R bootcamp - tentative syllabus/schedule

July 26, 2013

Comments

- 40-50 minutes per session for presentation/demo, 20 minutes for breakout (including answers)
- try to intersperse little coding questions/challenges within the presentation material
- ask participants to play with code as we go along and see if they can break it or tweak it in interesting ways (perhaps reporting this on the bSpace forum)

Schedule/syllabus

- Day 1 morning (8:30-12:30) (learning R)
  - Session 0: introduction, what is R, starting R, why R? why not R? (Chris P.) (10 minutes)
  - Session 1: basics of R, with Rstudio (Chris P.)
    - R as a calculator
    - helpful shortcuts: tab-complete, up arrow, Ctrl-{up arrow}
    - vectors and indexing and subset assignment
    - some basic functions; help()
    - vectorized calculations, comparisons
    - basic R objects: vectors, matrices, dataframes, lists
    - managing R objects, the R workspace, save.image(), str()
    - basic graphics
    - breakout [perhaps a basic plotting problem that involves some data manipulation]
  - Session 2: Working with data (Chris P.)

        * dataframes/matrices and subsetting, and subset assigning (LHS)
        * more advanced object types (including factors, date-time, classes), attributes, NAs
        * strings (a little bit) and cleaning data
        * reading/writing data; working directory, foreign package
        * breakout problem [work on reading a dataset in and manipulating a data frame]
    - Break (20 minutes)
    - Session 3: Manipulating data (Chris P.)
        * more on vectorized calculations and efficiency (pre-allocate, look-up tables by indexing, linear algebra/BLAS)
        * apply, lapply
        * data summary, aggregation, merge
        * breakout [work on a data aggregation problem]

- Day 1 afternoon (1:30-5:00) (programming and real-world work)
    - Session 4: R resources (Chris P.) (30 minutes)
        * packages - installing, loading, namespaces
        * getting help - R help, Stack Overflow, Google "in R", reproducible examples, sessionInfo() [could shift to last session Sunday]
    - Session 5: programming in R (Jacob)
        * loops, if-else
        * writing your own functions, function arguments, functions as objects
        * basic scoping and environments
        * intro to OOP in R: using existing S3 classes/methods (mention S4 exists)
        * breakout [write a function with arguments - perhaps a sort function using order()]
    - Break (20 minutes)
    - Session 6: doing useful stuff (Chris K.)
        * stratified analyses: groupwise operations (see plyr: subset, mutate, summarise, arrange); split-apply-combine
        * reshape
        * regression, GLMs
        * breakout [assign overnight homework - data analysis problem, perhaps with some programming - bootstrap or cross-validation]

- Day 2 morning (9-12:30) (more real-world work)
  - Session 7: doing more useful stuff (Chris P.)
    * go over homework
    * smoothing
    * optimization
    * simulation, sample()
    * breakout [a small simulation study perhaps]
  - Session 8: more graphics (Chris K.)
    * exporting graphics (vector/raster)
    * more advanced graphics including lattice/ggplot2 (see Knowles)
    * breakout [creating a nice lattice/ggplot2 graphic]
  - Break (20 minutes)
  - Session 9: Workflows, coding practices, and project management (perhaps Jarrod; Chris has draft of material already)
    * scripting, source(); separating data, code, figures
    * R in batch mode and command line mode
    * timing, memory use, debugging
    * reproducible research with knitr, Rmd
    * version control for code and data; Git
    * breakout [assess timing and memory use for some code?]
- Day 2 afternoon (1:30-4:30) (more advanced topics)
  - Session 10: quick tastes of advanced topics (Chris P.)
    * strings and regular expressions, encodings
    * OOP (S3, S4, ReferenceClasses)
    * computing on the language (using R to write and evaluate R code)
    * working with databases
    * breakout [set up some OOP code]
  - Break (20 minutes)
  - Session 11: parallel processing (Chris P.)

    * foreach
    * parApply and variants
    * RNG issues
    * breakout [do some parallel computation]
- Session 12: Wrapping up (Chris P.) (15 minutes)
    * R inconsistencies and different ways to do things (see Knowles slide)
    * Where to learn more and get help