Data visualization : Origins and Techniques*

Paul Peyssard[1]

[1]Côte d'azur University, Campus SophiaTech Lucioles
MSc 1 Data Science and Artificial Intelligence
`paul.peyssard@etu.univ-cotedazur.fr`
https://univ-cotedazur.eu/msc/msc-data-science-and-artificial-intelligence

Abstract. In this paper, we will explain the principle of data visualization, its strengths and weaknesses. There is a large amount of data visualization techniques, some are useful for a lot of things, some are specialized in certain areas and some are obsolete. It is therefore necessary to think about how we are going to use these techniques in order to have the best possible rendering. The most important thing is to create understandable data and to keep our values not to be brought to mislead people by manipulating them.

Keywords: Data visualization · Big data · Visualization technique· Personal data

1 What is Data visualisation ?

1.1 Introduction

For a long time there has been the need of visualize, understand and manipulate all the data we are faced with. Companies, individuals or new technologies generate more and more data every day. The amount of data is therefore increasing day by day and has become an important part of our lives, their use and storage becomes complicated. In our time, data has become a real field that can improve research, help understand other areas in which we are not necessarily experts and also be a source of revenue for companies to exploit.

1.2 Origins and uses of data visualization

Information visualization rising in 1990s was born in scientific visualization, which emphasizes to express the abstract information using computer graphics and images. The term of data visualization covering the fields of scientific visualization and information visualization was born in 2000s.[7] Data visualization is born from this need of displaying massive amounts in a way that is understandable for everyone. It can be useful in many fields as :

– Public Health : We can have a better understanding of what is happening, prevent virus from spreading and raise awareness more easily.

---

* Supported and supervised by Marco Winckler

- Renewable Energy : With the calculation of energy consumption in order to prevent shortage
- Fraud Detection : Companies can detect frauds more easily with graphs that links fraudulent account. For example, during my internship at the cybersecurity division of the CAF, developers were creating a new application called Graphaloc in which we could see, in different graphs and charts, all the accounts with the same phone number, IBAN or even addresses. It was most used by CAF's controllers in order to detect frauds.
- Communication : More and more companies uses data to communicate, to sell their products, to compare themselves to others and even to understand their strengths and weaknesses

Data visualization is also regarded as information visualization or scientific visualization. Human beings have always employed visualizations to make messages or information last in time. What cannot be touched, smelled or tasted can be represented visually. [6]

Data visualization has been used extensively to inform users. However, little research has been done to examine the effects of data visualization in influencing users or in making a message more persuasive. [5]

The downside of this field is that some people can use data to their benefit and manipulate others. Because with some confusing charts or scale effect (Logarithmic scale for example) we can lead people to make wrong interpretations. You can even compare two completely different data with no correlation to prove a point and some people will believe it, so it can become a communication weapon.

In my opinion, it is therefore a vast field with many uses which tends to develop more and more and is increasingly accessible and sought after but we have to use and interpret it carefully.

2   Different techniques of data visualization

2.1   Types of data visualizations

There are numerous types of data visualizations, each one as its uses, its strong and weak points. The most commons are :

- Line graph : Is compatible to show relationship between data or the evolution in a given period of time. For me, this type of visualization is not adapted to the visualization of percentages or proportions.
- Scatter plot : Shows the variations of two or multiple items. It is very good to show the correlations between two variables
- Pie chart : The best representation of shared data, to show proportions or percentages but has no other use.
- Bar chart : It is used to compare quantities of different categories. It can even represents categories that have nothing to do together so it is very useful when you compare data in large fields. But it can be used to mislead people by comparing data and create false correlations.
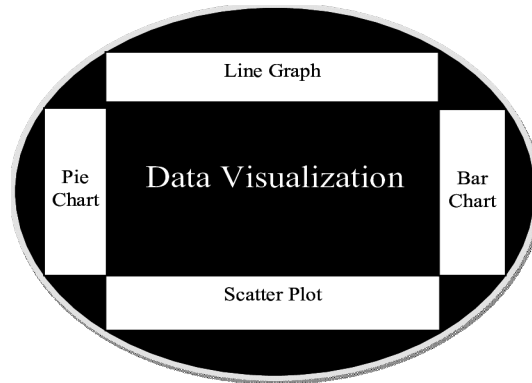
Fig. 1. Commonly used data visualization types. [6]

Each type of data visualizations is useful and you need to chose the best in order to make your visual suitable for your objectives.

2.2   Techniques of data visualizations

There are different types of techniques in visualization but the process is the same : A request is made by a user, the objective is described, then we must collect the data, analyze it in order to be able to model it as we wish.
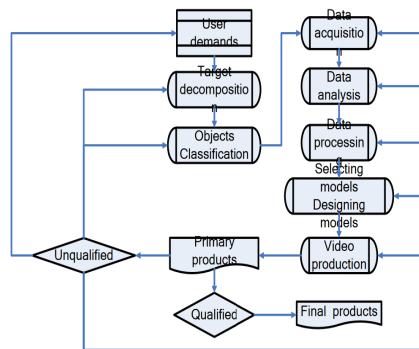

Fig. 2. Process of data visualization. [7]

In this section, we are going to see some examples of well used technique of data visualization.

One of the best software and language to manipulate and compute statistics is R. There are several useful packages and one of the best for the biology's searchers is SVGMapping framework. As High-throughput omic technologies are now commonly used in large-scale experimental biology [1], searchers were in need of something to manipulate this huge amount of complex data. In this package, you have all the template allowing you to visualize the shape of molecules such as proteins, DNA, RNA, etc. This kind of package is very well suited for the field it has been created for but has little use in others.

For lambda users, there is different techniques to manipulate the data such as learning python and R if you are familiar with coding and statistics, you will be able to create basic charts or graphics, but it can quickly become difficult to manipulate a large amounts of data or complex data. You can also use paid services such as Amazon AWS or Data Scientists, but again, it can quickly become expensive with large amount of data.

Table 1. Different techniques of Data visualization.

Visualization Technique	Field of Use
SVGMapping[1]	Biology (protein visualization
Visual Information-Seeking Mantra[2]	Interactive visualization
Linked data visualizer[4]	Semantic Web / Linked Data
GsmPlot[3]	Epigenome visualization

3   Conclusion

For me, data visualization is a very rich field with a large amount of techniques, some are useful, some less. It has many uses and it helps a lot of people to understand their proper field, fields that they are not expert in or even themselves. It helps some who doesn't have access to culture or information to understand many things with just a graph or some plots. The only bad point that we were able to observe in this field is the fact that some can manipulate others by changing the scales, the data or even creating fake data.

References

1. Champeimont, R., Leplat, C., Chauvat, F., Aude, J.C.: SVGMapping: an R package to map omic data sets onto pathways templates. In: 1ères Rencontres R. Bordeaux, France (Jul 2012), https://hal.archives-ouvertes.fr/hal-00717509
2. Conti, G.: Security Data Visualization: Graphical Techniques for Network Analysis. No Starch Press (2007), google-Books-ID: Cg2cEz10XpMC
3. Li, J., Yin, Y., Zhang, M., Cui, J., Zhang, Z., Zhang, Z., Sun, D.: GsmPlot: a web server to visualize epigenome data in NCBI. BMC Bioinformatics **21**(1), 55 (Feb 2020). https://doi.org/10.1186/s12859-020-3386-0, https://doi.org/10.1186/s12859-020-3386-0

4. Menin, A., Michel, F., Gandon, F., Gazzotti, R., Cabrio, E., Corby, O., Giboin, A., Marro, S., Mayer, T., Villata, S., Winckler, M.: Covid-on-the-Web: Exploring the COVID-19 scientific literature through visualization of linked data from entity and argument mining. Quantitative Science Studies **2**(4), 1301–1323 (Dec 2021). https://doi.org/10.1162/qss$_{a0}$0164, $https://doi.org/10.1162/qss_{a0}0164$

5. Pandey, A.V., Manivannan, A., Nov, O., Satterthwaite, M., Bertini, E.: The Persuasive Power of Data Visualization. IEEE Transactions on Visualization and Computer Graphics **20**(12), 2211–2220 (Dec 2014). https://doi.org/10.1109/TVCG.2014.2346419, conference Name: IEEE Transactions on Visualization and Computer Graphics

6. Sadiku, M., Shadare, A., Musa, S., Akujuobi, C., Perry, R.: DATA VISUALIZATION. International Journal of Engineering Research and Advanced Technology (IJERAT) **12**, 2454–6135 (Dec 2016)

7. Wu, Q., Li, X., Wang, D., Jiao, W., Han, X.: Study on the Theory and Practice of Data Visualization. In: Li, D. (ed.) 10th International Conference on Computer and Computing Technologies in Agriculture (CCTA). Computer and Computing Technologies in Agriculture X, vol. AICT-509, pp. 1–9. Springer International Publishing, Dongying, China (Oct 2016). https://doi.org/10.1007/978-3-030-06155-5$_1$, $https://hal.inria.fr/hal-02179978$