Projekt

Julia Janczyk, Julia Filewicz

2024-12-10

```
library(cluster)
library(readr)
```

```
## Warning: pakiet 'readr' został zbudowany w wersji R 4.3.2
```

```
library(ggplot2)
```

```
## Warning: pakiet 'ggplot2' został zbudowany w wersji R 4.3.2
```

```
library(dplyr)
```

```
## Warning: pakiet 'dplyr' został zbudowany w wersji R 4.3.2
```

```
##
## Dołączanie pakietu: 'dplyr'
```

```
## Następujące obiekty zostały zakryte z 'package:stats':
##
##     filter, lag
```

```
## Następujące obiekty zostały zakryte z 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
setwd("C:/Users/jula7/Desktop/ja/Mod danych prze/projekt/dane")
```

Odfiltrowanie ramki danych

```
data <- read.csv("janczyk_output_0.txt", sep = ";", header = TRUE)
nrow(data)
```

```
## [1] 29992
```

```
# oczyszczamy obserwacje z "x"
data_filtered <- data[data$Dip_dir != "x", ]
nrow(data_filtered)
```

```
## [1] 29388
```

```r
# wartości na liczbowe
data_filtered$Dip_dir <- as.numeric(as.character(data_filtered$Dip_dir))

write.table(data_filtered, "janczyk_output_filtered.txt", sep = ";", row.names = FALSE, quote = FALSE)
```

Oczyszczenie z błędów na brzegach

```r
data = read.csv("janczyk_output_filtered.txt", sep=";")
filtered = dplyr::filter(data, X_C >= min(data$X_C) + 0.05*(max(data$X_C) - min(data$X_C)))

filtered = dplyr::filter(filtered, X_C <= max(data$X_C) - 0.05*(max(data$X_C) - min(data$X_C)))

filtered = dplyr::filter(filtered, Y_C >= min(data$Y_C) + 0.05*(max(data$Y_C) - min(data$Y_C)))

filtered = dplyr::filter(filtered, Y_C <= max(data$Y_C) - 0.05*(max(data$Y_C) - min(data$Y_C)))

write.table(x=filtered, file = "janczyk_output_filtered.txt", sep = ";", row.names = FALSE)
```

Grupownie k - średnich

```r
dataN <- read.csv("janczyk_output_filtered.txt", sep =";", header=TRUE)

k <- 3
grupowanieN <- kmeans(
  as.matrix(dplyr::select(dataN, c("X_N", "Y_N", "Z_N"))),
  centers = k,
  nstart = 1,
  iter.max = 100000,
  algorithm = "Lloyd"
)

dataN$Cluster <- grupowanieN$cluster
write_delim(dataN, "janczyk_output_N.txt", delim = ";")
```

Normals

Zapisywanie klastrów dla Stereonet N

```r
df1 <- dplyr::filter(dataN, Cluster == 1)
df2 <- dplyr::filter(dataN, Cluster == 2)
df3 <- dplyr::filter(dataN, Cluster == 3)

write_delim(df1, "janczyk_output_N_cluster1.txt", delim = ",")
write_delim(df2, "janczyk_output_N_cluster2.txt", delim = ",")
write_delim(df3, "janczyk_output_N_cluster3.txt", delim = ",")
```

```
dataD <- read.csv("janczyk_output_filtered.txt", sep =";", header=TRUE)

k <- 3
grupowanieD <- kmeans(
  as.matrix(dplyr::select(dataD, c("X_D", "Y_D", "Z_D"))),
  centers = k,
  nstart = 1,
  iter.max = 100000,
  algorithm = "Lloyd"
)

dataD$Cluster <- grupowanieD$cluster
write_delim(dataD, "janczyk_output_D.txt", delim = ";")
```

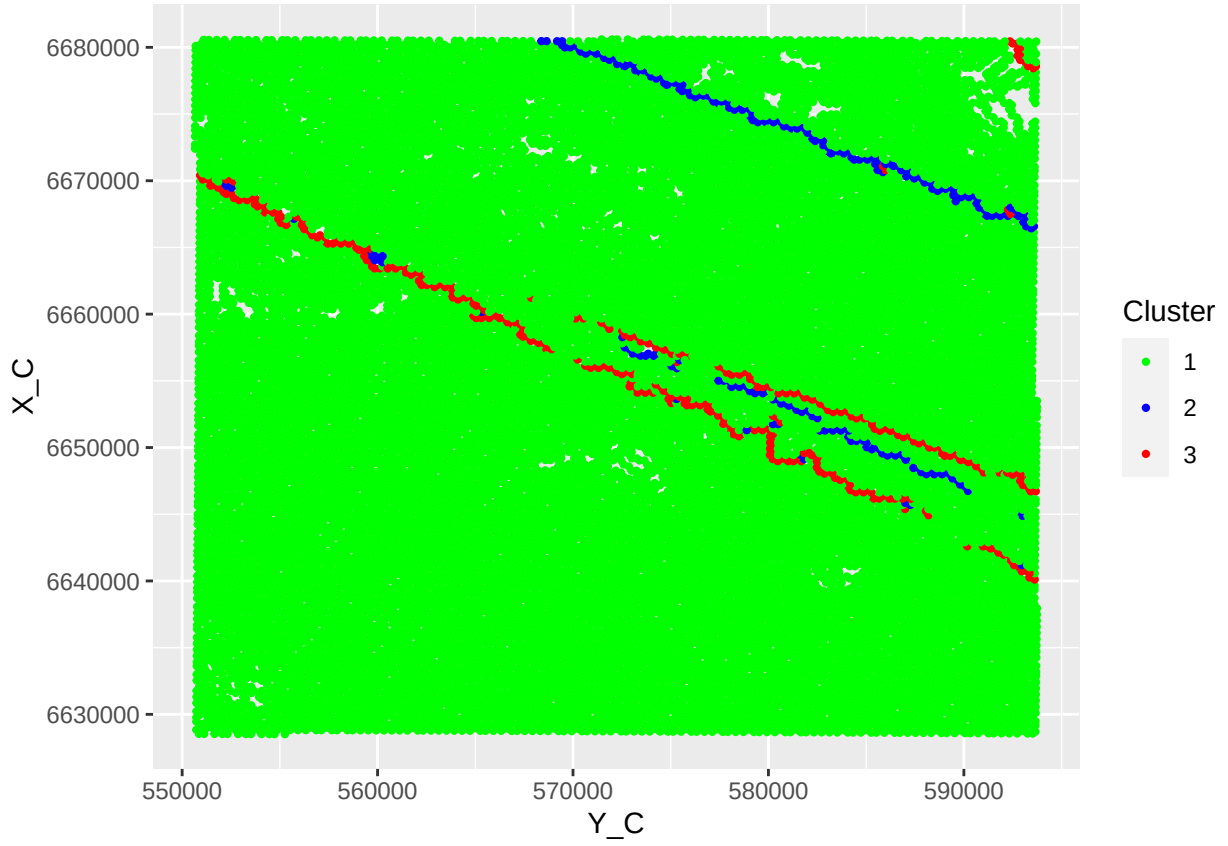
Dips

Zapisywanie klastrów dla Stereonet D

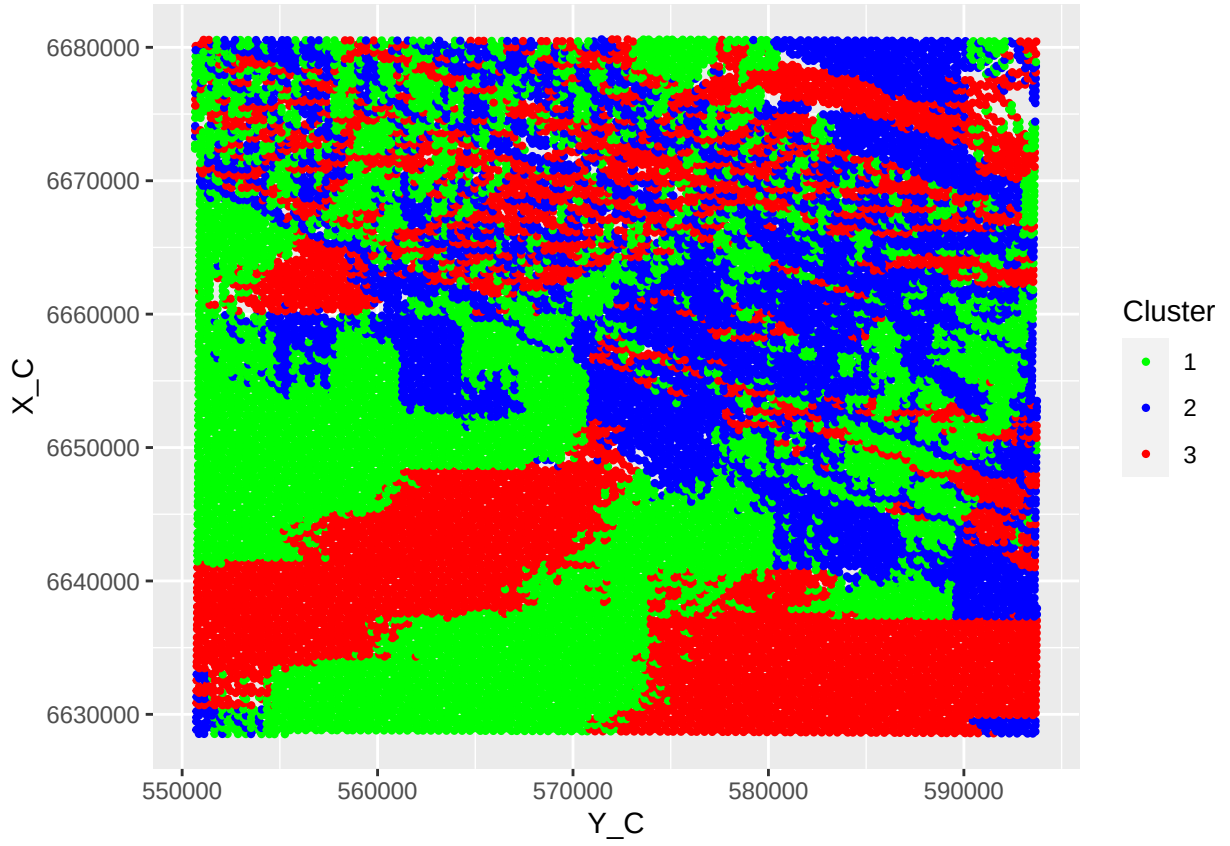
```
df1 <- dplyr::filter(dataD, Cluster == 1)
df2 <- dplyr::filter(dataD, Cluster == 2)
df3 <- dplyr::filter(dataD, Cluster == 3)

write_delim(df1, "janczyk_output_D_cluster1.txt", delim = ",")
write_delim(df2, "janczyk_output_D_cluster2.txt", delim = ",")
write_delim(df3, "janczyk_output_D_cluster3.txt", delim = ",")
```

Mapy grupowania trójkątów

```
grid_data <- read.csv("janczyk_grid_locate.txt", sep =";", header=TRUE)

triangles_N <- read.csv("janczyk_output_N.txt", sep =";", header=TRUE)
triangles_D <- read.csv("janczyk_output_D.txt", sep =";", header=TRUE)

gmergedN <- merge(x = grid_data, y = triangles_N, by = c("IDT1", "IDT2", "IDT3"), all.y = TRUE)
gmergedD <- merge(x = grid_data, y = triangles_D, by = c("IDT1", "IDT2", "IDT3"), all.y = TRUE)

palkmeans3<- c("green", "blue","red" )

ggplot(gmergedN, aes(x=Y_C, y=X_C, col=factor(Cluster) ))+
  geom_point(size=0.8)+ scale_color_manual("Cluster", values = palkmeans3)
```




```
ggplot(gmergedD, aes(x=Y_C, y=X_C, col=factor(Cluster) ))+
  geom_point(size=0.8)+ scale_color_manual("Cluster", values = palkmeans3)
```



Środki skupień

```
# f do obliczenia Dip Direction i Dip Angle
calculate <- function(x, y, z) {
  norm_vec <- c(x, y, z) / sqrt(x^2 + y^2 + z^2)
  dip_ang <- acos(norm_vec[3]) * (180 / pi) - 90
  while (dip_ang < 0) { dip_ang <- dip_ang + 90 }
  dip_dir <- atan2(norm_vec[2], norm_vec[1]) * (180 / pi)
  while (dip_dir > 360) { dip_dir <- dip_dir - 360 }
  while (dip_dir < 0) { dip_dir <- dip_dir + 360 }
  return(c(dip_ang, dip_dir))
}

dataN <- read.csv("janczyk_output_N.txt", sep = ";", header = TRUE)
dataD <- read.csv("janczyk_output_D.txt", sep = ";", header = TRUE)
data_norm <- data.frame()
data_dip <- data.frame()

for (x in 1:3) {
  # aktualny  klaster
  cluster_N <- dataN[dataN$Cluster == x, c("X_N", "Y_N", "Z_N")]
  cluster_D <- dataD[dataD$Cluster == x, c("X_D", "Y_D", "Z_D")]

  center_N <- colMeans(cluster_N)
```

```
  center_D <- colMeans(cluster_D)
  norm_result <- calculate(center_N[1], center_N[2], center_N[3])
  dip_result <- calculate(center_D[1], center_D[2], center_D[3])

  data_norm <- rbind(data_norm, norm_result)
  data_dip <- rbind(data_dip, dip_result)
}
colnames(data_dip) <- c("Dip_ang", "Dip_dir")
colnames(data_norm) <- c("Dip_ang", "Dip_dir")

write_delim(data_norm, "srodki_skupien_N.txt", delim = ",")
write_delim(data_dip, "srodki_skupien_D.txt", delim = ",")
```