Biodiversity knowledge synthesis: an introduction to meta-analyses and systematic reviews

Quantitative data extraction

October 2024 - Montpellier

Dakis-Yaoba Ouédraogo (PhD)
dakis.ouedraogo@gmail.com

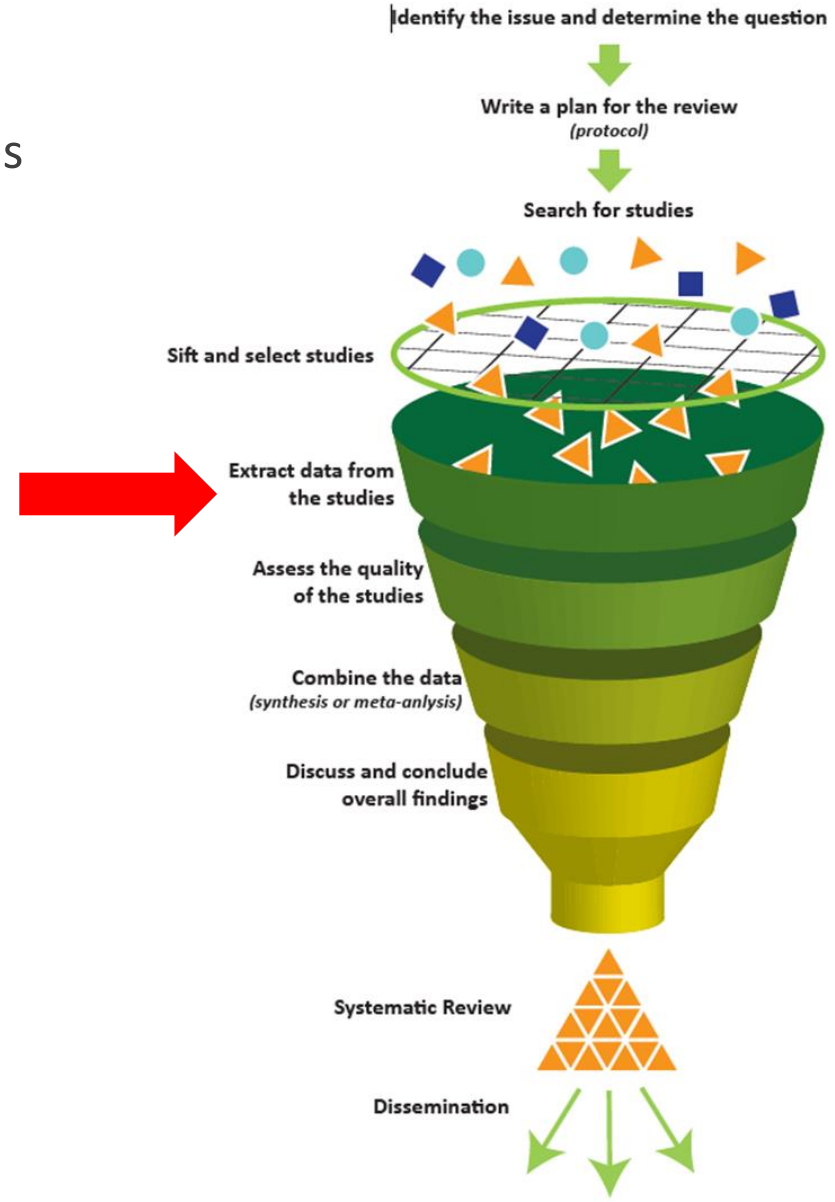
Extraction of quantitative data

Extraction of the data needed to calculate effect sizes (e.g. mean, sample size, sd/se/95% CI)

+ extraction of variables that could explain the heterogeneity of effect sizes (*effect modifiers*)

Extraction from
- text
- table
- figure
- supp. mat.

+ calculations may be needed (keep a record)



Jessica Kaufman, Cochrane Consumers & Communication Review Group, Centre for Health Communication and Participation, La Trobe University, 2011.

Extraction of quantitative data

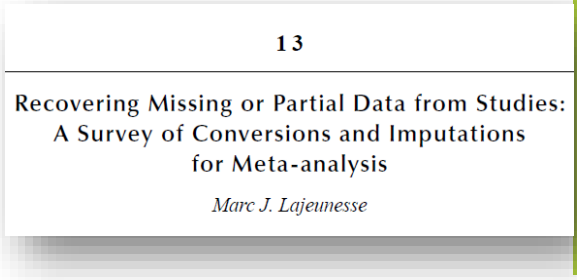
! Warning !

Data extraction is time-consuming: clearly define the extraction grid and the effect modifiers to be extracted

Importance of **testing** the extraction sheet on a sample of articles to check that it matches the content of the studies

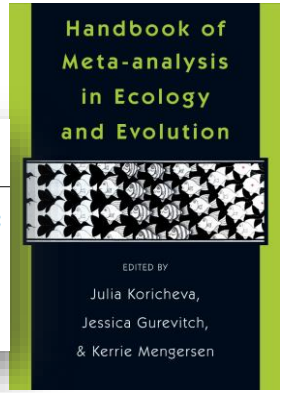
Document the work/decisions (transparency, repeatability)

Decide what to do in case of **missing information** ("missing data", contact the authors, imputation)


13

Recovering Missing or Partial Data from Studies: A Survey of Conversions and Imputations for Meta-analysis

Marc J. Lajeunesse



Consistency check

To be sure that data extraction is objective / robust:

- Data extraction of each study carried out **independently** by 2 people

- if several people share the work, **check the consistency of data extraction** between people on a sample before starting the actual extraction (discuss any disagreements)

- if only 1 person, have someone to check a sample of the extracted data at the start of the work (discuss any disagreements)

Example of extraction sheet

A study = a taxon × an exposure × an outcome

Case study level: ex. several concentrations of a chemical

IDdata	ID_map	author	...	taxon	Population_descri	Life_stage	Type_system	Tempera	pH
ScreenTA_9680	880	Cantin, N.E.	...	*Acropora tenuis*	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	880	Cantin, N.E.	...	*Acropora tenuis*	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	884	Cantin, N.E.	...	*Acropora valida*	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	884	Cantin, N.E.	...	*Acropora valida*	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	889	Cantin, N.E.	...	*Pocillopora damicornis*	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	889	Cantin, N.E.	...	*Pocillopora damicornis*	Colonies	Adult	500 L outdoor tank	27.5	NA

Treatment_description	Control_description	Solvent	Concentration_nom	Concentration_eff	Duration	Measured_variable	Time_after
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	67 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	67 days	Symbiodinium density / total protein	NA

Example of extraction sheet

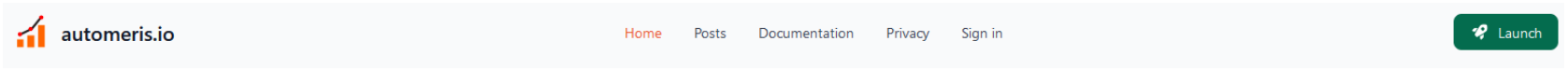
Treatment_description	Control_description	Solvent	Concentration_nom	Concentration_eff	Duration	Measured_variable	Time_after
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	67 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	67 days	Symbiodinium density / total protein	NA

Metaanalyse_data	unit	ID_experiment	ID_case	ID_common_control	N_c	Mean_c	Type_variation_c	Variation_c	N_t	Mean_t	Type_variation_t	Variation_t
OK (Fig3, SE, n=6)	x 10^6 / mg protein	1	3	1	6	4.2	sd	1.71464281	6	3.5142857	sd	0.979795897
OK (Fig3, SE, n=6)	x 10^6 / mg protein	1	4	1	6	4.2	sd	1.71464281	6	3.6	sd	0.524890659
OK (Fig3, SE, n=6)	x 10^6 / mg protein	2	7	2	6	0.928571	sd	0.45490523	6	1.4142857	sd	0.979795897
OK (Fig3, SE, n=6)	x 10^6 / mg protein	2	8	2	6	0.928571	sd	0.45490523	6	1.3142857	sd	0.699854212
OK (Fig3, SE, n=6)	x 10^6 / mg protein	3	11	3	6	1.714285	sd	0.69985421	6	2.2285714	sd	0.699854212
OK (Fig3, SE, n=6)	x 10^6 / mg protein	3	12	3	6	1.714285	sd	0.69985421	6	0.9142857	sd	0.244948974

Method_extraction	Source	Comment_extract	Name_data_extractio
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO

Tools for extracting data from figures

WebPlotDigitizer https://automeris.io/WebPlotDigitizer/



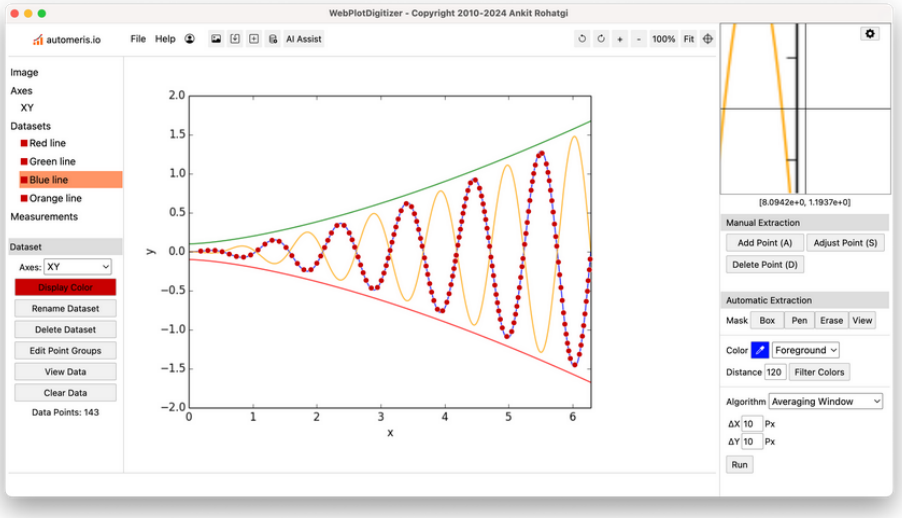
Extract data from charts

A large quantity of useful data is locked away in images of data visualizations. WebPlotDigitizer is a computer vision assisted software that helps extract numerical data from images of a variety of data visualizations.

Launch v5

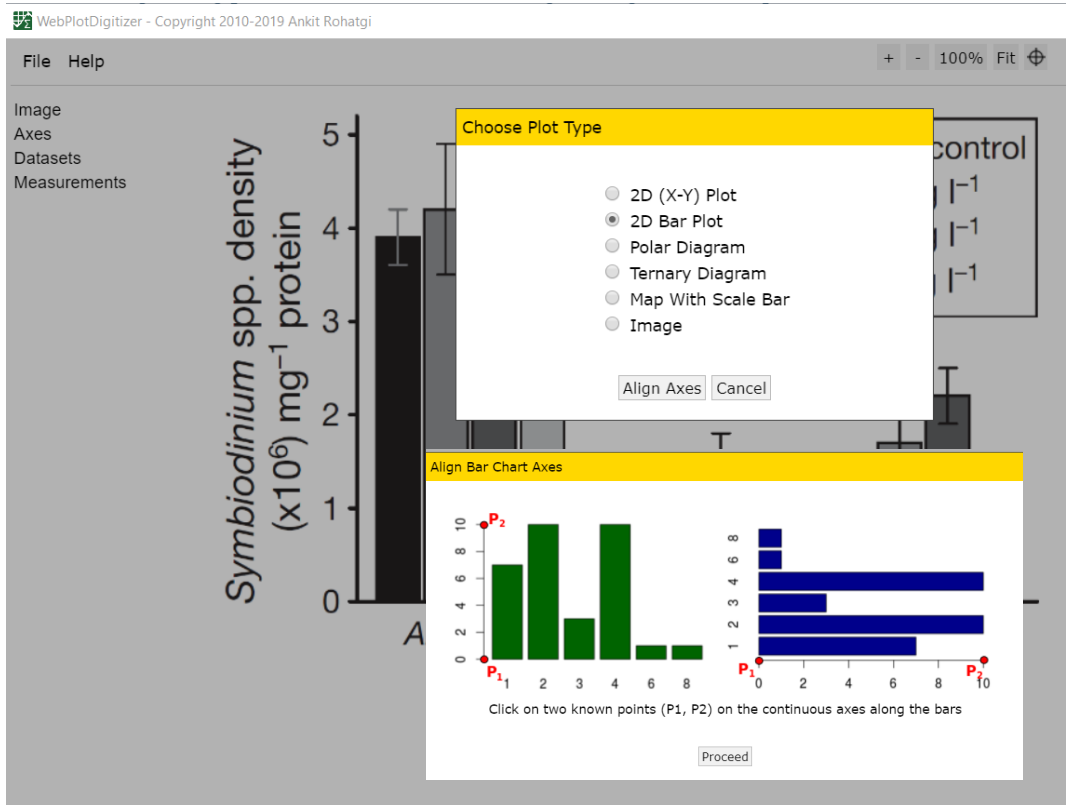
Learn more about the latest version (v5)

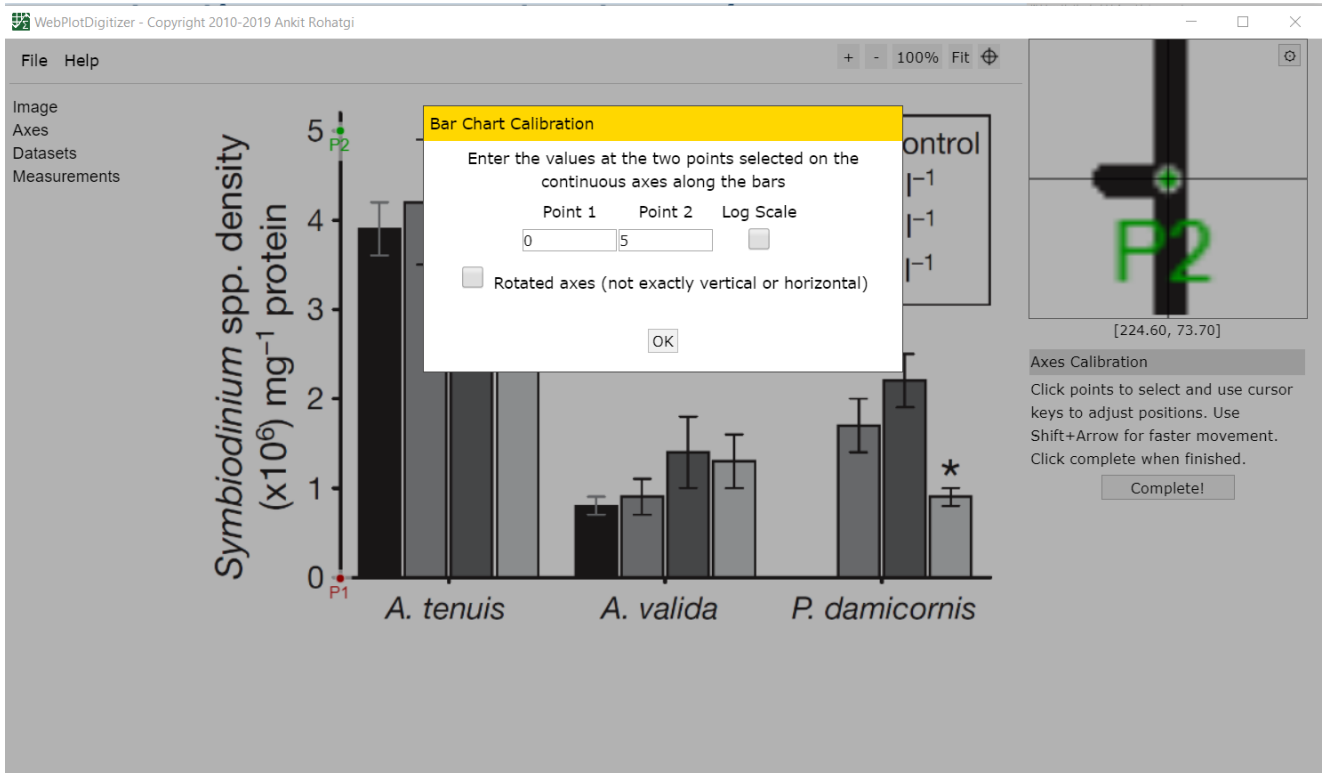
Access archived v4



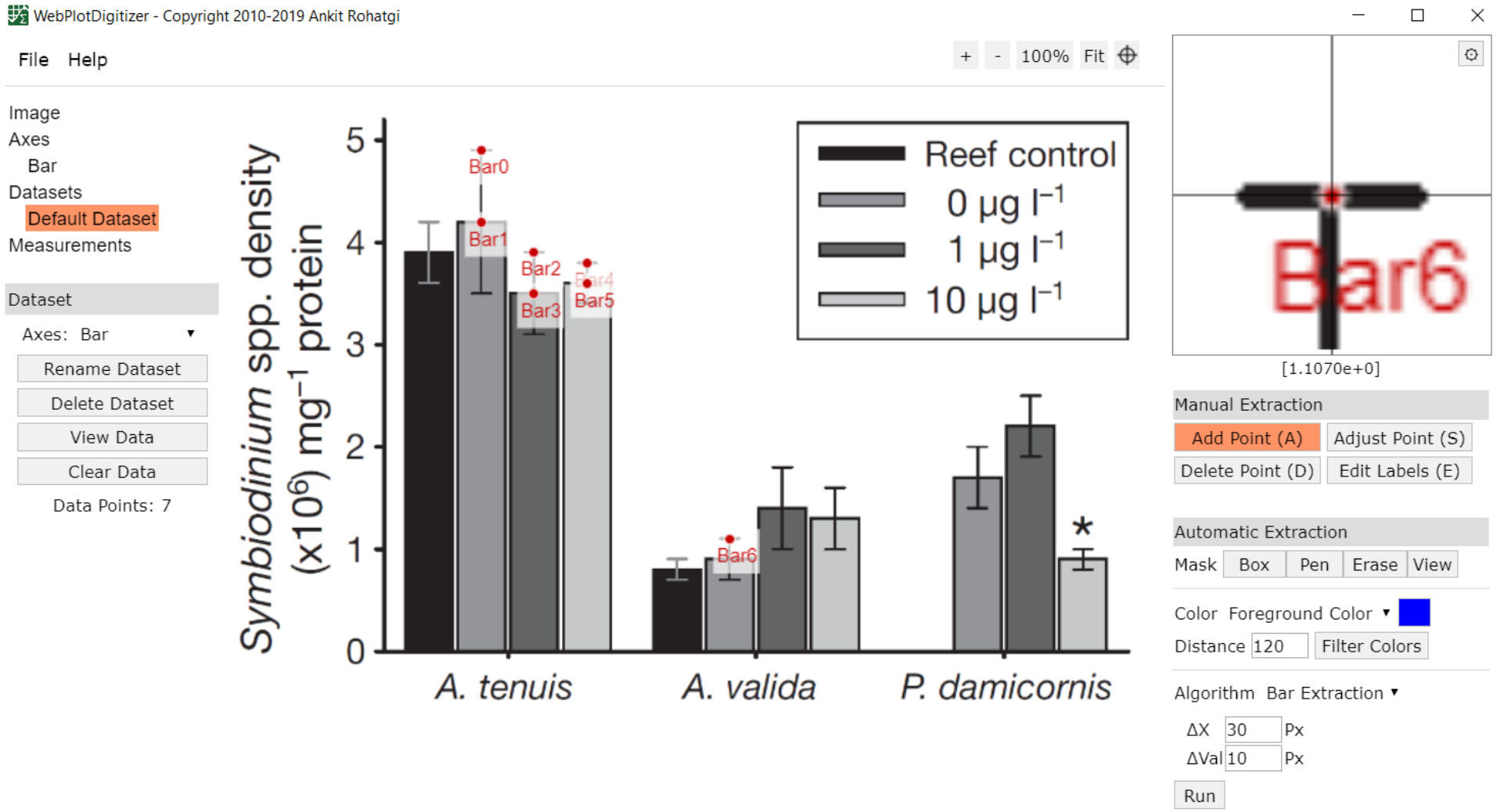
New version v5 (experimental) with AI Assist extraction

Tools for extracting data from figures



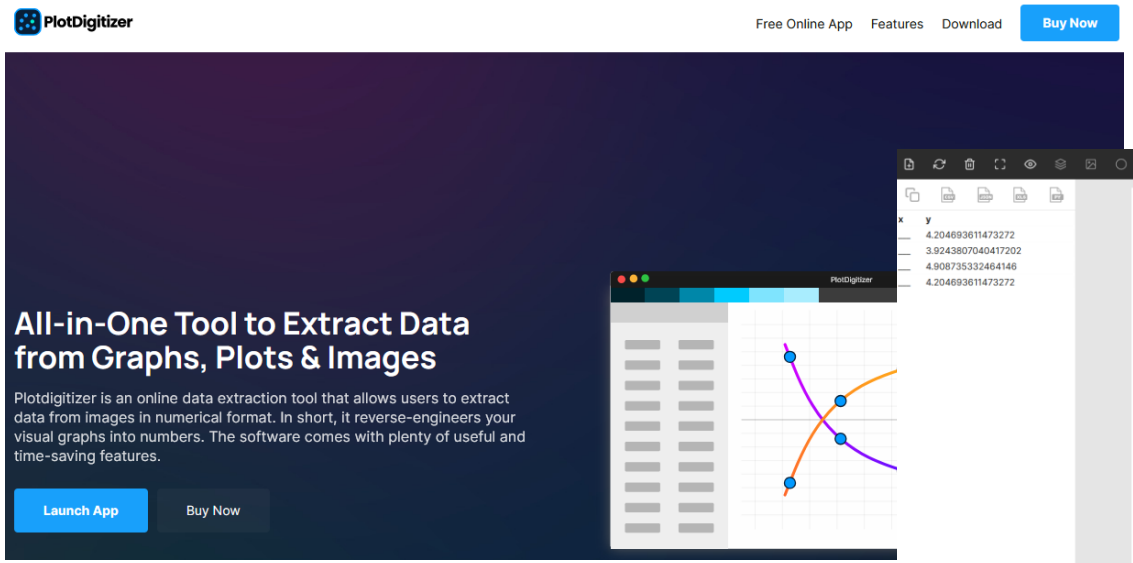


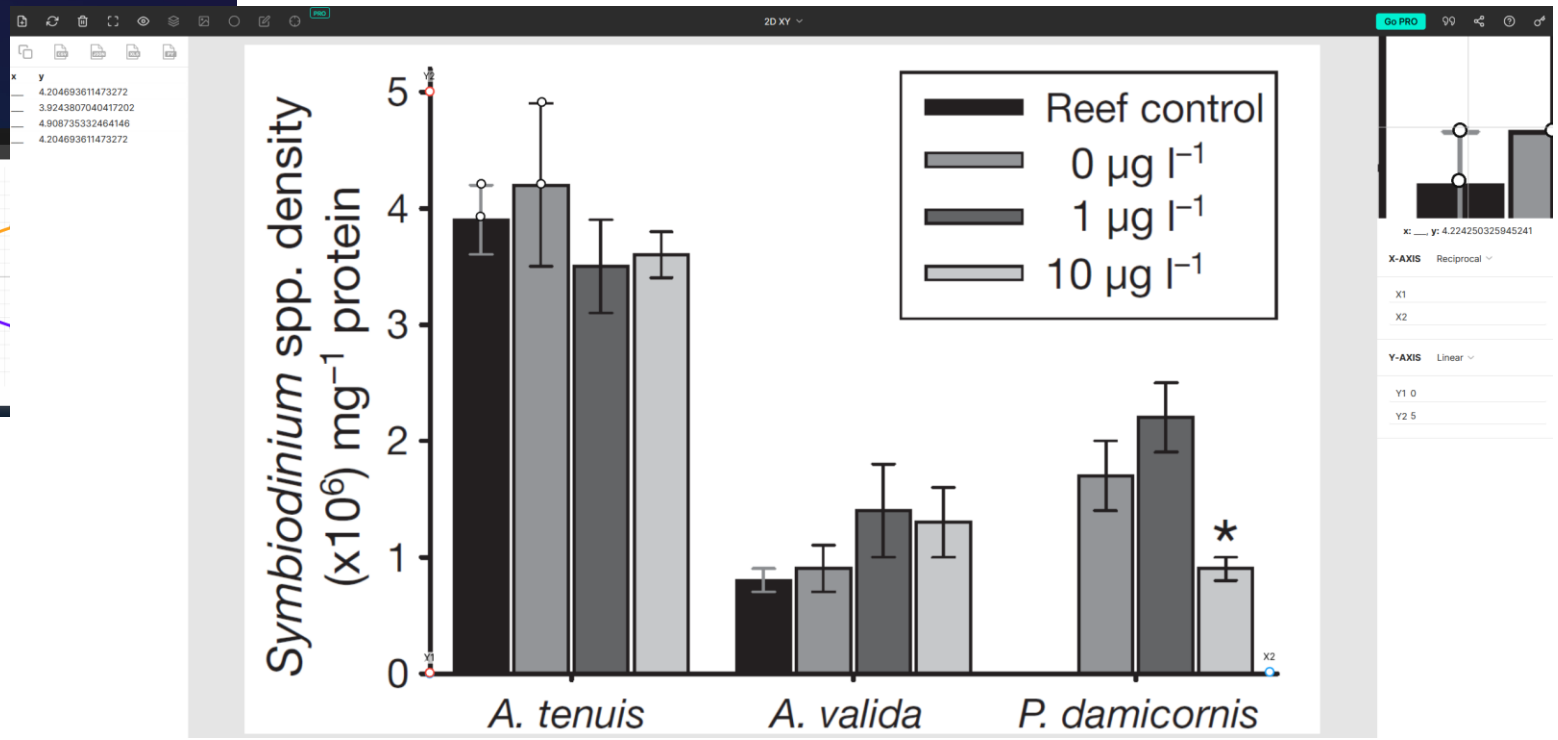
Tools for extracting data from figures



Tools for extracting data from figures

Plotdigitizer https://plotdigitizer.com/





Tools for extracting data from figures: metaDigitise

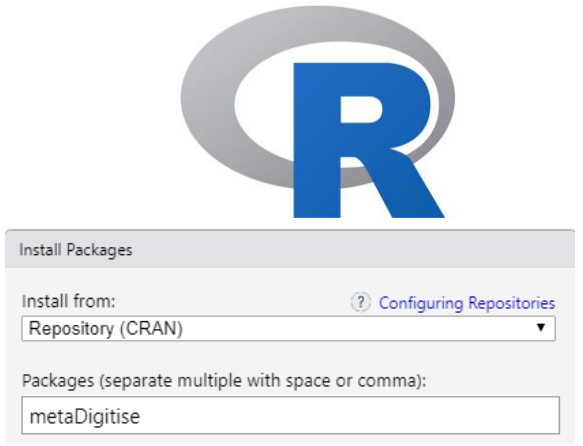
Received: 13 July 2018 | Accepted: 12 October 2018
DOI: 10.1111/2041-210X.13118

APPLICATION

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Reproducible, flexible and high-throughput data extraction from primary literature: The METADIGITISE R package

Joel L. Pick | Shinichi Nakagawa | Daniel W. A. Noble

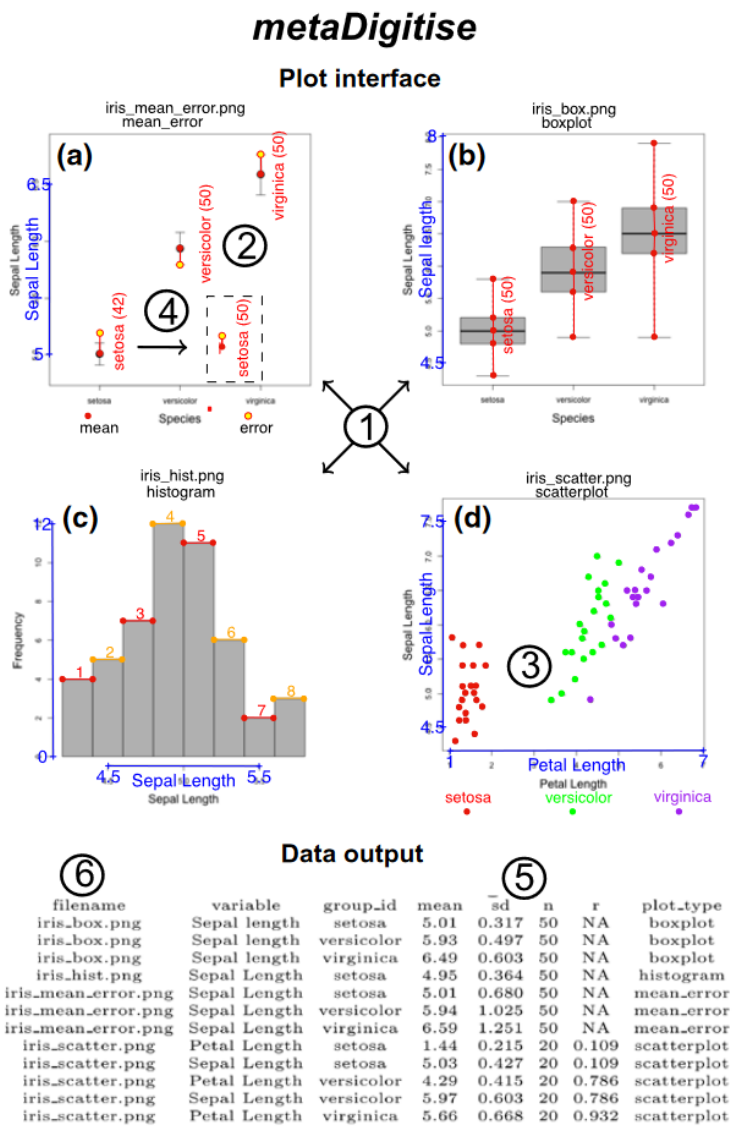


(+) possible to save, trace and modify data extraction

(-) no zoom

https://cran.r-project.org/web/packages/metaDigitise/vignettes/metaDigitise.html

Tools for extracting data from figures: metaDigitise



filename	variable	group_id	mean	sd	n	r	plot_type
iris_box.png	Sepal length	setosa	5.01	0.317	50	NA	boxplot
iris_box.png	Sepal length	versicolor	5.93	0.497	50	NA	boxplot
iris_box.png	Sepal length	virginica	6.49	0.603	50	NA	boxplot
iris_hist.png	Sepal Length	setosa	4.95	0.364	50	NA	histogram
iris_mean_error.png	Sepal Length	setosa	5.01	0.680	50	NA	mean_error
iris_mean_error.png	Sepal Length	versicolor	5.94	1.025	50	NA	mean_error
iris_mean_error.png	Sepal Length	virginica	6.59	1.251	50	NA	mean_error
iris_scatter.png	Petal Length	setosa	1.44	0.215	20	0.109	scatterplot
iris_scatter.png	Sepal Length	setosa	5.03	0.427	20	0.109	scatterplot
iris_scatter.png	Petal Length	versicolor	4.29	0.415	20	0.786	scatterplot
iris_scatter.png	Sepal Length	versicolor	5.97	0.603	20	0.786	scatterplot
iris_scatter.png	Petal Length	virginica	5.66	0.668	20	0.932	scatterplot

FUNCTIONALITY

1 *Different plot types*

Capable of handling A) mean error plots, B) boxplots, C) histograms and D) scatterplots

2 *Entry of Metadata*

Enter sample sizes variable and group names while digitising that are displayed on plot

3 *Grouped Data*

Enter as many groups as needed to capture descriptive statistics for sub-samples of data

4 *Digitise, edit or replot digitisations*

Simple user interface to guide user. Can digitise new images, edit digitisations or easily replot previous digitisations and metadata by cycling through images or choosing specific images

5 *Summarising data*

Get descriptive statistics automatically calculated for all plot types or use raw x,y data, if desired

6 *Multiple image processing*

Process as many images at once as needed and of varying types efficiently and quickly. New plots automatically plotted for digitisation

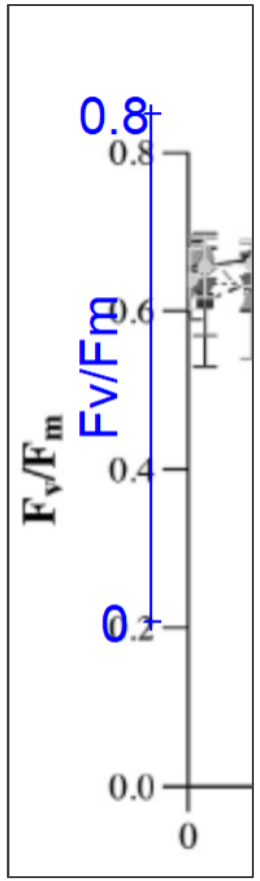
Tools for extracting data from figures: metaDigitise

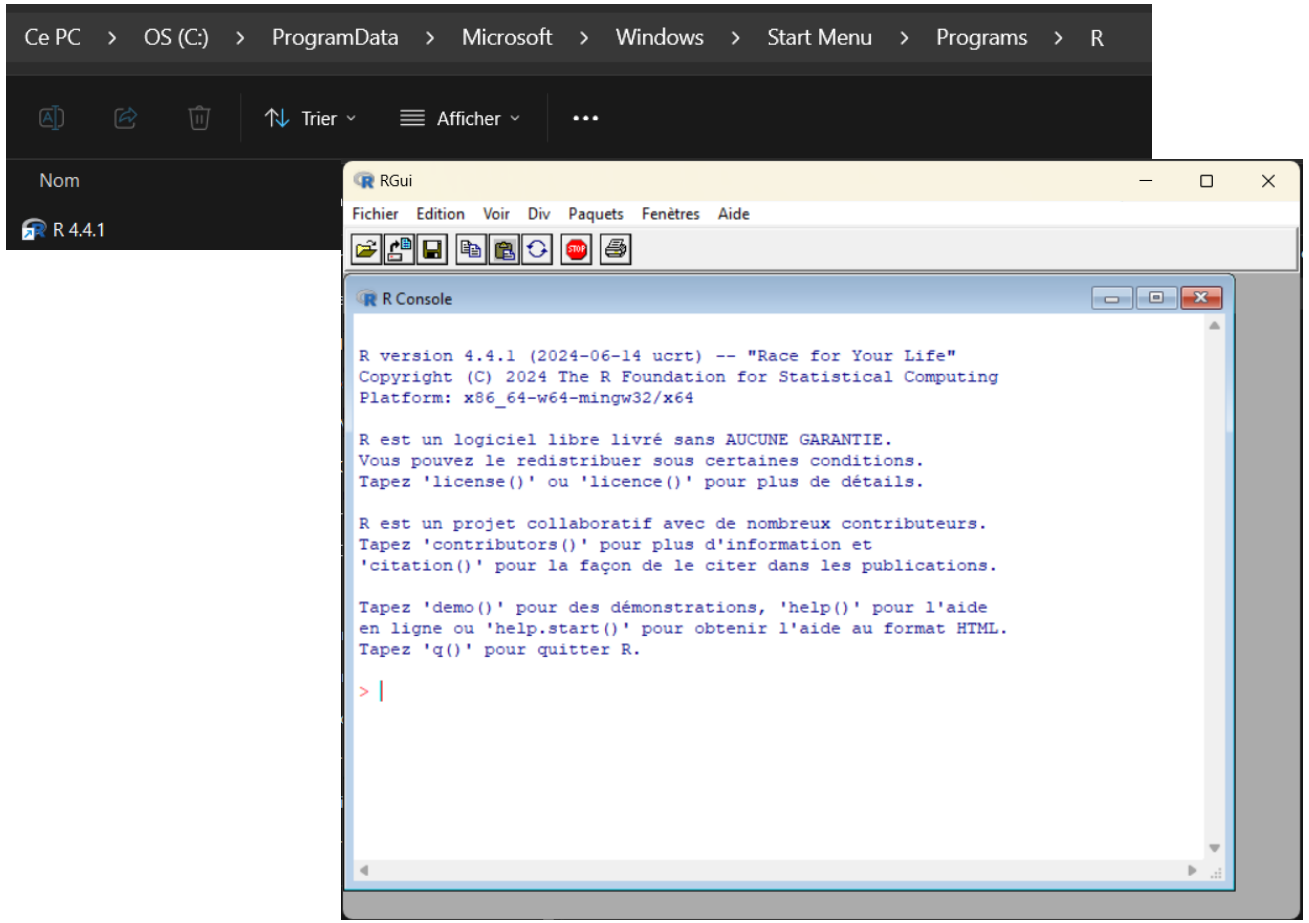
```
> dat <- metaDigitise(dir = "./figs")

Do you want to...

1: Process new images
2: Import existing data
3: Edit existing data

Sélection : 1

Are all plot types Different or the Same? (d/s)
s

Please specify the plot_type as either:

 m: Mean and error
 b: Box plot
 s: Scatter plot
 h: Histogram

 m

**** NEW PLOT ****


mean_error and boxplots should be vertically orientated
       _
       |
  I.E. o    NOT  |-o-|
       |
       _

If they are not then chose flip to correct this.

If figures are wonky, chose rotate.

Otherwise chose continue

Flip, rotate or continue (f/r/c)
c
```

```
What is the y variable?
Fv/Fm

On the Figure, click IN ORDER:
      y1, y2


    Step 1 ----> Click on known value on y axis - y1
  |
  |
  |
  |
  y1
  |_____________________________


    Step 2 ----> Click on another known value on y axis - y2
  |
  y2
  |
  |
  |
  |_____________________________


What is the value of y1 ?
0

What is the value of y2 ?
0.8

Re-calibrate? (y/n)
y
```

Tools for extracting data from figures: metaDigitise

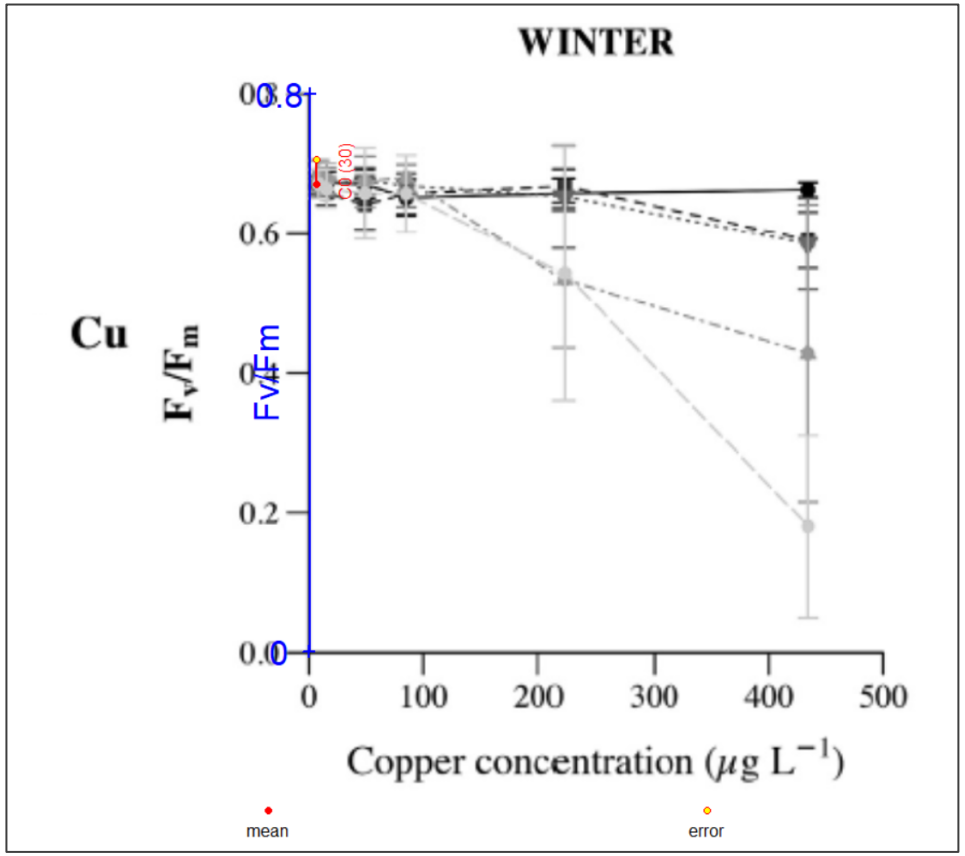
! If display problem -> exit Rstudio -> use Rgui





Tools for extracting data from figures: metaDigitise

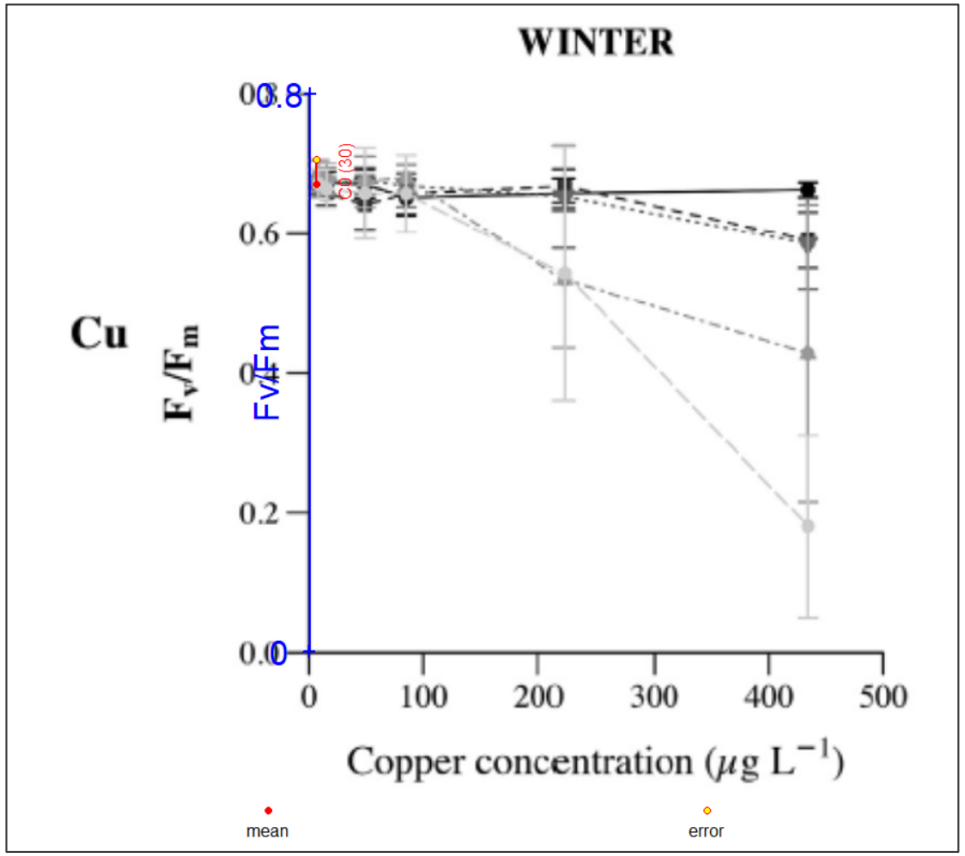
```
Do you know sample sizes? (y/n)
y

If there are multiple groups, enter unique group identifiers (otherwise press enter)
Group identifier:
C0

Group sample size:
30

Click on Error Bar, followed by the Mean

Add group, Edit Group, Delete group or Finish plot? (a/e/d/f)
```



Tools for extracting data from figures: metaDigitise

```
Do you know sample sizes? (y/n)
y

If there are multiple groups, enter unique group identifiers (otherwise press enter)
Group identifier:
C0

Group sample size:
30

Click on Error Bar, followed by the Mean

Add group, Edit Group, Delete group or Finish plot? (a/e/d/f)
```

```
Add group, Edit Group, Delete group or Finish plot? (a/e/d/f)
f

Type of error (se, CI95, sd):
sd


Do you want continue: 1 plots out of 2 plots remaining (y/n)
y
```

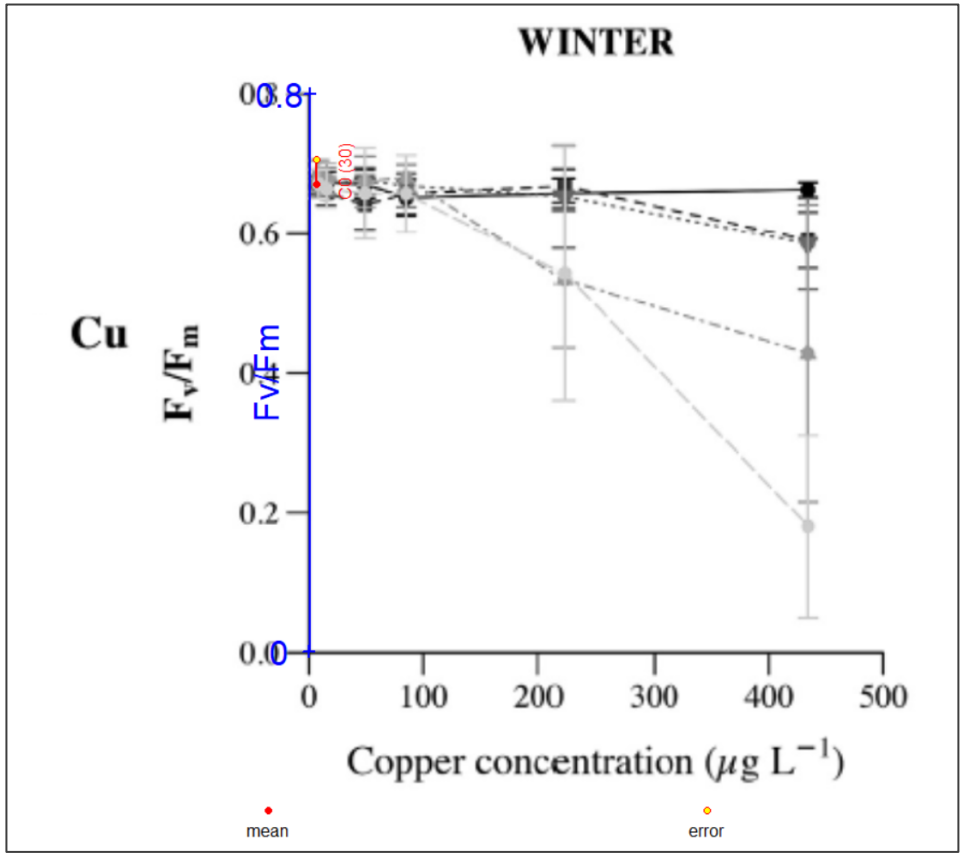


Tools for extracting data from figures: metaDigitise

```
Do you know sample sizes? (y/n)
y

If there are multiple groups, enter unique group identifiers (otherwise press enter)
Group identifier:
C0

Group sample size:
30

Click on Error Bar, followed by the Mean

Add group, Edit Group, Delete group or Finish plot? (a/e/d/f)
```

```
Add group, Edit Group, Delete group or Finish plot? (a/e/d/f)
f

Type of error (se, CI95, sd):
sd


Do you want continue: 1 plots out of 2 plots remaining (y/n)
y
```

```
Add group, Edit Group, Delete group or Finish plot? (a/e/d/f)
f

Type of error (se, CI95, sd):
sd

Congratulations! Looks like you have finished digitising all figures in this directory.
>
```



TD

Question: estimate the effects of chemicals on the photosynthetic performance (***maximum quantum yield***, **Fv/Fm)** of tropical reef-building corals

A **study** = combination of a taxon × an exposure × an outcome

A **case study** = one tested concentration-duration

Effect size = mean standardized difference

$$d = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} J,$$

where $J = 1 - \frac{3}{4(n_1 + n_2 - 2) - 1}$

<u>If monitoring over time, extraction of the longest duration of exposure</u>

TD

Files:

data_TD.xlsx (sheet 5)

Hedouin2016_Improving.pdf

Hedouin2016_Fig8_SummerCu.png, Hedouin2018_Fig8_WinterCu.png (folder « figs »)

Exercise: extract data using metaDigitise (15-20 min)

Prepare the figures with the data to be extracted (screenshot -> .png), put them in a "figs" folder in the "TD" working directory

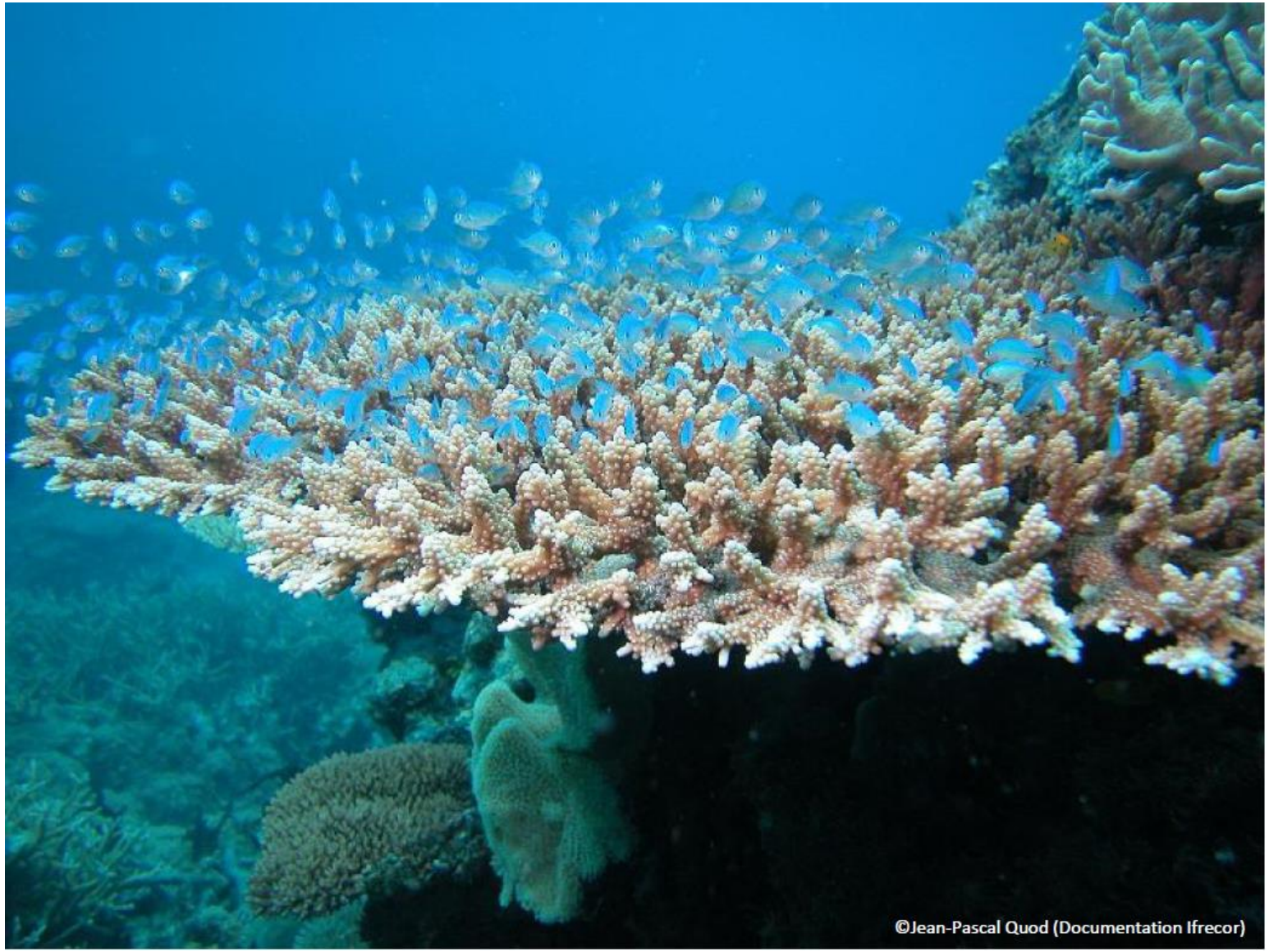
In R: Install & load metaDigitise package `library(metaDigitise)`

Set working directory `setwd("~/TD/TD_extraction")`

```
dat <- metaDigitise(dir = "./figs")


write.csv2(dat, "extracted_data_TD_metadigitise.csv", quote=F, row.names=F)
```

Extract data into data_TD.xlsx (sheet 5)

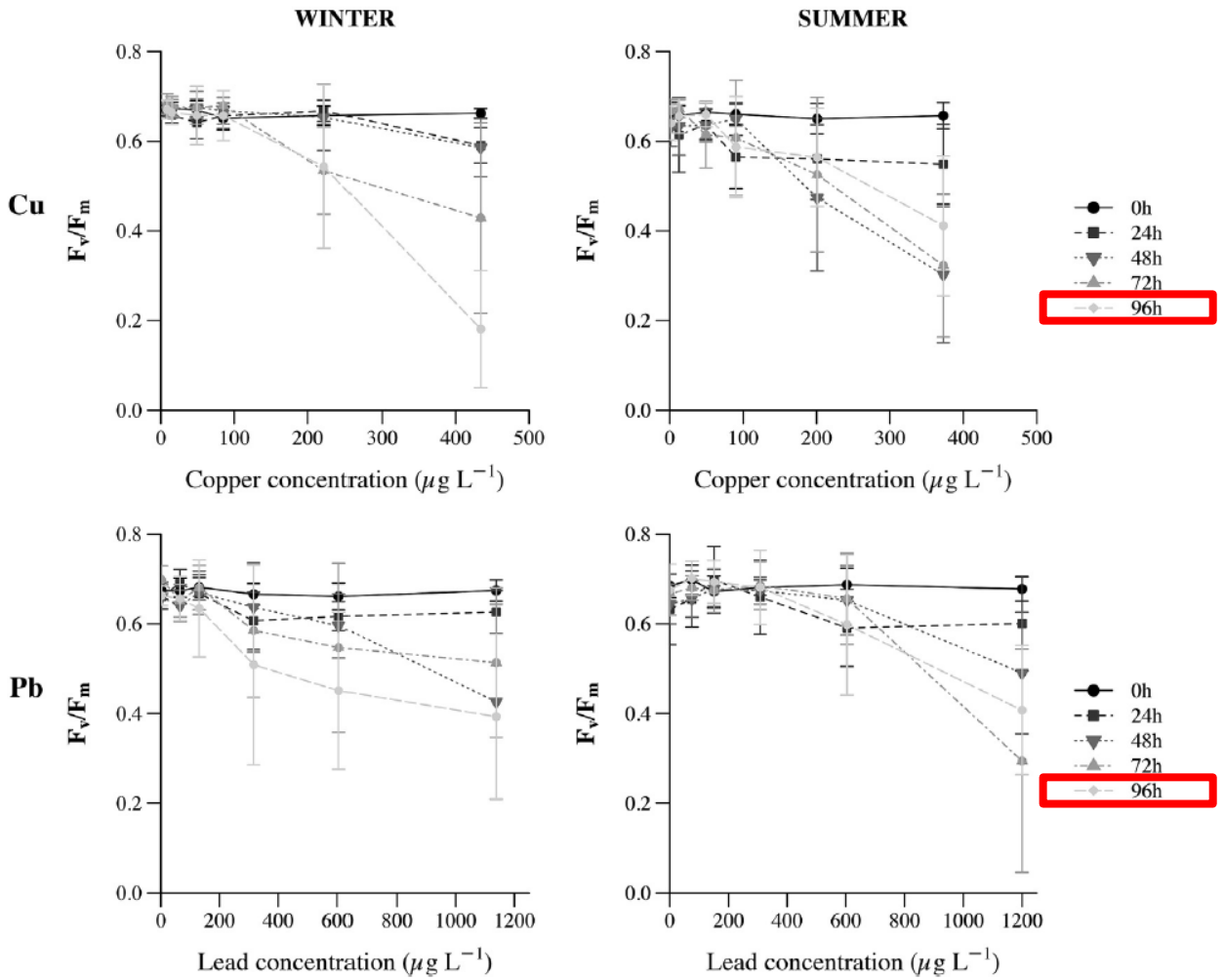
TD



Hédouin et al. 2016

Table 1
Nominal versus measured concentrations (mean ± standard deviation, n = 3) of Cu and Pb (μg L^{-1}) used in the toxicity experiments for the adult *Pocillopora damicornis* corals.

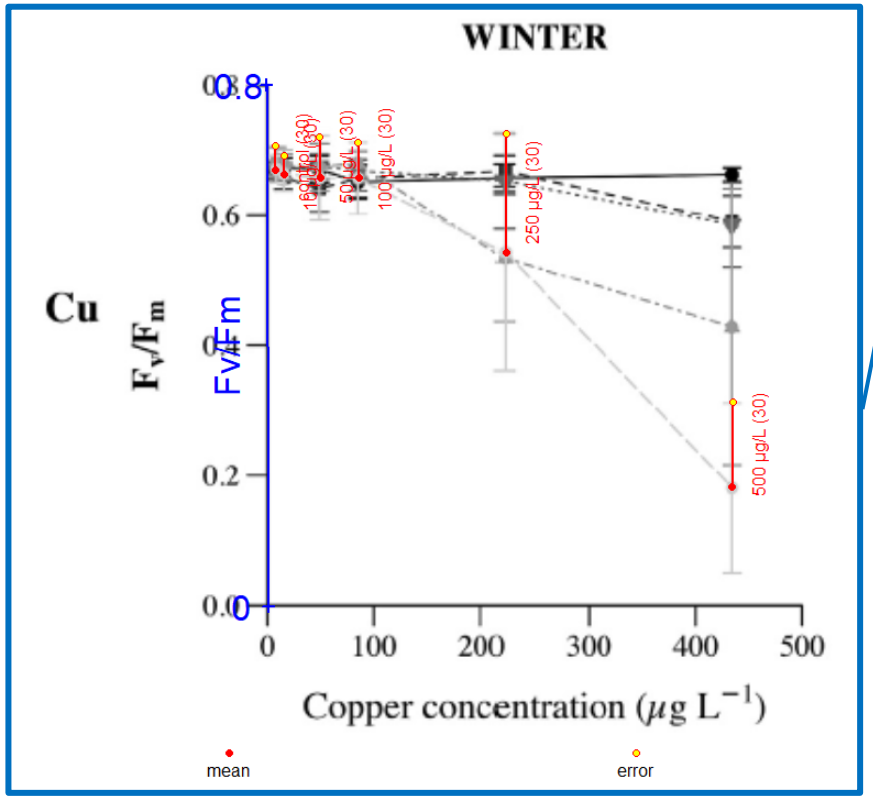
Condition	Concentration (μg L^{-1})		
	Nominal	Measured – summer	Measured-Winter
Cu			
Control – C_0	0	1.5 ± 0.6	8.1 ± 0.6
Concentration 1 – C_1	10	12.6 ± 2.8	15.0 ± 1.3
Concentration 2 – C_2	50	49.2 ± 11.3	48.9 ± 5.0
Concentration 3 – C_3	100	90.0 ± 15.3	85.2 ± 16.2
Concentration 4 – C_4	250	206 ± 33.6	222 ± 29.4
Concentration 5 – C_5	500	378 ± 49.8	434 ± 72.6
Pb			
Control- C_0	0	0.5 ± 0.7	1.9 ± 3.0
Concentration 1 – C_1	80	75.6 ± 4.2	67.9 ± 7.8
Concentration 2 – C_2	160	151 ± 8.2	133 ± 23.1
Concentration 3 – C_3	320	308 ± 10.0	316 ± 30.6
Concentration 4 – C_4	640	604 ± 37.1	605 ± 34.3
Concentration 5 – C_5	1280	1200 ± 92.2	1138 ± 57.4

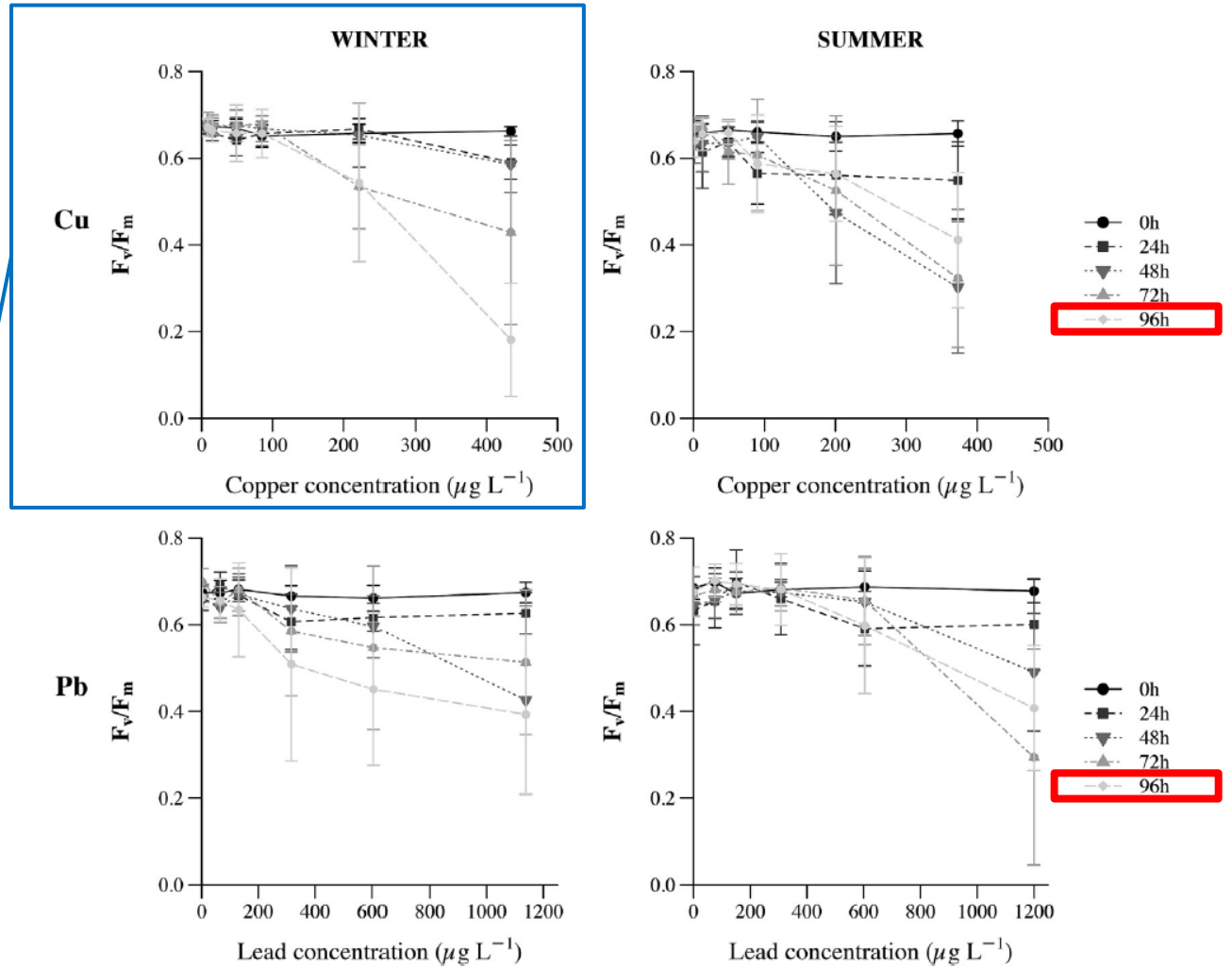


N = 30
SD

Fig. 8. Dark-adapted quantum yield (F_v/F_m, mean ± standard deviation, n = 30 nubbins per condition, -pooled data-) of *Pocillopora damicornis* corals exposed for 96 h to a range of dissolved Cu and Pb concentration in the summer and winter seasons.

Hédouin et al. 2016





N = 30
SD

Fig. 8. Dark-adapted quantum yield (F_v/F_m, mean ± standard deviation, n = 30 nubbins per condition, -pooled data-) of *Pocillopora damicornis* corals exposed for 96 h to a range of dissolved Cu and Pb concentration in the summer and winter seasons.

TD

Hédouin et al. 2016

```
> dat
                        filename variable group_id      mean         sd  n  r          se error_type  plot_type
1  Hedouin2016_Fig8_SummerCu.png    Fv/Fm       C0 0.6396917 0.04007707 30 NA 0.007317039         sd mean_error
2  Hedouin2016_Fig8_SummerCu.png    Fv/Fm       C1 0.6551060 0.03699422 30 NA 0.006754190         sd mean_error
3  Hedouin2016_Fig8_SummerCu.png    Fv/Fm       C2 0.6612717 0.02928709 30 NA 0.005347067         sd mean_error
4  Hedouin2016_Fig8_SummerCu.png    Fv/Fm       C3 0.5872832 0.11560694 30 NA 0.021106842         sd mean_error
5  Hedouin2016_Fig8_SummerCu.png    Fv/Fm       C4 0.5657033 0.10944123 30 NA 0.019981144         sd mean_error
6  Hedouin2016_Fig8_SummerCu.png    Fv/Fm       C5 0.4100193 0.15722543 30 NA 0.028705306         sd mean_error
7  Hedouin2018_Fig8_WinterCu.png    Fv/Fm       C0 0.6707692 0.03538462 30 NA 0.006460317         sd mean_error
8  Hedouin2018_Fig8_WinterCu.png    Fv/Fm       C1 0.6646154 0.02769231 30 NA 0.005055901         sd mean_error
9  Hedouin2018_Fig8_WinterCu.png    Fv/Fm       C2 0.6584615 0.06307692 30 NA 0.011516218         sd mean_error
10 Hedouin2018_Fig8_WinterCu.png    Fv/Fm       C3 0.6553846 0.05538462 30 NA 0.010111801         sd mean_error
11 Hedouin2018_Fig8_WinterCu.png    Fv/Fm       C4 0.5400000 0.18615385 30 NA 0.033986887         sd mean_error
12 Hedouin2018_Fig8_WinterCu.png    Fv/Fm       C5 0.1769231 0.13230769 30 NA 0.024155969         sd mean_error
>
```

TD

```
> dat
                        filename variable group id      mean        sd  n  r          se error_type  plot_type
1  Hedouin2016_Fig8_SummerCu.png     Fv/Fm      C0 0.6396917 0.04007707 30 NA 0.007317039         sd mean_error
2  Hedouin2016_Fig8_SummerCu.png     Fv/Fm      C1 0.6551060 0.03699422 30 NA 0.006754190         sd mean_error
3  Hedouin2016_Fig8_SummerCu.png     Fv/Fm      C2 0.6612717 0.02928709 30 NA 0.005347067         sd mean_error
4  Hedouin2016_Fig8_SummerCu.png     Fv/Fm      C3 0.5872832 0.11560694 30 NA 0.021106842         sd mean_error
5  Hedouin2016_Fig8_SummerCu.png     Fv/Fm      C4 0.5657033 0.10944123 30 NA 0.019981144         sd mean_error
6  Hedouin2016_Fig8_SummerCu.png     Fv/Fm      C5 0.4100193 0.15722543 30 NA 0.028705306         sd mean_error
7  Hedouin2018_Fig8_WinterCu.png     Fv/Fm      C0 0.6707692 0.03538462 30 NA 0.006460317         sd mean_error
8  Hedouin2018_Fig8_WinterCu.png     Fv/Fm      C1 0.6646154 0.02769231 30 NA 0.005055901         sd mean_error
9  Hedouin2018_Fig8_WinterCu.png     Fv/Fm      C2 0.6584615 0.06307692 30 NA 0.011516218         sd mean_error
10 Hedouin2018_Fig8_WinterCu.png     Fv/Fm      C3 0.6553846 0.05538462 30 NA 0.010111801         sd mean_error
11 Hedouin2018_Fig8_WinterCu.png     Fv/Fm      C4 0.5400000 0.18615385 30 NA 0.033986887         sd mean_error
12 Hedouin2018_Fig8_WinterCu.png     Fv/Fm      C5 0.1769231 0.13230769 30 NA 0.024155969         sd mean_error
>
```

L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
Concentration_nom	**Concentration_eff**	**Duration**	**Measured_outcome**	**ID_experiment**	**ID_case**	**N_c**	**Mean_c**	**Type_variatic**	**Variation_c**	**N_t**	**Mean_t**	**Type_variatic**	**Variation_t**	**Method_extr**	**Source**
10 µg/L	12.6 µg/L	96h	Fv/Fm (Maximum dark-	Summer_Exp	1	30	0,63969171	sd	0,04007707	30	0,65510597	sd	0,03699422	Figure	Figure 8
50 µg/L	49.2 µg/L	96h	Fv/Fm (Maximum dark-	Summer_Exp	2	30	0,63969171	sd	0,04007707	30	0,66127168	sd	0,02928709	Figure	Figure 8
100 µg/L	90 µg/L	96h	Fv/Fm (Maximum dark-	Summer_Exp	3	30	0,63969171	sd	0,04007707	30	0,58728324	sd	0,11560694	Figure	Figure 8
250 µg/L	206 µg/L	96h	Fv/Fm (Maximum dark-	Summer_Exp	4	30	0,63969171	sd	0,04007707	30	0,56570328	sd	0,10944123	Figure	Figure 8
500 µg/L	378 µg/L	96h	Fv/Fm (Maximum dark-	Summer_Exp	5	30	0,63969171	sd	0,04007707	30	0,41001927	sd	0,15722543	Figure	Figure 8
10 µg/L	15 µg/L	96h	Fv/Fm (Maximum dark-	Winter_Exp	6	30	0,67076923	sd	0,03538462	30	0,66461538	sd	0,02769231	Figure	Figure 8
50 µg/L	48.9 µg/L	96h	Fv/Fm (Maximum dark-	Winter_Exp	7	30	0,67076923	sd	0,03538462	30	0,65846154	sd	0,06307692	Figure	Figure 8
100 µg/L	85.2 µg/L	96h	Fv/Fm (Maximum dark-	Winter_Exp	8	30	0,67076923	sd	0,03538462	30	0,65538462	sd	0,05538462	Figure	Figure 8
250 µg/L	222 µg/L	96h	Fv/Fm (Maximum dark-	Winter_Exp	9	30	0,67076923	sd	0,03538462	30	0,54	sd	0,18615385	Figure	Figure 8
500 µg/L	434 µg/L	96h	Fv/Fm (Maximum dark-	Winter_Exp	10	30	0,67076923	sd	0,03538462	30	0,17692308	sd	0,13230769	Figure	Figure 8