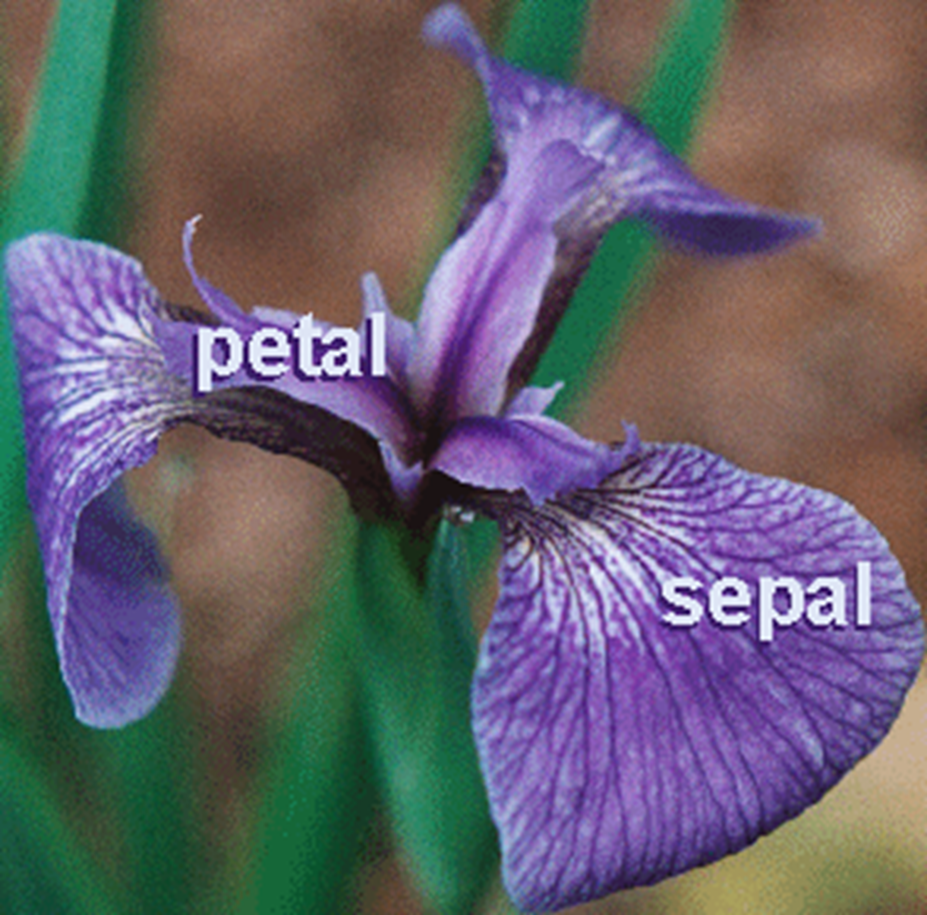
*eigen werk © HLO;HAN*

*De tijd, die je krijgt om deze opdracht te maken tijdens het tentamen is 60 minuten.*

**R- opdracht in tentamen:** (12 + 22 + 15 + 5 + 6 = 60 punten) maximaal 60 punten

De dataset iris wordt vaak gebruikt als voorbeeld voor classificering op basis van enkele kenmerken (patroonherkenning). Deze dataset bevat van 150 bloemen, allemaal irissen, gegevens over de lengte en breedte van zowel kelkblaadjes (sepal) als kroonblaadjes (petal). Er is gemeten in cm.



Haal deze dataset op in R. Hij zit in het package datasets.

1. Er zitten gegevens van drie verschillende soorten irissen in de dataset.

Bepaal **voor elke soort** van de vier andere variabelen: het minimum,

het maximum, het eerste kwartiel, het derde kwartiel en de mediaan.

Vul daarna de onderstaande tabel in (alleen voor Versicolor,steeds één cijfer achter de komma).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Versicolor** | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width |
| Minimum |  |  |  |  |
| 1ste kwartiel |  |  |  |  |
| mediaan |  |  |  |  |
| 3de kwartiel |  |  |  |  |
| Maximum |  |  |  |  |

1. Maak in R een **plaatje** met in één figuur de boxplots voor Sepal.Length;

Sepal.Width; Petal.Length en Petal.Width voor de soort **Setosa**.

Er mag geen boxplot van Species bij zitten! Voeg een titel toe en zorg voor duidelijke tekst langs beide assen.

1. Maak een **grafiek** met punten in een vlak, waarbij je langs de horizontale as

Petal.Length en langs de verticale as Petal.Width afzet voor alledrie de soorten

in één grafiek. Schrijf hieronder het **script** dat je gebruikt hebt.

(Denk aan titel, tekst langs de assen. Zet in de subtitel: bron: irisdataset van Fisher)

1. Probeer het script van onderdeel c zo aan te passen dat de punten, die bij “Setosa” horen blauw zijn en de punten, die bij “Versicolor” horen rood.

Voor de punten, die bij “Virginica” horen mag je een kleur naar keuze gebruiken.

Schrijf hieronder het **script** dat je gebruikt hebt (Denk aan legenda).

1. Welke irissoort kun je op grond van de grafiek bij onderdeel d gemakkelijk van de twee anderen onderscheiden? Toelichten! Als je geen grafiek hebt kunnen maken bij d mag je gegevens van onderdeel a gebruiken. Ook dan toelichten!

(12 + 22 + 15 + 5 + 6 = 60 punten) maximaal 60 punten

**Beoordeling R- opdracht**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Maximum | Oordeel student | Oordeel docent |
| Onderdeel a | **12** |  |  |
| Onderdeel b | **22** |  |  |
| Onderdeel c | **15** |  |  |
| Onderdeel d | **5** |  |  |
| Onderdeel e | **6** |  |  |
| **Totaal** | **60** |  |  |

**Antwoorden R-opdracht**

**a)** > iris

> setosa <-subset(iris,Species=="setosa")

> versicolor<-subset(iris, Species=="versicolor")

> virginica<-subset(iris, Species=="virginica")

Of

> iris

> setosa <-iris[1:50, ]

> versicolor<-iris[51:100, ]

> virginica<-iris[101:150, ]

Dan

> summary(virginica)

> summary(setosa)

> summary(versicolor)

**Versicolor**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width |
| minimum | 4,9 | 2,0 | 3,0 | 1,0 |
| 1ste kwartiel | 5,6 | 2,5 (2,525) | 4,0 | 1,2 |
| mediaan | 5,9 | 2,8 | 4,4 (4,35) | 1,3 |
| 3de kwartiel | 6,3 | 3,0 | 4,6 | 1,5 |
| maximum | 7,0 | 3,4 | 5,1 | 1,8 |

**b)** > boxplot(subset(setosa,select=-Species))

of

>boxplot(subset(setosa,select= c(Sepal.Length,Sepal.Width,Petal. Length,Petal.Width))

# plaatje, nog uitbreiden met tekst

> boxplot(subset(setosa,select=-Species),main="Metingen bij Iris Setosa",sub="bron: irisdataset Fisher",ylab="cm") # dit geeft een grafiek, tekst langs de x-as is al duidelijk

**c)** > plot(iris$Petal.Length,iris$Petal.Width) # basis, nog uitbreiden tot grafiek

>plot(iris$Petal.Length,iris$Petal.Width,main="Lengte en breedte van de kroonblaadjes voor drie soorten irissen”,sub="bron: irisdataset van Fisher" ,xlab="lengte in cm",ylab="breedte in cm")

# Er moeten 4 juiste teksten op de juiste plaats zijn toegevoegd om 10 punten voor de teksten te kunnen krijgen.

**d)** >plot(iris$Petal.Length,iris$Petal.Width,main="Lengte en breedte van de kroonblaadjes voor drie soorten irissen”,sub="bron: irisdataset van Fisher" ,xlab="lengte in cm",ylab="breedte in cm”) #dit had je al bij onderdeel c. Toevoegen:

>points(setosa$Petal.Length,setosa$Petal.Width,col=”blue”)

>points(versicolor$Petal.Length,versicolor$Petal.Width,col=”red”)

>points(virginica$Petal.Length,virginica$Petal.Width,col=”yellow”)

>legend(“topright”,legend=c(“Setosa”,”Versicolor”, “Virginica”),text.col=c(“blue”,”red”, “yellow”))

**e)** Iris **Setosa**, want Iris Setosa ligt in de grafiek “Petal.Length, Petal.Width”

geïsoleerd in de linkerhoek (beneden). Dit kun je door de kleur blauw zien.