Report - Data visualization

Joint Research Programme, BTO 2022.037 | June 2022

BTO-VO project - Data visualization (2022)

- **Researchers**: Dr. Xin Tian, Dr. Dirk Vries, Siddharth Seshan MSc, Dr. Tessa Pronk, Drs. Bernard Raterman, Janine de Wit MSc, Sharon Clevers MSc

- **Project number**: 402045/304

- **Project manager**: Dr. Geertje Pronk

- **Project quality controller**: Dr. Peter van Thienen

- **Client**: BTO exploratory research

- **Report number:** BTO 2022.037

PO Box 1072

3430 BB Nieuwegein

The Netherlands

T   +31 (0)30 60 69 511

F   +31 (0)30 60 61 165

E   info@kwrwater.nl

I   www.kwrwater.nl

Management summary

Good visualization can considerably improve the interpretability of data and the efficiency of communication. With the latest developments in graphic techniques, we are now able to visualize and dashboard many kinds of seemingly messy (high-dimensional) data in 2D/3D plots, networks, etc. More practically, many visualizations can be presented in an interactive way, in which graphs and data points can be manipulated and customized by users, giving great ease to check data, identify patterns, and present findings.

Importance:

Data visualization and dashboarding are the graphical display of abstract information for two purposes: sense-making (e.g., data analysis) and communication [1-3]. In other words, data visualization is the representation of data or information in a graph, chart, or other visual formats. It communicates the relationship of the data with images. This is important because it allows trends and patterns to be more easily seen [1]. With the rise of (more available) data upon us, we need to be able to interpret increasingly larger and more heterogeneous batches of data. Multiple tools make it easier to visualize and present data. Data visualization is not only important for data scientists and data analysts, but also for many roles in different fields, e.g., marketing, communication, tech, design, etc. Besides, interactive visualization (hereafter referred to as visualization) enables the exploration of data via the manipulation of data embedded in the graph, with the color, brightness, size, shape, and position of visual objects representing aspects of the dataset being analyzed. The use of interactive visualizations is becoming increasingly popular in scientific research and business intelligence, and is now a common part of most analytics suites (e.g., Plotly, Tableau, PowerBI), thanks to its ease of use and added value, which allows a more effective exploration of the data.

Method:

In this report, we first discuss why good data visualization is important for KWR and Dutch water utilities. Then we summarize different types of graphs suitable

for data science projects. Next, we list potential tools and software to make visualization and provide a list of top 5 of recommended tools. Last but not least, we also provide a user guide for quickly visualizing data based on Python.

Results:

The following concrete results have been produced in this project:

- Visualization tools and software (overview): ᴋᴡʀ <u>Tools for visualization</u>

- Types of graphs for visualization (overview): ᴋᴡʀ <u>Types of graphs and maps for visualizing (scientific) data</u>

- Code examples for quick visualization (examples): ᴋᴡʀ <u>Gallery of interactive graphs with easy codes</u>

- Data dashboard examples (examples): ᴋᴡʀ <u>Dashboard or interactive report examples</u>

Application and future directions

In this report, we have addressed data visualization and its use in our research activities. We expect readers can follow the guideline and use interactive data visualization to boost their research efficiency and improve internal/external communications. Following this project, we suggest the following direction and activities:

- Data dashboarding: To use dashboard and web apps to tell a complete and logical story of research

- Formatting visualization: To design a template with KWR color and font styles for visualization, together with colleagues from the Communication team

- Data science: To further use data visualization in data science projects, including Explainable Artificial Intelligence, Decision-making, etc.

This report:

This research is published in the report: Data visualization (BTO-402045-xxx). It becomes public after one year.

Table of Contents

1. Why informative visualization is needed

1.1 Data, information, and knowledge

The terms data, information, and knowledge are used extensively in the context of computer science. There are many definitions of these terms, often conflicting and inconsistent [1]. Generally speaking, the primary objective of data visualization is to gain insight (hidden truth) into the data, turn (sometimes not well organized) data into useful information and knowledge, and make informed decisions. Before talking about data visualization, we assume that readers already have some basic knowledge about data, information, and knowledge. Otherwise, please refer to [1] for details.

1.2 The value of informative and interactive data visualization

Data visualization is the graphical display of abstract information for two purposes: sense-making (e.g., data analysis) and communication [1-3]. Figure 1 (i.e., the bar charts info-graphically displayed in coffee cups) can quickly give you an overview of

different types of coffees. Without such a graph, you may need to read nine paragraphs about the particular decomposition of each type.

Figure 1. Illustration of different types of coffees based on visualization. (Source: https://lokeshdhakar.com/)

As a critical tool to convert messy data to useful information and deliver this information to the audience, data visualization has been widely used in various data-based projects wherever we see increased knowledge or value from data, which can be further determined by doing more data analysis and/or running through algorithms. Figure 2 shows an idea about how data can be presented and explained based on data visualization (and analyses).

Figure 2. From data to information. (Source: https://goalstoresults.com/category/data/)

In other words, data visualization is the representation of data or information in a graph, chart, or other visual formats. It communicates the relationship of the data with images. This is important because it allows trends and patterns to be more easily seen [1]. With the rise of (more available) data upon us, we need to be able to interpret increasingly larger and more heterogenous batches of data. Multiple tools make it easier to visualize and present data. Data visualization is not only important for data scientists and data analysts, but also for many roles in different fields, e.g., marketing, communication, tech, design, etc.

Besides, **interactive visualization** (hereafter referred to as visualization) enables the exploration of data via the manipulation of data embedded in the graph, with the color, brightness, size, shape, and position of visual objects representing aspects of the dataset being analyzed. The use of interactive visualizations is becoming increasingly popular in scientific research and business intelligence, and is now a common part of most analytics suites (e.g., Plotly, Tableau, PowerBI), thanks to its ease of use and added value.

1.3 Data visualization for KWR's researchers and projects

To understand the value of interactive graphs for KWR (and Dutch water utilities), we first need to consider how projects are done at KWR. After initializing a concrete

research question based on the actual need of project partners and ourselves, we need to first design a methodological framework and collect needed data, in which data visualization plays a key role in exploring data trends and patterns for data analyses. When data analysis is done, the results are to be shared with project partners. The use of interactive reports or dashboards can maintain the interactivity of (interactive) graphs and make end-users see exactly what we produce without the need to check codes. In other words, there is no need to re-run codes to see the interactive graphs for the results. For a higher-level use of figures, we can also customize and add controls (e.g., dropdowns, slider bars), for which we need to embed interactive graphs into dashboards. To let users be able to control the graphs as well, the last step is to deploy the dashboard onto a server. So users can get access to it, simply via a link. All in all, the values of interactive visualization for us are shown in the following aspects:

- As a powerful tool to explore data trends and patterns for data analyses in any data-related project (for both data analyzers and end-users)

- A better demonstration of the project outcome

- Enhancement of communication efficiency and effectiveness internally and externally

1.4 Building on past projects at KWR

BTO 2017.012 Verkennend Onderzoek naar Interactieve Grafieken

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/74f74f8b-dfa5-4284-ae04-0e6e19fb7b64/BTO_2017.012_Verkennend_Onderzoek_naar_Interactieve_Grafieken.pdf

Report-BTO2017-012

This report introduced Tableau and Plotly for the first time to KWR researchers, which has laid a good base for the use of interactive graphs. However, visualization techniques have evolved in recent years (e.g., the occurrence of Plotly.express, which is a high-level API to wrap Plotly).

Therefore, based on this BTO project, we further elaborate on recent advances, putting special focus on a guideline about usage, a gallery for reference, and the use of interactive reports.

2. How to make a clear and informative visualization?

2.1 Which type of visualization is clear and informative?

The classic example "Anscombe's quartet", which comprises four datasets with nearly identical simple statistical properties yet very different visual displays, demonstrates the importance of good data visualization prior to performing analyses.

A good visualization is not just a static picture that one can look at, like an exhibit in a museum. In [1], a good visualization is defined as a way that allows us to drill down and find more about data patterns. For example, view first, zoom and filter, change the values of some scale of the display, and view the results in an incremental way, e.g., explained in [2]. In other words, visualization is useful in both organizing and making sense of data, particularly when data are abundant. For doing it, we need interactive visualization, which emerges as a new form of communication and allows users to analyze the information and construct their own, new understanding of the data.

Good visualization helps the users to explore and understand data, providing value and deep insights. It is effective, visually appealing, scalable, and is easy to understand (good visualization does not have to be too complicated). Visualization is a central tool in finding patterns and trends in the data by carrying out research and analysis, using whichever one can answer questions about the data. The main principle behind a good and effective visualization is to identify the main message that we aim to deliver to the audience, recognize the level and background of the audience, accurately represent the data, and then create a clear presentation that conveys the message to that audience [1].

2.2 Visualization tools and software

Good visualization requires tools, which should provide us with an easier way to create interactive visual representations of (sometimes large) data sets. We consider tools for both programmers and non-programmers (at KWR or Dutch water companies) and summarize these tools as follows:

KWR <u>Tools for visualization</u>

💡 Please note that we use subpage in this report, which can be easily shared seperately. A tip is to click your mouse wheel to open it as a new page (rather than left clicking).

2.3 Different types of graphs for visualization

Today, there are over thirty different visual representations of data, each having a reason to represent data in that specific way. As the visualization methods progress, we have much more than just bar graphs and pie charts. Below we summarize the <u>main types of graphs and maps</u> used to visualize data:

KWR <u>Types of graphs and maps for visualizing (scientific) data</u>

2.4 Steps for data visualization

Below are key steps for making a good visualization.

1. Prepare data: data retrieval, cleaning, integration, reduction, and transformation (consult a data scientist about how to do these);

2. Be clear of the target for visualization (what information to deliver) and keep it in mind during the entire process of visualization;

3. Choose a familiar and powerful tool for your visualization from <u>2.2</u>

4. Think about a suitable type of graph(s) from <u>2.3</u> for visualization and sometimes try a few more graphs to find the best way to deliver your message from data;

5. Polish the graph(s) for colors, font sizes, etc. And then add necessary information (title, labels, etc.);

6. Check whether the visualization can meet the target from Point 2 and use the interactive graphs in the follow-up analyses;

7. Continue making other graphs to tell the entire storyline;

3. Check the ease of use for the interactivity of graphs, share with colleagues and end-users, and ask for their feedback.

Please note that typically, data visualization follows data collection and preprocessing. Although how to collect data is out of the scope of this project/webpage, we commend that researchers, during data collection, think about whether collected data are sufficient to present aimed information for visualization and analyses. Or consult a data scientist.

Also note that data visualization is actually a useful tool to help us analyze data, rather than just a means of showing the final result. In other words, we need to put ourselves in a loop depicted below and take the above-mentioned steps more than once if necessary.
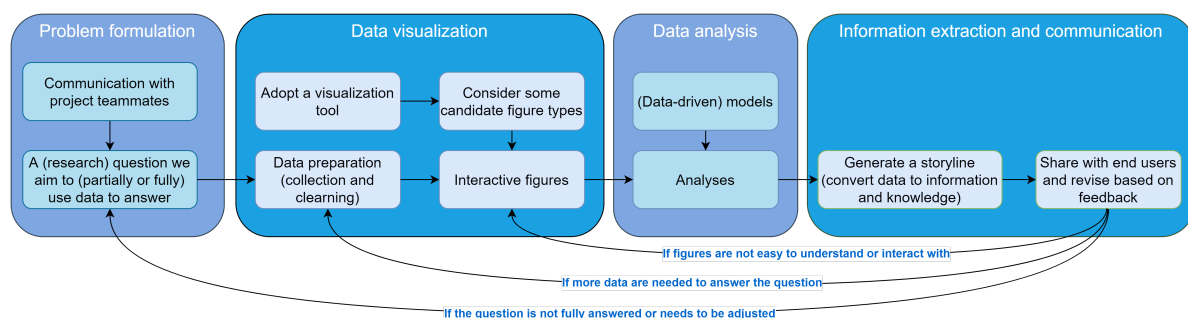

Figure 3. An example workflow for data science projects with involvement of data visualization.

3. Galleries

3.1 Gallery of graphs with example python code

In this section, we provide a reference gallery for Python programmers who are looking for an easy and quick way to visualize data. With this reference, one can expect to find how to visualize data using a specific type of diagrams and how to customerize diagrams (colors, font, etc.).

KWR Gallery of interactive graphs with easy codes

4. Why and how to dashboard data and results?

After making interactive graphs, we need to think of a cohesive set of images to convey a message. What we can do further is to embed these graphs into interactive dashboards and reports, which can add controls of graphs and give the controls to end-users too. The purpose is to allow users to independently conduct deeper mining and custom reports on demand, which allows the end users to better understand the (usually physical, chemical or biological) processes they are interested in. The advantages of interactive reports are apparent as follows.

For end-users viewing and analyzing data, they can do active exploration rather than passive consumption, without any coding required, particularly:

- When the (data) analyst wants to conduct a more in-depth analysis, they should be able to click to zoom in data for more specific information.

- In addition, the end-user can also modify the diagrams and export it to PDF, Excel, PNG, or email without asking for help from the developer.

For developers, interactive dashboards and reports make it possible for the high-efficiency development of reports.

- Interactive reports expand the scope of operations that non-technical users can perform without IT professionals' help.

- Meanwhile, developers can restrain the features available to end-users under permission and security concerns (e.g. disable data download).

5. Dashboard and interactive reports examples

After deploying the designed dashboard or interactive reports, they can be shared with the target audience. Since the deployment itself is out of scope for this project, we limit ourselves by listing a few examples to show deployed dashboards and reports, some of which were made by KWR.

KWR Dashboard or interactive report examples

6. Conclusions and recommendations

This project studied the modern interactive data visualization techniques. We expect readers can well understand the following topics and use interactive data

visualization to boost their research efficiency and improve internal/external communications.

- Visualization tools and software (overview): KWR Tools for visualization

- Types of graphs for visualization (overview): KWR Types of graphs and maps for visualizing (scientific) data

- Code examples for quick visualization (examples): KWR Gallery of interactive graphs with easy codes

- Data dashboard examples (examples): KWR Dashboard or interactive report examples

Following this project, we suggest the following direction and activities.

- Data dashboarding: To use dashboard and web apps to tell a complete and logical story of research

- Formatting visualization: To design a template with KWR color and font styles for visualization, together with colleagues from the Communication team

- Data science: To further use data visualization in data science projects, including Explainable Artificial Intelligence, Decision-making, etc.

7. References

[1] Raman, K. (2015). Mastering Python data visualization: generate effective results in a variety of visually appealing charts using the plotting packages in Python. Packt Publishing.

[2] http://www.mat.ucsb.edu/~g.legrady/academic/courses/11w259/schneiderman.pdf

[3] https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception

[4] https://www.nature.com/articles/d41586-018-01322-9