Presentation of Tables, Graphs and Maps

Alex Thomson

2021-07-06

Contents

Chapter 1

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.

Chapter 2

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:

2.1 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.

2.2 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.

2.3 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.

2.4 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.

2.5 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.

2.6 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.

Chapter 3

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.

Chapter 4

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.

4.1 Introduction to Flextable

This section of the guide will be supplemented by reproducible code examples from R. These will be focusing on the incredibly useful package flextable which has been designed to help create report ready tables directly in R. It is especially useful for those intending to write their reports or knit their documents into Word format. This guide will mostly cover some of the basic features, for extended guidance on the full capabilities of flextable, please see this guide

In general, when using flextable the idea is to use R code to manipulate your data into roughly the format you wish to present as a table. In other words you create your table as a data frame. For purposes of demonstration, all data manipulation shown shall be done using the tidyverse range of packages, particularly dplyr and tidyr.

Once you have your data in a desired format you can apply flextable functions to the data to create and design your table. Starting with `flextable()` to turn your data from a data.frame/tibble object into a flextable object.

A flextable object consists of 3 parts.

- header: the section containing any and all headers/titles (defaults to column names of data frames in a single row)
- body: this contains all of the data from the data frame
- footer: not present by default but can be used to add footnotes or additional content

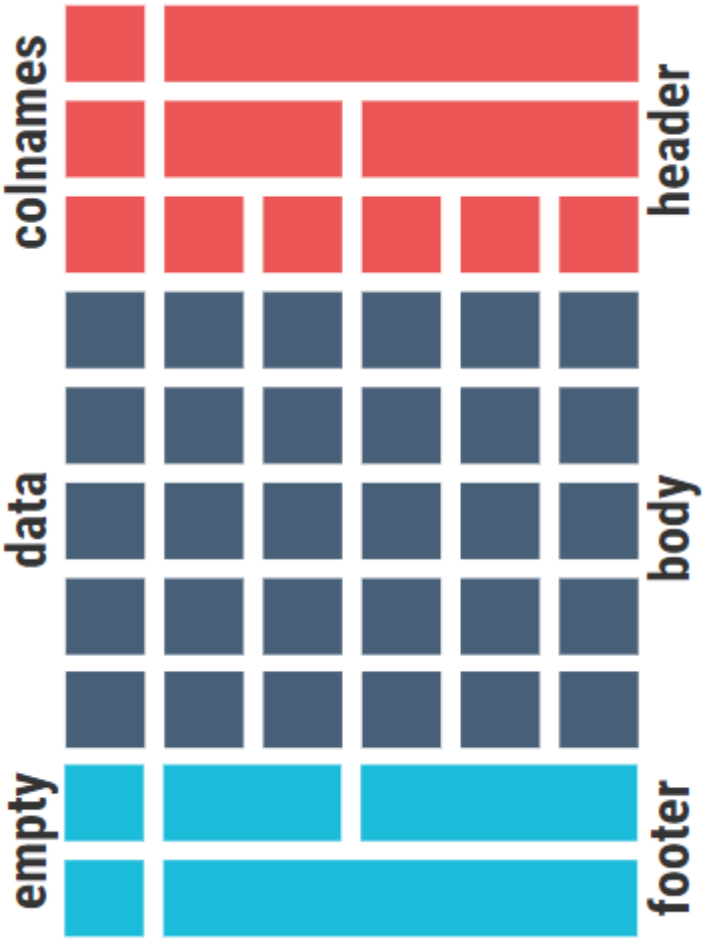

Figure 4.1: Structure of a flextable object

```
#install.packages("flextable")

library(flextable)
```

4.2 General Guidance

4.2.1 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

```
Data <- read.csv("data/MathsGrades.csv")
```

```
library(tidyverse)
```

```
Data%>%
  group_by(Sex)%>%
  summarise(Mock.mean = mean(Mock),
            Final.mean = mean(Final),
            Failures.mean = mean(Failures))%>%
  flextable()
```

```
Data%>%
  group_by(Sex)%>%
  summarise(Mock.mean = mean(Mock),
            Final.mean = mean(Final),
            Failures.mean = mean(Failures))%>%
```

Sex	Mock.mean	Final.mean	Failures.mean
F	51	54	0.29
M	55	57	0.34

Table 1a: Average Exam Results by Sex

Sex	Mock Result (Mean)	Final Result (Mean)	F
F	51	54	
M	55	57	

```
flextable()%>%
set_header_labels(Mock.mean = "Mock Result (Mean)",
                  Final.mean = "Final Result (Mean)",
                  Failures.mean = "Failures (Mean)")%>%
align(align = "center", part = "header")%>%
add_header_lines(values = "Table 1a: Average Exam Results by Sex")%>%
autofit()
```

Chapter 5

Graphs

Chapter 6

Maps

Chapter 7

Accessibility