

Portfolio Ghislaine

Ghislaine van Gilse

2022-12-28

Contents

Prerequisites	5
Curriculum Vitae	7
Name	7
1 Reproducible Research	9
1.1 C.elegans plate experiment	9

Prerequisites

This portfolio was created as an assignment for Data Science for Biology 2 at the HU University of Applied Sciences Utrecht.

14 Nov. 2022 - 8 Jan. 2023

Curriculum Vitae

Name

Adres postcode

Telefoon: Email: Geboortedatum: Geboorteplaats:

Werkervaring

Tijdens mijn middelbare school en HBO-opleiding heb ik enkele bijbanen gehad:

Action Augustus 2015 - februari 2016, Hilversum

Hier heb ik als vakkenvuller medewerkster gewerkt.

Sociëteit de Unie Augustus 2018 - februari 2020, Hilversum Bij deze herenso-ciëteit heb ik als ampulante kracht in de bediening gewerkt. Daarnaast heb ik ook achter de bar gestaan en incidenteel in de spoelkeuken gewerkt.

Kinderoppas September 2018 - mei 2020, Hilversum Opgepast op basis- en onderbouw (middelbare) schoolkinderen. Deze heb ik onder meer geholpen bij het maken van hun huiswerk en bij het leren voor toetsen.

Blokker Februari 2022 - heden, Hilversum Hier werk ik achter de kassa, maar ben ik ook in de winkel bezig met klanten, vakkenvullen en vracht.

Opleiding

Alberdingk Thijm College September 2013 - juli 2018, Hilversum Op deze school heb ik vijf jaar HAVO gedaan. Vakkenpakket Natuur & Gezondheid met natuurkunde en economie als keuzevakken.

ROC Midden-Nederland - VAVO Lyceum Utrecht / HAVO-diploma September 2018 - juli 2019, Utrecht Het vak natuurkunde gevolgd om mijn HAVO-diploma af te ronden.

Hogeschool Utrecht / Propedeuse September 2019 - heden, Utrecht Opleiding Life Science - Biologie en Medisch laboratorium onderzoek Propedeuse behaald: Januari 2021

Honours: Februari 2021: Mensenrechten & duurzame ontwikkeling

Vaardigheden

Labvaardigheden Onder andere: Celkweek, (licht- en fluorescentie)microscopie, SDS-page, Western-blot, (q)PCR, ELISA, (Agarosegel) elektroforese, kloonwerk

Data verwerking (Basis)kennis van excel en SPSS Basiskennis van het werken met Rstudio waaronder, Bash R

Talen Nederlands (moedertaal) Engels B1

Hobby's & Interesses

(Berg)wandelen Paardrijden Skiën Trainen met de hond Bakken Puzzelen

Nevenactiviteiten

Gedurende drie jaar op de middelbare school van het Alberdingk Thijm College ben ik actief geweest als lid van de leerlingenraad.

Vrijwilligerswerk Comic Con November 2022

Chapter 1

Reproducible Research

1.1 C.elegans plate experiment

For this experiment, adult *C.elegans* nematodes have been exposed to varying concentrations of different compounds. The data that is used for this exercise is supplied by J. Louter (INT/ILC).

At first we had to review the Excel file with all the data. The file is called `./data.CE.LIQ.FLOW.062_Tidydata.xlsx`. Something that stood out to me was that the different compounds were measured in nM. But the ethanol and the medium of the cells was measured in percentage.

The compounds used in this experiment are: 2,6-diisopropyl-naphthalene, Decane and Naphthalene with as positive control group 1,5% Ethanol in S-Medium and as negative control group just S-Medium.

```
elegansData <- read_excel(here("port_data/CE.LIQ.FLOW.062_Tidydata.xlsx"))
tibble(elegansData)
```

```
## # A tibble: 360 x 34
##   plateRow plateCo~1 vialNr dropC~2 expType expRe~3 expName expDate
##   <lg1>    <lg1>      <dbl> <chr>   <chr>      <dbl> <chr>   <dtm>
## 1 NA      NA          1 a      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
## 2 NA      NA          1 b      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
## 3 NA      NA          1 c      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
## 4 NA      NA          1 d      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
## 5 NA      NA          1 e      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
## 6 NA      NA          2 a      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
## 7 NA      NA          2 b      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
## 8 NA      NA          2 c      experi~    3 CE.LIQ~ 2020-11-30 00:00:00
```

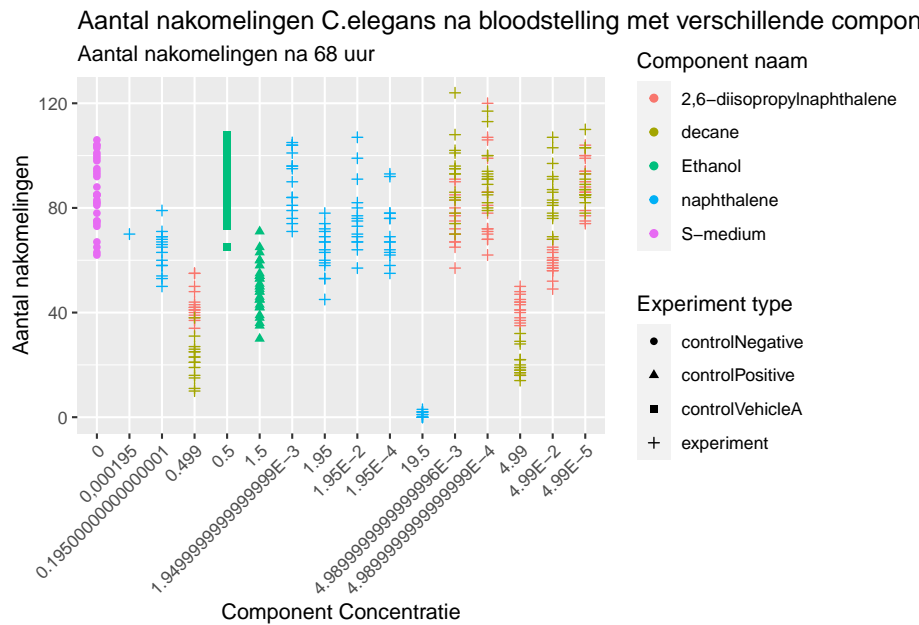
```
## 9 NA      NA      2 d      experi~      3 CE.LIQ~ 2020-11-30 00:00:00
## 10 NA     NA      2 e      experi~      3 CE.LIQ~ 2020-11-30 00:00:00
## # ... with 350 more rows, 26 more variables: expResearcher <chr>,
## #   expTime <dbl>, expUnit <chr>, expVolumeCounted <dbl>, RawData <dbl>,
## #   compCASRN <chr>, compName <chr>, compConcentration <chr>, compUnit <chr>,
## #   compDelivery <chr>, compVehicle <chr>, elegansStrain <chr>,
## #   elegansInput <dbl>, bacterialStrain <chr>, bacterialTreatment <chr>,
## #   bacterialOD600 <dbl>, bacterialConcX <dbl>, bacterialVolume <dbl>,
## #   bacterialVolUnit <chr>, incubationVial <chr>, incubationVolume <dbl>, ...
```

After loading the data into Rstudio we were asked to inspect the data types of the columns: RawData, compName and compConcentration and what types we expected. I expect RawData: Numeric/integer, compName: character/factor and compConcentration: numeric/double. The typeof function tells me that RawData is a double. It doesn't contain decimals so it should be integer but this is not a big problem in further analyses. compName is set as character compConcentration is not a double. This means the data types have not correctly been assigned during the importing of the data into R.

#D. Create a graph displaying a scatterplot for the CE.LIQ.FLOW.062_Tidydata.xlsx data, for the different compounds and the varying concentrations. Put the compConcentration on the x-axis, the DataRaw counts on the y-axis and assign a colour to each level in compName. Assign a different symbol (shape =) to each level in the expType variable. Try fixing the labels of the x-axis so that we can read them.

```
ggplot(elegansData, aes(x=compConcentration, y=RawData, shape=expType, colour=compName)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  labs(colour="Component naam", shape="Experiment type",
       x="Component Concentratie",
       y="Aantal nakomelingen",
       title="Aantal nakomelingen C.elegans na bloedstelling met verschillende componenten",
       subtitle="Aantal nakomelingen na 68 uur"
  )
```

```
## Warning: Removed 5 rows containing missing values (`geom_point()`).
```



#E. When creating the plot under C), what happened with the ordering of the x-axis labels. Explain why this happens. Look at the data-type of the compConcentration column in the data again to find a clue. ##De volgorde wordt gegroepeerd en niet meer van klein naar groot. Dit komt omdat hij het ziet als characters en dus op alfabetische volgorde zet. 4.99E-5 komt later in het alfabet dan 4.99 en 4 komt later dan 0 of 1

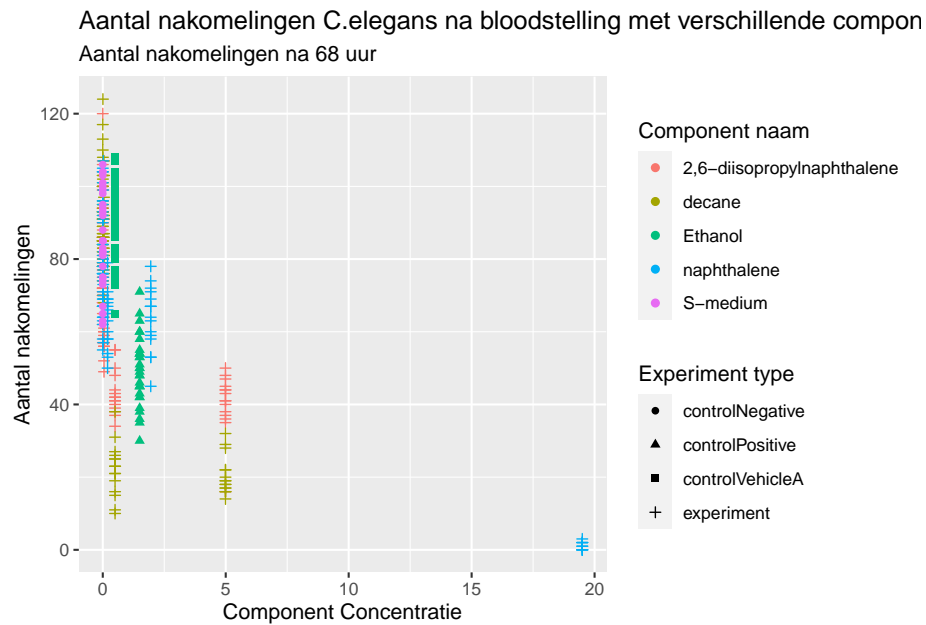
```
elegansData$compConcentration <- as.numeric(elegansData$compConcentration)
```

```
## Warning: NAs introduced by coercion
```

#F. Correct this and than look at the graph again.

```
ggplot(elegansData, aes(x=compConcentration, y=RawData, shape=expType, colour=compName, label=rawData)) +
  geom_point() +
  labs(colour="Component naam", shape="Experiment type",
       x="Component Concentratie",
       y="Aantal nakomelingen",
       title="Aantal nakomelingen C.elegans na bloodstelling met verschillende componenten",
       subtitle="Aantal nakomelingen na 68 uur")
```

```
## Warning: Removed 6 rows containing missing values (`geom_point()`).
```

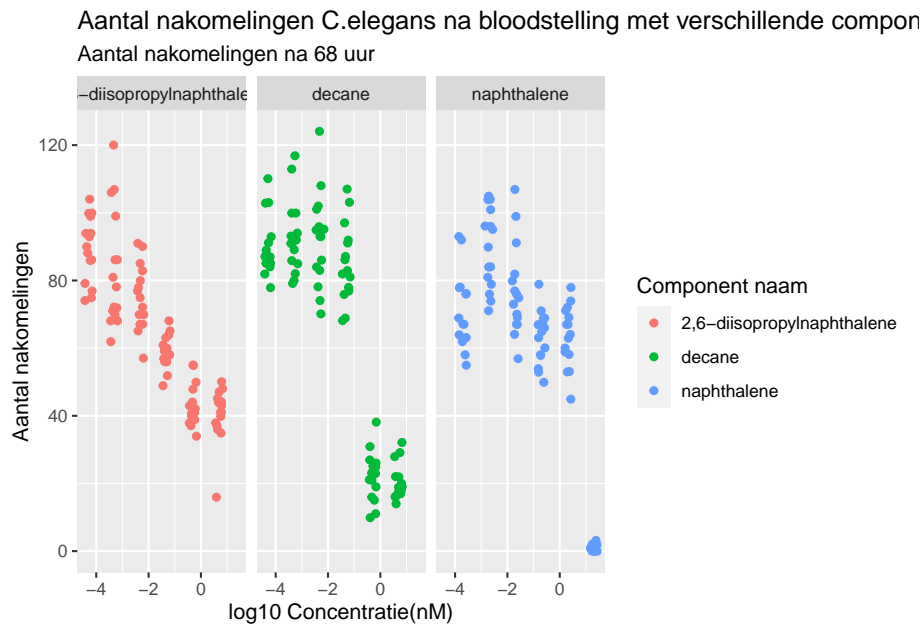


#F.2 Use a log10 transformation on the x-axis to get a clear graph. Also, add a bit of jitter to the points in the graph so that points are not overlapping (Google “ggplot jitter” if needed).

```
elegansData_nM <- elegansData %>% filter(compUnit == "nM")

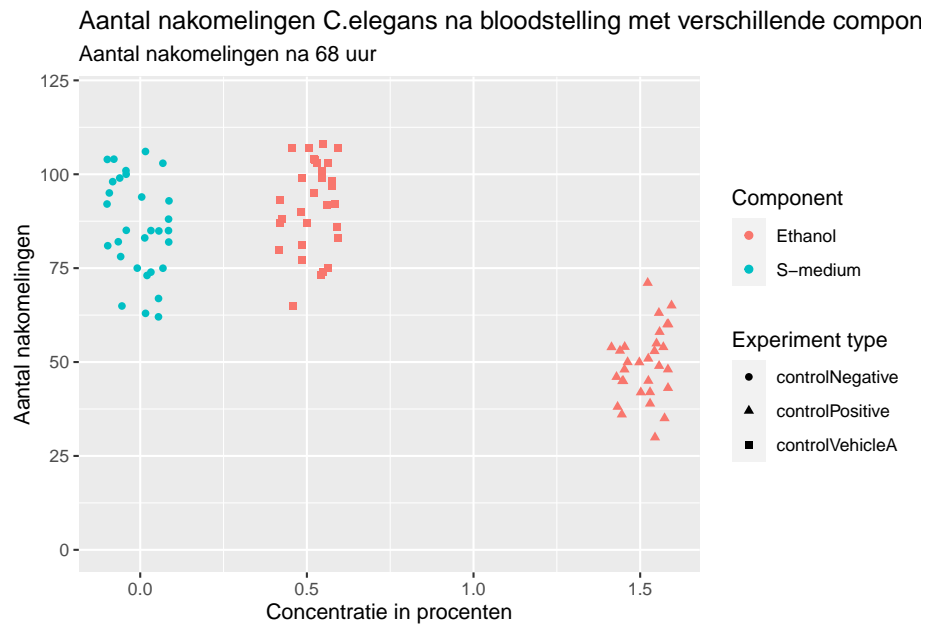
ggplot(elegansData_nM, aes(x=log10(compConcentration), y=RawData, colour=compName))+
  geom_point(position = position_jitter(h=0.15,w=0.15))+
  scale_x_continuous()+
  labs(colour="Component naam",
       x="log10 Concentratie(nM)",
       y="Aantal nakomelingen",
       title="Aantal nakomelingen C.elegans na bloodstelling met verschillende componen",
       subtitle= "Aantal nakomelingen na 68 uur"
  )+
  facet_wrap(~compName)
```

Warning: Removed 6 rows containing missing values (`geom_point()`).



```
elegansData_pct <- elegansData %>% filter(compUnit == "pct")

ggplot(elegansData_pct, aes(x=compConcentration, y=RawData, shape=expType, colour=compName))+
  geom_point(position = position_jitter(h=0.1,w=0.1))+
  coord_cartesian(ylim = c(0, 120))+
  labs(colour="Component", shape="Experiment type",
       x="Concentratie in procenten",
       y="Aantal nakomelingen",
       title="Aantal nakomelingen C.elegans na bloodstelling met verschillende componenten",
       subtitle= "Aantal nakomelingen na 68 uur"
  )
```



#Fill in: ##(G) The positive control for this experiments is ethanol (expType controlPositive). ##(H) The negative control for this experiment is s-medium (expType controlNegative).

#I. Think about how you would analyze this experiment to learn whether there is indeed an effect of different concentrations on offspring count and whether the different compounds have a different curve (IC50). Write down you analysis as a step-wise plan.

#J. Normalize the data for the controlNegative in such a way that the mean value for controlNegative is exactly equal to 1 and that all other values are expressed as a fraction thereof. Rerun your graphs with the normalized data.

#K. Why would you want to take the step under J?