Portfolio “Telling Data Stories”

Julia Undeutsch

Faculty of Philological and Cultural Studies, University of Vienna

UE 136024 Telling Data Stories - Diagrams, Graphs, Maps and other visual and physical representations of (research) data

Mag. Dr. Ludwig Maximilian Breuer, Mag. Markus Pluschkovits

February 18, 2023

Table of contents

Table of graphics

Introduction

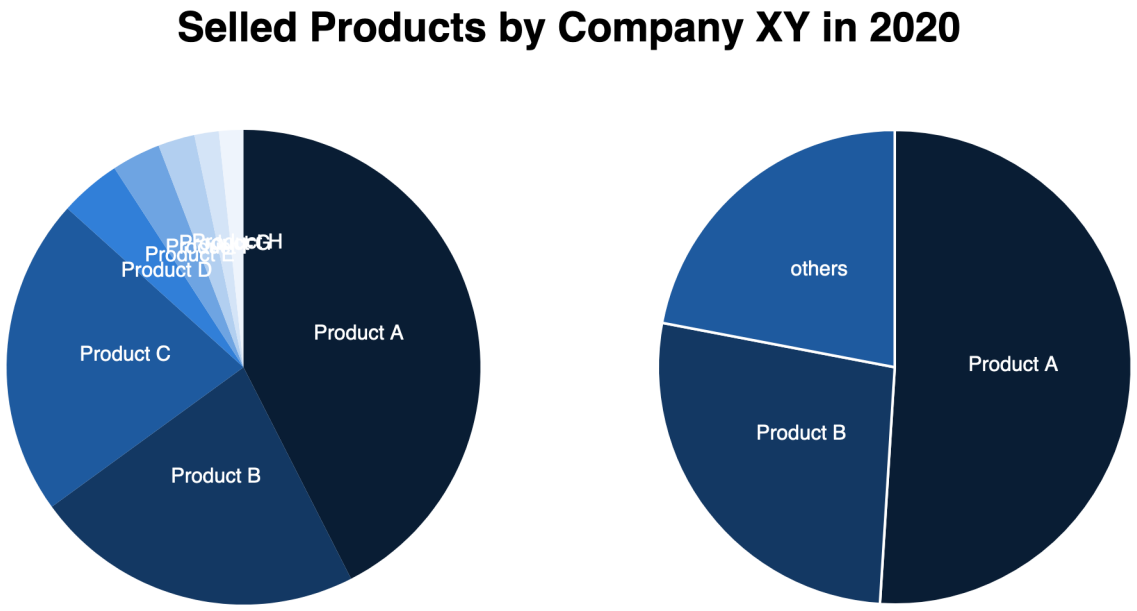
This course focused on learning about the different tools and forms such as QGIS, Jasp, R, Python, etc. that data scientists can use to visualize data. In order to successfully pass this course, students must create a portfolio in which they create a data graph for each session using different tools and techniques and reflect on its origin and impact, as well as explain good and bad practices in relation to their results.

1. Lecture: What is a visualization - a short intro of theory and history of data visualization

The first lecture was an introduction to data visualization and gave a first impression about good and bad charts and how to best use which charts (purpose). For example, although pie charts are the most well-known charts, they also bring certain difficulties [Ricks:2020] that many are not aware of.

I would like to explain these difficulties with the following examples.

These two pie charts were created using D3.js. The dataset was created manually and has no special meaning in either case. They serve purely for the representation of possible problems, which a Pie Chart can bring along in such a way.



Graphic 1: Two Pie Charts demonstrating pros and cons, by Julia Undeutsch.

As the headline shows, both charts are about the sold products of company XY in 2020. The chart on the left has too many slices, which makes it confusing and

impossible to distinguish between them in terms of size. The labels are difficult to read due to overlap and the chart as a whole is difficult to understand.

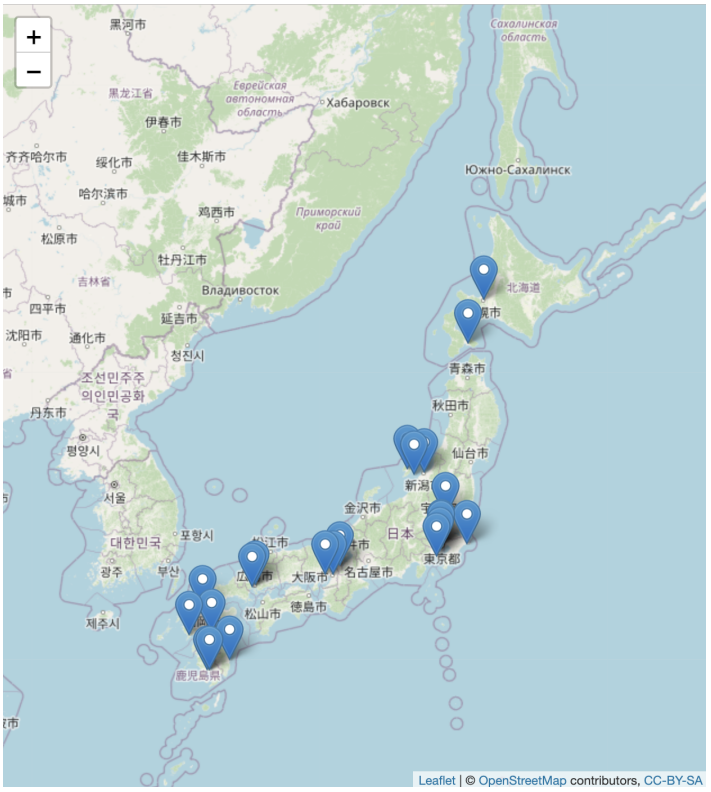
In comparison, the chart on the right highlights the two best-selling products and has the rest grouped into a third slice called "others", which makes it far more understandable. It also differs visually from the left chart in that all labels are legible and the slices are additionally separated by a white border.

One thing that is missing from the charts is a legend with the assignment of name, color and number in e.g. percent, which gives detailed information about the chart. The labels product A, B and so on in the chart alone do not give any indication of how much share they take out in the chart.

This introduction and illustration on simple graphs like the Pie Chart helped a lot to find easier access to more complex charts in the following lectures, to avoid common mistakes and to improve skills to use the right charts and tools for visualization.

2. Lecture: Map and area: Putting Data on Maps (Sprachatlas/LAS/SyHD) - Praxis quantitative: QGIS / Praxis qualitativ: Neatline

The second lesson was about maps and areas and the different ways of representing visual data, such as the frequency of an earthquake in a certain area or election results in a certain country.



Graphic 2: Places visited in Japan, by Julia Undeutsch.

To demonstrate the newfound knowledge of this lesson, I used R and the leaflet library to create a map of Japan and set markers for all the places I have already visited (unique locatable geographic information using latitude and longitude). It can be seen that the more northern it gets, i.e. the Tohoku region and Hokkaido, at times fell less into my itinerary than the central and southern parts (Kyushu) of Japan.

Reasons for this are mainly the times when traveling. When I traveled to Japan for Hanami (Cherry blossom season), the cherry blossoms in the north were already faded, when I traveled to Japan in late autumn, it was already too cold for me personally to stay there for any length of time.

For this map, a label area hardly makes sense, since the markers are all uniformly designed and represent the same value, namely the presence of me.

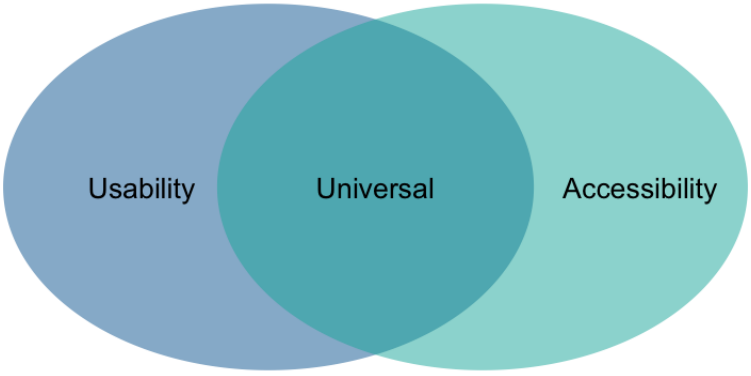
In Live View mode, clicking on a marker brings up a popup with the name of the respective city in Latin rather than Japanese script.

A heading, on the other hand, should be included to make it clear to the reader at a glance what the markers on the map representing Japan mean.

It was my first experience with the programming language R. The short introduction to R by the lecturer gave me enough knowledge to be able to create this map without any problems.

3. Lecture: So basic: descriptive graphs I (Pie, Stacked Bar, Bar and more) - practical examples with R (Python, D3.js)

Depending on how much data will be included in a diagram, it can be simpler or more complex. This session was all about the most common diagrams, how they are presented and best practices on how to provide the necessary data and amount of information so that it can be understood by any user.



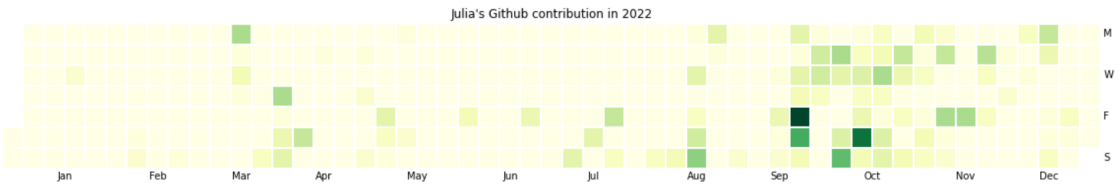
Graphic 3: Commonalities and differences of Usability and Accessibility, by Julia Undeutsch.

To illustrate this, I have created a venn diagram using R that shows the differences between usability and accessibility, as well as the overlap that accessibility experts refer to as universal design. A venn diagram is about the relationship of similar factors. Often the size ratio of the quantities can lead to difficulties (readability, and subsequently comprehension). My representation of a venn diagram is quite simple, but it can also become super complex, overlapping at many different endpoints.

My expertise is in accessibility, so I often use this diagram in presentations to make clear to work colleagues and UX designers how important the connection between usability and accessibility is.

4. Lecture: So special: descriptive graphs II (Dendrogramme, Heatmaps and alternatives) & dynamics - examples with R (Python, Jasp, ...)

I've always liked the charts on my GitHub profile. The green squares were often my motivation to be productive, participate in open source, and more. And now I know the name of this graph, too: Heatmap.



Graphic 4 : Heatmap showing my contribution in 2022 on GitHub, by Julia Undeutsch.

I created the graph using Python, and its libraries numpy, pandas, seaborn, and matplotlib, to create initial evaluations with the dataset I created. To create the

heatmap as it is also visible on GitHub, I installed the library calmap, which simplified the creation of this heatmap for me.

The heatmap visualizes the data (my daily contribution in 2022) by variations in the color scheme. In my case, color shading was chosen from light yellow (hardly any contribution) to dark green (a lot of contribution).

However, this example has some pitfalls to understanding the data properly. For example, the heatmap lacks a legend to read it properly. Numeric data needs a color scale that goes from one color (light yellow in my case) to another (dark green), to which values are assigned. The less contribution, the lighter the color. e.g. (1-5 light yellow, 6-10, medium yellow, 15 and above dark green, etc).

In addition, days on which there was no contribution should be grayed out, for example, to show that there was no data to read here. Without knowing the dataset, this heat map suggests to the user that there was at least 1 contribution every day, because every box is at least light yellow. This problem could be solved by looping through the dataset and assigning the color gray to each box if there is no data or the numeric value is less than 1.

5. Lecture: Telling more: Processes, Infographics & the medium - Info Visualisation (draw.io, ...)

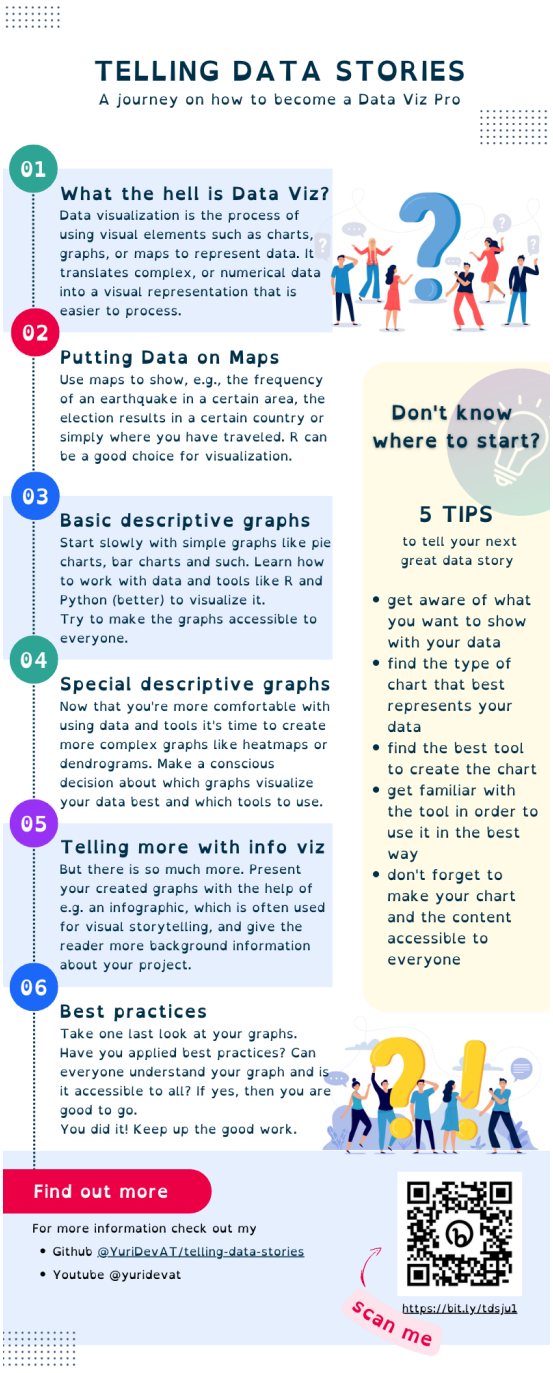
Since an infographic can be seen as a type of visual storytelling (Lankow et al., 2012), I chose to use a progress infographic to represent my experience with the course "Telling Data Stories." The infographic was created using Canva.

Like Tomboc explains in her blog post (2023), there are several bad practices someone should avoid when creating an infographic. These could be, e.g.

- Choosing the wrong charts
- Too long (endless scrolling) / Not enough white space

- No coherent visual language / No clear hierarchy
- Too many fonts / too many colors

Let's take a closer look at my infographic and analyze it for good and bad practices.



Graphic 5 : Infographic showing my progress with the lecture "Telling data stories", by Julia Undeutsch.

The choice of a process infographic was made because I wanted to show the process of the lecture. This is supported by the ascending numbering on the outer left side as well as visually represented by the two graphs on the outer right side (the first graph symbolizes my not knowing (question mark) about the topic in the course, which develops into more clarity over time (question mark, exclamation mark). It also supports the flow of reading (from top to bottom).

The actual display of the progress on the left is kept visually with the two background colors light blue and white, the font color with high color contrast in dark blue to the background to be easily read. For the font, a special one for people with dyslexia was chosen, so as to reach more readers. The last box offers the reader the possibility to get more information through provided links and QR code on interest.

The actual display of the progress on the left is kept visually with the two background colors light blue and white, the font color with high color contrast in dark blue to the background to be easily read. For the font, a special one for people with dyslexia was chosen, so as to reach more readers. The last box offers the reader the possibility to get more information through provided links and QR code on interest.

An info box on the right side is displayed in the complementary background color light yellow to distinguish from the progress. The graphic thus follows best practices throughout.

A time infographic is just one of several different types. There are timeline infographic, static infographic, informative infographic, comparative infographic, and many more.

6. Lecture: Back to the beginning: UI Design; Best/Worst Practices - Poster, presentation and more

The poster was created as a result of research using an eye tracker about web accessibility awareness in another lecture. Different tools were used to display the different graphs.

Are You Aware Of Web Accessibility?

An Eye Tracking Study On Web Accessibility.

Julia Undeutsch
Media Lab, Faculty of Philological and Cultural Studies, University of Vienna

universität wien

Introduction

The web is an increasingly important resource in many aspects of life: education, employment, commerce, and more. It is critical that everyone, whether living with a disability or not, is able to perceive, understand, navigate, and interact with the Web. Therefore, websites, tools, and technologies must be designed and developed so that people with disabilities can use them like people without disabilities. Unfortunately, many people are unaware of web accessibility and its tremendous impact.
This study investigates awareness of accessibility on the web. 14 people of different ages and backgrounds participated in this experiment and were asked to view screenshots of home pages in terms of their structure and accessibility.

Hypothesis

H1: Participants are not aware of accessibility on the web.
H2: Participants have misconceptions about accessibility on the web.
H3: Participants distinguish between UX and accessibility.

Method

In the MediaLab laboratory at the University of Vienna, 14 participants performed an eye-tracking measurement followed by a questionnaire on knowledge about accessibility on the web.

SR Research EyeLink 1000 Plus

- ✓ Eye gaze data were collected using the SR Research EyeLink 1000 Plus
- ✓ Each participant was presented with 10 different screenshots of homepages
- ✓ Each screenshot was viewed for 20 seconds using the EyeLink1000 Plus
- ✓ Participants observed each screenshot for structure and accessibility

After the eye-tracking experiment, each participant answered a questionnaire with the opportunity to view the screenshots again in printed form with no time limit.

Questionnaire

- ✓ Questions were asked about the structure, appearance, style, and accessibility of the homepages (Rating system: evaluation range in 5 levels, where 1 means "very bad" and 5 means "very good")
- ✓ Regarding accessibility, participant were asked about their awareness (open question answered with keywords)
- ✓ Finally, participants were asked to rate the top three homepages according to personal preference

Results

For H1, the questionnaire asked participants about their familiarity with accessibility. Contrary to expectations, more than 80% indicated that they had heard of or were aware of web accessibility.

	Not at all	I can only imagine	I have an idea	Absolutely yes
Do you know what "web accessibility is?	1	1	7	4

1 participant abstained from voting.

When examining screenshots of homepages for accessibility with the SR Research EyeLink 1000 Plus, participants seem to notice accessibility problems for the most part.
As an example, *Graph 1* shows a screenshot of a home page that was rated as one of the most accessible home pages. The Area of Interest (AoI) was set at the top right, both for the image/icon in the text and for the search field. As the heatmap shows, participants perceived the image within the text (which does not necessarily mean it is so for accessibility reasons), but not the low color contrast of text and background in the search box.

Graphic 1

Graphic 2

This result partially supports H2 that participants tend to have misconceptions about Internet accessibility. This is supported by the responses to the questions about how participants describe accessibility using keywords (*Graphic 2*). Regardless of how participants responded to the questions about whether or not they are aware of web accessibility, misconceptions could be found in all 4 familiarity groups (awareness of accessibility), mainly related to color.

Accessibility rate of participants compared to actual rate of accessibility measured with automated testing tool Axe DevTools: 5 means "very good", 1 means "very bad".

Participants rated websites according to how accessible they thought they were based on the user interface (UI). As only screenshots were shown, the user experience (UX), i.e. interaction with the website, could not be taken into account. The line graph shows that websites tend to be rated lower on accessibility than they actually are (H2).
Two websites were rated significantly lower than they actually are, website 7 and website 9, neither showing any unusual eye movements or fixations that would explain why participants rated these two websites lower. Only the keywords indicate that it was due to misconceptions about colors, that dark colors, "bad" colors and high color contrast are bad for accessibility (provided answers of the participants).
Against expectations (H3), when asked which website they personally prefer, participants chose websites 1, 2, and 8 (average score), which are also the ones that participants rated as most accessible.

Conclusion

- ✓ The results show that although not all participants claimed to know exactly what web accessibility means, the theoretical understanding is quite present. This is shown by providing keywords to answer the questions about what web accessibility means.
- ✓ There were deviations in the evaluation of the homepages for their accessibility and the actual accessibility. This can be explained by the misunderstanding of colors in relation to accessibility.

References

What is Web Accessibility? https://www.w3.org/WAI/fundamentals/accessibility-intro/ | EyeLink * Data Viewer User's Manual, Document Version 3.2.1, (2018). Mississauga, Ontario, Canada: SR Research Ltd. | SR Research Experiment Builder User Manual, Version 1.10.1630, (2004-2015) SR Research Ltd. Mississauga, Canada https://www.sr-support.com/forums/showthread.php?t=99

Graphic 6 : Poster about research project "Are you aware of web accessibility?", by Julia Undeutsch.

Since the University of Vienna is required by law to make its website and print materials accessible (color, font size, etc), I chose their poster template to present my findings.

Jasp was used to analyze the data, resulting in the line graph for the "Website rating on their accessibility" display, which can be seen in the middle right. The data table at the bottom left of the poster was created with Excel, and the word cloud at the top right was created with D3.js.

The format of a classic poster allows the reader to be given detailed information about a research project, as it uses text, visual divisions in color and format, and various visual representations. Information in this example is presented as a process, i.e., from the beginning of a research, through its execution, to the presentation of results and a summary. Posters can be overwhelming for some, as a lot of information is included and concentration or interest wanes as you read (especially if you don't find the topic appealing). Therefore, nowadays there are other poster formats, such as the BetterPoster, which advertises to be more user-friendly.

7. Conclusion

An understanding of different types of data charts, methods and tools could be acquired in this lecture. Creating the portfolio gave me the possibility to show off my new gained knowledge about data visualization as well as skills on how to use certain tools and techniques. This will be helpful for future projects in order to making the right choice on with which chart the data can be presented in a good way for everyone to understand as well which tools can be used to create the expecting outcome in an easier way - that the tool is here to help making charts easier to understand and visualize and not be a burden because you don’t know how to use it.

7.1. Key takeaways

Be aware of what you want to show with your data and choose the perfect chart to represent it. Ask yourself the question: What is being visualized? How successful is the visualization? How can the visualization be improved? It is of enormous importance to choose the right diagram depending on the data situation!

References

Drucker, J. (2021). *The Digital Humanities Coursebook. An Introduction to Digital Methods for Research and Scholarship*. Routledge.

Undeutsch, J. (2023, February 11). *Telling-Data-Stories* (GitHub Repository). https://github.com/YuriDevAT/telling-data-stories , last access: 2023, February 18

Storytelling with data. Chart guide, (n.d.). https://www.storytellingwithdata.com/chart-guide last access: 2023, February 18

Ricks, E. (2020, May 14). What is a pie chart? In: *Storytelling with data*. https://www.storytellingwithdata.com/blog/2020/5/14/what-is-a-pie-chart last access: 2023, February 18

What is data visualization? (n.d.). https://aws.amazon.com/de/what-is/data-visualization/ last access: 2023, February 18

Tomboc, K. (n.D.) What is an infographic? In: *Piktochart* https://piktochart.com/blog/what-is-an-infographic last access: 2023, February 18

Ritchi, J. Crooks, Ross and Lankow, J. (2012). *Infographics: The Power of Visual Storytelling*, John Wiley & Sons.