Plots - tool function experiment

Lisa Schunk

2021-02-04 15:05:12

# Goal of the script

This script plots all variables to see which ones should be used for further analysis.  
Scatterplot of each variable will be plotted.

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

Raw data must be located in ~/analysis/derived\_data/.  
Formatted data will be saved in ~/analysis/plots.

## The knit directory for this script is the project directory.

# Load packages

library(R.utils)  
library(ggplot2)

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

library(tools)  
library(tidyverse)

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

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

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

library(ggfortify)

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

# Get name, path and information of the file

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

The checksum (MD5 hashes) of the imported file is:

file checksum  
1 TFE\_use-wear.Rbin 433a3db62e03eee5450b37da5616b1a7

# Load data into R object

imp\_data <- loadObject(data\_file)

The imported file is: “~/analysis/derived\_data/TFE\_use-wear.Rbin”

# Prepare variables

## Define numeric variables

num.var <- 24:length(imp\_data)

The following variables will be used:

[24] Sq  
[25] Ssk  
[26] Sku  
[27] Sp  
[28] Sv  
[29] Sz  
[30] Sa  
[31] Smr  
[32] Smc  
[33] Sxp  
[34] Sal  
[35] Str  
[36] Std  
[37] Sdq  
[38] Sdr  
[39] Vm  
[40] Vv  
[41] Vmp  
[42] Vmc  
[43] Vvc  
[44] Vvv  
[45] Maximum.depth.of.furrows  
[46] Mean.depth.of.furrows  
[47] Mean.density.of.furrows  
[48] First.direction  
[49] Second.direction  
[50] Third.direction  
[51] Isotropy  
[52] epLsar  
[53] NewEplsar  
[54] Asfc  
[55] Smfc  
[56] HAsfc9  
[57] HAsfc81

# Plot each of the selected numeric variable

## (facet plot = 1 plot for flint, 1 plot for lydite)

# BottleRocket2   
custom.col5 <- data.frame(type = levels(imp\_data$Edge.angle), col = c("#046C9A", "#FAD510")) imp\_data$col <- custom.col5[imp\_data$Edge.angle, "col"]  
  
  
for (i in num.var){  
 #plot  
 range\_var <- range(imp\_data[[i]]) # gets the min/max range of the data set   
  
 p <- ggplot(data=imp\_data, aes\_string(x = "Task", y=names(imp\_data)[i],   
 colour = "Edge.angle")) +   
 # avoids overplotting  
 geom\_jitter(size = 3, position = position\_jitter(width = 0.4, seed = 1)) +   
 coord\_cartesian(ylim = range\_var) +   
 theme\_classic() +   
 scale\_colour\_manual(values = custom.col5$col) +  
 # removes the "." between "Edge.angle" in the legend   
 labs(colour = gsub("\\.", " ", "Edge.angle")) +   
 # removes the "." between the variable names  
 labs(y = gsub("\\.", " ", names(imp\_data)[i])) +   
 # combines the flint and the lydite plot into one   
 facet\_wrap(~Raw.material)   
 print(p)  
  
  
 #save to PDF  
 file\_out <- paste0(file\_path\_sans\_ext(info\_in[["file"]]), "\_plot\_",   
 names(imp\_data)[i], ".pdf")  
 ggsave(filename = file\_out, plot = p, path = dir\_out, device = "pdf")  
}

Error: <text>:2:94: unexpected symbol  
1: # BottleRocket2   
2: custom.col5 <- data.frame(type = levels(imp\_data$Edge.angle), col = c("#046C9A", "#FAD510")) imp\_data  
 ^

## Principal component analysis (without outliers)

# excludes the outliers   
# adds the indices as row numbers   
imp\_data <- imp\_data %>% mutate(id = row\_number())  
imp\_data2 <- imp\_data[-c(2, 10), ]  
  
# removes the rows with missing values   
data\_pca <- na.omit(imp\_data2)   
imp\_data.pca <- prcomp(data\_pca[, c(24:25, 42, 47, 51, 54, 56)], scale. = TRUE)   
  
# FantasticFox1   
custom.col4 <- data.frame(type = levels(data\_pca$Task), col = c("#E58601", "#B40F20"))   
data\_pca$col <- custom.col4[data\_pca$Task, "col"]  
  
  
# plots the task - PCA with convex hull   
# Using ggfortify  
a<- autoplot(imp\_data.pca, data = data\_pca, colour = "Task", size = 2,  
 loadings = TRUE, loadings.colour = "black", loadings.label = TRUE,  
 loadings.label.colour = "black",   
 loadings.label.size = 4, loadings.label.hjust = 1, loadings.label.vjust = 1,   
 frame = TRUE, frame.type = "convex", frame.colour = "Task", frame.alpha = 0) +   
 theme\_classic() +  
 scale\_colour\_manual(values = custom.col4$col)  
   
print(a)  
  
# saves the plot  
file\_out <- paste0(file\_path\_sans\_ext(info\_in[["file"]]), "TFE\_PCA\_Task", ".pdf")  
ggsave(filename = file\_out, plot = a, path = dir\_out, device = "pdf")  
  
  
  
# BottleRocket2   
custom.col5 <- data.frame(type = levels(data\_pca$Edge.angle), col = c("#046C9A", "#FAD510")) data\_pca$col <- custom.col5[data\_pca$Edge.angle, "col"]  
  
  
# plots the edge angle - PCA with convex hull   
b<- autoplot(imp\_data.pca, data = data\_pca, colour = "Edge.angle", size = 2,  
 loadings = TRUE, loadings.colour = "black", loadings.label = TRUE,  
 loadings.label.colour = "black",   
 loadings.label.size = 4, loadings.label.hjust = 1, loadings.label.vjust = 1,   
 frame = TRUE, frame.type = "convex", frame.colour = "Edge.angle",   
 frame.alpha = 0) +   
 theme\_classic() +  
 scale\_colour\_manual(values = custom.col5$col)  
   
print(b)  
  
# saves the plot  
file\_out <- paste0(file\_path\_sans\_ext(info\_in[["file"]]), "TFE\_PCA\_angle", ".pdf")  
ggsave(filename = file\_out, plot = b, path = dir\_out, device = "pdf")  
  
  
  
# Royal1   
custom.col7 <- data.frame(type = levels(data\_pca$Raw.material), col = c("#899DA4", "#DC863B"))   
data\_pca$col <- custom.col7[data\_pca$Raw.material, "col"]  
  
  
# plots the raw material - PCA with convex hull   
c<- autoplot(imp\_data.pca, data = data\_pca, colour = "Raw.material", size = 2,  
 loadings = TRUE, loadings.colour = "black", loadings.label = TRUE,  
 loadings.label.colour = "black",   
 loadings.label.size = 4, loadings.label.hjust = 1, loadings.label.vjust = 1,   
 frame = TRUE, frame.type = "convex", frame.colour = "Raw.material",   
 frame.alpha = 0) +   
 theme\_classic() +  
 scale\_colour\_manual(values = custom.col7$col)  
   
print(c)  
  
# saves the plot  
file\_out <- paste0(file\_path\_sans\_ext(info\_in[["file"]]), "TFE\_PCA\_raw.material", ".pdf")  
ggsave(filename = file\_out, plot = c, path = dir\_out, device = "pdf")

Error: <text>:34:94: unexpected symbol  
33: # BottleRocket2   
34: custom.col5 <- data.frame(type = levels(data\_pca$Edge.angle), col = c("#046C9A", "#FAD510")) data\_pca  
 ^

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] ggfortify\_0.4.11 forcats\_0.5.1 stringr\_1.4.0 dplyr\_1.0.3   
 [5] purrr\_0.3.4 readr\_1.4.0 tidyr\_1.1.2 tibble\_3.0.6   
 [9] tidyverse\_1.3.0 ggplot2\_3.3.3 R.utils\_2.10.1 R.oo\_1.24.0   
[13] R.methodsS3\_1.8.1  
  
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 readxl\_1.3.1   
[17] lifecycle\_0.2.0 munsell\_0.5.0 gtable\_0.3.0 cellranger\_1.1.0   
[21] rvest\_0.3.6 evaluate\_0.14 knitr\_1.31 broom\_0.7.4   
[25] Rcpp\_1.0.6 scales\_1.1.1 backports\_1.2.1 jsonlite\_1.7.2   
[29] fs\_1.5.0 gridExtra\_2.3 hms\_1.0.0 digest\_0.6.27   
[33] stringi\_1.5.3 grid\_4.0.2 cli\_2.3.0 magrittr\_2.0.1   
[37] crayon\_1.4.0 pkgconfig\_2.0.3 ellipsis\_0.3.1 xml2\_1.3.2   
[41] reprex\_1.0.0 lubridate\_1.7.9.2 assertthat\_0.2.1 rmarkdown\_2.6   
[45] httr\_1.4.2 rstudioapi\_0.13 R6\_2.5.0 compiler\_4.0.2

RStudio version 1.3.1073.

END OF SCRIPT