DEMOGRAPHIC RESEARCH

A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

Research Material

Geofaceting: Aligning small multiples for regions in a spatially meaningful way

Ilya Kashnitsky

José Manuel Aburto

Contents

Geofaceting: Aligning small multiples for regions in a spatially meaningful way

Ilya Kashnitsky[1]

José Manuel Aburto[2]

Abstract

BACKGROUND
Creating visualizations that include multiple dimensions of the data while preserving spatial structure and readability is challenging. Here we demonstrate the use of geofaceting to meet this challenge.

OBJECTIVE
Using data on young adult mortality in the 32 Mexican states from 1990 to 2015, we demonstrate how aligning small multiples for territorial units, often regions, according to their approximate geographical location – geofaceting – can be used to depict complex multi-dimensional phenomena.

RESULTS
Geofaceting reveals the macro-level spatial pattern while preserving the flexibility of choosing any visualization techniques for the small multiples. Creating geofaceted visualizations gives all the advantages of standard plots in which one can adequately display multiple dimensions of a dataset.

CONCLUSIONS
Compared to other ways of small multiples arrangement, geofaceting improves the speed of regions' identification and exposes the broad spatial pattern.

---

[1] Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, Odense, Denmark, and National Research University Higher School of Economics, Moscow, Russia. Email: ikashnitsky@sdu.dk.
[2] Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, Odense, Denmark, and Max-Planck Institute for Demographic Research, Rostock, Germany. Email: jmaburto@sdu.dk.

1. Introduction

In data visualization, it is often challenging to represent multiple relevant dimensions while preserving the readability of a plot. This is especially true when the task is to expose spatial variation of some complex phenomenon. In such a case, geographical maps are the natural choice for a visualization framework because they are meant to show spatial patterns. However, the usual limitation is that only one variable can be meaningfully represented with colors in a choropleth. So, what if the dataset at hand is more complex and demands a balanced exposure of several dimensions?

Usually, time is a dimension difficult to represent, yet it is very important for the story underlying certain phenomena. Visualizing time series with choropleths is challenging. One has to produce either small multiples for the years or animated pictures with maps for various years flashing sequentially. Both variants make it difficult to compare regions across time, which is the main goal of such visualizations. Furthermore, including additional variables (e.g., age) complicates the representation, and the basic choropleth visualization framework fails. An alternative to overcome these limitations is geofaceting.

The idea of geofaceting is simple: A 'normal' plot is produced for each of the regions, and then all the small panels are arranged according to their approximate geographic location, thereby making it easier to identify regions. The spatial logic of small multiples alignment helps to identify the units of analysis – usually regions of a country – faster. Moreover, it reveals the macro-level spatial pattern while preserving the flexibility of visualization technique choice for the small multiples. As a result, creating geofaceted visualizations gives all the advantages of standard plots in which one can easily display at least three dimensions of a dataset. The resulting map-like plots provide a unique opportunity to view multivariate spatial patterns at once.

Geofaceting has been reinvented multiple times. The use of small multiples arranged as grids can be found in the famous Galton's 1863 multivariate weather chart of Europe (Galton 1863; Friendly 2008). French geographers of the 19th century utilized another very closely related idea: They systematically overlaid small plots in geographical maps, providing additional information for the chosen locations (Palsky 1996). Geofaceting goes one step further by dropping the actual geographical map and just arranging the small multiples in line with the spatial pattern of the corresponding areas. This approach was recently formalized by Ryan Hafen, received its name, and was consistently implemented in the R package `geofacet` (Hafen 2019). The R package also provides tools for creating and publishing grids for custom territories, thus accumulating a library of community-contributed grids (Hafen 2018).

http://www.demographic-research.org

2. Data and methods

The application of our visualization proposal relies on the results from Aburto, Riffe, and Canudas-Romo (2018). These results are based on cause-of-death information available from the Mexican Statistical Office from 1990 to 2015 (Instituto Nacional de Estadística y Geografía 2015), and population estimates from the Mexican Population Council, and includes 32 Mexican states as geographical units. Data were disaggregated by single age, sex, and state. Population estimates were adjusted for age misstatement, undercounting, and interstate and international migration.

Cause-specific death rates were smoothed over age and time for each state and sex separately, using 2-d p-spline to avoid random variations (Camarda 2012). Smoothed death rates were then constrained to sum to the unsmoothed all-cause death rates. Period life tables were constructed for males from 1990 to 2015, following standard demographic methods (Preston, Heuveline, and Guillot 2001: Chapter 3). The average years lived between ages 15 and 49 – temporary life expectancy (Arriaga 1984) – were calculated with cause-specific contributions to the difference between state-specific temporary life expectancy and a low-mortality benchmark using standard decomposition techniques (Horiuchi, Wilmoth, and Pletcher 2008).

The low-mortality benchmark was calculated on the basis of the lowest observed mortality rates by age and cause of death, from among all states for a given sex and year. The resulting minimum mortality rate schedule has a unique age profile, and it determines a benchmark temporary life expectancy. The minimum mortality schedule can be treated as the best presently achievable mortality, assuming perfect diffusion of the best available practices and technologies in Mexico (Vallin, Meslé, and Divinagracia 2008; Canudas-Romo, Booth, and Bergeron-Boucher 2019).

There exists substantial regional variation in young male mortality across Mexican states. Therefore, to properly visualize mortality patterns, it is necessary to take into account the spatial dimension of the dataset, which we achieve with geofaceting. As there was no geofacet layout for Mexico, we created one. This produced grid for Mexican states was successfully submitted to the `geofacet` package (Kashnitsky 2017); however, at the revision stage of the paper, we switched to an improved layout of Mexican states (Zepeda 2018).

There is no way to efficiently represent in one plot both absolute and relative values. Thus, the first two figures complement each other: Figure 1 uses the stacked bar plot technique to reveal the variation of young adult mortality in Mexican states over time; Figure 2 shows the dominant cause of death with a colored tile plot on a standard Lexis surface, which can be seen as a categorical version of a heatmap (Schöley and Willekens 2017; Rau et al. 2018).

Focusing on one leading cause of death may mask its relative importance compared with the second, third, and others. Thus, in Figure 3 we apply the framework

of ternary color-coding, which was recently formalized and streamlined in the R package `tricolore` (Schöley and Kashnitsky 2018). Ternary color-coding maximizes the amount of information conveyed with colors by representing each element in a three-dimensional array of compositional data with a single color. Each part of the ternary composition is assigned a hue (color characteristic), and the amount of hue for each data element is proportional to its weight in the ternary composition. For more technical details on the method, check Schöley (forthcoming); for an indicative use case of ternary color-coding see Kashnitsky and Schöley (2018). Figure 4 facilitates comparison between Figures 2 and 3.

The figures presented in this paper can be reproduced easily by using the replication material that we provide openly (Kashnitsky and Aburto 2019). We used the R programming language (R Core Team 2018) for the analyses and data visualization; in addition, we used these packages: `tidyverse` (Wickham 2017), `tricolore` (Schöley and Kashnitsky 2018), `ggtern` (Hamilton and Ferry 2018), `hrbrthemes` (Rudis 2018), `extrafont` (Chang 2014), `RColorBrewer` (Neuwirth 2014), and `geofacet` (Hafen 2019).

3. Application

To show the usefulness of our proposal, we analyze the contribution of homicide, road traffic accidents, and suicide, medically amenable mortality, and causes amenable to health behavior to the gap in temporary life expectancy between ages 15 and 49 of each of 32 Mexican states with a low-mortality benchmark. The category 'Amenable to medical service' refers to those conditions that are susceptible to medical intervention, such as infectious and respiratory diseases, some cancers and circulatory conditions, and birth conditions, among others. For details on codes from International Classification of Diseases revision 10 included in this category, we refer the reader to the original classification (Aburto, Riffe, and Canudas-Romo 2018). These causes have emerged as leading among young people, and the first two recently had a sizable impact on life expectancy in Mexico (Aburto et al. 2016; Aburto and Beltrán-Sánchez 2019).

Three complementary geofaceted plots were created. Figure 1 shows the absolute impact of five causes of death on the difference between the observed life expectancy with the best-practice life expectancy (low benchmark) for young males. For example, it shows how the contribution of homicides (magenta) increased substantially after 2005, particularly in the north, reaching a peak in 2011 for Chihuahua, Sinaloa, and Durango, among others. It is also clear from this graph that the most affected state in the south since the early 1990s is Guerrero.

Figure 1: **Gap between observed and best-practice life expectancy for Mexican states: Years of life lost by cause of death across time (1990–2015)**
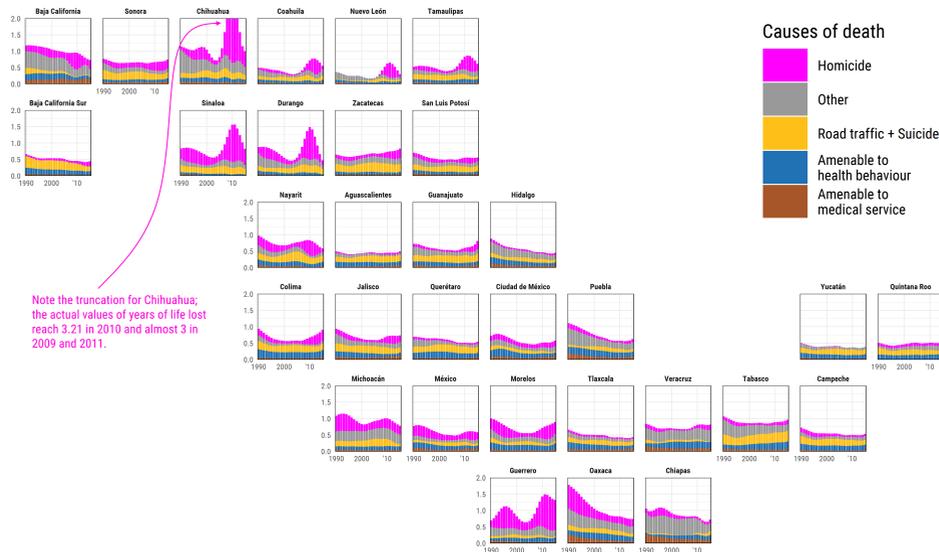
Figure 2 shows state-specific Lexis diagrams with the main cause of death at each age in a given year. It gives a full representation of the main cause of death by age and period, compromising on the actual values of the gap, i.e., Figure 1. For example, from this graph it is clear that homicides are contributing the most across ages between 15 and 49 in most states in the north. However, even though in Oaxaca (in the south) the contribution of homicide was decreasing (Figure 1), between ages 20 and 30 homicide remained the main contributor to the gap.

Figure 2:   **Gap between observed and best-practice life expectancy for Mexican states: Cause of death contributing the most by age (15–49) and time (1990–2015)**
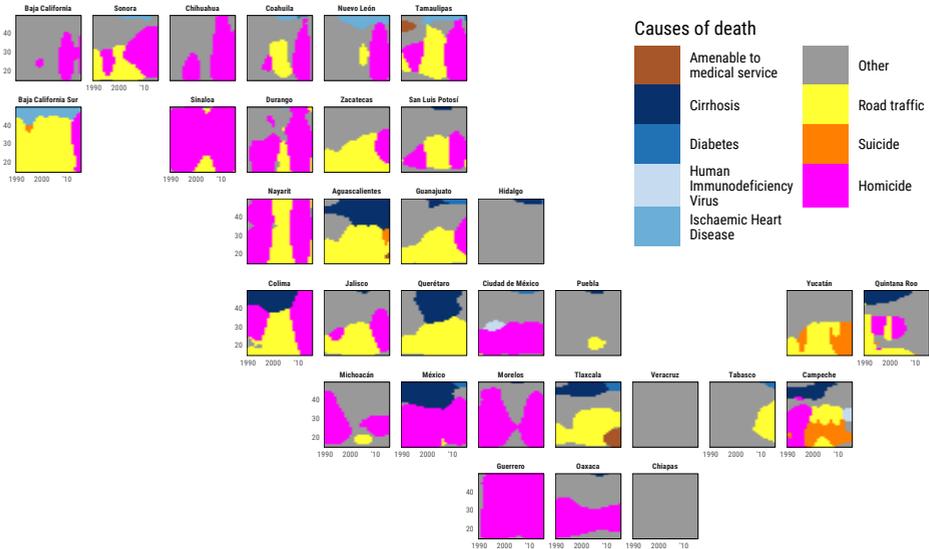
To enrich the plot with geofaceted Lexis surfaces (Figure 2), we use ternary color-coding of the three main groups of causes of death: homicides, road traffic, and suicides, and all other causes combined (Figure 3). These causes of death are known to be the main contributors in midlife and through the young mortality hump (Remund, Camarda, and Riffe 2018). This plot deepens the previous one by representing the relative importance of the two main causes of death compared with all others combined. For example, if we compare México state with the neighboring Guerrero, their mortality patterns at ages 20–30 seem very similar if we look at Figure 2 and focus only at the leading cause of death, homicide. Yet when we consider the relative importance of homicide in the mortality regime of the two states (Figure 3), it becomes clear that homicide is a bigger problem by far in the state of Guerrero.

Figure 3:   **Gap between observed and best-practice life expectancy for Mexican states: Color-coded ternary compositions of the three leading groups of causes of death by age (15–49) and time (1990–2015)**
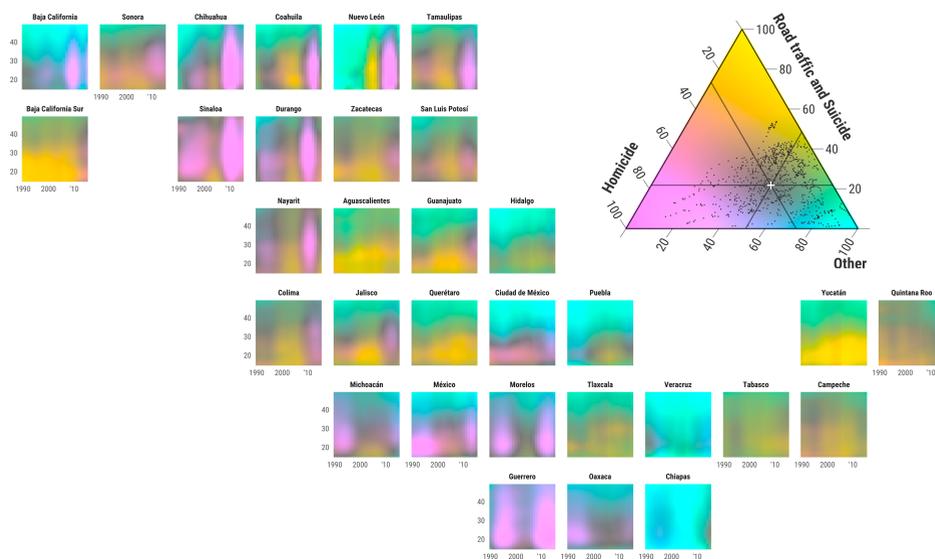
Figure 4 demonstrates further the usefulness of ternary color-coding as a way to highlight the story of the homicide crisis. Consider two states – Ciudad de México and Oaxaca (Figures 4c and 4d), which have very similar profiles when we look at the leading cause of death (Figure 4a). Once the relative importance of the leading causes of death is taken into account (Figure 4b), the differences in mortality regimes between the states come forward in both in time and age dimensions.

Figure 4:   **The comparison of the visualization approaches in Figures 2 and 3, panels a and b, respectively, for two selected states of Mexico, Ciudad de México and Oaxaca (The locations of the selected states on a geographical map, and the used geofacet grid are represented in panels c and d, respectively)**
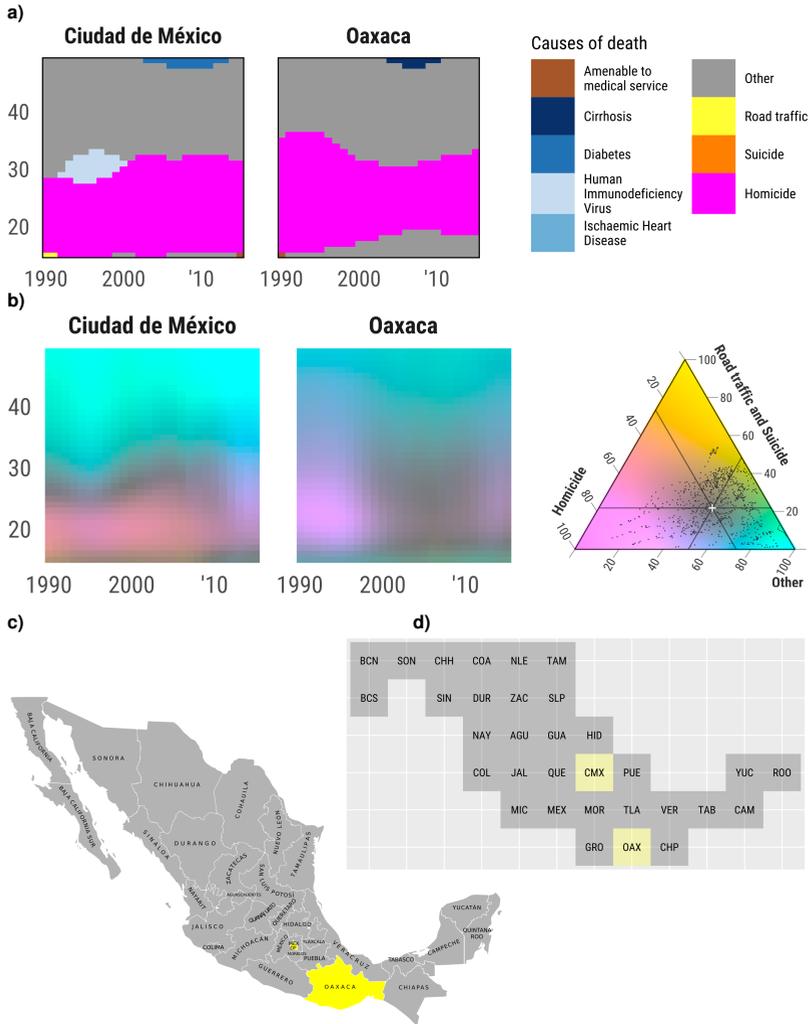
4. Discussion

Geofaceting is an elegant data visualization technique that helps to analyze different dimensions of a complex phenomenon across multiple regions, improving on their graphical representation. Essentially, the method proposes to arrange a multi-panel plot (often called 'small multiples') according to the geographical location of the regions. The approach preserves the flexibility of a standard plot for each of the regions allowing different dimensions of a dataset to be depicted. Geofaceting can be seen as an improved alternative to panel matrices or small multiples grids.

We demonstrate the usefulness of geofaceting by showing the specific case of Mexico and mortality patterns over a fairly large period, 1990–2015. The main advantage of our proposal is that the reader can easily interpret complex phenomena while being able to identify regional variations. Four dimensions (geography, cause of death, age, and time) can be summarized in a single figure. This benefit is particularly important in the case of young males in Mexico that have experienced an unprecedented period of rising homicidal mortality. Moreover, the changing dynamics of violence in the country is a dimension that is hard to represent graphically. Nevertheless, with the geofaceting framework, the reader can easily get a sense of this phenomenon. For example, while most of the historically violent states are in the northern part of the country (Figures 1, 2, and 3), an upsurge of violence in the south is clear, albeit with different intensities (i.e., the absolute gap between states and best-practice life expectancy). Being able to identify variations regionally, but also in terms of intensity, is a great advantage of the proposed visualization technique.

There are some limitations of the approach. For instance, if the number of regional divisions in a territory is too large or too small, geofaceting might not be the ideal approach to show complex phenomena. Moreover, if a territory is oddly shaped or the subregions are difficult to align geographically, getting a reasonable regional representation might be impossible. For example, a geofaceted plot of the provinces of Chile would look remarkably close to a one-columned vertical grid. Also, in most cases geofacet grids are unable to reflect exact relative positions of the regions because the boundaries of the neighboring regions can be rather complex. For example, if one region is encapsulated completely within another, the location of the inset region becomes completely arbitrary. Another case is when bordering regions get separated in the geofacet grid, as with Oaxaca and Puebla in our Mexican plots (see Figures 4c and 4d). Nevertheless, we believe that in the case shown here, geofaceting, combined with ternary color-coding, proved to be a useful tool that exposed a macro-representation of the explored phenomenon, i.e. it showed cause-specific contributions to the gap between states and best-practice life expectancy while still accounting for regional variations.

5. Acknowledgements

References

Aburto, J.M. and Beltrán-Sánchez, H. (2019). Upsurge of homicides and its impact on life expectancy and life span inequality in Mexico, 2005–2015. *American Journal of Public Health* 109(3): e1–e7. doi:10.2105/AJPH.2018.304878.

Aburto, J.M., Beltrán-Sánchez, H., García-Guerrero, V.M., and Canudas-Romo, V. (2016). Homicides in Mexico reversed life expectancy gains for men and slowed them for women, 2000–10. *Health Affairs* 35(1): 88–95. doi:10.1377/hlthaff.2015.0068.

Aburto, J.M., Riffe, T., and Canudas-Romo, V. (2018). Trends in avoidable mortality over the life course in Mexico, 1990–2015: A cross-sectional demographic analysis. *BMJ Open* 8(7): e022350. doi:10.1136/bmjopen-2018-022350.

Arriaga, E.E. (1984). Measuring and explaining the change in life expectancies. *Demography* 21(1): 83–96. doi:10.2307/2061029.

Camarda, C.G. (2012). MortalitySmooth: An R package for smoothing Poisson counts with P-splines. *Journal of Statistical Software* 50(1): 1–24. doi:10.18637/jss.v050.i01.

Canudas-Romo, V., Booth, H., and Bergeron-Boucher, M.P. (2019). Minimum death rates and maximum life expectancy: The role of concordant ages. *North American Actuarial Journal* 1–13. doi:10.1080/10920277.2018.1519448.

Chang, W. (2014). extrafont: Tools for using fonts [electronic resource]. Vienna: R Foundation for Statistical Computing. https://CRAN.R-project.org/package=extrafont.

Friendly, M. (2008). A brief history of data visualization. In: Chen, C.h., Härdle, W., and Unwin, A. (eds.). *Handbook of data visualization*. Berlin: Springer: 15–56. doi:10.1007/978-3-540-33037-0.

Galton, F. (1863). *Meteorographica, or, methods of mapping the weather: Illustrated by upwards of 600 printed and lithographed diagrams referring to the weather of a large part of Europe, during the month of December 1861*. London: Macmillan.

Hafen, R. (2018). Introducing geofacet [electronic resource]. Ryan Hafen. http://ryanhafen.com/blog/geofacet.

Hafen, R. (2019). geofacet: 'ggplot2' faceting utilities for geographical data [electronic resource]. Vienna: R Foundation for Statistical Computing. https://cran.r-project.org/package=geofacet.

Hamilton, N.E. and Ferry, M. (2018). ggtern: Ternary diagrams using ggplot2. *Journal of Statistical Software* 87(1): 1–17. doi:10.18637/jss.v087.c03.

Horiuchi, S., Wilmoth, J.R., and Pletcher, S.D. (2008). A decomposition method based on a model of continuous change. *Demography* 45(4): 785–801. doi:10.1353/dem.0.0033.

Instituto Nacional de Estadística y Geografía (2015). Deaths microdata [electronic resource]. Aguascalientes: INEGI.

Kashnitsky, I. (2017). New grid: 'mex_grid1' [electronic resource]. https://github.com/hafen/geofacet/issues/31.

Kashnitsky, I. and Aburto, J.M. (2019). Geofaceting: Align small-multiples for regions in a spatially meaningful way: Replication materials [electronic resource]. https://github.com/ikashnitsky/demres-2018-geofacet.

Kashnitsky, I. and Schöley, J. (2018). Regional population structures at a glance. *The Lancet* 392(10143): 209–210. doi:10.1016/S0140-6736(18)31194-2.

Neuwirth, E. (2014). RColorBrewer: ColorBrewer palettes [electronic resource]. Vienna: R Foundation for Statistical Computing. https://CRAN.R-project.org/package=RColorBrewer.

Palsky, G. (1996). *Des chiffres et des cartes: Naissance et développement de la cartographie quantitative française au XIXe siècle*. Paris: Ministère de l'enseignment supérieur et de la recherche, Comité des travaux historiques et scientifiques.

Preston, S.H., Heuveline, P., and Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford: Blackwell.

R Core Team (2018). R: A language and environment for statistical computing [electronic resource]. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rau, R., Bohk-Ewald, C., Muszyńska, M.M., and Vaupel, J.W. (2018). *Visualizing mortality dynamics in the Lexis diagram*. Cham: Springer. doi:10.1007/978-3-319-64820-0.

Remund, A., Camarda, C.G., and Riffe, T. (2018). A cause-of-death decomposition of young adult excess mortality. *Demography* 55(3): 957–978. doi:10.1007/s13524-018-0680-9.

Rudis, B. (2018). hrbrthemes: Additional themes, theme components and utilities for 'ggplot2' [electronic resource]. Vienna: R Foundation for Statistical Computing. https://CRAN.R-project.org/package=hrbrthemes.

Schöley, J. (forthcoming). The centered ternary balance scheme: A technique to visualize surfaces of unbalanced three part compositions. *Demographic Research* .

Schöley, J. and Kashnitsky, I. (2018). tricolore: A flexible color scale ternary compositions [electronic resource]. Vienna: R Foundation for Statistical Computing. https://cran.r-project.org/package=tricolore.

Schöley, J. and Willekens, F. (2017). Visualizing compositional data on the Lexis surface. *Demographic Research* 36(21): 627–658. doi:10.4054/DemRes.2017.36.21.

Vallin, J., Meslé, F., and Divinagracia, E. (2008). Minimum mortality: A predictor of future progress? *Population* 63(4): 557–590. doi:10.3917/popu.804.0647.

Wickham, H. (2017). tidyverse: Easily install and load the 'Tidyverse' [electronic resource]. Vienna: R Foundation for Statistical Computing. https://CRAN.R-project.org/package=tidyverse.

Zepeda, F.A. (2018). New grid: 'mx_state_grid3' [electronic resource]. https://github.com/hafen/geofacet/issues/135.