Project

Advanced Data Visualization (MECD) 2023/24

Introduction

The main goal of the information visualization project is the development and application of a set of skills that allow the construction of interactive visualizations capable of solving analytical problems and those related to the communication of information. The work is organised in a manner simulating the conditions that occur in a real-world project. This includes activities such as communication with the client (teacher) and obtaining feedback, working with data, identification of the problem(s) and possible solutions, sketching and prototyping, implementation of interactive visualization, and validation. It is also intended that the students demonstrate skills in analysis and synthesis, organisation and planning, critical reasoning, application of knowledge in practice, investigation, autonomous learning, and group work.

The project must integrate the following characteristics and requirements:
1. All the code should be produced by the students;
2. Students are free to use any visualization tools and frameworks only for the Exploratory Data Analysis phase;
3. Students should use Plotly/Dash framework and Python to implement the main visualization;
4. Students are free to use any additional frameworks to support the main code;
5. Students are free to choose the datasets, but they need to be approved by the teachers;
6. The visualization should be an analytical application or a communicative visualization, depending on the path the students choose;
7. The visualization should implement, at least, one advanced visualization technique (e.g., bar graph, line chart, pie charts are not advanced techniques);
8. The visualization should implement, at least, one advanced interaction technique (e.g., linking and brushing);
9. Include other interactive elements, such as hover, range selection, or any other covered during the semester;
10. The visualization should follow good practices of design (i.e., contain all the necessary graphical elements, such as legends, title, etc.).

Groups

The project should be developed in groups of **2 elements**. The groups should be defined by filling out this form: 🟩 VAD — Grupos

Milestones

The project is divided into several milestones with respective deadlines:

- **Group formation** (within the first week): We expect you to form a team as soon as possible. No formal submission is needed.
- **Project proposal** (deadline: **3 March**): For this milestone, you should have completed the following topics of your progress report (see Project section): **Authors, Introduction, Related Work, Design Requirements, and Data.** It is not expected that you have your final data acquired, but a significant part should be ready. This will be the first part of your progress report, and it is expected that you submit it in PDF format to the respective slot at the *Inforestudante* platform.
- **Midterm review and feedback** (deadline: **31 March**): For this milestone, you should have completed the Data, Exploratory Data Analysis, and Design steps of your project. It is not mandatory to entirely complete the design phase, but it is highly recommended. It is your opportunity to show your progress so far and get feedback from the teachers. You will submit your progress report to the *Inforestudante*. The feedback session will be held during the TP and T class the following week.
- **Final submission** (deadline: **19 May**): For the final submission you should have completed all the remaining steps of the progress report and must hand in the following items (see details in Project section): i) progress report, ii) video that showcases your project, and iii) the code.
- **Project presentation and defenses** (last week of the semester): There will be a final defense of your projects with the duration of 20 minutes. You should start with a formal presentation of your project, after which the defense continues with a Q&A session. Defences are mandatory. The projects will be evaluated according to predefined grading criteria (see respective Section). The defences require registration to a time slot, which will be opened in the *Inforestudante* closer to the defence date.

Evaluation

The grading for the project is equal to 60% of the final grade, which is equivalent to **12 out of 20. Important notice!** It is not possible to improve the grade of the project.
Another important notice! Submissions made after the deadline will not be accepted.

Grading criteria

The following elements will be evaluated:
- Data and analysis:
  - Data preparation (Does your final dataset include only the data items needed for visualization? Was the data stored in the appropriate format and structure? Was the data cleaned and filtered according to the good practices of data preparation?);
  - Exploratory Data Analysis (Did the EDA support the definition of analytical questions and the goals of the visualization? Did EDA help in the preparation of the dataset?).
- Solution:
  - Application of visualization principles and practices of good design (Was it designed following the principles studied in the course?);
  - Visual interaction (Does your visualization employs advanced interaction methods?);
  - Effectiveness and appropriateness of the solution (Is your visualization effective in answering the defined analytical questions? Is your visualization effective in communicating data/insights? Does your solution employ appropriate visualization methods and encodings?).
- Implementation: Quality of the implementation (Is your visualization robust? Is your application optimised to run in real-time? Does your code follow conventions and overall good programming practices?)
- Presentation and defence: Clear, engaging, and effective presentation. Did you answer all the questions during the defence? The score for this component is individual.
- Written report: Are your project and the progress well documented in the report?

Minimum grade

In order to get approval in the Advanced Data Visualization course, it is necessary to obtain a classification equal to, at least, 40% of the final grade of the project.

Project

The project consists of creating a visualization for data analysis and/or communication. The students are faced with real-world datasets with the goal of identifying analytical and/or communication problems and developing a visualization that should solve these problems. The visualization should be effective, should contain multiple views, and should incorporate appropriate visualization techniques for data analysis and/or communication, depending on the path each student chooses. The visualization should follow good practices of visual design, as well as the principles and theories studied in this course.

Progress report (the article)

Keeping track of your progress is an important part of your project. Remember, it is equally important how you get your final results! Therefore, each project should be documented in the progress report, providing details about each part of your project, including any additional information that you find relevant (e.g., alternative designs you tried, the problems that you faced, or the insights that you learned). It is also important to underline that your report book is not a logbook or diary. You should present your work in an informative and concise way, without unneeded details that do not contribute to the understanding of your project. Instead, include many figures in your progress report (e.g., different designs you tried, photos of your sketches, or sketches you made using different visualization tools, etc.). Ultimately, the goal is to convey the progress and rationale of your project in a clear and informative way.

The progress report should take the form of an **article** with a typical structure, which should include the following items:
- **Authors**: list the members of the team, including their email addresses.
- **Introduction**: Provide an overview of your project, its goals, and motivation. Try to define the problem being addressed, the target users, and what they would want to know about the data.
- **Related work**: Include and briefly describe any visualization you used as a reference or inspiration for your project (e.g., a paper, the visualization you found on the internet, examples you saw in class, etc).
- **Design requirements**: Provide a list of design requirements and the tasks you identified for the end users of your applications.
- **Data**: Describe the chosen data, its characteristics, the data type, the structure, and data sources, among others.
- **Exploratory Data Analysis (EDA)**: Provide a detailed description of the exploratory analysis of data, including the tools you used, the visualization methods tried, the hypothesis tested, the insights gained, and how these insights guided your design decisions, among other relevant questions.
- **Design**: Describe the rationale behind your designs, supported by the mockups/sketches of tested solutions. It is important to use conceptual and theoretical tools that you learned in the course to justify your choices, i.e., what design principles, perceptual theories, or colour theories informed your decisions.
- **Implementation**: Provide the key functionalities of your application and the interaction mechanisms employed. Describe the design elements (e.g., views, GUI elements, layout) and interaction components (e.g., sliders, buttons, dropdown components). You can resort to diagrams to explain the architectural aspects of your applications.
- **Evaluation**: Detail the method you used to evaluate your visualizations (e.g., user testing, qualitative questionnaire, critical discussion, among others). Try to identify what works well in your visualization and what does not, and how you would improve it. You can also provide the results of user testing.
- **Critical Reflection**: This section should include a critical reflection on your visualization. Provide an overall assessment of your final visualization. Was your visualization effective

in getting insights about your data? Was it effective in answering the defined questions? What could you learn about your data that wouldn't be possible without it? How would you improve your solutions?

The process report should not exceed 8 pages in length (plus references). We suggest including 3-4 pages of text and 5-4 pages of figures. Figures should be placed near their reference in the text.

Data

Students are free to use any data and datasource. As an example of the data (e.g., size, complexity) consider Airbnb listings around the world, an independent project that collects publicly available Airbnb data (http://insideairbnb.com/index.html). The respective datasets are available at http://insideairbnb.com/get-the-data.html. Students are allowed to use any additional datasources to enrich the datasets at hand, for instance, geographic information about places http://centraldedados.pt/codigos_postais/. Try to be creative and employ all the skills you have earned in the other courses.

Regarding the data preparation and Exploration Data Analysis, students are free to choose any API and programming language for the former, and any visualization tool and framework for the latter. The teachers may restrain to support only the tools they use. Consider discussing your choice before starting the project.

Recall that the Inforestudante platform is limited to 200MB of space, so in cases of large datasets **do not use** *Inforestudante*, instead use external file-sharing services (e.g. Dropbox, Wetransfer). Include the link in your report.

Demo video

Create a demo video to showcase your visualization. Try to be concise and clear about what you are presenting in your video. Employ good practices of narration. Focus the majority of your video on what is important in your project, rather than unneeded technical details. Remember, your project should be understood by the general public. The duration of your video should be at most 2 minutes in length.

To submit your video do not use Inforestudante, instead use external file-sharing services (e.g. Dropbox, Wetransfer) or video streaming platforms (e.g., Youtube, Vimeo). Include the link in your report.

Code

You should use Plotly/Dash and Python to implement the core functionalities of the project. Students are free to use any additional APIs and frameworks as supportive tools (e.g., to

compute graph layouts), but teachers are reserved to answer the questions regarding Plotly/Dash.

You should also submit all the code (for visualization and data preparation) to the *Inforestudante* platform.

Project activities (suggestion)

As a suggestion, this section lists several core activities that students may perform to accomplish the project. This includes the following:
- Definition of target users and a problem that the visualization should solve;
- Developing of analytical questions for data analysis and/or communication;
- Definition of data and tasks, as well as the specification of the requirements;
- Acquisition of the necessary data and its preparation;
- Identification of an appropriate visualization method for each of the views of the visualization application (e.g., a time-series representation, a map-based visualization);
- Implementation of the chosen views;
- Refinement and definition of visual encodings;
- Implementation of the data transformation functions;
- Refinement of the visualizations;
- Addition of supportive graphical elements, such as titles, captions, legends, etc.
- Implementation of the basic interaction, such as element selection/hovering, range selection, etc.;
- Implementation of advanced interaction methods (e.g., linking and brushing interaction method);
- Validation of the employed models using the chosen method (e.g., critical discussion, questionnaire, user study).

Dissemination

The projects developed in the context of this course are disseminated via FeedNPlay.

1. Projects developed in the context of this course may be submitted to be displayed in FeedNPlay, a multimedia device with a total width of more than 6 meters, which will be installed on the 1st Floor of the Department of Informatics Engineering. Students should follow the instructions available at the FeedNPlay website (https://feednplay.dei.uc.pt).

Plagiarism and fraud policy

According to the Article 28.º of the disciplinar reglament for the UC students (https://www.uc.pt/regulamentos/ga/vigentes/Regulamento_Disciplinar_Estudantes_UC)

Project

"2 — A fraude cometida em sede de avaliação de uma unidade curricular implica a anulação imediata dessa avaliação e leva à reprovação liminar do estudante na inscrição na unidade curricular em causa, devendo ser registada na plataforma informática da UC e averbada no processo individual do estudante."