Pohnpei Sohte Ehu

Quantitative methods for finding emergent heteroglossic patterns in language attitudes

Bradley Rentz

24 Oct 2017

Kalahngan en...

- The Bilinski Foundation
- My dissertation committee: Andrea, Amy, Christina, Emelihter, and David
- Carisma, Celine, Cartina, Rofino, Banae, Maynard, Diana, Jade, Carvy, and many others on Pohnpei

Language attitudes background

What are language attitudes?

- Language attitudes are the opinions, beliefs, and biases that we have about language use
- They are the product of social interaction (discourse) (Potter & Wetherell 1987)
- Can only be observed as an emergent product of discourse
- Highly contextualized and unstable
- Are also influenced by all of one's linguistic experiences (education, work, family, friends, geography, politics, etc.)

Heteroglossia

- An important aspect of the study of language attitudes is the notion of *heteroglossia*
- Heteroglossia according to Bakhtin (1981) is the idea that:
  - Every language has internal diversity (whether class, social, ethnic, political, historical, geographic, or otherwise)
  - A community may also have access to other languages as well with their own diversity
  - An individual has multiple voices, whereby they navigate through the complexity of the linguistic resources available to them
- Language attitudes are both a product of heteroglossia and a reinforcer of it
- Any language attitudes study must take this diversity into account

Quantitative approaches to language attitudes

Past quantitative methods

- Typical quantitative research question: what language attitudes do certain groups have and to what extent are they different from other groups
- Many past language attitudes studies have primarily used regressions and ANOVAs (or ANCOVA, MANOVA, etc.) to analyze group differences
- These methods require the use of demographic variables to define groups

What does regression assume?

- The variables used are a meaningful representation of reality
- There is a single consistent pattern within those groups (aka they approximate a specific normal distribution)
    - There can only be one mean, median, and mode (aka one norm) and deviation from that is noise
- The groups are isolated and are independent of each other
- Items can only be in one group

Problems with demographic variables

- Represent social constructs (e.g., age, gender, geographic place, education)
- Social constructs are messy and have lots of inherent variation
- Regression assumes an essentialist view of these variables
- They are often coded with overly simplistic and conservative values (i.e., binary gender)
- Western colonial constructs are often used instead of potentially more meaningful local categories
- Require arbitrary cut-offs and groups that may not have any actual behavioral or identity correlates (e.g., age)
- But can still be useful if done meaningfully

Example demographic variable: gender

- Do each of the genders exhibit different language attitudes?
- Imagine all the different women, men, and non-binary people you know
- How many different opinions, ideologies, and beliefs on a given topic can you think of for each gender?
- Maybe there are consistent differences at a small level, such as for a single question, but probably not
- What if there are several different trends that are specific to one gender?

Is regression useful for language attitude studies?

- Yes and no
- If the demographic variables used correspond to meaningful behaviors where there is a single observable trend for the given groups, then yes
- If the variables do not represent that, then probably no
- Regression can miss trends that occur across groups or multiple trends that occur within a single group

What can be done instead? MDS and cluster analyses

- There are other quantitative analyses that do not have these assumptions
- Multidimensional scaling (MDS) + cluster analysis allow for patterns to emerge from the data without a priori categories
- Allows for locally constructed patterns to become visible
- Shows variation that would have been missed otherwise

How does it work?

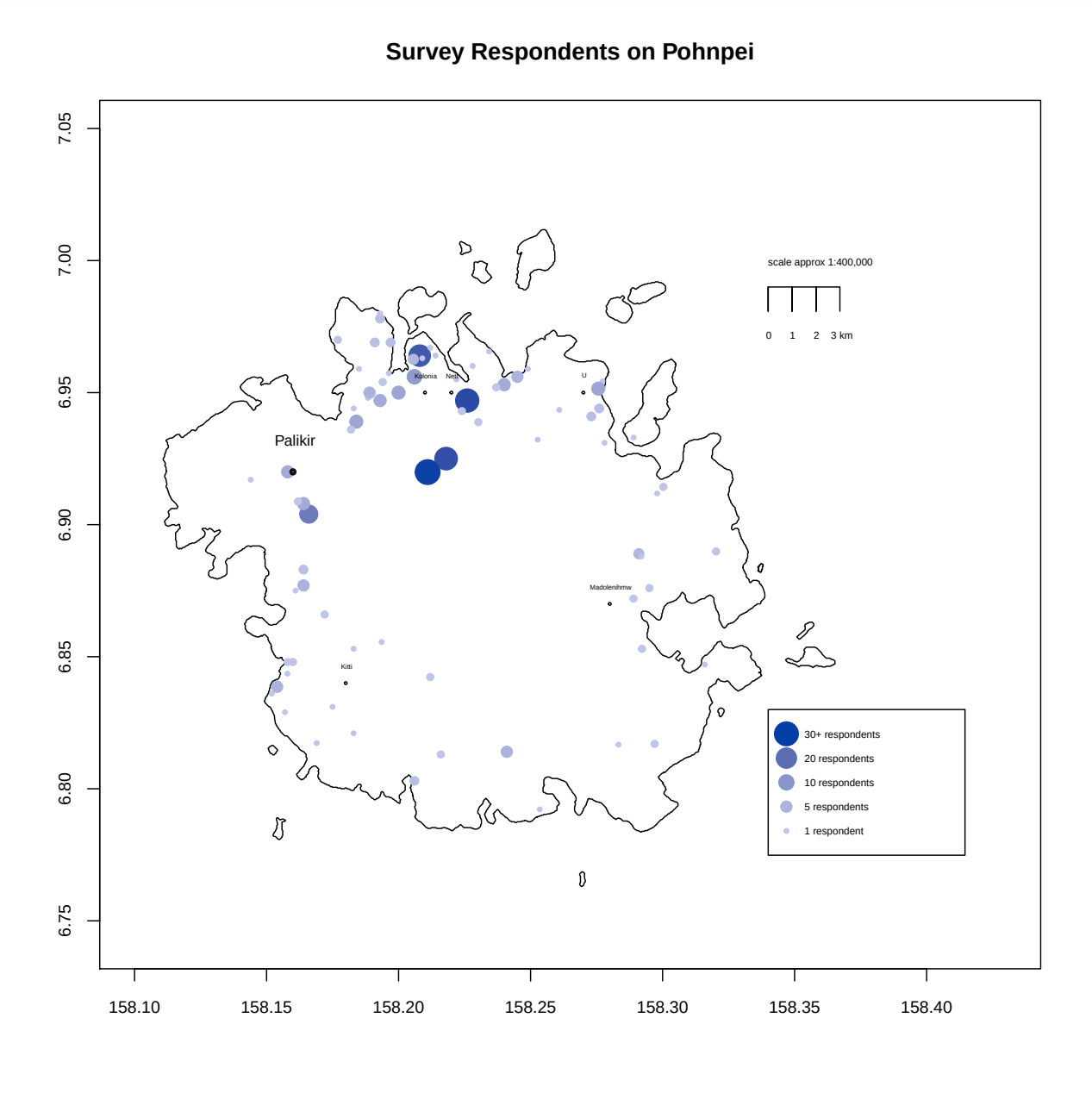
- 3 step process:
- 1.) Select the applicable response variables and create a dissimilarity matrix
    - This matrix indicate how dissimilar each respondent is to every other respondent (from 0–1) based on how they responded to those items
- 2.) Apply the MDS algorithm to the dissimilarity matrix
    - This takes the dissimilarity matrix and maps out all the respondents onto a two-dimensional plane
- 3.) Apply a clustering algorithm like Partitioning Around Mediods (PAM)
    - This groups the data into a pre-specified number (k) of clusters
    - k is determined based on the data by looking at the in-cluster variation vs. between-cluster variation and picking the number of clusters that maximizes the difference (Rousseeuw 1987)

Pohnpei case study

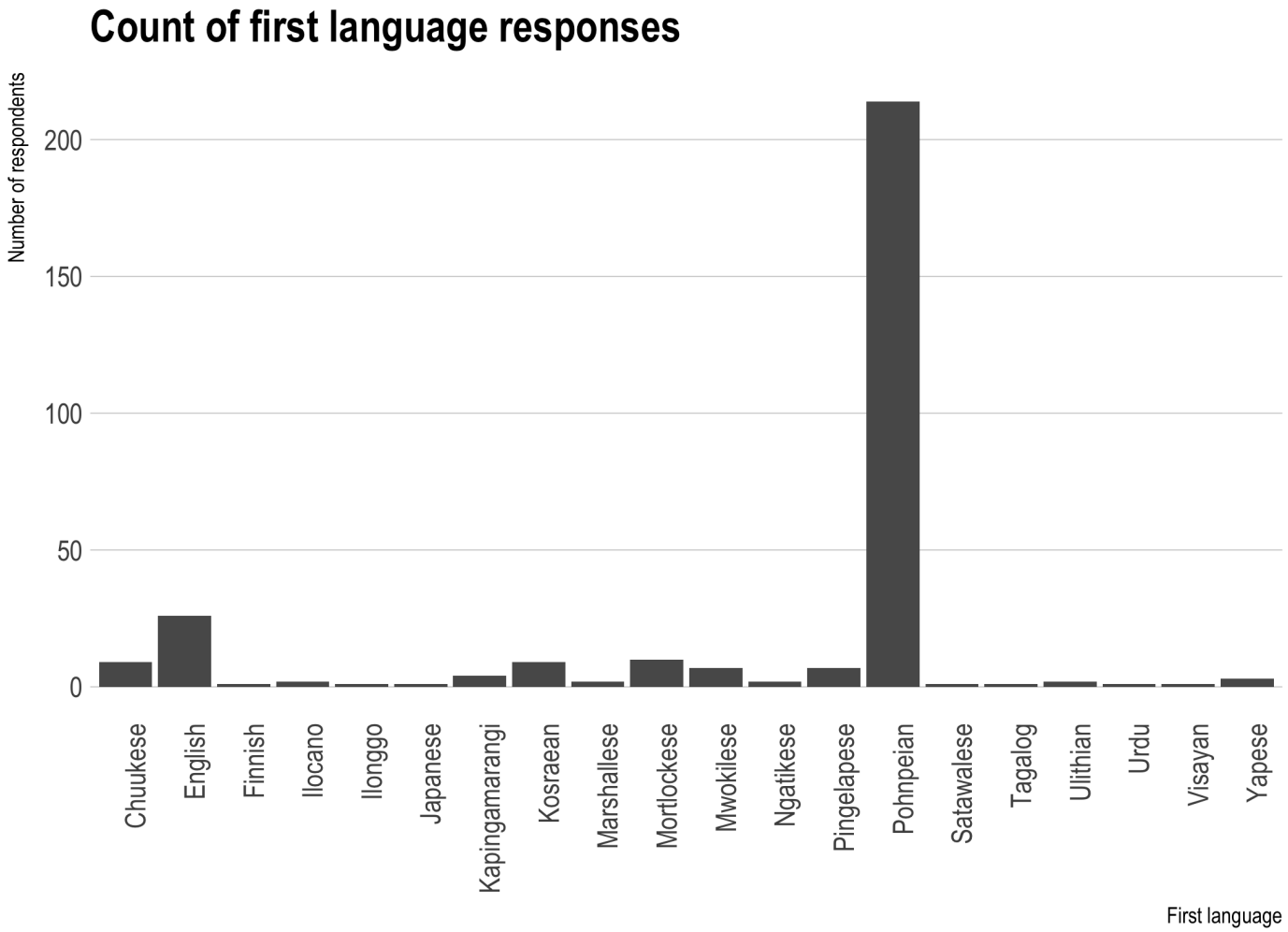
Pohnpei sohte ehu

- Pohnpei sohte ehu (Pohnpei is not one) is a common metaphor on the island
- Contrary to most views of ‘Micronesia’, Pohnpei is a place of high linguistic diversity and multilingualism
- This study (my dissertation) has 301 respondents (1.3% of the adult population)
- Respondents reported speaking over 30 languages
- Median number of languages spoken per respondent is 3

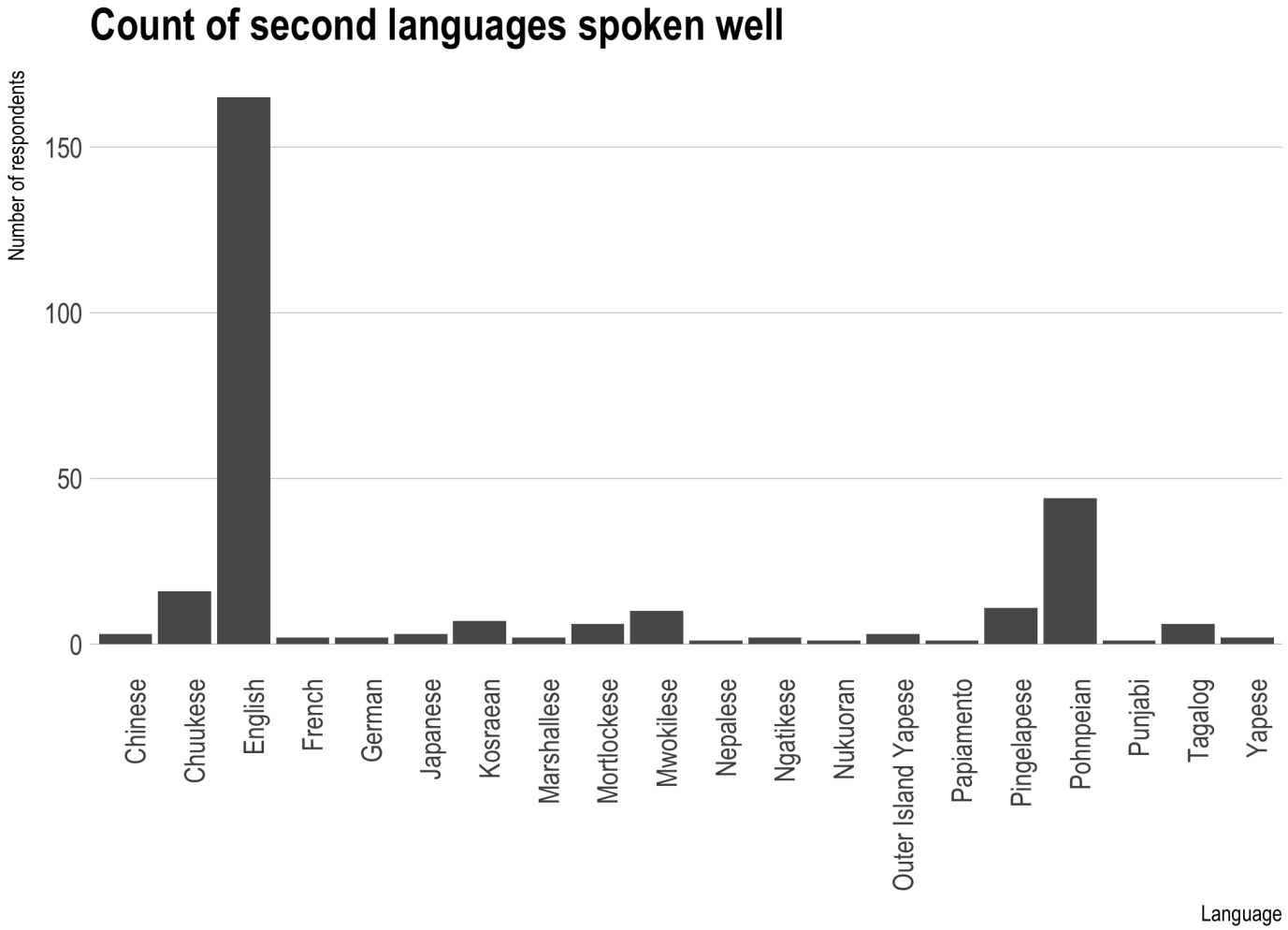
Map of respondents



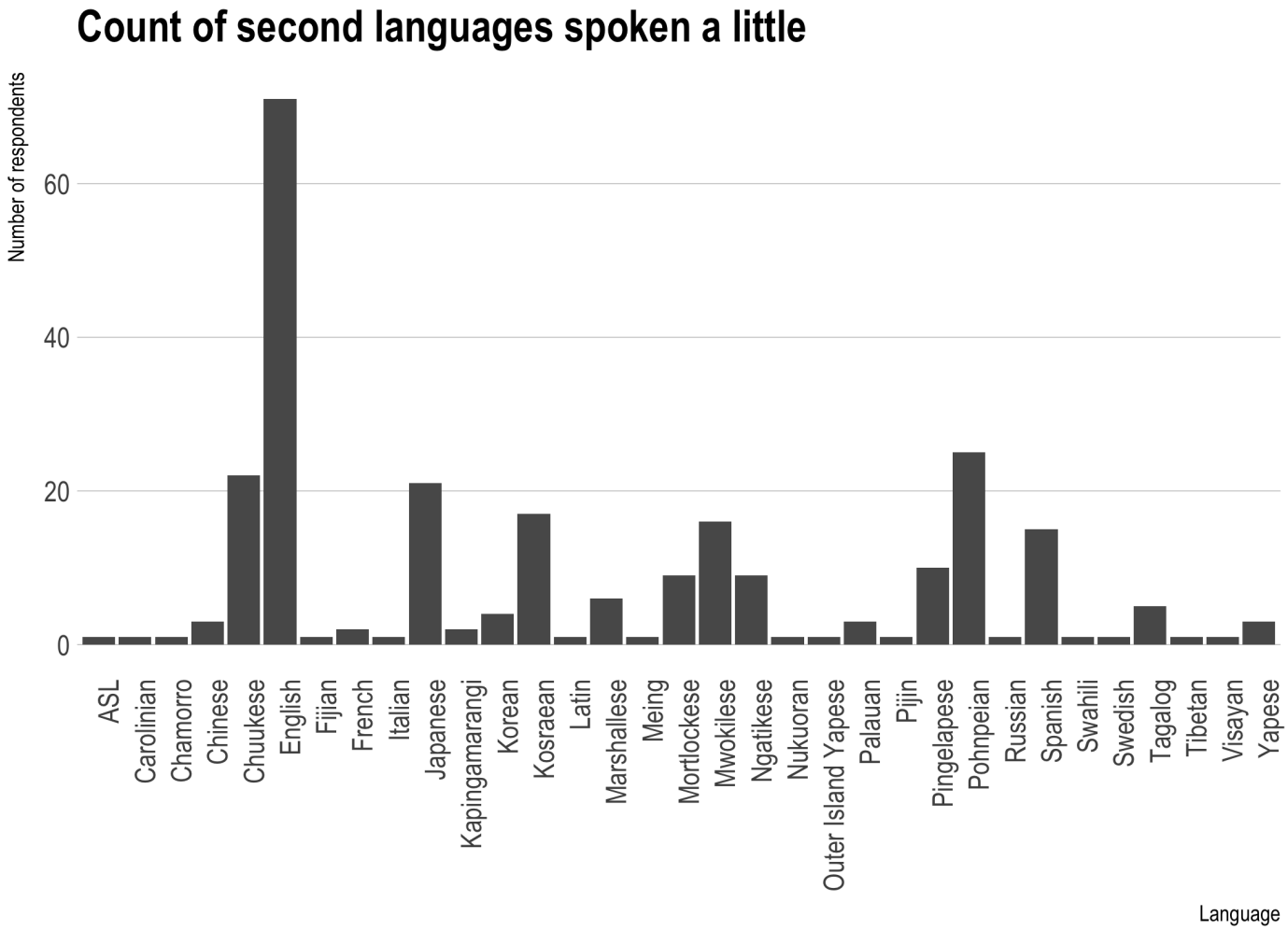
First languages of respondents



L2s spoken well by respondents



L2s spoken a little by respondents



Domain-based language attitudes

Domain-based language importance

- Respondents were asked to select 1 of 8 languages that is most important for 25 different domains
- These questions were analyzed via Bayesian hierarchical regression modeling and MDS + PAM

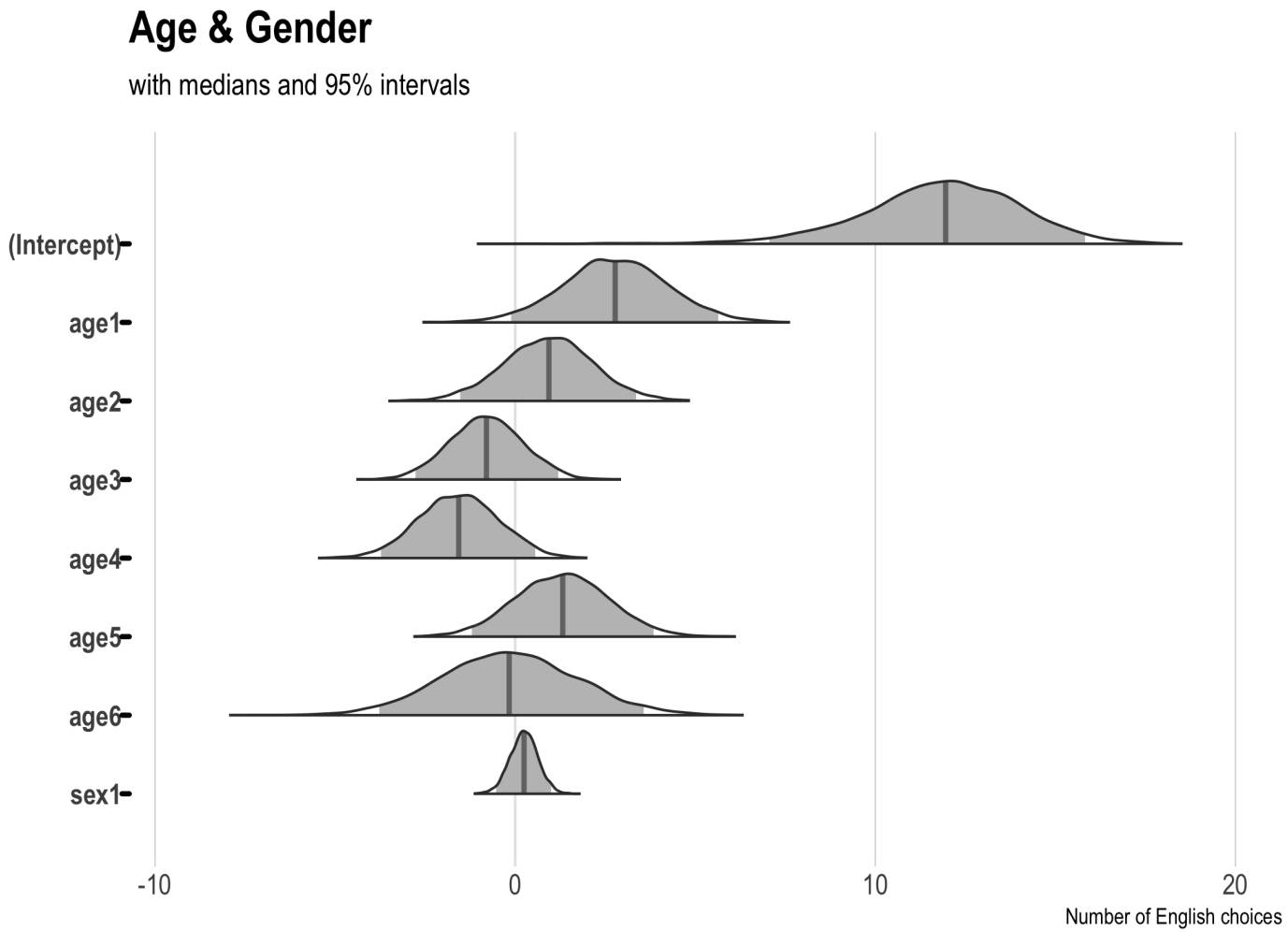
Regression

- Dependent variable: The number of times a respondent selected English for any of the 25 domains (0–25)
- Predictor variables: age, gender, birth location, travel abroad, years on Pohnpei, ability to speak the formal registers of Pohnpeian (Meing), education achievement level, elementary school type (public or private), high school type (public, private, or none)
- Mixed effects: Current section of island nested in current municipality
- Includes poststratification survey weights
- Predictors deviation coded

Posterior distributions for age and gender

Parameter	mean	sd	2.5%	97.5%	Meaning of predictor
(Intercept)	11.8	2.2	7.1	15.8	Grand mean
age1	2.8	1.5	-0.1	5.6	18–24 years old
age2	0.9	1.2	-1.5	3.4	25–34 years old
age3	-0.8	1.0	-2.8	1.2	35–45 years old
age4	-1.6	1.1	-3.7	0.6	45–54 years old
age5	1.3	1.3	-1.2	3.8	55–64 years old
age6	-0.2	1.9	-3.8	3.6	65–74 years old
sex1	0.2	0.4	-0.5	1.0	women

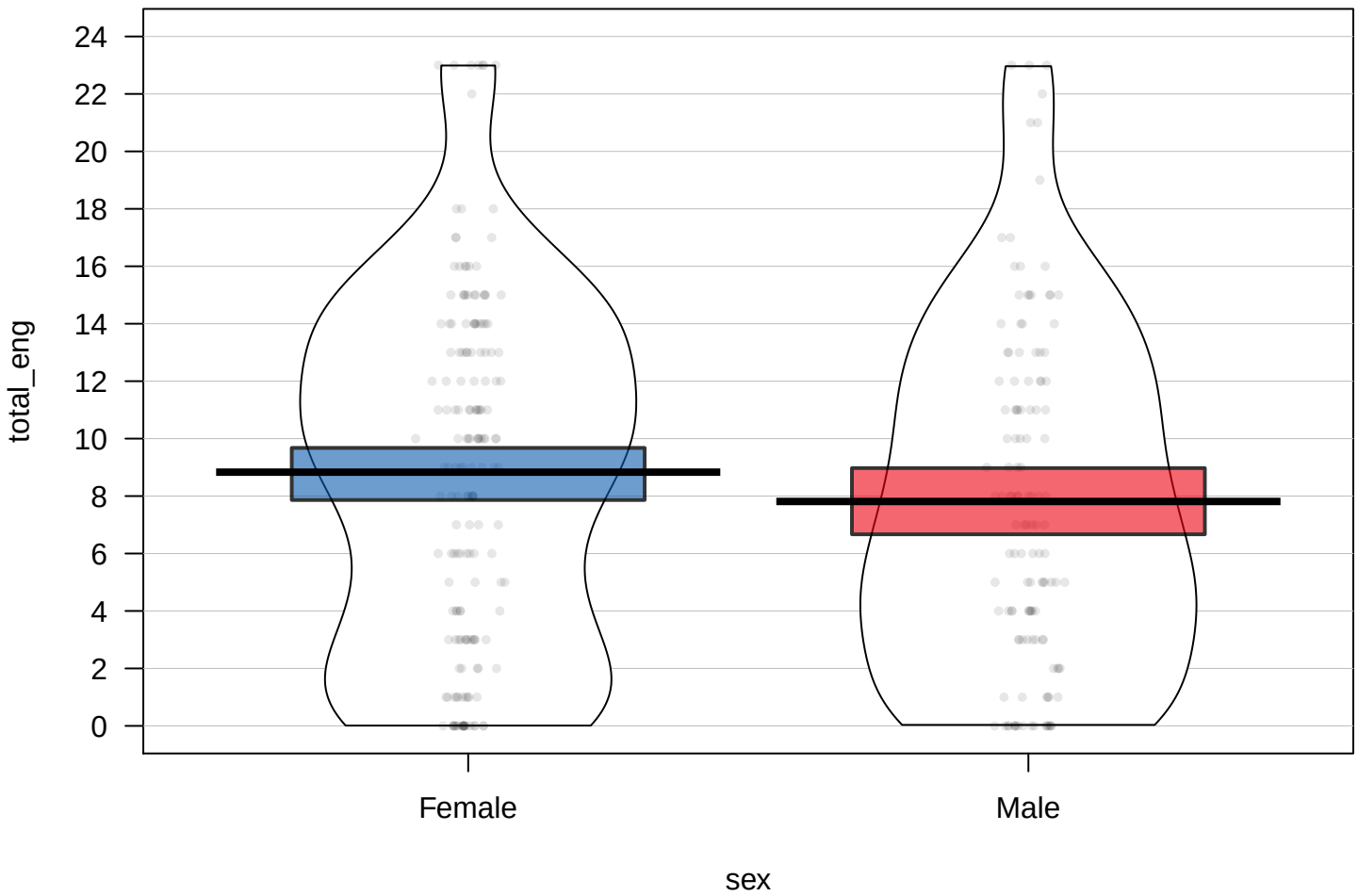
Posterior distributions for age and gender



Age and Gender

- There are some differences by age group but not a linear trend (stops with age5 and age6) and high variation
- Little to no difference based on gender
- Are these useful categories?
    - Do show some differences
    - But these are large groups with lots of inherent variation that is not seen
    - Gender has so much internal variation that it is no different from the grand mean

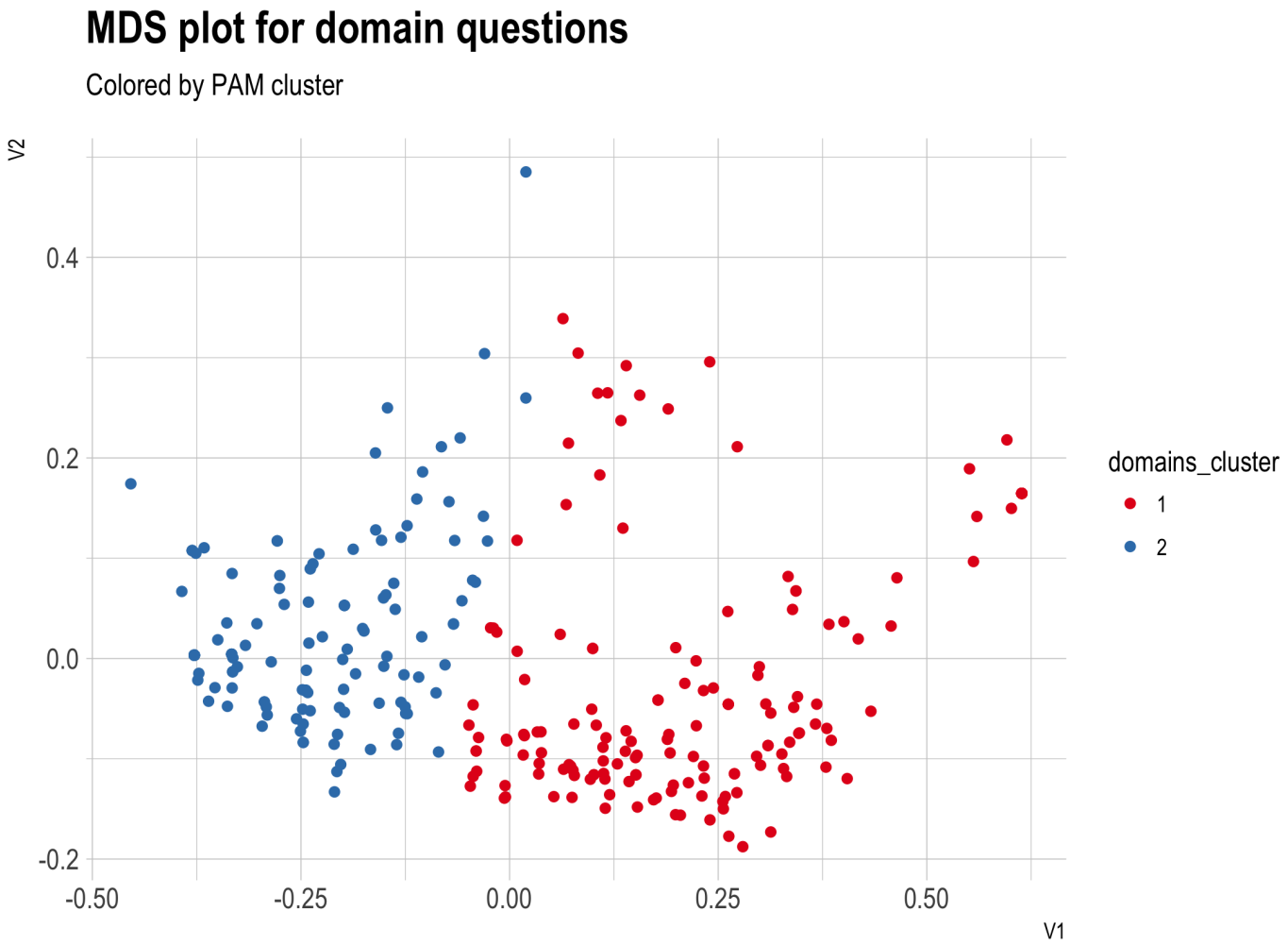
RDI plot of English choices by gender



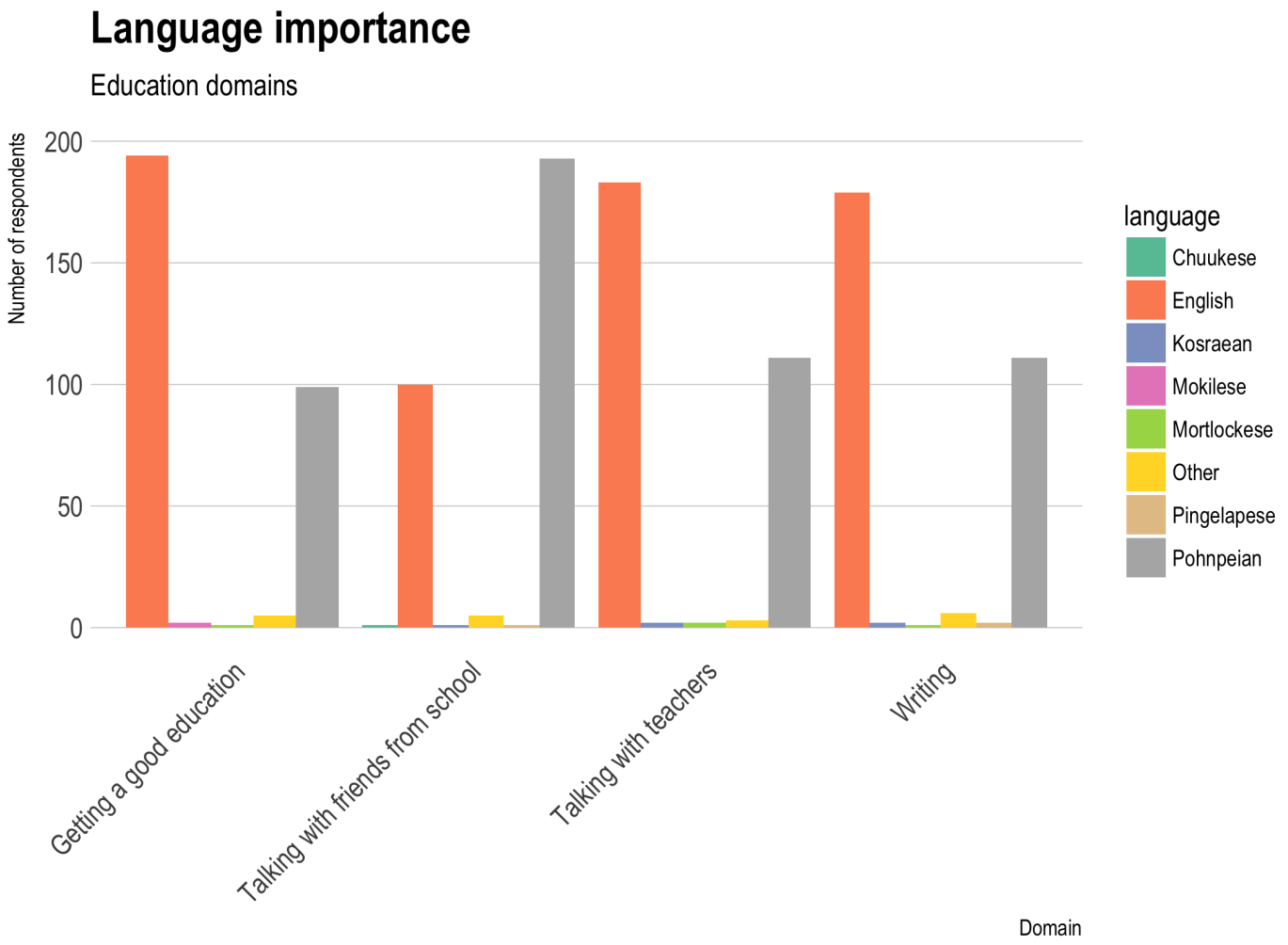
MDS and PAM

- Clustered all of the respondents into two groups based on their responses to the 25 questions
- Number of groups was determined by the data and not a priori
- Questions are grouped into 6 super-domains for display purposes: education, occupation, general, media, and social solidarity

MDS and PAM

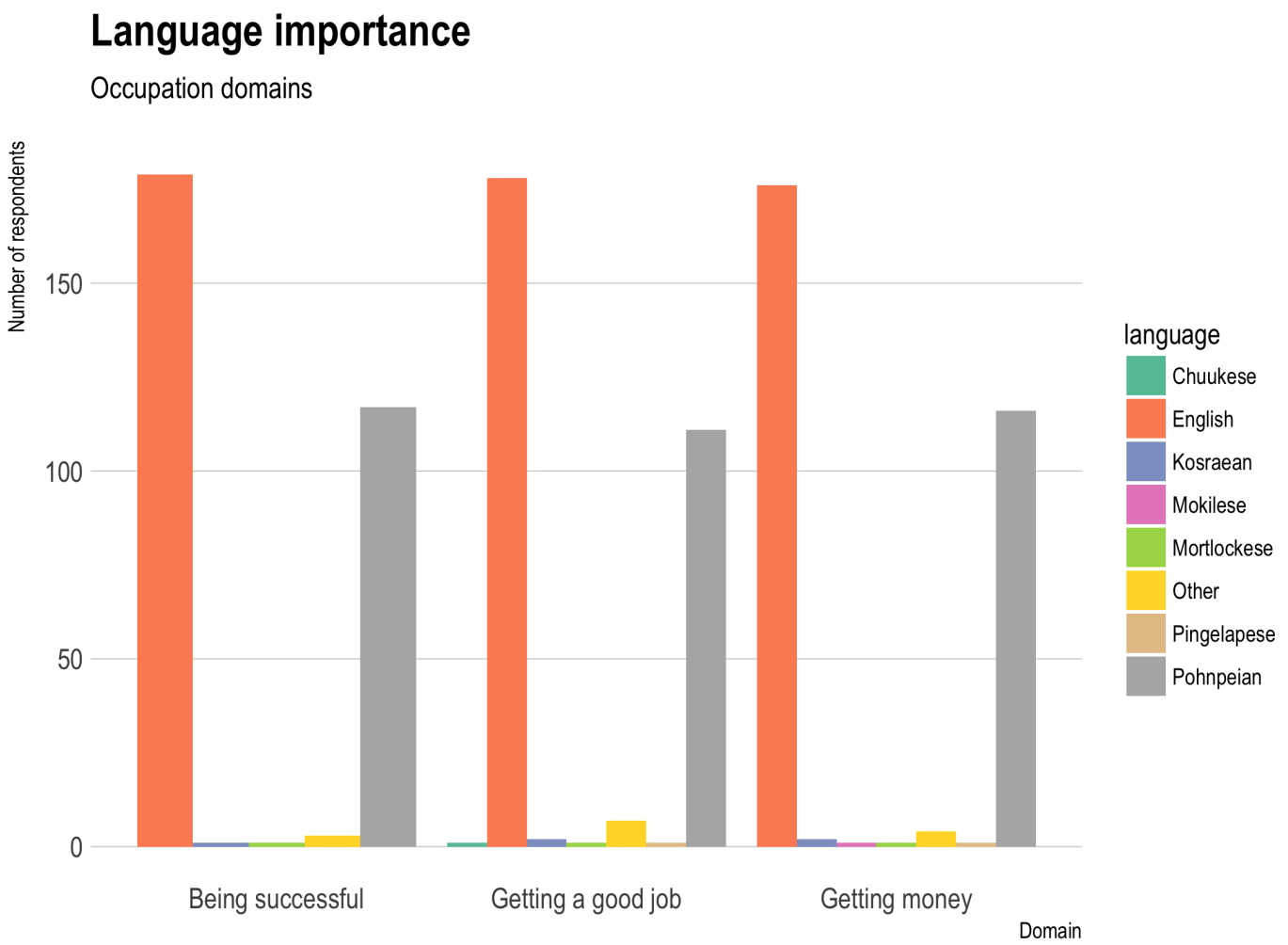

MDS plot for domain questions
Colored by PAM cluster
V2
0.4
0.2
0.0
-0.2
-0.50
-0.25
0.00
0.25
0.50
V1
domains_cluster
1
2

Education domains (original)

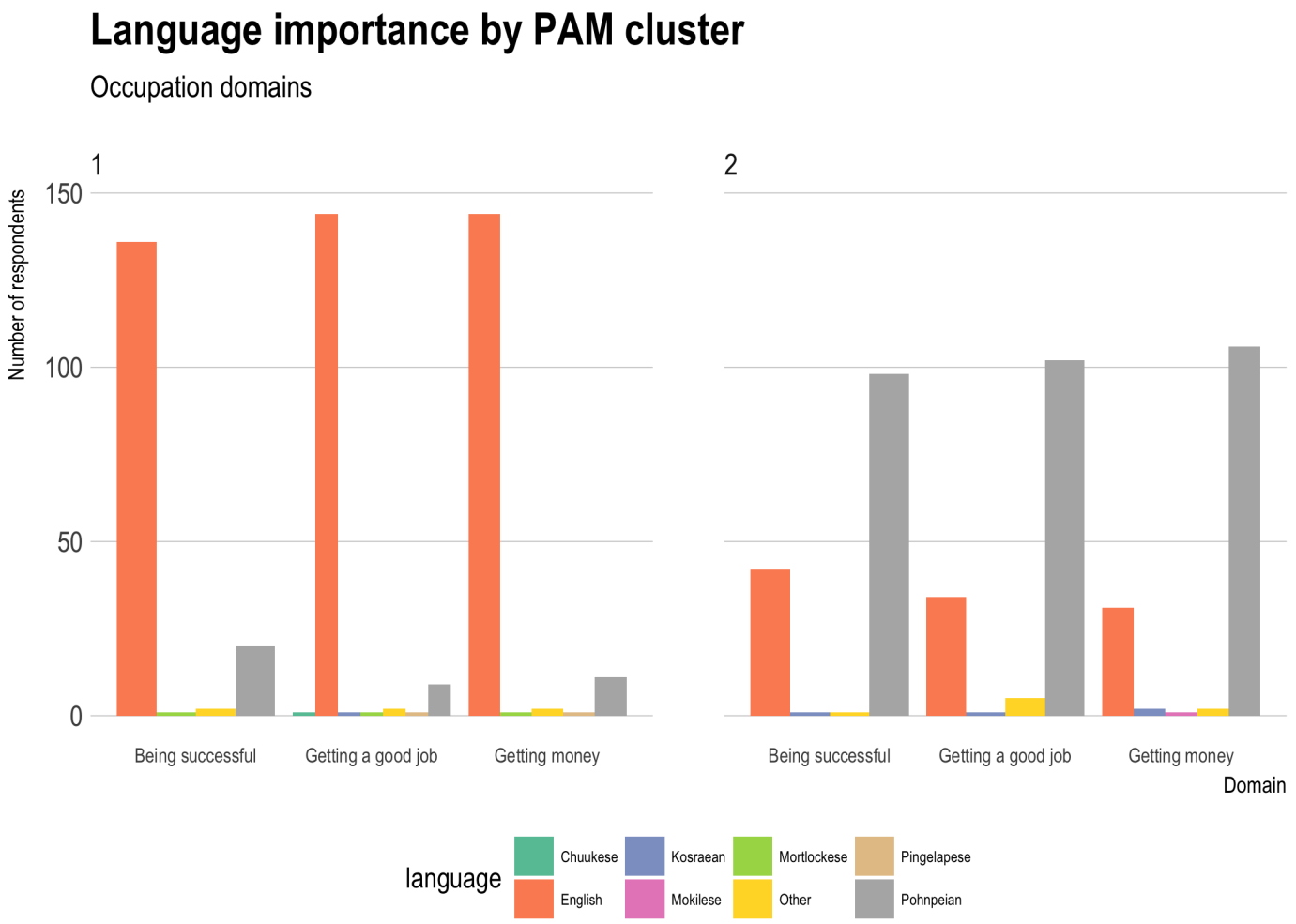


PAM Education domains

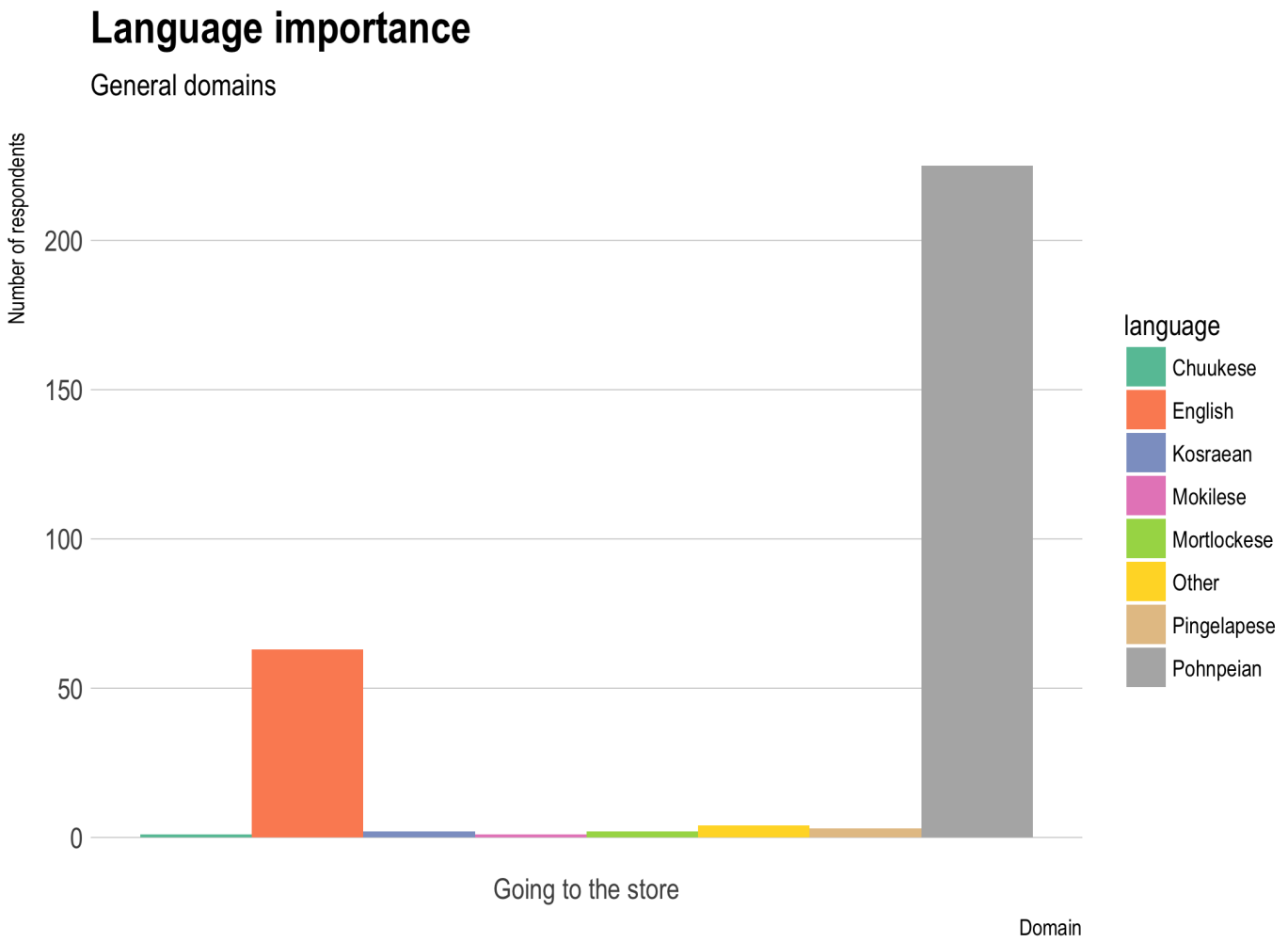
Occupation domains (original)



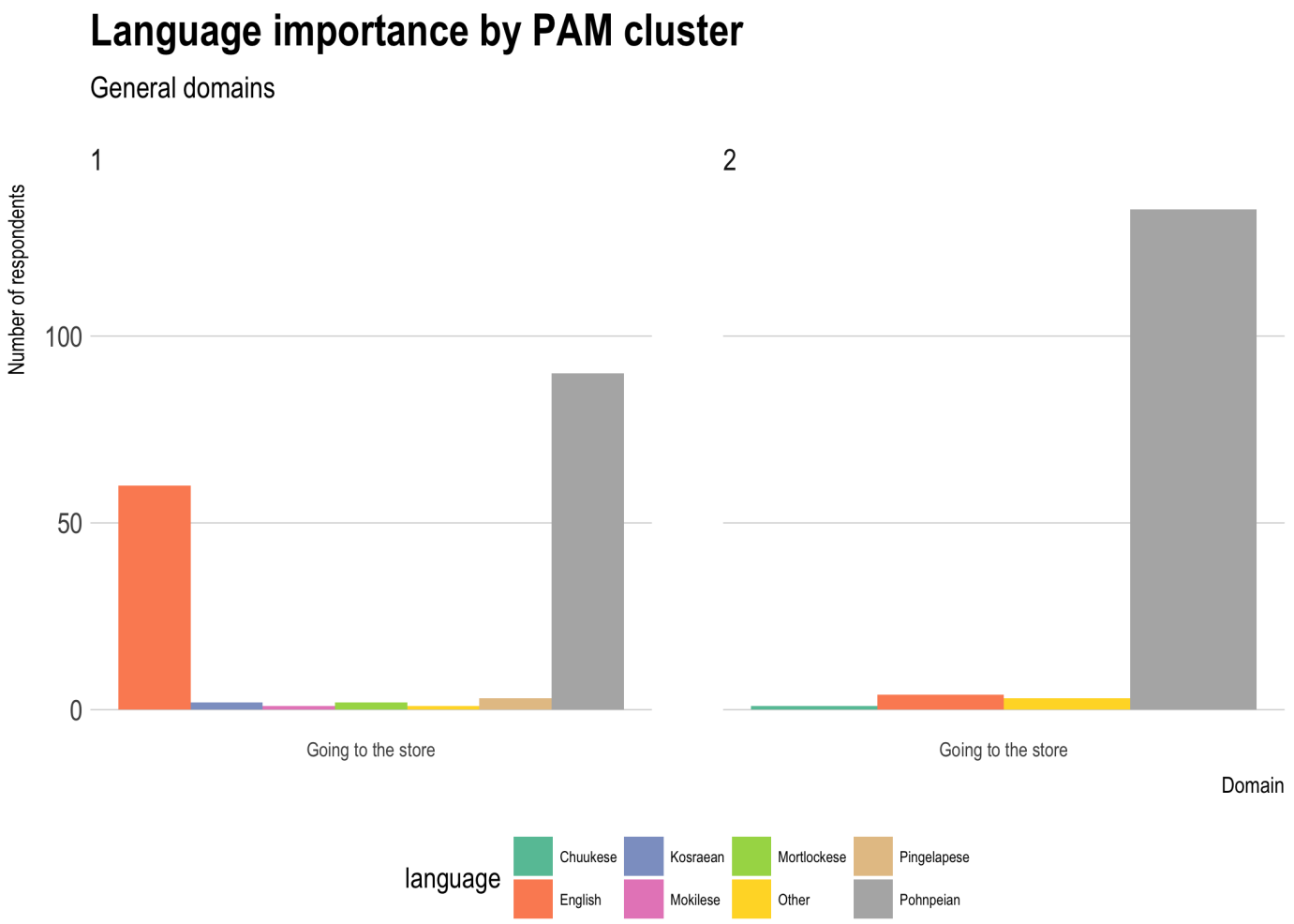
PAM Occupation domains



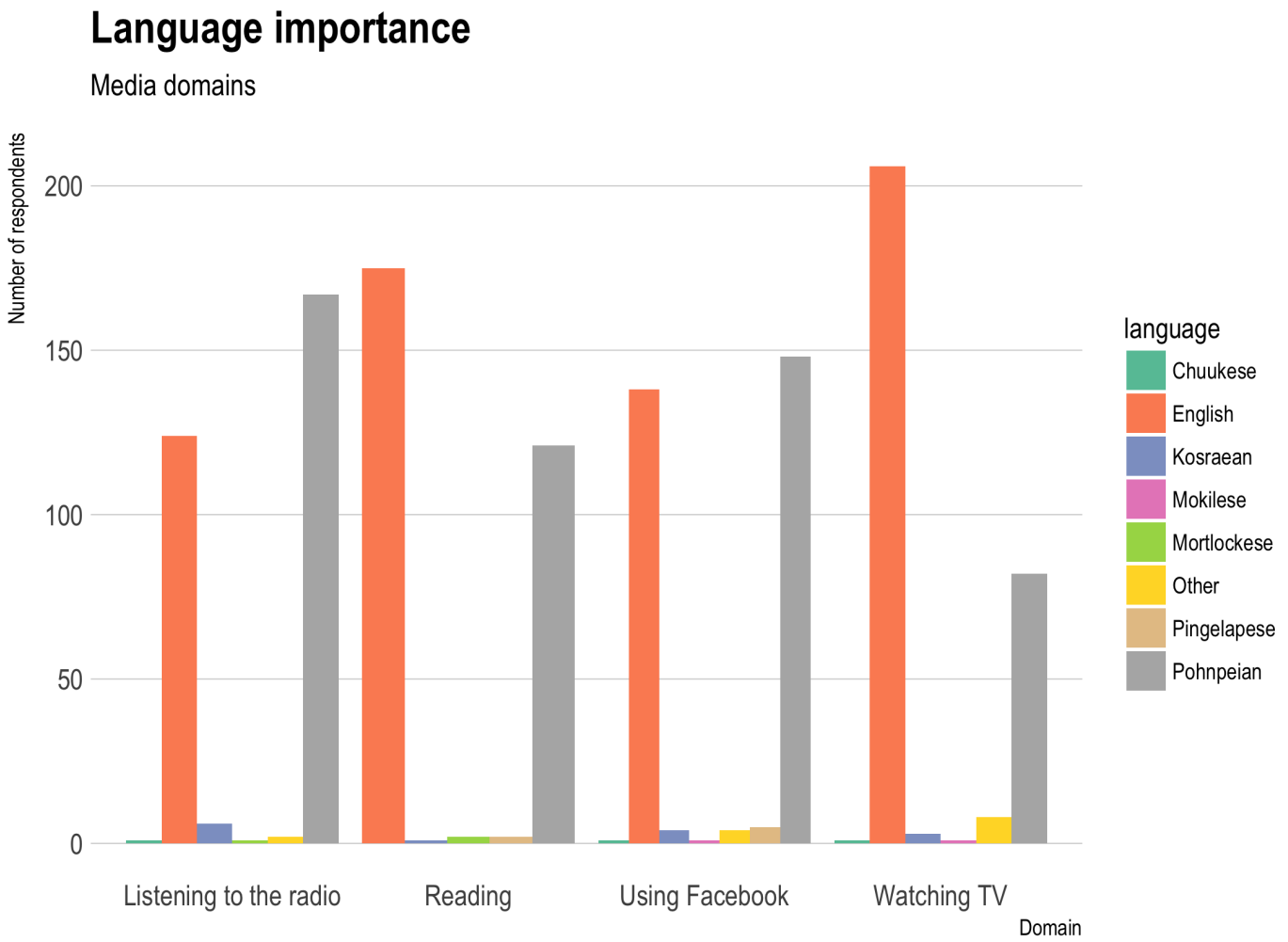
General domains (original)



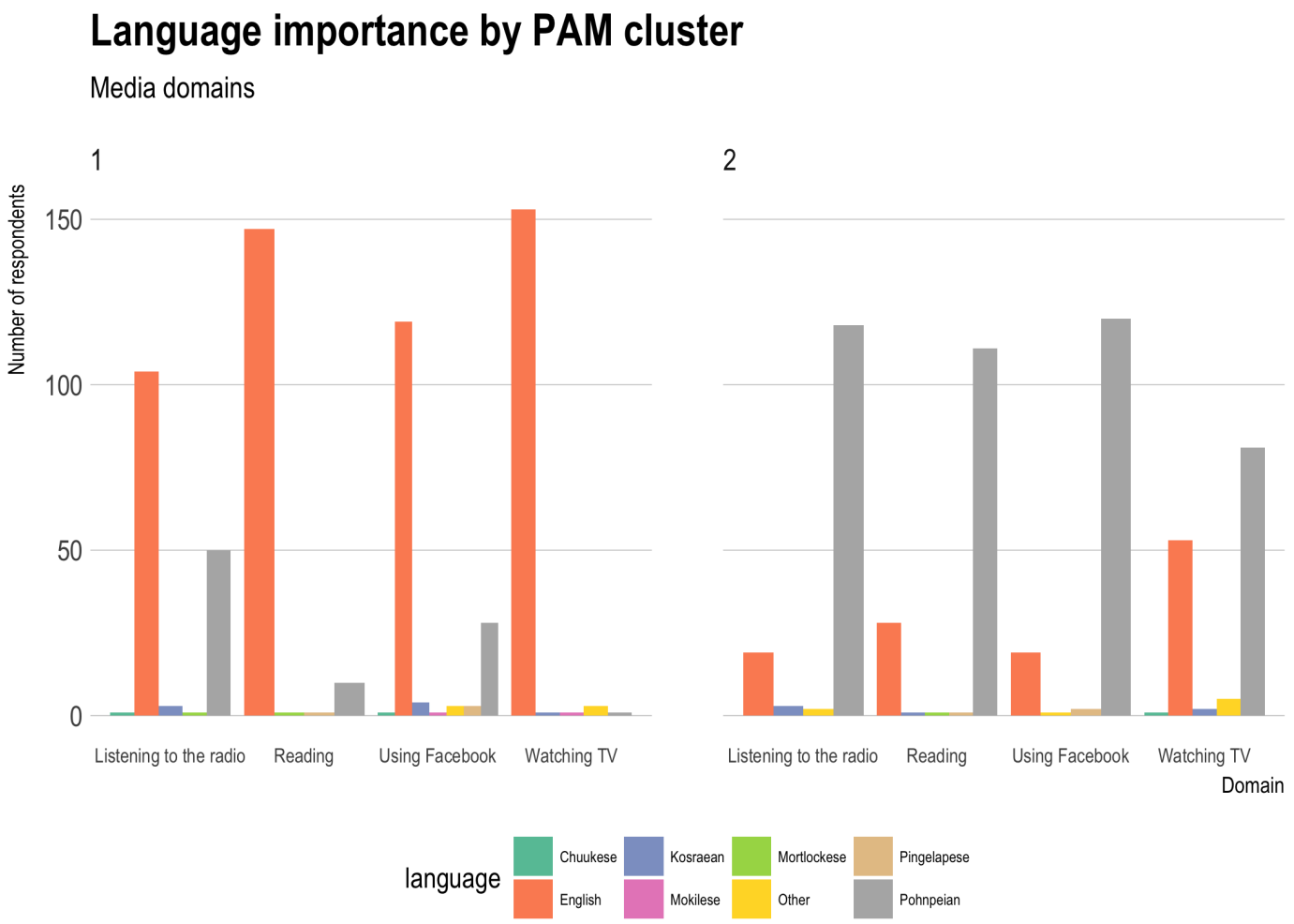
PAM General domains



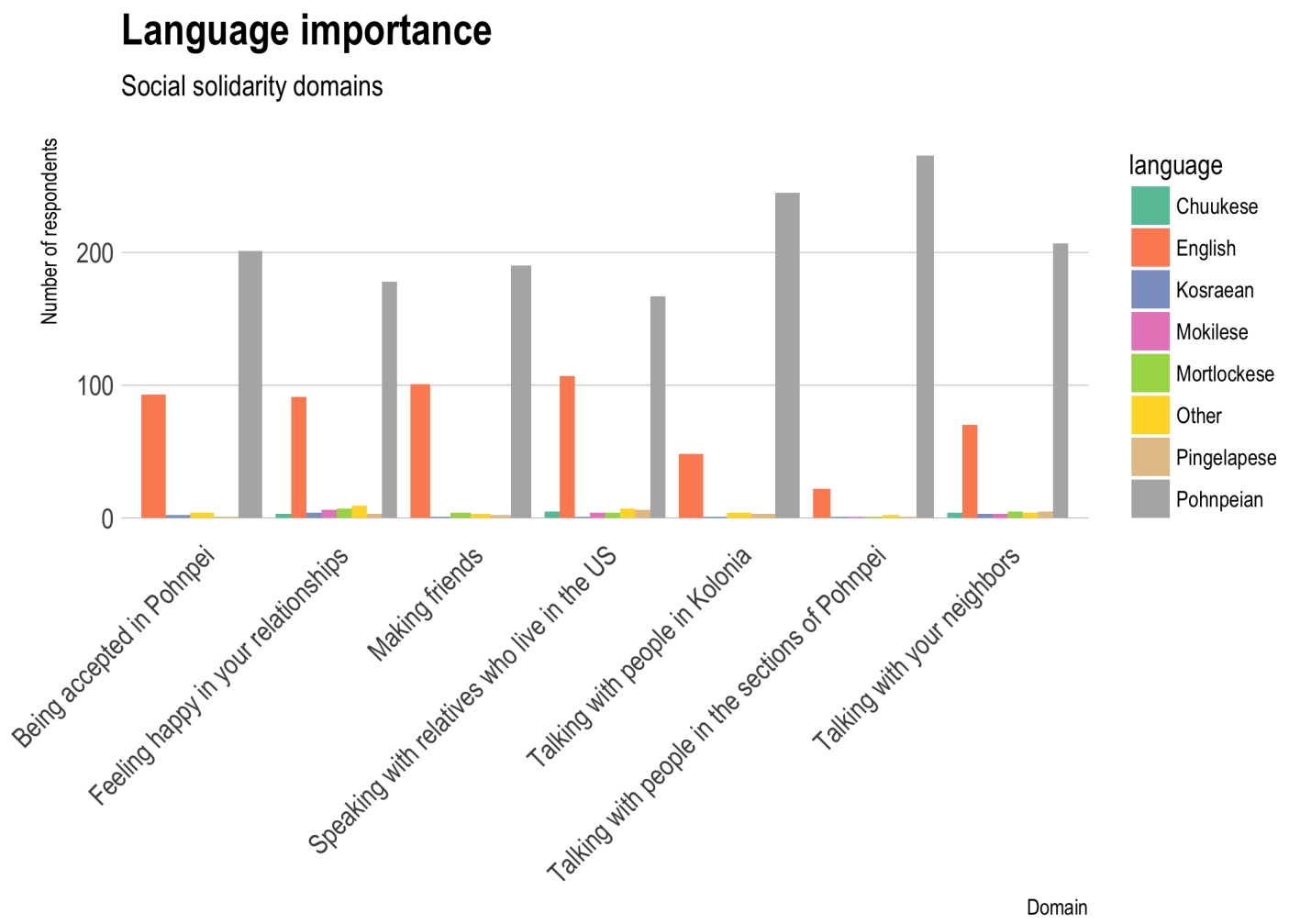
Media domains (original)



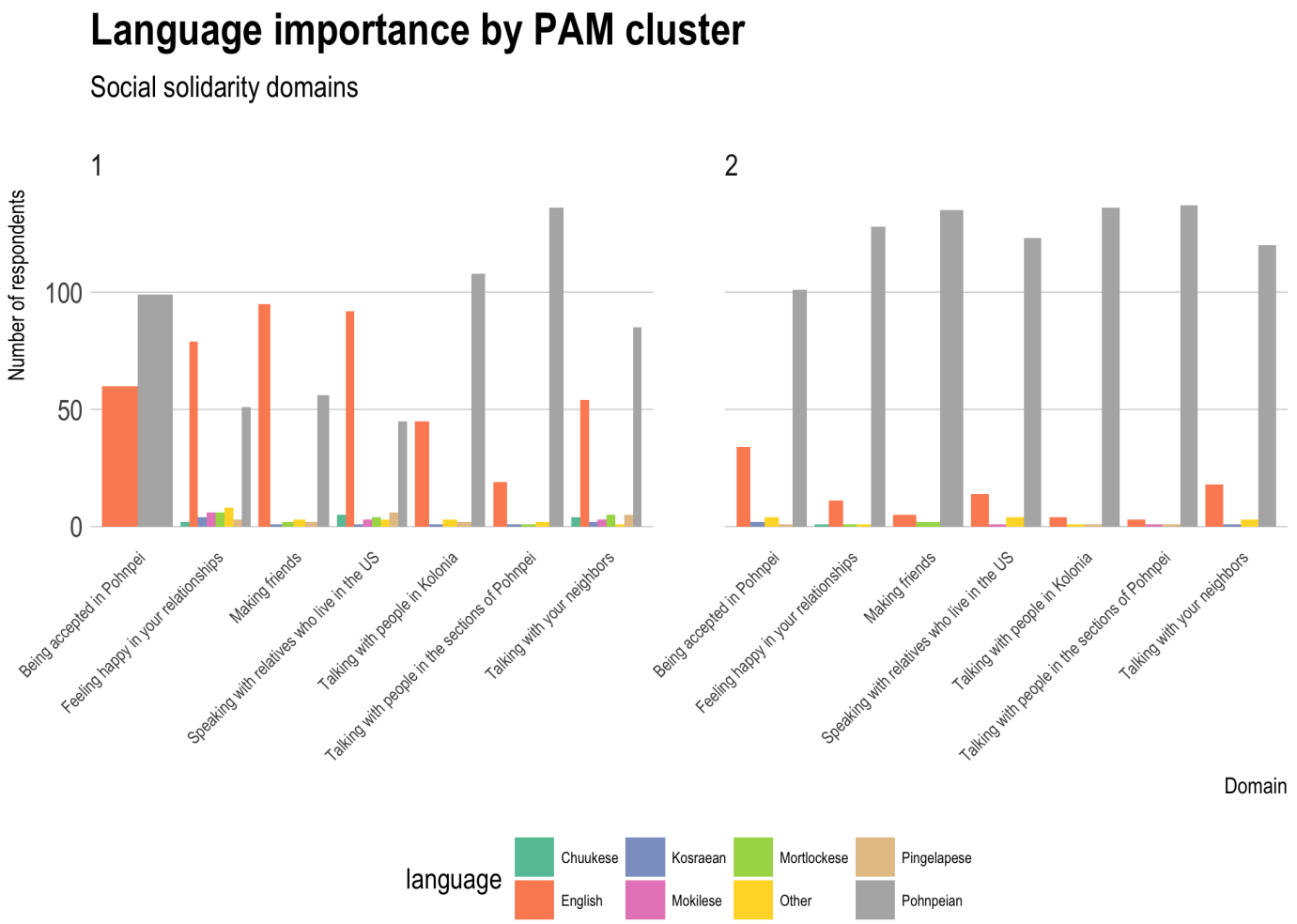
PAM Media domains



Social solidarity domains (original)



PAM Social solidarity domains



Benefits of MDS and PAM

- The data and analysis are not biased or limited by a priori groups
- Shows groups and patterns that were hidden in the data and not shown by regression analysis
- In this example shows two groups: one who values English over Pohnpeian and the other who values Pohnpeian over English
- Also shows that even the English valuing group preferred Pohnpeian for some social solidarity domains (not limited to a single trend)
- The resulting clusters can be analyzed post hoc to see if they correspond to specifc

Conclusions

- Language attitudes research (and linguistics research in general) is about finding the categories and ideas the people actually use
- One must be aware of the assumptions and limitations of the methods used, especially quantitative methods
- Using MDS and clustering analyses is one way to quantitatively allow local groups and categories to emerge from the data
- It also helps paint a picture of the heteroglossia inherent in every community

Kalahngan!

Questions?

References

Bakhtin, Mikhail M. 1981. *The dialogic imagination: Four essays*. Austin: University of Texas Press.

Potter, Jonathan & Margaret Wetherell. 1987. *Discourse and social psychology: Beyond attitudes and behaviour*. London: Sage Publication.

Rousseeuw, Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20. 53–65. doi:10.1016/0377-0427(87)90125-7.