**Presentation guidelines for STAT 992**

The purpose of these presentations is to give students an opportunity to explore computational statistics topics that they are interested in. Students will spend a few weeks learning on their own a particular topic and then present this topic to the class during a ½ class-period presentation. Along with the presentation, students will need to construct a homework assignment (approximately 10 points in length) that assesses other students’ comprehension of the topic. An answer key for this assignment needs to be written too and given to your instructor. During a class period after the presentation, we will discuss the assignment and other items about the presentation topic as needed.

Below are specifics about the presentation:

1. Topics

All topics must be computational statistics related. A list of potential topics are given later in this handout. If you have other ideas for topics, please talk to me individually about them. All topics need to be pre-approved by me!

1. Presentation styles

The presentation should be given in a way that all students in the class will understand the topic presented! Thus, the presentation should be in a teaching style rather than a seminar or JSM presentation style. One way to present a topic is to use a similar style to what I use in the courses that I teach.

1. Grading

40% of grade: Presentation quality

Did the presentation appear to be practiced?

Was there a logical flow to the presentation?
Were there transitions between items?

Were the slides or other visual tools readable?

40% of grade: Content

Was there a meaningful amount of items covered?

Was everything correct?

20% of grade: Inside and outside of class learning tools

Handouts for students to examine (e.g., lecture notes or slides)

Homework assignment and answer key

Computer programs (if needed)

Example presentations given by Aimee Schwab (this was for a whole class period) and Trevor Hefley are available on the graded materials web page of the course website.

**Potential topics**

1. Computer programming language – Present the basics of how to program in a language along with basic statistical aspects of it. One component of the presentation could be how to use the language with R (e.g., with C++, one can use the Rcpp package or use a .Call() function to an executable file). Of course, all students need to have access to the language and/or compiler in order to complete an assignment. Potential languages include:
	1. Julia
	2. Python
	3. C++
	4. Perl
	5. Ruby
2. R packages for computational statistics – Present the basics of how to use functions within an R package. For example, the High Performance Computing, Machine Learning, and Optimization task views (<http://cran.r-project.org/web/views>) provide listings of packages that may be appropriate. In particular, the Rmpi package gives another way to do parallel processing. An advantage of using this package is that you can use cores across nodes.
3. Using GPUs for statistical calculations – Rather than using CPUs for computations, graphical processing units (GPUs) can be used as well. The advantage to using these processors is that there are a large number of cores available on a single graphics card. Some R packages have been developed for using GPUs – see the “Parallel computing: GPUs” section at <http://cran.r-project.org/web/views/HighPerformanceComputing.html>.
4. Mathematical software – Present the basics of how to use the software along with the basic statistical aspects of it. In order to do this presentation, you will need to make sure that all students and the instructor can access the software that you are using. Potential software packages include:
	1. Matlab
	2. Scilab
	3. Octave
	4. Maple
	5. Mathematica
5. Linux operating systems – A presentation could continue my discussion from the HPC section. Alternatively, a presentation could focus on how to install (perhaps just on a USB drive or SD card) and then use desktop versions of Linux distributions, such as Ubuntu, Mint Linux, or Fedora, along with software available for it like R.
6. Kaggle (<http://www.kaggle.com>) – This is a company focused on data analysis competitions. They host competitions for many different entities from university courses to corporations that will pay for winning entries. A presentation could include the basics about kaggle and then information about how to enter and participate in a competition. Information on how to start your own competition should be included as well.
7. Big data aspects of computational statistics – A presentation could examine how to handle big data in R. For example, R by default puts data frames and other data storage objects in a computer’s memory. This is a BIG problem with BIG data. There have been a number of tools developed to deal with this problem. A presentation could examine these tools. Potential packages to examine for more information include biglm, bigmemory, and pbdBASE. Alternatively, a presentation could focus on Hadoop for handling big data.
8. Sampling for Bayesian methods – A presentation could focus on the basic aspects of Markov Chain Monte Carlo (MCMC) methods such as the Metropolis Hasting algorithm or Gibbs sampling. Two introductory papers on these topics are Chib and Greenberg (*American Statistician*, 1995, p. 327-335) and Casella and George (*American Statistician*, 1992, p. 167-174). Givens and Hoeting (2013) and Robert and Casella (2010) provide introductions to these topics too.
9. Topics can be obtained from the books given as suggested readings for this course. For example, topics from Givens and Hoeting (2013) include genetic algorithms, nonparametric density estimation, and a few bootstrap methods.

If none of these ideas interest you, I recommend looking at

* *The R Journal* ([http://journal.r-project.org](http://journal.r-project.org/)) and its predecessor *R News* ([http://www.r-project.org/doc/Rnews](http://www.r-project.org/doc/Rnews/index.html))
* *Journal of Statistical Software* (<http://www.jstatsoft.org>)
* <http://www.r-bloggers.com>

for additional ideas.