Learning R

Tomáš Křehlík

R is a general programming language with special emphasis on statistical an numerical problems. There are several quirks that the language has compared to other languages and generally as with any other languages, the learning curve is high in the beginning, but the skill you acquire is very, very valuable.

These days, applied statistical analysis is as much about using the right methods as about presenting your results. And not only in academia but also in business and anywhere else. Without good presentation even great result can fall behind in comparison. R is an environment where you can achieve both powerful analysis and powerful presentation. R is used by Citi Bank, Visa, Novartis, Google, Merck, and many others.[1] On the other hand New York Times uses R for creating their graphs and charts. These are only few examples from huge pile of companies that use R. And the number is growing every year...

Moreover, R is free and you can take it wherever you want and do whatever you want with it. All parts of R are distributed under GPL license, which means that you can use R for free for any activity you want. You can be making millions using R and nobody can charge you a dollar. This is in sharp contrast with other tools (including MATLAB, Stata, etc.) which can potentially cost a lot of money if you are using them in business.

In this document, I provide materials and observations that I found useful in learning R.

Basic concepts

There are three possible starting positions to learn R. One of them is that you have not seen Stata or any programming language, including MATLAB, before. This I hope does not apply in here as you should have had Stata as part of bachelor's programe.

Second one, probably the most common in here, is that you have worked with Stata but not with other programming language. For those people, R will probably seem quite unfriendly, as I have heard from friends. The truth is that R is about equally friendly to users, only it contains more brackets and equality signs. Useful reference for those who know Stata is at Princeton webpage.[2] What will be new for you is the common use of programming concepts in to code. By programming concepts I mean assigning a lot of variables, using loops and defining functions in almost any script that you will write.

The third is that you have seen and written some code in other programming languages and you know statistics. For you the learning will be probably the easiest. Because everything can be done using loops, it is just more laborious than looking for specialized code.

Basic strategies to learn

1. Complete a wonderful online introductory course to R.
2. Download and print for yourself a copy of cheat sheet. (Cheat sheets are used in

[1] Look at list of customers of Revolution R, which is a commercial blend of R.

[2] Even though, I personally think that some of the things can be done a bit more simply in R.

programmers community quite a lot, because one cannot remember names of all the functions in all the languages.)

3. Read some of the introductory literature provided below.

4. Open R and code and code and code. Conceptually, two types of people exist.

   - If you like real-world problems, then think of some real data, download them and analyze, plot and solve the problems you run into.

   - If you like set up puzzles, go here and solve them all.

   - Alternatively you might try to replicate things you have done in other languages, that is also very useful.

Basic recommendations, study sources and R packages

1. Be lazy! If you have a problem then probably someone have already thought of that before and there is answer in Google or Stackoverflow. Use "`?name_of_function`" to get help! Reinventing the wheel is often very costly so first search.

2. Very useful resource is UCLA. They have description of basic procedures in various languages. For econometricians, there is useful Youtube playlist. It covers a lot of things we will be doing during the seminars. You can also find Wikibook, that is not really exhaustive but sometimes provides useful examples. On the page of R, if you click on Manuals in the left pane you will get several manuals which are well written and are useful reference once you have already done something in R.

3. In many packages, there are so called vignettes which are typically a very good review of theory and implementation of the package in R. They are full of examples. To just mention a few that you might find useful: `gmm` (General Method of Moments package), `vars` (VAR, SVEC, SVAR and VECM models package), and `plm` (Panel Data Methods). I found them to be also good study material that is complementing the textbooks.

4. It is worth investing your time to learn several things in R. First is use of "`apply`" function and its variants. It makes many thing much easier and faster to calculate.

5. Basic way how to get expand R by functionality that is not exactly built into the basic package is to load (or build, it is not that hard!!!) a package that is in The Comprehensive R Archive Network (CRAN). There are thousands of packages and the hardest thing is to select the right one. One useful reference is the views page on CRAN which sorts the packages into different topics.

6. Here is little list of packages I find using myself again and again.

   - Econometrics: `vars`, `MASS`, `car`, `sandwich`, `urca`, `xts`, `rugarch`, `rmgarch`.

   - Plotting and tools: `plyr`, `lubridate`, `reshape`, `ggplot2`.

   - (Big data: `data.table`, `parallel`, `foreach`; this is only for those that will eventually want to deal with datasets that are 500MB and bigger.)