



ISAAK

## Classifying vessel shapes

using automated shape extraction and unsupervised classification

---

Martin Hinz<sup>1</sup>, Caroline Heitz<sup>2</sup>

March 2018

1. Institute of Pre- and Protohistoric Archaeology, Kiel University; 2. Institut für Archäologische Wissenschaften, Bern University

motivation

previous approaches

Shape Extraction

Shape Analysis 1: PCA & hclust

Shape Analysis 2: t-sne & hdbscan

Case study 1: Bell Beakers of the Iberian Peninsula

Case study 2: Neolithic Swiss Ceramic

Resume

## **motivation**

---

## shape vs. decoration for typology

# levels of pottery design

Tendencies:

criteria	shape	decoration
functional dependency	strong	weak
dependency on local ressources	stronger	weak
moment of determination	from the beginning	finished product
ad hoc changes	difficult	easy
more reflecting	practised reality	imagined reality
design decision	more collective	more individual
homogeneity	more convergent	more divergent
degrees of freedom	smaller	higher
category of style	more isochrestic	more emblemic
unit of analysis	ideal type	singularity

# Phases of ceramic production

## Design

- imagined and intended
- concept of a shape (incl. attached parts, eg. handles)
- ideal type
- **social bounded options for creativity**

## 'style of action'

- learned, habitual movements, manipulations, routines
- at all stages of the *chaîne opératoire*
- instantiation reflecting the characteristics of the material
- **limited or prescribed options for creativity**

## 'material style'

- physical traces and effects in the material
- Characteristics of the vessel
- finished object
- **material induced alterations of the original design possible**

**using vessel shapes to identify underlying large scaled ‘communities of practice’ in production and consumption**

- more related to covert communication structures, habitus and similar (economic) practises
- ‘better measurable’ from a practical quantitative perspective

## **previous approaches**

---

## using specific locations

- Koch 1998

## 'Holistic' approaches

- Mom 2005
- Chapman et al., 2006
- Keogh et al., 2009

## Shape Extraction

---

## 'by hand': workflow

---



load



clean



enhance contrast



close gaps



fill inside black



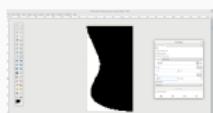
turn bw (bitmap)



rotate



crop



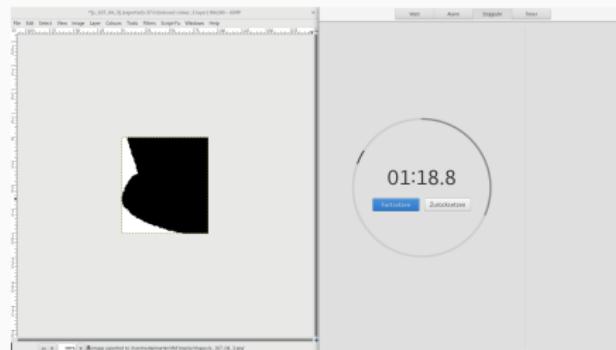
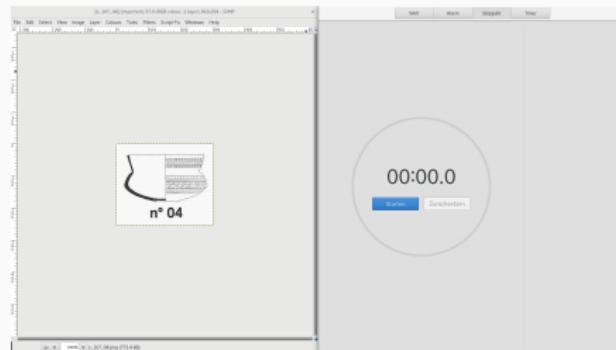
split in half



scale to resolution (100px)

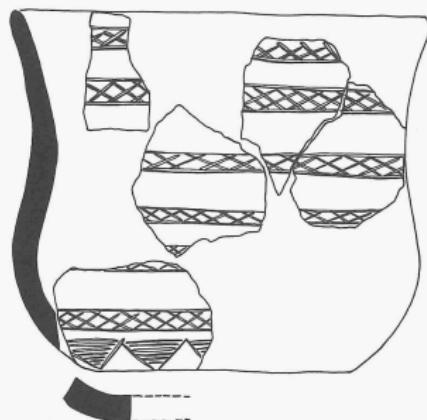
---

# 'by hand': speed



# Adaptive Contour Motivation

- 1:30 min for 800 images -> 20 hours
  - automation makes sense!
- scanned vessel drawings not trivial
  - might have holes (dashed lines)
  - might have additional image components beside the vessels
- no simple image segmentation with background color and floodfill possible



Bell Beaker (Harrison 1977)

## Adaptive Contour: what is it

***Active contour model***, also called ***snakes***, is a framework in computer vision for delineating an object outline from a possibly noisy 2D image. The snakes model is popular in computer vision, and snakes are greatly used in applications like object tracking, ***shape recognition***, segmentation, edge detection and stereo matching. Wikipedia

source: <https://github.com/pmneila/morphsnakes>

# Active contour with Bell Beakers

See shapAAR vignette

<https://github.com/ISAAKiel/shapAAR/blob/master/vignettes/object-extraction.md>

---

	Image preparation	bw, blur
	Image segmentation	active_contour
	Selecting the biggest object	EBImage
	Rectify and crop	get Bounding box, rotate upright, crop
	half and side mean	split in middle, mean left-right

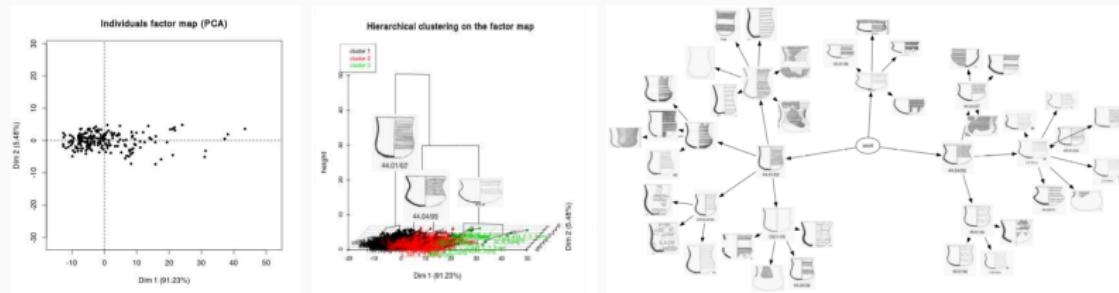
---

## Shape Analysis 1: PCA & hclust

---

# approach/workflow

- using profile distances + additional nominal variables as input
- conduct PCA
- using hclust (euclidean/Ward) to cluster
- package FactoMinR: HCPC with automatical cut (number of clusters according to higher relative loss of inertia)
- for individual clusters, repeat with cluster members as new dataset (usually 3 levels)



## problems

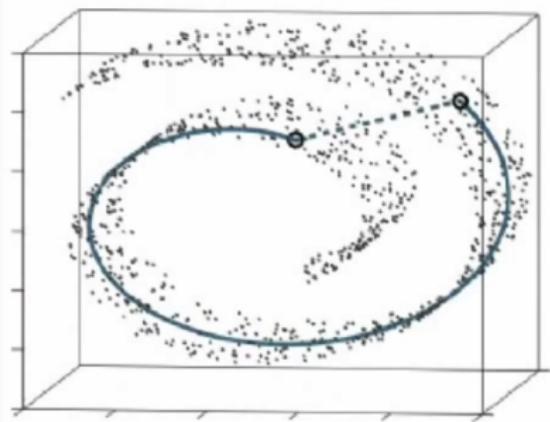
- non-metric (non-euclidean) variables not considered correctly
- all vessels had to belong to one cluster (hand made ceramics)
- global dissimilarity resulted in suboptimal separation

## Shape Analysis 2: t-sne & hdbSCAN

---

# t-sne

## t-Distributed Stochastic Neighbor Embedding

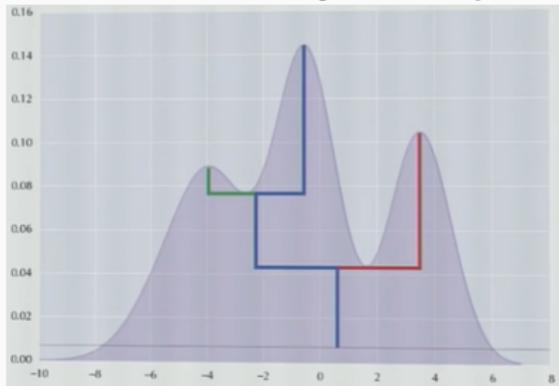


### Challenge

- high dimensional non-linear data distribution
- consider not the global, but the local neighbourhood (contrasting PCA)

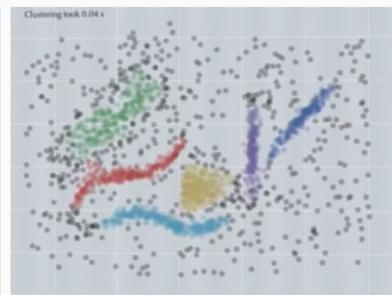
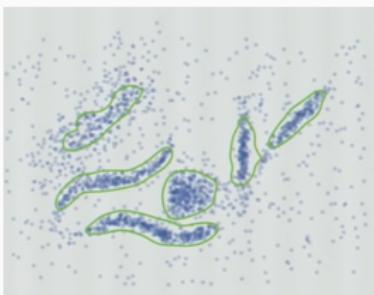
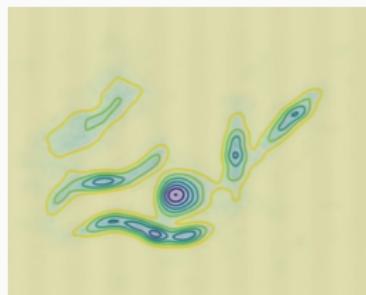
van der Maaten 2008; 2009; 2012; 2014;  
<https://lvdmaaten.github.io/tsne/>  
<https://www.youtube.com/watch?v=RJVL80Gg3IA>

## Hierarchical Density-Based Spatial Clustering of Applications with Noise



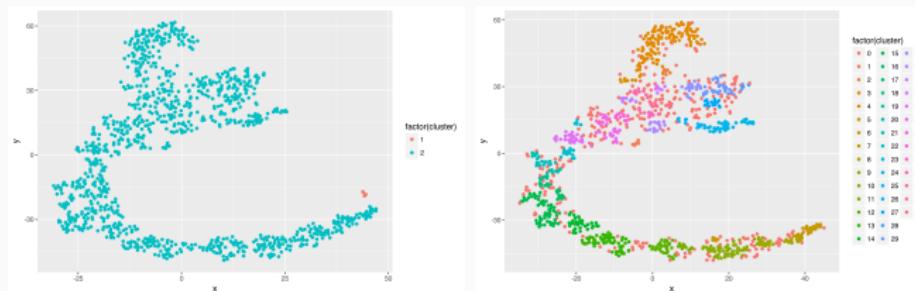
### Benefits:

- separating non-circular clusters
- 'identification' and exclusion of noise (hand made ceramics!)



## approach/workflow

- using profile distances + additional nominal variables as input
- conduct t-sne
- using hdbscan to cluster
- package Rtsne: conduct t-sne
  - `Rtsne(as.matrix(shapes), perplexity=30, dims=2, theta = 0)`
- package dbscan: conduct hdbscan on t-sne result
  - `hdbscan(this_res_sne, minPts = 5)`
- for individual clusters, repeat with cluster members as new dataset (usually only 2 levels)



## stochastic nature of t-sne

### Problem

- t-sne is a stochastic algorithm
  - results from two t-sne runs might differ
  - accordingly clustering might also differ

### solution

- consensus clustering
  - repeat clustering for n times (101)
  - mark co-occurrence of objects in cluster
  - make final cluster solution on majority voting

## **Case study 1: Bell Beakers of the Iberian Peninsula**

---

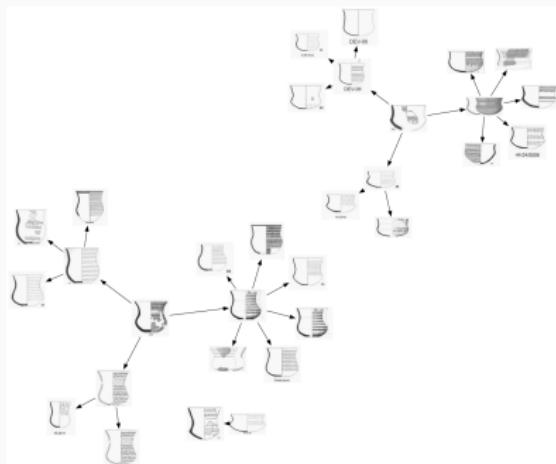
## Dataset

- 213 specimens
- Collected from 5 publications
  - Leisner/Leisner 1956
  - Harrison 1977
  - García Rivero 2008
  - Rojo-Guerra et al. 2005
  - Prieto-martinez/Salanova 2011
- Only ceramics with full (preserved or reconstructed) profiles
- Only beakers and carinated bowls
  - rather low diversity of ceramic shapes



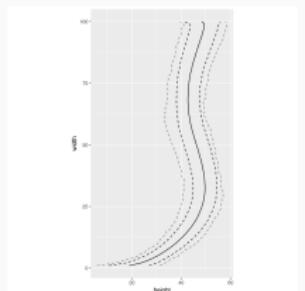
## Results HCPC

- Separation Beakers/Carinated Bowls on first Level
- additionally small group with low belly break

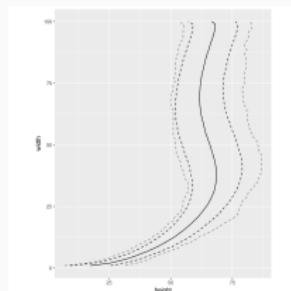


# mean profile per cluster HCPC

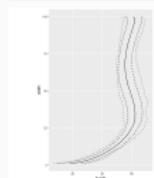
Cluster 1



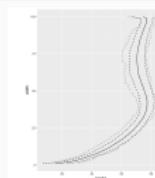
Cluster 2



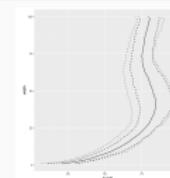
Cluster 2.1



Cluster 2.2



Cluster 2.3

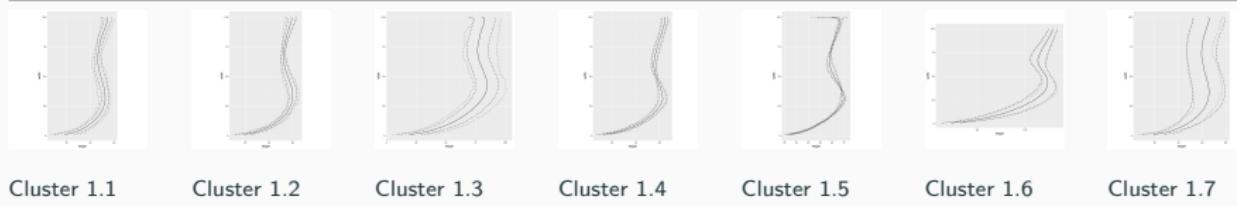
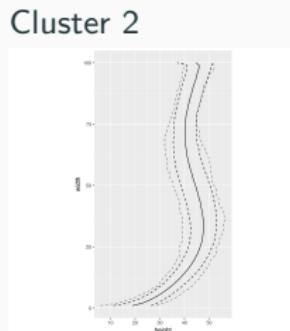
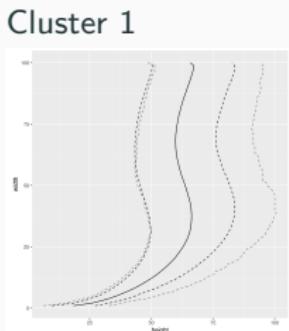


## Results tsne\_hdbscan

- Separation Slim Beakers/Stout Beakers + Carinated Bowls on first Level
- Separation Beakers/Bowls on second level



# mean profile per cluster tsne\_hdbscan



## Resume Case Study 1

with t-sne and hdbscan

- finer separation of classes early
- profile curvature has more impact

but in general at this complexity level both approaches are feasible/comparable

## **Case study 2: Neolithic Swiss Ceramic**

---

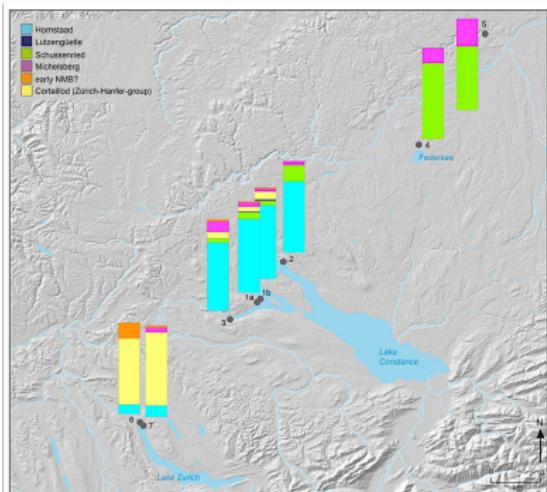
## Background

MET

# Mobilities, Entanglements, Transformations

What can be said about mobility, relationships and transformations in neolithic societies on the basis of ceramics?

- including form, but also nominal characteristics like
    - lugs, handles, eyelet
    - bottom characteristics
    - plastic attachments
    - surface treatment
  - much more diverse ceramic shapes
  - 1045 vessels of different ceramic traditions



## Ceramic spectrum at different Swiss and South German sites, 3950–3900 cal BCE

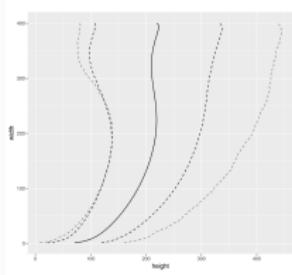
# Results HCPC

- Separation first Level
  - bowls
  - higher forms
  - special forms
- second level higher forms
  - beakers
  - higher and flatter bowls
- not bad

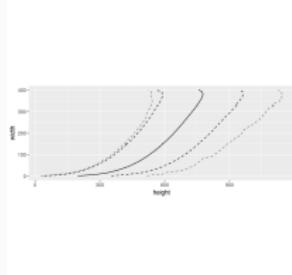


# mean profile per cluster HCPC

Cluster 1



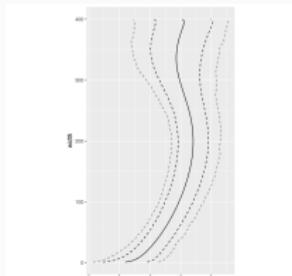
Cluster 2



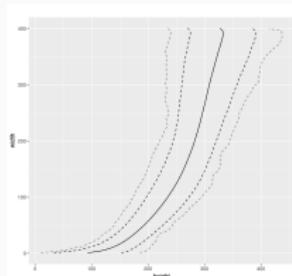
Cluster 3



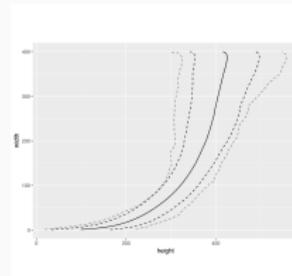
Cluster 1.1



Cluster 1.2



Cluster 1.3

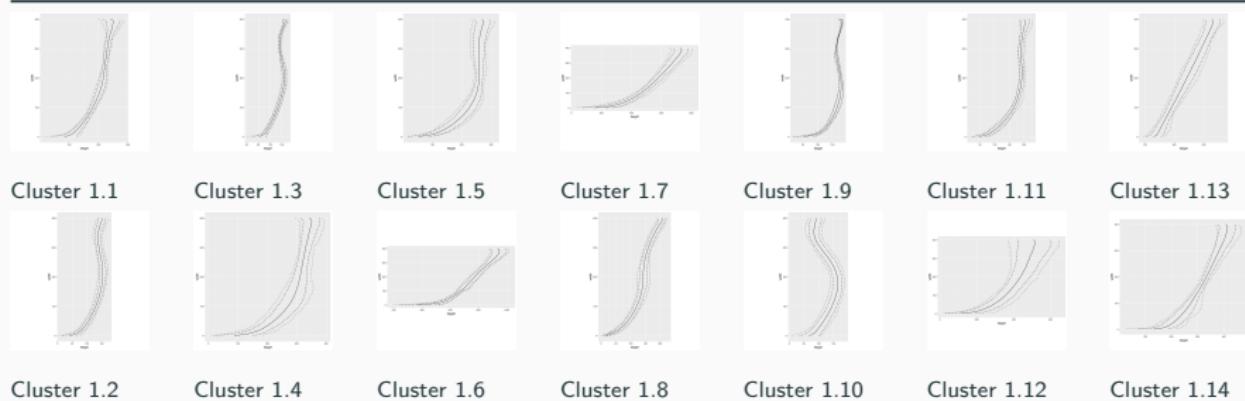
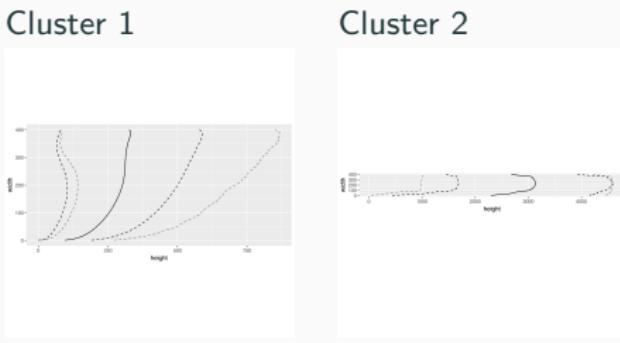


## Results tsne\_hdbscan

- First Level separation dishes / remainder
- on second level an explosion of 33 different classes



mean profile per cluster tsne\_hbscan



## Resume Case Study 2

with t-sne and hdbscan

- much finer separation of classes
- much greater homogeneity within the classes
- Classification clearly more meaningful based on prior archaeological knowledge
  - no mixing up of 'exotic shapes' into one class
  - eg. Michelsberg Beakers separated
    - even division between 'Beutelbecher' ('bag beaker') and 'Tulpenbecher' ('tulip beaker')

## **Resume**

---

## Results from case studies

- unsupervised classification (also as merely explorative tool) can open up new perspectives in archaeological interpretation
- automated object separation has the potential to speed up the process significantly
- in case of low diversity pca/hclust and t-sne/hdbscan perform similar
- in case of high diversity t-sne/hdbscan results in much more adequate classes

# shapAAR as (CRAN) Package

Watch this space: <https://github.com/ISAAKiel/shapAAR>

## shapAAR

[build](#) passing [CRAN](#) [not published](#) [coverage](#) 0%

This is an R package for the extraction, analysis and classification of (not only) archaeological objects derived from scanned images. Especially it aims at the analysis of the shapes/profiles of eg. ceramic vessels or arrow heads.

Currently the extraction is implemented using [active contour](#) to identify and extract the shape informations. You can get an idea about its abilities in a [vignette](#).

### Licence

shapAAR is released under the [GNU General Public Licence, version 3](#). Comments and feedback are welcome, as are code contributions.

### Installation

shapAAR is currently not on [CRAN](#), but you can use [devtools](#) to install the development version. To do so:

```
if(!require('devtools')) install.packages('devtools')
library(devtools)
install_github('ISAAKiel/shapAAR')
```

- object separation from scanned images already usable
- inclusion of classification in progress
- cran release expected this year (2018)

**Thank you**



**Martin Hinz**

[martin.hinz@ufg.uni-kiel.de](mailto:martin.hinz@ufg.uni-kiel.de)



**Caroline Heitz**

[caroline.heitz@iaw.unibe.ch](mailto:caroline.heitz@iaw.unibe.ch)