

# Workshop 9: Multivariate Analyses in R

## QCBS R Workshop Series

QCBS R Workshop Series<sup>1</sup>

Developed and maintained by the contributors of the QCBS R Workshop Series<sup>2</sup>

2021-02-16

<sup>1</sup>The QCBS R Workshop Series is part of the Québec Centre for Biodiversity Science, and is maintained by the series coordinators and graduate student, postdoctoral, and research professional members. <https://www.qcbs.ca>

<sup>2</sup>The contributors for this workshop can be accessed [here](#).



# Contents



## Part I

# QCBS R Workshop Series



# Preface

This series of 10 workshops walks participants through the steps required to use R for a wide array of statistical analyses relevant to research in biology and ecology. These open-access workshops were created by members of the QCBS both for members of the QCBS and the larger community.

The content of this workshop has been peer-reviewed by several QCBS members. If you would like to suggest modifications, please contact the current series coordinators, listed on the main Github page.

## **0.0.1 Series learning objectives**

## **0.0.2 Contributors**

## **0.0.3 Code of conduct**

The QCBS R Workshop Series and the QCBS R Symposium are venues dedicated to providing a welcoming and supportive environment for all people, regardless of background or identity.

Participants, presenters and organizers of the workshop series and other related activities accept this Code of Conduct when being present at any workshop-related activities.

We do not tolerate behaviour that is disrespectful or that excludes, intimidates, or causes discomfort to others.

We do not tolerate discrimination or harassment based on characteristics that include, but are not limited to, gender identity and expression, sexual orientation, disability, physical appearance, body size, citizenship, nationality, ethnic or social origin, pregnancy, familial status, genetic information, religion or belief (or lack thereof), membership of a national minority, property, age, education, socio-economic status, technical choices, and experience level.

It applies to all spaces managed by or affiliated with the workshop, including, but not limited to, workshops, email lists, and online forums such as GitHub, Slack and Twitter.

**0.0.3.0.1 Expected behaviour** All participants are expected to show respect and courtesy to others. All interactions should be professional regardless of platform: either online or in-person. In order to foster a positive and professional learning environment we encourage the following kinds of behaviours in all workshop events and platforms:

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

**0.0.3.0.2 Unacceptable behaviour** Examples of unacceptable behaviour by participants at any workshop event/platform include:

- written or verbal comments which have the effect of excluding people on the - basis of membership of any specific group;
- causing someone to fear for their safety, such as through stalking or intimidation;
- violent threats or language directed against another person;
- the display of sexual or violent images;
- unwelcome sexual attention;
- nonconsensual or unwelcome physical contact;
- insults or put-downs;
- sexist, racist, homophobic, transphobic, ableist, or exclusionary jokes;
- incitement to violence, suicide, or self-harm;
- continuing to initiate interaction (including photography or recording) with - someone after being asked to stop;
- publication of private communication without consent.

## 0.0.4 Contributing



# **(PART\*) Workshop 9: Multivariate Analyses in R**



# Chapter 1

## Preparing for the workshop

To prepare for this workshop, you must do the following steps:

Download the R script and data required for this workshop:

- R Script
- DoubsEnv data
- DoubsSpe data
- Coldiss R function

Make sure to load the following packages (see how in the R script):

- `vegan` (for multivariate analyses)
- `gclus` (for clustering graphics)
- `ape` (for phylogenetics)

```
install.packages("vegan")  
install.packages("gclus")  
install.packages("ape")
```

```
library(vegan)  
library(gclus)  
library(ape)
```

```
source(file.choose()) # use coldiss.R which you have downloaded to your own directory
```



## Chapter 2

# What is ordination?

Ordination is a set of methods for depicting samples in multiple dimensions (Clarke and Warwick 2011) and often feels like a catch-all term in ecological statistics. Ecologists are often told to “run a PCA” in the face of complex and messy multivariate data. R code for ordination techniques is readily available and relatively straight-forward to implement for most data. Interpretation of ordination analyses can be more difficult, especially if you are unsure of the specific questions you wish to explore with a particular ordination method. As such, while ordination methods are very useful for helping simplify and make sense of multivariate data, careful consideration of why the methods are being used and which ones are most appropriate is necessary for strong ecological analyses!

When you use an ordination method, you are taking a set of variables and creating new principal axes along which samples (sites etc.) are scored or ordered (Gotelli and Ellison 2004), in order to reduce or simplify the data, i.e. to create new axes that represent most of the variation in the data. As an example, a dataset with 24 variables may be reduced to five principal components that represent the main patterns of variation amongst samples. Unconstrained ordination methods are generally not frameworks for hypotheses testing, rather they are best suited for exploratory data analysis. The different types of ordination can be useful for many different questions; see (the Ordination Website) for an overview of different types of ordination).



## Chapter 3

# Getting started with data

We will use two main data sets in the first part of this workshop. “DoubsSpe.csv” is a data frame of fish community data where the first column contains site names from 1 to 30 and the remaining columns are fish taxa. The taxa columns are populated by fish abundance data (counts). “DoubsEnv.csv” is a data frame of environmental data for the same sites contained in the fish community data frame. Again, the first column contains site names from 1 to 30. The remaining columns contain measurements for 11 abiotic variables. Note that data used in ordination analyses is generally in wide-format.

```
#Species community data frame (fish abundance): "DoubsSpe.csv"
spe<- read.csv(file.choose(), row.names=1)
spe<- spe[-8,] #Site number 8 contains no species and so row 8 (site 8) is removed. Be careful to
#only run this command line once as you are overwriting "spe" each time.

#Environmental data frame: "DoubsEnv.csv"
env<- read.csv(file.choose(), row.names=1)
env<- env[-8,] #Remove corresponding abiotic data for site 8 (since removed from fish data).
#Again, be careful to only run the last line once.
```

