The centered ternary balance scheme

A technique to visualize surfaces of unbalanced three part compositions

Jonas Schöley

Abstract

I propose the centered ternary balance scheme as a technique to visualize unbalanced compositional data on a surface.

Problem description and proposed solution

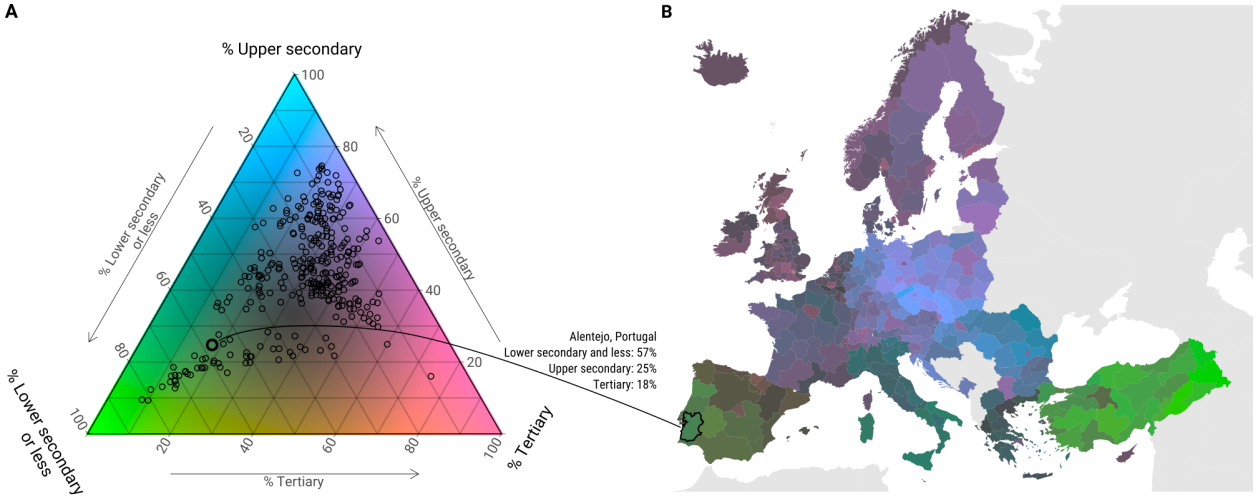
When it comes to compositional data, the number "three" is quite significant: the share of people working in the primary vs. secondary vs. tertiary sector, the proportion of total population change explained by migration vs. fertility vs. mortality, the relative population numbers in young age vs. working age vs. retirement age, the share of a cohort attaining primary vs. secondary vs. tertiary education degrees, the relative number of deaths due to prematurity vs. accidents vs. old age, the share of papers accepted as is vs. revised vs. rejected... It comes to no surprise that compositional data often comes in *ternary* form, i.e. featuring three parts, as the simplicity of a ternary classification scheme facilitates data collection and comparability. Ternary compositions are often available by period, age and/or geographical region leading to the challenge of *visualizing ternary compositions on a surface*, i.e. the surface of the Earth or the period-age Lexis surface.

The *ternary balance scheme* (Brewer 1994, see Schöley and Willekens (2017) or Dorling (2012) for applications of that scheme) is a color scale suited to the visualization of three part compositions on a surface. It works by expressing the relative shares among three parts as the mixture of three primary colors. Figure 1B shows the proportions of people with either primary, secondary or tertiary educational attainment in Europe in 2016. Primary degrees are mapped to green, secondary to blue, and tertiary to red. The deeper the green in a region, the higher the share of people with primary education in that region, the same logic applies for the two other education categories. The more grayish a region is colored, the more balanced the three proportions are with a perfect grey signifying an equal share of people in all three education categories. A ternary diagram is used as a color key (figure 1A).

While the ternary balance scheme allows for incredibly dense yet clear visualizations of *well spread out* ternary compositions the technique is less informative when used with highly *unbalanced data*. Figure 2A shows the regional labor force composition in Europe as of 2016. The map is nearly monochromatic, the intense pink signifying a working population which is concentrated in the tertiary (services) sector. Regions in Turkey and Eastern Europe show a somewhat higher concentration of workers in the primary (production) sector but overall the data shows little variation with regards to the *visual reference point*, i.e. the greypoint marking perfectly balanced proportions.

A remedy for analysing data which shows little variation in relation to some reference point is to *change the point of reference*. Figure 2B yet again shows the European regional labor force composition in 2016 but the color scale has been altered so that its greypoint – the visual point of reference – is positioned at the European annual average. Consequently the colors now show direction and magnitude of the deviation from the European average labor force composition. Green, blue and pink hues show a higher than average share of workers in the primary, secondary and tertiary sector respectively. The saturation of the colors show the magnitude of that deviation with perfect grey marking a region that has a labor force composition equal to the European average, i.e. the reference point.

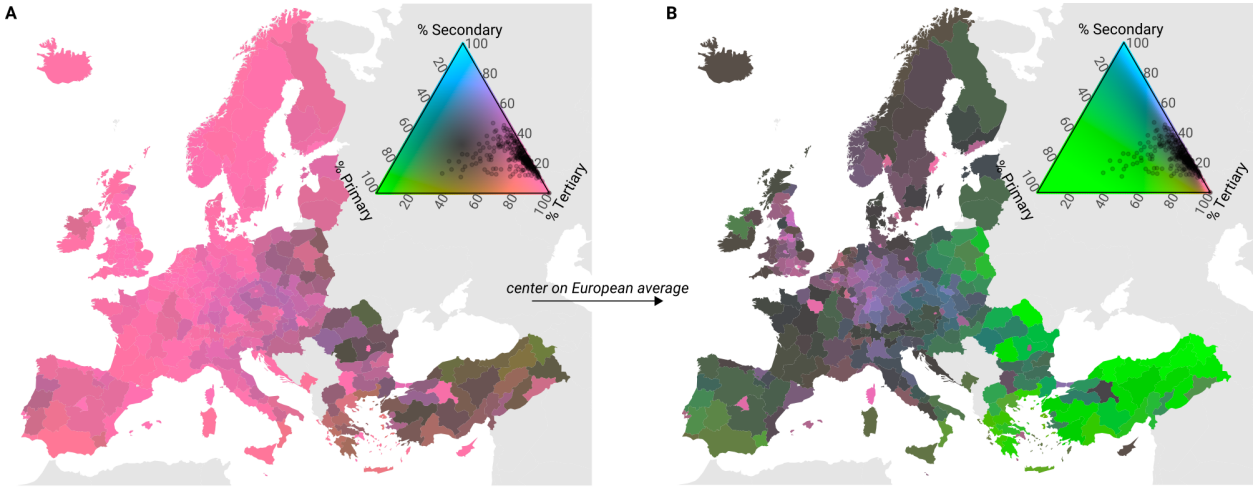
In the following I discuss how to center ternary compositional data to arbitrary reference points and how to plot the compositions using a corresponding *centered ternary balance scheme*.



A **Ternary diagram** showing the population composition by education level for each European NUTS-2 region in 2016 ages 25-64. The **colors correspond to the ternary balance scheme** used to color map B and show direction (via hue) and magnitude (via lightness and saturation) of the deviation from a perfectly balanced composition.

The regions in this map have been **color-coded with the ternary balance scheme** as displayed in figure A. Bright and vivid colors indicate regions where most people have the same education level whereas desaturated and dark colors indicate a more diverse population composition.

Figure 1: Demonstration of the ternary balance scheme showing the composition of educational attainment by region in Europe 2016. Data by eurostat.



The ternary balance scheme is less informative if the data is narrowly clustered. While this map of the regional workforce composition by economic sector clearly shows the dominance of the service sector Europe 2016, it largly fails to distinguish between regions with varying shares of primary and secondary workers.

The centered ternary balance scheme shows deviations from a reference composition. Centering the color scale around the average European labor force composition in 2016 (4% primary, 24% secondary, 72% tertiary) clearly highlights regions which have higher/ lower than expected shares of workers in each of the three sectors.

Figure 2: Demonstration of the *centered* ternary balance scheme in comparision with the non-centered scheme showing the labor force composition by region in Europe 2016. Data by eurostat.

Centering of compositional data

Say you have a series of temperature readings for each day of a year and want to show how each reading compares to the yearly average. By substracting the annual average from each data point you create a mean standardized variable representing direction and magnitude of the temperature deviation from the mean with a reference point at zero. Alternatively you could divide by the mean, putting the reference point at unity, to show the relative deviations. A similar standardization is possible for compositional data: the *pertubation by the inverse of the reference composition*.

For compositional data a natural reference point is a perfectly balanced composition, i.e. a composition of K parts with each part equal to $1/K$. For a ternary composition this reference point is located at $(1/3, 1/3, 1/3)$ – the barycenter of a ternary diagram – and coincides with the position of the greypoint in a ternary balance scheme. The operation of *pertubation* allows to transform the ternary diagram such that an arbitrary composition can be located at the center (Von Eynatten, Pawlowsky-Glahn, and Egozcue 2002), or in other words, *pertubation* centers compositional data around a new reference point. Let's demonstrate.

In 2016 the average European NUTS-2 region had 4% of the labor force working in the primary sector, 24% in the secondary and 72% in the tertiary sector. This composition $\mathbf{c} = (0.04, 0.24, 0.72)$ shall be the reference against which all other compositions are compared. Figure 3A shows the labor force composition of all European regions in a ternary diagram with the European average annotated by lines.

Let $\mathbf{p}_i = (p_1, p_2, p_3)_i$ be the labor force composition of region i. For each region we calculate $\mathbf{p}'_i = \left(\frac{p_1/c_1}{\sum}, \frac{p_1/c_1}{\sum}, \frac{p_1/c_1}{\sum}\right)_i$, with $\sum_i = p_1/c_1 + p_2/c_2 + p_3/c_3$, i.e. we divide each element of the composition by the reference composition and then *close* the result so that it sums to unity. This sequence of operations is called the pertubation of $\mathbf{p}_i$ by the inverse of $\mathbf{c}$ and *centers* $\mathbf{p}_i$ around $\mathbf{c}$ with $\mathbf{c}$ moving to the barycenter of the ternary diagram. Figure 3B shows the centered European regional labor force compositions.

The centered ternary balance scheme

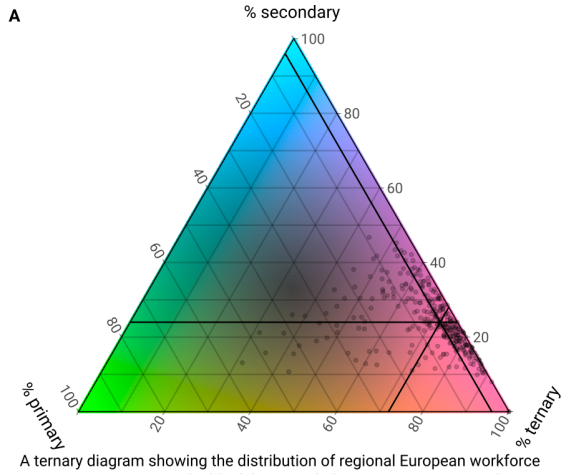
The construction of the centered ternary balance scheme is simple, one needs but two ingredients: the ability to colorize a ternary composition with the regular ternary balance scheme[1], and the ability to pertubate a ternary composition by the inverse of a reference composition as proposed by Von Eynatten, Pawlowsky-Glahn, and Egozcue (2002) and described above. One first performs the pertubation of the compositional data set, thereby centering the data on the reference composition, and then colorizes the pertubated data according to the regular ternary balance scheme.

The resulting colors show for each composition on the ternary diagram the direction and magnitude of deviation from the reference composition. The hue of the color encodes which part(s) of a three part composition are higher than the reference composition. The lightness and saturation of a color show how far away a composition is from the reference with the reference composition itself colored grey.

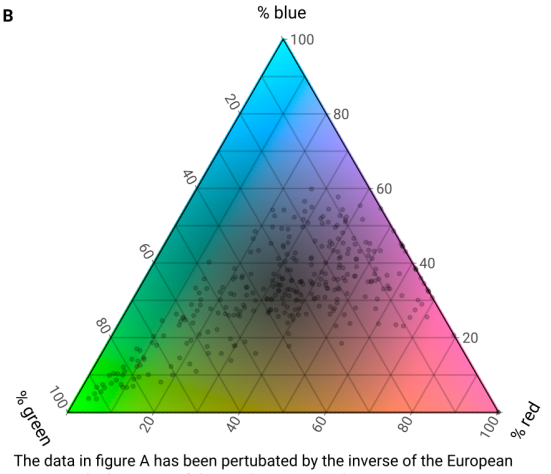
How to draw an informative color key for the centered ternary balance scheme? Simply plotting the centered compositions in a ternary diagram with a ternary balance scheme background (see figure 3B) – while correct – isn't very intuitive as the centered compositions can not be easily interpreted[2]. A better option is to plot the centered data in a ternary diagram but to label the scales of the diagram such that the original proportions can be read (see figure 3C). Because the centering operation skews the gridlines in a ternary diagram, centered gridlines have to be plotted (Von Eynatten, Pawlowsky-Glahn, and Egozcue 2002). The gridlines can also be labelled with the percent-point-difference to the reference composition. To do so one generates a set of gridlines crossing at the reference composition $\mathbf{c}$ and from there draws additional gridlines on each axis spaced 0.1 units apart in both directions. Substracting the reference composition from those gridline positions results in a grid that is centered at $(0, 0, 0)$ and shows the positive or negative percent point difference of a location

[1] The simplest way to achieve this is to interpret the composition as coordinates in the rgb color space. A more flexible method is described in Schoeley2017.

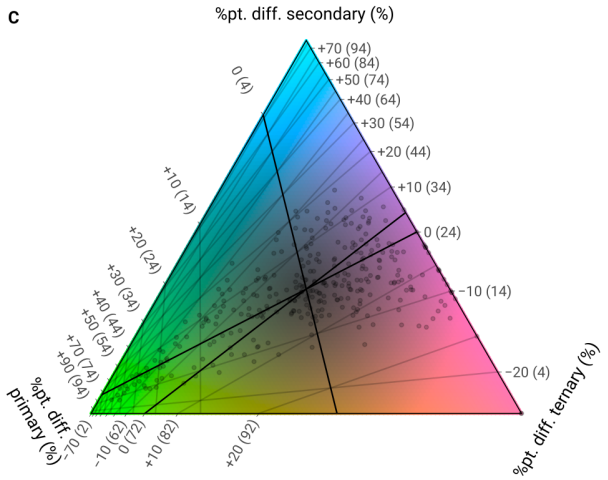
[2] The same situation arises when a logarithmic scale is labelled with the logged values as opposed to the values on the original scale.



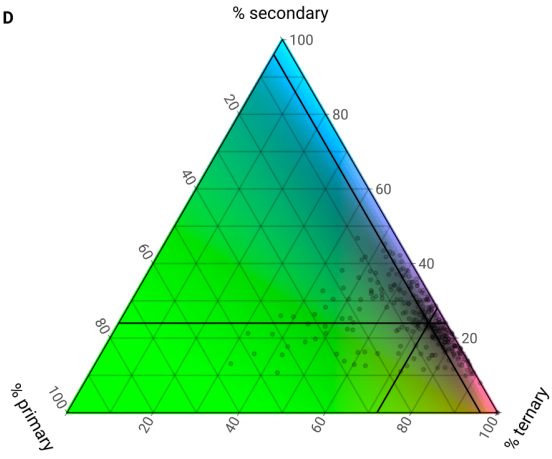
A ternary diagram showing the distribution of regional European workforce compositions in 2016. The black lines mark the European average. The background is colored with the ternary balance scheme.



The data in figure A has been pertubated by the inverse of the European average. The locations of the resulting centered compositions are interpreted as colors in the centered ternary balance scheme.



To facilitate interpretation the centered compositions can be labelled in terms of their original values and their percent point differences from the reference composition, here, the European average, which is now situated at the barycenter of the triangle.



The centered ternary balance scheme is a skewed standard ternary balance scheme with its greypoint shifted to the location of the reference composition. Note that this color key conveys the same information as figure C. Whereas in figure C the data has been centered on the color scale, here the color scale has been centered on the data.

Figure 3: Different representations of the color key for the (centered) ternary balance scheme showing the labor force composition by region in Europe 2016. Data by eurostat.

on the ternary surface to the reference composition. Each point on such a ternary surface sums up to 0. A third option is to plot the uncentered data in the standard ternary diagram and to pertube the background color key instead, shifting its greypoint from the barycenter of the triangle to the location of the reference composition (figure 3D).

Discussion

With the centered ternary balance scheme I've proposed a visualization technique capable of showing the divergence of a three part composition with respect to a reference composition and has – to my knowledge – never been proposed before. The technique can be utilized to show the internal variation of a data set which is narrowly clustered on a global scale (as demonstrated in figures 1c and d) or to compare a composition against a standard (i.e. by centering the regional distribution of educational attainment in plot 1b around the European average of 1980).

The centered ternary balance scheme is a straightforward synthesis of the three variable balance scheme as described by Brewer (1994) and the centering operation applied in the context of compositional data analysis in the ternary diagram (Von Eynatten, Pawlowsky-Glahn, and Egozcue 2002). There is future opportunity for further synthesis, i.e. direction and magnitude of *compositional change over time* can be visualized by pertubing each data point at t_1 by the corresponding composition at t_2 and colorizing the resulting pertubation using the standard ternary balance scheme.

The technique described in this paper has been implemented in the R package "tricolore". Given a three column matrix of three part compositions "tricolore" returns a vector of colors along with a suitable color key. I hope that this implementation encourages people to experiment with this novel visualizations technique. Three part compositions are plenty and surfaces, wether defined by longitude and latitude or by period and age, provide plenty of room to find interesting variation in the data.

References

Brewer, Cynthia A. 1994. "Color Use Guidelines for Mapping and Visualization." In *Visualization in Modern Cartography*, edited by Alan M. MacEachren and D. R. Fraser Taylor, 123–47. Modern Cartography. Oxford, UK: Pergamon.

Dorling, Daniel. 2012. *The Visualization of Spatial Social Structure*. Wiley Series in Computational and Quantitative Social Science. Chichester, UK: Wiley. https://sasi.group.shef.ac.uk/thesis/prints.html.

Schöley, Jonas, and Frans Willekens. 2017. "Visualizing compositional data on the Lexis surface." *Demographic Research* 36 (21): 627–58. doi:10.4054/DemRes.2017.36.21.

Von Eynatten, Hilmar, Vera Pawlowsky-Glahn, and Juan José Egozcue. 2002. "Understanding perturbation on the simplex: A simple method to better visualize and interpret compositional data in ternary diagrams." *Mathematical Geology* 34 (3): 249–57. doi:10.1023/A:1014826205533.