

```

#download relevante packages, installeer en activeer ze:
packages <- c("xlsx", "cluster")
for(i in 1 : length(packages)){
  if (!require(packages[i], character.only = TRUE)) install.packages(packages[i])
  library(packages[i], character.only = TRUE)
}

#Haal gegevens eerste week binnen:
download.file(url = "https://surfdrive.surf.nl/files/index.php/s/48oXoxszqmsNlo3/download", destfile = "Toets1.xlsx", mode = "wb")
#Merk op dat de studentnamen en nummers gefingeerd zijn
bestand <- read.xlsx(file = "Toets1.xlsx", sheetIndex = 1)

#Stel datum begin toets vast (handig voor rekenen met datums).
start1 <- "2017-09-05 18:00:00 UTC"

#Verwerk data en kijk naar correlaties
input <- bestand[, c("Grade", "Date", "Duration")]
input$Datex <- as.numeric(difftime(as.POSIXct(bestand$Date), as.POSIXct(start1), units = "days"))
cor(input[, c("Grade", "Datex", "Duration")], use = "p")

#DOE CLUSTERANALYSE
set.seed(123)
#Gebruik drie clusters
nclus <- 3
clara.1 <- clara(input, nclus, stand = TRUE, correct.d = TRUE)
plot(clara.1)

#Kijk naar de gemiddelden op de drie variabelen in elke cluster
by(input[, c("Grade", "Date", "Duration")], clara.1$cluster, summary)
#Kijk naar de correlaties tussen de drie variabelen binnen elk cluster
by(input[, c("Grade", "Datex", "Duration")], clara.1$cluster, cor, use = "p")
#Aantal studenten in elke cluster
table(clara.1$cluster)

#Maak plaatje van clusters; x-as is dag in de week dat toets is gemaakt
#y-as is cijfer op de toets, en grootte van kruisje geeft de benodigde tijd weer (groter = meer tijd)
plot(bestand$Date, Bestand$Grade, col=clara.1$cluster, cex=bestand$Duration/8, pch = 3)
legend("bottomleft", legend = 1:nclus, col = 1:nclus, pch = 3)

```