UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

RESEARCH AND POSTGRADUATE HEAD OFFICE

LEARNING MODULE PROGRAM

Identification Information			
School: Facultad de Ciencias Marinas e Instituto de Investigaciones Oceanológicas			
Program: PhD in Sciences in Coastal Oceanography		Study Program: 2021-1	
Name of Learning Module: Ecological Data in R			
Learning Module Number:		Type of Learning Module: Elective	
Class Hours (HC):	2	Field Practice Hours (HPC):	0
Workshop Hours (HT):	2	Clinical Hours (HCL):	0
Lab Hours (HL):	0	Extracurricular Hours (HE)	0
Credits (CR): 6			
Requirements:			

End of Program Profile

Upon completion of the PhD Program in Coastal Oceanography the student will be trained to develop original and independent research in marine sciences with top-level technical and methodological capabilities. The program will equip the student to push scientific knowledge forward and solve emerging problems related to the marine environment. Upon completion of the PhD Program in Coastal Oceanography the student will be able to:

Evaluate the oceanographic and climatological conditions in a comprehensive manner, through professionally applying the scientific method, including interdisciplinary and multidisciplinary work, as well as critical thinking implementing innovative strategies that resolve emerging regional and global problems to appropriately use and protect the marine environment, with honesty, social responsibility and respect for the environment.

Evaluate the effects of physical and climatological variability on chemical-biological variables that occur in the ocean, through the generation and application of multidisciplinary methodologies and techniques of biogeochemical analyses, for the implementation of innovative and comprehensive mitigation actions fostering the protection and sustainability of marine natural resources, with a proactive and innovative attitude, social responsibility and respect for the environment.

Evaluate the biological components of an ecosystem, their relationship and adaptation to the environmental physicochemical variables and their anthropic variations, through collaborating in interdisciplinary and multidisciplinary groups, as well as generating innovative biotechnological tools, to contribute to the implementation of conservation and management strategies of marine resources based on the environmental service value they provide to the ecosystem, with a proactive and innovative attitude, social responsibility and respect for the environment.

General Definitions of the Learning Module	
General Purpose of the Learning Module:	This learning unit aims to provide current computational tools to program data analysis routines that assess relationships among the biological components and the physicochemical variables of an ecosystem.
Competency of the Learning Module:	To program ecological data analysis routines, using data mining and statistical analyses tools in the R programming environment, to evaluate in a novel way the

	relationships among the biological components and the physicochemical variables of an ecosystem, with a purposeful and innovative attitude and social responsibility.
Learning Evidence (achievement or product to assess) of the Learning Module:	Portfolio of evidence: (1) Solving five practical exercises presented in an RStudio project format that integrates the data analysis routines, and (2) a final project that integrates data description, data exploration and data analyses routines. The data analysis routines refer to a detailed statistical analysis of explored variables.

Content *(add or delete rows as appropriate)*	
I. Name of the Module: Introduction to R and RStudio	**Hours: 2**
Competency of the Module: To apply basic R programming language commands, by using and exploring the RStudio interface, to get acquainted with the programming tools that will allow developing data analysis routines, objectively and with social responsibility.	
Topic and subtopics: **1.1.** Installing R and Rstudio **1.2.** How to communicate and visualize information in R and RStudio **1.3.** Basic command use **1.4.** Data types in R **1.5.** Creating, listing and removing objects from memory **1.6.** Troubleshooting in R and RStudio	
Practice (workshop): 1. Installing R, introduction to basic commands, use of information through RStudio. Creating, listing and removing objects from memory. Troubleshooting in R and RStudio.	**Hours: 2**

II. Name of the Module: Workflow in R	**Hours: 2**
Competency of the Module: To employ efficient workflow strategies in R, by using and applying computational tools such as RStudio projects, Rmarkdown documents, and the GitHub online repository, to develop efficient data analysis routines when using the R programming environment, with social responsibility and a critical attitude.	
Topic and subtopics: **2.1.** R Markdown use **2.2.** Producing scripts in Rmarkdown **2.3.** GitHub use **2.4.** How to share programs and files via GitHub	
Practice (workshop): ***Practical exercise No. 1 includes the following activities in R:*** 1. Use of commands to produce scripts in RMarkdown. Use of routines to produce and share programs and codes through the GitHub platform and version control integration in the R programming environment.	**Hours: 2**

III. Name of the Module: Database manipulation	**Hours: 6**
Competency of the Module: To analyze ecological databases in a reproducible way in the RStudio platform, using available packages in R (e.g. Tidyverse), to transform the array of variables into specific formats needed for analytical routines, with social responsibility and an innovative attitude.	

Topic and subtopics: **3.1.** Database manipulation with basic R programming commands **3.2.** Use of the package Tidyverse (with a focus on dplyr and tidyr)	
Practice (workshop): ***Practical exercise No. 2 includes the following activities in R:*** 1. Database handling as objects in the R programming environment. 2. Use of commands from the Tidyverse package with focus on dplyr and tidyr.	**Hours: 6**

IV. Name of the Module: Data visualization	**Hours: 8**
Competency of the Module: To analyze ecological databases in a reproducible way in the RStudio platform, using available packages in R (i.e. ggplot2), to graphically illustrate relationships among biological and environmental variables of an ecosystem, with professional responsibility and an innovative attitude.	
Topic and subtopics: **4.1.** Basic visualization commands in the R programming environment **4.2.** Advanced visualization packages in the R programming environment (i.e. ggplot2) **4.3.** Producing high quality plots in the R programming environment	
Practice (workshop): ***Practical exercise No. 3 includes the following activities in R:*** 1. Use of basic visualization commands in the R programming environment 2. Use of commands of the ggplot2 package in the R programming environment	**Hours: 8**

V. Name of the Module: Data analysis and linear models	**Hours: 10**
Competency of the Module: To investigate the relationships among biological and physicochemical variables in a reproducible way in the RStudio platform, using available packages in R, to implement simple, multiple and generalized linear regression models, with a purposeful and innovative attitude.	
Topic and subtopics: **5.1.** The 8 steps of the data exploration process **5.2.** Simple linear regression **5.3.** Multiple linear regression **5.4.** Model validation and model selection **5.6.** Generalized linear regression models	
Practice (workshop): ***Practical exercise No. 4 includes the following activities in R:*** 1. The 8 steps of the data exploration process in the R programming environment 2. Linear regression model commands in the R programming environment 3. Model validation and model selection commands in the R programming environment 4. Generalized linear regression model commands in the R programming environment	**Hours: 10**

VI. Name of the Module: Multivariate data analyses	**Hours: 4**
Competency of the Module: To investigate the relationships among biological and physicochemical variables in a reproducible way in the RStudio platform, using available packages in R, to implement ordinate and dissimilarity multivariate data analysis, with a purposeful and innovative attitude.	
Topic and subtopics: **6.1.** Introduction to multivariate data analyses **6.2.** The package vegan in the R programming environment	
Practice (workshop): ***Practical exercise No. 5 includes the following activities in R:*** 1. Ordination multivariate data analyses using the R package vegan 2. Dissimilarity multivariate data analyses using the R package vegan	**Hours: 4**

Learning Strategies used:

During workshops, the student will revise the data analyses routines previously prepared by the instructor. The student will apply theoretical and practical aspects implementing their use in the R programming environment. The student will solve exercises that integrate data analyses routines using the strategies and tools offered by the R programming environment.

Evaluation Criteria:

Portfolio of evidence with 5 exercises: 50%

Final project: 50%

Accreditation Criteria:

- The student must fulfill regulations stated in the current Academic Statute or other applicable regulation.
- The grade will be in a 0 to 100 point scale, with a minimum passing grade of 70.

Bibliography:

Abedin, J. & Mittal, H. V. (2014). *R Graphs Cookbook* (2a.ed.). Birmingham: Packt Publishing eBook Collection (EBSCOhost). [classic]

Abedin, J. & Das, K. K. (2015). *Data Manipulation with* R (2a. ed.). Birmingham: Packt Publishing eBook Collection (EBSCOhost).

Anderson, D., & Burnham, K. (2002). Avoiding Pitfalls When Using Information-Theoretic Methods. *The Journal of Wildlife Management, 66*(3), 912-918. [classic]

Buttigieg, P.L. & Ramette, A. (2014). A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiology Ecology*, 90 (3): 543–550. [classic]

Everitt, B.S. & Hothorn, T. (2010). *A handbook of statistical analysis using R*. Boca Raton: Chapman-Hall. [classic].

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statist. Med*., 27: 2865-2873. [classic]

Harrison, X.A., Donaldson, L., Correa-Cano M.E., Evans, J., Fisher, D.N., Goodwin, C.E.D., Robinson, B.S., Hodgson, D.J. & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:e4794.

Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54(2), 187-211. [classic]

O'Hara, R.B. & Kotze, D.J. (2010). Do not log–transform count data. *Methods in Ecology and Evolution*, 1: 118-122. [classic]

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, H., Szoecs, E. & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology*. 62(2):142–160. [classic]

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1: 103-113. [classic]

Long, J.D. & Teetor, P. (2019). *R Cookbook* (2a. ed.). Sebastopol: O'Reilly Media Inc. https://rc2e.com/

Toomey, D. (2014). *R for data science: learn and explore the fundamentals of data science with R*. Birmingham: Packt Publishing eBook Collection (EBSCOhost). [classic]

Venables, W. N. & Smith, D.M. (2009). *An Introduction to R*. United Kingdom: Network Theory Limited. [classic]

Weissgerber, T.L., Milic, N.M., Winham, S.J. & Garovic, V.D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology* 13(4): e1002128. [classic]

Wickham, H. & Grolemund G. (2017). *R for data science: visualize, model, transform, tidy, and import data*. O'Reilly Media, p. 518. ISBN: 978-1491910399. FREE online at: http://r4ds.had.co.nz/index.html.

Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Zuur, A., Ieno E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NW, USA, Springer-Verlag New York. [classic]

Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3-14. [classic]

Zuur, A.F. & Ieno, E.N. (2016). A protocol for conducting and presenting results of regression–type analyses. *Methods in Ecology and Evolution*, 7: 636-645.

NOTE: The professor must keep up to date the R packages references in order to use the most recent versions.

Date Created / Updated: August, 2020.

Professor Profile: Professor/Researcher with background on Biological Oceanography, Marine Biology or Marine Ecology and experience with data analysis and the use of the R programming environment.

Name(s) and signature(s) of the creator(s) of this Learning Module Program:

Dr. Luis Malpica Cruz
Investigador de Tiempo Completo
IIO, CA de Ecología, Conservación y Manejo de Recursos Marinos

Dr. Rodrigo Beas Luna
Profesor de Tiempo Completo
FCM, CA de Ecología, Conservación y Manejo de Recursos Marinos

Dr. Julio Lorda Solórzano
Profesor de Tiempo Completo
FC, CA de Ecología, Conservación y Manejo de Recursos Marinos

Name and signature of the person who authorizes this Learning Module Program: Dra. Lus Mercedes López Acuña Directora de la Facultad de Ciencias Marinas Profesor de Tiempo Completo FCM, CA de Biotecnología Acuícola Animal Dr. Alejandro Cabello Pasini Director del Instituto de Investigaciones Oceanológicas Investigador de Tiempo Completo IIO, CA de Botánica Marina
Name(s) and signature(s) of the person(s) who peer-reviewed the Learning Module Program: Dr. Rafael Hernández Walls Profesor de Tiempo Completo FCM, CA de Oceanografía Sinóptica Dra. Mónica Torres Beltrán Profesor por Asignatura UABC, Facultad de Ciencias Marinas