Equotip hardness measurements - ‘Initial experiment’

Lisa Schunk

2021-02-04

# Goal of the script

This script reads the xlsx file (measurements have been generated with the Equotip Leeb C rebound) and formats the data for a statistical analysis.  
The script will:

1. Read in the original xlsx file and organise the data
2. Plot the data
3. Write an XLSX-file and save an R object ready for further analysis in R

dir\_in <- "analysis/raw\_data/"  
dir\_out <- "analysis/plots"

Raw data must be located in ~/analysis/raw\_data/.  
Formatted data will be saved in ~/analysis/plots. 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(readr)

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

library(tools)  
library(ggplot2)

Warning: package 'ggplot2' was built under R version 4.0.3

library(readxl)  
library(tidyverse)

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

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

library(wesanderson)

Warning: package 'wesanderson' 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 IE\_hardness\_flint+lydite.xlsx e1decfc698cd4cad8eec271bd012b5d5

# Read in original xlsx-file

imp\_data <- read.xlsx(xlsxFile = data\_file, sheet = 1, startRow = 3,   
 colNames = TRUE,  
 rowNames = FALSE, skipEmptyRows = FALSE)   
  
# changes the mode of 'hardness in HLC' from character to numeric   
imp\_data$hardness.in.HLC <- as.numeric(imp\_data$hardness.in.HLC)  
str(imp\_data)

'data.frame': 36 obs. of 2 variables:  
 $ raw.material : chr "flint" "flint" "flint" "flint" ...  
 $ hardness.in.HLC: num 963 953 959 961 965 965 967 963 967 965 ...

flint <- imp\_data[1:20,]  
lydite <- imp\_data[31:35,]  
balve <- imp\_data[36,]

# Data analsysis - stats

# descriptive statistics   
# flint  
length(flint[["raw.material"]])

[1] 20

summary(flint[["hardness.in.HLC"]])

Min. 1st Qu. Median Mean 3rd Qu. Max.   
 953.0 960.8 963.0 962.2 965.2 969.0

# lydite  
length(lydite[["raw.material"]])

[1] 5

summary(lydite[["hardness.in.HLC"]])

Min. 1st Qu. Median Mean 3rd Qu. Max.   
 896.0 898.0 907.0 910.8 920.0 933.0

# balve (MU-278)  
length(balve[["raw.material"]])

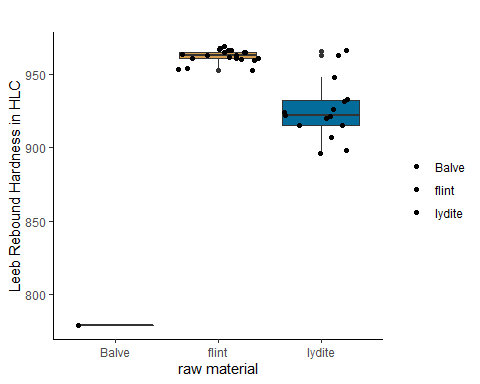
[1] 1

summary(balve[["hardness.in.HLC"]])

Min. 1st Qu. Median Mean 3rd Qu. Max.   
 779 779 779 779 779 779

# Data analsysis - plot

# boxplot  
data\_plot <- ggplot (imp\_data, aes(y = hardness.in.HLC, x = raw.material,   
 fill = raw.material)) +   
 theme\_classic() +  
 theme(legend.title = element\_blank()) +  
 geom\_boxplot(fill = c("#ECCBAE", "#D69C4E", "#046C9A")) +   
 geom\_jitter() + labs(x="raw material",   
 y="Leeb Rebound Hardness in HLC", title="")   
  
print(data\_plot)



# Save data

## Define output file name

file\_out <- paste0(file\_path\_sans\_ext(info\_in[["file"]]), "\_IE\_plot", ".pdf")  
ggsave(filename = file\_out, plot = data\_plot, path = dir\_out, device = "pdf")

The files will be saved as “~/analysis/plots.[ext]”.

# 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] wesanderson\_0.3.6 forcats\_0.5.1 stringr\_1.4.0 dplyr\_1.0.3   
 [5] purrr\_0.3.4 tidyr\_1.1.2 tibble\_3.0.6 tidyverse\_1.3.0   
 [9] readxl\_1.3.1 ggplot2\_3.3.3 readr\_1.4.0 openxlsx\_4.2.3   
  
loaded via a namespace (and not attached):  
 [1] tidyselect\_1.1.0 xfun\_0.20 haven\_2.3.1 colorspace\_2.0-0   
 [5] vctrs\_0.3.6 generics\_0.1.0 htmltools\_0.5.1.1 yaml\_2.2.1   
 [9] rlang\_0.4.10 pillar\_1.4.7 glue\_1.4.2 withr\_2.4.1   
[13] DBI\_1.1.1 dbplyr\_2.0.0 modelr\_0.1.8 lifecycle\_0.2.0   
[17] munsell\_0.5.0 gtable\_0.3.0 cellranger\_1.1.0 rvest\_0.3.6   
[21] zip\_2.1.1 evaluate\_0.14 labeling\_0.4.2 knitr\_1.31   
[25] highr\_0.8 broom\_0.7.4 Rcpp\_1.0.6 scales\_1.1.1   
[29] backports\_1.2.1 jsonlite\_1.7.2 farver\_2.0.3 fs\_1.5.0   
[33] hms\_1.0.0 digest\_0.6.27 stringi\_1.5.3 grid\_4.0.2   
[37] cli\_2.3.0 magrittr\_2.0.1 crayon\_1.4.0 pkgconfig\_2.0.3   
[41] ellipsis\_0.3.1 xml2\_1.3.2 reprex\_1.0.0 lubridate\_1.7.9.2  
[45] rstudioapi\_0.13 assertthat\_0.2.1 rmarkdown\_2.6 httr\_1.4.2   
[49] R6\_2.5.0 compiler\_4.0.2

RStudio version 1.3.1073.

END OF SCRIPT