

TEXT MINING ANALYSIS OF PERCEPTION IN ARCHAEOLOGICAL LANDSCAPES: THE CASE OF STONEHENGE

Haley Anne Schwartz
hschwartz@ub.edu
Universitat de Barcelona
Barcelona, Spain

When it comes to analyzing the archaeological record, surviving heritage, and landscapes of the past – how often does this include analyzing those discussing the past ?

Background

- Methodology
 - *Textual Analysis: Topic Modelling*
 - *Open-Source software: SQLite and R*
 - *Applying GIS technologies to Textual Sources (Murrieta-Flores; Gregory)*
- Theoretical lens
 - *Landscape Archaeology (Llobera)*
 - *Sense of Place (Tuan)*
 - *Transformation of Symbol (Kolodziejaska)*

Case Study



- Stonehenge
 - *Globally recognized archaeological and heritage site*
 - *Shifting role in local and global landscape*

Antiquity

- Archaeological Academic Journal
 - *Based in the UK*
- Single lens of academic writing pertaining to Stonehenge



<https://antiquity.ac.uk/>

Topic Modelling

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for a function `createCorpus` that processes a list of papers into a corpus, removing special characters and stopwords, and applying stemming and language detection.
- Environment:** Lists functions defined in the script, including `createCorpus`, `createDTM`, `getPresentIdx`, `leyColors`, `plotMetrics`, and `removeEmptyRows`.
- Console:** Shows the execution of `exampleTermData <- terms(ppLDA, 10)` and `exampleTermData[, 1:5]`, resulting in a matrix of terms for 10 topics across 5 dimensions.
- Plots:** Two line plots showing metrics for 4 to 20 topics. The top plot, labeled 'minimize', shows a decreasing trend in a metric (black circles) as the number of topics increases. The bottom plot, labeled 'maximize', shows an increasing trend in a metric (black triangles) as the number of topics increases.

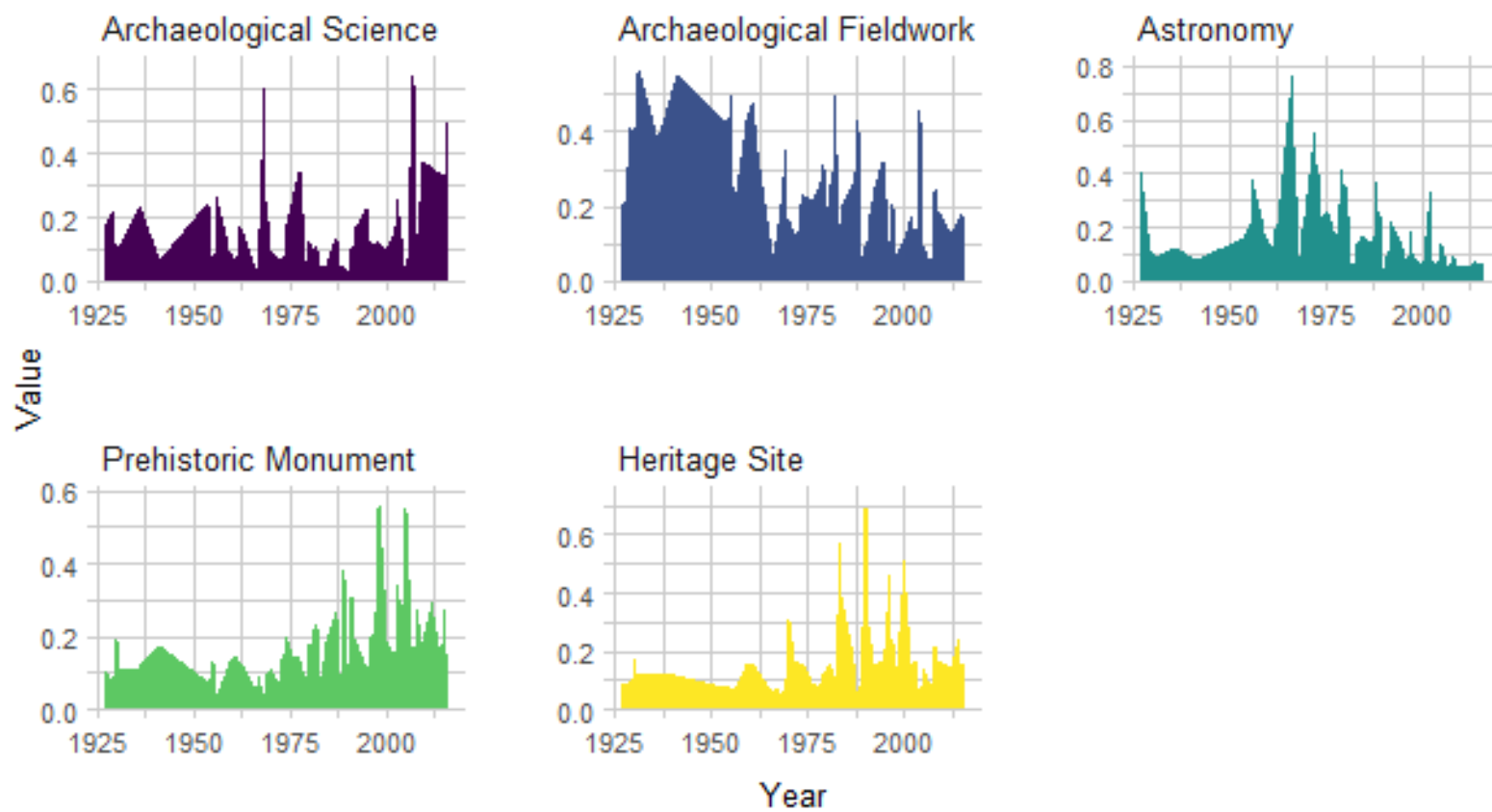
```
36
37 createCorpus <- function( papers )
38 {
39   # final list is custom words + standard english words
40   stopWords <- tm::stopwords("en")
41   stopWords <- c(stopWords, wordsToIgnore)
42
43   corpus <- Corpus(DataframeSource(papers))
44   processedCorpus <- tm_map(corpus, content_transformer(toLower))
45   removeSpecialChars <- function(x) gsub("[^a-zA-Z0-9 ]", "", x)
46   processedCorpus <- tm_map(processedCorpus, removeSpecialChars)
47   processedCorpus <- tm_map(processedCorpus, removeNumbers)
48   processedCorpus <- tm_map(processedCorpus, stemDocument, langua
49   processedCorpus <- tm_map(processedCorpus, removeWords, stopWor
50
51
```

Topic	Term 1	Term 2	Term 3	Term 4	Term 5
[1,]	"blueston"	"origin"	"point"	"neolith"	"heritag"
[2,]	"radiocarbon"	"prese"	"moon"	"survey"	"aveburi"
[3,]	"monument"	"bronz"	"observ"	"societi"	"visitor"
[4,]	"phase"	"museum"	"sun"	"ancestor"	"monument"
[5,]	"bone"	"plate"	"direct"	"monument"	"public"
[6,]	"pit"	"megalith"	"rise"	"bronz"	"landscap"
[7,]	"sarsen"	"photograph"	"align"	"field"	"local"
[8,]	"burial"	"similar"	"horizon"	"social"	"develop"
[9,]	"chalk"	"object"	"astronom"	"interpret"	"london"
[10,]	"aubrey"	"ancient"	"angl"	"landscap"	"interest"

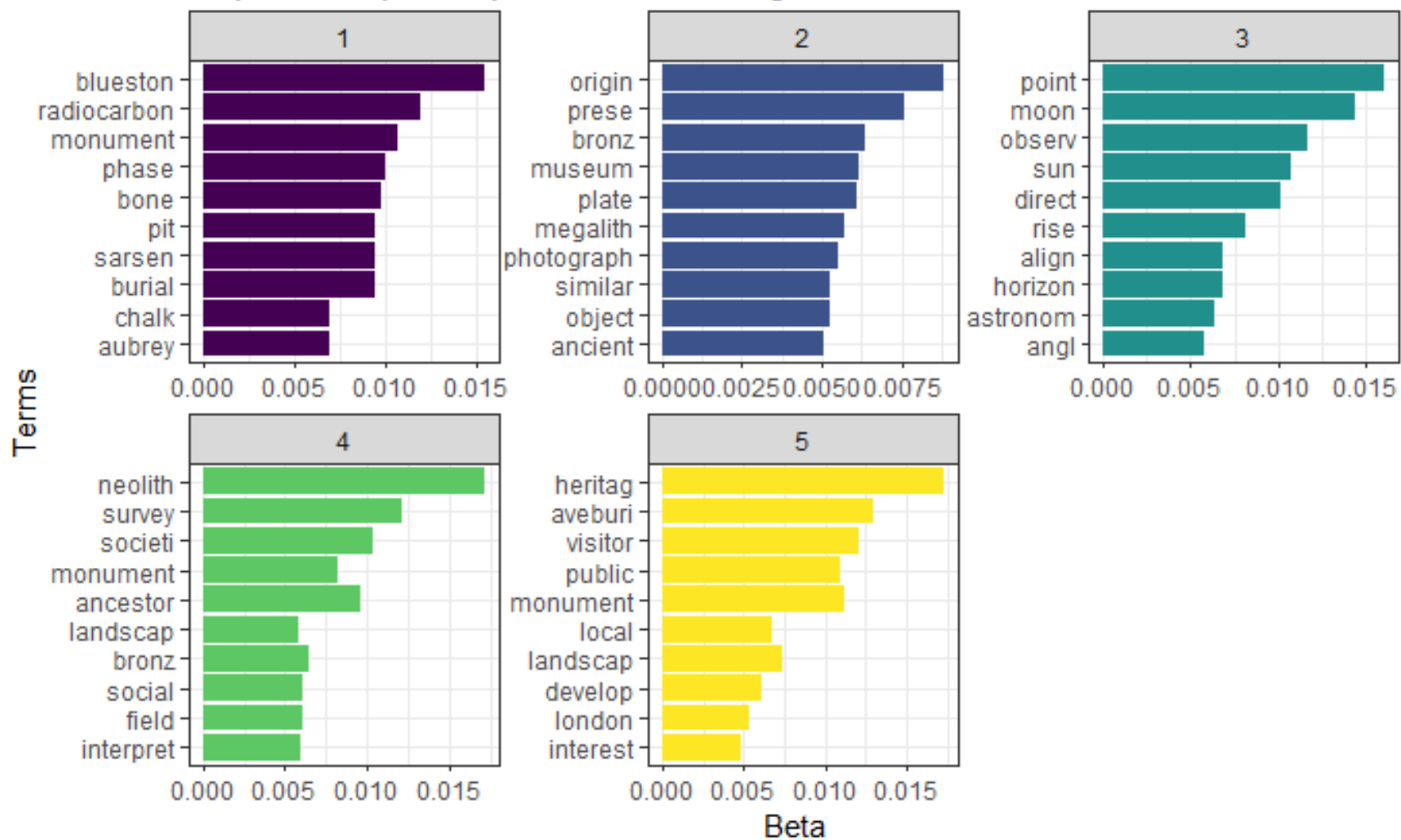
metrics:

- CaoJuan2009
- ▲ Deveaud2014

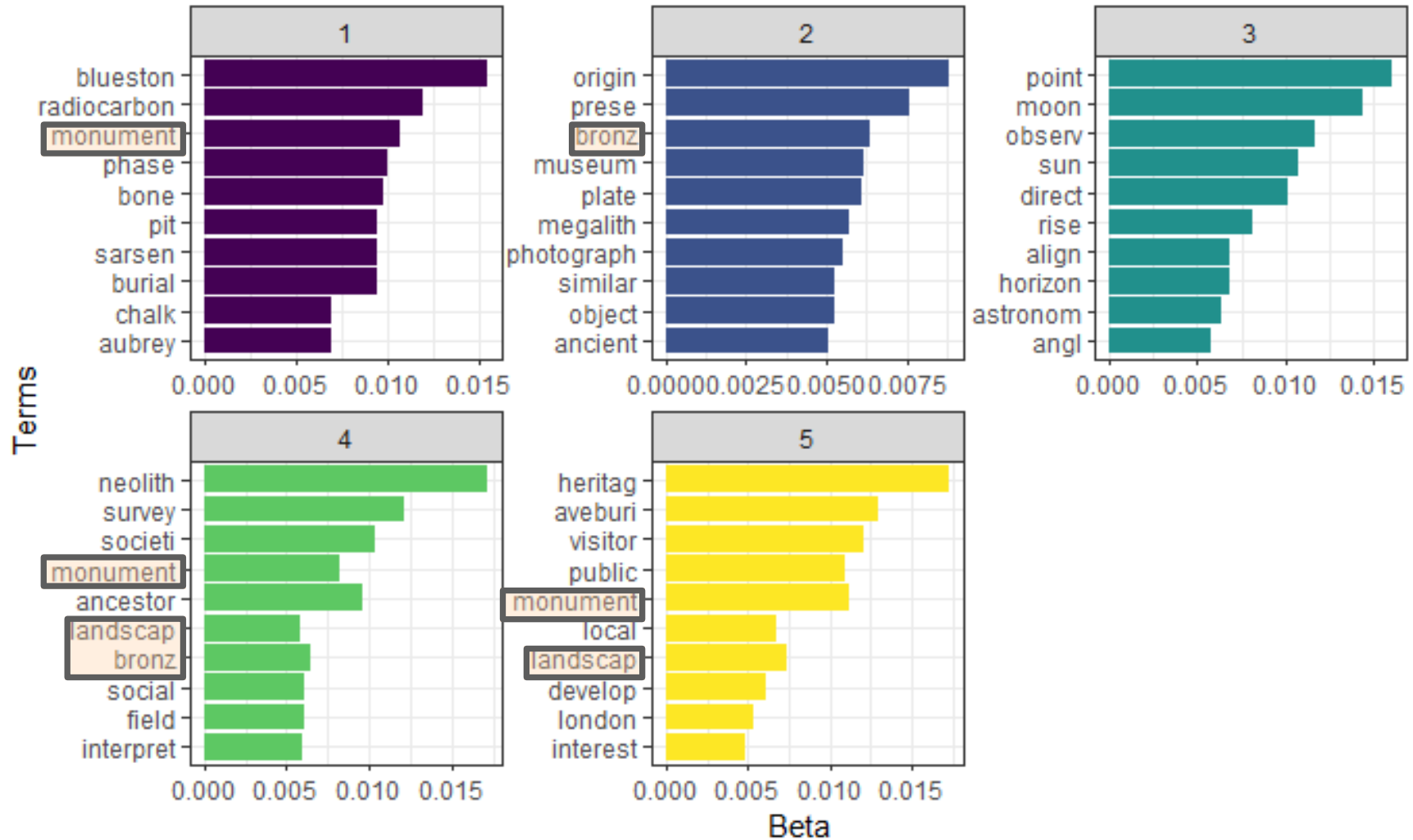
Topic Model of Stonehenge-Related Texts



Top Terms per Topics of Stonehenge-Related Texts



Top Terms per Topics of Stonehenge-Related Texts



Future Work

- Utilizing a Comparative Framework in other case studies
- Additional textual sources other than academic texts
- Additional types of heritage: tangible vs. intangible
- Geotextual Analysis: potential use of GTA

ANY QUESTIONS?

Thank you

Haley Anne Schwartz
hschwartz@ub.edu

References:

Murrieta-Flores, Patricia, Baron, A., Gregory, I., Hardie, A., & Rayson, P. (2015). Automatically Analyzing Large Texts in a GIS Environment: The Registrar General's Reports and Cholera in the 19th Century. (Report). 19(2), 296.

Gregory, I. (2014). Challenges and Opportunities for Digital History. *Frontiers in Digital Humanities*, 1. <https://doi.org/10.3389/fdigh.2014.00001>

Llobera, M. (2001). Building Past Landscape Perception With GIS: Understanding Topographic Prominence. *Journal of Archaeological Science*, 28(9), 1005–1014. <https://doi.org/10.1006/jasc.2001.0720>

Tuan, Y.-F. (1991). Language and the Making of Place: A Narrative-Descriptive Approach. *Annals of the Association of American Geographers*, 81(4), 684–696. <https://doi.org/10.1111/j.1467-8306.1991.tb01715.x>

Kolodziejska, M. (2013) Symbol of the Cross in Popular Culture: an Analysis of the Use and [1] Transformation of the Symbol in 'Machina Magazine'. *Polish sociological review*. (182), 209–222.