Approaches to Broadening the Statistics Curricula

Deborah Nolan

Duncan Temple Lang

Abstract

In spite of major changes in science and technology and their enormous impact on statistical theory and practice, the statistics curricula have hardly changed. Traditional statistics courses give inadequate focus to computing and essentially ignore the vital topic of "statistical thinking and experience". The authors offer 3 separate approaches to reduce the hurdles of changing the curricula by pooling resources from statistics communities. One approach provides resources for instructors to introduce computing into the curricula in innovative ways. The other involves infrastructure for transferring research results from the research community to the teaching community to provide guided statistical experiences for students. Also described is a new approach the authors have experimented with for teaching statistical thinking.

Keywords: Dynamic Documents, Reproducible Research, Statistical thinking, Statistical Computing

About the Authors

Deborah Nolan
Statistics Department
University of California, Berkeley
Email: nolan@stat.Berkeley.edu

Deborah Nolan is a professor of statistics at the University of California, Berkeley, USA. She is a Fellow of the Institute of Mathematical Statistics, co-author and editor of four books on teaching statistics and mathematics through case studies, activities, and projects, and the recipient of several awards for teaching and mentoring. Her current research interests are in the area of cyber-infrastructure for statistics education.

Duncan Temple Lang
Statistics Department
University of California, Davis
Email: duncan@wald.ucdavis.edu

Duncan Temple Lang is an associate professor of statistics at the University of California, Davis, USA. Before joining UC Davis in 2004, he was a researcher for 7 years in the Statistics and Data Mining Group at Bell Labs, Lucent Technologies. His current research focuses on developing new languages and environments for statistical computing, integrating information technologies and computer science research into the process of scientific and statistical research in general, automated ways.

Recently, there has been a lot of discussion about what a statistics curriculum should contain and which elements are important for different types of students. For the most part, attention has been understandably focused on the introductory statistics course. This course services thousands of students who take only one statistics course. In the USA, the course typically fulfills a general education requirement of the university or a degree program requirement. There has also been considerable activity regarding the use of computers to present statistical concepts and to leverage the Web and course management software to interact with students. Recently, there has been debate as to whether statisticians should make ambitious changes using resampling, the bootstrap, and simulation in place of the more traditional mathematical topics that are seen as the fundamentals or origins of the field (Cobb, 2007). It is unclear that we are achieving the goals of basic statistical literacy by focusing on formulae or even by concentrating almost exclusively on methodology. Instead, we believe the field and students would be significantly better served by showing the challenges and applicability of statistics to everyday life, policy and scientific decision making in many contexts and by teaching students how to think statistically and creatively.

In contrast to the activity at the introductory level, there has been much less attention paid to updating the statistics curricula for other categories of students. While smaller in number, these students—undergraduate majors and minors, master's and doctoral students—are very important as they are the ones who will use statistics to further the field. Other disciplines (e.g. biology, geography, political and social sciences) are increasingly appreciating the importance of statistics and including statistical material in their curricula. Further, statistics has become a broader subject and field. However, the statistics curricula at these levels have not changed much past the introductory courses. Students taking courses for just two years may not see any modern statistical methods, leading them to a view that the important statistical ideas have all been developed. More importantly, few students will see how these methods are really used and even fewer will know at the end of their studies what a statistician actually does. This is because statisticians very rarely attempt to teach this, instead laboring over the details of various methodologies. The statistics curricula are based on presenting an intellectual infrastructure in order to understand the statistical method. This has significant consequences. As the practice of science and statistics research continues to change, its perspective and attitudes must also change so as to realize the field's potential and maximize the important influence that statistical thinking has on scientific endeavors. To a large extent, this means learning from the past and challenging the status quo. Instead of teaching the same concepts with varying degrees of mathematical rigor, statisticians need to address what is missing from the curricula. In our work, we look at what statistics students might do and how statistics programs could change to allow graduates to attain their potential.

WHAT IS MISSING FROM THE STATISTICS CURRICULUM?

Efron noted (as cited in Rossman & Chance, 2003) that theoretical statistics courses are caught in a time warp that bores course instructors and subsequently their students. Cobb (2007, p. 7) supported this position, "We may be living in the early twenty-first century, but our curriculum is still preparing students for applied work typical of the first half of the twentieth century." For example, hypothesis testing takes up a lot of space in the current undergraduate curricula because statistics texts place so much emphasis on sets of rules developed in the 1950s for various test statistics (e.g., *z*-test, one sample *t*-test, two sample *t*-test, paired *t*-test, *t*-test with nonhomogeneous variances). As a result of the large number of formulae, too little attention goes to the main notions behind testing. Even worse, these same sets of rules for testing are taught over and over again in introductory, advanced undergraduate, and graduate courses. This approach fails to teach modern developments in statistics and fails to convey the excitement of statistical practice.

Cobb (2007, p. 6) posited the reason for this focus on particular tests stems from the way the curricula have developed: “What we teach was developed a little at a time, for reasons that had a lot to do with the need to use available theory to handle problems that were essentially computational.” However, modern computing offers alternatives to these approximations, and it offers the opportunity to break from the constraints of current curricula and to design syllabi from scratch using of the large collection of computational tools and statistical experiences currently available. If students are facile with computing, they will be able to actively explore methods and their characteristics and limitations in contrast to merely accepting mathematical statements about them. Also, computationally capable students will be able to work on interesting, topical, scientific problems and apply statistical ideas.

Computing is one dimension of the statistics curricula that can attract bright, talented students and educate them in a way that will broaden the focus and impact of statistics. We want graduates who not only know statistics but can do statistics; increasingly, that means nontrivial computation on rich, varied, and large datasets from many sources in collaboration with scientists from other fields. We must modernize the curricula to include computation, like mathematics, as an important medium for expressing statistical ideas—not only for being able to apply statistical concepts but also to be able to develop the computational tools needed for research.

In addition to addressing its computational inadequacies, we advocate teaching from the vantage point of statistical concepts flowing from contextual problem solving with data. Traditional courses do "not attempt to teach what we do, and certainly not why we do it" (Efron, as cited by Rossman & Chance, 2003, p. 3), yet intuition and experience of methodology in the scientific context are essential to learning how to think statistically (Wild & Pfannkuch, 1999). Statistical thinking and practice involves so many more aspects than selecting and fitting statistical methods to data. Yet most courses focus on statistical methodology—either the theory or the application—and very few discuss in any detail the skills needed to approach a scientific problem from a statistical perspective. What is missing is the experience of connecting these methods to actual applications and rounding the student into a scientific collaborator. An application is too often just an example of how to apply a particular statistical method to some manageable data—pre-selected by the instructor to illustrate the strength of the method—rather than a scientific application that students identify and evaluate based on relevant statistical methods. For those learning statistics, the intuition and experience that are necessary for good data analysis are the hardest things to learn; they involve a very different dimension of both learning and thinking than are used in mathematical thinking developed when teaching statistical methodology. Undoubtedly, students need to understand both methodology and statistical experience. At present, the focus is primarily, if not exclusively, on the former (Bryce, Gould, Notz, & Peck, 2001).

In this chapter, we describe activities that we believe will reduce some of the hurdles in achieving these changes. The curricula changes we believe are needed come from two aspects: (a) embracing computing as an essential building block of statistical creativity and practice; and (b) focusing on statistical experience, reasoning, and applications. We are not suggesting that the mathematical approaches be discarded. Instead, we propose creating more balanced, relevant, modern curricula for various levels of students that is determined by what they need for the future, rather than what is known from the past.

One project that the authors have embarked on seeks to first collect and then disseminate materials to help faculty members argue for, introduce, and teach computing within the statistics curricula. These materials include model or template syllabi, which are intended as discussion documents describing the different elements of statistical computing and how each is important for different types of students and to get courses adopted by departments; and lecture notes, exercises, projects, tutorials, textbook chapters, a textbook in data technologies, and workshops to assist faculty members teach statistical computing. By leveraging the existing, small community of those involved in statistical computing research, we are essentially trying to seed the statistical community with resources for teaching computing as a part of the undergraduate and graduate curricula.

A second project addresses the issue of statistical practice. The authors and a colleague (Hansen, Nolan, & Temple Lang, 2006) developed a model for a summer program in statistics where undergraduates were exposed to important, topical, scientific research problems presented by statisticians working on a team with scientists who were addressing a problem. The statisticians brought data for the students to creatively explore. The students gained a sense of the data and how the data might be used to address the problem. Based on evaluations from students and faculty participants, the approach developed in the program was successful in exciting the students about the possibilities that statistics holds.

Finally, relating to statistical thinking in the curriculum, the authors' third project (Nolan & Temple Lang, 2007) offers another approach for providing students with statistical experience.

This project aims to provide a flow of materials from statistics researchers involved in scientific research to the pedagogical community using reproducible, dynamic, interactive documents. The premise is to enable researchers to document their computations and analyses so that they can be reproduced for both themselves and others (e.g., peers, reviewers, editors, managers). Researchers would work in an environment that captures their writings, computations, and thought process in an electronic notebook. In essence, the notebook is a database of all the activities within the data analysis or simulation study; and it can be projected into different views to make the information it contains available for different audiences. These documents would provide resources to instructors to assist them in teaching in new ways because they would open up the thought process and experience behind a data analysis both to the instructor and the students. This technological approach would support a model for cooperation between statisticians active in research and consulting and the community of statistics educators. Instructors would then have libraries of real case studies that include data analysis projects and current research methodologies that show how statisticians think and work.

COMPUTING IN THE CURRICULA

The current approach to teaching statistics focuses almost exclusively on mathematics. Mathematics is not always the best medium in which to teach statistical concepts; unfortunately, a heavy reliance on mathematics restricts an instructor's ability to convey statistical concepts in a fuller light. Although many tasks are easier to convey through mathematics, others are more appropriately conveyed through a plot, a simulation study, or experience with data. That is, these computational approaches offer complementary means for presenting and understanding statistical concepts (Moore, 1997). In addition, not all students appreciate the insight mathematics offers, and not all uses of mathematics offer the best insight. With the computer, students can explore data to formulate scientific questions. They can explore statistical models to understand assumptions, operating characteristics, etc. and how they behave—and they can explore both together to answer scientific questions. Moreover, computing has revolutionized statistics; many modern statistical methods are feasible only because of today's computational opportunities. It is unimaginable that statisticians today would not be facile with the computer, for they are expected to be able to access data from various sources, apply the latest statistical methodologies, and communicate their findings to others. They should be encouraged to (a) create interesting presentations of statistical findings with important consequences (e.g., as exemplified by GapMinder www.gapminder.org), (b) influence developments in the digital world (e.g., the semantic web), and (c) increase the impact of good decision making with statistics.

What Do Statisticians Need To Know?

Clearly, computing is not a fad, but something vital to the field of statistics. "Computation is now regarded as an equal and indispensable partner, along with theory and experiment, in the advance of scientific knowledge" (Society for Industrial and Applied Mathematics (SIAM) Working Group on Computational Science and Engineering Education, 2001, p. 163). Although many agree that there should be more computing in the statistics curriculum and that statistics students need to be more computationally capable and literate, it can be difficult to determine how it should change because computing has many components or dimensions. These components need to be carefully considered and prioritized in order to understand where they might fit and which groups of students would benefit from a particular emphasis.

While statistics students must learn practical details of computing (e.g., programming language syntax), we must strive to teach higher level concepts including a vocabulary and computational thinking that will enable them to discuss computational problems precisely and clearly. Vocabulary is necessary to be able to communicate—understand, express, and reason—about computational issues. As computing and data technologies continue to evolve rapidly—especially as statistics enters the era of mainstream parallel and distributed computing for scientific computing—it is essential that students are provided a good foundation rather than a thin memorization of specifics so they are able to continue to learn new aspects of computation. Statisticians must not mistakenly think all that is needed to introduce computing into the curriculum is to teach students a particular programming language. We must aim higher and more generally, just as statistics is not taught as a collection of formulae and ad hoc tricks. It is helpful to look at three high-level components of statistical computing: programming languages–environments and paradigms, algorithms and data structures, and data technologies.

Programming Languages, Environments and Paradigms

The vast majority of, if not all, statisticians would agree that students need to learn a programming language (American Statistical Association, ASA, n.d.; Bryce et al., 2001). There will be different opinions about what language(s) should be taught. Some will want to cover the practical aspects of using common types of statistical software or packages, such as SAS or SPSS. However, we believe that computing should be viewed as a supporting skill for statistical practice and research and that courses should cover the concepts of computing as well as the specifics. That is, the teaching of computing needs to be approached in the same way as the teaching of mathematics or statistics. Students need to be able to transfer the concepts to the specifics of other languages and environments as they change from, for example, R to MATLAB or from MATLAB to Perl. When taught in isolation, programming languages are idiosyncratic and arcane. When taught more generally, the commonality and patterns emerge and provide a significantly simpler viewpoint and much more useful, general skills that will serve them well in the future. Students should be made aware that (a) different languages serve different roles and (b) learning just one language is likely to be quite limiting in the future.

For these reasons, we advocate that students learn a general purpose programming language with which they can create new algorithms and functionality and express statistical ideas and computations at a relatively high level. Some students will need to learn lower level languages (e.g., C or FORTRAN), but most will be well equipped with languages such as MATLAB or R. Our experiences with code written for student research (graduate and undergraduate) have included both the need for more sophisticated algorithms and better understanding of fundamental programming concepts. Teaching algorithms that are subtly different or whose applicability is somewhat subtle, before improving the basic programming skills of students, would seem to be misplaced. A course in computational statistics—essentially how to do statistical computations properly—is more appropriate as a follow-up to an introductory course in programming languages and environments.

In addition to programming languages, graduate students will need to learn new paradigms, such as parallel and distributed computing. These are no longer exotic, specialized topics but commonly used techniques for implementing real scientific computations. Similarly, as statisticians increasingly publish software implementing their methodological research, graduate students need to understand some essential principles of software engineering. Issues of portability, memory management, object-oriented programming, compilation and efficiency, version control, unit testing, and so on are very important in developing software for others to use. What might have been considered advanced computing a decade ago is becoming more important for doctoral students so that they can successfully function in the scientific community.

Algorithms and Data Structures

If one were to ask academic statisticians what computing should be taught to statistics graduate students, many would list linear algebra decompositions, numerical optimization, and aspects of approximation and numerical analysis (Gentle, 2004; Lange, 2004; Monahan, 2004). Interestingly, none of those topics requires a computer; however, they are methods for obtaining solutions efficiently or approximately, or both, and often become necessary in advanced research. These are undoubtedly good things for students to know. However, when prioritizing the importance of this material relative to other statistical and computational topics, their importance is less clear. For the most part, students will not implement the general algorithms as the implementations available in well-tested, widely available software are efficient and robust. Teaching the circumstances under which each algorithm might be best is undoubtedly useful, but it is not necessarily limited to a computational course. Rather, it should be part of a theory or methodology course in which the need for optimization is raised (e.g., maximum likelihood estimation, robust regression).

If such topics are to be taught in a computational course, it is imperative that the students have the skills to be able to express computations so that they can quickly perform experiments to explore and understand the characteristics of these algorithms. For the most part, these are topics more appropriate for graduate students than undergraduates—and not all graduate students will need such skills early in their research—as these topics only make sense when the student has studied statistical methods that require such computational approaches. On the other hand, all students will need to be able to program, perform simulations, and access and manipulate data. While the choice of algorithm is often critical for developing efficient code, poor understanding of programming concepts is often the primary cause of inefficiency. Furthermore, human time is expensive relative to computer cycles; therefore, optimizing performance may be a waste of precious resources. These considerations imply an order and a priority for the different computational topics.

Classical computational statistics topics are undoubtedly of importance and, all else being equal, students should master them. Statisticians need to question this legacy and consider new topics and their importance. We wish to provoke thought about their importance relative to other potential topics for a computational statistics course. Simulation, computer experiments, Markov chain Monte Carlo (MCMC), the Expectation-Maximization (EM) algorithm, and resampling methods (bootstrap, cross validation) are of greater importance from a pedagogical perspective than matrix calculations and optimization algorithms—because they are less amenable to general purpose implementations and so do not exist as well-tested implementations in common software environments. Furthermore, since matrix calculations and optimization algorithms are extremely well implemented in widely available libraries and environments, students should not write their own versions of these highly tuned implementations.

Data Technologies

It is much easier to teach more algorithmic, mathematical material, such as the topics found in many computational statistics courses, than it is to teach topics in data technologies. For many, merely defining data technologies may prove difficult. Instead, think of these topics as new computational tools, techniques, and standards that allow access to data and functionality from varied sources and the presentation of information, results, and conclusions in rich, dynamic ways. These technologies include regular expressions for text manipulation, relational database management systems for storing and accessing data, the eXtensible Markup Language (XML) used for many purposes, Web services for distributed functionality and methods, Web publication and asynchronous documents, interactive and dynamic graphics, etc. In fields such as bioinformatics, finance, and astronomy, these are essential tools by which researchers access data. They are becoming important for statisticians who work in these fields, and a handful of statistics departments around the world are beginning to teach these topics. However, they are much less amenable to the definition–theorem–corollary–application style of teaching. They require instructors to think about teaching in a different manner; thus, it is necessary to rebuild much of the usual infrastructure for teaching.

We argue that, while difficult to teach, the topic of data technologies is growing in importance in the field. As statisticians deal with larger amounts of data from many and varied sources, often the challenges to data analysis start well before the computational steps involved in model fitting. Rather, simply accessing the large volume of data and getting it into the programming environment in a manageable way (e.g., from a relational database) can pose a problem. The choice of data structure and understanding when and when not to copy data are examples of issues that may be far more important and immediate hurdles. Further, rather than being concerned with potential inefficiencies in an algorithm, it is often more productive to use profiling tools to determine where lie the bottlenecks in the code. These profiling tools, data structures, and management of large datasets may well be more important than learning about efficient algorithms that are needed primarily for worst-case situations.

In addition to these three core topics in the statistics curriculum, one other topic deserves mention and consideration: visualization. Visualization clearly offers an invaluable tool, and computing plays an important role in modern visualization techniques for data analysis. Skills in visualization may well be the most valuable of all computing skills when considering the ubiquity of visual presentations of data and the great potential for communicating complex data structures simply with appropriate images. However, when promoting one set of computational topics over another, we must be quite specific about both the goals and the audience, and their interests.

What Are The Challenges To Making This Change?

There is little doubt that statistical ideas and concepts are important topics for students to learn. Mathematics and computing are supporting tools that aid learning these concepts and provide complementary approaches to this end. Ideally, both approaches are mastered. However, unlike mathematics, if one does not have computational skills, one simply cannot engage in the application–practice of statistics regardless of one's knowledge of the concepts. Computation is the currency of statistical action while mathematics is typically the currency of statistical description. Since most statistics students go on to apply statistics rather than study it academically, computational skills are vital.

At the graduate level, most statistics students have studied mathematics for at least six years, but have taken at most one course in computing and have no experience in statistical or scientific computing. While their mathematical skills may not be as strong as instructors would like, most students do not have a vocabulary for computation and often arrive with bad habits from point-and-click applications. Statistics departments have historically admitted doctoral students solely on the basis of their mathematical background. More balance in graduate curricula is needed so that students can leverage both mathematics and computation to understand and practice statistics and play a more active role in current and future developments.

We have heard many reasons or explanations why computing is not a larger part of the curricula. In our opinion, the primary reasons computing is omitted are that (a) it is difficult and time consuming to teach and or re-tool to teach, and (b) the discipline is conservative and clings to its mathematical past. It is useful to consider these explanations because some are legitimate points of view and obstacles to change.

"We don't know that material."

One explanation sometimes offered for not having computing in the statistics curricula is that statistics faculty were never taught computing, they have not had the opportunity to learn it, they cannot teach it effectively and so do not. This is very unfortunate as it means that new students do not have the opportunity to learn it either. At some point, statisticians need to break the cycle and learn this material. The situation is improving as some students are learning this material on their own, albeit in an ad hoc fashion and often incorrectly. With a willingness to change, the cycle can be broken.

"We send our students to Computer Science to learn computing."

This approach seems reasonable until one tries to determine more precisely what statistics students should learn and then map these needs to the courses available in computer science (CS). For the most part, CS courses are justifiably concerned more with abstract, theoretical aspects of computer science and technology than statistics students need to learn. That is, computing is an important means to an end for statisticians, not a study in its own right. For example, statisticians generally need not worry about the optimizations performed in the execution of a relational database query; instead, they should understand the general principles of the relational database model and the common elements of the Structured Query Language (SQL) used to extract data from a database. This material is easily covered in several classes;

statistics students do not need to spend an entire course on other, less important topics. As for programming concepts, we would argue that the emphasis and tone of general programming courses are not appropriate. Rather, high-level, interpreted languages that use vectorized operations and provide garbage collection, such as R, S-PLUS and MATLAB, are more appropriate. This is very different from the more traditional, object-oriented languages, such as C, C++, and Java, used in introductory CS courses.

This is not to suggest that there are no CS courses that are relevant to statistics students. However, they need to acquire fundamental scientific computing knowledge and skills that are the prerequisites to the more advanced topics. For database design, a CS course in more detailed database topics would be valuable. Understanding algorithms is a very important skill; therefore, a data structures and algorithms course would be useful. For disseminating statistical methods as software, a software engineering course is important. Scientific visualization is another course that is highly recommended.

"We let the students learn it on their own."

Many of our colleagues advocate—or at least practice—the approach in which students are told to learn about computing by themselves, from each other, or from their teaching assistant. This sends a strong signal that the material is not of intellectual importance relative to the material covered in lectures. With this approach, students pick up bad habits, misconceptions, and, more importantly, the wrong concepts. They learn just enough to get what they need done, but they do not learn the simple ways to do things or take the time to abstract what they have learned and assimilate these generalities. For many, they are unaware of the possibilities that surround them and so continue to do everything in the same, limited way. They cannot learn about new topics as they lack a basic vocabulary. Their initial knowledge shapes the way they think in the future and typically severely limits them, making some tasks impossible. They lack the ability to deal with new problems, and they typically lack the necessary confidence to approach new tasks. Their lack of computational skills makes it difficult for them to work in a team where others are computationally capable, independent, and autonomous. The curricula must provide computing fundamentals; we believe that adding a small amount of structure and guidance would yield large professional gains for students, research assistants, and professionals.

One of the rather ironic aspects of this approach is the relative paucity of material with which the students can learn. There are very few textbooks on general aspects of statistical computing or programming for statistics. In contrast, there are hundreds of textbooks on each statistical topic that instructors present in lectures. There seems to be an inversion in teaching, where students are left with few aids in computing while valuable contact time is spent repeating what they can read for themselves.

"Students only need to learn basic programming."

Some statisticians think that graduate students only need to learn MATLAB or R, others think just SAS is needed as it is widely used, while others think a language such C or Java is the right choice because it is the common language of scientific computing and is easily transferred. The fact that there is a difference of opinion and various options illustrates that there are differing goals and needs for statistics students. However, it is a big leap from a single course in basic computing to embracing problem-solving methodologies and general computing principles; the latter should be taught and then fostered by the culture of a modern, vibrant department that contributes to advances in statistical methodology and application. Additionally, to omit data technologies (e.g., relational databases, XML, Web services, distributed and parallel computing, etc.) is a disservice to those students having only basic computational literacy skills when they work with others from different fields who are vastly more skilled in the practicalities and advanced skills of working with data.

"Computing is not as important as our core statistical concepts."

While this may have once been true and may still be relevant to those with a very narrow view of statistics, the growth in data analysis in all sciences and the relative intractability of complex models and methods makes computational skills of immense importance in a modern view of statistics. The goal of statistics education is statistical concepts and thinking. Mathematics and computing are supporting tools that aid learning these statistical concepts; both must be mastered, not just the former. To function in the practice of statistics, one must be capable of increasingly complex computation. Since most students go on to practice statistics rather than study it academically, computational skills are vital.

How Do We Make It Feasible To Teach Computing?

It is an immense amount of work for an individual instructor to integrate computing into an existing statistics curriculum. First, the instructor must socialize the idea with colleagues and foster their support to add or change a course, and argue about which course or topics can be discarded to make way for this new material. This change can be met with resistance or apathy, which can herald the end of the process for all but those who feel sufficiently strongly about the new direction to persevere. Then there is the need to create a syllabus, debate what should and should not be included, and outline the topics on a week-by-week or lecture-by-lecture basis for the course. The process will typically involve multiple iterations. With all this done, the course may be submitted to a campus-wide committee for approval. After this, one must still convince other relevant individuals that the course should be scheduled and taught and not simply listed in the course catalog.

Having cleared the typically lengthy administrative hurdles, the instructor now has to teach the material, which can involve the following steps:

- Decide what the basic programming language will be, for example, R, S-PLUS, MATLAB, SAS, Perl or Python, C/C++.
- Be familiar with the topics at a level that goes at least slightly beyond what is being taught.
- Prepare exercises and longer projects, which involves identifying and evaluating the main topics to be covered and deciding how problems can be combined to reach the overall goal.
- For projects, find interesting datasets with an associated scientific or social problem of interest and then create a sequence of doable tasks that lead to the pedagogical goal, which typically means trying three or four datasets to find one that fits all the necessary criteria.

This collection of hurdles makes it apparent why statistical computing is not taught more. However, computing is too important to merely accept the difficulties and continue along the current, traditional path. We advocate pooling resources so that the materials needed to clear the administrative hurdles and teach the topics are available as templates that can be quickly adapted and customized for different situations.

To this end, we are creating, gathering, and disseminating materials to help faculty members initiate new courses or modules on computing within the statistics curricula (Hansen, Nolan, & Temple Lang, n.d.). To promote discussions among faculty members and assist the decision-making process, these materials include model or template syllabi—discussion documents that describe the different elements of statistical computing, why each element is important for different programs of study, and why the topic was selected. To aid in teaching, the materials include lecture notes, exercises, case studies, projects, tutorials, textbook chapters, and a textbook on data technologies. In today's Web-based world of information exchange, we no longer need to think in units of textbooks but smaller units that can be combined creatively for different courses. In addition to these materials, we have organized one-week workshops for faculty on how to teach statistical computing and established an electronic forum for discussing aspects of teaching computing in statistics. Essentially, we hope to seed the statistical community with resources for introducing and teaching computing by leveraging the existing, small community of those involved in statistical computing research.

This work is part of a three-year, U.S. National Science Foundation-funded grant from its Division of Undergraduate Education. We began with a workshop that brought together experts in statistical computing to discuss different topics and evaluate the areas to be taught. The wiki (Hansen et al., n.d.) includes findings from the workshop, syllabi from computing courses taught by workshop participants and others, and an annotated bibliography on statistics curriculum reform. In July 2008, we held a workshop for faculty members from around the U.S. who teach or plan to teach statistical computing. The participants are from departments with a commitment to introducing or continuing to develop their computing courses. The workshop covered the basic material and discussed different teaching approaches. Participants worked through case studies and projects, thinking about how to get students involved and be creative.

Building for the Future

More recently, the statistical community has seen the advent of several systems developed by and for the community. XLisp-Stat and Quail are systems that explore different paradigms for statistical computing and have significant results and merits. The S language and its two implementations—S-PLUS® (Insightful Corporation, n.d.) and R (R Development Core Team, 2006)—have been very important for the practice of statistics and also for the development of over a thousand add-on packages providing cutting-edge methodologies, often before they are published in journals. This new form of disseminating work is a terrific, modern change. As excellent as R and other systems are, they are aging, relative to the dramatic changes and innovations of both hardware and software from the engineering and information technology communities and relative to the ever-increasing size of datasets of interest. It should be clear that statistics as a field must continually innovate and build new systems and infrastructure to handle the challenges and new directions for statistics so as to remain relevant. The popularity and impact that R has had on the field should encourage us to put more resources into development, rather than hold the misguided belief that we have all we will need. The infrastructure must adapt to the changing needs of statisticians by importing innovations in general computing and technology and helping transform and shape statistics.

Statisticians cannot depend on commercial entities to develop the tools they need. Nor can they rely on research laboratories (such as Bell Labs, where S was created and developed) to produce the next generation of computational innovations. Traditional university statistics departments do not necessarily provide the stimulating, supportive environment that encourages faculty members to conduct research in computational infrastructure. (The University of Auckland, New Zealand, which is the home of R, is a rare exception.) This must change. While the field needs only a few bright students to focus on computational infrastructure, it definitely needs them to be educated with the fundamentals so that they can easily specialize in this work.

More than just developing the infrastructure needed for statistics itself, statisticians must aim to influence and guide some of the technological innovations that are underway. For example, access to self-describing data with rich metadata describing its content and origins are at the heart of the semantic web. Statisticians should be helping to incorporate ideas that could revolutionize access to data for statistical decision-making in this effort.

Statistical Experience

Most statisticians would agree that students need to know how to apply statistical methods, which is in many regards the goal of statistics education (ASA, n.d.; Bryce et al., 2001; Cobb & Moore, 1997). However, while many courses teach methodology—either the mathematics or the applied heuristics—very few focus primarily on teaching the skills of approaching a scientific problem from a statistical perspective. Instead, courses often focus on understanding methods and their characteristics in the belief that providing students with a set of tools and an understanding of those tools is the necessary background for using statistics in an applied setting. This is a rather big leap as there are so many other skills needed. For as Wild (2007, p. 325) noted, “the biggest holes in our educational fabric, limiting the ability of graduates to apply statistics, occur where methodology meets context (i.e. the real world).”

Typically, the applications in statistics courses are merely examples where students have to identify the inputs and plug them into the method. Examples do not have the same rich, complicated context, extraneous information, and decision-making issues as real applications. Examples focus on methodology and ignore the many other dimensions needed to apply statistical ideas to a real problem. At other times, the applications are derived from data collected from students and, again, typically lack a real question and context. Applications should involve uncovering the relevant information, understanding the needs of the problem, drawing conclusions and understanding their limitations, incorporating less quantitative considerations, refining the goal, and communicating the essential findings. These steps are far broader than estimating parameters or performing an F-test. On a more specific level, one needs to break the task into steps, figure out how to combine the steps, and then perform each step; this involves very different skills than using a particular statistical method or tool. The methodology is one important detail in the bigger picture, but it is just one, and too often students miss this important experience.

While statistics students learn the mathematical fundamentals of the field, students in other fields are often learning to apply more modern statistical methods than we even teach our doctoral students (e.g., Gelman & Hill, 2007; Sekhon, n.d.). To ensure that statistics students remain relevant, they must have skills that cannot be replaced by reading a textbook of modern methods. Statistical experience is such a skill; it gives students the skills needed to become collaborators in scientific endeavors. That is, with these skills they can work as part of a team and bring a particular way of thinking about a problem, along with a different set of tools that they have experienced in real situations.

For students learning data analysis and statistics, the intuition and experience that are used in data analysis are the hardest things to learn and the elements that are least often taught. Of course, it is also the hardest thing to teach, being somewhat subjective, context-specific, and an art. But we cannot shy away from this difficulty, and we can attempt to distill the more objective aspects. At the very least, students need to be exposed to the paths that are followed during a real analysis and understand the statistician's thinking and decision-making process.

When and Where Do Students Encounter the Experience Component?

Intuition and experience of methodology in the scientific context are essential to this thought pattern. Ironically, these are rarely presented in books—in sharp contrast to the large collection of textbooks that offer similar formal descriptions of common methods. Many statisticians would do well to adopt the pedagogical technique where students read material outside of class and the professor spends time on material not in the book. Wild and Pfannkuch (1999, p. 224) noted that to teach statistical thinking we often simply "let them do projects."

Although a valuable exercise, as a single, unguided encounter with statistical thinking in a real setting, it is far from adequate. Another approach is to leave the statistical experience until after they have learned the basic, traditional tools of statistics, such as probability, hypothesis testing, estimation, and inference. This might be in a capstone course for undergraduate majors. Again, this single exposure pales in comparison to opportunities appearing in multiple courses earlier in their studies. For graduate students, consulting courses in which clients bring statistical problems to a class have potential. However, the problems are of varied quality and interest and somewhat random in the lessons taught.

Another statistical experience teaching venue is its integration into an existing methodology or applied course via case studies (Brady & Allen, 2002). Case studies of data analyses often hide much of the thought process that is required. In a case study, an analysis is typically presented as a completed work, but the thought process that led to the conclusions and choice of approaches and methods is usually omitted. There is rarely mention of alternative approaches that were not fruitful or that were not pursued and the reasons why. Also not typically identified are the alternative approaches that led, or would lead, to almost equivalent solutions to the one presented in the final analysis.

Another option is for the practical experience to be inserted as tangents in the flow of a course that show how a particular method came to be used. Again, it is difficult in this scenario to get away from the use of a particular method (i.e., the one just learned) to the data at hand. However, it can be effective if it truly emerges from a scientific problem and includes contrasts with other approaches and methods that may be applicable (Nolan & Speed, 1999). It takes time to prepare and to teach, so one must decide if it is worth it. Unfortunately, we may be allowing our decision making to be clouded by limited time rather than the good of the students.

Some graduate programs require students to take a minor subject for a year in which they apply statistics to problems in that field. This seems to be successful, but these schools are usually quite forward looking and already have a broader view of statistics. Many students, both graduates and undergraduates, take a minor in statistics while majoring in another subject. Why is statistics typically the minor? One reason is that students believe they can make more of a difference in their work in other fields; this partly comes from being exposed to actual applications of statistics and appreciating its challenges and impact.

Yet another opportunity for displaying statistical thinking and imparting experience is in the introductory statistics course. Instead of assuming that this course is the only chance to teach students statistics and so must cover a long list of fundamentals, a more novel and potentially more effective approach would be to teach backwards. That is, rather than students learning methods as formulae and applying them to draw conclusions, compelling scientific and social problems are presented for students to grapple with, debate, and make decisions based on data exploration. With a well-guided discussion that pushes students to justify opinions and conclusions, they can discover and deduce commonsense statistical concepts and methods. We might do well to recognize that (a) there are students who may be interested in studying statistics but who do not know much about it, and (b) these students can be attracted to the field by showing them the bigger picture of how statistics is used and in what ways it is important. When students grapple with intellectually demanding questions and discover personal expression and creativity in the statistical experience, rather than the drier material of a traditional introductory statistics course, they may be attracted to the field or at least gain an appreciation for it.

The occasional course that presents different aspects of a broader view of statistics might get students thinking in new and interesting ways, and foster activity and innovation. Courses entitled *Weird Science* and *Disruptive Technologies* at the University of Texas, Austin are both thought provoking and engaging. Similar experiments could be tried in statistics.

What Should Be Taught?

While there is much variability in how statisticians operate, a statistician often approaches a consulting problem or scientific collaboration in ways that can be abstracted. From interviewing practitioners and researchers, Wild and Pfannkuch (1999) identified four dimensions of statistical thinking: the investigative cycle, types of thinking, the interrogative cycle, and dispositions. Their framework would complement the mathematical models used in analysis and address areas of the process of statistical investigation that the mathematical models do not, particularly areas requiring the synthesis of problem-contextual and statistical understanding.

We offer a concrete list of these aspects of statistical thinking that captures the elements of a typical data analysis process: decompose the problem and identify key components; abstract and formulate different strategies; connect the original problem to the statistical framework; choose and apply methods; reconcile the limitations of the solution; communicate findings. There are many nuances that we have omitted; and it is a subjective, informal process. Yet, the overlap between our list and the National Research Council (US NRC, 1996) science education standards is notable:

> *Inquiry is a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results.*

We consider each of these aspects in more detail.

Decompose the Problem and Identify Key Components

The statistician asks the scientist questions about the general subject matter to get the context of the problem and to understand the goals. As they interact, there is a discussion that iterates between what the scientist wants and what the statistician needs to know more generally about the problem. The conversation often revisits earlier topics to get more information that earlier did not seem necessary or occur to the discussants. Students need to learn how to identify this information. It is hard to mimic an interaction with a scientist, but making the information available from this process in the document-database and summer statistics program (described earlier and also later) is important.

Abstract and Formulate Different Strategies

As more details are uncovered, the statistician is collecting potential approaches, identifying potential problems or additional information needed, and formulating strategies by mapping the scientific problem to statistical approaches. This high-level work involves classification, prediction, or parameter estimation. As more information becomes available, particular techniques come into the picture (e.g., CART or *k*-nearest neighbors, linear model or GLM). There are many possible methods for each high-level statistical goal, and there may be several statistical goals that lead to an approach to the more general scientific problem. Exposing the dynamic picture the statistician builds as the investigation proceeds would be very valuable to students.

Connect the Original Problem to the Statistical Framework

A key component in this stage is understanding what is really of interest. If the goal is to make a decision about a particular social phenomenon, the statistician must understand the decision space, that is, for what outcomes is there the same decision and where is the boundary between the different decisions. These will ultimately be determined by the available data and the statistical methodology that is used. However, the boundary may be quite invariant with respect to the different statistical techniques being used; differentiating between various methods may not be of importance. In other cases, the sensitivity of the decision process may be quite extreme and require very careful understanding of the choices. The decision-making process is not determined entirely by the statistical approach but by the context in which the decision is being made. The particular method may not be important or the actual parameter estimates but rather their impact further up the decision-making chain. Further, before pursuing a statistical approach, it is important to understand the accuracy that is needed. For example, there can be a large difference between estimating a value to within 2 meters or to within 2 centimeters. Quite different data or a different approach may be required for the latter whereas any, even an ad hoc, method may suffice for the former.

Choose and Apply Methods

With the array of possible directions, the statistician will prioritize these realizing that some are easy, some are more complex, and some are long shots that may not lead anywhere. When mapping the problem to a statistical formulation, it is useful to consider if the data will be able to answer the problem in this setting. The statistician considers the limitations of the data and whether other data are needed and accessible. After choosing a statistical approach, the statistician performs the necessary computations to obtain the results. In so many ways, this is really the easy part of the process. Unfortunately, picking and fitting the model is the common focus of most courses.

Reconcile the Limitations of the Solution

With the statistical results in hand, the statistician puts them back into the decision-making context to evaluate whether a conclusion can be made and to determine the possible limitations of that conclusion. This information leads to an updated formulation of the problem so that the statistician can iterate through the entire process to get better results leading to a final conclusion.

Communicate Findings

Finally, the statistician communicates the results in a meaningful, interesting way. While the effectiveness might be related to the subject matter, the presentation needs to be engaging, clear, and context-specific in telling an important story.

What Are The Challenges?

Some reasons for the lack of change in this direction in the statistical curricula are: (a) the difficulty in finding good problems that are both compelling and at the right level for students. (b) conservatism in the field coupled with a deep-rooted belief that mathematics is the most important topic for students in spite of dramatic technological changes, (c) the effort in preparing course materials (i.e., providing all of the details of a compelling analysis) and concern that they will be out of date quickly, and (d) unfamiliarity with teaching from this approach. There is no question that it takes energy, time, and creativity to successfully convey the statistical experience. Models of how to begin to think about making the change are needed. We present two different models here. We have experimented with a summer program where undergraduates engage in research problems and statistical thinking through data visualization. The second is an approach where research papers are augmented by instructors to create interactive, dynamic documents that guide students in data analysis and statistical thinking.

The Summer Statistics Program

In June of both 2005 and 2006, the authors, along with Mark Hansen of the University of California, Los Angeles (UCLA), organized and conducted a one-week workshop at UCLA. Each year, about 25 undergraduate students, 5 graduate students, 3 researchers, and 3 additional statisticians participated. The undergraduates ranged in background from statistics majors to promising students who had taken only one statistics course but who were keen to see a broader side of statistics. In the first day, we taught essential computing skills thus enabling students to manipulate data, extract subsets, and create numerical and graphical displays.

The workshop format involved different research statisticians introducing a scientific problem with which they were active contributors. This typically involved motivating the problem and its importance, discussing the challenges (e.g., what data were needed and available), what outcomes were feasible, and how to start thinking about the problem from a statistical perspective. Each high-level problem was typically broken into about six steps over two days. The students first familiarized themselves with the data and explored it. Typically, there were some suggestions of aspects to explore; as the researchers moved between groups and focused discussions, new ideas emerged. After about an hour, each group presented something of interest, which the class discussed and critiqued. After this more free-form exploration, the researcher guided the students through a particular statistical formulation of the problem and explained how this would help lead to a result that could be used to address the problem. The remainder of the session oscillated between explaining some statistical techniques and sending the students off in groups or individually to use these techniques on the data to solve the problem. Again, there were several different approaches to explore and students either selected those that interested them or groups agreed to try different approaches for comparison purposes.

Having a discussion at the end of each break-out session ensured the students were engaged and acted as participants in the problem-solving activity. They exhibited a sense of involvement and creativity to find an unusual perspective or hidden feature—in contrast to the more typical, computational data-analysis course laboratories where students complete tasks or exercises in a prescribed manner. (We were coincidentally fortunate to have the opportunity to observe this more traditional approach within one morning of the same workshop, where the students' reactions were markedly less enthused.)

There have been several outcomes from this summer school. Firstly, it exposed students to a very different aspect of statistics than they had experienced in their courses. We found that the focus on real problems with statistical thinking, as opposed to learning about methods, interested the students and motivated them to learn more about the methods. Secondly, the experience confirmed for us that one can teach statistics in this manner to good effect. The students were able to quickly master sufficient computing skills so that they could then work relatively independently and be in charge of exploring their own creative ideas. We found that, while statisticians often teach as if the mathematics provides intuition for students, in these workshops students were able to rapidly grasp statistical concepts and intelligently apply methodology when described more heuristically. For some, a subsequent mathematical description helped to clarify the idea. Furthermore, they were able to suggest adaptations of the methods and were unencumbered by mathematical formalities and the sense of there being a unique, correct answer. A sense of creativity and a can-do attitude, even if erroneous, are desirable attributes of statistics students.

One very important outcome of the summer school is the case studies of the presentations in the form of an extended laboratory. We often use one or two as exercises or entire projects in our regular courses. This flow from researcher to instructor to student, where there are some mutually beneficial gains for the researchers and instructors, leads to very rich and somewhat unique teaching materials. However, there is no doubt that it is time consuming to gather this material. Two of the three organizers worked with each presenter before the workshop so as to understand the scientific problem, reconstruct the analysis and the computations in R, discuss how to decompose the topics for the students, and often provide higher level functions and preprocessed data to expedite analysis within the short period available. This process consumed several person-weeks; it would be greatly beneficial to reduce this time and to access more potential applications without burdening the researcher directly. This is one motivation for the database document concept described in the next section.

Reproducible, Dynamic, Interactive Documents

A second, novel avenue that the authors are pursuing in the area of teaching statistical experience is to provide infrastructure that induces a passive flow of research and case-study documents from researchers to educators. The vehicle is a *database document* or a reproducible, dynamic, interactive document. The essential idea is to enable authors to document their actual computations and analyses, along with notes and explanations of their thoughts and decisions, so that the analytical process is reproducible. Authors use the document as an electronic notebook that captures their writings, computations, thought process, and notes or comments. This document is not what is intended for the readers, but a database or repository of all the activities within the analysis or task. The document can be projected into different views to make articles, papers, stories, code, and software available for distinct audiences. Readers can switch between the projected view and access the details at various resolutions, from seeing data to the general computational flow to specific lines of code. One can replay all the computations up to a specific point or change inputs and recalculate the tables and figures.

This style of documenting one's work aids researchers by allowing them to archive material in a structured manner rather than the more personal style currently in use. Critically from a pedagogical perspective, it makes real analyses and applications of statistics available to educators in a manner that can be easily used for teaching students about this subtle, elusive process. The collection of such documents has vague similarities to open-source software, which has served the statistical community very well with R and S-PLUS, and software repositories at Comprehensive R Archive Network (CRAN) and Statlib (http://lib.stat.cmu.edu/). Here, the idea is to share details of analyses across discrete communities, allow analyses to be used for different purposes, encourage greater verification and understanding of results, facilitate further extensions of approaches, and enable students to observe and participate in the statistical experience.

Instructors can take such a document and know that it has all the details involved in a real analysis. They can annotate the material with links to explanations of the science and the statistical terms. They can annotate the computations (either programmatically or manually) to identify the inputs and outputs of the different subtasks. Such annotations can be used to display interactive controls for students who can then control various aspects of the computations—set nuisance parameters to different values, remove subsets of the data, introduce alternative datasets, create new plots, or introduce entirely different ways of analyzing data (e.g., using a different classification method in one step of the overall analysis). This is the interactive aspect of the document, which allows for student control via graphical user interface (GUI) elements rendered when displaying a projection of the original document. It provides a semiguided exploration of the details that can go on to delve deeper and eventually go to free-form analysis.

Our goal is for students to experience the thought process of the *masters* in context, seeing their choices, approaches, and explorations. We want to avoid simplifying the scientific–data problems; instead, we want to simplify how students see these details initially while allowing them to gradually see them to their full extent to experience the reality of statistical practice. As Wild (2007) noted, these documents give instructors a mechanism to:

- control complexity and other environmental settings
- provide multiple pathways to explore
- focus attention on what is new and accelerate through what has already been mastered
- allow students to efficiently "unlock the stories in data"
- encourage students to "just try things out"

This system (see Nolan & Temple Lang, 2007) is based on widely used and standardized technologies and frameworks; it readily supports multiple and different programming languages. Because it is highly extensible, it allows adaptation and will accommodate future developments (e.g., different aspects of the analysis process). The approach is to create a programming and authoring environment designed for professional statisticians that supports communication of statistical results and the data analysis process. The document created by the statistician would be both interactive and dynamic—dynamic in this case meaning that the code for the analysis and plots is contained in the document, and this code is run to create the view of the document. The document is interactive in that the reader can control the embedded computations by, for example, dragging a slider that leads to code reevaluation and subsequent update and redisplay of the output.

Our prototype is based upon the R computational environment. The document is a collection of text, code, output, and other metadata and is written in XML—XML syntax is similar to HTML, having elements or nodes of the hierarchical form. The XML document-database can be converted to a variety of formats, such as PDF, HTML, and what we call interactive HTML. R packages provide the functionality to transform and display these XML documents. The interactive controls are provided by a general purpose GUI toolkit called wxWidgets, which is also available from within R via a plug-in package. Information can be programmatically queried and extracted from the document database via R commands that identify the XML nodes of interest. While XML underlies the representation of the document, these documents can be authored without any knowledge of XML using tools such as Microsoft Word. However, the richness, flexibility, extensibility, and generality emerge from the XML infrastructure.

One might think that this is yet an additional burden on the author and so is unlikely to be adopted. We are more optimistic because essentially this archiving of the actual computations and noting of ideas, decisions, and thoughts is what is done more informally in every analysis. Statisticians store code used in the computations in separate directories and files, adding comments to LaTEX or Word documents as notes to themselves. At the very simplest, we are describing a system that facilitates such archiving and provides ways to retrieve and manage the elements, allows extraction of notes and code for other uses, and simplifies the creation of documents from the centralized master document—an important feature as XML and related technologies continue to dominate in publishing. The adoption of Sweave (Leisch, 2002) within R for dynamic documents illustrates that people are willing to use such tools. Our approach is a more general notion of a document with Sweave essentially as a special case. We provide a much richer concept of a document acting as a database rather than merely as a dynamic, presentation-based document mechanism. The concept of having many other dimensions within the document makes it much richer. However, not all documents are required to have these extra dimensions; they can be added after the document is first authored, which allows authors to gradually move from Sweave-like use to leveraging these extended facilities as appropriate. The use of standard, ubiquitous technologies makes it more broadly applicable across different communities and more amenable to interesting extensions.

SUMMARY

Computing and statistical thinking and experience are very important elements of a statistics education. To bring these elements into statistics curricula, statisticians must think boldly, unconstrained by legacy, starting from a blank slate and bringing back the best of the existing curricula along with new important topics. For example, instead of re-teaching the same concepts at progressively higher levels of mathematical abstraction, the time gained could be used to teach other topics, including computing, statistical experience, and modern statistical methods. Perhaps more importantly and ambitiously, once the entire curricula is evaluated from the viewpoint of what is no longer needed because the available computational power is so much greater, many topics can be streamlined or eliminated entirely. As Cobb (2007, p. 7) stated:

> *.. a huge chunk of statistical theory was developed in order to compute things, or approximate things, that were otherwise out of reach. .... The computer has offered to free us and our students from that, but our curriculum is at best in the early stages of accepting the offer.*

Teaching computing and statistical thinking is very hard. We have outlined various approaches that attempt to make it easier for individual instructors to introduce and teach this material within the statistics curriculum. Common to all of them is the notion of pooling resources across one or more communities. For computing courses, we are working to create model syllabi and documents to discuss the importance of different topics for different types of students. Also, we are working to create an archive for tutorials, chapters, case studies, course notes, videos for use within courses; and we are holding workshops for faculty on how to teach this material. The intent is to enlarge the community of instructors capable of and willing to teach statistical computing by leveraging the existing small community of those who already do.

To aid the teaching of statistical reasoning and experience, we aim to unite the research community and instructors by providing a flow of real-world data analyses from the former to the latter. This is done by providing an infrastructure for reproducible results for the researcher that allows the capture of computational details and thought process in an electronic notebook that acts as a project database. While this is beneficial to individual researchers and their community of fellow researchers and reviewers, it is also useful to course instructors. These documents allow students to enter the world of the researcher and to engage in the research process. Much remains to be done before this approach is complete and effective; software must be written, and communities must be engaged. However, the infrastructure is in place to achieve these ends.

The suggestions in this chapter represent more than incremental changes motivated by constrained resources and conservatism. Computing, the Web, the digital world and interdisciplinary science present a change point for the field of statistics and require statisticians to think about what a modern statistics curriculum would look like if they had both the freedom to change and resources to implement. For too long, the field of statistics has acted more passively to such change points and responded by merely adding topics to courses and not seeking, considering, or embracing new paradigms. For statistics to flourish in this new era of science and technology and to have the impact that it could and should, educators must seize the opportunity to move the field of studies towards the modern needs of scientific research with data.

ENDNOTE REFERENCES – DO NOT CHANGE ANYTHING HERE.

American Statistical Association. (n.d.). *Curriculum guidelines for undergraduate programs in statistical science*. Retrieved May 14, 2008, from http://www.amstat.org/education/index.cfm?fuseaction=Curriculum_Guidelines

Brady, J. E., & Allen, T. T. (2002). Case study based instruction of DOE and SPC. *The American Statistician, 56*(4), 312-315.

Bryce, G. R., Gould, R., Notz, W. I., & Peck, R. L. (2001). Curriculum guidelines for bachelor of science degrees in statistical science. *The American Statistician, 55*(1), 7-13.

Cobb, G. W. (2007). The introductory statistics course: A ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1). Retrieved from http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art1/

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*, 801-823.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. **CITY NEEDED -- IS THIS CAMBRIDGE, MA OR CAMBRIDGE, UK??**: Cambridge University Press.

Gentle, J. E. (2004). Courses in statistical computing and computational statistics. *The American Statistician, 58*(1), 2-5.

Hansen, M., Nolan, D., & Temple Lang, D. (2006, June 17-24). *Data visualization and its role in the practice of statistics* [second annual Undergraduate Summer Program in Statistics]. Retrieved May 14, 2008, from http://summer.stat.ucla.edu/

Hansen, M., Nolan, D., & Temple Lang, D. (n.d.). Retrieved May 14, 2008, from Computing in statistics: Model courses and curricula Wiki: http://www.stat.berkeley.edu/twiki/Workshop/CompCurric

Insightful Corporation. (n.d.). S-PLUS® 8 Enterprise Developer [Computer software]. Seattle, WA: Author. Available from http://www.insightful.com/

Lange, K. (2004). Computational statistics and optimization theory at UCLA. *The American Statistician, 58*(1), 9-11.

Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle & B. Rönz (Eds.), *Compstat 2002: Proceedings in computational statistics* (pp. 575-580). Heidelberg, Germany: Physika Verlag.

Monahan, J. (2004). Teaching statistical computing at North Carolina State University. *The American Statistician, 58*(1), 6-8.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123-165.

Nolan, D., & Speed, T. P. (1999). Teaching statistics theory through applications. *The American Statistician, 53*(4), 370-375.

Nolan, D., & Temple Lang, D. (2007). Dynamic, interactive documents for teaching statistical practice. *International Statistical Review, 75*(3), 295-321.

R Development Core Team. (2006). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rossman, A., & Chance, B. (2003). *Notes from the JSM 2003 session "Is the Math Stat course obsolete?"*, Available from American Statistical Association Section on Statistical Education Website: http://www.amstat.org/sections/educ/SFPanelFinal.doc

Sekhon, J. S. (n.d.). Quantitative methodology in the social sciences seminar (Political Science 239 course syllabus), Available from http://sekhon.berkeley.edu/seminar/syllabus_ps239.pdf

Society for Industrial and Applied Mathematics (SIAM) Working Group on Computational Science and Engineering Education. (2001). Graduate education in computational science and engineering. *SIAM Review, 43*(1), 163-177.

United States National Research Council. (1996). *The national science education standards*. Washington, DC: The National Academies Press.

Wild, C. J. (2007). Virtual environments and the acceleration of experiential learning. *International Statistical Review, 75*(3), 322-335.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 37*(3), 223-248.