Authors’ Attempts:
Chapter 2, Section 2.4 Exercises

 **We have each had a go at producing suitable pie charts and tables based on the ‘Stretch your understanding’ data sets we provide in section 2.4. See below for our attempts and what the other thought of them.**

# UK Beach Waste

Graeme’s attempt(*with comments by Rosalind*)

I prefer a table to a pie chart, so let’s start with that.

We are free to order the data any way we like, but I would stick with the order it is presented in (from largest to smallest).

I could not help but check that the percentages add up to 100%, and they do.

I noticed that one value is anomalously given to a different level of precision; my guess is that this was a typing error, so I have changed ‘1’ to ‘1.0’ in my table.

The category labels seem easy to understand, but I thought it was tidier to simplify ‘Sewage-related debris’ to ‘Sewage’.

After that I followed the advice in the chapter pretty closely. I gave the table a number and made the caption short, adding other material that could have been in the caption in a footnote. Where I deviated from our advice is in using alternating shading. I actually like this a bit more than Rosalind does. I can’t pretend that it is really helpful in such a small table, but I don’t think it’s that distracting either—it looks visually appealing to me.

Table G.2.1: Origins of beach waste1

|  |  |
| --- | --- |
| Origin | % |
| Unknown | 46.2 |
| Public | 30.4 |
| Fishing | 10.8 |
| Sewage | 8.5 |
| Shipping | 2.9 |
| Fly tipping | 1.0 |
| Medical | 0.2 |

1 Marine Conservation Society survey of 339 UK beaches (BBC News 2017)

***Rosalind’s comments:*** *This is pretty similar to mine. We follow much the same steps.*

*My table differentiated the column headings from the content using emboldening alone, rather than alongside a darker background colour, but both are effective. I didn’t shorten the ‘Sewage-related debris’ category label in my table (as with my longer column heading, it didn’t stretch the column width too extremely) but I see why this would be more useful when drawing the pie chart. I don’t think Graeme’s use of alternating shading is a terrible crime, but I don’t think it particularly helps either. Some people might find it eye-catching, but I don’t think mine is stark by comparison. So, I do prefer my version because he and I both believe that most tables we see are too cluttered, and decluttering tables is a good thing. That said, I don’t think this is a major issue. Some issues about graphing and data presentation really are down to personal preference.*

Now, since you forced me, I produced a pie chart. See my code for how I did this.



Figure G.2.1: Origins of beach waste. From a Marine Conservation Society survey of 339 UK beaches (BBC News 2017).

Here we see a well-known problem with pie charts: the ‘medical’ slice is so thin that you cannot really see it. Also, enough of the slices were so small that I couldn’t see a simple way to get their labels to not overlap, so I ended up abandoning the labels and squeezing everything into the legend. But without the labels the pie chart really does become pointless.

***Rosalind’s comments:*** *I generally like the distinct colour choices in this pie chart, but I think the ‘medical’ slice would be better as a more contrasting colour to its neighbouring slices to help it be more visible (although it is really tiny). I agree that without labels associated with the slices we might as well have a table, but we both struggled to find a neat way of providing the percentages associated with slices of the pie, given the small values associated with some of the categories here.*

Rosalind’s attempt(*with comments by Graeme*)

I find that tables are always easier to read than pie charts, particularly when the data is nominal and can be arranged in descending order of value. With this data set, the categories are already given to us this way, so it does not take much effort to organize it into a smart-looking table.

Following the guidelines outlined in section 2.3.1 of the chapter, I started by drawing a vertically-oriented table (for easier comparison of values), made sure the values were all given to the same number of decimal places, right-aligned the column with numerical values and left-aligned the column with text data, and pulled in the column sides to minimize blank space. I then emboldened the column headings and minimized the number of lines/borders around cells, keeping only a few horizontal lines to structure the table. Finally, I gave the table a short caption that provided details on the source of the data.

Table R.2.1: Origins of beach waste in the UK. From a Marine Conservation Society of 339 beaches across the UK (BBC News 2017).

|  |  |
| --- | --- |
| Origin of waste | % |
| Unknown | 46.2 |
| Public | 30.4 |
| Fishing | 10.8 |
| Sewage-related debris | 8.5 |
| Shipping | 2.9 |
| Fly-tipping | 1.0 |
| Medical | 0.2 |

***Graeme’s comments:*** *We went for very similar tables. I went for alternating shading for visual appeal—but honestly, when I look at it, I prefer the clarity of Rosalind’s approach. In future, I will save alternating shading for much larger tables when the reader has to follow rows across several columns. Rosalind rightly picked me up for not proofreading my header for the ‘Attitudes to Alcohol’ example, but I am way too polite to suggest that she missed the word ‘survey’ after ‘Society’ in her caption. Or maybe I’m not that polite.* 😊

In the spirit of ‘giving it a go’ I also tried making a pie chart of this data.

Because there were quite a few categories, some of which had really small associated values, I thought I’d try to make the pie large so that all categories would be visible. To do this, I played with the margins of the plot a bit (see my code and section 7.2 of chapter 7 for details on how to do this) and made a bit of extra space to the right of the pie for the legend.

I originally intended to place the percentages as labels for each slice of the pie, but the tiny values of the smallest percentages made it tricky to do so without them overlapping. I tried one method, where the **locator** command in base R allows you to interact with the plot in the ‘Plots’ window of RStudio to manually position labels (see my code to give this a go), but this was tricky to do by hand and would not be reproducible as code alone. It may well be possible to arrange the labels for this data by plotting the percentages as text, but figuring out suitable coordinates would take a lot of laborious trial-and-error. Instead, I opted for the easy way out and just included the percentages in my legend. But this essentially made the legend a less readable version of my table, with the pie itself adding precious little!

In placing my legend, I used **inset** to alter how close it was positioned to the ‘right’ of the plot and **xpd=TRUE** to allow R to add text where it fell slightly outside the plotting region. I removed the automatic box around the legend with the command **bty=”n”**. Instead of typing out all the percentages after my waste origin category names, I also used a lazy shortcut to get R to add the % symbol for me using the list of values that I already gave it to produce the plot itself. The command in my code that reads **paste(waste,” (“,percent,”%”,”)”,****sep=””)** tells R to print, in turn, each of my waste category names, followed by a space and an opening bracket, then the associated percent value, then a % symbol, and then a closing bracket. **Sep=”” is** a character string used to separate the terms listed in the paste function, and here (as no space is included in the quotation marks) functions to prevent big gaps from being included between the terms. See sections 6.3.3, 7.3.5, and 8.6 for further tips on adding lists of text and mathematical symbols to plots.



Figure R.2.1: Origins of beach waste in the UK. From a Marine Conservation Society survey of 339 beaches across the UK (BBC News, 2017).

***Graeme’s comments:*** *Rosalind wins hands down here. She really tried some ingenious things to try and get this pie chart to work. But even then, in the end she finished up with something not so different from mine—and so very inferior to a table. However, her pie chart is still better than mine. The biggest improvement is in changing the borders to remove white space around the chart. Her legend is better too, for the space she made for it, for removing the box round it, and for the stylish way she automatically added the percentage values.*

# Attitudes to Alcohol

Graeme’s attempt(*with comments by Rosalind*)

My feeling is that very little precision is lost, and the data is easier to assimilate, in percentage form, so I got R to calculate the percentages for me, rounding them to one decimal place. Otherwise, I very much followed my line of thinking from example 1, using my alternating shading again.

Table G.2.2: Perceived societal attitude to alcohol1. Responses to the statement:

“*People in Scotland/England are generally discouraged or encouraged to drink alcohol*” (Li et al., 2017)

|  |  |
| --- | --- |
| Origin | % |
| Strongly Discouraged | 3.6 |
| Discouraged | 11.2 |
| Neither | 35.5 |
| Encouraged | 30.1 |
| Strongly Encouraged | 19.5 |

1 2017 study of 1700 English and Scottish drinkers

***Rosalind’s comments:*** *As before, I think this table is pretty sound design-wise, although I still wouldn’t use the alternating shading myself for such a simple table. Presenting the data in percentage form is a good call in terms of making comparisons easier. In my version, I included the raw frequencies as well, but whether these are useful to present may depend on the purpose and audience of the table. I think a final proofread of the column headings would have made this table very fit for purpose.* 😊

Now let’s try a pie chart.



Figure G.2.2: Perceived societal attitude to alcohol. Responses to the statement:

“*People in Scotland/England are generally discouraged or encouraged to drink alcohol*”

(From a 2017 study of 1700 English and Scottish Drinkers by Li et al.)

That worked a little better than the first pie chart—but I still had to fiddle around with font sizes to squeeze everything in. I went for a colour-diverging colour choice based on traffic lights but I am not sure how intuitive that turned out to be.

***Rosalind’s comments:*** *I like the idea behind the traffic-light colour choice—it conveys negative relative to positive attitudes quite well. But perhaps a lighter shade for the ‘Neither’ slice would make it more clearly a middle ground between the two. Otherwise, the chart looks smart, though it’s a shame that the automatic plot margins in R make pie charts very small.*

Rosalind’s attempt(*with comments by Graeme*)

For the table design I followed the same guidelines from section 2.3.1 as I did for the first example data set. However, as well as showing the raw frequencies from the original data set, I added an extra column to show the relative percentages of each response—these percentages were calculated and rounded to one decimal place using R. Whether it is useful to show both the raw frequency and the relative percentage might depend on the table’s intended audience and use.

Table R.2.2: Public perceptions of society’s attitude to alcohol consumption. The responses of 1700 English and Scottish drinkers to the statement: “*People in Scotland/England are generally discouraged or encouraged to drink alcohol*”. From a 2017 study into public attitudes to alcohol by Li et al.

|  |  |  |
| --- | --- | --- |
| Response | Frequency | % |
| Strongly discouraged | 62 | 3.6 |
| Discouraged | 191 | 11.2 |
| Neither | 603 | 35.5 |
| Encouraged | 512 | 30.1 |
| Strongly encouraged | 332 | 19.5 |

***Graeme’s comments:*** *Again, I agree that when I see Rosalind’s attempt, my use of alternating shading was not the best choice. I think, in the end, Rosalind’s caption is just a little more graceful than mine (avoiding the reader’s eye having to dart down to the footnote). I can see situations where folk might want to discuss frequencies as well as the percentages, so I think including both was a good call. I think if I had included both, then I would have made the last column a bit wider, just to visually separate the two columns of numbers a bit.*

When attempting a pie chart, I just used the percentages I had already calculated in R for the table, as these relative values are more useful for making comparisons when arranging categories as a pie.

This time the data was ordinal, so rather than ordering slices by value I kept the order running from the one extreme ‘Strongly discouraged’, through ‘Neither’, to the other extreme ‘Strongly encouraged’. Because the data ranged two extremes with a neutral middle-ground I went for a diverging colour palette, with a grey neutral ‘Neither’ slice, one extreme in shades of red and the other extreme in shades of blue. Red and green would probably feel more intuitive for negative and positive attitudes respectively, but this combination can be problematic for colour-blind viewers if shades are not selected carefully. See section 1.6 for more discussion of how to choose colours effectively.

I again decided to make the chart big (by customizing the plot margins, again see section 7.2 for details on **par** and **mar**) to improve readability, and used **inset** and **xpd=TRUE** to place my legend just where I wanted it. Fortunately, because none of the slices were so tiny this time, it was also much easier to assign the slices labels that did not overlap, and I used my lazy %-adding **paste** shortcut here this time.



Figure R.2.2: Public perceptions of society’s attitude to alcohol consumption. The responses of 1700 English and Scottish drinkers to the statement: “*People in Scotland/England are generally discouraged or encouraged to drink alcohol*”. From a 2017 study into public attitudes to alcohol by Li et al.

***Graeme’s comments:*** *Again, all the changes that made Rosalind’s first pie chart better than mine are true here too. Looking at this makes me think of yet another reason why I don’t like pie charts. Here we have five categories with a clear order to them. In the table the two extremely contrasting positions are naturally as far away from each other as they could be, whereas in the pie chart they end up side by side--that just seems wrong.*

# References:

BBC NEWS. 2017. *Seven charts that explain the plastic pollution problem* [Online]. Available: <https://www.bbc.co.uk/news/science-environment-42264788> [Accessed 26/10/2020].

LI, J., LOVATT, M., EADIE, D., DOBBIE, F., MEIER, P., HOLMES, J., HASTINGS, G. & MACKINTOSH, A. M. 2017. Public attitudes towards alcohol control policies in Scotland and England: Results from a mixed-methods study. *Social Science & Medicine,* 177, 177-189.