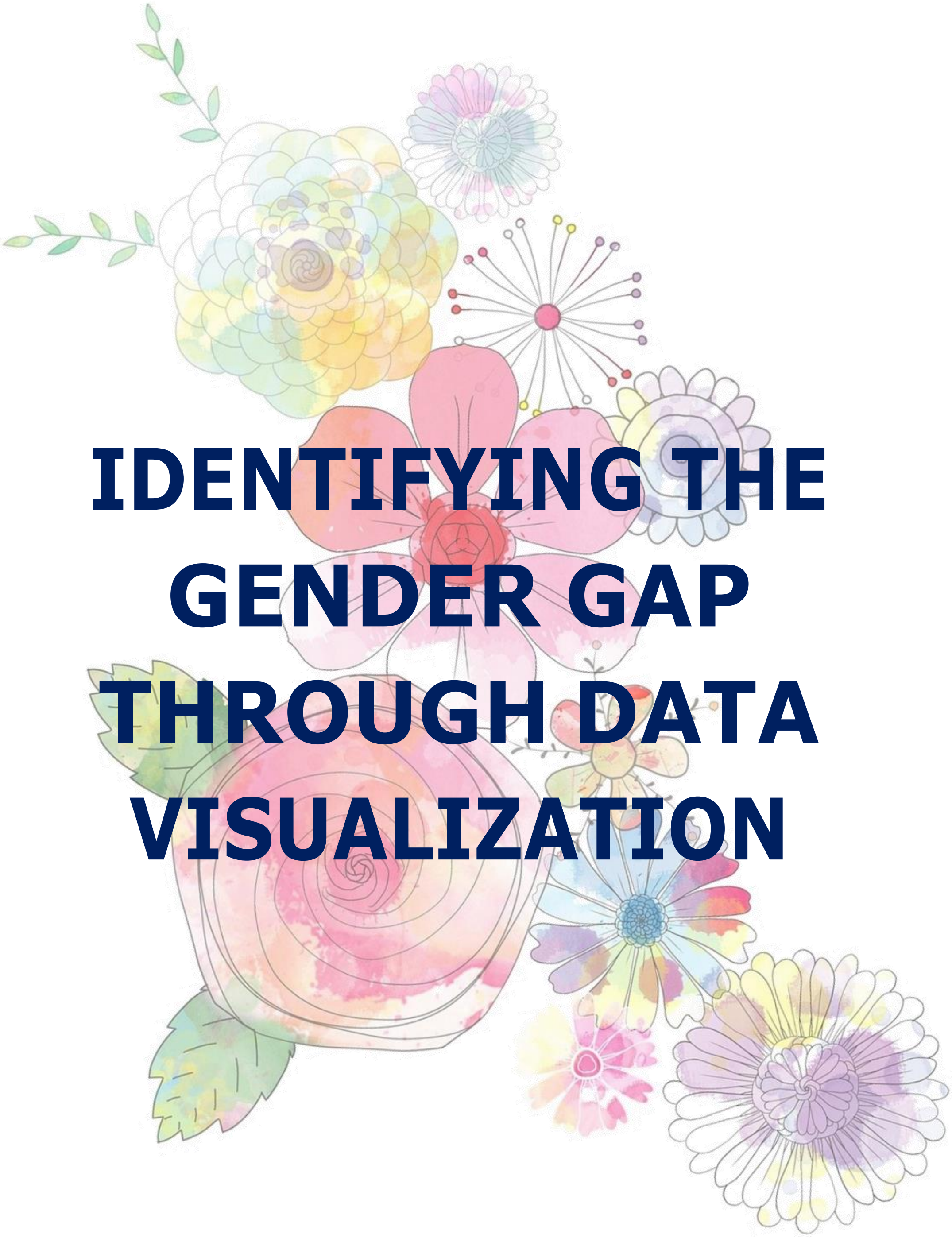
CHAPTER 5

IDENTIFYING THE GENDER GAP THROUGH DATA VISUALIZATION



Dra. Cristina Portalés

cristina.portales@uv.es

Universitat de València

2023 Cover image: Pixabay

ISBN: 978-84-9133-589-4
DOI: http://dx.doi.org/10.7203/PUV-OA-589-4

EXECUTIVE SUMMARY:

Data visualisation provides users with intuitive means to interactively explore and analyse massive datasets in a more intuitive way. Graphical representations not only allow us to visualise and analyse the message contained in the data, but also to remember it. Nevertheless, visualising data can be challenging, as there are many factors that might determine the type of visualisation that is optimal for a given dataset. For instance, it is relevant to know some basics on data visualisation, such as which aesthetics better represent better which variables. Also, it is important to know what degree of expertise the people that the graphs are directed to have. The medium also determines the visualisation (e.g. paper vs. screen), while the consideration of 2D vs. 3D is another factor.

Among the huge variety of purposes and topics that visualising data can tackle, the gender gap is one that has been gaining relevance over the last few years. At the European Union level, it is relevant to highlight the initiative of the European Institute of Gender Equality (EIGE) to compute the Gender Equality Index (GEI). This index is a composite indicator that measures the complex concept of gender equality and, based on the EU policy framework, assists in monitoring progress of gender equality across the EU over time. This indicator is provided with a set of graphical representations for better understanding. Yet, even with the provided visualisations, it is not straightforward to answer all questions about the gender gap. In this chapter, we will go deeper into different ways of visualising the GEI index, analysing the meaning that can be taken from each graph, analysing the pros and cons, and proposing different solutions to respond to specific questions.

Keywords: Data Visualisation, Gender Equality Index, Gender Gap, Charts, Graphs, Aesthetics.

1. Introduction

Data visualisation provides users with intuitive means to interactively explore and analyse massive datasets, which can be dynamic, noisy and heterogeneous, enabling them to effectively identify interesting patterns, infer correlations and causalities, and support sense-making activities (Bikakis, 2018), making it possible to amplify human cognition (Chan, 2006; Protopsaltis et al., 2020). Graphical displays not only allow us to visualise and analyse the message contained in the data, but also to remember it, since for most people, visual memory is more persistent than verbal or auditory memory (Zinovyev, 2010). For all these reasons, data visualisation is nowadays one of the cornerstones of Data Science, turning the abundance of Big Data being produced through modern systems into actionable knowledge (Andrienko et al., 2020), allowing tons of data to be synthesised in visual forms that humans are able to understand.

Among the huge variety of purposes and topics that visualising data can tackle, the gender gap is one that has been gaining relevance recent years. The gender gap can be defined as a gap in any area between women and men in terms of their levels of participation, access, rights, remuneration or benefits (*Gender Gap*, 2022). Many examples of charts visualising the gender gap can be found, for instance, in Flowingdata (Yau, 2022b), an independent site where people can share their charts and some of them are tagged as gender related (Yau, 2022c). The line chart on the decline of women in computer science (Yau, 2014a) shows that, while the percentage of women in other technical fields has risen in the last 30 years, the percentage ofwomen in computer science has declined. This decline coincides with when computers were mostly marketed towards boys in the 1980s. Another example is a bubble chart on PhD gender gaps around the world (Yau, 2014e), which shows that, in almost three quarters of the 56 considered nations, more men than woman receive a PhD.

In another work (Yau, 2022d), a combination of beeswarm, difference, and stacked area charts are used to depict the most female and male occupations since 1950. As more women entered the workforce, many occupations saw a shift from mostly male to a majority or more female, such as opticians, human resources assistants or bill and account collectors. Focusing on the last collected data (years 2000–2015), for those jobs involving the word “computer”, there are more men than women, while for those involving the word “education”, there are more women than men. A project by The Pudding results in a bubble-like artistic graph that summarises the common words used to describe men and women’s bodies in literature (Sposto, 2022). The graph reveals that, for describing parts of the head, the words hair, cheek, smile, face, lip, eyelid or eye are more often used for women, while the words brain, head, forehead brow, eyebrow, pupil, ear, nose, nostril, grin, jaw, mouth or tooth are more often used for men. It also shows how many times this is likely to happen, e.g., the word head is 2.23 times more likely to appear for women, while the word brain is 1.61 times more likely to appear for men.

Another example worth commenting on refers to the World Bank, which, as an effort to make gender inequalities more obvious, updated their Gender Data Portal with different visualisations (*How Data Can Accelerate Equality*, 2022): “*The World Bank Group has redesigned its Gender Data Portal with these audiences in mind by offering over 900 gender indicators in different formats, ranging from raw data to appealing visualizations and stories. Making sex-disaggregated data easier to analyze, interpret and visualize will bring into focus gender issues that are frequently invisible, including on topics such as digital development, transport, and water. It will highlight* *existing gender gaps as well as gaps in the availability of gender data.*" Thus, it recognises the relevance of data visualisation for identifying the gender gap.

At the European Union level, it is relevant to highlight the initiative of the European Institute of Gender Equality (EIGE) to compute the Gender Equality Index (GEI) (*Gender Equality Index | 2021*, 2021). This index is a composite indicator that measures the complex concept of gender equality and, based on the EU policy framework, assists in monitoring progress of gender equality across the EU over time. The visualisations are mainly composed of an interactive radar chart that summarises the GEI indicator for each country and for the EU, as shown in Figure 1. When clicking on one of the countries, pie charts for different domains (Work, Money, Health, etc.) are depicted; and when clicking on one of the pie charts, detailed information on the related domain is given and complemented with bar charts and tables.

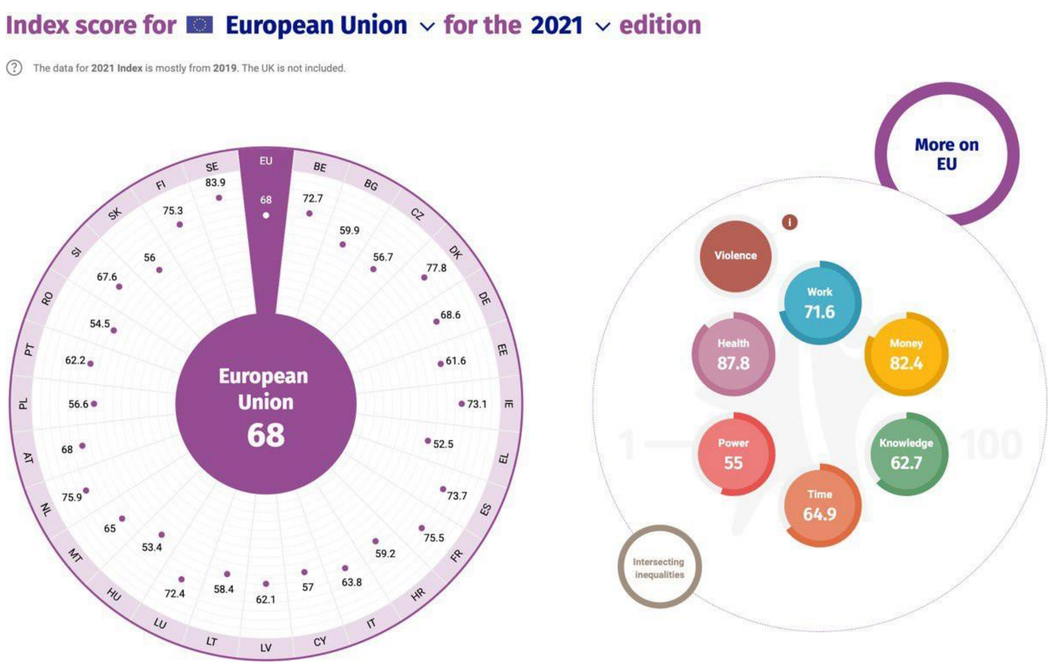


Figure 1. Snapshots taken from the EIGE site. Source: (*Gender Equality Index | 2021*, 2021).

Said graphs are easy to understand and provide a quick and synthesised access to the GEI indicator and domain values for each year and country. Yet, with the provided visual representations, it is not straightforward to answer all questions about gender gap, such as: How many countries are above the GEI average for a given year? What is the temporal GEI evolution for a given country? Overall, which domains present the smallest or the greatest gender gap? Are there relevant patterns in the data?

Because of the relevance of this index at the EU level, in the next section we will provide different visualisations to try to answer these questions. We will start with simple representations (e.g. bar charts) and then move on to more complex ones (e.g. heatmaps), adding a discussion to each chart and unveiling relevant characteristics of the GEI indicator, which are not evident from the inspection of the visualisations provided in (*Gender Equality Index | 2021*, 2021), thus contributing to the understanding of the gender gap in the European Union.

2. Towards visualizing the Gender Equality Index

Before visualizing the Gender Equality Index (GEI), it is important to understand how it is built. The GEI indicator is calculated according to six different domains – Work, Money, Knowledge, Time, Power, and Health – and each domain is computed according to different factors that measure the relationship between men and women. The relationship between the GEI indicator and the domains is not linear. Instead, the GEI indicator is calculated following the methodological approach described in (*Gender Equality Index 2017*, 2017), leading to equation (1):

$$GEI = Work^{0.19} * Money^{0.15} * Knowledge^{0.22} * Time^{0.15} * Power^{0.19} * Health^{0.1} \quad (1)$$

All data involved in the calculation of the GEI indicator for the years 2013, 2015, 2017, 2019, 2020 and 2021 is openly available at (*Gender Equality Index - dataset*, 2022) in the form of an Excel file. To derive the charts in this section, the relevant data in the Excel file was exported to a CSV file, with eleven variables and 168 features. In Figure 2, the names of the variables and the first and last features are depicted.

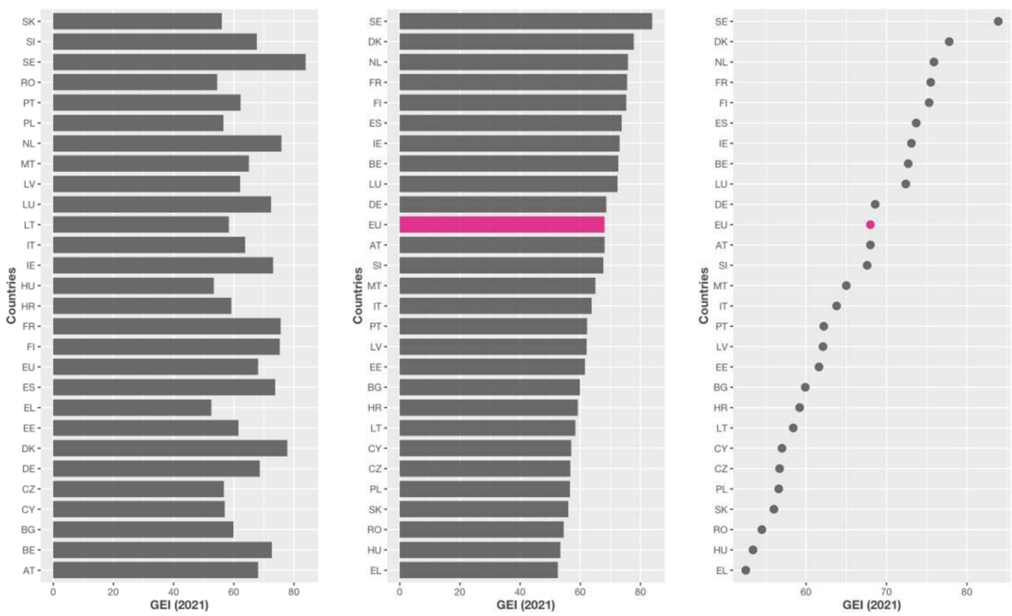
```
Index year;Reference year (main);Protocol order;Country;Gender
Equality Index;WORK;MONEY;KNOWLEDGE;TIME;POWER;HEALTH
2013;2010;0;EU;63,1;69,7;79,1;59,8;65,2;41,9;86,7
2013;2010;1;BE;69,3;72,7;85,5;70,6;70,3;47,9;86,5
2013;2010;2;BG;55,0;67,9;60,8;50,4;43,9;45,8;75,3
2013;2010;3;CZ;55,6;64,9;73,8;55,4;53,8;31,0;85,7
2013;2010;4;DK;75,2;79,8;83,6;73,2;80,4;58,0;90,3
...
2021;2019;23;RO;54,5;67,5;69,1;52,8;50,3;34,7;71,3
2021;2019;24;SI;67,6;73,0;83,7;56,6;72,9;53,0;87,8
2021;2019;25;SK;56,0;66,8;75,1;61,6;46,3;30,7;85,5
2021;2019;26;FI;75,3;75,5;87,9;61,9;77,4;74,3;89,5
2021;2019;27;SE;83,9;83,1;85,4;75,2;90,1;84,5;94,6
```

Figure 2. Summary of raw data used to derive the charts in this section.

The index is calculated based on 27 countries, which are listed with their country code: BE (Belgium), BG (Bulgaria), CZ (Czechia), DK (Denmark), DE (Germany), EE (Estonia), IE (Ireland), EL (Greece), ES (Spain), FR (France), HR (Croatia), IT (Italy), CY (Cyprus), LV (Latvia), LT (Lithuania), LU (Luxemburg), HU (Hungary), MT (Malta), NL (Netherlands), AT (Austria), PL (Poland), PT (Portugal), RO (Romania), SI (Slovenia), SK (Slovakia), FI (Finland), SE (Sweden). Also, the EU (European Union) average is provided.

Raw data is usually difficult to interpret, i.e. just looking at the numbers, we cannot easily extract meaning. Data visualisation can aid synthesising such information in charts, which are easy to understand, and thus it can help respond to specific questions, as will be seen in the following paragraphs. All the graphs in this section have been developed by the author of this chapter, making use of the R programming language. For further information on how to produce graphs with R, a detailed R graph gallery can be found at (Holtz, 2022).

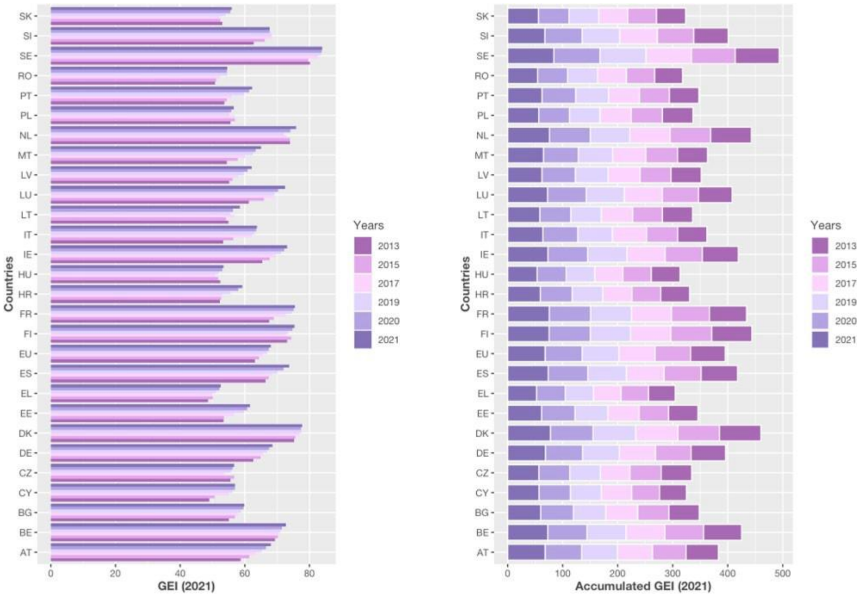
Let’s start with a basic example, a graph that shows the GEI indicator for the year 2021 for each of the countries. The easiest way to show quantities is a bar plot, as shown in Figure 3a. This figure shows the countries in the given order and, in addition to the countries, it includes the average index for the EU. Figure 3b is derived by rearranging the countries according to the index and highlighting the bar for the EU. Differently from the first bar plot, this one is easier to interpret as one can easily identify for which countries this index is smaller and for which it is bigger, and which countries are above and below the EU average. But we can still improve this graph. As the graphical representations of bars indicate areas, they always start at the origin of coordinates, which is one important limitation of this type of visualisation (Wilke, 2019). However, the GEI indicator is concentrated in the values around 50 and 85 points. Thus, another way of representing such amounts can be done with by placing dots at the appropriate locations along the x or y axis, as given in Figure 3c. In this figure, each individual country GEI (2021) indicator is shown with a grey dot, while for the EU it is highlighted with another color. By limiting the axis range to the interval from 45 to 85 points, the figure highlights the key features of this dataset: SE has the highest GEI among all the listed countries, with a difference of more than five points compared to the second country. We can also see that there are sets of countries withs similar GEI values (e.g. ES, IR, BE, LU) and that DE, AT and SI are quite close to the EU average. Also, it is evidenced that RO, HU, have much lower GEI than all other countries.



(a) (b) (c)

Figure 3. Charts showing the GEI indicator for the year 2021, using: (a) unordered bar chart;(b) ordered bar chart with a highlight; (c) ordered scatter plot. Source: own elaboration.

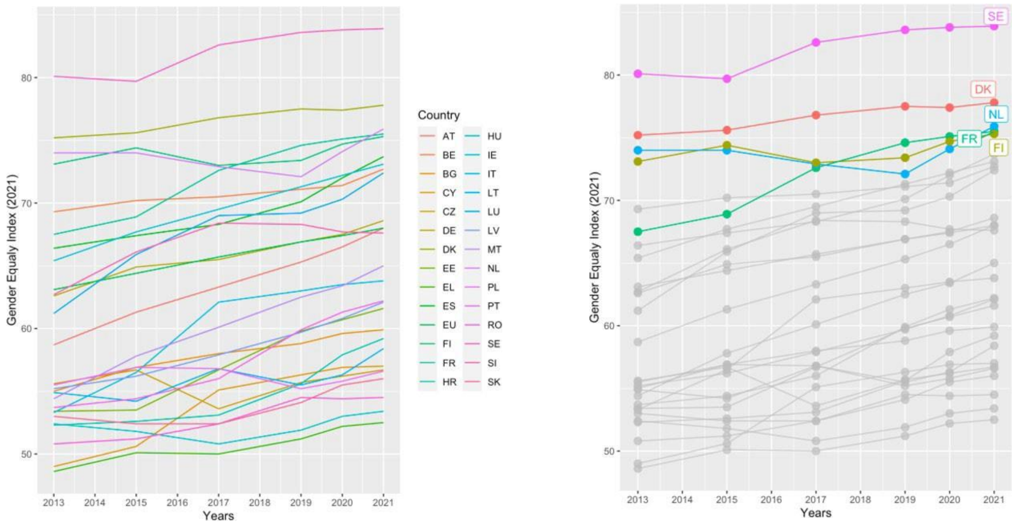
We can go deeper in the data by adding another dimension, for instance, the year. In this case, we could try to build grouped bar plots, but the results are not optimal, as seen in Figure 4a, because the bars are thin and comparison between different countries is difficult. Another option is to use stacked bar plots instead, as seen in Figure 4b, where the horizontal axis represents the accumulated GEI over the years. Although in this case the individual bars that represent each year are clearly seen, it is difficult to see the temporal evolution. For instance, looking at individual countries in Figure 4b, it is not clear in which year the GEI value is greater. Also, it is difficult to compare a single year between different countries because the bars do not share the same base line (except for the year 2021).



(a) (b)

Figure 4. Temporal evolution of the GEI indicator, using: (a) grouped bar plot; (b) stacked bar plot. Source: own elaboration.

Time series are usually represented by lines that connect observed data, as shown in Figure 5a. Line graphs are appropriate whenever one variable imposes an ordering on the data (Wilke, 2019), as shown in the example here. However, the generated graph is difficult to read. The problem here is that there are so many countries that it is difficult to discern which line corresponds to which country. An alternative graph is presented in Figure 5b, where, despite the lines being shown for all countries, only some of them are highlighted – the five countries with the greatest GEI values for the year 2021. Additionally, in Figure 5b, dots are included, which represent the observations, thus clearly showing that there are missing values for the years 2014, 2016 and 2018, which was not evident in Figure 5a. However, if we want to produce a graph that summarises all the countries, we need to explore other representations.



(a) (b)

Figure 5. Temporal evolution of the GEI indicator, using: (a) line chart; (b) line chart highlighting the five countries with the greatest GEI for the year 2021. Source: own elaboration.

An alternative to avoid a graph with so many lines is to represent the temporal evolution with heatmaps, as shown in Figure 6, where the countries are ordered according to the GEI values for the year 2021. This new graph is less busy and still can encode the three variables (country, year, and GEI) in a single graph. This kind of visualisation is good at representing large datasets and highlighting broader trends (Wilke, 2019).

However, while, in Figure 5, the GEI indicator is encoded using the “position” aesthetic along the vertical axis, in Figure 6, the GEI indicator is encoded with the “color” aesthetic, which is more difficult for humans to discern an exact value of on a continuous scale. For instance, looking at Figure 6, would you be able to say what the GEI value for SE is in the year 2015? Comparing the color in the corresponding cell with the color ramp given in the legend, one could say that it has a value somewhere between 75 and 80, for example. On the other hand, when looking at the graphs in Figure 5, one can clearly see that the value is slightly below 80. Therefore, with heatmaps it is harder to determine the exact data values.

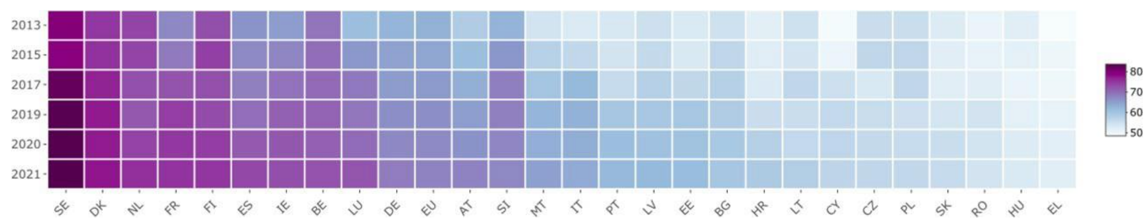


Figure 6. Heatmap for the GEI indicator, for each year and country. Source: own elaboration.

Let’s now explore another variable. Instead of time, we want to find visual representations to simultaneously explore the six domains (Health, Knowledge, etc.) contributing to the GEI indicator for the year 2021. One could think that producing a kind of stacked bars, as in Figure 4b, would be appropriate. We could even calculate the percentage that each domain is contributing to the GEI indicator, and draw the corresponding proportions, so a complete bar would represent the GEI indicator for a given country. The problem here is that the relationship between the GEI indicator and the domains is not linear, as explained above. Therefore, a better option would be to represent single bar charts, as shown in Figure 7.

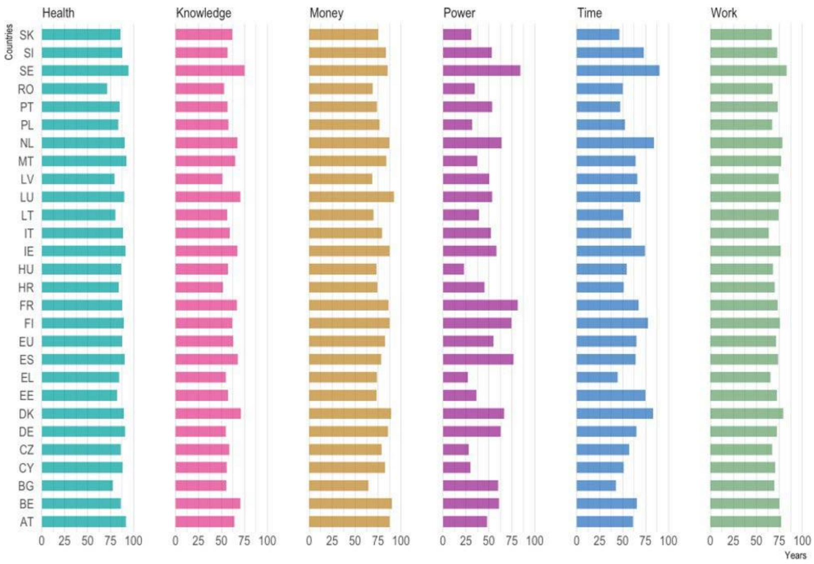


Figure 7. Bar charts for each domain that depict the GEI (2021) value for each country. Source: own elaboration.

We can also represent these data with line graphs, as given in Figure 8, but we already saw that this option was not efficient with the time variable (Figure 5), where we had six values to represent (2013, 2015, etc.), just as with the domains. Why should it work now? The key is how to choose the aesthetic for each variable, which depends on the story. In figure 5, we were interested in exploring the temporal evolution of GEI for each country, so we mapped the GEI indicator to the y-position and the year to the x-position, and thus we had one line per country encoded with the color aesthetic, giving a total of 28 lines (27 countries plus the EU). However, now we want to explore, within a given year (2021), how the domains behave for each country. Therefore, we can choose to map the GEI indicator to the y-position and the countries to the x-position, leaving the lines to represent each one of the domains, so we have only six lines to represent. An example is seen in Figure 8, which is readable (only six lines) and seems a valid representation for the purpose of identifying which domains present the smallest equality (Power) or the greatest equality (Health).

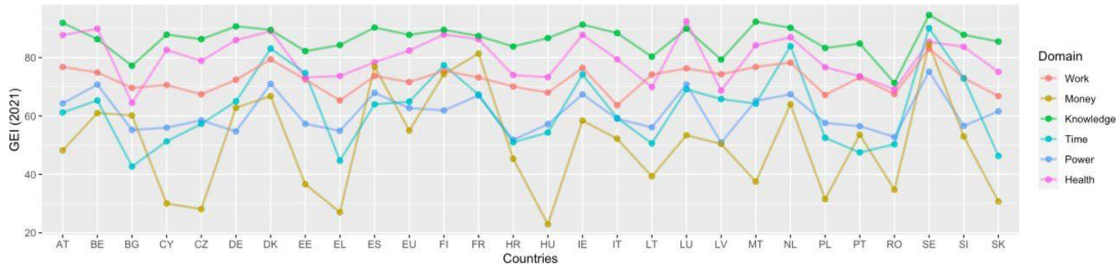
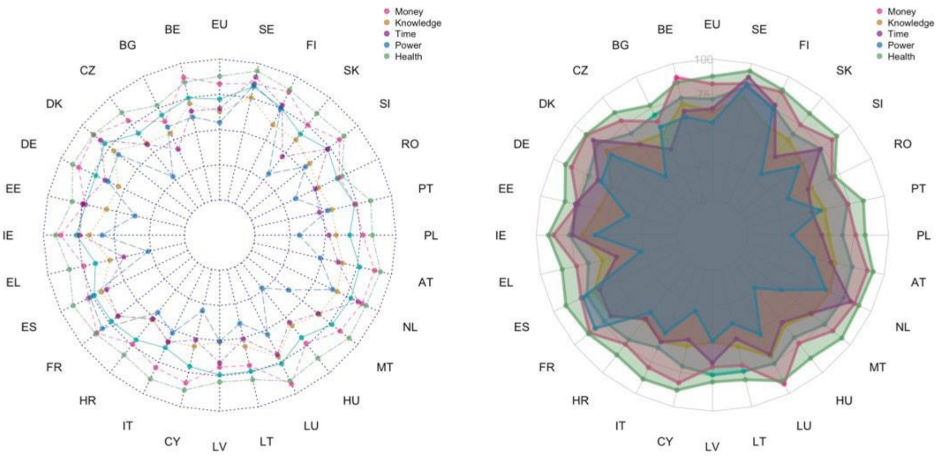


Figure 8. Line chart for the domains contributing to the GEI (2021); depicting the domain values for each country. Source: own elaboration.

Another option for visualising these lines in a single graph is a radar chart, also known as spider or web chart. A radar chart is a two-dimensional chart type designed to plot one or more series of values over multiple quantitative variables. Each variable has its own axis and all axes are joined in the center of the figure (Healy, 2022). An example is seen in Figure 9, where each variable is encoded with a line that forms a closed polygon. A variation is represented in Figure 9b, by filling the areas inside the polygons.



(a) (b)

Figure 9. Radar charts for the six domains contributing to the GEI (2021): (a) radar chart; (b) radar chart with filled polygons. Source: own elaboration.

However, with more than two or three series, it is good practice to use small multiples to avoid a cluttered figure. Such a representation is seen in Figure 10, where one can easily spot that the greatest differences are between the Power and the Health domains. Although, for our case, these graphs seem to work fine, radar charts have been criticised by different authors for reasons such as the huge impact that the category order has in the graphical representation, or the problem of over-evaluation of differences because the area of a shape in a radar chart increases quadratically rather than linearly, among other problems (Healy, 2022). Therefore, we shall explore more charts.

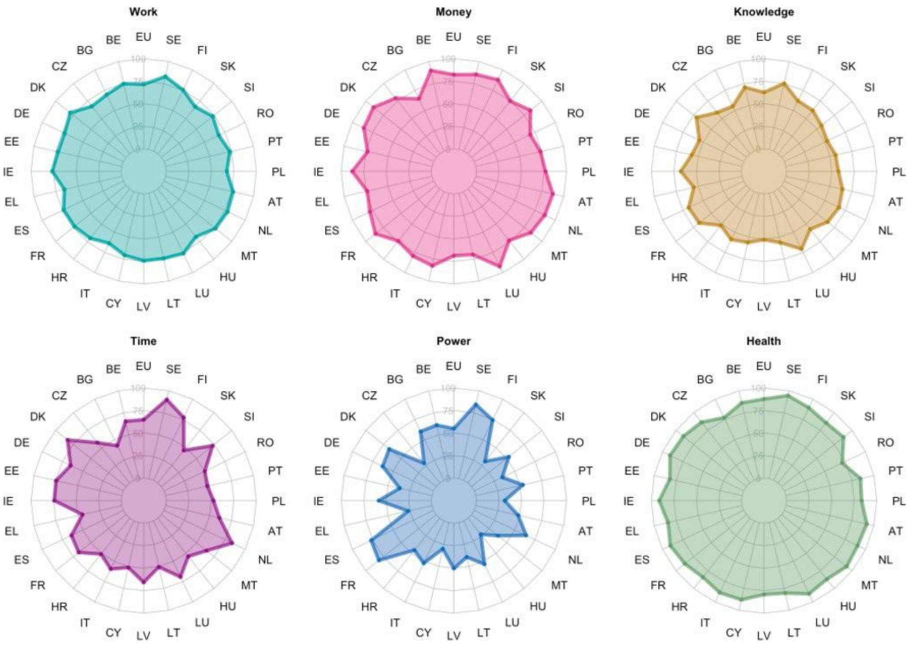
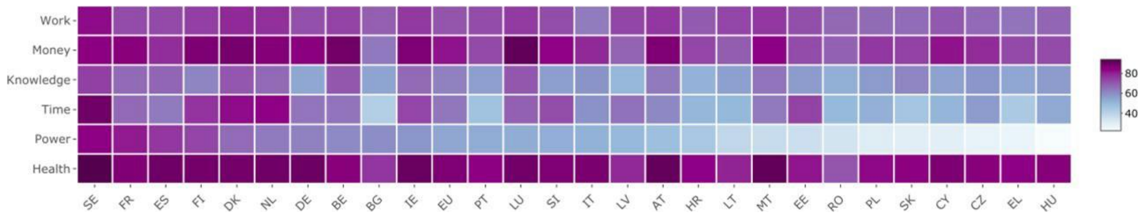
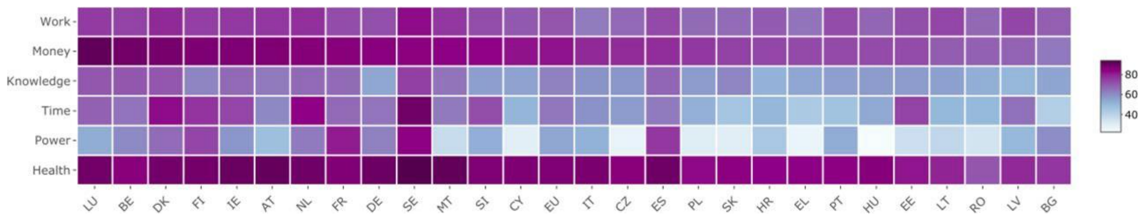


Figure 10. Individual radar charts for each domain contributing to the GEI (2021). Source: own elaboration.

We can also explore heatmaps, as seen in Figure 10a, where countries on the horizontal axis are arranged according to the value in the Power domain. If the values are rearranged according to another domain, the graph will look quite different, as seen in Figure 10b, where countries are arranged according to Money. This is similar for other graphs; in particular, the polygons in radar charts can look quite different after rearranging the countries. Therefore, it seems that the order in which we represent the countries is of relevance. But can we find a good convention to order the countries? Well, the fact is that the countries have a natural order in terms of geographical location, so let's explore this option.



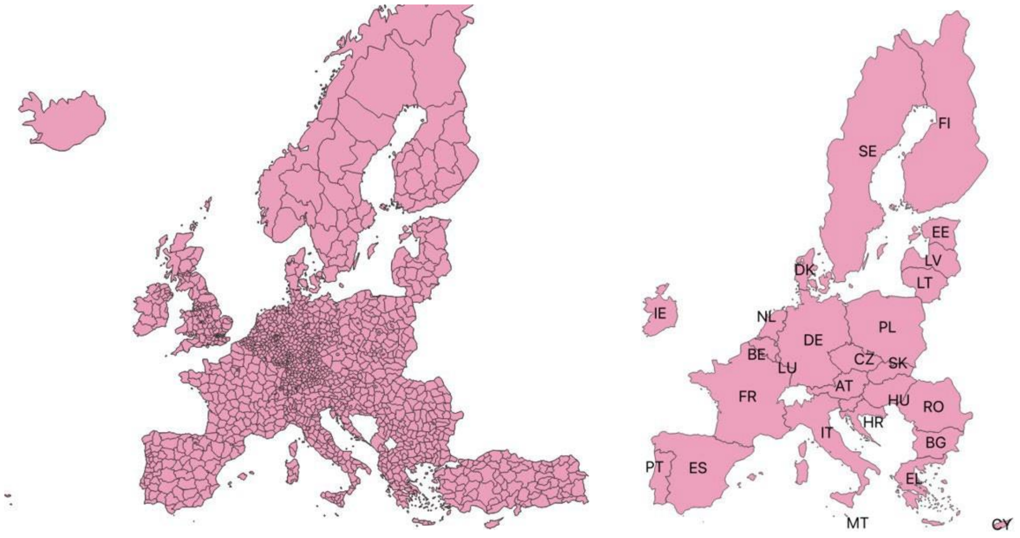
(a)



(b)

Figure 11. Heatmaps for the domains contributing to the GEI (2021) indicator: (a) heatmap ordered according to the values in the Power domain; (b) heatmap ordered according to the values in the Money domain. Source: own elaboration.

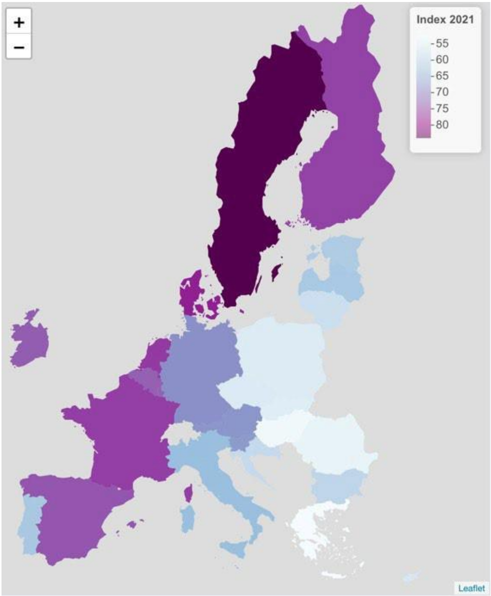
We can represent data in the form of choropleth maps, which display divided geographical areas or regions that are colored in relation to a numeric variable. For that, we first need to consider the geographical shape of each country, so we need a map with the countries' boundaries, which is not available in our original dataset. We can download such a map of the European countries from Eurostat (*NUTS - GISCO - Eurostat*, 2022), for instance. But this map has smaller boundaries than the ones we are interested in and involves more countries that are not part of the EU27, as shown in Figure 12a. Doing a little bit of spatial analysis in a Geographical Information System, such as QGIS, we can join those polygons that share the same country name and remove countries that are not part of the EU27. The result is shown in Figure 12b, where country labels are also depicted. Now we can call this map from our R script, merge it with our dataset, and produce a choropleth map, as shown in Figure 13a.



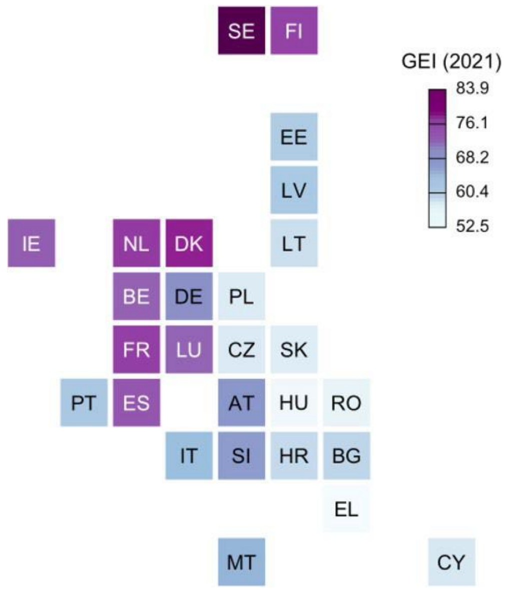
(a) (b)

Figure 12. Maps of Europe: (a) map provided by (*NUTS - GISCO - Eurostat*, 2022); (b) map with the EU27 countries. Source: own elaboration.

The map in Figure 13a shows the GEI (2021) indicator embedded in the color aesthetic. We could use the same technique to map each of the domains and produce a total of six maps. Before doing that, it is worth mentioning the problem of small areas, as is the case for MT or LU, which are difficult to see on the map. In interactive maps, where one can zoom in/out, this fact does not pose such a big problem, but if the map has a fixed size, as in the example shown here, this is indeed an issue to consider. As the size of a country does, in principle, not influence the indicators that we are analysing, we can, instead of using real boundaries, use other conventions to build a cartogram. A simple way is to replace countries' boundary by rectangles, while trying to arrange them according to their topological relationships (e.g. to the left of, on top of, etc.); this is, however, not fully possible in our example, so we must be aware of this limitation. Such a cartogram can be seen in Figure 13b, where the real size of the countries is not an issue when discerning the color; now we can clearly see the values corresponding to MT and LU.



(a)



(b)

Figure 13. Maps of the GEI (2021) value for each country: (a) choropleth map; (b) cartogram. Source: own elaboration.

Following the example of the cartogram in Figure 13b, we can derive one of them for each of the domains, as displayed in Figure 14. Note that, in these representations, the same values are used for the color ramp in all cases; as seen in the legend, it ranges from 22.9 to 94.6, which are the min and max values of all the considered variables. This scale is the same as the one considered in Figure 11, so a direct comparison between these two graphs is possible. Similarly, for the maps in Figure 13, the color ramp is fitted to the min and max values of the unique variable that is considered, GEI (2021), corresponding to 52.5 and 83.9. In doing so, it is easier to depict the greatest variations between the data, for that variable. But, why not use independent ranges for each of the cartograms in Figure 14? The answer is to allow easy comparisons between them. For instance, in Figure 14, it can easily be seen that Power is the domain with the greatest gender gap, while Health involves the least gender gap.

Figure 14. Cartograms for the six domains, with the same color ramp range (min=22.9, max=94.6). Source: own elaboration.

However, using different ranges for individual cartograms can be of interest when comparing the values of the different countries within a single domain. Such a solution is seen in Figure 15. Here, we can see, for instance, that for the domain Money, LU has a clear high value in comparison to the rest of countries, which was not as evident in Figure 14.

Figure 15. Cartograms for the six domains, with different color ramps ranges for each domain.

Source: own elaboration.

3. Future actions on visualising data

The amount and complexity of data that humans have access to through the Internet has increased enormously in recent years (Protopsaltis et al., 2020). For instance, public agencies and administrations provide open data to meet the demands of citizens for agile and flexible services, promoting transparency and citizen participation, optimising their resources and improving their efficiency (*How to promote improvements in public administration using open data | datos.gob.es*, 2021). Data visualisation has proven effective for presenting essential information and driving complex analysis with vast amounts of data (Keim et al., 2013). A clear example is the Gender Equality Index, which gives more visibility to areas that need improvement and ultimately supports policy makers to design more effective gender equality measures (*Gender Equality Index | 2021*, 2021).

However, as seen in the previous section, for a given dataset, finding out which visualisation is optimal for answering specific questions can be challenging, as the process of synthesising data into visual representations involves a set of human decisions that takes time and requires expertise.

Indeed, visualising data is not just a matter of having, for example, an Excel table and clicking one button to derive a beautiful chart; one needs to know what the best graphs (bar plot, pie chart, etc.) and aesthetics (color, shape, size, etc.) are to represent such data. But what are the factors that might condition such choices? The answer is: many. First, it is relevant to know some basics about data visualisation, such as how some aesthetics can represent both continuous and discrete data (position, size, line width, color), while others can usually only represent discrete data (shape, line type) (Wilke, 2019). Also, it is important to know who the people are that the graphs are directed to, as, depending on their expertise, they might be able to interpret more complex visual representations of data (e.g. violin plots). The medium also conditions the visualisation. It is not the same to produce a chart to be printed on a paper medium as it is to produce a digital graph to be visualised on a device. Firstly, because paper creates a physical restriction on the size that a screen might not have because of the possibility of zooming; and, secondly, because a printed chart is static, while a digital chart might be dynamic, interactive, or both. The consideration of 2D vs. 3D is another factor, and linked with this, there are a variety of new technologies and interaction paradigms that remain quite unexplored, such as virtual reality (VR) or augmented reality (AR).

Because of these reasons, currently it is difficult for designers to anticipate and test all possible combinations of interactive inputs which a visualisation might receive (Walny et al., 2019). For instance, (Hissitt, 2020) points out that there is an increased need for technical skills to first understand and translate the data, and then create visualisations around the results. To sum up, in order to represent data in visual forms to transfer knowledge to humans, data visualisation needs to be rediscovered to fit the demands of current and future data volumes and heterogeneity of data, technologies and interaction paradigms, taking into consideration human factors.

4. Conclusions

In this chapter, we have seen how data visualisation can aid in understanding the gender gap, specifically at the EU level. We have explored different charts to understand the Gender Equality Index (GEI), which is calculated taking in consideration six different domains. From the derived charts, some of the conclusions that can be derived are:

- Figure 3: For the GEI (2021), there are 10 countries above the EU average. From these countries, SE seems to have a significantly greater GEI – at least five points above the next one, which is SI. On the other side is EL, with the smallest GEI value, 15 points below the EU average.
- Figures 4, 5 and 6: Comparing the years 2013 and 2021, all countries have increased their GEI value. But, taking in consideration the rest of the years (2015, 2017, 2019 and 2020), we can see that there are fluctuations, so between consecutive years there was not always an increment.
- Figures 7, 8, 9 and 10: Overall, the domain with the greatest gender gap is Power, and the one with the least gender gap is Health. HU presents the greatest difference between these two domains. SE is the only country with all the domains above 70 points, and also has little variation between all of them. FR and DK have also small variations among the six domains, in comparison to the rest of countries.

- Figure 13: The countries with the greatest gender gap for the year 2021, as measured by the GEI indicator, are located in the central- and south-eastern part of the EU with the exception of PT, that also has a low GEI value.
- Figure 14: Power is the domain with the greatest gender gap, while Heath is the one with the smallest gender gap. Also, Health seems to have small variations between countries, while Power seems to have great differences. Related to the geographical distribution, all domains seem aligned with the GEI index, as shown in Figure 13.
- Figure 15: For each specific domain, the pattern of the gender gap is aligned with the one shown in Figure 13 with a few exceptions, such as the Knowledge domain for FI, which presents a lower value in comparison to the rest while being located to the North of the EU. SE has the first position in all the domains out of Money, where Luxemburg (LU) stands out.

These conclusions are richer that the ones that can be inferred from the charts available in (*Gender Equality Index | 2021*, 2021); thus, we can state that, in this chapter, we have contributed to the understanding of the gender gap at the European Union.

5. References

Andrienko, G., Andrienko, N., Drucker, S., Fekete, J.-D., Fisher, D., Idreos, S., Kraska, T., Li, G., Ma, K.-L., Mackinlay, J., Oulasvirta, A., Schreck, T., Schmann, H., Stonebraker, M., Auber, D., Bikakis, N., Chrysanthis, P., Papastefanatos, G., & Sharaf, M. (2020, marzo 30). *Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications*. BigVis 2020 - 3rd International Workshop on Big Data Visual Exploration and Analytics. https://hal.inria.fr/hal-02568845

Bikakis, N. (2018). Big Data Visualization Tools. *arXiv:1801.08336 [cs]*. http://arxiv.org/abs/1801.08336

Chan, W. W.-Y. (2006). *A Survey on Multivariate Data Visualization*. Hong Kong University of Science and Technology.

Gender Equality Index | 2021. (2021). European Institute for Gender Equality. https://eige.europa.eu/gender-equality-index/2021

Gender Equality Index 2017: Methodological Report. (2017). European Institute for Gender Equality. https://eige.europa.eu/publications/gender-equality-index-2017-methodological-report

Gender Equality Index—Dataset. (2022). https://eige.europa.eu/sites/all/modules/custom/eige_gei/app/content/downloads/gender-equality-index-2013-2015-2017-2019-2020-2021.xlsx

Gender gap. (2022). European Institute for Gender Equality. https://eige.europa.eu/thesaurus/terms/1178

Healy, Y. H. and C. (2022). *The Radar chart and its caveats*. https://www.data-to-viz.com/caveat/www.data-to-viz.com/caveat/spider.html

Hissitt, J. (2020). *The 5 Biggest Challenges Facing Data Visualization | Articles | Chief Data Officer*. https://channels.theinnovationenterprise.com/articles/the-5-biggest-challenges-facing-data-visualization

Holtz, Y. (2022). *The R Graph Gallery – Help and inspiration for R charts*. The R Graph Gallery. https://r-graph-gallery.com/index.html

How data can accelerate equality. (2022). https://blogs.worldbank.org/opendata/how-data- can-accelerate-equality

How to promote improvements in public administration using open data | datos.gob.es. (2021). https://datos.gob.es/es/node/8909

Ke m, D., Qu, H., & Ma, K.-L. (2013). Big-Data Visualization. *IEEE Computer Graphics and Applications*, *33*(4), 20-21. https://doi.org/10.1109/MCG.2013.54

N' TS - GISCO - Eurostat. (2022). https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

P otopsaltis, A., Sarigiannidis, P., Margounakis, D., & Lytos, A. (2020). Data visualization in internet of things: Tools, methodologies, and challenges. *Proceedings of the 15th International Conference on Availability, Reliability and Security*, 1-11. https://doi.org/10.1145/3407023.3409228

Sposto, L. (2022). *The Physical Traits that Define Men and Women in Literature*. The Pudding. https://pudding.cool/2020/07/gendered-descriptions

Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., & Willett, W. (2019). Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *arXiv:1908.00192 [cs]*. http://arxiv.org/abs/1908.00192

W lke, C. O. (2019). *Fundamentals of Data Visualization*. https://clauswilke.com/dataviz/

Yau, N. (2014a). *Decline of women in computer science | FlowingData*. https://flowingdata.com/2014/10/30/decline-of-women-in-computer-science/

Yau, N. (2014e). *PhD gender gaps around the world | FlowingData*. https://flowingdata.com/2014/09/18/phd-gender-gaps-around-the- world/

Yau, N. (2022b). *FlowingData | Data Visualization and Statistics*. https://flowingdata.com/ Yau, N. (2022c). *Gender | FlowingData*. https://flowingdata.com/tag/gender/

Yau, N. (2022d). *Most Female and Male Occupations Since 1950 | FlowingData*. https://flowingdata.com/2017/09/11/most-female-and-male-occupations-since-1950/

Zinovyev, A. (2010). Data visualization in political and social sciences. *arXiv:1008.1188 [cs]*. http://arxiv.org/abs/1008.1188