



# Spiekbrieftje R

Rekenen • Vectoren • Dataframes • CSVs lezen • Selecties maken • Plotten (base R)

## Starten

### Working directory & files

```
getwd()      # -> huidige wd (working directory)
setwd("pad/naar/map") # -> wd aanpassen
list.files()  # -> bestanden in wd laten zien
```

### CSV inlezen & checken

```
df <- read.csv("data.csv") # -> df wordt een dataframe
head(df, n)                # -> toont eerste n rijen
str(df)                   # -> toont kolomnamen + types
```

## Rekenen

Getallen en vectoren worden meestal **elementgewijs** bewerkt.

### Basis rekenwerk

```
2 + 3      # -> 2 plus 3
5 - 1      # -> 5 min 1
4 * 2      # -> 4 keer 2
10 / 4     # -> 10 gedeeld door 4
2^5        # -> 2 tot de macht 5
sqrt(9)    # -> wortel van 9
```

### Vectoren optellen en vermenigvuldigen

```
a <- c(1, 2, 3)      # -> vector (1,2,3)
b <- c(10, 20, 30)    # -> vector (10,20,30)

a + b                # -> elementgewijs, dus (11, 22, 33)
a * b                # -> elementgewijs, dus (10, 40, 90)
a + 1                # -> alles van a + 1, dus (2, 3, 4)
```

### Vector berekeningen

```
sum(a)              # -> som van alle waarden
mean(a)             # -> gemiddelde
min(a)              # -> kleinste waarde
max(a)              # -> grootste waarde
length(a)           # -> lengte van de vector
```

## Logica & selectie

### Vergelijken + selecteren

```
x <- c(3, 7, 1, 9)

x > 5              # -> (FALSE, TRUE, FALSE, TRUE)
x[x > 5]           # -> (7, 9) (selectie)
which(x > 5)       # -> (2, 4) (indexen waarvoor dit geldt)
```

### if / else en ifelse()

```
leeftijd_student <- 19
if(leeftijd_student >= 18){
  print("Student is volwassen") # uitgevoerd want waar
} else{
  print("Student is niet volwassen") # niet uitgevoerd
}
print("Einde") # dit wordt altijd uitgevoerd
```

## Dataframes

Een dataframe is een tabel zoals een spreadsheet: elke **kolom** (column) is één variabele (met een vast datatype), en elke **rij** (row) is één observatie van al die variabelen.

### Kolom of rij pakken

```
df$kleur          # -> geeft kolom met naam 'kleur' terug
df[1, ]           # -> alles op rij 1 (observatie #1)
df[, 1]           # -> kolom 1 (op index)
```

### Filteren (selecteer rijen)

```
df[df$leeftijd >= 19, ] # -> rijen met leeftijd >= 19
subset(df, leeftijd >= 19) # -> hetzelfde (leesbaarder)
```

### Nieuwe kolom maken

```
df$adult <- df$leeftijd >= 18 # -> TRUE/FALSE per rij
df$leeftijd2 <- df$leeftijd^2  # -> kwadraat per rij
```

## Plotten (base R)

### Snelle plots

```
plot(x, y)          # -> puntenplot
plot(x, y, type="l") # -> lijnplot
hist(x, breaks=20)  # -> histogram
boxplot(x)          # -> boxplot
```

### Labels & hulplijnen

```
plot(x, y, type="l",
      xlab="tijd", ylab="waarde",
      main="Titel") # -> labels + titel

abline(h=0)         # -> horizontale lijn op y=0
```

## Veelgemaakte fouten

### De symbolen = en == betekenen iets anders!

```
x <- 3      # toewijzen
x = 3       # ook toewijzen, maar is niet gebruikelijk in R
x == 3      # NIET toewijzen, maar "vragen" of dit waar is
```

### NA breekt berekeningen

```
x <- c(1, NA, 3)
mean(x)          # -> NA
mean(x, na.rm=TRUE) # -> 2
```

### Bestand inlezen

```
read.csv("data.csv") # foutmelding, wd goed ingesteld?
setwd("pad/naar/map") # wd aanpassen
read.csv("data.csv")  # nu werkt het wel

read.csv("data2.csv") # huh? slechts 1 kolom...?
read.csv("data2.csv", sep=";") # nu werkt het!
```