R in real life: analysing NHS data

Each winter, the NHS releases updates of how busy hospitals in England are on a daily basis[[1]](#footnote-1). In 2018–19, Fiona Grimm from the Health Foundation used R to analyse some of this data[[2]](#footnote-2). A simplified version of this code, along with the relevant other files, is contained in **WPA.zip**.

* Save **WPA.zip** to your computer and extract the contents.
* Open **WPA.R** and run all the code.   
  Hint: See the Code, Run Region menu. If you encounter problems, read the ‘Working directory’ section below.

As R completes this relatively long computation, you may see a red stop button in the top-right corner of the RStudio console. Being able to stop code that is running can be particularly useful if you accidentally write code that will never terminate.

The code will produce two outputs: a plot, and a **.png** file (located in the same folder).

* Open the **.png** file to see the result of the analysis.
* Look through the R code; see if you can work out what the different sections of code are doing.

The first part of the R code loads some **libraries**. These are blocks of code that have been written to provide more functionality in R.

One line of code replaces certain data-points with the value **NA**. This is a special value that tells R that a data-point is missing.

* Find this line of code.
* In what circumstances is a data-point being replaced by **NA**?
* What do you think would happen to the data analysis if these values had not been replaced?
* What problems were encountered while processing the data that had to be dealt with?
* Can you work out what the plot is showing you?   
  Hint: Find the corresponding plot command in the code, and check what is happening at that point of the analysis.
* Edit the code to process the 17–18 data instead of the 18–19 data.  
  Hint: There are three lines of code that need changing, one near the top and two near the bottom. You may also wish to edit one comment.
* Compare the output for 17–18 with that for 18–19. What possible factors might have resulted in the differences?

# Extension exercises

* What different file types are used? Where in the code do they appear?
* Change the boundaries of the categories the data is split into, and the colours used.
* Compare the 16–17 data file with those for 17–18 and 18–19. Can you see any difficulties that might arise in using the same code to process the 16–17 data?
* The full analysis also considered data on hospital beds occupied by long-stay patients. Looking at the original data files, can you see any potential difficulties with comparing this data over the years?

# Working directory

When using more than one file, R needs to know where to look for each file. Usually it is not a good idea to give the complete ‘address’ of every file (reason 1: the full address is usually quite long; reason 2: if you move your files, you don't want to have to change every reference to another file). Instead, R has an idea of which folder it thinks you are using, and tries to find files there; this is called the **working directory**.

If you open a **.R** file, R is probably going to take the folder it is in as the working directory. However, this might not always happen. You can check what the current working directory is by running the command **getwd()**. If the working directory is not set correctly, loading files will not work properly, as R will look in the wrong place for them. You can change the working directory using the command **setwd(** ‘address of new folder' **)**. If the character \ appears in the address, you need to replace it with /.

-------------------------------------------------

This worksheet is designed for students who have recently begun using R and RStudio, and so assumes a minimal understanding of these. It aims to give students a concrete example of R being used to address real-life concerns, while also touching on computer- and statistical-literacy, along with some R(Studio)-specific skills. It can be used in a classroom context or for self-study.

1. See the acute time series data at <https://www.england.nhs.uk/statistics/statistical-work-areas/winter-daily-sitreps/> . [↑](#footnote-ref-1)
2. See <https://towardsdatascience.com/using-r-to-track-nhs-winter-pressures-fedcccce0b06> . [↑](#footnote-ref-2)