Code

* [Show All Code](#gjdgxs)
* [Hide All Code](#gjdgxs)
* [Download Rmd](#gjdgxs)

# Oefeningen week 5

#### ***Jelle Becirspahic***

Opdracht 1

*Download de meest recente lijst met aardbevingen en geef de 5 plaatsen met de zwaarste aardbevingen.*

all\_induced <- read.csv("all\_induced.csv")  
highest\_values <- order(all\_induced$MAG, decreasing = T)[1:5]  
all\_induced$LOCATION[highest\_values]

[1] Huizinge Bergen Westeremden  
[4] Roswinkel Zeerijp   
180 Levels: 't Zandt Alkmaar Amen ... Zuidwolde

Opdracht 2

*In welke tien plaatsen hebben zich de meeste opgewekte aardbevingen voorgedaan met een sterkte van meer dan 1.5?*

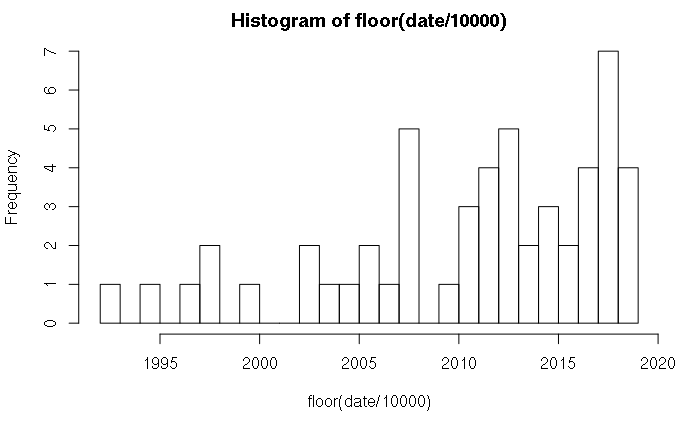
waarde <- order(all\_induced$MAG, decreasing = T)[sort(all\_induced$MAG, decreasing = T) > 1.5]  
names(sort(table(all\_induced$LOCATION[waarde]), decreasing = T)[1:10])

[1] "Roswinkel" "Zeerijp" "Appingedam" "Wirdum"   
 [5] "Garrelsweer" "Westeremden" "Noordzee" "Emmen"   
 [9] "Loppersum" "Middelstum"

Opdracht 3

*Is er in de gemeente Loppersum een trend in de frequentie van het optreden van aardbevingen zichtbaar?*

indices <- which(all\_induced$LOCATION == "Loppersum")  
date <- all\_induced$YYMMDD[indices]  
hist(floor(date/10000), breaks = 35)



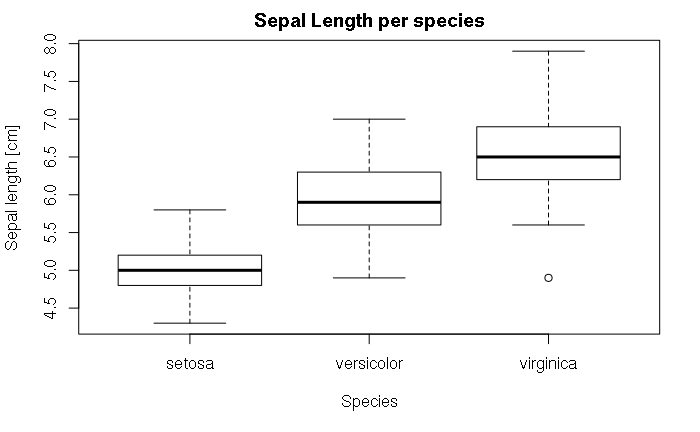
Opdracht 4

*De dataset bevat regels met metingen aan verschillende soorten irissen. Gemeten zijn de lengte en de breedte van de kelk- en de bloemblaadjes. Welke van deze gegevens kun je gebruiken om een iris die je in de tuin hebt te determineren? Maak een figuur c.q. de figuren die je hiervoor nodig hebt.*

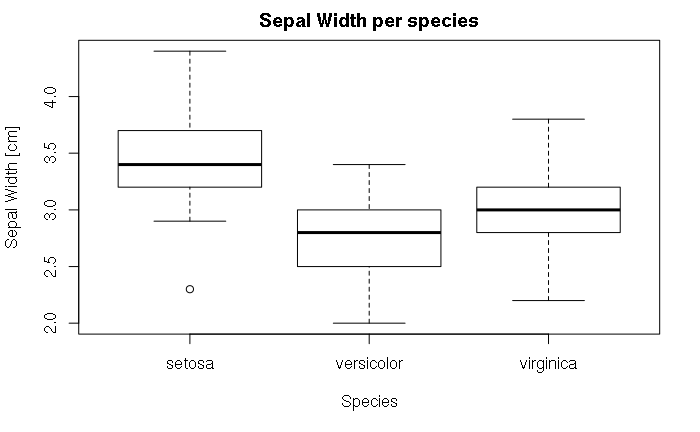
summary(iris)

Sepal.Length Sepal.Width Petal.Length Petal.Width   
 Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100   
 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300   
 Median :5.800 Median :3.000 Median :4.350 Median :1.300   
 Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
 Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500   
 Species   
 setosa :50   
 versicolor:50   
 virginica :50

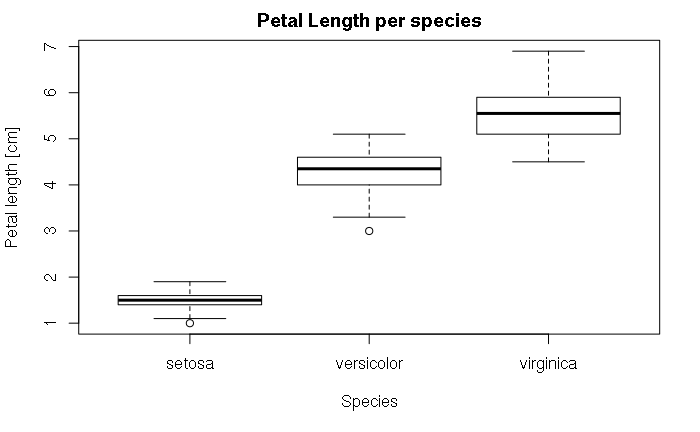
plot(iris$Species, iris$Sepal.Length, main="Sepal Length per species", ylab="Sepal length [cm]", xlab = "Species")



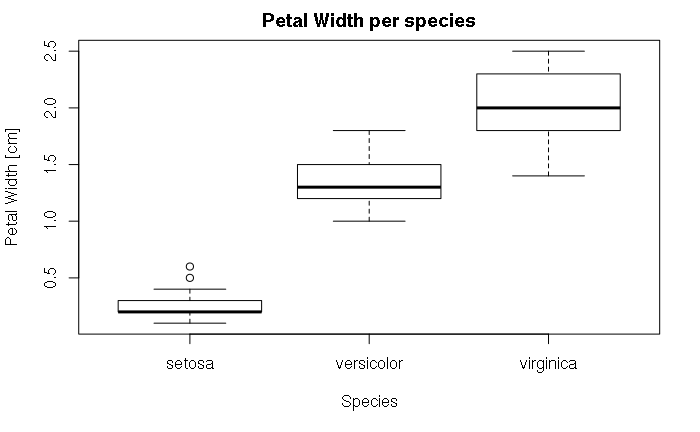
plot(iris$Species, iris$Sepal.Width, main="Sepal Width per species", ylab="Sepal Width [cm]", xlab = "Species")



plot(iris$Species, iris$Petal.Length, main="Petal Length per species", ylab="Petal length [cm]", xlab = "Species")



plot(iris$Species, iris$Petal.Width, main="Petal Width per species", ylab="Petal Width [cm]", xlab = "Species")



Bij de blad lengte is het grootste verschil te zien tussen de verschillende soorten. Hier zou dus het makkelijkst onderscheid in te maken zijn.

Opdracht 5

*Geef in een codeblok de R-code waarmee je een .xls in kunt lezen in R*

library("readxl")  
my.data <- read\_excel("my\_file.xls")

Error: `path` does not exist: ‘my\_file.xls’

Opdracht 6

*Je kunt een .xls ook converteren naar het .csv formaat. Leg uit (code en tekst) hoe je dit m.b.v. Python doet.*

import xlrd  
import csv  
  
 def csv\_from\_excel():  
 wb = xlrd.open\_workbook('your\_workbook.xls')  
 sh = wb.sheet\_by\_name('Sheet1')  
 your\_csv\_file = open('your\_csv\_file.csv', 'wb')  
 wr = csv.writer(your\_csv\_file, quoting=csv.QUOTE\_ALL)  
  
 for rownum in xrange(sh.nrows):  
 wr.writerow(sh.row\_values(rownum))  
  
 your\_csv\_file.close()

Opdracht 7

*Je kunt een .xls ook converteren naar het .csv formaat. Leg uit (code) hoe je dit m.b.v. LibreOffice kunt doen.*

libreoffice –headless –convert-to csv “my\_file.xls”

Opdracht 8

*Pas het commando uit opdracht 7 zo aan dat je alle .xls bestanden in één directory tegelijk converteert.*

libreoffice –headless –convert-to csv \*.xls

Opdracht 9

*In /homes/arne/tmp/data/ staan een aantal .xls bestanden die gegevens over een batterij zonnecellen bevatten. I staat voor stroomsterkte in Ampere V staat voor spanning in Volt F staat voor de (wisselstroom)frequentie in Herz. Pac staat voor het op dat moment afgegeven vermogen in Watt. pv staat voor ‘photo voltaic’ dus opgewekt door de zonnecellen. Er zijn hier twee groepen voor. ac staat voor ‘alternating current’ dus dit wordt doorgegeven aan het stroomnetwerk. Hiervoor zijn er 3 fasen. De overige velden zouden voor zich moeten spreken.* *Copiëer deze data naar een directory in je eigen home en lees ze in in R.*

read.xls.self <- function(x){  
 read\_xls(x, col\_names = T, skip = 2)  
}  
my.data <- lapply(Sys.glob("\*.xls"), read.xls.self)

-  
/  
   
  
-  
/  
   
  
-  
/  
   
  
-  
/  
   
  
-  
/  
   
  
-  
/  
   
  
-  
/

names(my.data) <- c(1:length(my.data))

Opdracht 10

*Gebruik de gegevens die je in opdracht 9 ingelezen hebt. Maak één figuur waarin je het totaal aantal KWh wat gegenereerd is uitzet tegen de tijd (datum). Gebruik hiervoor de gegevens uit alle bestanden. Laat de datum zien in het formaat yyyymmdd.*

tijd <- numeric(0)  
gegenereerd <- numeric(0)  
for(num in 1:length(my.data)){  
 tijd <- c(tijd, my.data[[num]]$Time)  
 gegenereerd <- c(gegenereerd, my.data[[num]]$`Total Generation(kWh)`)  
 rm(num)  
}  
tijd <- as.Date(tijd, format = "%m/%d/%Y")  
plot(tijd, gegenereerd)

