

# Linguistic Data Analysis - Final Project

Benedict Wuethrich

2024-07-09

Everything related to this project can be found in this GitHub repository.

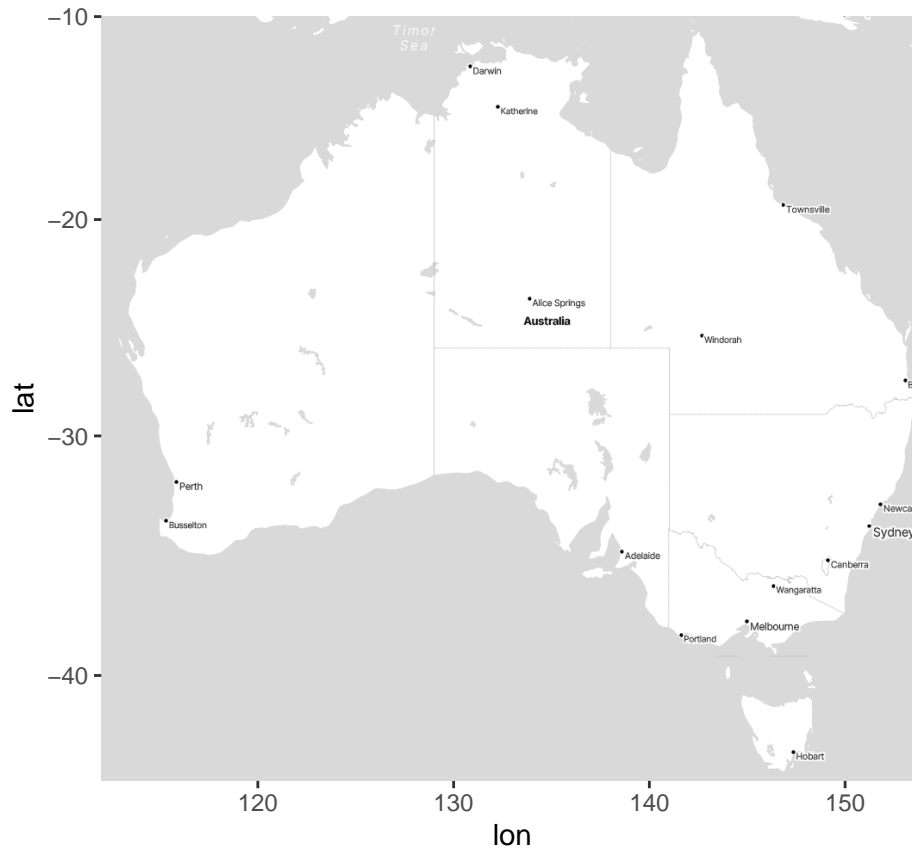
#Introduction

#Data Wrangling The data wrangling for this project was relatively simple, since the datasets provided by WALS and phoible respectively are quite clean already.

```
map_AUS <- get_stadiamap(bbox = c(left = 112,
                                   bottom = -44,
                                   right = 154,
                                   top = -10), #here the map boundaries given by ChatGPT earlier came in
                        zoom = 5,
                        maptype = "stamen_toner_lite",
                        color = "color")
```

## i © Stadia Maps © Stamen Design © OpenMapTiles © OpenStreetMap contributors.

```
ggmap(map_AUS)
```



```
map_AUS_family <- ggmap(map_AUS) +
  geom_point(data = wals_aus,
             aes(x = Longitude,
                 y = Latitude,
                 color = Family),
             size = 9) +
  scale_color_colorblind() + #better color
  theme_map() + #removes axes labels and puts the legend in bottom left corner of map
  theme(legend.text = element_text(size = 10),
        legend.title = element_text(size = 12)) +
  # legend.position.inside = c(0.9,0)) + #we can still modify it to our liking
  labs(color = "Families")
map_AUS_family
```

```
## Warning: This manual palette can handle a maximum of 8 values. You have
## supplied 24
```

```
## Warning: Removed 162 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



## Bibliography

- Bouckaert, Remco R., Claire Bowern & Quentin D. Atkinson (2018). The origin and expansion of Pama–Nyungan languages across Australia. *Nature Ecology & Evolution* (2), 741–749.
- Bowern, Claire / Koch, Harold (Hrsg.) (2004): *Australian Languages. Classification and the comparative method*. Amsterdam / Philadelphia: John Benjamins.
- Dixon, R.M.W. (1980): *The Languages of Australia*. Cambridge: Cambridge University Press.
- Dixon, R.M.W. (2004): *Australian Languages. Their Nature and Development*. Cambridge: Cambridge University Press.
- Kahle, D., Wickham, H. ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144–161. <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- Matthew S. Dryer. (2013) Order of Subject, Object and Verb. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *WALS Online (v2020.3)* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7385533> (Available online at <http://wals.info/chapter/81>, Accessed on 2024-04-16.)
- O’Grady, G. N. (1998). Toward a Proto-Pama-Nyungan Stem List, Part I: Sets J1–J25. *Oceanic Linguistics*, 37(2), 209–233.
- Schmidt, W. (1919). *Die Gliederung der australischen Sprachen: geographische, bibliographische, linguistische Grundzüge der Erforschung der australischen Sprachen*. Mechitharisten-Buchdruckerei.
- Wickham, H et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4 (43), 1686. [doi:10.21105/joss.01686](https://doi.org/10.21105/joss.01686) <https://doi.org/10.21105/joss.01686>
- Zuckermann, G. et al. (2021) LARA in the Service of Revivalistics and Documentary Linguistics: Community Engagement and Endangered Languages. *Proceedings of the Workshop on Computational Methods for Endangered Languages* (1), 13–23.