

Master Digital Humanities Data Science

Filippo Bergamasco

14/01/2017

Mississippian Pottery Analysis

Suppose to be a famous archaeologist willing to study long-distance trade routes in the southeastern United States. We will follow the work of [1] pretending to be the first to perform this kind of analysis trying to provide new interesting insight for the scientific community interested in this field. As suggested in the paper, there are historical evidence that trading pots was a widespread practice among the Indians. So, tracking the ceramic exchange based on the occasional discovery of non-local shreds can hopefully give useful informations on how these trade networks evolved over time.

After a tedious fieldwork period, we collected many different shreds from different locations of the southeastern United States. Each sample was chemically analyzed so that the percentage of 23 different elements was measured and recorded for each shred sample. Our goal is to find correlations between the chemical composition of each sample and the location in which it has been found so that the analysis of trading-routes samples may provide informations on their original provenance.

The collected data

The data of all the acquired samples can be found in the attached *mississippian_pottery_data.csv* CSV file. Each row represents a shred for which a record of different characteristics have been collected. Specifically, each shred is identified by a unique name, the region of finding (See Figure 1) and, most importantly, the concentration of the following 23 chemical elements:

(As) Arsenic	(La) Lanthanum	(Lu) Luthetium	(Sm) Samarium
(U) Uranium	(Yb) Ytterbium	(Ce) Cerium	(Co) Cobalt
(Cr) Chromium	(Cs) Cesium	(Eu) Europium	(Fe) Iron
(Hf) Hafnium	(Rb) Rubidium	(Sb) Antimony	(Sc) Scandium
(Ta) Tantalum	(Tb) Terbium	(Th) Thorium	(Zn) Zinc
(Ba) Barium	(K) Potassium	(Na) Sodium	

Chemical concentrations have already been corrected to remove the effect of shell dilution as explained in the original paper so you can expect the data being ready without additional pre-processing.

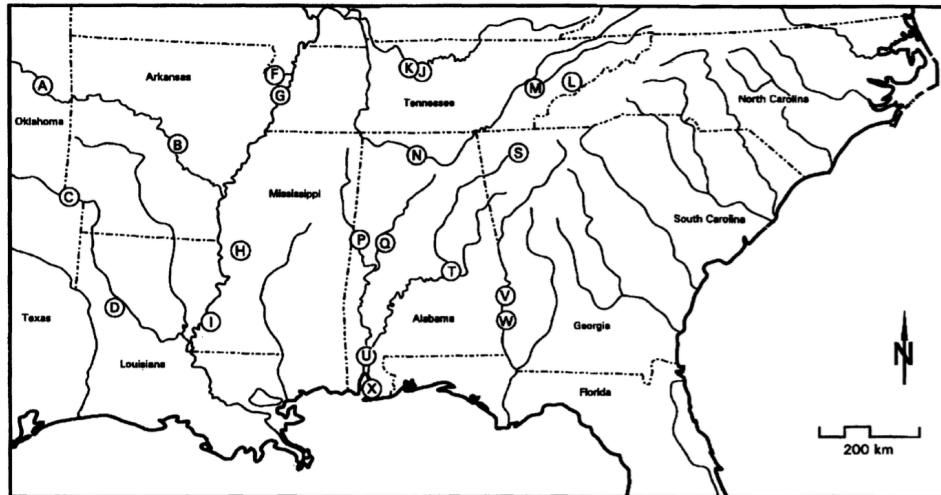


Figure 1. Map of the southeastern United States showing the regions used in the present study. A, Spiro region; B, Little Rock region; C, Great Bend region; D, Natchitoches region; F, Big Lake region; G, Pecan Point region; H, Lower Yazoo region; I, Natchez region; J, Nashville region; K, Lower Harpeth region; L, Sevierville region; M, Tellico region; N, Wheeler Lake region; P, Gainesville Lake region; Q, Black Warrior region; S, Carters Lake region; T, Wetumpka region; U, Mobile Delta region; V, Eufaula region; W, Fort Gaines region; X, Mobile Bay region.

Assignment

Using Matlab, perform the following tasks and answer the related questions in a short essay. Describe all the operations performed and justify each answer in detail. Whenever possible, don't hesitate to include plots and drawings to convince the reader on the validity of your claims.

T1

Load the provided *mississippian_pottery_data.csv* into a Matlab table. How many data samples are present? Are there any incomplete or missing values?

T2

Considering only its chemical composition, each data sample can be represented as a vector in a 23-dimensional space. Compute the first 3 principal components and project each vector in a 3-dimensional Euclidean space. Then, render the projected vectors in a 3D scatter plot. Can you visually identify interesting patterns in the data? Are the feature vectors uniformly distributed in the space or can you observe some clusters? If yes, how many of them?

T3

Perform the same analysis of the task T2 but, this time, considering the first 2 principal components and hence projecting on a 2-dimensional Euclidean space. Can you still identify the same number of clusters?

With respect of T2 we are now using a lower dimensional space. Is this giving us more or less insight on the structure of the data?

T4

Run the K-Means algorithm to extract the number of clusters you previously identified in T2. Evaluate the obtained result with respect of what can be observed visually after the projection on the 3D space. For each cluster, mark on the map the region of finding of each sample belonging to that cluster. How are the element of each cluster distributed on the map? Can we identify some clusters as representative of geographical regions?

T5

Compute the hierarchical linkage of the data (assuming Euclidean distance) and plot the obtained dendrogram. Cut the tree to obtain the same number of clusters you identified before and compare the result with respect to K-Means you computed in T4.

T6 (Optional)

In their original papers, the authors identified 4 different geographical groups: WESTERN, EASTERN, SOUTHERN, NORTHERN (each data sample contains this information in the loaded CSV file). Compare these geographical grouping with the clustering obtained in T4 and discuss differences and similarities.

References

- [1] Vincas P. Steponaitis, M. James Blackman, and Hector Neff. Large-scale patterns in the chemical composition of mississippian pottery. *American Antiquity*, 61(3):555–572, 1996.