

Portofolio

Timo Voskuilen

2022-05-09

Contents

1	C.elegans experiment	5
2	Guerilla analytics	9
2.1	Contact Info	12
2.2	Disclaimer	12
3	CV	13
3.1	Timo Voskuilen	13
3.2	Education	13
3.3	Research Experience	13
4	Looking ahead	15
4.1	Assignment 1:	16
4.2	Assignment 2:	16
4.3	Assignment 3:	16
4.4	Assignment 4:	16
4.5	Assignment 5:	16
4.6	Assignment 6:	16
4.7	Making this a bookdown:	16
4.8	Assignment 7:	16
4.9	Assignment 8:	17
4.10	Assignment 9:	17
4.11	Free assignment	17

Chapter 1

C.elegans experiment

And the data types of the variables RawData, compName and compConcentration:

```
## # A tibble: 1 x 3
##   RawData compName          compConcentration
##   <dbl> <chr>          <chr>
## 1      44 2,6-diisopropylnaphthalene 4.99
```

It would be expected that the data types of the columns are dbl, chr and dbl. compConcentration however, has the datatype chr. This means the format was imported incorrectly making it difficult to analyse the data.

after transforming the compConcentration data to double, the data was plotted in a scatterplot graph with the data for the different compounds on the y-axis (RawData) and the varying concentrations on the x-axis (compConcentration). Each variable got a colour matching the compName and a symbol matching the expType column-values. Jitter and log10 scaling of the X axis was added, for the “correct” graph

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

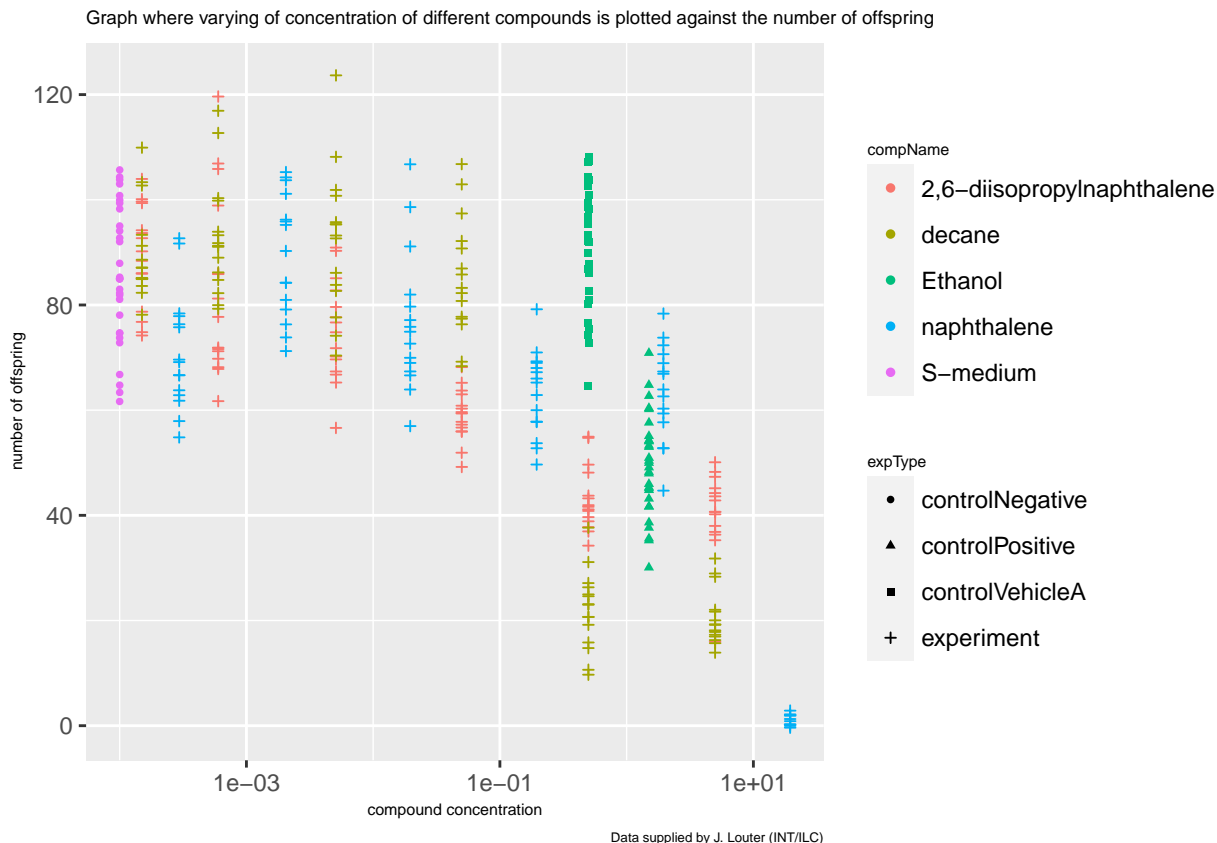


Figure 1: Scatterplot displaying the amount of offspring at different concentrations per compound (colour) and experiment type (symbol)

The positive control for this experiments is ethanol. The negative control for this experiment is S-medium.

To analyze this experiment and learn whether there is an effect of different concentrations on offspring count and the effects of different compounds i would take these steps: -group data per compound -check if data is normally distributed -Use the appropriate statistical tests on the data to see if there is a significant difference in offspring count when using different concentrations of compounds -Calculate IC50 and create a plot of the curve

For the next visualization I have normalized 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.

```
## 'summarise()' has grouped output by 'expType', 'compName'. You can override
## using the '.groups' argument.
```

expType	compName	compConcentration	mean_RawData_normalized
controlNegative	S-medium	0.00e+00	1.0000000
controlPositive	Ethanol	1.50e+00	0.5750873
controlVehicleA	Ethanol	5.00e-01	1.0690726
experiment	2,6-diisopropylnaphthalene	4.99e-05	1.0391929
experiment	2,6-diisopropylnaphthalene	4.99e-04	0.9670159

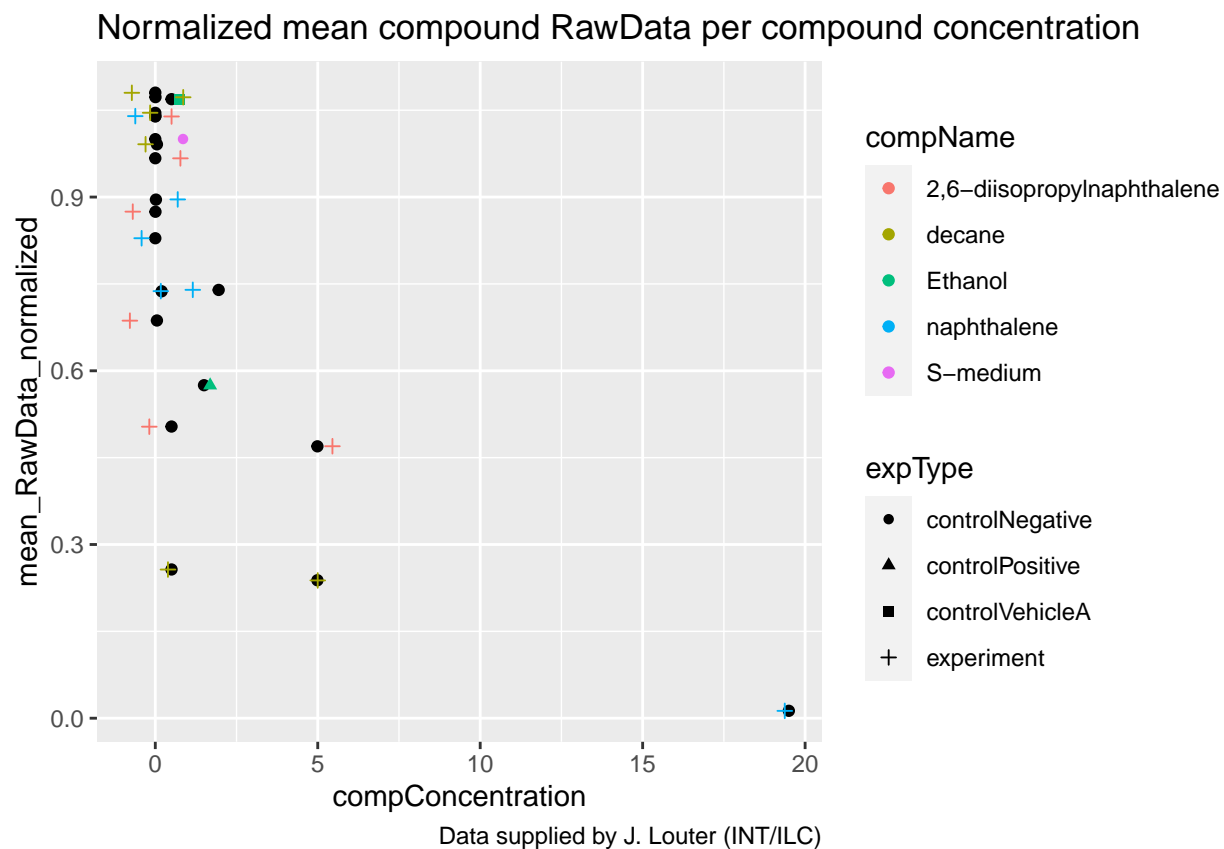


Figure 4: Scatterplot displaying the amount of offspring at different concentrations per compound (colour) and experiment type (symbol) normalized on the negative control.

this step was taken to get the result in relation to the “0” value, the amount of offspring from without any compound.

Chapter 2

Guerilla analytics

An important part of having a project repository is to keep it easily accesible and clean. This will allow anyone who uses the repository to be able to easily find their way around the files and data. The data structure in the repository used for this portfolio is based on the Guerilla Analytics Principle.

An example of this can be found below:

```
#create a tree for the DAUR2 structure  
fs::dir_tree(dir = "https://github.com/TimoVoskuilen/Portofolio/tree/main/data/data_Raw/003_data_management")
```

```
## .  
## +-- 001_C.elegans_exp.html  
## +-- 001_C.elegans_exp.Rmd  
## +-- 003_Guerillla_analytics.html  
## +-- 003_Guerillla_analytics.Rmd  
## +-- 004_CV.Rmd  
## +-- 005_looking_ahead.Rmd  
## +-- data  
## |   \-- data_Raw  
## |       +-- 001_C.elegans_exp  
## |           |   +-- CE.LIQ.FLOW.062_Tidydata.xlsx  
## |           |   \-- README  
## |           \-- 003_data_management  
## |               +-- daur2  
## |                   |   +-- daur2.Rproj  
## |                   |   +-- LICENSE  
## |                   |   +-- metagenomics  
## |                   |   |   +-- dataset1  
## |                   |   |       |   +-- HU1 MOCK1_L001_R1_001_fastqc.html  
## |                   |   |       |   +-- HU1 MOCK1_L001_R1_001_fastqc.zip  
## |                   |   |       |   +-- HU1 MOCK1_L001_R2_001_fastqc.html  
## |                   |   |       |   +-- HU1 MOCK1_L001_R2_001_fastqc.zip  
## |                   |   |       |   +-- mock1.bracken  
## |                   |   |       |   \-- README  
## |                   |   |   +-- datasetformative  
## |                   |   |       |   +-- HU2 MOCK2_L001_R1_001_fastqc.html  
## |                   |   |       |   +-- HU2 MOCK2_L001_R1_001_fastqc.zip  
## |                   |   |       |   +-- HU2 MOCK2_L001_R2_001_fastqc.html  
## |                   |   |       |   +-- HU2 MOCK2_L001_R2_001_fastqc.zip
```

```

## |           | | | | \-- README
## |           | | | +-- formatieveopdracht2.html
## |           | | | +-- formatieveopdracht2.Rmd
## |           | | | +-- metagenomics.html
## |           | | | +-- metagenomics.Rmd
## |           | | | \-- pictures
## |           | | |   +-- fastqc_mock1_R1_per_base_quality.png
## |           | | |   +-- fastqc_mock1_R2_per_base_quality.png
## |           | | |   +-- fastqc_mock2_R1_per_base_quality.png
## |           | | |   \-- fastqc_mock2_R2_per_base_quality.png
## |           | +-- README.md
## |           | \-- RNA seq
## |           |   +-- airway data
## |           |   | +-- README
## |           |   | +-- SRR1039508
## |           |   | +-- SRR1039509
## |           |   | +-- SRR1039512
## |           |   | +-- SRR1039513
## |           |   | +-- SRR1039516
## |           |   | +-- SRR1039517
## |           |   | +-- SRR1039520
## |           |   | \-- SRR1039521
## |           |   +-- eindopdracht.html
## |           |   +-- eindopdracht.Rmd
## |           |   +-- formatieveopdracht1.html
## |           |   +-- formatieveopdracht1.Rmd
## |           |   +-- formative data
## |           |   | +-- README
## |           |   | +-- SRR7866687
## |           |   | +-- SRR7866688
## |           |   | +-- SRR7866689
## |           |   | +-- SRR7866690
## |           |   | +-- SRR7866691
## |           |   | +-- SRR7866692
## |           |   | +-- SRR7866693
## |           |   | \-- SRR7866694
## |           |   \-- pictures
## |           |       +-- SRR7866703_1_per_b_sq.png
## |           |       +-- SRR7866703_1_per_s_qs.png
## |           |       +-- SRR7866703_2_per_b_sq.PNG
## |           |       \-- SRR7866703_2_per_s_qs.PNG
## |           \-- Guerilla analytics framework daur2.png
## +-- Index.Rmd
## +-- LICENSE
## +-- Portofolio.Rproj
## +-- README.md
## +-- _book
## | \-- _main.pdf
## +-- _main.Rmd
## \-- _main_files
##     \-- figure-latex
##         +-- normalized-1.pdf
##         \-- unnamed-chunk-2-1.pdf

```

```
knitr::include_graphics("data/data_Raw/003_data_management/Guerilla analytics framework daur2.png")
```

```
+-- daur2.Rproj
+-- metagenomics
|   +-- dataset1
|   |   +-- HU1 MOCK1_L001_R1_001_fastqc.html
|   |   +-- HU1 MOCK1_L001_R1_001_fastqc.zip
|   |   +-- HU1 MOCK1_L001_R2_001_fastqc.html
|   |   +-- HU1 MOCK1_L001_R2_001_fastqc.zip
|   |   +-- mock1.bracken
|   |   \-- README
|   +-- datasetFormative
|   |   +-- HU2 MOCK2_L001_R1_001_fastqc.html
|   |   +-- HU2 MOCK2_L001_R1_001_fastqc.zip
|   |   +-- HU2 MOCK2_L001_R2_001_fastqc.html
|   |   +-- HU2 MOCK2_L001_R2_001_fastqc.zip
|   |   \-- README
|   +-- formatieveopdracht2.html
|   +-- formatieveopdracht2.Rmd
|   +-- metagenomics.html
|   +-- metagenomics.Rmd
|   \-- pictures
|       +-- fastqc_mock1_R1_per_base_quality.png
|       +-- fastqc_mock1_R2_per_base_quality.png
|       +-- fastqc_mock2_R1_per_base_quality.png
|       \-- fastqc_mock2_R2_per_base_quality.png
\-- RNA seq
    +-- airway data
    |   +-- README
    |   +-- SRR1039508
    |   +-- SRR1039509
    |   +-- SRR1039512
    |   +-- SRR1039513
    |   +-- SRR1039516
    |   +-- SRR1039517
    |   +-- SRR1039520
    |   \-- SRR1039521
    +-- eindopdracht.html
    +-- eindopdracht.Rmd
    +-- formatieveopdracht1.html
    +-- formatieveopdracht1.Rmd
    +-- formative data
    |   +-- README
    |   +-- SRR7866687
    |   +-- SRR7866688
    |   +-- SRR7866689
    |   +-- SRR7866690
    |   +-- SRR7866691
    |   +-- SRR7866692
    |   +-- SRR7866693
    |   \-- SRR7866694
    \-- pictures
        +-- SRR7866703_1_per_b_sq.png
        +-- SRR7866703_1_per_s_qs.png
        +-- SRR7866703_2_per_b_sq.PNG
        \-- SRR7866703_2_per_s_qs.PNG
```

2.1 Contact Info

- github.com/TimoVoskuilen

2.2 Disclaimer

Last updated on 2022-05-09.

Chapter 3

CV

3.1 Timo Voskuilen

Student with experience working in the biotechnology industry. Skilled in techniques commonly used in the laboratory, such as cell culture, organoid culture, DNA isolation and PCR.

Also very interested and learning data sciences and programming (R, R shiny, Bash).

3.2 Education

3.2.1 ROC Midden nederland

Microbiological analyst

Utrecht, Nederland

2015

Thesis: Organoid dissociation optimization.

3.2.2 Hogeschool Utrecht

B.S. biomolecular research

Utrecht, Nederland

2019

3.3 Research Experience

3.3.1 Intern

Utrecht center of Tick Borne Diseases (UCTD), Universiteit Utrecht

Utrecht, Nederland

2017 - 2018

The purpose of my internship at Utrecht Center of Tick Borne Diseases (UCTD) was identifying hotspots for ticks and tick borne diseases. I was responsible for the taxonomic identification of send in ticks.

•

3.3.2 Intern

Hub Organoids.

Utrecht, Nederland

2018 - 2019

- The goal of my internship at HUB was the optimization of organoid dissociation to single cells for transfection. I was also responsible for making culture media, and culturing a recombinant HEK293T cell line.

3.3.3 Publications

“Tekenscanner”: a novel smartphone application for companion animal owners and veterinarians to engage in tick and tick-borne pathogen surveillance in the Netherlands,” Parasit Vectors, 2019

Chapter 4

Looking ahead

In 2 years time I want to be working as a data scientist specializing in data engineering. My first big step and also my first big inspiration in regards to bioinformatics and data science is creating this digital portfolio as a result of the minor Data Science For Biology.

The next skill I would like to learn is controlling a database and building interactive websites, so that people who do not know data science can still enjoy the gift as it is. I will be doing a projecticum where both of these skills will be used.

I will also be doing an internship where i will be learning to use a SQL database and translate the data to my coworkers through the magic of an interactive R.shiny website.

4.1 Assignment 1:

Goal of this assignment: Being able to use data from real experiments and being able to reproduce the results.

COMPLETED

4.2 Assignment 2:

Goal of this assignment: Being able to rank experiments on open science and reproducibility.
need to find a publication first

4.3 Assignment 3:

Goal of this assignment: Being able to store my files accordingly using the Guerilla analytics framework
COMPLETED

4.4 Assignment 4:

Goal of this assignment: Making a CV using RMarkdown
COMPLETED

4.5 Assignment 5:

Goal of this assignment: Being able to learn a new skill without holding the hands of the teachers
my skill to learn would be shiny in a metagenomics environment

4.6 Assignment 6:

Goal of this assignment: Being able to write an introduction using references and crediting those references accordingly

4.7 Making this a bookdown:

Goal of this assignment: Creating a bookdown to present my work easily

4.8 Assignment 7:

Goal of this assignment: Being able to use SQL to store datasets in databases

4.9 Assignment 8:

Goal of this assignment: Being able to create packages to use on a later date

4.10 Assignment 9:

Goal of this assignment: Being able to use parameters

4.11 Free assignment

For my free assignment i want to be specializing on shiny webapps. I find it very important to make data sciences accessible to the public, and to the people at the lab who do not have the skills of data science. An interactive shiny webapp enables you to communicate results via interactive charts, visualizations, text or tables. it is very easy for people to use and get used to.