaluation criteria: some results leaned more heavily on the statistical aspect, while other indulged more in visual metaphors to explain the algorithm. These two approaches are not mutually exclusive; however, they are a relevant indicator on how the students group tackled the topic assigned to them.

5. Conclusions and further developments

The presented didactical experience was aimed at introducing knowledge on the role that statistical learning plays in communication design to design students. Through this first experience, we can draw some lessons as well identify possible improvements.

The quality of the results (section 4.1) showed that it was a successful way to provide statistical knowledge to communication design students: they proved to have understood the assigned statistical method, and the ability to deal with related statistical concepts and lexicon. It has also become clear that the more the assigned topics were defined also in their technical details (e.g., "Simplicial Depth Measure", "PageRank") the more students were able to understand it and learn the surrounding concepts. The more the topic was generic (e.g., "Sentiment Analysis", "Control Charts") the more the students found difficult to frame it. A side consideration is the fact that it is fundamental to leverage the ability of communication design students to think visually: an example is probability distribution, that caused several misconceptions until we proposed a visual explanation of them.

The students involved in qualitative interviews (section 4.2) demonstrated understanding of the reason why they attended to a statistical module, and that the active engaging with it allowed them to start a broader reflection. The exercise could benefit from a framing that is more oriented towards the understanding of the biases embedded in methods and algorithms, and how they can affect results; a similar approach could help students in better realise that technologies are not neutral nor objective. Additionally, to the qualitative interviews with selected students it could be associated a quantitative questionnaire to the whole course, achieving a more comprehensive understanding of the method strengths and weaknesses. Furthermore, a by-product of the adoption of a research-led teaching approach is that the feeling of working on boundaries topics motivated them in deepening the state of the art and in finding efficient visual solutions.

From the faculty side, this positive experience allowed to reflect on limits and possible improvements. One evident issue is the fact that the complementary presence of teachers from both the disciplines is fundamental, making this experience more difficult to be reproduced in other educational contexts.

Given the large size of the class, a potential issue is the temporal sustainability of this approach: algorithms suited for this course are very specific and so also very limited. This might force as in the future to repropose the same algorithm, and the challenge will be to avoid the possibility of biasing the creative process of future students towards previous visualization.

Acknowledgements

We would like to express our gratitude to the master's students in Communication Design who attended the 2022 course "Final Synthesis Design Studio" at Politecnico di Milano for their commitment to the concept and design of the eight posters. Specifically we thank: Group 1 – Simplicial Depth Measure: Battistoni F., Boschis C., Inzani F., Meani F., Mertens M., Robuschi O.; Group 2 – Neural Networks: Casarotto M., Chiappini D., De Simone A., Ghebaur E., Gheli F., Ma H., Riccardelli R.; Group 3 – Sentiment Analysis: Alzhanova A., Astaghforellahi S., Coelho De Oliveira M. C., Foresti B., Luan Y., Saad N. S., Schwaighofer S. A.; Group 4 – Page Rank Algorithm: Agosta L., Antonini C. A., Bombardieri M., Busletta E., De Amicis C., Lucifora F.; Group 5 – Hierarchical Clustering: Bocchio A., Bruno M., Casolino M. C., Draisci L., Leccisotti V., Roncalli B., Zanardi S.; Group 6 – Control Charts: Arappali B., Bernini L., Caputo C., Casano I., Hamdani Y., Perico M., Perucchini D.; Group 7 – Bootstrap Estimation: Dell'Orto D., Francella M., Huang S., Husoschi O. D., Melillo M., Pini M. M., Quets A.; Group 8 – Classification Trees: Fernández De La Rosa M., Gamberini L. L., Llamas R. A., Martínez T. R., Salviato R., Rodrigues Cestaro A., Yu Q.

6. References

Braun, Virginia, and Victoria Clarke. 2006. 'Using Thematic Analysis in Psychology'. *Qualitative Research in Psychology* 3(2):77–101. doi: 10.1191/1478088706qp063oa.

Correll, Michael. 2019. 'Ethical Dimensions of Visualization Research'. Pp. 1–13 in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM.

Diehm, Cade. 2018. 'On Weaponised Design'. *Our Data Our Selves*. Retrieved (https://ourdataourselves.tacticaltech.org/posts/30-on-weaponised-design/).

Ehsan, Upol, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. 'Expanding Explainability: Towards Social Transparency in AI Systems'. Pp. 1–19 in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM.

Healey, Mick. 2005. 'Linking Research and Teaching to Benefit Student Learning'. *Journal of Geography in Higher Education* 29(2):183–201. doi: 10.1080/03098260500130387.

Kerren, Andreas, John T. Stasko, and Jason Dykes. 2008. 'Teaching Information Visualization'. Pp. 65–91 in *Information Visualization*. Vol. 4950, *Lecture Notes in Computer Science*, edited by A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.

Kolkman, Daan. 2020. 'The (in)Credibility of Algorithmic Models to Non-Experts'. *Information, Communication & Society* 1–17. doi: 10.1080/1369118X.2020.1761860.

Lo, Leo Yu-Ho, Yao Ming, and Huamin Qu. 2019. 'Learning Vis Tools: Teaching Data Visualization Tutorials'. Pp. 11–15 in *2019 IEEE Visualization Conference (VIS)*. Vancouver, BC, Canada: IEEE.

Mauri, Michele. 2020. 'INTRODUCING INFORMATION VISUALIZATION TO DESIGN STUDENTS'. Pp. 4442–48 in. Online Conference.

Mauri, Michele, Gabriele Colombo, M. D. L. A. Briones Rojas, and Paolo Ciuccarelli. 2019. 'Teaching the Critical Role of Designers in the Data Society: The DensityDesign Approach'. 13.

Miller, Tim. 2019. 'Explanation in Artificial Intelligence: Insights from the Social Sciences'. *Artificial Intelligence* 267:1–38. doi: 10.1016/j.artint.2018.07.007.

Özkar, Mine. 2007. 'Learning by Doing in the Age of Design Computation'. Pp. 99–112 in *Computer-Aided Architectural Design Futures (CAADFutures) 2007*, edited by A. Dong, A. V. Moere, and J. S. Gero. Dordrecht: Springer Netherlands.

Swanson, Gunnar. 2020. 'Educating the Designer of 2025'. *She Ji: The Journal of Design, Economics, and Innovation* 6(1):101–5. doi: 10.1016/j.sheji.2020.01.001.

Tufte, Edward R., Susan R. McKay, Wolfgang Christian, and James R. Matey. 1998. 'Visual Explanations: Images and Quantities, Evidence and Narrative'. *Computers in Physics* 12(2):146. doi: 10.1063/1.168637.

Valsecchi, Francesca, Paolo Ciuccarelli, Donato Ricci, and Giorgio Caviglia. 2010. 'The DensityDesign Lab: Communication Design Experiments among Complexity and Sustainability'. in Vol. Proceedings Cumulus Conference 2010 "Young Creators for Better City and Better Life".

Venturini, Tommaso. 2010. 'Diving in Magma: How to Explore Controversies with Actor-Network Theory'. *Public Understanding of Science* 19(3):258–73. doi: 10.1177/0963662509102694.

About the Authors:

Michele Mauri is the data visualization professor of the module and the scientific director of the DensityDesign Lab.

Simone Vantini is the statistics professor of the module, currently teaching at the Mathematical Engineering Department of Politecnico di Milano.

Beatrice Gobbo is a Ph.D. student at DensityDesign and teaching assistant of the module. Her research is positioned at the intersection between information design and Explainable Artificial Intelligence.

Tommaso Elli is a post-doc researcher at DensityDesign and teaching assistant of the module. He is interested in the relationship between information design and digital humanities.

Elena Aversa is a PhD student in Design at DensityDesign and a teaching assistant of the module. She is interested in investigating the relationship between data visualization, information disorders, social media, and the ethics of information design.

Andrea Benedetti is a Ph.D. student at DensityDesign and teaching assistant of the module. His research focuses on the role of data in society and how design can be an activator of reflection in this relationship.

Maria De Los Ángeles Briones Rojas is an adjunct professor of the course.

Gabriele Colombo is an adjunct professor of the course.