Report of the Training Course in the R environment

ICES Headquarters, 29 February – 4 March 2016

Einar Hjorleifsson, Bjarki Thor Elvarsson

ICES
CIEM

International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

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

Recommended format for purposes of citation:

ICES. 2016. Report of the Training Course in the R environment, ICES Headquarters, 29 February - 4 March 2016, Einar Hjorleifsson, Bjarki Thor Elvarsson. 14 pp.

Contents

1 Summary

The training course in the R environment took place at ICES Headquarters, in Copenhagen, Denmark from 28 February to 4. March 2016. The number of participants was 24, coming from 8 countries.

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 (landings data, catch data, survey data, and tagging data) as case examples. Emphasis was 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.

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. To that end, a GitHub repository was generated (https://github.com/fishvice/tcrenv2016) that contains all source documents (in the form of .Rmd and/or .Rnw) and data, with associated output documents (.html and/or .pdf format). A webpage (http://www.hafro.is/~einarhj/education/tcrenv2016/) based on that material was also 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, 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 have been included inside ready-made function, 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.

This course is aimed more 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 as well participate in social 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 the basics or are unaware of recent advancements in the R environment.

3 Course Programme, Product, Deliverance and Instructors

The course took place in Copenhagen, Denmark from 28 February to 4 March 2016. The number of participants was 25 coming from 8 countries (Annex 1).

3.1 Programme

The schedule as set up prior to the course was as follows:

3.1.1 Day 1 – Monday

- Introduction:
  - Rstudio and R projects
  - knitr, markdown, document writing (html, pdf, docx)
  - reproducible analysis using R
- Getting data into R and out of R:
  - from ones own computer (text files, excel, …)
  - from the web
  - from API's (ICES Webservices)
  - from databases (Ram's online Postgres database)
- The grammar of data and graphics:
  - Introduction to ggplot2
  - Introduction to dplyr

3.1.2 Day 2 – Tuesday

- The grammar of data and graphics - continued:
    - Exploratory data analysis and visualization
- Working with characters and dates
- The base R equivalence

3.1.3 Day 3 – Wednesday

- GIS in R
    - using ggplot2 and ggmap
    - using leaflet
- Applied project(s) - From "messy" data to a final report using reproducible approach based on case examples
    - candidate: abundance and biomass indices starting with ICES DATRAS data

3.1.4 Day 4 – Thursday

- Applied project - continued

3.1.5 Day 5 – Friday

- The fundamentals of functions and package writing
    - Fundamental of functions and documentation
    - Directory structure and a minimal example
    - Version control (git) and social coding (www.github.com)

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. To that end a GitHub repository was generated (https://github.com/fishvice/tcrenv2016) that contains all source documents (in the form of .Rmd and/or .Rnw) and data, with associated output documents (.html and/or .pdf format). A webpage (http://www.hafro.is/~einarhj/education/tcrenv2016/) based on that material was also generated and used throughout 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/tcrenv2016/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/tcrenv2016.git
```

During the development of the course, the ICES secretariat asked the instructors to consider the usage of the BlueBridge Virtual Research Environment (BVRE) as a platform for the course. BVRE is all encompassing environment for collaborative work such as projects in a university course. Projects can import powerful modules or tools that allow a particular analysis. Similarly, data can be imported through a set of fixed data import routines and made available to the projects participants. The BVRE is accessed through a browser and all the analysis is made on centralised servers. This approach therefore makes very little demands on the computer platform of the end user or available software, all that is required is a modern browser and a stable internet

connection. Very superficial exploration was made with respect to the usefulness of this platform and it was concluded that it most likely would not be any better than the approach as described above. Although BVRE has an impressive feature, it requires an additional investment and would add layer of confusion for the course participants. Although R is an available from within BVRE, its use is somewhat restricted. Notably it is assumed that most of the data handling is done through BVRE. In contrast, the course covered various ways of working with data with R from multiple sources. Additionally the benefit for the user is not obvious as the course participants are not likely to have access to the BRVE at their home institutions.

3.3   Deliverables

The major change from the planned schedule described above and the actual course deliverance was that the coverage of each topic, including practical assignments took longer than anticipated. This resulted in that the topics scheduled for Friday were more or less put on the back burner and were only covered with brief lectures without participant's hands on training.

Each day was split up into group discussion of the topics/assignments covered the previous day, introduction lectures of the day's topics followed by practical assignments. Emphasis was put on cooperative work and code sharing (including difficulties/stumbling blocks) among participants. The practical assignments were fairly diverse, ranging from simple summary statistics based on DATRAS data, to maps of fishing intensity and collating landing statistics. In addition, course participants were encouraged to work with their own data and problems using R. These included time-series analysis of fish price, automatic cruise reports and biomass estimation of zooplankton.

3.4   Lecturers

- Bjarki Þór Elvarsson, Marine Research Institute, Iceland

Bjarki is a statistician in the Fisheries Advisory Section of the Marine Research Institute, Reykjavík Iceland. He recently finished his PhD in 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 Research Institute, Iceland

Einar is a fisheries scientist in Fisheries Advisory Section of the Marine Research Institute, Reykjavík Iceland. He has been involved in various ICES works since 1996 that spans the whole spectrum from ACFM/ACOM membership up to working group participation. In the early 2000s, he taught stock assessment at ICES with Dankert Skagen for three consecutive years. 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   Recommendations

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

- Document organization

- Documents and data were distributed by the instructors in any of three different spaces (web-access, SharePoint, GitHub). It was recommended that if the course is taught again all material distributed via only one space. It should also be considered if course content could not be organized by the day covered.
- Order and scope of the course material:
    - Data importing, checking, tidying and "mining" should be put upfront in the course.
    - Question was raised if the usage of R-package development and version control (git) was outside the main scope of the course. If the course were to be trimmed the R-package development was considered to be the most obvious candidate to put on the chopping block. If version control was to be retained as a part of the course content, it was suggested that it is thought as a part of the assignments.
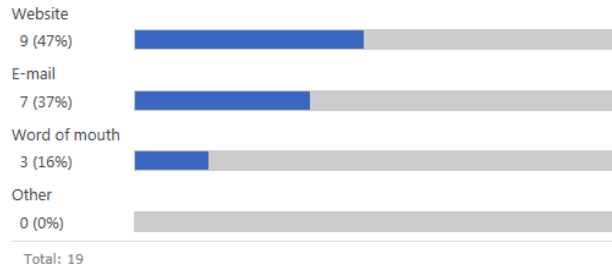
Annex 1: List of participants

Name	Country	E-mail address
Bjarki Thor Elvarsson	Iceland	bthe@hafro.is
Einar Hjorleifsson	Iceland	einarhj@hafro.is
Tone Vollen	Norway	tone.vollen@imr.no
Keno Ferter	Norway	Keno@imr.no
Anne Sell	Germany	anne.sell@ti.bund.de
Toyonobu Fujii	UK	t.fujii@abdn.ac.uk
Lies Vansteenbrugge	Belgium	lies.vansteenbrugge@ilvo.vlaanderen.be
Loes Vandecasteele	Belgium	loes.vandecasteele@ilvo.vlaanderen.be
Debbi Pedreschi	Ireland	debbi.pedreschi@marine.ie
Elvar H. Hallfredsson	Norway	elvarh@imr.no
James Murphy	Ireland	james.murphy@nuigalway.ie
Valentina Melli	Denmark	vmel@aqua.dtu.dk
Espen Strand	Norway	espen.strand@imr.no
Matthias Schaber	Germany	matthias.schaber@ti.bund.de
Martin Pastoors	-	mpastoors@pelagicfish.eu
Vaishav Soni	Denmark	vaishav.soni@ices.dk
Anna Osypchuk	Denmark	anna.osypchuk@ices.dk
Hans Mose Jensen	Denmark	hans.jensen@ices.dk
Sven Stötera	Germany	sven.stoetera@ti.bund.de
Aage Fotland	Norway	aage.fotland@imr.no
Ricardo Leite	-	ricardo.leite@efca.europa.eu
Didzis Ustups	Latvia	didzis.ustups@bior.lv
Arved Staby	Norway	arved@imr.no
Maksims Kovsars	Latvia	Maksims.Kovsars@bior.lv
Ruben Verkempynck	Netherlands	ruben.verkempynck@wur.nl
Maciej T. Tomczak	Sweden	maciej.tomczak@su.se

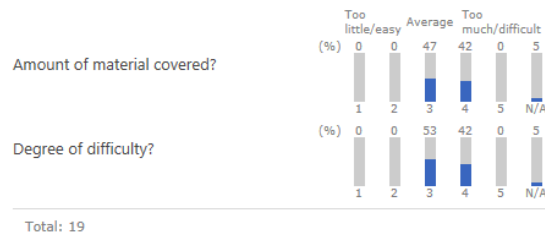## Annex 2: Results of course evaluation questionnaire
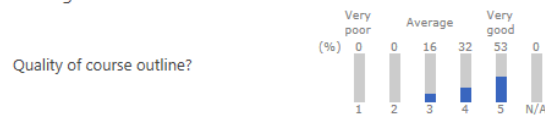
1. _____
How did you hear about this course?

Website	9 (47%)
E-mail	7 (37%)
Word of mouth	3 (16%)
Other	0 (0%)

Total: 19

2. _____
Course content

	Too little/easy		Average	Too much/difficult		
Amount of material covered? (%)	0	0	47	42	0	5
	1	2	3	4	5	N/A
Degree of difficulty? (%)	0	0	53	42	0	5
	1	2	3	4	5	N/A

Total: 19

3. _____
Course Organization

	Very poor		Average	Very good		
Quality of course outline? (%)	0	0	16	32	53	0
	1	2	3	4	5	N/A

Total: 19

4. _____
Teaching and Learning Support

	Low		Average	High		
Helpfulness of teaching staff? (%)	0	0	0	16	84	0
	1	2	3	4	5	N/A
Usefulness of course materials? (%)	0	0	0	32	68	0
	1	2	3	4	5	N/A
Clarity of presentation? (%)	0	0	16	37	47	0
	1	2	3	4	5	N/A
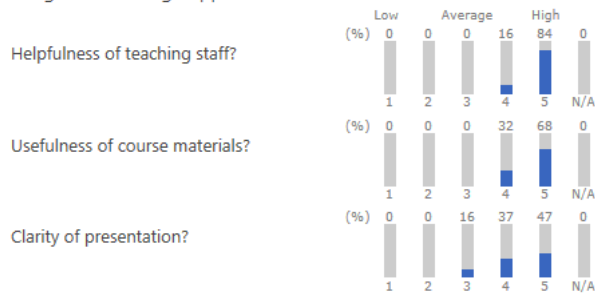
Total: 19

5. _____
Overall Evaluation

	Very poor		Average	Very good		
Overall, how would you rate the course content? (%)	0	0	5	37	58	0
	1	2	3	4	5	N/A
Overall, how would you rate the organisation of the course? (%)	0	0	21	37	37	0
	1	2	3	4	5	N/A
Overall, how would you rate the quality of the teaching? (%)	0	0	5	32	63	0
	1	2	3	4	5	N/A
Overall, how would you rate this course? (%)	0	0	5	37	58	0
	1	2	3	4	5	N/A
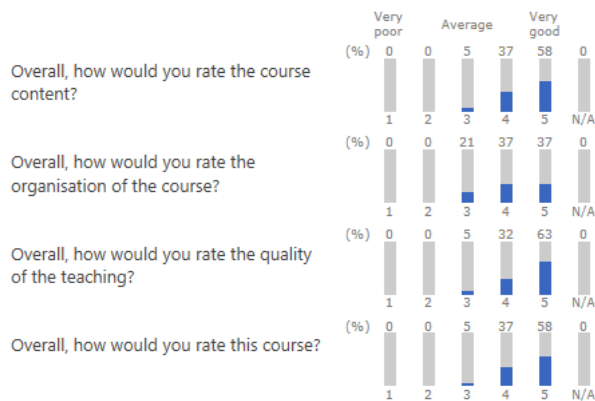
Total: 19

6. _____
Good features of this course/suggestions for improvement:

I did not answer the first question, because it did not cover my intended response: Yes, the amount of material covered in the course was extraordinarily high, but this was excellent - not too much: Because even for a beginner in R, there was always the opportunity to start with easy tasks and ask for help once stuck...

1 (11%)

Good features: practical skills to use R efficiently with large data sets skills to produce high quality graphics Suggestions: sometimes lecture went a bit too fast...

1 (11%)

Organizing the teaching material on one single platform would be handy,

1 (11%)

It was very useful to have time to practice what we learned. Very interesting, intense but fun course.

1 (11%)

Extensive help, support and encouragement. Also, good humour made it easier to tackle a challenging subject!

1 (11%)

Very interesting course! Variaty of experience levels within the group was dealt with really well. Everyone got to learn something applicable to their own field. Interesting tips and tricks for mapping and producing automatic reports out of R. This will be very useful. Lots of information, examples, manuals etc. were made available, although it was not always clear were to get everything. So my only remark is to make this a bit more clear and structured in the future. For the rest, very interesting course. I was happy to attend!

1 (11%)

I think the course had a good approach in explaining a complex environment like R but I would have preferred a bit more linear structure in the sequence of topics: the feeling was that we jump from one topic to the other and for that reason it was not always easy to follow the lesson. With some minor improvements the course may be a great tool for fisheries scientists!

1 (11%)

Possibly include the use of GitHub as a optional exercise for the ones interested

1 (11%)

I think the course could have benefited from a clearer structure, and availability of course material beforehand. The content covered was very applicable and relevant to the work I am doing, and both teachers were helpful in assisting.

1 (11%)

7. _____
Have you taken any other ICES training courses?

No, never
10 (53%)

Yes, I have participated in ICES training courses previously
9 (47%)

Total: 19

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

Yes
19 (100%)

No
0 (0%)

Total: 19

9. Comments

Yes, the icebreaker and joint dinner definitely served the networking well!

1 (25%)

I feel that many participants shared similar aims/purposes to attend this course and the atmosphere was therefore great in terms of networking.

1 (25%)

Ideal way to combine training and networking, especially since I am new in the world of ICES. Through this course, I got to make some interesting contacts
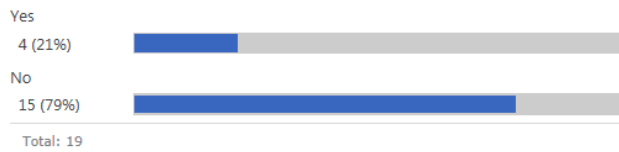
1 (25%)

I think it was a very good composition of participants. Because I had sick family members at home I could unfortunately not take full advantage of the afternoon networking chances.
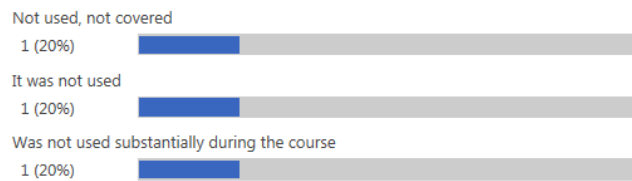
1 (25%)

Total: 4

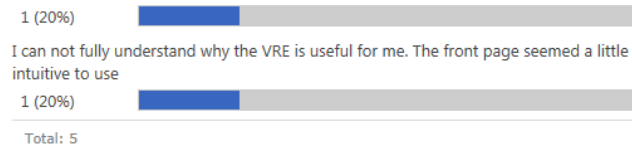10. Did you make use of the BlueBridge Virtual Research Environment (VRE)?

Yes
4 (21%)

No
15 (79%)

Total: 19

11. If so, did you find the VRE useful? Please share any comments or feedback below.

Yes, it was useful
2 (20%)

No, it was not useful
8 (80%)

Total: 10

12. Feedback or comments to the VRE

Not used, not covered
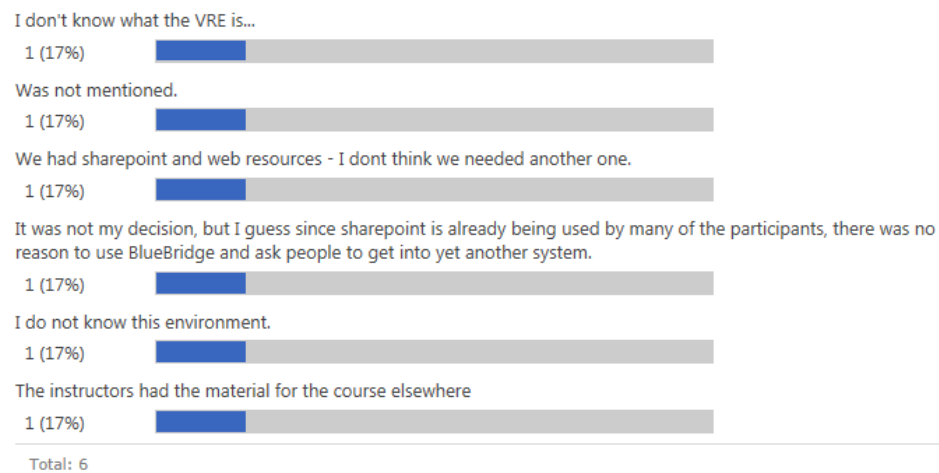1 (20%)

It was not used
1 (20%)

Was not used substantially during the course
1 (20%)

I am sure it can be useful. Documents for this course were shared on sharepoint GitHub and additionally on another website. I had a short look at the BlueBridge portal and I am sure it could do the job (sharing links and documents). I can see that it includes data management functionality, species information, spatial data and analytics. Perhaps one comment could be that all this functionality is going a bit too far. In my daily work I use a whole range of data management applications, GIS tools etc. It would not be possible to replace these tools inside BlueBridge as a webapp, so therefore perhaps it is not worth the effort even trying to include these things. Perhaps some users will appreciate it though... Same goes for the messaging functionalities. I do not see why I would want to use this, since I have my everyday email system at hand already.
1 (20%)

I can not fully understand why the VRE is useful for me. The front page seemed a little confusing, and was not at all intuitive to use
1 (20%)

Total: 5

13. If you did not make use of the VRE, why not?

I don't know what the VRE is...
1 (17%)

Was not mentioned.
1 (17%)

We had sharepoint and web resources - I dont think we needed another one.
1 (17%)

It was not my decision, but I guess since sharepoint is already being used by many of the participants, there was no reason to use BlueBridge and ask people to get into yet another system.
1 (17%)

I do not know this environment.
1 (17%)

The instructors had the material for the course elsewhere
1 (17%)

Total: 6

14. Did you participate in the course dinner on Thursday evening?

Yes
17 (89%)

No
2 (11%)

Total: 19

15. Would you prefer the course dinner to take place in the city centre, or at the course location (at own expense)?

At the course venue

4 (22%)

In the city centre

14 (78%)

Total: 18

16. More comments

The location (Rub & Stub) was great ! Would choose it again

1 (17%)

The dinner was excellent - thanks!

1 (17%)

I really enjoyed both the venue, the food and the ethos

1 (17%)

Thank you!!!!

1 (17%)

Great course that is worth to repeat if someone did not make it this time, or if new people want to get into R :)

1 (17%)

Unfortunately I got sick, so I could not participate in the course dinner.

1 (17%)

Total: 6