TCRENV REPORT 2017

# Report of the Training Course in the R environment

ICES Headquarters, 21-25 August 2017

Einar Hjorleifsson, Bjarki Thor Elvarsson

ICES
International Council for the Exploration of the Sea
CIEM
Conseil International pour l'Exploration de la Mer

**International Council for the Exploration of the Sea**
**Conseil International pour l'Exploration de la Mer**

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 20187. Report of the Training Course in the R environment. TCRENV Report 2017 ICES Headquarters, 21-25 August 2017. Einar Hjorleifsson, Bjarki Thor Elvarsson. 13 pp.

## Contents

# 1 Summary

The training course in the R environment took place at ICES Headquarters, in Copenhagen, Denmark from 21 to 28 August 2017. The number of participants was 17, coming from 13 different institutes.

The objective of the course was to provide participants with a solid foundation in efficient use of the R environment using various typical and familiar fisheries datasets as case examples. Emphasis was put on efficient data munging, data visualization using literate programming starting with "raw" data (individual stations, individual observation), and culminating with deliverance of publishable output produced from a single coded document file.

From the conception of the course through deliverance, all course material were developed and delivered using non-proprietary free software and made available as open source (https://github.com/fishvice/tcrenv2017).

A course webpage (http://www.hafro.is/~einarhj/education/tcrenv2017) based on the source code was generated and used throughout the course.

# 2 Background

## 2.1 Context

The R language is becoming the *Lingua franca* both in data science in general as well as within the ICES community. Recent advancements within R have resulted in that R can no longer be considered as a specific statistical programming language but as a general scientific working environment. This broader environment has resulted in the R has become a natural component of reproducible data analysis and document writing.

Various R packages (e.g. FLR, DATRAS, MSY, SURBAR, VMStools) have often been the backbone of ICES training course and/or workshops. These packages as well as courses are geared towards solving specific pending tasks that tend to come with requirements that the participants are reasonable proficient in basic R and that the input data are correctly formatted and available. Any of these requirements have been seen to pose problems.

The course is aimed at covering the fundamental/generic basis of the grammar of data and graphics as well reproducible document writing where R is used as the sole working medium. Recent developments in the R community that are of interest to fisheries science will also be described.

## 2.2 Objective

The objective of the course is to provide participants with a solid foundation in efficient use of the R environment using various typical and familiar fisheries datasets (landings data, catch data, survey data, and tagging data) as case examples. Emphasis will be put on data munging and literate programming starting with "raw" data (individual stations, individual fish measurements…) and culminating with deliverance of publishable output produced from a single coded document file.

By the end of the course, the participants:

- Will be able to import data from multitude of sources computer (i.e. own text files, excel, access, sql databases) and via the web.
- Will be able to clean, manipulate, explore, summarize, and graph data. This includes being able to:
  - Apply best practices in data preparation
  - Present results graphically, highlighting significant results
  - Merge, slice and dice various datasets
- Will be able to apply the principle of reproducible analysis and report writing from A through Z which are then deliverable through any of the current three common deliverable formats: .html, .pdf and .docx.
- Will be able to produce own functions and understand the principles of creating R packages and social version control coding (through www.github.com).

## 2.3 Level

The course is targeted at fisheries scientist with already have some basic experience in R but are yet not proficient enough to write fluently code for data manipulation, exploration and writing own functions. We believe that some part of the course would also be beneficial to those that are currently productively using R in fisheries science but may along the way have skipped some of the basics and/or are unaware of recent advancements in the R environment when it comes to efficient data handling and processing.

# 3 Course Programme, Product, Deliverance and Instructors

The training course in the R environment took place at ICES Headquarters, in Copenhagen, Denmark from 21 to 28 August 2017. The number of participants was 17, coming from 13 different institutes.

## 3.1 Programme

The schedule was as follows:

**Day 1 - Monday**

- Introduction
- Getting data into R
- Getting started with ggplot

**Day 2 - Tuesday**

- 09:00 - 10:00 - recapitulation of yesterday
- Getting started with dplyr
- Working with dates
- Reproducible research with R

**Day 3 - Wednesday**

- 09:00 - 10:00 - recapitulation of yesterday
- Plotting spatial data
- Relational data
- Exercises using DATRAS data
- Work on preassigned projects or on your own data or
  - Start thinking about creating a presentation of your application to the group. See e.g. this presentation from a "fellow" of last year's course.

**Day 4 - Thursday**

- 09:00 - 10:00 - recapitulation of yesterday
- Statistics in R
- Work on assignments or own data
- Participants presentation of work

**Day 5 - Friday**

- Generating a package
- Introduction to version control (git) and social coding (www.github.com)
- Participant presentation of work
- General discussion on the course

## 3.2 Course products

From the conception of the course through deliverance, all course material were developed and delivered using non-proprietary free software. In that spirit all the material was also made available as open source on a github repository (https://github.com/fishvice/tcrenv2017). The repository contains all source documents (lectures in the form of .Rmd and/or .Rnw). A webpage (http://www.hafro.is/~einarhj/education/tcrenv2017) based on that material was also generated and used in delivery of the course.

The above product allows anybody interest to make a full copy of the course, either as a compiled zip-document (https://github.com/fishvice/tcrenv2017/archive/master.zip) or by simply typing in a terminal window on computers where the git program has been installed:

```
git clone https://github.com/fishvice/tcrenv2017.git
```

The source code for course is shared via the creative commons licence, allowing anybody to reuse any or all of the material produced for the course.

## 3.3 Deliverables

The deliverance of the course deviated somewhat from the planned schedule because mainly because the coverage of each topic, including practical assignments took longer than anticipated. This resulted in that some of the topics (mainly function, packages and social coding) were only covered with brief lectures without participants' hands on training.

Each day started with a short discussion of the topics/assignments covered the previous day. During the deliverance emphasis was put on cooperative work and code sharing (including difficulties/stumbling blocks) among participants. Once the basic elements had been introduced within the first three days, course participants were encouraged to work with their own data and problems, the instructors being at hand to suggest solutions when issues arouse. The participants' data and issues tackled were quite diverse and included the automatic generation of cruise reports, animations of temporal changes in species distribution and interactions to big datasets located in a remote SQL database.

## 3.4 Lecturers

### Bjarki Þór Elvarsson, Marine and Freshwater Research Institute, Iceland

Bjarki is a statistician at the Demersal Division of the Marine and Freshwater Research Institute, Reykjavík Iceland. His main research focuses on statistical methods related to stock assessment models. He has been a member of the Icelandic delegation to the scientific council of International whaling commission since 2010 and participated in various ICES working group meetings since 2013. R has been in his main working environment since 2004.

### Einar Hjörleifsson, Marine and Freshwater Research Institute, Iceland

Einar is a fisheries scientist at the Data Division of the Marine and Freshwater Research Institute, Reykjavík Iceland. He has been involved in various ICES works since 1996 that spans the whole spectrum from ACFM membership up to working group chairing and participation. He has also been involved in the United Nation University Fisheries Training Program teaching stock assessment. R has been his primary working environment since 2009.

# 4 Recommendation

In the waning hour of the course a discussion focusing on the documentation, deliverance and what may have been handled better by the instructors. The main comments on the latter were:

- Provide guidance on naming of objects.
- Allow more time in lecture on the assigned exercises.
- Provide a better deliverance of the solution to the exercises (now stored within the Rmd-source documents).
- The prearranged projects be categorized with respect to the issues they address.
- Set the frame structure in RStudio up such that the source code is on the left and the console is on the right (instead of below).

## Annex 1: List of participants

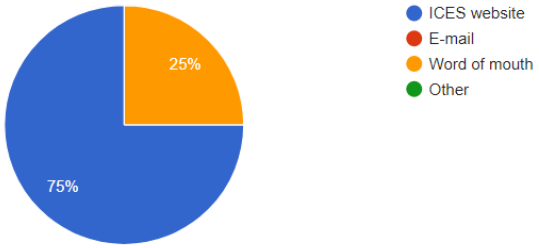
| Name | Institute | Email |
|---|---|---|
| **Bjarki Þór Elvarsson (Instructor)** | Marine Research Institute Iceland | bjarki.elvarsson@hafog-vatn.is |
| **Einar Hjörleifsson (Instructor)** | Marine Research Institute Iceland | haraldur.arnar.einars-son@hafogvatn.is |
| **Edgaras Ivanauskas** | Klaipeda university Lithuania | edgaras.iva@gmail.com |
| **Erik Berg** | Institute of Marine Research Norway | erik.berg@imr.no |
| **Floor Quirijns** | Pelagic Freezer-trawler Association The Netherlands | floorquirijns@gmail.com |
| **Franziska Schade** | Thuenen Institute Germany | fran-ziska.schade@tuenen.de |
| **Henrik Kjems-Niel-sen** | ICES | henrikkn@ices.dk |
| **Iratxe Rubio** | University of the Basque Country Spain | iratxe.rubio@bc3re-search.org |
| **Jette Fredslund** | ICES | jette.fredslund@ices.dk |
| **Julia Hoffmann** | Institute for Economics Kiel Univer-sity Germany | j.hoffmann@economics.uni-kiel.de |
| **Julie Davies** | DTU Aqua Denmark | coad.julie@gmail.com |
| **Karen Bekaert** | ILVO Belgium | karen.bekaert@ilvo.vlaan-deren.be |
| **Katie Thomas** | Marine Institute Ireland | katie.thomas@marine.ie |
| **Maciej Adamowicz** | National Marine Fisheries Research Institute Poland | madamowicz@mir.gdy-nia.pl |
| **Marijus Spegys** | Fisheries Service under the Ministry of Agriculture of the Republic of Lithuania | marijus.spegys@zuv.lt |
| **Marta Suska** | National Marine Fisheries Research Institute Poland | msuska@mir.gdynia.pl |
| **Ole Henriksen** | DTU Aqua Denmark | ohen@aqua.dtu.dk |
| **Ole Thomas Albert** | Institute of Marine Research Norway | oleta@imr.no |
| **Periklis Panagiotidis** | ICES | periklis@ices.dk |
| **Perttu Rantanen** | Natural Resources Institute Finland | perttu.rantanen@luke.fi |
| **Signe Bagger** | ICES | signe.bagger@ices.dk |

| Name | Institute | Email |
|---|---|---|
| **Yury Zablotski** | Thuenen Institute for Baltic See Fisheries Germany | yury.zablotski@thuenen.de |

## Annex 2: Results of course evaluation questionnaire

How did you hear about this course?

12 réponses




Comments

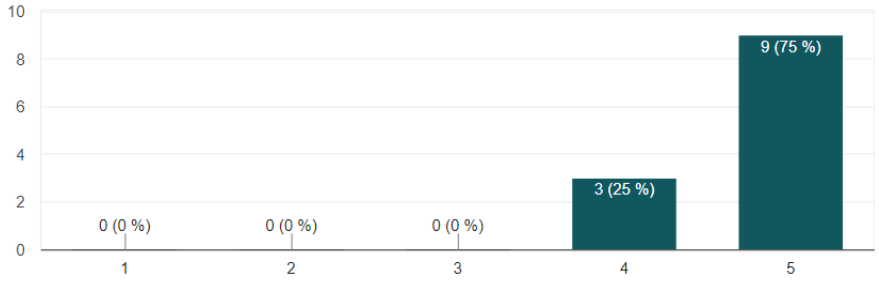
Une réponse

It was recommended by a participant from 2016

### Course content

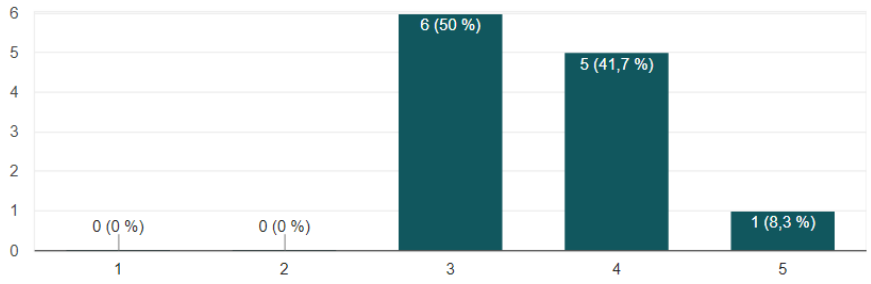
Did the Training course meet your expectations?

12 réponses



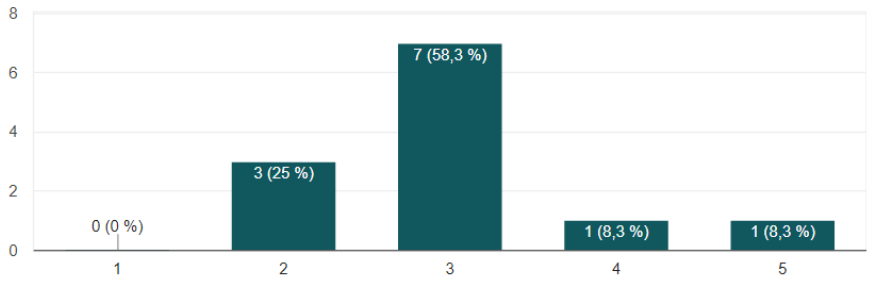
## Was the level of instruction appropriate?

12 réponses




## Was the length of the training course appropriate?

12 réponses




## Comments

3 réponses

It could have been reveiling to go through some case examples more detailled.. on how some perticular research issue(s) has/have been solved using some methods learned during the week.

The density of information was high, which was sometimes brain frying. I was glad with the amount of info that was shared. Perhaps it could be spread out a bit more over the days. At the end of the week there was a lot of time for working on our own data. Maybe that could be spread out a bit as well. So everyday new stuff, and the possibility to apply it to your own data (Or to simply process all new stuff in your head)

The instructions were to fast, but I think, otherwise, we wouldn't be able to cover everything. I'd rather had more time for exercising and using the knowledge, instead of just listening and not getting much done, because we moved forward fast. Otherwise, the course was amazing!

### What did you like best or find more useful about the training course?

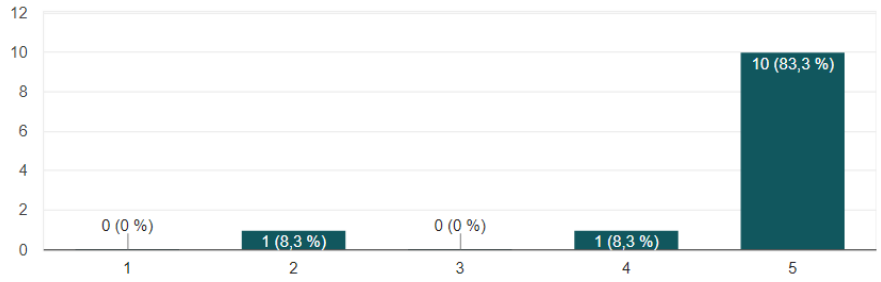
10 réponses

| |
|---|
| Applying directly to our own data what we learnd |
| All practical exercises were very useful and topics such as package writing and git were something new and very interesting. I highly appreciate as well all the books recommended by instructors. |
| Opportunity to use, apply my own data from the training course exercises and examples. |
| well structured and materials |
| The range of topics |
| Good approach that we had time to try the procedures ourselves. |
| I like that it is so applied. And I like the possibility to ask questions/find solutions for problems with my own data |
| The big picture of R and its most useful features for Data Scientists! It allowed me to connect some dots. |
| Many practical examples. Instructors available to help solving problems straight away. |
| to have time to apply newly learned stuff to own research projects under supervision |

## Course Organization

### Inscription to the training course and communication with organizers were efficient.

12 réponses




### Comments

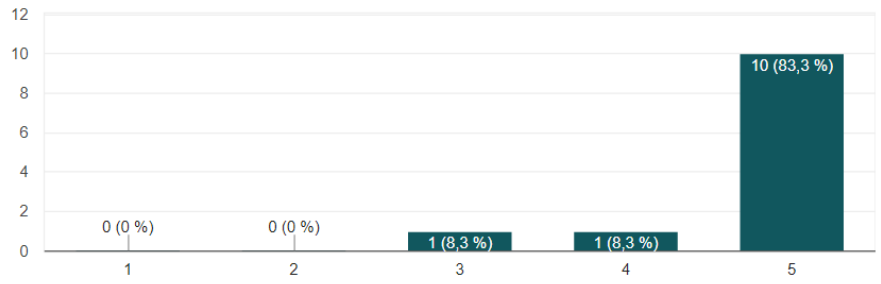
3 réponses

| |
|---|
| (What do you mean by inscription?). Communication about starting time/daily hours before and during the course could have been more specific (not that I asked for more information, I just could have). Also, I could have been better prepared for all the time given to work on our own data. Not sure if I overlooked communication about bringing own problems or if it wasn't communicated so clearly. |
| It was fine. I only would have liked to know earlier that the course would take place, instead of during summer. Prices of accommodations and tickets are high in summer and even higher if booked only a few weeks before the course |
| Almost 100% of the code worked, which is really not easy task to do in the class of people with different levels. |

## Teaching and Learning Support

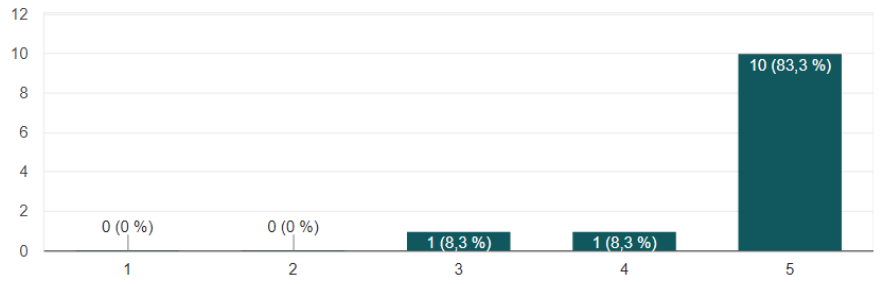
The instructors were helpful, informative, and approachable.

12 réponses




The working documents were presented in a way that facilitated learning.

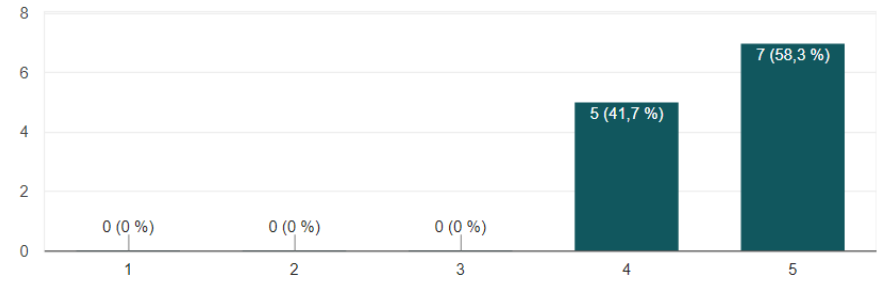
12 réponses




## Overall Evaluation

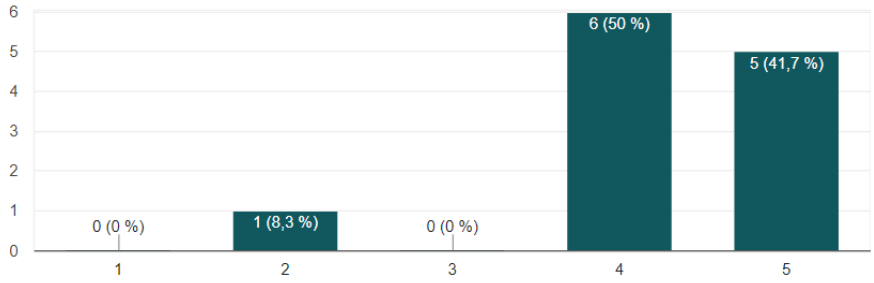
Overall, how would you rate this training course?

12 réponses



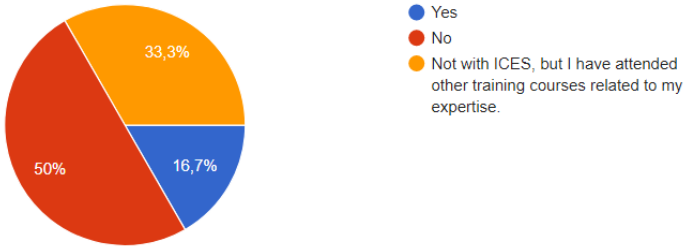
## Overall, how would you rate the quality of the teaching?

12 réponses



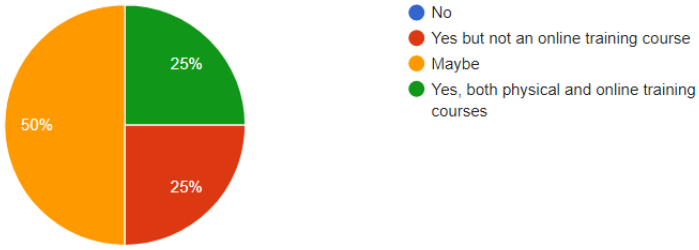

## Have you taken any other ICES training courses?

12 réponses




## Would you be interested in another training course within ICES?

12 réponses




## If yes, which topic would you be interested in?

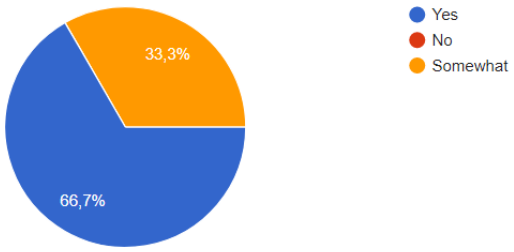
6 réponses

| |
|---|
| R, more advanced R, Statistics |
| fisheries management, assessment, common fisheries policy |
| Let's see what becomes relevant... |
| More R, but not yet sure what kind of subject would be useful |
| Statistics in R for non-mathematicians, SQL, SQL - R interaction, machine learning in R, scientific writing |
| Statistics - probability, estimations. (Elementary level) |

## Social Event

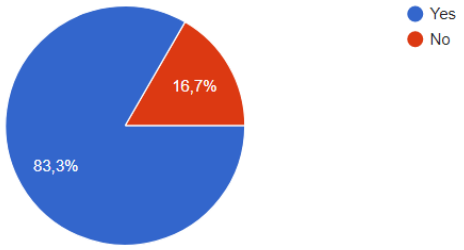
Do you feel that you have benefited from networking opportunities on the course?

12 réponses




Did you participated in the course dinner on Thursday evening ?

12 réponses




## General comments on the Training Course

7 réponses

| |
|---|
| Suggestions were already made during the evaluation the last day of the course. Einar will certainly transmit them to you. |
| Default RStudio setting should be used in the future also. I would guess most of the users use default settings. |
| To improve the quality of teaching I recommend to give more time to work on the exercises by yourself, instead of giving the answer 1 min later. I liked the range of topics a lot, but the instructors were rushing through the topics and scripts. |
| Very well prepared electronic material for the course. Easy to follow. Well though out data sets and example code. |
| Thanks for the organisation! It was a nice course, and I was happy to be able to follow it. |
| Thanks a lot for your amazing work! Always enjoy being at ICES :) |
| Thank you for organizing the training course. |