Authors’ Attempts:
Chapter 3, Section 3.8 Exercises

 **We have each had a go at producing suitable figures based on the ‘Stretch your understanding’ data sets we provide in Table 3.5 and Table 3.6. See below for our attempts and what the other thought of them.**

# Bee caste relative abundances

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

While I would usually prefer to plot multiple samples such as this as a grouped bar chart, here the percentages of trapped individuals in each caste add up neatly to 100% each month, and we’re interested in the changing overall caste composition over the months. So for these reasons, I thought this was a rare chance to use a stacked bar chart.

To do this, I inputted the data as three lists (one per caste) and plotted a bar chart with a data set of the combined lists, excluding the **beside=TRUE** needed for grouped bar charts. This added the bars for each caste on top of each other, so my y-axis was set to a maximum of 100%.

The counts of individuals in different castes are nominal qualitative data, so I wanted to select three distinct colours. To do this, I chose to load the package ‘viridis’, which provides an accessible colour palette, and got R to choose 3 colours from this in the code for both my bar chart and the accompanying legend.

Because each of the stacked bars fills the entire plot (i.e. runs the full length of the y-axis), the legend had to be positioned outside of the plotting area. To do this, I had to include **xpd=TRUE** in its code, which tells R that it is allowed to plot outside the span of the current figure. I wanted the legend positioned at the top of the plot, so I then had to make the **inset** value negative so that my **"top"** positioned legend would be higher than the maximum height of the stacked bars. I also removed the box of the legend (**bty= "n "**) and made the legend horizontal (**horiz=TRUE**).



***Graeme’s comments:*** *Rosalind picked much more attractive colours than me and found a much more stylish way to present the legend. However, I am horrified by the lack of a figure caption!*

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



Figure G3.1: Monthly percentages of bumblebee *Bombus terrestris* individuals of different types trapped in Hokkaido, Japan (source: Inari et al. (2005)).

I had to go with a stacked bar chart, really, because we heavily hinted when we introduced this example that that might be a good choice. Some things to note are that I think the chart looks a little better with an x-axis drawn in, which I did with **abline**. I also used **xlim** to squeeze the bars together to make space for the legend, and I thought the legend looked better with the box round it removed (see **bty = “n”).** Also, if you look at my code, you will see that I defined the list of colours at the start and referred to that list in both the barplot and the legend. That means if I decide to change colour, I only need to make one change in my code, not two. Lastly, note that I have made the font of the caption different from the main text here. I think that is quite a good thing to do in reports to help the reader know what is caption and what is main text.

***Rosalind’s comments:*** *I think the colours Graeme has picked out for Figure G3.1 actually work really well here—they are very distinct from each other. I also like the addition of an x-axis line, which joins the bars together as one figure and makes this version look less sparse. Personally, I’m not so keen on the legend sitting above blank space and an extended x-axis—I would rather the bars filled the plot space and the legend was off to the side without the axis being extended for it. But I really love the figure caption, which I shamefully neglected to accompany my version with!*

# Alzheimer’s deaths

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

This data set was nominal, in that there was no logical ordering to the causes of death, so I decided to order the bars in my bar chart by decreasing percentage change rather than sticking with the order in the table. Whichever order they were in, the different names of the various causes of death took up quite a lot of space along the x-axis, so I made the names involving two words run over two lines (using **\n** to indicate where R should start a new line) and made all of the names a slightly smaller font size than the default of 1 (using **cex.names**).

Because there were quite a few different bars, and thus the last bars added were quite a distance from the y-axis, I decided to add some major and minor grid lines to help viewers interpret their values. To further improve the readability of the y-axis, I also used the package ‘Hmisc’ to add minor ticks and used the argument **las** to orient tick labels horizontally.

Regarding colour choice, there was no reason to colour the bars for different causes of death in different colours (they were all showing % change), so I just chose a single, clear colour for the whole figure.



***Graeme’s comments:*** *This is really pretty similar to mine. I think Rosalind’s looks a little more polished with the added tick marks, and she has differentiated the major and minor horizontal lines better than me. In retrospect, my y-axis label is a bit too involved, and her briefer one is better. And it would have been better yet if combined with a caption.* 😊

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



Figure G3.2: Percentage changes in selected causes of death in the USA between 2000 and 2018. Based on a figure by the Alzheimer’s Association (2020) created from data from the US National Centre for Health Statistics (CDC WONDER, Tejada-Vera, 2013).

Really there isn’t much here that we haven’t covered in the book. To squeeze the names of the causes of death in I had to shrink the font size using **cex.names** and force new lines using **\n**. The only slight innovation was that I thought, because this data came from a study by the Alzheimer’s Association, that I should highlight Alzheimer’s Disease on the figure by making it a different colour.

***Rosalind’s comments:*** *There isn’t much difference between our attempts at this one, but I think the use of different colours in Figure G3.2 was a nice idea here, as the Alzheimer’s Association would clearly want to highlight one of these categories over the others. I generally prefer to change line colour and/or width to differentiate major and minor grid lines rather than line type, but they do still serve the purpose here. The y-axis label is quite busy compared to mine, but explaining what the data shows in both the y-axis label and a figure caption is certainly better than failing to do either…* 😊

## References:

ALZHEIMER’S ASSOCIATION 2020. Alzheimer's disease facts and figures. *Alzheimer's & Dementia,* 16**,** 391-460.

CDC WONDER. *CDC WONDER online database: About Underlying Cause of Death, 1999-2018* [Online]. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Available: <https://wonder.cdc.gov/ucd-icd10.html> [Accessed 07/05/2020].

INARI, N., NAGAMITSU, T., KENTA, T., GOKA, K. & HIURA, T. 2005. Spatial and temporal pattern of introduced *Bombus terrestris* abundance in Hokkaido, Japan, and its potential impact on native bumblebees. *Population Ecology,* 47**,** 77-82.

TEJADA-VERA, B. 2013. Mortality from Alzheimer’s disease in the United States: Data for 2000 and 2010. Hyattsville, MD: National Center for Health Statistics.