Import - Lithic analysis Balver Höhle

Lisa Schunk

2021-02-04

# Goal of the script

This script reads the xlsx file (database techno-typological analysis) generated with E4 and formats the data for a statistical analysis.  
The script will:

1. Reads in the original xlsx file
2. Changes and sort the data in order to do stats
3. Saves the data as a new xlsx file and R object

dir\_in <- "analysis/Balve/raw\_data/"  
dir\_out <- "analysis/Balve/derived\_data/"

Raw data must be located in “analysis/Balve/raw\_data/”.  
Formatted data will be saved in “analysis/Balve/derived\_data/”. The knit directory for this script is the project directory.

# Load packages

library(openxlsx)

Warning: package 'openxlsx' was built under R version 4.0.3

library(readxl)  
library(R.utils)  
library(tools)  
library(data.table)

Warning: package 'data.table' was built under R version 4.0.3

library(chron)  
library(dplyr)

Warning: package 'dplyr' was built under R version 4.0.3

# Get name, path and information of the file

data\_file <- list.files(dir\_in, pattern = "\\.xlsx$", full.names = TRUE)  
md5\_in <- md5sum(data\_file)  
info\_in <- data.frame(files = basename(names(md5\_in)), checksum = md5\_in,   
 row.names = NULL)

The checksum (MD5 hashes) of the imported files are:

files checksum  
1 Balver\_Höhle.xlsx 0fcdc965c57b8f87c33fd0f8407a0ea9

# Read in original xlsx-file

imp\_data <- read.xlsx(xlsxFile = data\_file, sheet = 1, startRow = 1, colNames = TRUE,  
 rowNames = FALSE, skipEmptyRows = FALSE)   
#select the columns to keep   
keep\_col <- c(1:2, 9:19, 21:35, 37:38, 40:43)  
data\_final <- imp\_data[,keep\_col]  
  
data\_final$length <- as.numeric(data\_final$length)  
data\_final$width <- as.numeric(data\_final$width)  
data\_final$thickness <- as.numeric(data\_final$thickness)  
data\_final$weight <- as.numeric(data\_final$weight)

Warning: NAs introduced by coercion

data\_final$perimeter.basis.back <- as.numeric(data\_final$perimeter.basis.back)  
data\_final$perimeter.arch <- as.numeric(data\_final$perimeter.arch)

Error in `$<-.data.frame`(`\*tmp\*`, perimeter.arch, value = numeric(0)): replacement has 0 rows, data has 347

data\_final$perimeter.active.edge <- as.numeric(data\_final$perimeter.active.edge)  
data\_final$perimeter.total <- as.numeric(data\_final$perimeter.total)  
data\_final$thickness.back <- as.numeric(data\_final$thickness.back)  
  
str(data\_final)

'data.frame': 347 obs. of 34 variables:  
 $ site : chr "Balver\_Höhle" "Balver\_Höhle" "Balver\_Höhle" "Balver\_Höhle" ...  
 $ ID : chr "HE-012" "HE-013" "HE-014" "HE-015" ...  
 $ raw.material : chr "baltic\_flint" "silicified\_schist" "silicified\_schist" "silicified\_schist" ...  
 $ technological.class : chr "Keilmesser" "Keilmesser" "Keilmesser" "Keilmesser" ...  
 $ artefact.state : chr "complete" "semifinished\_product" "complete" "complete" ...  
 $ blank : chr "core" "core" "flake" "core" ...  
 $ morpho.type : chr "Bockstein" "Bockstein" "Balve" "Pradnik" ...  
 $ cortex : chr "YES" "YES" "YES" "N/A" ...  
 $ cortex.percentage : chr "N/A" "N/A" "N/A" NA ...  
 $ cortex.location : chr "back" "back" "back" NA ...  
 $ morphology.back : chr "cortex/unworked" "cortex/unworked" "cortex/unworked" "partly\_retouched" ...  
 $ retouch.active.edge : chr "YES" "YES" "YES" "YES" ...  
 $ retouch.type.edge : chr "bifacial" "bifacial" "bifacial" "bifacial" ...  
 $ tip.morphology : chr "undeterminable" "undeterminable" "undeterminable" "undeterminable" ...  
 $ application.Pradnik.method : chr "YES" "NO" "NO" "NO" ...  
 $ frequency.application.Pradnik.method: chr "N/A" NA NA NA ...  
 $ type.lateral.sharpening.spall : chr NA NA NA NA ...  
 $ tool.lateralisation : chr "sin." "sin." "dex." "sin." ...  
 $ length : num 72.5 142.8 52.4 48.3 58.9 ...  
 $ width : num 41.6 69.7 38.2 34 30.5 ...  
 $ thickness : num 17.8 24.2 20.9 11.6 19.7 ...  
 $ weight : num 0.058 0.129 0.035 0.022 0.037 ...  
 $ perimeter.basis.back : num 10.6 18.7 5.7 4.1 9.6 7.6 6.3 8.5 4.6 10.8 ...  
 $ perimeter.distal.posterior.part : chr "1" "4" "4" " 4.6" ...  
 $ perimeter.active.edge : num 6.7 11.4 4.7 4.4 4.4 6.4 5.9 5.4 5.9 5.4 ...  
 $ perimeter.total : num 18.3 34.1 14.1 13.1 15.3 ...  
 $ thickness.back : num 17.16 22.07 19.03 7.06 20.03 ...  
 $ taphonomic.visual.inspection : chr "N/A" "N/A" "N/A" "N/A" ...  
 $ tool.edges.preservation : chr "N/A" "N/A" "N/A" "N/A" ...  
 $ macroscopically.visible.use-wear : chr "N/A" "N/A" "N/A" "N/A" ...  
 $ use-wear.analysis : chr "N/A" "N/A" "N/A" "N/A" ...  
 $ 3D-scan : chr "N/A" "N/A" "N/A" "N/A" ...  
 $ schistosity : chr "N/A" "N/A" "N/A" "N/A" ...  
 $ orientation.schistosity : chr NA NA NA NA ...

# Data analsysis - sorting

## Dimension

# keeps only columns relevant for dimensions and sorts them based on   
# their technological class  
keep\_col <- c(1:2, 4:5,19:21)  
dimensions <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_dimensions <- dimensions[1:191, ] %>% arrange(artefact.state)  
KM.point\_dimensions <- KM\_dimensions[159:179,]  
KM.only\_dimensions <- KM\_dimensions[1:158,]  
KM.complete\_dimensions <- KM\_dimensions[1:179,]  
PS\_dimensions <- dimensions[309:335,] %>% arrange(artefact.state)  
LSS\_dimensions <- dimensions[192:308,] %>% arrange(artefact.state)  
S\_dimensions <- dimensions[336:347,] %>% arrange(artefact.state)

## Perimeter

# keeps only columns relevant for perimeter measurements and sorts them   
# based on their technological class  
keep\_col <- c(1:2, 4:5,7,23:26)  
perimeter <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_perimeter <- perimeter[1:191, ] %>% arrange(artefact.state)  
PS\_perimeter <- perimeter[309:335,] %>% arrange(artefact.state)

## Weight

# keeps only columns relevant for weight measurements and sorts them based   
# on their technological class  
keep\_col <- c(1:2, 4:5, 22)  
weight <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_weight <- weight[1:191, ] %>% arrange(artefact.state)  
PS\_weight <- weight[309:335,] %>% arrange(artefact.state)  
LSS\_weight <- weight[192:308,] %>% arrange(artefact.state)  
S\_weight <- weight[336:347,] %>% arrange(artefact.state)

## Raw material

# keeps only columns relevant for raw material classification and sorts them   
# based on their technological class  
keep\_col <- c(1:2, 4:5, 3)  
raw\_material <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_raw\_material <- raw\_material[1:191, ] %>% arrange(artefact.state)  
PS\_raw\_material <- raw\_material[309:335,] %>% arrange(artefact.state)  
LSS\_raw\_material <- raw\_material[192:308,] %>% arrange(artefact.state)  
S\_raw\_material <- raw\_material[336:347,] %>% arrange(artefact.state)

## Cortex + blanks

# keeps only columns relevant for cortex and blank classification and sorts them   
# based on their technological class  
keep\_col <- c(1:2, 4:5, 6, 8:10)  
cortex\_blanks <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_cortex\_blanks <- cortex\_blanks[1:191, ] %>% arrange(artefact.state)  
PS\_cortex\_blanks <- cortex\_blanks[309:335,] %>% arrange(artefact.state)

## Back

# keeps only columns relevant for back modifications and sorts them based   
# on their technological class  
keep\_col <- c(1:2, 4:5, 11, 27)  
back <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_back <- back[1:191, ] %>% arrange(artefact.state)  
PS\_back <- back[309:335,] %>% arrange(artefact.state)

## Edge retouch

# keeps only columns relevant for edge retouch classification and sorts them   
# based on their technological class  
keep\_col <- c(1:2, 4:5, 12:13, 29)  
edge\_retouch <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_edge\_retouch <- edge\_retouch[1:191, ] %>% arrange(artefact.state)  
PS\_edge\_retouch <- edge\_retouch[309:335,] %>% arrange(artefact.state)  
LSS\_edge\_retouch <- edge\_retouch[192:308,] %>% arrange(artefact.state)  
S\_edge\_retouch <- edge\_retouch[336:347,] %>% arrange(artefact.state)

## Morpho type

# keeps only columns relevant for morpho type classification and sorts them   
# based on their technological class  
keep\_col <- c(1:2, 4:5, 7, 19:21)  
morpho.type <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_morpho.type <- morpho.type[1:191, ] %>% arrange(artefact.state)  
PS\_morpho.type <- morpho.type[309:335,] %>% arrange(artefact.state)

## Application ‘Pradnik method’

# keeps only columns relevant for 'morpho type 'Pradnik method' classification   
# and sorts them based on their technological class  
keep\_col <- c(1:2, 4:5, 15:16)  
Pradnik.method <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_Pradnik.method <- Pradnik.method[1:191, ] %>% arrange(artefact.state)  
PS\_Pradnik.method <- Pradnik.method[309:335,] %>% arrange(artefact.state)

## Lateralisation

# keeps only columns relevant for lateralisation and sorts them based on their  
# technological class  
keep\_col <- c(1:2, 4:5, 18)  
lateralisation <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
KM\_lateralisation <- lateralisation[1:191, ] %>% arrange(artefact.state)  
PS\_lateralisation <- lateralisation[309:335,] %>% arrange(artefact.state)

## Type lateral sharpening spall

# keeps only columns relevant for lateral sharpening spall classification and sorts   
# them based on their technological class  
keep\_col <- c(1:2, 4:5, 17:18)  
lss\_type <- data\_final[, keep\_col] %>% arrange(technological.class)  
  
LSS\_type <- lss\_type[192:308, ] %>% arrange(artefact.state)

# Save data

## Format name of output file

file\_out <- "Balve\_lithic\_analysis"

The files will be saved as “~/Balve\_lithic\_analysis.[ext]”.

## Write to XLSX

write.xlsx(list(data = data\_final, dimensions = dimensions, KM\_dimensions = KM\_dimensions, KM.point\_dimensions =  
 KM.point\_dimensions, KM.only\_dimensions = KM.only\_dimensions,  
 KM.complete\_dimensions = KM.complete\_dimensions, PS\_dimensions =  
 PS\_dimensions, LSS\_dimensions = LSS\_dimensions, S\_dimensions =  
 S\_dimensions, KM\_perimeter = KM\_perimeter, PS\_perimeter = PS\_perimeter,  
 KM\_weight = KM\_weight, PS\_weight = PS\_weight,LSS\_weight = LSS\_weight,  
 S\_weight = S\_weight, KM\_raw\_material = KM\_raw\_material,PS\_raw\_material =  
 PS\_raw\_material, LSS\_raw\_material = LSS\_raw\_material, S\_raw\_material =  
 S\_raw\_material, KM\_cortex\_blanks = KM\_cortex\_blanks, PS\_cortex\_blanks =  
 PS\_cortex\_blanks, KM\_back = KM\_back, PS\_back = PS\_back, KM\_edge\_retouch =  
 KM\_edge\_retouch, PS\_edge\_retouch = PS\_edge\_retouch, LSS\_edge\_retouch =  
 LSS\_edge\_retouch, S\_edge\_retouch = S\_edge\_retouch, KM\_morpho.type =  
 KM\_morpho.type, PS\_morpho.type = PS\_morpho.type, KM\_Pradnik.method =  
 KM\_Pradnik.method, PS\_Pradnik.method = PS\_Pradnik.method,  
 KM\_lateralisation = KM\_lateralisation, PS\_lateralisation =  
 PS\_lateralisation, LSS\_type = LSS\_type),  
 file = paste0(dir\_out, file\_out, ".xlsx"))

## Save R object

saveObject(data\_final, file = paste0(dir\_out, file\_out, ".Rbin"))

# sessionInfo() and RStudio version

sessionInfo()

R version 4.0.2 (2020-06-22)  
Platform: x86\_64-w64-mingw32/x64 (64-bit)  
Running under: Windows 10 x64 (build 19041)  
  
Matrix products: default  
  
locale:  
[1] LC\_COLLATE=German\_Germany.1252 LC\_CTYPE=German\_Germany.1252   
[3] LC\_MONETARY=German\_Germany.1252 LC\_NUMERIC=C   
[5] LC\_TIME=German\_Germany.1252   
  
attached base packages:  
[1] tools stats graphics grDevices utils datasets methods   
[8] base   
  
other attached packages:  
[1] dplyr\_1.0.3 chron\_2.3-56 data.table\_1.13.6 R.utils\_2.10.1   
[5] R.oo\_1.24.0 R.methodsS3\_1.8.1 readxl\_1.3.1 openxlsx\_4.2.3   
  
loaded via a namespace (and not attached):  
 [1] Rcpp\_1.0.6 knitr\_1.31 magrittr\_2.0.1 tidyselect\_1.1.0   
 [5] R6\_2.5.0 rlang\_0.4.10 stringr\_1.4.0 xfun\_0.20   
 [9] DBI\_1.1.1 ellipsis\_0.3.1 htmltools\_0.5.1.1 assertthat\_0.2.1   
[13] yaml\_2.2.1 digest\_0.6.27 tibble\_3.0.6 lifecycle\_0.2.0   
[17] crayon\_1.4.0 zip\_2.1.1 purrr\_0.3.4 vctrs\_0.3.6   
[21] glue\_1.4.2 evaluate\_0.14 rmarkdown\_2.6 stringi\_1.5.3   
[25] pillar\_1.4.7 compiler\_4.0.2 cellranger\_1.1.0 generics\_0.1.0   
[29] pkgconfig\_2.0.3

RStudio version 1.3.1056.

END OF SCRIPT