BE 434/535 Biosystems Analytics

The field of "bioinformatics" is biology plus data science. For the latter, most people find Unix-like operating systems to be the most efficient way to conduct research. Almost all our tools run in Unix, most of them from the command line, so the bioinformatician must know how to move data around, run programs, and chain the output of one program into another to create analysis pipelines.

Code

All the code examples presented here can be found at:

`https://github.com/hurwitzlab/biosys-analytics`

About The Author

I went to college (the University of North Texas, 1990) thinking I might study music and become a professional drummer. I decided against that particular career but didn't have an alternative. I changed my major a couple of times (business, communications) before deciding on an English literature degree just so I could finish Never did programming cross my mind, and biology was my most loathed science.

After college, I looked for a job that 1) would be interesting and 2) people would pay me to do. At my first job (1995), I learned Microsoft's Access database on Windows 3.1 and created my company's first website by writing HTML into Notepad and using a Windows FTP client to upload to our ISP. The die was cast. My next job was writing a technical manual for a piece of software written in VisualBasic. I was offered a position to learn VB and support the program and another Access program. I went on to learn another database (DBase IV) and language (Delphi). My next job was developing Windows desktop applications in Delphi/Interbase, but I was itching to get into web applications. So my next job (c. 1998) was writing VBScript in Microsoft ASP with SQLServer as a database. At this point, I was fed up with Microsoft and rediscovered Unix.

I discovered that my ISP offered "shell" accounts on their servers accessible by "telnet." In college I had gone to the computer labs where I used Unix programs like "pine" for email and "talk" for chatting, and so I remembered how to get around a shell. I started reading more about Unix on the Internet and how people were using the Perl programming language with CGI (common gateway interface) to create interactive web pages. The more I learned about Unix and Perl and "open source/free" software, the more I realized I'd found my tribe. At my next job at boston.com (1999) I moved into developing web apps on Linux

platforms using the Apache web server with the MySQL database and Perl (the "LAMP" stack).

Around 2001, I saw that a very celebrated Perl developer named Lincoln Stein was looking to hire people. I got hired to work on a comparative plant genomics database called "Gramene." Lincoln was a very important character in a fairly new field called "bioinformatics" (cf. "How Perl Saved the Human Genome Project") and he ran a research lab at Cold Spring Harbor Laboratory in Cold Spring Harbor, NY. Lincoln hired me to write a web-based visual comparative map application (CMap, PMID: 19648141) to augment existing web genome browsers like the UCSC browser, the Ensembl browser, and Lincoln's own Gbrowse. This was my entree into the world of biology and genomics. Around 2004, Lincoln hired Bonnie Hurwitz who left a few years later to earn her PhD from the University of Arizona. In 2014, Bonnie set up her new lab at the Universit of Arizona and hired me.

Outside of science and coding, I also enjoy biking, cooking, and playing music.