



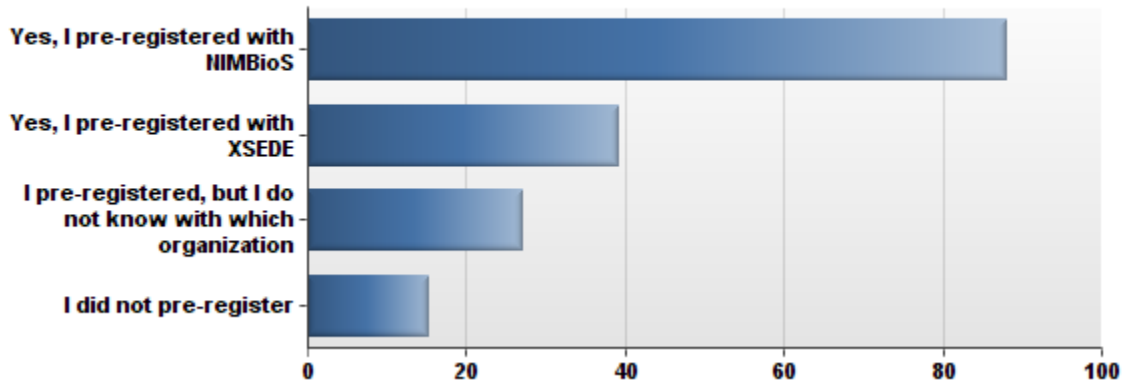
Evaluation Summary Report

Tutorial: *Using R for HPC*

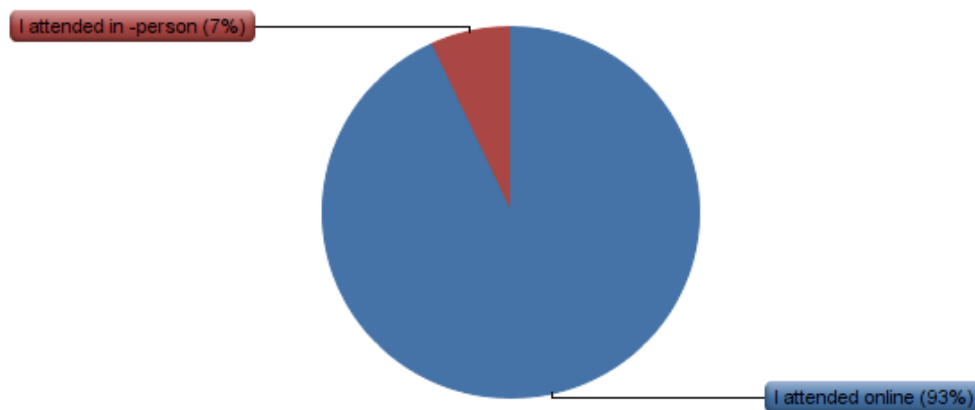
February 27, 2015

Pamela Bishop
Program Evaluation Manager
National Institute for Mathematical and Biological Synthesis

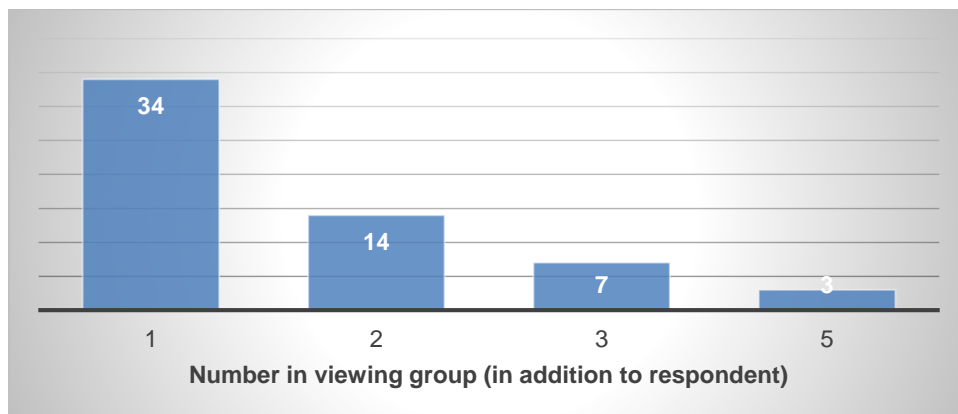
Did you pre-register online for the event?



Did you attend the tutorial online or in-person?



If you were watching the tutorial with a group, approximately how many individuals were watching with you?



I had no technical problems. I see the importance of including each angle—presenter (lip movement, gestures), room (presenter pointing to slides), and slides (since they aren't clear through the previous angle); but the inclusion of all three didn't seem to be much more effective than just using the first and the third of these, as many other online presentations do.

I liked being able to see the presenter and the slides

I was experiencing a lot of lag and chop on the video stream. Occasionally the stream would freeze and I would need to restart.

It stopped streaming after module one, but this may have been a connectivity issue on my end.

It was a little hard to see the video of the presentation slides, but since we were given the presentation, I just followed along from there. But all in all, it was great to be able to 'sit in' on a workshop like this remotely

It would be nice to have a stream that can be played with a standalone player like VLC, rather than a web-based player. This would allow partial recording/time-delayed streaming which is useful to enable remote attendance despite a busy schedule, and/or from other time zones. (This is more of a general comment, as I have attended previous remote workshops at NIMBioS that were not recorded and released on YouTube afterwards)

Jwplayer is blocked at our university. I would advertise in advance that you use it for streaming purposes

Most of the time the streaming worked well. It occurred several times that the video was frozen and I need to restart the webinar to get the video back to work.

No connection on any machines in my lab

Nothing that can be blame on NIMBioS. I had a slow connection on my end.

Other than a few blips and freezes, I had no issues

Some lag between voice and images on the screen and the video of the presenter - sometimes not in sync

Sound not great, but it may have been my local machine.

Stream did not work in Firefox

Technology was easy to use and the quality (sound, video) was good for me.

Technology was great.

Text on many slides was difficult to read (too small).

The codes on the screen was too small to see. This computer screen need to be bigger, and the other two screens can be smaller.

The link doesn't work for me I can't view it live

The Only problem was I was not able to be there and being online I had to go and take care of other jobs at work during the class

The presentation did not work on one of my computers. It did work on another one I had at hand. It would be nice if we could test that before the actual presentation.

The right technology for remote users. However the general view of the speaker in front of the screen wasn't too useful for the remote viewer. What the speaker was pointing on the screen could not be seen.

The stream kept pausing. It did help to load it directly in VLC.

The tech worked very well for the most part. One small glitch in the video feed. One item to consider addressing was for the view of the presenter in front of the screen and improving view of the screen, as it was nearly impossible to read the screen when the presenter was referencing/pointing to items on it.

The twitter feed was fun

The use of twitter

The video capturing the white board was difficult to read due to reflecting light.

The video froze about once every hour and I had to reset my browser

The video of the presenter with the screen wasn't helpful because it wasn't possible to read the screen. Given this, it seemed like a waste of viewing space to present that view. In our case, it would have been more helpful to have a larger view of the slide portion since we were also projecting the webinar on a screen.

There should have been a way to ask questions or make comments other than Twitter! Something within the web conferencing interface would have been a lot more inclusive of all remote participants.

This worked fine...and I could see the tweets as they came in on the side. I was able to get all of the example .pdf files and other materials just fine.

Video player did not work in my Linux workstation. Fortunately had a Windows box available

VLC did a great job for streaming. We lost the connection few times before being able to stream the rtsp with VLC

Was well done, esp. Since there was the pdf of the slides.

Watching through VLC was much better than watching in a browser

We got in troubles with one laptop because we lack the right plugins and/or codecs. Starting 10-15 minutes earlier didn't provide us with time enough to setup the computer in the room and had to change it for one of our personal laptop which worked fine. I would recommend to provide a list of software requirements in advance to be able to connect with the webinar system.

Wonderful!

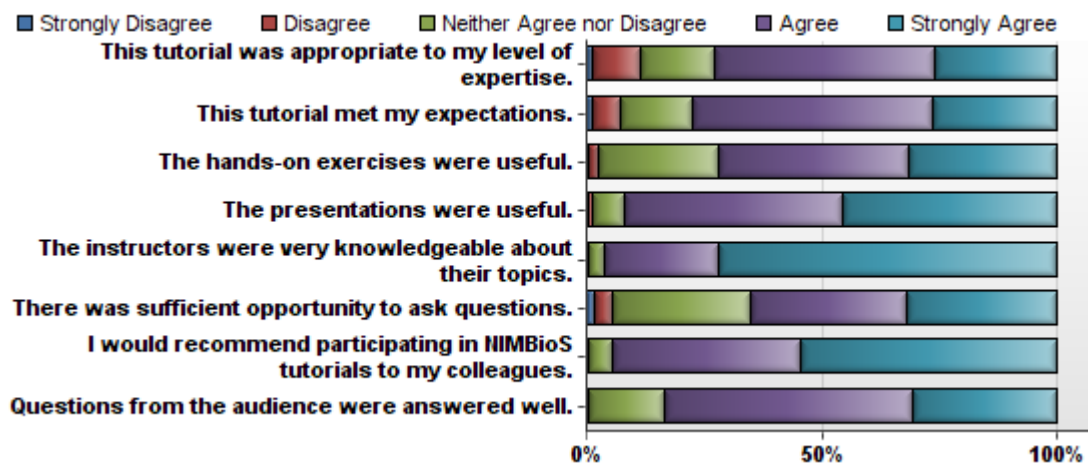
Worked fine

Worked fine for me

Worked smoothly as far as I could tell

Worked well - some of the details on the PowerPoint was a little hard to see because it only took up half my screen

Please indicate your level of agreement with the following statements about this tutorial:



All the topics are great, but I cannot digest, please spend more time, more hands-on next time....

Any important differences between writing R scripts and R packages; how to incorporate C++ code in packages

Basics of parallel programming and some hands-on experience- it was too rushed

Bioinformatics data analysis

Cloud computing

Debug

Deeper coverage of same topics. Some parts were a bit beyond my capabilities, especially toward the end. The first part about debugging and optimization could have been its own tutorial. But in general I feel like I was given a lot of good leads to explore these topics now on my own.

Exercises a bit more involved.

Foreign language interface, parallel programming

Go deeper into how to choose/combine the different methods of optimization (parallel, C++,)

Go over an example of linking a FORTRAN or C to an R code - in particular the compilation. And repeat this for the case of openMP in the Fortran/C code. And see how this compares with using the parallel part in R and not inside the Fortran/C

Going over the exercise problems that were given. Since I was participating online it was kind of a problem.

GPU computing

Handling big data. In- versus out-of-core processing.

Handling of larger-than-memory objects in R

Hands on examples in parallel computing using pbdR

How to write scripts to efficiently distribute jobs among processors in a remote computing cluster

I am just beginning to use R, so this question is not applicable to me.

I could not attend the entire tutorial, due to other responsibilities. I could use info about offline viewing afterwards.

I particularly like the discussion on debugging and optimization.

I think it was fine

I thought it was a great overview, not sure what I would add.

I thought it was going to be a bit more advanced given that the name was R for high performance computing.

I use R for a variety of modeling and statistical applications, which often involve the use of spatial data layers. Large rasters or other spatial data present special problems for efficient programming in R to obtain sufficient processing speeds. I am just learning some of these, but would really like to attend a workshop that addresses this aspect of R, which has an increasingly large user community aimed at spatial data analysis given the modeling capabilities.

I was disappointed that no information was given on how to run multiple R jobs on XSEDE. That is the key piece of information that I wish to learn!

I was expecting it to be basically 100% parallel discussion, getting down to talking about the guts of PbdR, but only the last 10 min were really about that.

I was hoping to hear more details about the R parallelization especially as compared to common approaches such as openMP (via e.g. For each construct) or MPI (via rMPI). As a new R user and experienced MPI and openMP programmer, I find the R parallelization options somewhat convoluted and hard to use perhaps that could be a topic for a future focused workshop.

I was unable to see any of the tutorial and could not participate in any way

I would have liked more focus and specifics on speeding things up in R.

I would have spent more time on parallel computing

I would like to have seen more in-depth presentation of profiling and optimization. I felt like the tutorial basically just covered what those things were, without actually showing HOW to do much.

I would like to have seen the topics in some more detail. Anyway, becoming aware of some tools, I can now look deeper into those I found most useful.

I would like to know how R can be connect to other database, such as MySQL, etc... It could be tricky based on my experience.

I'm actually an R newbie, so I don't even know what topics I would want or need to have covered, honestly (well, besides maybe an intro to using R for analytics ha-ha). I attended to expand my understanding of what R can do and what I should pay attention to when coding in R.

In depth tutorial for profiling

Interfacing with compiled code

Linking parallel R to a resource scheduler such as SLURM, PBS, or other

More about C++ for R, and more about parallel computing. This time it was more of an overview but not hand-on instruction.

More background and basics not about R but about defining the goals of each section and why you would use the new tool...the people in my group are comfortable using R but are not programmers and we were not able to keep up the all of the new acronyms and references to other platforms Also, more examples and detail on debugging would be great!

More depth into using source code where possible. Use of flat files.

More detail or examples of specific types of situations that can or cannot be easily parallelized

More details and examples on parallel programming. Should also discuss ff and big memory.

More discussion of the actual ways to use different approaches, rather than simply just showing that the approach could be beneficial

More discussion on how to optimize codes would be great.

More examples and practice refactoring "bad" R code into better, faster, cleaner R code.

More examples. Alternative R implementations (pqR, etc.)

More in parallel stuff, and also use of externalities such as admb, tmb,INLA AND parallel. Also, advertise that you are using jwplayer to stream. My university blocks it, so I had to scramble and run/drive/run home to follow you guys.

More on basic optimization and debugging technique. I have not really had to expand to the parallel methods, but then again I am really more of a tinkerer.

More on parallel programming and limitations in terms of observations, data size, model fitting etc. A lot of it was kind of silly stuff so it would have been nice to do some real model fitting.

More on parallelization. More on how to handle unavoidable cache misses by R.

More on profiling and performance coding

More on Rcpp - maybe a little bit on what R functionality can be used with Rcpp and what can't. For example, randomly sampling from non-uniform distributions, a comment about that.

More parallel and distributed details Scalability of solutions using R. Different ways to bring R solutions to production environment.

More parallelism in R.

More profiling examples.

More real use examples with genetics.

More time on the basic stuff. Maybe graphing

More time should be spent on the final topic, R for HPC. It felt rushed and the information a bit sparse, even though this was advertised as the theme of the workshop.

No additional topics, maybe just more depth if there was more time.

No new, but more in depth

Not more topics but more details, especially regarding Rcpp and parallelism

Optimization

Optimization, foreign language APIs

Parallelization

Parallel Computations/Computing More time for hands-on work on the topic dealt with

Parallel computing

Parallel computing

Parallel programming

Perhaps, a bit more on using compiled C++.

Profiling and parallel R

Profiling and performance analysis Optimization

Rcpp attributes

Rstudio

Spark/Hadoop

The amount of material covered is right. If there is more time, the pace should be a bit slower.

The basic method apply to genetic analysis.

This tutorial was a bit over my head given my R expertise (or lack of), but I expected that going in. So I was wishing for some more general/basic R information. But the speaker provided some great resources which I have begun going through.

Use of shared memory. Detailed examples of profiling in terms of memory management, Disk I/O

Using R with gpus more on optimizing serial R code (simple tricks and things to avoid) demonstration of creating and distributing your own R package more on R fundamentals- i.e. Data types, object types, etc.

Walking through a couple of extended examples and stopping to explain what to look for, how to interpret data, and how to make decisions based on the information (and balancing developer time vs running time) would be a helpful exercise. Maybe that's something worth doing separately.

Explanation and comparisons of R parallel libraries. Learning about C++ weaving was also very useful and something new to me.

Explanations for why certain methods increase performance

Foreign API

Foreign language APIs

Getting over the initial intimidation/unfamiliarity of Rcpp

Give me a general introduction for some advanced topics: debug, profiling, and parallel processing.

Going over general concepts. Some things I had not heard of.

Going through R code which was a pain because it had to be copied and pasted from the pdf! Upload some R code gosh darn it!

Good overview

Great overall introduction to R optimization.

Have a better understanding about the HPC tools available to use with R and how to use them appropriately.

Having the recorded session is great, and having a follow-up session is very helpful.

How to debug

I really gained a lot just seeing the use of the various profile techniques used in a hands on fashion. The examples of speed ups were really appreciated as well.

I think analyzing processing speeds and debugging were very useful and have already begun to use these tools. The last section on parallelization was also most useful to me.

Implementation of theory via R code

Importance of vectorization.

Included code, explanations and resources

Information about parallel programming with R

Intro to Rcpp

Intro to relevant software and a few simple how-to examples.

It provided good information for getting the most out of R.

It was a very good overview of the tools that are available and it was good to hear the presenter's opinions on how they compare.

It was good overview to being an R power user.

It was straight to the point. All key points are in slides.

It was useful as an introduction to the topic. In some aspects more details would have been, but it was excellent to know ups and downs of R and how to deal with them.

Learning how to debug.

Learning best practices for implementing particular types of parallel programming/debugging/etc.

Learning how to optimize code (e.g. Loop structures, vectoring) and benchmarking

Most useful was learning more about R programming.

No part stood out to me

Optimization

Optimization and debugging

Optimization section (though I'm a fairly low-level R user)

Optimization; the next day I increased speed of two of my scripts 15X.

Overview of best practices in R programming (for efficiency) and the trade-offs associated with different approaches. Places to look for more information.

Overview of different approaches to parallelization with concrete examples

Overview of issues and possible solutions

Overview of many topics!

Parallel

Parallel

Parallel R

Parallel R and hands on

Practical for my programming work

Presentation and hands-on material and examples

Profiling, parallel computing, other language APIs

Profiling

Profiling and Rcpp

Profiling, non-distributed parallel processing

Rcpp

Rcpp, I have seen some use of this before. Looks like a great package that I would really like to need; however since I currently don't I, again, won't get to spend the time to learn it.

Resources for more information and understanding of the default BLAS in R

Section on profiling

Showing how to use Rcpp, I use TMB, so it won't be overly used at the moment, but I feel to run simulations I would use it!

Some of the tips provided during the training and how they could impact performance... That was informative

Streaming is engaging. I am not sure if I could cover the same content if it was just reading.

That answers were provided with the exercises. This really adds value.

The basic overview of lots of tools, along with pdfs of the presentation and exercises. This way, I could listen to the presentation, highlight the parts that were most useful for me, and then go back and work on those exercises and use the information from the presentation pdfs to research them further.

The debugging portion

The examples presented and the exercises assigned.

The Exercises

The exercises.

The fact that there was good background info given followed by the more detail driven material then an opportunity to try it for myself.

The first half of the lectures were very useful

The first two sections.

The hands on activities.

Add more time

Add some more examples

Add some walk-throughs

Additional hands on exercises.

After the debugging and performance topics, most parts were too technical for my level of understanding.

Application process was a little confusing.

Breaks to work on assignments seemed too short for what we were given, though I am pretty sure part of this is by design.

Create other tutorial for more advance users

Demonstrations of code and exercises

Exercises to run R jobs in XSEDE should be provided.

Focus more on real applications using R in various scientific area, but just focusing R as a programming language and compare with other programing languages.

Font on slides should be larger. It was difficult to see text on slides, especially the examples. Having to have PDF of slides open as well as view presentation was cumbersome.

For all those that are newer to program, myself included, it would have been useful to dive a bit more into defining the more complicated topics and goals.

For me, less on the parallel and foreign language APIs. I mostly use CRAN packages that handle most of the language magic for me. Once again, I suspect that I am not representative of the intended audience. It was titled HPC w/ R.

Get rid of intro part. People attending this kind of tutorial are likely not beginners.

Given the interesting topics and their practical nature I would recommend to split the tutorial in two or just address the half of the topics with more examples of code and applications.

Go more into detail!

Hands on tutorials given out before the lecture

Have more examples allot more time to go into more detail on the covered topics (perhaps a multi-day tutorial)

Have the presenter talk slower

Host the live stream on YouTube/vimeo/any media player which allows online participants to pause and rewind the stream in the moment so that they can take advantage of the live question asking but also not miss anything due to distractions in the moment.

I actually thought it was good

I am comfortable in R, but don't have any background in C. The intro was mostly things I knew, the sections on debugging and profiling were very useful since I could follow along but the material was new. The later sections on Rcpp and parallel were mostly foreign and I had a harder time keeping up. So for me, going a little slower in that section would have been better. I appreciated that exercises were included, but didn't really have time during the tutorial to complete any of them.

I am not an experienced programmer, so compiling and using other programming languages is not as useful to me as other parts of the course, however I do see their utility and a need for expanding to other languages for HPC.

I could not attend the entire tutorial, due to other responsibilities. I could use info about offline viewing afterwards.

I do not Twitter, and as a consequence, I had no way to ask questions remotely. If there was a different way to submit questions via your web site or something that doesn't require I sign up for unwanted internet services would be better for me. I may be the last person on earth who doesn't Twitter, but there it is.

I liked it the way it was.

I think less material on the C++ integration would be better. It requires too much prior CS knowledge.

I thought it was very well presented and the material was very beneficial. Not sure I would have changed anything.

I would like to attend to more tutorials like this, is it possible to receive an email?

I would have left off the first 1/2, focused more on the parallel part or change the name. I felt that HPC was not a focus of the course despite the title.

I would just try to make it easier for those following up the webcast to participate. The twitter strategy did not get all questions answered.

I would like to have a more basic intro to R

I would make it more detailed and focused.

I would upload R code and I would spend more time in the latter subjects. With REAL examples, not replications of generating random numbers or multiplying matrices.

In the title I would put less emphasis on the parallelization and more on performance optimization to avoid false expectations such as mine.

It was hard to really try much of the hands on work in the periods allotted, especially since these breaks also had to serve as a time for me to eat, etc.

It was just fine but I had to re-load a few times because the video would stop working. But this problem might be on my side...

It was quite or really good for general public.

It was solid the way it was.

It's good

Last bit was too much like an advertisement

Less content, more examples, more time for questions.

Little to nothing.

Live feed link

Make it accessible. What software is required and please make it possible to check connections in advance to identify problems

Maybe add more contents.

Maybe split it into two short sessions since four-hour is a little long for me.

More about running R in a cluster environment.

More details about C++ for R.

More exercises could be included for experimentation afterward. The details of Rcpp could be better conveyed, perhaps through comparing different attempts at the same problem. A benchmark plot could be included in the exercises (hope I didn't just miss it...).

More hands-on exercises

More in-depth coverage of advanced topics

More time for hands on would have helped, or more presentation of solving problems.

More time on openMP in a Fortran/C code and linking it to the R

More time spend on the final topic, the actual discussion of HPC tools for R, and possibly providing some working codes. If a user environment could be set up for participants to actually test the MPI-related tools that would be amazing, but I realize there are difficulties in facilitating such exercises.

More time to go over the exercise problems.

Needed more basics of paralleled programming

Pdbbase does NOT work if R is built with MKL instead of BLAS (ticket open on GitHub). So users of Revolution R which is very popular in corporate world cannot use it. Also see my answer on the question about technology.

Perhaps break into 2 parts so topics could be covered a bit more in depth

Perhaps split it into two sessions and go more in-depth on a couple of the topics.

Precise the content beforehand. It only said that beginners could follow, however, I did not have the same definition of beginner. If I knew what was going to be talked about before registering, I would have known that it was not directed to me.

R intro could have been more comprehensive (e.g. Data frames) or dropped completely and made a prerequisite

Remove some of the more introductory aspects.

Replace exercises with proper working example. Working example could be prepared in the form of an R packaged distributed to the class

Simplify and reduce the time spending on the basic introduction part

Slides/ wiki page or some sort of material would have been helpful to get a better idea of what would be covered. Also, a schedule would have been good so that I could have tuned in when relevant information was going to be discussed.

Slightly slower pace, 200+ slides in 4 hours is fast...

Slow it down a bit. Our group of R users felt the tutorial went quickly but manageable through the debugging, backtracked to the loop/apply stuff (since we are familiar with these tools already) and zoomed through the rest so fast we were not able to follow.

Spend more time on the details of approaches. More focus on parallel processing

Split into two parts and expand both to a full afternoon

Spread to 2 classes

The pace should be a bit slower.

Too little time for hands-on: I would provide the exercises but not spend time on them during the tutorial

Use a whole day and more depth vis a vis the high end stuff, like parallel computing, also parallel computing on workstations

Use other mean than twitter to communicate with the speaker

Very long.

Would have been nice to go over at least one exercise per session at the end of the time period; especially for online attendees who cannot directly ask their question

Due to the slow connection, I drop out after the first session. This is partly the reason for my neutral evaluation. The instructor promised that I edited version will be posted. I will be looking forward to that.

Enjoyed the instructors enthusiasm

Everything fine

Good job, well done!

Great job and thanks for making it accessible!!!! Fun and interesting talk!

Great job!

Great job.

Great presentation! Thank you.

Great tutorial and knowledgeable presenter. I hope to see a focused R parallelization tutorial sometime in the future.

Great tutorial. Very useful and convenient.

Had a great experience! Thought the lectures were the right lengths. Thought the spacing of the tutorial was great, with the mix of lectures and breaks. Really appreciated the breaks after each lecture to let me rest my poor brain.

I found the lectures very useful. I will attend more as they become available.

I hope that NIMBioS will provide more tutorial like this!

I only tuned in for the parallelization portion but I found it helpful.

I really appreciated that this was open to the public. I found it to be a very informative session and the speaker was fantastic.

I really enjoyed the presentation, and I learned things that have already proven useful. However, I felt shut out from asking questions because I do not use twitter. Other on-line meetings I have "attended" provided a messaging system for asking questions.

I thought the topics covered were very useful. Perhaps more time could have been spent on each topic to get a more thorough understanding; more involved examples in addition to the toy examples (though useful) would've been good. Starting out with something on R basics (i.e. Object types, loops, and conditionals) might have been useful.

I wish it could be separated in a few weeks and each section has more examples.

I wish the information about this tutorial was more specific about the target audience! The target audience was computer programmers rather than users. If I had known that I probably would not have attended. I was looking to gain knowledge about making some R code I have run faster. While this was covered to some extent, it was a relatively small part of the tutorial. Then, really basic R things like "for" loops were covered in great detail, and translation between computer languages was covered in even more detail. Again, if this emphasis and the target audience was stated more explicitly up front, I would have been able to make a better assessment as to whether this tutorial would achieve what I wanted.

I would rather suggest to use YouTube live / etc... Which provide robust platform

It was a really nice seminar as an introduction, the instructor was very pleasant and knowledgeable. I would like there to be a follow up seminar that builds on this one to expand the horizon on high throughput computation

It was a very good workshop. Thanks.

It would be efficient if you solicited the questions for the follow up session ahead of time

It's a nice tutorial. Thank you very much!

Look forward to attending additional workshops based on this experience

Loved it, do it again!

Make sure the link is working before the webinar

More of these please!

Nice job! Thanks for offering this tutorial.

Nice workshop, I feel I have better awareness of HPC tools for R and a starting point to pursue learning more about their use.

Often the academic world does itself a great disservice when it strictly adheres to the philosophy of teaching to the exclusion of training. This was not the case in this tutorial; I greatly appreciated the content and the training approach.

Overall, I learned something useful from the workshop.

Please review your choice of medium and make this more accessible. I would try again but it was pretty frustrating not to be able to participate

Quite helpful, especially suggestion for further resources.

R code next time, real modeling applications, and more time on the latter sections. But overall a great learning experience. Thanks!

Speaker was funny, and clearly well suited to speak on the matter. Presentation was very well done.

Suggest organizing a webinar for beginners.

Thank You for doing this!!

Thank you for spreading your awesome R knowledge around to other departments!

Thank you for the magnificent tutorial

Thank you so much for this webinar.

Thank you!

Thank you! I appreciate the opportunity to participate in events like this remotely and consider it an important part of my professional development. I look forward to hearing about similar opportunities!

Thanks for this great tutorial!

The content of the webinar was at a much higher level than I was expecting given the minimum prerequisites. I would recommend either additional background and introduction, or suggesting a higher level of experience with R and other programming languages.

The speaker was very engaging and I enjoyed the webinar.

The tutorial was a little over my head, but conducted very well.

This was great that it was online like this so many people could attend and learn this information. It was also great that it was offered for free to those online.

This was well worth my time. I very much enjoyed the presentation of the material and the presenter was very good. I plan to look up some of the materials referenced during the workshop to keep building experience with R. I can't imagine what my work would be like without using R and the tremendous variety of packages and applications for them.

Very valuable presentation!

You did a great job!