BDMA - Decision Modeling

Jose Antonio Lorencio Abril

Fall 2023

Professor: Petra Isenberg

Student e-mail: jose-antonio.lorencio-abril@student-cs.fr

This is a summary of the course *Visual Analytics* taught at the Université Paris Saclay - CentraleSupélec by Professor Petra Isenberg in the academic year 23/24. Most of the content of this document is adapted from the course notes by Isenberg, [1], so I won't be citing it all the time. Other references will be provided when used.

Contents

1 Introduction

We all know about the increasing amount of data collected and handled by companies and organizations, but it is also important to understand that data is not the same as information: it is needed a process of analysis and understanding to derive information from data. When we have a question that we want to answer with our data, we **query** the data seeking for the pieces of data that might be relevant for our questions. On the other hand, when we aren't sure what we're looking for, we **explore** the data, looking for patterns that can give us insights and ideas we didn't thought about before.

Moreover, purely relying on automated analyses is not always effective due to potential unexpected results, usually because of edge cases and situations that we did not think about at the beginning or which did not even exist by then, and because data can be incomplete, inconsistent or deceptive. Therefore, human judgement and intervention is often needed, to provide background information, flexible analysis, modifiable to unintended directions and creativity. **Visual analytics** is then a field that provides different tools to have a human in the loop in analysis tasks.

In this course, we want to build a strong critical thinking with data, relying on visualizations that can help us to better understand data. We will delve into the topics of Data Collection, Data Cleaning, Exploratory Analysis and Visualization.

1.1 What is Visual Analytics?

1.1.1 What is Data Analysis

Traditionally, there is the vision that data analysis consists in applying statistics to analyze data collected from the real world, but a more accurate vision would be to define data analysis as the task of thinking carefully about evidence, represented by data. Nowadays, this view is more spread, and data analysis is now covering a wide range of activities and skills:

- Problem definition
- Disassembling problems and data into analyzable pieces
- Data evaluation & Conclusion making
- Decision recommendation

1.1.2 Visual Analytics

Visual Analytics is the science of analytical reasoning facilitated by interactive visual interfaces. It combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets.

The greatest challenge of visual analytics is to enable deep insights, allowing analysts to examine massive, multi-dimensional, multi-source and time-varying information, to make the right decisions in a time-critical manner. With this in mind, the method is to combine automated analysis with human intervention, and representing data visually to allow for interaction, insight generation, conclusions making and enabling for better decision making.

The field can be understood as a whole, but it can also be divided into:

- **Information Visualization**: visualizations that enable to transmit information to the general public in a clear way.
- **Scientific Visualization**: visualizations that show scientific work and discoveries with precision.
- **Infographics**: this represents a visual summary of a topic that aims at providing an easy-to-understand overview on a topic.

As mentioned before, there are basically two approaches towards data analysis:

- **Confirmatory Analysis**: starts with a hypothesis about the data and tries to confirm its validity. This kind of analysis focuses more on fully automated analysis methods.
- **Exploratory Analysis**: when there is no or little a-priori information about the data and we are not sure about which patterns and information can be present in the data, we can explore it to create hypotheses that will need to be confirmed later. It is in this area where visual analytics is most widely used.

We can also understand visual analytics as a process, involving the following steps:

1. Information (data) gathering
2. Data preprocessing
3. Knowledge representation
4. Interaction
5. Decision making

Therefore, the requirements for an interesting and efficient visualization analytics approach are the development and understaing of data transformations and analysis algorithms, analytical reasoning techniques, visual representations and interactions, and techniques for production, presentation and dissemination.

1.1.3 History of Visual Analytics

In the early 2000s, there was an outgrowth of the Scientific & Information Visualization community, which started with US National Visualization and Analytics Center (NVAC) at PNNL in 2004. This center developed the first research and development agenda "Illuminating the Path" sponsored initially by DHS (US Department of Homeland Security). At first, the goals of this center and of the field were analyzing terrorist threats, safeguarding boarders and ports, and preparing for and responding to emergencies.

The field has evolved since then to serve for larger research goals, specially since the first edition of the VAST symposium, a conference in visual analytics, science and technology, as part of the IEEE Visualizations Conference, in 2006; in addition to fundation of the VisMaster in 2008 by the EU. This represented a coordination action to join European academic and industrial R&D, with a focus in broader aplicability in different scientific fields (physics, astronomy, weather,...) rather than homeland security. From this point on, many centers in Europe have been created to research in this field.

1.1.4 Challenges of Visual Analytics

1. Human reasoning & decision making
   (a) Understanding and supporting how humans reason about data.

References

[1] Petra Isenberg. Visual analytics. Lecture Notes.

1. Support convergent and divergent thinking.
    (a) Create interfaces that are meaningful, clear, effective, and efficiente.
2. Adoption
    (a) Communicate benefits of developed tools to drive frequent use.
    (b) Make tools accepted by users.
3. Evaluation
    (a) Develop methods to compare novel tools to existing ones.
    (b) Asses how good a tool is, which is a very difficult task for measures other than time and error.
4. Problem interdependence
    (a) Analysis in the real world often does not consist of isolated problems or questions. Problems are usually correlated and how one is solved influences how one should approach another.
    (b) Synthesis of analyses is needed.
5. Integration of analysis methods
    (a) It is simple to do many isolated analyses, but it is hard to integrate them well into one tool or interface for human analysis.
6. **Scalability**
    (a) **Information scalability**: capability to extract relevant information from massive and possibly dynamically changing data streams. There are different methods to achieve this kind of scalability, among which we can find abstract data sets, filter & reduce data, or multi-resolution representation of data.
    (b) **Display scalability**: this referes to the capability of visualizations and tools to adapt to different types of displays (computer monitor, smartphone, smartwatch,...).
    (c) **Visual scalability**: refers to the capability of visualizations to effectively display massive datasets in terms of number of data items or data dimensions. It depends on the quality of the layout, the interaction techniques and the perceptual capabilities.
    (d) **Human scalability**: human skills don't scale, but the amount of people involved in the analysis task can, so we must seek to design techniques to scale from a single to multiple users.
    (e) **Software scalability**: software systems and algorithms must scale to larger and different data.
    (f) **Others**:
        i. Privacy and security in multi-user settings.
        ii. Collaboration across languages and borders.