Plots 3D data analysis

Eduardo Paixão, Tegenu Gossa, Walter Gneisinger, João Marreiros, Soren Tholen, Ivan Calandra, Erella

2024-01-19 15:47:02.468451

Contents

Content

This script reads and plots the data from the 3D volume loss and Leeb Rebound hardness. 3D volume loss was calculated using CloudCompare cloud-to-mesh distance tool. Hardness was measured on each rock using the device Equotip Leeb Impact device C (HLC). For details on the methods and data acquisition, please visit the Materials and Methods section of the paper. The knit directory for this script is the project directory.

Load packages

```r
library(R.utils)
```

```
## Warning: package ’R.utils’ was built under R version 4.3.1
```

```
## Loading required package: R.oo
```

```
## Loading required package: R.methodsS3
```

```
## R.methodsS3 v1.8.2 (2022-06-13 22:00:14 UTC) successfully loaded. See ?R.methodsS3 for help.
```

```
## R.oo v1.25.0 (2022-06-12 02:20:02 UTC) successfully loaded. See ?R.oo for help.
```

```
##
## Attaching package: ’R.oo’
```

```
## The following object is masked from ’package:R.methodsS3’:
##
##     throw
```

```
## The following objects are masked from ’package:methods’:
##
##     getClasses, getMethods
```

```
## The following objects are masked from ’package:base’:
##
##     attach, detach, load, save
```

```
## R.utils v2.12.3 (2023-11-18 01:00:02 UTC) successfully loaded. See ?R.utils for help.
```

```
##
## Attaching package: ’R.utils’
```

```
## The following object is masked from ’package:utils’:
##
##     timestamp
```

```
## The following objects are masked from ’package:base’:
##
##     cat, commandArgs, getOption, isOpen, nullfile, parse, warnings
```

```r
library(ggplot2)
```

```
## Warning: package ’ggplot2’ was built under R version 4.3.1
```

```r
library(tools)
library(tidyverse)
```

```
## Warning: package 'readr' was built under R version 4.3.1

## Warning: package 'dplyr' was built under R version 4.3.1

## Warning: package 'stringr' was built under R version 4.3.1

## Warning: package 'lubridate' was built under R version 4.3.1

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract() masks R.utils::extract()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(doBy)
```

```
## Warning: package 'doBy' was built under R version 4.3.1

##
## Attaching package: 'doBy'
##
## The following object is masked from 'package:dplyr':
##
##     order_by
```

```r
library(ggrepel)
```

```
## Warning: package 'ggrepel' was built under R version 4.3.1
```

```r
library(flextable)
```

```
## Warning: package 'flextable' was built under R version 4.3.1

##
## Attaching package: 'flextable'
##
## The following object is masked from 'package:purrr':
##
##     compose
```

```
library(readr)
library(ggtern)
```

```
## Registered S3 methods overwritten by 'ggtern':
##   method           from
##   grid.draw.ggplot ggplot2
##   plot.ggplot      ggplot2
##   print.ggplot     ggplot2
## --
## Remember to cite, run citation(package = 'ggtern') for further info.
## --
##
## Attaching package: 'ggtern'
##
## The following objects are masked from 'package:ggplot2':
##
##     aes, annotate, ggplot, ggplot_build, ggplot_gtable, ggplotGrob,
##     ggsave, layer_data, theme_bw, theme_classic, theme_dark,
##     theme_gray, theme_light, theme_linedraw, theme_minimal, theme_void
```

Import and preview data

```
hlcdata <- read_csv2("../rawdata/hlcdata.csv")
```

```
## i Using "','" as decimal and "'.'" as grouping mark. Use ‘read_delim()‘ for more control.

## Rows: 5 Columns: 30
## -- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr  (11): cubeid, Date, Probe, Probe serial, Probe verification, Operator, ...
## dbl  (17): Upper limit, Lower limit, Readings, Mean, Min, Max, Range, readin...
## num   (1): Std. dev.
## time  (1): Time
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
volumelossdata <- read_csv2("../rawdata/volumelossdata.csv")
```

```
## i Using "','" as decimal and "'.'" as grouping mark. Use ‘read_delim()‘ for more control.
## Rows: 1280 Columns: 7-- Column specification -------------------------------------------------------
## Delimiter: ";"
## chr (5): cubeid, Material, Cycle, Class start, Class end
## dbl (2): Class, Value
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```r
densitydata <- read_csv2("../rawdata/densitydata.csv")
```

```
## i Using "’,’" as decimal and "’.’" as grouping mark. Use ‘read_delim()‘ for more control.
## Rows: 5 Columns: 5-- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr (2): cubeid, Material
## dbl (3): mass, volume, density
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```r
str(hlcdata)
```

```
## spc_tbl_ [5 x 30] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ cubeid              : chr [1:5] "DW-G1-S3" "DW-G3-S3" "DW-G4-S3" "KW-G1-S3" ...
##  $ Date                : chr [1:5] "16/11/2022" "16/11/2022" "16/11/2022" "16/11/2022" ...
##  $ Time                : ’hms’ num [1:5] 15:55:29 15:47:15 15:52:31 15:50:56 ...
##   ..- attr(*, "units")= chr "secs"
##  $ Probe               : chr [1:5] "Equotip Leeb Impact Device C" "Equotip Leeb Impact Device C" "Equ
##  $ Probe serial        : chr [1:5] "IC51-004-0185" "IC51-004-0185" "IC51-004-0185" "IC51-004-0185" ..
##  $ Probe verification : chr [1:5] "--" "--" "--" "--" ...
##  $ Operator            : chr [1:5] "Paixao" "Paixao" "Paixao" "Paixao" ...
##  $ Device serial       : chr [1:5] "UP01-003-1680" "UP01-003-1680" "UP01-003-1680" "UP01-003-1680" ..
##  $ Direction (Equotip): chr [1:5] "Automatic" "Automatic" "Automatic" "Automatic" ...
##  $ Material            : chr [1:5] "basalt" "scoria" "pumaceousignimbrite" "glassyignimbrite" ...
##  $ Conversion Std.     : chr [1:5] "Default" "Default" "Default" "Default" ...
##  $ Scale               : chr [1:5] "HLC" "HLC" "HLC" "HLC" ...
##  $ Upper limit         : num [1:5] 528 528 528 528 528
##  $ Lower limit         : num [1:5] 498 498 498 498 498
##  $ Readings            : num [1:5] 10 10 10 10 10
##  $ Mean                : num [1:5] 769 717 528 907 672
##  $ Min                 : num [1:5] 743 604 456 789 591
##  $ Max                 : num [1:5] 813 834 605 938 741
##  $ Range               : num [1:5] 70 229 149 149 149
##  $ Std. dev.           : num [1:5] 23 86 455 412 445
##  $ reading1            : num [1:5] 769 644 500 907 591
##  $ reading2            : num [1:5] 750 688 600 920 609
##  $ reading3            : num [1:5] 750 834 456 920 690
##  $ reading4            : num [1:5] 743 824 518 888 671
##  $ reading5            : num [1:5] 783 653 511 921 654
##  $ reading6            : num [1:5] 813 614 488 925 741
##  $ reading7            : num [1:5] 747 701 509 925 718
##  $ reading8            : num [1:5] 800 604 605 789 707
##  $ reading9            : num [1:5] 755 810 532 932 651
##  $ reading10           : num [1:5] 782 797 562 938 685
##  - attr(*, "spec")=
##   .. cols(
##   ..   cubeid = col_character(),
##   ..   Date = col_character(),
##   ..   Time = col_time(format = ""),
##   ..   Probe = col_character(),
##   ..   ‘Probe serial‘ = col_character(),
##   ..   ‘Probe verification‘ = col_character(),
##   ..   Operator = col_character(),
##   ..   ‘Device serial‘ = col_character(),
##   ..   ‘Direction (Equotip)‘ = col_character(),
##   ..   Material = col_character(),
##   ..   ‘Conversion Std.‘ = col_character(),
##   ..   Scale = col_character(),
##   ..   ‘Upper limit‘ = col_double(),
##   ..   ‘Lower limit‘ = col_double(),
##   ..   Readings = col_double(),
##   ..   Mean = col_double(),
##   ..   Min = col_double(),
##   ..   Max = col_double(),
##   ..   Range = col_double(),
##   ..   ‘Std. dev.‘ = col_number(),
##   ..   reading1 = col_double(),
##   ..   reading2 = col_double(),
##   ..   reading3 = col_double(),
##   ..   reading4 = col_double(),
##   ..   reading5 = col_double(),
##   ..   reading6 = col_double(),
##   ..   reading7 = col_double(),
##   ..   reading8 = col_double(),
##   ..   reading9 = col_double(),
##   ..   reading10 = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(volumelossdata)
```

```
## spc_tbl_ [1,280 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ cubeid     : chr [1:1280] "DW-G1-S3" "DW-G1-S3" "DW-G1-S3" "DW-G1-S3" ...
##  $ Material   : chr [1:1280] "basalt" "basalt" "basalt" "basalt" ...
##  $ Cycle      : chr [1:1280] "0-1" "0-1" "0-1" "0-1" ...
##  $ Class      : num [1:1280] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Value      : num [1:1280] 22 30 19 15 25 29 27 14 18 20 ...
##  $ Class start: chr [1:1280] "0.500017583370" "0.501913751010" "0.503809918649" "0.505706086289" ...
##  $ Class end  : chr [1:1280] "0.501913751010" "0.503809918649" "0.505706086289" "0.507602253929" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   cubeid = col_character(),
##   ..   Material = col_character(),
##   ..   Cycle = col_character(),
##   ..   Class = col_double(),
##   ..   Value = col_double(),
##   ..   ‘Class start‘ = col_character(),
##   ..   ‘Class end‘ = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(densitydata)
```

```
## spc_tbl_ [5 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ cubeid  : chr [1:5] "DW-G1-S3" "KW-G1-S3" "MW6-G2-S3" "DW-G4-S3" ...
##  $ Material: chr [1:5] "basalt" "glassyignimbrite" "ignimbrite" "pumaceousignimbrite" ...
##  $ mass    : num [1:5] 55.2 46.8 39.5 36.3 46.6
##  $ volume  : num [1:5] 15.6 15.6 15.6 15.6 15.6
##  $ density : num [1:5] 3.53 3 2.53 2.32 2.98
##  - attr(*, "spec")=
##   .. cols(
##   ..   cubeid = col_character(),
##   ..   Material = col_character(),
##   ..   mass = col_double(),
##   ..   volume = col_double(),
##   ..   density = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Manipulate data

```
# HLC data
hlcdatalong <- hlcdata %>%
  pivot_longer(c(`reading1`, `reading2`,`reading3`, `reading4`, `reading5`, `reading6`, `reading7`, `rea

write_csv(hlcdatalong, "../deriveddata/hlcdatalong.csv")

# Volume loss data (3D)
volumeloss.expanded <- volumelossdata[rep(row.names(volumelossdata), volumelossdata$Value), 1:7]
write_csv(volumeloss.expanded, "../deriveddata/volumelossexpanded.csv")
```

Descriptive stats

```
# function
nminmaxmeanmedsd <- function(x){
    y <- x[!is.na(x)]
    n_test <- length(y)
    min_test <- min(y)
    max_test <- max(y)
    mean_test <- mean(y)
    med_test <- median(y)
    sd_test <- sd(y)
    out <- c(n_test, min_test, max_test, mean_test, med_test, sd_test)
    names(out) <- c("n", "min", "max", "mean", "median", "sd")
    return(out)
}

# computation

# hlc data
num.var <- 22:length(hlcdatalong)
hlc_stats <- summaryBy(.~Material + cubeid,  data=hlcdatalong [c("Material", "cubeid", names(hlcdatalon

# volume loss
num.var2 <- 5:length(volumeloss.expanded)
```

```
volumeloss_stats <- summaryBy(.~Material + Cycle +  cubeid, data=volumeloss.expanded[c("Material", "Cycl

# View results
hlc_stats
```

```
## # A tibble: 5 x 8
##   Material             cubeid  Hardness.n Hardness.min Hardness.max Hardness.mean
##   <chr>                <chr>        <dbl>        <dbl>        <dbl>         <dbl>
## 1 basalt               DW-G1-~         10          743          813          769.
## 2 glassyignimbrite     KW-G1-~         10          789          938          906.
## 3 ignimbrite           MW6-G2~         10          591          741          672.
## 4 pumaceousignimbrite  DW-G4-~         10          456          605          528.
## 5 scoria               DW-G3-~         10          604          834          717.
## # i 2 more variables: Hardness.median <dbl>, Hardness.sd <dbl>
```

```
volumeloss_stats
```

```
## # A tibble: 5 x 9
##   Material      Cycle cubeid Value.n Value.min Value.max Value.mean Value.median
##   <chr>         <chr> <chr>    <dbl>     <dbl>     <dbl>      <dbl>        <dbl>
## 1 basalt        0-1   DW-G1~    1931         1        34      15.3            17
## 2 glassyignimb~ 0-1   KW-G1~     313         1         4       2.09            2
## 3 ignimbrite    0-1   MW6-G~   36162         6       395     209.            181
## 4 pumaceousign~ 0-1   DW-G4~  183482        12      2063    1245.           1542
## 5 scoria        0-1   DW-G3~   11245         1       216      92.8            83
## # i 1 more variable: Value.sd <dbl>
```

```
# save the results
write_csv(hlc_stats, "../stats/stats_hlc.csv")
write_csv(volumeloss_stats, "../stats/stats_volumeloss.csv")
```

Plot hardness data

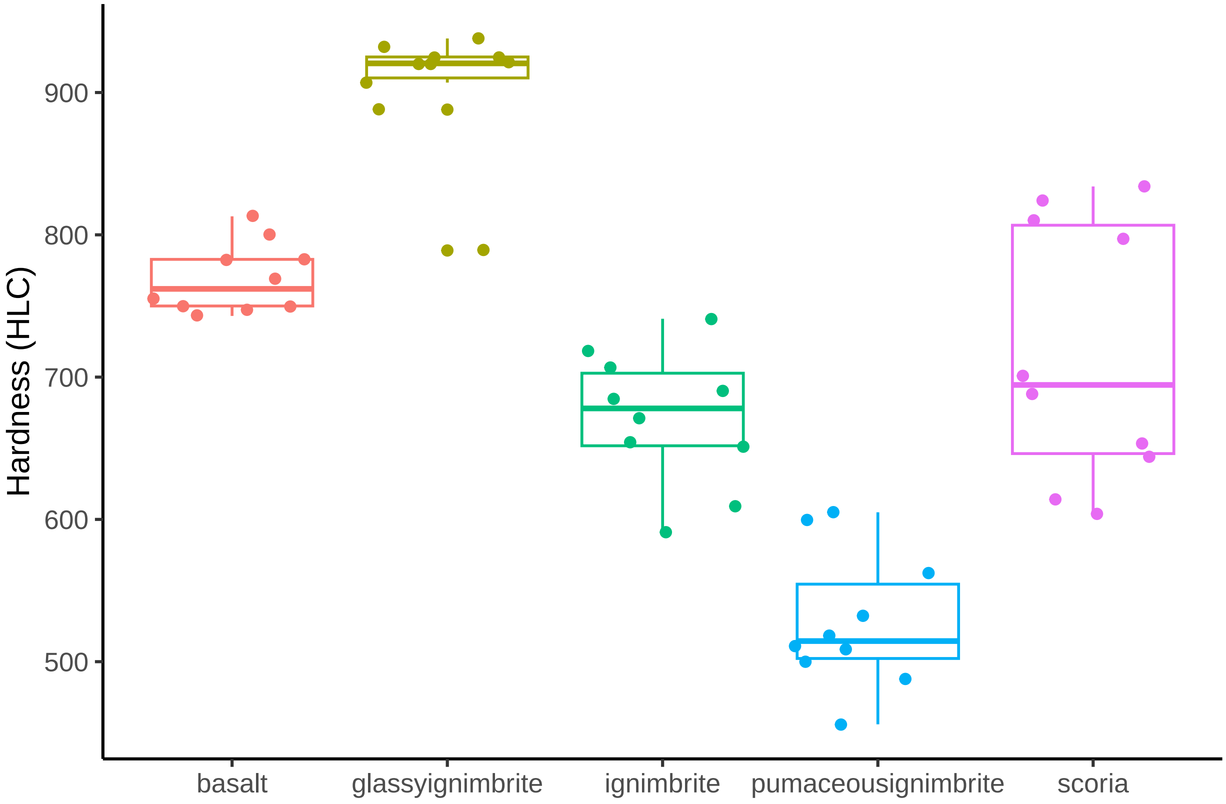
```
hlcplot <- ggplot(hlcdatalong, aes(Material, Hardness, colour = Material)) +
  theme_classic() +
  theme(legend.title = element_blank()) +
  geom_boxplot() +
  geom_jitter() +
  guides(color = FALSE)  +
  labs(y = "Hardness (HLC)", x = "", colour = "Raw Material")
```

```
## Warning: The ‘<scale>‘ argument of ‘guides()‘ cannot be ‘FALSE‘. Use "none" instead as
## of ggplot2 3.3.4.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```r
ggsave("../plots/hlcdata.png")
```

```
## Saving 6.5 x 4.5 in image
```

```r
print(hlcplot)
```

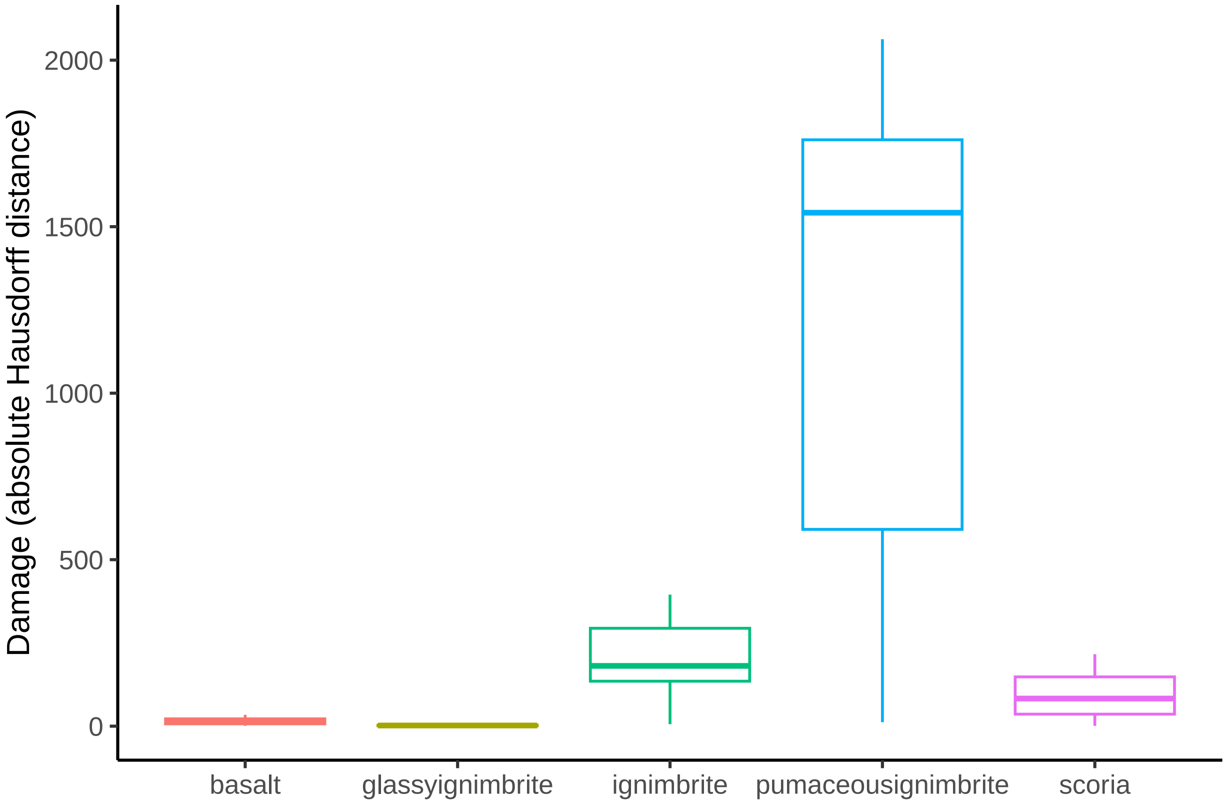


Plot volume loss data, organised by raw material and sample

```r
# Plot data

volumeplot <- ggplot(volumeloss.expanded, aes(Material, Value, colour = Material)) +
  theme_classic() +
  theme(legend.title = element_blank()) +
  geom_boxplot() +
  guides(color = FALSE)  +
  labs(y = "Damage (absolute Hausdorff distance)", x = "", colour = "Raw Material")

ggsave("../plots/damage_rawmaterial.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
print(volumeplot)
```

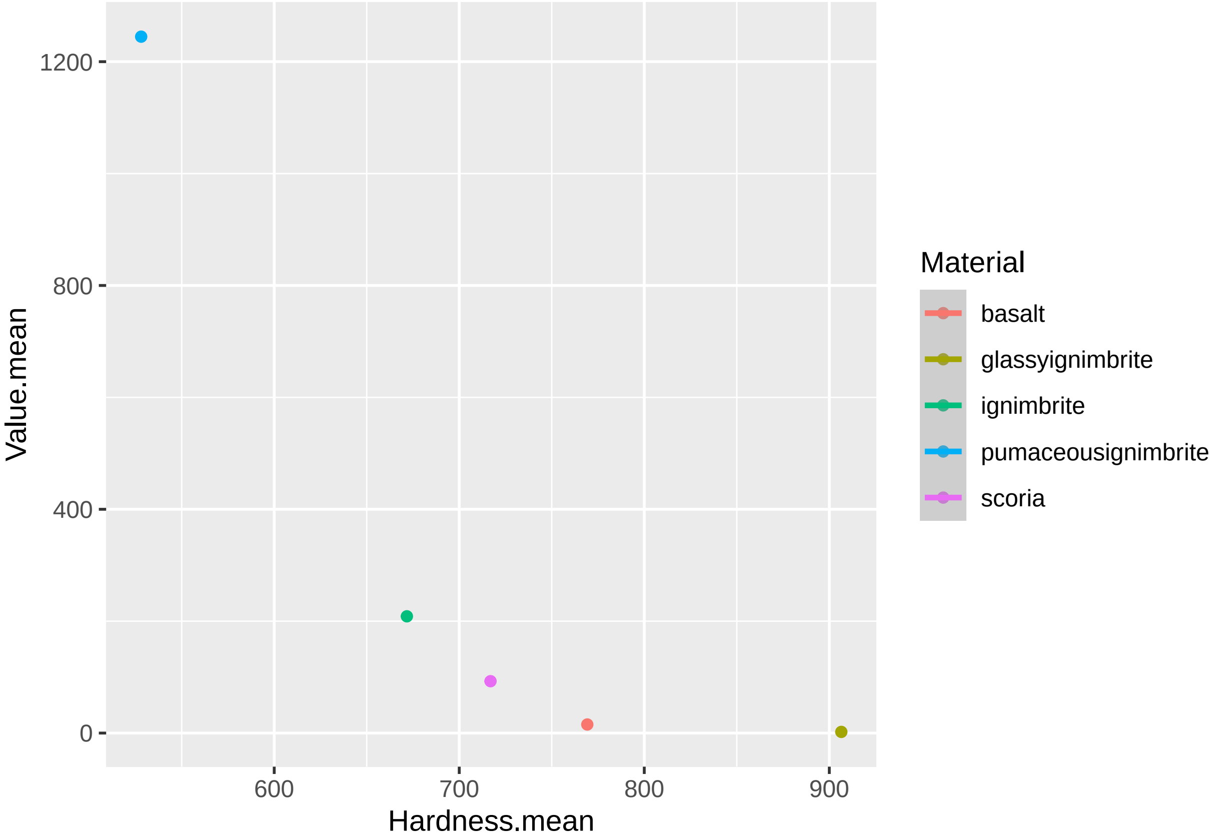


```
# do scatterplot XXXXXX
```

Correlation between hardness, density, and volume loss

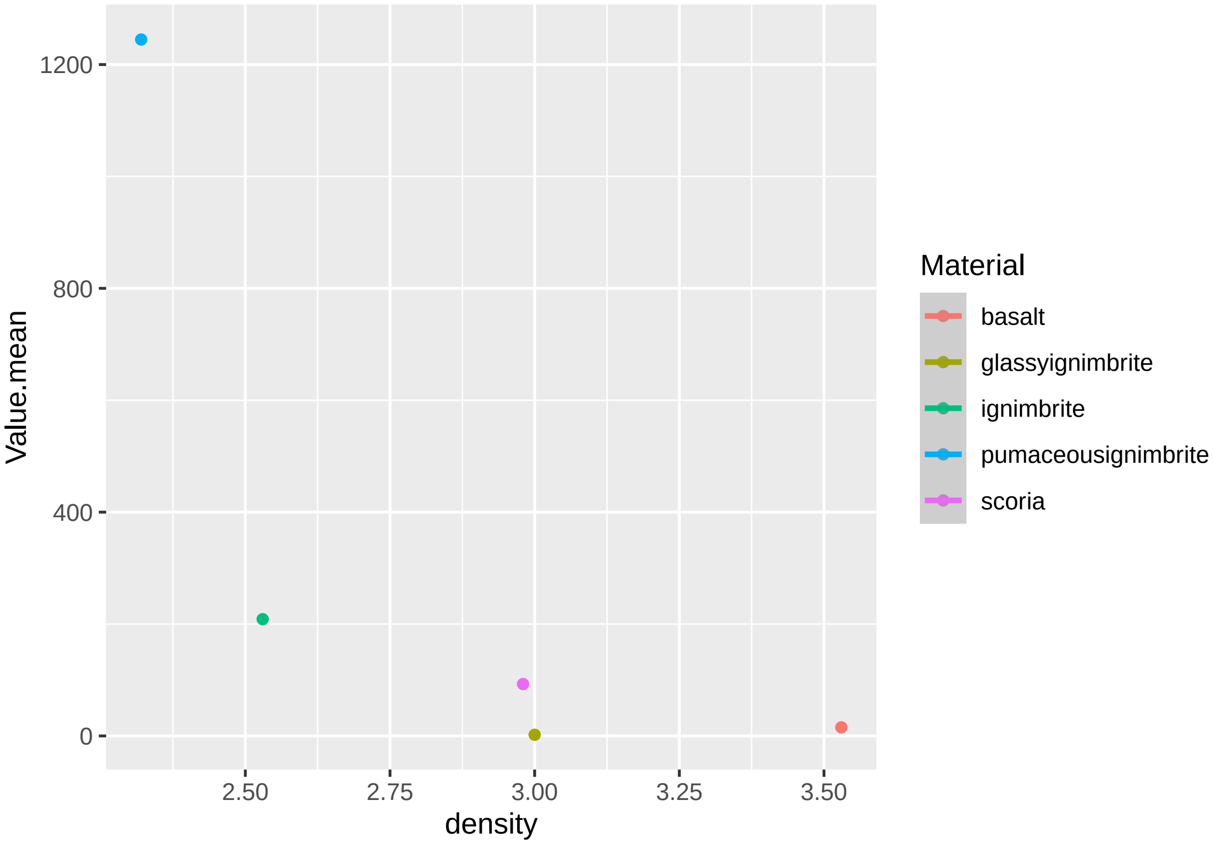
```
# join data tables

corr_data <- full_join(hlc_stats, densitydata,
              by = join_by("cubeid", "Material"))

corr_data2 <- full_join(corr_data, volumeloss_stats,
              by = join_by("cubeid", "Material"))

# save data
write_csv(corr_data2, "../deriveddata/full_data.csv")

# scatter plot to check correlation between variables (dependent and independent)

scat <- ggplot(corr_data2,aes(Hardness.mean, Value.mean, color = Material)) +
  geom_point() +
  geom_smooth(method='lm')

print(scat)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



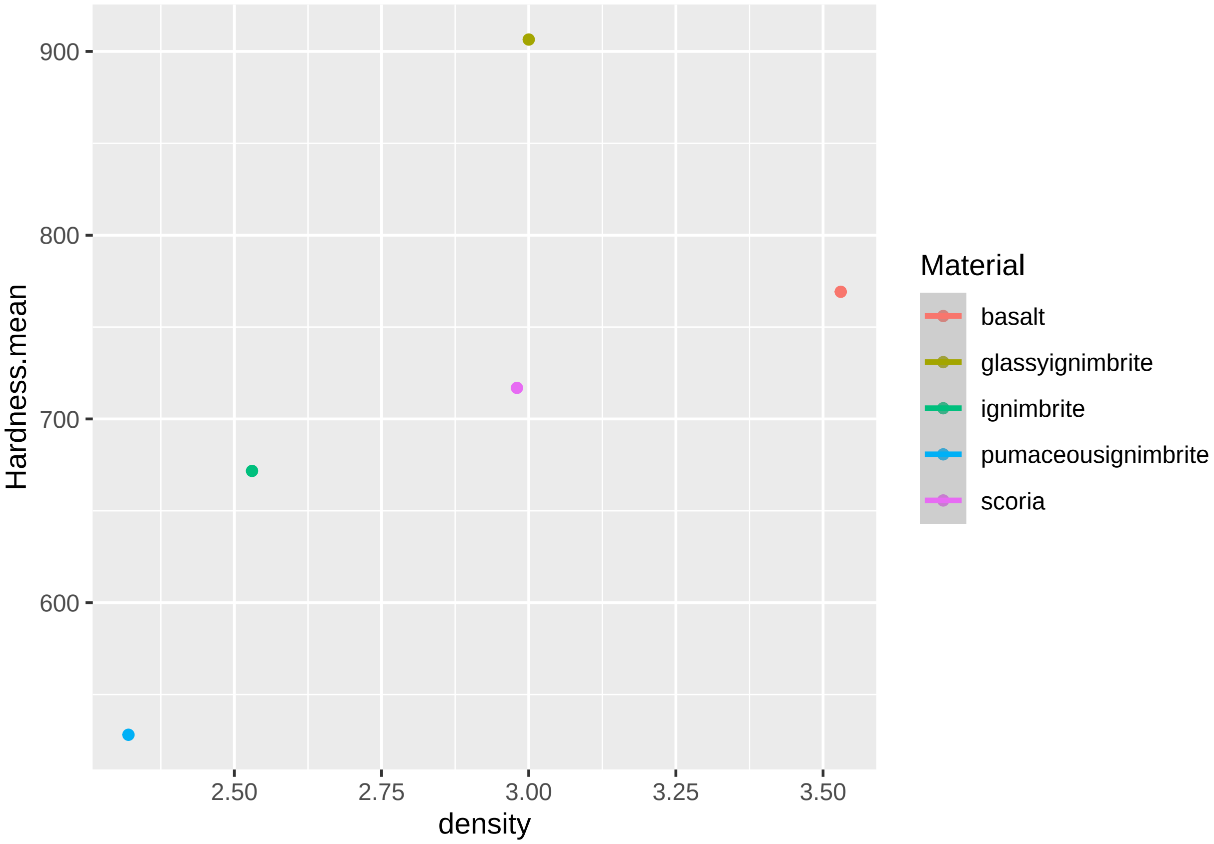
```
scat2 <- ggplot(corr_data2,aes(density, Value.mean, color = Material)) +
  geom_point() +
  geom_smooth(method='lm')

print(scat2)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```r
scat3 <- ggplot(corr_data2,aes(density, Hardness.mean, color = Material)) +
  geom_point() +
  geom_smooth(method='lm')

print(scat3)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Bootstrapping

Bootstrapped Mann-Whitney U p-value (95% confidence interval) for hardness by raw material.

```
##  2.5% 97.5%
## 0.406 0.406
```

Bootstrapped Mann-Whitney U p-value (95% confidence interval) for density by raw material.

```
##  2.5% 97.5%
## 0.406 0.406
```

Bootstrapped Mann-Whitney U p-value (95% confidence interval) for volume loss (damage) by raw material.

```
##  2.5% 97.5%
## 0.317 0.317
```

Kendall rank correlation between hardness and volume loss

```
##
```

```
##  Kendall’s rank correlation tau
##
## data:  mean_hardness$HLC and mean_hardness$‘Volume loss‘
## T = 0, p-value = 0.01667
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
##  -1
```

Kendall rank correlation between density and volume loss

```
##
##  Kendall’s rank correlation tau
##
## data:  mean_density$Density and mean_density$‘Volume loss‘
## T = 1, p-value = 0.08333
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##  tau
## -0.8
```

---

sessionInfo() and RStudio version

```r
sessionInfo()
```

```
## R version 4.3.0 (2023-04-21)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS 14.2.1
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK ve
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: Europe/Berlin
## tzcode source: internal
##
## attached base packages:
## [1] tools     stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] ggtern_3.4.2      flextable_0.9.4   ggrepel_0.9.5     doBy_4.6.20
##  [5] lubridate_1.9.3   forcats_1.0.0     stringr_1.5.1     dplyr_1.1.4
##  [9] purrr_1.0.2       readr_2.1.5       tidyr_1.3.0       tibble_3.2.1
## [13] tidyverse_2.0.0   ggplot2_3.4.4     R.utils_2.12.3    R.oo_1.25.0
## [17] R.methodsS3_1.8.2
##
## loaded via a namespace (and not attached):
##  [1] gridExtra_2.3          rlang_1.1.3            magrittr_2.0.3
##  [4] compiler_4.3.0         systemfonts_1.0.5      vctrs_0.6.5
##  [7] httpcode_0.3.0         pkgconfig_2.0.3        crayon_1.5.2
## [10] fastmap_1.1.1          backports_1.4.1        ellipsis_0.3.2
## [13] labeling_0.4.3         utf8_1.2.4             promises_1.2.1
## [16] rmarkdown_2.25         tzdb_0.4.0             ragg_1.2.7
## [19] bit_4.0.5              xfun_0.41              jsonlite_1.8.8
## [22] highr_0.10             later_1.3.2            uuid_1.2-0
## [25] Deriv_4.1.3            broom_1.0.5            parallel_4.3.0
## [28] R6_2.5.1               stringi_1.8.3          compositions_2.0-7
## [31] Rcpp_1.0.12            knitr_1.45             httpuv_1.6.13
## [34] Matrix_1.6-5           timechange_0.2.0       tidyselect_1.2.0
## [37] rstudioapi_0.15.0      yaml_2.3.8             curl_5.2.0
## [40] lattice_0.22-5         plyr_1.8.9             shiny_1.8.0
## [43] withr_3.0.0            askpass_1.2.0          evaluate_0.23
## [46] bayesm_3.1-6           zip_2.3.0              xml2_1.3.6
## [49] pillar_1.9.0           tensorA_0.36.2.1       generics_0.1.3
## [52] vroom_1.6.5            hms_1.1.3              munsell_0.5.0
## [55] scales_1.3.0           xtable_1.8-4           glue_1.7.0
## [58] gdtools_0.3.5          gfonts_0.2.0           hexbin_1.28.3
## [61] robustbase_0.99-1      data.table_1.14.10     grid_4.3.0
## [64] colorspace_2.1-0       proto_1.0.0            cli_3.6.2
## [67] latex2exp_0.9.6        textshaping_0.3.7      officer_0.6.3
## [70] fontBitstreamVera_0.1.1 fansi_1.0.6           gtable_0.3.4
## [73] DEoptimR_1.1-3         digest_0.6.34          fontquiver_0.2.1
## [76] crul_1.4.0             farver_2.1.1           htmltools_0.5.7
## [79] lifecycle_1.0.4        mime_0.12              fontLiberation_0.1.0
## [82] microbenchmark_1.4.10  openssl_2.1.1          bit64_4.0.5
## [85] MASS_7.3-60.0.1
```

END OF SCRIPT