

UpAndDownPlots: An R Package for Displaying Absolute and Percentage Changes

by Antony Unwin

Abstract UpAndDown plots display the ups and downs of sector changes that make up an overall change between two time points. They show percentage changes by height and absolute changes by area. Most alternative displays only show percentage changes. UpAndDown plots can visualise changes in indices, in consumer markets, in stock markets, in elections, showing how the changes for sectors or for individual components contribute to the overall change. Examples in this paper include the UK's Consumer Price Index, Northern Ireland population estimates, and the German car market.

1 Introduction

There are many situations when overall percentage changes between two time points are broken down by subgroups. When share indices go up or down, news reports refer to which sectors and shares moved most. Governments produce many official indices, for instance the Consumer Price Index or CPI, and readers want to know what has driven the overall change in the index, which components have changed most. When companies study the consumer markets they supply, such as energy or cars, they are particularly interested in how well their own products have done and how their market shares have moved. When changes are discussed for a country as a whole, then people usually want to know what changes occurred in the individual regions. In all these cases, the percentage changes of the components are generally considered first, although the importance of a particular component, how much it makes up of an index or market, is also highly relevant. A small company which leaps in value, say an IT startup, may be a positive surprise, but it does not move the stock market the way a smaller change for a big company like Microsoft or Apple can. The same goes for fringe products in a consumer market. A specialist chocolate bar may double its sales thanks to a promotional campaign without selling nearly as many additional bars as would match a 1% increase in Mars bar sales. Percentage changes for subgroups are often displayed using lengths, with one bar for each sector and all bars having the same width.

1.1 UpAndDown plots

Only displaying percentage changes ignores how important those changes might be. UpAndDown plots use bar heights to represent percentage change and bar widths to display the initial contributions of sectors. Bar areas then represent absolute changes, so that amounts can be compared as well as percentages. Plotting in this way means that the length of the horizontal axis represents the sum of the individual sector contributions at the first time point, rather like a horizontal stacked barchart of no width.

It turns out that area in these examples has a valuable conservation property: if the bar rectangle for a sector is split up horizontally into subsectors and individual bars drawn for these, then the total area of the bars for the subsectors equals the area of the original bar for the whole sector. This means that you can drill down across UpAndDown plots. Displays of multiple levels are consistent, a valuable property when interpreting them.

A new package is required for drawing UpAndDown plots because of the need to draw bars whose width depends on the value of a variable. Flexible as it is, [ggplot2](#) does not offer this functionality directly. [UpAndDownPlots](#) is based on [ggplot2](#), adding the capability to handle hierarchies as well as tools for ordering and sorting the plots.

1.2 Related graphics

UpAndDown plots use both length and area to represent changes. Length is preferable, as [Cleveland and McGill \(1987\)](#) discussed long ago, but area has been used effectively in several displays, especially for different kinds of mosaicplot ([Hofmann, 2003](#)) and for treemaps ([Shneiderman, 1992](#)). Area is necessary together with length in UpAndDown plots because two different values are displayed.

Two early examples of area-bar charts or sky-line charts can be found in [Karsten \(1923\)](#). One, of average income by occupation, even includes a second level. These examples only have positive

lengths, but [Brinton \(1939\)](#) includes an example in a chapter on area-bar charts of negative values: price distortion by company sales volume. Some recent examples of area-bar charts can be found in [MacKay \(2009\)](#).

Doubledecker plots ([Hofmann, 2008](#)), a version of mosaicplots, are used for displaying proportions in subgroups by shading the appropriate height of a bar representing the whole group. All bars are the same height and bar widths show the subgroup size. Doubledecker plots are not suitable for datasets with negative values or values of over 100%. Proportions are highlighted as part of a whole. Usually there are small gaps between the bars to separate the subgroups. UpAndDown plots could be used for these kind of data, but the upper limit of 100% would not be emphasized. On the other hand they would be effective for data sets with low percentage rates such as rare diseases or unemployment rates where the upper limit is not relevant. Treemaps are useful for displaying overall structure but can only show negative values using colour, e.g., as with [treemap \(Tennekes, 2023\)](#).

Some demographers have used the term Skyline plots to refer to plots of population size over time, where the size is often constant over periods of time ([Pybus et al., 2000](#)). These plots generally have no vertical lines, the heights are never negative, and they may include uncertainty intervals. UpAndDown plots can look similar, but include vertical lines, the heights can be negative, and the horizontal axis represents shares not time.

The data structure discussed so far has assumed that only two fixed time points are considered, an initial base point and a final point. There may be several time points over a longer period and then a display of a succession of changes is wanted. Time series graphics could be used, possibly with line thickness representing amounts. For seasonal data, such as ice-cream, there might be interest in monthly changes, differences to the same month last year or differences to year-to-date last year.

1.3 Outline of article

The type of data that can be visualised by UpAndDown plots is discussed first, including what classification hierarchies can arise. The next section presents the mathematics underlying the plots and shows how a particular version of the plot can be used to display changes in brand shares in a market. There is a section on using the package and one on varying UpAndDown plots using different orderings of levels, different sortings within the levels, and colour. There are three detailed applications with annotated code snippets to show how the package can be applied in practice. A summary section concludes the paper.

2 What kind of data can be visualised in an UpAndDown plot?

To display changes between two time points, data are needed for both end points for all items. The simplest example is a consumer market in which performance is measured by unit sales. An UpAndDown plot can be drawn if there is a complete list of the products in the market and the unit sales of the products at the initial and end time points are available. All products are treated equally and have the same weight. There are two main advantages of UpAndDown plots. They show both percentage and absolute changes for each product and they allow the user to drill down through classifications to show changes within levels as well as across levels.

More complicated examples involve unequal weightings and multiple grouping variables. Share indices are commonly constructed by weighting share prices by the numbers of issued shares, i.e. it is the company capitalisations that make up the index not the raw share prices. UpAndDown plots display the changes in capitalisations. As long as the numbers of shares do not change (because of a rights issue or a company buyback plan) the percentage change in a company's capitalisation will be the same as the percentage change in share price. A further weighting variant arises with governmental or other indices that are weighted sums of components. The CPI is an example. An individual component's contribution to the percentage change in the index depends on the weighted value at the first time point not just on the weight.

Sometimes a set of components is changed, for instance when the CPI is redefined or when stocks are removed from a financial index and others added or when new products are introduced into a consumer market, for example a new chocolate bar. In all these situations it is not easy to make comparisons at the component level, as corrective adjustments are necessary. UpAndDown plots display changes and the changes must be for comparable data (or for data that has been made comparable). The CPI uses chaining to maintain consistency ([ONS, 2014](#)).

2.1 Multiple classification levels and nesting

Subgroups and components can be hierarchical (nested), as in the CPI, which has Sector, Subsector, and Component levels, or have no predetermined order. The Northern Ireland Statistics and Research Agency produces population estimates (NISRA, 2019). They look at population over time broken down by four classifications: age (four groups), gender (two), LGD2014_name (Local Government Districts or LGDs, eleven), and area_name (District Electoral Areas or DEAs, eighty). Each DEA is a subarea of one LGD, so those two are nested (or form a hierarchy) while the others do not. If both levels of a nesting are to be plotted, then the higher level must come first.

If levels form a complete hierarchy, as with the CPI, then there is only one possible ordering of levels, but many different orderings are possible within the levels. With a partial hierarchy, as with the Northern Ireland data, and assuming the three levels to be plotted were gender, LGD2014_name and area_name, then the ordering of the levels would have to have gender either first or last, keeping the nested levels together. If there is no nesting and p classification variables then there are $p!$ orderings of the levels and, again, many different orderings within the levels. In these situations it makes sense to order consistently within the levels. Suppose age and then gender are plotted for the Northern Ireland data. Plotting male and then female for the three youngest age groups, but female and then male for the oldest age group would just be confusing.

An additional, unusual form of nesting is “double-nesting”, where one grouping variable is separately nested in each of two other variables. This can arise with car sales, if models are nested in both market segment and manufacturer. Suppose there are 3 sectors and 5 manufacturers, where each manufacturer offers several models across the sectors. There is a hierarchical relationship between sectors and models (each model is in only one sector) and another between manufacturers and models (each model is produced by only one manufacturer). The model level should then be placed last.

The package’s `ud_prep` function includes a function that checks what kind of nesting, if any, exists in the data. It also checks that the ordering of levels specified by the user is permissible, and calculates the relevant statistics. One of the package vignettes discusses nesting and includes plot examples.

2.2 Dataset structure

An UpAndDown plot dataset must contain at least two numeric variables representing the item values at the start and end points. There should be at least one classification variable defining the groups. If items are not equally weighted, then there has to be a weight variable. There is no need to provide differences or percentage changes, as the package calculates these from the data as required. The CPIuk dataset includes component index values for August 2017 and August 2018, three classification variables (Sector, Subsector, and Component), and a weighting variable. The AutoSalesX dataset for the German car market includes sales in 2017 and 2018, the three classification variables of Sector, Segment, and Manufacturer, and no weighting variable. Here is an example taking the first few lines of AutoSalesX.

```
#>   Sector Segment Manufacturer sales17 sales18
#> 1   Car Compact    ALFA ROMEO   1171    1008
#> 2   Car Compact         AUDI  49820   45901
#> 3   Car Compact         BMW   86843   82332
#> 4   Car Compact    CITROEN   8390    6308
#> 5   Car Compact     DACIA   31920   32659
#> 6   Car Compact     FIAT   14187    8181
```

3 Mathematics of UpAndDown plots

All the bars in a barchart of percentage changes have the same width. In an UpAndDown plot each bar for a sector has a width proportional to the sector’s importance at the initial time point. This allows users to drill down consistently to study changes within sectors. The key property is that area is conserved, i.e., that the sum of the areas of subsector bars equals the area of the sector’s bar. Consider a dataset of sales of cars and vans. If the cars subsector goes up by 10% and the vans subsector by 20%, how much would the combined sector of cars and vans together change? If nine times more cars than vans were sold last year then it would be 11%. UpAndDown plots display changes for sectors and subsectors appropriately. Regardless of how the sectors are broken down or grouped, the total area remains constant. Other plots commonly in use cannot manage this. Area conservation can be shown as follows.

Assume an index is made up of m components with weight w_i for component i and values v_{i1} and v_{i2} recorded at two time points, t_1 and t_2 .

The index value at time t_j is

$$T_j = \sum_{i=1}^m w_i * v_{ij}$$

The overall absolute change over the interval (t_1, t_2) is

$$C = \sum_{i=1}^m w_i * (v_{i2} - v_{i1}) = T_2 - T_1$$

and the overall percentage change is

$$pC = 100 * \frac{C}{T_1}$$

The corresponding overall and percentage changes for the individual components are

$$c_i = (v_{i2} - v_{i1})$$

and

$$pc_i = 100 * \frac{c_i}{v_{i1}}$$

3.1 Area conservation

A single bar with the overall percentage change, $y = 100 * \frac{T_2 - T_1}{T_1}$, as height and last year's index value as width would have an area of

$$100 * (T_2 - T_1)$$

Individual bars can be drawn for each component i with widths $w_i * v_{i1}$, the components' contributions to the index value at time t_1 , so that their total width is

$$\sum_{i=1}^m w_i * v_{i1} = T_1$$

the same as the width of the single bar for the whole index. If the individual bars are drawn with heights pc_i , the percentage changes for the components, then the area of each bar is

$$a_i = (w_i * v_{i1}) * pc_i = 100 * w_i * (v_{i2} - v_{i1})$$

the weighted absolute change for component i . The sum of the areas of the m individual bars is

$$\sum_{i=1}^m a_i = 100 * \sum_{i=1}^m w_i * (v_{i2} - v_{i1}) = 100 * (T_2 - T_1)$$

the same as the area of the bar for the whole index.

3.2 Scales

The default vertical scale for an UpAndDown plot depends on the largest positive and negative percentage changes. Sometimes there are small components with very large percentage changes that are not of major importance and the resulting scale masks the details of the rest of the display. One way to get round this is to limit the scale using the `vscale` option in the `ud_plot` function. This will trim any bars with larger heights at that value. (Using the `ggplot2` option `ylim` would exclude components with larger heights.) Any large percentage changes still stand out, but no longer affect the bulk of the display.

3.3 Change of baseline and market share change

UpAndDown plots are drawn by default with a baseline of 0 marking no change. In some cases, it is more informative to use an alternative baseline, such as the overall market change. With that choice, the plot emphasizes which components performed better than the market and which performed worse. Area is still conserved if the baseline is set differently to 0, as all bar widths remain as they were and all heights are changed by the same amount.

If the baseline is set at the overall percentage change $y\%$, then the area representing the total market change is 0. The sum of the areas representing the components' changes is

$$\sum_{i=1}^m (w_i * v_i) * (pc_i - y) = 100 * (T_2 - T_1) - y * \sum_{i=1}^m (w_i * v_i) = 0$$

The baseline influences the look of the graphic and also the interpretations of the heights and areas of the bars. Using the default baseline of 0 means that a bar's height represents the percentage change between t_1 and t_2 while the bar's area represents the absolute change. If the overall percentage change is used as the baseline then a bar's height represents the percentage change relative to the overall change and the bar's area represents the absolute change relative to the overall absolute change. In fact there is more to it than that. The bar's area is proportional to the component's share change.

In consumer markets there can be more interest in changes in market share than in absolute changes. Given sales for brand i of v_{ij} in period j , the size of the market is $T_j = \sum_{i=1}^m v_{ij}$ and the market share of brand i is

$$s_{ij} = \frac{v_{ij}}{T_j}$$

while the change in market share between the two periods is

$$s_{i2} - s_{i1} = \left(\frac{v_{i2}}{T_2} - \frac{v_{i1}}{T_1} \right) = \frac{v_{i1}}{T_2} \left(\frac{v_{i2}}{v_{i1}} - \frac{T_2}{T_1} \right)$$

If the baseline of each brand's bar is set to be the overall change in sales in the market, $\frac{T_2 - T_1}{T_1}$, then a bar going up shows a performance better than the market and means a gain in market share while a bar going down indicates a performance worse than the market and means a loss in market share. The height of the bar of brand i is the performance difference compared with the market

$$\left(\frac{v_{i2} - v_{i1}}{v_{i1}} \right) - \left(\frac{T_2 - T_1}{T_1} \right) = \left(\frac{v_{i2}}{v_{i1}} - \frac{T_2}{T_1} \right)$$

and the area of the bar is

$$v_{i1} * \left(\frac{v_{i2}}{v_{i1}} - \frac{T_2}{T_1} \right)$$

which is T_2 times the change in market share, so the bar areas using the baseline of overall changes in sales represent the market share changes. The height of a rectangle in the UpAndDown plot shows the percentage brand change less the market percentage change, while the area of the rectangle reflects the change in a brand's market share. There is an application to the German car market in §8.

4 The package

The **UpAndDownPlots** package is available on CRAN (Unwin, 2024b). There are two main functions, `ud_prep` and `ud_plot`. `ud_prep` prepares the data for plotting. First it checks the input parameters and data. It then calculates the statistics needed for sorting, and sorts the (at most) three levels of classification. There are five sorting options that may be chosen at each level and an option as to whether they should be displayed in reverse or not. `ud_plot` draws the plot. After checking the input parameters, it calculates the cumulative statistics needed for drawing bars with different widths. The layered structure of **ggplot2** is used to draw a barchart layer for each of the levels specified. Two plots are prepared, a horizontal plot to display increases going upwards and decreases going downwards, and a vertical one, so that labels can be added to the bars of one of the levels. Finally, a dataset combining the statistics derived for drawing the plots and the original data is constructed.

Three further functions are included: `ud_colours` to allow the user to specify the colours used in plots; `sort5` to draw a set of plots for one level of changes to compare the outputs of the five sorting methods available; `dgroup` to compare the effects of different grouping variables by drawing a one-level plot for each one. The package help covers the functions in detail and there are several illustrative vignettes.

Three datasets are included: `NIpop` provides Northern Ireland population estimates over seven years from 2011 to 2017 by age, gender, Local Government District, and District Electoral Area; `CPIuk` provides Consumer Price Index values for the UK from 2017 and 2018 by Sector, Subsector, and Component; `AutoSales` (and the cleaned subset `AutoSalesX`) provide data from the German car market for the years 2017 and 2018 by Sector, Segment, `ModelSeries`, `Manufacturer`, and `Model`.

4.1 Drawing an UpAndDown plot

Drawing a plot with several layers is easy in `ggplot2` (Wickham, 2022), provided the layers are independent of one another. In UpAndDown plots they are not. Each level adds more detail, conditional on the higher levels already specified.

Ordering, sorting, and arranging are key elements in UpAndDown plots. They determine how attractive and informative a plot is. The classifying variables are ordered either by definition (as with a fully nested hierarchy) or to display particular conditional groupings. Classification variable categories are sorted to support comparisons between them. Finally, the graphic layers comprising the plot are arranged to convey the information as effectively as possible.

There are seven steps in preparing and drawing an UpAndDown plot:

1. Order the grouping levels, respecting any nestings that exist.
2. Calculate the statistics at each level that are used for sorting within levels.
3. Sort each level, respecting the natural orders of any grouping variables (e.g., age).
4. Calculate the data needed for drawing bars in levels. As bars have different widths, the bases for them are calculated cumulatively from left to right. The values will depend on how the level has been sorted.
5. Decide whether to fill the bars of one level with colour and define an appropriate colour palette if the default does not suit.
6. Prepare individual graphic layers. Given the order within a level and the colour choices across levels, each layer can be drawn.
7. Choose in which order to draw the levels. The default is to draw the levels in grouping order (`drawFrom="BigToSmall"`) but the software allows the reverse order ("`SmallToBig`"). Layers drawn last are more visible. Layers with filled bars may cover information in other layers if not drawn first.

Although drawing more than three levels is not currently offered, as plots become overloaded, it would be possible. Statistic calculations and sortings of bars could be carried out for any number of levels. There are three possible alternative graphics for each individual layer, one with grey-filled bars, one with colour-filled bars and one with transparent bars. These would be prepared in step 6 to provide all options needed to put the layers together in step 7.

5 Varying UpAndDown plots

By default, an UpAndDown plot includes a red dotted line to indicate the overall percentage change. Using this as the baseline for the bars gives the plot a different interpretation, covered in §3.3. Other variants can be drawn by reordering classification levels and using colour. Drawing more than one level of subdivision on the same plot requires choosing suitable shading and transparency to keep the different levels perceptually separated. It is usually most effective to draw the lowest level first, but it can depend on which features are to be emphasized. As so often with graphics, the exception proves the rule. In principle, when there are few groups at each level, more levels could be displayed, but this can become confusing and the software is currently restricted to displaying at most three levels.

5.1 Orderings of UpAndDown plots

The same data can give rise to many different UpAndDown plots and it is worth choosing carefully from amongst a selection of versions, as different information will be conveyed by each display, sometimes more effectively, sometimes less. Figure 1 shows the Consumer Price Index in the default order provided by the UK Office for National Statistics. Figure 2 shows the index ordered by percentage changes of sectors and by percentage changes of components within sectors.

The individual categories of a level can be ordered in several ways using the R package UpAndDownPlots:

- `orig` original, as components arise in the dataset. Consistent orderings are important when comparing plots.
- `base` by initial size (bar width). This orders in terms of the initial importance of the components, emphasizing how the biggest and smallest changed.
- `final` by final size. This orders in terms of the final importance of the components.
- `perc` by ascending percentage change (bar height). Which components went up and which went down? Which components had the biggest relative change?
- `abs` by ascending absolute change (bar area). Which components had the biggest absolute change?

Other orderings could be considered too, using properties of the components, or even just an alphabetic ordering. They can be calculated in advance and used to order the dataset before applying the functions in the package. Setting the sort option to `orig` retains an ordering calculated in advance. This was done for the plot on the right of Figure 6, so that it could be sorted by the size of the changes in market share.

If the components are grouped into higher levels as with the CPI's subsector classification, then the components can be sorted by their subsector and then, within that, by one of the five approaches listed already. It is usually best to use the same sorting for each subsector for consistency of interpretation. A different situation arises when the variables defining the levels do not form a hierarchy. For instance, in the car sales dataset vehicles are split by Sector and by Manufacturer. A display could have a top level of Manufacturer and then split by Sector or order the levels the other way round. There is no nested hierarchy and however the top level is ordered, it is best to have the lower level ordered the same way for each top level grouping. Having the lower level consistently ordered for each top level category makes comparisons easier. A possible lower level ordering would be to take whatever ordering you would use if the lower level were the top level.

If inconsistent lower level orderings are desired, then it is best to define the required nested hierarchy by renaming the lower level categories accordingly. For the cars data this would mean using new classifications `car_VW`, `car_Mercedes`, `van_VW` etc if you split on Sector first, or `VW_car`, `VW_van`, `Mercedes_car` etc if you split on Manufacturer first.

If there is a percentage change of zero then a bar has no height or area, but the width shows its size. If a new component is introduced after the initial time, for instance a new car, then there is no initial weight and an infinitely tall bar. Cases like this must be excluded and a note should be added to the display.

5.2 Colour in UpAndDown plots

The default is to fill the bars of the top level with grey and draw transparent unfilled bars with either blue or brown borders on top. The percentage change for the whole system is marked with a red dotted line. Different colours may be used to fill bars for one of the levels drawn. In that case the top level bars are not filled with grey. Colour schemes are helpful in distinguishing sectors and particularly informative when associated with different groups (e.g., for political parties, as in Chapter 26 of [Unwin \(2024a\)](#)). Colours are specified in the function `ud_colours` and the default colour choices can be changed there as required. The help page for `ud_plot` gives the details.

5.3 Limitations of UpAndDown plots

UpAndDown plots are designed for comparing changes between two time points. They have not been extended to handle changes over several time points.

Cases with missing data have to be excluded. Cases with initial values of 0 cannot be handled as any positive value at the end time point will give an infinite percentage change. Examples might be a new model of car, a new chocolate bar or even a new political party. Cases with very low initial values and high final ones can also lead to very high percentage changes, which then have to be trimmed as described in §3.2. Redefinitions of classification variables (e.g., in the car market when SUVs became common) or reclassifications of items (e.g., deciding a car now belongs to a different group) can cause problems. These are typical issues for any dataset involving data over time and are not restricted to UpAndDown plots.

All three applications in the following sections include large numbers of individual items or components. This means that plots of the full dataset have to be drawn very large or only display data at higher grouping levels. For some applications (e.g., the car market in §8), smaller manufacturers can be grouped together into a group called `Other`.

6 Visualising changes in the UK Consumer Price Index

The Consumer Price Index (CPI) summarises how prices change for a basket of a wide range of products and services. It is an important measure of inflation and influences both government policy and public attitudes. The basket is continually reviewed and revised as it is supposed to cover typical costs of living. It is said that Margaret Thatcher, when she was UK Prime Minister, supported the basket being amended yearly to reflect as quickly as possible changes in the public's spending habits. This meant that the public's switching to cheaper alternatives was identified and included more quickly, having a lowering effect on the index.

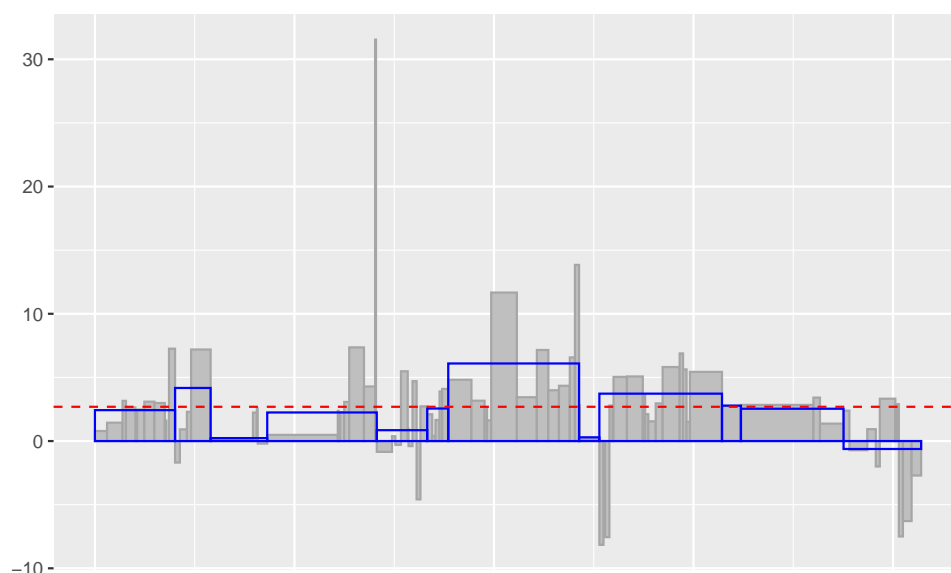


Figure 1: UK Consumer Price Index changes August 2017 to August 2018. The red dotted line shows the overall percentage change. The rectangles with blue borders show the changes of the main sectors. Heights are percentage changes, widths are index values in August 2017, so that the areas of the rectangles are each group contributions to the overall change. The grey bars show the percentage and absolute changes for the individual components.

Every country produces its own CPI and they have many different ways of displaying information on how their index changes between two time points. Some use barcharts (e.g., [National Bureau of Statistics of China \(2024\)](#) and [US Bureau of Labor Statistics \(2023\)](#)). France also has stacked bars ([Banque de France \(2017\)](#)). Germany includes a kaleidoscope plot ([Statistisches Bundesamt \(2024\)](#)). The UK includes multiple time series ([Office for National Statistics \(2024\)](#)) as well as different kinds of barchart. Most of these displays just show percentage changes. A few have accompanying displays of how big the sectors comprising the index are. What is needed is a display showing both individual percentage changes and their contributions to the change in the index (i.e., the individual absolute changes). This is what UpAndDown plots do.

In the UK's CPI of 2017 there were 12 main groups with Transport having the biggest weight (156 out of 1000) ([ONS, 2019](#)). All of these groups, except Education, were further subdivided into up to 7 subgroups and the subgroups were subdivided yet again into up to 9 individual components. Overall there were 85 individual items with weights ranging from 1 (e.g., solid fuels) to 86 (Restaurants & Cafes). Figure 1 is an UpAndDown plot of the CPI changes over one year plotted horizontally for comparison with the kind of standard barchart that is often used. Horizontal plots are useful for giving an overview of the data before deciding what to concentrate on in detail.

To draw the plot, first load the package with

```
library(UpAndDownPlots)
```

Prepare the data for plotting using the `ud_prep` function. `CPIuk` is the dataset, `weight` is the vector of individual component weightings, `v1` is the initial component value and `v2` the final one. Two levels are to be plotted, `Sector` and `Component`, both are to be left in their original order.

```
yp <- ud_prep(CPIuk, weight="Weight", v1="Aug2017", v2="Aug2018", levs=c("Sector",
"Component"), sortLev=c("orig", "orig"))
```

`ud_plot` takes the data and prepares the plots and data summaries. The parameter `drawFrom` specifies in which order the levels are plotted, in this case the lower `Component` level first and the higher `Sector` level on top.

```
yd <- ud_plot(yp, drawFrom="SmallToBig")
```

A horizontal unlabelled UpAndDown plot has been chosen.

```
yd$uad
```

It can be seen that only one sector declined, the one on the far right that turns out to be Miscellaneous goods and services, that one component (liquid fuels) had a much higher jump in price, over 30%, than any other, and that the components of sectors did not change uniformly. Plots showing only percentage changes would show neither the relative importance of the separate parts of the index nor that liquid fuels are only a small component of the index. They would emphasize neither the

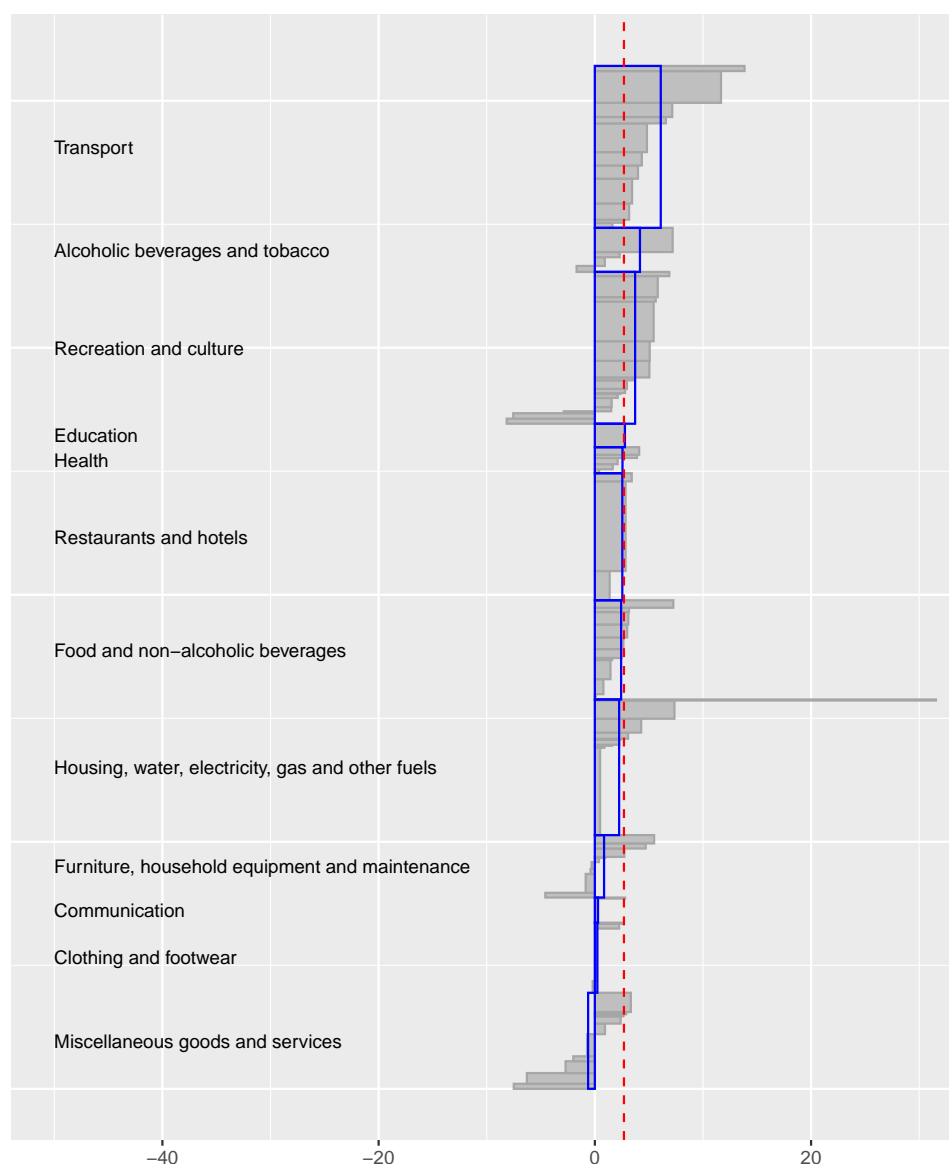


Figure 2: CPI changes ordered by percentage changes of sectors and by percentage changes of components within sectors. Two sectors, Transport and Recreation and Culture, made the biggest contributions to change. Prices increased for all components in most sectors.

hierarchical structure of the index nor the relative importance of components within sectors.

There are a number of ways UpAndDown plots can be varied and extended to make them more informative: the plot can be drawn vertically with labels for either sectors or their components; the sectors and their components can be sorted in various ways; colour can be employed to distinguish sectors. Figure 2 shows a version ordered by percentage changes of sectors and by percentage changes of components within the sectors. The plot has labels for the sectors and is drawn vertically to make those labels readable.

The preparatory code with `ud_prep` defines the sorting of the levels using the parameter `sortLev`.

```
yq <- ud_prep(CPIuk, weight="Weight", v1="Aug2017", v2="Aug2018", levs=c("Sector",
"Component"), sortLev=c("perc", "perc"))
```

Labels are added for the sectors using the parameter `labelvar`. Limits have been set for the percentage change axis to add sufficient space for the labels using the parameter `vscale`.

```
yf <- ud_plot(yq, labelvar="Sector", drawFrom="SmallToBig", vscale=c(-30, 30))
```

A vertical labelled UpAndDown plot has been chosen.

```
yf$uadl
```

There are 85 components at the lowest level of the CPI, too many to label effectively unless the plot is drawn very tall. Plots like Figure 2 (and indeed Fig 1) are drawn to give an overview to help

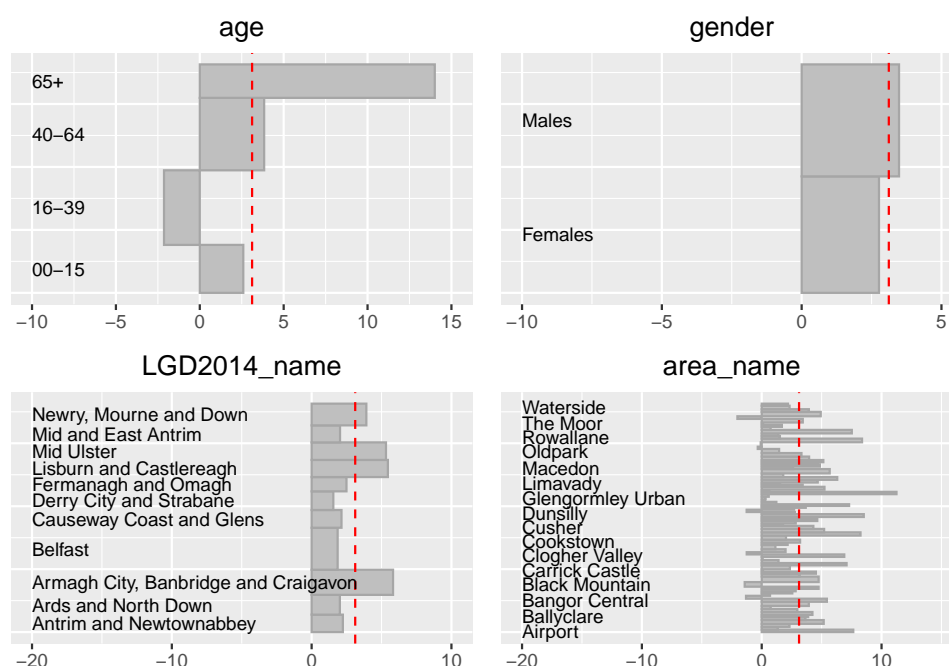


Figure 3: Percentage changes in population between 2011 and 2017 by age, gender, district, and area.

decide what final plot or plots to draw. Perhaps only changes at the Sector level might be displayed, perhaps changes for the Componentlevel of one sector, say Recreation and culture, might be displayed.

7 Visualising population changes in Northern Ireland

National Statistics Offices study demographic changes to understand how their country's population is developing: which groups are growing or declining and how areas are changing. The Nipop data in the package come from the Northern Ireland Statistics and Research Agency and include population estimates for seven years. This section considers the changes between 2011 and 2017.

Figure 3 shows UpAndDown plots for each of the possible grouping variables. The code uses the `dgroup` function to produce two sets of UpAndDown plots for each of the individual classification variables specified. One set is labelled, the other not. The `grid.arrange` function from [gridExtra\(Augue, 2017\)](#) is used to plot the graphics together in a single display.

Age groups displayed the most variation with a high percentage increase amongst the 65+ group and a decline in the 16-39 group. Male and female changes were similar. All the districts increased in population, while some of the 80 areas actually declined in population.

```
library(gridExtra)
g4 <- dgroup(Nipop, byvars=c("age", "gender", "LGD2014_name", "area_name"),
             v1="y2011", v2="y2017")
grid.arrange(g4$uadgl)
```

There are many alternatives using more than one classification variable. If you plot the Northern Ireland population data changes grouped first by gender and then by age, there will be two bars in the gender layer and eight in the age layer. If you plot the changes grouped first by age and then by gender, there will be four bars in the age layer and eight in the gender layer. In both plots the eight bars in the lowest layer will be the same, but they will be grouped differently. Table 1 shows the absolute and percentage changes for the data: firstly for gender and then age by gender; secondly for age and then gender by age. Some conclusions can be drawn from the tables by careful inspection. More can be seen with graphics.

Figure 4 is a graphic for age and gender together. It reveals that there was a much bigger percentage increase in numbers of older men than in numbers of older women (albeit from a lower base), that there was little difference between males and females in the youngest group, and that the decline in 16-39 year-olds was greater for females than for males.

```
ag <- ud_prep(Nipop, v1="y2011", v2="y2017", levs=c("age", "gender"),
```

Table 1: Absolute and percentage changes in N. Ireland population estimates by age and gender.

			gender	age	change	% change
			Females	00-15	4950	2.67
			Females	16-39	-8790	-2.95
			Females	40-64	12960	4.45
			Females	65+	16470	10.97
			Males	00-15	4930	2.52
			Males	16-39	-3840	-1.31
			Males	40-64	9120	3.21
			Males	65+	20770	17.94
			age	gender	change	% change
			00-15	Females	4950	2.67
			00-15	Males	4930	2.52
			16-39	Females	-8790	-2.95
			16-39	Males	-3840	-1.31
			40-64	Females	12960	4.45
			40-64	Males	9120	3.21
			65+	Females	16470	10.97
			65+	Males	20770	17.94

gender	change	% change
Females	25590	2.77
Males	30980	3.48

age	change	% change
00-15	9880	2.59
16-39	-12630	-2.13
40-64	22080	3.84
65+	37240	14.01

```
sortLev=c("orig", "perc"))
kag <- ud_plot(ag, labelvar="age")
kag$uadl

These results may vary by the 11 districts and a plot of changes by district, age, and gender could be drawn. For the purposes of the article, three districts have been chosen: the cities of Belfast and Derry and the district of Newry, Mourne and Down. Figure 5 shows the plot. The two younger groups declined in population for both males and females in Derry and Strabane. In Belfast the number of male 16-39 year-olds actually increased, but the main difference to the other districts is the somewhat lower increase in the male 65+ group and the lack of any increase in the female 65+ group.

NIPopW <- NIPop %>% filter(LGD2014_name %in% c("Belfast", "Derry City and Strabane",
"Newry, Mourne and Down"))
zdag <- ud_prep(NIPopW, v1="y2011", v2="y2017", levs=c("LGD2014_name", "age", "gender"),
sortLev=c("perc", "orig", "perc"))
zdag2 <- ud_plot(zdag, labelvar="LGD2014_name", drawFrom="SmallToBig")
zdag2$uadl
```

In all the above three displays the levels have been sorted by percentage change, except for the age groups where, for obvious reasons, their original order has been retained.

8 Visualising brand share change in German car sales

An example of displaying brand share change in UpAndDown plots is shown in Figure 6 for the German car industry between 2017 and 2018. The patchwork package (Pedersen, 2024) is used to plot more than one graphic in a figure.

```
library(patchwork)
Only the compact market sector is to be displayed.
AutoSalesXcomp <- AutoSalesX %>% filter(Segment=="Compact")
The aim is to compare brand shares by Manufacturer, so that level is specified.
yxp <- ud_prep(AutoSalesXcomp, v1="sales17", v2="sales18", levs=c("Manufacturer"),
sortLev=c("perc"))
The ud_plot function prepares graphics for changes in sales.
yM <- ud_plot(yxp, labelvar="Manufacturer")
```

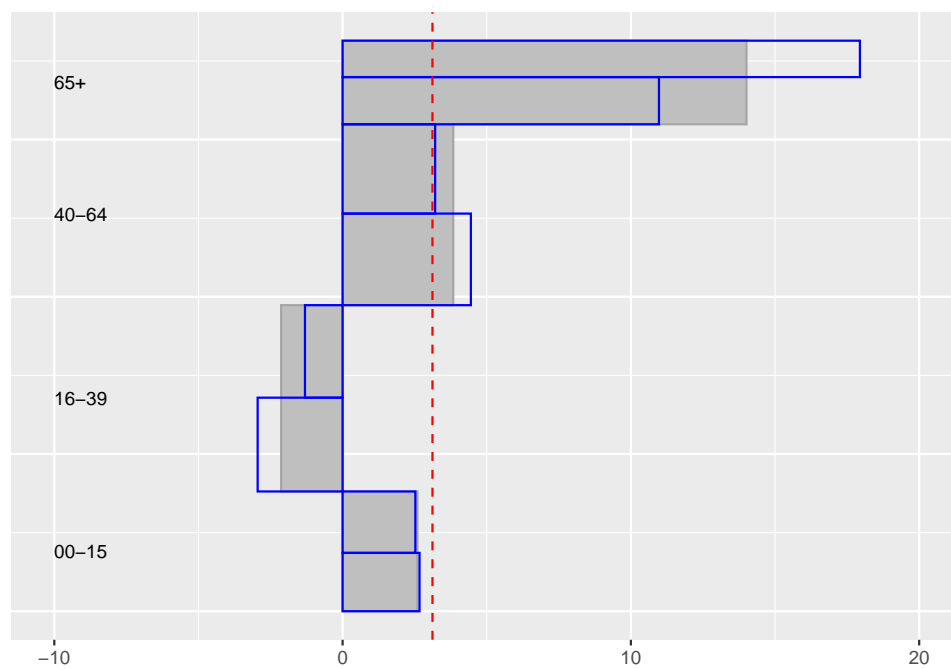


Figure 4: Percentage changes in population between 2011 and 2017 by age and then gender (males above, females below).

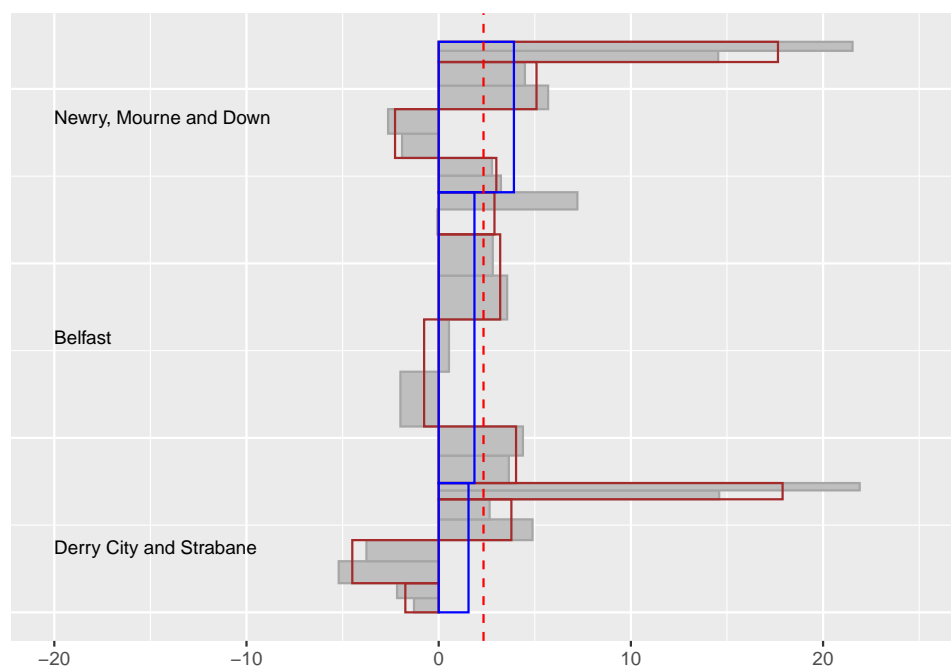


Figure 5: Percentage changes in population between 2011 and 2017 for three districts (changes outlined in blue), four age groups (outlined in brown), and gender (filled bars).

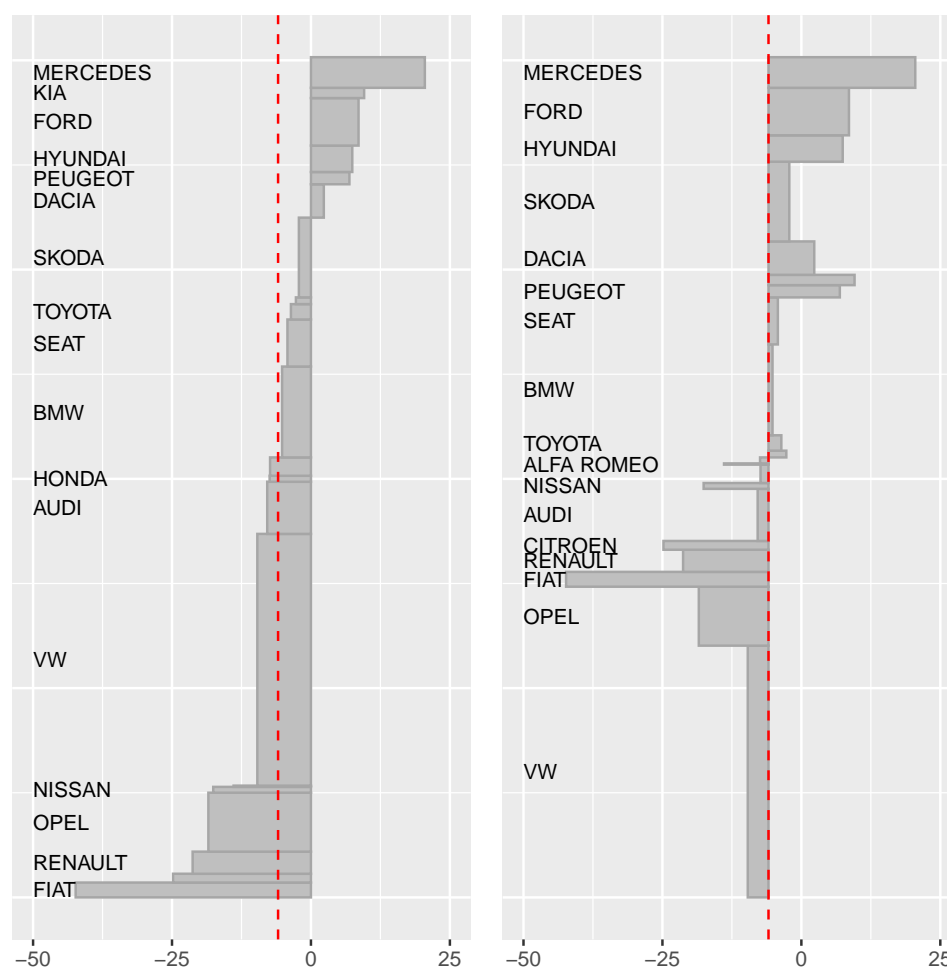


Figure 6: Compact car sales changes in Germany by manufacturer between 2017 and 2018. The plot on the left is drawn with a baseline of zero change and ordered by percentage changes. The plot on the right is drawn with a baseline of the overall market change and ordered by changes in market share.

The next code snippet calculates brand shares by manufacturer and the changes in them between the two years. It orders the data by these changes in brand share.

```
AutoSalesXcomp <- AutoSalesXcomp %>% mutate(S17=sum(sales17), S18=sum(sales18),
p17=100*sales17/S17, p18=100*sales18/S18, msch=p18-p17) AutoSalesXcompS <- AutoSalesXcomp
%>% arrange(msch)
```

The brand share data are prepared in the usual way but using the `b` parameter to set the baseline value at the overall percentage change in the market.

```
yxs <- ud_prep(AutoSalesXcompS, v1="sales17", v2="sales18", levs=c("Manufacturer"),
sortLev=c("orig")) yS <- ud_plot(yxs, b=yM$TotPerc, labelvar="Manufacturer")
```

The two labelled graphics are drawn side by side using [patchwork](#).

```
yM$uad1 + yS$uad1
```

Vehicles selling less than 1000 in both years have been reclassified as `Other` and the data have been aggregated by manufacturer. To avoid overlapping, not all manufacturers are labelled. The plot on the left uses a baseline of 0 and shows that the whole segment declined. It has been ordered by bar height, i.e. percentage changes. Of the bigger manufacturers, Mercedes and Ford had year-on-year increases. The plot on the right uses a baseline of the overall market change, a little under -5.9% . It has been ordered by bar area, using the areas proportional to market share changes. The calculations were carried out in advance and the manufacturers sorted accordingly, as can be seen in the code.

Mercedes and Ford gained the most market share and VW and Opel lost the most. Dacia gained in sales and market share, while Skoda lost sales, but still gained more market share than Dacia because they had a bigger starting value and performed better than the market. It would be difficult to compare the bar area sizes in the right plot, so ordering is essential (if those are the comparisons you want to make).

Area conservation holds for share changes as well. If a producer i sells k_i different products, then the sum of their bar areas using total market percentage change as the baseline equals the equivalent area for producer i as a whole.

Graphical displays are always relative. The visual representation of a statistic is proportional to the value of the statistic. Comparisons should only be made between lengths or areas that are on the same scales. The widths of bars in UpAndDown plots are always proportional to initial values (sales, market shares, volumes etc.), while the heights are proportional to percentages representing relative changes, and the areas are proportional to amounts representing absolute changes (if the baseline is 0) or share changes (if the base line is the total percentage change).

9 Summary

Displaying percentage and absolute changes between two time points in the same plot is very helpful, especially when individual components are of quite different sizes. Being able to drill down (or aggregate up) through multiple levels thanks to area conservation makes UpAndDown plots a powerful descriptive and exploratory tool. For the UK CPI dataset they showed that one sector actually declined in price and that individual components in other sectors also did. They showed that Transport and Recreation and Culture were the biggest drivers of inflation. The UpAndDown plots for the example of Northern Ireland population data showed the declines in the numbers of younger people and increases in the numbers of older ones, the increases in population across all districts, the different age patterns for males and females, and the decline in numbers of young people in the Derry Local Government District. The final example of the German car sales data showed that UpAndDown plots can be used to display brand share changes too. The conclusions were that although the overall market declined, two of the biggest manufacturers increased their sales. They also showed that the biggest loss in market share was for a manufacturer with a medium percentage loss of sales, because that manufacturer started from a high initial level of sales.

The display of market share changes in an UpAndDown plot is a surprising and powerful additional feature. In principle it would be possible to draw all three changes (percentage, absolute, and share) in the same plot. Be that as it may, it would not be a good idea. What is a good idea is drawing a separate UpAndDown plot for the market share changes, when this is the kind of change of most interest. The fact that the relevant bar areas also have the property of area conservation underlines the strength of the basic concept of UpAndDown plots.

As graphics become more complex, tools for adjusting and adapting them become more important. UpAndDown plots allow the flexible ordering of classifying levels for non-hierarchical data, multiple sorting methods for the elements of individual levels, and a rearranging of the order of drawing levels, all to make the displays more informative.

Future goals include developing related displays for multiple time points and adding interactivity. Studying an UpAndDown plot would be easier if it were interactive. Querying, changing the order of levels, sorting in different ways, zooming, linking to other graphics, would all be valuable tools (as they would be for every kind of graphic).

Acknowledgements

Thanks to Bill Venables for coding assistance and to Nick Cox for pointing out two older references.

References

- B. Auguie. *gridExtra: Miscellaneous Functions for "Grid" Graphics*, 2017. URL <https://CRAN.R-project.org/package=gridExtra>. R package version 2.3. [p188]
- Banque de France. Services prices in france and oil prices, 2017. URL <https://www.bls.gov/opub/btn/volume-13/a-year-in-review-exploring-consumer-price-trends-in-2023.htm>. (accessed 13.11.2024). [p186]
- W. C. Brinton. *Graphic Presentation*. Brinton Associates, New York, 1939. URL <https://archive.org/details/graphicpresentat00brinrich>. [p180]
- W. S. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society A*, 150(3):192–229, 1987. URL <https://doi.org/10.2307/2981473>. [p179]

- H. Hofmann. Constructing and reading mosaicplots. *Computational Statistics & Data Analysis*, 43(4): 565–580, 2003. URL [https://doi.org/10.1016/S0167-9473\(02\)00293-1](https://doi.org/10.1016/S0167-9473(02)00293-1). [p179]
- H. Hofmann. Mosaic plots and their variants. In C. Chen, W. Haerdle, and A. Unwin, editors, *Handbook of Data Visualization*, pages 617–642. Springer, 2008. URL https://doi.org/10.1007/978-3-540-33037-0_24. [p180]
- K. Karsten. *Charts and Graphs: An Introduction to Graphic Methods in the Control and Analysis of Statistics*. Prentice-Hall, Cambridge, 1923. URL <https://archive.org/details/chartsgraphsintr0000karl/page/n7/mode/2up>. [p179]
- D. MacKay. *Sustainable Energy — without the hot air*. UIT, Cambridge, 2009. URL <https://www.withouthotair.com>. [p180]
- National Bureau of Statistics of China. Consumer price index for may 2024, 2024. URL https://www.stats.gov.cn/english/PressRelease/202406/t20240625_1955163.html. (accessed 13.11.2024). [p186]
- NISRA. 2017 mid-year population estimates for district electoral areas, 2019. URL <https://www.nisra.gov.uk/publications/2017-mid-year-population-estimates-district-electoral-areas>. [p181]
- Office for National Statistics. Consumer price inflation, uk: September 2024, 2024. URL <https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/consumerpriceinflation/september2024>. (accessed 13.11.2024). [p186]
- ONS. Consumer price indices technical manual, 2014. URL https://doc.ukdataservice.ac.uk/doc/7022/mrdoc/pdf/7022_technical_manual_2014.pdf. [p180]
- ONS. Consumer price inflation tables, 2019. URL <https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceinflation/current>. (accessed 18.03.2019). [p186]
- T. Pedersen. *patchwork: The Composer of Plots*, 2024. URL <https://CRAN.R-project.org/package=patchwork>. R package version 1.3.0. [p189]
- O. Pybus, A. Rambaut, and P. Harvey. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, 155:1429–1437, 2000. URL <https://doi.org/10.1093/genetics/155.3.1429>. [p180]
- B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.*, 11(1):92–99, 1992. URL <https://doi.org/10.1145/102377.115768>. [p179]
- Statistisches Bundesamt. Price kaleidoscope, 2024. URL <https://service.destatis.de/Voronoi/PriceKaleidoscope.svg>. (accessed 13.11.2024). [p186]
- M. Tennekes. *treemap*, 2023. URL <https://CRAN.R-project.org/package=treemap>. R package version 2.4-4. [p180]
- A. Unwin. *Getting (more out of) Graphics*. Chapman & Hall/CRC, Boca Raton, Florida, 2024a. URL <https://doi.org/10.1201/9781003131212>. [p185]
- A. Unwin. *UpAndDownPlots: Displays Percentage and Absolute Changes*, 2024b. URL <https://CRAN.R-project.org/package=UpAndDownPlots>. R package version 0.5.0. [p183]
- US Bureau of Labor Statistics. BEYOND THE NUMBERS, 2023. URL <https://www.bls.gov/opub/btn/volume-13/a-year-in-review-exploring-consumer-price-trends-in-2023.htm>. (accessed 13.11.2024). [p186]
- H. Wickham. *ggplot2*, 2022. URL <https://ggplot2-book.org>. (accessed 01.08.2022). [p184]

Antony Unwin
University of Augsburg
Mathematics Institute
University of Augsburg
86135 Augsburg, Germany
unwin@math.uni-augsburg.de