

# R Markdown Tutorial

Setta

Last compiled on 01th Mar 2024

## Goal:

This RMarkdown tutorial will walk through different components of an R Markdown file and how to process data, create figures and save in one pdf or html file with notes and observations.

## General Notes about R project & R Markdown Files:

### R projects:

Opening the R project file (R\_MarkdownTutorial) I created or creating your own R project file will allow you to save figures, data, and output files for later use and sharing with others.

In the R project R Markdown Tutorial there is an example dataset (ExampleData.csv), a picture to add to the R Markdown file (this can be a .jpg or .png format) and the R Markdown File we will be working with.

### R Markdown files:

R markdown files are an easy way to take notes during data analysis including online tutorials you are working with, observations, and results. These will be saved as an output file of your choosing (html or pdf). Afterwards, you will have your data analysis methods stored in one file along with figures you produced during data analysis.

Notice at the top of this file we've saved the name of the R markdown file, the creator of the file and the date it was last compiled (to html or pdf format).

R markdown works by creating code chunks in between text in your script. You will notice the R markdown chunks by the “`````” and curly brackets at the beginning of a set of code. Here we are using r code, so you will see an r after the brackets but you can insert other types of code. If you just want to run a chunk of code before compiling, press the play button in the top right hand corner.

You can format text in between code chunks by changing the headers using the hash mark before text. The more hash marks indicate subheadings.

You can also create bullet points with asteriks and numbered lists. Below is an example of what I generally keep in an R markdown file during data analysis.

*Note* When you open a new R markdown file a short set of instructions is included, like the following:

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

## Set working directory:

The r code below sets the working directory and the knitr settings to wrap text in the output file. Notice we've added `include=FALSE` to the brackets, therefore this code chunk is not included in the R markdown file.

## Description of data and project:

### Dataset description:

Today we will be working with made up data (ExampleData.csv) for my diatom-diazotroph experiment growing cultures in three treatment types:

- WNH - 10 uM Ammonium
- WNO - 10 uM Nitrate
- CON - no added nitrate/ammonium

Notice the above list will create bullet points in your output file. Make sure to include an empty line between the bulleted list and the text above.

The example data also has two timepoints:

- 1) Initial Timepoint (T0)
- 2) Final Timepoint (TF)

Notice the above will produce a numbered list. Make sure to include an empty line between the list and the text above.

### Goal of analysis:

We are comparing the fluorescence and cell counts of the DDA for the final timepoint to see if they differ across treatments.

R markdown files also allows you to include images in case you have notes or calculations you'd like to keep with the associated code.

Here we are using a dataset for the DDA *Hemiaulus-Richelia* a diatom host (*Hemiaulus hauckii*) and nitrogen fixing cyanobacterial symbiont (*Richelia intracellularis*), seen in the image below.

Removed image to publish script on github.

The text inside the square brackets gives a caption to the image file included in parentheses.

### Load libraries

The r code below loads libraries needed for this tutorial. I also like to include a code chunk at the end to list of the version of the packages I'm using to make the methods write up for my manuscript easier.

```
# list of packages to install in case you don't have them.
# install.packages('BiocManager')
# BiocManager::install(c('tidyverse', 'knitr', 'ggplot2',
# 'viridis', 'kableExtra', 'dplyr', 'ggsci')) load
# packages:
library("tidyverse")
library("knitr")
library("kableExtra")
library("viridis")
library("dplyr")
library("ggsci")
```

Chunk output can be customized with knitr options, arguments set in the `{ }` of a chunk header. Above, we use five arguments:

- `include = FALSE` prevents code and results from appearing in the finished file. R Markdown still runs the code in the chunk, and the results can be used by other chunks.
- `echo = FALSE` prevents code, but not the results from appearing in the finished file. This is a useful way to embed figures.
- `message = FALSE` prevents messages that are generated by code from appearing in the finished file.
- `warning = FALSE` prevents warnings that are generated by code from appearing in the finished.
- `fig.cap = "..."` adds a caption to graphical results.

See the R Markdown Reference Guide for a complete list of knitr chunk options.

## Load Data

```
# read in data
DDA_CellCounts_RFU <- read_csv("ExampleData.csv")

# Change Flask, Timepoint, and Treatment to factors:
DDA_CellCounts_RFU$Flask <- as.factor(DDA_CellCounts_RFU$Flask)
DDA_CellCounts_RFU$Timepoint <- as.factor(DDA_CellCounts_RFU$Timepoint)
DDA_CellCounts_RFU$Treatment <- as.factor(DDA_CellCounts_RFU$Treatment)

# change order of Treatments:
DDA_CellCounts_RFU$Treatment <- factor(DDA_CellCounts_RFU$Treatment,
  levels = c("CON", "WNH", "WNO"))

# Keep only the final timepoint:
DDA_Final <- DDA_CellCounts_RFU %>%
  filter(Timepoint == "TF")
```

## Cell Count Statistics & Figure:

Compare cell counts between treatment at the final timepoint:

```
# Host Cell Summary:
group_by(DDA_Final, Treatment) %>%
  summarise(mean = mean(CellCount_perL, na.rm = TRUE), sd = sd(CellCount_perL,
    na.rm = TRUE))
```

```
## # A tibble: 3 x 3
##   Treatment mean    sd
##   <fct>      <dbl> <dbl>
## 1 CON        3197.  124.
## 2 WNH        4200   216.
## 3 WNO        3262.  206.
```

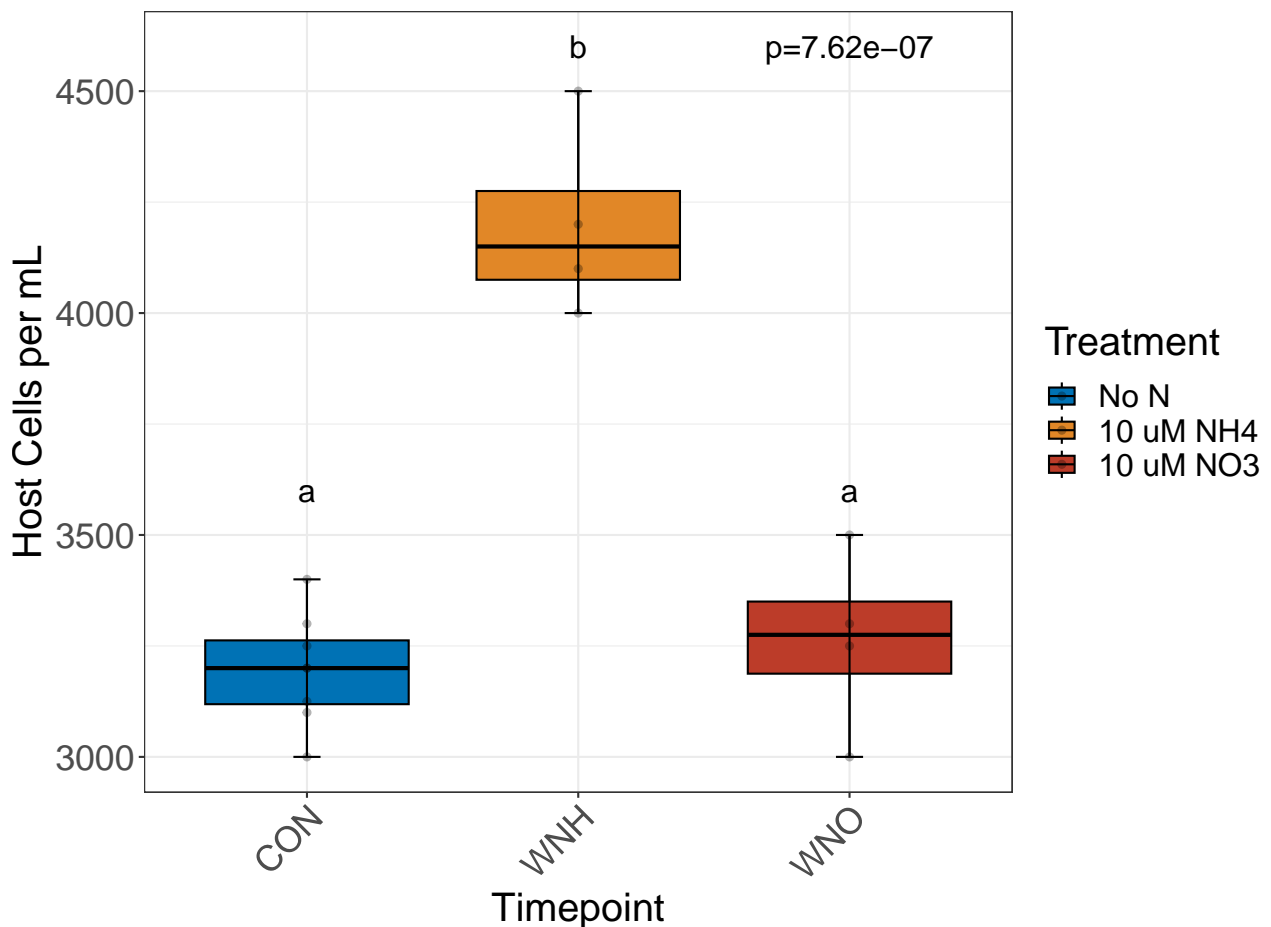
```
# Compute the analysis of variance
Counts.aov <- aov(CellCount_perL ~ Treatment, data = DDA_Final)
# Summary of the analysis
summary(Counts.aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment     2 2900039 1450020   50.28 7.62e-07 ***
## Residuals    13  374922    28840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Multiple comparison to see which group is significant:
TukeyHSD(Counts.aov)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = CellCount_permL ~ Treatment, data = DDA_Final)
##
## $Treatment
##          diff          lwr          upr      p adj
## WNH-CON 1003.125    728.5315 1277.7185 0.0000008
## WNO-CON   65.625   -208.9685   340.2185 0.8059617
## WNO-WNH -937.500 -1254.5732 -620.4268 0.0000082
```

When creating chunks of code in r markdown to use with figures, you need to make sure to change the figure numbers for each. If you accidentally include two of the same figure numbers the R code won't compile.



### RFU Statistics & Figure:

Compare cell counts between treatment at the final timepoint:

```
# Host Cell Summary:
group_by(DDA_Final, Treatment) %>%
  summarise(mean = mean(RFU_Turner, na.rm = TRUE), sd = sd(RFU_Turner,
    na.rm = TRUE))
```

```
## # A tibble: 3 x 3
##   Treatment mean    sd
##   <fct>      <dbl> <dbl>
## 1 CON        7.76 0.450
## 2 WNH       10.8  0.957
## 3 WNO        8.5  0.363

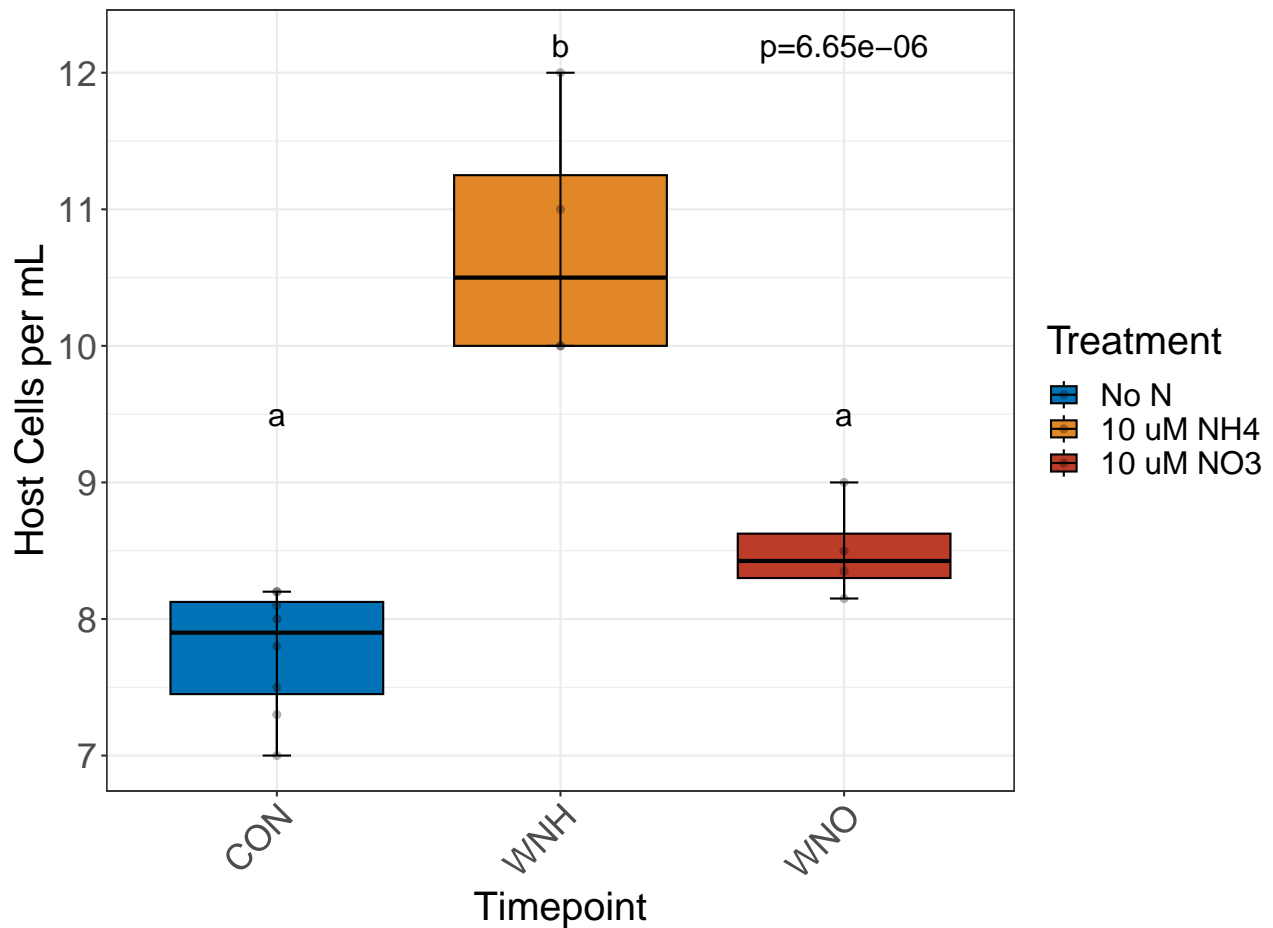
# Compute the analysis of variance
RFU.aov <- aov(RFU_Turner ~ Treatment, data = DDA_Final)
# Summary of the analysis
summary(RFU.aov)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment      2 24.001  12.000    34.18 6.65e-06 ***
## Residuals     13  4.564   0.351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Multiple comparison to see which group is significant:
TukeyHSD(RFU.aov)

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = RFU_Turner ~ Treatment, data = DDA_Final)
##
## $Treatment
##           diff             lwr             upr             p adj
## WNH-CON    2.9875    2.0294665    3.945534    0.0000046
## WNO-CON    0.7375   -0.2205335    1.695534    0.1435884
## WNO-WNH   -2.2500   -3.3562418   -1.143758    0.0003492
```

When creating chunks of code in r markdown to use with figures, you need to make sure to change the figure numbers for each. If you accidentally include two of the same figure numbers the R code won't compile.



## Conclusions:

Growth is significantly different in the added ammonium treatment (WNH) compared to the control and added nitrate treatments.

Next Steps: Were the cells able to take up added nutrients?

## Version of packages used for this R markdown tutorial:

```
packageVersion("rlang")
```

```
## [1] '1.1.1'
```

```
packageVersion("tidyverse")
```

```
## [1] '2.0.0'
```

```
packageVersion("geosphere")
```

```
## [1] '1.5.18'
```

```
packageVersion("knitr")
```

```
## [1] '1.44'
```

```
packageVersion("ggplot2")
```

```
## [1] '3.4.3'
```

```
packageVersion("viridis")
```

```
## [1] '0.6.4'
```

```
packageVersion("kableExtra")
```

```
## [1] '1.3.4.9000'
```

```
packageVersion("dplyr")
```

```
## [1] '1.1.3'
```