

A Book

Frida Gomam

Contents

1	Hello World	5
2	Assignment 1.1	7

Chapter 1

Hello World

Hi.

Bye.

Toedels!

Chapter 2

Assignment 1.1

Engelse namen aanpassen naar Nederlands, verschuif Y-as labels verder van as vandaan

Het *C. elegans* kweekplaat experiment

Volwassen *C. elegans* nematoden zijn blootgesteld aan verschillende concentraties naftaleen, 2,6-di-isopropylnaftaleen en decaan, waarna is onderzocht of deze blootstellingen en negatief effect hebben op het aantal nakomelingen. S-medium dient als negatieve controle voor dit effect, 1,5% ethanol dient als positieve controle.

De belangrijkste variabelen in de verkregen dataset (`.data_raw/data01_assignment_1.1_c.elegans/CE.LIQ.FLOW.0`) zijn:

`expType`: de behandelconditie.

`RawData`: het aantal nakomelingen geteld na de incubatietijd.

`compName`: de generieke naam van de stof.

`compConcentration`: de concentratie van de stof.

Gebruikte CRAN packages voor de analyse: `{ggplot2}`

A.

Het Excel bestand heeft een tidy opmaak, maar bevat verschillende variabelen die voor elke onderzochte vial hetzelfde zijn. Zoals: `elegansStrain`, `bacterialStrain`, `incubateTemperature` etc. Het bestand bevat één sheet waarin de aanpassingen aan de parameter 'levels' zijn gedocumenteerd en één sheet met alle unieke 'level' opties per parameter. Er staan ook 'levels' tussen die niet zijn gebruikt in dit experiment; zo is er geïncubeerd bij 20 (waarschijnlijk) graden Celsius maar 15, 16, 25 en RT(?) graden Celcius hadden ook gekund? Er

mist een exacte beschrijven van de parameters en hun ‘levels’: zijn de graden inderdaad in Celsius? Betekent het pct ‘level’ van compConcentration procent?

5 MISSENDE WAARDEN

B.

#Upload bestand CE.LIQ.FLOW.062_Tidydata.xlsx naar de ./data_raw/data01_assignment_1.1

#Lees het bestand in in R als tibble:

```
elegans<-read_excel("C:/Users/steph/DSFB2/workflows_portfolio_OPNIEUW/data_raw/data01_1.xlsx")
summary(elegans)
```

```
##  plateRow      plateColumn      vialNr      dropCode
##  Mode:logical  Mode:logical  Min.   :1.000  Length:360
##  NA's:360      NA's:360      1st Qu.:1.000  Class :character
##                                     Median :2.000  Mode  :character
##                                     Mean   :2.375
##                                     3rd Qu.:3.000
##                                     Max.   :6.000
##
##  expType      expReplicate      expName      expDate
##  Length:360   Min.   :3.00  Length:360   Min.   :2020-11-30
##  Class :character 1st Qu.:3.00  Class :character 1st Qu.:2020-11-30
##  Mode  :character Median :3.00  Mode  :character Median :2020-11-30
##                                     Mean   :3.75  Mean   :2020-11-30
##                                     3rd Qu.:3.75  3rd Qu.:2020-11-30
##                                     Max.   :6.00  Max.   :2020-11-30
##
##  expResearcher      expTime      expUnit      expVolumeCounted
##  Length:360         Min.   :68  Length:360         Min.   :50
##  Class :character  1st Qu.:68  Class :character  1st Qu.:50
##  Mode  :character Median :68  Mode  :character Median :50
##                                     Mean   :68  Mean   :50
##                                     3rd Qu.:68  3rd Qu.:50
##                                     Max.   :68  Max.   :50
##
##  RawData      compCASRN      compName      compConcentration
##  Min.   : 0.0  Length:360      Length:360      Length:360
##  1st Qu.:51.5  Class :character  Class :character  Class :character
##  Median :72.0  Mode  :character  Mode  :character  Mode  :character
##  Mean   :68.1
##  3rd Qu.:88.0
##  Max.   :124.0
##  NA's    :5
##  compUnit      compDelivery      compVehicle      elegansStrain
##  Length:360     Length:360      Length:360      Length:360
```



```
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##   elegansInput bacterialStrain   bacterialTreatment bacterialOD600
## Min.   :25      Length:360       Length:360         Min.    :0.743
## 1st Qu.:25      Class :character   Class :character   1st Qu.:0.743
## Median :25      Mode  :character   Mode  :character   Median :0.743
## Mean   :25                                     Mean   :0.743
## 3rd Qu.:25                                     3rd Qu.:0.743
## Max.   :25                                     Max.   :0.743
##
## bacterialConcX bacterialVolume bacterialVolUnit   incubationVial
## Min.    :8      Min.    :300      Length:360       Length:360
## 1st Qu.:8      1st Qu.:300      Class :character   Class :character
## Median :8      Median :300      Mode  :character   Mode  :character
## Mean   :8      Mean    :300
## 3rd Qu.:8      3rd Qu.:300
## Max.   :8      Max.    :300
##
## incubationVolume incubationUnit   incubationMethod   incubationRPM
## Min.    :1000     Length:360       Length:360         Min.    :35
## 1st Qu.:1000     Class :character   Class :character   1st Qu.:35
## Median :1000     Mode  :character   Mode  :character   Median :35
## Mean   :1000
## 3rd Qu.:1000
## Max.   :1000
##
##   bubble      incubateTemperature
## Mode:logical  Min.    :20
## NA's:360      1st Qu.:20
##              Median :20
##              Mean   :20
##              3rd Qu.:20
##              Max.   :20
##
```

#Er missen 5 RawData waarden in het Excel bestand en dus ook in de tibble. De bijbehorende rijen

C.

Na het importeren van de data naar R is variabele RawData van het data type double geworden, dit moet integer zijn. compName is van het data type character, dit is prima. De variabele moet wel tot factor gemaakt worden voor gebruik in een R ggplot grafiek. Variabele compConcentration is van het data

type character, dit moet double zijn.

D.

```
elegans%>%ggplot(aes(x=as.double(compConcentration),y=as.integer(RawData)))+
  geom_point(aes(colour=as.factor(compName),shape=as.factor(expType)),size=1.5)+
  labs(x="Stof concentratie (nM/pct)",y="Aantal nakomelingen na blootstelling",colour=
  theme_bw()
```

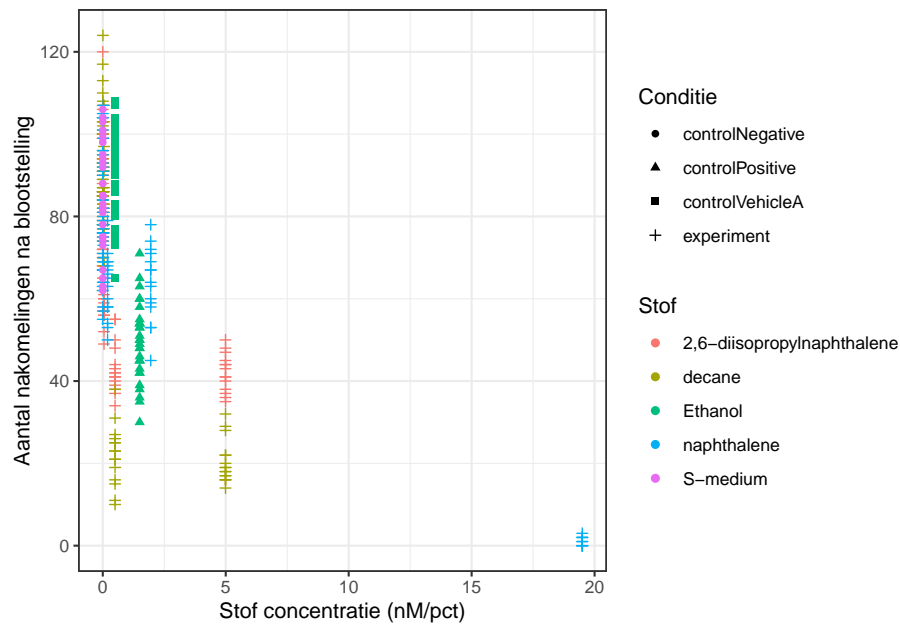


Figure 2.1: (#fig:1.1 D) **Figuur 1: Aantal *C. elegans* nakomelingen na verschillende blootstellingen versie 1.** Het aantal nakomelingen is geteld na incubatie voor een onbekende tijd van nematoden in S-medium en bij verschillende concentraties naftaleen, 2,6-di-isopropylnaftaleen, decaan en ethanol.

E.

Indien de compConcentration variabele, van het data type character, niet als double wordt opgegeven, ontstaat een foute X-as.

Verder: de stof concentraties worden gegeven in nM maar de concentratie ethanol wordt gegeven in procent. Reken als volgt het percentage ethanol om naar nM:

$$\text{nMol/L} = (\text{percentage ethanol} * \text{dichtheid in g/L}) / (\text{molaire massa in g/mol} * 100) * 1000000$$

*De dichtheid van het gebruikte ethanol is nodig om de percentages om te rekenen naar nM, maar niet gegeven. Daarom wordt uitgegaan van de standaard

789 g/L.

1.5% ethanol = 256891.686564 nmol/L

0.5% ethanol = 85630.562188 nmol/L

#Verander de ethanol concentraties van procent naar nM:

```
elegans_new<-elegans
```

```
elegans_new$compConcentration<-elegans_new$compConcentration%>%str_replace_all("1.5","256891.686564")
```

```
elegans_new$compConcentration<-elegans_new$compConcentration%>%str_replace_all("0.5","85630.562188")
```

```
elegans_new$compUnit<-elegans_new$compUnit%>%str_replace_all("pct","nM")
```

F.

Zet de stof concentraties op een log10-schaal en voeg variatie toe aan de punten tegen overlap.

```
elegans_new%>%ggplot(aes(x=log10(as.double(compConcentration)),y=as.integer(RawData)))+
  geom_point(aes(colour=as.factor(compName),shape=as.factor(expType)),size=1.5,position="jitter")+
  labs(x="Log10(stof concentratie) (nM)",y="Aantal nakomelingen na blootstelling",colour="Stof",shape="Conditie")+
  theme_bw()
```

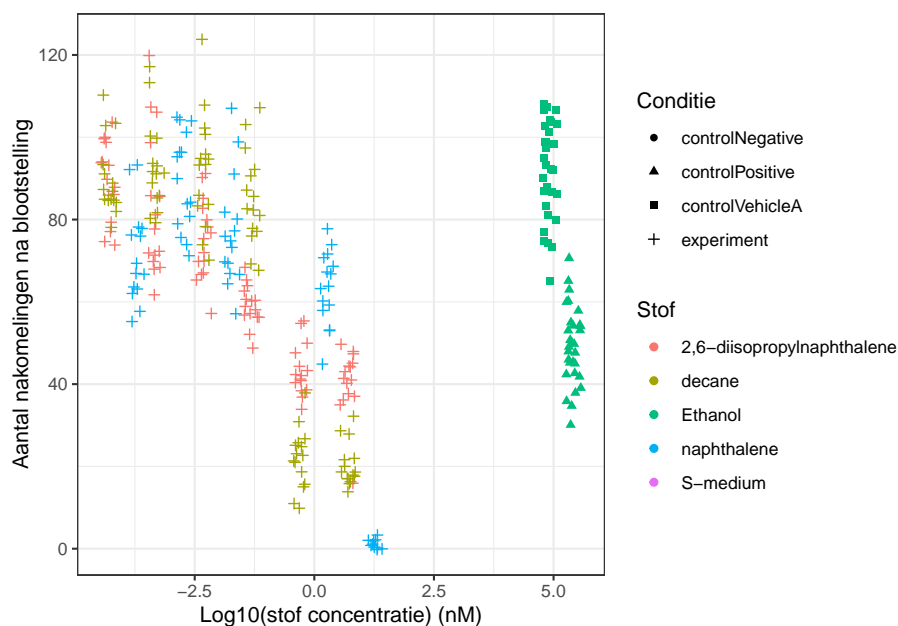


Figure 2.2: (#fig:1.1 F) **Figuur 2: Aantal *C. elegans* nakomelingen na verschillende blootstellingen versie 2.**

G.

De positieve controle voor dit experiment is dus 1.5% ethanol, waarvan bekend

is dat het zorgt voor minder nematode nakomelingen. De negatieve controle voor dit experiment is dus S-medium, met een compConcentration gezet op 0 pct. Vandaar dat S-medium niet meer in de versie 2 grafiek verschijnt, de log10 van 0 bestaat niet.

NOG MAKEN: I.

```
https://storage.googleapis.com/plos-corpus-prod/10.1371/journal.pone.0146021/1/pone.0146021.s001.pdf?X-Goog-Algorithm=GOOG4-RSA-SHA256&X-Goog-Credential=wombat-sa%40plos-prod.iam.gserviceaccount.com%2F20220421%2Fauto%2Fstorage%2Fgoog4_request&X-Goog-Date=20220421T100713Z&X-Goog-Expires=86400&X-Goog-SignedHeaders=host&X-Goog-Signature=cc73877fc2f7c1cb9ea8b2a7eda54fa0dd715d217d2e519b9675b415a3afd033dc22a8adb9
```

J.

```
controlNegative<-c(92,101,106,104,93,73,94,104,78,103,63,85,62,75,82,88,85,67,65,81,85)
mean(controlNegative)
```

```
## [1] 85.9
```

#De ruwe data van controlNegative delen door 85.9 maakt dat het gemiddelde gelijk is a

```
elegans_new<-elegans_new%>%mutate(RawData_normalized=RawData/mean(controlNegative))
```

```
elegans_new%>%ggplot(aes(x=log10(as.double(compConcentration)),y=as.integer(RawData_no
  geom_point(aes(colour=as.factor(compName),shape=as.factor(expType)),size=1.5,position
  labs(x="Log10(stof concentratie) (nM)",y="Fractie nakomelingen na blootstelling",col
  theme_bw()
```

#Er zijn alleen negatieve fracties door de jitter optie!

#Waarom zijn er zonder jitter geen punten tussen fractie 0 en fractie 1? Worden alle f

Voor de positieve controle met 1,5% ethanol geldt, dat er duidelijk minder nematoden geboren worden na incubatie ten opzichte van na incubatie in S-medium (fractie « 1).

Bij stof concentraties lager dan ~ 0.06 nM (= log10's lager dan -1.25) worden er evenveel, minder of meer nematoden geboren. Bij deze concentraties is er dus geen effect op het aantal nakomelingen. Voor incubatie bij hogere concentraties van de drie geteste stoffen is zichtbaar dat er minder nematoden geboren worden, ten opzichte van incubatie in S-medium (fractie « 1)

K.

We normaliseren de data omdat we geïnteresseerd zijn in hoeveel nematoden er meer of minder worden geboren na incubatie met een stof ten opzichte van de normale situatie (S-medium zonder stof). Dit is niet af te lezen aan de grafiek met de X-as log10-schaal van niet-genormaliseerde data, omdat het aantal nakomelingen van de negatieve controle conditie wegvallen uit de grafiek (zie

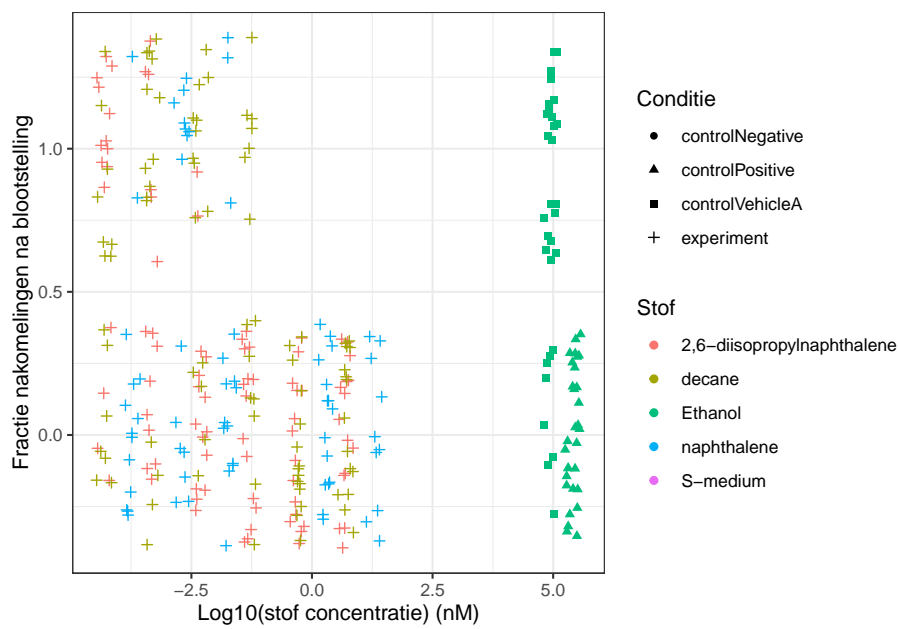


Figure 2.3: (#fig:1.1 J)**Figuur 3: Fracties *C. elegans* nakomelingen na verschillende blootstellingen.** Voor elke blootstelling aan één van de stoffen in een bepaalde concentratie is de fractie nakomelingen berekend, relatief een het aantal nakomelingen binnen de negatieve controle (incubatie in S-medium).

G.).