Presentation of Tables, Graphs and Maps

Alex Thomson

December 2020

Access our full collection of guides, materials and more at:
stats4sd.org/resources

This document should be cited as:

Thomson, A (2020). *Presentation of Tables, Graphs and Maps.* Statistics for Sustainable Development, Reading, UK

Stats4SD

Contents

Introduction

Visualising data is an essential part of communicating messages and results to any form of audience. An ineffective visualisation of data can communicate a very misleading message.

Building skills in data visualisation can help you to understand and see important results in other people's tables, graphs, and maps. This is in addition to enabling you to create informative visualisations of your own.

The aim of this document is to provide comprehensive guidance on the presentation of data in tables, graphs, and maps. This will include both general guidance and more specific advice on different types of visualisations. We intend to provide some principles of good graphical, tabular, and cartographic practice. By providing this advice, we hope to assist anyone in their future work, especially when it comes to the writing up of research results for an audience.

This guide is intended for anyone who wishes to develop their data visualisation and reporting skills. The advice presented here will be applicable to a wide variety of situations and is not specific to certain topics. Additionally, we hope that users of all ability levels will be able to take this advice to mind in their future projects and their everyday interactions with data.

This resource will start by exploring some general guidance on the presentation of data before going into more specific detail on the use of tables, graphs and maps (an increasingly popular method of presenting data). It then provides advice on ensuring your visualisations are accessible, with consideration on the use of colour.

General Guidance

Through developing your data visualisation skills, you can generate a wide variety of graphical/cartographic/tabular representations of data. This could be anything from simple bar graphs and line graphs to complicated cartograms. However, regardless of the complexity of your chosen data visualisation technique, there are certain principles that should always be followed:

- **Effectiveness**

  By effectiveness, we mean you should be ensuring that you are using the right type of visualisation for your objectives and priorities. This is the first crucial step in making sure what you produce is effective at displaying the message you intend to show. If you pick the wrong method, your visualisation will not be effective regardless of its quality.

  Maps are of course for displaying data which have some form of geographic component.

  Tables are suited for presenting structured numerical information; consider tables of means across some groups, frequencies, or some statistical information. This makes them ideal for when the message is in the specific numbers and potentially the relationship between them.

  Graphs are quite multi-purpose; there is a type of graph for almost any message you could be wanting to convey. In general, we would choose to use them for indicating trends, making broad comparisons, or showing relationships.

- ***Readability***

  All elements of your visualisation should be legible, understandable, and coherent. In a word, readable. While this largely relates to any textual elements of your visualisations, the principle is applicable to the whole visualisation.

  This includes having titles and headings which concisely explain the content. It should be informative without being overly long and confusing. The same goes for any further labels such as axis labels for graphs, column headings for tables and geographic labels on a map.

  Details to consider mentioning include measurement units, geographical coverage, time, the source of the data and any relevant statistics.

Of course, the elements of a visualisation will vary depending on what visual you produce, but they should always be easy to read and understand. You can achieve this by avoiding language beyond the scope of your target audience, providing the necessary information needed to read your visual and presenting the element in a simple and tidy manner. You will find further guidance on specific elements in each of the subsequent sections of this guide.

- **Tidiness**

A visualisation should never be cluttered. This follows on from readability, although more specifically relates to positioning and spacing of elements as well as avoiding using unnecessary elements.

This includes making sure no elements are overlapping; there should be adequate spacing between them without there being so much that it makes the visualisation look empty. This can also be described as making good use of "white space".

There is much more to be said on this topic, but these are mostly specific to the type of visualisation you are using. The general principle of ensuring your visual is neat and organised is always applicable.

- **Accessibility**

Accessibility has become an increasingly important aspect of data presentation in recent years. Ensuring good practice in accessibility will help in getting an even wider audience to see our research and use our results. The Government Statistical Service makes content accessible to those with impairments to their vision, hearing, mobility, and thinking/understanding skills. For our purposes, we are mostly concerned with visual impairments.

There are some general principles on accessibility, including making sure you explain any uncommon abbreviations, avoiding clutter and keeping information concise. However, we are focusing on the use of colour. Further guidance on this is included in the accessibility section of this document. This includes considerations of colour blindness, cultural context, and the use of saturation/hue/luminance.

- **Consistency**

This is mostly relevant when you are intending to use multiple visualisations across your report. When doing so, it is important to ensure you maintain a level of internal consistency.

This involves many aspects. For instance, if you intend to disaggregate your visuals by the levels of a variable, pay attention to the order you put these categories in. They should be kept to a logical or ascending/descending order and this order should be kept the same for the sake of consistency and readability.

The same goes for when using colours to indicate certain characteristics of the data; keep the meaning of the colours consistent.

Of course, this is also important for all the smaller details such as the font, size and face (bold/italic) of text. In essence, try to keep the formatting between visualisations as similar as possible. Generate your personal visual style and stick to it. Changing things up too much will just confuse your audience and reduce your visual's readability.

- **Informative**

A good data visualisation serves to succinctly show a message about our findings. We aim to inform our reader. Usually, it would be accompanied by some text which helps to interpret the visualisation, placing it into a wider context or providing more formal details such as the results of a relevant statistical analysis.

However, a good data visualisation should be self-explanatory and should be able to serve as a stand-alone piece. The reader should be able to understand the message without constantly referring to the text. Much of this can be accomplished by sticking to the particulars of keeping your visual tidy and readable.

Whenever creating a table, graph, or a map, you should include the source of the information from which the visualisation was created. This aids the credibility of your visualisation but also ensures a properly informed audience. An exception is when all information that is used for visualisations in a report comes from the same source. In this case, you should clearly indicate the source in advance of your visualisations.

This also means making sure that your visual is necessary in the first place. Consider the following: Can you achieve the same message with some simple text? Can a visualisation accurately demonstrate your results, or would it be distracting? Are your results too complex to visualise in isolation?

These six principles are relevant regardless of which visualisation you choose to create. In the following sections you will find guidance that is more specific to tables, graphs, and maps. While the guidance is specific to the different forms, they all tie into the central principles described here in this first section.

Considering the Message

Two of the above principles go beyond the specifics of what you put into your data visualisations: effectiveness and being informative. Consideration of these two principles does not start when you plot your variables. They are principles which should guide your entire research process, including presentation.

This guide will make regular reference to considering what is appropriate for your message, your results and your purpose. For your data visualisations to be effective and informative, you need to think hard about the message you want them to convey. This will often come back to an original research question. These research questions should always be guiding you in the creation of data visualisations. Effective data presentation needs to have something to say, and what it says should be relevant.

Consider this as a process:

1. We start with our research questions that we want to help answer through our research.
2. We can break these up and consider how we will answer them. What are going to be the key points we will need to investigate to answer these questions?

   For example, say we want to research the prevalence of a disease across areas within a country. We can decide that we are going to need to make points about the overall prevalence, the geographical variation, the explanations, compounding variables. We could look at these as the building blocks of our messages. Our messages are what we want people to remember and they will all stack up to help answer our bigger questions.
3. After this, we conduct our analysis and pick out our key findings. These key findings will similarly be informed by our existing research questions and pre-conceived ideas about what our messages will be. However, they should always be flexible; an unexpected result should not be ignored.
4. We now need to update our messages based on what we have observed. Our messages should always strive to be important, relevant, and interesting. Also consider novelty; repeating a message we have heard many times over and over will not result in a very interesting data visualisation.
5. These updated messages and key findings will inform the creation of our presentable data visualisations. These visualisations along with our messages help to answer our initial research questions.

Therefore, we think about being effective and informative throughout the research process. If your messages and questions are not effective or informative then you cannot expect your visualisations to be.

Bringing effectiveness and informativeness into your data visualisations requires careful consideration of the messages you have drafted. This leads to questions you will need to ask of yourself, including:

- **What variables should I include?**

  You should not be including more variables than are necessary. Think about the specifics of your intended message and only include the variables which are relevant and necessary for effectively showing this message. You should also avoid including variables which are uninformative. If adding in a variable does not add any explanatory value, then drop it from your visualisation.

- **Which variables should I split by?**

  Disaggregating your findings by certain groups is a common practice. What variables you use to do this splitting should largely be informed by your research questions and messages.

  Consider the example of geographical variation of disease prevalence. Explicitly, we know we will need to look at how our results vary by geography, but we may also want to consider variables which could help to explain the geographical variation. So, we could split by levels of economic deprivation or rurality. It always comes back to keeping relevant and important.

- **Which graph/table/map should I use?**

  This involves thinking carefully about the type of message you want to show.

  Does your message mostly concern changes over time? If so, then a table probably is not suitable unless it is quite a short time frame with few points.

  A few graphical options would be suitable including line graphs, column charts, slope charts etc. From here, the choice would now be dependent on the types of variables you want to show. Concerned with totals? Then consider column charts. Concerned with averages/rates? Consider a line chart.

  Deciding on the right type of graph is a process, starting from your overall message and working down:

  1. What is the purpose of your message? (Change over time? Distribution? Spatial? Correlation? Etc.)

2. What variables are you plotting and how many? (Categorical? Continuous?)
3. How much data needs displaying? A lot or a little?
4. What are the measurement units? (Averages? Totals? Rates? Proportions?)

Considering these questions should help you to narrow down what type of data visualisation is most appropriate. There are too many possible variations to consider here, but these points should help guide your thinking process.

Tables

When thinking of data visualisations, tables may not be your first choice as they are not as visually remarkable as a graph. However, tables are a crucial tool in presenting data and results as they have the advantage of much greater specificity than graphs and are usually simple to understand. Generally, it is harder to read patterns in tables than in graphs. Therefore, graphs should be used when you want to focus on patterns, trends and relationships that do not necessarily require the exact values to be understood.

A table would therefore be more appropriate than a graph or map if:

- You are asking the audience to compare individual values directly
- You are wanting to include both the values and some derived measures such as percentages or indices. These are harder to show succinctly all together on one graph.
- You want to include summary statistics such as means or totals
- You need to show values with very different magnitudes together.
- If users may want to use the data for their own analysis or reference.

Reference tables contain extensive information for people to look up.

- They are useful for archival purposes rather than analysis.
- They should include detailed metadata about the information presented: what, where and when of the data.
- They usually appear as appendices.

Demonstration tables are probably what you think of when we mention tables for research purposes.

- They are intended to reinforce a point by showing statistics or values that can be quickly assimilated by the reader.

- They are included within the text to allow readers to follow the general argument and without having to flip back and forth to refer to the relevant information.
- It is important they are clear and well-presented, usually using reasonable approximations to keep figures to a few significant figures.
- Very large demonstration tables can be confusing and intimidating. If all the information is truly required, it should be split across multiple smaller tables.

The following guidance mostly concerns the formatting of demonstration tables although the general principles are applicable to both forms.

Reference tables however are not designed to draw attention to specific numbers, patterns, or comparisons and therefore advice on topics such as ordering of columns and rows are not especially relevant.

General Guidance

Title, column headers and labels

Titles and labels are very important to the design of a table as they help users understand what is being presented. The titles and labels make sure the table works on its own and can be read within a different context than its original presentation.

You should consider including the following information in tables within either titles, labels, headings or possible footnotes, the choice of which points depends on your data and how important the details are to understanding the information:

- Analysis units (people, households etc.)
- Types of statistics (totals, means etc.)
- Units (thousands, kg, $)
- Geographical coverage
- Time period
- Source of data
- Key quality information

The table below (1a) contains many errors, some of which we will come back to shortly. However, let us firstly address its labelling issues.

- It is not clear what the table means by diet and what the difference between the levels of diet are.
- These labels are very uninformative and fail to be specific. We should not have to look to footnotes to define the contents of a column.
- The measurement units have not been specified.
- The title does not effectively stand out compared to the rest of the table

Table 1a: Mean growth rates and water intake across four diets.

	Diet[1]		
Variable	I	II	III
Growth rate	89	145.32	97.128
Water intake	108.4	121.29	121
[1]Diet I = Control; Diet II = Alternative			

Let us go step by step and fix these labels and titles first. We have:

- Centered and put the title in bold face so that it stands out and specified a better meaning for diet.
- Renamed diet to supplement as this is a more appropriate heading.
- Directly named the supplements rather than through footnotes.
- Included the units of measurement for the variables so the numbers can be understood.

However, we can still make further improvements.

Table 1b: Mean growth rates and mean water intakes for four dietary supplements

	Supplement		
Variable	None	Lucerne	Sesbania
Growth rate (g/day)	89	145.32	97.128
Water intake ($ml/kg^{0.75}$)	108.4	121.29	121

Comparing numbers (rounding, decimal places, and alignment)

Tables will require the reader to compare numbers. If these numbers are differently rounded or contain differing levels of significant figures/decimal places, then comparing them becomes more difficult. Here are some things you can do to make this process easier:

- The same level of precision should be used within each variable. The precision can vary between them, because different measures or ranges will require different levels of precision to make an accurate comparison, but it should be consistent within each variable.
- It is best to minimise the number of decimal places such that comparisons can be effectively made without any loss of information.
    - It is uncommon to need more than 3 decimal places. If you are dealing with incredibly small precise values, then consider using scientific notation. However, bear in mind that scientific notation is harder to understand for most readers.
- Rounding larger numbers is also advisable depending on your purpose. Demonstration tables usually use suitably rounded numbers that effectively illustrate the message. Reference tables tend to use a higher level of precision as users typically require a more exact number.
- Using commas to separate large numbers can make these numbers more easily readable, although if you are presenting large numbers you should also consider standardising the numbers into thousands, millions etc.
- Generally, numbers should also be right aligned, as should the column headings. The decimal point should line up.
- Decimal numbers between 0 and 1 (or 0 and -1 if negative) should start with a 0 and not a decimal point.

With this guidance in mind, we can see that in our example we still have some issues:

- We are inconsistent on how many decimal places we use within our variables, making comparison difficult.
  This is especially true between water intake for Lucerne and Sesbania as the mismatched precision means we cannot tell which is larger.
- We have not aligned our numbers correctly as the decimal points do not match up with each other.

Table 1c: Mean growth rates and mean water intakes for four dietary supplements

	Supplement		
Variable	None	Lucerne	Sesbania
Growth rate (g/day)	89	145	97
Water intake ($ml/kg^{0.75}$)	108.4	121.3	121.0

We have helped alleviate these problems by:

- Ensuring our level of precision is consistent within our variables.
- Minimising this precision to the level required. Adding decimal places to growth rate increases the precision of the individual values but is not required for comparison. While, as two of our groups both round down to 121 for water intake, we need at least one decimal place to make a proper comparison.
- We have right aligned our numbers and column headers.

Orientation

A table's orientation can significantly affect its readability. It is much easier to compare numbers within a column than within a row. Therefore, if we intend to compare numbers across groups according to several variables, the variables should define the columns and the groups should define the rows.

This is true of both demonstration and reference tables.

In our continuously improving example, we are using variables to define our rows. This is making it more difficult to compare between the groups. Here is the updated version:

Table 1d: Mean growth rates and mean water intakes for four dietary supplements

Supplement	Growth rate (g/day)	Water intake ($ml/kg^{0.75}$)
None	89	108.4
Lucerne	145	121.3
Sesbania	97	121.0

Order of rows and columns

Another way to improve the layout is to consider the ordering of rows and columns. If there is some logical ordering to the groups, maybe because it is an ordinal variable, then you should keep them in this order. However, if there is no logical order, it is advised to order them according to the most important variable. The most important variable will depend on your data and objectives.

Additionally, in cases when one of your groups is "none" or "other", it is often sensible to put these as the bottom rows. A "none" group often serves as a useful baseline to compare all other groups against. An "other" group is usually a combination of rare instances and lacks specified information so is rarely useful to a table's overall message.

Ordering of rows and columns is generally not so important when creating reference tables, as these are less likely be used for comparisons or to spot patterns. However, keeping the rows in some form of logical order will likely still help with the table's readability.

In our example, we can keep our columns in their current order as there is no natural reason put one before the other. However, our rows are not ordered and therefore we can still make comparisons easier by ordering them by growth rate. Conveniently, this puts our "none" row last.

Table 1e: Mean growth rates and mean water intakes for four dietary supplements

Supplement	Growth rate (g/day)	Water intake ($ml/kg^{0.75}$)
Lucerne	145	121.3
Sesbania	97	121.0
None	89	108.4

Borders

Borders should be using sparingly and only when necessary. They can be used to help separate parts of a table or groups of rows. However, using them too much just makes the table look cluttered and can interrupt numerical comparisons.

Therefore, borders should be avoided within the main body of the table and there should be no vertical lines. Horizontal lines should only be used to separate out a table's header and footer from the main table body and the page itself. Horizontal borders are effective between column headers if there is a hierarchical grouping between them.

In our example, we have put borders around every single cell. This is completely unnecessary.

In table 1f, we have improved our table by removing all vertical lines. These are not needed since there is enough white space between the values and they have been properly aligned, making the columns appropriately distinct.

The only borders we have kept are those that separate the table from the page and one that splits the column headers from the table body.

Table 1f: Mean growth rates and mean water intakes for four dietary supplements

Supplement	Growth rate (g/day)	Water intake ($ml/kg^{0.75}$)
Lucerne	145	121.3
Sesbania	97	121.0
None	89	108.4

We now have a table which is tidy, readable, effective, consistent, and informative. However, there are some other areas of advice which may be useful.

Font

Be consistent with your font and ensure it is professional. It is recommended to use sans serif fonts such as Open Sans, Arial, Helvetica, Tahoma, or Verdana. Bold should only be used for titles and headings. Keep changes in font size to a minimum and avoid small fonts.

Grouping of rows and columns

Grouping is often useful to maximise the amount of information displayed while maintaining the table's effectiveness and readability. For example, you may use levels of more than one categorical variable to define your rows or you may wish to present more than one measure for a variable (such as wishing to show the mean, the sample size, and the standard deviation).

Below is an example blank table that may be created for such a purpose.

While there may be a temptation to put horizontal borders between the different groups of rows and vertical borders between the groupings of columns, this should be avoided. Instead, using white space between the groupings is a much neater alternative that effectively separates out the information and keeps focus within the groups.

Table 1g: Skeleton table example for grouping of columns and rows

	Growth rate (g/day)			Water intake ($ml/kg^{0.75}$)		
	n	mean	sd	n	mean	se
Supplement						
Lucerne	-	-	-	-	-	-
Sesbania	-	-	-	-	-	-
None	-	-	-	-	-	-
Site						
Site 1	-	-	-	-	-	-
Site 2	-	-	-			
Overall	-	-	-	-	-	-

Summary rows and columns

Summary rows and columns are quite useful for providing extra information that may be useful for interpretation. These should be placed at the bottom or right of the table unless they are the primary message of your table, in which case putting them first and then disaggregating is acceptable.

Other

Some other general tips include:

- You can use footnotes to provide additional contextual information including:
    - Source
    - Units of measurement
    - Statistical information (such as level of significance)
    - Any mitigating information that helps with interpreting the figures.
- If the table spans multiple pages, include the table's heading at the top of each page.

- Do not put the table in the middle of text. Ensure an effective and neat layout between your table and your text.

- If your table would only need two or fewer columns and a handful of rows, consider just writing this information out in text.

Statistical Information

In more formal scientific papers, it is often required to include statistical information within tables and present the results of statistical analysis. This following tips offer guidance on presenting descriptive statistics, measures of precision and identifying statistical significance in tables:

- In tables, we often want to present useful summary statistics including the mean, the sample size, and the standard deviation.
- A sample size is usually presented in its own column which can be called “n” if we are addressing a statistically literate audience.
- A standard deviation is commonly included inside parenthesis after the mean or sometimes within its own column. The choice may depend on the number of columns already being used.
    - If you require standard deviations across multiple columns of means, then giving each their own column and putting the deviation in italic should make it stand out more (see table 1g). This will also be neater and more readable than using lots of parenthesis.
- Measures of precision such as standard deviation or error are usually presented with one more decimal place than the mean, although this is not a strict rule.

Graphs

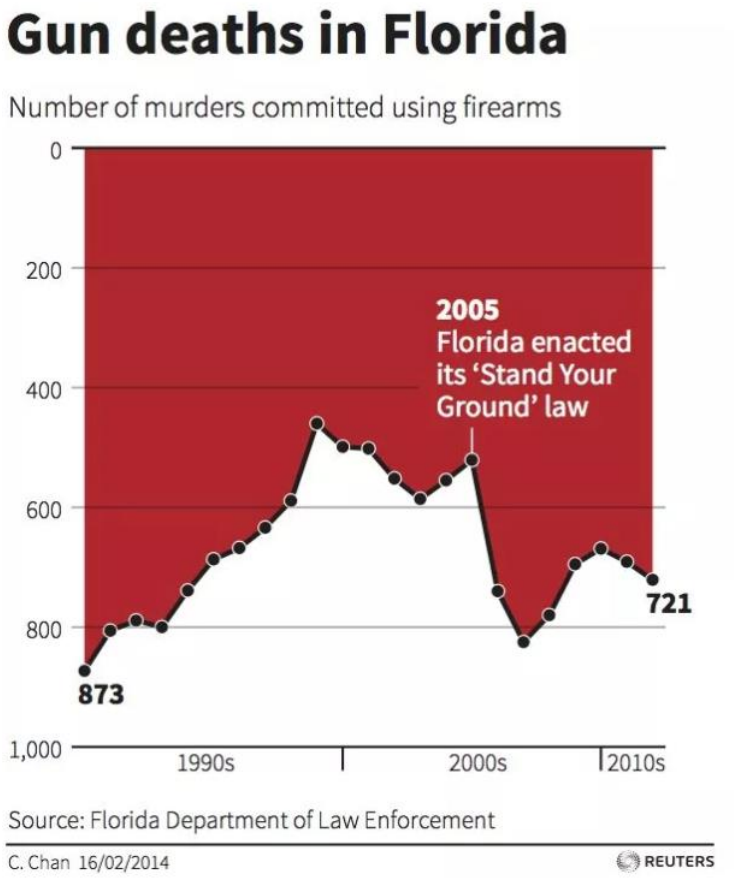
While a table can be very effective at displaying a clear and interesting message, a graph can deliver this message more succinctly and provide greater visual emphasis.

It can sometimes be difficult to spot the pattern in tables even if it is well formatted and accompanied with clear explanations. However, with a graph, we can make this pattern the focal point.

Additionally, we can visualise the entirety of our data, while tables tend to be limited to summaries or statistical results. Therefore, graphs allow us to highlight a much greater range of messages about our data. Graphs can potentially reveal insights about our data and results that would be hidden by a table.

As with all visualisations, a good graph should make sense when seen out of context, with its message intact. A poorly formatted graph can, whether intentionally or not, mislead its viewers into thinking patterns are more/less significant than they are, especially when taken out of context. Regrettably, we often see graphs being intentionally used to mislead and we should be sure to avoid committing the errors that such graphs commit.

For example, the graph below has become one of the most notorious recent examples of a misleading graph:



Generally, we would expect the y axis to start with 0 at the bottom and increase upwards, not the other way around. At first glance, it appears there was a significant drop in gun deaths after the introduction of "Stand Your Ground" laws. However, upon reading the y axis labels, we can see that the truth is the exact opposite.

Choosing the right graph

Choosing the right graph to fit your message is arguably the most important step in creating effective data visualisations. As mentioned in the section on Considering the Message, the first question to ask yourself is 'what is purpose of the message I want to show?'

A good starting point is to consider if your message focuses on any of the following purposes:

- Highlighting the distribution of a variable
- Highlighting change over time
- Highlighting a relationship between variables
- Ranking categories by a variable of interest
- Comparing values across groups
- Exploring a part to whole relationship
- Showing a geographical pattern
- Exploring deviation

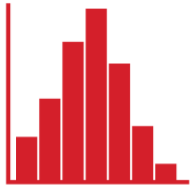
Different plot types have different purposes. Some will have the same purpose but show it in a different way or are useful for different kinds of variables. For instance, a line chart and a column chart are both suitable for showing changes over time, but the latter is more suited to totals while the former is better for rates and averages.

Some plots can have multiple purposes. For instance, a stacked bar chart can show both the magnitude of numbers comparatively and a part to whole relationship, depending on the scale used.

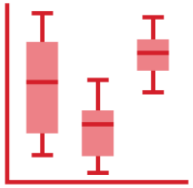
Graphs and purposes are not mutually exclusive and the list presented below is not exhaustive. Other graph types do exist, and you may have other types of messages in mind from your data analysis. This section offers an overview of some key purposes of messages as well as graph types

Distribution

The purpose of distribution graphs is to show the variation or shape of a continuous variable within your dataset to either show the frequency or distribution of all values in your data, or to highlight key measures of the distribution i.e. the range, median, maximum, and minimum. They are useful ways of highlighting lack of uniformity or equality in the data.



Histograms – The standard graph for this purpose. Uses bars to show the distribution of the data with the bars representing frequency. Keep the columns tight together to keep the users focus on the shape.



Boxplots – Summarise the data using a box to represent the inter-quartile range and lines to represent the median, the maximum, and the minimum. Outliers are often presented separately as dots. Useful for comparing **multiple distributions**.

Violin plots – Similar to boxplots except they show the entire distribution of the data, making it easier to spot specific spikes and crowding of observations. Effective with data with more **complex distributions**.

Dot plots – Very simple graph designed to show the range of data. Dots representing the minimum and maximum are connected using a line. Useful for a **non-statistically literate** audience.

Jitter plots – Visually similar to scatterplots. Each individual observation is plotted using a point with the continuous variable on one axis and a categorical on the other. A slight random variation is applied to avoid overplotting.

Some less common distribution graphs:

Dot strip plots – Show individual values in a distribution across a horizontal line. Can be difficult to read when too many values are the same

Barcode plots – like dot strip plots but uses vertical lines instead

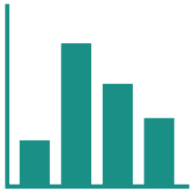
Population pyramids – Essentially back-to-back histograms. Normally reserved for breaking down a population by age and sex.

Change Over Time

In some cases, we want to show how variables have changed over time, often to emphasise trends or patterns.

Line graph –The most common way to present time series data. Very versatile, they can also be used for other forms of data, including distributions, ranks, and proportions.

- The data can be split into multiple lines by some other variable or the lines are separate variables altogether.
- It is acceptable to break the y axis on a line graph so long as this is clearly marked.



Column – Using columns to represent values within given period. Quite useful if the variable of interest is a total. Usually best for shorter single time series.

- Can be combined with a line to show relationship between an amount (the column) and a rate (line)

Slope - A simple line plot connecting just a few points. Useful for showing changing data so long as the pattern can be concisely summarised into just 2 or 3 points.

Area chart – Useful for showing how contributions to totals by different categories change over time as well as the trend of the of the total overall. However, can be difficult to read all but the bottom category.

Streamgraph – Similar to area charts but designed to showcase changes in proportions between different categories over time. Like area charts, they are visually striking, and the general patterns are easily readable but exact values are likely not of importance to the audience.

Calendar heatmap – Often each box will present a day, week, or month, depending on your time scale. Each box is coloured according to the value of another variable or frequency of observations. Useful for keeping track of recent trends or trends over a short period of time.

Magnitude

Another common reason to create a graph is to compare values, usually frequencies/totals within groups. Many of the following graphs can also be used to plot values such as comparing means/totals between different groups as well as within groups.

Bar chart – One of the most standard graphs used in data visualisation. Simple to use and can be quite versatile. Can use either vertical or horizontal bars. Horizontal bars often useful when you have lots of categories or long category labels that do not fit well under vertical bars.

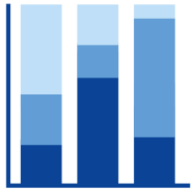
- The gap between bars should be slightly narrower than the width of a single bar
- You can add value labels to a bar chart although if you do, a table should be a valid alternative.

Clustered bar chart – allows you to show multiple series of bars by splitting the categories into subgroups. For instance, using countries across your x axis and then splitting the country bars into separate bars for male and female.

- Choice of which variable to use on the x axis and which to use as sub-groups will depend your purposes as it is easier to compare within groups and the general pattern rather than across subgroups along the x axis.

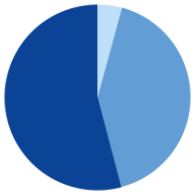
Part to Whole

It is very common to want to show a proportional breakdown for each level of a categorical variable, sometime split across another categorical variable such as location or gender.



Stacked column or bar chart – Similar to a paired bar/column chart but the groups of bars are stacked on top of each other rather than next to each other.

- Either can have the chunks equal the frequency and therefore the whole bar is equal to the total frequency.
- Alternatively, use individual chunks adding up to 100% so the chunks represent the proportion it contributes to the total.



Pie chart – Commonly used chart with segments of a circle being representative of the proportion of the whole the category accounts for.

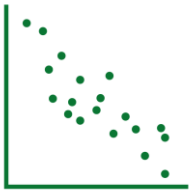
Donut chart – Same as a pie chart but with a space in the middle that can be used to provide additional information such as a total.

Grid plot – A series of boxes, usually 100, with the amount coloured in a certain colour being equal to the proportion of the total that category accounts for.

Tree plot – A square separated into multiple boxes to visualise a part-to-whole relationship which is hierarchical while also showing the frequencies of each level of the categories. Can be hard to read and designed for a more statistically or data literate audience.

Correlation

Often, we wish to demonstrate the relationship between two or more variables. It is quite simple to do this with basically any type of graph, usually by using colour to denote some extra categorical variable or creating separate lines/bars. There are however some graphical methods which can be used to prioritise the visualisation of this relationship. As previously mentioned, combining a column and a line graph is a good way to show relationship between a total and a rate over time.



Scatterplot – Standard method for demonstrating the relationship between two continuous variables with a point representing each observation.

Bubble – Like a scatterplot but the points are sized according to a third continuous variable. Probably best not to also use colour for a fourth variable as this will make your graph cluttered and hard to read.

XY heatmap – Equivalent to a contingency table with categories of one variable along the y and categories of another along the x and a square representing each possible combination. Coloured to represent frequency.

Ranking

Sometimes an item's position in an ordered list is more important than the absolute value. Often this may also be a secondary message from our data and therefore we can often use the same graphs as we use for other means.

Ordered bar/column charts – Essentially just a simple bar/column chart but make sure the bars are ordered according to the value they represent.

Lollipop chart – Similar to the above but draws more attention to the value as well as the rank.

Bump – Used to show changes in ranks over time with a line per category moving up or down, crossing over with other lines to demonstrate the ranking of the categories at each point in time. Avoid using too long a time scale.

Slope – You can also use the slope chart to show changes in ranks over time across different categories.

Deviation

Deviation graphs are used to emphasises variation from a fixed point of reference such as 0, a target value, or the overall/long-term average. Also, useful to show the results of Likert scale questions. Another way to incorporate this onto virtually any plot

Diverging bar chart – A simple bar chart but usually presented with horizontal bars with values being able to take both negative and positive values.

Diverging stacked bar chart – Usually used for presenting Likert scale results with the middle representing the number/proportions of observations in the neutral category. Negative categories are then below this group and positive categories are above.

Spine – Essentially two back-to-back bar charts with the charts split by a contrasting category. For instance, the left side of the bars representing females and the right side being male.

Spatial Data

Sometimes our data has a spatial component, and we may wish to plot the geographic relationship. If there are only a few geographic locations, a table or a bar graph may be sufficient. Consider creating a graph/map if you have lots of locations or the geographical pattern in the data is the most important message

See section in this document on map types.

General Guidance

Tidiness

A graph can easily become cluttered and messy. It can be very tempting to add more information to explain more of our data or to find more patterns. However, you should always use restraint. Focus on the key messages and do not try to show too many all at once.

A good graph should have 1 or 2 key results to show (usually the general pattern or relationship). It is acceptable for other minor messages to be visualised, so long as they do not deter attention away from the focus and do not require significant extra effort to plot.

Some key tips to keep your plots tidy include:

- Avoid using "Junk" features.

  These are extra visual features which serve little to no purpose in aiding the data visualisation other than formatting.

  - Shaded backgrounds
  - Borders
  - Boxes around legends
  - Patterns, textures, and shadows
  - 3-D graphics
  - Data makers on line charts (unless used to annotate specific key data points, in which case do not mark every single point)
  - Thick or dark gridlines

- Avoid Overplotting

  Overplotting occurs when data points or labels overlap, making it difficult to read individual points. Usually this is because there are too data points with the same/similar values, or because there are only a limited number of unique values. It can lead to graphs that are difficult to interpret or show misleading results.

  You can avoid overplotting by:

  - Reducing the size of the points
  - Showing just a subset of the data
  - Changing the plotting symbols

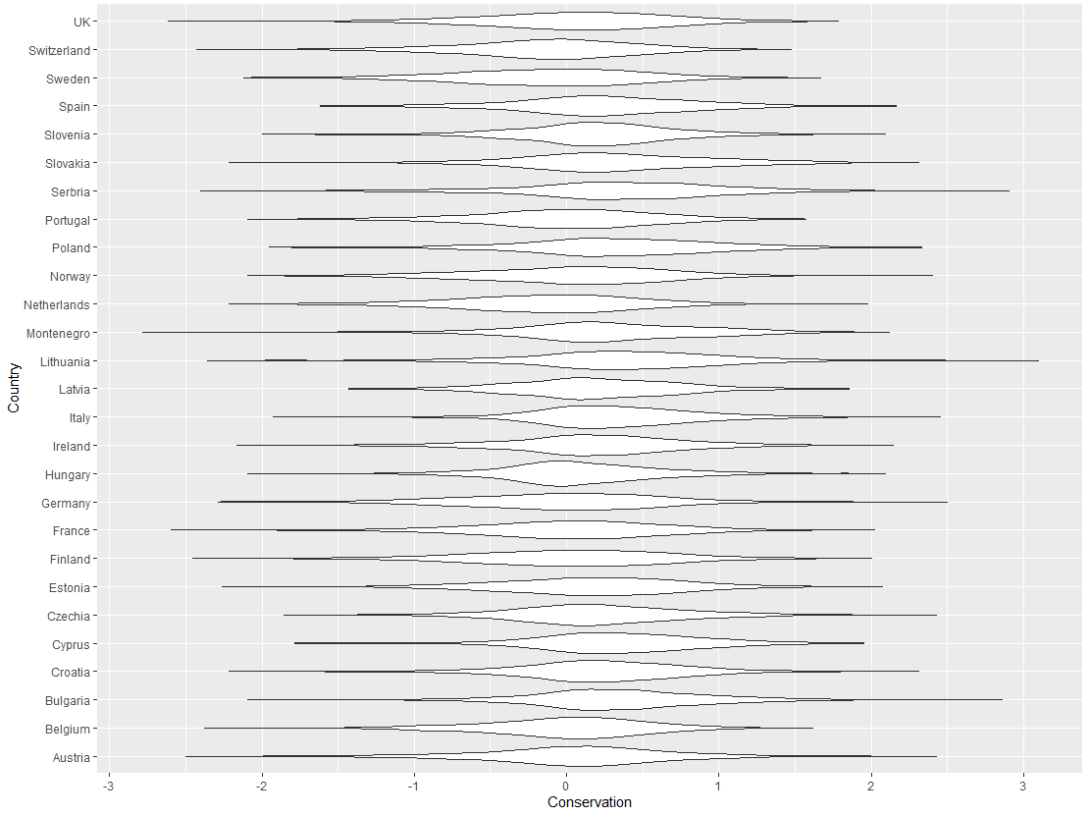
- Using transparent symbols – this is the preferred option as the areas with lots of points become clearly darker than the rest of the plot.
- Jittering – adds slight random movement of the data points around their true value. Makes it easier to read points where there are only a small number of values.
- For more information, see https://www.displayr.com/what-is-overplotting/

- Do not use too many categories

  The exact guidance on this will depend on your data and thus the type of plot you want to make, but the key points are as follows:

  - Do not plot too many lines on your line graph.
    A key advantage of a line graph is that we can plot multiple lines (either split by a variable or another variable entirely). However, we do not want to use too many because otherwise we cannot follow each line independently.
  - Similarly, with any plot with splits something by a categorical variable, avoid using too many categories. Again, this can make your graph incredibly large and make it difficult to compare specific groups.

  In the graph below, there are 27 violins placed on top of one another and we can barely tell anything about one, let alone compare them.



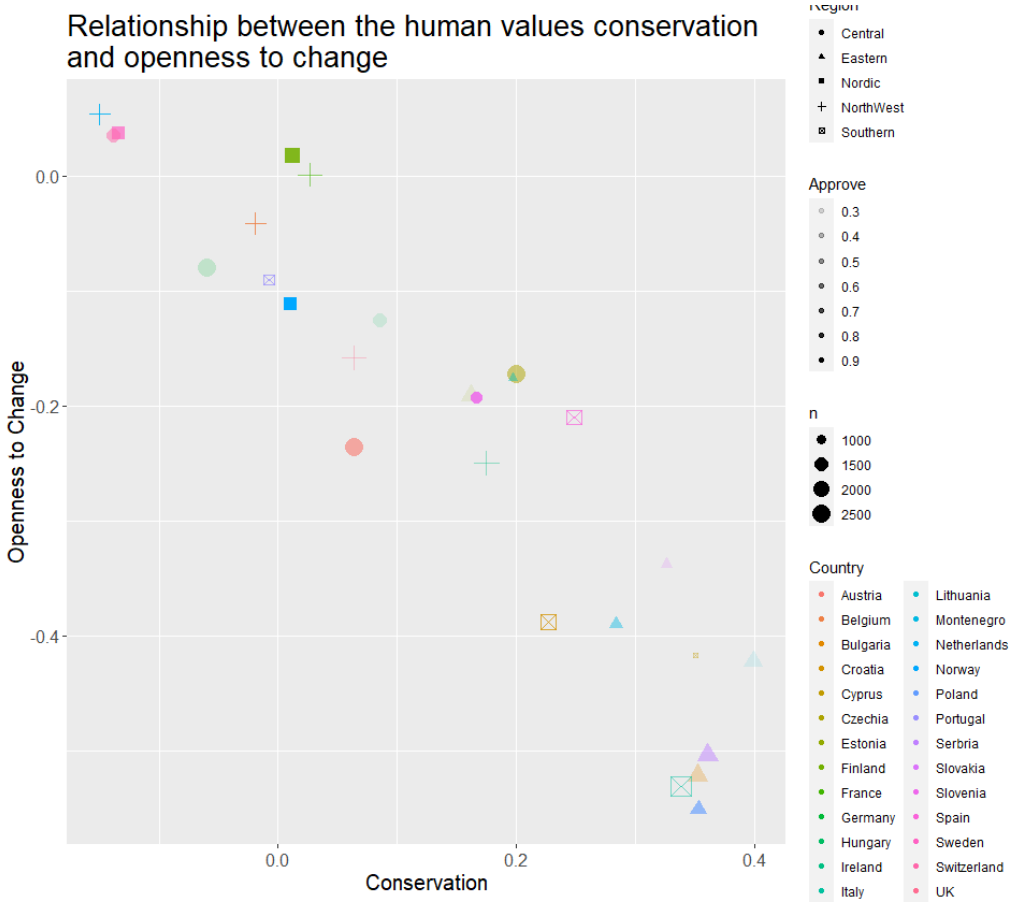
- Do not use too many variables

  When we conduct an analysis, we often want to account for as many variables as possible. Consider how many variables we may include in a model to explain some other key indicator. However, we want to avoid doing the same for our graphs; we only have so many axes and ways to visually split our data.

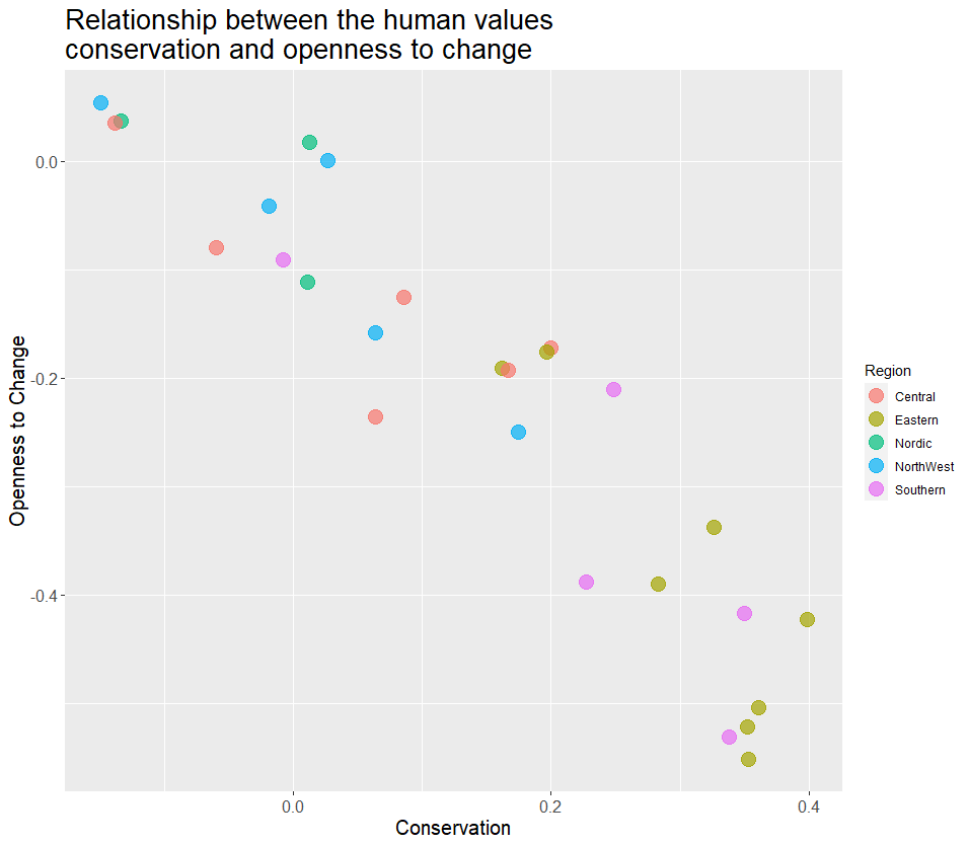
  Just because we have an x axis, a y axis (sometimes 2 of them), a z axis (not recommended), colours, sizes, shapes, line types etc. available to us for plotting data, this does not mean we should try to use them all at once

  Only use as many axes/dimensions as needed to remain informative. If you can present your information using just one dimension, then do not use a second. If can be done in 2, do not use a third. If you need more than 3, consider separating them over a series of graphs.

  Consider the graph below. So many dimensions have been used that the graph has become far too noisy and can barely be understood.



A better option would be to stick to just the 2 or 3 axes/dimensions that are most important to your messages. In this case, let's focus on the relationship between the x and the y axis and how countries differ between regions (the latter represented by colour).



Title

See previous general guidance. (pg.10)

Source

If it is likely that your graph could be shared externally or repurposed by others, include a source note in your graph so that others are properly informed about the data's origin and the context of the graph.

Scales

When plotting variables, both continuous and categorical, we need to consider how are going to go about it. What dimensions and axes are available to us? Which of these should we use? How should we construct them?

All variables will be plotted on some form of scale and getting this scale right is very important to ensuring that your graph is effective. An ineffective scale can make a graph unreadable or, at worst, misleading.

- **Axes**

  Axes can be used inappropriately to show patterns that are not there or inflate existing differences Therefore, you need to ensure that your choice of scale on your x and y axes are not misleading and authentically represent your data.

  For any axis:

  - Label the axes appropriately so that it is clear what the graph is showing

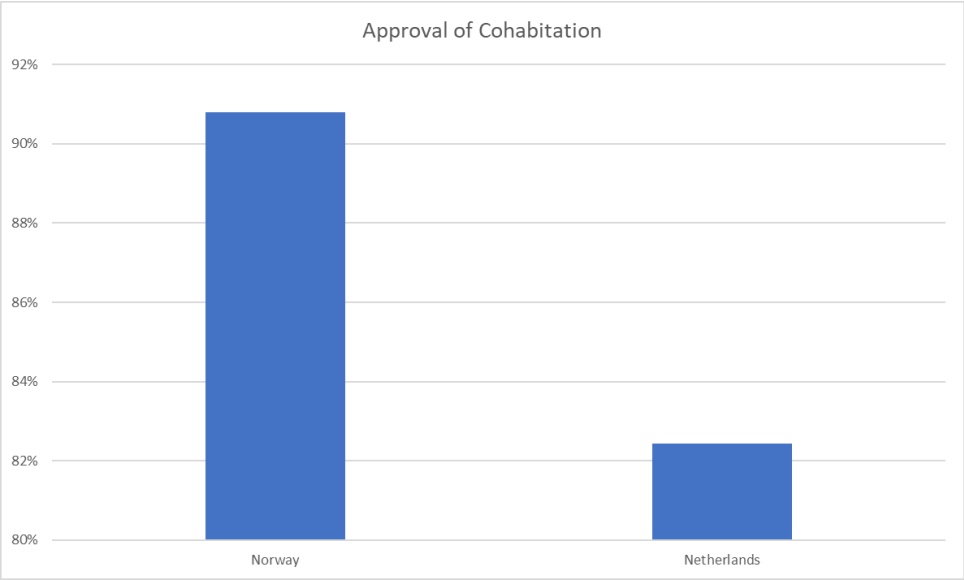
  For continuous axes:

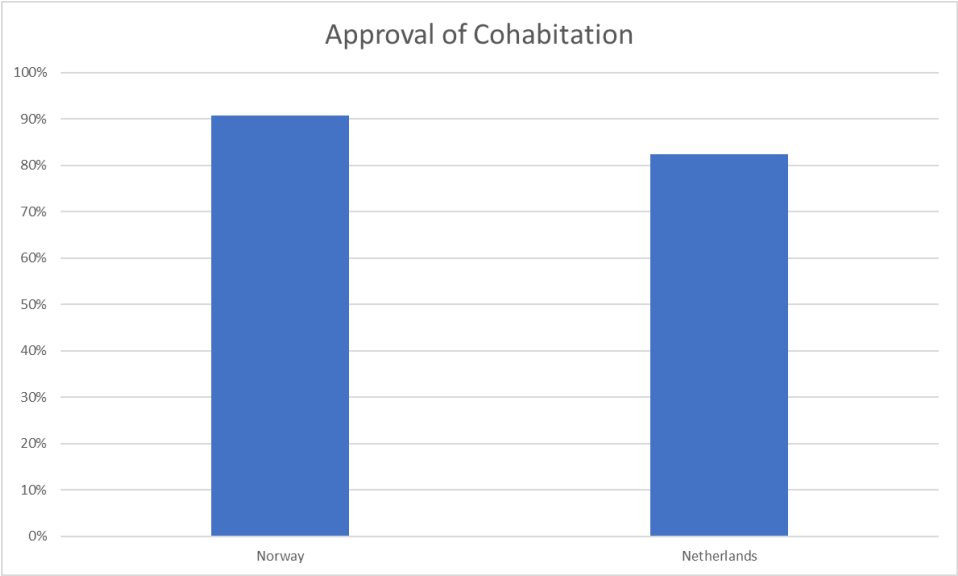
  - Use equally sized breaks along the scale
  - Label every key break, axes can become quite cluttered if every single break is labelled
  - Do not leave large gaps in your axes
    - For example, if you create a line graph with time on your axis but there is some period in the middle with no data, do not just leave a gap in your axis.
    - Keep the axis together. If the gap is particularly long you could consider an axis break.
    - You should not connect the lines either side of the missing period as this is misleading about the actual trend of your data.

  On a bar chart, you should:

  - Always start your axis at 0.
  - Never start your axis at anything else, and do not break it either
    - The height of the bars relative to the axis gives a clear indication of the magnitude of the category or value.
    - If the axis does not start at 0 or is broken, then the height becomes misleading and differences will be inaccurately visualised.

Compare the next two graphs. One starts the x axis at 80% while the other starts it at 0%. The first makes Norway look almost 4 or 5 times higher than the Netherlands when the true difference is less than 10 percentage points.





- **Dimensions**

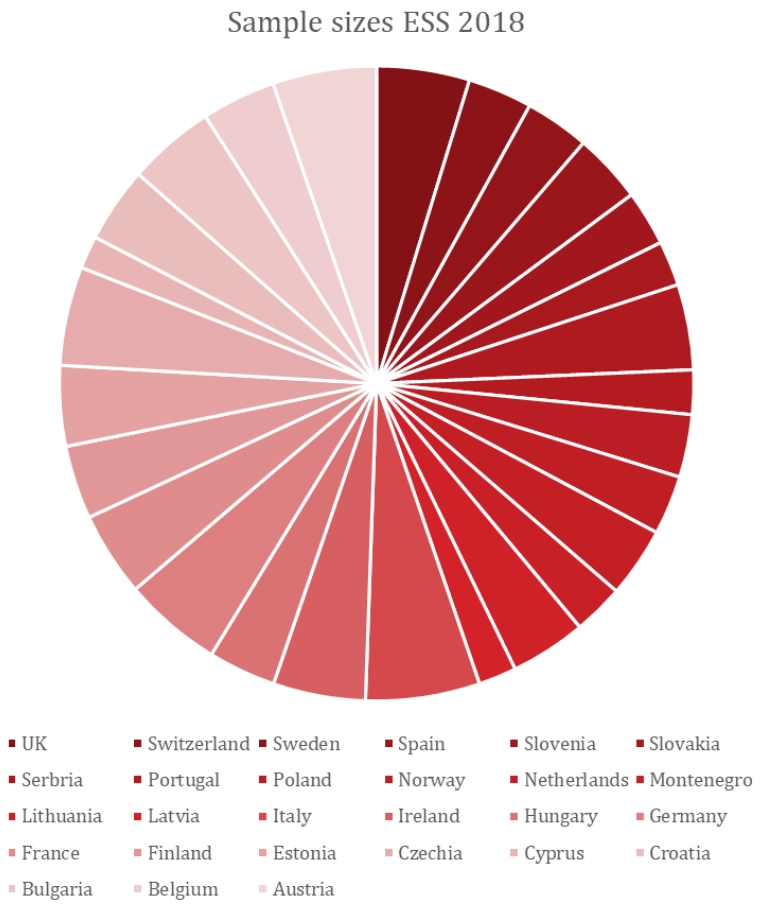
  The standard X and Y axes are of course not the only way to plot data. Other plotting dimensions also need to be used appropriately. These dimensions include, but are not limited to: colour, shape, size, area, and transparency.

  - Do not use more dimensions that are needed to be informative and readable. Using more than 3 or 4 dimensions is often pushing the limits of keeping a graph readable.

- As with axes, differences between "values" of these dimensions should be measurable and informative. For instance, detecting small differences in transparency can be difficult. Therefore, using transparency is not ideal for a continuous variable.
- Area is not necessarily a common dimension to use, but one type of graph is quite dependent on it: a pie chart. While proportional bar charts can convey the same message using area, these have the advantage of an x/y axis showing the proportions, unlike pie charts.

  While pie charts are common, they are rarely the most appropriate choice. It is easy to find examples of using too many categories or mismeasured segments, as seen below. It is also harder to read the area of a circle than the height of a bar.

  Therefore, pie charts are only useful with 4 or 5 categories and when the differences between them are quite large. A bar chart will show small differences much more clearly.



- **Orientation and ordering of your scales**

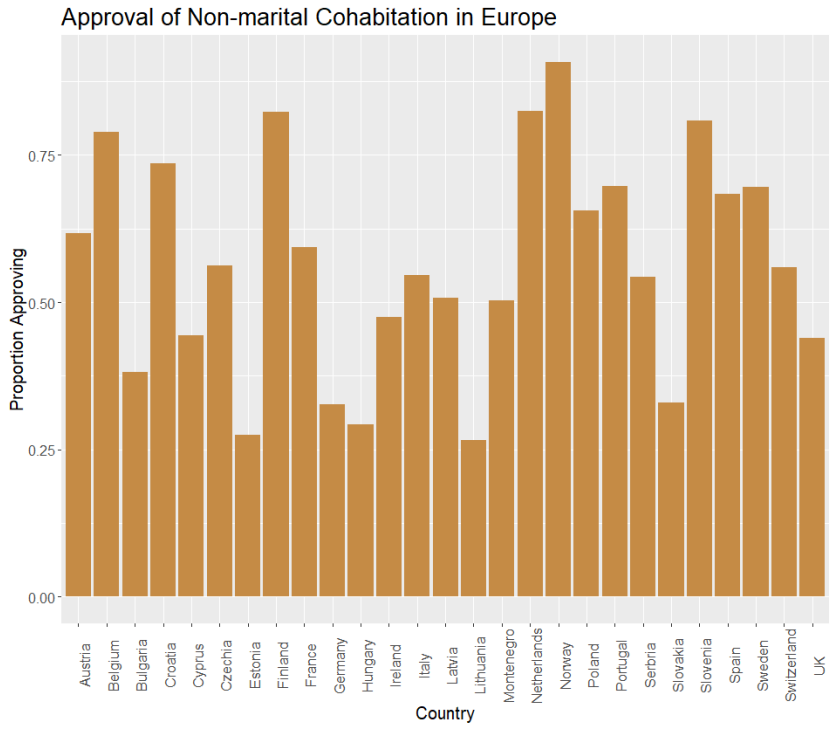
  As with tables, the orientation and ordering within your graph can make a big difference to readability and can provide additional information.

  Bar charts, box plots, violin plots etc. are often plotted with a continuous variable along the y axis and categorical along the x axis. However, there are advantages to switching this the other way around.

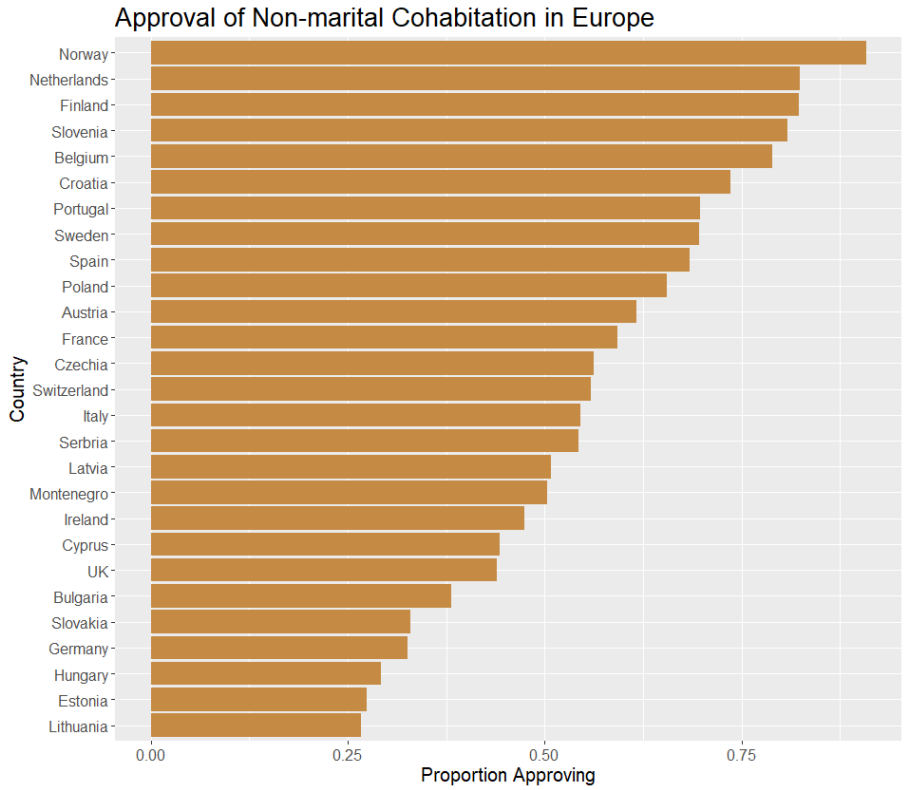
  When dealing with large numbers of categories or long category labels, flipping the x and y axis can improve the readability of your graph as you can write the long labels without overlapping or having to turn them at an angle.

  Also, consider the order of these categories the same way you would in a table. Is there a logical order? If so, then stick to it. If not, consider putting them in ascending or descending order by value. This makes the graph more informative, as now the ranking will also be displayed.

  Consider the next graph. The countries are in alphabetical order and the rotated labels are hard to read. Comparing two bars of similar heights and singling out the best and worst performers becomes challenging.



This has been fixed by flipping the axes and ordering the countries by descending order of the value on the x axis:



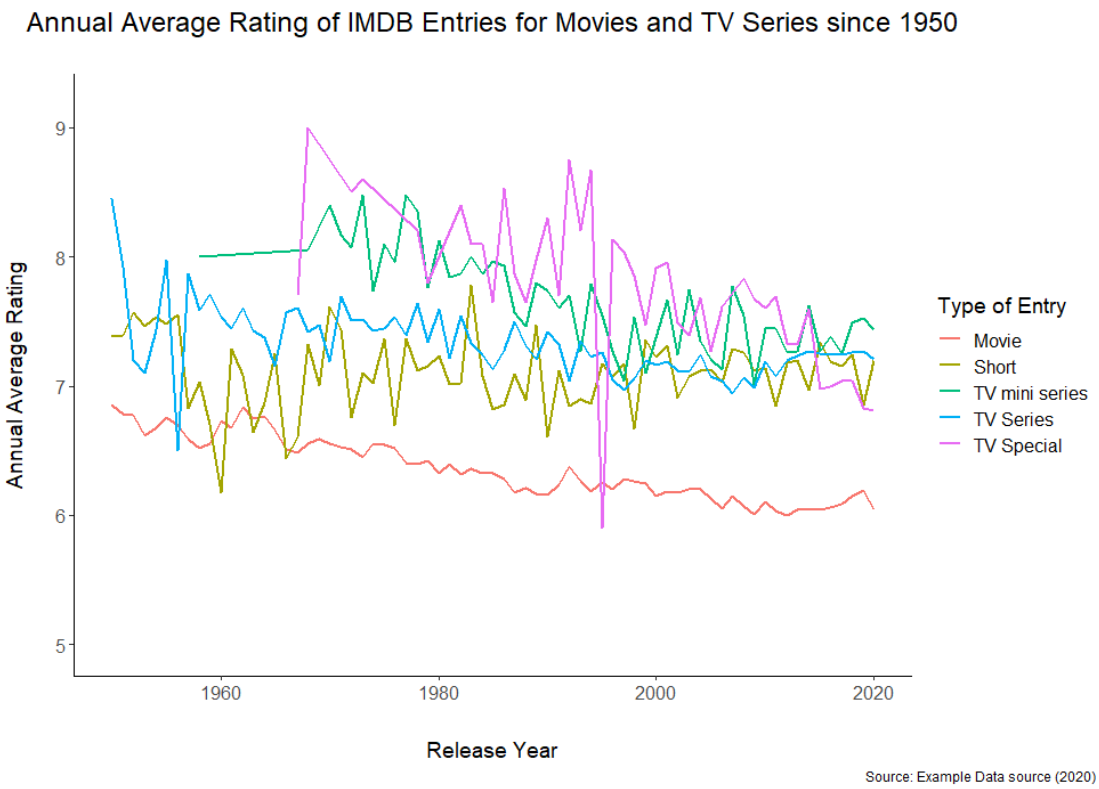
- **Facets**

  A similar concept to scales is facets. This involves splitting a graph into smaller, separate plots. These would be split by a categorical variable (maybe two). You could think of facets as another scale - just one that splits up your plotting space.

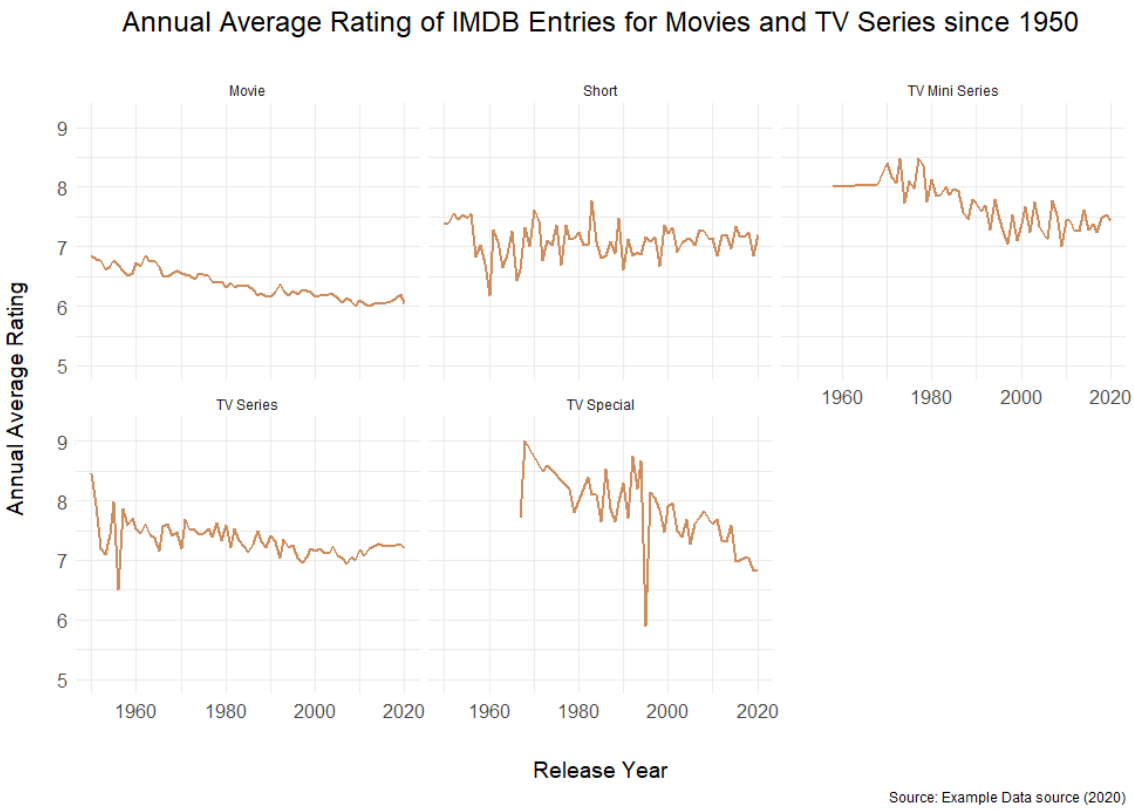
  Using facets is a good way to avoid bars/lines etc. becoming too squashed together or overlapping to the point where reading them becomes difficult. They can also provide another layer of information to your results, as you can compare across groups.

  This means you could potentially use graph types which are limited in terms of how many dimensions they offer. For instance, it is not advised to use more than the x and y axis for a histogram, but you could split the histogram over multiple smaller histograms and allow for comparability.

  In the example on the next page, there are too many lines occupying the same space, so it is difficult to pick out individual lines or make comparisons. The graph is too visually noisy.



We can instead split these lines by the type variable and plot five smaller, individual line graphs which are easier to compare and single out.



- **Legends**

  See page 46

- **Annotations**

  Placing annotations within the graph can be useful to help tell the story of your data. They are especially useful for explaining unusual or interesting data points such as outliers or spikes, as well as providing context or avoiding misinterpretation.

- **Using Colour**

  On any form of graph, colour is always an effective tool to differentiate between different plotting elements. Often, we use colour to distinguish between the levels of a categorical variable which may have also been plotted along the x or y axis. However, if the variable has been plotted on an axis already then adding colour is unnecessary and could just make your graph visually noisy.

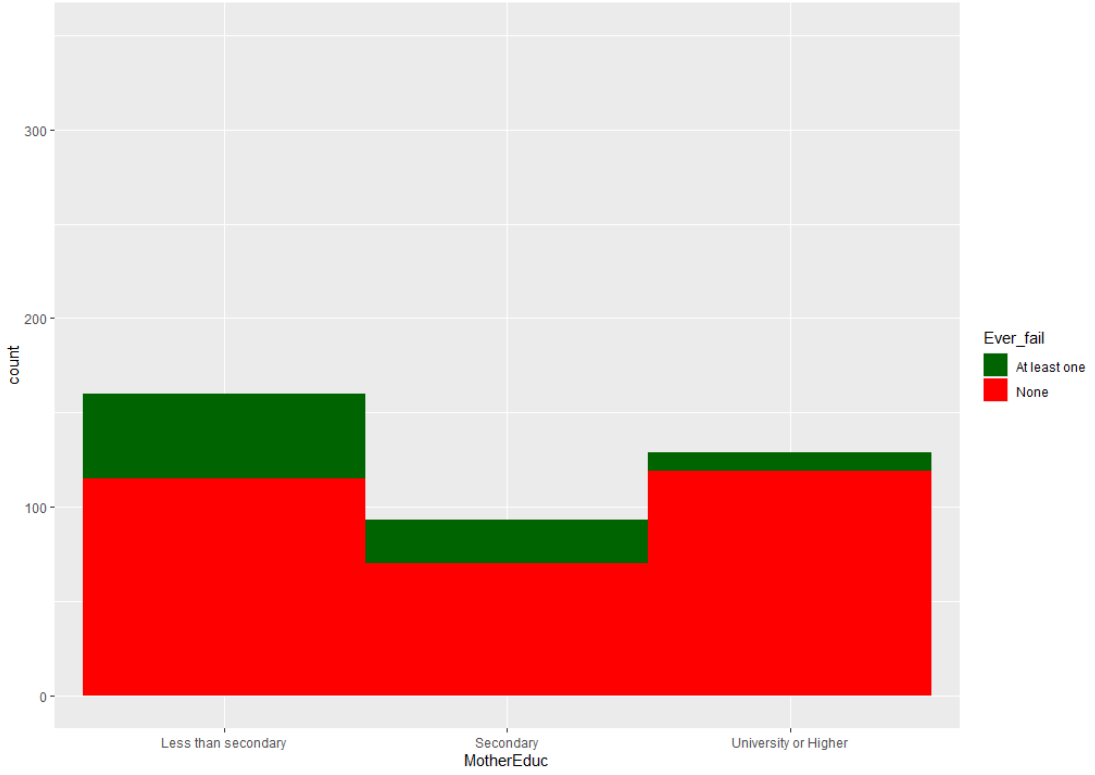
  More guidance on colour will be provided in the section on accessibility.

  - Using colour to highlight
    - You can use colour to highlight key elements to aid interpretation. This could be useful for showing quite specific messages about your data.
      For example, you may have multiple lines plotted on a line graph, but you want to make a specific point about one or two of them while the others are used to provide context or are useful but not key to your message.
    - Use a distinct colour in the foreground for the data you want to draw attention too while shading the rest in a much more muted colour, such as a pastel or shades of grey. This will mean the other elements are still visible without them taking attention away from your main point.
  - Use colour with restraint. Using colour too much will make your graph incredibly messy and create too much visual noise.
    - Also, consider how your graph might be repurposed. Colour should never be the only way meaning is added to your data. If people print the document in black and white or are colour-blind, will your message still be clear?

- Use colour to mark helpful differences in your data. Different colours suggest there is a meaning to those colours and if there is not a useful meaning, then the audience wastes time trying to figure out something that isn't there.
- If you are creating a bar graph, it is easier to compare the bars when they are the same colour. Colouring each bar differently is unnecessary.

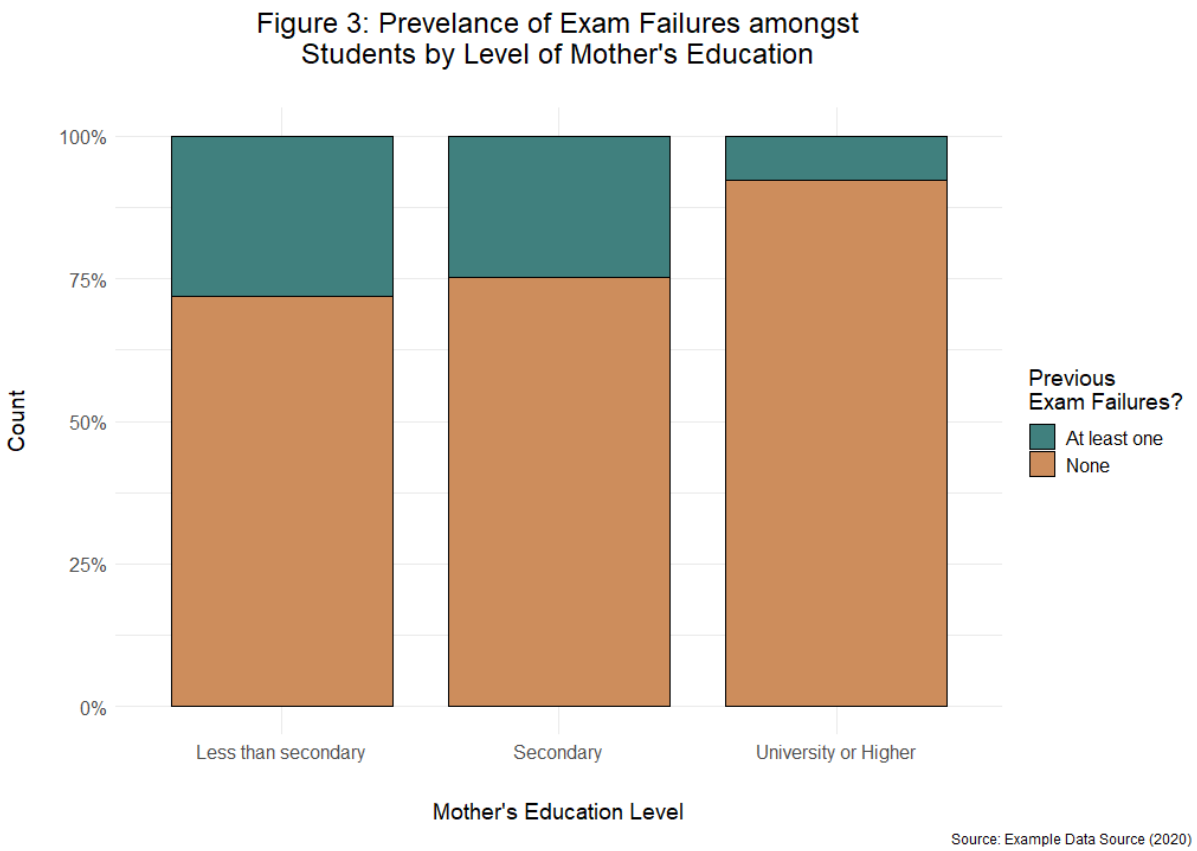
More Examples

Example 1: Poorly formatted and presented stacked bar chart



- No spacing between bars make comparisons difficult.
- No title.
- Uninformative axis labels.
- Uninformative legend.
- Red and green together is difficult to read and the colours are too saturated.
- Y axis is too large, making differences look smaller than they are.
- Harder to compare both within and across groups at the same time because of use of counts and stacked bars.
- Poor spacing between axis and axis labels.

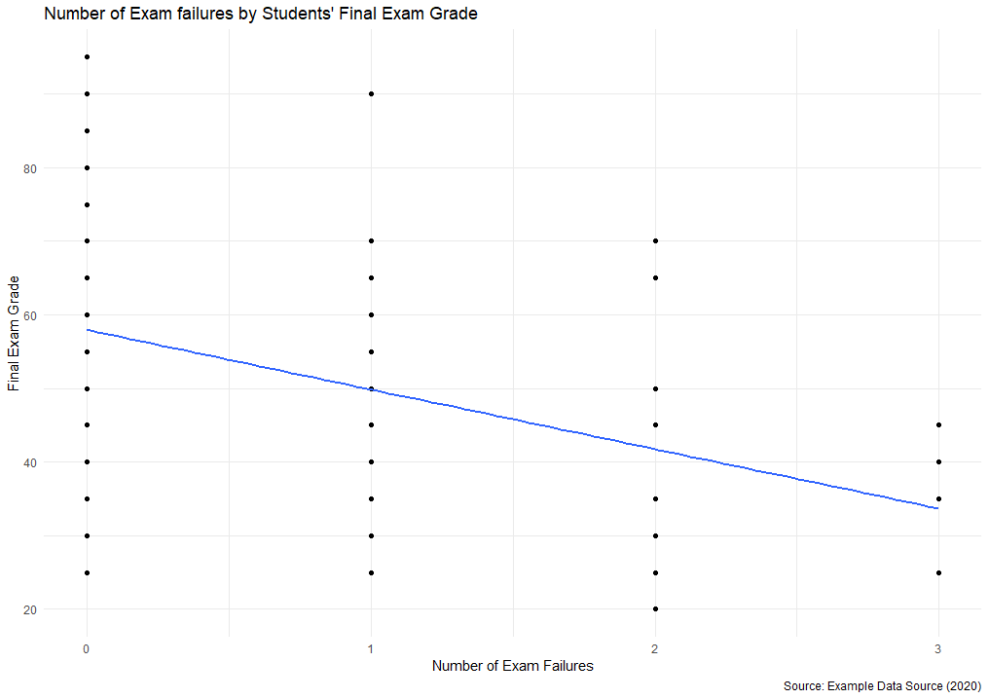
Example 2: Improved alternative



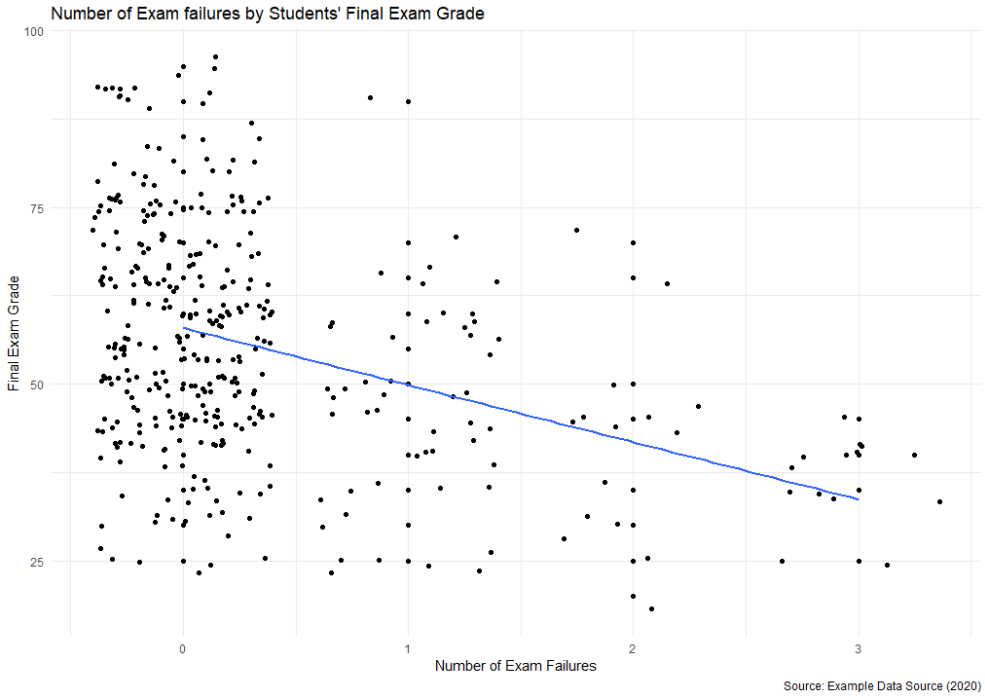
- Informative title.
- Informative axis labels.
- Adequate spacing between plot and all textual elements.
- Colour-blind friendly colour palette.
- Reduced colour saturation.
- Black borders around bars to improve differentiation between bars and colours.
- Use of percentages rather than counts, making it easier to compare both between and within groups.
- Use of proportional stacked bars rather than stacks of counts. These are easier to read when dealing with a categorical variable with two levels.
- More appropriate y axis scale.
- Increased spacing between bars to make them distinct.
- Included a source note in the bottom right of the plot.

Example 3: Overplotting through limited number of unique values

Too many values crowd at the same point because there are only 4 possible unique values along the x axis:



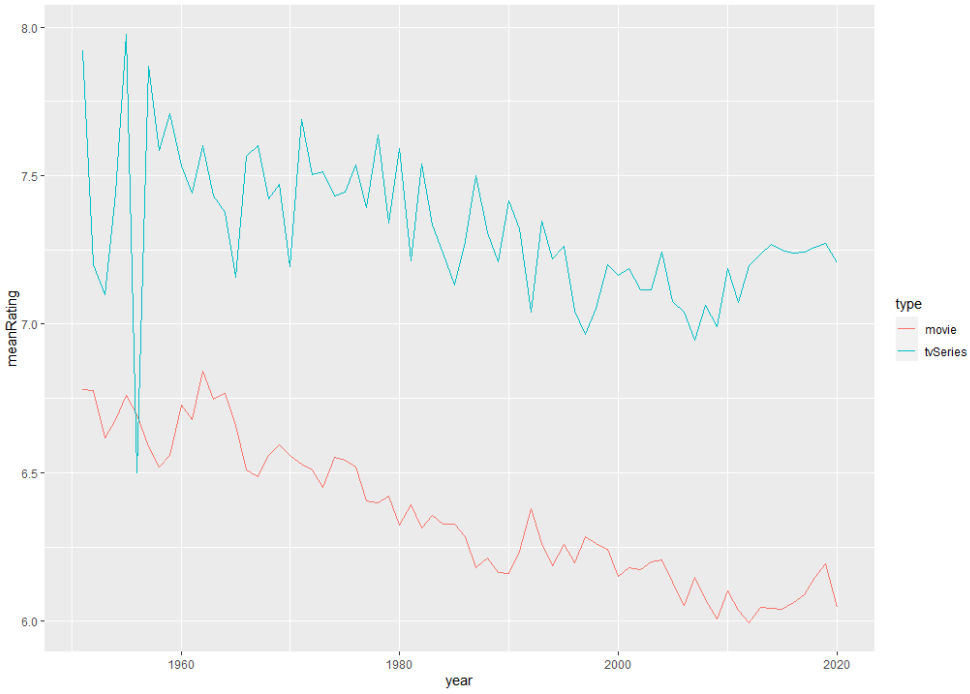
Example 4: Overplotting improved using jittered data points



This visualisation could also be improved by using alternative graph types such as a boxplot, violin plot or a series of histograms.

Examples 5 – 7: Making a line graph informative and readable

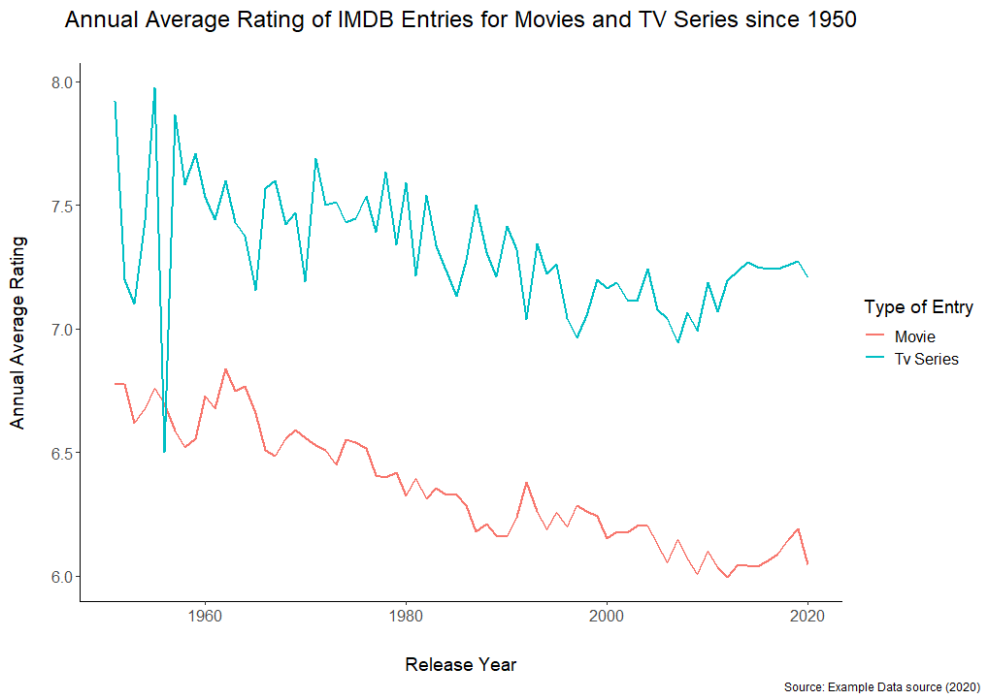
At first glance, we can still somewhat understand the patten and trend this graph is showing us, but we are missing a lot of key information that allows this graph to be understood out of context.



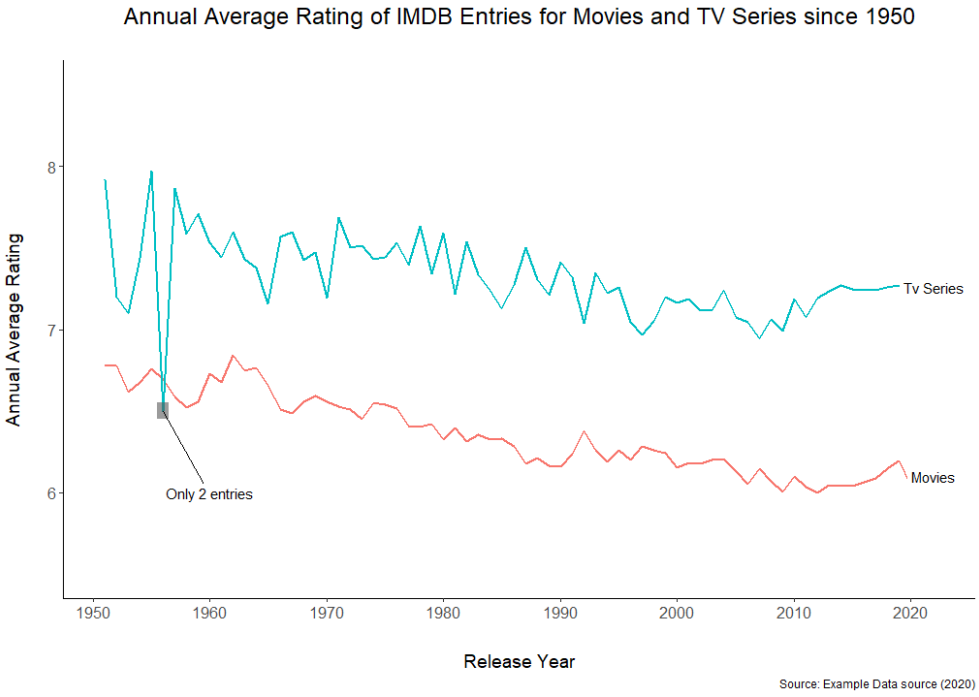
- No title
- Uninformative axis and legend labels
- Grey used as background colour
- Difficult to read text
- Lines are quite thin, making them harder to see clearly

Below, we have improved this example considerably by correcting these mistakes according to the general guidance for graphs.

- Included a source
- Informative labelling and titles
- Better use of colour thanks to a white background
- Clearer axis lines
- Increased the thickness of the lines to stand out more



This is a reasonably good example of a line graph but there are a couple more refinements we can make.



- Directly labelling the lines has given us more plotting room.
  An annotation has supplied additional contextual information that allows the reader to see that this average was based off only 2 tv shows. This helps prevent misinterpretation.

Maps

Maps have become an increasingly common sight in our daily lives, being used to show the variation in some statistic across the world or just a local area. Generally, there are 2 main kinds of maps: reference maps and thematic maps. We are more concerned with the latter, as reference maps are used to show an area's geographic extent and characteristics while thematic maps are used to display the geographic distribution/variation in each variable.

In this sense, thematic maps are just another form of graph, albeit with a few unique elements. They are useful when the geographic patterns of locations of your data points are the most important part of the message to your readers. They can be used for many different data types, including categorical and continuous data, as well as aggregate measures (counts/percentages/densities) at some level of geographical boundaries. (e.g., % unemployed within UK constituencies).

Before getting into the specifics of the maps themselves, it is worth considering a few things about your data and how this could affect the map that you produce.

- Consider the level of geographic aggregation you use for your map. Of course, this may depend on what geographic variables you have within your data. However, if you do have multiple levels available, consider which is best to show your results. Aggregation at higher levels may smooth out differences at lower levels.
- Check you have enough data for your chosen level of aggregation. If some areas have very few observations, you could be subject to extreme values making interesting but inaccurate geographic patterns.

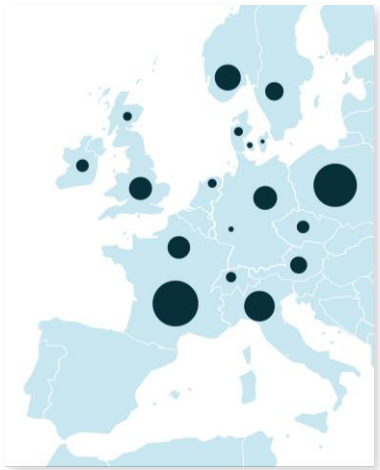
Map Types

Choropleth Maps

The most common type of thematic maps. Data is summarised at a sensible level of geographic aggregation. Values are then represented by shading in the different geographic areas in different colours. Tends not to be used for totals as these will be affected by the size of a geographic area.

Dot Maps

One of the few thematic map types which does not involve geographic aggregation of your data. However, it will require very low-level geographic information e.g., geographic coordinates or a postcode. Here each observation will be a point on the map and are a good way to provide an idea of the geographic density of your data. These tend to be used when your observations are locations or events to show how they cluster in certain areas.



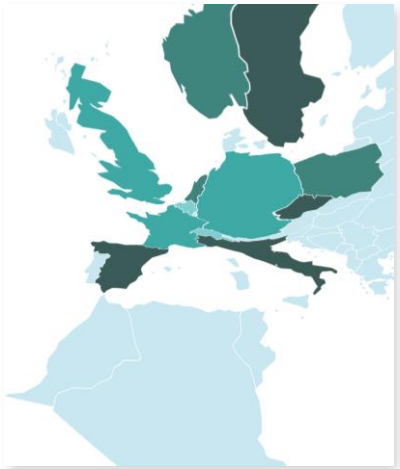
Proportional Symbol Maps

Symbols are placed in each geographic area of the map and are scaled in size according to a variable of interest. These tend to be hard to read as it can be difficult to compare relative sizes of the symbols. This can be helped by using a well-organised legend.

Cartograms

These maps **distort map geometry** to show a specific message about the data. They can be quite effective if displayed alongside a more standard choropleth map. However, being less common they can take some explanation when presented.

There are two distinct types:



Area cartograms

Map geometry is distorted to scale up areas which are high in the variable of interest and scale down areas with low values. The size of an area is proportional to its value.

Equal area cartograms

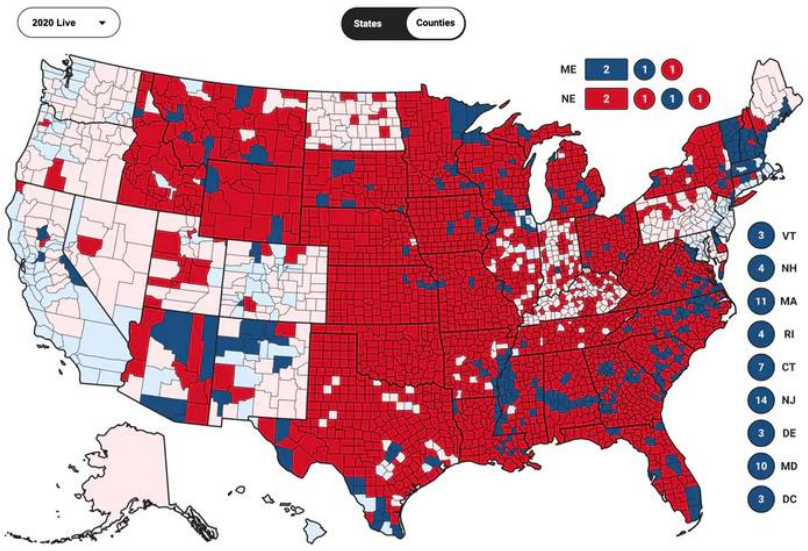
All geographic areas are changed into shapes of the same size and positioned as best as possible to approximate their geographic relationships. Useful when the areas are very different in size, but area is irrelevant to your message, as larger areas will dominate the map. Appropriate labels may be needed as it becomes difficult to identify specific areas.

Examples:

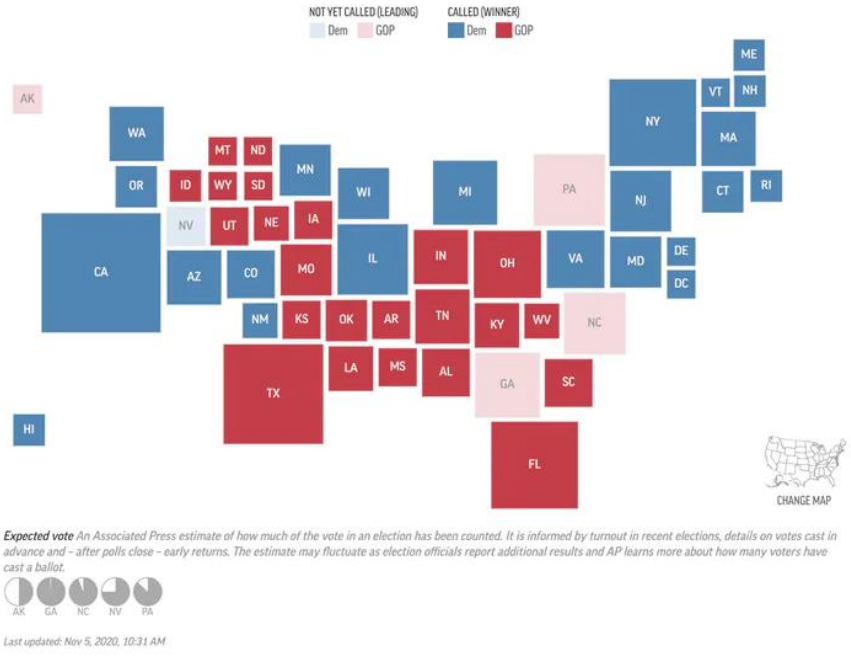
Choropleth Map vs Cartogram - The examples here show how important it is to use an appropriate style of map depending on the message you intend to show.

This is a US election choropleth map showing the results at the county level. It offers a clear message that most counties were won by Republican candidates.

However, this visual message is misleading, as election results are not decided at the county level, but rather at the state level. Land does not vote; people do.



The following cartogram offers a much more accurate picture of the US election. The cartogram changes the states into coloured boxes which are sized proportionally to the number of electoral colleges votes the states contribute. This shows that the result is much closer than the county level map suggests and that there *not* a significant lead in favour of the Republican Party.



A cartogram that appeared in The *Globe and Mail* of U.S. election results as of the morning of Nov. 5, 2020. The Globe and Mail

Using colour and categorising values

The advice presented in this section is tailored towards using colour scales in choropleth maps or cartograms and would also be true of certain types of graphs. Further advice on the use of colours will be provided in the accessibility section.

- If the variable of interest is on a continuous scale you should use different shades of the same colour to fill in the geographic areas. Light shades should signal low values and dark colours should represent high values.

- If you intend to map areas according to deviation from a given value such as the average of the whole area, you should use a diverging colour scale. For instance, using shades of red for below this value and blue for values above.

- If the variable is categorical and there is no natural order to them, use different colours for each level.

- You should not use white except to represent 0, “no data” or the centre of a diverging scheme.

Of course, using a colour scheme is only effective if you are breaking up the range of values in a meaningful way. There are multiple ways you can achieve this, and the choice will depend on your data. However, bear in mind that the choice of method will greatly alter the appearance of the map. Generally, you should limit the scale to avoid too many categories.

- **Equal intervals** – divide the data into equal class sizes. Useful with normally distributed data to highlight a small number of areas. Not useful for skewed data.
- **Quantiles** – assign an equal number of observations per class. Produces a balanced map for non-uniformly distributed data. However, classes could have vastly different ranges.
- **Standard deviation** – measure distance from the mean. Useful when the data is normally distributed.

- **Manual adjustments** – can be used to make more intuitive levels but may make comparability across a series of maps difficult. May include some target level to highlight areas below and above this value

Outliers should be highlighted as a unique category if they are important to the messages from your data. If they are not important and are simply noise, they can be hidden by widened ranges.

Map Elements

Title

As with graphs and tables, a map title should be concise but informative. It should be larger than the rest of the text elements and centred at the top of the map. For a map, the title will be required to include the "where" of the data as the geography is central to the visualisation.

Legend

A legend is likely to be even more essential to a map than it was for a graph. When creating a thematic map, we will almost always be using a colour axis or some other symbology to tell a message about our data therefore a legend will almost always be required. The following tips are also applicable to graph legends.

- If using proportional symbols, these should be the exact sizes they are on the map.
- Preferably keep the symbol/colour to the left and the description on the right
- If any values are decimals below 1, then these should have a leading 0 and not just a decimal point. (E.g., 0.52 rather than .52)
- Numbers in ranges should be separated using hyphens or the word 'to'
    - Avoid hyphens if said numbers are negative numbers. '-30 - -20' is difficult to read.
- If using different scales (e.g., both colour and proportional symbols), you can organise them into groups separated by headings.

Scale

Maps are designed to be representative of a geographic area and we want our visual to be as accurate to real geography as possible. Therefore, we use a representation of the geographic area scaled down by some proportional relationship. These are less important for thematic maps, but the scale of a map should still be acknowledged.

- You can use either a scale bar (a bar highlighting certain points showing the real-life distance it represents.) or a textual representation (a sentence describing the relationship between distance on the page and real-life distances) or a representative fraction
    - For a thematic map, a textual scale or representative faction are arguably more appropriate and could be relegated to a footnote
- If using a scale bar.
    - Avoid decimal numbers
    - Make the bar long enough to be useful
    - Round to easy to use, meaningful numbers
- The scale should be subtle and small, never drawing too much focus away from the map body as it is there to provide context to the audience.

Inset Map/Locator Map

Inset or Locator maps are smaller secondary maps that are used to provide some additional geography.

Inset maps are maps of related areas that are in some way separate from the main geography, either because they do not border any of the other areas shown (such as maps of the US separating out Alaska and Hawaii) or the areas on the inset map are within the main map body but are too small to be adequately visualised on the larger map body.

The latter is often seen on maps using UK constituencies to show the numerous parts of London as there are many over a small geographic area and therefore would be difficult to show on a map of the entire nation. They will use the same colour scale and symbology as the primary map.

Locator maps are used to show the geographic context of the primary map. These will tend to be off in the corner of the visual pane and will serve to place the main geographic areas within a more familiar geographic context. For instance, if your main map is areas of Bolivia, a locator map may be included to show where Bolivia is within South America.

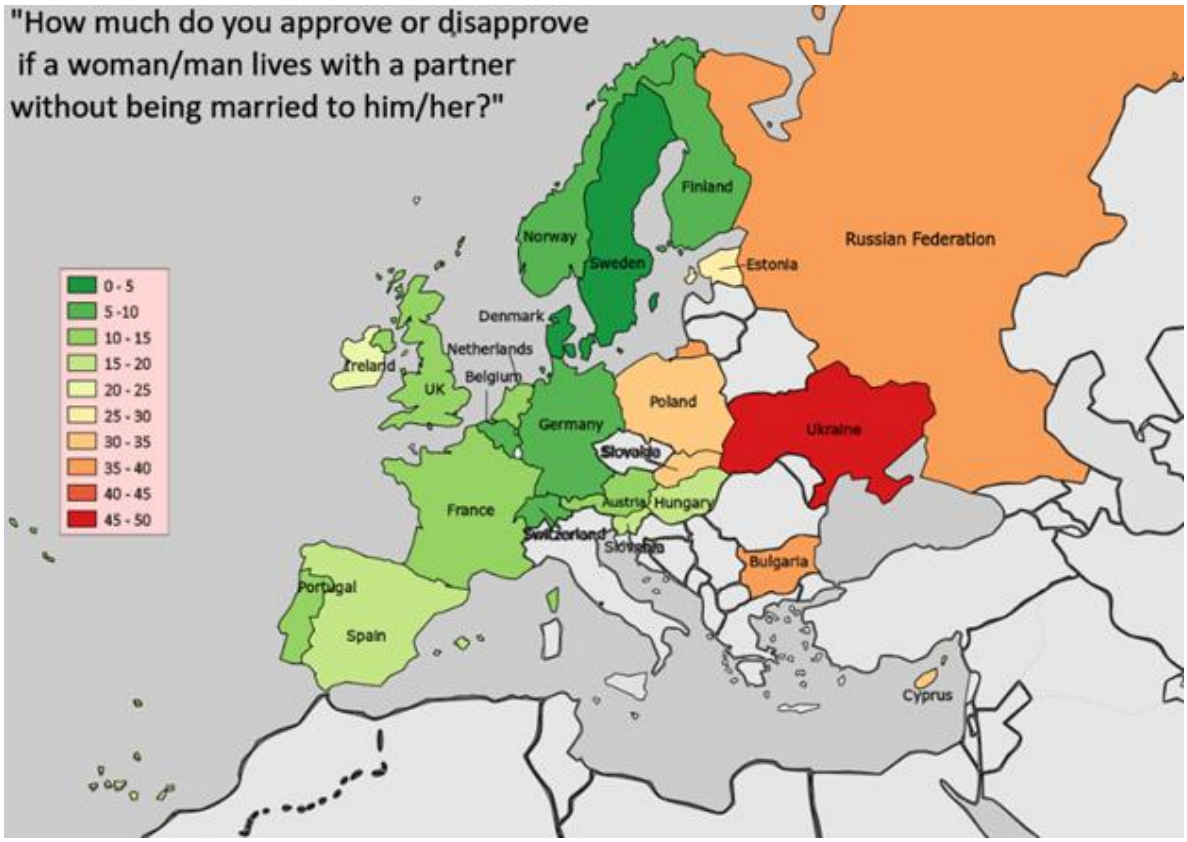
Using a locator map is useful when the area being shown is likely to be unfamiliar to the map reader. It is useful to consider the potential geographic literacy of your audience.

Example

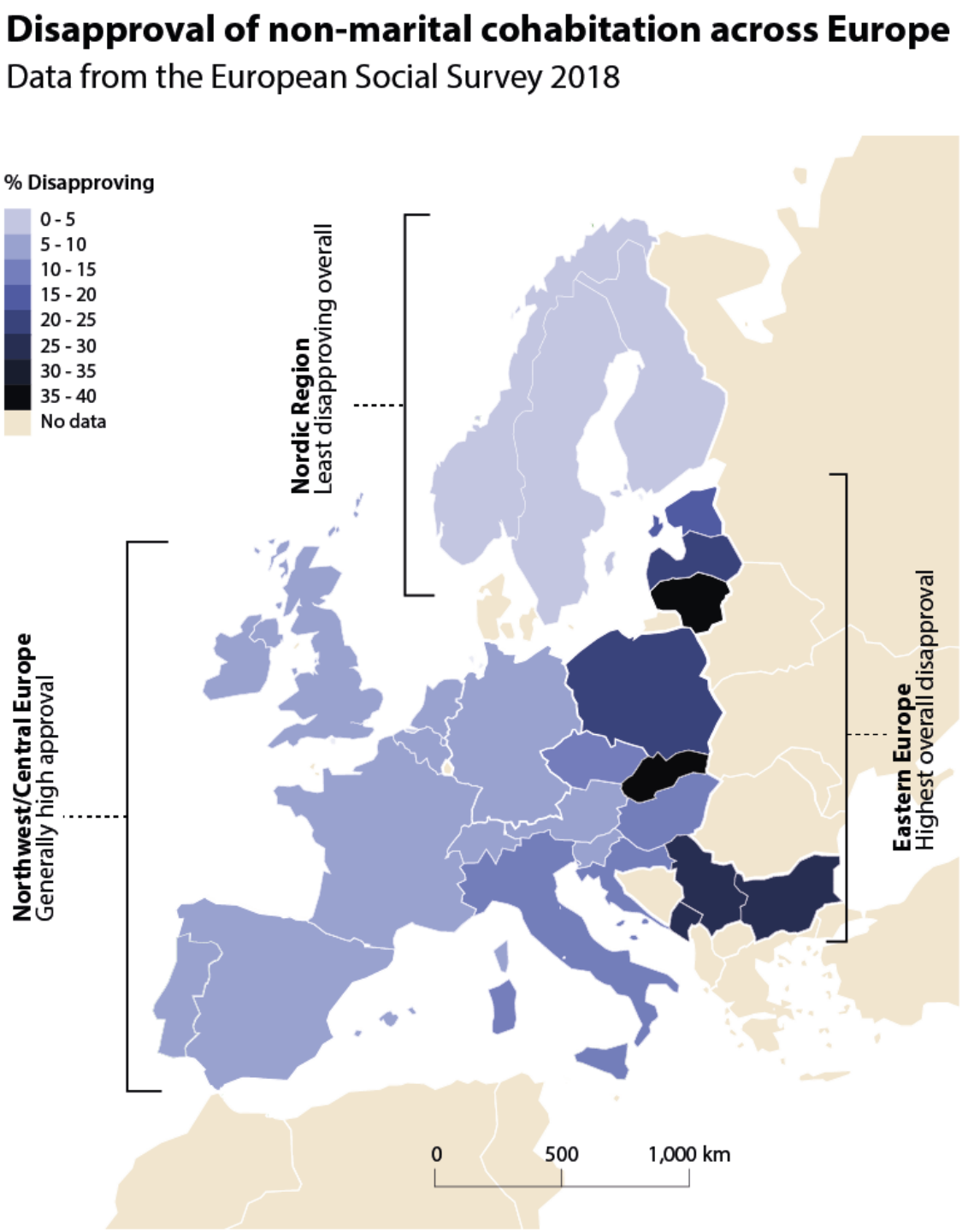
Bad Example

The example presented here makes some quite basic mistakes, mainly in the use of colour and separation.

- A non-colour blind friendly colour scheme has been used to represent the thematic data.
- Areas not being mapped thematically are a similar colour to bodies of water so there is poor separation between the land and sea.
- Nation labels often overlap with lines representing national borders, making them hard to read.
- The title is not very informative.
- The legend is untitled and uninformative.
- No scale included.
- Bold border lines draw unnecessary attention away from the main mapped areas.



Improved Example[1]



[1] Different source of data but using the same variable is plotted. Bad example = ESS 2006, Improved example = ESS 2018

This improved version makes a considerable number of changes to create a much neater and professional looking map.

- There is a bigger contrast of colours between the unmapped areas and bodies of water.
- An improved colour scheme, which focuses solely on changes in luminance.
- A scale bar has been included.
- A much more explanatory title.
- A much neater and more informative legend.
- Annotations help provide additional interpretation of the material.
- Removed the nation labels to avoid clutter.

Accessibility

The following advice is adapted from advice on accessibility published by the Government Statistical service (GSS), which can be found at the following web address: https://gss.civilservice.gov.uk/policy-store/introduction-to-data-visualisation/#section-8

While we may have specific target audiences in mind when we write up our results and produce data visualisations, we should always aim to ensure inclusivity by making our content accessible to those with certain impairments. The GSS specifies impairments to vision, hearing, mobility and thinking & understanding as key areas to consider. For data visualisations, this primarily concerns those with vision and thinking impairments. These tips are designed to make your results as clear and readable as possible in general.

Tables

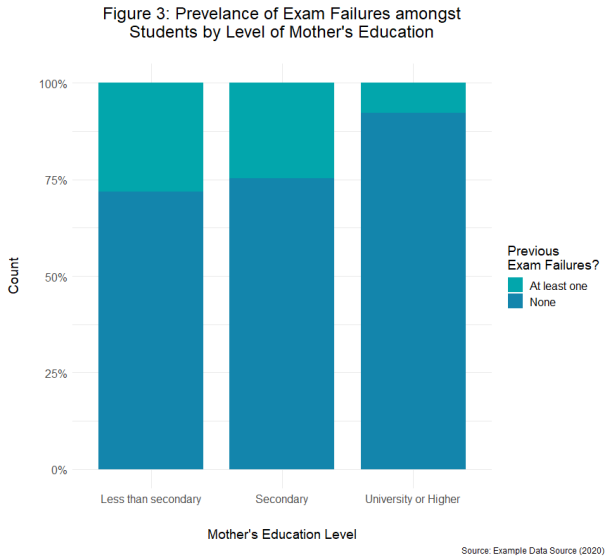
- Use column headers which explain the content of the columns, including measurement units where applicable.
- Include derived variables (e.g., totals) at the end of columns or rows.
    - Try to use more rows than columns.
- Write out or clearly explain any acronyms.
- Use table footnotes/captions to provide extra important information that cannot fit in the main body of the table.
- If you do not need to use exact numbers, consider rounding larger numbers.

Graphs/Maps

- Write out or clearly explain any acronyms.
- Make sure there is a clear distinction between lines on a line graph.
- Do not use red and green together as it is difficult for colour-blind people to distinguish between them.

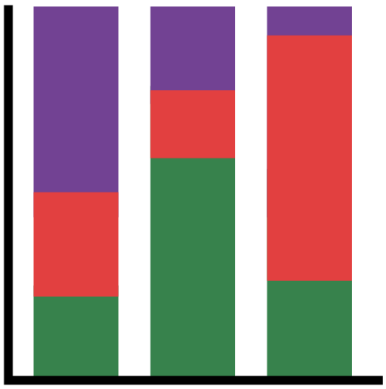
Colour

Colour is one of the most useful tools for supplying extra information to maps and graphs, and sometimes even tables. It can be used to clearly highlight patterns and relationships that could be missed by a monochromatic visualisation. Adding colour can make visualisations more effective, but this only works if viewers can tell which colour is which. For instance, the graph presented here uses colours which are far too similar.

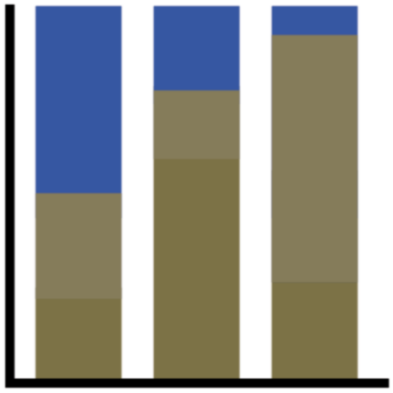


Previous advice already presented on the use of colour in graphs and maps (using it to highlight, use it sparingly etc.) are still relevant and contribute to ensuring accessibility. The advice presented here tends to focus on the choice of colours themselves.

- Ensure the colours are accessible.
    - Colour blindness affects an individual's ability to distinguish between certain colours. It affects men more commonly than women.
    - Most commonly, it affects the ability to distinguish reds and greens. Avoid using these colours together. Less commonly, but to be considered, is blue-yellow colour blindness. The two graphs below demonstrate how a colour-blind person may see a graph using red and green. The colours are much harder to distinguish.



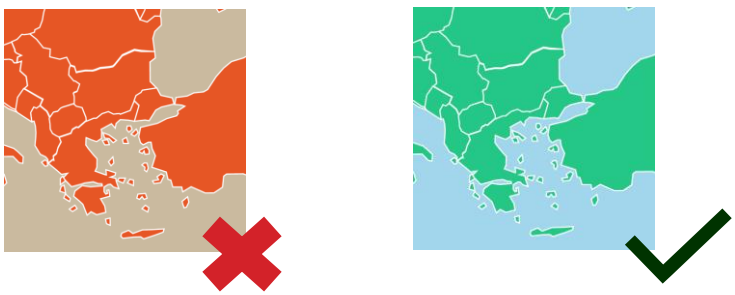
Original graph.



Graph as a colour-blind person might see it.

- Red-green colour palettes may also not be clear when printed in greyscale.
- Blue palettes are a safe default starting point as they are colour blind safe and are visible in grayscale, as shown in the picture below.

- Choose colours carefully.
  - Consider cultural context. Colours often have some inherent culturally defined associations. For instance, using colours people associate with familiar concepts can improve the quality and speed of information processing, such as using blue for water on a map.

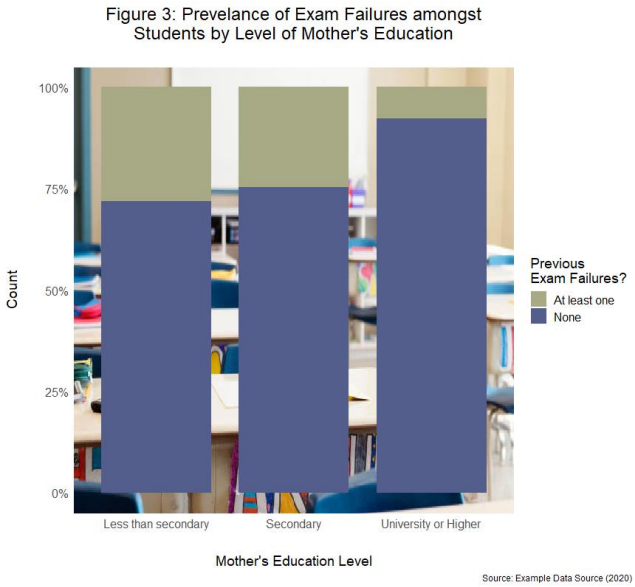


- Understand the digital colour palette.
    - Colours are represented using several common schemes. The most useful of these considers hue, saturation, and luminance. This scheme allows us to intuitively define unique colours.
    - Hue – Hues are colours. They do not have a natural order and therefore users cannot assign a logical order to them. Small changes in hue are easy to detect although colour blindness can affect how well people can detect these differences.

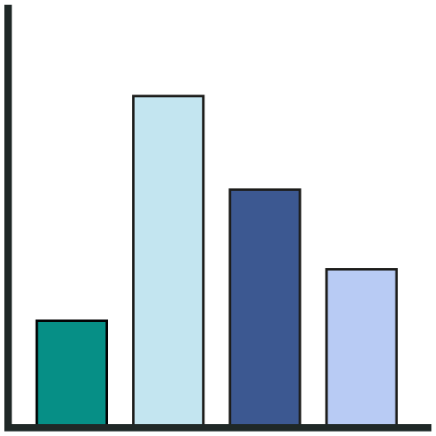
    - Saturation – This is the intensity of a colour, ranging from grey/white (no saturation) to rich, vibrant, almost glowing colour. Saturation is perceived on a continuous scale, although it is difficult to detect small changes. High saturation can also cause issues for those with certain visual/light sensitivity problems.

    - Luminance – the brightness of colour. Also perceived as continuous ordered scale from dark to light. This natural order can help us optimize colour schemes for maximum distinction and differentiation.
        - Changes are easier to detect
        - It is easier to distinguish between bars even if luminance is the only difference.
        - Changes in luminance need to be larger if creating a line graph as the white space between the lines makes it harder to quickly compare.

- Never use an image as a background. This looks messy and cluttered and can make it difficult to read the graph.

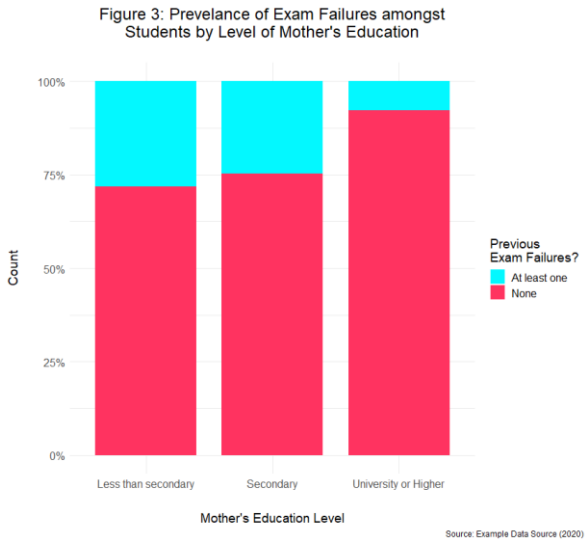


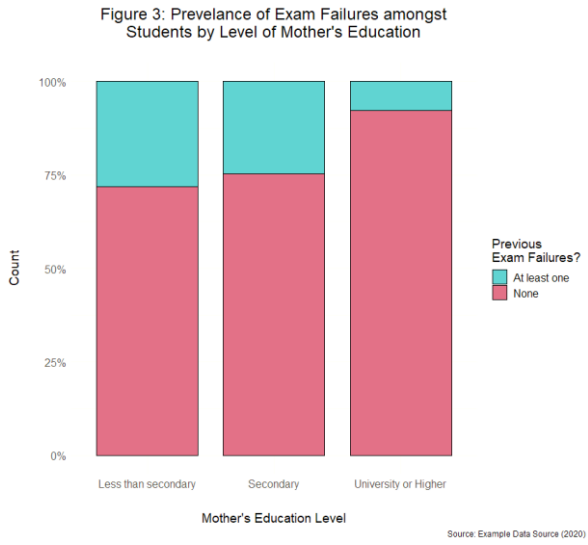
- Know how to use colour effectively
    - Alternate colours – consider alternating dark and light for categorical data to improve clarity and distinction.



    - Use borders – adding thin borders to the edges of bars can enhance clarity/separation.
    - Avoid overuse of saturated colours.
        - Mid to low saturations are preferred
        - Only use bold saturated colour to draw attention to a specific piece of information or hard to see, small elements.

- Bold, saturated colours can have visual side-effects. They may appear to glow for many users, can generate after-images and can affect how other colours appear.
- The first graph below uses very saturated and bright colours which are quite straining to the eye. The second keeps the same hue but reduces the saturation and luminance and adds in some border lines for better separation.





- Use colour logically and consistently. If using sequences of colours, ensure that they progress in a logical manner that the user would expect, such as increasing luminance like in the map on page 50.
  - If creating multiple graphs, use the same colour to mean the same thing. Changing what they mean can confuse the user.
- Use a white background.
  - Most palettes are designed to appear on top of a white background. It provides a helpful reference for the colour scale being used.

Advice on Software

This guide has avoided specific guidance for creating visuals in any particular software, as preferences will vary between readers and over time.

However, we will set out some general advice for you to consider based on the guidance offered throughout this resource. We have repeatedly stressed the importance of design, readability, and consistency. Therefore, there are a few characteristics your software should preferably adhere to.

Your chosen software should offer the ability to code, program, or edit all elements of the graph, so you can quickly draw, modify, or share the visualisation. Preferably, it should also be open source/free so that you can learn from others and keep your costs down.

As such, we generally recommend using R/RStudio and particularly the package ggplot2, which is designed for effective creation of graphs inspired by good practice and principles. For creating tables and maps, other software may be better suited for your purposes, although R is also capable of producing these through well-documented packages such as FlexTable or Leaflet (for maps).

Checklist

For all your data visualisations, consider these key points as a form of a checklist, making sure to mark off all points which are relevant to the work you are producing.

General Guidance

- ☐ **Effectiveness**
  - Are you using the right kind of visualisation?
- ☐ **Readability**
  - Are your titles and labels clear and informative?
  - Is it suitable for your target audience?
- ☐ **Tidiness**
  - Have you avoided clutter?
  - Are the elements well spaced out and positioned?
- ☐ **Accessibility**
  - Have you explained all abbreviations?
  - Have you used colour blind friendly schemes?
- ☐ **Consistency**
  - Have you been consistent between your tables and graphs?
    - Kept categories in same order?
    - Kept the meaning of colours the same?
- ☐ **Informative**
  - Does your visualisation make sense out of context?
  - Have you included the source?

Tables

- ☐ Have you kept the level of precision consistent?
- ☐ Have you kept the number of decimals to the minimum required for your results to be informative?
- ☐ If dealing with very big or small numbers have you considered standardisation or scientific notation?
- ☐ Remember it is easier to compare numbers in columns than in rows
- ☐ Is the order of your rows?

  - Logical?
  - Ascending/Descending?
  - Have none/other categories been put at the bottom of the table?

- ☐ Have you only used border sparingly?
  - Avoided putting them within the main body of the table?
  - Only used horizontal lines?
- ☐ Been consistent and professional with font
- ☐ Grouped columns or rows where possible?

Graphs

- ☐ Are you using the right kind of graph?
- ☐ Are all elements well-spaced out and not cluttered?
- ☐ Have you avoided "junk" features?
- ☐ Have you avoided overplotting?
- ☐ Have you avoided plotting too much information?
- ☐ Have you used and appropriate number of dimensions/variables?
- ☐ Are your axes well labelled and appropriate for the type of graph you are making?
- ☐ Is there a clear distinction between the different levels of other dimensions you may be using?
- ☐ Would annotations be additionally informative to your viewers?
- ☐ **Have you used colour effectively?**
  - Consider:
    - Highlighting.
    - Have you used restraint?
    - Is it colour blind friendly?
    - Is it easy on the eyes (not overly saturated)?

Maps

- ☐ Are you using the right kind of map?
- ☐ Are the colour scale and categories of values you are using appropriate and well labelled?

- ☐ Is your legend accurate and informative?
- ☐ Have you included a well formatted scale bar?
- ☐ Would interpretation be aided using a locator or inset map?
    - Are these appropriately positioned and understandable?
- ☐ Do the most important elements (the thematic symbols) stand out the most?

Other Resources

General

GSS – Introduction to Data visualisation
https://gss.civilservice.gov.uk/policy-store/introduction-to-data-visualisation/#section-7

Informative Presentation of Tables, Graphs and Statistics
https://stats4sd.org/resources/412

Data visualisation examples
https://stats4sd.org/resources/59

Tips on effective use of tables and figures in research papers
https://www.editage.com/insights/tips-on-effective-use-of-tables-and-figures-in-research-papers

Tufte Principles
https://sites.google.com/site/tufteondesign/home/six-fundamental-principles-of-design

Tables

Exporting Tables from R using Flextable, Kable and gt
https://stats4sd.org/resources/506

Preparing tables for research papers
https://www.manuscriptedit.com/scholar-hangout/preparing-tables-research-papers/

Formatting Tables in MS word
http://www.docs.is.ed.ac.uk/skills/documents/3575/3575.pdf

Graphs

R Graphics Cookbook
https://r-graphics.org/

Financial Times – Visual Vocabulary
https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf

Maps

ESRI lecture – Map Elements and Design Principles

https://www.arcgis.com/apps/Cascade/index.html?appid=e739c503a1f04d38839834a0fe4ca6d4

ESRI – Make Maps People Want to Look At
https://www.esri.com/news/arcuser/0112/make-maps-people-want-to-look-at.html#:~:text=Five%20of%20the%20main%20design,map%20and%20on%20the%20page

Carto – 6 Design Principles for Making Maps on the Web
https://carto.com/blog/six-design-principles-making-maps-on-the-web/

GISLounge – Ten Things to Consider When Making a Map
https://www.gislounge.com/ten-things-to-consider-when-making-a-map/

GSS – Effective Maps in official statistics
https://gss.civilservice.gov.uk/wp-content/uploads/2020/06/Effective_Maps_temporary_May_2020.pdf

Accessibility

ONS – Web accessibility
https://style.ons.gov.uk/writing-for-the-web/web-accessibility/introduction-3/

Tableau – Colour Blindness
https://www.tableau.com/about/blog/2016/4/examining-data-viz-rules-dont-use-red-green-together-53463

Sources:

USA election maps:

https://theconversation.com/how-to-read-u-s-election-maps-as-votes-are-being-counted-149251

Misleading Gun death graph

https://www.livescience.com/45083-misleading-gun-death-chart.html