

Geographic display of linguistic data

Michael Cysouw

September 25, 2024

Contents

1	Visualisation in comparative linguistics	1
2	Methods	2
3	Projection	2
4	Weighting	6
5	Interpolation	6
6	Prediction	9
7	Conclusion	9
8	Appendix	9
	8.1 Projections	9
	References	9

1 Visualisation in comparative linguistics

visualisation in comparative linguistics is mainly used to inspect data and formulate hypotheses and generalisations

Previous research (Rohrdantz et al. 2012; McNew, Derungs & Moran 2018; Comrie 2020)

dialect maps (Rabanus 2018; Rabanus 2020)

2 Methods

In the main text of this paper only a high-level outline of the visualisations will be presented. An appendix is added to the paper with more detailed information on how to create the visualisations. However, even this appendix will still be rather abstract in the sense that it will not contain any code or detailed instructions on how to use specific software. The actual code that has been used to prepare the visualisations are available in a separate online repository.

3 Projection

The surface of the world is curved and for that reason a decision is needed how to project this curved surface onto a flat surface like a screen or a piece of paper. The problem of how to use such geographical projections is known for centuries and many different projections have been developed. Modern geographical software, either in form of a dedicated GIS or of geographical libraries for general programming languages like Python or R, make it really easy to switch between different projections. Mostly it takes just a single click or a single line of code. So there is really no excuse not to spend at least a few thoughts on the choice of the projection when trying to represent geographical data.

When dealing with linguistic diversity, the most detailed geographical data to

be represented is dialectal variation. The graphical representation of such local variation is actually not strongly influenced by different projections. For example, consider three different projections for the outline of Germany shown in Figure 1. Although clear differences are visible, the effects of different projections do not seem to be very meaningful for the interpretation of linguistic variation in such an area.

The situation is quite different when dealing with global data. The choice of projection can have a strong impact on the interpretation of the data. As an example, consider the projections shown in Figure 2. The rectangular longitude-latitude projections (top left) is often used (e.g. in the online version of WALS, Dryer & Haspelmath 2013) because it allows for the direct specification of longitude-latitude coordinates, so it might feel like a natural unopinionated display. However, the use of this visual characterisation is a decision and it is actually rather unfitting for the display of worldwide linguistic variation. Landmasses close to the North Pole are relatively large, while these areas generally contain little linguistic variation. Note that the specific projection in Figure 2 is already a slightly changed longitude-latitude projection, because it is rotated 11.5 degrees to the west to put the cut-off in the Bering Strait.

The other three projections shown in Figure 2 are proposed here are better choices for the projection of worldwide linguistic variation. All three are area-preserving projections, which means that the size of the areas in the display is proportional to the area of the surfaces on the globe. The projection in the top-right of Figure 2 is a Mollweide projection, which is a compromise between the rectangular longitude-latitude projection and an equal-area projection. Specifically, shown here is an “interrupted” Mollweide projection, which breaks open

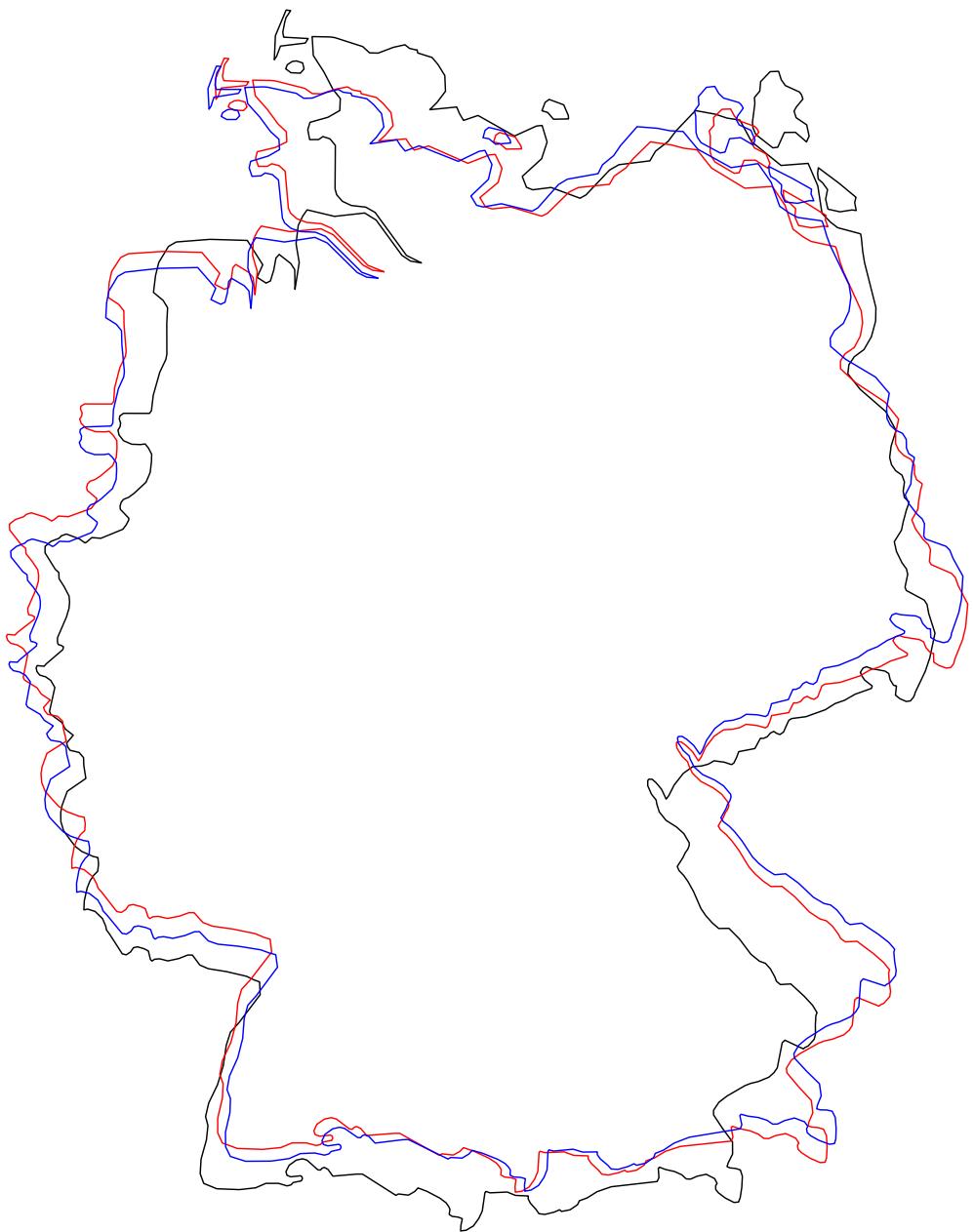


Figure 1: Different projections of Germany, centralised and scaled for comparison: a rectangular longitude-latitude plot (EPSG:3857 in black), an equal-area Mollweide projection (ESRI:54009 in red), and an azimuthal equal-area projection (ESRI:54032 in blue).

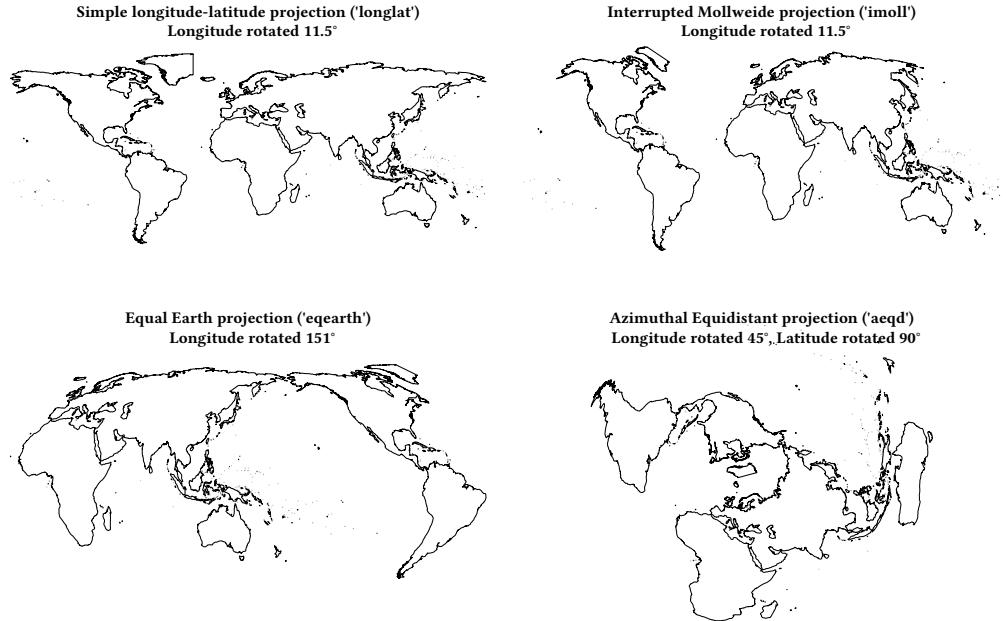


Figure 2: Different projections of the World

the projection in the oceans to more faithfully represent the outline of the landmasses. Like the previous projection, this Mollweide projection is rotated to put the cut-off in the Bering Strait.

Shown in the bottom-left of Figure 2 is a so-called Equal Earth projection (Šavrič, Patterson & Jenny 2019). This is a relatively new projection intended as an improvement on the Robinson projection (which was used in the original printed version of the WALS, Haspelmath et al. 2005). This projection seems particularly useful for a pacific-centered display, because it has still a reasonably recognizable outline of the extreme left and right coastlines.

Finally, a somewhat unusual but promising projection is shown bottom-right in Figure 2. This is an azimuthal equidistant projection, for example used in the United Nations Emblem. The biggest drawback of this projection is the rather

strong distortion of the form of Australia.

4 Weighting

Even when using a suitable projection, there is still another problem with the display of world-wide linguistic variation. The number of languages spoken around the world is very unevenly distributed geographically. There are some regions in which very many different languages are spoken in close proximity, while other areas only few different languages are attested. Any display that does not take this effect into account will have an implicit visual bias in they display of the world-wide linguistic variation.

For example, when using equally-sized symbols in linguistic maps (like in WALS) this often leads to overlapping symbols. The hidden portions of symbols induces less visual prominence for these languages and are thus an implicit distortion of the visual display. This problem was explicitly addressed by McNew, Derung & Moran (2018), who suggest a Thiessen/Voronoi-tesselation as an alternative. In such a display

Distortion to visually weight the prominence of certain areas.

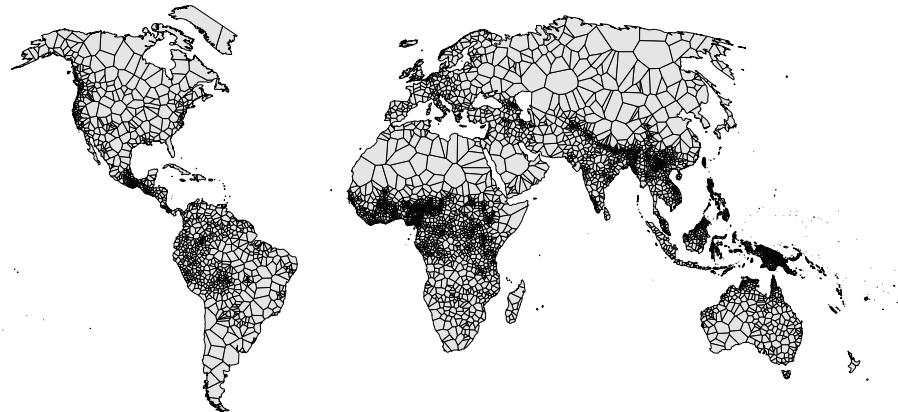
To be able to distort

Populationdata from (Amano et al. 2014)

5 Interpolation

(Thebpanya, Hatfield & Lee 2020)

Voronoi diagram of all glottolog languages

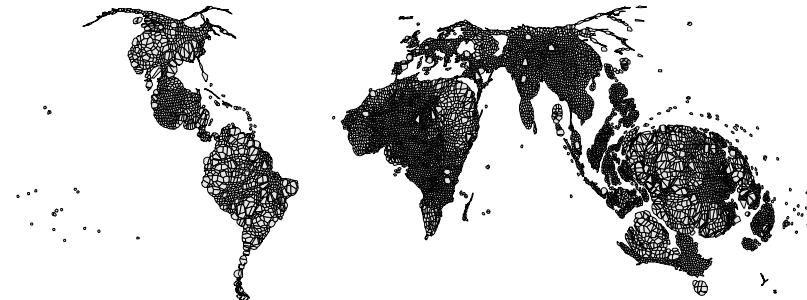


Weighting all polygons to have equal surface

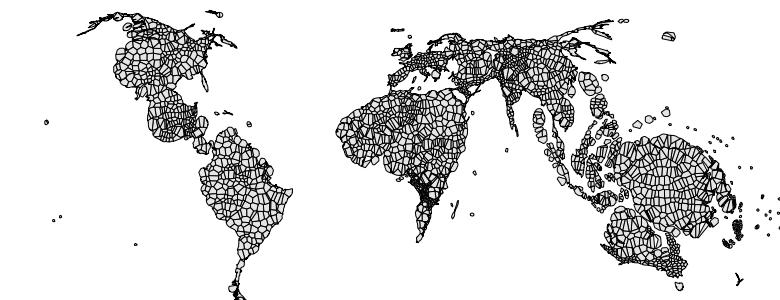


Figure 3: Cartogram with equal area for all languages in the Glottolog

Weighting all families from Glottolog equally



Weighting all genera from WALS equally



Weighting languages by logarithm of number of speakers

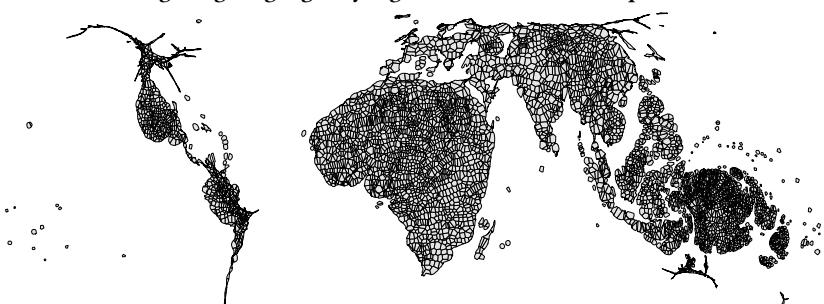


Figure 4: Cartogram with different kinds of weighting

6 Prediction

7 Conclusion

8 Appendix

8.1 Projections

Different projections can be easily applied by using the algorithms and specifications from PROJ (PROJ contributors 2024). The following PROJ-strings can be used to create the projections shown in Figure 2:

- Longitude-latitude projection: “+proj=longlat +long_0=11.5”
- Interrupted Mollweide projection: “+proj=imoll +long_0=11.5”
- Equal Earth projection: “+proj=eqearth +lon_0=151”
- Azimuthal Equal Distance projection: “+proj=aeqd +lat_0=90 +lon_0=45”

Similar to the Azimuth projection, the Fuller projection is a way to solve the distortion of Australia. Such a Fuller projection is for example used for the display of human migrations based on genetic data on Wikipedia.¹ Because of copyright reasons the Fuller projection is currently not available in PROJ.

References

Amano, Tatsuya, Brody Sandel, Heidi Eager, Edouard Bulteau, Jens-Christian Svenning, Bo Dalsgaard, Carsten Rahbek, Richard G. Davies & William J.

¹https://commons.wikimedia.org/wiki/File:World_map_of_prehistoric_human_migrations.jpg, accessed 21 August 2014.

- Sutherland. 2014. Global distribution and drivers of language extinction risk. *Proceedings of the Royal Society B: Biological Sciences* 281(1793). 20141574. doi:10.1098/rspb.2014.1574.
- Comrie, Bernard. 2020. Mapping the world's languages: From data via purpose to representation. In Stanley D. Brunn & Roland Kehrein (eds.), *Handbook of the changing world language map*, 61–76. Springer. doi:10.1007/978-3-030-02438-3_131.
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. v2020.3 edn. Leipzig: Max Planck Institute for Evolutionary Anthropology. doi:10.5281/zenodo.7385533.
- Haspelmath, Martin, Matthew S. Dryer, Bernard Comrie & David Gil. 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- McNew, Garland, Curdin Derungs & Steven Moran. 2018. Towards faithfully visualizing global linguistic diversity. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, et al. (eds.), *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA). <https://aclanthology.org/L18-1129>.
- PROJ contributors. 2024. *PROJ coordinate transformation software library*. Open Source Geospatial Foundation. doi:10.5281/zenodo.5884394. <https://proj.org/>.
- Rabanus, Stefan. 2018. Dialect maps. In Charles Boberg, John Nerbonne & Dominic Watt (eds.), *The handbook of dialectology*, 348–367. Oxford: Wiley. doi:10.1002/9781118827628.ch20.
- Rabanus, Stefan. 2020. Language mapping worldwide: Methods and traditions. In Stanley D. Brunn & Roland Kehrein (eds.), *Handbook of the changing world*

- language map*, 103–129. Springer. doi:10.1007/978-3-030-02438-3_151.
- Rohrdantz, Christian, Michael Hund, Thomas Mayer, Bernhard Wälchli & Daniel A. Keim. 2012. The world's languages explorer: Visual analysis of language features in genealogical and areal contexts. *Computer Graphics Forum* 31(3pt1). 935–944. doi:10.1111/j.1467-8659.2012.03086.x.
- Šavrič, Bojan, Tom Patterson & Bernhard Jenny. 2019. The Equal Earth map projection. *International Journal of Geographical Information Science* 33(3). 454–465. doi:10.1080/13658816.2018.1504949.
- Thebpanya, Paporn, Sudarat Leerabhandh Hatfield & Jay Lee. 2020. Visualizing dialect variation on a 3-D interpolated map: A case study in Chiang Mai, Thailand. In Stanley D. Brunn & Roland Kehrein (eds.), *Handbook of the changing world language map*, 189–200. Springer. doi:10.1007/978-3-319-73400-2_90-1.