Mapping the endangerment status of the world's languages

Andrew Caines, apc38@cam.ac.uk
Department of Theoretical & Applied Linguistics, University of Cambridge

I approach the challenge of visualising the endangerment status of the world's known past and present languages, with data taken from the Glottolog linguistic database[1]. Each of the 'languoids' (languages and dialects; any distinction between the two is fuzzy) in the database has been assigned a status, from 'living' to 'vulnerable' to 'endangered' and ultimately 'extinct'.

The obvious method of display is a world map: I demonstrate how this is possible using R[2] (e.g. Figure 1). However, with more than 7000 datapoints to display, and a high density of points in certain regions such as sub-Saharan Africa and south-east Asia, the visualisation would in this case be greatly enhanced by interactivity. I therefore consider what various R libraries can offer in terms of interactivity, including Shiny[3], ggvis[4], and Leaflet[5]. I also discuss the shortcomings of each approach, and the ideal future outcome of this work.

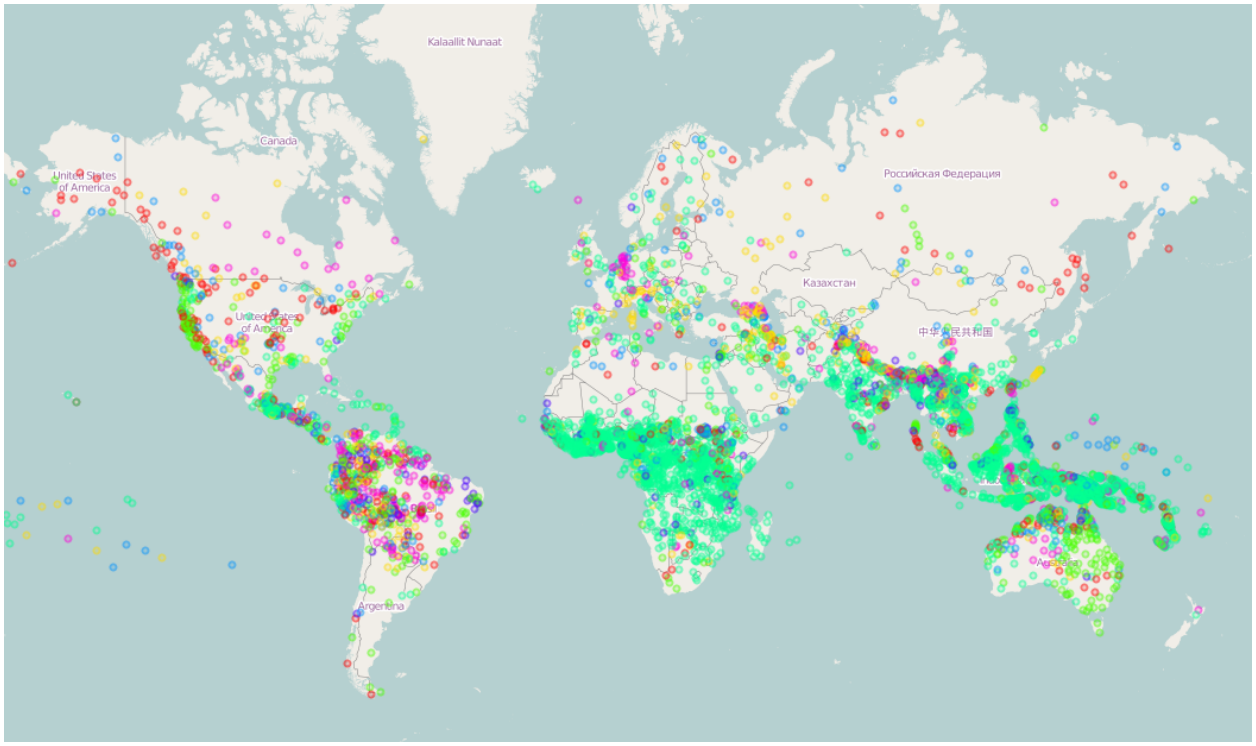

Figure 1: world map of 7429 Glottolog languoids, drawn using the R 'Leaflet' package

[1] http://glottolog.org
[2] http://cran.r-project.org
[3] http://shiny.rstudio.com
[4] http://ggvis.rstudio.com
[5] http://rstudio.github.io/leaflet