

## What is R?:

R is a powerful and free programming language and software environment designed specifically for statistical computing and graphics. It is widely used by data analysts, statisticians, and researchers for data analysis, data visualization, and statistical modelling.

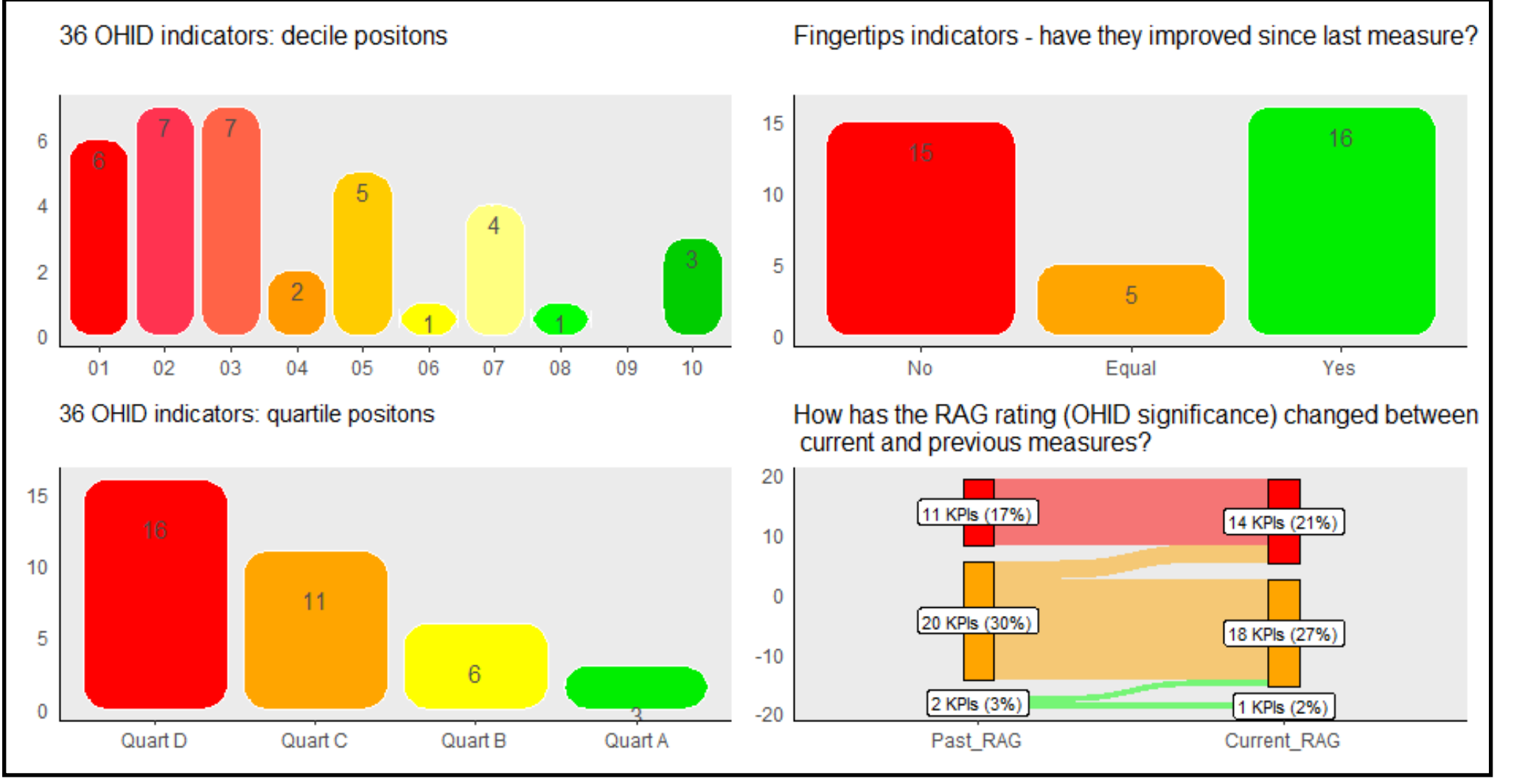
## What have I used R for?:

I have used R for various applications and found it particularly useful for analysing Fingertips data using the application programming interface (API). R has enabled me to analyse multiple indicators from every local authority, create integrated heat maps, calculate quartile positions, and understand positions in relation to statistical neighbours. While the need to use code can be a downside, once the code is written, it is very easy to tweak variables (e.g., gender, age, indicator ID) to report on new Fingertips indicators. The purpose this poster is to showcase some of R's potential in a public health setting.

## Monthly surveillance:

I have used R to provide monthly reports showing which indicators have been updated, whether they have improved, and what their quartile / decile position is. This format allows all Fingertips indicators to be reviewed regularly and reported on a monthly basis. The script is about 750 lines long and draws monthly data directly from Fingertips, automatically creating the table and four summary charts. In September, 36 indicators were updated and could be analysed by quartiles. To achieve the results below, R processed 182,555 lines of data.

Indicator Information	Quartiles	Performance	Evaluation
<b>UID</b>	<b>Q1</b> <b>Q2</b> <b>Q3</b>	<b>Prev1</b> <b>NL</b> <b>Eng</b> <b>Impr</b> <b>Quart</b> <b>Pos</b> <b>Decile</b> <b>Past_RAG</b> <b>Current_RAG</b>	<b>Not compared</b> <b>Not compared</b>
91818_All_Age_Persons	6.0 8.8 14.5	2.9 3.1 10.4	Not compared
93542_15-yr_Male	2.0 3.0 5.0	0.0 0.0 34.3	Not compared
93544_15-yr_Female	3.3 4.8 17.1	0.0 0.0 46.4	Not compared
92712_16-24-yr_Persons	44.1 52.2 60.3	63.9 52.4	Not compared
92713_25-49-yr_Persons	60.4 65.0 69.9	67.2 65.4	Not compared
93052_16-49-yr_Male	6.0 7.9 10.5	6.0 6.8 7.9	Not compared
93074_16-49-yr_Female	6.0 11.4 20.6	12.7 10.7 20.9	Not compared
93982_16-64-yr_Persons	1.5 2.0 2.8	2.3 1.8 2.9	Not compared



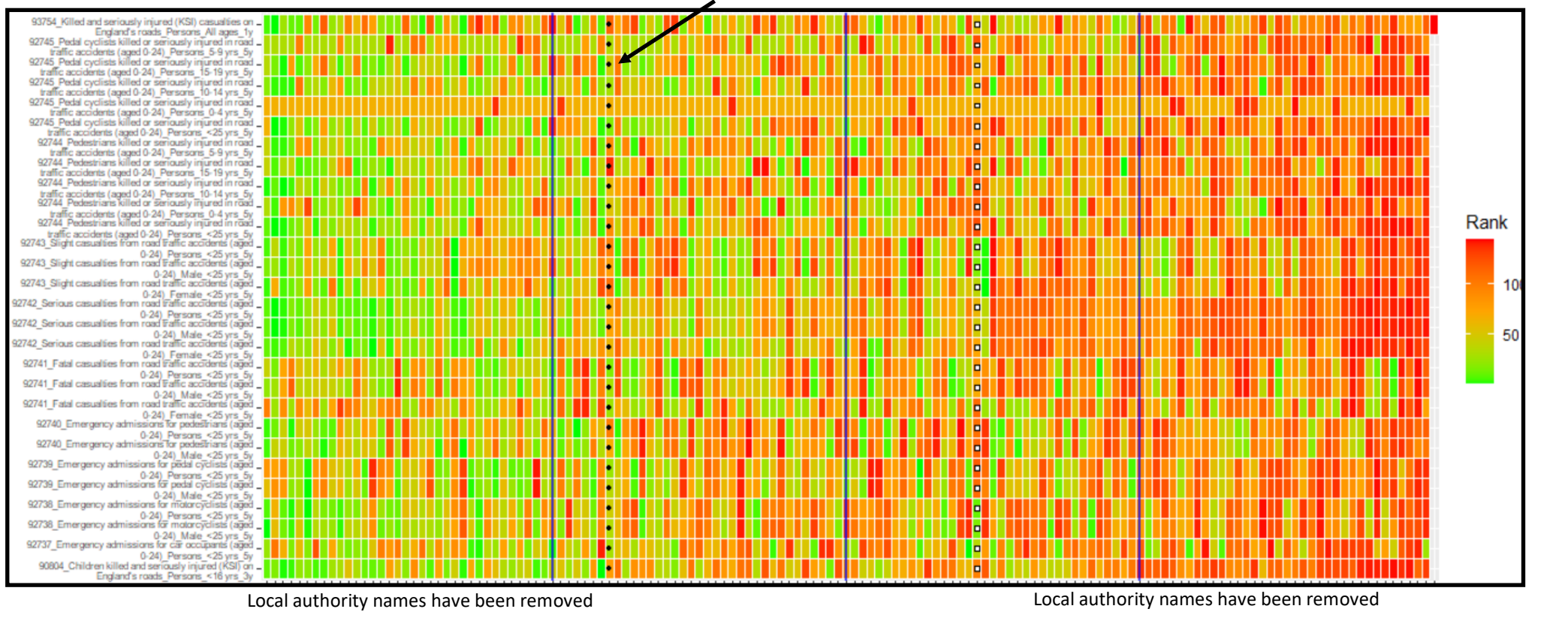
## Performance Reports :

The code used for the surveillance reports has also been modified to conduct a deep dive analysis of all Fingertips KPIs. This helps, for example, identify how many indicators a local authority has in each quartile and decile. A recent review of indicators analysed over 2.6 million records using R. The table below shows a few rows looking at all indicators in decile 10. R is particularly efficient at harvesting data from Fingertips, enabling this kind of analysis. Data can be formatted using a library called RMarkdown to produce reports.

Indicator Information	quartiles	Context	Evaluation
91818_All_Age_Persons	1524 1 1021 1	2341 3281 107 7	Not compared
91819_15-yr_Male	1724 1414 1781	1712 2461 545	Not compared
91820_15-yr_Female	161 166 201	26 28 20	Not compared
90771_16-24-yr_Persons	161 166 201	26 28 20	Not compared
90772_25-49-yr_Persons	161 166 201	26 28 20	Not compared
93542_All_Age_Persons	35 55 77	34 23 12	Not compared

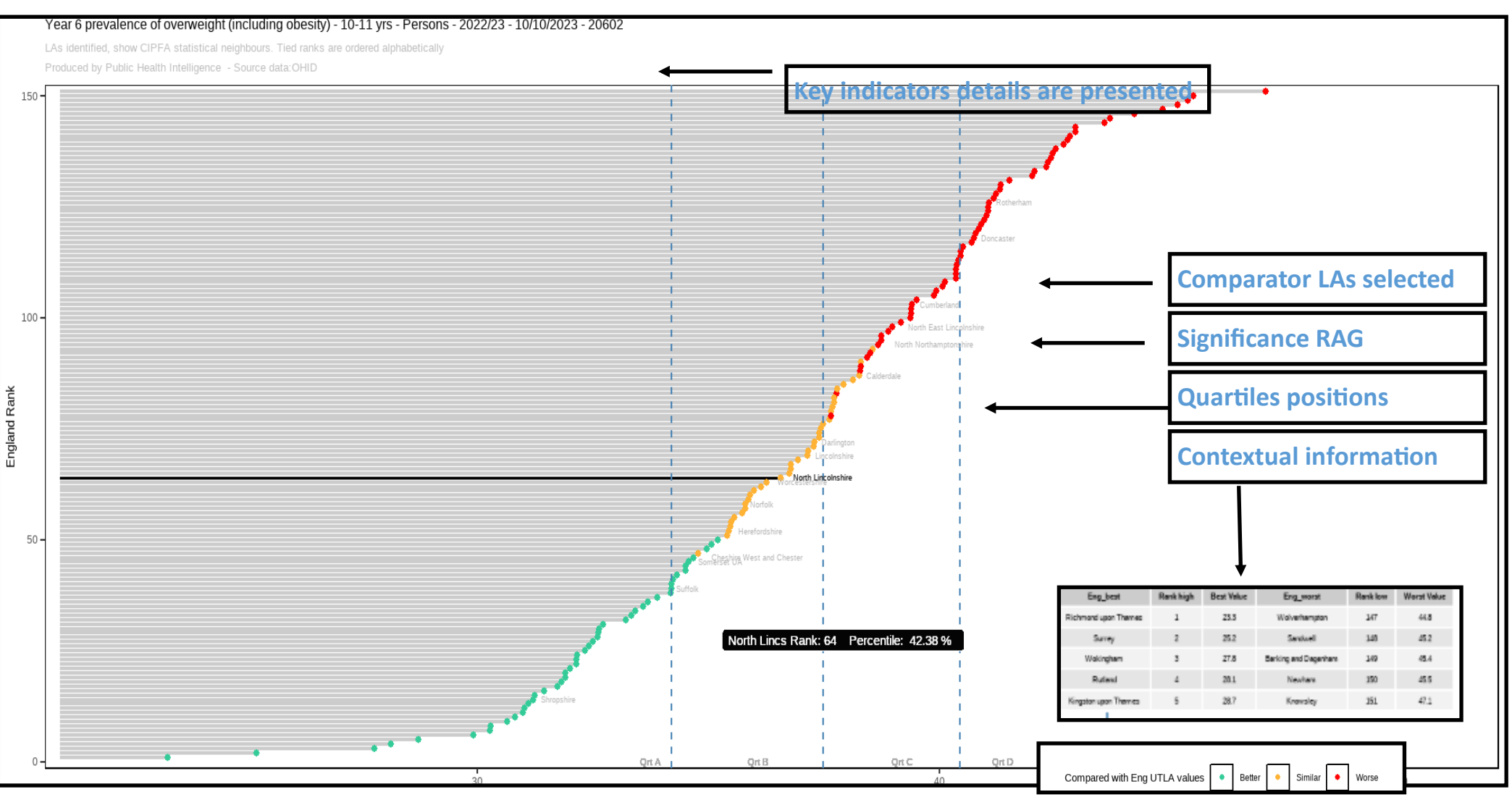
## Complex patchwork / indices:

This heatmap or patchwork analysis allows multiple indicators to be selected from Fingertips, creating an index that shows the relative performance of each local authority as a colour-coded patchwork. Each local authority's performance is ranked, with the data aggregated to display higher-performing local authorities on the left (better). R calculates the polarity for each indicator and ranks the data accordingly (from 1 to about 152, depending on the number of local authorities). The illustration below uses approximately 4,500 data points.

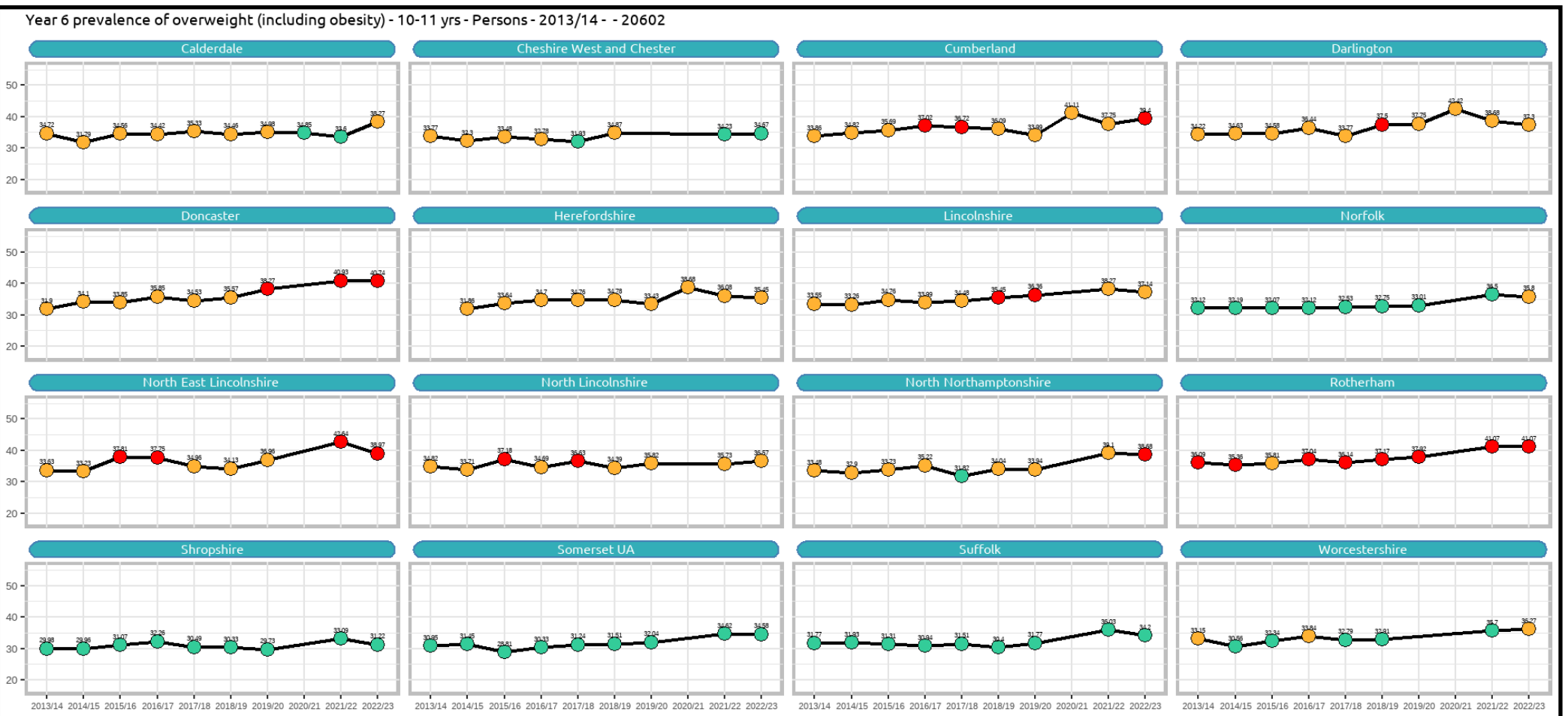


## Benchmarking reports:

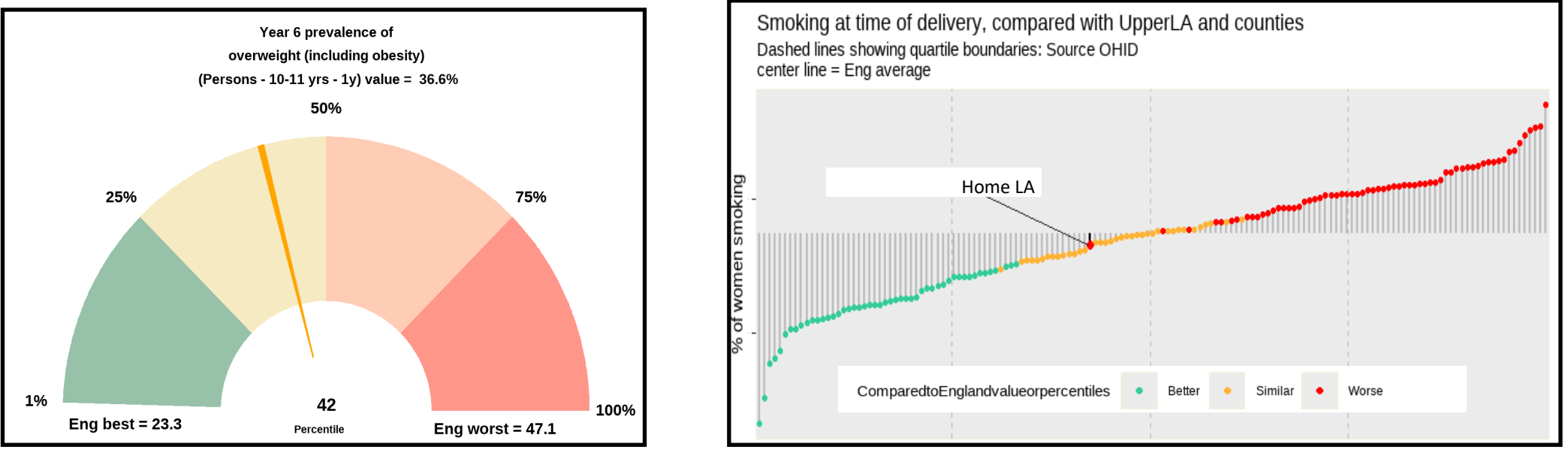
The charts below shows two benchmarking reports that clearly evaluate the relative position of any selected local authorities. The top chart displays the home local authority (black line) and comparator local authorities. Any comparators can be selected, but I have chosen statistical neighbours. The data are then separated into quartiles and RAG rated according to significance. The code takes into account the polarity of the indicator and formats the quartile boundaries accordingly. The header is generated from Fingertips meta data and presents key information about the indicator definition.



The line chart provides a trend view of how well the home local authority is performing compared to a selection of other local authorities over time. The data visualisation capabilities in R are excellent for displaying multiple comparison graphs in an ordered format. Producing this report, compared to a single line graph, only required a couple of extra lines of code.

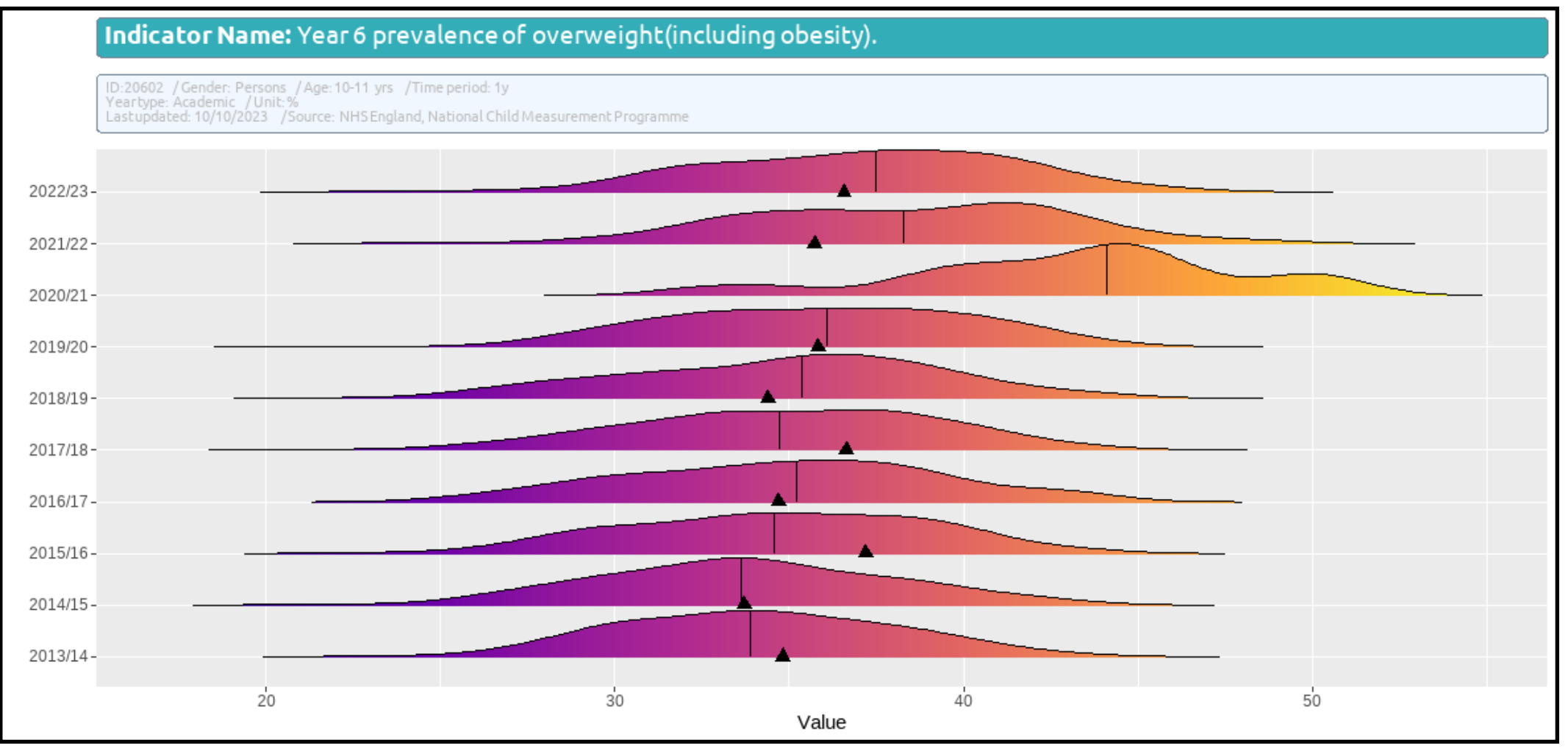


The dial chart below uses three variables from Fingertips (age, gender, and Indicator ID). By changing these variables, a dial can be produced for most Fingertips KPIs. The dial shows the quartile position, significance (needle colour), and the best and worst outturn results for English local authorities. The lollipop chart on the right is a specific benchmarking report that shows whether the outturn is better or worse than the England average

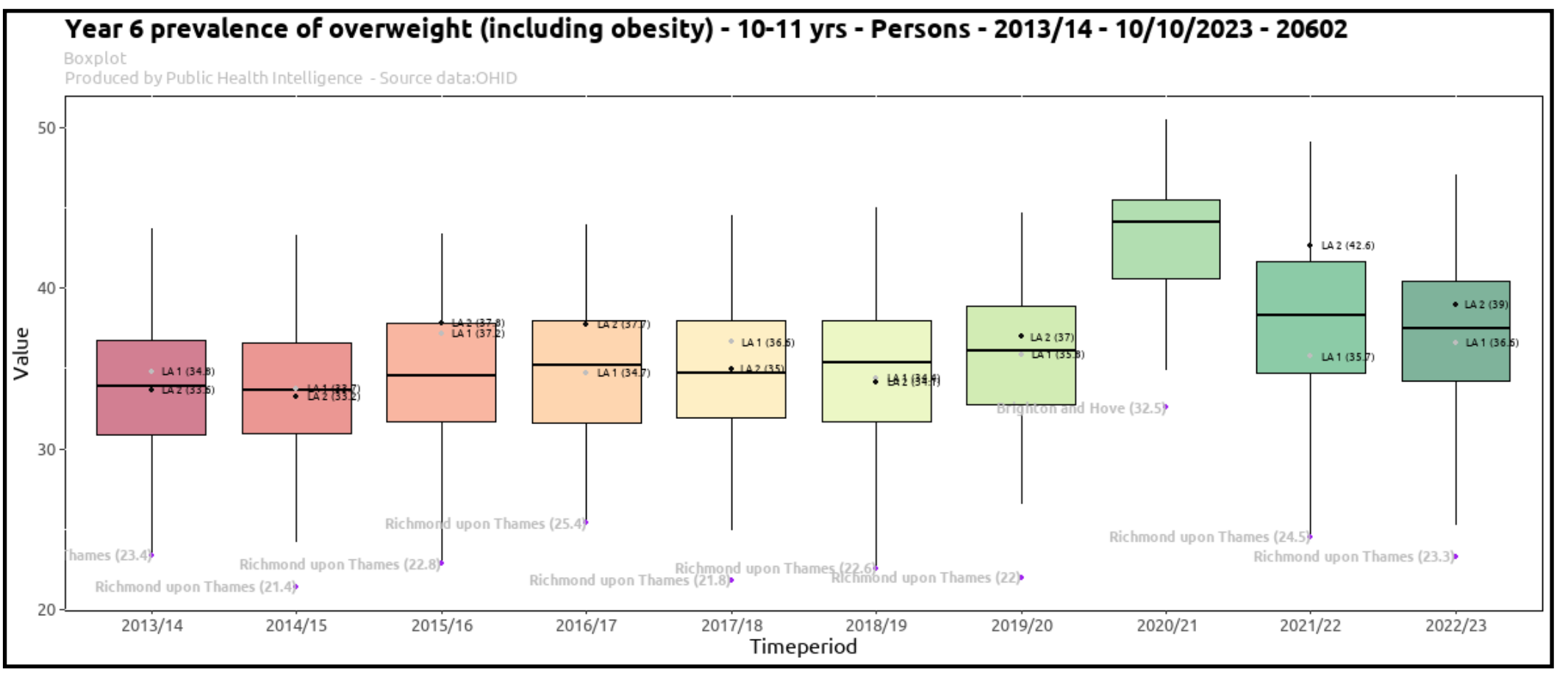


## Distributions

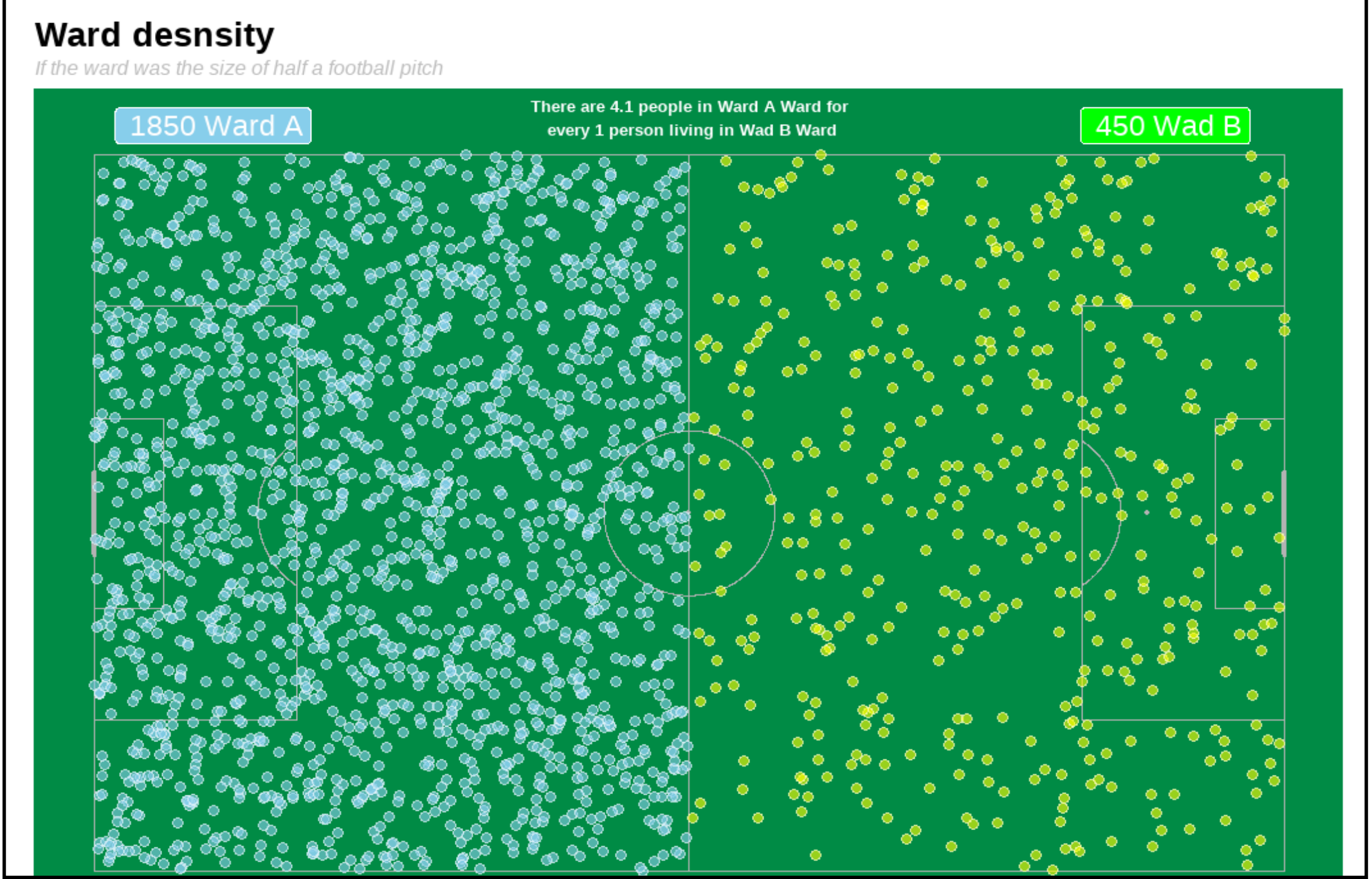
This column shows how I have used R to illustrate various distributions. The ridgeline graph below shows the English distribution results for overweight Y6 pupils (fingertips). This highlights how the mean England score has shifted over 10 years—the triangle is selected LA.



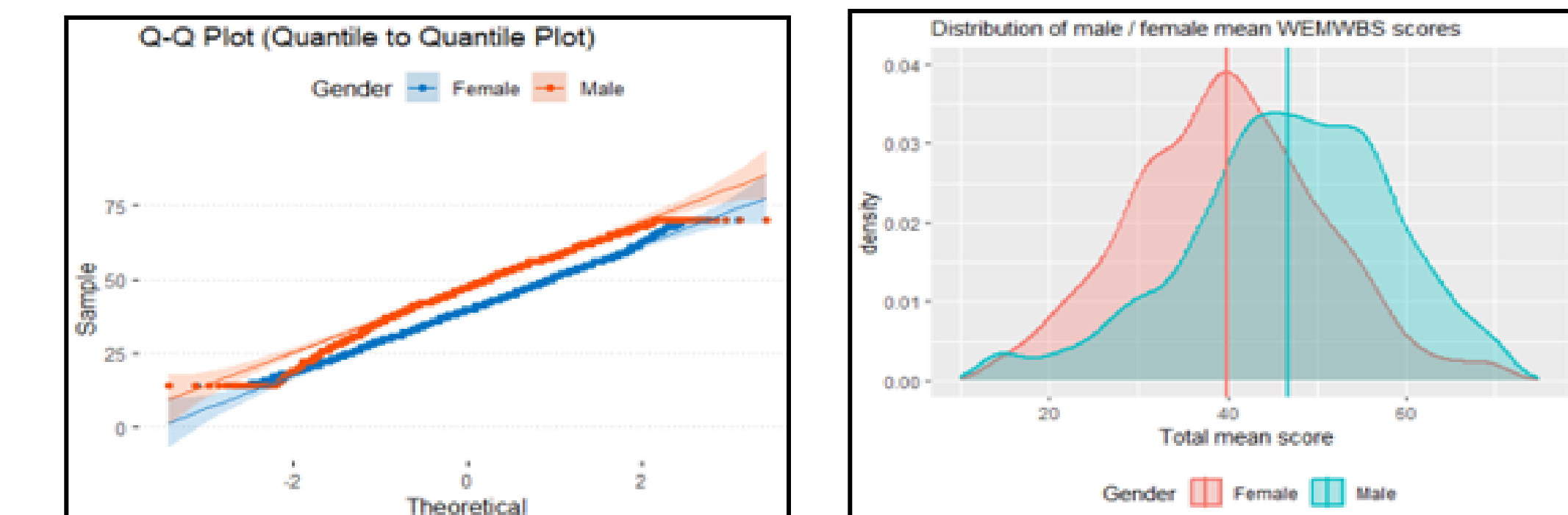
The same data are plotted in a box and whisker chart with the advantage that the inter quartile range is easier to identify. Bespoke data points can also be added. I have added the best performing LA and two LA's with their respective outturn positions.



This simple spatial data visualisation below requires two values to be input: (a) ward name and (b) population density. R then produces this visualisation to illustrate the differences in population densities. This uses a specific library called "Soccer", which provides template for this data visualization. The dots in this visualization have been made transparent, so its easier to see overlapping dots. The position of the dots are randomly generated.

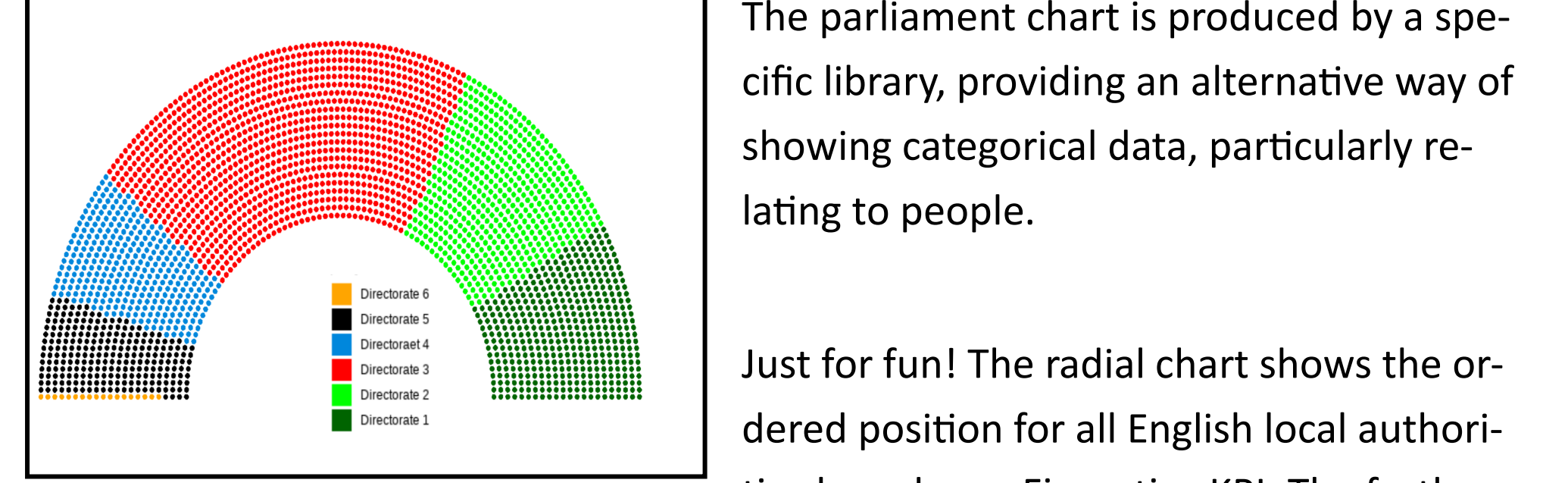
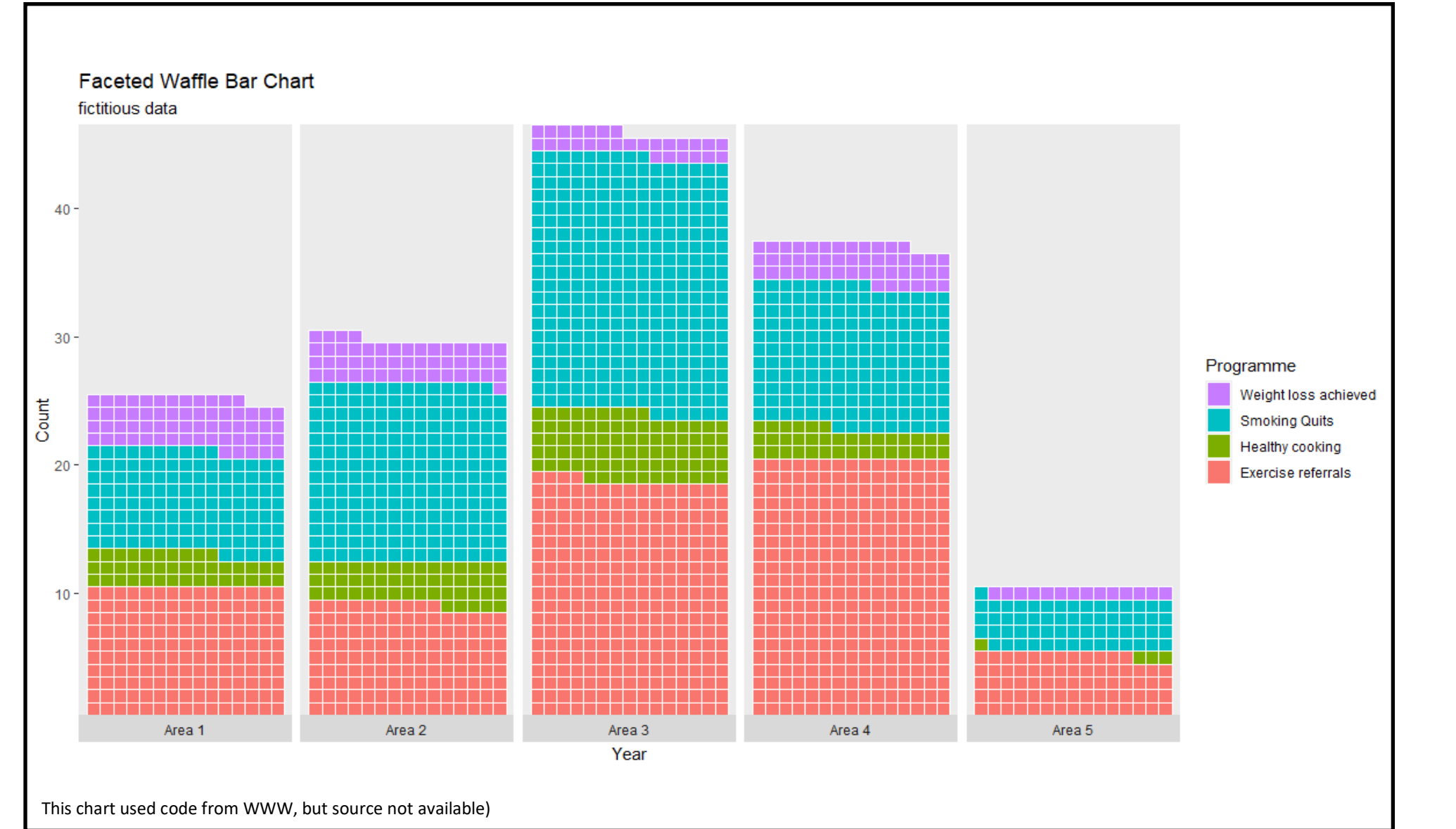


The two graphs below demonstrate how R was used to visually assess a population sample for skewness. Using transparency colours, the distributions were clearly illustrated, making it easier to see how the two distributions varied. Additionally, a Q-Q plot test was conducted in R to test for skewedness.

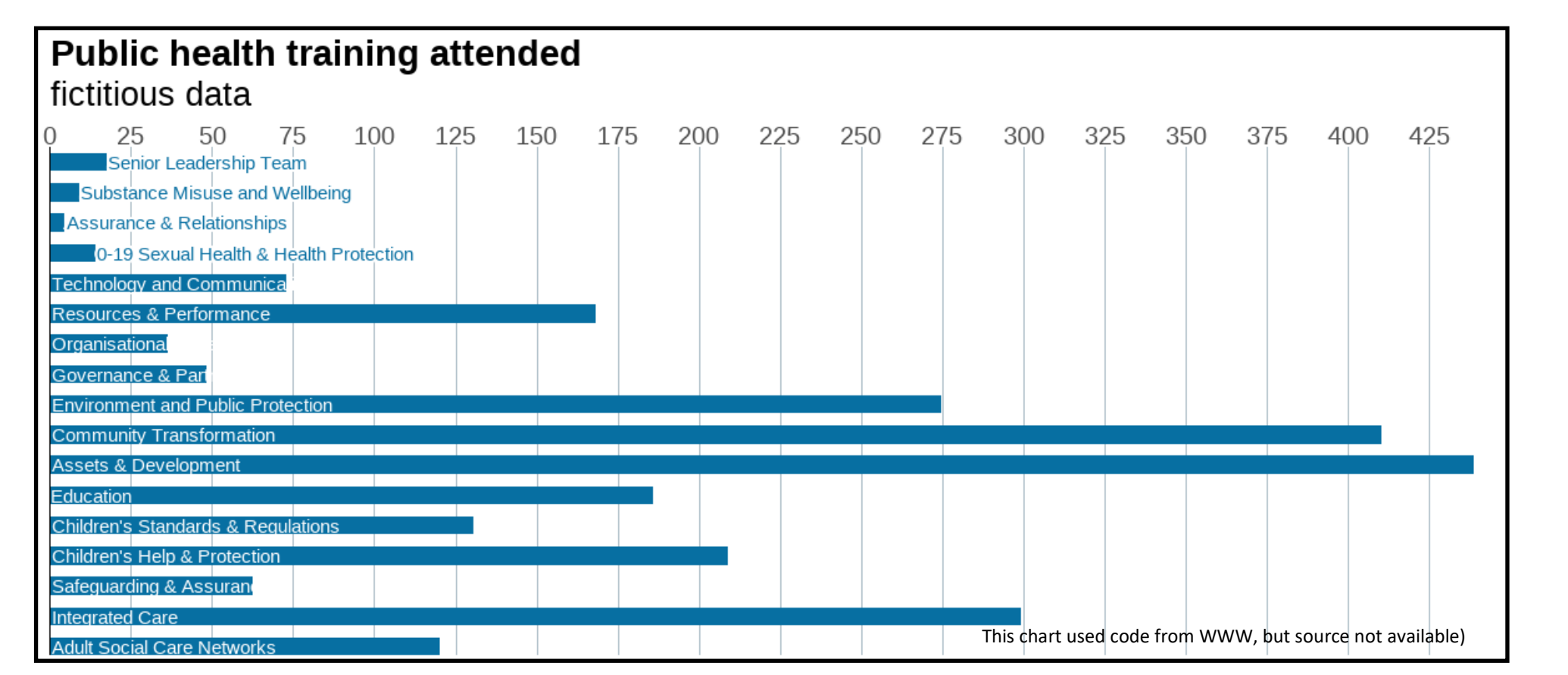
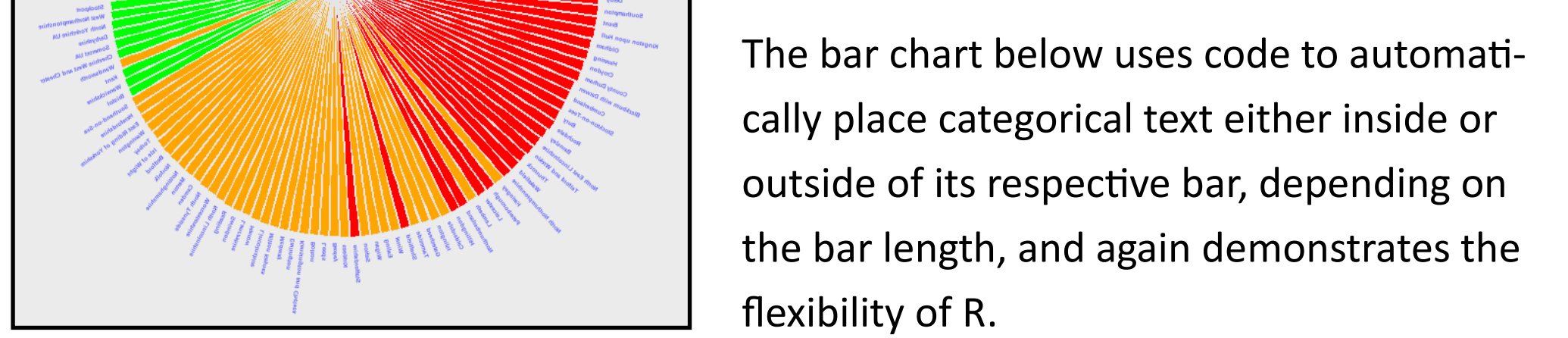


## Categorical Data

This column demonstrates how I have used R to illustrate categorical data. The first chart is known as a "Waffle Chart". I have set up this template in R, allowing the variables to be easily changed and the chart to be reproduced. This format is excellent for representing people or units as individual squares.



The parliament chart is produced by a specific library, providing an alternative way of showing categorical data, particularly relating to people.



## Conclusion

I have found R enjoyable but challenging to learn, especially the basics. It requires coding skills, which makes it quite different from Excel. However, the work and outputs can be tailored very precisely. Power BI is certainly easier for building interactive dashboards, and Excel's pivot tables allow for quicker data manipulation. However, R possesses an array of tools and packages for importing, cleaning, and wrangling data, making it a versatile tool. For example, R was able to analyse a complete survey with just a few lines of code.

I like that data can be wrangled using R without ever altering the source data. Cleaning data, creating new columns, and performing calculations is a relatively straightforward process. One of the main advantages I have found is R's ability to handle and analyse very large datasets. This allows for the analysis of Fingertips data, which is otherwise not available on the website. There are numerous web resources available to support learning and using R effectively