Prerequisites for the Baruch MFE "Big Data in Finance" course

Document name: "BDiF2015 – TN0000.pdf"
Date/revision: Mon 26th Jan 2015 / Rev A.
Author: Andrew Sheppard, © 2015.

---

1. Introduction

First, let me tell you a little about the course.

Everyone talks about Big Data; in this course, you will learn how to use Big Data to enhance your career or start a new career in finance. The course is taught by an experienced Big Data in finance practitioner and designed for current and future practitioners.

The course has 7 classes over 7 weeks and provides access to real-world financial data and high performing computing resources and direct interaction with the instructor.

The course is practical and hands-on. That is not to say that I skimp on the conceptual stuff, but there is a strong emphasis on building Big Data tools and working with Big Data. You can find out more about the rationale and philosophy behind the course by reading the article, "Big Data in Finance", in the QuantNet Guide.

I expect students to join the course from very different backgrounds. Some will have the skills necessary to start the course, others will need some help. This technical note describes the materials that are available to bring all students up-to-speed before they start the course.

2. Course Prerequisites

So, what do you need to know before you can begin the course?

The "Big Data in Finance" course is a fast-paced course that covers a lot of material in just 7 classes. Prerequisites have been kept to a minimum, but understanding the concepts and tools below will make the course both easier and more enjoyable:

- ***Programming***: Some experience with programming. Ideally in C/C++ (just the basic language and syntax), but experience in any language will do. Familiarity with some dynamic language, such as Python, would be a plus, but is not essential.

- ***Mathematics***: linear algebra, matrices, calculus, and basic statistical concepts. All at the high school level or above.
- ***Command line & scripting***: Familiarity working at the command line, whether on Linux, Windows or Mac.
- ***Data analysis software***: Some familiarity working with data and graphing data in Excel, R, or other software package.
- ***Technical writing***: An ability to express yourself succinctly, and clearly, in written English; and you must be comfortable reading documentation of various kinds: course notes, assignment specifications, academic papers and articles, software specifications (APIs) and other technical documentation.

The course requires the students to build software tools, which means you must know how to program. As indicated above, the starting requirements are quite modest and beyond those basics I'll teach you what you need to know.

The course has a strong, practical, hands-on approach and most assignments require writing code and building software tools. By the end of the course, not only will you have Big Data skills, you can expect to have significantly improved your programming skills also. Moreover, the assignments will form a portfolio of projects that you can use to impress your current boss for a raise, or a prospective employer for a job; it's always better to "show" than "tell"!

Appendix I suggests some resources for you to start learning the prerequisite skills for the course.

3. Computer Requirements

In terms of prerequisites, the computer requirements have been kept to an absolute minimum. I don't want you to get bogged down with installing software; rather, I want you to start running programs and learning; specifically learning how to program in parallel. Early concepts are introduced using an online web based application that will compile and run your code. You can find that application by following this link: http://coliru.stacked-crooked.com/

To use this online application you will only have to have access to a web browser and an Internet connection. I tested and found the application to work in these browsers ("or above" in terms of version number):

- Google Chrome, Version Version 31.0.1650.63.
- Firefox, Version 26.0.
- Safari, Version 6.1.

Basically, you shouldn't have problems if you are using a relatively new version of your favorite browser.

4. **Prerequisite Materials**

In addition to the resources given in **Appendix I**, I have prepared these technical notes to help with some of the key skills you will need in advance of working with financial Big Data during the course:

- "BDiF2015 – TN0000.pdf" : Prerequisites for the Baruch MFE "Big Data in Finance" course (this document).
- "BDiF2015 – TN0001.pdf" : Running your first multithread C++11 program.
- "BDiF2015 – TN0002.pdf" : Running your first multithread simulation using C++11.

Appendix I

Here are some suggestions for free online resources and books you can buy to learn the basic prerequisites for the course:

- ***Programming***:
  - ***Online***: C++ Beginner's Tutorial
  - ***Book***: "Programming: Principles and Practice Using C++" by Bjarne Stroustrup, 2008, Addison-Wesley, ISBN 0321543726.
- ***Mathematics***:
  - ***Online***: Concrete Mathematics
  - ***Book***: "Concrete Mathematics: A Foundation for Computer Science" by Ronald Graham, Donald Knuth, and Oren Patashnik, 1994, Addison-Wesley, ISBN 0201558025.
- ***Command line & scripting***:
  - ***Online***: 1) Windows ~ Windows Command Prompt in 15 minutes / Windows Powershell ~ PowerShell Tutorials, 2) Linux ~ LinuxCommand.org, and 3) Mac OSX ~ Introduction to the OS X Unix Command Line.
  - ***Books***: 1) Windows ~ I'm afraid I don't think there is a book I would recommend / Windows Powershell ~ "Windows PowerShell in Action" by Bruce Payette, 2011, Manning Publications, ISBN 9781935182139, 2) Linux ~ "The Linux Command Line: A Complete Introduction" by William Shotts, 2012, No Starch Press, ISBN 1593273894, and 3) Mac OSX ~ "Learning Unix for OS X: Going Deep With the Terminal and Shell" by Dave Taylor, 2012, O'Reilly Media, ISBN 1449332315.
- ***Data analysis software***:
  - ***Online***: 1) Excel ~ Using MS Excel 2003 to Analyze Data: An Introductory Tutorial, and 2) R ~ Quick-R.
  - ***Books***: 1) Excel ~ "Statistical Analysis with Excel For Dummies" by Joseph Schmuller, 2009, For Dummies, ISBN 0470454067, and 2)

R ~ “Learning R” by Richard Cotton, 2013, O’Reilly Media, ISBN 1449357105.

- ***Technical writing***:
  - ***Online***: Guide to Technical Writing
  - ***Book***: “The Elements of Style” by Strunk and White, 1999, Longman Publishing, ISBN 020530902X.
