Import raw data

Joao Marreiros

2021-04-22 14:20:55

Brief description of the script

This R markdown document imports and formats the output of the resulting CSV file from the computing ISO 25178-2 parameters in ConfoMap. These data is part of the manuscript: *Dubreuil et al. A 'family of wear': Exploring use-wear patterns on ad hoc burnishing tools*

The script includes three steps:

1. Reads in the original CSV-file
2. Formats the data
3. Writes an XLSX-file and save an R object ready for further analysis in R

This R project and respective scripts follow the procedures described by Marwick et al. 2017.

The authors would like to thank Ivan Calandra and Lisa Schunk for their help and contribution on several chunks of code included here in the script (pieces of code are also adapated from Calandra et al. 2019, Pedergnana et al. 2020a, 2020b).

To compile this markdown document do not delete or move files from their original folders.

For any questions, comments and inputs, please contact:

Joao Marreiros, marreiros@rgzm.de

Load libraries

```
library(openxlsx)
library(tools)
library(R.utils)
library(chron)

dir_in <- "analysis/raw_data/"
dir_out <- "analysis/derived_data/"
```

Get file names, path and info

```
data_file <- list.files(dir_in, pattern = "\\.csv$", full.names = TRUE)
md5_in <- md5sum(data_file)
```

Import and read the original CSV-file

```
imp_data <- read.csv(data_file, header = FALSE, na.strings = "*****", encoding = "latin1")
str(imp_data)
```

```
'data.frame':   33 obs. of  105 variables:
 $ V1  : chr  "#" "#" "#" "2021/04/15" ...
 $ V2  : chr  "#" "#" "#" "14:06:02" ...
 $ V3  : chr  "#" "#" "#" "D:\\Dropbox\\jmmarreiros_dropbox\\Work\\Papers and articles\\2021 - Debreuil
 $ V4  : chr  "OPERATOR:1" "X-axis rotation angle" "°" "1.389601332" ...
 $ V5  : chr  "OPERATOR:1" "Y-axis rotation angle" "°" "-3.540040607" ...
 $ V6  : chr  "OPERATOR:2" "a0" "µm" "-3.327279468" ...
 $ V7  : chr  "OPERATOR:2" "ax" "µm" "-0.007136344206" ...
 $ V8  : chr  "OPERATOR:2" "ax2" "µm" "6.08357491E-06" ...
 $ V9  : chr  "OPERATOR:2" "ax3" "µm" "-1.075360198E-09" ...
 $ V10 : chr  "OPERATOR:2" "ay" "µm" "0.004476105667" ...
 $ V11 : chr  "OPERATOR:2" "axy" "µm" "6.779964383E-07" ...
 $ V12 : chr  "OPERATOR:2" "ax2y" "µm" "-1.114150037E-09" ...
 $ V13 : chr  "OPERATOR:2" "ay2" "µm" "2.420272666E-06" ...
 $ V14 : chr  "OPERATOR:2" "axy2" "µm" "5.305851231E-10" ...
 $ V15 : chr  "OPERATOR:2" "ay3" "µm" "-1.457157799E-09" ...
 $ V16 : chr  "6" "Name" "<no unit>" "Kremasti4_LSM_50x_natural_a_Topo" ...
 $ V17 : chr  "6" "Created on" "<no unit>" "4/15/2021 10:24:41 AM" ...
 $ V18 : chr  "6" "Studiable type" "<no unit>" "Surface" ...
 $ V19 : chr  "6" "Axis name - X" "<no unit>" "X" ...
 $ V20 : chr  "6" "Axis length - X" "µm" "255.4748064" ...
 $ V21 : chr  "6" "Axis size - X" "points" "3000" ...
 $ V22 : chr  "6" "Axis spacing - X" "nm" "85.18666436" ...
 $ V23 : chr  "6" "Axis name - Y" "<no unit>" "Y" ...
 $ V24 : chr  "6" "Axis length - Y" "µm" "255.4748064" ...
 $ V25 : chr  "6" "Axis size - Y" "points" "3000" ...
 $ V26 : chr  "6" "Axis spacing - Y" "nm" "85.18666436" ...
 $ V27 : chr  "6" "Axis name - Z" "<no unit>" "Z" ...
 $ V28 : chr  "6" "Layer type - Z" "<no unit>" "Topography" ...
 $ V29 : chr  "6" "Axis length - Z" "µm" "40.71166716" ...
 $ V30 : chr  "6" "Axis size - Z" "digits" "65532" ...
 $ V31 : chr  "6" "Axis spacing - Z" "nm" "0.6212486596" ...
 $ V32 : chr  "6" "NM-points ratio - Z" "%" "0" ...
 $ V33 : chr  "15" "Name" "<no unit>" "Kremasti4_LSM_50x_natural_a_Topo > Leveled (LS-plane) > Form rem
 $ V34 : chr  "15" "Created on" "<no unit>" "4/15/2021 10:24:41 AM" ...
 $ V35 : chr  "15" "Studiable type" "<no unit>" "Surface" ...
 $ V36 : chr  "15" "Axis name - X" "<no unit>" "X" ...
 $ V37 : chr  "15" "Axis length - X" "µm" "255.4748064" ...
 $ V38 : chr  "15" "Axis size - X" "points" "3000" ...
```

```
$ V39 : chr  "15" "Axis spacing - X" "nm" "85.18666436" ...
$ V40 : chr  "15" "Axis name - Y" "<no unit>" "Y" ...
$ V41 : chr  "15" "Axis length - Y" "µm" "255.4748064" ...
$ V42 : chr  "15" "Axis size - Y" "points" "3000" ...
$ V43 : chr  "15" "Axis spacing - Y" "nm" "85.18666436" ...
$ V44 : chr  "15" "Axis name - Z" "<no unit>" "Z" ...
$ V45 : chr  "15" "Layer type - Z" "<no unit>" "Topography" ...
$ V46 : chr  "15" "Axis length - Z" "µm" "12.32495216" ...
$ V47 : chr  "15" "Axis size - Z" "digits" "198390" ...
$ V48 : chr  "15" "Axis spacing - Z" "nm" "0.06212486596" ...
$ V49 : chr  "15" "NM-points ratio - Z" "%" "0" ...
$ V50 : chr  "17" "Sq" "µm" "1.584021467" ...
$ V51 : chr  "17" "Ssk" "<no unit>" "-0.6095035235" ...
$ V52 : chr  "17" "Sku" "<no unit>" "4.87649139" ...
$ V53 : chr  "17" "Sp" "µm" "5.465869957" ...
$ V54 : chr  "17" "Sv" "µm" "6.859082201" ...
$ V55 : chr  "17" "Sz" "µm" "12.32495216" ...
$ V56 : chr  "17" "Sa" "µm" "1.125056973" ...
$ V57 : chr  "17" "Smr (c = 1 µm below highest peak)" "%" "0.4839345421" ...
$ V58 : chr  "17" "Smc (p = 10%)" "µm" "1.693314186" ...
$ V59 : chr  "17" "Sxp (p = 50%  q = 97.5%)" "µm" "3.99094246" ...
$ V60 : chr  "17" "Sal (s = 0.2)" "µm" "19.51575135" ...
$ V61 : chr  "17" "Str (s = 0.2)" "<no unit>" "0.480092466" ...
$ V62 : chr  "17" "Std (Reference angle = 0°)" "°" "42.24936206" ...
$ V63 : chr  "17" "Sdq" "<no unit>" "0.3826728109" ...
$ V64 : chr  "17" "Sdr" "%" "6.003094867" ...
$ V65 : chr  "17" "Vm (p = 10%)" "µm³/µm²" "0.08953456587" ...
$ V66 : chr  "17" "Vv (p = 10%)" "µm³/µm²" "1.782866126" ...
$ V67 : chr  "17" "Vmp (p = 10%)" "µm³/µm²" "0.08953456587" ...
$ V68 : chr  "17" "Vmc (p = 10%  q = 80%)" "µm³/µm²" "1.129500923" ...
$ V69 : chr  "17" "Vvc (p = 10%  q = 80%)" "µm³/µm²" "1.496020162" ...
$ V70 : chr  "17" "Vvv (p = 80%)" "µm³/µm²" "0.2868459635" ...
$ V71 : chr  "18" "Maximum depth of furrows" "µm" "7.217169929" ...
$ V72 : chr  "18" "Mean depth of furrows" "µm" "1.568997462" ...
$ V73 : chr  "18" "Mean density of furrows" "cm/cm2" "3750.457168" ...
$ V74 : chr  "19" "First direction" "°" "89.9771993" ...
$ V75 : chr  "19" "Second direction" "°" "45.01151178" ...
$ V76 : chr  "19" "Third direction" "°" "179.9901922" ...
$ V77 : chr  "20" "Texture isotropy" "%" "73.9724373" ...
$ V78 : chr  "21" "Length-scale anisotropy (Sfrax) (epLsar)" "<no unit>" NA ...
$ V79 : chr  "21" "Length-scale anisotropy (NewEplsar)" "<no unit>" NA ...
$ V80 : chr  "22" "Fractal complexity (Asfc)" "<no unit>" "9.933672877" ...
$ V81 : chr  "22" "Scale of max complexity (Smfc)" "nm²" "6281984.839" ...
$ V82 : chr  "22" "HAsfc9 (HAsfc9)" "<no unit>" "0.5388613394" ...
$ V83 : chr  "22" "HAsfc81 (HAsfc81)" "<no unit>" "0.8700496629" ...
$ V84 : chr  "128" "Name" "<no unit>" "Kremasti4_LSM_50x_natural_a_Topo > Leveled (LS-plane) > Form re
$ V85 : chr  "128" "File path" "<no unit>" "D:\\Dropbox\\jmmarreiros_dropbox\\Work\\Papers and article
$ V86 : chr  "128" "Created on" "<no unit>" "4/15/2021 10:24:41 AM" ...
$ V87 : chr  "128" "Studiable type" "<no unit>" "Surface" ...
$ V88 : chr  "128" "Axis name - X" "<no unit>" "X" ...
$ V89 : chr  "128" "Axis length - X" "µm" "255.4748064" ...
$ V90 : chr  "128" "Axis size - X" "points" "3000" ...
$ V91 : chr  "128" "Axis spacing - X" "nm" "85.18666436" ...
$ V92 : chr  "128" "Axis offset - X" "µm" "0" ...
```

```
 $ V93 : chr  "128" "Axis name - Y" "<no unit>" "Y" ...
 $ V94 : chr  "128" "Axis length - Y" "µm" "255.4748064" ...
 $ V95 : chr  "128" "Axis size - Y" "points" "3000" ...
 $ V96 : chr  "128" "Axis spacing - Y" "nm" "85.18666436" ...
 $ V97 : chr  "128" "Axis offset - Y" "µm" "-255.4748064" ...
 $ V98 : chr  "128" "Axis name - Z" "<no unit>" "Z" ...
 $ V99 : chr  "128" "Layer type - Z" "<no unit>" "Topography" ...
  [list output truncated]
```

Format data

Keep only interesting columns and rows

```
# keeps only the columns and rows of interest for the analysis
data_keep_col <- c(1:2, 16:17, 20:22, 24:26, 29:32, 50:83)
data_keep_rows <- which(imp_data[[1]] != "#")
data_keep <- imp_data[data_keep_rows, data_keep_col]
```

Add headers

```
head_data_keep <- unlist(imp_data[2, data_keep_col])
colnames(data_keep) <- gsub("\\.+", "\\.", make.names(head_data_keep))
colnames(data_keep) <- gsub("\\.$", "", colnames(data_keep))
```

Identify results using frame numbers

```
# combines the results from the different analysis based on the column numbers
# (ID from MountainsMAp)
frames <- as.numeric(unlist(imp_data[1, data_keep_col]))
```

```
Warning: NAs introduced by coercion
```

```
ID <- which(frames == 6)[-(1:2)]
ISO <- which(frames == 17)
furrow <- which(frames == 18)
diriso <- which(frames %in% 19:20)
SSFA <- which(frames %in% 21:22)
```

Shorten the names for parameters

```
# keeps only the important information of the headers
colnames(data_keep)[ISO] <- sapply(strsplit(names(data_keep)[ISO], ".", fixed = TRUE), `[[`, 1)
colnames(data_keep)[SSFA] <- gsub("^([A-Za-z0-9]+\\.)+", "", colnames(data_keep)[SSFA])
```

Save units

```
# takes the units which were part of the headers and separates them; creates a data frame
var_num <- c(ID, ISO, furrow, diriso, SSFA)
# extracts 'unit' line for considered columns
units_var <- unlist(imp_data[3, data_keep_col])[var_num]
# gets names associated to the units
names(units_var) <- head_data_keep[var_num]
# puts all of it into a data.frame
units_var_table <- data.frame(variable = names(units_var), unit = units_var)
```

Convert to numeric

```
for (i in var_num) data_keep[[i]] <- as.numeric(data_keep[[i]])
```

Split the column 'Name' into several columns

```
# these lines extract the artefact ID out of the path name
stud_name <- gsub(".* --- ", "", data_keep[["Name"]])
split_name <- do.call(rbind, strsplit(stud_name, "_"))

# splits the ID in the separat information
data_final <- data.frame(split_name[,1], split_name[,2], split_name[,3], split_name[,4], split_name[,5]
colnames(data_final)[1:9] <- c("Sample.ID", "Microscope", "Objective", "PolishType", "Surface", "Topo",
```

Check the result

```
str(data_final)
```

```
'data.frame':   30 obs. of  53 variables:
 $ Sample.ID              : chr  "Kremasti4" "Kremasti4" "Kremasti4" "Kremasti4" ...
 $ Microscope             : chr  "LSM" "LSM" "LSM" "LSM" ...
 $ Objective              : chr  "50x" "50x" "50x" "50x" ...
 $ PolishType             : chr  "natural" "natural" "natural" "natural" ...
 $ Surface                : chr  "a" "b" "c" "d" ...
 $ Topo                   : chr  "Topo" "Topo" "Topo" "Topo" ...
 $ Acquisition.Date       : chr  "2021/04/15" "2021/04/15" "2021/04/15" "2021/04/15" ...
 $ Analysis.Date          : chr  "14:06:02" "14:08:39" "14:11:12" "14:13:40" ...
 $ Analysis.Time          : chr  "4/15/2021 10:24:41 AM" "4/15/2021 10:51:24 AM" "4/15/2021 11:45:47 AM
 $ Axis.length.X          : num  255 255 255 255 255 ...
 $ Axis.size.X            : num  3000 3000 3000 3000 3000 3000 3000 3000 3000 3000 ...
 $ Axis.spacing.X         : num  85.2 85.2 85.2 85.2 85.2 ...
 $ Axis.length.Y          : num  255 255 255 255 255 ...
 $ Axis.size.Y            : num  3000 3000 3000 3000 3000 3000 3000 3000 3000 3000 ...
 $ Axis.spacing.Y         : num  85.2 85.2 85.2 85.2 85.2 ...
 $ Axis.length.Z          : num  40.7 49.9 92.6 31.8 29.3 ...
```

```
 $ Axis.size.Z             : num  65532 65532 65531 65532 65531 ...
 $ Axis.spacing.Z          : num  0.621 0.761 1.413 0.485 0.447 ...
 $ NM.points.ratio.Z       : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sq                      : num  1.58 4.09 1.47 2.24 1.77 ...
 $ Ssk                     : num  -0.61 -0.391 -0.274 -0.049 -0.929 ...
 $ Sku                     : num  4.88 2.53 6.23 3.44 5.87 ...
 $ Sp                      : num  5.47 10.44 5.15 9.02 7.91 ...
 $ Sv                      : num  6.86 12.48 8.2 7.06 8.52 ...
 $ Sz                      : num  12.3 22.9 13.4 16.1 16.4 ...
 $ Sa                      : num  1.13 3.4 1.08 1.77 1.25 ...
 $ Smr                     : num  0.484 0.239 0.604 0.207 0.126 ...
 $ Smc                     : num  1.69 4.56 1.72 2.87 1.66 ...
 $ Sxp                     : num  3.99 9.02 2.67 4.65 4.7 ...
 $ Sal                     : num  19.5 32.4 23.5 30.6 20.7 ...
 $ Str                     : num  0.48 NA NA 0.614 0.813 ...
 $ Std                     : num  42.2 93.2 33 25.3 62 ...
 $ Sdq                     : num  0.383 0.658 0.521 0.403 0.403 ...
 $ Sdr                     : num  6 15.94 8.53 6.73 6.56 ...
 $ Vm                      : num  0.0895 0.1529 0.1057 0.0994 0.0945 ...
 $ Vv                      : num  1.78 4.72 1.82 2.97 1.75 ...
 $ Vmp                     : num  0.0895 0.1529 0.1057 0.0994 0.0945 ...
 $ Vmc                     : num  1.13 4.14 1.09 1.84 1.28 ...
 $ Vvc                     : num  1.5 4.24 1.65 2.68 1.41 ...
 $ Vvv                     : num  0.287 0.478 0.17 0.292 0.343 ...
 $ Maximum.depth.of.furrows: num  7.22 10.88 9.43 6.76 8.68 ...
 $ Mean.depth.of.furrows   : num  1.57 3.13 1.47 1.9 1.54 ...
 $ Mean.density.of.furrows : num  3750 3056 4011 3480 3423 ...
 $ First.direction         : num  89.9772 90.014 45.0229 0.0123 44.9941 ...
 $ Second.direction        : num  45 135 180 26.5 63.5 ...
 $ Third.direction         : num  180 45 33.7 90 90 ...
 $ Texture.isotropy        : num  74 82.7 77.8 90.3 92.3 ...
 $ epLsar                  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ NewEplsar               : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Asfc                    : num  9.93 25.92 17.49 11.21 10.47 ...
 $ Smfc                    : num  6281985 10723090 4628049 7318909 11574299 ...
 $ HAsfc9                  : num  0.539 0.39 1.927 0.603 0.546 ...
 $ HAsfc81                 : num  0.87 0.638 2.369 0.728 0.848 ...
```

```
head(data_final)
```

```
  Sample.ID Microscope Objective PolishType Surface Topo Acquisition.Date
4 Kremasti4        LSM       50x    natural       a Topo       2021/04/15
5 Kremasti4        LSM       50x    natural       b Topo       2021/04/15
6 Kremasti4        LSM       50x    natural       c Topo       2021/04/15
7 Kremasti4        LSM       50x    natural       d Topo       2021/04/15
8 Kremasti4        LSM       50x    natural       e Topo       2021/04/15
9 Kremasti4        LSM       50x      type1       a Topo       2021/04/15
  Analysis.Date          Analysis.Time Axis.length.X Axis.size.X Axis.spacing.X
4      14:06:02 4/15/2021 10:24:41 AM      255.4748        3000       85.18666
5      14:08:39 4/15/2021 10:51:24 AM      255.4748        3000       85.18666
6      14:11:12 4/15/2021 11:45:47 AM      255.4748        3000       85.18666
7      14:13:40 4/15/2021 12:01:50 PM      255.4748        3000       85.18666
8      14:16:06 4/15/2021 12:16:18 PM      255.4748        3000       85.18666
9      14:18:42  4/13/2021 3:00:34 PM      255.4748        3000       85.18666
```

```
  Axis.length.Y Axis.size.Y Axis.spacing.Y Axis.length.Z Axis.size.Z
4      255.4748        3000       85.18666      40.71167       65532
5      255.4748        3000       85.18666      49.89209       65532
6      255.4748        3000       85.18666      92.57281       65531
7      255.4748        3000       85.18666      31.79712       65532
8      255.4748        3000       85.18666      29.31425       65531
9      255.4748        3000       85.18666      24.97545       65531
  Axis.spacing.Z NM.points.ratio.Z       Sq         Ssk      Sku        Sp
4      0.6212487                 0 1.584021 -0.60950352 4.876491  5.465870
5      0.7613393                 0 4.090417 -0.39145394 2.527444 10.438800
6      1.4126567                 0 1.471140 -0.27367641 6.226369  5.151535
7      0.4852151                 0 2.240846 -0.04897288 3.435836  9.022964
8      0.4473341                 0 1.767716 -0.92850679 5.866180  7.914458
9      0.3811242                 0 1.434170 -1.65028211 6.832436  3.630703
         Sv       Sz       Sa       Smr      Smc      Sxp      Sal       Str
4  6.859082 12.32495 1.125057 0.4839345 1.693314 3.990942 19.51575 0.4800925
5 12.482615 22.92141 3.402968 0.2388669 4.564882 9.019519 32.42663        NA
6  8.201602 13.35314 1.084219 0.6035167 1.717278 2.666714 23.54556        NA
7  7.063908 16.08687 1.765024 0.2065847 2.872713 4.647387 30.56610 0.6142008
8  8.517286 16.43174 1.248673 0.1263085 1.657583 4.698921 20.72937 0.8132318
9  7.463707 11.09441 1.009955 0.5506006 1.290805 4.246005 18.83861 0.6902008
       Std       Sdq       Sdr          Vm       Vv        Vmp      Vmc
4  42.24936 0.3826728  6.003095 0.08953457 1.782866 0.08953457 1.129501
5  93.24467 0.6584596 15.937044 0.15288943 4.717775 0.15288943 4.141595
6  33.00070 0.5207327  8.532379 0.10571506 1.823054 0.10571506 1.087074
7  25.25098 0.4030433  6.729032 0.09941242 2.972141 0.09941242 1.843508
8  62.00315 0.4031913  6.562433 0.09447176 1.752063 0.09447176 1.276085
9 148.49476 0.3498352  5.156392 0.04783352 1.338637 0.04783352 0.982558
      Vvc       Vvv Maximum.depth.of.furrows Mean.depth.of.furrows
4 1.496020 0.2868460                7.217170              1.568997
5 4.240142 0.4776329               10.880719              3.134721
6 1.653056 0.1699981                9.434004              1.469384
7 2.679862 0.2922788                6.760114              1.903889
8 1.408635 0.3434273                8.684991              1.535972
9 1.015295 0.3233419                6.789918              1.217140
  Mean.density.of.furrows First.direction Second.direction Third.direction
4                3750.457     89.97719930         45.01151       179.99019
5                3056.415     90.01401251        134.99001        45.00416
6                4010.997     45.02286722        179.99523        33.71723
7                3479.585      0.01225454         26.47753        89.99378
8                3422.848     44.99412580         63.51278        90.01651
9                3631.833    179.98336160         89.99780       135.00711
  Texture.isotropy epLsar NewEplsar      Asfc     Smfc    HAsfc9   HAsfc81
4         73.97244     NA        NA  9.933673  6281985 0.5388613 0.8700497
5         82.66691     NA        NA 25.923849 10723090 0.3895934 0.6379997
6         77.84103     NA        NA 17.485180  4628049 1.9267261 2.3693875
7         90.31548     NA        NA 11.212214  7318909 0.6032577 0.7279640
8         92.32249     NA        NA 10.466204 11574299 0.5456690 0.8480940
9         69.99147     NA        NA  9.047928  6281985 0.2340000 0.8845739
```

Save data

Format name of output file

```
file_out <- "data"
```

The files will be saved as "~/data.[ext]".

Write to XLSX

```
write.xlsx(list(data = data_final, units = units_var_table),
           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 4.0.4 (2021-02-15)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Catalina 10.15.7

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] tools     stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
[1] chron_2.3-56      R.utils_2.10.1    R.oo_1.24.0       R.methodsS3_1.8.1
[5] openxlsx_4.2.3

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.6        digest_0.6.27     rprojroot_2.0.2   magrittr_2.0.1
 [5] evaluate_0.14     zip_2.1.1         rlang_0.4.10      stringi_1.5.3
 [9] rmarkdown_2.7     stringr_1.4.0     xfun_0.22         yaml_2.2.1
[13] compiler_4.0.4    htmltools_0.5.1.1 knitr_1.32
```