The *UpAndDownPlots* Package: 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. They can visualise changes in indices, in consumer markets, in stockmarkets, 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, 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 it is generally the percentage changes of the components that are 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 they do not move the stockmarket 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. 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. 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 sector This means that UpAndDown plots with multiple levels are consistent, a valuable property when interpreting them.

2 Visualising changes in the 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 sooner, having a lowering effect on the index.

Every country produces its own CPI and they have many different ways of displaying information on how the their index changes between two time points. Some use barcharts of percentage changes, New Zealand offers interactive piecharts, Germany includes a kaleidoscope plot, first suggested by the New York Times, and Malaysia has a display like petals of a flower.

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.

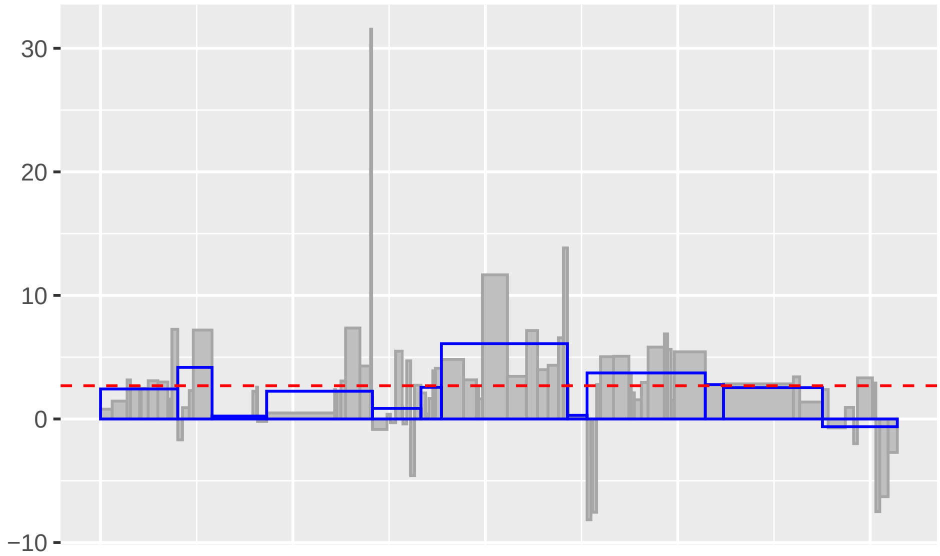

Figure 1: UK Consumer Price Index changes August 2017 to August 2018. The red dotted line shows the overall percentage change. The rectangles with blue borders show the changes of the main sectors. Heights are percentage changes, widths are index values in August 2017, so that the areas of the rectangles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

```
yp <- ud_prep(CPIuk, weight="Weight", v1="Aug2017", v2="Aug2018",
  levs=c("Sector", "Component"), sortLev=c("orig", "orig"))
yd <- ud_plot(yp, drawFrom="SmallToBig")
yd$uad
```

It can be seen that only one sector declined (Miscellaneous goods and services, the last one), 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. 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 an ordered, sector-labelled version. The plot is drawn vertically to make the labels readable.

```
yq <- ud_prep(CPIuk, weight="Weight", v1="Aug2017", v2="Aug2018",
  levs=c("Sector", "Component"), sortLev=c("perc", "perc"))
yf <- ud_plot(yq, labelvar="Sector", drawFrom="SmallToBig", vscale=c(-30, 30))
yf$uadl
```

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

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.

More complicated examples involve unequal weightings. 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

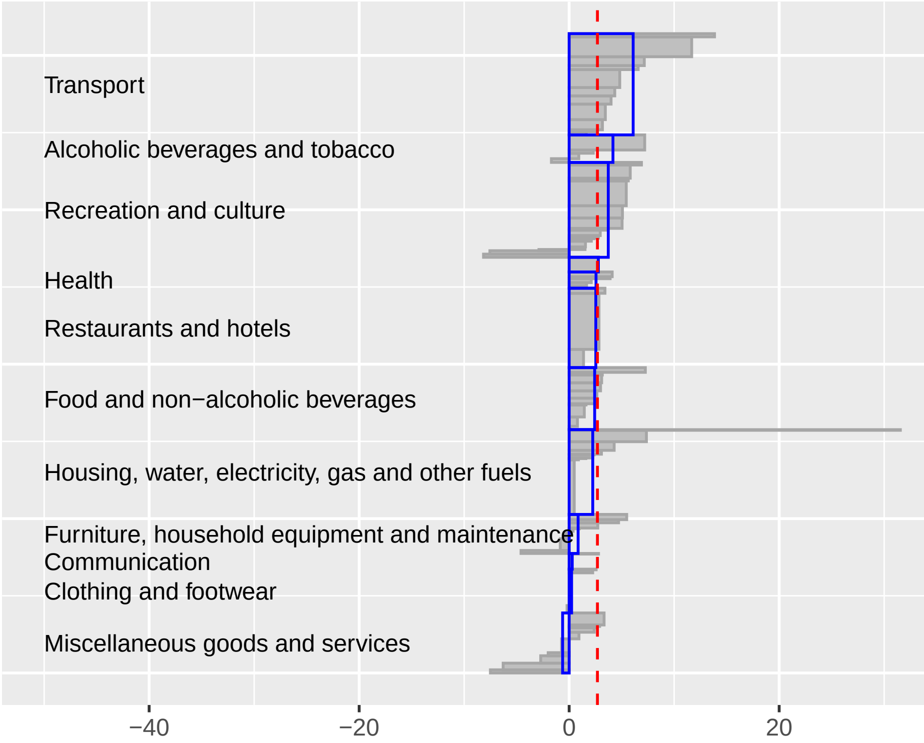

Figure 2: CPI changes ordered by percentage changes of sectors and by percentage changes of components within sectors. Two sectors, Transport and Recreation and Culture, made the biggest contributions to change. Prices increased for all components in most sectors.

comparisons at the component level, as corrective adjustments are necessary. The CPI uses chaining to maintain consistency (ONS 2014).

Multiple classification levels and nesting

Subgroups and components can be hierarchical (nested), as in the CPI, 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), Local Government Districts (LGDs, eleven), and District Electoral Areas (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 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, LGD and DEA, 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.

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. The models level should then be placed last.

Dataset structure

A dataset for an UpAndDown plot must contain at least two numeric variables giving the item values at the starting 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. The AutoSalesX dataset for the German car market in `UpAndDownPlots` includes sales in 2017 and 2018, the three classification variables of Sector, Segment, and Manufacturer, and no weighting variable. The CPIuk dataset includes component index values for August 2017 and August 2018, three classification variables (Sector, Subsector, and Component), and a weighting variable.

4 Mathematics of UpAndDownPlots

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

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

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

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

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

Scales and units

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 and censor any bars with heights larger than this at that value.

Multiplying all values by the same constant will change the numbers but not the graphics. It would just amount to a rescaling, for instance using dollars instead of euros.

Change of baseline and market share change

UpAndDown plots are drawn by default with a baseline of 0 marking no change. Sometimes it is more informative to use another baseline, for instance the overall market change. With that choice the plot emphasises 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. An example is shown in Figure 3.

```
library(patchwork)
AutoSalesXcomp <- AutoSalesX %>% filter(Segment=="Compact")
yxp <- ud_prep(AutoSalesXcomp, v1="sales17", v2="sales18",
  levs=c("Manufacturer"), sortLev=c("perc"))
yM <- ud_plot(yxp, labelvar="Manufacturer")
AutoSalesXcomp <- AutoSalesXcomp %>% mutate(S17=sum(sales17),
  S18=sum(sales18), p17=100*sales17/S17, p18=100*sales18/S18, msch=p18-p17)
AutoSalesXcompS <- AutoSalesXcomp %>% arrange(msch)
yxs <- ud_prep(AutoSalesXcompS, v1="sales17", v2="sales18",
  levs=c("Manufacturer"), sortLev=c("orig"))
yS <- ud_plot(yxs, b=yM$TotPerc, labelvar="Manufacturer")
yM$uadl + yS$uadl
```

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

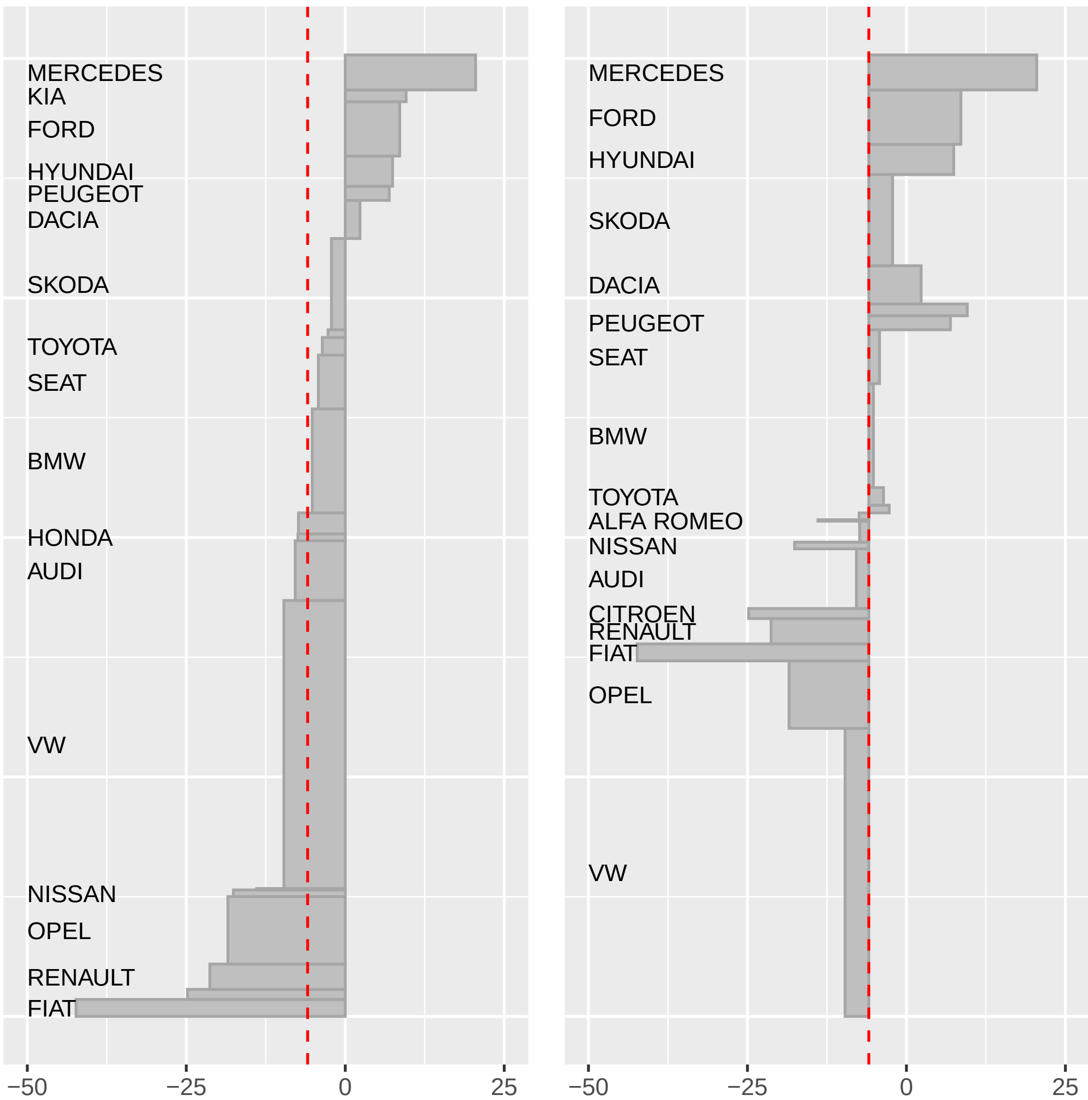

Figure 3: Compact car sales changes in Germany by manufacturer between 2017 and 2018. The plot on the left is drawn with a baseline of zero change and ordered by percentage changes. The plot on the right is drawn with a baseline of the overall market change and ordered by changes in market share.

the 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, ...), 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).

5 Varying UpAndDown plots

A default UpAndDown plot includes a dotted line showing the overall percentage change. Using this as the baseline for the bars, as just discussed, gives the plot a different interpretation.

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 emphasised. 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.

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.

The individual components of the CPI index, the lowest level in the dataset, 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.
- `final` by final size. This orders in terms of the final importance of the components.
- `perc` by ascending percentage change (bar height). 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 the precalculated ordering. This was done for the plot on the right of Figure 3 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 US Cars dataset there is information splitting the vehicles into cars or trucks and by manufacturer. A display could have a top level of manufacturer and then split by vehicle or have 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 US Cars data this would mean using car_Ford, car_GM, truck_Ford etc if you split on vehicle first, or Ford_car, Ford_truck, GM_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, for instance a new car, then there is no initial weight and an infinitely tall bar. Currently such special cases must be excluded and a note should be added to the display.

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). The default colours are provided in the function ud_colours and can be respecified with that function as required.

6 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 (Johnson and Shneiderman 1991). Area is necessary here together with length 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) 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 emphasised. 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.

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, Rambaut, and Harvey 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. For seasonal data, such as ice-cream, there might be interest in monthly changes, year-to-date changes, differences to the same month last year or differences to year-to-date last year.

7 Drawing an UpAndDown plot

Drawing a plot with several layers is easy in **ggplot2**, provided the layers are independent of one another. In UpAndDown plots they are not as 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.

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

gender	change	% change
Females	25590	2.77
Males	30980	3.48

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

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

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 choose an appropriate colour palette.

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 too 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.

8 Visualising population changes in Northern Ireland

If you plot the Northern Ireland population data changes grouped first by gender and then 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 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 Northern Ireland population 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 shows UpAndDown plots for each of the possible grouping variables. 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)
```

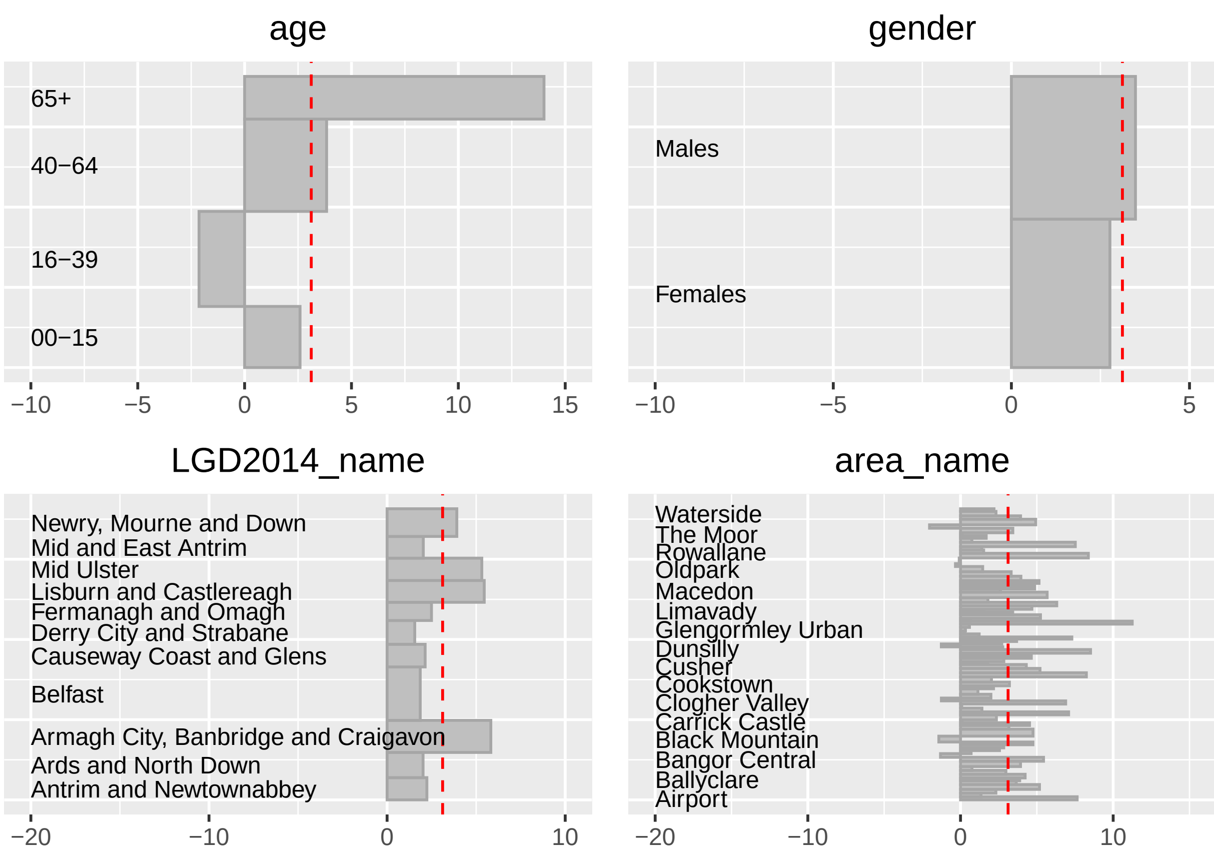

Figure 4: Percentage changes in population between 2011 and 2017 by age, gender, district, and area.

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

It is interesting to draw a graphic for age and gender together, Figure 5. This reveals that there was a much bigger increase in older men than in 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 males.

```
ag <- ud_prep(NIpop, v1="y2011", v2="y2017", levs=c("age", "gender"),
              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 6 shows the plot. All the four younger groups declined in population 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.

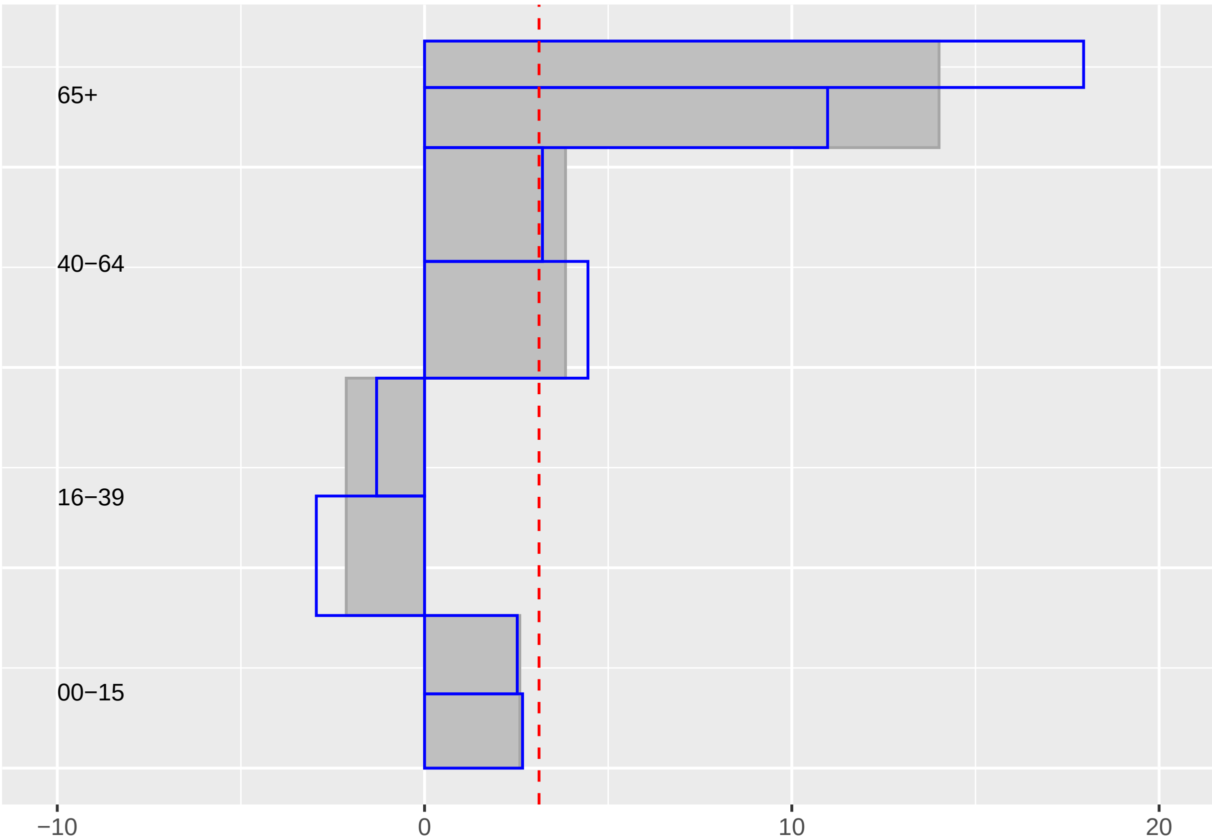

Figure 5: Percentage changes in population between 2011 and 2017 by age and then gender.

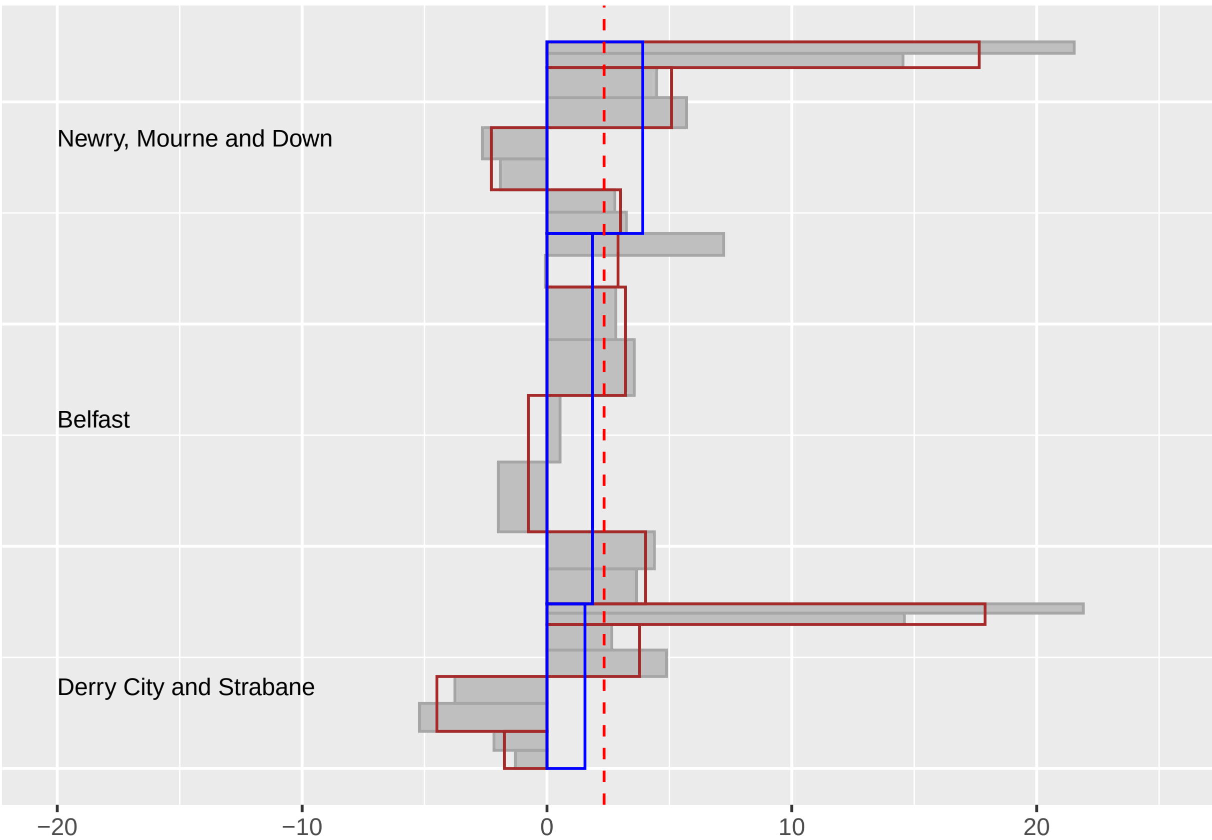

Figure 6: Percentage changes in population between 2011 and 2017 for three districts (changes outlined in blue), four age groups (outlined in brown), and gender (filled bars).

9 Using the package

There are two main functions in the `UpAndDownPlots` package, `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, one plot horizontal to display increases going upwards and decreases going downwards. The other is vertical, so that labels can be added for the bars of one of the levels. Finally, a dataset combining the statistics derived for drawing the plots and the original data is provided.

Three further functions are provided: `ud_colours` to allow the user to specify the colours used in particular plots; `sort5` to draw a set of plots for one level of changes to compare the output of all five sorting methods available; `dgroup` to compare the effects of different grouping variables by drawing a one-level plot for each one.

Three datasets are included in the package: `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 item; `AutoSales` (and the cleaned subset `AutoSalesX`) provide data from the German car market for the years 2017 and 2018 by sector, segment, model series, manufacturer, and model.

10 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.

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.

Acknowledgements

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

References

Brinton, Willard C. 1939. *Graphic Presentation*. New York: Brinton Associates.

Cleveland, William S., and Robert McGill. 1987. "Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data." *Journal of the Royal Statistical Society A* 150 (3): 192–229.

Hofmann, Heike. 2003. "Constructing and Reading Mosaicplots." *Computational Statistics & Data Analysis* 43 (4): 565–80.

———. 2008. "Mosaic Plots and Their Variants." In *Handbook of Data Visualization*, edited by C. H. Chen, W. Haerdle, and A. Unwin, 617–42. Springer.

Johnson, B., and Ben Shneiderman. 1991. "Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures." In *Proceedings of the IEEE Conference on Visualization '91*, 284–91.

Karsten, Karl. 1923. *Charts and Graphs: An Introduction to Graphic Methods in the Control and Analysis of Statistics*. Cambridge: Prentice-Hall.

MacKay, David. 2009. *Sustainable Energy — Without the Hot Air*. Cambridge: UIT.

NISRA. 2019. "2017 Mid-Year Population Estimates for District Electoral Areas." Northern Ireland Statistics; Research Agency.

ONS. 2014. "Consumer Price Indices Technical Manual." Office for National Statistics.

———. 2019. "Consumer Price Inflation Tables." https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceinflation/current%0A.

Pybus, O., A. Rambaut, and P. Harvey. 2000. "An Integrated Framework for the Inference of Viral Population History from Reconstructed Genealogies." *Genetics* 155: 1429–37.

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