Multivariate Humanities Data sets

Pieter M. Kroonenberg *https://www.springer.com/gp/book/9783030691493*

Faculty of Social and Behavioural Sciences, Leiden University

# Introduction

The data sets available on the site of the book [Multivariate Humanities](https://www.springer.com/gp/book/9783030691493) are supplied as they are and may be freely used, provided proper acknowledgements are given to the book and the original publications mentioned therein. There is no guarantee on the data, either from the author or from the suppliers of the data. The latter have all given permission to have their data sets digitally published here.

The data sets are only available in the standard SPSS format. However, there are various other software programs that can read such files, especially common statistical software such as Stata, SAS, MS Excel, and R: The appropriate information on how to do this can be found on the internet. Alternatively, the data can be copy-pasted from the spreadsheet-like format.

Details about each data set are contained in the relevant book chapters and will not be repeated *in extenso* in this document. If desired, the relevant chapters can be purchased separately via the [website of the book](https://www.springer.com/gp/book/9783030691493).

# Book chapters with available data sets

## Theology & Bible studies

05. Similarity data: Bible translations (supplier: Bleemer)

06. Stylometry: Authorship of the Pauline Epistles (compiled: PMK)

## History & Archeology

07. Economic history: Agriculture development on Java (compiled: Boomgaard, PMK)

08. Seriation: Graves in the Münsingen-Rain burial site (source: Kendall)

## Arts

09. Complex response data: Evaluating Marian art (supplier: Polzella)

10. Rating scales: Craquelure and pictorial stylometry (supplier: Bucklow)

11. Pictorial similarity: Rock art images across the world (supplier: Barry)

12. Questionnaires: Public views on deaccessioning (supplier: Piazzai;. Vecco)

## Linguistics

13. Stylometry: The Royal Book of Oz: Baum or Thompson? (compiled: PMK)

14. Linguistics: Accentual prose rhythm in mediæval Latin (source: Janson, PMK)

15. Linguistics: Chronology of Plato’s works (source: Kalusha, Brandwood)

16. Binary judgements: Reading preferences (supplier: Vermunt)

## Music

17. Music appreciation: The Chopin Preludes (supplier: Murakami, PMK;   
also available from // http://three-mode.leidenuniv.nl/)

18. Musical stylometry: Characterisation of music (supplier: Herremans)

# Theology & Bible studies

## 05. Similarity data: Bible translations

* ***Abstract***:
  + *Main topics.* Many Bible translations from different times are based on the same *Masoretic text.* How alike are these translations, and how many differences have emerged over time?
  + *Data.* The data for this study were collected by Zachary Bleemer from British, American, and German translations. First, Bible verses were matched, and subsequently parallel words were counted.
  + *Research questions*. How can the similarities between texts based on the word counts be analysed and graphed, so that conclusions can be drawn with respect to influence between the various translations?
* ***File*:** 05\_Bleemer\_BibleTranslations.sav
* ***Reference***: Bleemer, Z. (2016). *Old testament, new tricks: Using Biblical translation to examine word sense and popular belief.* (Please contact Dr. [Bleemer](mailto:bleemer@berkeley.edu) for a copy)

## 06. Stylometry: Authorship of the Pauline Epistles

* **Abstract**
  + *Main topic:.* When the authorship of a text is in doubt, statistics can assist in shedding light on this question.
  + *Data:* Frequently occurring words in the Pauline Epistles are the base for an investigation into the authorship of the Epistles.
  + *Research questions:* Did Paul write all Epistles assumed to be his?
* ***File****:* 06\_Morton\_St\_Paul.sav
* ***Reference****:* Morton, A. Q. (1965). The authorship of Greek prose. *Journal of the Royal Statistical Society* (*Series A)*, *128*, 169—233; doi=10.2307/2344178.

## 07. Economic history: Agriculture development on Java

* **Abstract**
  + *Main topic:.* The analysis of historical data measured at different points in time and at several locations. One of the issues is how to analyse such data, given that data from the same variables and the same locations are not always available over time, and that some variables have different measurement types.
  + *Data:* Historical data about the state of agriculture on Java in 1880 and 1920.
  + *Research questions:* How can the agricultural landscape in Java be made visible? What similarities and differences can be observed between the four Java regions and their administrative units, the residencies? What changes can be observed between the two time points?
* ***File****: 07\_Boomgaard\_AgriculturalHistoryJava.sav  
  Data courtesy Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS).*
* ***Reference****:* Boomgaard, P. and Kroonenberg, P. M. (2017). Rice, sugar, and livestock in Java, 1820-1940: Geertz's agricultural involution. In Bray, F., Coclanis, P. A. , Fields-Black, E. L & Schäfer, D..(Eds.), *Rice: Global networks and new histories*. New York, NY, USA: Cambridge University Press: pp. 56—83; isbn=978-1-107-04439-5.

## 08. Seriation: Graves in the Münsingen-Rain burial site

* **Abstract**
  + *Main topic:.* The burial site of Münsingen-Rain, Switzerland is situated on a ridge and contains a large number of graves, many of which contain artefacts. The graves have been dated by their distance from the village by Hodson. Is it also possible to statistically validate this dating from the artefacts in the graves?
  + *Data:* An occurrence table of graves by artefacts consisting of ones (artefact present in the grave) and zeroes (artefact absent from the grave).
  + *Research questions:* Is it possible to validate the archeological chronology of the graves by applying statistical procedures on the table of occurrences?
* ***File****:08\_Hodson\_Munsingen.sav*
* ***Reference****:* Kendall, D. G. (1970). A mathematical approach to seriation, *Philosophical Transactions of the Royal Society of London (Series A*), *269*, 125—135; doi=10.1098/rsta.1970.0091.

# Arts

## 09. Complex response data: Evaluating Marian art (supplier: Polzella)

* **Abstract**
  + *Main topic:* Four categories of paintings of the Virgin Mary in particular two different types of content and executed in two different styles, were evaluated by student judges.
  + *Data:* Each of 89 students judged 24 paintings on 10 bipolar scales. The words in the scales were descriptive, evaluative, or emotional
  + *Research questions:* The central question was whether the judgements were influenced by *Style* (pre-Renaissance and post-Renaissance) and *Type of Content* (event and devotional). An additional question was whether students' Age and Gender affected their judgements of the paintings. Another question was the structure of the scales, in particular whether the scales functioned in the same way for all four painting categories.
* ***File****:09\_Polzella\_MarianArt.sav*
* ***Reference****:* Polzella, J. P., Roten, J. G., & Parker, C. W. (1998). Images of Mary: Effects of style and content on reactions to Marian art.,10th Annual Convention of the American Psychological Society, Washington, DC, USA; url=https://udayton.edu/imri/mary/u/undergraduates-psychological-reactions-to-marian-art.php

## 10. Rating scales: Craquelure and pictorial stylometry (supplier: Bucklow)

* **Abstract**
  + *Main topic:.* Paintings acquire cracks over time, but the way in which they do varies with the surface on which they were painted, and the painting materials used. Cracks in paintings from different traditions in Europe were characterised by expert judges.
  + *Data:* Small isolated segments of 40 paintings from four different areas in Europe were photographed and used for the judgements. No information about the paintings was provided. The 50cm2 parts were judged on seven characteristics by the twenty-seven judges, creating a 40 x 7 x 27 data block.
  + *Research questions:* Can art-historical categories from different countries be distinguished via the judges' subjective scores? Do the different materials on which the paintings were made influence the cracks in such a way that their differences can be recovered from the judgements?
* ***File****: 10\_Bucklow\_CraquelureExperts.sav*
* ***Reference****:* Bucklow, S.(1998). A stylometric analysis of craquelure. *Computers and the Humanities*, *31*, 503-521.

## 11. Pictorial similarity: Rock art images across the world (supplier: Barry)

* **Abstract**
  + *Main topic:.* The aim of this chapter is to show how rock art images all over the world can be analysed so that (dis)similarities of images from different regions can be assessed.
  + *Data:* A large collection of human rock art images were scored on a number of binary characteristics pertaining to both the person pictured and accompanying attributes, such as clothes, objects, etc. A large part of the images described are of Australian origin.
  + *Research questions:* Regional differences were the centre of attention. Images were compared from specifically: (1) Algeria and the Kimberley (Australia) – data not available, (2) Zimbabwe, India, and Algeria, (3) three regions of Northern Australia: Arnhem Land, the Kimberley, and Pilbara.
* ***Files****:*
  + *11\_Barry\_RockArt\_NorthAustralia.sav*
  + *11\_Barry\_RockArt\_SampledDataset\_ZIA.sav*
* ***Reference****:* Barry, M. & White, J. P (2004). ‘Exotic Bradshaws' or Australian `Gwion': (An archaeological test). *Australian Aboriginal Studies, 1*, 37-44; https://protect-au.mimecast.com/s/M4HrCGv0oyCGWgNAh77OOO?domain=questia.com.

## 12. Questionnaires: Public views on deaccessioning (supplier: Piazzai; Vecco,)

* **Abstract**
  + *Main topic:.* In how far do Italian museum visitors appreciate their museums selling part of their stock (deaccessioning)? Of primary interest was whether people think this should be done at all, what objects are eligible, and who was allowed to buy them.
  + *Data:* 295 Italians visiting open-air museums in Rome filled out the Italian questionnaire of 22 questions on such topics as which objects could be sold, under which conditions they could be sold, who should be allowed to buy them, and what could be done with the proceeds from the sales.
  + *Research questions:* Next to an interest in opinions about deaccessioning, a major question was whether the structure of the questionnaire was recognisable in the answers of the respondents. In addition, we would like to know whether within the subdomains the items form a single scale.
* ***File****: 12\_Vecco\_Deaccession.sav*
* ***Reference****:* Vecco, M. ***,***Srakar, A., & Piazzai, M. (2017).Visitor attitudes toward deaccessioning in Italian public museums: An econometric analysis, *Poetics*, *63*, 33-46; doi= 10.1016/j.poetic.2017.05.001

# Linguistics

## 13. Stylometry: The Royal Book of Oz: Baum or Thompson? (extracted: PMK)

* **Abstract**
  + *Main topic:.* Linguistic stylometry is concerned with the quantitative analysis of language and style. Authors have idiosyncratic styles in the way they build their sentences, which often makes it possible to determine with some probability who has written a particular text. In this chapter the focus is on one particular book from the Wizard of Oz series
  + *Data:* Occurrences of 50 frequently used function words were recorded for 21 Wizard of Oz books. The first 14 were written by L. Frank Baum, the last 6 by Ruth Plumly Thompson, and the authorship of one (the 15th in the series) was contested at the time
  + *Research questions:* The central question is who of the two authors wrote the 15th book - *The Royal Book of Oz*, which appeared two years after Baum's death
* ***Files****:* 
  + *13\_Baum\_WizardOz\_Functionwords.sav*
  + *13\_Baum\_WizardOz\_Transposed.sav*
* ***Reference****:* Binongo, J. N. G. (2003). Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution. *Chance, 16,* 9-17; doi=10.1080/09332480.2003.10554843.

## 14. Linguistics: Accentual prose rhythm in mediæval Latin (source: Janson)

* **Abstract**
  + *Main topic:.* In this chapter the focus is on accentual rhythms at the end of prose sentences. In mediaeval Latin prose rhythm is a style element used differently by

different authors in different countries and periods.

* + *Data: Cursus* frequencies from various mediaeval authors writing in Latin.
  + *Research questions:* What kinds of accentual patterns were used by and popular with various authors? Are these patterns the same through time and place?
* ***Files****:* 
  + *14-Janson\_BasicFrequencies.sav*
  + *14\_Janson\_AuthorFrequencies.sav*
* ***Reference****:* Janson, T. (1975). *Prose rhythm in medieval Latin from the 9th to the 13th century*. Stockholm, Sweden: Almqvist & Wiksell International; issn=0491-2764;20

## 15. Linguistics: Chronology of Plato’s works (source: Kalusha, Brandwood)

* **Abstract**
  + *Main topic:.* In Greek and Latin, sentence endings often carry a special meaning, especially through their rhythm. It has been said that changing use of specific sentence endings over time can be used to put an author's work in chronological order.
  + *Data:* Plato's sentence endings in 45 of his texts collected by Kaluscha in 1904
  + *Research questions:* Is it possible to use the rhythm of sentence endings for the seriation of Plato's works?
* ***Files****: 15\_Kalusha\_Plato.sav.*
* ***Reference****:* Brandwood, L. (1990). *The chronology of Plato's dialogues.* Cambridge, UK: Cambridge University Press; isbn=978-0521106559.

## 16. Binary judgements: Reading preferences (supplier: Vermunt)

* **Abstract**
  + *Main topic:.* People have varying preferences in what they read. The aim of a Dutch study was to investigate what types of publications people preferred to read, and which personal factors played a role in their choice.
  + *Data:* The data consisted of binary scores indicating whether or not the respondents usually read each of six types of printed material: *Quality* newspapers versus *Popular newspapers* (including regional newspapers), *Literary books* versus *Non-literary books*, and  *Opinion magazines* versus *Family magazines*. In addition, five background variables were available to characterise the different groups of readers.
  + *Research questions:* How many and which groups of readers could be distinguished, and which background variables were relevant to distinguish between the various groups?
* ***File****: 16\_Vermunt\_Readers.sav*
* ***Reference****:* Van Rees, K., Vermunt, J., & Verboord, M. (1999). Cultural classifications under discussion. Latent class analysis of highbrow and lowbrow reading. *Poetics, 26,* 349—365; doi=10.1016/S0304-422X(99)00019-4.

# Music

## 17. Music appreciation: The Chopin Preludes (supplier: Murakami)

* **Abstract**
  + *Main topic:.* Musical pieces may be characterised via a set of semantic differential scales. It proves possible to link evaluative judgements to technical properties of the music.
  + *Data:* Japanese students judged Chopin's preludes on a specially designed set of semantic differential scales, and in doing so produced a three-way dataset of Scales × Preludes × Students.
  + *Research questions:*
* ***File****:17\_Murakami\_ChopinPreludes.sav*
* ***Reference****:* Murakami, T.& Kroonenberg, P. M. (2003). Three-mode models and individual differences in semantic differential data. *Multivariate Behavioral Research, 38,* 247—283; doi=10.1207/S15327906MBR3802\_5.

## 18. Musical stylometry: Characterisation of music (supplier: Herremans)

* **Abstract**
  + *Main topic:.* Midi files contain digitised music and are widely available. This opens the possibility to digitally examine similarities between composers and between different types of compositions.
  + *Data:* Musical features scored from midi files of compositions by Bach, Beethoven, and Haydn.
  + *Research questions:* Is it possible to use these features to characterise, identify, and discriminate between the compositions of the three composers? Can the features be considered musical `function words' or do they also indicate genre?
* ***File****: 18\_Herremans\_MusicStyles.sav*
* ***Reference****:* Herremans, D., Martens, D. & Sörensen, K. (2016). Composer classification models for music-theory building. In D. Meredith (Ed.), *Computational music analyses* (pp. 369--392)*.* Cham, Switzerland: Springer; doi=(10.1007/978-3-319-25931-4-14).