22S:166
Computing in Statistics

Introduction

Lecture 1
August 22, 2016

Kate Cowles
374 SH, 335-0727
kate-cowles@uiowa.edu

Statistical endeavors

- three branches
  - applied statistics and data analysis
  - development of statistical methods and software
  - research in statistical theory
- computing essential to all of them

> I keep saying that the sexy job in the next 10 years will be statisticians. And I'm not kidding.

Hal Varian, chief economist at Google (New York Times, Aug. 6, 2009)
`http://www.nytimes.com/2009/08/06/technology06stats.html`

Goals of this course are to develop:

- intelligent use of appropriate computing tools for both statistical endeavors
  - R/Splus
  - SAS
  - database management software and concepts
- understanding of important statistical computing algorithms
  - Newton's method
  - EM algorithm
  - the bootstrap
- ability to design and implement simulation studies

- communication of statistical ideas in words, numbers, and graphics
  - LaTeX
  - format of scientific reporting

Types of computer products

This section and the next 2 borrow heavily from Chapter 1 of the course notes for "Statistical Computing and Graphics" by Frank Harrell
`hesweb1.med.virginia.edu/biostat/teaching/statcomp`

- operating systems: make the computer itself work
  e.g. Linux, Windows, Unix, MacOS
- applications: perform specific tasks
  e.g. Microsoft Word, Excel, S-Plus, OpenOffice, R, SAS, ...
- commercial systems
  - code and lists of bugs are secret
  - expensive
  - require upgrading and relicensing
  - Microsoft products, S-Plus, SAS, SPSS, Unix, etc.
- free Open Source systems

  - revolution in software availability and function from the open source movement
  - can see all code, change it, learn from it
  - quality generally quite good
    * often better than that of commercially-developed software because Open Source software has been tested by more people under more different conditions
  - more rapid updates
  - most products have an active and helpful user news group
  - generally lack some fancy features like extensive GUI
  - Linux, LaTeX , R

User interfaces: graphical vs command line

- graphical (GUI, mouse, menus)
  - easier to learn
  - less flexible
  - repetitive when the same tasks have to be repeated
  - hard to document the exact steps taken
  - hard to reproduce results
- command line interfaces
  - harder to learn
  - more flexible and powerful
  - can save commands in scripts to replay when the same tasks have to be performed repeatedly
  - can write generic commands to facilitate running different analyses with the same structure

Types of user files

- text
- binary
- graphics files

Linux history

The material in this section borrows heavily from Section 1.1 of *Introduction to Linux: A Hands on Guide* by Machtelt Garrels.

`http://www.tldp.org/LDP/intro-linux/intro-linux.pdf`

- Unix
  - 1969: team of developers at Bell Laboratories began work on solution to problem of software incompatibility
    - at that time, every model of computer had different operating system
    - software was customized to specific purposes, and ran on only one type of computer system
  - UNIX operating system needed only small piece of code specific to one type of computer: the *kernel*
  - operating system (and all other functions) built around kernel

  - higher-level programming language C specially developed for creating UNIX
  - at first used only in very large computing environments — universities, government, large corporations with mainframes and minicomputers
- developments in 1970's and 1980's
  - continued development of UNIX
  - support of UNIX in products of increasing numbers of hardware and software vendors
  - invention of personal computers
  - by end of 1980's, several versions of UNIX available for PC architecture, but not free

- Linus Torvalds and Linux
  - computer science student at University of Helsinki
  - goal: to create a freely-available operating system that was compliant with original UNIX
  - began working on it in early 1990's
  - other coders jumped aboard to develop drivers to make Linux usable with more and more hardware
  - 12000 Linux users by 1993
  - all features of UNIX added over few more years

- Linux today
  - only operating system in the world that runs on as wide a range of hardware
    - desktop workstations
    - mid- and high-end servers
    - PDAs, netbooks, experimental wristwatches, etc.
  - well known as a stable and reliable platform for servers
  - examples of users
    - Amazon (Internet book seller)
    - United States Post Office
    - German army
    - high-energy physics grid

Logging in, activating the user interface, and logging out

- PC-based Linux systems have two basic modes: graphical and text
- in graphical mode (usually the default)
  - login requires inputting user name in one window and password in another
  - make sure mouse pointer is in the login window; press Enter after entering the username and after entering the password
  - open a terminal window by left clicking on icon of a computer screen at bottom of screen
  - log out by
    - closing all terminal windows and applications
    - clicking "System" at bottom of screen and selecting "Log Out"

  - in text mode
    - whole screen is black, with white characters
    - you are prompted for username and password; press enter after typing in each one
    - log out by entering **logout** command and pressing enter

Linux student computer labs

- UI Mathematical Sciences educational computer laboratories:
  - 346 SH (primarily for statistics students)
  - B5 MLH and 301 MLH (primarily for math and c.s.)
- must bring your own printer paper if you wish to print
- to access Internet, click Firefox icon (world with a flaming fox) the bottom of the screen
- to log out, click icon of a running person and select "Log Out"

There are two ways to access the Linux computers remotely.

- Select "NX Client/CSG Linux Desktop" from the "All Programs" menu. Use it to log into the gateway to the DIVMS Linux network.
- Alternatively, use any browser to connect to: `https://fastx.divms.uiowa.edu/` and log in using your hawkid and password. Click the plus sign when instructed to start a new session. Select the user interface that you wish to use (Mate or XFCE).